Temporal, Geospatial & Multivariate Data

COMP8503

Advanced Topics in Visual Analytics

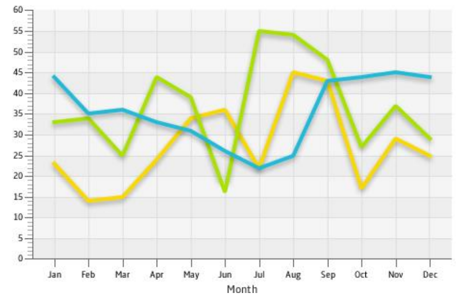
Types of Data

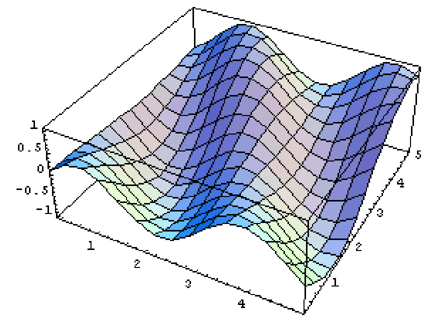
- One-dimensional — linear data
  - Includes sequential data such as text, program source codes
- Two-dimensional: planar or map data
  - Includes geographical maps, floor plans, newspaper layouts
- Three-dimensional: real-world objects
  - e.g medical scans
- Temporal
  - e.g. timelines
- Multidimensional or Multivariate
- Tree — hierarchical data
- Network — relational data with more complex structure than tree

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.

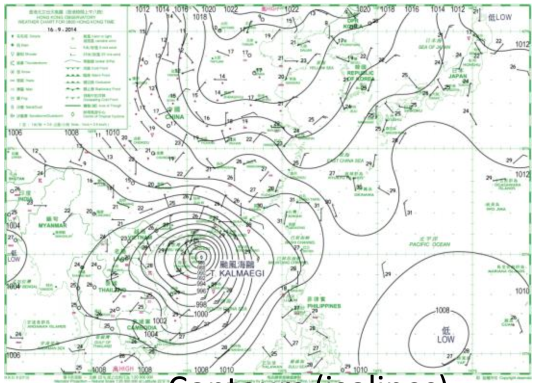
Scalar Function Visualization

- Univariate
  - a plot v = f(x)
- Bivariate
  - a surface v = f(x,y)
- Trivariate
  - a volume v = f(x, y, z)
- Multivariate or high-dimensional
  - nD data for n > 2
  - HiD data

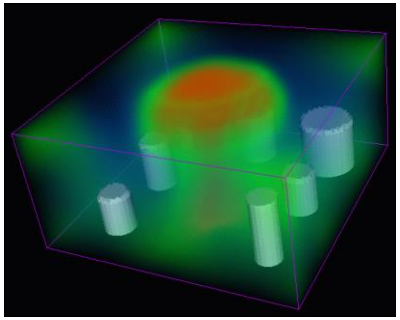




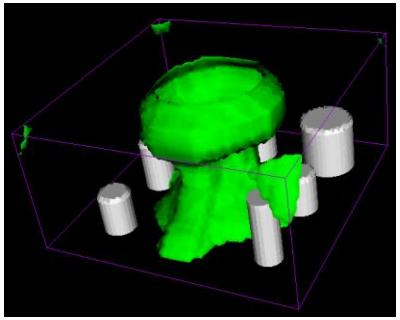
2D surface



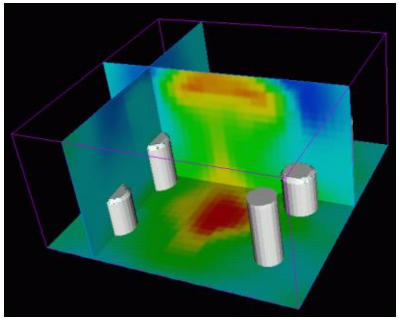
Contours (isolines)



Volume rendering



isosurface



2D slices

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

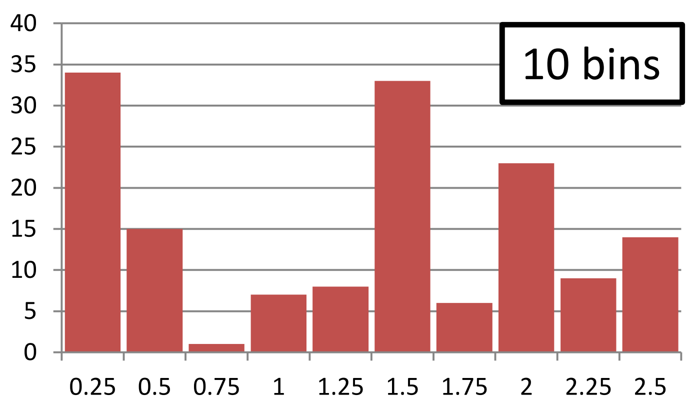
Iris Virginica

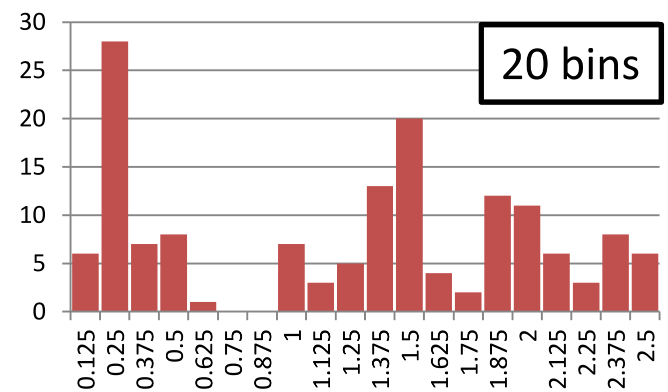
[wikipedia]

Basic Plots

Basic Plots - Bar Charts / Histograms

- For showing distribution of values of a single variable
- Values are divided into bins and a bar plot is used to show the number of objects in each bin
- Height of each bar indicates the number of objects
- Shape of histogram depends on the number of bins

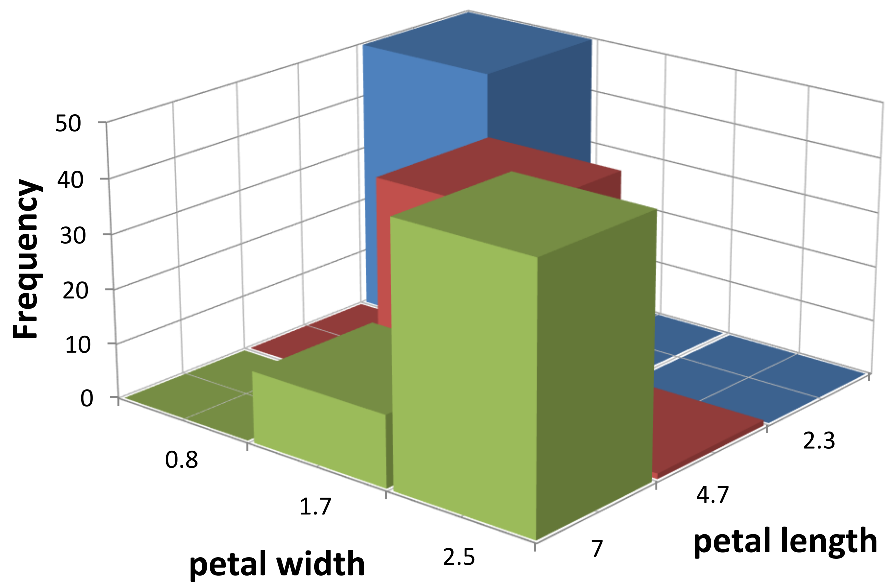




Iris's data – petal width

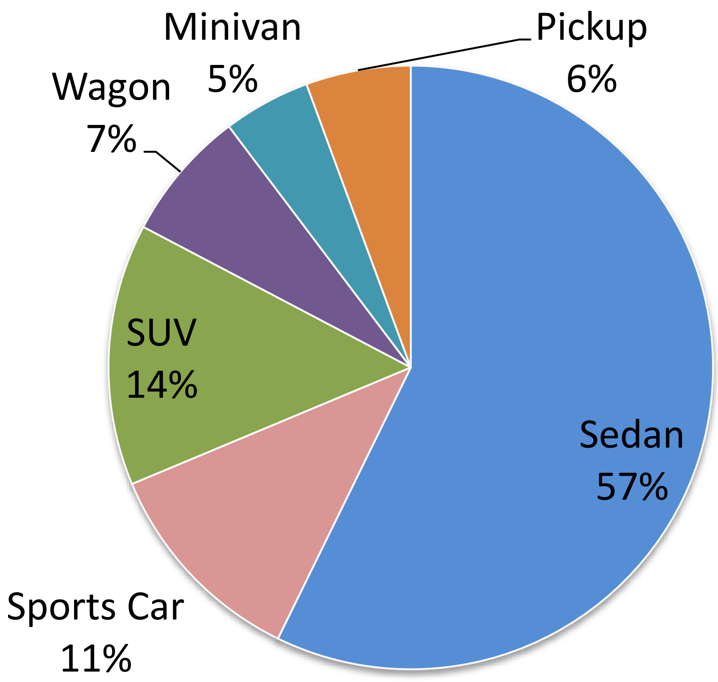
Basic Plots – 2D Bar Charts

- For showing the joint distribution of the values of two variables



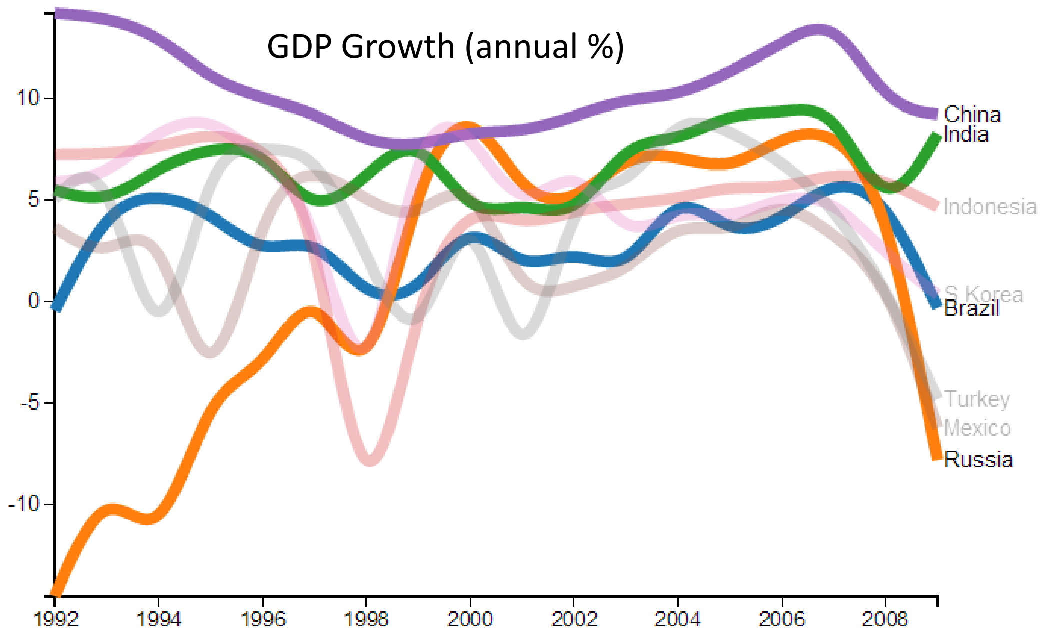
Basic Plots – Pie Charts

- A circular chart divided into sectors, with each sector showing the relative size of each value



