DS 501 - Introduction to Data Science

Visualization

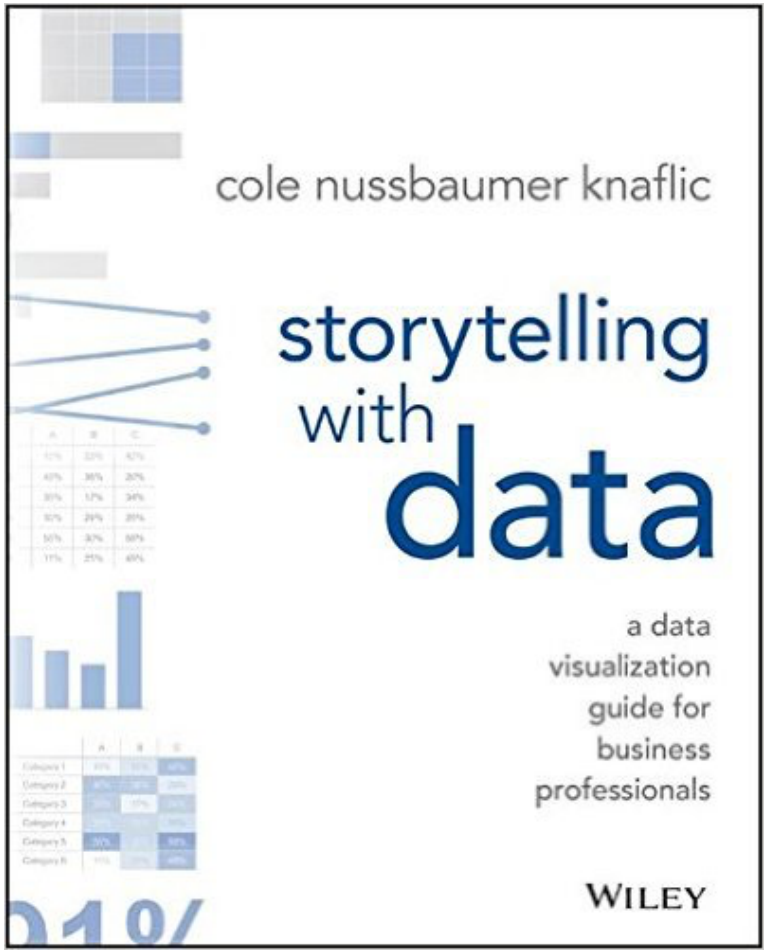
Narahara Chari Dingari, PhD

WPI

Case Study 2 - Analyzing data from MovieLens

- R Code/R Markdown (R/Rmd)
- Report (Word/PDF)
- Slides (Keynote/PPT)

Story Telling With Data



http://www.storytellingwithdata.com

Overview

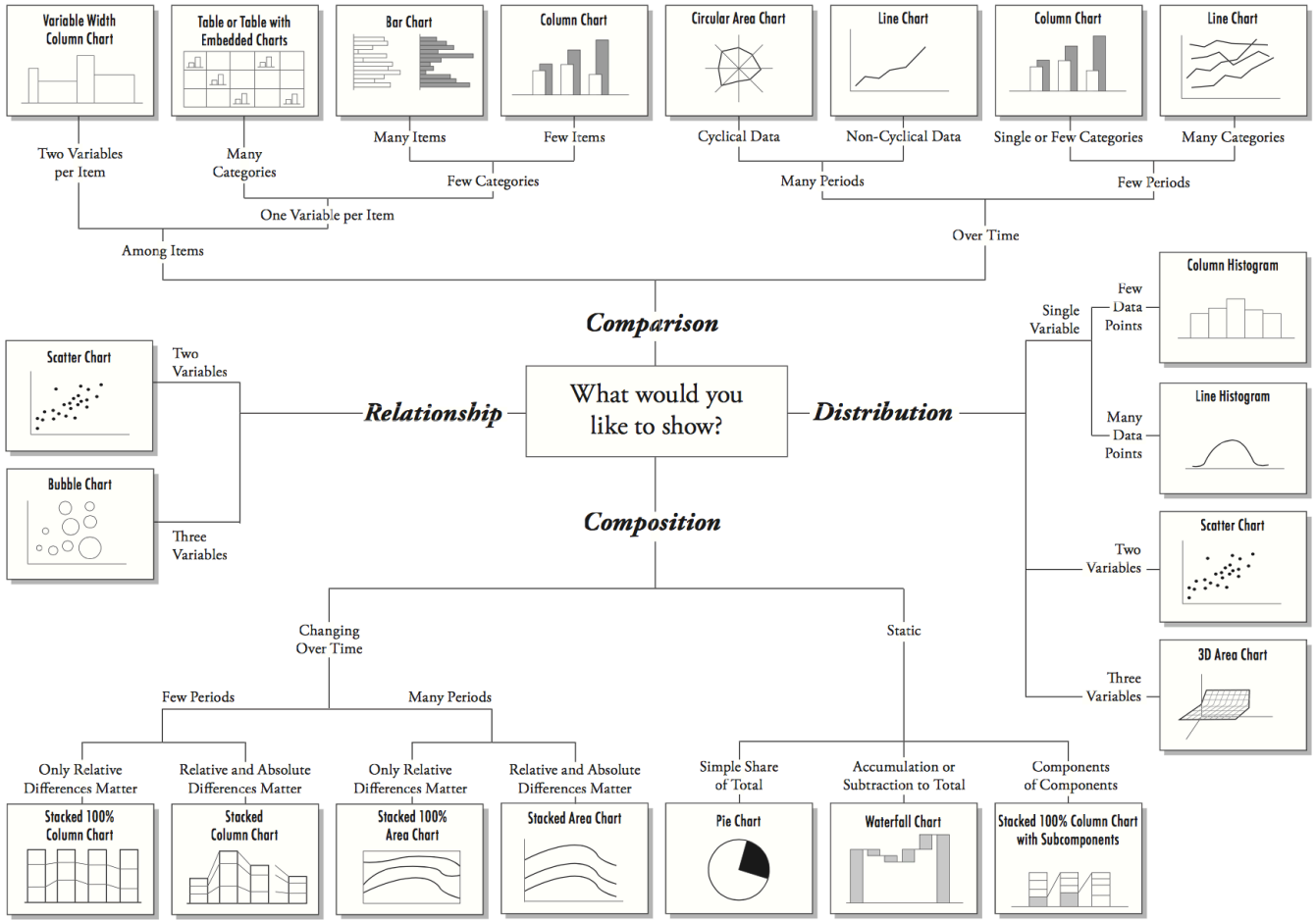
Open Source

- R
  - Base package
  - ggplot2
  - googleVis
  - rCharts
  - ggvis
  - highcharter
- Gnuplot
- Gephi
- Plot.ly
- Python
  - Bokeh
  - Vispy
  - matplotlib

