Temporal & Multivariate Data

COMP7507
Visualization & Visual Analytics

Types of Data

- One-dimensional — linear data
  - e.g., age distribution
  - includes sequential data such as text, program source codes
- Two-dimensional: planar or map data
  - e.g., geographical maps, floorplans, newspaper layouts
- Three-dimensional: real-world objects
  - e.g medical scans, architectural design
- Temporal
  - e.g. timelines, weather

Types of Data

- Multidimensional or Multivariate
  - e.g., financial data, customer behaviour
- Tree (hierarchical data)
  - e.g., file structure, evolution
- Network (relational data)
  - e.g., social network, air traffic
- The above classification is not mutually exclusive
  - E.g., how about air traffic data?
  - multivariate, geographical, network, and temporal

Shneiderman, B., "The eyes have it: a task by data type taxonomy for information visualizations," Proc. IEEE Symposium on Visual Languages, 1996, pp.336,343.

The Iris Sample Data Set

- Created by R.A. Fisher
- Possibly the best known data set in the pattern recognition community
- 3 classes (types of iris)
- 50 objects in each class
- 5 attributes
  - sepal length & width (cm)
  - petal length & width (cm)
  - class (Iris Setosa, Iris Versicolour, Iris Virginica)

Iris Setosa

Iris Versicolour

Iris Virginica

[wikipedia]

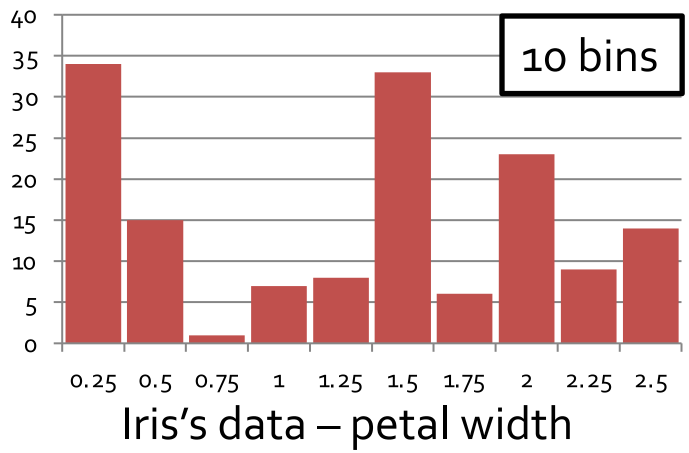
The Iris Sample Data Set

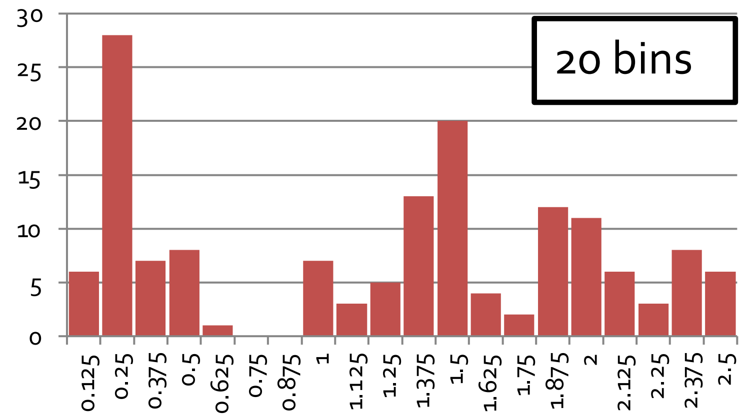
iris.data Sepal Length	iris.data Sepal Width	iris.data Petal Length	iris.data Petal Width	iris.data Class
5.10000	3.50000	1.40000	0.20000	Iris-setosa
4.90000	3.00000	1.40000	0.20000	Iris-setosa
4.70000	3.20000	1.30000	0.20000	Iris-setosa
4.60000	3.10000	1.50000	0.20000	Iris-setosa
5.00000	3.60000	1.40000	0.20000	Iris-setosa
5.40000	3.90000	1.70000	0.40000	Iris-setosa
4.60000	3.40000	1.40000	0.30000	Iris-setosa
5.00000	3.40000	1.50000	0.20000	Iris-setosa
4.40000	2.90000	1.40000	0.20000	Iris-setosa
4.90000	3.10000	1.50000	0.10000	Iris-setosa

Some Basic Plots

Bar Charts / Histograms

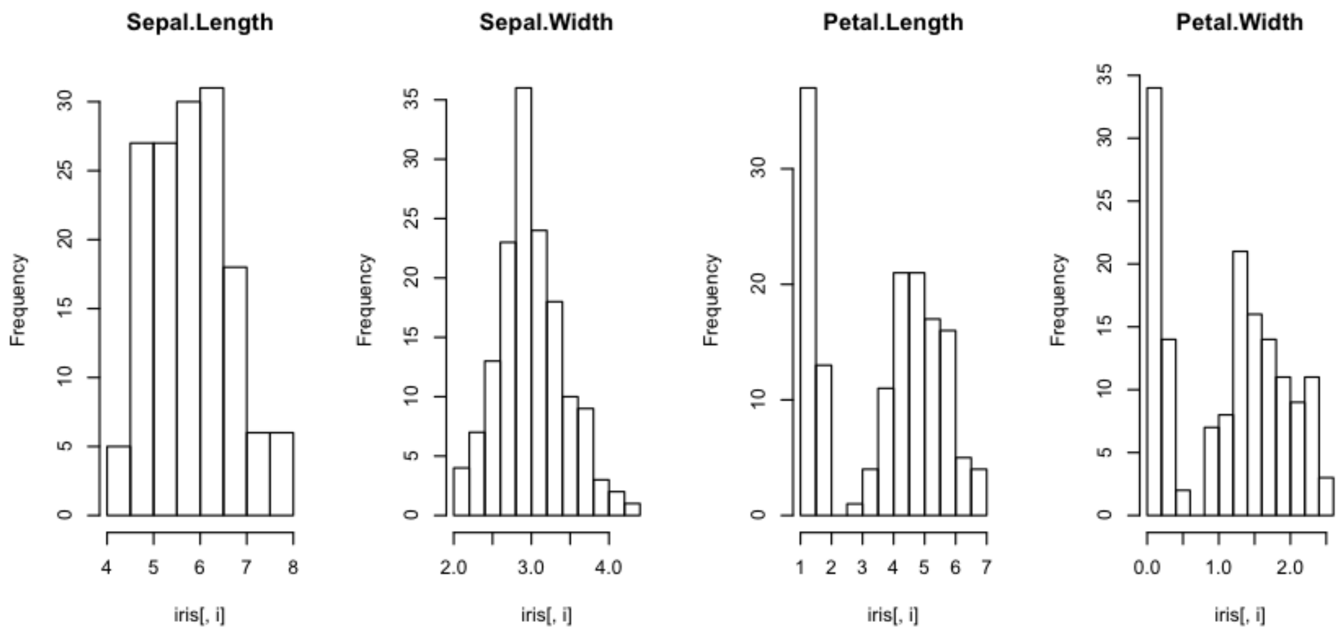
- To show distribution of values of a single variable
- Values are divided into bins
- A bar plot is used so that the height of each bar indicates the number of objects in each bin
- Shape of histogram depends on the number of bins





Box Plots

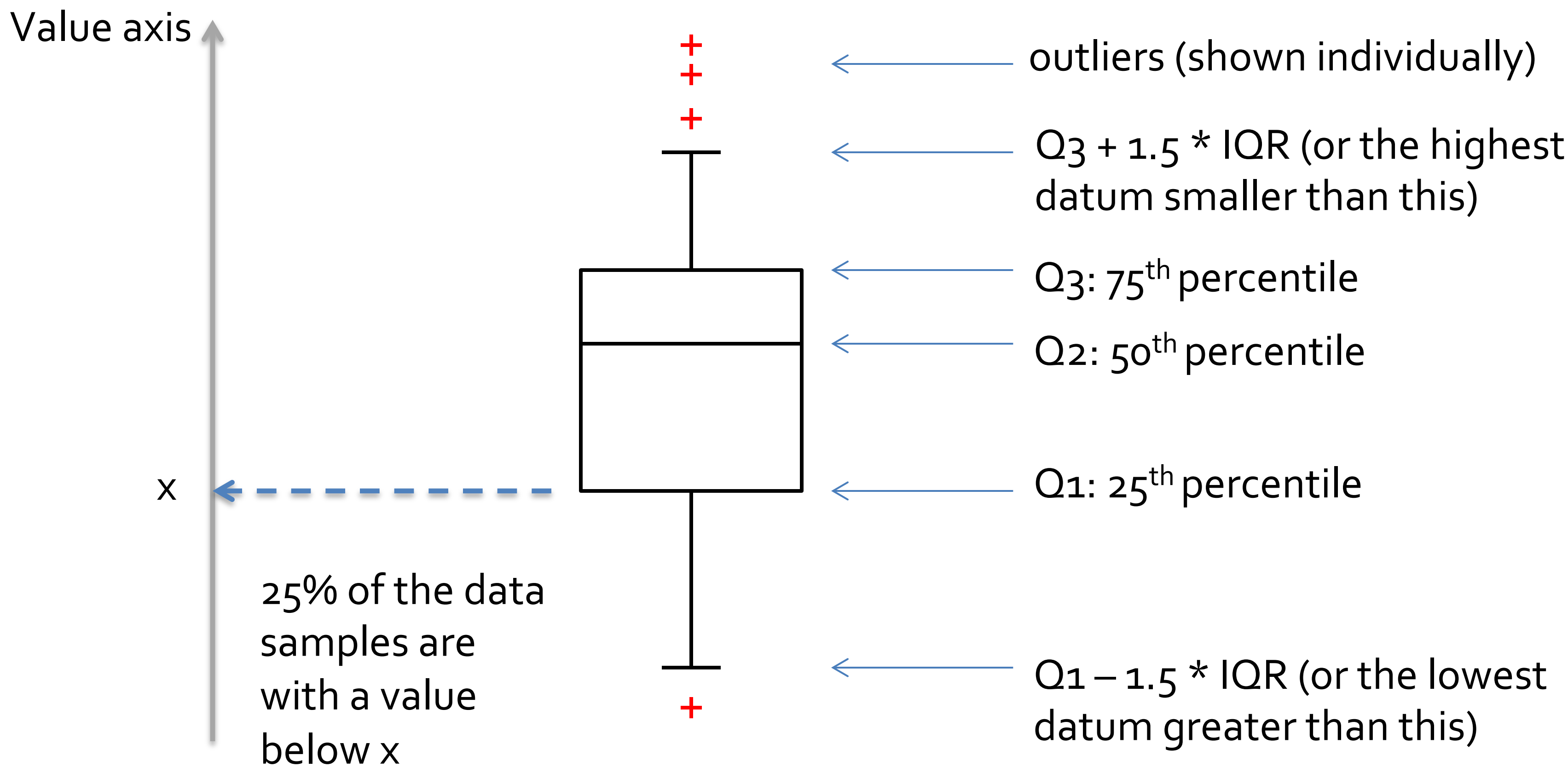
- Conventional histograms / distribution plots might not be space-efficient.