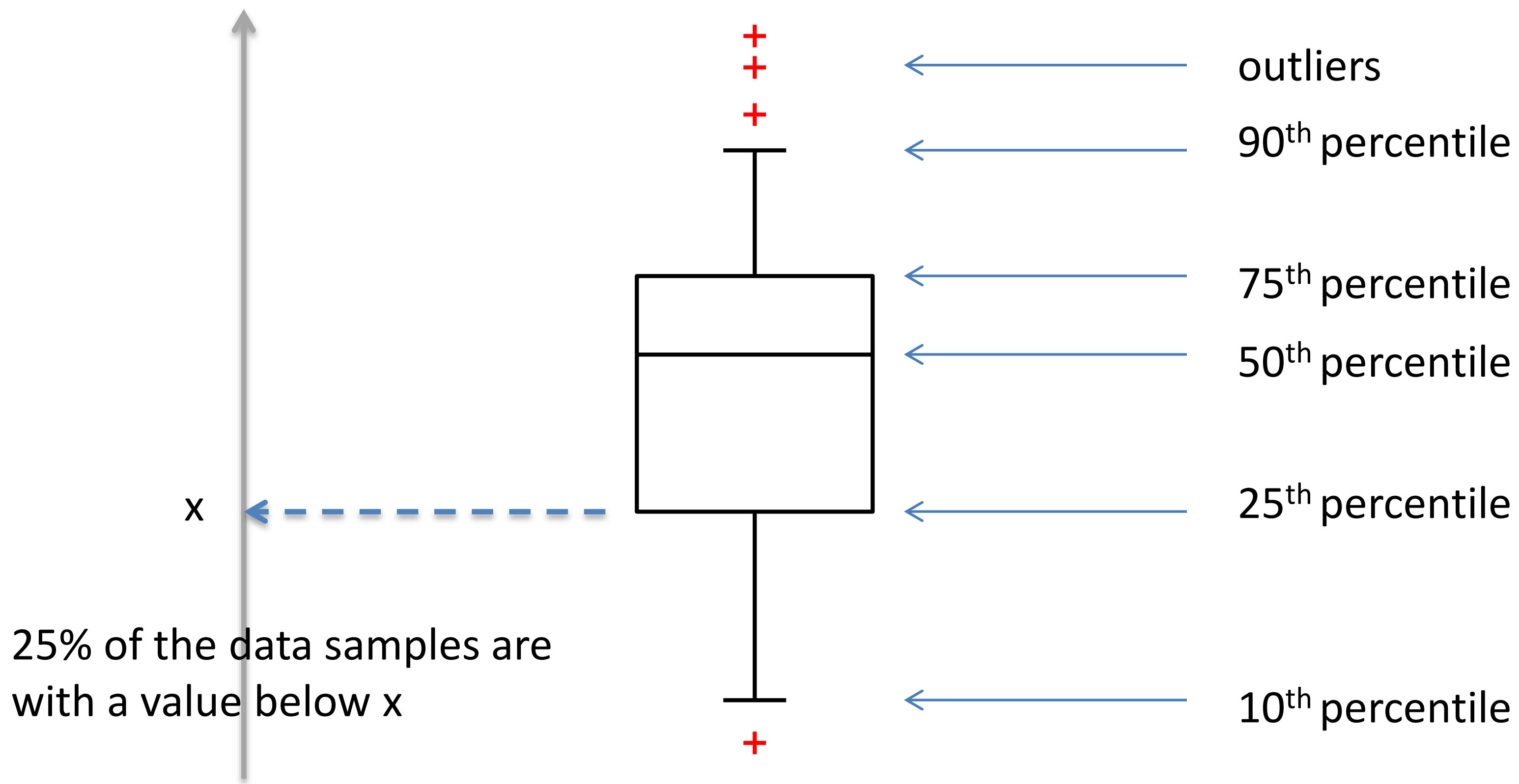
Basic Plots – Line Graphs

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

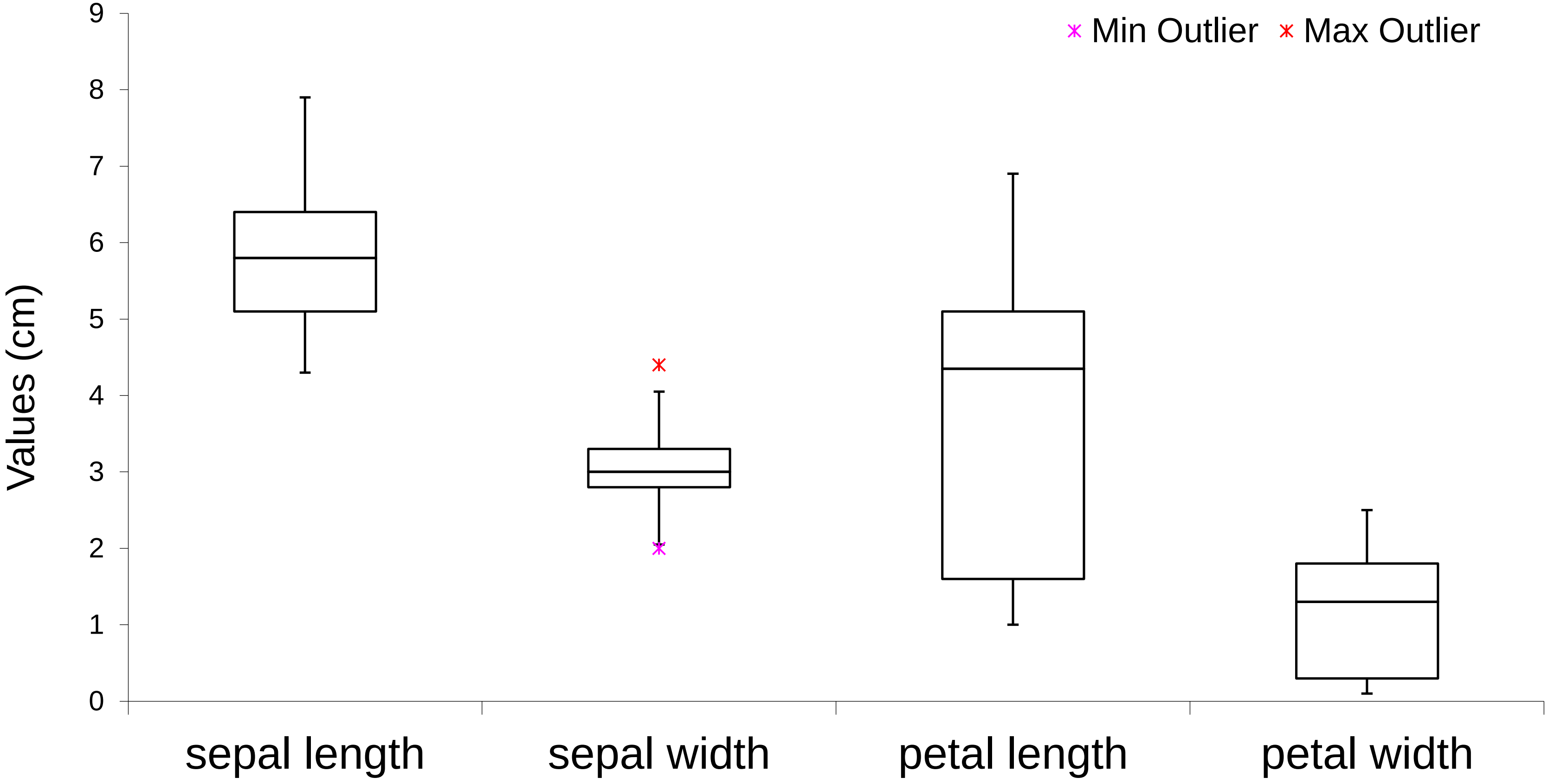


Basic Plots – Box Plots

- For showing the quantitative distribution

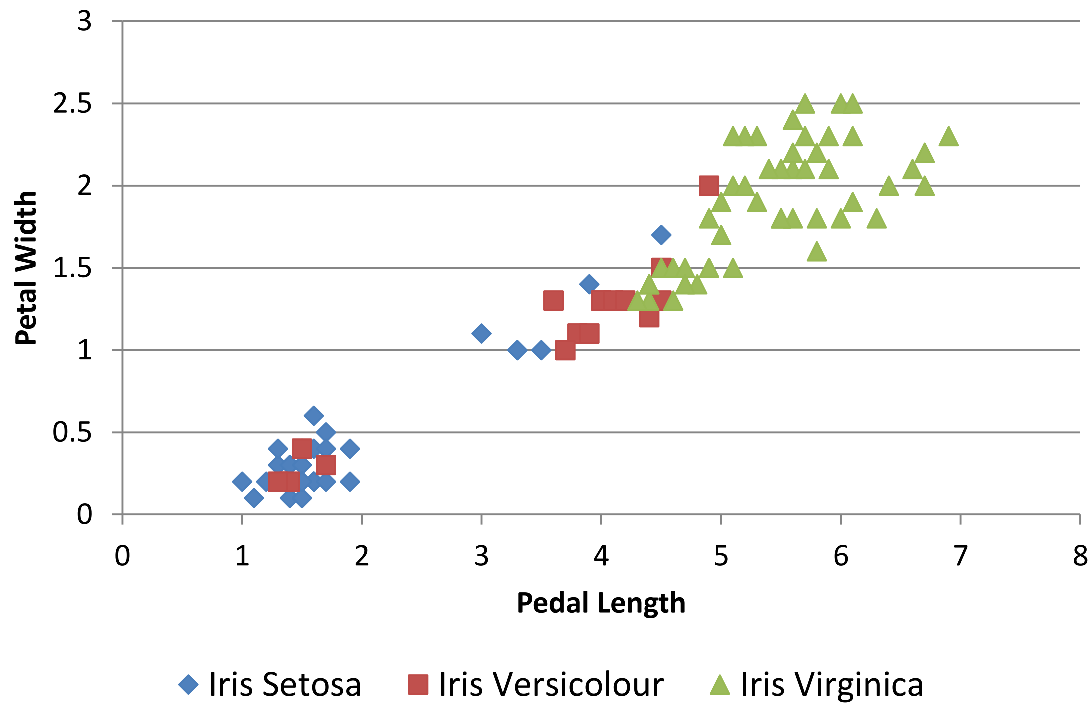


Basic Plots – Box Plots



Basic Plots – Scatter Plots

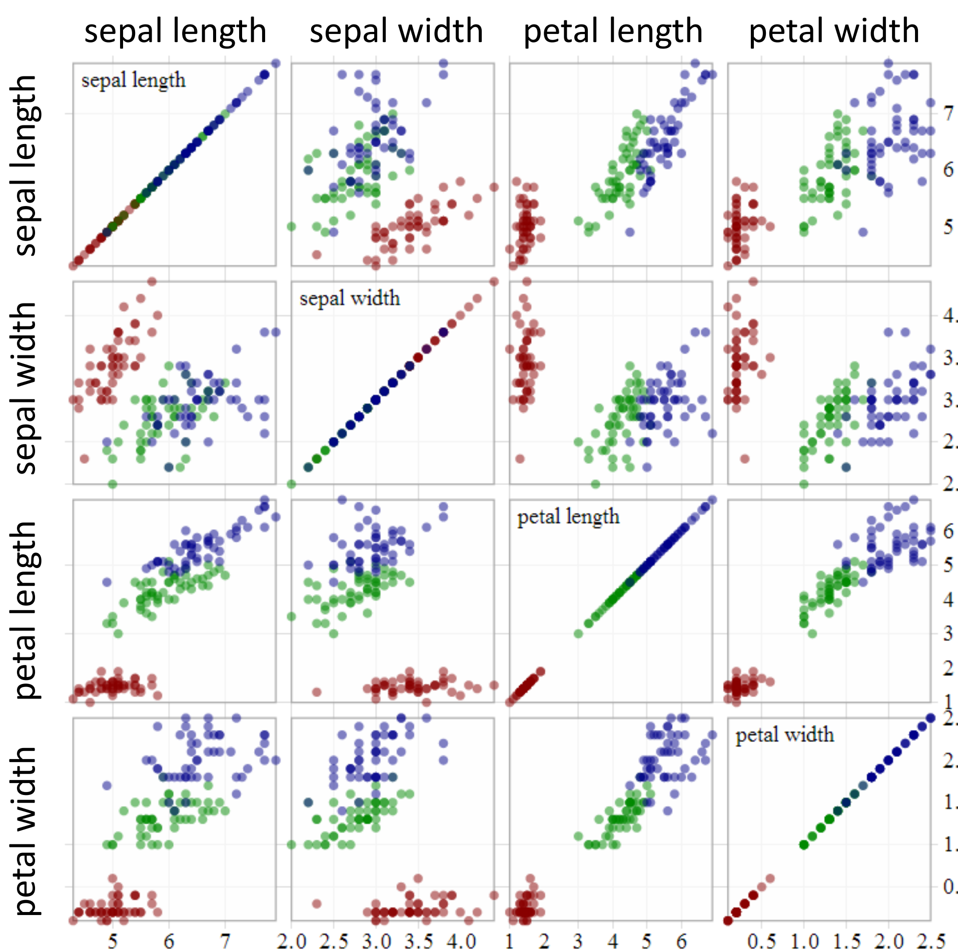
- A plot of points showing the relationship between two attributes or variables



Basic Plots – Scatter Plots

- Point position determined by attribute values
- 2D scatter plots are commonly used, still there are also 3D scatter plots
- Additional attributes can be marked by size, shape, or color for each item
- Useful to have arrays of scatter plots; can compactly summarize the relationships of several pairs of attributes

Basic Plots – Scatter Plot Matrices

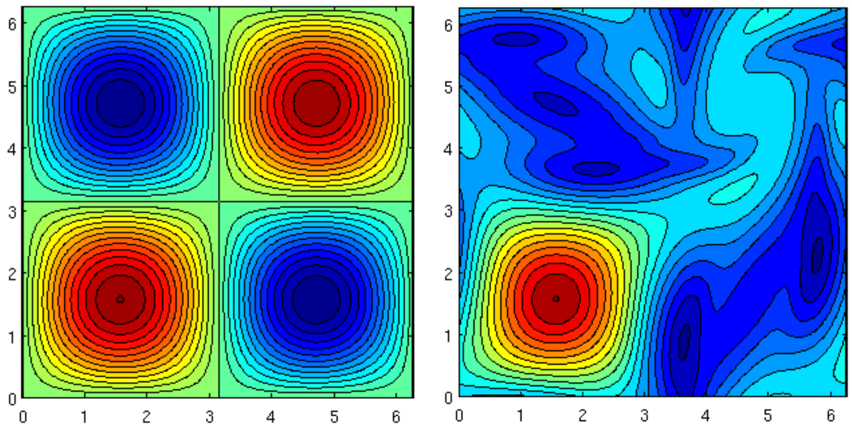


- *Iris setosa*
- *Iris versicolor*
- *Iris virginica*

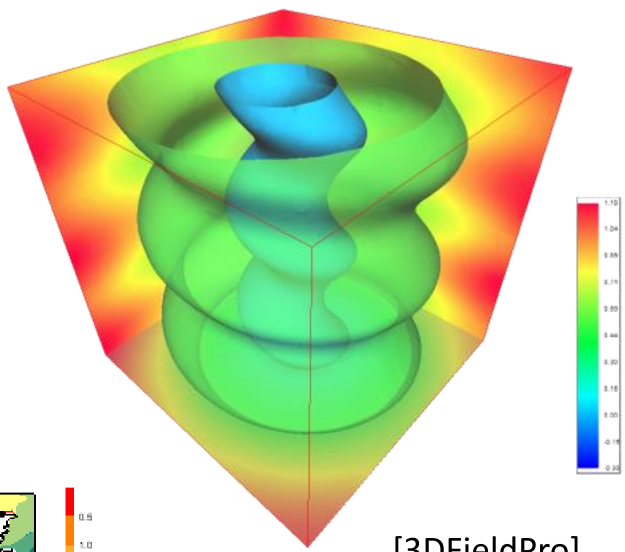
Basic Plots – Contour Plots

- Useful for showing continuous attributes measured on a spatial grid
- Partition the plane into regions of similar values
- Contour lines forming the boundaries of the regions are iso-value lines, or isolines
- Common examples: height fields, temperature, rainfall, etc.

