Human perception

- How are visualizations perceived by different humans?
- How do we know that a given visualization is correctly interpreted?

Perception:

- Recognizing
- Organizing (gathering, storing)
- Interpreting (binding to knowledge)

Perception mechanism

- Preattentive
  - Fast (250 ms)
  - Performed in parallel
- Attentive
  - Slow
  - Uses short term memory
  - Transforms simple visual features into structured objects
  - Compares to memory models (ex. door)

Preattentive features

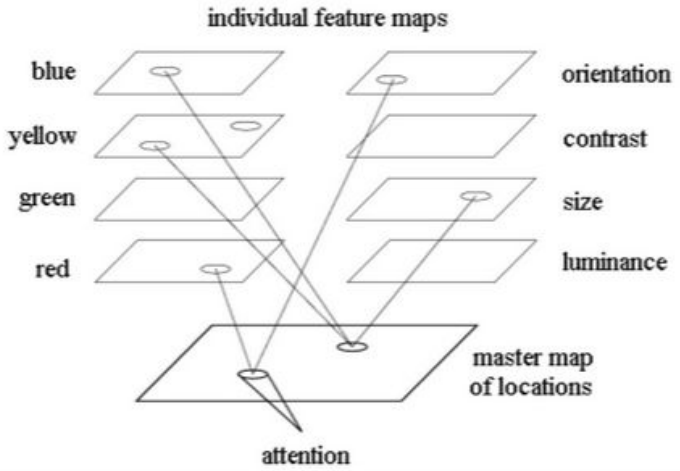
- Length
- Width
- Size
- Curvature (shape)
- Hue
- Intensity
- Flicker
- Direction of motion
- 3D depth
- Lighting direction

Preattentive visual tasks

- Presense or absense of object with a unique visual feature among distractors is detected preattentively
- Boundary between two groups of elements with the same visual feature is detected preattentively
- Movement of an object with a unique visual feature is tracked preattentively
- Amount of elements with a unique visual feature is estimated preattentively

Treisman's theory of preattentive processing

- A figure is processed in parallel by checking individual feature maps
- A specific preattentive task is performed in each feature map
- Conjuction of features requires searial search between maps - takes time



Metrics

Relative judgement: comparing two values of a feature

Errors (in increasing order)

- Position along a common scale
- Length
- Angle
- Area
- Volume
- Color hue

–> *Pie charts* are less effective than *Bar Charts*

Principles of good visualization

- Use intuitive mapping to aesthetics
  - Visualization type is adopted to user's background
  - Geographical coordinates –> X,Y, temperature–>color
  - Use correct mapping
    - Ordinal variables- X,Y, saturation, orientation
    - Nominal variables - shape, texture, hue
- Support view modifications
  - Scrolling, zooming
  - Color map
  - Mapping aesthetics
  - Scales
  - Level of details

Density plots and box plots

What should be analysed?

- Density plot, histogram, violin plots
  - Mean value or typical value
  - Symmetry
  - Variation
  - Whether reminds some distribution
  - Heavy/Light tailed
  - One ore more modes
  - Skewness

Density plots and box plots

What should be analysed?

- Box plot
  - Median
  - Variation
  - Outliers
  - Symmetry
  - Quantiles

Scatter plot

- Y: dependent variable, X: independent variable
- Smoother is a good idea to have

Analysis:

- Shape (data=true+error, true=linear, quadratic, cubic, exponential, .., empirical)
  - How to find the right model?
  - Fitting the data (regression)
  - Analysis of residuals or model selection methods
- Strength (how close observations to a hypothized model)
  - If linear, Correlation r or coefficient of determination R2

Scatter plot

Analysis:

- Direction (if monotonic, decreasing or increasing; if not monotonic, which parts increasing, which decreasing)
- Density (dense areas, sparse areas)
- Outliers
- Clusters

Scatter plot

Analysis:

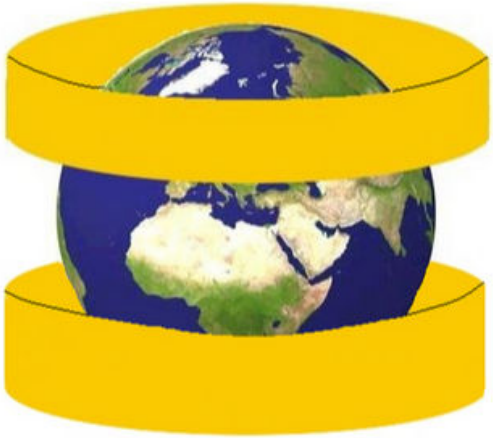
- Direction (if monotonic, decreasing or increasing; if not monotonic, which parts increasing, which decreasing)
- Density (dense areas, sparse areas)
- Outliers
- Clusters

What is map?

- Map coordinates:
  - longitude $\lambda = [-180, 180]$, negative=west
  - lattitude $\phi = [-90, 90]$, negative= south
- Challenge: $[\lambda, \phi] \rightarrow [x, y]$
- Different map projections
  - Conformal projection: retains angles (shapes) but not area
  - Equal area: retains areas but not angles (shapes)
  - ...

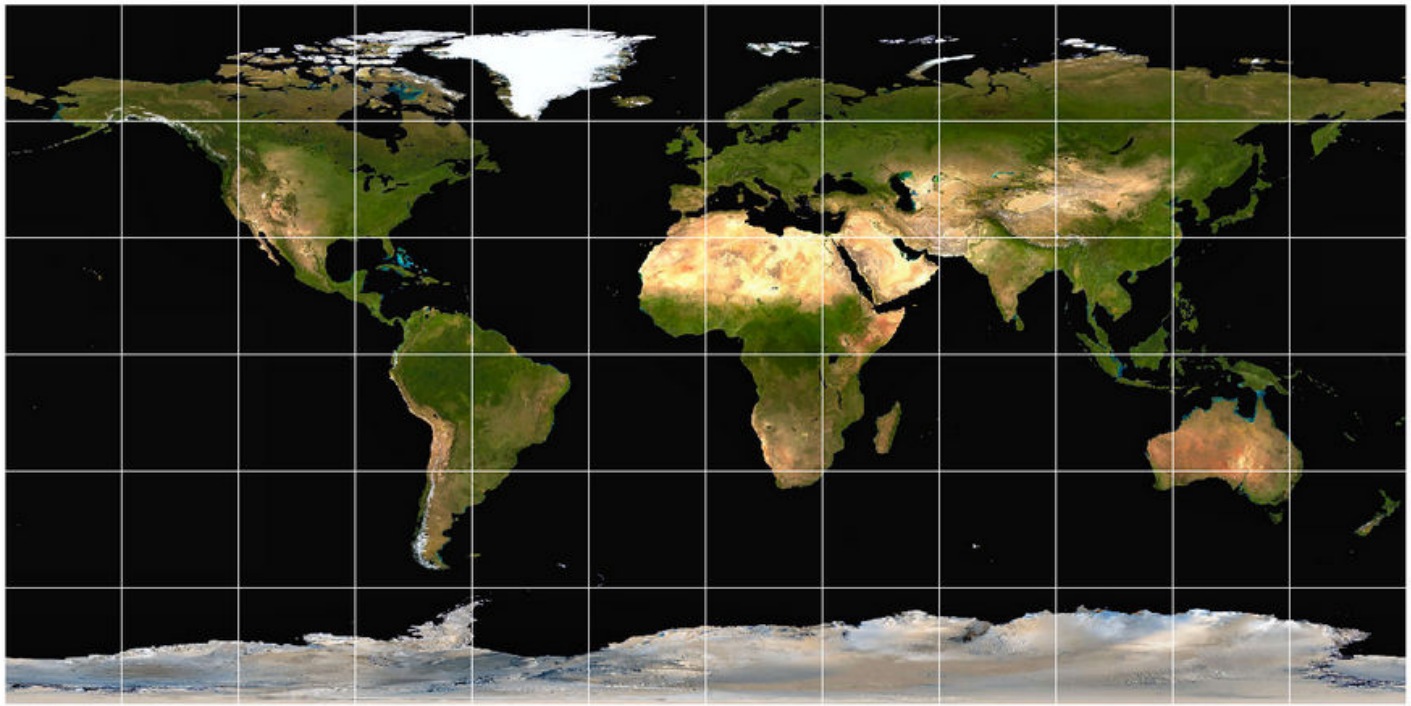
What is map?