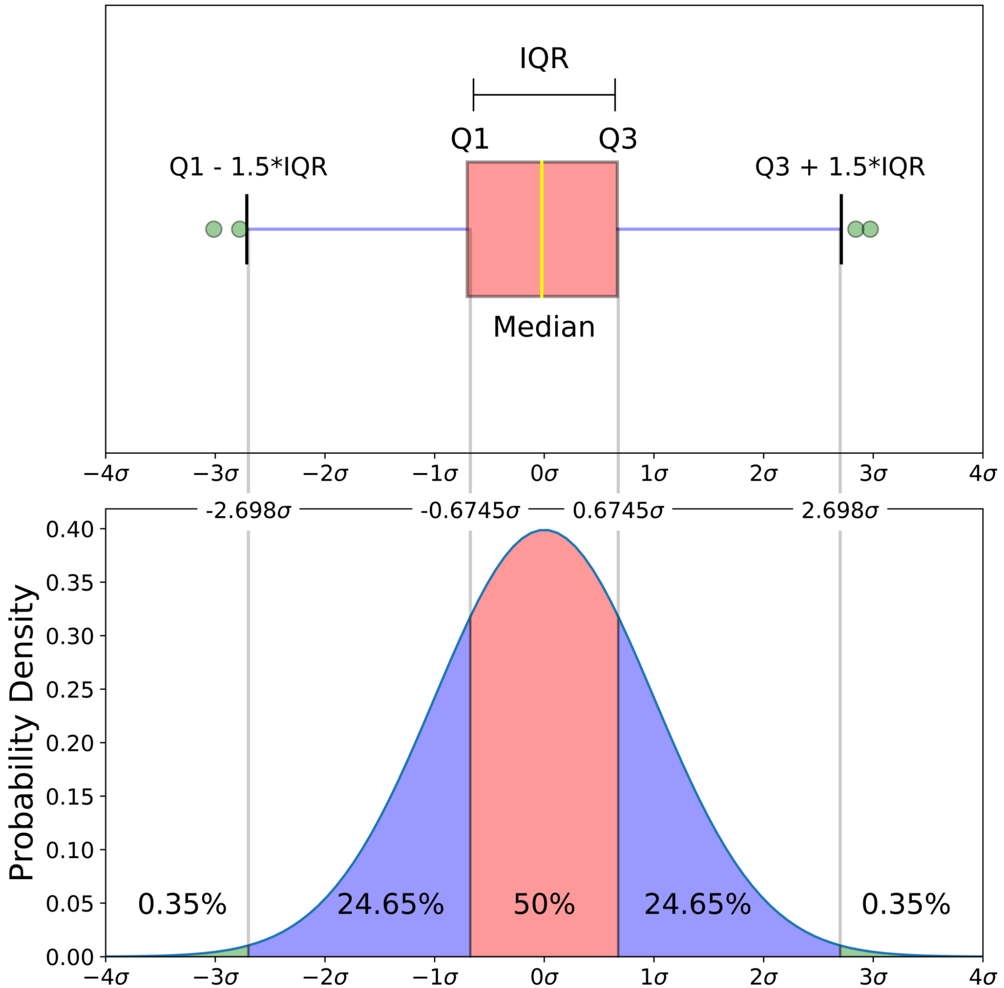
https://machinelearningmastery.com/data-visualization-in-r/

Box Plots

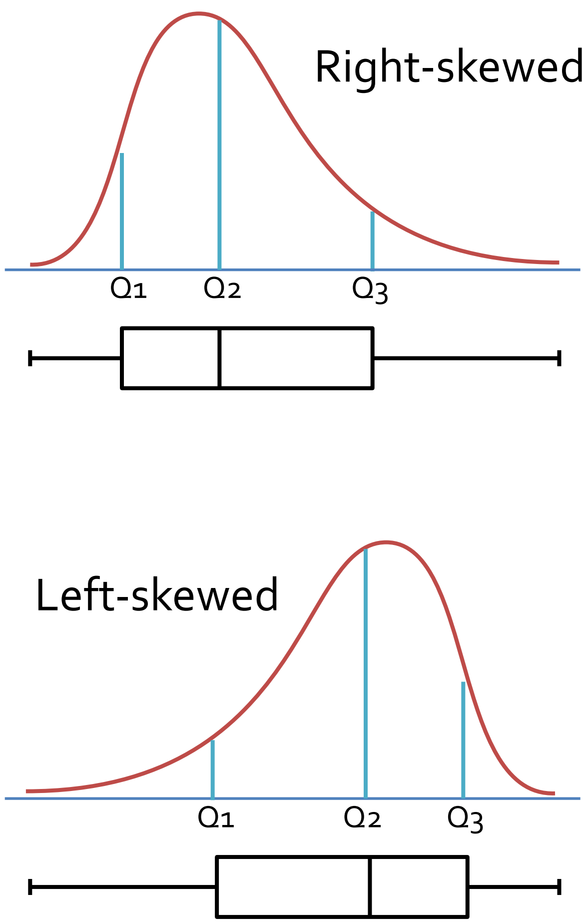
- An efficient way to show quantitative distribution of 1D data



Box Plots



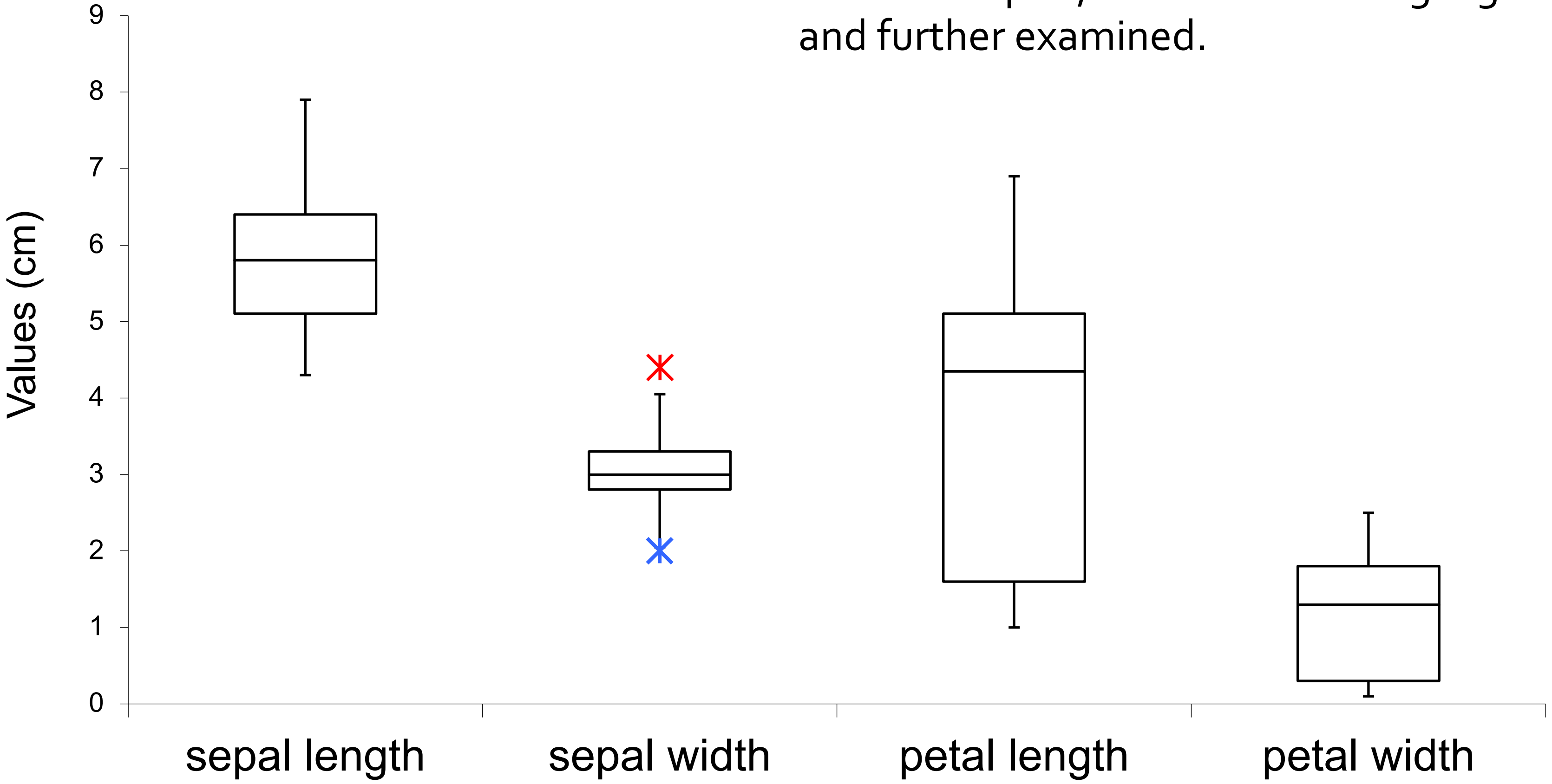
For normal distribution (symmetric)



https://builtin.com/data-science/boxplot

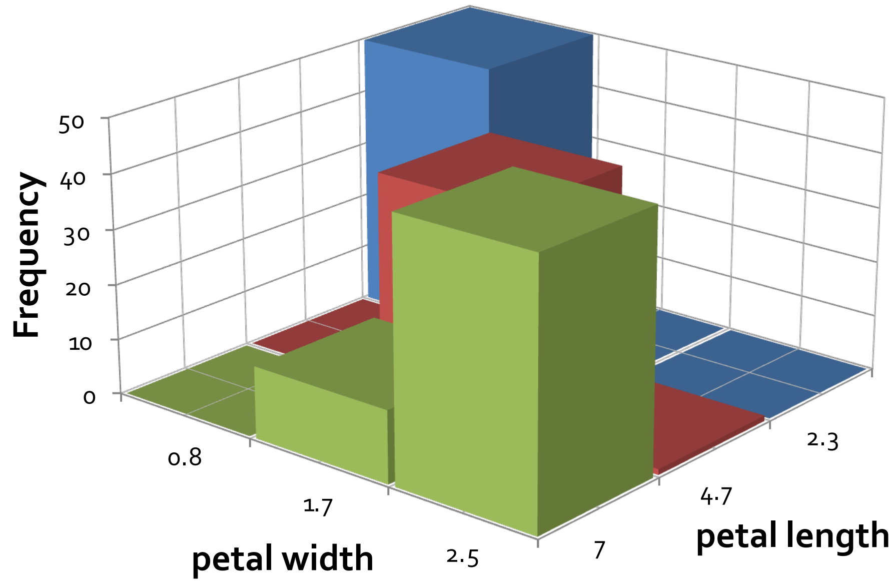
Box Plots

With a box plot, outliers can be highlighted and further examined.



