

High-D User Guide

Macrofocus GmbH

Version 2019.9.3

Table of Contents

Introduction	1
Getting started	2
User interface	3
Menu and toolbars	3
Status bar	8
Loading data	9
File-based data sources	9
Database connectivity	14
On-line data sources	15
Automatic default configuration	15
Data types	16
Axes panel	18
Visibility	19
Scale	19
Filter	19
Distribution	20
Axis reordering	20
Configuration panel	21
Color	21
Rendering	26
Legend	27
Parallel Coordinates view	28
Probing and selection	28
Filtering	28
Axes reordering	29
TablePlot view	30
Probing and selection	30
Distributions view	31
Probing and selection	31
Scatter Plot Matrix view	32
Parallel Coordinates Matrix view	33
Probing and selection	33
Scatter Plot view	34
Configuration	34
Zooming	34
Probing and selection	35
Multidimensional Scaling view	36
Computation	36

Zooming	36
Probing and selection	37
TreeMap view	38
Probing and selection	38
CartoPlot view	39
Zooming	39
Probing and selection	39
Saving settings, data, and graphics	40
Exporting graphics	40
Exporting data	41
Import/export of settings	41
Printing	41
Invoking High-D through the command line	42
Command line options	42
Bibliography	44

Introduction

Welcome to the documentation for Macrofocus High-D.

The "[Getting Started](#)" chapter provides a brief introduction to the most important features of the application and is a good place to start for new users.

All of the features are explained in detail in the subsequent chapters.

If you don't find what you are looking for in these pages, then you might want to look at our [FAQ](#) list, and of course you may also contact [High-D support](#).

Getting started

This chapter will introduce the core features of Macrofocus High-D and is intended to help new users get started with analyzing their data.

High-D is still lacking proper documentation. Meanwhile, we can suggest the following references:

- [Parallel coordinates Wikipedia article](#)
- [Multivariate Analysis Using Parallel Coordinates](#)
- [InfoViz.net, Parallel Coordinates article](#)
- [Better Know a Visualization: Parallel Coordinates](#)
- [Parallel Coordinates By Robert Kosara](#)
- [Alfred Inselberg's Homepage](#)

After you played around a bit, you will have found that it is very easy to quickly access specific values for specific objects. It is much faster than looking it up in a big table or issuing a database query. Just click on an object in any of the views and there you go. Individual values can not only be easily found, but they are also embedded in the overall context and you immediately see how they relate to other objects.

In addition to quick data access, the different views provide various ways of revealing interesting patterns in the data and allow you to make sense of it.

User interface

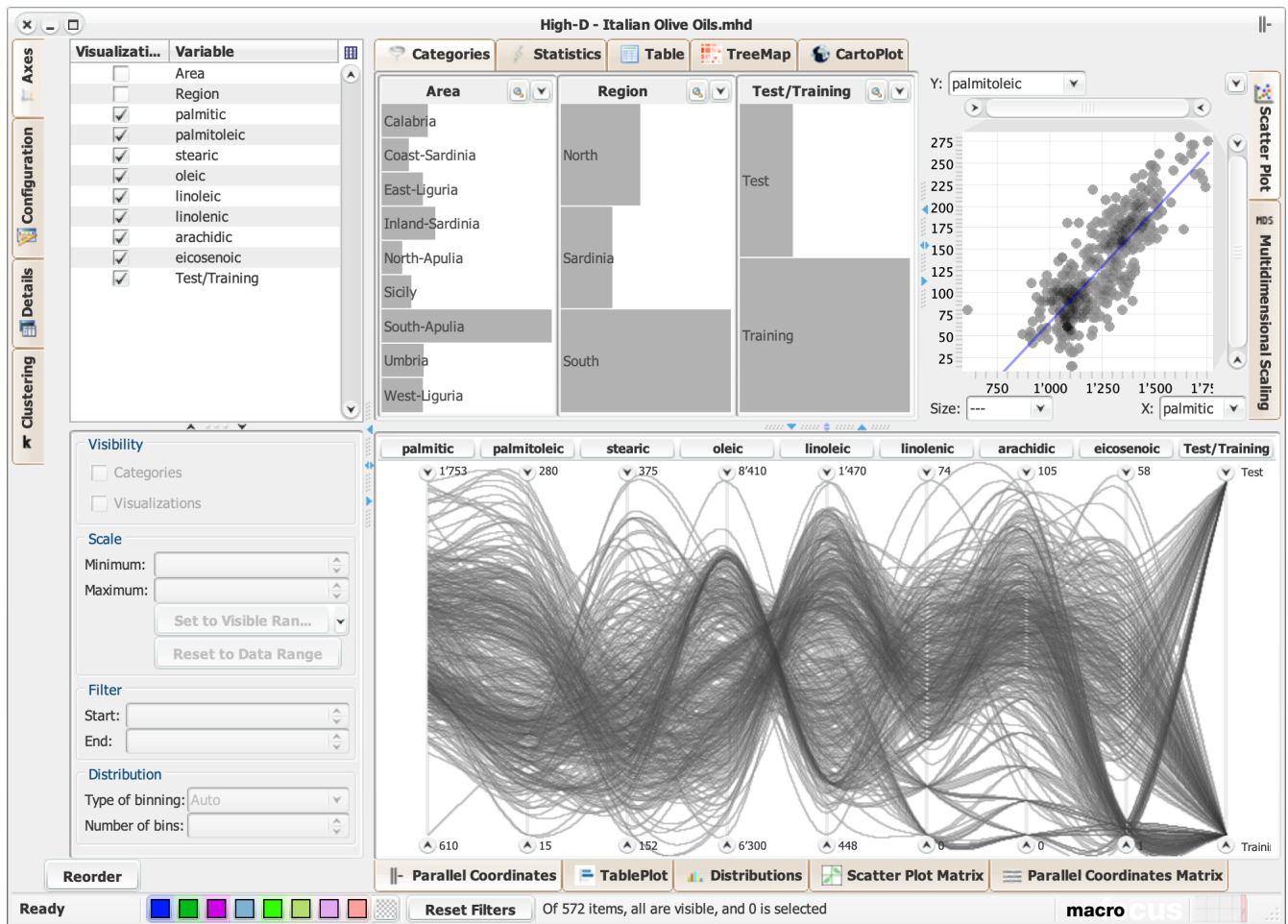


Figure 1. High-D user interface

Menu and toolbars

File menu

New

Creates a new empty window.

Open...

Load a data file in one of the [supported format](#).

Open URL...

Load a data file in one of the [supported format](#) from a remote location.

Open Database...

Load a database table or query from one of the [supported database](#) system.

Open Directory...

Create a dataset based on the directory structure.

Open Google Spreadsheet...

Load data from Google Spreadsheet.

Open Dataset

Load a dataset from High-D Server.

Open Recent

Load one of the previously opened dataset.

Reload

Reload the currently opened dataset, possibly retrieving updated data.

Save

Save the active window in native High-D format.

Save As...

Save the active window in native High-D format and give it a new file name.

Export Graphics...

The current view is exported in vector or raster form in one of the following supported formats:

PDF (Portable Document Format) (*.pdf)

The resulting document is ideal for printing or inclusion in a report. It is a vector format and therefore resolution independent.

Scalable Vector Graphics (*.svg)

The resulting document is ideal for further editing and for inclusion into another document. It is a vector format and therefore resolution independent. Scalable Vector Graphics (SVG) can be displayed by many web browsers with an embedded SVG viewer, or edited by any application supporting SVG (such as Adobe Illustrator).

Postscript (*.ps)

A common vector format and therefore resolution independent. Can be used for printing.

EMF (Enhanced Metafile) (*.emf)

A resolution independent format common on the Windows platform.

PNG (Portable Network Graphics) (*.png)

A raster format.

JPEG (*.jpg)

A raster format.

Compuserve GIF (*.gif)

A raster format.

TIFF (Tagged Image File Format) (*.tiff)

A raster format.



All the raster export format allow for setting the desired DPI for high-quality output.

Export Data...

The data visible in the Table view can be exported with **File › Export Data...** for further processing in spreadsheet programs or other applications. The following formats are supported:

CSV (Comma Delimited) (*.csv)

The comma-separated values (CSV) format stores tabular data (numbers and text) in plain-text form. Most spreadsheet and data management software are able to import data in this format.

Text (Tab Delimited) (*.txt;*.tsv;*.tab;*.raw)

The tab-separated values format is a popular method of data interchange among databases and spreadsheets. It stores tabular data (numbers and text) in plain-text form.

Microsoft Excel Workbook (*.xls;*.xlsx;*.xlsm)

The comma-separated values (CSV) format stores tabular data (numbers and text) in plain-text form. Most spreadsheet and data management software are able to import data in this format.

Apache Arrow (*.arrow)

The Arrow format stores tabular data (numbers and text) in form that allow data access without serialization overhead.

Apache Parquet (*.parquet)

The Parquet format stores tabular data (numbers and text) in a compressed, efficient columnar data representation. It is popular in the Hadoop ecosystem.

Import Settings...

All the settings can be exported using **File › Export Settings...**

Export Settings...

Settings saved using **File › Export Settings...** can be applied to another dataset using **File › Import Settings...**

Page Setup...

Setup the formatting of the page.

Print...

Print the current window.

Close

Close the current window.

Exit

Quit the High-D application.

Edit menu

Reset

Reset the views to their default.

Select menu

All

Select every non-filtered object.

Inverse

Inverse the selection.

None

Select nothing.

Filter menu

Selected

Filter out the selected object

None

Unfilter what has been previously filtered

Reset

Reset the filtering

Paint menu

Color

Paint the selected objects with the given color.

Reset

Reset the coloring

Interaction menu

Mode

Selection

Selection mode.

Filter

Filtering mode.

Toggle

Toggle selection mode.

DoNothing

Disabled interaction.

Options menu

Rendering

Density

Density-based drawing scheme.

AlphaBlended

Alpha-blended drawing scheme.

Opaque

Opaque drawing scheme.

Antialiasing

Turn antialiasing on or on

Show Filtered

Show filtered objects

Geometry

Polylines

Connect the points in the Parallel Coordinates view can be connected using polylines.

Steps

Connect the points in the Parallel Coordinates view can be connected using steps.

Polycurves

Connect the points in the Parallel Coordinates view can be connected using polycurves.

Look and Feel

Change the look and feel of the application

Create menu

Scatter Plot

Create an additional [Scatter Plot](#).

Control Chart

Create a Control Chart.

Window menu

Full Screen

Go into full-screen mode

Help menu

High-D Help

Read the High-D documentation

Check for Update...

Check for new version of the software

Register...

Register the license key

About High-D...

Obtains information about the current version of High-D

Status bar

Loading data

High-D offers the possibility of loading data in various formats and from multiple data sources. The most common ways of importing your own data is to use [tab-delimited](#) or [comma-separated files](#), as well as [Excel workbooks](#). Connectivity to common relational databases and some on-line data providers is also provided.

