Statistical Discovery.™ From SAS.

Version 16

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

16.1

The correct bibliographic citation for this manual is as follows: SAS Institute Inc. 2020–2021. *JMP® 16 Essential Graphing*. Cary, NC: SAS Institute Inc.

JMP® 16 Essential Graphing

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

March 2021

August 2021

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
- the JMP user community, resources for users including examples of add-ins and scripts, a forum, blogs, conference information, and so on

https://www.jmp.com/getstarted

Contents

Essential Graphing

Chapter 1

Learn about JMP

Documentation and Additional Resources

Learn about JMP documentation, such as book conventions, descriptions of each JMP document, the Help system, and where to find additional support.

Contents

Formatting Conventions in JMP Documentation

These 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. 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 **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" 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. You can also download the Documentation PDF Files add-in if you prefer searching individual PDF files of each document in the JMP library. Download the available add-ins from https://community.jmp.com.

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 • Model Screening • 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 • Legacy Control Charts • Pareto Plot • Diagram • Manage Spec Limits • OC Curves

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:

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

JMP 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\16\Samples\Data

On macOS: \Library\Application Support\JMP\16\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

For help with statistical terms, select Help > Statistics Index. For help with JSL scripting and examples, select **Help > Scripting Index**.

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.

JMP Tooltips

JMP provides descriptive tooltips (or *hover labels*) when you hover 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.

JMP New User Welcome Kit

The JMP 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**.

JMP 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 “Graph Builder” and “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 “Bubble Plots”.
- Scatterplot Matrix shows an ordered collection of bivariate graphs. See “Scatterplot Matrix”.
- Parallel Plot draws connected line segments that represent each row in a data table. See “Parallel Plots”.
- Cell Plot draws a rectangular array of cells where each cell corresponds to a data table entry. See “Cell Plots”.
- 3D Scatterplot shows the values of numeric columns in the associated data table in a rotatable, three-dimensional view. See “Scatterplot 3D”.
- Contour Plot constructs contours of a response in a rectangular coordinate system. See “Contour Plots”.
- Ternary Plot display the distribution and variability of three-part compositional data. See “Ternary Plots”.
- Maps can be used in Graph Builder, but also in other platforms, as background maps. See “Maps”.
- Treemaps, charts, and overlay plots are deprecated platforms and can be created using Graph Builder. See “Legacy Platforms”.

Chapter 3

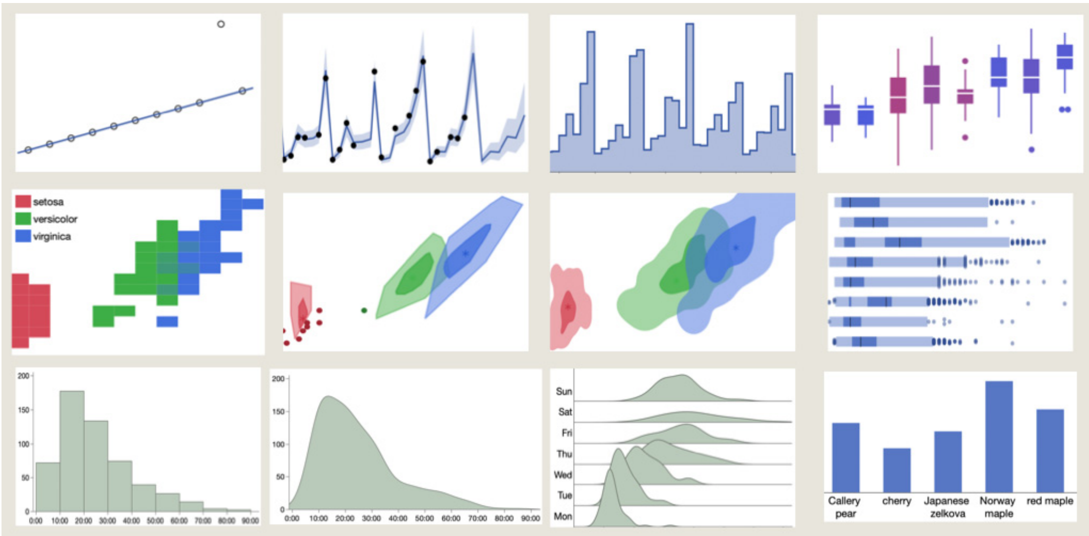
Graph Builder

Interactive Data Visualization

Use Graph Builder to interactively explore your data. You can quickly create and experiment with plots until you find the one you want. Then, you can share your results with others.

This chapter shows you how to use Graph Builder. For detailed examples using Graph Builder, see “Graph Builder Examples”.

Figure 3.1 Graphs Created in Graph Builder



Contents

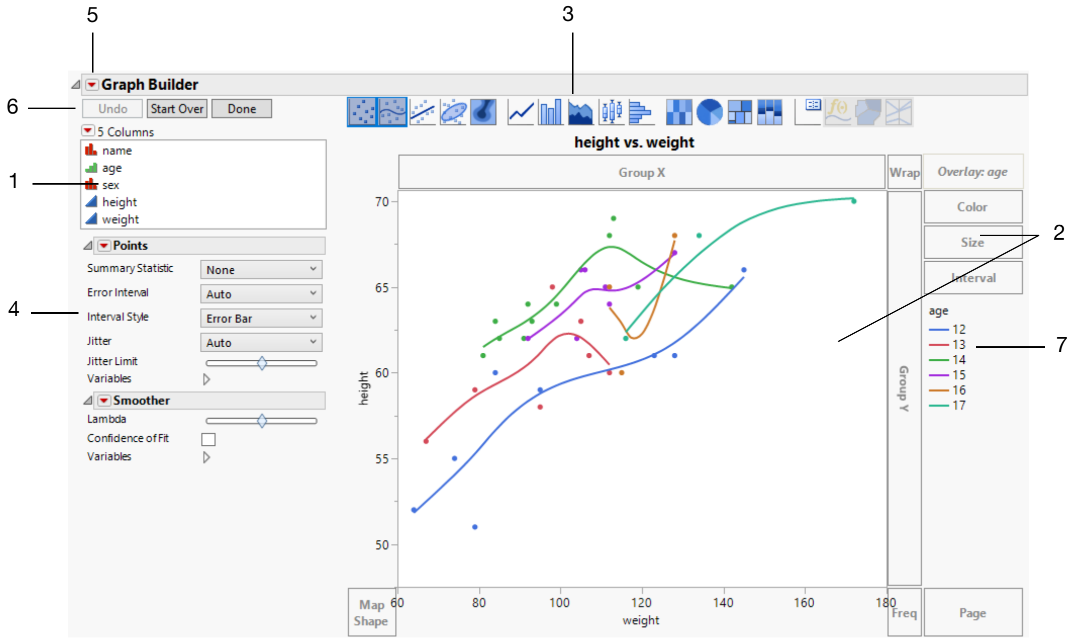
How to Use Graph Builder

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" and "Graph Zones".
4. Click element type icons to choose different types of graphs or elements. See "Element Types and Options".
5. (Optional) Customize the selected element types. See "Element Types and Options".
6. (Optional) Customize the legend. See "Legend Options".
7. When you are 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



Number	Description
1	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 *Using JMP*.
2	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". **Note:** Any excluded rows are hidden in graphs.
3	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".
4	Customize the graph with specific options for the type of graph you have selected. See "Element Types and Options".
5	Customize the Graph Builder window. See "Red Triangle Options".
6	Undo actions, start over, and more. See "Graph Builder Buttons".
7	If your graph has a legend, it appears here. To customize the legend, see "Legend Options".

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

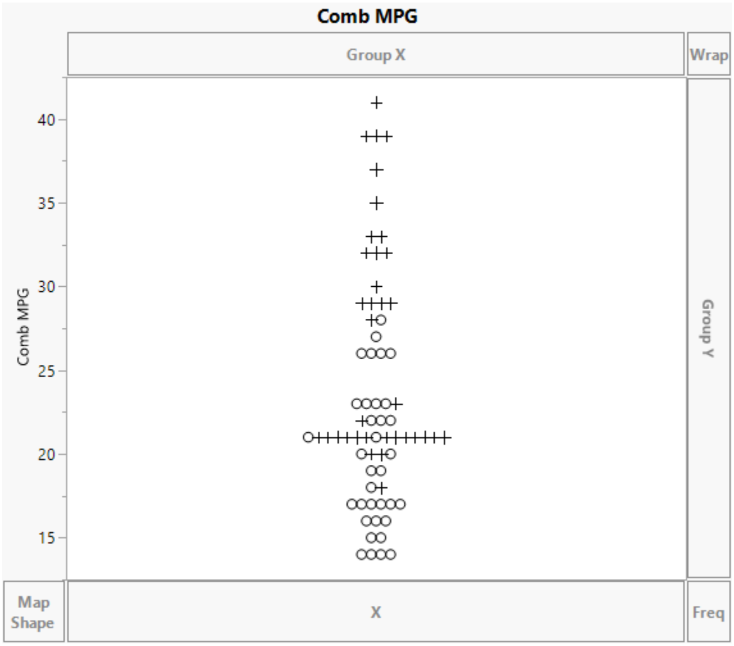
The following example gets you started using some of the features in Graph Builder. This example uses fuel economy data collected for hybrid and non-hybrid cars. You want to get an understanding of miles per gallon (MPG), a measure of fuel economy, and to see which factors might be influencing MPG. This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

- “Get an Overall Picture of Combined Miles per Gallon”
- “Plot Mileage Stratified by Engine and Driving Type”
- “Find the Relationship between Hwy and City MPG by Engine Type”
- “Find the Relationship between Hwy and City MPG by Engine Type”

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.

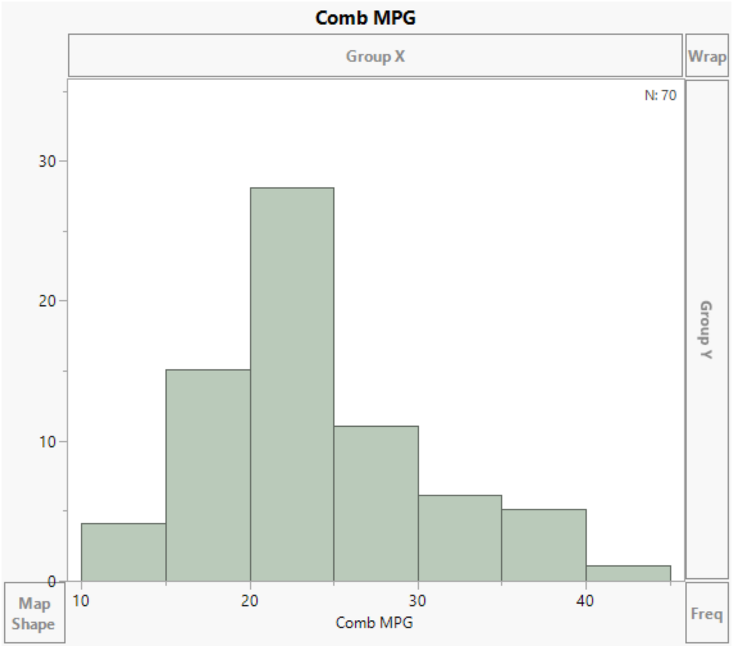
You can get a cleaner picture of the distribution by plotting a histogram.

4. Click the Histogram icon .

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

5. Hover 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



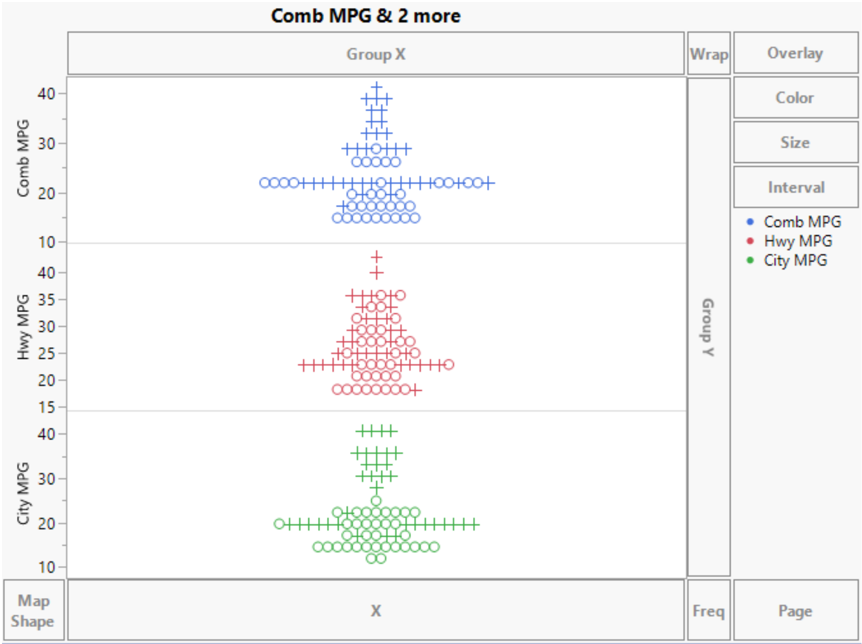
The histogram shows that Comb MPG can range widely, from about 10 to 45. However, a lot of vehicles get mileage between 15 and 30. The mean MPG is 23.3286.

Plot Mileage Stratified by Engine and Driving Type

1. Click **Start Over**.
2. Select City MPG, Hwy MPG, and Comb MPG.
3. Hover over the Y zone, hold down the Shift key 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



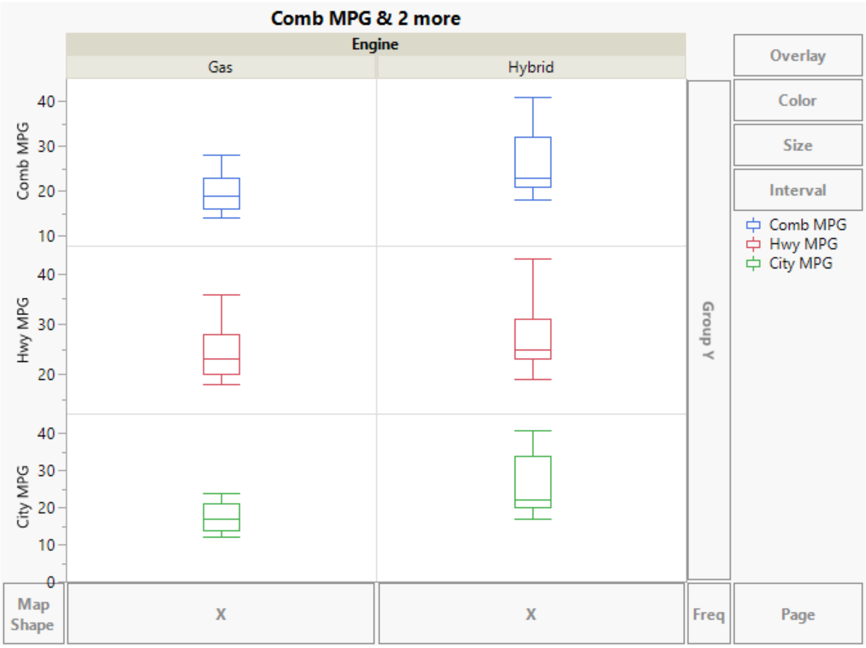
In the legend at right, markers are colored according to the driving type.

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  .

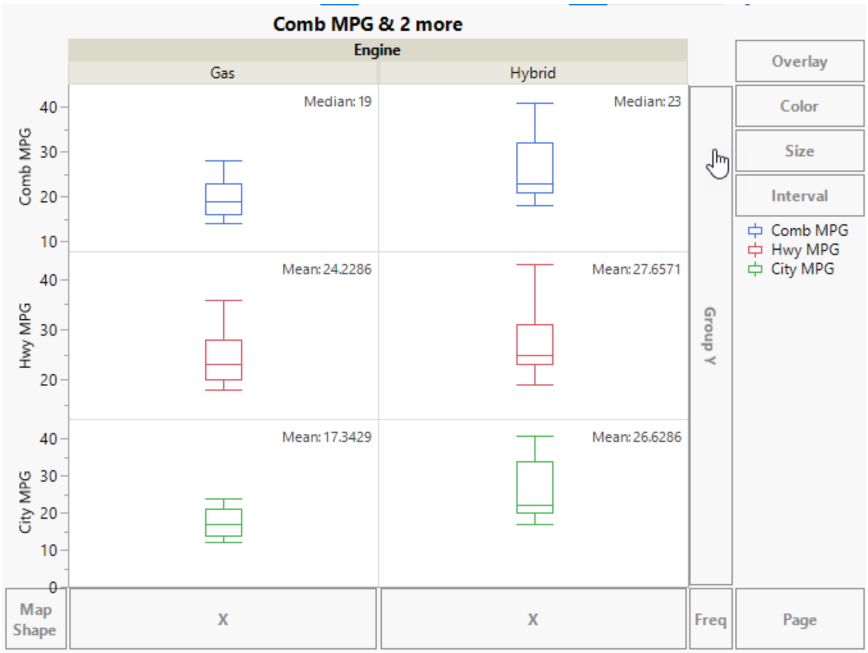
Figure 3.7 Box Plots of MPG Variables by Engine Type



The box plots 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 box plots, 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 Box Plots of MPG Variables with Captions

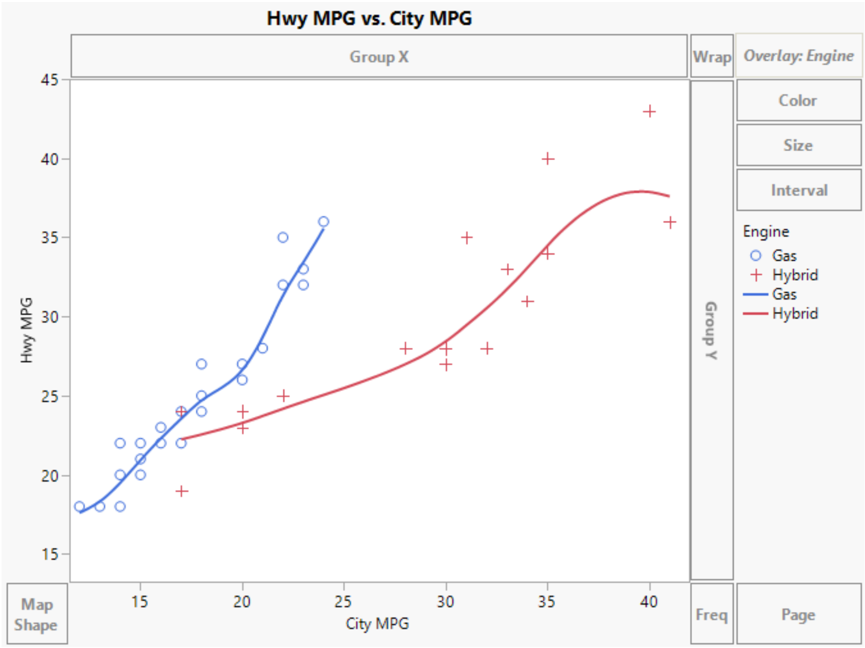


The median city MPG for gas engines is 17 and for hybrid engines is 22.

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**.

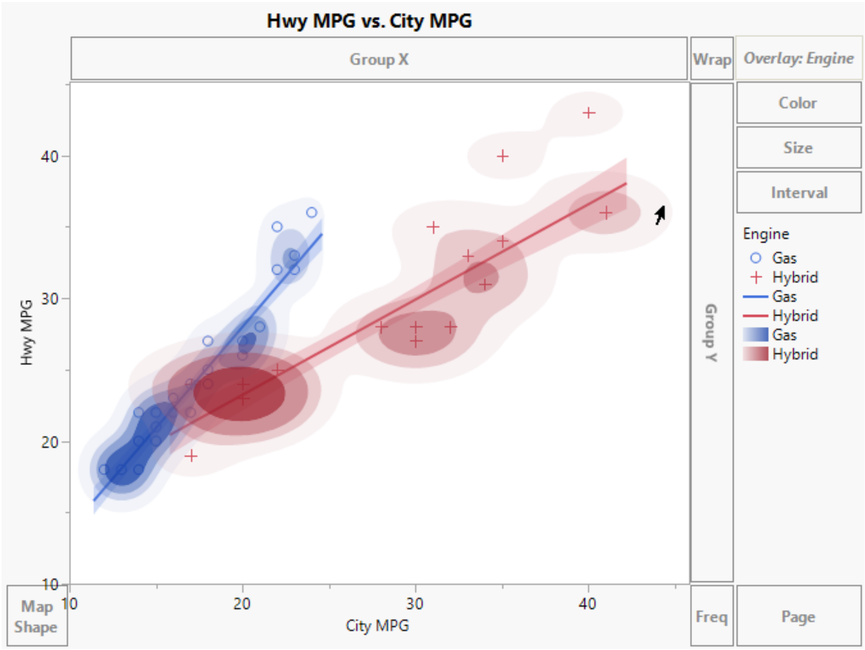
6. Drag a rectangle around the “o” marker with the smallest value of City MPG and check the number of rows selected in the Rows panel of the data table.

   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".

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

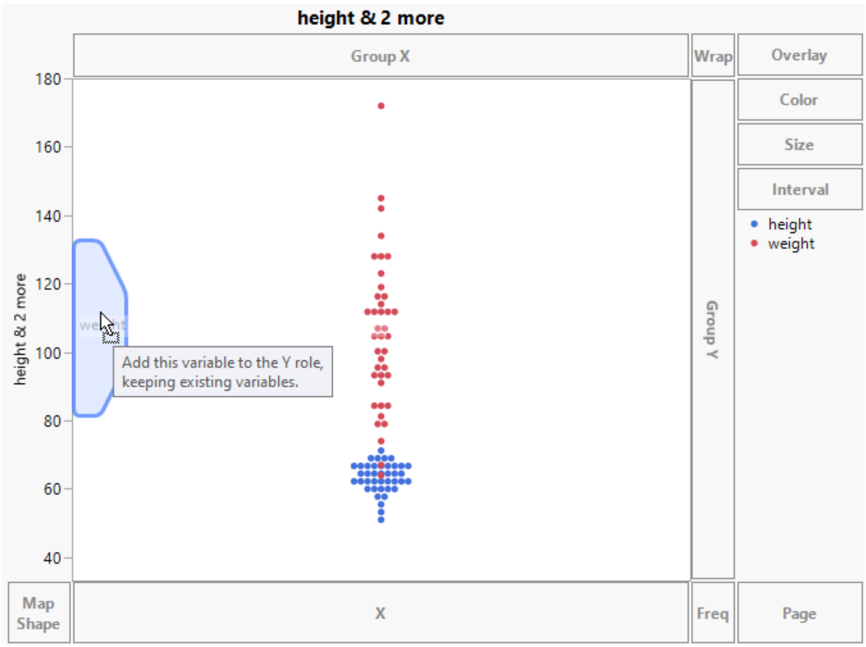
- "Merge Variables on a Common Axis"
- "Create Separate Axes in the X and Y Zones"
- "Create a Second Y Axis"
- "Create Nested Axes for Character Variables"
- "Order the Levels of a Categorical Variable"

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".

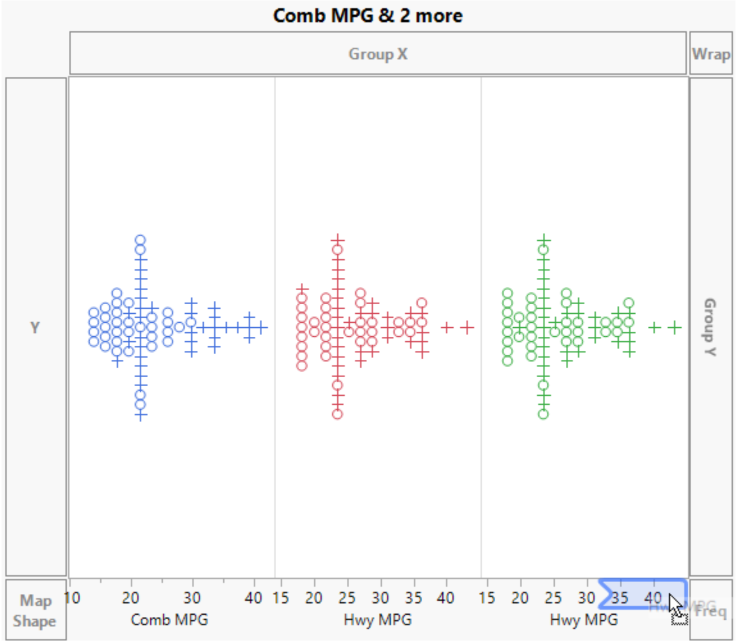
Create Separate Axes in the X and Y Zones

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



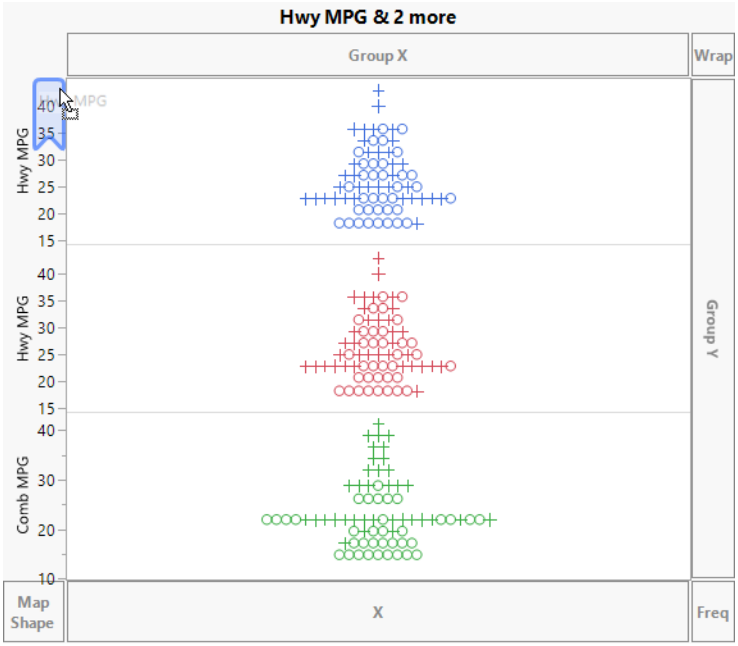
A single polygonal shape indicates when your variable is in the appropriate target zone.

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 Few (2008).

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

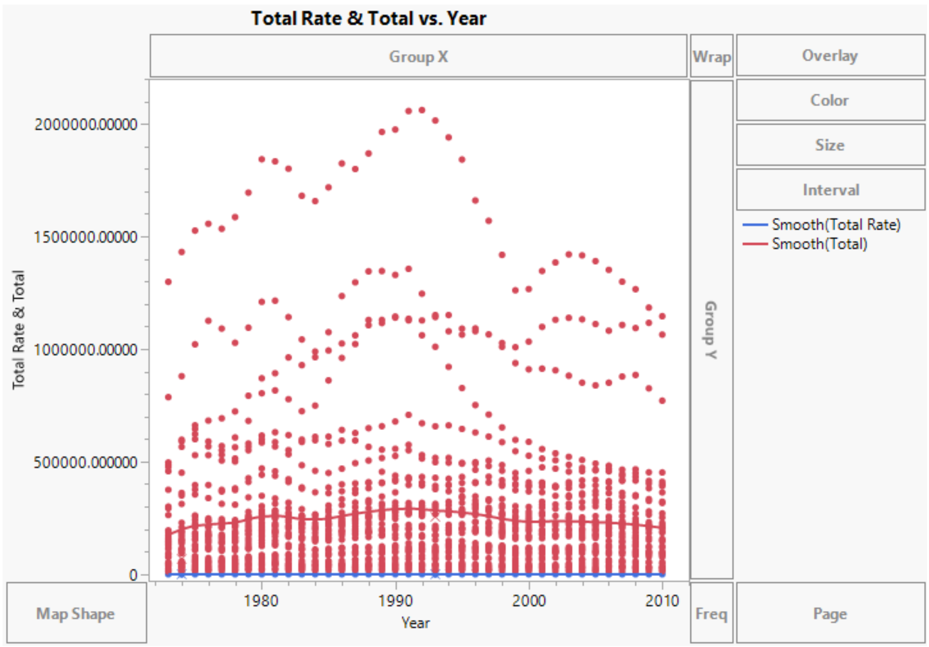
When you click an element type, it applies only to the variables on the left Y axis. To apply an element to the variables on the right Y axis, right-click and change the second element, which corresponds to the right axis.

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

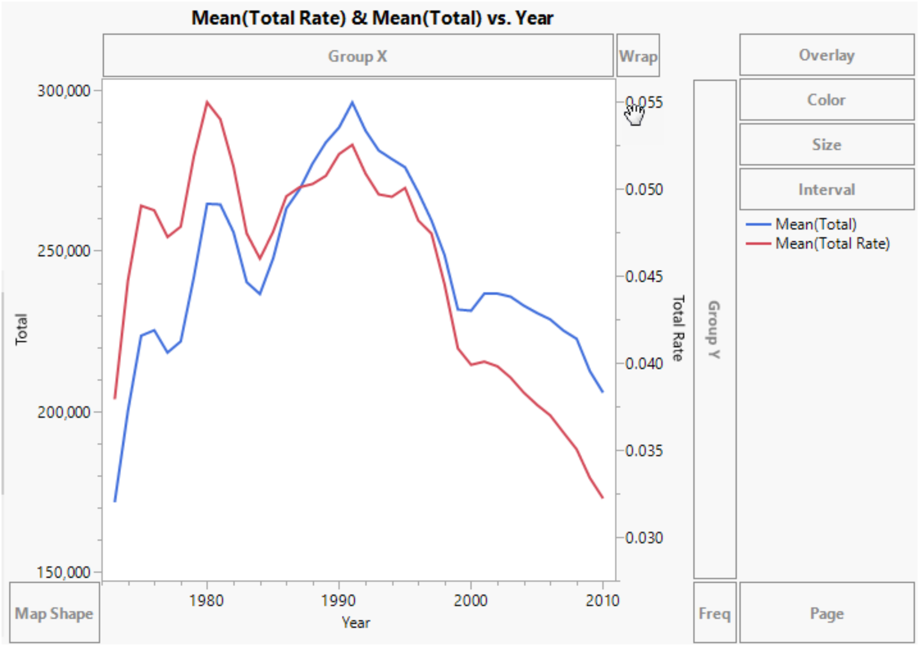


The Total Rate values are all between 0 and 1, and they are barely visible. Calculate yearly means and then use a second Y axis to make these values visible.

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.

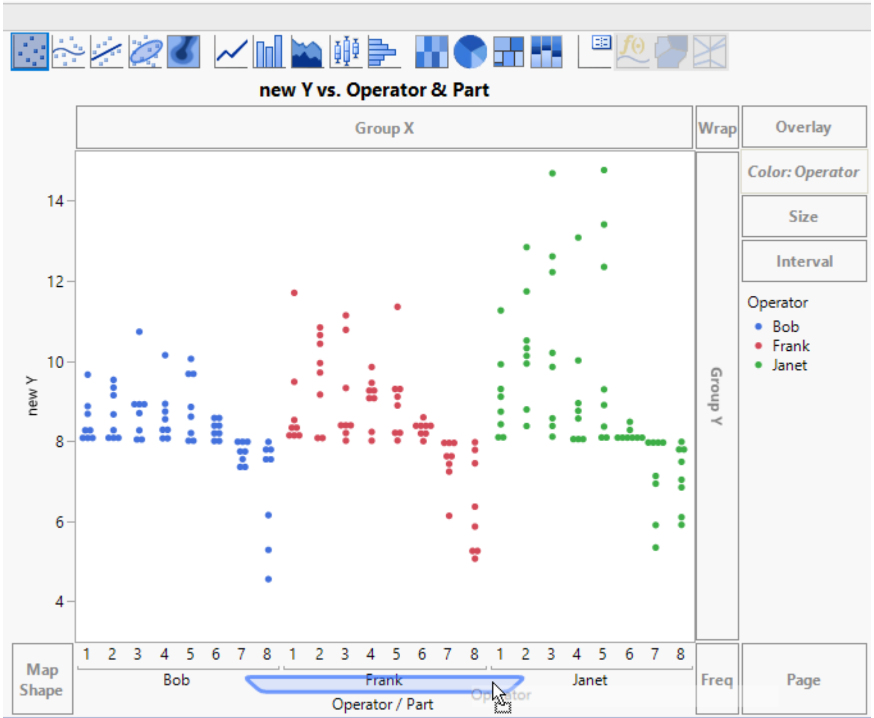
1. Select **Help > Sample Data Library** and open Variability Data/3 Factors Crossed.jmp.
2. Select **Graph > Graph Builder**.
3. Select new Y and drag it to the **Y** zone.
4. Select Part and drag it to the X zone.

   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

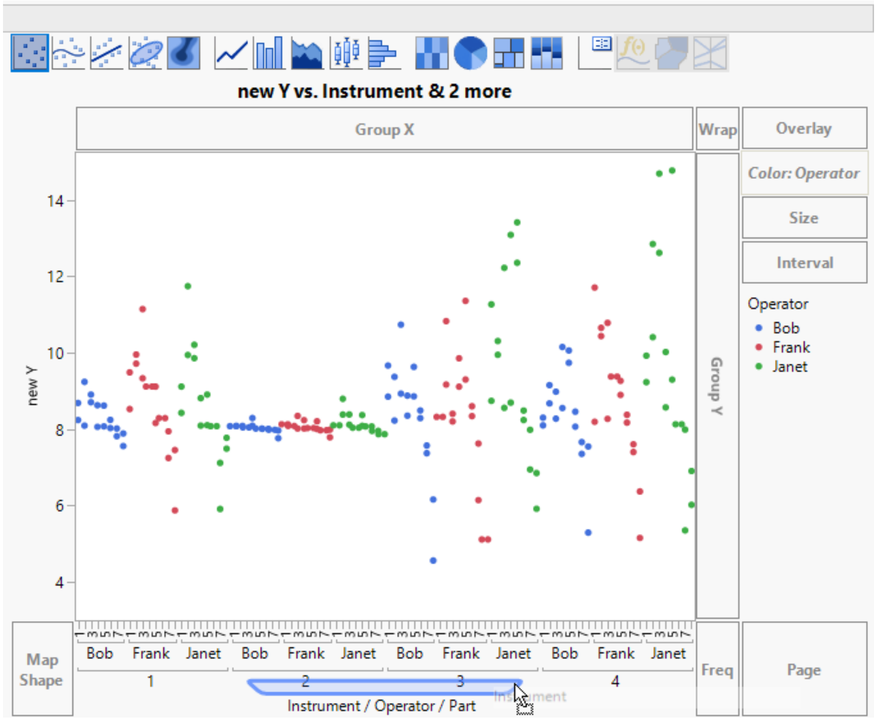


Now it is easier to see that Janet tends to obtain higher measurements for the same parts than do Bob and Frank. But what about the effect of Instrument?

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.

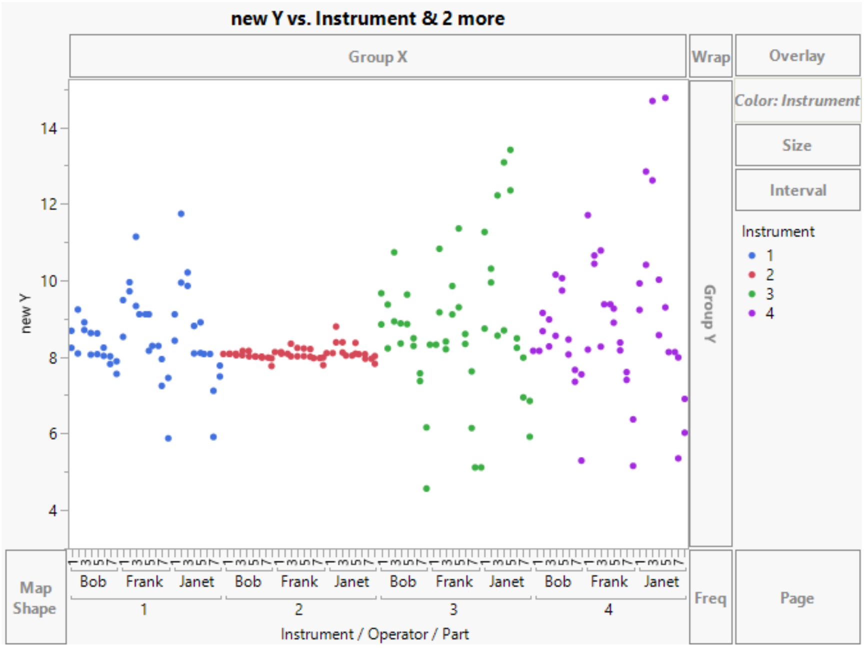
Figure 3.17 Drag Instrument to Add a Third Axis



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.

To change the order of the levels of a categorical variable on a graph axis, click and hold a level on the axis to enter edit mode. In the edit mode, denoted by boxes around each label, you can drag one or more labels (using shift-click for multiple selection) to rearrange them. As you rearrange labels, the value order column property is updated. You can also edit labels, which updates the value label column property. Press Esc to exit the edit mode.

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 *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”.

Change the order of the levels or the statistic:

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

   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”.

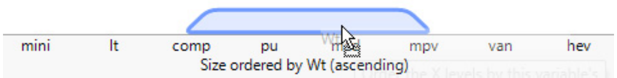
Example of Ordering the Levels of a Categorical Variable Using a Numeric Variable

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.