2D Bar Charts

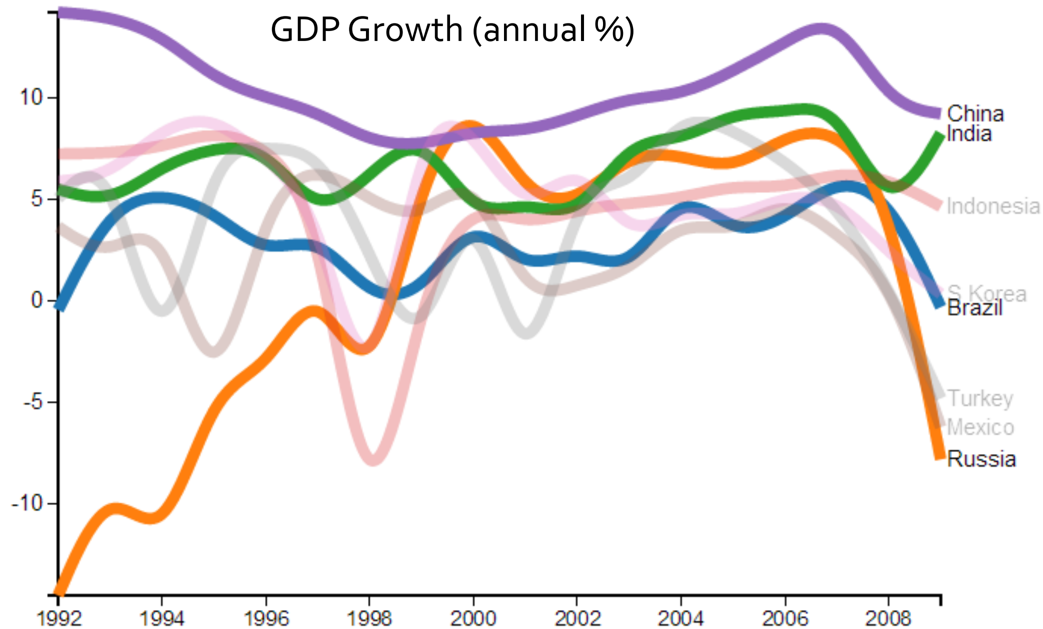
- To show the joint distribution of the values of two variables



3D effect not good in showing the exact values, but the correlation can be seen clearly

Line Graphs

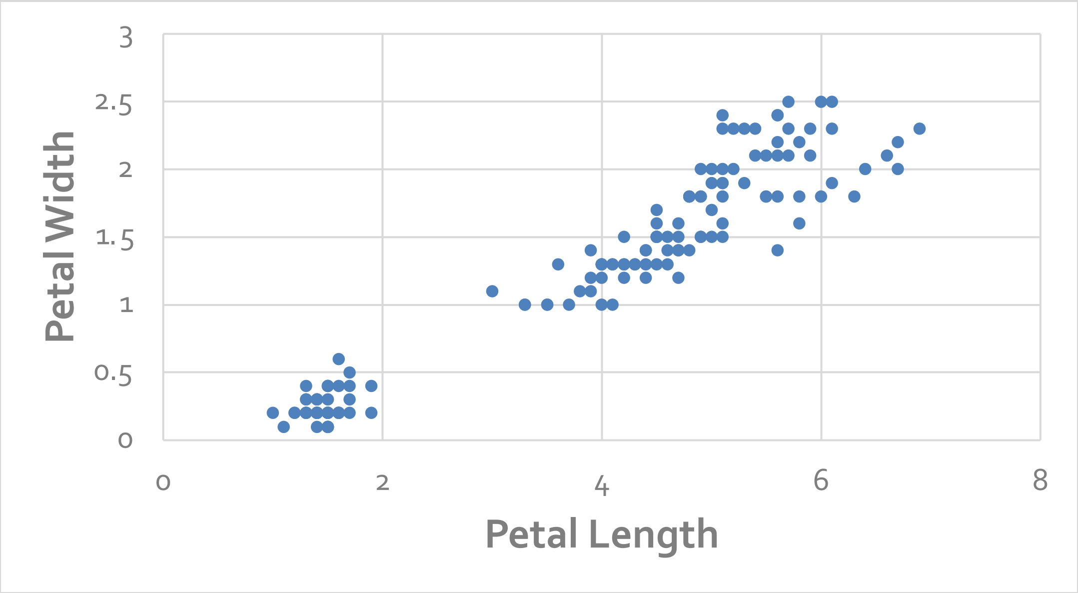
- Points connected by lines to show how something changes in value (usually over time)



Issues:
- occlusion when too many lines,
- when to use piecewise or smooth lines

Scatter Plots

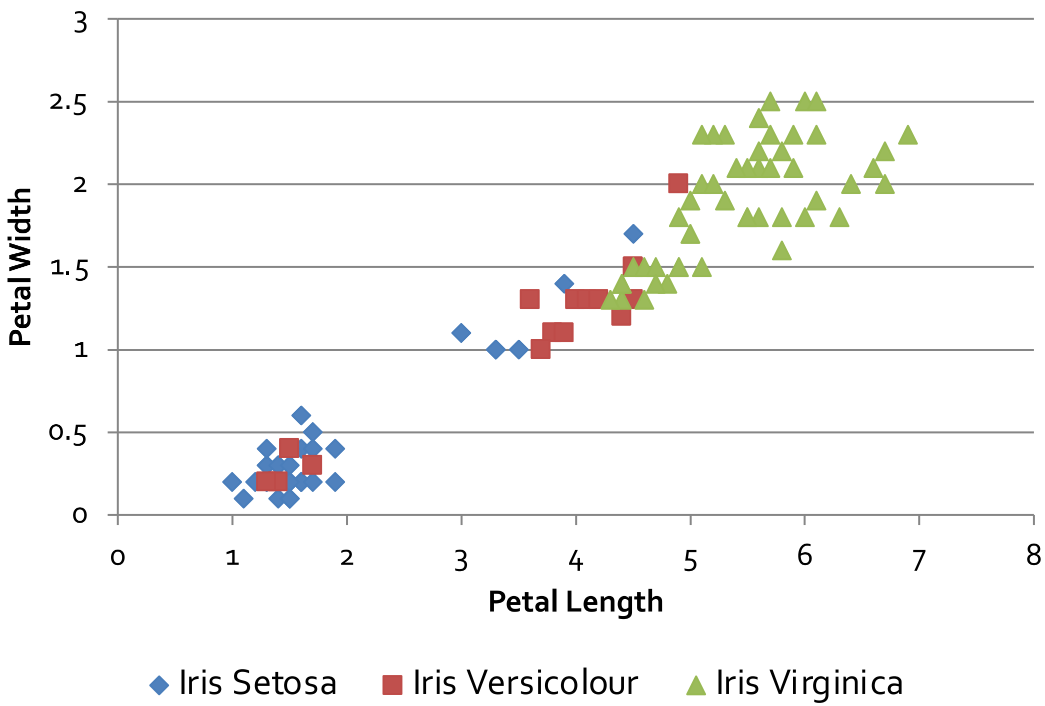
- A plot of points showing the relationship between two variables of a set of data



- Point position determined by attribute values

Scatter Plots

- A plot of points showing the relationship between two variables of a set of data

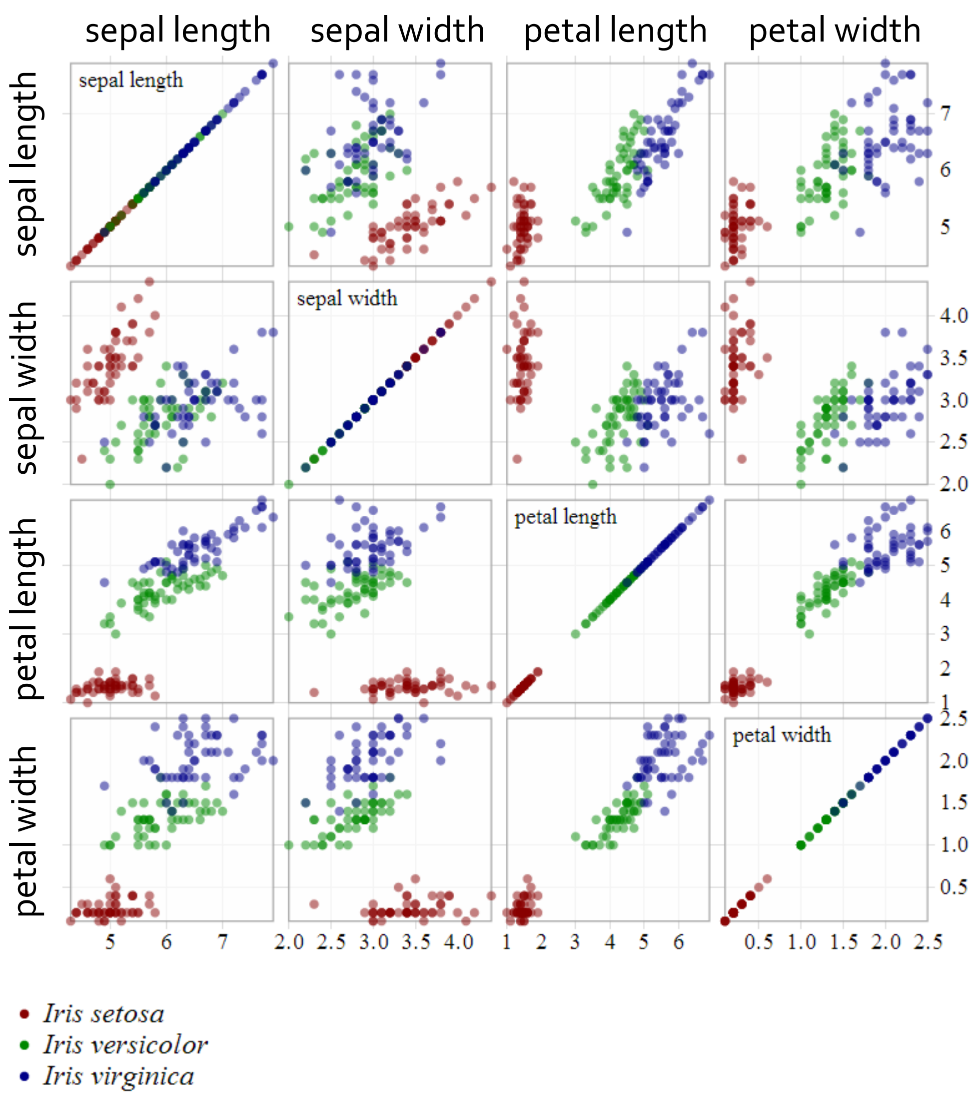


- Point position determined by attribute values
- Additional attributes can be marked by size, shape, or color for each item