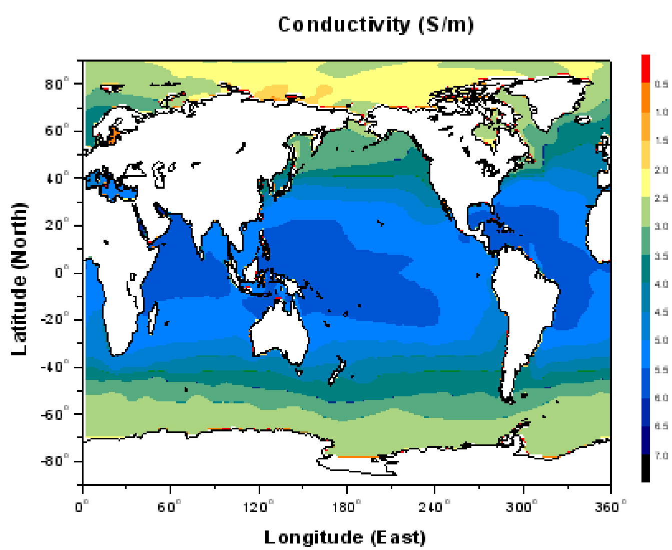
Basic Plots – Contour Plots



[Math, NYU]



[3DFieldPro]



[OriginLab]

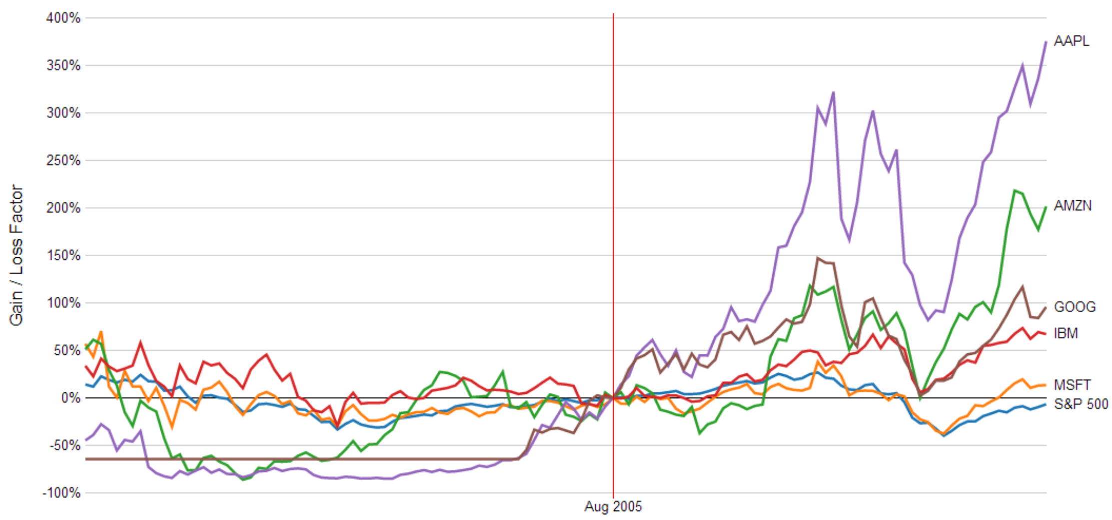
Temporal Data

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)
- Requirements:
  - The ability to compare many time series simultaneously
  - The options of using 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

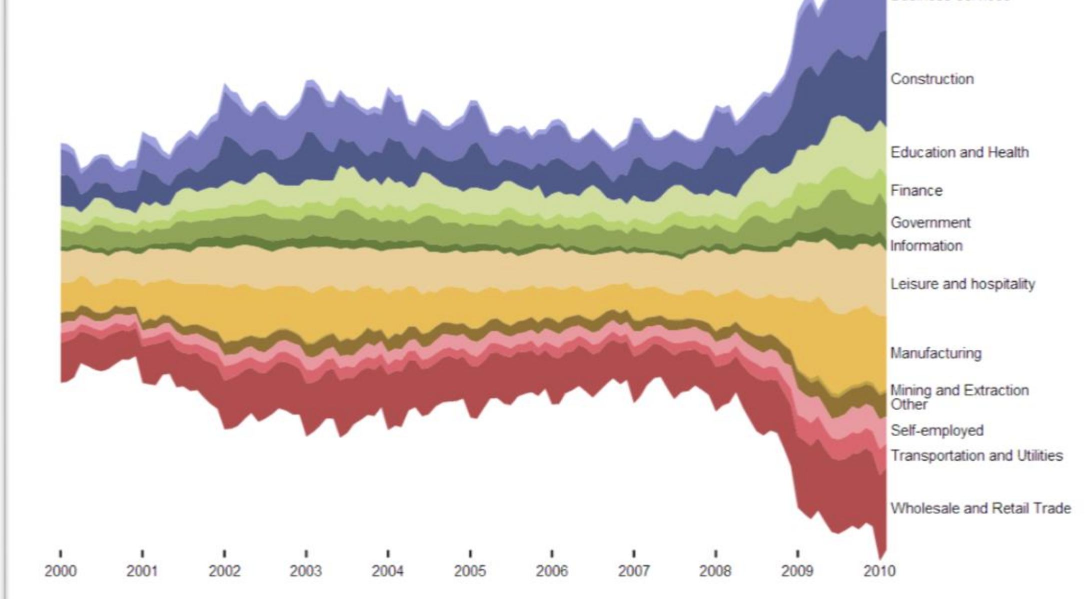
[Heer et al., 2010]

Stacked Graphs

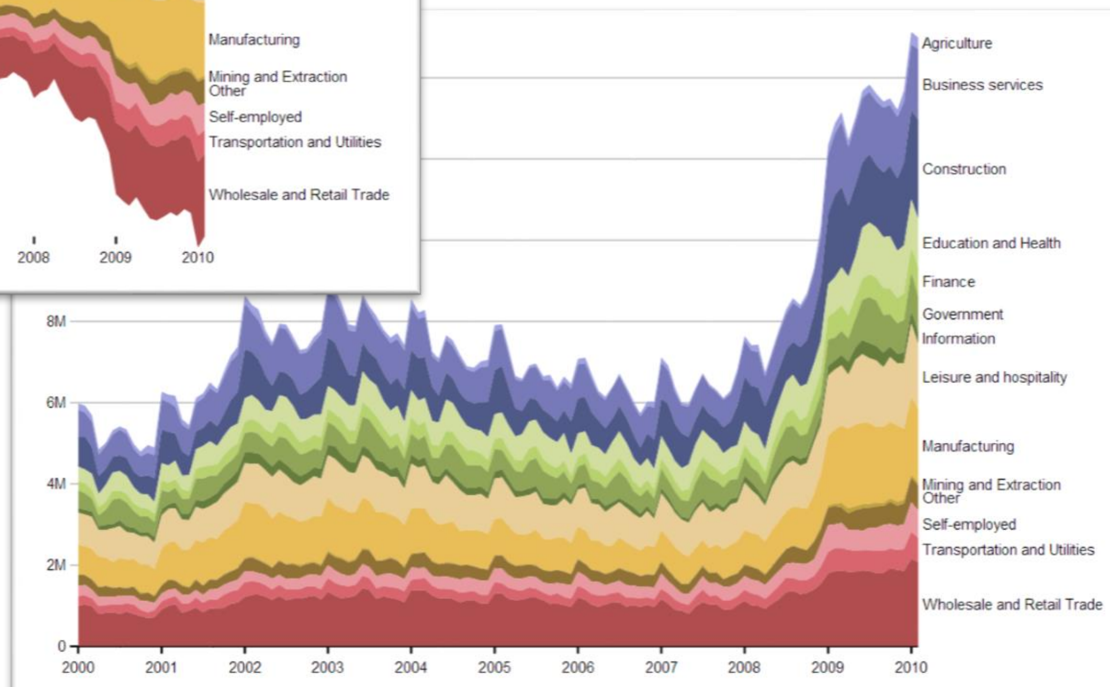
- Stack area charts on top of each other
- Useful for showing summation of time-series values (aggregation)
- Limitation:
  - negative numbers not supported
  - difficult to interpret trends accurately
  - meaningless for some kind of data (e.g., temperatures)

Stacked Graphs

Centered



Zero-Basedline



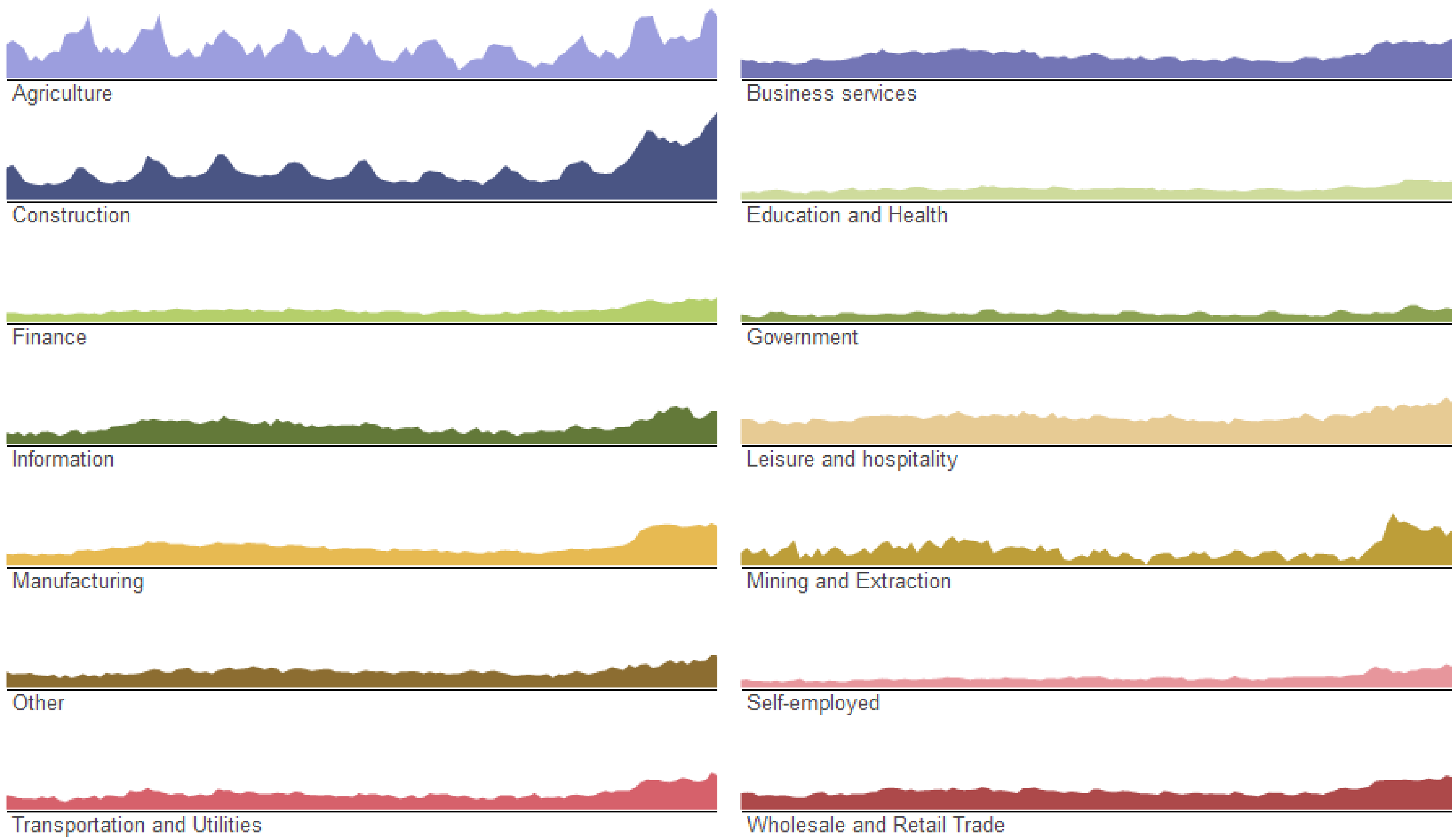
Total counts of unemployed US workers per industry, 2000-2010.

[Heer et al., 2010]

Small Multiples

- To show each series in its own chart
- Useful for avoiding overlapping of multiple curves
- Applies also to other visualization, e.g., bar charts, pie charts, maps, etc. (ref. scattered plot matrix)
- Limitation: Take up a lot of space

Small Multiples



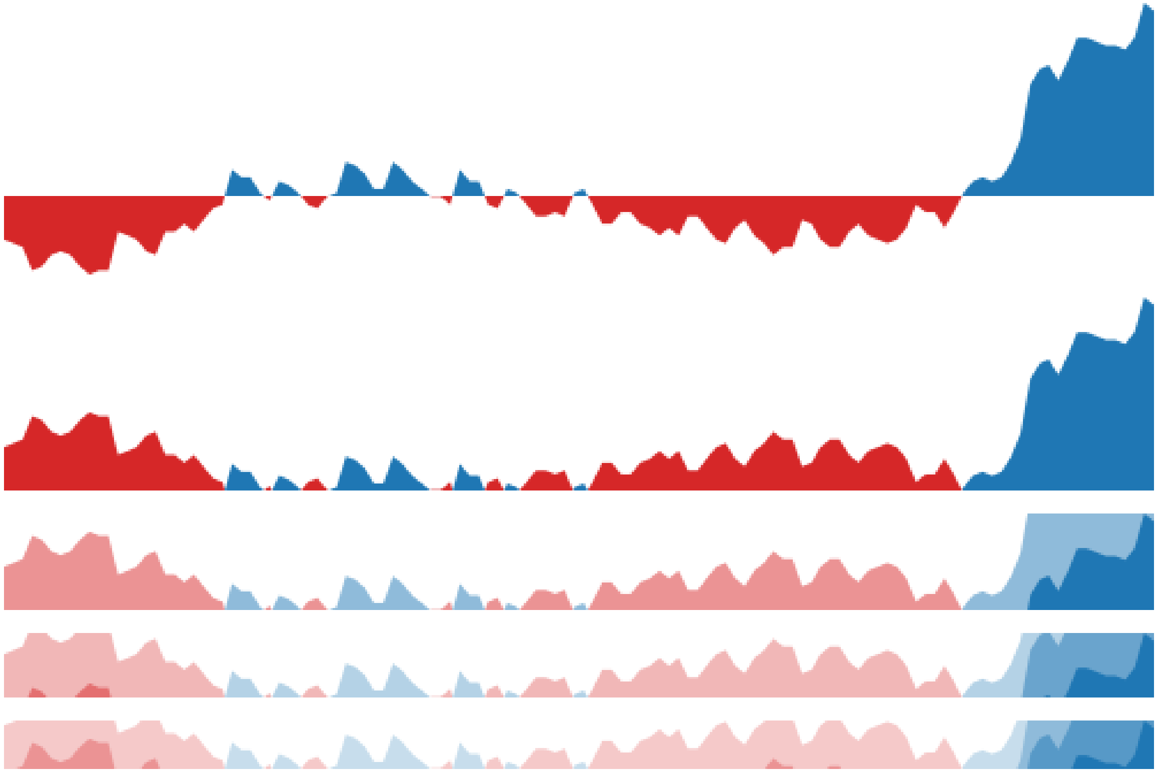
Unemployment rate of US workers per industry, 2000-2010.
Data normalized within each category.