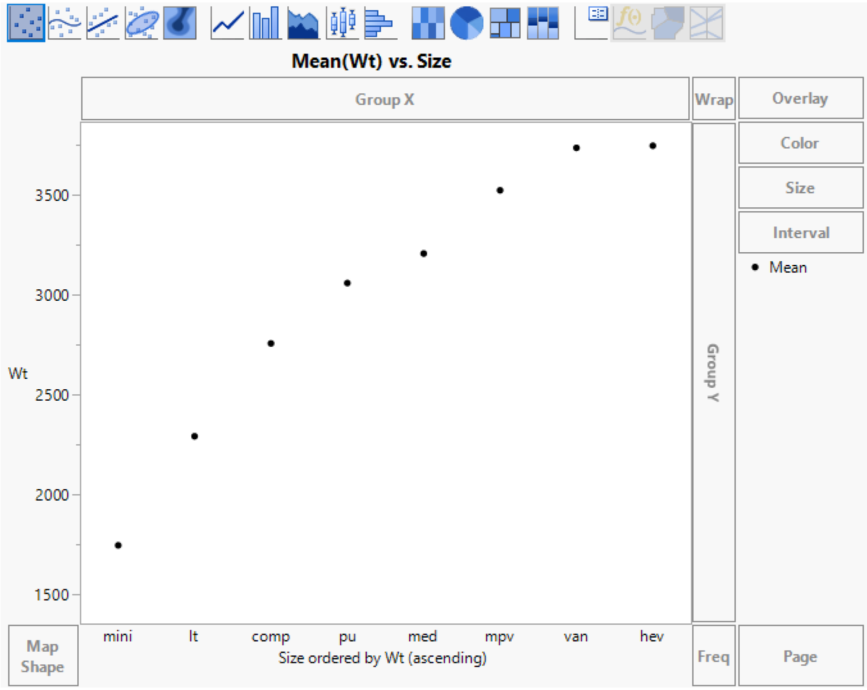
Figure 3.19 Merging Wt and Size



   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.

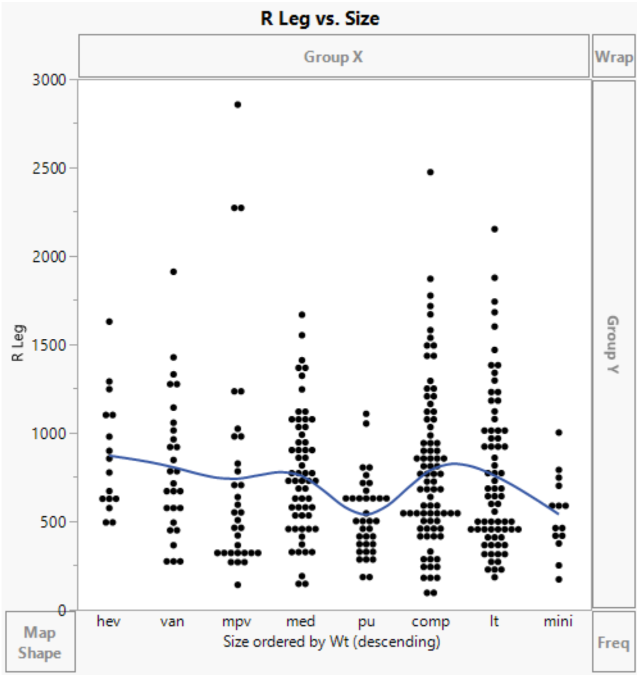
9. Select R Leg and drag it to the Y zone.

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



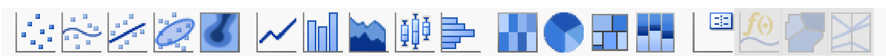
R Leg seems unrelated to the weight classes.

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

You can change how your data appear by clicking an element type icon, such as a bar chart, line, or histogram.

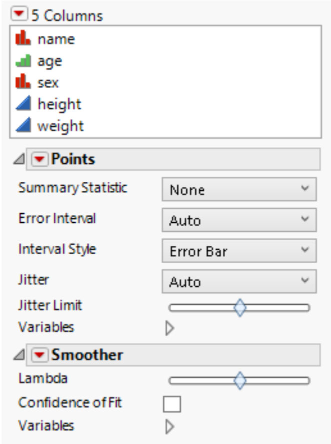
Figure 3.22 Element Type Icons



Note: Supported element types vary depending on the variable type and zones that are selected. Element types that are not applicable appear dimmed.

For each element that you add to a graph, an outline appears in the Properties area at left. This is where you specify and change properties for each element type.

Figure 3.23 Properties Areas for Points and Smoother

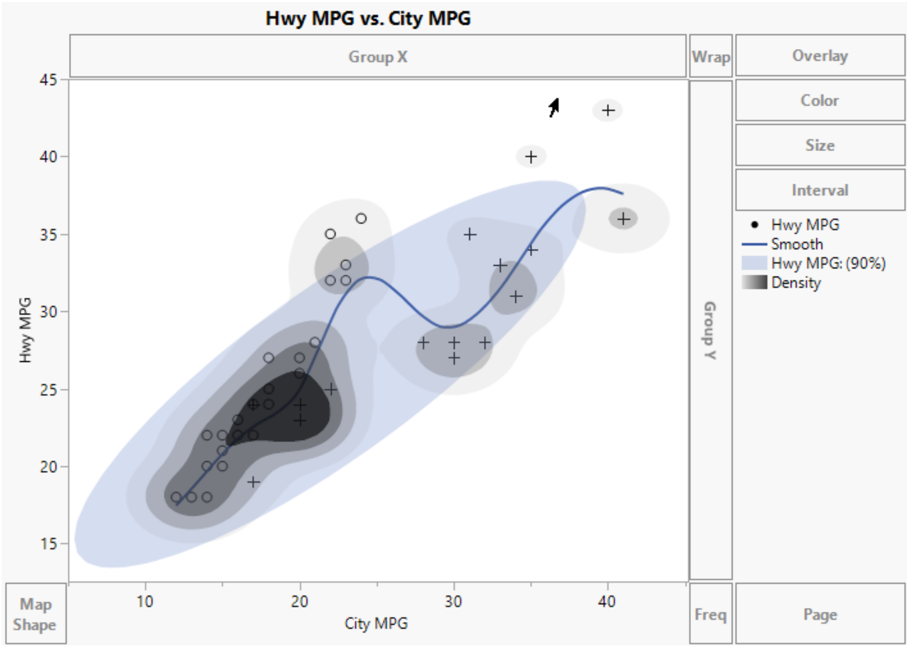


You can also right-click in any graph to change the element, or customize elements in the graph.

Select Multiple Elements at Once

To select and overlay multiple elements, hold down the Shift key 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 hover over a point, a label appears showing information about the corresponding row. Hover 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 *Using JMP*.

Hidden and Excluded Rows

Hide and exclude are two different row states:

- The goal of a hidden row is to impact visibility but not analytical results.
- The goal of an excluded row is to impact the analytical result.

For graphs, an analytical value is usually part of the appearance. Therefore, excluding rows typically impacts the graph appearance and any analytical result included as part of the graph. Hiding a row typically impacts only the graph appearance.

For example, consider a graph with points and a line of fit:

- If rows are hidden, those rows are included in the calculations to obtain the line of fit but they are not plotted.
- If rows are excluded, those rows are excluded from both the calculations and the display.

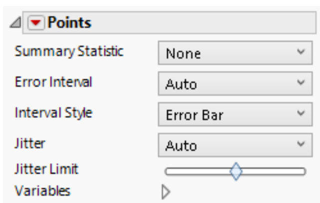
In the two scenarios, the number of points that appear is the same, but the line of fit is different.

When graphing groups, if all rows in a group are hidden, then the visual for that group is also hidden.

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.

Density Random For one-dimensional jitter, this option places markers randomly within the bounds of a violin plot. For two-dimensional jitter, this option places markers randomly within an oval.

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.

For an example using Variables, see "Example of an Area and Line Chart".

Red Triangle Options for Points

Response Axis Changes the axis that is associated with the response variable to X (horizontal), Y (vertical), or automatically determines the response direction.

Set Shape Expression Enables the use of a JSL expression to define a marker shape.

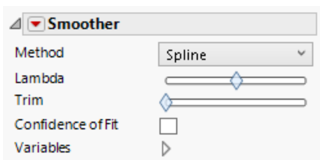
Set Shape Column Sets a column to use as the marker shape.

Smoother

The **Smoother** element shows a smooth curve through the data.

Tip: To label a smoother, right-click on the smoother element in the legend.

Figure 3.26 Smoother Options



Method Specifies the method used for generating the smooth fit. Options include Spline, Local Kernel, Savitzky-Gloay, Moving Average, and Moving Box.

Lambda (Available for Spline.) Enables you to adjust the value of lambda. The default value is 0.05. For more information about lambda, see *Basic Analysis*.

Degree (Available for Savitzky-Golay.) Specifies the degree of the smoother.

Local Width (Not available for Spline.) Enables you to control the smoother fit in terms of a local to a global fit. Use the red triangle Local Region option for further control.

Trim Enables you to trim the sections of the smoother that are based on incomplete or imbalanced data ranges. The slider is scaled from 0 to 1. At 1, all of the incomplete or imbalanced regions are trimmed with the exception of the spline where 1 corresponds to the penultimate edge knots.

Confidence of Fit Shows or hides 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.

For an example using Variables, see "Example of an Area and Line Chart".

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).

Local Region Specifies options for controlling the region used for the local width option.

Local Weighting Specifies weighting options.

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

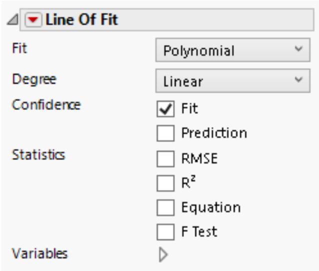
Save Formula Saves fit formulas and upper and lower confidence intervals for a single variable to the data table.

Line of Fit

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

Tip: To label a line of fit, right-click on the line element in the legend.

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. The smoothing model is selected from a subset of state space smoothing models defined by Hyndman et al. (2008). See *Predictive and Specialized Modeling*.

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.

For an example using Variables, see “Example of an Area and Line Chart”.

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.

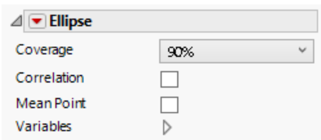
Adapt to Axis Scale For log and other axis transformations, applies computations on the transformed coordinates.

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.

For an example using Variables, see "Example of an Area and Line Chart".

Red Triangle Option for Ellipse

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

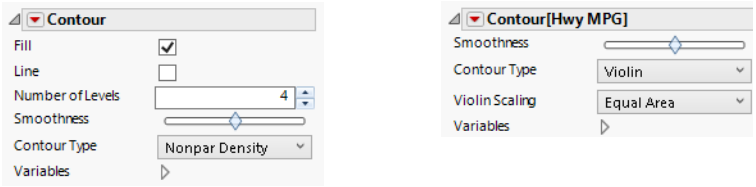
Contour

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.

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 might be undesirable.

Smoothness (Not available for Bagplots.) Smooths the boundaries of the contour plots. The smoothness value is normalized between -1 and 1. This value can be interpreted as a smoothing kernel radius. The original data are 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 (Available only for bivariate plots.) A smooth bivariate nonparametric density surface.

Bagplot Draws a Bagplot, also known as a bivariate box plot. 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 highest density region rectangles for univariate and contours for bivariate data. The lighter shaded region represents the 99% probability region; the darker shaded region represents the 50% probability region. (Note that the regions can be noncontiguous and may not contain exactly 50% or 99% of the input data.) The density mode within the 50% probability region is represented by a line (univariate data) or an asterisk (bivariate data). If you remove points from the plot, the points that remain represent outliers relative to the 99% probability region. You can adjust the smoothness of the regions using the Smoothness option.

The HDR regions are based on a nonparametric density estimated by applying a Gaussian kernel to the data after the points have been interpolated to a grid. The default bandwidth for determining the smoothness of the HDR regions is computed as follows:

$$\frac{0.9s}{\sqrt[5]{n}}$$

The resulting equally spaced bins are used to compute the density regions. The contours are evenly spaced within the density regions that separate the space into equal-probability regions. For more information about highest density regions, see Hyndman (1996).

Violin (Available only for univariate plots.) Draws a violin plot of the density of the data by plotting symmetric kernel densities around a common vertical axis. The densities are defined using the same procedure as the HDR regions.

Jitter Turns jitter on or off. Jitter adds random noise to data, and can help reduce over-plotting. For descriptions of the specific jitter options, see “Points”.

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.

For an example using Variables, see "Example of an Area and Line Chart".

Red Triangle Options for Contour

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

Contour Placement (Available with grouping variables.) Enables you to align or offset the contours.

Line

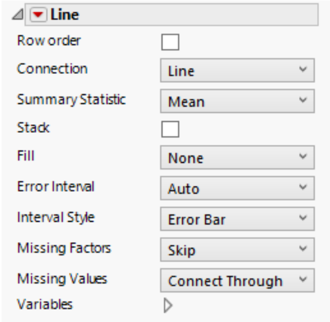
The **Line** element  uses line segments to connect summary statistic values (or raw values) of the variable plotted on the Response Axis across the values or categories of another variable.

For an example that combines a line chart with an area plot to emphasize a line, see "Example of an Area and Line Chart".

Notes:

- 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.
- To label a line, right-click on the line element in the legend.

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.

For an example using Variables, see “Example of an Area and Line Chart”.

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

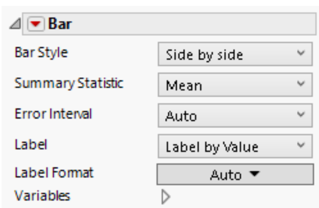
If the variable on the Response Axis has the Character data type, then this variable is assigned numeric values:

- 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 an Overlaid Histogram and Ridgeline Chart” and “Example of a Packed Bar Chart”.

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 and 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 in the bar chart. You can label by value, percent of total value, or by row. When **Label by Value** is selected the values are determined by the summary statistic setting. Click and drag to reposition a label.

Label Format (Available for value and percent labels.) Enables you to set the format for the cell labels.

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.

For an example using Variables, see "Example of an Area and Line Chart".

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.

Overlap Specifies the amount of overlap (Auto, None, Half, Full) between multiple bars.

Table 3.1 Bar Styles and Descriptions

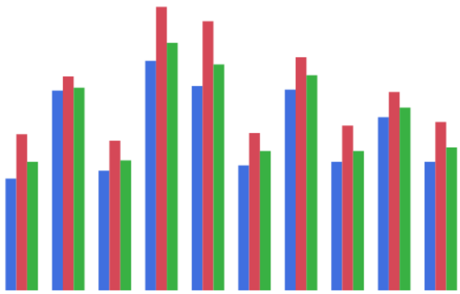
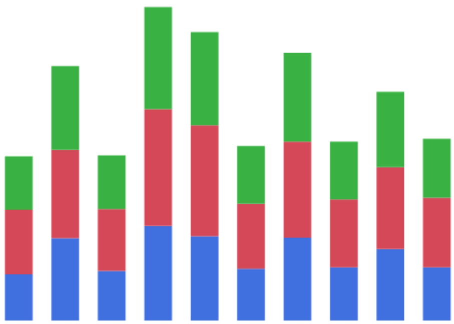
Bar Style	**Description**
Side by side	When you have two or more merged variables on the response axis, a bar is plotted for each of the merged variables. The bars are arranged next to each other for each level of the stratifying variables. They are colored and a legend appears. 
Stacked	When you have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

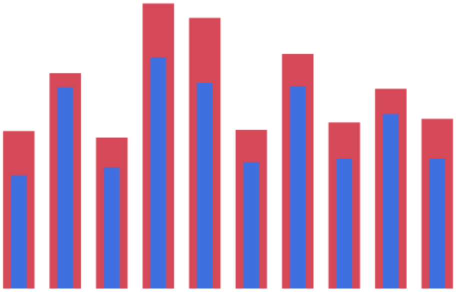
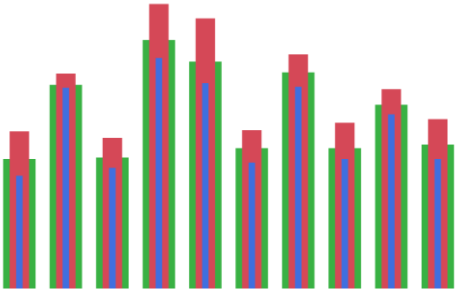
Bar Style	Description
Bullet	When you have two or more merged variables on the response axis, a narrow bar is plotted within a wider and taller bar. The widths of the bars depend on the order of the variables. The part of the bar corresponding to each merged variable is colored and a legend appears. 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)*

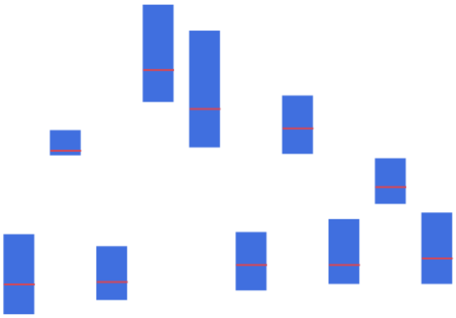
Bar Style	**Description**
Range	When you have two or more merged variables on the response axis, a rectangle and a line are plotted for each level of the stratifying variables. For each level of the stratifying variables: • The values of the summary statistic are calculated for each of the merged variables. • 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)*

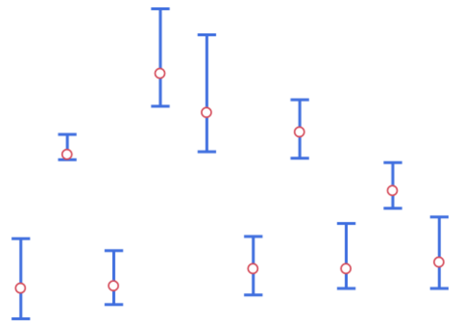
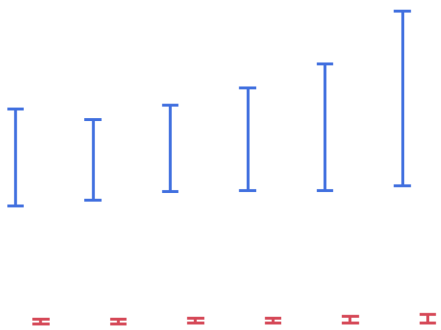
Bar Style	Description
Interval	When you have two or more merged variables on the response axis, a line with boundaries and a circle is plotted for each level of the stratifying variables. For each level of the stratifying variables: • The values of the summary statistic are calculated for each of the 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. 
Side by side intervals	In this style, each pair of variables defines an interval. 

Table 3.1 Bar Styles and Descriptions *(Continued)*

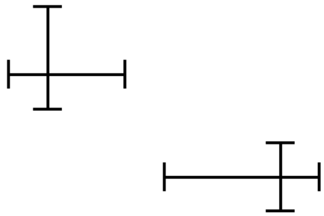
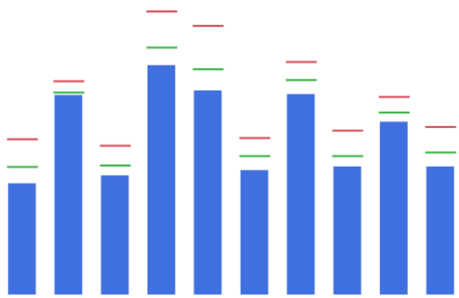
Bar Style	**Description**
Two-way interval	This style is for two-dimensional bar charts with bars in each direction. 
Single	For each level of the stratifying variables, a bar is plotted. When you have two or more merged variables on the response axis, the bar represents the value of the summary statistic 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)*

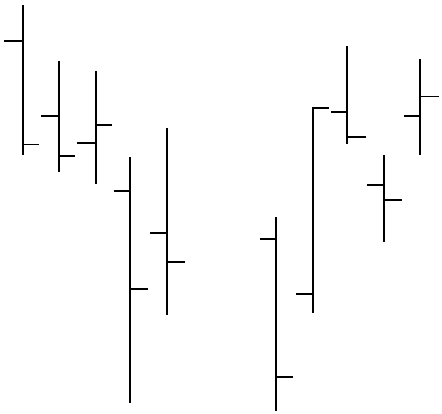
Bar Style	Description
Stock	When you have two or more merged variables on the response axis, line segments are plotted at the levels of the stratifying variables. For each level of the stratifying variables: • The values of the summary statistic are calculated for each of the merged variables. • A line 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)*

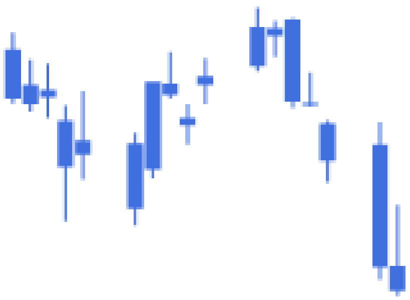
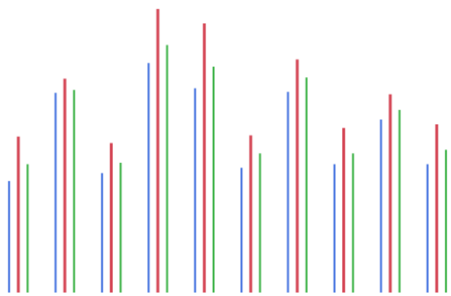
Bar Style	Description
Box Plot	When you have two or more merged variables on the response axis, box plots are shown for the levels of the stratifying variables. For each level of the stratifying variables: • The values of the summary statistic are calculated for each of the merged variables. • An outlier 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. 

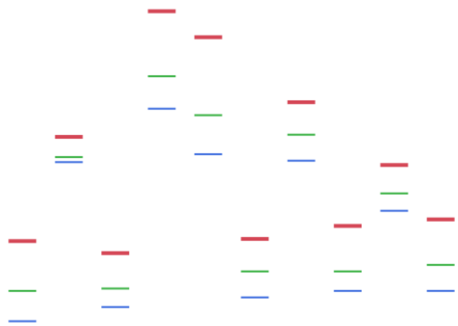
Table 3.1 Bar Styles and Descriptions *(Continued)*

<table>
<tr><th>Bar Style</th><th>Description</th></tr>
<tr><td>Float</td><td>

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.

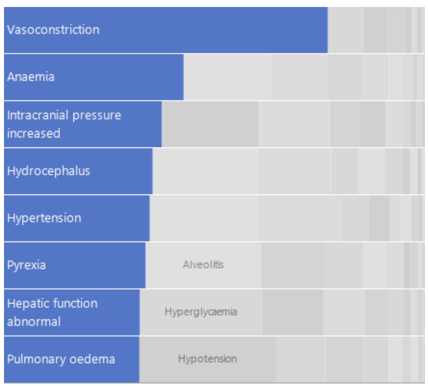
This style is useful when you have many merged variables.



</td></tr>
<tr><td>Packed</td><td>

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.



</td></tr>
</table>

Area

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

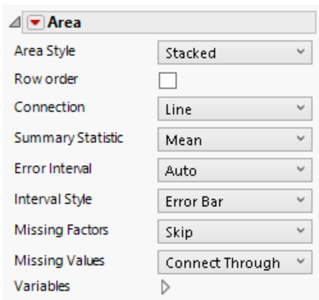
The **Area** element shows a contiguous area that represents the area beneath summary statistic values of a variable across the values or categories of another variable. The Area element can be considered a continuous representation of the Bar element.

- 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 "Example of an Area and Line Chart".

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.

For an example using Variables, see “Example of an Area and Line Chart”.

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

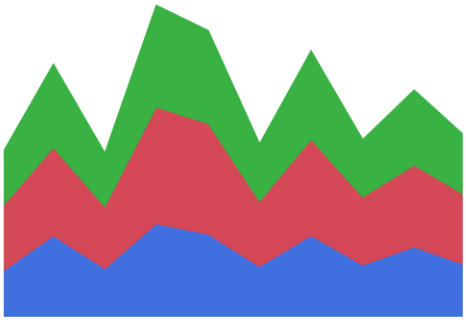
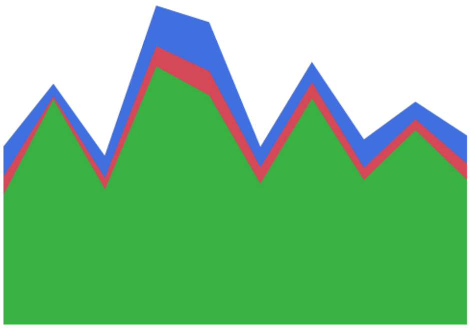
Area Style	Description
Stacked	Shows areas that correspond to summary statistics for each of the merged variables. The plots are colored and a legend appears in the Legend zone. 
Overlaid	Shows the area plots for the merged variables superimposed over each other. The plots are colored and a legend appears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

Box Plot

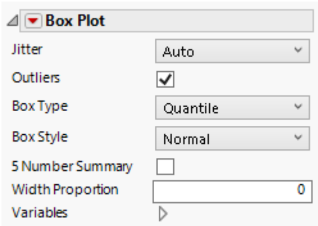
The **Box Plot** element shows outlier or quantile box plots. A box plot provides a compact view of a distribution of values. The box extends from the 25th percentile to the 75th percentile where the distance between the 75th and 25th percentiles is the interquartile range (IQR). The median is marked within the box. For outlier box plots, whiskers extend to the last point that is within 1.5*IQR from the ends of the box. For quantile box plots the whiskers cover the full range of observations with tick marks at specific quantiles. For more information about outlier and quantile box plots, see *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".

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.

For an example using Variables, see "Example of an Area and Line Chart".

Red Triangle Options for Box Plots

Confidence Diamond Shows or hides a confidence diamond about the mean. The diamond covers the 95th confidence interval for the mean.

Notched Displays a box plot that is notched at the median where the notches span:

$$\text{median} \pm (1.58 \times \text{IQR})/(\sqrt{n})$$

Fences Shows or hides vertical lines, or fences, at the end of the box plot whiskers.

Shortest Half Shows or hides a bracket that indicates the location of the shortest interval that contains 50% of the data.

Shortest Half Color Enables you to select a color for the shortest half bracket.

Response Axis (Available only when the X and Y axes variables both have a continuous modeling type.) 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.

Box Placement (Available with grouping variables.) Enables you to align or offset the box plots.

Tip: When you have a categorical grouping variable for your box plot, right-click on the categorical axis and select **Size By > Count** to draw box plots that are proportional to the number of observations in each category.

Table 3.3 Box Styles and Descriptions

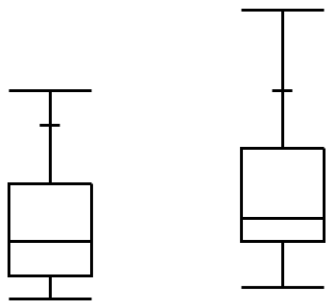
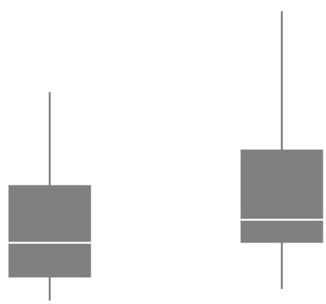
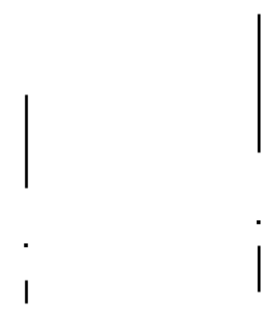
Box Style	Description
Normal	Shows the box defined by the quartiles as hollow with only a line at the median. 
Solid	Shows the box defined by the quartiles with solid fill. The line at the median is shown by white space. 

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 *Basic Analysis*.

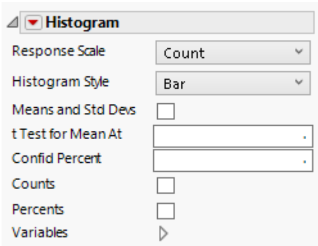
Note: The bars of the histogram align with the tick marks on the axis for the zone containing the variable. If you customize the axis, the histogram adjusts to match the new tick marks.

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 scatterplot 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 an Overlaid Histogram and Ridgeline Chart".

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.

For an example using Variables, see "Example of an Area and Line Chart".

Red Triangle Option 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". The intensity legend shows counts or average values:

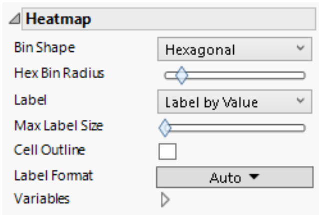
- If no Color variable is assigned, the counts in each bar or rectangle define the scale for the intensity legend.

- 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: Hover 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.

Figure 3.35 Heatmap Options



Bin Shape (Available for continuous variables). Specifies the shape of the bin.

Hex Bin Radius (Available for hexagonal bins.) Enables you to adjust the radius of the bin.

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

Max Label Size Enables you to adjust the size of the label.

Cell Outline (Available when a size variable is used.) Enables you to add outlines for the original bin sizes. This highlights the size scaling on the bins.

Label Format (Available for value and percent labels.) Enables you to set the format for the cell labels.

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.

For an example using Variables, see “Example of an Area and Line Chart”.

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.

In the Heatmap panel, under Variables, deselect all check boxes except the one for Color.

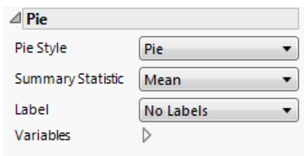
Pie

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.

Label Format (Available for value and percent labels.) Enables you to set the format for the cell labels.

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.

For an example using Variables, see "Example of an Area and Line Chart".

Table 3.4 Pie Style Descriptions

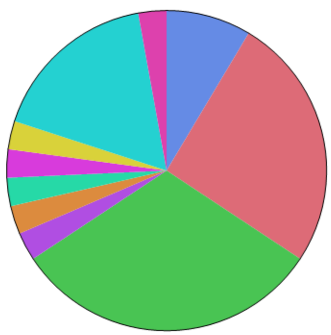
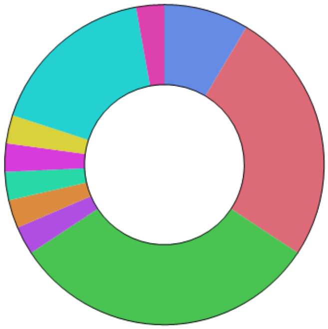
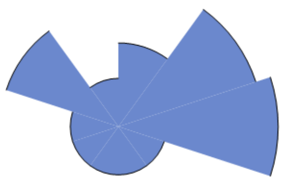
Pie Style	Description
Pie	Traditional pie chart with each slice sized by the Summary Statistic. 

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 **Size** zone across the levels or values of one or more X variables.

To construct a treemap that shows summary values across the levels of two or more X variables, merge the variables in the X zone. The rectangles are sized to represent the values of the summary statistic without leaving unused space in the overall rectangular layout.

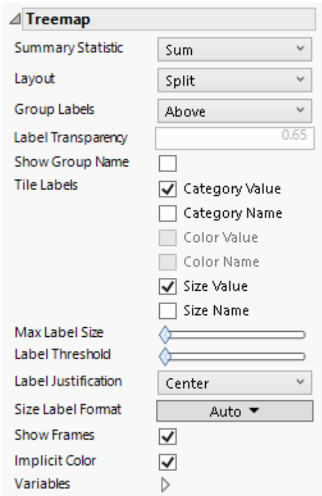
By default, the data table order is used when ordering the tiles in the treemap using the Split layout. Alternatively, you can specify one or two continuous variables as Y variables for ordering:

- If one Y variable is specified, it orders both the X and Y axes.
- If two Y variables are specified, the first orders the X axis and the second orders the Y axis.

If the Squarify layout is specified, then the order of the tiles is determined by the size, and ordering variables are ignored.

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

Figure 3.37 Treemap Options



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 (Available when you have more than one X variable.) Shows or hides the additional group labels above each category or floating in the center of each category.

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

Label Transparency (Available with floating group labels.) Specifies the transparency of the floating group labels.

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.

Color Name The name of the Color variable.

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

Size Name The name of the Size variable, or the Y variable if no Size variable exists (Size Name).

Tip: Hover 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.

Color Label Format (Available when a color variable is used.) Enables you to set the format for the cell labels based on the color variable.

Size Label Format (Available when Size Value is selected.) Enables you to set the format for the cell labels based on the size value variable.

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.

For an example using Variables, see “Example of an Area and Line Chart”.

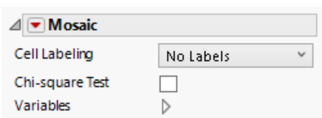
Mosaic

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 *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.

Figure 3.38 Mosaic Options



Cell Labeling Add labels to cells based on counts, percents, values, or row.

Label Format (Available for percent or value labels.) Enables you to set the format for the cell labels.

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.

For an example using Variables, see "Example of an Area and Line Chart".

Red Triangle Option 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 summary statistics 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.

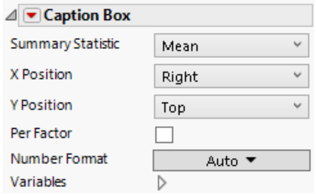
Summary Statistic for Categorical Variables

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".

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".

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.

Number Format Enables you to set the format of the summary statistic.

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.

For an example using Variables, see "Example of an Area and Line Chart".

Red Triangle Option 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.

Figure 3.40 Formula Option

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.

For an example using Variables, see “Example of an Area and Line Chart”.

Red Triangle Option 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”.

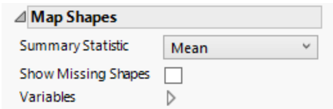
If you have a variable in the Map Shape zone, the X and Y zones disappear. You can add information from other variables using Color and Size.

- 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 "Maps". For examples, see "Examples of Creating Maps", 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.

For an example using Variables, see "Example of an Area and Line Chart".

Parallel

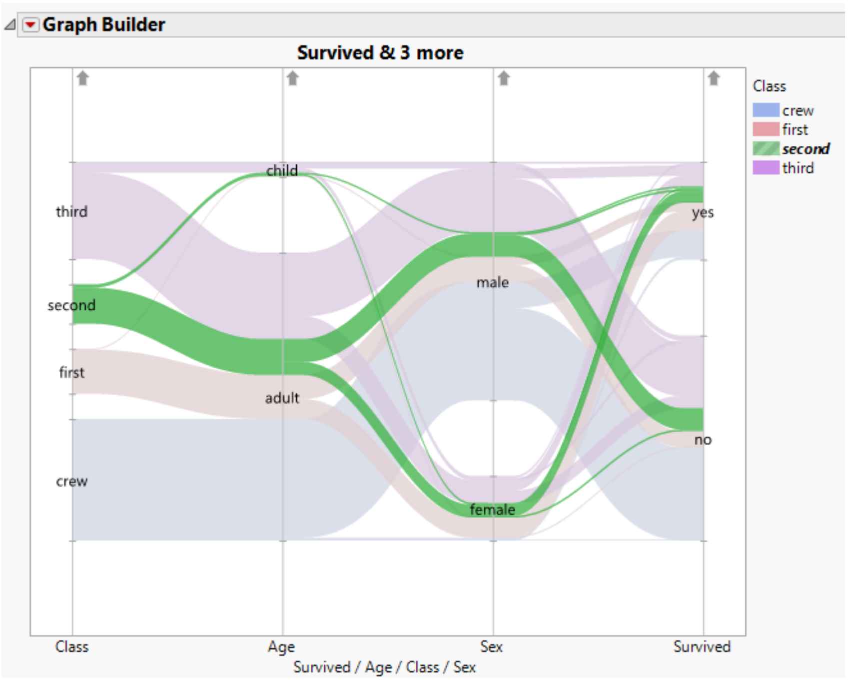
The **Parallel** element connects the values in a row across two or more variables. Drag two or more variables together to either the X or Y zone. The variable names appear as axis labels in the zone to which they were dragged.

- 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.

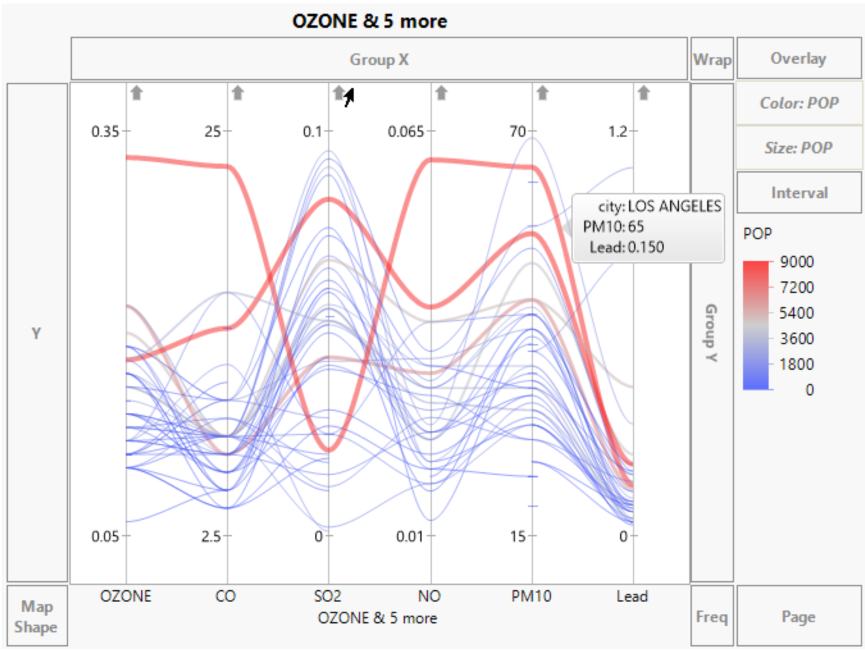
- To color the curves by the values of a variable, drag the column of interest to the **Color** zone. The categorical or continuous color theme selected in your Preferences is shown in the Legend zone.

- 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: Hover 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



Tips:

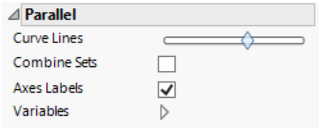
- You can add a line element to a parallel plot to show a summary relationship across the parallel plot.
- You can add reference lines for specification limits. For information, see *Using JMP*.

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.

For an example using Variables, see "Example of an Area and Line Chart".

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:

Table 3.5 Graph Builder Zones

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.

Table 3.5 Graph Builder Zones *(Continued)*

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, Axes Variable Labels, or Graph Titles”.

Table 3.5 Graph Builder Zones *(Continued)*

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 Option for Formula" and "Graph Builder".
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". **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. Right-click on legend items for customization options.

Table 3.5 Graph Builder Zones *(Continued)*

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. A legend appears to the right of the plot. Right-click on legend items for customization options.
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.

Table 3.5 Graph Builder Zones *(Continued)*

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 *Using JMP*. **Tips:** • 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. • To copy individual graphs using the selection tool, first save the Graph Builder output to a journal (Edit > Journal).

For more information about the legend, see “Legend Options”.

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

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

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

Legend Shows or hides the legend.

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

X Axis Shows or hides the X axis.

Y Axis Shows or hides the Y axis.

X Axis Title Shows or hides the X-axis title.

Y Axis Title Shows or hides the Y-axis title.

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 *Using JMP*.

