Information Visualization

Multi-dimensional Data (Hypervariate Data, Multivariate Data)

SCHMALKALDEN
UNIVERSITY
OF APPLIED SCIENCES

Graphics+Interaction
Laboratory
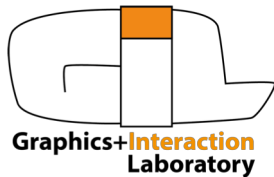
Sample Data

▸ Famous car data set

   ▸ Car models released from 1970 to 1982

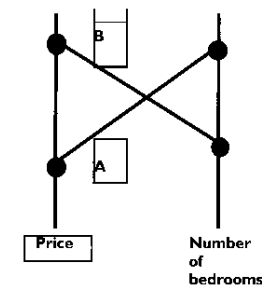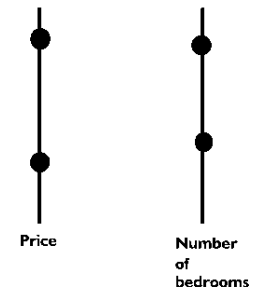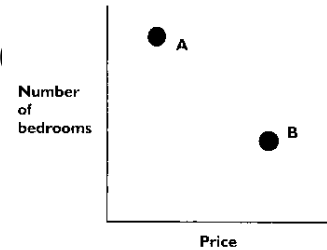|	MPG	Cylinders	Horsepower	Weight	Acceleration	Year	Origin
1	MPG	Cylinders	Horsepower	Weight	Acceleration	Year	Origin
2	18	8	130	3504	12	70	USA
3	15	8	165	3693	11.5	70	USA
4	18	8	150	3436	11	70	USA
5	16	8	150	3433	12	70	USA
6	17	8	140	3449	10.5	70	USA
7	15	8	198	4341	10	70	USA
8	14	8	220	4354	9	70	USA
9	14	8	215	4312	8.5	70	USA
10	14	8	225	4425	10	70	USA
11	15	8	190	3850	8.5	70	USA
12	15	8	170	3563	10	70	USA
13	14	8	160	3609	8	70	USA
14	15	8	150	3761	9.5	70	USA
15	14	8	225	3086	10	70	USA
16	24	4	95	2372	15	70	Europe
17	22	6	95	2833	15.5	70	USA
18	18	6	97	2774	15.5	70	USA
19	21	6	85	2587	16	70	USA

Parallel Coordinates (PC)

- N-dimensional data tuples $x_1, x_2, x_3 \ldots x_n$ depicted polyline, where the points $x_1, x_2, x_3 \ldots x_n$ are connected via n parallel Y-axes
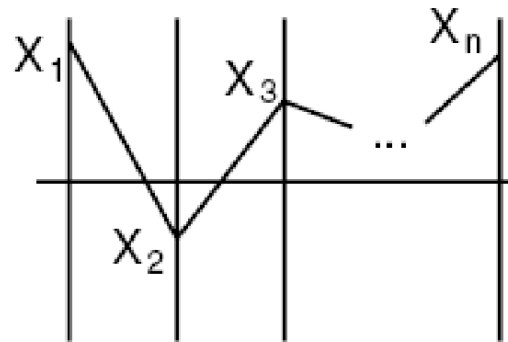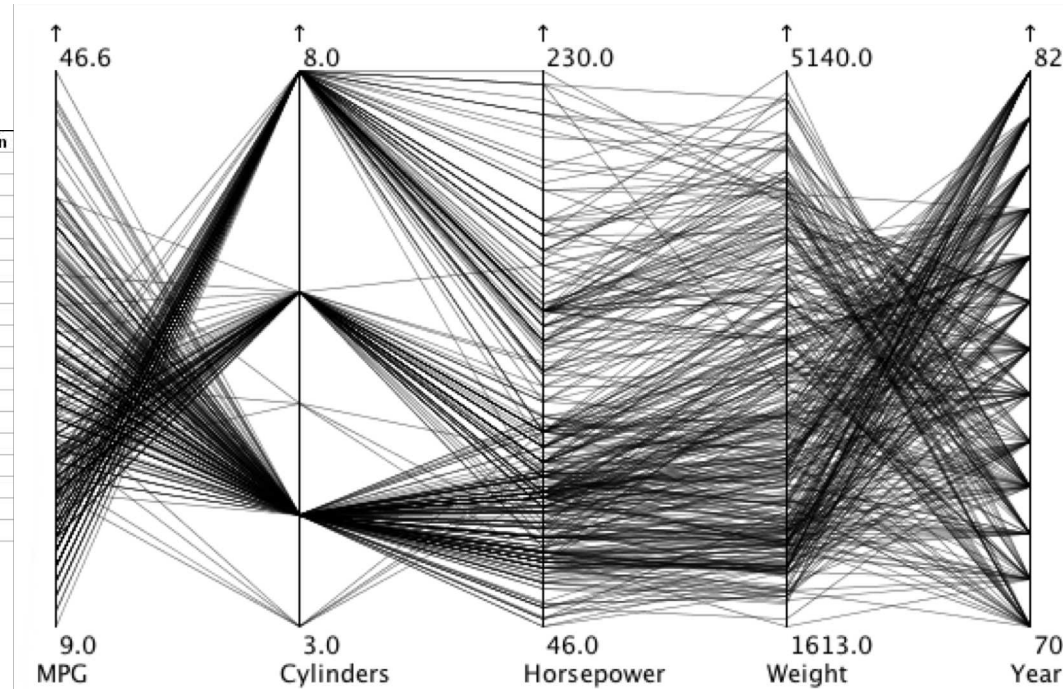