File-based data sources

To load data files, either

- use the **File > Open...** menu entry. This will open a dialog to select the file to open:

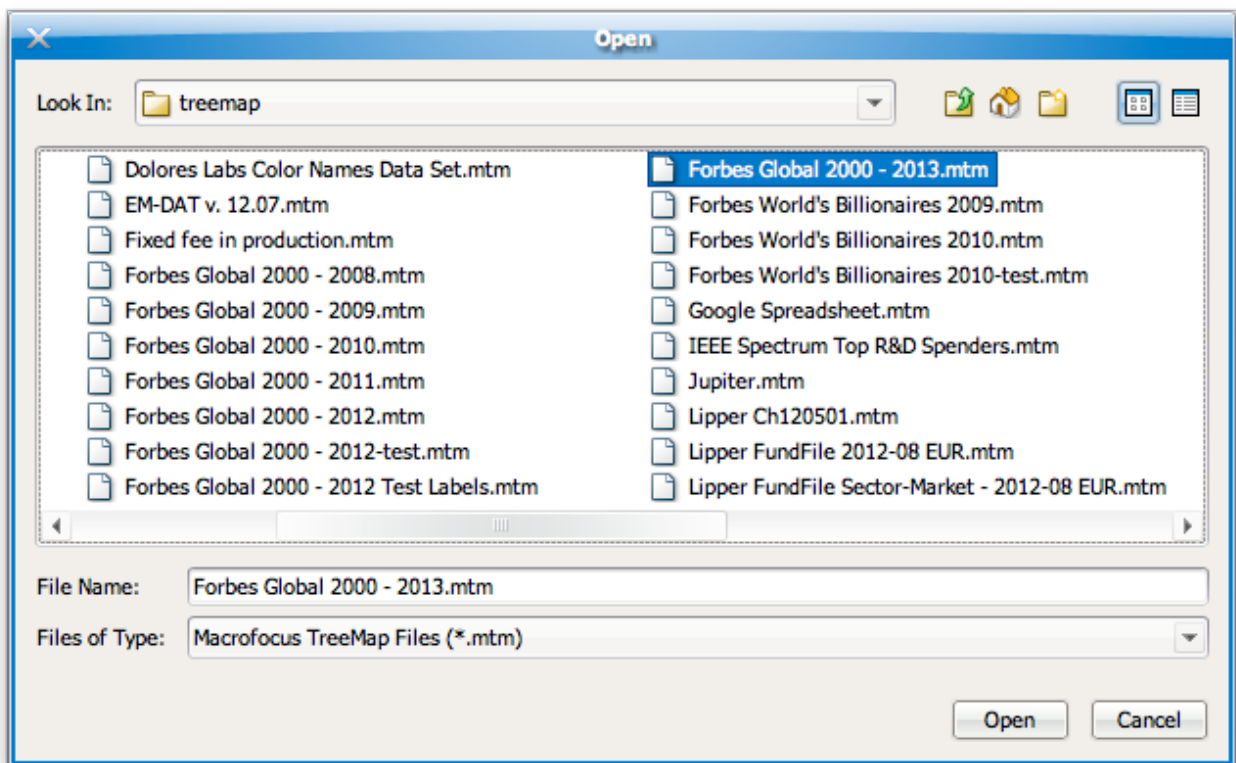


Figure 2. File chooser dialog for selecting a data file

- drag and drop a file with a known file extension onto the High-D application frame,
- or double-click on the file if its extension is registered to High-D.

Macrofocus High-D (*.mhd)

This is the native format used by High-D. It can be used to store both a copy of the actual data, its original data source, as well as all the configurations made using the High-D application. The data are stored in a highly compressed binary format to reduce the file size and all the configuration information in XML format. For a detailed technical specification of the data format, please [contact us](#).

Text (Tab delimited) (*.txt;*.tsv;*.tab;*.raw)

Loading data from tab-delimited text files should be pretty straightforward. High-D expects the first line to contain the name of each column, using the tab character to separate each column.

The tab-separated values format is a popular method of data interchange among databases and spreadsheets. It stores tabular data (numbers and text) in plain-text form. While it is a loosely defined format (even though IANA attempts to standardize it), High-D automatically detects its encoding, the type of data values, and handles smoothly all the most common causes of errors. Tab-delimited files are processed similarly to [comma-delimited files](#), except that they use the tabulator character to separate each column.

High-D expects the first line as a header to contain names corresponding to the columns in the file. These values will be used to name each of the variables. Each record is then located on a separate line. The values between each column are delimited by tabs. Each record "should" contain the same number of tab-separated fields. Any field may be quoted (with double quotes). Fields containing a line-break, double-quote, and/or tab should be quoted. A (double) quote character in a field must be represented by two (double) quote characters.

After the file has been loaded, High-D will attempt to detect the [data type](#) of each column. Automatically recognized types are text (String), numbers (Integer and Double) and some more specialized types such as dates (supported formats are "MM/dd/yyyy", "MM/dd/yy", "yyyy-MM-dd", "dd.MM.yyyy HH:mm:ss"), URLs, geometries (in WKT format), and binary data (in Base64 format).

As an example, the following text file

```
Planet Region Spherical area Radius in km Discovery date Wikipedia article
Mercury Inner Solar System 18688458.19 2439
http://en.wikipedia.org/wiki/Mercury_(planet)
Venus Inner Solar System 115066184.2 6052 http://en.wikipedia.org/wiki/Venus
Earth Inner Solar System 127796483.1 6378 http://en.wikipedia.org/wiki/Earth
Mars Inner Solar System 36274097.98 3398 http://en.wikipedia.org/wiki/Mars
Jupiter Outer Solar System 16014816458 71398
http://en.wikipedia.org/wiki/Jupiter
Saturn Outer Solar System 11309733553 60000
http://en.wikipedia.org/wiki/Saturn
Uranus Outer Solar System 2026829916 25400 3/13/1781
http://en.wikipedia.org/wiki/Uranus
Neptune Outer Solar System 1855079046 24300 9/23/1846
http://en.wikipedia.org/wiki/Neptune
Pluto Outer Solar System 7547676.35 1550 2/18/1930
http://en.wikipedia.org/wiki/Pluto
```

will result in the following table being loaded in High-D:

Planet	Region	Spherical area	Radius in km	Discovery date	Wikipedia article
String	String	Double	Integer	Date	URL

Mercury	Inner Solar System	18688458.19	2439		http://en.wikipedia.org/wiki/Mercury_(planet)
Venus	Inner Solar System	115066184.2	6052		http://en.wikipedia.org/wiki/Venus
Earth	Inner Solar System	127796483.1	6378		http://en.wikipedia.org/wiki/Earth
Mars	Inner Solar System	36274097.98	3398		http://en.wikipedia.org/wiki/Mars
Jupiter	Outer Solar System	16014816458	71398		http://en.wikipedia.org/wiki/Jupiter
Saturn	Outer Solar System	11309733553	60000		http://en.wikipedia.org/wiki/Saturn
Uranus	Outer Solar System	2026829916	25400	3/13/1781	http://en.wikipedia.org/wiki/Uranus
Neptune	Outer Solar System	1855079046	24300	9/23/1846	http://en.wikipedia.org/wiki/Neptune
Pluto	Outer Solar System	7547676.35	1550	2/18/1930	http://en.wikipedia.org/wiki/Pluto

While High-D will autodetect the character encoding used for representing international and special characters beyond ASCII characters, it is recommended to use the [Unicode](#) standards (typically UTF-8 or UTF-16).

To force High-D to parse values for a specific [data type](#), an optional second header line can be inserted. The second line can optionally contain information about the type of values to be expected for each column. Possible types are "String" for any type of textual information, "Integer" for numbers without a fractional or decimal component, "Float" and "Double" for single and double precision floating-point numbers, and "Color" to provide color information. Each subsequent lines should contain the respective values for each of the columns.

As an example, you can download the [Forbes Global 2000 dataset](#) in this format.

After the data file has been loaded into High-D, it will automatically attempt to create a [default configuration](#).

CSV (Comma delimited) (*.csv)

The comma-separated values (CSV) format stores tabular data (numbers and text) in plain-text form. Most spreadsheet and data management software are able to export data in this format. While it is a loosely defined format (even though [RFC 4180](#) attempts to standardize it), High-D automatically detects its encoding, the type of data values, and handles smoothly all the most common causes of errors. Comma-delimited files are processed similarly to [tab-delimited files](#), except that they use a comma (or semicolon) to separate each column.

High-D expects the first line as a header to contain names corresponding to the columns in the file. These values will be used to name each of the variables. Each record is then located on a separate line. The values between each column are delimited by commas (or semicolons). Each record "should" contain the same number of comma-separated fields. Any field may be quoted (with double quotes). Fields containing a line-break, double-quote, and/or commas should be quoted. A (double) quote character in a field must be represented by two (double) quote characters.

After the file has been loaded, High-D will attempt to detect the [data type](#) of each column. Automatically recognized types are text (String), numbers (Integer and Double) and some more specialized types such as dates (supported formats are "MM/dd/yyyy", "MM/dd/yy", "yyyy-MM-dd", "dd.MM.yyyy HH:mm:ss"), URLs, geometries (in WKT format), and binary data (in Base64 format).

As an example, the following text file

```
Planet,Region,Spherical area,Radius in km,Discovery date,Wikipedia article
Mercury,Inner Solar
System,18688458.19,2439,,http://en.wikipedia.org/wiki/Mercury_(planet)
Venus,Inner Solar System,115066184.2,6052,,http://en.wikipedia.org/wiki/Venus
Earth,Inner Solar System,127796483.1,6378,,http://en.wikipedia.org/wiki/Earth
Mars,Inner Solar System,36274097.98,3398,,http://en.wikipedia.org/wiki/Mars
Jupiter,Outer Solar System,16014816458,71398,,http://en.wikipedia.org/wiki/Jupiter
Saturn,Outer Solar System,11309733553,60000,,http://en.wikipedia.org/wiki/Saturn
Uranus,Outer Solar
System,2026829916,25400,3/13/1781,http://en.wikipedia.org/wiki/Uranus
Neptune,Outer Solar
System,1855079046,24300,9/23/1846,http://en.wikipedia.org/wiki/Neptune
Pluto,Outer Solar System,7547676.35,1550,2/18/1930,http://en.wikipedia.org/wiki/Pluto
```

will result in the following table being loaded in High-D:

Planet	Region	Spherical area	Radius in km	Discovery date	Wikipedia article
String	String	Double	Integer	Date	URL
Mercury	Inner Solar System	18688458.19	2439		http://en.wikipedia.org/wiki/Mercury_(planet)

Venus	Inner Solar System	115066184.2	6052		http://en.wikipedia.org/wiki/Venus
Earth	Inner Solar System	127796483.1	6378		http://en.wikipedia.org/wiki/Earth
Mars	Inner Solar System	36274097.98	3398		http://en.wikipedia.org/wiki/Mars
Jupiter	Outer Solar System	16014816458	71398		http://en.wikipedia.org/wiki/Jupiter
Saturn	Outer Solar System	11309733553	60000		http://en.wikipedia.org/wiki/Saturn
Uranus	Outer Solar System	2026829916	25400	3/13/1781	http://en.wikipedia.org/wiki/Uranus
Neptune	Outer Solar System	1855079046	24300	9/23/1846	http://en.wikipedia.org/wiki/Neptune
Pluto	Outer Solar System	7547676.35	1550	2/18/1930	http://en.wikipedia.org/wiki/Pluto

While High-D will autodetect the character encoding used for representing international and special characters beyond ASCII characters, it is recommended to use the [Unicode](#) standards (typically UTF-8 or UTF-16).

To force High-D to parse values for a specific [data type](#), an optional second header line can be inserted. The second line can optionally contain information about the type of values to be expected for each column. Possible types are "String" for any type of textual information, "Integer" for numbers without a fractional or decimal component, "Float" and "Double" for single and double precision floating-point numbers, and "Color" to provide coloring information. Each subsequent line should contain the respective values for each of the columns.

As an example, you can download the [Forbes Global 2000 dataset](#) in this format.

After the data file has been loaded into High-D, it will automatically attempt to create a [default configuration](#).

Microsoft Excel Workbook (*.xls;*.xlsx;*.xlsm)

High-D can read files produced by Microsoft Excel, including the recent Office Open XML format, even without having Excel installed on the local computer. The first row is expected to contain the name of each column. If the workbook contains multiple sheets, a dialog allows to choose which

one should be loaded by High-D.

To force High-D to parse values for a specific [data type](#), an optional second header line can be inserted. The second line can optionally contain information about the type of values to be expected for each column. Possible types are "String" for any type of textual information, "Integer" for numbers without a fractional or decimal component, "Float" and "Double" for single and double precision floating-point numbers, and "Color" to provide color information. Each subsequent line should contain the respective values for each of the columns.

As an example, you can download the [Forbes Global 2000 dataset](#) in this format.

ODF Spreadsheet (*.ods)

High-D can read files in the native OpenOffice and LibreOffice format.

SPSS (*.sav)

High-D can read files in the native SPSS format.

SAS (*.sas7bdat)

High-D can read files in the native SAS format.