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

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

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 Specifies 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 *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 deselecting 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

Numerous options are contained in Graph Builder right-click menus. The specific menu options depend on where in a graph you click, and the types and number of variables in your graph. Descriptions for the right click menu options are grouped in the following categories.

- "Graph Options"
- "Grouping Zones Options"
- "Options for Columns in the Variables Panel"
- "Options for Axes, 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*.

Tip: Select **Rows > Name Selection in Column** to create a new column to name and label selected points in a graph.

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".

Order By Orders the levels of a variable. See "Order the Levels of a Categorical Variable".

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".

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".

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 *Using JMP*.

Options for Axes, Axes Variable Labels, or Graph Titles

Right-clicking on axes, axes variable labels, or graph titles show options for customizing your graph. Select **Show Properties** from a right-click menu to access the properties panel for additional customizations, such as tick line thickness or color. The following options are included in the right-click menus:

Graph Titles

Alignment Changes the graph title alignment.

Span Changes the width of the graph title span.

Additional options for customizing fonts, font colors, and the visibility of the title are also available in the right-click menu.

Axes

Axes have two levels of right-click menus, one for the axis title and one for the axis labels.

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.

Date/Time Bin (Appears only for a date or time axis variable.) Enables you to select a level of granularity to bin the data. For example, you might choose to bin at a high level, by year, or at a low level, by second.

Date/Time Wrap (Appears only for a date or time axis variable.) Enables you to select a cyclic period for your date data.

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

Remove Removes the variable from the zone.

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.

Move Left (Appears only after you have moved a variable to the right.) Moves the variable back to the left.

Align Ticks with Left Axis Aligns the major tick marks with those on the left axis.

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".

Note: Select **Other** to indicate a continuous variable as the ordering variable.

Order Statistic Enables you to select a statistic to use for the ordering.

Save Value Order (Available once an ordering has been defined.) Saves the defined order to the Value Order column property.

Note: You can edit value ordering from Graph Builder. Right-click on an axis and select **Edit Value Order**.

Size By (Available for a categorical axis.) Enables you to size your box plots by the count of observations within a group or by a numerical column.

Size Statistic (Available when the Size By option indicates a numerical column as the sizing variable.) Enables you to select a statistic to use for the sizing.

For descriptions of the general axis options below the line, see *Using JMP*.

Note: The **Revert Axis** option reverts the axis back to the default settings. The **Revert Scale** option reverts the scale-specific items: min, max, major interval, n minor.

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 elements, such as lines, fill colors, and gradients. Right-click on legend elements for customization properties. These properties include color and labeling options when available. Additional legend settings are found in 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 graphics 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”.

Revert Legend Returns the legend to its initial settings.

Move the Legend

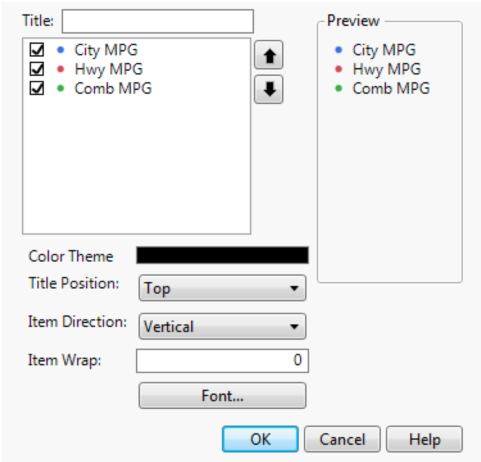
To move the position of the legend, click and drag to the desired location. The location can be a hotspot corresponding to the locations accessible from the Graph Builder red triangle menu or the legend or it can be any other location within the graph.

To use the menu options, click the Graph Builder red triangle and select **Legend Position**. 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 *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 to the item.

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".

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

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:

None Density contours use one transparency level for all levels.

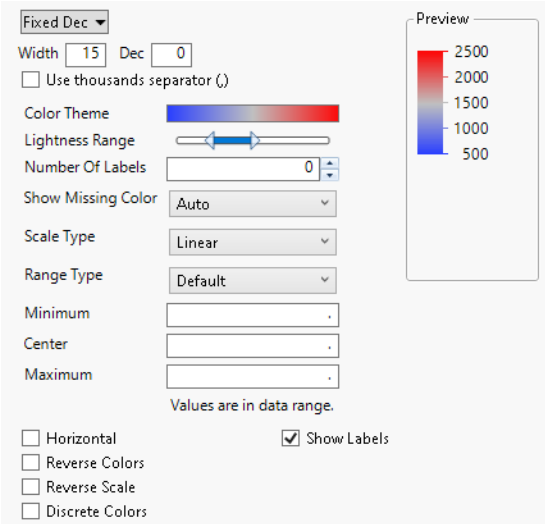
Linear Density contours get gradually more transparent at lower densities.

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 *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.

Show Missing Color Specifies whether to show missing color in the legend. Auto shows the color only when there are missing values, Off never shows the color, On always shows the color.

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.

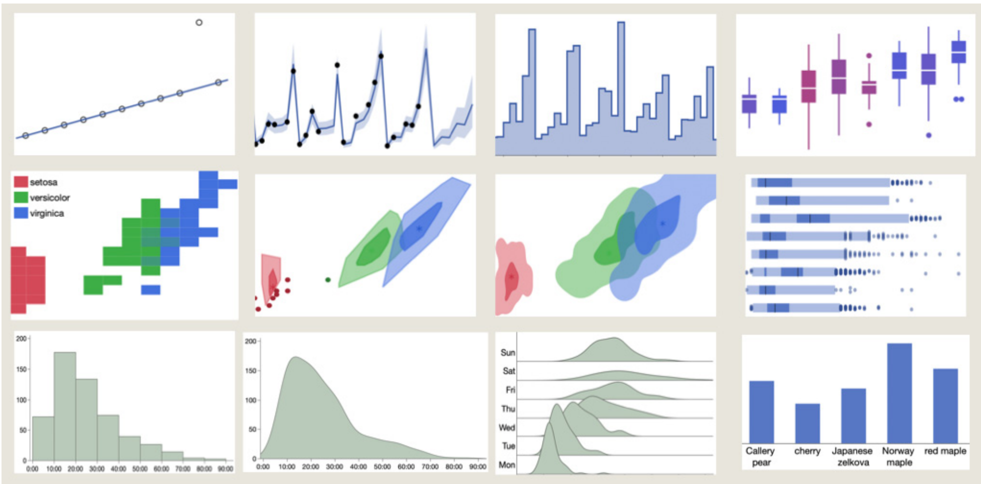
Chapter 4

Graph Builder Examples

Interactive Data Visualization

Graph Builder provides you with a variety of ways to visualize and explore your data. This chapter provides detailed examples of data exploration using bar charts, histograms, area plots, maps, and other types of plots.

Figure 4.1 Graphs Created in Graph Builder



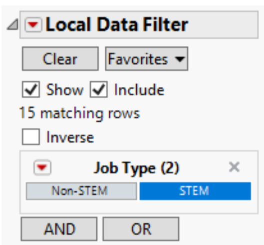
Contents

Example of an Ordered Bar Chart

Data	This example uses data from college graduates in 2012. The table contains counts of degrees for graduates in 15 fields of study. Each degree is subdivided by the graduate's job type. There are categorical columns for degree, degree type, and job type (where type is STEM or non-STEM related). The continuous column Number contains the number of graduates from each unique combination of degree and job type. **Note:** These data come from the U.S. Census Bureau (2012).
Techniques	This example uses color, a local data filter, and sorting options.
Goal	The goal of this example is to visualize the percentage of graduates who are employed in STEM-related jobs, based on their degree type.

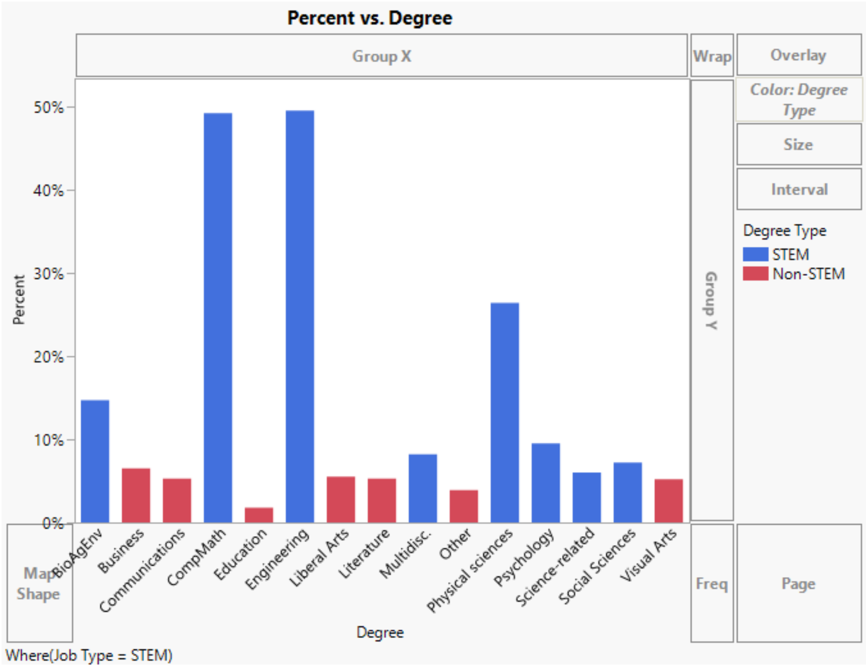
1. Select **Help** > **Sample Data Library** and open STEM Jobs.jmp.
2. Select **Graph** > **Graph Builder**.
3. Select Degree and drag it to the **X** zone.
4. Select Percent and drag it to the **Y** zone.
5. Click the Bar element .
6. Click the Graph Builder red triangle and select **Local Data Filter**. In the Local Data Filter, perform these steps:
   a. Click Job Type and then click .
   b. Click **STEM**.

Figure 4.2 Selections in the Local Data Filter



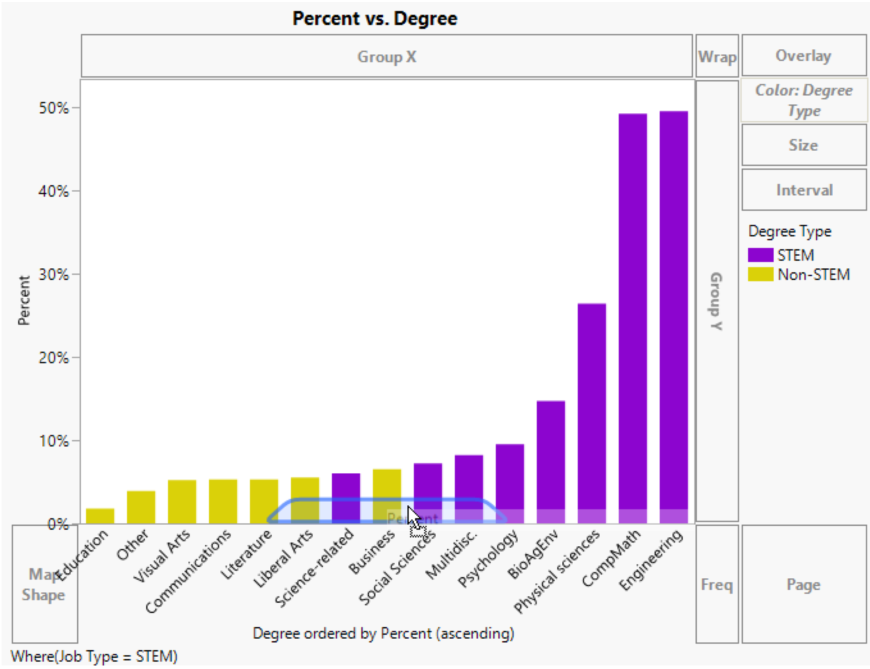
7. In the Graph Builder column list, select Degree Type and drag it to the **Color** zone.

Figure 4.3 Bar Chart Colored by Degree Type



8. In the legend, change the bar colors according to degree type:
   – Right-click **STEM** and change the fill color to purple.
   – Right-click **Non-STEM** and change the fill color to yellow.
9. In the Graph Builder column list, select Percent and drag it above the X axis. See Figure 4.4.

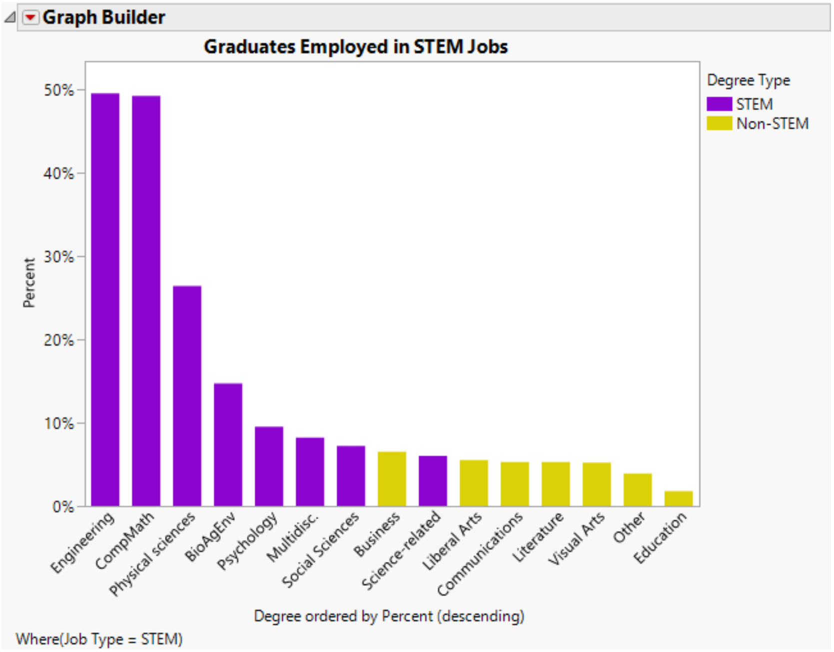
Figure 4.4 Ordering Degree by Percent



This orders the degree by percent in ascending order. However, you want it to be in descending order.

10. Right-click the X axis and select **Order By** > **Percent, descending**.
11. Click the graph title and type Graduates Employed in STEM Jobs and press Enter.
12. Click **Done**.

Figure 4.5 Completed Bar Chart



This graph displays the percent of graduates employed in STEM jobs by their degree type. Nearly 50% of Engineering and CompMath graduates are employed in STEM jobs, as compared to less than 2% of Education graduates.

Example of an Area and Line Chart

Data	This example uses data based on website traffic. It includes hourly measurements and historical range information.
Techniques	This example uses Recode to improve data labels, multiple graph elements (area and line), and legend customization. It also shows how to remove variables from a graphing element when multiple variables are plotted on a single axis.
Goal	The goal of this example is to display trend lines for website traffic within an overall range.

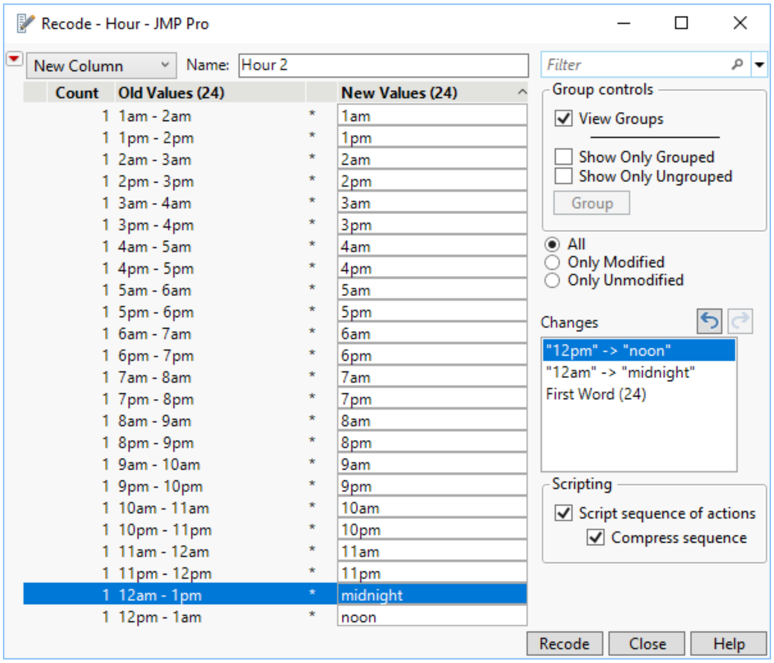
1. Select **Help** > **Sample Data Library** and open Bands.jmp.

2. Select the Hour column, and then select **Cols** > **Recode**.
3. In the Recode window, make these changes:
   a. Click the red triangle and select **First Word**.

      Notice that the New Values include only the first hour.

   b. Under New Values, click **12am** and enter **midnight**. Click **12pm** and enter **noon**.

Figure 4.6 Completed Recode Window



   c. Click **Recode**.

      A new column called Hour 2 is added to the data table.

Tip: Recode enables you to recode your data to a new column, as a formula in a new column, or to overwrite your current data. For more information about Recode, see *Using JMP*.

---

4. Select **Graph** > **Graph Builder**.
5. Select Hour 2 and drag it to the **X** zone.

Note: The ordering of the X axis is logical, going from midnight to 11pm. This is because the Hour column has a value ordering column property that was copied to the Hour 2 column when you created it with Recode.

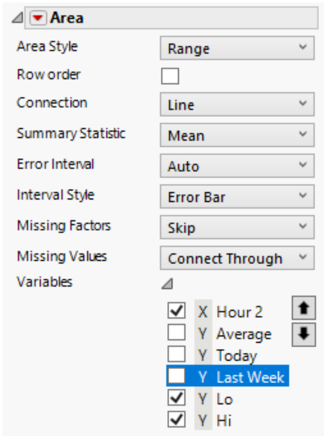
---

6. Select Average, Today, Last Week, Lo, and Hi and drag them to the **Y** zone.

7. Click the Area element .
8. In the Area properties panel, make these selections:
   a. Change the Area Style to **Range**.
   b. Open Variables and deselect Average, Today, and Last Week.

   Now the area element is applied only to the Lo and Hi variables.

Figure 4.7 Completed Area Properties Panel



9. Hold down the Shift key and click the Line element .

Tip: Holding down the Shift key and clicking adds additional elements to the graph, without replacing any existing elements.

10. In the Line properties panel, open Variables and deselect Lo and Hi.

    Now the line element is applied only to the Avg, Today, and Last Week Y variables.

Figure 4.8 Completed Line Properties Panel

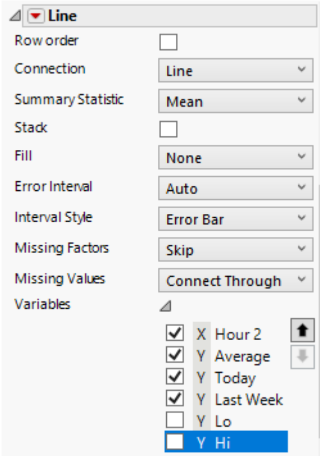
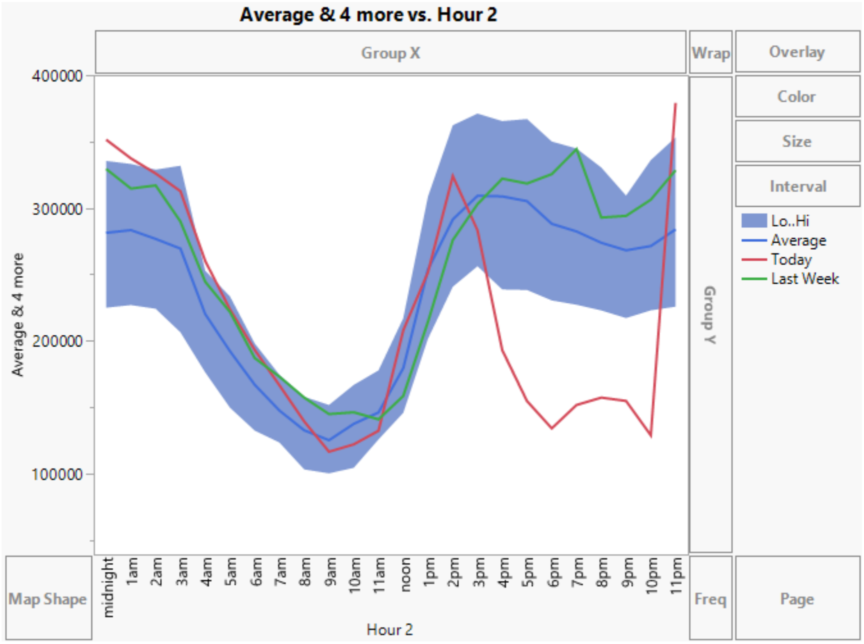


Figure 4.9 Area and Line Graphs

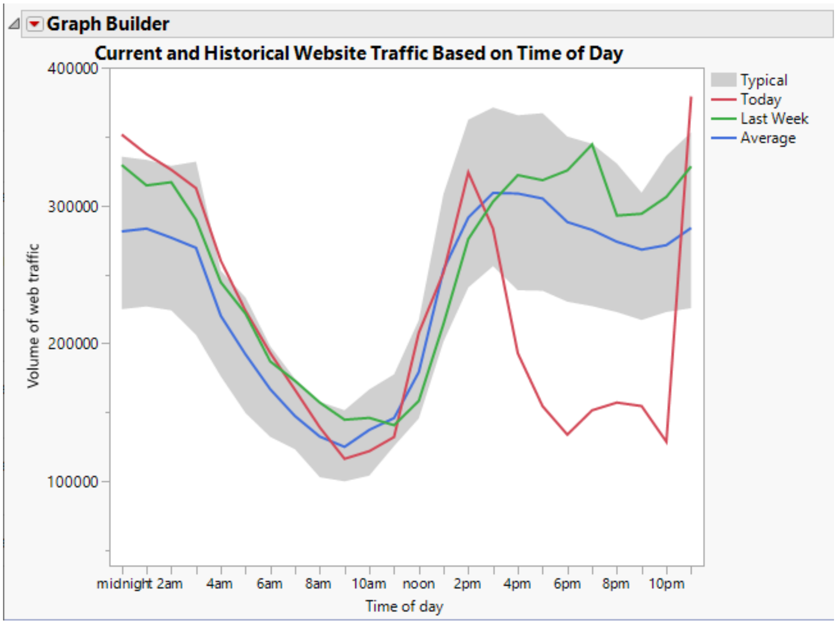


11. Customize the graph elements using the legend:
    a. In the legend at right, double-click **Lo..Hi**.
    b. In the Legend Settings window, double-click **Lo..Hi** and rename it to **Typical.**
    c. Highlight Today and Last Week, and click the up arrow to move them up one, so that the order is: Typical, Today, Last Week, and Average. Click **OK**.

   d. Right-click Typical > **Fill Color** and select light gray.

12. Right-click the X axis > **Axis Settings**. For Label Orientation, click **Horizontal**, and then click **OK**.
13. Change the titles of the axes and graph:
    a. Click the Y-axis title and enter Volume of web traffic.
    b. Click the X-axis title and enter Time of day.
    c. Click the graph title and enter Current and Historical Website Traffic Based on Time of Day.
14. Click **Done**.

Figure 4.10 Completed Area and Line Graph



Tip: If an item in the legend is still highlighted, click it to deselect it.

The graph shows that the afternoon website traffic fell to unusually low levels today. The red trend line that represents today's traffic peaks at 1pm and then falls to outside of the gray band that represents the range of expected daily traffic for this website by about 3pm.

Example of a Map Chart

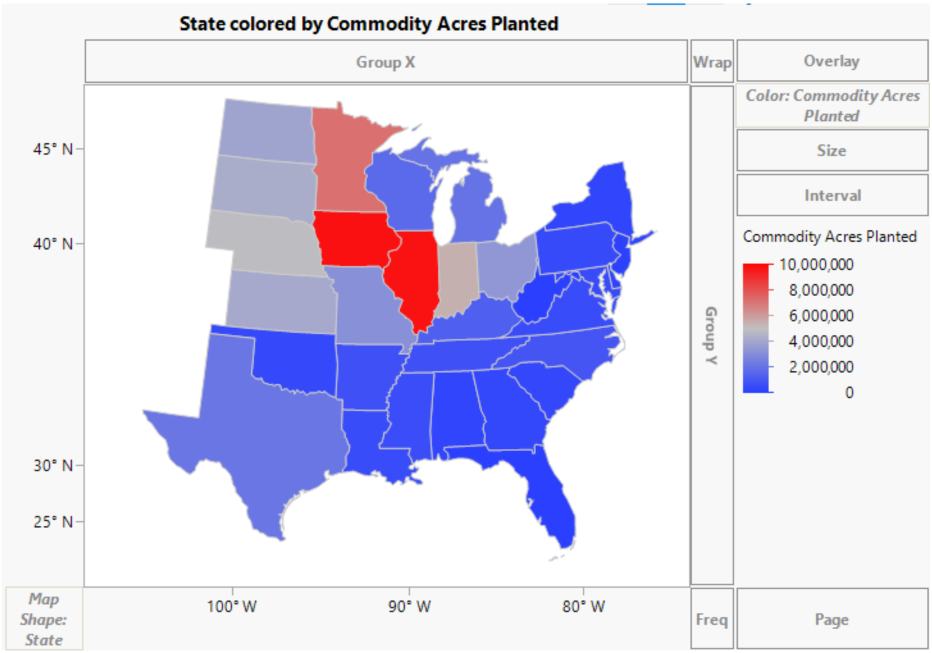
Data	This example uses data from 2017 of commodity acres planted in states within the US that grow corn, wheat, and soybeans. **Note:** These data come from the United States Department of Agriculture (2017).
Techniques	This example uses map shapes and legend customization.
Goal	The goal of this example is to show the median number of commodity acres that have been planted per state.

1. Select **Help** > **Sample Data Library** and open Corn Wheat Soybean Production.jmp.
2. Select **Graph** > **Graph Builder**.
3. Select State and drag it to the **Map Shape** zone.

   Note: JMP recognizes US State names as map shapes. For more information, see "Map Shape".

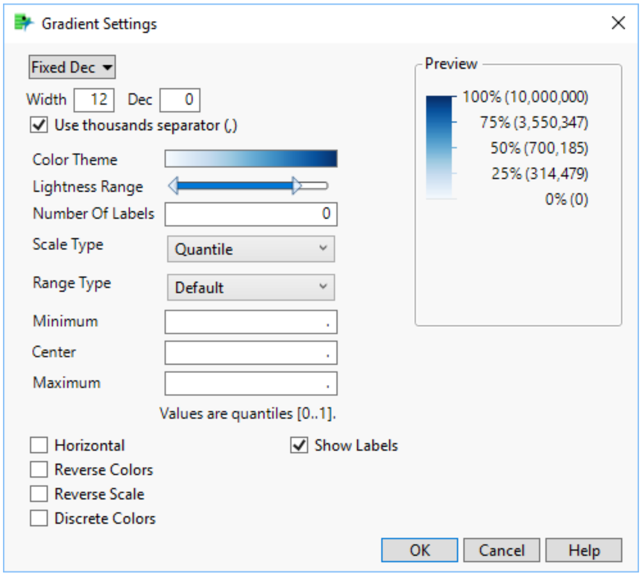
4. Select Commodity Acres Planted and drag it to the **Color** zone.
5. In the Map Shapes properties panel, change the Summary Statistic to **Median**.

Figure 4.11 Map Showing Median Commodity Acres Planted



6. Customize the graph elements in the legend:
   a. In the legend color bar, right-click and select **Gradient**.
   b. Next to Width, enter **12**.
   c. Next to Color Theme, click the color bar. Under Sequential, click the 2nd option (white to blue) and then click **OK**.
   d. From the Scale Type list, choose **Quantile**.

Figure 4.12 Completed Gradient Settings Window



e. Click **OK**.

7. Click the title above the graph and type **State Colored by Median Commodity Acres Planted**.

8. Right-click in the graph and select **Hover Label > Bar**.

   This adds a bar chart hover graph. When you hover over a state, you can see the proportions of the commodities grown in that state in a bar chart.

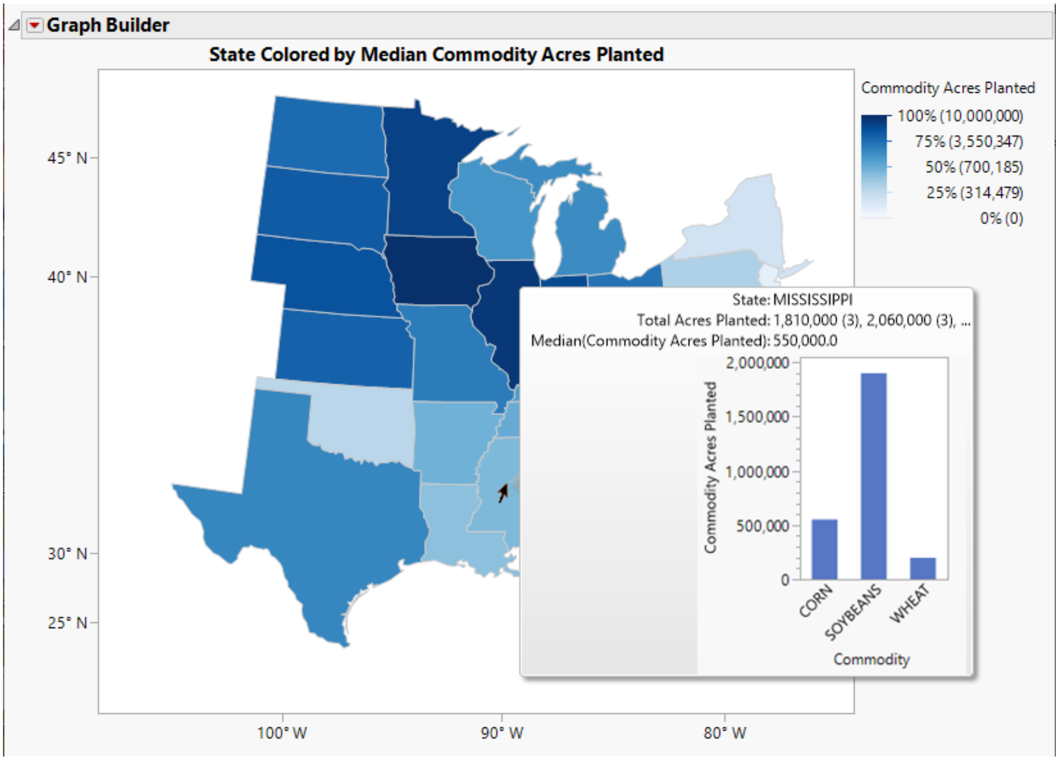
Note: For more information about Hover Labels, see *Using JMP*.

---

9. Click **Done**.

10. Hover over a state (in this case, Mississippi).

Figure 4.13 Completed Map Colored by Median Commodity Acres Planted



The coloring on the map highlights the states with the highest number of acres of wheat, corn, and soybeans. The dark blue states are known as America's Breadbasket. The hover graphs enable you to explore the proportions of the three crops grown in a state. This example shows that in Mississippi, soybeans are the prominent crop.

Tip: To see other examples using Graph Builder, run the additional scripts in the Corn Wheat Soybean Production.jmp data table.

Example of a Butterfly Chart

Data	This example uses summarized survey responses to 20 questions all using the same 5-point Likert scale.
Techniques	This example uses a bar chart with variable ordering and color customizations.

Goal	The goal of this example is to create a butterfly chart of the responses to 20 survey questions. This provides a summary visual of the results from a large number of survey questions in one chart.

1. Select **Help** > **Sample Data Library** and open Likert Survey.jmp.

   Notice the following about this data:

   – The data has a row for each question and columns with the counts of responders for each different response.

   – The order of the survey questions in the question column is not sequential. When you use this column in Graph Builder, JMP orders the rows sequentially.

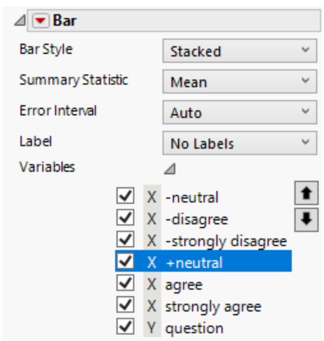
   – Formula columns are included in the table to create negative counts for negative responses. The neutral counts are split so that half contribute to a negative column and half contribute to a positive column.

2. Select **Graph** > **Graph Builder**.
3. Select question and drag it to the **Y** zone.
4. Select agree, strongly agree, -strongly disagree, -disagree, -neutral, and +neutral and drag them to the **X** zone.
5. Click the Bar element .
6. In the Bar properties panel, do the following:

   – Change the Bar Style to **Stacked**.

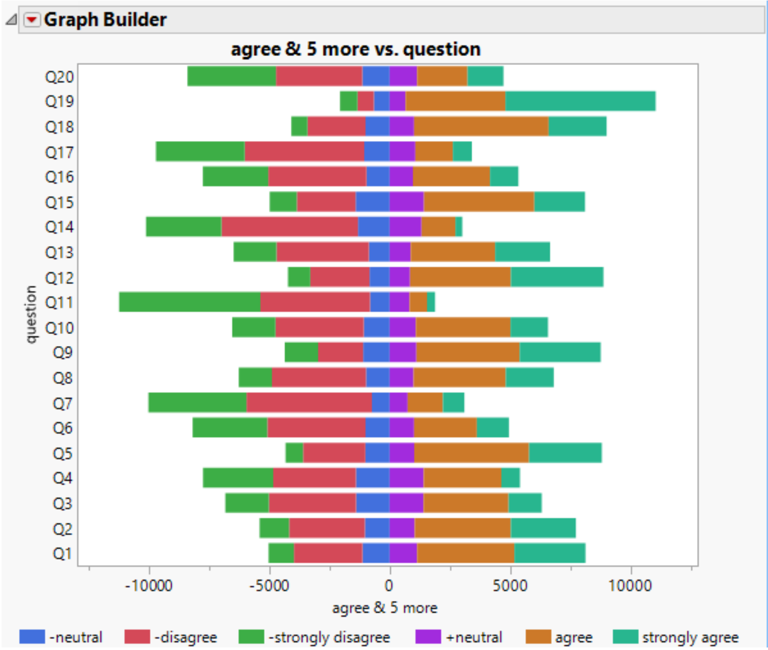
   – Open **Variables**, and then highlight a variable and use the arrow icons to move them into this order, from top to bottom: -neutral, -disagree, -strongly disagree, +neutral, agree, strongly agree.

Figure 4.14 Completed Bar Properties Panel



7. Click the Graph Builder red triangle and select **Legend Position** > **Bottom**.
8. Click **Done**.

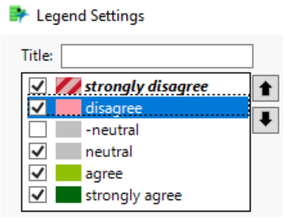
Figure 4.15 Graph Prior to Legend Customizations



Customize the Legend Colors and Ordering

1. Right-click on each color in the legend and change the Fill Color:
   - Light gray for +neutral and -neutral
   - Light pink for -disagree
   - Red for -strongly disagree
   - Light green for agree
   - Dark green for strongly agree
2. Hover over any item in the legend. The cursor should appear as a hand. Double-click to open the Legend Settings window.
3. Highlight a variable and use the arrow icons to change the order from top to bottom: -strongly disagree, -disagree, -neutral, +neutral, agree, strongly agree.
4. Double-click +neutral and rename it to **neutral**.
5. Deselect the box next to -neutral so that it no longer appears in the legend, because it is a duplicate.
6. Double-click -strongly disagree and -disagree and remove the - sign.

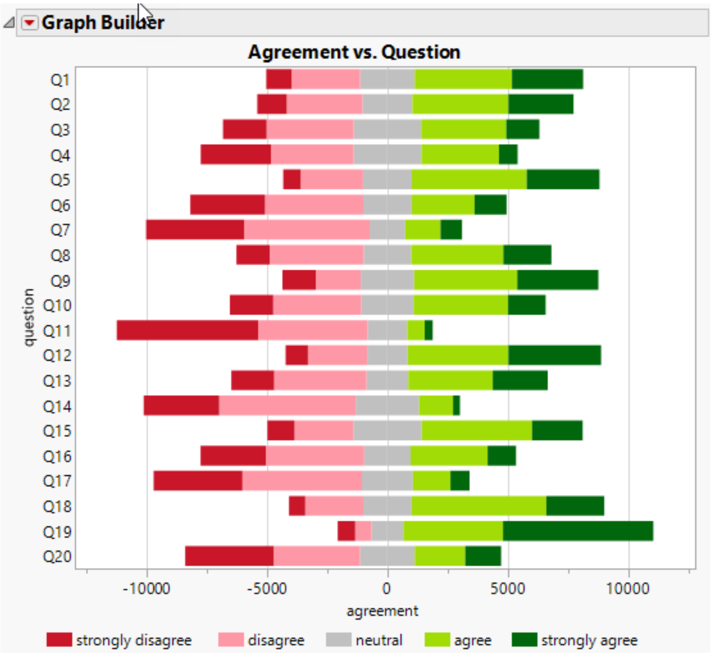
Figure 4.16 Completed Legend Settings



7. Click **OK**.
8. Change the titles of the X axis and graph:
   a. Click the graph title and enter Agreement vs. Question.
   b. Click the X-axis title and enter agreement.
9. Double-click in the Y axis and select the **Reverse Order** box, and then click **OK**.
10. Double-click in the X axis and select the **Major Grid Lines** box, and then click **OK**.

    The grid lines help you visually judge the differences between question results.

Figure 4.17 Final Likert Scale Graph



Tip: If an item in the legend is still highlighted, click it to deselect it.

This graph provides a visual summary of 20 survey questions, ignoring the non-responses. Note that the non-responses are less than 1% of the total responders for any question. Therefore, leaving them out of the visualization does not impact the conclusions. Question 19 has the most positive responses, and Question 11 has the least favorable responses.

Example of a Scatterplot

Data	This example uses data from 442 diabetic patients. The data includes baseline clinical and laboratory data as well as a binary measure of diabetes disease progression obtained one year after each patient's initial visit. This measure classifies disease progression as Low or High.
Techniques	This example uses two scatterplots with a shared X axis, axis customization, and annotation.
Goal	The goal of this example is to compare and understand the predicted probabilities of High disease progression from two classification models.

