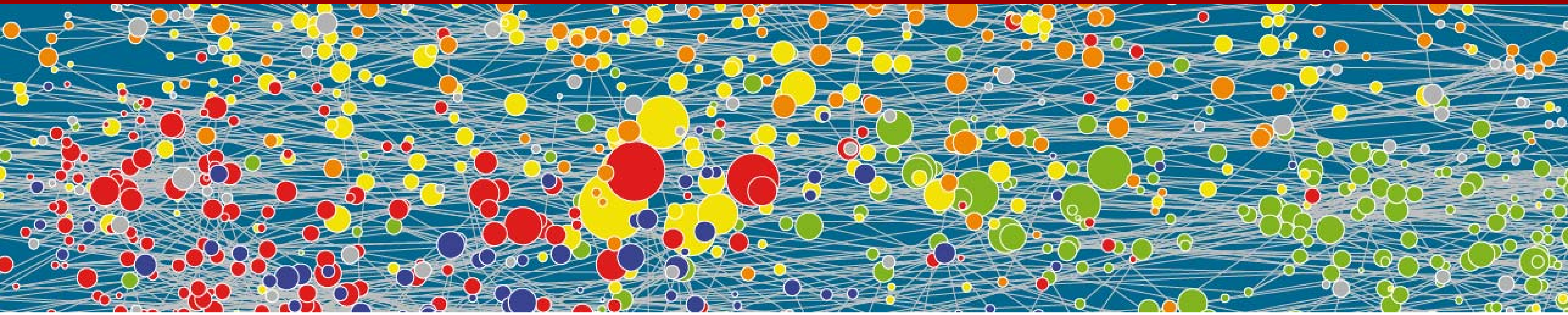
Carnegie Mellon



Visualizing Social Networks

How to Create Meaningful and Compelling Network Drawings

Jürgen Pfeffer

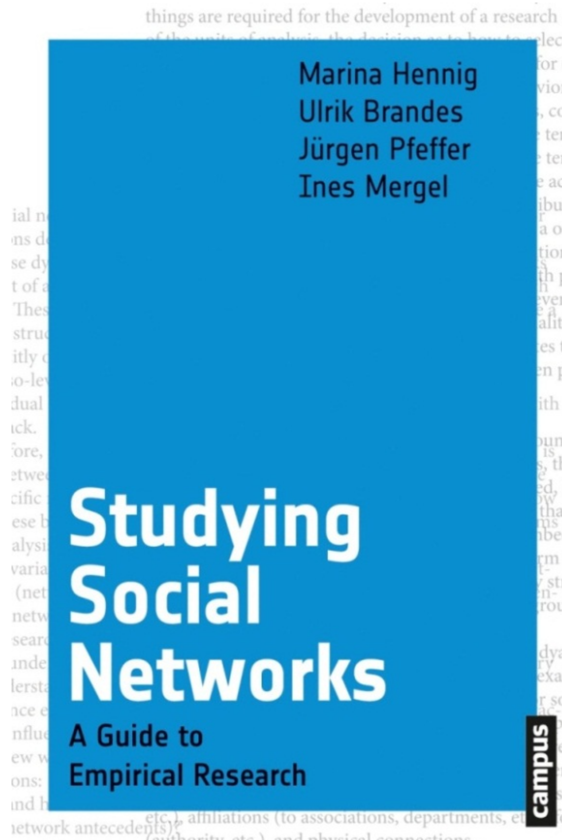
jpfeffer@cs.cmu.edu · @JurgenPfeffer

Sunbelt 2014

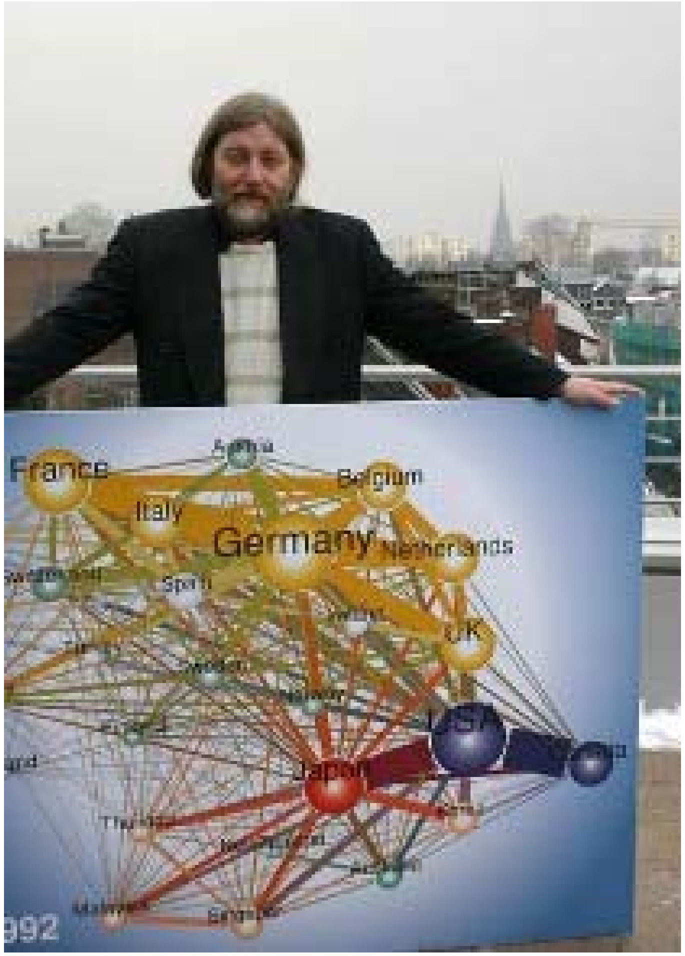
St. Pete Beach, Florida

isr institute for SOFTWARE RESEARCH

- Assistant Research Professor
  School of Computer Science
  Carnegie Mellon University
- Research focus:
  - Computational analysis of social systems
  - Special emphasis on large-scale systems
  - Methodological and algorithmic challenges
- Methods:
  - Network analysis theories and methods
  - Information visualization, GIS, simulations



Little Helpers



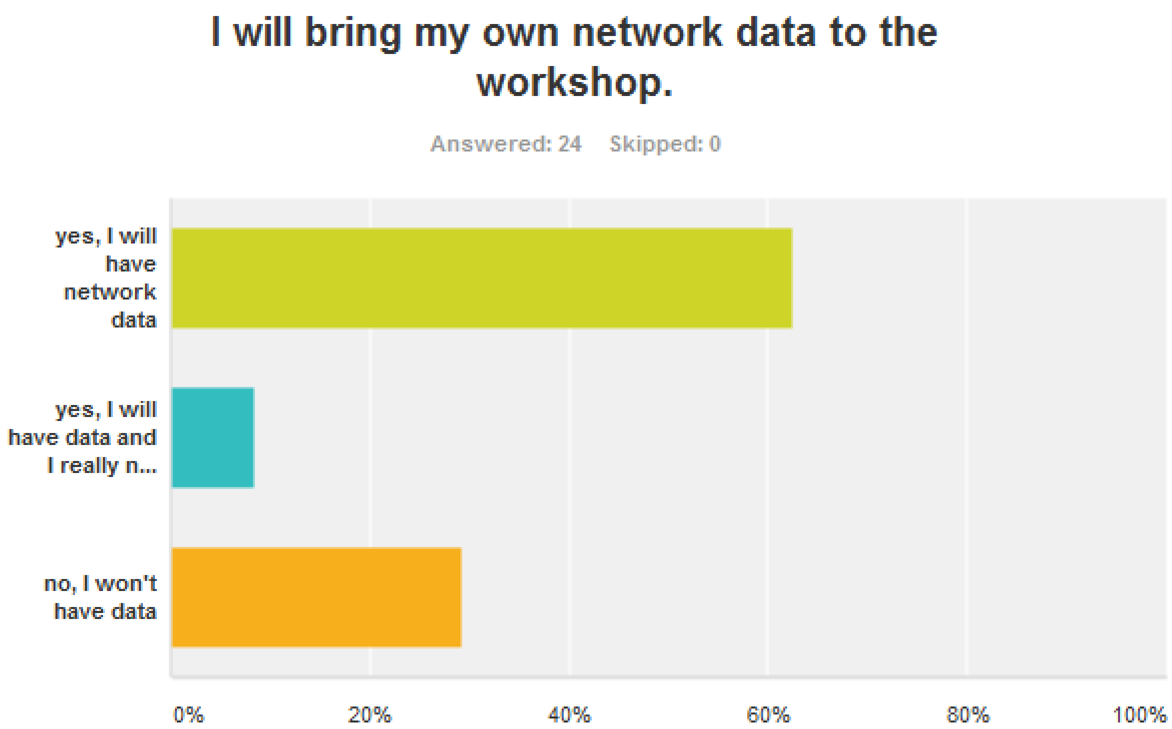
Lothar Krempel

Ju-Sung Lee
(Juice)

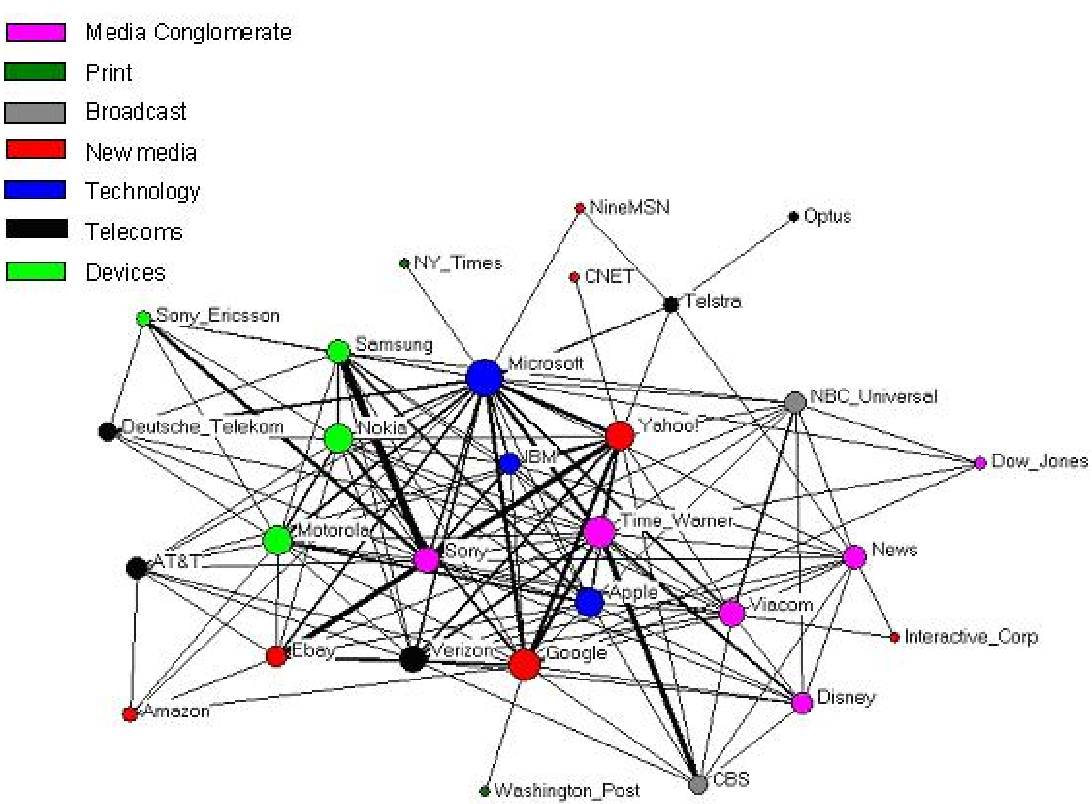
Ian McCulloh

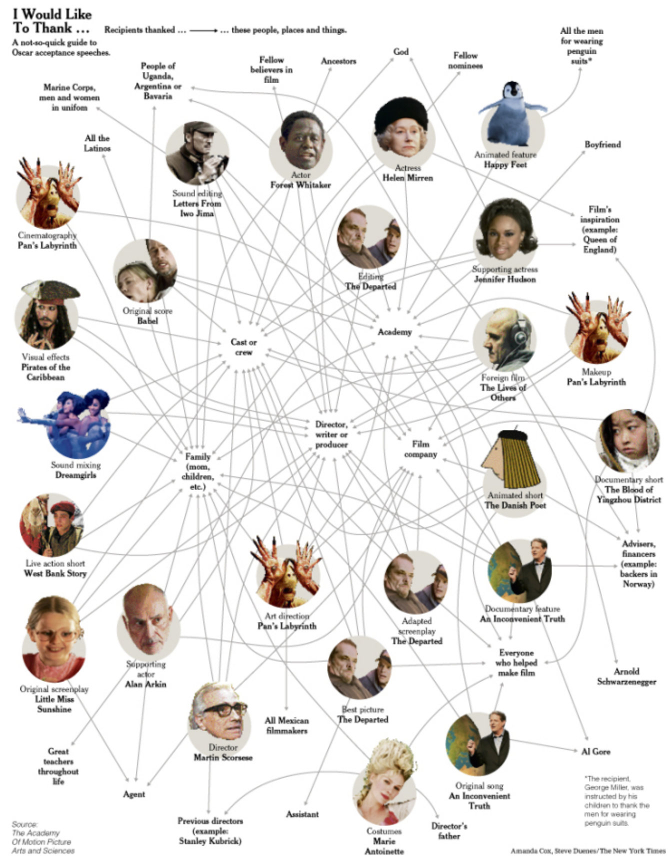
Schedule

- 8:00am – 11:00am
- No breaks ;)
- Last 30 minutes for discussion your data



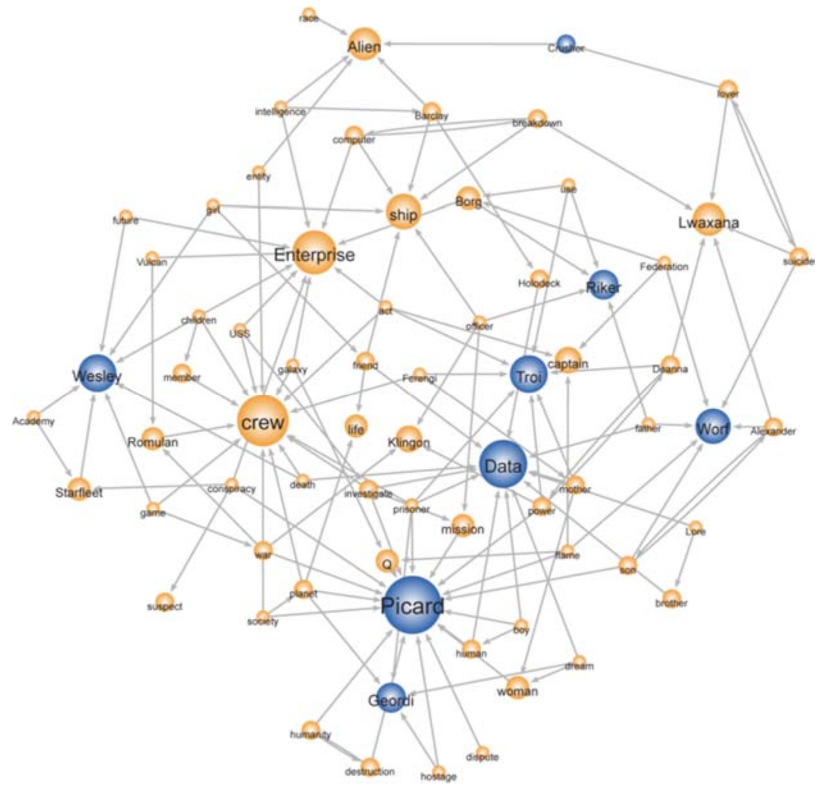
Issues Related to Network Figures?





