Statistical Discovery.™ From SAS.

Version 13

Discovering JMP

"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

The correct bibliographic citation for this manual is as follows: SAS Institute Inc. 2016. *Discovering JMP 13®*. Cary, NC: SAS Institute Inc.

Discovering JMP 13®

ISBN 978-1-62960-473-2 (Hardcopy)
ISBN 978-1-62960-570-8 (EPUB)
ISBN 978-1-62960-571-5 (MOBI)

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

September 2016

Get the Most from JMP®

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

Visit JMP.com to find the following:

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

http://www.jmp.com/getstarted/

Contents

Discovering JMP

3 Work with Your Data

4 Visualize Your Data

5 Analyze Your Data

Index

About This Book

Discovering JMP provides a general introduction to the JMP software. This guide assumes that you have no knowledge of JMP. Whether you are an analyst, researcher, student, professor, or statistician, this guide gives you a general overview of JMP's user interface and features.

This guide introduces you to the following information:

- Starting JMP
- The structure of a JMP window
- Preparing and manipulating data
- Using interactive graphs to learn from your data
- Performing simple analyses to augment graphs
- Customizing JMP and special features

This guide contains six chapters. Each chapter contains examples that reinforce the concepts presented in the chapter. All of the statistical concepts are at an introductory level. The sample data used in this book are included with the software. Here is a description of each chapter:

- Chapter 2, "Introducing JMP" provides an overview of the JMP application. You learn how content is organized and how to navigate the software.
- Chapter 3, "Work with Your Data" describes how to import data from a variety of sources, and prepare it for analysis. There is also an overview of data manipulation tools.
- Chapter 4, "Visualize Your Data" describes graphs and charts that you can use to visualize and understand your data. The examples range from simple analyses involving a single variable, to multiple-variable graphs that enable you to see relationships between many variables.
- Chapter 5, "Analyze Your Data" describes many commonly used analysis techniques. These techniques range from simple techniques that do not require the use of statistical methods, to advanced techniques, where knowledge of statistics is useful.
- Chapter 6, "The Big Picture" shows you how to analyze distributions, patterns, and similar values in several platforms.
- Chapter 7, "Save and Share Your Work" describes sharing your work with non JMP users in PowerPoint presentations, interactive HTML, and Adobe Flash. Saving analyses as scripts and saving work in journals and projects for JMP users are also covered.

- Chapter 8, “Special Features” describes how to automatically update graphs and analyses as data changes, how to use preferences to customize your reports, and how JMP interacts with SAS.

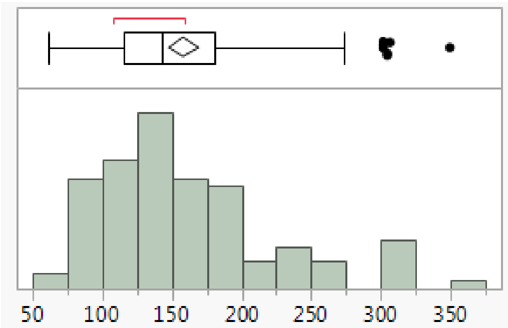
After reviewing this guide, you will be comfortable navigating and working with your data in JMP.

JMP is available for both Windows and Macintosh operating systems. However, the material in this guide is based on a Windows operating system.

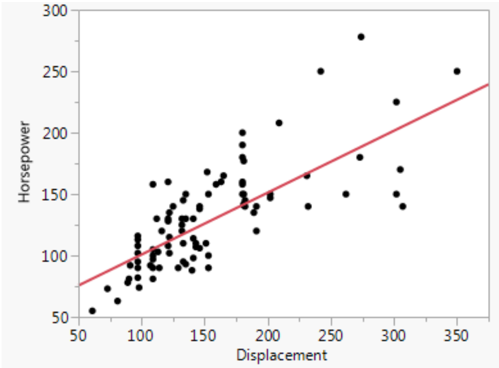
Gallery of JMP Graphs

Various Graphs and Their Platforms

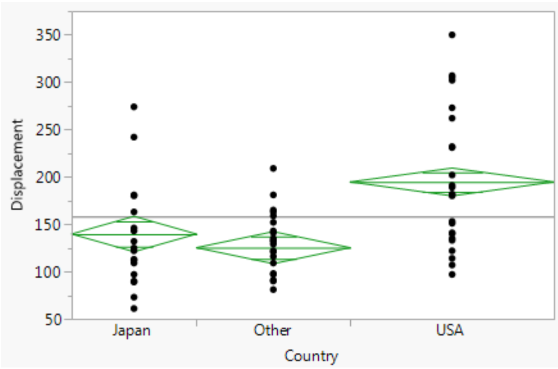
Here are pictures of many of the graphs that you can create with JMP. Each picture is labeled with the platform used to create it. For more information about the platforms and these and other graphs, see the documentation on the **Help > Books** menu.



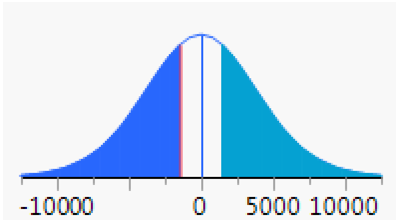
Histogram
Analyze > Distribution



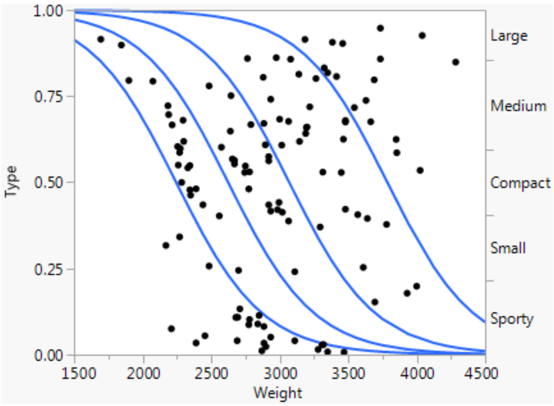
Bivariate
Analyze > Fit Y by X



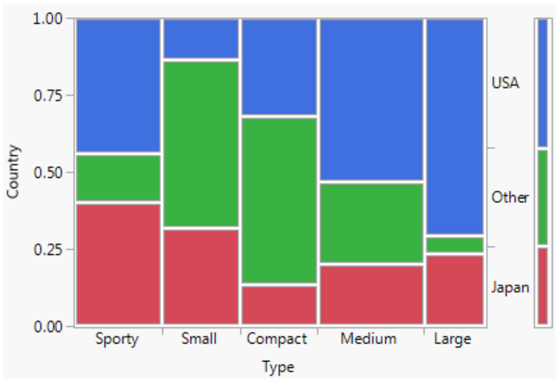
Oneway
Analyze > Fit Y by X



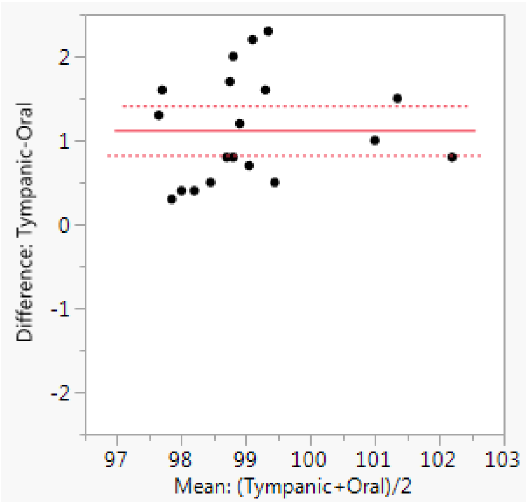
Oneway t Test
Analyze > Fit Y by X



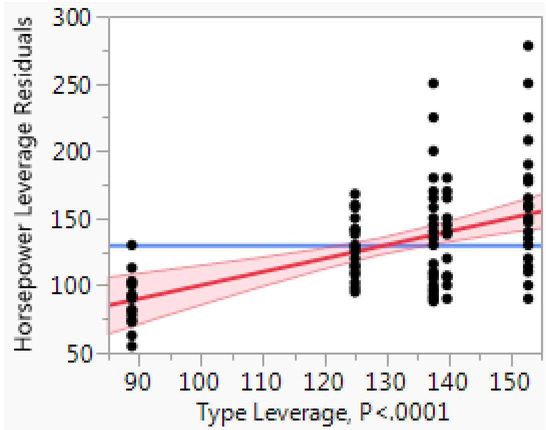
Logistic
Analyze > Fit Y by X



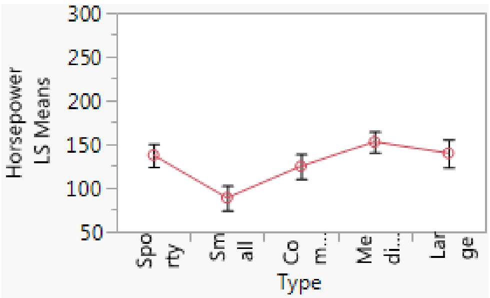
Mosaic Plot
Analyze > Fit Y by X



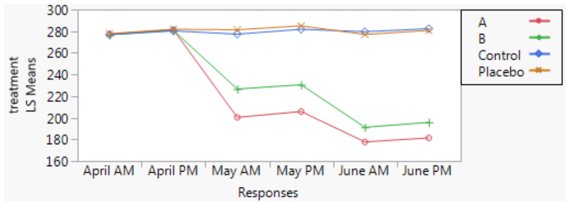
Matched Pairs
Analyze > Specialized Modeling > Matched Pairs



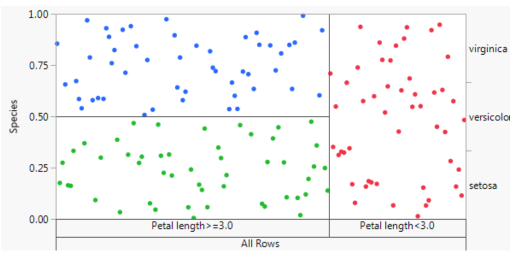
Leverage Plot
Analyze > Fit Model



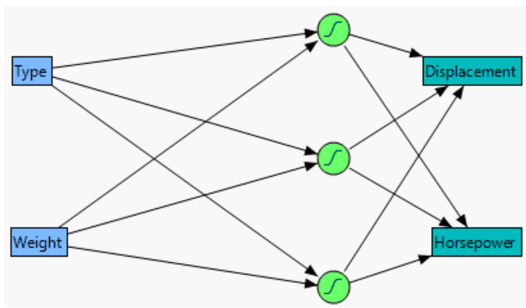
LS Means Plot
Analyze > Fit Model



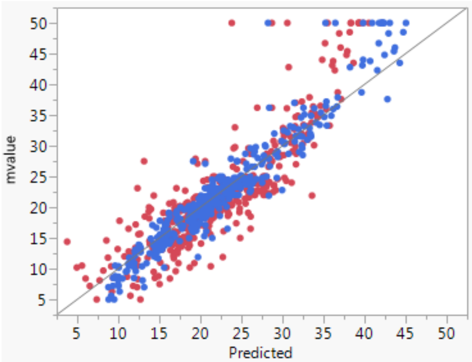
MANOVA
Analyze > Fit Model



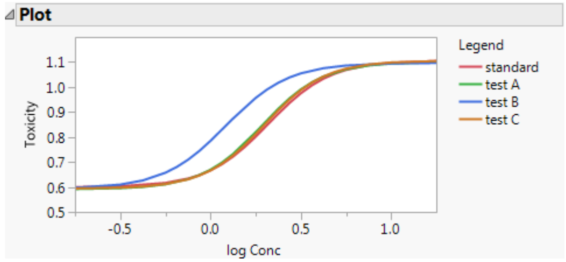
Partition
Analyze > Predictive Modeling > Partition



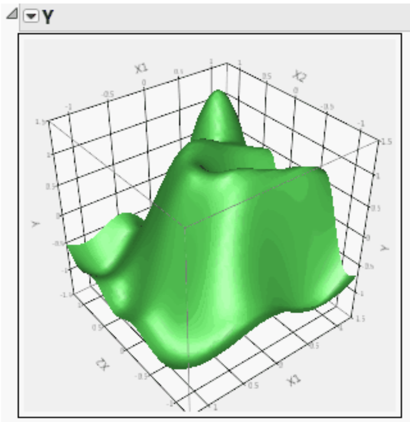
Neural Diagram
Analyze > Predictive Modeling > Neural



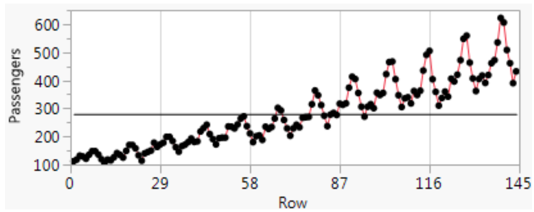
Actual by Predicted
Analyze > Predictive Modeling > Model Comparison



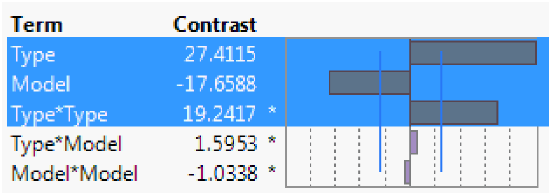
Nonlinear Fit
Analyze > Specialized Modeling > Nonlinear



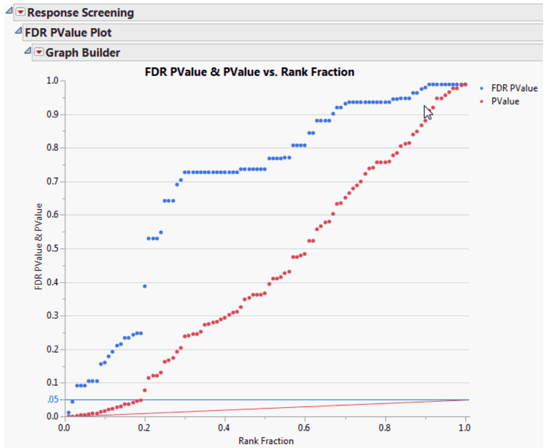
Surface Profiler
Analyze > Specialized Modeling > Gaussian Process



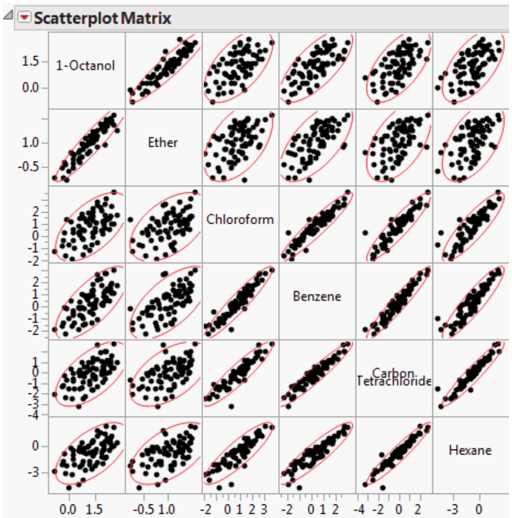
Time Series
Analyze > Specialized Modeling > Time Series



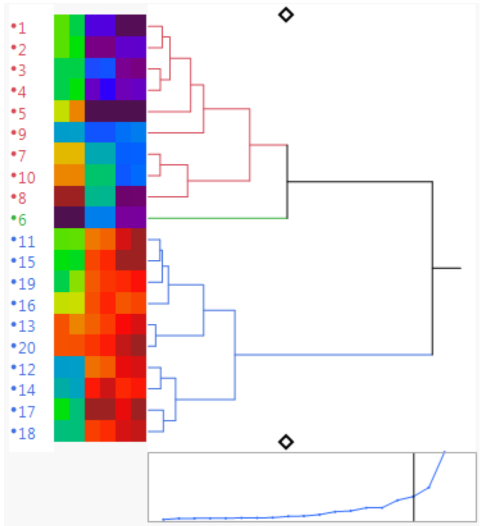
Screening
Analyze > Specialized Modeling > Specialized DOE Models > Fit Two Level Screening



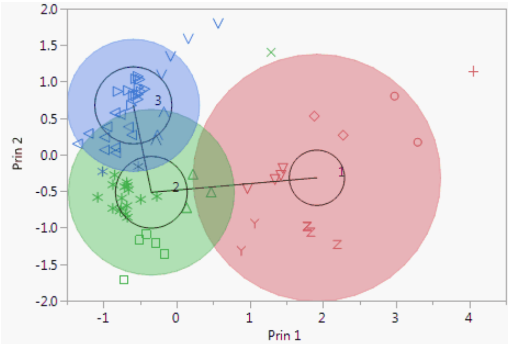
FDR pValue Plot
Analyze > Screening > Response Screening



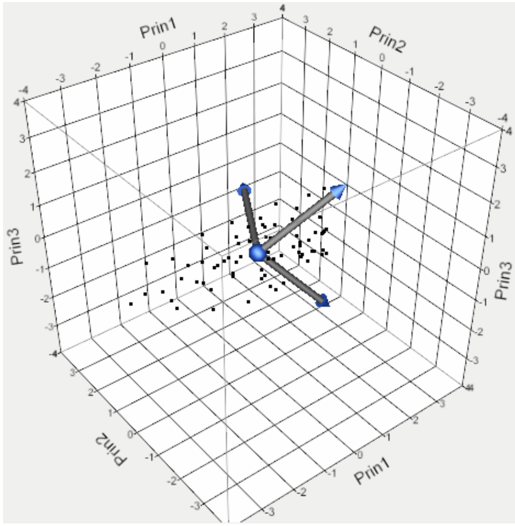
Scatterplot Matrix
Analyze > Multivariate Methods > Multivariate



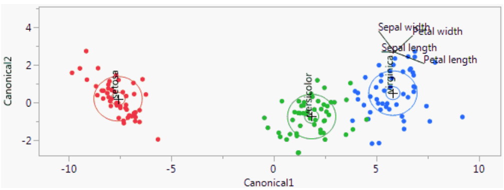
Dendrogram
Analyze > Clustering > Hierarchical Cluster



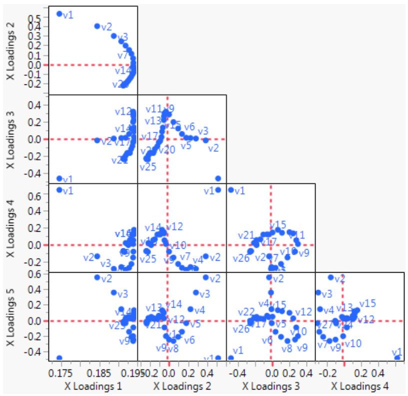
Self Organizing Map
Analyze > Clustering > K Means Cluster



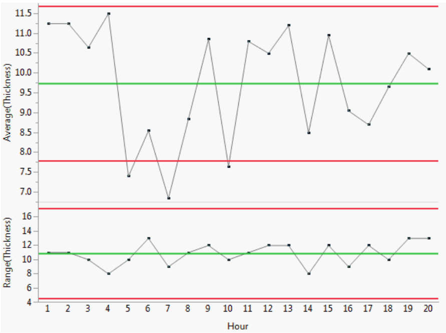
Principal Components
Analyze > Multivariate Methods > Principal Components



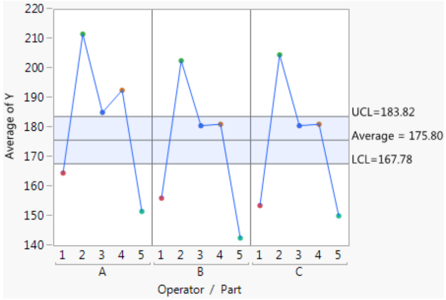
Canonical Plot
Analyze > Multivariate Methods > Discriminant



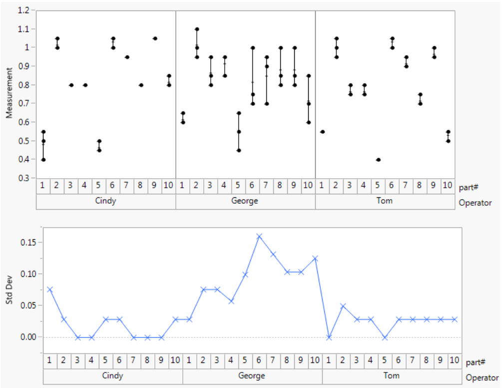
Loadings Plot
Analyze > Multivariate Methods > Partial Least Squares



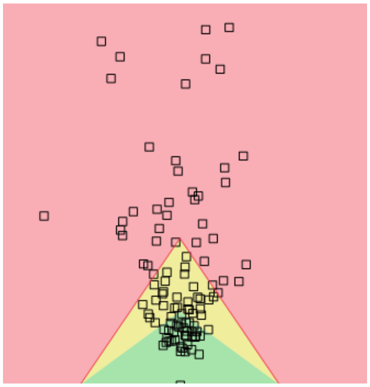
XBar and R Charts
Analyze > Quality and Process > Control Chart Builder



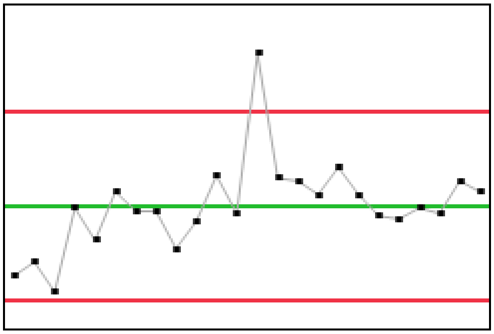
Average Chart
Analyze > Quality and Process > Measurement Systems Analysis



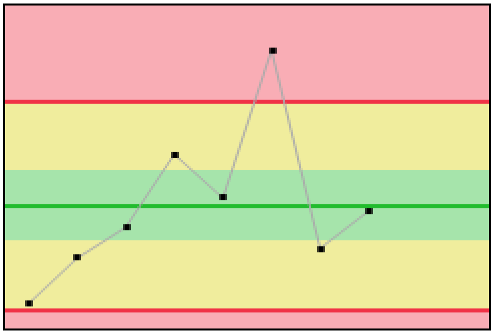
Variability Chart
Analyze > Quality and Process > Variability/Attribute Chart



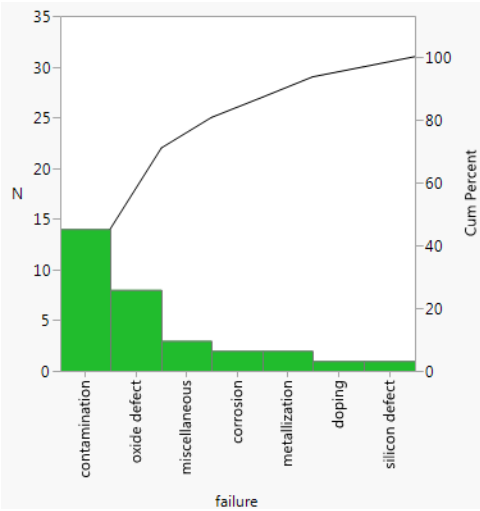
Goal Plot
Analyze > Quality and Process > Process Capability



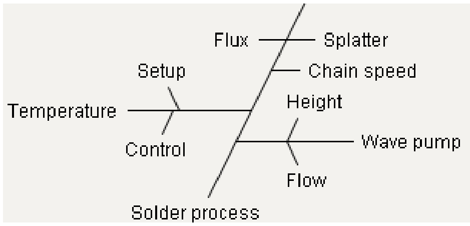
Individual Measurement Chart
Moving Range Chart
Analyze > Quality and Process > Control Chart > IR



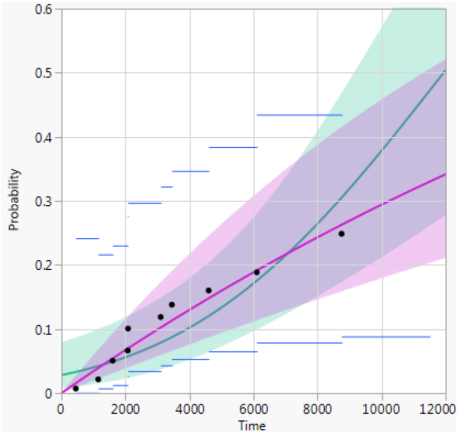
XBar Chart
Analyze > Quality and Process > Control Chart > XBar



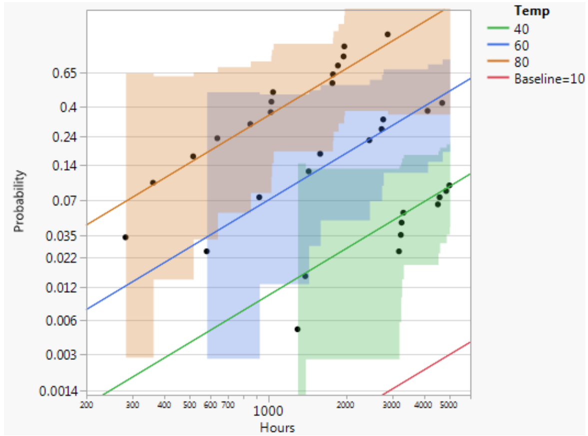
Pareto Plot
Analyze > Quality and Process > Pareto Plot



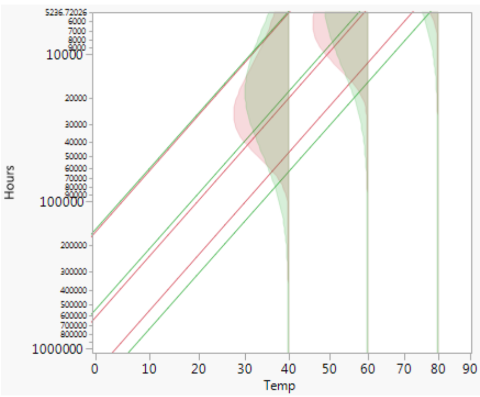
Ishikawa Chart
Fishbone Chart
Analyze > Quality and Process> Diagram



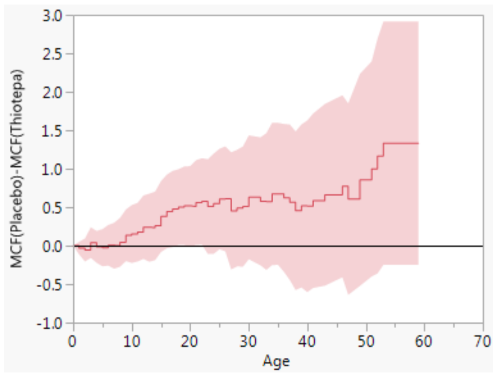
Compare Distributions
Analyze > Reliability and Survival > Life Distribution



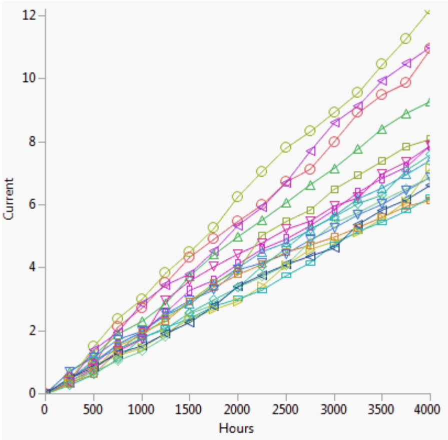
Nonparametric Overlay
Analyze > Reliability and Survival > Fit Life by X



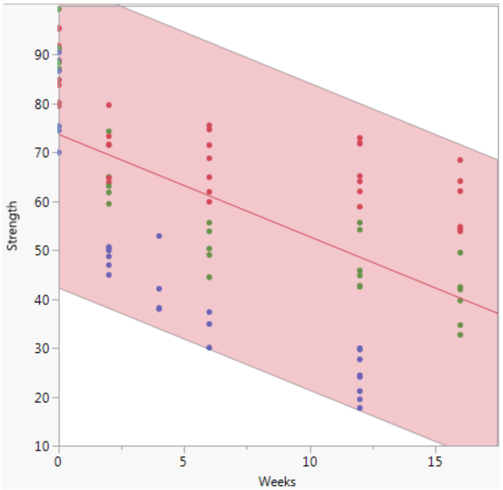
Scatterplot
Analyze > Reliability and Survival > Fit Life by X



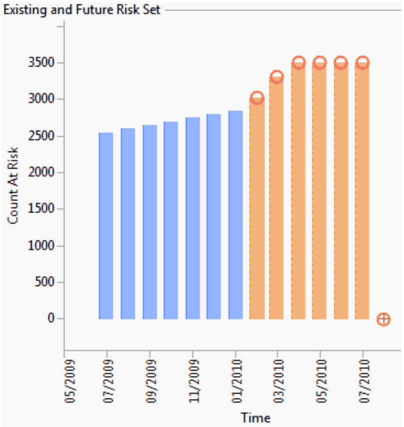
MCF Plot
Analyze > Reliability and Survival > Recurrence Analysis



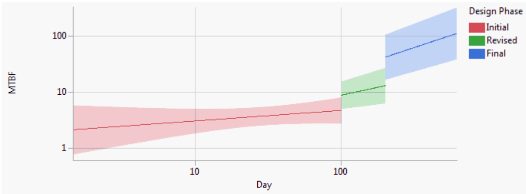
Overlay
Analyze > Reliability and Survival > Degradation



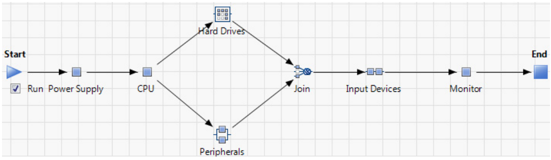
Prediction Interval
Analyze > Reliability and Survival > Destructive Degradation



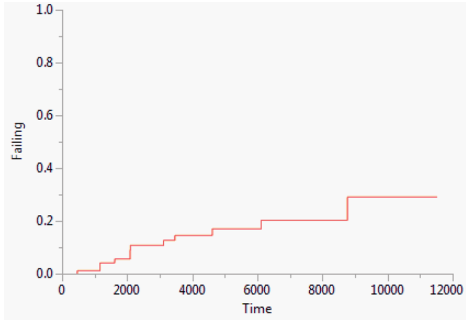
Forecast
Analyze > Reliability and Survival > Reliability Forecast



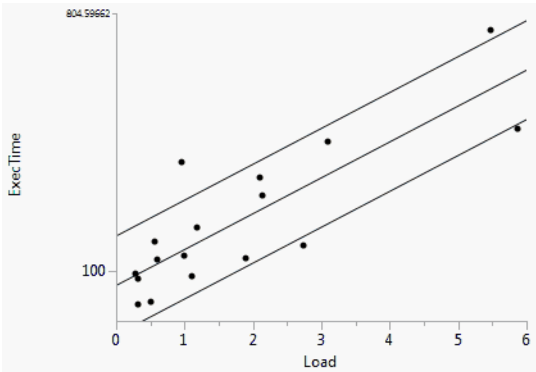
Piecewise Weibull NHPP
Analyze > Reliability and Survival > Reliability Growth



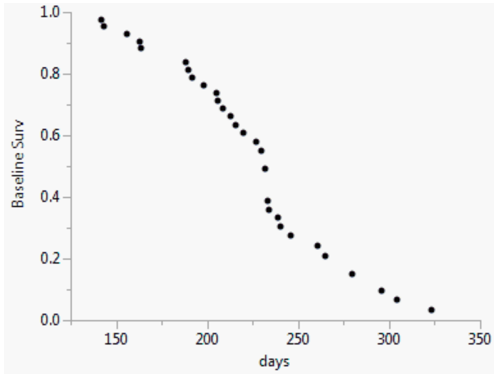
Reliability Block Diagram
Analyze > Reliability and Survival > Reliability Block Diagram



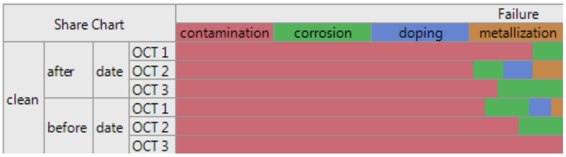
Failure Plot
Analyze > Reliability and Survival > Survival



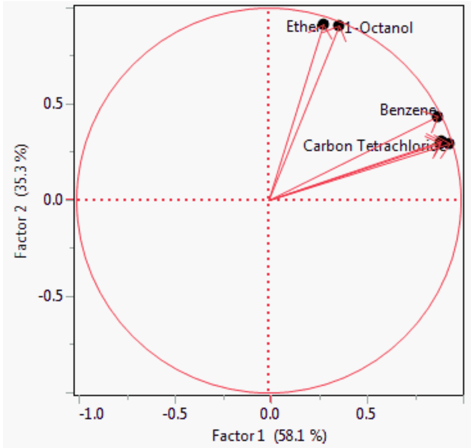
Survival Quantiles
Analyze > Reliability and Survival > Fit Parametric Survival



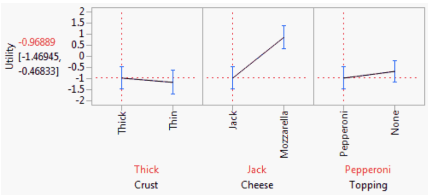
Baseline Survival
Analyze > Reliability and Survival > Fit Proportional Hazards



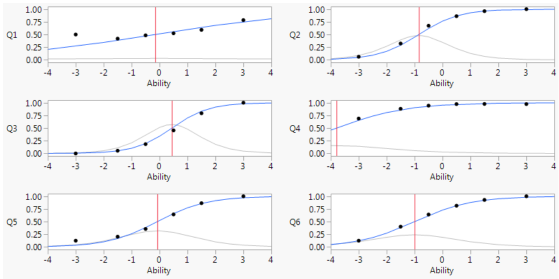
Mixture Profiler
Analyze > Consumer Research > Categorical



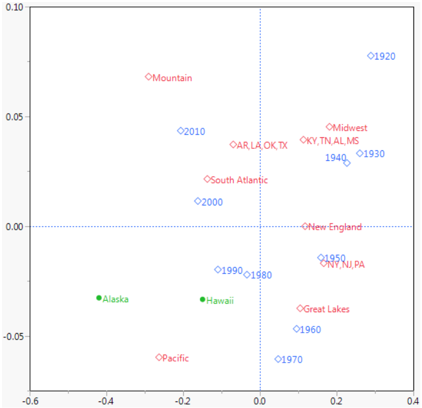
Factor Loading Plot
Analyze > Consumer Research > Factor Analysis



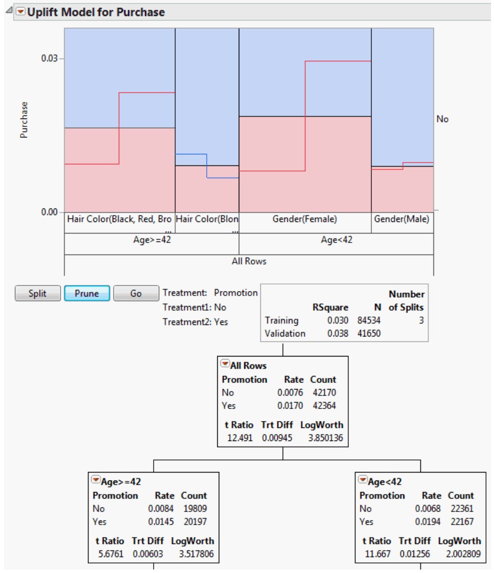
Prediction Profile
Analyze > Consumer Research > Choice



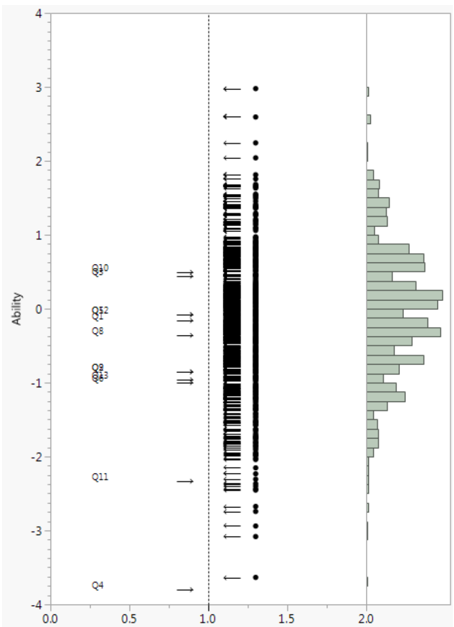
Characteristic Curves
Analyze > Consumer Research > Item Analysis



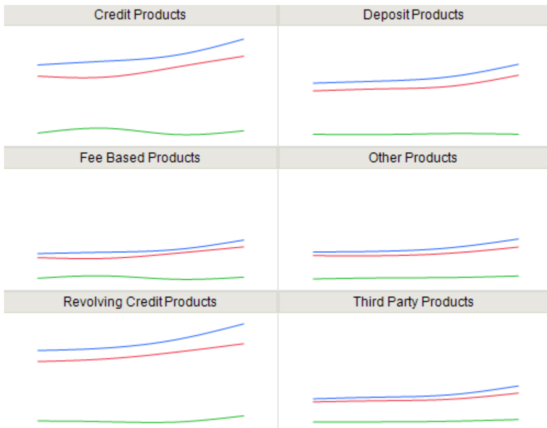
Multiple Correspondence Analysis
Analyze > Consumer Research > Multiple Correspondence Analysis



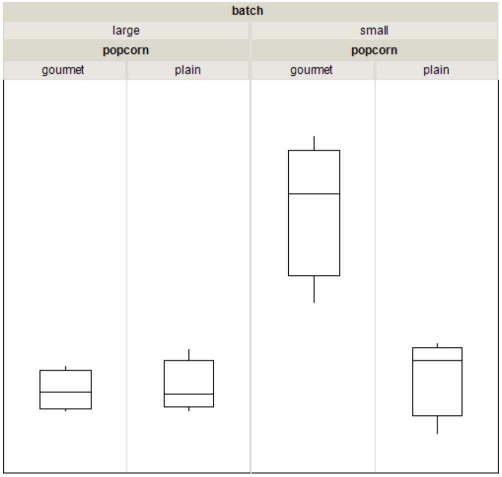
Uplift Model
Analyze > Consumer Research > Uplift



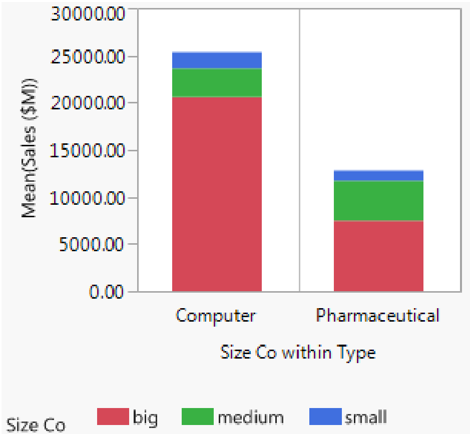
Dual Plot
Analyze > Consumer Research > Item Analysis



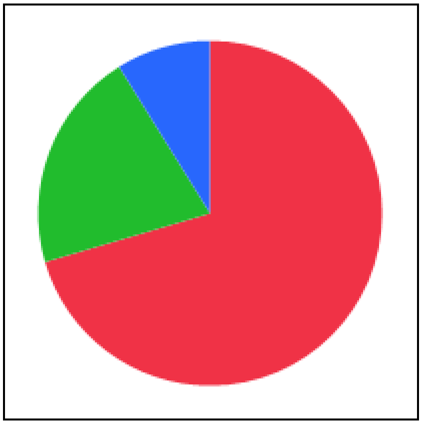
Line Graphs
Graph > Graph Builder



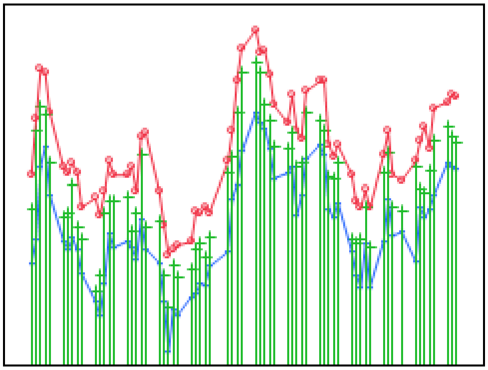
Box Plots
Graph > Graph Builder



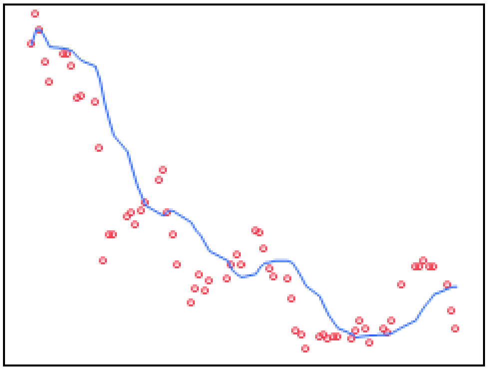
Stacked Bar Chart
Graph > Chart



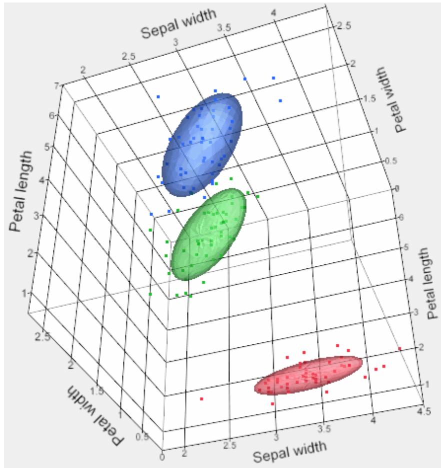
Pie Chart
Graph > Chart



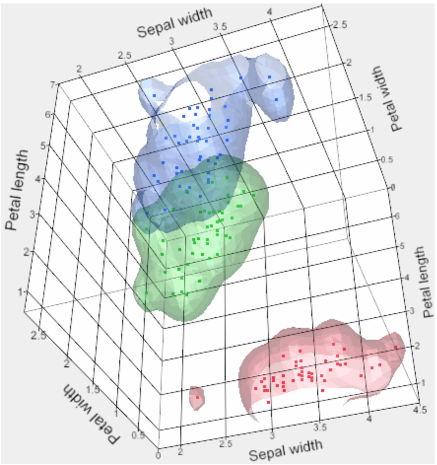
Needle and Line Chart
Graph > Overlay Plot



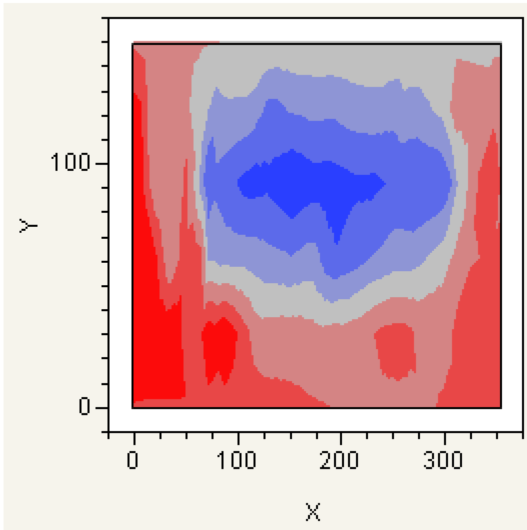
Dot and Line Chart
Graph > Overlay Plot



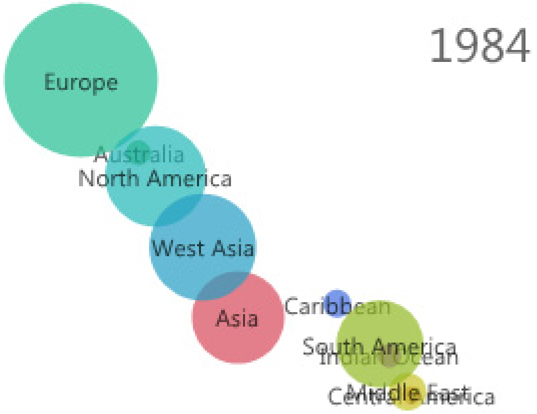
Three Dimensional Scatterplot
Graph > Scatterplot 3D



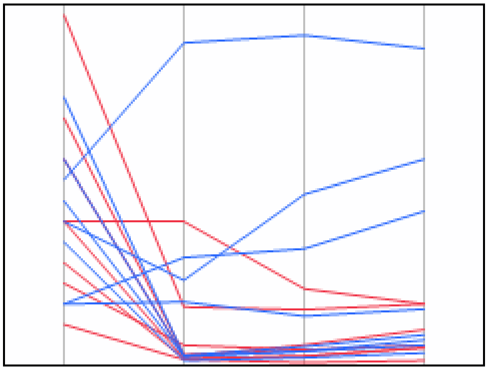
Three Dimensional Scatterplot
Graph > Scatterplot 3D



