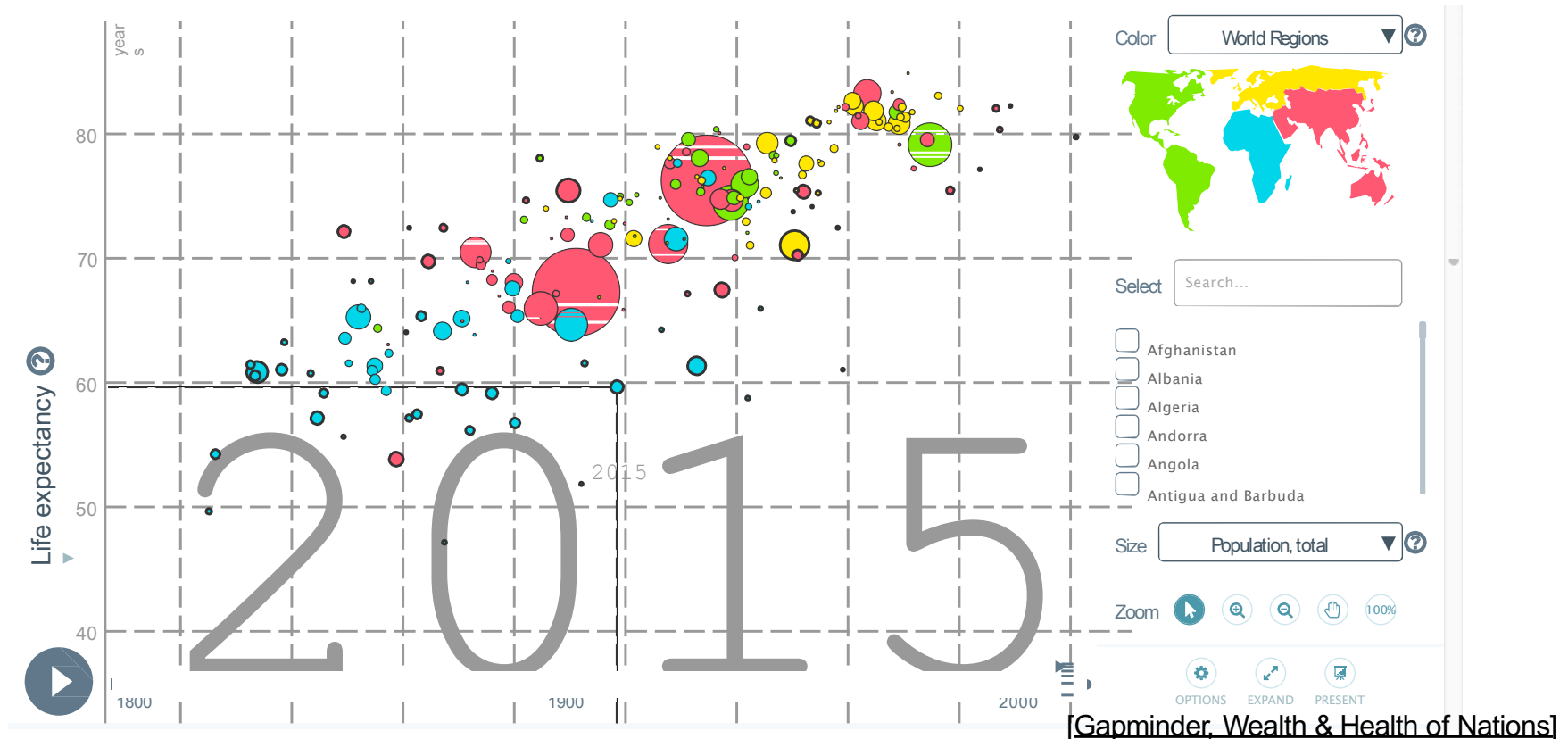# Exploratory Data Analysis (cont..) + *Visualization of* Multidimensional Data

DATA ANALYSIS & VISUALIZATION

FALL 2021

Dr. Muhammad Faisal Cheema

FAST NU

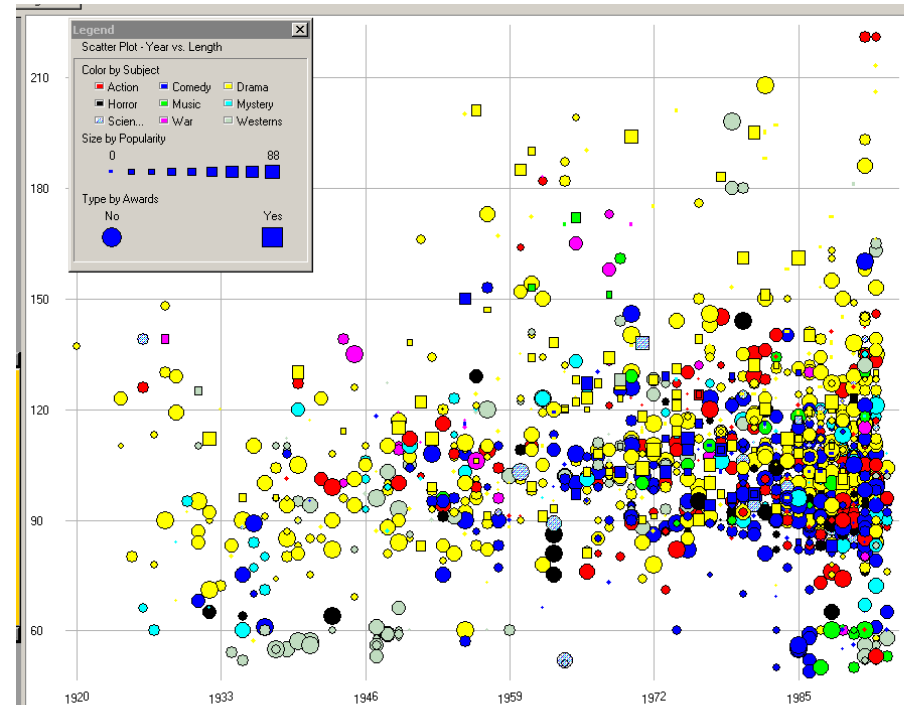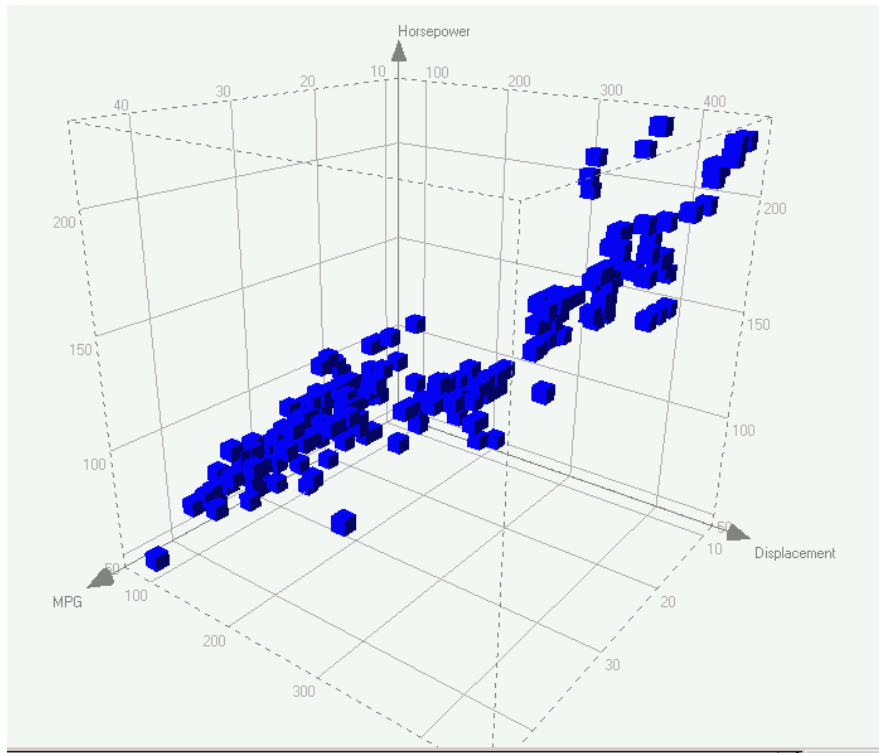De we know any way (visualizations) to represent Hypervariate/Multivariate/Multi-Dimensional data?