Commercial

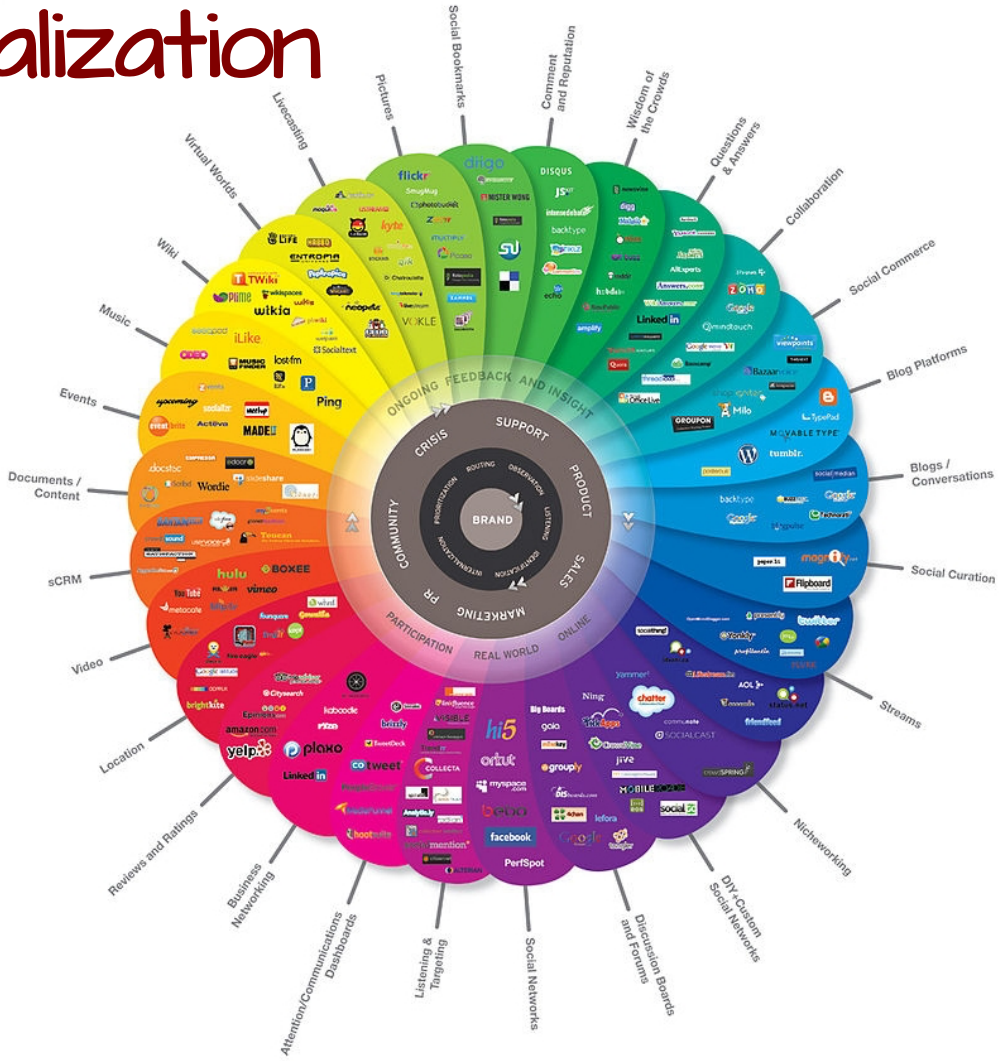
- Tableau
- Spotfire (Tibco)
- Qlikview
- Adobe illustrator

Chart Suggestions—A Thought-Starter



Perception and Visualization

- "A picture is worth a thousand words"
- We are visual beings
  - Sight is the key sense for information understanding
  - Have been using visuals for many centuries

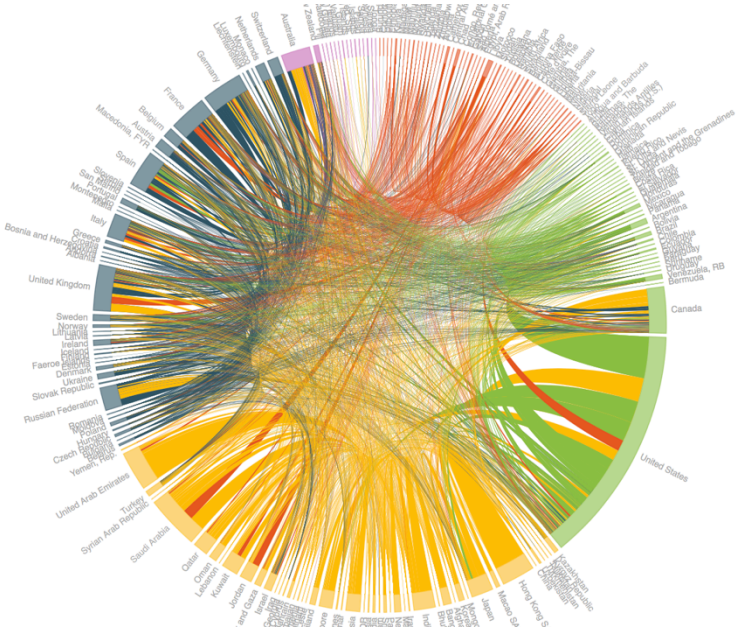


Why data visualization is important?