▸ Each column mapped onto virtual axis

▸ Nice intro on
http://eagereyes.org/techniques/parallel-coordinates

[Spence 2001]

Parallel Coordinates Example:
Car Dataset

	MPG	Cylinders	Horsepower	Weight	Acceleration	Year	Origin
1	MPG	Cylinders	Horsepower	Weight	Acceleration	Year	Origin
2	18	8	130	3504	12	70	USA
3	15	8	165	3693	11.5	70	USA
4	18	8	150	3436	11	70	USA
5	16	8	150	3433	12	70	USA
6	17	8	140	3449	10.5	70	USA
7	15	8	198	4341	10	70	USA
8	14	8	220	4354	9	70	USA
9	14	8	215	4312	8.5	70	USA
10	14	8	225	4425	10	70	USA
11	15	8	190	3850	8.5	70	USA
12	15	8	170	3563	10	70	USA
13	14	8	160	3609	8	70	USA
14	15	8	150	3761	9.5	70	USA
15	14	8	225	3086	10	70	USA
16	24	4	95	2372	15	70	Europe
17	22	6	95	2833	15.5	70	USA
18	18	6	97	2774	15.5	70	USA
19	21	6	85	2587	16	70	USA

ParVis – Java Implementation from 2000

Parallel Coordinates Limitations

▸ Overplotting for large data sets

  ▸ Transparency

  ▸ Sampling

    ▸ Random sampling
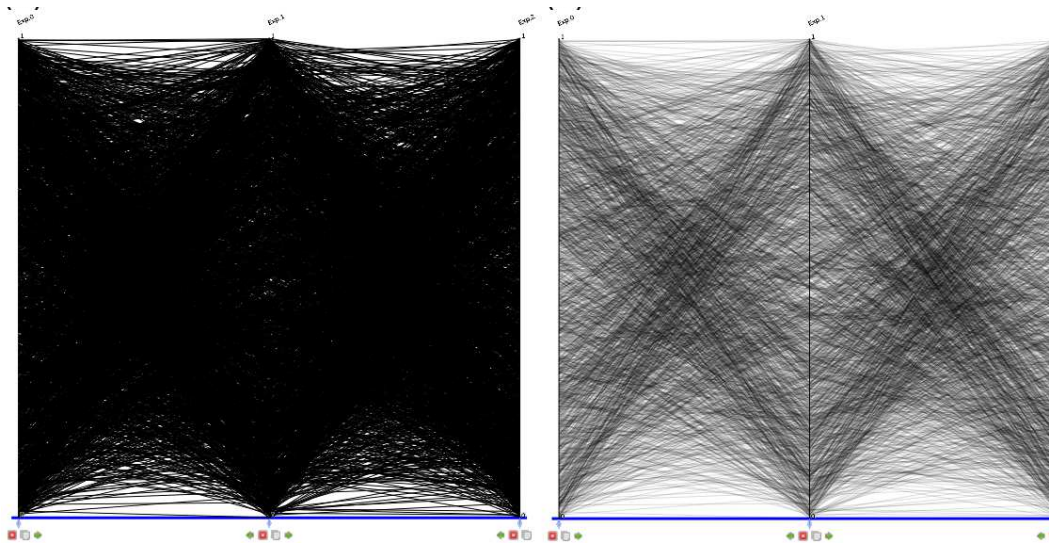
Parallel Coordinates Limitations (2)

