Version 15

Essential Graphing

"The real voyage of discovery consists not in seeking new landscapes, but in having new eyes."

Marcel Proust

JMP, A Business Unit of SAS
SAS Campus Drive
Cary, NC 27513

15.0

The correct bibliographic citation for this manual is as follows: SAS Institute Inc. 2019. *JMP® 15 Essential Graphing*. Cary, NC: SAS Institute Inc.

JMP® 15 Essential Graphing

Copyright © 2019, SAS Institute Inc., Cary, NC, USA

All rights reserved. Produced in the United States of America.

U.S. Government License Rights; Restricted Rights: The Software and its documentation is commercial computer software developed at private expense and is provided with RESTRICTED RIGHTS to the United States Government. Use, duplication or disclosure of the Software by the United States Government is subject to the license terms of this Agreement pursuant to, as applicable, FAR 12.212, DFAR 227.7202-1(a), DFAR 227.7202-3(a) and DFAR 227.7202-4 and, to the extent required under U.S. federal law, the minimum restricted rights as set out in FAR 52.227-19 (DEC 2007). If FAR 52.227-19 is applicable, this provision serves as notice under clause (c) thereof and no other notice is required to be affixed to the Software or documentation. The Government's rights in Software and documentation shall be only those set forth in this Agreement.

SAS Institute Inc., SAS Campus Drive, Cary, North Carolina 27513-2414.

September 2019

SAS® and all other SAS Institute Inc. product or service names are registered trademarks or trademarks of SAS Institute Inc. in the USA and other countries. ® indicates USA registration.

Other brand and product names are trademarks of their respective companies.

SAS software may be provided with certain third-party software, including but not limited to open-source software, which is licensed under its applicable third-party software license agreement. For license information about third-party software distributed with SAS software, refer to http://support.sas.com/thirdpartylicenses.

Get the Most from JMP®

Whether you are a first-time or a long-time user, there is always something to learn about JMP.

Visit JMP.com to find the following:

- live and recorded webcasts about how to get started with JMP
- video demos and webcasts of new features and advanced techniques
- details on registering for JMP training
- schedules for seminars being held in your area
- success stories showing how others use JMP
- a blog with tips, tricks, and stories from JMP staff
- a forum to discuss JMP with other users

https://www.jmp.com/getstarted/

Contents

Essential Graphing

Chapter 1

Welcome to JMP

Learn about JMP

Documentation and Additional Resources

This chapter includes details about JMP documentation, such as book conventions, descriptions of each JMP document, the Help system, and where to find other support.

Contents

Formatting Conventions

The following conventions help you relate written material to information that you see on your screen:

- Sample data table names, column names, pathnames, filenames, file extensions, and folders appear in Helvetica (or sans-serif online) font.
- Code appears in `Lucida Sans Typewriter` (or monospace online) font.
- Code output appears in *`Lucida Sans Typewriter`* italic (or monospace italic online) font and is indented farther than the preceding code.
- **Helvetica bold** formatting (or bold sans-serif online) indicates items that you select to complete a task:
  - buttons
  - check boxes
  - commands
  - list names that are selectable
  - menus
  - options
  - tab names
  - text boxes
- The following items appear in italics:
  - words or phrases that are important or have definitions specific to JMP
  - book titles
  - variables
- Features that are for JMP Pro only are noted with the JMP Pro icon JMP PRO. For an overview of JMP Pro features, visit https://www.jmp.com/software/pro/.

Note: Special information and limitations appear within a Note.

Tip: Helpful information appears within a Tip.

JMP Help

JMP Help in the Help menu enables you to search for information about JMP features, statistical methods, and the JMP Scripting Language (or *JSL*). You can open JMP Help in several ways:

- Search and view JMP Help on Windows by selecting the **Help > JMP Help**.
- On Windows, press the F1 key to open the Help system in the default browser.
- Get help on a specific part of a data table or report window. Select the Help tool from the **Tools** menu and then click anywhere in a data table or report window to see the Help for that area.
- Within a JMP window, click the **Help** button.

Note: The JMP Help is available for users with Internet connections. Users without an Internet connection can search all books in a PDF file by selecting **Help > JMP Documentation Library**. See “JMP Documentation Library” on page 14 for more information.

JMP Documentation Library

The Help system content is also available in one PDF file called *JMP Documentation Library*. Select **Help > JMP Documentation Library** to open the file. If you prefer searching individual PDF files of each document in the JMP library, download the files from https://www.jmp.com/documentation.

The following table describes the purpose and content of each document in the JMP library.

Document Title	Document Purpose	Document Content
Discovering JMP	If you are not familiar with JMP, start here.	Introduces you to JMP and gets you started creating and analyzing data. Also learn how to share your results.
Using JMP	Learn about JMP data tables and how to perform basic operations.	Covers general JMP concepts and features that span across all of JMP, including importing data, modifying columns properties, sorting data, and connecting to SAS.

Document Title	Document Purpose	Document Content
Basic Analysis	Perform basic analysis using this document.	Describes these Analyze menu platforms: • Distribution • Fit Y by X • Tabulate • Text Explorer Covers how to perform bivariate, one-way ANOVA, and contingency analyses through Analyze > Fit Y by X. How to approximate sampling distributions using bootstrapping and how to perform parametric resampling with the Simulate platform are also included.
Essential Graphing	Find the ideal graph for your data.	Describes these Graph menu platforms: • Graph Builder • Scatterplot 3D • Contour Plot • Bubble Plot • Parallel Plot • Cell Plot • Scatterplot Matrix • Ternary Plot • Treemap • Chart • Overlay Plot The book also covers how to create background and custom maps.
Profilers	Learn how to use interactive profiling tools, which enable you to view cross-sections of any response surface.	Covers all profilers listed in the Graph menu. Analyzing noise factors is included along with running simulations using random inputs.

Document Title	Document Purpose	Document Content
Design of Experiments Guide	Learn how to design experiments and determine appropriate sample sizes.	Covers all topics in the DOE menu.
Fitting Linear Models	Learn about Fit Model platform and many of its personalities.	Describes these personalities, all available within the Analyze menu Fit Model platform: • Standard Least Squares • Stepwise • Generalized Regression • Mixed Model • MANOVA • Loglinear Variance • Nominal Logistic • Ordinal Logistic • Generalized Linear Model

Document Title	Document Purpose	Document Content
Predictive and Specialized Modeling	Learn about additional modeling techniques.	Describes these Analyze > Predictive Modeling menu platforms: • Neural • Partition • Bootstrap Forest • Boosted Tree • K Nearest Neighbors • Naive Bayes • Support Vector Machines • Model Comparison • Make Validation Column • Formula Depot Describes these Analyze > Specialized Modeling menu platforms: • Fit Curve • Nonlinear • Functional Data Explorer • Gaussian Process • Time Series • Matched Pairs Describes these Analyze > Screening menu platforms: • Modeling Utilities • Response Screening • Process Screening • Predictor Screening • Association Analysis • Process History Explorer

Document Title	Document Purpose	Document Content
Multivariate Methods	Read about techniques for analyzing several variables simultaneously.	Describes these Analyze > Multivariate Methods menu platforms: • Multivariate • Principal Components • Discriminant • Partial Least Squares • Multiple Correspondence Analysis • Structural Equation Models • Factor Analysis • Multidimensional Scaling • Item Analysis Describes these Analyze > Clustering menu platforms: • Hierarchical Cluster • K Means Cluster • Normal Mixtures • Latent Class Analysis • Cluster Variables
Quality and Process Methods	Read about tools for evaluating and improving processes.	Describes these Analyze > Quality and Process menu platforms: • Control Chart Builder and individual control charts • Measurement Systems Analysis • Variability / Attribute Gauge Charts • Process Capability • Model Driven Multivariate Control Chart • Pareto Plot • Diagram • Manage Spec Limits

Document Title	Document Purpose	Document Content
Reliability and Survival Methods	Learn to evaluate and improve reliability in a product or system and analyze survival data for people and products.	Describes these Analyze > Reliability and Survival menu platforms: • Life Distribution • Fit Life by X • Cumulative Damage • Recurrence Analysis • Degradation • Destructive Degradation • Reliability Forecast • Reliability Growth • Reliability Block Diagram • Repairable Systems Simulation • Survival • Fit Parametric Survival • Fit Proportional Hazards
Consumer Research	Learn about methods for studying consumer preferences and using that insight to create better products and services.	Describes these Analyze > Consumer Research menu platforms: • Categorical • Choice • MaxDiff • Uplift • Multiple Factor Analysis
Scripting Guide	Learn about taking advantage of the powerful JMP Scripting Language (JSL).	Covers a variety of topics, such as writing and debugging scripts, manipulating data tables, constructing display boxes, and creating JMP applications.
JSL Syntax Reference	Read about many JSL functions on functions and their arguments, and messages that you send to objects and display boxes.	Includes syntax, examples, and notes for JSL commands.

Additional Resources for Learning JMP

In addition to reading JMP help, you can also learn about JMP using the following resources:

- "Tutorials"
- "Sample Data Tables"
- "Learn about Statistical and JSL Terms"
- "Learn JMP Tips and Tricks"
- "Tooltips"
- "JMP User Community"
- "Free Online Statistical Thinking Course"
- "New User Welcome Kit"
- "Statistics Knowledge Portal"
- "JMP Training"
- "JMP Books by Users"
- "The JMP Starter Window"

Tutorials

You can access JMP tutorials by selecting **Help > Tutorials**. The first item on the **Tutorials** menu is **Tutorials Directory**. This opens a new window with all the tutorials grouped by category.

If you are not familiar with JMP, start with the **Beginners Tutorial**. It steps you through the JMP interface and explains the basics of using JMP.

The rest of the tutorials help you with specific aspects of JMP, such as designing an experiment and comparing a sample mean to a constant.

Sample Data Tables

All of the examples in the JMP documentation suite use sample data. Select **Help > Sample Data Library** to open the sample data directory.

To view an alphabetized list of sample data tables or view sample data within categories, select **Help > Sample Data**.

Sample data tables are installed in the following directory:

On Windows: C:\Program Files\SAS\JMP\15\Samples\Data

On macOS: \Library\Application Support\JMP\15\Samples\Data

In JMP Pro, sample data is installed in the JMPPRO (rather than JMP) directory.

To view examples using sample data, select **Help > Sample Data** and navigate to the Teaching Resources section. To learn more about the teaching resources, visit https://jmp.com/tools.

Learn about Statistical and JSL Terms

The **Help** menu contains the following indexes:

Statistics Index Provides definitions of statistical terms.

Scripting Index Lets you search for information about JSL functions, objects, and display boxes. You can also edit and run sample scripts from the Scripting Index and get help on the commands.

Learn JMP Tips and Tricks

When you first start JMP, you see the Tip of the Day window. This window provides tips for using JMP.

To turn off the Tip of the Day, clear the **Show tips at startup** check box. To view it again, select **Help > Tip of the Day**. Or, you can turn it off using the Preferences window.

Tooltips

JMP provides descriptive tooltips (or *hover labels*) when you place your cursor over items, such as the following:

- Menu or toolbar options
- Labels in graphs
- Text results in the report window (move your cursor in a circle to reveal)
- Files or windows in the Home Window
- Code in the Script Editor

Tip: On Windows, you can hide tooltips in the JMP Preferences. Select **File > Preferences > General** and then deselect **Show menu tips**. This option is not available on macOS.

JMP User Community

The JMP User Community provides a range of options to help you learn more about JMP and connect with other JMP users. The learning library of one-page guides, tutorials, and demos is a good place to start. And you can continue your education by registering for a variety of JMP training courses.

Other resources include a discussion forum, sample data and script file exchange, webcasts, and social networking groups.

To access JMP resources on the website, select **Help > JMP User Community** or visit https://community.jmp.com/.

Free Online Statistical Thinking Course

Learn practical statistical skills in this free online course on topics such as exploratory data analysis, quality methods, and correlation and regression. The course consists of short videos, demonstrations, exercises, and more. Visit https://www.jmp.com/statisticalthinking.

New User Welcome Kit

The New User Welcome Kit is designed to help you quickly get comfortable with the basics of JMP. You'll complete its thirty short demo videos and activities, build your confidence in using the software, and connect with the largest online community of JMP users in the world. Visit https://www.jmp.com/welcome.

Statistics Knowledge Portal

The Statistics Knowledge Portal combines concise statistical explanations with illuminating examples and graphics to help visitors establish a firm foundation upon which to build statistical skills. Visit https://www.jmp.com/skp.

JMP Training

SAS offers training on a variety of topics led by a seasoned team of JMP experts. Public courses, live web courses, and on-site courses are available. You might also choose the online e-learning subscription to learn at your convenience. Visit https://www.jmp.com/training.

JMP Books by Users

Additional books about using JMP that are written by JMP users are available on the JMP website. Visit https://www.jmp.com/books.

The JMP Starter Window

The JMP Starter window is a good place to begin if you are not familiar with JMP or data analysis. Options are categorized and described, and you launch them by clicking a button. The JMP Starter window covers many of the options found in the Analyze, Graph, Tables, and File menus. The window also lists JMP Pro features and platforms.

- To open the JMP Starter window, select **View** (**Window** on macOS) **> JMP Starter**.
- To display the JMP Starter automatically when you open JMP on Windows, select **File > Preferences > General**, and then select **JMP Starter** from the Initial JMP Window list. On macOS, select **JMP > Preferences > Initial JMP Starter Window**.

Technical Support

JMP technical support is provided by statisticians and engineers educated in SAS and JMP, many of whom have graduate degrees in statistics or other technical disciplines.

Many technical support options are provided at https://www.jmp.com/support, including the technical support phone number.

Chapter 2

Introduction to Interactive Graphing

Overview of Data Visualization

Essential Graphing describes all of the different graphs and elements you can use to visualize your data:

- Graph Builder interactively creates many different types of graphs. See Chapter 3, “Graph Builder” and Chapter 4, “Graph Builder Examples”.
- Bubble Plot creates a scatterplot that represents its points as circles, or bubbles. Bubble plots can be dynamic (animated over time) or static (fixed bubbles that do not move). See Chapter 5, “Bubble Plots”.
- Scatterplot Matrix shows an ordered collection of bivariate graphs. See Chapter 6, “Scatterplot Matrix”.
- Parallel Plot draws connected line segments that represent each row in a data table. See Chapter 7, “Parallel Plots”.
- Cell Plot draws a rectangular array of cells where each cell corresponds to a data table entry. See Chapter 8, “Cell Plots”.
- 3D Scatterplot shows the values of numeric columns in the associated data table in a rotatable, three-dimensional view. See Chapter 9, “Scatterplot 3D”.
- Contour Plot constructs contours of a response in a rectangular coordinate system. See Chapter 10, “Contour Plots”.
- Ternary Plot display the distribution and variability of three-part compositional data. See Chapter 11, “Ternary Plots”.
- Maps can be used in Graph Builder, but also in other platforms, as background maps. See Chapter 12, “Maps”.

Chapter **3**

Graph Builder

Create and Explore Graphs

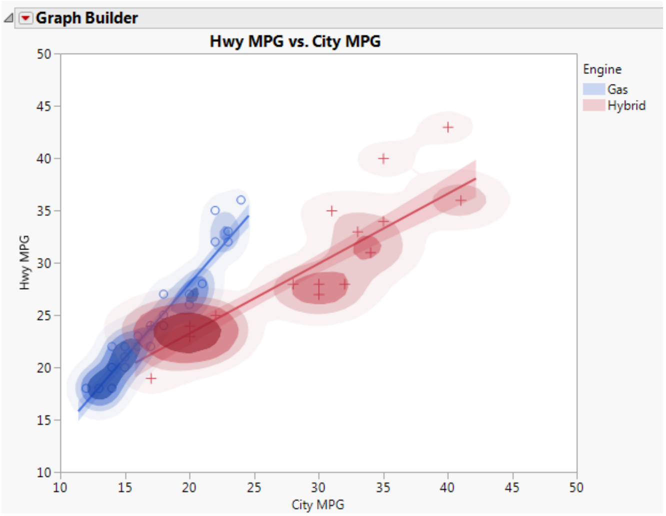
Use Graph Builder to explore multidimensional relationships easily and flexibility. You can quickly create and modify plots using Graph Builder's interactive interface. Select the variables that you want to graph and drag and drop them into zones.

Use Graph Builder to do the following:

- Graph data in a multitude of ways, including line charts, bar charts, histograms, maps, contour plots, and many more.
- Experiment by changing the graph type with the click of a button.
- Compare groups with overlays or trellis layouts.
- Examine and illustrate relationships between several variables of different modeling types.
- Use color to highlight aspects of your data.

This chapter shows you how to use Graph Builder to work with plots. For detailed examples using Graph Builder, see the Graph Builder Examples chapter.

Figure 3.1 Example Using Graph Builder



Contents

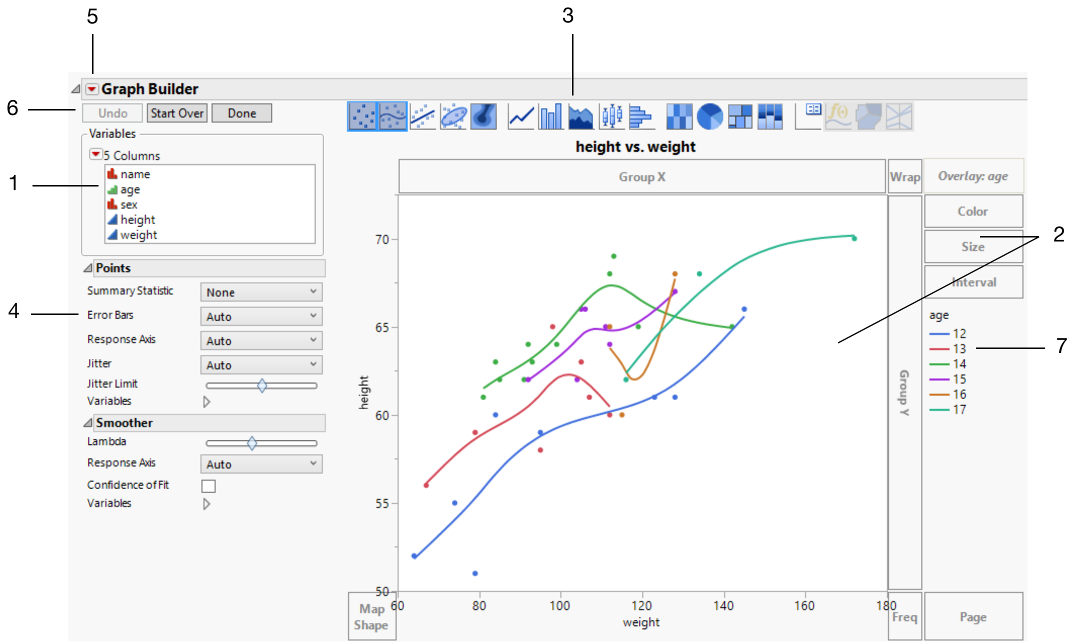
How to Use Graph Builder

1. Open the data table containing the data that you want to graph.
2. Select **Graph** > **Graph Builder**.
3. Drag columns from the Variables list into zones. See “Move or Remove Variables in Zones” on page 39 and “Graph Zones” on page 99.
4. Click element type icons to choose different types of graphs or elements. See “Element Types and Options” on page 52.
5. (Optional) Customize the selected element types. See “Element Types and Options” on page 52.
6. (Optional) Customize the legend. See “Legend Options” on page 108.
7. When you’re satisfied with the graph, click **Done**.

About the Graph Builder Window

To launch Graph Builder, open your data table and then select **Graph > Graph Builder**. The Variables list contains the columns in your data table, which you can drag into zones.

Figure 3.2 Graph Builder Window for Big Class.jmp



<table>
<tr><th>Number</th><th>Description</th></tr>
<tr><td>1</td><td> The columns in your data table are listed in the Variables panel. Drag columns from the Columns list into zones, or select the column in the Variables list and click the empty zone. You can create a virtual column by right-clicking a column name in the Variables list and selecting from the options provided. The transformed column appears at the bottom of the list in italics.
You can drag that column into Graph Builder zones.
Right-click the transformed column for further options. Tip: For descriptions of the column options in the Variables red triangle menu, see the Get Started chapter in Using JMP. </td></tr>
<tr><td>2</td><td> The graph area and zones are the main body of Graph Builder. Your graph appears in the middle, and you drag variables into different zones. See “Graph Zones” on page 99. Note: Any excluded rows are hidden in graphs. </td></tr>
<tr><td>3</td><td>Click an element type icon to choose the type of graph that you want. You can apply several types of elements to your data simultaneously (using the Shift key). See “Element Types and Options” on page 52.</td></tr>
<tr><td>4</td><td>Customize the graph with specific options for the type of graph you have selected. See “Element Types and Options” on page 52.</td></tr>
<tr><td>5</td><td>Customize the Graph Builder window. See “Red Triangle Options” on page 103.</td></tr>
<tr><td>6</td><td>Undo actions, start over, and more. See “Graph Builder Buttons” on page 105.</td></tr>
<tr><td>7</td><td>If your graph has a legend, it appears here. To customize the legend, see “Legend Options” on page 108.</td></tr>
</table>

To re-create and modify the graph in Figure 3.2, follow these steps:

1. Select **Help** > **Sample Data Library** and open Big Class.jmp.
2. Select **Graph** > **Graph Builder**.
3. Select height and drag it to the **Y** zone.
4. Select weight and drag it to the **X** zone.
5. Select age and drag it to the **Overlay** zone.

The graph in Figure 3.2 appears. The relationship between height and weight for each value of age is shown using color in a single graph.

Now you want to see separate graphs for each value of age.

6. Select age in the **Overlay** zone and drag it to the **Wrap** zone.

Figure 3.3 Height and Weight Grouped by Age

For each value of age, the relationship between height and weight is shown in a separate graph.

Example of Features in Graph Builder

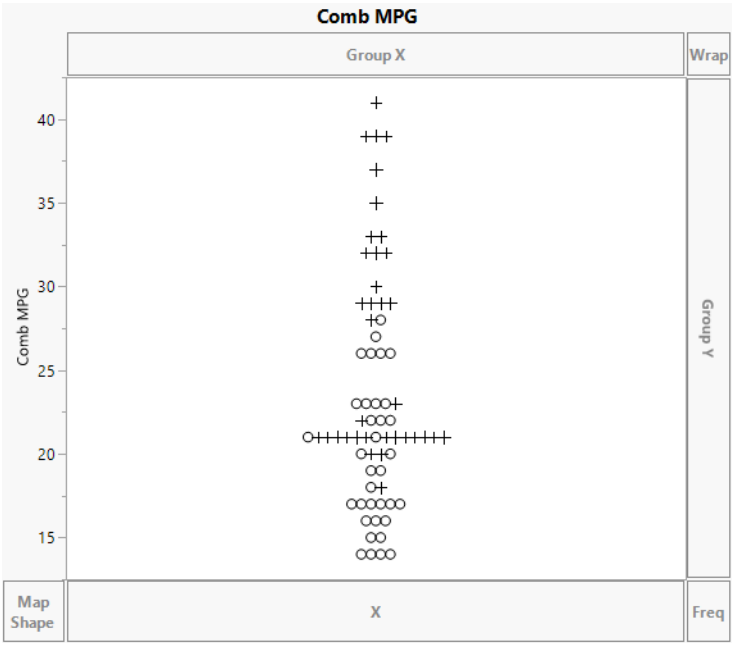
The following example gets you started using some of the features in Graph Builder. This example uses fuel economy data collected for hybrid and non-hybrid cars. You want to get an understanding of miles per gallon (MPG), a measure of fuel economy, and to see which factors might be influencing MPG. This example shows you how to plot several factors of interest:

- "Get an Overall Picture of Combined Miles per Gallon" on page 33
- "Plot Mileage Stratified by Engine and Driving Type" on page 34
- "Find the Relationship between Hwy and City MPG by Engine Type" on page 37
- "Find the Relationship between Hwy and City MPG by Engine Type" on page 37

Get an Overall Picture of Combined Miles per Gallon

1. Select **Help > Sample Data Library** and open Hybrid Fuel Economy.jmp.
2. Select **Graph > Graph Builder**.
3. Select Comb MPG (combined miles per gallon) and drag it to the **Y** zone.

Figure 3.4 Initial Graph of Combined MPG



Because markers are assigned to rows in the data table, points are represented by those markers. The value of Comb MPG for each row is plotted at its value on the vertical axis. To avoid over-plotting points that correspond to observations with the same Comb MPG values, JMP automatically jitters the points.

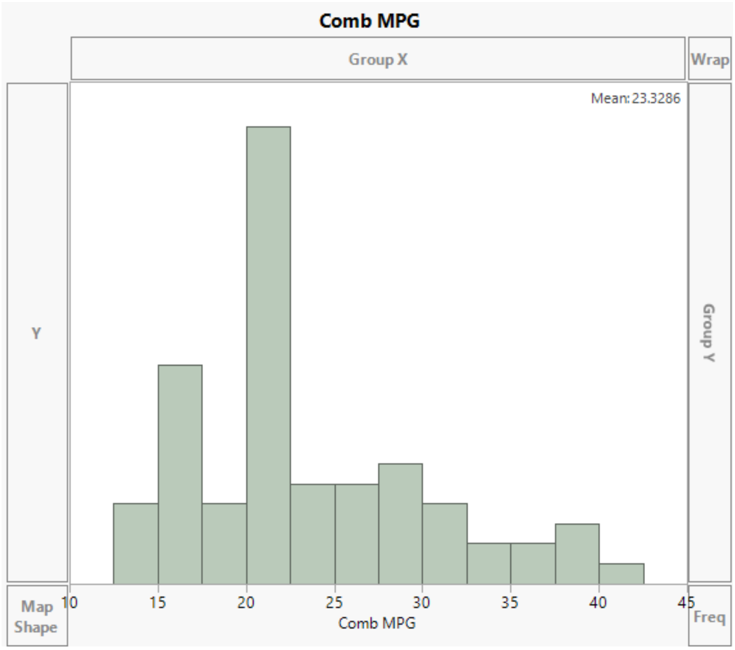
You can get a cleaner picture of the distribution by plotting a histogram.

4. Click the Histogram icon .

   You want the histogram to be vertical, so you need to move Comb MPG to the X zone.

5. Place your cursor over Comb MPG in the Y axis. The cursor turns into a hand. Select and drag Comb MPG into the **X** zone.
6. Click the Caption Box icon  to show the mean Comb MPG on the graph.

Figure 3.5 Histogram of Combined MPG



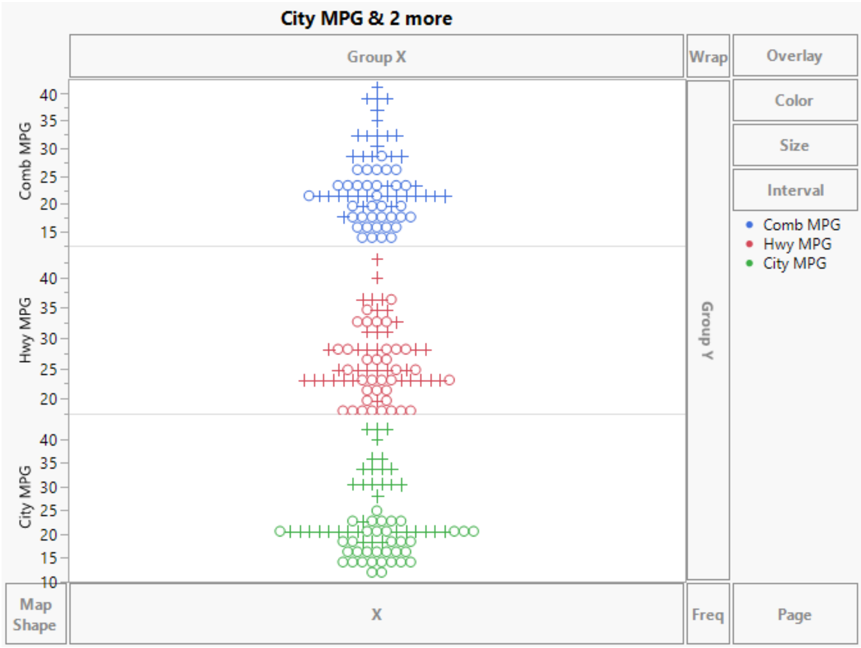
The histogram shows that Comb MPG can range widely, from about 10 to 45. However, a lot of vehicles get mileage between 15 and 30. The mean MPG is 23.3286.

Plot Mileage Stratified by Engine and Driving Type

1. Click **Start Over**.
2. Select City MPG, Hwy MPG, and Comb MPG.
3. Place your cursor over the Y zone with your cursor, press Shift, and click in the Y zone.

   This creates separate Y axes for City MPG, Hwy MPG, and Comb MPG.

Figure 3.6 MPG Variables with Separate Y Zone Axes

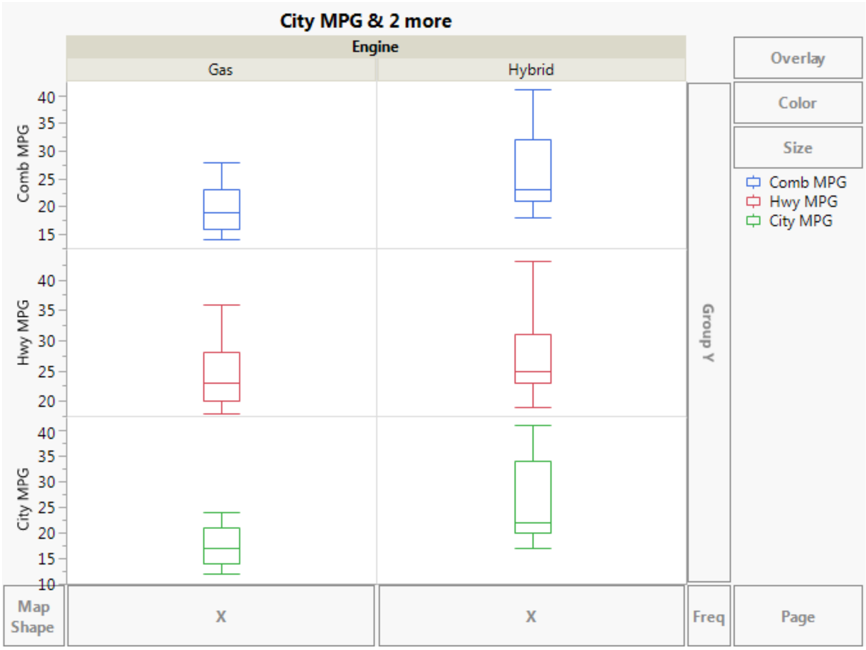


In the legend at right, markers are colored according to the driving type.

Tip: To change the color assigned to a driving type, right-click the marker in the legend.

4. Drag Engine to the Group X zone.
5. Click the Box Plot icon .

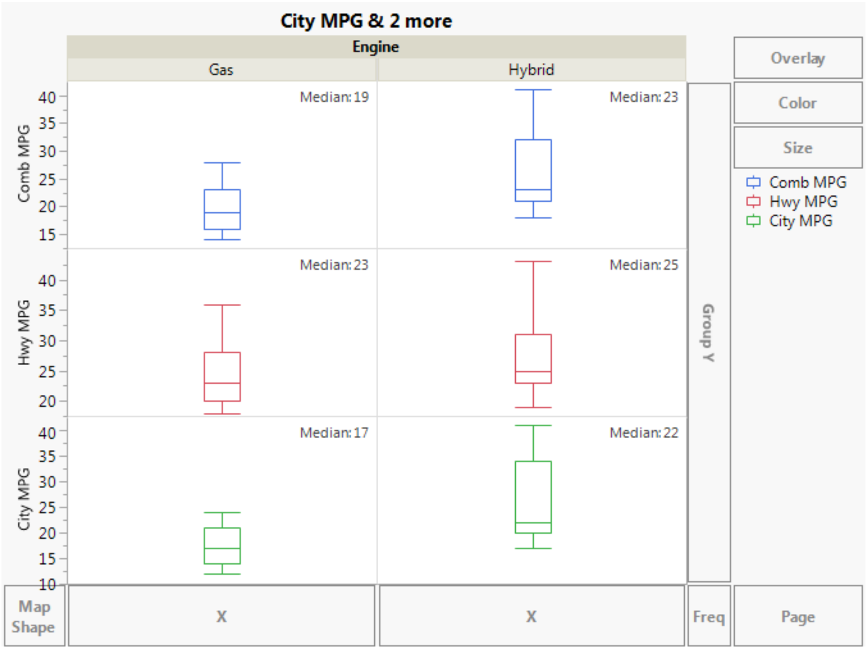
Figure 3.7 Boxplots of MPG Variables by Engine Type



The boxplots show the distributions in a very compact form. You can see that all three types of MPG values are substantially lower for the gas vehicles than for the hybrids in the study. From the horizontal lines in the centers of the two City MPG boxplots, you can estimate that the median difference is about five miles per gallon. However, you can use the Caption element to be sure.

6. Click the Caption Box icon .
7. In the Caption Box options at the left of the plot, select **Median** from the Summary Statistic list for Comb MPG, Hwy MPG, and City MPG.

Figure 3.8 Boxplots of MPG Variables with Captions

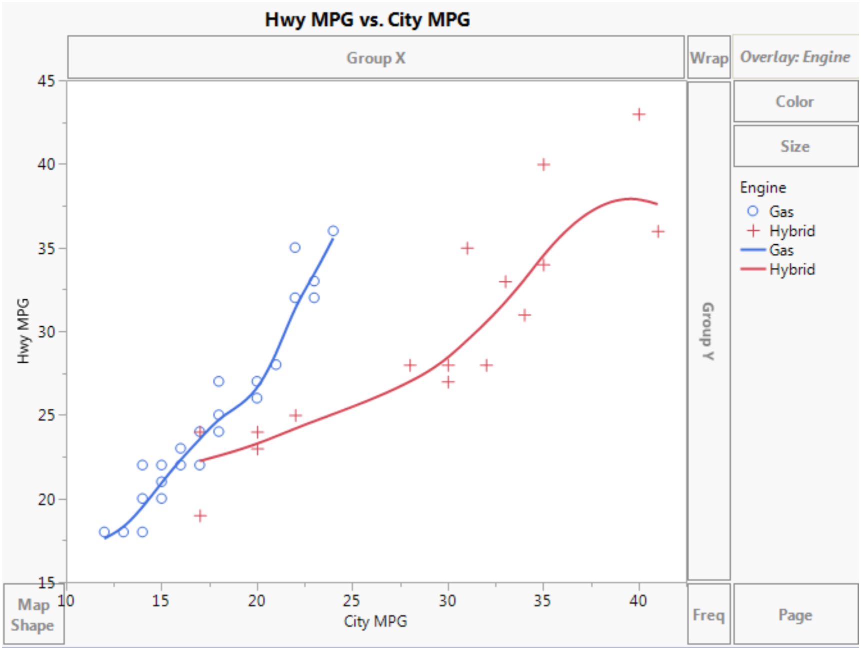


The median city MPG for gas engines is 17 and for hybrid engines is 22.

Find the Relationship between Hwy and City MPG by Engine Type

1. Click **Start Over**.
2. Select City MPG and drag it into the **X** zone.
3. Select Hwy MPG and drag it into the **Y** zone.
4. Select Engine and drag it into the **Overlay** zone.

Figure 3.9 Hwy and City MPG by Engine Type



A Smoother appears for each type of engine, Gas, and Hybrid.

Tip: To change the line properties (color, width, and so on), right-click a line in the legend.

5. Click the Line of Fit icon .

   For each engine type, the smoother is removed and a least squares line, together with confidence bands for the predicted mean, is added.

Tip: To add the y = x line, right-click in the graph and select **Customize** > **+** > **Templates** > **Y Function**. Replace `_function_of_x_` with `y = x` and click **OK**.

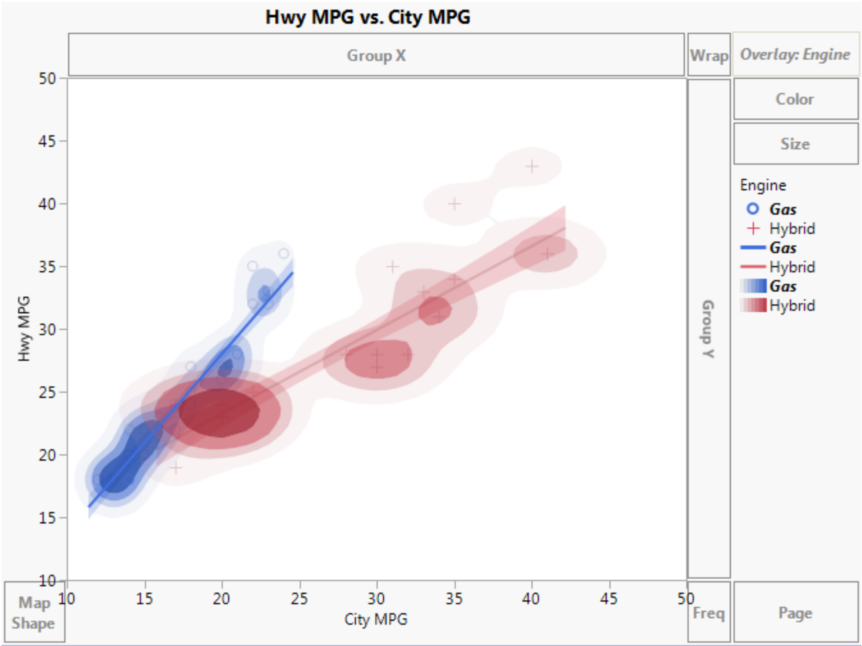
6. Drag a rectangle around the “o” marker with the smallest value of City MPG and check the number of rows selected in the Rows panel of the data table.

   You see that two rows are represented by this marker. You can check that over-plotting happens for other markers as well. When there is over-plotting of observations, namely when several observations are represented by a single point on a plot, density contour plots can help you see the density of points.

7. Drag the Contour icon  into the graph.

   Dragging the Contour icon into the graph retains the Line of Fit for each engine type.

Figure 3.10 Density Contour for Hwy and City MPG by Engine Type



The contours indicate that for both engine types, there are more vehicles in the study with lower MPG values than higher MPG values.

Move or Remove Variables in Zones

There are several ways that you can move a variable into a zone:

- To move a variable from one zone to another, click and drag it from the old zone into the new zone.
- To replace a variable in a zone, drag the new variable over the old one.
- To switch a variable from one zone to another, right-click the variable in the zone that you want to move and select **Swap**. Then, select the variable that you want to switch places with.
- To merge variables on the same axis, see "Merge Variables on a Common Axis" on page 40.

To remove a variable, drag it into empty space, or right-click the name of the variable in the zone and select **Remove**.

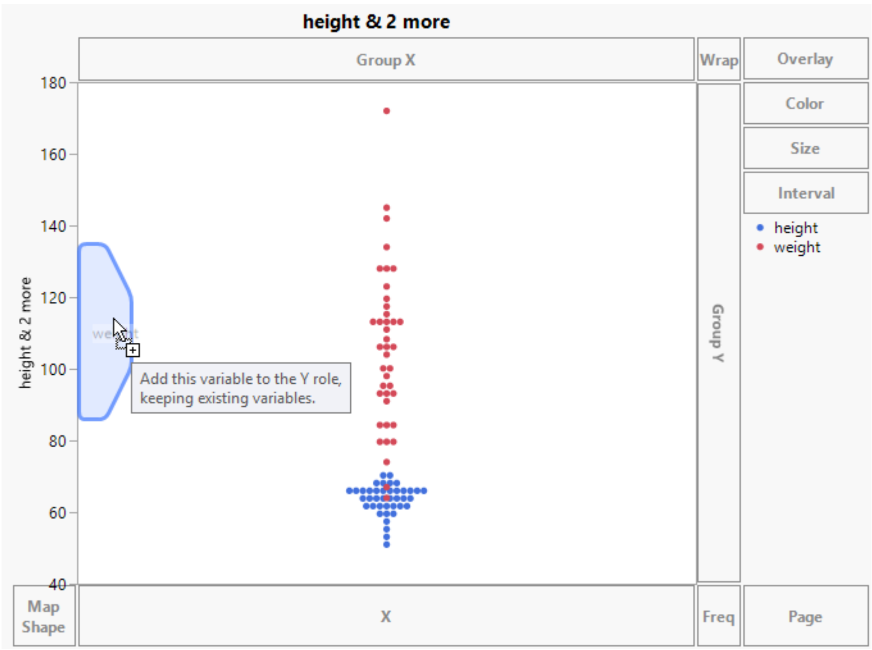
Work with Axes

Merge Variables on a Common Axis

Merging variables places both variables on the same axis and creates a single graph for both variables. Elements in the plot are colored by the variables, and a legend appears to the right of the plot. When points are plotted, they are randomly jittered for each value or category defined by the combination of the axes.

- If you are starting with an empty zone, highlight multiple variables from the list and drag them into the zone at the same time.
- If you already have variables in the zone and want to add more, drag the new variables to the inner side of the zone, in the graph area, until a symmetric polygonal shape appears.

Figure 3.11 New Variable Merged with Existing Variable



Results Based on Data and Modeling Type

- If only variables with continuous modeling types are merged, their values are plotted against a common axis.
- If variables with categorical (nominal or ordinal) modeling types are merged with a continuous variable, the values of all variables are plotted against a common axis.

  Before you can merge a variable with the character data type with numeric variables, the levels of the character variable have to be assigned numeric values. Integer values are assigned based on the Value Order column property or the default value ordering. These integer values range from 0 (highest level in value ordering) to the number of levels minus one (lowest value in value ordering). These integer values are then plotted against the common axis.
- If only variables with categorical modeling types are merged, then separate, *nested*, axes are constructed for each variable. See "Create Nested Axes for Character Variables" on page 45.

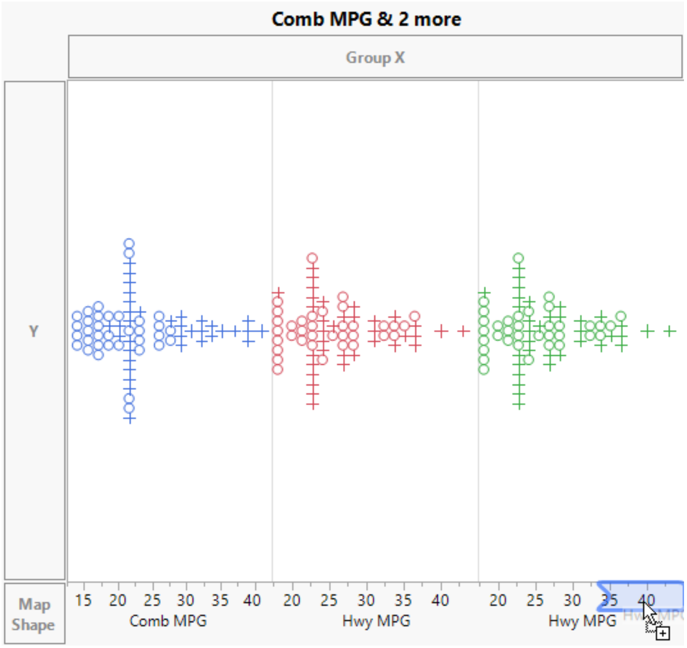
Create Separate Axes in the X and Y Zones

To quickly plot multiple variables against separate axes in the X zone:

1. Select the columns in the Variables list.
2. Click **Shift** and then click the **X** zone.

If you want more control over where the variables appear, drag them individually. Drag additional variables above, below, or between existing variables in the X zone.

Figure 3.12 Dragging a Variable to the Right of the Existing Variable



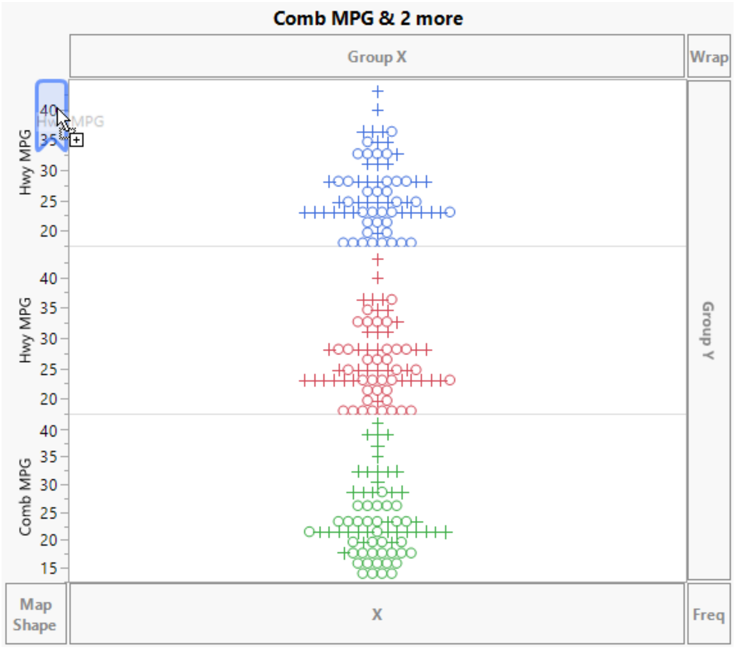
A single polygonal shape indicates when your variable is in the appropriate target zone.

To quickly plot multiple variables against separate axes in the Y zone:

1. Select the columns in the Variables list.
2. Click **Shift** and then click the **Y** zone.

If you want more control over where the variables appear, drag them individually. Drag additional variables above, below, or between existing variables in the Y zone.

Figure 3.13 Dragging a Variable above the Existing Variable



A single polygonal shape indicates when your variable is in the appropriate target zone.

Create a Second Y Axis

If you have two or more Y variables on the same axis, you can reflect the scaling of a second set of variables by creating a second Y axis. This can be useful when two variables measure the same underlying quantities, but have different scales. In general, it is unwise to use a second Y axis in any other situation. See Stephen Few's paper on Dual-Scaled Axes in Graphs.

To create a second Y axis:

1. Right-click one of the Y variable names in the Y zone and select **Move Right**.
2. Select the variable or variables that you want to move to the new axis.

   The new axis is scaled according to the values of the specified variable, and the selected variable is plotted against this axis.
3. Repeat the process to plot additional variables against the new axis.

   The new axis adjusts to accommodate the values of the additional variables.

Change the Graph for a Second Y Axis

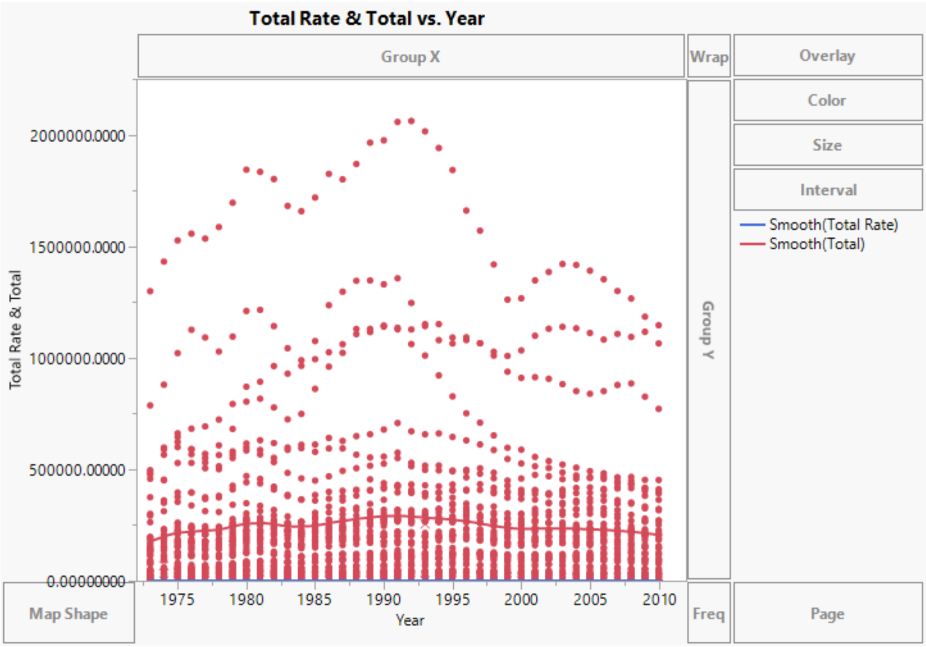
When you click an element type, it applies only to the variables on the left Y axis. To apply an element to the variables on the right Y axis, right-click and change the second element, which corresponds to the right axis.

Example of Creating a Second Y Axis

The CrimeData.jmp sample data table contains data on various types of crime for the 50 US states yearly, from 1973 to 2010. The Total column sums the number of incidents, and the Total Rate column gives a population-adjusted rate. You want to show these two variables on a single chart.

1. Select **Help > Sample Data Library** and open CrimeData.jmp.
2. Select **Graph > Graph Builder**.
3. Select Total Rate and Total and drag them to the Y zone.
4. Select Year and drag it to the X zone.

Figure 3.14 Total Rate and Total Merged

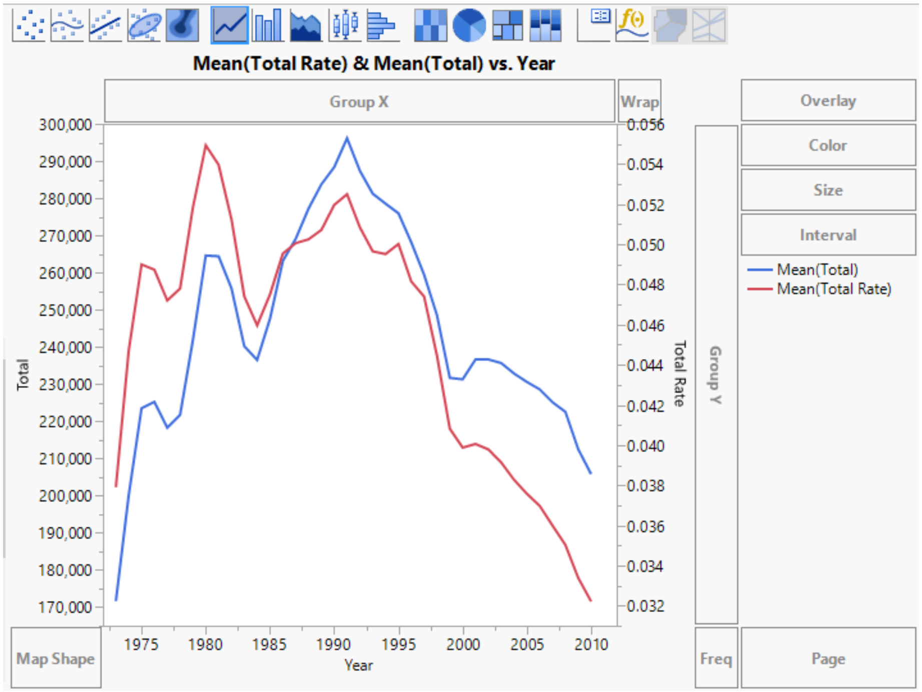


The Total Rate values are all between 0 and 1, and they are barely visible. Calculate yearly means and then use a second Y axis to make these values visible.

5. In the options panel for Points, select **Mean** next to **Summary Statistic**.
6. Click the Line element .

7. Right-click the variable names in the Y zone and select **Move Right** > **Total Rate**.

Figure 3.15 Second Y Axis Added



An axis for Total Rate is added on the right, and the axes rescale to show both sets of values. You can now see relationships between the two measures. For example, both measures began to decrease in 1991.

8. (Optional) Click **Done**.

Create Nested Axes for Character Variables

If you merge variables that all have categorical modeling types, then an individual axis is constructed for each variable. The outermost axis corresponds to the first variable selected, the next to the second, and so on.

To nest the axes in a specific order, drag the variables to the zone individually. First, drag the variable for the innermost axis to the zone. Then drag the variable for the next axis to the outside of the preceding variable and drop the variable once a trapezoid shape appears.

The following example illustrates nested axes.

After changes are made to improve a measurement process, a measurement systems analysis study is conducted to study repeatability and reproducibility with Part, Operator, and Instrument as factors. Each of three operators measures each of eight parts with four instruments. Of particular interest is the consistency of the instruments. The measured quantity is called new Y.

1. Select **Help > Sample Data Library** and open Variability Data/3 Factors Crossed.jmp.
2. Select **Graph > Graph Builder**.
3. Select new Y and drag it to the **Y** zone.
4. Select Part and drag it to the X zone.

   The plot shows variation in the values that are measured for each of the eight parts. There are systematic differences among the parts, which is to be expected. For example, measurements for parts 7 and 8 are lower than those for parts 1 through 6.

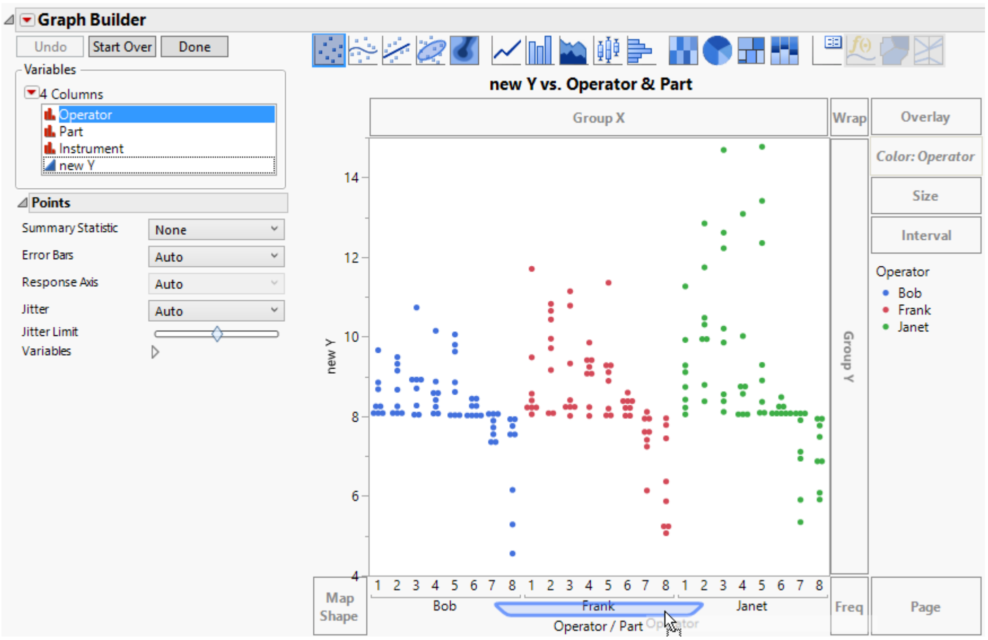
5. Drag Operator to the **Color** zone.

   The new Y values are colored by Operator, using the legend at the right of the graph. It appears that Janet might be measuring higher values than the other two operators on most parts. But the Operator effect is not easy to visualize, so you create a separate Operator axis.

6. Select Operator and drag it beneath Part in the X zone.

   The label Operator/Part appears, indicating that Part is associated with the topmost axis, and Operator is associated with the lower axis.

Figure 3.16 Drag Operator to Add a Second Axis

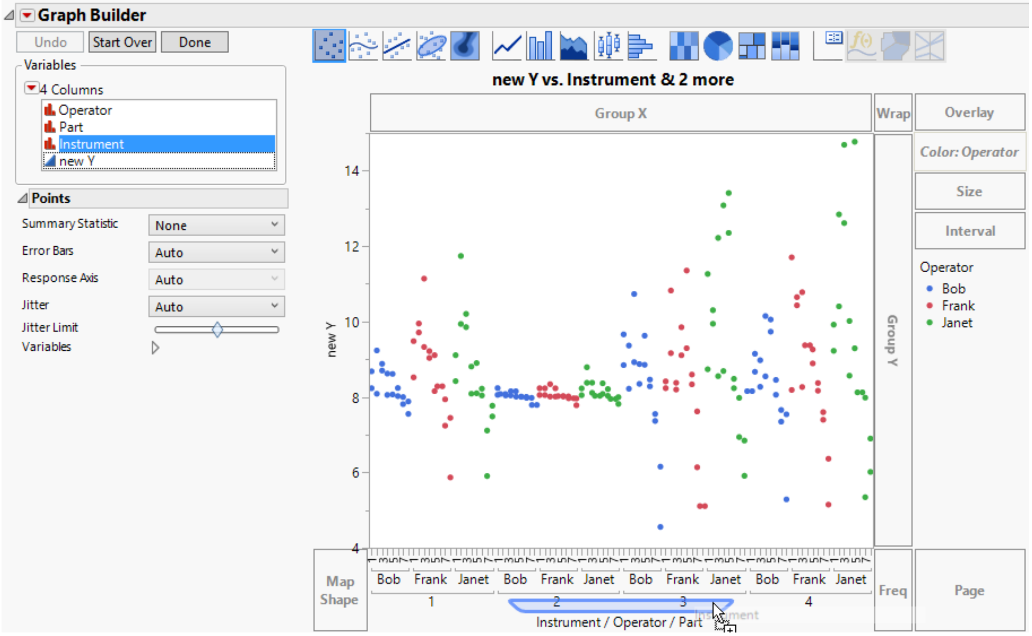


Now it is easier to see that Janet tends to obtain higher measurements for the same parts than do Bob and Frank. But what about the effect of Instrument?

7. Select Instrument and drag it beneath Operator/Part in the X zone.

   The label Instrument/Operator/Part appears, indicating that a third axis for Instrument has been added beneath the Operator axis.

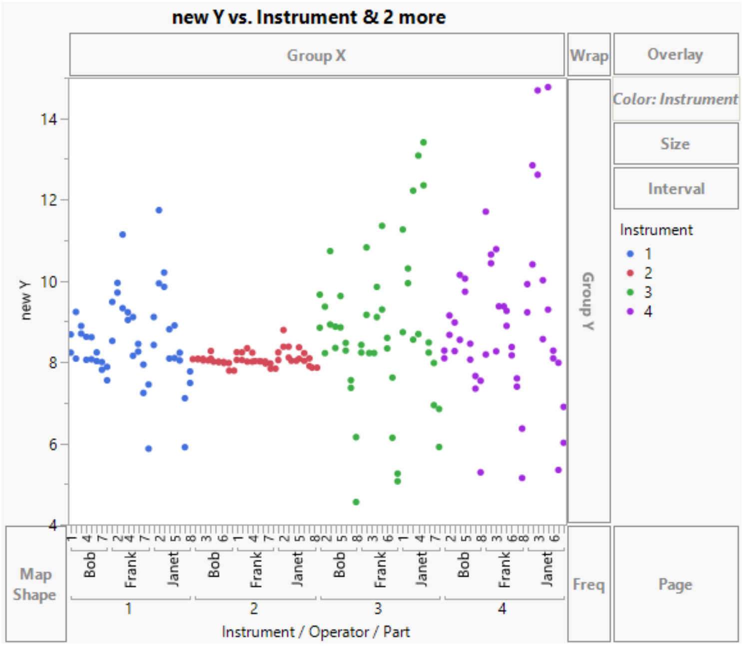
Figure 3.17 Drag Instrument to Add a Third Axis



It is clear that Instrument 2 leads to much more consistent measurements than the other three instruments. For Instrument 2, there is comparatively little variation between or within operators.

8. Select Instrument and drag it to the **Color** zone.

Figure 3.18 Three Nested Axes



Now the new Y values are colored by Instrument, and it is easy to see Instrument differences.

By nesting the axes for the three factors in study, you are able to obtain a visual understanding of the variation attributable to the factors.

Order the Levels of a Categorical Variable

JMP orders the levels of a categorical variable according to the following rules:

- Numeric, nominal data are sorted numerically.
  - White space around numbers is compared: “vt 1” is sorted before “vt1”.
- Character data that are only digits (numbers) are sorted numerically.
- Character data are sorted alphabetically, with the following exceptions:
  - Months and days of the week are in chronological order.
  - Ratings are sorted from low to high:

    Low to high: Very Low, Low, Medium Low, Medium, Medium High, High, Very High

    Agreement: Strongly Disagree, Disagree, Neutral, Indifferent, Agree, Strongly Agree

    Bad or good: Failing, Unacceptable, Very Poor, Poor, Bad, Acceptable, Average, Good, Better, Very Good, Excellent, Best

- Character data that have a character prefix and a numeric suffix are sorted first by prefix and then by suffix. For example, lots 1 through 12 are ordered “lot1”, “lot2”, “lot3”,...”lot10”, “lot11”, and “lot12”.
- If the numeric suffix starts with a “0”, it does not follow numeric ordering by suffix. (“lot1” is ordered *after* “lot02”).

Note: The preceding rules apply only to graphs with categorical axes.

You can also order the levels of a categorical (nominal or ordinal) variable on axes using the Value Order column property, which takes precedence over all other ordering rules. To change the ordering, you can do one of the following:

- Use the Value Order column property to ensure that values are ordered as you intend. For more information about the Value Order column property, see The Column Info Window chapter in *Using JMP*.
- Use the values of a numeric variable already in the graph. In Graph Builder, right-click the categorical axis and select **Order By**.
- Use the values of an arbitrary numeric variable. In Graph Builder, drag the numeric variable next to the categorical variable in the zone until a trapezoid appears and then drop the variable. The axis label changes to <categorical variable> ordered by <numeric variable>. See “Example of Ordering the Levels of a Categorical Variable Using a Numeric Variable” on page 49.

To change the order of the levels or the statistic, proceed as follows:

1. Right-click the axis label and select **Order By**.
2. Select one of the ascending or descending options.

   You can order by a variable’s Order Statistic (these options are prefixed by the variable’s name) or by the number of observations in each level of the categorical variable (Count).
3. The default ordering statistic is the mean. To use another statistic, right-click the axis label and select **Order Statistic**.

Note: If you try to order the values of a numeric variable using another numeric variable, JMP merges the variables. See “Merge Variables on a Common Axis” on page 40.

Example of Ordering the Levels of a Categorical Variable Using a Numeric Variable

To order a nominal or ordinal variable by a numeric variable, consider data about vehicle types. You want to see the vehicle types arranged in a meaningful order.

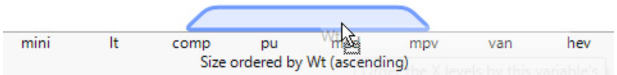
1. Select **Help > Sample Data Library** and open Cars.jmp.
2. Select **Graph > Graph Builder**.

3. Select Size and drag it into the **X** zone.

   This variable represents the type of the vehicle. Eight levels are listed alphabetically on the X axis: compact (comp), heavy (hev), lightweight (lt), medium (med), mini, multi-purpose (mpv), pick-up truck (pu), and van. Since the levels are listed alphabetically, they are not ordered in a meaningful way. For example, heavy comes before mini and lightweight. You want to order the levels by Wt (weight).

4. Select Wt and drag it to the middle of the X axis. Drop it just above the **X** axis. Before you drop the variable, a blue quadrilateral appears.

Figure 3.19 Merging Wt and Size

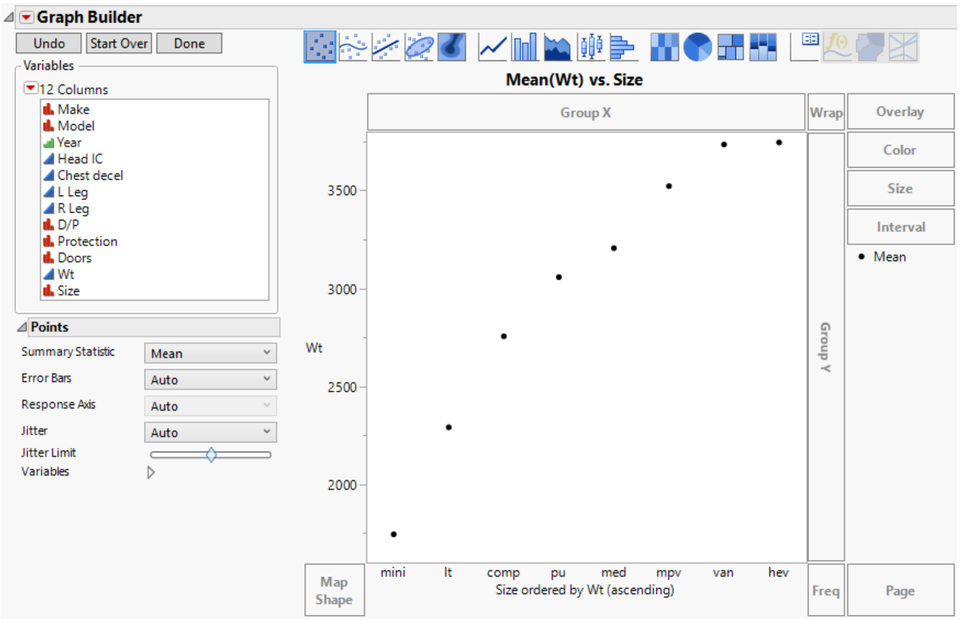


The levels of Size are now arranged in increasing order according to the average Wt of all vehicles in the levels. Notice that mini and lt (lightweight) are now ordered before hev (heavy). The axis label is updated, signifying that an ordering variable is in use.

Next, verify that Size is actually ordered by Wt.

5. Select Wt and drag it to the **Y** zone.
6. Select **Mean** from the Summary Statistic list.

Figure 3.20 Example of Size Ordered by Wt, Ascending



You can see that the average Wt increases from left to right.

Change the order from ascending to descending.

7. Right-click in the **X** zone and select **Order By** > **Wt, descending**.

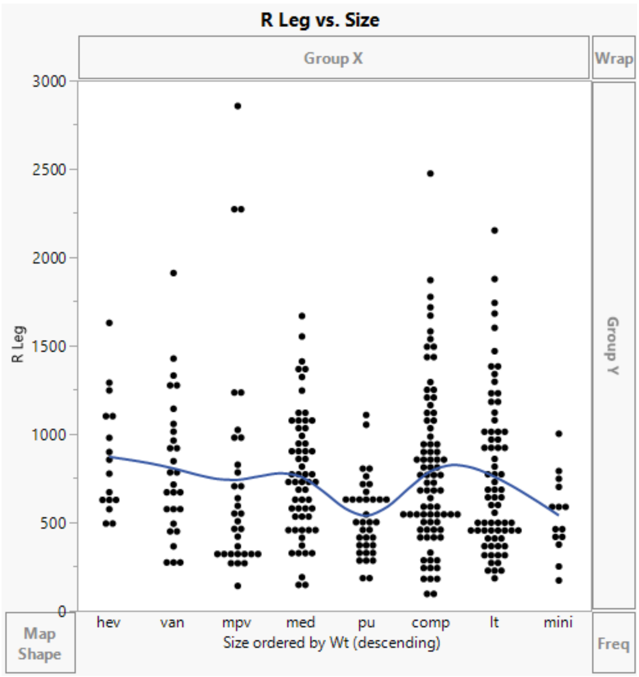
   Now the levels of Size are arranged in decreasing order of mean Wt. You want to see whether the right femur load (R Leg) decreases with vehicle weight.

8. Right-click Wt in the Y zone and select Remove.
9. Select R Leg and drag it to the Y zone.
10. Select **None** from the Summary Statistic list.

    This selection replaces the means with points for all the individual observations.

11. Click the Smoother element  .

Figure 3.21 Example of R Leg Ordered by Wt, Descending



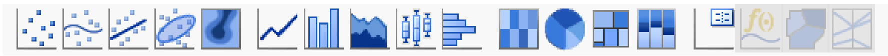
R Leg seems unrelated to the weight classes.

The default ordering statistic is the mean. To use another statistic, right-click in the **X** zone, select **Order Statistic** and change it to the statistic that you want.

Element Types and Options

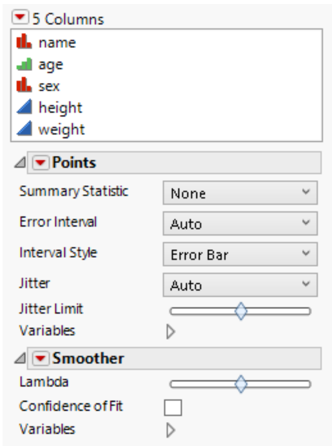
When you enter variables into zones, element types that are not applicable are not available and appear dimmed. You can change how your data appear by clicking on an element type icon.

Figure 3.22 Element Type Icons



You can specify and change properties for each element type in the Properties area under the Variables panel. For every element type that you apply in a graph, an outline whose name is the element type appears in the Properties area.

Figure 3.23 Example of the Properties Areas for Points and Smoother

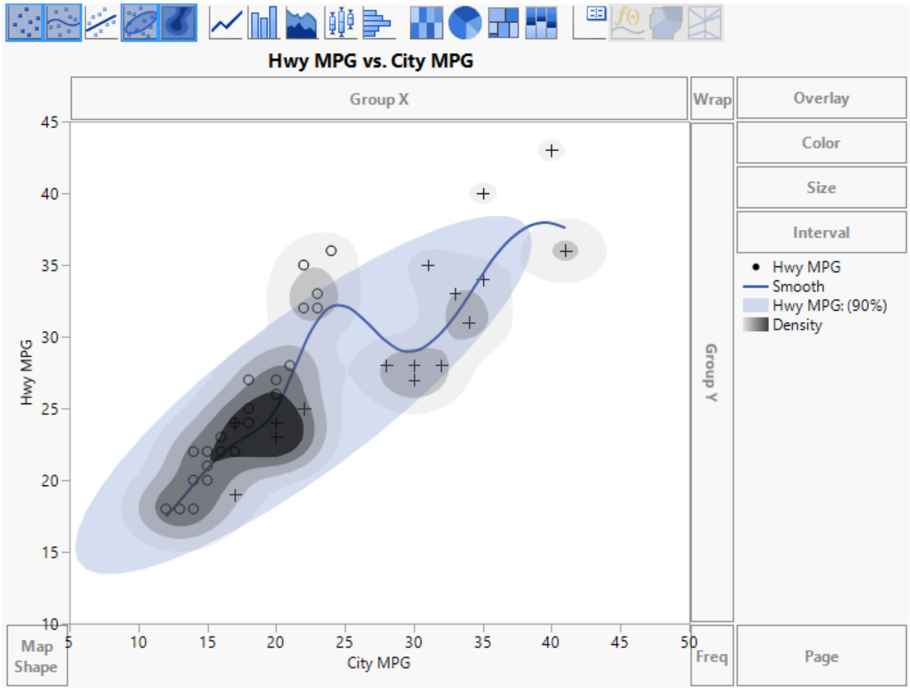


You can also right-click in any graph to change the element, or customize elements in the graph.

Select Multiple Elements at Once

To select and overlay multiple elements, press Shift and click the elements. Alternatively, you can drag the elements into the graph one at a time.

Figure 3.24 Graph Showing Points, Smoother, Ellipse, and Contour Elements



Labels for Points

In most graphs that display points, when you place your cursor over a point, a label appears showing information about the corresponding row. Place your cursor over the label and right-click for more options. You can pin the label to the graph, copy the contents, or close to remove the label. You can also pin the label by clicking the pin icon in the upper right of the label. See the JMP Reports chapter in *Using JMP*.

Hidden but Unexcluded Rows

As a general rule, when you hide rows but do not exclude them, elements in the plot that are in one-to-one correspondence with the hidden rows are then hidden in the plot. When elements in the plot are *not* in one-to-one correspondence with rows that are hidden but not excluded, the effect on the plot depends on the element type. For more information about a graph element, see the specific element's section.

The following example illustrates how plot elements that are not in one-to-one correspondence with rows might be entirely hidden when one constituent row is hidden.

1. Select **Help > Sample Data Library** and open Big Class.jmp.
2. Select **Graph** > **Graph Builder**.
3. Select age and drag it to the **X** zone.

4. Go back to **Big Class.jmp**, right-click row 8, and select **Hide/Unhide**.

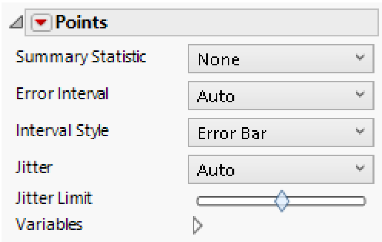
   In Graph Builder, the corresponding point disappears.
5. In Big Class.jmp, select **Rows** > **Clear Row States**.
6. In Graph Builder, click and drag the Bar icon into the plot.
7. In Big Class.jmp, right-click row 8, and select **Hide/Unhide**.

   In Graph Builder, the entire bar disappears.

Points

The **Points** element shows data values as points.

Figure 3.25 Points Options



Summary Statistic Changes the statistic being plotted. The statistic is calculated using the variable on the Response Axis. None is the default setting, indicating that the data values themselves are plotted.

Error Interval Adds or removes specified error interval in the graph.

Interval Style Draws error bars or error bands for the selected error interval.

Jitter Jitter adds random noise to values to reduce over-plotting. Choose from the following types of jitter:

None No adjustments are made.

Auto Adds various types of jitter when categorical variables are involved. No jitter is added when you have only continuous variables.

Random Uniform Random offset with uniform distribution.

Random Normal Random offset with Gaussian distribution.

Packed Places markers tightly to preserve any non-jittered dimensions.

Centered Grid Similar to Packed, but adjusts the non-jittered dimensions to fall into a grid.

Positive Grid Similar to Center Grid, but in the positive direction.

You can also customize jitter by right-clicking and selecting **Customize** > **Marker**.

Tip: To create a dot plot, select the Positive Grid option. You can move the dot plot to start at the bottom by changing the Y-axis values to 0-1. You can also resize the markers to a larger value, such as 20, to adjust the binning of the dots.

Jitter Limit Controls the spread or amount of overlap.

Variables Shows or hides graph elements for variables, or re-orders the display of variables.

Note: These options do not apply to variables in the Group X, Group Y, Wrap, or Page zones.

Check boxes are followed by the zone designation and the name of the variable. Use check boxes to do the following:

– Show or hide the elements corresponding to a variable in a zone.

– Add or remove the effect of applying the Color, Size, Shape, or Freq variable to the variable in the zone.