- Data is easier to read in visual form
- Helps discover new knowledge
- Applies to any domain
- Assist in analysis and communication

Remittance flows

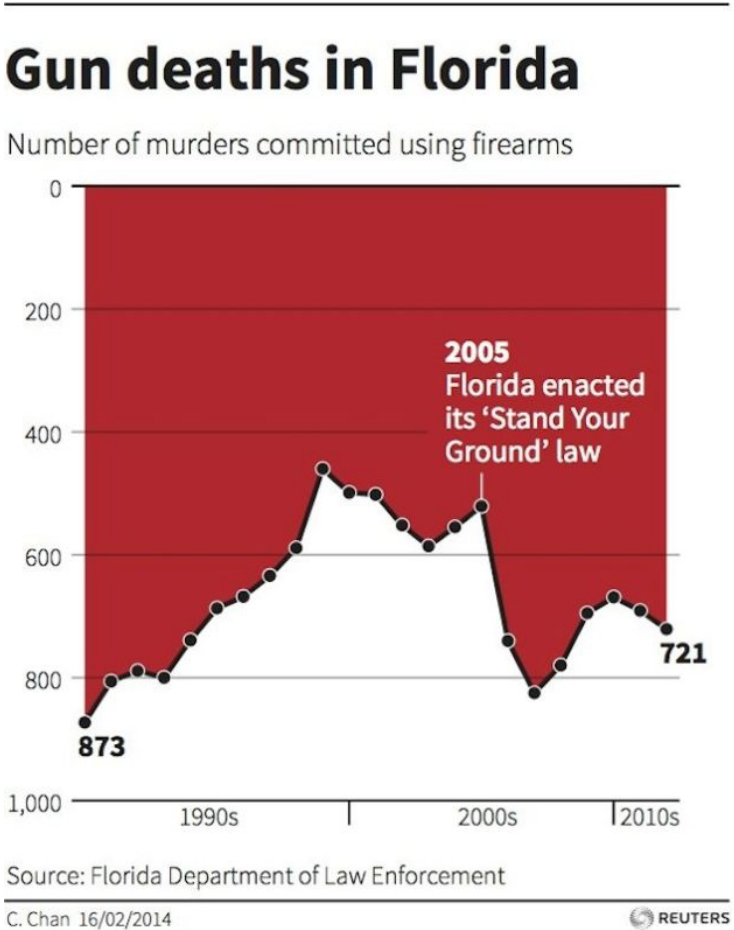
Remittances Sent

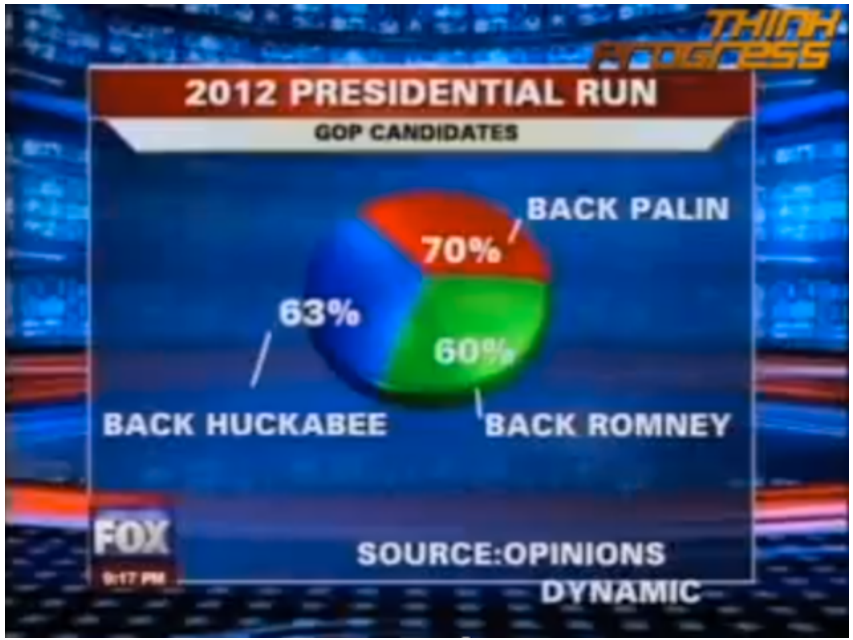


Remittances Received

http://www.torre.nl/remittances/#menu

Some bad examples





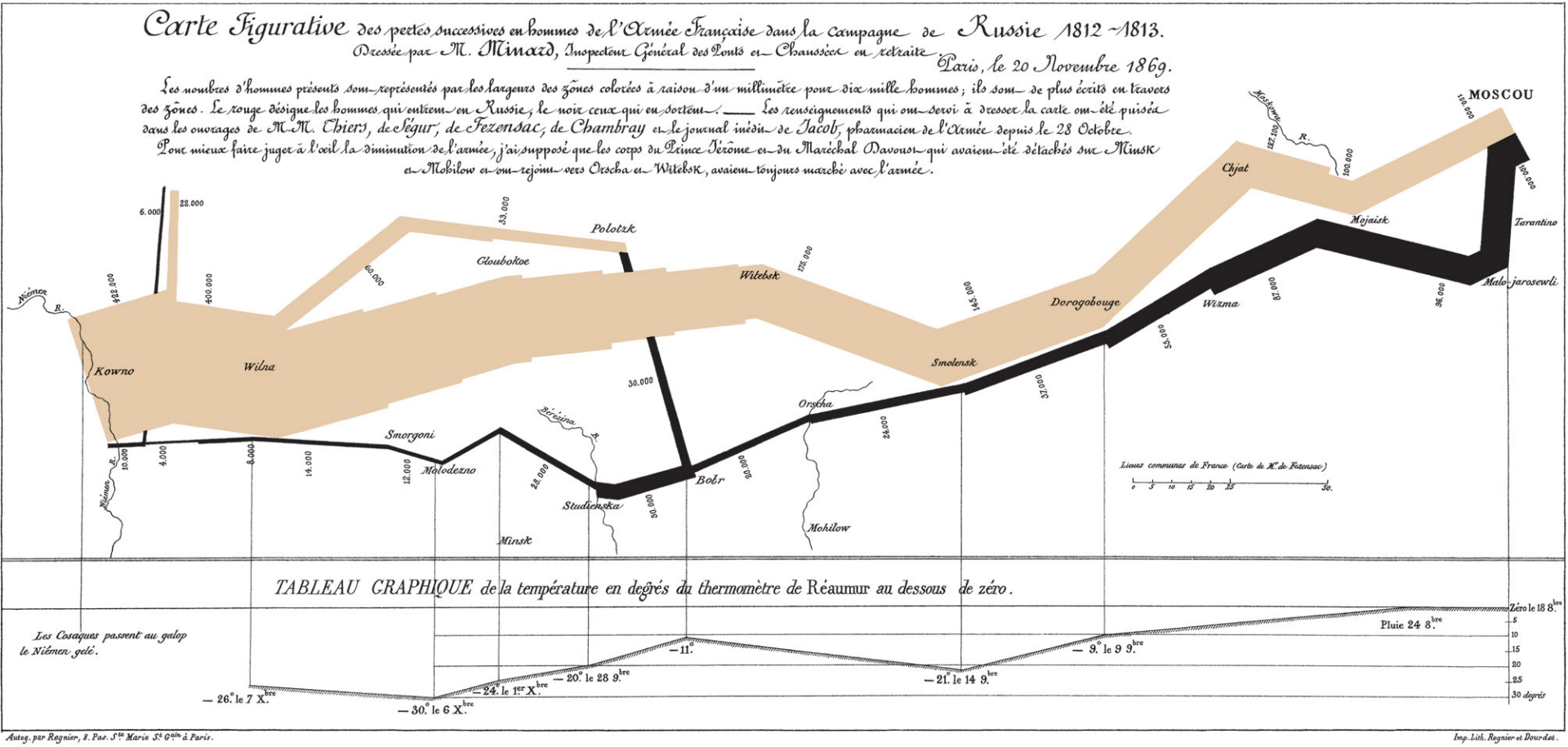
WPI

What is Data Visualization?

- Data visualization is a general term used to describe any technology that lets corporate executives and other end users "see" data in order to help them better understand the information and put it in a business context.
