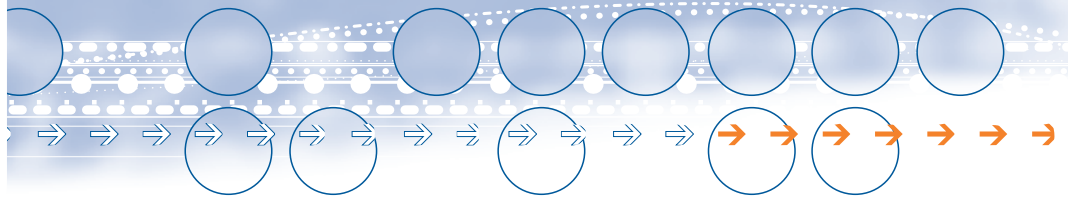




# Spotfire.net<sup>TM</sup>

## Desktop 5



**Version 5.0**

July 2000

The content of this publication may be changed without further notice. No part of this documentation may be reproduced or transmitted in any form or by any means, electronic or mechanical, for any purpose, without prior written consent from Spotfire AB and Spotfire, Inc.

© 2000 Spotfire AB. All rights reserved.

Spotfire Pro is protected by U.S. Patent No. 6,014,661. Other patent(s) pending.

Spotfire and Spotfire Discovery Server are registered trademarks, and Spotfire.net, Spotfire Pro and Spotfire Structure Visualizer are trademarks of Spotfire, Inc.

All other brand or product names are trademarks or registered trademarks owned by their respective companies or organizations.

### **Other resources**

If you have any questions about a Spotfire product, you may try the following:

- Use the built-in help function.
- Consult the README files, which come with the product.
- Visit the Spotfire web site <http://www.spotfire.com/eSupport>, where you can send questions to our support department.

### **Support from Spotfire AB and Spotfire, Inc.**

Spotfire offers a variety of support agreements. Contact Spotfire for more information about email and phone support, training, consulting agreements, and our Product Subscription Programs.

#### **Contacting Spotfire:**

Spotfire AB  
Första Långgatan 26  
SE-413 28 Göteborg  
Sweden

Phone: +46 31 704.1500  
Fax: +46 31 704.1501

Spotfire, Inc.  
60 Hampshire Street  
Cambridge, MA 02139  
USA

Phone: +1 617 621.0340  
Fax: +1 617 621.0381

WWW: <http://www.spotfire.com>

Email: [support@spotfire.com](mailto:support@spotfire.com)



# Contents

|   |           |
|---|-----------|
| <b>Contents</b>                           | <b>i</b>  |
| <b>1 Introduction</b>                     | <b>5</b>  |
| <b>2 Spotfire.net Desktop Overview</b>    | <b>9</b>  |
| Access, Analyze, Publish                  | 9         |
| The Spotfire.net Desktop                  | 10        |
| New Features in Spotfire.net Desktop      | 12        |
| <b>3 Installation and Setup</b>           | <b>15</b> |
| Hardware Requirements                     | 15        |
| Software Requirements                     | 16        |
| Spotfire.net API Documentation            | 17        |
| <b>4 Principles of Visual Data Mining</b> | <b>19</b> |
| Exploring Data Sets                       | 19        |
| Disparate Data                            | 20        |
| Outliers, Anomalies, Trends and Clusters  | 20        |
| Databases and Data Sets                   | 22        |
| Dynamic Queries                           | 23        |
| Query Devices                             | 25        |
| Details-on-Demand                         | 27        |
| <b>5 Getting Started</b>                  | <b>29</b> |
| Exploring a Simple Data Set               | 29        |
| Checklist for Exploring New Data          | 35        |
| <b>6 Working with Query Devices</b>       | <b>37</b> |
| Different Types of Query Devices          | 37        |
| Initial Query Devices                     | 45        |
| Changing the Type of Query Devices        | 46        |
| Moving Individual Query Devices           | 48        |
| Hiding a Query Device                     | 49        |
| Transforming Axis Scales                  | 50        |
| Quick Access to Color and Size Properties | 51        |
| Moving the Query Device Window            | 52        |

## CONTENTS

|           |   |            |
|-----------|---|------------|
| <b>7</b>  | <b>Managing Visualizations</b>                | <b>53</b>  |
|           | Overview of Visualizations                    | 53         |
|           | 2D Scatter Plots                              | 55         |
|           | 3D Scatter Plots                              | 61         |
|           | Histograms and Bar Charts                     | 66         |
|           | Pie Charts                                    | 69         |
|           | Line Charts                                   | 73         |
|           | Profile Charts                                | 82         |
|           | Layout Using Trellis Plots                    | 86         |
|           | Annotations                                   | 93         |
|           | Data and Background                           | 95         |
| <b>8</b>  | <b>Handling Multiple Visualizations</b>       | <b>99</b>  |
|           | Arranging Windows                             | 101        |
|           | Workbook Mode                                 | 105        |
|           | Customizing Visualizations                    | 106        |
|           | Customizing Toolbars                          | 107        |
|           | Advanced Options                              | 109        |
| <b>9</b>  | <b>Marked, Active and Highlighted Records</b> | <b>111</b> |
|           | Methods for Choosing a Sub Set                | 111        |
|           | Actions Involving the Marked Set              | 114        |
|           | Information about the Data Set                | 116        |
| <b>10</b> | <b>The Portal Window</b>                      | <b>117</b> |
|           | What is the Portal Window?                    | 117        |
|           | Spotfire.net Content                          | 118        |
|           | Spotfire.net Applications                     | 119        |
| <b>11</b> | <b>The Details-on-Demand Window</b>           | <b>121</b> |
|           | Details for Different Visualizations          | 121        |
| <b>12</b> | <b>View Tip</b>                               | <b>131</b> |
|           | What is View Tip?                             | 131        |
|           | Ranking Visualizations                        | 131        |
|           | Using View Tip                                | 133        |
| <b>13</b> | <b>Binning</b>                                | <b>135</b> |
|           | Different Binning Methods                     | 137        |
| <b>14</b> | <b>Calculating Columns</b>                    | <b>141</b> |
|           | Calculating Columns—An Example                | 141        |
|           | Advanced Calculations                         | 144        |
|           | Calculation Errors                            | 147        |

|           |   |            |
|-----------|---|------------|
| <b>15</b> | <b>Background Images and Maps</b>             | <b>149</b> |
|           | Image Files and Background Maps               | 149        |
|           | Adding a Background Image                     | 151        |
|           | Adding MapInfo Background Files               | 152        |
|           | Spotfire.net Map Files                        | 155        |
| <b>16</b> | <b>Printing</b>                               | <b>157</b> |
|           | Print Layout                                  | 157        |
|           | Printing the Document                         | 158        |
| <b>17</b> | <b>Importing Data</b>                         | <b>159</b> |
|           | Selecting a Data Source                       | 161        |
|           | Data Conditioning                             | 165        |
|           | Favorites                                     | 169        |
|           | Connecting to a data source using OLE DB      | 171        |
|           | Database Size Limitations                     | 173        |
|           | Sampling a Database                           | 174        |
| <b>18</b> | <b>Exporting Data</b>                         | <b>177</b> |
|           | Details-on-Demand HTML                        | 178        |
|           | Visualization                                 | 179        |
|           | SQL Queries                                   | 180        |
|           | Legend Images                                 | 181        |
|           | Data  | 182        |
| <b>19</b> | <b>Spotfire.net File Types</b>                | <b>183</b> |
|           | SFS Files—Spotfire Interactive Reports        | 183        |
|           | XSF Files                                     | 187        |
|           | SFT Files                                     | 187        |
|           | SFA Files                                     | 188        |
| <b>A</b>  | <b>Reference Guide to Regular Expressions</b> | <b>189</b> |
|           | Introduction to Regular Expressions           | 189        |
|           | List of Regular Expression Symbols            | 191        |
| <b>B</b>  | <b>Glossary</b>                               | <b>193</b> |
| <b>C</b>  | <b>Quick Reference</b>                        | <b>199</b> |
|           | The Menus                                     | 199        |
|           | The Toolbar                                   | 206        |
|           | Mouse Shortcuts                               | 208        |
| <b>D</b>  | <b>Text File Formatting</b>                   | <b>215</b> |
| <b>E</b>  | <b>Data Types</b>                             | <b>219</b> |
|           | General specifications                        | 219        |
|           | Time  | 221        |

## CONTENTS

|   |            |
|---|------------|
| Date                                    | 222        |
| DateTime                                | 224        |
| Integer                                 | 226        |
| Real                                    | 227        |
| Double                                  | 227        |
| String                                  | 228        |
| Data Type Constraints                   | 229        |
| <b>References</b>                       | <b>231</b> |
| Literature on Exploratory Data Analysis | 231        |
| <b>Index</b>                            | <b>233</b> |



# Introduction



*This chapter introduces the Spotfire.net documentation set, explaining how the information is structured and formatted. We recommend that you read this chapter before installing or using the Spotfire.net software.*

## How to read the manual

To get the most out of this manual, you should start by reading the Spotfire.net Overview (chapter 2) to gain a basic understanding of how Spotfire.net works. Then proceed with the setup instructions.

## Intended audience

The *Spotfire.net Desktop User's Guide and Reference Manual* is intended for end users of Spotfire.net Desktop. It explains the basic concepts of visual data mining, and guides you through the features of the product in a step-by-step fashion.

This manual assumes that you:

- are familiar with working in the Windows operating systems, including Windows 95, 98 and/or Windows NT, and
- know how to find and open files



The Windows operating systems provide information about using online Help, as well as finding and opening files.

---

## How this book is organized

Chapter 2 explains what Spotfire.net is and what you can do with it. It also describes the main components of the user interface.

Chapter 3 lists the system requirements for running Spotfire.net. It also describes how to install the software from a CD-ROM, and how to download it from the Internet

Chapter 4 is an introduction to visual data mining, explaining the basic principles and how to take advantage of them.

Chapter 5 teaches you how to get started with Spotfire.net after you have completed installation.

Chapters 6 to 15 explain how to open and analyze data sets and databases making use of all features contained in Spotfire.net.

Chapters 16 to 19 contain information on how to import and export data with Spotfire.net, and how to print reports for sharing information.

Appendix A serves as an introduction to using regular expressions.

Appendix B contains a glossary explaining some of the terms and notations used in this manual.

Appendix C contains a guide to toolbar buttons and short cuts—handy keyboard commands that speed up your work.

Appendices D and E deal with data types and text file formatting.




## Online Help

The information included in this manual is also available in the online help system of the Spotfire.net Desktop application. Online help lets you search for appropriate pieces of information using index words, full-text search, and a table of contents.

- ▶ To access Online Help, select **Help Topics** from the **Help** menu.

## Documentation conventions

The following conventions are used throughout this guide:

| Convention  | Indicates  |
|---|--|
|  | A reference to related topics in this guide or other books that provide more information about the topic.              |
|  | Notes provide important information or warnings related to the concept discussed in the text.                          |
|  | Useful tips are designated with this symbol. These may be shortcuts or smart methods that help you speed up your work. |
| Ctrl + R  | Key combinations for which the user must press and hold down one key and then press another.                           |
| <i>Laser.csv</i>  | File names and variable expressions, such as parameters. Also used for general accentuation.                           |
| <b>File &gt; Open</b>   | Names of commands and menu items.  |

## Acknowledgements

The technology behind Spotfire.net was a result of six years of doctoral research conducted by Spotfire's founder, Dr. Christopher Ahlberg, at the University of Maryland's Center for Human-Computer Interaction under the supervision of Dr. Ben Shneiderman. Since 1991, research on information visualization has been done at the University of Maryland as well as at Chalmers University of Technology in Göteborg, Sweden. This work eventually led to the development of the Spotfire.net products.

## Contact us!

The authors of this manual would like to hear your comments and suggestions on the quality and usefulness of this publication. We also welcome any comments regarding our documentation delivery format. Your input is important for future revisions.

Please send your comments to *support@spotfire.com*

# Spotfire.net Desktop Overview



*This chapter explains what the Spotfire.net Desktop is and what you can do with it. It also explains the main components of the user interface.*

## Access, Analyze, Publish

Spotfire.net is a solution for accessing, analyzing, and reporting on data. It delivers immediate value to researchers, scientists, process engineers, IT professionals and statisticians by facilitating rapid identification of trends, anomalies, outliers and patterns. Spotfire.net Desktop is the core application within Spotfire.net.

Spotfire.net Desktop can access data residing in databases, in a local or networked file system, or on the desktop. It is database independent and extracts data (1,000,000 records or more) from commercial data sources such as Oracle, SQL-Server, Informix, and Sybase.

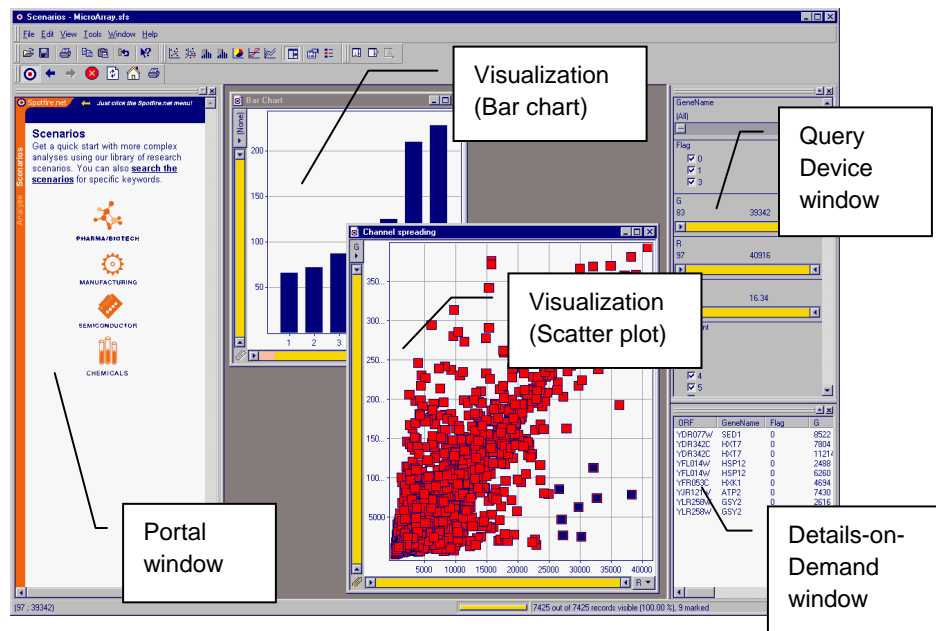
Spotfire.net Desktop uses the patented Spotfire Pro visualization technology, allowing users to interactively query data and instantly visualize results as scatter plots, bar charts, profile charts and a number of other modes.

Spotfire.net Desktop stores data internally in a proprietary data format, allowing for rapid response times to user interaction. It has a series of built in heuristics and algorithms that helps you find interesting alternative views of data.

Spotfire.net Desktop allows users to share their work. Analyses can be easily presented as structured reports for printing or Internet publishing.

## The Spotfire.net Desktop

The image below shows the main parts of the Spotfire.net user interface. The entire application window is referred to as the Spotfire.net Desktop. It is the core of the Spotfire.net solution, which encompasses a suite of applications for database access, statistical analysis, data preprocessing and domain-specific tasks.



The Spotfire.net Desktop

## Visualizations



### 7: Managing Visualizations

Visualizations are the key to analyzing data in Spotfire.net. A variety of visualization types can be used to provide the best view of the data:

- 2D scatter plots
- 3D scatter plots
- Histograms
- Bar charts

- Line charts
- Profile charts
- Pie charts

Different types of visualizations can be shown simultaneously. They are linked, and are updated dynamically when the query devices are manipulated (see below).

Visualizations can be made to reflect high-dimensional data by letting values control visual attributes such as size, color, shape, rotation and text labels.

## The Query Device window

Query devices are used to filter your data. Query devices appear in several forms, and you can easily select a type of query device that best suits your needs (e.g. check boxes, sliders, etc).

When you manipulate a query device by moving a slider or checking a box, all visualizations are immediately updated to reflect the new selection of data.



6: Working with Query Devices

## The Details-on-Demand window

The Details-on-Demand window can be used to show the exact values of a record or group of records. By clicking a marker in a visualization, or marking several records by drawing a rectangle around them, you can study the numerical values in a tabular form.



11: The Details-on-Demand window

## The Portal window

The Portal window is the gateway to a range of services such as online technical support, discussion groups, domain-specific scenarios, free download of plug-ins, etc.

The Portal window is also where you will run your Spotfire.net applications. These peripheral applications replace previous plug-ins. They include a range of features such as statistical analysis, data preprocessing and genomics tools.



10: The Portal Window

## New Features in Spotfire.net Desktop 5



17: Importing  
Data

### Access

- New import dialog for simplified data access and data conditioning
- Faster text import
- Data reshape (aggregation and pivoting) during import
- Support for several new data formats
- Support for new XML based file format: XSF.



7: Managing  
Visualizations

### Analyze

- Profile charts for representing records as lines, with column names on the X-axis.
- Trellis layout for splitting visualizations into multiple plots
- White Details-on-Demand for better readability
- Record in Details-on-Demand can be clicked to activate the corresponding record in visualizations
- Annotations for individual visualizations and columns
- Improved Columns tab in the Properties dialog
- Warning appears before changing the set of marked records
- Customizable marker shapes, allowing mapping of specific shapes to individual values
- New Advanced Options tab in the Customize dialog

### Publish

- Text file export as CSV, TXT or SKV files
- New tabbed dialog replacing the Page Setup Wizard
- Improved quality of printouts
- New Export Visualization dialog for controlling various features of bitmap export



## Other technical notes

- Colon is no longer one of the default column separators accepted when loading text files or pasting text data. (However, by using the Import Data option, you may choose any character as a separator, see page 162)
- When exported, CSV files use Windows default list separator as column separator (set in Regional Settings of Windows)
- OLE DB is not available in standard installation
- Treatment of quotation marks in string fields has changed (see page 228).



# Installation and Setup



*This section lists the system requirements for running Spotfire.net Desktop. It also describes how to install the software from a CD-ROM, and how to download it from the Internet.*

## Hardware Requirements

Spotfire.net in itself does not require a high-powered computer to run smoothly. However, if your data sets are large they will require both a certain amount of RAM, and will of course benefit from high processor speeds.

|                    | <b>Pentium®<br/>clock speed<br/>(MHz)</b> | <b>Random Access<br/>Memory<br/>(Mb)</b> | <b>Free hard<br/>disk space<br/>(Mb)</b> |
|--------------------|---|--|--|
| <b>Minimum</b>     | 100                                       | 32                                       | 15                                       |
| <b>Recommended</b> | 250 or higher                             | 64 or more                               | 22 or more                               |

For data sets under 10 000 records and 100 000 fields, a 100 MHz Pentium® processor running Windows® 95 will suffice if you have 32Mb of RAM. We recommend at least a 250 MHz processor with 64Mb RAM for larger data sets.

When 3D visualizations are used for analyzing, you will find that the graphics card of the computer is a critical component. OpenGL support is required and hardware acceleration beneficial to allow the processor to devote its work to other things than presenting the display with information. Due to the great variety of graphics cards manufacturers, Spotfire.net Desktop has not been tested on all of them. If you experience difficulties or slow performance, check that the most recent drivers for your hardware are installed and that the manufacturer's recommendations are met.

## Software Requirements

During the installation process the setup program checks that your computer contains all necessary operating system components. Most components are mandatory, while some may be missing without the installation being aborted. If this is the case you will be warned about the missing system component.

### **The following system components are mandatory:**

- One of the operating systems Windows 95, Windows 98, Windows NT 4.0 with Service Pack 4 or higher, or Windows 2000
- On a Windows 95 system, the DCOM component
- OLE Automation 2.30 or later
- Common Controls 4.70 or later
- MDAC 2.0 or later
- Shfolder.dll 5.0 or later

### **The following system components are checked and if missing, the user is warned:**

- VBScript 5.0 or later (required for opening files with embedded scripts and for using regular expressions)
- On a system with Internet Explorer, the XML parser should be version 5.0 or later (required for most features of the Import Data dialogs and for opening XSF files)

The Spotfire.net Desktop will run without these components, but with the limitations mentioned above.

The system components can be installed using the separate product Spotfire Operating System Upgrader. This product installs all the system components mentioned above. In addition, it allows you to select what version of MDAC to install (2.0, 2.1 or 2.5) and whether to install Internet Explorer 4.01 SP2. All these system components are in English.



You must have Administrator privileges to execute the Spotfire.net Desktop setup or the Spotfire Operating System Upgrader setup.

---

## Downloading Spotfire.net from the Internet

If you participate in the Spotfire.net Subscription Program—i.e. you have a support agreement with Spotfire for the Spotfire.net product—you may download the latest version of Spotfire.net Desktop from the Internet. New versions and bug fixes are continuously made available there. Please ask your Spotfire representative for the exact location, or contact Spotfire AB for more information.

Enter your name and the password provided by Spotfire AB, then select the software you want to download. When the Spotfire.net distribution file download is complete, double-click the EXE file and the setup process will start automatically. Follow the instructions of the setup program.

## Installing Spotfire.net from CD-ROM

### ▶ To install Spotfire.net Desktop on Windows 95, 98 or Windows NT 4.0:

1. Close all Windows programs.
2. Activate the **Control Panel** and double-click **Add/Remove programs**. Click the **Install...** button and follow the on-screen instructions.

## Removing Spotfire.net

In Windows 95, 98 or Windows NT 4.0, use **Add/Remove programs** in the Control panel. Select Spotfire.net from the list of programs that can be automatically uninstalled. The uninstall program will remove all files installed by Spotfire.net, which are no longer used by any other program in the system. Any data files you may have added in the Spotfire.net directories will be left intact.

## Spotfire.net API Documentation

Spotfire.net may be remote controlled by OLE automation. The API (Application Programming Interface) supports functions such as database loading, visualization manipulation, and window manipulation.

Spotfire Software Laboratory provides the full documentation required to work with the API. For more information about this product, please contact Spotfire.



# Principles of Visual Data Mining



*Visual data mining using Spotfire.net involves interactive visualization and tightly coupled dynamic queries. This chapter explains the concept and the methods used for exploring large, complex data sets.*

The following concepts used in Spotfire.net are introduced:

- dynamic queries
- query devices
- interactive visualizations
- tight coupling
- details-on-demand

## Exploring Data Sets

A data set may be experimental data in a Microsoft Excel or Lotus 1-2-3 spreadsheet. It may also be the results of a series of tests performed on a matrix of chemical compounds or even a list of addresses or information collected from a genetic experiment stored in an Oracle or Sybase database. Analyzing, exploring, searching, understanding, and presenting such material is a complex task. Nevertheless it is performed on a daily basis by market researchers, scientists, engineers, and many others.

The objective may be to understand the financial markets, to search for relations between medical treatments and results, to identify and validate drug targets, to convey understanding for new company strategies, and so on.



17: Importing  
Data

Even when the problem is well understood it may be hard to find trends, clusters, anomalies, and other patterns that may be concealed in a data set related to the task.

Although there are many statistical methods for confirming a predefined hypothesis, there are very few adequate methods for creating hypotheses about complex data sets. To make an analysis with few or no predefined goals may well be likened to looking for a needle in a haystack—in fact, it is one of the most complex undertakings when dealing with large data sets. The problem is open-ended—exploratory rather than confirmatory in nature, and therefore much harder to solve.

Exploring and understanding large data sets are difficult tasks. Tools utilized for these activities may be spreadsheet programs, database systems, GIS, OLAP tools, presentation graphics packages, statistical packages, etc. They are typically complex, difficult to use and not very helpful for performing visual data mining tasks.

## Disparate Data

Visual Data Mining uses the extraordinary power of the human visual system in combination with an amazing neural network, our brain, to reveal patterns and correlations.

Exploration of databases encompasses a variety of tasks. For certain kinds of data, such as a data set of environmental measurements, analysis is the key. The driving force is the need to develop an intuitive understanding of a set of observations. For other kinds of data, such as a Structure Activity Relationship database, interest is concentrated on searching for and locating particular chemicals, such as biologically active compounds. Developers of pharmaceuticals need to analyze huge collections of experiments performed with a vast array of compounds and correlate them with numerous other experiments in order to find compounds with the appropriate properties. Spotfire.net allows you to perform all these activities.

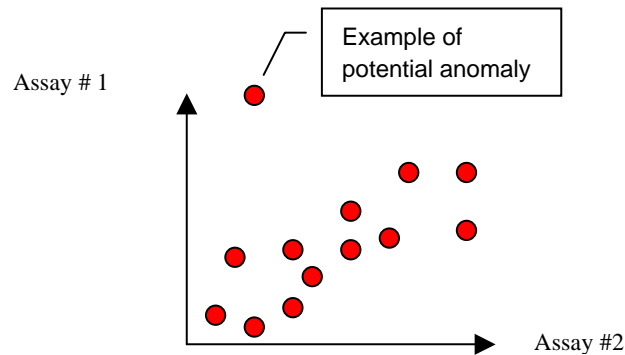
## Outliers, Anomalies, Trends and Clusters

When analyzing data you typically want to remove erroneous data, identify trends, patterns, clusters, anomalies, outliers and other features of the data.



## Outliers and anomalies

When a data set is studied it is quite normal to find anomalies. The irregularity may stem from the actual values being extreme or perhaps from an error made when collecting data.

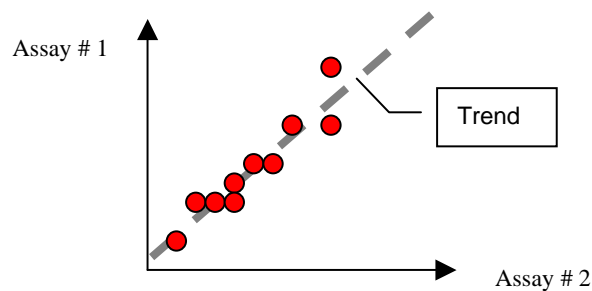


Visualization indicating an anomaly

In the figure above, a possible outlier is apparent; an assay result with a large response in Assay # 1, and a low response in Assay # 2. Anomalies and outliers that should not be included in the continued analysis may be removed from the data set Spotfire.net has in memory.

## Trends and patterns

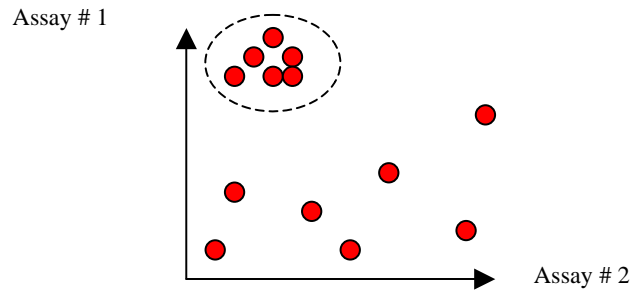
A typical and straightforward trend for a data set such as a pharmacological analysis data set may be a correlation between two assays. Hypotheses concerning other patterns, or lack of patterns, can also be confirmed.



A possible correlation between Assay # 1 and Assay # 2

## Clusters

A typical cluster pattern in a market analysis may be the way purchases are grouped in certain product categories.



Indication of a cluster

In the figure above, a cluster of data is visible: results with high values in Assay # 1 and low values for Assay # 2.

## Databases and Data Sets

A database can be organized in different ways – e.g. as a hierarchical database or as a spreadsheet. Spotfire.net works with all data sources that can be transformed into a flat (two-dimensional) data format. A database extraction loaded into Spotfire.net is referred to as a data set. Such a set consists of a number of items (or records), all with the same structure. Each record consists of a number of fields (also referred to as values or attributes). Each field is identified by a name and a type. Spotfire.net recognizes a number of standard field types, e.g. integers and floating-point numbers, text strings, and dates. The Importing Data chapter describes this in more detail.

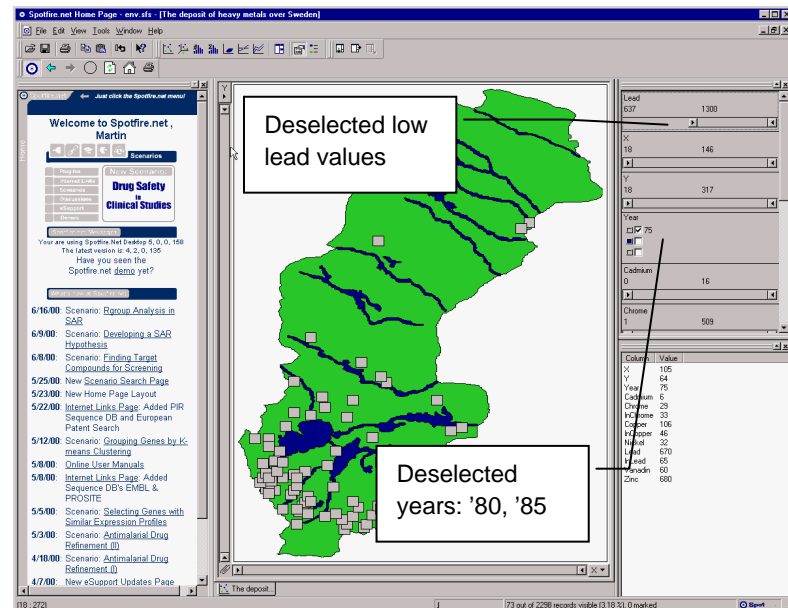
## Dynamic Queries

Spotfire.net does away with many of the problems of manually composing database queries in a query language such as SQL. This is achieved by using dynamic query devices—a technology for visually seeking information about a collection of data. Queries are built and executed by manipulating graphic objects, known as query devices.

Results are presented visually and continuously at every step of the query formulation process, thus guiding you through the process. All actions are reversible.

Some of the concepts pertaining to dynamic queries, such as interactive visualizations and query devices, are exemplified in this section. For a detailed description of using queries, see the chapters beginning with Working with Query Devices.

On the following pages are examples in which Spotfire.net has been used to visualize a data set of results from environmental sampling. Each record provides information about airborne deposits of ten heavy metals, the location, and the year in which the observation was made.



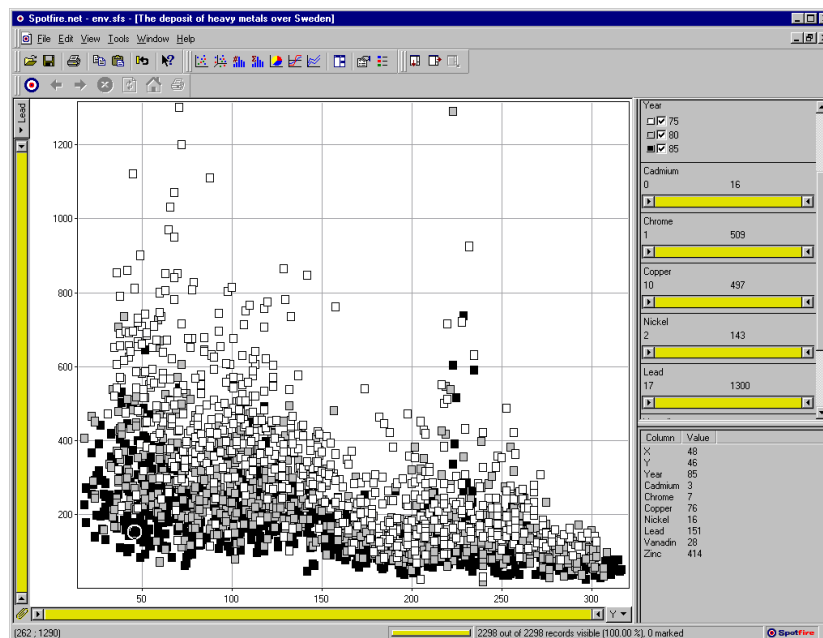
Sample setup of Spotfire.net showing elevated lead deposits in 1975

#### 4 • PRINCIPLES OF VISUAL DATA MINING



#### 7: Managing Visualizations

In the visualization above, the data is displayed as a two-dimensional interactive scatter plot; a marker represents each record in the data set. These markers can be given a large number of attributes, e.g. color, size, shape, orientation in order to add more dimensions to the visualization. By moving the query devices located to the right, the records displayed in the visualization are filtered. In the above example the slider for lead has been adjusted so that low lead deposit levels have been filtered out (deselected) from the visualization. In the figure above a map has been added as a background to the scatter plot in order to visually relate the markers to the location of the measurement.



2D-scatter plot with shading (color) indicating a third dimension

In the scatter-plot above, shade is used to differentiate between different sampling years. This shows at a glance that levels of lead deposits have decreased from 1975 to 1985 (lightest gray = 1975, middle gray = 1980, black = 1985)

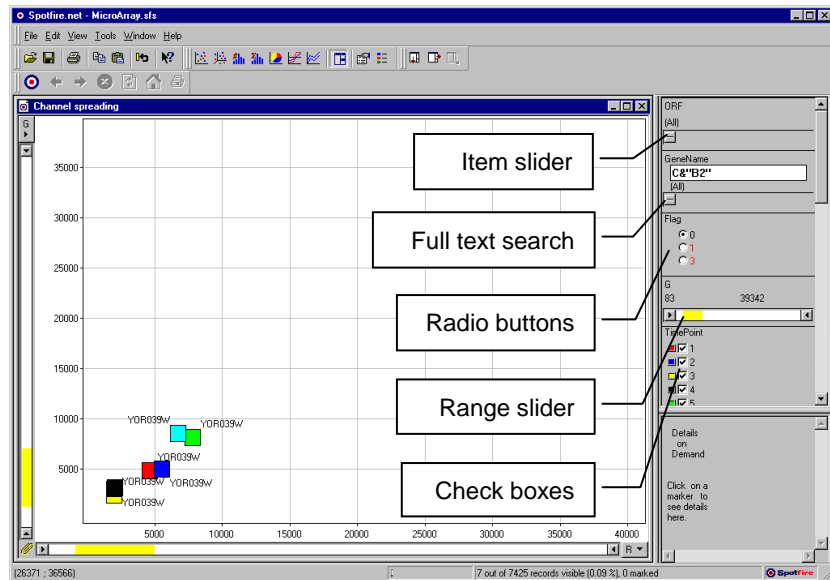
## Query Devices



### 6: Working with Query Devices

Query devices are sliders, check boxes and other graphic controls used for filtering the data shown in the visualization. A query device is always associated with a certain data column. The different query device types are described below.

A **range slider** is used to select records with values within a certain range. An **item slider** is used to select individual values in a column. **Check boxes**, one for each value appearing in the corresponding column in the data set, are used to select or deselect the values that should or should not appear in the visualization. Check boxes are typically used when a record field holds just a few distinct values. **Radio buttons** are check boxes that enable one and only one choice among the alternatives. **Full text search** query devices permit the search for a specific string of alphanumeric characters with the use of Boolean operators.



Five types of query devices used to filter the data set

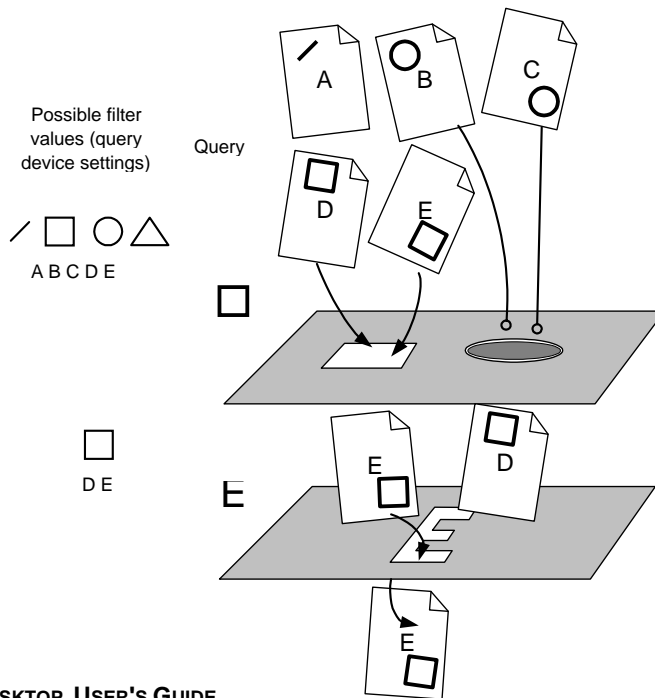
## Tight coupling

Tight coupling is a technique that simplifies the formulation of queries.

As an example, suppose that you are exploring a database containing information about films on video. Assume that Ingmar Bergman is selected as the director. When browsing through different actresses, the tight coupling technique causes only the ones appearing in Ingmar Bergman films to be selectable; the other ones are already disqualified by the constraint of having selected Bergman as director. The same restrictions apply to sliders when selecting film category and film length; only categories for which there exist Ingmar Bergman films are selectable, and the film length will be restricted to the running times of Bergman films.

When an actress is selected, the set of possible actors is reduced even further: it is now only possible to choose those who have worked with the chosen actress in a Bergman film.

The figure below illustrates this concept: initially, there are documents containing a character and a graphical object. Any character or graphical object may be selected for filtering. When the square graphical object is chosen, only two documents remain. Since these contain the letters D and E, the tight coupling mechanism ensures that only these characters can be chosen for the next filtering process.



In Spotfire.net, tight coupling is also reflected in the appearance of the query devices themselves. For example, a range slider will have a deeper yellow color denoting the values for that parameter that are visible in the active plot. Change the focus to include other records and the deeper yellow part of the bar will change size as well.

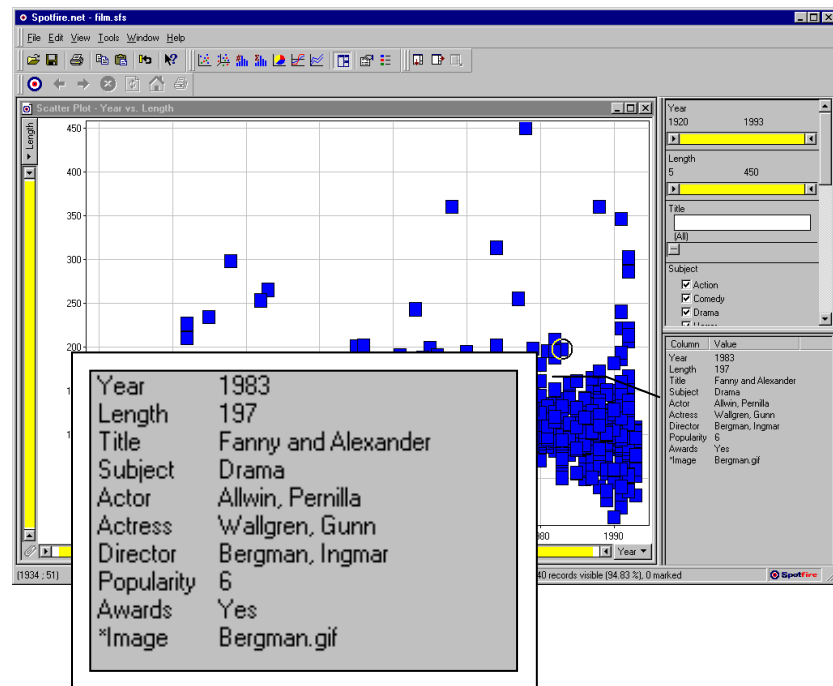
## Details-on-Demand

In an interactive visualization system it is not only important to get an overview of large data sets. It is also important to be able to see the full details of a single record, such as the entire customer transaction in the market analysis database, or the full measurement description in the environmental database.



11: The Details-on-Demand Window

Simply click on a marker in two- or three-dimensional scatter plots to display details, i.e. the full data record. The figure below shows such an example. To get more information about the different ways to display data, please read the Details-on-Demand chapter.



The Details-on-Demand Window





# Getting Started



*This chapter provides guidance on how to get started with Spotfire.net, outlining the basic steps required for embarking on an analysis. This includes accessing the data set, setting up an informative visualization, and filtering the data using query devices.*

## Exploring a Simple Data Set

### The Periodic table data set

As an example of simple, yet interesting information, the periodic table of the elements will be used. The data set holds 75 records with 9 fields each. The columns include the name of the element, six important properties, and the coordinates of each element when the periodic table is printed in the familiar layout first suggested by Dimitri Mendeleev. The text file contains the column names, column types (integer or string) and the data itself.

```
*P1,*P2, Element, Atomic Number, Atomic Mass, Atomic Radius,  
-  
-Ionic Radius, Ionization Energy, Electronegativity, ↵  
INT,INT,STRING,INT,INT,INT,INT,INT, INT ↵  
140,0,Ac,89,227,200,126,51,11,62,56 ↵  
630,80,Ag,47,107,144,129,75,18,124,40 ↵  
750,160,Al,13,27,143,67,60,16,28,25 ↵  
.....
```



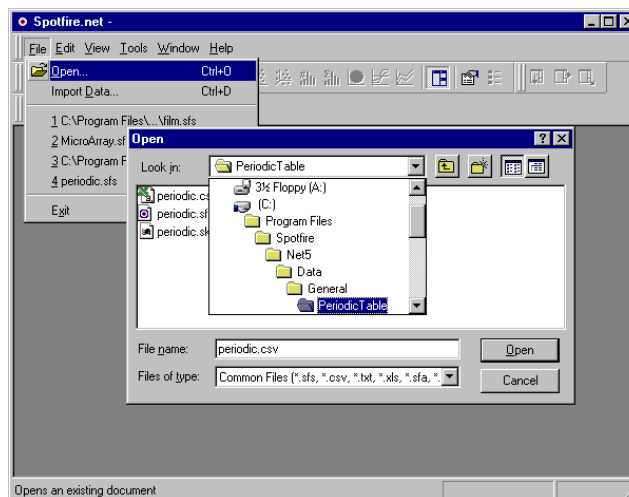
17: Importing data

This database is stored in one of the formats recognized by Spotfire.net. For more information about these formats, see Importing Data. This particular sample can be found in the directory named *Data/General/Periodic Table* in the Spotfire.net distribution as *periodic.csv*.

## Loading a data set

▶ **To load a data set into Spotfire.net:**

1. Start Spotfire.net by selecting Spotfire.net from **Programs** in the **Windows Start** menu, or by double-clicking the Spotfire.net icon on the desktop
2. Select the **Open** option from the **File** menu in the Spotfire.net main window. The **File open** dialog will appear.