[Heer et al., 2010]

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

Horizon Graphs



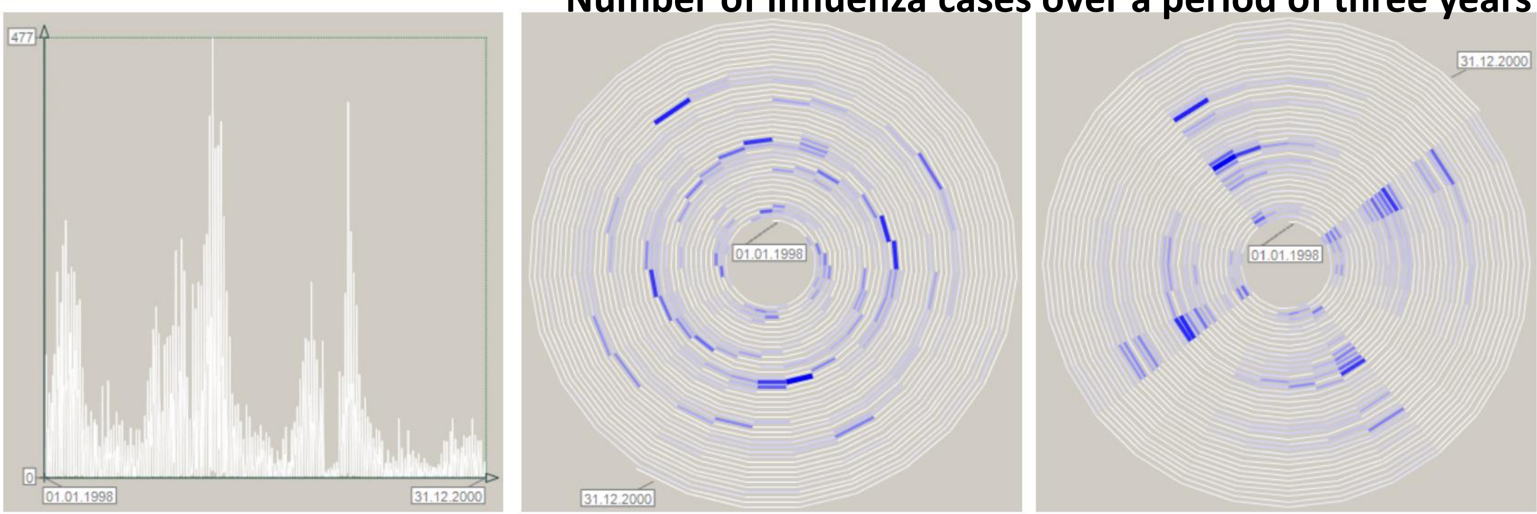
US unemployment rate, 2000-2010.
Positive values: above average unemployment
Negative values: below average unemployment

[Heer et al., 2010]

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.]

Geospatial Data

Geospatial Data

- Data refers to a specific location in the world.
  - e.g., population, health data, traffic, etc.
- Visualization techniques used intensively in geographic information systems (GIS), cartography.
- Issues:
  - Map projection
  - Geographical aggregation
    - Recall the London Cholera Case

Map Projections

- A mapping from a position on Earth (spherical surface) to a position on screen (a flat plane)
- From longitude+latitude pair (λ,φ) to screen coordinates (x,y)

meridians and circles of latitude

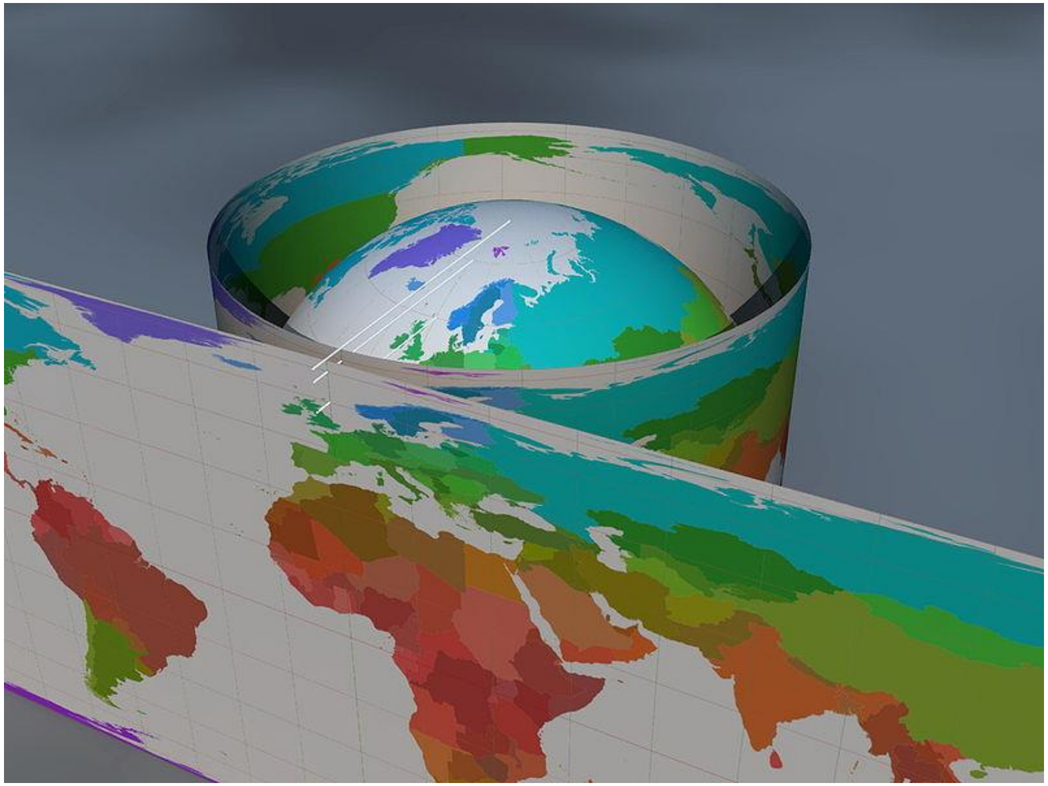
[An Album of Map Projections, U.S. Geological SurveyProfessional Paper 1453]

Map Projections

- Properties of mapping depend on projection methods:
  - Conformal (preserves local angle; not area-preserving)
  - Equivalent (preserves area; shape can change)
  - Equidistance (preserves distance from a specific point or line)
  - Others: Gnomonic (great circles as straight lines), Azimuthal/Retrozimuthal (preserves direction from/to a point)

Cylindrical Projection

- Each point on the sphere surface is projected outward on a cylinder that is put around the sphere.
- Two common cylindrical map projections:
  - Equirectangular projection
  - Lambert cylindrical projection

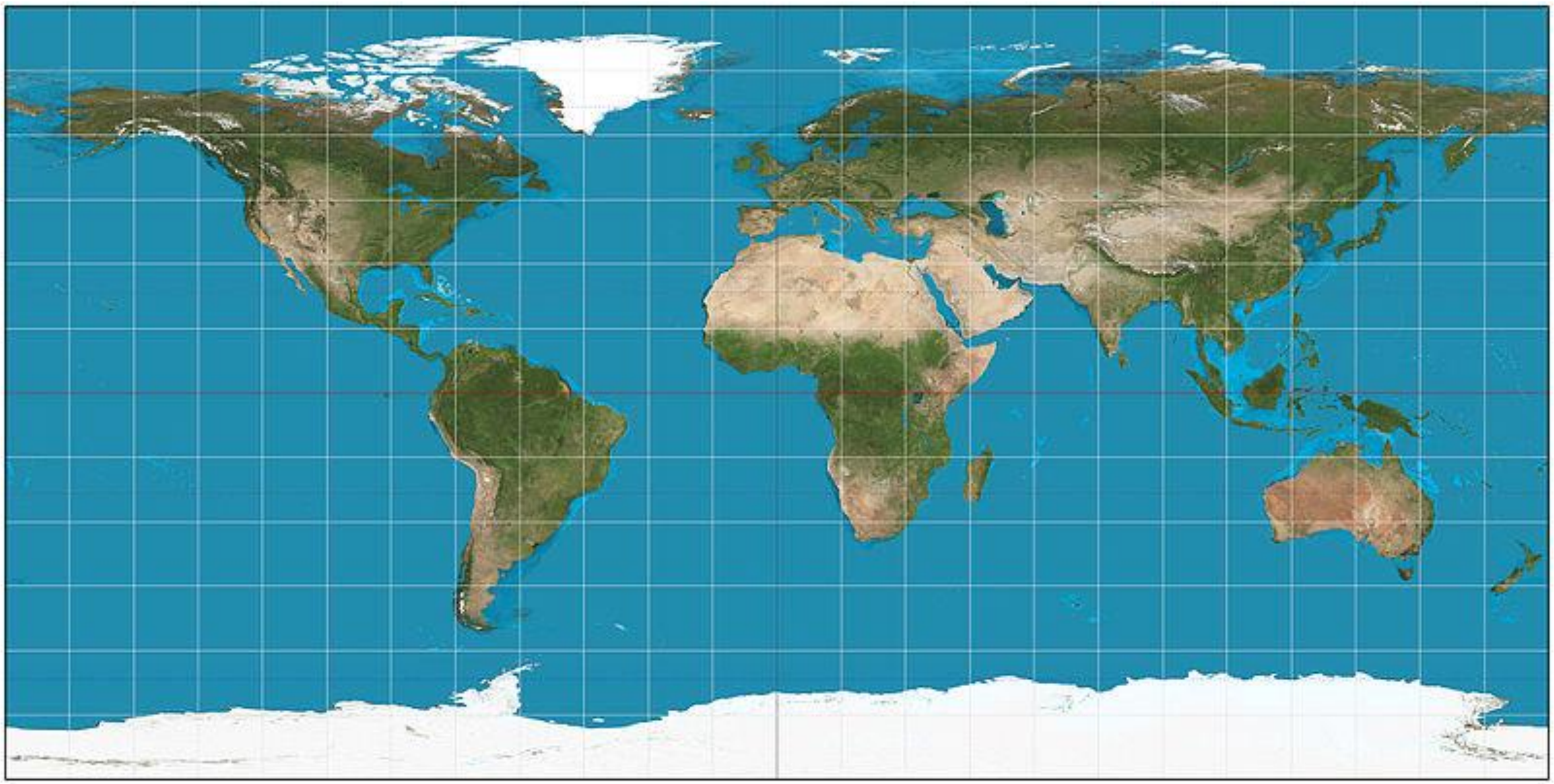


[Wikipedia]

Equirectangular Projection

- Cylindrical Projection
- Meridians are mapped to equally spaced vertical straight lines
- Circles of latitude are mapped to equally spaced horizontal straight lines
- Mapping: $x = \lambda$, $y = \varphi$
- Neither conformal nor equal area, i.e., much distortion
- Use often in thematic mapping, e.g., choropleth map

Equirectangular Projection

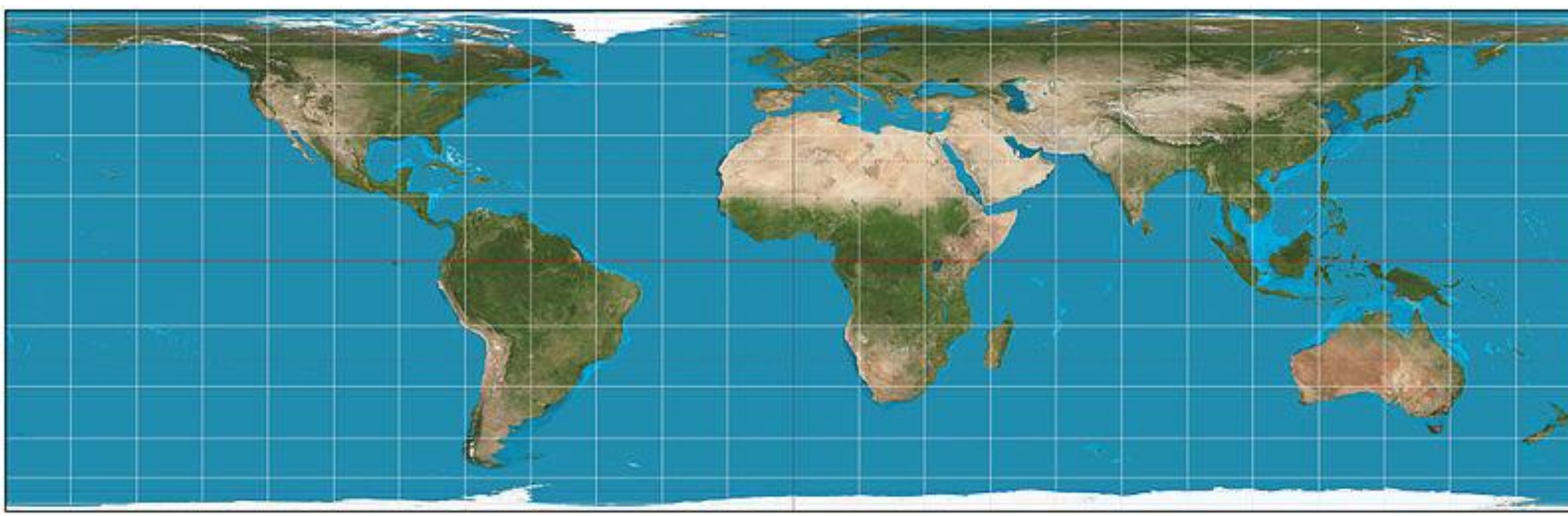


[Wikipedia]

Lambert Cylindrical Projection

- Cylindrical Projection
- Mapping: $x = \lambda$, $y = \sin \varphi$
- Area preserving
- Undistorted along equator, but highly distorted near the poles.