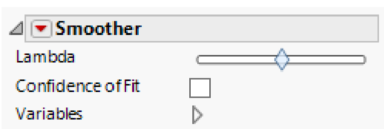
Tip: If you have multiple graphs, you can color or size each graph by different variables. Drag a second variable to the **Color** or **Size** zone, and drop it in a corner. In the Variables option, select the specific color or size variable to apply to each graph.

Use arrows to re-order the display if there are multiple variables in a zone. Highlight a variable name and click an arrow to reposition it.

Smoother

The **Smoother** element  shows a smooth curve through the data. The smoother is a cubic spline with a default lambda of 0.05 and standardized *X* values. You can adjust the degree of smoothness by changing the value of lambda using the slider.

Figure 3.26 Smoother Options



Lambda Change the default lambda of 0.05 using the slider. For more information about lambda, see the Bivariate chapter in *Basic Analysis*.

Confidence of Fit Shows the bootstrap confidence region for each fit.

Variables Shows or hides graph elements for variables, or re-orders the display of variables.

Note: These options do not apply to variables in the Group X, Group Y, Wrap, or Page zones.

Check boxes are followed by the zone designation and the name of the variable. Use check boxes to do the following:

– Show or hide the elements corresponding to a variable in a zone.

– Add or remove the effect of applying the Color, Size, Shape, or Freq variable to the variable in the zone.

Tip: If you have multiple graphs, you can color or size each graph by different variables. Drag a second variable to the **Color** or **Size** zone, and drop it in a corner. In the Variables option, select the specific color or size variable to apply to each graph.

Use arrows to re-order the display if there are multiple variables in a zone. Highlight a variable name and click an arrow to reposition it.

Red Triangle Options for Smoother

Response Axis Changes the axis that is associated with the response variable to X (horizontal), Y (vertical), or Auto (the Y axis for Smoother).

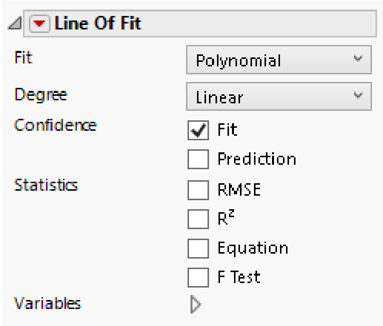
Adapt to Axis Scale For log and other axis transformations, applies computations on the transformed coordinates.

Save Formula Saves fit formulas and upper and lower confidence intervals for a single variable to the data table.

Line of Fit

The **Line of Fit** element shows a linear regression line with confidence intervals for the fit.

Figure 3.27 Line of Fit Options



Fit Specifies the type of fit:

Polynomial A simple linear regression.

Robust Cauchy A linear regression assuming Cauchy distributed residuals, to de-emphasize outliers.

Time Series Smoothing for equally spaced X values with optimal seasonality. Includes options for a forecast model and the number of seasonal and forecast periods.

Degree Specifies the polynomial degree of the linear regression fit, which can be linear, quadratic, or cubic.

Confidence Shows or hides confidence intervals for the predicted value (Fit) or for individual values (Prediction). Both types of intervals are fixed at 95% confidence.

Statistics Shows various selections on the graph. You can show the root mean square error (RMSE), R-square, the equation of the regression line, and the F Test value.

Variables Shows or hides graph elements for variables, or re-orders the display of variables.

Note: These options do not apply to variables in the Group X, Group Y, Wrap, or Page zones.

Check boxes are followed by the zone designation and the name of the variable. Use check boxes to do the following:

– Show or hide the elements corresponding to a variable in a zone.

– Add or remove the effect of applying the Color, Size, Shape, or Freq variable to the variable in the zone.

Tip: If you have multiple graphs, you can color or size each graph by different variables. Drag a second variable to the **Color** or **Size** zone, and drop it in a corner. In the Variables option, select the specific color or size variable to apply to each graph.

Use arrows to re-order the display if there are multiple variables in a zone. Highlight a variable name and click an arrow to reposition it.

Red Triangle Options for Line of Fit

Response Axis Specifies the axis for the variable that is used as the response in the calculation of the linear regression line. Because standard least squares typically uses the Y (vertical) axis as the response axis, the Auto setting defaults to the Y axis.

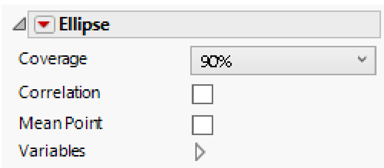
Adapt to Axis Scale For log and other axis transformations, applies computations on the transformed coordinates.

Save Formula Saves fit formulas and upper and lower confidence intervals for a single variable to the data table.

Ellipse

The **Ellipse** element  shows a bivariate normal density ellipse.

Figure 3.28 Ellipse Options



Coverage Specifies the coverage percentage (99%, 95%, 90%, or 50%) for the density ellipse.

Correlation Shows the Pearson correlation coefficient for the X and Y variables on the graph.

Mean Point Shows the mean on the graph.

Variables Shows or hides graph elements for variables, or re-orders the display of variables.

Note: These options do not apply to variables in the Group X, Group Y, Wrap, or Page zones.

Check boxes are followed by the zone designation and the name of the variable. Use check boxes to do the following:

– Show or hide the elements corresponding to a variable in a zone.

– Add or remove the effect of applying the Color, Size, Shape, or Freq variable to the variable in the zone.

Tip: If you have multiple graphs, you can color or size each graph by different variables. Drag a second variable to the **Color** or **Size** zone, and drop it in a corner. In the Variables option, select the specific color or size variable to apply to each graph.

Use arrows to re-order the display if there are multiple variables in a zone. Highlight a variable name and click an arrow to reposition it.

Red Triangle Options for Ellipse

Adapt to Axis Scale For log and other axis transformations, applies computations on the transformed coordinates.

Contour

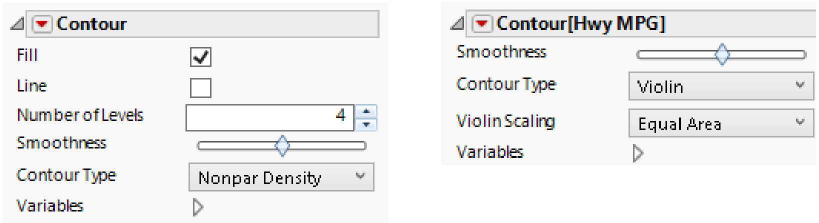
The **Contour** element shows regions of density (or value contours when used with a Color variable). Density contours are useful when you have a scatterplot with many points where the mass of points makes it difficult to see patterns in density. Multiple contour types are available. The default is a smooth bivariate nonparametric density surface that is fit to reflect the density of the data points. The nonparametric density surface estimates the bivariate probability density function at each point, providing a continuous analog of a bivariate histogram.

- For two continuous variables, four contours are plotted by default. These contours are 100%, 75%, 50%, and 25% density contours. You can specify the number of contour levels to display. Alternatively, you can select a Bagplot or High Density Region (HDR) contours.
- For only one continuous variable, a *violin* plot appears instead of a contour plot. A violin plot illustrates the density of the data by plotting symmetric kernel densities around a common vertical axis. The kernel density estimates the probability density function at each point, providing a continuous analog of the histogram. The violin plot is similar to a box plot with symmetric kernel densities replacing the box and whiskers. Alternatively, you can select High Density Region (HDR) contours.
- If you add a Color variable to a Contour plot, the plot shows *value contours* that reflect the levels of the Color variable. The value contours are computed using Delaunay triangulation. You can select an option (Transform) to show a plot where the X and Y ranges have been normalized. See “Example of a Contour Plot with a Color Variable” on page 118 in the “Graph Builder Examples” chapter.

For an example of a contour plot, see “Example of Wafer Maps Based on a Cluster Analysis” on page 146 in the “Graph Builder Examples” chapter. For an example of a violin plot, see “Example of a Violin Plot” on page 120 in the “Graph Builder Examples” chapter.

Contour Options

Figure 3.29 Contour Options for a Contour Plot or a Violin Plot



Fill (Not available for Bagplots.) Fills in the contours.

Line (Not available for Bagplots.) Adds lines around the contours.

Number of Levels (Available for Nonpar Density.) For density contours, specifies the number of contours that appear. The number can be between 1 and 1000, the default is 4 contours.

Boundary (Available only when you have a Color variable.) Adds a line around the outside boundary of the contour.

Alpha (Available only when you have a Color variable.) Controls the hull of value contours. Increasing alpha can eliminate some of the long, skinny, or large triangles where interpolation may be undesirable.

Smoothness (Not available for Bagplots.) Smooths the underlying data and the contours.The smoother is a normalized value between -1 and 1. The value given can be interpreted as a smoothing radius. The original data is interpolated to a grid, and then a Gaussian smoother is applied.

Transform (Available only when you have a Color variable.) Transforms the triangulation to use a normalized scale for X and Y by selecting Range Normalized. This causes both the X and Y values to be scaled to [0,1] before computing the Delaunay triangulation. This option might be desirable in cases where the X and Y units are very different. In these cases, if points are stretched in one dimension and not the other, Delaunay triangulation tries to minimize long, skinny triangles, which can obscure some features.

Contour Type Enables selection of contour types

Nonpar Density A smooth bivariate nonparametric density surface.

Bagplot Draws a Bagplot, also known as a bivariate boxplot. A Bagplot consists of two polygons, a set of outlier points, and a median point. All computations are based on first computing the Tukey depth (bivariate depth) of each point in the data. The median point is the average of all points at maximum depth, which is plotted as an asterisk. The inner polygon is the *bag*, which contains at most 50% of the data points. Not shown in the plot is the *fence*, which is the bag polygon inflated three times relative to the median point. The outer polygon is the convex hull of all points contained within the fence. This is referred to as the *loop*. Points that lie outside the fence are designated outliers, and are shown as points on the plot. For more information about Bagplots, see Rousseeuw (1999).

HDR Draws a high density region contour. These are shaded regions for the 99% probability region and the 50% probability region, Outliers are shown for items that fall outside of the threshold. For more information about high density regions see Hyndman (1996).

Violin Draws violin plot of the density of the data by plotting symmetric kernel densities around a common vertical axis.

Outliers (Available for Bagplot and HDR contours.) Plots outliers.

Violin Scaling For multiple violin plots, choose a scaling option. Equal area or width means that the areas or maximum width of the violins are the same. Weighted area means that the areas of the violins are proportional to the number of observations in each violin.

Variables Shows or hides graph elements for variables, or re-orders the display of variables.

Note: These options do not apply to variables in the Group X, Group Y, Wrap, or Page zones.

Check boxes are followed by the zone designation and the name of the variable. Use check boxes to do the following:

- Show or hide the elements corresponding to a variable in a zone.
- Add or remove the effect of applying the Color, Size, Shape, or Freq variable to the variable in the zone.

Tip: If you have multiple graphs, you can color or size each graph by different variables. Drag a second variable to the **Color** or **Size** zone, and drop it in a corner. In the Variables option, select the specific color or size variable to apply to each graph.

Use arrows to re-order the display if there are multiple variables in a zone. Highlight a variable name and click an arrow to reposition it.

Red Triangle Option for Contour

Adapt to Axis Scale For log and other axis transformations, applies computations on the transformed coordinates.

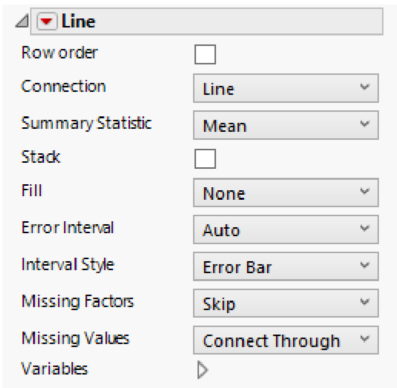
Line

The **Line** element uses line segments to connect summary statistic values (or raw values) of the variable plotted on the Response Axis across the values or categories of another variable.

If you want to emphasize the lines more than the filled areas, you can also create an Area plot using the Line element. For an example, see "Show Behavior over Time - Line Element" on page 138.

Note: If any of the rows used in calculating a point on the plot are hidden and not excluded, the point is not shown, but the position that it would occupy remains. The line segment connects the adjoining points.

Figure 3.30 Line Options



Row order Connects points with line segments in the order of their row numbers.

Connection Changes the connection type to a line, curve, step, centered step, horizontal, or vertical line. If you select a curve, you can also change the smoothness.

Smoothness Changes the smoothness of a Curve connection. To use this option, select Curve for the Connection type.

Summary Statistic Specifies the statistic that is plotted. The statistic is computed for observations at each distinct value of the variable in the X zone (or the Y zone, if X is specified as the Response Axis).

Stack (Available for multiple Y variables.) Accumulates, or stacks, the Y values. Use to build an area chart.

Fill Specify how to fill the area below or between the lines.

Error Interval Adds or removes specified error interval in the graph.

Interval Style Draws error bars or error bands for the selected error interval.

Missing Factors Provides options for connections when all values are missing for a setting of the variable on the X axis:

Skip Skips over the missing factor value, connecting the previous factor to the next factor.

Treat as Missing Skips over the missing factor value, connecting the previous factor to the next factor. Draws the connection in the style specified for Missing Values.

Treat as zero Treats the missing factor values as zero.

Missing Values Provides options for connections when all values are missing for a setting of the variable on the Response axis:

Connect Through Connects the points adjacent to the missing values with the connector type used for the remaining points.

Connect Faded Connects the points adjacent to the missing values with a faded connection.

Connect Dashed Connects the points adjacent to the missing values with a dashed connection.

No Connection Shows no connection between the points adjacent to the missing values.

Variables Shows or hides graph elements for variables, or re-orders the display of variables.

Note: These options do not apply to variables in the Group X, Group Y, Wrap, or Page zones.

Check boxes are followed by the zone designation and the name of the variable. Use check boxes to do the following:

– Show or hide the elements corresponding to a variable in a zone.
– Add or remove the effect of applying the Color, Size, Shape, or Freq variable to the variable in the zone.

Tip: If you have multiple graphs, you can color or size each graph by different variables. Drag a second variable to the **Color** or **Size** zone, and drop it in a corner. In the Variables option, select the specific color or size variable to apply to each graph.

Use arrows to re-order the display if there are multiple variables in a zone. Highlight a variable name and click an arrow to reposition it.

Red Triangle Options for Line

Response Axis Specifies the axis for the variable that is used as the response in plotting the line segments. The Auto setting is the Y axis.

Save Formula Saves fit formulas and upper and lower confidence intervals for a single variable to the data table.

Bar

The **Bar** element  shows a bar chart of summary statistic values of one or more variables across the values or categories of one or more stratifying variables. The graph can display a single variable or multiple variables of any type arranged in the X and Y zones.

- If you add a single categorical variable, then each bar shows the count of observations in the corresponding level of the variable.
- If you add two variables, with one in each zone, each bar represents the value of the specified summary statistic for the variable on the specified Response Axis. There is a bar for each value or level of the stratifying variable.
- If you add multiple variables in each zone, multiple bars are plotted for each value or combination of levels of the stratifying variables.

Response Axis Variables with Character Data Type

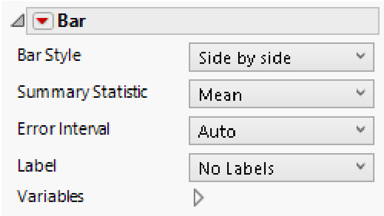
If the variable on the Response Axis has the Character data type, then this variable is assigned numeric values as follows:

- Integer values are assigned based on the Value Order column property or the default value ordering.
- These integer values range from 0 (highest level in value ordering) to the number of levels minus one (lowest value in value ordering).
- These integer values are used in calculating the Summary Statistic.

Note: If any of the rows used in constructing a bar on the plot are hidden and not excluded, the bar is not shown, but the position that it would occupy remains.

For an example of a graph that uses the bar element, see "Example of a Stacked Bar Chart" on page 121 in the "Graph Builder Examples" chapter.

Figure 3.31 Bar Options



Bar Style You can change the appearance of the bars that describe the levels of the stratifying variable. Many bar styles are applicable only when there are merged variables on the response axis, or if an overlay variable is present. For bar style descriptions, see Table 3.1. To change bar widths and space, right-click in the graph and select **Customize** > **Bar**.

Packed Bar Style Options

Packed Primaries Specifies the number of primary bars.

Packed Placement Specifies where the secondary bars are stacked. Options include the following:

First Stack—Start on the top row and fill it before moving to the next row.

Smallest Stack—Place each bar at the stack that is currently the smallest.

Separate Stack—Make a stack called Other where all secondary bars appear.

Packed Ordering Specifies the order in which secondary bars are placed. Can be by size. The biggest bars appear first or by label, which is in natural order, usually alphabetical.

Packed Coloring Specifies how the secondary bars are colored. Options include the following:

Bar color—Applies the same color as the primary bar color.

Faded bar color—Applies a faded version of the primary bar color.

Grays—Applies random shades of light gray.

Packed Labeling Controls the percentage of secondary bars that are labeled.

Packed Primary Labels Specifies where primary labels should be placed, either inside bars or on the axis.

Summary Statistic Specifies the statistic that is used as bar labels and controls the response axis scale. Available statistics depend on the type of bar chart used.

Error Interval Adds or removes error intervals in the graph when applicable, otherwise it is not available. To customize error bars, right-click in the graph and select **Customize** > **Error Bar**.

Label Adds or removes labels of various types in the bar chart. When **Label by Value** is selected the values are determined by the summary statistic setting.

Variables Shows or hides graph elements for variables, or re-orders the display of variables.

Note: These options do not apply to variables in the Group X, Group Y, Wrap, or Page zones.

Check boxes are followed by the zone designation and the name of the variable. Use check boxes to do the following:

– Show or hide the elements corresponding to a variable in a zone.

– Add or remove the effect of applying the Color, Size, Shape, or Freq variable to the variable in the zone.

Tip: If you have multiple graphs, you can color or size each graph by different variables. Drag a second variable to the **Color** or **Size** zone, and drop it in a corner. In the Variables option, select the specific color or size variable to apply to each graph.

Use arrows to re-order the display if there are multiple variables in a zone. Highlight a variable name and click an arrow to reposition it.

Red Triangle Options for Bar

Response Axis Specifies the axis for the variable that is used as the response in plotting the bars. This option is available only when each axis contains at least one variable with a Continuous modeling type. If only one variable is Continuous, its axis is the response axis. The Auto setting is the Y axis.

Save Formula Saves fit formulas and upper and lower confidence intervals for a single variable to the data table.

Table 3.1 Bar Styles and Descriptions

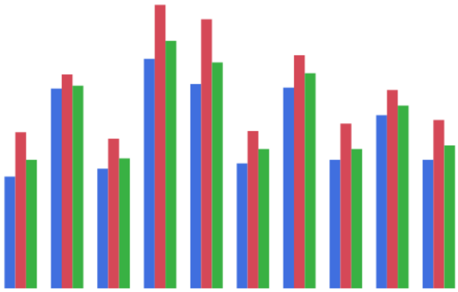
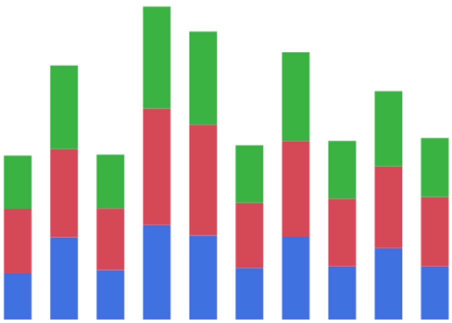
Bar Style	Description
Side by side	When you have two or more merged variables on the response axis, a bar is plotted for each of the merged variables. The bars are arranged next to each other for each level of the stratifying variables. They are colored and a legend appears. 
Stacked	When you have two or more merged variables on the response axis, a single bar is plotted for each level of the stratifying variables. Each bar stacks the bars for each of the merged variables. The part of the bar corresponding to each merged variable is colored and a legend appears. 

Table 3.1 Bar Styles and Descriptions *(Continued)*

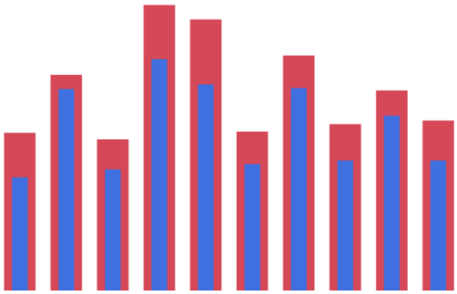
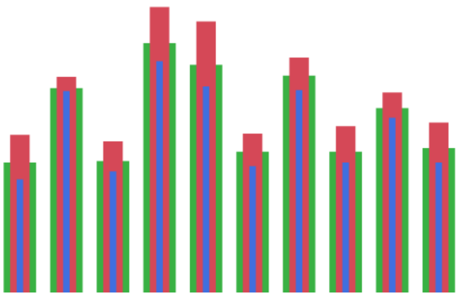
Bar Style	Description
Bullet	When you have two or more merged variables on the response axis, a narrow bar is plotted within a wider and taller bar. The widths of the bars depend on the order of the variables. The part of the bar corresponding to each merged variable is colored and a legend appears. If you have only a single variable, this style plots narrow bars. 
Nested	When you have two or more merged variables on the response axis, the bars for each variable are nested within each other. The narrowest bar corresponds to the first variable listed in the legend, the next narrowest to the second, and so on. **Tip:** You can change the legend order using the arrows in the Variables panel. 

Table 3.1 Bar Styles and Descriptions *(Continued)*

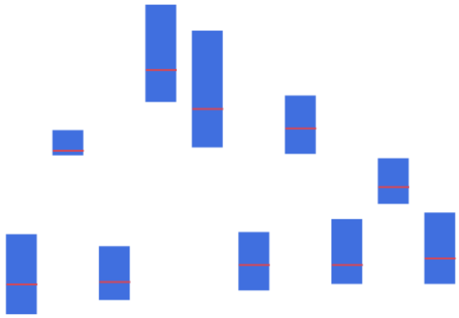
Bar Style	Description
Range	When you have two or more merged variables on the response axis, a rectangle and a line are plotted for each level of the stratifying variables. For each level of the stratifying variables: • The values of the summary statistic are calculated for each of the merged variables. • A rectangle is plotted with bounds equal to the summary statistic values for the first two of variables, as indicated in the legend. • Lines are plotted at the values of the summary statistic for the other variables. 

Table 3.1 Bar Styles and Descriptions *(Continued)*

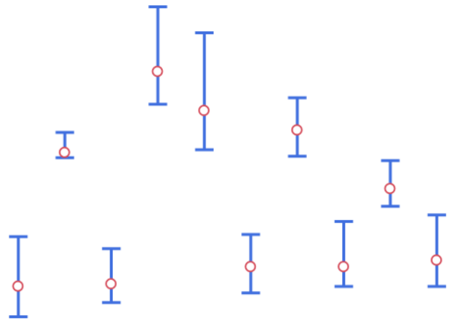
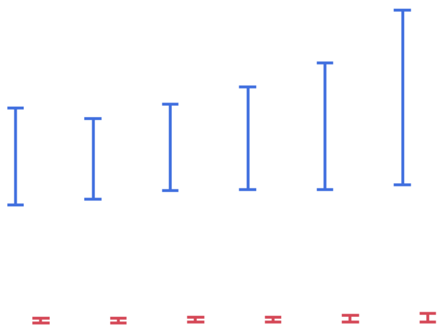
<table>
<tr><th>Bar Style</th><th>Description</th></tr>
<tr><td>Interval</td><td>When you have two or more merged variables on the response axis, a line with boundaries and a circle is plotted for each level of the stratifying variables. For each level of the stratifying variables:
• The values of the summary statistic are calculated for each of the merged variables.
• A line with boundaries is plotted with bounds equal to the summary statistic values for the first two of variables, as indicated in the legend.
• A circle is plotted at the overall values of the summary statistic for the other variables.
</td></tr>
<tr><td>Side by side intervals</td><td>In this style, each pair of variables defines an interval.
</td></tr>
</table>

Table 3.1 Bar Styles and Descriptions *(Continued)*

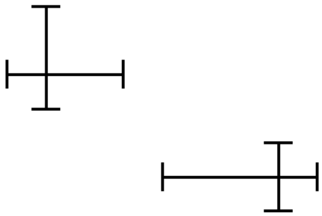
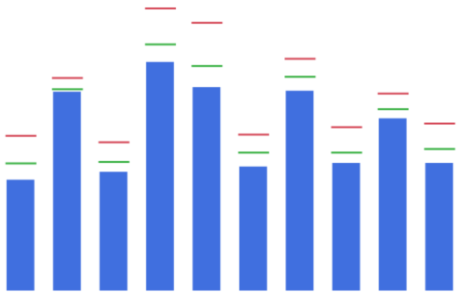
Bar Style	Description
Two-way interval	This style is for two-dimensional bar charts with bars in each direction. 
Single	For each level of the stratifying variables, a bar is plotted. When you have two or more merged variables on the response axis, the bar represents the value of the summary statistic for the first variable in the legend and lines are shown for the values of the summary statistic for the remaining variables. **Tip:** You can change the legend order using the arrows in the Variables panel. 

Table 3.1 Bar Styles and Descriptions *(Continued)*

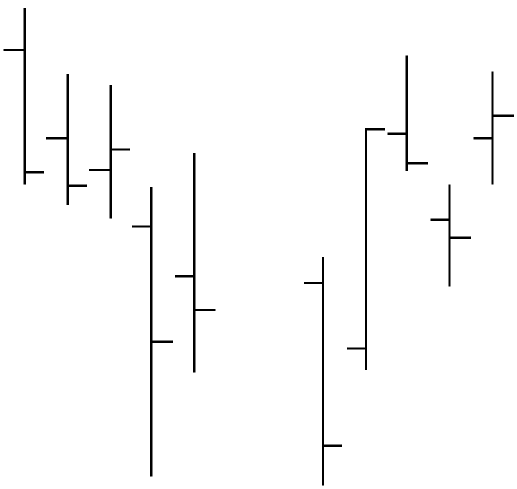
Bar Style	Description
Stock	When you have two or more merged variables on the response axis, line segments are plotted at the levels of the stratifying variables. For each level of the stratifying variables: • The values of the summary statistic are calculated for each of the merged variables. • A line segment connecting the summary statistic values for two of the variables, as indicated in the legend. • Bars are plotted at right angles to the line segment at the remaining values of the summary statistic. 

Table 3.1 Bar Styles and Descriptions *(Continued)*

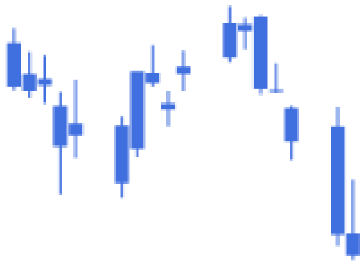
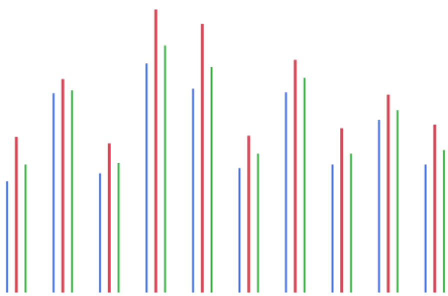
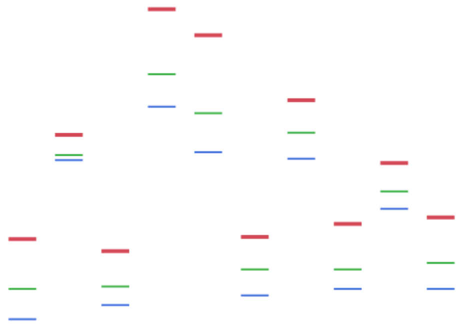
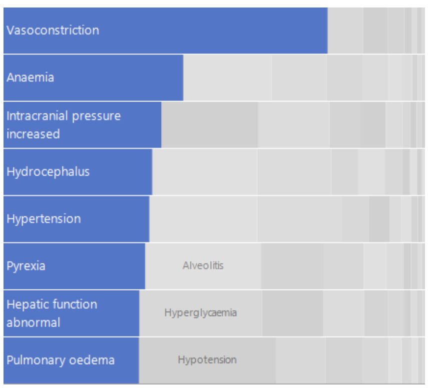
Bar Style	Description
Box Plot	When you have two or more merged variables on the response axis, box plots are shown for the levels of the stratifying variables. For each level of the stratifying variables: • The values of the summary statistic are calculated for each of the merged variables. • An outlier box plot is constructed for these values. 
Needle	When you have two or more merged variables on the response axis, needle-shaped bars are plotted side-by-side for each of the merged variables. The bars are colored and a legend appears. This style is useful when you have many levels of the stratifying variables. 

Table 3.1 Bar Styles and Descriptions *(Continued)*

Bar Style	Description
Float	When you have two or more merged variables on the response axis, bars are plotted at the levels of the stratifying variables. For each level of the stratifying variables: • The values of the summary statistic are calculated for each of the merged variables. • A bar is plotted at each of these values. This style is useful when you have many merged variables. 
Packed	This bar style shows the top categories as a bar chart and stacks the other categories as bars with an approximately rectangular fill. The top categories are the focus and the other bars provide context. This style is useful when you have many categories and skewed response data. 

Area

Note: You can also create an area plot using the Line element.

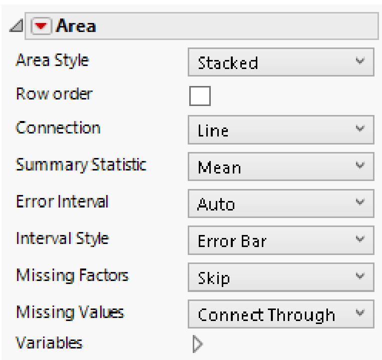
The **Area** element shows a contiguous area that represents the area beneath summary statistic values of a variable across the values or categories of another variable. The Area element can be considered a continuous representation of the Bar element.

- Both variables might be nominal or ordinal. The summary statistic represents the count of observations in a level of the variable on the axis that is not specified as the Response Axis.
- One of the variables might be continuous and the other nominal or ordinal. The summary statistic represents the value of the specified summary statistic for the continuous variable for the corresponding level of the categorical variable.
- If both variables are continuous, the summary statistic is computed for the continuous variable on the axis that is specified as the Response Axis.

Note: The plot is not affected in any way if rows are hidden and not excluded.

For an example of a graph that uses the Area element, see “Examples of Area Plots” on page 135 in the “Graph Builder Examples” chapter.

Figure 3.32 Area Options



Area Style Changes how the area is displayed when there are merged variables on the response axis. For area style descriptions, see Table 3.2.

Row order Connects points with line segments in the order of their row numbers.

Connection Changes the connection type to a line, curve, step, centered step, horizontal, or vertical line. If you select a curve, you can also change the smoothness.

Summary Statistic Specifies the statistic that is plotted. The statistic is computed for observations at each distinct value of the variable in the X zone (or the Y zone, if X is specified as the Response Axis).

Error Interval Adds or removes specified error interval in the graph.

Interval Style Draws error bars or error bands for the selected error interval.

Missing Factors Provides options for connections when all values are missing for a setting of the variable on the X axis:

Skip Skips over the missing factor value, connecting the previous factor to the next factor.

Treat as Missing Skips over the missing factor value, connecting the previous factor to the next factor. Draws the connection in the style specified for Missing Values.

Treat as zero Treats the missing factor values as zero.

Missing Values Provides options for connections when all values are missing for a setting of the variable on the Response axis:

Connect Through Connects the points adjacent to the missing values with the connector type used for the remaining points.

Connect Faded Connects the points adjacent to the missing values with a faded connection.

Connect Dashed Connects the points adjacent to the missing values with a dashed connection.

No Connection Shows no connection between the points adjacent to the missing values.

Variables Shows or hides graph elements for variables, or re-orders the display of variables.

Note: These options do not apply to variables in the Group X, Group Y, Wrap, or Page zones.

Check boxes are followed by the zone designation and the name of the variable. Use check boxes to do the following:

– Show or hide the elements corresponding to a variable in a zone.
– Add or remove the effect of applying the Color, Size, Shape, or Freq variable to the variable in the zone.

Tip: If you have multiple graphs, you can color or size each graph by different variables. Drag a second variable to the **Color** or **Size** zone, and drop it in a corner. In the Variables option, select the specific color or size variable to apply to each graph.

Use arrows to re-order the display if there are multiple variables in a zone. Highlight a variable name and click an arrow to reposition it.

Red Triangle Options for Area

Response Axis Specifies the axis for the variable that is used as the response in plotting the areas. Available only when both variables have a Continuous modeling type. If only one variable is Continuous, its axis is the response axis. The Auto setting is the Y axis.

Save Formula Saves fit formulas and upper and lower confidence intervals for a single variable to the data table.

Table 3.2 Area Styles and Descriptions

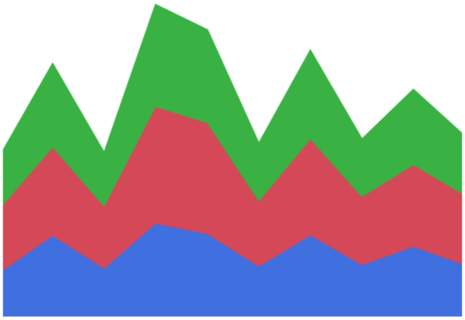
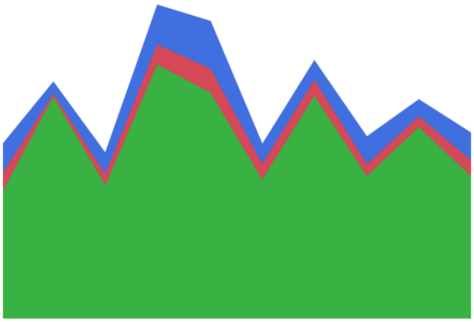
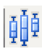
Area Style	Description
Stacked	Shows areas that correspond to summary statistics for each of the merged variables. The plots are colored and a legend appears in the Legend zone. 
Overlaid	Shows the area plots for the merged variables superimposed over each other. The plots are colored and a legend appears in the Legend zone. 

Table 3.2 Area Styles and Descriptions *(Continued)*

Area Style	Description
Range	Shows the area corresponding to the difference in the summary statistic for the merged variables. 

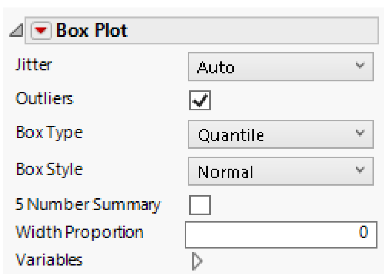
Box Plot

The **Box Plot** element  shows outlier or quantile box plots. A box plot provides a compact view of a distribution of values. The box plot element is useful when variables have a Numeric data type. For more information about outlier and quantile box plots, see the Distributions chapter in *Basic Analysis*.

- If one of the variables is continuous and the other is nominal or ordinal, box plots are constructed for values of the continuous variable. There is a box plot for each level of the categorical variable.
- If both variables are continuous, box plots are constructed for values of the continuous variable on the axis specified as the Response Axis. There is a box plot for each level of the other continuous variable.

For an example of a graph using box plots, see "Example of Features in Graph Builder" on page 32.

Figure 3.33 Box Plot Options



Jitter Turns jitter on or off. Jitter adds random noise to data, and can help reduce over-plotting.

Outliers Shows or hides values that extend beyond the whiskers.

Box Type Specifies whether the box plot is an outlier box plot or a quantile box plot.

Box Style Changes the style the box plot. For box style descriptions, see Table 3.3.

5 Number Summary Adds statistics like the median, maximum, minimum, and quantiles 1 and 3 to the graph.

Width Proportion adjusts the width of the box plots.

Variables Shows or hides graph elements for variables, or re-orders the display of variables.

Note: These options do not apply to variables in the Group X, Group Y, Wrap, or Page zones.

Check boxes are followed by the zone designation and the name of the variable. Use check boxes to do the following:

– Show or hide the elements corresponding to a variable in a zone.

– Add or remove the effect of applying the Color, Size, Shape, or Freq variable to the variable in the zone.

Tip: If you have multiple graphs, you can color or size each graph by different variables. Drag a second variable to the **Color** or **Size** zone, and drop it in a corner. In the Variables option, select the specific color or size variable to apply to each graph.

Use arrows to re-order the display if there are multiple variables in a zone. Highlight a variable name and click an arrow to reposition it.

Red Triangle Options for Box Plot

Response Axis Specifies the axis for the variable that is used as the response in constructing the box plots. Available only when both variables have a Continuous modeling type. If only one variable is Continuous, its axis is the response axis. The Auto setting is the Y axis.

Table 3.3 Box Styles and Descriptions

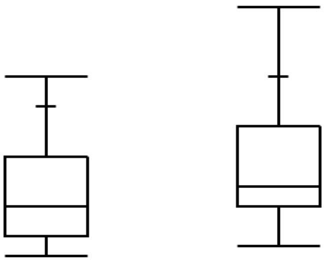
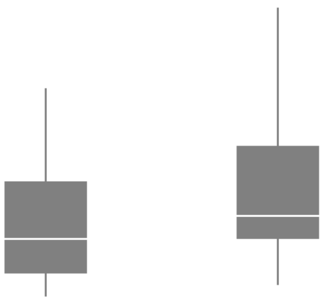
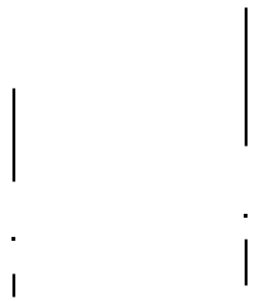
Box Style	Description
Normal	Shows the box defined by the quartiles as hollow with only a line at the median. 
Solid	Shows the box defined by the quartiles with solid fill. The line at the median is shown by white space. 

Table 3.3 Box Styles and Descriptions *(Continued)*

Box Style	Description
Thin	The box defined by the quartiles is not shown. It is discernible due to the whiskers. The median is denoted by a point. 

Histogram

The **Histogram** element  shows a variable's distribution using binning. For more information about histograms, see the Distributions chapter in *Basic Analysis*.

Note: The bars of the histogram align with the tick marks on the axis for the zone containing the variable. If you customize the axis, the histogram adjusts to match the new tick marks.

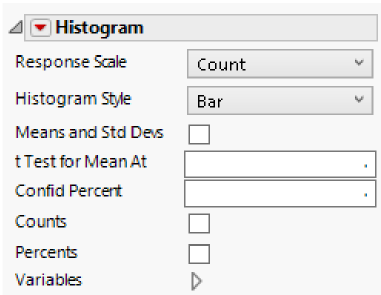
You can change the primary direction of the graph to X (horizontal), Y (vertical), or Auto using the Response Axis option.

- If both variables are nominal or ordinal, the histogram shows the count of observations in a level of the variable on the axis that is not specified as the Response Axis.
- If one of the variables is continuous and the other is nominal or ordinal, the plot shows a histogram for the continuous variable for each level of the categorical variable.
- If the same variable is specified for both *X* and *Y*, then the *Y* role is ignored and a single histogram appears.
- If both variables are continuous, then the variable on the X axis is treated as discrete. You might need to rescale the X axis to view the graph clearly. For an alternative visual, use a scatter plot with the Contour element.

Tip: You can overlay histograms with transparent color so that you can see the individual histograms. To do this, assign the variable of interest as X or Y. Assign your overlay variable to the **Overlay** zone. Then, click the **Histogram** element icon. See "Example of Overlaying Histograms with Transparency" on page 144 in the "Graph Builder Examples" chapter.

For an example illustrating the histogram element, see "Example of Features in Graph Builder" on page 32 and "Example of Wafer Maps Based on a Cluster Analysis" on page 146 in the "Graph Builder Examples" chapter.

Figure 3.34 Histogram Options



Response Scale Specifies the scale for the response axis. The options are count, percent, or fill. Fill has no response axis label, fills the display space, and scales multiple histograms independently.

Overlap (Available with a categorical response.) Adjusts the amount of histogram overlap between vertical categories.

Histogram Style Specifies the histogram style.

Bar Default traditional histogram bars with a height based on the number or percent of observations that fall within each bar or bin.

Polygon Connects the peaks of each histogram bar to construct a polygon representation of the distribution of the data.

Kernel Density Density curve with a smoother control.

Shadowgram Overlays histograms with different bin widths.

Means and Std Devs Shows the means and standard deviations for the levels of the variables in the X or Y zone.

t Test for Mean At Performs a *t* test for the specified mean.

Confid Percent Specifies the confidence interval for the mean.

Counts Displays counts on the histogram bars.

Percents Displays percents on the histogram bars.

Variables Shows or hides graph elements for variables, or re-orders the display of variables.

Note: These options do not apply to variables in the Group X, Group Y, Wrap, or Page zones.

Check boxes are followed by the zone designation and the name of the variable. Use check boxes to do the following:

– Show or hide the elements corresponding to a variable in a zone.

– Add or remove the effect of applying the Color, Size, Shape, or Freq variable to the variable in the zone.

Tip: If you have multiple graphs, you can color or size each graph by different variables. Drag a second variable to the **Color** or **Size** zone, and drop it in a corner. In the Variables option, select the specific color or size variable to apply to each graph.

Use arrows to re-order the display if there are multiple variables in a zone. Highlight a variable name and click an arrow to reposition it.

Red Triangle Options for Histogram

Response Axis Specifies the axis for the variable that is used as the response in constructing the histograms. Available only when both variables have a Continuous modeling type. If only one variable is Continuous, the other axis is the response axis.

Heatmap

The **Heatmap** element shows the counts or average values for groups of observations using a color intensity scale. If you plot a single variable as either Y or X, your plot appears as bars. If you plot a cross tabulation of two variables Y and X, your plot shows rectangles.

- For a categorical variable, the levels define the groups of observations.
- For a continuous variable, Graph Builder constructs non-overlapping intervals of values. These intervals define the groups of observations.

Note: If all of the rows used in constructing a bar or rectangle on the plot are hidden and not excluded, the rectangle is not shown, but the position that it would occupy remains.

When you create a heatmap, an intensity legend appears in the Legend area. See “Discrete or Continuous Legend Items” on page 110. The intensity legend shows counts or average values:

- If no Color variable is assigned, the counts in each bar or rectangle define the scale for the intensity legend.
- If a Color variable is assigned, the average of that variable for the observations in each bar or rectangle define the scale for the intensity legend.

Tip: Place your cursor over a cell to see labels. Click a cell to select the corresponding rows.

To color or size a heatmap by two variables, drag the first variable into the Color or Size zone and then drag the second variable into the top or bottom corner of the zone.

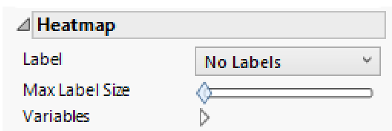
Use a Heatmap to Apply Background Colors

You can also use the heatmap to control the background color when you have plot elements that use the Wrap, Group X, or Group Y zones. To do so:

1. Select the variable for the background color and drag it to the **Color** zone.
2. Select the Heatmap element .
3. Construct the plot elements.
4. In the Heatmap panel, under Variables, de-select all check boxes except the one for Color.

For examples illustrating the heatmap element, see “Example of Wafer Maps Based on a Cluster Analysis” on page 146 in the “Graph Builder Examples” chapter and “Example Using a Heatmap to Apply Background Colors” on page 148 in the “Graph Builder Examples” chapter.

Figure 3.35 Heatmap Options



Label Adds labels to the heatmap. You can label by value, percent of total, or by row.

Max Label Size Changes the size of the label.

Variables Shows or hides graph elements for variables, or re-orders the display of variables.

> **Note:** These options do not apply to variables in the Group X, Group Y, Wrap, or Page zones.

Check boxes are followed by the zone designation and the name of the variable. Use check boxes to do the following:

– Show or hide the elements corresponding to a variable in a zone.

– Add or remove the effect of applying the Color, Size, Shape, or Freq variable to the variable in the zone.

Tip: If you have multiple graphs, you can color or size each graph by different variables. Drag a second variable to the **Color** or **Size** zone, and drop it in a corner. In the Variables option, select the specific color or size variable to apply to each graph.

Use arrows to re-order the display if there are multiple variables in a zone. Highlight a variable name and click an arrow to reposition it.

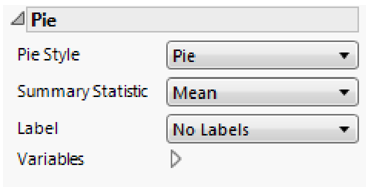
Pie

The **Pie** element  plots summary statistics for groups of observations, representing their values as the size of the slices or rings on a pie chart.

Note: If any of the rows used in constructing a slice of the pie chart are hidden and not excluded, the slice is not shown, but the position that it would occupy remains.

- For a single variable, the slice sizes the number of observations in each category.
- For two variables, the variable in the Y zone is used to size the slices according to the selected Summary Statistic. A legend appears in the Legend zone showing the values of the variable in the X zone.

Figure 3.36 Pie Options



When you add an Overlay variable, a ring chart is constructed for each level of the Overlay variable. The outer ring represents the smallest value in the value ordering, and the inner ring represents the largest value.

Pie Style Changes the appearance of the pie chart. For pie style descriptions, see Table 3.4.

Summary Statistic Changes the statistic that is plotted.

Label Adds or removes labels in the pie chart.

Variables Shows or hides graph elements for variables, or re-orders the display of variables.

Note: These options do not apply to variables in the Group X, Group Y, Wrap, or Page zones.

Check boxes are followed by the zone designation and the name of the variable. Use check boxes to do the following:

– Show or hide the elements corresponding to a variable in a zone.

– Add or remove the effect of applying the Color, Size, Shape, or Freq variable to the variable in the zone.

Tip: If you have multiple graphs, you can color or size each graph by different variables. Drag a second variable to the **Color** or **Size** zone, and drop it in a corner. In the Variables option, select the specific color or size variable to apply to each graph.

Use arrows to re-order the display if there are multiple variables in a zone. Highlight a variable name and click an arrow to reposition it.

Table 3.4 Pie Style Descriptions

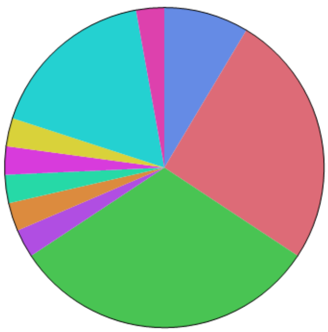
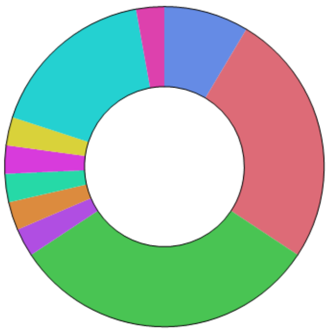
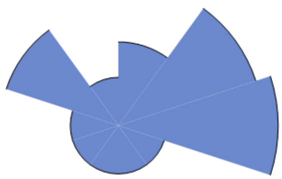
Pie Style	Description
Pie	Traditional pie chart with each slice sized by the Summary Statistic. 

Table 3.4 Pie Style Descriptions *(Continued)*

Pie Style	Description
Ring	Each variable or level of a stratifying variable is represented by a concentric ring. The sections are sized by the Summary Statistic. Ring charts can help you visualize hierarchical data, using concentric rings. 
Coxcomb	The central angles for all slices are equal. The size of each slice is determined by the Summary Statistic. Also known as an *exploded pie chart*, a coxcomb chart can help you see smaller areas. 

Treemap

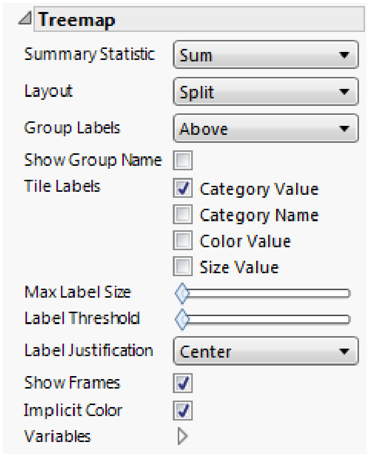
The **Treemap** element  shows the levels or values of a variable as rectangles within a rectangular display. The rectangle sizes represent summary statistic values of the variable in the Y zone across the levels or values of one or more X variables. The rectangles are sized to represent the values of the summary statistic without leaving unused space in the overall rectangular layout.

To construct a treemap that shows summary values across the levels of two or more X variables, merge the variables in the X zone.

Note: The plot is not affected in any way if rows are hidden and not excluded.

For an example illustrating the Treemap element, see “Example of a Treemap” on page 151 in the “Graph Builder Examples” chapter.

Figure 3.37 Treemap Options



Note: Some of these options are specific to Graph Builder and are not available in the Treemap platform.

Summary Statistic Specifies the summary statistic, corresponding to the variable in the Y zone, that is used to size the rectangles.

Layout Arranges rectangles to the extent possible by the value ordering of the X variable or by the size of the rectangle.

> **Split** Arranges the rectangles according to the value ordering of the levels or values of the X variable. Split is the default setting.
>
> **Squarify** Arranges the rectangles according to the values of the summary statistic, sorted in descending order. This places the largest rectangles in the top left corner of the plot and the smallest in the bottom right corner.
>
> **Mixed** Applies only when you have two or more X variables. It applies Split to the outermost variable and Squarify to the other variables. It follows that the large groupings are ordered according to their value ordering, whereas the inscribed rectangles are ordered according to the summary statistic values.

Group Labels When you have more than one X variable, show the additional group labels above each category, or floating (default) in the center of each category. You can also remove group labels.

Tip: To filter data in a group, click a group label.

Show Group Name Adds the variable title to the group labels.

Tile Labels Shows or hides the following labels:

Category Value The values of the X variable.

Category Name The name of the X variable.

Color Value The values of the Color variable.

Size Value The values of the Size variable, or the Y variable if no Size variable exists (Size Value).

Tip: Place your cursor over a rectangle to see details.

Max Label Size Increases or decreases the size of the labels.

Label Threshold Removes labels based on the size of the rectangle. By default, all labels are shown. Move the slider to the right to remove labels for progressively larger rectangles.

Label Justification Places the label in the center, left, or right.

Show Frames Shows or hides the borders around the rectangles.

Tip: Use the treemap option in the right-click customize menu to adjust frame color, style, and thickness.

Implicit Color Shows or hides the color. If no variable is specified in the Color zone, all rectangles have the same color.

Variables Shows or hides graph elements for variables, or re-orders the display of variables.

Note: These options do not apply to variables in the Group X, Group Y, Wrap, or Page zones.

Check boxes are followed by the zone designation and the name of the variable. Use check boxes to do the following:

– Show or hide the elements corresponding to a variable in a zone.

– Add or remove the effect of applying the Color, Size, Shape, or Freq variable to the variable in the zone.

Tip: If you have multiple graphs, you can color or size each graph by different variables. Drag a second variable to the **Color** or **Size** zone, and drop it in a corner. In the Variables option, select the specific color or size variable to apply to each graph.

Use arrows to re-order the display if there are multiple variables in a zone. Highlight a variable name and click an arrow to reposition it.

Mosaic

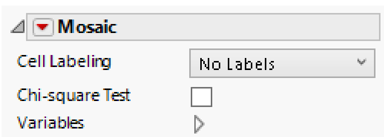
The **Mosaic** element uses rectangle sizes to represent the proportion of observations in categories for one variable across categories of another variable. Although variables can be categorical or continuous, they are treated as categorical. For more information about mosaic plots, see the Contingency Analysis chapter in *Basic Analysis*.

The default Response Axis is the Y axis. This means that, for each level of the X variable, vertical stacked bars show the proportion of values in each level of the Y variable. You can change the primary direction of the graph to X (horizontal).

Note: The plot is not affected in any way if rows are hidden and not excluded.

For an example illustrating the Mosaic element, see "Example of a Mosaic Plot" on page 153 in the "Graph Builder Examples" chapter.

Figure 3.38 Mosaic Options



Cell Labeling Add labels to cells based on counts, percents, and so on.

Chi-square Test Performs a Pearson chi-square test and shows results in a label on the graph. This option appears if each side has 2 or more levels.

Variables Shows or hides graph elements for variables, or re-orders the display of variables.

Note: These options do not apply to variables in the Group X, Group Y, Wrap, or Page zones.

Check boxes are followed by the zone designation and the name of the variable. Use check boxes to do the following:

– Show or hide the elements corresponding to a variable in a zone.

– Add or remove the effect of applying the Color, Size, Shape, or Freq variable to the variable in the zone.

Tip: If you have multiple graphs, you can color or size each graph by different variables. Drag a second variable to the **Color** or **Size** zone, and drop it in a corner. In the Variables option, select the specific color or size variable to apply to each graph.

Use arrows to re-order the display if there are multiple variables in a zone. Highlight a variable name and click an arrow to reposition it.

Red Triangle Options for Mosaic

Response Axis Changes the axis that is associated with the response variable to X (horizontal), Y (vertical), or Auto (the Y axis for Mosaic). The variable on the Response Axis is used to calculate the proportion of observations in each level of the variable on the other axis.

Caption Box

The **Caption Box** element  displays a summary statistic value on the plot. The summary statistic corresponds to the variable that corresponds to the axis that you specify as the Response Axis in the Caption Box options.

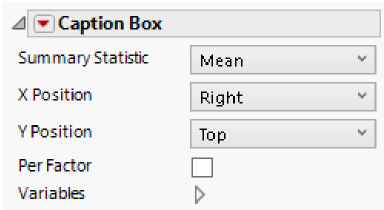
Summary Statistic for Categorical Variables

For a categorical variable, the Summary Statistic that appears is calculated by assigning numerical values to the levels of the categorical variable. The levels are placed in the order defined by their value ordering. They are assigned integer values ranging from 0 to the number of levels minus one.

Note: When using categorical values in zones associated with summarization roles like Y and Color, summary statistics are calculated based on numerical values assigned to the variable levels. (The integer values range from 0 to the number of levels minus one.) The final result is then mapped back as a level to a corresponding categorical value or value range. This approach allows the use of ordinal variables in summarization, but might not lead to clear results for nominal variables.

For an example illustrating the Caption Box element, see "Example of Features in Graph Builder" on page 32.

Figure 3.39 Caption Box Options



Summary Statistic Specifies the summary statistic to be displayed for the variable defined by the Response Axis setting. The default Summary Statistic is the Mean. For more information about how summary statistics are computed for categorical variables, see "Summary Statistic for Categorical Variables" on page 91.

X Position Specifies the horizontal position of the caption.

Y Position Specifies the vertical position of the caption.

Per Factor Shows a caption for each X, or each Y if horizontal.

Variables Shows or hides graph elements for variables, or re-orders the display of variables.

> **Note:** These options do not apply to variables in the Group X, Group Y, Wrap, or Page zones.

Check boxes are followed by the zone designation and the name of the variable. Use check boxes to do the following:

– Show or hide the elements corresponding to a variable in a zone.

– Add or remove the effect of applying the Color, Size, Shape, or Freq variable to the variable in the zone.

> **Tip:** If you have multiple graphs, you can color or size each graph by different variables. Drag a second variable to the **Color** or **Size** zone, and drop it in a corner. In the Variables option, select the specific color or size variable to apply to each graph.

Use arrows to re-order the display if there are multiple variables in a zone. Highlight a variable name and click an arrow to reposition it.

Red Triangle Options for Caption Box

Response Axis Specifies the axis for the variable whose summary statistic is displayed. The Y axis is the default. Notice that this axis can differ from the Response Axis specified to define the plot.

Formula

The **Formula** element shows the graphs of a function, its inverse, or a parametric curve. The function or functions are defined by column formulas.

- $y = f(x)$:
  - To plot a function of a single argument across the values of its single argument, drag the column that contains the formula to the Y zone and the column that contains the values of its single argument to the X zone.
  - Make sure that the Response Axis is set to Auto or Y.
- $y = f^{-1}(x)$:
  - To plot the inverse of a function of a single argument for a column of values, drag the column that contains the formula to the X zone. Drag the column that contains the values of interest to the Y zone.
  - Set the Response Axis to X.
- $x = f(t)$ and $y = g(t)$: To plot the parametric curve defined by two functions, drag one to the X zone and the other to the Y zone.

Note: Because the Formula element plots the values of a function, the plot is not affected in any way if rows are hidden and not excluded.

For an example illustrating the Formula element, see "Example Using a Formula" on page 154 in the "Graph Builder Examples" chapter.

Figure 3.40 Formula Options

Variables Shows or hides graph elements for variables, or re-orders the display of variables.

> **Note:** These options do not apply to variables in the Group X, Group Y, Wrap, or Page zones.

Check boxes are followed by the zone designation and the name of the variable. Use check boxes to do the following:

- Show or hide the elements corresponding to a variable in a zone.
- Add or remove the effect of applying the Color, Size, Shape, or Freq variable to the variable in the zone.

Tip: If you have multiple graphs, you can color or size each graph by different variables. Drag a second variable to the **Color** or **Size** zone, and drop it in a corner. In the Variables option, select the specific color or size variable to apply to each graph.

Use arrows to re-order the display if there are multiple variables in a zone. Highlight a variable name and click an arrow to reposition it.

Red Triangle Options for Formula

Response Axis Specifies the axis for the formula variable.

Map Shapes

The Map Shapes element  shows areas defined by the variable in the Map Shape zone.

When a column in your data table contains the names of geographical regions (such as countries, states, provinces, counties), you can assign the column to the Map Shape zone. When a variable is dropped in the Map Shape zone, Graph Builder searches for internal maps that correspond to the values in the column. If it finds an appropriate map, it draws the map in the graph area.

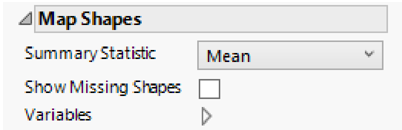
Alternatively, you can define the Map Role column property for the column of interest. This property tells JMP how to connect the values in the column with map shape data. It is especially useful when you create custom maps. See “Custom Map Files” on page 273 in the “Maps” chapter.

If you have a variable in the Map Shape zone, the X and Y zones disappear. You can add information from other variables using Color and Size.

- To color the map shapes by the values of a summary statistic, drag the column of interest to the **Color** zone. The categorical or continuous color theme selected in your Preferences is applied to each shape.
- To size the map shapes by the values of a summary statistic, drag the column of interest to the **Size** zone. This scales the map shapes according to the summary statistic value of the size variable, minimizing distortion.

See the “Maps” chapter on page 263. For examples, see “Examples of Creating Maps” on page 286 in the “Maps” chapter, or run the associated scripts in these sample data tables: PopulationByMSA.jmp or SAT.jmp.

Figure 3.41 Map Shapes Options



Summary Statistic Changes the statistic that is plotted as the Color or Size variable.

Show Missing Shapes Shows or hides missing data from the map.

Variables Shows or hides graph elements for variables, or re-orders the display of variables.

> **Note:** These options do not apply to variables in the Group X, Group Y, Wrap, or Page zones.

Check boxes are followed by the zone designation and the name of the variable. Use check boxes to do the following:

– Show or hide the elements corresponding to a variable in a zone.

– Add or remove the effect of applying the Color, Size, Shape, or Freq variable to the variable in the zone.

> **Tip:** If you have multiple graphs, you can color or size each graph by different variables. Drag a second variable to the **Color** or **Size** zone, and drop it in a corner. In the Variables option, select the specific color or size variable to apply to each graph.

Use arrows to re-order the display if there are multiple variables in a zone. Highlight a variable name and click an arrow to reposition it.

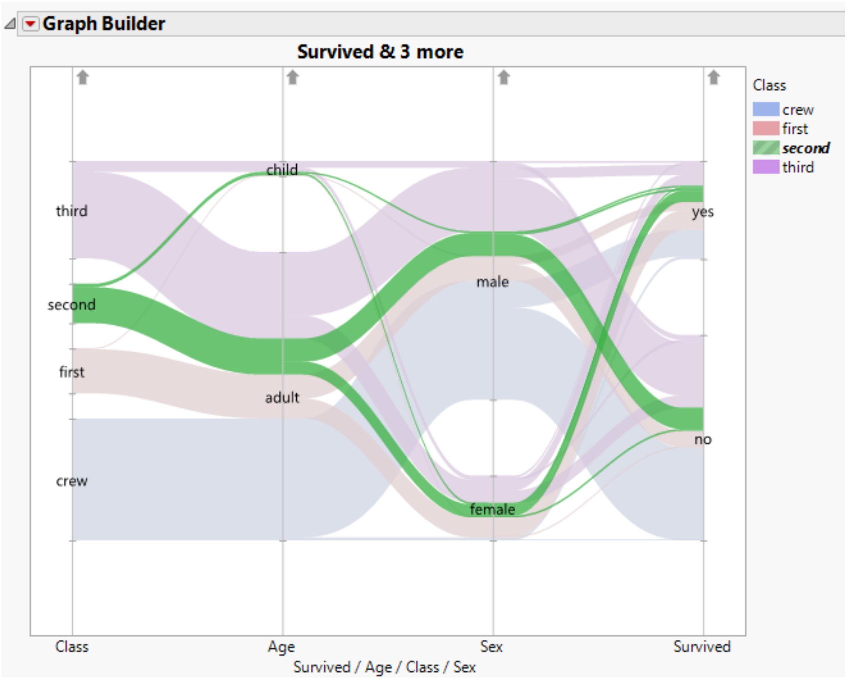
Parallel

The **Parallel** element  connects the values in a row across two or more variables. Drag two or more variables together to either the X or Y zone. The variable names appear as axis labels in the zone to which they were dragged.

- Values for each continuous variable are plotted along lines (axes) parallel to the other axis.
- Levels of categorical variables are represented by intervals on lines (axes) parallel to the other axis. The sizes of the intervals are proportional to the number of observations in each level of the categorical variable.
- The values (for continuous variables) or bands (for categorical variables) are joined with continuous curves.

When all variables are categorical, there is a band for every combination of levels of the categorical variables. The bands split as they move from left to right. In the first interval corresponding to a categorical variable, there is a band for every level of that categorical variable. In the last interval corresponding to categorical variables, there is a band for every possible combination of the categorical variables.

Figure 3.42 Example of Categorical Bands Using Titanic.jmp



In Figure 3.42, the band containing all second-class passengers is selected. The parallel plot shows that most were adults, there were more males than females, and slightly fewer survived than did not survive.

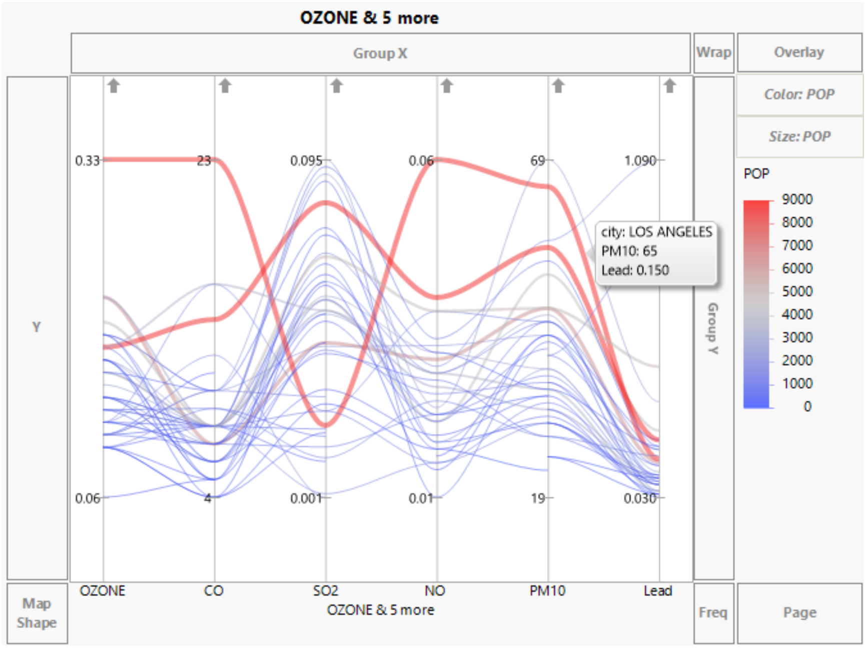
The values or levels of the variables are connected with lines or curves using the Curve Lines option.

- To color the curves by the values of a variable, drag the column of interest to the **Color** zone. The categorical or continuous color theme selected in your Preferences is shown in the Legend zone.
- To size the curves by the values of a variable, drag the column of interest to the **Size** zone.
- To change the direction of a variable's axis, click the arrow at the top of the vertical line denoting the variable's axis.
- To move a variable and its axis, click and drag the axis.

Tip: Place your cursor over a curve to see a label giving information about the corresponding row.

Figure 3.43 shows a parallel plot for six variables in the Cities.jmp data table. The variable POP is used both as a Color and Size variable. The curve for Los Angeles is labeled.

Figure 3.43 Parallel Plot for Pollution Data in Cities.jmp



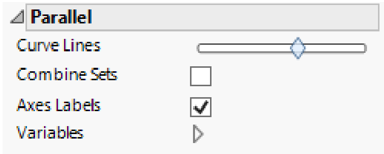
Tip: You can add reference lines for specification limits. For information, see The Column Info Window chapter in *Using JMP*.

Scaling

By default, the scales for the values of the variables are adjusted so that the minimum and maximum values are plotted at the same level. For example, in Figure 3.43, the values of each of the variables have an identical vertical spread. Each vertical line is labeled by the minimum and maximum values of the variables.

In Figure 3.43, the scales for CO and PM10 differ greatly from the scales of the other variables. When your variables are measured on very different scales, this scaling enables you to see differences clearly.

Figure 3.44 Parallel Options



Curve Lines Adjusts the amount of curvature of the curves that connect points. Place the slider all the way to the left for lines. As you move the slider to the right, the degree of curvature increases.

Combine Sets (Applicable only if you have three or more categorical factors.) For categorical variables, when Combine Sets is not selected, bands split as they move from left to right. In the last interval that corresponds to a categorical variable, there is a band for every possible combination of the categorical variables. Selecting the Combine Sets option causes the bands not to split. Each interval that follows a categorical variable shows a band for each level of that variable.

Axes Labels Removes the axis labels from the display.

Variables Shows or hides graph elements for variables, or re-orders the display of variables.

Note: These options do not apply to variables in the Group X, Group Y, Wrap, or Page zones.

Check boxes are followed by the zone designation and the name of the variable. Use check boxes to do the following:

– Show or hide the elements corresponding to a variable in a zone.

– Add or remove the effect of applying the Color, Size, Shape, or Freq variable to the variable in the zone.

Tip: If you have multiple graphs, you can color or size each graph by different variables. Drag a second variable to the **Color** or **Size** zone, and drop it in a corner. In the Variables option, select the specific color or size variable to apply to each graph.

Use arrows to re-order the display if there are multiple variables in a zone. Highlight a variable name and click an arrow to reposition it.

Graph Zones

The main element in the Graph Builder window is the graph area. The graph area contains zones, and you drag and drop variables from the Variables box into the zones.

Tip: If you drop variables into the center of the graph, JMP guesses which drop zone to put them into, based on whether the variables are continuous, ordinal, or nominal.

There are two types of zones into which you can drag variables:

- *Data zones* include **X**, **Y**, **Map Shape**, **Freq**, **Color**, **Size**, and **Interval**. The X, Y, and Map Shape zones are positional, and influence the types of graph elements that are available. The Freq, Color, Size, and Interval zones modify certain graph elements.
- *Grouping zones* partition the data into subsets that repeat the graph for each possible subset. Grouping zones include **Group X**, **Group Y**, **Wrap**, **Overlay**, and **Page**.

The following zones are available in Graph Builder:

Zone	Description
X and Y	Constructs plots that are based on the values, levels, or categories of the variables that you drop into these zones. Use the X zone for a horizontal layout and the Y zone for a vertical layout.

Group X and Group Y	Stratifies your data by the levels of the nominal or ordinal variable and constructs a plot for each level. Group X produces a horizontal display, and Group Y produces a vertical display. The type of variable that you put in Group X or Y also determines aspects of the display: • Adding a continuous variable creates non-overlapping interval subgroups and plots for the values in each interval. • Adding a categorical variable creates plots for each level of the categorical variable. You can add multiple variables to Group X or Group Y: • To add a variable *above* a variable in one of the Group zones, drag it to the left of that variable in the Group X zone or above that variable in the Group Y zone. • To add a variable *below* a variable in one of the zones, drag it to the right of that variable in the Group X zone or below that variable in the Group Y zone. To stratify by the levels of two variables, drag a variable into both Group X and Group Y. To change the number of levels that appear, right-click in the axis border and select Levels in View. For a description of all options, see “Options for Axes, Variable Labels, or Graph Titles” on page 107.

Map Shape	Adds a map. Your variable must contain geographic place names, such as countries, regions, states, or provinces, or you must provide information in a custom shape file. Note the following: • If you drag and drop columns labeled **Latitude** and **Longitude**, JMP automatically assigns them to the Y and X roles (respectively). • When you drop a variable into the **Map Shape** zone, the X and Y zones disappear. • The Map Shape element can be used with the Points element to place points within the shape. For more information about creating map shapes, see “Red Triangle Options for Formula” on page 94 and “Graph Builder” on page 267 in the “Maps” chapter.
Wrap	Creates rows of plots for the levels of a categorical variable or for the interval subgroups of a continuous variable. Use Wrap when your stratifying variable has many levels. To change the number of levels that appear, right-click in the axis border and select **Levels in View**. For a description of all options, see “Grouping Zones Options” on page 106. **Note:** Once a variable is placed here, no variable can be placed in Group X.
Overlay	Subgroups and colors the plot elements according to the levels of the Overlay variable. • Adding a nominal or ordinal variable stratifies and colors the plot elements by the levels of the variable. • Adding a continuous variable creates interval subgroups. The plot elements are stratified and colored according to the subgroups. A legend appears to the right of the plot.

Color	Adds color to points, map shapes, and other objects. • Adding a nominal or ordinal variable colors the plot elements by the variable's levels. • Adding a continuous variable colors the plot elements according to an intensity scale. A legend appears to the right of the plot. Right-click in the legend to change colors or to customize the intensity scale. **Tip:** Once you have added a variable in the Color zone, you can show or hide color using the Variables option in the element properties panel.
Size	Sizes graph elements by a summarizing statistic or other size variable.
Interval	Add interval variables to draw custom error bars. • A single interval variable is treated as a delta to compute upper and lower ends of the interval. • Two interval variables are treated as lower and upper values for the interval.
Freq	Add a frequency or weight variable. When all values are integer values, the Freq variable is treated as a frequency, otherwise it is treated as a weight. A footer note indicates how the variable is treated. The Freq variable affects summary statistics.
Page	Drop a By group variable to the Page zone to show each level of the group on a separate graph. By default, the Page zone is limited to 200 pages. The first 200 categories are displayed. Adjust this limit in the Graph Builder preferences. See "Platforms" on page 663 in the "JMP Preferences" chapter. **Tip:** Once you click Page, you can select the red triangle menu option Link Page Axes. Use this option to link or unlink graph axis scales across levels of the By group variable in the Page zone.

For more information about the legend, see "Legend Options" on page 108.

Note: When you use categorical values in zones associated with summarization roles like Y and Color, summary statistics are calculated based on numerical values assigned to the variable levels (integer values ranging from 0 to the number of levels minus one). The final result is then mapped back as a level to a corresponding categorical value or value range. This approach allows the use of ordinal variables in summarization, but might not lead to clear results for nominal variables.

Categorical and Continuous Variables in Group X or Y Zones

When a categorical variable is added to the Group X or Group Y zone, a partition is created for each level of the variable.

When a continuous variable is added to a grouping zone, Graph Builder uses quantiles of the data to divide the variable into five groups. To change the number of groups:

1. Right-click the grouping variable label and select **Number of Levels**.
2. Type in the number of levels that you want to display.
3. Click **OK**.

Move Group X and Group Y Labels

Grouping variable labels can be relocated to another position on the graph. The Group X labels can be either on the top or the bottom of the graph. The Group Y labels can be either on the right or the left of the graph.

- To relocate a Group X label, right-click the variable in the Group X zone and select **X Group Edge > Top** or **Bottom**.
- To relocate a Group Y label, right-click the variable in the Group Y zone and select **Y Group Edge > Left** or **Right**.

Red Triangle Options

The red triangle menu for Graph Builder contains these options:

Show Control Panel Shows or hides the platform buttons, the Select Columns box, and the drop zone borders.

Show Legend Shows or hides the legend.

Legend Position Sets the position of the legend. The legend appears on the right by default. Putting the legend at the bottom places it in the center below the graph. The legend items then appear horizontally instead of vertically. Or, you can put the legend inside the graph, on the left or right.

Legend Settings Opens a window where you can modify legend settings, such as the title, font, and title position.

Color Settings Contains the following options:

Continuous Color Theme Selects the color theme for continuous variables.

Categorical Color Theme Selects the color theme for categorical variables.

Note: For more information about color themes, see the Enter and Edit Data chapter in *Using JMP*.

Lighten large fills Automatically lightens colors for elements that fill large areas, such as pie, treemap, and mosaic charts.

Use row colors for levels When every level has a unique color, initialize legend levels with row colors.

Show Title Shows or hides the graph title. Right-click to change the alignment or the span of the title.

Show Subtitle Shows or hides the graph subtitle. Right-click to change the alignment or the span of the subtitle.

Show Footer Shows or hides the footer, which contains informative messages such as missing map shapes, error bar notes, frequency notes, and WHERE clauses.

Lock Scales Prevents axis scales and gradient legend scales from automatically adjusting in response to data or filtering changes.

Error Bar Offset Adjust the offset for overlaid error bars.

Link Page Axes Links or unlinks graph axis scales across levels of the By group variable in the Page zone.

Fit to Window Determines whether the graph is resized as you resize the JMP window. The default setting is Auto, which is usually set to On except when the window is shared with other content or if there is a Page variable. To always fit the graph inside the window, keep the setting at On. You can also maintain the aspect ratio. To prevent the graph from resizing, change the setting to Off.

Sampling Uses a random sample of the data to speed up graph drawing. If the sample size is zero, or greater than or equal to the number of rows in the data table, then sampling is turned off.

Graph Spacing Sets the amount of space between graph panels.

Include Missing Categories Enables a graph to collect and display missing values for categorical variables.

Set Alpha Level Enables you to set the Alpha level for graphs and statistics that use an alpha value for confidence intervals.