Loading a file into Spotfire.net

3. The folder **Data** will be shown in the Spotfire.net catalog
4. Browse to the folder *Data\General\Periodic Table*. Select the file *periodic.csv* and click the **Open** button



There are other ways to load a raw data set:

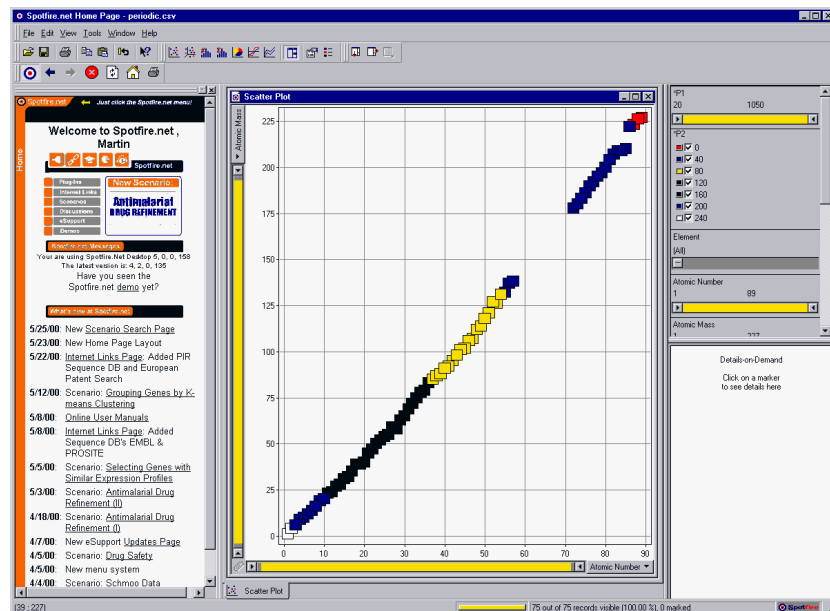
- Locate *periodic.csv* with Windows Explorer, and drag the file icon to an open instance of Spotfire.net, or drop it onto the Spotfire.net icon on your desktop
- Open *periodic.csv* in Excel; copy the content [Ctrl+A] [Ctrl+C] from Excel and paste [Ctrl+V] it into Spotfire.net



## 7: Managing Visualizations: Initial visualizations

### Initial setup

Spotfire.net will automatically set up an initial visualization where each record is represented by a marker, along with a number of query devices for manipulating the visualization.

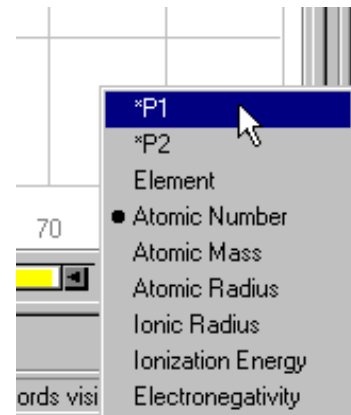


Initial visualization of the periodic table data set

### Changing visualization axes

To change the column to be mapped to an axis, click the button next to the axis, and select a column from the pop-up menu. To see the Mendeleevian layout of the elements, click on the X-axis selector box and select P1 and click the Y-axis selector to choose P2 for the Y-axis.

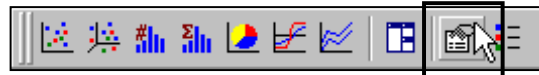
Changing the column to use as X-axis in the visualization



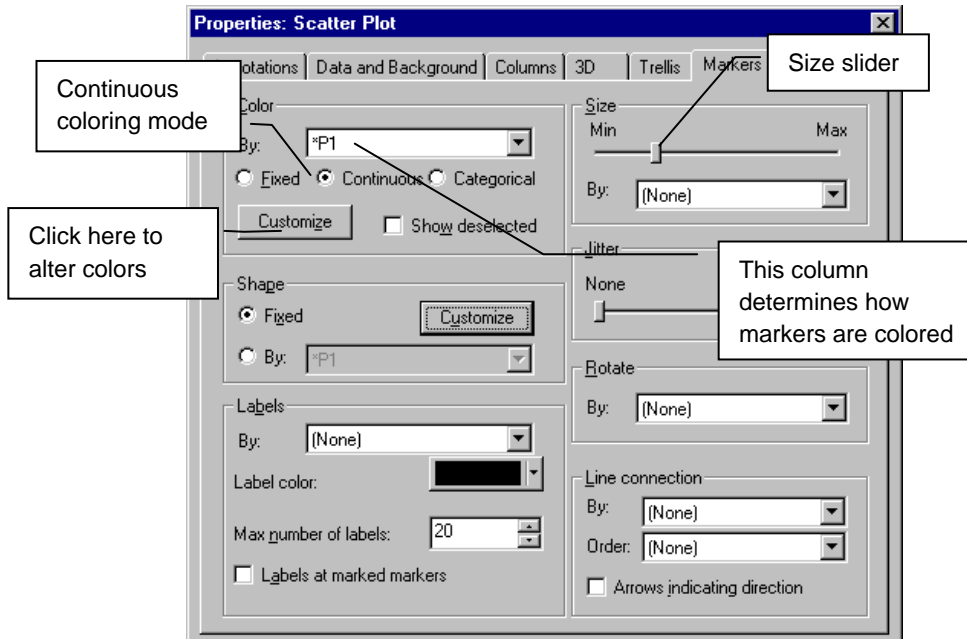
## Setting visualization properties

▶ **To adjust the appearance of a visualization:**

1. Select **Properties** from the **Edit** menu. (Alternatively, press [Alt+Enter] or use the toolbar button.)



2. Go to the **Markers** tab
3. Move the **Size** slider towards Max to make the markers larger
4. Set the **Color** control to **Continuous**, and the **By** control to *Electronegativity*. This will cause the markers to be colored according to their electronegative potential. The elements with the lowest values will be red and the highest blue. You may alter these colors with a dialog box that opens under **Customize**.



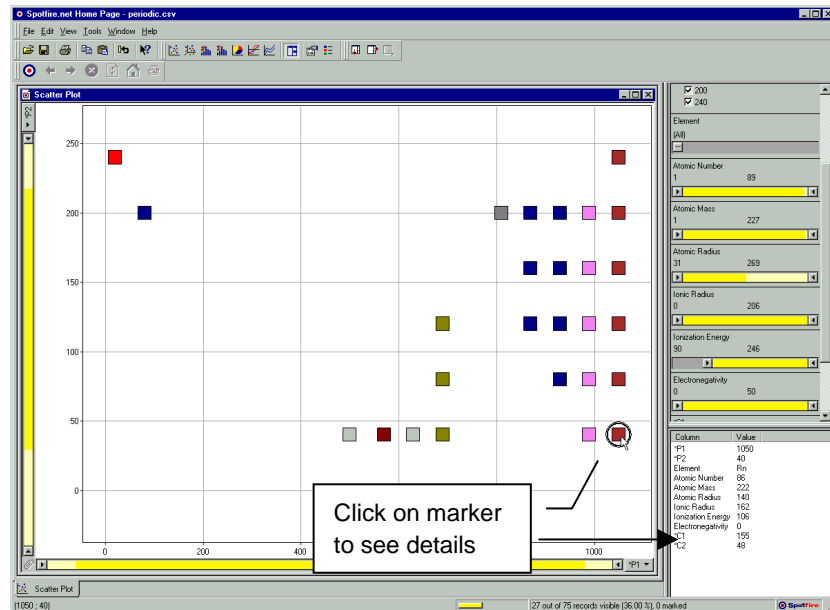
Markers **Properties** dialog box for scatter plots



## 6: Working with Query Devices

### Querying the data set

To select a subset of markers to display in the visualizations, change the settings of the query device sliders found in the area on the right. For example, move the left drag box of the Ionization Energy slider to select only those elements with high ionization energies. We can now see that the elements with high ionization energy are more or less the same as those with high electro negativity since there are few markers with a color representing low electronegative potential in the visualization.



Elements with high ionization energies

Click on a marker with the left mouse button to activate that record. Information in the database about that element will appear in the Details-on-Demand window in the bottom right hand corner.

## Creating a new visualization

▶ **To create a three-dimensional visualization:**

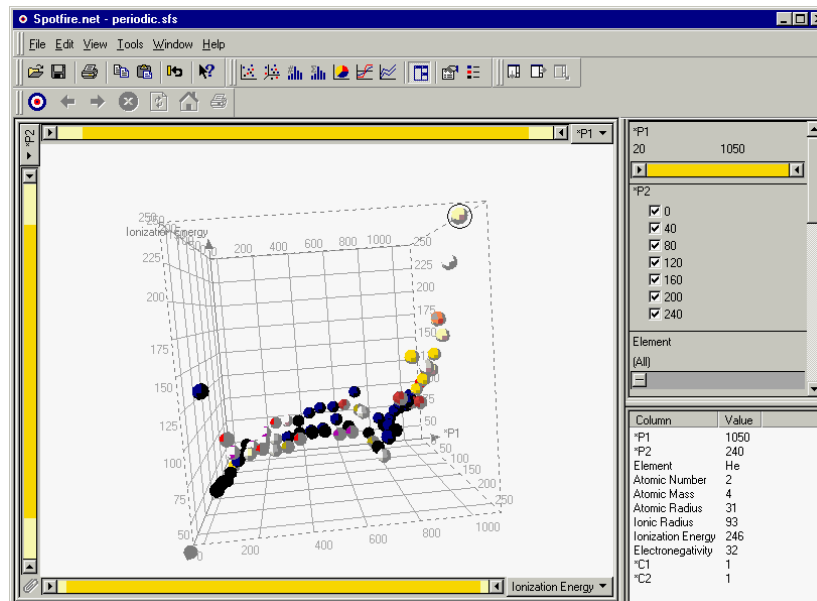
1. Select **New 3D Scatter plot** from the **Window** menu, or use the keyboard shortcut: [Ctrl+2].
2. Reassign the X-axis to *P1*, the Y-axis to *P2* and the Z-axis (along the top of the visualization) to *Ionization Energy*.
3. View the plot from different angles by dragging with the right mouse button.



8: Handling  
Multiple  
Visualizations

The 3D plot will cover the 2D plot but you can show both plots by selecting **Window > Auto Tile**.

If you move your mouse pointer over one of the markers in the 2D plot, the corresponding marker will also be highlighted in the 3D view. If you activate one element in the 2D plot (by clicking on it) you can move the selection using the arrow keys.



3D plot of the periodic table of the elements.

## Checklist for Exploring New Data

The following steps are typically helpful when exploring a new data set, especially if its content is unknown:



12: View Tip

1. Start off by selecting different combinations of visualization axes to reveal patterns, trends, and anomalies. Use View Tip to browse through different combinations of axes. The anomalies should be scrutinized to determine if they are due to erroneous data.
2. Write down particularly interesting combinations of axes, or save the settings so that you can retrieve interesting visualizations later.
3. Color the visualization objects, first by category. Again, write down or save variables that produce interesting groupings and/or patterns.
4. Color visualization objects by continuous variables, to reveal areas where interesting features such as consistently high or low values occur.
5. Go through the range sliders to reveal markers with values occurring in a certain range and observe if they are grouped in some kind of cluster.
6. When interesting clusters are identified, these may be either zoomed in on or marked, followed by changing to another view to reveal whether the markers are clustered for a particular reason.





# Working with Query Devices

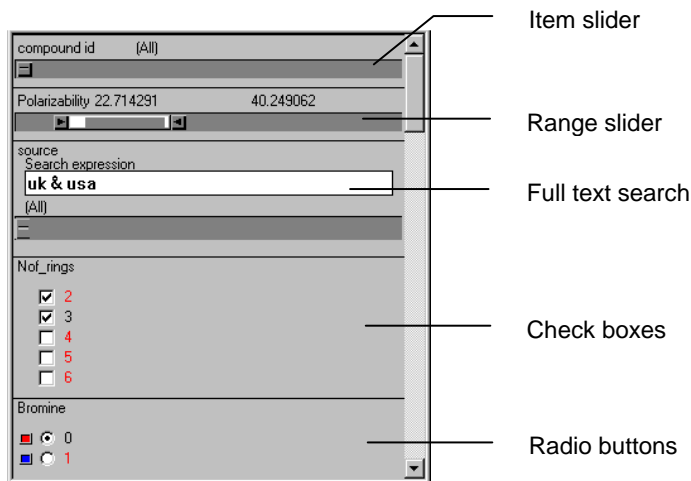


*Spotfire.net automatically creates query devices when a data set is loaded. One device is created for each column of data. This chapter describes Spotfire.net's query devices, and how they can be used. It also explains the application's initial choice of query devices, how to change query device types, and how to rearrange and hide the query devices.*

## Different Types of Query Devices

A Spotfire.net query device is a visual tool for performing dynamic queries against an underlying data set.

Spotfire.net has five types of query devices: range sliders, item sliders, full text search, check boxes, and radio buttons. In the default layout, query devices are found in the window at the top right.



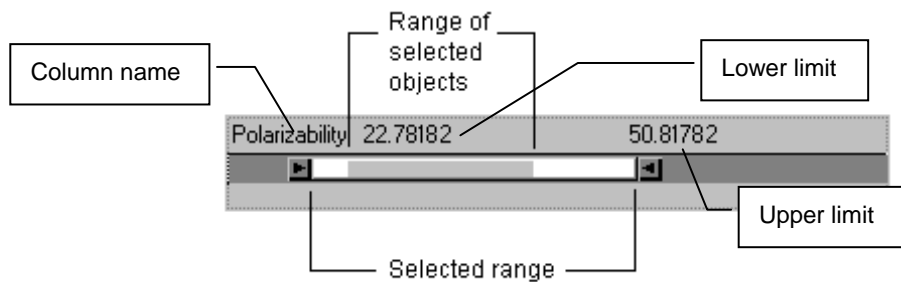
The full assortment of query devices available in Spotfire.net

## Range slider

Range sliders allow the selection of a range of values. The left and right drag box can be used to change the lower and upper limit of the range – meaning that only records with values within the chosen range are selected and therefore visible in the visualization. Labels above the slider indicate the selected span. The range can also be adjusted with the arrow keys when the query device is active: left and right arrows move the lower limit (left drag box), and up and down arrow keys move the upper limit. The slider automatically jumps to values in the data set (not necessarily the visible or selected records).

The currently selected interval of the range slider can be grabbed and moved to pan the selected range – this provides a powerful way of sweeping over different “slices” of a data set. Click and drag the yellow portion of the range slider to do this. Observing the reactions the other sliders to such a sweep can give some interesting clues to correlation between parameters in the data set.

If other query devices impose further restrictions, then the result may be that parts of the interval of the range slider are unpopulated. This area is indicated with a pale yellow color, as opposed to the bright yellow color that indicates the populated interval.

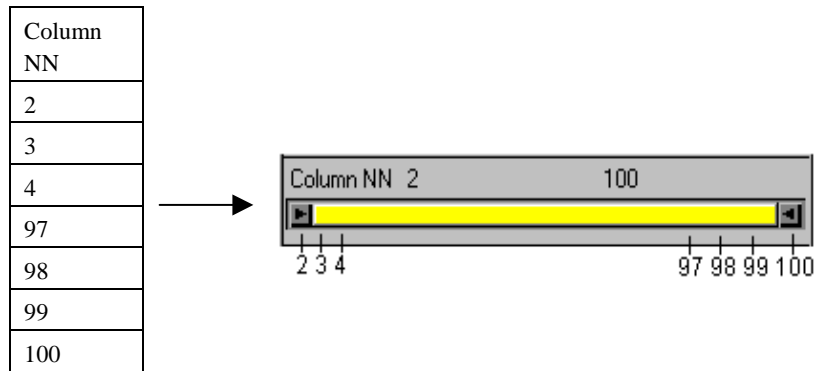


A range slider query device

An important feature of the range slider is that the values are distributed on a linear scale according to the values of the data. Thus if values are unevenly distributed, this will be reflected in the range slider.



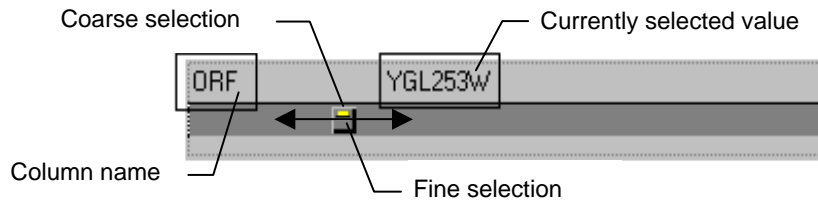
This is not the case with item sliders, where values are evenly distributed along the range of the slider, regardless of what values appear in the column.



The scope of the range slider reflects the values of the data in the column

### Item slider

In an item slider query device, data items are evenly distributed on a continuous linear scale. However, the item slider selects only a single item at a time. The selected value is displayed as a label above the slider. As a special case, all items are selected when the slider handle is at the extreme left of the scale.



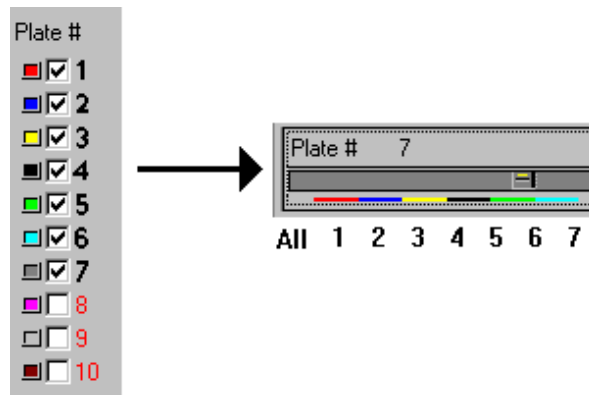
An item slider query device

The item slider features two levels of granularity—a coarse grain slider for moving rapidly between items, and a fine grain slider for moving more slowly between adjacent items.



The scope of an item slider is dependent on the settings of other query devices. This means that the item slider range constantly changes as you manipulate the query devices.

## 6 • WORKING WITH QUERY DEVICES



Values not present in the set of currently selected records—in this case 8, 9, and 10—will not be included in the scope of the *Plate#* item slider

If you are working with large data sets—consisting of perhaps several hundreds of thousands of records—it can be quite tricky to get the slider to stop at the desired record. The solution to this problem is to use the fine-resolution handle together with the keyboard.

### ▶ To pin-point a certain entry with the item slider:

1. Click the upper handle of the slider, and drag the box to the approximate position of the entry by keeping an eye on *Currently selected value* above the slider.
2. Fine tune with the fine resolution handle.
3. When the input focus is set on the slider (marked by a dotted line), you can use the arrow keys on your keyboard to adjust the slider to the exact position of the entry. Up and right arrows move to the next record, down and left to the previous one.

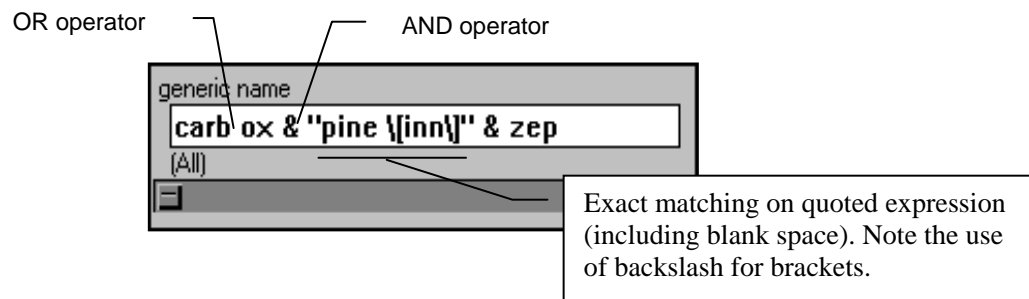
When the item slider drag box is moved to its leftmost position, all values for the slider are selected. This is indicated by the label **(All)** above the slider.

## Full Text Search

The full text search query device allows you to locate records that include matching expressions and patterns. It also allows the use of *regular expressions*, which provide a mechanism to select specific strings from a set of character strings.

### Normal text search

By *normal text search*, we refer to search strings that don't contain regular expressions. The search can be made arbitrarily complex by use of logical operators AND (&) and OR (blank space). Search expressions are evaluated from left to right.



Once the search string has been entered, press Enter on your keyboard to execute the search. All records matching the search criteria will be shown in the visualization window.

### Using the full text slider

The slider integrated in the full text search query device is really just an item slider with one important difference: the *scope* of the slider corresponds to the result of the full text search. This means that if you set the slider in its leftmost position (All), all records *that match the search criteria* will be shown (which doesn't necessarily mean all the records in the data set).

Use the slider to select individual records among those that match the search criteria. Drag the upper portion and the scrolling will be rapid. If the search has resulted in many hits, some will not appear with the coarse selector. When in the appropriate region of hits, drag the fine-tuning lower portion to locate the specific value you want.

As with other sliders, once activated they can be adjusted with the keyboard arrows. Keyboard selection is always fine grained—record by record.

### Example

Let's look at the example in the figure above. Note that the brackets have to be preceded by a backslash since they have a special meaning in regular expressions (see Regular Expressions below).

**carb ox & "pine \[inn]" & zep**

Spotfire.net will evaluate the search expression in the following way

**(carb OR ox) AND ("pine [inn]") AND zep**

In plain English this means: “Show me all records where the column Generic Name contains names that include the sub-strings **carb** or **ox**, and also contains the full string **pine [inn]** as well as the sub string **zep**”.

This means that **OXCARBAZEPINE [INN]** would be matched, but not **AMOXAPINE [U;INN]** or **CARBAMAZEPINE [U;INN]**.

The operators allowed in normal searches are:

|               |  |
|---------------|--|
| (blank space) | logical OR   |
| &             | logical AND  |
| " "           | search for the exact sub-string within double quotes |



The preselected **Show Empty Values** option, available by right-clicking the query device, may give you unintended search results, since it also accepts records with null values in the column you are searching. If this is not your intention, right-click the text search slider, and uncheck the **Show Empty Values** option.

---



A: Reference Guide  
to Regular  
Expressions

## Regular expressions

*Regular Expressions* (RegExp) are used by many utilities to aid in the matching of strings. The use of regular expressions empowers you to search for very specific patterns, regardless of complexity. Regular expressions may contain simple characters, that is any alphanumeric character, but also special characters and sequences used in writing patterns for regular expressions.

Below you find a list of meta-characters used in regular expressions. See Appendix A for more information on the Regular expression syntax, or search the Internet for useful tips. Microsoft's MSDN site provides the full documentation of regular expressions as used in Spotfire.net.

- A period (.) matches any character.
- A caret (^) forces matching at the beginning of an entry.
- A dollar sign (\$) forces matching at the end of a line.
- An asterisk (\*) represents zero or more occurrences of a match of a regular expression. An asterisk following a period will match any string of characters.
- A plus sign (+) matches one or more occurrences of a preceding character.
- A question mark (?) matches zero or one occurrence of the preceding character.
- Square brackets ([]) define a character that matches any single character within the brackets. Z[ABC]Z will match any entry containing the strings ZAZ, ZBZ, or ZCZ.

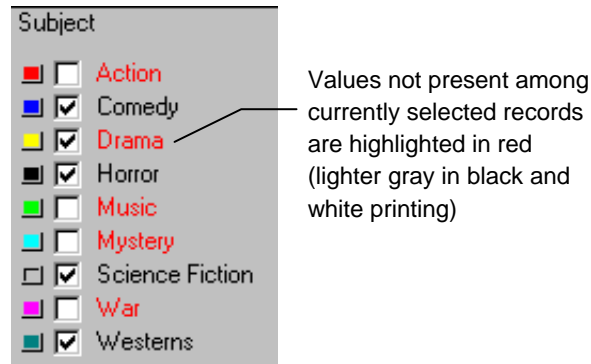
When you write a regular expression, you're actually writing a grammar for a little language. The interpreter takes your grammar and compares it to the entries in the column and returns the records that can be parsed with your grammar.



An ordinary character is a character that matches itself. Any *Regular expression special character*, for instance ], [ , } , { , \* , + , . , \ , can be seen as an ordinary character if it is preceded by a backslash, i.e. if you want to search for records containing a backslash, you would have to type "\\".

## Check boxes

In a *check box query device*, each unique value is represented by a check box, which is either checked (selected in the local context) or unchecked (deselected). If all records with a certain value are deselected (by this or some other query device) the label of that value becomes red.



A check box query device applied to the column *Subject* of the *Film database*. Coloring is set to categorical; ticking a check box causes all records of that particular color to show (unless they are deselected by another query device).

By default Spotfire.net assigns check boxes to any column containing ten values or less. Initially all boxes are checked, which makes all records in the data set visible. For quick checking or un-checking of all the values, right click on the check boxes query device, and select **All** or **None** from the pop-up menu.

Like radio buttons, check boxes provide options that are either On or Off. Check boxes differ from radio buttons in that you typically use check boxes for independent or nonexclusive choices.



## Radio buttons

In a *radio button query device*, a radio button represents each unique value. Radio buttons, also referred to as option buttons, represent a single choice within a limited set of mutually exclusive choices. That is, in any group of option buttons, only *one* option in the group can be set. If all records with a certain value are deselected (by this or some other query device) the label of that value becomes red.



Radio buttons assigned to a column with only two values

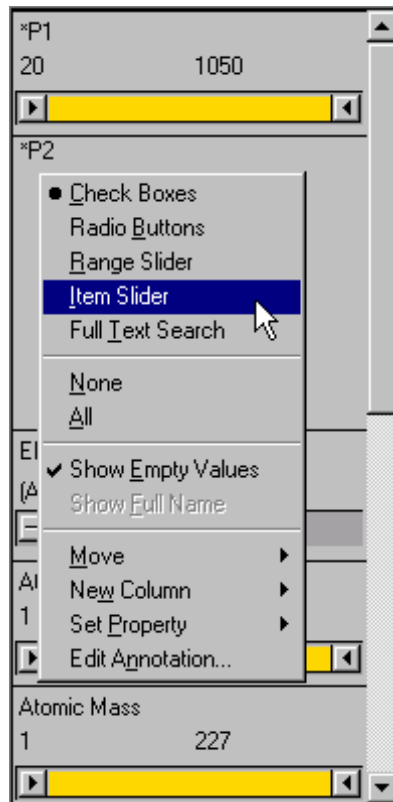
## Initial Query Devices

Spotfire.net's default choice of query devices is based on the column content and the number of unique values present in the data set for that attribute.

- If a column contains 10 unique values or less, check boxes will be assigned as query device.
- For columns containing more than 10 values, an item slider is chosen for alphanumeric (string) attributes, such as names and descriptions. Range sliders are assigned to numeric columns like *date*, *time*, and *decimal or integer values*.

## Changing the Type of Query Devices

As mentioned above, Spotfire.net makes an initial suggestion as to what query device to use for each column in the data set. You can then change the type of query device to use for the column, with one restriction: *check boxes* and *radio buttons* can only be used for columns having less than 500 unique values.



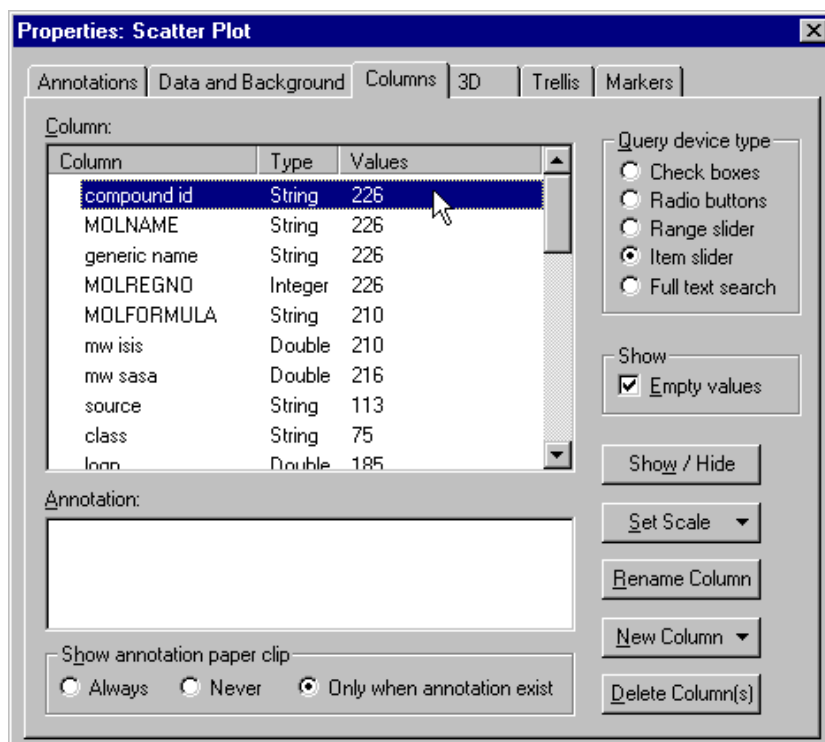
### To change the type of query device:

- Right-click the query device to make the pop-up menu appear. Select the appropriate query device option from the pop-up menu

or

- Select the **Columns** tab of the **Properties** dialog. This tab contains a list of all the columns in the data set. Mark a column and select the type of query device to use for that column.

The query device pop-up menu. The currently selected query device is marked with a bullet



The **Columns** tab of the **Properties** dialog



### 13: Binning

From the Columns tab of the Properties dialog you can also make new columns from expressions or by binning, as well as delete columns from the Spotfire.net internal database.

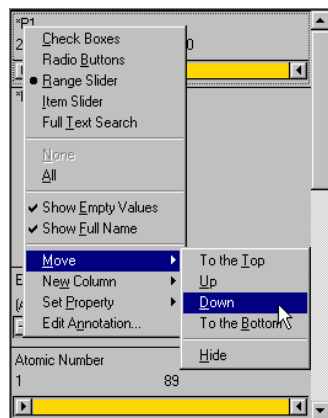
## Moving Individual Query Devices

Working with data sets containing a large number of columns does not necessarily mean that you need to manipulate them all at once. A more useful strategy is to select a few parameters, to see how they interact with each other. For this reason you may want to regroup them and rearrange their order to avoid having to scroll up and down to keep track of the changes.

The initial order of the Query Devices depends on the structure of the data set loaded into Spotfire.net or the SQL query that was used to acquire data. You can change this as needed by rearranging columns in the originating spreadsheet program or writing the SQL query in a certain order.

▶ **To change the order of the query devices:**

1. Right-click on the query device to be moved. The following pop-up menu will open:



Changing the order of the query devices

2. Place the cursor over the **Move** option. A sub menu will appear.
3. Choose the command that reflects the direction in which to move the device. The options **Up** and **Down** move it one step upwards or downwards respectively. The options **To the Top** and **To the Bottom** make the query device the topmost or bottommost device respectively.

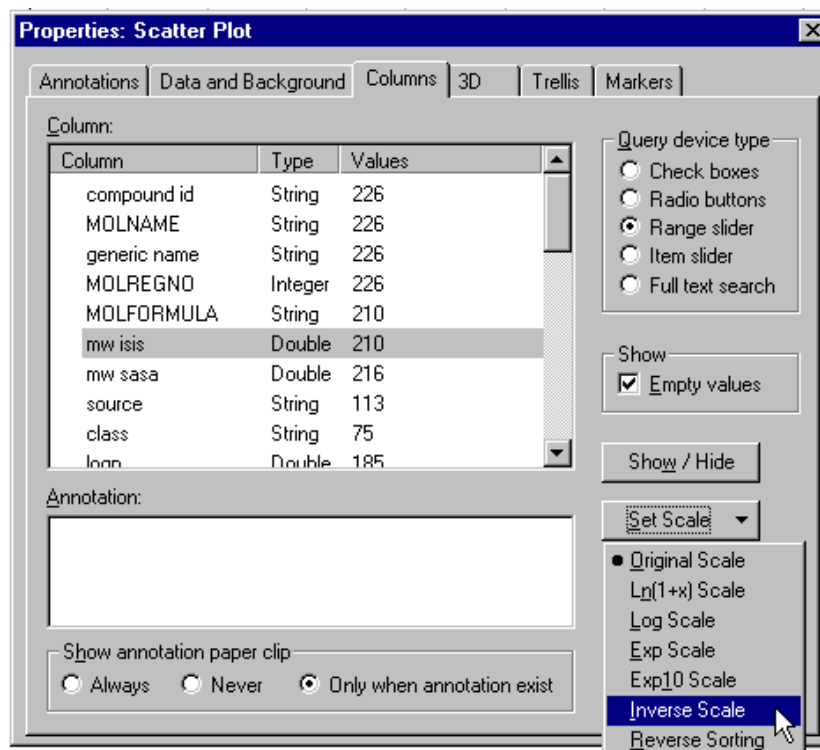
## Hiding a Query Device

▶ **To hide a query device:**

1. Go to the **Columns** tab of the **Properties** dialog.
2. Highlight the desired column name.
3. Click the **Show/Hide** button. A red cross indicates that a query device is hidden. Click **Show/Hide** again to bring the device back.

To hide the Query device window altogether, uncheck **Query devices** in the **View** menu.

## Transforming Axis Scales



Setting the scales from the Columns tab in the Properties dialog

A numerical column can be set to display its values using a variety of scales instead of the linear scale, which is default: Natural Log Scale ( $\ln(x+1)$ ), Logarithmic Scale (Log), Exponential to the bases of  $e$  (Exp) and 10 (Exp10), Inverse Scale ( $1/x$ ), and Reverse Sorting. The last option changes the sort order for the selected column. This is accomplished by right-clicking on the query device representing the column, and from the **Set Property** sub-menu selecting the transformation desired.

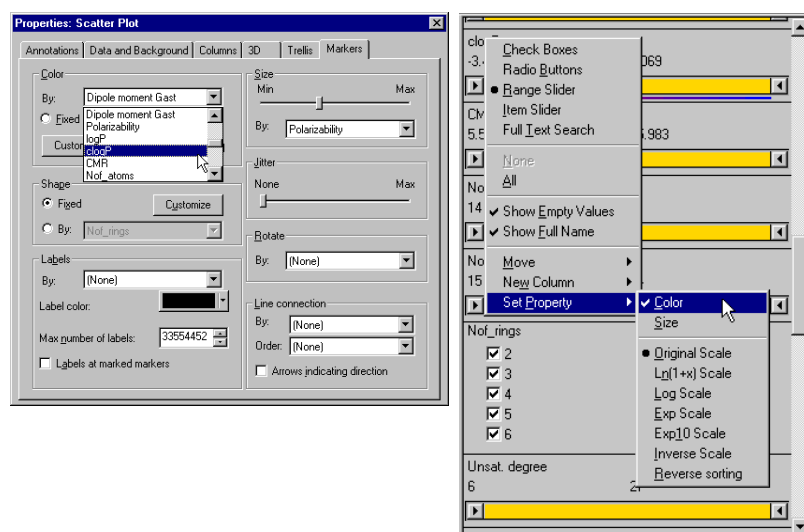
Another way to change the scale is from the **Columns** tab of the **Properties** dialog, by marking the column to re-scale and selecting the transformation from the **Set Scale** pull down box. The transformation can be removed and the scale reset to linear by choosing *Original Scale*.



## 7: Managing Visualizations: Coloring

### Quick Access to Color and Size Properties

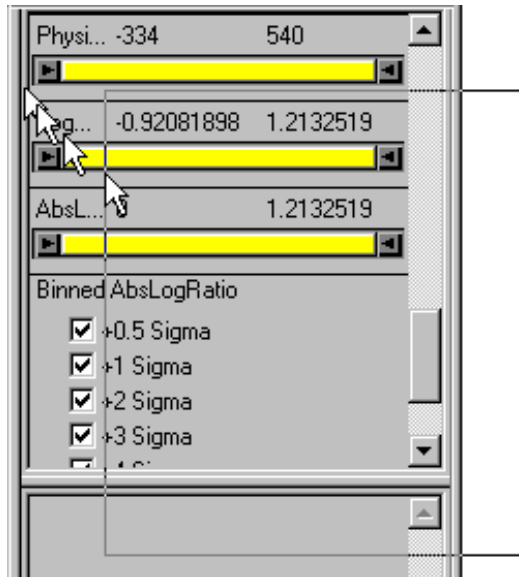
A quick way of setting the color or size property in the current visualization is to right-click on the query device representing the column, and select the appropriate property from the **Set Property** sub menu that appears. If there are ten or fewer values in the column, Spotfire.net will assign categorical colors to the markers. These colors will be very different from each other. If there are many different values in the column the coloring will be set to continuous and the markers colored depending on the value of the record in that column. The colors used can be modified from the **Customize** box of the **Properties** dialog.




Two ways of setting the Color property. Through the **Columns** tab of the **Properties** dialog box (left) and by right clicking the query device associated with the column according to which marker color is to be assigned (right).

## Moving the Query Device Window

The query devices are gathered in a window that can easily be moved around, inside as well as outside the application window.



Grab the frame of the window and reposition it. The area that can be grabbed is around 3 pixels wide, so it requires some accuracy, especially if you have a high-resolution monitor.

If the window with the query devices is placed entirely outside the Spotfire.net application window, it receives a title bar with which it can be repositioned by dragging and dropping. The title bar also has an  at the right-hand edge, which, if clicked, hides the query device window. The devices can be displayed again by checking the **Query Device** alternative under the **View** menu.



# Managing Visualizations



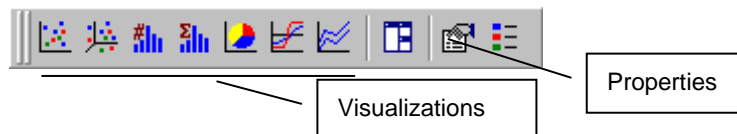
*Spotfire.net can display seven types of visualizations, each one providing a unique view of the data. By setting properties such as colors, shapes and size, each visualization can be tailored to personal taste and the specific task at hand. This chapter provides detailed information on how to configure and manipulate visualizations in Spotfire.net.*

## Overview of Visualizations

Spotfire.net can display the following visualization types:

- 2D scatter plot
- 3D scatter plot
- Bar chart
- Histogram
- Pie chart
- Line chart
- Profile chart

New visualizations are created by clicking one of the toolbar buttons, or selecting a visualization from the **Window** menu. Visualization properties are controlled from the **Properties** dialog. Click the toolbar button, or select **Edit > Properties** to go to this dialog.



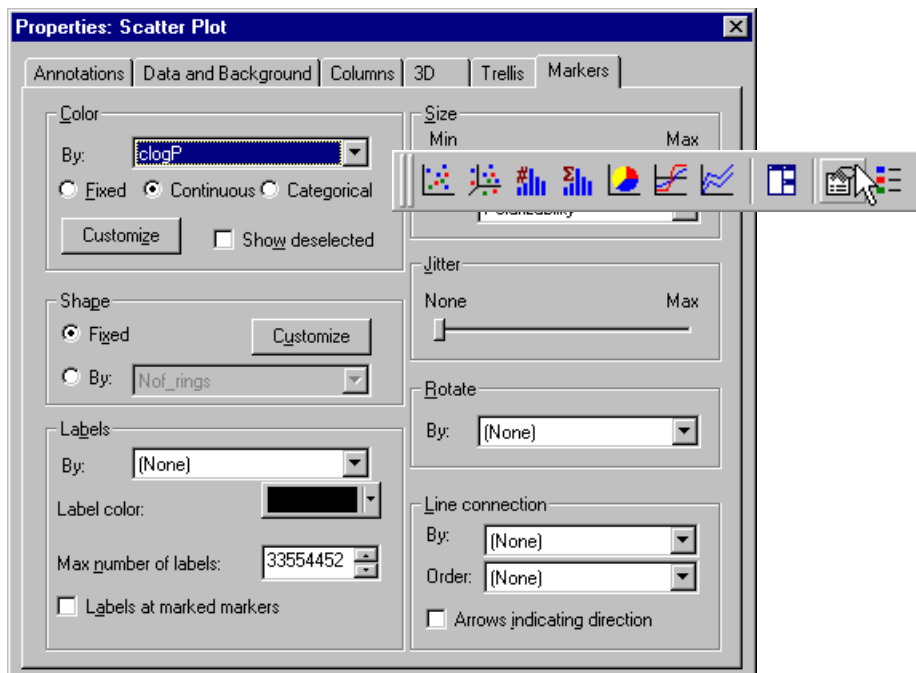
## Initial visualizations

When a data set is loaded in Spotfire.net, the program produces an initial visualization, either a scatter plot or a histogram depending on the number of records in the data set. If there are over 30 000 records, an initial histogram is created.

The columns to be displayed as X- and Y-axes initially are suggested by the program. The algorithm first looks for a column of type *time* or *date*, not containing less than 50 values. Should this fail, it tries to find a column not comprised of strings and that has more than 50 unique values, and finally it looks for *string* columns with more than 50 unique values. For coloring the markers in the initial scatter plot, Spotfire.net first looks for a column with less than 10 unique values and assigns categorical coloring according to that column. If none is found, the program will assign a fixed color to the markers.




Use the **View Tip** command on the **Tools** menu to see miniature visualizations of the columns with the highest correlation plotted against each other.



The Properties dialog

## 2D Scatter Plots

A two-dimensional scatter plot is the default visualization of many data sets. New two-dimensional scatter plots can be created in a number of ways:

- Press [Ctrl+1]
- Click the 2D button on the toolbar: 
- Select **New 2D Scatter Plot** from the **Window** menu

## Zooming and panning

Zooming can be done in two ways: by using the zoom bars or by using the mouse.

| Action   | Mouse maneuver      | Comment                                |
|----------|---------------------|--|
| Zoom in  | Right-click         | The scroll wheel has the same function |
| Zoom out | Shift + Right-click |  |

### Zoom bars

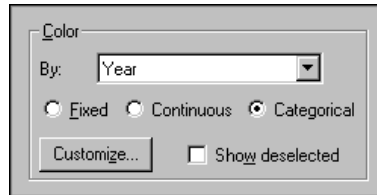
Dragging the end arrows of the zoom bars (along the edges of the visualization window) zooms in on a portion of the visualization. Dragging the bar itself (by placing the mouse pointer on the yellow bar and dragging) pans across different areas of the entire visualization.