ESRI Shapefile (*.shp)

This is a popular geospatial vector data format for geographic information systems (GIS) software. Shapefiles spatially describe features: points, lines, and polygons, representing, for example, water wells, rivers, and lakes. Each item usually has attributes that describe it, such as name or temperature.

Apache Arrow (*.arrow)

High-D can read files in the Apache Arrow format.

Apache Parquet (*.parquet)

High-D can read files in the Apache Parquet format.

Microsoft Access (*.mdb;*.accdb)

Access database tables can directly be loaded into High-D. However, this is only supported on the Windows platform and requires Microsoft Access or the Microsoft Access Database Engine to be installed.

Database connectivity

High-D can directly import data from popular relational database servers installed on the local computer or on a remote machine. Currently supported are:

- MySQL
- Oracle
- Microsoft SQL Server
- PostgreSQL
- IBM DB2
- SAP MaxDB
- PostGIS

Please [contact support](#) if your database system is not currently supported. Any data source queryable through a JDBC driver can easily be integrated into High-D.

[Microsoft Access](#) is also supported, but as a file-based data source.

To start importing data from a database, go to **File > Open Database....** This will open a dialog to define the required parameters:

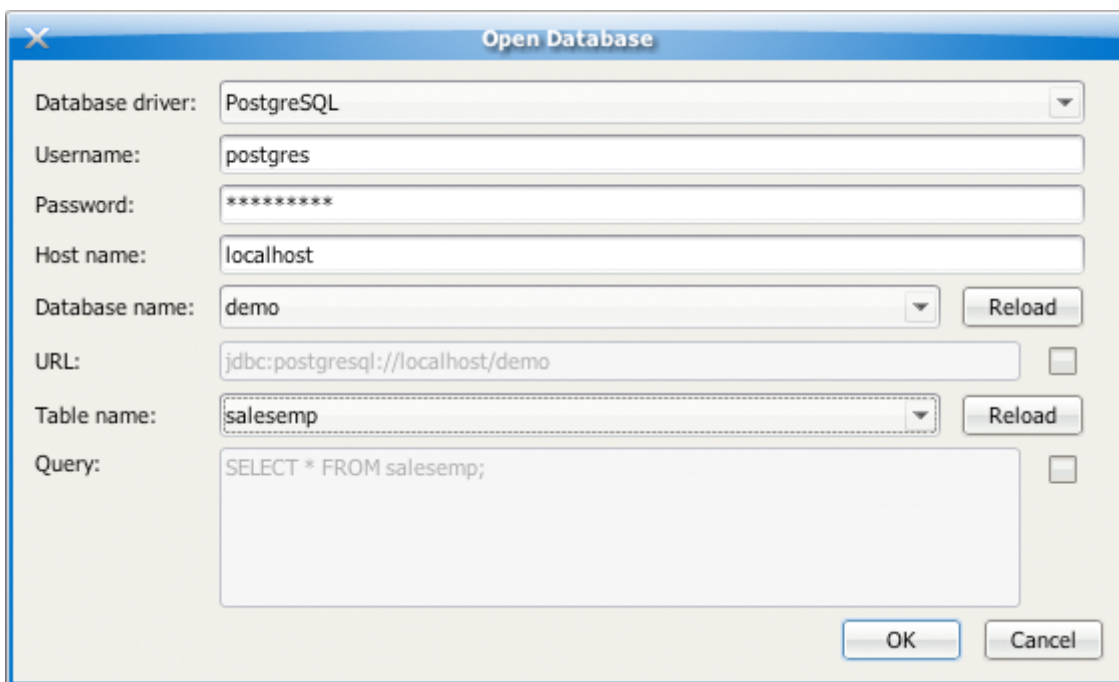


Figure 3. Database query dialog

On-line data sources

Stock quotes data from Yahoo Finance can directly be access through the **File > Open Dataset** submenu, as well as all the example datasets provided on our website. This menu entry also provides integration with High-D Server.

Automatic default configuration

By default, High-D automatically assigns the first categorical variable to the label, the second categorical variable (if available) to the grouping, the first numerical variable to the size, and the second numerical variable (if available) to the color.

Data types

All data types support null (blank) values. Supported types are:

Text

String

Represents character strings such as "abc".

StringPath

Represents an array of character strings. Values should be delimited by commas.

HtmlString

Represents a tagged string in HTML format.

Numbers

Byte

The Byte data type is an 8-bit signed two's complement integer. It has a minimum value of -128 and a maximum value of 127 (inclusive).

Short

The short data type is a 16-bit signed two's complement integer. It has a minimum value of -32,768 and a maximum value of 32,767 (inclusive).

Integer

The Integer data type is a 32-bit signed two's complement integer. It has a minimum value of -2,147,483,648 and a maximum value of 2,147,483,647 (inclusive). For integral values, this data type is generally the default choice unless there is a reason (like the above) to choose something else. This data type will most likely be large enough for the numbers your program will use, but if you need a wider range of values, use Long instead.

Long

The Long data type is a 64-bit signed two's complement integer. It has a minimum value of -9,223,372,036,854,775,808 and a maximum value of 9,223,372,036,854,775,807 (inclusive). Use this data type when you need a range of values wider than those provided by Integer.

Float

The Float data type is a single-precision 32-bit IEEE 754 floating point. Its range of values is beyond the scope of this discussion, but it can typically handle more than 7 decimal digits. This data type should never be used for precise values, such as currency. For that, you will need to use the BigDecimal type instead.

Double

The Double data type is a double-precision 64-bit IEEE 754 floating point. Its range of values is beyond the scope of this discussion, but it can typically handle more than 15 decimal digits. For decimal values, this data type is generally the default choice. As mentioned above, this data type should never be used for precise values, such as currency.

BigDecimal

An arbitrary-precision signed decimal number.

StringDouble

A Double data type with support for formatting patterns.

Others

Boolean

The Boolean data type has only two possible values: true and false. Use this data type for simple flags that track true/false conditions. This data type represents one bit of information.

Date

Represents a specific instant in time, with millisecond precision.

Color

The Color data type is used to encapsulate colors in the default sRGB color space. Every color has an implicit alpha value of 1.0 or an explicit one provided in the constructor. The alpha value defines the transparency of a color and can be represented by a float value in the range 0.0 - 1.0 or 0 - 255. An alpha value of 1.0 or 255 means that the color is completely opaque and an alpha value of 0 or 0.0 means that the color is completely transparent. When constructing a Color with an explicit alpha or getting the color/alpha components of a Color, the color components are never premultiplied by the alpha component.

Icon

A small fixed size picture, typically used to decorate components.

Image

Represents graphical images.

URL

The URL data type represents a Uniform Resource Locator, a pointer to a "resource" on the World Wide Web. A resource can be something as simple as a file or a directory, or it can be a reference to a more complicated object, such as a query to a database or to a search engine. More information on the types of URLs and their formats can be found in the [URL Specification](#).

File

A representation of file and directory pathnames.

byte[]

For binary data.

Geometry

Represents geometric information, such as points, lines, and polygons.

Axes panel

The Axes panel allows the customization of each of the available variables.

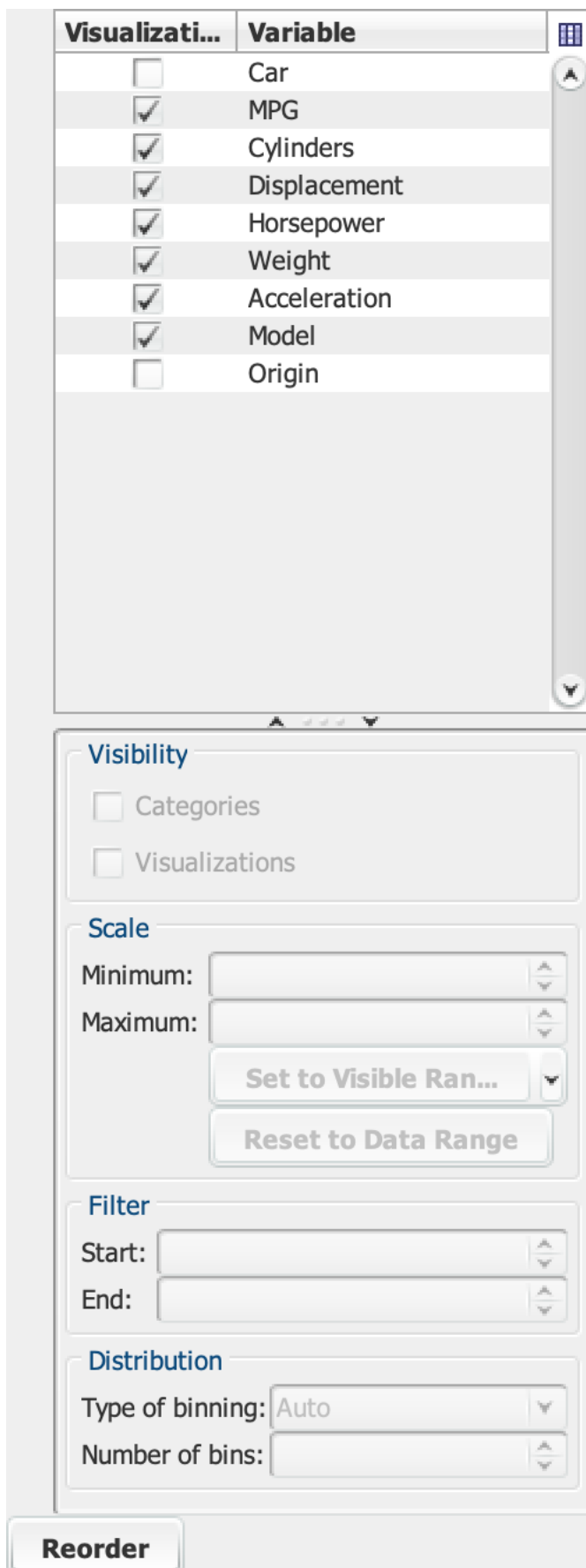


Figure 4. The Axes panel

In the upper part, all the variables included in the dataset are listed.

Selecting one variable can be performed by clicking on its row. Adding to or removing variables from the selection can be done by holding the **Ctrl** key down while selecting them. Multiple adjoining variables can be selected using the mouse while holding down the **Alt** key.

At the bottom part, can apply customization that will be applied to each of the selected variables:

Visibility

Categories

Whether to show the variables in the Categories view.

Visualizations

Whether to show the variable in the Parallel Coordinates, Table Plot, Distributions, Scatter Plot Matrix, Parallel Coordinates Matrix views.

Scale

Minimum

The lower end of the scale.

Maximum

The upper end of the scale.

Filter

Start

Items with a value below this threshold will be filtered out.

End

Items with a value above this threshold will be filtered out.

[Set to Visible Range]

Set to scale to the minimum and maximum values of the non-filtered items.

[Set to Range Slider]

Set the scale to the current range of the Parallel Coordinates sliders.

[Make Common Range]

Compute the overall minimum and maximum values of the selected variables.

[Make Symmetrical around Mean]

The center of the scale will be the mean value.

[Make Symmetrical Range around 0]

The center of the scale will be at 0.

[Round Range Values]

Will round the scale using power of 10.

[Reset to Data Range]

Will set the scale to the minimum and maximum values found in the data.

Distribution

In order to classify values in a certain amount of bins,



The following settings apply to the Distributions view only.

Type of binning

Auto will automatically attempt to find a good number of bins, or will use the number specified below. The size of each bin will be distributed evenly. On the other hand, with **Sigma** will have the size of each bin vary depending on the standard deviation. A value of 6 will yield the typical Six Sigma (6σ) split often found in process improvement analysis.

Number of bins

With a value higher than 0, then fix the number of bins, otherwise the number of bins will be determined empirically.

Axis reordering

You can reorder the selected axis based on their similarity. To do so, select a some variables and hit the [**Reorder**] button.



You can also manually reorder the axis by dragging them in the Parallel Coordinates view.