Agenda

- Fundamentals of information visualization
- Visual elements for drawing networks
- Multivariate information visualization with networks
- Communicating with colors
- Human perception
- Reducing visual complexity
- Post-processing for print
- Check-lists for better figures



Goal:

Giving you the ability to assess the quality of network visualizations and to draw better network pictures by yourself.

Why Pictures?

- “Efficient communication of information” [Tufte 2001]
- Graphical representation of multivariate (high dimensional) data [Tufte 2001]:
  - visual evidence
  - visual reasoning
  - visual understanding
- „Effective translation of information to a system of visual elements“ [Bertin 1984]
- „The faster the information is understood, the more effective is the visualization“ Krempel 2005]

Why Pictures?

1. Parallel processing of pictures

- Language and writing required sequential coding
- Graphical communication has a high „bandwidth“

2. Relational cognition of the human brain

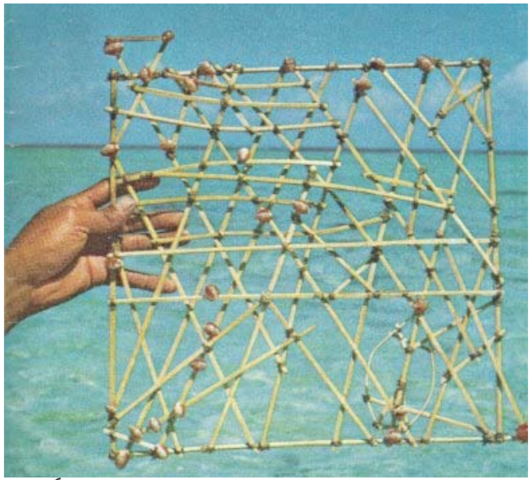
- We think in pictures
- Mental models

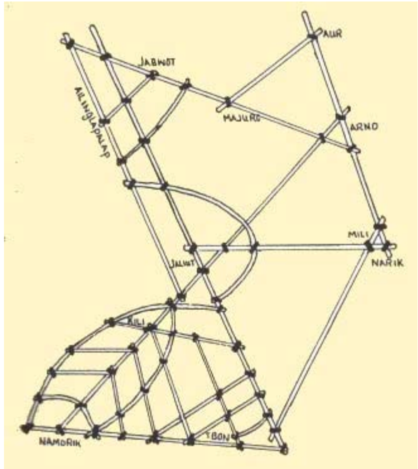
Imagine a map of the large number of islands of Polynesia (Oceania)

Geographical Visualizations

- Ancient Stick Charts
- Used in Micronesia, Polynesia
- Constructed by palm ribs bound by coconut fiber
- Shells used to represent the islands

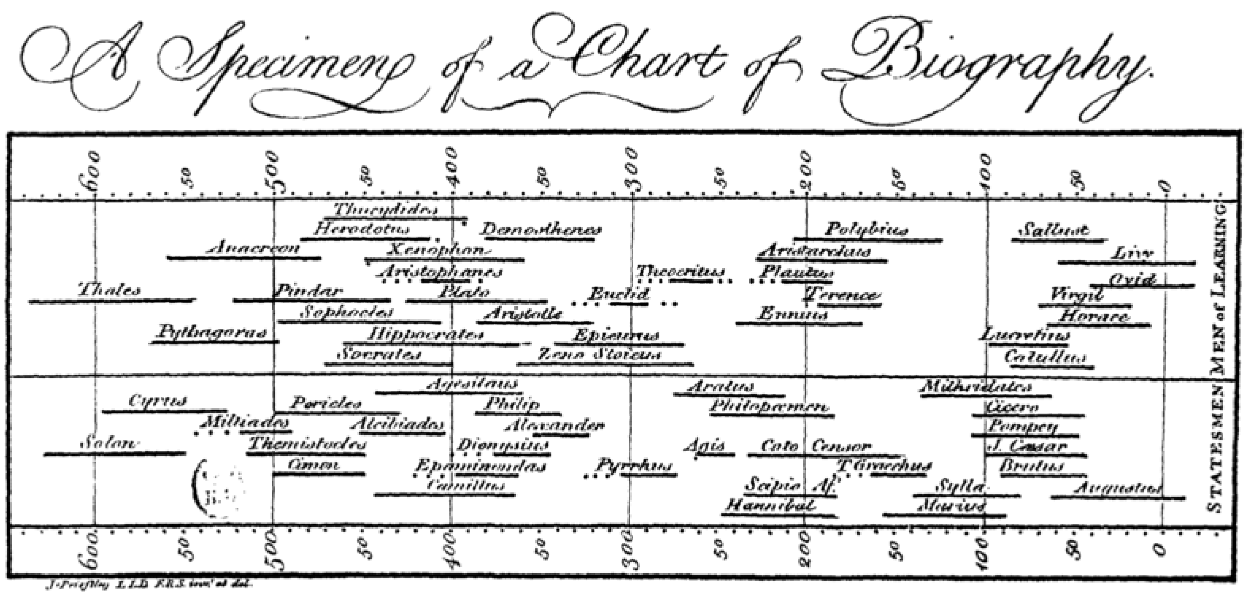
"Rebbilib" stick chart of the Marshall Islands:





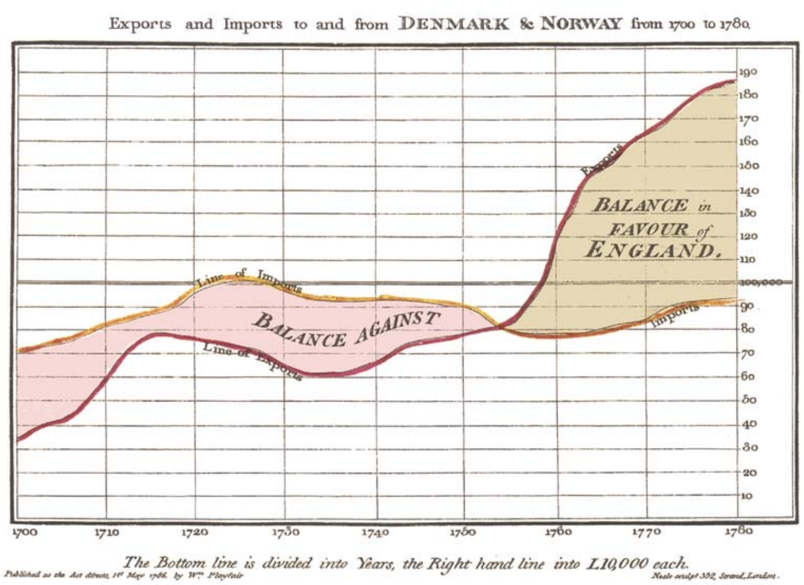
Timeline Charts

- Joseph Priestley
  English theologian, Dissenting clergyman, natural philosopher, chemist, educator, and political theorist ;)
- First timeline charts [1765]
- Lines to visualize the life span of a person
- Compare the life spans of multiple persons



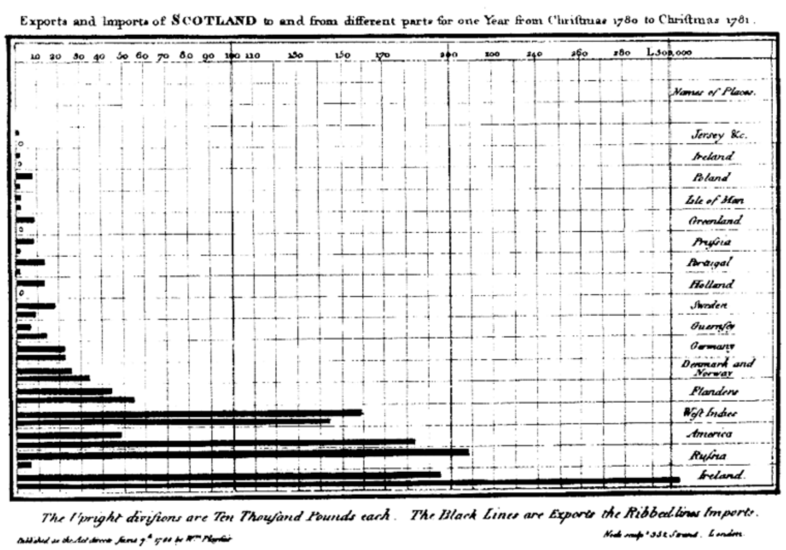
William Playfair

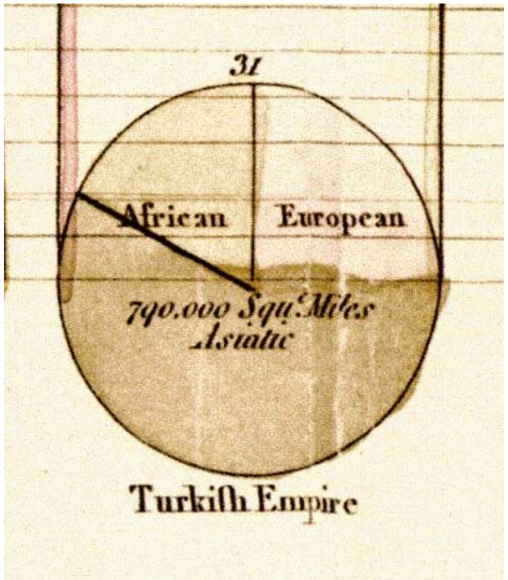
- William Playfair (1759-1823)
- Founder of graphical statistics
- “the increasing complexity of modern commercial life”
- Commercial and Political Atlas, 1786
- 1786: The line graph (trade-balance time-series chart)



William Playfair

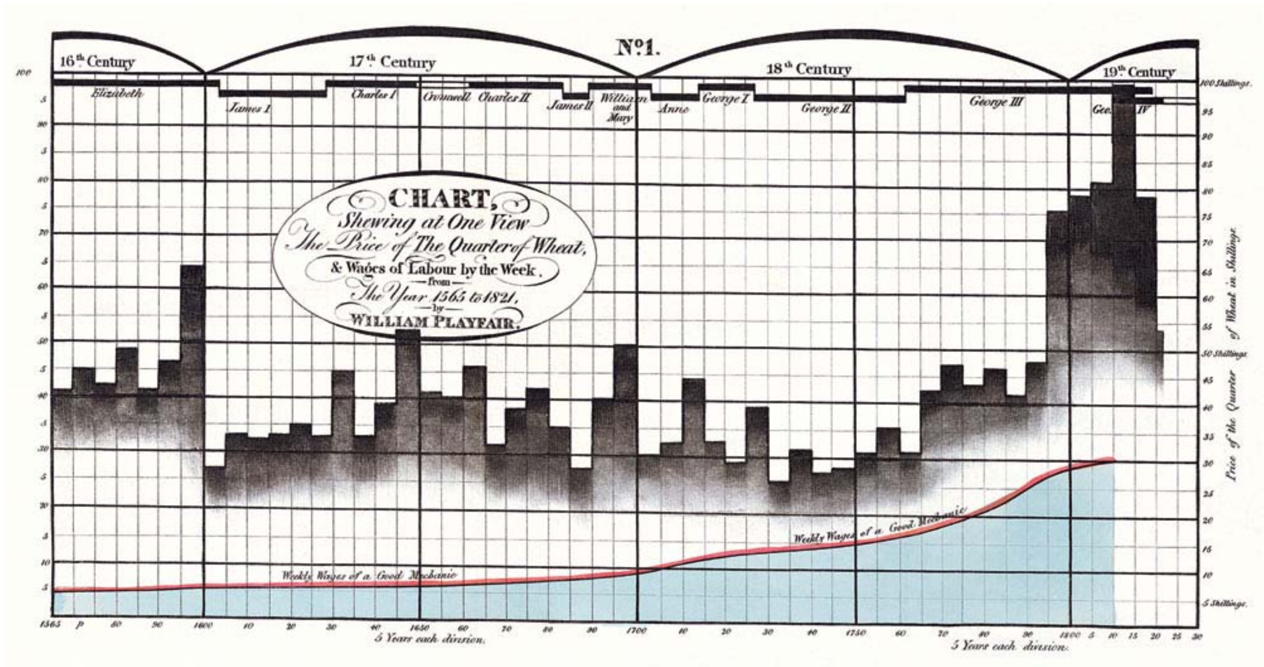
- 1786: The bar chart (Scotland's imports and exports from and to 17 countries in 1781)
- 1801: The pie chart and circle graph (the proportions of the Turkish Empire located in Asia, Europe and Africa before 1789)





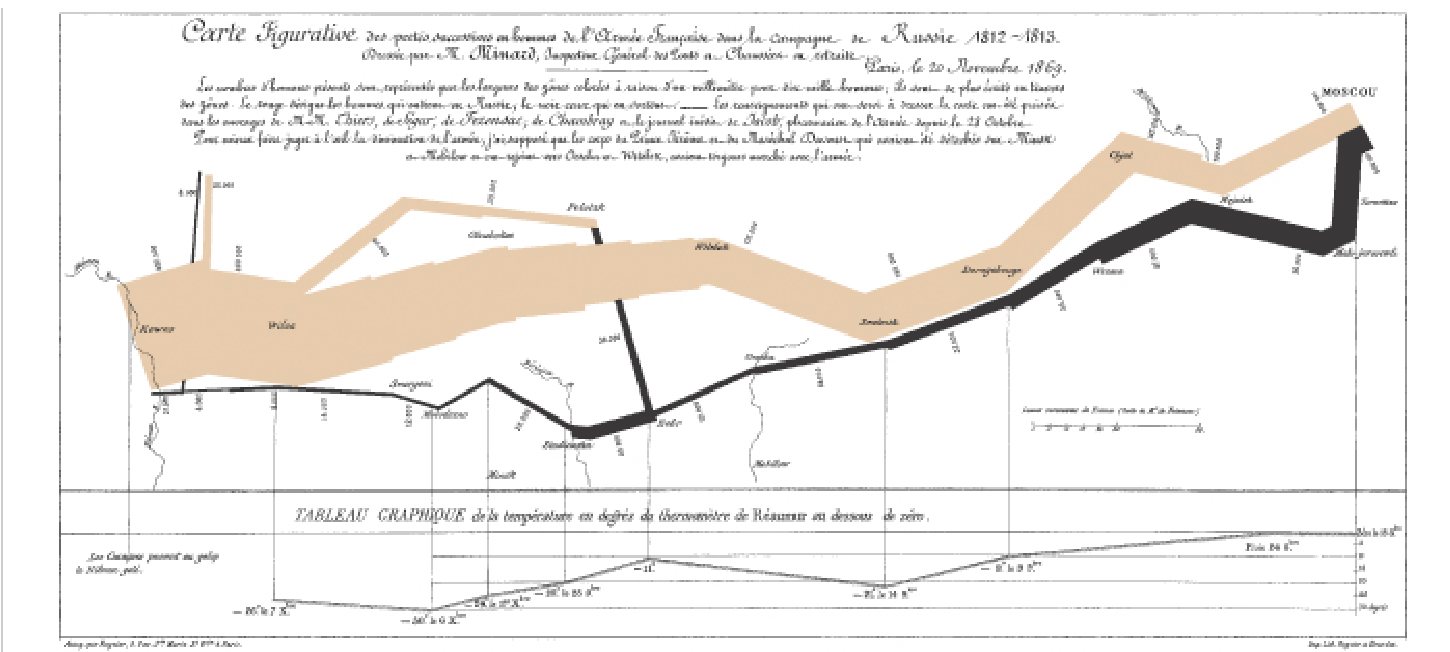
William Playfair

- Multivariate visualizations
- 1821: The "weekly wages of a good mechanic" and the "price of a quarter of wheat" over time
- Visualizations as propaganda