Scatter Plots

- 2D scatter plots are commonly used, and there are also 3D scatter plots
- Can compare two attributes at a time only. What if we have a lot of pairwise comparisons to show?
  - have an array (a matrix) of scatter plots

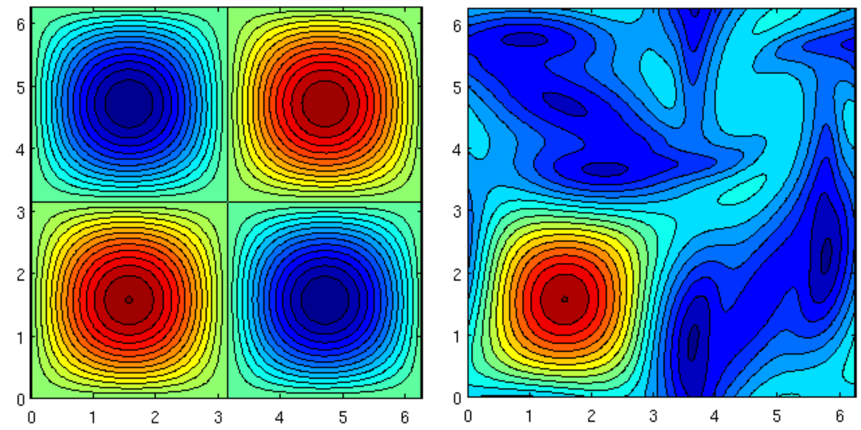
Scatter Plot Matrices



- Diagonal plot show the distribution of the 1D data
- Can also make use of the diagonal space for other kinds of 1D visualization, e.g., a histogram
- Matrix plot can be used not only for scatter plot, but anything that can deal with a bivariate plot

Contour Plots

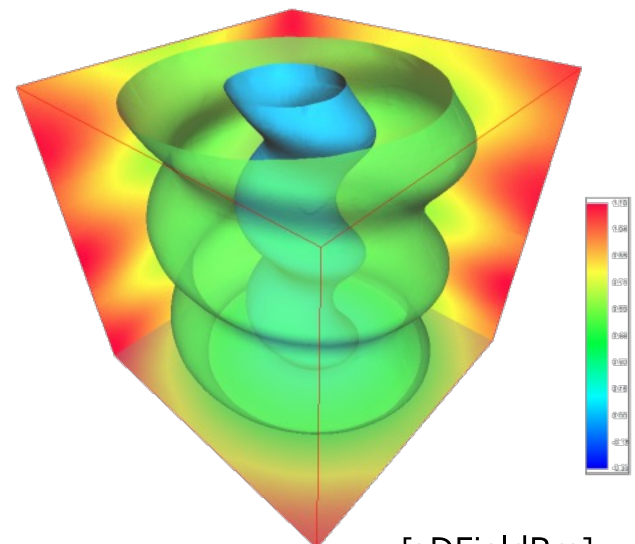
- To show continuous attributes measured on a spatial grid
- Partition the space into regions of similar values; boundaries of regions are contour lines called iso-value lines, or isolines.



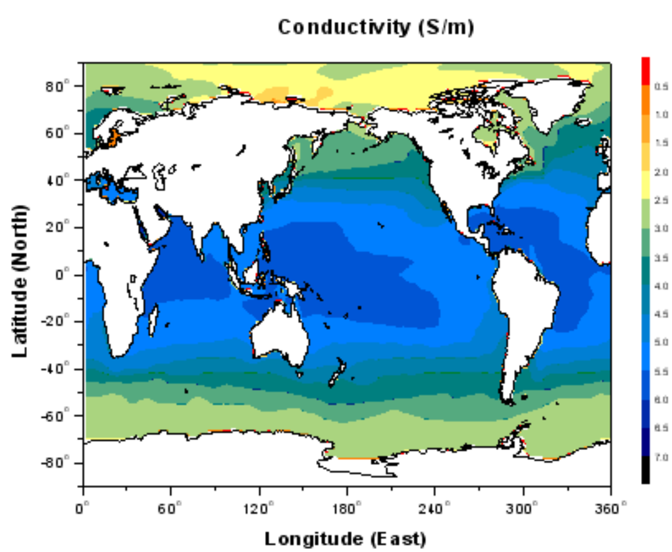
[Math, NYU]

Contour Plots

- Commonly used in scientific visualization
- Examples: height fields, temperature, rainfall, etc.



[3DFieldPro]



[OriginLab]

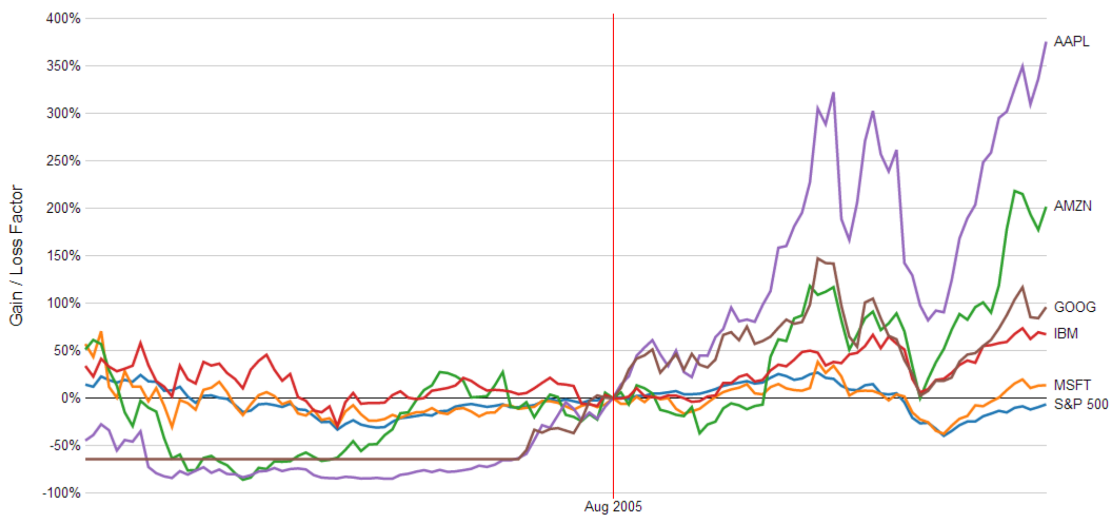
Temporal Data

Time-Series Data

- Set of values that change over time
- Examples:
  - Finance (stock prices, exchange rates)
  - Science (temperatures, pollution levels, electric potentials)
  - Public policy (crime rates, public health)
- Common requirements:
  - Able to compare many time series simultaneously
  - Able to use different visualizations in combination

Index Charts

- Interactive line chart showing % change based on a selected index point
- Useful for showing relative changes

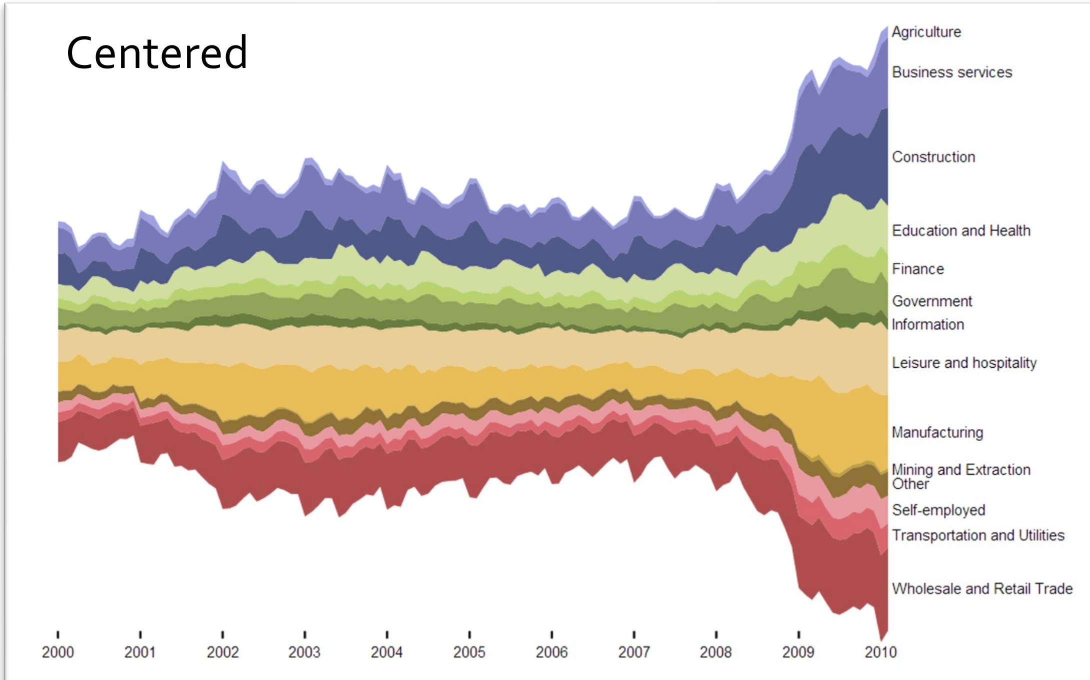


percentage change of selected stock prices according to the day of purchase

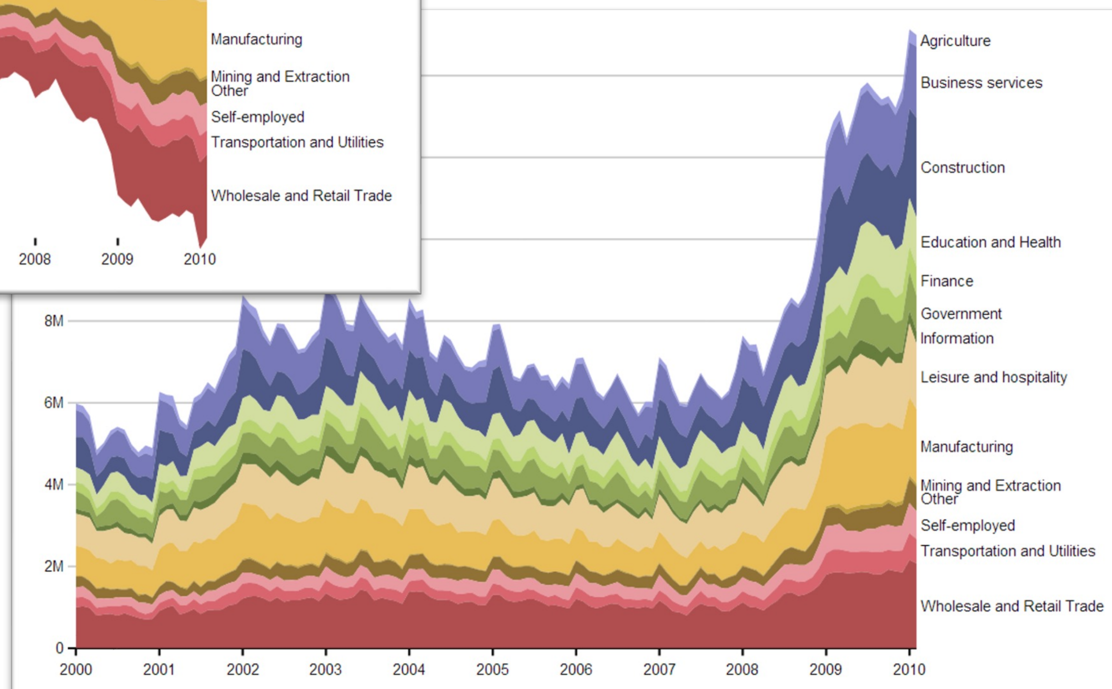
[Heer et al.]