Configuration panel

High-D possesses a powerful layout, data processing, and rendering engine that offers a vast choice of customization possibilities. The configuration panel gives instant access to all the key settings, where each section can be further expanded to expose the full palette of choices to fine-tune the appearance of the various views.

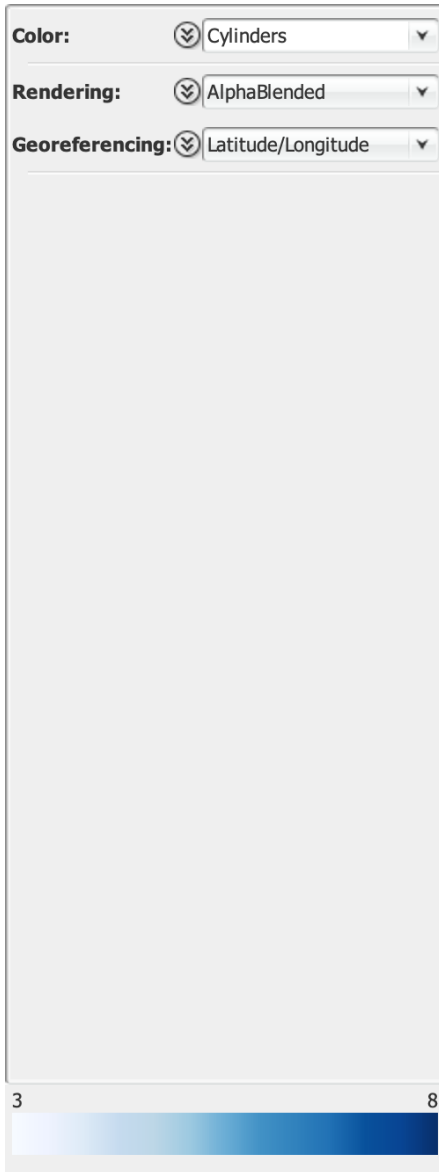


Figure 5. The configuration panel in its default unexpanded form

Color

The Color drop down list gives the possibility of selecting which variable should be used for coloring the shapes.

Import Colormap...

Import a colormap definition and apply it to the currently selected color variable.

Export Colormap...

Export the colormap definition of the currently selected color variable.

Copy Graphics

Copy the colormap to the clipboard.

Export Graphics...

Export the colormap to a raster or vector-based graphic format.

Print...

Print the colormap.

Categorical colormap

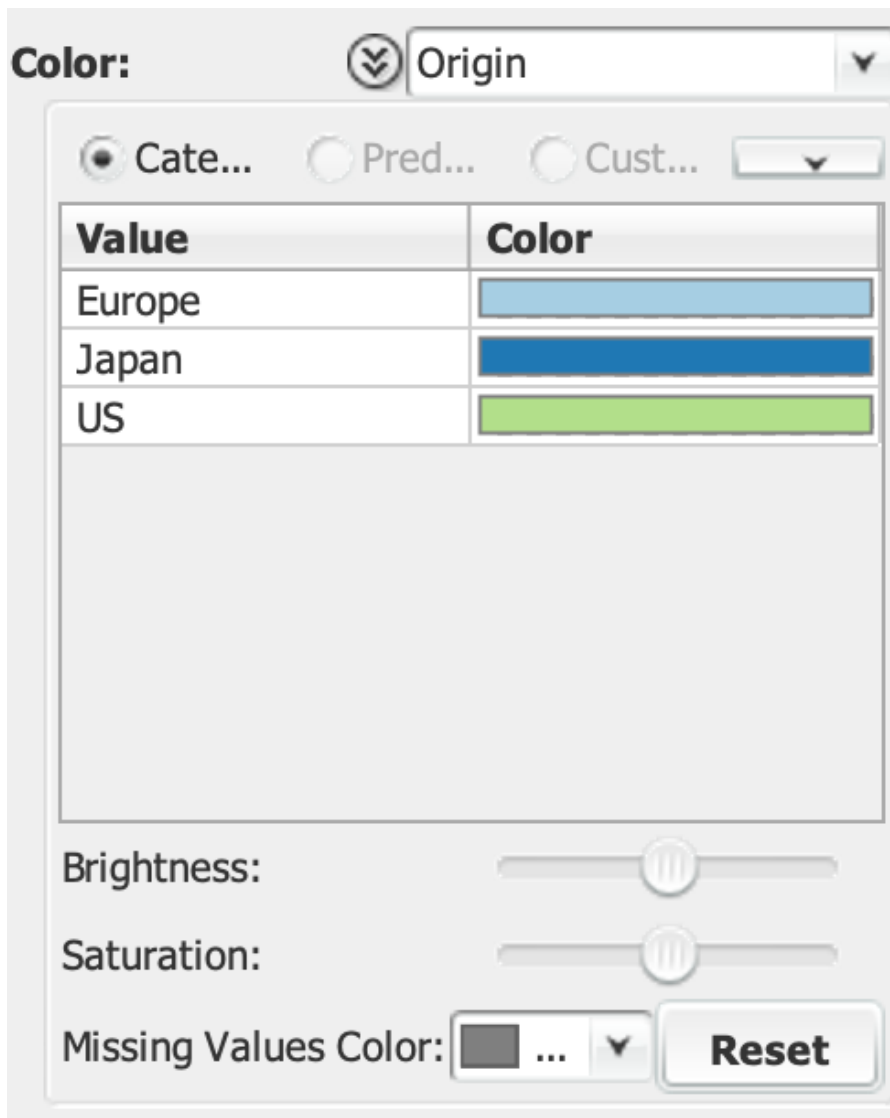


Figure 6. The color pane expanded with a categorical colormap

If a categorical variable is selected, then colors are automatically assigned to each of the value. Each color can be individually customized by clicking on the color itself. Each color can be individually changed by clicking on the color cell.

Missing Value Color

If the data contains missing values, then their color can be edited here.

Reset

Allows to reassign all the values to their default color.

Predefined colormap

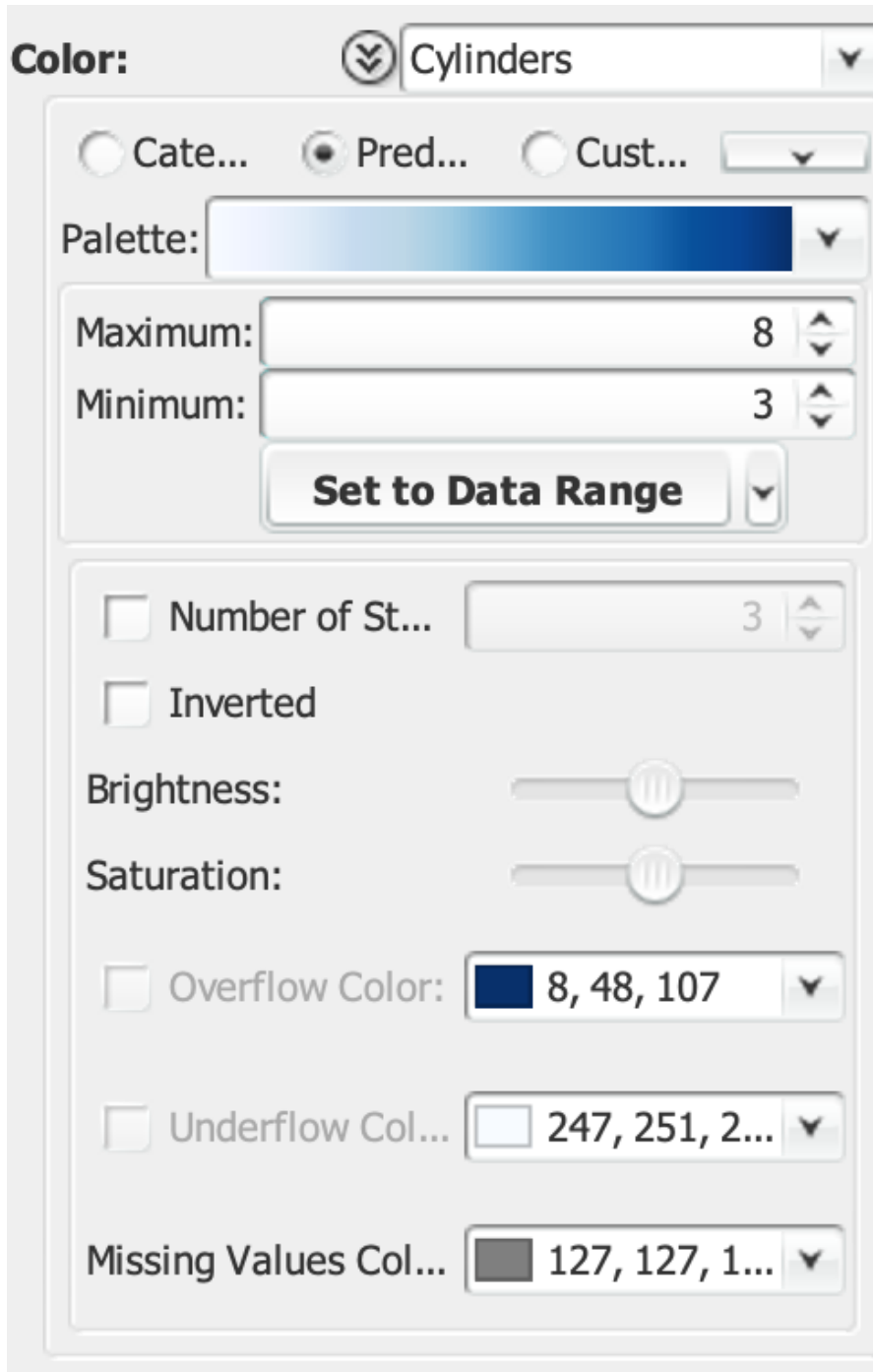


Figure 7. The color pane expanded with a predefined colormap

If a numerical variable is selected, High-D offers the possibility of setting the lowest and highest values that should be mapped to the selected colormap. If the variable contains negative values, the range is automatically made symmetric.

Palette

A color palette can be selected from a wide range of predefined color palettes

Maximum

Sets the upper bound of the colormap.

Minimum

Sets the lower bound of the colormap.

Set to Data Range

Set the minimum and maximum of the colormap to the minimum and maximum values of the data.

Set to Symmetrical Range around 0

Will make the colormap symmetrical should it contain negative values.

Set to Rounded Range

Will round the minimum and maximum values to their next power of 10 value.

Number of Steps

Can be used to segment the palette into a specified number of discrete colors.

Inverted

Invert the colormap.

Brightness

The color luminance can be adjusted by increasing or decreasing its brightness.

Saturation

The color intensity can be adjusted by increasing or decreasing its saturation.

Overflow Color

If some of the data values fall above the upper threshold, then their color can be edited here.

Underflow Color

If some of the data values fall below the lower threshold, then their color can be edited here.

Missing Value Color

If the data contains missing values, then their color can be edited here.