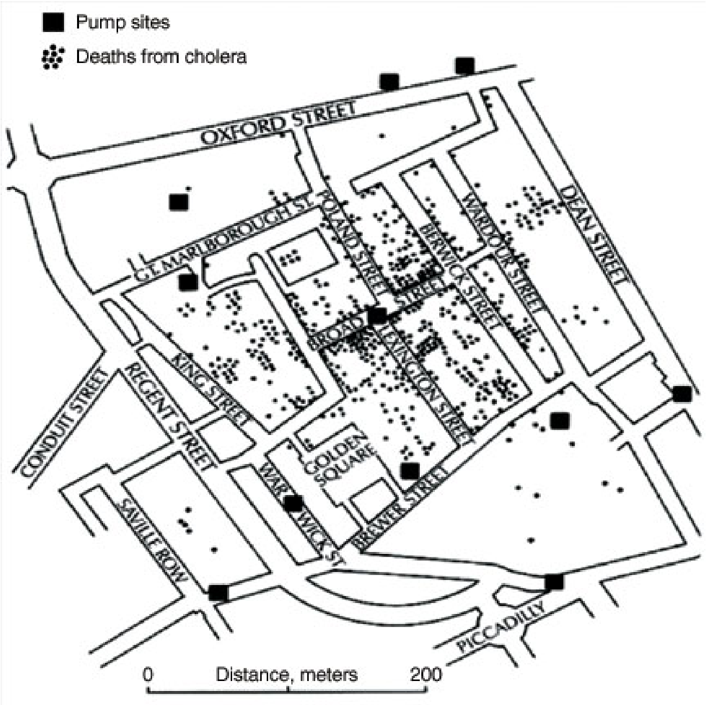
Time & Space & Data

- Charles Joseph Minard
- Map of Napoleon's March to Moscow. The War of 1812–1813”
- „It may well be the best statistical graphics ever draw.“ [Tufte 2001]



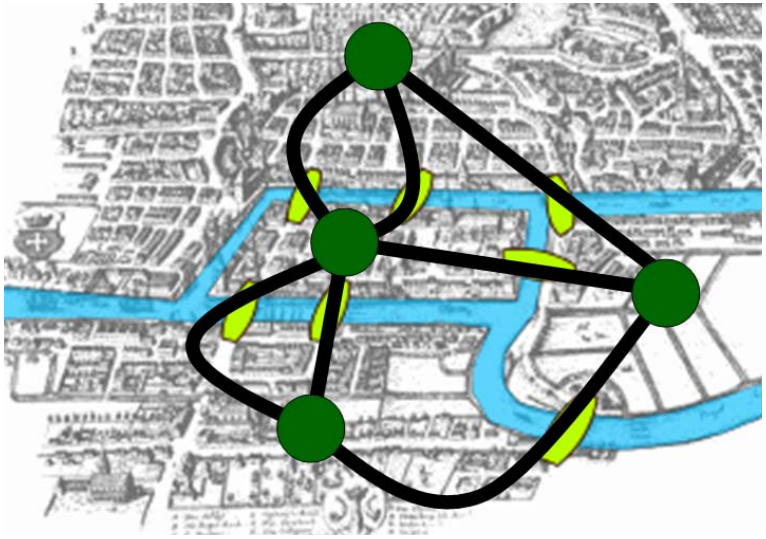
Visual Reasoning

- London, a 10-day period in September 1854
- More than 500 people died of cholera
- Map showing the disease convinced authorities to close the Broad Street water pump



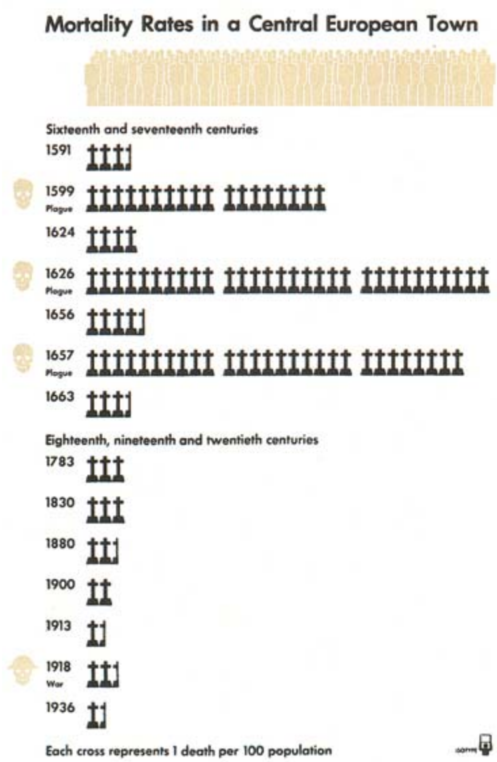
Abstraction

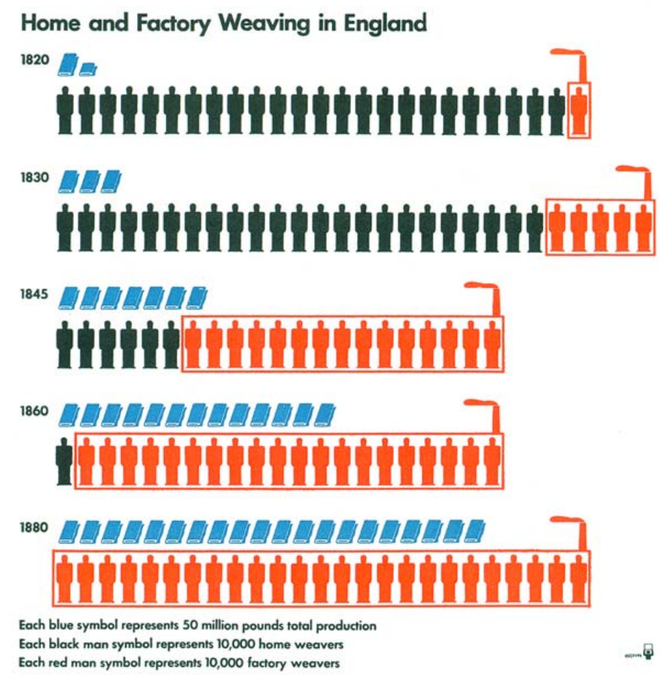
- Königsberg in Prussia (now Kalinigrad, Russia)
- Question: Is it possible to find a round trip through the city by passing every one of the seven bridges over the river Pregel?
- Leonhard Euler 1736: Just the structure is important, not the details



Symbols

- Otto Neurath, 1927
- ISOTYPE: International System of Typographic Picture Education

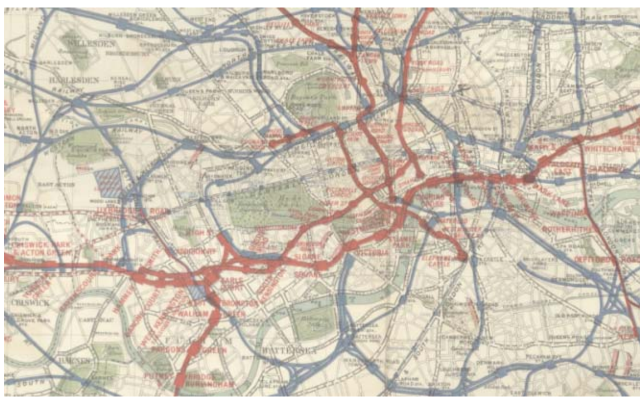




Abstraction

- London Underground
- From geographical visualization to data visualization

1905



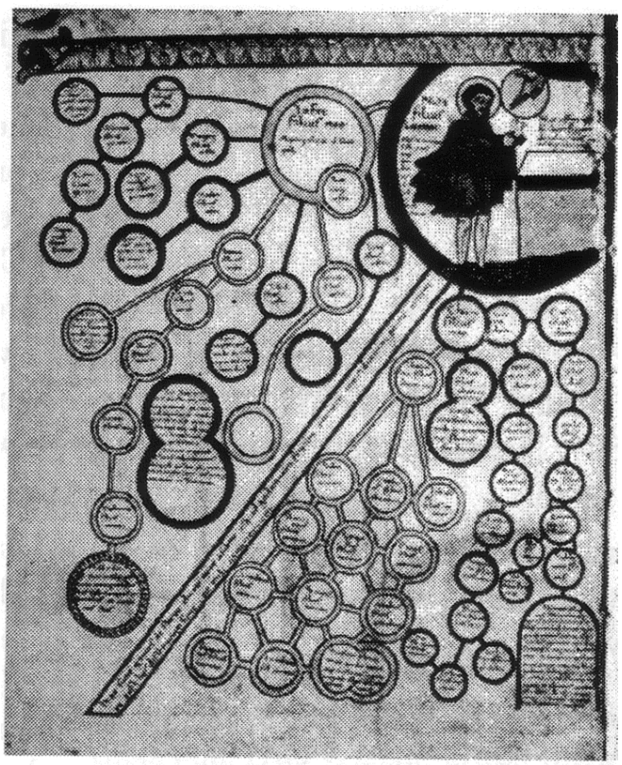
1908

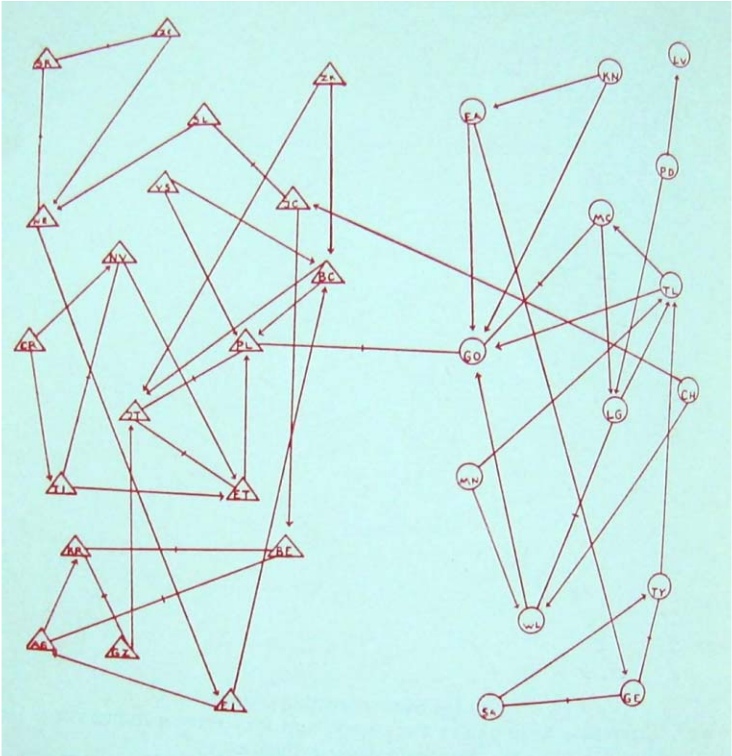
1921

1933

Network Visualizations

- Family Trees (medieval)
- Sociometry, Moreno (1934)





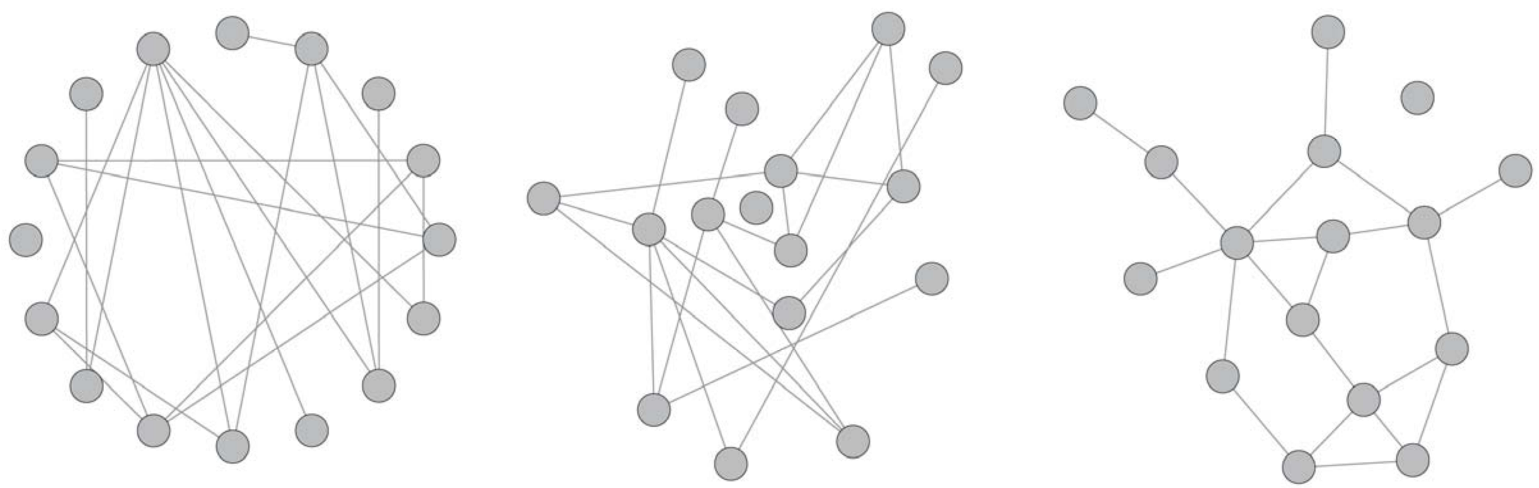
Visualizing Networks

- Explorative visualizations – find something
  - First impressions of the data
  - Validate your network data
- Information visualization – show something
  - What is the information that you want to visualize (substance)?
  - How is it possible to represent this information with your network in a useful way (design)?
  - How to realize this with satisfying approaches (algorithm)?

Network Visualization

Technical but also aesthetical criteria for good networks:

- Show structure
- Optimize distribution on the surface
- Minimize line crossings, maximize angles, and optimize length of lines
- Optimize path distances

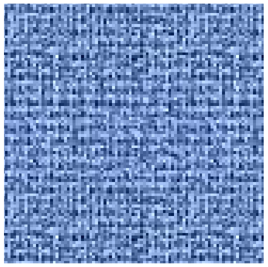


Drawing networks is more than positioning the nodes

Perception

- Preattentive perception…

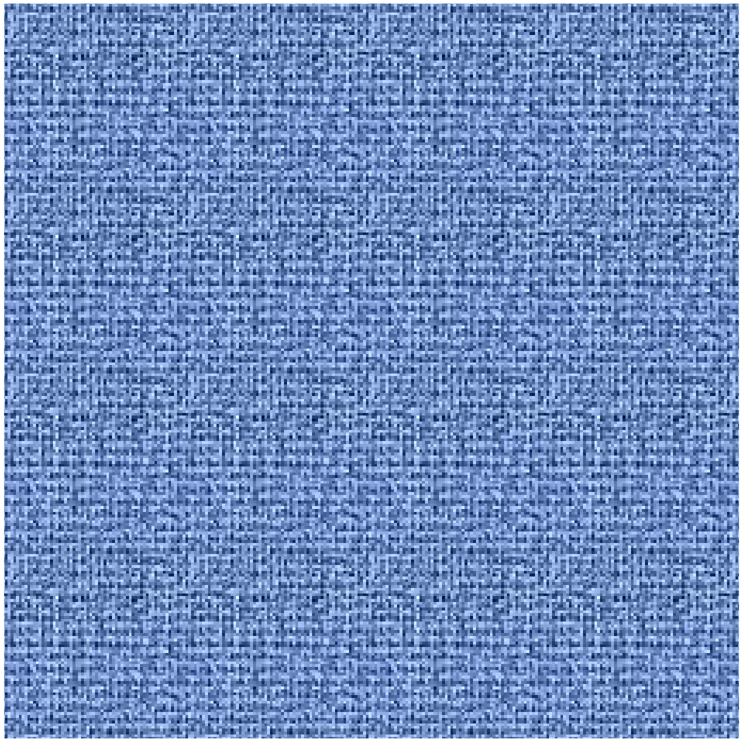
Request for attention!



What Did You See?