Stacked Graphs

Centered



Zero-Basedline

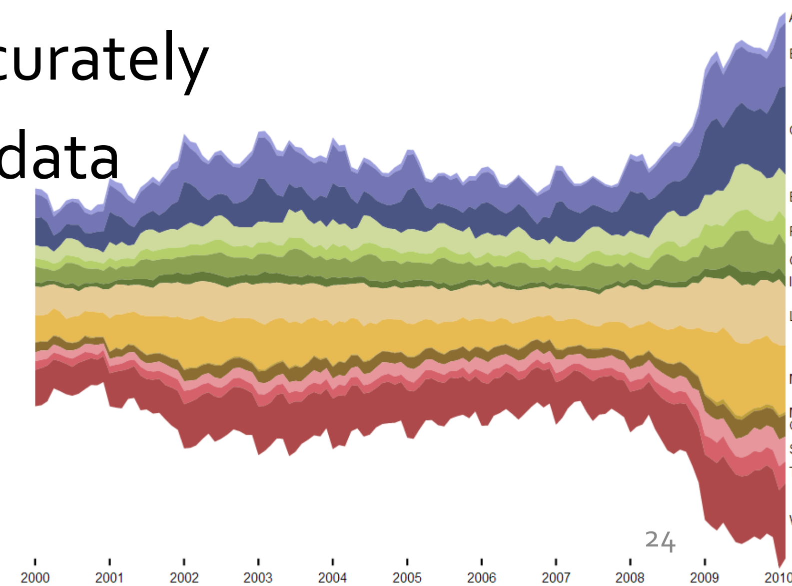


Total counts of unemployed US workers per industry, 2000-2010.

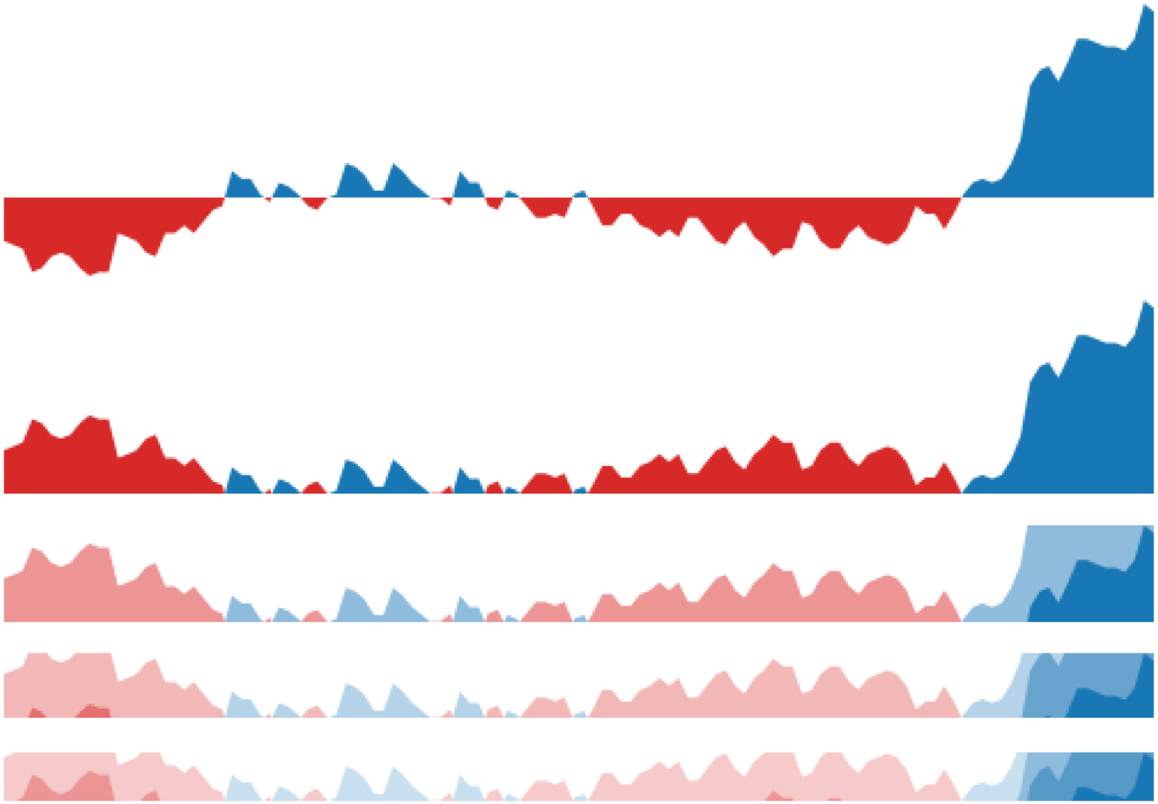
[Heer et al.]

Stacked Graphs

- Stack area charts on top of each other
- Useful for showing summation of time-series values (aggregation)
- Limitation:
  – negative numbers not supported
  – difficult to interpret trends accurately
  – meaningless for some kind of data (e.g., temperatures)



Horizon Graphs



US unemployment rate, 2000-2010.
Positive values: above average unemployment
Negative values: below average unemployment

[Heer et al.]

[http://homes.cs.washington.edu/~jheer//files/zoo/ex/time/horizon.html]

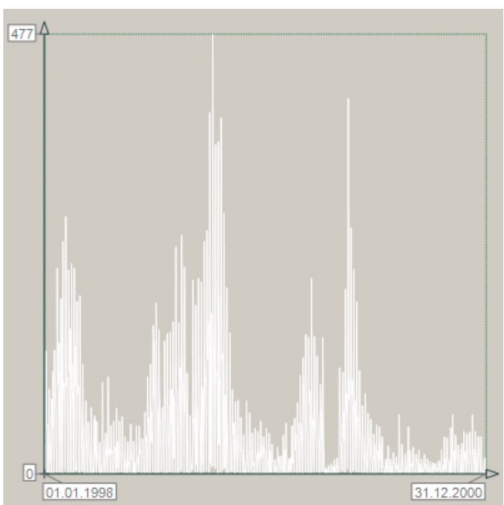
Horizon Graphs

- To divide the area plot into horizontal bands and layer them over each others.
- Useful for increasing the data density (i.e. save space) without sacrificing resolution.
- Limitation: Not intuitive and takes time to learn

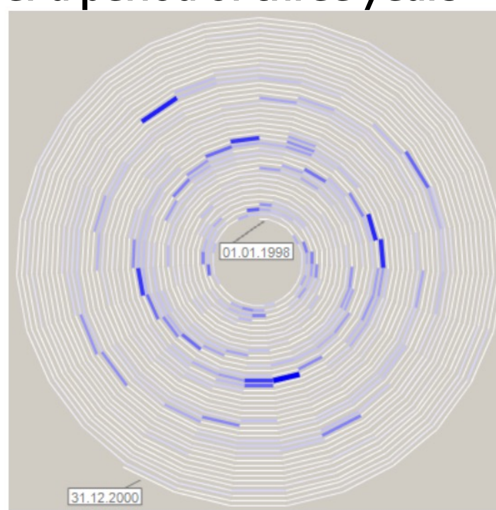
Spiral Graphs

- Use a spirally shaped time axis
- Good for showing or identifying periodic structure of data

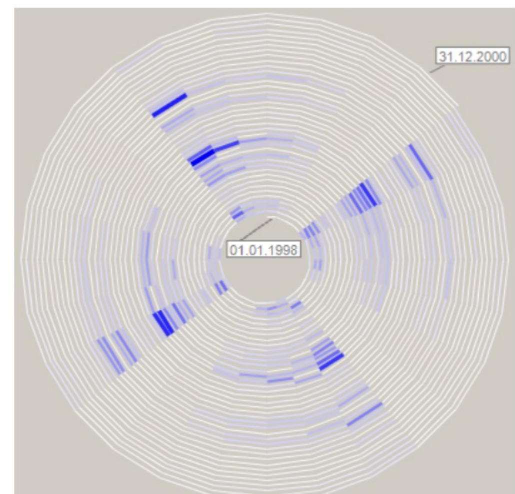
Number of influenza cases over a period of three years



Time series



27 days per cycle



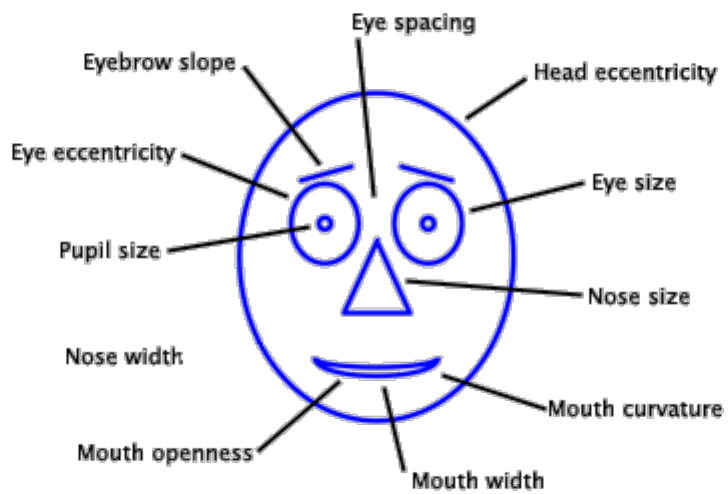
28 days per cycle

[Aigner et al., "**Visual Methods for Analyzing Time-Oriented Data**", *IEEE TVCG*, 2008.]