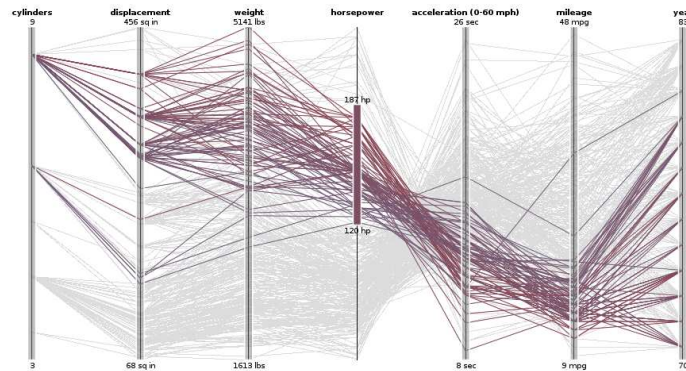▸ Only adjacent axes can be related to each other

▸ Number of axes is limited

[Yang 2003]

ProtoVis PC Implementation
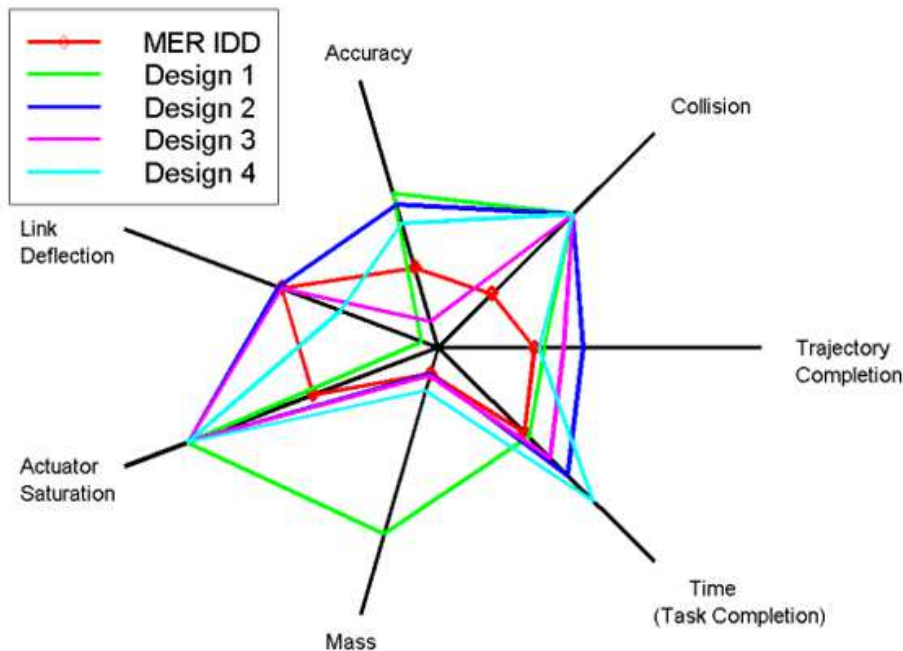
▸ Available on http://vis.stanford.edu/protovis/

▸ Java script and SVG

▸ Native web-visualization

▸ Cars dataset in ProtoVis
http://vis.stanford.edu/protovis/ex/cars.html

Star Plot

- ▸ Similar to parallel coordinates

- ▸ Radiate from a common origin

[Coekin1969]