Preattentive Perception

- Preattentive elements:
  - Position
  - Size
  - Shape
  - Color
  - Saturation
  - Texture

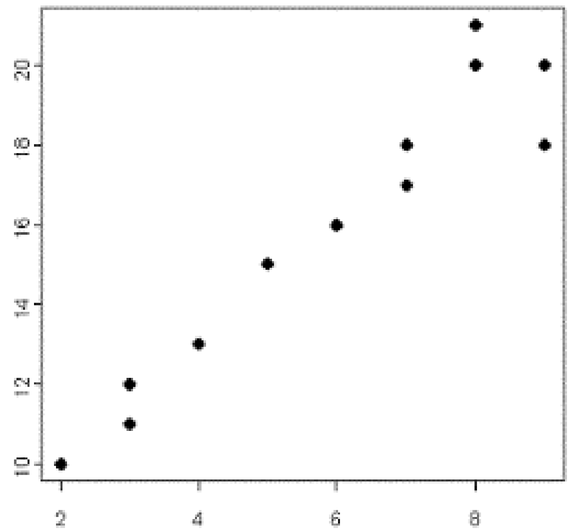


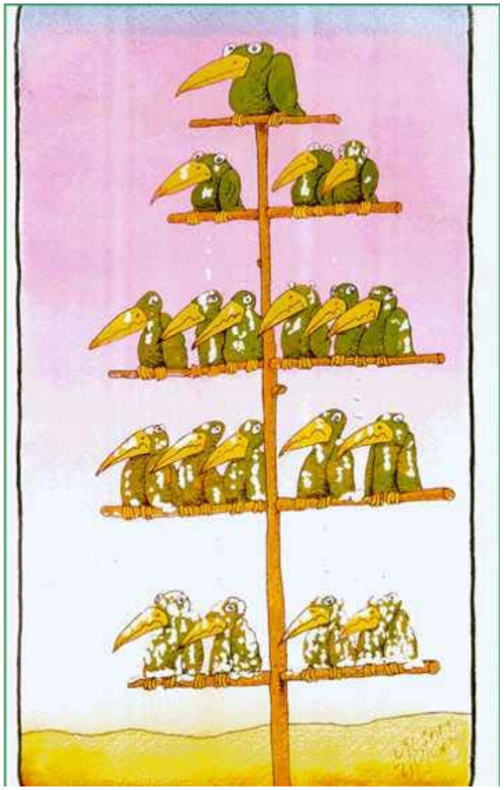
Preattentive Perception

- Preattentive perception:
  - Unconscious collection of information
  - Nervous system can react, no brain activity
  - All information we see, hear,…
  - 200-250 msec.
- Attentive perception
  - Conscious processing of information
  - Analyzing and interpretation

Position

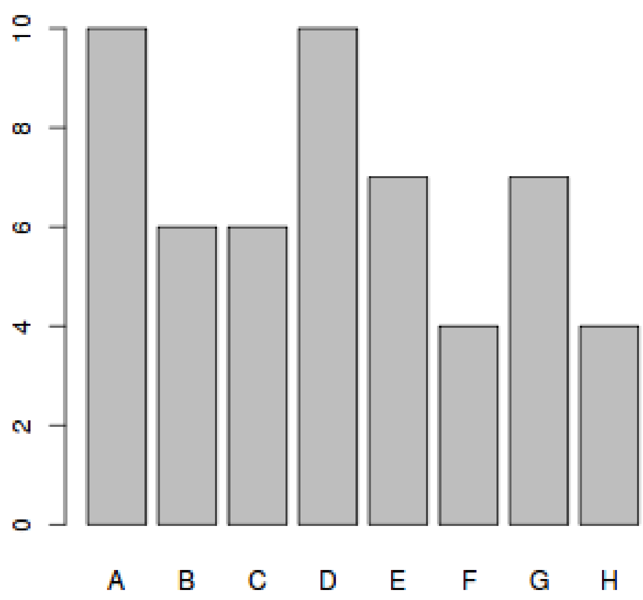
- x-axis, y-axis of elements
  - Left, right, top, bottom
  - Central, peripheral

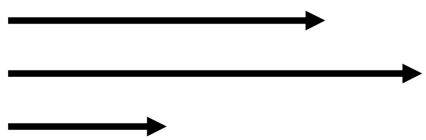




Length

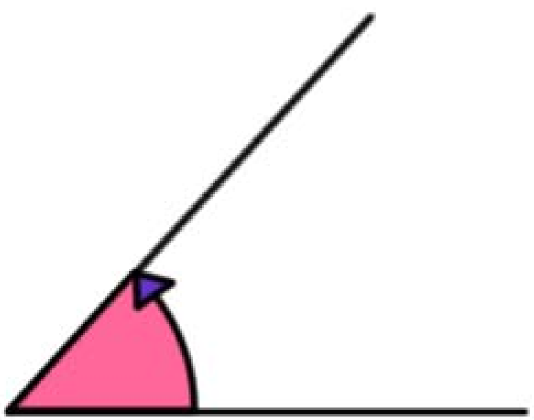
- Length, width, height of elements
  - Longer, shorter, taller

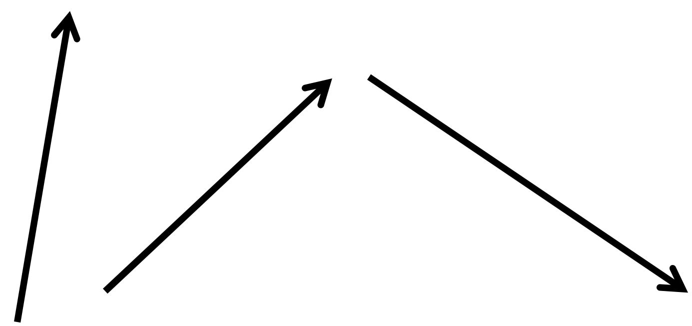


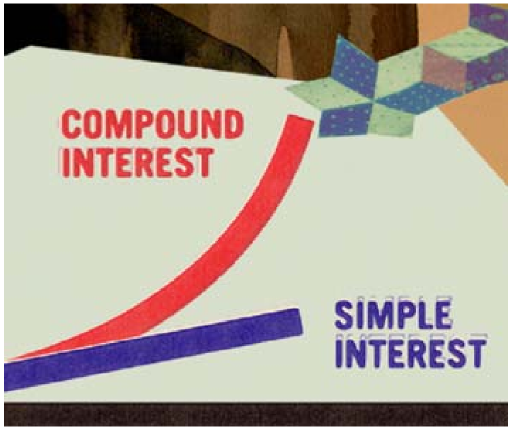


Angle/Slope

- The slope of an element (normally a line)
- The Angle created from two lines
  - Steep, flat, up, down, obtuse angle, …

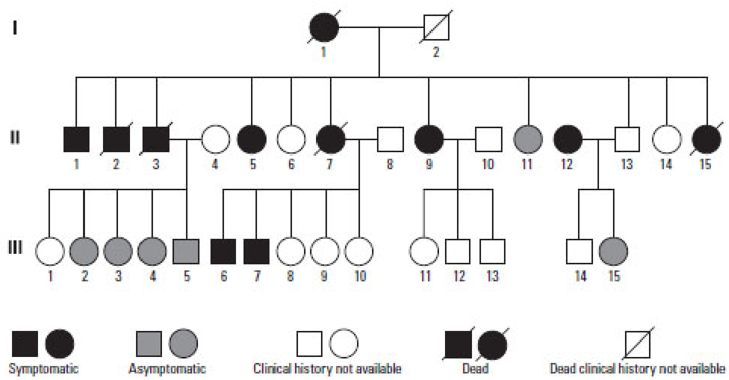






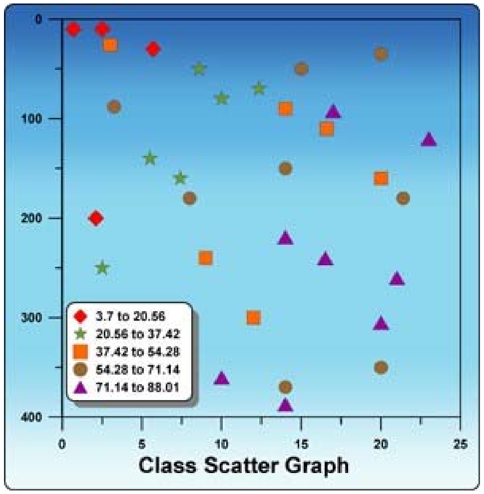
Shape

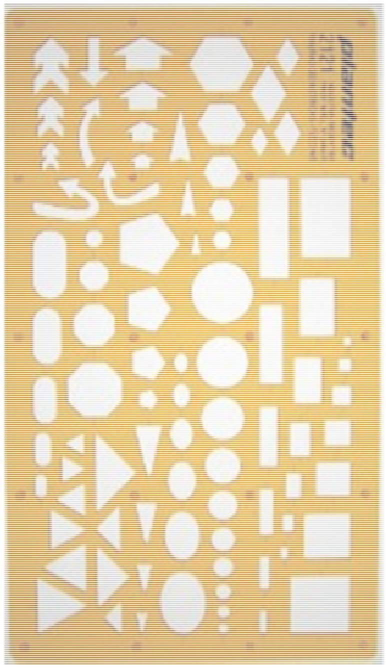
- The form of the elements
  - Squares, circles, triangles,…



II-1 (52 years old); II-11 (64 years); III-5 (38 years); III-6 (45 years); III-7 (52 years); and II-15 (41 years). III-4 (44 years) left the study in 2007 and patient II-11 (64 years) entered in 2008.

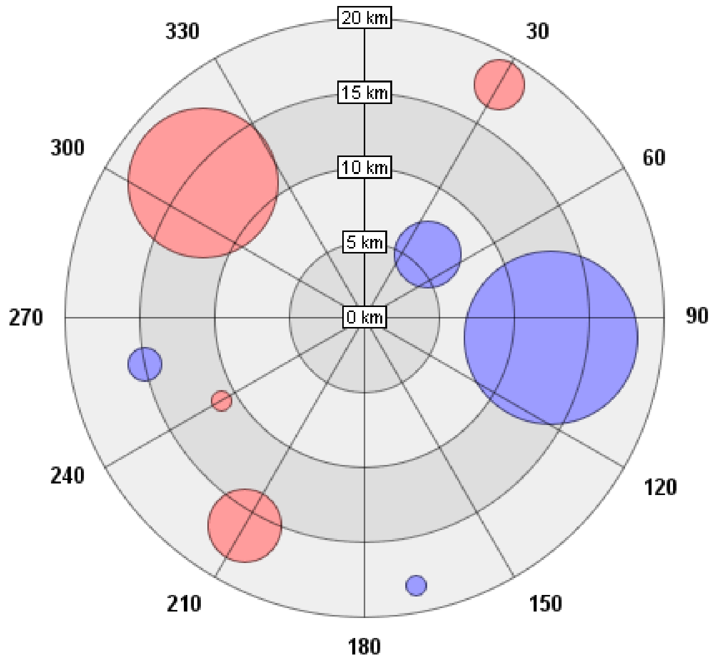
Fig 1. Heredogram of the study family.