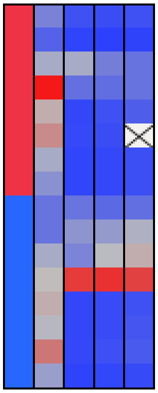
Contour Plot
Graph > Contour Plot



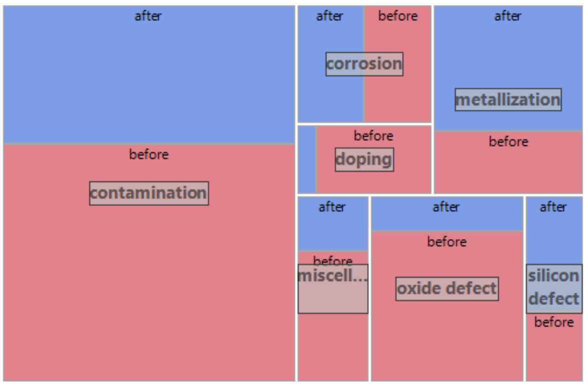
Bubble Plot
Graph > Bubble Plot



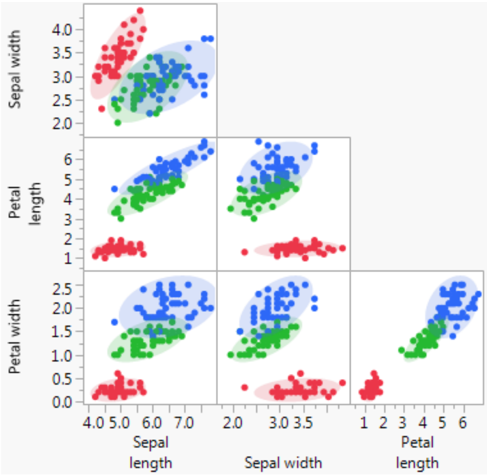
Parallel Plot
Graph > Parallel Plot



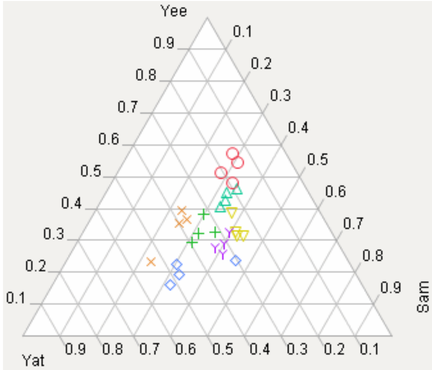
Cell Plot
Graph > Cell Plot



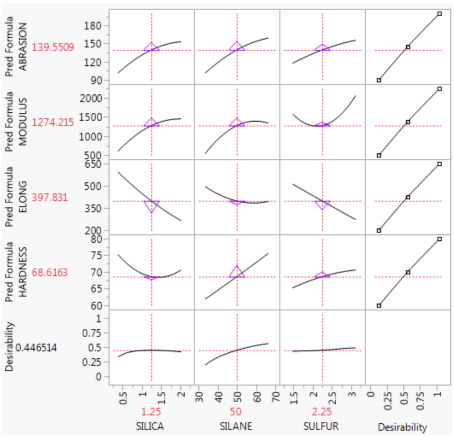
Treemap
Graph > Treemap



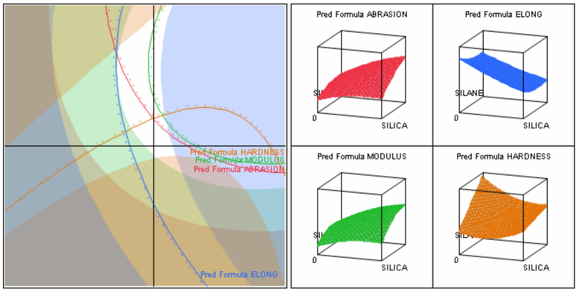
Scatterplot Matrix
Graph > Scatterplot Matrix



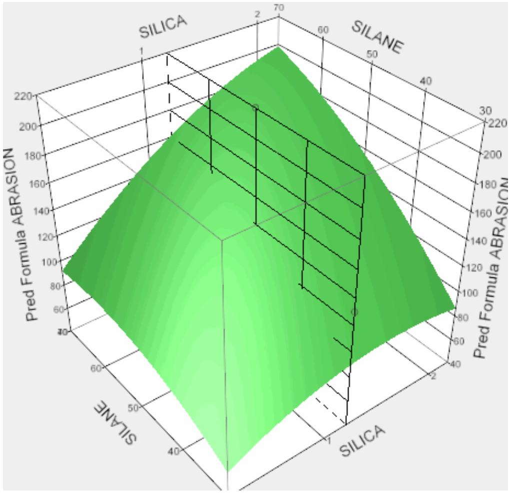
Ternary Plot
Graph > Ternary Plot



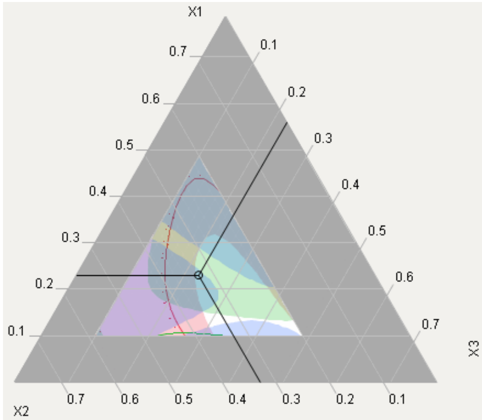
Prediction Profiler
Graph > Profiler



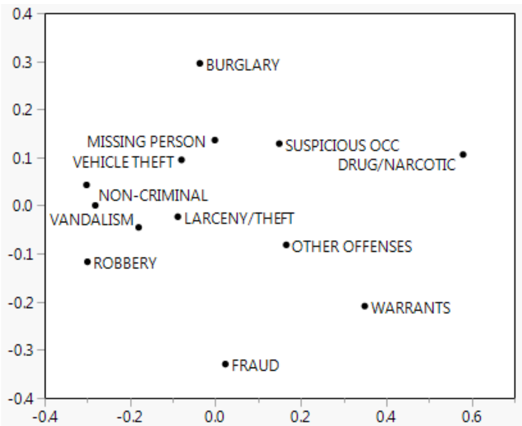
Contour Profiler
Graph > Contour Profiler



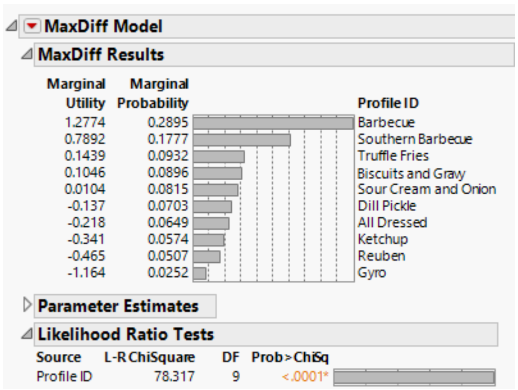
Surface Plot
Graph > Surface Plot



Mixture Profiler
Graph > Mixture Profiler



Multidimensional Scaling
Analyze > Consumer Research > Multidimensional Scaling



MaxDiff Model

MaxDiff Results

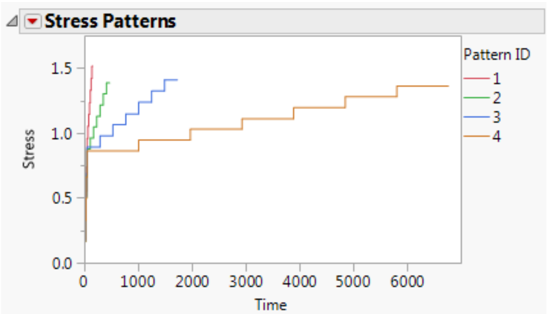
Marginal Utility	Marginal Probability	Profile ID
1.2774	0.2895	Barbecue
0.7892	0.1777	Southern Barbecue
0.1439	0.0932	Truffle Fries
0.1046	0.0896	Biscuits and Gravy
0.0104	0.0815	Sour Cream and Onion
-0.137	0.0703	Dill Pickle
-0.218	0.0649	All Dressed
-0.341	0.0574	Ketchup
-0.465	0.0507	Reuben
-1.164	0.0252	Gyro

Parameter Estimates

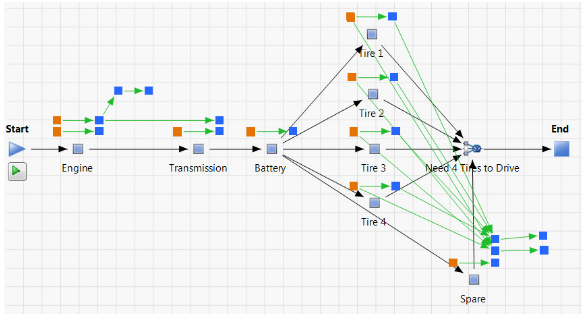
Likelihood Ratio Tests

Source	L-R ChiSquare	DF	Prob>ChiSq
Profile ID	78.317	9	<.0001*

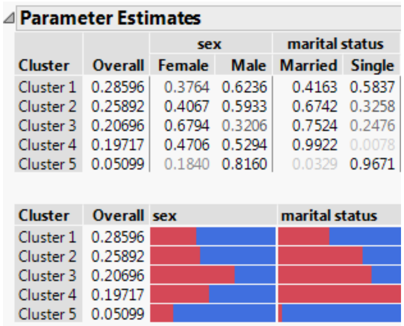
MaxDiff
Analyze > Consumer Research > MaxDiff



Stress Patterns Plot
Analyze > Reliability and Survival > Cumulative Damage



Repairable Systems Simulation
Analyze > Reliability and Survival > Repairable Systems Simulation

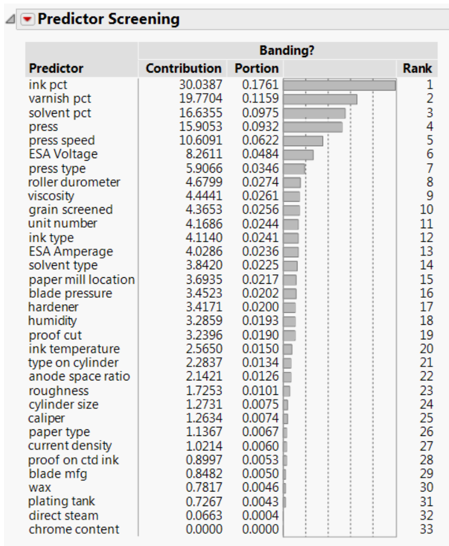


Parameter Estimates

Cluster	Overall	sex: Female	sex: Male	marital status: Married	marital status: Single
Cluster 1	0.28596	0.3764	0.6236	0.4163	0.5837
Cluster 2	0.25892	0.4067	0.5933	0.6742	0.3258
Cluster 3	0.20696	0.6794	0.3206	0.7524	0.2476
Cluster 4	0.19717	0.4706	0.5294	0.9922	0.0078
Cluster 5	0.05099	0.1840	0.8160	0.0329	0.9671

Cluster	Overall	sex	marital status
Cluster 1	0.28596		
Cluster 2	0.25892		
Cluster 3	0.20696		
Cluster 4	0.19717		
Cluster 5	0.05099		

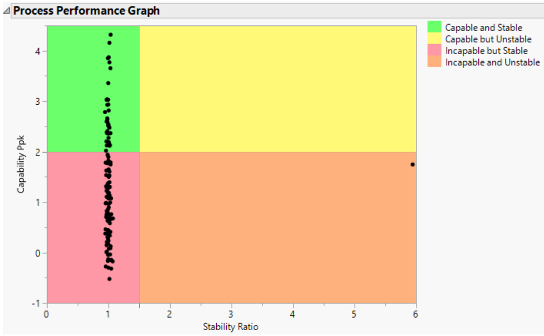
Latent Class Analysis
Analyze > Clustering > Latent Class Analysis



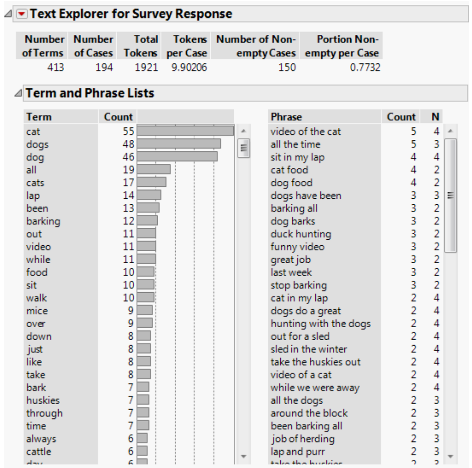
Predictor Screening

Predictor	Contribution	Portion	Rank
ink pct	30.0387	0.1761	1
varnish pct	19.7704	0.1159	2
solvent pct	16.6355	0.0975	3
press	15.9053	0.0932	4
press speed	10.6091	0.0622	5
ESA Voltage	8.2611	0.0484	6
press type	5.9066	0.0346	7
roller durometer	4.6799	0.0274	8
viscosity	4.4441	0.0261	9
grain screened	4.3653	0.0256	10
unit number	4.1686	0.0244	11
ink type	4.1140	0.0241	12
ESA Amperage	4.0286	0.0236	13
solvent type	3.8420	0.0225	14
paper mill location	3.6935	0.0217	15
blade pressure	3.4523	0.0202	16
hardener	3.4171	0.0200	17
humidity	3.2859	0.0193	18
proof cut	3.2396	0.0190	19
ink temperature	2.5650	0.0150	20
type on cylinder	2.2837	0.0134	21
anode space ratio	2.1421	0.0126	22
roughness	1.7253	0.0101	23
cylinder size	1.2731	0.0075	24
caliper	1.2634	0.0074	25
paper type	1.1367	0.0067	26
current density	1.0214	0.0060	27
proof on ctd ink	0.8997	0.0053	28
blade mfg	0.8482	0.0050	29
wax	0.7817	0.0046	30
plating tank	0.7267	0.0043	31
direct steam	0.0663	0.0004	32
chrome content	0.0000	0.0000	33

Predictor Screening
Analyze > Screening > Predictor Screening



Process Screening
Analyze > Screening > Process Screening



Text Explorer for Survey Response

Number of Terms	Number of Cases	Total Tokens	Tokens per Case	Number of Non-empty Cases	Portion Non-empty per Case
413	194	1921	9.90206	150	0.7732

Term and Phrase Lists

Term	Count
cat	55
dogs	48
dog	46
all	19
cats	17
lap	14
been	13
barking	12
out	11
video	11
while	11
food	10
sit	10
walk	10
mice	9
over	9
down	8
just	8
like	8
take	8
bark	7
huskies	7
through	7
time	7
always	6
cattle	6

Phrase	Count	N
video of the cat	5	4
all the time	5	3
sit in my lap	4	4
cat food	4	2
dog food	4	2
dogs have been	3	3
barking all	3	2
dog barks	3	2
duck hunting	3	2
funny video	3	2
great job	3	2
last week	3	2
stop barking	3	2
cat in my lap	2	4
dogs do a great	2	4
hunting with the dogs	2	4
out for a sled	2	4
sled in the winter	2	4
take the huskies out	2	4
video of a cat	2	4
while we were away	2	4
all the dogs	2	3
around the block	2	3
been barking all	2	3
job of herding	2	3
lap and purr	2	3

Text Explorer
Analyze > Text Explorer

Chapter 1

Learn about JMP

Documentation and Additional Resources

This chapter includes the following information:

- book conventions
- JMP documentation
- JMP Help
- additional resources, such as the following:
  - other JMP documentation
  - tutorials
  - indexes
  - Web resources
  - technical support options

Formatting Conventions

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

- Sample data table names, column names, pathnames, filenames, file extensions, and folders appear in Helvetica font.
- Code appears in `Lucida Sans Typewriter` font.
- Code output appears in *`Lucida Sans Typewriter`* italic font and is indented farther than the preceding code.
- **Helvetica bold** formatting 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
  - script output
- Features that are for JMP Pro only are noted with the JMP Pro icon . For an overview of JMP Pro features, visit http://www.jmp.com/software/pro/.

Note: Special information and limitations appear within a Note.

Tip: Helpful information appears within a Tip.

JMP Documentation

JMP offers documentation in various formats, from print books and Portable Document Format (PDF) to electronic books (e-books).

- Open the PDF versions from the **Help > Books** menu.
- All books are also combined into one PDF file, called *JMP Documentation Library*, for convenient searching. Open the *JMP Documentation Library* PDF file from the **Help > Books** menu.
- You can also purchase printed documentation and e-books on the SAS website:

  http://www.sas.com/store/search.ep?keyWords=JMP

JMP Documentation Library

The following table describes the purpose and content of each book 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.
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.
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.

Document Title	Document Purpose	Document Content
Essential Graphing	Find the ideal graph for your data.	Describes these Graph menu platforms: • Graph Builder • Overlay Plot • Scatterplot 3D • Contour Plot • Bubble Plot • Parallel Plot • Cell Plot • Treemap • Scatterplot Matrix • Ternary Plot • Chart 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.
Design of Experiments Guide	Learn how to design experiments and determine appropriate sample sizes.	Covers all topics in the DOE menu and the Specialized DOE Models menu item in the Analyze > Specialized Modeling menu.

Document Title	Document Purpose	Document Content
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: • Modeling Utilities • Neural • Partition • Bootstrap Forest • Boosted Tree • K Nearest Neighbors • Naive Bayes • Model Comparison • Formula Depot Describes these Analyze > Specialized Modeling menu platforms: • Fit Curve • Nonlinear • Gaussian Process • Time Series • Matched Pairs Describes these Analyze > Screening menu platforms: • Response Screening • Process Screening • Predictor Screening • Association Analysis The platforms in the Analyze > Specialized Modeling > Specialized DOE Models menu are described in *Design of Experiments Guide*.

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 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 • Pareto Plot • Diagram

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 and 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 • Multiple Correspondence Analysis • Multidimensional Scaling • Factor Analysis • Choice • MaxDiff • Uplift • Item 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.

Document Title	Document Purpose	Document Content
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.

Note: The **Books** menu also contains two reference cards that can be printed: The *Menu Card* describes JMP menus, and the *Quick Reference* describes JMP keyboard shortcuts.

JMP Help

JMP Help is an abbreviated version of the documentation library that provides targeted information. You can open JMP Help in several ways:

- On Windows, press the F1 key to open the Help system window.
- 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.
- Search and view JMP Help on Windows using the **Help > Help Contents**, **Search Help**, and **Help Index** options. On Mac, select **Help > JMP Help**.
- Search the Help at http://jmp.com/support/help/ (English only).

Additional Resources for Learning JMP

In addition to JMP documentation and JMP Help, you can also learn about JMP using the following resources:

- Tutorials (see "Tutorials" on page 44)
- Sample data (see "Sample Data Tables" on page 44)
- Indexes (see "Learn about Statistical and JSL Terms" on page 44)
- Tip of the Day (see "Learn JMP Tips and Tricks" on page 44)
- Web resources (see "JMP User Community" on page 45)
- JMPer Cable technical publication (see "JMPer Cable" on page 45)
- Books about JMP (see "JMP Books by Users" on page 46)
- JMP Starter (see "The JMP Starter Window" on page 46)

- Teaching Resources (see “Sample Data Tables” on page 44)

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, then 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\13\Samples\Data

On Macintosh: \Library\Application Support\JMP\13\Samples\Data

In JMP Pro, sample data is installed in the JMPPRO (rather than JMP) directory. In JMP Shrinkwrap, sample data is installed in the JMPSW 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 http://jmp.com/tools.

Learn about Statistical and JSL Terms

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

Statistics Index Provides definitions of statistical terms.

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

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. See the *Using JMP* book for details.

Tooltips

JMP provides descriptive tooltips when you place your cursor over items, such as the following:

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

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

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/.

JMPer Cable

The JMPer Cable is a yearly technical publication targeted to users of JMP. The JMPer Cable is available on the JMP website:

http://www.jmp.com/about/newsletters/jmpercable/

JMP Books by Users

Additional books about using JMP that are written by JMP users are available on the JMP website:

http://www.jmp.com/en_us/software/books.html

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 the Macintosh) **> 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 Macintosh, select **JMP > Preferences > Initial JMP Starter Window**.

Technical Support

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

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

Chapter 2

Introducing JMP

Basic Concepts

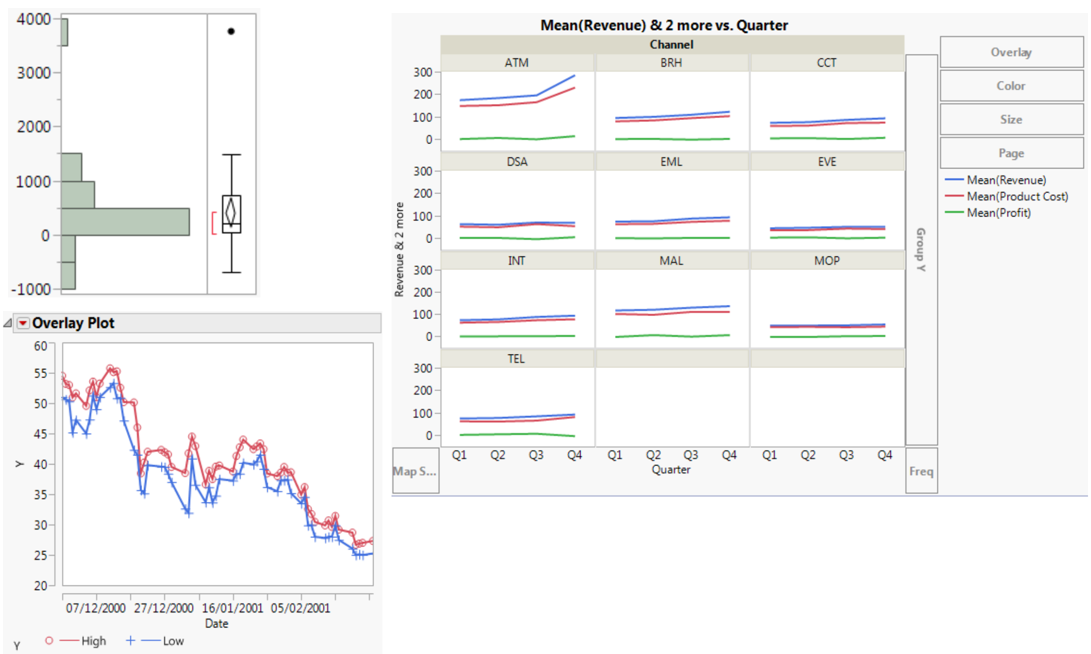
JMP (pronounced *jump*) is a powerful and interactive data visualization and statistical analysis tool. Use JMP to learn more about your data by performing analyses and interacting with the data using data tables, graphs, charts, and reports.

JMP enables researchers to perform a wide range of statistical analyses and modeling. JMP is equally useful to the business analyst who wants to quickly uncover trends and patterns in data. With JMP, you do not have to be an expert in statistics to get information from your data.

For example, you can use JMP to do the following:

- Create interactive graphs and charts to explore your data and discover relationships.
- Discover patterns of variation across many variables at once.
- Explore and summarize large amounts of data.
- Develop powerful statistical models to predict the future.

Figure 2.1 Examples of JMP Reports



Concepts That You Should Know

Before you begin using JMP, you should be familiar with these concepts:

- Enter, view, edit, and manipulate data using JMP *data tables*.
- Select a *platform* from the **Analyze**, **Graph**, or **DOE** menus. Platforms contain interactive windows that you use to analyze data and work with graphs.
- Platforms use these windows:
  – *Launch windows* where you set up and run your analysis.
  – *Report windows* showing the output of your analysis.
- Report windows normally contain the following items:
  – A graph of some type (such as a scatterplot or a chart).
  – Specific *reports* that you can show or hide using the *disclosure button* .
  – Platform *options* that are located within *red triangle menus* .

How Do I Get Started?

The general workflow in JMP is simple:

1. Get your data into JMP.
2. Select a platform and complete its launch window.
3. Explore your results and discover where your data takes you.

This workflow is described in more detail in "Understanding the JMP Workflow" on page 53.

Typically, you start your work in JMP by using graphs to visualize individual variables and relationships among your variables. Graphs make it easy to see this information, and to see the deeper questions to ask. Then you use analysis platforms to dig deeper into your problems and find solutions.

- The "Work with Your Data" chapter on page 61 shows you how to get data into JMP.
- The "Visualize Your Data" chapter on page 89 shows you how to use some of the useful graphs JMP provides to look more closely at your data.
- The "Analyze Your Data" chapter on page 123 shows you how to use some of the analysis platforms.
- The "The Big Picture" chapter on page 157 shows you how to analyze distributions, patterns, and similar values in several platforms.

Each chapter teaches through examples. The following sections in this chapter describe data tables and general concepts for working in JMP.

Starting JMP

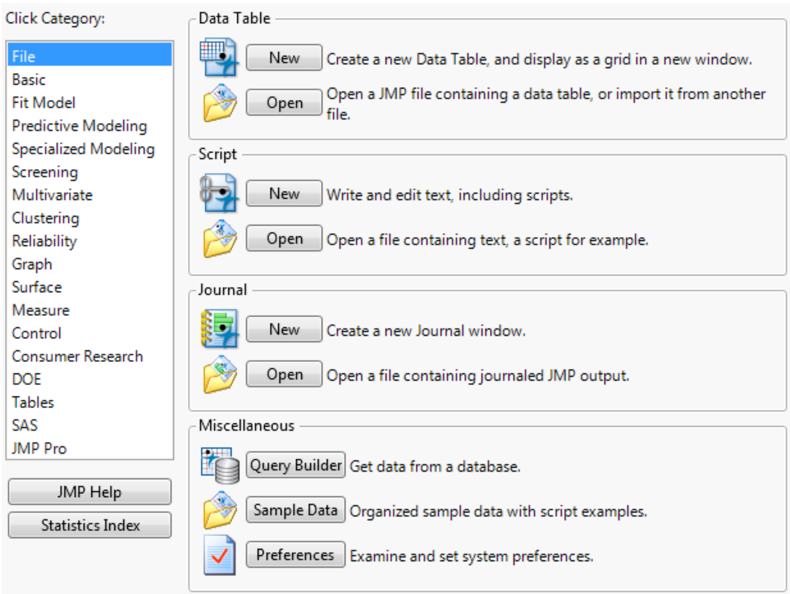
Start JMP in two ways:

- Double-click the JMP icon, normally found on your desktop. This starts JMP, but does not open any existing JMP files.
- Double-click an existing JMP file. This starts JMP and opens the file.

The initial view of JMP includes the Tip of the Day window and the Home Window on Windows; on Macintosh, the Tip of the Day and JMP Starter, and Home Window initially appear.

The JMP Starter window classifies actions and platforms using categories.

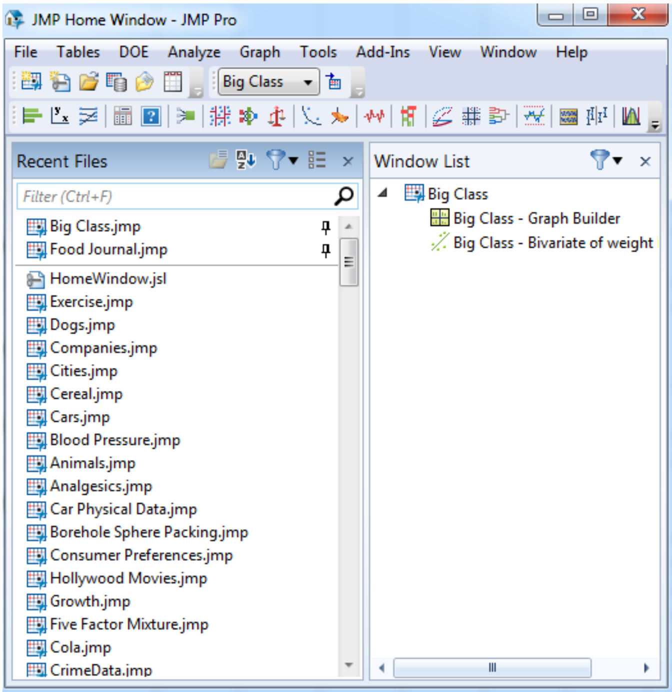
Figure 2.2 The JMP Starter



On the left is a list of categories. Click a category to see the features and the commands related to that category. The JMP Starter also lists JMP Pro features and platforms.

The Home Window helps you organize and access files in JMP.

Figure 2.3 The Home Window on Windows



To open the Home Window on Windows, select **View > Home Window**. On Macintosh, select **Window > JMP Home**. The Home Window includes links to the following:

- the data tables and report windows that are currently open
- files that you have opened recently

For more details about the Home Window, see the Get Started chapter in the *Using JMP* book.

Almost all JMP windows contain a menu bar and a toolbar. You can find most JMP features in three ways:

- using the menu bar
- using the toolbar buttons
- using the buttons on the JMP Starter window

About the Menu Bar and Toolbars

The menus and toolbars are hidden in many windows. To see them, place your mouse pointer over the blue bar under the window's title bar. The menus in the JMP Starter window, the Home Window, and all data tables are always visible.

Using Sample Data

The examples in this book and the other JMP books use sample data tables. The default location on Windows for the sample data is:

C:/Program Files/SAS/JMP/13/Samples/Data

C:/Program Files/SAS/JMPPro/13/Samples/Data

C:/Program Files/SAS/JMPSW/13/Samples/Data

The Sample Data Index groups the data tables by category. Click a disclosure button to see a list of data tables for that category, and then click a link to open a data table.

Macintosh sample data is installed in /Library/Application Support/JMP/13/Samples/Data.

Opening a JMP Sample Data Table

1. From the **Help** menu, select **Sample Data**.
2. Open the **Data Tables used in Discovering JMP** list by clicking on the disclosure button next to it.
3. Click the name of the data table to use it in the examples in this book.

Sample Import Data

Use files from other applications to learn how to import data into JMP.

The default location on Windows for the sample import data is:

C:/Program Files/SAS/JMP/13/Samples/Import Data

C:/Program Files/SAS/JMPPro/13/Samples/Import Data

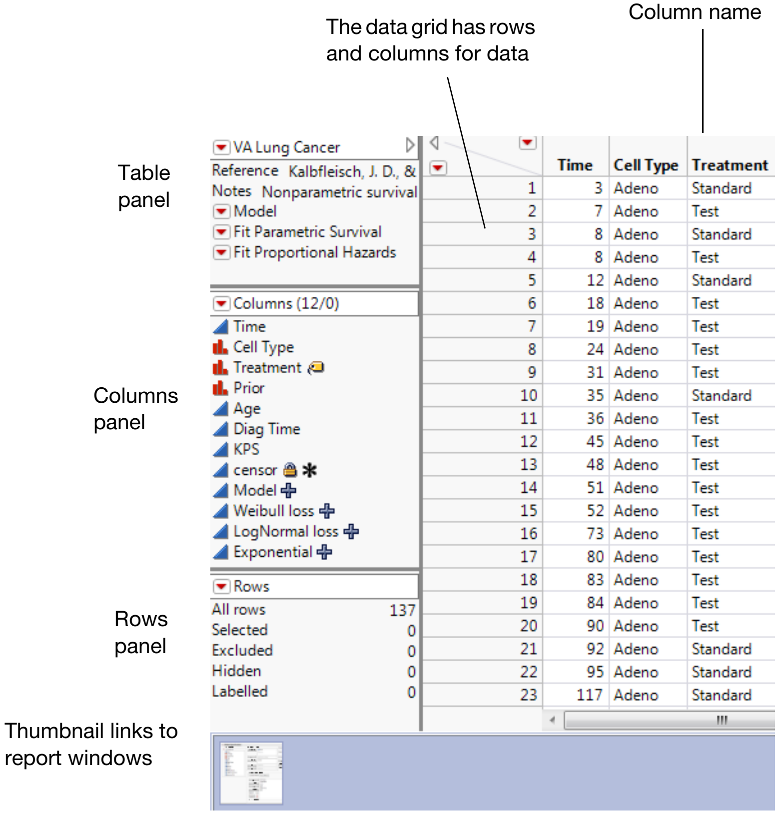
C:/Program Files/SAS/JMPSW/13/Samples/Import Data

Understanding Data Tables

A data table is a collection of data organized in rows and columns. It is similar to a Microsoft® Excel® spreadsheet, but with some important differences that are discussed in "How is JMP Different from Excel?" on page 58. A data table might also contain other information like notes, variables, and scripts. These supplementary items are discussed in later chapters.

Open the VA Lung Cancer data table to see the data table described here.

Figure 2.4 A Data Table



A data table contains the following parts:

Data grid The data grid contains the data arranged in rows and columns. Generally, each row in the data grid is an observation, and the columns (also called variables) give information about the observations. In Figure 2.4, each row corresponds to a test subject, and there are twelve columns of information. Although all twelve columns cannot be shown in the data grid, the Columns panel lists them all. The information given about each test subject includes the time, cell type, treatment, and more. Each column has a header, or name. That name is not part of the table's total count of rows.

Table panel The table panel can contain table variables or table scripts. In Figure 2.4, there is one saved script called **Model** that can automatically re-create an analysis. This table also has a variable named Notes that contains information about the data. Table variables and table scripts are discussed in a later chapter.

Columns panel The columns panel shows the total number of columns, whether any columns are selected, and a list of all the columns by name. The numbers in parentheses (12/0) show that there are twelve columns, and that no columns are selected. An icon to the left of each column name shows that column's modeling type. Modeling types are described in

"Understand Modeling Types" on page 126 in the "Analyze Your Data" chapter. Icons to the right show any attributes assigned to the column. See "View or Change Column Information" on page 73 in the "Work with Your Data" chapter for more information about these icons.

Rows panel The rows panel shows the number of rows in the data table, and how many rows are selected, excluded, hidden, or labeled. In Figure 2.4, there are 137 rows in the data table.

Thumbnail links to report windows This area shows thumbnails of all reports based on the data table. Place your mouse pointer over a thumbnail to see a larger preview of the report window. Double-click a thumbnail to bring the report window to the front.

Interacting with the data grid, which includes adding rows and columns, entering data, and editing data, is discussed in the "Work with Your Data" chapter on page 61. If you open multiple data tables, each one appears in a separate window.

Understanding the JMP Workflow

Once your data is in a data table, you can create graphs or plots, and perform analyses. All features are located in platforms, which are found primarily on the **Analyze** or **Graph** menus. They are called platforms because they do not just produce simple static results. Platform results appear in report windows, are highly interactive, and are linked to the data table and to each other.

The platforms under the **Analyze** and **Graph** menus provide a variety of analytical features and data exploration tools.

The general steps to produce a graph or analysis are as follows:

1. Open a data table.
2. Select a platform from the Graph or Analyze menu.
3. Complete the platform launch window to set up your analysis.
4. Click **OK** to create the report window that contains your graphs and statistical analyses.
5. Customize your report by using report options.
6. Save, export, and share your results with others.

Later chapters discuss these concepts in greater detail.

The following example shows you how to perform a simple analysis and customize it in four steps. This example uses the Companies.jmp file sample data table to show a basic analysis of the variable Profits ($M).

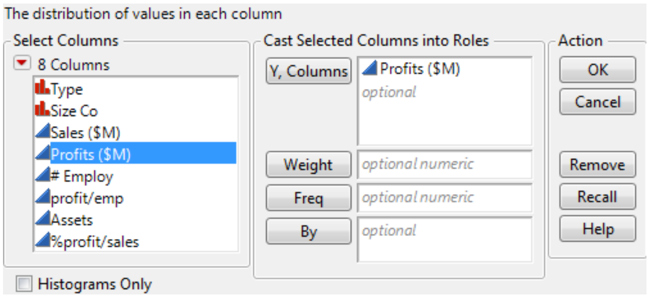
Step 1: Launching a Platform and Viewing Results

1. Select **Help > Sample Data Library** and open Companies.jmp.
2. Select **Analyze > Distribution** to open the Distribution launch window.
3. Select Profits ($M) in the Select Columns box and click the **Y, Columns** button.

   The variable Profits ($M) appears in the **Y, Columns** role. See Figure 2.5 for the completed window.

   Another way to assign variables is to click and drag columns from the Select Columns box to any of the roles boxes.

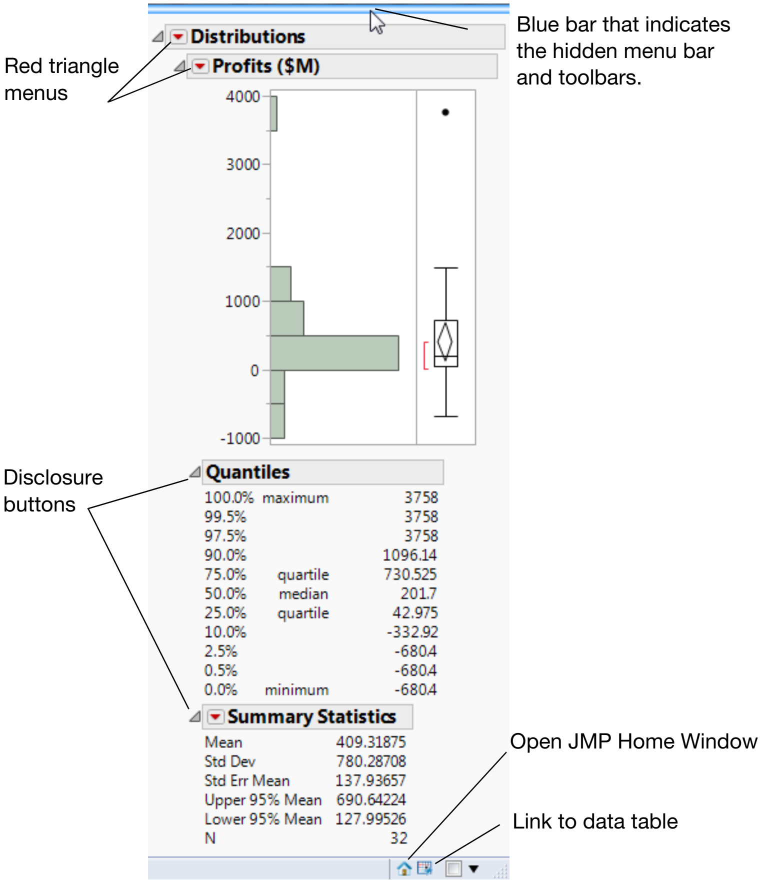
Figure 2.5 Assign Profits ($M)



4. Click **OK**.

   The Distribution report window appears.

Figure 2.6 Distribution Report Window on Windows



The report window contains basic plots or graphs and preliminary analysis reports. The results appear in an outline format, and you can show or hide any report by clicking on the disclosure button.

Red triangle menus contain options and commands to request additional graphs and analyses at any time.

- On Windows, place your mouse pointer over the blue bar at the top of the window to see the menu bar and the toolbars.
- On Windows, click the data table button in the lower right corner to view the data table that was used to create this report. On Macintosh, click the **Show Data Table** button in the upper right corner of the report window.
- On Windows, click the **JMP Home Window** button in the lower right corner to view the Home Window. On Macintosh, select **Window > JMP Home**.

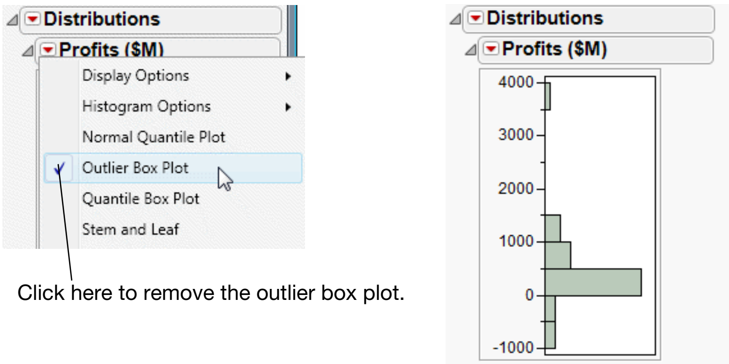
Step 2: Removing the Box Plot

Continue using the Distribution report that you created earlier.

1. Click the red triangle next to Profits ($M) to see a menu of report options.
2. Deselect **Outlier Box Plot** to turn the option off.

   The outlier box plot is removed from the report window.

Figure 2.7 Removing the Outlier Box Plot



Step 3: Requesting Additional Output

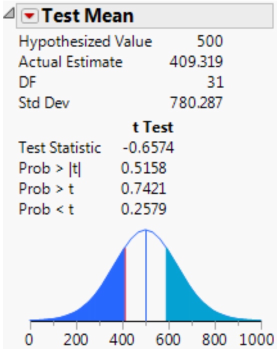
Continue to use the same report window.

1. From the red triangle menu next to Profits ($M), select **Test Mean**.

   The Test Mean window appears.
2. Enter 500 in the **Specify Hypothesized Mean** box.
3. Click **OK**.

   The test for the mean is added to the report window.

Figure 2.8 Test for the Mean



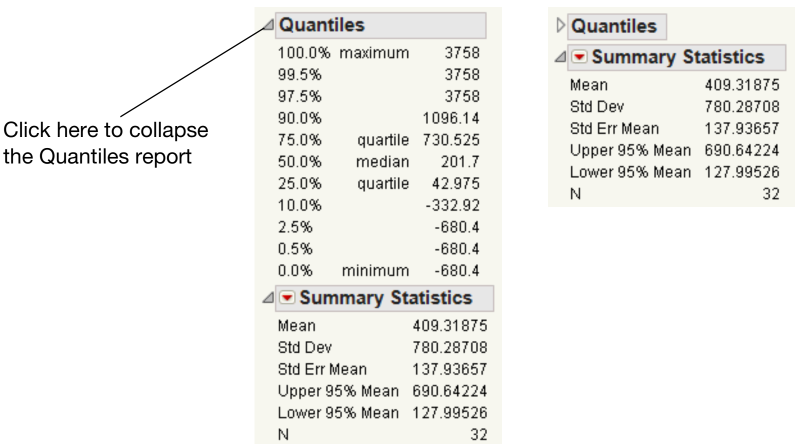
Step 4: Interacting with Platform Results

All platforms produce results that are interactive. For example:

- Reports can be shown or hidden.
- Additional graphs and statistical details can be added or removed to suit your purposes.
- Platform results are connected to the data table and to each other.

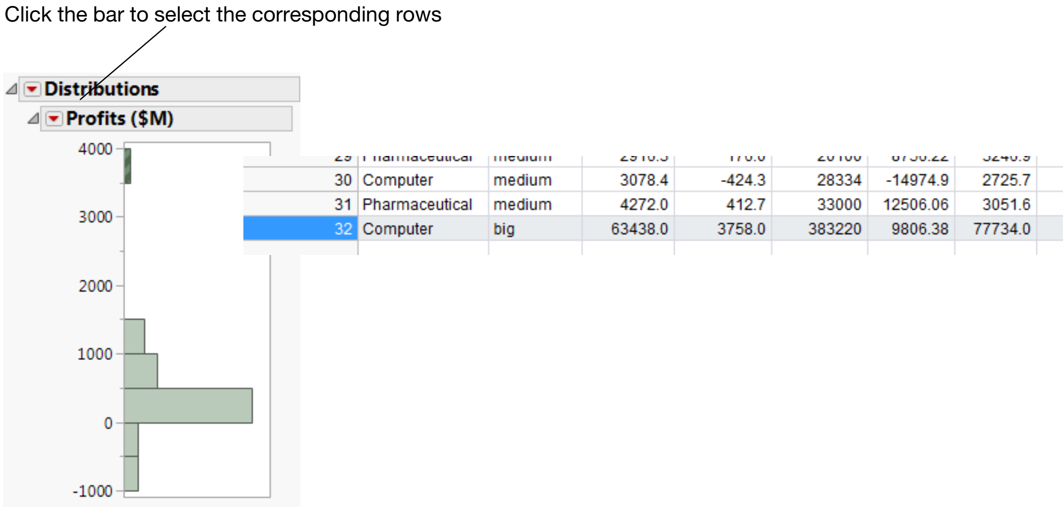
For example, to close the **Quantiles** report, click the disclosure button next to **Quantiles**.

Figure 2.9 Close the Quantiles Report



Platform results are connected to the data table. The histogram in Figure 2.10 shows that a group of companies makes a much higher profit that the others. To quickly identify that group, click on the histogram bar for them. The corresponding rows in the data table are selected.

Figure 2.10 Connection between Platform Results and Data Table



In this case, the group includes only one company, and that one row is selected.

How is JMP Different from Excel?

There are a number of important differences between JMP and Excel or other spreadsheet applications.

Table 2.1 JMP and Excel Differences

Formulas	
Excel	Formulas are applied to individual cells.
JMP	Formulas are applied only to entire columns. “Calculate Values with Formulas” on page 75 in the “Work with Your Data” chapter describes how to use formulas.
Column Names	
Excel	Column names are part of the grid. Numbered rows and labeled columns extend past the data. Numeric and character data reside in the same column.