Multivariate Data

Chernoff Faces

- Relate data to facial features, something which we find easy to differentiate
- Each feature, e.g., mouth, encode a data dimension by their shape, size, placement and orientation



10 facial features, each corresponds to a parameter in [0,1]

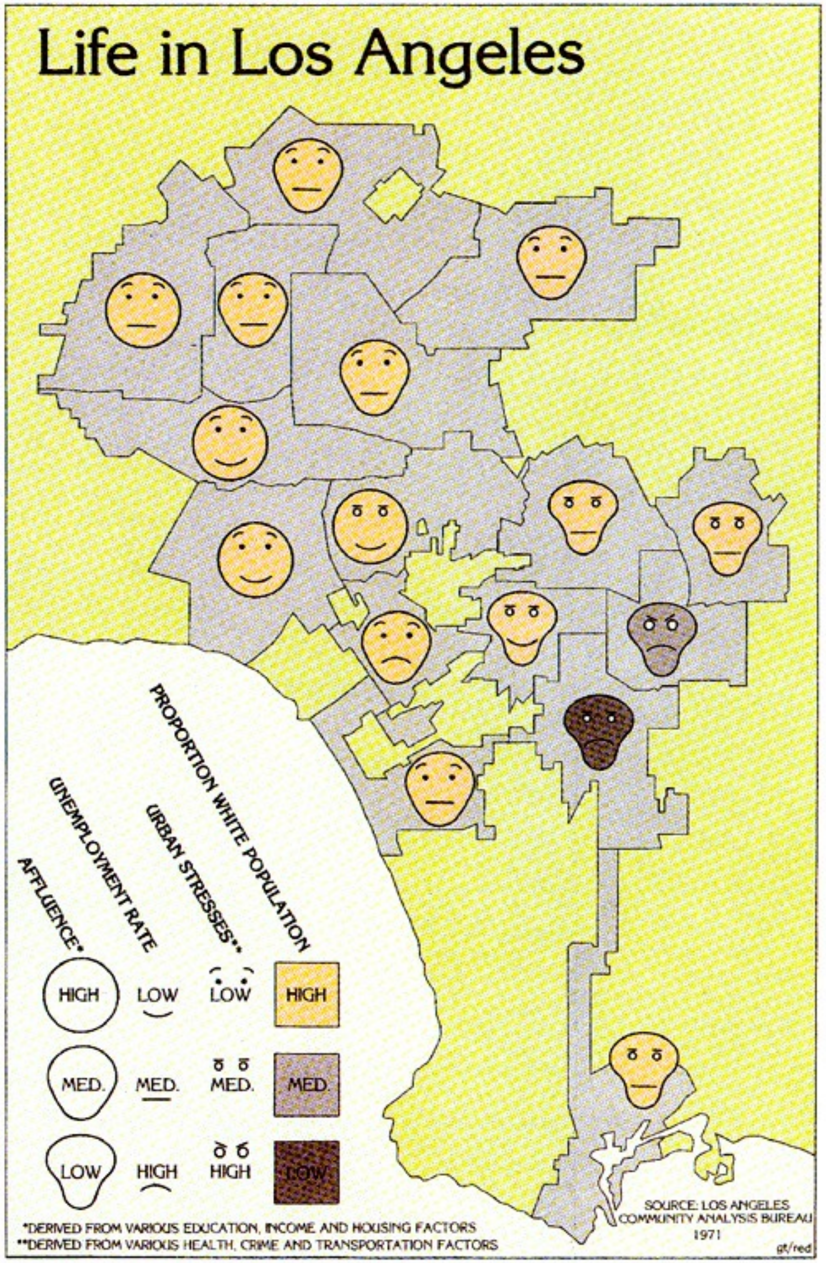
All 0

All 0.5

All 1

Random parameters

[https://courses.isds.tugraz.at/ivis/surveys/ss2013/g3-survey-chernoff.pdf]

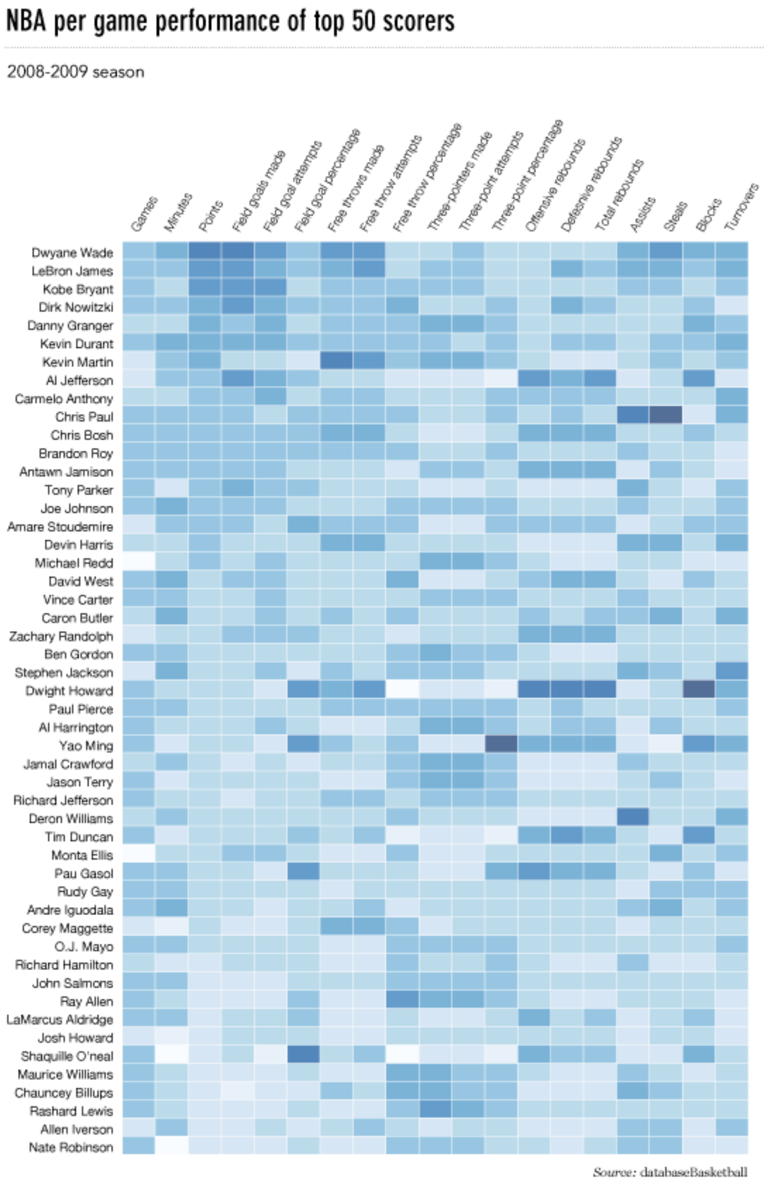


[https://courses.isds.tugraz.at/ivis/surveys/ss2013/g3-survey-chernoff.pdf]

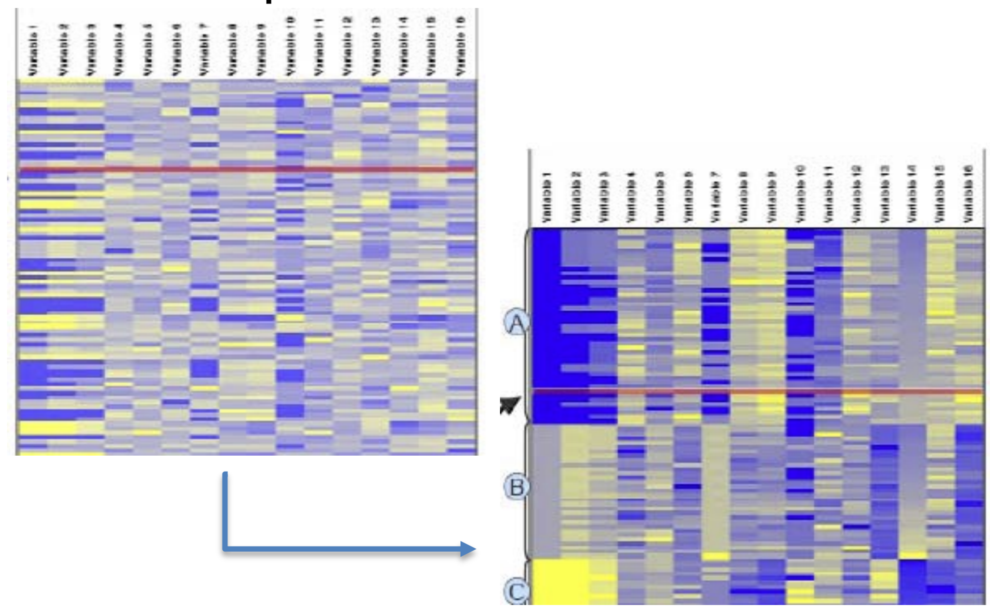
Chernoff Faces

- Represent only trends but not actual values
- Drawback: Affected by our perceived importance of a facial feature