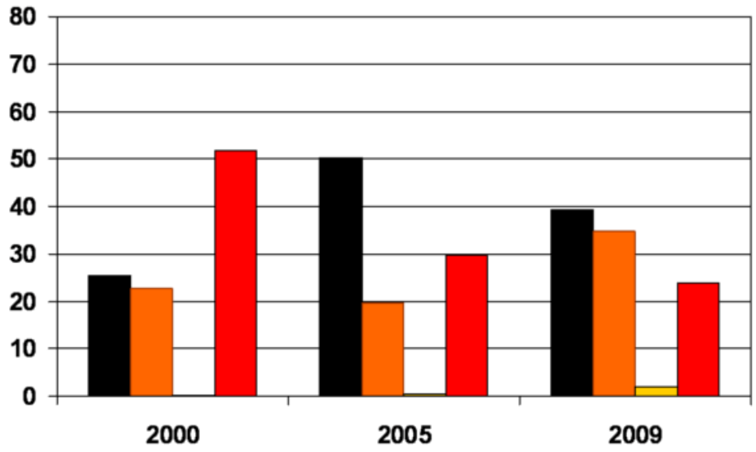
Area/Volume

- Size of elements
  - Larger, smaller



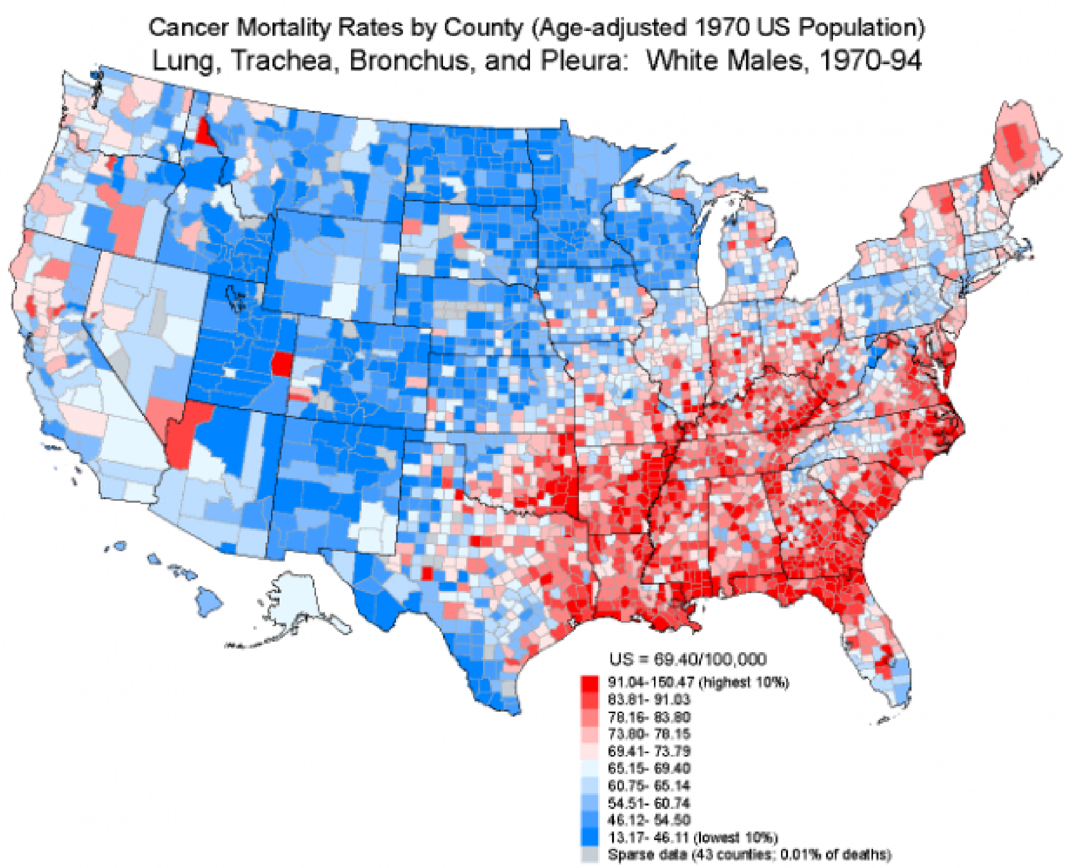
Color Hue

- Color of elements
  - Red, black, blue,…



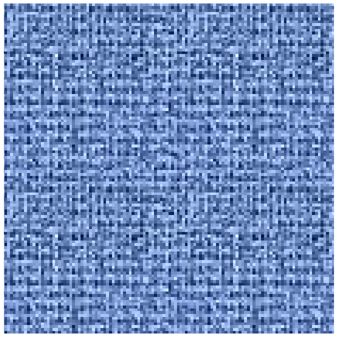
Color Saturation

- Saturation of colors of elements
  - Light, dark, color gradient



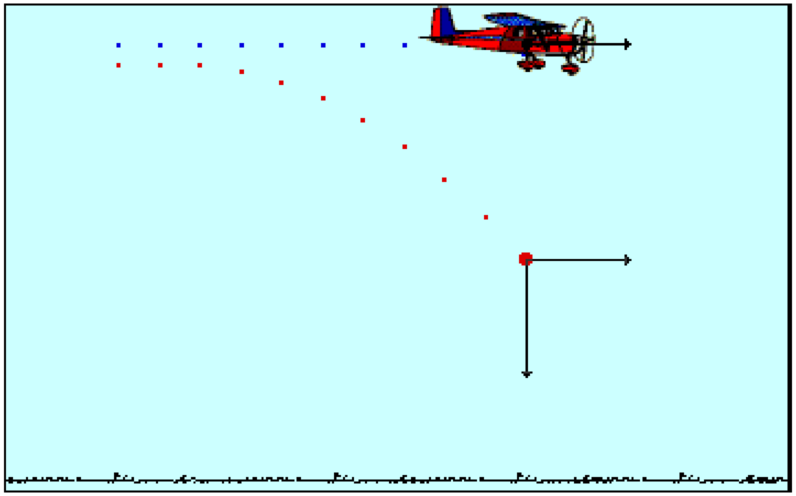
Texture

- Texture
  - Plaid, striped, …



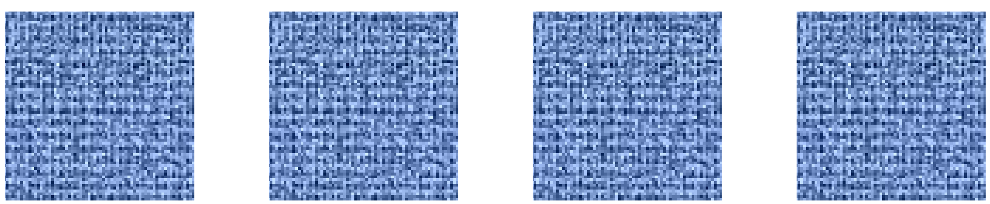
Additional Elements

- Connections, labels
- 3D: z-axis
- Motion: direction, velocity, acceleration



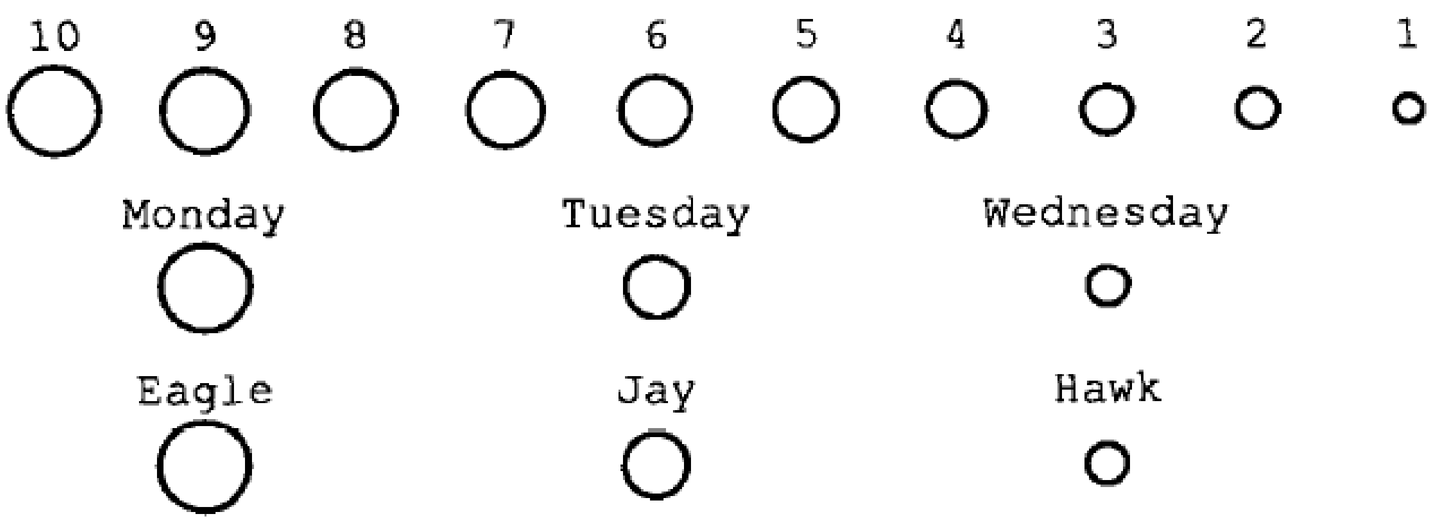
Graphical Elements

- Which graphical elements exist?
- Which are suitable to visualize nominal, ordinal, and quantitative data?
- Which elements help that similar information is perceived as associated?
- Which elements help to distinguish?
- Which elements conserve size differences and which elements distort size differences?



Elements and Data Type

- Quantitative data
- Ordinal data
- Nominal data



[Mackinlay 1986]

Relevance of Elements

Quantitative Information

Position
Length
Angle
Slope
Area/Volume
Color Saturation
Color Hue
Texture
Shape

Angle
Color Hue
Color Saturation
Length
Position
Shape

Relevance of Elements

Ordinal Information

Position
Color Saturation
Color Hue
Texture
Length
Angle
Slope
Area/Volume
Shape

Angle
Area/Volume
Color Saturation
Position
Shape

Relevance of Elements

Nominal Information

- Position
- Color Hue
- Texture
- Color Saturation
- Shape
- Length
- Angle
- Slope
- Area/Volume

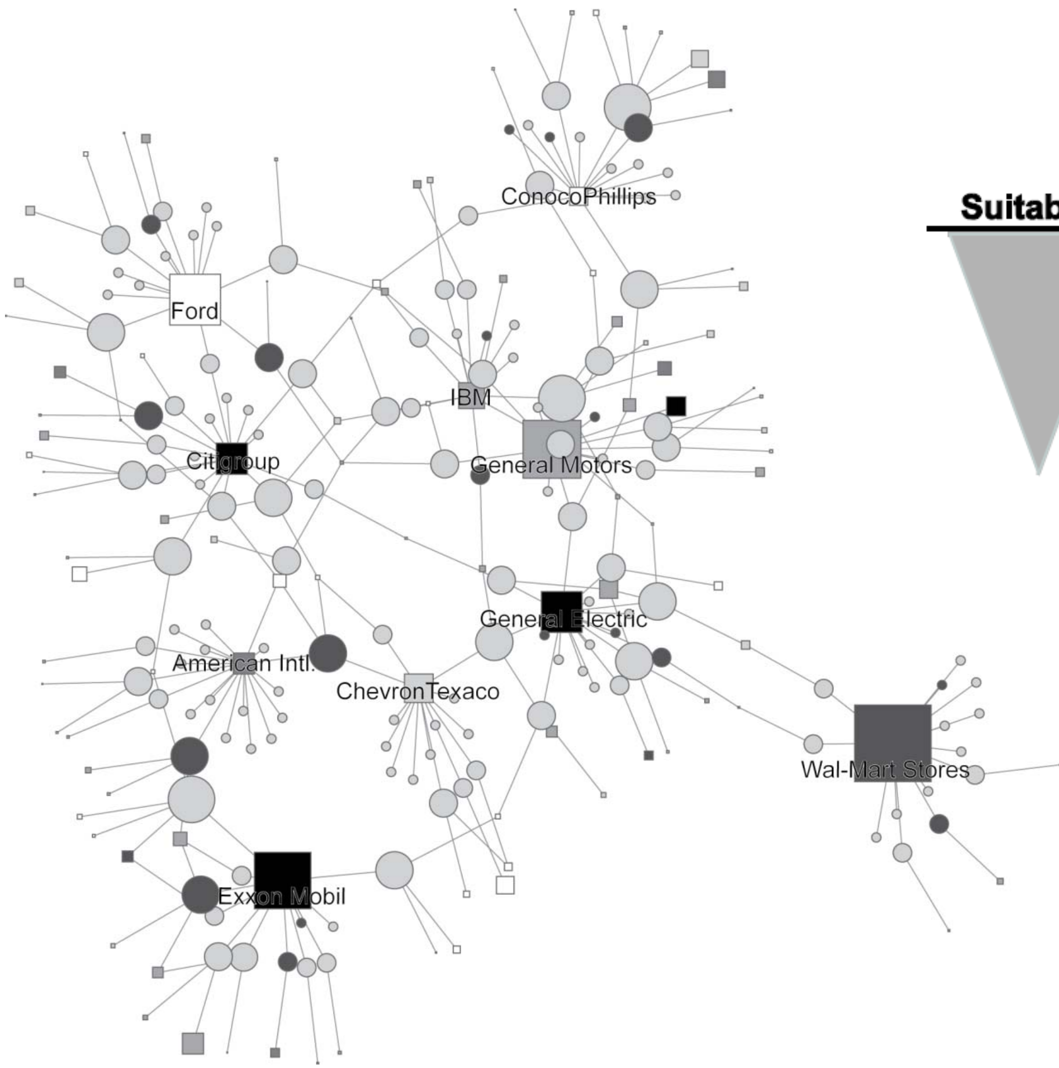
←

- Area/Volume
- Color Hue
- Shape
- Texture

Relevance of Elements

Quantitative	Ordinal	Nominal
Position	Position	Position
Length	Color Saturation	Color Hue
Angle	Color Hue	Texture
Slope	Texture	Color Saturation
Area/Volume	Length	Shape
Color Saturation	Angle	Length
Color Hue	Slope	Angle
Texture	Area/Volume	Slope
Shape	Shape	Area/Volume

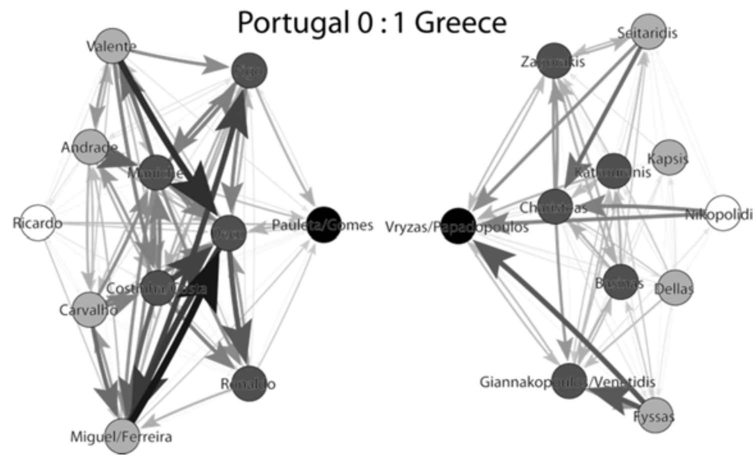
Multivariate Network Visualizations

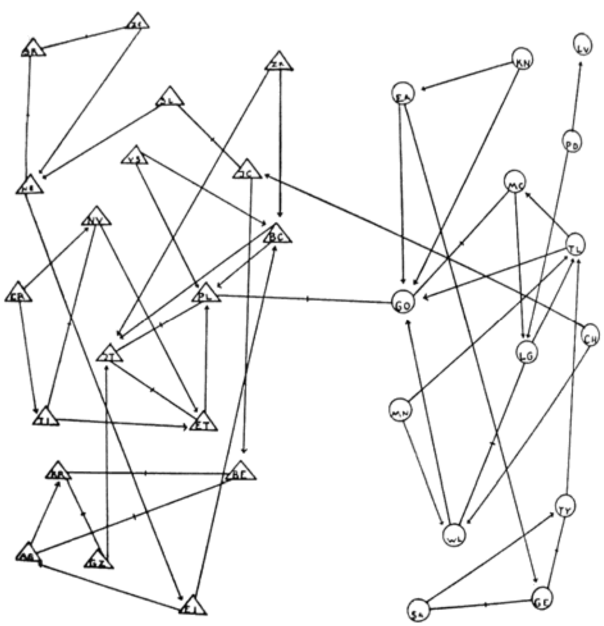


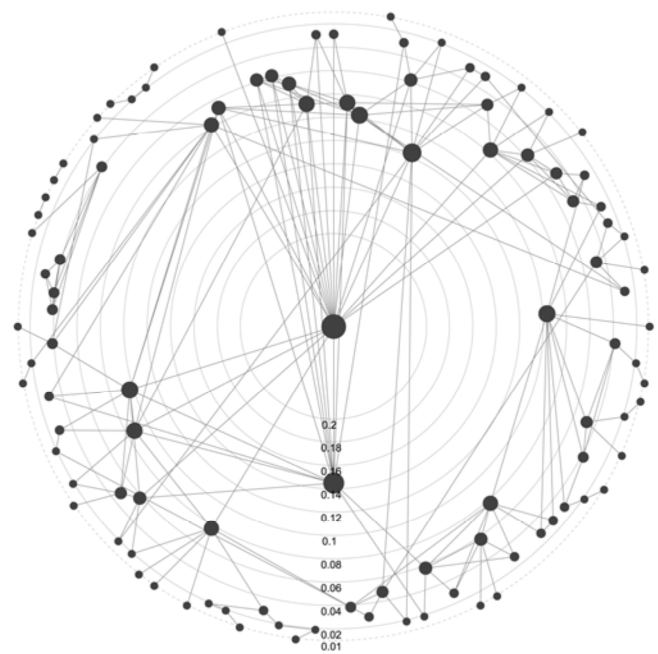
Suitability	Quantitative	Ordinal	Nominal
	Position	Position	Position
	Size	Color Saturation	Color Hue
	Color Saturation	Color Hue	Texture
	Color Hue	Texture	Color Saturation
	Texture	Size	Shape
	Shape	Shape	Size

Special: Substance-Based Layout

- Predefined Layout
- Status
- Centrality
- Attribute Grouping







Special: Scaling Problem

- Psychophysical power law [Stevens 1975]
- Difference between perceived and actual magnitude

$$p = k\ a^{\alpha}$$

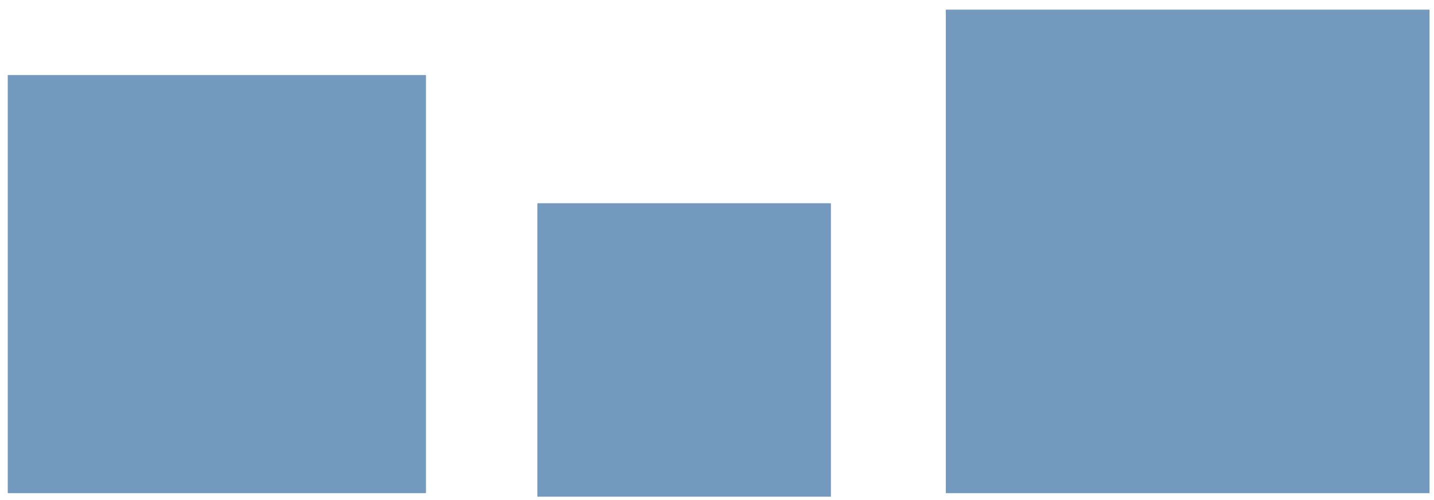
Perceived and Actual Magnitude

Continuum	Exponent	2^{α}	Stimulus
Visual length	**1**	**2.0**	**Projected line**
Visual area	**0.7**	**1.6**	**Projected square**
Redness (saturation)	**1.7**	**3.2**	**Red-gray mixture**
Loudness	0.67	1.6	3000 hertz tone
Lightness	1.2	2.3	Reflectance of gray papers
Taste	1.4	2.6	Salt
Taste	0.8	1.7	Saccharine
Smell	0.6	1.5	Heptane
Cold	1	2.0	Metal contact on arm
Warmth	1.6	3.0	Metal contact on arm
Heaviness	1.45	2.7	Lifted weights
Viscosity	0.42	1.3	Stirring silicone fluids
Duration	1.1	2.1	Noise stimuli