Table 2.1 JMP and Excel Differences *(Continued)*

JMP	Column names are not part of the grid. There are no rows and columns beyond the existing data. The grid is only as big as the data. A column is either numeric or character. If a column contains both character and numeric data, the entire column's data type is character, and the numbers are treated as character data.
	"Understand Modeling Types" on page 126 in the "Analyze Your Data" chapter describes how data type influences platform results.
Tables and Worksheets	
Excel	A single spreadsheet contains several tables, or worksheets.
JMP	JMP does not have the concept of worksheets. Each data table is a separate .jmp file and appears in a separate window.
Data Grid	
Excel	Data can be located anywhere in the data grid.
JMP	Data always begins in row 1 and column 1.
Analysis and Graph Reports	
Excel	All data, analyses, and graphs are placed inside the data grid.
JMP	Results appear in a separate window.

Chapter 3

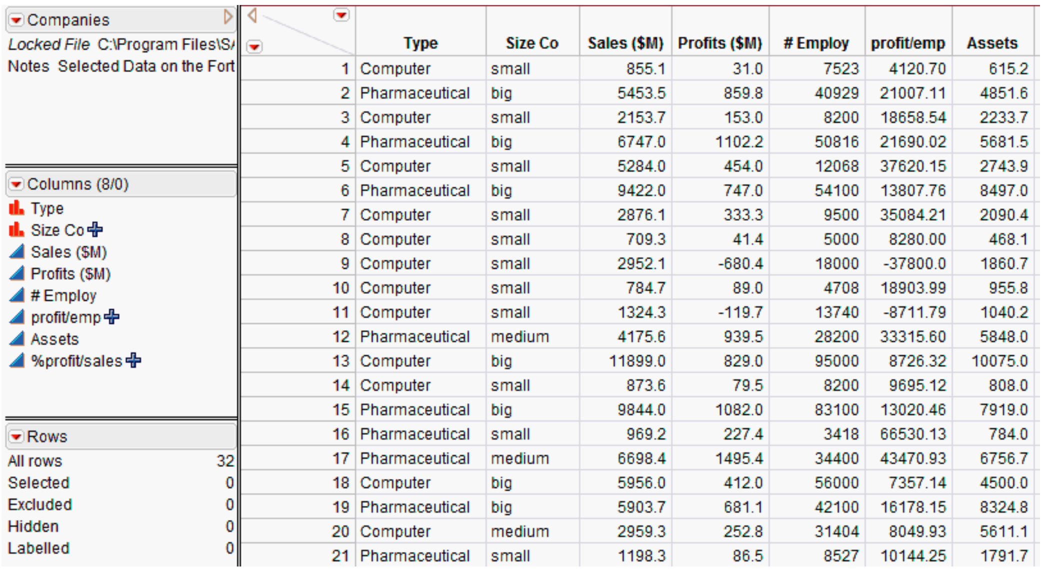
Work with Your Data

Prepare Your Data for Graphing and Analyzing

Before graphing or analyzing your data, the data has to be in a data table and in the proper format. This chapter shows some basic data management tasks, including the following:

- Creating new data tables
- Opening existing data tables
- Importing data from other applications into JMP
- Managing your data

Figure 3.1 Example of a Data Table



	Type	Size Co	Sales ($M)	Profits ($M)	# Employ	profit/emp	Assets
1	Computer	small	855.1	31.0	7523	4120.70	615.2
2	Pharmaceutical	big	5453.5	859.8	40929	21007.11	4851.6
3	Computer	small	2153.7	153.0	8200	18658.54	2233.7
4	Pharmaceutical	big	6747.0	1102.2	50816	21690.02	5681.5
5	Computer	small	5284.0	454.0	12068	37620.15	2743.9
6	Pharmaceutical	big	9422.0	747.0	54100	13807.76	8497.0
7	Computer	small	2876.1	333.3	9500	35084.21	2090.4
8	Computer	small	709.3	41.4	5000	8280.00	468.1
9	Computer	small	2952.1	-680.4	18000	-37800.0	1860.7
10	Computer	small	784.7	89.0	4708	18903.99	955.8
11	Computer	small	1324.3	-119.7	13740	-8711.79	1040.2
12	Pharmaceutical	medium	4175.6	939.5	28200	33315.60	5848.0
13	Computer	big	11899.0	829.0	95000	8726.32	10075.0
14	Computer	small	873.6	79.5	8200	9695.12	808.0
15	Pharmaceutical	big	9844.0	1082.0	83100	13020.46	7919.0
16	Pharmaceutical	small	969.2	227.4	3418	66530.13	784.0
17	Pharmaceutical	medium	6698.4	1495.4	34400	43470.93	6756.7
18	Computer	big	5956.0	412.0	56000	7357.14	4500.0
19	Pharmaceutical	big	5903.7	681.1	42100	16178.15	8324.8
20	Computer	medium	2959.3	252.8	31404	8049.93	5611.1
21	Pharmaceutical	small	1198.3	86.5	8527	10144.25	1791.7

Get Your Data into JMP

- To copy and paste data from another application, see "Copy and Paste Data" on page 62.
- To import data from another application, see "Import Data" on page 62.
- To enter data directly into a data table, see "Enter Data" on page 65
- To open a data table, double-click on the file, or use the **File > Open** command.

You can also import data into JMP from a database. For more information, see the Import Your Data chapter in the *Using JMP* book.

This chapter uses sample data tables and sample import data that is installed with JMP. To find these files, see "Using Sample Data" on page 51 in the "Introducing JMP" chapter.

Copy and Paste Data

You can move data into JMP by copying and pasting from another application, such as Excel or a text file.

1. Open the VA Lung Cancer.xls file in Excel. This file is located in the Sample Import Data folder.
2. Select all of the rows and columns, including the column names. There are 12 columns and 138 rows.
3. Copy the selected data.
4. In JMP, select **File > New > Data Table** to create an empty table.
5. Select **Edit > Paste with Column Names** to paste the data and column headings.

   If the data that you are pasting into JMP does *not* have column names, then you can use **Edit > Paste**.

Import Data

You can move data into JMP by importing data from another application, such as Excel, SAS, or text files. The basic steps to import data are as follows:

1. Select **File > Open**.
2. Navigate to your file's location.
3. If your file is not listed in the Open Data File window, select the correct file type from the **Files of type** menu.
4. Click **Open**.

Example of Importing a Microsoft Excel File

1. Select **File > Open**.
2. Navigate to the Samples/Import Data folder.
3. Select Team Results.xls.

   Note the rows and columns on which the data begin. The spreadsheet also contains two worksheets. In this example, you import the Ungrouped Team Results worksheet.
4. Click **Open**.

   The spreadsheet opens in the Excel Import Wizard, where a preview of the data appears along with import options.

   Text from the first row of the spreadsheet are column headings. However, you want text in row 3 of the spreadsheet to be converted to column headings.
5. Next to **Column headers start on row**, type 3, and press **Enter**. The column headings are updated in the data preview. The value for the first row of data is updated to 4.
6. Save the settings only for this worksheet:
   – Deselect **Use for all worksheets** in the lower left corner of the window.
   – Select **Ungrouped Team Results** in the upper right corner of the window.
7. Click **Import** to convert the spreadsheet as you specified.

When you import Excel files, JMP predicts whether columns headings exist, and if the column names are on row one. The copy and paste method is recommended for the following situations:

- If the column names are located in a row other than row one
- If the file does not include column names and the data does not start in row one
- If the file contains column names and the data does not start in row two

See "Copy and Paste Data" on page 62 and the Import Your Data chapter in the *Using JMP* book for more information about importing Excel files.

Example of Importing a Text File

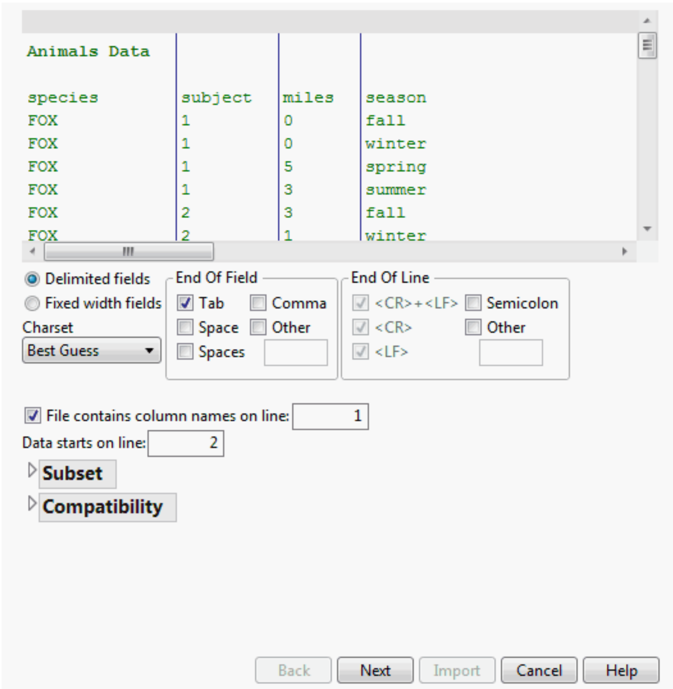
One way to import a text file is to let JMP assume the data's format and place the data in a data table. This method uses settings that you can specify in Preferences. See the JMP Preferences chapter in the *Using JMP* book for information about setting text import preferences.

Another way to import a text file is to use a Text Preview window to see what your data table will look like after importing, and make adjustments. The following example shows you how to use Text Import Preview window.

1. Select **File > Open**.
2. Navigate to the Samples/Import Data folder.

3. Select Animals_line3.txt.
4. At the bottom of the Open window, select **Data with Preview**.
5. Click **Open**.

Figure 3.2 Initial Preview Window



This text file has a title on the first line, column names on the third line, and the data starts on line four. If you opened this directly in JMP, the Animals Data line would be the first column name, and all the column names and data afterward would be out of sync. The Preview window lets you adjust the settings before you open the file, and see how your adjustments affect the final data table.

6. Enter 3 in the **File contains column names on line** field.
7. Enter 4 in the **Data starts on line** field.
8. Click **Next**.

   In the second window, you can exclude columns from the import and change the data modeling of the columns. For this example, use the default settings.
9. Click **Import**.

The new data table has columns named species, subject, miles, and season. The species and season columns are character data. The subject and miles columns are continuous numeric data.

Enter Data

You can enter data directly in a data table. The following example shows you how to enter data that was collected over several months into a data table.

Scenario

Table 3.1 shows the data from a study that investigated a new blood pressure medication. Each individual's blood pressure was measured over a six-month period. Two doses (300mg and 450mg) of the medication were used, along with a control and placebo group. The data shows the average blood pressure for each group.

Table 3.1 Blood Pressure Data

Month	Control	Placebo	300mg	450mg
March	165	163	166	168
April	162	159	165	163
May	164	158	161	153
June	162	161	158	151
July	166	158	160	148
August	163	158	157	150

Enter Data in a New Data Table

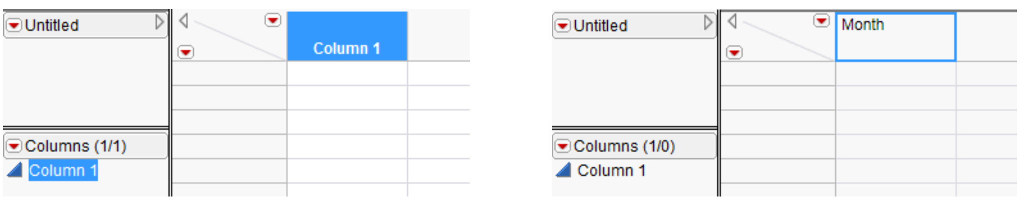
1. Select **File > New > Data Table** to create an empty data table.

   A new data table has one column and no rows.

2. Select the column name and change the name to Month. See Figure 3.3

   Note: To rename a column, you can also double-click the column name or select the column and press Enter.

Figure 3.3 Entering a Column Name



Click once to select the column.

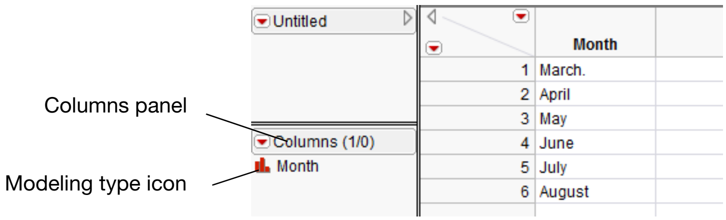
Then type “Month”.

3. Select **Rows > Add Rows**.

   The Add Rows window appears.

4. Since you want to add six rows, type 6.
5. Click **OK**. Six empty rows are added to the data table.
6. Enter the Month information by clicking in a cell and typing.

Figure 3.4 Month Column Completed



   In the columns panel, look at the modeling type icon to the left of the column name. It has changed to reflect that Month is now nominal (previously it was continuous). Compare the modeling type shown for Column 1 in Figure 3.3 and for Month in Figure 3.4. This difference is important and is discussed in "View or Change Column Information" on page 73.

7. Double-click in the space on the right side of the Month column to add the Control column.
8. Change the name to Control.
9. Enter the Control data as shown in Table 3.1. Your data table now consists of six rows and two columns.
10. Continue adding columns and entering data as shown in Table 3.1 to create the final data table with six rows and five columns.

Change the Data Table Name

1. Double-click on the data table name (Untitled) in the Table Panel.
2. Type the new name (Blood Pressure).

Figure 3.5 Changing the Data Table Name

Transfer Data from Excel

You can use the JMP Add In for Excel to transfer a spreadsheet from Excel to JMP:

- a data table
- Graph Builder
- Distribution platform
- Fit Y by X platform
- Fit Model platform
- Time Series platform
- Control Chart platform

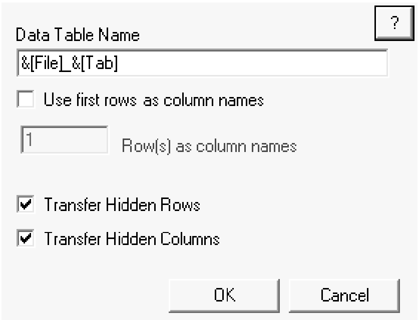
Set JMP Add In Preferences in Excel

To configure JMP Add In Preferences:

1. In Excel, select **JMP > Preferences**.

   The JMP Preferences window appears.

Figure 3.6 JMP Add In Preferences



2. Accept the default **Data Table Name** (File name_Worksheet name) or type a name.
3. Select to **Use the first rows as column names** if the first row in the worksheet contains column headers.
4. If you selected to use the first rows a column headers, type the number of rows used.
5. Select to **Transfer Hidden Rows** if the worksheet contains hidden rows to be included in the JMP data table.
6. Select to **Transfer Hidden Column** if the worksheet contains hidden columns to be included in the JMP data table.
7. Click **OK** to save your preferences.

Transfer to JMP

To transfer an Excel worksheet to JMP:

1. Open the Excel file.
2. Select the worksheet to transfer.
3. Select **JMP** and then select the JMP destination:
   – Data Table
   – Graph Builder
   – Distribution platform
   – Fit Y by X platform
   – Fit Model platform
   – Time Series platform
   – Control Chart platform

   The Excel worksheet is opened as a data table in JMP and the selected platform's launch window appears.

Work with Data Tables

This section contains the following information:

- "Edit Data" on page 68
- "Select, Deselect, and Find Values" on page 70
- "View or Change Column Information" on page 73
- "Calculate Values with Formulas" on page 75
- "Filter Data" on page 77

Tip: Consider setting the Autosave Time Out value in the General preferences to automatically save open data tables at the specified number of minutes. This autosave value also applies to journals, scripts, projects, and reports.

Edit Data

You can enter or change data, either a few cells at a time or for an entire column. This section contains the following information:

- "Change Values" on page 69
- "Recode Values" on page 69

- "Create Patterned Data" on page 69

Change Values

To change a value, select a cell and type the change. You can also double-click a cell to edit it.

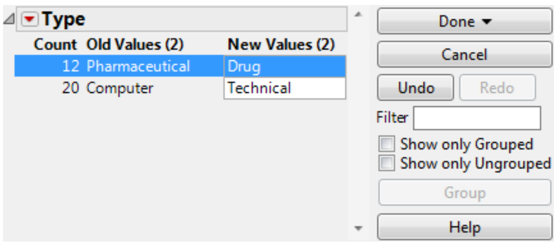
Note: Double-clicking in a cell is not the same as selecting a cell. A single click selects a cell. You can select more than one cell at the same time, and you can perform certain actions on selected cells. Double-clicking only lets you edit a cell. For more information about selecting rows, columns, and cells, see "Select, Deselect, and Find Values" on page 70.

Recode Values

Use the recoding tool to change all of the values in a column at once. For example, suppose that you are interested in comparing the sales of computer and pharmaceutical companies. Your current company labels are Computer and Pharmaceutical. You want to change them to Technical and Drug. Going through all 32 rows of data and changing all the values would be tedious, inefficient, and error-prone, especially if you had many more rows of data. Recode is a better option.

1. Select **Help > Sample Data Library** and open Companies.jmp.
2. Select the Type column by clicking once on the column heading.
3. Select **Cols > Utilities > Recode.**
4. In the New Value column of the Recode window, type Technical in the Computer row and Drug in the Pharmaceutical row.
5. Click **Done** and select the **In Place** option from the list.

Figure 3.7 Recode Window



All cells are updated automatically to the new values.

Create Patterned Data

Use the Fill options to populate a column with patterned data. The Fill options are especially useful if your data table is large, and typing in the values for each row would be cumbersome.

Example of Filling a Column with the Pattern

1. Add a new column.
2. Enter 1 in the first cell, 2 in the second cell, and 3 in the third cell.
3. Select the three cells, and right-click anywhere in the selected cells to see a menu.
4. Select **Fill > Repeat sequence to end of table**.

   The rest of the column is filled with the sequence (1, 2, 3, 1, 2, 3, ...).

To continue a pattern instead of repeating it (1, 2, 3, 4, 5, 6, ...), select **Continue sequence to end of table**. This command can also be used to generate patterns like (1, 1, 1, 2, 2, 2, 3, 3, 3, ...).

The Fill options can recognize simple arithmetic and geometric sequences. For character data, the Fill options only repeat the values.

Select, Deselect, and Find Values

You can select rows, columns, or cells within a data table. For example, to create a subset of an existing data table, you must first select the parts of the table that you want to subset. Also, selecting rows can make data points stand out on a graph. Select rows and columns manually by clicking, or select rows that meet certain search criteria. This section contains the following information:

- "Select and Deselect Rows" on page 70
- "Select and Deselect Columns" on page 71
- "Select and Deselect Cells" on page 72
- "Search for Values" on page 72

Select and Deselect Rows

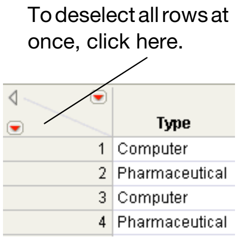
Table 3.2 Selecting and Deselecting Rows

Task	Action
Select rows one at a time	Click on the row number.
Select multiple adjacent rows	Click and drag on the row numbers. or Select the beginning row, and then hold down the Shift key and click the last row number.
Select multiple non-adjacent rows	Select the first row, and then hold down the Ctrl key and click the other row numbers.
Deselect rows one at a time	Hold down the Ctrl key and click the row numbers.

Table 3.2 Selecting and Deselecting Rows *(Continued)*

Task	Action
Deselect all rows	Click in the lower-triangular space in the top left corner of the table. See Figure 3.8.

Figure 3.8 Deselecting Rows

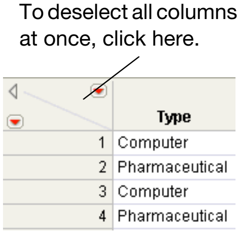


Select and Deselect Columns

Table 3.3 Selecting and Deselecting Columns

Task	Action
Select columns one at a time	Click the column heading.
Select multiple adjacent columns	Click and drag across the column headings. or Select the beginning column, and then hold down the Shift key and click the last header.
Select multiple non-adjacent columns	Select the first column, and then hold down the Ctrl key and click the other column headings.
Deselect columns one at a time	Hold down the Ctrl key and click the column heading.
Deselect all columns	Click in the upper-triangular space in the top left corner of the table. See Figure 3.9.

Figure 3.9 Deselecting Columns



Select and Deselect Cells

Table 3.4 Selecting and Deselecting Cells

Task	Action
Select cells one at a time	Click each cell individually.
Select multiple adjacent cells	Click and drag across the cells. or Select the beginning cell, and then hold down the Shift key and click the last cell.
Select multiple non-adjacent cells	Select the first cell, and then hold down the Ctrl key and click the other cells.
Deselect all cells	Click in the upper and lower triangular spaces in the top left corner of the table.

Search for Values

In a data table that has thousands or tens of thousands of rows, it can be difficult to locate a particular cell by scrolling through the table. If you are looking for specific information, use the Search feature to find it. If data is matches the search criteria, the cell is selected and the data grid scrolls to show it in the window. For example, the Companies.jmp data table contains information about a company that has total sales of $11,899. Use the Search feature to find that cell.

Example of Searching for a Value

1. Select **Edit > Search > Find** to launch the Search window.
2. In the **Find what** box, enter 11899.
3. Click **Find**. JMP finds the first cell that has 11,899 in it, and selects it.

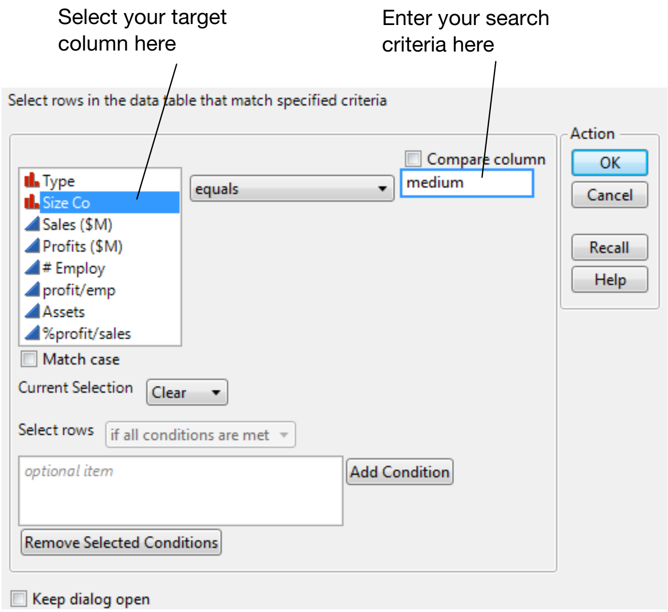
If multiple cells meet the search criteria, click **Find** again to find the next cell that matches the search term.

You can also search for multiple rows at once, with each row matching some criteria.

Example of Select All Rows That Correspond to Medium-Sized Companies

1. Select **Rows > Row Selection > Select Where** to open the **Select rows** window.
2. In the column list box on the left, select Size Co.
3. In the text box on the right, enter medium.
4. Click **OK**.

Figure 3.10 Select Rows Window



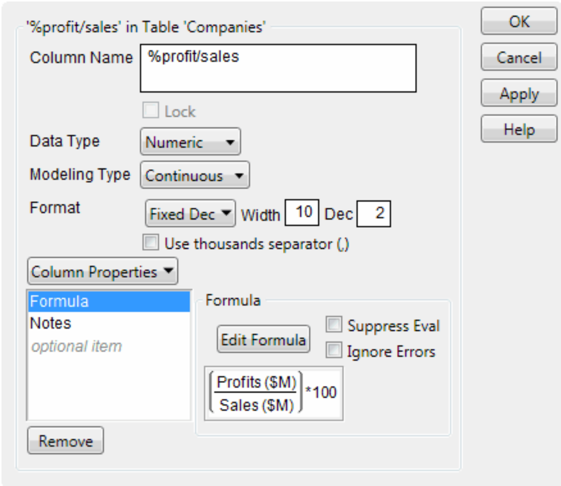
JMP selects all of the rows that have Size Co equal to medium. There are seven.

View or Change Column Information

Information about a column is not limited to the data in the column. Data type, modeling type, format, and formulas can also be set.

To view or change column characteristics, double-click the column heading. Or, right-click the column heading and select **Column Info**. The Column Info window appears.

Figure 3.11 Column Info Window



Column Name Enter or change the column name. No two columns can have the same column name.

Data Type Select one of the following data types:

– **Numeric** specifies the column values as numbers.

– **Character** specifies the column values as non-numeric, such as letters or symbols.

– **Row State** specifies the column values as row states. This is an advanced topic. See The Column Info Window chapter in the *Using JMP* book.

Modeling Type Modeling types define how values are used in analyses. Select one of the following modeling types:

– **Continuous** values are numeric only.

– **Ordinal** values are either numeric or character, and are ordered categories.

– **Nominal** values are either numeric or character, but not ordered.

Format Select a format for numeric values. This option is not available for character data. Here are a few of the most common formats:

– **Best** lets JMP choose the best display format.

– **Fixed Dec** specifies the number of decimal places that appear.

– **Date** specifies the syntax for date values.

– **Time** specifies the syntax for time values.

– **Currency** specifies the type of currency and decimal points that are used for currency values.

Column Properties Set special column properties such as formulas, notes, and value orders. See The Column Info Window chapter in the *Using JMP* book.

Lock Lock a column, so that the values in the column cannot be changed.v

Calculate Values with Formulas

Use the Formula Editor to create columns that contain calculated values.

Scenario

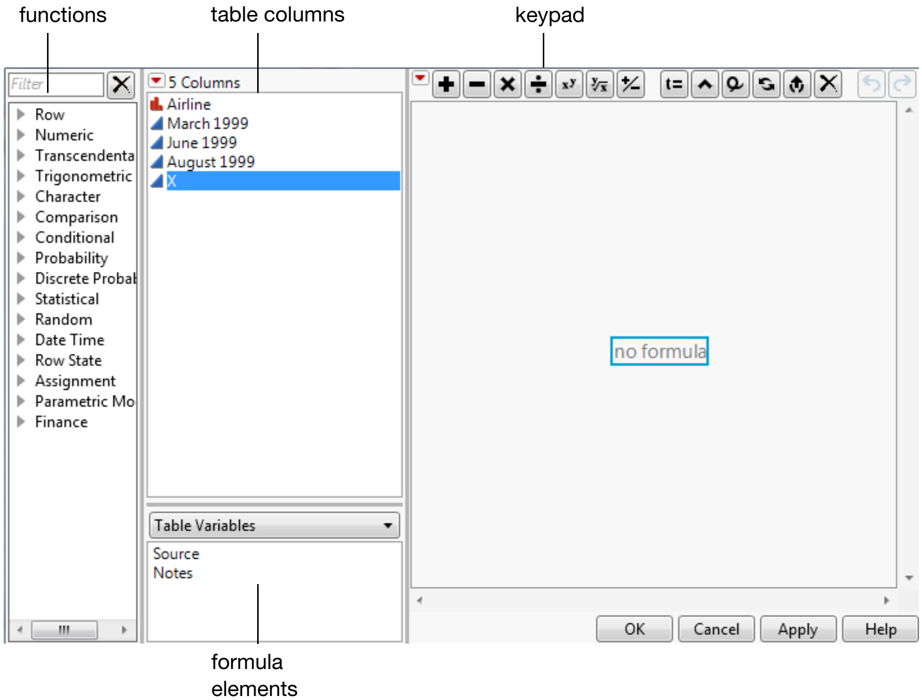
The sample data table On-Time Arrivals.jmp reflects the percent of on-time arrivals for several airlines. The data was collected for March, June, and August of 1999.

Create the Formula

Suppose that you want to create a new column containing the average on-time percentage for each airline.

1. Add a new column.
2. Right-click the column heading of the new column and select **Formula**. The Formula Editor window appears.

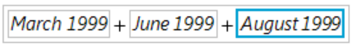
Figure 3.12 Formula Editor



Create the formula for the average on-time percentage of each airline:

3. From the Columns list, select March 1999.
4. Click the button on the keypad.
5. Select June 1999, followed by another sign.
6. Select August 1999.

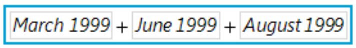
Figure 3.13 Sum of the Months



Notice that only August 1999 is selected (has the red box around it).

7. Click on the box surrounding the entire formula.

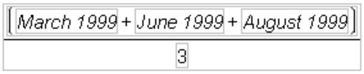
Figure 3.14 Entire Formula Selected



8. Click the button.

9. Type a 3 in the denominator box, and then click outside of the formula in any of the white space.

Figure 3.15 Completed Formula



10. Click **OK**

    The new column contains the averages.

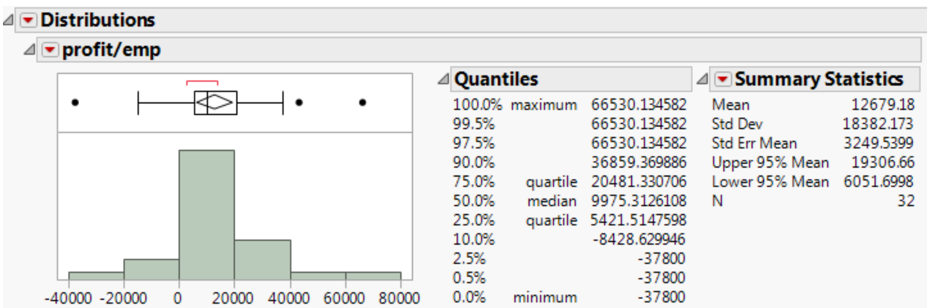
The Formula Editor has many built-in arithmetic and statistical functions. For example, another way to calculate the average on-time arrival percentage is to use the Mean function in the Statistical functions list. For details about all of the Formula Editor functions, see the Formula Editor chapter in the *Using JMP* book.

Filter Data

Use the **Data Filter** to interactively select complex subsets of data, hide these subsets in plots, or exclude them from analyses. For example, look at profit per employee for computer and pharmaceutical companies.

1. Select **Help > Sample Data Library** and open Companies.jmp.
2. Select **Analyze** > **Distribution**.
3. Select profit/emp and click **Y, Columns**.
4. Click **OK**.
5. From the red triangle menu for profit/emp, select **Display Options > Horizontal Layout**.

Figure 3.16 Distribution of profit/emp



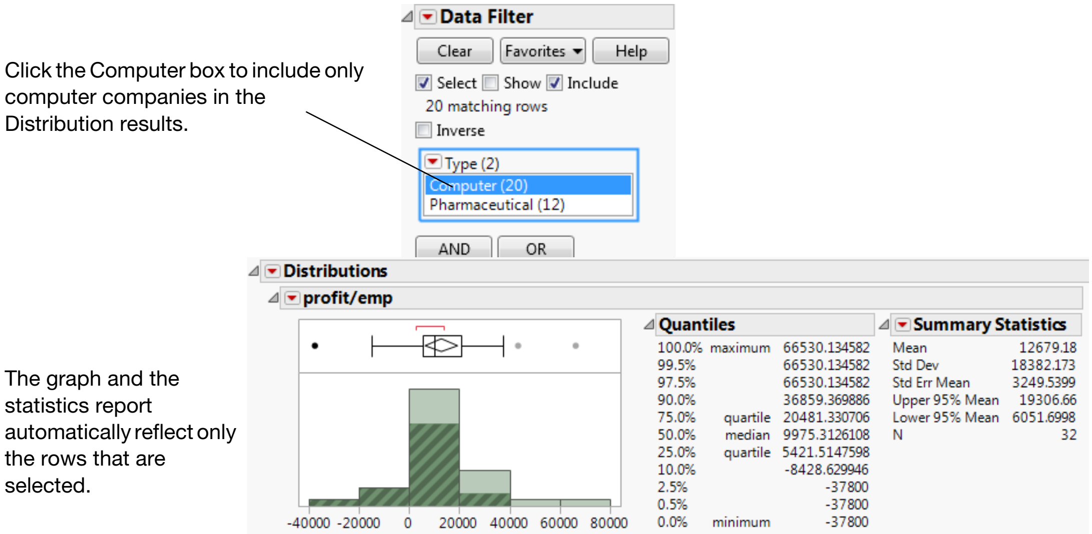
6. Turn on Automatic Recalc by selecting **Redo > Automatic Recalc** from the red triangle menu for **Distributions**.

When this option is on, every change that you make (for example, hiding or excluding points) causes your report window to automatically update itself.

7. In the data table, select **Rows** > **Data Filter**.
8. Select Type and click **Add**.
9. Make sure that Select and Include are selected.
10. To filter out the Pharmaceutical companies from the Distribution results, and include only the Computer companies, click the **Computer** box in the Data Filter window.

    The distribution results update to only include Computer companies.

Figure 3.17 Filter for Computer Companies



Conversely, to change the Distribution results to include only the Pharmaceutical companies, click the **Pharmaceutical** button on the Data Filter window.

Manage Data

The commands on the **Tables** menu (and Tabulate on the **Analyze** menu) summarize and manipulate data tables into the format that you need for graphing and analyzing. This section describes five of these commands:

Summary Creates a table that contains summary statistics that describe your data.

Tabulate Provides a drag and drop workspace to create summary statistics.

Subset Creates a table that contains a subset of your data.

Join Joins the data from two data tables into one new data table.

Sort Sorts your data by one or more columns.

For complete details about these and the other Tables menu commands, see the Reshape Data chapter in the *Using JMP* book.

View Summary Statistics

Summary statistics, such as sums and means, can instantly provide useful information about your data. For example, if you look at the annual profit of each company out of thirty-two companies, it's difficult to compare the profits of small, medium, and large companies. A summary shows that information immediately.

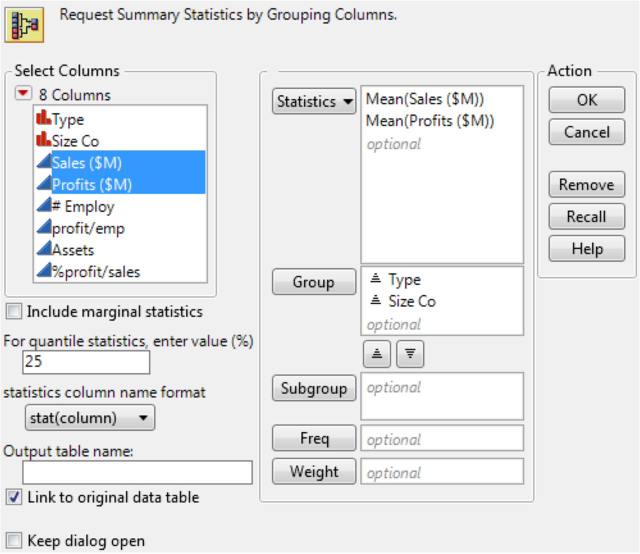
Create summary tables by using either the **Summary** or **Tabulate** commands. The **Summary** command creates a new data table. As with any data table, you can perform analyses and create graphs from the summary table. The **Tabulate** command creates a report window with a table of summary data. You can also create a table from the Tabulate report.

Summary

A summary table contains statistics for each level of a grouping variable. For example, look at the financial data for computer and pharmaceutical companies. Suppose that you want to calculate the mean of sales and the mean of profits, for each combination of company type and size.

1. Select **Help > Sample Data Library** and open Companies.jmp.
2. Select **Tables > Summary**.
3. Select Type and Size Co and click **Group**.
4. Select Sales ($M) and Profits ($M) and click **Statistics > Mean**.

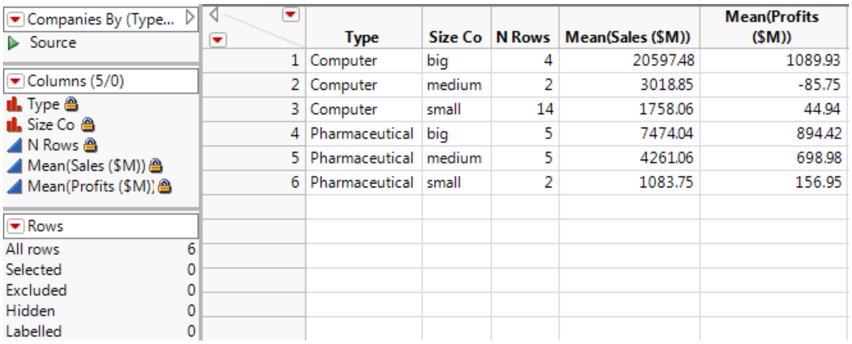
Figure 3.18 Completed Summary Window



5. Click **OK**.

JMP calculates the mean of Sales ($M) and the mean of Profit ($M) for each combination of Type and Size Co.

Figure 3.19 Summary Table



	Type	Size Co	N Rows	Mean(Sales ($M))	Mean(Profits ($M))
1	Computer	big	4	20597.48	1089.93
2	Computer	medium	2	3018.85	-85.75
3	Computer	small	14	1758.06	44.94
4	Pharmaceutical	big	5	7474.04	894.42
5	Pharmaceutical	medium	5	4261.06	698.98
6	Pharmaceutical	small	2	1083.75	156.95

The summary table contains the following:

- There are columns for each grouping variable (in this example, Type, and Size Co).
- The N Rows column shows the number of rows from the original table that correspond to each combination of grouping variables. For example, the original data table contains 14 rows corresponding to small computer companies.

- There is a column for each summary statistic requested. In this example, there is a column for the mean of Sales ($M) and a column for the mean of Profits ($M).

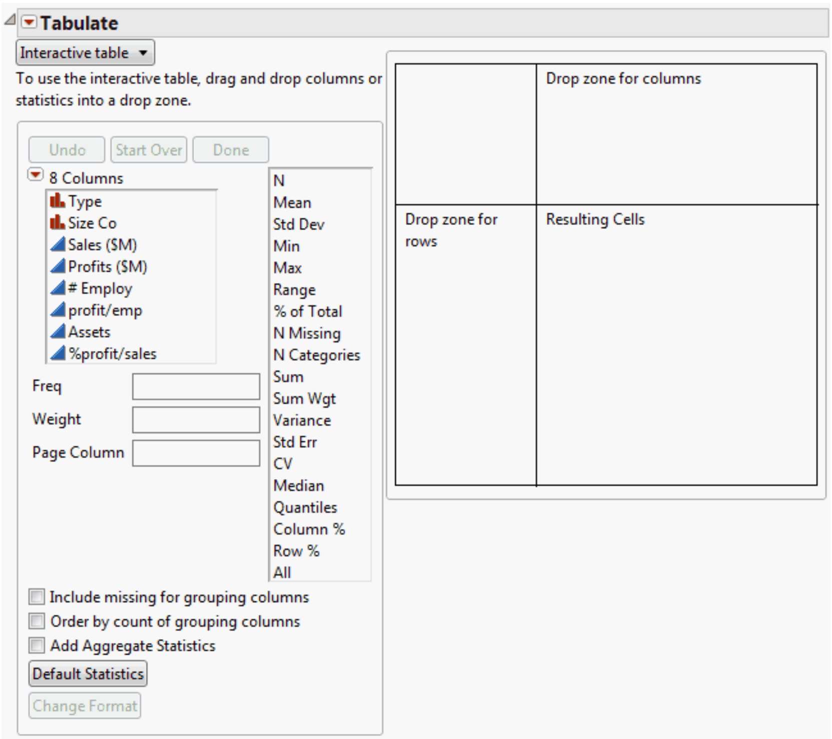
The summary table is linked to the source table. Selecting a row in the summary table also selects the corresponding rows in the source table.

Tabulate

Use the Tabulate command to drag columns into a workspace, creating summary statistics for each combination of grouping variables. This example shows you how to use Tabulate to create the same summary information that you just created using Summary.

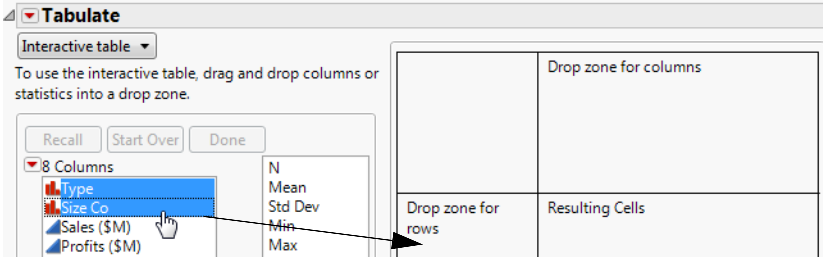
1. Select **Help > Sample Data Library** and open Companies.jmp.
2. Select **Analyze** > **Tabulate**.

Figure 3.20 Tabulate Workspace



3. Select both Type and Size Co.
4. Drag and drop them into the **Drop zone for rows**.

Figure 3.21 Dragging Columns to the Row Zone



5. Right-click a heading and select **Nest Grouping Columns**.

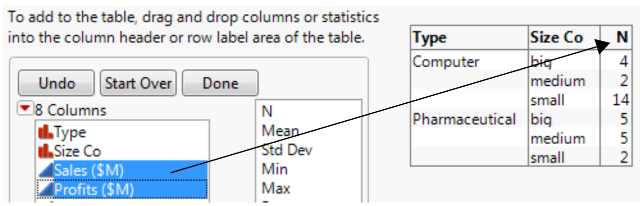
   The initial tabulation shows the number of rows per group.

Figure 3.22 Initial Tabulation

Type	Size Co	N
Computer	big	4
	medium	2
	small	14
Pharmaceutical	big	5
	medium	5
	small	2

6. Select both Sales ($M) and Profits ($M), and drag and drop them over the **N** in the table.

Figure 3.23 Adding Sales and Profit



The tabulation now shows the sum of Sales ($M) and the sum of Profits ($M) per group.

Figure 3.24 Tabulation of Sums

		Sales ($M)	Profits ($M)
Type	Size Co	Sum	Sum
Computer	big	82389.9	4359.7
	medium	6037.7	-171.5
	small	24612.8	629.1
Pharmaceutical	big	37370.2	4472.1
	medium	21305.3	3494.9
	small	2167.5	313.9

7. The final step is to change the sums to means. Right-click **Sum** (either of them) and select **Statistics** > **Mean**.

Figure 3.25 Final Tabulation

		Sales ($M)	Profits ($M)
Type	**Size Co**	**Mean**	**Mean**
Computer	big	20597.48	1089.9
	medium	3018.85	-85.75
	small	1758.06	44.94
Pharmaceutical	big	7474.04	894.42
	medium	4261.06	698.98
	small	1083.75	156.95

The means are the same as those obtained using the Summary command. Compare Figure 3.25 to Figure 3.19.

Create Subsets

If you want to look closely at only part of your data table, you can create a subset. For example, suppose that you have already compared the sales and profits of big, medium, and small computer and pharmaceutical companies. Now you want to look at the sales and profits of only the medium-sized companies.

Creating a subset is a two-step process. First select the target data, and then extract the data into a new table.

Subset with the Subset Command

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

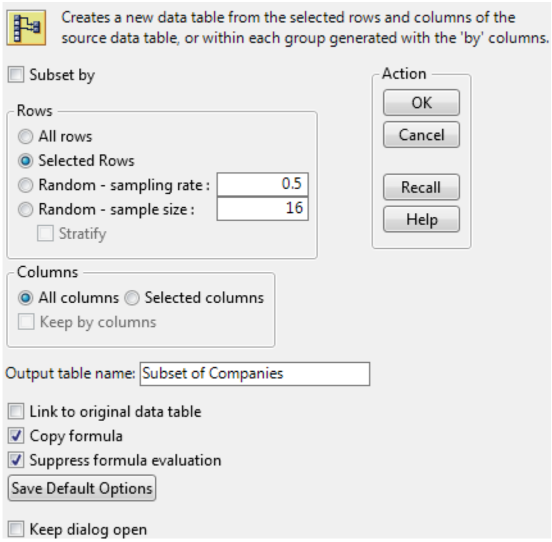
Selecting the Rows and Columns That You Want to Subset

2. Select **Rows > Row Selection > Select Where**.
3. Select Size Co in the column list box on the left.
4. Enter medium in the text enter box.
5. Click **OK**.
6. Hold down the Ctrl key and select the Type, Sales ($M), and Profits ($M) columns.

Creating the Subset Table

7. Select **Tables > Subset** to launch the Subset window.

Figure 3.26 Subset Window



8. Select **Selected columns** to subset only the columns that you selected. You can also customize your subset table further by selecting additional options.
9. Click **OK**.

The resulting subset data table has seven rows and three columns. For complete details about the Subset command, see the Reshape Data chapter in the *Using JMP* book.

Subset with the Distribution Platform

Another way to create subsets uses the connection between platform results and data tables.

