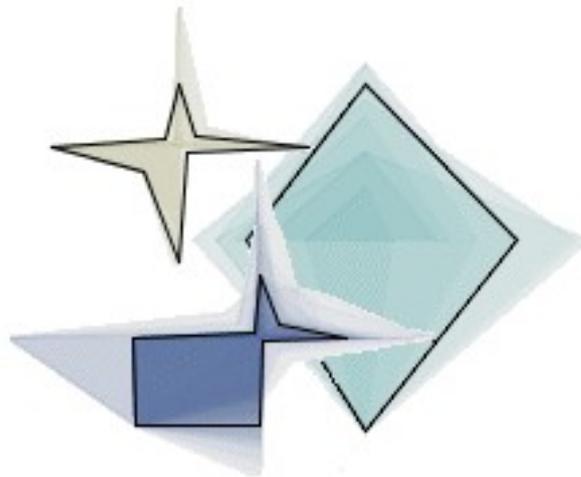


# Gauguin

Interactive Glyph Analysis with R

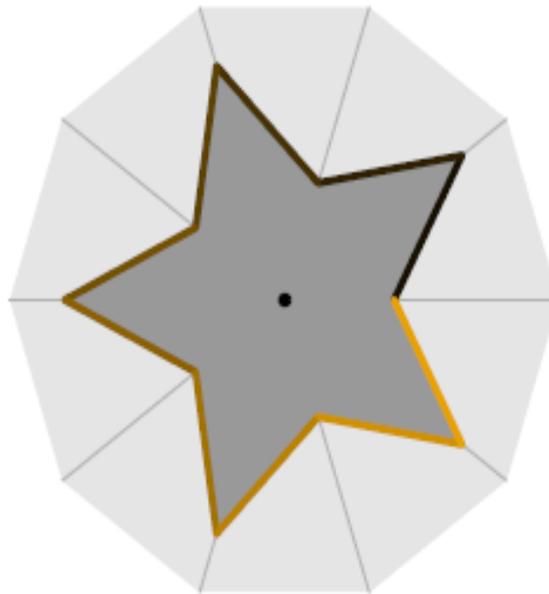


Alexander Gribov, Antony Unwin (University of Augsburg)

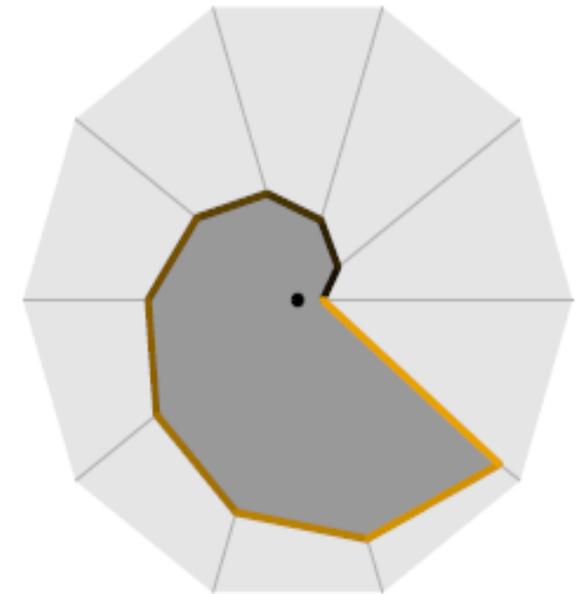
# Glyphs

- geometric shapes scaled by the values of multivariate data.
- each glyph represents usually one high-dimensional data point (or average of data points).

4,8,4,8,4,8,4,8,4,8



1,2,3,4,5,6,7,8,9,10



Two simple examples for star glyphs. Each glyph represents ten numbers between 0 and 10 as shown in the titles.

# Glyphs

Gauguin offers four different glyph shapes:

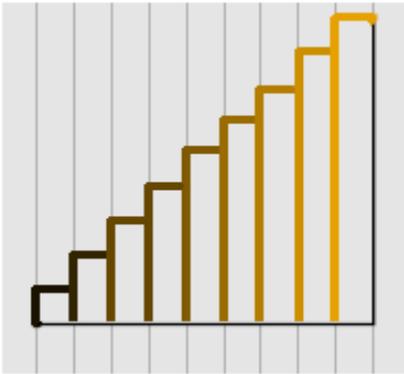
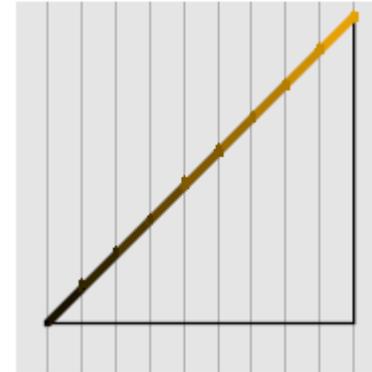
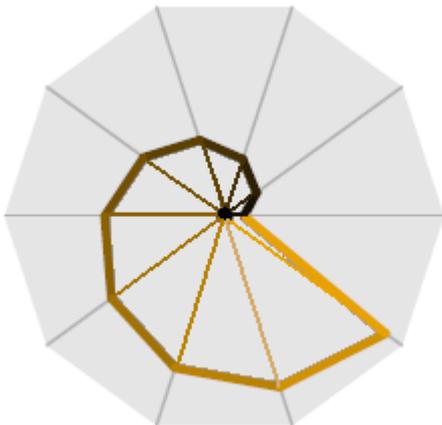