YES: More dimensions as Visual Channels in 2D graphs

[Gapminder, Wealth & Health of Nations]

more examples..
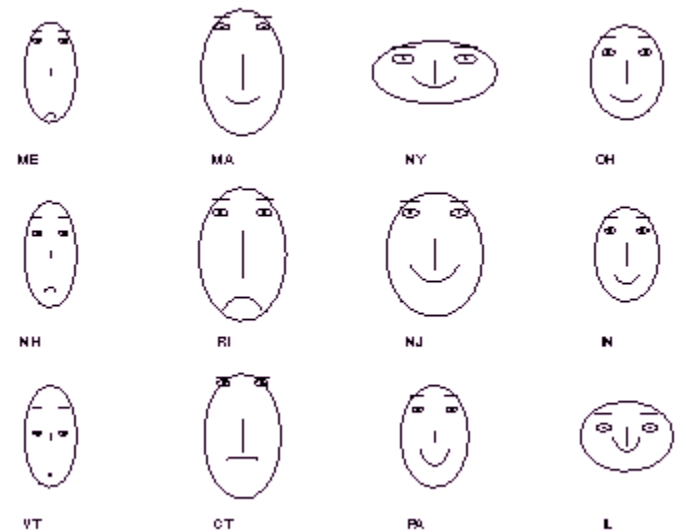
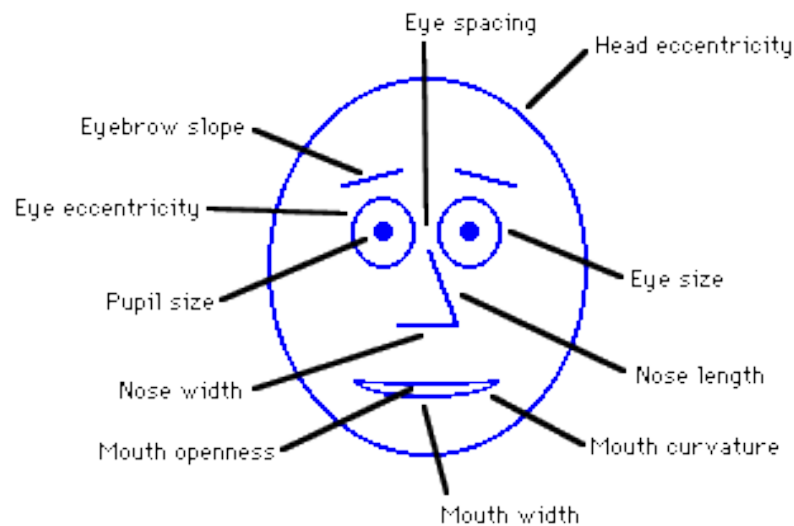3D scatterplot, spin plot
2D plot + size (or color…)

But there are other ways to show Multi-dimensional data too

Glyphs

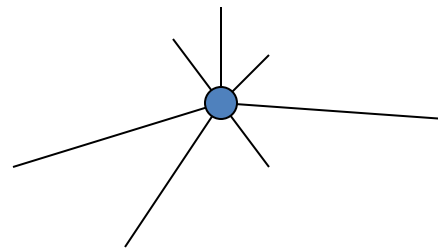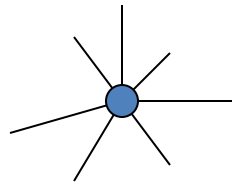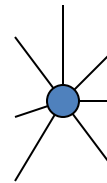# Glyphs: Chernoff Faces

Encode different variables' values in characteristics of human face

Glyphs: Stars
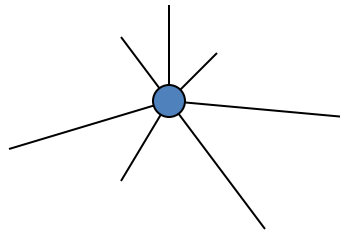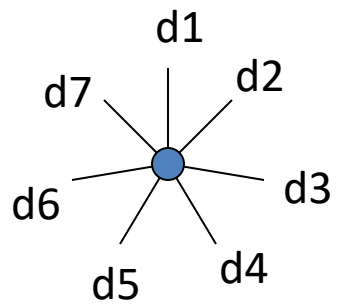
Multiple views of different dimensions

Multiple Views

Give each variable its own display

	A	B	C	D	E
1	4	1	8	3	5
2	6	3	4	2	1
3	5	7	2	4	3
4	2	6	3	1	5

1

2

3

4

A B C D E
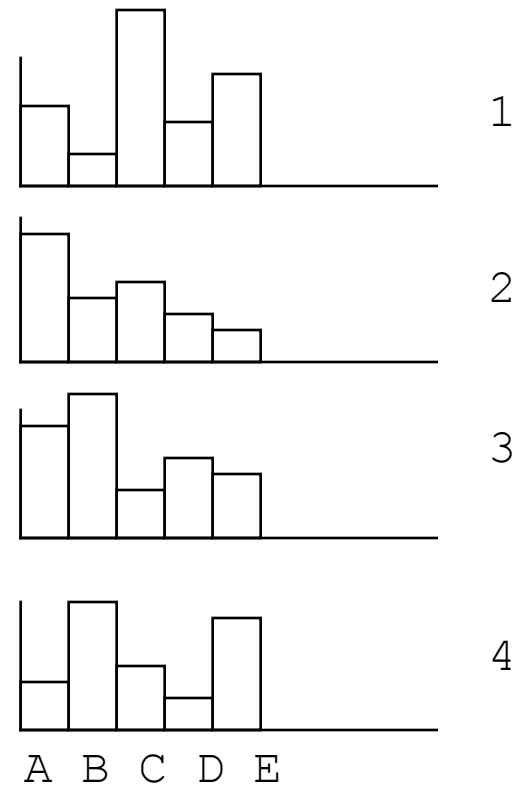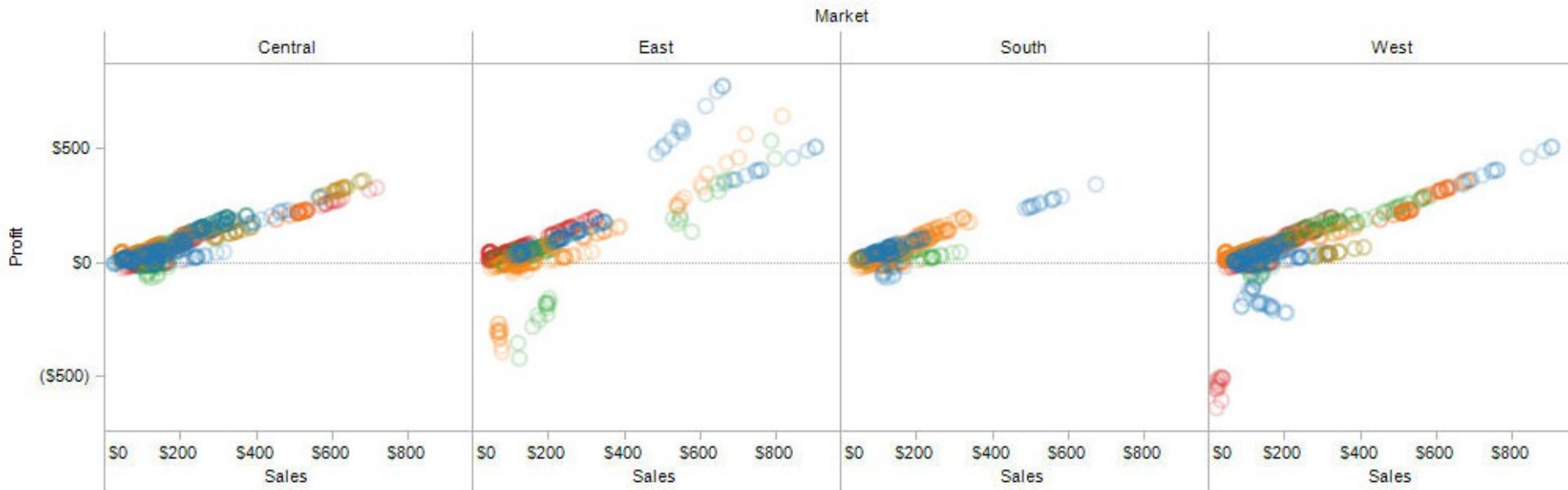
Trellis Plots

A *trellis plot* subdivides space to enable comparison across multiple plots.

Typically nominal or ordinal variables are used as dimensions for subdivision.