[Wikipedia]



Flow Maps

- For showing the movement of a quantity in space.
- Flow lines encode multivariate information in attributes such as path points, direction, line thickness, color, etc.

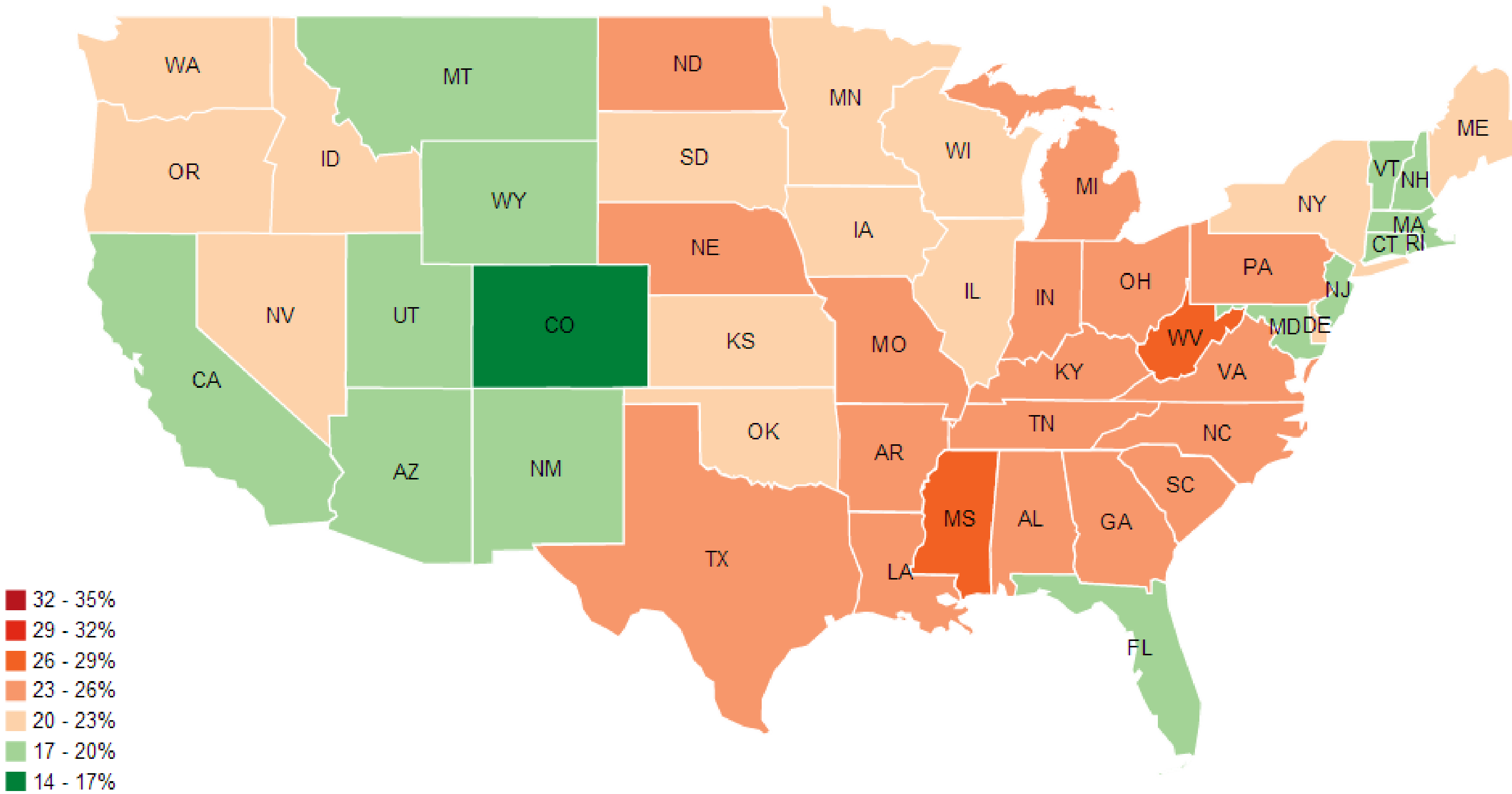
The French invasion on Google map

[Heer et al., 2010]

Choropleth Maps

- For showing data collected or aggregated by geographical areas
- In Greek: choro = area, pleth = value
- Use color to encode values for a region
- Be careful about normalization of data and color mapping
- Problem: tends to highlight patterns in large areas, while highly populated but small areas might be of more interest

Choropleth Maps



Obesity in the US, 2002

[Heer et al., 2010]

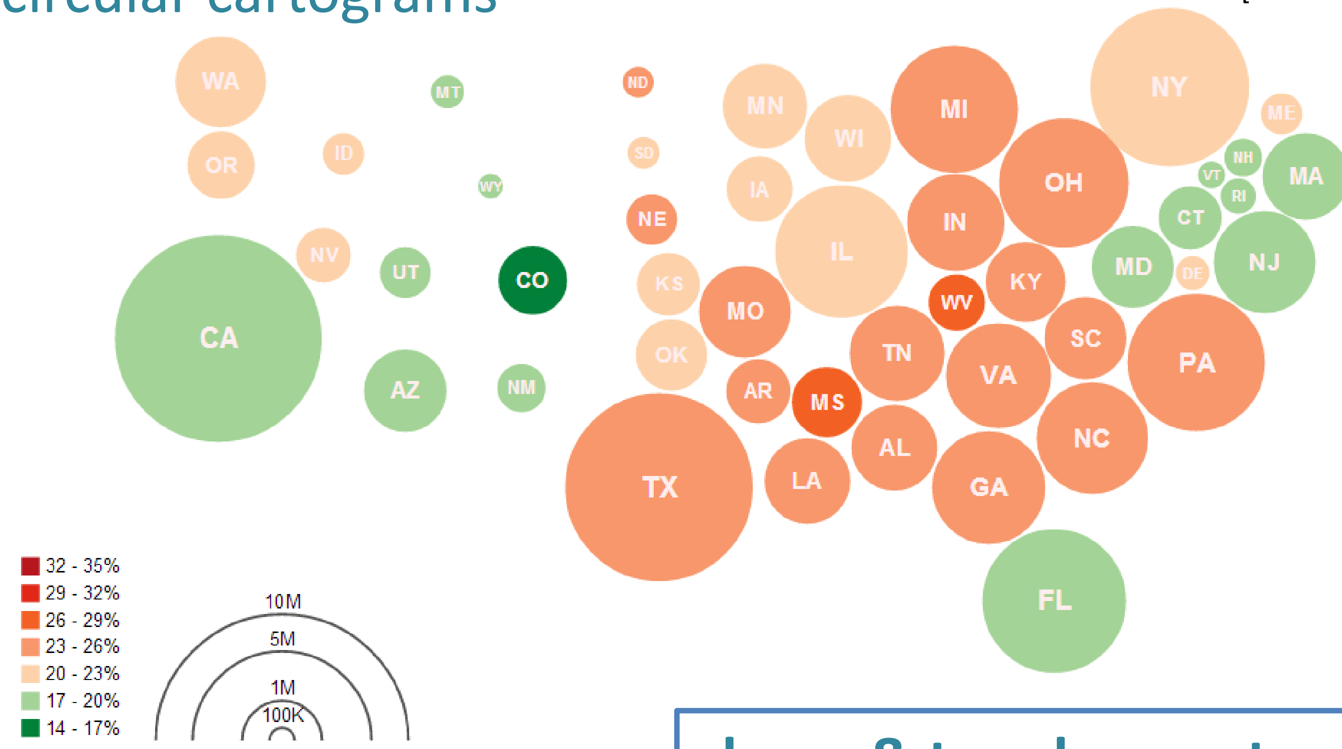
Cartograms

- Regions are resized so that the area directly encodes a data variable
- Cartograms differ by the properties of:
  - Shape preservation
  - Exact area correspondence
  - Topology preservation (i.e., region connectivity)
- An optimization problem to find a good compromise between the above conflicting criteria

Dorling Cartograms

a.k.a. circular cartograms

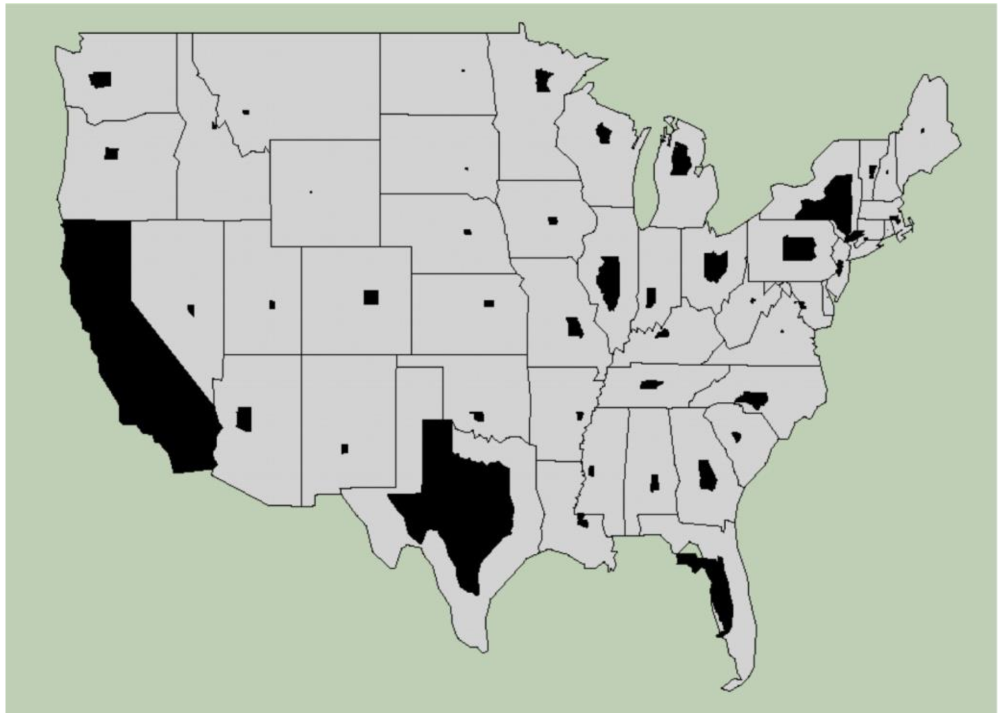
[Heer et al., 2010]



shape & topology not retained

Obesity in the US, 2002
Color: % of obese people
Circle size: absolute number of obese people

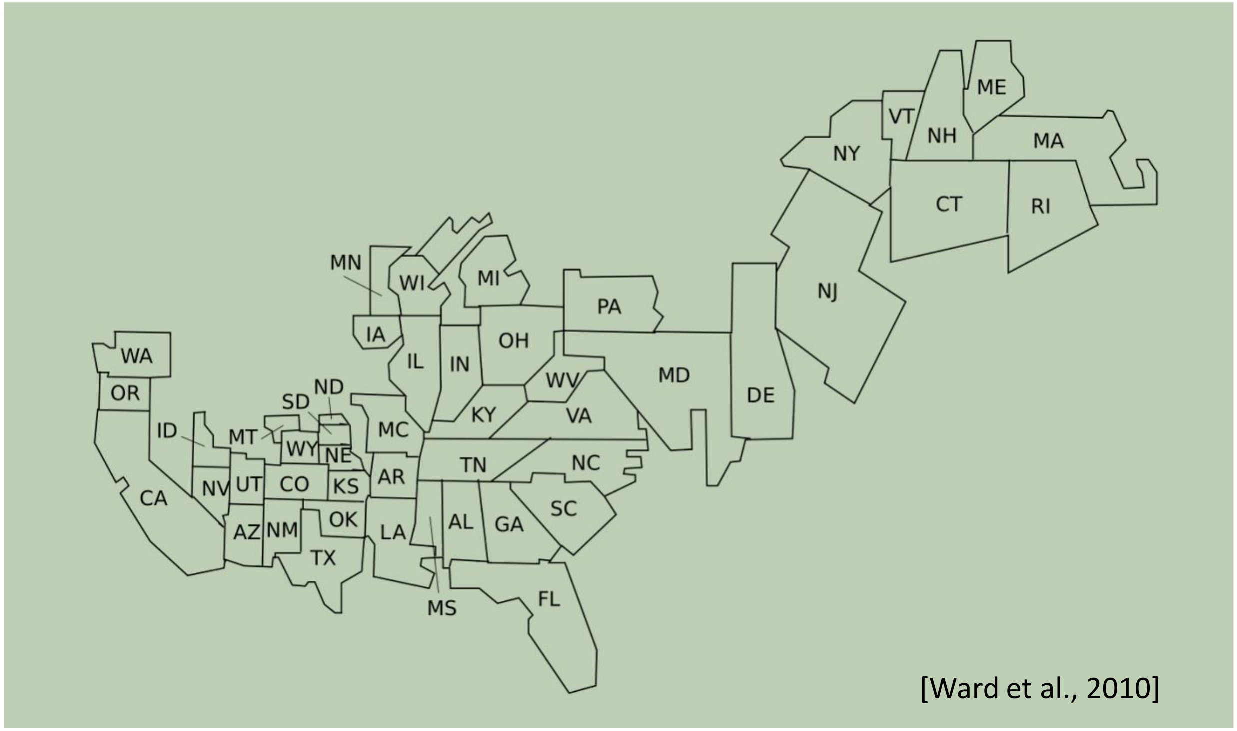
Noncontinuous Cartograms



[Ward et al., 2010]