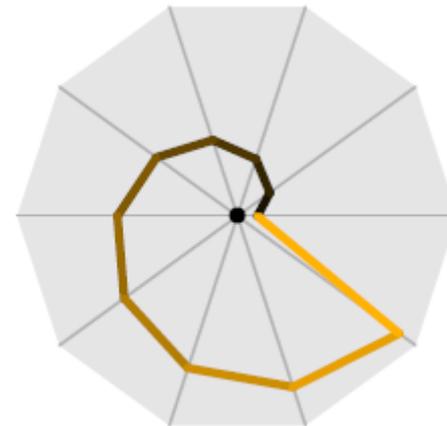
chart glyph



line glyph



filled star glyph



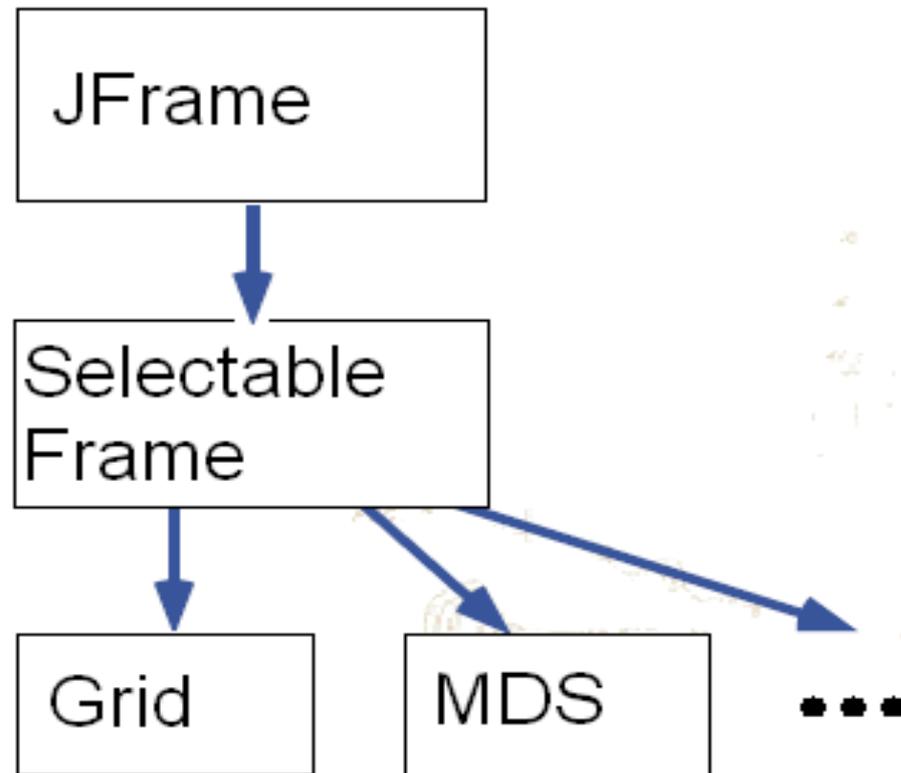
unfilled star glyph

# What is Gauguin?

- Project for the interactive visual exploration of multivariate data sets
- Supports a variety of methods for displaying flat-form and hierarchically clustered data
- Developed in Java
- Multiplatform support (Windows, Linux, Mac)
- Connection to R via Rserve for calculations

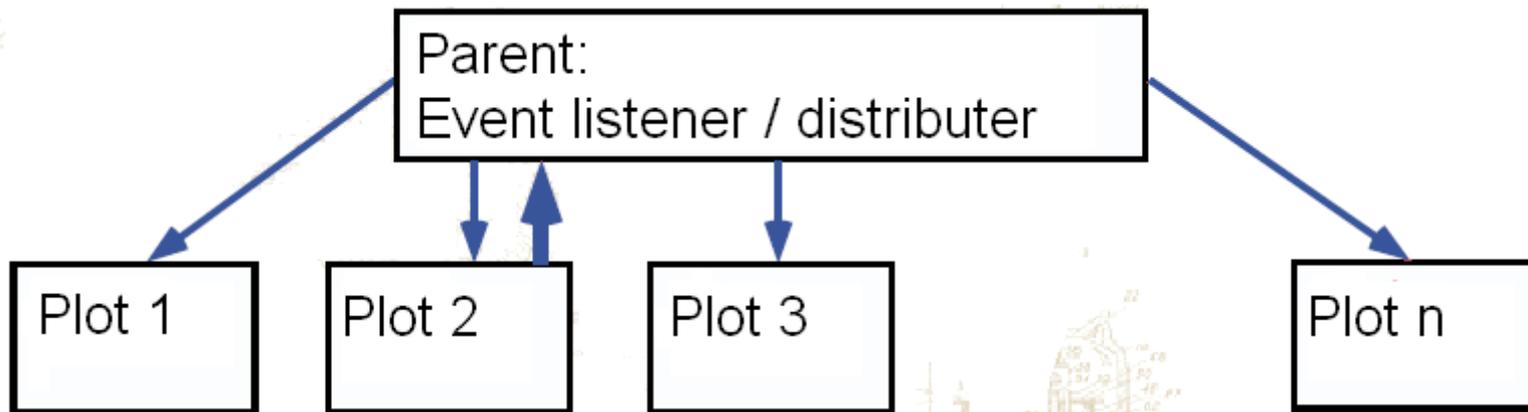
# Design Choices

- Graphs



# Event Handling

- Events in Gauguin:
  - Selection Event
  - DataChanged Event
- Example: Selection Event in Plot 2