- Data visualization schematically abstracts information to bring about a deeper understanding of the data, wrapping it in an element of awe – Bloomberg Business Weekly
- The presentation of statistics with images that depict the meaning of the statistics – census
- The use of computer-supported, interactive, visual representations of abstract data to amplify cognition – Card *et.al.* 1999

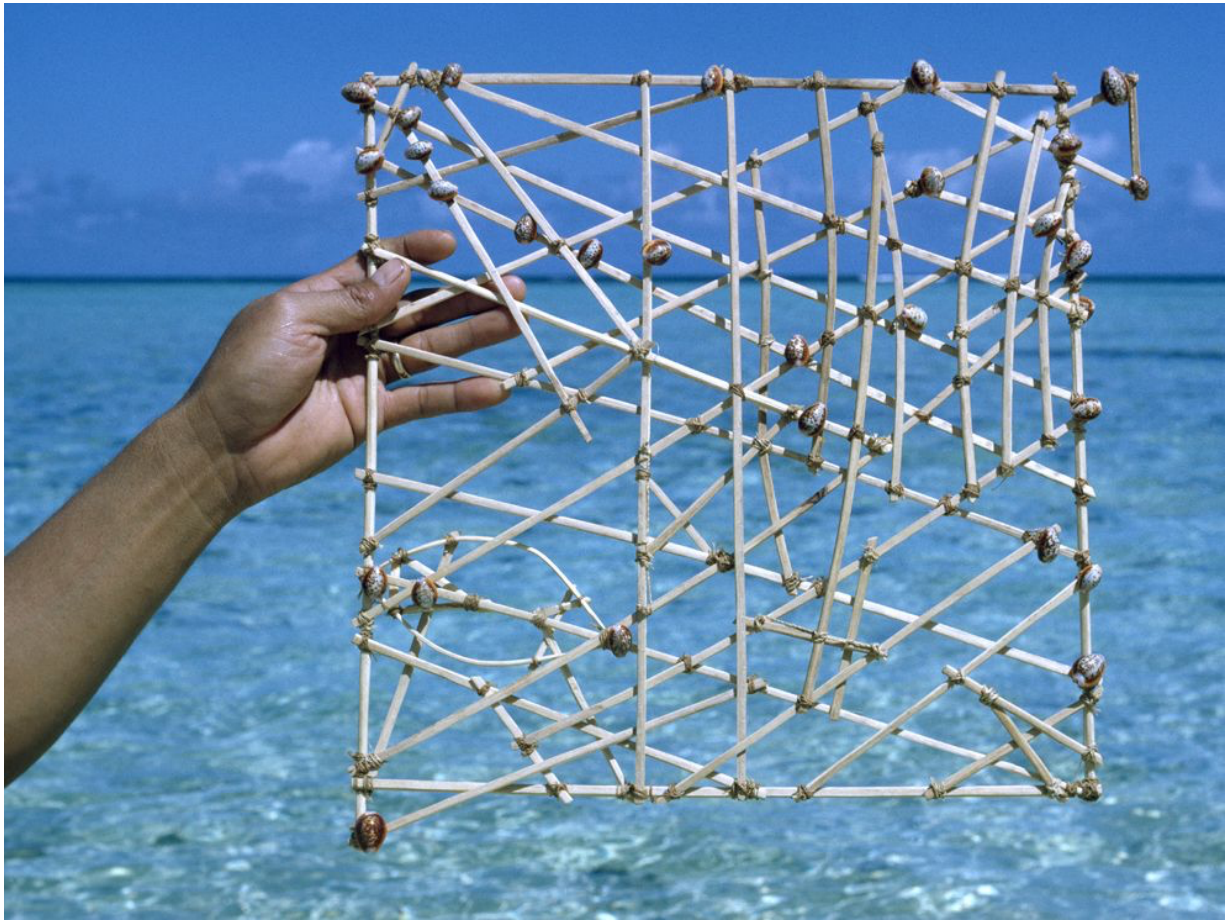
Minard's flow map (Napoleon's march)

Displays army size, geo coordinate, date and weather



Micronesian Stick Chart

Can you navigate the Pacific?

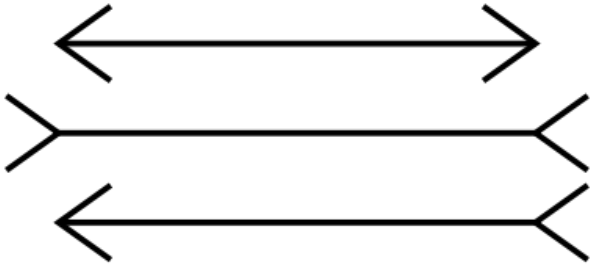


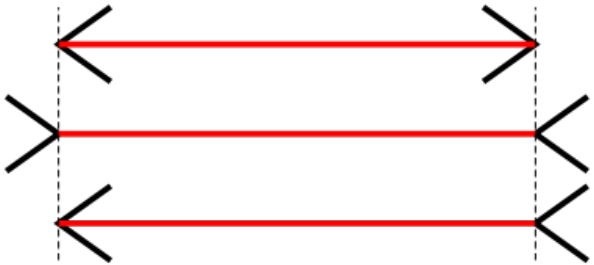
WPI

http://education.nationalgeographic.com/education/media/micronesian-stick-chart/?a