- Cylindrical projection, plane projection and cone projection
- Cylindrical projection used by Google, standard now



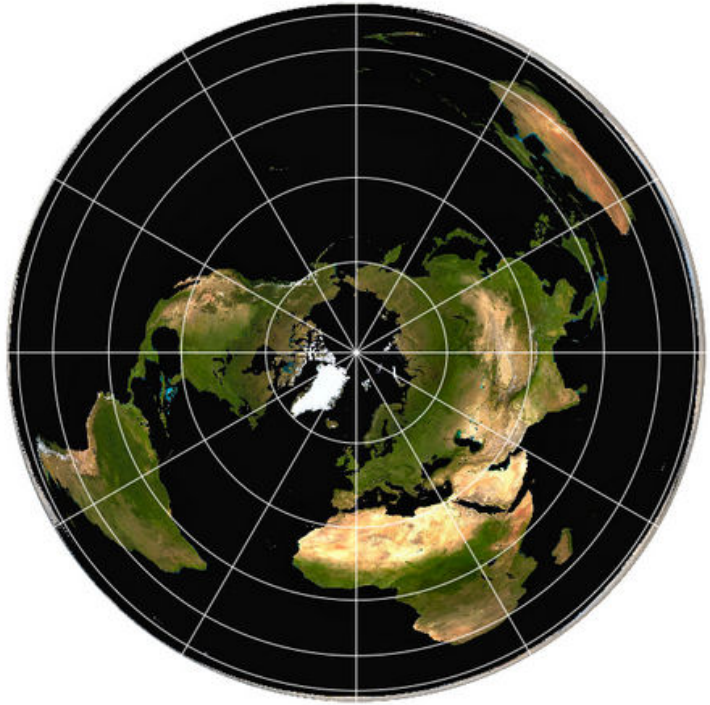
Cylindrical projection

- Conformal projection: far northern/far southern areas inflated
- Defined by $x = \lambda, y = \phi$



Cone projection

- Albers Equal-area projection
  - Preserves areas
  - Shapes or distances are not correct



Symbol/dot maps

- **Analysis:**
  - Density in geogr areas and between geogr areas
  - Spatial pattern of density (north, south)
  - Clusters, outliers
- **Problems:**
  - Overplotting in highly populated ares
  - If several observations have the same coordinate
  - Size aesthetics used–> perception problem
  - Perceived size depends on local neighborhood (Ebbinghaus illusion)
- Color used: color perception problems

Choropleth maps

- Analysis:
  - Find clusters of regions that are similar
  - Find unusual regions (compared to neighbor regions)
  - Find patterns on the map
- Problems affecting perception:
  - Color/grayscale mapping
  - Choice of regions (county, state,…)
  - Larger region with the same color looks dominating
  - Patterns in small/densely populated areas hard to see

Software for geospatial visualization

- Plenty of commercial/Noncommercial software
  - ArcGIS, Google, Yahoo, Microsoft map API
- Plotly
  - plot_geo()
  - Using MapBox + plot_mapbox()
- To use Mapbox:
  - Register with your email, find your token
  - Run in R `Sys.setenv('MAPBOX_TOKEN' = 'your_mapbox_token_here')`
- Ggplot2
  - geom_sf()
  - ggmap