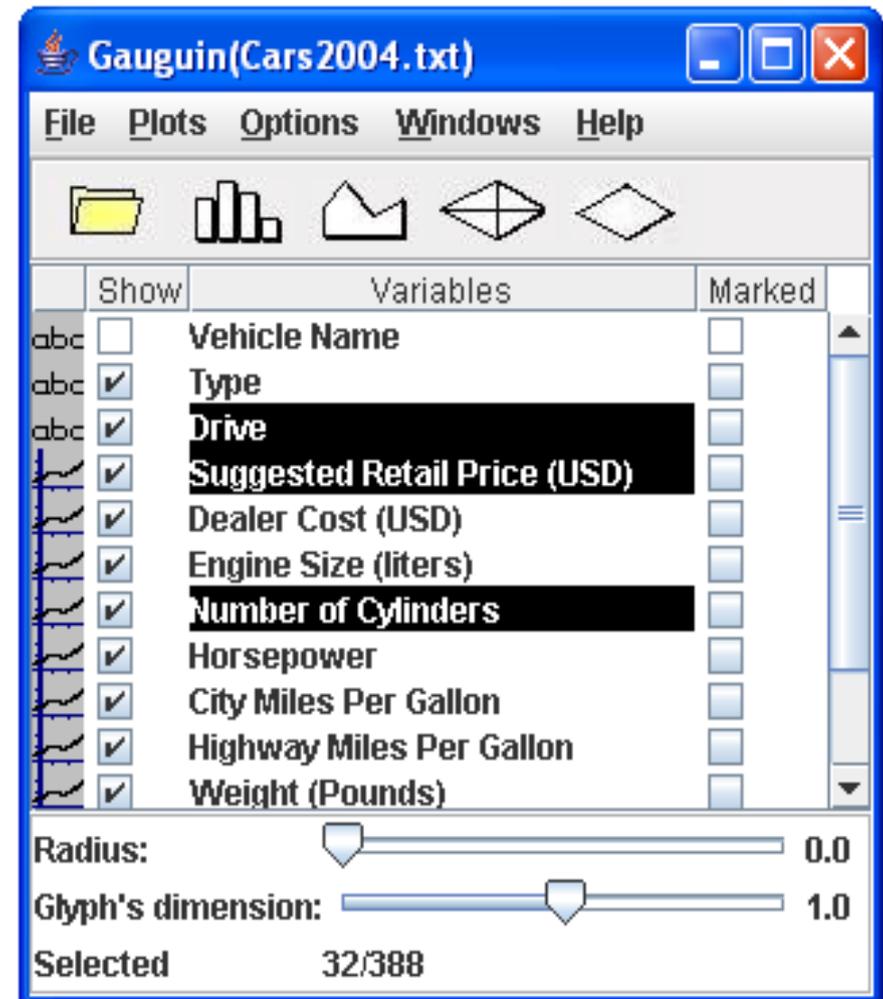
# Main Window

Allows the user:

- change glyph mode
- enable,disable,mark or delete variables
- change the order of variables
- specify the glyph's size

Selection is relevant for plots and queries.

Radius allows to highlight the cases similar to an individual or to the group of selected glyphs.



# Plots in Gauguin

All support:

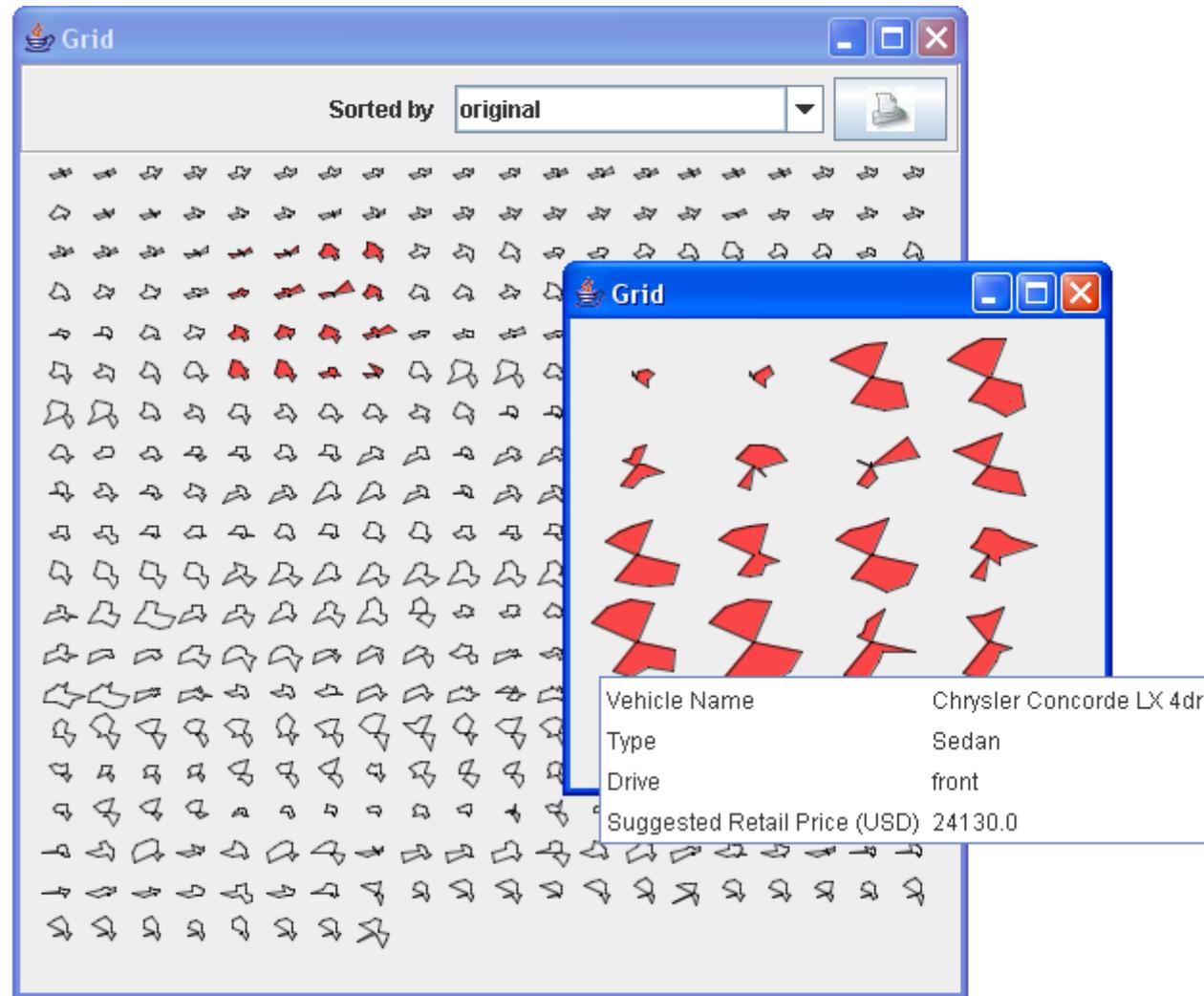
- selection and linked highlighting
- zooming, if there is a suitable coordinate system
- interrogation of objects