Altering the length of the X-axis zoom bar

The pale yellow area represents the selected range of values, whereas the bright yellow area represents the range of existing values within the selected range.

## Coloring



Markers can be colored to reflect the value of a particular attribute. There are three modes for coloring: Fixed, Continuous and Categorical.

### Fixed

Fixed coloring means that all markers are the same color (except deselected, empty and marked).

### Categorical

Categorical coloring means that each value in the chosen column is given its own color. Categorical coloring makes most sense if there are less than ten unique values. To control which color is assigned to each value, click **Customize...**

### Continuous

Continuous coloring means that the maximum and minimum values in the selected column are each assigned a color. Intermediate values are then assigned colors on a scale ranging between the two extreme colors. In scatter plots, any column can be used for continuous coloring.

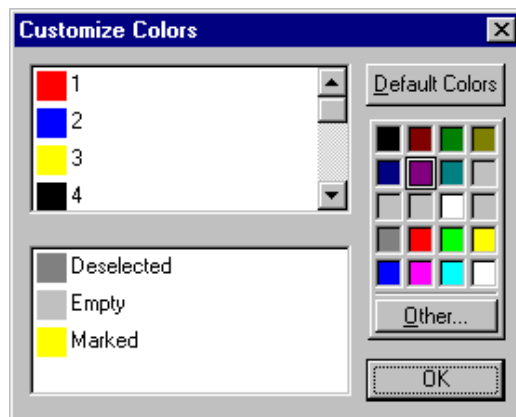
Colors, representing minimum and maximum values are set with the **Customize...** dialog. **Begin** and **End** categories define the color limits. When one of the categories is selected, you can choose which color will represent that end of the value range. A line with the color scale is displayed below the corresponding query device.

### Show deselected

By default, deselected records (i.e. records that have been filtered out using the query devices) are invisible. It is possible to keep them visible, but colored differently. Check the box labeled **Show deselected**. Set the color by pressing **Customize...**

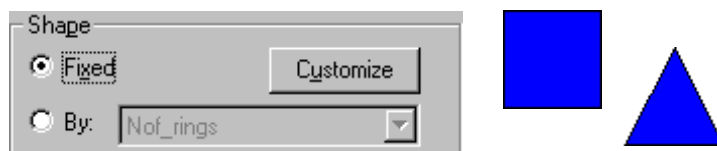
## Customize Colors

Regardless of coloring mode, the choice of colors can be controlled by clicking **Customize...** on the **Markers** tab of the **Properties** dialog. Depending on the current coloring mode, the top-most list will display the fixed color, Begin and End colors (continuous mode), or the color associated to each category (categorical mode). The other list displays colors associated with deselected, empty and marked records. ("Empty" refers to records for which no value is specified in the column used for coloring.)



- ▶ To change a color, click the category that you want to modify, then click a color in the palette.
- ▶ To revert to default coloring, click **Default Colors**. To select a color from the complete palette, click **Other...**

## Marker shape



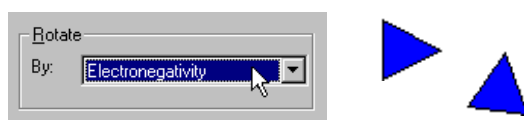
The shape of markers can be fixed, or made to reflect the value of a particular column. Click **Fixed** or **By** to alternate between these modes. Only columns with less than 16 distinct values can be used for controlling shapes. Click **Customize** to choose appropriate shapes for each value.

## Marker size



The size of markers can be made to reflect the value of a particular column. Select a column from the drop-down list under **Size**. Moving the slider changes the size of all markers, while maintaining the size ratio of different markers.

## Marker rotation

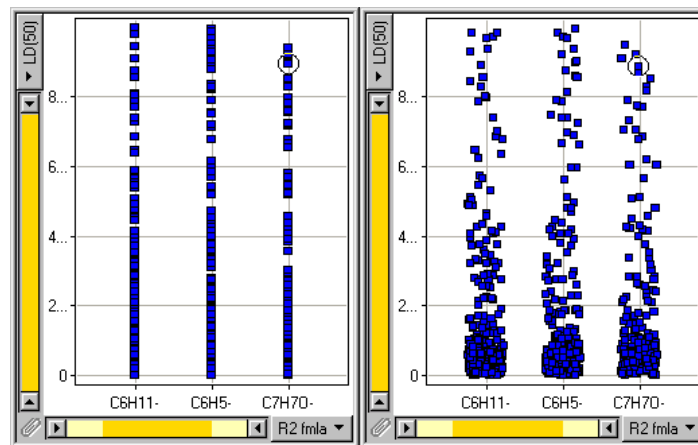


The rotation of markers can be made to reflect the value of a column. Using asymmetrical markers and rotating them according to a chosen variable takes advantage of the human ability to visually detect regions with different surface structure. The maximum rotation is 90° and so the most useful shapes for this effect are the arrow, the minus, and the bar.

## Jittering

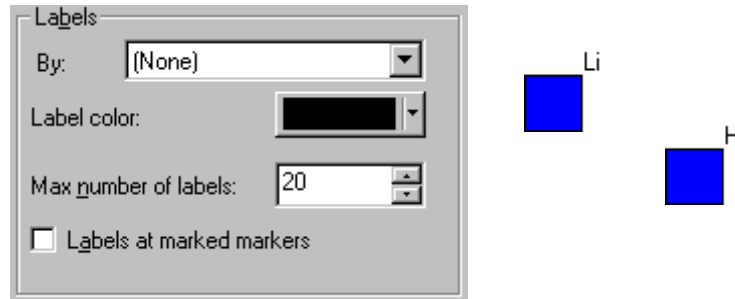


Jittering displaces markers slightly and randomly, thereby making aggregations stand out visually (below, right). Attention is brought to areas where many records overlap. Such regions can then be investigated further, by zooming, changing axes, etc. Use the slider to increase or decrease jittering.



Jittering the data set (right) reveals density of markers

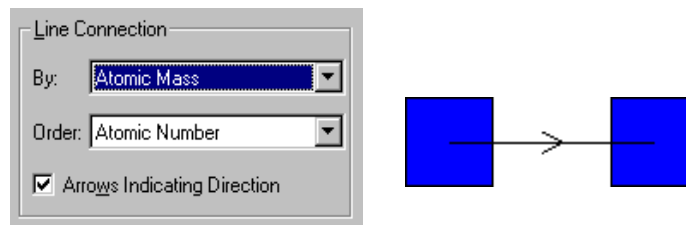
## Marker labels



It is possible to tag each marker with a label, showing the value of a particular column.

To avoid cluttering the visualization, a limit to the number of labels can be specified. Note that if this number is less than half the total number of visible markers, then no labels are shown at all.

## Line connection



Line connection means that a line is drawn between records that share some property. To follow a progression, markers can also be connected in ascending order according to some other property. The order of lines can be made clearer by use of arrows indicating direction (pointing from lower to higher value in the **Order** column).



7: Managing Visualizations:  
Layout using Trellis plots

## Trellis plots


You can break up a scatter plot into multiple panes, and organize them in a grid layout using the Trellis tab of the properties dialog. See page 86 for more information.



## 3D Scatter Plots

3D scatter plots allow even more information to be encoded into visualizations. They are especially useful when analyzing data that is not clustered along any of the axes (columns) of the data set.





Make a new 3D scatter plot in one of the following ways:

- Press [Ctrl+2]
- Click the 3D button on the toolbar: 
- Select **New 3D Scatter Plot** from the **Window** menu

3D visualizations have the Z-axis selector and zoom bar along the top edge of the visualization. The Z-axis initially points towards the viewer, out from the screen.

## Navigating 3D visualizations

In 3D, the zoom bars are used as in 2D. Additionally, the following means of navigation are available.

| Action                                      | Mouse maneuver            | Pointer icon  | Comment  |
|---|---------------------------|---|--|
| Rotate                                      | Right-drag                |  | Rotation is around the visualization center  |
| Zoom  | Shift + Right-drag        |  | Moves in the direction of view   |
| Roll  | Shift + Ctrl + Right-drag |  | Rotation is in the plane of the screen; downwards and right cursor motion rotate clockwise |
| Zoom while maintaining size of bounding box | Ctrl + Right-drag         |  | Keeps the selected marker in the same place on screen                                      |

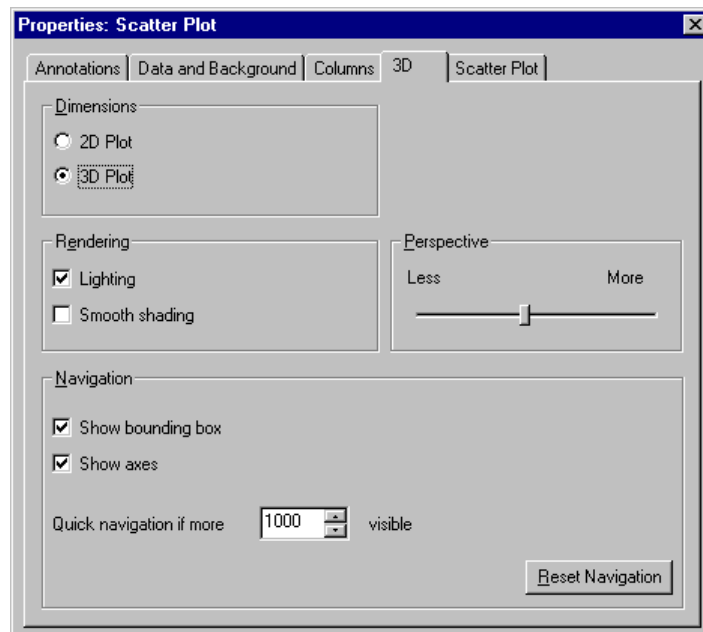
To zoom in on a specific marker, highlight it by clicking on it. Use the **Ctrl** key while you right-click with the mouse and drag vertically in the window (upward zooms in, downward out).



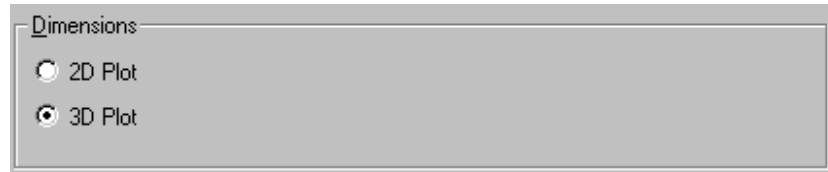
Clicking the **Reset Navigation** button on the **3D** tab of the **Properties** dialog will undo all navigation adjustments.

## The 3D tab

3D scatter plots can be controlled using the **Markers** tab, which is largely similar to that of 2D scatter plots. Additionally, the **3D** tab of the **Properties** dialog provides settings that are specific to 3D scatter plots, several of which have direct bearing on the speed and smoothness of the presentation of data.



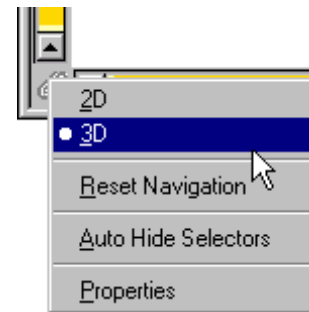
## Dimensions



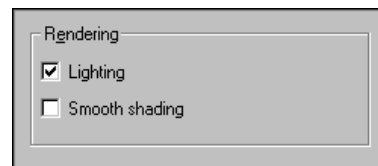
A 3D visualization can be ‘flattened’ into a 2D scatter plot or, conversely, a 2D visualization converted to 3D. In the latter case, Spotfire.net will suggest a third axis using the same algorithm as axis selection for 2D scatter plots.



By right-clicking in the bottom left-hand corner (or anywhere in the gray area surrounding a visualization) you will bring up a pop-up menu that makes it easy to switch between 2D and 3D.



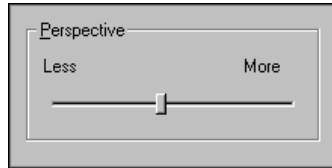
## Rendering



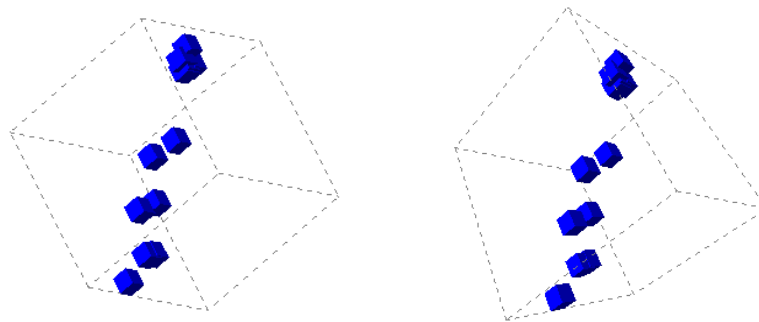
With the **Lighting** box unchecked, the markers will be flat, as if illuminated evenly from all sides, a sphere will appear as a filled circle while a cube from certain angles will look like a hexagon and from other angles like a square.

The **Smooth shading** alternative will diminish the faceted appearance of the glyphs and thus make the surface appear smoother; mostly useful when the filled sphere is used; this makes the sphere appear rounder.

## Perspective



The perspective slider modifies the ‘field of view’ of the 3D visualization. Using the mouse, drag the slider towards the right for a ‘wide angle’ perspective. With the slider in the left-most position the view becomes orthogonal, a projection where the size of the markers is retained no matter the distance. The keyboard arrow keys allow fine-tuning of the perspective when the slider has been activated.



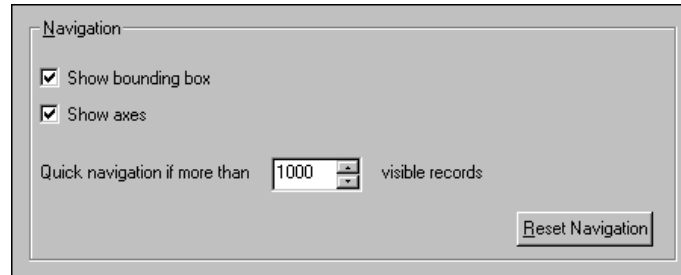
Orthogonal view (left) and exaggerated perspective (right).



When viewing 3D visualizations, perspective will make markers that are closer to the viewer look bigger. Therefore, care must be taken if you also want size to reflect some attribute. Orthogonal projection where markers retain size irrespective of distance may be preferable.

---

## Navigation



**Show bounding box** draws a dotted cube around the edges of the visualization volume, which often aids in locating a marker in the 3D space.

**Show axes** draws the arrows representing the X, Y and Z-axes.

**Quick navigation if more than ... visible records** lets you speed up rotation, zooming, etc by drawing markers as dots while navigating. Markers are restored to their normal shapes as soon as you stop navigating. Use this feature if the lag times become too great during 3D navigation.

The **Reset Navigation** button resets the orientation of the visualization—X (horizontal), Y (vertical), and Z (depth)—but does not affect zooming or perspective.



The Reset button in the toolbar will reset not only navigation but also all query devices.



Pg. 86:  
Layout using  
Trellis plots


## Trellis plots

You can break up your scatter plot into multiple panes, and organize them in a grid layout using the **Trellis** tab of the **Properties** dialog.


## Histograms and Bar Charts

Histograms and bar charts can effectively analyze very large data sets. Histograms are the default visualization when the data set is comprised of more than 30 000 records.

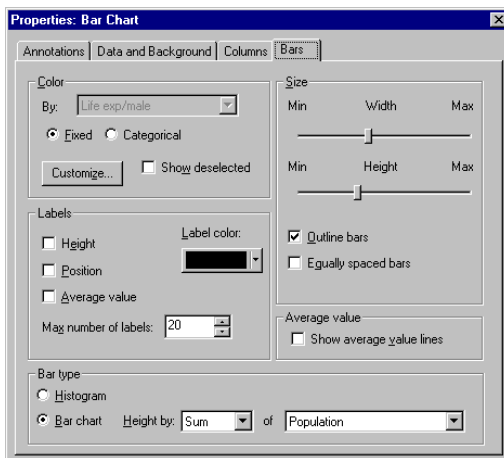
New Histograms are made in one of the following ways:

- Press [Ctrl+3]
- Click the Histogram button on the toolbar: 
- Select **New Histogram** from the **Window** menu

Bar charts are created in the following manners:

- Press [Ctrl+4]
- Click the Bar Chart button on the toolbar: 
- Select **New Bar Chart** from the **Window** menu

In traditional bar charts the height of the bar is the sum of the values of the records in a certain column. In histogram-type visualizations heights of bars are proportional to the number of records specified by the X-axis column. The attributes of both histograms and bar charts are altered from the **Bars** tab in the **Properties** dialog.

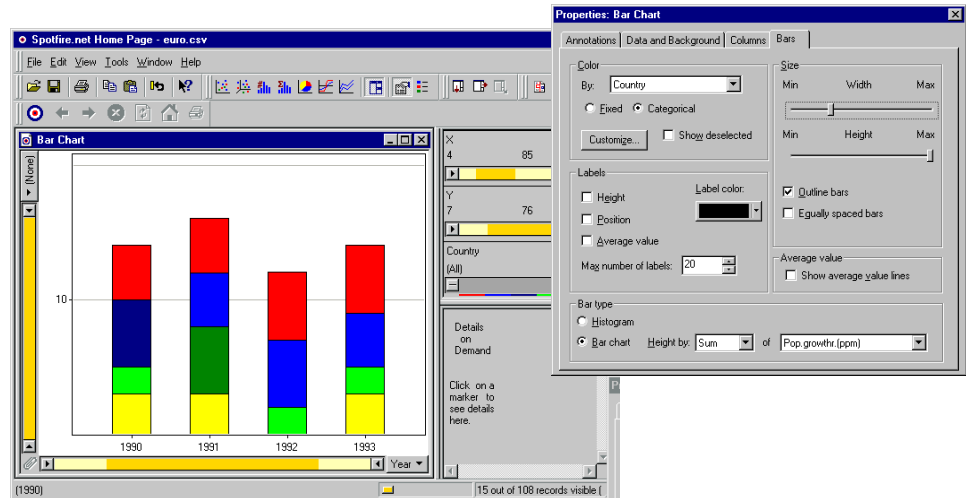


The **Bars** tab of the **Properties** dialog box

## Coloring

### Categorical

When using categorical coloring for histograms and bar charts, the bars are divided into sections according to the relative content of the column selected under **Color By**.



Bar chart colored to show relative contents

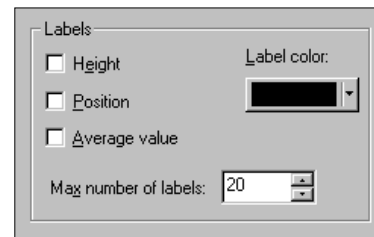
The categorical coloring scheme and the colors for empty and deselected records can be altered in the same manner as for 2D scatter plots (see page 57).

### Continuous

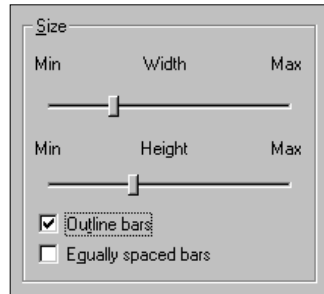
For histograms, bar charts - and pie charts - continuous coloring can be used only for columns with numerical content. Each bar will be colored according to the average of the values in the selected column.

### Labels

Under **Labels**, in the **Properties** dialog box, it is possible to choose to display **Height**, **Position** and **Average** values as labels next to the bars. The color of the label text is also selected here.

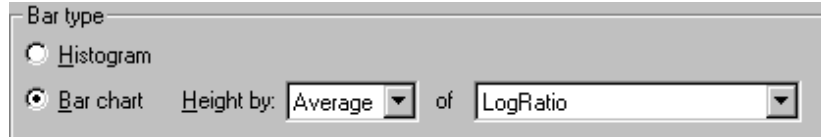


## Size



Width and height are controlled using the two sliders. The **Outline bars** option displays the bars with a thin black line around their periphery. The **Equally spaced bars** option, when checked, will place all bars at equal intervals along the axis irrespective of the values in the column; i.e. if the values on the X-axis are 2, 3, 5, 16, 18 and the box unchecked, the five columns will be visualized with a large gap in the center. When checked, the bars will be dispersed evenly along the baseline.

## Bar types




The radio buttons allow you to switch between bar chart and histogram. In histograms, the height represents the number of records in the bar. With bar charts you can also select which column to visualize, and which aggregation method to use (sum or average).

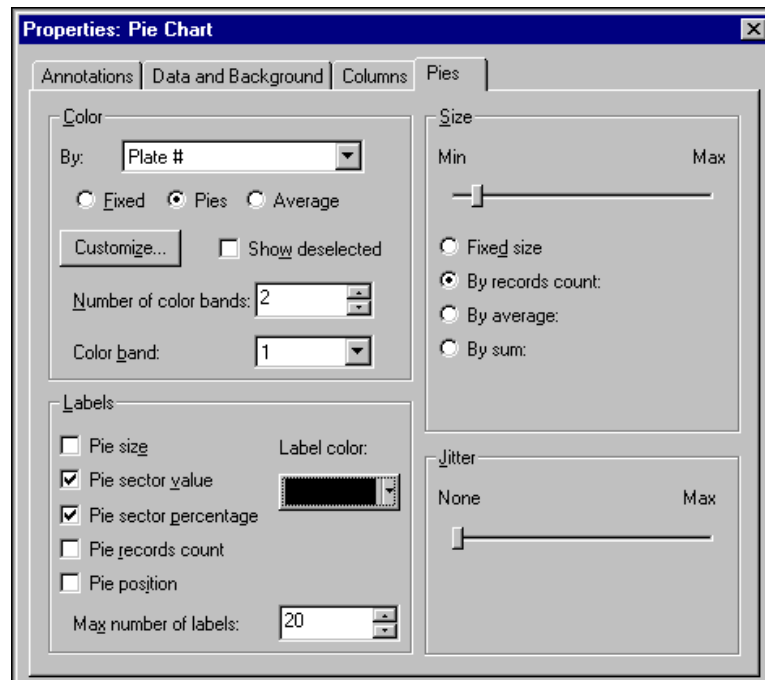


## Pie Charts

New pie charts are made in one of the following ways:

- Press [Ctrl+5]
- Click the Pie chart button on the toolbar: 
- Select **New Pie Chart** from the **Window** menu

Pie chart properties are set using the **Pies** tab of the **Properties** dialog:

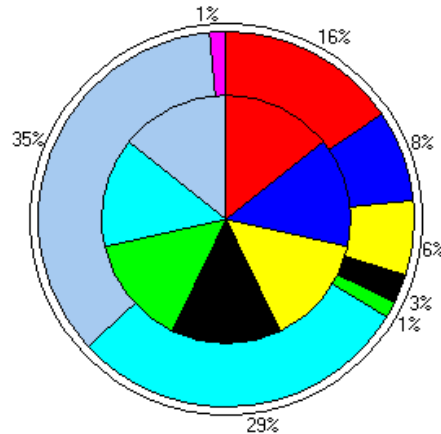


The **Pies** tab of the **Properties** dialog

A pie chart visualization can consist of one or more pies. To generate multiple pies, change the X-axis and/or Y-axis selectors to anything but **None**.

## Coloring

A pie chart can consist of one or more bands, each reflecting the distribution of values in a certain column. Set **Number of color bands** to the desired number, then select one band at a time using **Color band**. Color settings are set for each band individually.



Multiple bands in a pie chart.

For whole pies or bands, three coloring modes are available:

1. **Fixed** – One color is assigned to the entire pie or band.
2. **Pies** – The normal way of coloring: the band to be colored is divided into sections whose size is determined according to the distribution of the selected column.
3. **Average** – Assigns one color to each band or pie. The color is given according to the average value of the records in the chosen column for that pie/band.

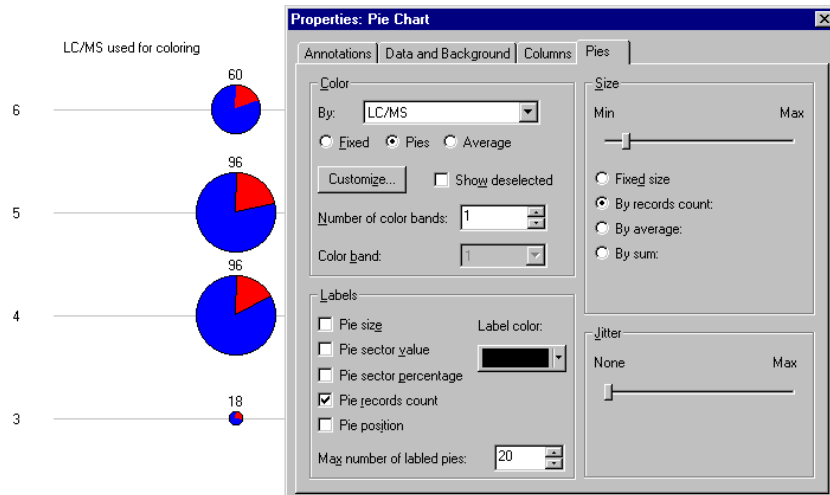
Use the **Customize...** button to choose color. The colors can be customized in the same way as when working with scatter plots—see page 57.

Use **Show deselected** if all records in the data set should be included in the pies. The deselected records are displayed as a separate sector. The color of this sector can be altered with the color selector under **Customize...** The color of the sector representing empty markers and the marking color can be changed in the same way.

## Size

The maximum size of the pies in the pie chart can be varied with the **Size** slider. Under the **Size** slider there are four radio buttons with the following functions:

1. **Fixed size** – pies are sized all the same.
2. **By count** – pies are sized according to the number of records in each pie.
3. **By average** – pies are sized according to the (absolute) average value of the records in the pie.
4. **By sum** – pies are sized according to the (absolute) sum of the records in the pie.

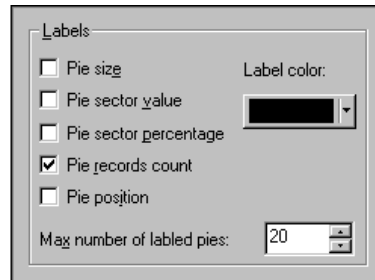


Pies sized by count showing the value above the pie



When high *negative* values are used for sizing (By average or By sum), pies will appear large. A red contour will indicate that the value used for sizing is negative. If the sum or average is zero, this will yield invisible pies.

## Labels



Five different labels can be added to the pie charts:

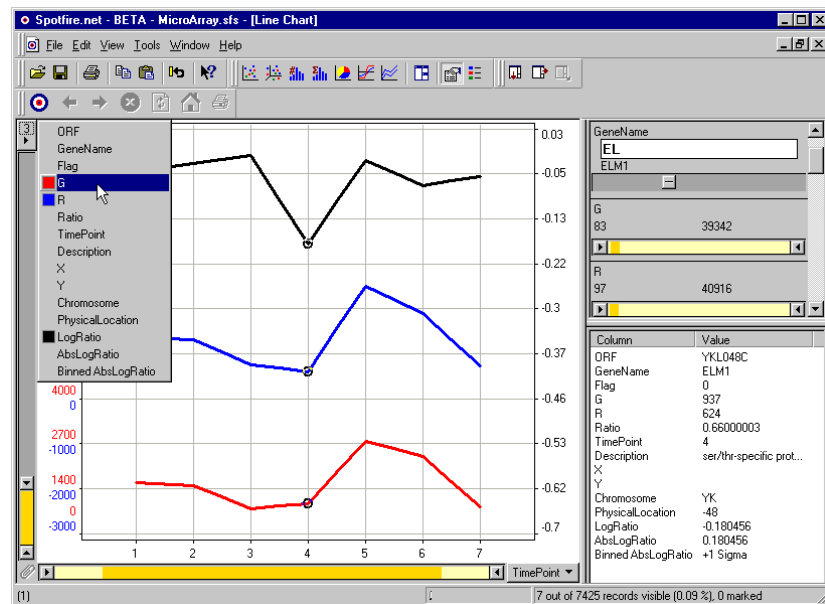
1. **Pie size** – shows the number of records associated with the pie, or average or sum, depending on the settings under **Size**.
2. **Pie sector value** – shows the value associated with each sector in the outermost band.
3. **Pie sector percentage** – shows the percentage of the pie that each sector covers in the outermost band.
4. **Pie records count** – shows the number of records in the pie.
5. **Pie position** – shows (in brackets) the position of the pie. Requires the use of axes.

## Jittering

The **Jitter** function, at the bottom right of the **Properties** dialog, can be used in the same way with pie charts as when working with scatter plots. The overlapping pies will be randomly displaced so that hidden information will be made visible. At least one axis has to be in use.


## Line Charts

Line charts are useful for visualizing data that behaves like a function with respect to the variable on the X-axis, for example time series data. This visualization type is ideal for showing trends or changes in data over a period of time, at even intervals. Line charts emphasize time flow and rate of change rather than the amount of change.

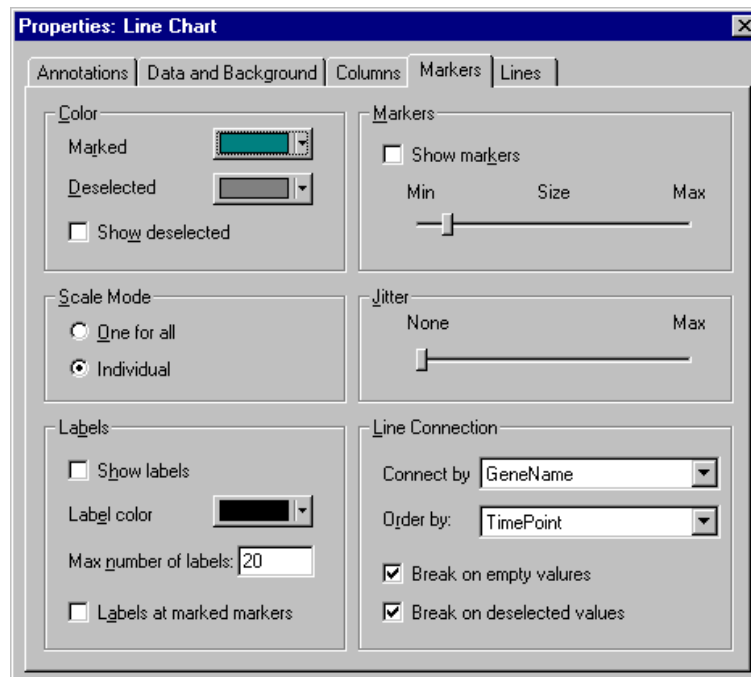


Line Chart visualization with multiple Y-axes

You can create a line chart in any of the following ways:

- Press [Ctrl+6]
- Click the Line Chart button on the toolbar: 
- Select **New Line Chart** from the **Window** menu

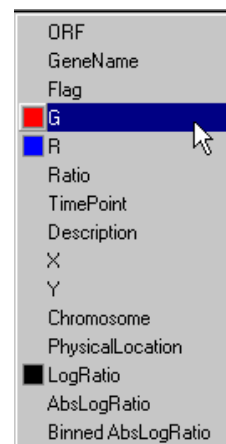
Line charts are controlled from two tabs in the **Properties** dialog; the **Markers** tab, which controls the overall properties of the plot, and the **Lines** tab on which the properties of the individual lines are edited.



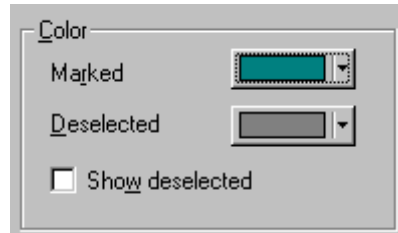
The **Markers** tab of the Line Chart **Properties** dialog

### Selecting columns to plot on the Y-axis

The axis selector of the line chart differs from other Spotfire.net views in that it is possible to select more than one column to plot against the Y-axis. The axis selector for the Y-axis therefore lets you select any number of columns to be visualized. Each of the selected columns will have a colored square next to it in the axis selector menu.



## Coloring



The colors used for drawing marked and deselected records are edited here. Whether deselected records should be drawn or not is determined with the **Show Deselected** check box.



Note that the color used to plot the records is the same as the connecting line, and thus configured on a per column basis on the **Lines** tab.

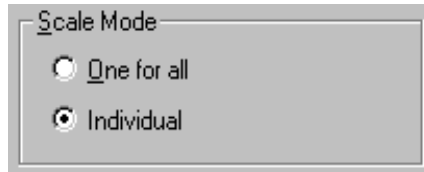
---

## Labels



For line charts, labels show the values of the column being plotted. Labels are turned on by clicking the **Show Labels** check box. Otherwise functionality is the same as for scatter plots—see page 60.

## Scale mode



The line chart works in two modes: one in which all columns are plotted against a common Y-axis, and one where each column has a Y-axis of its own.

### One for all

In this mode there is a single Y-axis scale for all columns. The range of this axis is automatically adjusted to fit all the columns currently being visualized.



Since the values of the plotted columns may be of different magnitude, some lines might not be visible due to their low amplitude in relation to the scale of the y-axis. Select Individual scale mode, or use the zoom slider on the y-axis to zoom in on curves with low amplitude.

---

### Individual

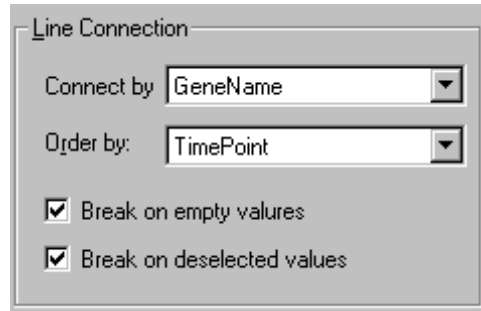
In this mode each column has a Y-axis scale of its own. This is useful when you want to visualize data of different types and magnitude in one plot, or when the shapes of the curves are more interesting than the absolute values. In Individual mode you can control the vertical position and scale of each plotted column using the sliders on the **Lines** tab.

## Markers

The **Markers** check box determines if markers should be drawn or not. The size of markers is controlled with the **Size** slider. The shape and color of the markers are controlled on a column basis on the **Lines** tab.

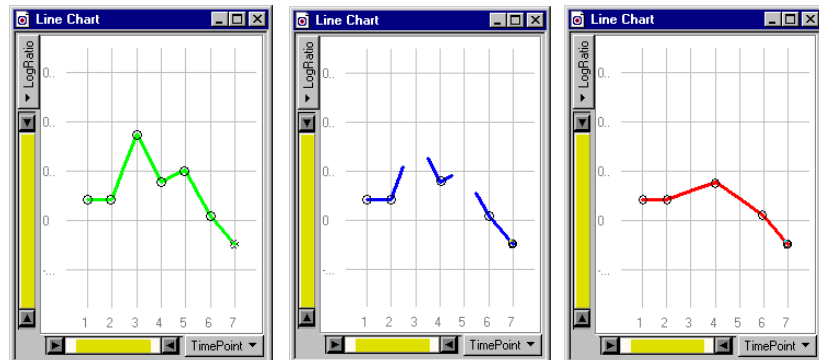


## Line connection



You can choose which column to use for discriminating between lines (**Connect by**), and the order in which data points are connected within each line (**Order by**). The order can be set to any column, to *Row Number*, or to *None*. The column on the X-axis is used by default. Selecting *None* turns line connection off.

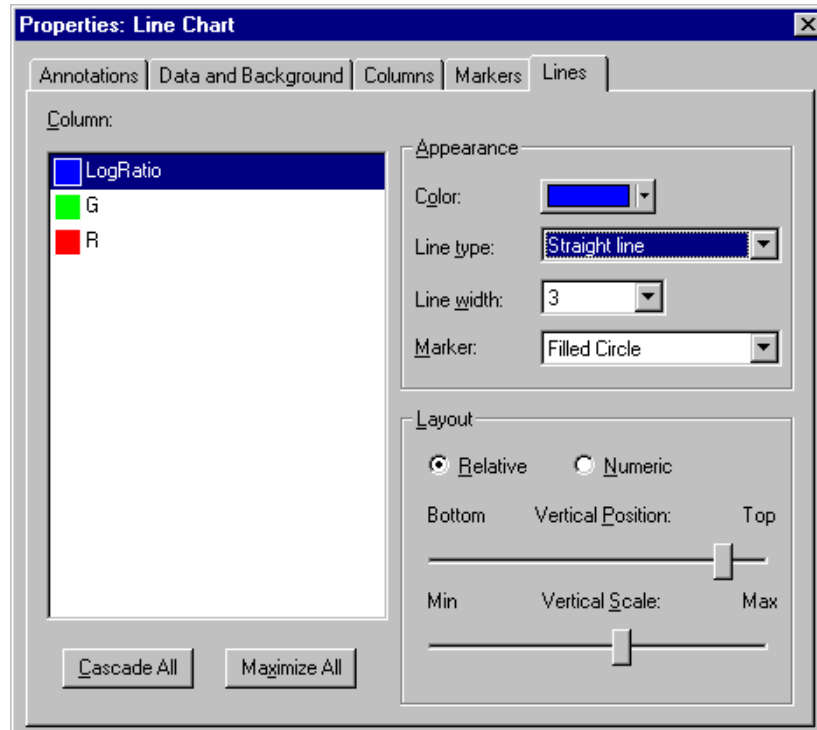
Line connection works in a slightly different way for line charts compared to scatter plots. In scatter plots, points are connected no matter how many intermediate points are deselected (filtered out). In a line chart, you can control how you want to handle values that are missing or deselected.



In the first chart, no values are deselected. In the second, high values have been deselected and **Break on deselected values** has been turned on. Broken lines then extend half way to the deselected value. In the third chart, high values have been deselected and **Break on deselected** values turned off.

## Appearance

On the **Lines** tab you set the properties for each column, i.e. each line. Select the column you want to edit, and view the properties in the list boxes on the right. Adjusting the controls will change the appearance of the selected column.



The controls on the **Lines** tab determine how the selected column will appear in the Line chart

## Color

The color used for this column. The column is used for lines and markers as well as for axis labels in the Individual Scale mode.

### Line type

Determines the type of the line used to connect points. Alternatives:

- **None** – no line will be drawn
- **Straight line** – points will be connected with a straight line
- **Step** – points will be connected with a right angle. The first line is horizontal

### Line width

Indicates the width of the line in pixels. The default setting is 1.



Thick lines take considerably longer time to draw—therefore when viewing large data sets, it's a good idea use a low value for Line width.

---

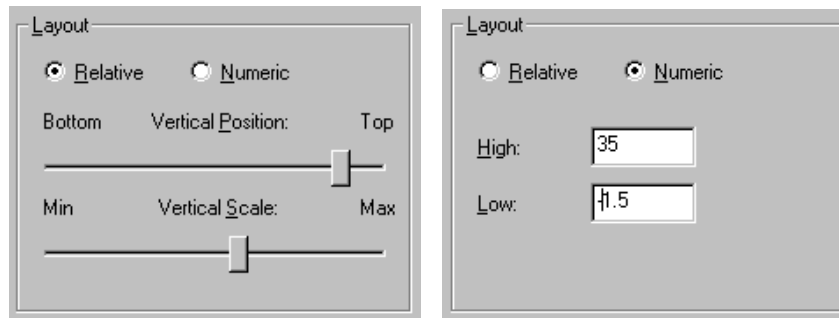
### Marker

Allows you to select a shape to use for markers. If **Marker** is set to None, no markers are shown in the plot.

Please note that no markers will be drawn unless **Show Markers** is turned on in the **Markers** tab.

## Layout

When **Scale Mode** is set to *Individual* (see **Markers** property page) it is possible to control the vertical position and amplitude of the plotted curves. This makes it easy to compare the shapes of the curves. Position and scale can be set either using sliders, or by entering numerical values. You can at any time switch between the two modes by pressing either of the two radio buttons labeled Relative or Numeric.

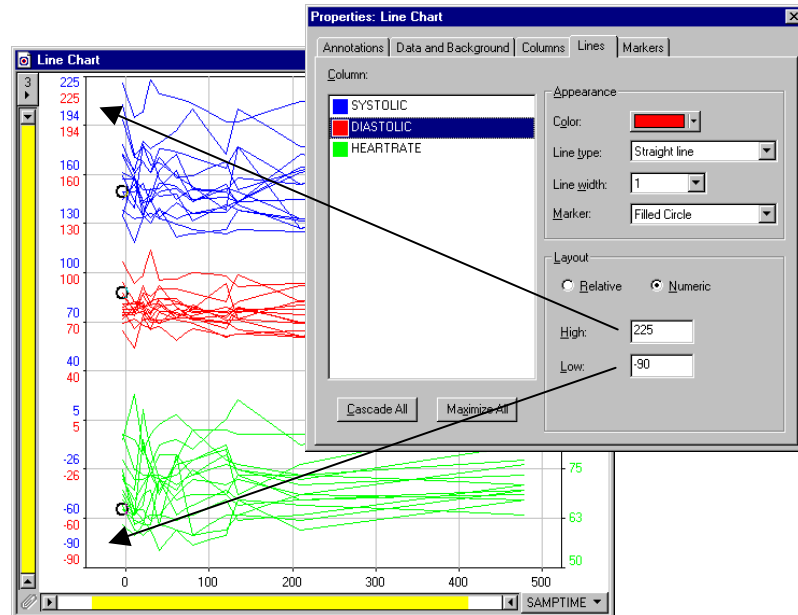


### Vertical Position (Relative)

Moves the selected curve up and down in the visualization window. This may be desirable when you want to compare two curves by superimposing them on one another.

### Vertical Scale (Relative)

Scales the selected curve. If you move the slider to *Min* the curve will appear almost flat. When set to *Max* the curve amplitude will equal the window height of the visualization.



### High (Numeric)

Sets the highest value of the visible range.

### Low (Numeric)

Sets the lowest value of the visible range.

### Cascade all

Sets the Scale and Position of all columns so that each of them will occupy a separate slot of the window height.

### Maximize all

Set the Scale and Position on every column so they utilize the whole height of the window.