Example of Creating a Subset Using the Distribution Command

1. Select **Help > Sample Data Library** and open Companies.jmp.
2. Select **Analyze > Distribution**.
3. Select Type and click **Y, Columns**.
4. Click **OK**.
5. Double-click on the histogram bar that represents Computer to create a subset table of the Computer companies.

Caution: This method creates a *linked* subset table. This means if you make any changes to the data in the subset table, the corresponding value changes in the source table.

Join Data Tables

Use the Join option to combine information from multiple data tables into a single data table. For example, suppose that you have a data table containing results from an experiment on popcorn yields. In another data table, you have the results of a second experiment on popcorn yields. To compare the two experiments or to analyze the trials using both sets of results, you need to have the data in the same table. Also, the experimental data was not entered into the data tables in the same order. One of the columns has a different name, and the second experiment is incomplete. This means that you cannot copy and paste from one table into another.

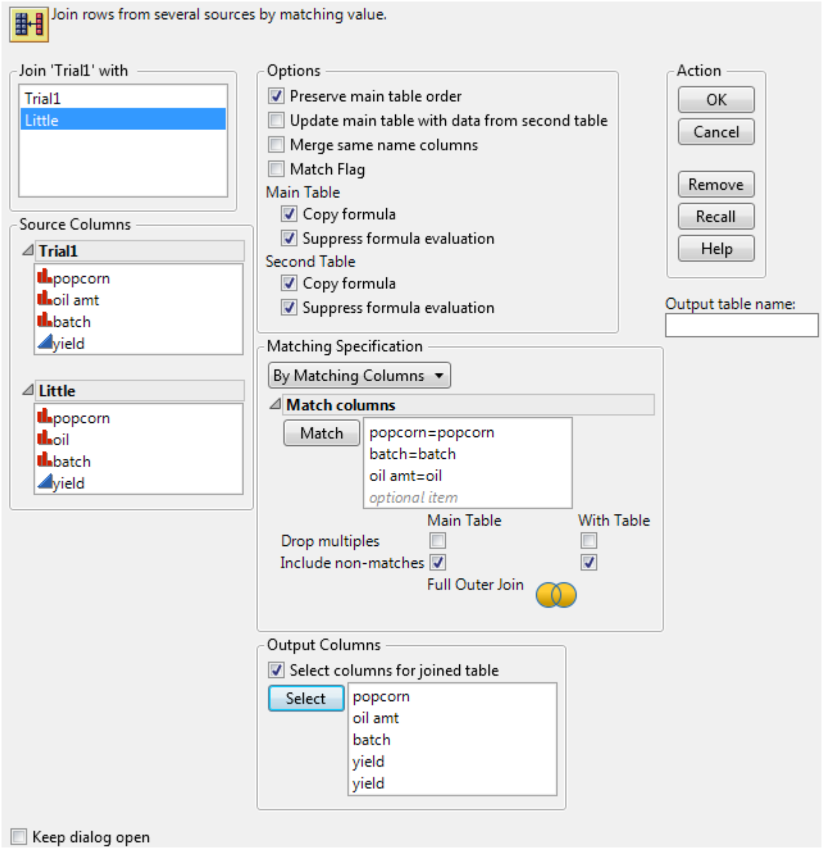
Example of Joining Two Data Tables

1. Select **Help > Sample Data Library** and open Trial1.jmp and Little.jmp.
2. Click on Trial1.jmp to make it the active data table.
3. Select **Tables > Join**.
4. In the **Join 'Trial1' With** box, select Little.
5. From the **Matching Specification** menu, select **By Matching Columns**.
6. In the **Source Columns** boxes, select popcorn in both boxes, and then click **Match**.
7. In the same way, match batch and oil amt to oil in both boxes.

   Your matching columns do not have to have the same name.
8. Select **Include non-matches** for both tables.

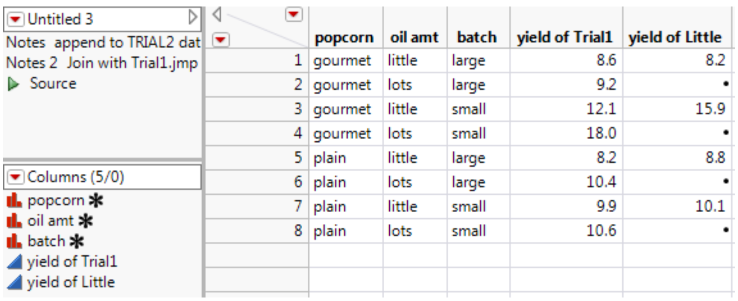
   Since one experiment is partial, you want to include all rows, including any with missing data.
9. To avoid duplicate columns, select the **Select columns for joined table** option.
10. From Trial1, select all four columns and click **Select**.
11. From Little, select only yield and click **Select**.

Figure 3.27 Completed Join Window



12. Click **OK**.

Figure 3.28 Joined Table



	popcorn	oil amt	batch	yield of Trial1	yield of Little
1	gourmet	little	large	8.6	8.2
2	gourmet	lots	large	9.2	•
3	gourmet	little	small	12.1	15.9
4	gourmet	lots	small	18.0	•
5	plain	little	large	8.2	8.8
6	plain	lots	large	10.4	•
7	plain	little	small	9.9	10.1
8	plain	lots	small	10.6	•

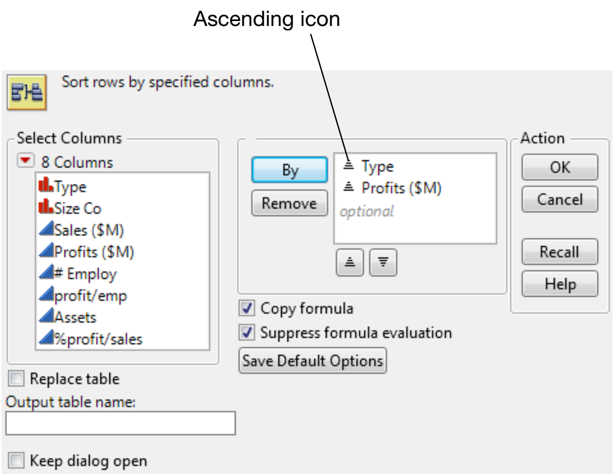
Sort Tables

Use the Sort command to sort a data table by one or more columns in the data table. For example, look at financial data for computer and pharmaceutical companies. Suppose that you want to sort the data table by Type, then by Profits ($M). Also, you want Profits ($M) to be in descending order within each Type.

1. Select **Help > Sample Data Library** and open Companies.jmp.
2. Select **Tables** > **Sort**.
3. Select Type and click **By** to assign Type as a sorting variable.
4. Select Profits ($M) and click **By**.

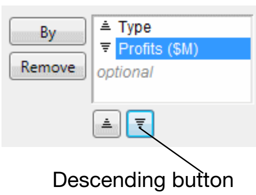
At this point, both variables are set to be sorted in ascending order. See the ascending icon next to the variables in Figure 3.29.

Figure 3.29 Sort Ascending Icon



5. To change Profits ($M) to sort in descending order, select Profits ($M) and click the descending button.

Figure 3.30 Change Profits to Descending



The icon next to Profits ($M) changes to descending.

6. Select the **Replace Table** check box.

   When selected, the **Replace Table** option tells JMP to sort the original data table instead of creating a new table with the sorted values. This option is not available if there are any open report windows created from the original data table. Sorting a data table with open report windows might change how some of the data is displayed in the report window, especially in graphs.

7. Click **OK**.

The data table is now sorted by type alphabetically, and by descending profit totals within type.

Chapter 4

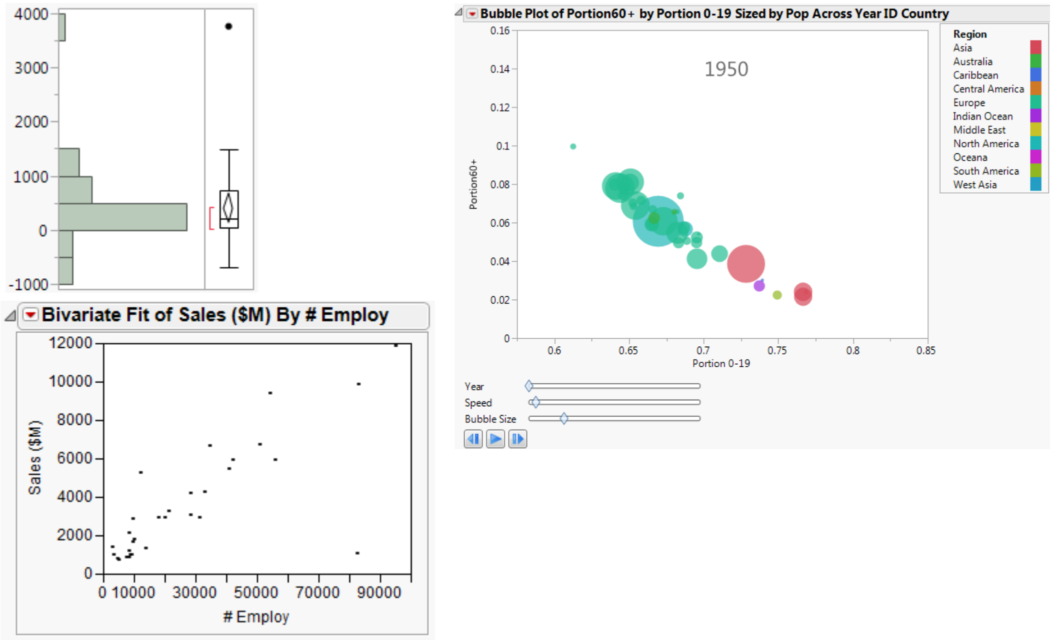
Visualize Your Data

Common Graphs

Visualizing your data is an important first step. The graphs described in this chapter help you discover important details about your data. For example, histograms show you the shape and range of your data, and help you find unusual data points.

This chapter presents several of the most common graphs and plots that enable you to visualize and explore data in JMP. This chapter is an introduction to some of JMP's graphical tools and platforms. Use JMP to visualize the distribution of single variables, or the relationships among multiple variables.

Figure 4.1 Visualizing Data with JMP



Analyze Single Variables

Single-variable graphs, or *univariate* graphs, let you look closely at one variable at a time. When you begin to look at your data, it's important to learn about each variable before looking at how the variables interact with each other. Univariate graphs let you visualize each variable individually.

This section covers two graphs that show the distribution of a single variable:

- "Histograms" on page 90, for continuous variables
- "Bar Charts" on page 92, for categorical variables

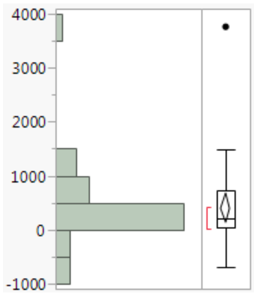
Use the Distribution platform to create both of these graphs. Distribution produces a graphical description and descriptive statistics for each variable.

Histograms

The histogram is one of the most useful graphical tools for understanding the distribution of a continuous variable. Use a histogram to find the following in your data:

- the average value and variation
- extreme values

Figure 4.2 Example of a Histogram



Scenario

This example uses the Companies.jmp data table, which contains data on profits for a group of companies.

A financial analyst wants to explore the following questions:

- Generally, how much profit does each company earn?
- What is the average profit?

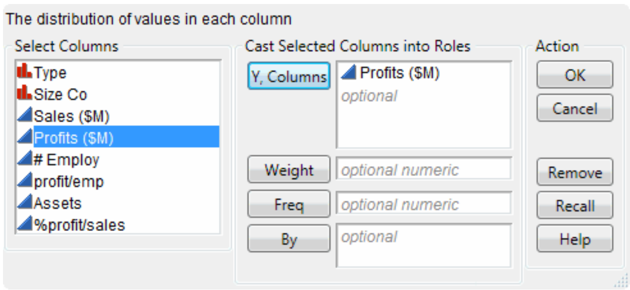
- Are there any companies that earn either extremely high or extremely low profits compared to the other companies?

To answer these questions, use a histogram of Profits ($M).

Create the Histogram

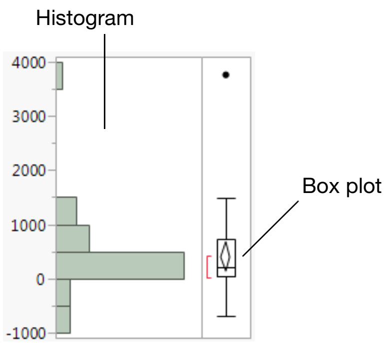
1. Select **Help > Sample Data Library** and open Companies.jmp.
2. Select **Analyze** > **Distribution**.
3. Select Profits ($M) and click **Y, Columns**.

Figure 4.3 Distribution Window for Profits ($M)



4. Click **OK**.

Figure 4.4 Histogram of Profits ($M)



Interpret the Histogram

The histogram provides these answers:

- Most companies' profits are between $-1000 and $1500.

  All the bars except for one are located in this range. Also, more companies' profits range from $0 to $500 than any other range. The bar representing that range is much longer than the others.

- The average profit is a little less than $500.

  The middle of the diamond in the box plot indicates the mean value. In this case, the mean is slightly lower than the $500 mark.
- One company has significantly higher profits than the others, and might be an *outlier*. An outlier is a data point that is separated from the general pattern of the other data points.

  This outlier is represented by a single, very short bar at the top of the histogram. The bar is small and represents a small group (in this case, a single company), and it is widely separated from the rest of the histogram bars.

In addition to the histogram, this report includes the following:

- The box plot, which is another graphical summary of the data. For detailed information about the box plot, see the Graph Builder chapter in the *Essential Graphing* book.
- **Quantiles** and **Summary Statistics** reports. These reports are discussed in "Analyze Distributions" on page 130 in the "Analyze Your Data" chapter.

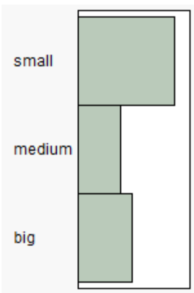
Interact with the Histogram

Data tables and reports are all connected in JMP. Click on a histogram bar to select the corresponding rows in the data table.

Bar Charts

Use a bar chart to visualize the distribution of a categorical variable. A bar chart looks similar to a histogram, since they both have bars that correspond to the levels of a variable. A bar chart shows a bar for every level of the variable, whereas the histogram shows a range of values for the variable.

Figure 4.5 Example of a Bar Chart



Scenario

This example uses the Companies.jmp data table, which contains data on the size and type of a group of companies.

A financial analyst wants to explore the following questions:

- What is the most common type of company?
- What is the most common size for a company?

To answer these questions, use bar charts of Type and Size Co.

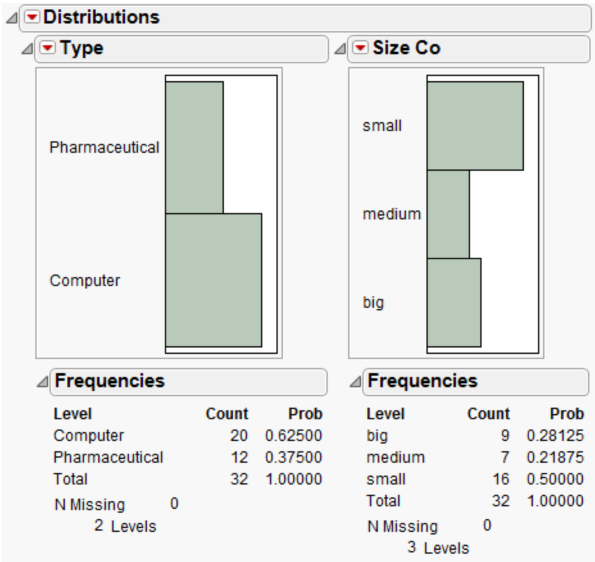
Create the Bar Chart

1. Select **Help > Sample Data Library** and open Companies.jmp.
2. Select **Analyze** > **Distribution**.
3. Select Type and Size Co and click **Y, Columns**.
4. Click **OK**.

Figure 4.6 Bar Charts of Type and Size Co

Bar Charts

Summary Information



Interpret the Bar Charts

The bar charts provide these answers:

- There are more computer companies than pharmaceutical companies.

  The bar that represents computer companies is larger than the bar that represents pharmaceutical companies.

- The most common company size is small.

  The bar that represents small companies is larger than the bars that represent medium and big companies.

The additional summary output gives detailed frequencies. This report is discussed in “Distributions of Categorical Variables” on page 133 in the “Analyze Your Data” chapter.

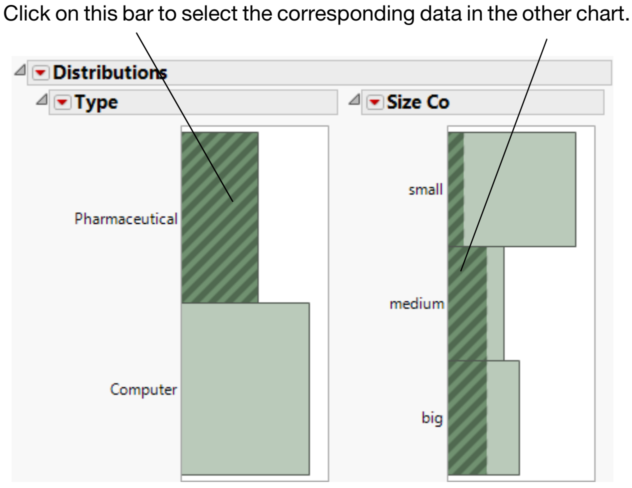
Interact with the Bar Charts

As is the case with histograms, click on individual bars to highlight rows of the data table. If more than one graph is created, clicking on a bar in one bar chart highlights the corresponding bar or bars in the other bar chart.

For example, suppose that you want to see the distribution of company size for the pharmaceutical companies. Click the Pharmaceutical bar in the Type bar chart, and the pharmaceutical companies are highlighted on the Size Co bar chart. Figure 4.7 shows that although most companies in this data table are small, most of the pharmaceutical companies are medium or big.

Also, the corresponding rows in the data table are selected.

Figure 4.7 Clicking Bars



Compare Multiple Variables

Use multiple-variable graphs to visualize the relationships and patterns between two or more variables. This section covers the following graphs:

Table 4.1 Multiple-Variable Graphs

“Scatterplots” on page 95	Use scatterplots to compare two continuous variables.

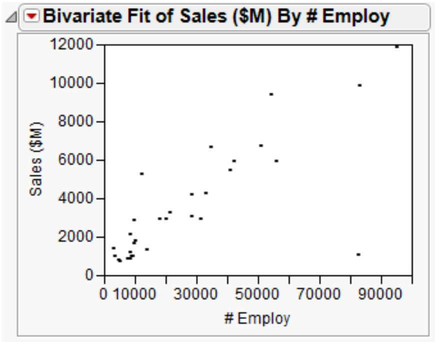
Table 4.1 Multiple-Variable Graphs *(Continued)*

"Scatterplot Matrix" on page 99	Use scatterplot matrices to compare several pairs of continuous variables.
"Side-by-Side Box Plots" on page 102	Use side-by-side box plots to compare one continuous and one categorical variable.
"Overlay Plots" on page 105	Use overlay plots to compare one or more variables on the Y-axis to another variable on the X-axis. Overlay plots are especially useful if the X variable is a time variable, because you can compare how two or more variables change across time.
"Variability Chart" on page 108	Use variability charts to compare one continuous Y variable to one or more categorical X variables. Variability charts show differences in means and variability across several categorical X variables.
"Graph Builder" on page 111	Use Graph Builder to create and change graphs interactively.
"Bubble Plots" on page 117	Bubble plots are specialized scatterplots that use color and bubble sizes to represent up to five variables at once. If one of your variables is a time variable, you can animate the plot to see your other variables change through time.

Scatterplots

The scatterplot is the simplest of all the multiple-variable graphs. Use scatterplots to determine the relationship between two continuous variables and to discover whether two continuous variables are *correlated*. Correlation indicates how closely two variables are related. When you have two variables that are highly correlated, one might influence the other. Or, both might be influenced by other variables in a similar way.

Figure 4.8 Example of a Scatterplot



Scenario

This example uses the Companies.jmp data table, which contains sales figures and the number of employees of a group of companies.

A financial analyst wants to explore the following questions:

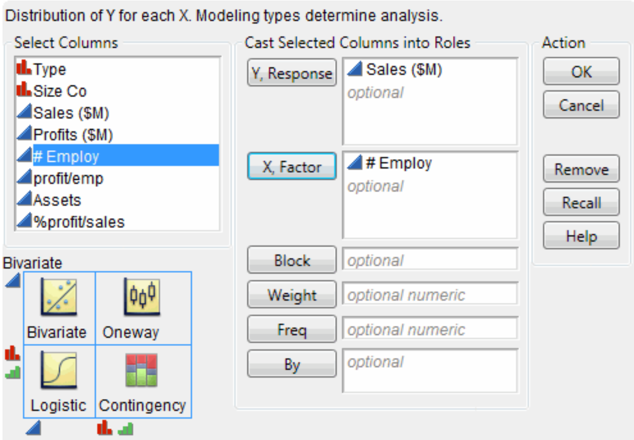
- What is the relationship between sales and the number of employees?
- Does the amount of sales increase with the number of employees?
- Can you predict average sales from the number of employees?

To answer these questions, use a scatterplot of Sales ($M) versus # Employ.

Create the Scatterplot

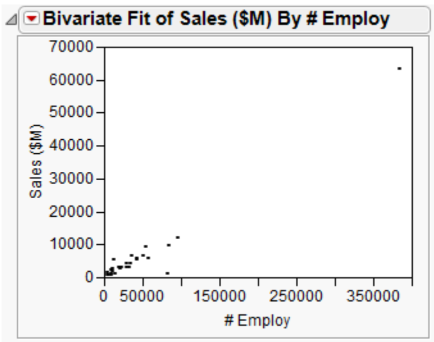
1. Select **Help > Sample Data Library** and open Companies.jmp.
2. Select **Analyze > Fit Y by X**.
3. Select Sales ($M) and **Y, Response**.
4. Select # Employ and **X, Factor**.

Figure 4.9 Fit Y by X Window



5. Click **OK**.

Figure 4.10 Scatterplot of Sales ($M) versus # Employ



Interpret the Scatterplot

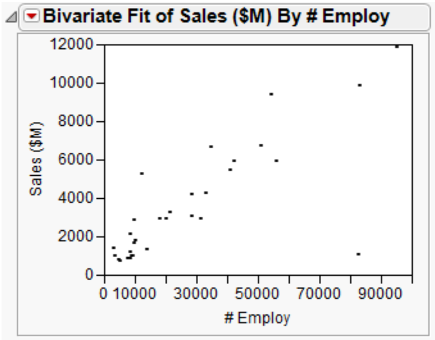
One company has a large number of employees and high sales, represented by the single point at the top right of the plot. The distance between this data point and all the rest makes it difficult to visualize the relationship between the rest of the companies. Remove the point from the plot and re-create the plot by following these steps:

1. Click on the point to select it.
2. Select **Rows > Hide and Exclude**. The data point is hidden and no longer included in calculations.

Note: The difference between hiding and excluding is important. Hiding a point removes it from any graphs but statistical calculations continue to use the point. Excluding a point removes it from any statistical calculations but does not remove it from graphs. When you both hide and exclude a point, you remove it from all calculations and from all graphs.

3. To re-create the plot without the outlier, select **Redo > Redo Analysis** from the red triangle menu for Bivariate. You can close the original report window.

Figure 4.11 Scatterplot with the Outlier Removed



The updated scatterplot provides these answers:

- There is a relationship between the sales and the number of employees.

  The data points have a discernible pattern. They are not scattered randomly throughout the graph. You could draw a diagonal line that would be near most of the data points.

- Sales do increase with the number of employees, and the relationship is linear.

  If you drew that diagonal line, it would slope from bottom left to top right. This slope shows that as the number of employees increases (left to right on the bottom axis), sales also increases (bottom to top on the left axis). A straight line would be near most of the data points, indicating a linear relationship. If you would have to curve your line to be near the data points, there would still be a relationship (because of the pattern of the points). However, that relationship would not be linear.

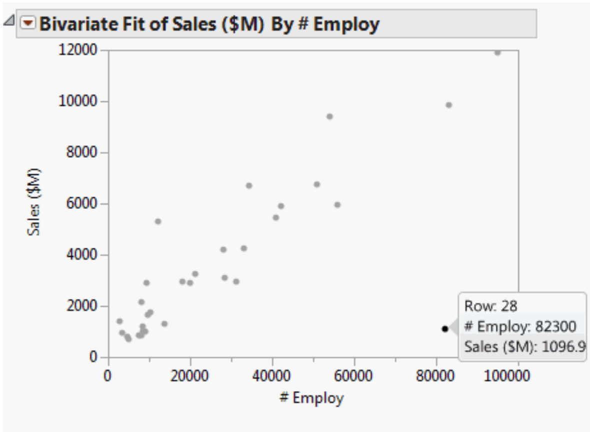
- You can predict average sales from the number of employees.

  The scatterplot shows that sales generally increase as the number of employees does. You could predict the sales for a company if you knew only the number of employees of that company. Your prediction would be on that imaginary line. It would not be exact, but it would approximate the real sales.

Interact with the Scatterplot

As with other JMP graphics, the scatterplot is interactive. Place your mouse pointer over the point in the bottom right corner with the mouse to reveal the row number and the x and y values.

Figure 4.12 Place Your Mouse Pointer Over a Point



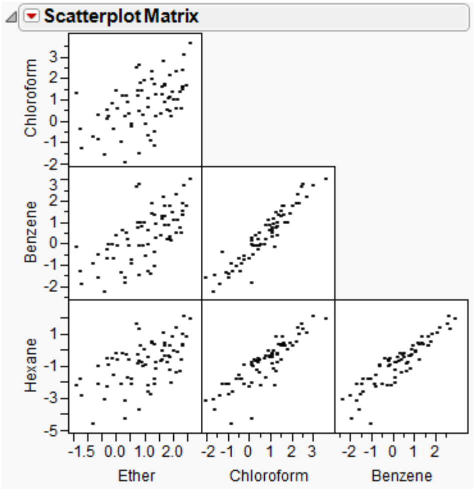
Click on a point to highlight the corresponding row in the data table. Select multiple points by doing one of the following:

- Click and drag with the mouse around the points. This selects points in a rectangular area.
- Select the lasso tool, and then click and drag around multiple points. The lasso tool selects an irregularly shaped area.

Scatterplot Matrix

A scatterplot matrix is a collection of scatterplots organized into a grid (or matrix). Each scatterplot shows the relationship between a pair of variables.

Figure 4.13 Example of a Scatterplot Matrix



Scenario

This example uses the Solubility.jmp data table, which contains data for solubility measurements for 72 different solutes.

A lab technician wants to explore the following questions:

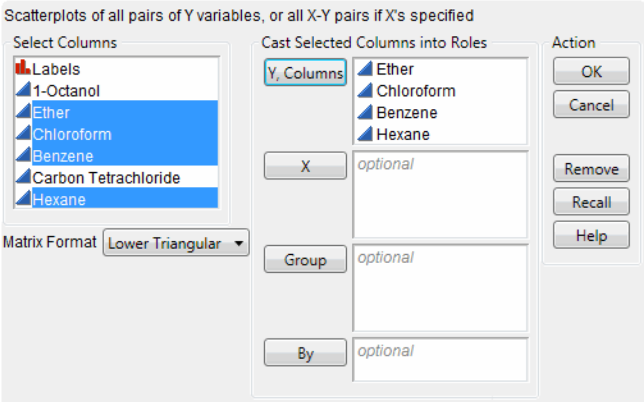
- Is there a relationship between any pair of chemicals? (There are six possible pairs.)
- Which pair has the strongest relationship?

To answer these questions, use a scatterplot matrix of the four solvents.

Create the Scatterplot Matrix

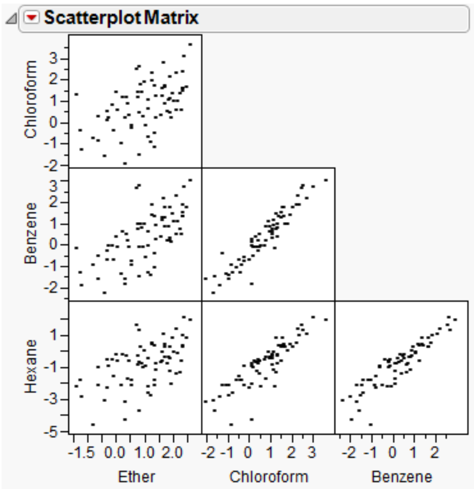
1. Select **Help > Sample Data Library** and open Solubility.jmp.
2. Select **Graph** > **Scatterplot Matrix**.
3. Select Ether, Chloroform, Benzene, and Hexane, and click **Y, Columns**.

Figure 4.14 Scatterplot Matrix Window



4. Click **OK**.

Figure 4.15 Scatterplot Matrix



Interpret the Scatterplot Matrix

The scatterplot matrix provides these answers:

- All six pairs of variables are positively correlated.

  As one variable increases, the other variable increases too.

- The strongest relationship appears to be between Benzene and Chloroform.

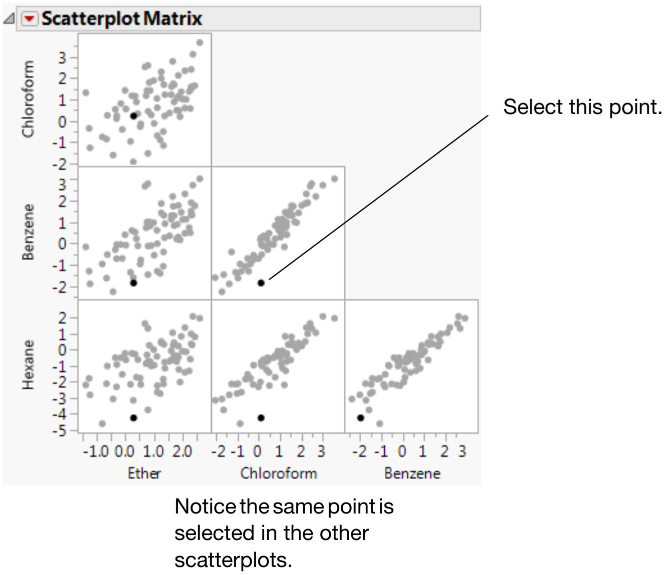
  The data points in the scatterplot for Benzene and Chloroform are the most tightly clustered along an imaginary line.

Interact with the Scatterplot Matrix

If you select a point in one scatterplot, it is selected in all the other scatterplots.

For example, if you select a point in the Benzene versus Chloroform scatterplot, the same point is selected in the other five plots.

Figure 4.16 Selected Points

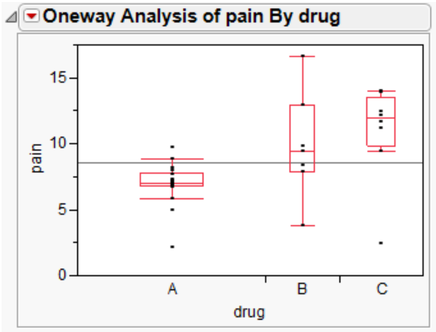


Side-by-Side Box Plots

Side-by-side box plots show the following:

- the relationship between one continuous variable and one categorical variable
- differences in the continuous variable across levels of the categorical variable

Figure 4.17 Example of Side-by-Side Box Plots



Scenario

This example uses the Analgesics.jmp data table, which contains data on pain measurements taken on patients using three different drugs.

A researcher wants to explore the following questions:

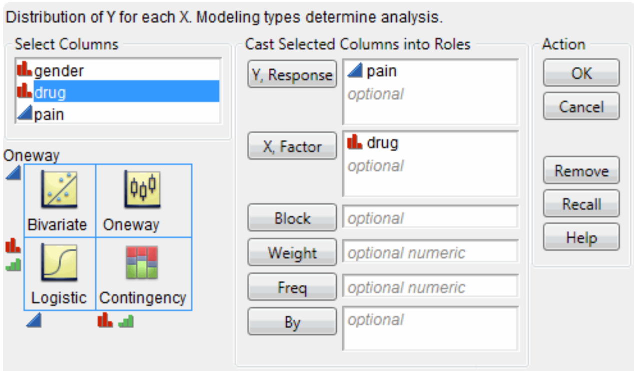
- Are there differences in the average amount of pain control among the drugs?
- Does the *variability* in the pain control given by each drug differ? A drug with high variability would not be as reliable as a drug with low variability.

To answer these questions, use a side-by-side box plot for the pain levels and the drug categories.

Create the Side-by-Side Box Plots

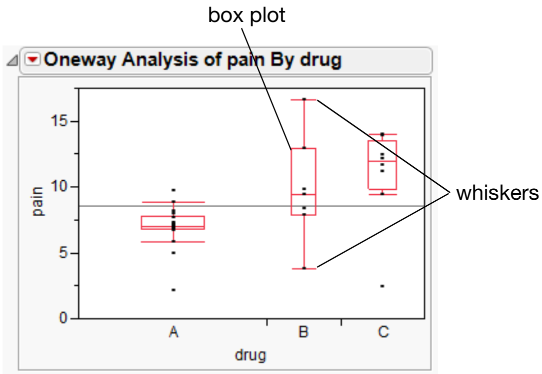
1. Select **Help > Sample Data Library** and open Analgesics.jmp.
2. Select **Analyze** > **Fit Y by X**.
3. Select pain and click **Y, Response**.
4. Select drug and click **X, Factor**.

Figure 4.18 Fit Y by X Window



5. Click **OK**.
6. From the red triangle menu, select **Display Options > Box Plots**.

Figure 4.19 Side-by-Side Box Plots



Interpret the Side-by-Side Box Plots

Box plots are designed according to the following principles:

- The line through the box represents the median.
- The middle half of the data is within the box.
- The majority of the data falls between the ends of the whiskers.
- A data point outside the whiskers might be an outlier.

The box plots in Figure 4.19 show these answers:

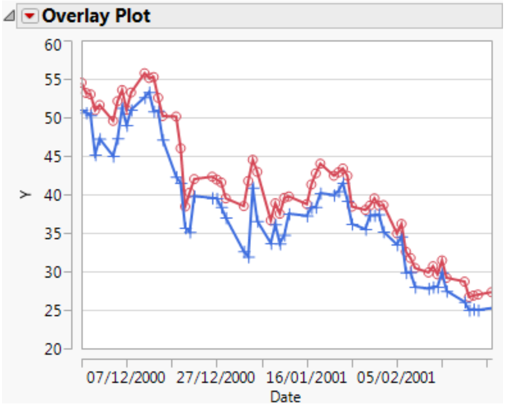
- There is evidence to believe that patients on drug A feel less pain, since the box plot for drug A is lower on the pain scale than the others.
- Drug B appears to have higher variability than Drugs A and C, since the box plot is taller.

There is one point for drug C that is a lot lower than the other points for drug C. Place your mouse pointer over it with your mouse to see that it is row 26 of the data table. That point looks like it is more similar to the data in drug group A or B. The information in row 26 deserves investigation. There might have been a typographical error when the data was recorded.

Overlay Plots

Like scatterplots, overlay plots show the relationship between two or more variables. However, if one of the variables is a time variable, an overlay plot shows trends across time better than scatterplots do.

Figure 4.20 Example of an Overlay Plot



Note: To plot data over time, you can also use Graph Builder, bubble plots, control charts, and variability charts. For complete details about Graph Builder and bubble plots, see the Graph Builder chapter in the *Essential Graphing* book. Refer to the Control Chart Builder chapter and the Variability Gauge Charts chapter in the *Quality and Process Methods* book for information about control charts and variability charts.

Scenario

This example uses the Stock Prices.jmp data table, which contains data on the price of a stock over a three-month period.

A potential investor wants to explore the following questions:

- Has the stock's closing price changed over the past three months?

  To answer this question, use an overlay plot of the stock's closing price over time.

- How do the stock's high and low prices relate to each other?

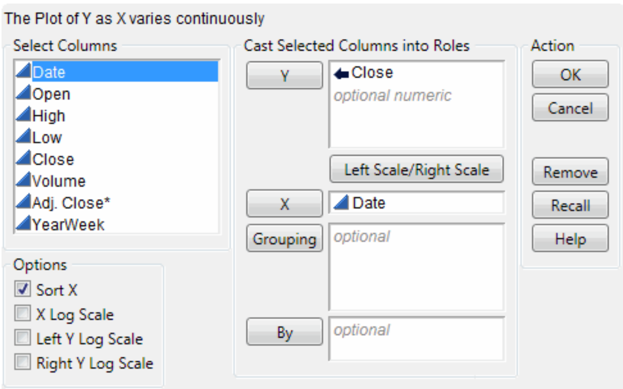
To answer this question, use another overlay plot of the stock's high and low prices over time.

Create the first overlay plot to answer the first question, and then create a second overlay plot to answer the second question.

Create the Overlay Plot of the Stock's Price over Time

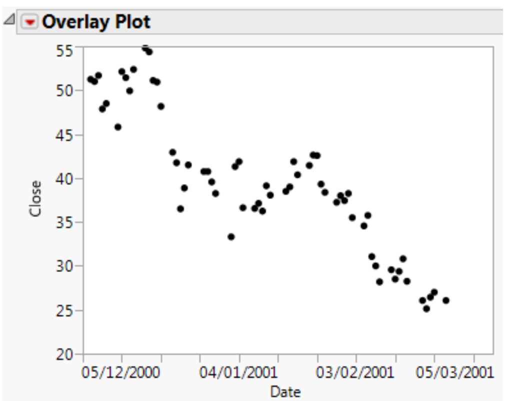
1. Select **Help > Sample Data Library** and open Stock Prices.jmp.
2. Select **Graph** > **Overlay Plot**.
3. Select Close and click **Y**.
4. Select Date and click **X**.

Figure 4.21 Overlay Plot Window



5. Click **OK**.

Figure 4.22 Overlay Plot of the Closing Price over Time

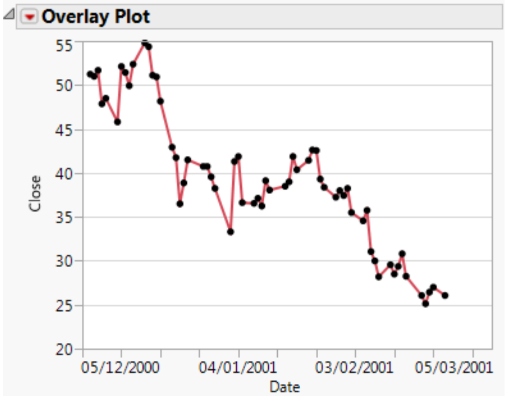


Interpret and Interact with the Overlay Plot

The overlay plot shows that the closing stock price has been decreasing over the last several months. To see the trend more clearly, connect the points and add grid lines.

1. From the red triangle menu, select **Connect Thru Missing**.
2. Double-click the Y axis.
3. Select the **Major Grid Lines** check box.
4. Click **OK**.

Figure 4.23 Connected Points and Grid Lines



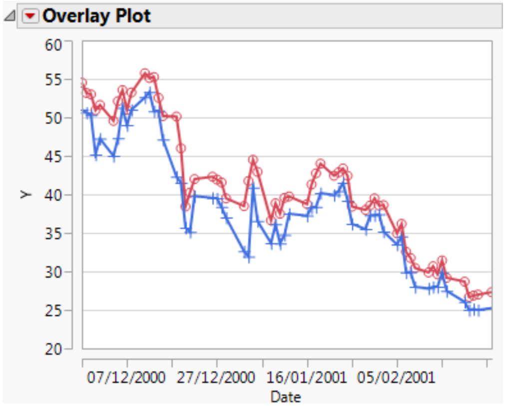
The potential investor can see that although the stock price has gone up and down over the past three months, the overall trend has been downward.

Create the Overlay Plot of the Stock's High and Low Prices

Use an overlay plot to plot more than one Y variable. For example, suppose that you want to see both the high and the low prices on the same plot.

1. Follow the steps in "Create the Overlay Plot of the Stock's Price over Time" on page 106, this time assigning both High and Low to the **Y** role.
2. Connect the points and add grid lines as shown in "Interpret and Interact with the Overlay Plot" on page 107.

Figure 4.24 Two Y Variables



The legend at the bottom of the plot shows the colors and markers used for the High and Low variables in the graph. The overlay plot shows that the High price and Low price track each other very closely.

Answer the Questions

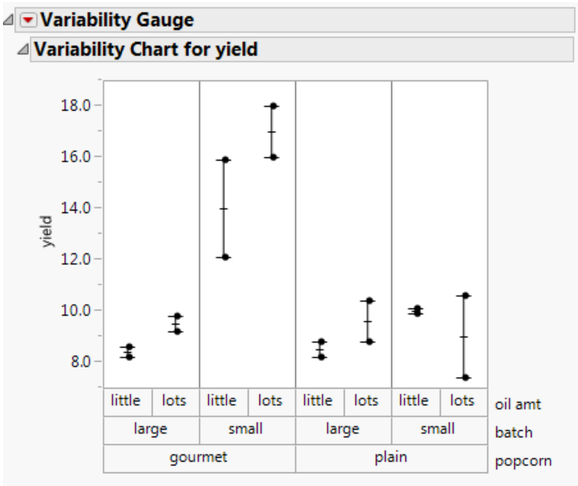
Both of the overlay plots answer the two questions asked at the beginning of this example.

- The first plot shows that the price of this stock has not remained the same, but has been decreasing.
- The second plot shows that the high and low prices of this stock are not very different from each other. The stock price does not vary wildly on any given day.

Variability Chart

In the graphs described so far, you specified only a single X variable. Use a variability chart to specify multiple X variables and see differences in means and variability across all of your variables at once.

Figure 4.25 Example of a Variability Chart



Scenario

This example uses the Popcorn.jmp data table with data from a popcorn maker. The yield (the volume of popcorn for a given measure of kernels) was measured for each combination of popcorn style, batch size, and amount of oil used.

The popcorn maker wants to explore the following question:

- Which combination of factors results in the highest popcorn yield?

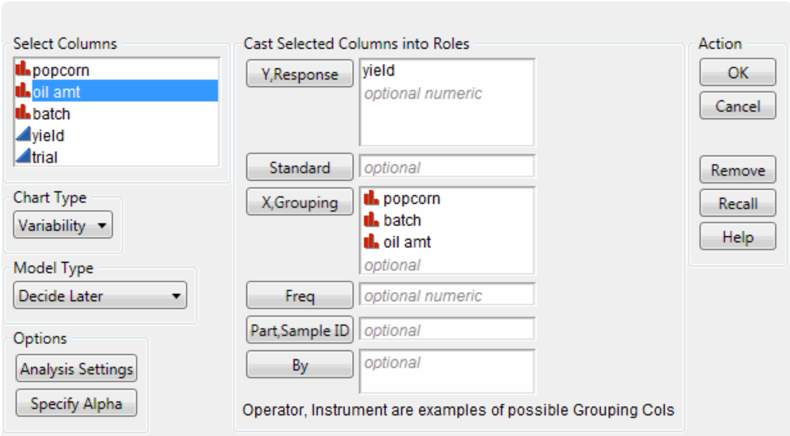
To answer this question, use a variability chart of the yield versus the style, batch size, and oil amount.

Create the Variability Chart

1. Select **Help > Sample Data Library** and open Popcorn.jmp.
2. Select **Analyze > Quality and Process > Variability/Attribute Gauge Chart**.
3. Select yield and click **Y, Response**.
4. Select popcorn and click **X, Grouping**.
5. Select batch and click **X, Grouping**.
6. Select oil amt and click **X, Grouping**.

Note: The order in which you assign the variables to the **X, Grouping** role is important, because the order in this window determines their nesting order in the variability chart.

Figure 4.26 Variability Chart Window

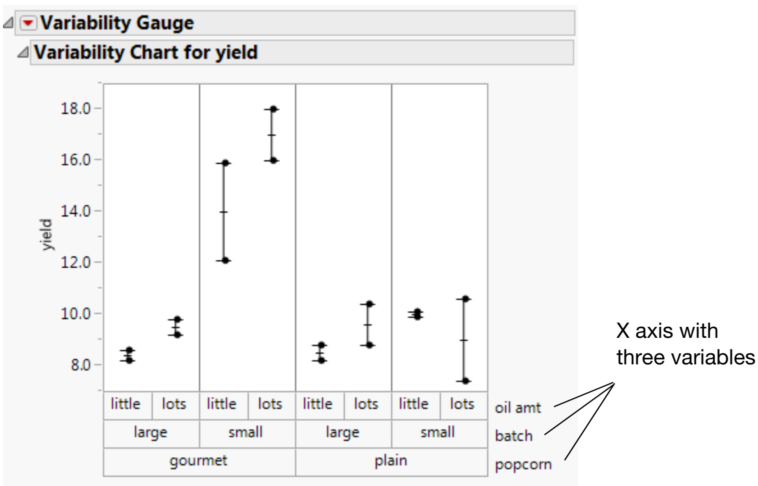


7. Click **OK**.

   The top chart is the variability chart, showing the yield broken down by each combination of the three variables. The bottom chart shows the standard deviation for each combination of the three variables. Since the bottom chart does not show the yield, hide it.