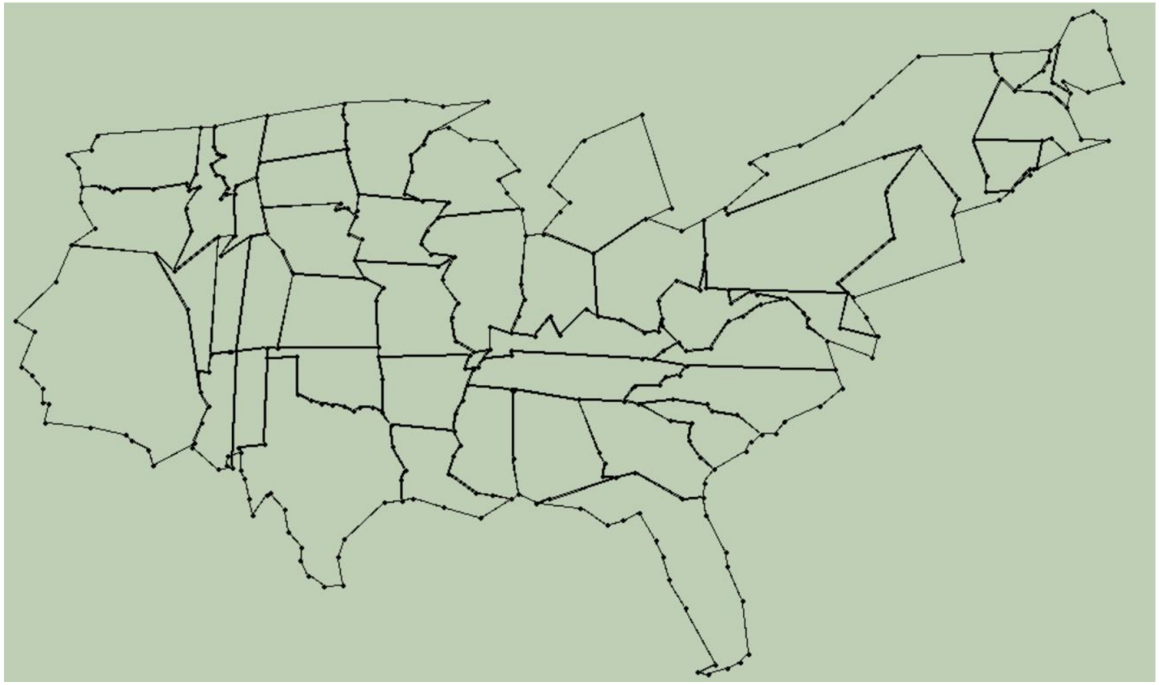
- Exact area, preserves shape
- Not preserving topology, map perception still ok
- Size limited by the maximum scaling w.r.t. map region

Noncontiguous Cartograms



- Exact area, preserves shape as much as possible
- Not preserving topology
- Map perception more difficult

Continuous Cartograms

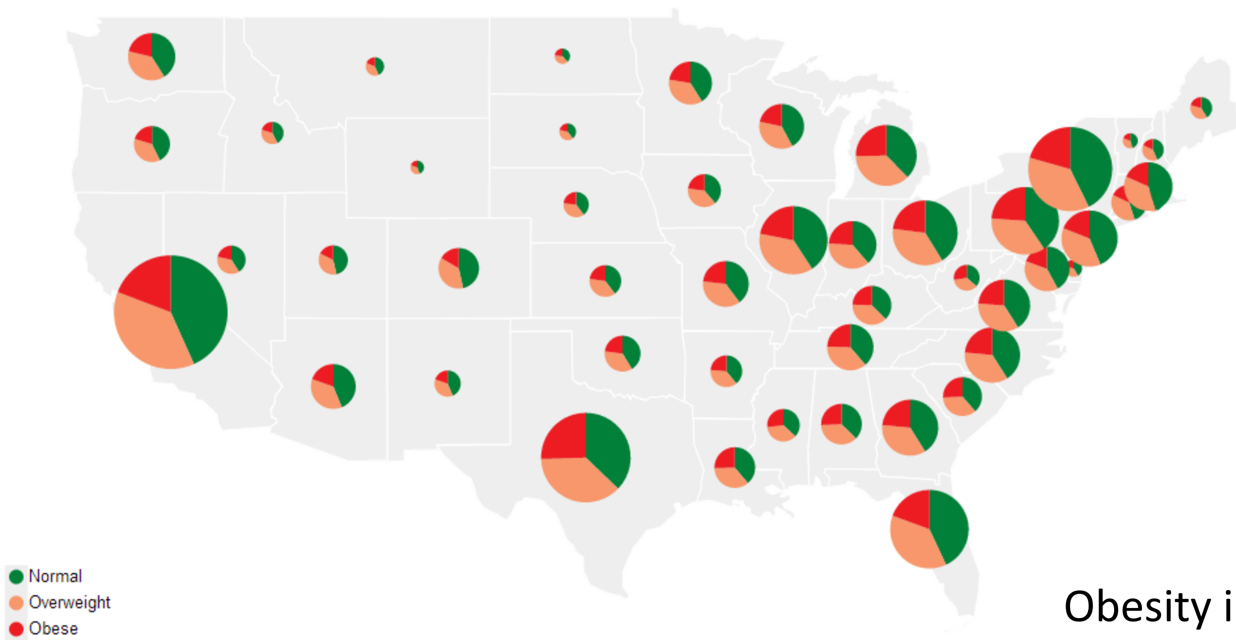


[Ward et al., 2010]

- Preserves topology
- Not preserving area & shape
- Takes a long time to compute the visualization, interactive data change is not possible

Graduated Symbol Maps

- Data is showed by placing symbols over a map
- More dimensions can be visualized by encoding with the attributes of the symbol



[Heer et al., 2010]

Obesity in the US, 2002

Multivariate Data

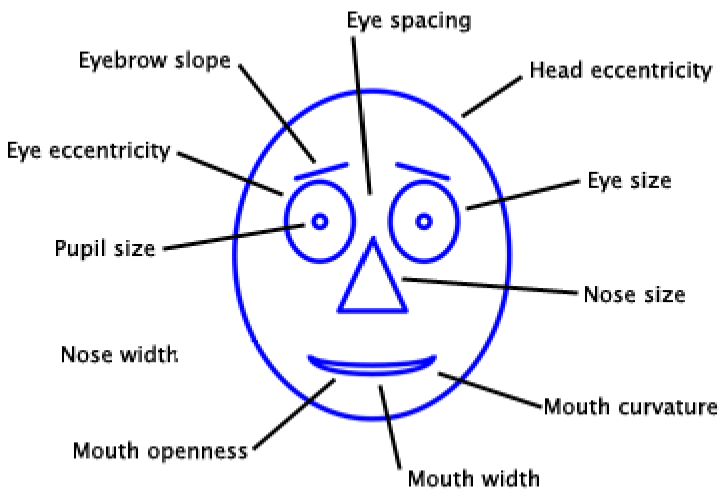
A Variety of Methods

- Icon-based
  - Chernoff faces, …
- Table-based
  - Heatmaps, Table Lens, …
- Geometric
  - Parallel coordinates, scatterplot matrices, …
- Hierarchical
  - Dimensional stacking, worlds-within-worlds, …

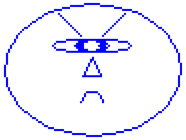
Chernoff Faces

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

Chernoff Faces

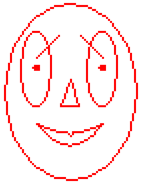


10 facial features, each corresponds to a parameter in [0,1]



All 0

All 0.5

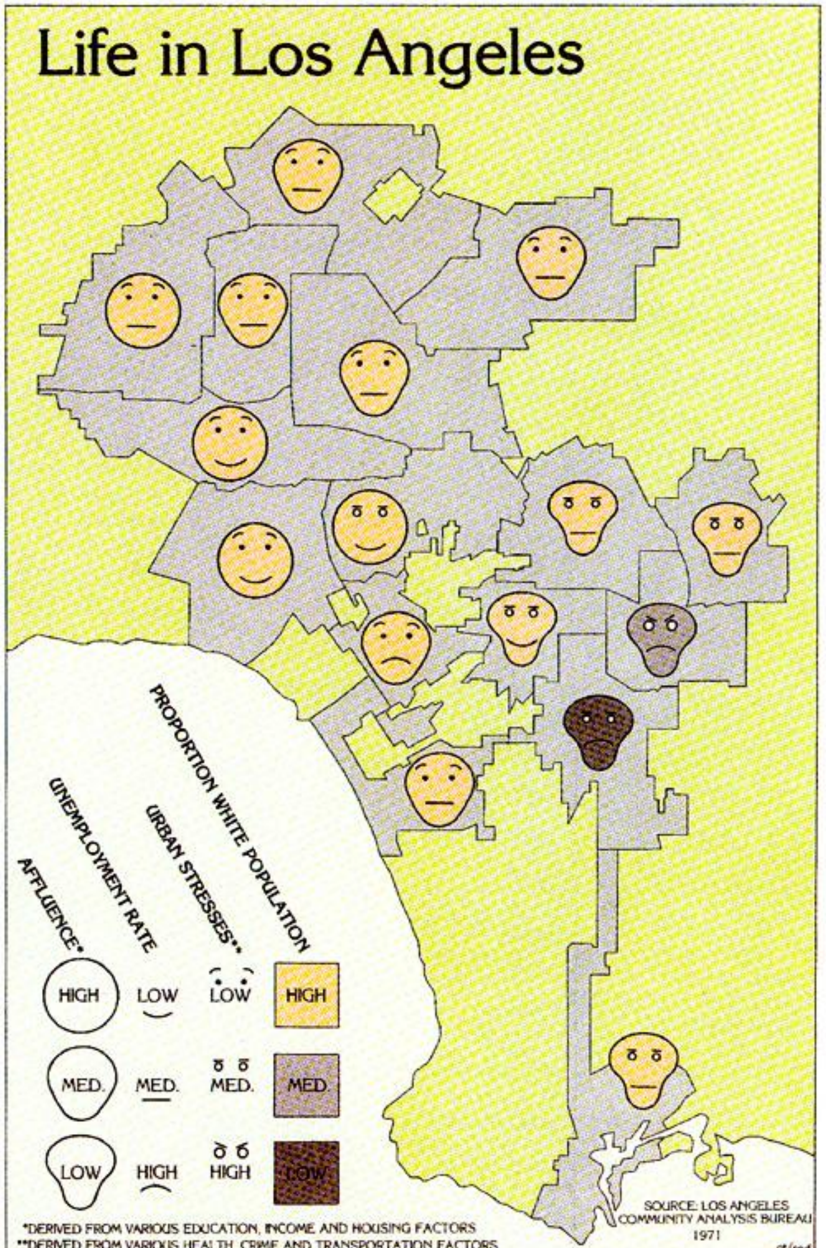


All 1

Random parameters

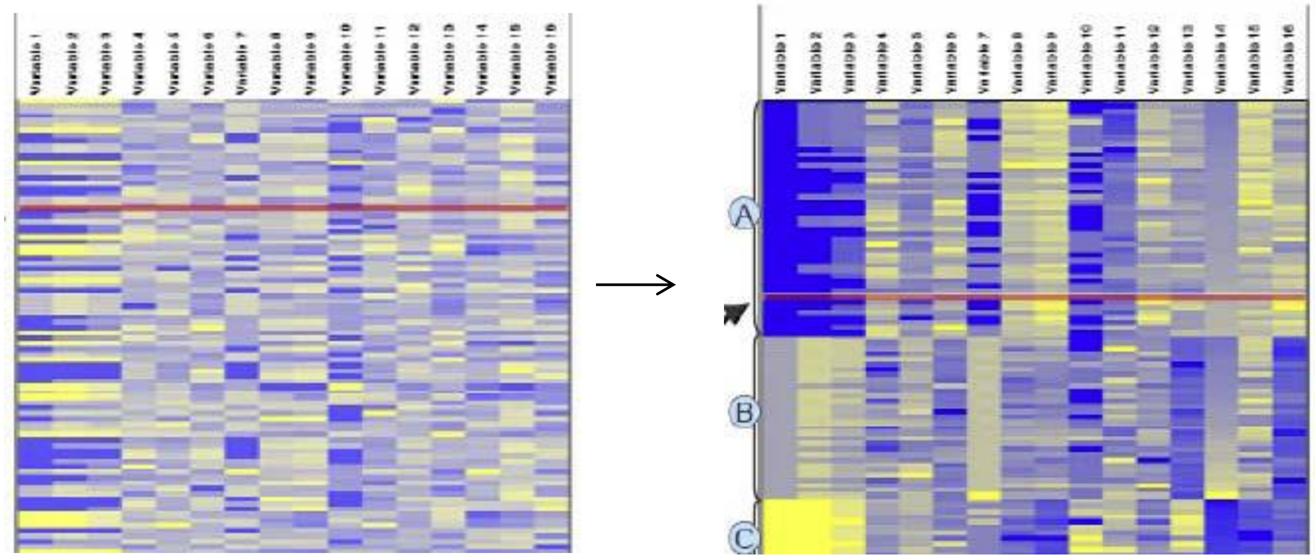
[http://kspark.kaist.ac.kr/Human%20Engineering.files/Chernoff/Chernoff%20Faces.htm]

[http://kspark.kaist.ac.kr/Human%20Engineering.files/Chernoff/Chernoff%20Faces.htm]