Custom colormap

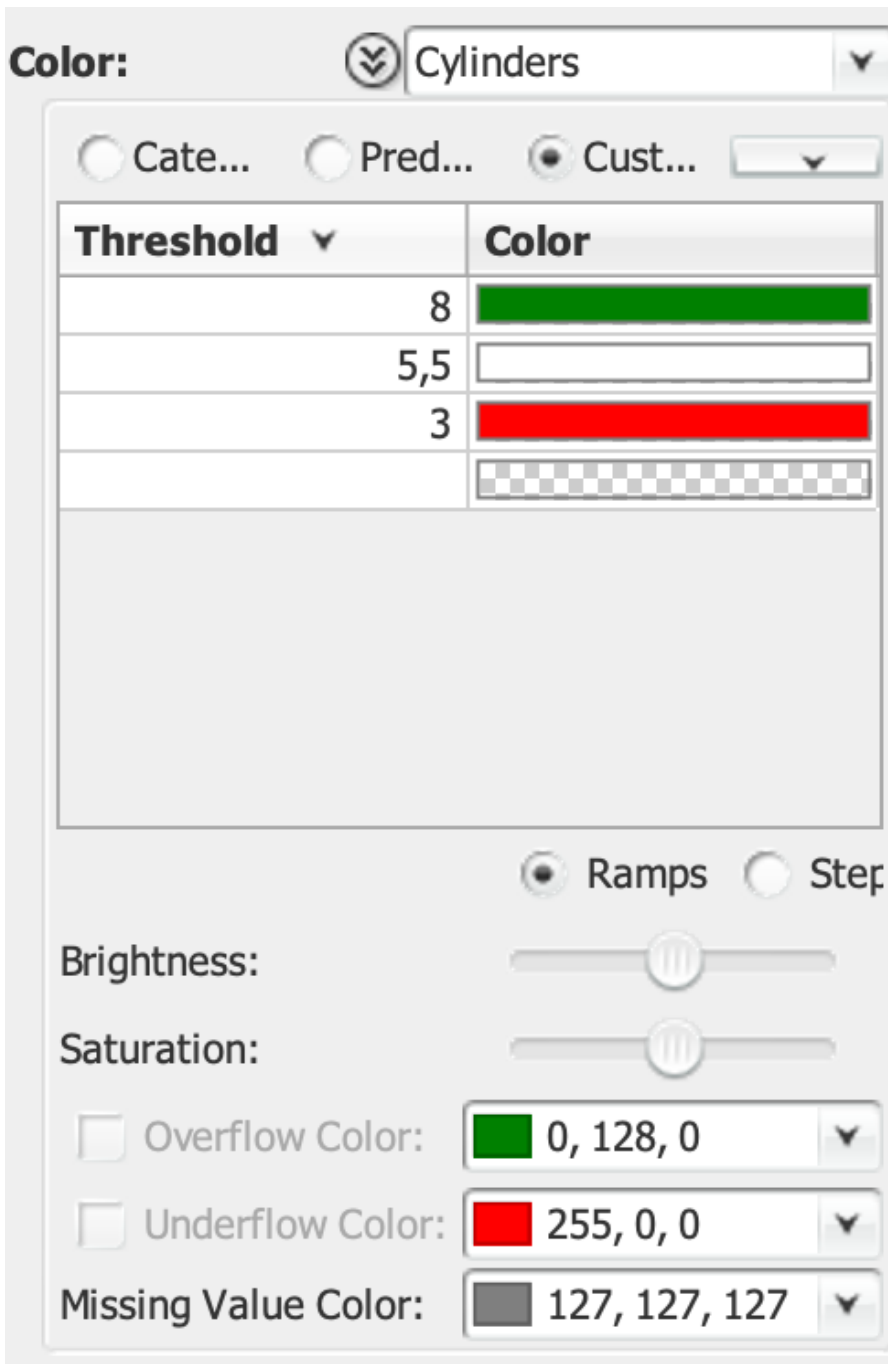


Figure 8. The color pane expanded with a custom colormap

For more customization possibilities, it is also possible to define a custom colormap by setting thresholds at given values. High-D will take care of interpolating the colors if Ramps mode is selected, or will make them valid for the whole range in Steps mode.

Threshold

Define the threshold for which values equal or above its value will be assigned the associated color. Removing a threshold can be accomplished by setting its color to None. New thresholds can be added by specifying a value in the last entry of the table. The color associated to the new threshold will be automatically extrapolated from the current colormap definition.

Color

The color associated with each threshold.

Ramps/Steps

Indicates whether the values should be interpolated within the threshold ranges (Ramps) or made discrete (Steps).

Brightness

The color luminance can be adjusted by increasing or decreasing its brightness.

Saturation

The color intensity can be adjusted by increasing or decreasing its saturation.

Overflow Color

If some of the data values fall above the upper threshold, then their color can be edited here.

Underflow Color

If some of the data values fall below the lower threshold, then their color can be edited here.

Missing Value Color

If the data contains missing values, then their color can be edited here.

Rendering

More options can be customized in the Rendering pane. Each visualization can be rendered using **AlphaBlended**, **Density**, or **Opaque** drawing schemes.

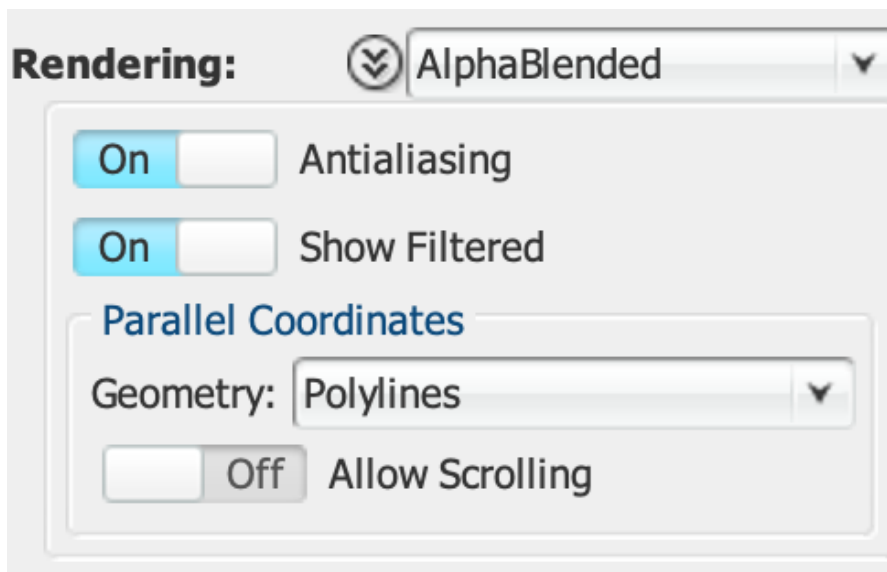


Figure 9. The expanded rendering pane

Antialiasing

Gives the possibility to disabled antialiased drawing.

Show Filtered

Allows to make filtered items visible.

Geometry

Points in the parallel coordinates plots can be connected using **Polylines**, **Steps**, or **Polycurves**.

Legend

A graphical depiction of the color scale as well as a textual description of the main options that have been selected. The legend can be exported accessing the context menu (by right clicking the mouse):

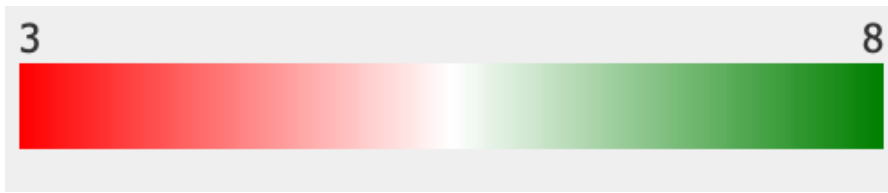


Figure 10. The legend

Copy Graphics

Copy the legend to the clipboard.

Export Graphics...

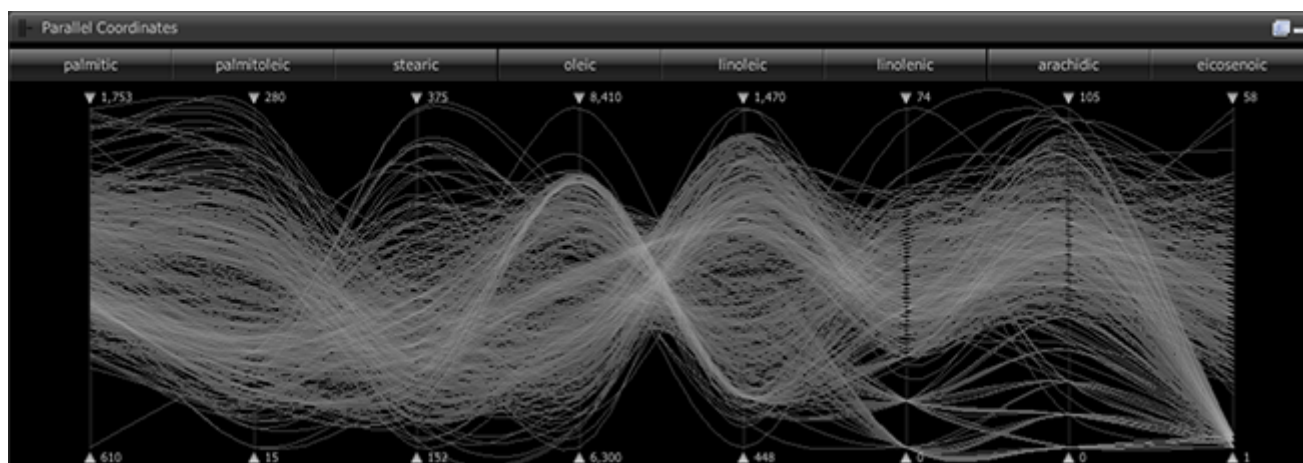
Export the legend to a raster or vector-based graphic format.

Print...

Print the legend.

Parallel Coordinates view

Parallel coordinates works by having vertical axis per data column and each row is displayed as a series of connected points along the axes. Using our innate pattern-recognition abilities, it enables spotting multivariate relations in a blink. Parallel coordinates has been popularised and systematically developed by Alfred Inselberg [Inselberg2009]. Thanks to the unique approach taken by High-D to use density-based rendering to avoid overplotting, and the choice between straight and curved geometries, relations and trends emerge immediately.



At the bottom of the user interface, you will find the Parallel Coordinates view that corresponds to the chosen settings in the Configuration and Axes panels. Each item is represented by a polyline or polycurve.

Moving the mouse over a shape will display a pop-up window (also called a tooltip) that shows the values configured in the "Labels" section of the "Configuration" panel.

Probing and selection

Selection can be performed by clicking on a shape. Adding to or removing items from the selection can be done by holding the `Ctrl` key down while clicking on shapes.

Multiple adjoining items can be selected using a rubberband which is activated by dragging the mouse while holding down the `Alt` key.

Filtering

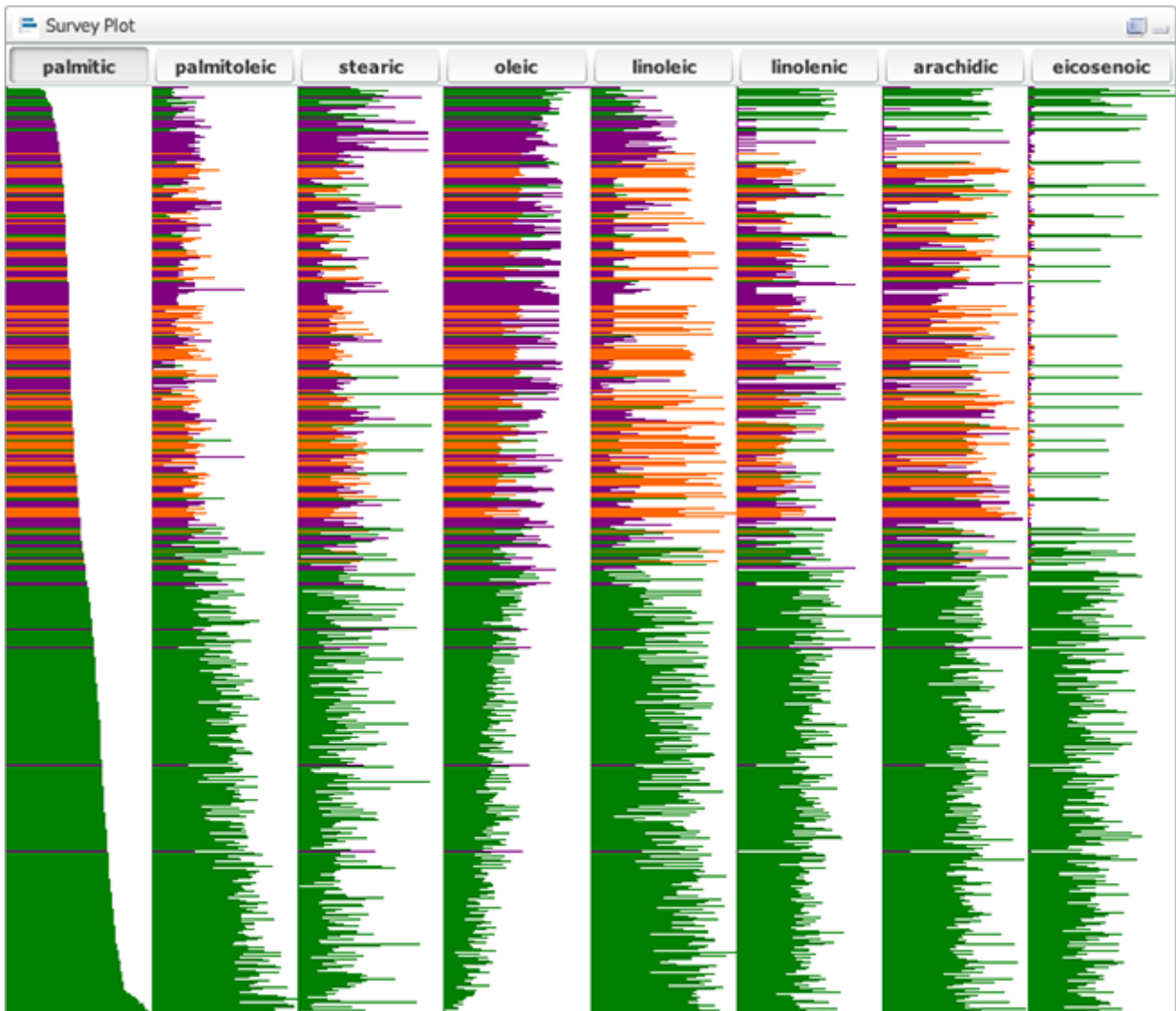
Items can be filtered by using the range sliders embedded in the Parallel Coordinates view. The range of an attribute can be specified by moving the handles on the top and bottom of the corresponding range slider. Items whose value for that attribute falls outside of the specified range, are filtered out and can not be interacted with anymore. Their "ghosts" remain visible though and they appear greyed-out. Use a combination of range sliders to dynamically formulate complex queries.