Example: Simple Plot
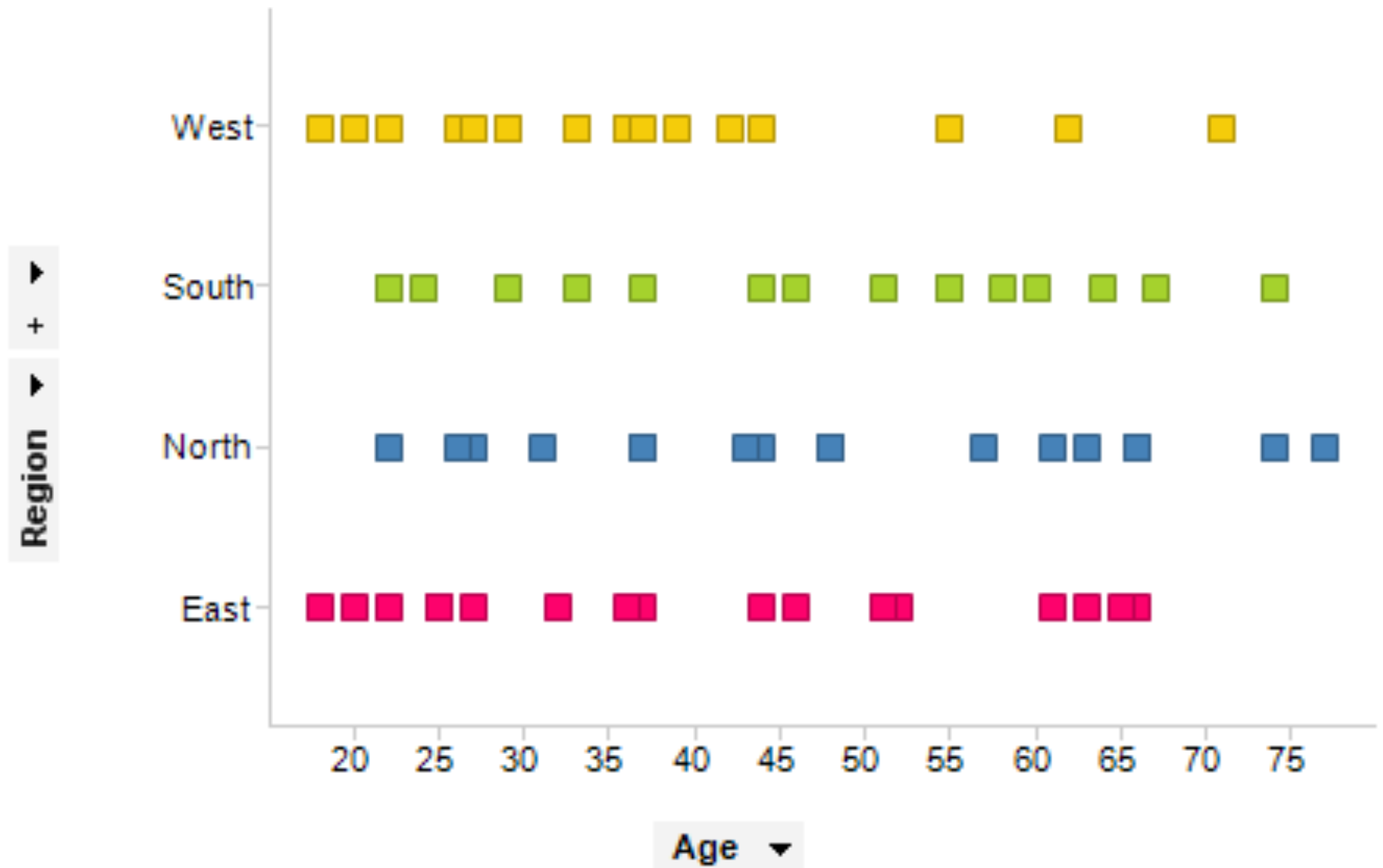
Example: Trellis Plots
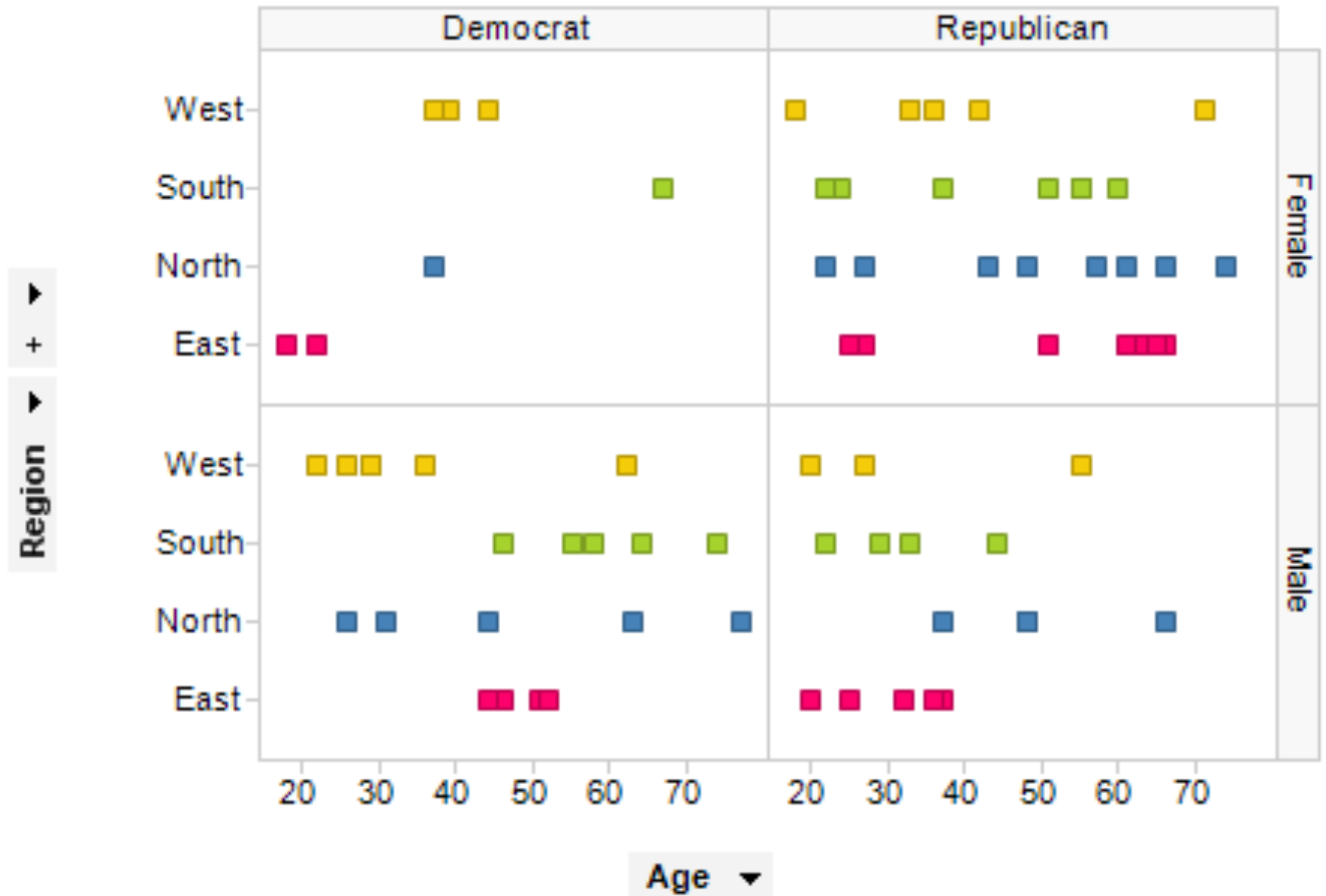
Small Multiples

[MacEachren '95, Figure 2.11, p. 38]

Small Multiples
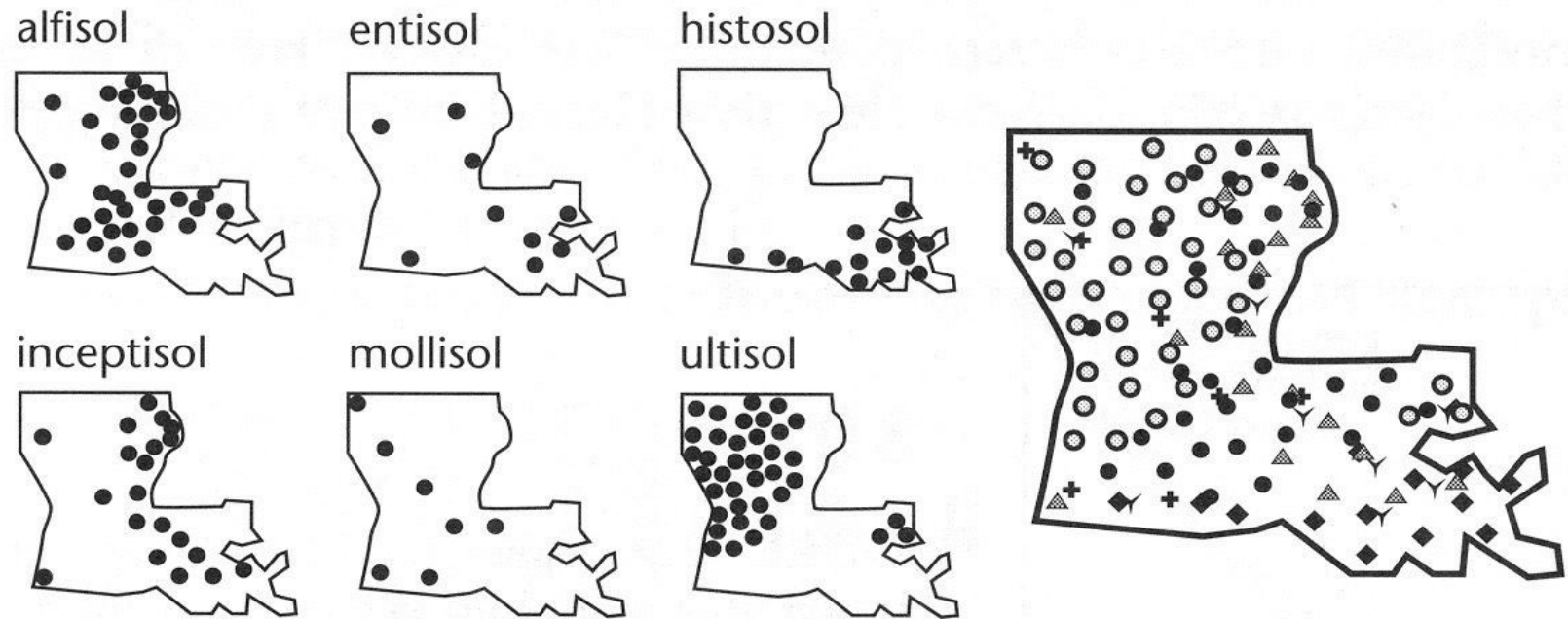
[MacEachren '95, Figure 2.11, p. 38]