8. Deselect **Std Dev Chart** on the red triangle menu.

Figure 4.27 Results Window



Interpret the Variability Chart

The variability chart for yield indicates that small, gourmet batches produce the highest yield.

To be more specific, the popcorn maker might ask this additional question: Is the yield high because those batches are small, or because those batches are gourmet?

The variability chart shows the following:

- The yield from small, plain batches is low.
- The yield from large, gourmet batches is low.

Given this information, the popcorn maker can conclude that only the combination of small and gourmet at the same time results in batches with high yield. It would have been impossible to reach this conclusion with a chart that only allowed a single variable.

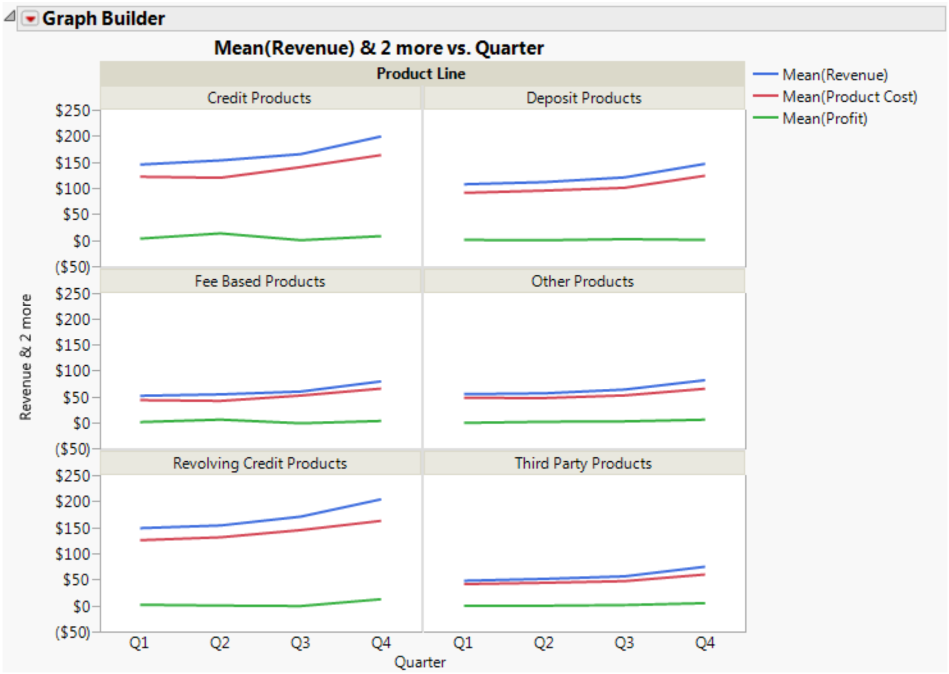
Graph Builder

Use Graph Builder to interactively create and modify graphs. So far, all of the graphs have been created by launching a platform and specifying variables. To create a different type of graph, you must launch a different platform. In Graph Builder, you can change the variables and change the graphs at any time.

Use Graph Builder to accomplish the following tasks:

- Change variables by dragging and dropping them in and out of the graph.
- Create a different type of graph with a few mouse clicks.
- Partition the graph horizontally or vertically.

Figure 4.28 Example of a Graph That Was Created with Graph Builder



Note: Only some of the Graph Builder features are covered here. For complete details, see the Graph Builder chapter in the *Essential Graphing* book.

Scenario

This example uses the Profit by Product.jmp data table, which contains profit data for multiple product lines.

A business analyst wants to explore the following question:

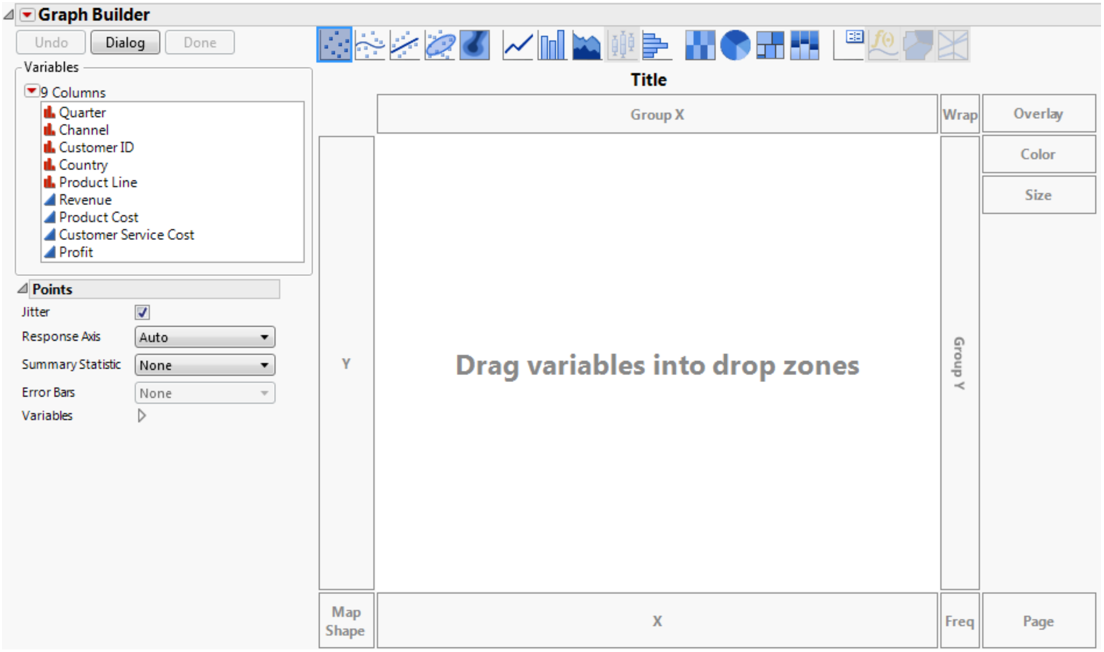
- How is the profitability different between product lines?

To answer this question, use a line plot that displays revenue, product cost, and profit data across different product lines.

Create the Graph

1. Select **Help > Sample Data Library** and open Profit by Product.jmp.
2. Select **Graph** > **Graph Builder**.

Figure 4.29 Graph Builder Workspace

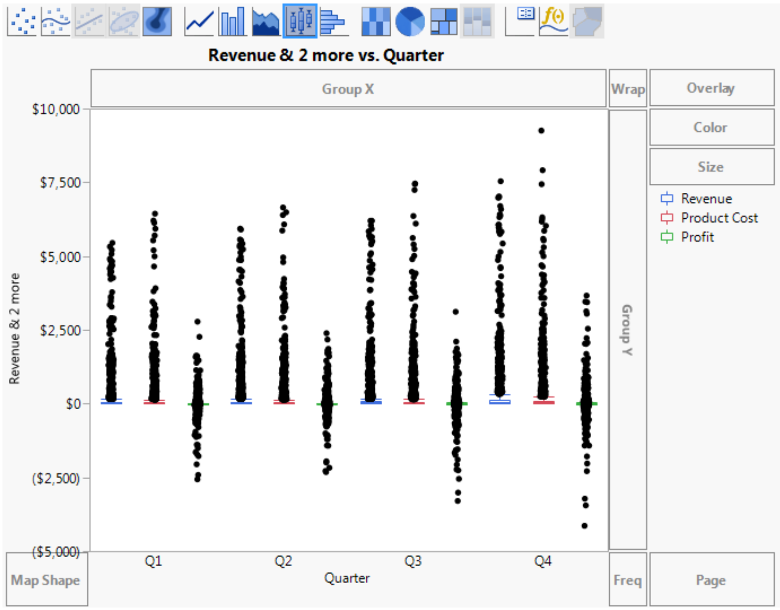


3. Click Quarter and then drag and drop it onto the X zone to assign Quarter as the X variable.
4. Click Revenue, Product Cost, and Profit, and drag and drop them onto the Y zone to assign all three variables as Y variables.

   The X and Y zones are now axes.

Note: You can also click on variables and then click a zone to assign them. However, after a zone becomes an axis, drag and drop additional variables onto the axis rather than clicking on the variables and axis.

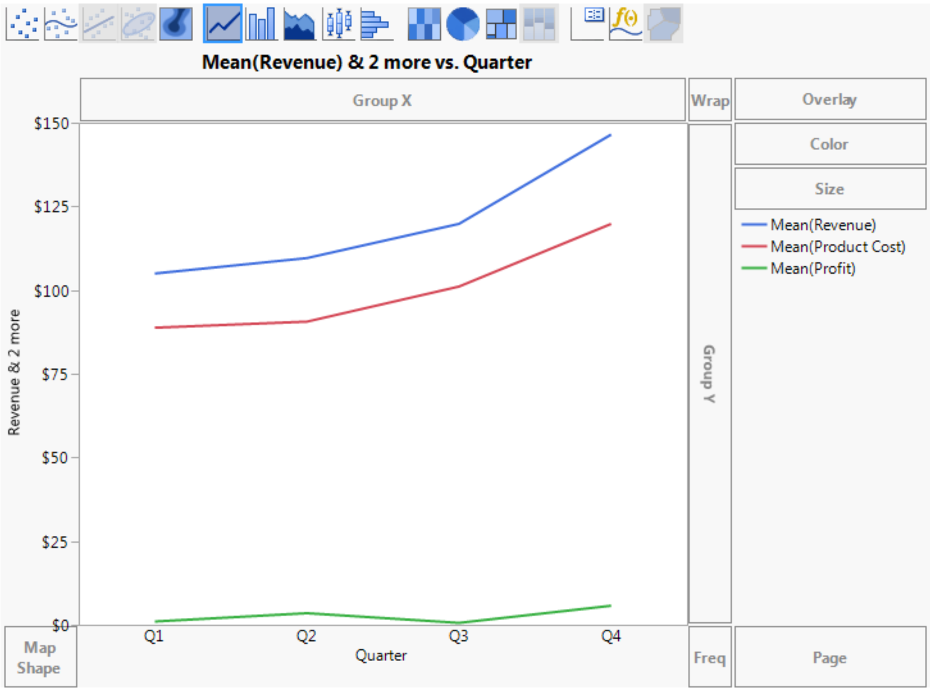
Figure 4.30 After Adding Y and X Variables



Based on the variables that you are using, Graph Builder shows side-by-side box plots.

5. To change the box plots to a line plot, click the Line  icon.

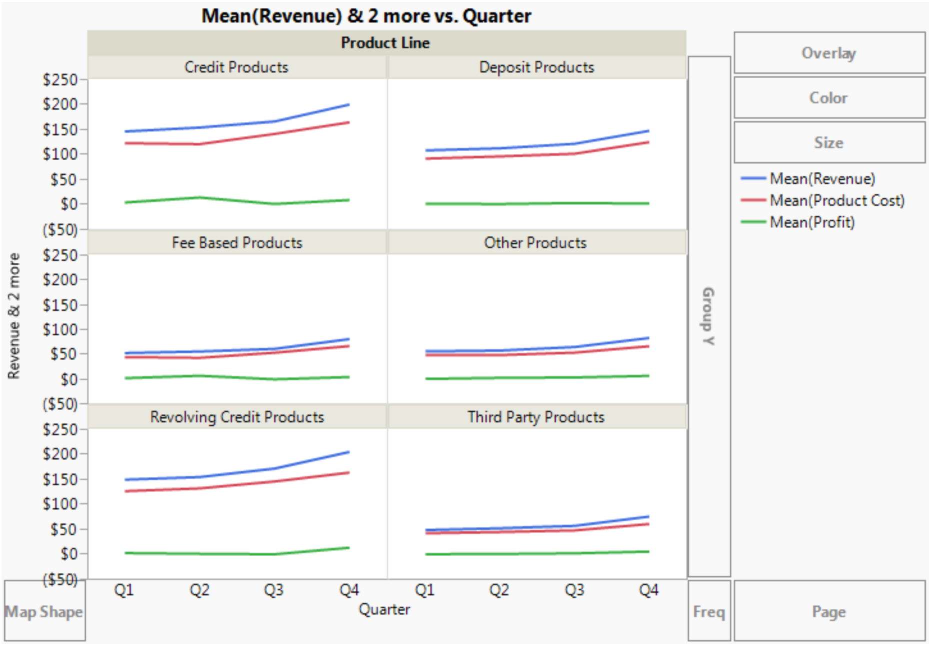
Figure 4.31 Line Plot



6. To create a separate chart for each product, click Product Line, and drag and drop it into the **Wrap** zone.

   A separate line plot is created for each product.

Figure 4.32 Final Line Plots



Interpret the Graph

Figure 4.32 shows revenue, cost, and profit broken down by product line. The business analyst was interested in seeing the difference in profitability between product lines. The line plots in Figure 4.32 can provide some answers, as follows:

- Credit products, deposit products, and revolving credit products produce more revenue than fee-based products, third-party products, and other products.
- However, the profits of all the product lines are similar.

The data table also includes data on sales channels. The business analyst wants to see how revenue, product cost, and profit differ between different sales channels.

1. To remove Product Line from the graph, click the title of the graph (Product Line) and drag and drop it into any empty area within Graph Builder.
2. To add Channel as the wrap variable, click Channel and drag and drop it into the **Wrap** zone.

Figure 4.33 Line Plots Showing Sales Channels

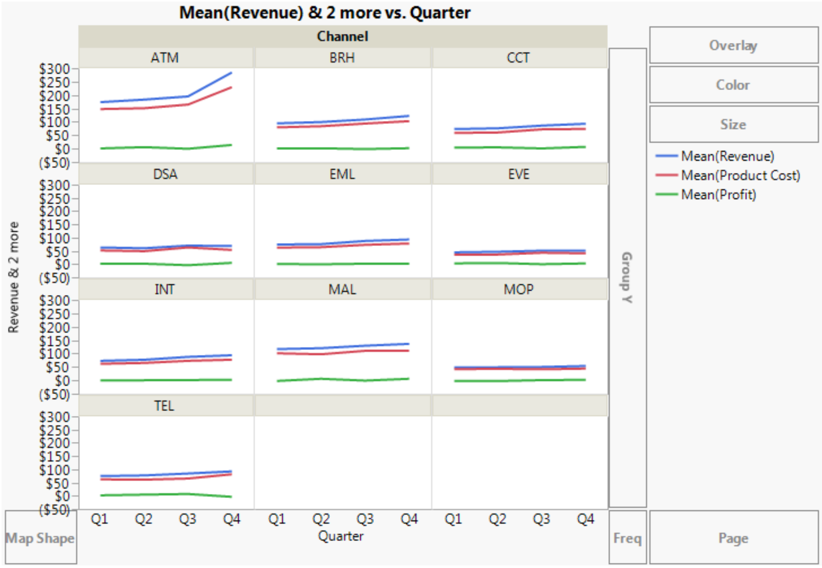
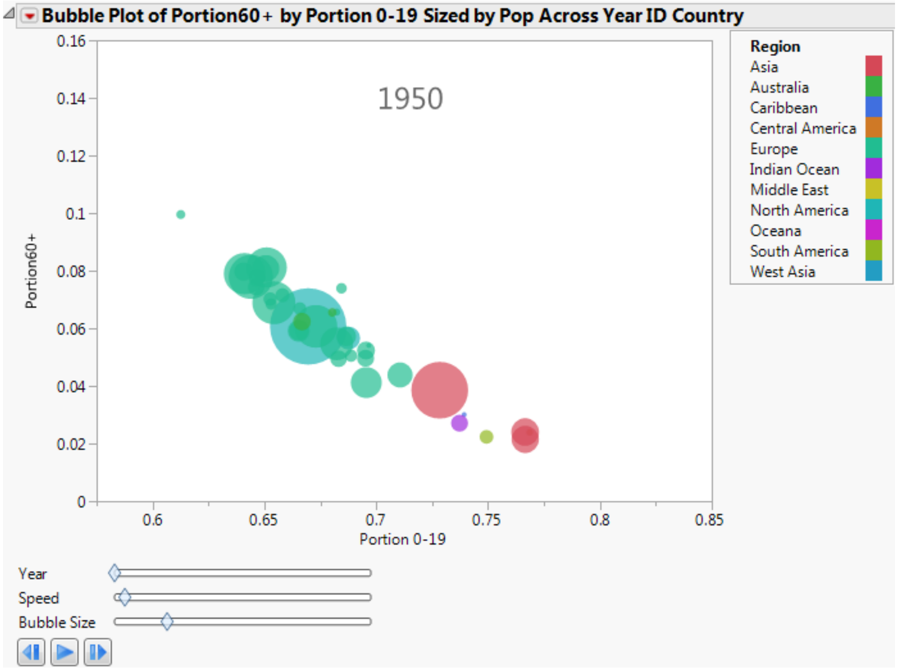


Figure 4.33 provides this answer: revenue and product cost for ATMs are the highest and are growing the most quickly.

Bubble Plots

A bubble plot is a scatterplot that represents its points as bubbles. You can change the size and color of the bubbles, and even animate them over time. With the ability to represent up to five dimensions (*x* position, *y* position, size, color, and time), a bubble plot can produce dramatic visualizations and make data exploration easy.

Figure 4.34 Example of a Bubble Plot



Scenario

This example uses the PopAgeGroup.jmp data table, which contains population statistics for 116 countries or territories between the years 1950 to 2004. Total population numbers are broken out by age group, and not every country has data for every year.

A sociologist wants to explore the following question:

- Is the age of the population of the world changing?

To answer this question, look at the relationship between the oldest (more than 59) and the youngest (younger than 20) portions of the population. Use a bubble plot to determine how this relationship changes over time.

Create the Bubble Plot

1. Select **Help > Sample Data Library** and open PopAgeGroup.jmp.
2. Select **Graph** > **Bubble Plot**.
3. Select Portion60+ and click **Y**.

   This corresponds to the *Y* variable on the bubble plot.
4. Select Portion 0-19 and click **X**.

This corresponds to the *X* variable on the bubble plot.

5. Select Country and click **ID**.

   Each unique level of the ID variable is represented by a bubble on the plot.

6. Select Year and click **Time**.

   This controls the time indexing when the bubble plot is animated.

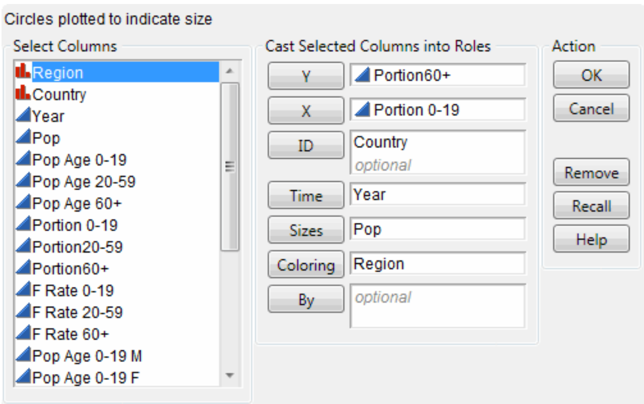
7. Select Pop and click **Sizes**.

   This controls the size of the bubbles.

8. Select Region and click **Coloring**.

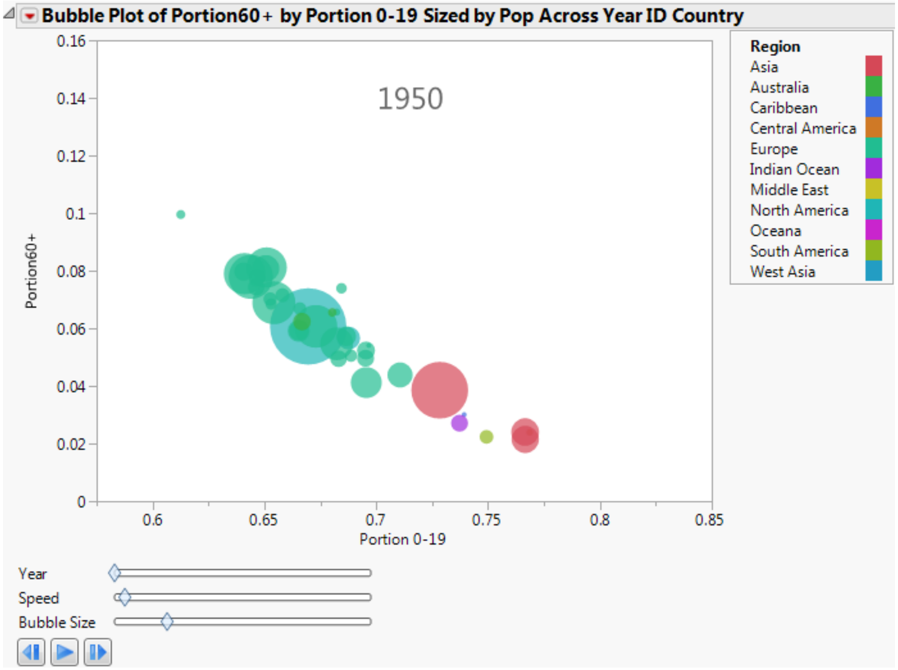
   Each level of the Coloring variable is assigned a unique color. So in this example, all the bubbles for countries located in the same region have the same color. The bubble colors that appear in Figure 4.36 are the JMP default colors.

Figure 4.35 Bubble Plot Window



9. Click **OK**.

Figure 4.36 Initial Bubble Plot



Interpret the Bubble Plot

Because the time variable (in this case, year) starts in 1950, the initial bubble plot shows the data for 1950. Animate the bubble plot to cycle through all the years by clicking the play/pause button. Each successive bubble plot shows the data for that year. The data for each year determines the following:

- The *X* and *Y* coordinates
- The bubble's sizes
- The bubble's coloring
- Bubble aggregation

Note: For detailed information about how the bubble plot aggregates information across multiple rows, see the Bubble Plots chapter in the *Essential Graphing* book.

The bubble plot for 1950 shows that if a country's proportion of people younger than 20 is high, then the proportion of people more than 59 is low.

Click the play/pause button to animate the bubble plot through the range of years. As time progresses, the Portion 0-19 decreases and the Portion60+ increases.

 plays the animation, turns to a pause button after you click it.

 pauses the animation.

 manually controls the animation back one unit of time.

 manually controls the animation forward one unit of time.

Year is used to change the time index manually.

Speed controls the speed of the animation.

Bubble Size controls the absolute sizes of the bubbles, while maintaining the relative sizes

The sociologist wanted to know how the age of the world's population is changing. The bubble plot indicates that the population of the world is getting older.

Interact with the Bubble Plot

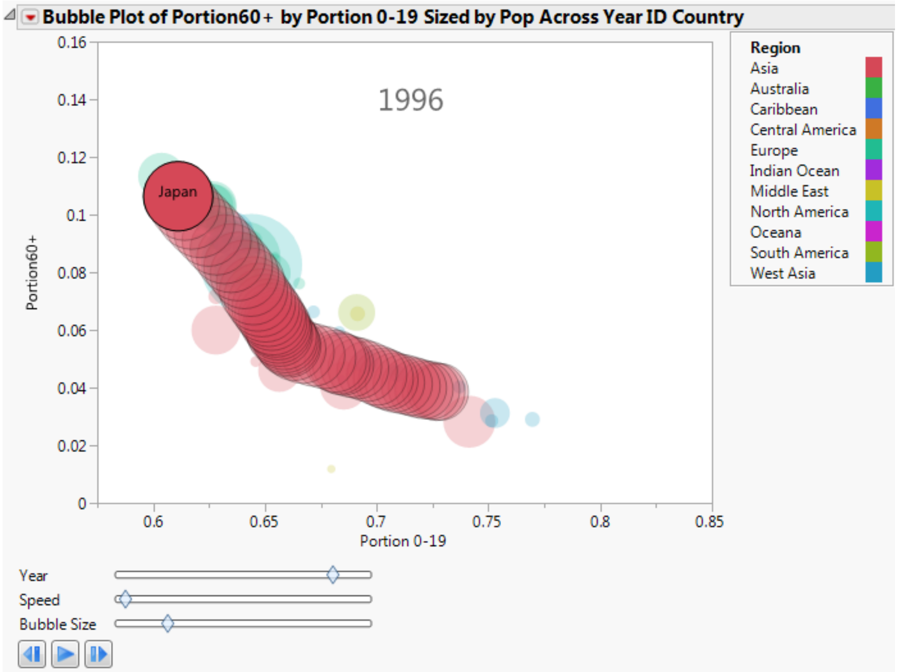
Click to select a bubble to see the trend for that bubble over time. For example, in the 1950 plot, the large bubble in the middle is Japan.

To See the Pattern of Population Changes in Japan through the Years

1. Click in the middle of the Japan bubble to select it.
2. From the red triangle menu, select **Trail Bubbles > Selected**.
3. Click the play button.

As the animation progresses through time, the Japan bubble leaves a trail of bubbles that illustrates its history.

Figure 4.37 Japan's History of Population Shifts



Focusing on the Japan bubble, you can see the following over time:

- The proportion of the population 19 years old or less decreased.
- The proportion of the population 60 years old or more increased.

Chapter 5

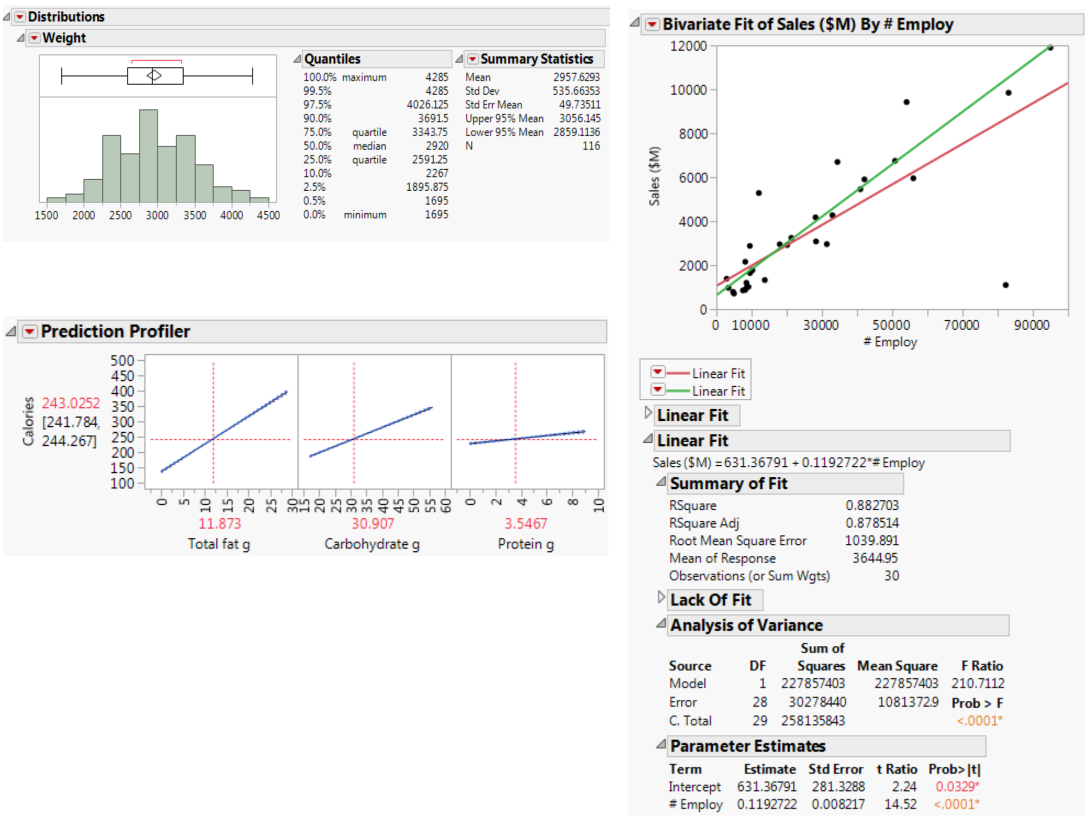
Analyze Your Data

Distributions, Relationships, and Models

Analyzing your data helps you make informed decisions. Data analysis often involves these actions:

- Examining distributions
- Discovering relationships
- Hypothesis testing
- Building models

Figure 5.1 Analysis Examples



About This Chapter

Before you analyze your data, review the following information:

- "The Importance of Graphing Your Data" on page 124
- "Understand Modeling Types" on page 126

The rest of this chapter shows you how to use some basic analytical methods in JMP:

- "Analyze Distributions" on page 130
- "Analyze Relationships" on page 136

For a description of advanced modeling and analysis techniques, refer to the following JMP books:

- *Fitting Linear Models*
- *Multivariate Methods*
- *Predictive and Specialized Modeling*
- *Consumer Research*
- *Reliability and Survival Methods*
- *Quality and Process Methods*

The Importance of Graphing Your Data

Graphing, or visualizing, your data is important to any data analysis, and should always occur before the use of statistical tests or model building. To illustrate why data visualization should be an early step in your data analysis process, consider the following example:

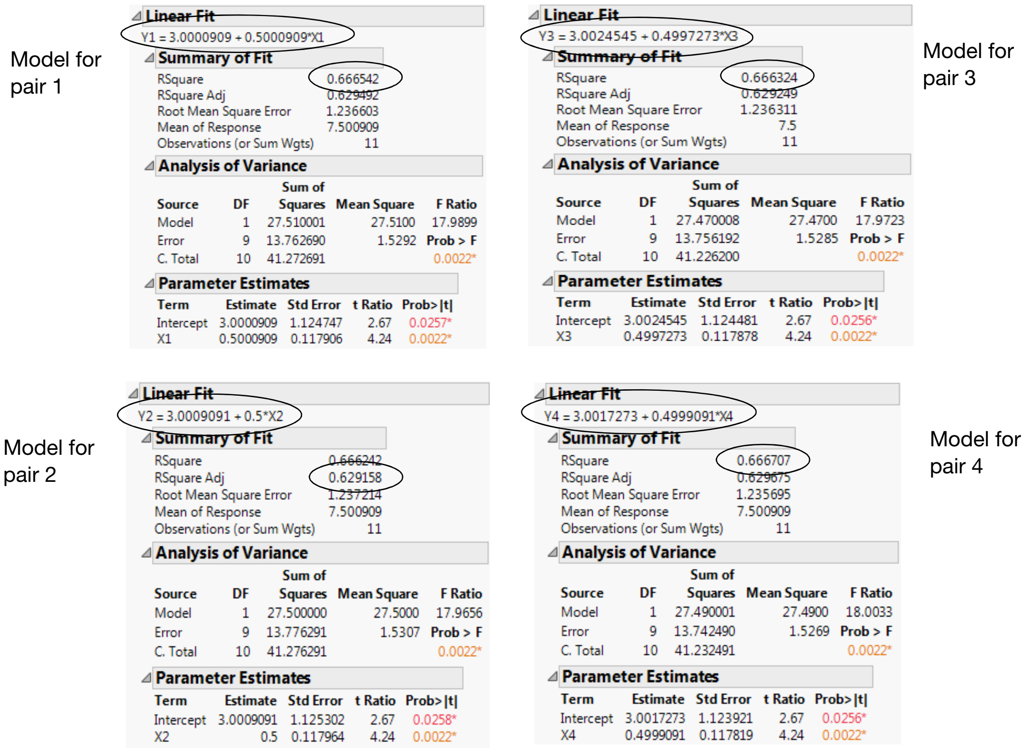
1. Select **Help > Sample Data Library** and open Anscombe.jmp (F. J. Anscombe (1973), *American Statistician*, 27, 17-21).

   This data consists of four pairs of *X* and *Y* variables.

2. In the Table panel, click the green triangle next to the **The Quartet** script.

   The script creates a simple linear regression on each pair of variables using **Fit Y by X**. The **Show Points** option is turned off, so that none of the data can be seen on the scatterplots. Figure 5.2 shows the model fit and other summary information for each regression.

Figure 5.2 Four Models



Linear Fit

Y1 = 3.0000909 + 0.5000909*X1

Summary of Fit

RSquare	0.666542
RSquare Adj	0.629492
Root Mean Square Error	1.236603
Mean of Response	7.500909
Observations (or Sum Wgts)	11

Analysis of Variance

Source	DF	Sum of Squares	Mean Square	F Ratio
Model	1	27.510001	27.5100	17.9899
Error	9	13.762690	1.5292	Prob > F
C. Total	10	41.272691		0.0022*

Parameter Estimates

Term	Estimate	Std Error	t Ratio	Prob>|t|
Intercept	3.0000909	1.124747	2.67	0.0257*
X1	0.5000909	0.117906	4.24	0.0022*

Linear Fit

Y3 = 3.0024545 + 0.4997273*X3

Summary of Fit

RSquare	0.666324
RSquare Adj	0.629249
Root Mean Square Error	1.236311
Mean of Response	7.5
Observations (or Sum Wgts)	11

Analysis of Variance

Source	DF	Sum of Squares	Mean Square	F Ratio
Model	1	27.470008	27.4700	17.9723
Error	9	13.756192	1.5285	Prob > F
C. Total	10	41.226200		0.0022*

Parameter Estimates

Term	Estimate	Std Error	t Ratio	Prob>|t|
Intercept	3.0024545	1.124481	2.67	0.0256*
X3	0.4997273	0.117878	4.24	0.0022*

Linear Fit

Y2 = 3.0009091 + 0.5*X2

Summary of Fit

RSquare	0.666242
RSquare Adj	0.629158
Root Mean Square Error	1.237214
Mean of Response	7.500909
Observations (or Sum Wgts)	11

Analysis of Variance

Source	DF	Sum of Squares	Mean Square	F Ratio
Model	1	27.500000	27.5000	17.9656
Error	9	13.776291	1.5307	Prob > F
C. Total	10	41.276291		0.0022*

Parameter Estimates

Term	Estimate	Std Error	t Ratio	Prob>|t|
Intercept	3.0009091	1.125302	2.67	0.0258*
X2	0.5	0.117964	4.24	0.0022*

Linear Fit

Y4 = 3.0017273 + 0.4999091*X4

Summary of Fit

RSquare	0.666707
RSquare Adj	0.629675
Root Mean Square Error	1.235695
Mean of Response	7.500909
Observations (or Sum Wgts)	11

Analysis of Variance

Source	DF	Sum of Squares	Mean Square	F Ratio
Model	1	27.490001	27.4900	18.0033
Error	9	13.742490	1.5269	Prob > F
C. Total	10	41.232491		0.0022*

Parameter Estimates

Term	Estimate	Std Error	t Ratio	Prob>|t|
Intercept	3.0017273	1.123921	2.67	0.0256*
X4	0.4999091	0.117819	4.24	0.0022*

Notice that all four models and the RSquare values are nearly identical. The fitted model in each case is essentially $Y = 3 + 0.5X$, and the RSquare value in each case is essentially 0.66. If your data analysis took into account only the above summary information, you would likely conclude that the relationship between X and Y is the same in each case. However, at this point, you have not visualized your data. Your conclusion might be wrong.

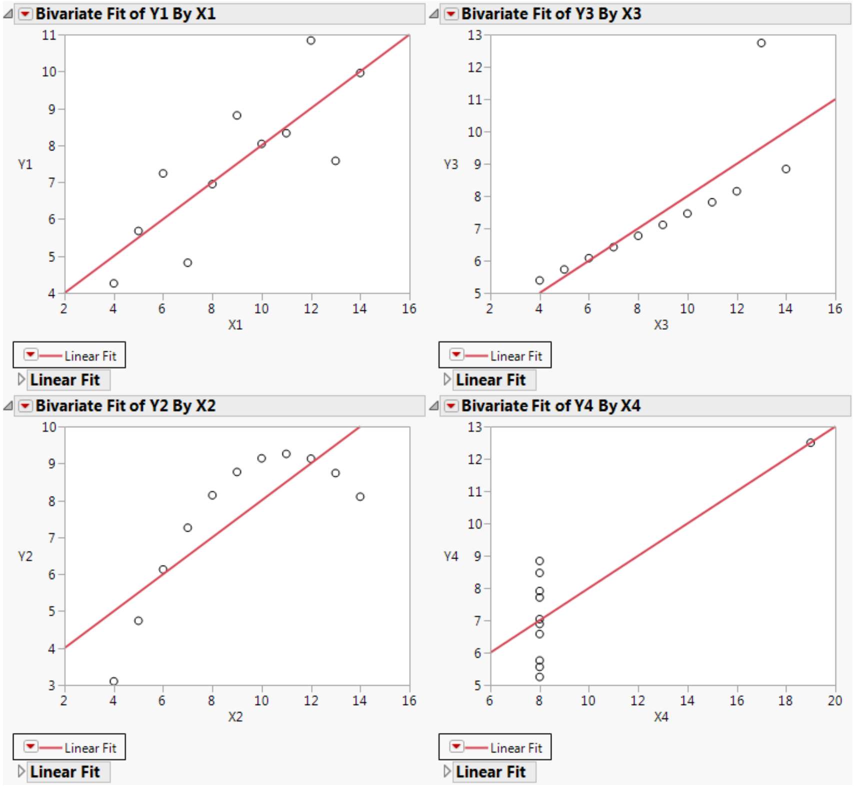
To Visualize the Data, Add the Points to All Four Scatterplots

1. Hold down the Ctrl key.
2. From the red triangle menu for any one of the Bivariate Fits, select **Show Points**.

Figure 5.3 Scatterplots with Points Added

Plot for pair 1

Plot for pair 3



Plot for pair 2

Plot for pair 4

The scatterplots show that the relationship between *X* and *Y* is not the same for the four pairs, although the lines describing the relationships are the same:

- Plot 1 represents a linear relationship.
- Plot 2 represents a non-linear relationship.
- Plot 3 represents a linear relationship, except for one outlier.
- Plot 4 has all the data at x = 8, except for one point.

This example illustrates that conclusions that are based on statistics alone can be inadequate. A visual exploration of the data should be an early part of any data analysis.

Understand Modeling Types

In JMP, data can be of different types. JMP refers to this as the modeling type of the data. Table 5.1 describes the three modeling types in JMP.

Table 5.1 Modeling Types

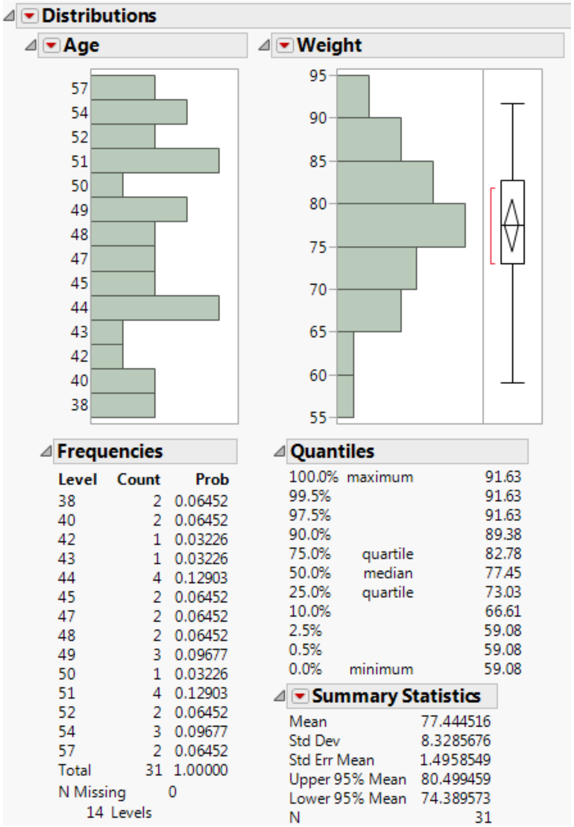
Modeling Type and Description	Examples	Specific Example
Continuous Numeric data only. Used in operations like sums and means.	Height Temperature Time	The time to complete a test might be 2 hours, or 2.13 hours.
Ordinal Numeric or character data. Values belong to ordered categories.	Month (1,2,...,12) Letter grade (A, B,...F) Size (small, medium, large)	The month of the year can be 2 (February) or 3 (March), but not 2.13. February comes before March.
Nominal Numeric or character data. Values belong to categories, but the order is not important.	Gender (M or F) Color Test result (pass or fail)	The gender can be M or F, with no order. Gender categories can also be represented by a number (M=1 and F=2).

Example of Viewing Modeling Type Results

Different modeling types produce different results in JMP. To see an example of the differences, follow these steps:

1. Select **Help > Sample Data Library** and open Linnerud.jmp.
2. Select **Analyze** > **Distribution**.
3. Select Age and Weight and click **Y, Columns**.
4. Click **OK**.

Figure 5.4 Distribution Results for Age and Weight



Although Age and Weight are both numeric variables, they are not treated the same. Table 5.2 compares the differences between the results for weight and age.

Table 5.2 Results for weight and age

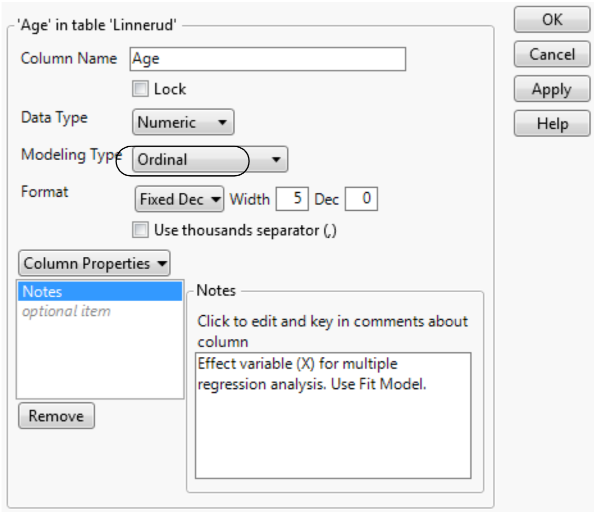
Variable	Modeling Type	Results
Weight	Continuous	Histogram, Quantiles, and Summary Statistics
Age	Ordinal	Bar chart and Frequencies

Change the Modeling Type

To treat a variable differently, change the modeling type. For example, in Figure 5.4, the modeling type for Age is ordinal. Remember that for an ordinal variable, JMP calculates frequency counts. Suppose that you wanted to find the average age instead of frequency counts. Change the modeling type to continuous, which shows the mean age.

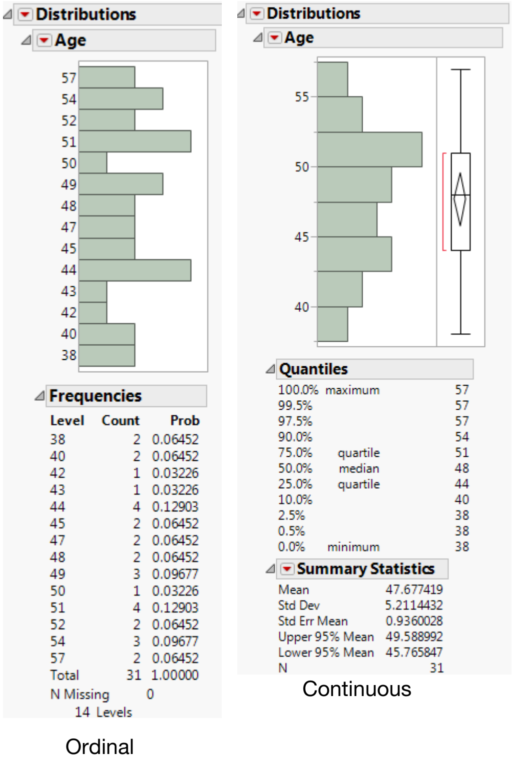
1. Double-click the Age column heading. The Column Info window appears.
2. Change the Modeling Type to **Continuous**.

Figure 5.5 Column Info Window



3. Click **OK**.
4. Repeat the steps in the example (see "Example of Viewing Modeling Type Results" on page 127) to create the distribution. Figure 5.6 shows the distribution results when Age is ordinal and continuous.

Figure 5.6 Different Modeling Types for Age



When age is ordinal, you can see the frequency counts for each age. For example, age 48 appears two times. When age is continuous, you can find the mean age, which is nearly 48 (47.677)

Analyze Distributions

To analyze a single variable, you can examine the distribution of the variable, using the Distribution platform. Report content for each variable varies, depending on whether the variable is categorical (nominal or ordinal) or continuous.