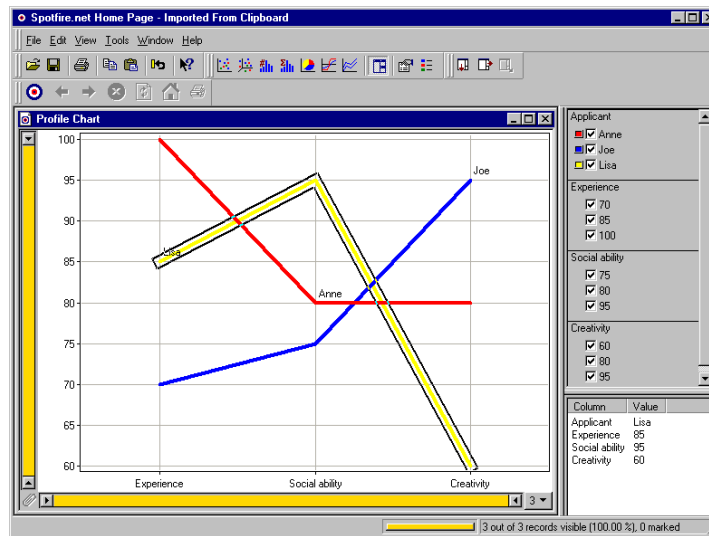
## Profile Charts

A profile chart maps each record as a line, or *profile*. Each attribute of a record is represented by a point on the line. This makes profile charts similar in appearance to line charts, but the way data is translated into a plot is substantially different.

Consider, for example, a data set with three different applicants for a job position. Through tests or interviews, the recruiter may have provided percentage estimates of each candidate in terms of experience, social competence, and creativity. The data may look as follows:

| Applicant | Experience | Social competence | Creativity |
|-----------|------------|-------------------|------------|
| Lisa      | 85         | 95                | 60         |
| Joe       | 70         | 75                | 95         |
| Anne      | 100        | 80                | 80         |

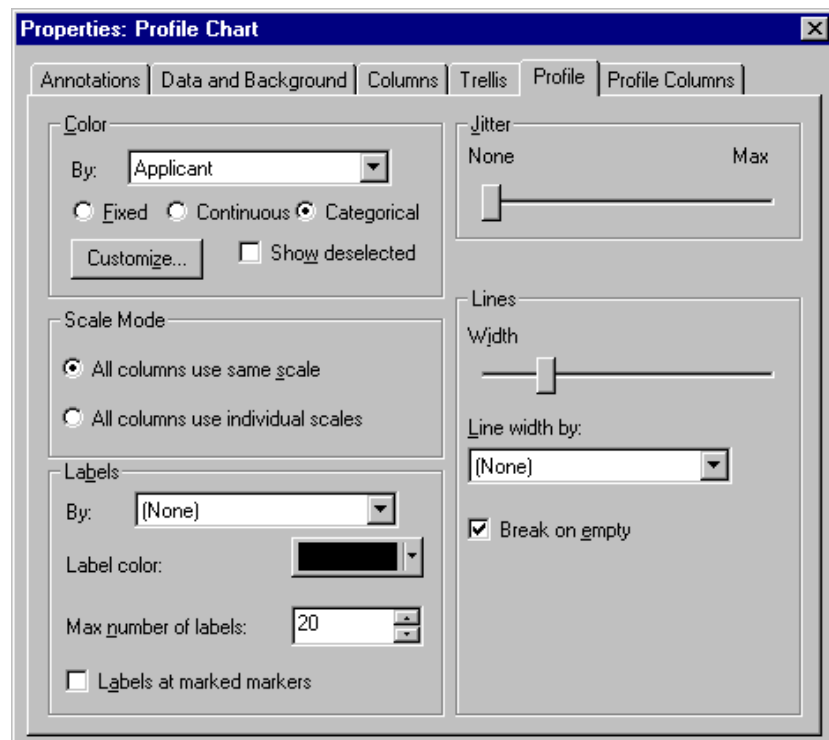
For each applicant, it is now possible to plot a personality profile. This will give a good (if somewhat simplified) overview of characteristics, and a possibility to visually match an applicant against a "model profile" for the job. The profile chart would look as follows:



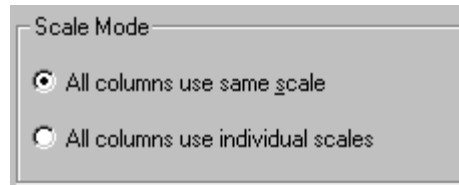
Profile charts are perfectly suited for investigating time-related parameters, where each record consists of an identifier and a sequence of values (with a column for each time-point).

► **To create a profile chart:**

1. Press [Ctrl+7], or click the **New Profile Chart** button on the toolbar, or select **Window > New Profile Chart**.
2. Go to the axis selector of the X-axis and uncheck solitary columns that you do not want to include in the chart such as identifier columns, or go to the **Profile Columns** tab if you want to change multiple columns.
3. Use the **Properties** dialog to adjust the various properties of the chart.



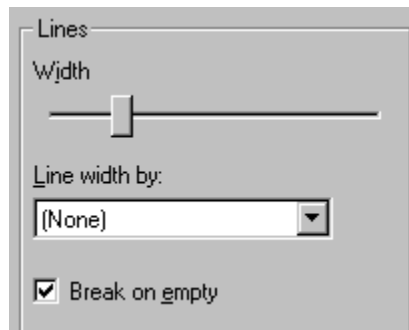
## Scale mode



The vertical dimension in a profile chart represents the value of each attribute. The scale can either be the same for all columns, or individual for each column.

In the first case, the scale used for all columns ranges from the lowest to the highest value in *all* columns being plotted. This should be used when the values in the plotted columns are of the same unit and similar magnitude. To use this mode, go to the **Profile** tab of the **Properties** dialog, and select **All columns use same scale**. In the second case, the scale for each column ranges from the lowest value in the column to the highest value in the column. To use this mode, select **All columns use individual scales**.

## Lines



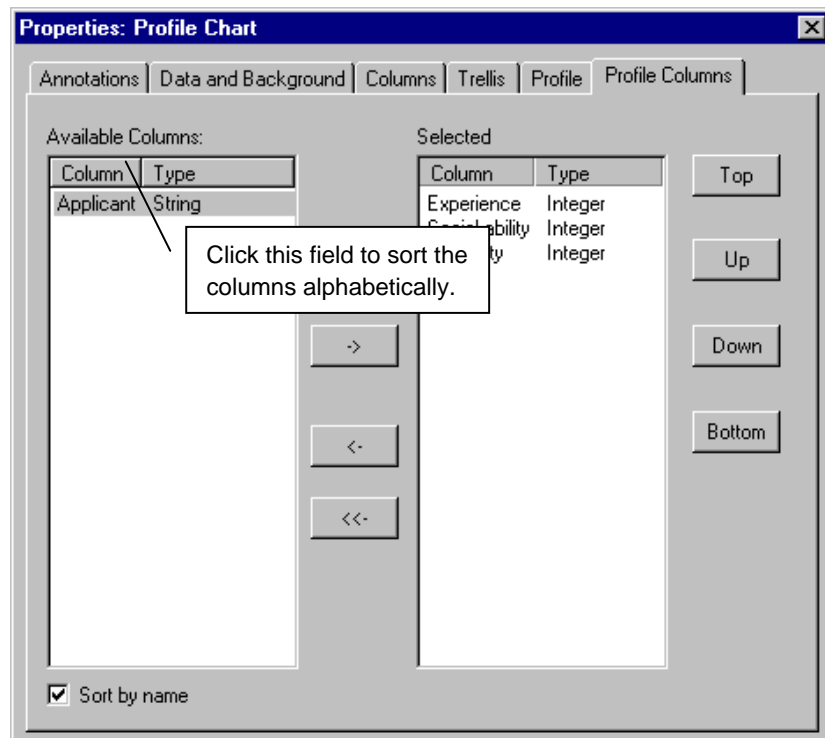
The width of a line is adjusted by moving the **Width** slider. Select a column under **Line width by** to associate the width of individual lines with the value of a particular attribute.

Empty values can be dealt with in two ways: either by leaving a gap in the profile, or by connecting the values on each side with a straight line. To leave a gap, check **Break on empty**.



## Profile columns

The **Profile Columns** tab of the **Properties** dialog is used to organize the columns in the profile chart. The list on the left-hand side shows the columns that are not included in the profile, while the one on the right shows those that are. Use the arrow buttons to move column names between the two lists. The double arrow moves all the selected columns back to the list of available columns.



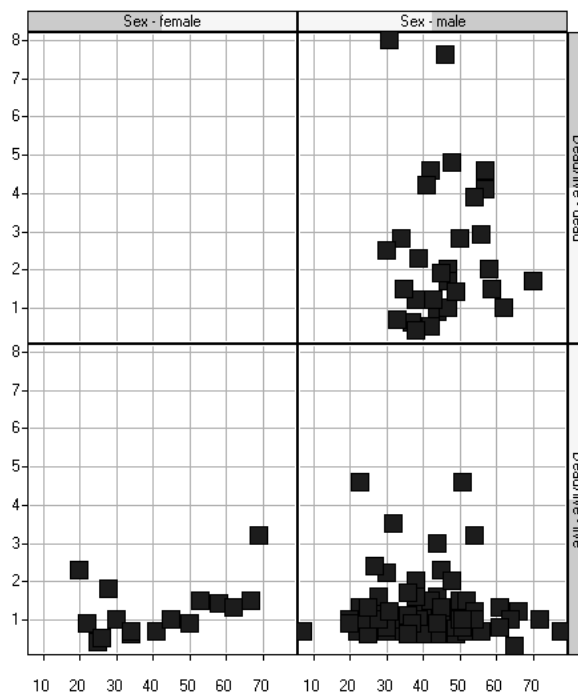
## Sorting columns

The list of available columns can be sorted by checking the box labeled **Sort by name**, or by clicking the **Column** field in the column heading.

The list of selected columns can be sorted manually by moving individual columns. Select a column and click **Top**, **Up**, **Down** or **Bottom** to alter the position of a column in the list.

## Layout Using Trellis Plots

Trellis plots provide the means to create a systematic layout for scatter plots and profile charts. Each individual pane in a Trellis plot displays a subset of the original data set, where the subset is defined by certain *conditions*. By this we mean that all the records in a particular subset have the same value for one or more attributes. A condition specifies one attribute.



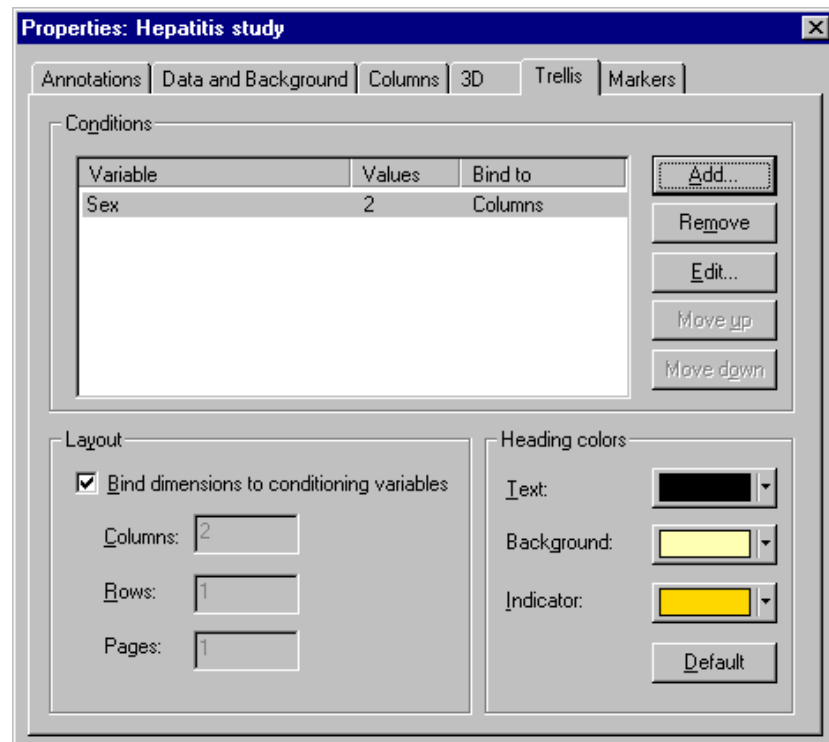
A Trellis plot with two conditions: Sex and Dead/live

Trellis plots enable the user to quickly recognize similarities or differences between different subsets.

► **To create a Trellis plot:**

1. Create a new scatter plot or profile chart.
2. Open the **Properties** dialog.
3. Go to the **Trellis** tab.

4. In the **Conditions** box, click **Add...**
5. In the **Trellis condition** dialog, select the variable by which you want to separate the plots.
6. Click **Ok**.



## Conditions

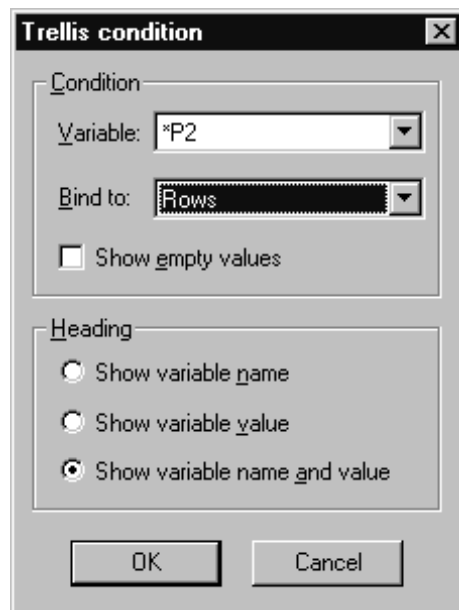
A condition specifies a variable by which the data is divided into subsets. Each value of the conditioning variable will yield a separate subset. Adding a second condition will split each subset, etc.

If, for example, two Boolean variables are used as conditioning variables, this will result in four separate plots representing the combinations True-True, True-False, False-True and False-False. If a Boolean variable is used in conjunction with a variable that has five different values, this will yield ten plots. From this follows that variables with a continuous distribution and a wide range of values (e.g. Real

values) should be binned before forming a Trellis plot. Otherwise the number of plots quickly becomes unmanageable.

▶ **To add a condition:**

1. In the **Scatter Plot Properties** dialog, under the **Trellis** tab, click **Add...**
2. In the **Trellis condition** dialog, select a variable and a dimension to bind it to.
3. Optionally, use the radio buttons in the **Heading** box to control what information to show in the heading for this variable.
4. Click **OK**.



### Empty values

Checking the **Show empty values** option will generate an extra row, column or page (depending on what the variable has been bound to) containing records for which the conditioning variable is undefined.

▶ **To remove a condition:**

1. Select a condition from the list on the **Trellis** tab.
2. Click **Remove**.

▶ **To modify an existing condition:**

1. Select a condition from the list.
2. Click **Edit...**
3. In the **Trellis condition** dialog, make the desired changes.
4. Click **OK**.

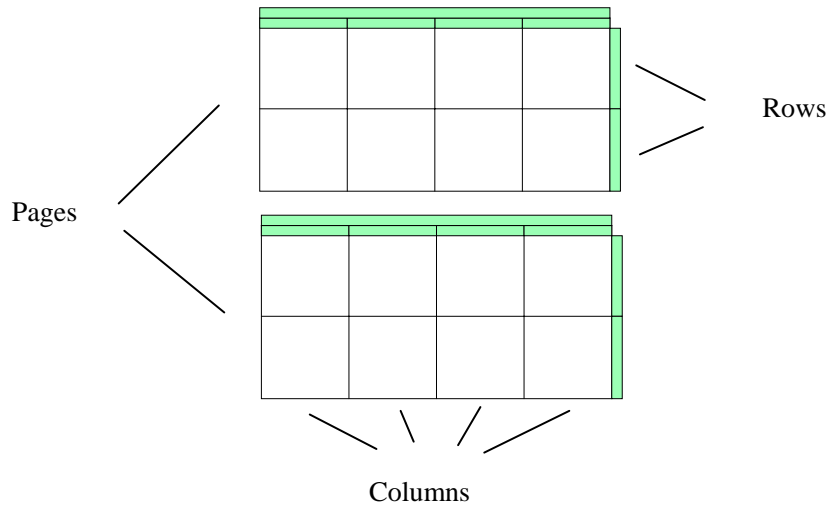
▶ **To change the order of conditions:**

1. Select a condition from the list.
2. Click **Move up** or **Move down**.

## Layout

The user can control the layout of the plots in terms of rows, columns and pages. There are two ways of doing this:

- Automatically, by binding a dimension to each conditioning variable
- Manually, by specifying the exact dimensions of the Trellis plot.



Example of three conditioning variables bound to rows, columns and pages.

### ▶ To arrange the plots automatically:

1. In the Scatter Plot **Properties** dialog, under the **Trellis** tab, make sure the box labeled **Bind dimensions to conditioning variables** is checked.
2. Click **Add...**
3. In the **Trellis condition** dialog, select a variable and a dimension to bind it to.
4. Repeat steps 2 and 3 for any additional conditions.



Experiment with different layouts while still in the Trellis condition dialog. The visualization is updated dynamically, so you can immediately see the results of your current settings.

---

▶ **To set layout manually:**

1. In the Scatter Plot **Properties** dialog, under the **Trellis** tab, make sure the box labeled **Bind dimensions to conditioning variables** is not checked.
2. Enter the number of columns and rows in the desired layout.



The number of pages is set automatically to make sure there is room for enough plots.

---

### Empty panels

It is possible to insert empty panels anywhere in the layout. This is useful when manually configuring a multi-page layout. Normally in such a layout, a page flows to the next as soon as it is full. This may not always be appropriate.

Consider, for example, a situation with twenty-four panels arranged in a three-by-three layout. This would yield two pages with nine panels on each, and a final page with only six panes. By inserting an empty panel on each page, the result would be three pages with eight panels on each.

▶ **To insert an empty panel:**

1. Right-click in the heading field of a panel.
2. Select **Insert Empty Panel**.

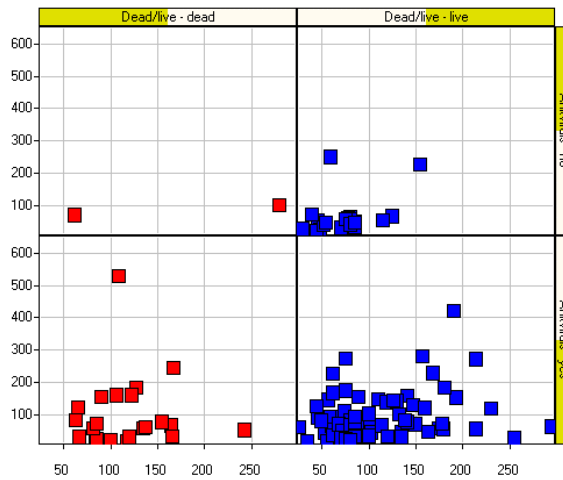
If an empty panel is inserted on the first page, it will appear at the same position on each subsequent page. An empty panel inserted on the second page will appear on each even page, etc.

## Heading colors

The colors used in the headings of columns, rows and pages can be set manually from the **Trellis** tab of the **Properties** dialog.

- **Text** is the color of the text in all headings.
- **Background** is the background color of all headings.
- **Indicator** is the color of the bar that indicates the relative size and position of the range of values.

Note that the indicator bar represents the size of the range of *values* represented by a column, row or page, not the number of *records* with that value. The Trellis plot below illustrates this. The top-left plot only has two values, but the indicator bar shows 50% because the column represents one out of two possible values.



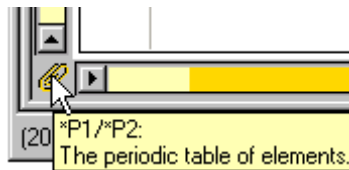
Right-clicking in a heading field will bring up a pop-up menu that provides quick access to features available for that panel.



## Annotations

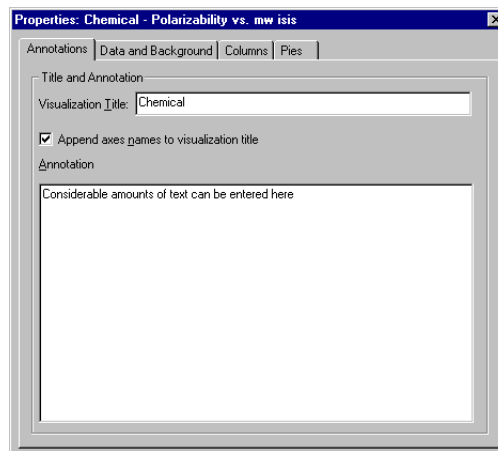
### Annotating visualizations

You can give any visualization a title and an annotation. The title will appear as the caption of the window. It can also appear in the heading of printouts. The annotation will appear as a tool tip when you place the mouse pointer over the paper clip at the bottom-left corner of the visualization.



▶ **To set title and annotation:**

1. Go to the **Annotations** tab of the **Properties** dialog.
2. Enter a title and/or an annotation.
3. Check **Append axes name to visualization title** if you want the current axes to be appended to the title.



The **Annotations** tab of the **Properties** dialog

You can type in a lot of text in the **Annotation** field, as well as cut and paste to and from other Windows applications—however, you can only do this with text. This is a good place for adding information about the purpose and methods used in the visualization, features to look for and other helpful information concerning the data. The annotation can be printed together with the visualization, by selecting the **Include annotation** option available in the **Print Layout Options** on the **File** menu.

## Annotating columns

It is also possible to annotate individual columns. Column annotations will appear when you place the mouse pointer over the paper clip next to a query device.

- ▶ **To enter an annotation for a column.**
  1. Go to the **Columns** tab of the **Properties** dialog.
  2. Select a column.
  3. Enter your text in the **Annotation** field.
  4. Select **Always** or **Only when annotation exists** from the radio buttons under **Show annotation paper clip**, if you want to be able to access the annotations through the paper clips.

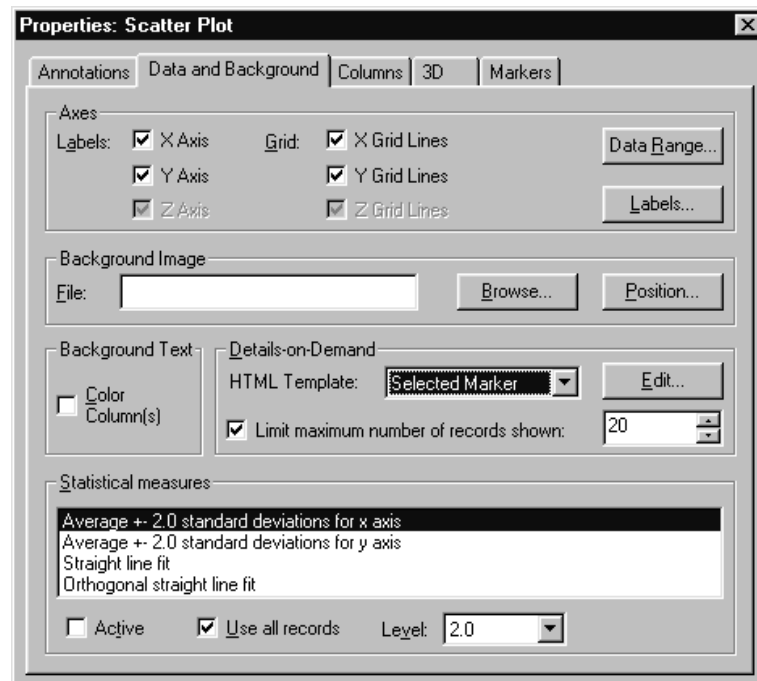


You can also edit annotations for a column by double-clicking the paper-clip next to the corresponding query device, or by selecting **Edit Annotations...** from the query device pop-up menu.

---

## Data and Background

The **Data and Background** tab is divided into five sections that are devoted to adjusting and setting various aspects of the visualization environment and display of information.

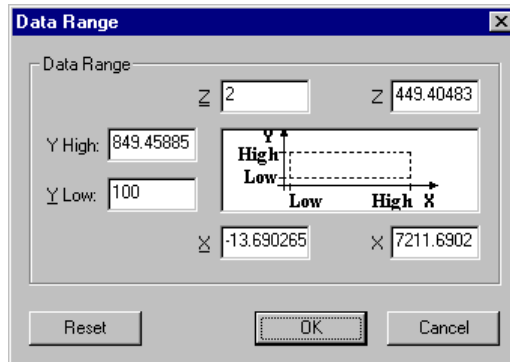


Data and Background tab of the Properties dialog

### Axes

Here you define several label, axis and grid attributes. The check boxes allow you to control which labels and grid lines to display. Under the **Data Range** dialog box, the minimum and maximum values of the axes can be set. This allows you to get more space around the data whose range otherwise is set to the range of the records of the data set.

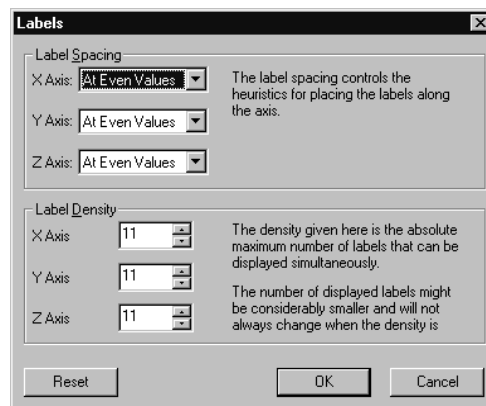
## 7 • MANAGING VISUALIZATIONS



Data Range dialog

If you click the **Labels** button you can set label placement and label density for each of the axes. For placement you can select to show axis labels in four ways:

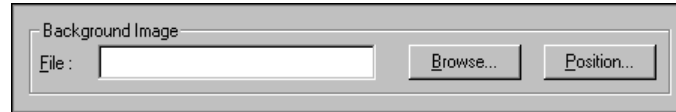
- **Even Distribution** – will attempt to distribute labels between highest and lowest values
- **At Values** – will not place a label where there is no value
- **At All Values** – will place labels at all values
- **At Even Values** – will attempt to place labels at evenly spaced whole numbers.



The **Labels** dialog


The **Label Density** setting will override **Label Spacing** settings so that even if **At All Values** is selected for spacing; the density value will limit the number of lines and labels in the visualization.

## Background Image




**Background Image** section of **Data and Background** tab

A background image can be loaded to show under the visualization—see the chapter on Background Images and Maps. **Browse...** to the location of a bitmap to place behind your visualization. Put the image accurately in place with the controls found under the **Position...** button. The image is stretched or shrunk between the corner data coordinates. You cannot crop images in this dialog.

 15: Background Images and Maps

## Details-on-Demand

The information in each record can be displayed in an external or internal HTML browser. This section of the **Data and Background** tab gives access to the templates used for formatting the HTML display of record details.

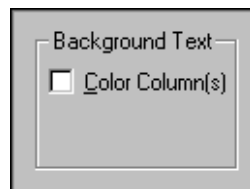
 11: The Details-on-Demand window



Details-on-Demand templates are accessed through this portion of the **Data and Background** tab

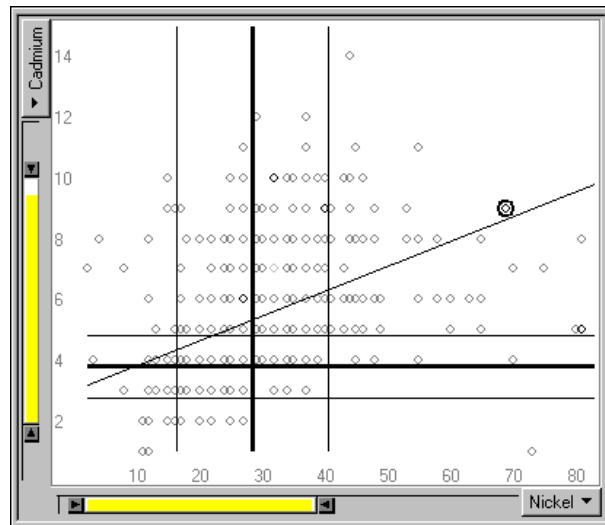
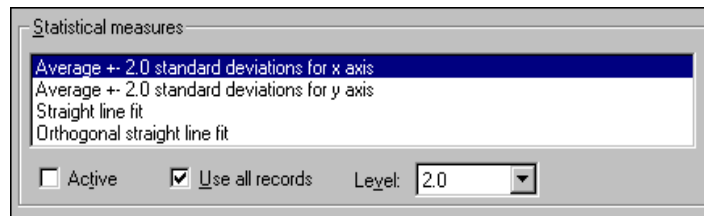
## Background Text

The Color Columns check box shows or hides the text in the visualization window describing which column is used for coloring.



## Statistical measures

Some statistics can be visualized directly in 2D scatter plots. This section of the dialog offers control over the X-axis and Y-axis standard deviation marking lines, and the two types of straight line fit: *least squares fit* and *orthogonal*. The latter is more accurate in cases where records approach a vertical line.



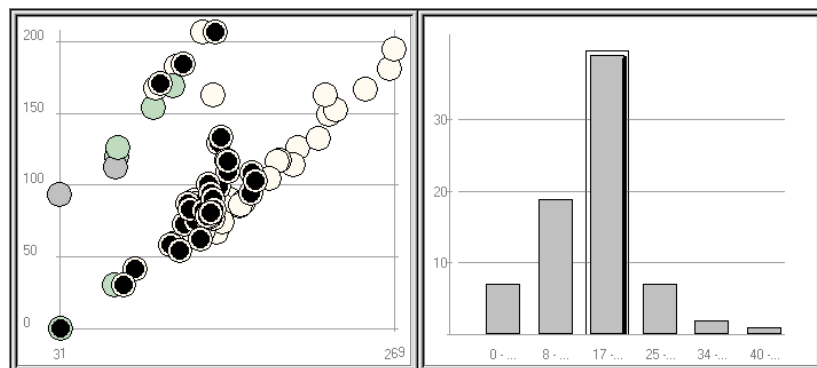
A scatter plot with standard deviation statistics and straight line fit added

The **Active** check box hides and shows the selected statistics. The **Use all records** checkbox toggles between using the only the visible records, or using all the data in the data set at all times, regardless of query device setting and zoom. The Level of standard deviation interval can be set from the **Level** combo-box, or by using the arrow keys. Thin black lines delineate the chosen interval in the visualization.

## Handling Multiple Visualizations



*Discovery often involves working with different views of the same information. Bar charts and histograms are powerful tools for analyzing aggregate data while scatter plots can reveal trends and correlation. This chapter describes how to rearrange the layout of windows when working with multiple scatter plots, histograms, bar charts, etc.*



A histogram and a scatter plot displaying the same data set—the dark scatter plot markers are the records in the tallest bar

Spotfire.net is able to show multiple visualizations, each one as a window presenting the same data, but in different ways. The visualizations may have dissimilar coloring or axes—or even be of different types—one a 3D scatter plot, another a bar chart. Each visualization can fill the entire window, all can be seen simultaneously, or each can reside on their own tab of a workbook.

When you operate the query devices, all visualizations are simultaneously updated, showing alterations in all visualizations when a factor is changed. When a marker is highlighted, it is highlighted in all visualizations simultaneously.

## Creating new visualizations

New visualizations are created by selecting the appropriate item a command from the **Window** menu, or by using the keyboard shortcuts shown in the same menu.

## Arranging visualizations

There are several ways to reposition windows; the commands governing these functions all reside in the **Window** menu:

- **Auto Hide Axes Selector** – hides the zoom bars and the selectors for column association to the axes
- **Hide Window Frame** – hides the title bar giving more space to the visualizations—only available when several visualizations have been tiled
- **Auto Tile** – arranges all the windows on screen according to an internal algorithm. The active visualization will be given leftmost, uppermost, and size priority
- **Cascade** – arranges the visualization windows so that they partially overlap each other, leaving each window accessible by clicking on the title bar
- **Tile Horizontal** – splits the window area horizontally according to the number of visualizations, giving each visualization equal area
- **Tile Vertical** – splits the window area vertically according to the number of visualizations, giving each visualization equal area



An alternative way of organizing visualizations is to use a Trellis plot. This allows you to split an existing plot into panes, and organize these in a grid. Note that this feature is currently available only for scatter plots (2-D and 3-D) and profile charts. See page 86 for more information.

---



## Arranging Windows

The Spotfire.net Desktop consists of several windows. It can be desirable to rearrange the application's layout to organize visualizations and information for more effective analysis.

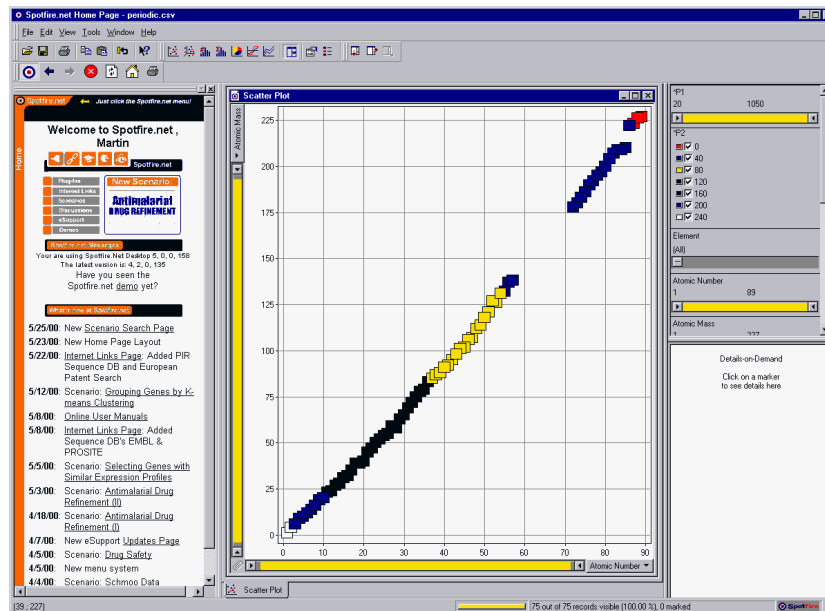
### Showing and hiding windows

The **View** menu displays a list of the Spotfire.net windows. Each window can be shown or hidden by ticking or un-ticking the name of the window, respectively. The windows are listed and described below:

- **Toolbars** – the row of clickable icons right below the menus. Appendix C has a complete list and description of functions available in the toolbar.
- **Status Bar** – the information field at the bottom of the application window. The status bar displays information about pointer position, number and portion of records that are visible and marked, etc.
- **Workbook** – when this option is checked, each visualization is given a separate tab
- **Details-on-Demand (Text)** – the window where information about the brushed objects, the highlighted scatter plot marker, bar or pie object appears
- **Details-on-Demand (HTML)** – the window where information about the highlighted or marked scatter plot marker(s), bar or pie object(s) appears, formatted according to the HTML templates stored in the current SFS file
- **Details-on-Demand (External Browser)** – a separate web browser showing the same information as described above
- **Query Devices** – the window that contains all the query devices
- **Legend** – the window that explains the significance of the current visualization settings, e.g. which columns are used for size, rotation, color, etc.
- **Portal Window** – the browser window used to access Spotfire.net web content, and to run Spotfire.net applications (plug-ins).

## Rearranging windows

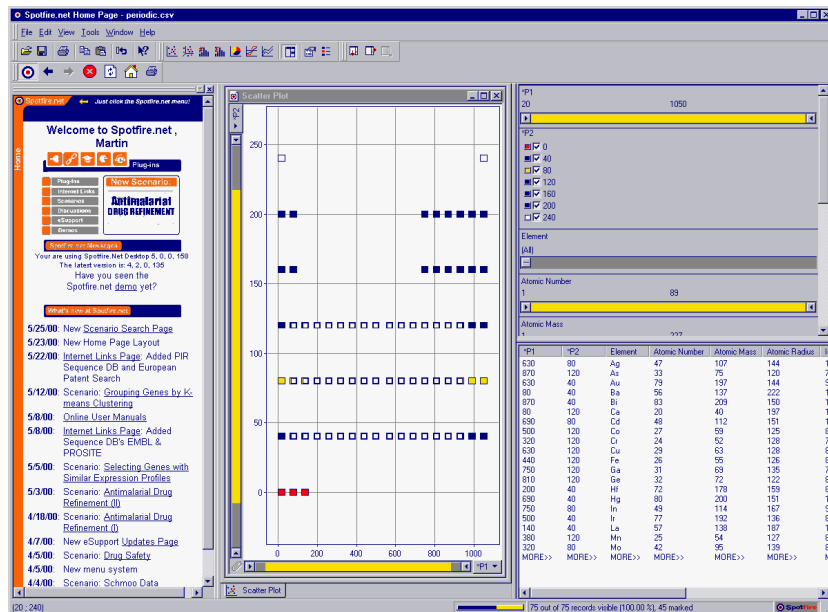
Spotfire.net Desktop has a certain default setup. The visual analysis area contains the visualizations; the Portal window appears on the left, while the Query device window and the Details-on-Demand window are on the right side.



The default setup of the Spotfire.net Desktop.

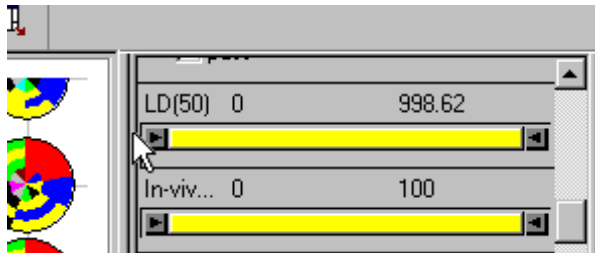
Since this may not be the most effective layout for all purposes, it is possible to resize the different parts by clicking on a separating border and moving it until the adjacent windows have the desired size.

In the figure below, the vertical separating border has been moved to the left to make the query device window wider—allowing finer control over the sliders. The horizontal border, below the query devices, has been moved downwards to display more query devices.



Spotfire.net with enlarged query device window

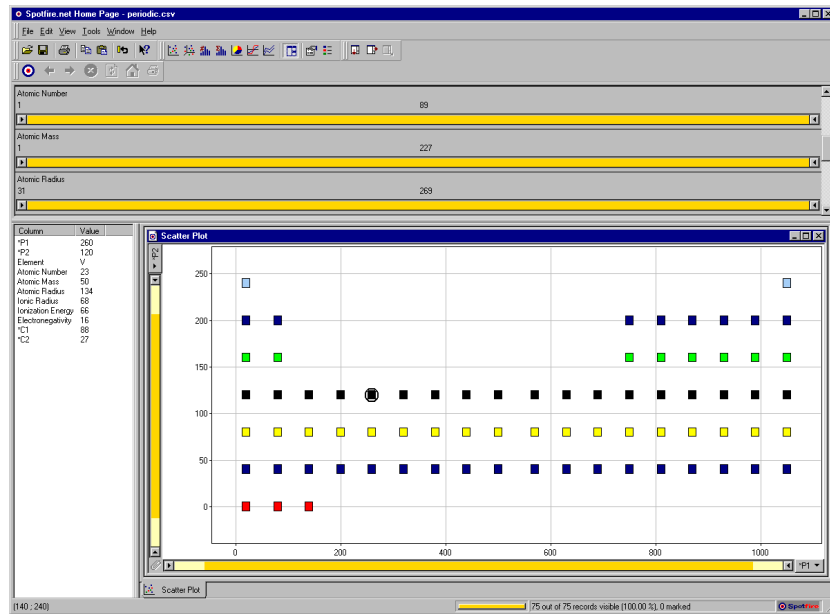
It is also possible to move an entire window: Press the left mouse key while pointing at the left or right edge—there is a thin, 3 pixel band—grab the window for repositioning.



Grab side edge of window to reposition it

While keeping the left mouse key down, move the pointer to where the window is to be placed. While moving the window, an outline will reshape to show how the window would end up, if put in that particular position. The window can later be resized with the methods described above.

## 8 • HANDLING MULTIPLE VISUALIZATIONS



Relocated query device and Details-on Demand windows. Note that the Portal window has been closed.

## Workbook Mode

Workbook mode puts each visualization on its own tab. A click on a tab activates the corresponding visualization and allows manipulation and modification of its attributes.

- ▶ To put Spotfire.net into Workbook mode, select **Workbook** from the **View** menu.



Workbook tabs

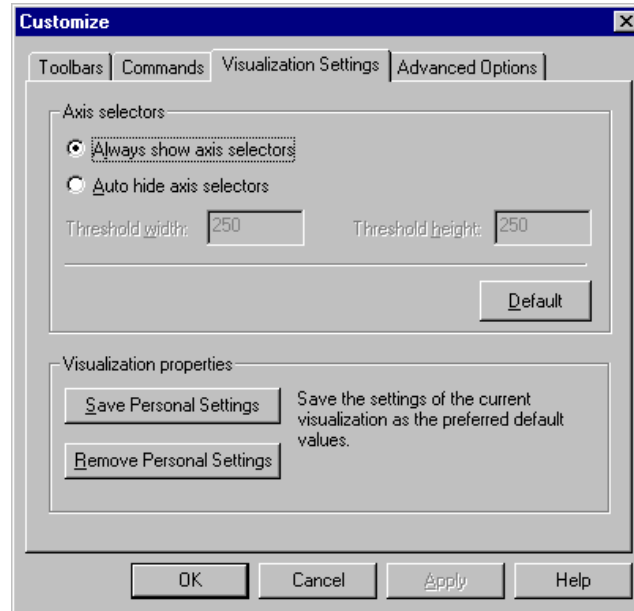
## Combining Workbook with Auto tile

With **Auto Tile** off, pages in the Workbook retain the size that they had been given; decreasing the size of the application window will crop portions of some visualizations.

With the **Auto Tile** function on, clicking on a Workbook tab gives the chosen visualization priority position and size.

## Customizing Visualizations

The **Visualization Settings** tab of the **Customize** dialog allows you to control certain aspects of the visualizations.



The **Visualization Settings** tab is available only if data is loaded

### Hiding axis selectors

Axis selectors can be made to automatically disappear when the visualization size is less than a certain threshold value. Click **Auto Hide Axes Selectors** and set threshold values to enable this function. The **Default** button restores default threshold values and disables Auto-hide.

### Controlling default settings for new visualizations

The **Save personal settings** button on the **Visualization Settings** tab of the **Customize** dialog allows you to set the preferred initial settings for new visualizations.