Axes reordering

An axis can be moved to a different position by dragging its label and dropping it at the desired position. Automatic reordering can be performed through the Axes panel.

TablePlot view

The TablePlot view works by having rows sorted by one variable to visually spot how the increase in value is correlate to the values of other variables. To obtain the complete picture, sorting through each variable is necessary.



At the bottom of the user interface, you will find the Table Lens view that corresponds to the chosen settings in the Configuration and Axes panels. For each axis, each item is represented by a line those width is proportional to its value.

Moving the mouse over a shape will display a pop-up window (also called a tooltip) that shows the values configured in the "Labels" section of the "Configuration" panel.

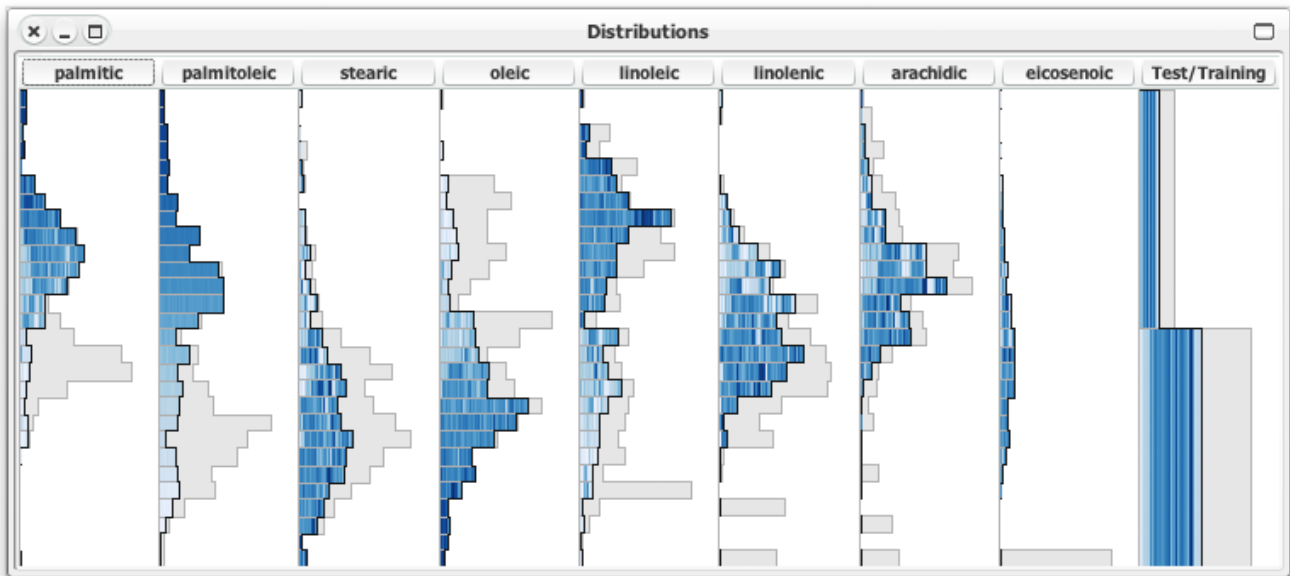
Probing and selection

Selection can be performed by clicking on a shape. Adding to or removing items from the selection can be done by holding the **Ctrl** key down while clicking on shapes.

Multiple adjoining items can be selected using a rubberband which is activated by dragging the mouse while holding down the **Alt** key.

Distributions view

The Distributions view shows how values are distributed for each variable.



At the top of the user interface, you will find the Distributions view that corresponds to the chosen settings in the Configuration and Axes panels. For each axis, items are grouped into bins whose width is proportional to the number of values.

Moving the mouse over a shape will display a pop-up window (also called a tooltip) that shows the values configured in the "Labels" section of the "Configuration" panel.

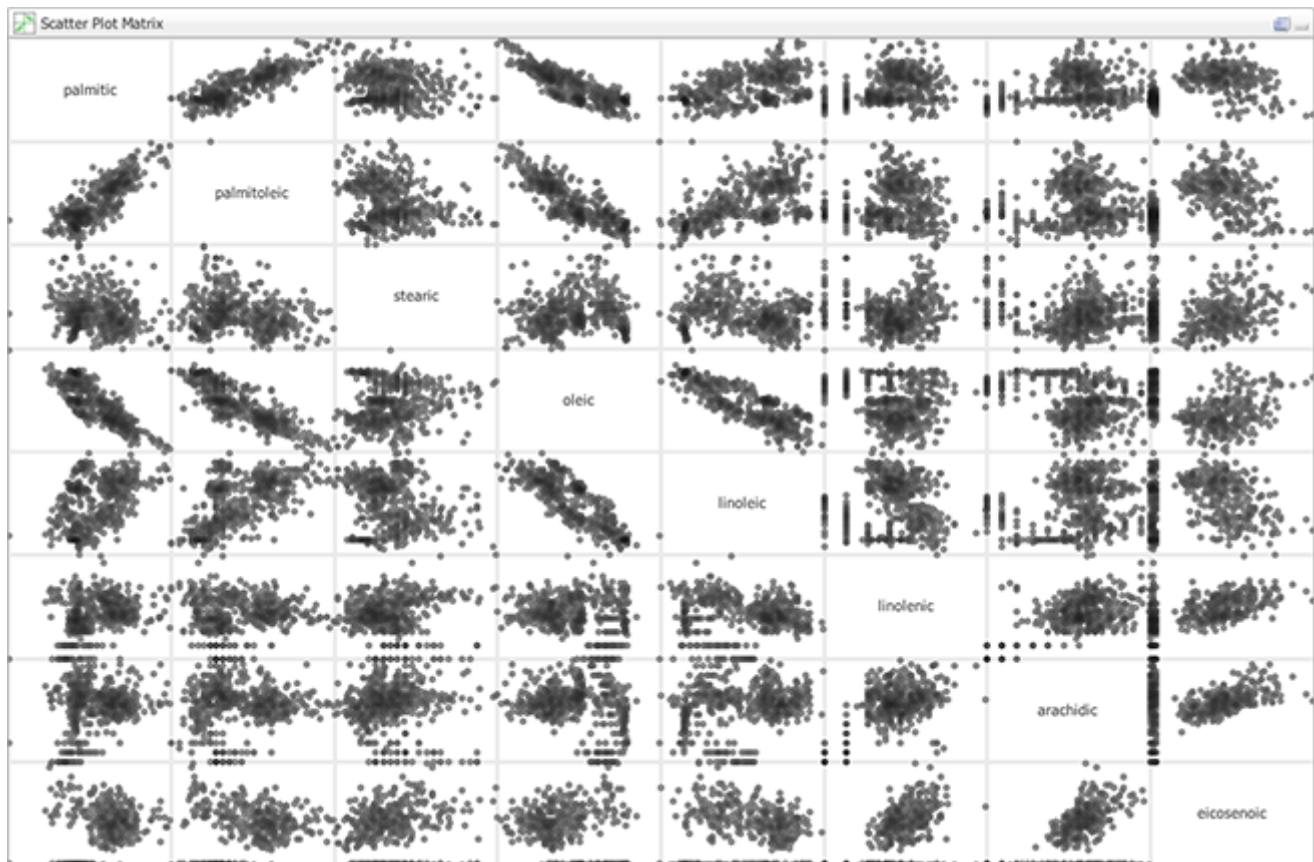
Probing and selection

Selection can be performed by clicking on a shape. Adding to or removing items from the selection can be done by holding the **Ctrl** key down while clicking on shapes.

Multiple adjoining items can be selected using a rubberband which is activated by dragging the mouse while holding down the **Alt** key.

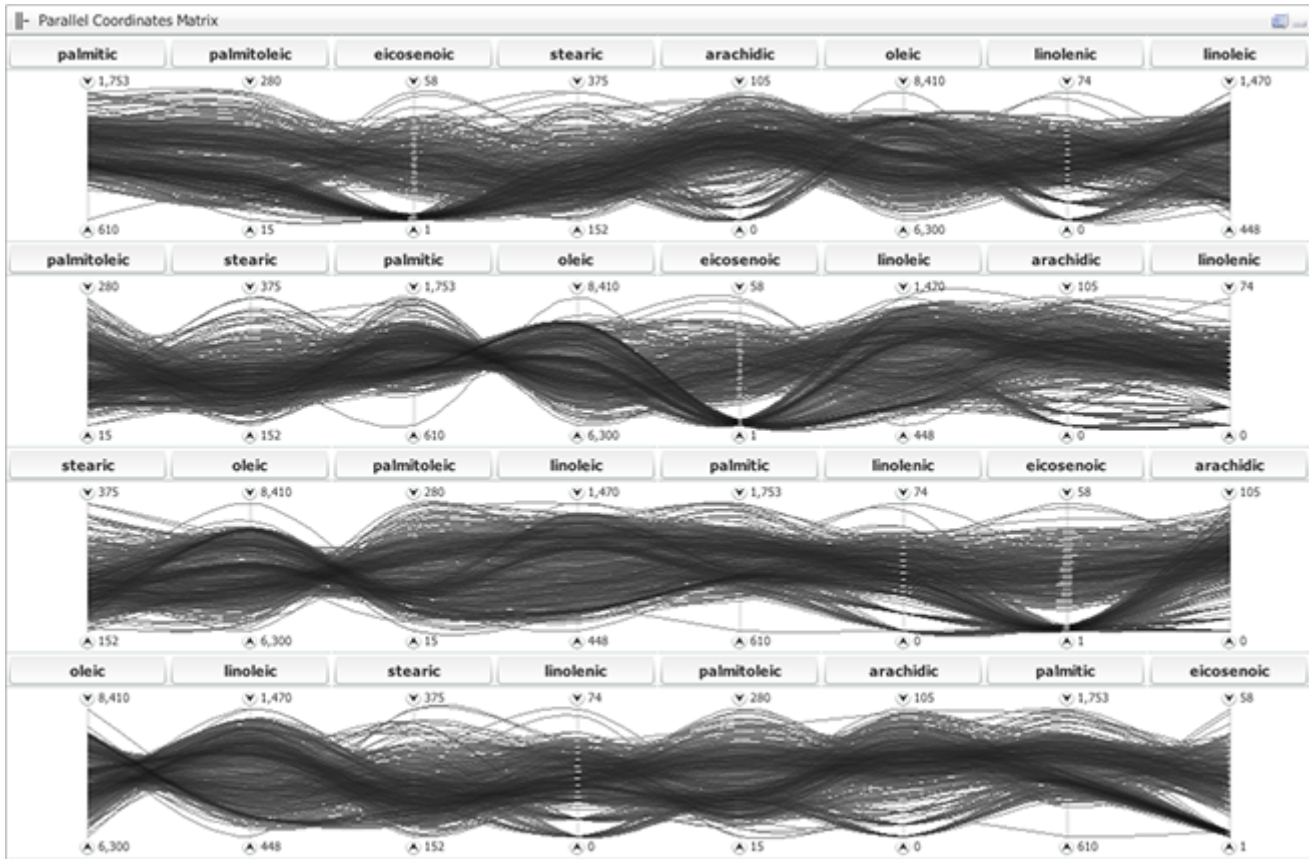
Scatter Plot Matrix view

The Scatter Plot Matrix view allows you to create a view containing a scatter plot for each pairwise combination of variables.