Note: For complete details about the **Distribution** platform, see the Distributions chapter in the *Basic Analysis* book.

Distributions of Continuous Variables

Analyzing a continuous variable might include questions such as the following:

- Does the shape of the data match any known distributions?
- Are there any outliers in the data?
- What is the average of the data?
- Is the average statistically different from a target or historical value?
- How spread out are the data? In other words, what is the standard deviation?
- What are the minimum and maximum values?

You can answer these and other questions with graphs, summary statistics, and simple statistical tests.

Scenario

This example uses the Car Physical Data.jmp data table, which contains information about 116 different car models.

A planning specialist has been asked by a railroad company to determine the possible issues involved in transporting cars by train. Using the data, the planning specialist wants to explore the following questions:

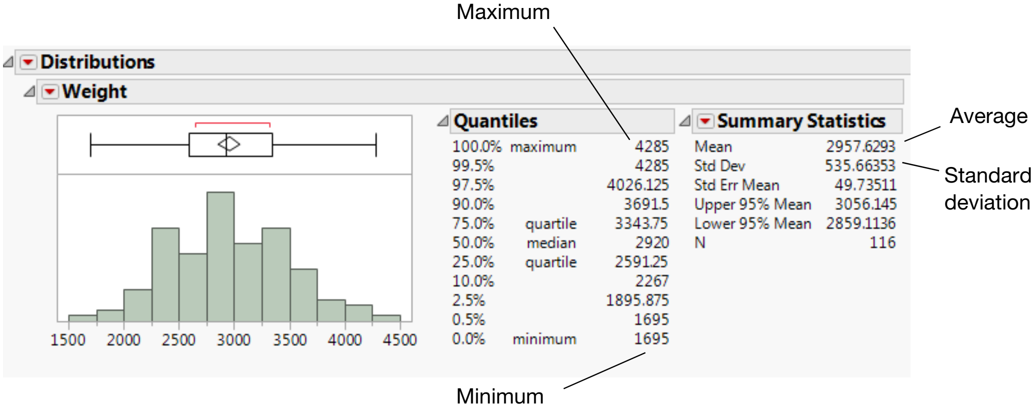
- What is the average car weight?
- How spread out are the cars' weights (standard deviation)?
- What are the minimum and maximum weights of cars?
- Are there any outliers in the data?

Use a histogram of weight to answer these questions.

Create the Histogram

1. Select **Help > Sample Data Library** and open Car Physical Data.jmp.
2. Select **Analyze** > **Distribution**.
3. Select Weight and click **Y, Columns**.
4. Click **OK**.
5. To rotate the report window, select **Display Options** > **Horizontal Layout** from the red triangle menu next to **Weight**.

Figure 5.7 Distribution of Weight



The report window contains three sections:

- A histogram and a box plot to visualize the data.
- A Quantiles report that shows the percentiles of the distribution.
- A Summary Statistics report that shows the mean, standard deviation, and other statistics.

Interpret the Distribution Results

Using the results presented in Figure 5.7, the planning specialist can answer the questions.

What is the average car weight? The Histogram shows a weight of around 3,000 lbs.

How spread out are the weights (standard deviation)? The Summary Statistics show a weight of around 2,958 lbs. The Summary Statistics show a standard deviation of around 536 lbs.

What are the minimum and maximum weights? The Histogram shows a minimum of around 1,500 lbs. and a maximum of around 4,500 lbs. The Quantiles show a minimum of around 1,695 lbs. and a maximum of around 4,285 lbs.

Are there any outliers? No.

The default report window in Figure 5.7 provides a minimal set of graphs and statistics. Additional graphs and statistics are available on the red triangle menu.

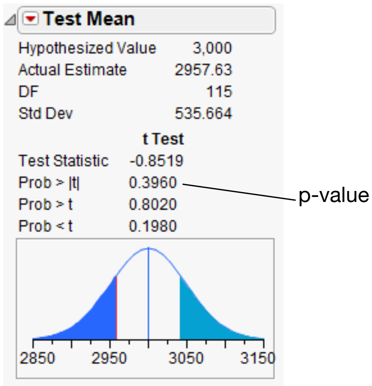
Draw Conclusions

Based on other research, the railroad company has determined that an average weight of 3000 pounds is the most efficient to transport. Now, the planning specialist needs to find out whether the average car weight in the general population of cars that they might transport is 3000 pounds. Use a t-test to draw inferences about the broader population based on this sample of the population.

Test Conclusions

1. From the red triangle menu for **Weight**, select **Test Mean**.
2. In the window that appears, type 3000 in the Specify Hypothesized Mean box.
3. Click **OK**.

Figure 5.8 Test Mean Results



Interpret the t-Test

The primary result of a t-test is the p-value. In this example, the p-value is 0.396 and the analyst is using a significance level of 0.05. Since 0.396 is greater than 0.05, you cannot conclude that the average weight of car models in the broader population is significantly different from 3000 pounds. Had the p-value been lower than the significance level, the planning specialist would have concluded that the average car weight in the broader population *is* significantly different from 3000 pounds.

Distributions of Categorical Variables

Analyzing a categorical (ordinal or nominal) variable might include questions such as the following:

- How many levels does the variable have?
- How many data points does each level have?
- Is the data uniformly distributed?
- What proportions of the total do each level represent?

Scenario

See the scenario in "Distributions of Continuous Variables" on page 130.

Now that the railroad company has determined that the average weight of the cars is not significantly different from the target weight, there are more questions to address.

The planning specialist wants to answer these questions for the railroad company:

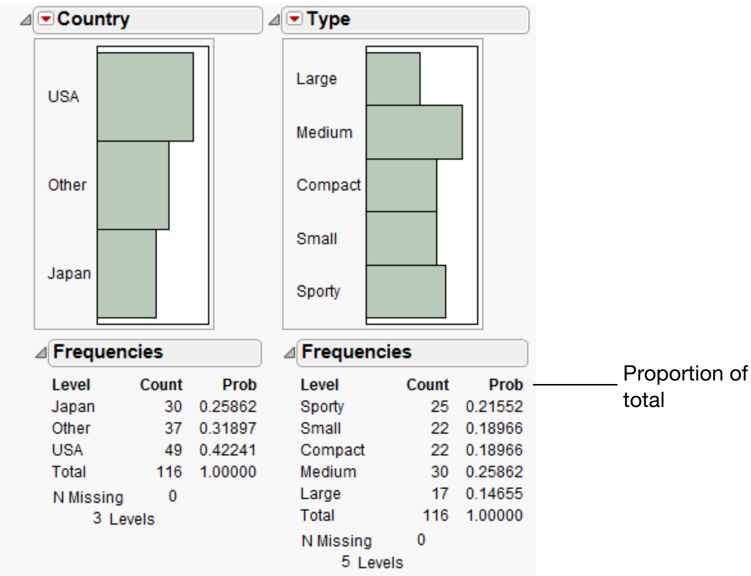
- What are the types of cars?
- What are the countries of origin?

To answer these questions, look at the distribution for Type and Country.

Create the Distribution

1. Select **Help > Sample Data Library** and open Car Physical Data.jmp.
2. Select **Analyze** > **Distribution**.
3. Select Country and Type and click **Y, Columns**.
4. Click **OK**.

Figure 5.9 Distribution for Country and Type



Interpret the Distribution Results

The report window includes a bar chart and a Frequencies report for Country and Type. The bar chart is a graphical representation of the frequency information provided in the Frequencies report. The Frequencies report contains the following:

- Categories of data. For example, Japan is a category of Country, and Sporty is a category of Type.

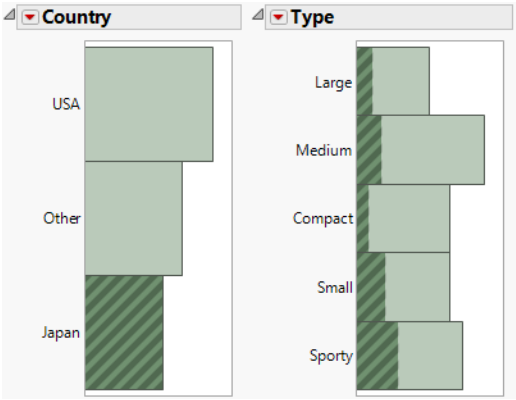
- Total counts for each category.
- Proportion of the total each category represents.

For example, there are 22 compact cars, or about 19% of the 116 observations.

Interact with the Distribution Results

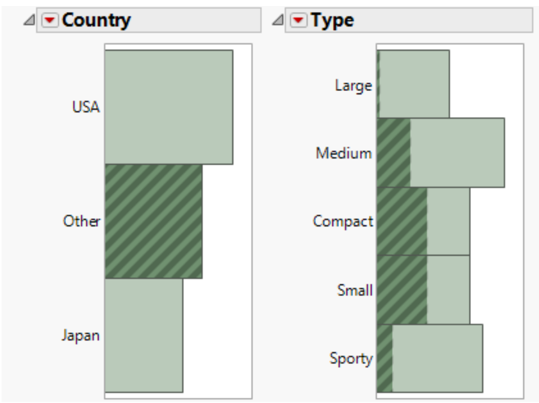
Selecting a bar in one chart also selects the corresponding data in the other chart. For example, select the Japan bar in the Country bar chart to see that a large number of Japanese cars are sporty.

Figure 5.10 Japanese Cars



Select the Other category to see that a majority of these cars are small or compact, and almost none are large.

Figure 5.11 Other Cars



Analyze Relationships

Scatterplots and other such graphs can help you visualize relationships between variables. Once you have visualized relationships, the next step is to analyze those relationships so that you can describe them numerically. That numerical description of the relationship between variables is called a *model*. Even more importantly, a model also predicts the average value of one variable (*Y*) from the value of another variable (*X*). The *X* variable is also called a predictor. Generally, this model is called a *regression* model.

With JMP, the **Fit Y by X** platform and the **Fit Model** platform creates regression models.

Note: Only the basic platforms and options are covered here. For complete details and explanations of all platform options, see *Basic Analysis*, *Essential Graphing*, and the books listed in "About This Chapter" on page 124.

Table 5.3 shows the four primary types of relationships.

Table 5.3 Relationship Types

X	Y	Section
Continuous	Continuous	• "Use Regression with One Predictor" on page 136 • "Use Regression with Multiple Predictors" on page 150
Categorical	Continuous	• "Compare Averages for One Variable" on page 141 • "Compare Averages for Multiple Variables" on page 146
Categorical	Categorical	"Compare Proportions" on page 144
Continuous	Categorical	Logistic regression is an advanced topic. See the Logistic Analysis chapter in the *Basic Analysis* book.

Use Regression with One Predictor

Scenario

This example uses the Companies.jmp data table, which contains financial data for 32 companies from the pharmaceutical and computer industries.

Intuitively, it makes sense that companies with more employees can generate more sales revenue than companies with fewer employees. A data analyst wants to predict the overall sales revenue for each company based on the number of employees.

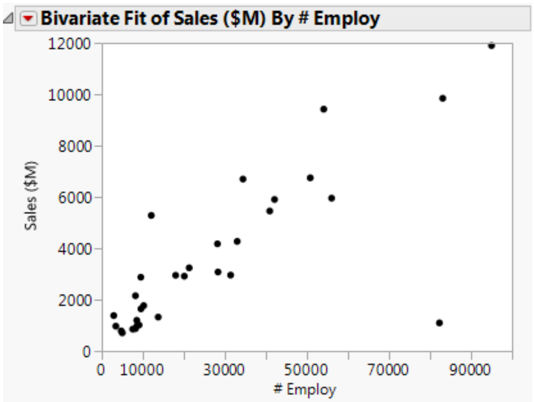
To accomplish this task, do the following:

- "Discover the Relationship" on page 137
- "Fit the Regression Model" on page 137
- "Predict Average Sales" on page 139

Discover the Relationship

First, create a scatterplot to see the relationship between the number of employees and the amount of sales revenue. This scatterplot was created in "Create the Scatterplot" on page 96 in the "Visualize Your Data" chapter. After hiding and excluding one outlier (a company with significantly more employees and higher sales), the plot in Figure 5.12 shows the result.

Figure 5.12 Scatterplot of Sales ($M) versus # Employ

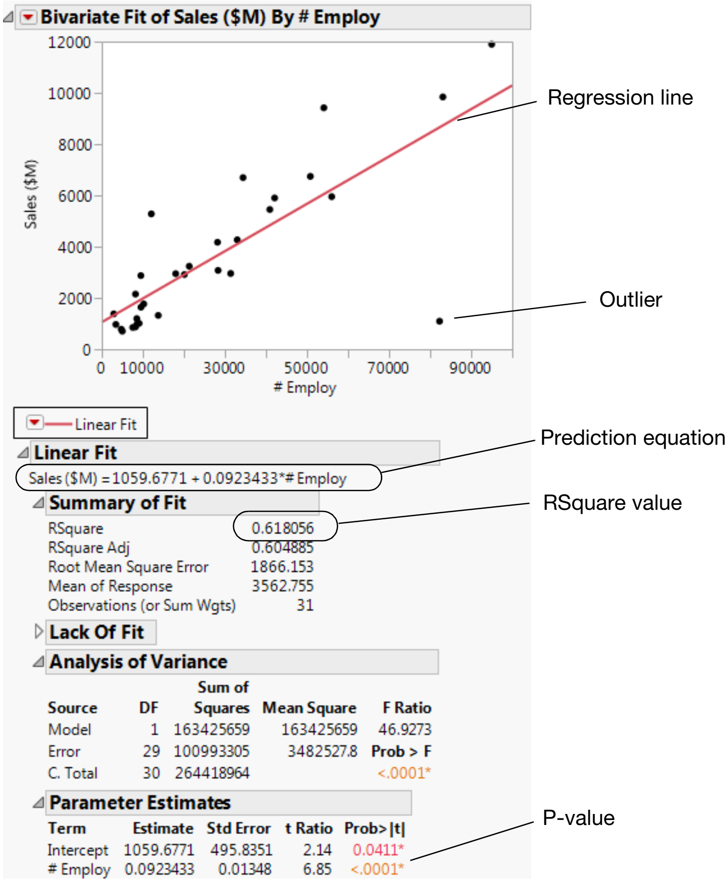


This scatterplot provides a clearer picture of the relationship between sales and the number of employees. As expected, the more employees a company has, the higher sales that it can generate. This visually confirms the data analyst's guess, but it does not predict sales for a given number of employees.

Fit the Regression Model

To predict the sales revenue from the number of employees, fit a regression model. From the red triangle for **Bivariate Fit**, select **Fit Line**. A regression line is added to the scatterplot and reports are added to the report window.

Figure 5.13 Regression Line



Within the reports, look at the following results:

- the p-value of <.0001
- the RSquare value of 0.618

From these results, the data analyst can conclude the following:

- The p-value is less than the significance level of 0.05. Therefore, including the number of employees in the prediction model significantly improves the ability to predict average sales.
- Since the RSquare value in this example is large, this confirms that a prediction model based on the number of employees can predict sales revenue. The RSquare value shows the strength of a relationship between variables, also called the *correlation*. A correlation of 0 indicates no relationship between the variables, and a correlation of 1 indicates a perfect linear relationship.

Predict Average Sales

Use the regression model to predict the average sales a company might expect if they have a certain number of employees. The prediction equation for the model is included in the report:

Average sales = 1059.68 + 0.092*employees

For example, in a company with 70,000 employees, the equation is as follows:

$7,499.68 = 1059.68 + 0.092*70,000

In the lower right area of the current scatterplot, there is an outlier that does not follow the general pattern of the other companies. The data analyst wants to know whether the prediction model changes when this outlier is excluded.

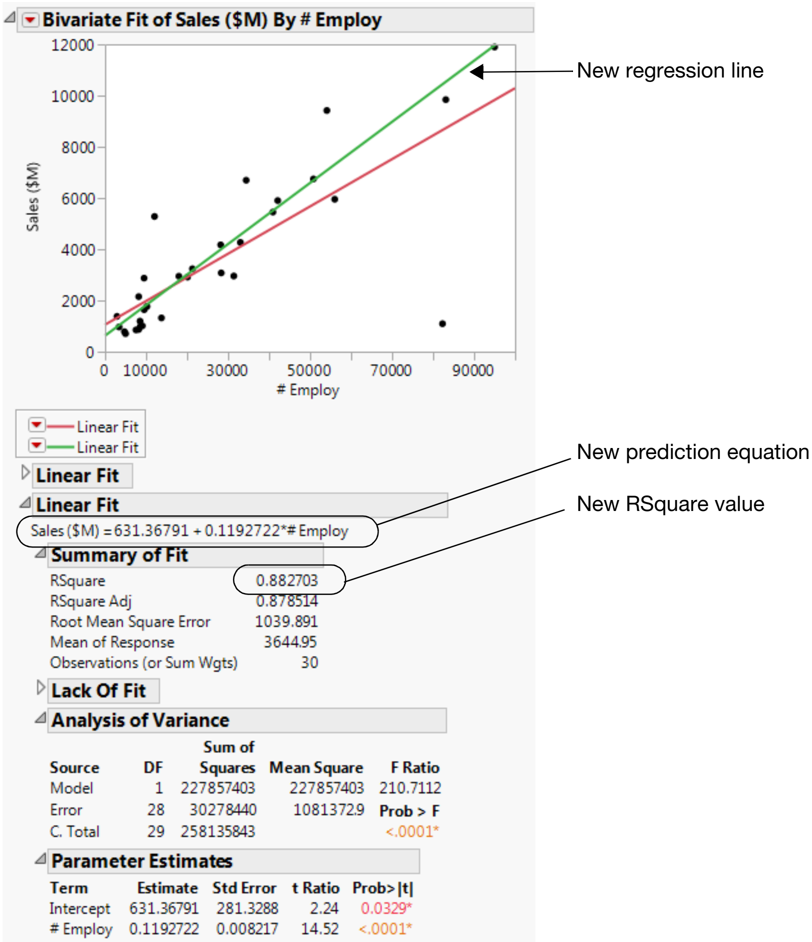
Exclude the Outlier

1. Click on the outlier.
2. Select **Rows** > **Exclude/Unexclude**.
3. Fit this model by selecting **Fit Line** from the red triangle menu for **Bivariate Fit**.

The following are added to the report window (see Figure 5.14):

- a new regression line
- a new Linear Fit report, which includes:
  - a new prediction equation
  - a new RSquare value

Figure 5.14 Comparing the Models



Interpret the Results

Using the results in Figure 5.14, the data analyst can make the following conclusions:

- The outlier was pulling down the regression line for the larger companies, and pulling the line up for the smaller companies.
- The new model fits the data better, since the new RSquare value (0.88) is closer to 1 than the first RSquare value (0.618).

Draw Conclusions

Using the new prediction equation, the predicted average sales for a company with 70,000 employees can be calculated as follows:

$8961.37 = 631.37 + 0.119*70,000

The prediction for the first model was $7499.68, so this model predicts a higher sales total by $1461.69.

The second model, after removing the outlier, describes and predicts sales totals based on the number of employees better than the first model. The data analyst now has a good model to use.

Compare Averages for One Variable

If you have a continuous Y variable, and a categorical X variable, you can compare averages across the levels of the X variable.

Scenario

This example uses the Companies.jmp data table, which contains financial data for 32 companies from the pharmaceutical and computer industries.

A financial analyst wants to explore the following question:

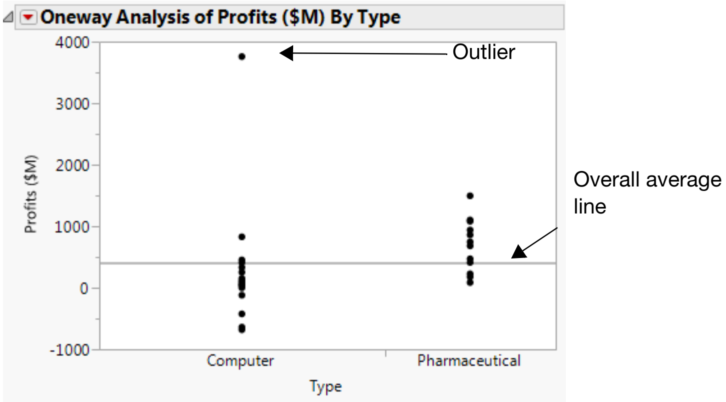
- How do the profits of computer companies compare to the profits of pharmaceutical companies?

To answer this question, fit Profits ($M) by Type.

Discover the Relationship

1. Select **Help > Sample Data Library** and open Companies.jmp.
2. If you still have the Companies.jmp sample data table open, you might have rows that are excluded or hidden. To return the rows to the default state (all rows included and none hidden), select **Rows** > **Clear Row States**.
3. Select **Analyze** > **Fit Y by X**.
4. Select Profits ($M) and click **Y, Response**.
5. Select Type and click **X, Factor**.
6. Click **OK**.

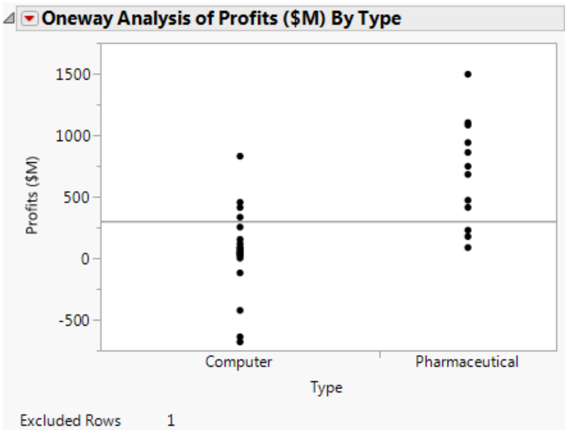
Figure 5.15 Profits by Company Type



There is an outlier in the Computer Type. The outlier is stretching the scale of the plot and making it difficult to compare the profits. Exclude and hide the outlier:

1. Click on the outlier.
2. Select **Rows** > **Exclude/Unexclude**. The data point is no longer included in calculations.
3. Select **Rows** > **Hide/Unhide**. The data point is hidden from all graphs.
4. To re-create the plot without the outlier, select **Redo** > **Redo Analysis** from the red triangle menu for Oneway Analysis. You can close the original Scatterplot window.

Figure 5.16 Updated Plot

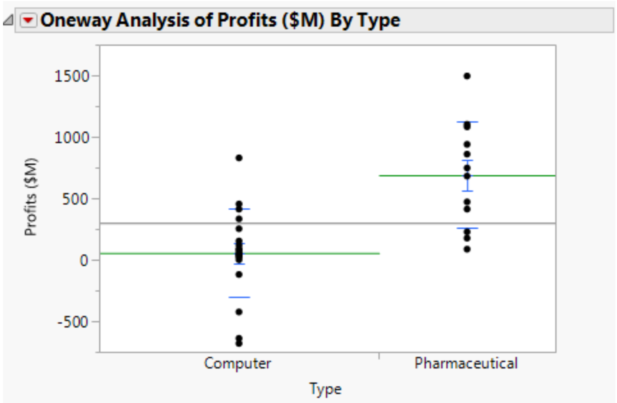


Removing the outlier gives the financial analyst a clearer picture of the data.

5. To continue analyzing the relationship, select these options from the red triangle menu for Oneway Analysis:
   – **Display Options** > **Mean Lines**. This adds mean lines to the scatterplot.

– **Means and Std Dev**. This displays a report that provides averages and standard deviations.

Figure 5.17 Mean Lines and Report



Excluded Rows 1

Means and Std Deviations

Level	Number	Mean	Std Dev	Std Err Mean	Lower 95%	Upper 95%
Computer	19	55.753	357.909	82.11	-116.8	228.26
Pharmaceutical	12	690.075	429.988	124.13	416.9	963.28

Interpret the Results

The financial analyst wanted to know how the profits of computer companies compared to the profits of pharmaceutical companies. The updated scatterplot shows that pharmaceutical companies have higher average profits than computer companies. In the report, if you subtract one mean value from the other, the difference in profit is about $635 million. The plot also shows that some of the computer companies have negative profits and all of the pharmaceutical companies have positive profits.

Perform the T-test

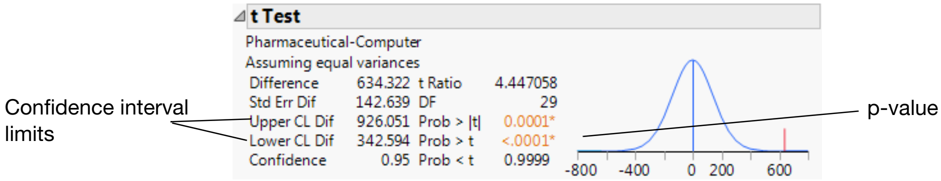
The financial analyst has looked at only a sample of companies (the companies in the data table). The financial analyst now wants to examine these questions:

- Does a difference exist in the broader population, or is the difference of $635 million due to chance?
- If there is a difference, what is it?

To answer these questions, perform a two-sample t-test. A t-test lets you use data from a sample to make inferences about the larger population.

To perform the t-test, select **Means/Anova/Pooled t** from the red triangle for Oneway Analysis.

Figure 5.18 t Test Results



The p-value of 0.0001 is less than the significance level of 0.05, which indicates statistical significance. Therefore, the financial analyst can conclude that the difference in average profits for the sample data is not due to chance alone. This means that in the larger population, the average profits for pharmaceutical companies are different from the average profits for computer companies.

Draw Conclusions

Use the confidence interval limits to determine how much difference exists in the profits of both types of companies. Look at the **Upper CL Dif** and **Lower CL Dif** values in Figure 5.18. The financial analyst concludes that the average profit of pharmaceutical companies is between $343 million and $926 million higher than the average profit of computer companies.

Compare Proportions

If you have categorical *X* and *Y* variables, you can compare the proportions of the levels within the Y variable to the levels within the X variable.

Scenario

This example continues to use the Companies.jmp data table. In "Compare Averages for One Variable" on page 141, a financial analyst determined that pharmaceutical companies have higher profits on average than do computer companies.

The financial analyst wants to know whether the size of a company affects profits more for one type of company than the other? However, before examining this question, the financial analyst needs to know whether the populations of computer and pharmaceutical companies consist of the same proportions of small, medium, and big companies.

Discover the Relationship

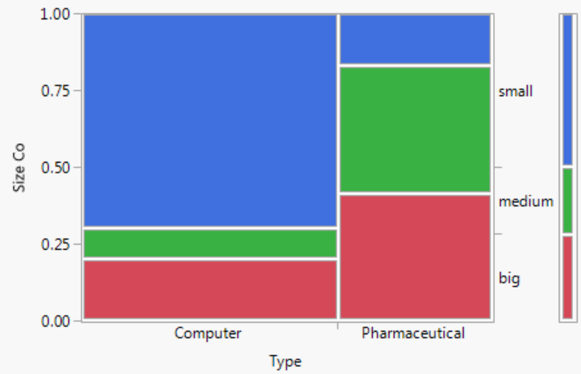
1. Select **Help > Sample Data Library** and open Companies.jmp.
2. If you still have the Companies.jmp data file open from the previous example, you might have rows that are excluded or hidden. To return the rows to the default state (all rows included and none hidden), select **Rows** > **Clear Row States**.
3. Select **Analyze** > **Fit Y by X**.

4. Select Size Co and click **Y, Response**.
5. Select Type and click **X, Factor**.
6. Click **OK**.

Figure 5.19 Company Size by Company Type

Contingency Analysis of Size Co By Type

Mosaic Plot



Contingency Table

Table legend

Type	Size Co: big	medium	small	
Count Total % Col % Row %				
Computer	4 12.50 44.44 20.00	2 6.25 28.57 10.00	14 43.75 87.50 70.00	20 62.50
Pharmaceutical	5 15.63 55.56 41.67	5 15.63 71.43 41.67	2 6.25 12.50 16.67	12 37.50
	9 28.13	7 21.88	16 50.00	32

Tests

N	DF	-LogLike	RSquare (U)
32	2	4.7711597	0.1439

Test	ChiSquare	Prob>ChiSq
Likelihood Ratio	9.542	0.0085*
Pearson	8.957	0.0114*

Pearson test

Warning: 20% of cells have expected count less than 5, ChiSquare suspect.

The Contingency Table contains information that is not applicable for this example. From the red triangle menu for **Contingency Table** deselect **Total %** and **Col %** to remove that information. Figure 5.20 shows the updated table.

Figure 5.20 Updated Contingency Table

Contingency Table

		Size Co			
	Count Row %	big	medium	small	
Type	Computer	4 20.00	2 10.00	14 70.00	20
	Pharmaceutical	5 41.67	5 41.67	2 16.67	12
		9	7	16	32

Interpret the Results

The statistics in the Contingency Table are graphically represented in the Mosaic Plot. Together, the Mosaic Plot and the Contingency Table compare the percentages of small, medium, and big companies between the two industries. For example, the Mosaic Plot shows that the computer industry has a higher percentage of small companies compared to the pharmaceutical industry. The Contingency Table shows the exact statistics: 70% of computer companies are small, and about 17% of pharmaceutical companies are small.

Interpret the Test

The financial analyst has looked at only a sample of companies (the companies in the data table). The financial analyst needs to know whether the percentages differ in the broader populations of all computer and pharmaceutical companies.

To answer this question, use the p-value from the Pearson test in the **Tests** report. See Figure 5.19. Since the p-value of 0.011 is less than the significance level of 0.05, the financial analyst concludes the following:

- The differences in the sample data are not due to chance alone.
- The percentages differ in the broader population.

Now the financial analyst knows that the proportions of small, medium, and big companies are different, and can answer the question: Does the size of company affect profits more for one type of company than the other?

Compare Averages for Multiple Variables

The section "Compare Averages for One Variable" on page 141, compared averages across the levels of a categorical variable. To compare averages across the levels of two or more variables at once, use the *Analysis of Variance* technique (or ANOVA).

Scenario

The financial analyst can answer the question that we started to work through in the Comparing Proportions section, which is: Does the size of the company have a larger effect on the company's profits, based on type (pharmaceutical or computer)?

To answer this question, compare the company profits by these two variables:

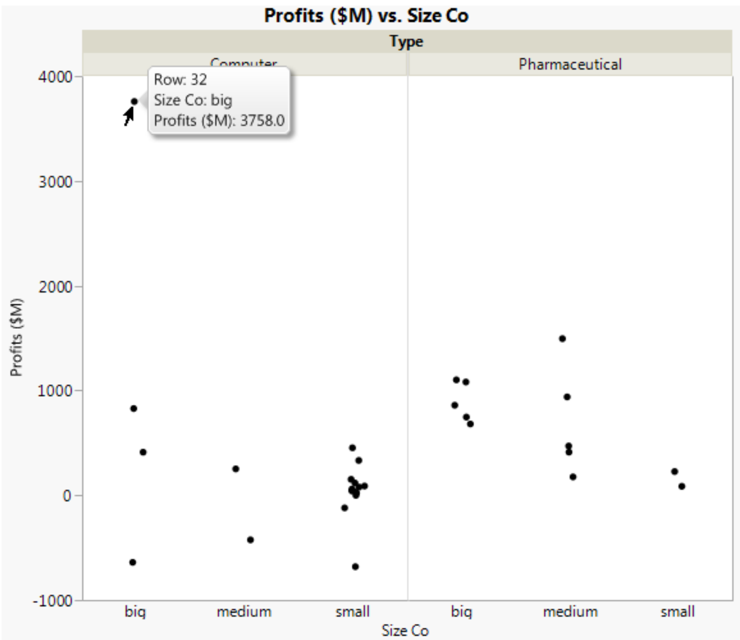
- Type (pharmaceutical or computer)
- Size (small, medium, big)

Discover the Relationship

To visualize the differences in profit for all of the combinations of type and size, use a graph:

1. Select **Help > Sample Data Library** and open Companies.jmp.
2. Select **Graph** > **Graph Builder**. The Graph Builder window appears.
3. Click Profits ($M) and drag and drop it into the **Y** zone.
4. Click Size Co and drag and drop it into the **X** zone.
5. Click Type and drag and drop it into the **Group X** zone.

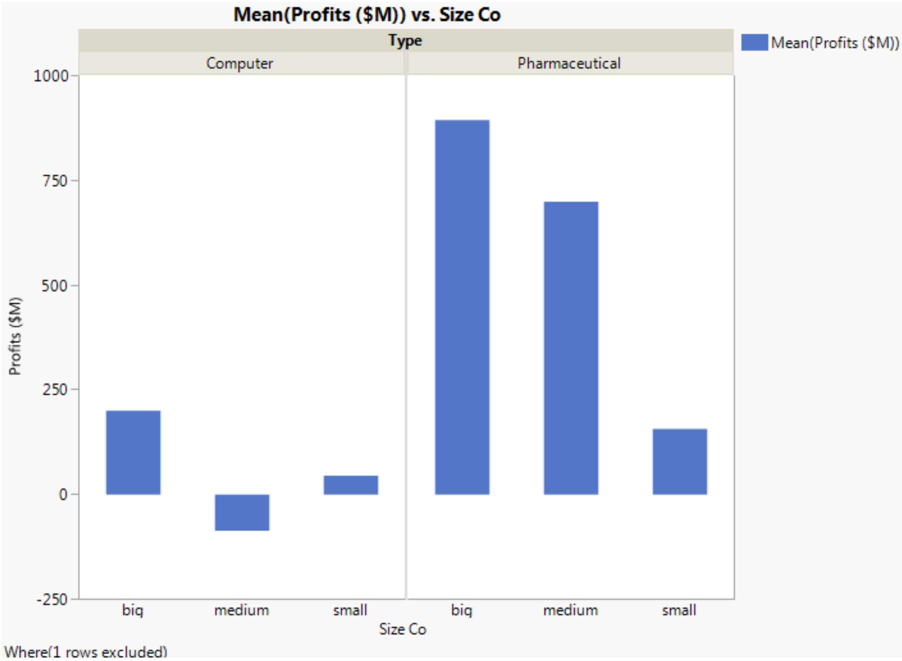
Figure 5.21 Graph of Company Profits



The graph shows that one big computer company has very large profits. That outlier is stretching the scale of the graph, making it difficult to compare the other data points.

6. Select the outlier, then right-click and select **Rows > Row Exclude**. The point is removed, and the scale of the graph automatically updates.
7. Click on the Bar  icon. Comparing mean profits is easier with bar charts than with points.

Figure 5.22 Graph with Outlier Removed



The updated graph shows that pharmaceutical companies have higher average profits. The graph also shows that profits differ between company sizes for only the pharmaceutical companies. When the effect of one variable (company size) changes for different levels of another variable (company type), this is called an *interaction*.

Quantify the Relationship

Because this data is only a sample, the financial analyst needs to determine the following:

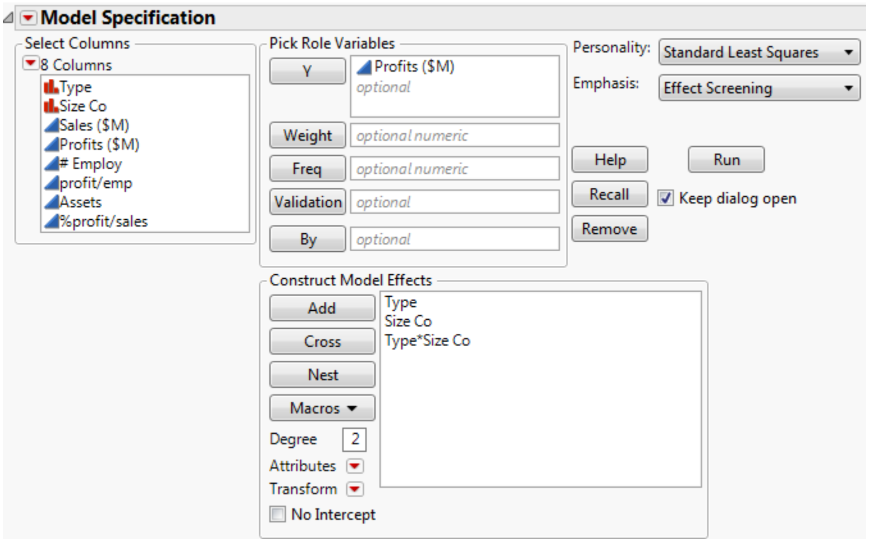
- if the differences are limited to this sample and due to chance

  or
- if the same patterns exist in the broader population

1. Return to the Companies.jmp sample data table that has the data point excluded. See "Discover the Relationship" on page 147.
2. Select **Analyze** > **Fit Model**.

3. Select **Profits ($M)** and click **Y**.
4. Select both Type and Size Co.
5. Click the **Macros** button and select **Full Factorial**.
6. From the Emphasis menu, select **Effect Screening**.
7. Select the **Keep dialog open** option.

Figure 5.23 Completed Fit Model Window



8. Click **Run**. The report window shows the model results.

To decide whether the differences in profits are real, or due to chance, examine the **Effect Tests** report.

Note: For complete details about all of the **Fit Model** results, see the Model Specification chapter in the *Fitting Linear Models* book.

View Effect Tests

The Effect Tests report (see Figure 5.24) shows the results of the statistical tests. There is a test for each of the effects included in the model on the Fit Model window: Type, Size Co, and Type*Size Co.

Figure 5.24 Effect Tests Report

Effect Tests

Source	Nparm	DF	Sum of Squares	F Ratio	Prob > F
Type	1	1	1401847.4	10.1368	0.0039*
Size Co	2	2	724616.2	2.6198	0.0927
Type*Size Co	2	2	448061.5	1.6200	0.2180

First, look at the test for the interaction in the model: the Type*Size Co effect. Figure 5.22 showed that the pharmaceutical companies appeared to have different profits between company sizes. However, the effect test indicates that there is no interaction between type and size as it relates to profit. The p-value of 0.218 is large (greater than the significance level of 0.05). Therefore, remove that effect from the model, and re-run the model.

1. Return to the Fit Model window.
2. In the Construct Model Effects box, select the **Type*Size Co** effect and click **Remove**.
3. Click **Run**.

Figure 5.25 Updated Effect Tests Report

Effect Tests

Source	Nparm	DF	Sum of Squares	F Ratio	Prob > F
Type	1	1	1,356,298	9.3768	0.0049*
Size Co	2	2	434,161.3	1.5008	0.2410

The p-value for the Size Co effect is large, indicating that there are no differences based on size in the broader population. The p-value for the Type effect is small, indicating that the differences that you saw in the data between computer and pharmaceutical companies is not due to chance.

Draw Conclusions

The financial analyst wanted to know whether the size of the company has a larger effect on the company's profits, based on type (pharmaceutical or computer).The financial analyst can now answer the question as follows:

- There is a real difference in profits between computer and pharmaceutical companies in the broader population.
- There is no correlation between the company's size and type and its profits.

Use Regression with Multiple Predictors

The section "Use Regression with One Predictor" on page 136 showed you how to build simple regression models consisting of one predictor variable and one response variable.

Multiple regression predicts the average response variable using two or more predictor variables.

Scenario

This example uses the Candy Bars.jmp data table, which contains nutrition information for candy bars.

A dietitian wants to predict calories using the following information:

- Total fat
- Carbohydrates
- Protein

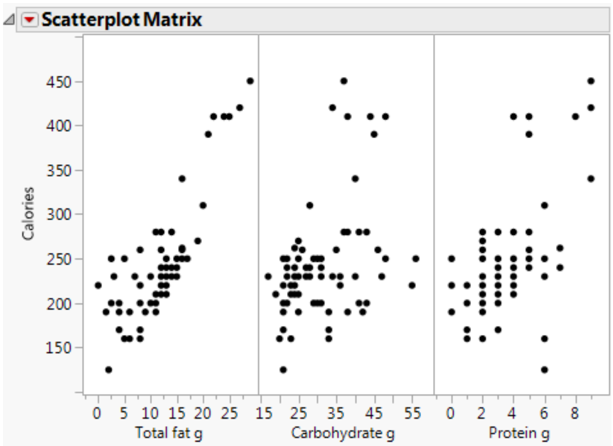
Use *multiple regression* to predict the average response variable using these three predictor variables.

Discover the Relationship

To visualize the relationship between calories and total fat, carbohydrates, and protein, create a scatterplot matrix:

1. Select **Help > Sample Data Library** and open Candy Bars.jmp.
2. Select **Graph** > **Scatterplot Matrix**.
3. Select Calories and click **Y, Columns**.
4. Select Total fat g, Carbohydrate g, and Protein g, and click **X**.
5. Click **OK**.

Figure 5.26 Scatterplot Matrix Results



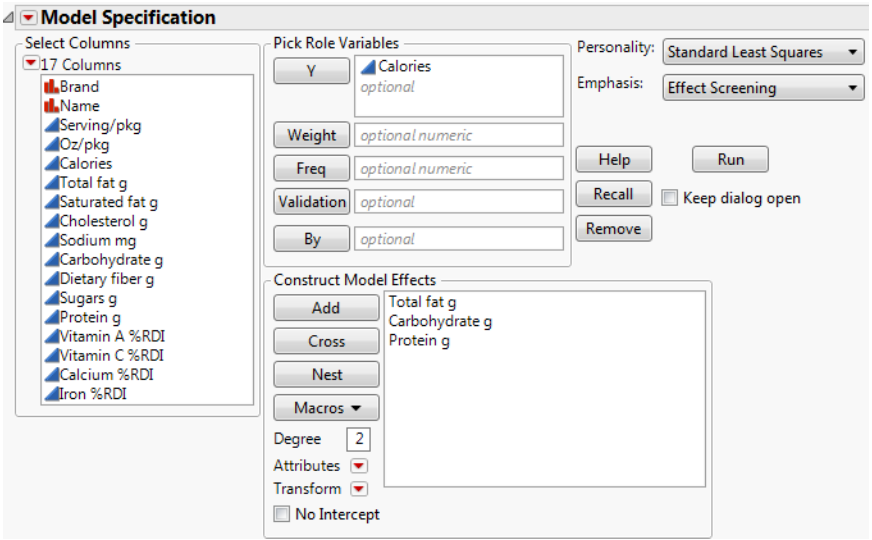
The scatterplot matrix shows that there is a positive correlation between calories and all three variables. The correlation between calories and total fat is the strongest. Now that the dietitian knows that there is a relationship, the dietitian can build a multiple regression model to predict average calories.

Build the Multiple Regression Model

Continue to use the Candy Bars.jmp sample data table.

1. Select **Analyze** > **Fit Model**.
2. Select Calories and click **Y**.
3. Select Total Fat g, Carbohydrate g, and Protein g and click **Add**.
4. Next to Emphasis, select **Effect Screening**.

Figure 5.27 Fit Model Window



5. Click **Run**.

The report window shows the model results. To interpret the model results, focus on these areas:

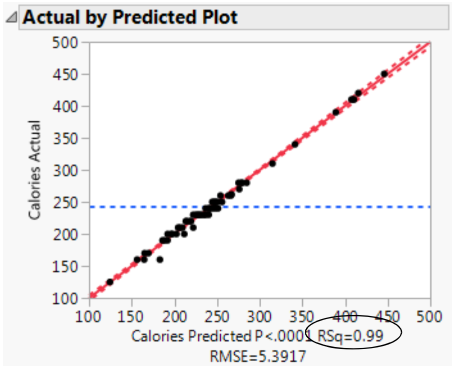
- "View the Actual by Predicted Plot" on page 153
- "Interpret the Parameter Estimates" on page 153
- "Use the Prediction Profiler" on page 154

Note: For complete details about all of the model results, see the Model Specification chapter in the *Fitting Linear Models* book.

View the Actual by Predicted Plot

The Actual by Predicted Plot shows the actual calories versus the predicted calories. As the predicted values come closer to the actual values, the points on the scatterplot fall closer around the red line. See Figure 5.28. Because the points are all very close to the line, you can see that the model predicts calories based on the chosen factors well.

Figure 5.28 Actual by Predicted Plot



Another measure of model accuracy is the RSq value (which appears below the plot in Figure 5.28). The RSq value measures the percentage of variability in calories, as explained by the model. A value closer to 1 means a model is predicting well. In this example, the RSq value is 0.99.

Interpret the Parameter Estimates

The Parameter Estimates report shows the following information:

- The model coefficients
- P-values for each parameter

Figure 5.29 Parameter Estimates Report

Model coefficients P-values

Parameter Estimates

Term	Estimate	Std Error	t Ratio	Prob>|t|
Intercept	-5.964301	2.899986	-2.06	0.0434*
Total fat g	8.9899516	0.144981	62.01	<.0001*
Carbohydrate g	4.097505	0.071025	57.69	<.0001*
Protein g	4.4013313	0.39785	11.06	<.0001*

In this example, the p-values are all very small (<.0001). This indicates that all three effects (fat, carbohydrate, and protein) contribute significantly when predicting calories.

