

CASSATT-Documentation

Abstract

The interactive software CASSATT has been developed to visualise high-dimensional, multivariate data using parallel coordinates. CASSATT is platform independent and widens the range of methods for analysing parallel coordinate data.

Parallel coordinate plots are a generalisation of two-dimensional scatterplots. The orthogonal axes are replaced by parallel axes, such that a point from a scatterplot is represented by a line in the parallel coordinate plot. This method allows you to plot multidimensional data in a two-dimensional display.

Analysis of high-dimensional data sets with parallel coordinate plots requires special techniques. CASSATT uses interactive tools to extend the basic graphic displays.

Using CASSATT

● Download and Installation

CASSATT can be downloaded from http://www1.math.uni-augsburg.de/~winklers/Cassatt_Download.html.

Both PC and Macintosh versions are available.

● Macintosh:

Copy the "Cassatt" folder onto your hard disk.

● Windows (fast processor recommended):

Copy the "Cassatt" folder onto your hard drive, e.g. to "C:\Cassatt", retaining the given structure of subdirectories.

● Starting CASSATT

CASSATT can run on every platform since it is implemented in Java.

● Macintosh:

Go to "Startup" folder in the "Cassatt" folder. Double-click the CASSATT icon.

● Windows:

Go to the main Cassatt folder (i.e. "C:\Cassatt" in the example above). Then double-click the file "Start_C.bat".

If you have installed CASSATT on a hard drive with a different name, say "D:", use "Start_D.bat" instead.

If this does not work, follow the instructions of the read_me file.

After CASSATT has started, you may choose a dataset in the dialog box which initially appears. If you close the dialog box without choosing a dataset, you can select 'open' from the file menu in the main window.

(A hint for Java experts: The class containing the main method is called "Cassatt" and is a part of "AppClasses.jar")

● Data Format

CASSATT supports the standard text matrix file, i.e. tab separated columns with a row of labels at the top. **Unfortunately, at this stage of implementation all data have to be numeric!** Other data will generate a number format exception.

● Data Sets

CASSATT comes with a number of data sets for testing and demonstration purposes.

You can select 'open' from the file menu in the main window. Note that CASSATT can handle many data sets at the same time. The data sets loaded are listed in the Data Set window and may be selected by clicking on their names.

● Plots

To produce a plot select a suitable number of variables in the variables list. Then choose a plot from the Plots menu or press one of the buttons for a Parallel Coordinates Plot (PC Plot). Parallel Coordinate Group Plots (or Group Plot for short) require a group to be selected in the group list.

● Parallel Coordinates Plots

A PC Plot can handle many variables at once.

By default, the parallel dotplot with a standardised scale appears first and all variables are visible in the diagram window. Depending on the number of variables the PC Plot displays, you may want to alter the separation between variables (using shift and + or -) to produce a clearer image. The PC Plot then offers various features:

● Shortcuts

- add and remove lines, boxes, menu bar ...

- toggle selection

- undo the last selection

- alter the point size and the distance between the variables

-

● Hot spots and mouse features: note the altered cursor appearance(see picture below)

- invert the axis of a variable

- drag a variable to another position

- exchange two variables by drag and drop (this takes place only when the second variable turns blue)

- angle selection

- query of points and variable names

- remove selection by clicking on the plot

-

● Menu

- switch from a standardised scale for each variable to a common scale for all variables and vice versa

- sort variables according to various criteria: mean, median, minimum, maximum

- change the colours of the selected or unselected points/lines

- hide selected or unselected points and lines

-

For more information on shortcuts and features press "Help" in the menu of the PC Plot.

● Groups

To create a group press the "Make Group" button in the tool frame.

All selected items will be a member of this new group. In the dialog window which appears you can specify a name and choose a colour for this group. Subsequent to this, the name and the number of members of the new group appear in the group list in the main window.

For an application of groups see example below.