Using maps

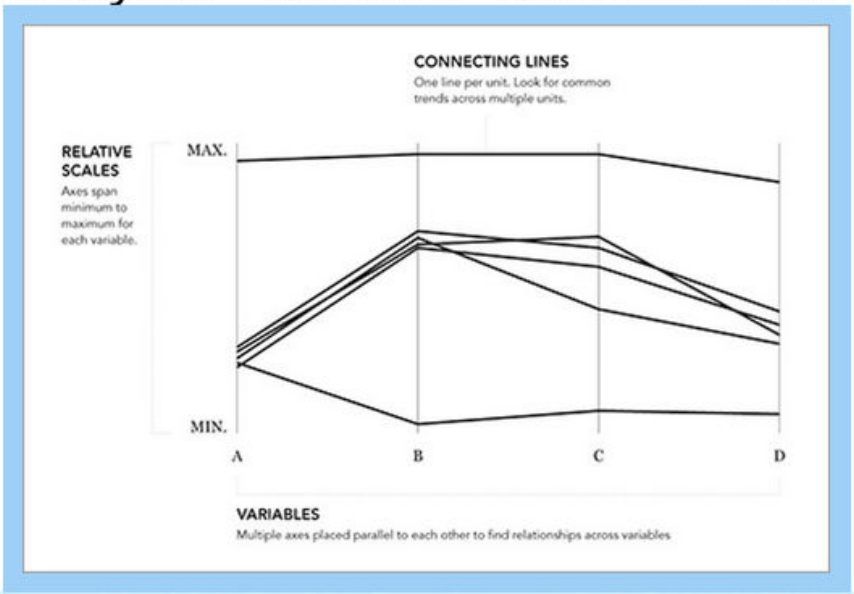
- A few countries available through plotly
- Downloading map of a country:
  - Finding a country map http://gadm.org/
  - Decide what level of detalization is needed (region, county,…)
  - Download R(sf) file.
  - Load the file to R using readRDS function
    - e.g. rds <- readRDS('filename.rds')
  - Use with ggplot()+geom_sf(data=rds)
  - Use with Plotly: plot_ly()+add_sf(data=rds)

Parallel coordinates

Construction:

- Vertical axis: Values
- Horizontal Axis : Variables
- 1 trace line = 1 observation

Analysis: - Clusters - Outliers - Correlated variables



Parallel coordinates

Analysis

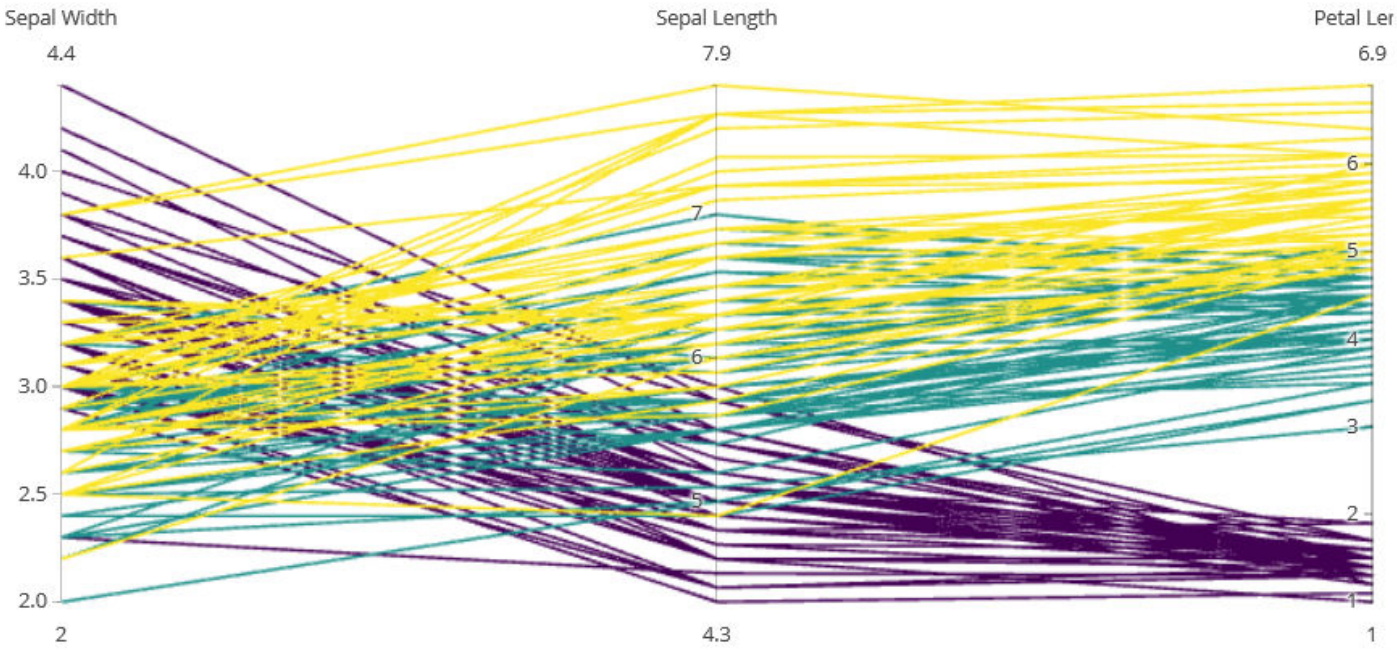
- Positive correlation between two adjacent variables: almost all segments are parallel to each other
- Clusters in some variable space: several trace lines that are near each other and have similar pattern
- Outliers: trace lines that have unusual pattern and/or fall out outside the common plot area