You can use the model coefficients to predict the value of calories for particular values of fat, carbohydrate, and protein. For example, suppose that you want to predict the average calories for any candy bar that has these characteristics:

- Fat = 11 g
- Carbohydrate = 43 g
- Protein = 2 g

Using these values, you can calculate the predicted average calories as follows:

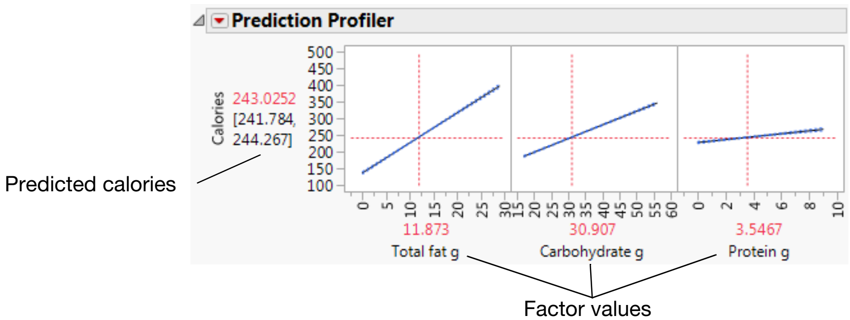
277.92 = -5.9643 + 8.99*11 + 4.0975*43 + 4.4013*2

The characteristics in this example are the same as the Milky Way candy bar (on row 59 of the data table). The actual calories for the Milky Way are 280, showing that the model predicts well.

Use the Prediction Profiler

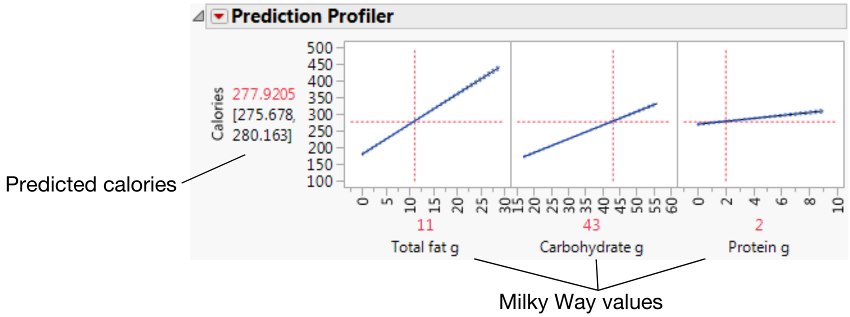
Use the Prediction Profiler to see how changes in the factors affect the predicted values. The profile lines show the magnitude of change in calories as the factor changes. The line for Total fat g is the steepest, meaning that changes in total fat have the largest effect on calories.

Figure 5.30 Prediction Profiler



Click and drag the vertical line for each factor to see how the predicted value changes. You can also click the current factor values and change them. For example, click on the factor values and type the values for the Milky Way candy bar (row 59).

Figure 5.31 Factor Values for the Milky Way



Note: For complete details about the Prediction Profiler, see the Profiler chapter in the *Profilers* book.

Draw Conclusions

The dietitian now has a good model to predict calories of a candy bar based on its total fat, carbohydrates, and protein.

Chapter 6

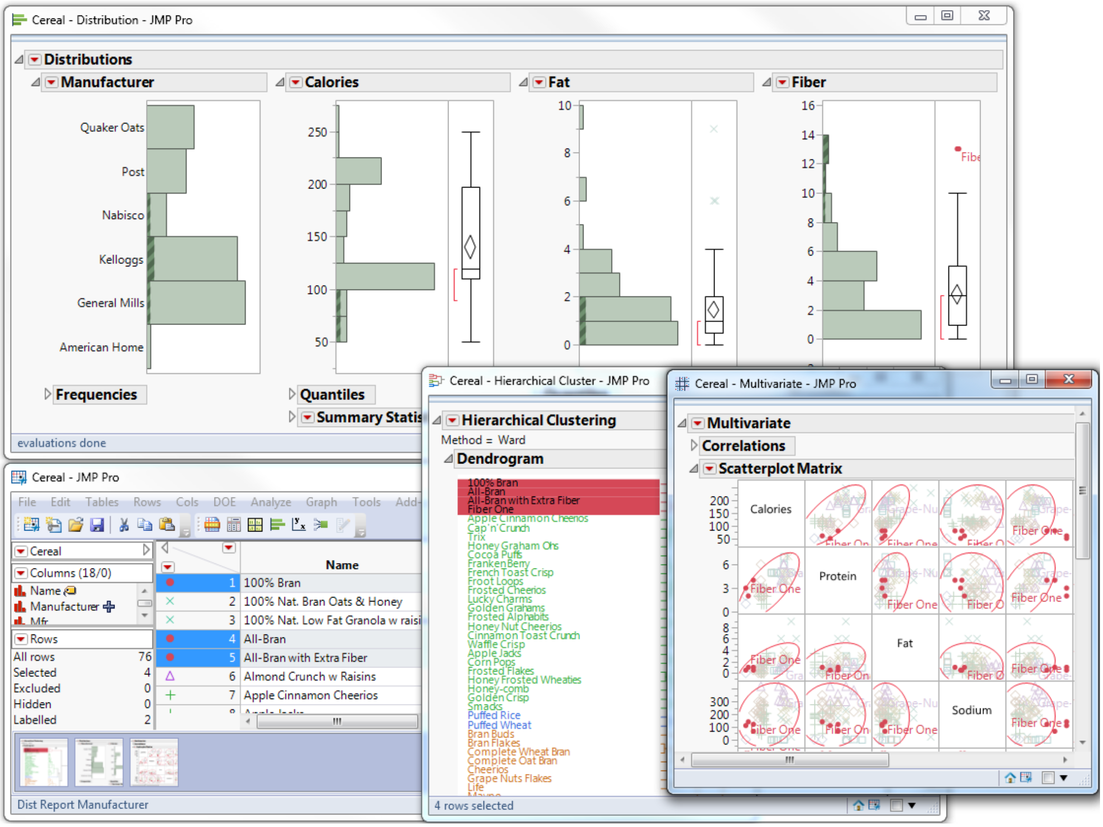
The Big Picture

Exploring Data in Multiple Platforms

JMP provides a host of statistical discovery platforms to help you explore different aspects of your data. You might start with a simple look at individual variables in histograms and then progress to multivariate and cluster analyses to get a deeper look. Each step of the way, you learn more about your data.

This chapter steps through an analysis of the Cereal.jmp sample data table that is installed with JMP. You learn how to explore the data in the Distribution, Multivariate, and Hierarchical Clustering platforms.

Figure 6.1 Multiple Analyses in JMP



Before You Begin

One of the powerful features in JMP is its linked analyses. The graphs and reports that you create are linked to each other through the data table. As shown in Figure 6.1, data that are selected in the data table are also selected in the three report windows. As you work through the examples in this chapter, keep the JMP windows open to see these interactions yourself.

Explore Data in Multiple Platforms

Which cereals are part of a healthy diet? The Cereal.jmp sample data (real data gathered from boxes of popular cereals) provides statistics on fiber content, calories, and other nutritional information. To identify the healthiest cereals, you step through interpreting histograms and descriptive statistics, correlations and outlier detection, scatterplots, and cluster analysis.

Analyze Distributions

The Distribution platform illustrates the distribution of a single variable (*univariate* analysis) using histograms, additional graphs, and reports. The word *univariate* simply means involving one variable instead of two (bivariate) or many (multivariate). However, you can examine the distribution of several individual variables within a single report. The report content for each variable changes depending on whether the variable is categorical (nominal or ordinal) or continuous.

- For categorical variables, the initial graph is a histogram. The histogram shows a bar for each level of the ordinal or nominal variable. The reports show counts and proportions.
- For continuous variables, the initial graphs show a histogram and an outlier box plot. The histogram shows a bar for grouped values of the continuous variable. The reports show selected quantiles and summary statistics.

Once you know how your data are distributed, you can plan the appropriate type of analysis going forward.

Note: For details about the Distribution platform, see the Distributions chapter in the *Basic Analysis* book.

Scenario

You want to view the nutritional values of cereals so that you can eat more healthily. Analyzing distributions of cereal data reveals answers to the following questions:

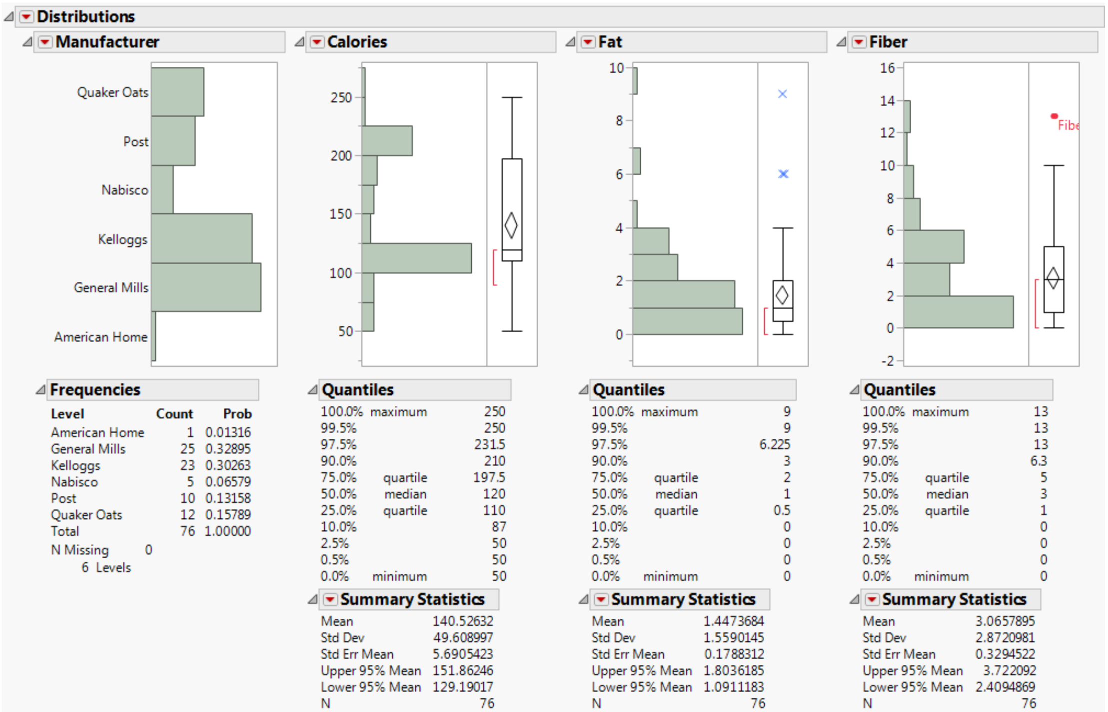
- Which cereals contain the highest amount of fiber?
- What is the average, minimum, and maximum number of calories?

- What is the average amount of fat?
- Which cereal contains the most fat?
- Are there any outliers in the data?

Create the Distributions

1. Select **Help > Sample Data Library** and open Cereal.jmp.
2. Select **Analyze > Distribution**.
3. Press Ctrl and click Manufacturer, Calories, Fat, and Fiber.
4. Click **Y, Columns** and then click **OK**.

Figure 6.2 Distributions for Manufacturer, Calories, Fat, and Fiber



In the Fiber distributions, notice the following:

– Fiber One and All-Bran with Extra Fiber contain the most fiber as shown in the Fiber box plot. These cereals are outliers in terms of fiber content.

  To display the name of the cereal next to a data point in graphs, the row that contains Fiber One in Cereal.jmp is labeled. To see the entire label, drag the right-most vertical border to the right. Place your cursor over the unlabeled data point to see "All Bran with Extra Fiber".

In the Fat distributions, notice the following:

– Place your cursor over the top data point (the x marker) in the Fat box plot to see that 100% Nat. Bran Oats & Honey is the highest in fat.

– In the Fat Quantiles report, the median amount of fat is 1 gram.

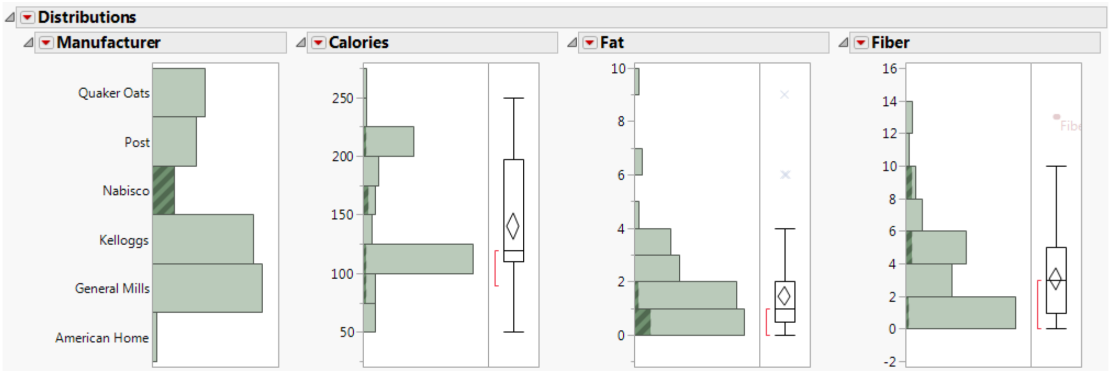
In the Calories Quantiles report, notice the following:

– The maximum number of calories is 250.

– The minimum number of calories is 50.

5. In the Manufacturer histogram, click on the bar for Nabisco.

Figure 6.3 Distributions for Nabisco Cereals

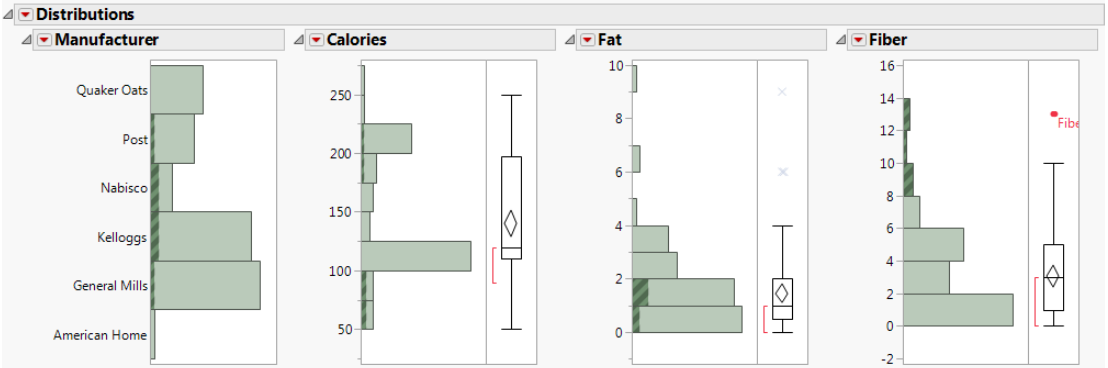


The Calories, Fat, and Fiber distributions for Nabisco cereals are highlighted in the other histograms. You can view the Calories, Fat, and Fiber distributions for the Nabisco cereals relative to the Calories, Fat, and Fiber distributions for the overall data. For example, the Fat distribution of Nabisco cereals seems to be lower than the Fat distribution for the overall data.

6. Click below the last Fiber bar to deselect all bars.

7. Press Shift and click all histogram bars in the Fiber histogram with a value above 8.

Figure 6.4 High-Fiber Cereals

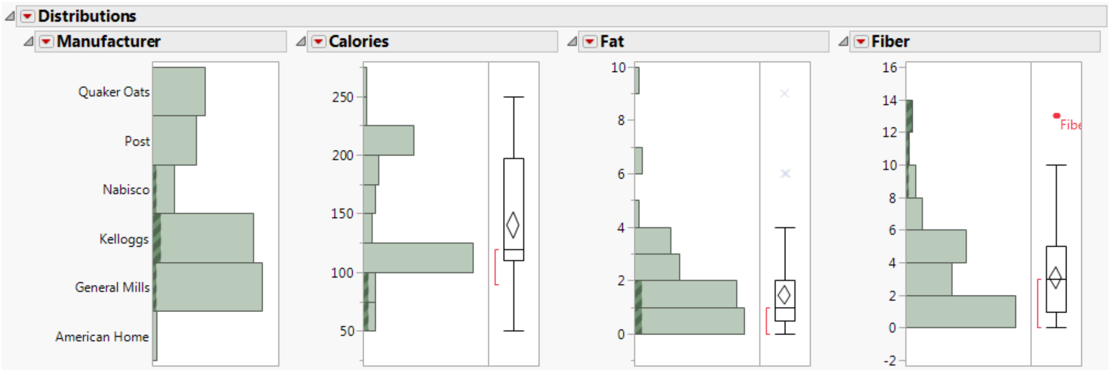


The highest-fiber cereals are highlighted in the Calories and Fat histograms. Because the histograms are linked, note that some of the high-fiber cereals are also low in fat.

8. Press Ctrl and Shift and deselect the two Calories histogram bars that are at or near 200.

   High calorie cereals are eliminated from the histograms.

Figure 6.5 High-Fiber and Low-Calorie Cereals



Tip: Leave the Distributions report open. You will use it later in a cluster analysis. See "Analyze Similar Values" on page 166.

Interpret the Results

Looking at the results, you can answer the following questions:

Which cereals are highest in fiber? The Fiber box plot shows that All-Bran with Extra Fiber and Fiber One have the highest amount of fiber. These two cereals are outliers.

What is the average, minimum, and maximum number of calories? The Calories histogram shows that the number of calories range from 50 to 275. The Calories Quantiles show that the number of calories range from 50 to 250, and the median number of calories is 120. The distribution is not uniform.

What is the average amount of fat? The Fat Quantiles report shows that the median amount of fat is 1 gram.

Which cereal contains the most fat? The Fat box plot shows that 100% Nat. Bran Oats & Honey is the highest in fat. This cereal is an outlier.

Draw Conclusions

To increase the amount of fiber in your diet, you decide to try All-Bran with Extra Fiber and Fiber One. These cereals also happen to have a low number of calories and fat. Most cereals do not greatly increase the amount of fat in your diet, but you plan to avoid the high fat 100% Nat. Bran Oats & Honey. And although most cereals are relatively low in fat, they are not necessarily low in calories.

Analyze Patterns and Relationships

Now that you have identified which cereals to eat or avoid, you want to see how the cereal variables relate to each other. The Multivariate platform enables you to observe patterns and relationships between variables. From the Multivariate report, you can do the following:

- summarize the strength of the linear relationships between each pair of response variables using the Correlations table
- identify dependencies, outliers, and clusters using the Scatterplot Matrix
- use other techniques to examine multiple variables, such as partial, inverse, and pairwise correlations, covariance matrices, and principal components

Note: For details about the Multivariate platform, see the Correlations and Multivariate Techniques chapter in the *Multivariate Methods* book.

Scenario

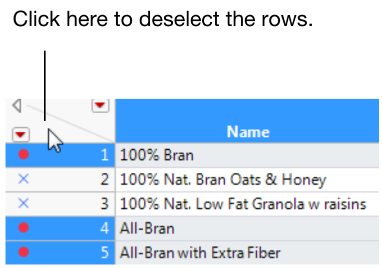
You want to see the relationships between variables such as fat and calories. Analyzing the cereal data in the Multivariate platform reveals answers to the following questions:

- Which pairs of variables are highly correlated?
- Which pairs of variables are not correlated?

Create the Multivariate Report

1. In the Cereal.jmp data table, click the bottom triangle at the top of the Columns panel to deselect the rows (Figure 6.6).

Figure 6.6 Deselecting Rows



2. Select **Analyze > Multivariate Methods > Multivariate.**
3. Select Calories through Potassium, click **Y, Columns**, and then click **OK**.

   The Multivariate report appears. The report contains the Correlations report and Scatterplot Matrix by default. The Correlations report is a matrix of correlation coefficients that summarizes the strength of the linear relationships between each pair of response (Y) variables. The dark numbers indicate a higher degree of correlation.

Figure 6.7 Correlations Report

Correlations

	Calories	Protein	Fat	Sodium	Fiber	Complex Carbos	Tot Carbo	Sugars	Calories fr Fat	Potassium
Calories	1.0000	0.7041	0.6460	0.1996	0.1953	0.6688	0.9076	0.5060	0.6709	0.4451
Protein	0.7041	1.0000	0.4080	0.0050	0.5470	0.6486	0.6937	-0.0010	0.4288	0.6782
Fat	0.6460	0.4080	1.0000	-0.0768	0.1824	0.1037	0.3860	0.4148	0.9013	0.3420
Sodium	0.1996	0.0050	-0.0768	1.0000	-0.0448	0.2619	0.3066	0.1767	0.0572	0.0459
Fiber	0.1953	0.5470	0.1824	-0.0448	1.0000	0.1769	0.3668	-0.1264	0.2553	0.8326
Complex Carbos	0.6688	0.6486	0.1037	0.2619	0.1769	1.0000	0.7773	-0.1601	0.1558	0.2693
Tot Carbo	0.9076	0.6937	0.3860	0.3066	0.3668	0.7773	1.0000	0.4263	0.4636	0.5375
Sugars	0.5060	-0.0010	0.4148	0.1767	-0.1264	-0.1601	0.4263	1.0000	0.4369	0.1166
Calories fr Fat	0.6709	0.4288	0.9013	0.0572	0.2553	0.1558	0.4636	0.4369	1.0000	0.3694
Potassium	0.4451	0.6782	0.3420	0.0459	0.8326	0.2693	0.5375	0.1166	0.3694	1.0000

   Note the following:

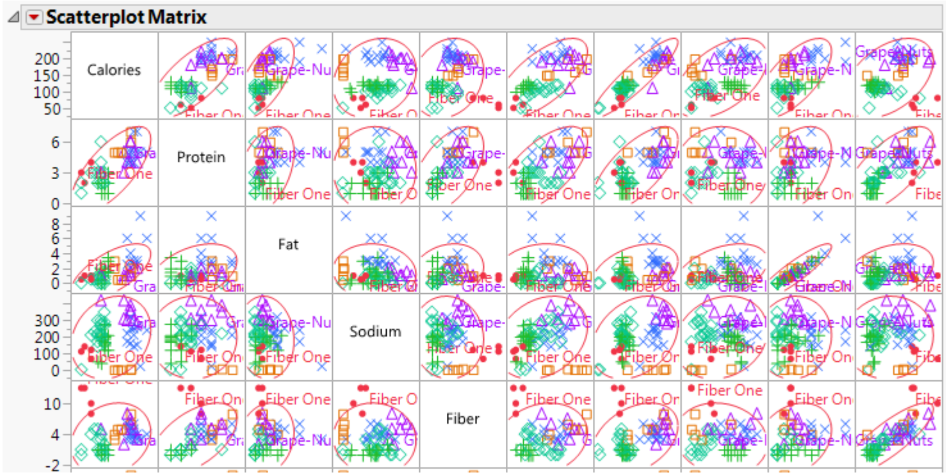
   – In the Calories column, the number of calories is highly correlated with all variables except for sodium and fiber.

   – In the Fiber column, fiber and potassium appear to be highly correlated.

   – In the Sodium column, sodium is not highly correlated with the other variables.

   The density ellipses in the Scatterplot Matrix further illustrates relationships between variables. Figure 6.8 shows a portion of the plot.

Figure 6.8 Portion of the Scatterplot Matrix



By default, a 95% bivariate normal density ellipse is in each scatterplot. Assuming that each pair of variables has a bivariate normal distribution, this ellipse encloses approximately 95% of the points. If the ellipse is fairly round and is not diagonally oriented, the variables are uncorrelated. If the ellipse is narrow and diagonally oriented, the variables are correlated.

Note the following:

– The ellipses are fairly round in the Sodium row. This shape indicates that Sodium is uncorrelated with other variables.

– The blue x markers, which represent Nat. Bran Oats & Honey, Cracklin' Oat Bran, and Banana Nut Crunch, appear outside the ellipses in the Fat row. This placement indicates that the data is an outlier (due to the amount of fat in the cereal).

You will further explore a scatterplot matrix later.

4. Select **Pairwise Correlations** from the Multivariate red triangle menu to display the Pairwise Correlations report.

Figure 6.9 Portion of the Pairwise Correlations Report

Pairwise Correlations

Variable	by Variable	Correlation	Count	Lower 95%	Upper 95%	Signif Prob	-.8 -.6 -.4 -.2 0 .2 .4 .6 .8
Protein	Calories	0.7041	76	0.5690	0.8022	<.0001*	
Fat	Calories	0.6460	76	0.4922	0.7607	<.0001*	
Fat	Protein	0.4080	76	0.2011	0.5801	0.0003*	
Sodium	Calories	0.1996	76	-0.0270	0.4068	0.0838	
Sodium	Protein	0.0050	76	-0.2207	0.2302	0.9655	
Sodium	Fat	-0.0768	76	-0.2971	0.1513	0.5099	
Fiber	Calories	0.1953	76	-0.0315	0.4030	0.0908	
Fiber	Protein	0.5470	76	0.3668	0.6877	<.0001*	
Fiber	Fat	0.1824	76	-0.0449	0.3918	0.1147	
Fiber	Sodium	-0.0448	76	-0.2675	0.1825	0.7010	

The Pairwise Correlations report lists the Pearson product-moment correlations for each pair of Y variables. The report also shows significance probabilities and compares the correlations in a bar chart.

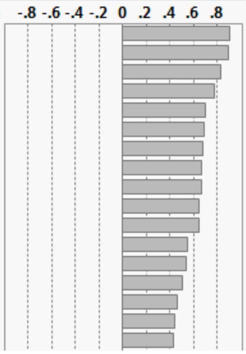
5. To quickly see which pairs are highly correlated, right-click in the report and select the **Sort by Column**, **Signif Prob**, **Ascending** checkbox, and then click **OK**.

   The most highly correlated pairs appear at the top of the report. The small *p*-values for the pairs indicate evidence of correlation. The most significant correlation is between Tot Carbo (total carbohydrates) and Calories.

Figure 6.10 Small *p*-values for Pairs

Pairwise Correlations

Variable	by Variable	Correlation	Count	Lower 95%	Upper 95%	Signif Prob
Tot Carbo	Calories	0.9076	76	0.8576	0.9406	<.0001*
Calories fr Fat	Fat	0.9013	76	0.8483	0.9365	<.0001*
Potassium	Fiber	0.8326	76	0.7474	0.8908	<.0001*
Tot Carbo	Complex Carbos	0.7773	76	0.6691	0.8532	<.0001*
Protein	Calories	0.7041	76	0.5690	0.8022	<.0001*
Tot Carbo	Protein	0.6937	76	0.5551	0.7949	<.0001*
Potassium	Protein	0.6782	76	0.5344	0.7838	<.0001*
Calories fr Fat	Calories	0.6709	76	0.5248	0.7786	<.0001*
Complex Carbos	Calories	0.6688	76	0.5221	0.7771	<.0001*
Complex Carbos	Protein	0.6486	76	0.4956	0.7625	<.0001*
Fat	Calories	0.6460	76	0.4922	0.7607	<.0001*
Fiber	Protein	0.5470	76	0.3668	0.6877	<.0001*
Potassium	Tot Carbo	0.5375	76	0.3551	0.6805	<.0001*
Sugars	Calories	0.5060	76	0.3167	0.6566	<.0001*
Calories fr Fat	Tot Carbo	0.4636	76	0.2659	0.6238	<.0001*
Potassium	Calories	0.4451	76	0.2442	0.6094	<.0001*
Calories fr Fat	Sugars	0.4369	76	0.2345	0.6029	<.0001*



Interpret the Results

Looking at the results, you can answer the following questions:

Which pairs of variables are highly correlated? The Correlations report and Scatterplot Matrix show that the number of calories is highly correlated with all variables except for sodium and fiber. The Pairwise Correlations report shows that Tot Carbo (total carbohydrates) and Calories is the most correlated pair of variables.

Which pairs of variables are not correlated? The Correlations report and Scatterplot Matrix show that Sodium is not correlated with the other variables.

Draw Conclusions

You confirm the previous decision to avoid the high fat 100% Nat. Bran Oats & Honey. Trying All-Bran with Extra Fiber and Fiber One was also a smart decision. These two high-fiber cereals have the added benefit of contributing a lower number of calories, fat, and sugars and a higher amount of potassium. You also decide to avoid high-carbohydrate cereals because they likely contain a large number of calories.

Analyze Similar Values

Clustering is a multivariate technique that groups observations together that share similar values across a number of variables. Hierarchical clustering combines rows in a hierarchical sequence that is portrayed as a tree. Cereals with certain characteristics, such as high-fiber cereals, are grouped in clusters so that you can view similarities among cereals.

Note: For details about hierarchical clustering, see the Hierarchical Cluster chapter in the *Multivariate Methods* book.

Scenario

You want to know which cereals are similar to each other and which ones are dissimilar. Analyzing clusters of cereal data reveals answers to the following questions:

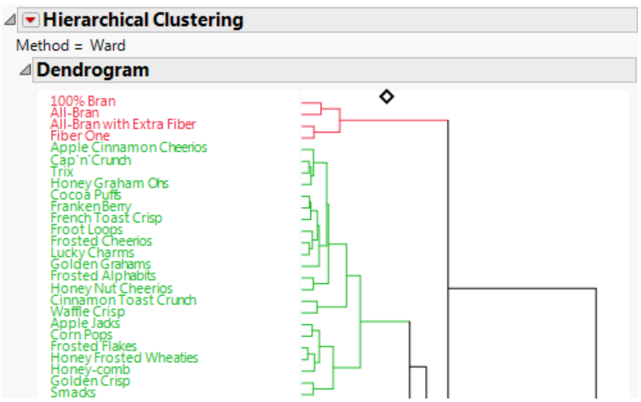
- Which cluster of cereals provides little nutritional value?
- Which cluster of cereals is high in vitamins and minerals and contains a low amount of sugar and fat?
- Which cluster of cereals contains high fiber and low calories?

Create the Hierarchical Cluster Graph

1. With Cereal.jmp displayed, select **Analyze > Clustering > Hierarchical Cluster**.
2. Select Calories through Enriched, click **Y, Columns**, and then click **OK**.

   The Hierarchical Clustering report appears. Figure 6.11 shows a portion of the report. The clusters are colored according to the data table row states.

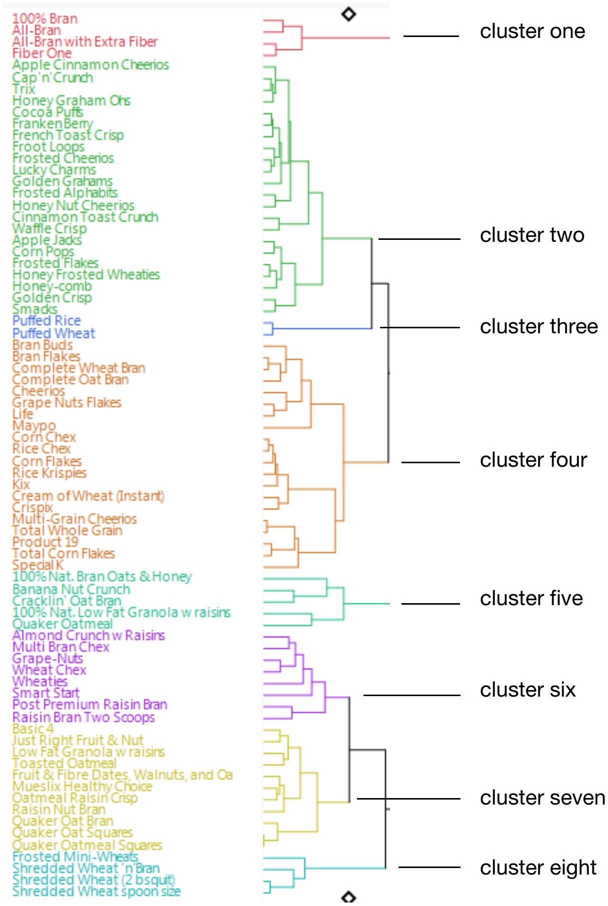
Figure 6.11 Portion of the Hierarchical Clustering Report



3. Select **Color Clusters** from the Hierarchical Clustering red triangle menu.

The clusters are colored according to their relationships in the dendrogram.

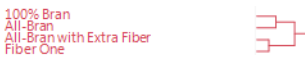
Figure 6.12 Colored Clusters



The cereals have similar characteristics within each cluster. For example, judging by the names of the cereals in cluster one, you guess that the cereals are high in fiber.

Notice how All-Bran with Extra Fiber and Fiber One are grouped in cluster one. These cereals are more similar to each other than the other two cereals in the cluster.

Figure 6.13 Similar Cereals in Cluster One



4. To select cluster one, click on the red horizontal line on the right.

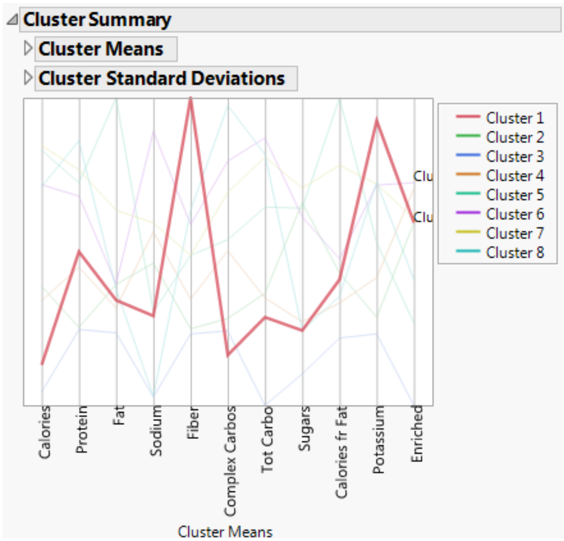
   The four cereals are highlighted in red.

Figure 6.14 Selecting a Cluster

5. To see the similar characteristics in the cluster, select **Cluster Summary** from the red triangle menu.

   The Cluster Summary graph at the bottom of the report shows the mean value of each variable across each cluster. For example, the cereals in this cluster contain more fiber and potassium than cereals in other clusters.

Figure 6.15 Cluster Summary

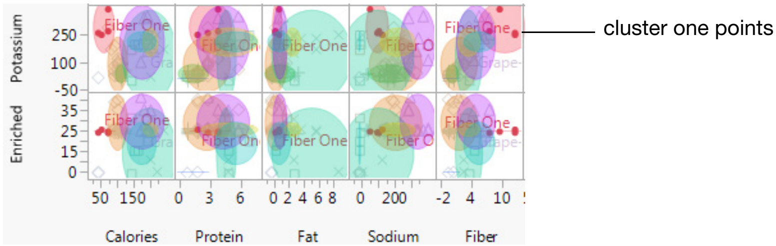


6. Select Scatterplot Matrix from the red triangle menu.

   This option is an alternative to creating a scatterplot matrix in the Multivariate platform.

   Note the Fiber plot in the Potassium row. The selected cereals are located on the right side of the plot between 8 and 13 grams. This location indicates that the cereals in cluster one are high in fiber and potassium.

Figure 6.16 Cluster One Characteristics



Note: The points are also selected in the previous scatterplot matrix that you created if it is still open.

Interpret the Results

Clicking through the clusters and looking at the Cluster Summary report, you can see the following characteristics:

- Cluster one cereals, such as Fiber One and All-Bran, contain high fiber and potassium and low calories.
- Cluster two, which contains many favorite children's cereals, is high in sugar and low in fiber, complex carbohydrates, and protein.
- Cluster three cereals (Puffed Rice and Puffed Wheat) are low in calories but provide little nutritional value.
- Cluster four cereals, such as Total Corn Flakes and Multi-Grain Cheerios, provide 100% of your daily requirement of vitamins and minerals. They are low in fat, fiber, and sugar.
- Cluster five cereals are high in protein and fat and low in sodium. The cluster consists of cereals such as Banana Nut Crunch and Quaker Oatmeal.
- Cluster six cereals are low in fat and high in sodium and carbohydrates. Traditional cereals such as Wheaties and Grape-Nuts are in this cluster.
- Cluster seven cereals are high in calories and low in fiber. Many cereals that include dried fruit are in this cluster (Mueslix Healthy Choice, Low Fat Granola w Raisins, Oatmeal Raisin Crisp, Raisin Nut Bran, and Just Right Fruit & Nut).
- Cluster eight cereals are low in sodium and sugar, and high in complex carbohydrates, protein, and potassium. Shredded Wheat and Mini-Wheat cereals are in this cluster.

By looking at the joins in the dendrogram, you can see which cereals in each cluster are most similar.

- In cluster one, Fiber One is similar in nutritional value to All-Bran with Extra Fiber. 100% Bran and All-Bran are also similar. Each pair of similar cereals are made by different companies, so the cereals are competing against each other.

- In cluster two, Frosted Flakes and Honey Frosted Wheaties are similar even though one is a corn flake and the other is a wheat flake. Lucky Charms and Frosted Cheerios are similar. Cap'n'Crunch and Trix are also similar.

Draw Conclusions

Based on your desire to eat more fiber and fewer calories, you decide to try the cereals in cluster one. You will avoid cereals in cluster three, which consists of puffed wheat and puffed rice and have little nutritional value. And you will try cereals in the highly nutritious cluster four.

Chapter 7

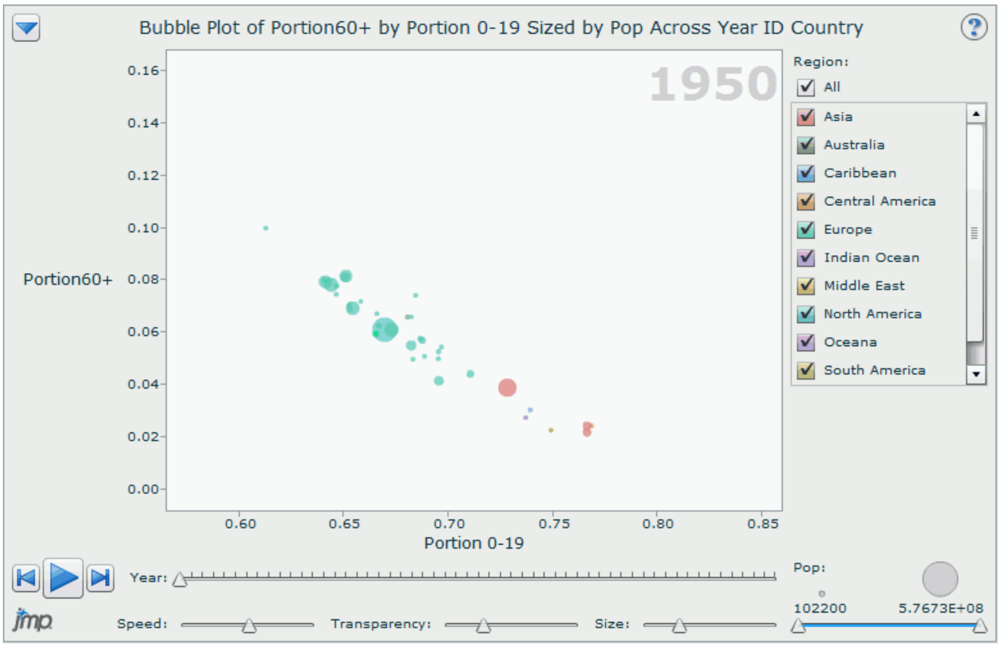
Save and Share Your Work

Save and Re-create Results

Once you have generated results from your data, JMP provides you with multiple ways to share your work with others. Here are some of the ways that you can share your work:

- Saving platform results as journals or projects
- Saving results, data tables, and other files in projects
- Saving scripts to reproduce results in data tables
- Creating Adobe Flash (SWF) versions of platform results
- Saving results as Interactive HTML (.htm, html)
- Saving results as a PowerPoint presentation (.pptx)
- Sharing results in a dashboard

Figure 7.1 Example of an Adobe Flash (SWF) Bubble Plot



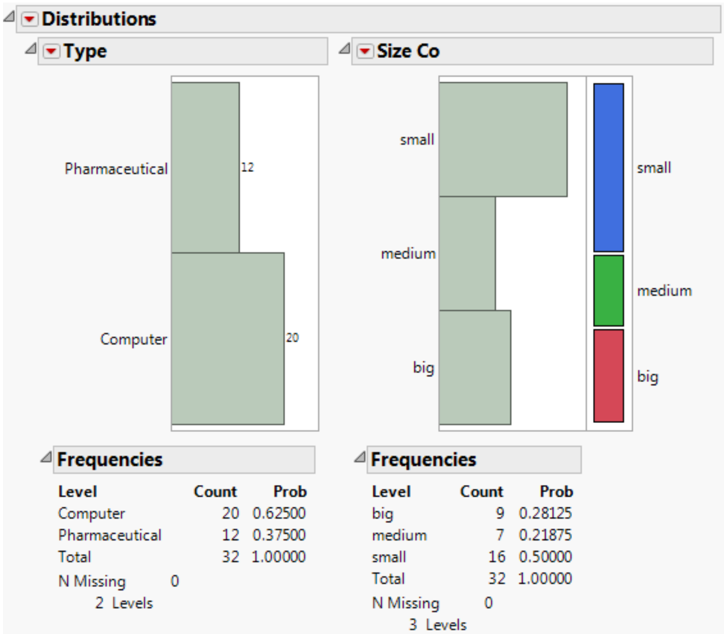
Save Platform Results in Journals

Save platform reports for future viewing by creating a journal of the report window. The journal is a copy of the report window. You can edit or append additional reports to an existing journal. The journal is not connected to the data table. A journal is an easy way to save the results from several report windows in a single report window that you can share with others.

Example of Creating a Journal

1. Select **Help > Sample Data Library** and open Companies.jmp.
2. Select **Analyze** > **Distribution**.
3. Select both Type and Size Co and click **Y, Columns**.
4. Click **OK**.
5. From the red triangle menu for **Type**, select **Histogram Options** > **Show Counts**.
6. From the red triangle menu for **Size Co**, select **Mosaic Plot**.
7. Select **Edit** > **Journal** to journal these results. The results are duplicated in a journal window.

Figure 7.2 Journal of Distribution Results



The results in the journal are not connected to the data table. In the Type bar chart, if you click the Computer bar, no rows are selected in the data table.

Since the journal is a copy of your results, most of the red triangle menus do not exist. A journal does have a red triangle menu for each new report that you add to the journal. This menu has two options:

Rerun in new window If you have the original data table that was used to create the original report, this option runs the analysis again. The result is a new report window.

Edit Script This option opens a script window that contains a JSL script to re-create the analysis. JSL is a more advanced topic that is covered in the *Scripting Guide* and *JSL Syntax Reference*.

Add Analyses to a Journal

If you perform another analysis, you can add the results of the analysis to the existing journal.

1. With a journal open, select **Analyze** > **Distribution**.
2. Select profit/emp and click **Y, Columns**.
3. Click **OK**.
4. Select **Edit** > **Journal**. The results are appended to the bottom of the journal.

Create Projects

Save multiple JMP file types (such as data tables, reports, journals, and scripts) in a single file by creating a project. The project file contains all the information needed to re-open all of the included files.

Example of Creating a Project

Create a Report

1. Select **Help > Sample Data Library** and open Companies.jmp.
2. Select **Analyze** > **Distribution**.
3. Select Profits ($M) and profit/emp, and click Y, Columns.
4. Click **OK**.

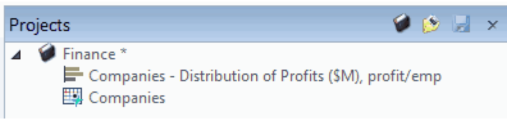
Create a Project and Add the Data Table and the Report to It

1. To start a new project, select **File** > **New** > **Project**. A window appears that shows the untitled project.

Figure 7.3 Initial Project Window on Windows

2. Right-click the project (Untitled) and select **Rename**, and then enter a new name (Finance).
3. To add the Distribution results to the project, right-click on the project name and select **Add Window**.
4. In the Add Windows to Project window, select the Distribution results.
5. Click **OK**. The Distribution results are added to the project.
6. To add the Companies.jmp data table to the project, repeat step 3 and select the Companies data table from the window.
7. Click **OK**. The data table is added to the project.

Figure 7.4 The Final Project



You can double-click on the links in the project to open the data table and re-create the Distribution results.

Save and Run Scripts

Most platform options in JMP are scriptable, meaning that most actions that you perform can be saved as a JMP Scripting Language (JSL) script. You can use a script to reproduce your actions or results at any time.