Launch Analysis Launches the Fit Model platform with the variables on the graph placed into roles. It launches the Distribution platform when only one variable is placed.

Make into Data Table Creates a new data table that contains the results from the graph.

See the JMP Reports chapter in *Using JMP* for more information about the following options:

Local Data Filter Shows or hides the local data filter that enables you to filter the data used in a specific report.

Redo Contains options that enable you to repeat or relaunch the analysis. In platforms that support the feature, the Automatic Recalc option immediately reflects the changes that you make to the data table in the corresponding report window.

Save Script Contains options that enable you to save a script that reproduces the report to several destinations.

Graph Builder Buttons

The Graph Builder window contains these buttons:

Recall Populates the Graph Builder window with the last graph that you created. This button changes into the Undo button once you perform an action.

Dialog Opens the Graph Builder launch window. This button changes into the Start Over button once you perform an action. You can add a By variable in the Graph Builder launch window.

Done Hides the buttons, Columns list, and all drop zone outlines. Equivalent to de-selecting the Show Control Panel option.

Use this presentation-friendly format for pasting into other applications. To copy the graph, select **Edit > Copy**. To restore the window to the interactive mode, click the Graph Builder red triangle and select **Show Control Panel**.

Undo Reverses the last change made to the window.

Start Over Returns the window to the default condition, removing all data and graph elements from the window, and all variables from the drop zones.

Right-Click Menus

- "Graph Options"
- "Grouping Zones Options"
- "Options for Columns in the Variables Panel"
- "Options for Axes, Variable Labels, or Graph Titles"
- "Options for the Dividing Line between Multiple Graphs"
- "Legend Options"

For multiple graphs (created with grouping zones or the Wrap zone), note the following:

- Generally, any options that you apply to one graph element apply to all graphs, across all grouping and wrap variables.
- To apply an option to only one graph, right-click in that graph area and select the option from the Graph menu.
- To apply an option to all of your multiple graphs simultaneously, first hold down the Ctrl key and right-click and then select the option from the Graph menu.

Graph Options

Right-clicking in a graph shows the following:

1. Element types that appear in the plot (such as Points, Line, and so on). Sub-menus provide options that are specific to each element type, and general options. Alternatively, you can change element-specific options in each element's properties panel, below the Variables panel.
2. The Add option, where you can select other elements to add to the graph. Alternatively, you can add more elements to the graph by dragging the desired element icons.
3. The Rows, Graph, Customize, and Edit menus. For descriptions of these options, see *Using JMP*.

Grouping Zones Options

Right-clicking in the X Group, Y Group, Wrap, Color, Size, and Overlay zones can show the following options:

Levels in View Changes the number of levels of the grouping variable that are visible. Use the arrow buttons to scroll forward and backward.

Number of Levels Changes the number of levels. See "Categorical and Continuous Variables in Group X or Y Zones" on page 103.

Order By Orders the levels of a variable. See “Order the Levels of a Categorical Variable” on page 48.

Show Title Shows or hides the variable title.

Title Orientation Changes the orientation of the variable text to horizontal or vertical.

Level Orientation Changes the orientation of the level values to horizontal or vertical.

Levels per Row Changes the number of columns included in the graph. Use with a **Wrap** variable.

Color Changes the background color of the grouping zone.

X or Y Group Edge Moves the grouping variable labels. See “Move Group X and Group Y Labels” on page 103.

Summary Statistic Applies color and or size variables based on a summary statistic. Available only for a summarized Y response.

Swap Swaps the position of two variables. See “Move or Remove Variables in Zones” on page 39.

Remove Removes a variable.

Options for Columns in the Variables Panel

Right-clicking on a column in the Variables panel shows options that depend on the column’s modeling type. For more information about these menus, see the Enter and Edit Data chapter in *Using JMP*.

Options for Axes, Variable Labels, or Graph Titles

Right-clicking on axes, variable labels, or graph titles shows options that are specific to Graph Builder and general options: Graph Builder options appear above the line, and general options appear below the line. The following options are specific to Graph Builder:

Alignment Changes the graph title alignment.

Span Changes the width of the graph title span.

Combine Scales Combines scales when you have multiple variables in the same zone. Auto does the following:

- If there are 4 or more variables on one axis and no variables on the other axis, elements that support parallel mode are combined as parallel scales. If all variables are continuous and overlapping, Parallel Merged is used.

- If all variables are categorical, use Nested.

In addition, consider the following:

  – Merged creates one scale that covers all variables.
  – Nested creates a nested scale with one level per variable.
  – Parallel Merged is for parallel coordinates with a shared scale. You can use this option with the following elements: Points, Contour (Violin), Histogram, Box Plot and Parallel.
  – Parallel Independent is for parallel coordinates with independent scales. You can use this option with the following elements: Points, Contour (Violin), Histogram, Box Plot and Parallel.

Remove Removes the variable from the zone.

Swap Switches a variable from one zone to another. Select the variable that you want to switch places with.

Move Right (Appears only if you have more than one variable in the Y zone.) Creates a second Y axis (on the right) for the variable that you select.

Order By (Appears only if you have categorical variables in the X or Y zone.) Orders the levels of a categorical variable. See "Order the Levels of a Categorical Variable" on page 48.

For descriptions of the general options below the line, see the JMP Reports chapter in *Using JMP*.

Options for the Dividing Line between Multiple Graphs

This menu appears only if you have two or more graphs in the graph area. Right-click the line that separates graphs. Options appear that change the formatting of the line, such as color, transparency, spacing, and borders.

Tip: If the line between the graphs is too thin, you can resize it. Click the Graph Builder red triangle, select **Graph Spacing**, and enter a larger value.

Legend Options

A legend consists of items, such as lines, fill colors, and gradients. Each item has its own set of properties that can be changed using a right-click menu or the Legend Settings window.

To change legend settings or revert to initial settings, right-click to the right of the legend or on the title (if applicable). To undo recent changes to the legend, click Undo.

Tip: When you have both an overlay and a color variable, the color variable levels are shown by rounded color icons in the legend.

Legend Settings Change certain aspects of the graphic elements (line, bars, and so on) such as color, fill, transparency, and gradient. You can also add a title to the legend. See "Legend Settings Window" on page 109.

Revert Legend Returns the legend to its initial settings.

Move the Legend

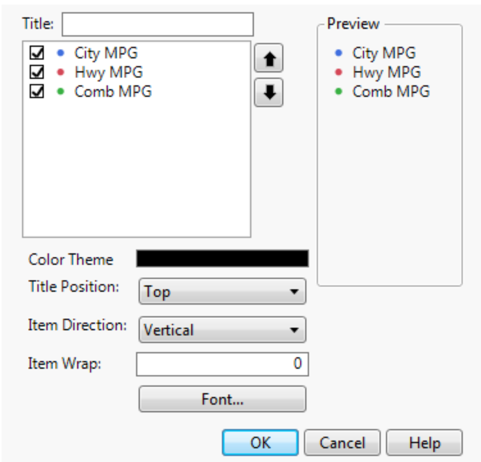
To move the position of the legend, click the Graph Builder red triangle and select **Legend Position**. The legend appears on the right by default, but you can move it to the bottom of the graph, or inside the graph on the left or right.

To make the legend vertical or horizontal, change the Item Direction in the Legend Settings Window.

Legend Settings Window

Change aspects of the legend through the Legend Settings window.

Figure 3.45 Example of the Legend Settings Window Using Hybrid Fuel Economy.jmp



Title The name of the legend.

Check boxes Shows or hides items in the legend. JMP hides items that appear to be redundant, but you can make them appear by selecting them. If your legend has a large number of entries, some might be hidden in the Graph Builder window to ensure sufficient space for the graph. However, all of the entries appear here, so you can select the ones that you want to show. To select or deselect all entries, right-click and select **Toggle check box**.

Up and down arrows Changes the order of items in the legend.

Color Theme Select a different color theme. For more information about color themes, see the Enter and Edit Data chapter in *Using JMP*.

Title Position Places the legend title on top or to the left of the items in the legend.

Item Direction Displays the legend horizontally or vertically.

Item Wrap Sets the legend to be *n* items tall (if vertical) or *n* items wide (if horizontal).

Font Changes the font type, style, and size.

Preview Shows your changes to the legend.

OK Commits your changes to the legend.

Cancel Cancels your changes to the legend.

Help Opens the online Help.

Discrete or Continuous Legend Items

Legends can list discrete values or show a scale of continuous values:

- Discrete legends can contain the levels of a categorical variable, or a list of variables.
- Continuous legends describe a color intensity scale for a range of numerical values.

Right-click an item in either legend to see the following options:

Line Color (Discrete legends only) Changes the color of the element associated with the item.

Marker Changes the type of marker associated with the item.

Marker Size Changes the size of the marker associated with the item.

Pattern or Style Changes the fill or pattern of the element associated with the item.

Line Width Changes the width of the line associated with the item.

Fill Color Assigns a color for the contours.

Color by Theme Colors multiple contours by levels in a color theme. This option sets line colors and fill colors for density contours.

Transparency Changes the marker or label transparency. Enter the level of transparency to draw markers (points) on the graph. The degrees of opacity ranges from 0 (clear) to 1 (opaque).

Gradient (Continuous legends only) Changes the gradient settings. See “Gradient Settings” on page 111.

Density Gradient (Contour element only) Changes the gradient colors as follows:

Fade to White The highest density contour uses the fill color, and colors fade to white at lower densities.

Fade to Gray The highest density contour uses the fill color, and colors fade to gray at lower densities.

Full Color Density contours use a predefined or customized color theme with full color control.

Gradient Transparency (Contour element only) Changes the transparency of the gradient as follows:

None Density contours use one transparency level for all levels.

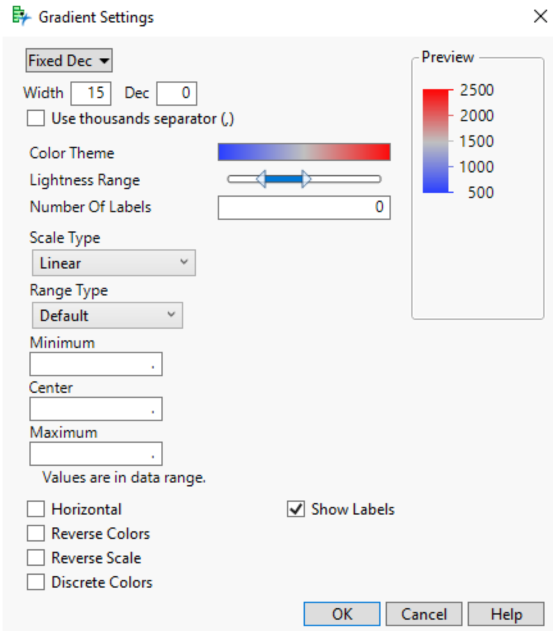
Linear Density contours get gradually more transparent at lower densities.

Revert Restores default settings.

Gradient Settings

You can change various aspects of the gradient legend in the Gradient Settings window.

Figure 3.46 Gradient Settings Window



To format the labels in the legend, use the menu, Width and Dec boxes, and the **Use thousands separator (,)** check box in the top left of the window.

Color Theme Change the color theme or define a custom color theme. For more information about color options, see the Enter and Edit Data chapter in *Using JMP*.

Lightness Range Set the range of intensities for the gradient coloring.

Number of Labels Specify the number of labels for your legend. The value of zero provides the default number of labels.

Scale Type Sets the scale for the gradient coloring. Look below the Maximum option for a description of the scale type.

Linear The scale is piecewise linear between the Minimum and Center values and between the Center and Maximum values. This is the default scaling.

Quantile The scale is piecewise linear for quantiles of the variable represented by the legend.

Standard Deviation The legend range is divided into offsets from the mean determined by standard deviation increments. The scale is piecewise linear between these offset settings. The number of offsets is determined by the Number of Labels.

Log The scale is linear for the logarithm of the values represented by the legend.

Log Offset The scale is linear for the offset of the base 10 logarithm of the values represented by the legend.

Range Type Sets the range of the legend values.

Default Unless you specify values for the Minimum and Maximum, the minimum and maximum values include the range of the data and are chosen to be appropriately rounded values.

Exact Data Range The minimum and maximum values are the exact minimum and maximum values for the data, or they are the values that you specify as Minimum and Maximum.

Middle 90% The minimum and maximum values are the 5^{th} and 95^{th} quantiles, or they are the values that you specify as Minimum and Maximum. This option is resistant to outliers.

Minimum, Center, Maximum Specify the smallest, middle, and largest values to use for your legend scaling.

Note: These values can override the Range Type values.

Horizontal Changes the orientation of the legend to horizontal.

Reverse Colors Reverses the colors in the color scheme.

Reverse Scale Reverses the scale for the color theme.

Discrete Colors Changes the color scheme from a continuous gradient to a stepped gradient with discrete colors.

Show Labels Shows or hides labels for the legend.

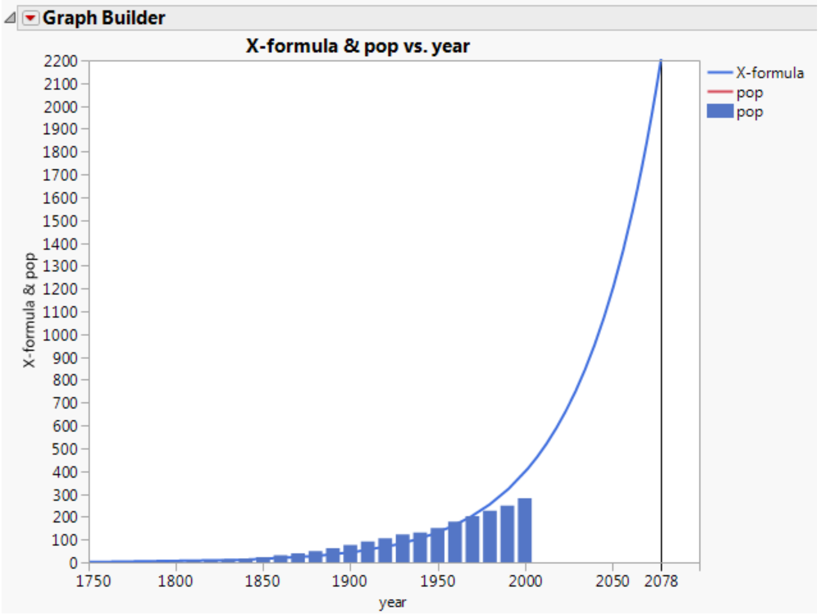
Chapter 4

Graph Builder Examples

Visualize Data Interactively

Graph Builder provides you with a variety of flexible ways to explore your data. This chapter shows examples of examining data using contour plots, bar charts, histograms, area plots, treemaps, and other types of plots.

Figure 4.1 Example Using a Formula, Bars, and a Reference Line



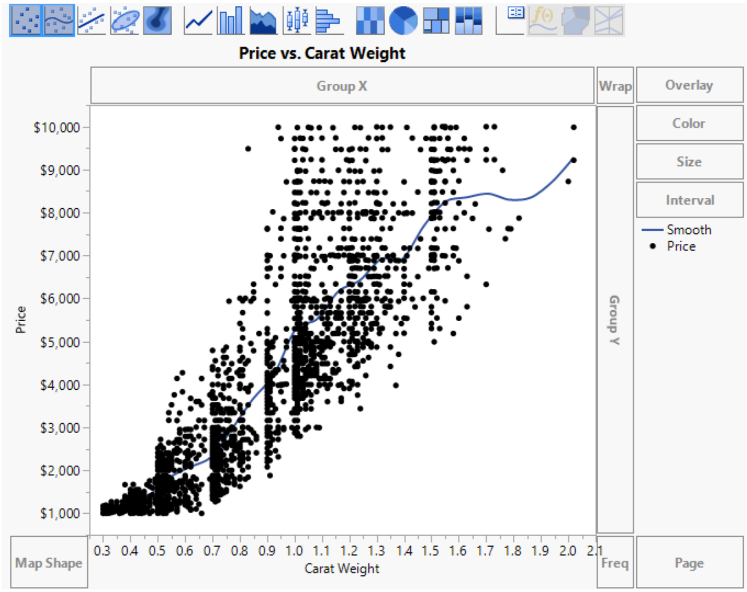
Contents

Example of a Contour Plot

You have data about diamonds, including their carat weight and price. Examine the relationship between carat weight and price.

1. Select **Help > Sample Data Library** and open Diamonds Data.jmp.
2. Select **Graph > Graph Builder**.
3. Select Price and drag it to the **Y** zone.
4. Select Carat Weight and drag it to the **X** zone.

Figure 4.2 Points Showing Diamond Characteristics

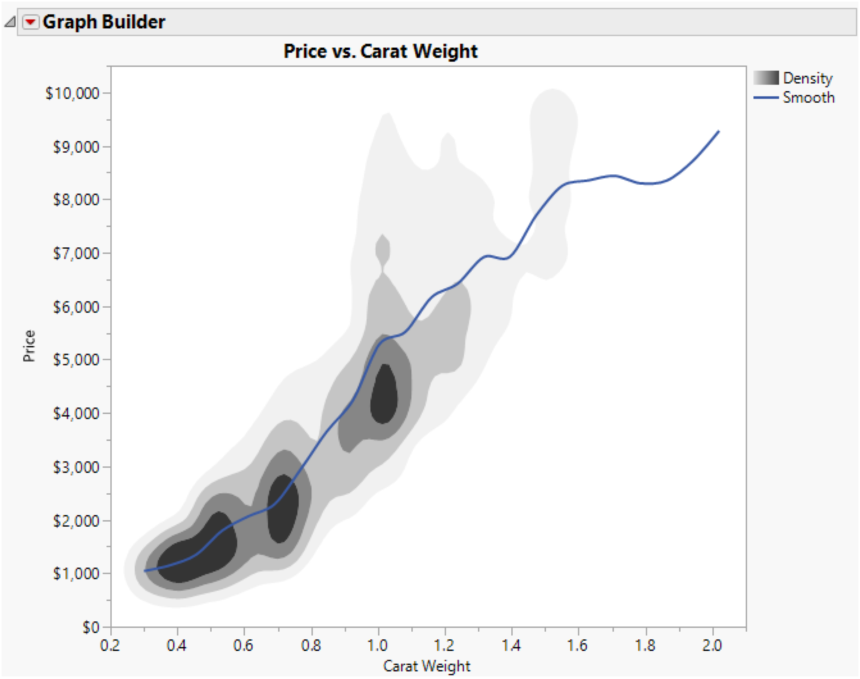


You can see that the points are difficult to interpret. Some points overlap, making the density unclear.

Replace the points with a contour plot of their density.

5. Right-click the plot and select **Points > Change to > Contour**.
6. (Optional) Click **Done**.

Figure 4.3 Contour Plot of Diamond Characteristics



The darker areas indicate a higher density of observations. Use the crosshairs tool to verify the following:

- Most people in the sample purchased diamonds with a carat weight of about 0.34 to 0.57, 0.67 to 0.75, and 0.97 to 1.05.
- Most people who purchased diamonds with a carat weight around 0.5 paid about $100-$2,000.
- Most people who purchased diamonds with a carat weight of 0.70 paid about $1,600-$2,700.
- Most people who purchased diamonds with a carat weight of 1.0 paid about $3,800-$4,800.

Example of a Contour Plot with a Color Variable

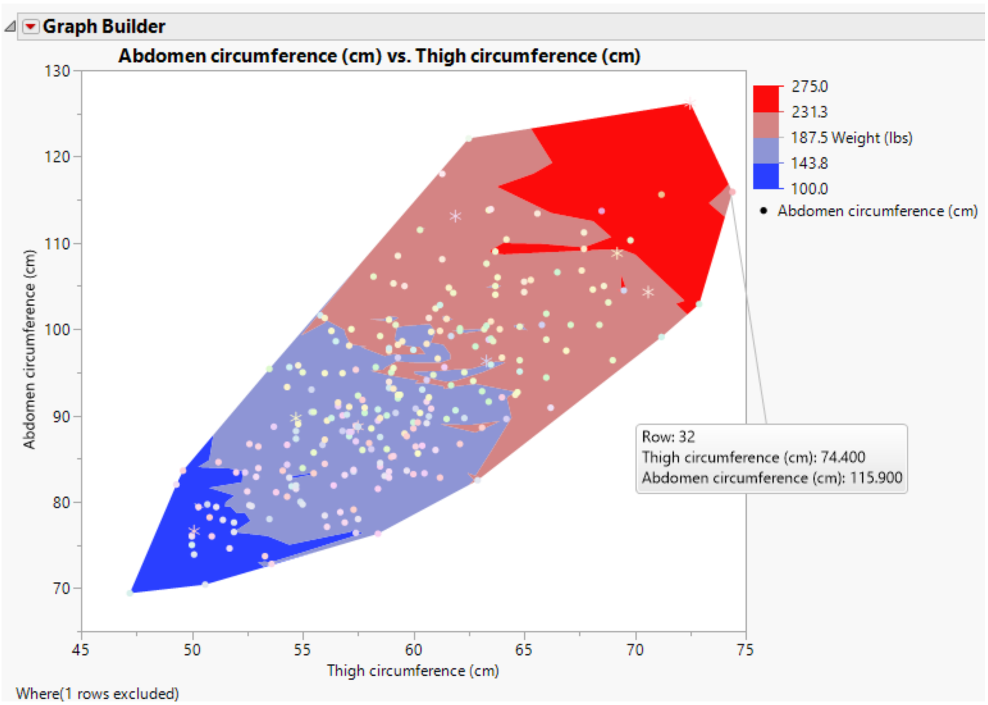
This example uses body fat data collected from men. You want to get an understanding of how the relationship between abdomen circumference and thigh circumference varies by weight.

1. Select **Help > Sample Data Library** and open Body Fat.jmp.
2. Right-click row 1 and select **Hide and Exclude**.

The data in this row is an outlier, which is why we want to hide and exclude it from the graph.

3. Select **Graph > Graph Builder**.
4. Select Abdomen circumference and drag it to the **Y** zone.
5. Select Thigh circumference and drag it to the **X** zone.
6. Select Weight (lbs) and drag it to the **Color** zone.
7. Click the Contour element .
8. Drag the Points element into the plot.
9. In the Points options under Variables, de-select **Color Weight (lbs)**.
10. (Optional) Click **Done**.

Figure 4.4 Contour Plot with a Color Variable



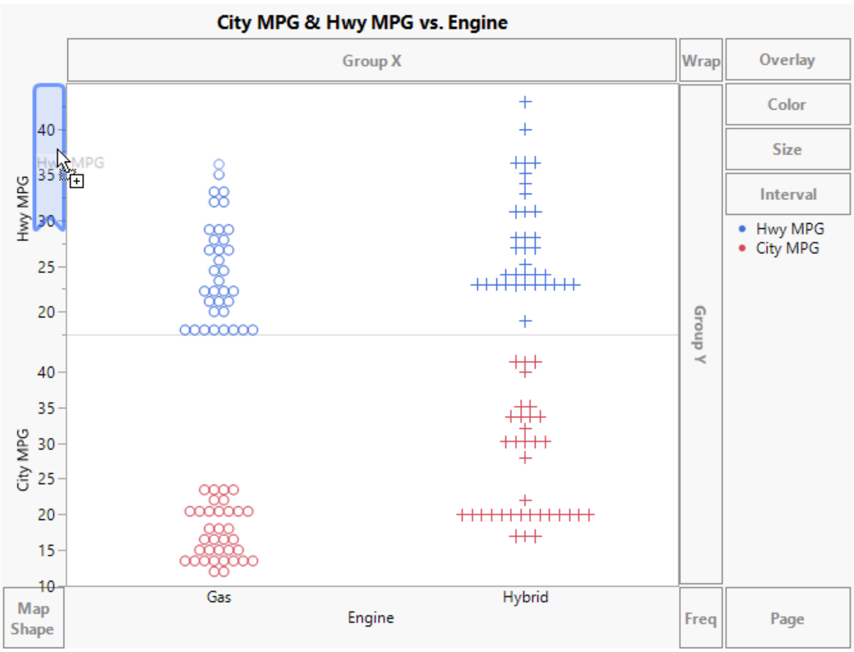
The contour colors correspond to the weight, as shown in the Weight (lbs) legend. The graph shows that thigh and abdomen circumference are related; as one increases, so does the other. The colored contours indicate how weight fits in the relationship. Individuals with large measurements of abdomen and thigh circumference are also generally the heaviest. Notice that one individual (row 32) does not quite fit the pattern.

Example of a Violin Plot

This example uses fuel economy data collected for hybrid cars. You want to get an understanding of how the distributions of highway miles per gallon (Hwy MPG) and city miles per gallon (City MPG) differ by engine type.

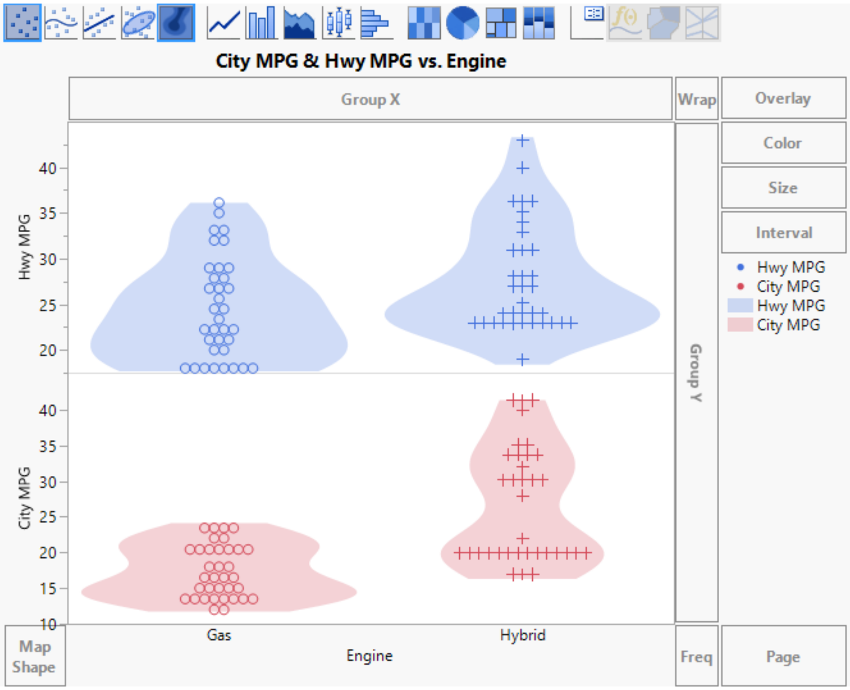
1. Select **Help > Sample Data Library** and open Hybrid Fuel Economy.jmp.
2. Select **Graph > Graph Builder**.
3. Select Engine and drag it to the **X** zone.
4. Select City MPG and drag it to the **Y** zone.
5. Select Hwy MPG and drag it to the top of the **Y** zone.

Figure 4.5 Dragging Hwy MPG to the Y Zone



6. Press Ctrl, select the Contour element , and drag it into the graph.

Figure 4.6 Points and Violin Plot



7. (Optional) Click **Done**.

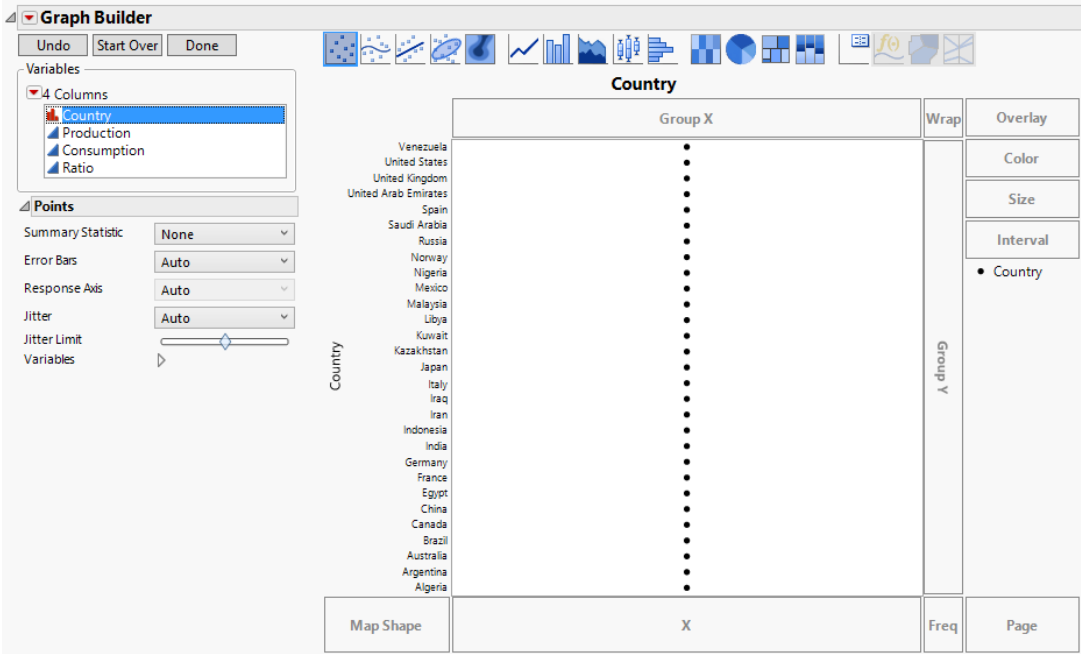
   The violin plots are essentially one-dimensional contour plots. They show density outlines of the observations and give information similar to what a histogram or box plot would provide. In this example, the violin plots show you the individual distributions for city and highway MPG for both engine types. You can see that the difference in City MPG for gas and hybrid vehicles is greater than the difference in Hwy MPG.

Example of a Stacked Bar Chart

In this example, you have data about oil consumption and production for selected countries. You want to visualize oil consumption and production in a meaningful way.

1. Select **Help > Sample Data Library** and open Oil Use.jmp.
2. Select **Graph > Graph Builder**.
3. Select Country and drag it to the **Y** zone.

Figure 4.7 Country Assigned to the Y Zone



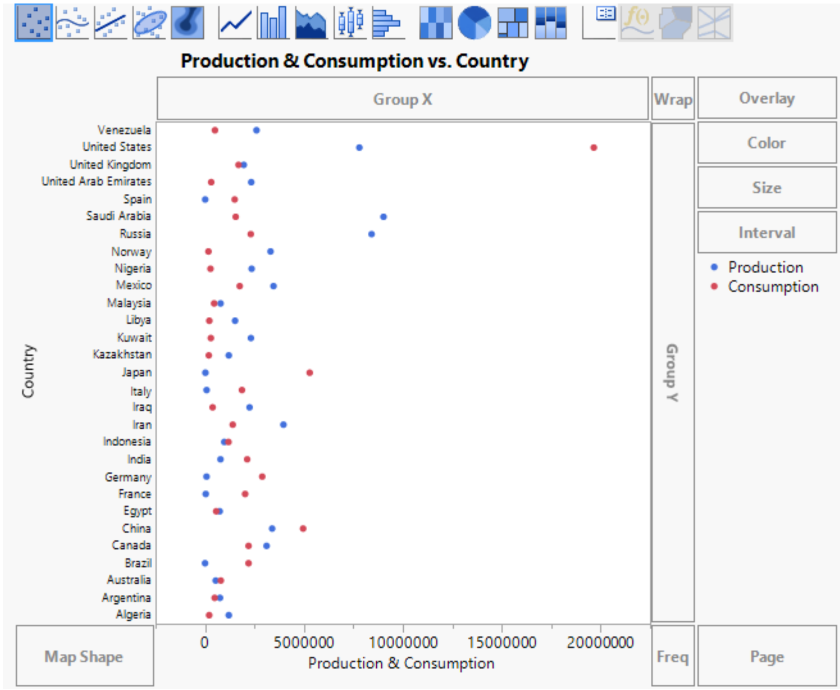
Tip: The countries are ordered in ascending alphabetical (starting point is at the bottom). To change the sorting order, use the Value Order column property or Row Order Levels option. See The Column Info Window chapter in *Using JMP*. Alternatively, you can sort using the values of a continuous variable. See "Order the Levels of a Categorical Variable" on page 48 in the "Graph Builder" chapter.

---

4. Select Production and Consumption and drag them to the **X** zone.

Markers appear on the graph for both variables with a legend identifying the two colors.

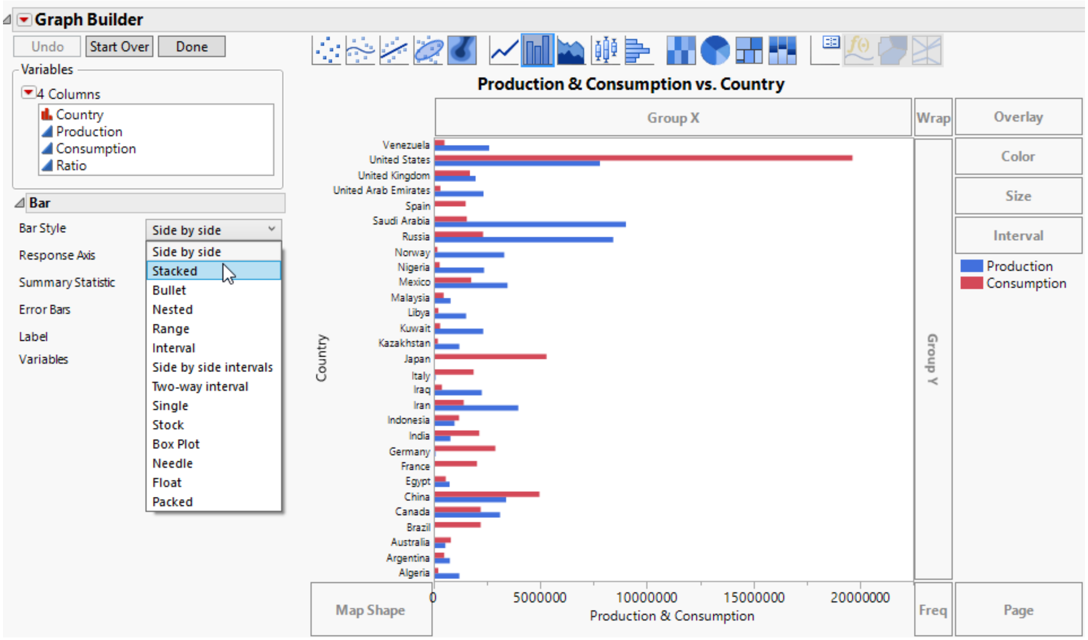
Figure 4.8 Country versus Production and Consumption



Because the default graph element is set to points, it is difficult to relate the labels to the points. Change the points to bars to make the graph easier to interpret.

5. Click the Bar element  .

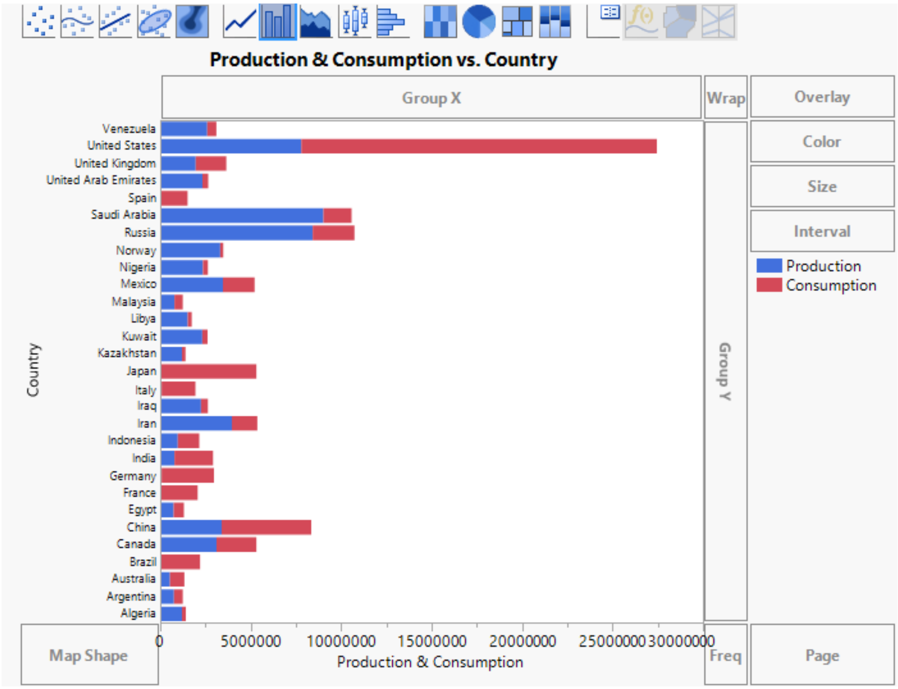
Figure 4.9 Side-by-Side Bars for Production and Consumption



Experiment with the presentation of the bar chart. Change the side-by-side bars to stacked bars.

6. Next to Bar Style, select **Stacked** (Figure 4.9).

Figure 4.10 Stacked Bars for Production and Consumption



Because production is the opposite of consumption, you would like your plot to reflect this by having the production and consumption bars point in opposite directions.

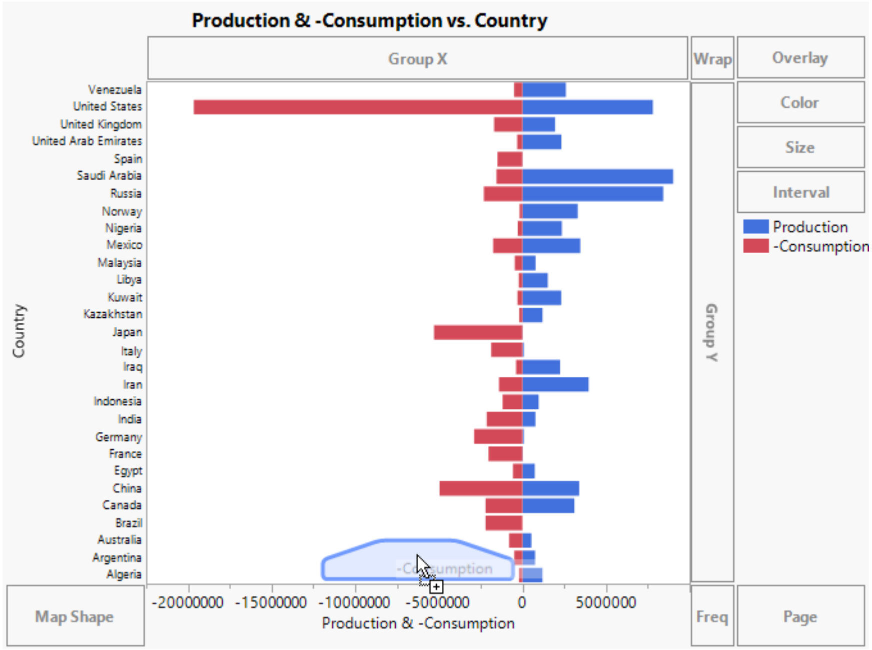
7. In the Variables list, right-click the column Consumption and select **Transform > Negation.**

   This makes Consumption negative. A column called -Consumption appears in the list of columns. The column name is italicized to indicate that it is a virtual column. Use this column to plot the consumption bar in terms of negative values.

8. Right-click the X axis label, Production & Consumption, and select **Remove** > **Consumption**.

9. Select -Consumption and drag it to just above the X axis.

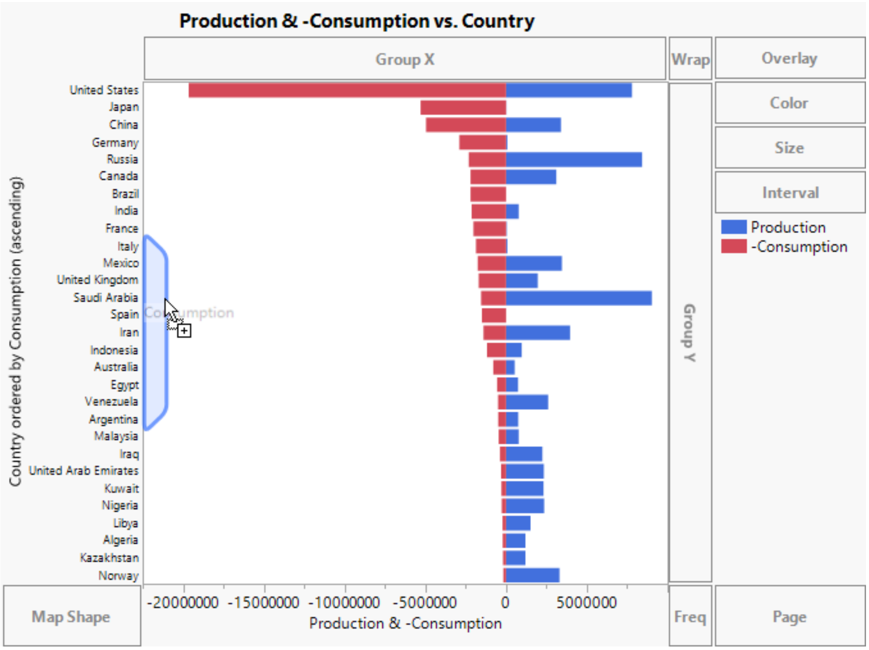
Figure 4.11 Stacked Bars for Production and Negative Consumption



Negative consumption (-Consumption) is plotted on the left in red, and Production is on the right in blue. Next, use the Consumption column as an ordering variable to sort the countries by their oil consumption.

10. Select Consumption and drag it to the right of the **Y** axis. Do not drop the variable until a blue polygon appears.

Figure 4.12 Polygon for Ordering Countries by Consumption



11. (Optional) Click **Done**.

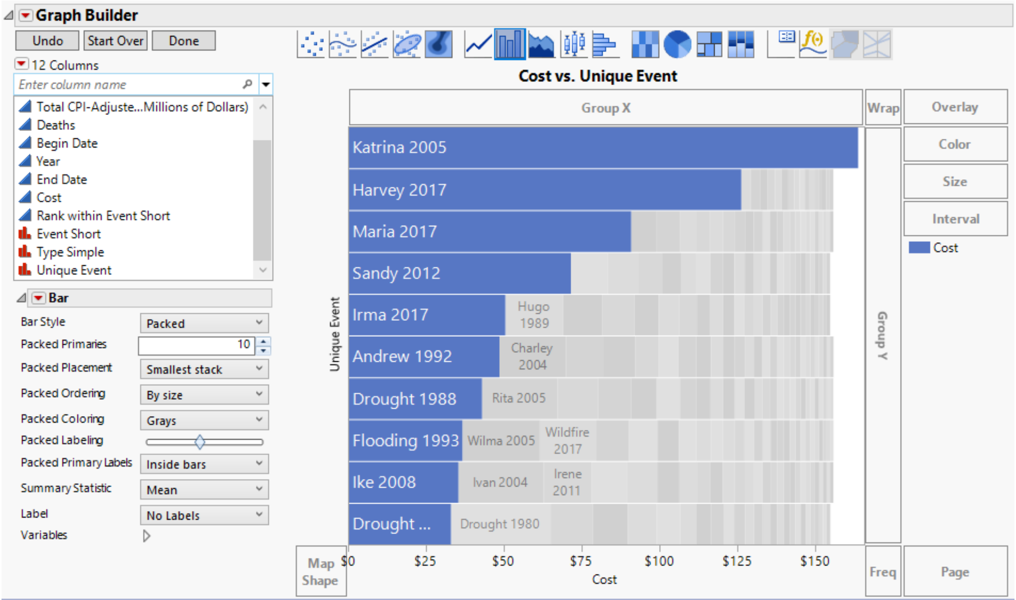
    The levels of Country are now ordered by Consumption. The countries that consume the most oil appear at the top of the graph. You can also see each country's oil production.

Example of a Packed Bar Chart

The following example uses NOAA data on weather and climate events in the U.S. that had the greatest economic impact during the years 1980 to 2018.

1. Select **Help > Sample Data Library** and open Billion Dollar Events.jmp.
2. Select **Graph > Graph Builder**.
3. Select Unique Event and drag it to the **Y** zone.
4. Select Cost and drag it to the **X** zone.
5. Select the Bar element  .
6. In the Bar options panel:
    a. For the Bar Style, select **Packed**.
    b. Change Packed Primaries to 10.
    c. Move the Packed Labeling slider down until it's about halfway.

Figure 4.13 Most Costly Events in a Packed Bar Chart



The top categories appear as a bar chart with blue bars. Secondary categories are labeled and in gray. You can hover over any event for more details.

In this example, you can clearly see that Katrina in 2005 had the largest impact, followed by Harvey and Maria in 2017.

Example of Custom Error Bars

You have conducted an experiment to study popcorn yield. The factors that you studied are popcorn type (plain or gourmet), how much oil was used, and the batch size (small or large). The response is the yield. You want to determine how the factors popcorn type, oil, and batch size affect the popcorn yield.

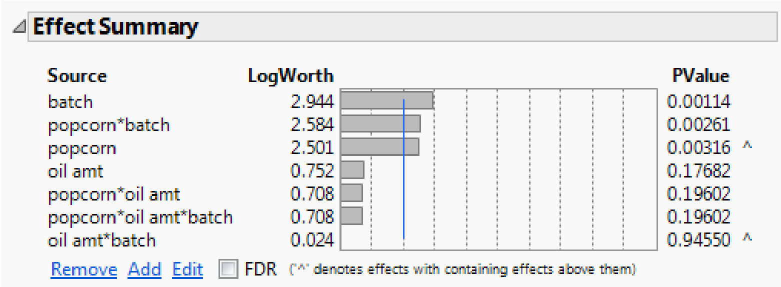
There are eight possible factor level combinations and each combination is tested twice, resulting in 16 values for yield.

Note: These data are artificial, but inspired from an experiment reported in Box, Hunter, and Hunter (1978).

1. Select **Help > Sample Data Library** and open Popcorn.jmp.
2. In the Tables panel at top left, click the green triangle next to **Full Factorial Model**.

   A full factorial model is fit to the data.

Figure 4.14 Effect Summary Report


Effect Summary

Source	LogWorth	PValue	
batch	2.944	0.00114	
popcorn*batch	2.584	0.00261	
popcorn	2.501	0.00316	^
oil amt	0.752	0.17682	
popcorn*oil amt	0.708	0.19602	
popcorn*oil amt*batch	0.708	0.19602	
oil amt*batch	0.024	0.94550	^

Remove Add Edit FDR ('^' denotes effects with containing effects above them)

In the Effect Summary report, notice that the popcorn*batch interaction has a small *p*-value (0.00261). From this, you conclude that there is a significant interaction between popcorn and batch.

3. Click the Response yield red triangle and select **Save Columns** > **Prediction Formula**.

   This saves a column that contains the prediction formula to the data table. The new column is **Pred Formula yield**.

4. Click the Response yield red triangle and select **Save Columns** > **Mean Confidence Interval**.

   This adds two new columns to the data table: **Lower 95% Mean yield** and **Upper 95% Mean yield**.

   Now you can use Graph Builder to visualize the interaction between popcorn and batch.

5. Select **Graph > Graph Builder**.

6. Select these columns and drag them to the **Y** zone:

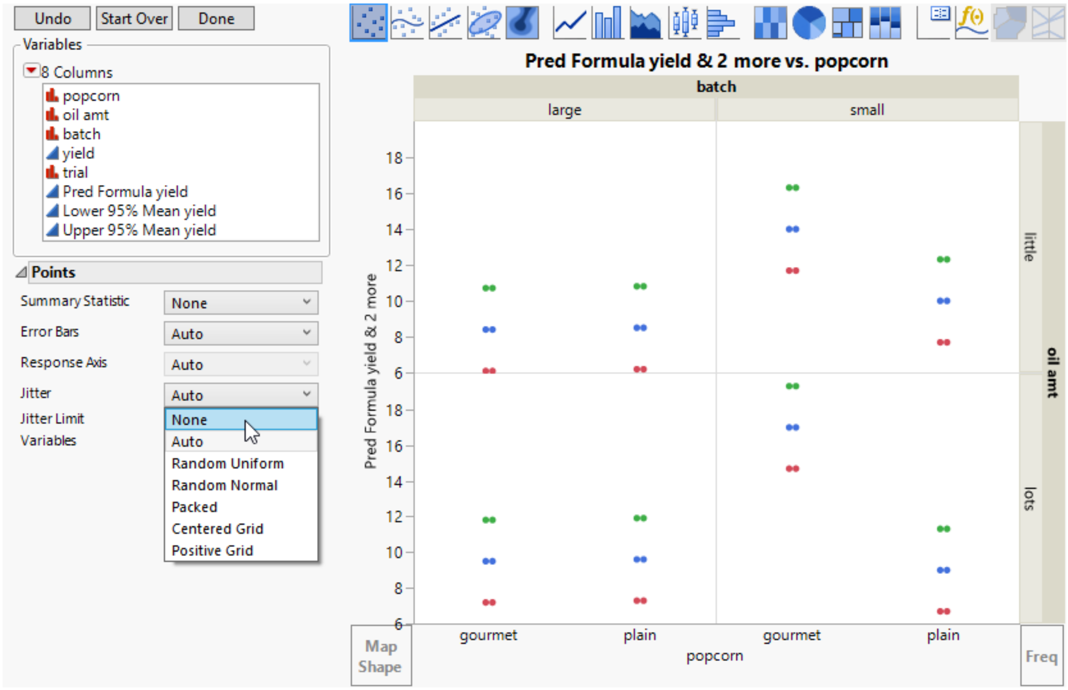
   – Pred Formula yield
   – Lower 95% Mean yield
   – Upper 95% Mean yield

7. Select popcorn and drag it to the **X** zone.

8. Select batch and drag it to the **Group X** zone.

9. Select oil amt and drag it to the **Group Y** zone.

Figure 4.15 Yield Values for the Eight Factor Combinations



Recall that there are two observations at each of eight possible combinations of factors levels. The prediction formula gives the same prediction for each of the two sets of identical factor level combinations. Because the Jitter option is selected, you see two points for each combination of popcorn, batch, and oil amt.

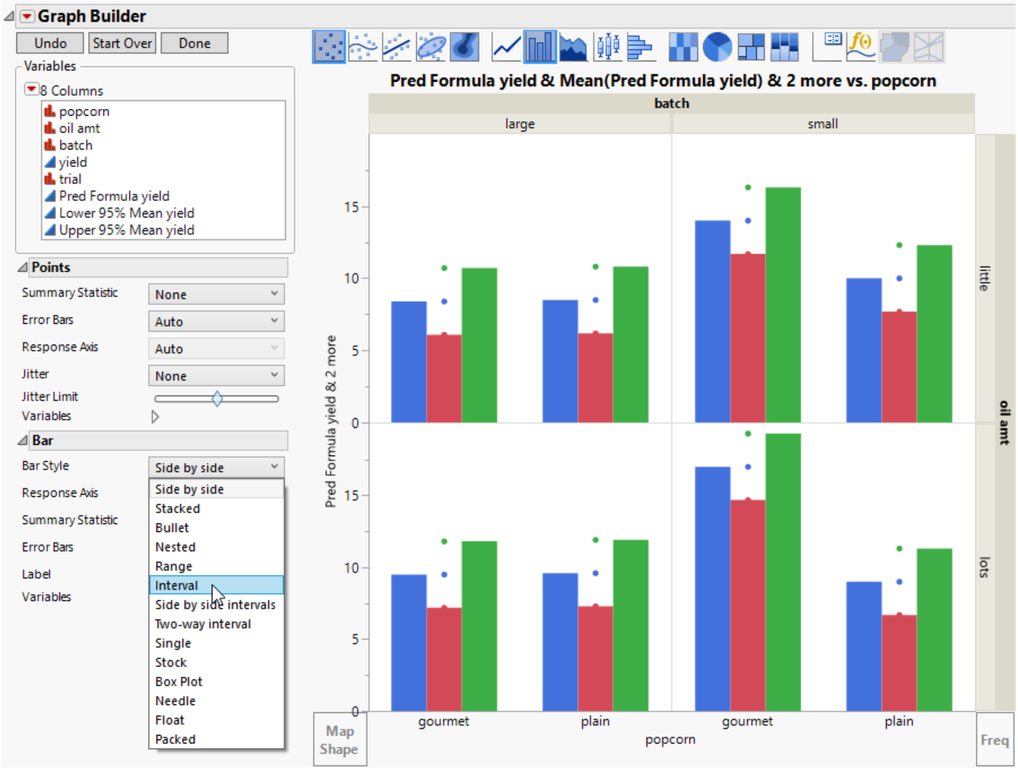
10. Change Jitter to **None** (Figure 4.15).

    Format the graph to see interval bars for Lower 95% Mean yield and Upper 95% Mean yield, and to see points for Pred Formula yield.

11. Select the Bar element  and drag it to the plot.

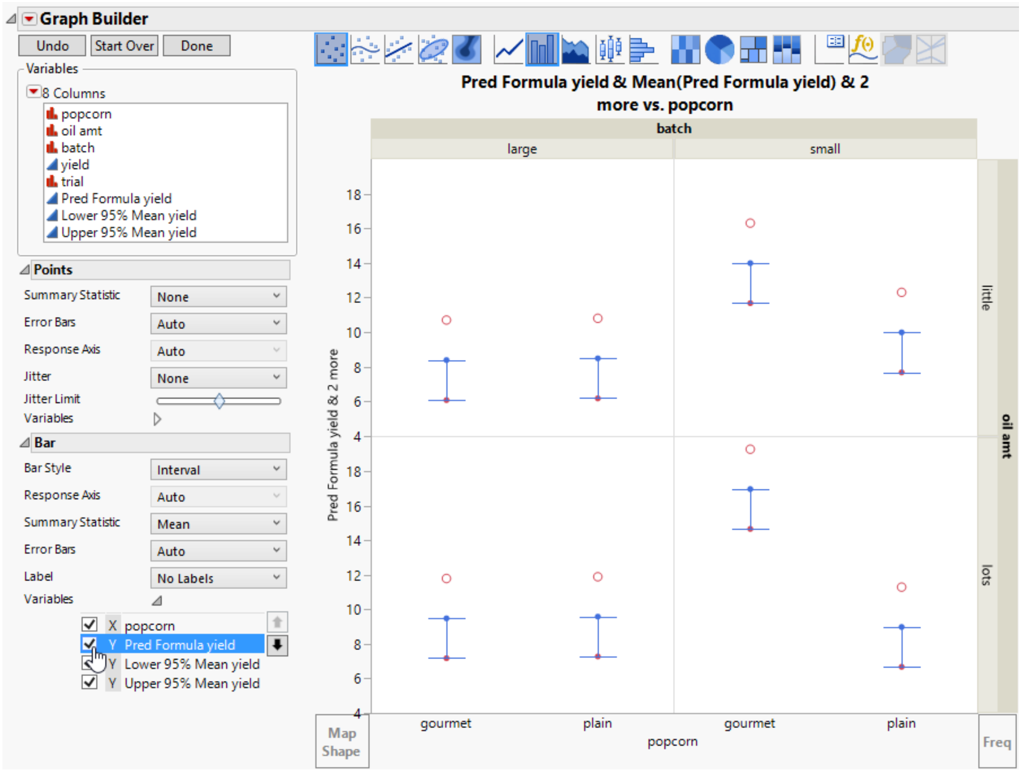
    This adds bars for all three Y-axis variables.

Figure 4.16 Confidence Limits and Means Plotted as Side-by-Side Bars



12. In the Bar options panel, select **Interval** from the Bar Style menu (Figure 4.16).

Figure 4.17 Interval Bar Style Applied to Three-Y-Axis Variables

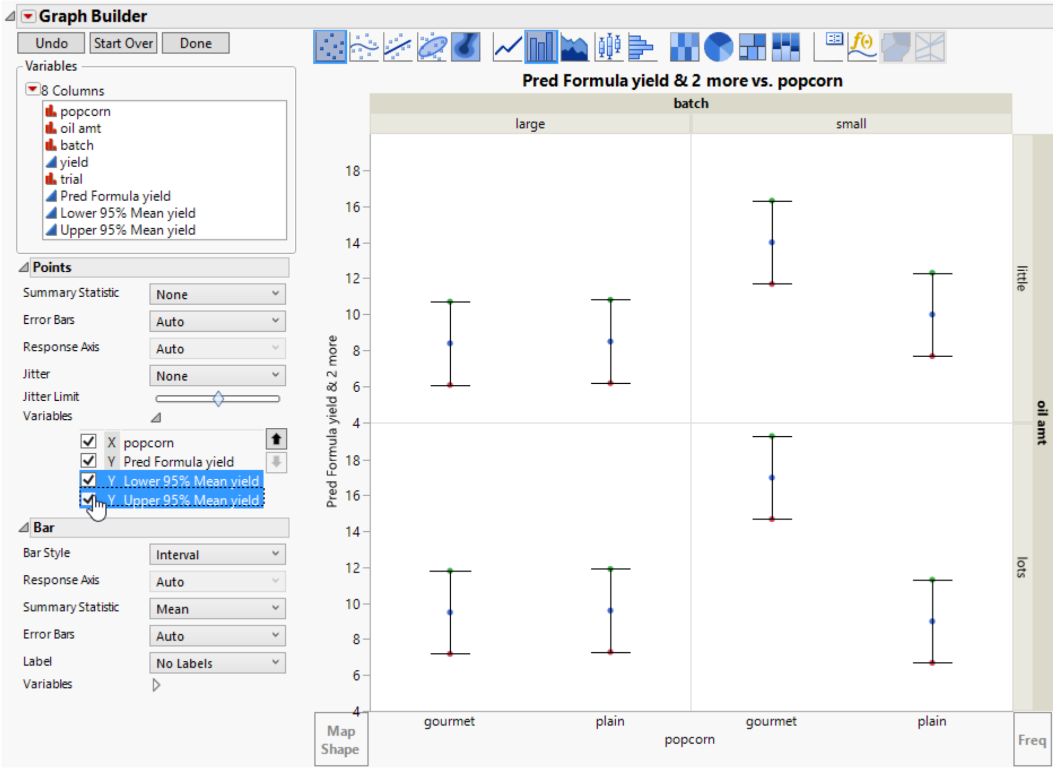


The interval bar style currently spans from Lower 95% Mean yield to Pred Formula yield, but you want it to span up to Upper 95% Mean yield. Remove the bar element for Pred Formula yield.

13. In the Bar options under Variables, deselect Y Pred Formula yield (Figure 4.17.)

    Now the confidence intervals span from their lower to upper values.

Figure 4.18 Interval Span for Confidence Boundaries



Remove the point graph element for Lower 95% Mean yield and Upper 95% Mean yield.

14. In the Points options under Variables, deselect Y Lower 95% Mean yield and Y Upper 95% Mean yield (Figure 4.18).

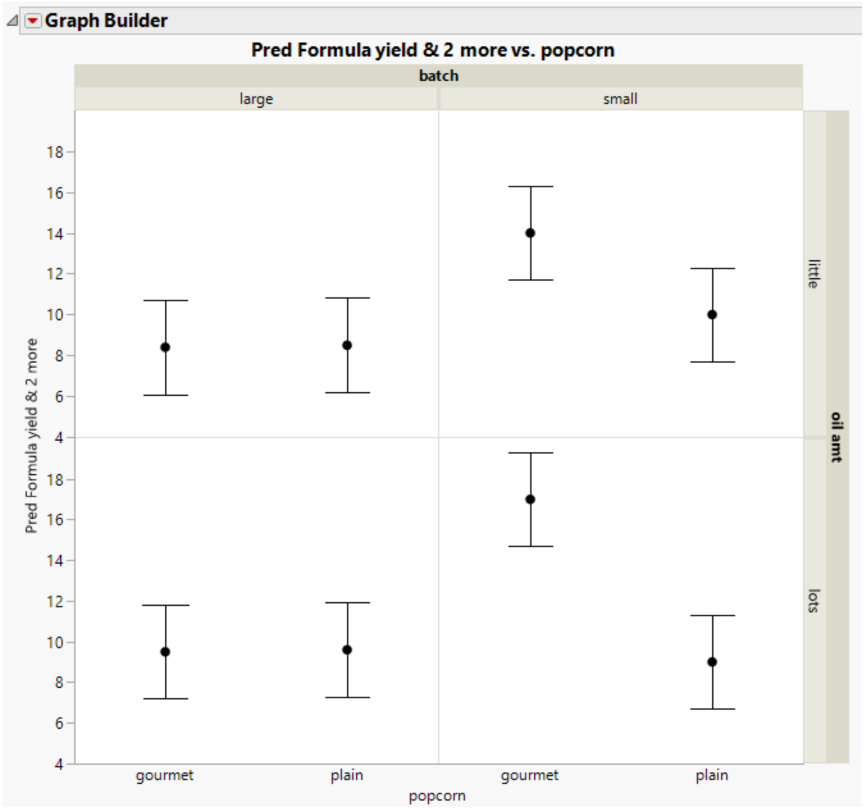
    To make the predicted values easier to see, increase the size of the points.

15. Hold down the Ctrl key and right-click in any quadrant. Select **Graph** > **Marker Size > XXL**.

> **Tip:** You can also change the graph title and labels for the X and Y zones. Click the label and type in the new text.

16. (Optional) Click **Done**.

Figure 4.19 Predicted Means and Confidence Intervals



You can see the following relationships:

- For large batches, there is no difference between plain and gourmet popcorn.
- For small batches, the gourmet popcorn has a higher yield than the plain popcorn.
- For each level of oil amount, the relationship between popcorn type and batch size is the same. There is no evidence of a three-way interaction, as you would expect, since the Effect Summary report indicates that the three-way interaction is not statistically significant.

Examples of Area Plots

The examples in this section use area plots to show behavior over a period of time, and to show upper and lower bounds around group means.

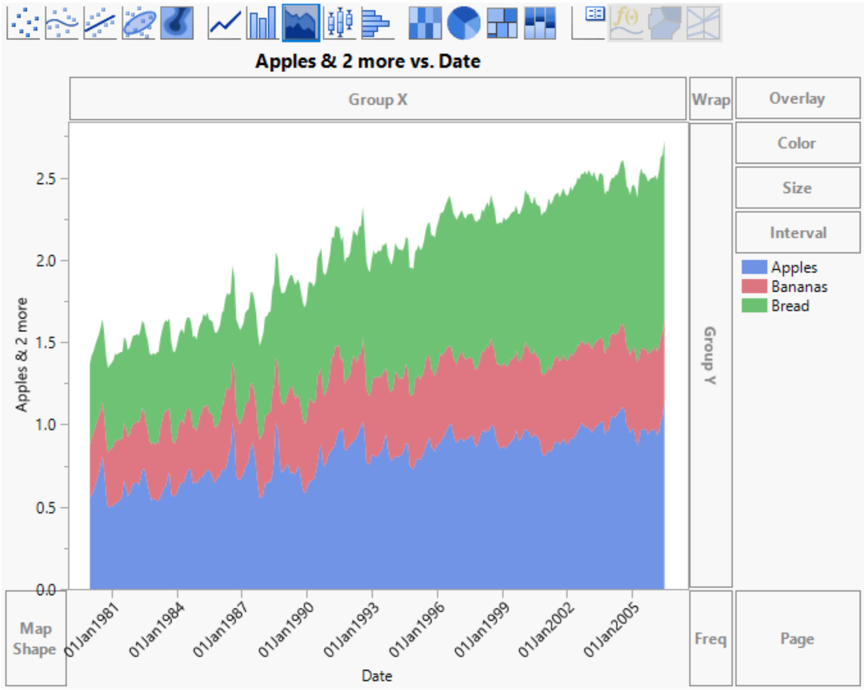
Show Behavior over Time

This example shows the behavior of average consumer prices for three items from January 1980 to July 2006. If you want to emphasize the lines in the area plot rather than the filled areas, you can create this plot using the Line element. See “Show Behavior over Time - Line Element” on page 138.

1. Select **Help > Sample Data Library** and open Consumer Prices.jmp.
2. Click the green triangle next to the **Split into Series** script.

   A new data table called Consumer Prices Split by Series appears. This is the data table to use for the remainder of the example.
3. Select **Graph > Graph Builder**.
4. Select Date and drag it to the **X** zone.
5. Select Apples, Bananas, and Bread and drag them to the **Y** zone.
6. Click the Area element .

Figure 4.20 Area Plot for Apples, Bananas, and Bread by Date



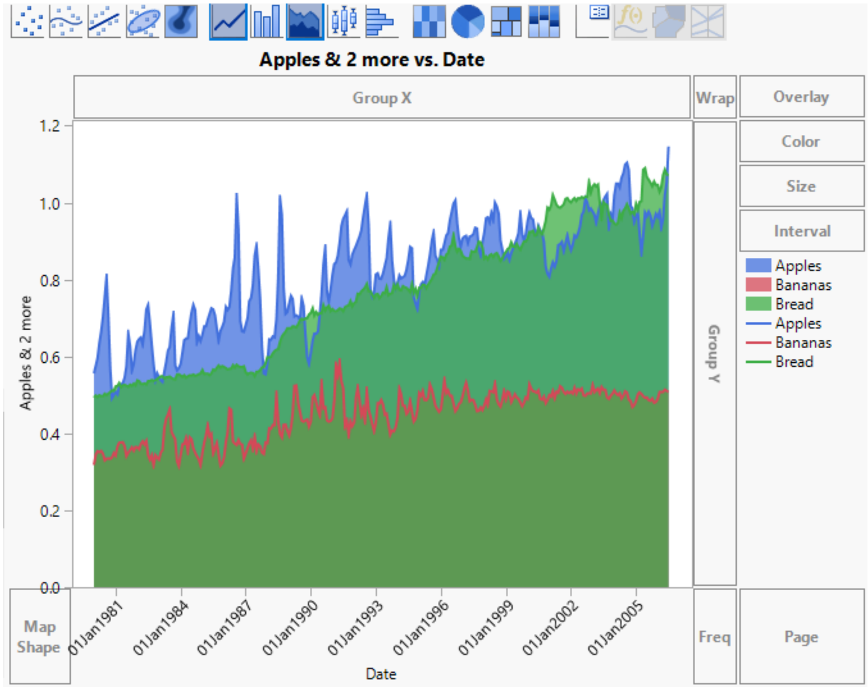
Stacked area plots appear for all three responses. You can see that apple and bread prices have increased over time, and that banana prices have increased, but to a smaller degree. You would like to be able to compare the three series in a way that makes comparisons easier.

7. Change the Area Style to **Overlaid**.

   Because banana prices are uniformly smaller than prices for apples and bread, the layer for Bananas is at first entirely hidden behind the other two layers. The apples price is hidden behind the bread price for some months. Use the Line element to see prices for all three clearly.

8. Select the Line element  and drag it into the graph.

Figure 4.21 Overlaid Area Style



The layers for the plot are ordered by the order of entry of variables, as shown in the legend. The graph shows that banana prices are uniformly smaller than both apple and bread prices. On average, apple prices dominate bread prices up to the year 2000. After this, bread prices are generally higher than apple prices. Also note that the price of bread increases at a smoother rate than prices for apples and bananas. The line plots for Apples and Bananas are very jagged, with a lot of spikes and dips from year to year.

9. In the Area options, click the triangle next to Variables. Select Y Bananas so that it is highlighted, and click the down arrow once.

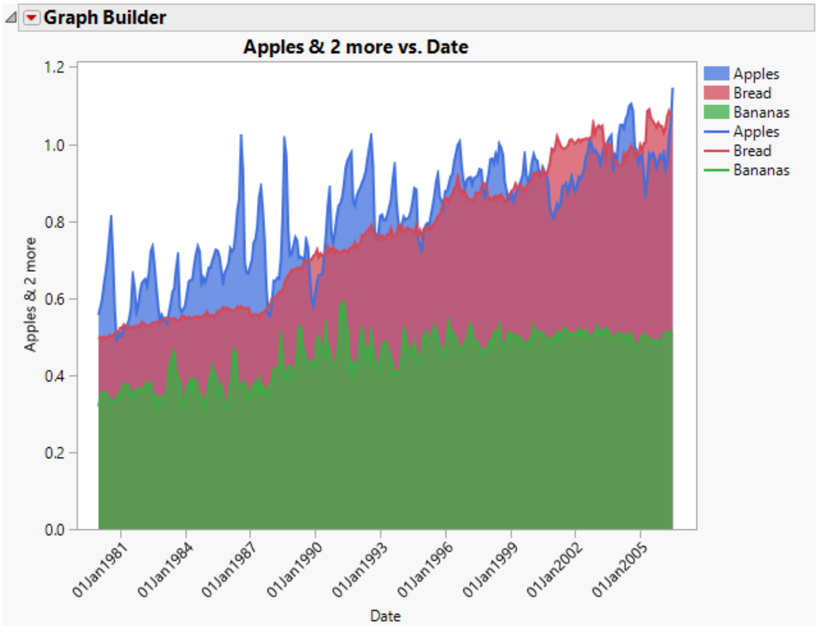
Note: When selecting Y Bananas, be sure to leave the check box marked. Otherwise, the variable is removed from the graph.

---

10. In the Line options, click the triangle next to Variables. Select Y Bananas so that it is highlighted, and click the down arrow once.

11. (Optional) Click **Done**.

Figure 4.22 Overlaid Area Showing All Layers



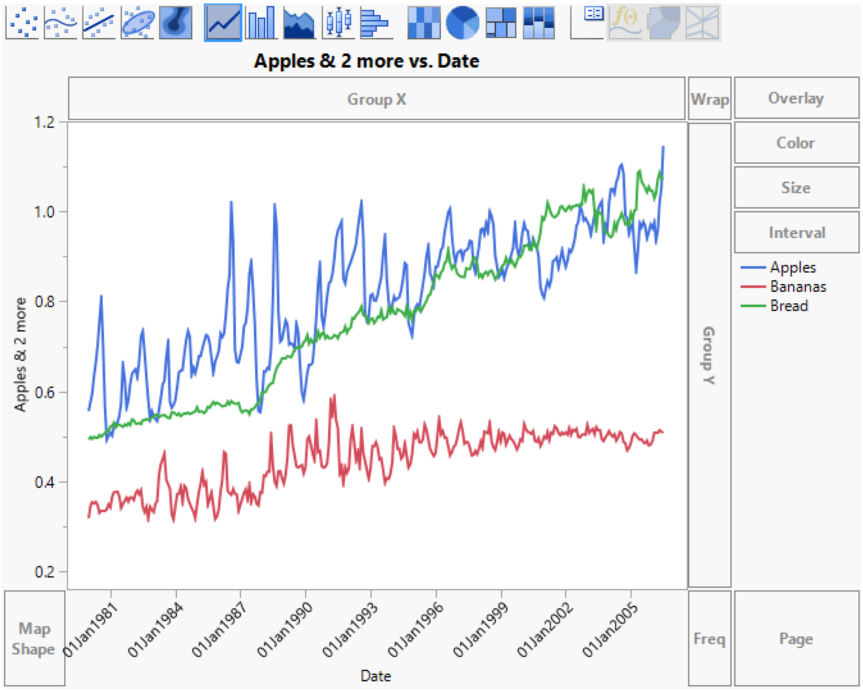
By re-ordering the variables in the legend and using the Line element, you can show prices for the three series with area plots in a way that does not obscure any of the underlying data.

Show Behavior over Time - Line Element

1. Select **Help > Sample Data Library** and open Consumer Prices.jmp.
2. Click the green triangle next to the **Split into Series** script.

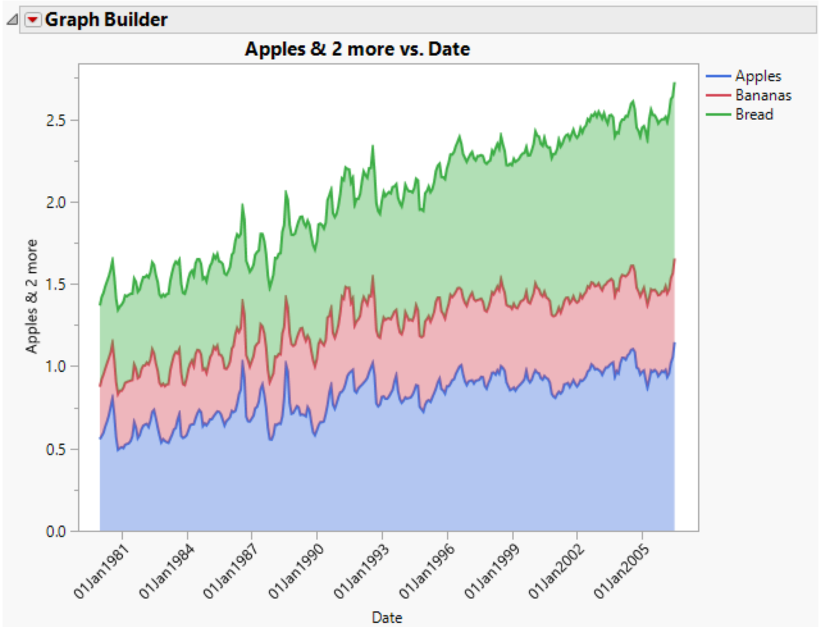
   A new data table called Consumer Prices Split by Series appears. This is the data table to use for the remainder of the example.
3. Select **Graph > Graph Builder**.
4. Select Date and drag it to the **X** zone.
5. Select Apples, Bananas, and Bread and drag them to the **Y** zone.
6. Click the Line element .

Figure 4.23 Line Plot for Apples, Bananas, and Bread by Date



7. In the Line options, select **Stack**, and change the Fill to **Fill Below**.
8. (Optional) Click **Done**.

Figure 4.24 Area Plot for Apples, Bananas, and Bread by Date using Line Element



Show Bounds around Group Means

You are interested in how the height of teenagers varies by age. In this example, you use the Area element to show upper and lower bounds. You use custom transformations to create virtual columns, and virtual columns to create a graph.

Create Formula Columns Using Transform

To create upper and lower bounds for height by age group, use the Formula option.

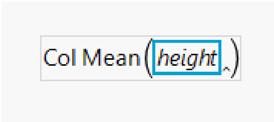
- Define your upper bound to be the standard deviation within an age group added to the mean of that age group.
- Define the lower bound to be the standard deviation within an age group subtracted from the mean of that age group.