Problems

- Trace lines overlap each other -> difficult to find patterns, difficult to follow a specific trace line
- Analysis depends much on the order of variables (correlation, clusters) -> a proper reordering may improve the analysis

Parallel coordinates

- Sometimes clusters overlap with categories given by some variable
  - Non-mixing groups is not the same as clustering!



Ordering problem

Solution

- early approaches (for ex. Ankerst et al. 1998):

1. Choose a distance (proximity) matrix $D = \{d_{ij} = d(x^i, x^j)\}$ between variables (columns)
   - Euclidian distance on scaled columns
   - 1- correlation
2. This defines graph with vertices $1, \ldots, p$ and edge weights d_{ij} -> **Hamiltonian path** (Traveling Salesman Problem)

$\min_\Psi \sum_{j=1}^{p-1} d'_{j,j+1}$

- TSP is NP-complete -> Approximate solutions are used

Heatmaps

Analysis:

- Compare the values of a parameter for different observations (row)
- Compare the values for a single observation (column)
- Compare the patterns for different rows or columns
- Find similar observations (areas with the same color intensity)
- Find which variables define similarity for a group
- Find correlated variables (similar pattern within a column)

Radar charts

If juxtaposed, analyse:

- Clusters
- Outliers
- Outlying directions

If superimposed,

- Comparing variable length
- Seeing similar and outlying observations

Trellis plots / facets

- Faceting = one more aesthetics
- What can be analysed?
  - Patterns within/between plots
  - Conditional dependence $Y \sim X|Z$
  - Variable interaction, additivity

–> Useful tool for modeling!

- Compare : 3D- scatter plots

Interactive graphics

- Key tool for visual analytics
- Much more efficient than static graphics

Examples:

- Navigation (panning, rotation, zooming)
- Selection (highlighting)
- Connecting (linked views)
- Filtering (sample)
- Reconfiguring (change aesthetics)
- …

Interaction operators

Navigation operator:

- Camera location
- Viewing direction
- Level of details (e.g. hierarchical representations)

Selection operators

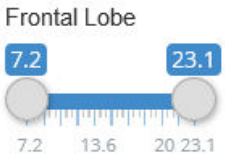
- User isolates a subset of objects
  - Highlighting
  - Masking
  - Focusing
- How to implement?
  - Click
  - Click+hold
  - Bounding box, lasso