● Parallel Coordinates Group Plots

A Group Plot is a special form of a PC Plot, so you can use the same features as in the PC Plot.

The only difference is that this plot ignores all data not belonging to the group, i.e. only group members appear in the plot and scaling and selection relate only to the group members.

● Selection and Modes

Selecting and interactive highlighting data are the main features of CASSATT: there are different ways of selecting cases.

● Points:

- Drag Box: Click and drag a rectangle in which all points are selected. (Dotplot, Boxplot and Scatterplot)

- Vertical Drag: Click and drag near the line of the axis of the variable. Note the altered cursor appearance. (PC-Plot, Group Plot)

● Lines (PC-Plot and Group Plot only):

- Drag Box: Click and drag a rectangle in which all lines are selected.

- Angle Selection: Click to the left beneath the variable axis. (Note the altered cursor.) A dialog box will pop up in which you can select angles graphically or by typing in the exact angle range desired.

In the tool frame at the top of the screen you can select different selection modes:

● Replace Mode (Standard):

A new selection replaces the old one. Selecting with the shift key pressed activates exclusion mode for the current selection but retains Replace Mode for further selections.

● Intersection Mode:

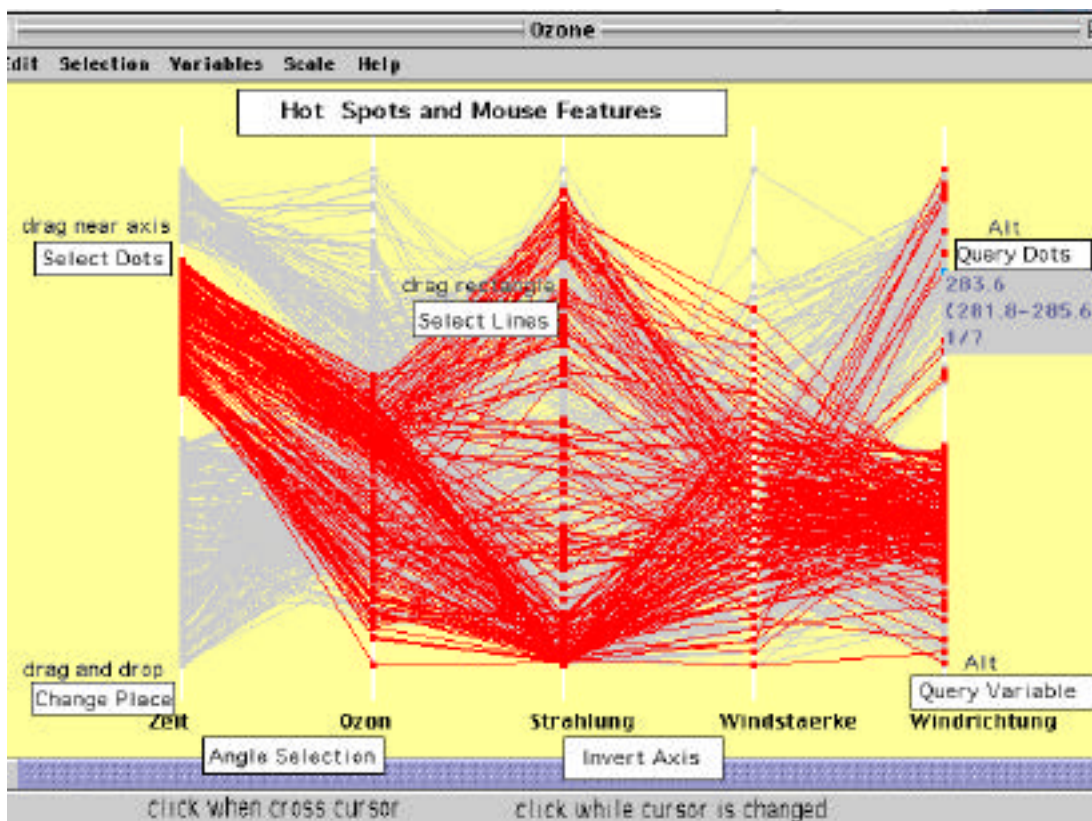
Cases are only selected from the current selection.