1	Price
2	Mileage (MPG)
3	1978 Repair Record (1 = Wo
4	1977 Repair Record (1 = Wo
5	Headroom
6	Rear Seat Room
7	Trunk Space
8	Weight
9	Length

Star Plot of MER IDD and Automated Designs

http://start1.jpl.nasa.gov/caseStudies/autoTool.cfm

1979 AUTOMOBILE ANALYSIS

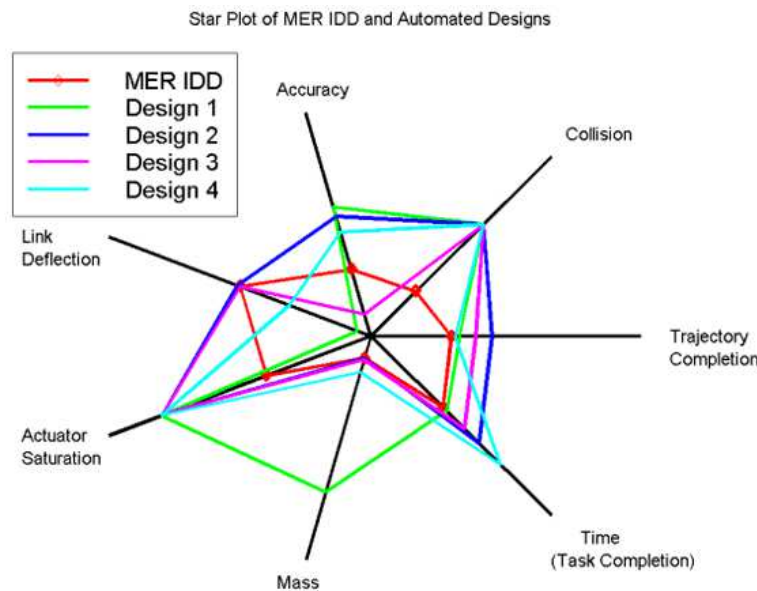http://www.itl.nist.gov/div898/handbook/eda/section3/starplot.htm
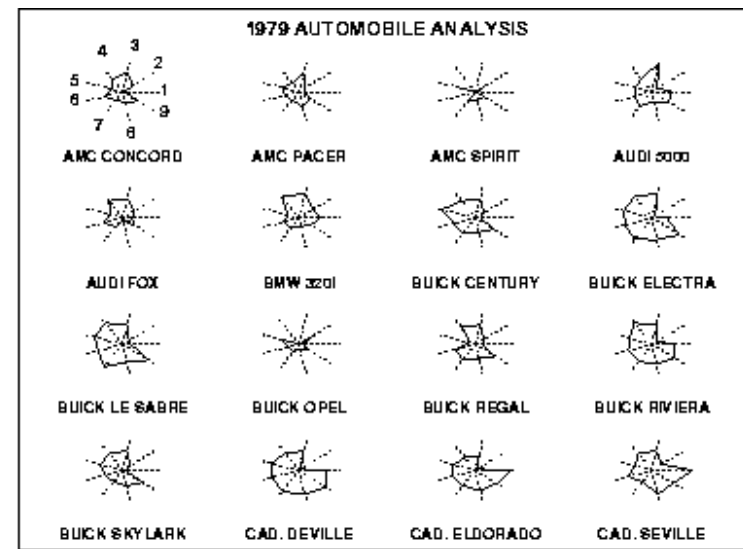
Star Plot