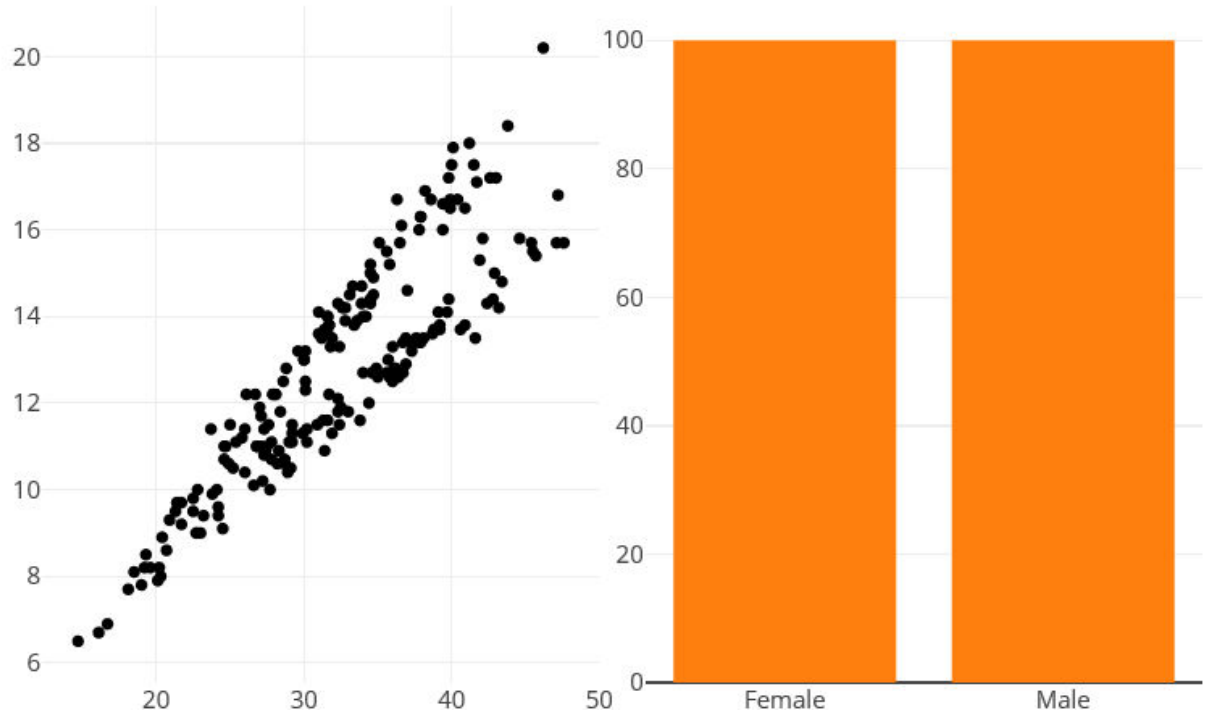
Connection operators

- Related observations are linked in corresponding views
- Selection operator+Connection operator = Brushing
  - Persistent and transient