... all plots are aware of big datasets

# Plots in Gauguin

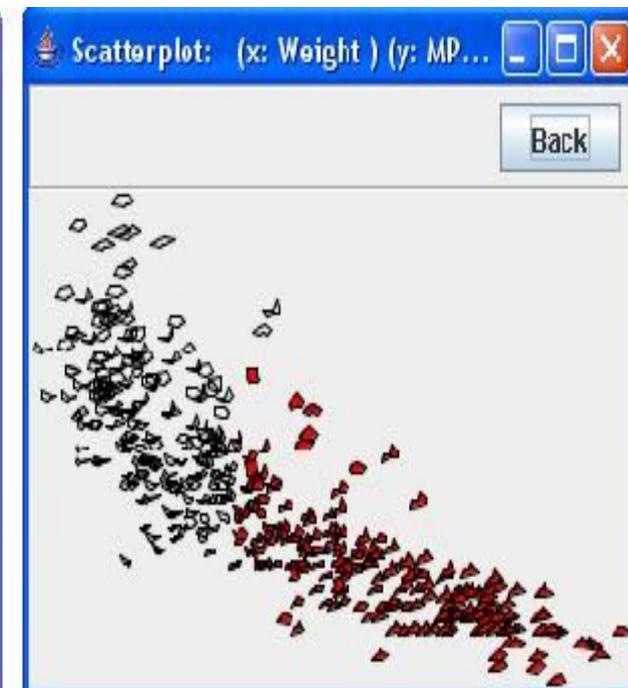
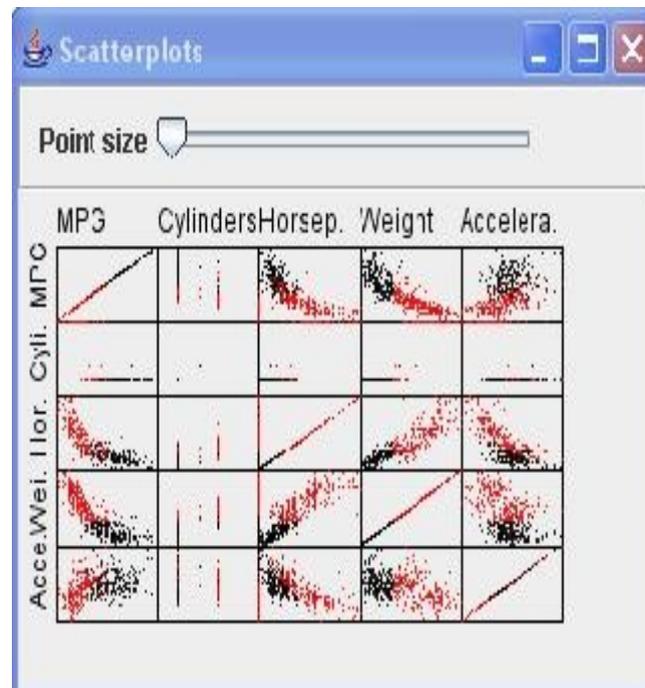
- Grid