The settings of the current visualization are used as a template for initial visualizations of similar type. In this context, 2D and 3D scatter plots are linked, and Histograms are linked to bar charts, i.e. default settings for 3D visualizations will also apply to 2D scatter plots, etc

The following settings are saved:

- **Color** – e.g. beginning and ending colors for continuous coloring, brushed and deselected marker color, etc.—but not which column governs coloring
- **Size** – but not which column governs size
- **Jittering** – random displacement of markers
- **Shape** – which marker shape or glyph is used
- **Labels** – the number of labels displayed and their color

No column-specific settings are saved, e.g. **Rotate by** is not saved since it refers to a specific column. Settings from the **3D** tab such as navigation, rendering and perspective are not saved; neither are **Data and Background** tab settings as they are column specific.

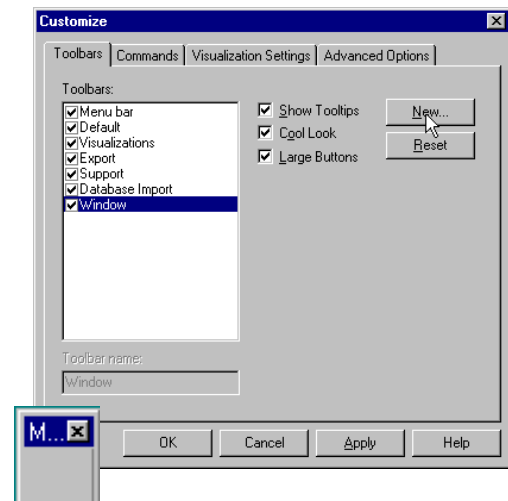
The **Remove personal settings** button will erase the configuration that you have previously saved and restore the default values.

## Customizing Toolbars

You can tailor Spotfire.net's toolbars to allow you to rapidly access the features you use regularly.

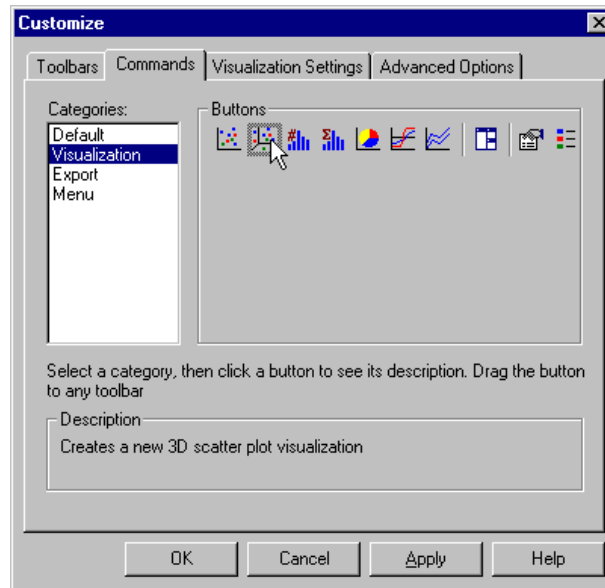
### ► To customize toolbars:

1. Select **Tools > Customize...**
2. Go to the **Toolbars** tab.
3. Click **New...**
4. Enter a name for the new toolbar and click **OK**.
5. Make sure the new toolbar is checked in the list. Your toolbar should be visible as a small box on your screen.



## 8 • HANDLING MULTIPLE VISUALIZATIONS

1. Go to the **Commands** tab.
2. Select a category of commands.



3. Drag any button to your new toolbar (or to any of the standard toolbars).
4. Click **OK**.



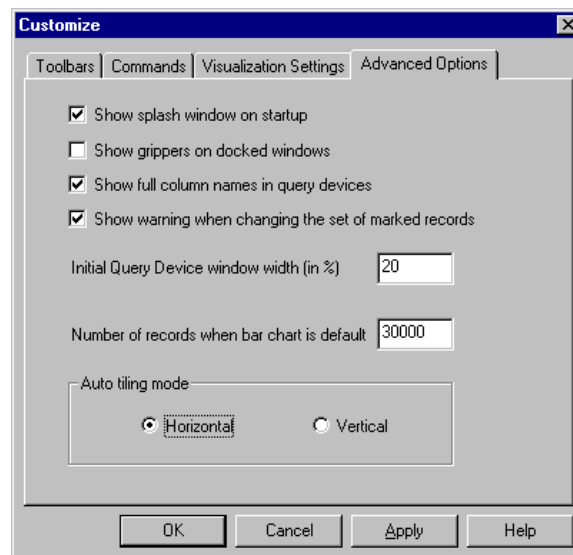
## Advanced Options

The **Advanced Options** tab of the **Customize** dialog provides some options for fine-tuning your Spotfire.net environment. Use the check boxes to enable or disable any of the following features (the first three only take effect after Spotfire.net is restarted):

- Splash window - the image shown while the Spotfire.net Desktop is being launched.
- Grippers – gray bars that can be used when dragging a window to a new position.
- Full column names in query devices
- Warning when changing marked records – the dialog that appears when you mark a new set of records

You can also set:

- The relative width of the query device window
- The number of records at which bar chart becomes default visualization type
- Vertical or horizontal auto-tiling





# Marked, Active and Highlighted Records



*This chapter describes how to choose a subset of the records in a data set and how to copy, delete and in other ways manipulate these records.*

## Methods for Choosing a Sub Set

There are four ways of choosing a sub set of records:

- Marking
- Activating
- Highlighting
- Selecting (see chapter 6)

### Marking records

The purpose of marking records may be to view details, or to distinguish records in order to copy, erase, or in other ways manipulate them.

▶ **To mark one or more records:**

1. Click and hold the left mouse button, and drag the mouse: the pointer will delineate a rectangle.
2. Enclose the markers of interest in this rectangle. When the mouse button is released, the marked objects will be given a different color and outline.



Marking a pie or bar is equivalent to marking all the records that are included in that pie or bar.

---

▶ **To add more records to the marked set:**

1. Press and hold **Ctrl**.
2. Click on markers you want to add, one by one. Alternatively, click and drag to mark several markers using the rectangle (while still pressing **Ctrl**).

### Marking with the lasso

A set of graphical objects, be they pies, bars or scatter plot markers, can be "lasso selected", i.e. be surrounded by a line drawn in an arbitrary shape. Press the **Shift** key and the left mouse button, then draw a line capturing the markers. Note that a straight line connects the beginning and the endpoints of the lasso marquee. The line delineates the area that will be included in the selection.



For a record to be marked, its exact coordinates must be within the bounds of the selecting line. Just slicing through part of an enlarged marker will not suffice.

---

### Setting colors for marked objects

Marked records will be given an identifying color in all visualizations. You can define this color through the **Customize...** color button of the **Properties** dialog.

### Unmarking

Re-marking any marked object(s) while **Ctrl** is pressed will unmark them. If **Ctrl** is not pressed, only the new selection will be marked. Unmark all data by selecting **Edit > Marked Record(s) > Reset**; or by pressing [**Ctrl+E**]. Alternatively you can unmark everything by marking an empty area of any visualization.

### Message box when marking

If a set of markers is already marked, and you mark a new set, a message box will appear, warning you that you will lose the previous marking. You can switch this feature off by checking the box labeled **Never show alert again**. If, at a later time, you would like this feature back, select **Edit > Marked Record(s) > Reset**, and check the box in the dialog. You can also use the **Customize** dialog.



11: The Details-on-Demand window

## Activating

Clicking on a scatter plot marker will make a persistent circle surround the corresponding marker in all scatter plot visualizations. The details of an active marker will be displayed in the Details-on-Demand window.

If a pie or bar is activated, the Details-on-Demand window will display summary information about all the records included in that object.

## Highlighting

Moving the mouse pointer close to or on top of a marker will *highlight* it. A thin circle will cling to the marker until the pointer is moved a certain distance away. A scatter plot marker that is highlighted in one visualization is highlighted in all scatter plot visualizations.



Highlighted marker

Highlighting shows where in each visualization the data point lies, because the same record is highlighted in each visualization. The arrow keys will move the highlighting to the marker in the visualization that is closest in the direction of the arrow. The Tab key will move the highlighting to the next record in the data set.




---

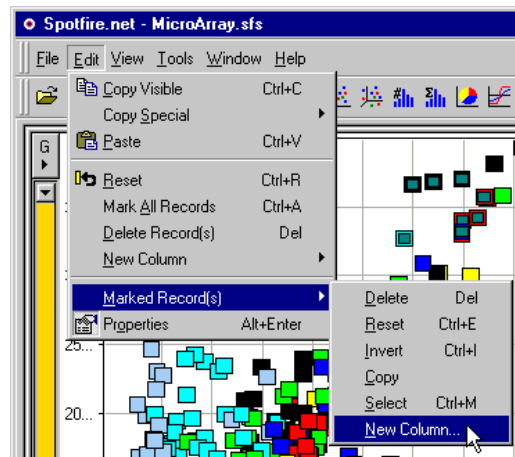
By **selected** we mean records that are visible, in other words not actively deselected by manipulating the query devices. See also chapter 6: *Working with Query Devices*.

---

## Actions Involving the Marked Set

The following commands concern the set of marked records. They can all be found in the **Edit** menu.

- **Marked Record(s) > Delete** deletes marked records from memory. [Del]
- **Marked Record(s) > Reset** unmarks all records [Ctrl+E]
- **Marked Record(s) > Invert** inverts the marking so that all unmarked records become marked and vice versa [Ctrl+I]
- **Marked Record(s) > Copy** copies marked records to the clipboard.
- **Marked Record(s) > Select** hides unmarked records [Ctrl+M]. To show them again, select **Edit > Reset**.
- **Marked Record(s) > New Column...** creates a new column containing two values, one value for the marked records and another for the unmarked records. You will be prompted for names for the column and for the values.



Deleting is irreversible and immediately affects the data set. The original data file, however, will remain unchanged until you explicitly overwrite it by saving. Data sets imported from a database, using ODBC, OLE DB, etc., cannot be overwritten. Save data in a file with another name if there is any chance you will want to revert to the entire original data set.

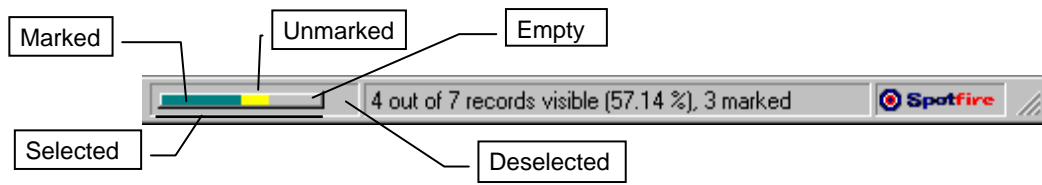
## Copying records

These commands under the **Edit** menu copy the data stored in the loaded data set, or a sub-set of it, to the clipboard. The copied information can then be exported, pasted into a spreadsheet program, such as Microsoft Excel or another program, e.g. Notepad.

- **Copy Visible [Ctrl+C]** copies all the records represented by graphical objects in the active visualization. This means that if some records cannot be represented in the visualization because they lack values in the necessary data fields (those fields are empty or null), they will be omitted. The same applies to records hidden using the **Edit > Marked Records > Select** command. The data is copied to the Windows clipboard in text format.
- **Copy Special > Selected Records** copies only the records within the limits set by the query devices to the Windows clipboard in text format
- **Copy Special > All Records** copies all the records in the active data set to the Windows clipboard in text format including the ones that happen to be invisible due to the settings of the query devices or zoom bars
- **Copy Special > Marked Records** copies the records corresponding to the markers, bars and pies brushed with the lasso or rectangle marking tool to the Windows clipboard in text format
- **Copy Special > Current Visualization** copies the active visualization display as a Windows Metafile image. This image can be pasted into other Windows compatible applications such as MS Word, Paint, PowerPoint etc. As Metafile images are vector-based graphics they can be enlarged or shrunk at will without losing detail. Axes, labels and grid lines are included in the image. These images can be manipulated from within the new host applications.

## Information about the Data Set

The record meter in the status bar at the bottom of the Spotfire.net Desktop gives continuous information about visualization statistics. You can see at a glance what portion of all the data is currently visible, marked, etc. The same information is displayed as text.



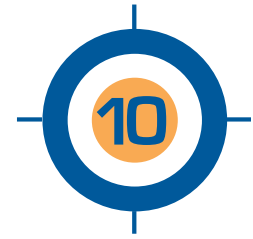
Colors should be read as follows:

- Green – Records that are selected and marked
- Yellow – Records that are selected and unmarked
- Grey – Records that are selected, but lack a value required for representation on the current axes, and hence are invisible.

The sum of the above makes up the set of selected records. The length of the record meter in relation to the box in which it sits is an indication of the relative size of the current selection.



# The Portal Window



*This chapter describes the Portal Window, and some of the wide range of information and applications it provides. These services are constantly being developed, so this chapter focuses on giving an introduction to the main features, and how to navigate between them.*

## What is the Portal Window?

The Portal Window is a simple web browser, fully integrated into the Spotfire.net environment. The most straightforward use of the window is displaying HTML-coded web pages. Like any modern browser, the window can also run script code (Java, for instance). This means that the Portal Window is an ideal platform for running customized applications for virtually any task that involves interaction with the data set being analyzed. These applications are referred to as Spotfire.net applications.

The Portal Window is open by default when you launch Spotfire.net. You can close and open it by clicking the button with the Spotfire logotype in the Navigation toolbar.



## The Navigation toolbar

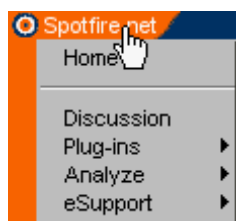
The Navigation toolbar is used to navigate the content of the Portal window. It has the same basic features as most web browsers, i.e. Back, Forward, Stop, Refresh, Home and Print.

The toolbar can be shown or hidden by selecting **View > Navigation toolbar**.

## Spotfire.net Content

As mentioned, the Portal Window can display any web page. In this section, however, we will focus on a small selection of the content provided or hosted by Spotfire.

From any page in Spotfire.net, click on the "Spotfire.net" text in the top right-hand corner. This brings up a menu from which you can reach any of the services described in the following sections.



### Discussion forum

In the forum, you can follow discussions relating to Spotfire products and the various kinds of research carried out using these tools. Discussion threads typically include end-users as well as Spotfire staff.

- ▶ To go to the discussion forum, select **Spotfire.net > Discussion**.

### Plug-ins

Plug-ins are small applications that are hooked into the Spotfire.net framework to perform specific tasks. Spotfire offers a wide selection of plug-ins for external software integration to data manipulation and user interface enhancements. The plug-ins can be downloaded for free.

- ▶ To go to Plug-ins, select **Spotfire.net > Plug-ins > Spotfire**.

### Scenarios

Scenarios are examples of successful use of Spotfire.net within various types of research, such as Pharma/Biotech, Manufacturing, Semiconductor and Chemicals. Scenarios provide ideas and inspiration for effective use of Spotfire's analytical tools. They also give you direct access to the underlying data of each example, allowing you to experiment on your own.

- ▶ To go to Scenarios, select **Spotfire.net > Analyze > Scenarios**.

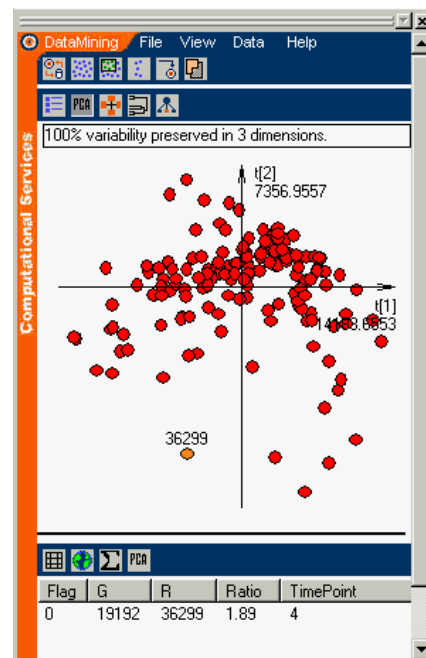
## Support

Spotfire eSupport provides a range of services, including the possibility to upgrade your software to the latest version. You can also search Spotfire's support database for tips and solutions to a range of problems, or read the list of Frequently Asked Questions. If this does not solve your problem, you can get in touch with Spotfire's support staff through these pages, and receive a reply within 24 hours.

- ▶ To go to eSupport, select **Spotfire.net > eSupport**.

## Spotfire.net Applications

Spotfire.net applications form the new generation of specialized software, which runs in the Portal Window. Once a license has been issued, the purchased software is available for immediate use. Because Spotfire.net applications are written as web pages, they require no installation. They communicate directly with the visualization engine of the Spotfire.net Desktop, thus taking advantage of the full power of Spotfire.net.



Spotfire.net applications typically enhance the capacity of the Spotfire.net desktop with features such as advanced statistical measures, Principal Component Analysis, or features specially designed for particular research fields.

Spotfire.net applications are available from Spotfire, Inc. You can also develop your own applications using Spotfire Software Laboratory.

Principal Component Analysis using the Computational Services Spotfire.net application.



# The Details-on-Demand Window



*This chapter describes the different ways in which Spotfire.net can present details about a specific visualization object, such as a marker, bar, or pie, depending on what type of visualization is used.*

## Details for Different Visualizations

Although many aspects of a data set can be shown graphically, there is often a need to inspect detailed information about a single visualization object or a group of records. The information presented in the Details-on-Demand window is different for each visualization type.



9: Marked, Active and Highlighted Records

- **Scatter plots** (2D or 3D), **line charts**, and **profile charts** – All fields of the *active* record, or all fields of all *marked* records, presented in table form
- **Bar charts** and **histograms** – the height, position and number of records in the *active* bar, or all fields of records included in the *marked* bars
- **Pie charts** – the size of the *active* pie and its individual sectors, or all fields of records included in the *marked* pies



---

Remember: to **activate** a record, bar or pie, click on it. To **mark** records, bars or pies, click and drag a rectangle around the desired objects.

---

## Presenting the Details-on-Demand

There are three different ways in which the information can be presented:

- As plain text in the details frame
- As an HTML document based on Microsoft Internet Explorer displayed in the details frame (if IE is installed)
- As an HTML based document in an external browser, Netscape or Internet Explorer

The HTML based details information displayed for marked or highlighted graphical objects, i.e. the data stored for the corresponding record, can be formatted using the templates in the HTML tab of the **Properties** dialog. Each type of visualization (scatter, bar, pie) has its own HTML template. The controls for these templates are to be found on the **Properties** dialog **Data and Backgrounds** tab. The HTML based presentation can use all functions provided by modern browsers, e.g. images, tables, links to the Internet and even Java applets. Further manipulation of the HTML template is described below.

### Plain text



8: Handling Multiple Visualizations: Arranging windows

| Column            | Value |
|-------------------|-------|
| *P1               | 810   |
| *P2               | 200   |
| Element           | C     |
| Atomic Number     | 6     |
| Atomic Mass       | 12    |
| Atomic Radius     | 77    |
| Ionic Radius      | 30    |
| Ionization Energy | 113   |
| Electronegativity | 25    |
| *C1               | 82    |
| *C2               | 1     |

By default, Spotfire.net displays data in plain text. The details window is placed in the lower right corner of the main window. It can of course be moved to another location if desirable. The details for a single highlighted marker are presented in a vertical table, those for several marked markers in a horizontal table.

Details-on-Demand presentation for a highlighted object (above) and for several marked records (below).

| *P1 | *P2 | Element | Atomic Number | Atomic Mass | Atomic Radius | Ionic Radius | Ionization Energy | Electronegativity | *C1 | *C2 |
|-----|-----|---------|---------------|-------------|---------------|--------------|-------------------|-------------------|-----|-----|
| 500 | 120 | Co      | 27            | 59          | 125           | 83           | 79                | 18                | 120 | 30  |
| 320 | 120 | Cr      | 24            | 52          | 128           | 75           | 68                | 17                | 91  | 28  |
| 440 | 120 | Fe      | 26            | 55          | 126           | 83           | 79                | 18                | 115 | 32  |
| 380 | 120 | Mn      | 25            | 54          | 127           | 89           | 74                | 15                | 113 | 31  |
| 320 | 80  | Mo      | 42            | 95          | 139           | 83           | 70                | 22                | 115 | 38  |
| 500 | 80  | Rh      | 45            | 102         | 134           | 80           | 75                | 22                | 128 | 33  |
| 440 | 80  | Ru      | 44            | 101         | 134           | 82           | 74                | 22                | 114 | 39  |
| 380 | 80  | Tc      | 43            | 98          | 136           | 78           | 73                | 18                | 113 | 37  |

The width of the columns in the Details-on-Demand text display can be adjusted to suit the purpose. By default the width is as wide as the longest string in the column be it the header or one of the values in the records. To change the width, place the mouse pointer on the separating line in the column header, it will change shape as shown in the illustration below. Click-and-drag the separator to the desired position. If you double-click on the separator, the column width will collapse or expand to fit the longest value in the table.

| *P1 | *P2 | Elem... | Atomic Number |
|-----|-----|---------|---------------|
| 500 | 120 | Co      | 27            |
| 320 | 120 | Cr      | 24            |
| 440 | 120 | Fe      | 26            |
| 380 | 120 | Mn      | 25            |
| 320 | 80  | Mo      | 42            |

Column width adjustment in the Details-on-Demand window



Notice that columns can be completely hidden from view and that if you do not know where they are they can be difficult to find! The order of the columns in the Details-on-Demand is always the same as in the database itself, checking the **Columns** tab of the **Properties** dialog will reveal the location of the column with respect to its neighbors.

---

## HTML

If Microsoft Internet Explorer, version 3 or higher (or more precisely the HTML browser ActiveX control) is installed on the computer, Spotfire.net can display the data for scatter plot markers as an HTML formatted document in the details frame. This is initialized by checking **View > Details-on-Demand (HTML)**.

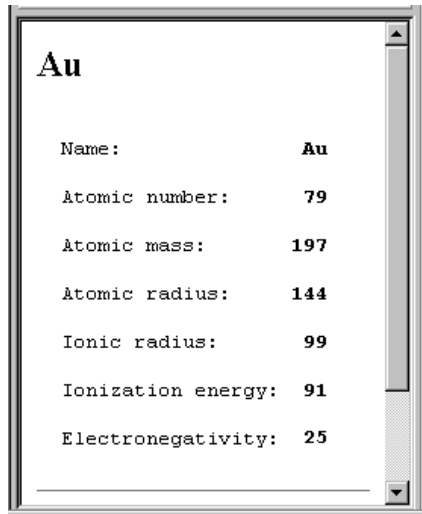
An external browser, Netscape or Internet Explorer, can also be launched to view the details of the data by selecting **View > Details-on-Demand (External)**.



With some versions of IE older than 4.0 each selection of a record will open a new browser window, this could eventually result in a large number of windows cluttering your workspace.

---

Spotfire.net uses the HTML template specified for the data set to layout the information. If none has been defined, the application will provide a default template for the type of visualization that is active.



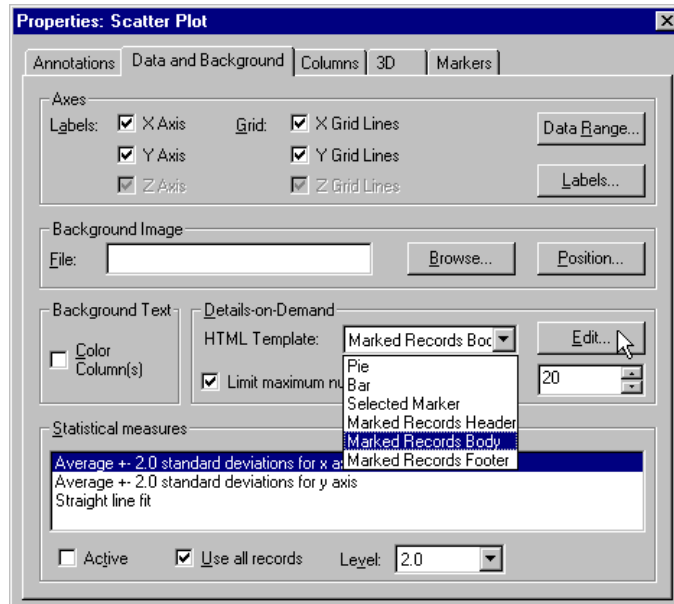
HTML formatted details for a single active element

The HTML templates used for formatting marker details are accessed and edited through the **Data and Background** tab of the **Properties** dialog box. There are six editable templates:

- Pie—used when a pie is activated (clicked)
- Bar— used when a bar is activated (clicked)
- Active Marker—used when individual records in scatter plots or single bars or pies have been clicked
- Marked Records Header
- Marked Records Body
- Marked Records Footer

When some records have been marked, the details of those records are presented according to the three "Marked records" templates.

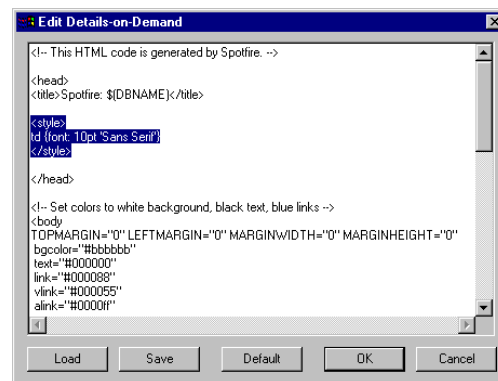




The **Data and Background** tab pull down menu for HTML template access

## Editing HTML templates

The **Edit...** button brings up a dialog box in which you can edit the selected template. On the HTML editing window there you have several possibilities for creating and modifying the template.



The **HTML Edit** window for the **Marked records Header** with a highlighted method of changing the typeface of all text in the tables.

The **Load** button allows you to browse for an external file and use it as the template. The **Save As...** button saves the changes in an external HTML file that can be used to format information in other Spotfire.net

sessions. The **Default** button will revert to the default settings and remove any changes made to the templates. **OK** applies the editing changes and displays the Details-on Demand if that option has been selected in the **View** menu. **Cancel** removes any changes you have made since opening the **Edit** window.

When a new data set is loaded into Spotfire.net, the program creates default HTML templates that can be edited to suit your needs. To apply the templates made for another data set to the presently loaded data set, choose **File > Apply Template** and select the SFS file that contains the desired settings. Spotfire.net will make a guess as to the allocation of attributes so it is preferable for the two data sets to have structures that are as similar as possible.

### Marked Records Header

The header template will be applied to the pages presented by the browser. To conserve space however the header will not be seen in the internal browser frame. In the editing window shown on the previous page a style block has been added to the Marked Records Header template, transforming all table text into a straight—sans serif—typeface.

### Variables

In the HTML templates, it is possible to use variables that refer to the value of a specific data column, or to other information. These variables will then be substituted with the values of any marked or active record.

| HTML variable | Refers to   |
|---------------|---|
| \$Name        | Value of a particular record in column 'Name'   |
| \$DBDIR       | The full path to the location of the directory of the current database (file), excluding a trailing backslash |
| \$DBPATH      | The full path, including the file name, of the currently loaded database                                      |
| \$DBFILE      | The name of the currently loaded database, including the file extension (.skv, .csv, etc.)                    |
| \$DBNAME      | The name of the currently loaded database, excluding the file extension                                       |

Additional variables for displaying descriptive statistics about the data set are described below.

## Example: Changing an HTML template

There is a plethora of literature on HTML programming. This example is simply a short demonstration of changing the colors and typeface of the displayed details.

### ▶ To alter the appearance of the Details-on-Demand presentation

1. Load a data set, in this example we will use the periodic table data
2. From the **Properties** dialog box [Alt] +[Enter] select the **Data and Background** tab
3. From the drop-down list of the HTML Template section—near the right central portion of the tab—select **Marked Records Header** and click the **Edit** button
4. Add the following lines of code inside the header of the HTML page to present text in a straight typeface:

```
<style>
td {font: 10pt 'Helvetica'}
</style>
```



Notice that this should be placed above the body of the template to affect the entire presentation. Notice also the use of 'curly brackets'.

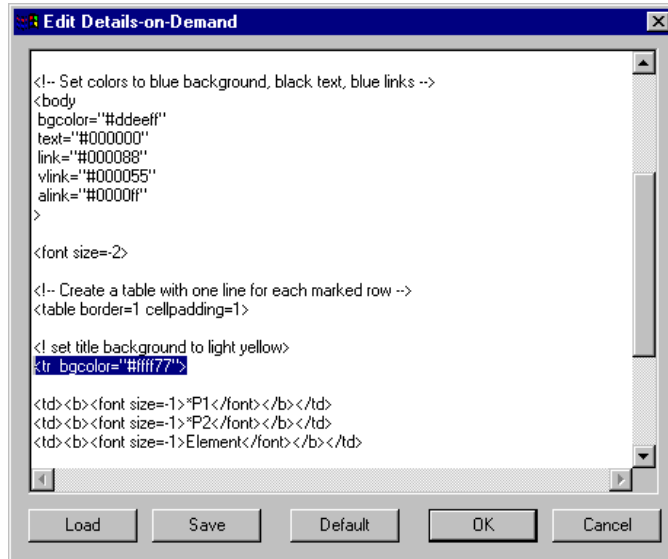
5. Further down in the body, change the background color of the body to light blue by changing the color definition from #ffffff—white—to light blue

```
bgcolor="#ddeeff"
```

6. Add the following line of code above the titles to make the title line appear with a light yellow background:

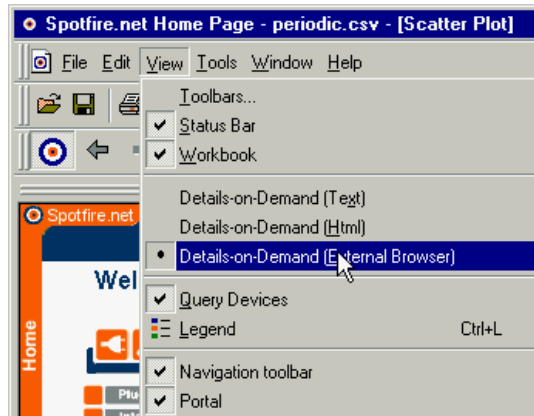
```
<tr bgcolor="#ffff77">
```

Table title color background defined as light yellow in the **Marked Records**



Header template

7. Click **OK**.
8. Select **View > Details-on-Demand (External Browser)**



Displaying **Details-on Demand** in an external browser

Each time you mark a number of markers in the scatter plot the details will be presented in a horizontal table in your web browser

| *P1 | *P2 | Element | Atomic Number | Atomic Mass | Atomic Radius | Ionic Radius | Ionization Energy | Electronegativity | *C1 | *C2 |
|-----|-----|---------|---------------|-------------|---------------|--------------|-------------------|-------------------|-----|-----|
| 760 | 160 | Al      | 13            | 27          | 143           | 67           | 60                | 16                | 28  | 25  |
| 990 | 160 | Cl      | 17            | 35          | 100           | 167          | 130               | 32                | 173 | 47  |
| 80  | 160 | Mg      | 12            | 24          | 160           | 86           | 75                | 12                | 78  | 30  |
| 20  | 160 | Na      | 11            | 23          | 186           | 116          | 50                | 9                 | 69  | 54  |
| 870 | 160 | P       | 15            | 31          | 110           | 58           | 105               | 22                | 102 | 13  |
| 930 | 160 | S       | 16            | 32          | 103           | 170          | 103               | 25                | 149 | 45  |
| 810 | 160 | Si      | 14            | 28          | 117           | 54           | 81                | 18                | 91  | 19  |

periodic.csv  
Database C:\Program Files\Spotfire\Pro4\Data\General\Periodic Table\periodic.csv

Details presented in MS Internet Explorer

## Descriptive statistics measurements

These variables may be added into the HTML template to display basic descriptive statistics for a data set. It is necessary to activate the statistical measurement in the **Data and Background** tab in the **Properties** dialog corresponding to the variables to be displayed.

The LINEFIT\_A and LINEFIT\_B variables return the intercept and slope of the least squares fit line respectively.

| Measurement | Variable       | Explanation                       | Activate           |
|-------------|----------------|-----------------------------------|--------------------|
| LINEFIT_A   | \$(LINEFIT_A)  | $y = A + B \times X$              | Straight Line Fit  |
| LINEFIT_B   | \$(LINEFIT_B)  | $y = A + B \times X$              | Straight Line Fit  |
| AVERAGE_X   | \$( AVERAGE_X) | Average for X-axis                | Average +... for X |
| AVERAGE_Y   | \$( AVERAGE_Y) | Average for Y-axis                | Average +... for Y |
| STDDEV_X    | \$(STDDEV_X)   | Standard deviation for X-axis     | Average +... for X |
| STDDEV_Y    | \$(STDDEV_Y)   | Standard deviation for Y-axis     | Average +... for Y |
| STDDEV2_X   | \$(STDDEV2_X)  | 2 × standard deviation for X-axis | Average +... for X |
| STDDEV2_Y   | \$(STDDEV2_Y)  | 2 × standard deviation for Y-axis | Average +... for Y |
| VARIANCE_X  | \$(VARIANCE_X) | Variance for X-axis               | Average +... for X |
| VARIANCE_Y  | \$(VARIANCE_Y) | Variance for Y-axis               | Average +... for Y |

## 11 • THE DETAILS-ON-DEMAND WINDOW

|          |              |                      |               |
|----------|--------------|----------------------|---------------|
| VISIBLE  | \$(VISIBLE)  | # of visible points  | Always active |
| MARKED   | \$(MARKED)   | # of marked points   | Always active |
| SELECTED | \$(SELECTED) | # of selected points | Always active |
| ALL      | \$(ALL)      | # total              | Always active |

Statistic measurements available in the HTML template and the checkbox to activate if the browser is to display the chosen variable

An example of how this can be used is found in the HTML template for the *store.sfs* demo file that is included with the Spotfire.net distribution.

## View Tip



*The View Tip function in Spotfire.net is an aid in finding interesting visualizations. It enables you to scan through and create visualizations, based on the data type and the number of unique values of each column. This chapter describes how to use View Tip.*

### What is View Tip?

View Tip is a tool for rapid assessment of potentially interesting visualizations. It suggests visualizations based on the correlation between data in the columns, the number of unique values, and the distribution of data. In some cases View Tip also suggests colors.

View Tip is a convenient way of quickly getting to know a data set. As you browse through the different miniature plots, you realize that View Tip is a natural first step when you start exploring unknown data.

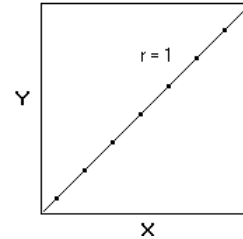
### Ranking Visualizations

View Tip calculates the correlation between different combinations of columns, and orders the suggested visualizations according to the degree of linear correlation.

The most common measure of correlation is the Pearson Product Moment Correlation (called Pearson's correlation for short). When computed in a sample, it is designated by the letter  $r$  and is sometimes called *Pearson's  $r$* . Pearson's correlation reflects the degree of linear relationship between two variables. *Pearson's  $r$*  ranges from +1 to -1.

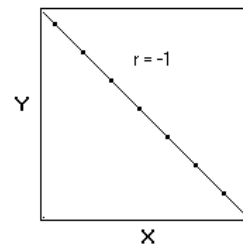
## Positive correlation

A correlation of +1 means that there is a perfect positive linear relationship between variables. This scatter plot depicts such a relationship. It is a positive relationship because high scores on the X-axis are associated with high scores on the Y-axis.



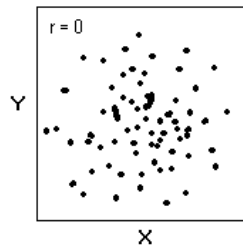
## Negative correlation

A correlation of -1 means that there is a perfect negative linear relationship between variables. The scatter plot to the right depicts a negative relationship. It is a negative relationship because high scores on the X-axis are associated with low scores on the Y-axis.



## Zero-correlation

A correlation of 0 means there is no linear relationship between the two variables, as shown in this graph. Correlation is rarely, if ever 0, 1, or -1.



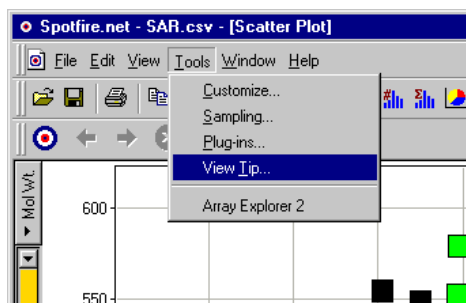
With behavioral data, there is almost never a perfect linear relationship between two variables. The more the points tend to fall along a straight line, the stronger the linear relationship. View Tip calculates the absolute value for  $r$ , and then determines the rank of the scatter plots accordingly.



## Using View Tip

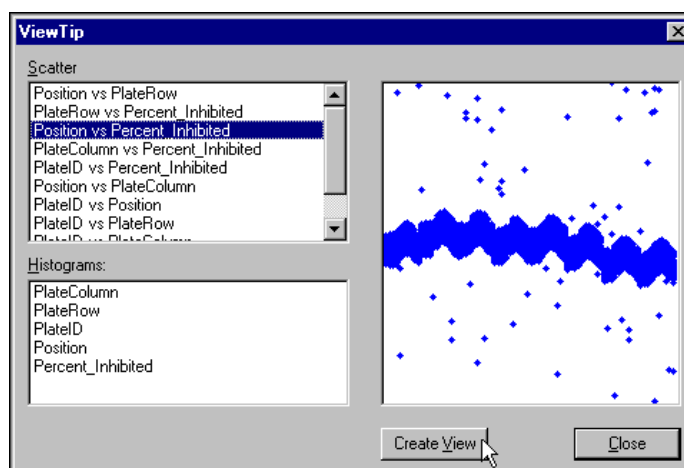
► **To invoke the View Tip function:**

1. Open the data set in Spotfire.net, and select **View Tip...** from the **Tools** menu



The fast way to interesting visualizations

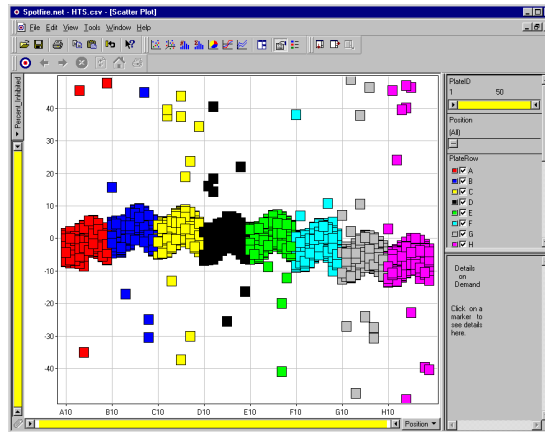
2. Spotfire.net uses heuristics to pair the columns with the strongest correlation, thereby creating potentially interesting plot suggestions. Click on the different View Tips, or use the arrow keys to quickly browse the scatter plot list.



View Tip window for rapid assessment of 2D visualizations

## 12 • VIEW TIP

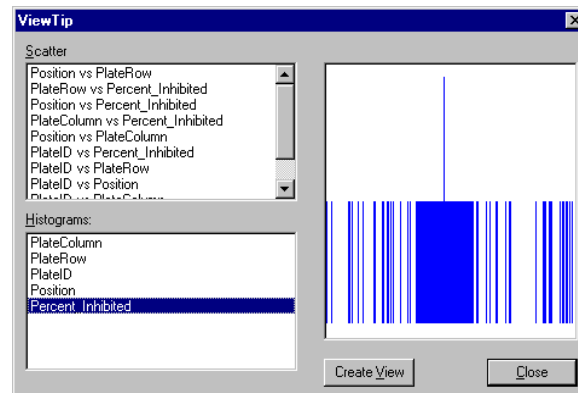
3. When you have found an interesting visualization that you want to investigate further, click **Create View** to generate the corresponding visualization in Spotfire.net.



The visualization created with View Tip.

A new visualization will be created using the columns selected as axes. View Tip will remain open and you can continue the search for meaningful correlation.

4. Go down to the **Histogram** list box to find correlating histograms.



Scatter plots vs. histograms is a powerful combination when analyzing data

5. When you want to examine the visualizations you have made more closely click on View Tip's **Close** button.
6. You can now examine each plot by itself, or all the plots together (choose **Auto Tile** from the **Window** menu).

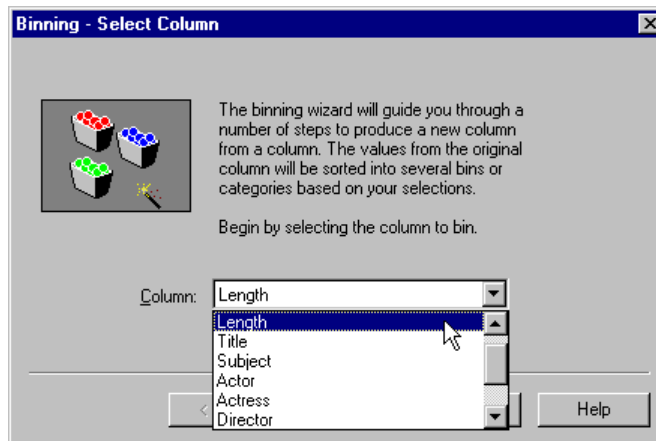
# Binning



*Binning means grouping values in categories according to some unifying property. This can be useful when you want a column to control marker shape, Trellising or some other property that can only take a limited number of unique values.*

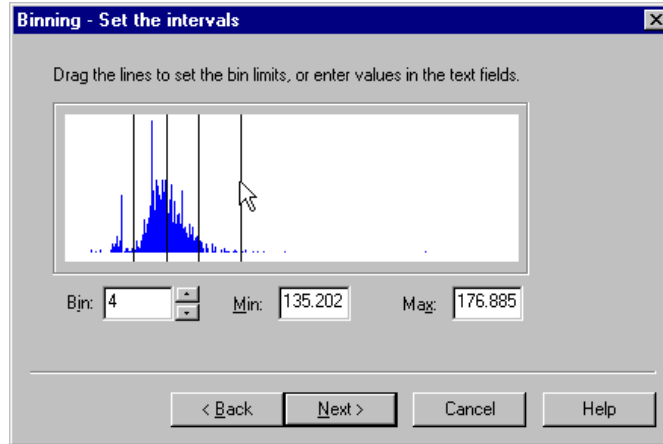
Binning is best described using an example. We will load a data set with information about a number of films. For each film, the length is given in minutes. Using binning, we will create a new column with the categories Very Short, Short, Normal, Long and Very Long.