▸ Find similarities, find max/min values

▸ Problem: Does not scale well

Star Plot of MER IDD and Automated Designs

1979 AUTOMOBILE ANALYSIS

http://start1.jpl.nasa.gov/caseStudies/autoTool.cfm

http://www.itl.nist.gov/div898/handbook/eda/section3/starplot.htm
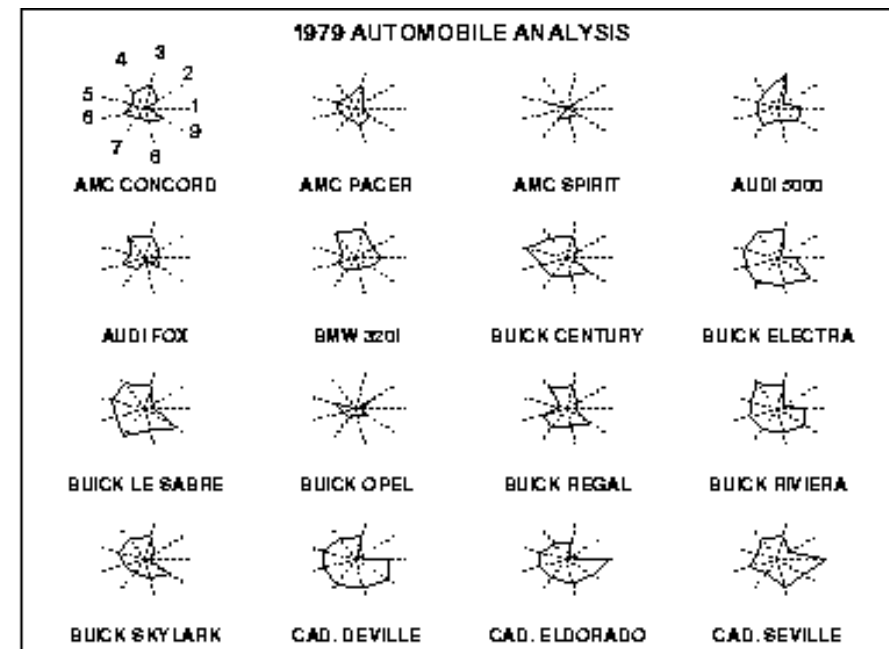
Visualizing Complex Data
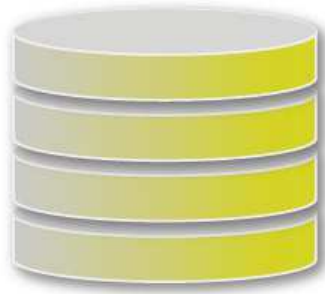
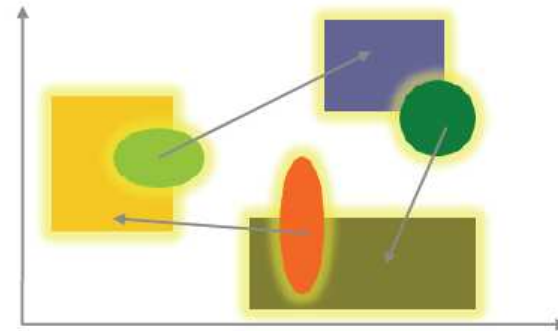**Complex** visualization
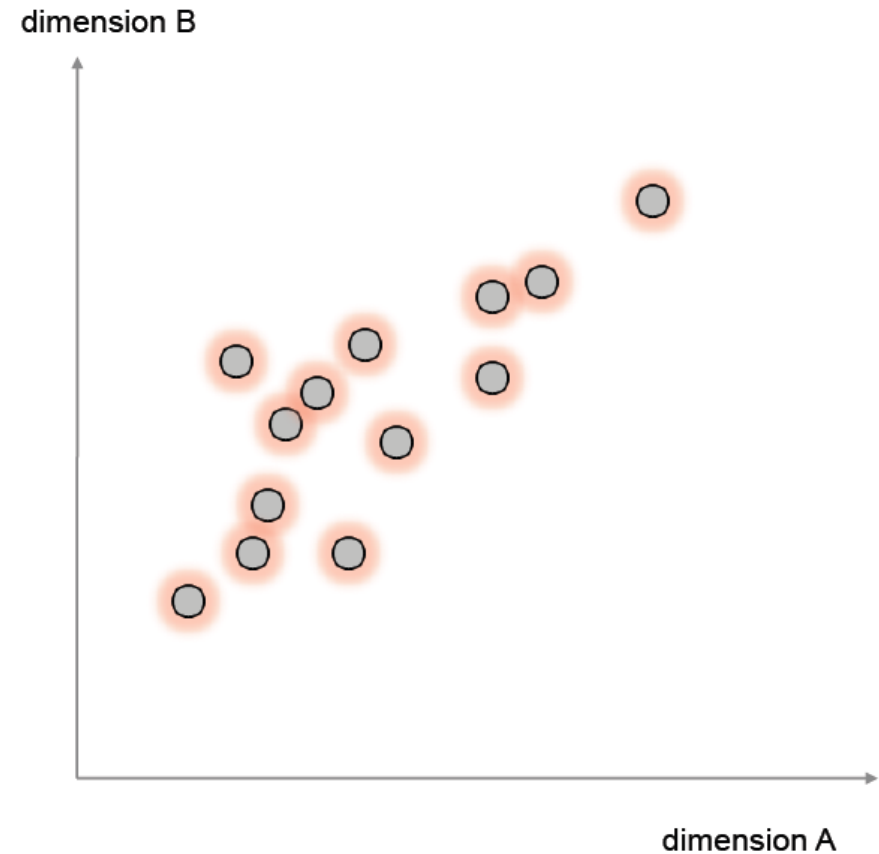
Complex dataset
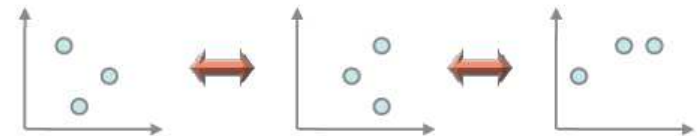
Many simple visualizations