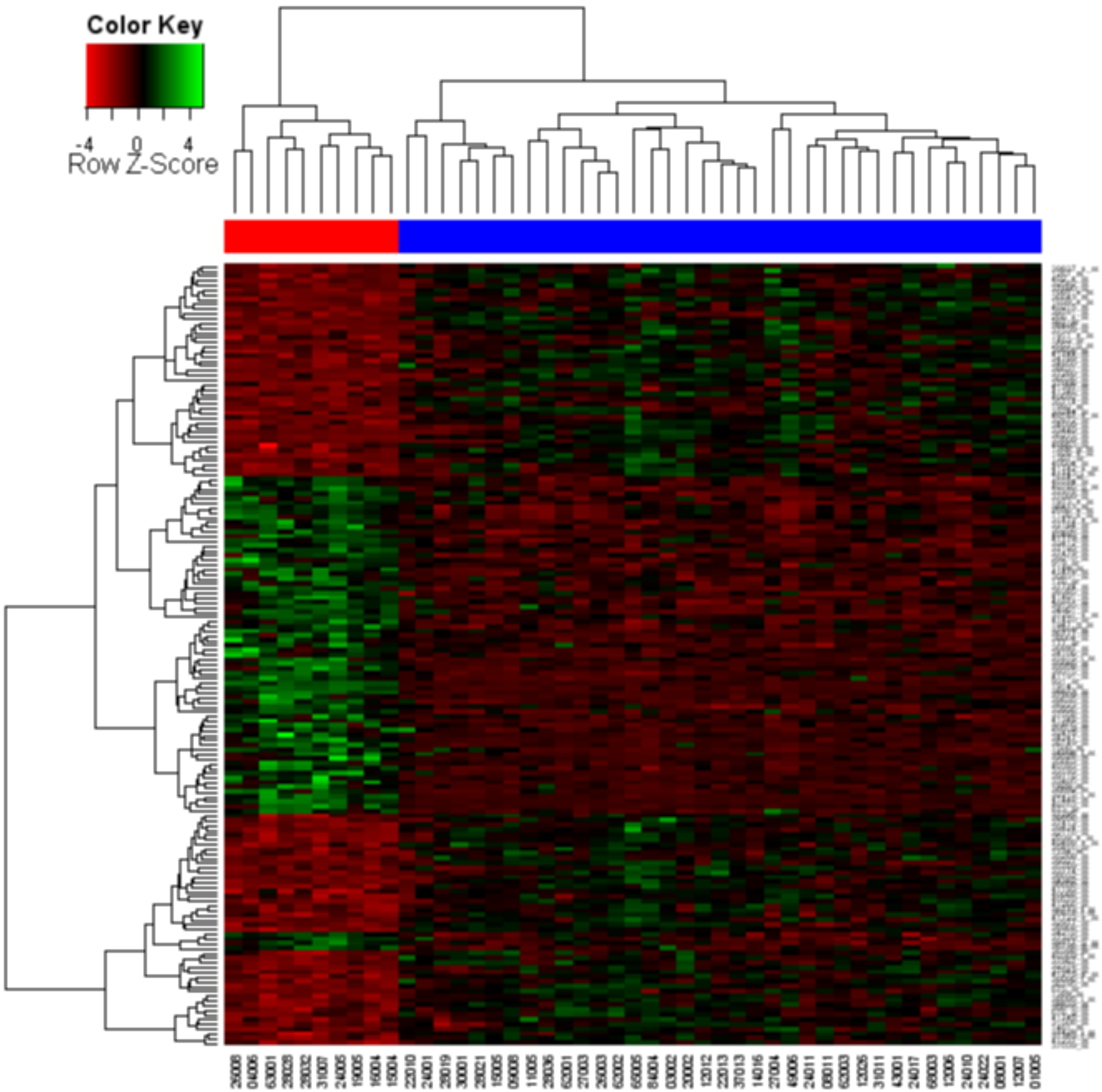
Heat Maps



- Encode values stored in table entries as colors
- Rows and columns can be reordered to better expose features.



[M. Ward]



column: patient
row: gene

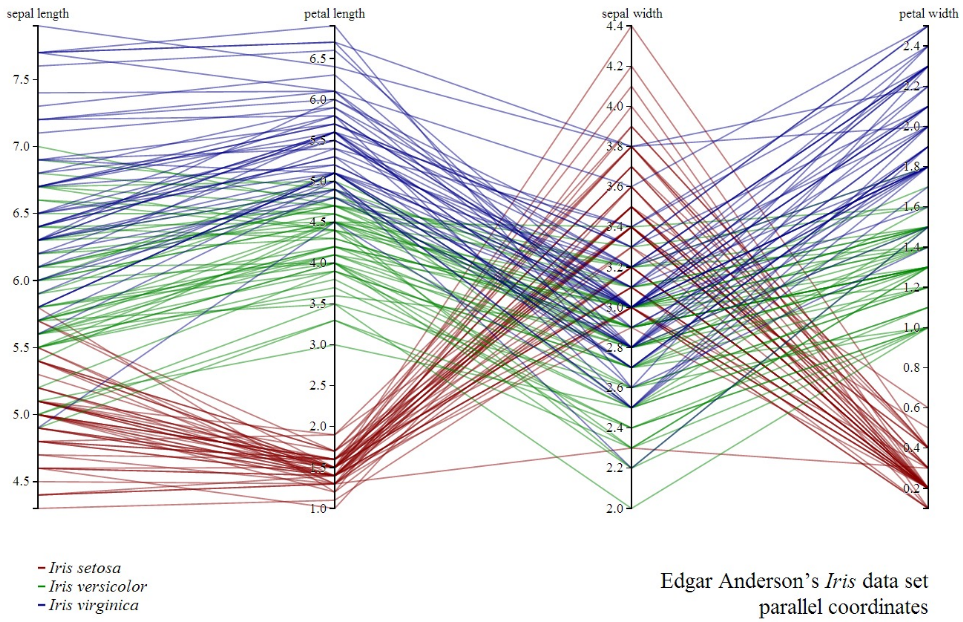
Heatmap from DNA microarray data showing genes expressed differently for two types of leukemia.

[Warwick, http://www2.warwick.ac.uk/fac/sci/moac/people/students/peter_cock/r/heatmap/]

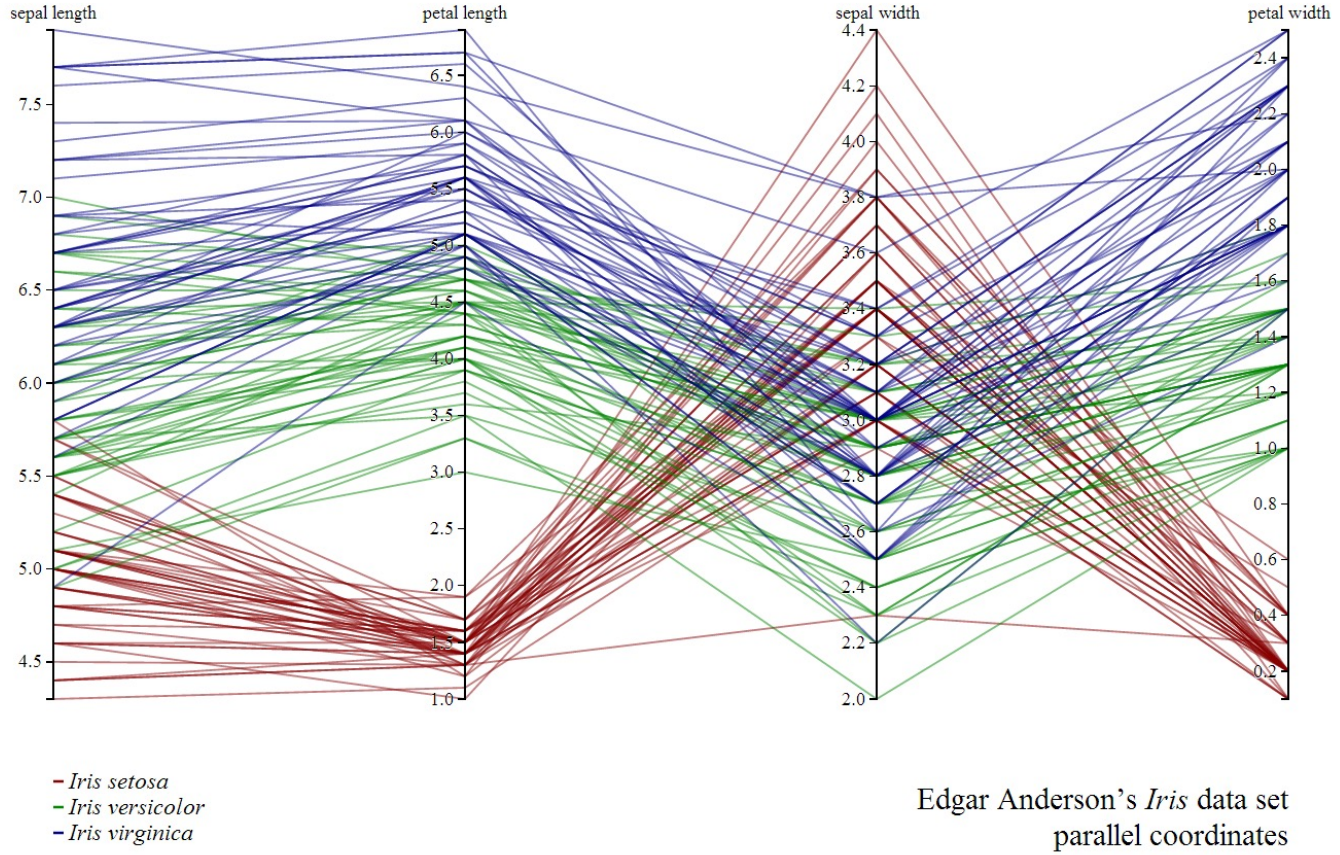
Interactive examples: http://amp.pharm.mssm.edu/clustergrammer/

Parallel Coordinates

- How to present all *n* axes of the *n* dimensions on a 2D plane?
- Use parallel axes instead of orthogonal axes

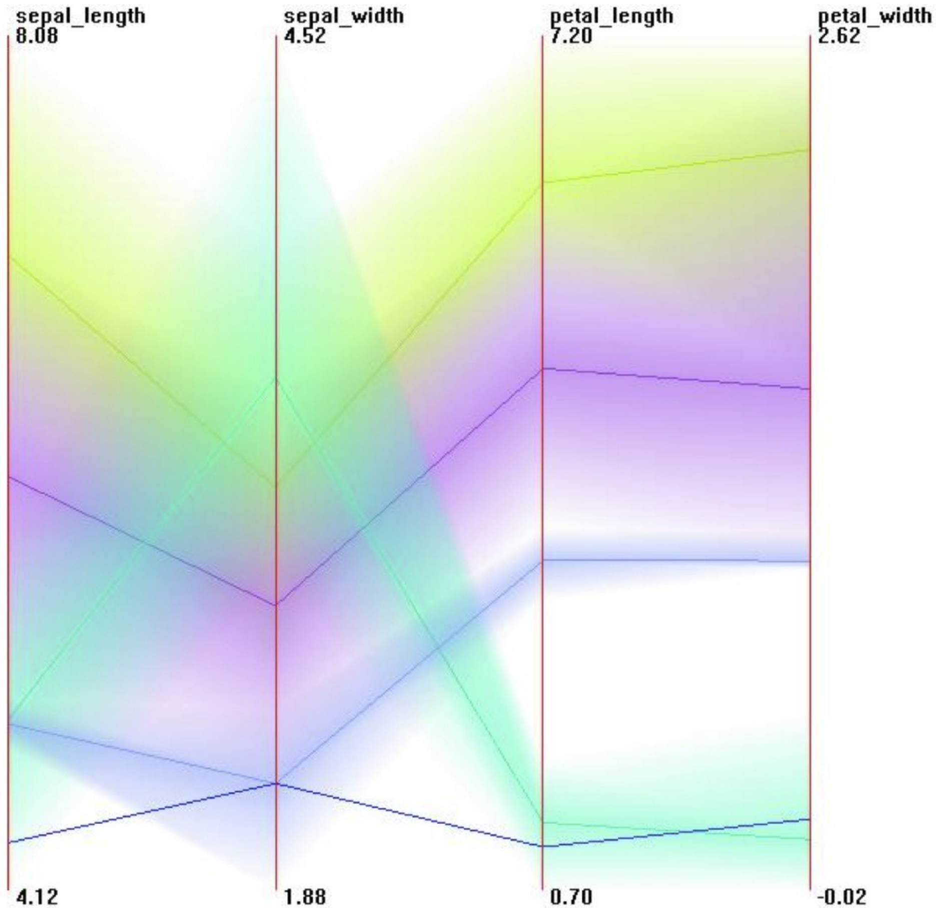


[http://mbostock.github.io/d3/talk/20111116/iris-parallel.html]