Parallel Coordinates Matrix view

The Parallel Coordinates Matrix view extends the parallel coordinates idea by providing a view of each pairwise relations between variables. Using our innate pattern-recognition abilities, it enables spotting correlations in a blink. Thanks to its unique density-based approach to avoid overplotting, and the choice between straight and curved geometries, relations emerge immediately.



At the bottom of the user interface, you will find the Parallel Coordinates Matrix view that corresponds to the chosen settings in the Configuration and Axes panels. Each item is represented by a polyline or polycurve.

Moving the mouse over a shape will display a pop-up window (also called a tooltip) that shows the values configured in the "Labels" section of the "Configuration" panel.

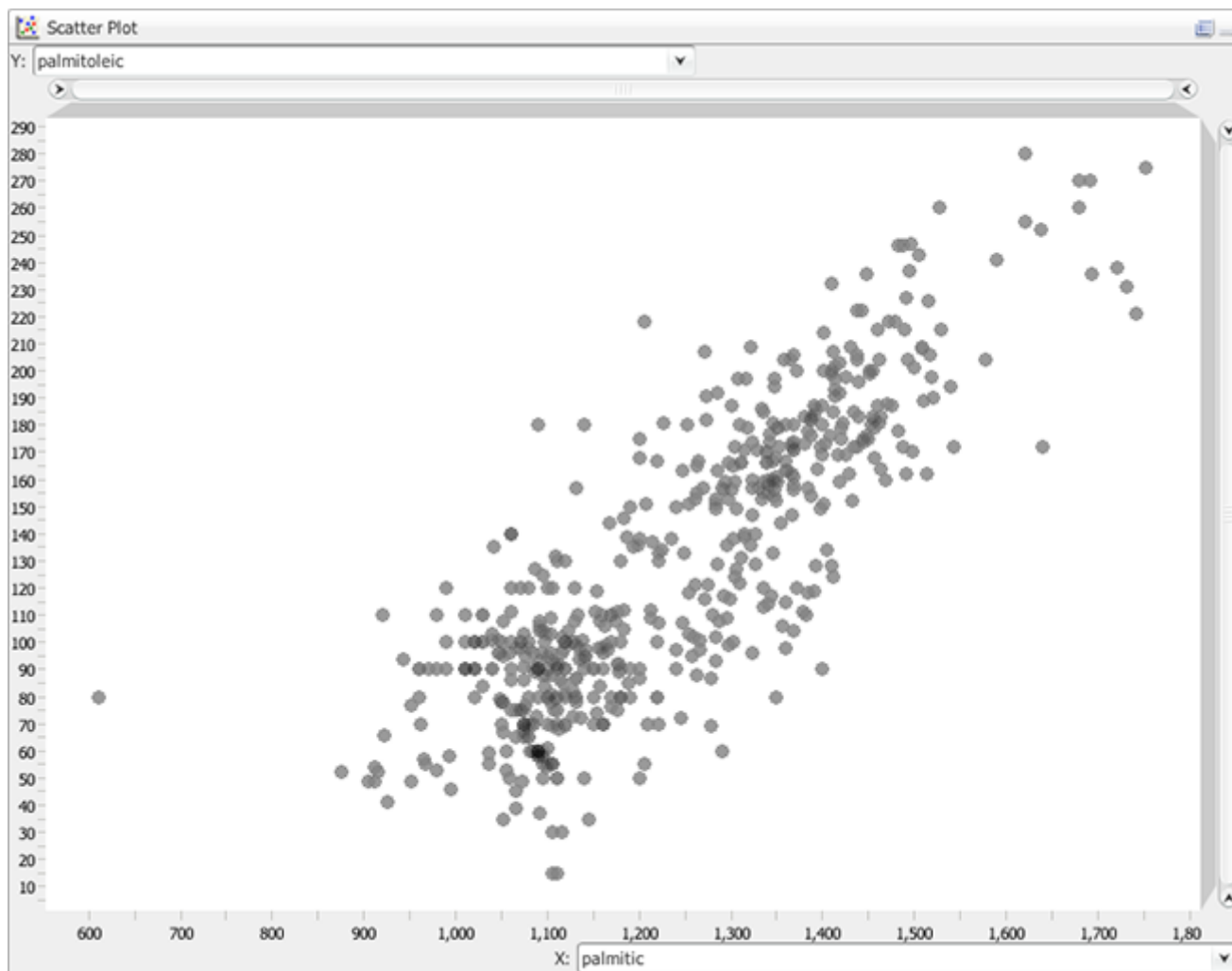
Probing and selection

Selection can be performed by clicking on a shape. Adding to or removing items from the selection can be done by holding the **Ctrl** key down while clicking on shapes.

Multiple adjoining items can be selected using a rubberband which is activated by dragging the mouse while holding down the **Alt** key.

Scatter Plot view

The Scatter Plot view allows you to create a scatter plot of the data. Any combination of numerical variables can be used to map to the x- and y-axes as well as size and color of the glyphs.



Configuration

To configure which of the numerical variables should be mapped to the x- and y-axes, use the drop-down lists located at the end of the axes.

The color and size of the markers are determined the same way as for the other views, i.e. the definitions in the drop-down lists in the Configuration panel.

Zooming

You can zoom in by using the range sliders on the top and to the right of the display area. You can zoom out by double-clicking anywhere on the slider.

And of course the mouse wheel also works.

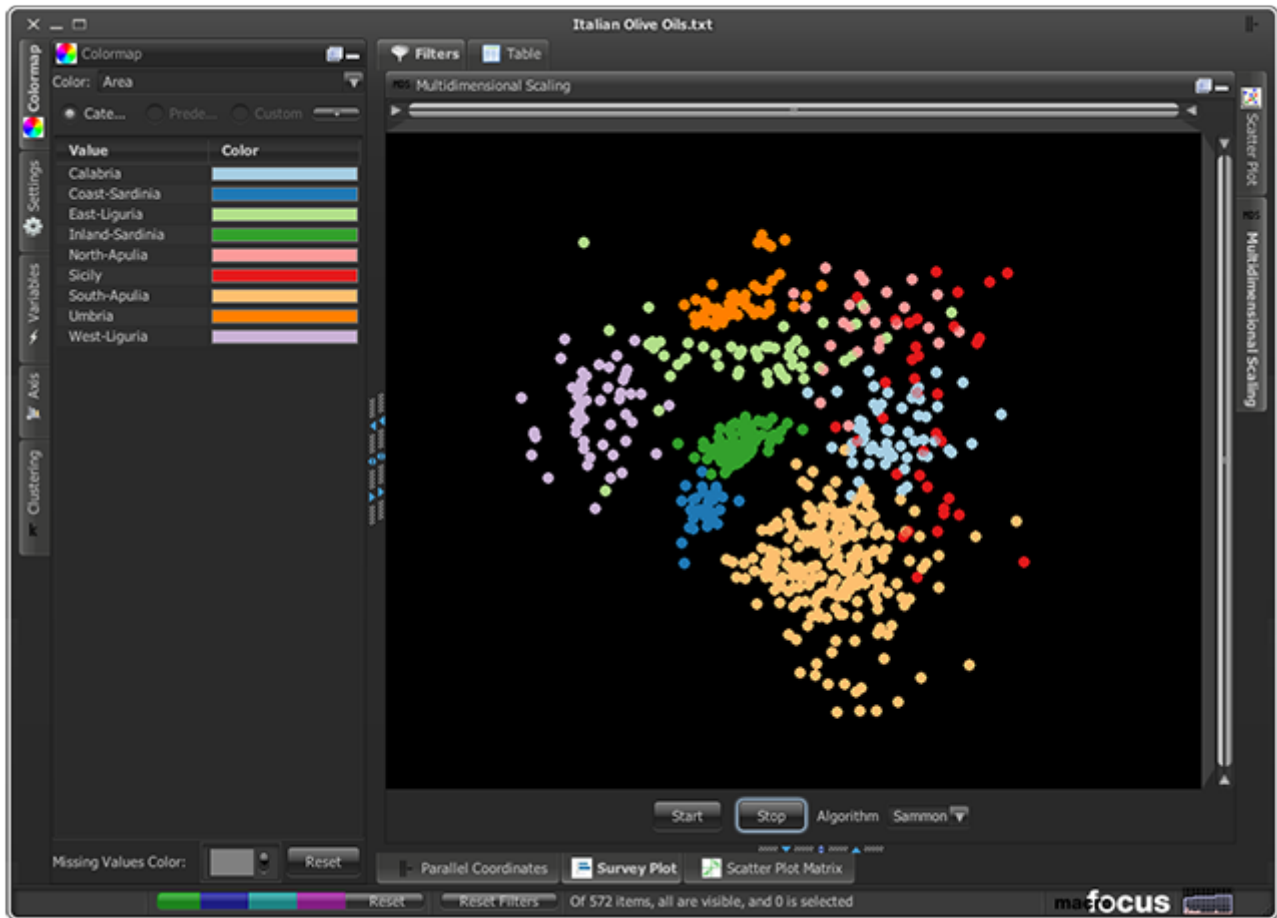
Probing and selection

Selection can be performed by clicking on a marker. Adding to or removing items from the selection can be done by holding the `Ctrl` key down while clicking on markers.

Multiple adjoining items can be selected using a rubberband which is activated by dragging the mouse while holding down the `Alt` key.

Multidimensional Scaling view

The Multidimensional Scaling view allows you to create a two-dimensional projection of the multidimensional data that attempts to capture the main relationships: items close together in the view are similar in the high-dimensional space while dissimilar one will be further away.



Computation

The process is iterative can be started using the Start button. When the layout has reached the desired stability, hitting the Stop button will terminate the computation. Two layout algorithms, Spring and Sammon are currently provided.

Several dimensionality reduction algorithms are provided, including **Sammon**'s mapping [Sammon1969], a **Spring**-based layout [Fruchterman1991], t-Distributed Stochastic Neighbor Embedding (**t-SNE**) [Maaten2008], and Principal Component Analysis (**PCA**) [Pearson1901].

Zooming

You can zoom in by using the range sliders on the top and to the right of the display area. You can zoom out by double-clicking anywhere on the slider.

And of course the mouse wheel also works.