1. Open the file *film.csv*. The default location is `\Net5\Data\General\Film`.
2. Right-click the **Length** query device.
3. Select **Edit > New Column > By Binning**.
4. Select **Length** from the drop-down list.

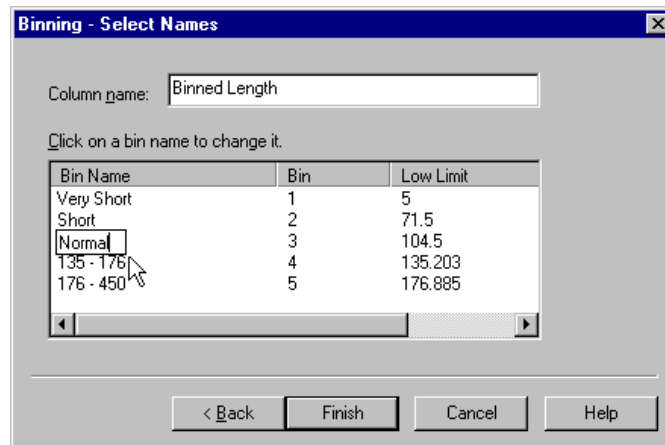


5. Click **Next >**.

6. Select **Even Distribution** and **Number of bins: 5**.
7. Click **Next >**.
8. If desired, adjust the limits by dragging the lines in the graph, or by selecting a bin and entering values in the number fields.

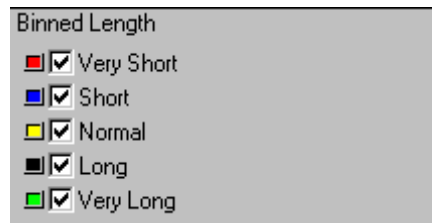


9. Click **Next >**.
10. Enter a name for the new column and for each category.



11. Click **Finish**.

The steps above will result in a new query device appearing:



## Different Binning Methods

Depending on the type of values in the columns—numeric or alpha-numeric, continuous or discrete—you may need to resort to different methods to rearrange the data. Spotfire.net provides a variety of binning methods:

### Even Intervals

The sub-ranges suggested are equal in scope. This is the default method, dividing the value range into equal intervals.

### Even Distribution

The suggested division is made so that the bins each contain an equal number of markers.

### Divide by Standard Deviation

The range is divided into sections as described by the selected standard deviation multipliers. In the example below, the range is divided into the following six subsections ( $\mu$  denoting the average value for the column and  $s$  the corresponding standard deviation):

lower limit  $\rightarrow (\mu - 3s)$

$(\mu - 3s) \rightarrow (\mu - s)$

$(\mu - s) \rightarrow \mu$

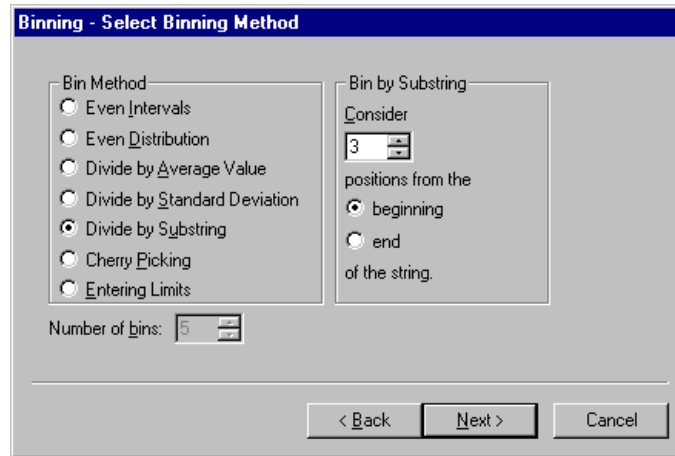
$\mu \rightarrow (\mu + s)$

$(\mu + s) \rightarrow (\mu + 3s)$

$(\mu + 3s) \rightarrow$  upper limit

## Divide by Substring

This method groups records by the initial or final sequence of characters in each value. The exact length of this sequence must be supplied. Suppose the column to be binned contains family names, beginning with Adams and ending with Winter. To bin the records according to the first letter in the name, use Divide by Substring considering *one* character from the beginning.



Binning by Substring

## Cherry Picking

This method allows you to manually select which value to put in each bin. Select **Cherry Picking** as method and click **Next**. A list of all values in the column is shown. Initially all values belong to a bin named *Bin 1*.

Binning by Cherry picking

|              |   |       |
|--------------|---|-------|
| 1            | 1 | Bin 1 |
| 2            | 1 | Bin 1 |
| 3            | 1 | Bin 1 |
| 4            | 1 | Bin 1 |
| 5            | 1 | Bin 1 |
| 6            | 1 | Bin 1 |
| 7            | 1 | Bin 1 |
| Empty Values | 1 | Bin 1 |

Select one or more values, then right-click on the selection. From the pop-up menu either choose an existing bin, or create a new one. If you

create a new bin it will initially have the same name as the last value in the selection.

When all values have been assigned a bin number, click **Next**. The **Select names** dialog will enable you to give each bin a descriptive name.

## Entering Limits

This feature allows explicitly entering the values of the limits to use for each bin. You can enter the interval limits explicitly, or graphically by dragging the interval dividers to the desired value.





# Calculating Columns



*New columns with numerical values can be computed from the current data set by using mathematical expressions. This chapter describes how to create and evaluate such expressions.*

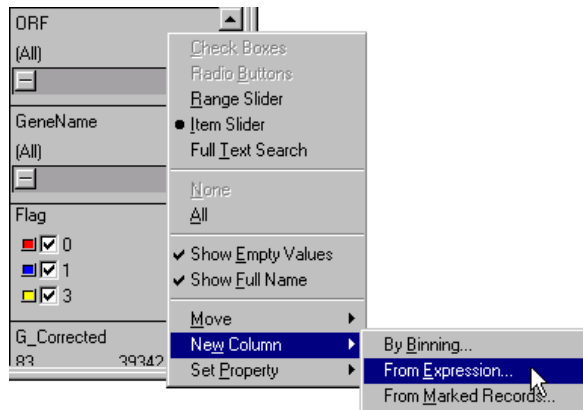
## Calculating Columns—An Example

Occasionally the columns included in a data set do not allow you to perform all necessary operations, or to create the visualizations needed to fully explore the data set. Still, in many cases the necessary information can be computed from existing columns. Spotfire.net provides the option to calculate new columns by applying mathematical operators to existing values.

To illustrate this, we will use the *euro.csv* data set. Among other things, it contains information about population size and exports. Suppose, however, that we are interested in exports per capita. Such a column does not exist, but can be created by dividing the Exports column by the Population column. The following example demonstrates how it can be done.

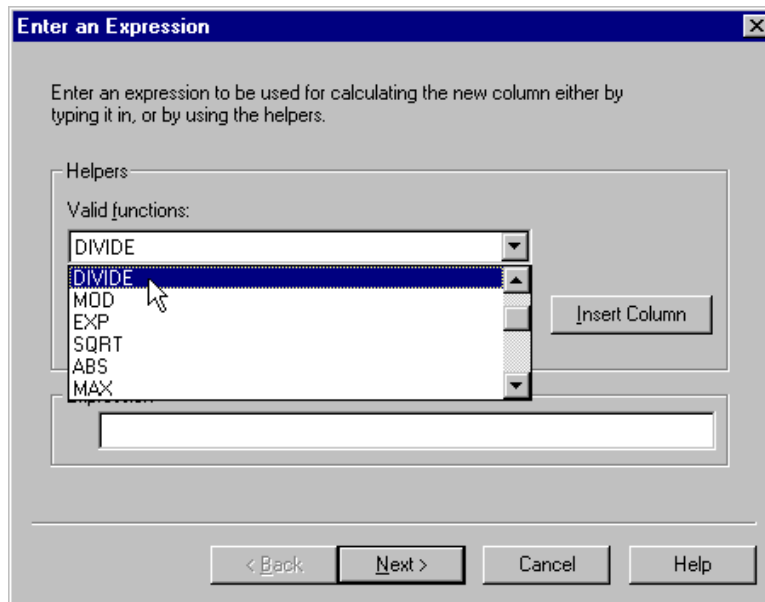
▶ **To divide two columns:**

1. Start fresh by opening *euro.csv* located under `\Net5\Data\General\EuroDemograph`
2. Click on any query device, and then right-click to bring up the context menu. Select **New Column > From Expression**.



Use the context menu to create a new column from expression

3. From the **Valid functions** drop-down list, select **Divide**.



4. For **Argument 1**, select **Exports m\$**, and for **Argument 2** select **Population**.
5. Click **Insert Function**.

Returns argument 1 divided by argument 2. If argument 2 evaluates to zero for any record, an option to leave the field empty is given.

Argument 1:

Argument 2:

6. Click **Next >**.
7. Enter a name for the new column.
8. Click **Finish**.

A new query device is added at the bottom of the Query Device window, showing exports per capita for each record.

| Exports per Capita |             |
|--------------------|-------------|
| 0                  | 0.016071518 |



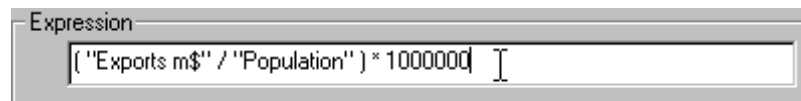
You can also create new columns from within the **Properties** dialog. Go to the **Columns** tab and click **New Column**.

## Advanced Calculations

The example above illustrates how a mathematical expression is created using the built in mathematical functions. Each of these will launch a dialog where you are prompted to enter arguments. Upon clicking **Insert function**, the specified function is inserted in the expression at the position of the cursor.

It is also possible to enter text directly in the **Expression** field, and to insert column names as variables. The following example illustrates this method. We will improve the calculation described in the previous section by multiplying by a million (since exports figures are given in millions, and per capita export never reaches that order of magnitude).

1. Perform steps 1 to 5 above.
2. Place the cursor in the Expression field, and enter \* **1000000** after the parentheses.



3. Complete steps 6 to 8 above.

Similarly, more advanced expressions are written by combining one or more built-in functions with inserted column names and manually entered text. The following sections provide further details.

## Valid functions

Spotfire.net supports the following functions in expressions used for calculating new columns:

| FUNCTION NAME | Syntax                                   | Description                            |
|---------------|--|--|
| ADD           | ADD( Arg1, Arg2 ) or<br>Arg1 + Arg2      |  |
| SUBTRACT      | SUBTRACT( Arg1, Arg2 )<br>or Arg1 - Arg2 |  |
| MULTIPLY      | MULTIPLY( Arg1, Arg2 )<br>or Arg1 * Arg2 |  |
| DIVIDE        | DIVIDE( Arg1, Arg2 ) or<br>Arg1 / Arg2   |  |
| MOD           | MOD( Arg1, Arg2)                         | the remainder from<br>integer division |
| EXP           | EXP( Arg1, Arg2 ) or<br>Arg1 ^Arg2       | Arg1 to the power of<br>Arg2           |
| SQRT          | SQRT( Arg1 )                             | square root                            |
| ABS           | ABS( Arg1 )                              | the unsigned value of<br>Arg1          |
| MAX           | MAX( Arg1, Arg2 )                        | Maximum                                |
| MIN           | MIN( Arg1, Arg2 )                        | Minimum                                |
| LOG           | LOG( Arg1 )                              | Base 10 logarithm                      |
| LN            | LN( Arg1 )                               | natural logarithm                      |
| FLOOR         | FLOOR( Arg1 )                            | Arg1 rounded down                      |
| CEIL          | CEIL( Arg1 )                             | Arg1 rounded up                        |
| SIN           | SIN( Arg1 )                              | sine, the argument is in<br>radians    |
| COS           | COS( Arg1 )                              | cosine, the argument is<br>in radians  |
| TAN           | TAN( Arg1 )                              | tangent, the argument is<br>in radians |

Spotfire.net also supports the two mathematical constants  $e$  and  $\pi$ , written as **E** and **PI**. These are rounded off to 15 digits.

## Entering expressions manually

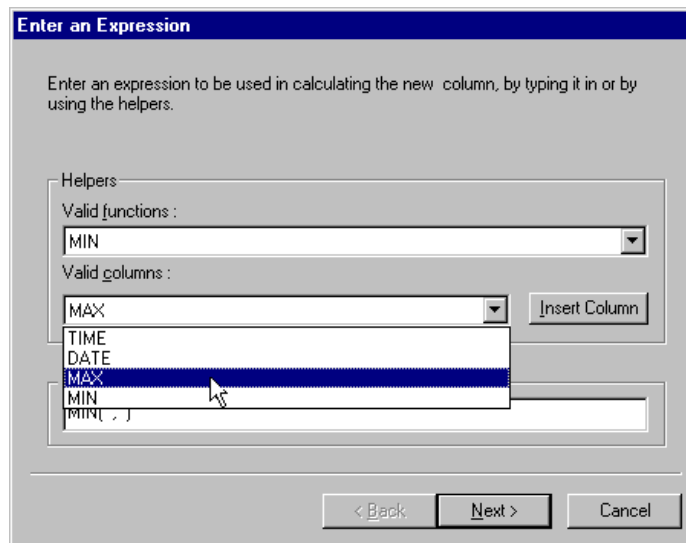
As was mentioned earlier, it is possible to type the desired expression into the **Expression** field in the **Enter an Expression** dialog. When doing so one must be careful about conflicts between column and function names—for example, if the data set includes a column called *MAX*, and the following expression is used:

**MIN( MAX , Length )**

Spotfire.net will interpret *MAX* as the *function* MAX that takes two arguments. The solution is to enclose the column name in quotation marks:

**MIN( "MAX" , Length )**

A safer way to do this would be to use the helpers: set the insertion point in the correct position, select the column from the list box, and finally click **Insert column**. This will automatically add double quotes.



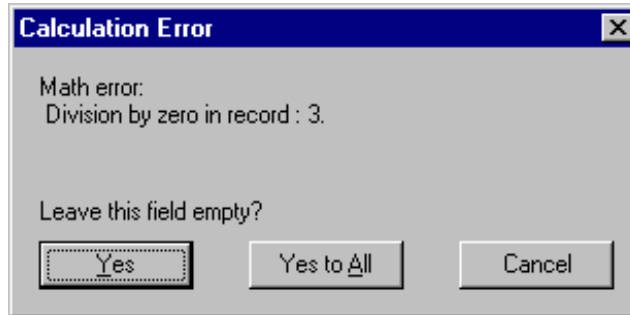
By using the Insert column feature quotes are added automatically



Column names containing spaces, or any of the following characters:  
( ) , + - / \* ^  
must also be enclosed in quotation marks in order to avoid misinterpretations. To avoid potential problems it is therefore best to put quotation marks around all column names used.

## Calculation Errors

If a mathematical error should occur during the evaluation, for example, a division by zero, you are prompted with the option to leave the record's entry empty in the new column. Pressing **Yes to All** leaves all fields presenting the same error empty.



The **Calculation Error** dialog



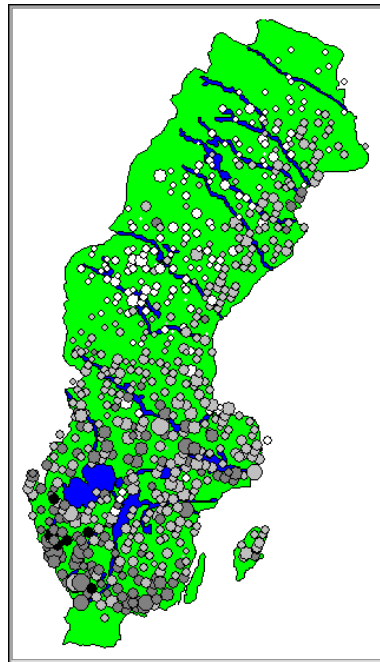


# Background Images and Maps



*This chapter describes how to add background maps, images, and figures to a Spotfire.net visualization.*

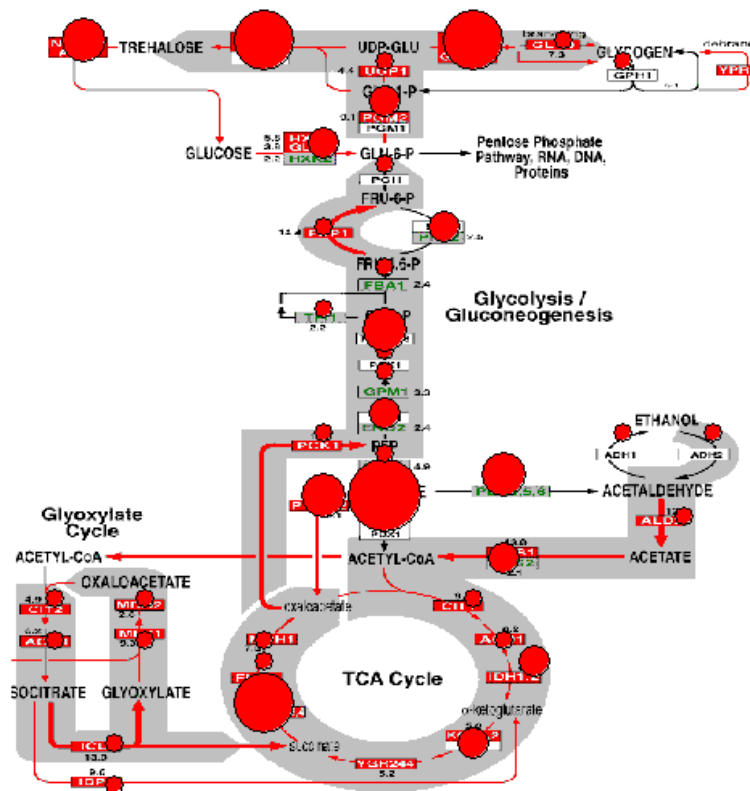
## Image Files and Background Maps



Background maps, images, and figures may be displayed along with visualization markers in a Spotfire.net visualization. Such images can ease the interpretation of the data by providing a suitable graphic context. For example, some aspects of the environmental database described earlier can be examined more easily if data is seen together with a map of Sweden. Another example where an image provides context to the data is the metabolic pathway for yeast genes used in the MicroArray sample file included in the Spotfire.net.

A map of Sweden used as background

## 15 • BACKGROUND IMAGES AND MAPS



The well known metabolic pathway for yeast used as background for the MicroArray data found in the Spotfire.net distribution

## File formats for background images and maps

Supported background image formats:

- Windows bitmap images (\*.bmp)
- JPEG images (\*.jpg)

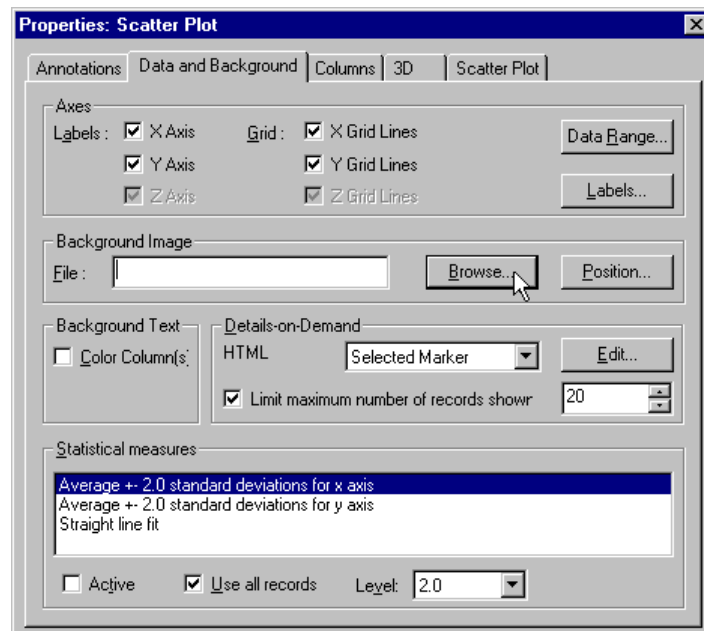
Supported map formats are:

- MapInfo Interchange Format files (\*.mif). These are typically tagged with for example zip codes, which Spotfire.net can use to position visualization markers.
- Spotfire.net mapping files (\*.sfm)

## Adding a Background Image

► **To add a background image to the current visualization:**

1. Open the **Data and Backgrounds** tab from the **Properties** dialog



The **Data and Background** tab in the **Properties** dialog

2. First enter the name of the image file to use in the **File** box. The **Browse...** button can be used to search the file system for appropriate files.
3. Click **Position...** in the **Background Image** section to adjust the positioning of the image relative to the markers.
4. Finally adjust the area to be visible in the visible data range section. This is necessary to get the full picture area even if no markers are to be displayed at the edges.

## Adding MapInfo Background Files

MapInfo is a leading geographic information system software package. MapInfo Interchange Format (MIF) is an ASCII format developed by Mapping Information Systems Corporation (<http://www.mapinfo.com/>) to facilitate the transfer of data from other geographical information system packages to MapInfo.

Spotfire.net can import MIF files to use as background maps. This allows you to position visualization markers in two dimensions based on a single column such as zip code, area code, or area name.

Suppose, for example, that you have a data file with information about various countries, such as name, population, language, etc. However, the data set includes no information about geographical coordinates. By using a MIF file as background, you get not only a background map, but also two new columns of data. Each record is given map coordinates depending on which country it represents.

▶ **To use a MIF file as background:**

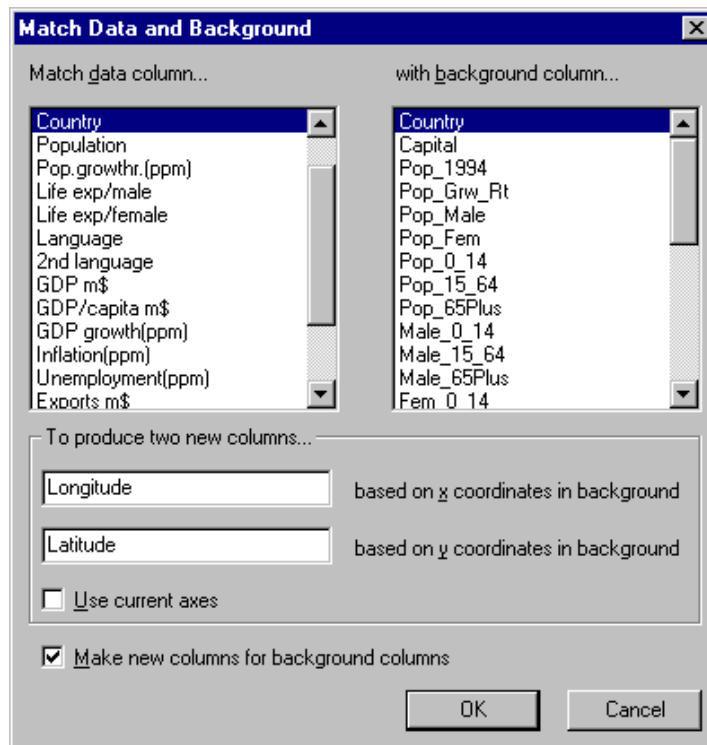
1. Create a scatter plot visualization.
2. Activate the **Properties** dialog and select the **Data and Background** tab.
3. Click the **Browse** button by the **Background Image** field and select **MapInfo Interchange Files**, then browse to the file you want as a background.
4. Click **Open**.

Spotfire.net will display the **Match Data and Background** dialog, where it is possible to match a database column with a tag in the MIF file.



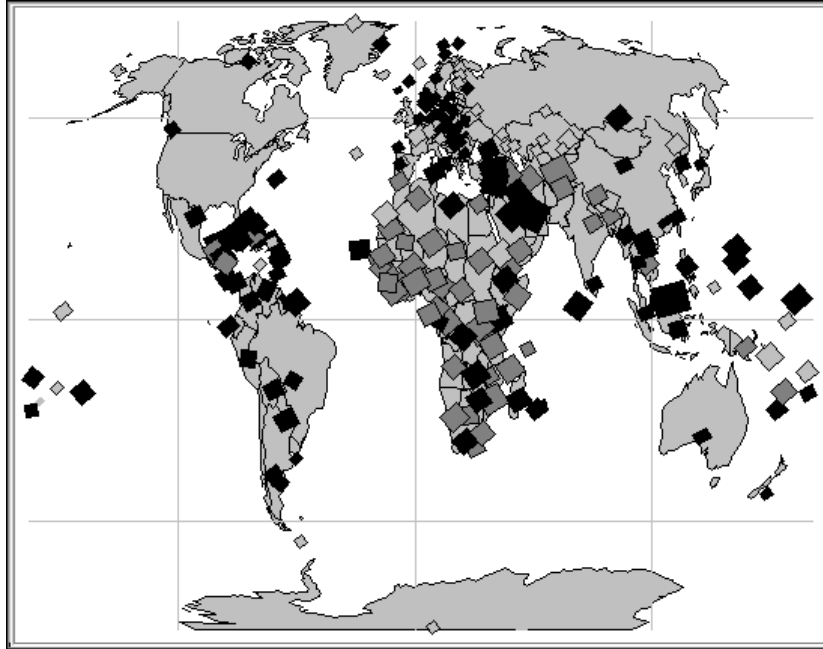
If you try to paste a background image or map into Spotfire.net, the image will not be interpreted as a picture, but as raw data. This will generally not produce any meaningful result.

---



5. In the left field, select the column in your data set that identifies geographical position, in this case **Country**.
6. In the right field, select the MapInfo tag in the MIF file that relates to the same information, in this case **Country**. (The names are not necessarily the same.)
7. In the two text fields, enter names for the new columns that will be added to your data set, specifying X and Y position.
8. Make sure **Use current axes** is unchecked. If you check this option, then no new coordinates will be extracted from the MIF file. You will then only get a background image.
9. If you want to add *all* the information in the MIF file (i.e. not just the coordinates) to your data set, then check **Make new columns for background columns**.
10. Click **OK**. Spotfire.net will try to match data and background.

15 • BACKGROUND IMAGES AND MAPS



Visualization markers positioned based on latitude and longitude

## Spotfire.net Map Files

The Spotfire.net map file (\*.sfm) format first establishes a coordinate system, which relates the data set's coordinate system to that of the polygon set being defined. After that, a number of polygons are defined. A polygon is defined by a fill color, an origin point, a border color, the number of points and the X, Y, and Z coordinates of the points.

The Z coordinates are ignored, however they must be included.

The format of the file as follows (with variables in italics):

```
COORDSYS xmin xmax ymin ymax zmin zmax

ADD
POLY fillcolor xorigin yorigin zorigin
!yes linecolor linewidth
numberofpoints
x1 y1 z1
x2 y2 z2
...
```

A sample file could look like the following text:

```
COORDSYS 0.0 1.0 0.0 1.0 0.0 1.0

ADD
POLY LimeGreen 0.0 0.0 -0.02 0.0
!yes Black 3
5
0.677419 0.976261 0.0
0.680645 0.977745 0.0
0.680645 0.983680 0.0
0.677419 0.985163 0.0
0.677419 0.988131 0.0
```

The *env.sfm* file in the Spotfire.net distribution serves as an example of this background map format.

▶ **To map a data file to a Spotfire.net mapping file:**

1. In your visualization, select the two axes from the data that serve as coordinates.
2. In the **Data and Background** tab in the **Properties** dialog, select an SFM file which has a coordinate system matching the coordinates in the axes file.



# Printing

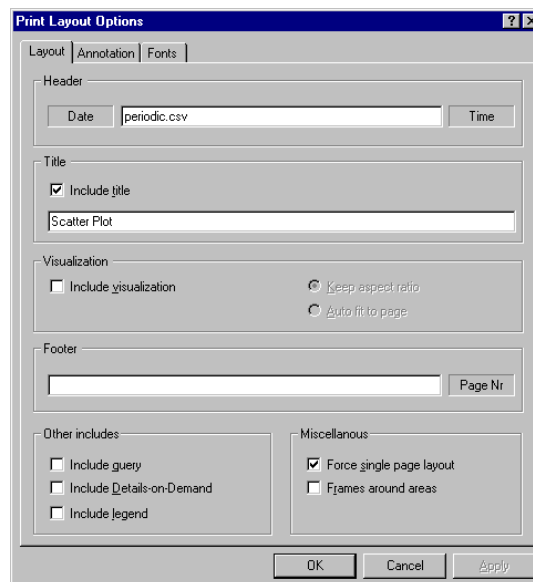


*Spotfire.net users working on the same type of data sets typically want to share analyses and results among each other. Spotfire.net provides excellent facilities for sharing data in a structured way. This chapter describes how to generate printed reports from a Spotfire.net session.*

## Print Layout

Spotfire.net printouts may contain any of the information available in the visualizations.

- ▶ To control which information to include select **Print Layout Options...** from the **File** menu.



## Layout

On the **Layout** tab, you can choose which elements to include. Time, date and page number are always included. You can also include the active visualization, header, footer, title, SQL query, Details-On-Demand or legend.

## Annotation

Under the **Annotation** tab, you can type in any text that you want to include in the printed report.

## Fonts

Under the **Fonts** tab, you can control which font to use for each part of the report.



Use the **Print Preview** option on the **File** menu to see how your printout will appear.

---

## Printing the Document

### ▶ To print the document with the current Print Layout settings:

1. Select **File > Print...**
2. Select a printer from the drop-down list.
3. Optionally, click **Properties** and adjust your printer settings.
4. Click **OK**.

# Importing Data



*This chapter describes how to import data into the program. Spotfire.net stores data in its internal database format to allow instant updates of the visualization when query devices are manipulated.*

Spotfire.net can access data from a variety of sources:

- Spotfire Interactive Reports (SFS)
- Text files (TXT, CSV, SKV)
- Microsoft Excel Workbooks (XLS)
- Databases (OLE DB or ODBC)
- The Windows clipboard

SFS files and text files can be opened directly using the **Open** option on the **File** menu. This is the quickest way to access unambiguously formatted data from a file. Similarly, properly formatted data on the clipboard can be pasted directly into Spotfire.net.

However, if your data resides in a database, or if it needs some kind of preprocessing before being visualized in Spotfire.net, then use **Import Data**. This dialog is designed to help you access your data, whatever the source and whatever preprocessing you require.

Data import is divided into three main steps:

1. Select a data source, such as a file or database.
2. Select methods for data conditioning, such as pivoting, aggregation, or normalization.
3. Optionally, save all settings for data source and/or data conditioning as a favorite. This will allow you to reuse the same data without having to configure data source and data conditioning each time.

The details of each of these steps vary depending on which type of data source or data conditioning method is chosen. A sequence of dialogs will guide you through the required interaction in each case. For example, accessing a database involves setting parameters that are not required when working with a text file. The following sections will deal with each of the three steps above, describing in detail the different cases that may arise.

- ▶ To import data, select **File > Import Data...**

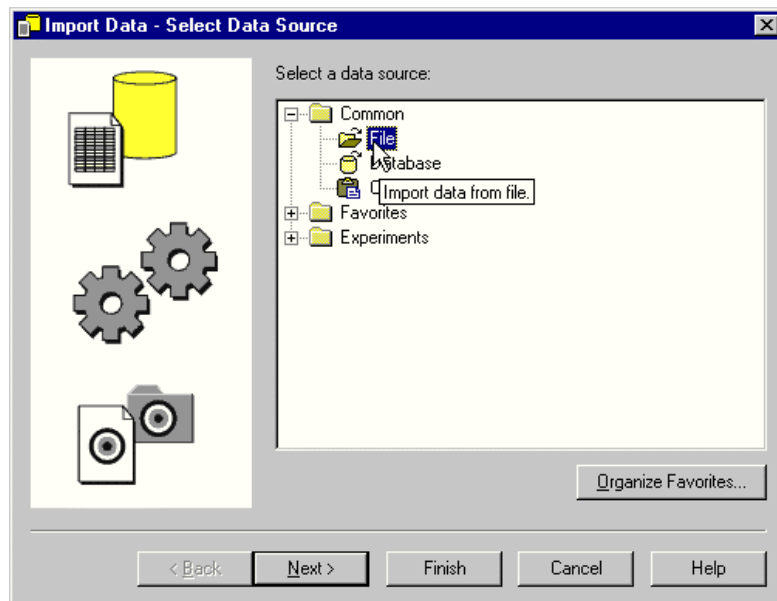
## Selecting a Data Source

The first step in the Import Data dialog is to specify a data source. A data source can be one of the following:

- A file
- The clipboard
- A database
- A data source defined in an earlier session

▶ **To select a data source:**

1. Select **File > Import Data**.



2. Double-click a data source, or select an item and click **Next >**.

## Importing from a file

If the data source is a file, you will be prompted to locate the file. What happens next depends on the format of the selected file. (SFS files require no further steps, since this is Spotfire.net's native format.)

### ▶ To import a text file:

1. Select a file to import (see above).
2. In the Specify Delimiter dialog, select a column delimiter.
3. Click **Next >**.
4. Click on a data type in the table head to change it.
5. Click **Finish**.

### ▶ To import an Excel file

1. Select a file to import (see above).
2. In the MS JET OLE DB dialog, click **OK**.
3. In the Database dialog specify which Excel worksheets and which columns to include. (An Excel file is treated as a database).
4. Click **OK**.

## Importing from the clipboard

### ▶ To import data from the Windows clipboard:

1. Mark a number of cells in a text window, a spreadsheet program or another program providing a tabular presentation of data, e.g. Statistica, Microsoft Excel, Word, etc.
2. Select **Edit > Copy** or hit [Ctrl+C] to copy them to the clipboard.
3. In Spotfire.net, select **File > Import Data...**
4. From the list of data sources, select **Clipboard**.
5. Click **Next >**. This triggers the same dialog as when importing text files.

## Importing from a database

By default, Spotfire.net connects to external sources of data by means of ODBC (Open Database Connectivity). This allows you to import data from virtually any kind of database commercially available.

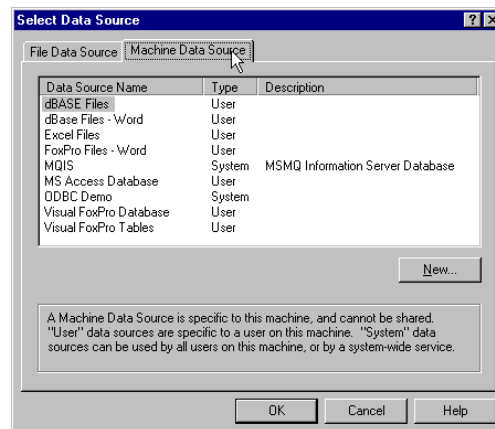
To learn how to set up an ODBC data source, please refer to the database vendor's documentation and the Windows documentation on ODBC. It might be necessary to install ODBC driver software particular to the database used before being able to utilize the ODBC option.

### ▶ To connect to a database using ODBC:

1. Select **File > Import Data....**
2. Select **Database** from the list of data sources.

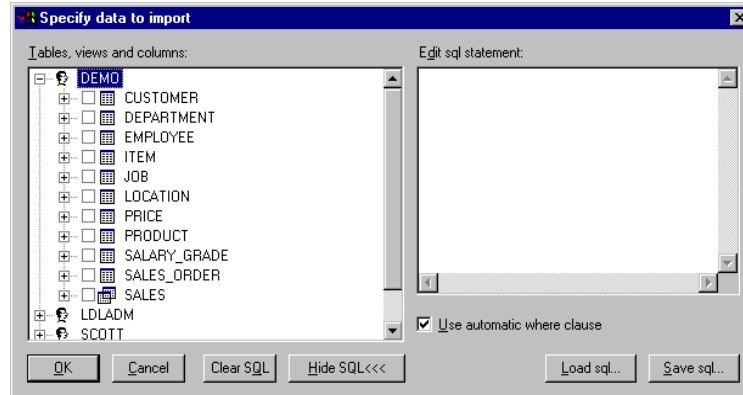


3. Click **Next >**.



4. In the **Select Data Source** dialog, go to the **Machine Data Source** tab. ODBC Select Data Source dialog
5. Select the source from which data should be loaded.

6. Click **OK**.
7. Enter login name and password to log on to the database.



ODBC data import specification

8. Select the database fields to be imported by checking the corresponding check boxes
9. Edit the SQL statement manually if required. This may be necessary for example when joining tables.
10. Click **OK**.

## Importing using a Favorite

Using a Favorite requires that you have defined a data source and preprocessing methods in a previous session and have saved these settings (see the section on Favorites below).

### ▶ To use a Favorite:

1. Select **File > Import Data....**
2. Select a favorite from the list of data sources.
3. Click **Finish** to load the data set, or **Next >** to edit data conditioning before loading.

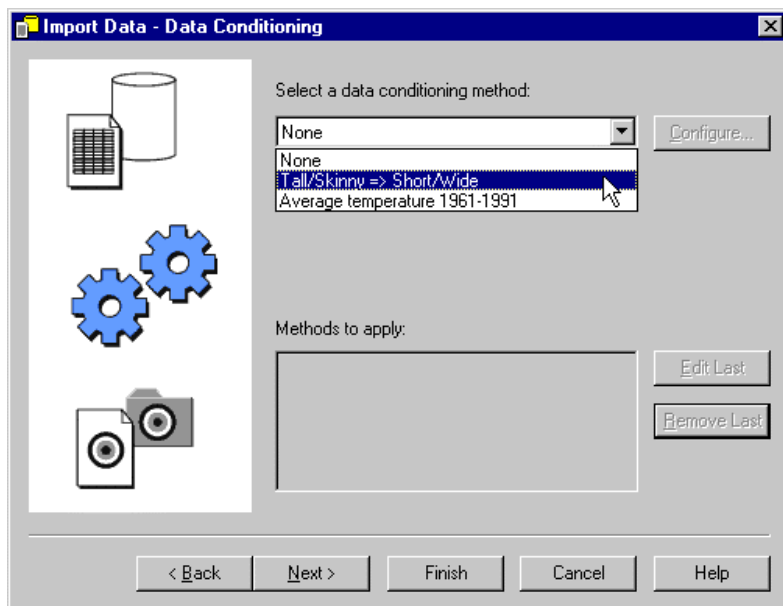


## Data Conditioning

Once you have specified a data source, it is possible to manipulate the data in a variety of ways. The available methods are listed in the drop-down list. Standard methods include the "Tall/Skinny => Short/Wide" conversion (described in detail below), but custom methods designed by Spotfire or a third party may also be available.

▶ **To apply a conditioning method:**

1. Select **File > Import Data**.
2. Select a data source and click **Next >**.



3. Select a method from the drop-down list.
4. Click **Configure...**
5. Make the required settings. (These are specific to each method. An example is provided below.)
6. Click **OK**.

## Tall/Skinny to Short/Wide conversion

Tall/Skinny is a common data format, which is best described using an example. Consider a data set based on a series of temperature measurements over time, taken at various locations. In a Tall/Skinny format, the data may look as follows:

| Location | Month    | Date | Temp |
|----------|----------|------|------|
| London   | May      | 5    | 14   |
| New York | May      | 5    | 23   |
| Tokyo    | May      | 5    | 17   |
| London   | May      | 10   | 16   |
| New York | May      | 10   | 27   |
| Tokyo    | May      | 10   | 16   |
| London   | May      | 15   | 20   |
| etc...   | ...      | ...  | ...  |
| London   | November | 5    | 7    |
| New York | November | 5    | 12   |
| Tokyo    | November | 5    | 11   |
| etc...   | ...      | ...  | ...  |

As we add more observations, the table grows taller, but remains four columns wide. While useful during data collection, this format may not be appropriate for certain types of calculations or visualizations. For example, the entities that interest us are the different cities, so we may want a representation with a single record for each city.

By converting to Short/Wide format, we can create a table like this:

| Location | Avg(Temp) for May | Avg(Temp) for November |
|----------|-------------------|------------------------|
| London   | 17                | 8                      |
| New York | 25                | 14                     |
| Tokyo    | 16.5              | 11                     |

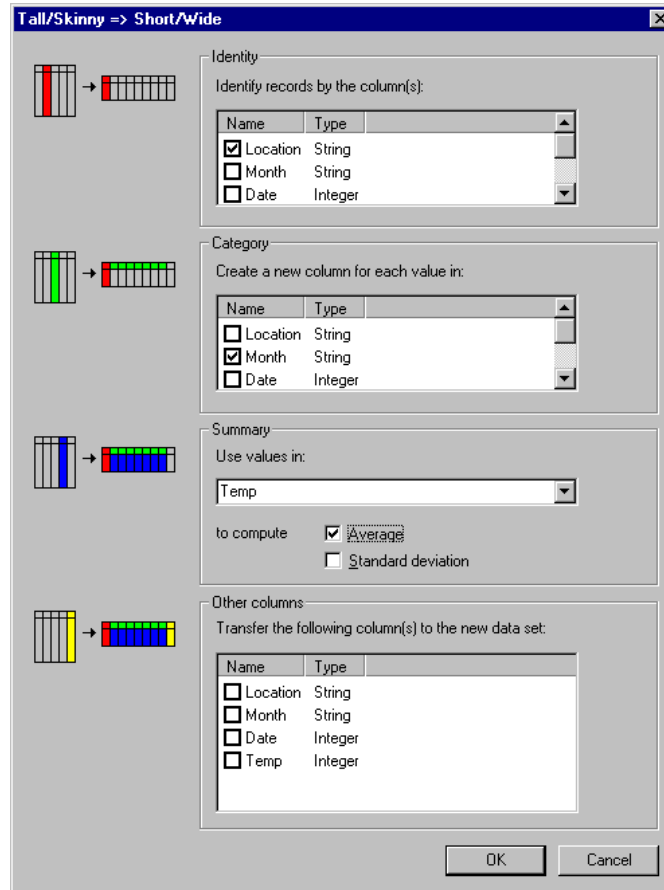
We have thus achieved a more compact, aggregated data format. Each city is represented by a single record, which makes this format very suitable for Spotfire.net profile charts. This example demonstrates four steps:

- Creating a new record for each unique value in **Location**.
- Creating a new column for each unique value in **Month**.
- Summarizing **Temp** for each cell in the resulting grid.
- Selecting **Avg** (average) as the summary method.