Example: Small Multiples

Agriculture

Construction

Finance

Information

Manufacturing

Other

Transportation and Utilities

Business services

Education and Health

Government

Leisure and hospitality

Mining and Extraction

Self-employed

Wholesale and Retail Trade
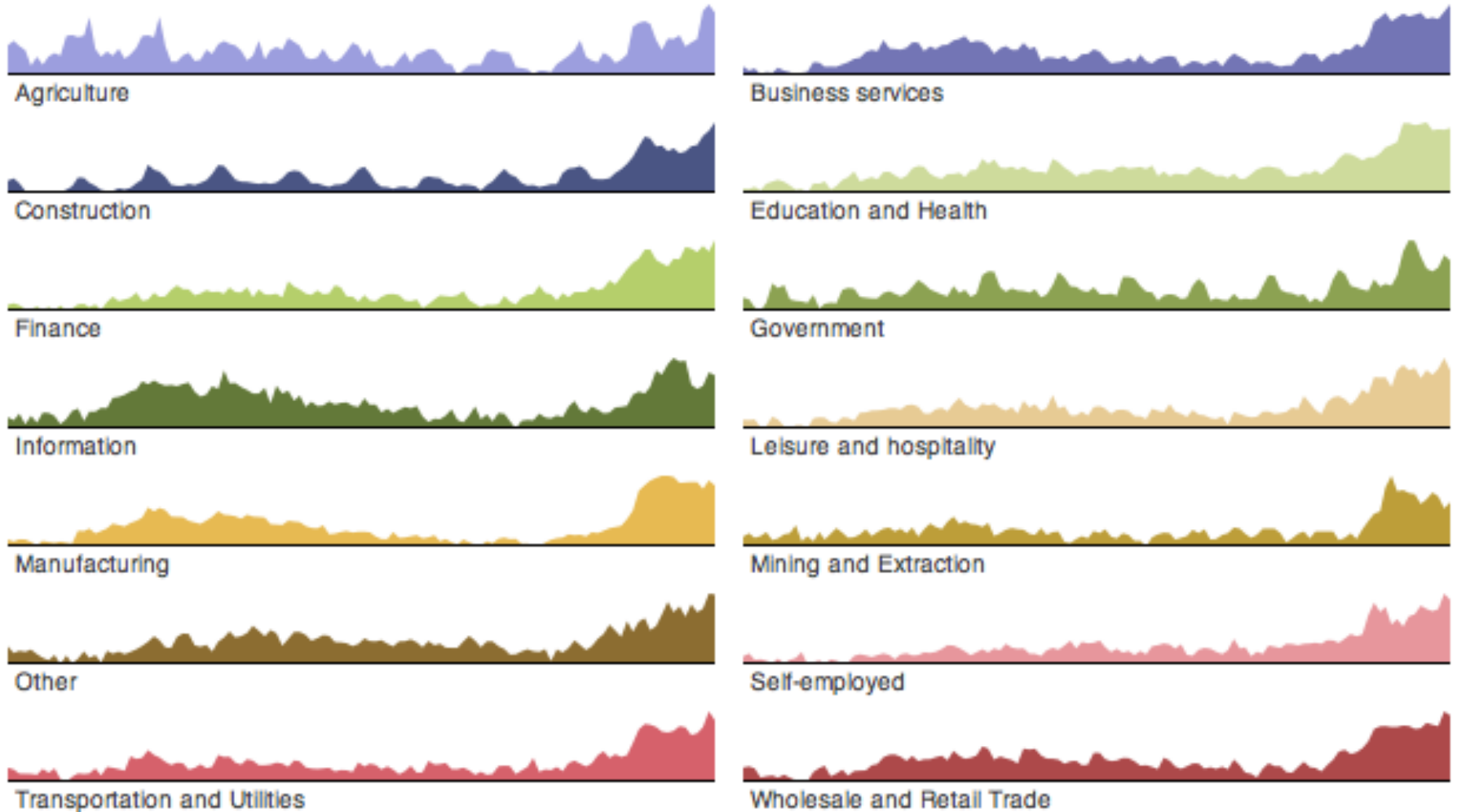
Example: Small Multiples
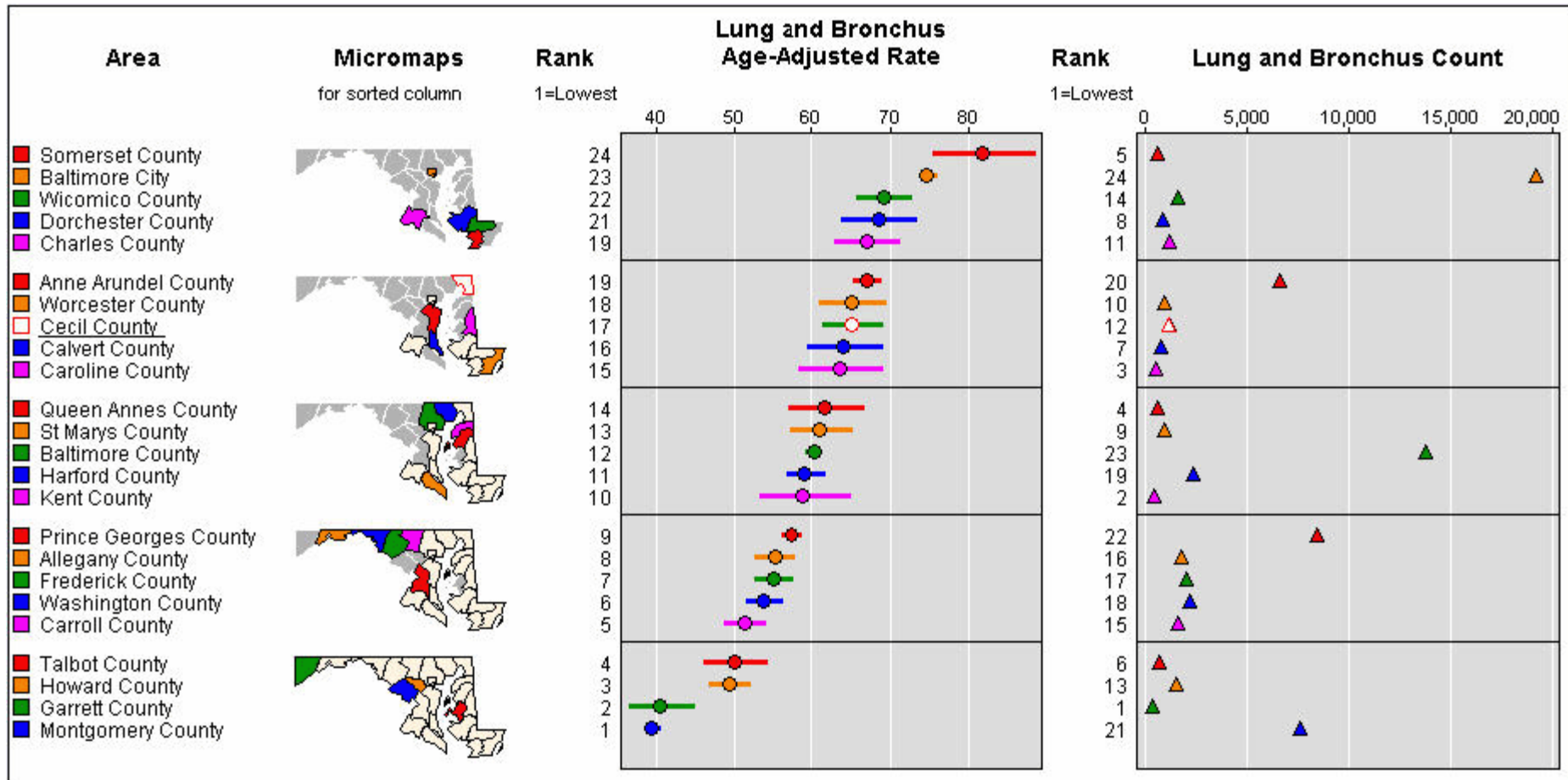
How are small multiples and Trellis plots different?

How are small multiples and Trellis plots different?