Run the Classification Models

You want to construct a classification model to predict the disease progression based on clinical and laboratory variables. You will build two different classification models, save prediction formulas, and compare their predicted classifications.

1. Select **Help > Sample Data Library** and open Diabetes.jmp.
2. Click the green triangle next to the **Decision Tree of Y Binary** script to build a decision tree prediction model.
3. Click the Partition for Y Binary red triangle and select **Save Columns > Save Prediction Formula**. You can close this window.

   This saves the probability formulas to the data table.
4. In the data table, right-click the Prob(Y Binary= = High) column and select **Column Info**.
5. Change the column name to Partition Prob High and click **OK**.
6. Click the green triangle next to the **Neural of Y Binary** script to build a neural net prediction model.
7. Click the Model NTanH(3) red triangle and select **Save Profile Formulas**. You can close this window.

   This saves the probability formulas to the data table.
8. In the data table, right-click the Probability(Y Binary=High) column and select **Column Info**.

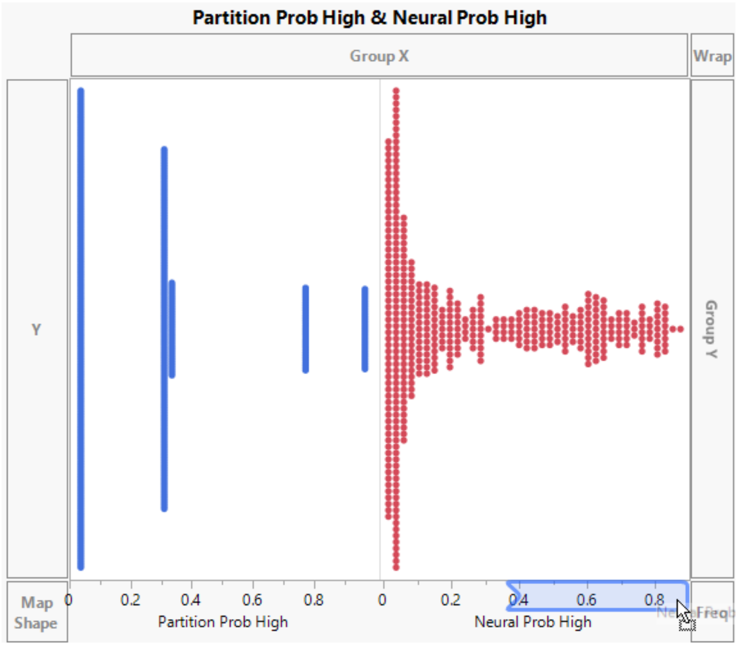
9. Change the column name to Neural Prob High and click **OK**.

Create the Initial Graph

You will compare the model predicted probabilities of a patient having a High disease progression.

1. Select **Graph > Graph Builder**.
2. Select Partition Prob High and drag it to the **X** zone.
3. Select Neural Prob High and drag it to the **X** zone, to the right of Partition Prob High. This creates a second X axis.

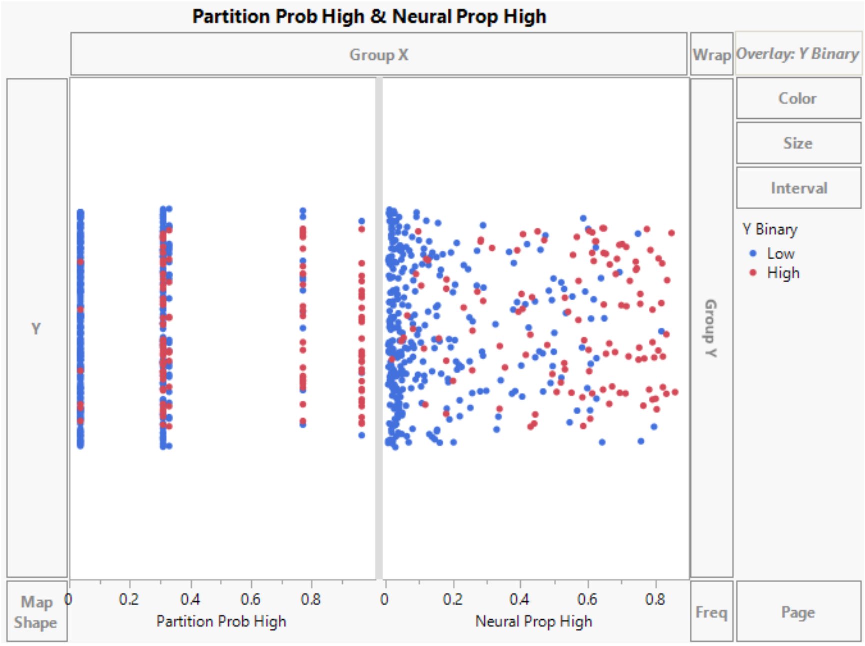
Figure 4.18 Drag Neural Prob High to the Right of Partition Prob High



4. Select Y Binary and drag it to the Overlay zone.
5. Click the Graph Builder red triangle and select **Graph Spacing**.
6. Type 6 next to Graph Spacing and click **OK**.

   This increases the spacing between the two X axes.

Figure 4.19 Initial Graph of Model Probabilities

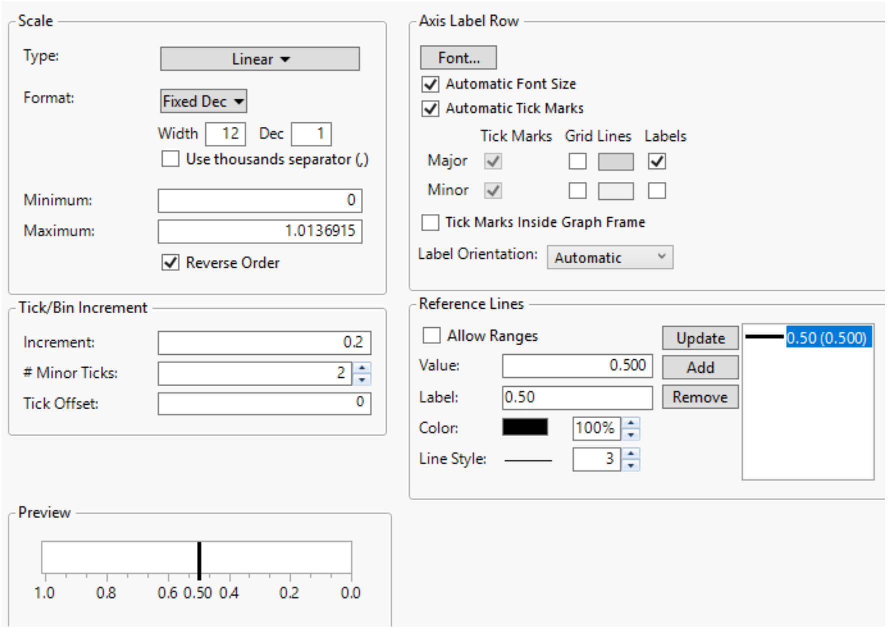


Customize the Graph

In both modeling platforms, the default threshold value is 0.50. This means that if a patient has a predicted probability greater than 0.50 of "High", the model predicts their classification as "High". Use reference lines to show the threshold value in the graph.

1. Right-click the Partition Prob High X axis and select **Axis Settings**.
2. In the Scale section, click the box next to **Reverse Order**.
3. In the Tick/Bin Increment section, set **# Minor Ticks** to 2.
4. In the Reference Lines section, type 0.50 in the boxes next to **Value** and **Label**.
5. In the box next to Line Style, enter 3. The 3 indicates the thickness of the reference line.
6. Click **Add**.

Figure 4.20 X Axis Settings

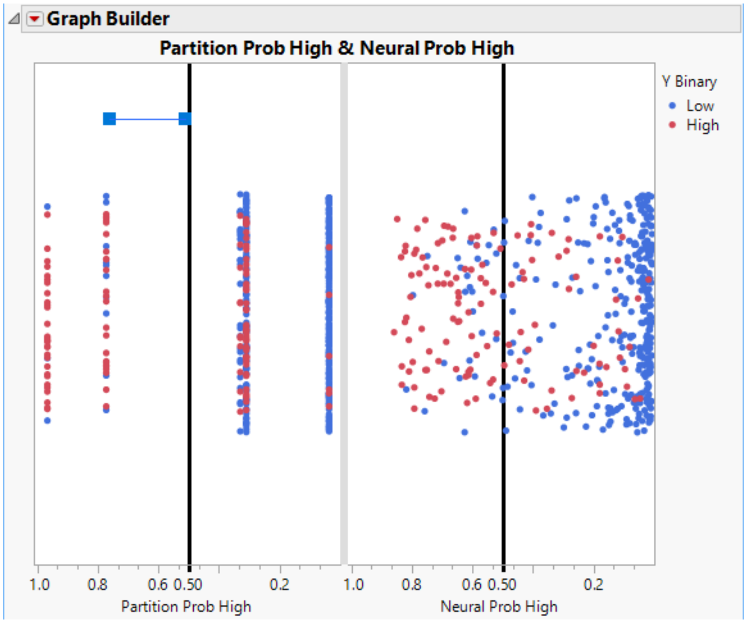


7. Click **OK**.
8. Right-click on the Partition Prob High X axis and select **Edit** > **Copy Axis Settings**. Right-click on the Neural Prob High X axis and select **Edit** > **Paste Axis Settings**.
9. Click **Done**.

Annotate the Graph

1. In the main menu, click **Tools** > **Line**. Draw a horizontal line to the left of the reference line in the Partition Prob High graph.

Figure 4.21 Line Drawn to the Left of the Partition Prob High Reference Line

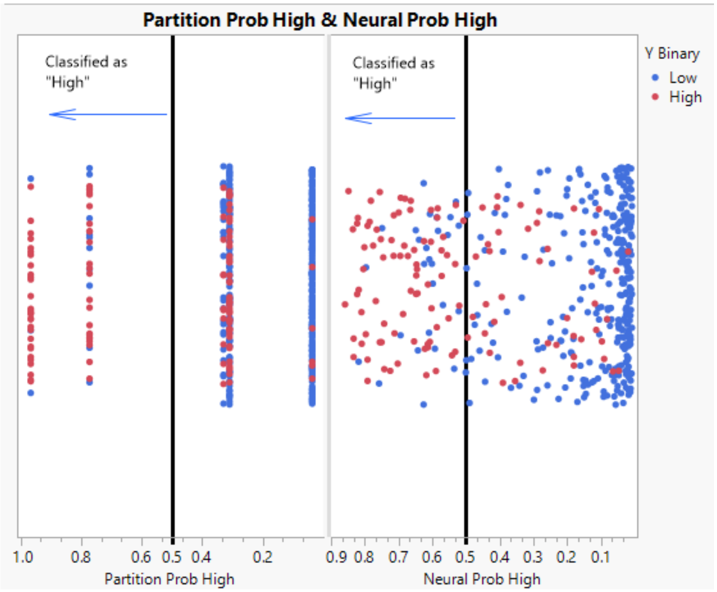


2. Right-click the line and select **Point to**.

   The arrow should be pointing away from the reference line.

3. In the main menu, click **Tools** > **Annotate** and click on the graph above the arrow.
4. Type **Classified as "High"** in the text box. See Figure 4.22.
5. Repeat step 1 through step 4 in the Neural Prob High graph.

Figure 4.22 Distribution of Predicted Probabilities



The graph shows that the distribution of the predicted probabilities differs between the two models. The partition model has five predicted score levels while the neural model scores are dispersed across the score range. For both models, there is a cluster of blue data points on the far right of each plot. These are low subjects that both models classify correctly as they fall to the right of the threshold.

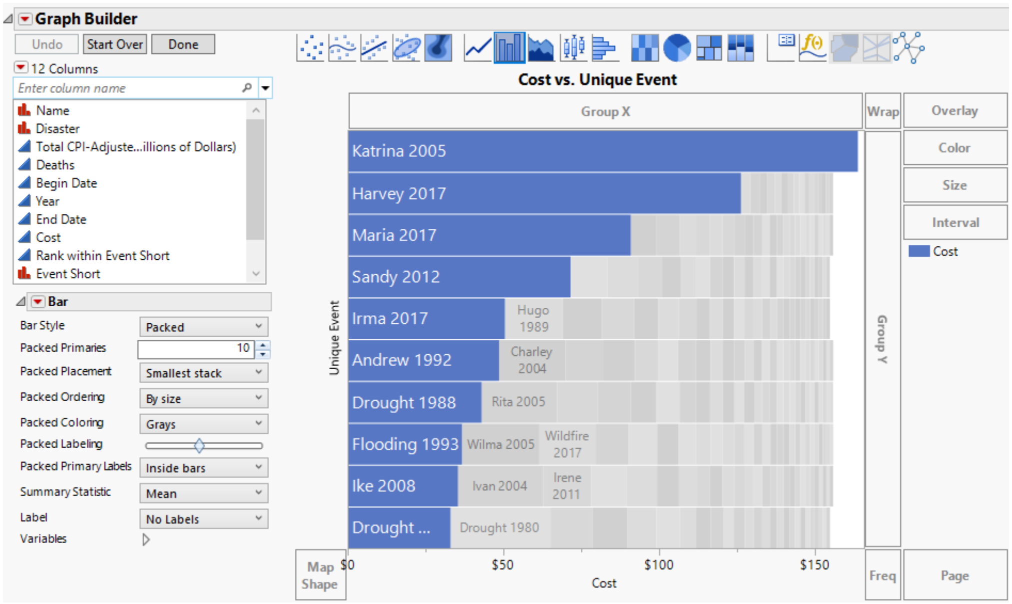
Example of a Packed Bar Chart

Data	This example uses data from the National Oceanic and Atmospheric Administration on the economic impact of weather and climate events in the U.S. during the years 1980 to 2018.
Techniques	This example uses a packed bar chart with 10 primary categories.
Goal	The goal of this example is to create a packed bar chart showing the events that had the greatest economic impact.

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 to the left until it is about halfway.

Figure 4.23 Weather Events With Large Economic Impacts in a Packed Bar Chart



The top 10 categories appear as a bar chart with blue bars. These are the “Packed Primaries” that you set to 10 in the options. Secondary categories are labeled and in gray. You can hover over any bar for more details.

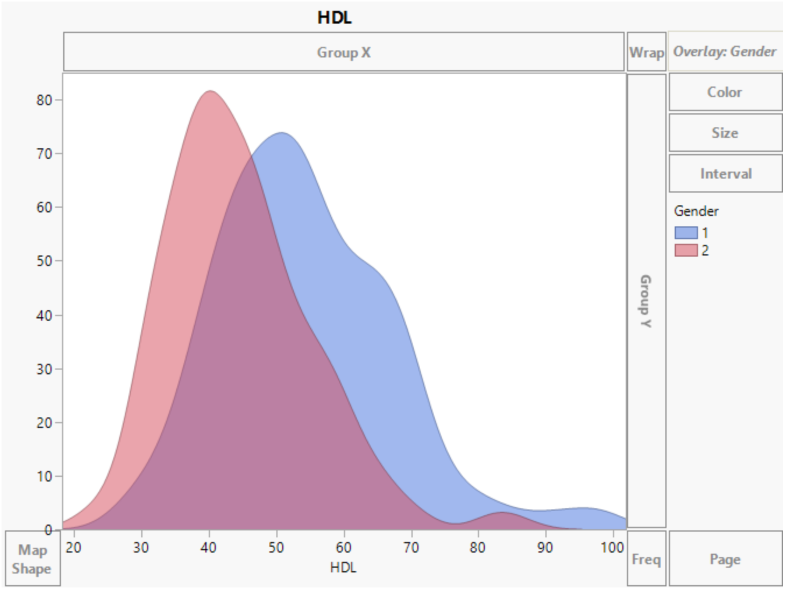
In this graph, you can clearly see that Katrina in 2005 had the largest financial impact of any storm that hit the US between 1980 and 2018. The 2017 hurricanes, Harvey and Maria, also had large financial impacts.

Example of an Overlaid Histogram and Ridgeline Chart

Data	This example uses data from 442 diabetic patients. The data include baseline clinical and laboratory data as well as a binary measure of diabetes disease progression obtained one year after each patient's initial visit. This measure classifies disease progression as Low or High.
Techniques	This example uses overlaid histograms and a ridgeline chart.
Goal	The goal of this example is to explore the distribution of a continuous variable grouped by categories. Specifically, you are interested in exploring how the distribution of HDL varies by gender and then by disease progression.

1. Select **Help > Sample Data Library** and open Diabetes.jmp.
2. Select **Graph > Graph Builder**.
3. Select HDL and drag it to the **X** zone.
4. Select Gender and drag it to the **Overlay** zone.
5. Click the **Histogram** element icon .
6. In the Histogram properties panel, from the Histogram Style list, select **Kernel Density**.
7. Hover over the Y axis near the top of the scale until the cursor turns into a horizontal hand. Click and drag until the maximum value is about 85.

Figure 4.24 Overlaid HDL Histograms by Gender

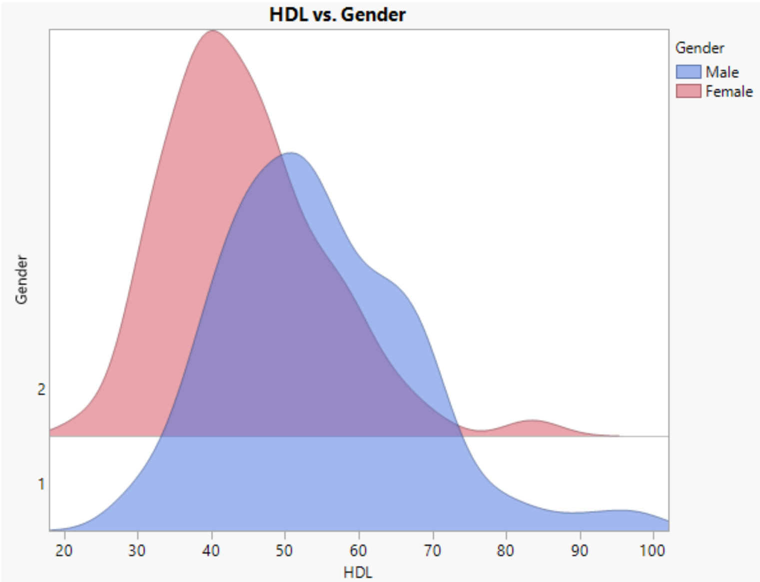


The histograms indicate that HDL levels are higher in males (blue) than females (pink). The blue distribution falls to the right, or higher on the HDL scale, than the pink distribution. Now change the graph to view the differences in gender in a ridgeline plot.

8. Select Gender and drag it to the Y zone.
9. In the Histogram properties panel, move the Overlap slider to overlap the histograms (until the slider is about 3/4 up).
10. Click **Done**.
11. Double-click in the legend. In the Legend Settings window, double-click 1 and type male. Double-click 2 and type female.
12. Click **OK**.

Tip: If an item in the legend is still highlighted, click it to deselect it.

Figure 4.25 Ridgeline Chart of HDL by Gender and Diabetes Level



The ridgeline chart shows the gender distributions offsets.

Example of a Bullet Bar Chart

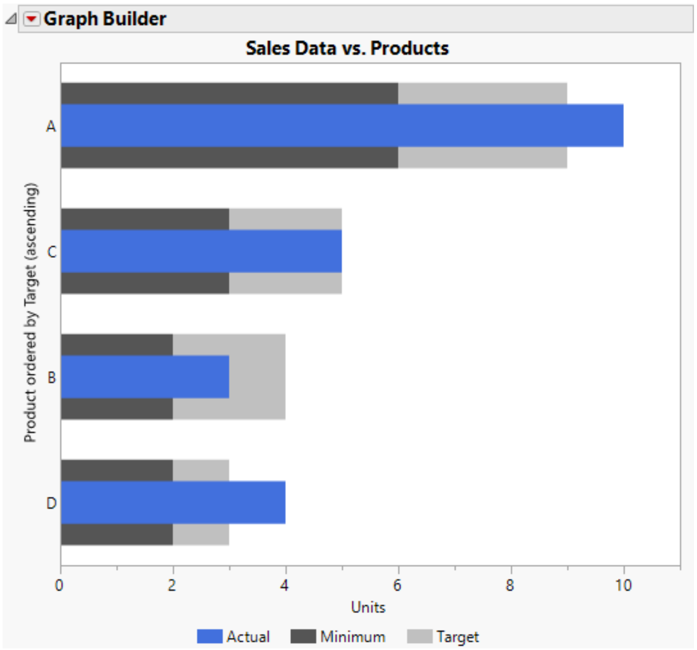
Data	This example uses fabricated sales data for four products.
Techniques	This example uses a bullet chart.
Goal	The goal of this example is to display actual units sold compared to a minimum and target number of units.

1. Select **Help** > **Sample Data Library** and open Bullet Measures.jmp.
2. Select **Graph** > **Graph Builder**.
3. Select Actual, Target, and Minimum and drag them to the **X** zone.
4. Select Product and drag it to the **Y** zone.
5. Click the **Bar** element icon .
6. In the Bar properties panel, perform these actions:
   – From the Bar Style list, select **Bullet**.

- Open **Variables**, and then highlight Minimum and click the Up arrow once.

7. Click the Graph Builder red triangle and select **Legend Position** > **Bottom**.
8. Right-click on a color in the legend and change the Fill Color:
   - No change to Actual
   - Change Minimum to dark gray
   - Change Target to light gray
9. Right-click in the Y axis > **Order By** > **Target, ascending**.
10. Change the titles of the X axis and graph:
    a. Click the graph title and enter Sales Data vs. Products.
    b. Click the X-axis title and enter Units.
11. Click **Done**.

Figure 4.26 Bullet Chart of Actual, Target, and Minimum Units Sold



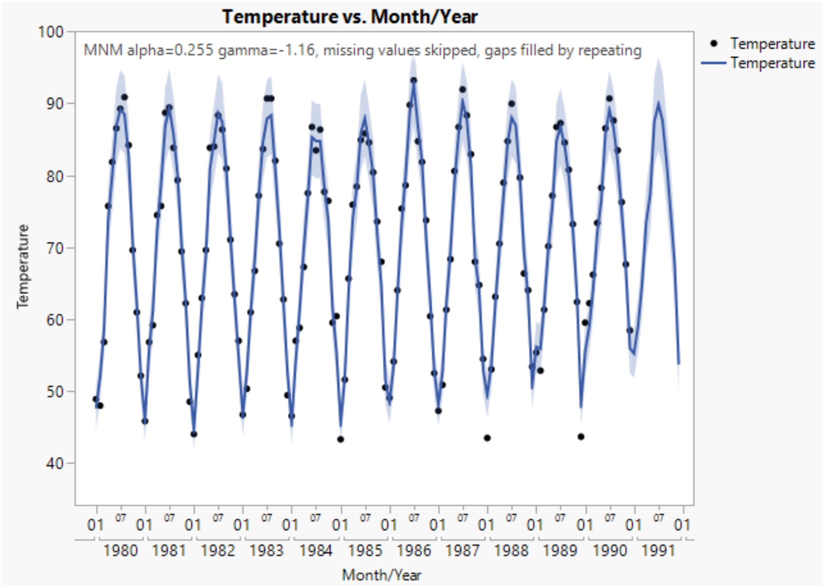
The products are ordered by the target number of units as shown by the light gray bars. Products A and D have actual sales that are above their targets; the blue bars extend beyond the light gray target bars. Sales for Product C are on target. Sales for Product B are below target but above the minimum number of units.

Example of a Forecast Plot

Data	This example uses maximum monthly temperature data from the US city of Raleigh, NC that were recorded from 1980 to 1990.
Techniques	This example uses a time series line of fit.
Goal	The goal of this example is to forecast the maximum monthly temperature for the following year.

1. Select **Help** > **Sample Data Library** and open Time Series/RaleighTemps.jmp.
2. Select **Graph** > **Graph Builder**.
3. Select Month/Year and drag it to the **X** zone.
4. Select Temperature and drag it to the **Y** zone.
5. Click the **Line of Fit** element icon .
6. In the Line of Fit properties panel, do these actions:
   – From the Fit list, select **Time Series**.
   – In the box next to Seasonal Period, type 12.
   – In the box next to Forecast Periods, type 12.
   This forecasts 12 months out. Since the data ends at 1990, it forecasts through 1991.
7. Right-click the X axis and select **Axis Settings**.
   – Under Tick/Bin Increment, change the Increment to 1.
   – Under Tick/Bin Increment, change the Label Row Nesting to 2.
   – Under Axis Label Rows, on the Label Row 1 tab, select **Labels** for the minor tick marks.
   – Click **OK**.
8. (Optional) To save the predicted forecast values and the lower and upper points of the forecast interval as new columns in the data table, click the Line of Fit red triangle and select **Save Formula**.
9. Click **Done**.

Figure 4.27 Forecast Plot of Maximum Monthly Temperature



The graph shows the data points used to estimate the time series function that extends through the forecast period. Note the following about this graph:

- The solid blue line represents the forecast model. The shaded blue area is the 95% confidence interval around the predicted values.
- The forecast period covers 1991, which is 12 months past the last data point. The forecast period is the right most portion of the plot that has no data points.
- The peak forecast high for 1991 is in July with an estimated value of 90°F. This estimate has a confidence interval of about 83°F to 96°F represented by the limits of the blue region above and below the estimate.
- The text at the top indicates the type of model that was fit, the parameter values for that model, and information about missing values. To remove this text from your graph, deselect the Forecast Model option in the Line of Fit properties panel.

Chapter 5

Bubble Plots

View Patterns in Multidimensional Data Using Bubble Plots

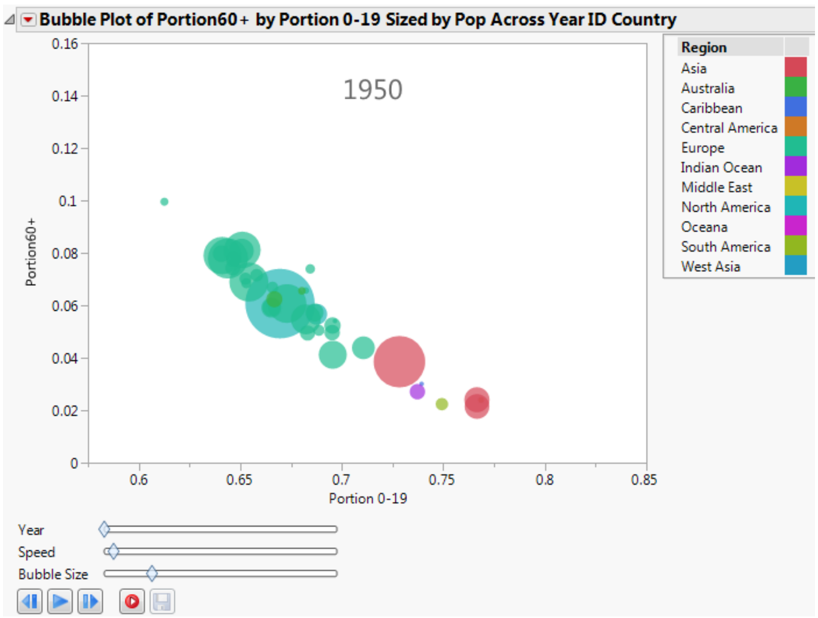
A bubble plot is a scatterplot that represents its points as circles, or bubbles. Bubble plots can be dynamic (animated over time) or static (fixed bubbles that do not move). Use bubble plots to:

- 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

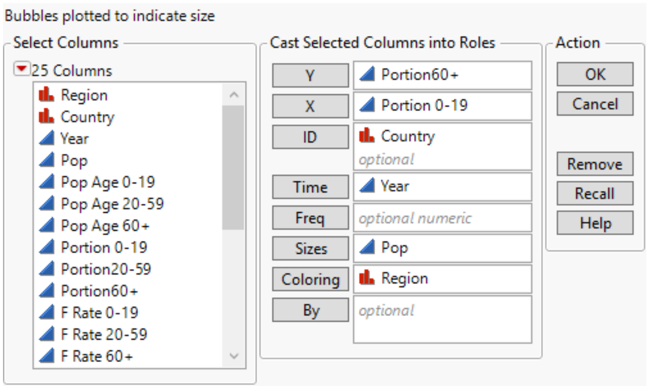
Example of a Dynamic Bubble Plot

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**.

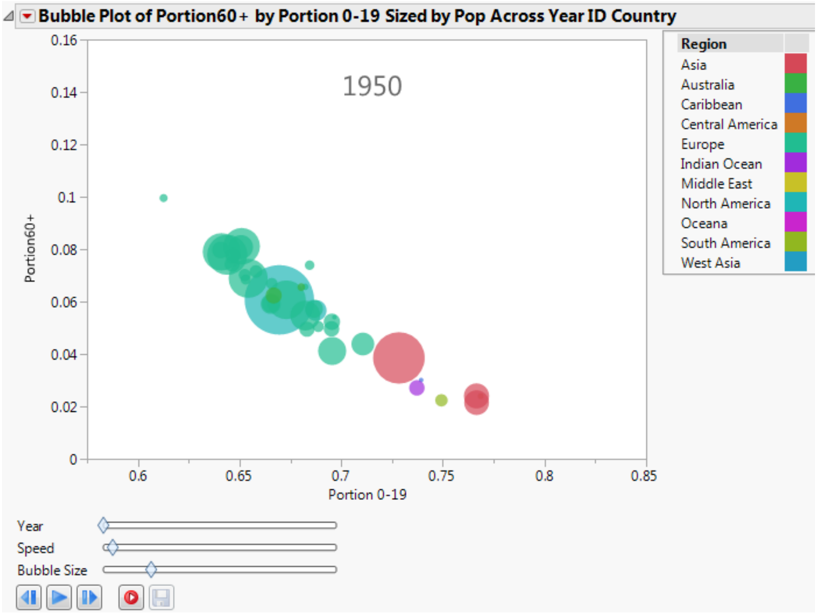
   The bubble plot shows a unique plot for each year's data.
7. Select Pop and click **Sizes**.

   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



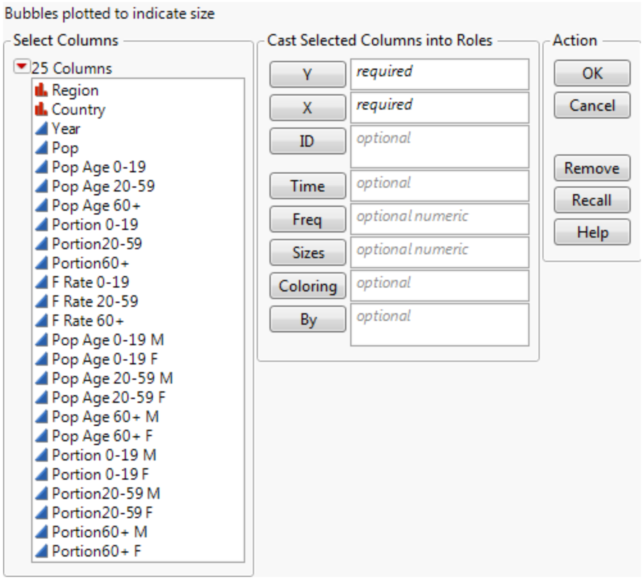
10. Click the play button to see the animated, dynamic report. Alternatively, you can click the previous button to move forward by one year.
11. (Optional) To view a legend that identifies each color with its region, click the Bubble Plot red triangle and select **Legend**.

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”.

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”.

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"
- "Example of Specifying Only a Time Variable"

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".
- *Dynamic mode*, where the bubbles are animated over time (a **Time** variable is specified). See "Example of a Dynamic Bubble Plot".

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

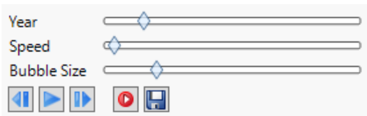


Table 5.1 Descriptions of the Animation Controls

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 *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 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”.

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”.

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”.

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 on the bubble plot.

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 *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.

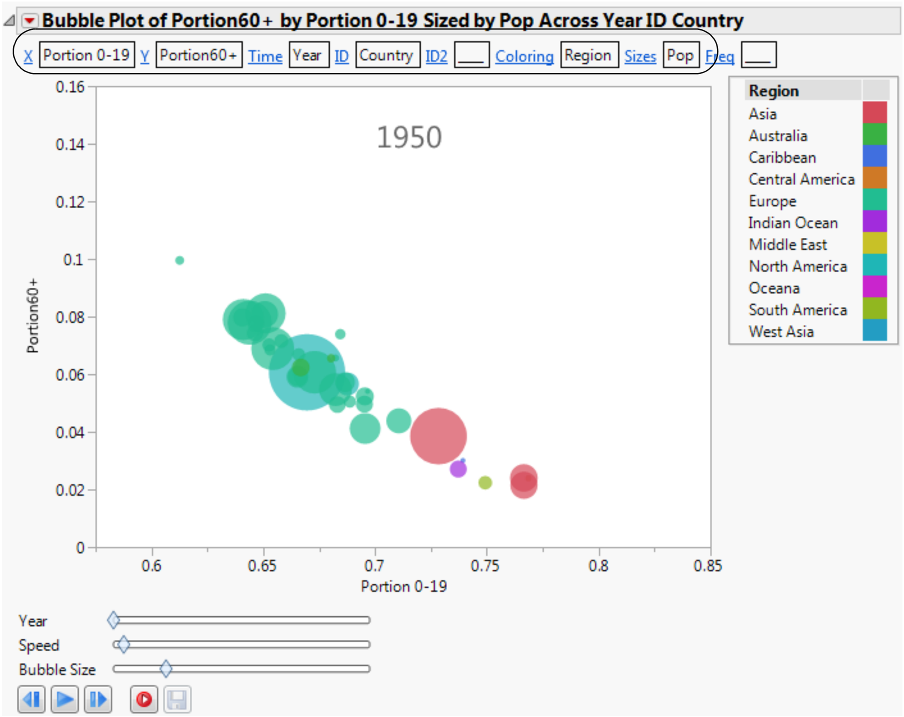
Save By-Group Script Contains options that enable you to save a script that reproduces the platform report for all levels of a By variable to several destinations. Available only when a By variable is specified in the launch window.

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" 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, perform these steps to change the **Coloring** variable in Figure 5.6:

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, perform these steps to remove the **Sizes** variable (Pop) in Figure 5.6:

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".
- 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

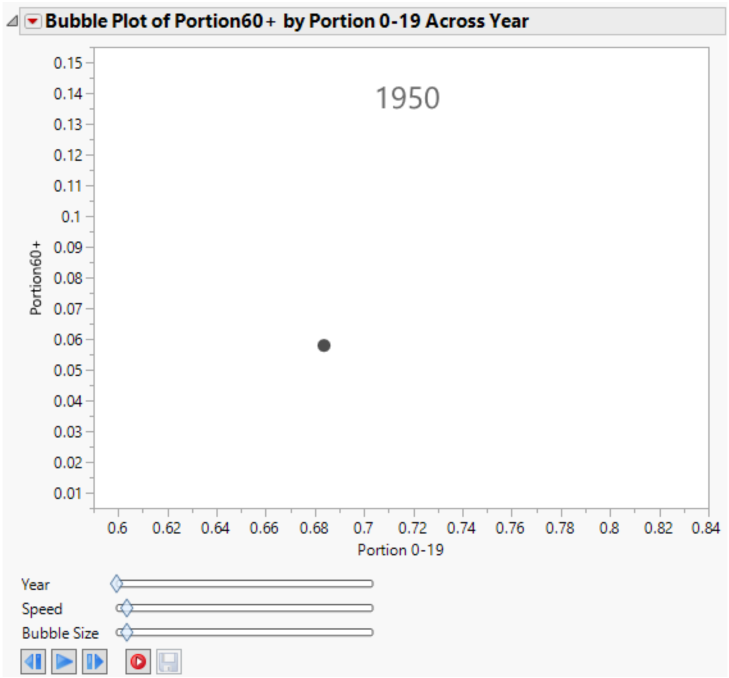
Example of Specifying Only a Time Variable

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

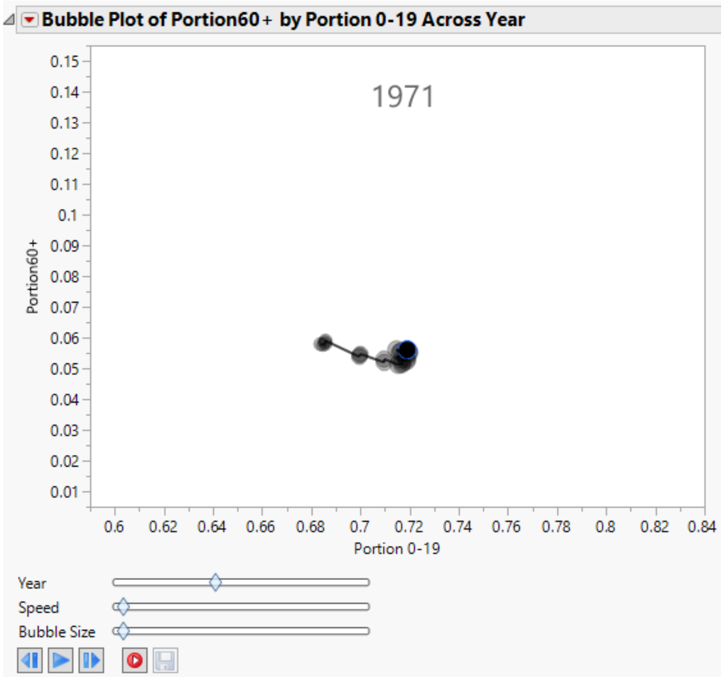


7. Click the bubble to select it.

   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



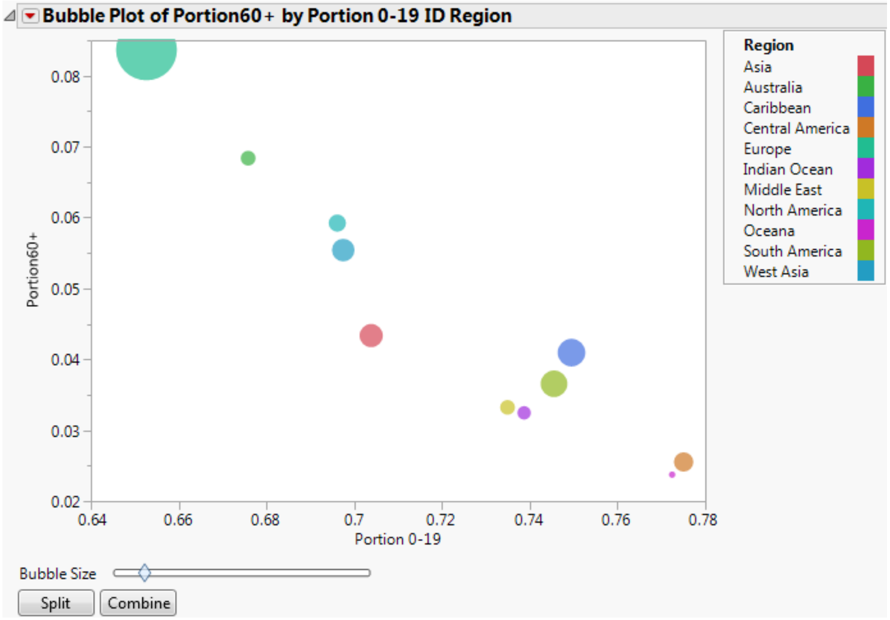
Example of Specifying Only ID Variables and Splitting a Bubble

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

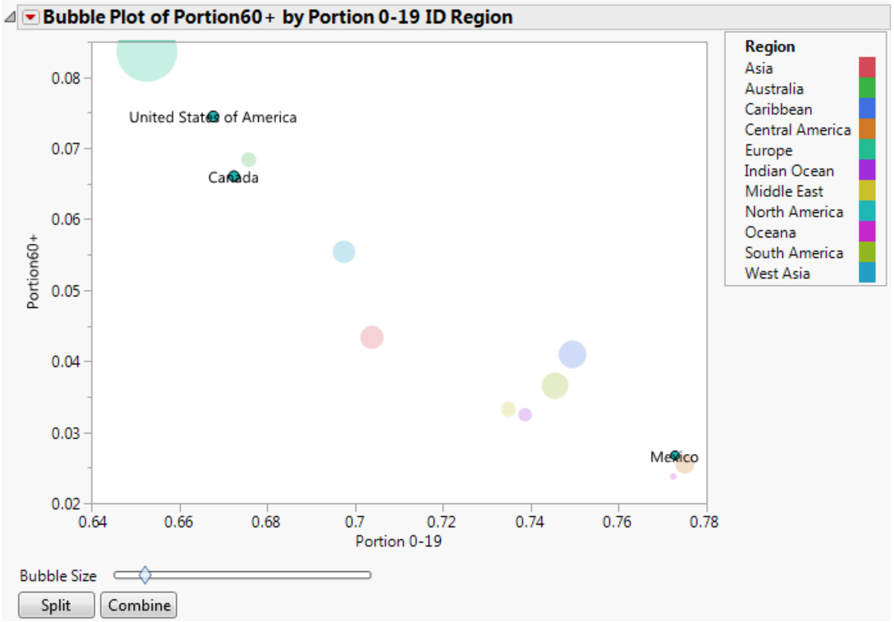


Split the bubble representing the region of North America into countries.