- overview over the data.
- supports sorting
- locally scaled zooming



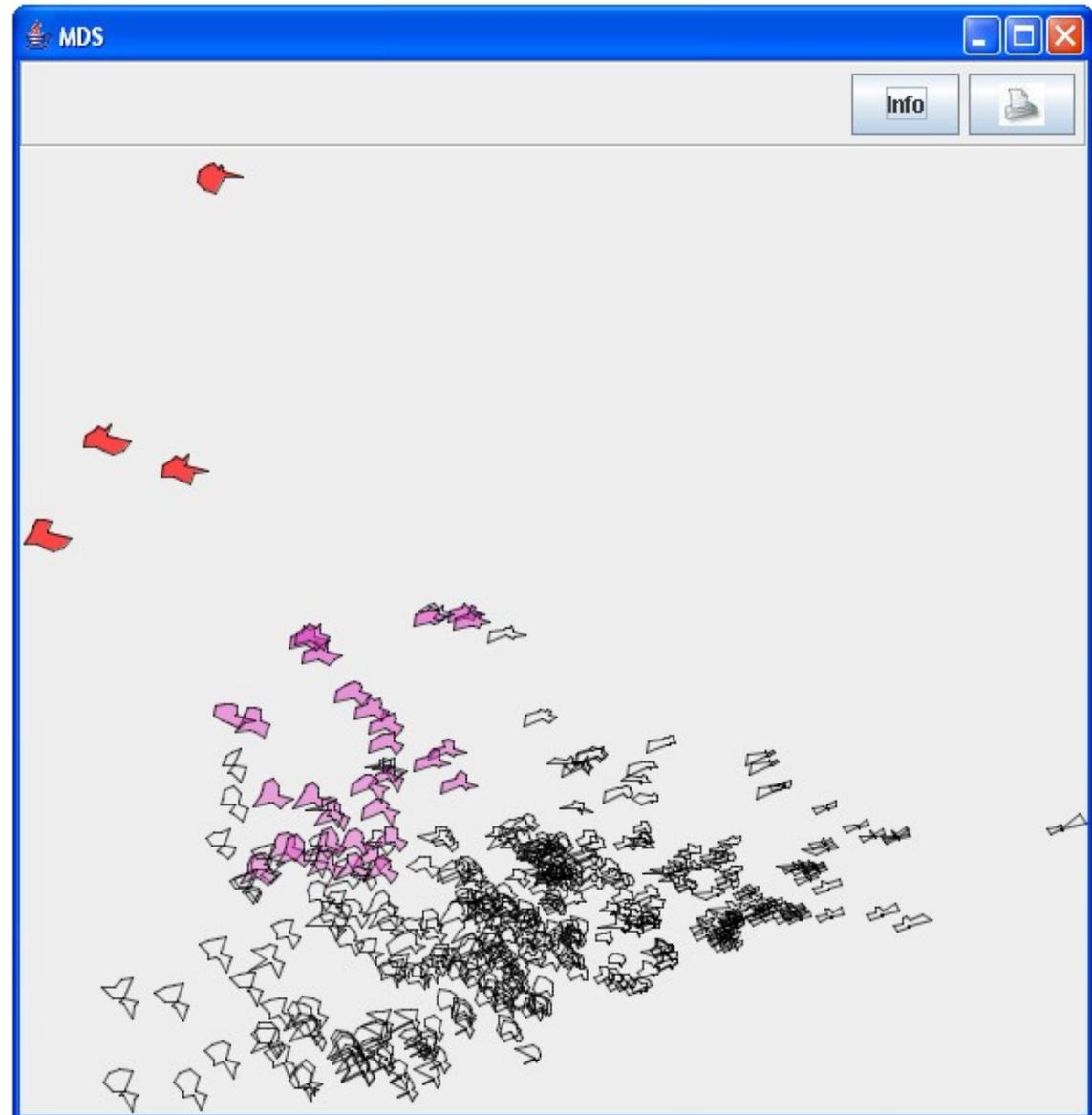
# Plots in Gauguin

- Scatterplots
  - all pairwise scatterplots of variables selected from the list in the main window
  - zooming for detailed view
  - tonal highlighting if points are plotted



# Plots in Gauguin

- MDS  
(Multidimensional scaling)
  - offers different methods
    - isoMds
    - sammon
    - cmdscale
  - central view