Create the Upper Bound for Height

1. Select **Help > Sample Data Library** and open Big Class.jmp.
2. Select **Graph > Graph Builder**.
3. Right-click the column height and select **Formula.**
4. Select **Statistical > Col Mean**.

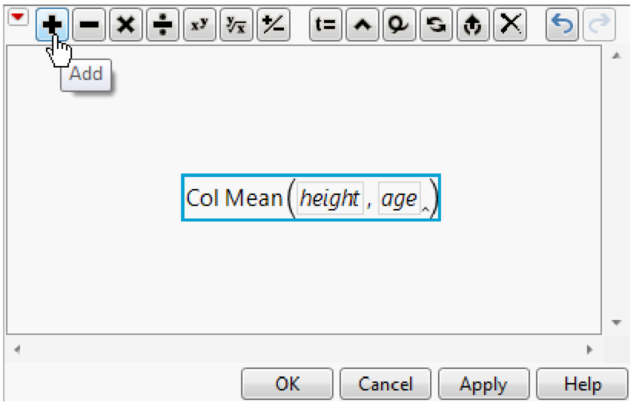
5. In the formula, click height so that the blue box is around only the variable name.

Figure 4.25 Select the height Variable



6. Enter a comma, and a byVariable box appears. Click age.
7. In the formula, select the outer box and click the plus sign.

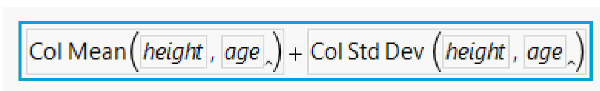
Figure 4.26 Click the Plus Sign



8. Click **Statistical > Col Std Dev**.
9. Click height.
10. Enter a comma and a byVariable box appears. Click age.

    Your formula for the upper bound should look like Figure 4.27.

Figure 4.27 Formula for Upper Bound



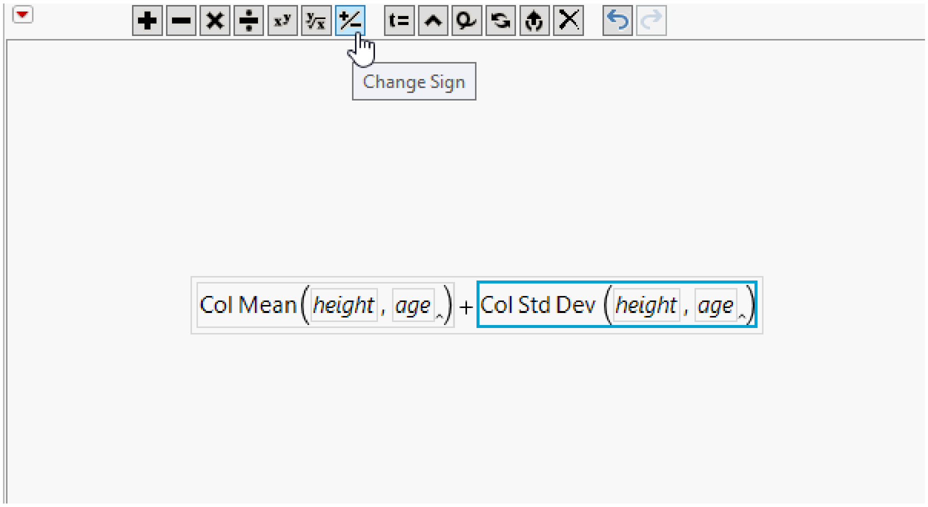
11. Select the entire formula, right-click the outer box, and select **Copy**.
12. Click **OK**.
13. In the Variables list, right-click the Transform[height] column and select **Rename**.
14. Type upper and click **OK**.

Create the Lower Bound for Height

1. In the Variables list, right-click the column height and select **Formula.**
2. In the formula, right-click height and select **Paste**.

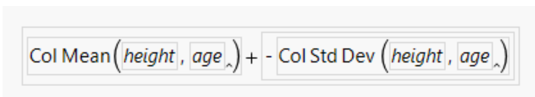
   The formula that you created for the upper bound appears.
3. Select the box around Col Std Dev (*height*, *age*) and click the +/- sign.

Figure 4.28 Selecting Col Std Dev and Making It Negative



   Your formula for the lower bound should look like Figure 4.29.

Figure 4.29 Formula for Lower Bound



4. Click **OK**.
5. In the Variables list, right-click the Transform[height] column and select **Rename**.
6. Type lower and click **OK**.

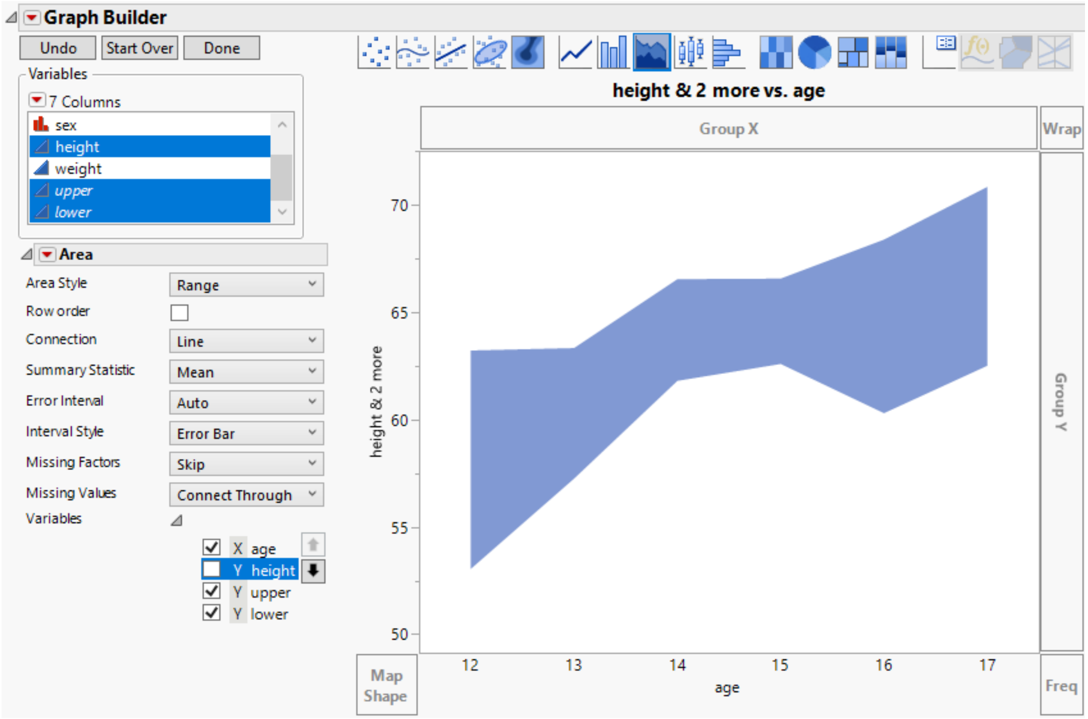
Build Graph

Using a combination of elements in Graph Builder, you can create a line graph of the mean height by age group, with shading to indicate the upper and lower bounds.

1. Select age and drag it to the **X** zone.
2. Select height, upper, and lower and drag them to the **Y** zone.

3. Click the Area element 
4. In the Area options, change the Area Style to **Range**.
5. In the Area options, click the triangle next to Variables. De-select Y height.

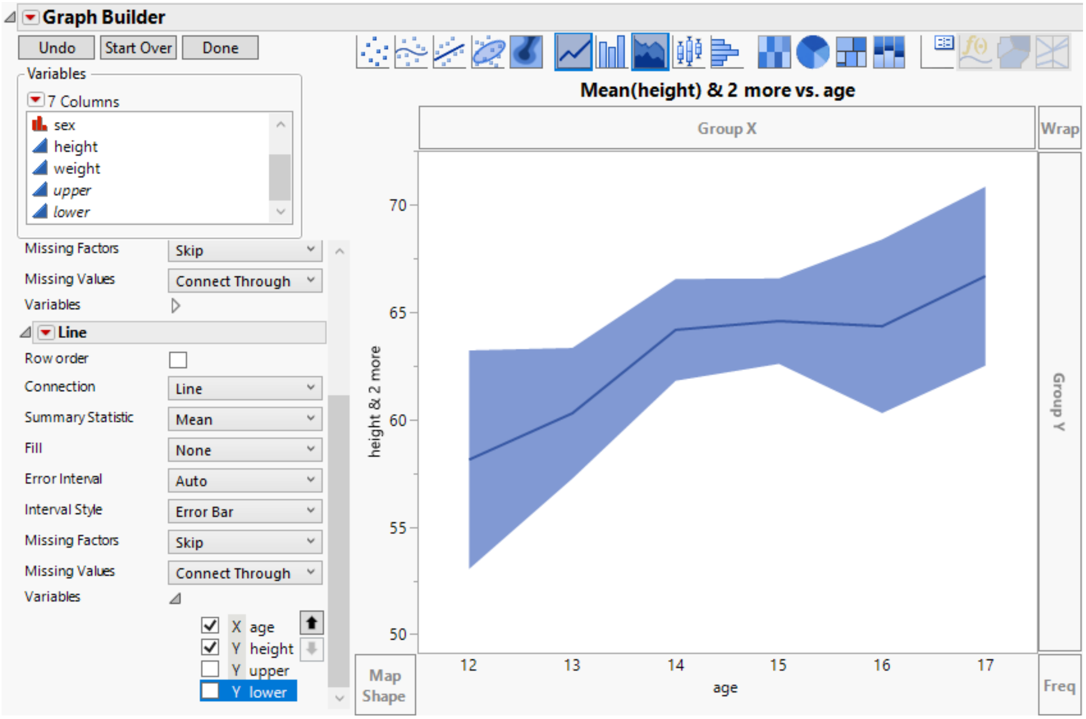
Figure 4.30 Area Plot



The shading indicates the range between the lower bound and upper bound for each age group. However, as is, you cannot see the means are for each group.

6. Press Ctrl and drag the Line element , into the graph.
7. In the Line options, click the triangle next to Variables. De-select Y upper and Y lower.

Figure 4.31 Area Plot with Mean Line



8. (Optional) Click **Done**.

   The dark blue line connects the mean heights for each age group. As expected, height tends to increase as age increases. The shaded area shows a range of one standard deviation above and one standard deviation below the mean height at each age group.

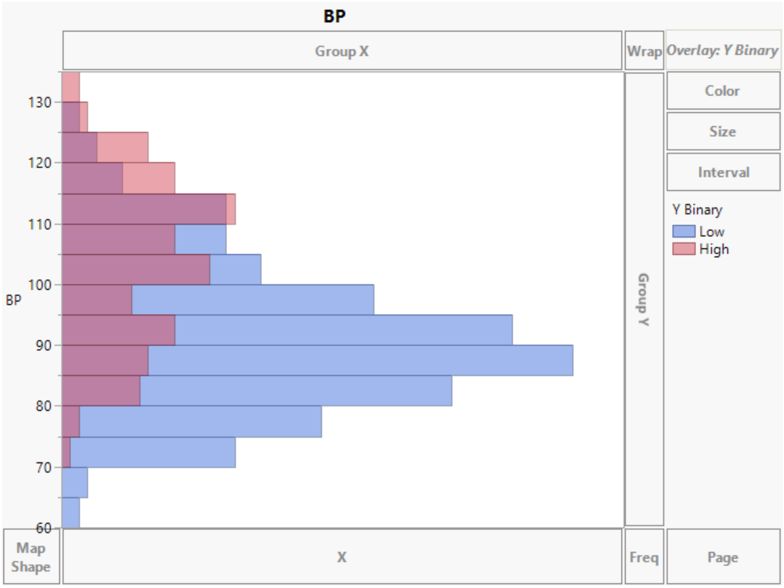
Example of Overlaying Histograms with Transparency

You are interested in finding out how blood pressure is related to a measure of diabetes disease progression. Progression can be Low or High. To see the relationship between blood pressure and disease progression, overlay histograms for these two responses.

1. Select **Help > Sample Data Library** and open Diabetes.jmp.
2. Select **Graph > Graph Builder**.
3. Select BP and drag it to the **Y** zone.
4. Select Y Binary and drag it to the **Overlay** zone.

5. Click the **Histogram** element icon  .

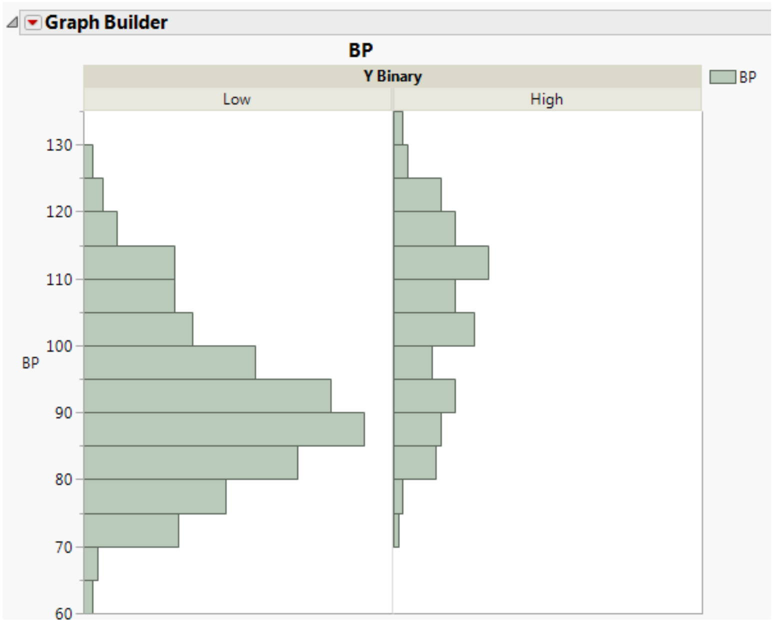
Figure 4.32 Blood Pressure Histograms Overlaid by the Levels of Y Binary



The histograms indicate that higher blood pressure readings are associated with the High level of disease severity.

6. Click the Overlay zone and drag **Y Binary** to the Group X zone.
7. (Optional) Click **Done**.

Figure 4.33 Individual Blood Pressure Histograms for the Levels of Y Binary



Now you see individual histograms for each of the severity levels.

Overlaid histograms can be useful in seeing overlap, as well as differences, in responses over a small number of categories.

Example of Wafer Maps Based on a Cluster Analysis

In this example, you have defect data on wafers. The X Die and Y Die columns give a location on the die, and Defects gives the number of defects at each location. Because die defects often occur in patterns, you use cluster analysis to cluster wafers into groups that might have the same type of defect pattern. Once you have done this, you want to visualize the defect patterns for each cluster.

1. Select **Help > Sample Data Library** and open Wafer Stacked.jmp.
2. Run the **Spatial Cluster of Defects** script.

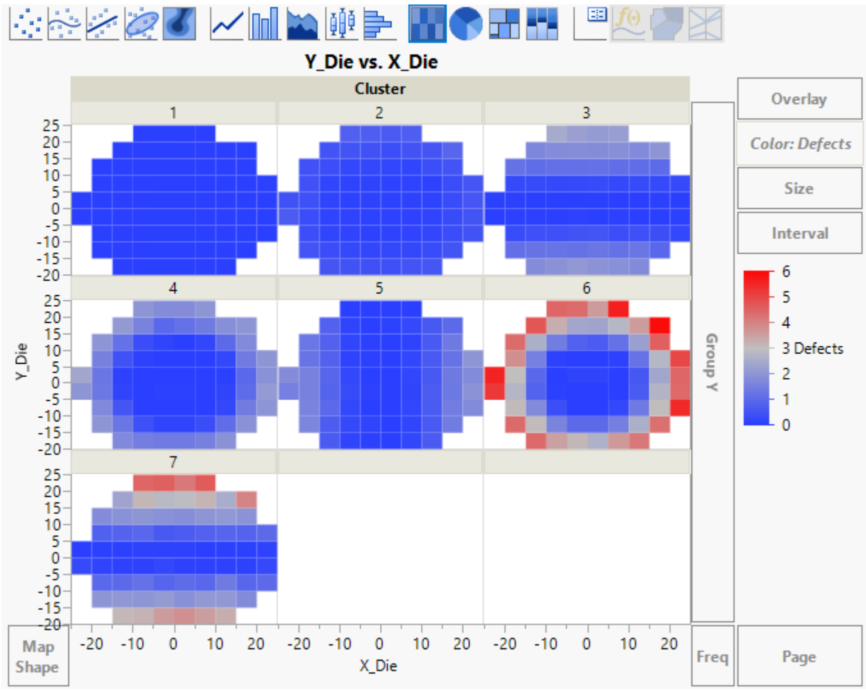
   This script fits a hierarchical cluster analysis model that includes spatial measures (Angle, Radius, Streak Angle, and Streak Distance). See the Hierarchical Cluster chapter in *Multivariate Methods*.
3. Click the Hierarchical Clustering red triangle and select **Save Clusters**.

   Seven clusters are constructed. The cluster numbers are saved to a column called Cluster in the Wafer Stacked.jmp data table.

4. Go back to Wafer Stacked.jmp.
5. Select **Graph > Graph Builder**.
6. Select X_Die and drag it to the **X** zone.
7. Select Y_Die and drag it to the **Y** zone.
8. Deselect the Smoother element  and click the Points element .
9. Select Cluster and drag it to the **Wrap** zone.

   Because there are seven clusters, there are seven plots.
10. Select Defects and drag it to the **Color** zone.
11. Click the Heatmap element .

Figure 4.34 Locations Colored by Mean Defects for Each Cluster



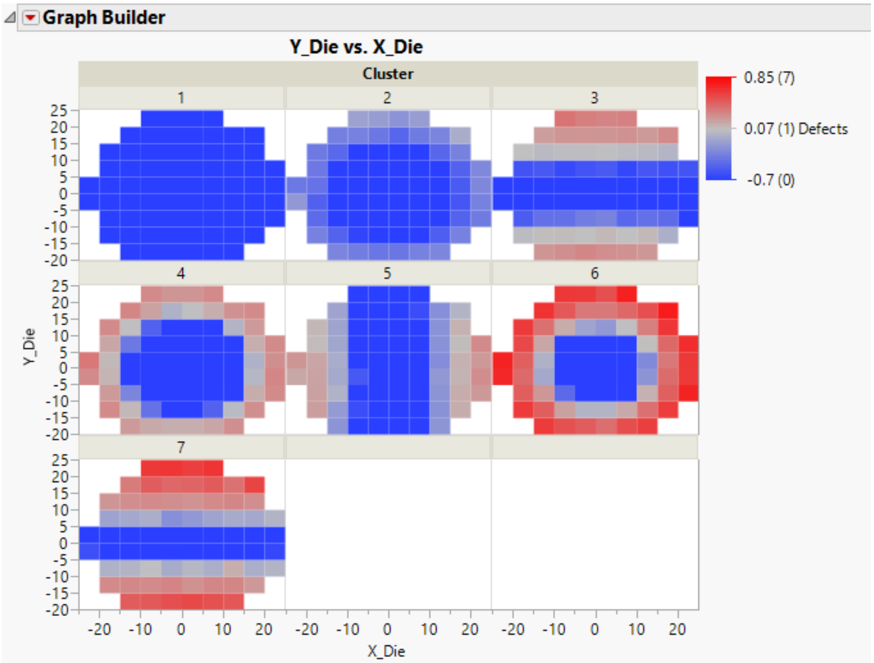
The heatmap is colored by the mean number of Defects.

12. Right-click the color bar in the legend and select **Gradient**.
13. Select **Log** for Scale Type.

    This scale is linear for the logarithm of Defects. Because the distribution of Defects is highly skewed, using the Log scale highlights the variation patterns, but at the expense of being able to easily compare magnitudes in the original scale.
14. Click **OK**.

15. (Optional) Click **Done**.

Figure 4.35 Heatmap for Wafer Clusters



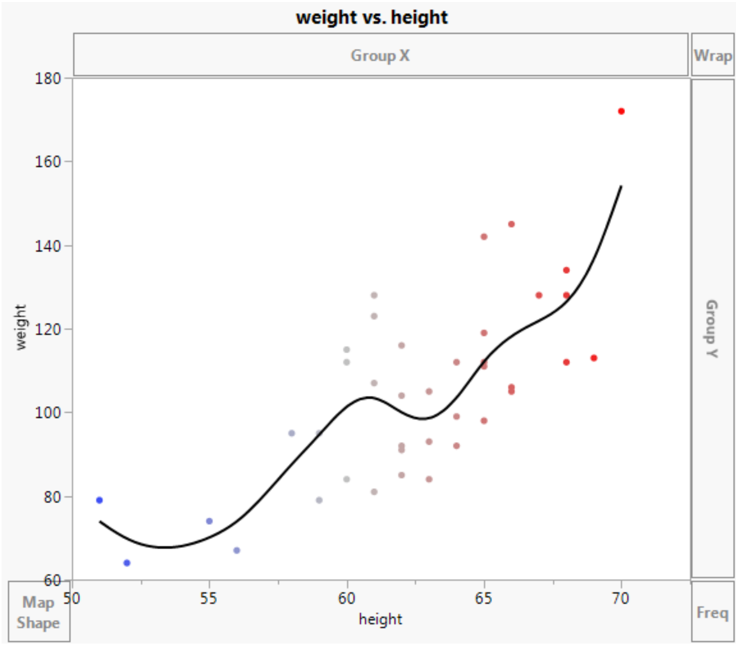
The plots show various patterns of defects. Notice that these same plots also appear in the Hierarchical Clustering report's Cluster Summary outline.

Example Using a Heatmap to Apply Background Colors

You can use a heatmap to apply a background color when you use the Wrap or Group zones. In this example, using height, weight, and age data from children, you color age panels using a height legend.

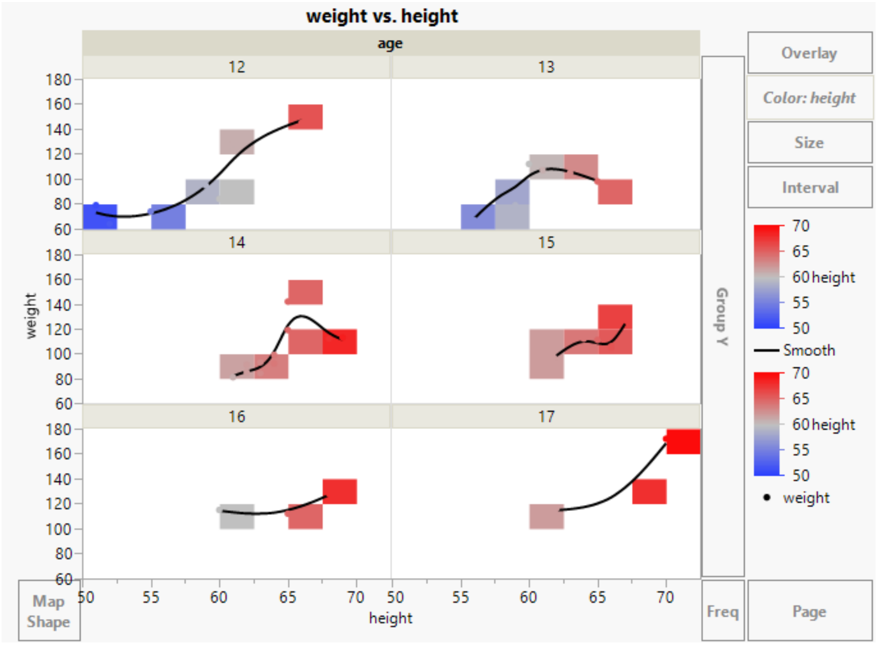
1. Select **Help > Sample Data Library** and open Big Class.jmp.
2. Select **Graph** > **Graph Builder**.
3. Select height and drag it to the **X** zone.
4. Select weight and drag it to the **Y** zone.
5. Select height and drag it to the **Color** zone.

Figure 4.36 Graph of height and weight Colored by height



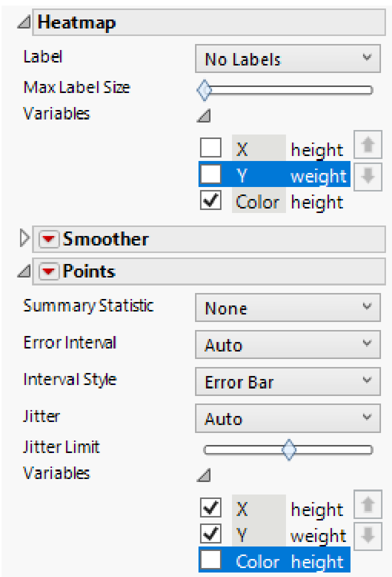
6. Select the Heatmap element .
7. Drag the Smoother element  into the plot.
8. Drag the Points element  into the plot.
9. Select age and drag it to the **Wrap** zone.

Figure 4.37 Heatmap, Smoother, and Points, Wrapped by age



10. In the Heatmap options, click **Variables**. De-select X height and Y weight.
11. In the Points options, click **Variables**. De-select Color height.

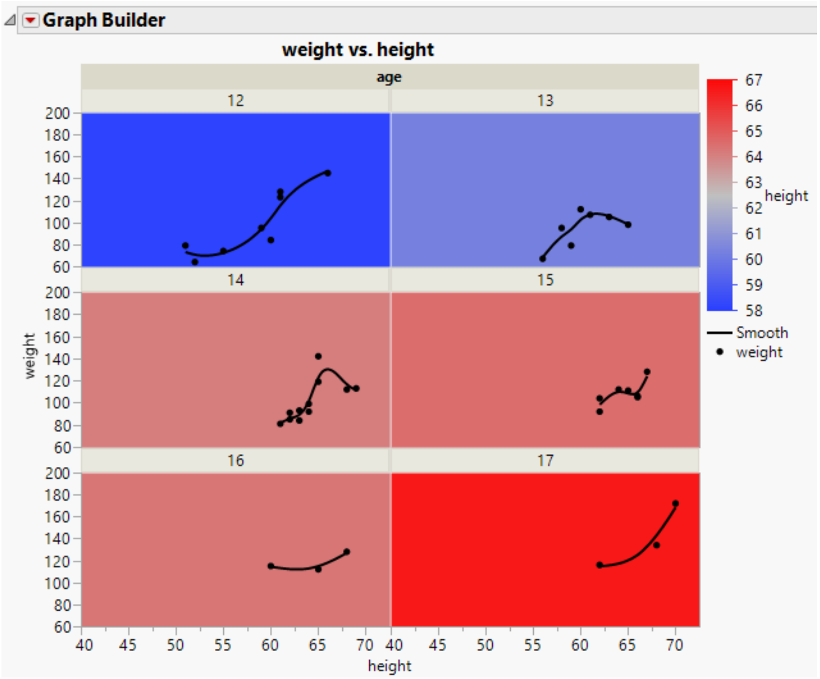
Figure 4.38 Variables Options



12. Double-click the X axis for height. In the X Axis Settings window, change the Minimum to 40, and click **OK**.
13. Double-click the Y axis for weight. In the Y Axis Settings window, change the Maximum to 200, and click **OK**.

13. (Optional) Click **Done**.

Figure 4.39 Scatterplots by Age with Background Color Showing Height



The panel colors are based on the average heights for students in each age group. As expected, the tallest students are the 17-year-olds. The plot in that panel shows how their weights vary by height.

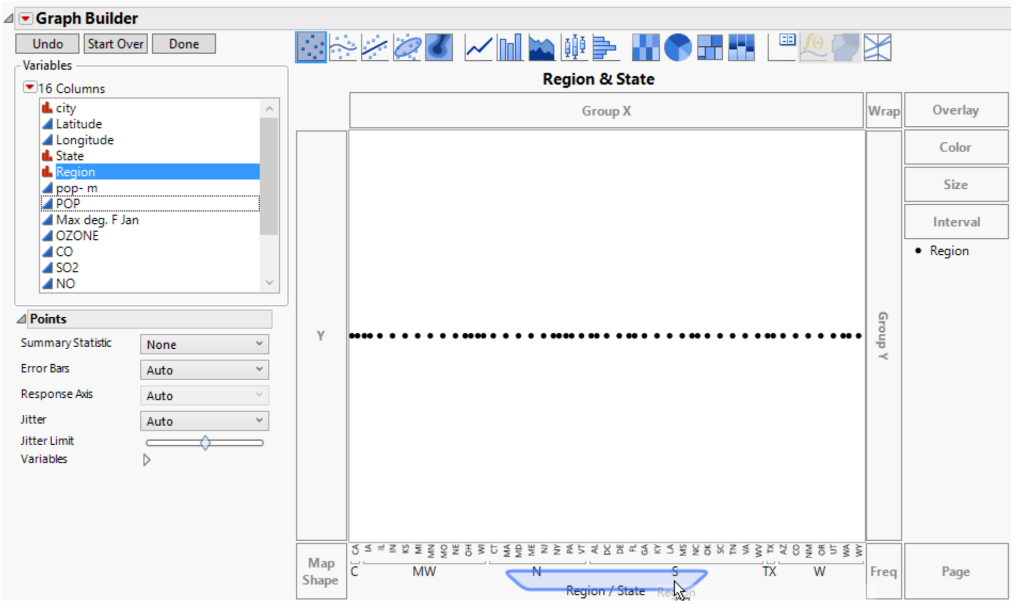
Example of a Treemap

In this example, you are interested in comparing the population of various states by region.

1. Select **Help > Sample Data Library** and open Cities.jmp.
2. Select **Graph > Graph Builder**.
3. Select State and drag it to the **X** zone.
4. Select Region and merge it into the **X** zone.

This makes Region a supercategory, so that states are grouped within Region.

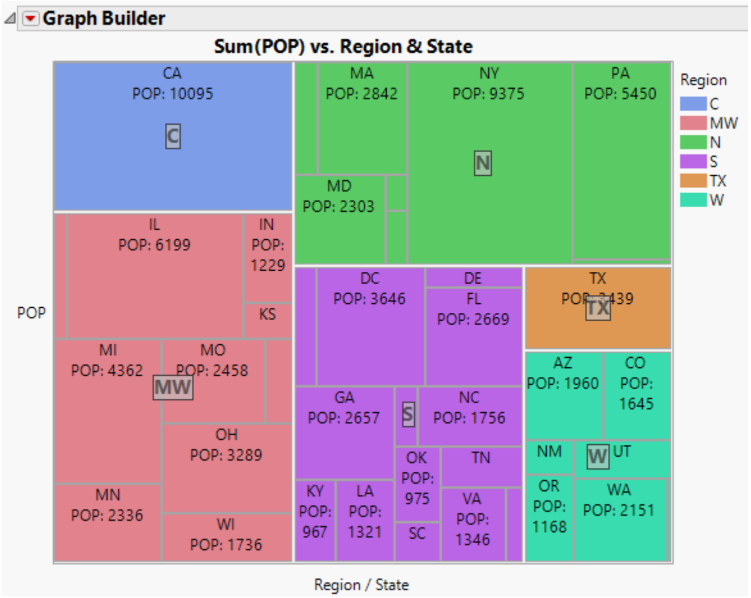
Figure 4.40 Region Merged with State



5. Select POP and drag it to the **Y** zone.
6. Click the Treemap element  .
7. Select Region and drag it to the **Color** zone.
8. Under the Treemap options, check the box next to **Size Value**.

   The Summary Statistic is set to Sum by default. This means that the rectangles are sized according to the sum of the population values in the cities within each state and region. According to this criterion, the most populated states within each region have the largest rectangles. The rectangle sizes for the regions also show their relative sizes in terms of population.
9. (Optional) Click **Done**.

Figure 4.41 City Populations by State Grouped within Region



You can easily compare the population of each state according to region.

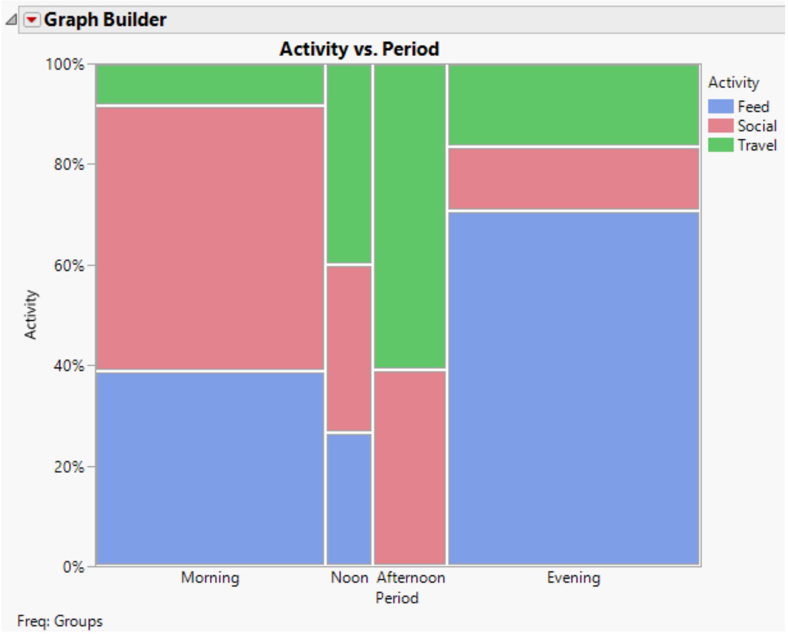
Tip: To deselect a Tree Map selection, use control-click (Windows) or command-click (macOS).

Example of a Mosaic Plot

In this example, you are interested in whether there is a relationship between the time of day and the activity of dolphins. Your data consist of Activity and Period (time of day) combinations, and the numbers of Groups of dolphins observed for each combination. Notice that Groups is a frequency column.

1. Select **Help > Sample Data Library** and open Dolphins.jmp.
2. Select **Graph > Graph Builder**.
3. Select Period and drag it to the **X** zone.
4. Select Activity and drag it to the **Y** zone.
5. Select Groups and drag it to the **Freq** zone.
6. Click the Mosaic element .
7. (Optional) Click **Done**.

Figure 4.42 Mosaic Plot for Dolphin Activity



You can draw many conclusions from this graph. For example, the largest numbers of groups were observed in the morning or evening. There was proportionately more feeding activity in the evening than at any other time of day. No groups were observed feeding in the afternoon. Social behavior tended to occur in the morning.

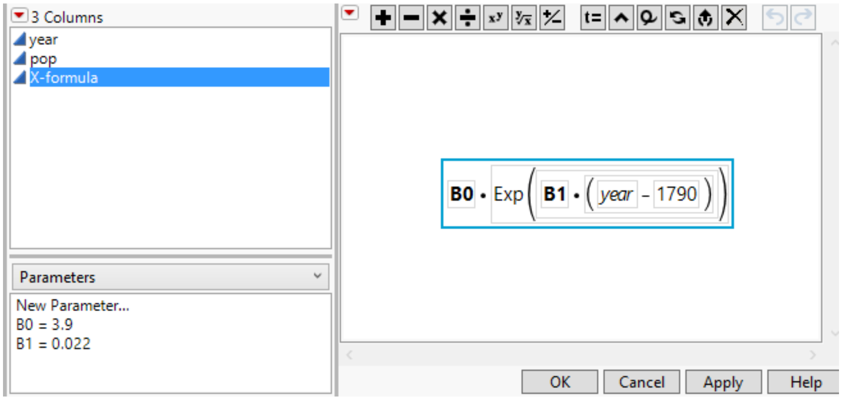
Example Using a Formula

In this example, you have data on the US population between 1790 and 2000. You fit a nonlinear model to the US population as a function of year and you want to plot this nonlinear formula.

1. Select **Help > Sample Data Library > Nonlinear Examples** and open US Population.jmp.
2. In the Columns panel, click the formula icon  next to X-formula.

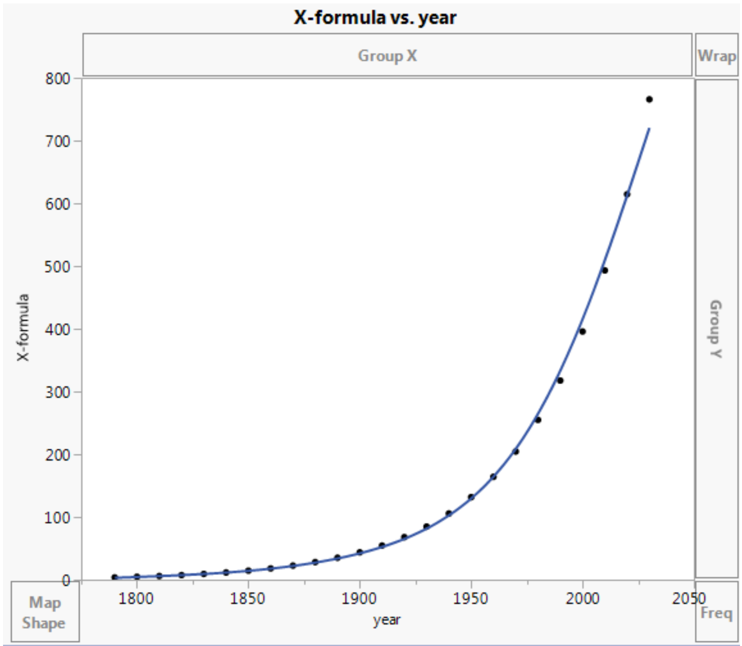
   The formula appears in the formula editor panel. The values of the parameters B0 and B1 appear in the Parameters panel on the left. Use the function element to plot this formula.

Figure 4.43 Formula for Nonlinear Fit to Population



3. Click **Cancel**.
4. Select **Graph > Graph Builder**.
5. Select year and drag it to the **X** zone.
6. Select X-formula and drag it to the **Y** zone.

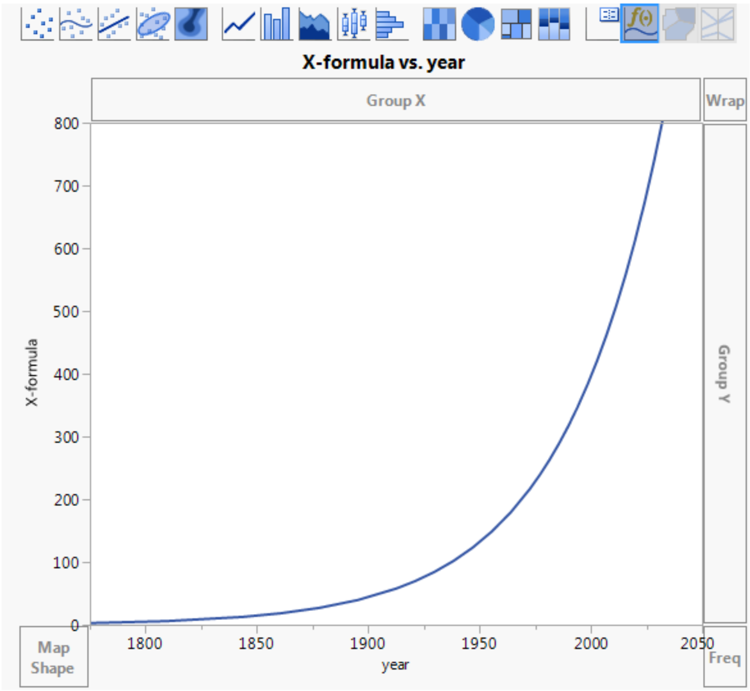
Figure 4.44 Points for Function Evaluated at Specified Years, with Smoother



Points are plotted for the years up to 2030. The points are the values of the function, X-formula, evaluated at the years in the year column. The function can be evaluated at the years for which pop is missing: 2010, 2020, and 2030. The smoother that is shown is only an approximation to the function. You want to see the function itself.

7. Click the Formula element .

Figure 4.45 Graph of Function

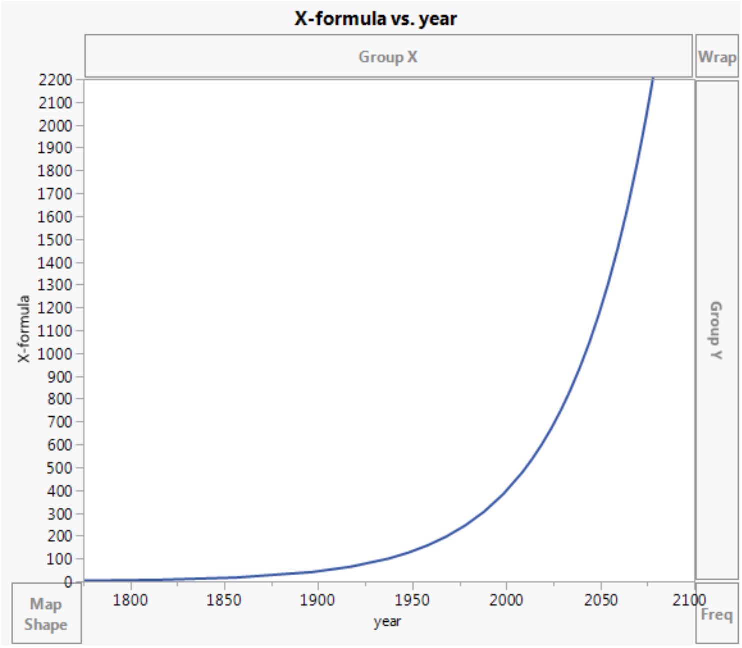


The Smoother and points are removed and a plot of the function itself appears. The function extends indefinitely to future years.

Next, rescale the axes to see the function plotted even further into the future.

8. Double-click the Y axis.
9. In the Y Axis Settings window, enter 2200 for the **Maximum** value, and then click **OK**.
10. Double-click the X axis.
11. In the X Axis Settings window, enter 2100 for the **Maximum** value, and then click **OK**.

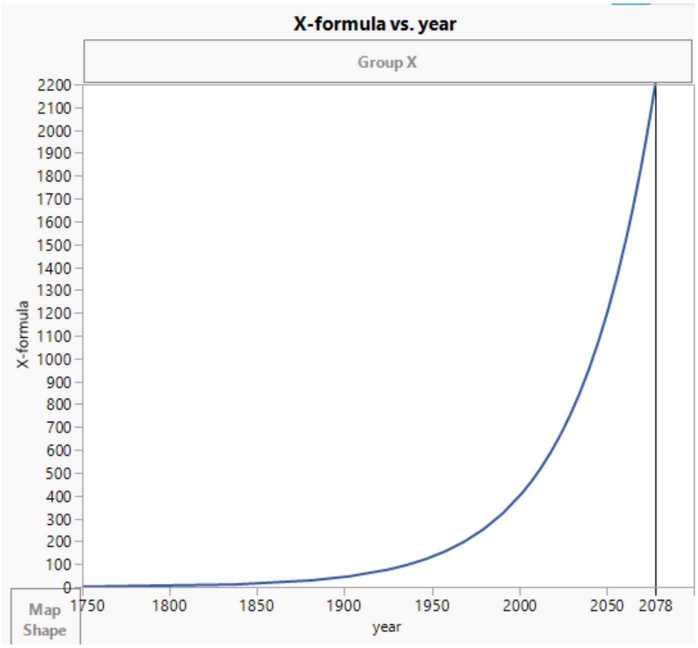
Figure 4.46 Graph of Function to Year 2100



To see that this function predicts the US population to be about 2,200 million people in the year 2078, you could use the crosshairs tool. But to see this more clearly, plot a reference line at year 2078.

12. Double-click the X axis.
13. In the Reference Lines panel, type 2078 next to **Value** and also next to **Label**.
14. Click **Add**.
15. Click **OK**.

Figure 4.47 Projected US Population at Year 2078

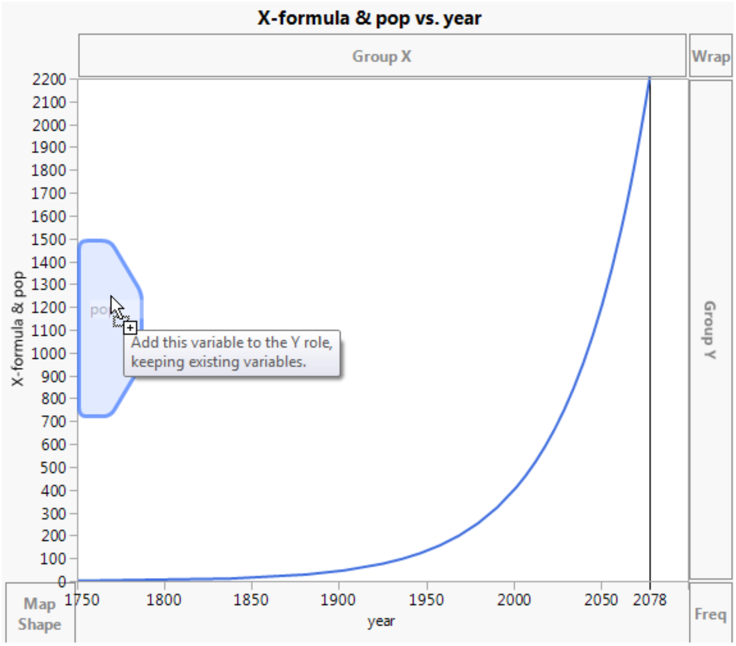


With the reference line, it is easy to see that at year 2078, the US population is projected to be about 2,200 million.

But you realize that making predictions based on this model for the year 2078 is dangerous extrapolation much beyond the range of the data. To see this clearly, add the population values for the years for which you have data to your plot.

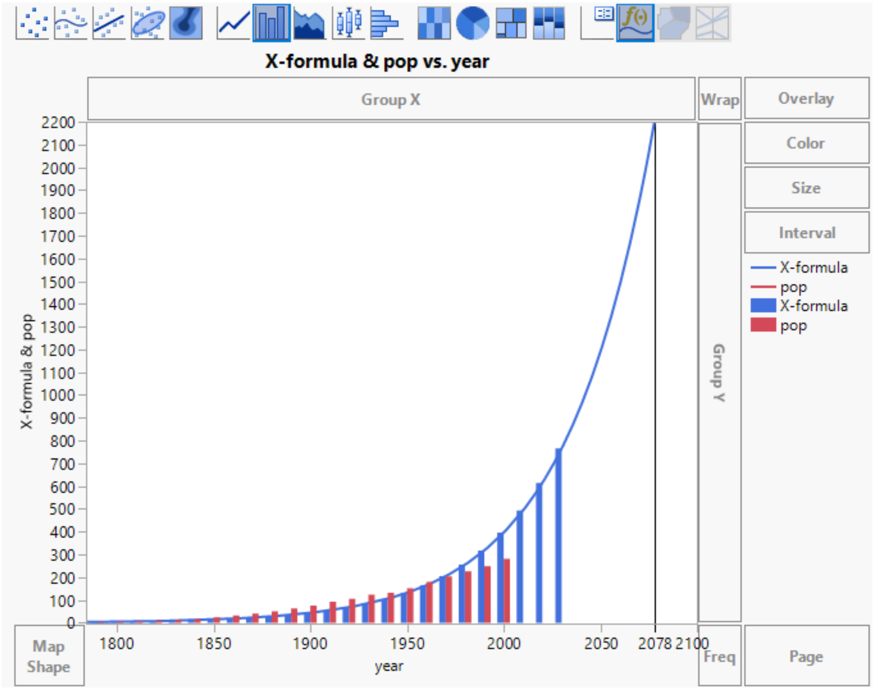
16. Drag pop just inside the Y axis to merge it with X-formula in the Y zone.

Figure 4.48 Merging pop with X-formula in the Y Zone



17. Drag the Bar element  into the plot.

Figure 4.49 Bars Showing Values for pop and X-formula



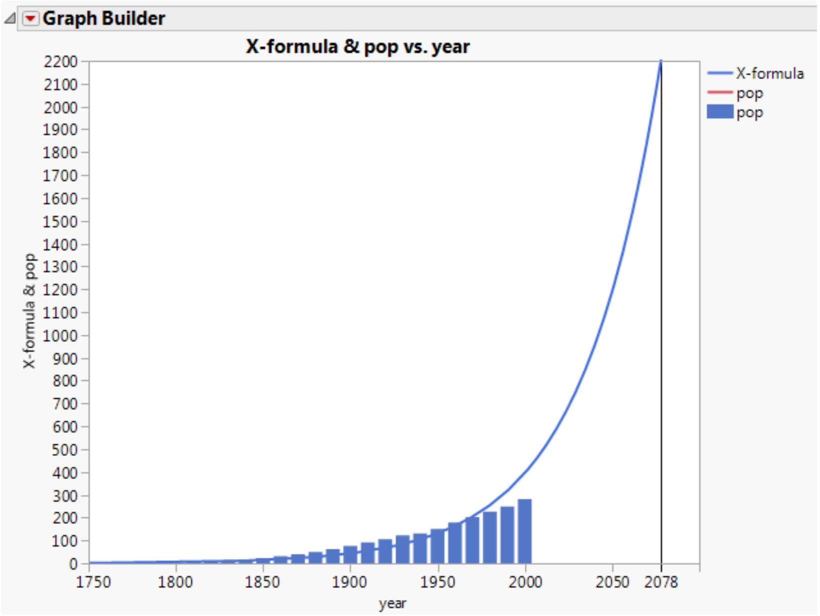
The height of each blue bar is the projected value of the US population for the given year. The height of each red bar is the actual US population. These are plotted for the years where you have data. Notice that the year 2078 is well beyond the range of years for which you have data.

18. In the Bar options, click **Variables** and de-select **Y X-formula**.

    The plot now shows only the function and bars that represent the actual population in the years for which you have data.

19. (Optional) Click **Done**.

Figure 4.50 Model Predictions to Year 2078 and Bars for Years with Data



You plotted the nonlinear function that was fit to the data from 1790 to 2000. You added bars showing the population for years for which you have data, which helps you see that using this function to predict values in the distant future is unwise.

Example Using Images as Markers

If you want to use images as markers or points in a graph, you can add an expression column containing the images to your data table, then use that column in Graph Builder. For an example, see the JMP Reports chapter in *Using JMP*.

Chapter 5

Bubble Plots

View Patterns in Multidimensional Data Using Bubble Plots

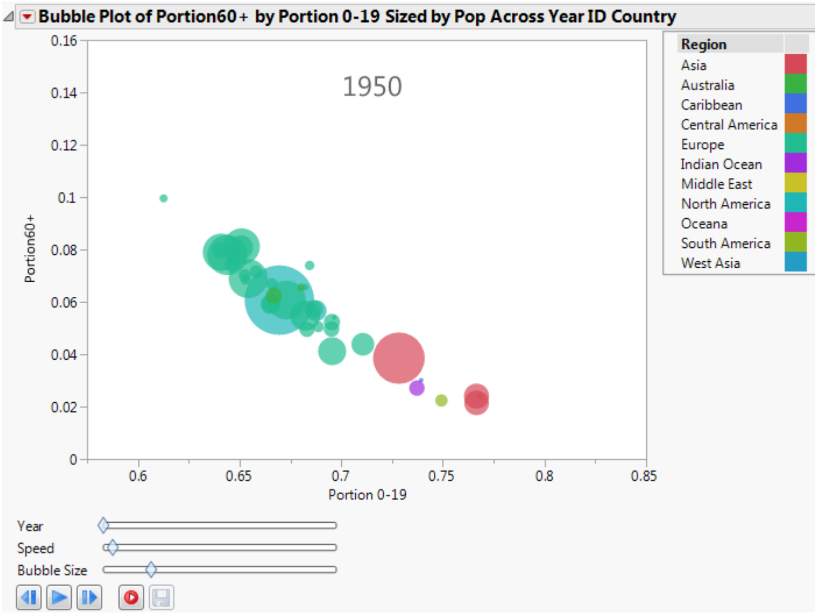
A bubble plot is a scatter plot that represents its points as circles, or bubbles. Bubble plots can be dynamic (animated over time) or static (fixed bubbles that do not move). Use bubble plots to:

- dynamically animate bubbles using a time variable, to see patterns and movement across time
- use size and color to clearly distinguish between different variables
- aggregate data (rows) into a single bubble, simplifying the bubble plot

Because you can see up to five dimensions at once (x position, y position, size, color, and time), bubble plots can produce dramatic visualizations and readily show patterns and trends.

Note: Dynamic bubble plots were pioneered by Hans Rosling, Professor of International Health, Karolinska Institutet, and the people involved in the Gapminder.org project.

Figure 5.1 Example of a Bubble Plot



Contents

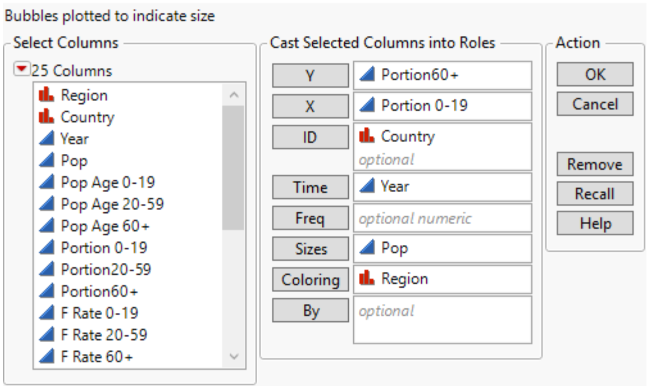
Example of a Dynamic Bubble Plot

This example uses the PopAgeGroup.jmp sample data table, which contains population data for countries and regions around the world. Examine the relationship between the proportion of younger and older people in the sample populations.

1. Select **Help > Sample Data Library** and open PopAgeGroup.jmp.
2. Select **Graph > Bubble Plot**.

   The launch window appears.

Figure 5.2 The Bubble Plot Launch Window



3. Select Portion60+ and click **Y**.

   The portion of the population that are 60 years or older becomes the *y* coordinate.
4. Select Portion 0-19 and click **X**.

   The portion of the population that are 0-19 years becomes the *x* coordinate.
5. Select Country and click **ID**.

   All the rows for each country are aggregated into a single bubble.
6. Select Year and click **Time**.

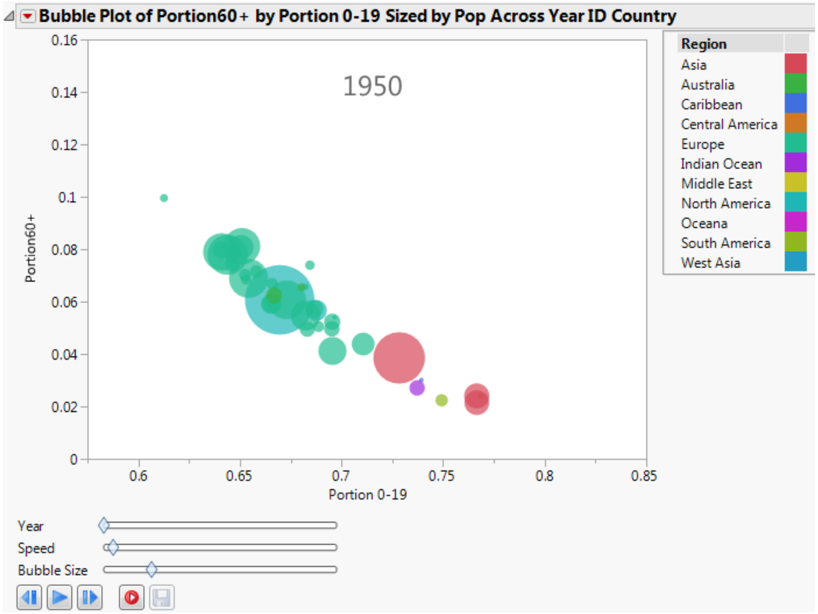
   The bubble plot shows a unique plot for each year’s data.
7. Select Pop and click **Sizes**.

   The sizes of the bubbles reflect the overall population values.
8. Select Region and click **Coloring**.

   Bubbles for different regions are assigned different colors. See the Region legend in Figure 5.3. The colors shown in the plot are JMP default colors.
9. Click **OK**.

The report window appears.

Figure 5.3 The Bubble Plot Report Window



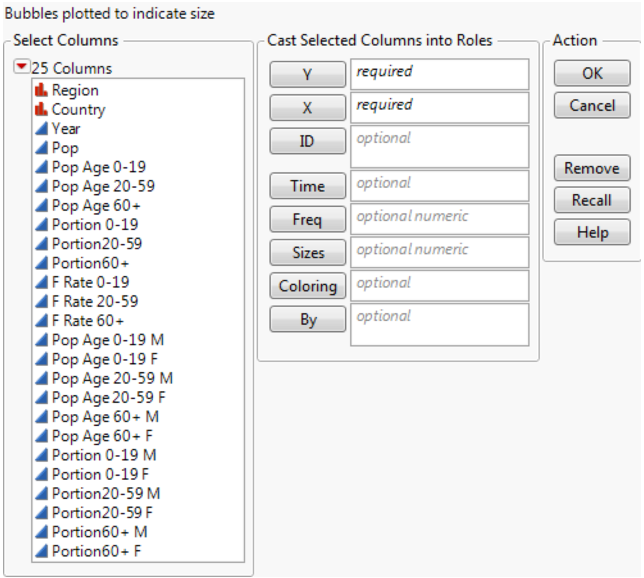
10. Click the play button to see the animated, dynamic report. Alternatively, you can click the previous button to move forward by one year.
11. (Optional) To view a legend that identifies each color with its region, click the Bubble Plot red triangle and select **Legend**.

As time progresses, you can see that the portion of the population that is 0-19 years decreases, and the portion of the population that is 60 years or more increases.

Launch the Bubble Plot Platform

Launch the Bubble Plot platform by selecting **Graph > Bubble Plot**.

Figure 5.4 The Bubble Plot Launch Window



Y, X The Y and X columns become the *y* and *x* coordinates of the bubbles in the plot. These values can be continuous or categorical (nominal or ordinal).

ID ID variables identify rows that should be aggregated and shown as a single bubble. The default coordinates of each bubble are the averaged *x* and *y* values, and the default size of each bubble is the sum of the sizes of all aggregated members. See “Specifying Two ID Variables” on page 168.

Time Maintains separate coordinates, sizes, and colors for each unique time period. The bubble plot shows these values for a single time period. For example, if the **Time** column contains years, the bubble plot is updated to show data by each year. See “Specifying a Time Variable” on page 168.

Freq Weights computations when aggregating bubbles using an ID variable.

Note: Negative frequency values are ignored.

Sizes Controls the size of the bubbles. The area of the bubbles is proportional to the **Size** value. There is a minimum bubble size, to keep bubbles visible, even if the size value is zero. If **Size** is left blank, the default bubble size is proportional to the number of rows in that combination of **Time** and **ID**.

Coloring Colors the bubbles according to the selected variable. If the selected variable is categorical (nominal or ordinal), each category is assigned a unique color. If the selected variable is continuous, a gradient of colors is used. You can set a preference for a color theme by selecting File > Preferences > Graphs and making a selection in the Color Themes panel.

By Place a column here to produce a separate bubble plot for each level of the variable.

After you click **OK**, the Bubble Plot report window appears.

Specifying Two ID Variables

Specifying a second **ID** variable provides a hierarchy of categories, but the bubbles are not split by the second category until they are selected and split interactively. In the launch window, if you specify a second **ID** variable, **Split** and **Combine** buttons appear in the report window.

For example, you might specify a country as the first **ID** variable, resulting in a separate aggregated bubble for each country. A second **ID** variable, perhaps designating regions within each country, would further split each country when the interactive **Split** button under the graph is pressed.

Specifying a Time Variable

Maintains separate coordinates, sizes, and colors for each unique time period. The bubble plot shows these values for a single time period. For example, if the **Time** column contains years, the bubble plot is updated to show data by each year.

To move the time label on the plot, click and drag the label.

If data is missing within a time period, the value is linearly interpolated. If data is missing for the first or last time period, the value is not estimated, but left as missing.

Related Information

- "Control Animation for Dynamic Bubble Plots" on page 169
- "Example of Specifying Only a Time Variable" on page 175

Interact with the Bubble Plot

Note: If *all* of the rows used in constructing a bubble plot shape are hidden and not excluded, the corresponding bubble plot shape is not shown in the bubble plot. If rows are excluded in the data table, the bubble plot is constructed without the excluded rows.

Use the Bubble Plot platform in one of two modes:

- *Static mode*, where the bubbles are fixed and do not animate over time (no **Time** variable is specified). See “Example of a Static Bubble Plot” on page 179.
- *Dynamic mode*, where the bubbles are animated over time (a **Time** variable is specified). See “Example of a Dynamic Bubble Plot” on page 165.

You interact with both static and dynamic bubble plots in different ways.

Control Animation for Dynamic Bubble Plots

Use sliders and buttons to control the animation of dynamic bubble plots.

Figure 5.5 Animation Controls

Table 5.1 Descriptions of the Animation Controls

Slider or Button	Description
<Time variable>	Controls which time values appear in the bubble plot. You manually drag the slider to see a progression of time. Click and drag on the time variable in the bubble plot to move its position. Appears only if you specified a variable for **Time**.
Speed	Adjusts the speed of the animation. Appears only if you specified a variable for **Time**.
Bubble Size	Adjusts the size of the bubbles. The bubbles maintain their relative size, but their absolute size can be adjusted. Appears on all bubble plots.

Table 5.1 Descriptions of the Animation Controls *(Continued)*

Slider or Button	Description
	Adjusts the time value by one unit and shows the previous time value. Appears only if you specified a variable for **Time**.
 	Press play to animate the bubble plot. Moves through all of the time values in order, and loops back to the beginning when the last time period is reached. Press pause to stop the animation. Appears only if you specified a variable for **Time**.
	Adjusts the time value by one unit and shows the next time value. Appears only if you specified a variable for **Time**.
	Records the animation (Windows only).
	Saves the animation as an animated GIF file (Windows only).
Split	Separates the bubble represented by the first, larger **ID** variable into its smaller constituent parts, which are defined by the second, smaller **ID** column. Select the bubble and click **Split**. Appears only if you specified two **ID** variables.
Combine	Reverses the action of the **Split** button by recombining the smaller bubbles back into their original bubble. Select any of the smaller bubbles in the group and click **Combine**. Appears only if you specified two **ID** variables.

Select Bubbles

Click a bubble to select it. Note the following:

- Visually, selected bubbles become darker or brighter, and non-selected bubbles are more transparent.
- If the bubble was not filled initially, selection fills it.
- If no bubbles are selected, all of the bubbles are semi-transparent.

When you select a bubble, all of the rows in the data table that correspond to the selected bubble are highlighted. Note the following:

- If the bubble is an aggregate based on an **ID** column, all of the rows for that **ID** are highlighted. Otherwise, the one row represented by that bubble is highlighted.
- If you specify an **ID** and a **Time** variable, selecting a bubble highlights all of the rows for that **ID**, across all of the **Time** levels.

If you select a row from the data table, it is selected in the associated bubble plot. Note the following:

- If you have not specified a **Time** variable, selecting one row from the data table highlights the corresponding bubble in the plot.
- If you have specified a **Time** variable, selecting one row from the data table highlights the corresponding bubble for only that time period in the dynamic bubble plot.

Use the Brush Tool

Use the brush tool  to temporarily select bubbles and obtain more information about the selected bubbles. When you select bubbles with the brush tool, the corresponding rows are highlighted in the associated data table.

Note: For a more granular examination of the highlighted rows, use the **Tables > Subset** command or the Row Editor. See the Reshape Data chapter in *Using JMP*.

Bubble Plot Platform Options

The Bubble Plot red triangle menu provides the following options:

Draw Applies a fill or outline.

Filled Fills all of the bubbles.

Outlined Outlines all of the bubbles.

Filled and Outlined Fills and outlines all of the bubbles.

Set Shape Change the shape of the bubble.

You can create a custom shape using JSL. The **Custom** option opens the custom shape. If no custom shape has been created, the **Custom** option uses the default circle shape. For more information about creating custom shapes, see the Scripting Graphs chapter in the *Scripting Guide*.

Orient Shapes Orients the shapes as they move in particular directions over time, following the shape of the data.

This option appears only if you have specified a variable for **Time**.

Trail Bubbles Shows the past history of bubbles as a semi-transparent trail. See "Example of Specifying Only a Time Variable" on page 175.

Note the following:

– This option appears only if you have specified a variable for **Time**.

– If you do not want to see the bubble labels, select the **Label > None** option.

Trail Lines Shows the past history of bubbles as connected line segments. See "Example of Specifying Only a Time Variable" on page 175.

Note the following:

– This option appears only if you have specified a variable for **Time**.

– If you do not want to see the bubble labels, select the **Label > None** option.

Label Changes the labels on the plot.

None Labels none of the bubbles in the plot.

All Labels all of the bubbles in the plot.

Selected Labels bubbles only when you select them.

Note: Click and drag on a label to move it.

Color Theme Change the colors representing the high, middle, and low values of the color variable.

This option appears only if you have specified a variable for **Coloring**.

Revert Color Theme Reverts back to the original color theme.

This option appears only if you have applied a color theme.

Legend Shows a legend that describes the colors in the bubble plot.

This option appears only if you have specified a variable for **Coloring**.

Selectable Across Gaps If a bubble is selected, this option keeps the bubble selected during time periods where data is missing. Otherwise, the bubble is not selected during time periods where data is missing.

Show Roles Shows the variables that are used in the bubble plot. You can change and delete the variables. See "Show Roles" on page 174.

Split All Splits all bubbles into their constituent parts. Unlike the **Split** button, the bubbles do not have to be selected.

This option appears only if you have specified two **ID** variables.

Combine All Combines all constituent bubbles within a group into their larger bubble. Unlike the **Combine** button, the bubbles do not have to be selected.

This option appears only if you have specified two **ID** variables.

Show Time Annotation In bubble plots that contain a time variable, the Show Time Annotation option shows or hides the time that is displayed in the upper left corner.

Lock Scales Prevents axis scales and gradient legend scales from automatically adjusting in response to data or filtering changes.

Fit to Window Determines whether the plot is resized as you resize the JMP window. The default setting is Auto, which bases the scaling on the contents of the plot. For example, a plot with By variables does not stretch to fit the resized window; the plot extends beyond the viewing area. Change the setting to On to always fit the plot inside the window. Change the setting to Off to prevent the plot from resizing.

Aggregation Options Alters how the **X**, **Y**, and **Sizes** roles are computed. By default, the values are calculated using means for **X** and **Y**, and sums for **Sizes**.

X as Sum or Y as Sum Computes the **X** and **Y** values using sums.

Size as Sum Deselecting this option computes **Size** values using means.

Color as Sum Computes the sum of the data values and maps to a color. This option appears only for continuous variables.

See the JMP Reports chapter in *Using JMP* for more information about the following options:

Local Data Filter Shows or hides the local data filter that enables you to filter the data used in a specific report.

Redo Contains options that enable you to repeat or relaunch the analysis. In platforms that support the feature, the Automatic Recalc option immediately reflects the changes that you make to the data table in the corresponding report window.

Save Script Contains options that enable you to save a script that reproduces the report to several destinations.

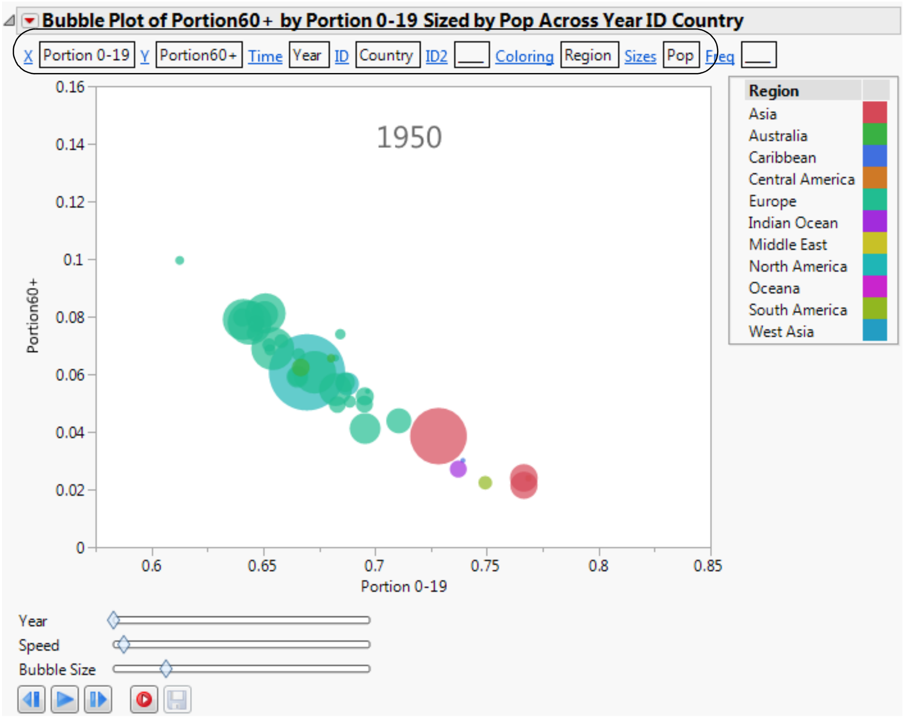
Save By-Group Script Contains options that enable you to save a script that reproduces the platform report for all levels of a By variable to several destinations. Available only when a By variable is specified in the launch window.

Show Roles

Using the **Show Roles** option in the red triangle menu, you can make changes to your existing variables without having to relaunch the platform and start your analysis over.

Follow the instructions in "Example of a Dynamic Bubble Plot" on page 165 to produce the report window shown in Figure 5.6. (The colors are the JMP default colors.)

Figure 5.6 Example of Bubble Plot with Show Roles Selected



Change the Variable Assigned to a Role

To change the variable assigned to a role, click a blue underlined role name. For example, in Figure 5.6, to change the **Coloring** variable from Region to Country, proceed as follows:

1. Click the **Coloring** link.

   The Select column for Coloring window appears.

2. Click Country.
3. Click **OK**.

   Country now replaces Region as the **Coloring** variable in the bubble plot.

Remove a Variable

To remove an existing variable from the bubble plot, make sure that nothing is selected in the Select column for <*Role*> window, and click **OK**. For example, in Figure 5.6, to remove the **Sizes** variable (Pop), proceed as follows:

1. Click the Sizes link.

   The Select column for Sizes window appears.

2. Ensure that nothing is selected. If a variable is selected, deselect it by holding down the Ctrl key and clicking on the variable.

3. Click **OK**.

   The Sizes role now appears with an empty box.

Note: The X and Y variables can be changed only and cannot be removed.

Add a Variable

Once you have removed an existing variable from the bubble plot, there are two ways to add a new variable:

- Click the blue underlined role name. See "Change the Variable Assigned to a Role" on page 174.
- In the data table, click the variable in the column panel, and drag it into the empty role box.

Additional Examples of the Bubble Plot Platform

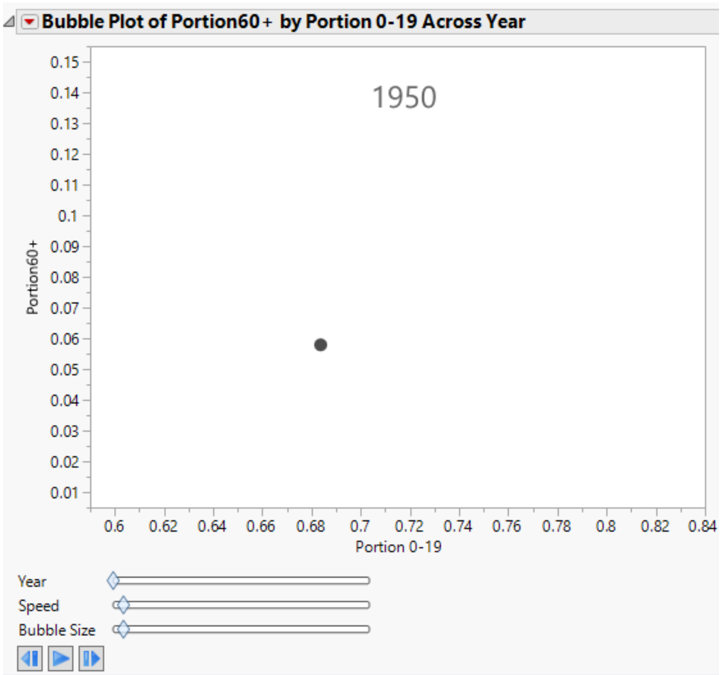
Example of Specifying Only a Time Variable

For dynamic bubble plots, you might specify only a **Time** variable and no **ID** variable. The resulting bubble plot contains a single moving bubble that tracks the series as the **Time** value changes.

1. Select **Help > Sample Data Library** and open PopAgeGroup.jmp.

2. Select **Graph > Bubble Plot**.
3. Select Portion60+ and click **Y**.
4. Select Portion 0-19 and click **X**.
5. Select Year and click **Time**.
6. Click **OK**.

Figure 5.7 The Initial Report Window with a Time Variable

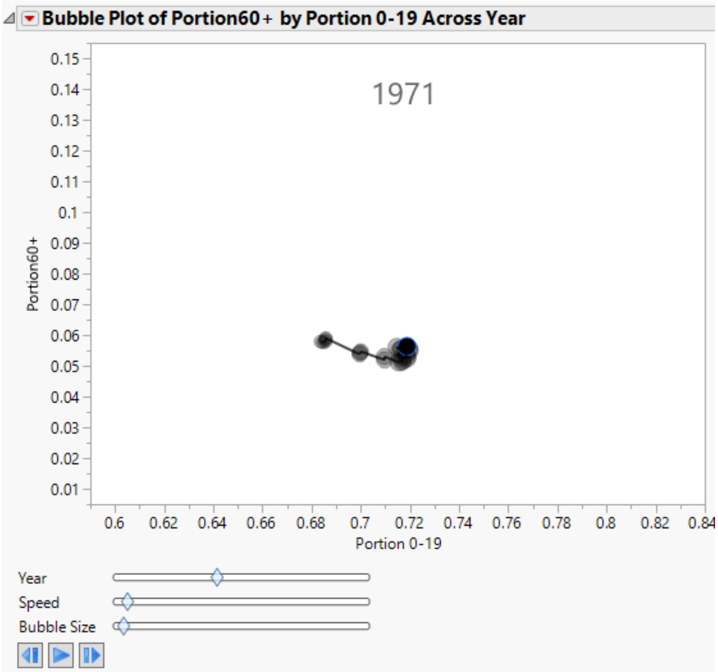


7. Click the bubble to select it.

   All rows in the data table are also highlighted.
8. Click the Bubble Plot red triangle and select **Trail Bubbles > All** and **Trail Lines > All**.
9. Click the play button.