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



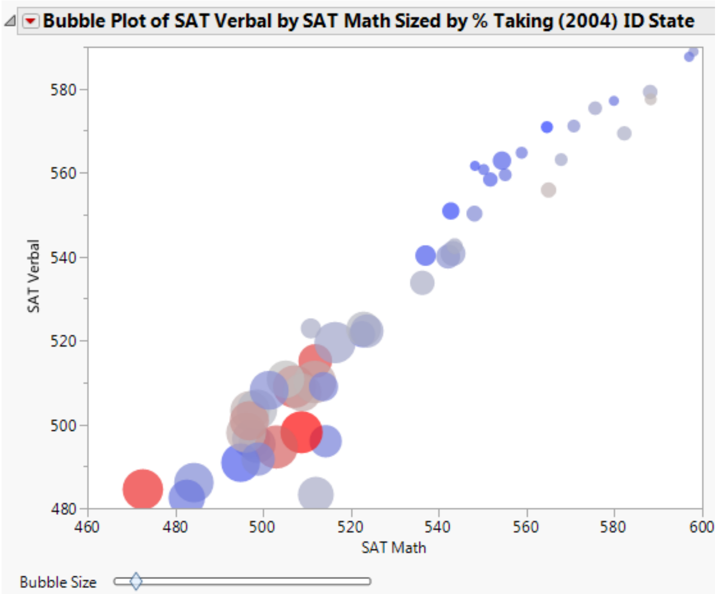
Example of a Static Bubble Plot

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.

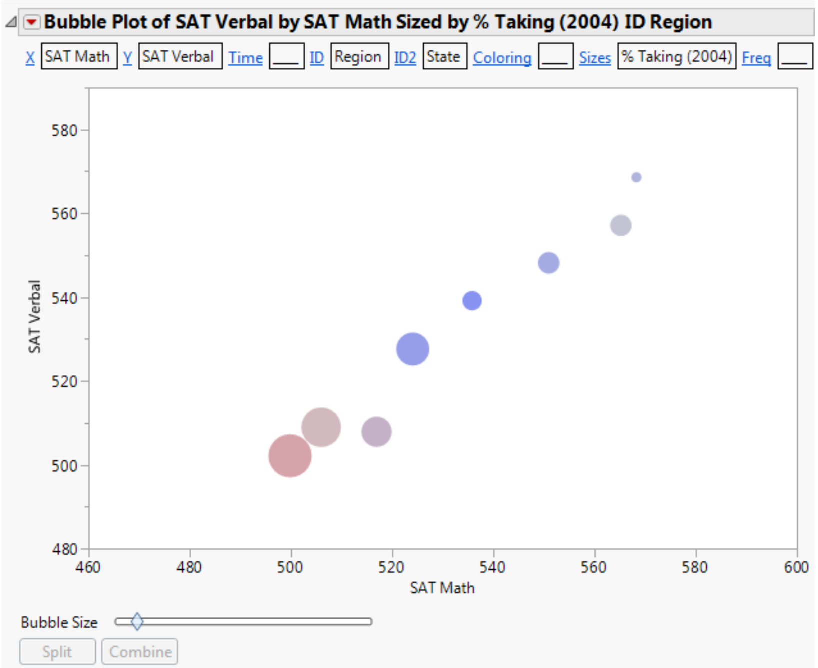
Instead of grouping the bubbles primarily by state, group the bubbles primarily by region:

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.

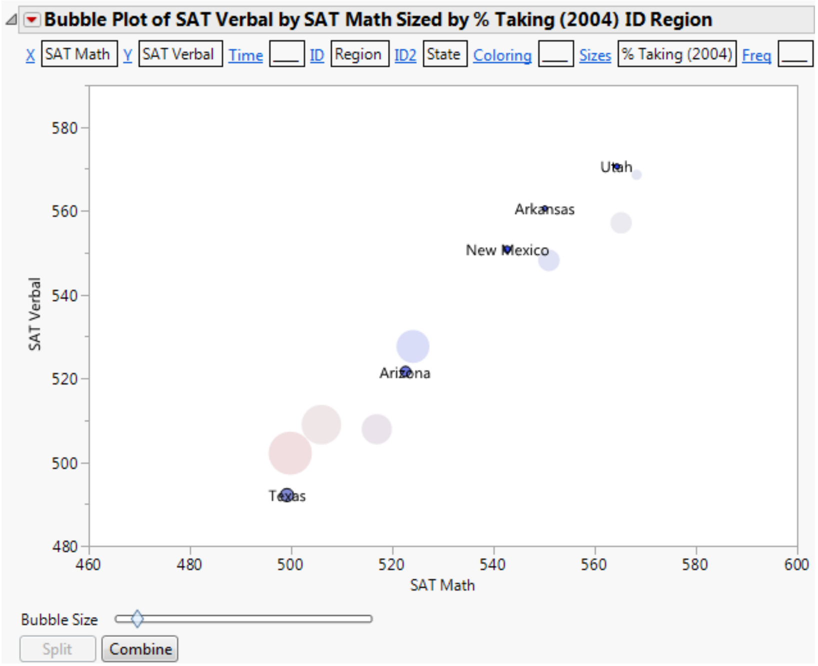
Figure 5.12 Example of Bubble Plot Grouped by Region and State



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.

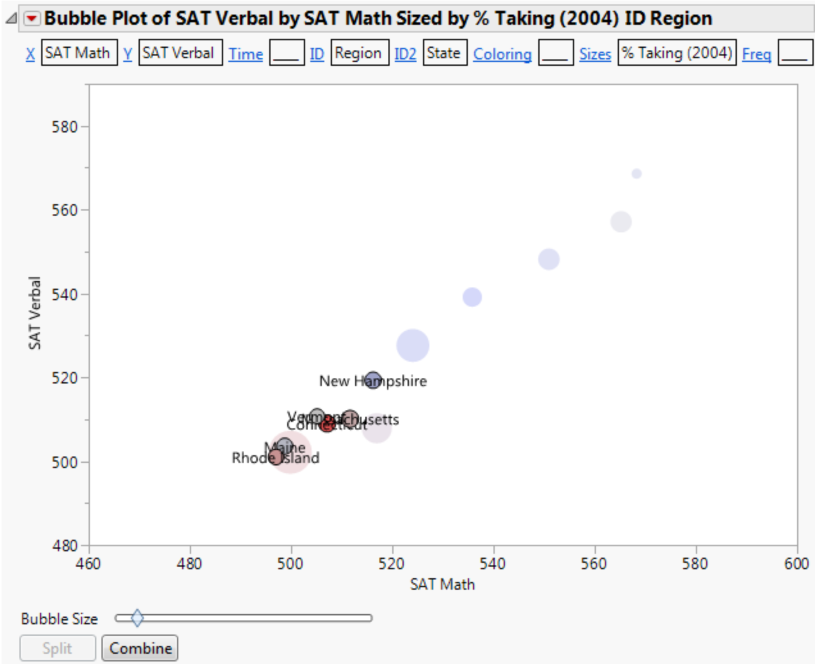
Figure 5.13 Example of Southwest Region Split by State



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.

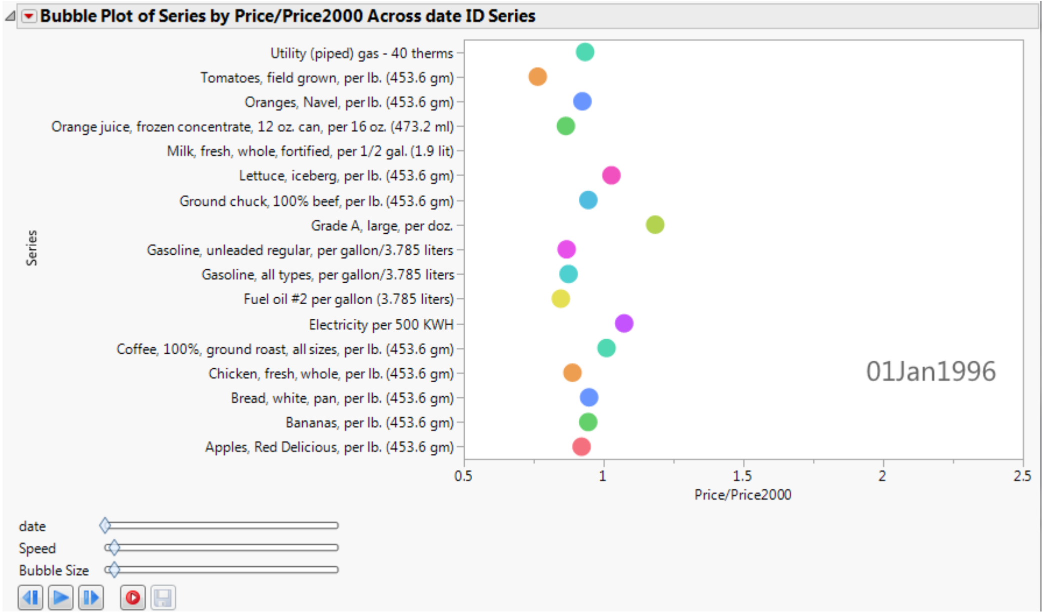
This example uses the blsPriceData.jmp sample data table, which shows the price of commodities over several years. Because the value of the US dollar changes over time, a column named Price/Price2000 shows the ratio of a commodity's price at any given time to the price in the year 2000.

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:

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.

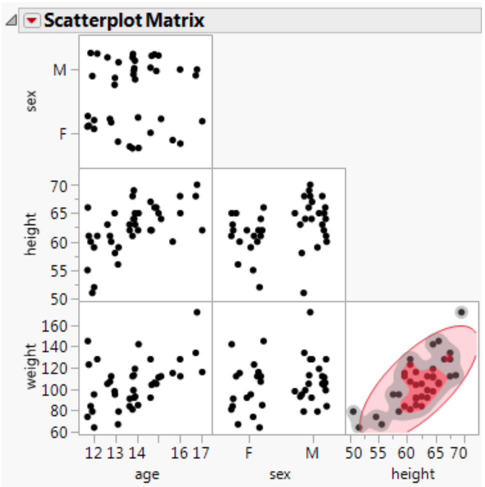
Chapter 6

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



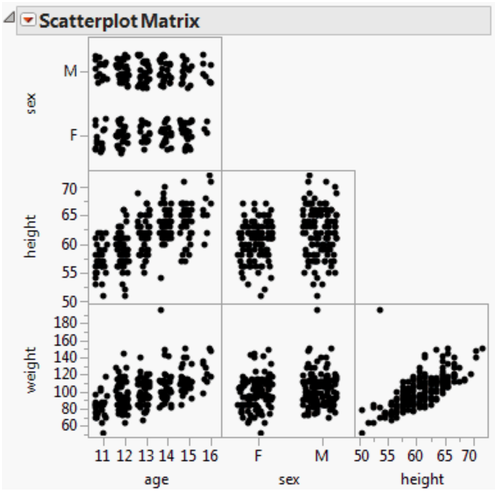
Contents

Example of a Scatterplot Matrix

This example shows you how to create a scatterplot matrix.

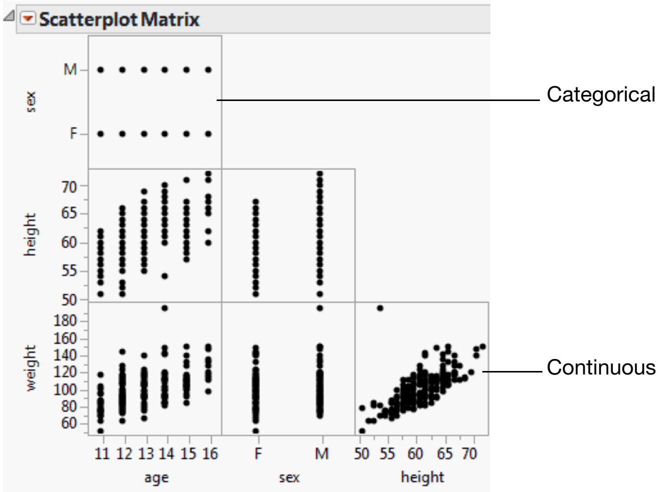
1. Select **Help > Sample Data Library** and open Students.jmp.
2. Select **Graph > Scatterplot Matrix**.
3. Select age, sex, height, and weight and click **Y, Columns**.
4. Click **OK**.

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

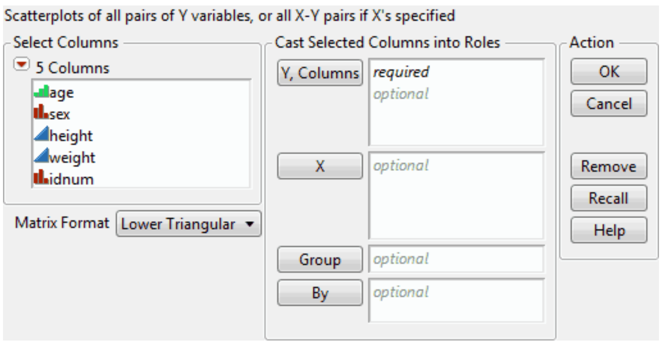


The weight versus height graph shows continuous data, and the sex versus age graph shows categorical data.

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".

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.

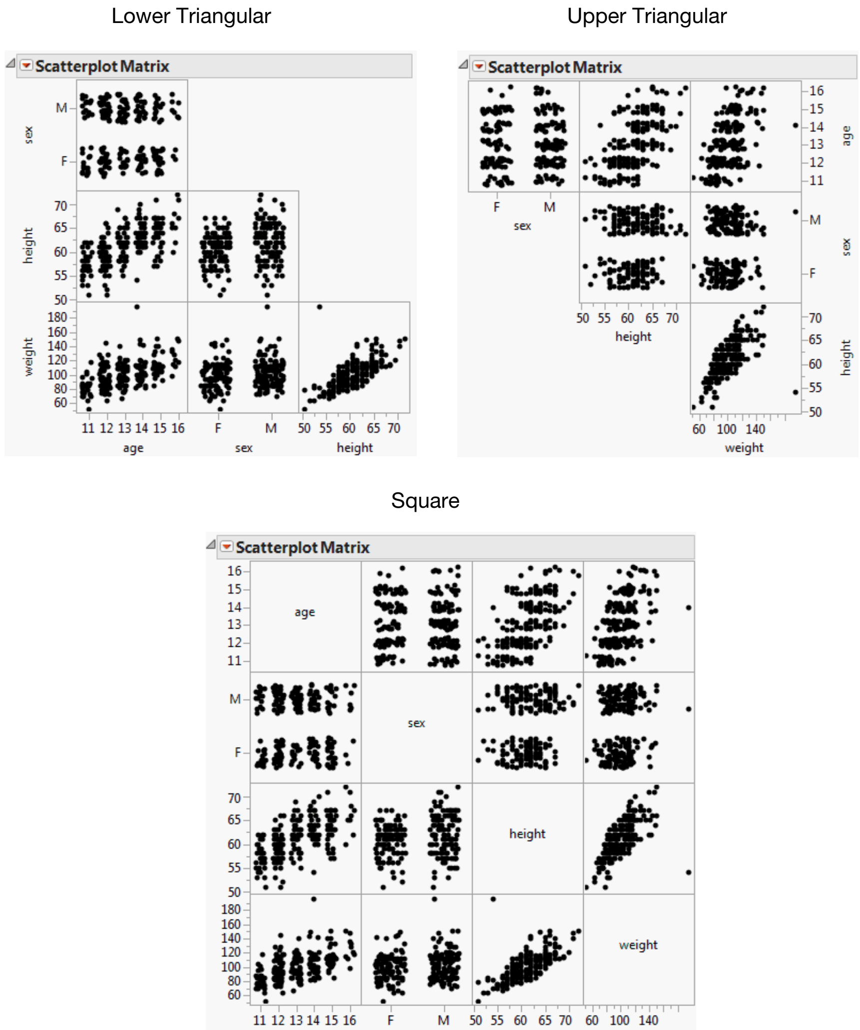
Matrix Format The **Matrix Format** can be one of three arrangements: **Upper Triangular**, **Lower Triangular**, or **Square**. See "Change the Matrix Format".

After you click **OK**, the Scatterplot Matrix window appears. See "The Scatterplot Matrix Window".

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



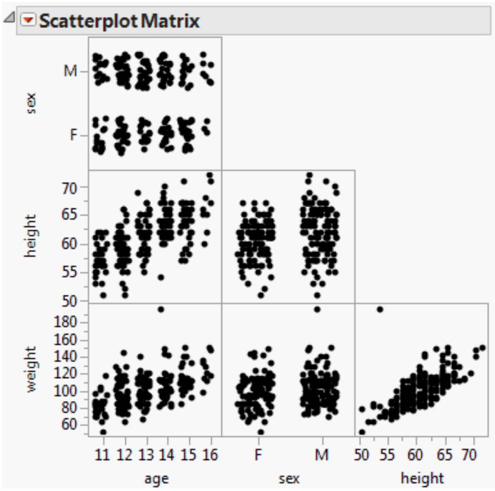
The Scatterplot Matrix Window

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" to produce the plot shown in Figure 6.6.

Note: For information about additional options, see "Scatterplot Matrix Platform Options".

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”.

Shaded Ellipses Colors the area within each ellipse. See “Example Using a Grouping Variable”.

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”.

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

See *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.

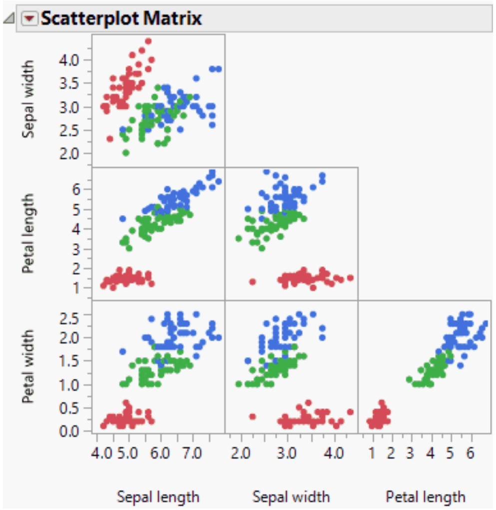
Save By-Group Script Contains options that enable you to save a script that reproduces the platform report for all levels of a By variable to several destinations. Available only when a By variable is specified in the launch window.

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**.

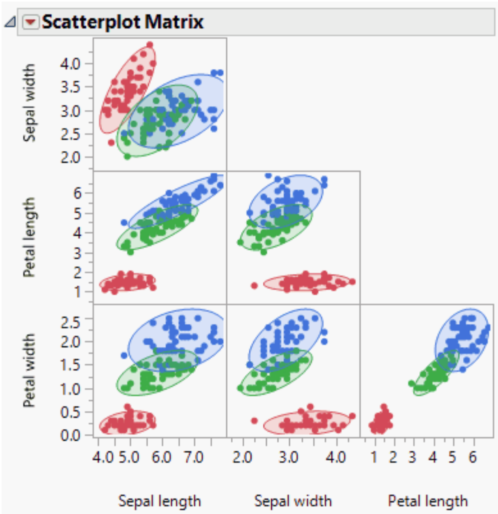
Figure 6.7 Initial Example Using a Grouping Variable



Make the groupings stand out by adding ellipses and shading:

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



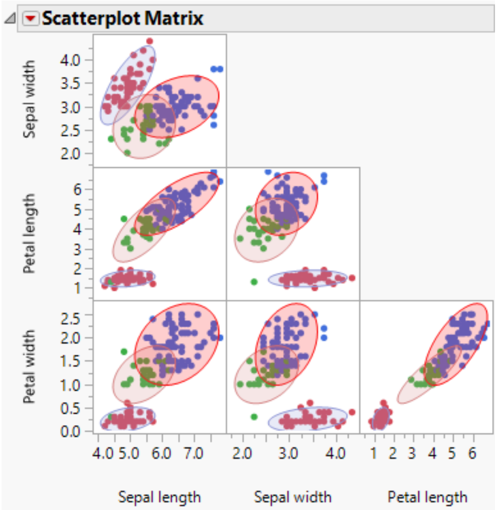
Create a Grouping Variable

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.

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 “Example Using a Grouping Variable”, but use Cluster as the grouping variable.

Figure 6.9 Example of a Scatterplot Matrix Using a Cluster Variable



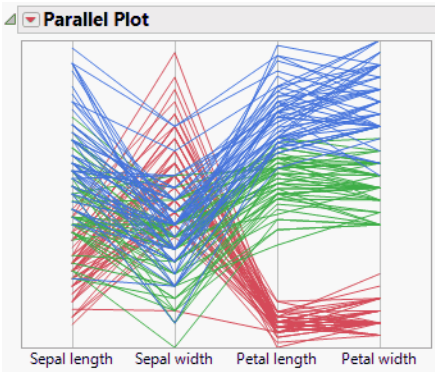
Chapter 7

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.

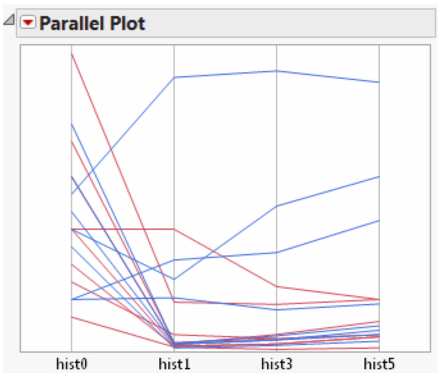
4. Click **OK**.

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



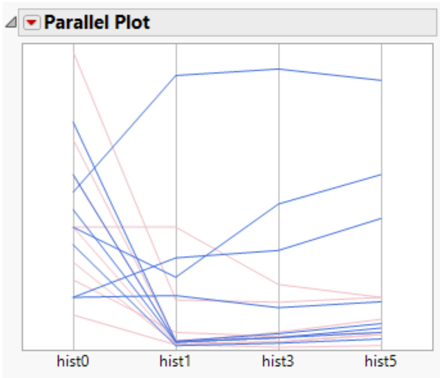
Each connected line segment represents a single observation. Click a line segment to see which observation (or row) it corresponds to in the data table.

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



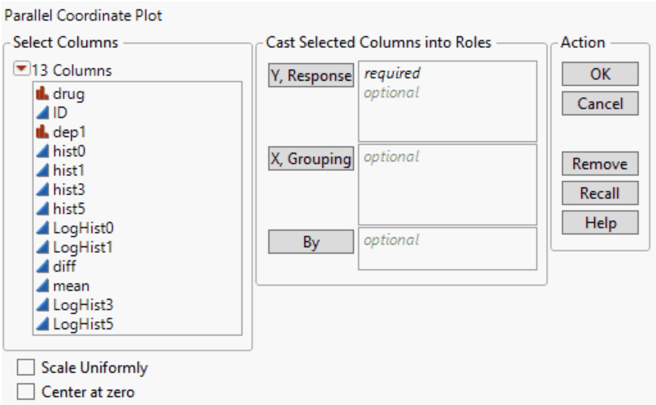
You observe the following about the histamine levels for dogs given trimeth:

- 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.

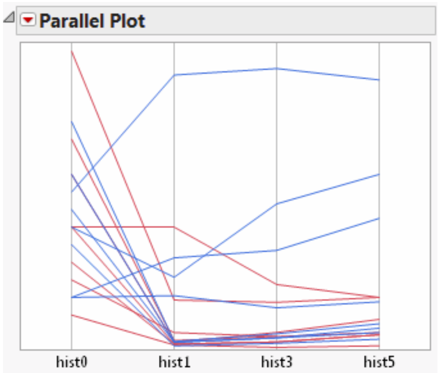
Center at zero Centers the parallel plot (not the variables) at zero.

After you click **OK**, the Parallel plot appears.

The Parallel Plot

To produce the plot shown in Figure 7.5, follow the instructions in "Example of a Parallel Plot".

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 *Using JMP*.

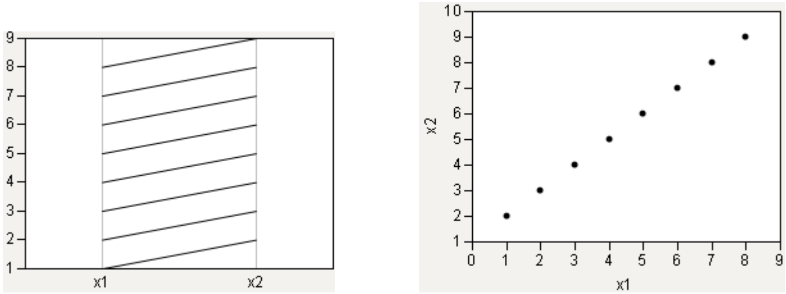
Interpreting Parallel Plots

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.

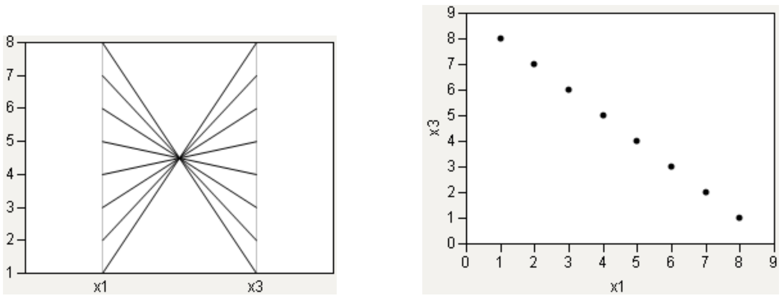
Figure 7.6 Strong Positive Correlation



Strong Negative Correlation

A strong negative correlation, by contrast, shows a narrow neck in the parallel plot.

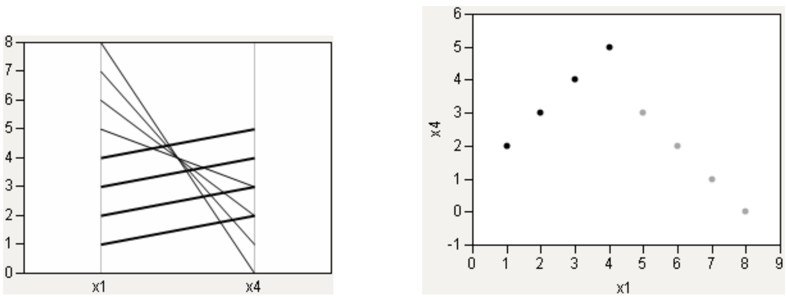
Figure 7.7 Strong Negative Correlation



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.

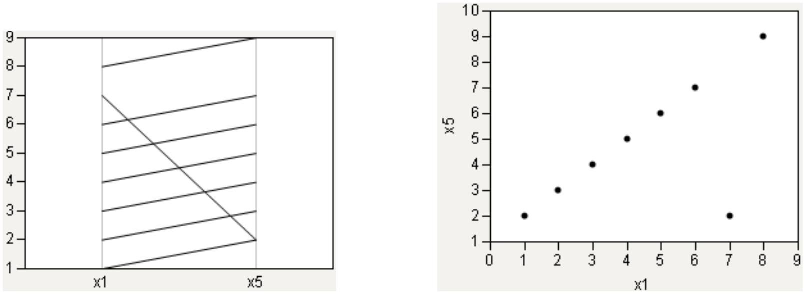
Figure 7.8 Collinear Groups: Parallel Plot and Scatterplot



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"

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 *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

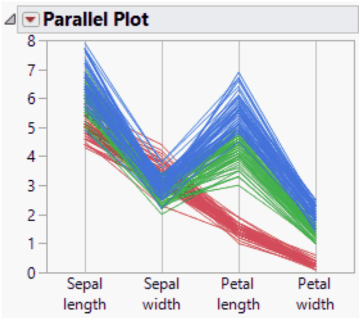
The following example uses the Fisher's Iris data set (Mardia, Kent, and Bibby 1979). The Iris.jmp sample data table contains measurements of the sepal length and width and petal length and width in centimeters for three species of Iris flowers: setosa, versicolor, and virginica. To find characteristics 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:

- Blue corresponds to virginica.
- Green corresponds to versicolor.
- Red corresponds to setosa.

From the parallel plot, you observe the following:

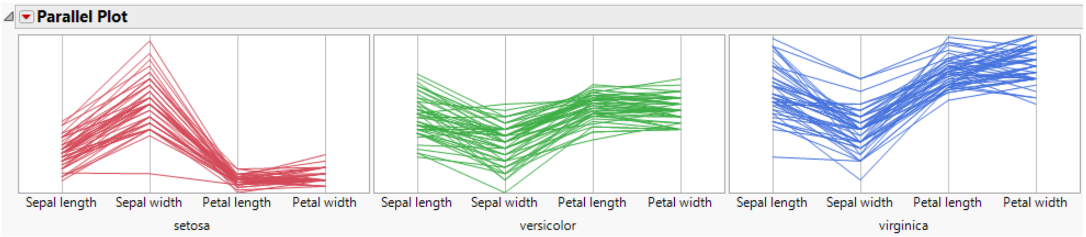
- For sepal width, the setosa values appear to be higher than the virginica and versicolor values.
- For petal width, the setosa values appear to be lower than the virginica and versicolor values.

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



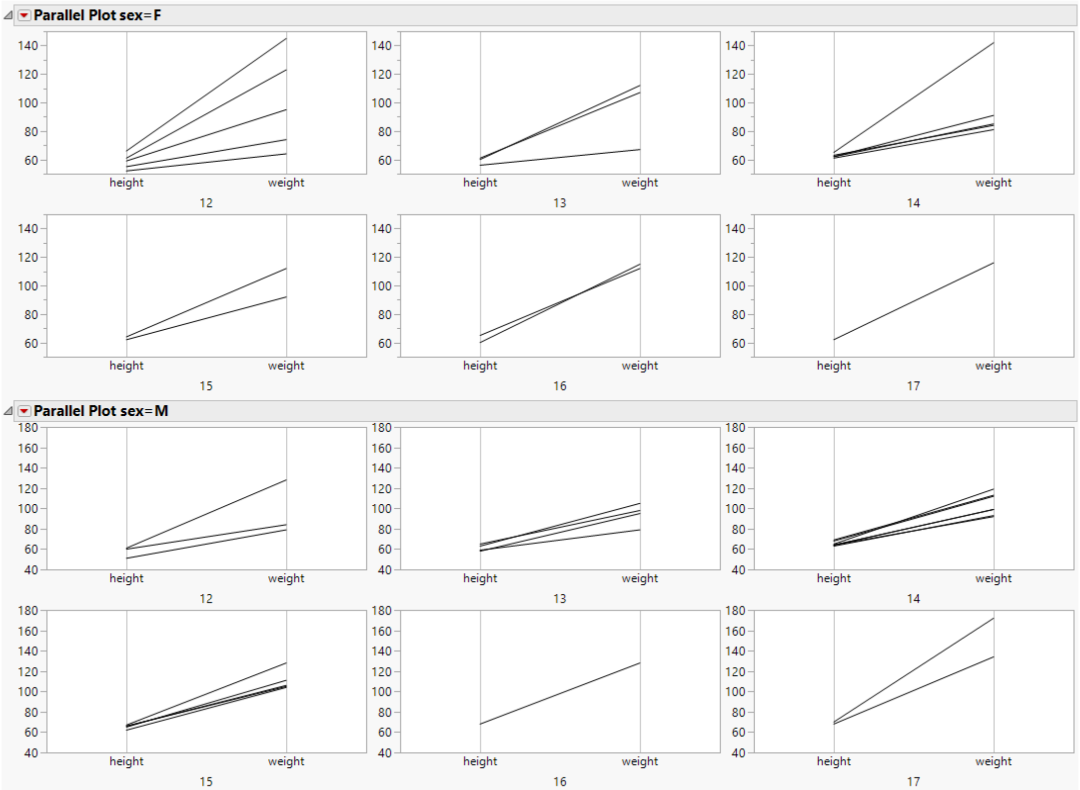
Each species is represented in a separate parallel plot.

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.

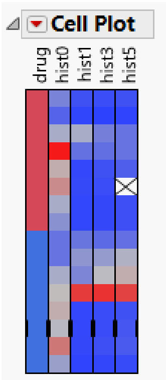
Chapter 8

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

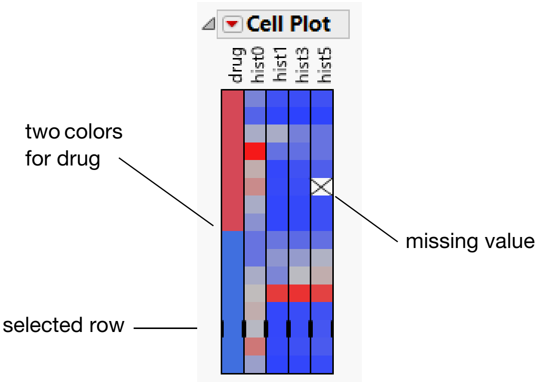
Example of a Cell Plot

Learn how to create and analyze a cell plot in JMP. 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



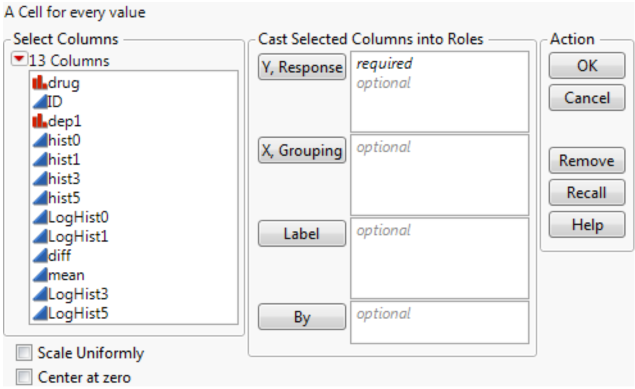
Notice the following:

- 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”.

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.

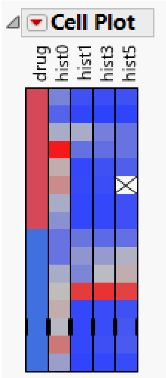
After you click **OK**, the Cell Plot window appears.

The Cell Plot

The cell plot appears with a one-to-one correspondence of a colored cell representing each data table entry. Colors are assigned to each cell based on the range and type of values found in the column.

Note: Any rows that are excluded in the data table are also hidden in the cell plot.

Figure 8.4 The Cell Plot



Note: To produce the plot shown in Figure 8.4, follow the instructions in "Example of a Cell Plot".

Note the following information about cell plots:

- 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.

Related Information

- "Additional Example of the Cell Plot Platform"

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 *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.

Right-Click Menu for Cell Plots

Right-click the cell plot to find the following options:

Graph Type Determines the appearance of the graph. See "Change the Graph Type".

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 *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".

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.

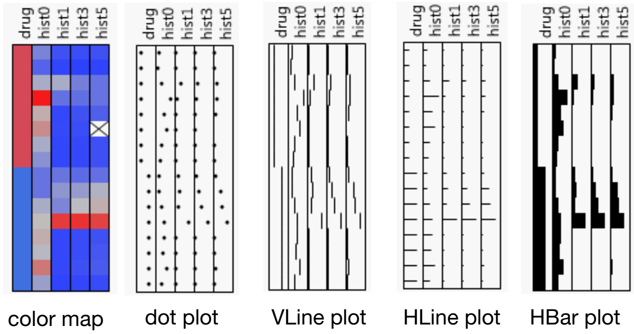
No Separator Lines Draws or removes lines separating the columns.