- Each attribute value of a data item corresponds to a point on a coordinate axis, and the data item is represented as a polyline connecting these points
- A distinct class of objects can sometimes be seen as a group of lines on some axes
- Ordering of axes is important for seeing patterns

Parallel Coordinates



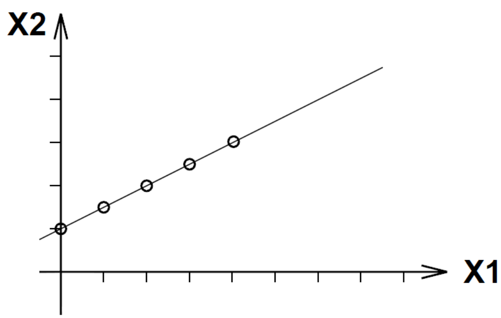
[Ward]

Parallel coordinate plot showing the distribution (i.e., centers and extents) of clusters

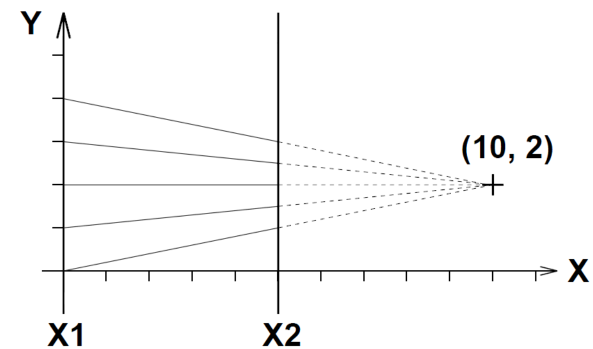
Parallel Coordinates

- Parallel correlation

X1 & X2
proportional

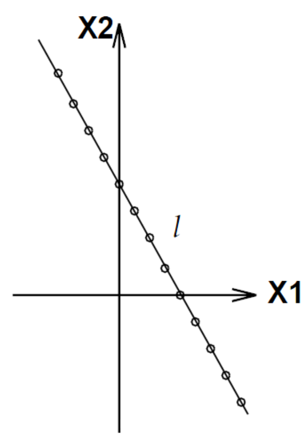


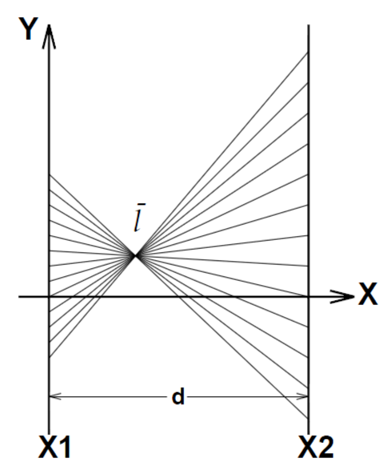
Cartesian point plot



PC plot

X1 & X2
inversely proportional

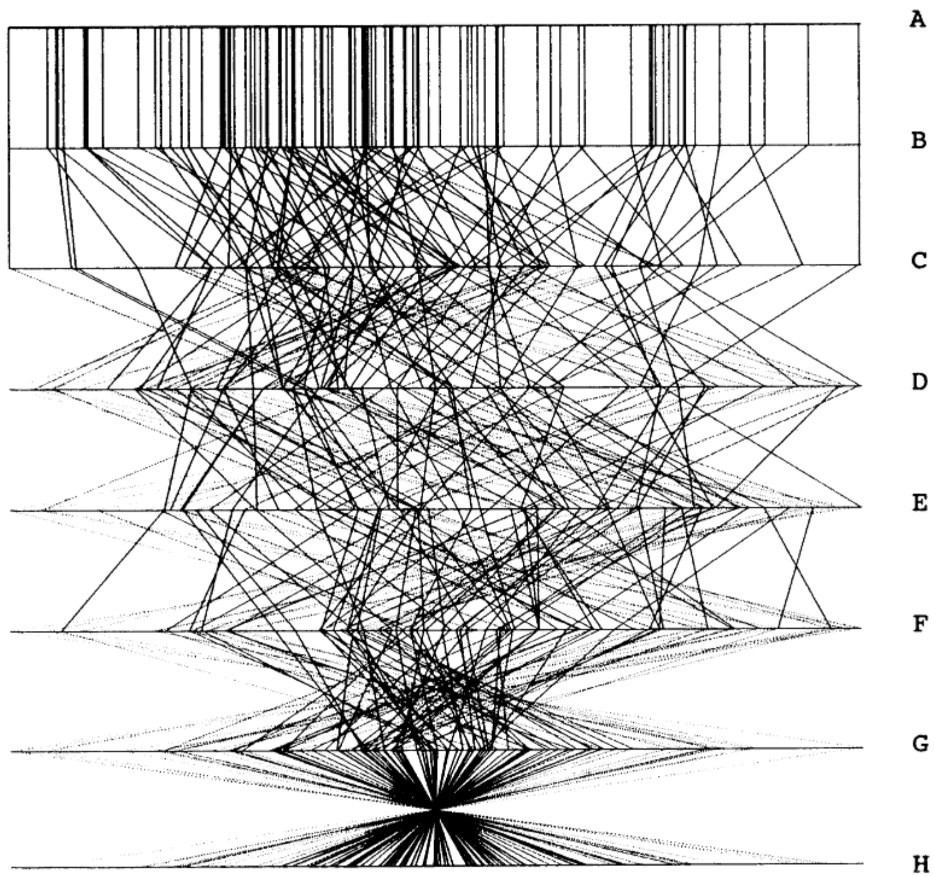




[Wong and Bergeron, "30 Years of Multidimensional Multivariate Visualization," *Scientific Visualization: Overviews, Methodologies & Techniques*, 1997.]

Parallel Coordinates

- Parallel correlation

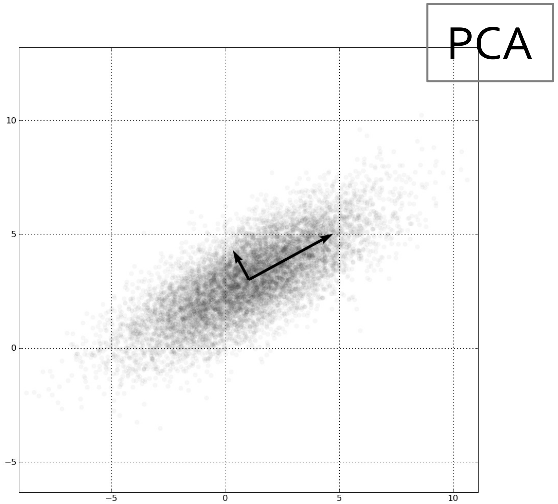


Parallel coordinate plot of six-dimensional data illustrating correlations of ρ (correlation coefficient) =1, .8, .2, 0, -.2, -.8 and -1.

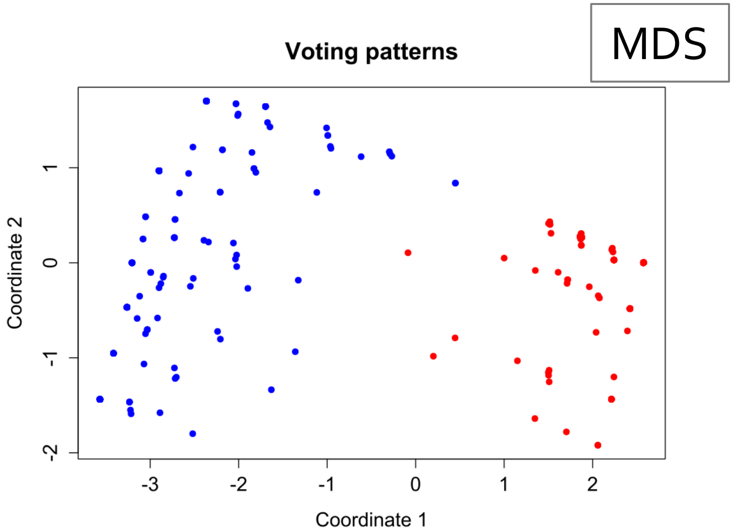
[Wegman, "Hyperdimensional Data Analysis Using Parallel Coordinates", Journal of the American Statistical Association, 1990.]

Dimension Reduction

- To remove some of the dimensions out from the display to avoid cluttering
  - Examples: Principle Component Analysis (PCA), Multidimensional Scaling (MDS), Self Organizing Maps (SOM)
- Issue: Resulting dimensions are not the original ones, not intuitive to users



[http://commons.wikimedia.org/wiki/File:GaussianScatterPCA.png#mediaviewer/File:GaussianScatterPCA.png]



[http://commons.wikimedia.org/wiki/File:RecentVotes.svg#mediaviewer/File:RecentVotes.svg]

Dimension Ordering

- Crucial for the effectiveness of many visualization techniques
- Relationship among adjacent dimensions are easier to detect than relationship among those positioned far apart, e.g., Parallel Coordinates, Heat Maps
- Use for attribute mapping to highlight important dimensions, e.g., Chernoff face
- An NP-complete problem equivalent to the Travelling Salesman Problem (TSP)
- Use approximation to compute ordering or by manual ordering (interaction needed)

Visualization Gallery

- Take a look at:
  - Tableau Public (https://public.tableau.com/s/gallery)
  - D3.js (http://d3js.org/)
  - Google Charts (https://developers.google.com/chart/interactive/docs/gallery)
- Try visualize the Iris data set with the different techniques taught in this class using the above tools.
- What can/cannot be done by these tools?

Reference

- Jeffrey Heer, Michael Bostock, and Vadim Ogievetsky. 2010. A tour through the visualization zoo. *Commun. ACM* 53, 6 (June 2010), 59-67. (http://hci.stanford.edu/jheer/files/zoo/)
- Matthew Ward, Georges Grinstein and Daniel Keim, "*Interactive Data Visualization: Foundations, Techniques, and Applications*", 2010 [Chapters 6 & 7]