● Union Mode:

New selections are added to the previous selection.

● Exclusion Mode:

Cases selected from the previous selection are de-selected, newly selected cases are added to the previous selection, i.e. the previous and the current selection are XORed.

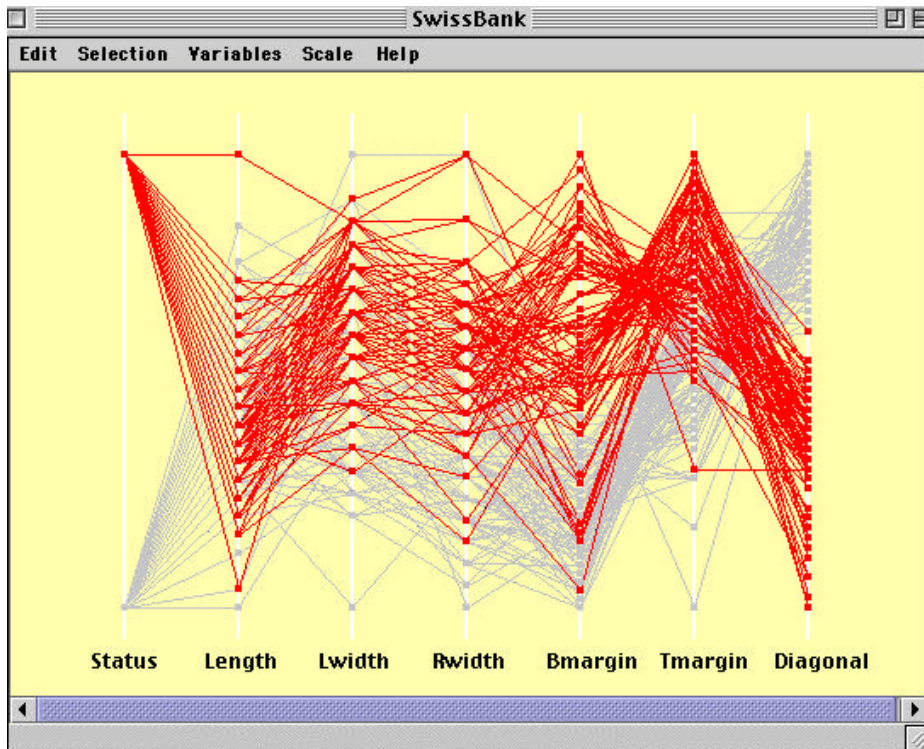


Example of a data analysis with groups in CASSATT