   The bubble plot animates, showing a trail for the single bubble.

Figure 5.8 Animated Bubble Plot over Time



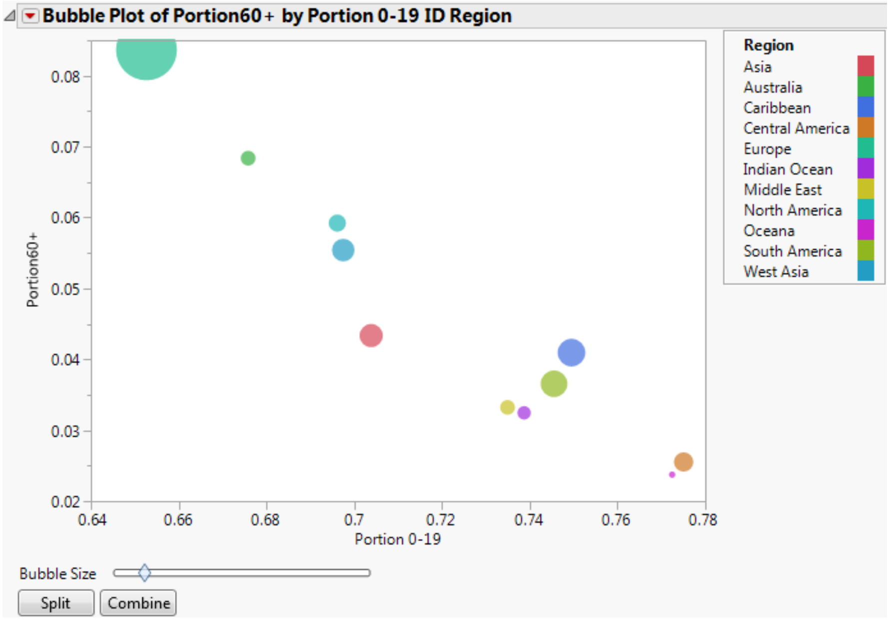
Example of Specifying Only ID Variables and Splitting a Bubble

For static bubble plots, you might specify one or two **ID** variables and no **Time** variable. The resulting bubble plot contains a bubble at each **ID** value. Note that although this bubble plot is static, you can perform splitting on bubbles.

1. Select **Help > Sample Data Library** and open PopAgeGroup.jmp.
2. Select **Graph > Bubble Plot**.
3. Select Portion60+ and click **Y**.
4. Select Portion 0-19 and click **X**.
5. Select Region and Country and click **ID**.
6. Select Region and click **Coloring**.
7. Click **OK**.

   The initial report window appears. (Figure 5.9 uses the default JMP colors.)

Figure 5.9 Initial Report Window with ID Variables

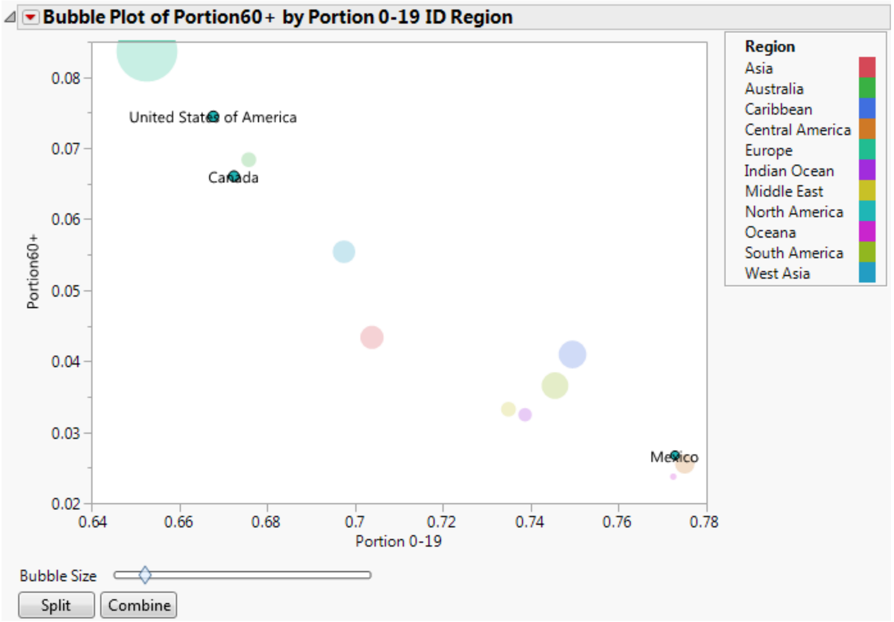


Split the bubble representing the region of North America into countries.

8. Click the bubble representing North America (hover over a bubble to see its label, or use the legend to find the color of North America.)
9. Click **Split**.

   You see that the North America bubble has split into three bubbles, representing the countries within the region of North America (the United States of America, Canada, and Mexico).

Figure 5.10 Splitting the North America Bubble



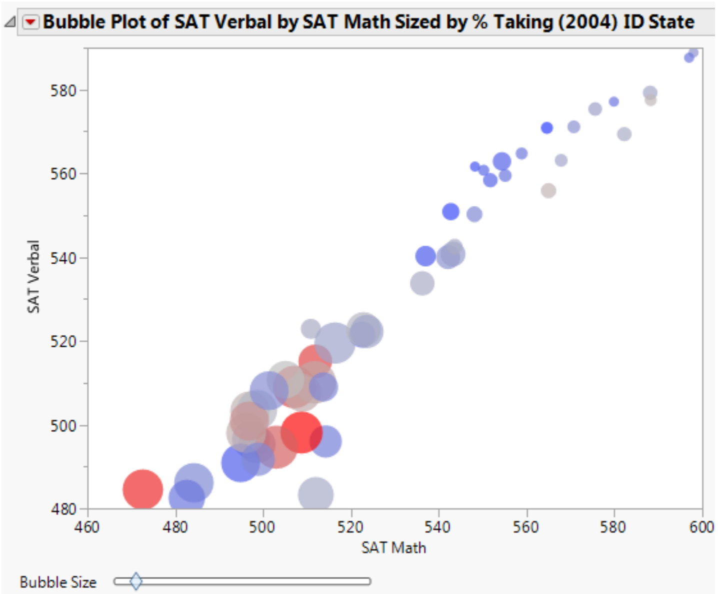
Example of a Static Bubble Plot

This example uses the SATByYear.jmp sample data table, which contains SAT verbal and math test scores for a selection of the US population in 2004.

1. Select **Help > Sample Data Library** and open SATByYear.jmp.
2. Select **Graph > Bubble Plot**.
3. Select SAT Verbal and click **Y**.
4. Select SAT Math and click **X**.
5. Select State and click **ID**.
6. Select % Taking (2004) and click **Sizes**.
7. Click **OK**.

   The report window appears. (Figure 5.11 uses the default JMP colors.)

Figure 5.11 The Static Bubble Plot Report Window



You draw the following conclusions:

- Higher verbal scores appear to be associated with higher math scores, since the two track very closely in the bubble plot. This signifies a correlation between verbal and math scores.
- The larger bubbles represent the US states that have a high percentage of individuals taking the SAT test in 2004. These larger bubbles are all grouped together in the lower left of the graph. This shows that when a state has a high percentage of individuals taking the test, both the math and verbal scores are low.

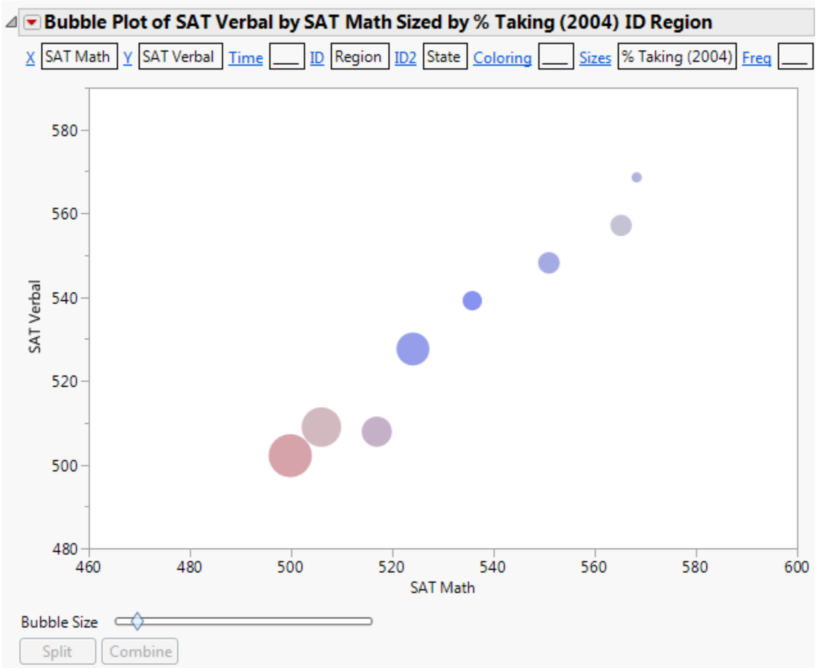
Instead of grouping the bubbles primarily by state, group the bubbles primarily by region as follows:

1. Click the Bubble Plot red triangle and select **Show Roles**.
2. Click the **ID** link.
3. Select Region, and click **OK**.

   Region is now the primary **ID** variable.
4. Click the **ID2** link.
5. Select State, and click **OK**.

   State is now the secondary **ID** variable.

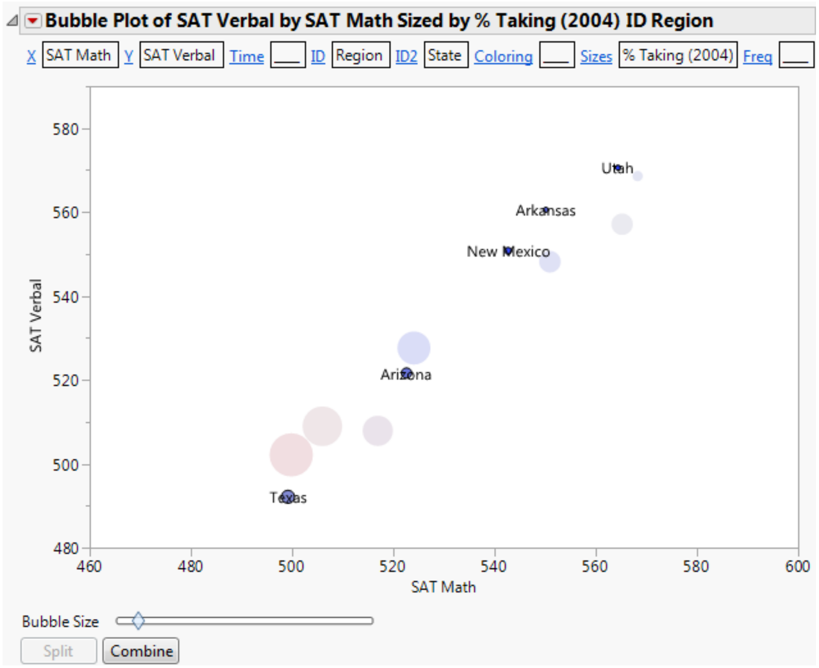
Figure 5.12 Example of Bubble Plot Grouped by Region and State



6. Click the bubble that represents the Southwest region (hover over a bubble or click it to see its label).
7. Click **Split**.

   Now the bubbles are split by the secondary **ID** variable, which is **State**. You now see each state within the Southwest region.

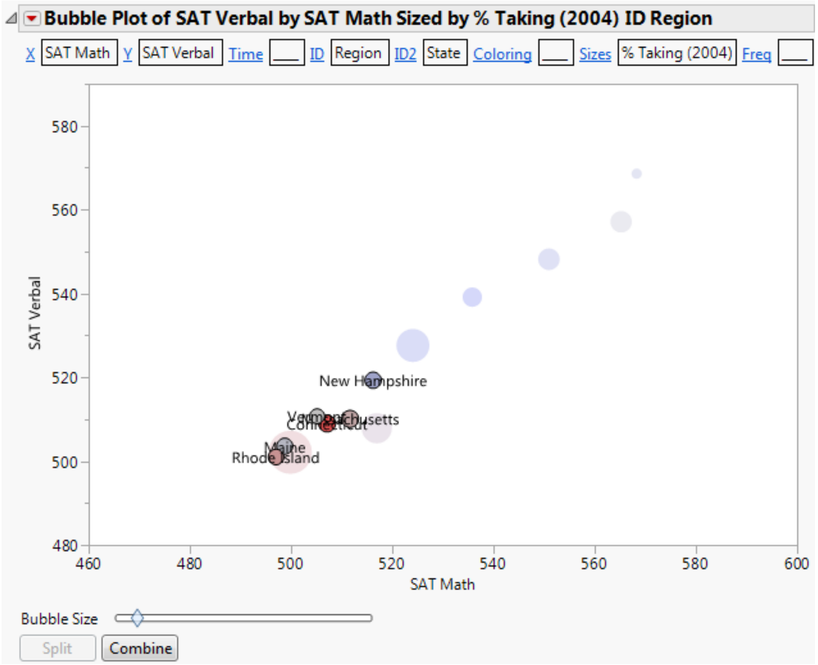
Figure 5.13 Example of Southwest Region Split by State



You see that there is significant variation between the scores from the Southwest states.

8. Click **Combine** to combine the southwest states again.
9. To do a comparison, click the New England bubble (hover over a bubble or click it to see its label).
10. Click **Split**.

Figure 5.14 Example of New England Region Split by State



You see that the New England states do not have as much variation as the Southwest states.

Example of a Bubble Plot with a Categorical Y Variable

All of the examples shown so far use continuous Y variables. If you use a categorical (nominal or ordinal) Y variable, the bubble plot appears differently.

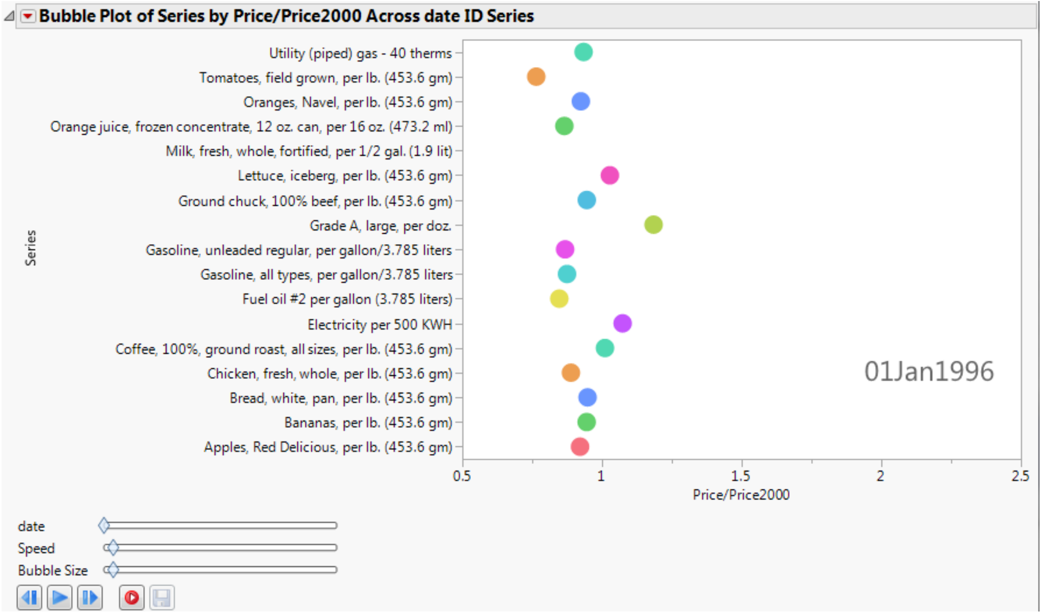
This example uses the blsPriceData.jmp sample data table, which shows the price of commodities over several years. Because the value of the US dollar changes over time, a column named Price/Price2000 shows the ratio of a commodity's price at any given time to the price in the year 2000.

1. Select **Help > Sample Data Library** and open blsPriceData.jmp.
2. Select **Graph > Bubble Plot**.
3. Select Series and click **Y**.
4. Select Price/Price2000 and click **X**.
5. Select date and click **Time**.
6. Click **OK**.

   The report window appears. (Figure 5.15 uses the JMP default colors.)

This produces a bubble plot that, when animated by clicking the play button, shows the price bubbles moving side to side according to their price ratio.

Figure 5.15 Static Example of Animated Bubbles



For easier readability, add grid lines as follows:

7. Double-click the categorical axis.
8. In the Y Axis Settings window, select **Show Grid**.
9. Click **OK**.

To animate the bubble plot, click the play button. The price bubbles move side to side, according to their price ratio.

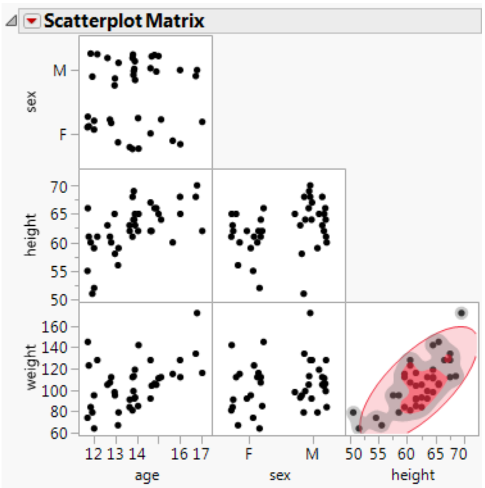
Chapter 6

Scatterplot Matrix

View Multiple Bivariate Relationships Simultaneously

Using the Scatterplot Matrix platform, you can assess the relationships between multiple variables simultaneously. A scatterplot matrix is an ordered collection of bivariate graphs. For further analysis, you can customize the scatterplots with density ellipses for all of your data, or for only groups of your data.

Figure 6.1 Example of a Scatterplot Matrix



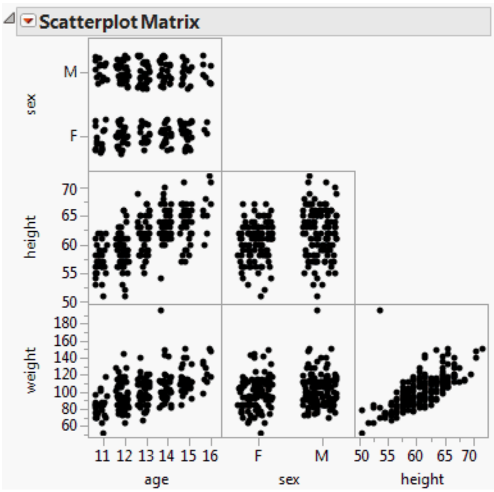
Contents

Example of a Scatterplot Matrix

This example shows you how to create a scatterplot matrix.

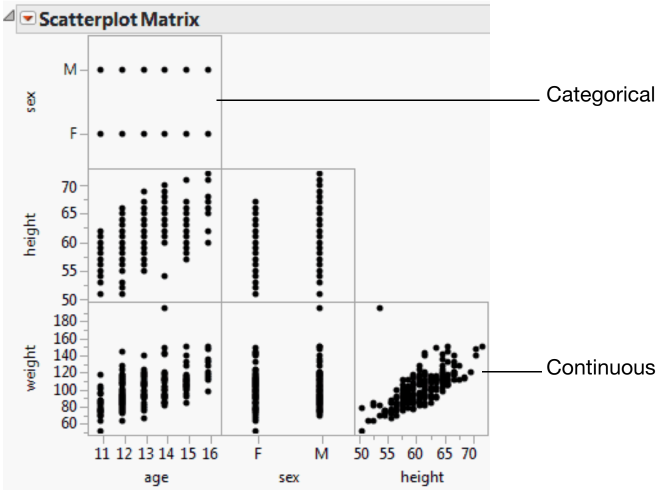
1. Select **Help > Sample Data Library** and open Students.jmp.
2. Select **Graph > Scatterplot Matrix**.
3. Select age, sex, height, and weight and click **Y, Columns**.
4. Click **OK**.

Figure 6.2 Example of a Scatterplot Matrix



In this example, you can see that the graph for weight versus height is different from the graph for sex versus age. If you turn off jitter by clicking on the red triangle menu and selecting **Points Jittered**, the difference becomes even more pronounced.

Figure 6.3 Example of a Scatterplot Matrix with No Jitter

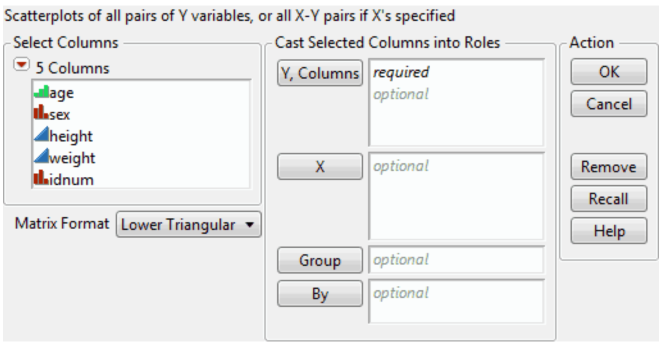


The weight versus height graph shows continuous data, and the sex versus age graph shows categorical data.

Launch the Scatterplot Matrix Platform

Launch the Scatterplot Matrix platform by selecting **Graph > Scatterplot Matrix**.

Figure 6.4 The Scatterplot Matrix Launch Window



Y, Columns, X Specify columns for *Y* and *X*.

– If you assign variables to the **Y, Columns** role only, they appear on both the horizontal and vertical axes.

– If you assign variables to both the **Y, Columns** and **X** role, then the **Y, Columns** variables appear on the vertical axis. The **X** variables appear on the horizontal axis. This approach enables you to produce rectangular matrices, or matrices that have different, yet overlapping, sets of variables forming the axes of the matrix.

Group If you assign a variable to the **Group** role, you can add shaded density ellipses for each level of the **Group** variable. See "Example Using a Grouping Variable" on page 193.

By This option produces a separate scatterplot matrix for each level of the **By** variable. If two **By** variables are assigned, a separate graph for each possible combination of the levels of both **By** variables is produced.

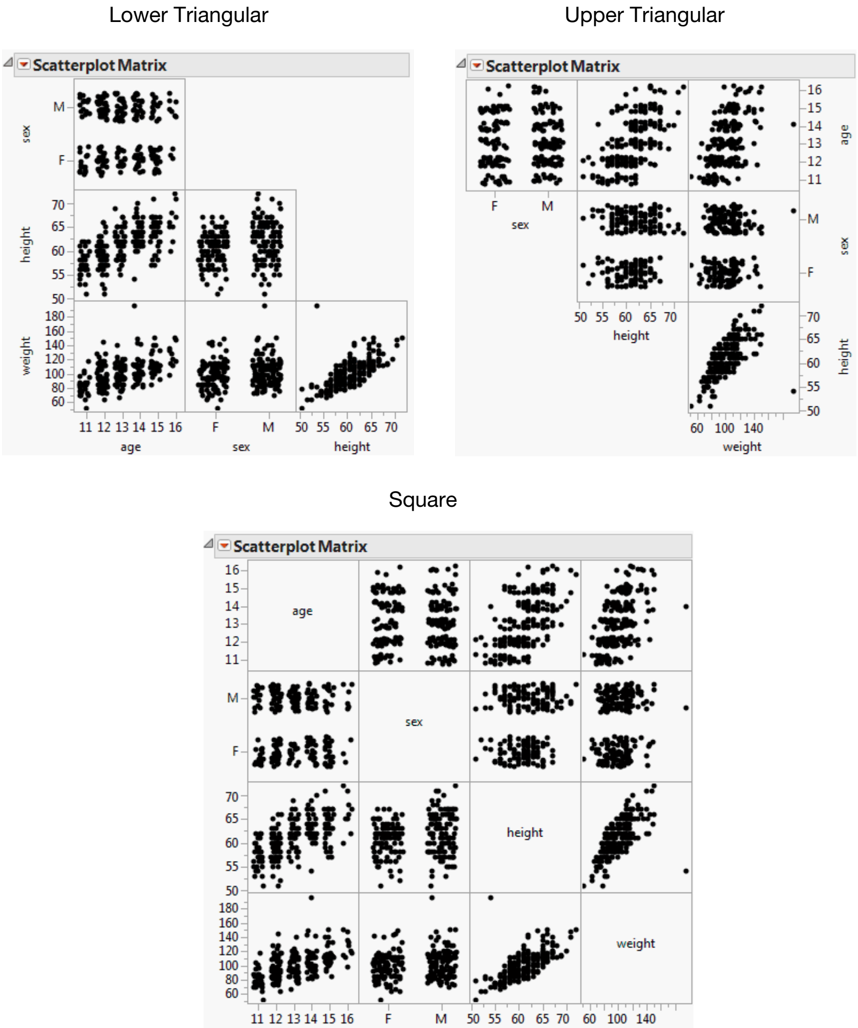
Matrix Format The **Matrix Format** can be one of three arrangements: **Upper Triangular**, **Lower Triangular**, or **Square**. See "Change the Matrix Format" on page 190.

After you click **OK**, the Scatterplot Matrix window appears. See "The Scatterplot Matrix Window" on page 191.

Change the Matrix Format

The **Matrix Format** can be one of three arrangements: **Upper Triangular**, **Lower Triangular**, or **Square**.

Figure 6.5 Examples of Matrix Formats



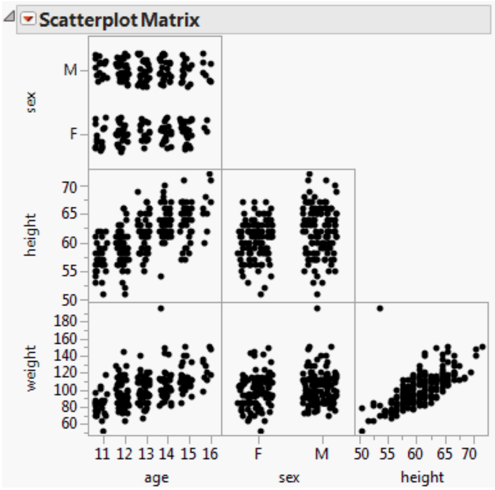
The Scatterplot Matrix Window

The Scatterplot Matrix window shows an ordered grouping of bivariate graphs. In each graph, you can examine the relationships between each pair of variables.

Follow the instructions in "Example of a Scatterplot Matrix" on page 187 to produce the plot shown in Figure 6.6.

Note: For information about additional options, see "Scatterplot Matrix Platform Options" on page 191.

Figure 6.6 Example of a Scatterplot Matrix Window



Replace variables in the plot by dragging and dropping a variable, in one of two ways: swap existing variables by dragging and dropping a variable from one axis to the other axis; or, click a variable in the Columns panel of the associated data table and drag it onto an axis. This feature is not available for matrices in the Square format.

Scatterplot Matrix Platform Options

The Scatterplot Matrix red triangle menu provides the following options:

Show Points Shows or hides the points in the scatterplots.

Points Jittered Turns the jittering of the points in the scatterplot on or off. This option is available when at least one variable is either ordinal or nominal.

Fit Line Fits a simple regression line and its mean confidence interval to the scatterplots.

Density Ellipses Shows or hides the outline and area of the density ellipses. See "Example Using a Grouping Variable" on page 193.

Shaded Ellipses Colors the area within each ellipse. See "Example Using a Grouping Variable" on page 193.

Ellipses Coverage Enables you to select an α-level for the ellipses to cover.

Ellipses Transparency Enables you to select the transparency of the shaded ellipses, where 0 is completely transparent and 1 is completely opaque.

Ellipses Color Enables you to select a color for the outline and the area within an ellipse.

Nonpar Density Shows or hides the nonparametric density, which represents the areas where the data points are the most dense. The nonparametric density estimation is helpful when you have a lot of points and the density of the points is difficult to see.

There are two quantile density contours. One contour includes 50% of the smoothed density, and the other contour includes 90% of the smoothed density. Since the percentage is based on the smoothed density, the percentage may not agree with the actual proportion of points within a contour.

Group By In the Group By window, you can perform the following actions:

– If you did not select a **Group** variable in the launch window, you can add one now.

– If you did select a **Group** variable in the launch window, you can remove the existing **Group** variable, or you can replace the **Group** variable.

See "Example Using a Grouping Variable" on page 193.

Lock Scales Prevents axis scales and gradient legend scales from automatically adjusting in response to data or filtering changes.

See the JMP Reports chapter in *Using JMP* for more information about the following options:

Local Data Filter Shows or hides the local data filter that enables you to filter the data used in a specific report.

Redo Contains options that enable you to repeat or relaunch the analysis. In platforms that support the feature, the Automatic Recalc option immediately reflects the changes that you make to the data table in the corresponding report window.

Save Script Contains options that enable you to save a script that reproduces the report to several destinations.

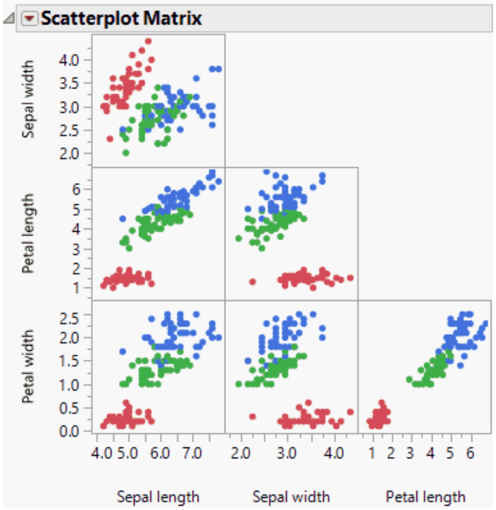
Save By-Group Script Contains options that enable you to save a script that reproduces the platform report for all levels of a By variable to several destinations. Available only when a By variable is specified in the launch window.

Example Using a Grouping Variable

This example shows you how to create a scatterplot matrix using a grouping variable.

1. Select **Help > Sample Data Library** and open Iris.jmp.
2. Select **Graph > Scatterplot Matrix**.
3. Select Sepal length, Sepal width, Petal length, and Petal width and click **Y, Columns**.
4. Select Species and click **Group**.
5. Click **OK**.

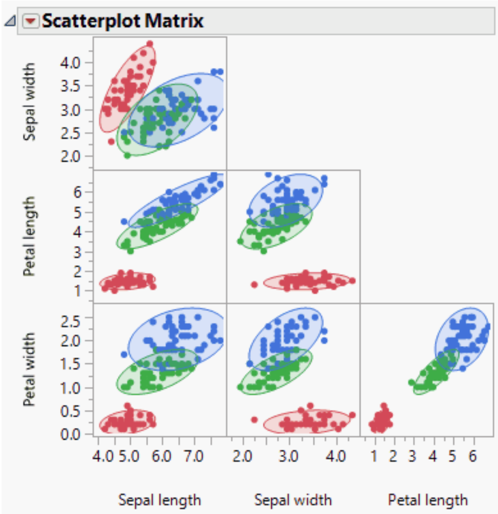
Figure 6.7 Initial Example Using a Grouping Variable



To make the groupings stand out, proceed as follows:

6. Click the Scatterplot Matrix red triangle and select **Density Ellipses > Density Ellipses**.
7. Click the Scatterplot Matrix red triangle and select **Density Ellipses > Shaded Ellipses**.

Figure 6.8 Example of a Scatterplot Matrix with Ellipses



Create a Grouping Variable

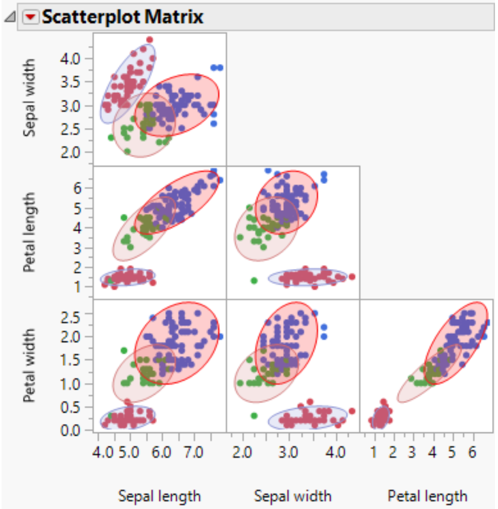
If your data does not already have a grouping variable, you can create one using the Cluster platform. Using the Iris.jmp data, assume that the Species column does not exist. You know that the data comes from three species of Iris flowers, so you want to create three clusters within a group.

Proceed as follows:

1. Using the Iris.jmp sample data table, select **Analyze > Clustering > Hierarchical Cluster**.
2. Select Sepal length, Sepal width, Petal length, and Petal width and click **Y, Columns**.
3. Click **OK**.
4. Click the Hierarchical Clustering red triangle and select **Number of Clusters**.
5. Type 3 to represent the three different Iris species.
6. Click **OK**.
7. Click the Hierarchical Clustering red triangle and select **Save Clusters**.
8. Close the Hierarchical Cluster report window, and go back to the Iris.jmp data table.

   You can see that a Cluster column has been added to the Iris.jmp data table.
9. Perform the Scatterplot Matrix analysis. Follow the instructions in the section "Example Using a Grouping Variable" on page 193, but use Cluster as the grouping variable.

Figure 6.9 Example of a Scatterplot Matrix Using a Cluster Variable



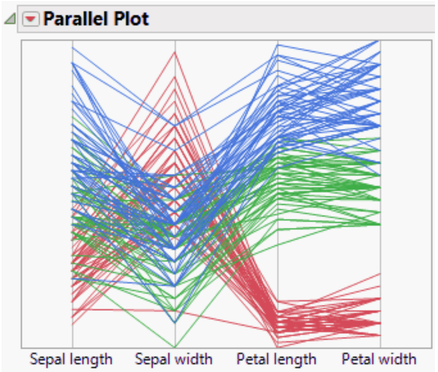
Chapter 7

Parallel Plots

View Patterns in Multidimensional Data by Plotting Parallel Coordinates

Using parallel plots, you can visualize each cell in a data table. Parallel plots draw connected line segments that represent each row in a data table. Parallel plots were initially developed by Inselberg (1985) and later popularized by Wegman (1990).

Figure 7.1 Example of a Parallel Plot



Contents

Example of a Parallel Plot

This example uses the Dogs.jmp sample data table, which contains histamine level measurements for 16 dogs that were given two different drugs. The histamine levels were taken at zero, one, three, and five minutes. Examine the variation in the histamine levels for each drug.

1. Select **Help > Sample Data Library** and open Dogs.jmp.

   To see the differences by drug, color the parallel plot lines by drug:

2. Select **Rows > Color or Mark by Column**.
3. Select drug.

   If the selection shown in your Colors menu is **JMP Default**, morphine is assigned the color red and trimeth is assigned the color blue.

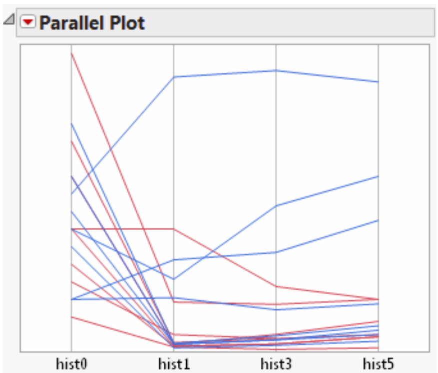
4. Click **OK**.

Create the parallel plot:

5. Select **Graph > Parallel Plot**.
6. Select hist0, hist1, hist3, and hist5 and click **Y, Response**.
7. Click **OK**.

   The report window appears.

Figure 7.2 Parallel Plot of Histamine Variables



Each connected line segment represents a single observation. Click a line segment to see which observation (or row) it corresponds to in the data table.

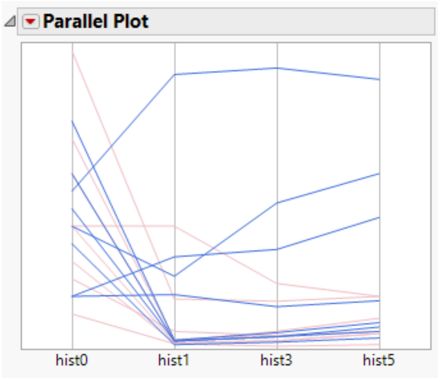
For further exploration, isolate the trimeth values:

8. Select **Rows > Data Filter**.

9. Select drug and click **Add**.
10. Select trimeth.

    Only the trimeth values are highlighted in the parallel plot.

Figure 7.3 Trimeth Values Highlighted



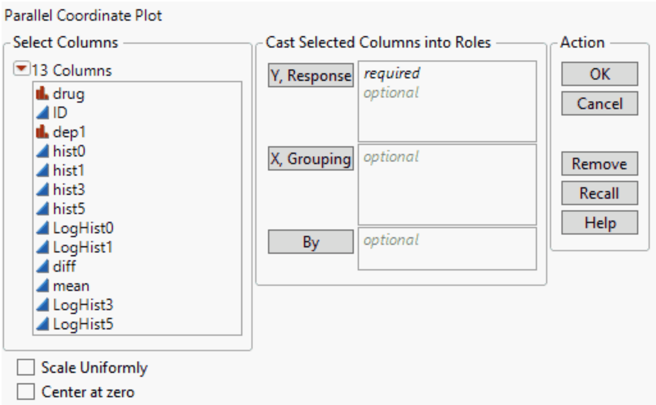
You observe the following about the histamine levels for dogs given trimeth:

- For most of the dogs, the histamine levels had a sharp drop at one minute.
- For four of the dogs, the histamine levels remained high, or rose higher. You might investigate this finding further, to determine why the histamine levels were different for these dogs.

Launch the Parallel Plot Platform

Launch the Parallel Plot platform by selecting **Graph > Parallel Plot**.

Figure 7.4 The Parallel Plot Launch Window



Y, Response Variables appear on the horizontal axis of the parallel plot. These values are plotted and connected in the parallel plot.

X, Grouping Produces a separate parallel plot for each level of the variable.

By Identifies a column that creates a report consisting of separate analyses for each level of the specified variable.

Scale Uniformly Represents all variables on the same scale, adding a *y*-axis to the plot. Without this option, each variable is on a different scale.

To allow for proper comparisons, select this option if your variables are measured on the same scale.

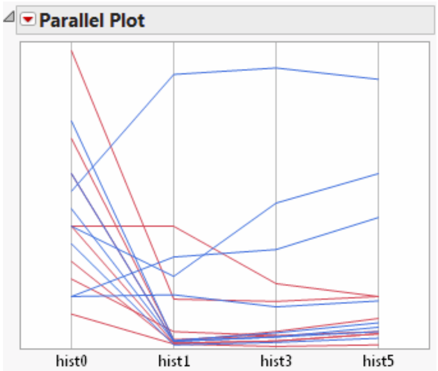
Center at zero Centers the parallel plot (not the variables) at zero.

After you click **OK**, the Parallel plot appears. See "The Parallel Plot" on page 202.

The Parallel Plot

To produce the plot shown in Figure 7.5, follow the instructions in "Example of a Parallel Plot" on page 199.

Figure 7.5 The Parallel Plot Report



A parallel plot is one of the few types of coordinate plots that show any number of variables in one plot. However, the relationships between variables might be evident only in the following circumstances:

- when the variables are side by side
- if you assign a color to a level of a variable to track groups
- if you select lines to track groups

Tip: You can add reference lines for specification limits. For information, see The Column Info Window chapter in *Using JMP*.

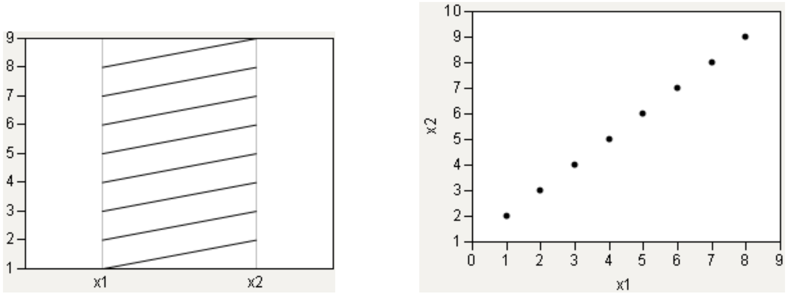
Interpreting Parallel Plots

To help you interpret parallel plots, compare the parallel plot with a scatterplot. In each of the following figures, the parallel plot appears on the left, and the scatterplot appears on the right.

Strong Positive Correlation

The following relationship shows a strong positive correlation. Notice the coherence of the lines in the parallel plot.

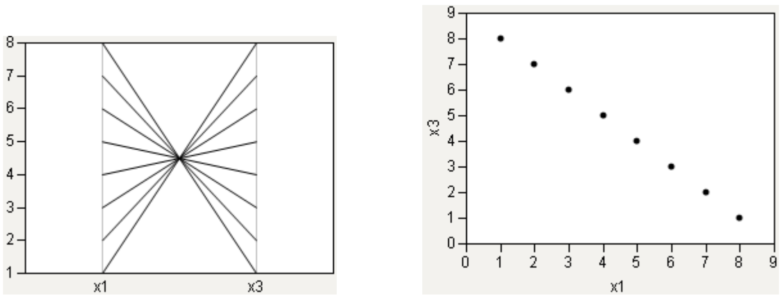
Figure 7.6 Strong Positive Correlation



Strong Negative Correlation

A strong negative correlation, by contrast, shows a narrow neck in the parallel plot.

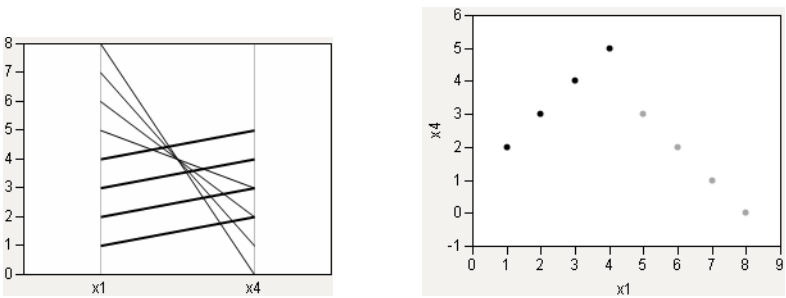
Figure 7.7 Strong Negative Correlation



Collinear Groups

Now, consider a case that encompasses both situations: two groups, both strongly collinear. One has a positive slope, the other has a negative slope. In Figure 7.8, the positively sloped group is highlighted.

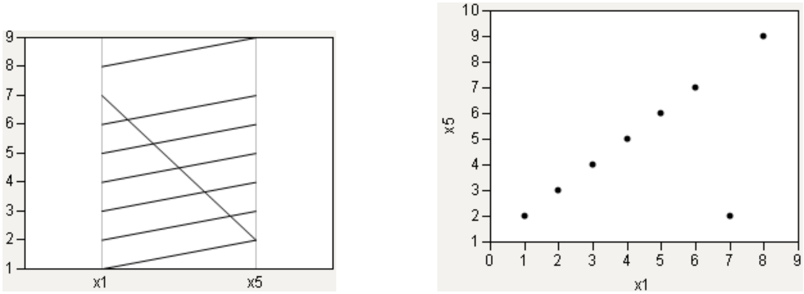
Figure 7.8 Collinear Groups: Parallel Plot and Scatterplot



Single Outlier

Finally, consider the case of a single outlier. The parallel plot shows a general coherence among the lines, with a noticeable exception.

Figure 7.9 Single Outlier: Parallel Plot and Scatterplot



Related Information

- "Additional Examples of the Parallel Plot Platform" on page 205

Parallel Plot Platform Options

The Parallel Plot red triangle menu provides the following options:

Show Reversing Checkboxes Reverses the scale for one or more variables.

See the JMP Reports chapter in *Using JMP* for more information about the following options:

Local Data Filter Shows or hides the local data filter that enables you to filter the data used in a specific report.

Redo Contains options that enable you to repeat or relaunch the analysis. In platforms that support the feature, the Automatic Recalc option immediately reflects the changes that you make to the data table in the corresponding report window.

Save Script Contains options that enable you to save a script that reproduces the report to several destinations.

Save By-Group Script Contains options that enable you to save a script that reproduces the platform report for all levels of a By variable to several destinations. Available only when a By variable is specified in the launch window.

For more information about the context menu options that appear when you right-click a parallel plot, see *Using JMP*.

Additional Examples of the Parallel Plot Platform

- "Examine Iris Measurements"
- "Examine Student Measurements"

Examine Iris Measurements

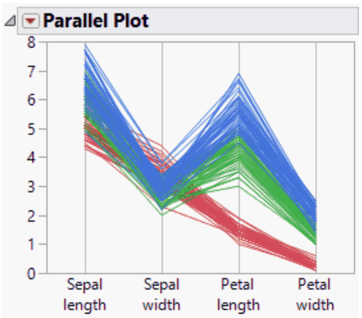
The following example uses the Fisher's Iris data set (Mardia, Kent, and Bibby 1979). The Iris.jmp sample data table contains measurements of the sepal length and width and petal length and width in centimeters for three species of Iris flowers: setosa, versicolor, and virginica. To find characteristics that differentiate the three species, examine these measurements.

Examine Three Species in One Parallel Plot

1. Select **Help > Sample Data Library** and open Iris.jmp.
2. Select **Graph > Parallel Plot**.
3. Select Sepal length, Sepal width, Petal length, and Petal width and click **Y, Response**.
4. Select the **Scale Uniformly** check box.
5. Click **OK**.

   The report window appears.

Figure 7.10 Three Species in One Parallel Plot



In this parallel plot, the three species are all represented in the same plot. The colors correspond to the three species, as follows:

- Blue corresponds to virginica.
- Green corresponds to versicolor.
- Red corresponds to setosa.

From the parallel plot, you observe the following:

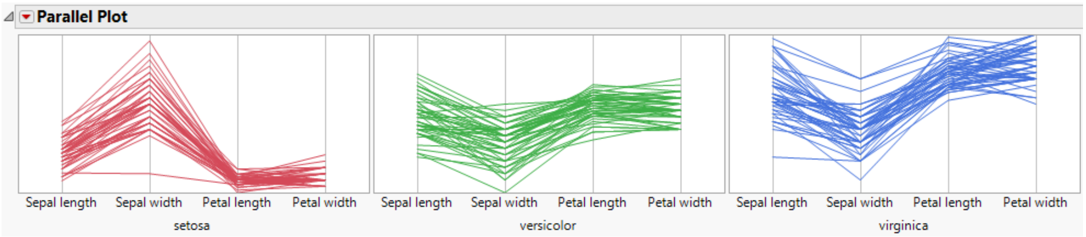
- For sepal width, the setosa values appear to be higher than the virginica and versicolor values.
- For petal width, the setosa values appear to be lower than the virginica and versicolor values.

Examine Three Species in Different Parallel Plots

1. From the Iris.jmp sample data table, select **Graph > Parallel Plot**.
2. Select Sepal length, Sepal width, Petal length, and Petal width and click **Y, Response**.
3. Select Species and click **X, Grouping**.
4. Click **OK**.

   The report window appears.

Figure 7.11 Three Species in Different Parallel Plots



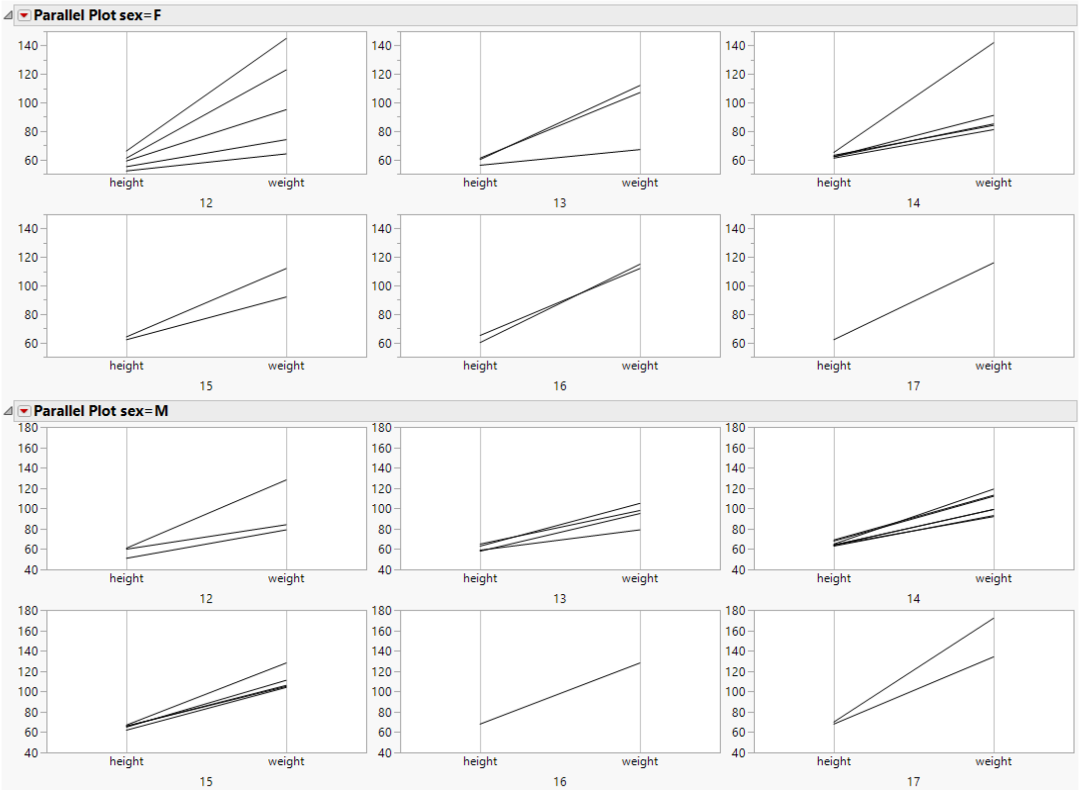
Each species is represented in a separate parallel plot.

Examine Student Measurements

The following example uses the Big Class.jmp sample data table, which contains data on age, sex, height, and weight for 40 students. Examine the relationships between different variables.

1. Select **Help > Sample Data Library** and open Big Class.jmp.
2. Select **Graph > Parallel Plot**.
3. Select height and weight and click **Y, Response**.
4. Select age and click **X, Grouping**.
5. Select sex and click **By**.
6. Select the **Scale Uniformly** check box.
7. Click **OK**.

Figure 7.12 Height and Weight by Sex, Grouped by Age



You observe the following:

- Among the 13-year-old females, one female’s weight is lower than the other females in her age group. If you click the line representing the lower weight, the respective individual (Susan) is highlighted in the data table.
- Among the 14-year-old females, one female’s weight is higher than the other females in her age group. If you click the line representing the higher weight, the respective individual (Leslie) is highlighted in the data table.

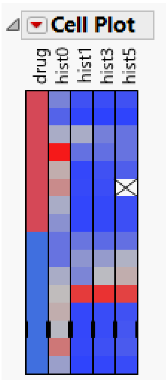
Chapter **8**

Cell Plots

View Color-Intensity Plots of Variables

Using cell plots, you can visualize each cell in a data table. Cell plots are direct representations of a data table, since they draw a rectangular array of cells where each cell corresponds to a data table entry. Cell plots were popularized by genomics applications to browse large numbers of values for gene expression levels.

Figure 8.1 Example of a Cell Plot



Contents

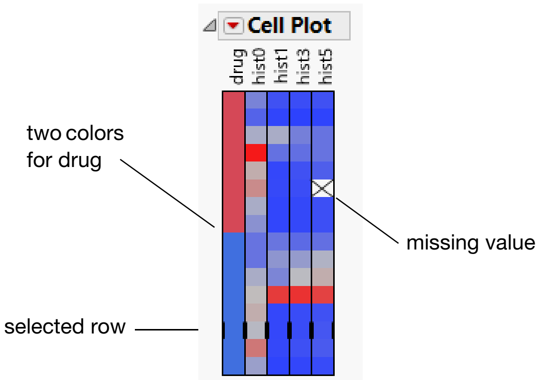
Example of a Cell Plot

This example uses the Dogs.jmp sample data table, which contains histamine level measurements for 16 dogs that were given two different drugs. The histamine levels were taken at zero, one, three, and five minutes. Examine the variation in the histamine levels for each drug.

1. Select **Help > Sample Data Library** and open Dogs.jmp.
2. Select the third row from the bottom (row 14).
3. Select **Graph > Cell Plot**.
4. Select drug, hist0, hist1, hist3, and hist5 and click **Y, Response**.
5. Click **OK**.

   The report window appears.

Figure 8.2 Dogs.jmp cell plot



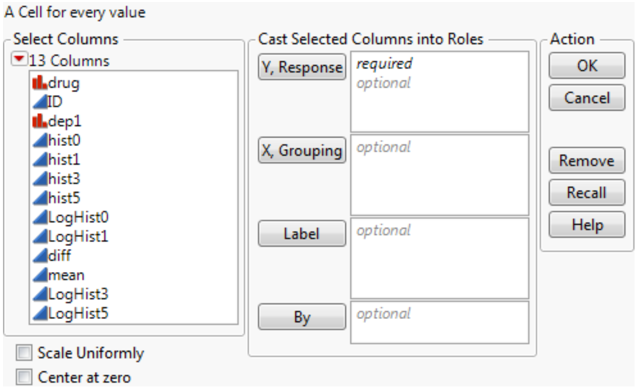
Notice the following:

- There are two types of drugs, represented by two distinct colors.
- Histamine levels are assigned colors from a gradient of blue to red.
- Any missing values are delineated by an X.
- The third row from the bottom is selected, and black lines appear next to the cells.

Launch the Cell Plot Platform

Launch the Cell Plot platform by selecting **Graph > Cell Plot**.

Figure 8.3 The Cell Plot Launch Window



Y, Response Variables appear on the horizontal axis of the cell plot. Each cell represents a value.

X, Grouping Produces a separate cell plot for each level of the variable.

Label Labels each row by the specified variable. See "Additional Example of the Cell Plot Platform" on page 215.

By Identifies a column that creates a report consisting of separate analyses for each level of the variable.

Scale Uniformly Represents all variables on the same scale. Without this option, each variable is on a different scale.

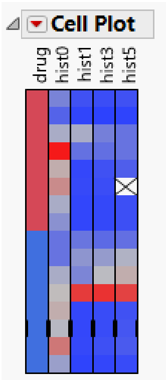
Center at zero Centers the cell plot at zero.

After you click **OK**, the Cell Plot window appears. See "The Cell Plot" on page 213.

The Cell Plot

To produce the plot shown in Figure 8.4, follow the instructions in "Example of a Cell Plot" on page 211.

Figure 8.4 The Cell Plot Report Window



Note: Any rows that are excluded in the data table are also hidden in the cell plot.

Cell plots are direct representations of a data table, drawn as a rectangular array of cells with each cell corresponding to a data table entry. Colors are assigned to each cell based on the range and type of values found in the column.

- Nominal variables use a distinct color for each level. You can customize nominal and ordinal colors using the **Value Colors** property of data columns, available through the **Column Info** command.
- Continuous variables are assigned a gradient of colors to show the smooth range of values in the variable.
- Ordinal variables are scaled like continuous variables in order.
- When some outliers are present, the scale uses all but the extreme categories for the 90% middle of the distribution, so that the outliers do not overly influence the scale.

The cell plot appears with a one-to-one correspondence of a colored cell representing each data table entry.

Related Information

- "Additional Example of the Cell Plot Platform" on page 215

Cell Plot Platform Options

The Cell Plot red triangle menu provides the following options:

Legend Shows or hides a legend.

Arrange Plots Specifies how many plots to put on the same row before starting the next row of plots.

This option is available only if you specify an **X, Grouping** variable.

See the JMP Reports chapter in *Using JMP* for more information about the following options:

Local Data Filter Shows or hides the local data filter that enables you to filter the data used in a specific report.

Redo Contains options that enable you to repeat or relaunch the analysis. In platforms that support the feature, the Automatic Recalc option immediately reflects the changes that you make to the data table in the corresponding report window.

Save Script Contains options that enable you to save a script that reproduces the report to several destinations.

Save By-Group Script Contains options that enable you to save a script that reproduces the platform report for all levels of a By variable to several destinations. Available only when a By variable is specified in the launch window.

Pop-Up Menu for Cell Plots

Right-click the cell plot to find the following options:

Graph Type Determines the appearance of the graph. See "Graph Type" on page 215.

Color Theme Shows a list of color themes that affect continuous variables in color maps. The default color theme is **Blue to Gray to Red** (corresponding to small values to middle values to large values). Use **White to Black** to create a gray-scale plot.

Note: To see custom colors, you must first create them. Select **File > Preferences > Graphs**. In the Color Themes area, click the type of color theme that you want to create, click **New**, and then change the colors. See the Enter and Edit Data chapter in *Using JMP* for more information about creating custom color themes.

Sort Ascending Sorts the rows of the plot from lowest to highest by the values of a column. To sort, right-click in the plot under a column and select **Sort Ascending**. The entire plot is rearranged to accommodate the sorting. See "Additional Example of the Cell Plot Platform" on page 215.

Sort Descending Sorts the rows of the plot from highest to lowest by the values of a column. To sort, right-click in the plot under a column and select **Sort Descending**. The entire plot is rearranged to accommodate the sorting.

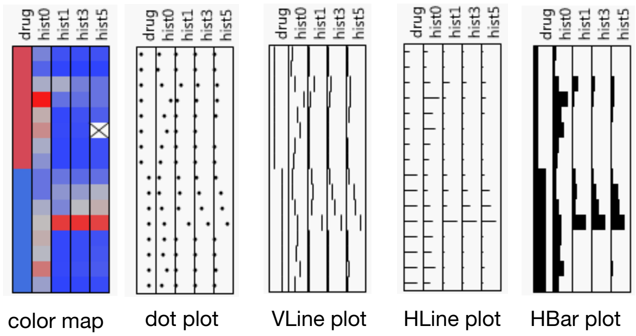
No Separator Lines Draws or removes lines separating the columns.

Note: For more information about the pop-up options that appear when you right-click labels, see the JMP Reports chapter in *Using JMP*.

Graph Type

Use the **Graph Type** option to change the appearance of the graph.

Figure 8.5 Graph Types



Additional Example of the Cell Plot Platform

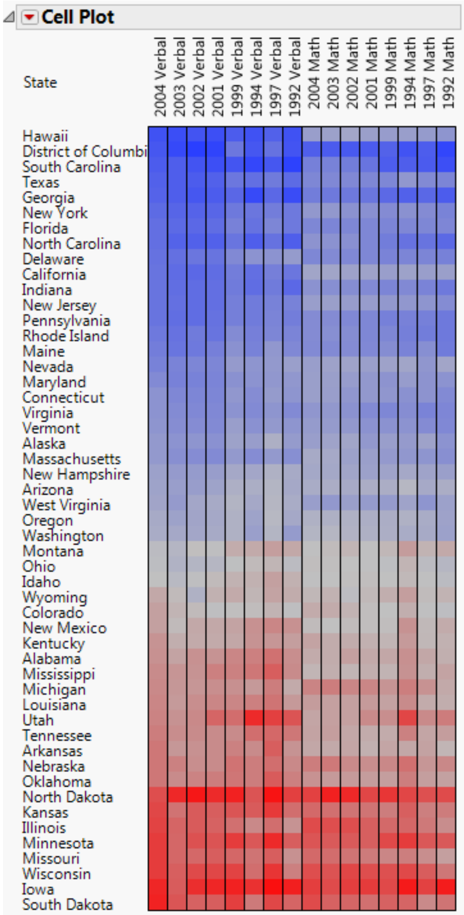
This example uses the SAT.jmp sample data table, which contains SAT test scores (divided into verbal and mathematics portions) for all 50 United States.

1. Select **Help > Sample Data Library** and open SAT.jmp.
2. Select **Graph > Cell Plot**.
3. Select all of the Verbal scores for all of the years, and click **Y, Response**.
4. Select all of the Math scores for all of the years, and click **Y, Response**.
5. Select State and click **Label**.
6. Click **OK**.

   The report window appears.
7. Right-click the plot under 2004 Verbal (the top left cell) and select **Sort Ascending**.

   This sorts the cell plot by the verbal scores for 2004.

Figure 8.6 Cell Plot for SAT Scores



You notice the following:

- Hawaii has the lowest verbal scores for 2004, and South Dakota has the highest verbal scores for 2004.
- There is a contrast between Hawaii's math and verbal scores. Hawaii has average math scores (represented by gray color values) but low verbal scores (represented by blue color values). Hawaii appears to be an outlier, since it has a strikingly different pattern for its math scores and its verbal scores.
- There is very little contrast between North Dakota's math and verbal scores. North Dakota's math and verbal scores are generally high (represented by red color values).

For a description of color themes, see "Pop-Up Menu for Cell Plots" on page 214.

Chapter 9

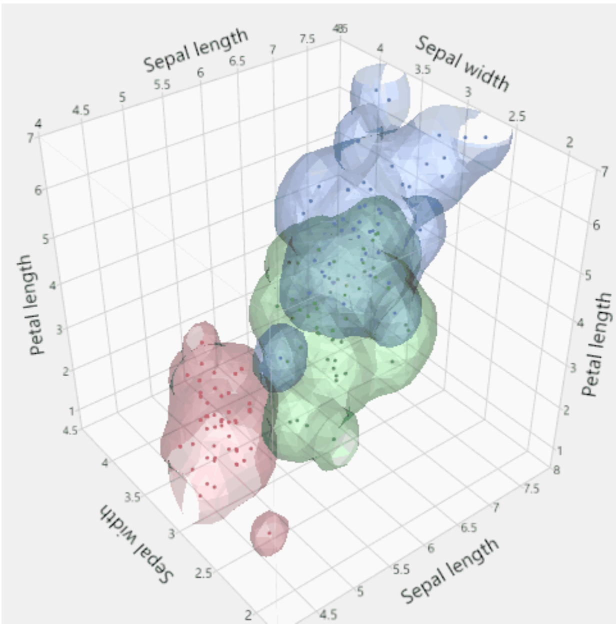
Scatterplot 3D

Create a Rotating Three-Dimensional View of Data

The Scatterplot 3D platform shows the values of numeric columns in the associated data table in a rotatable, three-dimensional view. Up to three columns that you select from the associated data table are displayed at one time.

To help visualize variation in higher dimensions, the 3D scatterplot can show a biplot representation of the points and variables when you request principal components. The most prominent directions of data are displayed on the 3D scatterplot report.

Figure 9.1 Example of a 3D Scatterplot



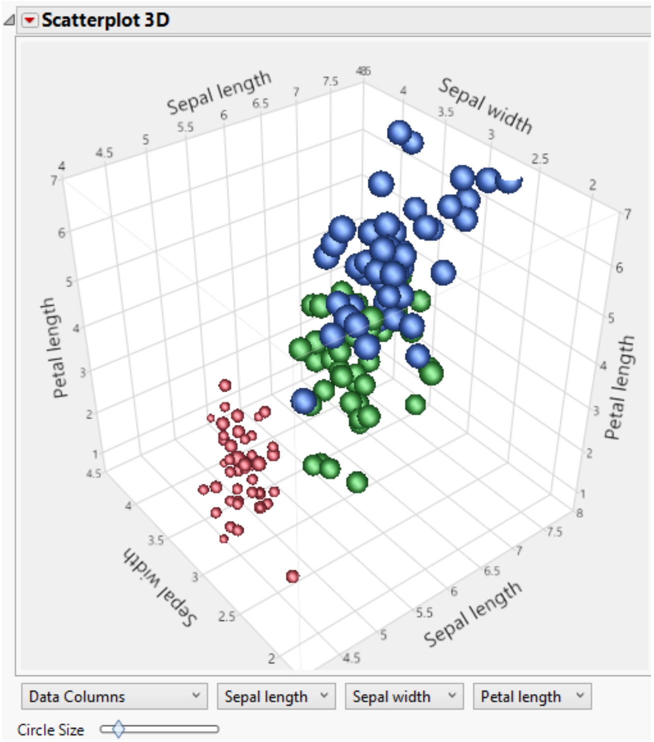
Contents

Example of a 3D Scatterplot

This example uses the Iris.jmp sample data table, which includes measurements of sepal length, sepal width, petal length, and petal width for three species of iris.

1. Open the Iris.jmp sample data table.
2. Select **Graph > Scatterplot 3D**.
3. Select **Sepal length**, **Sepal width**, and **Petal length**, and click **Y, Columns**.
4. Select **Petal width** and click **Weight**.
5. Click **OK**.

Figure 9.2 Example of an Initial 3D Scatterplot

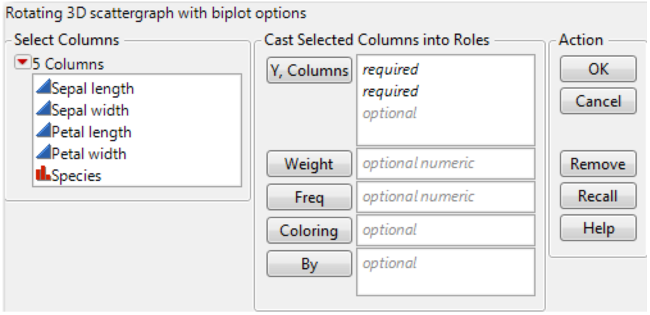


Now you can spin the 3D scatterplot to see the relationships between the variables. In this example, the data points are formatted in blue, red, and green. You might want to spin the scatterplot to see more clearly the relationships between the red and green points.

Launch the Scatterplot 3D Platform

Launch the Scatterplot 3D platform by selecting **Graph > Scatterplot 3D**.

Figure 9.3 The Scatterplot 3D Launch Window



Y, Columns Select the variables to plot on the 3D scatterplot. The order in which you select the variables determines where the data points appear on the axes:

– The first variable appears on the *x* axis.

– The second variable appears on the *y* axis.

– The third variable appears on the *z* axis.

You can assign the remaining variables interactively through the drop-down menus below the scatterplot.

Weight Use the **Weight** variable to:

– Assign a weight (importance or influence) to the data

– Visualize a fourth variable that sizes the points

Note: Red triangle options account for the **Weight** variable. If you do not want this variable accounted for in your analyses, remove it from the launch window.

When you specify a **Weight** variable, JMP draws the points as balls. The balls are scaled so that their volume represents the weight value. You click and drag the **Circle Size** slider below the scatterplot to resize the balls.

Freq Identifies the data table column whose variables assign a frequency to each row. This option is useful when a frequency is assigned to each row in summarized data.

Coloring Colors the points according to the selected variable. If the selected variable is categorical (nominal or ordinal), each category is colored distinctly. If the selected variable is continuous, a gradient of colors is used.

By Produces a separate 3D scatterplot for each By variable value. When two By variables are assigned, a separate graph is produced for each combination of both By variables.

After you click **OK**, the Scatterplot 3D report window appears. See "The Scatterplot 3D Report" on page 221.

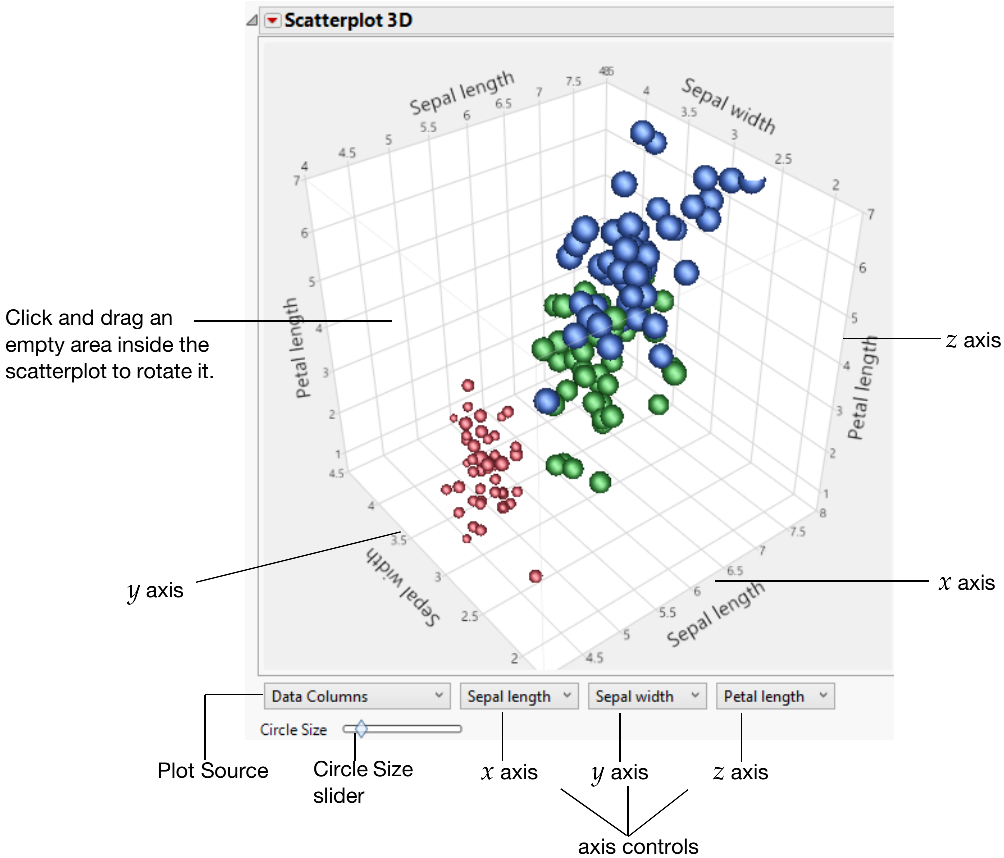
The Scatterplot 3D Report

To produce the 3D scatterplot shown in Figure 9.4, follow the instructions in "Example of a 3D Scatterplot" on page 219.

The Scatterplot 3D report shows a three-dimensional spinnable view of your data. See Figure 9.4. In the launch window, you select the variables and then create the report. The variables are displayed on the 3D scatterplots' x, y, and z axes. Up to three variables can be displayed at a time.

Note: The Crosshairs tool is not supported in 3D scatterplots.

Figure 9.4 Example of Information Displayed on the Scatterplot 3D Report



Note: Any rows that are excluded in the data table are also hidden in the 3D scatterplot.

Plot source The plot source box indicates the source of the data in the plot.

Circle Size slider Click and drag the **Circle Size** slider to resize the balls while maintaining their relative sizes.

> **Note:** The **Circle Size** slider appears only if you have specified a **Weight** variable.

Axis controls Select which variable appears on each axis. Choose the **Other** option to add a new variable.

Next Axis Set Cycles through the axis controls for any hidden variables. See "Change Variables on the Axes" on page 224.

Note: The **Next Axis Set** button appears only if your analysis contains more than three variables.

After you create a 3D scatterplot, you can add features such as displaying ellipses around specific data points, showing separate principal components, rotating components, connecting points, and more. See "Scatterplot 3D Platform Options" on page 225.

You can also assign colors and symbols (or markers) to data points either on the 3D scatterplot itself or in the associated data table. See "Assign Colors and Markers to Data Points" on page 225 and "Assign Colors and Markers in the Data Table" on page 225.

Spin the 3D Scatterplot

You spin the 3D scatterplot report in four ways:

- Click and drag an empty area on the 3D scatterplot. The 3D scatterplot spins in the direction you dragged the mouse.

  Note: Click and drag on an empty area on the 3D scatterplot, not on an axis or data point. Dragging the axis rescales the axis. Dragging a data point only selects the point.

- Slide the mouse wheel. The 3D scatterplot spins up and down only.
- Hold down an arrow key. (Before using an arrow on the number keypad, verify that NUM LOCK is turned off.)
- Hold down ESC. The 3D scatterplot spins left and right only.

In each case, the 3D scatterplot spins as long as you hold down the mouse button, arrow key, or Esc key. The spinning also continues as you slide the mouse wheel.

You can also spin the 3D scatterplot continuously as follows:

- Click and drag: Hold down Shift, click and drag an empty area on the plot, and release Shift. The faster you drag the mouse, the faster the 3D scatterplot spins.
- Mouse wheel: Hold down Shift, slide the wheel, and release the wheel. The 3D scatterplot spins up and down only.
- Arrow keys: Hold down Shift press the arrow key, and release Shift.
- Esc key: Hold down Shift and press Esc. The 3D scatterplot spins left and right only.

In addition to automatically spinning the plot, you can oscillate the plot. Hold down Shift and Ctrl and then click and drag the plot. The plot shakes up and down or left to right, depending on the direction in which you dragged the plot.

To stop the spinning or oscillating, click the plot or press Esc.

Change Variables on the Axes

The variables on each axis are determined by the order in which you select the variables in the launch window. For example, the first variable that you select is displayed on the x axis. The second variable is displayed on the y axis, and the third variable is displayed on the z axis.

After you create a 3D scatterplot, you can change the variable assigned to an axis, plot a different set of variables, or sequence through all combinations of the variables.

1. To change the variable on a specific axis, select the axis control drop-down menu and select a different variable.
2. To add a different variable, click an axis control drop-down menu, select **Other**, select the variable, and then click **OK**.
3. To sequence through combinations of all variables, click the **Next Axis Set** button until the variables that you want to plot are displayed.

Adjust the Axes

You can manually move or rescale the axis coordinates by clicking and dragging the axis. This option shows a different set of coordinates on the 3D scatterplot. It also lets you change the space displayed between the coordinates (or the coordinate scaling).

You can also specify axis properties by double-clicking the axis and modifying settings in the specifications window.

To Move the Coordinates on the Axis

1. Place your cursor over the middle of the axis.
2. Click and drag the axis.

To Modify Coordinate Scaling

1. Place your cursor over the end of the axis.
2. Click and drag the axis.

To Rescale an Axis Precisely

1. Place your cursor over the middle of the axis (the axis, not the label).
2. Double-click the axis.
3. Change the minimum and maximum coordinate values.

For more information about updating axes, see the JMP Reports chapter in *Using JMP*.

Assign Colors and Markers to Data Points

Each point in the 3D scatterplot corresponds to a row in the associated data table. To highlight points on the 3D scatterplot, you assign colors and markers to the points. The colors and markers are then displayed on the 3D scatterplot and in the data table.