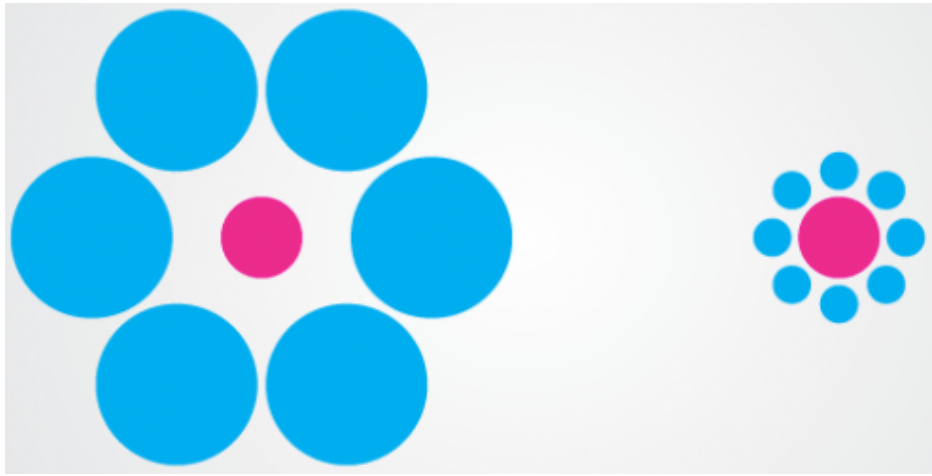
Perceptual Illusions

Müller-Lyer Illusion

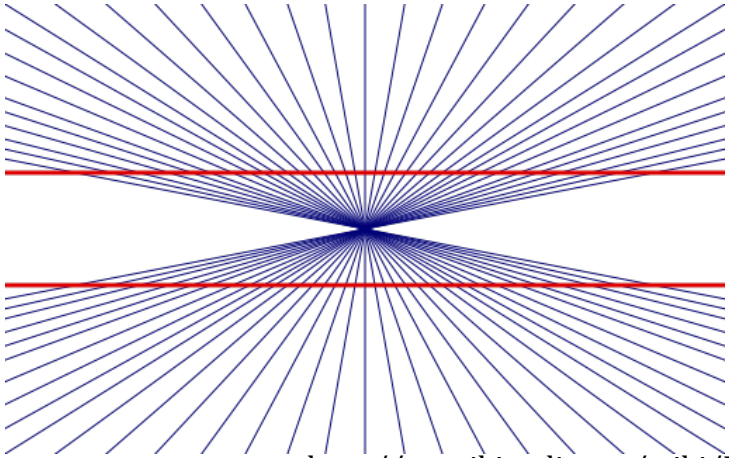




http://en.wikipedia.org/wiki/Müller-Lyer_illusion

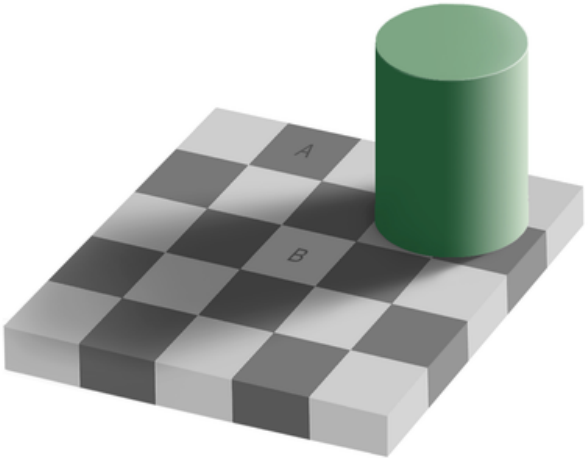


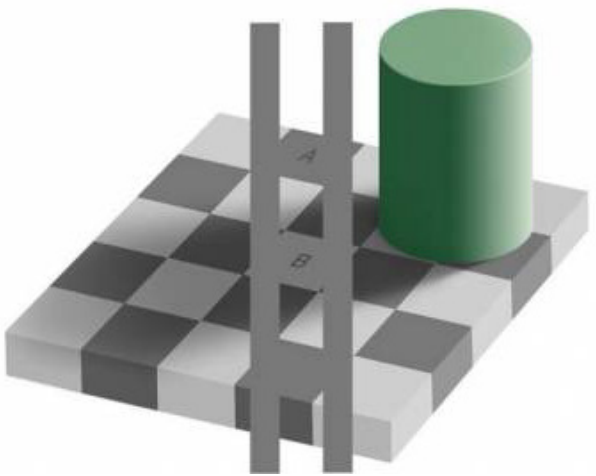
Hering Illusion

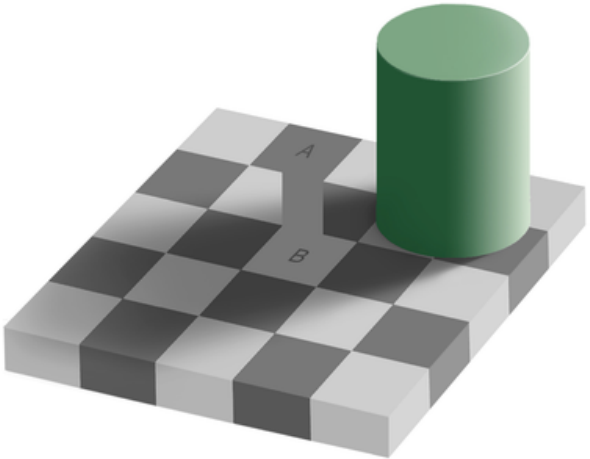


http://en.wikipedia.org/wiki/Hering_illusion

Edward Adelson - Checker Shadow Illusion







http://en.wikipedia.org/wiki/Checker_shadow_illusion

Preattentive Processing

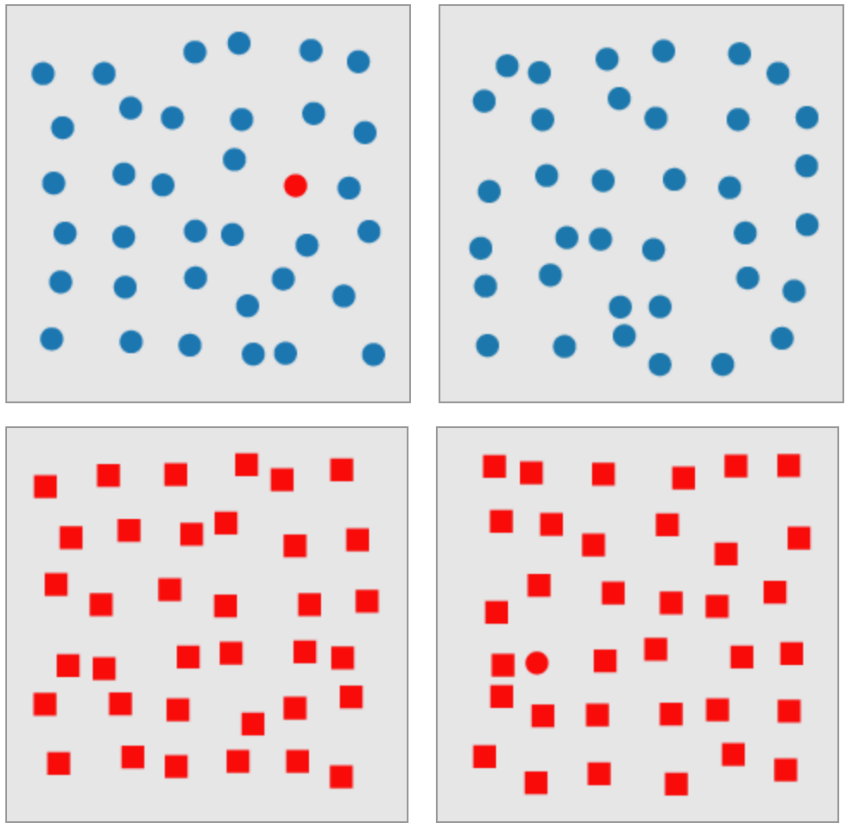
How many 5's are there?

1561321203658413076510374627
4173127527327592732990709742
1703707774179527931749270973
4019743217909370945179279417

1561321203658413076510374627
4173127527327592732990709742
1703707774179527931749270973
4019743217909370945179279417

Preattentive processing

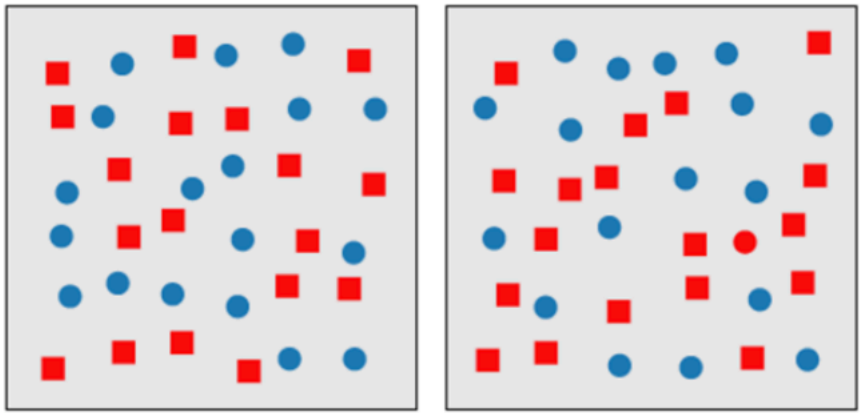
Find the target red circle?



http://www.csc.ncsu.edu/faculty/healey/PP/index.html

Conjunction target

- Target made up of combination of non-unique features
- Can not be detected pre-attentively
- Find the target red circle



http://www.csc.ncsu.edu/faculty/healey/PP/index.html

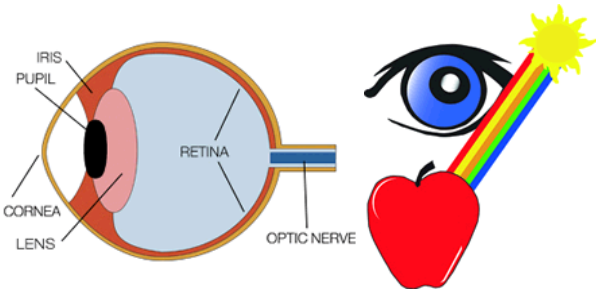
Perception study

- How do humans perceive data?
- How visualization is perceived?