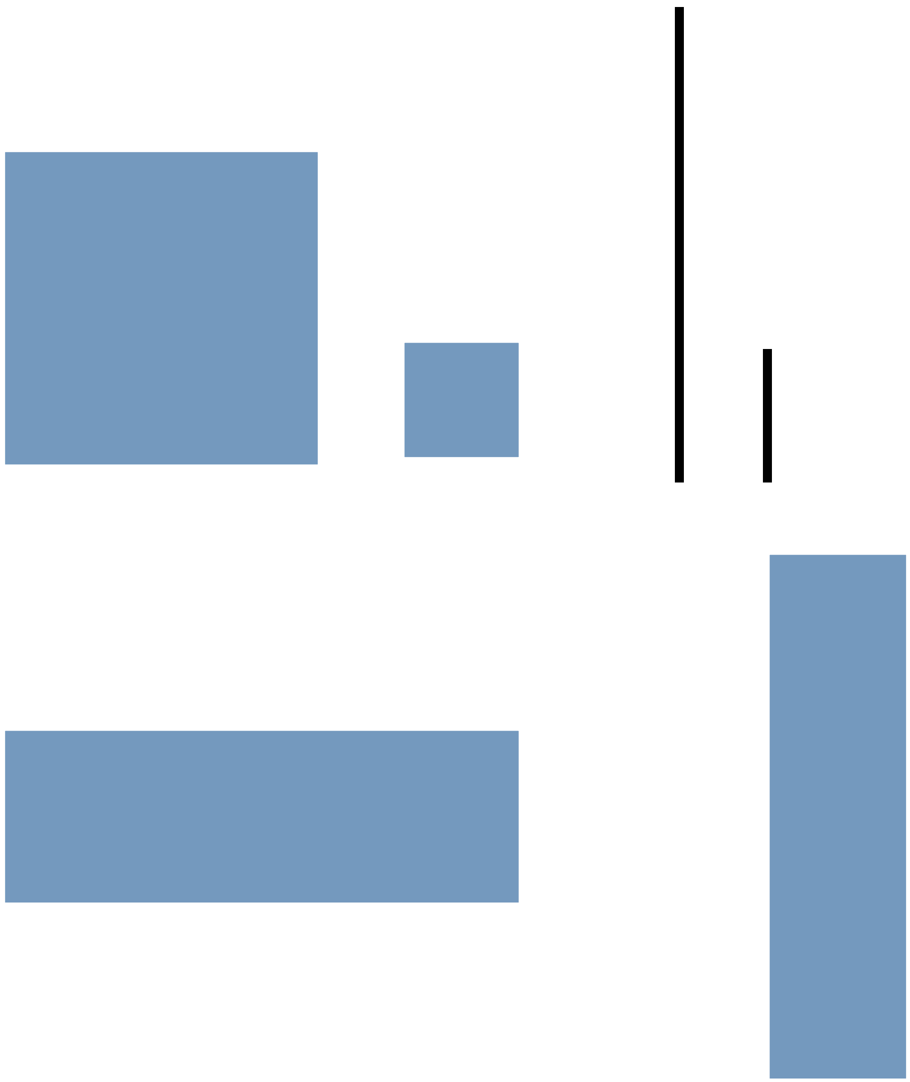
[Lodge 1981]

Perceived and Actual Magnitude

- Visual area:
  - Perceived magnitude: 2.00
  - Actual magnitude: 2.69



1-Dimensional Data, 2-Dimensional Element

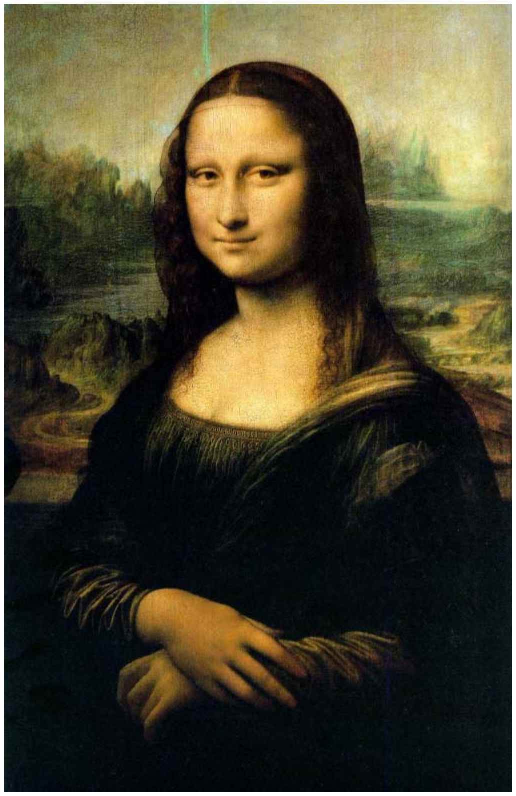


Betweenness Centrality

In-Degree
Out-Degree

Colors

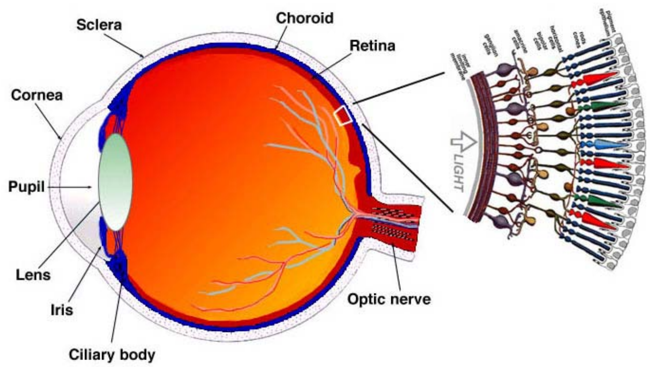
- Similarities: Leonardo da Vinci, Mona Lisa
- Differences: Tizian, Mary’s Assumption

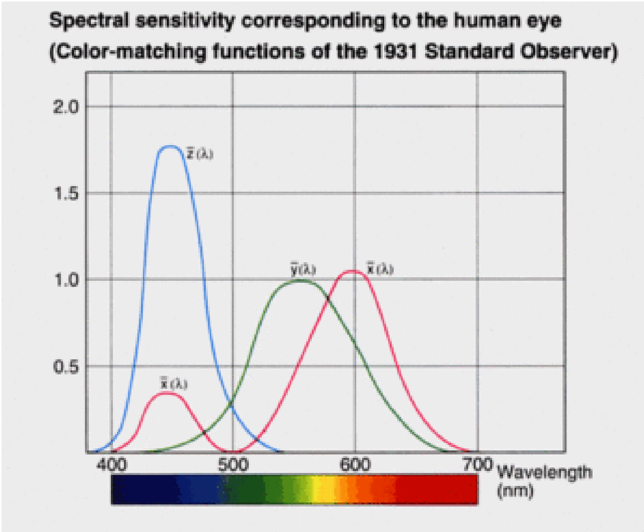


Colors & Differences

Eye/Brain

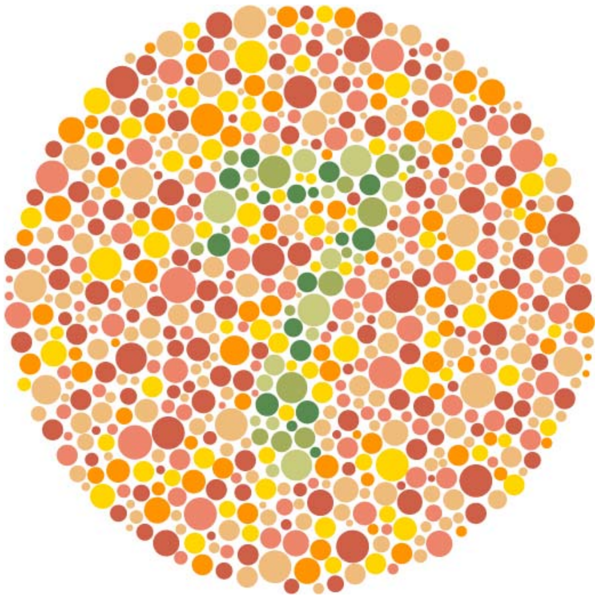
- The lense of the eye focusses the light to the retina
- Retina has color sensitive photoreceptors:
  - Rods: bright and dark differences
  - Cones: colors
- Colors = different wave length of the light:
  - 430 nm (blue)
  - 530 nm (green)
  - 560 nm (red)
- Brain analyzes:
  - Brightness
  - Red, blue, and green color intensities

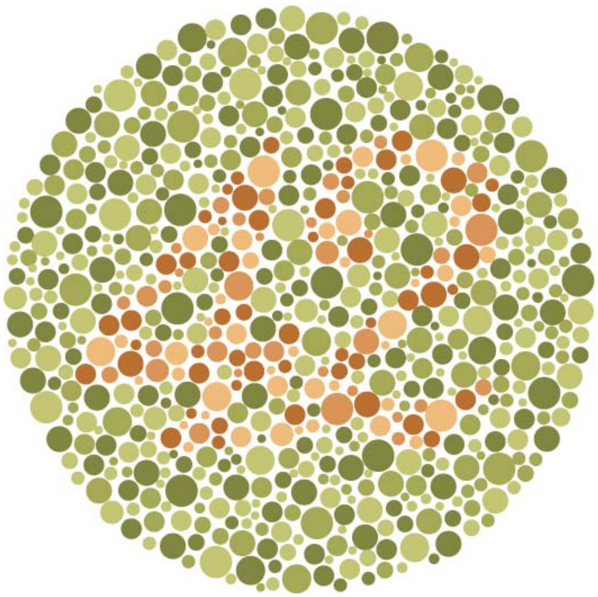




Color Blindness

- Problems with color sensitive photoreceptors
- Approximately 8% of men and 0.5% of women have a genetic condition which causes a typical color perception
- Large majority: red-green



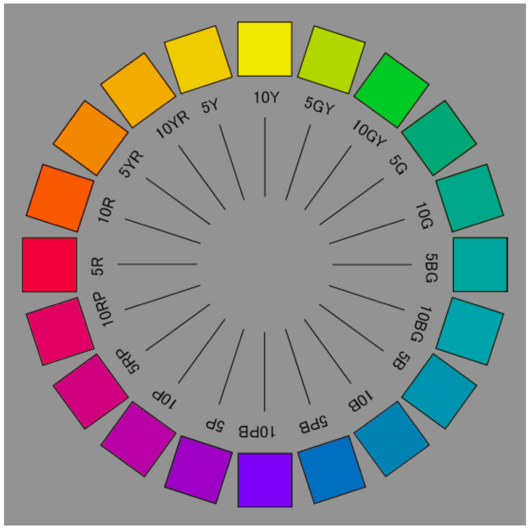


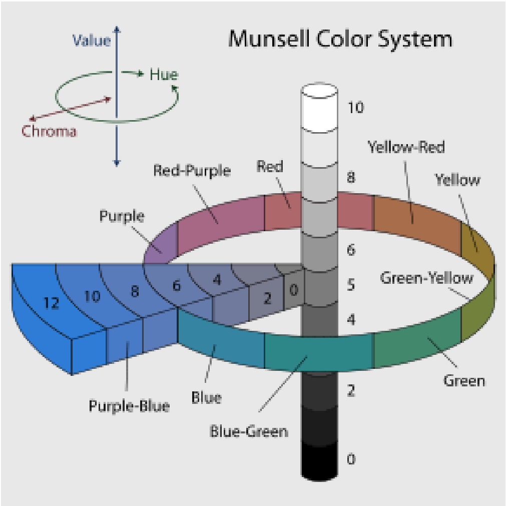
Goethe‘s Color Wheel

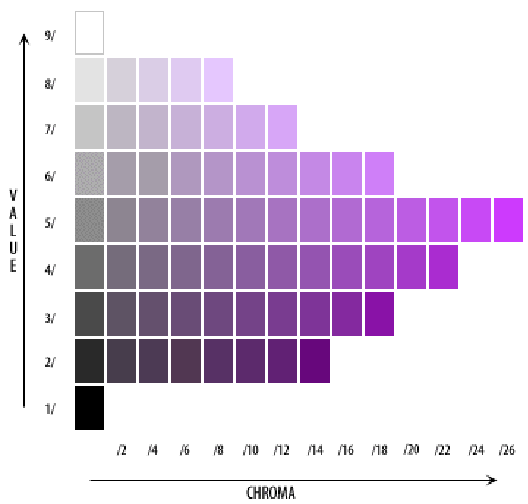
- “That I am the only person in this century who has the right insight into the difficult science of colors, that is what I am rather proud of, and that is what gives me the feeling that I have outstripped many.” (Goethe, 1810)
- Problem: Not perceptually uniform

Perception Oriented Colors

- A. H. Munsell, A Color Notion:
  - Hue (color tint)
  - Chroma (saturation)
  - Value (brightness)
- Create perceptually uniform distributed differences

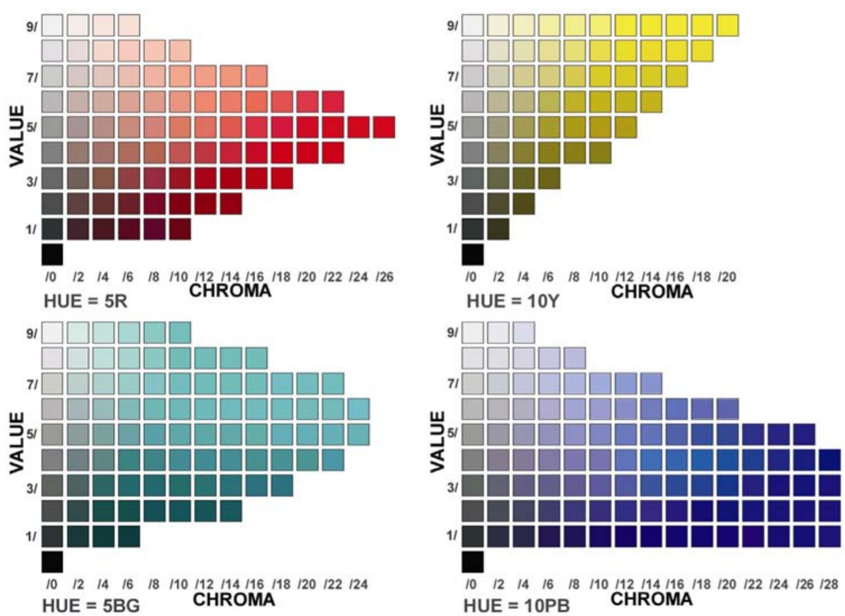






Munsell Color System

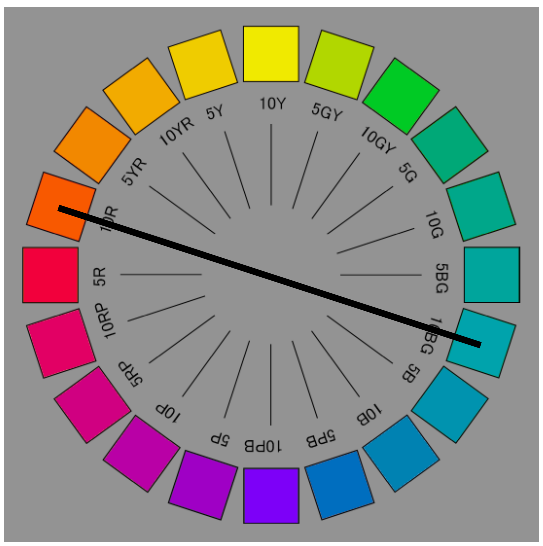
- Perceptually uniform distributed:
- You can change one parameter without the need of changing the other two
- Calculating with colors
- E.g. color saturation (to color ordinal data)