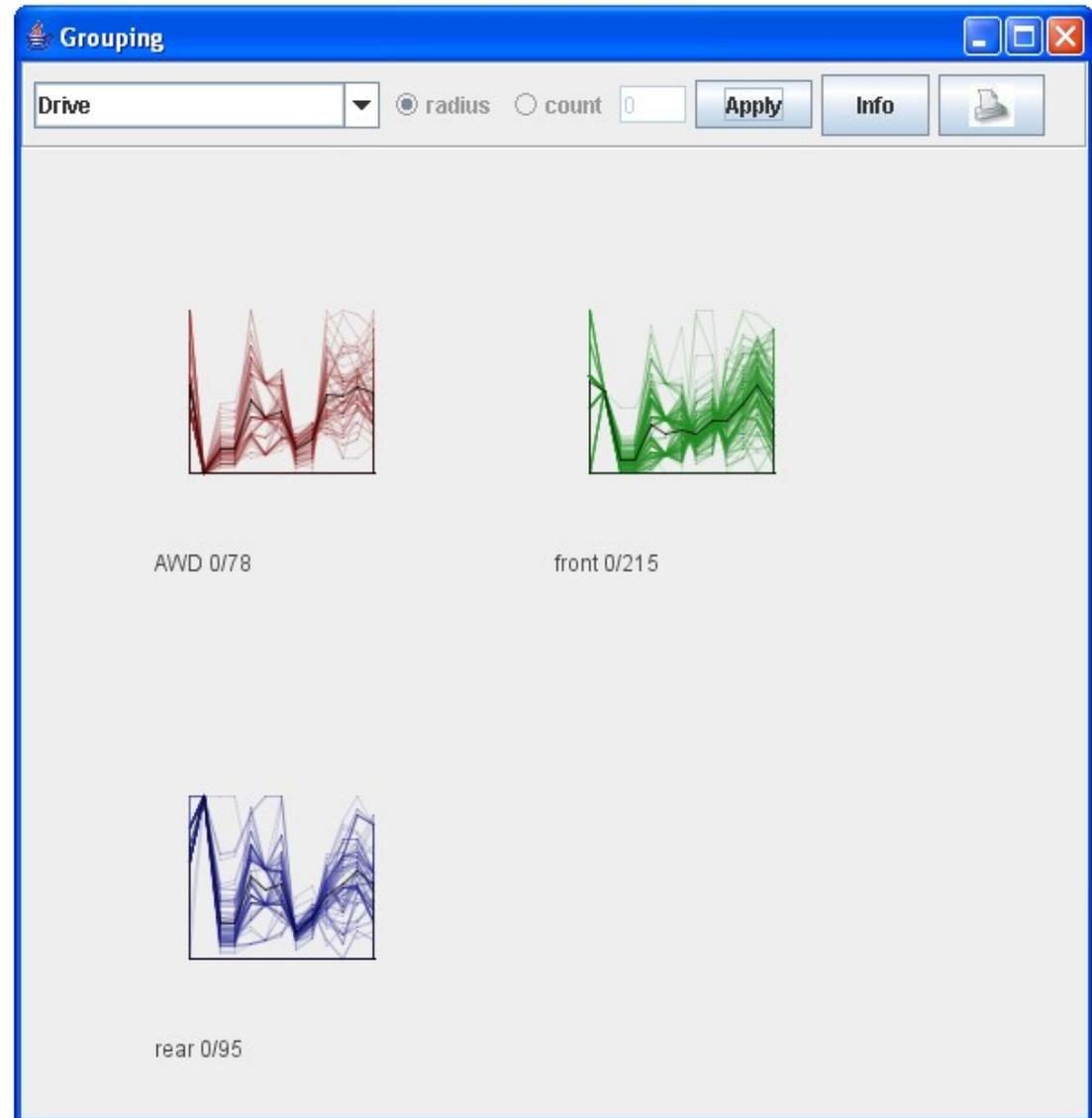


# Plots in Gauguin

- Grouping

- represented as the average and the band of the glyphs in the group

- grouping by:
  - category
  - radius
  - count



# Plots in Gauguin

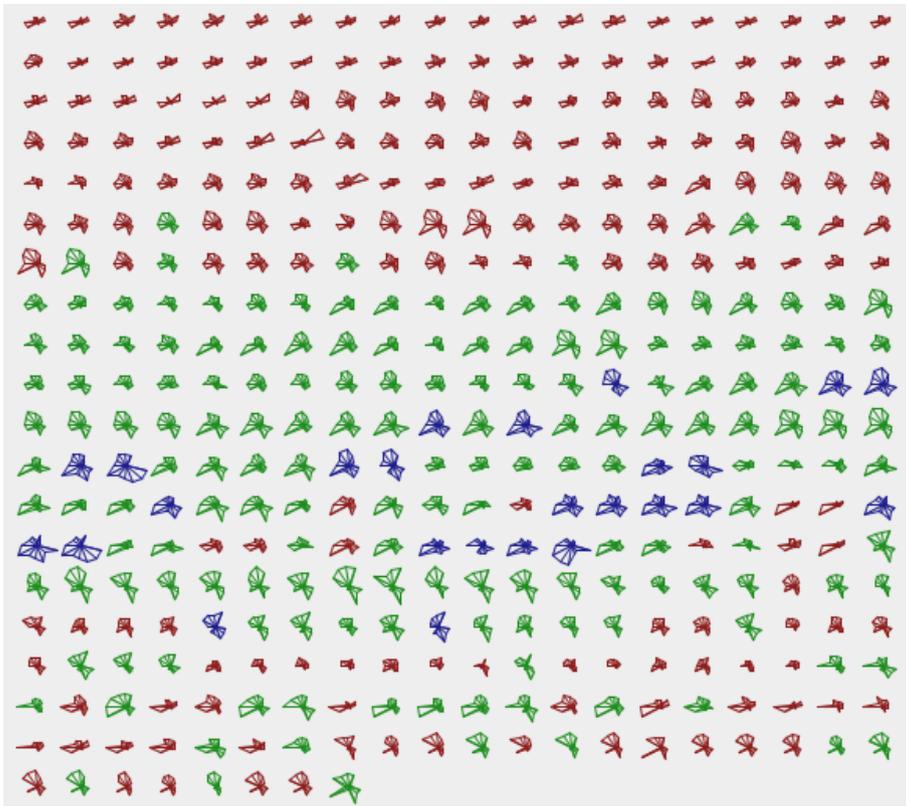
- Clustering
  - the same representation as by grouping
  - different methods:
    - kmeans
    - centroid
    - ward
    - ...