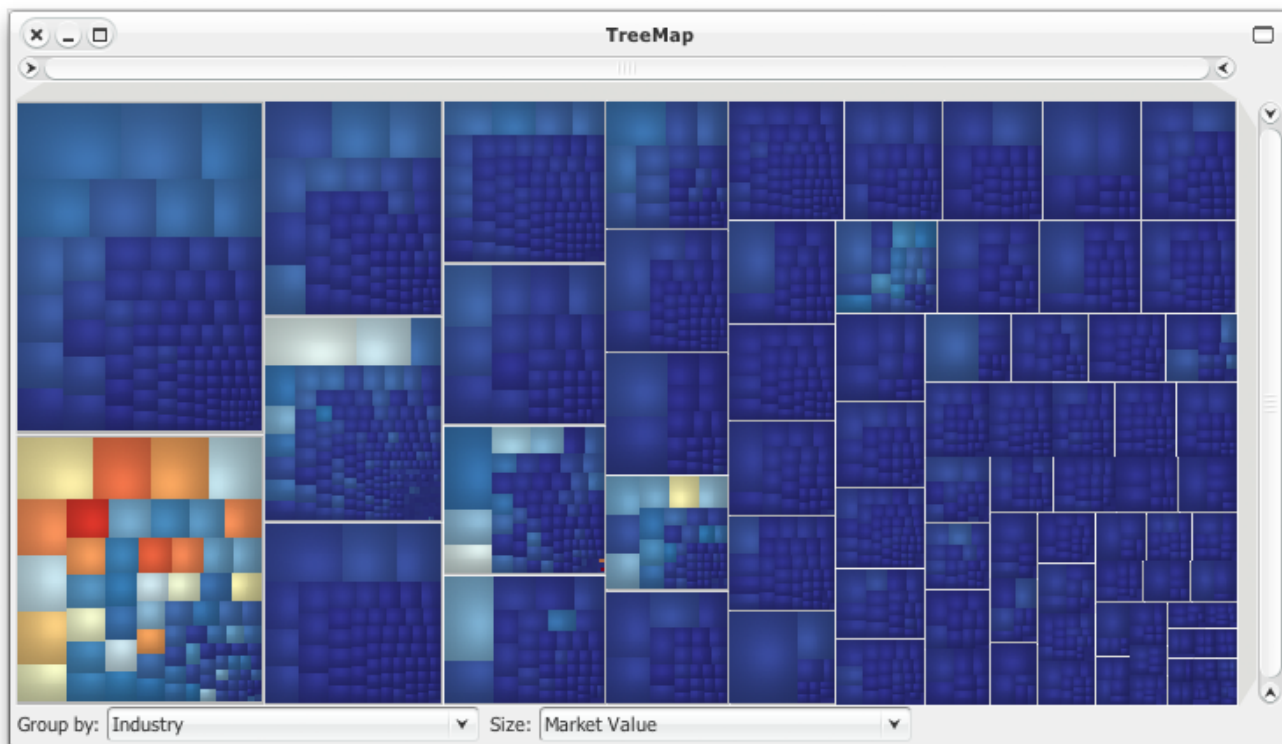
Probing and selection

Selection can be performed by clicking on a marker. Adding to or removing items from the selection can be done by holding the `Ctrl` key down while clicking on markers.

Multiple adjoining items can be selected using a rubberband which is activated by dragging the mouse while holding down the `Alt` key.

TreeMap view

The TreeMap view shows how values are distributed for each variable.



At the top of the user interface, you will find the TreeMap view that corresponds to the chosen settings in the Configuration and Axes panels. For each axis, items are grouped into bins whose width is proportional to the number of values.

Moving the mouse over a shape will display a pop-up window (also called a tooltip) that shows the values configured in the "Labels" section of the "Configuration" panel.

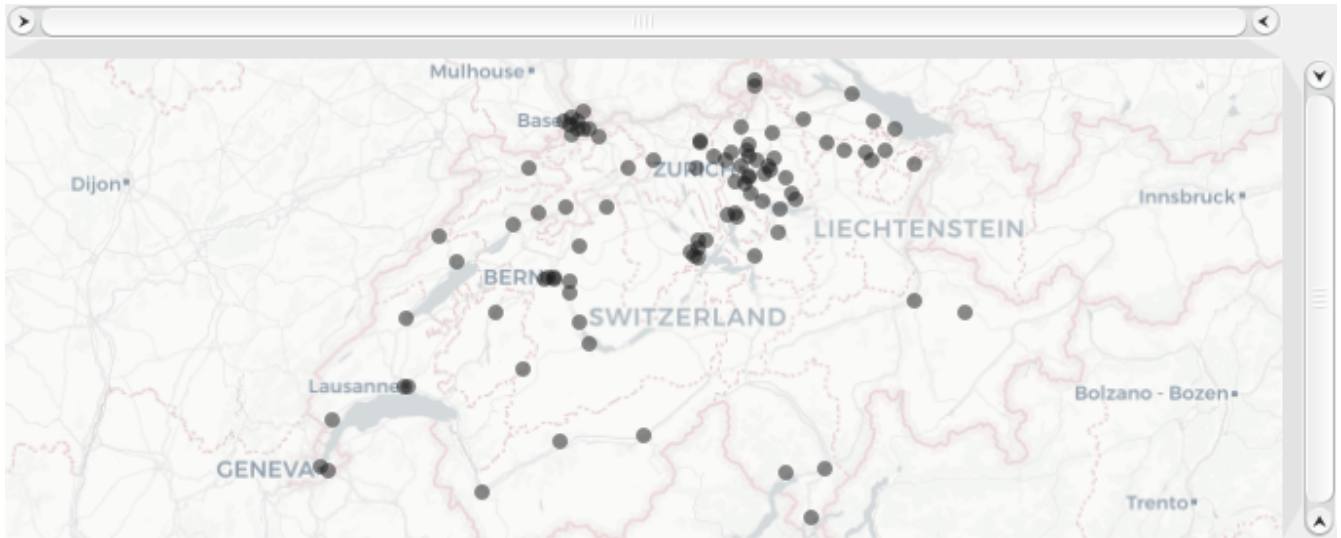
Probing and selection

Selection can be performed by clicking on a shape. Adding to or removing items from the selection can be done by holding the **Ctrl** key down while clicking on shapes.

Multiple adjoining items can be selected using a rubberband which is activated by dragging the mouse while holding down the **Alt** key.

CartoPlot view

When the data contains geographical features (Longitude/Latitude) coordinates, or geometrical objects such as lines and polygons), the CartoPlot view shows the items on top of geographical tiles obtained online map services.



Zooming

You can zoom in by using the range sliders on the top and to the right of the display area. You can zoom out by double-clicking anywhere on the slider.

And of course the mouse wheel also works.

Probing and selection

Selection can be performed by clicking on a marker. Adding to or removing items from the selection can be done by holding the `Ctrl` key down while clicking on markers.

Multiple adjoining items can be selected using a rubberband which is activated by dragging the mouse while holding down the `Alt` key.

Saving settings, data, and graphics

High-D can save the data along with the settings applied to the visualization in its own data format (file with `.mhd` extension). For this, use the **File** > **Save** or menu:File]Save As...] menu entries. To only save the settings and have the data file referenced instead of being embedded, you can produce such a file by doing:

1. Open a data file (Excel for example)
2. Modify all the parameters as desired
3. Do **File** > **Export Settings...**
4. Select **Macrofocus High-D (*.mhd)** as file type

When you open the resulting file (e.g. using **File** > **Open...**), it will read all the data from the referenced data file and apply all the settings. You can also see how the settings are stored in the resulting `.mhd` file. It can be opened using any text editor.

Exporting graphics

You can also export the currently active High-D view using the following schemes:

- Using **File** > **Export Graphics...**: the current view is exported in vector or raster form in one of the following supported formats:

PDF (Portable Document Format) (*.pdf)

The resulting document is ideal for printing or inclusion in a report. It is a vector format and therefore resolution independent.

Scalable Vector Graphics (*.svg)

The resulting document is ideal for further editing and for inclusion into another document. It is a vector format and therefore resolution independent. Scalable Vector Graphics (SVG) can be displayed by many web browsers with an embedded SVG viewer, or edited by any application supporting SVG (such as Adobe Illustrator).

Postscript (*.ps)

A common vector format and therefore resolution independent. Can be used for printing.

EMF (Enhanced Metafile) (*.emf)

A resolution independent format common on the Windows platform.

PNG (Portable Network Graphics) (*.png)

A raster format.

JPEG (*.jpg)

A raster format.

Compuserve GIF (*.gif)

A raster format.

TIFF (Tagged Image File Format) (*.tiff)

A raster format.

All the raster export format allow for setting the desired DPI for high-quality output.

- Using **Edit > Copy Graphics**: the current view is put into the clipboard in bitmap format (and can be pasted into applications such as Microsoft Powerpoint).

Exporting data

The data visible in the TreeTable view can be exported with **File > Export Data...** for further processing in spreadsheet programs or other applications. The following formats are supported:

CSV (Commad Delimited) (*.csv)

The comma-separated values (CSV) format stores tabular data (numbers and text) in plain-text form. Most spreadsheet and data management software are able to import data in this format.

Text (Tab Delimited) (*.txt;*.tsv;*.tab;*.raw)

The tab-separated values format is a popular method of data interchange among databases and spreadsheets. It stores tabular data (numbers and text) in plain-text form.

Microsoft Excel Workbook (*.xls;*.xlsx;*.xlsm)

The comma-separated values (CSV) format stores tabular data (numbers and text) in plain-text form. Most spreadsheet and data management software are able to import data in this format.

Apache Arrow (*.arrow)

The Arrow format stores tabular data (numbers and text) in form that allow data access without serialization overhead.

Apache Parquet (*.parquet)

The Parquet format stores tabular data (numbers and text) in a compressed, efficient columnar data representation. It is popular in the Hadoop ecosystem.

Import/export of settings

All the settings can be exported using **File > Export Settings...** to be applied to another dataset using **File > Import Settings...**

Printing

Using **File > Print** to get a printout of the active High-D view (note that the resulting print job can also be redirected to a file)

Invoking High-D through the command line

High-D can be invoked from the command line, typically for automating and batch processing the production of several visualizations.

If you intend to use this scripting possibility in unattended and automated batch jobs (typically a night job running on a remote build server): non-human devices that utilize our software without user interaction are counted as users and you would then need to order the appropriate number of licenses.

High-D comes bundled with its own optimized Java runtime (that can be found in the `jre` directory), which is order of magnitude faster for certain operations than the standard Java runtime. Nevertheless, High-D is fully compatible with Java 11. The invocation of High-D from the command line can be done as follows:

Windows

Start Command Prompt application and then type:

```
cd "C:\Program Files\High-D"  
jre\bin\java -jar lib/high-d-swing.jar --uiscaling 1.2 data.mhd`
```

macOS

Start Terminal application

```
cd /Applications/High-D/  
./install4j/jre.bundle/Contents/Home/bin/java -jar lib/high-d-swing.jar --uiscaling  
1.2 data.xls
```

Linux

```
cd /usr/local/TreeMap  
jre/bin/java -jar lib/high-d-swing.jar --uiscaling 1.2 data.csv
```

Command line options

High-D can be invoked from the command line with the following options::

-h, --help

Show the help

-e, --expert

Run High-D in expert mode

-f, --lf <argument>

Set the look and feel

-u, --uiscaling <argument>

Scale the UI

Bibliography

- [Fruchterman1991] Thomas M. J. Fruchterman, Edward M. Reingold (1991), "Graph Drawing by Force-Directed Placement", *Software – Practice & Experience*, Wiley, 21 (11): 1129–1164, [doi:10.1002/spe.4380211102](https://doi.org/10.1002/spe.4380211102).
- [Inselberg2009] Alfred Inselberg (2009). "Parallel Coordinates: Visual Multidimensional Geometry and its Applications". Springer. ISBN 978-0-387-68628-8.
- [Maaten2008] L.J.P. van der Maaten and G.E. Hinton (2008). "Visualizing High-Dimensional Data Using t-SNE", *Journal of Machine Learning Research* 9 (Nov): 2579-2605. https://lvdmaaten.github.io/publications/papers/JMLR_2008.pdf.
- [Pearson1901] Karl Pearson (1901). "On Lines and Planes of Closest Fit to Systems of Points in Space", *Philosophical Magazine*. 2 (11): 559–572. [doi:10.1080/14786440109462720](https://doi.org/10.1080/14786440109462720).
- [Sammon1969] John W. Sammon (1969). "A nonlinear mapping for data structure analysis", *IEEE Transactions on Computers*. 18 (5): 401–409, [doi:10.1109/t-c.1969.222678](https://doi.org/10.1109/t-c.1969.222678).