Both are same:
> *series of similar graphs or charts using the same scale and axes,*
> *easy comparisons*
> *It uses multiple views to show different partitions of a dataset.*

Scatterplot Matrix (SPLOM)
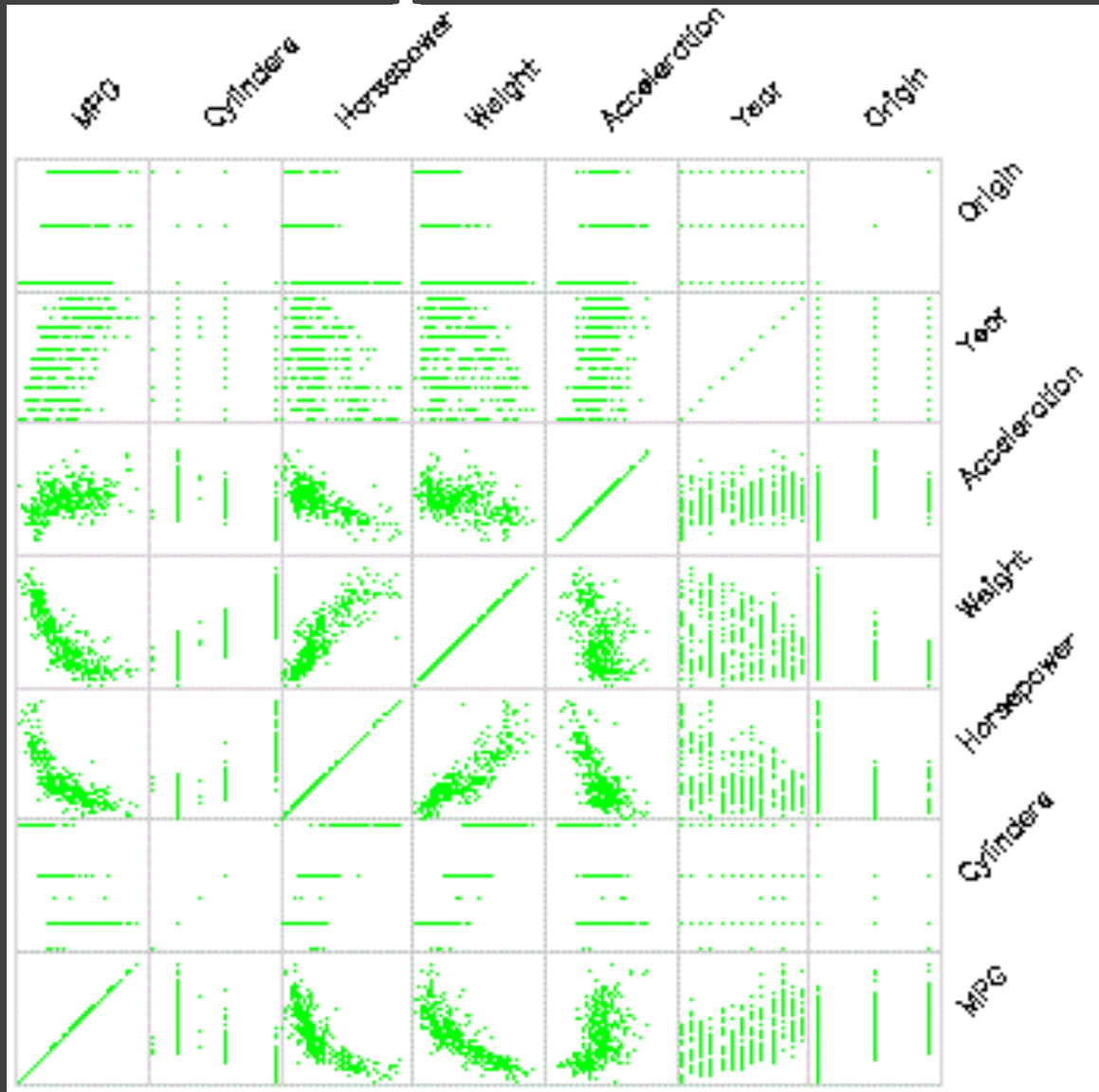
Scatter plots for pairwise comparison of each data dimension.

Scatterplot Matrix (SPLOM)

- Data: Many quantitative attributes
- Derived Data: names of attributes
- Task: Find correlations, trends, outliers
- How: Scatterplots in matrix alignment

- Visualizations in a visualization: at high level, marks are themselves visualizations…

Iris Data (red=setosa,green=versicolor,blue=virginica)

[Wikipedia]
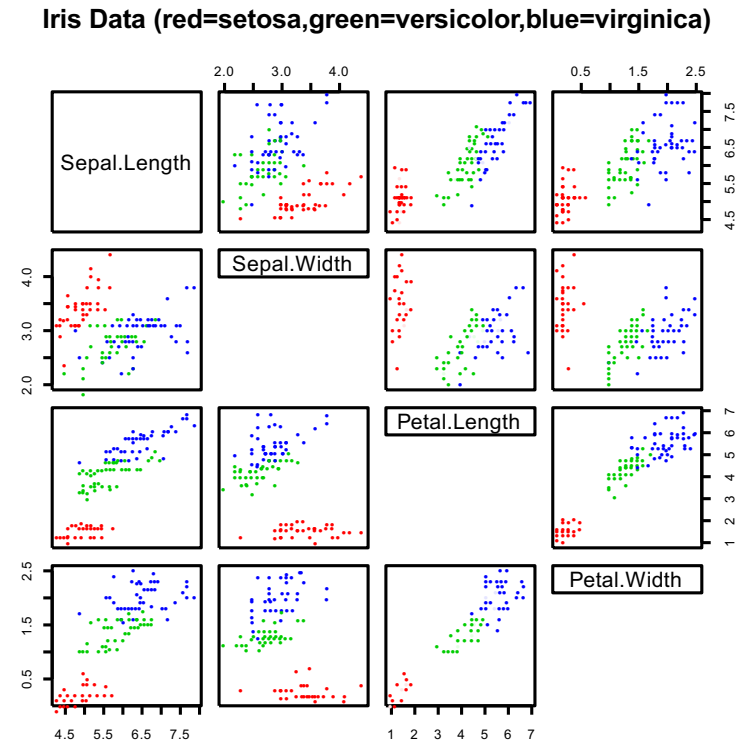
Non-orthogonal axis

Spatial Axis Orientation

- So far, we have seen the vertical and horizontal axes (a rectilinear layout) used to encode almost everything

- What other possibilities are there for axes?

Spatial Axis Orientation

- So far, we have seen the vertical and horizontal axes (a rectilinear layout) used to encode almost everything

- What other possibilities are there for axes?

  - Parallel axes

[Munzner (ill. Maguire), 2014]

Spatial Axis Orientation

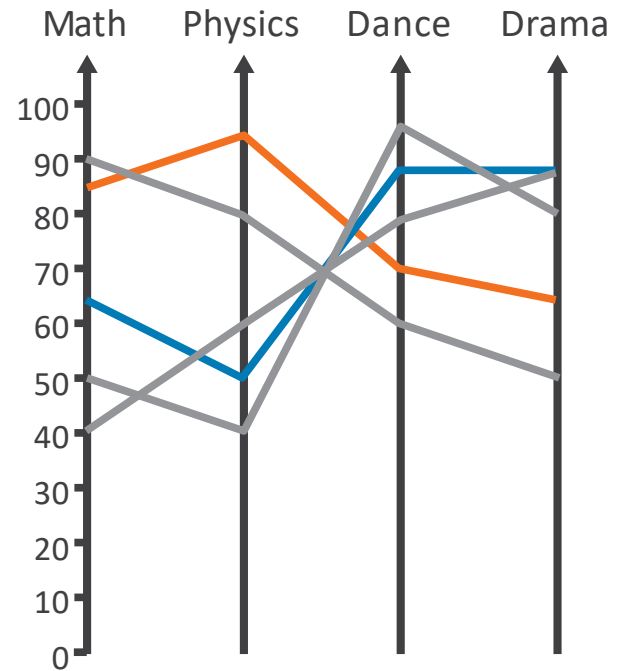- So far, we have seen the vertical and horizontal axes (a rectilinear layout) used to encode almost everything

- What other possibilities are there for axes?

  - Parallel axes

  - Radial axes

    - Also known as **Polar Coordinates** (angle + position along the line at that angle)
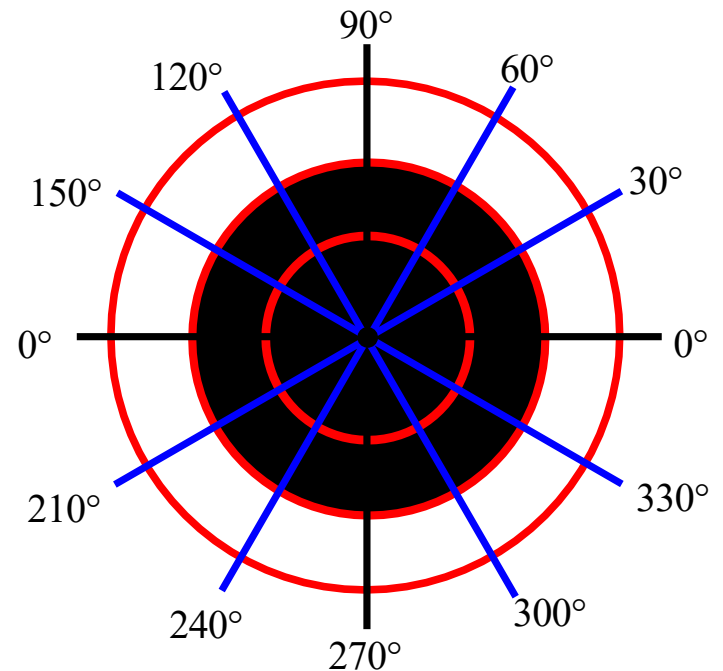
[Munzner (ill. Maguire), 2014]