Note: For more information about the pop-up options that appear when you right-click labels, see *Using JMP*.

Change the Graph Type

Use the **Graph Type** option to change the appearance of the cell plot.

Figure 8.5 Graph Types



Additional Example of the Cell Plot Platform

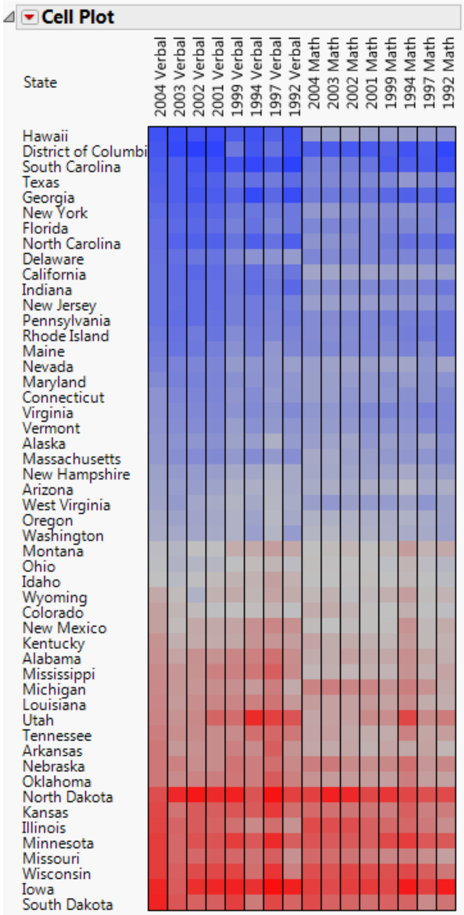
This example uses the SAT.jmp sample data table, which contains SAT test scores (divided into verbal and mathematics portions) for all 50 United States.

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 "Right-Click Menu for Cell Plots".

Chapter 9

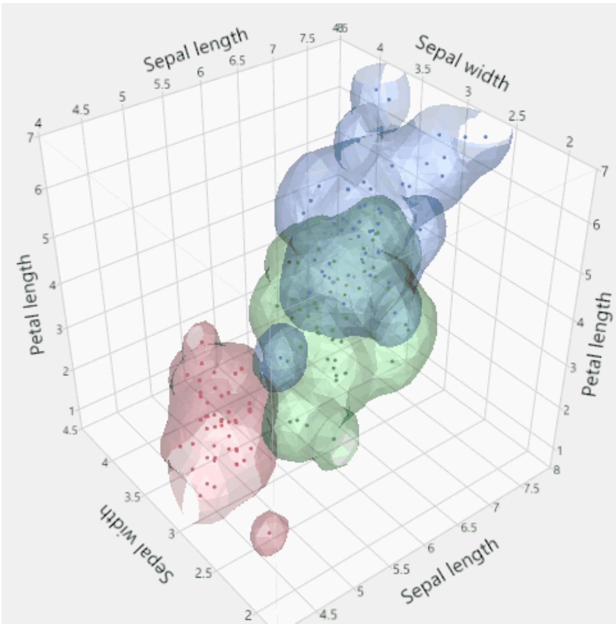
Scatterplot 3D

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



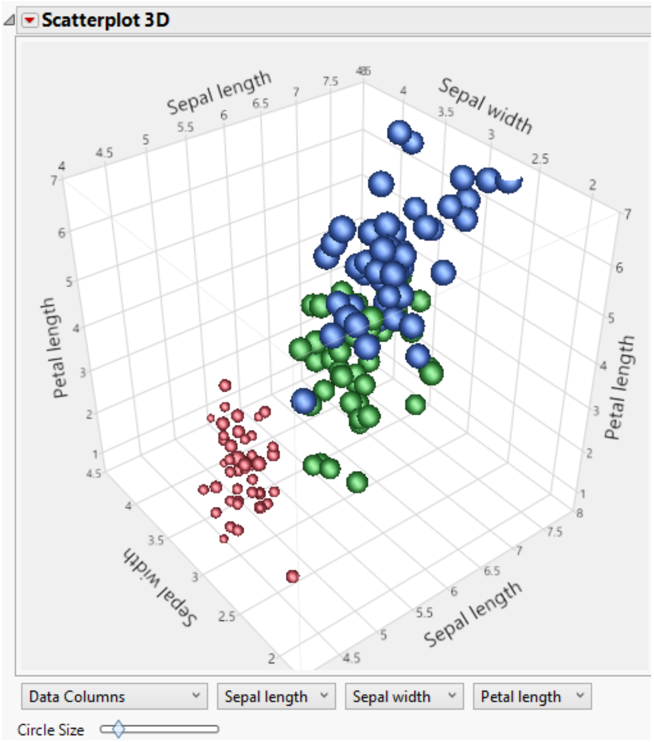
Contents

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

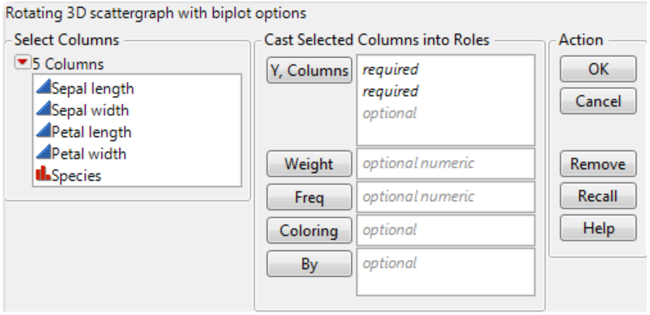


Now you can spin the 3D scatterplot to see the relationships between the variables. In this example, the data points are formatted in blue, red, and green. You might want to spin the scatterplot to see more clearly the relationships between the red and green points.

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.

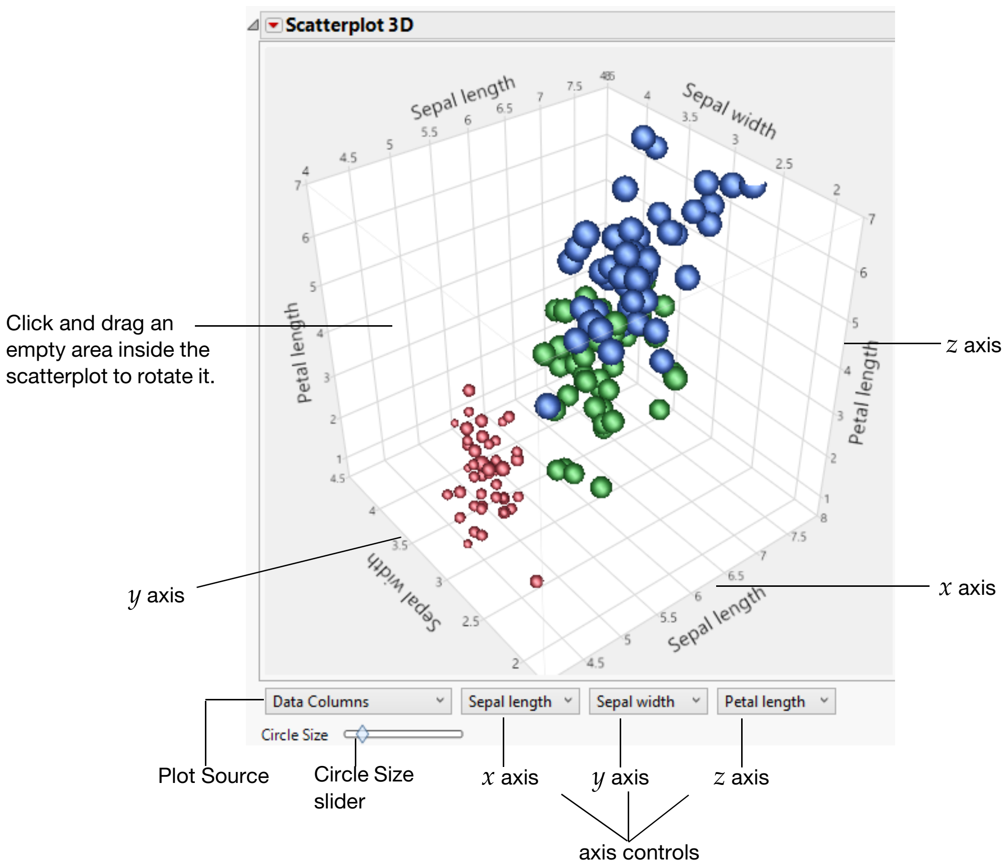
The Scatterplot 3D Report

To produce the 3D scatterplot shown in Figure 9.4, follow the instructions in "Example of a 3D Scatterplot".

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".

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".

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" and "Assign Colors and Markers in the Data Table".

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 selects only 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 the Esc key. 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 in any of these ways:

- Click and drag: Hold down the Shift key, 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 the Shift key, slide the wheel, and release the wheel. The 3D scatterplot spins up and down only.
- Arrow keys: Hold down the Shift key and press the arrow key, and then release Shift.
- Esc key: Press Shift+Esc. The 3D scatterplot spins left and right only.

In addition to automatically spinning the plot, you can oscillate the plot. Press Shift+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 the Esc key.

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. Hover over the middle of the axis.
2. Click and drag the axis.

To Modify Coordinate Scaling

1. Hover over the end of the axis.
2. Click and drag the axis.

To Rescale an Axis Precisely

1. Hover 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 *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.

Assign a color or marker to selected data points:

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") only highlights them for the current open graphs.

See *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".

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".

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".

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 *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 *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 *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 *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"
- "Example of Grouped Normal Contour Ellipsoids"

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".

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.

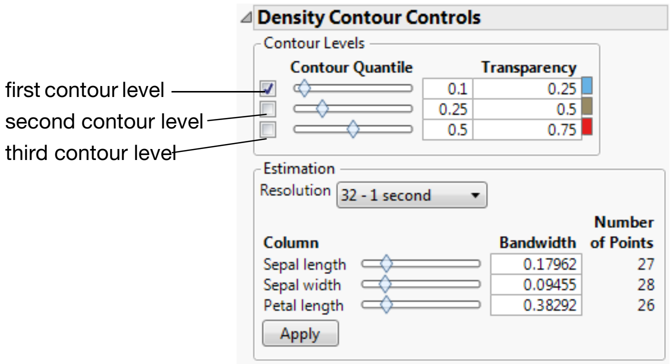
Related Information

- "Example of a Grouped Nonparametric Density Contour"

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.

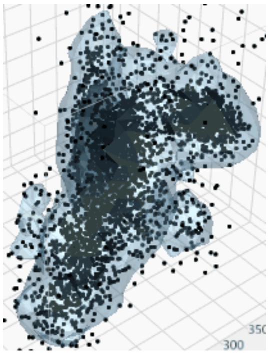
Type a new bandwidth for each variable, or click and drag the sliders. Click **Apply** to display your changes.

Optimizing a Dense Nonparametric Density Contour

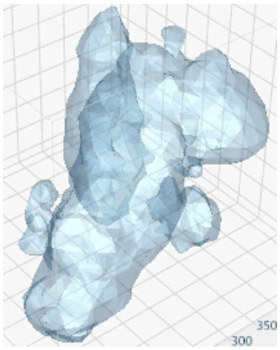
When you have many points on a 3D scatterplot, the contours can be so dark that you cannot see the structure. In this situation, you remove the points so that only the contours are displayed.

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".

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".

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 *Using JMP*.

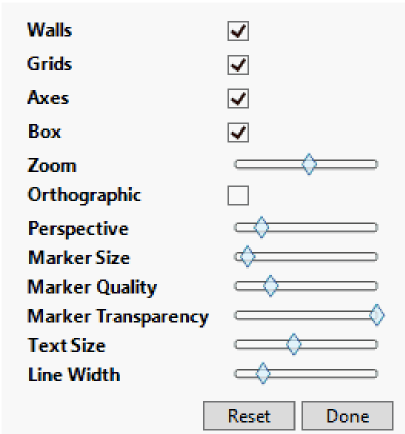
Use Hardware Acceleration Turns hardware acceleration on or off for machines that support acceleration. This option might display the scatterplot faster. If not, try updating your graphics drivers.

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

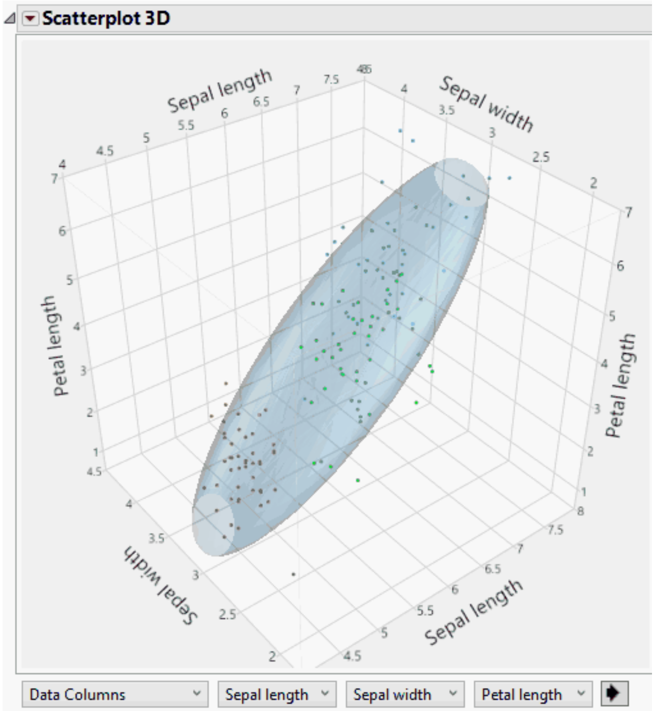
- "Example of an Ungrouped Normal Contour Ellipsoid"
- "Example of Grouped Normal Contour Ellipsoids"
- "Example of a Grouped Nonparametric Density Contour"

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

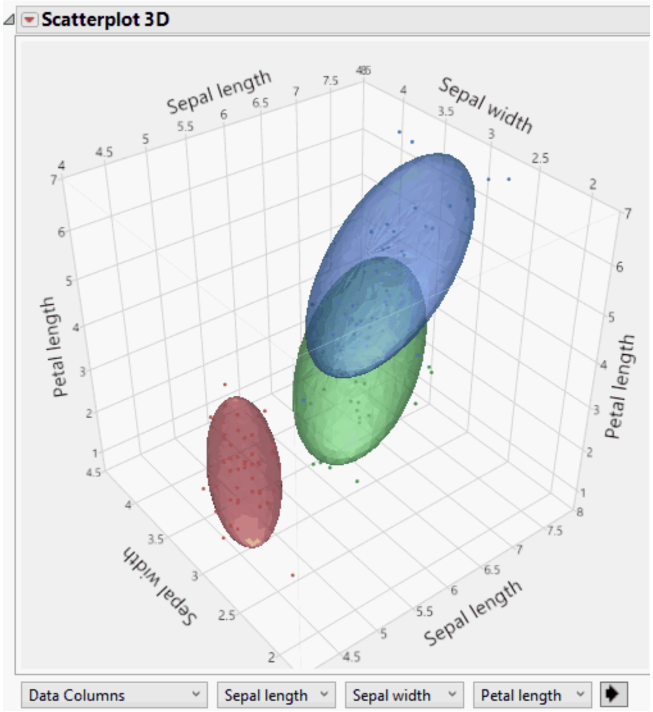


Example of Grouped Normal Contour Ellipsoids

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

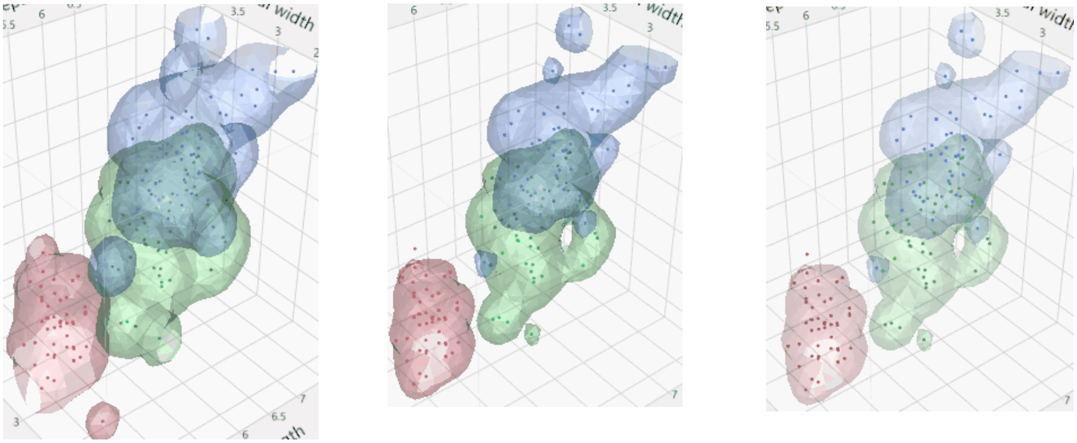


Example of a Grouped Nonparametric Density Contour

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.

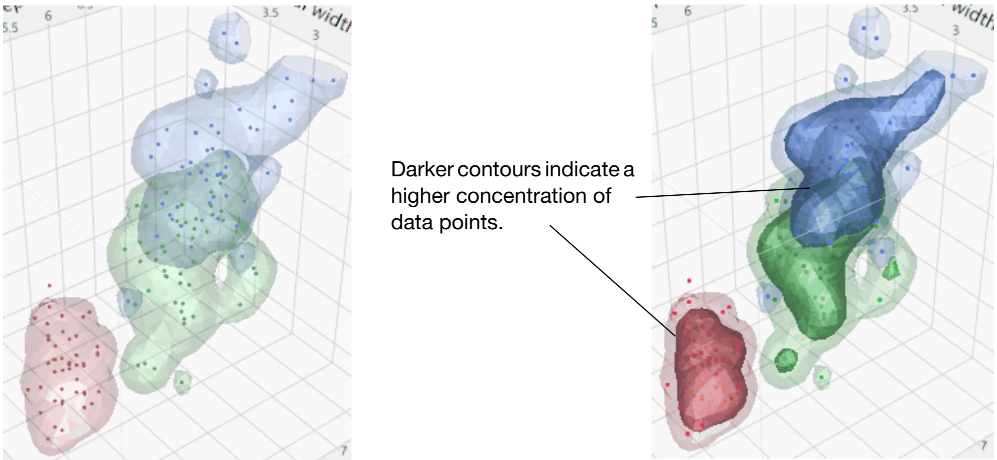
Figure 9.10 Changing the Nonparametric Density Contour Transparency and Density



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” for options.

Chapter **10**

Contour Plots

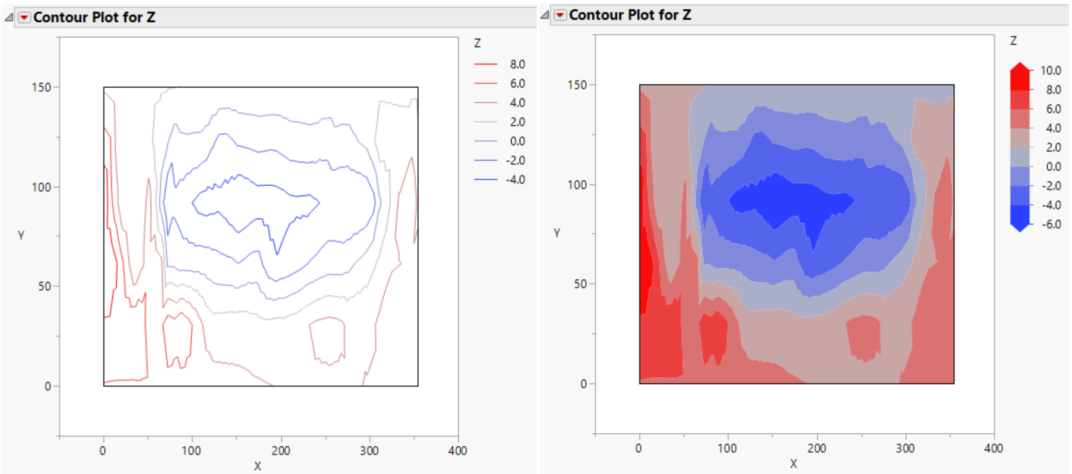
View Multidimensional Relationships in Two Dimensions

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".
- To see the example using Graph Builder, see "Example of a Contour Plot in Graph Builder".

Example Using the Contour Platform

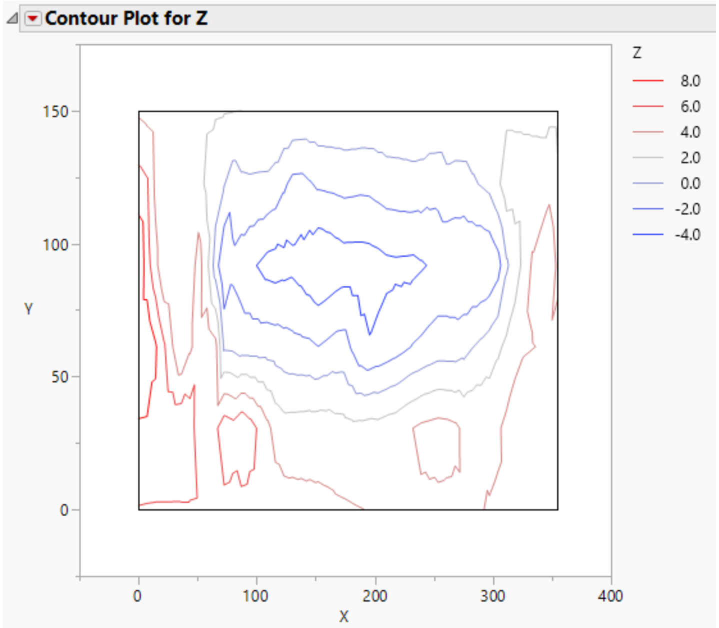
To create a contour plot, you need two variables for the *x*- and *y*-axes and at least one more variable for contours. You can also use several *y*-variables. This example uses the Little Pond.jmp sample data table. X and Y are coordinates of a pond. Z is the depth.

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



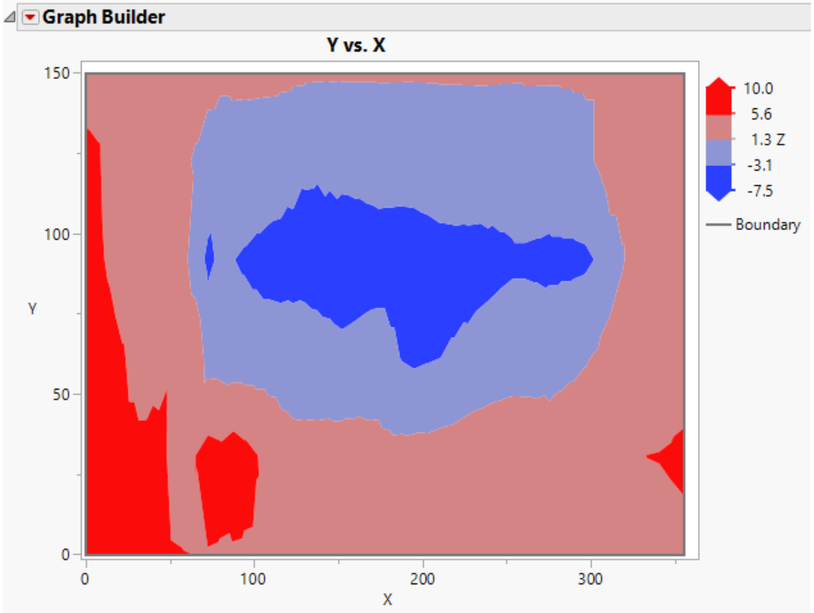
The *x*- and *y*-axes are coordinates and the contour lines are defined by the depth variable. This contour plot is essentially a map of a pond showing depth. To see the contours more clearly, click the red triangle next to Contour Plot for Z and select **Fill Areas**.

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



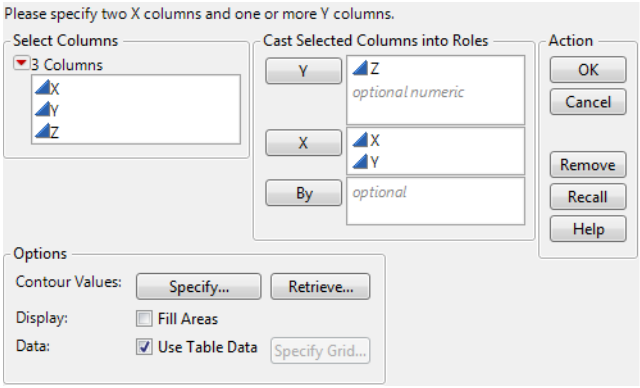
Notice the plot produced from the Contour platform looks slightly different than the plot produced using Graph Builder.

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”.

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”.

After you click **OK**, the Contour plot appears.

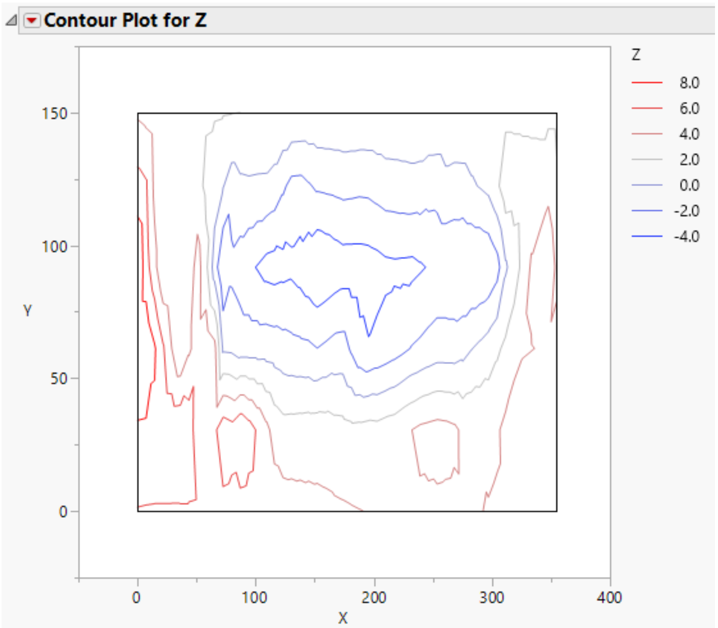
The Contour Plot

Follow the instructions in "Example of a Contour Plot" 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".

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".

> **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".

Label Contours Shows or hides the label (*z*-value) of the contour lines.

Fit to Window Specifies the stretching behavior of the report.

Change Contours Set your own number of levels and level increments. See "Contour Specification".

Save This menu has options to save information about contours, triangulation, and grid coordinates to data tables. See "Contour Plot Save Options".

See *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.

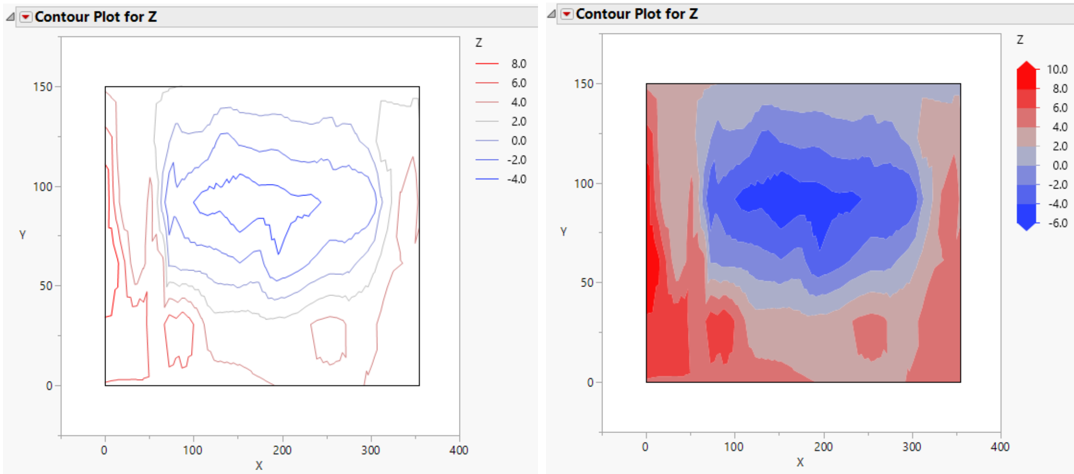
Save Script Contains options that enable you to save a script that reproduces the report to several destinations.

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

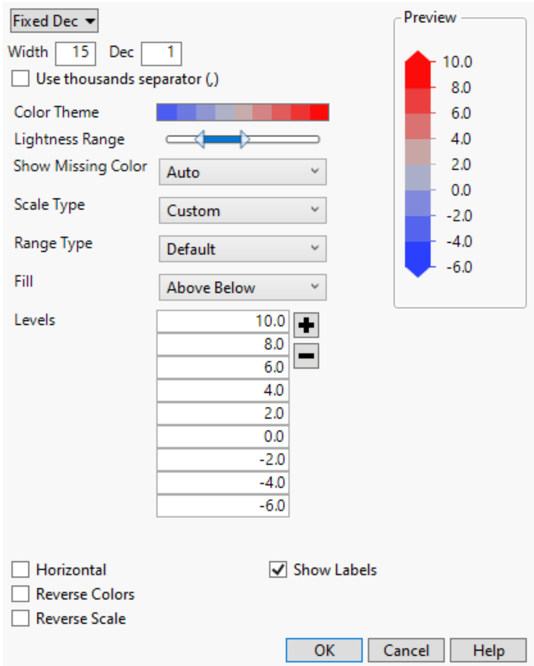


Areas are filled from low to high values on a color gradient. An additional color is added in the filled contour plot for the level above the last, and highest, contour line.

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 the contour color in the Contour Plot legend. You can customize the entire contour color gradient by right-clicking on the Contour Plot legend and selecting Gradient to open the Gradient Settings.

Figure 10.7 Gradient Settings Window



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

Color Theme Changes the color theme or defines a custom color theme. For more information about color options, see *Using JMP*.

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

Number of Labels (Not available if the Scale Type is Custom.) Specifies the number of labels for your legend. The value of zero provides the default number of labels.

Show Missing Color Specifies whether to show missing color in the legend. Auto shows the color only when there are missing values, Off never shows the color, On always shows the color.

Scale Type Sets the scale for the gradient coloring. A description of the scale type is found below the Maximum value setting.

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 gradient variable.

Standard Deviation The gradient variable range is divided into standard deviation offsets from the mean. 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 gradient variable.

Log Offset The scale is linear for the offset of the base 10 logarithm of the gradient variable.

Custom The scale labels are specified by the user in the Levels setting.

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.

Fill Determines how the endpoints of the range are specified in the contours.

Between The contours stop at the endpoint and any data beyond the endpoint is not shown.

Above Values above the range are included in the top contour level.

Below Values below the range are included in the bottom contour level.

Above Below Values at either end are included in their respective contour levels.

Levels (Available only when the Custom Scale Type is specified.) Enables you to specify the number of labels and the label values for the legend.

Horizontal Sets the orientation of the legend to horizontal.

Reverse Colors Reverses the colors in the color scheme.

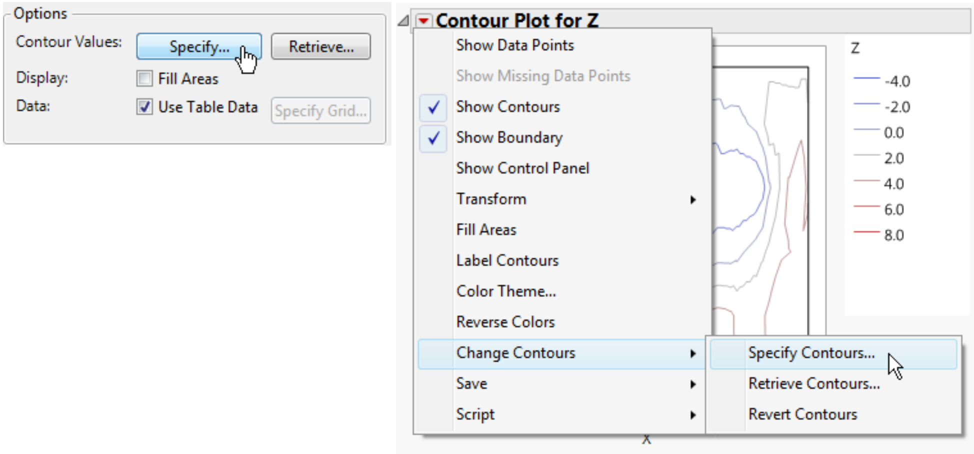
Reverse Scale Reverses the scale for the color theme.

Show Labels Shows or hides labels for the legend.

Specify

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). Selecting this option displays the Contour Specification window.

Figure 10.8 Example of Contour Specification: Launch Window (on the left) and Menu (on the right)

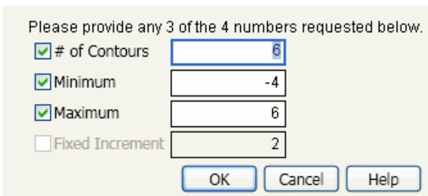


You can use the Contour Specification window to do the following:

- 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 10.9 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

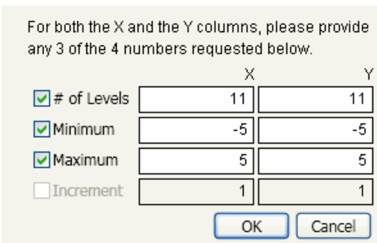
Save Triangulation Creates a new JMP data table that lists coordinates of each triangle used to construct the contours. By default, JMP uses Delaunay triangulation to connect the nearest data points to form triangles. The resulting set of triangles are calculated so that no other data points are inside 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 10.10 appears.

Figure 10.10 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.

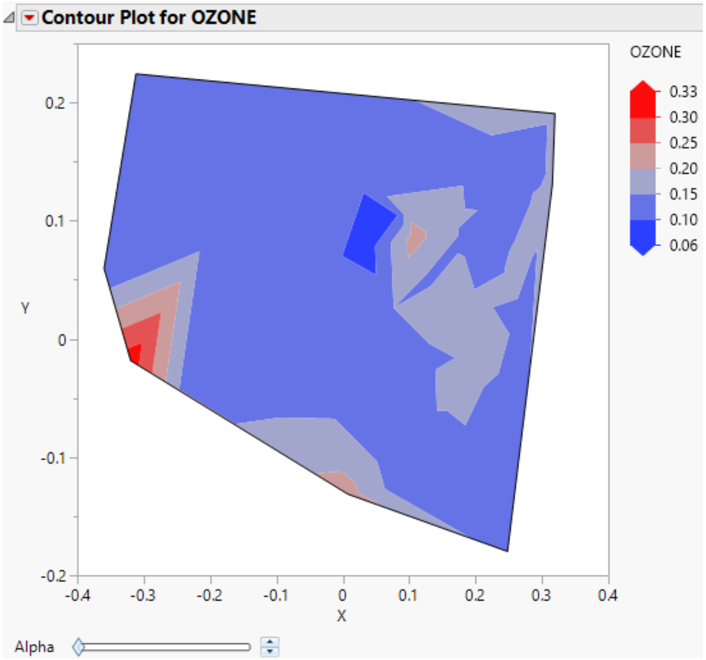
1. Select **Help > Sample Data Library** and open Cities.jmp.
2. Select **Graph > Contour Plot**.
3. Select OZONE and click **Y**.
4. Select X and Y and click **X**.
5. Select to **Fill Areas**.
6. Click **OK**.
7. Click the red triangle next to Contour Plot for OZONE and select the following options:
   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

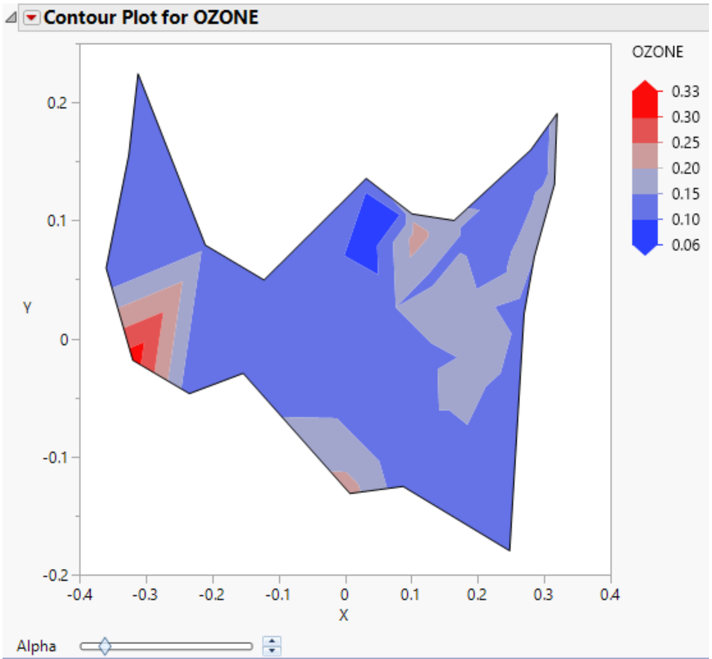
      The **Alpha** slider appears.

Figure 10.11 Contour Plot for OZONE



8. Click and move the **Alpha** slider to the right.

Figure 10.12 Alpha Shapes Filter



Using the Alpha slider filters out the larger Delaunay triangulation areas.