Example of Saving and Running a Script

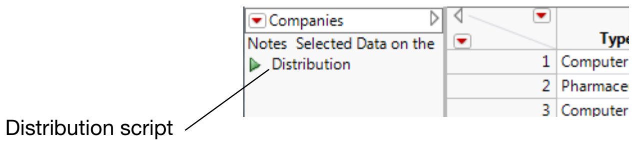
Create a Report

1. Select **Help > Sample Data Library** and open Companies.jmp.
2. Select **Analyze** > **Distribution**.
3. Select Type and profit/emp and click **Y, Columns**.
4. Click **OK**.
5. From the red triangle menu for Type, select these options:
   – **Histogram Options** > **Show Counts**
   – **Confidence Interval** > **0.95**
6. From the red triangle menu for profit/emp, select these options:
   – **Outlier Box Plot**, to remove the outlier box plot
   – **CDF Plot**
7. From the red triangle menu for Distributions, select **Stack**.

Save the Script to the Data Table and Run It

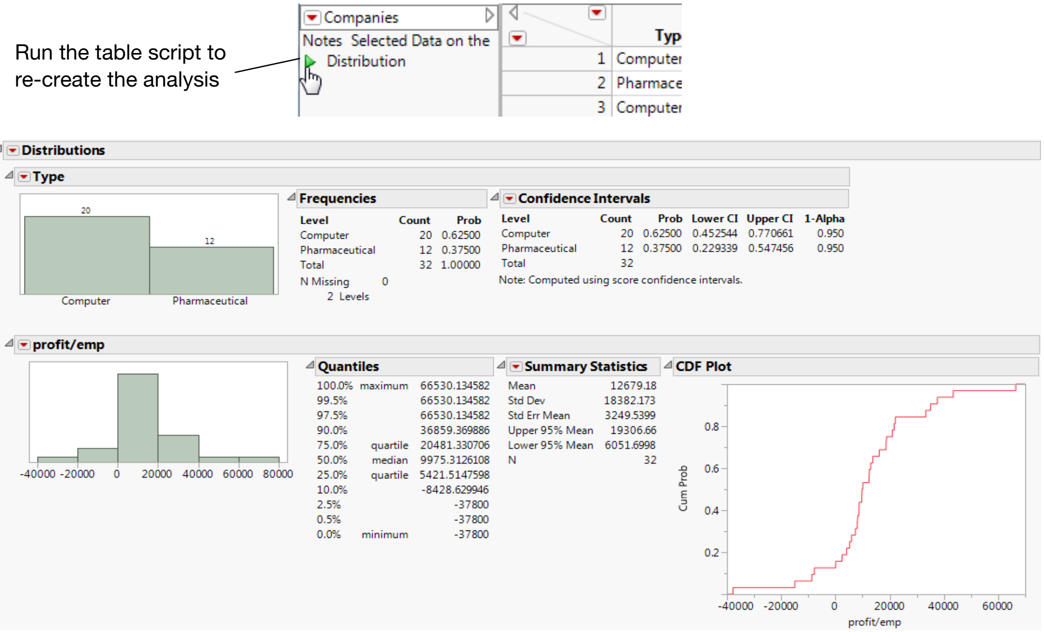
1. To save this analysis, select **Save Script > To Data Table** from the red triangle menu for Distributions. The Distribution script appears in the Table panel.

Figure 7.5 Distribution Script



2. Close the Distribution report window.
3. To re-create the analysis, click the green triangle next to the **Distribution** script.

Figure 7.6 Running the Distribution Script



Tip: Right-click the table script to view more options.

About Scripts and JSL

The script that you saved in this section contains JMP Scripting Language (JSL) commands. JSL is a more advanced topic that is covered in the *Scripting Guide* and *JSL Syntax Reference*.

Save Reports as Interactive HTML

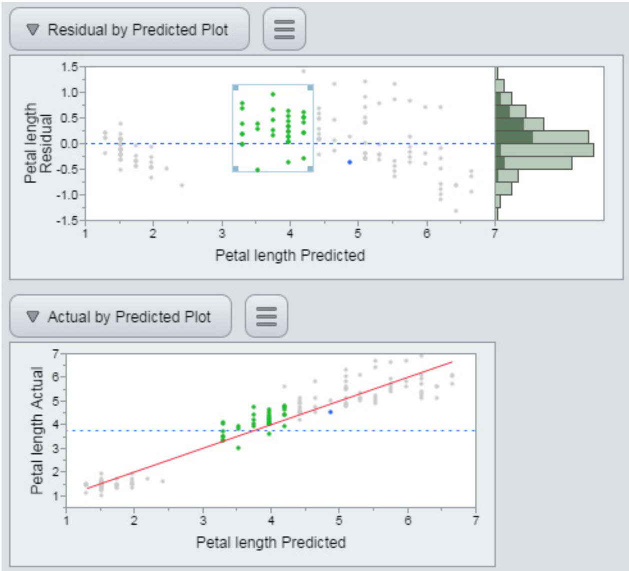
Interactive HTML enables JMP users to share reports that contain dynamic graphs so that even non JMP users can explore the data. The JMP report is saved as a web page in HTML 5 format, which you can email to users or publish on a website. Users then explore the data as they would in JMP.

Interactive HTML provides a subset of features from JMP:

- Explore interactive graph features, such as selecting histogram bars and viewing data values.
- View data by brushing.
- Show or hide report sections.

- Place your cursor over the report for tooltips.
- Increase the marker size.

Figure 7.7 Brushing Data in Interactive HTML



Many changes that you make to the graphs, such as ordered variables, horizontal histograms, background colors, and colored data points, are saved in the web page. Graphs and tables that are closed when you save the content remain closed on the web page until the user opens them.

Interactive HTML Contains Data

When you save reports as interactive HTML in JMP, your data are embedded in the HTML. The content is unencrypted, because web browsers cannot read encrypted data. To avoid sharing sensitive data, save your results as a non-interactive web page. (Select **File > Save As > HTML File** on Windows, or **File > Export > HTML** on Macintosh.)

Example of Creating Interactive HTML

Create a Report

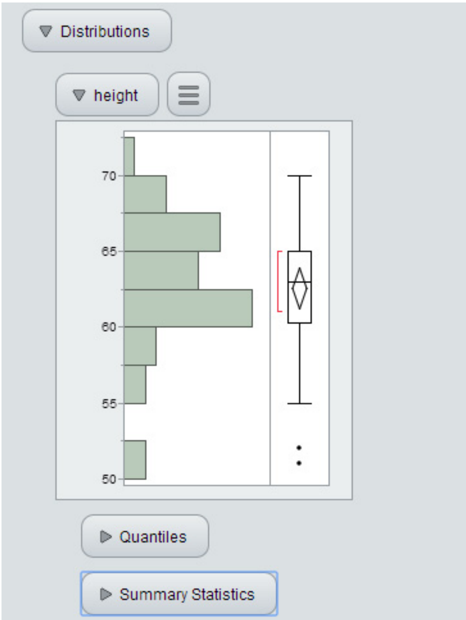
1. Select **Help > Sample Data Library** and open Big Class.jmp.
2. Select **Analyze > Distribution**.
3. Select height and click **Y, Columns**.
4. Click **OK**.

Save as Interactive HTML

1. On Windows, select **File > Save As** and select **Interactive HTML with Data** from the Save as type list. On Macintosh, select **File > Export > Interactive HTML with Data**.
2. Name and save the file (or export on Macintosh).

   The output appears in your default browser.

Figure 7.8 Interactive HTML Output



For information about exploring interactive HTML output, visit
http://www.jmp.com/support/help/InteractiveHTML/13/ShareJMPReports.shtml.

Save a Report as a PowerPoint Presentation

Create a presentation by saving JMP results as a Microsoft PowerPoint presentation (.pptx). Rearrange JMP content and edit text in PowerPoint after saving as a .pptx file. Sections of a JMP report are exported into PowerPoint differently.

- Report headings are exported as editable text boxes.
- Graphs are exported as images. Certain graphical elements, such as legends, are exported as separate images. Images resize to fit the slide in PowerPoint.

Use the selection tool to select the sections that you want to save in your presentation. Delete unwanted content once after you open the file in PowerPoint.

Note: On Windows, PowerPoint 2007 is the minimum version required to open .pptx files created in JMP. On Macintosh, at least PowerPoint 2011 is required.

1. In JMP, create the report.
2. (Windows) Select **File > Save As** and select **PowerPoint Presentation** from the Save as type list.
3. (Macintosh) Select **File > Export > PowerPoint Presentation** and click **Next**.
4. Select a graphic file format from the list.

   On Windows, EMF is the default format. On Macintosh, PDF is the default format.
5. Name and save the file (or export on Macintosh).

   The file opens in Microsoft PowerPoint because **Open presentation after save** is selected by default.

Note: The native EMF graphics produced on Windows are not supported on Macintosh. The native PDF graphics produced on Macintosh are not supported on Windows. For cross-platform compatibility, change the default graphics file format by selecting **File > Preferences > General**. Then, change the **Image Format for PowerPoint** to either PNG or JPEG.

Create Adobe Flash Versions of the Profiler, Bubble Plot, or Distribution Platforms

To share the interactive results of a profiler, bubble plot, or distribution outside of JMP, you can export an SWF file and use the Adobe Flash Player to view it. You can import the SWF file into presentations and applications. You can also save the results as an HTML page with the SWF output embedded.

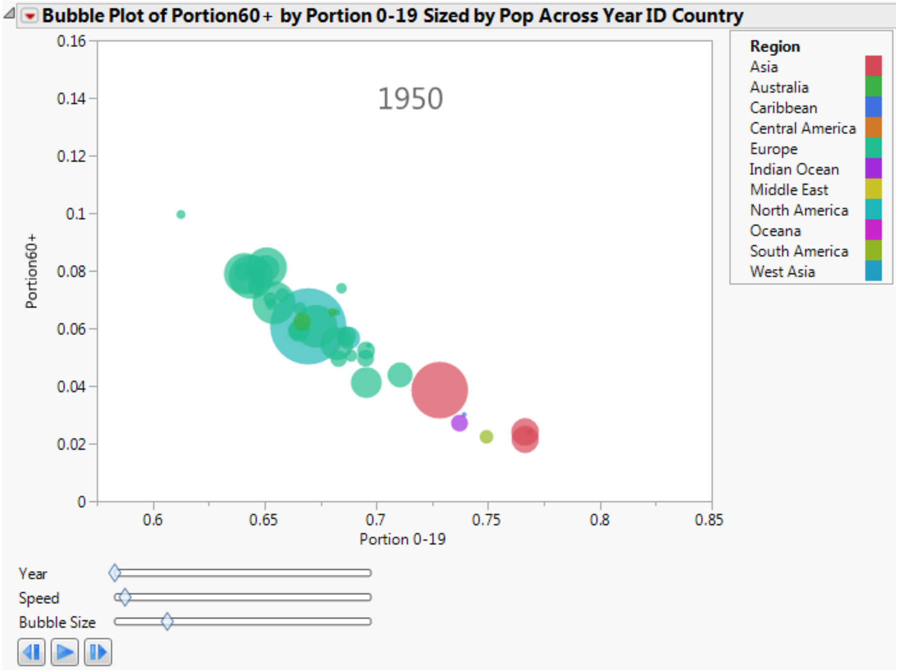
Example of Saving an Adobe Flash Version of a Bubble Plot

Create a Bubble Plot in JMP

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 Country and click **ID**.

6. Select Year and click **Time**.
7. Select Pop and click **Sizes**.
8. Select Region and click **Coloring**.
9. Click **OK**.

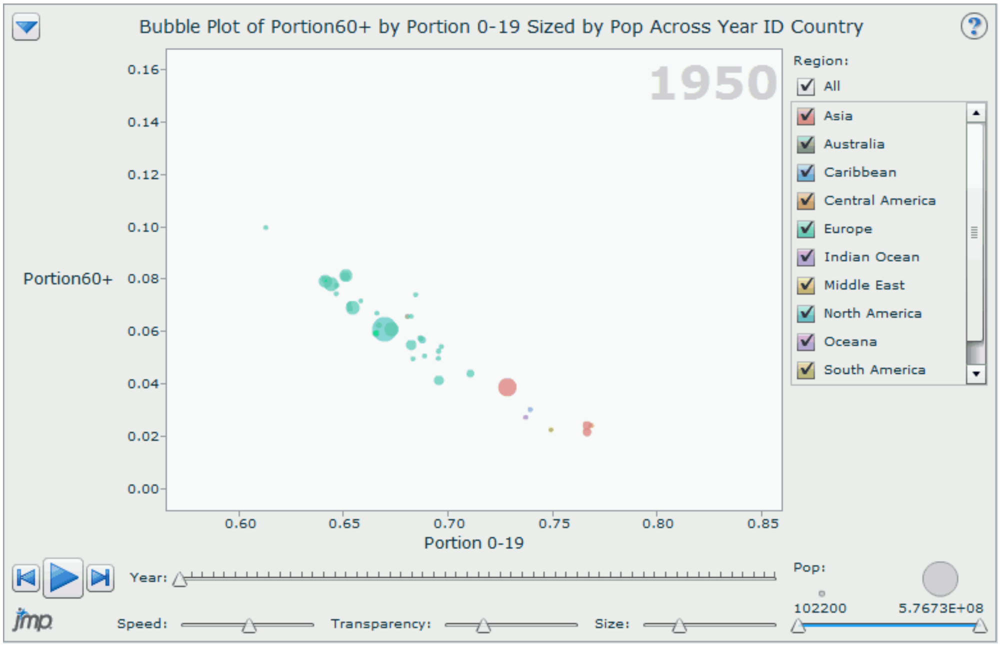
Figure 7.9 Initial Bubble Plot



Save the Bubble Plot as an SWF File

1. Select **Save for Adobe Flash Platform (SWF)** from the red triangle menu.
2. In the Save As SWF window, select the location to which you want to save the file.
3. Click **Save**. The Adobe Flash version of the Bubble Plot is saved as HTML and appears in a web browser.

Figure 7.10 Adobe Flash Version of Bubble Plot



Additional Information

The JMP website has additional information about the following:

- more details about Adobe Flash versions of the Profiler, Bubble Plot, and Distribution platforms
- instructions on importing the Adobe Flash versions into Microsoft PowerPoint

Visit www.jmp.com/support/swfhelp/en/ for details.

Create Dashboards

A dashboard is a visual tool that lets you run and present reports on a regular basis. You can show reports, data filters, selection filters, data tables, and graphics on a dashboard. The content shown on the dashboard is updated when you open the dashboard.

Example of Combining Windows

You can quickly create dashboards by merging several open windows in JMP. Combining windows provides options to view a summary of statistics and include a selection filter.

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

2. Run the Distribution and Bivariate table scripts.
3. Select **Window > Combine Windows**.

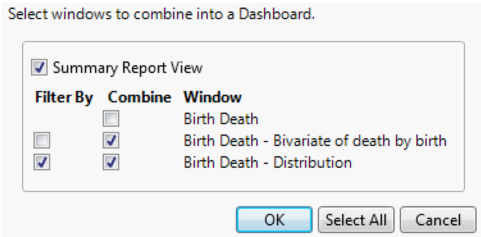
   The Combine Windows window appears.

   Tip: On Windows, you can also select Combine Windows from the Arrange Menu option in the lower right corner of JMP windows.

   ---

4. Select **Summary Report View** to display the graphs and omit the statistical reports
5. In the Combine column, select **Birth Death - Bivariate of death by birth** and **Birth Death - Distribution.**
6. In the Filter By column, select **Birth Death - Distribution**.

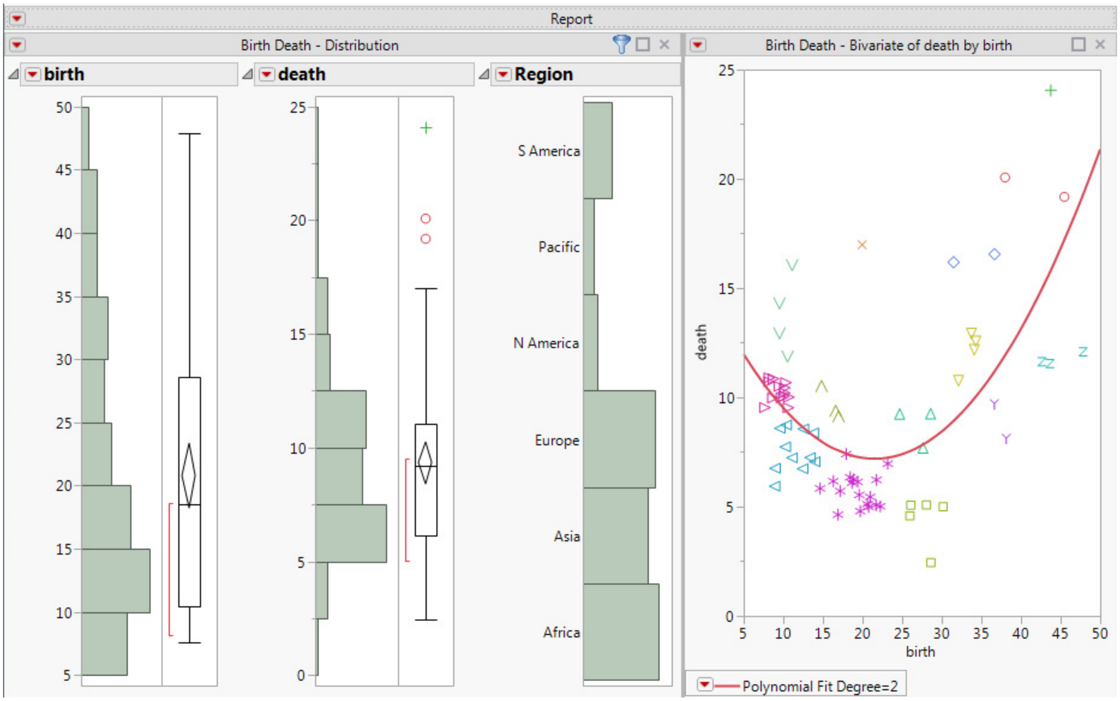
Figure 7.11 Combine Windows Options



7. Click **OK**.

   The two reports are combined into one window. Notice the filter icon  at the top of the Distribution report. When you select a bar in one of the histograms, the corresponding data in the Bivariate graph are selected.

Figure 7.12 Combined Windows



Example of Creating a Dashboard with Two Reports

Suppose that you created two reports and want to run the reports again the next day against an updated set of data. This example shows how to create a dashboard from the reports in Dashboard Builder.

1. Select **Help > Sample Data Library** and open Hollywood Movies.jmp.
2. Run the table scripts named "Distribution: Profitability by Lead Studio and Genre" and "Graph Builder: World and Domestic Gross by Genre".
3. From any window, select **File > New > Dashboard**.

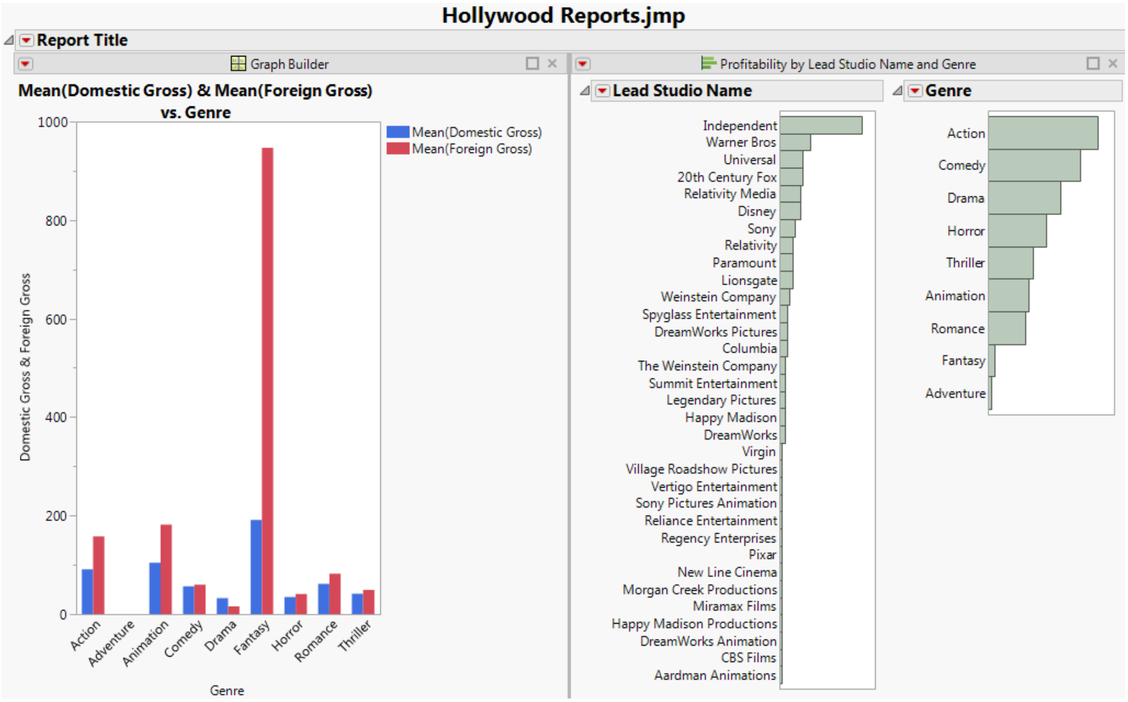
   Templates for common layouts appear.
4. Select the **2x1 Dashboard** template.

   A box with room for two reports appears on the workspace.
5. In the Reports list, double-click the report thumbnails to put them on the dashboard.
6. Select **Preview Mode** from the Dashboard Builder red triangle menu.

   A preview of the dashboard appears. Notice that the graphs are linked to each other and the data table. They also have the same red triangle options as the Distribution and Graph Builder platforms.

7. Click **Close Preview**.

Figure 7.13 Dashboard with Two Reports



For more information about creating dashboards, see the Extend JMP chapter in the *Using JMP* book.

Save the Dashboard as an Add-In

A JMP add-in is a JSL script that you can run anytime from the JMP Add-Ins menu. Saving a dashboard as an add-in enables you to share a dashboard with another JMP user.

After creating the dashboard as shown in "Example of Creating a Dashboard with Two Reports" on page 183, follow these steps to save it as an add-in.

1. From the Dashboard Builder red triangle menu, select **Save Script > To Add-In**.

   Add-In Builder opens.

2. Next to Add-In name, type Hollywood Movies Dashboard (the add-in filename).
3. Click the **Menu Items** tab and type Hollywood Movies Dashboard next to **Menu item name** (the name of the add-in menu item).
4. Click **Save** and save the add-in to your desktop.

   The add-in is saved and installed in your Add-Ins menu.

5. Click **Close** on the Add-In Builder window.
6. From the JMP main menu, select **Add-Ins** and select **Hollywood Movies Dashboard**.

   The Graph Builder and Distribution reports are generated from Hollywood Movies.jmp.

For more information about Add-In Builder, see the Creating Applications chapter in the *Scripting Guide*.

Chapter 8

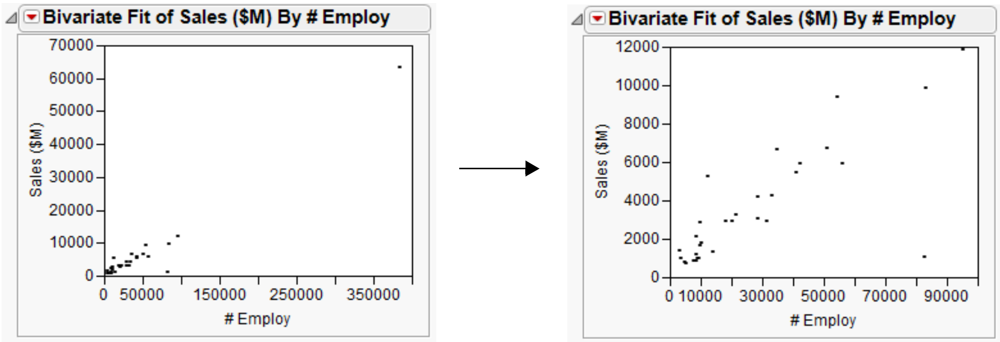
Special Features

Automatic Analysis Updates and SAS Integration

Using some of the special features in JMP, you can do the following:

- Update analyses or graphs automatically
- Customize platform results
- Integrate with SAS to use advanced analytical features

Figure 8.1 Examples of Special Features




```
DATA Candy_Bars; INPUT  Calories Total_fat_g Carbohydrate_g Protein_g; Lines;
310  20  28  6
230  12  27  4
220  12  24  3
170  8  21  3
200  2.5  43  1
260  16  26  5
190  1.5  42  2
190  11  21  2
230  12  28  3
;
RUN;

PROC GLM  DATA=Candy_Bars ALPHA=0.05;
MODEL  Calories =  Total_fat_g Carbohydrate_g Protein_g;
RUN;
```

Automatically Update Analyses and Graphs

When you make a change to a data table, you can use the Automatic Recalc feature to automatically update analyses and graphs that are associated with the data table. For example, if you exclude, include, or delete values in the data table, that change is instantly reflected in the associated analyses or graphs. Note the following information:

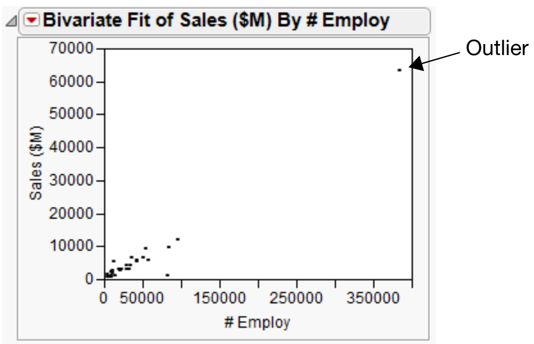
- Some platforms do not support Automatic Recalc. For more information, see the JMP Reports chapter in the *Using JMP* book.
- For the supported platforms in the **Analyze** menu, Automatic Recalc is turned off by default. However, for the supported platforms in the **Quality and Process** menu, Automatic Recalc is turned on by default, except for the Variability/Attribute Gauge Chart, Capability, and Control Chart.
- For the supported platforms in the **Graph** menu, Automatic Recalc is turned on by default.

Example of Using Automatic Recalc

This example uses the Companies.jmp sample data table, which contains financial data for 32 companies from the pharmaceutical and computer industries.

1. Select **Help > Sample Data Library** and open Companies.jmp.
2. Select **Analyze** > **Fit Y by X**.
3. Select Sales ($M) and click **Y, Response**.
4. Select # Employ and click **X, Factor**.
5. Click **OK**.

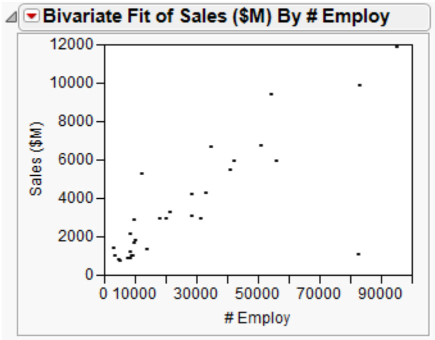
Figure 8.2 Initial Scatterplot



The initial scatterplot shows that one company has significantly more employees and sales than the other companies. You decide that this company is an outlier, and you want to exclude that point. Before you exclude the point, turn on Automatic Recalc so that your scatterplot is updated automatically when you make the change.

6. Turn on Automatic Recalc by selecting **Redo** > **Automatic Recalc** from the red triangle menu.
7. Click on the outlier to select it.
8. Select **Rows** > **Exclude/Unexclude**. The point is excluded from the analysis and the scatterplot is automatically updated.

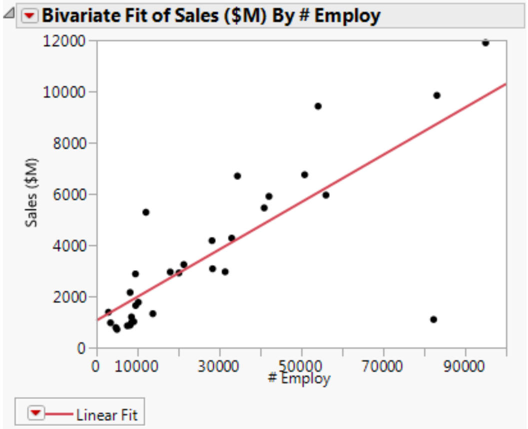
Figure 8.3 Updated Scatterplot



If you fit a regression line to the data, the point in the lower right corner is an outlier, and influences the slope of the line. If you then exclude the outlier with Automatic Recalc turned on, you can see the slope of the line change.

9. Fit a regression line by selecting **Fit Line** from the red triangle menu. Figure 8.4 shows the regression line and analysis results added to the report window.

Figure 8.4 Regression Line and Analysis Results



Linear Fit

Sales ($M) = 1059.6771 + 0.0923433*# Employ

Summary of Fit

RSquare	0.618056
RSquare Adj	0.604885
Root Mean Square Error	1866.153
Mean of Response	3562.755
Observations (or Sum Wgts)	31

Lack Of Fit

Analysis of Variance

Source	DF	Sum of Squares	Mean Square	F Ratio
Model	1	163425659	163425659	46.9273
Error	29	100993305	3482527.8	**Prob > F**
C. Total	30	264418964		<.0001*

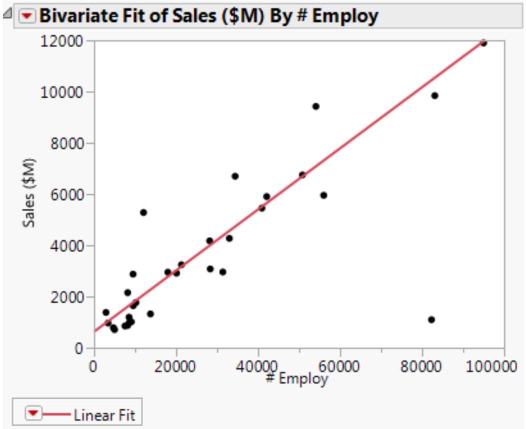
Parameter Estimates

Term	Estimate	Std Error	t Ratio	Prob>|t|
Intercept	1059.6771	495.8351	2.14	0.0411*
# Employ	0.0923433	0.01348	6.85	<.0001*

10. Click on the outlier to select it.
11. Select **Rows** > **Exclude/Unexclude**. The regression line and analysis results are automatically updated, reflecting the exclusion of the point.

Tip: When you exclude a point, the analyses are recalculated without the data point, but the data point is not hidden in the scatterplot. To also hide the point in the scatterplot, select the point, and then select **Rows** > **Hide and Exclude**.

Figure 8.5 Updated Regression Line and Analysis Results



Linear Fit

Sales ($M) = 631.36791 + 0.1192722*# Employ

Summary of Fit

RSquare	0.882703
RSquare Adj	0.878514
Root Mean Square Error	1039.891
Mean of Response	3644.95
Observations (or Sum Wgts)	30

Lack Of Fit

Analysis of Variance

Source	DF	Sum of Squares	Mean Square	F Ratio
Model	1	227857403	227857403	210.7112
Error	28	30278440	1081372.9	**Prob > F**
C. Total	29	258135843		<.0001*

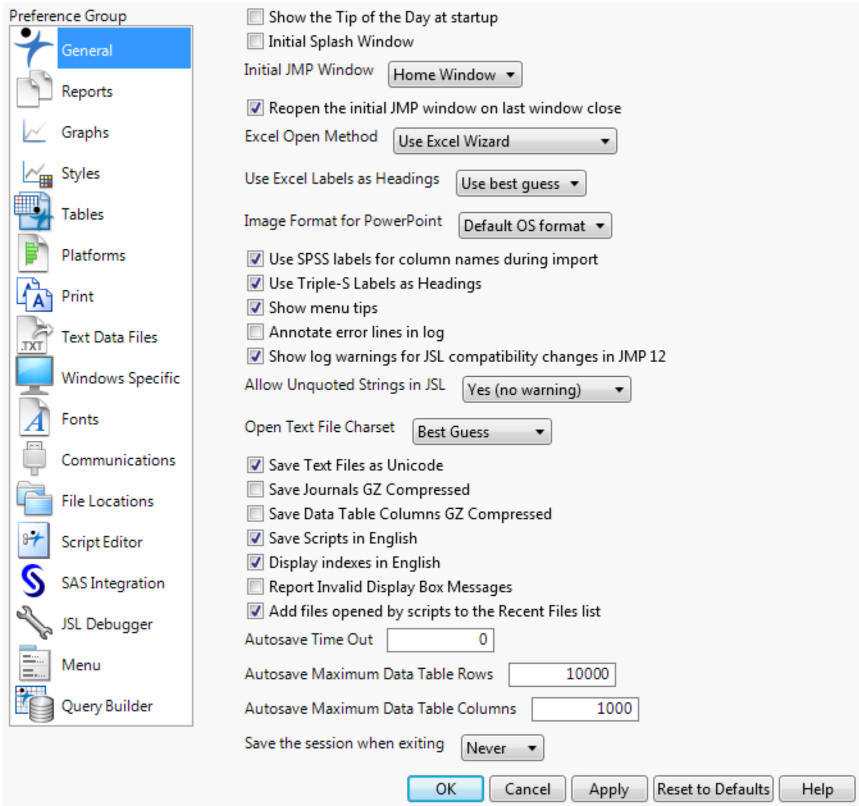
Parameter Estimates

Term	Estimate	Std Error	t Ratio	Prob>|t|
Intercept	631.36791	281.3288	2.24	0.0329*
# Employ	0.1192722	0.008217	14.52	<.0001*

Change Preferences

You can change preferences in JMP using the Preferences window. To open the Preferences window, select **File** > **Preferences**.

Figure 8.6 Preferences Window



On the left side of the Preferences window is a list of Preference groups. On the right side of the window are all of the preferences that you can change for the selected category.

Example of Changing Preferences

Every platform report window has options that you can turn on or off. However, your changes to these options are not remembered the next time you use the platform. If you want JMP to remember your changes every time you use the platform, change those options in the Preferences window.

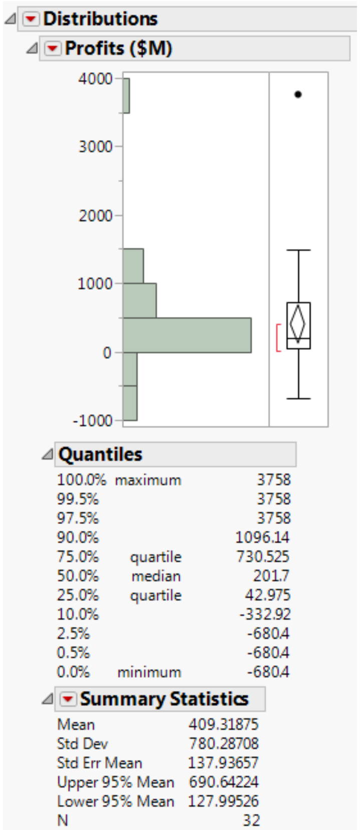
This example shows how to set the Distribution platform so that an Outlier Box Plot is not added to the initial report.

Create a Distribution Using the Default Preference Setting

1. Select **Help > Sample Data Library** and open Companies.jmp.
2. Select **Analyze** > **Distribution**.

3. Select Profits ($M) and click **Y, Columns**.
4. Click **OK**.

Figure 8.7 Distribution Report Window

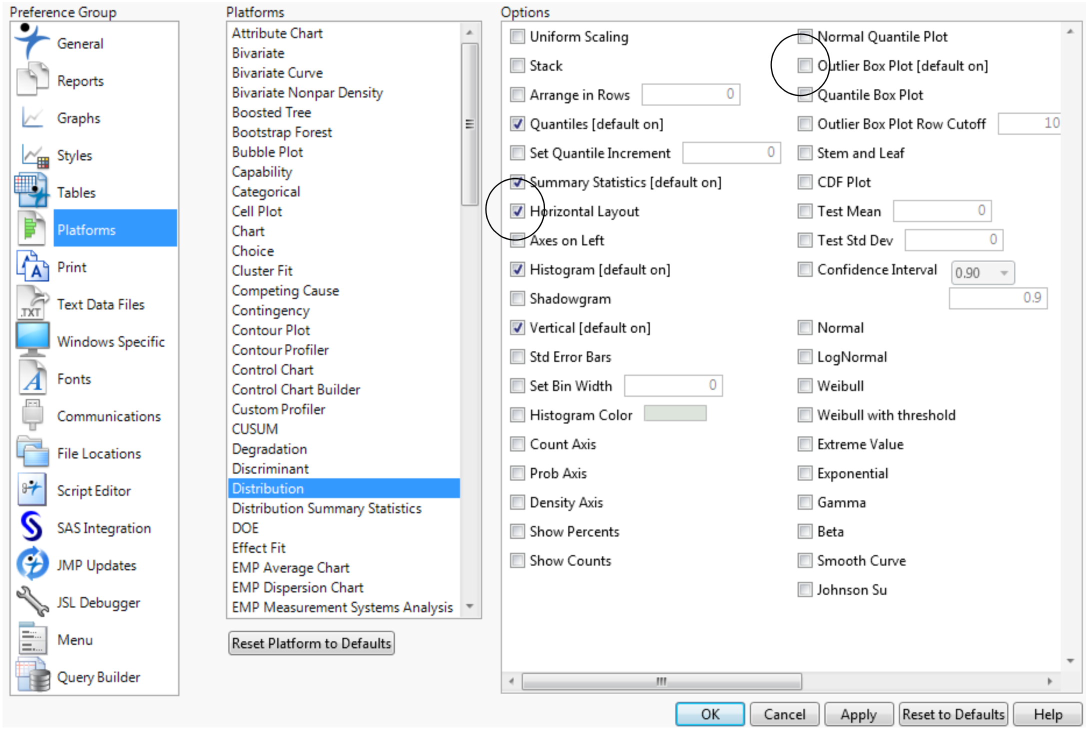


The histogram is vertical, and the graphs includes an outlier box plot. To change the histogram to horizontal and remove the outlier box, select the appropriate options from the red triangle menu for Profits ($M). However, if you want those preferences to be in effect every time you use the platform, then change them in the Preferences window.

Change the Preference for the Outlier Box Plot and Run Distribution Again

1. Select **File** > **Preferences**.
2. Select **Platforms** from the preference group.
3. Select **Distribution** from the Platforms list.
4. Select the **Horizontal Layout** option to turn it on.
5. Deselect the **Outlier Box Plot** option to turn it off.

Figure 8.8 Distribution Preferences



6. Click **OK**.
7. Repeat the Distribution analysis. See "Create a Distribution Using the Default Preference Setting" on page 192.

The histogram is now horizontal and the outlier box plot does not appear. These preferences remain the same until you change them.

For details about all of the preferences, see the JMP Preferences chapter in the *Using JMP* book.

Integrate JMP and SAS

Note: You must have access to SAS, either on your local machine or on a server, to use SAS through JMP.

Using JMP, you can interact with SAS as follows:

- Write or create SAS code in JMP.
- Submit SAS code and view the results in JMP.
- Connect to a SAS Metadata Server or a SAS Server on a remote machine.

- Connect to SAS on your local machine.
- Open and browse SAS data sets.
- Retrieve and view data sets generated by SAS.

For complete details about integrating JMP and SAS, see the Import Your Data chapter in the *Using JMP*.

Example of Creating SAS Code

This example uses the Candy Bars.jmp sample data table, which contains nutrition data for candy bars.

1. Select **Help > Sample Data Library** and open Candy Bars.jmp.
2. Select **Analyze** > **Fit Model**.
3. Select Calories and click **Y**.
4. Select Total fat g, Carbohydrate g, and Protein g, and click **Add**.
5. From the red triangle menu for Model Specification, select **Create SAS Job**.

Figure 8.9 shows the SAS code. (Not all of the data is shown.)

Figure 8.9 SAS Code

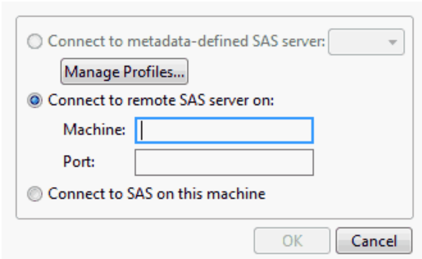
```
DATA Candy_Bars; INPUT  Calories Total_fat_g Carbohydrate_g Protein_g; Lines;
310  20  28  6
230  12  27  4
220  12  24  3
170  8  21  3
200  2.5  43  1
260  16  26  5
190  1.5  42  2
190  11  21  2
230  12  28  3
;
RUN;

PROC GLM  DATA=Candy_Bars ALPHA=0.05;
MODEL  Calories =  Total_fat_g Carbohydrate_g Protein_g;
RUN;
```

Example of Submitting SAS Code

1. Select **Help > Sample Data Library** and open Candy Bars.jmp.
2. Select **Analyze** > **Fit Model**.
3. Select Calories and click **Y**.
4. Select Total fat g, Carbohydrate g, and Protein g, and click **Add**.
5. From the red triangle menu for Model Specification, select **Submit to SAS**.
6. In the **Connect to SAS Server** window (see Figure 8.10), choose a method to connect to SAS (if you are not already connected). For this example, select **Connect to SAS on this machine**.

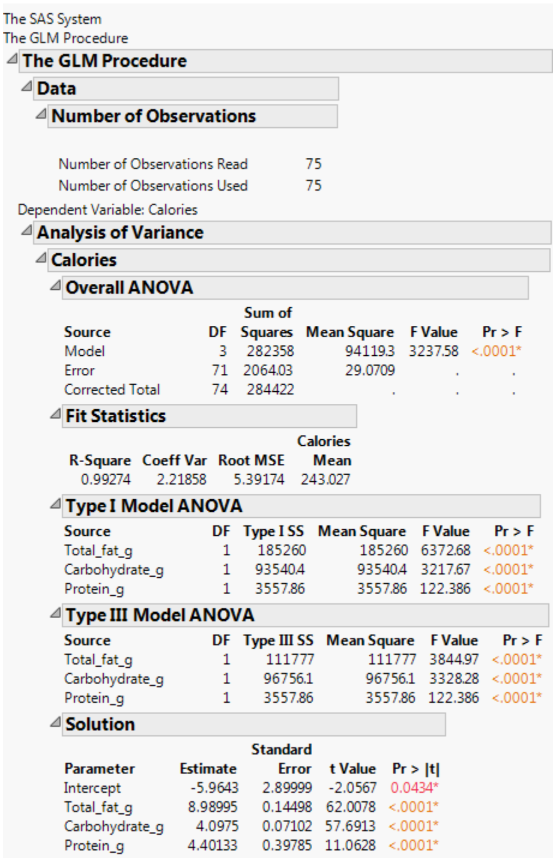
Figure 8.10 Connect to SAS Server



7. Click **OK**.

JMP connects to SAS. SAS runs the model and sends the results back to JMP. The results can appear as SAS output, HTML, RTF, PDF, or JMP report format (you can choose the format using JMP Preferences). Figure 8.11 shows the results formatted as a JMP report. For details, see the Import Your Data chapter in the *Using JMP* book.

Figure 8.11 SAS Results Formatted as a JMP Report


The SAS System
The GLM Procedure

The GLM Procedure

Data

Number of Observations

Number of Observations Read	75
Number of Observations Used	75

Dependent Variable: Calories

Analysis of Variance

Calories

Overall ANOVA

Source	DF	Sum of Squares	Mean Square	F Value	Pr > F
Model	3	282358	94119.3	3237.58	<.0001*
Error	71	2064.03	29.0709	.	.
Corrected Total	74	284422	.	.	.

Fit Statistics

R-Square	Coeff Var	Root MSE	Calories Mean
0.99274	2.21858	5.39174	243.027

Type I Model ANOVA

Source	DF	Type I SS	Mean Square	F Value	Pr > F
Total_fat_g	1	185260	185260	6372.68	<.0001*
Carbohydrate_g	1	93540.4	93540.4	3217.67	<.0001*
Protein_g	1	3557.86	3557.86	122.386	<.0001*

Type III Model ANOVA

Source	DF	Type III SS	Mean Square	F Value	Pr > F
Total_fat_g	1	111777	111777	3844.97	<.0001*
Carbohydrate_g	1	96756.1	96756.1	3328.28	<.0001*
Protein_g	1	3557.86	3557.86	122.386	<.0001*

Solution

Parameter	Estimate	Standard Error	t Value	Pr > |t|
Intercept	-5.9643	2.89999	-2.0567	0.0434*
Total_fat_g	8.98995	0.14498	62.0078	<.0001*
Carbohydrate_g	4.0975	0.07102	57.6913	<.0001*
Protein_g	4.40133	0.39785	11.0628	<.0001*

Index

Discovering JMP

G

H

I

J

L

M

S

T

V

W-Z