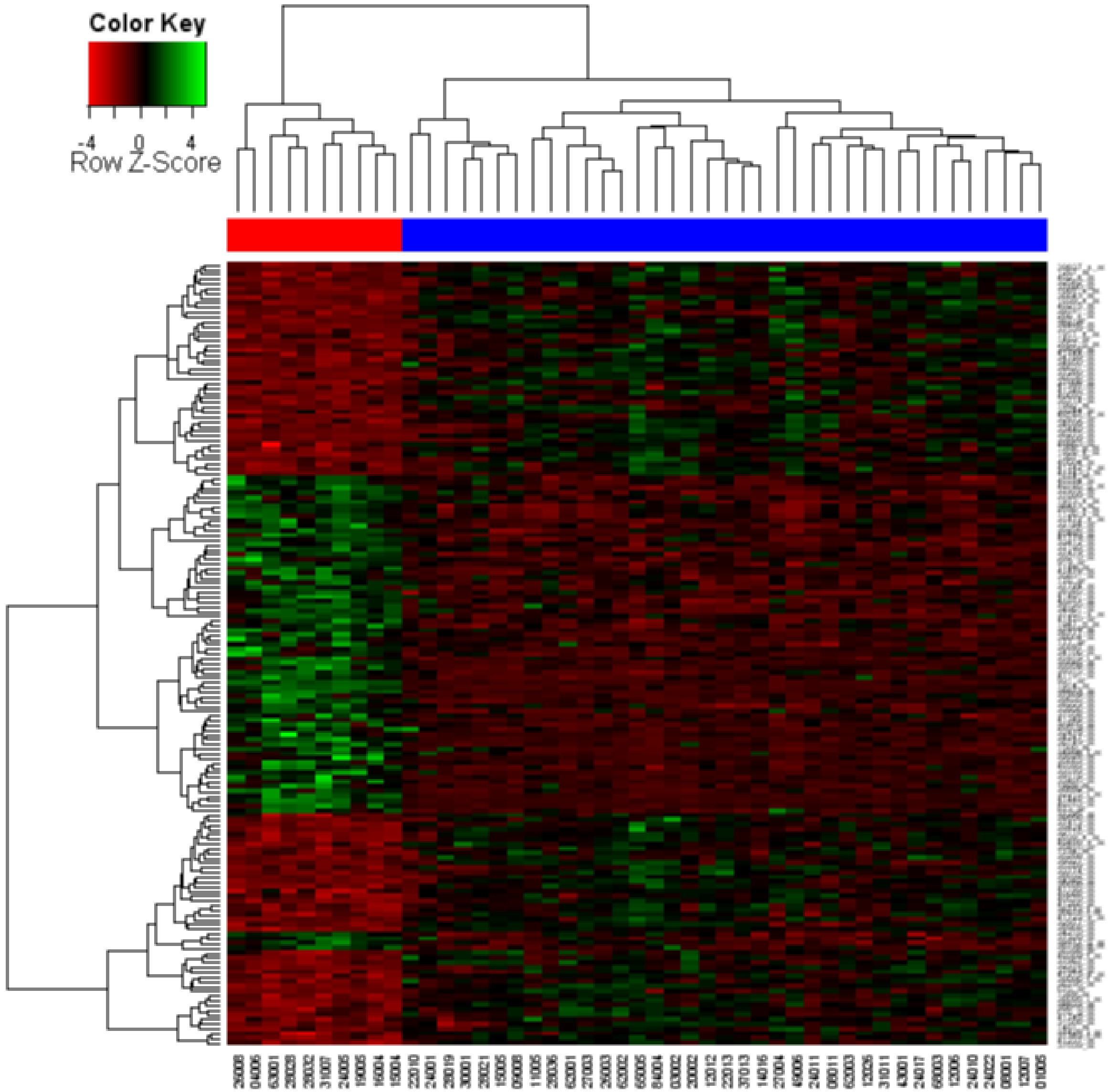


Heat Maps

- Map values of table entries to colors
- Rows and columns can be reordered to expose features.



[M. Ward]



column: patient
row: DNA test

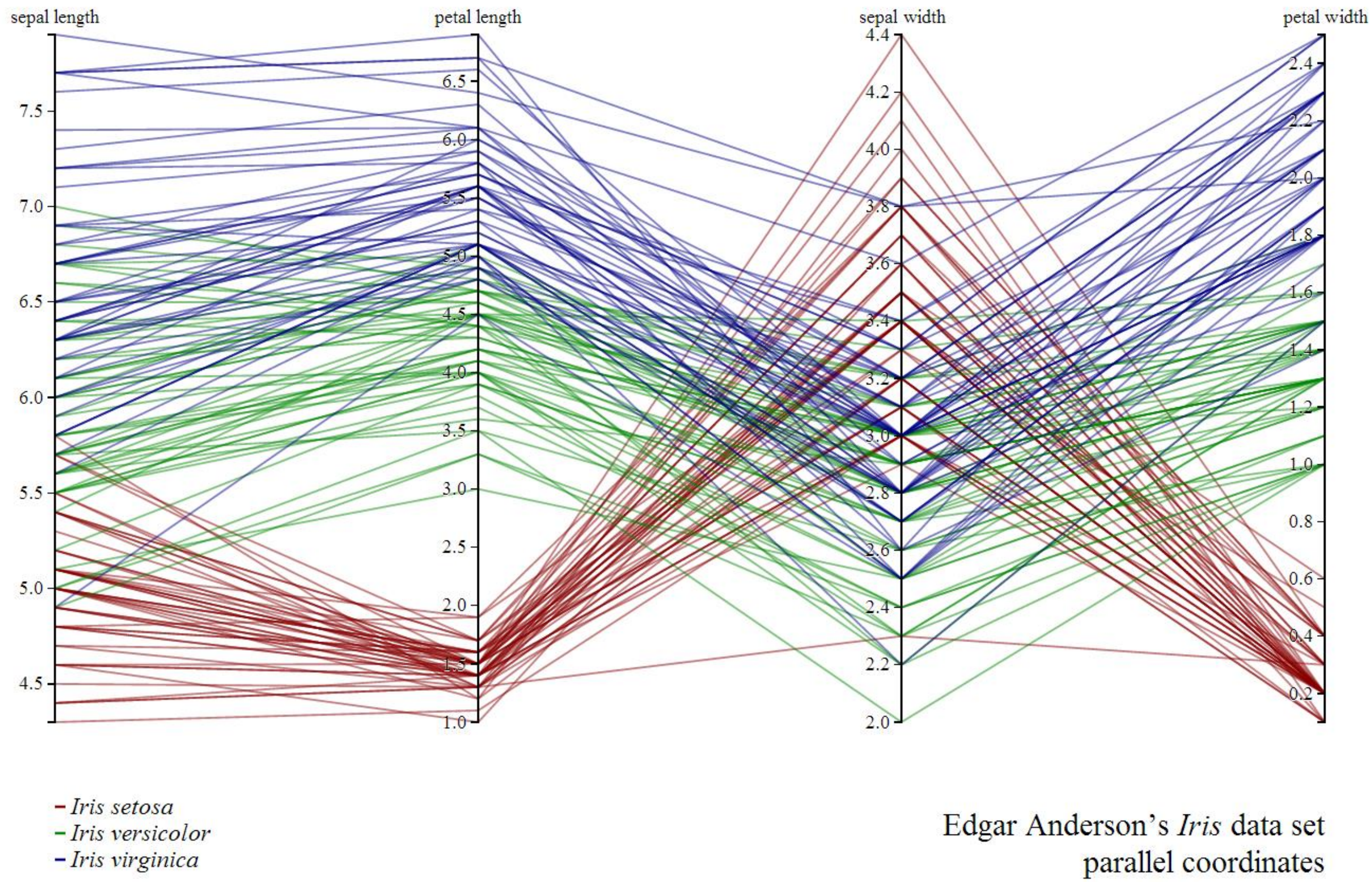
Heatmap from DNA microarray data showing genes expressed differentially for two types of leukemia.

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

Parallel Coordinates

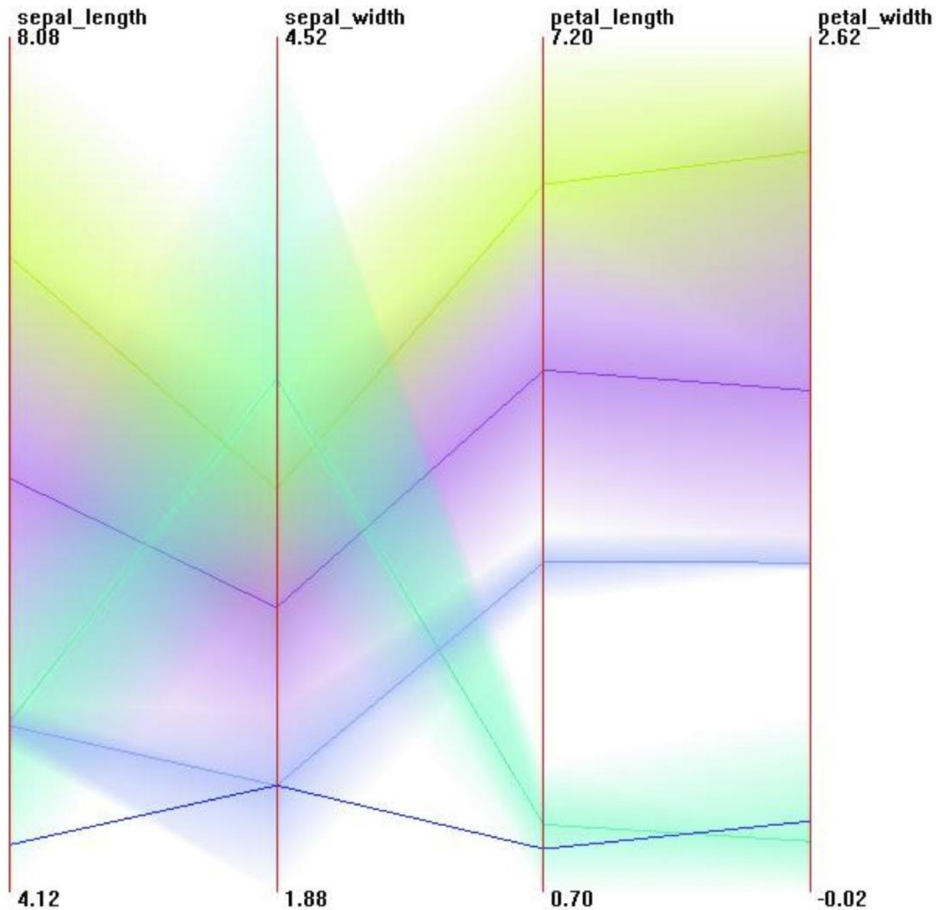
- How to present all *n* axes of the *n* dimensions on a 2D plane?
- Use parallel axes instead of orthogonal axes
- Each attribute value of a data item corresponds to a point on a coordinate axis, and the data item is represented as a polyline connecting these points
- A distinct class of objects can sometimes be seen as a group of lines on some axes
- Ordering of axes is important

Parallel Coordinates



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

Parallel Coordinates

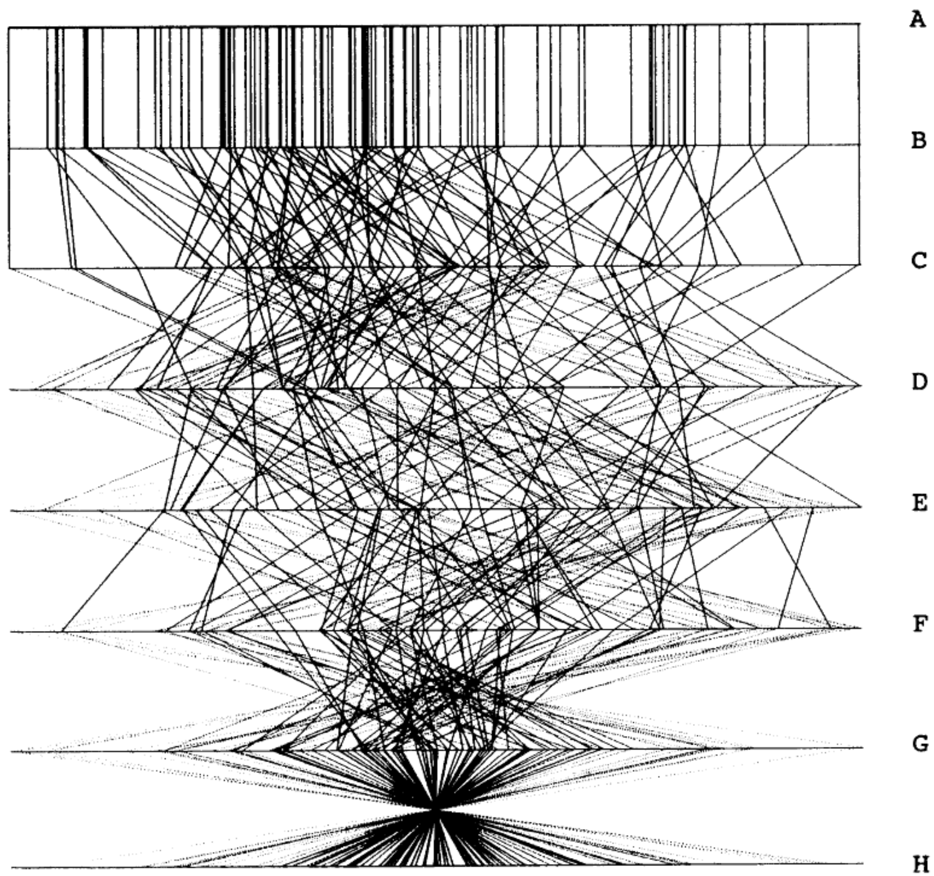


[Ward 2010]