Chapter 11

Ternary Plots

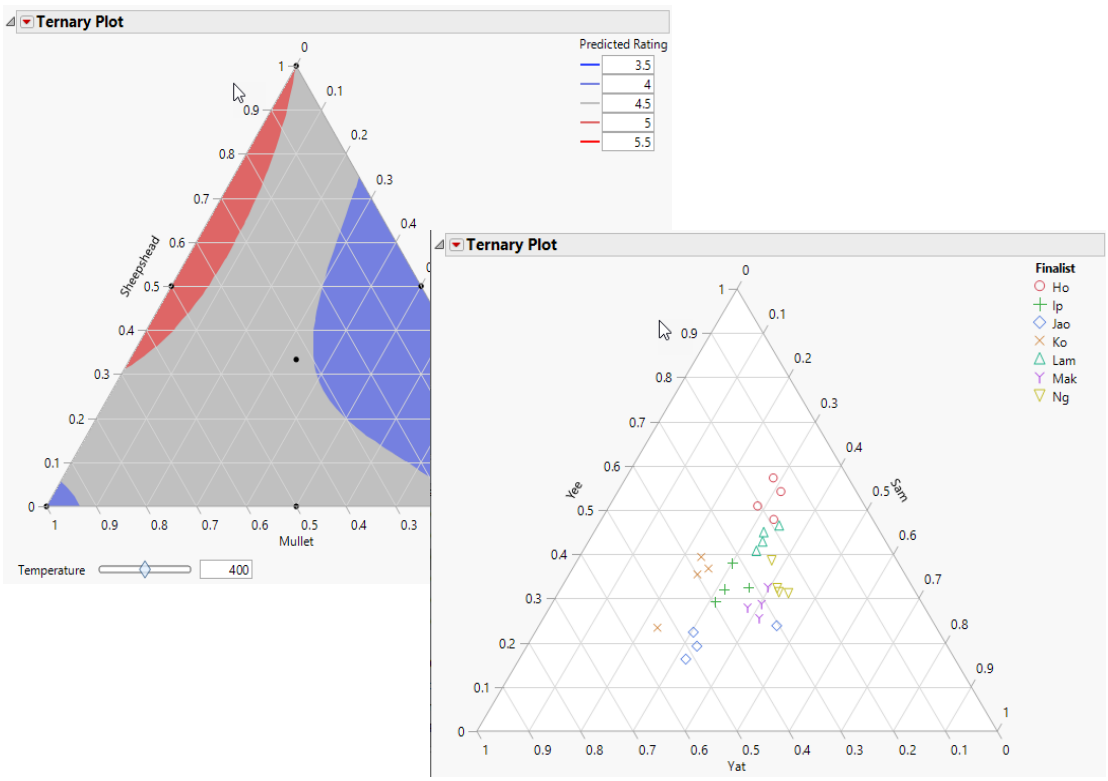
View Plots for Compositional or Mixture Data

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



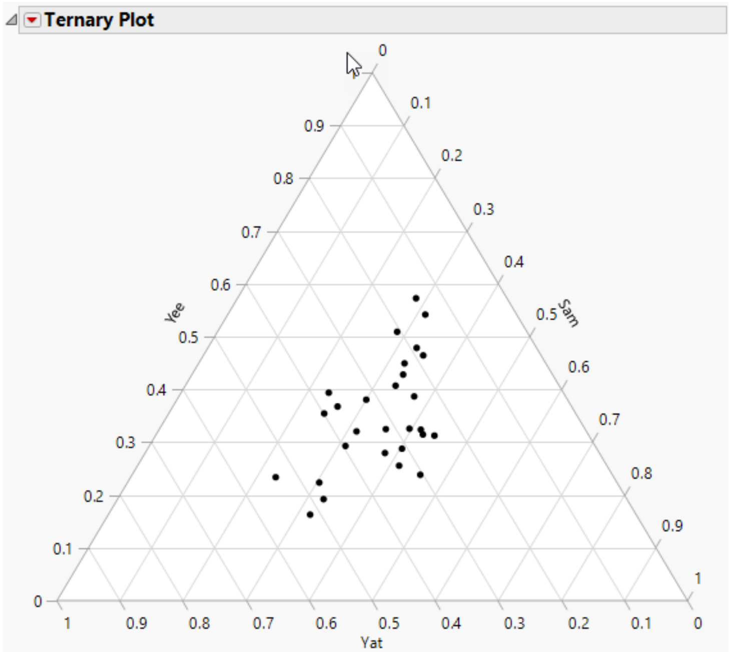
Contents

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*.

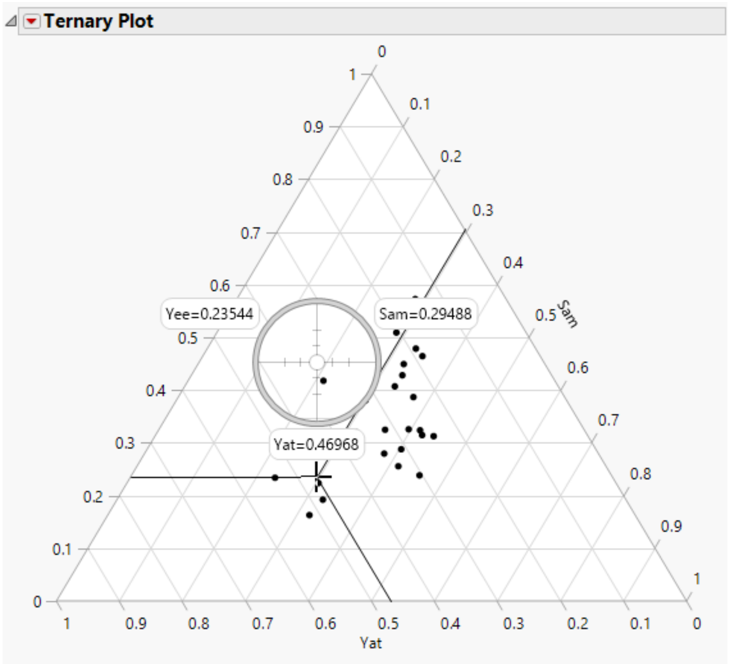
1. Select **Help > Sample Data Library** and open Pogo Jumps.jmp.
2. From the **Graph** menu, select **Ternary Plot**.
3. Select Yat, Yee, and Sam and click **X, Plotting**.
4. Click **OK**.

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

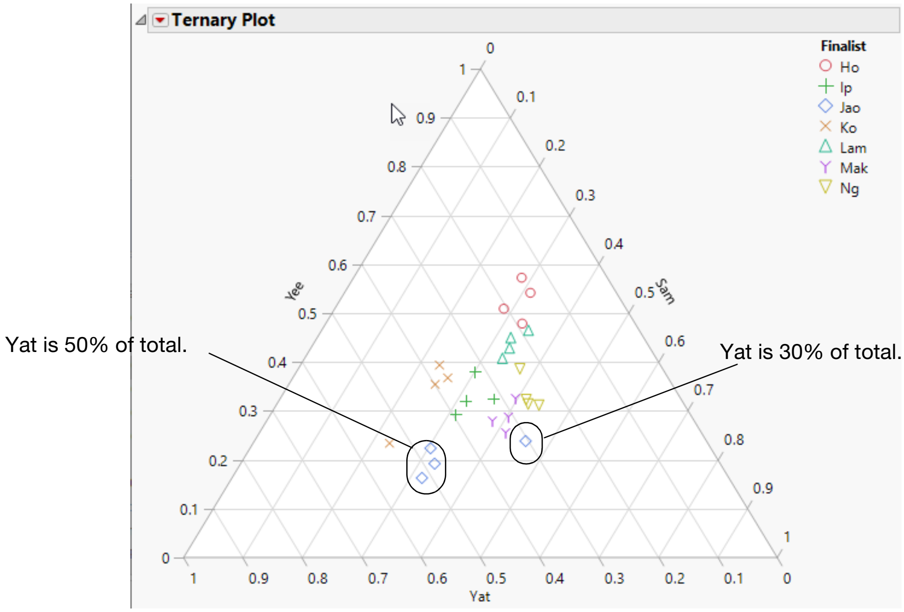


To get a better idea of how the three bounces contribute to total distance, assign each contestant's points a different color and marker.

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

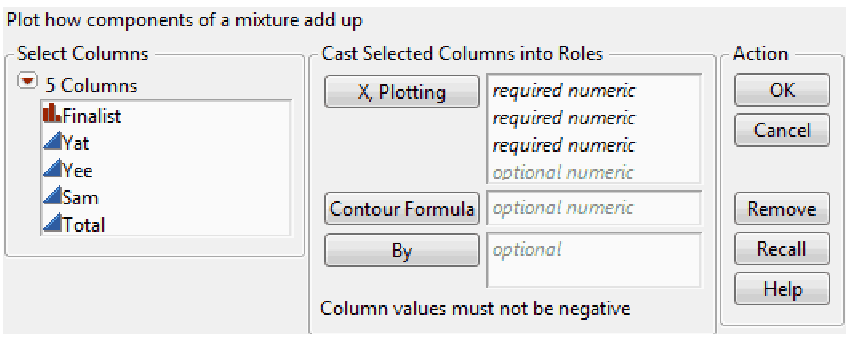


Note that most of the finalists are consistent in the composition of total distance. However, two finalists, Jao and Ko, both have one jump that is not consistent with their other jumps. For example, for three of Jao's jumps, the Yat composed about 50% of the total distance, but for the other jump, the 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".

By This option produces a separate graph for each level of the By variable.

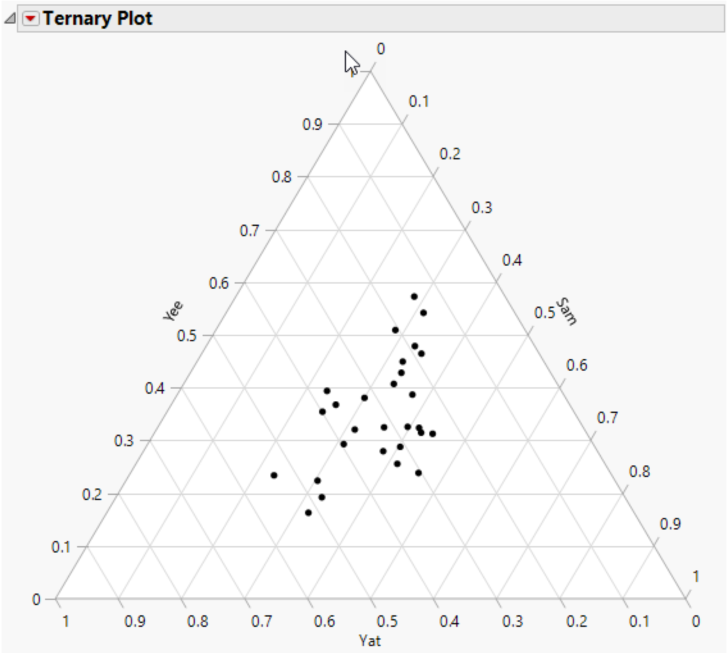
After you click **OK**, the Ternary Plot window appears.

The Ternary Plot

Follow the instructions in "Example of a Ternary Plot" 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 *Using JMP*.

Related Information

- "Example Using Mixture Constraints"

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.

Fit to Window Specifies 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.

See *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"

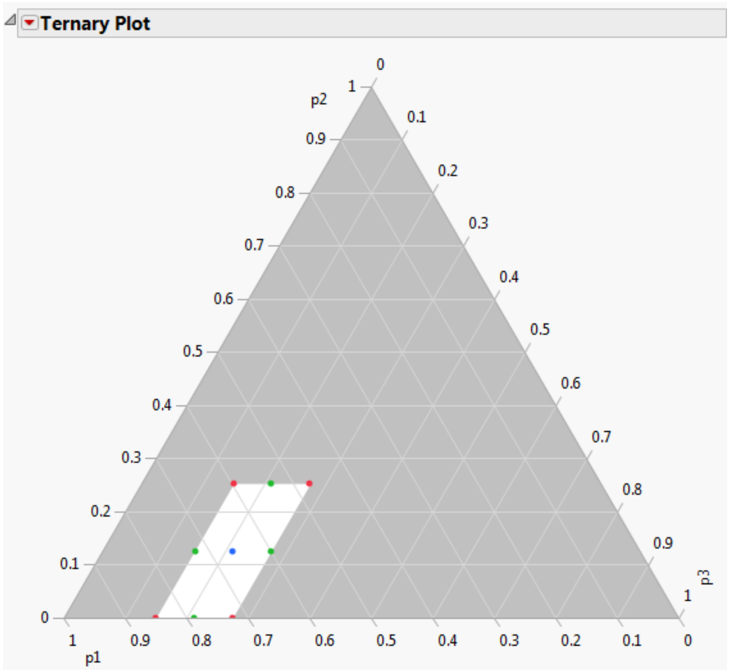
Example Using Mixture Constraints

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



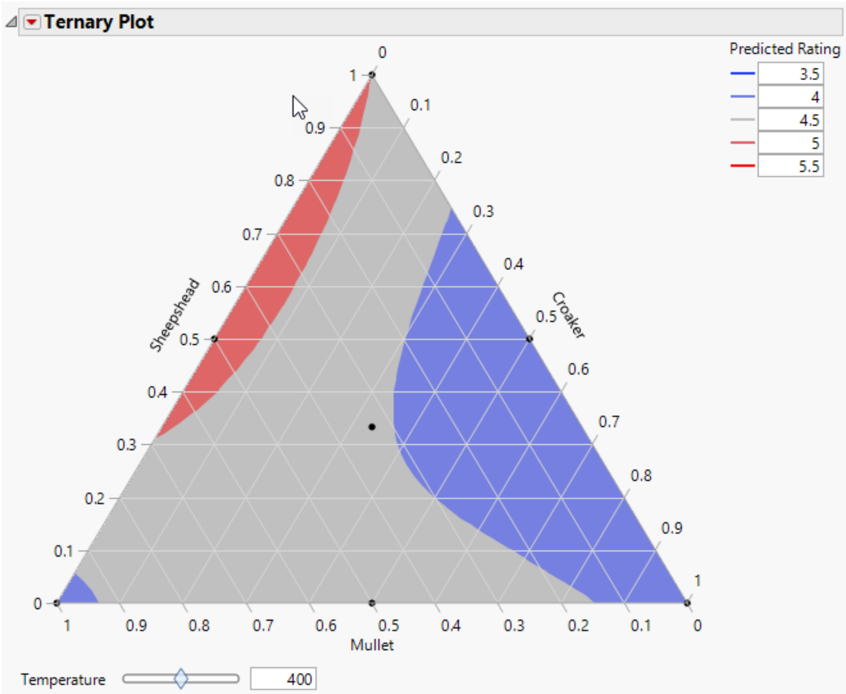
For more information about mixtures, see *Profilers*.

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 *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.

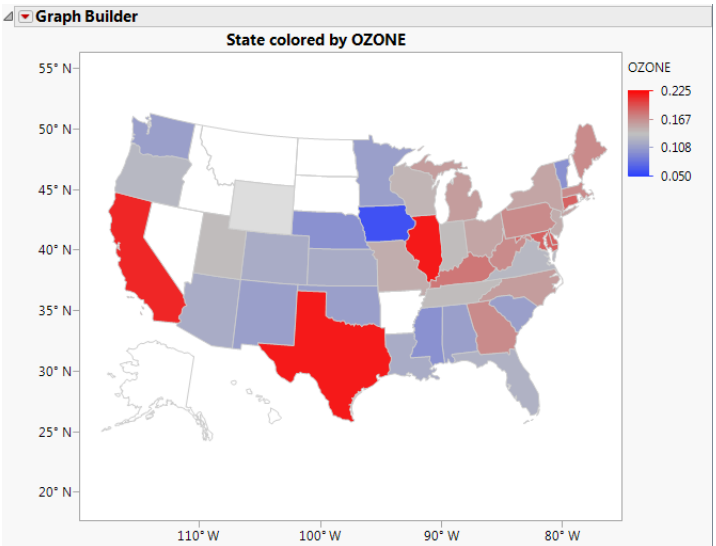
Chapter 12

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:

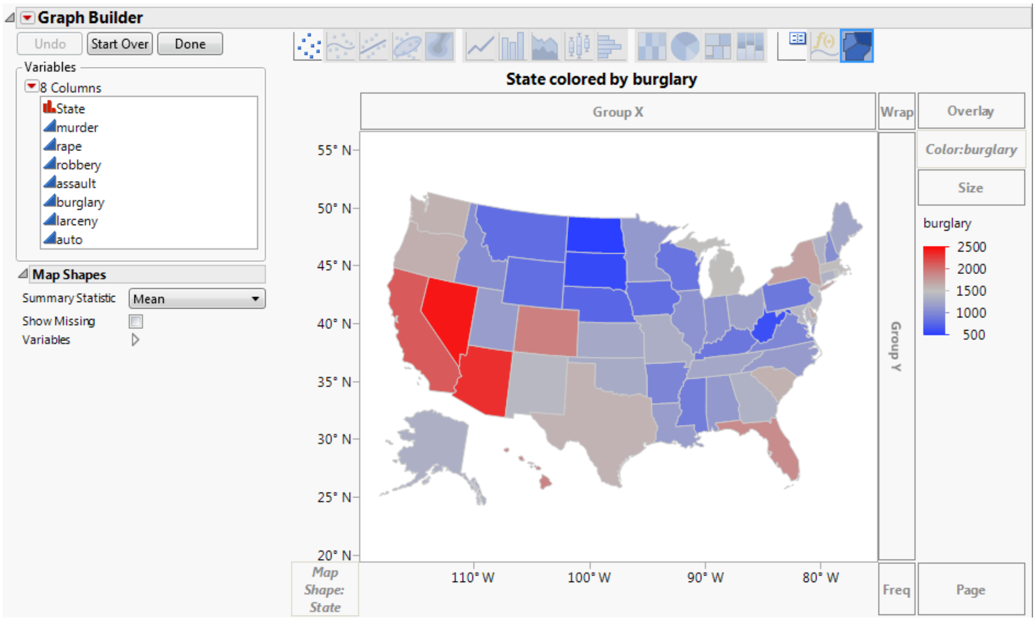
- **Simple Earth** and **Detailed Earth** maps are installed with JMP.
- **NASA server** provides maps using a WMS to show their most up-to-date maps.
- **Street Map Service** provides street maps. The OpenStreetMap and Open Database License links provide further information on the Street Map Service.
- **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



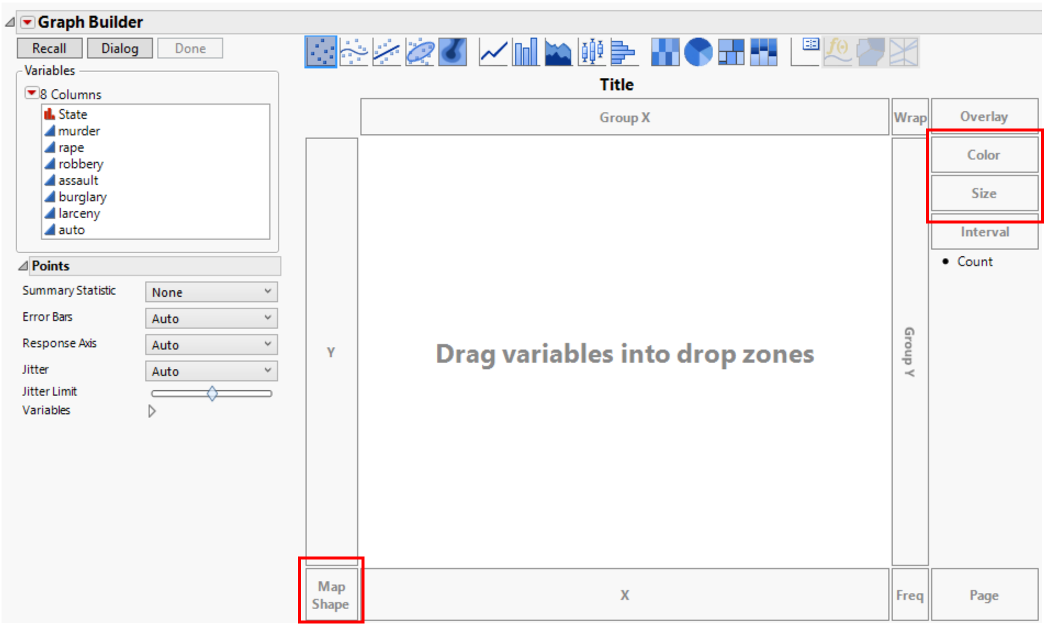
Note the following:

- 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.

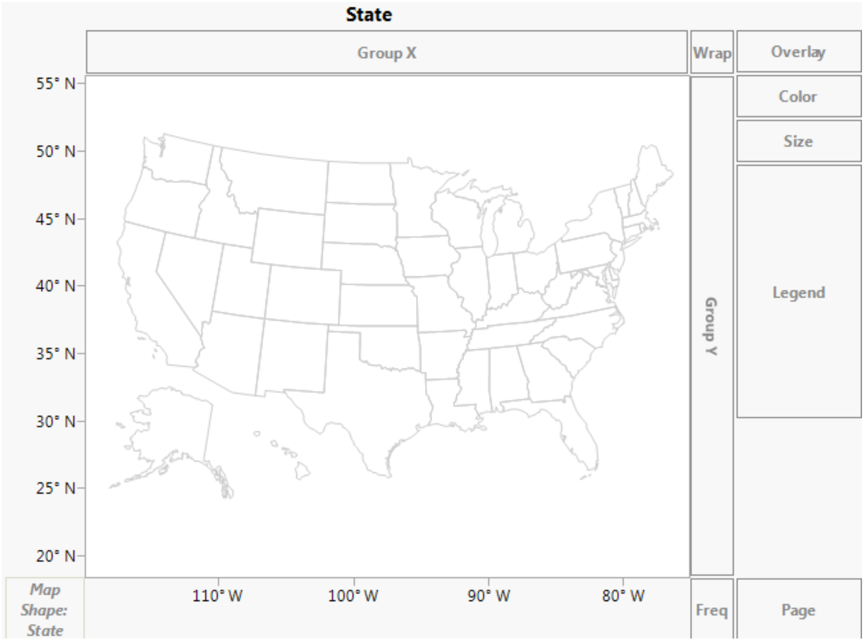
Figure 12.3 The Graph Builder Window



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 JMP 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 part 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

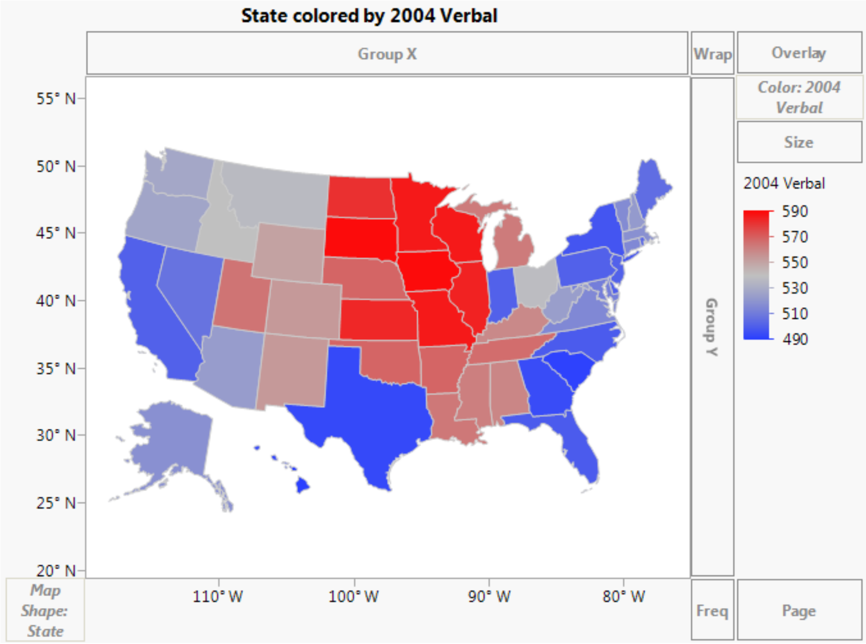
	Shape ID	Part ID	X	Y
1	1	1	85.6052° W	34.9847° N
2	1	1	85.1844° W	32.8613° N
3	1	1	84.9630° W	32.4242° N
4	1	1	85.0071° W	32.3284° N
5	1	1	84.8909° W	32.2615° N
6	1	1	85.0611° W	32.1341° N
7	1	1	85.1407° W	31.8575° N
8	1	1	85.0413° W	31.5410° N
9	1	1	85.1075° W	31.1865° N
10	1	1	85.0024° W	31.0007° N
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.

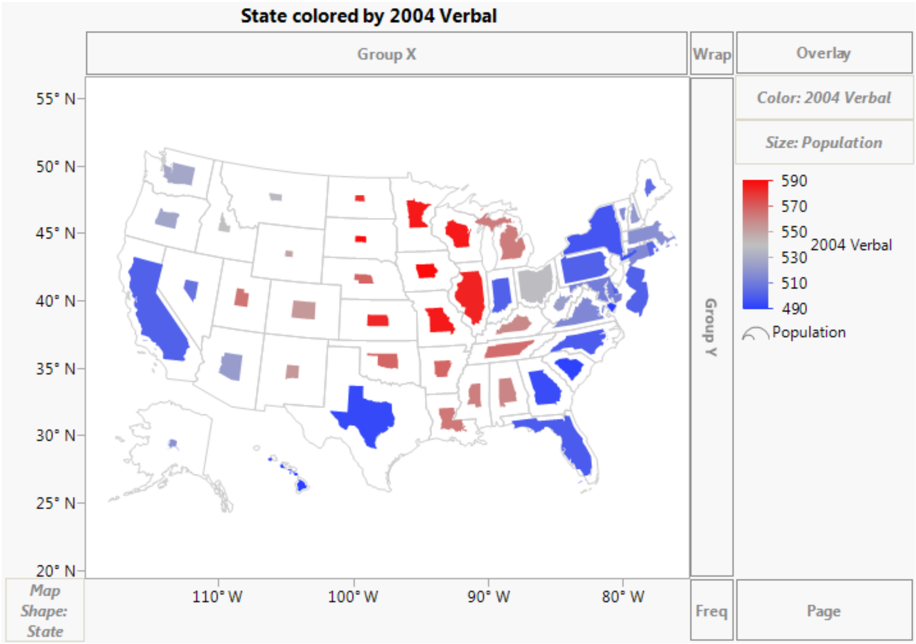
Figure 12.7 Example of SAT.jmp After Dragging 2004 Verbal to Color



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).

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.

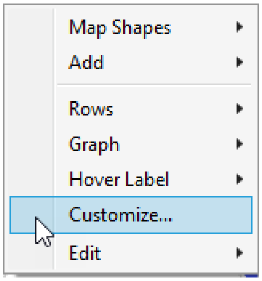
Continuous variables use a color gradient.

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" and *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.

To change the shape file:

1. Right-click the Map Shapes zone and select **Set Shape File**.
2. Navigate to the shape file that contains the name data.
3. Click **Open**.
4. (Optional) To revert the changes, right-click the Map Shapes zone and select **Reset Shape File**.

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".
- SAS/GRAPH software includes a number of map data sets that can be used with JMP. See "SAS/GRAPH Map Data Sets".

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.

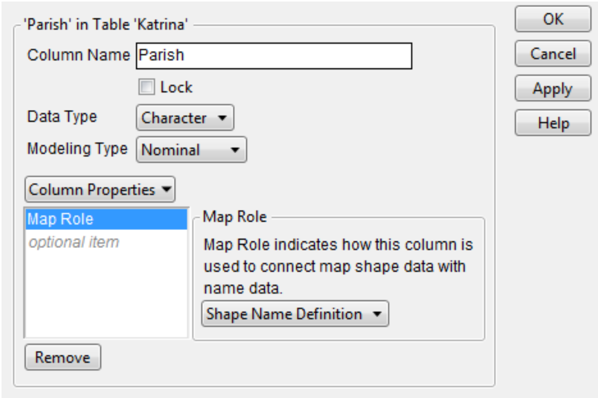
Note the following:

- 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.

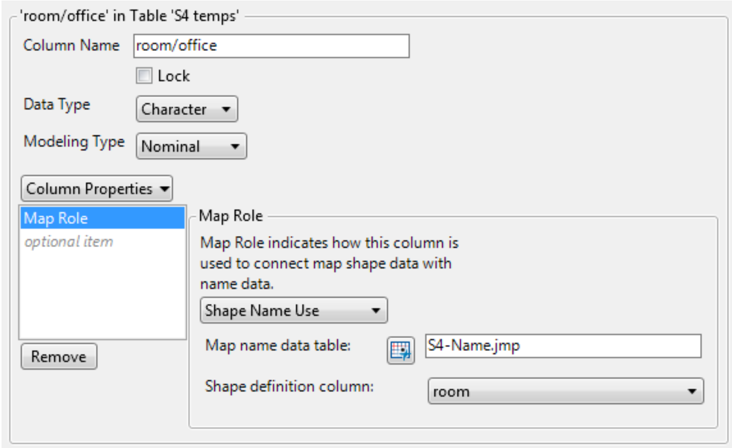
---

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/16/Maps/
- macOS: /Library/Application Support/JMP/16/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.

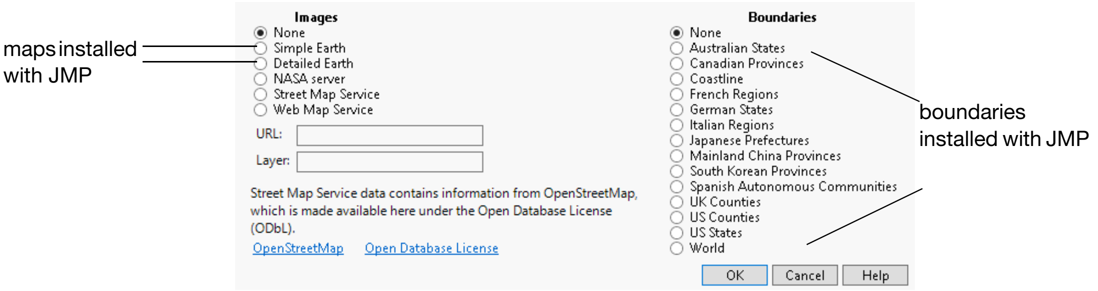
Background Maps

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.

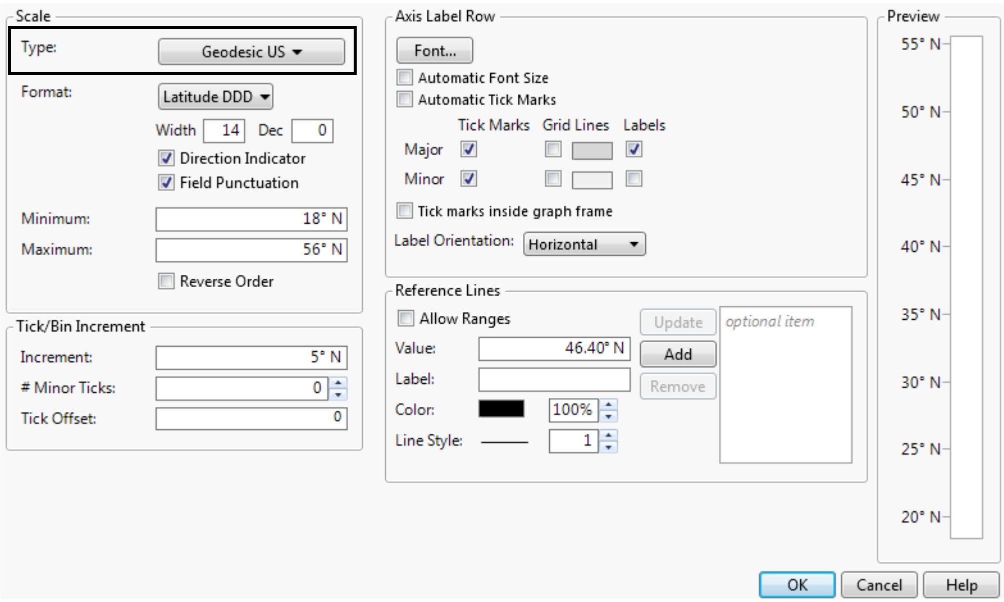
Another feature of Simple Earth and Detailed Earth is the ability to wrap. The Earth is round, and when you cross 180° longitude, the Earth does not end. The longitudinal value continues from -180° and increases. The map wraps continuously in the horizontal direction, much as the Earth does. The background map does not wrap in the vertical direction.

Simple Earth and Detailed Earth both support a geodesic scaling. In Figure 12.13, 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:

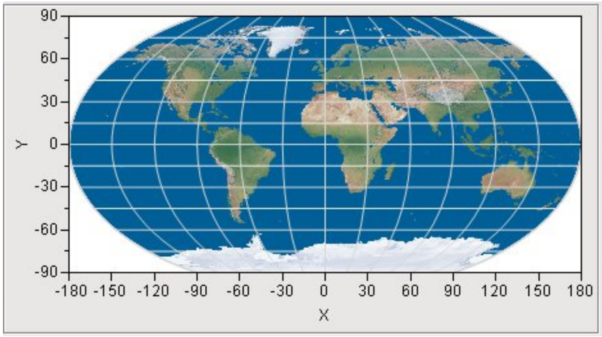
1. Right-click the X or Y axis and then select **Axis Settings.**
2. Change the Scale Type to **Geodesic** or **Geodesic US**.

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.

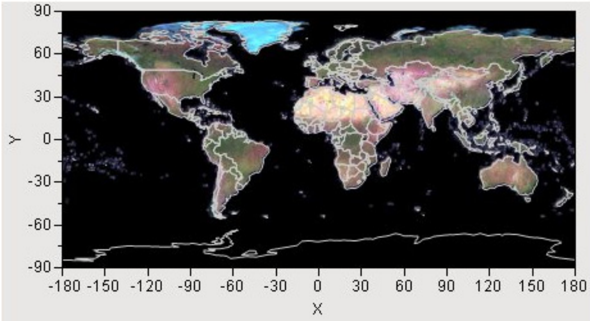
Figure 12.16 Examples of NASA and WMS Maps

NASA image

WMS image from the MetaCarta™ Server

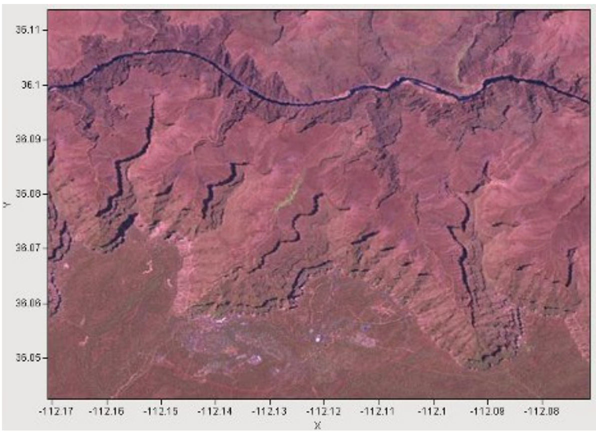
The NASA server provides maps for the entire Earth. The following figure displays the Earth using the NASA server as its source for the background map. The boundary map shows the outlines of the countries.

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 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.

1. In JMP, open the Parishes.shp file from the following default location:
   – On Windows: C:/Program Files/SAS/JMP/16/Samples/Import Data
   – On macOS: /Library/Application Support/JMP/16/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.

1. Open the Parishes.dbf file from the following default location:
   – On Windows: C:/Program Files/SAS/JMP/16/Samples/Import Data
   – On macOS: /Library/Application Support/JMP/16/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. Right-click the PARISH column and select **Column Info**.
3. Select **Column Properties > Map Role**.
4. Select **Shape Name Definition**.
5. Click **OK**.

6. 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
7. 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.

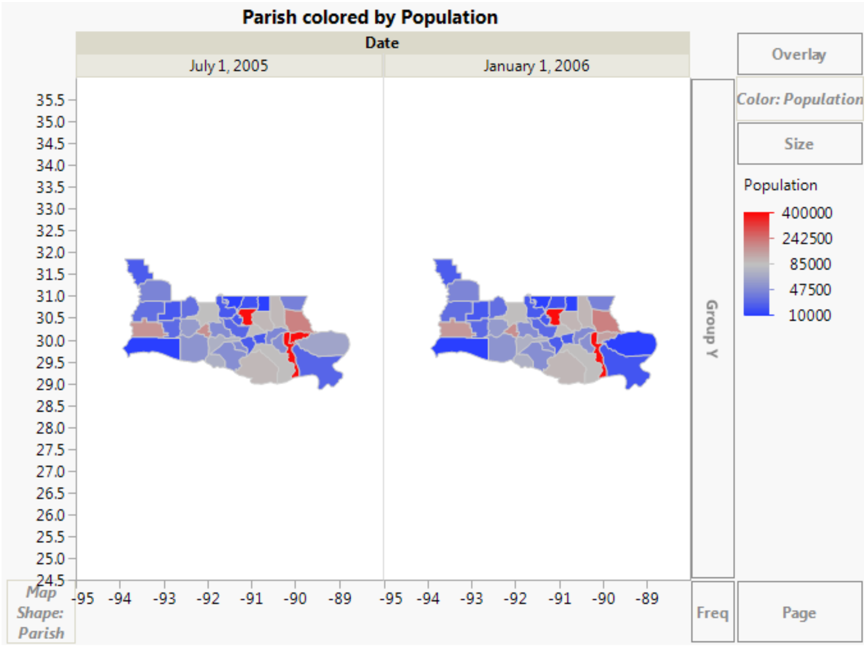
1. Select **Help > Sample Data Library** and open Katrina.jmp.
2. Right-click the Parish column and select **Column Properties > Map Role**.
3. Select **Shape Name Use**.
4. Click the **Map name data table** button and browse to select Parishes-Name.jmp, which you previously created.

   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. Click **OK**.
7. Select **Graph > Graph Builder**.
8. Drag and drop Parish into the **Map Shape** zone.

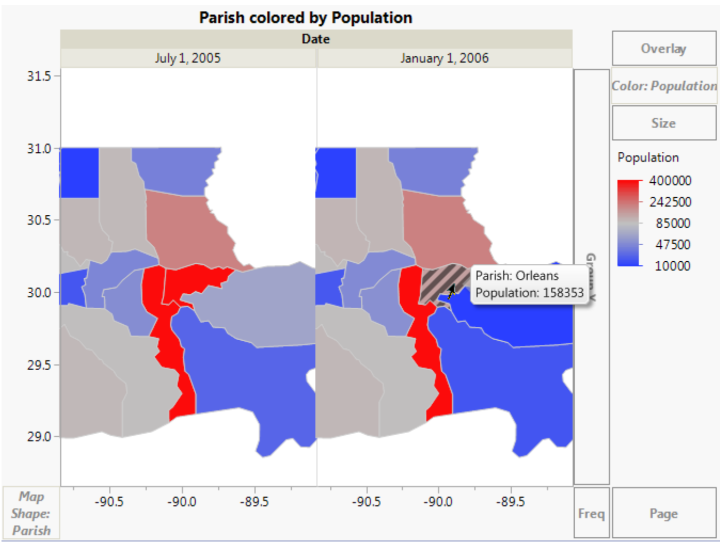
   The map appears automatically, since you defined the Parish column using the custom map files.
9. Drag and drop Population into the **Color** zone.
10. Drag and drop Date into the **Group X** zone.