The steps below describe how the transformation outlined above is achieved using the Import Data dialog.

▶ **To apply a Tall/Skinny—Short/Wide conversion to your original data:**

1. Select **File > Import Data**.
2. Select a data source, and click **Next >**.
3. Select **Tall/Skinny -> Short/Wide** from the drop-down list.
4. Click **Configure...**
5. Under **Identity**, check the column that you want to use to identify records. In the example described above, we used **Location**.
6. Under **Category**, check the column that you want to use to generate columns in the new table. In the example above, we used **Month**.
7. Under **Summary** select the column that you want to summarize. In the example above, we used **Temp**.
8. Select **Average** or **Standard deviation** as your summary method. In the example, we used **Average**.
9. Under **Other Columns**, check any other columns that you want to include in the new table. None were used in the example.
10. Click **OK**.



## Applying multiple conditioning methods

When applying several transformations on top of each other, it is important to decide in which order the methods should be applied. Once one conversion has been applied, the next conversion will apply to the output from the previous one, and so on.

### ▶ To apply multiple conditioning methods:

1. Select a method from the drop-down list and click **Configure**.
2. Configure the method and click **OK**.
3. Repeat from step 1 as many times as required.

## Favorites

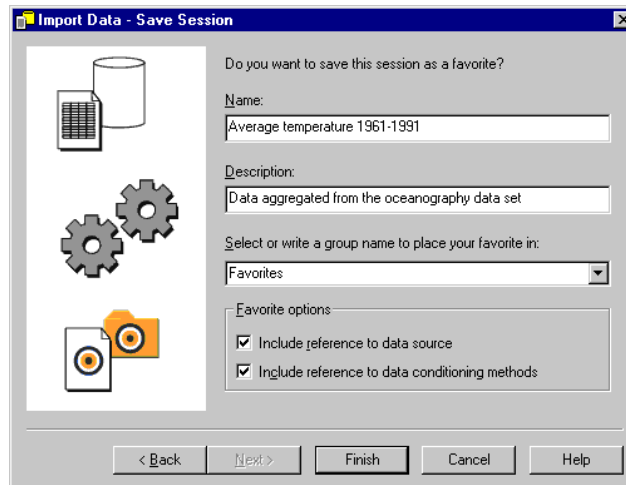
Favorites are combinations of data sources and data conditioning methods. A favorite stores all settings made during a session with the Import Data dialog. A favorite may, for example, specify a database, all the information required to log on, and a data conditioning method.

Favorites can also specify just a data source, or just a preprocessing method.

## Creating favorites

### ▶ To create a favorite:

1. Define a data source as described in "Selecting a data source" above.
2. Define a data conditioning method as described in "Data conditioning".
3. Click **Next**.



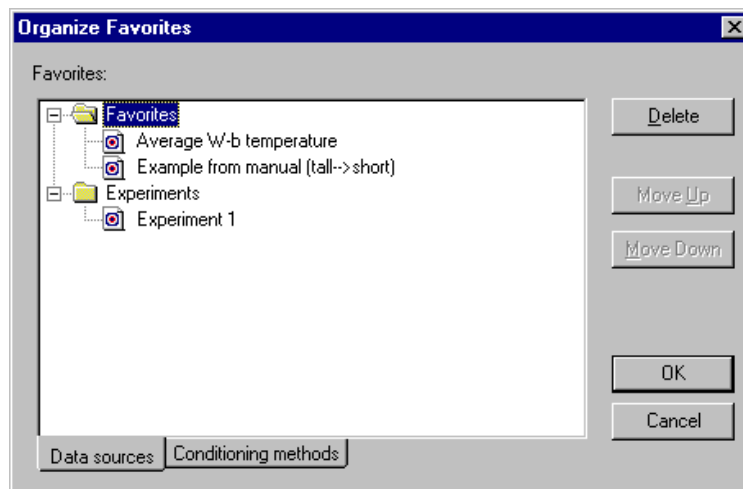
4. Specify a name for your favorite, e.g. "Average temperature 1961-1991". You can also add a description.
5. Under **Favorite options**, select both **Include reference to data source** and **Include reference to data conditioning methods**.
6. Click **Finish**.



Your choice of Favorite options affects how the favorite can be used. If a data source is included, then the favorite can be opened directly from the first step of the Import Data session. If only the method is saved, then you must first select a data source, and then in the second step apply the favorite as a data conditioning method.

## Organizing favorites

Favorites can be organized in folders. Click **Organize Favorites** on the first page of your Import Data session.



Use the **Move Up** and **Move Down** buttons to sort the favorites or move them between folders. Click **Delete** to remove a favorite.

The **Data source** tab contains favorites that are linked to data. These can be loaded directly from the first dialog. The **Conditioning methods** tab contains favorites that are not linked to a data source. These can only be applied from the second step (i.e. you must specify a data source before using them).

## Using favorites to access data

Once a favorite has been defined, it will appear as an item in the list of available data sources in the first Import Data dialog (or as a conditioning method if it contains no data source reference). Select a favorite, then click **Finish** to load the data directly. Click **Next** if you want to add or edit data conditioning methods before loading the data.

## Connecting to a data source using OLE DB

An OLE DB data provider allows native access to data, such as a SQL Server or Oracle database. Using an OLE DB provider, Spotfire.net can retrieve and manipulate data from a wide variety of data sources, not just relational databases. When you create a data link, you specify the OLE DB provider that is designed to work with your data. You can create data links for any Microsoft or third party OLE DB provider. The following providers are included with the Microsoft data access components:

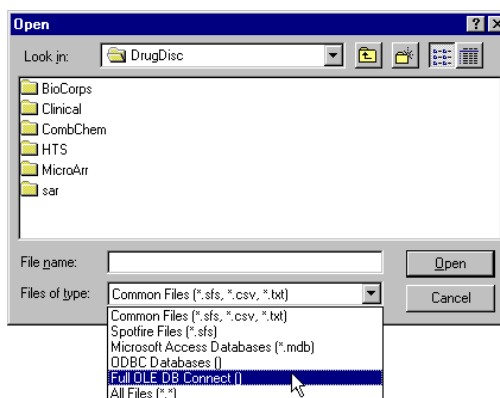
- Microsoft Jet 3.51 OLE DB Provider
- OLE DB Provider for Oracle
- OLE DB Provider for SQL Server
- OLE DB Provider for ODBC Drivers



For more information about OLE DB providers, see the OLE DB Programmer's Reference. This documentation is available in the Microsoft Data Access SDK.

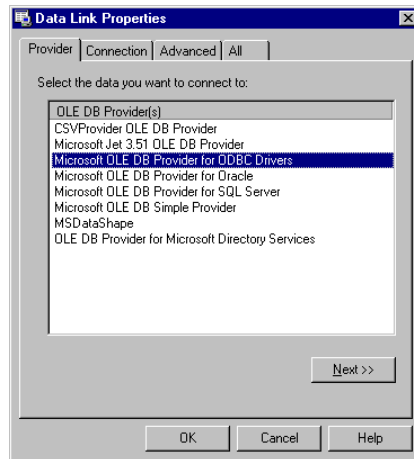
### ▶ To import data using an OLE DB provider:

1. Click **Open** on the **File** menu, and select **Full OLE DB Connect** from the **Files of type** drop down list



Connecting to an external data source using OLE DB

2. Use the **Data Link Properties** dialog box to specify initialization properties for your OLE DB provider. Data link properties may vary depending on your OLE DB provider. **Next** navigates to the **Connection** tab for the selected OLE DB provider.

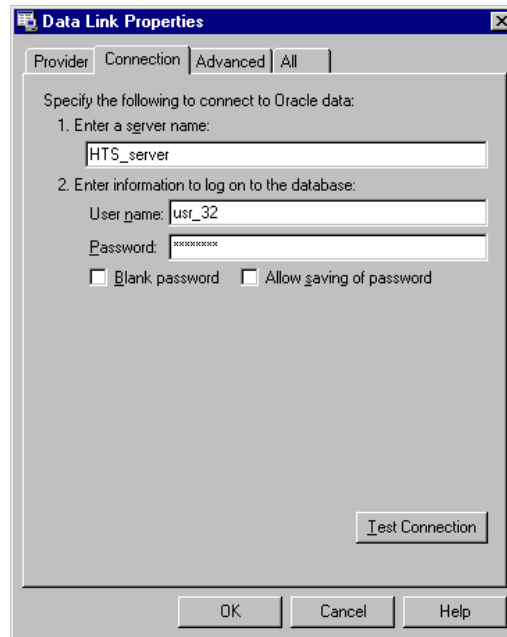


The Data Link Properties dialog

Before you can access data from OLE DB, you must provide specific connection information such as:

- The type of *data* that you want to access
  - The *server* on which the data resides
  - The *database* in which the data is stored
3. Use the **Connection** tab to specify how to connect to Oracle data. Enter the server where the database you want to access is located. Click **Test Connection** to attempt a connection to the specified data source. If the connection fails, ensure that the settings are correct. For example, spelling errors and case sensitivity may cause a connection to fail. Click **OK** to connect.





The provider specific **Connection** tab; Oracle



For more information about advanced initialization properties, see the documentation provided with your OLE DB provider. Use the **All** tab to view and edit all of the OLE DB initialization properties that are available for your OLE DB provider.

## Database Size Limitations

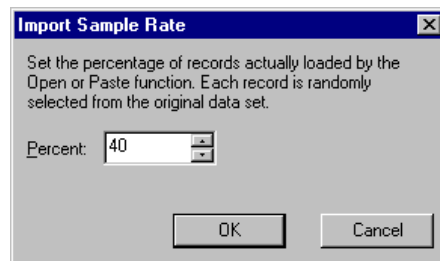
In principle, Spotfire.net can deal with databases of any size. However, because of computer memory and processor requirements, performance will degrade as the size of the data set grows.

Data sets of up to 60 000 – 70 000 records with 10 to 15 columns pose no problems on a 150 MHz Pentium with 32 Mb RAM. Even larger data sets can be handled if some latency in visualization updating can be tolerated.

If speed is of some importance, and if finding an approximate trend is acceptable, then the database should be sampled (see below).

## Sampling a Database

When the database to be imported is very large, greater than say 200 000 to 300 000 records (depending on the computer's memory and processor), it may be necessary to extract a subset of the data, in other words: sample the database. Spotfire.net allows for this through the **Sampling** item in the **Tools** menu. When **Sampling** is selected the dialog below is displayed.



Sampling dialog

By entering a percentage or using the arrows, you specify the probability that a single record in a data set will be loaded into Spotfire.net. If the sampling rate is set to 40%, Spotfire.net will load only 40% of the records of the database, regardless if the source is a text file, an ODBC connection, or the Windows clipboard. Spotfire.net applies the probability to each single record as it loads the data set; no patterns are used.

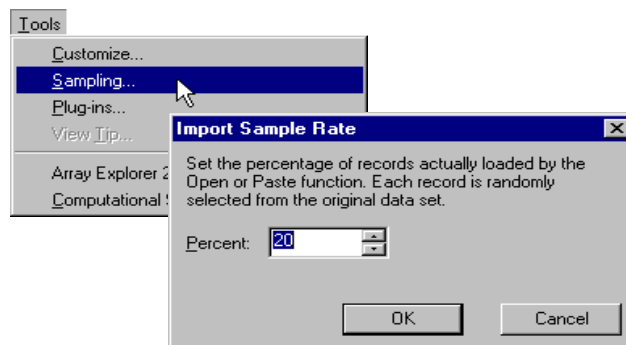
The default probability is 100% - all records are included. Once a particular probability has been specified, this will be used in all subsequent data loads. The sampling probability is in effect until a new probability is specified. Spotfire.net stores the current probability in the registry so the probability is stored from session to session.

## Handling very large data sets

Exceptionally large data sets (1,000,000 records or more) are cumbersome, because they take a long time to load, require a lot of memory, and do not respond instantly to queries. It is therefore recommended that such data is first sampled for a preliminary analysis, and that only a sub set is loaded for a more detailed study.

▶ **To analyze a large data set:**

1. Set a sampling rate of 20%



2. Load the data set.
3. Try to find interesting sections of the data using the query devices
4. Export the query device settings as an SQL-query (see page 180).
5. Restore the sampling rate to 100%.
6. Apply the exported query to the whole data set, thus loading only the interesting section.



The example above will work only if the data source is a database, since you must be able to use SQL during import.

---



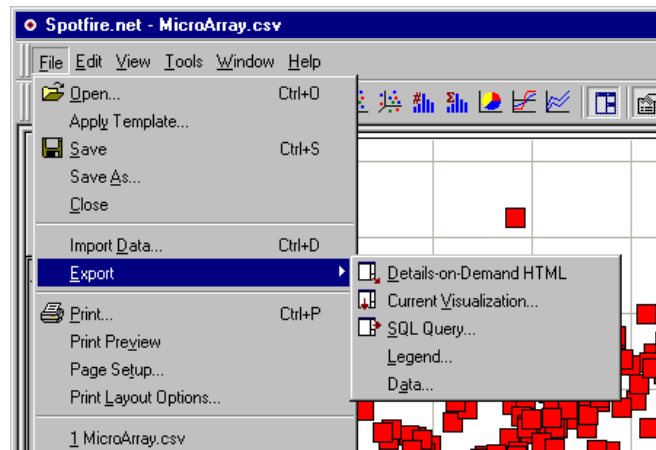
# Exporting Data



*This chapter describes how to export various parts of a Spotfire.net analysis.*

Under the **File** menu, **Export** sub-menu, there are five commands:

- Details-on-Demand
- Current Visualization
- SQL Query
- Legend
- Data



Available export options accessible from the **File** menu

Each of these enables the exportation of some portion of the Spotfire.net data. The first three of these export possibilities are also available by clicking buttons on the toolbar.

## Details-on-Demand HTML



Spotfire.net provides the possibility to export the information about the marked or the active markers. If the **Details-on-Demand (HTML)** option is checked under the **View** menu, the **Export > Details-on-Demand** option is available. Selecting it will launch an external browser, Netscape or Internet Explorer, in which the details of the records will be presented according to the templates that can be edited under the **Properties > Data and Background** tab.



10:  
The Details-on-Demand window

| ID  | Mol Wt. | Plate # | Row | Column | LC/MS | LD (50) | In-vivo Activity | Assay #1 (IC50) | Assay #2 (IC50) | Efficacy | Kd #1 | Kd #2  | Half life (min) | Bioavailability | cLogP | Pka | Rf |
|-----|---------|---------|-----|--------|-------|---------|------------------|-----------------|-----------------|----------|-------|--------|-----------------|-----------------|-------|-----|----|
| 9   | 346.9   | 1       | A   | 2      | pass  | 1.45    | 35               | 18.705          | 113.96          | 35       | 14.36 | 200.38 | )               | 58              | -0.44 | 4.7 | H  |
| 407 | 385.51  | 5       | G   | 3      | pass  | 49.52   | 54               | 32.669          | 95.163          | 54       | 9.62  | 54.14  | )               | 55              | -3.89 | 4.8 | C  |
| 442 | 428.54  | 5       | B   | 8      | pass  | 977.06  | 46               | 24.833          | 142.499         | 46       | 14.6  | 212.87 | 276             | 49              | -2.74 | 7.4 | C  |
| 460 | 496.06  | 5       | D   | 10     | pass  | 41.28   | 30               | 21.44           | 113.519         | 30       | 1.46  | 143.19 | )               | 21              | 0.49  | 3.4 | C  |

**SAR1.sfs**

Database D:\Work\Spotfire\SAR1.sfs

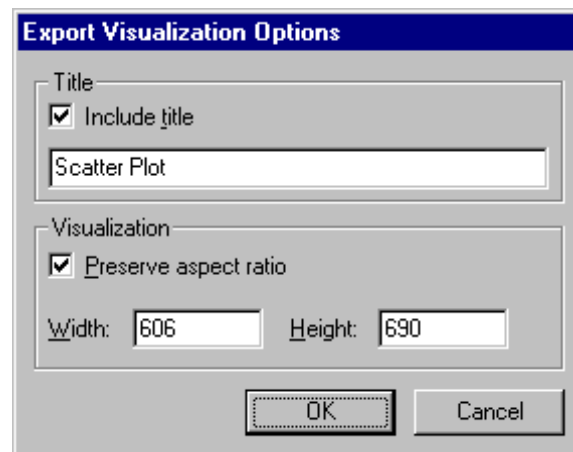
Details-on-Demand exported to Internet Explorer

## Visualization

- ▶ To save an image of the current visualization to a file:



1. Select **Export > Current Visualization**.



2. Enter the title to be included at the top of the image.
3. Enter the size of the exported image. If you check **Preserve aspect ratio**, the height-width ratio of the exported image will be the same as that of the visualization on screen.
4. Click **OK**.
5. Select a location, name and format for the file. The formats available are Windows BMP and JPEG .

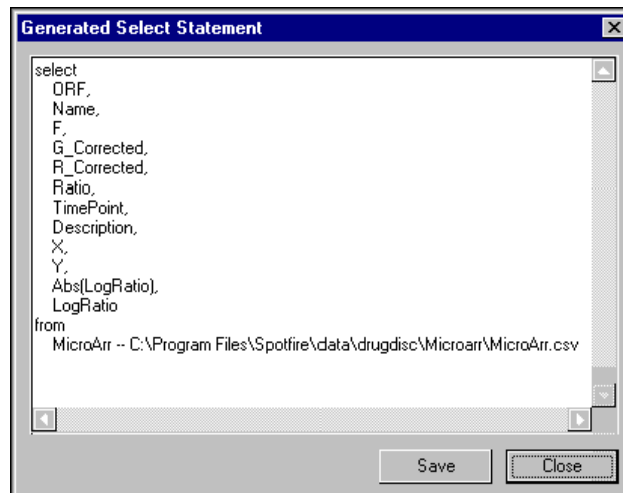


The current visualization can be copied to the Windows Clipboard as a vector graphic (Windows Metafile) to be pasted into another application. This is accomplished through the **Copy Special > Current Visualization** command under the **Edit** menu. In the Metafile format each block of text, marker, line, etc. is an independent object that can be manipulated.

## SQL Queries



The current query device settings can be exported as an SQL database query. This is done by selecting the **Export > SQL** query option from the **File** menu. When you do this, a dialog containing the query appears. The query can be edited at this stage by clicking in the edit box and entering or erasing text. By clicking **Save As** the query can be saved to the file system.




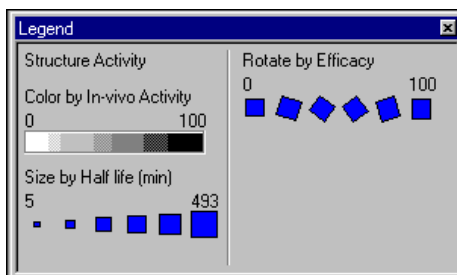
Transforming the query device settings into an SQL query



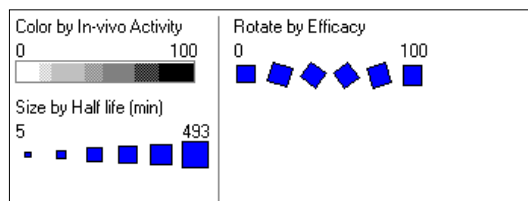
## Legend Images

This option will save an image of the legend describing the columns that are used for size, color, rotation, etc. To make this option available you must first display the Legend. There are three ways to do this:

- Select **Legend** command from the **View** menu
- Use the keyboard shortcut [Ctrl+L]
- Click the  button on the toolbar



The legend in the Details-on-Demand window (above) and the bitmap exported (below)



When you select the **Export > Legend** option, a **Save As...** dialog opens and you can browse to where you want to save the legend image; you may choose .bmp or .jpg format.

## Data

All the data loaded into memory and used by Spotfire.net can be exported using this option. Records that have been removed by deleting brushed markers will not be included in the exported data. All other records—even those not currently portrayed in visualizations, or those deselected by query devices—will be exported.

When you select the **Export > Data** option, a **Save As...** dialog opens and you can browse to the location where you want to save the file, and save it under any viable name you wish. Available formats are TXT (tab separated columns), CSV (default list separator) and SKV (semicolon).

# Spotfire.net File Types



*This chapter describes Spotfire's native file formats, and how they are used in Spotfire.net.*

Apart from supporting standard text file formats, Spotfire.net also handles a number of native file types:

|     |  |
|-----|--|
| SFS | Spotfire Interactive Report file, stores data and visualization settings in binary format                          |
| XSF | Same content as SFS, but in XML based format   |
| SFT | Template file, stores visualization settings only  |
| SFA | Spotfire Application file, similar to SFS, but when loaded, the document does not become associated with the file. |

## SFS Files—Spotfire Interactive Reports

All information and settings in a Spotfire.net session can be saved as a Spotfire Interactive Report. This file has the file extension *.sfs*, and contains information such as:

- Current settings of the query devices, both selected values and the query device types
- Visualizations
- Coloring schemes
- Information on how to regenerate derived columns, i.e. those created by calculation and binning

- Attached background image files
- Data (either as a reference to a data source, or embedded in the SFS file)

By saving your current session as an SFS file, you can pick up your work later exactly where you left it off.

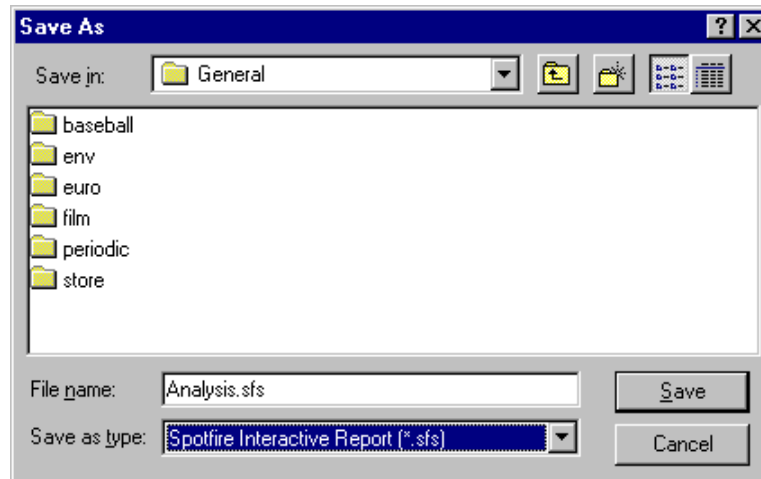


Spotfire Interactive Report file icon

The handling of Spotfire Interactive Reports differs depending on whether you include the data set in the report (static SFS file), or reference the data with a link (dynamic SFS file).

## Saving SFS files

The Spotfire Interactive Report (SFS) is saved by selecting **Save As...** in the **File** menu. Make sure that **Spotfire Interactive Report (\*.sfs)** is selected in the **Save as type** list box.

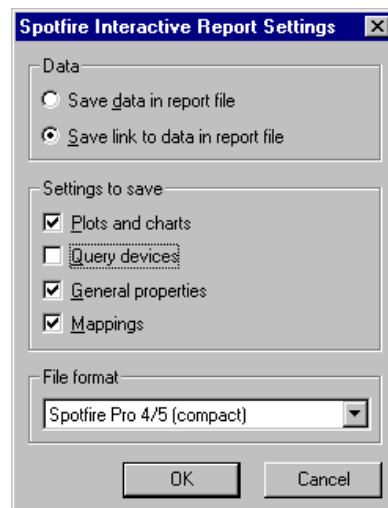


The **Save As** dialog

## How the data is stored

If the data was loaded from a source other than the clipboard, you will get an option on how to save the actual data used in the Spotfire.net session. There are two options available:

1. Saving a link to the data in a *dynamic* report file—no actual data is stored in the SFS file, merely a reference describing how to retrieve, and how to view the data
2. Saving the data in a *static* report file—the actual data used in the session is saved in the SFS file. This is the only option available if the data was loaded from the clipboard.



**Spotfire Interactive Report Settings** dialog



For a number of reasons, saving the data as a dynamic SFS file should be the preferred method, since it:

1. Ensures data integrity – avoids inconsistencies
2. Reduces the file size – essential if you want to distribute the data
3. Ensures that the data is up-to-date

However, since a link only references its data source, the dynamic SFS file is no good if the actual data is removed or damaged.

The saved link includes the file name if the data was imported from a text file, or the ODBC data source name and the SQL query used if the data was retrieved from a database using ODBC.

The way the data was saved has some implications on how the Spotfire Interactive Report can be used and how it will behave:

- If the data itself is saved in the report file, the session becomes static—in other words, if the source for the data is updated, the interactive report will still show the old data. If only a link to the data was saved, the *.sfs* file reflects all changes made to the data source each time the file is opened.
- If the report is to be distributed, the recipients must have access to the data source to be able to link to the data. Static report files containing the actual data create no distribution problems.
- A static SFS file is considerably larger than a dynamic SFS file, since the latter only contains a link to the data

### Settings to save

When saving a **Spotfire Interactive Report** you have the option to include or exclude various properties of your Spotfire.net session.

- **Plots and charts** – all visualizations and their settings (e.g. coloring, shapes, size, rotation, Trellising, annotations, etc)
- **Query Devices** – the current query device settings
- **General Properties** – properties associated with the Spotfire.net Desktop, such as the settings on the View menu (Details-on-Demand mode, Legend, etc) and Window menu (Auto-tile, etc), sampling rate, etc
- **Mappings** – new columns created by binning, calculation or marking, new column names, etc



If the sampling rate is stored in an SFS file, and the file is then opened, the sampling probability is read first, and affects the loading of the actual data referenced/contained in the *.sfs* file. The visualizations will show another set of records but sampled at the same frequency as the first visualizations.

---

## File format

When saving an SFS file, you have the option to save it in a format compatible with Spotfire Pro, versions 3 or 4. To do this, select **Spotfire Pro 3 & 4 Compliant** in the list box under File Format. However, if the file will only be used by Spotfire.net users, we suggest that you use the **Spotfire Pro4 /.net 5** option, as this is more compact and more efficient.

## XSF Files

XSF files store exactly the same information as SFS files (see above), but in an XML based format. Since the files are in text format rather than binary, they can be searched from a file system. XML also makes the structure of the content more clear, making it easy for developers of third party software to support this file type.

## SFT Files

SFT (Template) files are like SFS files but without any data content. SFT files are used as templates for quickly applying a whole range of visualization settings to a new data set.

### ▶ To create a template:

1. Open a data file.
2. Make the required visualization settings (query devices, coloring, multiple visualizations, etc).
3. Select **File > Save As...**
4. Enter a name for the template.
5. From **Save as type**, select **Spotfire Template Files**.
6. Click **OK**.
7. Check the settings that you want to include.
8. Click **OK**.



For a template file to work properly, it should only be applied to data with the same column names as the data set used to create the template. If applied to a data set with different column names, only those settings that are independent of column name will be used (e.g. visualization types, fixed color, fixed shape, size and jittering, but not query device settings, Color by, etc).

---

▶ **To apply a template:**

1. Open a data file.
2. Select **File > Apply Template**.
3. Select an SFT file.
4. Select which settings to apply.
5. Click **OK**.



You can use any SFS file in the same way as a template file, Applying the settings in the SFS file to your loaded data set. Use the procedure described above, but select an SFS file instead of SFT.

---

## SFA Files

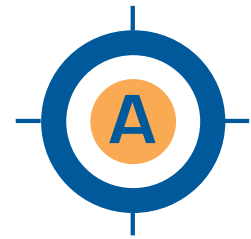
In their simplest form, SFA (Spotfire Application) files are identical to SFS files. You can save your analysis as an SFA file and open that file again just like an SFS file. The only difference is that when you open an SFA file, the resulting analysis is not automatically associated with that same file. You can only save it by using **Save As** and enter a name. This helps you avoid accidentally over-writing the application file.

The reason for this behavior is that SFA files are designed to incorporate an application that is launched when opening the file. This application, implemented using scripts and possibly involving a visible user interface, could for instance establish database connections and perform preprocessing. Although this is possible with SFS files too, SFS files may be accidentally over-written when you attempt to save your data.

Please contact Spotfire for more information on how to integrate code in SFA files.



# Reference Guide to Regular Expressions



*This appendix serves as a quick reference guide to writing search strings using regular expressions. It is not a complete description of available regular expressions, but a brief introduction to a language that looks foreign, but is very powerful.*

## Introduction to Regular Expressions

Regular expressions constitute the standard pattern matching language, used heavily for searching and parsing text in most UNIX and Perl software.

Regular expressions allow you to construct a template or pattern for what you're searching for. You can then use your pattern template to determine whether a given text fragment matches your pattern.

For example, suppose you wanted to find all records beginning with the letter *n*. You would then want to construct a regular expression that would have the meaning "has a letter *n* as the first character on the line". The regular expression for this is:

```
^n
```

The ^ in regular expression syntax means "beginning of line". Therefore ^n means "line beginning with the letter n".

Next, we want to find records whose first word begins with *n*, and whose second word begins with the letter *g*. Let's say we also don't care whether the letters are upper or lower case. Here is the regular expression meaning "all lines whose first word begins with *n* and whose second word begins with *g*":

```
^n.\s+g.+
```

The three main components of the pattern are 'n', '\s', and 'g'. The *n* means the letter *n*, the *g* the letter *g*, and the \s is a special character

meaning "white space", or "any spaces", which could be a space, a tab, or a line break. Notice that both the n and the g are followed by a period. The period in regular expression syntax means "any character". Therefore, n. or g. means n followed any character and g followed by any character, respectively.

Finally, notice the 3 plus signs. A plus sign (+) in regular expressions means "one or more" and it always applies to the character immediately preceding it. Therefore, a+ means "one or more a's", "\s+" means "one or more white spaces", etc. Since . means "any character", then, .+ means "one or more of any character", and "n.+" means "the letter n followed by one or more letters".

Therefore, ^n.+ \s+g.+ means "an n at the beginning of the line followed by one or more characters followed by one or more spaces, tabs, etc., followed by a g, followed by one or more characters".

For a more thorough tutorial on regular expressions, a good source is the O'Reilly & Associates nutshell guide to either Sed & Awk or Perl. These are UNIX programming languages which are based heavily on regular expression pattern matching, and the O'Reilly guide for each of these languages contains a chapter devoted to regular expressions.

Microsoft's MSDN web site offers a complete reference to the symbols that can be used with Spotfire.net full-text query devices.

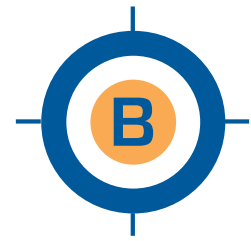
## List of Regular Expression Symbols

| Symbol       | Description   |
|--------------|---|
|              | <b>Positions</b>  |
| ^            | Only match the beginning of a string.<br>"^T" matches the first "T" in "To Be Or Not To Be" |
| \$           | Only match the ending of a string.<br>"t\$" matches the last "t" in "A cat in the hat"      |
| \b           | Matches any word boundary<br>"ly\b" matches "ly" in "probably not"                          |
|              | <b>Literals</b>   |
| Alphanumeric | Matches alphabetical and numerical characters literally.                                    |
| \n           | Matches a new line  |
| \f           | Matches a form feed   |
| \r           | Matches carriage return   |
| \t           | Matches horizontal tab  |
| \v           | Matches vertical tab  |
| \?           | Matches ?   |
| \*           | Matches *   |
| \+           | Matches +   |
| \.           | Matches .   |
| \            | Matches   |
| \{           | Matches {   |
| \}           | Matches }   |
| \\           | Matches \   |

**A • REFERENCE GUIDE TO REGULAR EXPRESSIONS**

|        |  |
|--------|--|
| \[     | Matches [  |
| \]     | Matches ]  |
| \(     | Matches (  |
| \)     | Matches )  |
|        | <b>Character Classes</b>   |
| [xyz]  | Match any one character enclosed in the character set.<br>"[a-e]" matches "b" in "basketball".                           |
| [^xyz] | Match any one character not enclosed in the character set.<br>"[^a-e]" matches "s" in "basketball".                      |
| .      | Match any character except \n.   |
| \w     | Match any word character. Equivalent to [a-zA-Z_0-9].  |
| \W     | Match any non-word character. Equivalent to [^a-zA-Z_0-9].   |
| \d     | Match any digit. Equivalent to [0-9].  |
| \D     | Match any non-digit. Equivalent to [^0-9].   |
| \s     | Match any space character. Equivalent to [ \t\r\n\v\f].  |
| \S     | Match any non-space character. Equivalent to [^ \t\r\n\v\f].   |
|        | <b>Repetition</b>  |
| {x}    | Match exactly x occurrences of a regular expression.<br>"\d{5}" matches 5 digits.  |
| {x,}   | Match x or more occurrences of a regular expression.<br>"\s{2,}" matches at least 2 space characters.                    |
| {x,y}  | Matches x to y number of occurrences of a regular expression.<br>"\d{2,3}" matches at least 2 but no more than 3 digits. |
| ?      | Match zero or one occurrences. Equivalent to {0,1}.<br>"a?s?b" matches "ab" or "a b".                                    |
| *      | Match zero or more occurrences. Equivalent to {0,}.  |
| +      | Match one or more occurrences. Equivalent to {1,}.   |

# Glossary



## **API**

The term API stands for **A**pplication **P**rogrammer's **I**nterface, and is simply a set of functions available to a programmer. These functions cover a wide range of tasks, such as controlling visualization settings, importing and exporting data, etc.

This functionality allows strong integration with legacy systems and novel applications.

## **Bar Charts**

A visualization displaying the data as bars. The value from each record is added to the bar's total value/ Y-axis height.

## **Binning**

The process of grouping values together in a way that each *Bin* contains values that, in some important respect, are similar.

## **Browser**

A program able to display pages encoded in HTML.

## **Categorical Coloring**

Each unique value in a column is associated with a specific color.

## **Check Boxes**

A Query Device enabling toggling of each data value, to control which will be shown in the visualization and which will not. This query device is only available for columns that contain no more than 500 unique values.

## **Column**

Corresponding fields in a set of records constitute a column. (Records go by the name 'rows', in spreadsheet programs).

## **Continuous Coloring**

Coloring whereby the values in the column determine the shade displayed. The (dissimilar) Begin and End colors selected in the Customize dialog represent the minimum and maximum values in a column. The markers representing records having intermediate values are colored with a blend of the two; depending on their value.

**CSV**

**Column Separated Values.** This is a standardized way to save data sets. Each data record is stored as text, using the default list separator as column separator, and a line feed and return character to mark the end of each record. The default list separator is a property of the Regional Settings of your Windows system.

**Data Conditioning**

Any kind of processing performed prior to visualizing the data, such as normalization, pivoting, de-pivoting, etc. These features are accessed through the Import Data option of the File menu.

**Data Source**

A database, a file, or the clipboard.

**Deselected Records**

The records that do not fulfill the constraints of the query devices.

**Details-on-Demand**

A window that shows the data in a marked record. The information can be displayed as plain text, HTML styled text in a window or as styled text in an external HTML browser.

**Docked Window**

By default, the Query Device window, the Details-on-Demand window, and the Legend window are docked; they stick to one of the outer borders of the Spotfire.net main window but can be moved.

**Empty Value**

A field which is undefined (has a null value).

**Field**

One of the values in a record. A single 'cell' in a spreadsheet.

**Full Search Query Device**

A query device with which any string of characters can be located in the column under investigation. Boolean logic can be used to refine the search.

**Glyph**

One of the shapes representing a data record in a 3D scatter plot.

**Histogram**

A visualization displaying the data as bars, the heights of the bars represent the number of records.

**HTML**

**Hyper Text Markup Language.** At its simplest, a language for formatting text. One of the primary reasons for the success of HTML and the World Wide Web is the ability to use links within or between documents.

**Item Slider**

A query device with which to select a specific value in a range.

**Jitter**

An option that displaces the visualization objects randomly in the display window, to make overlapping records visible.

**Lasso Function**

Enclosing a set of markers with the mouse pointer while pressing Shift on the keyboard. This brushes (marks) the records. The resulting encircling can have any shape.

**Marker**

One of the shapes representing a data record in a scatter plot. The 3D shapes are also called glyphs.

**Marked Record**

A marker in a scatter plot becomes marked when brushed, i.e. having been surrounded by the lasso or rectangle method (left mouse button pressed while moving pointer). Marked records are given a definable color to distinguish them from the rest of the data.

**Menu Bar**

The row of commands visible at the upper border of a Windows program. Clicking any of the menu titles will display sub menus with additional choices.

**Menu Item**

One of the options available in the menu presented by one of the commands in the menu bar.

**Null Value**

Empty. Cell with no value.

**ODBC**

**Open Data Base Connectivity**, a common method of accessing databases.

**Pie Chart**

A visualization showing data in aggregated “pie segments”, which can be split, drilled down into and taken apart.

**Portal Window**

A browser window integrated into the Spotfire.net environment, able to communicate with the Spotfire.net visualization engine.

**Profile Chart**

A visualization type where column names are plotted along the X-axis, and the value in each column on the Y-axis. This means that each record is presented as a line.

**Properties**

The dialog box enabling changes to query devices, visualization objects—markers, pies, bars, backgrounds and HTML code. Located in the Edit menu. Accessible with [Alt+Enter] **Edit > Properties** and a toolbar button.

**Query Device**

A tool that selects data having certain values for visualization. There are five kinds of query devices: range sliders, item sliders, full text searches, check boxes and radio buttons.

**Query Device Window**

The window where the query devices are located. By default, this window is placed in the upper right corner.

**Range Slider**

A Query Device used to select a range of data. The default device for numerical data. A range slider enables narrowing, broadening and moving the interval of the selected data of the designated column.

**Record**

One entry or line in a database or data set. Sometimes called a row.

**Scatter Plot**

A visualization type showing the records as markers in a two-dimensional or three-dimensional diagram.

**Selected Records**

The records explicitly chosen for display by manipulating the query devices and zoom bars. Do not confuse with the *Marked Records*, those that have been brushed (lassoed or rectangle enclosed).

**SFS**

The file extension Spotfire.net applies when saving data set in its proprietary format.

**SQL**

Structured Query Language, a common language used for formulating queries when investigating databases.



**Tight Coupling**

Spotfire.net visualizations, as well as query devices, are continuously updated to indicate chosen search intervals. For example, range slider intervals not applicable have a lighter shade of yellow than the valid ones.

**Toolbar**

The buttons located under the menu bar go by the collective name “toolbar”.

**Trellis plot**

A way of splitting a visualization into panes and organizing these in a grid layout.

**Visualization**

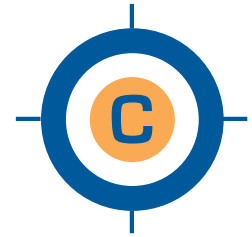
Visual representation of database records, e.g. scatter plot, pie chart, etc.

**Zoom Bar**

The bars along the edges of the visualization window. By manipulating them it is possible to zoom in on details of the visualization.