Scatterplots

- Assign dimensions to graphical axes
  - Two (typical) or three
- Data cases as points in the space defined by axes
- Additional dimensions
  - Point color
  - Point size
  - Point shape
  - …
- **Limited** number of displayed dimensions!

dimension B

dimension A

Problem

- Many (most?) interesting real-world datasets have **many** dimensions
    - Certainly more than a scatterplot can show!
- Two possible solutions:
    1. More complex visual representation
        - Parallel coordinates, DataMeadow, etc…
        - Lose simplicity of scatterplots
    2. Create many scatterplots (one per combination of dimensions)
   - But how to visualize them?

Scatterplot Matrices

▸ Create matrix of all data dimensions

  ▸ Each column/row is a data dimension

  ▸ Each cell in the matrix is a scatter plot

  ▸ Reordering of columns/rows (manual/automatic)

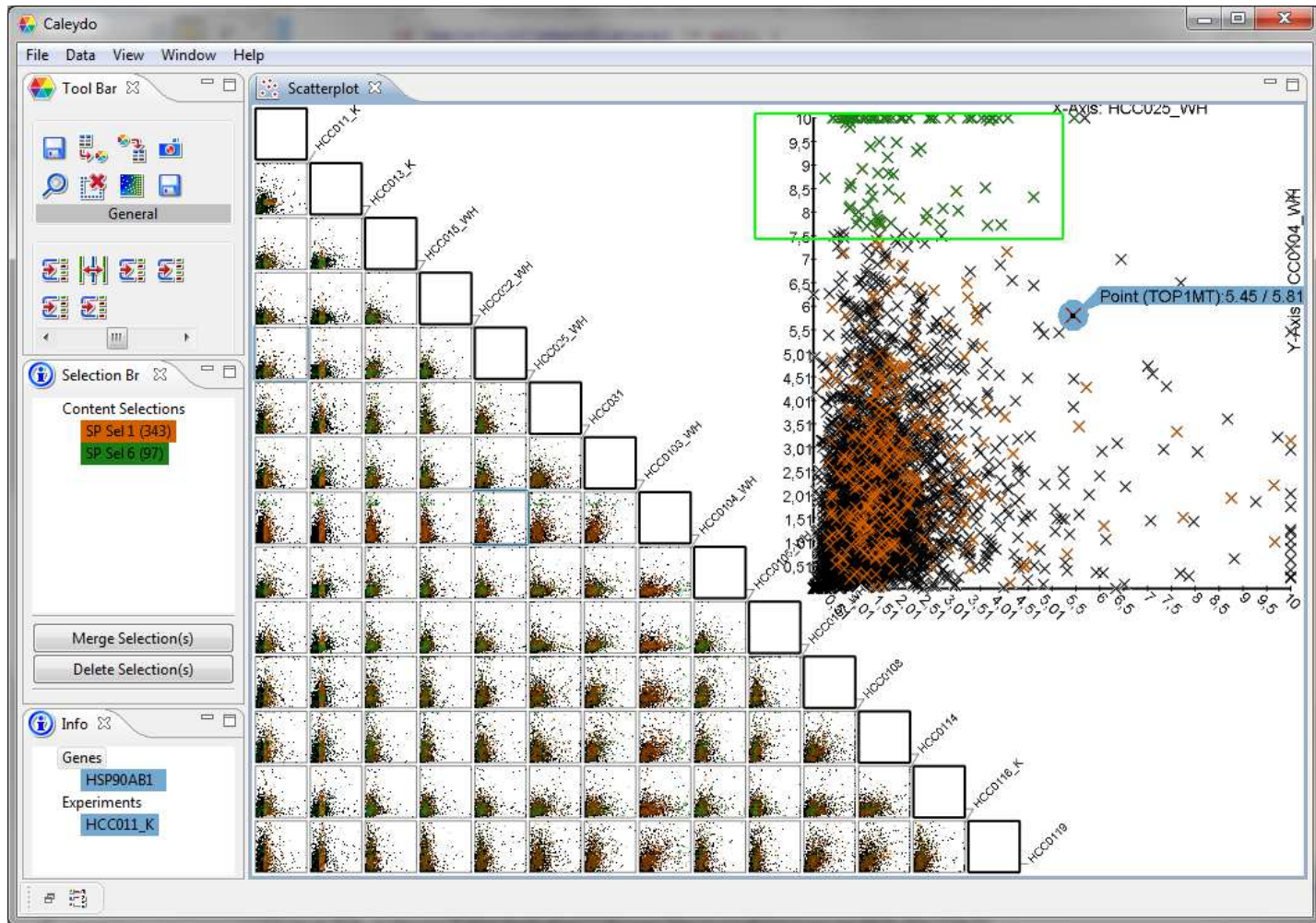▸ Complex data can be visualized through a sequence of simple visualizations