Parallel Coordinates

Parallel Coordinates (4D)

Forget about Cartesian orthogonal axes
$(0,1,-1,2)=$

Parallel Coordinates (2D)

- Encode variables along a horizontal row
- Vertical line specifies values
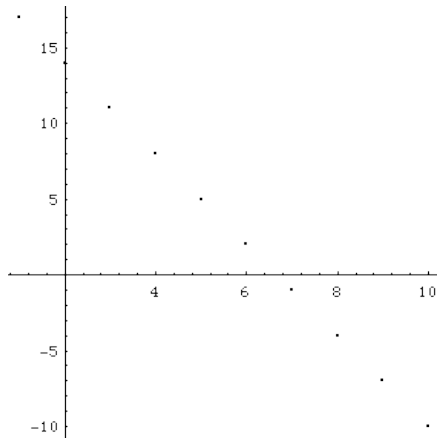
Dataset in a Cartesian graph
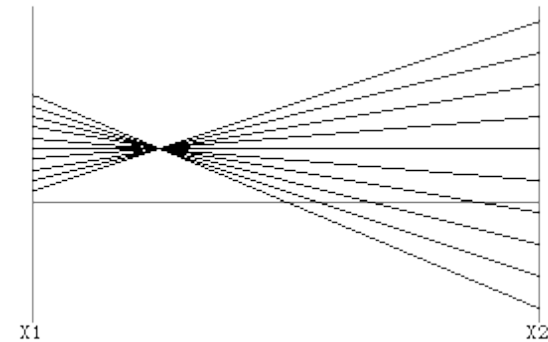
Same dataset in parallel coordinates

Parallel Coordinates Example
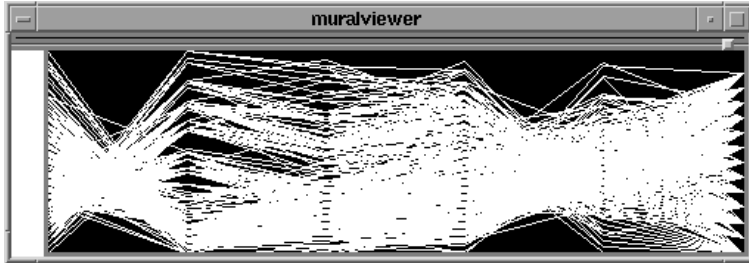
Basic
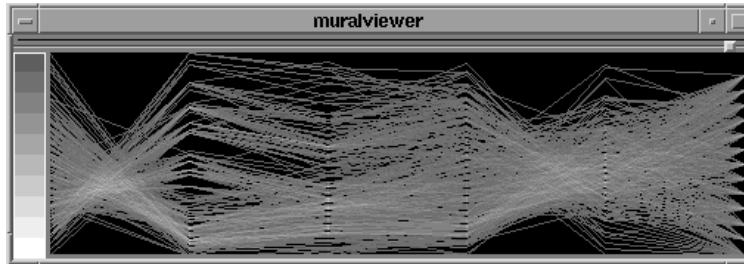
Grayscale
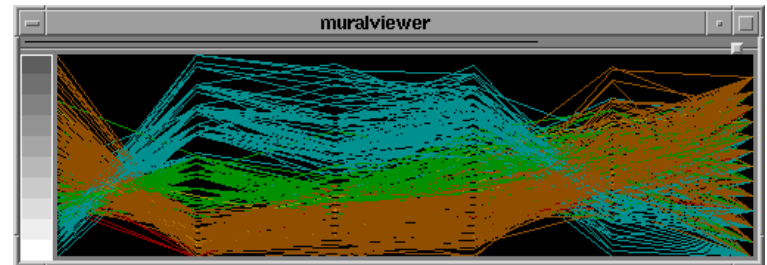
Color

Parallel Coordinates [Inselberg]
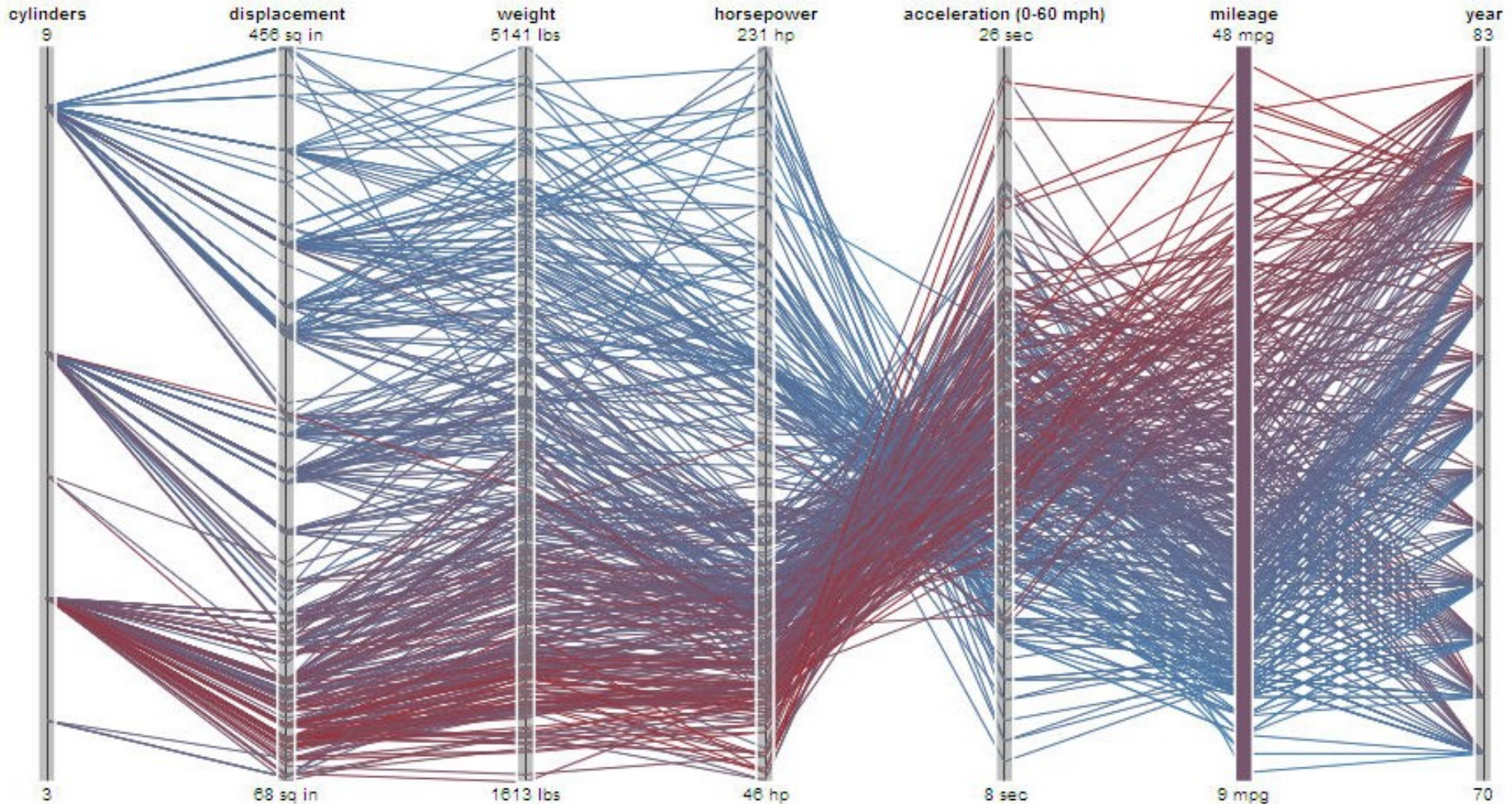
Parallel Coordinates [Inselberg]

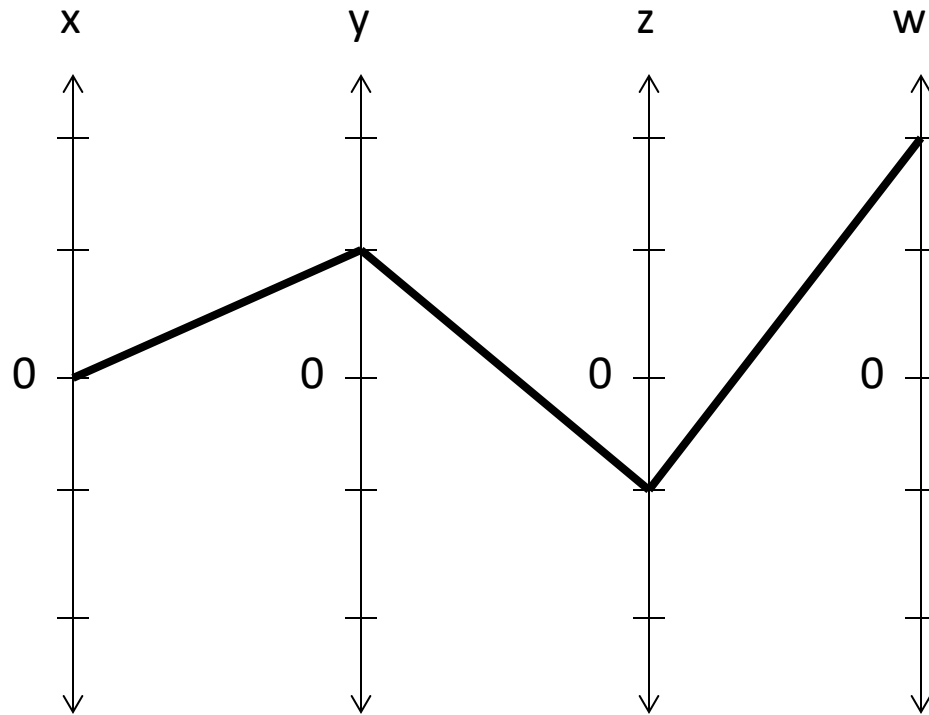Visualize up to ~two dozen dimensions at once

1. Draw parallel axes for each variable

2. For each tuple, connect points on each axis

Between adjacent axes: line crossings imply neg. correlation, shared slopes imply pos. correlation.