When you click a point, the following items are selected:

- the point in the 3D scatterplot
- the corresponding row in the associated data table
- the point in any other opened graphs, if applicable

To select one point, click the point.

To select several points, double-click the 3D scatterplot and drag the cursor over the points. A box is displayed to indicate which points are selected.

To deselect points, double-click the 3D scatterplot.

To assign a color or marker to selected data points, proceed as follows:

1. To assign a color to the selected point, select **Rows > Colors** and then select the color.
2. To assign a marker to the selected point, select **Rows > Markers** and then select the marker.

Assign Colors and Markers in the Data Table

You can assign colors and markers to rows in the data table. The colors and markers appear next to the row number in the data table and on the 3D scatterplot. This option distinguishes points for each variable, and you can save the settings in the data table. Assigning colors and markers to specific data points (as described in "Assign Colors and Markers to Data Points" on page 225) only highlights them for the current open graphs.

See the Enter and Edit Data chapter in *Using JMP* for more information about assigning colors and markers in the data table. For more information about changing the size, quality, or transparency of markers, see "Scatterplot 3D Settings" on page 231.

Scatterplot 3D Platform Options

The red triangle menu next to Scatterplot 3D contains options to customize the display and to compute, rotate, and save principal or rotated components.

Show Points Shows or hides the data points on the graph.

Show Controls Shows or hides the source and axis controls displayed beneath the 3D scatterplot (Figure 9.4).

Drop Lines Draws lines from each point to the plane created by the *x* and *z* variables that you selected on the launch window.

Connect Points Connects the points with a line. Points can be connected on the data as a whole or in groups. You can also group data by a specific variable.

Jitter Displays small spaces between the data points so that you can see each point more clearly.

Normal Contour Ellipsoids Draws one or more normal contour ellipsoids, that is, three-dimensional ellipses that encompass a specified portion of points. You specify whether you want an ellipsoid for all of the data or for each group. You can also control the size and transparency of the ellipsoids. See "Normal Contour Ellipsoids" on page 228.

Ellipsoid Coverage Changes the size of normal contour ellipsoids. Type a value *between* 0 and 1, where the greater the value creates a bigger the ellipsoid. The actual values "0" and "1" produce no ellipsoid, so a warning appears if you try to use those values.

This option only appears after you add a normal contour ellipsoid to the 3D scatterplot.

Ellipsoid Transparency Changes the surface of normal contour ellipsoids. The greater the value, the more opaque the ellipsoid. This option only appears after you add a normal contour ellipsoid to the 3D scatterplot.

Nonpar Density Contour Draws nonparametric density contours, which approximately encompass a specified proportion of the points. You specify whether you want a density contour for all of the data or for each group. See "Nonparametric Density Contours" on page 228.

Drop Line Thickness Changes the width of drop lines. This option only appears after you add drop lines to the 3D scatterplot.

Principal Components Calculates principal components on all variables. This changes the axes of the plot to have principal component scores.

Biplot rays are displayed by default. You can remove them by selecting **Biplot Rays** from the red triangle menu. For more information about principal components, see the Principal Components chapter in *Multivariate Methods*.

Std Prin Components Calculates principal components (as with the **Principal Components** option) but scales the principal component scores to have unit variance. If this option is not selected, the scores have variance equal to the corresponding eigenvalue.

With standardized principal components, the correlation between the variables and the principal component scores is equal to the values in the eigenvector. This helps you quickly assess the relative importance of the variables. See the Principal Components chapter in *Multivariate Methods*.

Select this option if you want GH' rather than JK' biplots. GH' biplots try to preserve relationships between variables; JK' biplots try to preserve relationships between observations. The interpoint distance shown by GH' biplots is less meaningful, but the angles of the GH' biplot rays measure correlations better.

Rotated Components Specifies the number of factors that you want to rotate and the rotation method. You rotate components to better align the directions of the factors with the original variables so that the factors might be more interpretable. See the Principal Components chapter in *Multivariate Methods*.

Biplot Rays Shows or hides biplot rays that correspond to the principal components. You must have already selected Principal Components, Std Prin Components, or Rotated Components for this option to appear.

Show Ray Labels Shows or hides labels for the biplot rays. You must have already selected Biplot Rays for this option to appear.

Remove Prin Comp Removes principal components, standardized principal components, and rotated components from the scatterplot 3D report. The 3D scatterplot reverts to its original display before principal components were selected. This option, however, does not remove any saved principal components from the data table.

This option only appears after you add principal, standard, or rotated components to the 3D scatterplot.

Save Prin Components Saves the specified number of current principal component scores as new columns in the current data table. These columns also include the formulas used for the principal components. For *n* variables in the components list, *n* principal component columns are created and named Prin1, Prin2, ... Prinn.

This option only appears after you add principal, standard, or rotated components to the 3D scatterplot.

Save Rotated Components Saves all rotated component scores as columns in the current data table. These columns also include the formulas that were used. If you requested *n* rotated components, then *n* rotated component columns are created and named Rot1, Rot2, ... Rotn.

This option only appears after you add rotated components to the 3D scatterplot.

See the JMP Reports chapter in *Using JMP* for more information about the following options:

Local Data Filter Shows or hides the local data filter that enables you to filter the data used in a specific report.

Redo Contains options that enable you to repeat or relaunch the analysis. In platforms that support the feature, the Automatic Recalc option immediately reflects the changes that you make to the data table in the corresponding report window.

Save Script Contains options that enable you to save a script that reproduces the report to several destinations.

Save By-Group Script Contains options that enable you to save a script that reproduces the platform report for all levels of a By variable to several destinations. Available only when a By variable is specified in the launch window.

Normal Contour Ellipsoids

A normal contour ellipsoid is a 3-dimensional ellipse that encompasses a specified portion of points. The ellipsoid is computed from a contour of the multivariate normal distribution fit to the points. The ellipsoid is a function of the means, standard deviations, and correlations of variables on the plot. See *Multivariate Methods* for more information about multivariate normal distributions.

When you add an ellipsoid, two formatting options are available:

- *Coverage* changes the portion of data points covered by the ellipsoid. The larger the value, the bigger the ellipsoid.
- *Transparency* changes the surface of the ellipsoid from transparent to opaque. The larger the value, the more opaque the ellipsoid.

The coverage and transparency options also appear in the red triangle menu after you add the ellipsoid.

When you add normal contour ellipsoids to a 3D scatterplot, you specify whether you want an ellipsoid for all of the data or for a specific group of data. The ellipsoid for each set of grouped data is color-coded to differentiate one group from another.

You display and remove normal contour ellipsoids by selecting and deselecting **Normal Contour Ellipsoids** from the red triangle menu.

Related Information

- "Example of an Ungrouped Normal Contour Ellipsoid" on page 233
- "Example of Grouped Normal Contour Ellipsoids" on page 234

Nonparametric Density Contours

The nonparametric density contour shows contours that approximately encompass a specified proportion of the points. You add nonparametric density contours to see patterns in point density when the scatterplot is darkened by thousands of points.

This feature is particularly valuable when you have many points on a 3D scatterplot; the contours can be so dark that you cannot see the structure. In this situation, you remove the points so that only the contours are displayed. See "Optimizing a Dense Nonparametric Density Contour" on page 230.

When you add nonparametric density contours to a 3D scatterplot, you specify whether you want a contour for all of the data or for a specific group of data. The contour for each set of grouped data is color-coded to differentiate one group from another.

You display and remove nonparametric density contours by selecting and deselecting **Nonpar Density Contours** from the red triangle menu.

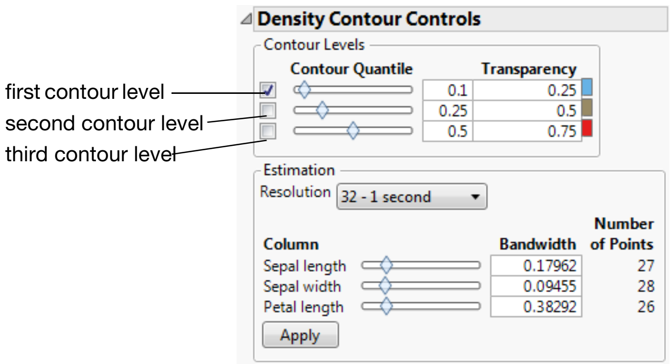
Related Information

- "Example of a Grouped Nonparametric Density Contour" on page 235

Density Contour Controls

The Density Contour Controls options are displayed below the 3D scatterplot. These options let you select additional contours and change each contour's formatting.

Figure 9.5 The Density Contour Controls Window



Contour Quantile Controls which contours are shown and lets you customize the contour formatting.

– *Density level* represents the volume and density of the points. As the contours go from smaller to larger values, the contours cover less volume but more dense areas. A 0.9 contour represents the 10% densest part of the total, where the points are closest together. Click and drag the slider below "Contour Quantile," or enter a value next to the slider.

– *Transparency* changes the surface of density contours. The greater the value, the more opaque the contour. Enter a value in the box.

– *Color* changes the color of the contour. Click the colored box and select a different color. (This option only appears for ungrouped density contours.)

Changes to these settings take effect immediately.

Resolution Changes the resolution of the contours. A higher resolution results in a less granular drawing of the contours but takes more time to display.

Column Bandwidth Changes the smoothness of the fitted density. A higher bandwidth results in a smoother fitted density.

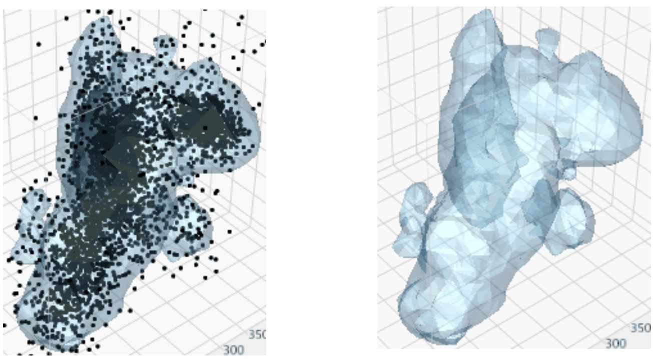
Type a new bandwidth for each variable, or click and drag the sliders. Click **Apply** to display your changes.

Optimizing a Dense Nonparametric Density Contour

When you have many points on a 3D scatterplot, the contours can be so dark that you cannot see the structure. In this situation, you remove the points so that only the contours are displayed.

To remove points from a 3D scatterplot, select **Show Points** from the red triangle menu. You can further optimize the contours by changing their size, color, and transparency. See "Scatterplot 3D Platform Options" on page 225.

Figure 9.6 Example of Optimizing a Dense Nonparametric Density Contour



3D scatterplot with density contour and points

3D scatterplot with density contour and no points

Pop-Up Menu

Right-click the 3D scatterplot to find the following options.

Show Legend Shows and hides the color legend for the 3D scatterplot.

Note: This option requires the **Coloring** role.

Reset Returns the orientation of the scatterplot to its original state.

Settings Provides options to change the appearance of the 3D scatterplot. See “Scatterplot 3D Settings” on page 231.

Hide Lights Border Shows and hides a border that displays the lights. The lights highlight different portions of the 3D scatterplot.

Right-click a light to turn it on or off and to change the color.

Wall Color Changes the color of the 3D scatterplot.

Background Color Changes the color surrounding the 3D scatterplot.

Rows You can color, mark, exclude, hide, and label points that correspond to rows in the associated data table. You must select the points before selecting this option. See the JMP Reports chapter in *Using JMP*.

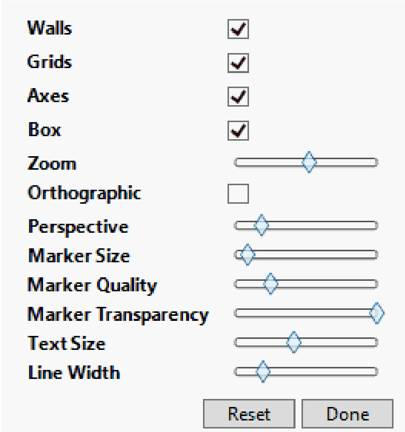
Use Hardware Acceleration Turns hardware acceleration on or off for machines that support acceleration. This option might display the scatterplot faster. If not, try updating your graphics drivers.

Show ArcBall Shows and hides a globe around the 3D scatterplot. This option helps you visualize the rotation of the scatterplot. Select whether you want the ArcBall to appear always, only when you drag the scatterplot, or never.

Scatterplot 3D Settings

To customize properties such as the marker size, text size, and grid lines, right-click the 3D scatterplot and select **Settings**. The Settings window appears. As you modify the settings, a preview appears on the 3D scatterplot.

Figure 9.7 The Scatterplot 3D Settings Window



Note the following:

- Move the sliders left to decrease the selected property or to the right to increase the selected property.
- To move the Settings window around the scatterplot, click and drag the top portion of the window.

The following options are available:

Reset Resets the default settings.

Done Closes the window.

Walls Adds or removes the 3D scatterplot walls. Without walls, the background color of the 3D scatterplot is displayed.

Grids Shows or hides the coordinate lines.

Axes Shows or hides the variable names that appear above each axis.

Box Shows or hides the box. Without the box, the 3D scatterplot is displayed as an open plot.

Zoom Enlarges or shrinks the 3D scatterplot.

Orthographic Changes the view of the scatterplot from 3-dimensional to an orthographic projection. In the orthographic view, the walls of the scatterplot do not converge to a vanishing point. This means that you can compare near and far distances and see the structure between data points.

Note: If you turn off orthographic view and completely decrease the perspective, the walls of the scatterplot do not converge. This is the same effect that you get when you turn on orthographic view.

Perspective Increases or decreases the perspective. Large values create a view that is unnaturally large and visually disorienting. In this case, you need to resize the scatterplot window to show the entire plot.

Marker Size Increases or decreases the size of the data point markers.

Marker Quality Increases and decreases the data marker quality. For example, when you increase the marker quality, some markers have an opaque center. Other symbol markers are formatted in bold. Increase the zoom to see these changes in quality.

Marker Transparency Increases or decreases the transparency of the data markers.

Text Size Increases or decreases the text size.

Line Width Changes the width of the coordinate and axes lines.

Additional Examples of the Scatterplot 3D Platform

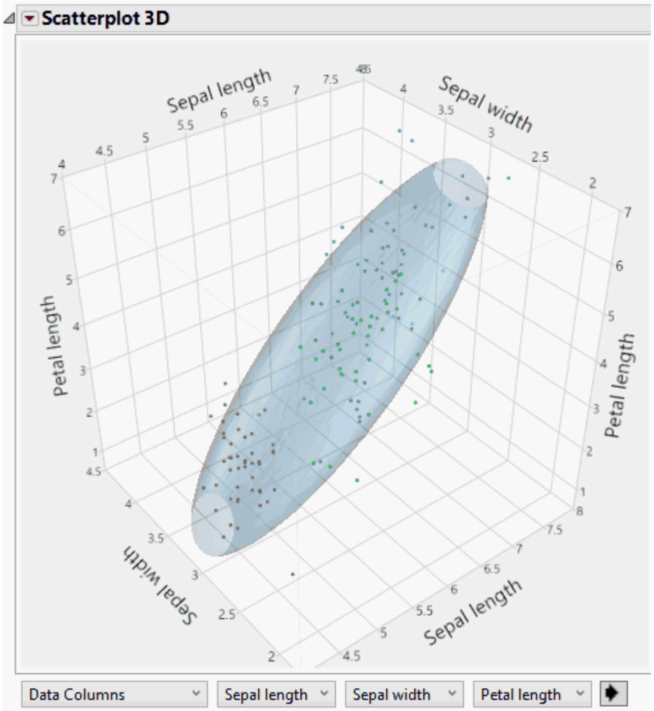
- "Example of an Ungrouped Normal Contour Ellipsoid"
- "Example of Grouped Normal Contour Ellipsoids"
- "Example of a Grouped Nonparametric Density Contour"

Example of an Ungrouped Normal Contour Ellipsoid

This example shows how to add a normal contour ellipsoid to more than 75% of the data points. The ellipsoid is 25% transparent.

1. Select **Help > Sample Data Library** and open Iris.jmp.
2. Select **Graph > Scatterplot 3D**.
3. Select Sepal length, Sepal width, Petal length, and Petal width and click **Y, Columns**.
4. Click **OK**.
5. Click the Scatterplot 3D red triangle and select **Normal Contour Ellipsoids**. Notice that **Ungrouped** is already selected.
6. Type 0.75 next to Coverage.
7. Type 0.25 next to Transparency.
8. Click **OK**.

Figure 9.8 Example of an Ungrouped Normal Contour Ellipsoid

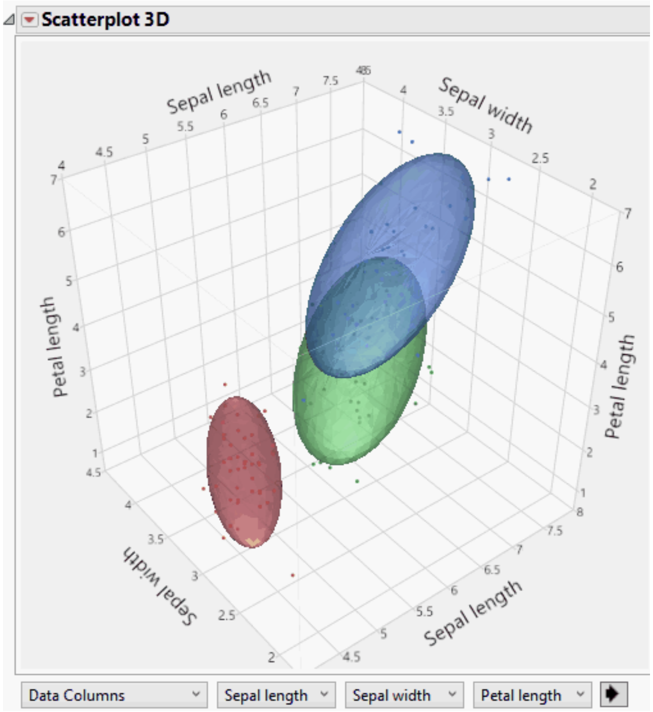


Example of Grouped Normal Contour Ellipsoids

This example shows how to group measurements by species and to format each group with a normal contour ellipsoid. The ellipsoids cover 75% of the data points and are 50% transparent. The contours are color-coded (using JMP default colors) based on species.

1. Select **Help > Sample Data Library** and open Iris.jmp.
2. Select **Graph > Scatterplot 3D**.
3. Select Sepal length, Sepal width, Petal length, and Petal width and click **Y, Columns**.
4. Click **OK**.
5. Click the Scatterplot 3D red triangle and select **Normal Contour Ellipsoids**.
6. Select **Grouped by Column**.
7. Select **Species**.
8. Type 0.75 next to Coverage.
9. Type 0.5 next to Transparency.
10. Click **OK**.

Figure 9.9 Example of Grouped Normal Contour Ellipsoids

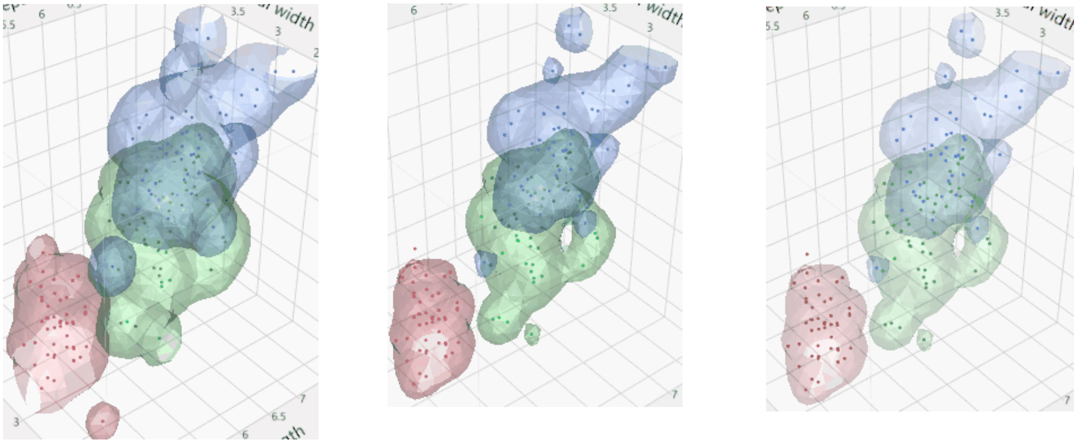


Example of a Grouped Nonparametric Density Contour

This example shows how to group data points and format each nonparametric density contour.

1. Select **Help > Sample Data Library** and open Iris.jmp.
2. Select **Graph > Scatterplot 3D**.
3. Select Sepal length, Sepal width, Petal length, and Petal width and click **Y, Columns**.
4. Click **OK**.
5. Click the Scatterplot 3D red triangle and select **Nonpar Density Contour**.
6. Select **Grouped by Column**.
7. Select **Species** and click **OK**. A different colored contour is displayed for each of the three species.
8. Type 0.25 in the first Contour Quantile box. 25% of the data points appear outside the contour surfaces, which results in smaller contours.
9. Type 0.15 in the first Transparency box. The contours are 15% opaque.

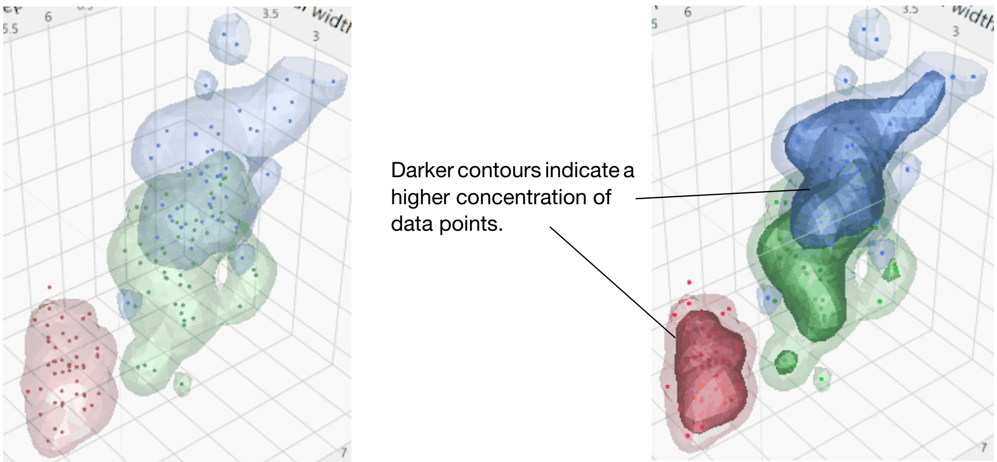
Figure 9.10 Changing the Nonparametric Density Contour Transparency and Density



original contours decreased density decreased transparency

10. Select the second check box. The contour quantiles are the same (.25), so the new contours overlap the first ones.
11. Type 0.5 in the second Contour Quantile box. 50% of the data points appear outside the contour surfaces. A second set of contours appears within the first, to further illustrate the density of the data points.

Figure 9.11 Adding a Second Nonparametric Density Contour



one level of contours two levels of contours

You can now format the second levels of contours and turn on the third level of contours.

The options for formatting the grouped and ungrouped nonparametric density contours are similar. The only difference is that you cannot change the color of each *grouped* nonparametric density contour. See "Scatterplot 3D Platform Options" on page 225 for options.

Chapter 10

Contour Plots

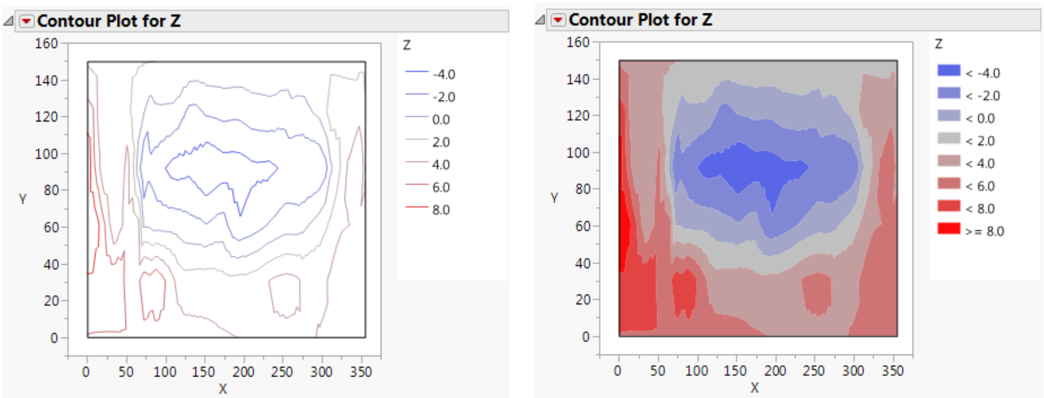
View Multidimensional Relationships in Two Dimensions

The **Contour Plot** command in the **Graph** menu constructs contours of a response in a rectangular coordinate system. A contour plot shows a three-dimensional surface in two dimensions. Contours delineate changes in the third dimension.

Here are some of the options available with the Contour platform:

- specify the number of contour levels
- choose to plot contour lines or filled contours
- show or hide data points
- label contours with response values
- define and use a custom coloring scheme

Figure 10.1 Examples of Contour Plots



Contents

Example of a Contour Plot

To create a contour plot, you can use either the Contour platform or Graph Builder. This section provides examples for both.

- To see the example using the Contour platform, see "Example Using the Contour Platform" on page 241.
- To see the example using Graph Builder, see "Example of a Contour Plot in Graph Builder" on page 242.

Example Using the Contour Platform

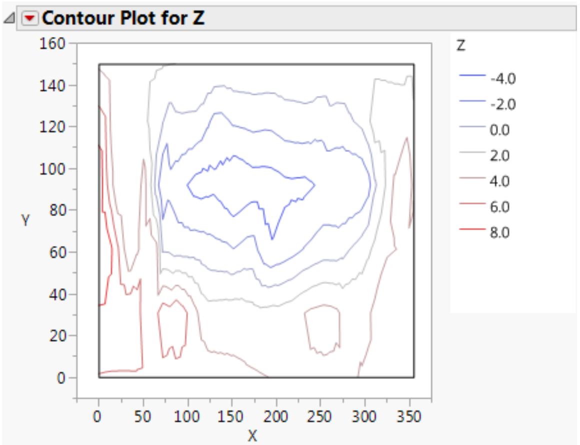
To create a contour plot, you need two variables for the *x*- and *y*-axes and at least one more variable for contours. You can also use several *y*-variables. This example uses the Little Pond.jmp sample data table. X and Y are coordinates of a pond. Z is the depth.

1. Select **Help > Sample Data Library** and open Little Pond.jmp.
2. Select **Graph > Contour Plot**.
3. Select the X and Y coordinates and click **X**.
4. Select the depth, Z, and click **Y**.

Note: In a contour plot, the X1 and X2 roles are used for the X and Y axes.

5. Click **OK**.

Figure 10.2 Example of a Contour Plot with Legend



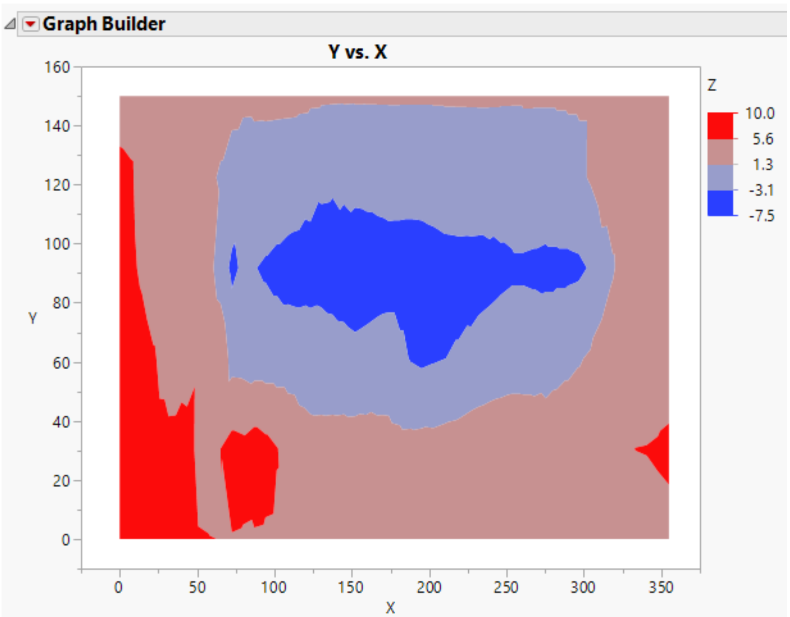
The *x*- and *y*-axes are coordinates and the contour lines are defined by the depth variable. This contour plot is essentially a map of a pond showing depth. To see the contours more clearly, click the red triangle next to Contour Plot for Z and select **Fill Areas**.

Example of a Contour Plot in Graph Builder

You can create the same contour plot shown in Figure 10.2 using Graph Builder.

1. Select **Help > Sample Data Library** and open Little Pond.jmp.
2. Select **Graph** > **Graph Builder**.
3. Click and drag the X coordinate to the **X** zone.
4. Click and drag the Y coordinate to the **Y** zone.
5. Click and drag depth, Z, to the **Color** zone.
6. Click the Contour  icon.

Figure 10.3 Contour Plot in Graph Builder



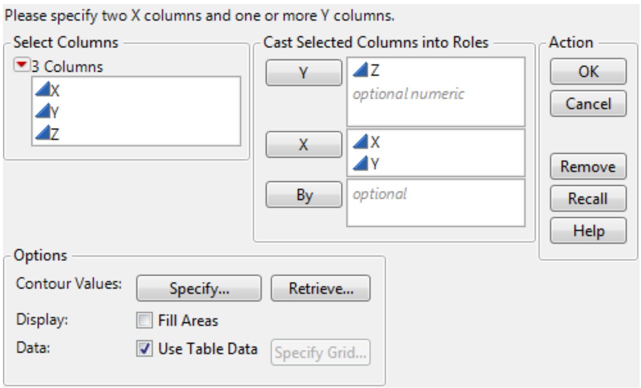
Notice the plot produced from the Contour platform looks slightly different than the plot produced using Graph Builder.

Launch the Contour Plot Platform

Launch the Contour Plot platform by selecting **Graph > Contour Plot**.

By default, the contour levels used in the plot are values computed from the data. You can specify your own number of levels and level increments in the Launch window before you create the plot. You can also do so in the red triangle menu for Contour Plot after you create the plot. You can use a column formula to compute the contour variable values.

Figure 10.4 The Contour Plot Launch Window



Y Columns assigned to the **Y** role are used as variables to determine the contours of the plot. You must specify at least one, and you can specify more than one.

You can also assign a column with a formula to this role. If you do so, the formula should be a function of exactly two variables. Those variables should be the *x* variables entered in the Launch window.

X Columns assigned to the **X** role are used as the variables for the *x*- and *y*-axes. You must specify exactly two columns for **X**.

By This option produces a separate graph for each level of the By variable. If two By variables are assigned, a separate graph for each possible combination of the levels of both By variables is produced.

Options

Contour Values Specify your own number of levels and level increments. See "Contour Specification" on page 247.

Fill Areas Fill the areas between contour lines using the contour line colors.

Use Table Data and Specify Grid Most often, you construct a contour plot for a table of recorded response values. In that case, **Use Table Data** is selected and the **Specify Grid** button is unavailable.

However, if a column has a formula and you specify that column as the response (**Y**), the **Specify Grid** button becomes available. When you click **Specify Grid**, you can define the contour grid in any way, regardless of the rows in the existing data table. This feature is also available with table templates that have one or more columns defined by formulas but no rows. See "Use Formulas for Specifying Contours" on page 249.

After you click **OK**, the Contour plot appears. See "The Contour Plot" on page 244.

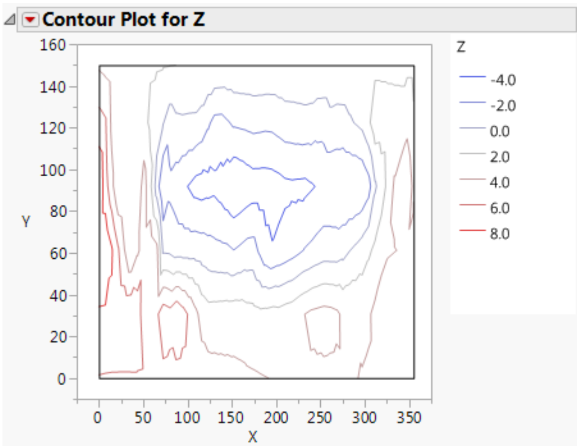
The Contour Plot

Follow the instructions in "Example of a Contour Plot" on page 241 to produce the plot shown in Figure 10.5.

The legend for the plot shows individual markers and colors for the *Y* variable. Replace variables in the plot by dragging and dropping a variable, in one of two ways: swap existing variables by dragging and dropping a variable from one axis to the other axis; or, click a variable in the Columns panel of the associated data table and drag it onto an axis.

For information about additional options for the report, see "Contour Plot Platform Options" on page 245.

Figure 10.5 The Contour Plot Report



Contour Plot Platform Options

Using the options in the red triangle menu next to Contour Plot, you can tailor the appearance of your contour plot and save information about its construction.

Show Data Points Shows or hides (x, y) points. The points are hidden by default.

Show Missing Data Points Shows or hides points with missing y values. Available only if **Show Data Points** is selected.

Show Contours Shows or hides the contour lines or fills. The contour lines are shown by default.

Show Boundary Shows or hides the boundary of the total contour area. The boundary is shown by default.

Show Control Panel Shows or hides the **Alpha** slider that allows you to change the **Alpha** shapes filter.

Transform If the contour plot includes a **Color** role, the **Transform** option is enabled. See "Additional Example of Contour Plots" on page 250.

None The triangulation is computed without any scaling to coordinates using Delaunay triangulation. Delaunay triangles are computed to maximize the minimum angle of the triangles in the triangulation. This value is selected by default.

Range Normalized The X1/X2 values are both scaled to [0,1] prior to computing the triangulation. If the X1/X2 limits are different, then this is a non-uniform scale. This option may be more desirable in cases where the X1/X2 units are very different.

Fill Areas Fills the areas between the contours with a solid color. It is the same option that is available in the Launch window. If you leave it deselected in the Launch window, you can see the line contours before filling the areas. See "Fill Areas" on page 246.

Label Contours Shows or hides the label (z-value) of the contour lines.

Color Theme Select another color theme for the contours.

Reverse Colors Reverses the order of the colors assigned to the contour levels.

Change Contours Set your own number of levels and level increments. See "Contour Specification" on page 247.

Save This menu has options to save information about contours, triangulation, and grid coordinates to data tables. See "Contour Plot Save Options" on page 249.

See the JMP Reports chapter in *Using JMP* for more information about the following options:

Local Data Filter Shows or hides the local data filter that enables you to filter the data used in a specific report.

Redo Contains options that enable you to repeat or relaunch the analysis. In platforms that support the feature, the Automatic Recalc option immediately reflects the changes that you make to the data table in the corresponding report window.

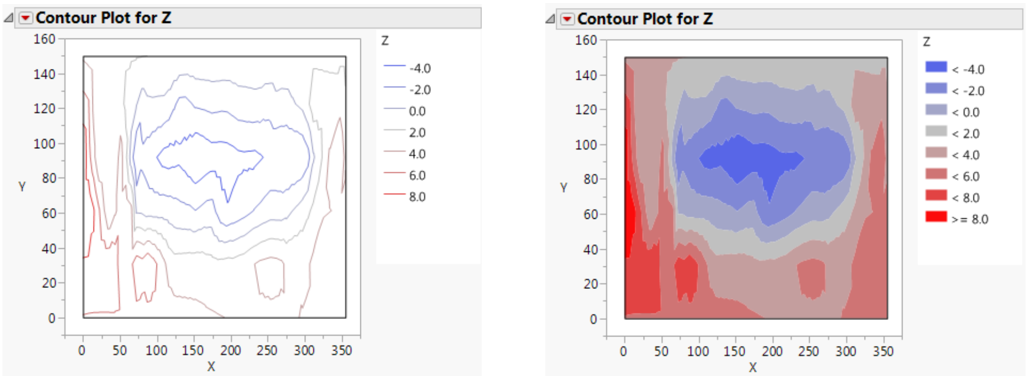
Save Script Contains options that enable you to save a script that reproduces the report to several destinations.

Save By-Group Script Contains options that enable you to save a script that reproduces the platform report for all levels of a By variable to several destinations. Available only when a By variable is specified in the launch window.

Fill Areas

If you select **Fill Areas**, the areas between contour lines are filled with the contour line colors. This option is available in the Launch window and in the red triangle menu for Contour Plot. Figure 10.6 shows a plot with contour lines on the left and a plot with the contour areas filled on the right.

Figure 10.6 Comparison of Contour Lines and Area Fills



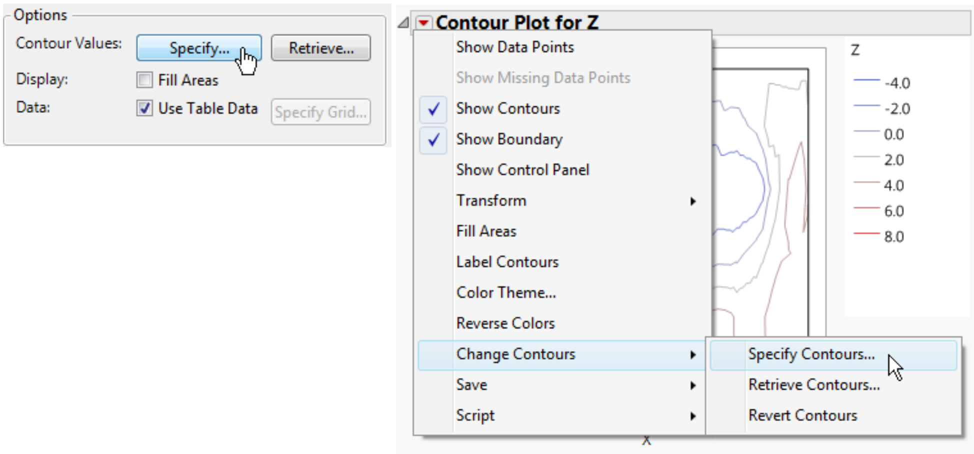
Areas are filled from low to high values. An additional color is added in the filled contour plot for the level above the last, and highest, contour line. The fill area colors shown in the plot are the JMP default colors.

Contour Specification

If you do not select options in the Launch window, the default plot spaces the contour levels equally within the range of the *Y* variable. You can specify a color theme by selecting a Continuous Color Theme in **File > Preferences > Graphs**. You can customize colors for individual contours by right-clicking on an item in the Contour Plot legend.

You can specify contour levels either in the Launch window (the **Specify** button) or in the report window from the red triangle menu for Contour Plot (the **Specify Contours** option).

Figure 10.7 Example of Contour Specification: Launch Window (on the left) and Menu (on the right)



Specify

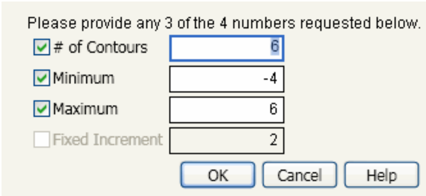
This option is both in the Launch window and on the red triangle menu for Contour Plot (the **Specify Contours** option).

Selecting this option displays the Contour Specification window (Figure 10.8). Using this window, you can do the following:

- change the number of contours
- specify minimum and maximum values to define the range of the response to be used in the plot
- change the increment between contour values

You supply any three of the four values, and the remaining value is computed for you. Click the check box to deselect one of the numbers and automatically select the remaining check box.

Figure 10.8 The Contour Specification Window



Colors are automatically assigned and are determined by the number of levels in the plot. After the plot appears, you can right-click (press CONTROL and click on macOS) on any contour in the plot legend and choose from the JMP color palette to change that contour color.

Retrieve

This option is both in the Launch window (the **Retrieve** button) and on the red triangle menu for Contour Plot (the **Retrieve Contours** option).

Note: Neither the button nor the menu option are active unless there is an open data table *in addition to* the table that has the contour plotting values. When you click **Retrieve** or select **Retrieve Contours**, a window with a list of open data tables appears.

Using this option, you can retrieve the following from an open JMP data table:

- the number of contours
- an exact value for each level
- a color for each level

From the list of open data tables, select the data table that contains the contour levels.

For level value specification, the Contour Plot platform looks for a numeric column with the same name as the response column that you specified in the Launch window. The number of rows in the data table defines the number of levels.

If there is a row state column with color information, those colors are used for the contour levels. Otherwise, the default platform colors are used.

Revert Contours

This option appears only on the red triangle menu for Contour Plot.

If you have specified your own contours, selecting this option reverts your Contour Plot back to the default contours.

Contour Plot Save Options

This menu has options to save information about contours, triangulation, and grid coordinates to data tables.

Save Contours Creates a new JMP data table with columns for the following:

– the x- and y-coordinate values generated by the Contour platform for each contour

– the response computed for each coordinate set

– the curve number for each coordinate set

The number of observations in this table depends on the number of contours you specified. You can use the coordinates and response values to look at the data with other JMP platforms. For example, you can use the Scatterplot 3D platform to get a three-dimensional view of the pond.

Generate Grid Displays a window that prompts you for the grid size that you want. When you click **OK**, the Contour platform creates a new JMP data table with the following:

– the number of grid coordinates you requested

– the contour values for the grid points computed from a linear interpolation

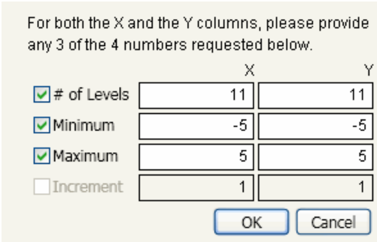
Save Triangulation Creates a new JMP data table that lists coordinates of each triangle used to construct the contours. By default, JMP uses Delaunay triangulation to connect the nearest data points to form triangles. The resulting set of triangles are calculated so that no other data points are inside a triangle's circumscribed circle, that is, the circle that passes through the three vertices of the triangle. To change the triangulation to a normalized scale, select **Transform > Range Normalized**.

Use Formulas for Specifying Contours

Most often you construct a contour plot for a table of recorded response values such as the Little Pond data table. In that case, in the launch window, **Use Table Data** is checked and the **Specify Grid** button is unavailable. However, if a column has a formula and you specify that column as the response (**Y**), the **Specify Grid** button becomes active.

When you click **Specify Grid**, the window shown in Figure 10.9 appears.

Figure 10.9 Example of the Contour Specification for Formula Column



You can complete the Specify Grid window and define the contour grid in any way, regardless of the rows in the existing data table. This feature is also available with table templates that have one or more columns defined by formulas but no rows.

Additional Example of Contour Plots

This example illustrates how to create a triangulation data table, to transform the triangulation to use Delaunay triangles, and to filter Alpha shapes of the triangles.

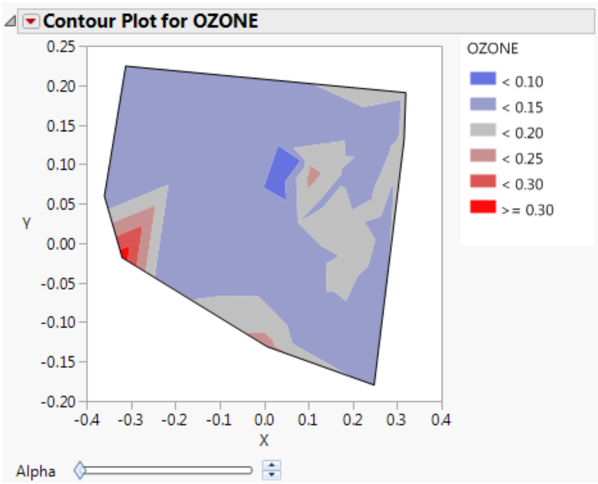
1. Select **Help > Sample Data Library** and open Cities.jmp.
2. Select **Graph > Contour Plot**.
3. Select OZONE and click **Y**.
4. Select X and Y and click **X**.
5. Select to **Fill Areas**.
6. Click **OK**.
7. Click the red triangle next to Contour Plot for OZONE and select the following options:
   a. **Transform** > **Range Normalized**

      Instead of Delaunay triangulation, this changes the method for calculating the triangulations to a normalized scale ([0,1]) in both X and Y.

   b. **Show Control Panel**

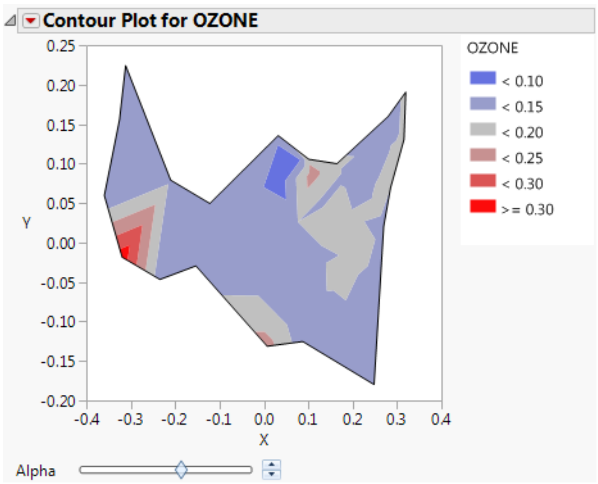
      The **Alpha** slider appears.

Figure 10.10 Contour Plot for OZONE



8. Click and move the **Alpha** slider to the right.

Figure 10.11 Alpha Shapes Filter



Using the Alpha slider filters out the larger Delaunay triangulation areas.

Chapter 11

Ternary Plots

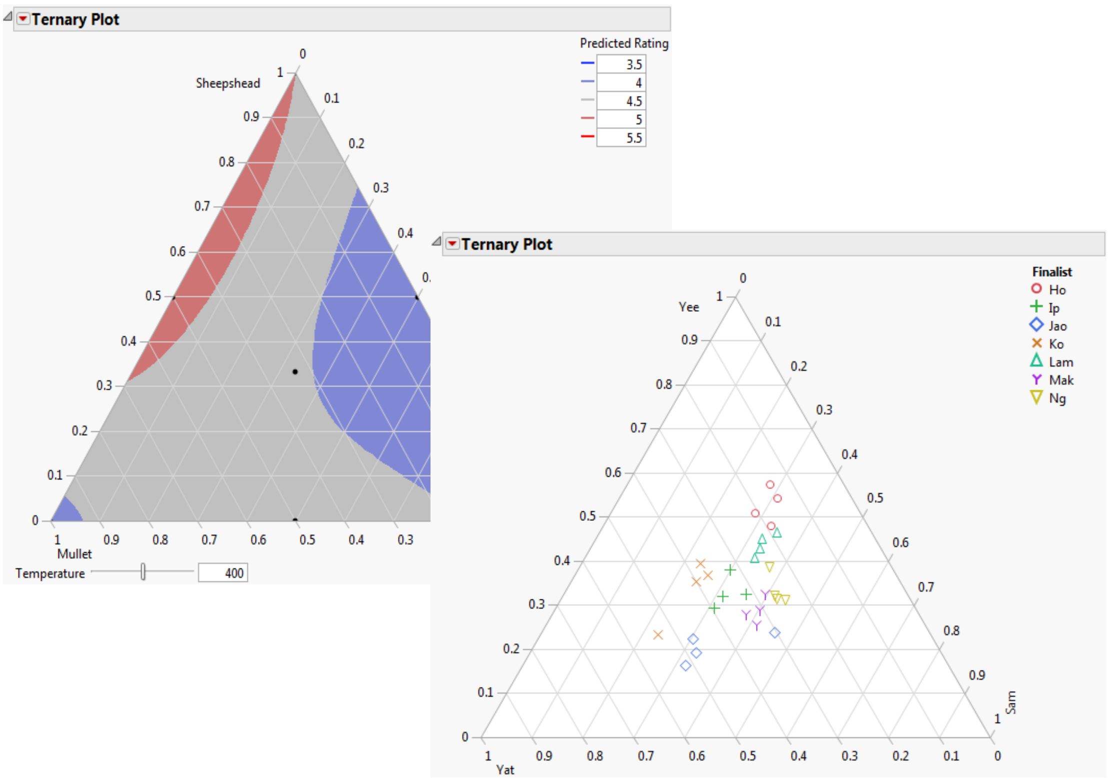
View Plots for Compositional or Mixture Data

The Ternary Plot command in the Graph menu produces a three-axis plot.

Ternary plots are a way of displaying the distribution and variability of three-part compositional data. (For example, the proportion of sand, silt, and clay in soil or the proportion of three chemical agents in a trial drug.) You can use data expressed in proportions or use absolute measures.

The ternary display is a triangle with sides scaled from 0 to 1. Each side represents one of the three components. A point is plotted so that a line drawn perpendicular from the point to each leg of the triangle intersect at the component values of the point.

Figure 11.1 Examples of Ternary Plots



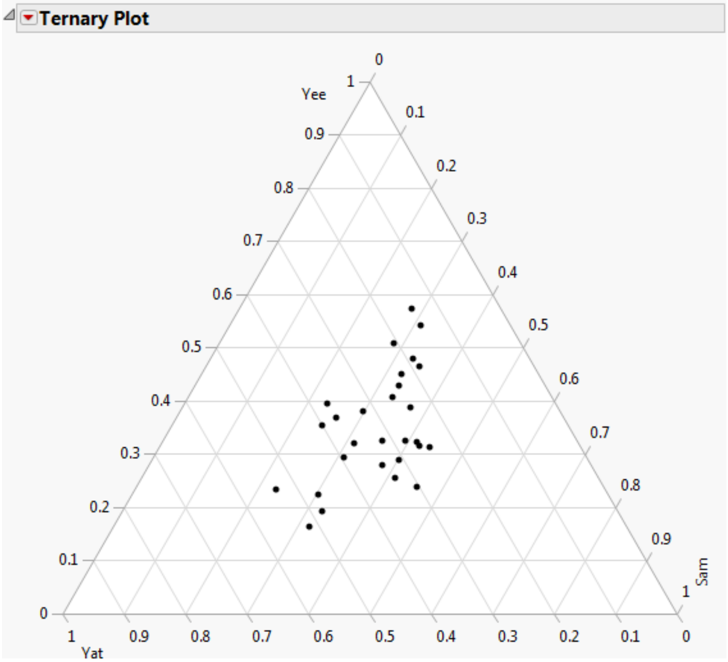
Contents

Example of a Ternary Plot

This example uses the Pogo Jumps.jmp sample data table. The data, adapted from Aitchison (1986), show measurements for pogo jumps of seven finalists in the 1985 Hong Kong Pogo-Jump Championship. A single pogo jump is the total jump distance in three consecutive bounces, referred to as *yat*, *yee*, and *sam*.

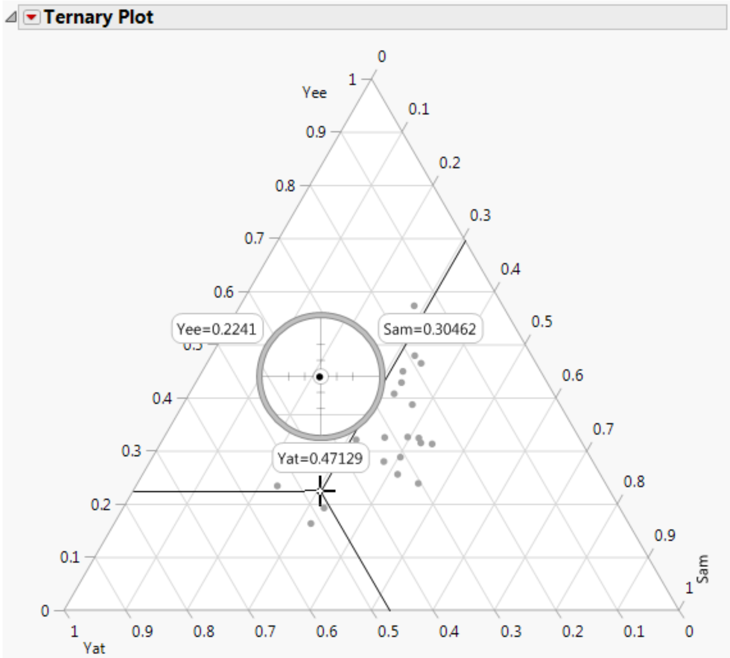
1. Select **Help > Sample Data Library** and open Pogo Jumps.jmp.
2. From the **Graph** menu, select **Ternary Plot**.
3. Select Yat, Yee, and Sam and click **X, Plotting**.
4. Click **OK**.

Figure 11.2 Example of a Ternary Plot



Use the crosshairs tool to determine exact coordinates of points within the plot.

Figure 11.3 Using the Crosshairs Tool

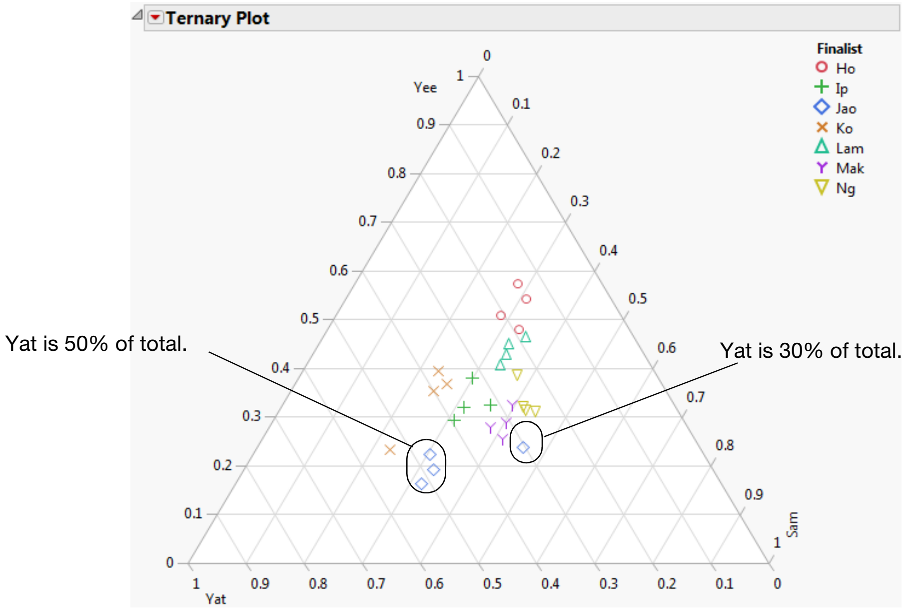


To get a better idea of how the three bounces contribute to total distance, assign each contestant's points a different color and marker.

1. Right-click the plot and select **Row Legend**.
2. Select Finalist in the column list box.

   Colors should be automatically set to **JMP Default**.
3. Select **Standard** from the **Markers** menu.
4. Click **OK**.

Figure 11.4 Pogo Data Colored by Finalist

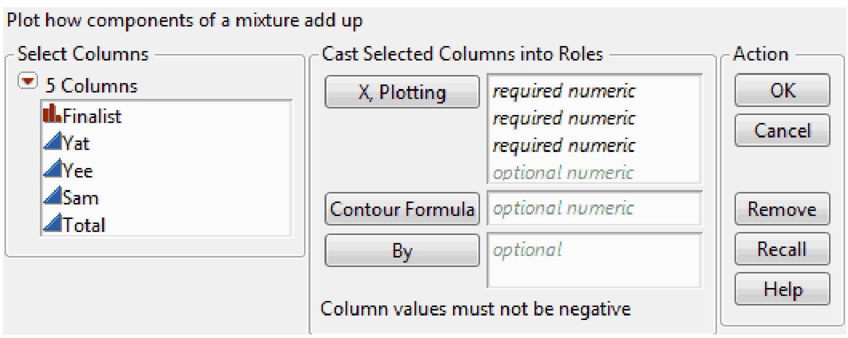


Note that most of the finalists are consistent in the composition of total distance. However, two finalists, Jao and Ko, both have one jump that is not consistent with their other jumps. For example, for three of Jao's jumps, the Yat composed about 50% of the total distance, but for the other jump, the Yat composed only 30% of the total distance. That jump is not consistent with the others. A similar observation can be made about Ko's jumps.

Launch the Ternary Plot Platform

Launch Ternary Plot by selecting **Graph > Ternary Plot**.

Figure 11.5 The Ternary Plot Launch Window



X, Plotting Assign three columns to generate a ternary plot.

If you assign more than 3 variables to the **X, Plotting** role, a matrix of ternary plots is created. A separate variable is assigned to the first two axes of a plot, with the third axis being the sum of the other variables. If necessary, the variables are scaled so they sum to 1.

Contour Formula To plot contours of a response surface, assign a column containing a formula to the **Contour Formula** role. If you have variables in a Contour formula that are not listed as **X, Plotting** variables, JMP appends sliders below the plot so that the values can be interactively adjusted. See "Example Using a Contour Function" on page 261.

By This option produces a separate graph for each level of the By variable.

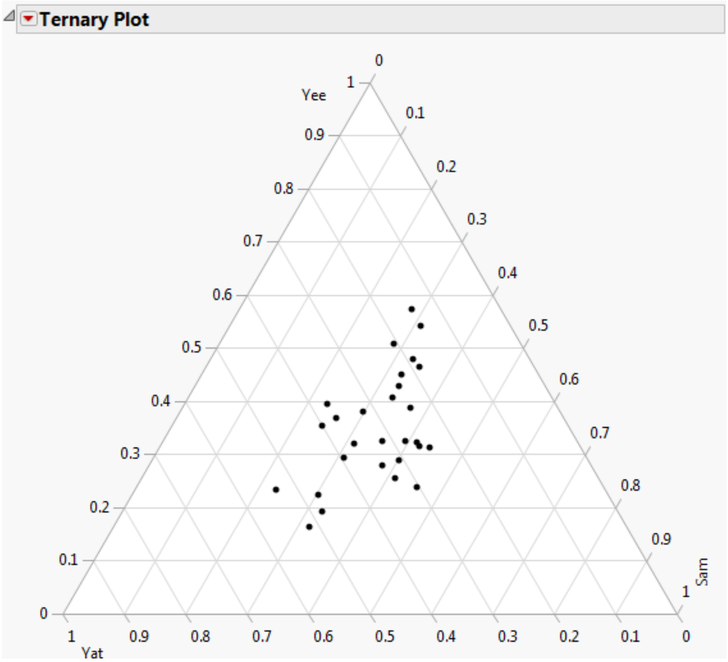
After you click **OK**, the Ternary Plot window appears. See "The Ternary Plot" on page 258.

The Ternary Plot

Follow the instructions in "Example of a Ternary Plot" on page 255 to produce the plot shown in Figure 11.6.

Each of the three sides of a ternary plot represents a proportion of 0%, with the point of the triangle opposite that base representing a proportion of 100%. As a proportion increases in any one sample, the point representing that sample moves from the base to the opposite point of the triangle.

Figure 11.6 The Ternary Plot



Mixtures and Constraints

Ternary Plot uses the Mixture column property to shade the portion of the graph that is out of bounds. The Ternary plot recognizes the mixture sum, mixture bounds, and general linear constraints. For information about setting the Mixture column property in the Column Info window, see The Column Info Window chapter in *Using JMP*.

Related Information

- "Example Using Mixture Constraints" on page 260

Ternary Plot Platform Options

The Ternary Plot red triangle menu contains options to modify the plot.

Note: To view more detailed options, right-click the plot.

Show Constraints Shows or hides the constraints on the plot. The default plot shows the constraints.

Contour Fill Allows filling of contours if a contour formula is specified in the plot. You can select Lines Only, Fill Above, or Fill Below. The default platform shows lines only.

Color Theme Allows you to select a color theme. The default plot shows the Blue to Gray to Red color theme.

Show Points Shows or hides the plotted points. The default plot shows the points.

See the JMP Reports chapter in *Using JMP* for more information about the following options:

Local Data Filter Shows or hides the local data filter that enables you to filter the data used in a specific report.

Redo Contains options that enable you to repeat or relaunch the analysis. In platforms that support the feature, the Automatic Recalc option immediately reflects the changes that you make to the data table in the corresponding report window.

Save Script Contains options that enable you to save a script that reproduces the report to several destinations.

Save By-Group Script Contains options that enable you to save a script that reproduces the platform report for all levels of a By variable to several destinations. Available only when a By variable is specified in the launch window.

Additional Examples of the Ternary Plot Platform

- "Example Using Mixture Constraints"
- "Example Using a Contour Function"

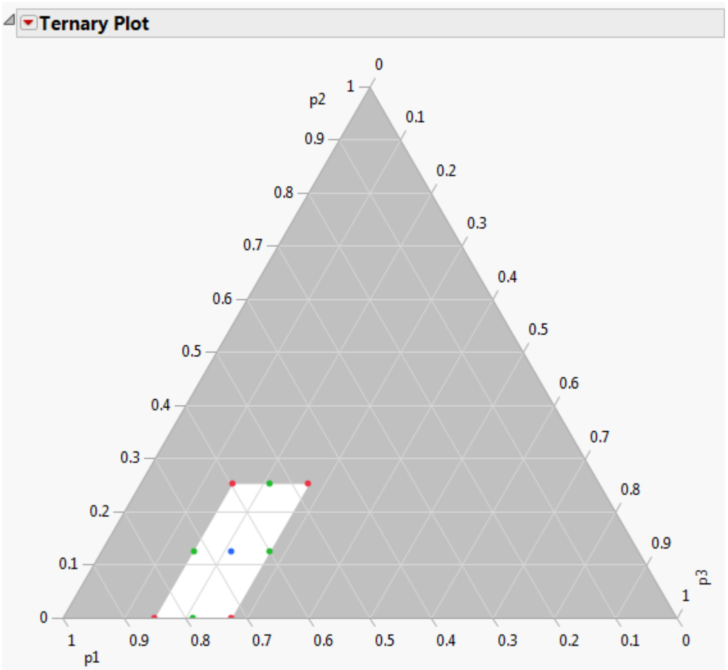
Example Using Mixture Constraints

1. Select **Help > Sample Data Library** and open Plasticizer.jmp.

   The p1, p2, and p3 columns all have Mixture Column Properties defined.
2. From the **Graph** menu, select **Ternary Plot**.

   Notice that p1, p2, and p3 appear next to **X, Plotting**.
3. Click **OK**.

Figure 11.7 Mixture Constraints in a Ternary Plot



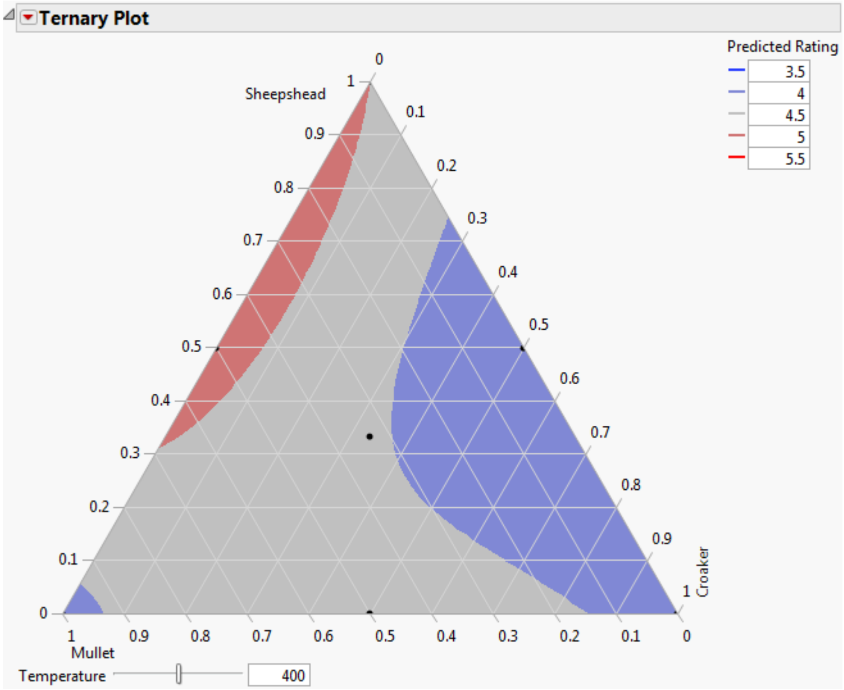
For more information about mixtures, see the Mixture Profiler chapter in *Profilers*.

Example Using a Contour Function

The data in Fish Patty.jmp is adapted from Cornell (1990) and comes from an experiment to optimize the texture of fish patties. The columns Mullet, Sheepshead, and Croaker represent what proportion of the patty came from those fish types. The column Temperature represents the oven temperature used to bake the patties. The column Rating is the response and is a measure of texture acceptability, where higher is better. A response surface model was fit to the data and the prediction formula was stored in the column Predicted Rating. See the Mixture Profiler chapter in *Profilers*.

1. Select **Help > Sample Data Library** and open Fish Patty.jmp.
2. From the **Graph** menu, select **Ternary Plot**.
3. Select Mullet, Sheepshead, and Croaker and click **X, Plotting**.
4. Select Predicted Rating and click **Contour Formula**.
5. Click **OK**.
6. Click the Ternary Plot red triangle and select **Contour Fill > Fill Above**.

Figure 11.8 Ternary Plot with a Contour Formula



The manufacturer wants the rating to be at least 5. You can drag the slider for Temperature and see the contours for the Predicted Rating change. Each point represents a mixture of the three fish. Any given mixture of fish types receives different ratings according to the temperature at which the patties are baked.

In this example, the red shaded area shows the mixture of fish that results in a rating of 5 to 5.5. Any purple areas show the mixture of fish that results in a rating of 4 and below. At 400 degrees, a mixture of mostly sheepshead and mullet with very little croaker results in a rating of 5 and above.

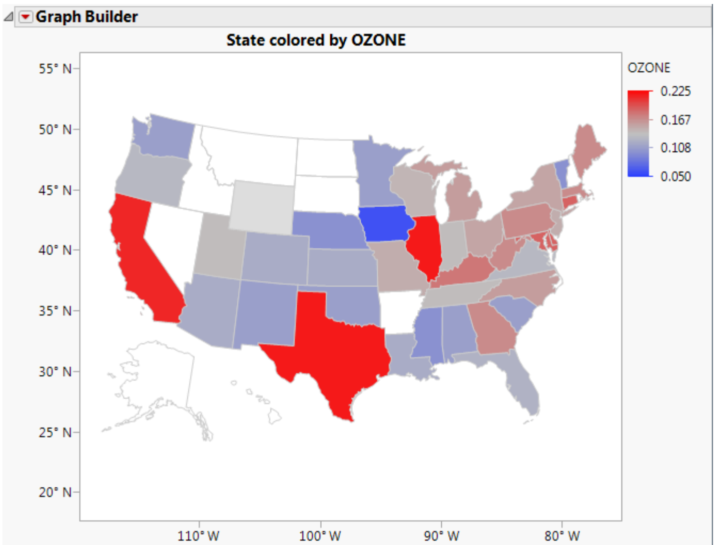
Chapter 12

Maps

Add Maps or Custom Shapes to Enhance Data Visualization

JMP transforms numbers and geographic data into compelling images, and turns simple tables of numbers into captivating pictures that bring the story in your data to life. JMP can help you display your data on geographical maps. Choose from built-in high-quality images. Select Street Map Service or Web Map Service to get custom map images from the Internet. JMP includes shape files for borders or many geographic regions and lets you add your own custom shapes, such as for a manufacturing plant or campus.

Figure 12.1 Example of a Map



Contents

Overview of Mapping

There are two types of map support in JMP: one where a map shows the data (Graph Builder) and one where a map provides context for the data (Background Maps). You can also create your own maps.

Graph Builder

You can interact with Graph Builder to create compelling visualizations of your data. JMP includes graphical support to display analyses using background maps and shape files. You can add color and geographical boundaries to maps through the following zones:

- The **Map Shape** zone assigns geographical boundaries to a map based on variables in the data table. The map shape value determines the x and y axes.

  Boundaries such as U.S. state names, Canadian provinces, and Japanese prefectures are installed with JMP. You can also create your own boundaries (geographical or otherwise) and specify them as a **Map Role** column property in the data table.
- The **Color** zone applies color based on a variable to geographical shapes.
- The **Size** element scales map shapes according to the size variable, minimizing distortion.

Background Maps

You can add background maps to any JMP graph through the Set Background Map window. You can use built-in background maps or connect to a Web Map Service (WMS) to display specialty maps like satellite images, radar images, or roadways. Right-click in a graph and select **Graph > Background Map** to choose from the following images and boundaries:

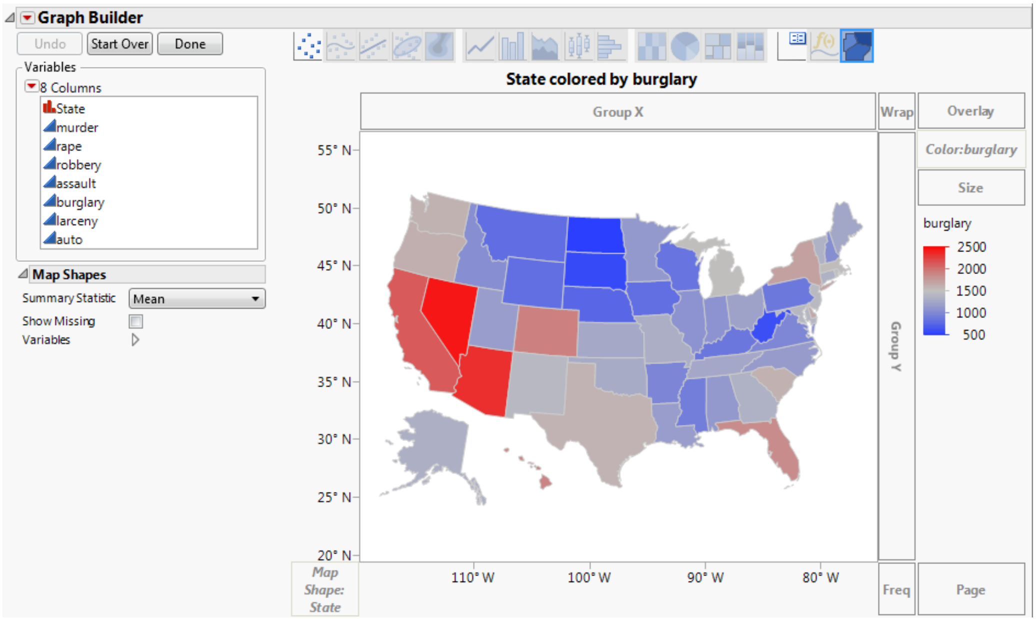
- **Simple Earth** and **Detailed Earth** maps are installed with JMP.
- **NASA server** provides maps using a WMS to show their most up-to-date maps.
- **Street Map Service** provides street maps. The OpenStreetMap and Open Database License links provide further information on the Street Map Service.
- **Web Map Service** lets you enter the URL for a website that provides maps using the WMS protocol. You can also specify the map layer.
- Boundaries for various regions.

Example of Creating a Map in Graph Builder

This example uses the Crime.jmp sample data table, which contains data on crime rates for each US state.

1. Select **Help > Sample Data Library** and open Crime.jmp.
2. Select **Graph > Graph Builder**.
3. Drag and drop State into the **Map Shape** zone.
4. Drag and drop Burglary into the **Color** zone.

Figure 12.2 Example of Burglary by State



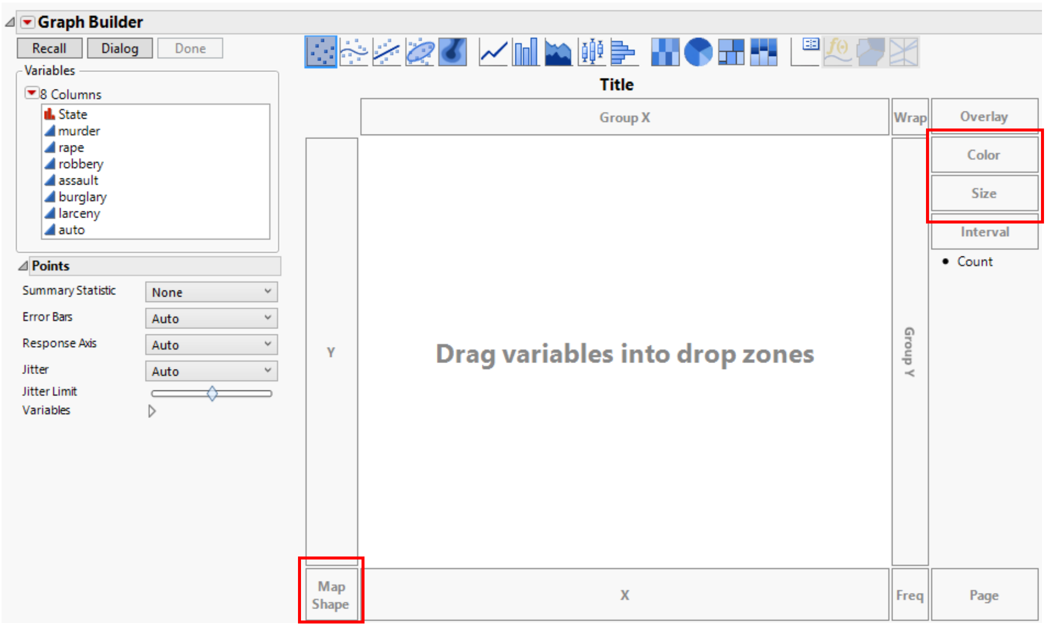
Note the following:

- The latitude and longitude appear on the Y and X axes.
- The legend shows the colors that correspond to the burglary rates. Since Burglary is a continuous variable, the colors are based on a continuous color theme. Figure 12.2 uses the JMP default continuous theme. You can change the theme under File > Preferences > Graphs.
- The map is projected so that relative areas are not distorted (the 49th parallel across the top of the US is not a straight line).

Graph Builder

Open a data table that contains geographic data. Launch Graph Builder by selecting **Graph** > **Graph Builder**. The primary element in the Graph Builder window is the graph area. The graph area contains drop zones (Map Shape, Color and Size), and you can drag and drop variables into the zones. From here you can map shapes for data tables that include place names.