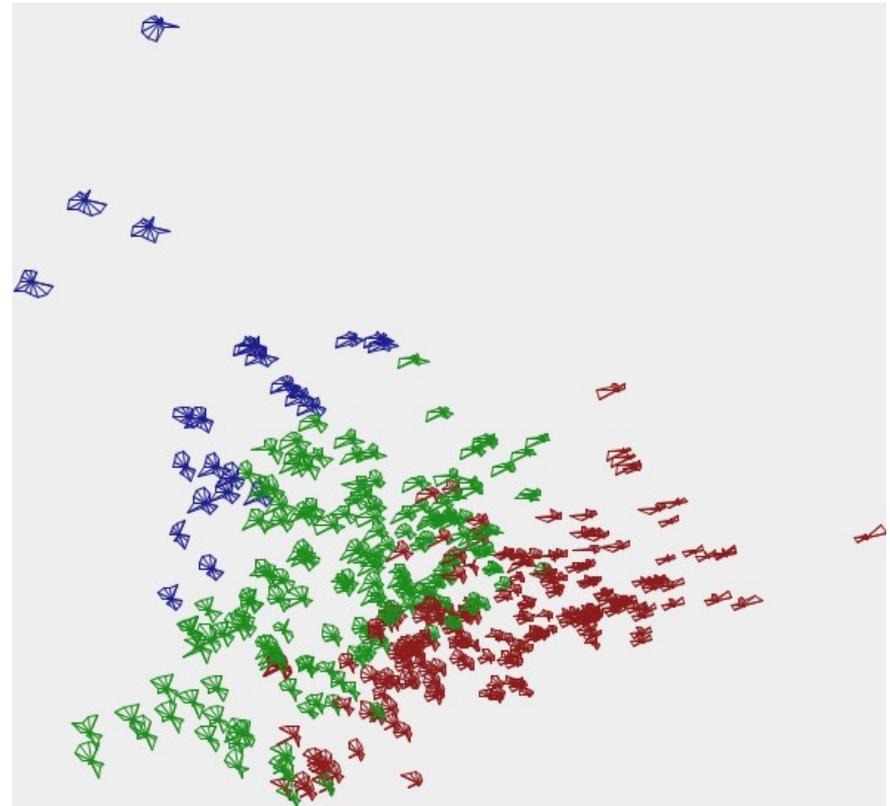
# Plots in Gauguin

- Other plots by grouping or clustering

Grid



MDS

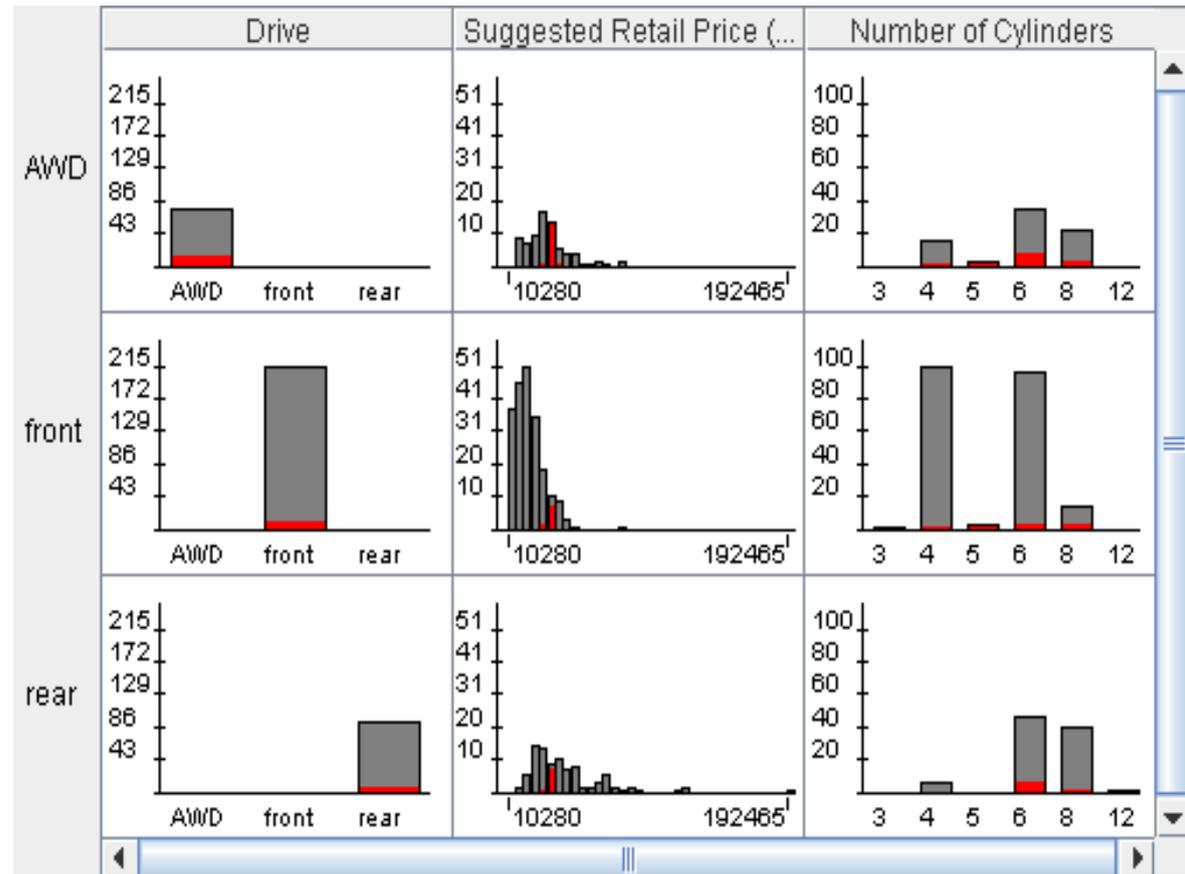


# Plots in Gauguin

- Groupsplot

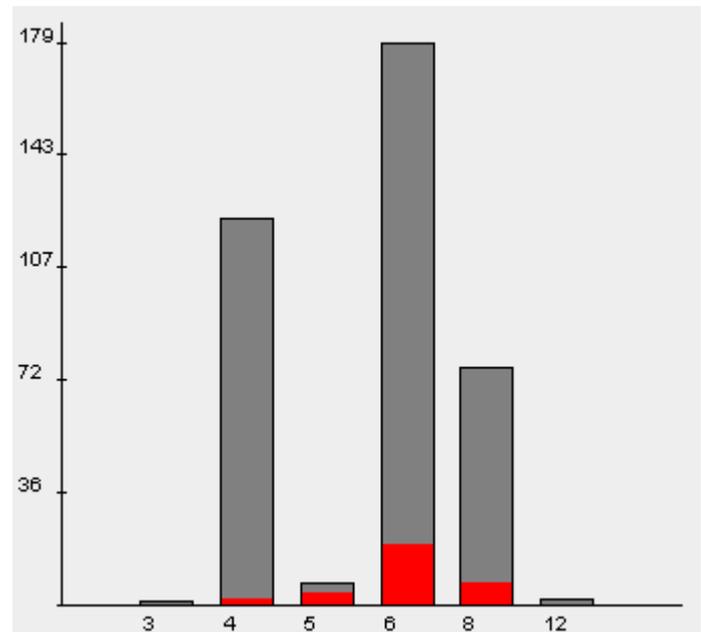
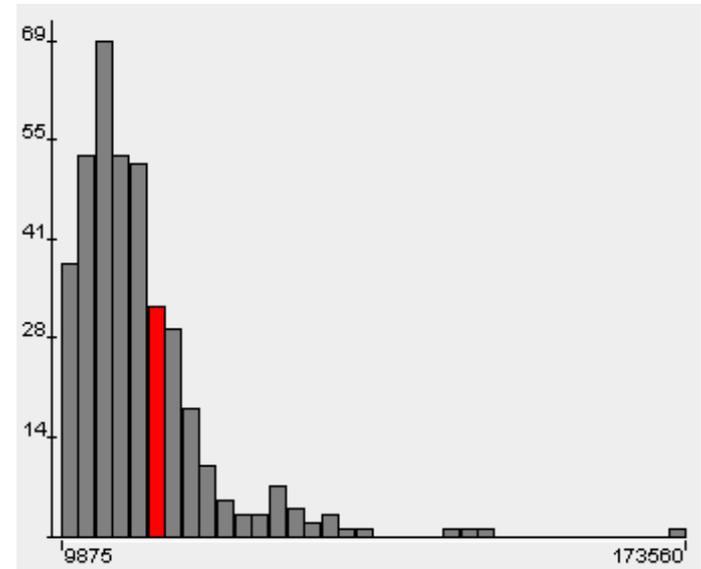
- All plots in the same column are common scaled.

- Values for the start point and bin width can be individually set for each histogram column



# Plots in Gauguin

- Histograms
  - exact values for the start point and bin width can be set via slider
  - bin width can be also easy changed by simple keyboard control
- Barcharts