Filtering operators

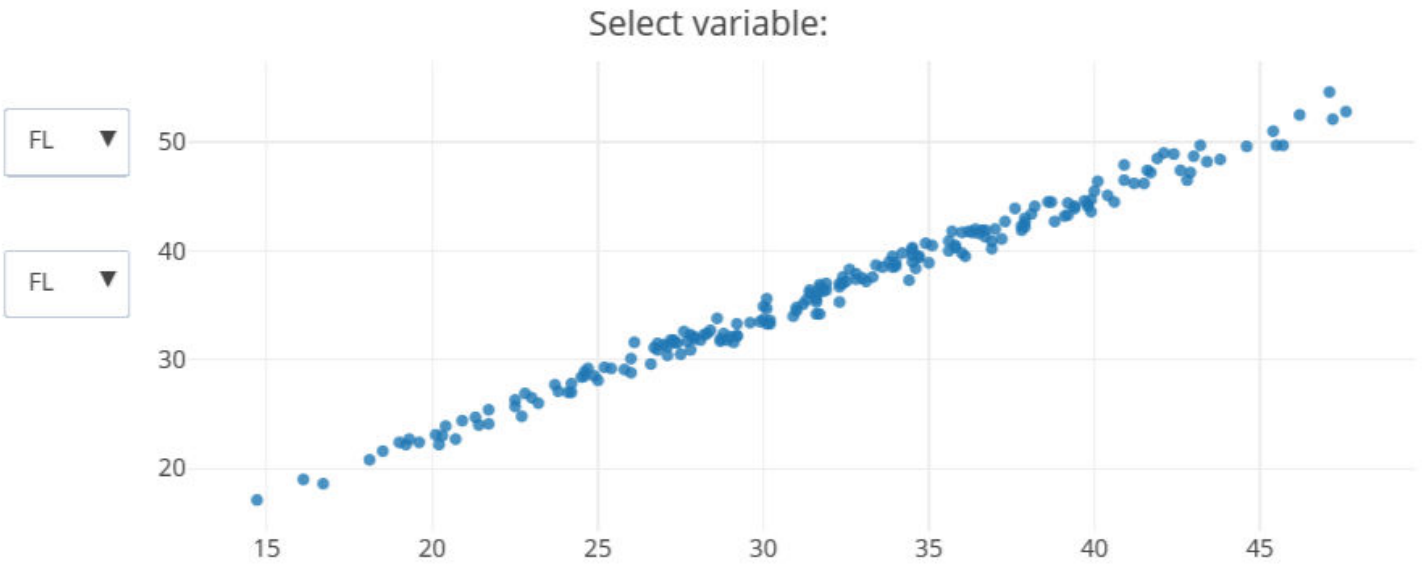
- Reducing data acc. to specifications





Reconfiguring operators

- Transforming data
  - Sorting rows, reorder columns, MDS
  - Change aesthetics mapping

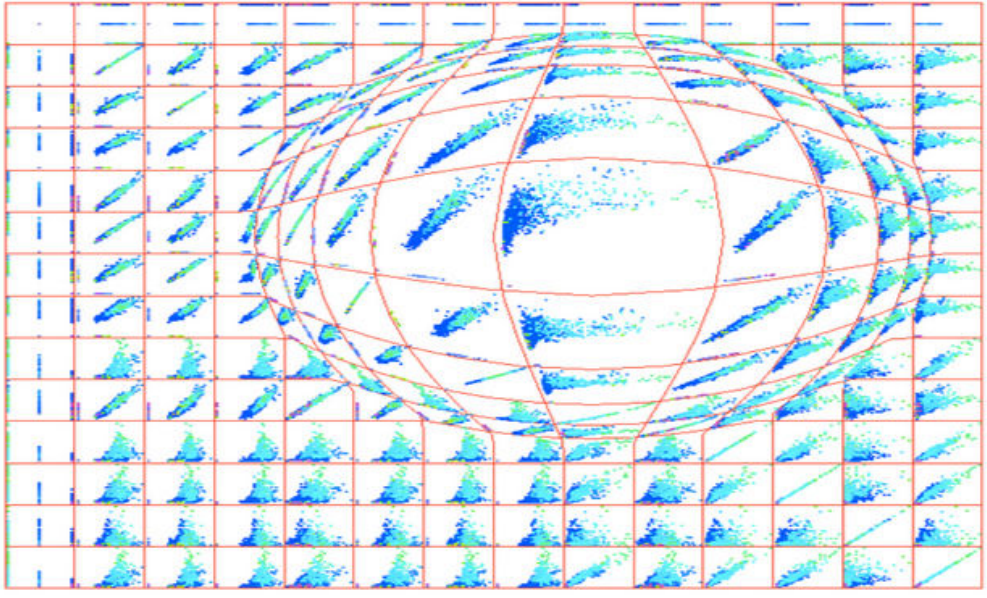


Encoding operators

- Changing the visualization type
- Changing aesthetics
- Another color map
- Change shapes
- …

Abstraction operators

- Distoring objects locally or globally



Interaction operands

Data value space

- Operate observations instead of pixels
- Navigating: translate the axis range
- Zooming: increase/decrease axis range
- Filtering: sample the data, sample dimensions
- Reconfiguring: sorting observations, dimensions, nonlinear transformations

Interactive visualization in Plotly

Without Shiny: key ingridients

1. `highlight`-function:
   - applied to Plotly object
   - parameters:
     - on: 'plotly_click', 'plotly_selected',...
     - persistent: TRUE/FALSE
     - dynamic:T/F - enables color selector
     - selectize: T/F text field for selection

Text visualization

Applications:

- Articles, books
- Emails, blogs, websites
- Program Logs
- Collections (corpus) of books, blogs,..