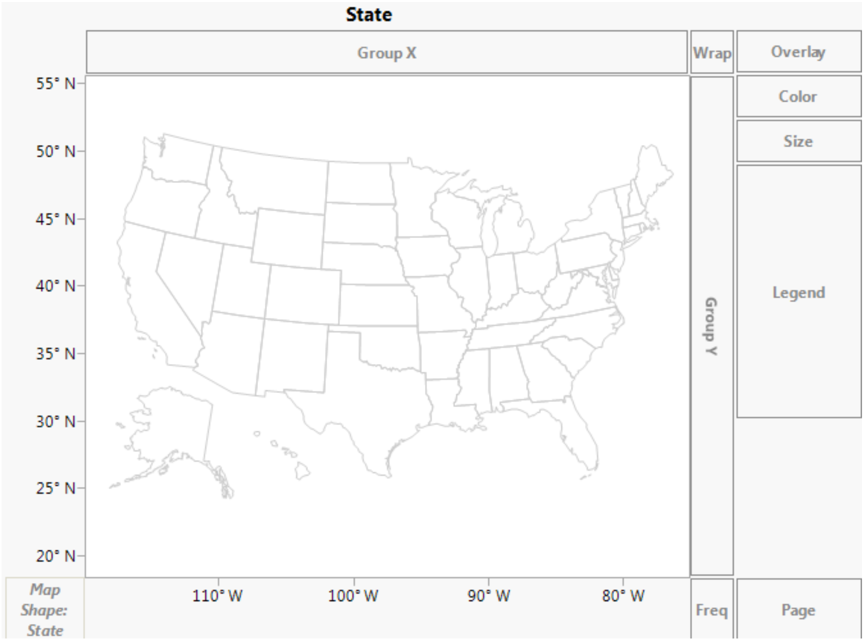
Figure 12.3 The Graph Builder Window



Map Shape

When a column contains the names of geographical regions (such as countries, regions, states, provinces, counties), you can assign the column to the **Map Shape** zone. When a variable is dropped in **Map Shape**, Graph Builder looks for map shapes that correspond to the values of the variable and draws the corresponding map. The variable can have a column property that tells JMP where to find the map data. If not, JMP looks through all known map files. If you have a variable in the Map Shape zone, the X and Y zones disappear. The Map Shape zone is positional and influences the types of graph elements that are available.

Figure 12.4 Example of Cities.jmp After Dragging State to Map Shape



For each map there are two .jmp files; one for the name data (one row per entity) and one for coordinate data (many rows per entity). They are paired via a naming convention; xxx-Name.jmp and xxx-XY.jmp, where "xxx" is some common prefix. Some examples of sample files that are shipped with the product are:

- World-Name.jmp
- World-XY.jmp
- US-State-Name.jmp
- US-State-XY.jmp

Map Name Files

Each xxx-Name.jmp can contain any number of shape name columns, which are identified with a column property. Multiple name columns support localizations and alternate names styles (such as abbreviations), but a given graph usage uses only one column of names. The first column of the Name file contains unique Shape ID numbers in ascending order. JMP creates this columns for you. The column values match those in the -XY.jmp map data table Shape ID column.

Note: The Shape ID column can also be named Shape.

Figure 12.5 Example of US-State-Name.jmp

	Shape ID	Name	State FIPS	USPS Code	Abbr	Max Grow ...	Max Grow X Offset	Max Grow Y Offset
1	1	Alabama	01	AL	Ala.	1	0	0
2	2	Alaska	02	AK	Alaska	1	0	0
3	3	Arizona	04	AZ	Ariz.	1	0	0
4	4	Arkansas	05	AR	Ark.	1	0	0
5	5	California	06	CA	Calif.	3	-1.7	-3
6	6	Colorado	08	CO	Colo.	1	0	0
7	7	Connecticut	09	CT	Conn.	4	0.8	-1.1
8	8	Delaware	10	DE	Del.	3	0.5	-0.3
9	9	District of Columbia|Washi...	11	DC	D.C.	70	2	-2.5
10	10	Florida	12	FL	Fla.	4	2	-2
11	11	Georgia	13	GA	Ga.	1	0	0

Map XY Files

Each xxx-XY.jmp file has four columns. Each row is a coordinate in some shape. Each shape is made of one or more parts. Each shape is a closed polygon. The first column is the same Shape ID as in the xxx-Name file. The second column is the Part ID. The next two columns are X and Y.

Figure 12.6 Example of US-State-XY.jmp

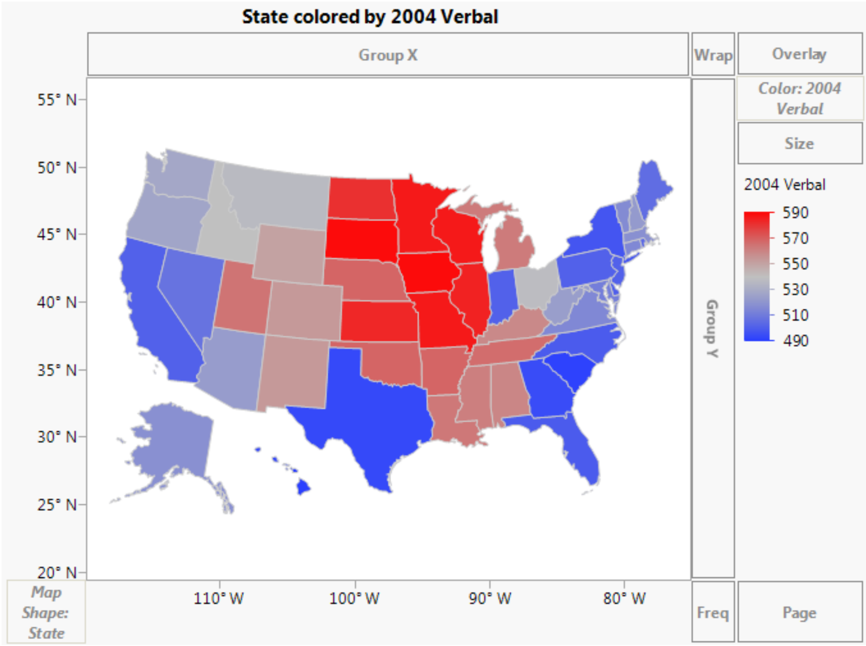
	Shape ID	Part ID	X	Y
1	1	1	85.6052° W	34.9847° N
2	1	1	85.1844° W	32.8613° N
3	1	1	84.9630° W	32.4242° N
4	1	1	85.0071° W	32.3284° N
5	1	1	84.8909° W	32.2615° N
6	1	1	85.0611° W	32.1341° N
7	1	1	85.1407° W	31.8575° N
8	1	1	85.0413° W	31.5410° N
9	1	1	85.1075° W	31.1865° N
10	1	1	85.0024° W	31.0007° N
11	1	1	87.5989° W	30.9975° N

Color

The Graph Builder platform lets you adds color to create choropleth maps. A choropleth map shows statistical differences in a geographic area while maintaining the proportion of the statistical variable.

Drag a column containing geographic place-names, like countries, regions, states, or provinces, into the **Map Shape** zone and create a map. Then drag a column to the **Color** zone to color the map by that column. The categorical or continuous color theme selected in your Preferences is applied to each shape.

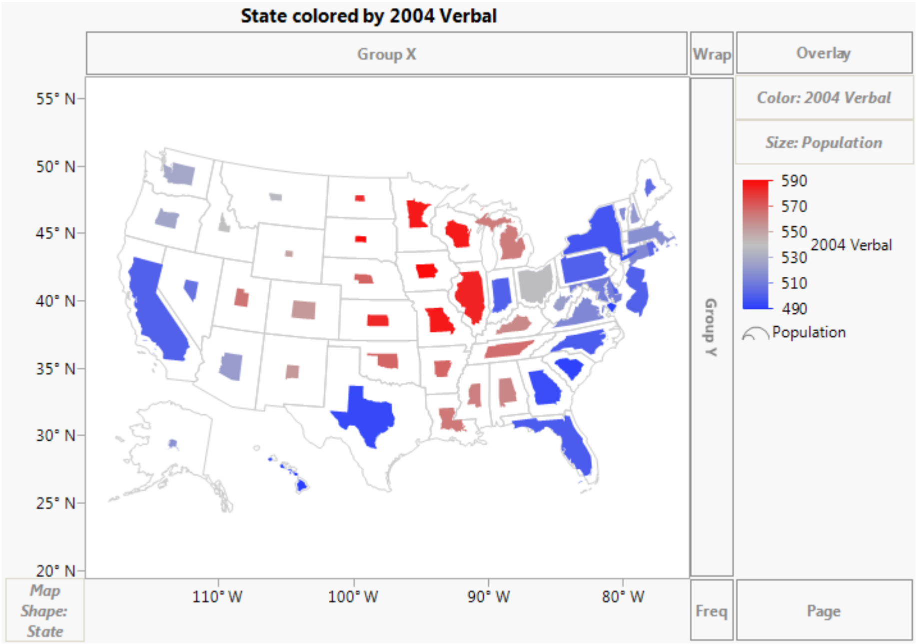
Figure 12.7 Example of SAT.jmp After Dragging 2004 Verbal to Color



Size

Use the **Size** element to scale map shapes according to the size variable, minimizing distortion.

Figure 12.8 Example of SAT.jmp After Dragging Population to Size



Customizing Graphs

To change colors and transparency for a map, right-click the color bar in the legend. The right-click options vary, depending on whether the color variable is continuous or categorical (nominal or ordinal). However, for both types of variables, you can change the transparency.

To change the transparency of a graph:

1. Right-click the color of the variable level on the color bar that you want to change and select **Transparency**.
2. Specify the transparency between 0 (clear) and 1 (opaque).
3. Click **OK**.

You can also change the transparency of images (for example, Simple Earth and Detailed Earth). To set the transparency, right-click over the graph and select **Customize...**. This brings up the Customize Graph window, where you can select the Background Map and assign a value for transparency. A valid value for transparency goes from 0.0 (completely transparent) to 1.0 (completely opaque). Within Graph Builder, you can also right-click over the graph and select **Graph > Transparency**.

Categorical (nominal or ordinal) variables use a singular coloring system, where each level of the variable is colored differently.

To change the color of one of the variable levels:

1. Right-click the color of the variable level that you want to change and select **Fill Color**.
2. Select the new color.

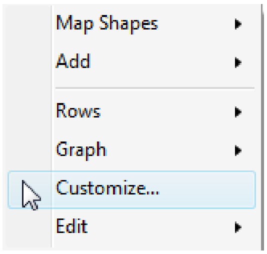
Continuous variables use a color gradient.

To change the color theme:

1. Right-click the color bar and select **Gradient**.
2. In the Gradient Settings window, select a different **Color Theme**.

Graphs consist of markers, lines, text, and other graphical elements that you can customize. If you right-click an image, there are several options for working with the graph. The options differ based on what you clicked. See "Gradient Settings" on page 111 in the "Graph Builder" chapter and the JMP Reports chapter in *Using JMP*. Below are a few options.

Figure 12.9 Right-click Menu for Graphics



- Map Shapes:
  – Change To - Provides options for changing the map into another graph type.
  – Summary Statistics - Provides options for changing the statistic being plotted.
  – Show Missing Shapes - Shows or hides missing data from a map (turned off by default). Missing Shape means that there are some shape names that exist in the map file but not in the data table for analysis.
  – Remove - Removes the current map shape.
- Customize - You can change the properties of the graph such as contents, grid lines, or reference lines. The graphical elements that you can customize differ for each graph. Select **Background Map** to change the transparency of a background map or **Map Shape** to change the line color, line style and width, fill color, missing shape fill or missing value fill. Click **Help** in the Customize Graph window for a more detailed explanation of the customize options.

Custom Map Files

You can create your own map files by following the same pattern as the built-in files. To add your own map files, you need two things: a series of XY coordinates for the vertices of the polygons that describe the shape, and a set of names for each polygon. Data and shape attributes are required to map custom shapes so that you can add your own shapes to JMP. There are two common sources for data like this: Esri shapefiles and SAS/GRAPH map data sets.

In order for JMP to automatically find your files, place them in the following directory:

- On Windows: C:/Users/<user name>/AppData/Roaming/SAS/JMP/Maps
- On macOS: /Users/<user name>/Library/Application Support/JMP/Maps

Note: On Windows, in JMP Pro, the "JMP" folder is named "JMPPro". Or, you can link the map files to your data files explicitly with the **Map Role** column property.

Note the following when creating map files:

- Each set of map files that you create must contain a -Name file and a -XY file.
- The first column in both files must be the ascending, numeric Shape ID variable. The -Name file can contain any other columns. The shapes are built by rows. The XY coordinates have to go around the shape rather than just define the convex hull of the shape.
- For the **Map Role** column property, columns that are marked with the **Shape Name Definition** are searched for shape identification and must contain unique values.
- If you import an Esri SHP file, it is opened in the correct format. -Name files commonly have a .dbf extension. See "Esri® Shapefiles" on page 276.
- SAS/GRAPH software includes a number of map data sets that can be used with JMP. See "SAS/GRAPH® Map Data Sets" on page 277.

You might want to create choropleth maps of other non-geographic regions (for example, a floor of an office building). Simply, add the two shape files for your non-geographic space. If you do not have XY coordinates, but you do have a graphic image of the space, you can use the Custom Map Creator add-in for JMP. With this add-in, you can trace the outlines of the space and JMP creates the -XY and -Name files for you. You can download this add-in from the JMP File Exchange page.

Map Role

You can specify the attributes and properties of a column in a data table within the Column Info window in Column Properties. The **Map Role** property is set for a column like other column properties in the Column Info window.

If you have created your own data table that contains boundary data (such as countries, regions, states, provinces, or counties) and you want to see a corresponding map in Graph Builder, use the **Map Role** property within Column Properties. Each pair of map files that you create must contain a -Name file and a -XY file.

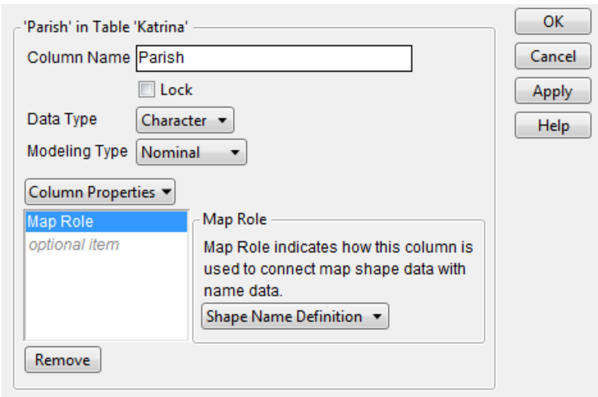
Note the following:

- If the custom boundary files reside in the default custom maps directory, then you need to specify only the **Map Role** property in the -Name file.
- If the custom boundary files reside in an alternate location, specify the **Map Role** property in the -Name file and in the data table that you are analyzing.
- The columns that contain the **Map Role** property must contain the same boundary names, but the column names can be different.

To add the Map Role property into the -Name data table:

1. Right-click the column containing the boundaries and select **Column Properties > Map Role**.
2. Select **Shape Name Definition** below **Map Role**.
3. Click **OK**.
4. Save the data table.

Figure 12.10 Shape Name Definition Example



To add the Map Role property into the data table that you are analyzing:

Note: Perform these steps only if your custom boundary files do not reside in the default custom maps directory.

---

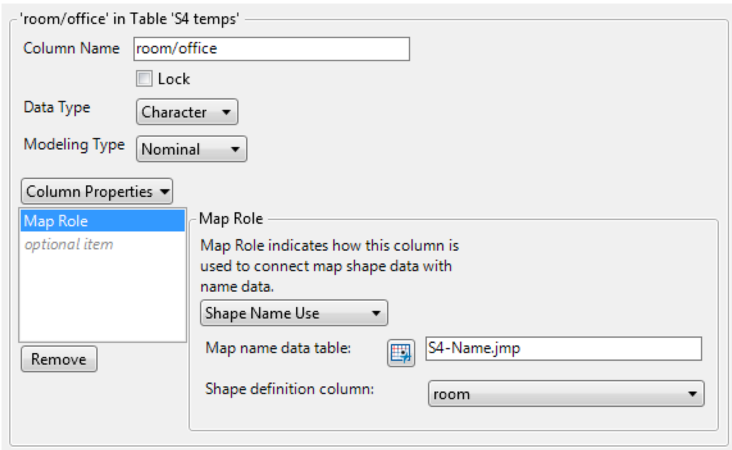
1. Right-click the column containing the boundaries and select **Column Properties > Map Role**.

2. Select **Shape Name Use** below **Map Role**.
3. Next to **Map name data table**, click to browse to a **-Name** map data table. You can enter the relative or absolute path.

   If the map data table is in the same folder, enter only the filename. Quotation marks are not required when the path contains spaces.
4. From the **Shape definition column** list, select the column in the map data table whose values match those in the selected column.

   Figure 12.11 shows an example of the **room/office** column in the **S4 Temps.jmp** sample data table.

Figure 12.11 Shape Definition Column Example



5. Click **OK.**
6. Save the data table.

When you generate a graph in Graph Builder and assign the modified column to the **Map Shape** zone, your boundaries appear on the graph.

For numeric columns, the Format Menu appears in the Column Info window. Specify the format to tell JMP how to display numbers in the column. Latitude and Longitude for geographic maps are located under **Format > Geographic** when customizing axes and axes labels.

Geographic Shows latitude and longitude number formatting for geographic maps. Latitude and longitude options include the following:

– DDD (degrees)

– DMM (degrees and minutes)

– DMS (degrees, minutes, and seconds)

In each format, the last field can have a fraction part. You can specify the direction with either a signed degree field or a direction suffix. To show a signed degree field, such as *-59°00′00″*, deselect **Direction Indicator**. To show the direction suffix, such as *59°00′00″ S*, select **Direction Indicator**.

To use spaces as field separators, deselect **Field Punctuation**. To use degrees, minutes, and seconds symbols, select **Field Punctuation**.

Esri® Shapefiles

The Esri shapefile is a vector data format that contains data about geographic features such as terrain and oceans. It is developed and regulated by Esri as a specification for geographic mapping software.

Each shapefile is a set of files with the same name and different extensions.

main file (.shp)

The .shp file contains sequences of points that make up polygons. When opened with JMP, a .shp file is imported as a JMP table.

- The Shape column is added during import to uniquely identify each geographic region. Each coordinate point is in a separate row.
- The Part column to indicate discontiguous regions, and the XY coordinates (in latitude and longitude degrees).

JMP supports two-dimensional .shp files (no elevation information).

dBase Table (.dbf)

You add a Shape ID column to the .dbf table, which maps to the Shape column in the .shp file. Add any number of columns that provide common names or values to refer to specific regions.

To convert an Esri shapefile to a JMP map file:

1. Open the .shp file in JMP.
2. Make sure that the Shape column is the first column in the .shp file. Add formatting and axis settings for the X and Y columns (optional). Graph Builder uses those settings for the X and Y axes.
3. Save the .shp file as a JMP data table to the Maps folder with a name that ends in -XY.jmp.
4. Open the .dbf file.

5. Add a Shape ID column as the first column in the table. This column should be the row numbers from 1 to n, the number of rows in the data table.

Note: You can use **Cols > New Columns > Initialize Data > Sequence Data**) to fill the column with sequential numbers.

6. Assign the **Map Role** column property to any column that you use for place names in the Shape role of Graph Builder. To do this, right-click at the top of the column and select **Column Properties > Map Role**.
7. Select **Shape Name Definition** from the drop-down box in the property definition.
8. Save the table as a JMP data table with a name that matches the earlier table and that ends in -Name.jmp.

JMP looks for these files in two locations. One location is shared by all users on a machine. This location is:

- Windows: C:/Program Files/SAS/JMP/15/Maps/
- macOS: /Library/Application Support/JMP/15/Maps

The other location is specific for an individual user:

- On Windows: C:/Users/<user name>/AppData/Roaming/SAS/JMP/Maps
- On macOS: /Users/<user name>/Library/Application Support/JMP/Maps

Note: On Windows, in JMP Pro, the “JMP” folder is named “JMPPro”.

SAS/GRAPH® Map Data Sets

SAS/GRAPH software includes a number of map data sets that can be converted for use with JMP. The data sets are in the Maps library. The traditional map data sets contain the XY coordinate data and the feature table contains the common place names. You need to convert both of these files to JMP data tables for use with JMP.

Most of the traditional map data sets have unprojected latitude and longitude variables in radians. The data sets can be used with JMP once they have been converted to degrees and the longitude variable has been adjusted for projection. The following is a DATA step that shows the conversion process for the Belize data set.

```
data WORK.BELIZE;
keep id segment x y;
rename segment=Part;
set maps.belize;
if x NE .;
if y NE .;
y=lat*(180/constant('pi'));
```

```
x=-long*(180/constant('pi'));
run;
```

You can now import the converted file and save it as Belize-XY.jmp.

The next step is to import the matching feature data set (in this case: MAPS.BELIZE2). After importing the feature data set, move the ID column to the first position in the data table. Then assign the Map Role column property to the columns that you use for place names in the Shape role of Graph Builder. To do this, right-click the top of the column and select **Column Properties** > **Map Role**. Then select **Shape Name Definition** from the drop-down box in the property definition. For MAPS.BELIZE2, use the IDNAME column. Save the feature data table as Belize-Name.jmp.

To convert SAS maps, download the SAS to JMP Map Converter add-in from the JMP File Exchange page. For each map, the add-in reads the data from the two SAS map tables, rearranges and formats the data and then places it into the two JMP map tables.

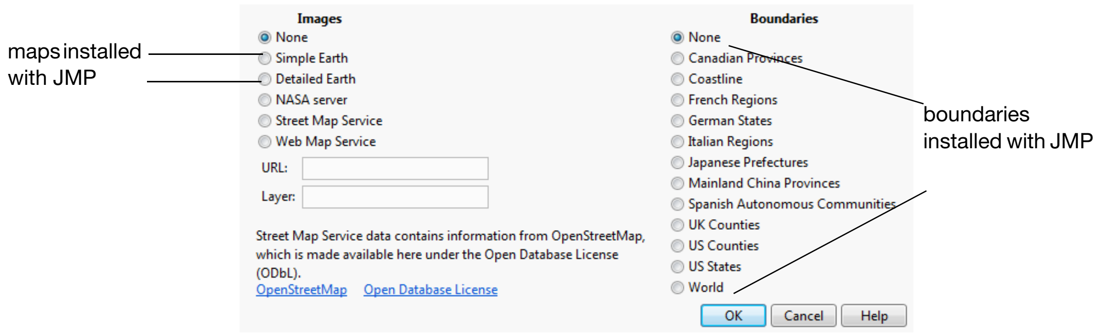
Background Maps

Adding map images and boundaries to graphs provides visual context to geospatial data. Affixing a background map generates an appealing map, providing your data a geographic context and giving you a whole new way to view your data. For example, you can add a map to a graph that displays an image of the U.S. Another option is displaying the boundaries for each state (when data includes the latitudes and longitudes for the U.S.). There are different types of background maps. Some maps are built into JMP and are delivered as part of the JMP install. Other maps are retrieved from an Internet source, and still other maps are user-defined.

The data should have latitudinal and longitudinal coordinates. Otherwise, the map has no meaning in the context of the data. The X and Y axes also have range requirements based on the type of map. These requirements are described in the following sections. Simply plot longitude and latitude on the X and Y axes, and then right-click within the graph and select **Graph** > **Background Map**.

The Background Map window shows two columns of choices: Images and Boundaries. On the left of the window that you can select from two built-in map images, or you can connect to a Web Map Service to retrieve a background image. On the right side of the window, you can select political boundaries for a number of regions.

Figure 12.12 Example of Background Map Options



The following Background Map options are available:

Images

None Removes the background map that you selected in the Images column.

Simple Earth Shows a map of basic terrain. For examples, run the associated scripts in these sample data tables: SAS Offices.jmp and Aircraft Incidents.jmp.

Detailed Earth Shows a high-resolution map with detailed terrain.

NASA Server Shows a map from the NASA server. Requires an Internet connection.

Street Map Service Shows a map with an appropriate amount of detail based on the display's zoom level. This enables you to zoom down to the street level. For example, run the associated script in the sample data table SAS Offices.jmp.

Web Map Service Shows a map from the Uniform Resource Locator (URL) and the layer that you specify. Requires an Internet connection.

Boundaries

None Removes the boundaries that you selected in the Boundaries column.

Boundaries for various regions Shows borders for the map regions, such as Canadian provinces, U.S. counties, U.S. States, and world countries. The list varies based on your location. The maps that you created from Esri shapefiles are also listed here.

Two tools are especially helpful when you are viewing a map:

- The grabber tool () lets you scroll horizontally and vertically through a map.
- The magnifier tool () lets you zoom in and out.

Images in Maps

Every flat map misrepresents the surface of the Earth in some way. Maps cannot match a globe in truly representing the surface of the entire Earth. A map projection is used to portray all or part of the round Earth on a flat surface. This cannot be done without some distortion. Every projection has its own set of advantages and disadvantages. A map can show one or more, but not all, of the following: true direction, distance, area, or shape. JMP uses a couple of projections (Albers Equal Area Conic and Kavrayskiy VII) for its maps. Within Images, you can select from two built-in map images, or you can connect to a Web Map Service to retrieve a background image.

Earth Images Installed with JMP

JMP provides two levels of earth imagery; simple and detailed. Both maps show features such as bodies of water and terrain. However, detailed maps show more precise terrain. And with detailed maps, you can zoom in farther, and the map features remain clear. Image maps are raster images. The maps wrap horizontally, so you continue to see map details as you scroll from left to right. The maps do not wrap vertically. Beyond the -90 and 90 y-axis range, a plain background appears instead of the map.

Figure 12.13 Examples of Simple and Detailed Maps

Simple Earth Image   Detailed Earth Image

As its name suggests, Simple Earth is a relatively unadorned image of the earth's geography. It does not show clouds or arctic ice, and it uses a green and brown color scheme for the land and a constant deep blue for water. Detailed Earth has a softer color scheme than Simple Earth, lighter greens and browns for the land, as well as variation in the blue for the water. Detailed Earth also has a slightly higher resolution than Simple Earth. The higher resolution lets you zoom into a graph further with Detailed Earth than with Simple Earth before the quality of the background image begins to blur.

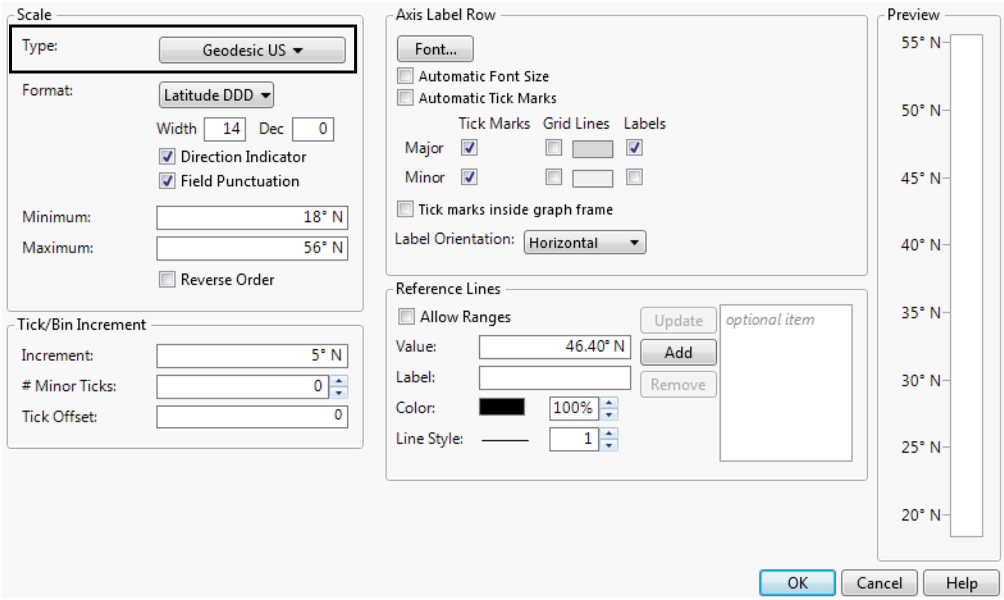
Another feature of Simple Earth and Detailed Earth is the ability to wrap. The Earth is round, and when you cross 180° longitude, the Earth does not end. The longitudinal value continues from -180° and increases. The map wraps continuously in the horizontal direction, much as the Earth does. The background map does not wrap in the vertical direction.

Simple Earth and Detailed Earth both support a geodesic scaling. In Figure 12.13 on page 280, the Earth appears as a rectangle, where the width is twice as wide as the height. If we were to take this rectangle and roll it up, we would have a cylinder. In reality, we know that the Earth does not form a cylinder, but rather a sphere. You can use a geodesic scaling, which transforms the map to a more realistic representation of the Earth. To use the geodesic scaling, change the type of scale on the axes.

To change the axes scale:

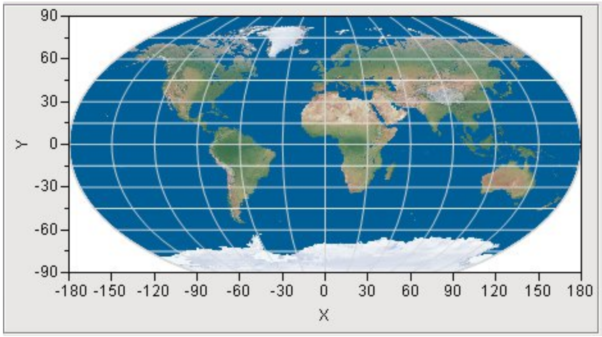
1. Right-click the X or Y axis and then select **Axis Settings.**
2. Change the Scale Type to **Geodesic** or **Geodesic US**.

Figure 12.14 Y Axis Setting Window



Both choices transform the map to a geodesic scaling. Use Geodesic US if you are viewing a map of the continental US and you want Alaska and Hawaii to be included in the map. It is important to note that you must set the scale to geodesic for both axes to get the transformation. You will not see a change in the map after setting only one of the axes. In the following figure, Simple Earth is used as the background map with the axes set to use a geodesic scale. The axes lines are turned on as well. Notice the longitudinal lines are now curved, instead of straight.

Figure 12.15 Simple Earth Background Map - Axes Set to Geodesic Scale



Since Detailed and Simple Earth are built into JMP, these options work anytime, without a network connection. However, these images might not be all that you want, or they might not be detailed at the resolution that you need. If this is the case, and if you have an Internet connection, you can connect to a Web Map Service to retrieve a map image that meets your needs.

Maps from the Internet

The National Aeronautics and Space Administration (NASA) and other organizations provide map image data using a protocol called Web Map Service (WMS). These maps have the advantage of showing the most up-to-date geographical information. However, the display of the maps can be slow depending on the response time of the server, and the sites can change or disappear at any time. An Internet connection is required to access the information.

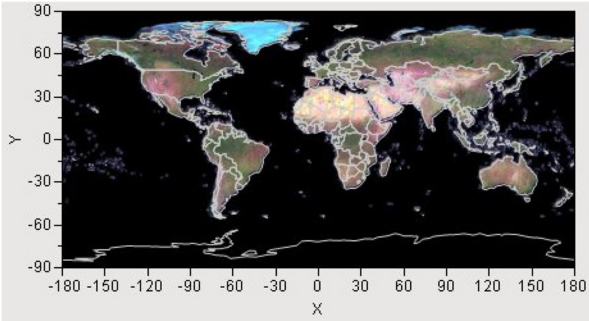
Figure 12.16 Examples of NASA and WMS Maps

NASA image

WMS image from the MetaCarta™ Server

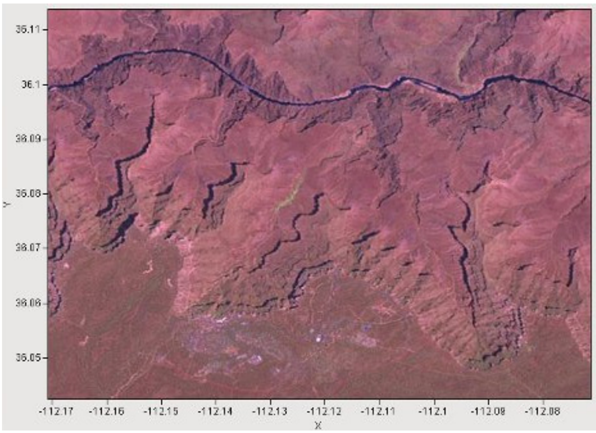
The NASA server provides maps for the entire Earth. The following figure displays the Earth using the NASA server as its source for the background map. The boundary map shows the outlines of the countries.

Figure 12.17 NASA Server Map Example



Not only does this server cover the entire Earth, but you can also zoom in on a much smaller area of the Earth and still get a reasonable map. The following figure displays the Colorado River running through the Grand Canyon in Arizona. The Grand Canyon Village is visible in the bottom of the map.

Figure 12.18 NASA Server Map Example - Zoom In on Colorado



If you look at the axes values, you can see that the area is less than 1/10° by 1/10°. The Simple Earth and Detailed Earth background maps do not display that type of resolution. The NASA server provides a fairly detailed view of any land mass on Earth. Water, however, is simply filled in as black. The NASA server is free to access, but it is also limited in availability. If the server is temporarily unavailable or becomes overloaded with requests, it delivers an error message instead of the requested map.

Another Internet-based option for background maps is a Web Map Service (WMS). The WMS option enables you to specify any server that supports the WMS interface. The NASA server is an example of a WMS server, but we have provided the URL and a layer name for you. With the WMS option, you must know the URL to the WMS server and a layer name supported by the server. Most WMS servers support multiple layers. For example, one layer can show terrain, another layer can show roads, and still another layer can include water, such as rivers and lakes. By specifying the URL for the server and the layer, JMP can make a request to the server and then display the map that is returned.

Unlike with simple and detailed maps, WMS maps do not wrap. You can scroll horizontally and vertically. However, beyond the -180 to 180 (x axis) and -90 to 90 (y axis) ranges, a plain background appears instead of the map. The limits of the axes are used to define the limits of the map that is displayed.

In order to use the WMS option for a background map, you need to decide which WMS server to use. There are many WMS servers freely available from the Internet. Most of them provide maps only for a particular area of the world, and each of them supports their own layers. So you have to search for the appropriate WMS server for your particular situation.

You can search for WMS servers on the Internet using your favorite search engine. Once you find one, you need to discover the layers that it supports. For this, you can use the WMS Explorer add-in. The WMS Explorer add-in generates a list of all the layers available on a server. You can select a layer from the list to see what it looks like. You can download the WMS Explorer add-in from the JMP File Exchange page.

Note: To use the WMS Explorer add-in and the WMS background map capabilities of JMP, your computer must be connected to the Internet.

To locate a server, launch the add-in through the menu items **Add-Ins > Map Images > WMS Explorer**. The add-in presents a text box for entering the url of a known WMS server. Alternatively, you can make a selection from a drop-down list of pre-discovered WMS servers (the list can be out of date). After specifying a WMS server, select **Get Layers**. Using Get Layers is not necessary if selecting from the drop-down list or if clicking **Enter** after entering a URL. This sends a request to the WMS server for a list of layers that the server supports. The returned list appears in the list box on the left, labeled **Layers**. A map of the world appears as an outline in the graph to the right. Selecting a layer makes a request to the WMS server to return a map, using the specified layer, that represents the entire earth. Selecting a different layer generates a different map.

The default maps do not cover the entire earth (for example, some WMS servers might provide mapping data for a particular county, within a state). In that case, it is likely that selecting a layer does not generate any visible map. You might have to zoom in on the appropriate area before any image map is visible. The standard JMP toolbar is available in the add-in window and the zoom tool works just like it does in any JMP window.

The graph is a typical graph in JMP, which means that all the regular JMP controls are available to you. You can adjust the axes or use the zoom tool (found on the hidden menu bar) just as you would in JMP. You can also right-mouse-click to select **Size/Scale > Size to Isometric** to return the graph to a proper aspect ratio. You can also select Background Map, where you can adjust the boundary map.

Once a desirable map is determined, note the URL in the text box at the top and the selected layer in the Layers list. This is the information that you need to enter in the background map window when WMS is selected as the type of image background map.

Because requests are being made to a server across the Internet, there are a number of conditions that can generate an error. WMS servers often have limited availability and sometimes are not available at all. Occasionally a WMS server might return a name of a layer that it no longer supports. In these types of cases (and others), a server usually returns an error message in lieu of a map. If that happens, the error message is displayed below the Layers list in an area labeled **Errors**.

Boundaries

JMP can display boundaries (such as U.S. states or French region boundaries). These boundaries draw an outline around a defined area and can be displayed alone on a graph or combined with image data. Several boundaries are installed with JMP. Alternatively, you can create your own boundaries from Esri shapefiles or from scratch. Because of this, the list of Boundaries that you see in the Set Background Map window can be different.

When you add shape files to the built-in locations in JMP, they are available for the Graph Builder platform and for the Boundaries option in the Background Map window. In this way, you can add more political boundaries for use with background maps. Boundary-style maps are vector-based shapes.

Figure 12.19 Example of U.S. State Boundaries

Add a Background Map and Boundaries

To add a background map and boundaries:

1. Right-click a blank area on the graph and select **Background Map** (or select **Graph > Background Map** in Graph Builder).

   The Set Background Map window appears (Figure 12.12).
2. To display a background, do one of the following:
   – Select **Simple Earth**, **Detailed Earth**, **NASA server**, or **Street Map Service** in the Images column.
   – Select **Web Map Service** and paste a WMS URL next to URL. Enter the layer identifier next to **Layer**.
3. To display geographic borders on the map, select an option in the Boundaries column (If you installed your own boundary shapefiles, they are also listed in this column).
4. Click **OK**.

   If the NASA map, Street Map, or WMS map does not appear after you add it, the map server might not be available. View the error log to verify the problem.

Examples of Creating Maps

• "Louisiana Parishes Example"
• "Hurricane Tracking Examples"
• "Office Temperature Study"
• "Boundary Map with Clipped Elements"

Louisiana Parishes Example

In this example you work with custom map files and then create custom maps in two different ways:

• Set up custom map files initially and save them in the predetermined location. JMP finds and uses them in the future with any appropriate data.
• Point to specific predefined map files directly from your data. This step might be required each time you want to specify custom maps.

Set Up Automatic Custom Maps

Suppose that you have downloaded Esri shapefiles from the Internet and you want to use them as your map files in JMP. The shapefiles are named Parishes.shp and Parishes.dbf. These files contain coordinates and information about the parishes (or counties) of Louisiana.

Note: Pathnames in this section refer to the “JMP” folder. On Windows, in JMP Pro, the “JMP” folder is named “JMPPro”.

Save the .shp File

Save the .shp file with the appropriate name and in the correct directory, as follows:

1. In JMP, open the Parishes.shp file from the following default location:
   – On Windows: C:/Program Files/SAS/JMP/15/Samples/Import Data
   – On macOS: /Library/Application Support/JMP/15/Samples/Import Data

   Note: If you cannot see the file, you might need to change the file type to **All Files**.

   JMP opens the file as Parishes. The .shp file contains the *x* and *y* coordinates.
2. Save the Parishes file with the following name and extension: Parishes-XY.jmp. Save the file here:
   – On Windows: C:/Users/<user name>/AppData/Roaming/SAS/JMP/Maps
   – On macOS: /Users/<user name>/Library/Application Support/JMP/Maps
3. Close the Parishes-XY.jmp file.

Save the .dbf File

Perform the initial setup and save the .dbf file, as follows:

1. Open the Parishes.dbf file from the following default location:
   – On Windows: C:/Program Files/SAS/JMP/15/Samples/Import Data
   – On macOS: /Library/Application Support/JMP/15/Samples/Import Data

   Note: If you cannot see the file, you might need to change the file type to **All Files**.

   JMP opens the file as Parishes. The .dbf file contains identifying information.
2. In the Parishes file, add a new column. Name it Shape ID. Drag and drop it to be the first column.
3. In the first three rows of the Shape ID column, type 1, 2, and 3 (Note - You can also use **Cols > New Columns > Initialize Data > Sequence Data**).
4. Select all three cells, right-click, and select **Fill > Continue sequence to end of table**.

Figure 12.20 Shape ID Column in Parishes File

	Shape ID	Shape	PARISH	FIPS	PARISH_FIP
1	1	1	Webster	22119.000000000...	119
2	2	2	East Carroll	22035.000000000...	035
3	3	3	West Carroll	22123.000000000...	123
4	4	4	Morehouse	22067.000000000...	067
5	5	5	Union	22111.000000000...	111
6	6	6	Lincoln	22061.000000000...	061
7	7	7	Ouachita	22073.000000000...	073
8	8	8	Richland	22083.000000000...	083
9	9	9	Bienville	22013.000000000...	013
10	10	10	Franklin	22041.000000000...	041

5. Right-click the PARISH column and select **Column Info**.
6. Select **Column Properties > Map Role**.
7. Select **Shape Name Definition**.
8. Click **OK**.
9. Save the Parishes file with the following name and extension: Parishes-Name.jmp. Save the file here:
   – On Windows: C:/Users/<user name>/AppData/Roaming/SAS/JMP/Maps
   – On macOS: /Users/<user name>/Library/Application Support/JMP/Maps
10. Close the Parishes-Name.jmp file.

Create the Map in Graph Builder

Once the map files have been set up, you can use them. The Katrina.jmp data table contains data on Hurricane Katrina's impact by parish. You want to visually see how the population of the parishes changed after Hurricane Katrina. Proceed as follows:

1. Select **Help > Sample Data Library** and open Katrina.jmp.
2. Right-click the Parish column and select **Column Properties > Map Role**.
3. Select **Shape Name Use**.
4. Click the **Map name data table** button and browse to select Parishes-Name.jmp, which you previously created.

   This tells JMP where the data tables containing the map information reside.
5. Select PARISH from the **Shape definition column** list.

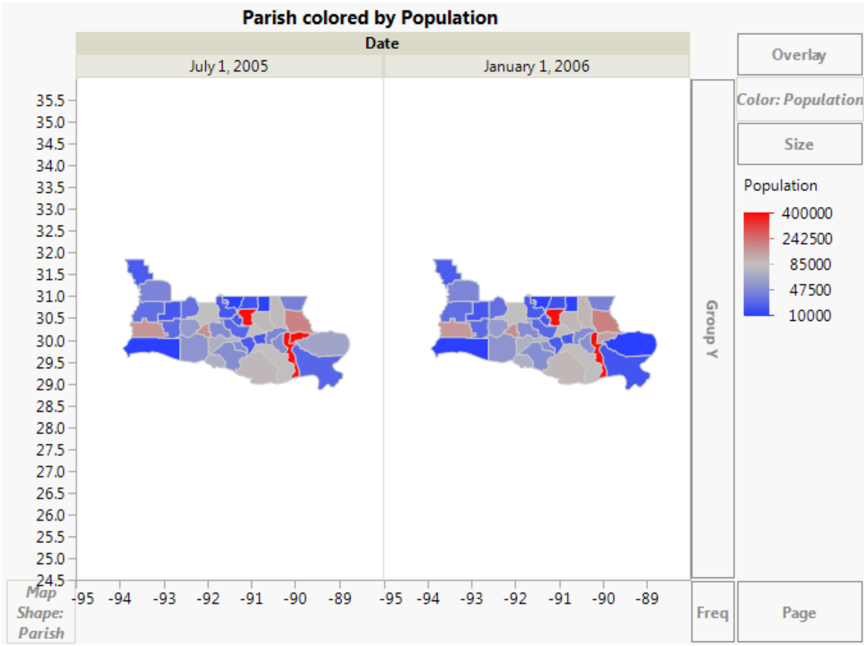
   In Parishes-Name.jmp, the PARISH column has the **Shape Name Definition Map Role** property assigned. The column consists of map shape data for each parish.
6. Select **Graph > Graph Builder**.

7. Drag and drop Parish into the **Map Shape** zone.

   The map appears automatically, since you defined the Parish column using the custom map files.

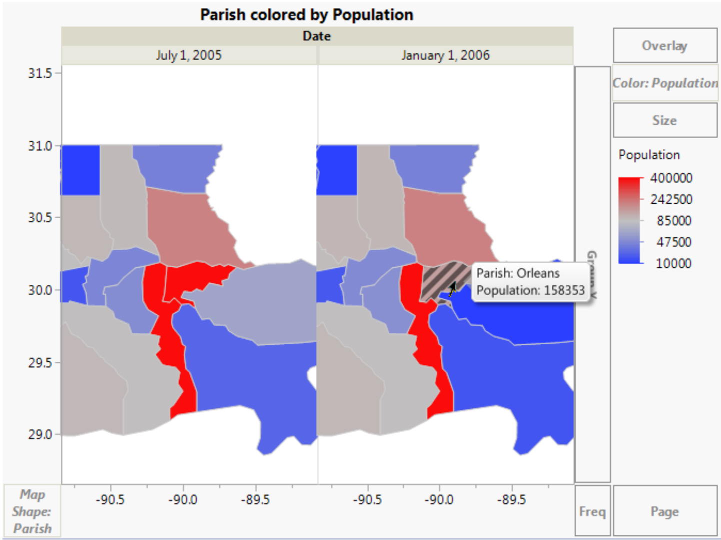
8. Drag and drop Population into the **Color** zone.
9. Drag and drop Date into the **Group X** zone.

Figure 12.21 Population of Parishes Before and After Katrina



10. Select the **Magnifier** tool to zoom in on the Orleans parish in both maps (Figure 12.22)

Figure 12.22 Orleans Parish



You can clearly see the drop in population as a result of Hurricane Katrina. The population of the Orleans parish went from 437,186 in July 2005 to 158,353 in January 2006.

Point to Existing Map Files Directly from Your Data

Suppose that you already have your custom map files and they are named appropriately. Your map files are US-MSA-Name.jmp and US-MSA-XY.jmp. They are saved in the sample data folder.

The PopulationByMSA.jmp data table contains population data from the years 2000 and 2010 for the metropolitan statistical areas (MSAs) of the United States. This example shows how the data table has been set up to create a map.

Add the Map Role Column Property

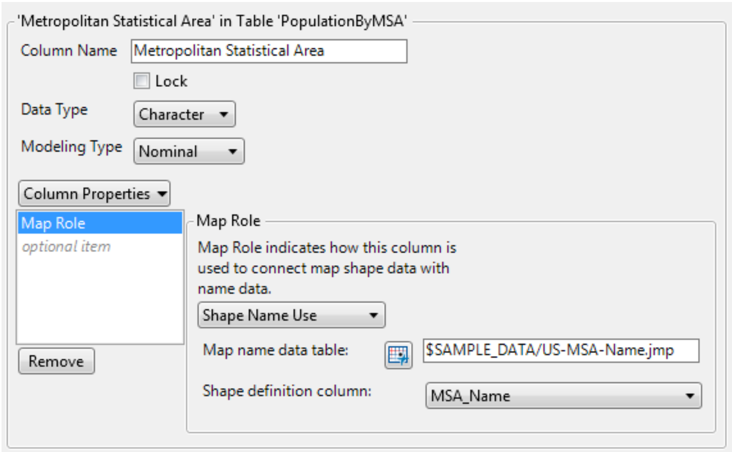
1. Select **Help > Sample Data Library** and open PopulationByMSA.jmp.
2. Right-click the **Metropolitan Statistical Area** column and select **Column Info**.
3. Select **Column Properties > Map Role**.
4. Select **Shape Name Use**.
5. Next to the **Map name data table**, type $SAMPLE_DATA/US-MSA-Name.jmp.

   This tells JMP where the data tables containing the map information reside.
6. Select MSA_Name from the **Shape definition column** list.

   MSA_Name is the specific column within the US-MSA-Name.jmp data table that contains the unique names for each metropolitan statistical area. Notice that the MSA_Name column has the **Shape Name Definition Map Role** property assigned, as part of correctly defining the map files.

Note: Remember, the Shape ID column in the -Name data table maps to the Shape ID column in the -XY data table. This means that indicating where the -Name data table resides links it to the -XY data table, so that JMP has everything that it needs to create the map.

Figure 12.23 Map Role Column Property



7. Click **OK**.

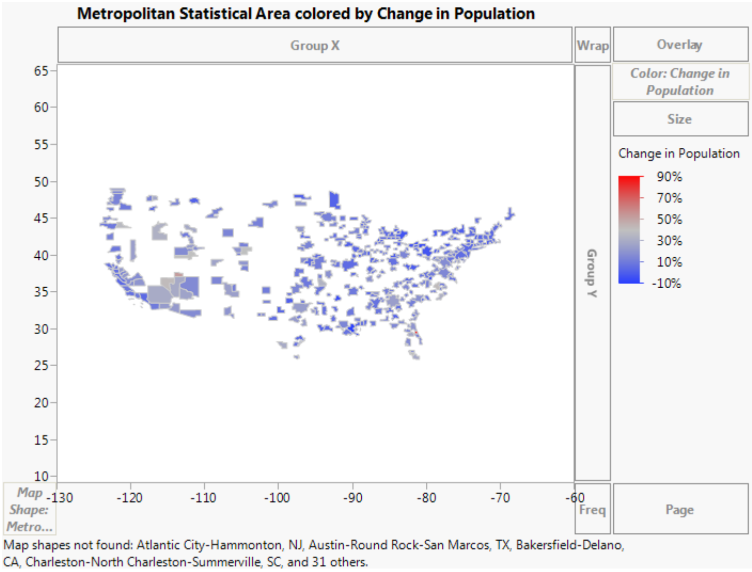
Create the Map in Graph Builder

Once the **Map Role** column property has been set up, you can perform your analysis. You want to visually see how the population has changed in the metropolitan statistical areas of the United States between the years 2000 and 2010.

1. Select **Graph > Graph Builder**.
2. Drag and drop **Metropolitan Statistical Area** into the **Map Shape** zone.

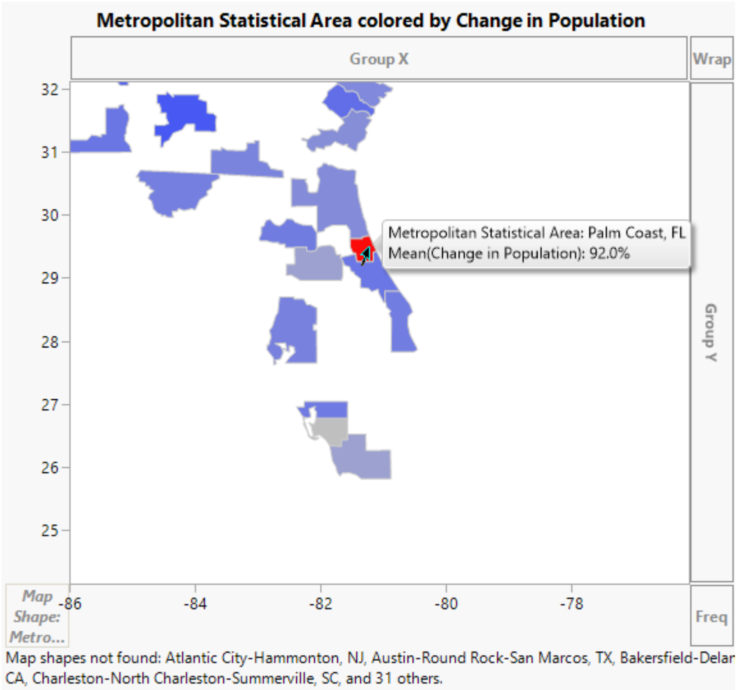
   Since you have defined the **Map Role** column property on this column, the map appears.
3. Drag and drop **Change in Population** to the **Color** zone.

Figure 12.24 Change in Population for Metropolitan Statistical Areas



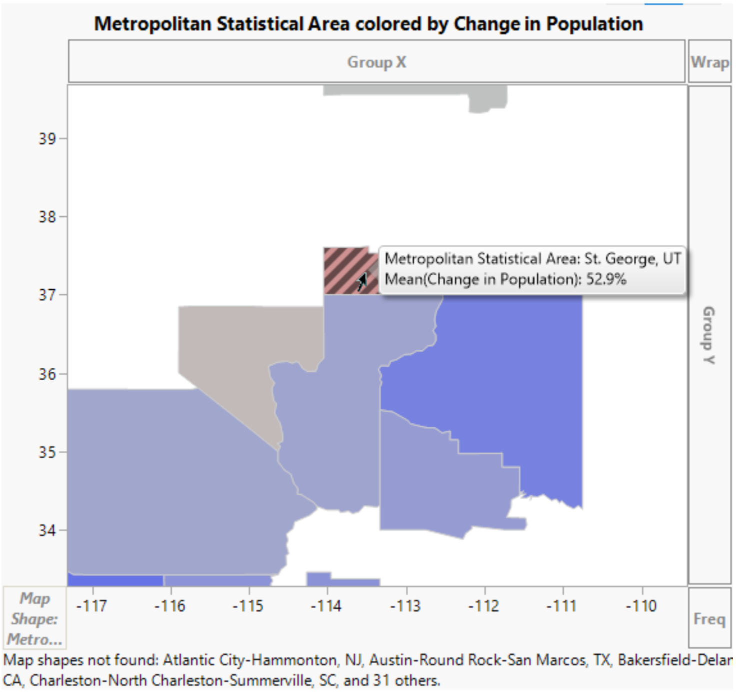
4. Select the **Magnifier** tool to zoom in on the state of Florida.
5. Select the **Arrow** tool and click the red area.

Figure 12.25 Population Change of Palm Coast, Florida



6. Select the **Magnifier** tool and hold down the Alt key while clicking on the map to zoom out.
7. Select the **Magnifier** tool and zoom in on the state of Utah.
8. Select the **Arrow** tool and click the area that is slightly red.

Figure 12.26 Population Change of St. George, Utah



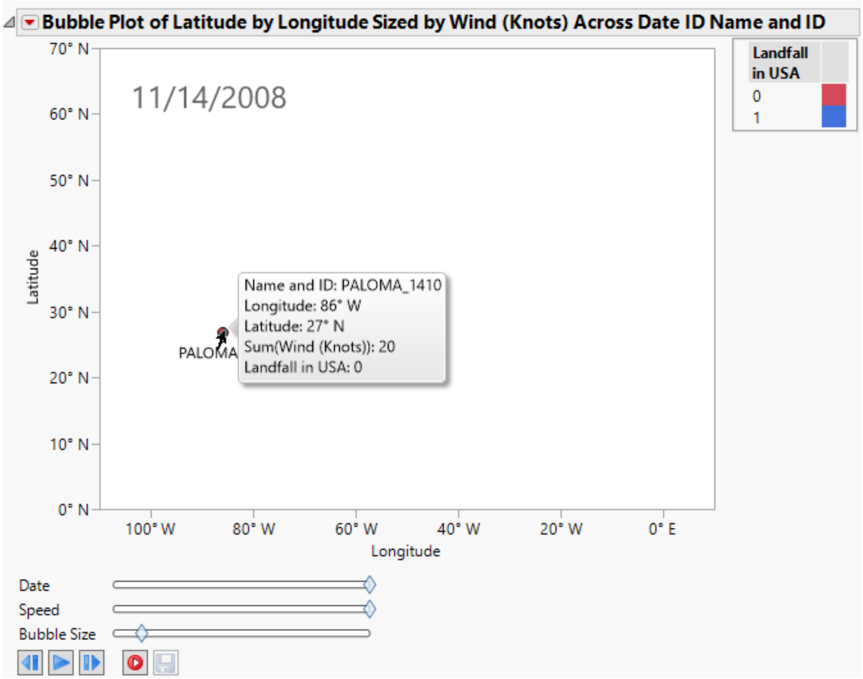
You can see that the areas of Palm Coast, Florida, and St. George, Utah had the most population change between 2000 and 2010. The Palm Coast area saw a population change of 92%, and the St. George area saw a population change of about 53%.

Hurricane Tracking Examples

This example uses the Hurricanes.jmp sample data table, which contains data on hurricanes that have affected the east coast of the United States. Adding a background map helps you see the areas the hurricanes affected. A script has been developed for this example and is part of the data table.

1. Select **Help > Sample Data Library** and open Hurricanes.jmp.
2. In the Table panel, click the green triangle next to the **Bubble Plot** script.
3. Drag the Date slider to the right as shown in Figure 12.27.
4. Click the red dot to display the name of the hurricane. The date appears in the upper left corner of the window. The red dot shows the location of Hurricane Paloma on November 14, 2008.

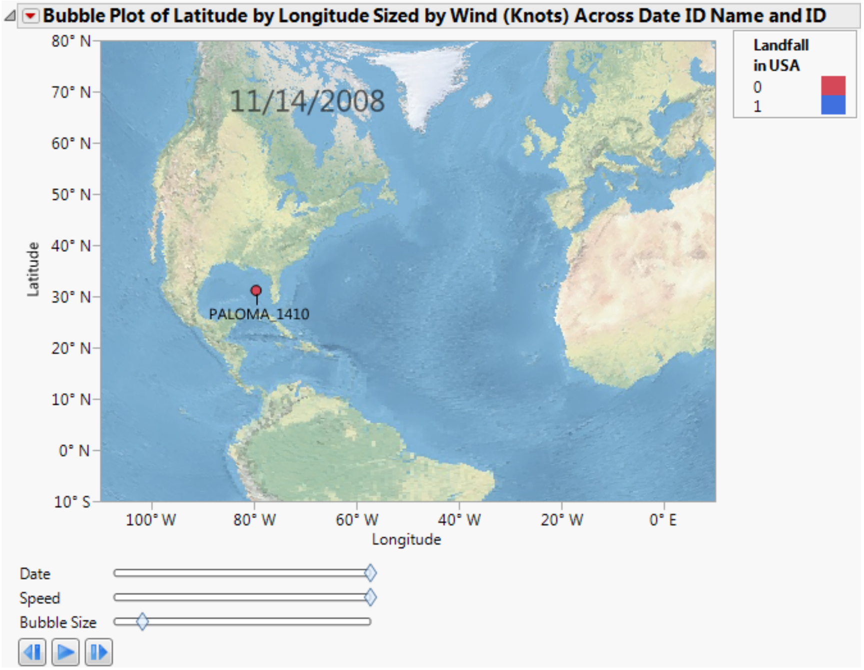
Figure 12.27 Bubble Plot of Hurricanes.jmp



Note that even though the location of the hurricane is plotted, it does not really tell us where it is. The axes information is there (27° North latitude and 86° West longitude), but we need a little more context. It is most likely over the middle of the Atlantic, but is it over a small island? This could make a big difference, especially for the inhabitants of the small island. Obviously, a map in the background of our graph would add a good deal of information.

5. Right-click the graph and select **Background Map**. The Set Background Map window appears (Figure 12.12).
6. Select **Detailed Earth** and click **OK**.

Figure 12.28 Bubble Plot of Hurricanes.jmp with Background Map

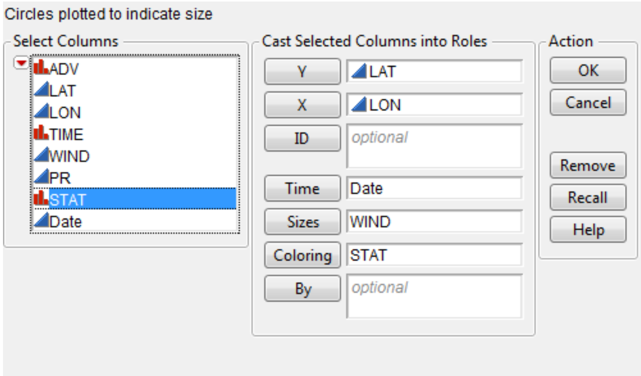


Now the coordinates make geographic sense. Click **Run** to view the animation of the hurricane data moving over the background map. Experiment with different options and view the displays. Adjust the axes or use the zoom tool to change what part of the world you are viewing. The map adjusts as the view does. You can also right-click the graph and select Size/Scale->Size to Isometric to get the aspect ratio of your graph to be proportional.

The next example uses the Katrina Data.jmp sample data table, which contains data on hurricane Katrina such as latitude, longitude, date, wind speed, pressure, and status. Adding a background map helps you see the path the hurricane took and impact on land based on size and strength. A script has been developed for this example and is part of the data table.

1. Select **Help > Sample Data Library** and open Katrina Data.jmp.
2. Select **Graph** > **Bubble Plot**.
3. Select LAT and click **Y**.
4. Select LON and click **X**.
5. Select Date and click **Time**.
6. Select WIND and click **Sizes**.
7. Select Stat and click **Coloring**.

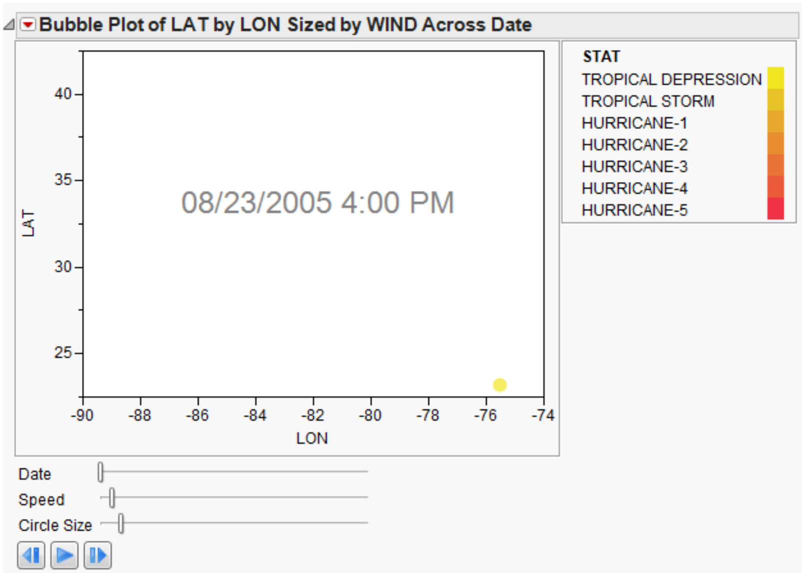
Figure 12.29 Bubble Plot Setup of Katrina Data.jmp



8. Click **OK**.

   The following image appears. The yellow dot shows the location of Tropical Depression Katrina on August 23, 2005.

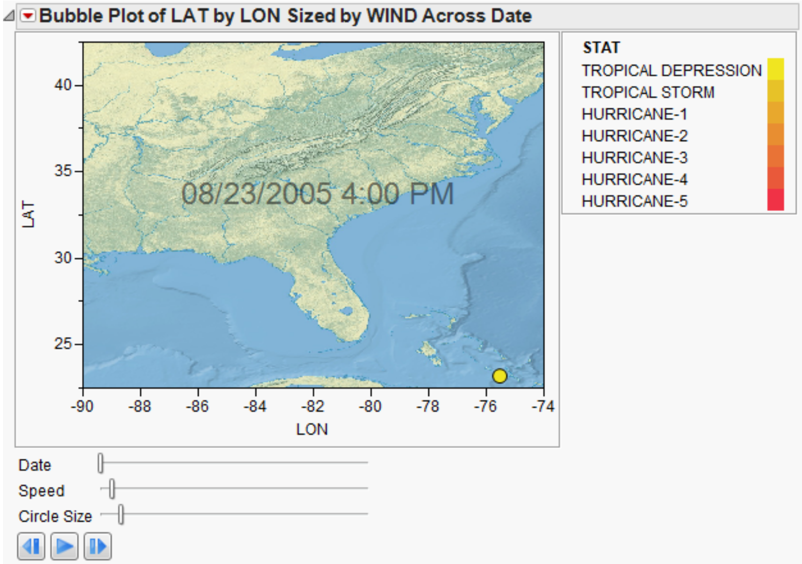
Figure 12.30 Bubble Plot of Katrina Data.jmp



   Note that even though the location of the storm is plotted, it does not really tell us where it is. To add more context, add a map in the background.

9. Right-click the graph and select **Background Map**. The Set Background Map window appears.
10. Select **Detailed Earth** and click **OK**.

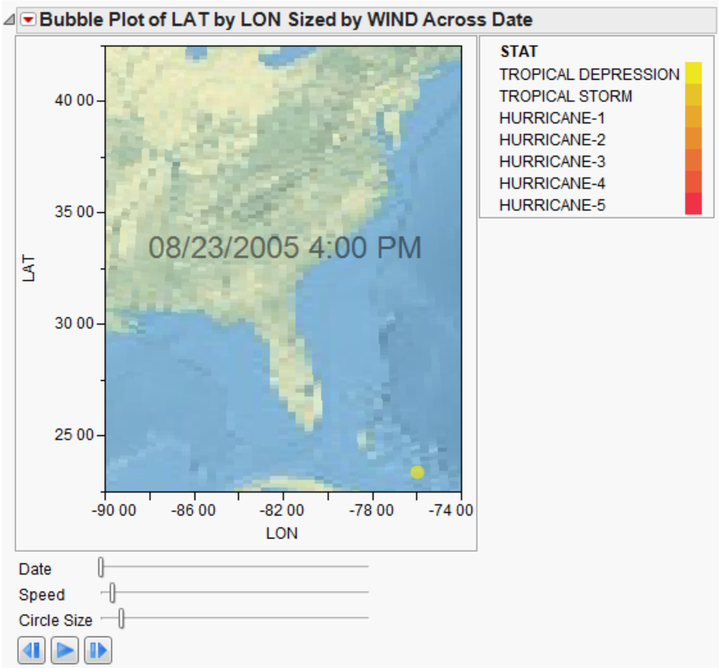
Figure 12.31 Bubble Plot of Katrina Data.jmp with Background Map



Now the coordinates make geographic sense. You can edit the axes and the size/scale to change how the graph appears.

11. Right-click the X axis (LON) and select **Axis Settings**. The X Axis Specification window appears.
12. Select **Scale** > **Geodesic US**.
13. Select **Format** > **Geographic** > **Longitude DMM**.
14. Click **OK**.
15. Repeat the same for the Y axis (LAT) except select **Format** > **Geographic** > **Latitude DMM**.
16. Right-click the map and select **Size/Scale** > **Size to Isometric.**

Figure 12.32 Bubble Plot of Katrina Data.jmp with Background Map



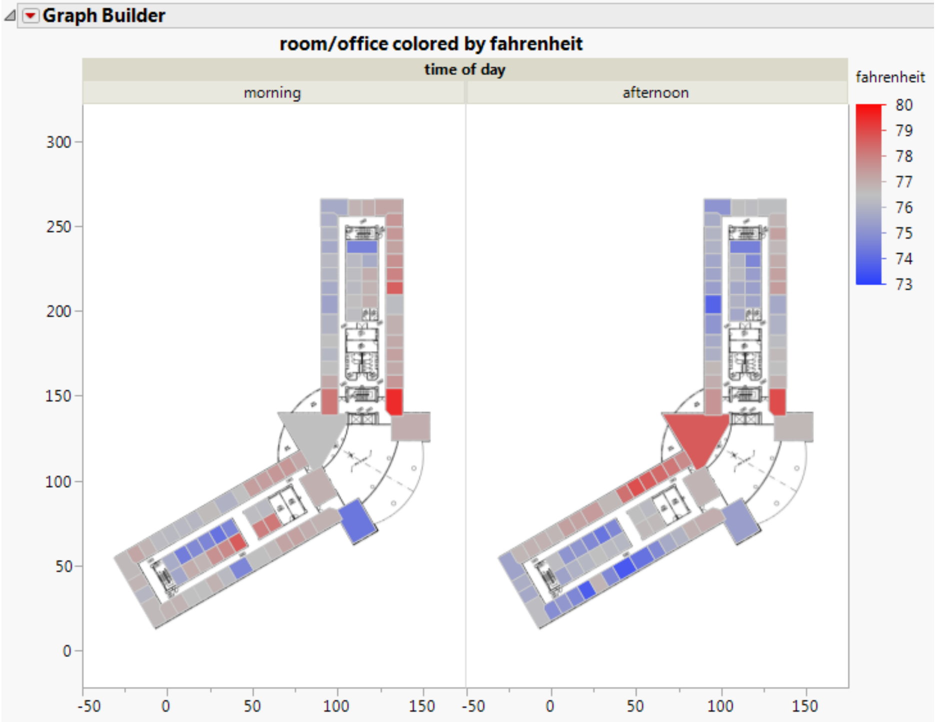
Click **Run** to view the animation of the hurricane data moving over the background map. You can manipulate the speed and the bubble size. Experiment with different options and view the displays. Adjust the axes or use the zoom tool to change what part of the area you are viewing. Add boundaries to the states. The map adjusts as the view does.

Office Temperature Study

This example demonstrates the creation of a custom background map for an office temperature study and how JMP was used to visualize the results. Data was collected concerning office temperatures for a floor within a building. A map was created for the floor using the Custom Map Creator add-in from the JMP File Exchange (https://community.jmp.com/docs/DOC-6218). Using Graph Builder, the office temperature results were then analyzed visually.

The map shown below is the floor, grouped by *time of day*. The color reflects the Fahrenheit value. Exploring data visually in this way can give hints as to what factors are affecting office temperature. Looking at this map, it appears the offices on the east side of the building are warmer in the mornings than they are in the afternoons. On the western side of the building, the opposite appears to be true. From this visualization, we might expect that both of these variables are affecting office temperatures, or perhaps that the interaction between these terms is significant. Such visuals help guide decision-making during the analysis.

Figure 12.33 Room/Office Colored by Fahrenheit and Grouped by Time of Day



First, data was collected and input into a data table (S4 Temps.jmp). Note the Room/Office column. It contains the unique names for each office and was assigned the Map Role to correctly define the map files.

Figure 12.34 Partial View of the S4 Temps.jmp Data Table

14/0 Cols	day	time of day	thermometer	sector	room/office
1	1	morning	T4	sector 1	S4068
2	1	morning	T4	sector 1	S4069
3	1	morning	T4	sector 1	S4070
4	1	morning	T4	sector 1	S4072
5	1	morning	T4	sector 1	S4074
6	1	morning	T4	sector 1	S4076
7	1	morning	T4	sector 1	S4078
8	1	morning	T4	sector 1	S4082
9	1	morning	T4	sector 1	S4084
10	1	morning	T4	sector 1	S4086
11	1	morning	T4	sector 1	S4088
12	1	afternoon	T2	sector 1	S4068
13	1	afternoon	T2	sector 1	S4069

Then, a map of the floor was created using the Custom Map Creator add-in, which you can download from the JMP File Exchange at https://community.jmp.com/docs/DOC-6218. The add-in creates two tables to define the shapes; an XY table and a Name table. The instructions below describe how it was built.

Create a Map of the Floor

1. Launch the add-in through the menu items **Add-Ins > Map Shapes > Custom Map Creator**. Two tables open in the background followed by the Custom Map Creator Window.
2. Drag a background image into the graph frame. An image of the floor plan was available.
3. Perform any resizing on the background image and graph the frame.
4. Name the table (for example, S4).
5. Click **Next**.
6. Name the shape that you are about to define. For this example, each office was individually named for the map (for example, S4001).
7. Within the graph frame, use your mouse to click all of the boundaries of the shape that you want to define. A line appears that connects all of the boundary points.
8. As soon as you finish defining the boundaries of the shape, click **Next Shape**. Continue adding shapes until you have completed the floor plan. Note that you do not need to connect the final boundary point; the add-in automatically does that for you when you click **Next Shape**.
9. The line size and color can be changed. In addition, checking **Fill Shapes** fills each shape with a random color.
10. Click **Finish**.

The custom map files were created and named appropriately. The map files are S4-Name.jmp and S4-XY.jmp and have been saved in the JMP Samples/Data folder.

Add the Map Role Column Property

Note: Pathnames in this section refer to the “JMP” folder. On Windows, in JMP Pro, the “JMP” folder is named “JMPPro”.

The S4 Temps.jmp data table contains office data over a three-day period. Set up the **Map Role** column property in the data table, as follows:

1. Select **Help > Sample Data Library** and open S4 Temps.jmp.
2. Right-click the **Room/Office** column and select **Column Info**.
3. Select **Column Properties > Map Role**.
4. Select **Shape Name Use**.

5. Click the icon next to **Map name data table** and browse to the S4-Name.jmp file (located in the JMP Samples/Data folder).

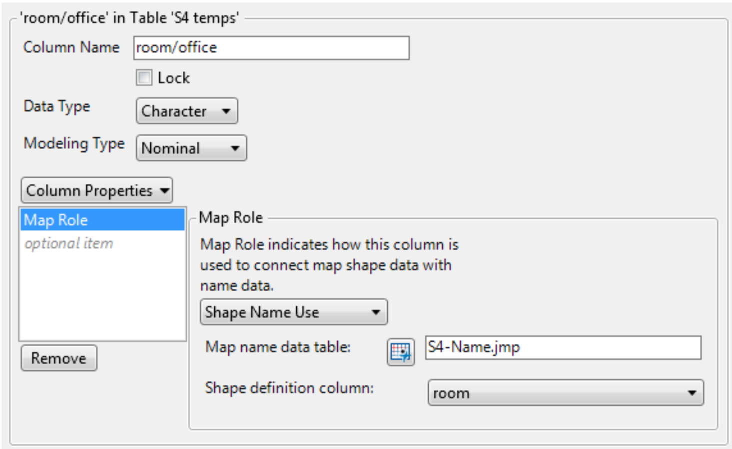
   This tells JMP where the data tables containing the map information reside.

6. Select room from the **Shape definition column** list.

   Room is the specific column within the S4-Name.jmp data table that contains the unique names for each office. Notice that the room column has the **Shape Name Definition Map Role** property assigned, as part of correctly defining the map files.

   Note: Remember, the Shape ID column in the -Name data table maps to the Shape ID column in the -XY data table. This means that indicating where the -Name data table resides links it to the -XY data table, so that JMP has everything that it needs to create the map.