PC plot showing the centers and extents of clusters

Parallel Coordinates

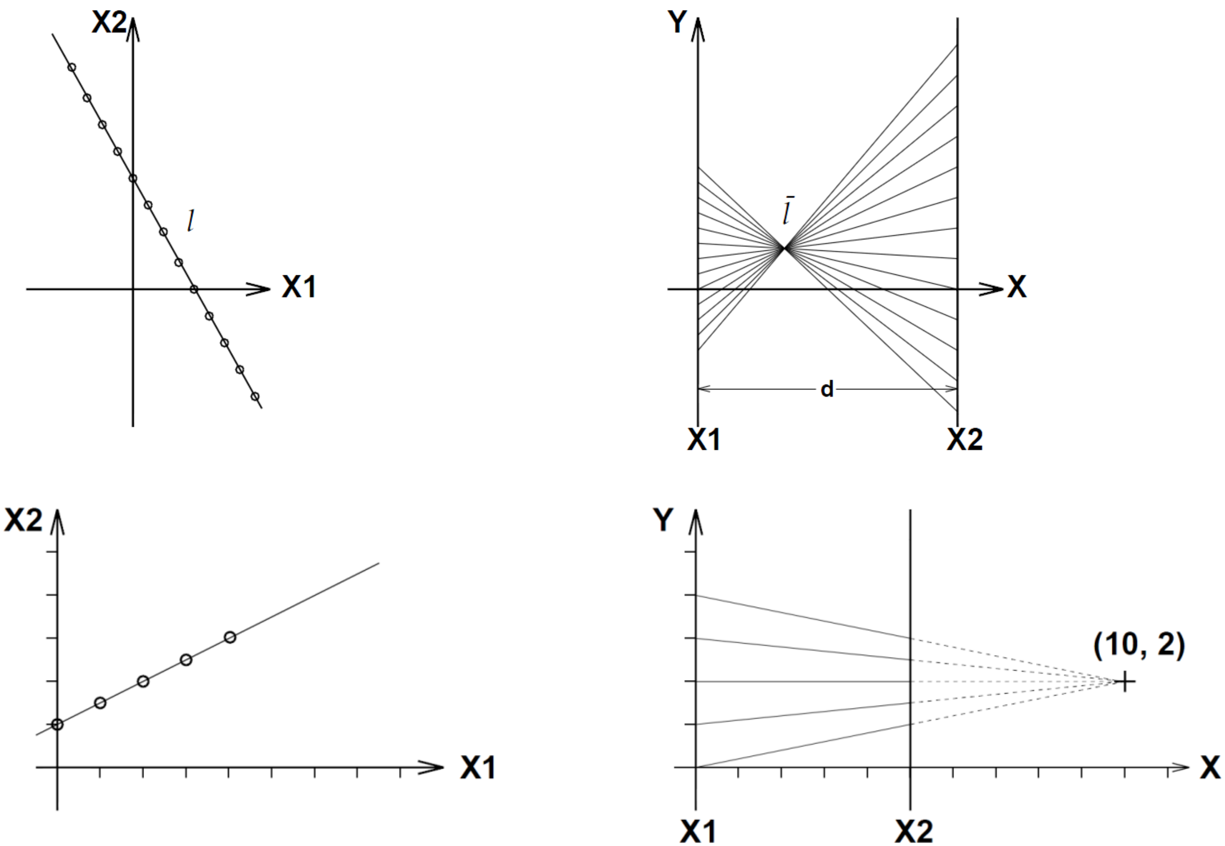
- Parallel correlation



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

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

Parallel Coordinates



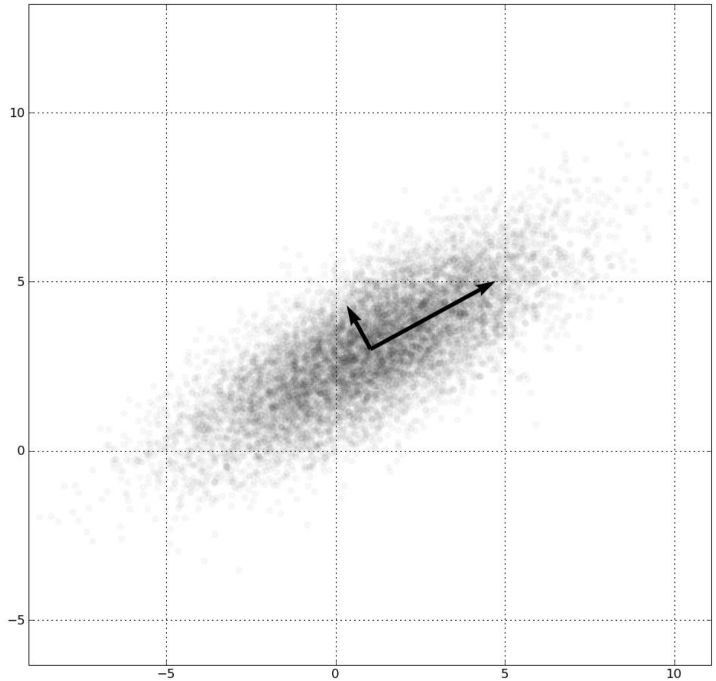
[Wong and Bergeron, “30 Years of Multidimensional Multivariate Visualization,” *Scientific Visualization: Overviews, Methodologies & Techniques*, 1997.]

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

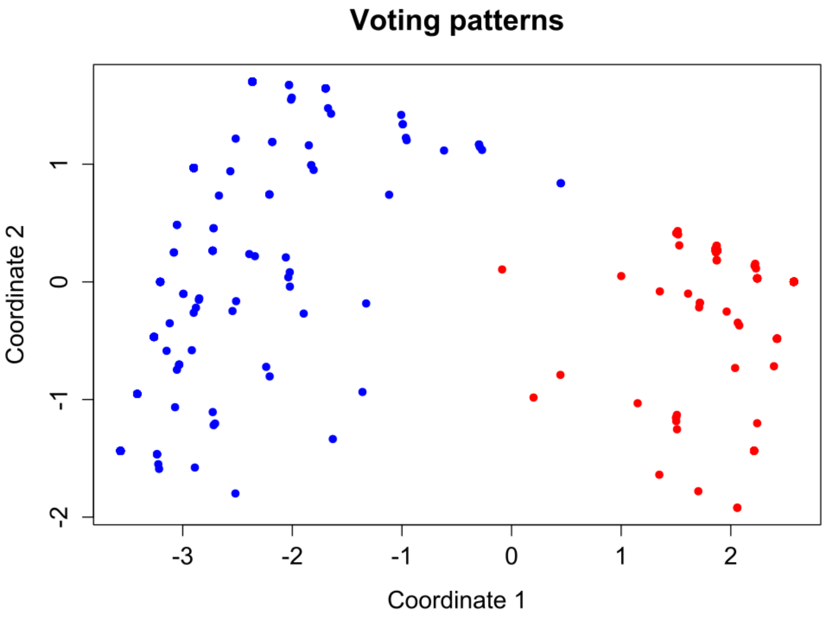
Examples

PCA



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

MDS



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

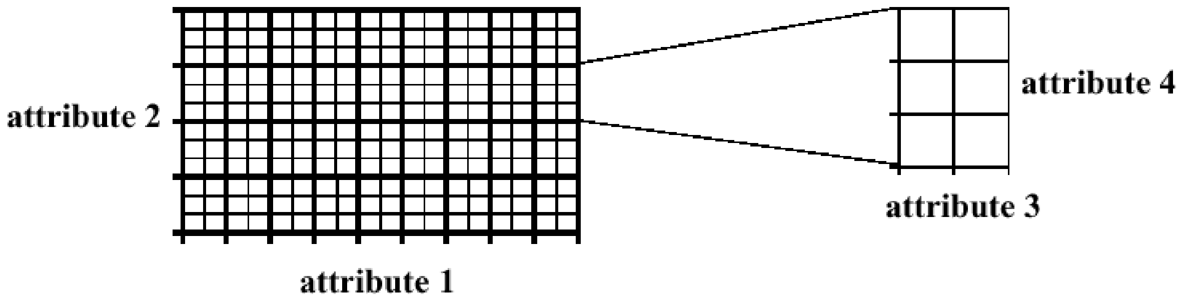
Dimension Ordering

- 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
- 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

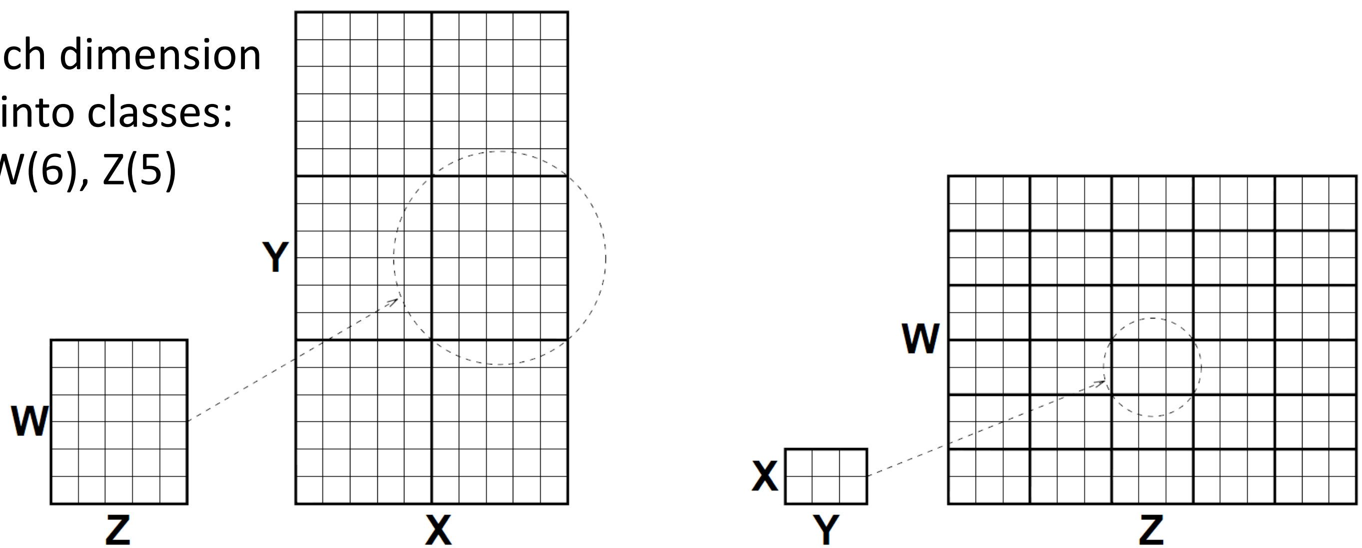
Dimensional Stacking

- The nD attribute space is partitioned in 2D subspaces which are recursively embedded (or ‘stacked’) into each other
- The range of the attribute values in each dimension is partitioned into classes
- Important attributes should be used on the outer levels
- Adequate especially for data with ordinal attributes of low cardinality
- Can be viewed as an nD histogram if the color of a cell is set proportional to the number of values that map to it

Dimensional Stacking

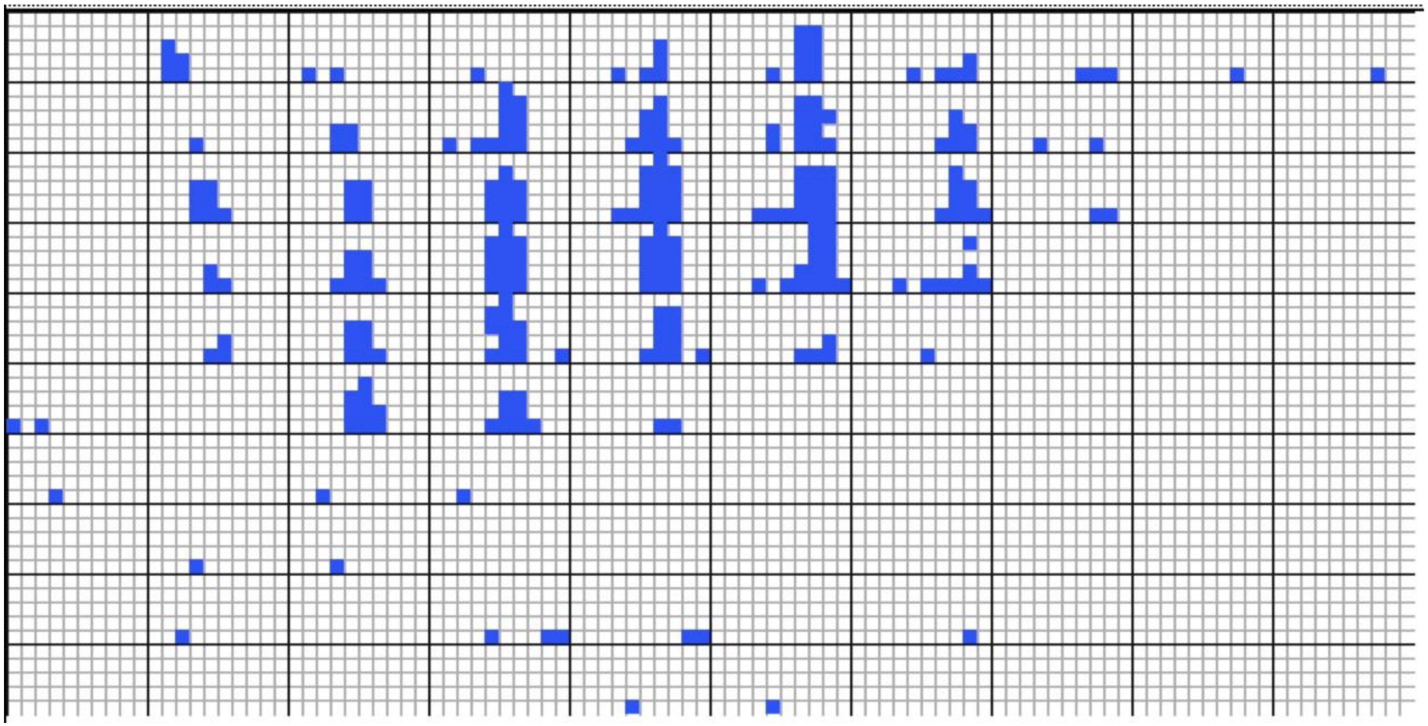


Values in each dimension are divided into classes: X (2), Y(3), W(6), Z(5)



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

Dimensional Stacking



[Ward 2010]

Visualization of oil mining data with longitude and latitude mapped to the outer x-, y-axes and ore grade and depth mapped to the inner x-, y-axes

Visualization Gallery

- Take a look at:
  - IBM Many Eyes
    (http://www-958.ibm.com/software/analytics/manyeyes/)
  - Tableau Public
    (http://www.tableausoftware.com/public/gallery)
  - Google Charts
    (https://developers.google.com/chart/interactive/docs/gallery)
  - D3.js
    (http://d3js.org/)
  - R
    (http://www.r-project.org/)
- You may 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]