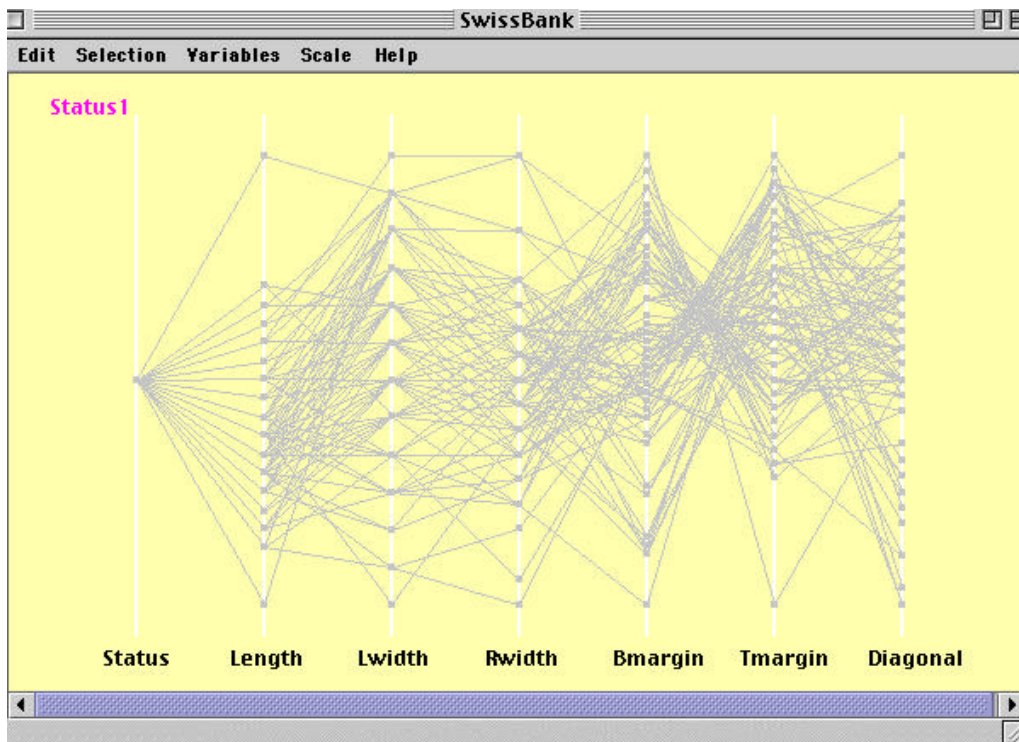
Data set: SwissBank

This data set contains the data on 200 bank notes. The Status says whether a bank note is a forgery or not.

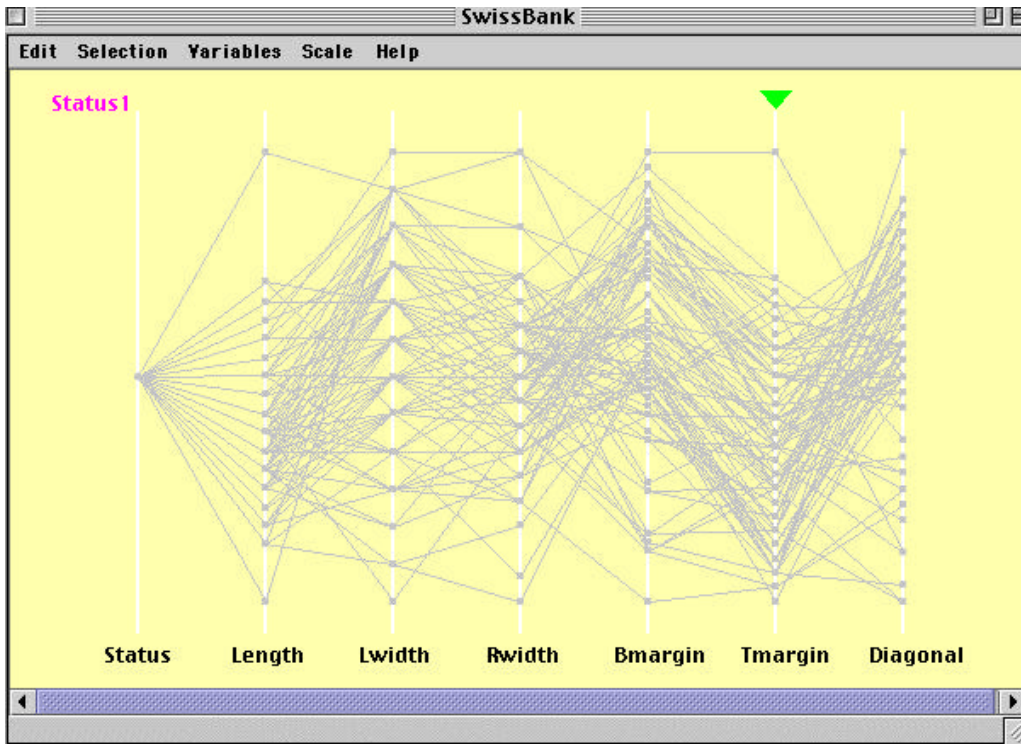
Produce a PC Plot with all variables, add lines by pressing "shift" and "L". Select all cases with Status=1:



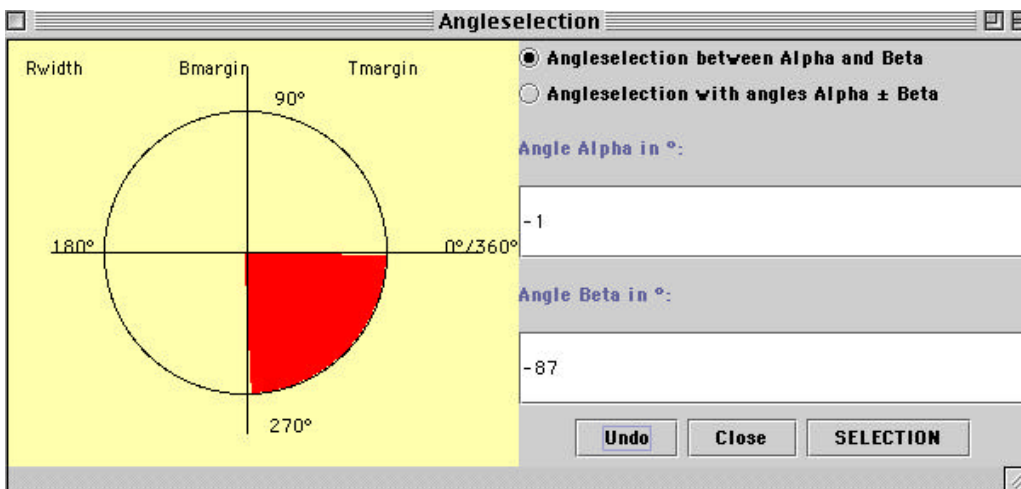
Produce a new group by pressing the "Make Group" button. Set the name and choose a colour. Plot all variables in a Group Plot by selecting the new group and pressing the "Group Plot" button. Add Lines as in the PCPlot.



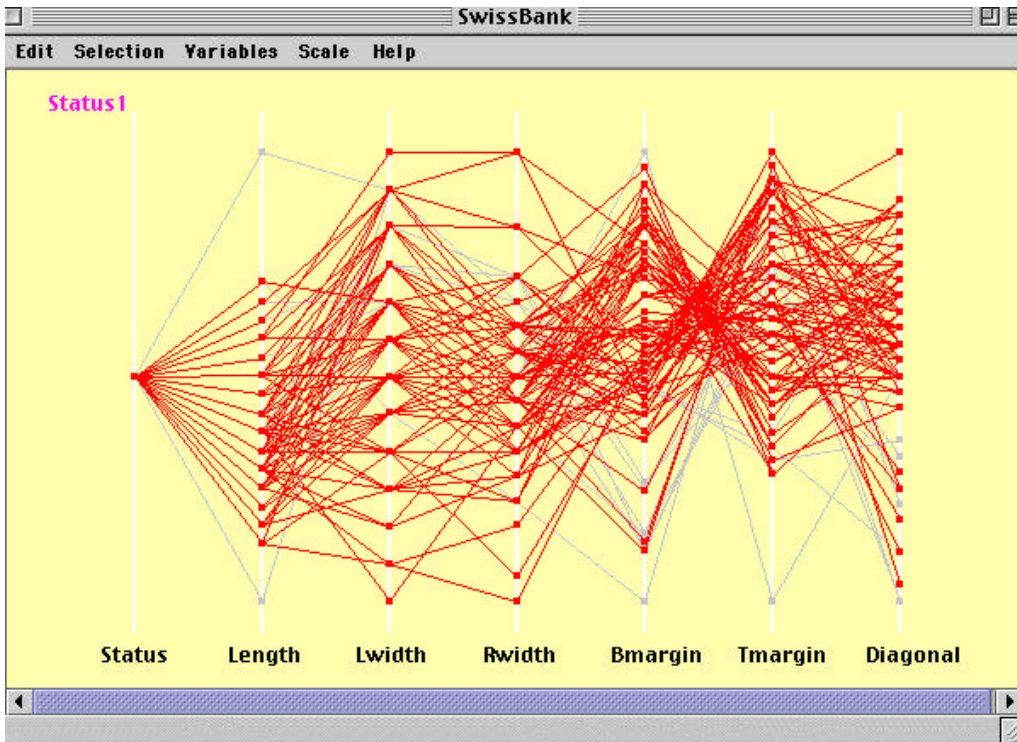
As you can see there are one or maybe two nodes between the variables Bmargin and Tmargin. Invert the axis of Tmargin and you can easily see a lot of parallel lines..