Figure 12.35 Map Role Column Property



7. Click **OK**.

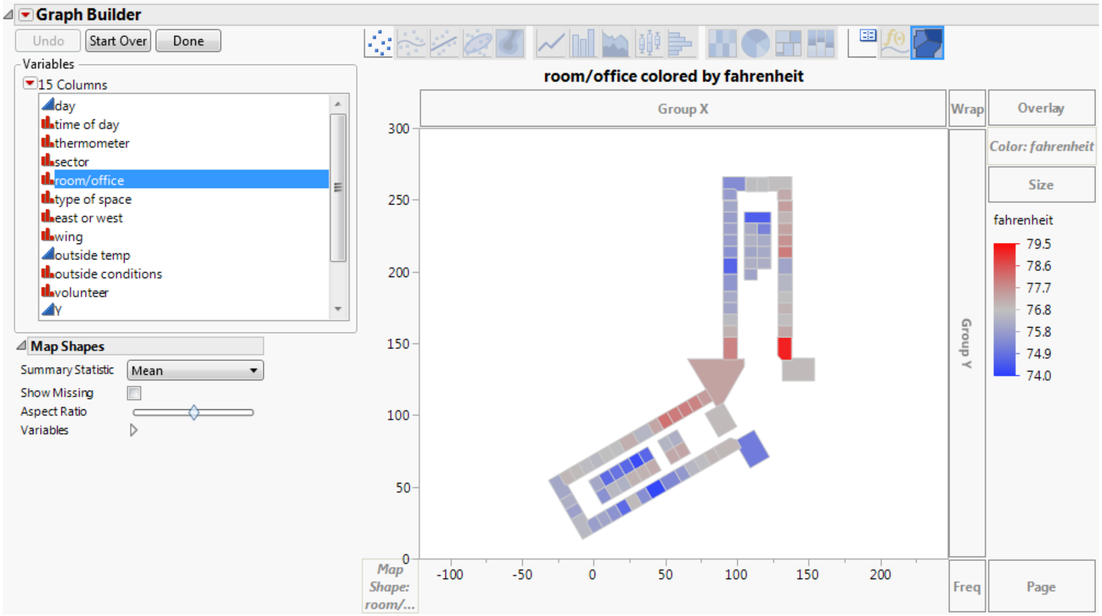
Once the **Map Role** column property has been set up, you can perform your analysis. You want to visually see the differences in office temperatures throughout the floor.

1. Select **Graph > Graph Builder**.
2. Drag and drop room/office into the **Map Shape** zone.

   Since you have defined the **Map Role** column property on this column, the map appears.

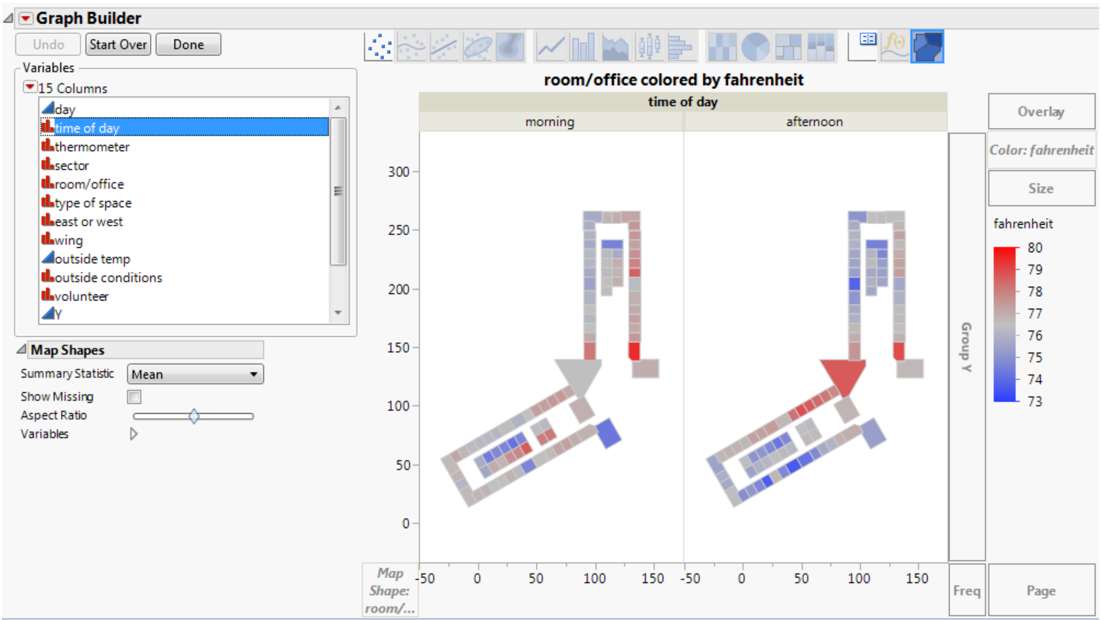
3. Drag and drop Fahrenheit to the **Color** zone.

Figure 12.36 Room/Office Colored by Fahrenheit



4. Drag and drop Time of Day onto Group X zone.

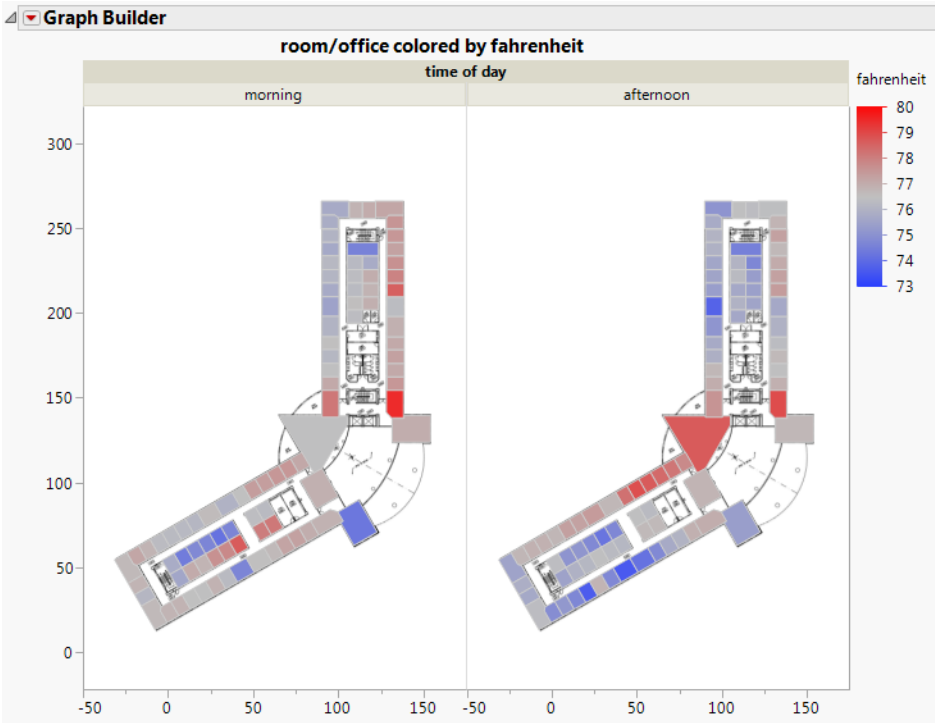
Figure 12.37 Room/Office Colored by Fahrenheit and Grouped by Time of Day



Note that only the offices that were part of the study and were created using the Custom Map Creator add-in are displayed. To add the entire floor plan image, the original floor plan graphic was dragged and dropped onto the Graph Builder window to create Figure 12.38.

To view Figure 12.38, select **Help > Sample Data Library** and open S4 Temps.jmp and run the **by Time of Day** script.

Figure 12.38 Room/Office Map with Original Floor Plan

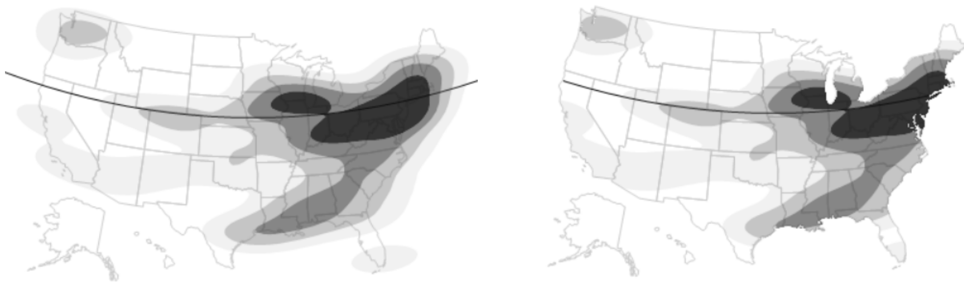


There are several scripts provided with the data table that you can run to view the various analysis and modeling that can be performed and visually displayed.

Boundary Map with Clipped Elements

Graphical elements (display segments such as reference, grid lines, and contours) can be clipped to conform to the boundaries of a geographical map. The Customize Graph window provides Clip Shape options, where you can select a boundary around which to clip the shape.

Figure 12.39 Unclipped (Left) and Clipped (Right) Reference Lines and Contours

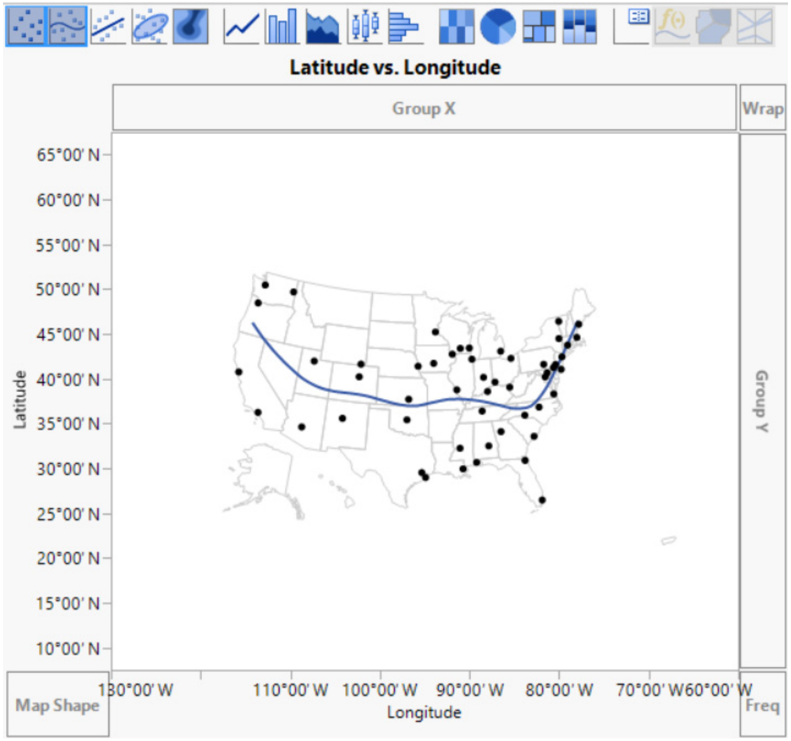


Create the Background Map

1. Select **Help > Sample Data Library** and open Cities.jmp.
2. Select **Graph > Graph Builder**.
3. Drag the Latitude column to the Y axis.
4. Drag the Longitude column to the X axis.
5. Right-click the graph and select **Graph > Background Map**.
6. In the Boundaries column, select **US States** and click **OK**.

   A map of the United States appears on the graph.

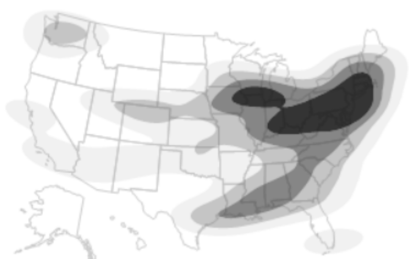
Figure 12.40 Background Map



7. Click the Points icon  to remove the points.
8. Click the Smoother icon  to remove the smoother.
9. Click the Contour icon .

   A contour plot appears on the map.

Figure 12.41 Unclipped Contours on Background Map



Add a Reference Line

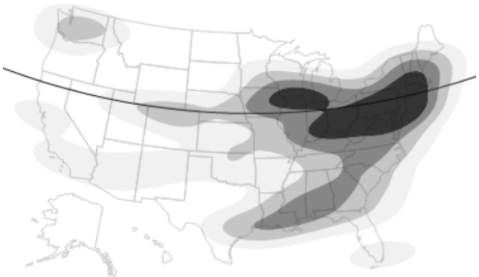
1. On the Latitude axis, double-click **45°00’ N** to open the Y Axis Settings window.

2. In the Reference Lines pane, change the first number in the Value box to “40”.

   For example, in 44°40.11′ N, change “44” to “40”.

3. Click **Add** and then **OK**.

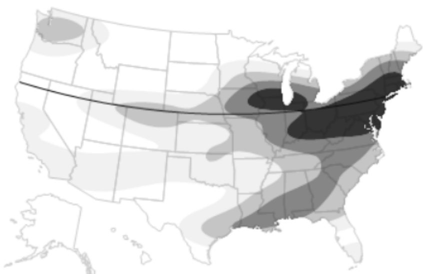
Figure 12.42 Unclipped Reference Line



Assign a Clip Shape

1. Right-click the graph and select **Customize**.
2. Select **Reference Lines** and select **US States** from the Clip Shape list.
3. Select **Contour**, select **US States** from the Clip Shape list, and click **OK**.

Figure 12.43 Clipped Reference Line and Contours



Tips:

- The boundaries that appear in the Background Map window are installed in the JMP Maps installation folder. See “Custom Map Files” on page 273 for more information about adding your own boundaries.
- Scripting provides additional options, such as specifying the clipping path in a matrix or string. See the Scripting Graphs chapter in the *Scripting Guide*.

Appendix A

Legacy Platforms

The following platforms are legacy platforms. A legacy platform is an older platform that contains features that can be accessed in a newer platform. The charts and plots in this appendix are all available in the Graph Builder platform:

Table A.1 Legacy Platforms

Platform	Graph Builder References
Treemap	• "Example of a Treemap" on page 151 in the "Graph Builder Examples" chapter • "Treemap" on page 87 in the "Graph Builder" chapter
Chart	• "Example of a Stacked Bar Chart" on page 121 in the "Graph Builder Examples" chapter • "Bar", "Line", or "Pie" on page 85 in the "Graph Builder" chapter
Overlay Plot	• "Example of Overlaying Histograms with Transparency" on page 144 in the "Graph Builder Examples" chapter • "Graph Builder" chapter on page 27

Contents

Treemaps

Treemaps are useful for observing patterns among groups that have many levels. Treemaps are like bar charts that have been folded over in two dimensions so that there is no unused space. Rather than drawing a single bar for each measurement, a treemap can show the magnitude of a measurement by varying the size or color of a rectangular area. Use treemaps when your data contains many categories, to visualize many groups.

Treemaps were named and popularized by Ben Schneiderman, who has an extensive website about the idea (https://www.cs.umd.edu/hcil/treemap-history/index.shtml).

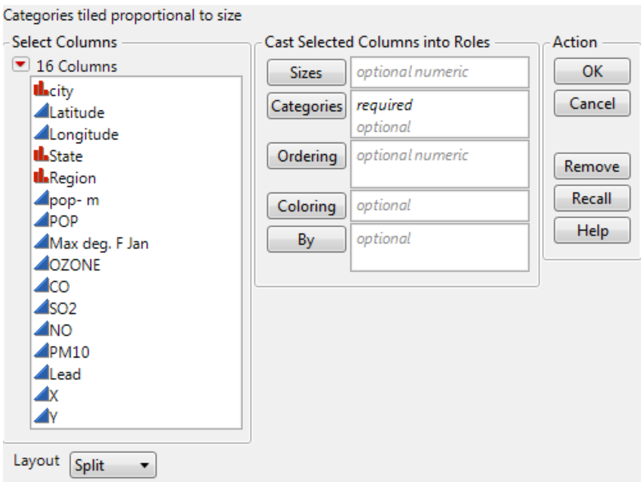
Example of Treemaps

To create a treemap, consider using Graph Builder. See “Example of a Treemap” on page 151 in the “Graph Builder Examples” chapter.

Launch the Treemap Platform

Launch the Treemap platform by selecting **Graph > Legacy > Treemap**.

Figure A.1 The Treemap Launch Window



Sizes Determines the size of the rectangles based on the values of the specified variable. See “Sizes” on page 312.

Categories Specifies the category that comprises the treemap. See “Categories” on page 312.

Ordering Changes the ordering from alphabetical (where values progress from the top left to the lower right) to order by the specified variable. You can specify more than one ordering variable. See “Ordering” on page 312.

Coloring Colors the rectangles corresponding to the levels of the specified variable.

– If the variable is continuous, the colors are based on a continuous color spectrum.

– If the variable is categorical, the default colors are selected in order from JMP's color theme.

See “Coloring” on page 313.

By Identifies a column that creates a report consisting of separate treemaps for each level of the variable.

Layout Determines the layout of the rectangles. See “Layout” on page 315.

After you click **OK**, the Treemap window appears. See “The Treemap Window” on page 314.

Sizes

If you want the size of the rectangles to correspond to the levels of a variable, specify a **Sizes** variable. The rectangle size is proportional to the sum of the **Sizes** variable across all of the rows corresponding to a category. If you do not specify a **Sizes** variable, the rectangle size is proportional to the number of rows for each category.

Categories

The only required variable role for the Treemap platform is **Categories**. If you specify only a **Categories** variable and no other variables, the rectangles in the treemap have these attributes:

- They are colored from a rotating color theme.
- They are arranged alphabetically.
- They are sized by the number of occurrences in each group.

If you specify two **Categories** variables, the treemap is grouped by the first variable, and sorts within groups by the second variable. For example, using the Cities.jmp sample data table, specify Region and city (in that order) as the **Categories** variables.

Ordering

By default, the rectangles in a treemap appear in alphabetical order. Values progress from the top left to the lower right. To change this ordering, specify an **Ordering** variable. When an **Ordering** variable is specified, the rectangles appear with the values progressing from the bottom left to the upper right.

Note: If you use the Squarify option, the rectangles are ordered by size and the Ordering variable is ignored.

If you specify a single **Ordering** variable, the rectangles are clustered, with the high levels or large values together, and the low levels or small values together.

If you specify two **Ordering** variables, the treemap arranges the rectangles horizontally by the first ordering variable, and vertically by the second ordering variable. This approach can be useful for geographic data.

For example, suppose the Cities.jmp data is ordered by US state. If you order the data by latitude and or longitude, the Treemap rectangles are drawn ordered by the location of the US state.

Coloring

If you specify a **Coloring** variable, the colors of the rectangles correspond to the levels of the variable.

- If the variable is continuous, the colors are based on a continuous color theme setting. The default color theme is **Blue to Gray to Red**. The color of each value is based on the average value of all of the rows. Blue represents the lowest values, and red represents the highest values. The color is most intense at the extremes of the variable, and paler colors correspond to levels that are close to the mean. For example, see "The Treemap Window" on page 314.
- If the variable is categorical, the colors are based on a categorical color theme. The default color theme is **JMP Default**.

Note: If you have used the **Value Colors** column property to color a column, that property determines the colors of the categories.

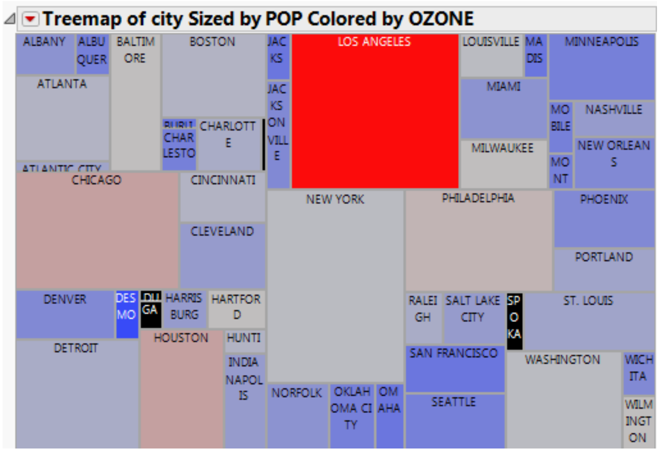
- If you do not specify a **Coloring** variable, colors are chosen from a rotating color palette.

Related Information

- "Treemap Platform Options" on page 315

The Treemap Window

Figure A.2 The Treemap Window



Tip: To zoom in on the treemap, use the **Magnifier** tool or press the Z key.

Treemap rectangles can have the following attributes:

- Categories add labels to the rectangles. You can specify one or two categories. You can show or hide labels using the **Show Labels** option. See "Treemap Right-Click Menu" on page 316.
- Rectangle size is determined by one of the following:
  - The **Sizes** variable, if you specify one.
  - If you do not specify a **Sizes** variable, size is determined by the frequency of the category.
- Rectangle color is determined by one of the following:
  - If the variable is continuous, the colors are based on a continuous color theme setting. The default color theme is **Blue to Gray to Red**. The color of each value is based on the average value of all of the rows. Blue represents the lowest values, and red represents the highest values. The color is most intense at the extremes of the variable, and paler colors correspond to levels that are close to the mean.
  - If the variable is categorical, the colors are based on a categorical color theme. The default color theme is **JMP Default**.
  - If you do not specify a **Coloring** variable, colors are chosen from a rotating color theme.
  - If you have used the **Value Colors** column property to color a column, that property determines the colors of the categories.

- The order of the rectangles is determined by one of the following:
  - The **Ordering** variable, if you specify one.
  - If you do not specify an **Ordering** variable, the order is alphabetical by default. Values progress from the top left to the lower right.

Treemap Platform Options

The Treemap red triangle menu provides the following options:

Change Color Column Change the column that is currently used to color the rectangles.

Color Theme Change the colors representing the high, middle, and low values of the color column. This option is available only if you have specified a **Coloring** variable.

Color Range Specify the range that you want applied to the color gradient. The default low value is the column minimum, and the default high value is the column maximum. This option is available only if you have specified a continuous column as the **Coloring** variable.

Legend Shows or hides a legend that defines the coloring used on the treemap. This option is available only if you have specified a **Coloring** variable.

Layout Arranges rectangles by order of the variable or by size of the rectangle.

Split Preserves the order of the data. Split is the default setting.

Squarify Sorts the data first. The largest value is in the top left corner. The rectangle sizes decrease diagonally to the lower right corner. The order of the variables is not preserved, but visually comparing variables is easier.

Mixed Preserves the order of the data for the main category and then sorts within the subcategory. Applies only when you select a category and subcategory.

See the JMP Reports chapter in *Using JMP* for more information about the following options:

Local Data Filter Shows or hides the local data filter that enables you to filter the data used in a specific report.

Redo Contains options that enable you to repeat or relaunch the analysis. In platforms that support the feature, the Automatic Recalc option immediately reflects the changes that you make to the data table in the corresponding report window.

Save Script Contains options that enable you to save a script that reproduces the report to several destinations.

Save By-Group Script Contains options that enable you to save a script that reproduces the platform report for all levels of a By variable to several destinations. Available only when a By variable is specified in the launch window.

Treemap Right-Click Menu

Right-click a treemap to use the following options:

Suppress Box Frames Suppresses the black lines outlining each box.

Ignore Group Hierarchy Flattens the hierarchy and sorts by the **Ordering** columns without using grouping, except to define cells.

Show Labels Shows or hides the **Categories** labels. If you have specified two **Categories**, the secondary labels are hidden or shown. If you hide the **Categories** labels, place your mouse pointer over a rectangle to show the primary or secondary category label.

Show Group Labels Shows or hides the group labels. This option is available only if more than one variable is assigned to the **Categories** role.

Group Label Background Adjust the transparency of the group label. This option is available only if more than one variable is assigned to the **Categories** role.

Charts

The Chart platform on the **Graph** menu charts continuous variables versus categorical variables. The continuous variables are summarized for each categorical level. Chart supports several chart types, such as bar charts, pie charts, and line charts. Chart is similar to the **Tables > Summary** command and is useful for making graphical representations of summary statistics.

If you want to make a plot of individual data points (rather than summaries of data points), we recommend using an overlay plot instead.

Example of the Chart Platform

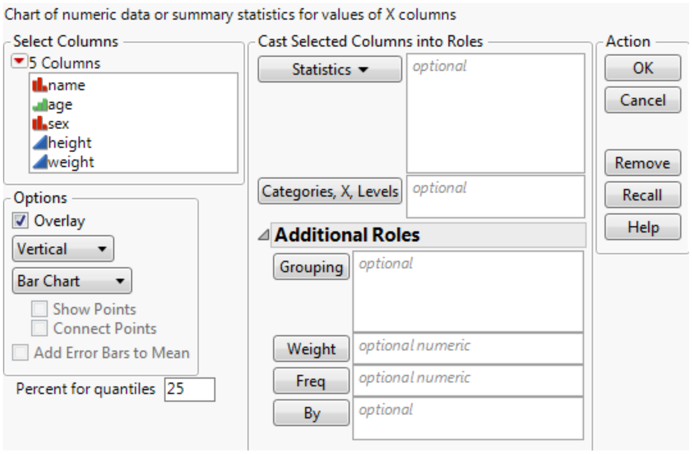
To create a chart, consider using Graph Builder. See “Example of a Stacked Bar Chart” on page 121 in the “Graph Builder Examples” chapter.

If you use the Chart platform, your chart appears in Graph Builder.

Launch the Chart Platform

Launch the Chart platform by selecting **Graph > Legacy > Chart**.

Figure A.3 The Chart Launch Window



In the Chart launch window, you can assign the following:

- Up to two X variables, which appear on the *x*-axis in the same order that you assign them in the launch window.
- As many Y variables (statistics) as you want. If the data is already summarized, select **Data** as the statistics option.

The following options are available:

Cast Selected Columns Into Roles

Statistics Use this menu to select the statistic to chart for each Y variable.

See "Plot Statistics for Y Variables" on page 319.

Categories, X, Levels Select up to two X variables whose levels are categories on the *x*-axis. The Chart platform produces a bar for each level or combination of levels of the X variables. If you do not specify an X variable, the chart has a bar for each row in the data table.

See "Use Categorical Variables" on page 320.

Additional Roles

Grouping For one or more **Grouping** variables, independent results are produced for each level or combination of levels of the grouping variables. These results appear in the same report window, but in separate plots. See "Use Grouping Variables" on page 321.

Weight Assigns a variable to give the observations different weights.

Freq Assigns a frequency variable. This is useful if you have summarized data.

By By variables cause plots to be created in separate outline nodes.

Options

Overlay If this option is selected, all Y variables are plotted in one graph. If this option is not selected, each Y variable is plotted in its own graph. This option is selected by default where available.

This option is available for all chart types except **Pie Chart**.

Chart Orientation Select Vertical for a vertical chart or Horizontal for a horizontal chart. Vertical is the default option.

This option is available for all chart types except **Pie Chart**.

Chart Type Select the type of chart that you want. Available chart types are **Bar Chart**, **Line Chart**, **Pie Chart**, **Needle Chart**, and **Point Chart**. Selecting a chart controls which of the other options are available.

You can always change these options after the chart appears. See "Y Options" on page 326.

Show Points Shows the points in the plot. This option is selected by default where available.

This option is available for all chart types except **Bar Chart** and **Pie Chart**.

Connect Points Connects the points in the plot. Show Points does not have to be selected to connect points. This option is selected by default where available.

This option is available only for **Line Chart**.

Add Error Bars to Mean Adds error bars when the **Mean** statistic is selected for at least one Y variable and at least one X variable is assigned. This option is not selected by default.

This option is available for **Line Chart** and **Bar Chart**, and additional options are added to the Chart launch window. See "Adding Error Bars" on page 321.

Percent for quantiles Sets the specific quantile when the **Quantiles** statistic is selected for at least one Y variable. The default value is 25. Specify a different quantile:

1. Type the value in the **Percent for quantiles** box.

2. Select a column.
3. Click **Statistics**.
4. Select **Quantiles** from the menu of statistics.

For more information about the launch window, see the Get Started chapter in *Using JMP*.

After you click **OK**, the Chart report window appears. See "The Chart Report" on page 322.

Plot Statistics for Y Variables

You can plot the raw data for Y variables, or you can plot as many statistics as you want on the *y*-axis. The **Statistics** menu in the Chart launch window lists the available statistics. To specify the *y*-axis, highlight one or more numeric columns in the Select Columns list and select from the list of statistics. If all the statistics requested are counting statistics (for example, **N**) for the same column, that column is used as the category variable.

The statistics in the Chart platform are the same as those computed by statistical platforms in the **Analyze** menu and the **Summary** command in the **Tables** menu. The following statistics are available:

Data The value of each row in the data table when there is no categorical variable. If there is a categorical variable, **Data** produces a point plot within the variable's levels.

N The number of nonmissing values. Also used to compute statistics when there is no column assigned as a weight variable. The Chart platform shows **N** for each level of a categorical variable.

Mean The arithmetic average of a column's values. The mean is the sum of nonmissing values divided by the number of nonmissing values.

Std Dev The sample standard deviation computed for each level of a categorical variable. It is the square root of the variance of the level values.

Min The least value, excluding missing values, in the level of a categorical variable.

Max The greatest value in the level of a categorical variable.

Range The difference between the maximum and minimum values in each level of a categorical variable.

% of Total The percentage of the total number of rows represented by each level of the **Categories, X, Levels** variable. If summary statistics are requested on a continuous variable, then the **% of Total** equals the proportion of the sum represented by each level of the **Categories, X, Levels** variable.

N Missing The number of missing values in each level of a categorical variable.

N Categories Also known as n-categories. The combinatorial or algebraic models of directed spaces.

Sum The sum of all values in each level of a categorical variable.

Sum Wgt The sum of all values in a column assigned as Weight. Also used instead of **N** to compute other statistics. **Chart** shows the sum of the weight variable for each level of a categorical variable.

Variance The sample variance computed for each level of a categorical variable.

Std Err The standard error of the mean of each level of a categorical variable. It is the standard deviation, **Std Dev**, divided by the square root of **N** for each level. If a column is assigned a weight variable, then the denominator is the square root of the sum of the weights.

CV The coefficient of variation of a column's values. The CV is computed by dividing the column standard deviation by the column mean and multiplying by 100.

Median The middle value in each level of a categorical variable. Half of the values in the level are greater than or equal to the median and half are less than the median.

Geometric Mean The nth root of the product of the data. For example, geometric means are often used to calculate interest rates. The statistic is also helpful when the data contains a large value in a skewed distribution.

Note: Negative values result in missing numbers, and zero values (with no negative values) result in zero.

Interquartile Range The measure of statistical dispersion (difference between the upper and lower quartiles) often used to find outliers in data. Also known as the midspread or middle fifty.

Quantiles Divides a data set so that n% of the data is below the n^{th} quantile. To compute a specific quantile, enter the quantile value in the box located in the lower left of the Chart launch window *before* requesting **Quantile** from the menu.

Use Categorical Variables

You can assign zero, one, or two X variables whose levels are categories on the *x*-axis. The Chart platform produces a bar (or a needle, or a pie slice, and so on) for each level or combination of levels of the X variables. If you do not specify any X variable, the chart has a bar for each row in the data table.

The following table shows what type of chart to expect based on the number of X and Y variables.

X	Y	Type of Chart
none	one or more	If you do not specify a variable for categories, most statistics produce a bar (or a needle, or a pie slice, and so on) for each observation in the data table. This is useful when your data is already summarized. In that case, you usually specify **Data** as the statistic to plot. Each bar reflects the value of the Y variable.
one or two	none	Plots the counts for each level of the X variable. For two X variables, the counts for each level of both X variables are included (or overlaid) in a single chart.
one or two	one or more	Plots the selected statistics for each level of the X variable. For two X variables, the selected statistics for each level of the X variables are included (or overlaid) in a single chart.

Use Grouping Variables

If you specify one grouping variable, the result is a separate chart for each level of the grouping variable. All charts are under the same outline title. If you used the same variable as a By variable instead, the same separate charts are produced, but each chart is under its own outline title.

If you specify two or more grouping variables, the result is a matrix of charts. Each chart shows a combination of one level from each of the grouping variables.

If there are multiple statistics, the **Overlay** option is checked by default, and the Y variables (statistics) are plotted on the same chart for each level of the grouping variable. However, the levels of the grouping variable cannot be overlaid into the same plot frame. For example, if the levels of your grouping variable are Male and Female, the **Overlay** option cannot be used to combine the two Male and Female graphs into one graph. To see that type of result, use **Categories, X, Levels** instead of **Grouping** variables.

Adding Error Bars

Error bars are available when the **Mean** statistic is selected for at least one Y variable, and at least one X variable is assigned. Error Bars are not available for pie charts. Selecting **Add Error Bars to Mean** causes additional options to appear in the Chart launch window.

After the option is checked, select a type of error bar from the menu that appears. Some of the types of error bar have an additional numeric field. The following types of error bars are available:

Range Creates error bars based on the range of the data.

Standard Error Creates error bars based on the standard error of the mean. You can specify the number of standard errors.

Standard Deviation Creates error bars based on the standard deviation of the data. You can specify the number of standard deviations.

Confidence Interval Creates error bars based on a confidence interval of the mean. The standard deviation used for the confidence interval is separate for each bar. You can specify the level of confidence.

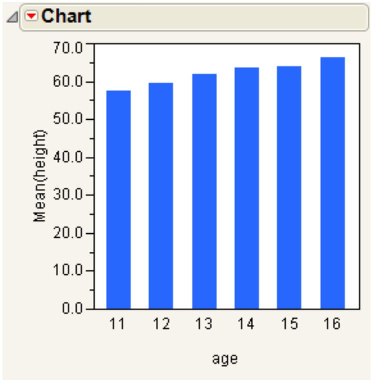
Confidence Interval (pooled) Creates error bars based on a confidence interval of the mean. The standard deviation used for the confidence interval is based on the pooled standard deviation. This option is not available if you have more than one category variable. You can specify the level of confidence.

The Chart Report

Follow the instructions in "Example of the Chart Platform" on page 316 to produce the report shown in Figure A.4.

Charts can be bar charts, pie charts, line charts, needle charts, point charts, and range charts. Figure A.4 shows a standard bar chart.

Figure A.4 The Initial Chart Report Window

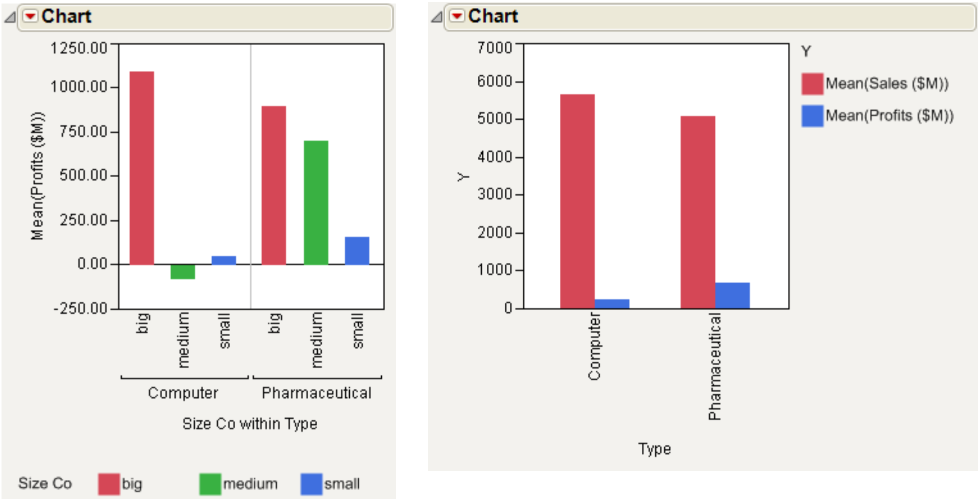


For information about additional options for the report, see "Chart Platform Options" on page 324.

Legends

Legends are shown as needed. If your chart uses different colors or markers to show levels of one or two X variables, a legend below the chart shows them. If your chart uses different colors or markers to show more than one statistic, a legend to the right of the chart shows them.

Figure A.5 Examples of an X Legend (left) and Y Legend (right)



Ordering

By default, the Chart platform orders the bars using one of the common orders supported by JMP (months, days of the week, and so on). If the grouping column has a Value Order column property, it uses that order. With no Value Order property, bars are drawn in alphanumeric order.

Coloring Bars in a Chart

There are a few ways to color bars after the chart has been created.

Manually Set the Color of All Bars

1. Ensure that no bars are selected.
2. Click the Chart red triangle and select **Level Options > Colors**.
3. Select a color from the color palette that appears.

Set the Color of a Single Bar

1. Select a bar in the chart.

2. Click the Chart red triangle and select **Level Options > Colors**.
3. Select a color from the color palette that appears.

Note: If you assigned *both* a category variable *and* a grouping variable in your chart, all of the bars are colored even if you selected only one bar.

Set the Color of a Single Bar Using the Legend

1. Select the legend bar color.
2. Right-click and select **Colors**.
3. Select a color from the color palette that appears.

Automatically Assign a Color to a Level

1. Select the column in the data table.
2. Select **Cols > Column Info**.
3. Assign colors using **Value Colors** in the **Column Properties** menu. For more information about the Value Colors property, see The Column Info Window chapter in *Using JMP*.

Chart Platform Options

The basic Chart report is shown in Figure A.4 on page 322.

The Chart platform has plotting options on the red triangle menu on the Chart title bar. When you select one of these options at the platform level, it affects all plots in the report if no legend levels are highlighted. If one or more plot legend levels are highlighted, the options affect only those levels. There is also a single-plot options menu for each Y variable, which appears when you highlight a Y variable legend beneath the plot and right-click.

The individual plot options are the same as those in the **Y Options** submenu at the platform level. See "Y Options" on page 326.

General Platform Options

When you select one of the following options at the platform level, it affects all plots in the report if no legend levels are highlighted. If one or more plot legend levels are highlighted, the options affect only those plots.

Overlay Displays a single overlaid chart when you have more than one Y (statistics) variable. Each statistic can be assigned a different type of chart (such as line and bar) and overlaid in a single plot. **Overlay** is selected by default. The axis notation only shows for the last chart displayed if the charts are not overlaid. When **Overlay** is not selected, the platform shows duplicate axis notation for each chart.

Vertical Chart, Horizontal Chart Changes horizontal charts to vertical charts (**Vertical**), or vertical charts to horizontal charts (**Horizontal**). Affects all charts in the report. Pie charts are converted to bar chats.

Pie Chart Changes a horizontal or vertical chart type to a pie chart.

Range Chart Displays a range chart. You can change any chart that includes at least two statistics in a single plot into a range chart.

Add Error Bars to Mean Adds error bars to charts based on means. A window opens, prompting you to select the type of error bar. If error bars already exist on a chart, you can change the error bar type. See "Adding Error Bars" on page 321 for a description of error bar types.

Stack Bars Stacks the bars from levels of a subgroup end-to-end. To use this option, you need two **Categories, X, Levels** variables and a statistic. See "Example of a Stacked Bar Chart" on page 121 in the "Graph Builder Examples" chapter for an example of stacking bars.

Y Options Contains the options described in "Y Options" on page 326. To apply these options to a single Y variable, highlight that variable in the legend first.

Level Options Selects colors and markers. If no levels (bars, points, or pie slices) are selected, the color or marker that you select is applied to all levels. If you select one or more levels, the color or marker that you select is applied only to the selected levels. See "Coloring Bars in a Chart" on page 323.

Label Options Attaches labels to your plots. In the **Label Options** menu, the first two options (**Show Labels** and **Remove Labels**) turn labels on and off. The last three options (**Label by Value**, **Label by Percent of Total Values**, **Label By Row**) specify what label should appear. Only one label can be shown at a time. Label options are also available by right-clicking in the chart.

Thick Connecting Line Toggles the connecting line in a line chart to be thick or thin.

Show Y Legend Shows the Y legend of the plot. This option is on by default for overlaid charts.

Show Level Legend Shows the level legend of the plot. This option is on by default when the **Show Separate Axes** option is selected.

Show Separate Axes Duplicates the axis notation for each chart when there are multiple charts. By default, the axis notation only shows for the last chart displayed if the charts are not overlaid. This option is not available for grouped charts.

Ungroup Charts Moves level identifiers from the right side of the charts to beneath the charts for individual charts when a grouping variable is specified.

See the JMP Reports chapter in *Using JMP* for more information about the following options:

Local Data Filter Shows or hides the local data filter that enables you to filter the data used in a specific report.

Redo Contains options that enable you to repeat or relaunch the analysis. In platforms that support the feature, the Automatic Recalc option immediately reflects the changes that you make to the data table in the corresponding report window.

Save Script Contains options that enable you to save a script that reproduces the report to several destinations.

Save By-Group Script Contains options that enable you to save a script that reproduces the platform report for all levels of a By variable to several destinations. Available only when a By variable is specified in the launch window.

Y Options

This section describes the **Y Options** submenu. These commands apply to all Y variables, unless you have a legend level highlighted, then they apply to only the highlighted Y variable.

Click the legend within a plot to highlight a Y. If you right-click a highlighted legend level, the commands to modify that Y appear. The commands then affect only the highlighted Y.

Bar Chart Displays a bar for each level of the chart variables. The default chart is a bar chart.

Line Chart Replaces a bar chart with a line chart and connects each point with a straight line. Select the **Show Points** option to show or hide the points. **Line Chart** is also available as a platform option, which then applies to all charts at once.

Needle Chart Replaces each bar with a line drawn from the axis to the plotted value. **Needle Chart** is also available as a platform option, which then applies to all charts at once.

Point Chart Shows only the plotted points, without connecting them.

Show Points Toggles the point markers on a line or needle chart on or off.

Connect Points Toggles the line connecting points on or off.

Show Error Bars Toggles the error bars on plots of means. Note that this option is available only for plots that involve means of variables.

Overlay Color Assigns a color to statistics (y-axis) to identify them in overlaid charts.

Overlay Marker Assigns a marker to statistics, to identify them in overlaid charts.

Pen Style Selects a line style for connecting lines.

Label Format Specifies the format, field width, and number of decimals for labels. Thousands separators can be turned on or off. Enter the values in the window that appears.

Note: To see the impact of this change, you must turn on labels. Click the Chart red triangle and select **Label Options > Show Labels.**

Overlay Plots

The Overlay Plot option produces plots of a single X column and one or more numeric Ys and does not accept non-numeric values for the *y*-axis. Curves can also be shown as separate plots for each Y with a common *x*-axis. Plots can be modified with range and needle options, color, log axes, and grid lines. Curves with two different scales can be overlaid on the same plot with the addition of a right axis.

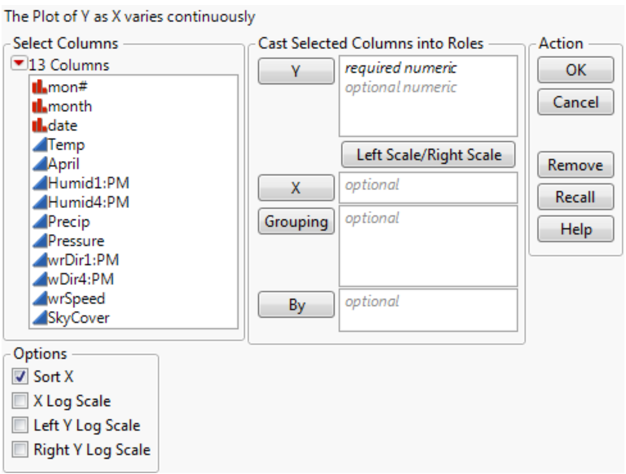
Example of an Overlay Plot

To create an overlay plot, you can use the Overlay zone in Graph Builder. See "Example of Overlaying Histograms with Transparency" on page 144 in the "Graph Builder Examples" chapter.

Launch the Overlay Plot Platform

Launch Overlay Plot by selecting **Graph > Legacy > Overlay Plot**.

Figure A.6 The Overlay Plot Launch Window



In the Overlay Plot Launch window, you assign the following:

- one X variable of any modeling type
- as many numeric Y variables as you want

Cast Selected Columns Into Roles

X, Y You can graph many numeric Y variables against a single X variable.

Left Scale/Right Scale The columns assigned to the **Y** role have a left- or right-pointing arrow to the left of the column name. This arrow designates on which vertical axis (on the left or right of the plot) the variable appears. Change the designation by highlighting the column in the **Y** list and clicking the **Left Scale/Right Scale** button.

Grouping This option produces a matrix of graphs for each Grouping variable.

By This option produces a separate graph for each level of the By variable. If two By variables are assigned, a separate graph for each possible combination of the levels of both By variables is produced.

Options

Sort X This option causes the points to be connected in order of ascending X values. Otherwise, the points are connected in row order. This option is selected by default.

X Log Scale This option applies a log scale to the *x*-axis.

Left Y Log Scale This option applies a log scale to the left *y*-axis. It is available only if one or more *Y* variables are left-scaled. (See **Left Scale/Right Scale**.)

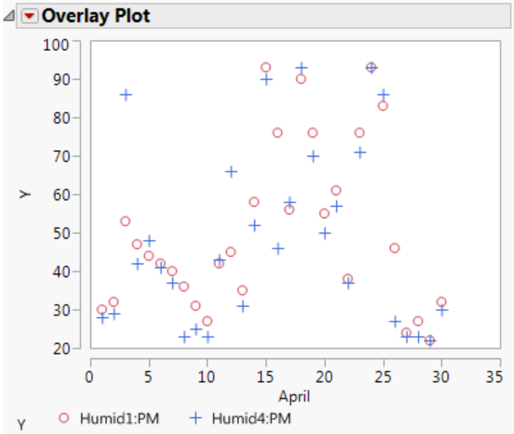
Right Y Log Scale This option applies a log scale to the right *y*-axis. It is available only if one or more Y variables are right-scaled. (See **Left Scale/Right Scale**.)

After you click **OK**, the Overlay plot appears. See "The Overlay Plot" on page 328.

The Overlay Plot

Initially, this platform overlays all specified Y columns. The legend below the plot shows individual markers and colors that identify each Y column. For information about additional options for the plot, see "Overlay Plot Options" on page 329.

Figure A.7 The Overlay Plot



Overlay Plot Options

The basic Overlay Plot is shown in Figure A.7.

The Overlay Plot platform has plotting options accessed from the red triangle menu on the Overlay Plot title bar. When you select one of these options at the platform level, it affects all plots in the report if no legend levels are highlighted. If one or more plot legend levels are highlighted, the options affect only those plots. There is also a single-plot options menu for each Y variable, which appears when you highlight a Y variable legend beneath the plot and right-click.

The individual plot options are the same as those in the **Y Options** submenu at the platform level. See “Y Options” on page 332.

General Overlay Platform Options

When you select one of these options at the platform level, it affects all plots in the report if no legend levels are highlighted. If one or more plot legend levels are highlighted, the options affect only those plots.

Overlay Plots Contains options for overlaying:

Overlay Y’s Overlays all variables assigned to the Y role on one plot. This option is on by default and unavailable if only one Y variable is assigned.

Overlay Groups Overlays groups and produces a legend. This option is off by default and unavailable if no grouping variables are assigned. See “Overlay Groups” on page 330.

No Overlay Turns off overlaying for both *Y*s and groups. Creates a separate plot for each Y and each group. This option is off by default unless only one Y variable is assigned and no grouping variables are assigned. In this case, no overlaying options are available.

Separate Axes Assigns each plot its own set of *xy*-axes. If **Separate Axes** is off, the vertical axis is shared across the same row of plots and the horizontal axis is shared on the same column of plots. The default setting is on (except when multiple plots exist). See "Separate Axes" on page 331.

Uniform Y Scale Uses the same Y scale for all grouped plots. The default setting is off.

Connect Thru Missing Connects adjacent points in the plot, regardless of missing values. The default setting is off.

Range Plot Connects the lowest and highest points at each *x* value with a line with bars at each end. The **Needle** and **Range Plot** options are mutually exclusive.

Y Options Contains options for the Y variables. See "Y Options" on page 332.

Ungroup Plots Creates a separate chart for each level of a grouping variable.

Arrange Plots Enables you to specify the number of plots in each row.

See the JMP Reports chapter in *Using JMP* for more information about the following options:

Local Data Filter Shows or hides the local data filter that enables you to filter the data used in a specific report.

Redo Contains options that enable you to repeat or relaunch the analysis. In platforms that support the feature, the Automatic Recalc option immediately reflects the changes that you make to the data table in the corresponding report window.

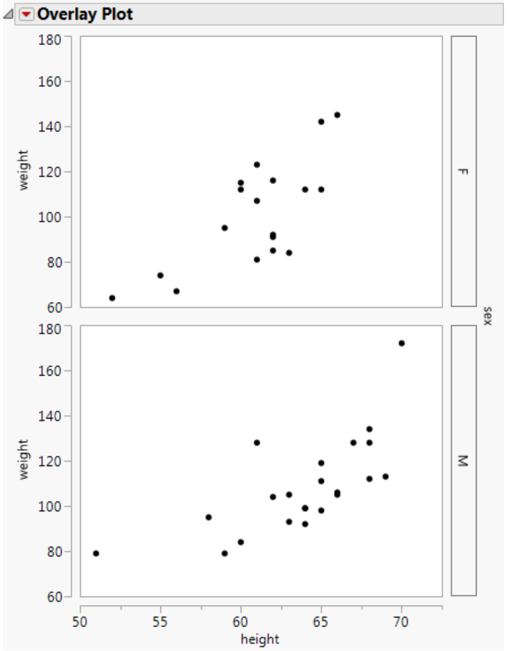
Save Script Contains options that enable you to save a script that reproduces the report to several destinations.

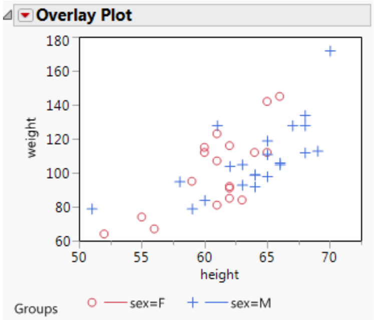
Save By-Group Script Contains options that enable you to save a script that reproduces the platform report for all levels of a By variable to several destinations. Available only when a By variable is specified in the launch window.

Overlay Groups

Figure A.8 shows the effect that the **Overlay Groups** option has on an overlay plot with one Y variable, one X variable, and a grouping variable. The grouping variable has two levels. The plot on the left has **Overlay Groups** turned off, so a separate graph is produced for the two levels of the grouping variable. The plot on the right has **Overlay Groups** turned on, so there is a single graph that uses colors and markers to show the two levels of the grouping variable. A legend describing the levels is added under the graph.

Figure A.8 Overlay Groups: Off (left) and On (right)

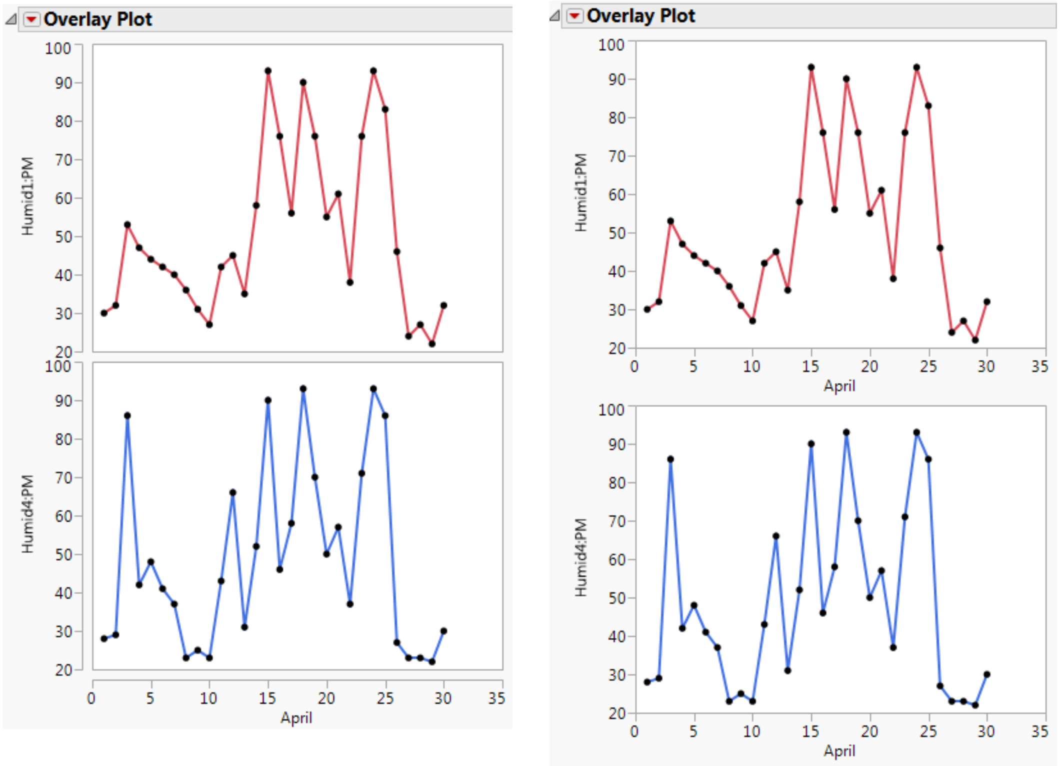




Separate Axes

Figure A.9 shows the effect that the **Separate Axes** option has on an overlay plot with two Y variables and one X variable. The **Overlay Y's** option is turned off, so a separate plot is produced for each Y variable. The plot on the left has **Separate Axes** turned off, so the two graphs share a single *x*-axis. The plot on the right has **Separate Axes** turned on, so both graphs have their own *x*-axis.

Figure A.9 Separate Axes: Off (left) and On (right)



Y Options

Each Y variable is labeled in a legend beneath the plot. The Y options are available from the **Y Options** menu from the Overlay Plot red triangle menu. You can also access the **Y Options** menu by right-clicking on any Y variable in the legend.

Note: If no Y variables are selected, any Y options that you select affect all Y variables. If one or more of the Y variables are selected, any Y options that you select affect only those you have selected.

Selecting and Deselecting Y Variables in the Legend

- Hold the Shift key and click to select multiple contiguous legend levels.
- Hold the Ctrl key and click to select multiple discontiguous legend levels.
- Hold the Ctrl key and click a selected legend level to deselect it.

Show Points A toggle that either shows or hides points in the graph.

Connect Points A toggle that either connects the points with lines or turns the connecting lines off. You can use **Connect Points** without showing points.

Needle Draws a vertical line from each point to the *x*-axis.

Step Draws a horizontal line from each point to the *x* value of the following point, and then a vertical line to that point. You can use **Step** without showing points.

Note: The **Connect Points**, **Needle**, and **Step** options are mutually exclusive.

Function Plot Plots a formula (stored in the Y column) as a smooth curve. To use this function, store a formula in a column that is a function of a single X column. Assign the formula to the Y role.

Connect Color Displays the JMP color palette for assigning colors to connecting lines.

Overlay Marker Displays the JMP marker palette for assigning markers to plotted points.

Overlay Marker Color Assigns a color to all points of the selected variable.

Line Style Enables the choice of dashed, dotted, or other line styles.

Line Width Enables the choice of line widths.

Appendix B

References

Aitchison, J. (1986). *The Statistical Analysis of Compositional Data*. London: Chapman & Hall.

Box, G. E. P., Hunter, W. G., and Hunter, J. S. (1978). *Statistics for Experimenters*. New York: John Wiley & Sons.

Cornell, J. A. (1990). *Experiments with Mixtures*. 2nd ed. New York: John Wiley & Sons.

Hyndman, R. J. (1996). "Computing and Graphing High Density Regions."*The American Statistician* 50: 120 - 126.

Inselberg, A. (1985). "The Plane with Parallel Coordinates." *Visual Computing* 1: 69–91.

Mardia, K. V., Kent, J. T., and Bibby J. M. (1979). *Multivariate Analysis*. New York: Academic Press.

Rousseeuw, P. J., Ruts, I., and Tukey, J. W. (1999). "The Bagplot: A Bivariate Boxplot." *The American Statistician* 53: 382 - 387.

Wegman, E. J. (1990). "Hyperdimensional Data Analysis using Parallel Coordinates." *Journal of the American Statistical Association* 85: 664–675.

Appendix C

Technology License Notices

- Scintilla - Copyright © 1998-2017 by Neil Hodgson <neilh@scintilla.org>.

  All Rights Reserved.

  Permission to use, copy, modify, and distribute this software and its documentation for any purpose and without fee is hereby granted, provided that the above copyright notice appear in all copies and that both that copyright notice and this permission notice appear in supporting documentation.

  NEIL HODGSON DISCLAIMS ALL WARRANTIES WITH REGARD TO THIS SOFTWARE, INCLUDING ALL IMPLIED WARRANTIES OF MERCHANTABILITY AND FITNESS, IN NO EVENT SHALL NEIL HODGSON BE LIABLE FOR ANY SPECIAL, INDIRECT OR CONSEQUENTIAL DAMAGES OR ANY DAMAGES WHATSOEVER RESULTING FROM LOSS OF USE, DATA OR PROFITS, WHETHER IN AN ACTION OF CONTRACT, NEGLIGENCE OR OTHER TORTIOUS ACTION, ARISING OUT OF OR IN CONNECTION WITH THE USE OR PERFORMANCE OF THIS SOFTWARE.
- Progressë® Telerik® UI for WPF: Copyright © 2008-2019 Progress Software Corporation. All rights reserved. Usage of the included Progress® Telerik® UI for WPF outside of JMP is not permitted.
- ZLIB Compression Library - Copyright © 1995-2005, Jean-Loup Gailly and Mark Adler.
- Made with Natural Earth. Free vector and raster map data @ naturalearthdata.com.
- Packages - Copyright © 2009-2010, Stéphane Sudre (s.sudre.free.fr). All rights reserved.

  Redistribution and use in source and binary forms, with or without modification, are permitted provided that the following conditions are met:

  Redistributions of source code must retain the above copyright notice, this list of conditions and the following disclaimer.

  Redistributions in binary form must reproduce the above copyright notice, this list of conditions and the following disclaimer in the documentation and/or other materials provided with the distribution.

  Neither the name of the WhiteBox nor the names of its contributors may be used to endorse or promote products derived from this software without specific prior written permission.

  THIS SOFTWARE IS PROVIDED BY THE COPYRIGHT HOLDERS AND CONTRIBUTORS "AS IS" AND ANY EXPRESS OR IMPLIED WARRANTIES, INCLUDING, BUT NOT LIMITED TO, THE IMPLIED WARRANTIES OF MERCHANTABILITY AND FITNESS FOR A PARTICULAR PURPOSE ARE DISCLAIMED. IN NO EVENT SHALL THE COPYRIGHT OWNER OR CONTRIBUTORS BE LIABLE FOR ANY DIRECT, INDIRECT, INCIDENTAL, SPECIAL, EXEMPLARY, OR CONSEQUENTIAL DAMAGES

(INCLUDING, BUT NOT LIMITED TO, PROCUREMENT OF SUBSTITUTE GOODS OR SERVICES; LOSS OF USE, DATA, OR PROFITS; OR BUSINESS INTERRUPTION) HOWEVER CAUSED AND ON ANY THEORY OF LIABILITY, WHETHER IN CONTRACT, STRICT LIABILITY, OR TORT (INCLUDING NEGLIGENCE OR OTHERWISE) ARISING IN ANY WAY OUT OF THE USE OF THIS SOFTWARE, EVEN IF ADVISED OF THE POSSIBILITY OF SUCH DAMAGE.

- iODBC software - Copyright © 1995-2006, OpenLink Software Inc and Ke Jin (www.iodbc.org). All rights reserved.

  Redistribution and use in source and binary forms, with or without modification, are permitted provided that the following conditions are met:

  - Redistributions of source code must retain the above copyright notice, this list of conditions and the following disclaimer.
  - Redistributions in binary form must reproduce the above copyright notice, this list of conditions and the following disclaimer in the documentation and/or other materials provided with the distribution.
  - Neither the name of OpenLink Software Inc. nor the names of its contributors may be used to endorse or promote products derived from this software without specific prior written permission.

  THIS SOFTWARE IS PROVIDED BY THE COPYRIGHT HOLDERS AND CONTRIBUTORS “AS IS” AND ANY EXPRESS OR IMPLIED WARRANTIES, INCLUDING, BUT NOT LIMITED TO, THE IMPLIED WARRANTIES OF MERCHANTABILITY AND FITNESS FOR A PARTICULAR PURPOSE ARE DISCLAIMED. IN NO EVENT SHALL OPENLINK OR CONTRIBUTORS BE LIABLE FOR ANY DIRECT, INDIRECT, INCIDENTAL, SPECIAL, EXEMPLARY, OR CONSEQUENTIAL DAMAGES (INCLUDING, BUT NOT LIMITED TO, PROCUREMENT OF SUBSTITUTE GOODS OR SERVICES; LOSS OF USE, DATA, OR PROFITS; OR BUSINESS INTERRUPTION) HOWEVER CAUSED AND ON ANY THEORY OF LIABILITY, WHETHER IN CONTRACT, STRICT LIABILITY, OR TORT (INCLUDING NEGLIGENCE OR OTHERWISE) ARISING IN ANY WAY OUT OF THE USE OF THIS SOFTWARE, EVEN IF ADVISED OF THE POSSIBILITY OF SUCH DAMAGE.

- bzip2, the associated library “libbzip2”, and all documentation, are Copyright © 1996-2010, Julian R Seward. All rights reserved.

  Redistribution and use in source and binary forms, with or without modification, are permitted provided that the following conditions are met:

  Redistributions of source code must retain the above copyright notice, this list of conditions and the following disclaimer.

  The origin of this software must not be misrepresented; you must not claim that you wrote the original software. If you use this software in a product, an acknowledgment in the product documentation would be appreciated but is not required.

  Altered source versions must be plainly marked as such, and must not be misrepresented as being the original software.

The name of the author may not be used to endorse or promote products derived from this software without specific prior written permission.

THIS SOFTWARE IS PROVIDED BY THE AUTHOR "AS IS" AND ANY EXPRESS OR IMPLIED WARRANTIES, INCLUDING, BUT NOT LIMITED TO, THE IMPLIED WARRANTIES OF MERCHANTABILITY AND FITNESS FOR A PARTICULAR PURPOSE ARE DISCLAIMED. IN NO EVENT SHALL THE AUTHOR BE LIABLE FOR ANY DIRECT, INDIRECT, INCIDENTAL, SPECIAL, EXEMPLARY, OR CONSEQUENTIAL DAMAGES (INCLUDING, BUT NOT LIMITED TO, PROCUREMENT OF SUBSTITUTE GOODS OR SERVICES; LOSS OF USE, DATA, OR PROFITS; OR BUSINESS INTERRUPTION) HOWEVER CAUSED AND ON ANY THEORY OF LIABILITY, WHETHER IN CONTRACT, STRICT LIABILITY, OR TORT (INCLUDING NEGLIGENCE OR OTHERWISE) ARISING IN ANY WAY OUT OF THE USE OF THIS SOFTWARE, EVEN IF ADVISED OF THE POSSIBILITY OF SUCH DAMAGE.

- R software is Copyright © 1999-2012, R Foundation for Statistical Computing.
- MATLAB software is Copyright © 1984-2012, The MathWorks, Inc. Protected by U.S. and international patents. See www.mathworks.com/patents. MATLAB and Simulink are registered trademarks of The MathWorks, Inc. See www.mathworks.com/trademarks for a list of additional trademarks. Other product or brand names may be trademarks or registered trademarks of their respective holders.
- libopc is Copyright © 2011, Florian Reuter. All rights reserved.

  Redistribution and use in source and binary forms, with or without modification, are permitted provided that the following conditions are met:

  - Redistributions of source code must retain the above copyright notice, this list of conditions and the following disclaimer.
  - Redistributions in binary form must reproduce the above copyright notice, this list of conditions and the following disclaimer in the documentation and / or other materials provided with the distribution.
  - Neither the name of Florian Reuter nor the names of its contributors may be used to endorse or promote products derived from this software without specific prior written permission.

  THIS SOFTWARE IS PROVIDED BY THE COPYRIGHT HOLDERS AND CONTRIBUTORS "AS IS" AND ANY EXPRESS OR IMPLIED WARRANTIES, INCLUDING, BUT NOT LIMITED TO, THE IMPLIED WARRANTIES OF MERCHANTABILITY AND FITNESS FOR A PARTICULAR PURPOSE ARE DISCLAIMED. IN NO EVENT SHALL THE COPYRIGHT OWNER OR CONTRIBUTORS BE LIABLE FOR ANY DIRECT, INDIRECT, INCIDENTAL, SPECIAL, EXEMPLARY, OR CONSEQUENTIAL DAMAGES (INCLUDING, BUT NOT LIMITED TO, PROCUREMENT OF SUBSTITUTE GOODS OR SERVICES; LOSS OF USE, DATA, OR PROFITS; OR BUSINESS INTERRUPTION) HOWEVER CAUSED AND ON ANY THEORY OF LIABILITY, WHETHER IN CONTRACT, STRICT LIABILITY, OR TORT (INCLUDING NEGLIGENCE OR OTHERWISE) ARISING IN ANY WAY OUT OF THE USE OF THIS SOFTWARE, EVEN IF ADVISED OF THE POSSIBILITY OF SUCH DAMAGE.

- libxml2 - Except where otherwise noted in the source code (e.g. the files hash.c, list.c and the trio files, which are covered by a similar license but with different Copyright notices) all the files are:

  Copyright © 1998 - 2003 Daniel Veillard. All Rights Reserved.

  Permission is hereby granted, free of charge, to any person obtaining a copy of this software and associated documentation files (the “Software”), to deal in the Software without restriction, including without limitation the rights to use, copy, modify, merge, publish, distribute, sublicense, and/or sell copies of the Software, and to permit persons to whom the Software is furnished to do so, subject to the following conditions:

  The above copyright notice and this permission notice shall be included in all copies or substantial portions of the Software.

  THE SOFTWARE IS PROVIDED “AS IS”, WITHOUT WARRANTY OF ANY KIND, EXPRESS OR IMPLIED, INCLUDING BUT NOT LIMITED TO THE WARRANTIES OF MERCHANTABILITY, FITNESS FOR A PARTICULAR PURPOSE AND NONINFRINGEMENT. IN NO EVENT SHALL DANIEL VEILLARD BE LIABLE FOR ANY CLAIM, DAMAGES OR OTHER LIABILITY, WHETHER IN AN ACTION OF CONTRACT, TORT OR OTHERWISE, ARISING FROM, OUT OF OR IN CONNECTION WITH THE SOFTWARE OR THE USE OR OTHER DEALINGS IN THE SOFTWARE.

  Except as contained in this notice, the name of Daniel Veillard shall not be used in advertising or otherwise to promote the sale, use or other dealings in this Software without prior written authorization from him.

- Regarding the decompression algorithm used for UNIX files:

  Copyright © 1985, 1986, 1992, 1993

  The Regents of the University of California. All rights reserved.

  THIS SOFTWARE IS PROVIDED BY THE REGENTS AND CONTRIBUTORS “AS IS” AND ANY EXPRESS OR IMPLIED WARRANTIES, INCLUDING, BUT NOT LIMITED TO, THE IMPLIED WARRANTIES OF MERCHANTABILITY AND FITNESS FOR A PARTICULAR PURPOSE ARE DISCLAIMED. IN NO EVENT SHALL THE REGENTS OR CONTRIBUTORS BE LIABLE FOR ANY DIRECT, INDIRECT, INCIDENTAL, SPECIAL, EXEMPLARY, OR CONSEQUENTIAL DAMAGES (INCLUDING, BUT NOT LIMITED TO, PROCUREMENT OF SUBSTITUTE GOODS OR SERVICES; LOSS OF USE, DATA, OR PROFITS; OR BUSINESS INTERRUPTION) HOWEVER CAUSED AND ON ANY THEORY OF LIABILITY, WHETHER IN CONTRACT, STRICT LIABILITY, OR TORT (INCLUDING NEGLIGENCE OR OTHERWISE) ARISING IN ANY WAY OUT OF THE USE OF THIS SOFTWARE, EVEN IF ADVISED OF THE POSSIBILITY OF SUCH DAMAGE.

  1. Redistributions of source code must retain the above copyright notice, this list of conditions and the following disclaimer.

  2. Redistributions in binary form must reproduce the above copyright notice, this list of conditions and the following disclaimer in the documentation and/or other materials provided with the distribution.

3. Neither the name of the University nor the names of its contributors may be used to endorse or promote products derived from this software without specific prior written permission.

- Snowball - Copyright © 2001, Dr Martin Porter, Copyright © 2002, Richard Boulton.

  All rights reserved.

  Redistribution and use in source and binary forms, with or without modification, are permitted provided that the following conditions are met:

  1. Redistributions of source code must retain the above copyright notice, this list of conditions and the following disclaimer.

  2. Redistributions in binary form must reproduce the above copyright notice, this list of conditions and the following disclaimer in the documentation and/or other materials provided with the distribution.

  3. Neither the name of the copyright holder nor the names of its contributors may be used to endorse or promote products derived from this software without specific prior written permission.

  THIS SOFTWARE IS PROVIDED BY THE COPYRIGHT HOLDERS AND CONTRIBUTORS \"AS IS\" AND ANY EXPRESS OR IMPLIED WARRANTIES, INCLUDING, BUT NOT LIMITED TO, THE IMPLIED WARRANTIES OF MERCHANTABILITY AND FITNESS FOR A PARTICULAR PURPOSE ARE DISCLAIMED. IN NO EVENT SHALL THE COPYRIGHT HOLDER OR CONTRIBUTORS BE LIABLE FOR ANY DIRECT, INDIRECT, INCIDENTAL, SPECIAL, EXEMPLARY, OR CONSEQUENTIAL DAMAGES (INCLUDING, BUT NOT LIMITED TO, PROCUREMENT OF SUBSTITUTE GOODS OR SERVICES; LOSS OF USE, DATA, OR PROFITS; OR BUSINESS INTERRUPTION) HOWEVER CAUSED AND ON ANY THEORY OF LIABILITY, WHETHER IN CONTRACT, STRICT LIABILITY, OR TORT (INCLUDING NEGLIGENCE OR OTHERWISE) ARISING IN ANY WAY OUT OF THE USE OF THIS SOFTWARE, EVEN IF ADVISED OF THE POSSIBILITY OF SUCH DAMAGE.

- Pako - Copyright © 2014–2017 by Vitaly Puzrin and Andrei Tuputcyn.

  Permission is hereby granted, free of charge, to any person obtaining a copy of this software and associated documentation files (the “Software”), to deal in the Software without restriction, including without limitation the rights to use, copy, modify, merge, publish, distribute, sublicense, and/or sell copies of the Software, and to permit persons to whom the Software is furnished to do so, subject to the following conditions:

  The above copyright notice and this permission notice shall be included in all copies or substantial portions of the Software.

  THE SOFTWARE IS PROVIDED “AS IS”, WITHOUT WARRANTY OF ANY KIND, EXPRESS OR IMPLIED, INCLUDING BUT NOT LIMITED TO THE WARRANTIES OF MERCHANTABILITY, FITNESS FOR A PARTICULAR PURPOSE AND

NONINFRINGEMENT. IN NO EVENT SHALL THE AUTHORS OR COPYRIGHT HOLDERS BE LIABLE FOR ANY CLAIM, DAMAGES OR OTHER LIABILITY, WHETHER IN AN ACTION OF CONTRACT, TORT OR OTHERWISE, ARISING FROM, OUT OF OR IN CONNECTION WITH THE SOFTWARE OR THE USE OR OTHER DEALINGS IN THE SOFTWARE.

- HDF5 (Hierarchical Data Format 5) Software Library and Utilities Copyright 2006 –2015 by The HDF Group. NCSA HDF5 (Hierarchical Data Format 5) Software Library and Utilities Copyright 1998-2006 by the Board of Trustees of the University of Illinois. All rights reserved. DISCLAIMER: THIS SOFTWARE IS PROVIDED BY THE HDF GROUP AND THE CONTRIBUTORS “AS IS” WITH NO WARRANTY OF ANY KIND, EITHER EXPRESSED OR IMPLIED. In no event shall The HDF Group or the Contributors be liable for any damages suffered by the users arising out of the use of this software, even if advised of the possibility of such damage.
- agl-aglfn technology is Copyright © 2002, 2010, 2015 by Adobe Systems Incorporated. All Rights Reserved.

  Redistribution and use in source and binary forms, with or without modification, are permitted provided that the following conditions are met:

  – Redistributions of source code must retain the above copyright notice, this list of conditions and the following disclaimer.

  – Redistributions in binary form must reproduce the above copyright notice, this list of conditions and the following disclaimer in the documentation and/or other materials provided with the distribution.

  – Neither the name of Adobe Systems Incorporated nor the names of its contributors may be used to endorse or promote products derived from this software without specific prior written permission.

  THIS SOFTWARE IS PROVIDED BY THE COPYRIGHT HOLDERS AND CONTRIBUTORS “AS IS” AND ANY EXPRESS OR IMPLIED WARRANTIES, INCLUDING, BUT NOT LIMITED TO, THE IMPLIED WARRANTIES OF MERCHANTABILITY AND FITNESS FOR A PARTICULAR PURPOSE ARE DISCLAIMED. IN NO EVENT SHALL THE COPYRIGHT HOLDER OR CONTRIBUTORS BE LIABLE FOR ANY DIRECT, INDIRECT, INCIDENTAL, SPECIAL, EXEMPLARY, OR CONSEQUENTIAL DAMAGES (INCLUDING, BUT NOT LIMITED TO, PROCUREMENT OF SUBSTITUTE GOODS OR SERVICES; LOSS OF USE, DATA, OR PROFITS; OR BUSINESS INTERRUPTION) HOWEVER CAUSED AND ON ANY THEORY OF LIABILITY, WHETHER IN CONTRACT, STRICT LIABILITY, OR TORT (INCLUDING NEGLIGENCE OR OTHERWISE) ARISING IN ANY WAY OUT OF THE USE OF THIS SOFTWARE, EVEN IF ADVISED OF THE POSSIBILITY OF SUCH DAMAGE.
- dmlc/xgboost is Copyright © 2019 SAS Institute.

Licensed under the Apache License, Version 2.0 (the "License"); you may not use this file except in compliance with the License. You may obtain a copy of the License at

http://www.apache.org/licenses/LICENSE-2.0

Unless required by applicable law or agreed to in writing, software distributed under the License is distributed on an "AS IS" BASIS, WITHOUT WARRANTIES OR CONDITIONS OF ANY KIND, either express or implied. See the License for the specific language governing permissions and limitations under the License.