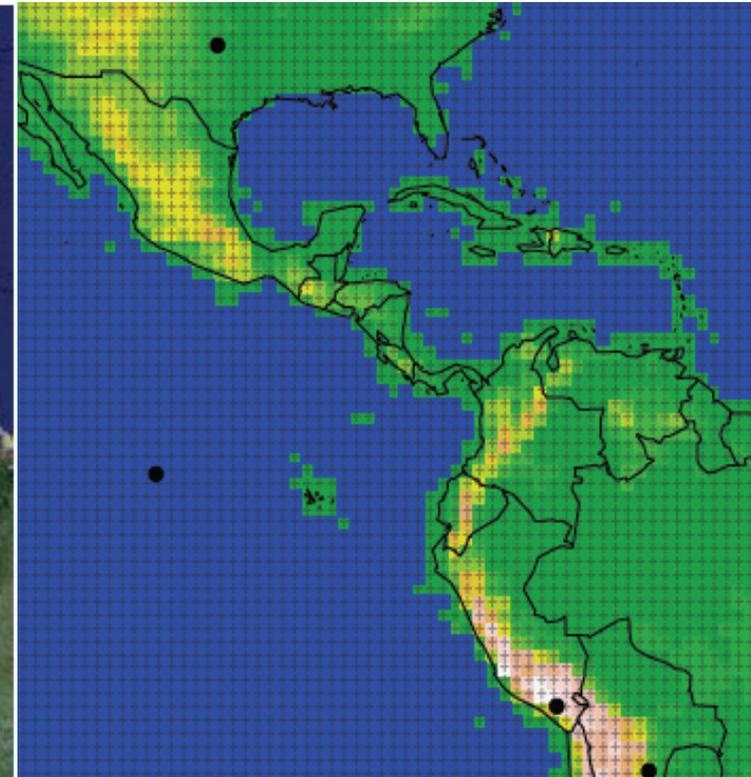


# Summary

- project for the interactive visual exploration of multivariate data sets
- cases and averages of groups and clusters are represented as glyphs
- connection to R via Rserve for calculations
- interactive and linked plots

# Data Expo 2006

- Pressure and temperature measurements over the area of Central America.



# Data Expo 2006

24 \* 24 points of observations over Central America

Variables:

- Latitude: between 55.5 W and 114.5W
- Longitude: between 36.5 N to 21.5 S
- Temperature and pressure: were taken once per month from Jan 1995 to Dec 2000
- Elevation: is reported for Jan 1998