Example: Scatterplot Matrix

▸ Find pairwise relationships
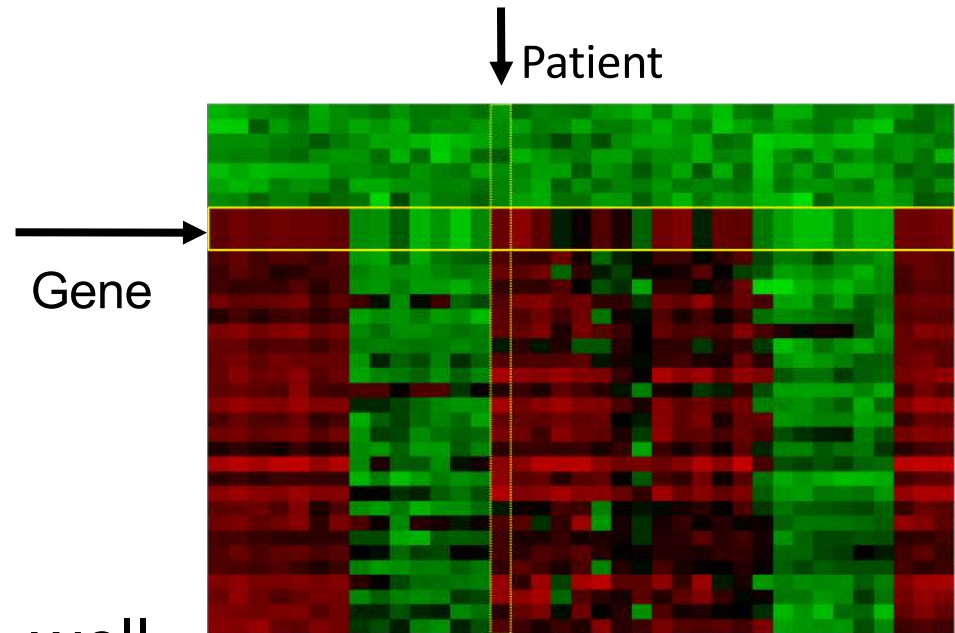
▸ Clustering in groups of data