Full plot can be cluttered. **Interactive selection** can be used to assess multivariate relationships.

Highly sensitive to axis **scale** and **ordering**.

Expertise required to use effectively!

Radar Plot / Star Graph
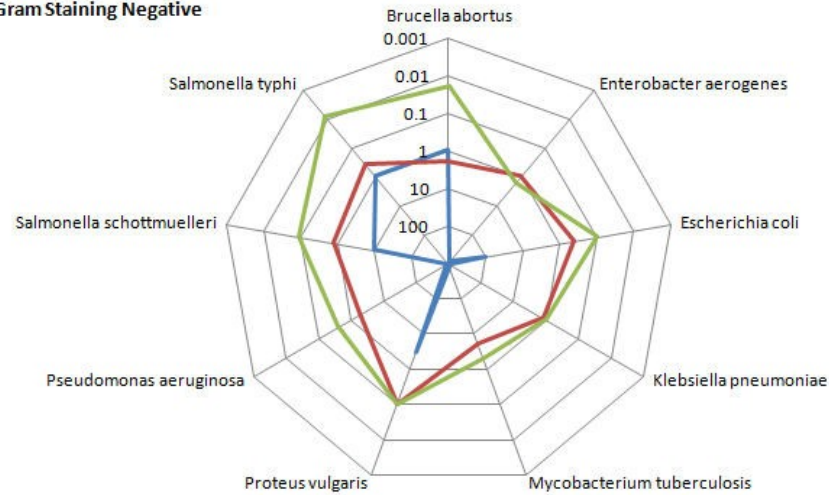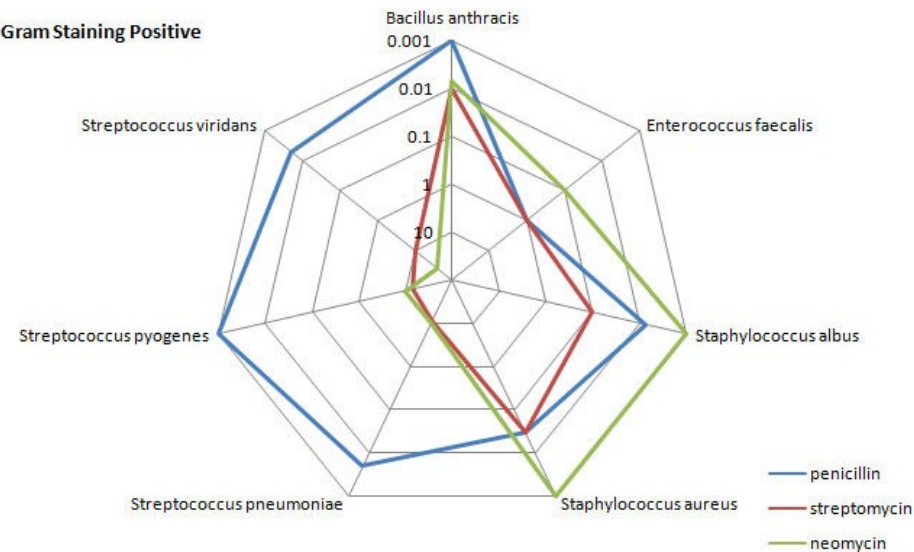
"Parallel" dimensions in polar coordinate space

Best if same units apply to each axis

Radar Plot / Star Graph
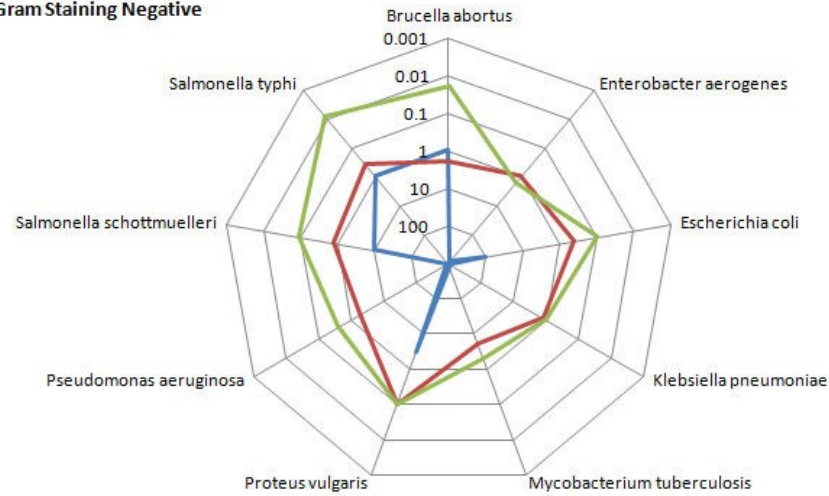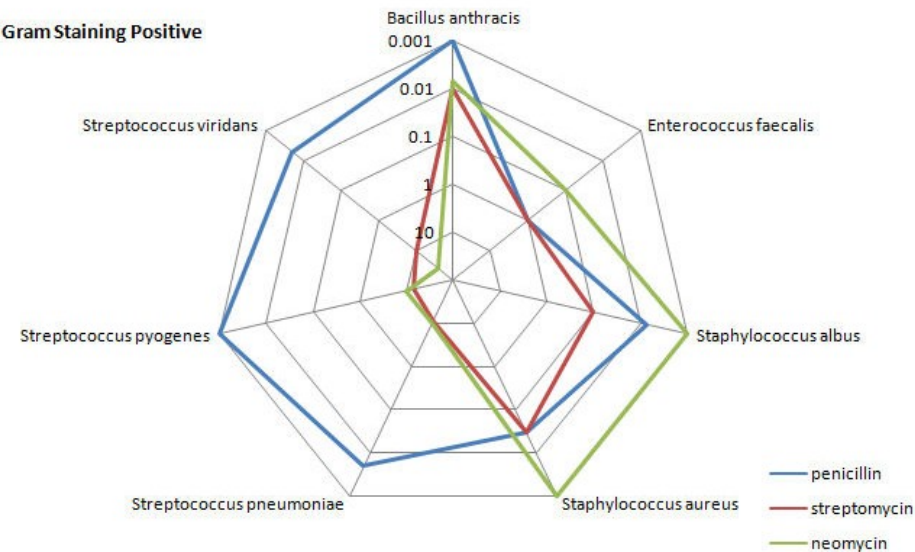
"Parallel" dimensions in polar coordinate space