- How do we know our visualization is not perceived differently by different viewers?

Perception study

- Study helps the design of sensory environments
- Build artificial devices to aid sense organ defects
- Build devices to trick the senses such as virtual realities
- Research on repairing broken sensory organs and nervous systems, or preventing defects

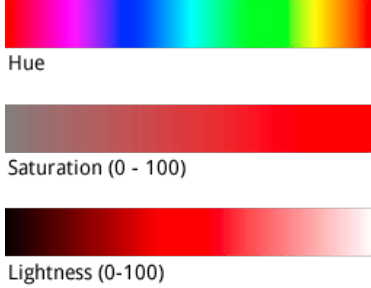
Color Model

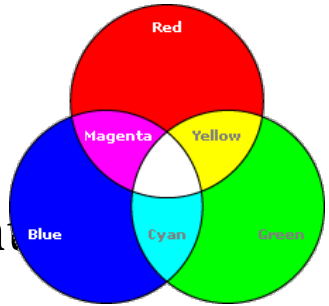


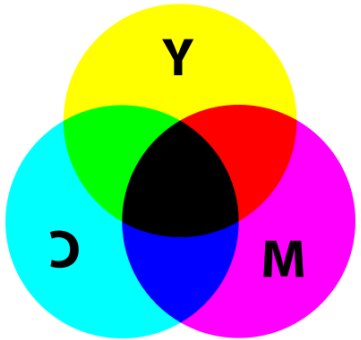
- The way human eye perceives colors suggests that color is a 3D entity
- Color model
  - Is an abstract 3D mathematical system for representing color
  - Defines three primary colors along 3D
  - Is typically limited in the range of colors they can represent and often can't represent all colors in the visible spectrum
- Example color models
  - Perception based (LHS)
  - Additive systems (RGB)
  - Subtractive systems (CMY)

Color Model - Examples

- LHS/HSB (lightness-hue-saturation)
  - Perception based
  - Define colors using Hue, Saturation and Lightness/Brightness
- RGB
  - An additive system
  - Uses red, green, and blue as the primary colors
  - A color can be obtained by combining different amount these three primaries
- CMY
  - A subtractive system
  - Uses cyan, magenta, and yellow as the primary colors