Prof. Hartmut Seichter, PhD | seichter@fh-sm.de | Schmalkalden University of Applied Sciences
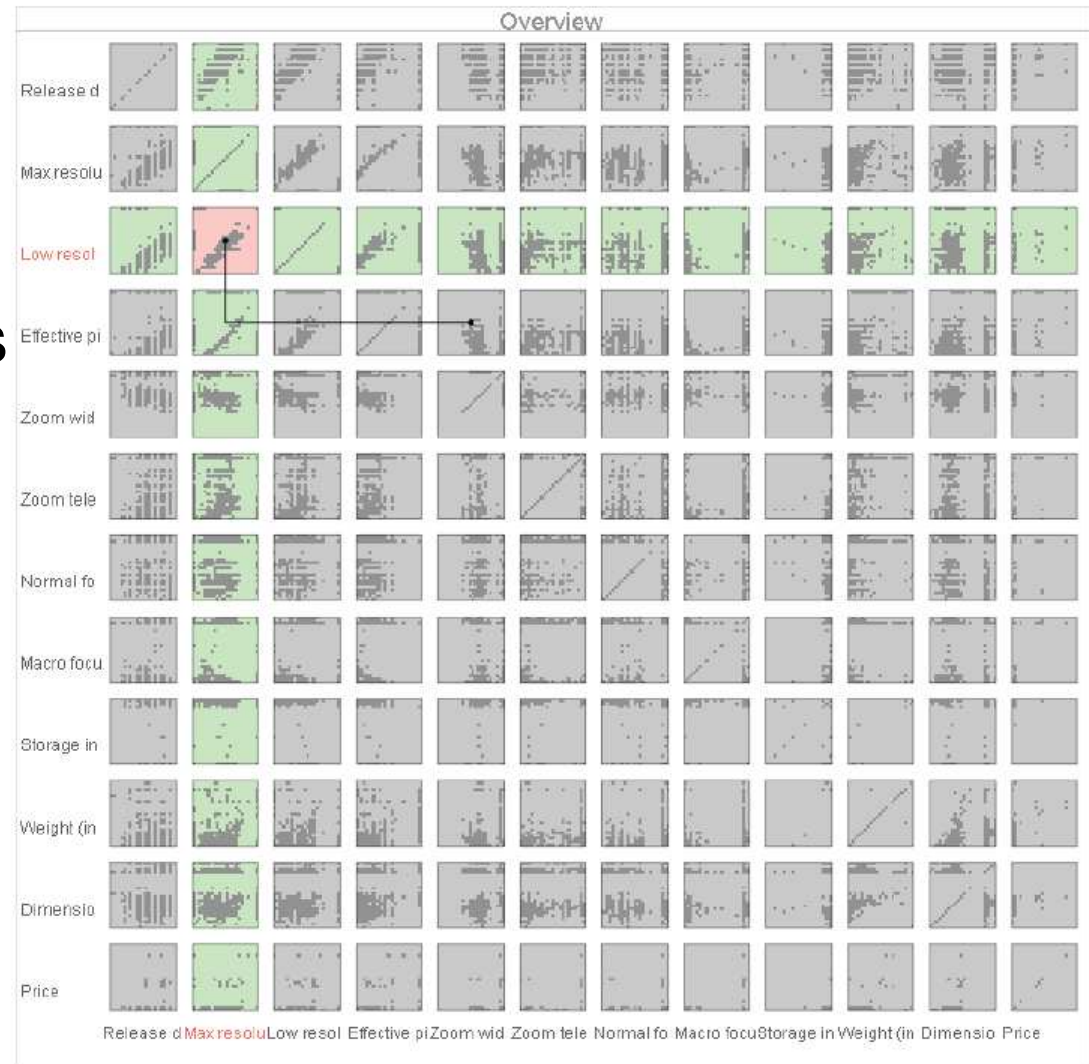
Scatterplot Matrix in Caleydo

Heat Map

▸ Common representation

▸ Color coded regulation

▸ Filtering and Clustering

Patient

Gene

▸ Problem: does not scale well

▸ Solution: hierarchical approach

Focus and Context Heat Map

[Lex, PacificVis 2010]