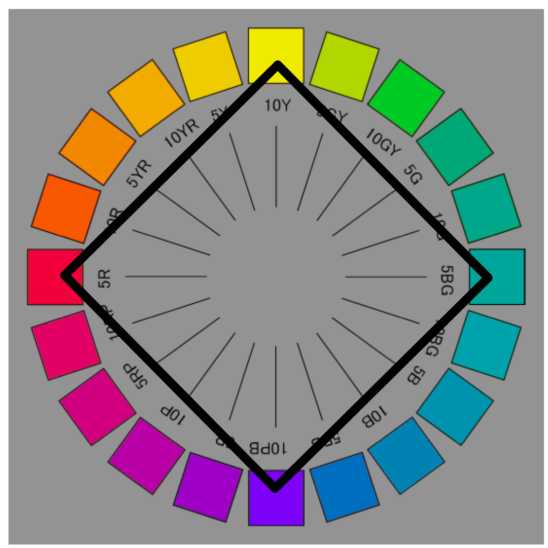
Select Color Hue

- E.g. to color nominal data
- Draw geometric figures into color wheel

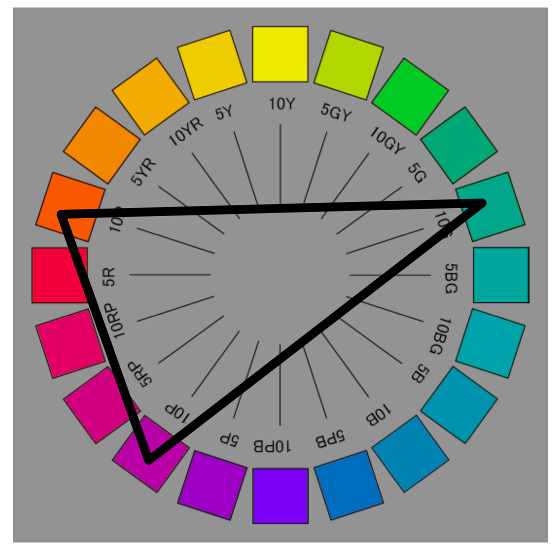
2 colors
Uniform distance



4 colors
Uniform distance



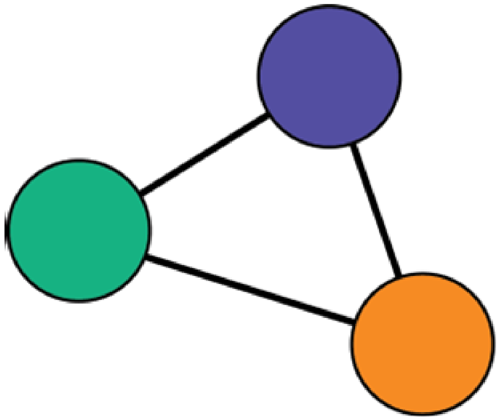
3 colors
Non-uniform distance

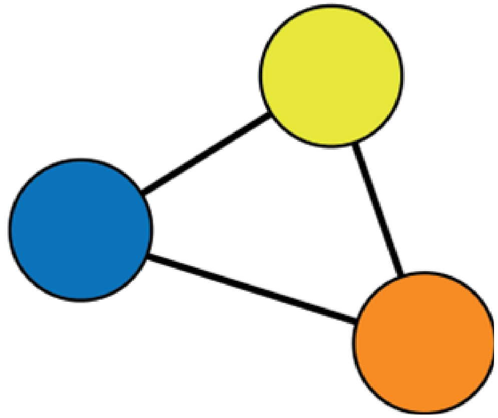


Colors: (Non-)Uniform Distances

a)
- Bank
- Insurance
- Steel Mill

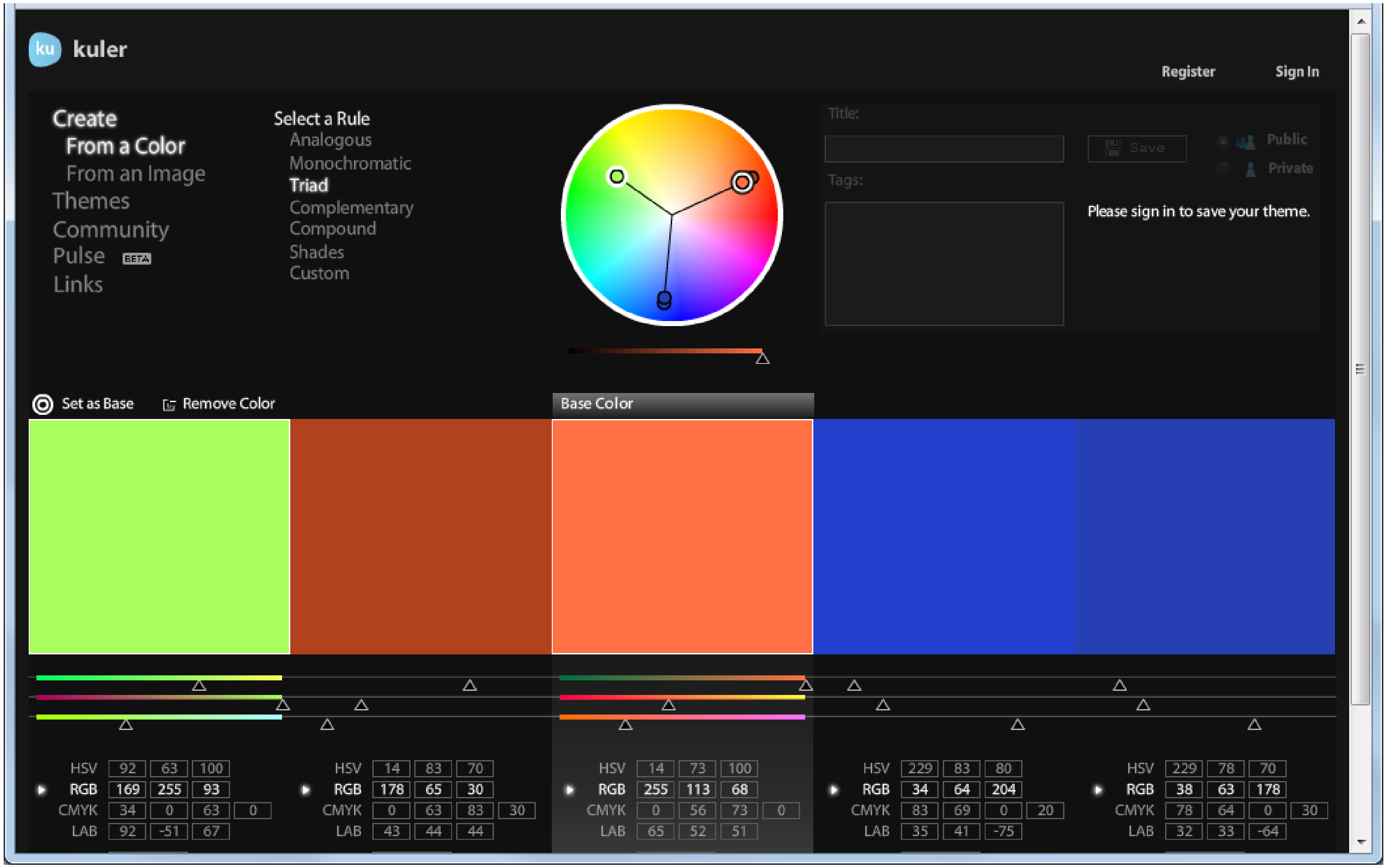
b)
- Bank
- Steel Mill
- Super Market

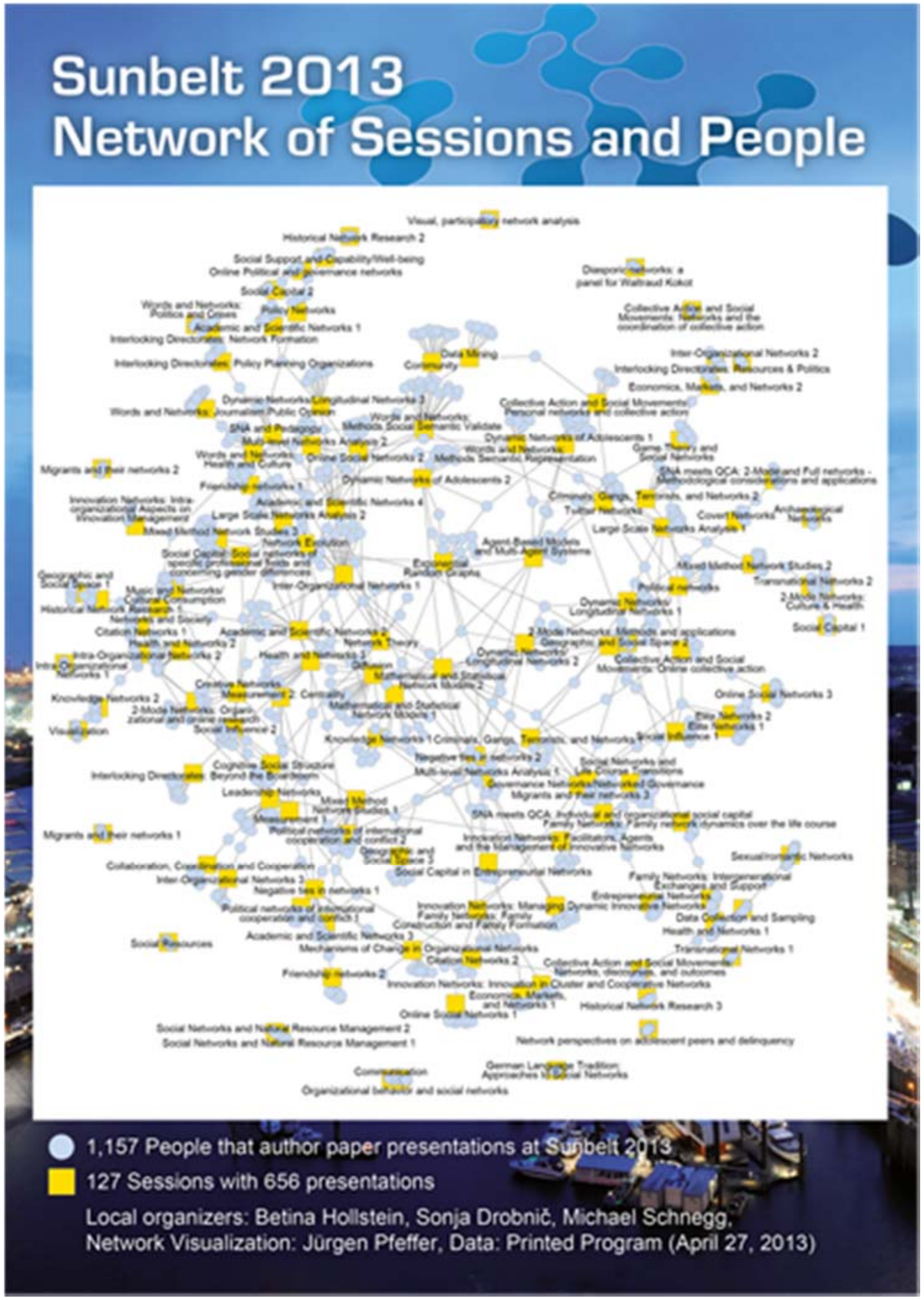




Color Calculator

Adobe Kuler, Software: http://kuler.adobe.com/



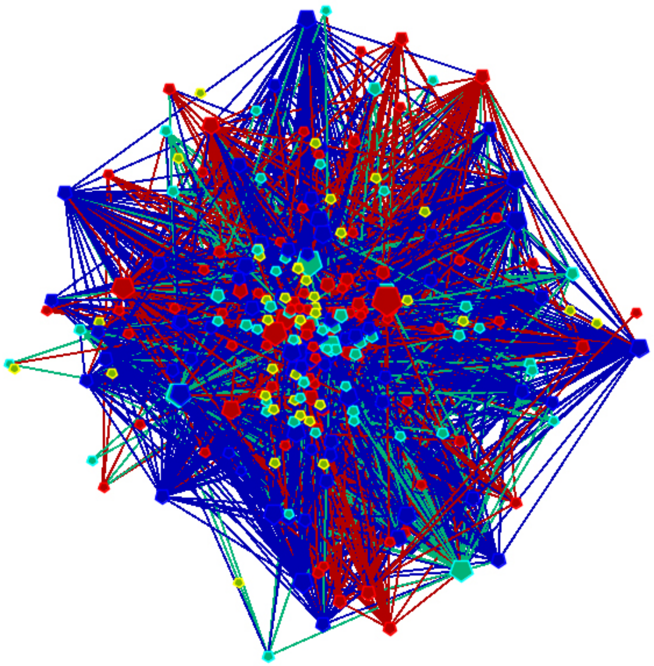

Sunbelt 2013
Network of Sessions and People
1,157 People that author paper presentations at Sunbelt 2013
127 Sessions with 656 presentations
Local organizers: Betina Hollstein, Sonja Drobnič, Michael Schnegg.
Network Visualization: Jürgen Pfeffer, Data: Printed Program (April 27, 2013)