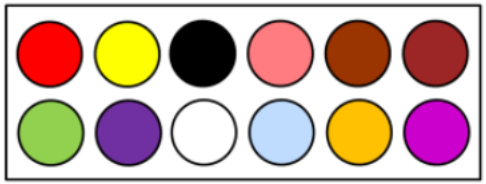


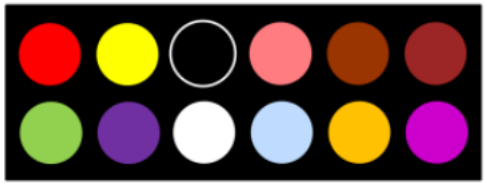


Perception of colors - nominal data

- Do colors naturally represent an order (low -> high)

- Categorical colors
  - Humans can distinguish at most 8-14 colors
    - Dependent on Objects size
    - Dependent on the background color
  - Do not use too many colors

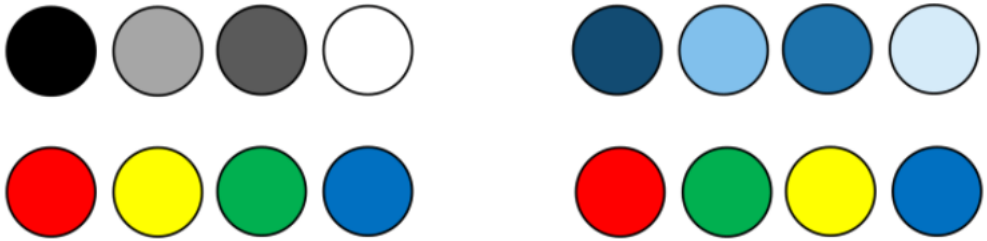




Perception of colors - ordinal data

- Good color scheme
  - Represent order with brightness or saturation

- Bad color scheme
  - Perceptually not ordered



Color advice

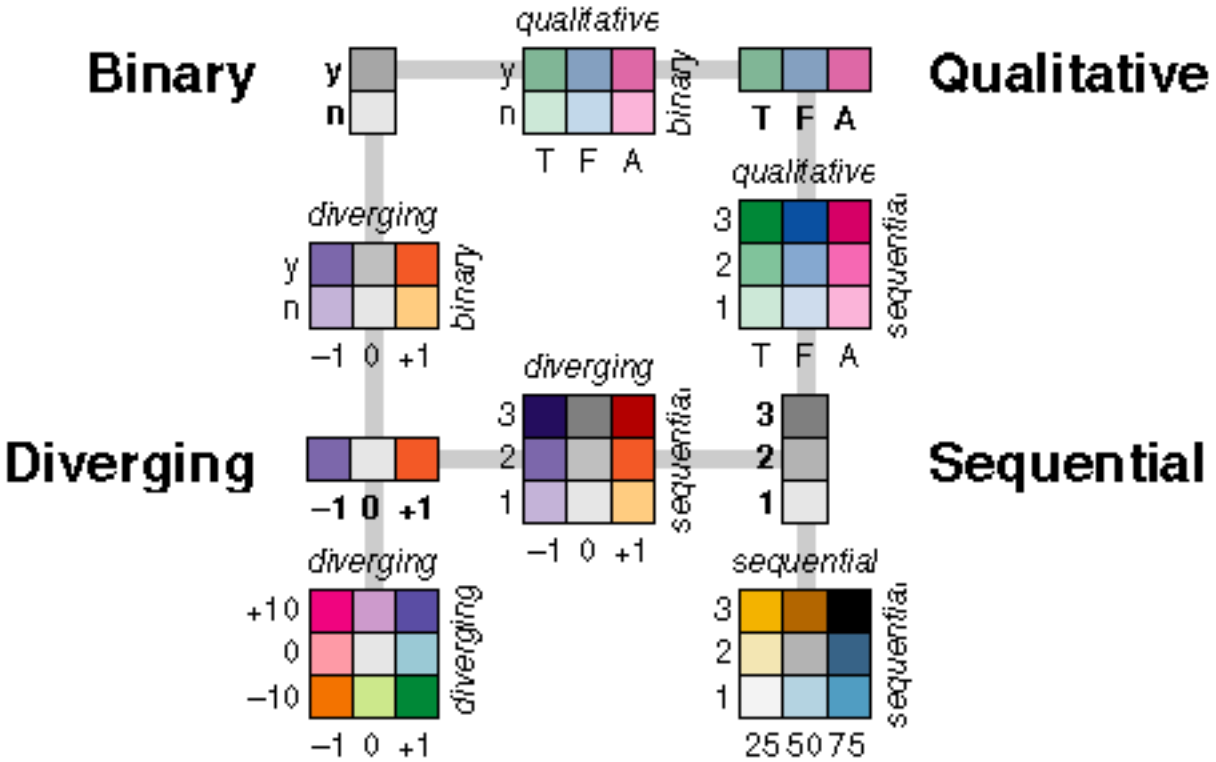
- Binary (True/False)

- Diverging (-2,-1,0,1,2)

- Qualitative

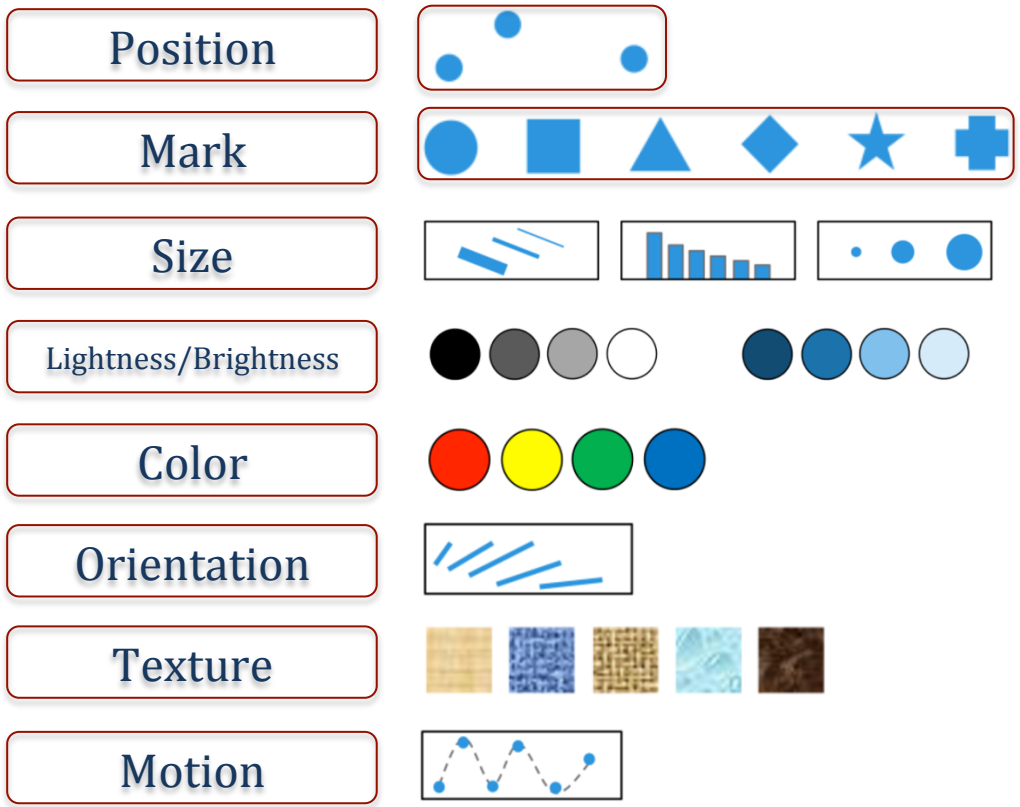
- Sequential (5,4,3,2,1)