Analysis:

- Understanding structure/context of text
- Group similar documents
- Developent of contents/topics over time

Word cloud

Issues:

- Stopwords need to be removed
- Words sharing the same stem aggregated
- Synonyms
- "Satisfied"/"not satisfied" example
- Incorrect spelling?
- Hyphens and apostrophes
- Size mapping inaccurate (long words)
- Does not show the structure

Phrase net

- Size of the word = word freq
- Thickness of the connection =co-occurance freq
- Color: dark colors -> word often to the left
- Close on map -> often close in the document

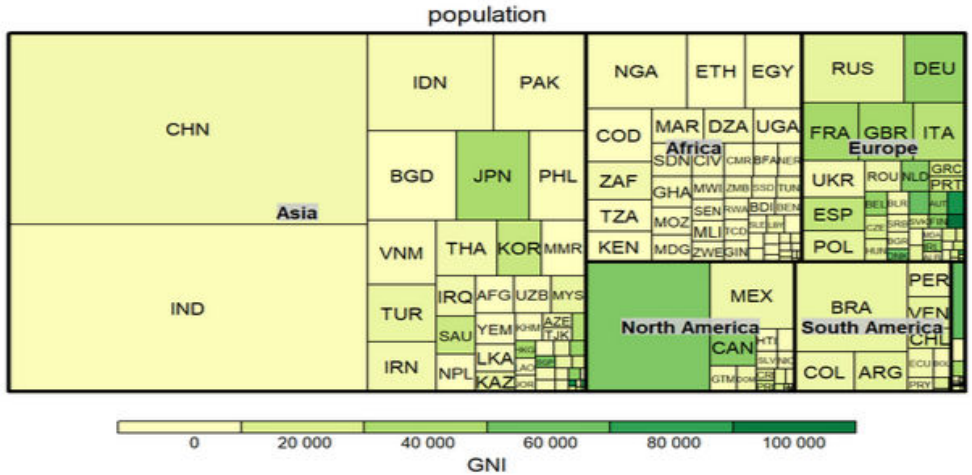
Analysis:

- Analyse most frequent
- Analyse connections and paths
- Analyse strength of paths and intensity of colors
- Click a specific word and look at the respective paths
- Zoom in and out

Treemaps

- Analysis: Comparing size/color of hierarchies between/within different levels

Example: World population



Tree

- Aesthetics: Node/link size, node/link color
- Can be used for classification result analysis
- Analysis
  - Analyse each branch and terminal nodes
  - Compare information between nodes
  - Compare information between branches

Graph visualization

Analysis:

- Which nodes are hubs?
- Which links are strongest?
- Which nodes are n steps away from some node?
- Components with strong connections?
  - Relationships to groups?
- How are different groups connected?
- Shortest path between nodes
- Community detection (densely connected nodes)
- Other interesting substructures

Animation

- Eye is drawn to similar motions and outlying motions
- Advantages
  - Effective at attracting attention
  - Time=one extra aesthetics
  - Easily perceived in peripheral vision ->many features can be captured at one time point
- Disadvantages
  - Unappropriate transformation (transition) -> false conclusions
  - Speed of the graphics may hide important details/make boring

Animation- recommendations

- Maintain valid data during transitions
  - Example: using splines
- Be careful when using interpolations: use appropriate models
- Group similar transitions
- Minimize occlusion
- Use simple transitions
- If trajectory is stable, use ease-in, ease-out
- Make transition as long as needed but no longer.

2D-tours

- Idea
  - Investigate various projections
  - Connect them into animation
- What to analyze?
  - Same as in scatterplots plus
  - Which variables contribute to projections
    - Patterns in lower dimensional manifold?
- Types of 2d-tours: grand tour and guided tour