Select these lines with the help of angle selection. You can do it by dragging the mouse in the circle or by typing angles into the textfields and then pressing SELECTION.

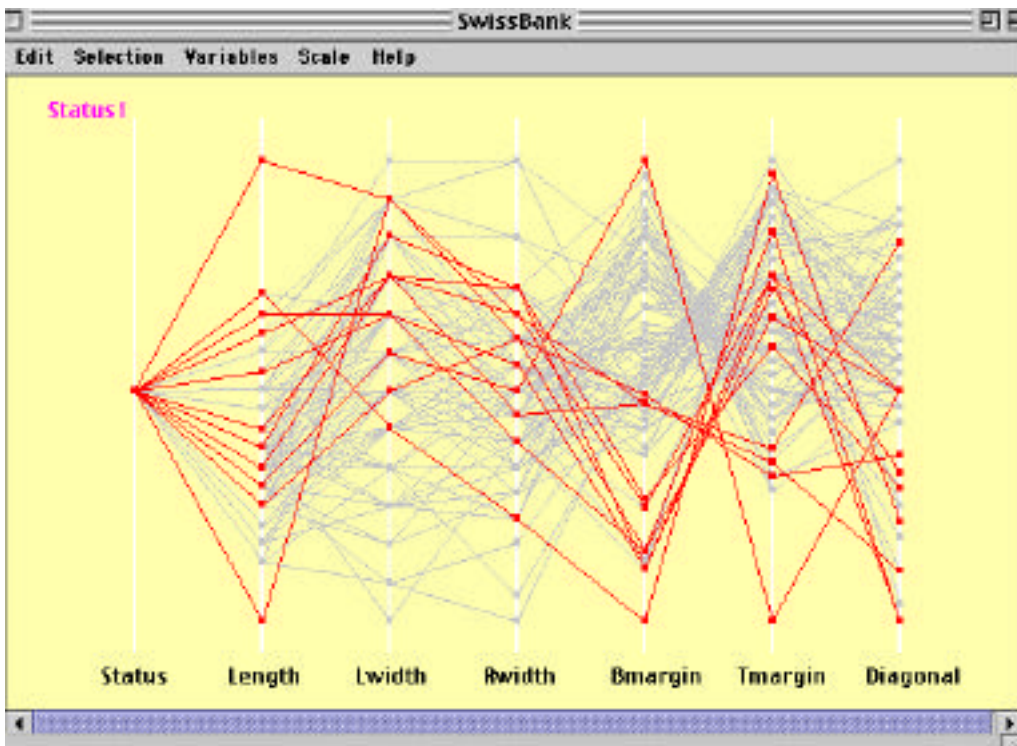


Invert the axis again to get the original picture. Now you can see a second node below the big one.



Toggle the data by pressing "shift" and "T". Now all other cases are selected. To get only the selected cases of the group change the selection mode in the toolframe to "Intersection".

Now select all cases in the Group Plot, e.g. Dotselection at the first variable Status.



Produce a Scatterplot for the variables Bmargin and Tmargin by selecting these two variables in the variable list. (Plot menu!) The red cases are our third group.