Color advice



http://www.personal.psu.edu/faculty/c/a/cab38/ColorSch/Schemes.html

Visual Variables



Characterizing visual variables

- Selective
  - Is a change enough to allow us to select it from a group
- Associative
  - Is a change enough to allow us to perceive them as a group
- Ordinal
  - Are changes in this variable perceived as ordered?
- Quantitative
  - Is there a numerical reading obtainable from changes in this variable?
- Separating
  - Across how many changes in this variable are distinctions perceptible?

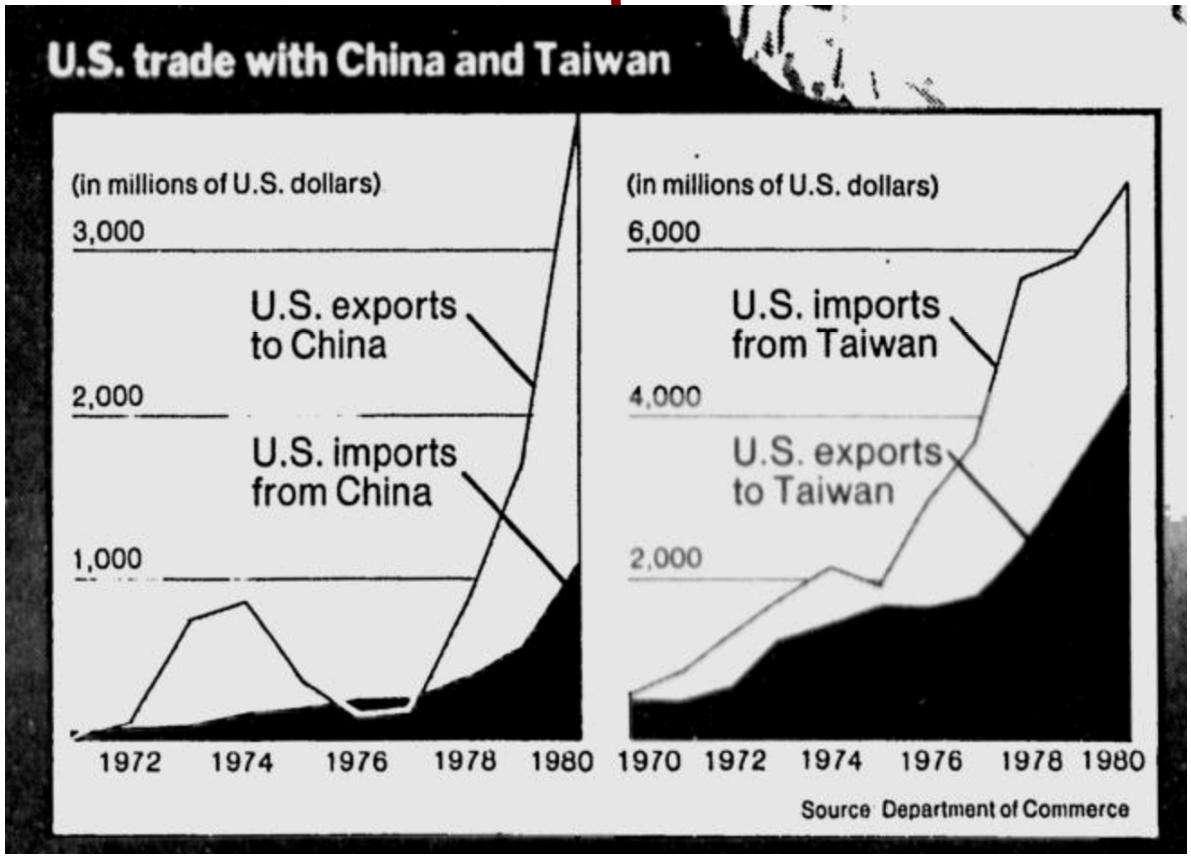
Effective use of colors

- Use color consistently
  - within and between displays
- Be consistent with use expectations (or assumptions) of colors
  - Choose wisely between categorical colors and ordinal colors
- Use color for emphasis, otherwise minimize the use of color
  - If gray scale works, use it
  - Avoid the use of high contrast colors, such as red and blue
  - Avoid large areas of saturated colors for backgrounds
- Use colors to convey messages clearly
  - Always provide a legend
  - Make color redundant with other visual variables
  - Select color for text to contrast with background
- Be aware of skewed distributions in your data

Principles

- Visualize the data
- Thought provoking
- Stick to what the data says
- Effective presentation of multiple parameters on one graphic
- Make large data sets coherent
- Encourage eyes to compare data
- Reveal data at several levels of detail – Big picture and then underlying details
- Serve a reasonably clear purpose

Another bad example



New York Times June 14th, 1981

Some R Visualization Tools

- https://plot.ly/r/
- http://docs.ggplot2.org/current/
- http://www.htmlwidgets.org/showcase_leaflet.html
- http://ramnathv.github.io/rCharts/
- http://cran.r-project.org/web/packages/googleVis/vignettes/googleVis_examples.html
- http://ggvis.rstudio.com

Thank You

WPI