Best if same units apply to each axis

Chord Diagram

Close-Ups of the Genome, Species by Species by Species

Scientists are sequencing the genomes of more than 70 organisms. The availability of these sequences has given rise to the field of comparative genomics, which seeks to answer questions about one animal's genome using information derived from another. A Canadian genomics scientist, Martin Krzywinski, has created a computer program called Circos that aids in visualizing and comparing the data. The large diagram below illustr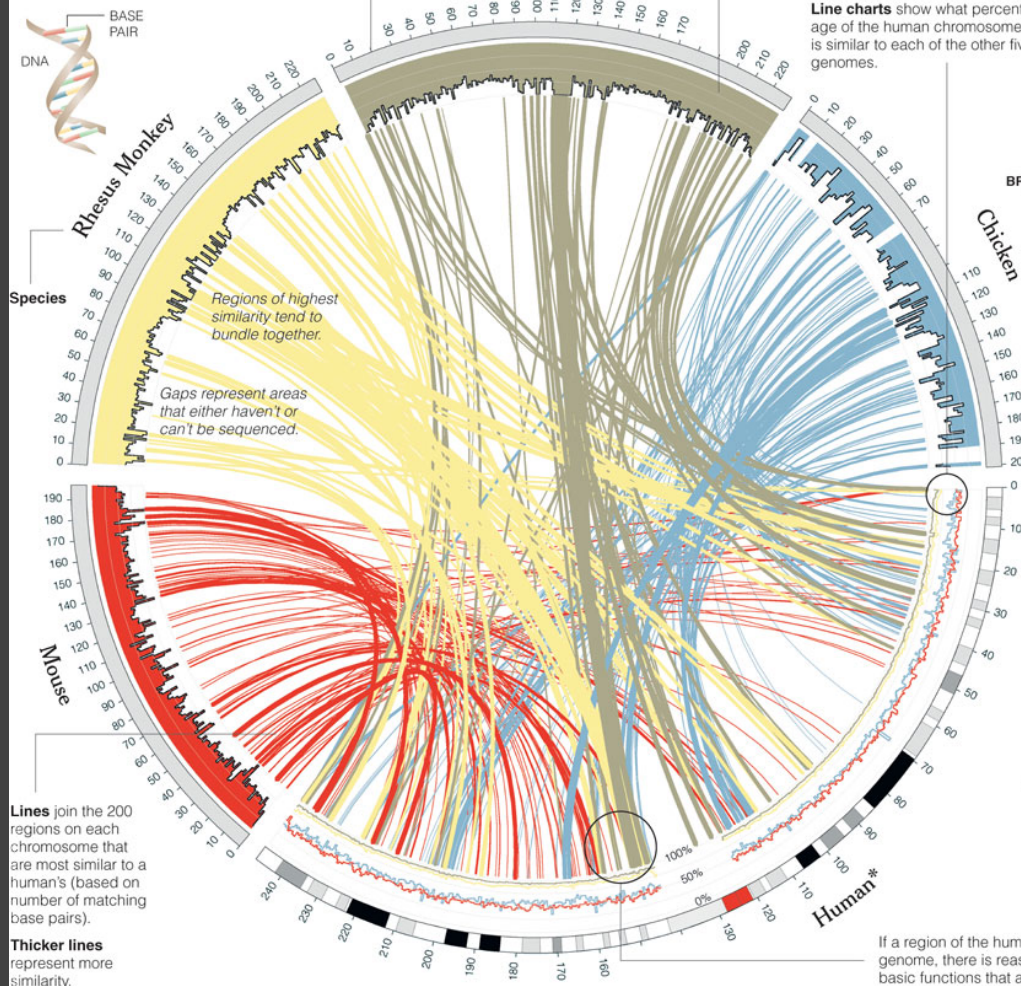ates the large degree of similarity between the first chromosomes of four animals to that of a human. Not surprisingly, the humans' is closest to the chimp's.
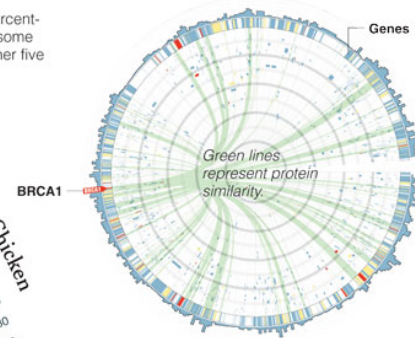
DAVID CONSTANTINE

COMPARING CHROMOSOMES 1

**Outer band** represents each species' first chromosome. Numbers represent millions of base pairs on the chromosome.

BASE PAIR

DNA

Species

Regions of highest similarity tend to bundle together.

Gaps represent areas that either haven't or can't be sequenced.

Lines join the 200 regions on each chromosome that are most similar to a human's (based on number of matching base pairs).

Thicker lines represent more similarity.

Chimp

Rhesus Monkey

Mouse

Chicken

Human*

Bar charts tell how many base pairs, 0 to 1 million, match part of the human chromosome.

Line charts show what percentage of the human chromosome is similar to each of the other five genomes.

OTHER TYPES OF COMPARISONS

To download the free program or view other examples: http://mkweb.bcgsc.ca/circos/

Genes

BRCA1

Green lines represent protein similarity.

The chart above shows the similarity of the BRCA1 protein, implicated in early breast cancer, to other genes on human chromosome 17.

Human chromosome 4

7

2

1

Lines connect areas of similarity.
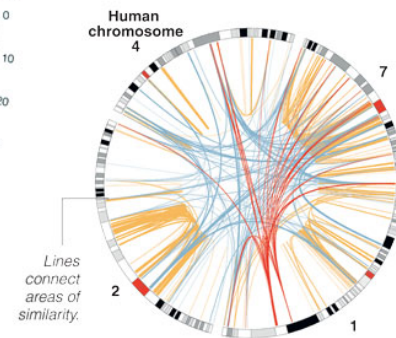
The image above illustrates the duplication within the human genome. Here, chromosomes 1, 2, 4 and 7 are shown (arbitrarily chosen).

If a region of the human genome is very similar to a region in another's genome, there is reason to suspect that these two regions both generate basic functions that are vital to both species and do not permit variation.

Source: Martin Krzywinski, Michael Smith Genome Sciences Center, Canada

*Length shown at 200% compared to other species. Shadings on outer band represent how chromosome looks when stained.

The New York Times

Tabular Layout and Multiple Coordinated Views (dashboards)

Table Lens

League	Players	Home ...	At Bats	Hits
National	Dale Murphy	29	614	163
	Eric Davis	27	415	115
	Darryl Strawb...	27	475	123
American	Jesse Barfield	40	589	170
	Dave Kingman	35	561	118
	Gary Gaetti	34	596	171

Table Lens

Idea: Make the text more visual and symbolic
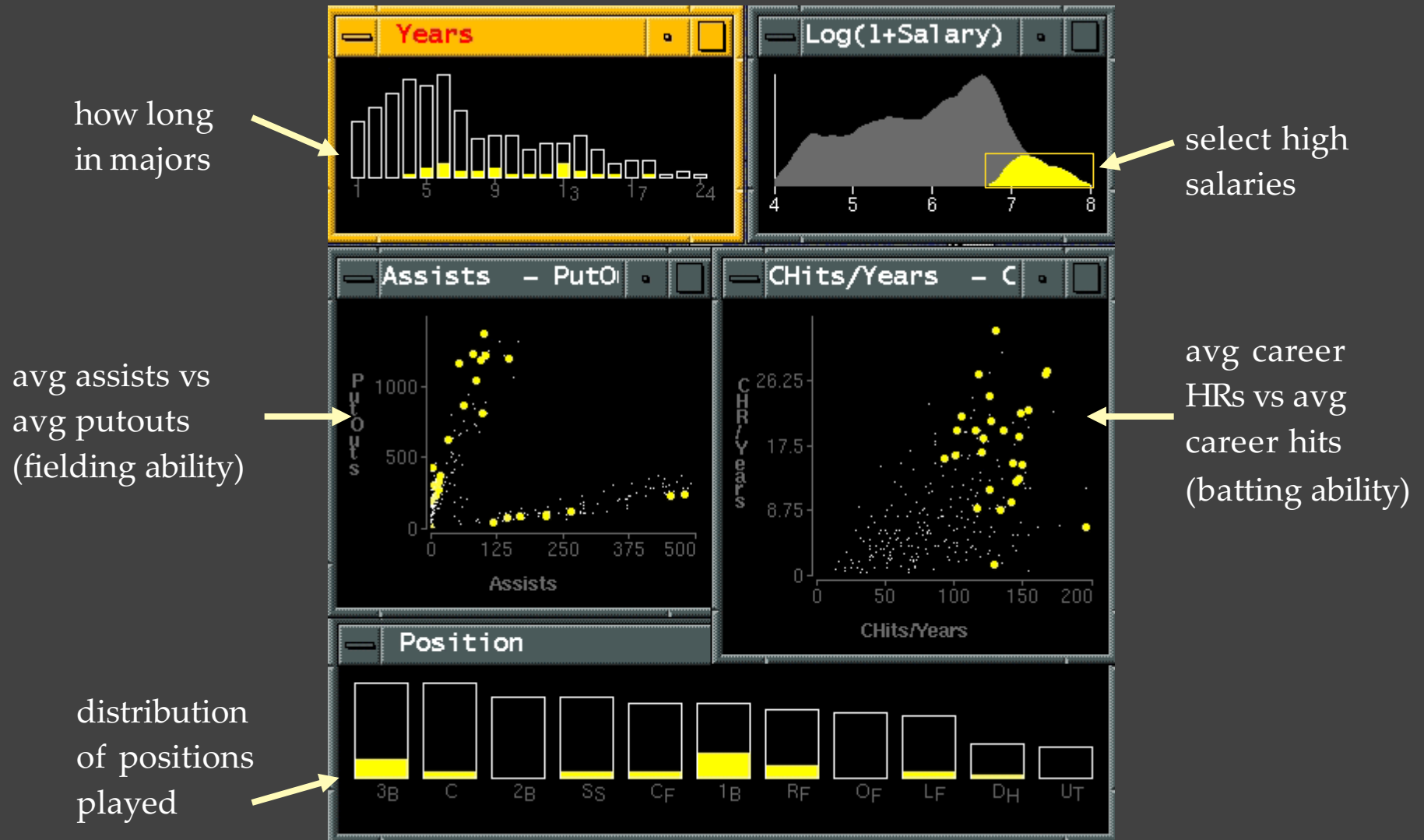Just leverage basic bar chart idea

Characteristics

Can sort on any attribute (row)
Focus on an attribute value (show only cases having that value) by doubleclicking on it
Can type in queries on different attributes to limit what is presented to.   Note this is main contribution: dynamic control (selection/change/querying/filtering) of individual attributes.

Multiple Coordinated Views

how long in majors

select high salaries

avg assists vs avg putouts (fielding ability)

avg career HRs vs avg career hits (batting ability)

distribution of positions played

Also known as: Dashboards