# Quick Reference



*The following list covers all items on the menu bar and its subsequent menus. The available functions/options are presented. Underlined letters indicate quick-key shortcuts [Alt+ key].*

## The Menus

| <u>F</u> ile                         |   |
|--------------------------------------|---|
| <u>O</u> pen   Ctrl+O                | Open an existing document. A browse dialog opens where you can search for files in various formats to open in Spotfire.net.   |
| <u>A</u> pply<br><u>T</u> emplate... | Apply settings from an SFT or SFS file. A dialog with check boxes opens allowing you to select which attributes to import from another SFS file. HTML templates for presenting Details-on-Demand will be imported from the external SFS file. |
| <u>S</u> ave   Ctrl+S                | Save the active document. If the document has never been saved before, this will open the same dialog as <b>Save As...</b>  |
| <u>S</u> ave <u>A</u> s...           | Save the document with another name, in another format, or in another location.   |
| <u>C</u> lose                        | Close the active document. You will be prompted for saving changes.   |

|  |  |
|--|--|
| <b>Import Data...</b><br><b>Ctrl+D</b> | Allows you to import data in any supported format (ODBC, text files, clipboard, etc), and to apply data conditioning.  |
| <b>Export &gt;</b>                     |  |
| <b>Details-on-Demand HTML</b>          | Export Details. The information on the object that is highlighted or the objects that have been marked is exported as text.  |
| <b>Current Visualization</b>           | Export the current visualization. An image of the currently active visualization is exported as a *.jpg or *.bmp image file. A Save As dialog is opened and you can navigate to where you want to save your image.                           |
| <b>SQL Query</b>                       | Exports the settings of the query devices as an SQL query.   |
| <b>Legend</b>                          | Export the legend. An image of the legend —the description of which columns are used for size, color, etc. is exported as a *.jpg or *.bmp image file. A Save As dialog is opened and you can navigate to where you want to save your image. |
| <b>Data</b>                            | Export data as Tab Separated Values. A Save As... dialog opens and you can save all the records in the data set as tab separated text.   |

|                                |  |
|--------------------------------|--|
| <b>Print... Ctrl+P</b>         | Print the active visualization. A Print dialog is opened where you can choose printer and printer settings.          |
| <b>Print Preview</b>           | Display full pages. Displays a preview of the visualization as it will be printed with the settings that will apply. |
| <b>Print Setup...</b>          | Change the printer and printing options.   |
| <b>Print Layout Options...</b> | Change the report layout.  |
| <b># Most recent files</b>     | The four most recently opened documents can be opened from the list that is displayed.                               |
| <b>Exit</b>                    | Quit the application. You are prompted to save open documents.   |

**Edit**

|   |  |
|---|--|
| <b><u>C</u>opy Visible</b><br><b>Ctrl+C</b>     | Copy visible records to the Clipboard. The data in the records represented by all the visible objects in the active visualization is copied to Windows Clipboard in text format. From there it can be pasted into an empty instance of Spotfire.net, or other application such as MS Excel or Notepad. |
| <b><u>C</u>opy <u>S</u>pecial</b>               |  |
| <b><u>S</u>electe<u>d</u> Records</b>           | Copy selected records to the Clipboard. Data in the records that fulfill all query device constraints is copied to the Clipboard in text format.   |
| <b><u>A</u>ll Records</b>                       | Copy all records to the Clipboard. Data in all the records in the data set is copied to the Clipboard in text format.  |
| <b><u>M</u>arked Records</b>                    | Copy set of marked records to the Clipboard. Data in all the records that have been marked —brushed— is copied to the Clipboard in text format.  |
| <b><u>C</u>urrent Visualization</b>             | Copy the current visualization to the Clipboard. An image in vector graphics format of the active visualization is copied to the Clipboard. This image can be pasted into another application as a Windows Metafile.   |
| <b><u>P</u>aste</b> <b>Ctrl+V</b>               | Insert Clipboard contents. The contents of the clipboard are pasted into a new Spotfire.net visualization. The Clipboard should contain data as text in one of the formats recognized by Spotfire.net.   |
| <b><u>R</u>eset</b> <b>Ctrl+R</b>               | Reset all query devices and zoom bars to show all records. Color, size, brushing and other properties are unaffected.  |
| <b>Mark <u>A</u>ll Records</b><br><b>Ctrl+A</b> | Mark all visible (selected) records. Records not selected by the current query device settings are not marked.   |
| <b><u>D</u>elete Record(s)</b><br><b>Del</b>    | Delete marked records. The marked (brushed) records are removed from the data set in Spotfire.net's memory. This operation cannot be undone but does not affect the original data unless that file is intentionally overwritten with the diminished data.  |

|                                   |                  |  |
|-----------------------------------|------------------|--|
| <b><u>N</u>ew Column</b>          |                  |  |
| <b>By <u>B</u>inning</b>          |                  | Create new column through binning. A new column and query device is made from an existing column by grouping the data according to algorithms found in subsequent dialog boxes.  |
| <b>From <u>E</u>xpression</b>     |                  | Create new column from an expression. A new column and query device is made from existing column(s) by performing mathematical operations on the records in the data. A series of dialog boxes guides you through the construction of the mathematical expression.   |
| <b>From <u>M</u>arked Records</b> |                  | Make a new column from marked records. A new column and checkboxes query device with two boxes is made from the data. The check boxes allow selection of the marked (brushed) data and/or the rest of the records.   |
| <b><u>M</u>arked Record(s)</b>    |                  |  |
| <b><u>D</u>elete</b>              | <b>Del</b>       | Delete the set of marked records. The marked (brushed) records are removed from the data set in Spotfire.net's memory. This operation cannot be undone but does not affect the original data unless that file is intentionally overwritten with the diminished data. |
| <b><u>R</u>eset</b>               | <b>Ctrl+E</b>    | Reset the set of marked records. Un-marks all marked records.  |
| <b><u>I</u>nv<u>e</u>rt</b>       | <b>Ctrl+I</b>    | Invert the set of marked records. Marks all unmarked records and un-marks the set of marked records.   |
| <b><u>C</u>opy</b>                |                  | Copy the set of marked records. All data in the records whose objects are marked is copied in text form to the Windows Clipboard.  |
| <b><u>S</u>elect</b>              | <b>Ctrl+M</b>    | Sets the selection to the marked records. Creates a temporary filter so that only the marked records are visible.  |
| <b>New Column...</b>              |                  | Make a new column from marked records. A new column and checkboxes query device with two boxes is made from the data. The check boxes allow selection of the marked (brushed) data and/or the rest of the records.   |
| <b><u>P</u>roperties...</b>       | <b>Alt+Enter</b> | Set document properties. The <b>Properties dialog</b> opens in which parameters governing the appearance of the visualization can be set, annotations made, etc.   |

## View

Checkmarks in front of the options indicate that they are active; selection toggles this status.

|  |  |
|--|--|
| <b><u>T</u>oolbars...</b>                          | Show or hide the toolbar.                                      |
| <b><u>S</u>tatus Bar</b>                           | Show or hide the status bar.                                   |
| <b><u>W</u>orkbook</b>                             | Show visualizations with tabs.                                 |
| <b>Details-on-Demand (<u>T</u>ext)</b>             | Show or hide text Details-on-Demand.                           |
| <b>Details-on-Demand (<u>H</u>TML)</b>             | Show or hide HTML Details-on-Demand.                           |
| <b>Details-on-Demand (<u>E</u>xternal Browser)</b> | Launches an external browser displaying the Details-on-Demand. |
| <b><u>Q</u>uery Devices</b>                        | Show or hide the query devices.                                |
| <b><u>L</u>egend <b>Ctrl+L</b></b>                 | Show or hide the Legend.                                       |
| <b>Navigation toolbar</b>                          | Show or hide the Navigation toolbar                            |
| <b>Portal</b>                                      | Show or hide the Portal window                                 |

## Tools

|                            |  |
|----------------------------|--|
| <b><u>C</u>ustomize...</b> | Customize menu and toolbars. A dialog box opens with four tabs on which you can choose appearances and customize toolbars, commands and menus. Here you can also choose to save personal settings.                               |
| <b><u>S</u>ampling...</b>  | Set the sampling rate. You are prompted to set the percent of records in the original data to load into Spotfire.net. Each record is randomly selected each time a data set is loaded.   |
| <b><u>P</u>lug-ins...</b>  | Manage Plug-ins. You can purchase additional software to augment the functionality of Spotfire.net; these are administered from the dialog boxes under this sub-menu. Each Plug-in will add one or more items to the tools menu. |
| <b>View <u>T</u>ip...</b>  | Invoke the View Tip Dialog. Thumbnail views of visualizations are presented and can be quickly analyzed.   |

## Window

|   |   |
|---|---|
| <b>C</b> lose Active Visualization<br>Ctrl+F4 | Close the active visualization.   |
| <b>N</b> ew <b>S</b> catter Plot<br>Ctrl+1    | Create a new scatter plot visualization. A two-dimensional visualization is created.  |
| <b>N</b> ew <b>3D</b> Scatter Plot<br>Ctrl+2  | Create a new scatter plot visualization. A three-dimensional visualization is created.  |
| <b>N</b> ew <b>H</b> istogram<br>Ctrl+3       | Create a new histogram visualization.   |
| <b>N</b> ew <b>B</b> ar Chart<br>Ctrl+4       | Create a new bar chart visualization.   |
| <b>N</b> ew <b>P</b> ie Chart<br>Ctrl+5       | Create a new pie chart visualization.   |
| <b>N</b> ew <b>L</b> ine Chart<br>Ctrl+6      | Create a new bar chart visualization.   |
| <b>N</b> ew <b>P</b> rofile Chart<br>Ctrl+7   | Create a new bar chart visualization.   |
| <b>A</b> uto Hide Axes Selectors              | Hide axes selectors. When hidden this gives greater room for the visualizations.  |
| <b>H</b> ide Window Frame                     | Hides window frame. With multiple visualizations displayed, greater room is available for the visualizations when frame and title bar are hidden. |
| <b>A</b> uto Tile                             | Auto tiles the visualizations. Arranges all the open visualizations to fill the visualization window.   |
| <b>C</b> ascade                               | Arranges windows so they overlap. Each window is staggered so that all title bars are accessible.   |
| <b>T</b> ile <b>H</b> orizontal               | Arranges visualizations as non overlapping horizontal tiles.  |
| <b>T</b> ile <b>V</b> ertical                 | Arranges visualizations as non overlapping vertical tiles.  |
| <b>(#)</b> Visualizations                     | Pressing the [Alt] key plus the number in front of the visualization title, activates that visualization.   |














## Help
















|   |   |
|---|---|
| <b><u>T</u>ip of the day</b>                  | View Tip of The Day. Helpful tips and shortcuts are presented.  |
| <b><u>H</u>elp Topics</b> <b>F1</b>           | List Help topics. Online Help is launched where you can search for topics.  |
| <b>What's <u>T</u>his?</b><br><b>Shift+F1</b> | Point to an object and access Online Help describing it.  |
| <b><u>e</u>Support</b>                        | Go to the eSupport Web Page. Launches web browser and connects to:<br><a href="http://www.spotfire.com/esupport">http://www.spotfire.com/esupport</a> . |
| <b><u>A</u>bout Spotfire.net</b>              | Display program information, version number and copyright. License information is also found here.  |

## The Toolbar

*The following table describes the toolbar and shortcuts that you can use to create, open, close, save, and manipulate your visualizations. Look up the action you want to take, and perform one of the alternatives shown.*

Most actions can be performed with hot-keys from the keyboard (Keyboard Shortcut), from the menu (Menu Command, or Mouse), or by clicking a toolbar button (Toolbar Button). The menus can be reached by pressing the Alt key in combination with the underlined letter key. Subsequent underlined letter keys activate the menu items.

| Action   | Keyboard Shortcut | Menu Command, or Mouse                 | Toolbar Button  |
|--|-------------------|--|---|
| Open an existing file or import a database                       | <b>Ctrl+O</b>     | <u>F</u> ile > <u>O</u> pen            |    |
| Save the present visualization in the same format as when opened | <b>Ctrl+S</b>     | <u>F</u> ile > <u>S</u> ave            |   |
| Print the active visualization                                   | <b>Ctrl+P</b>     | <u>F</u> ile > <u>P</u> rint           |  |
| Copy the visible records to the clipboard as text                | <b>Ctrl+C</b>     | <u>E</u> dit > <u>C</u> opy visible    |  |
| Paste the clipboard contents into Spotfire.net                   | <b>Ctrl+V</b>     | <u>E</u> dit > <u>P</u> aste           |  |
| Reset all query devices and zoom bars                            | <b>Ctrl+R</b>     | <u>E</u> dit > <u>R</u> eset           |  |
| Display Help contents  | <b>F1</b>         | <u>H</u> elp > <u>H</u> elp Topics     |   |
| Display context-sensitive help                                   | <b>Shift+F1</b>   | Click arrow on item you want help with |  |
| Create a new 2D scatter plot                                     | <b>Ctrl+1</b>     | <u>W</u> indow > New Scatter Plot      |  |
| Create a new 3D scatter plot                                     | <b>Ctrl+2</b>     | <u>W</u> indow > New 3D Scatter Plot   |  |
| Create a new histogram   | <b>Ctrl+3</b>     | <u>W</u> indow > New Histogram         |  |
| Create a new bar chart   | <b>Ctrl+4</b>     | <u>W</u> indow > New Bar Chart         |  |

|  |                  |   |   |
|--|------------------|---|---|
| Create a new pie chart   | <b>Ctrl+5</b>    | <u>W</u> indow > <u>N</u> ew<br><u>P</u> ie Chart                           |    |
| Create a new line chart  | <b>Ctrl+6</b>    | <u>W</u> indow > <u>N</u> ew<br><u>L</u> ine Chart                          |    |
| Create a new profile chart   | <b>Ctrl+7</b>    | <u>W</u> indow > <u>N</u> ew<br><u>P</u> rofile Chart                       |    |
| Open the <b>Properties</b> dialog.   | <b>Alt+Enter</b> | <u>E</u> dit > <u>P</u> roperties   |    |
| Display legend in separate window  | <b>Ctrl+L</b>    | <u>V</u> iew > <u>L</u> egend   |    |
| Export the active visualization as an image (*.bmp or *.jpg) to disk                 |                  | <u>F</u> ile > <u>E</u> xport ><br><u>C</u> urrent<br><u>V</u> isualization |    |
| Save the current device settings as an SQL query to disk                             |                  | <u>F</u> ile > <u>E</u> xport ><br><u>S</u> QL Query                        |    |
| Open browser and display values for the currently marked records                     |                  | <u>F</u> ile > <u>E</u> xport ><br><u>D</u> etails                          |    |
| Show Portal window.  |                  | <u>V</u> iew ><br>Spotfire.net  |    |
| Go to previous page in Portal Window.  |                  |   |    |
| Go to next page in Portal Window.  |                  |   |  |
| Stop loading page.   |                  |   |  |
| Refresh contents. This will also reset any Spotfire.net application that is running. |                  |   |  |
| Go to Spotfire.net start page  |                  |   |  |
| Print Portal Window contents.  |                  |   |  |

## Mouse Shortcuts

### Bar charts, histograms and pie charts

#### Mouse over

An outline highlights the object and corresponding objects in similar visualizations.

#### Left-click a bar, or pie

A persistent outline surrounds the object. Information about the bar or pie that was clicked displays in the Details-on-Demand window.

#### Left-drag

Marks the pies or bars that are inside the rectangle delineated by the diagonal movement of the pointer. If no bars are surrounded all objects will be unmarked.

#### Shift + left-drag

Enables marking of bars or pies with the lasso tool. Draw an arbitrary shape around the objects you want to mark. If none are surrounded all objects will be unmarked.

#### Ctrl + left-drag

Adds bars or pies to the set of marked objects using the rectangle 'rubber band box' tool.

#### Ctrl + Shift + left-drag

Adds bars or pies to the set of marked objects using a lasso function to the set of marked records.

#### Right-click in visualization

Zooms in on the bar chart or pie chart to reveal more detail.

#### Shift + right-click

Zooms out from the bar or pie chart.

#### Right-click in gray border outside visualization

Displays a pop-up menu.

## Scatter plots and line charts

### Mouse over marker

A thin ring highlights the marker and all markers containing the record in other visualizations.

### Left-click a marker

A persistent ring surrounds the object. Information about the record corresponding to the marker that was clicked displays in the Details-on-Demand window.

### Left-drag

Marks the records that are inside the 'rubber band box' delineated by the pointer's diagonal movement. These markers will be given another color. A table of details will be displayed in the Details-on-demand window. If no markers are surrounded all previously marked objects will be unmarked.

### Shift + left-drag

Enables brushing of markers with the lasso tool. Draw an arbitrary shape around the markers you want to mark. If none are surrounded all objects will be unmarked.

### Ctrl + left-drag

Adds markers to the set of marked objects using the rectangle 'rubber band box' tool.

### Ctrl + Shift + left-drag

Adds markers to the set of marked objects using a lasso function to the set of marked records.

### Right-click

Zooms in on the scatter plot. Repeated clicks or holding the right button down will zoom in until there is only one marker left in the visualization.

### Shift + right-click

Zooms out.

### Right-click in gray border outside scatter plot

Displays a pop-up menu.

## 3D scatter plots

### Mouse over marker

An outline highlights the marker and, all corresponding markers in other scatter plots, both 2D and 3D, will be highlighted with a thin ring.

### Left mouse button

Same as 2D scatter plots.

### Right-drag

Rotates the 3D plot.

### Ctrl + right-drag

Zooms in on—or out from—the highlighted marker: the one surrounded by a persistent ring. The marker retains its position and size on screen, the scale and surroundings change. If no marker is highlighted the scales of the axes will be enlarged and greater detail of the central volume of the visualization will be revealed.

### Shift + Ctrl + right-drag

Rotates the visualization in the plane of the screen.

### Right-click in gray border outside scatter plot

Displays a pop-up menu .

## Profile charts

### Mouse over profile

A thin contour highlights the profile (record), as well as all markers associated with the record in other visualizations.

### Left-click a profile

A persistent contour surrounds the profile. Information about the profile appears in the Details-on-Demand window.

### Left-drag

Marks the profiles that are fully or partially inside the 'rubber band box' delineated by the pointer's diagonal movement. These profiles will become thicker. A table of details will be displayed in the Details-on-demand window. If no profiles are surrounded all previously marked objects will be unmarked.

### Ctrl + left-drag

Adds profiles to the marked set using the rectangle 'rubber band box' tool.

### Shift + left-drag

Enables marking of records with the lasso tool. Draw an arbitrary shape around the profile you want to mark. If none are surrounded all profiles will be unmarked.

### Ctrl + Shift + left-drag

Adds profiles, by using a lasso function, to the set of marked records.

### Right-click

Zooms in on the profile chart. Repeated clicks or holding the right button down will zoom in until there is only one profile left in the visualization.

### Shift + right-click

Zooms out.

### Right-click in gray border outside scatter plot

Displays a pop-up menu.

## Right-click a Query device

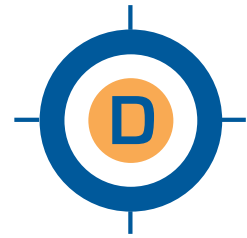
|                                       |   |
|---------------------------------------|---|
| <b><u>C</u>heck Boxes</b>             | If the column contains less than 500 unique values you can choose to have check boxes as query device. Values deselected by other query device constraints have red labels. |
| <b><u>R</u>adio <u>B</u>uttons</b>    | If the column contains less than 500 unique values you can choose to have radio buttons as query device. Only one of the values may be checked at a time.                   |
| <b><u>R</u>ange Slider</b>            | Range sliders can be used to select numeric data.   |
| <b><u>I</u>tem Slider</b>             | Item sliders can be used to select alphanumeric data.   |
| <b><u>F</u>ull <u>T</u>ext Search</b> | Full text search can search a string for any combination of characters using Boolean logic.   |
| <b><u>N</u>one</b>                    | Select none of the records with values in this column.  |
| <b><u>A</u>ll</b>                     | Select all records with values in this column.  |
| <b><u>S</u>how <u>E</u>mpy Values</b> | Display records even though values are empty (null).  |
| <b><u>S</u>how <u>F</u>ull Name</b>   | The query device is given another row with the entire name of the column.   |
| <b><u>M</u>ove</b>                    | Moves the query device.   |
| <b><u>T</u>o the <u>T</u>op</b>       | Moves the query device to the top of the window.  |
| <b><u>U</u>p</b>                      | Moves the query device up one position in the window.   |
| <b><u>D</u>own</b>                    | Moves the query device down one position in the window.   |
| <b><u>T</u>o the <u>B</u>ottom</b>    | Moves the query device to the bottom of the window.   |
| <b><u>H</u>ide</b>                    | Makes the query device invisible. Selecting the Show Query Device option on the Columns tab of the Properties dialog box makes the device visible again.                    |
| <b><u>N</u>ew <u>C</u>olumn</b>       |   |
| <b><u>B</u>y <u>B</u>inning</b>       | Create a new column by binning, a dialog box allows you to choose column to be binned and then a wizard guides you through the binning process.                             |



|                            |   |
|----------------------------|---|
| <b>From Expression</b>     | Create a new column from mathematical operations on data, a wizard guides through the construction of the expression. |
| <b>From Marked Records</b> | Create a new column. Each record is given a value indicating whether it is marked or not.                             |
| <b>Set Property</b>        |   |
| <b>Color</b>               | The colors of the markers, pies or bars are governed by the values of the records in the column.                      |
| <b>Size</b>                | The sizes of the markers, pies or bars are governed by the values of the records in the column.                       |
| <b>Original Scale</b>      | All changes to the scale of the column are removed.   |
| <b>Ln(1+x) Scale</b>       | Logarithm base e scale $[\ln(x+1)]$ .   |
| <b>Log Scale</b>           | Logarithm base 10 scale $[\log_{10}(x)]$ .  |
| <b>Exp Scale</b>           | Exponential scale base e $[e^x]$ .  |
| <b>Exp10 Scale</b>         | Exponential scale base 10 $[10^x]$ .  |
| <b>Inverse Scale</b>       | Inverse scale $[1/x]$ .   |
| <b>Reverse Sorting</b>     | Reverse scale   |
| <b>Edit Annotation...</b>  | Go to the dialog that lets you edit the annotation for this column.   |



# Text File Formatting



## Text file formats supported by Spotfire.net

The following standardized text file formats are supported:

- CSV – values are separated by the default list separator. (This is a Regional Setting property of your system. Comma is used in the US.)
- TXT – values are separated by tabs
- SKV – values are separated by semicolons

All these file formats are simple and text based—they don't contain any formatting or product specific information. The CSV and SKV formats can be exported from many spreadsheet and database programs. Many research tools produce output in these formats as well.

## General specifications

A data set imported into Spotfire.net is comprised of the following:

- One line containing the name of each column
- One declaration line describing the type of data in each column (optional)
- One line per data record
- [Carriage return] followed by [Line feed] end each row
- The first two single lines comprise the header. If the original data set has no header (i.e. it consists of raw data) Spotfire.net will name the columns Column1, Column2, ...

The following is an example of a well-formatted text file:

```
Name,Age,Address,Income
STRING,INT,STRING,INT
John,32,Göteborg,30000
Mary,28,Boston,90000
Steve,50,Stockholm,55000
Lisa,40,London,60000
```

## Column separators

Individual fields (column names, type strings, and values) are delimited by column separators—usually commas, semicolons or tabs, but the | (pipe) symbol is also recognized.

Spotfire.net automatically decides which character is interpreted as separator. To do this, Spotfire.net checks the first line in the database for the following characters:

| Separator | Name                                 | ASCII code |
|-----------|--------------------------------------|------------|
| →         | Horizontal tab (invisible on screen) | 9          |
| ,         | Comma                                | 44         |
| ;         | Semicolon                            | 59         |
|           | Pipe                                 | 124        |

The most common character of the ones above is selected as separator (characters enclosed in quotation marks are not considered). All following lines are assumed to use this most common character as column separator.



To load text data that uses a column separator different from those mentioned above, use the Import Data option. This will allow you to specify manually which character to interpret as separator. (See page 162.)



Values and column headers may contain any character (including white space) except the column separator. To include a separator (e.g. semi-colon), enclose the value or header with "double quotation marks".

## Name and type information

The first one or two lines in the Spotfire.net database may be interpreted as name and type information, according to the following rules.

1. If the second line appears to be type information, i.e. contains columns with strings like INT, STRING, DATE etc. it is used as type information. The first line is then expected to contain name information. For example, a small database with personal data and both name and type information looks like the table below. Here, the second line is interpreted as type information and the first as name information.

| Data        | Spotfire.net interpretation |
|-------------|-----------------------------|
| Name, Age   | name info                   |
| STRING, INT | type info                   |
| Peter, 26   | data                        |
| John, 35    | data                        |

2. Spotfire.net tries to find type information in lines one and two. If it cannot, line one is used as name information, and the program makes a qualified guess of the types contained, by analyzing the information in line two. The following is a small database with personal data, but no type information. Here, the first line is interpreted as name information, and the types are set to STRING and INT.

| Data      | Spotfire.net interpretation |
|-----------|-----------------------------|
| Name, Age | Type guess: STRING, STRING  |
| Peter, 26 | Type guess: STRING, INT     |
| John, 35  | Data                        |

3. If 1 and 2 fail, the program tries to guess the type by analyzing the information on line one. The column names are then automatically generated as "Column1", "Column2", etc. For example, for a small database with personal data with no type or name information, the

## D • TEXT FILE FORMATTING

type information is set to STRING and INT, and the names are set to Column1 and Column2:

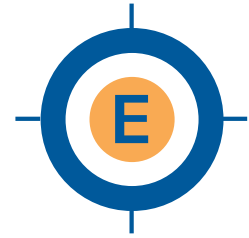
| Data     | Spotfire.net interpretation |
|----------|-----------------------------|
| Peter,26 | Type guess: STRING, INT     |
| John,35  | Data                        |



If a column is declared as being of one type, but some values in the column are of a different type, then the incongruous values will be set to null.

---

# Data Types



## General specifications

White spaces and ASCII control characters (i.e. all code characters less than 32) are always ignored at the start of a string. Type strings (strings that declare the data type) are not case sensitive.

In the descriptions of data types in this Appendix, the following abbreviations are used:

|             |  |
|-------------|--|
| HH          | Hour, in one or two digits   |
| MM          | Month, in one or two digits  |
| SS          | Seconds, in one or two digits  |
| Frac        | Fraction of a second   |
| DD          | Day of month, in one or two digits   |
| MM          | Month, in one or two digits  |
| YY          | Year, in two or four digits (e.g. <b>89</b> or <b>1989</b> )                                     |
| MonthString | Name of month spelled out, or abbreviated to three letters (e.g. <b>February</b> or <b>Feb</b> ) |

## E • DATA TYPES

If you are uncertain about whether or not Spotfire.net supports a data type, or how it must be formatted to be recognized during import, use a text editor to write a short file with sample data in the formats that you want to test. Copy and paste the text directly into Spotfire.net, and use Details-on-demand to verify the results.

The following tab-separated file can be used as a template for such a file:

| int_col | time_col    | date_col | str_col            |
|---------|-------------|----------|--------------------|
| int     | sfTime      | date     | string             |
| 1       | 11:01:32 PM | 990807   | "hello"            |
| 2       | 110132.6    | 08/07/99 | hello              |
| 3       | 11:01 AM    | 99-08-07 | ""Hello!", I said" |



## Time

### Format

HH:MM[:SS[.Frac]][ AM|PM]

### Type strings

time, sfTime

### Remarks

- The format requires that hour and minute be specified. Optionally seconds and fraction of second can be included. The fraction must be in the range of real values, and must begin with a point. Note, however, that Spotfire.net currently cannot handle the fraction. This means that the **decimal part is ignored** (truncated).
- In addition to a colon separator (as above), it is also possible to use a point or no character at all.
- AM or PM can optionally be specified at the end of a time string.
- Time must be within legal values. For example, **13:00 AM** is not supported. Negative times are not supported.

### Examples

| Accepted                             | Not accepted |
|--------------------------------------|--------------|
| 14:05:32.6 (interpreted as 14:05:32) | 14:05:32:6   |
| 14.05                                | 25.05        |
| 140532                               | 140532,6     |
| 2.05.32 PM                           | 14.05.32 PM  |
| 2.05.32 AM                           | 14.05.32 AM  |
| 2.5.2                                |              |

## Date

### Formats

MM/DD/YY (priority format, see below)

MonthString DD, YYYY

DD/MM/YY

DD.MM.YY

DD MonthString YY

DD-MonthString-YY

YY/MM/DD

YYMMDD

YY-MM-DD

### Type strings

date, sfDate

### Remarks

- The year, YY, can always be specified in two or four digits, e.g. as **89** or **1989**. If only two digits are given for the year, the date is assumed to belong to the 20th century if the number representing the years is higher than the dividing year, and to the 21st century if equal to the dividing year or below. Microsoft Window's settings provide the dividing year. If alterations have been made there, they will be reflected in the way Spotfire.net calculates dates. Windows default dividing year is 31.
- The priority status of the MM/DD/YY format means that a date given as 03/04/56 is understood to be the fourth of March 1956 and not the third of April 1956. If there is no conflict this is the format Spotfire.net assumes is in use.
- Months can be written using the initial three letters. If spelled out, names should be in U.S. English. Month names are not case sensitive.
- Spotfire.net performs a leap year control when importing data of type *Date*.
- Once dates are loaded into Spotfire.net Desktop they will be presented in localized format (according to your system's Regional Settings).

- Years may range from 00 to 9999.

### Examples

| Date       | Interpreted as    |
|------------|-------------------|
| 44-04-03   | April 3, 1944     |
| 25-07-21   | July 21, 2025     |
| 891224     | December 24, 1989 |
| 05/07/1973 | May 7, 1973       |
| 25/07/1973 | July 25, 1973     |
| 25 feb 68  | February 25, 1968 |

## DateTime

### Formats

|                 |                          |
|-----------------|--------------------------|
| MM/DD/YY        | HH:MM[:SS[.Frac]][AM PM] |
| MM/DD/YY        | HH.MM[:SS[.Frac]][AM PM] |
| DD/MM/YY        | HH:MM[:SS[.Frac]][AM PM] |
| DD.MM.YY        | HH:MM[:SS[.Frac]][AM PM] |
| DD MonthName YY | HH:MM[:SS[.Frac]][AM PM] |
| YY/MM/DD        | HH:MM[:SS[.Frac]][AM PM] |
| YY-MM-DD        | HH:MM[:SS[.Frac]][AM PM] |
| YY-MM-DD        | HH.MM[:SS[.Frac]][AM PM] |
| YYMMDD          | HH:MM[:SS[.Frac]][AM PM] |
| YYMMDD          | HHMM[SS[.Frac]][AM PM]   |

### Type strings

`datetime, timedate, timestamp, sfTimeStamp`

### Remarks

- Generally speaking, *DateTime* consists of a supported *Date* format, followed by a supported *Time* format. However, in order to maintain performance, only some of these combinations are allowed (see above).
- The *DateTime* format is used to combine date and time in a single data column. In previous documentation it has been referred to as *TimeStamp*.
- A value is accepted as *DateTime* even if it lacks the *Time* component.

**Examples**

| <b>DateTime</b>        | <b>Interpreted as</b>    |
|------------------------|--------------------------|
| 99-08-07 11:01:32.6    | August 7, 1999, 11:01:32 |
| 99-08-07 11.01.32      | August 7, 1999, 11:01:32 |
| 08/07/99 11:01         | August 7, 1999, 11:01:00 |
| 08/07/99 11.01.32.6 PM | August 7, 1999, 23:01:32 |
| 99/08/07 11:01         | August 7, 1999, 11:01:00 |
| 990807 11:01:32.6      | August 7, 1999, 11:01:32 |
| 990807 110132 AM       | August 7, 1999, 11:01:32 |
| 07 aug 99 11:01        | August 7, 1999, 11:01:00 |

## Integer

### Format

x.yyy.zzz[%]

### Type strings

int, integer, long, mapstring, sfInt

### Remarks

- An integer can include a grouping symbol for thousands, millions, etc. Permitted symbols are comma, dot, and white-space.
- Optionally, the %-character can be appended at the end.

### Examples

| Accepted |
|----------|
| 6843     |
| 6 483    |
| 6,483    |
| 6.483%   |

## Real

### Format

x.yyy,zz[%]

### Type strings

real, float, real1, real2, real3, real4, real5

### Remarks

4 byte floating point number. Written as an integer, optionally followed by a decimal part. The decimal separator can be a comma or dot, but must be different from the thousands separator.

### Examples

| Accepted  | Not accepted                       |
|-----------|------------------------------------|
| 6,843     |                                    |
| 6 843.27  | 6,843,27                           |
| 6.843,27% | 6.843.27%                          |
| 6.843,271 | 6.843.271 (interpreted as integer) |

## Double

### Format

(as Real above)

### Type strings

dbl, num, dreal, double, dfloat, sfReal

### Remarks

An 8 byte floating point number. If no type string is specified, the default type for any float will be Double.

## String

### Format

Any string of characters, with or without quotation marks

### Type strings

`string`, `text`, `str`, `sfString`

### Remarks

- Matching pairs of quotation marks at the beginning and end of the string are removed.
- Unicode files are currently not supported.
- Column separators appearing within outermost quotes are not interpreted as separators (see the last example below).
- Columns beginning or ending with empty noise (control characters except the separator currently being used) will be stripped of that noise.
- Strings are not case sensitive, e.g. `a12` is treated as identical to `A12`.

### Examples

| String            | Interpreted as                   |
|-------------------|----------------------------------|
| Hello             | Hello                            |
| "Hello"           | Hello                            |
| ""Hello""         | Hello                            |
| "Hello "there""   | Hello "there"                    |
| "3" 5"            | 3" 5'                            |
| one, two, three   | one two three (separate columns) |
| "one, two, three" | one, two, three (one column)     |
| ""Hello", I said" | "Hello", I said                  |



## Data Type Constraints

| <b>Int</b> i.e. 32 bits signed |             |
|--------------------------------|-------------|
| Maximum (signed) int value     | 2147483647  |
| Minimum (signed) int value     | -2147483648 |

| <b>Double</b> 64 bits   |                         |
|---|-------------------------|
| Number of decimal digits of precision   | 15                      |
| Smallest such that (1.0+DBL_EPSILON !=1.0)<br>The minimum difference between numbers that are to be calculated as having different values | 2.2204460492503131e-016 |
| Number of bits in mantissa  | 53                      |
| Maximum value   | 1.7976931348623158e+308 |
| Maximum decimal exponent  | 308                     |
| Maximum binary exponent   | 1024                    |
| Minimum positive value  | 2.2250738585072014e-308 |
| Minimum decimal exponent  | (-307)                  |
| Minimum binary exponent   | (-1021)                 |
| Exponent radix  | 2                       |
| Addition rounding: near   | 1                       |

| <b>Real</b> i.e. Float 32 bits  |                  |
|---|------------------|
| Number of decimal digits of precision   | 6                |
| Smallest such that (1.0+FLT_EPSILON !=1.0)<br>The minimum difference between numbers that are to be calculated as having different values | 1.192092896e-07F |
| Number of bits in mantissa  | 24               |
| Maximum value   | 3.402823466e+38F |
| Maximum decimal exponent  | 38               |
| Maximum binary exponent   | 128              |
| Minimum positive value  | 1.175494351e-38F |
| Minimum decimal exponent  | (-37)            |
| Minimum binary exponent   | (-125)           |
| Exponent radix  | 2                |
| Addition rounding: near   | 1                |



# References

## Literature on Exploratory Data Analysis

A fair number of interesting and helpful books exist on exploratory data analysis with a certain focus on visualization. Perhaps the most well known one is:

- “Exploratory Data Analysis” by John W. Tukey (Addison-Wesley, Reading M.A, 1977)

Two very good books, which indicate a lot of interesting relations between statistics and visualization, are:

- “Visualizing Data” by William Cleveland, (Hobart Press, Summit N.J, 1993)
- “The Elements of Graphing Data” by William Cleveland, (Hobart Press, Summit N.J, 1994)

Three highly readable comprehensive books on information visualization dealing mostly with paper designs are:

- “The Visual Display of Quantitative Information” by Edward Tufte, (Graphics Press, 1983)
- “Envisioning Information” by Edward Tufte, (Graphics Press, 1990)
- “Visual Explanations” by Edward Tufte, (Graphics Press 1997)

And a new 686-page book containing 43 key papers and extensive commentaries:

- “Readings in Information Visualization: Using Vision to Think” by S. K. Card, J. Mackinlay and B Shneiderman, (Morgan Kaufman Publishers 1999)

For obvious reasons these books do not deal specifically with the Spotfire.net system. However, they are very helpful in supplying information on how to explore data collections visually.



# Index

## 2

2D Scatter plots, 55

## 3

3D

- Dimensions, 63
- Navigation, 61
- Perspective, 64
- Rendering, 63

## A

Activating markers, 113  
Advanced options, 109  
Annotations, 93  
Anomalies, 21  
Appearance, setting for line charts, 78  
Application files, 188  
Arrow keys

- Axis selection, 34
- Fine tuning slider, 38
- Moving highlight ring, 113

Auto Tile, 100  
Average label

- Bar charts, Histograms, 67

Axes

- Change column, 31
- Scale transformation, 50

Axis properties, 95

## B

Background

- Image, 97, 149
- Map, 149

Text, 97

Bar charts, 66

- Bar size, 68

Bar types, 68  
Binning, 135  
Browsing, using Portal window, 117  
Brushing. *See* Marking records

## C

Calculating columns, 141  
Cascade, 100  
Categorical coloring, 56  
Check Boxes, 25, 44  
Checklist, new data, 35  
Cherry Picking. *See* Binning  
Clipboard, 162  
Clock speed, required, 15  
Clusters, 22  
Coloring

- Categorical, 56
- Continuous, 56
- Customize, 57
- Fixed, 56
- Histograms, Bar Charts, 67
- Pie charts, 70
- Quick access to, 51
- Scatter plots, 56

Column

- New, from expression, 141
- Separators, 216

Columns tab, 46  
Commands, 199  
Comments, 93  
Conditions, in Trellis plots, 87  
Control panel. *See* Query devices  
Copying records, 115

- Correlation, 131
- CSV, 215
- Customizing
  - Colors, 57
  - Toolbars, 107
  - Visualizations, 106

**D**

- Data and Background tab, 95
- Data conditioning
  - Applying, 165
  - Tall/Skinny to Short/Wide, 166
- Data Range, 95
- Data set
  - Sampling, 174
  - Size limitations, 173
- Data source, 161
- Data type constraints, 229
- Data types, 217, 219
- Data, exporting, 182
- Database, 22
  - Importing from, 163
- date, data type, 222
- dateTime, data type, 224
- dbl, data type, 227
- Details-on-Demand, 101, 121
  - Export, 178
  - HTML, 123
- dfloat, data type, 227
- Discussion forum, 118
- Divide by Substring. *See* Binning
- double, data type, 227
- dreal, data type, 227
- Dynamic queries, 23

**E**

- Edit menu
  - Copy, 115
  - Marked records, 114
- Empty panels, 91
- Even distribution. *See* Binning
- Even Intervals. *See* Binning
- Exponential axis scale, 50

- Export
  - Current Visualization, 179
  - Data, 182
  - Details-on-Demand, 178
  - Legend, 181
  - Queries, 180
- Expression
  - Manual entry, 146
  - New column from, 141

**F**

- Favorites, 164, 169
- File menu
  - Apply template, 126
  - Export, 177
  - Save As..., 184
- File types, 183, 215
- float, data type, 227
- Full text search, 41, 189
  - Regular expressions, 43
- Functions, 145

**G**

- Glossary, 193
- Graphics card, 15
- Grid lines, 95

**H**

- Hard disk space, required, 15
- Hardware requirements, 15
- Heading colors, 92
- Help, 6, 205
- Hide
  - Axis Selector, 100
  - Window Frame, 100
- Highlighting markers, 113
- Histograms, 66
  - Bar Size, 68
- HTML, 123
  - Access to templates, 97
  - Details-on-Demand, 122
  - Statistics, 129
  - Templates, 124

**I**

Images  
 Background, 149  
 Exporting legend, 181  
 Exporting visualization, 179

Importing data, 159  
 From clipboard, 162  
 From database, 163  
 From file, 162  
 OLE DB, 171  
 Using a favorite, 164

Initial query devices, 45

Installation, 15

int, data type, 226

Integer, data type, 226

Interactive reports, 183

Inverse axis scale, 50

Item slider, 39

**J**

Jittering, 59

**L**

Labels, 60, 95  
 Histograms, Bar charts, 67  
 Pie charts, 72  
 Placement, density, 96

Lasso, 112

Layout, 101  
 Line charts, 80

Legend, saving, 181

Limits of bins. *See* Binning

Line charts, 73  
 Appearance, 78  
 Layout, 80

Line connection, 60

Lines, in profile charts, 84

Linking. *See* Highlighting markers

Literature, 231

Loading, 30

Logarithmic axis scale, 50

long, data type, 226

**M**

MapInfo, 152

Maps, 149

mapstring, data type, 226

Markers  
 Jittering, 59  
 Labels, 60  
 Rotation, 58  
 Shape, 2D, 58  
 Size, 58

Marking records, 111

Menus, 199

Meter, 116

Mouse, 208

Multiple visualizations, 99

**N**

Navigation, 65

Navigation toolbar, 117

New Column, 141

New features, 12

num, data type, 227

**O**

ODBC, 163

OLE DB, 171

Online help, 6, 205

Operating system, 16

Option buttons. *See* Radio buttons

Outliers, 21

**P**

Panels, empty, 91

Panning, 55

Pasting an image, 152

Pearson's correlation, 131

Periodic table, 29

Perspective, 64

Pictures, 149

Pie Charts, 69

Plain text, 122

Plug-ins, 118

## INDEX

- Portal window, 117
- Printing, 157
- Profile charts, 82
- Profile columns, 85
- Properties
  - Coloring, 56
  - Line connection, 60
  - Setting visualization, 32
  - Shape, 58
  - Size, 58
- Q**
- Query devices, 25, 33, 37
  - Change type, 46
  - Hiding, 49
  - Moving, 48, 52
  - Order, 48
- Quick access
  - Color & Size, 51
- R**
- Radio buttons, 45
- Random Access Memory, required, 15
- Range slider, 38
- Range, of axis, 95
- real, data type, 227
- Record meter, 116
- Regular expressions, 43, 189
- Related reading, 231
- Rendering, 63
- Reports, printed, 157
- Reposition, 100
- Reset, 201, 206
  - Axis scale, 50
- Reset navigation, 65
- Reshape, 165
- Reverse sorting, 50
- Rotation, 58
- S**
- Sampling, 174, 186
- Saving, 184
- Scale mode, 76
- Scale, transformation, 50
- Scatter plots
  - 2D, 55
  - 3D, 61
- Scenarios, 118
- Selecting columns, 74
- Separators. *See* Column; Separators
- Setup, 15
- SFA files, 188
- sfDate, data type, 222
- sfInt, data type, 226
- sfReal, data type, 227
- SFS files, saving, 184
- sfString, data type, 228
- SFT files, 187
- sfTime, data type, 221
- sfTimeStamp, data type, 224
- Shape, 58
- Shortcuts, 199
- Size, 58
  - Pie charts, 71
  - Quick access, 51
- SKV, 215
- Sliders
  - Item slider, 39
  - Range slider, 38
- Software requirements, 16
- Source, 161
- Spotfire.net
  - Downloading, 17
  - Installing, 17
- Spotfire.net applications, 119
- Spotfire.net content, 118
- SQL queries, 23
- SQL, exporting, 180
- Standard Deviation. *See* Binning
- Statistical measures, 98
- Statistics, 98
  - HTML, 129
- Status bar, 116
- str, data type, 228
- Straight line fit, 98
- string, data type, 228



Support, 119  
System requirements, 15

## T

Tab key, highlighting with, 113  
Tall/Skinny to Short/Wide, 166  
Technical support, 119  
Template files, 187  
Text file formats, 215  
Text search, 189  
text, data type, 228  
Tight coupling, 26  
Tile, 100  
time, data type, 221  
timedate, data type, 224  
timestamp, data type, 224  
Toolbars, 206  
    Customizing, 107  
    Navigation, 117  
Tools menu, sampling, 174  
Transformations, 165  
Trellis plots, 86  
Trends, 21  
TXT, 215  
Type declarations, 219

## U

Uninstalling, 17  
User interface, 10

## V

Value ranges, 229  
View menu, 101

Workbook, 105  
View Tip, 131  
Visual data mining, 19  
Visualizations  
    2D, 55  
    3D, 61  
    Arranging, 100  
    Creating, 100  
    Customizing, 106  
    Default settings, 106  
    Export, 179  
    Initial, 54  
    Line charts, 73  
    Overview, 53  
    Ranking, 131  
    Statistics, 98

## W

Window menu, 100  
Windows  
    Arranging, 101  
    Hiding, 101  
Workbook, 105

## X

XML, 187  
XSF files, 187

## Z

Zoom bars, 55  
Zooming  
    2D, 55  
    3D, 61