Figure 12.20 Population of Parishes Before and After Katrina



11. Select the **Magnifier** tool to zoom in on the Orleans parish in both maps (Figure 12.21)

Figure 12.21 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

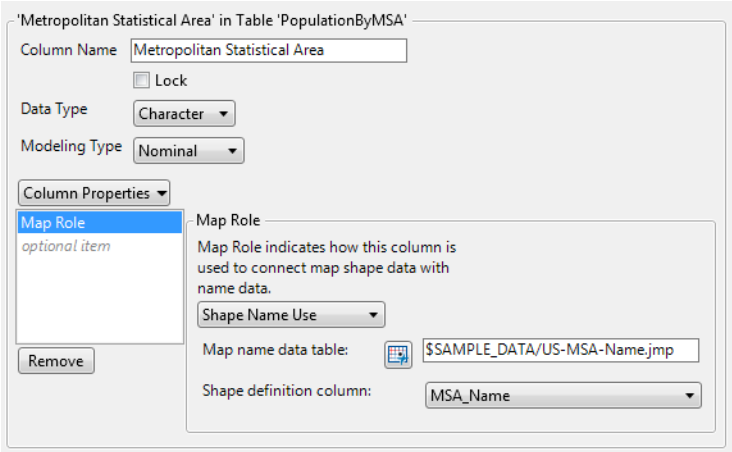
1. Select **Help > Sample Data Library** and open PopulationByMSA.jmp.
2. Right-click the **Metropolitan Statistical Area** column and select **Column Info**.
3. Select **Column Properties > Map Role**.
4. Select **Shape Name Use**.
5. Next to the **Map name data table**, type $SAMPLE_DATA/US-MSA-Name.jmp.

   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 12.22 Map Role Column Property



7. Click **OK**.

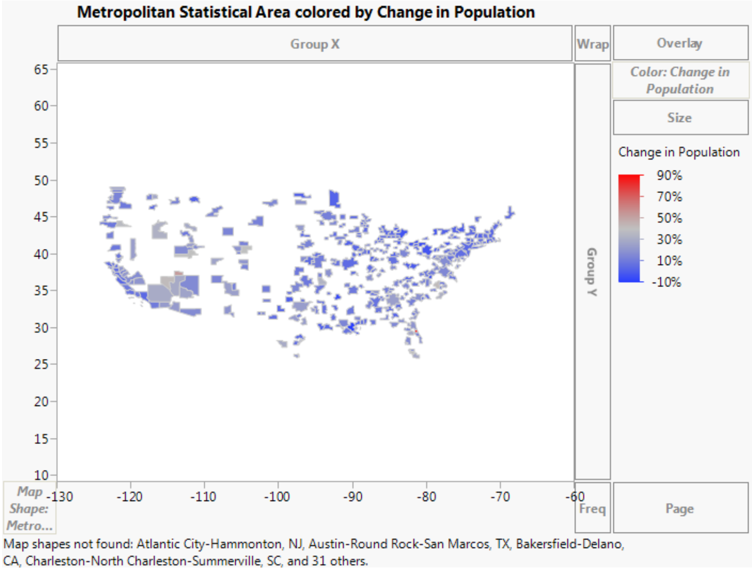
Create the Map in Graph Builder

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.

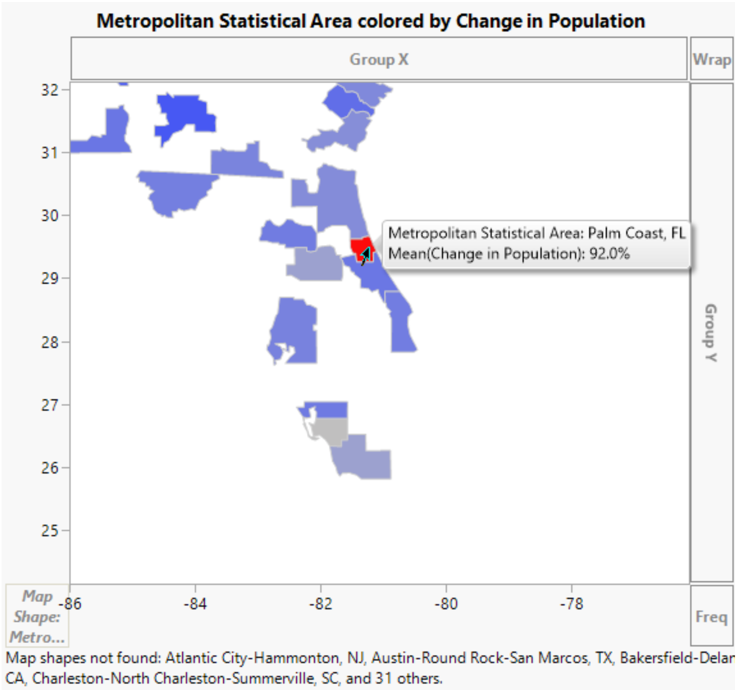
   Since you have defined the **Map Role** column property on this column, the map appears.
3. Drag and drop **Change in Population** to the **Color** zone.

Figure 12.23 Change in Population for Metropolitan Statistical Areas



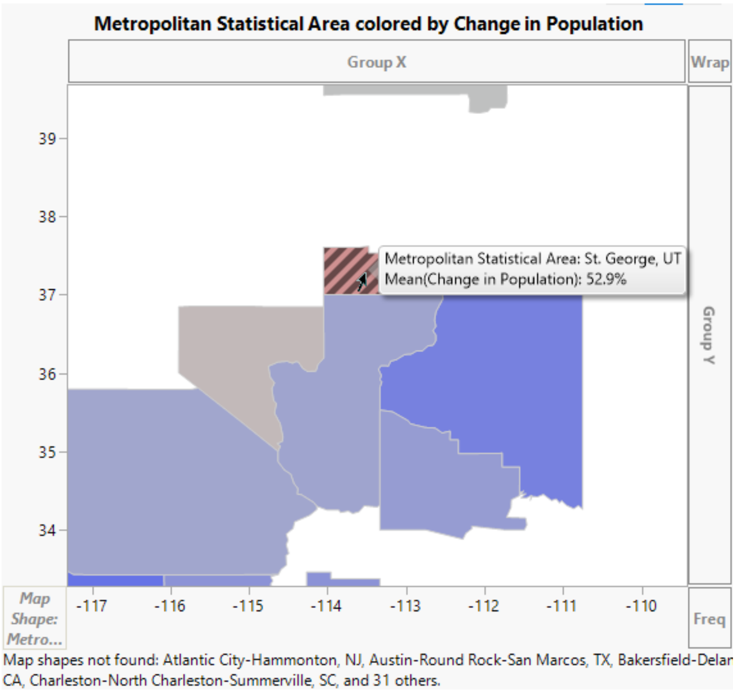
4. Select the **Magnifier** tool to zoom in on the state of Florida.
5. Select the **Arrow** tool and click the red area.

Figure 12.24 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 12.25 Population Change of St. George, Utah



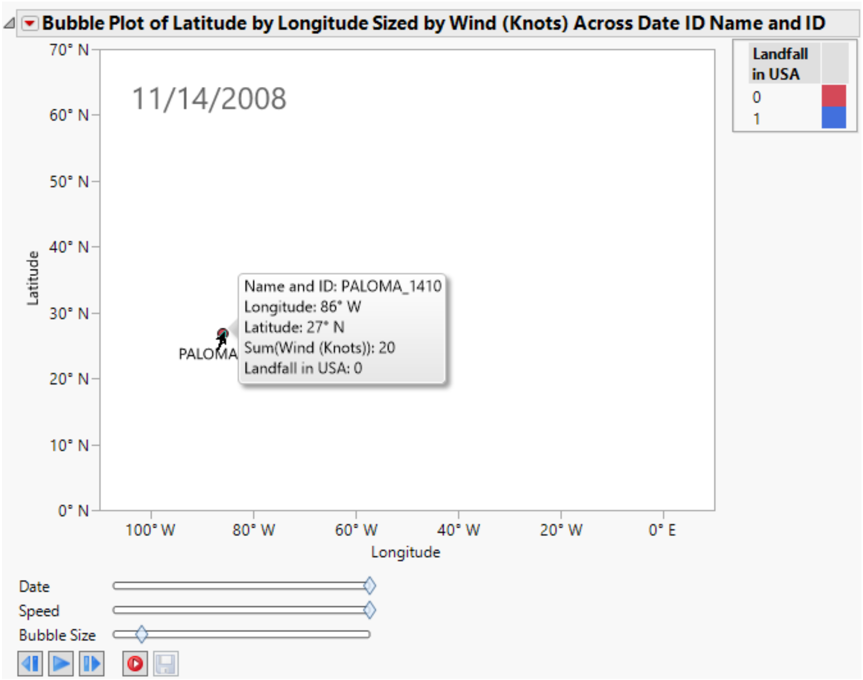
You can see that the areas of Palm Coast, Florida, and St. George, Utah had the most population change between 2000 and 2010. The Palm Coast area saw a population change of 92%, and the St. George area saw a population change of about 53%.

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 12.26.
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.

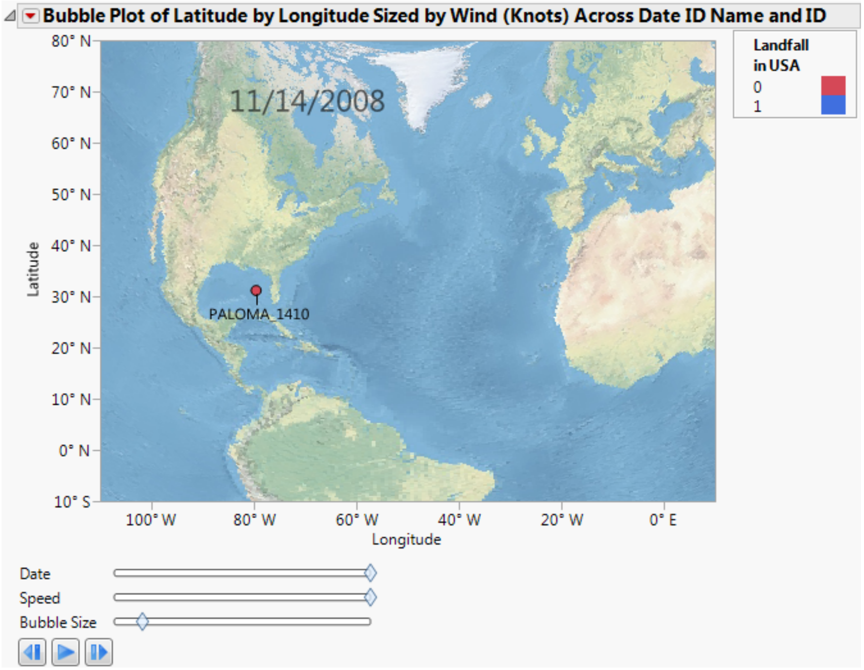
Figure 12.26 Bubble Plot of Hurricanes.jmp



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 12.27 Bubble Plot of Hurricanes.jmp with Background Map

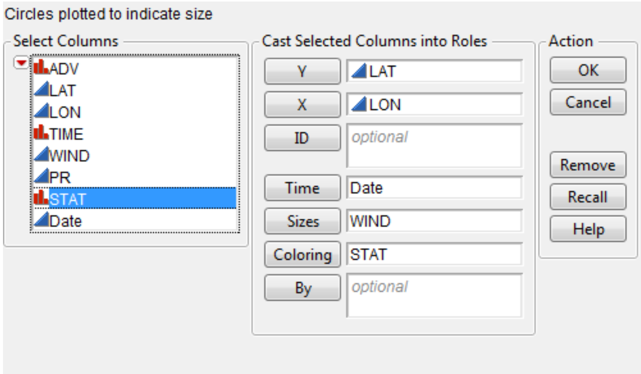


Now the coordinates make geographic sense. Click **Run** to view the animation of the hurricane data moving over the background map. Experiment with different options and view the displays. Adjust the axes or use the zoom tool to change what part of the world you are viewing. The map adjusts as the view 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**.

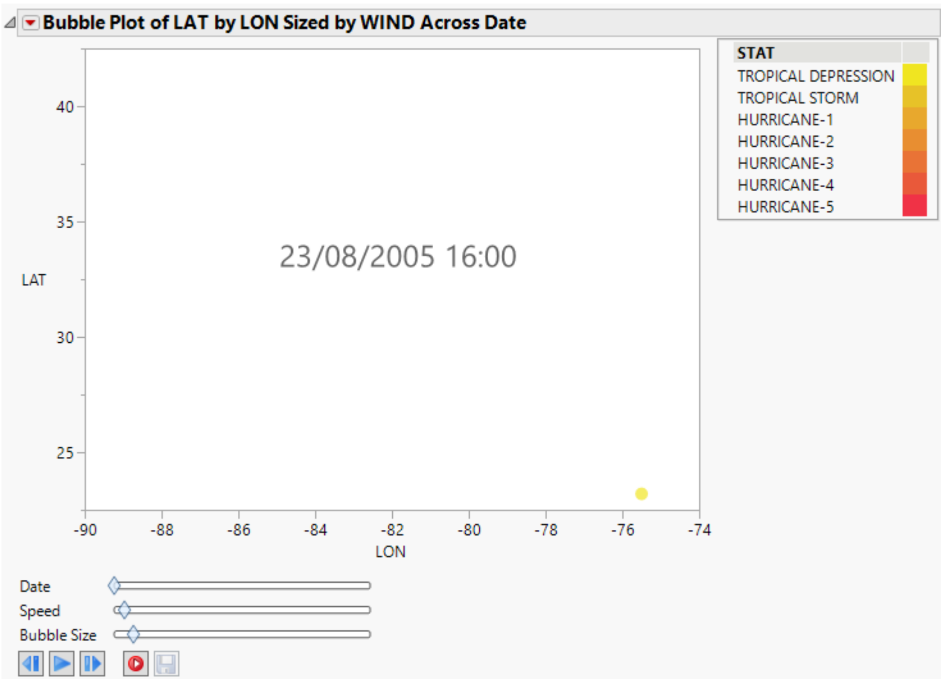
Figure 12.28 Bubble Plot Setup of Katrina Data.jmp



8. Click **OK**.

   The following image appears. The yellow dot shows the location of Tropical Depression Katrina on August 23, 2005.

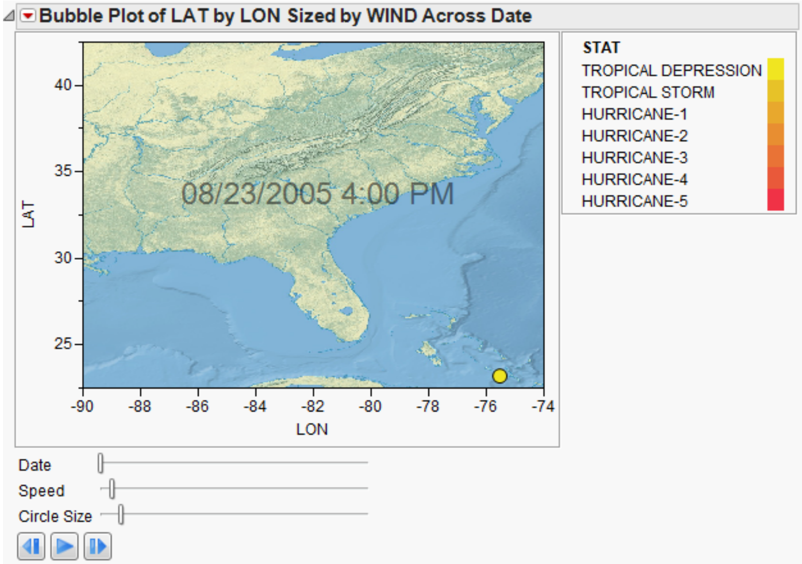
Figure 12.29 Bubble Plot of Katrina Data.jmp



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**.

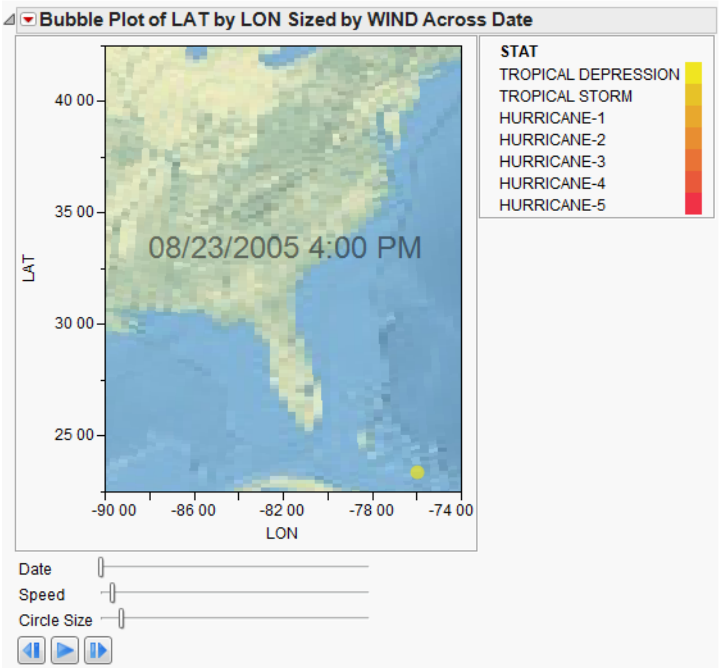
Figure 12.30 Bubble Plot of Katrina Data.jmp with Background Map



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 12.31 Bubble Plot of Katrina Data.jmp with Background Map



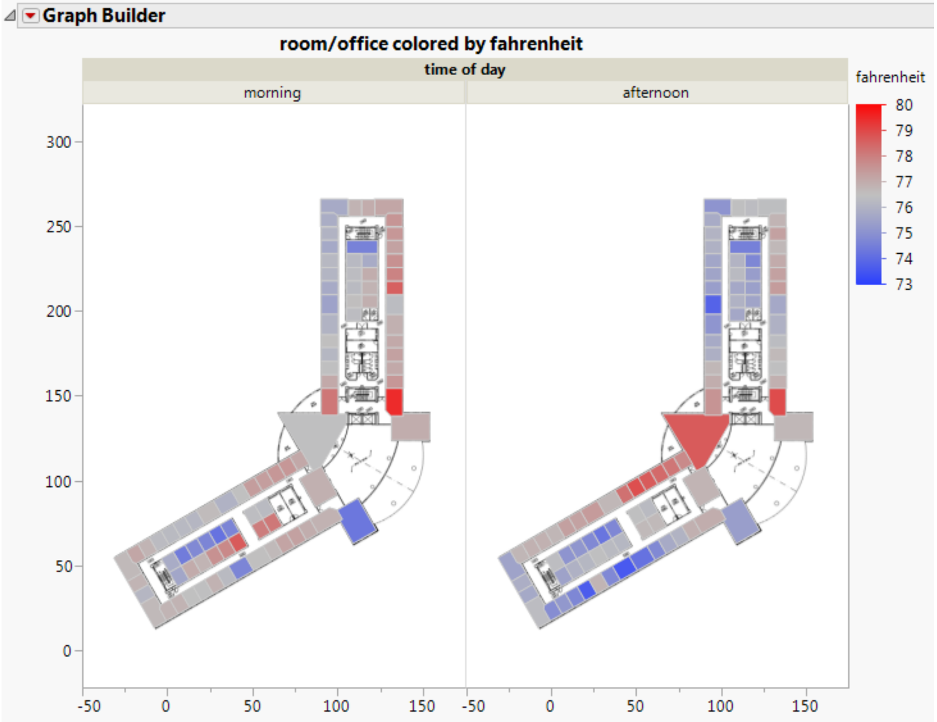
Click **Run** to view the animation of the hurricane data moving over the background map. You can manipulate the speed and the bubble size. Experiment with different options and view the displays. Adjust the axes or use the zoom tool to change what part of the area you are viewing. Add boundaries to 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 12.32 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 12.33 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:

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).

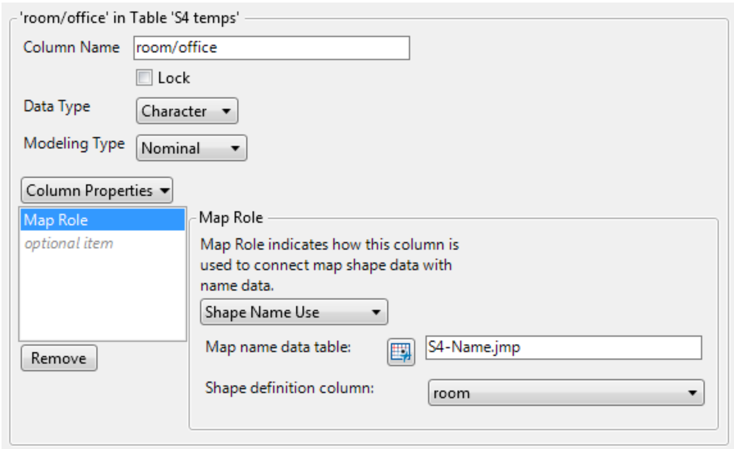
   This tells JMP where the data tables containing the map information reside.

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 12.34 Map Role Column Property



7. Click **OK**.

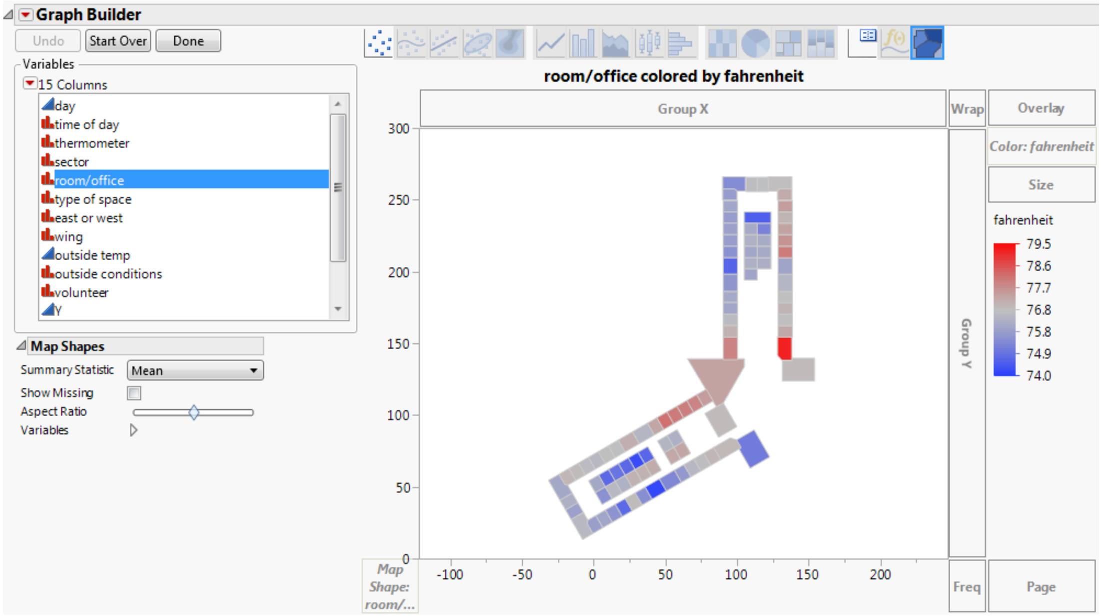
Once the **Map Role** column property has been set up, you can perform your analysis. You want to visually see the differences in office temperatures throughout the floor.

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.

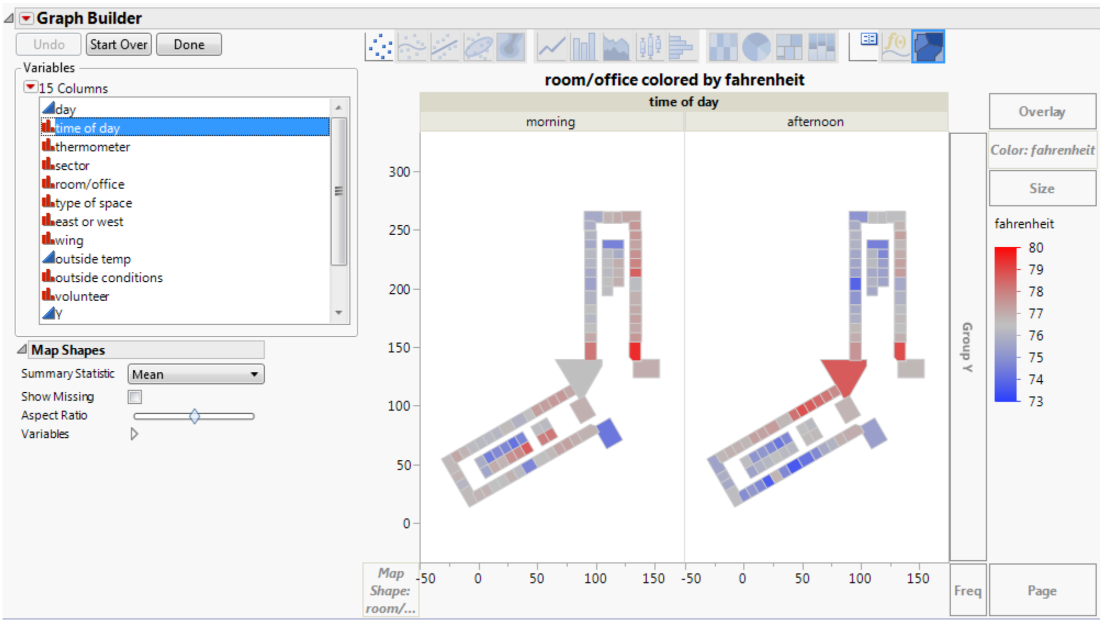
3. Drag and drop Fahrenheit to the **Color** zone.

Figure 12.35 Room/Office Colored by Fahrenheit



4. Drag and drop Time of Day onto Group X zone.

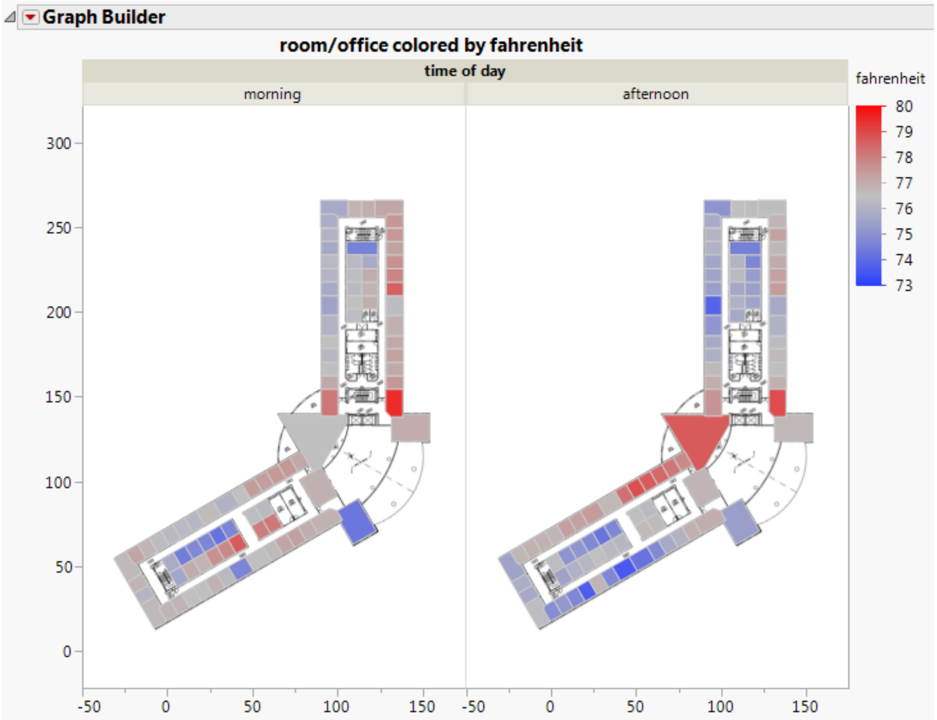
Figure 12.36 Room/Office Colored by Fahrenheit and Grouped by Time of Day



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 12.37.

To view Figure 12.37, select **Help > Sample Data Library** and open S4 Temps.jmp and run the **by Time of Day** script.

Figure 12.37 Room/Office Map with Original Floor Plan

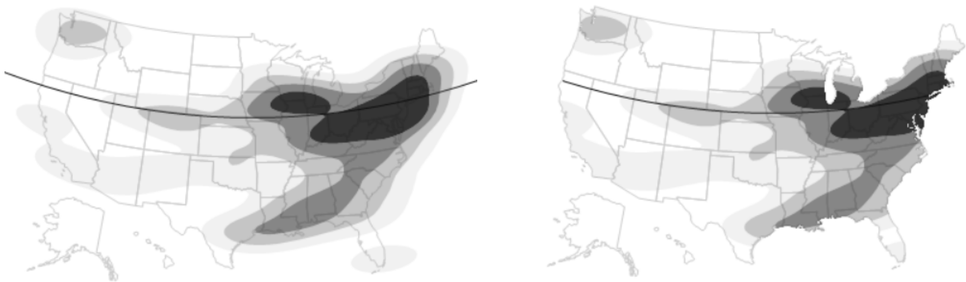


There are several scripts provided with the data table that you can run to view the various analysis and modeling that can be performed and visually displayed.

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 12.38 Unclipped (Left) and Clipped (Right) Reference Lines and Contours

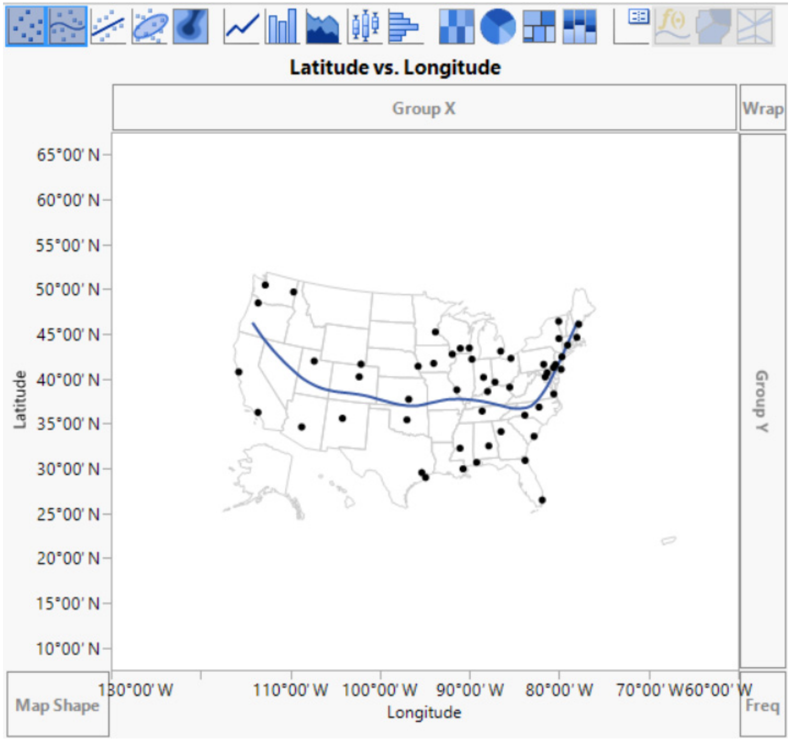


Create the Background Map

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.

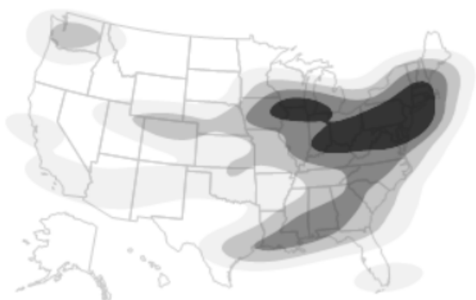
Figure 12.39 Background Map



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 12.40 Unclipped Contours on Background Map

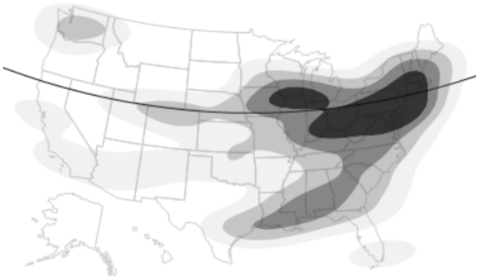


Add a Reference Line

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**.

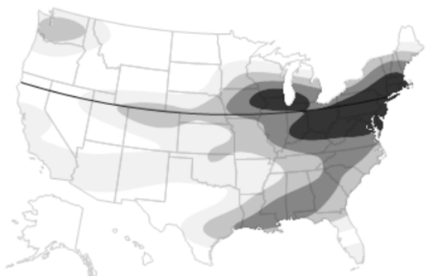
Figure 12.41 Unclipped Reference Line



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 12.42 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” 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 Guide*.

Appendix A
Legacy Platforms

The following legacy platforms are deprecated. A deprecated 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. These legacy platforms might not be available in future releases of JMP.

Table A.1 Legacy Platforms

Platform	Graph Builder References
Treemap	• "Treemap"
Chart	• Bar chart examples in "Graph Builder Examples" • "Bar", "Line", or "Use a Heatmap to Apply Background Colors"
Overlay Plot	• "Example of an Overlaid Histogram and Ridgeline Chart" • "Graph Builder"

Treemaps

Treemaps are useful for observing patterns among groups that have many levels. Use treemaps when your data contains many categories, to visualize many groups.

Tip: To create a treemap, consider using Graph Builder. See "Treemap".

Launch the Treemap platform by selecting **Graph > Legacy > Treemap**.

When you click **OK** in the launch window, a treemap in Graph Builder is created by default. To produce a report in the legacy Treemap platform, you must first deselect the Use Graph Builder platform preference under Files > Preferences > Platforms > Treemap. For documentation on the legacy Treemap platform, see the Legacy Platforms section of the JMP 15.2 Help.

Charts

The Chart platform on the **Graph** menu charts continuous variables versus categorical variables. The continuous variables are summarized for each categorical level.

If you want to make a plot of individual data points (rather than summaries of data points), we recommend using an overlay plot instead.

Tip: To create a chart, consider using Graph Builder. See “Example of an Overlaid Histogram and Ridgeline Chart”.

Launch the Chart platform by selecting **Graph > Legacy > Chart**.

When you click **OK** in the launch window, a chart in Graph Builder is created by default. To produce a report in the legacy Chart platform, you must first deselect the Use Graph Builder platform preference under Files > Preferences > Platforms > Chart. For documentation on the legacy Chart platform, see the Legacy Platforms section of the JMP 15.2 Help.

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.

Tip: To create an overlay plot, consider using the Overlay zone in Graph Builder. See “Example of an Overlaid Histogram and Ridgeline Chart”.

Launch Overlay Plot by selecting **Graph > Legacy > Overlay Plot**.

When you click **OK** in the launch window, an overlay plot in Graph Builder is created by default. To produce a report in the legacy Overlay Plot platform, you must first deselect the Use Graph Builder platform preference under Files > Preferences > Platforms > Overlay Plot. For documentation on the legacy Overlay Plot platform, see the Legacy Platforms section of the JMP 15.2 Help.

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.

Few, Stephen (2008). "Dual-Scaled Axes in Graphs." *Visual Business Intelligence Newsletter*. http://www.perceptualedge.com/articles/visual_business_intelligence/dual-scaled_axes.pdf.

Hyndman, R. J. (1996). "Computing and Graphing High Density Regions."*The American Statistician* 50: 120 - 126.

Hyndman, R. J., Koehler, A. B., Ord, J. K., and Snyder, R. D. (2008). *Forecasting with Exponential Smoothing: The State Space Approach*. Berlin: Springer-Verlag.

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 is 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 is Copyright © 1995-2005, Jean-Loup Gailly and Mark Adler.
- Made with Natural Earth. Free vector and raster map data @ naturalearthdata.com.
- Packages is 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

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 is 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 is 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.

- 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.

- libzip is Copyright © 1999–2019 Dieter Baron and Thomas Klausner.

  This file is part of libzip, a library to manipulate ZIP archives. The authors can be contacted at <libzip@nih.at>.

  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. The names of the authors may not be used to endorse or promote products derived from this software without specific prior written permission.

  THIS SOFTWARE IS PROVIDED BY THE AUTHORS ``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 AUTHORS 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.

- OpenNLP 1.5.3, the pre-trained model (version 1.5 of en-parser-chunking.bin), and dmlc/xgboost Version .90 are licensed under the Apache License 2.0 are Copyright © January 2004 by Apache.org.

  You may reproduce and distribute copies of the Work or Derivative Works thereof in any medium, with or without modifications, and in Source or Object form, provided that You meet the following conditions:

- You must give any other recipients of the Work or Derivative Works a copy of this License; and
- You must cause any modified files to carry prominent notices stating that You changed the files; and
- You must retain, in the Source form of any Derivative Works that You distribute, all copyright, patent, trademark, and attribution notices from the Source form of the Work, excluding those notices that do not pertain to any part of the Derivative Works; and
- If the Work includes a "NOTICE" text file as part of its distribution, then any Derivative Works that You distribute must include a readable copy of the attribution notices contained within such NOTICE file, excluding those notices that do not pertain to any part of the Derivative Works, in at least one of the following places: within a NOTICE text file distributed as part of the Derivative Works; within the Source form or documentation, if provided along with the Derivative Works; or, within a display generated by the Derivative Works, if and wherever such third-party notices normally appear. The contents of the NOTICE file are for informational purposes only and do not modify the License. You may add Your own attribution notices within Derivative Works that You distribute, alongside or as an addendum to the NOTICE text from the Work, provided that such additional attribution notices cannot be construed as modifying the License.
- You may add Your own copyright statement to Your modifications and may provide additional or different license terms and conditions for use, reproduction, or distribution of Your modifications, or for any such Derivative Works as a whole, provided Your use, reproduction, and distribution of the Work otherwise complies with the conditions stated in this License.

• LLVM is Copyright © 2003–2019 by the University of Illinois at Urbana-Champaign.

  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.

• clang is Copyright © 2007–2019 by the University of Illinois at Urbana-Champaign.

  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", WITHOUT WARRANTIES OR CONDITIONS OF