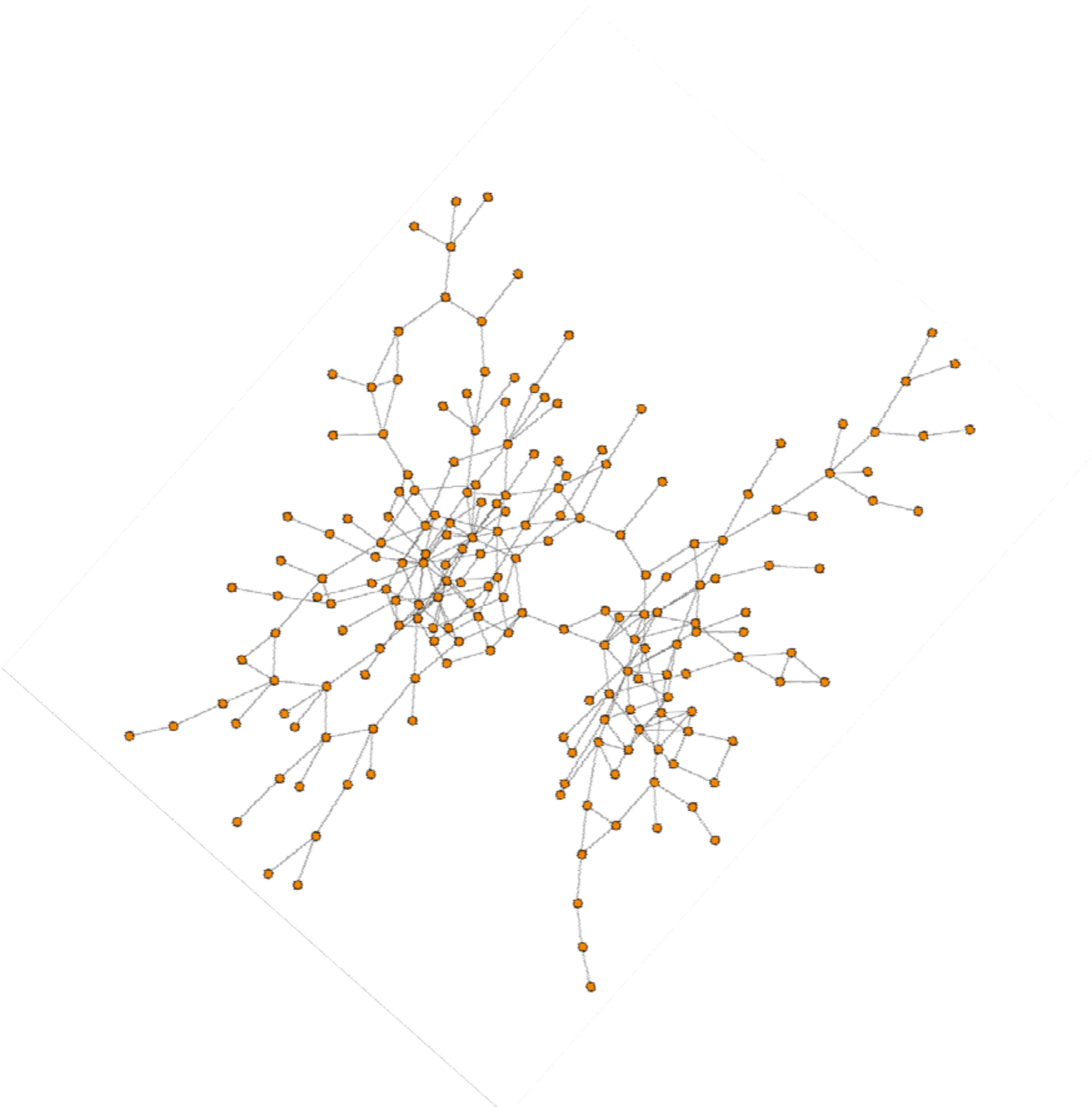
Reducing Visual Complexity

Dense Networks

<u>Problems:</u>

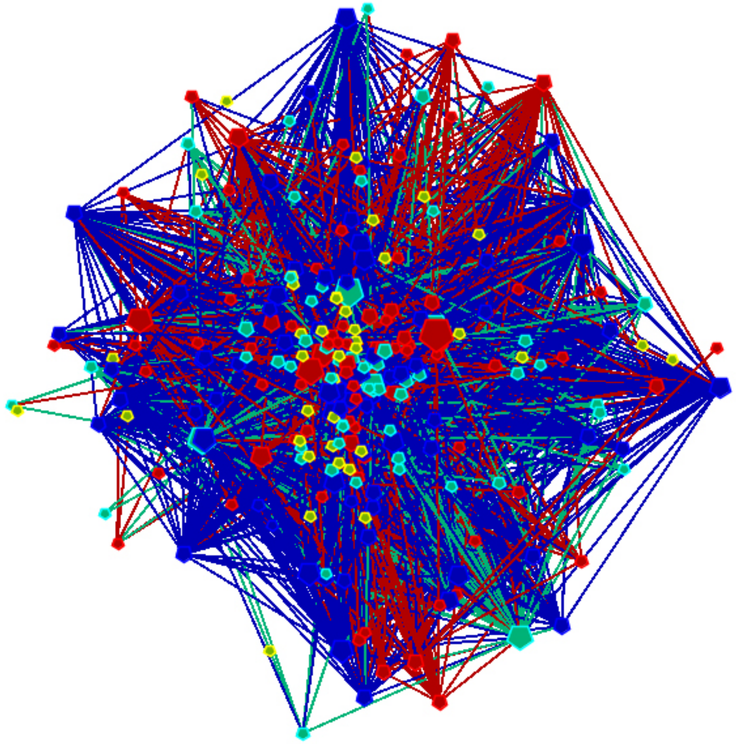
- Impact on measures
- Harder to interpret
- Hard to visualize

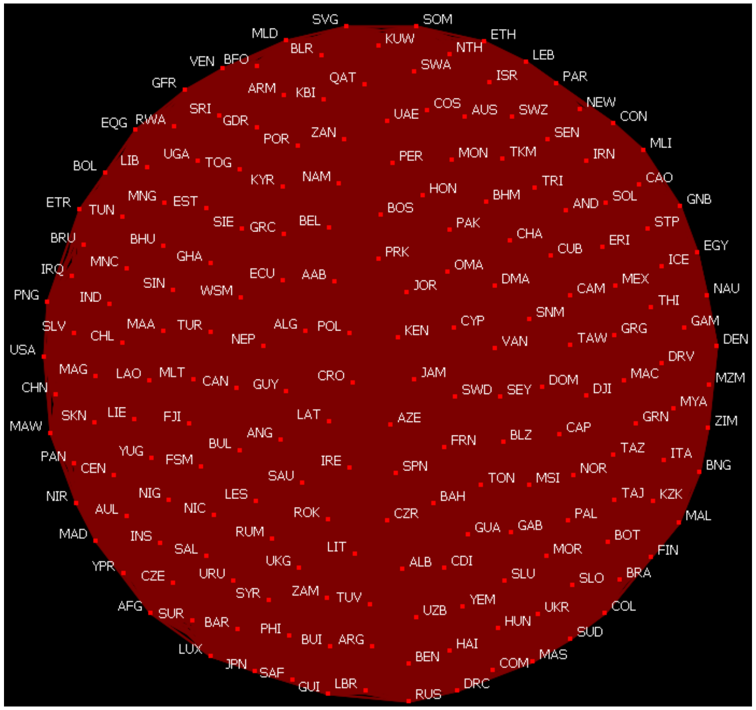




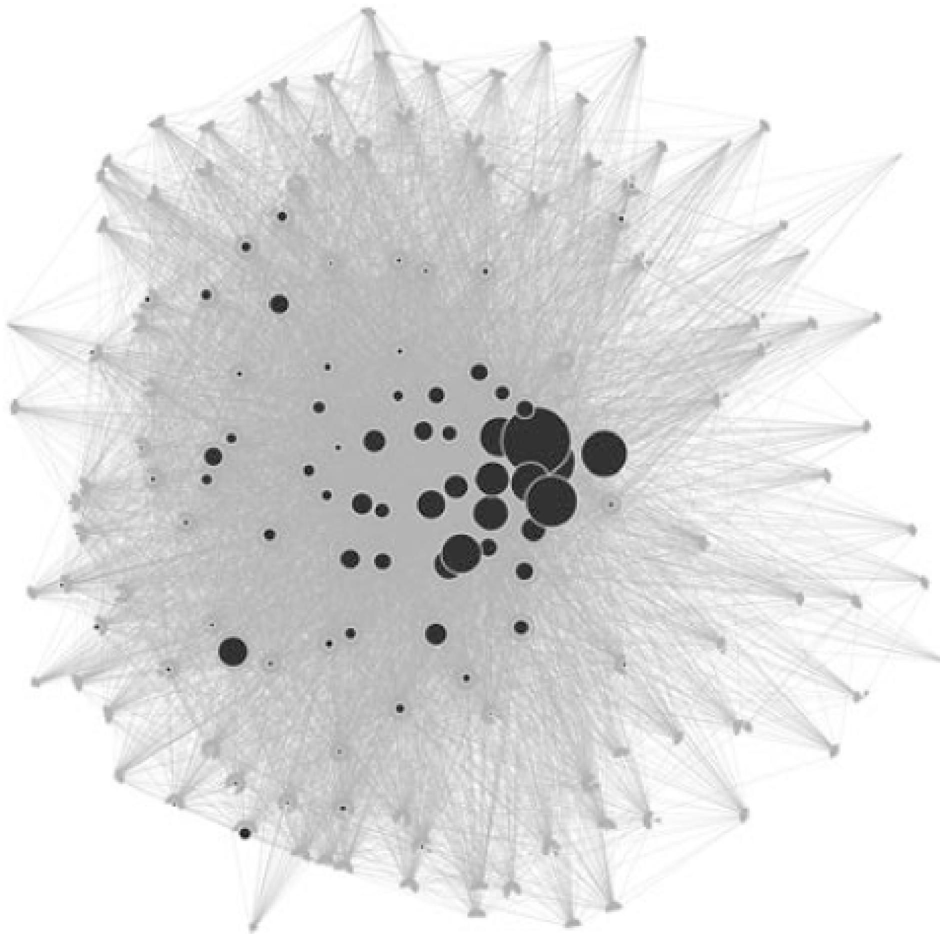
Very Dense Networks

- Useless: Betweenness + Closeness
- Useful: Degree + Eigenvector
- Option: Reduce lines in case of weighted networks





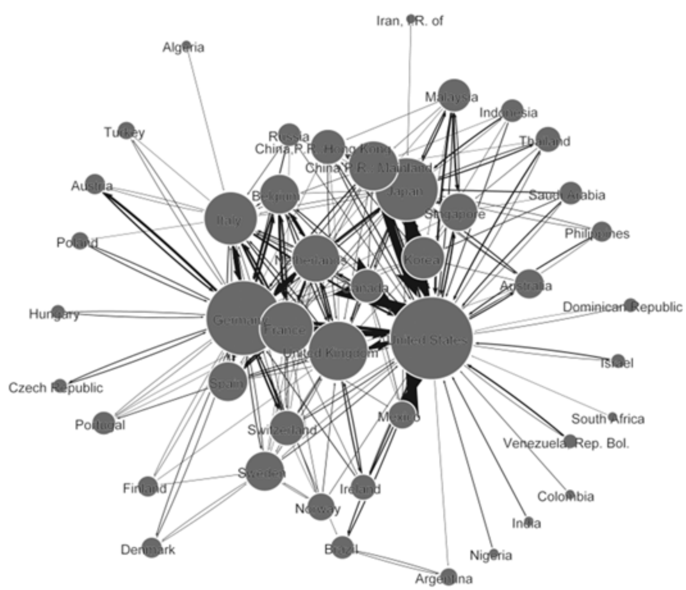
Reducing Complexity



Trade Network of 157 Countries

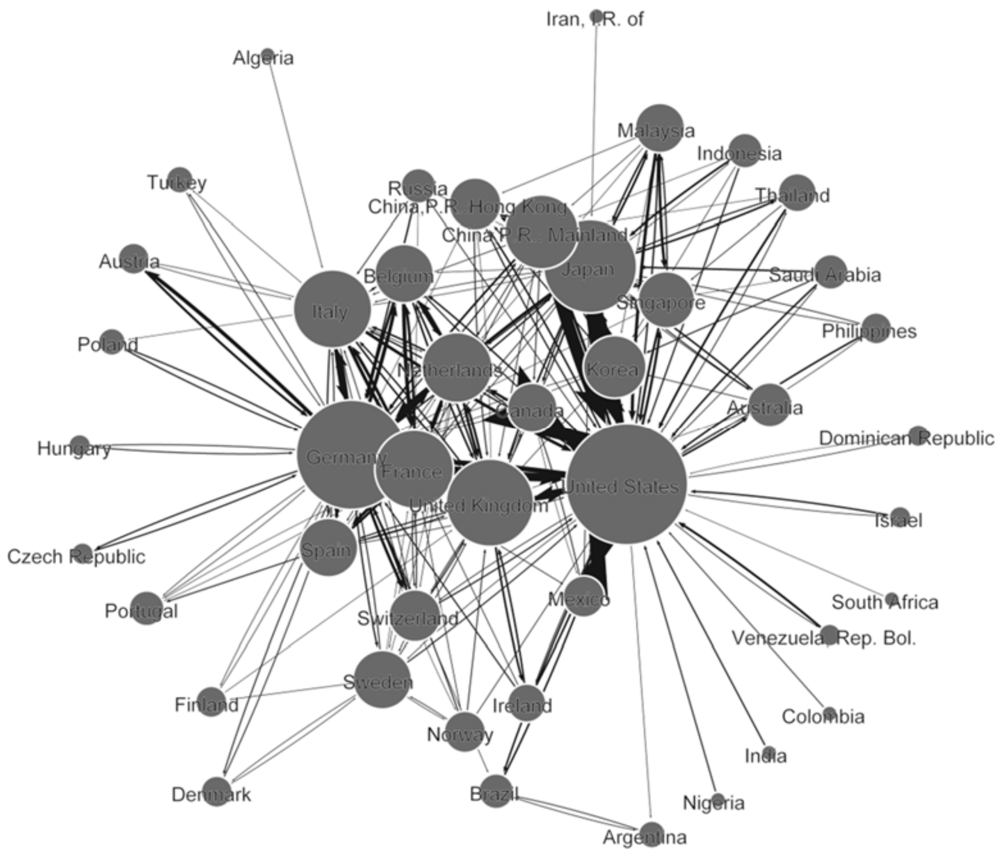
Reducing Nodes and Lines

- Weighted networks (each line is described by a value)
- A: Remove all lines lower a defined value (threshold)
  - Creates Isolates, focus on center
- B: Remove all but the most important lines for each node
  - Removes globally important lines



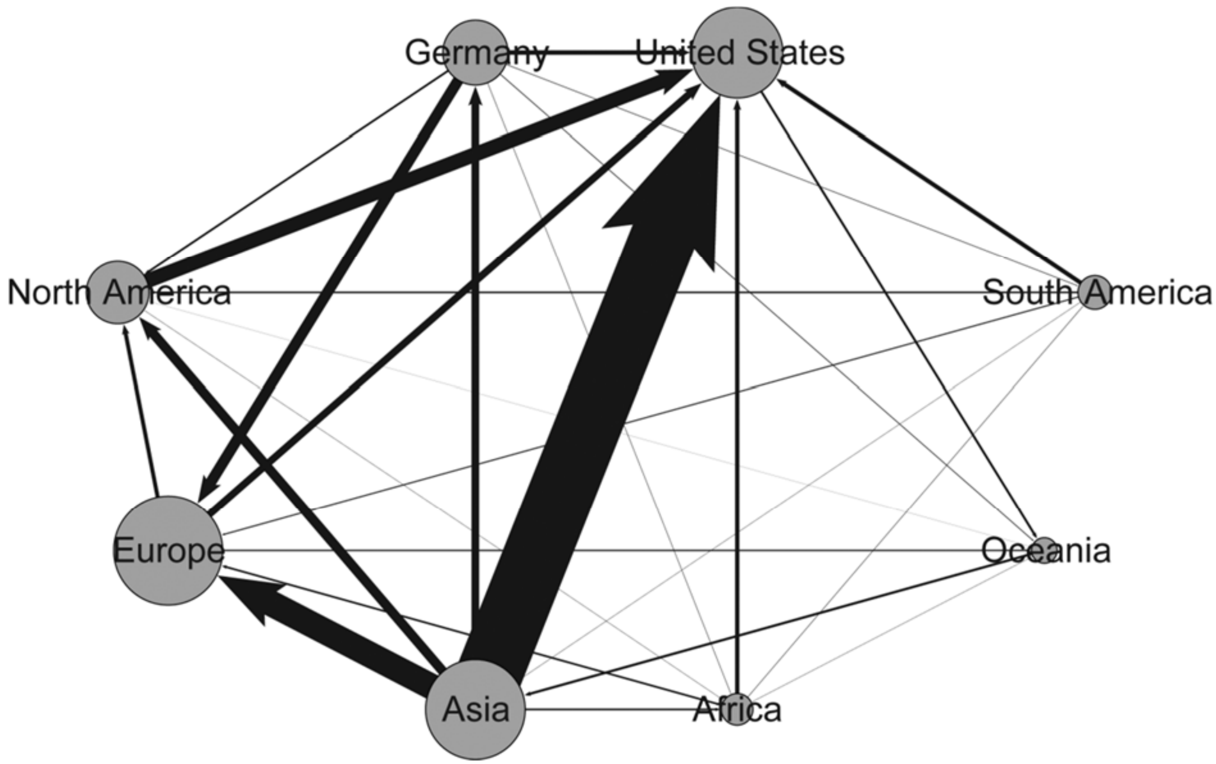
Attention

- Reducing nodes and lines changes the structure
- Try different thresholds



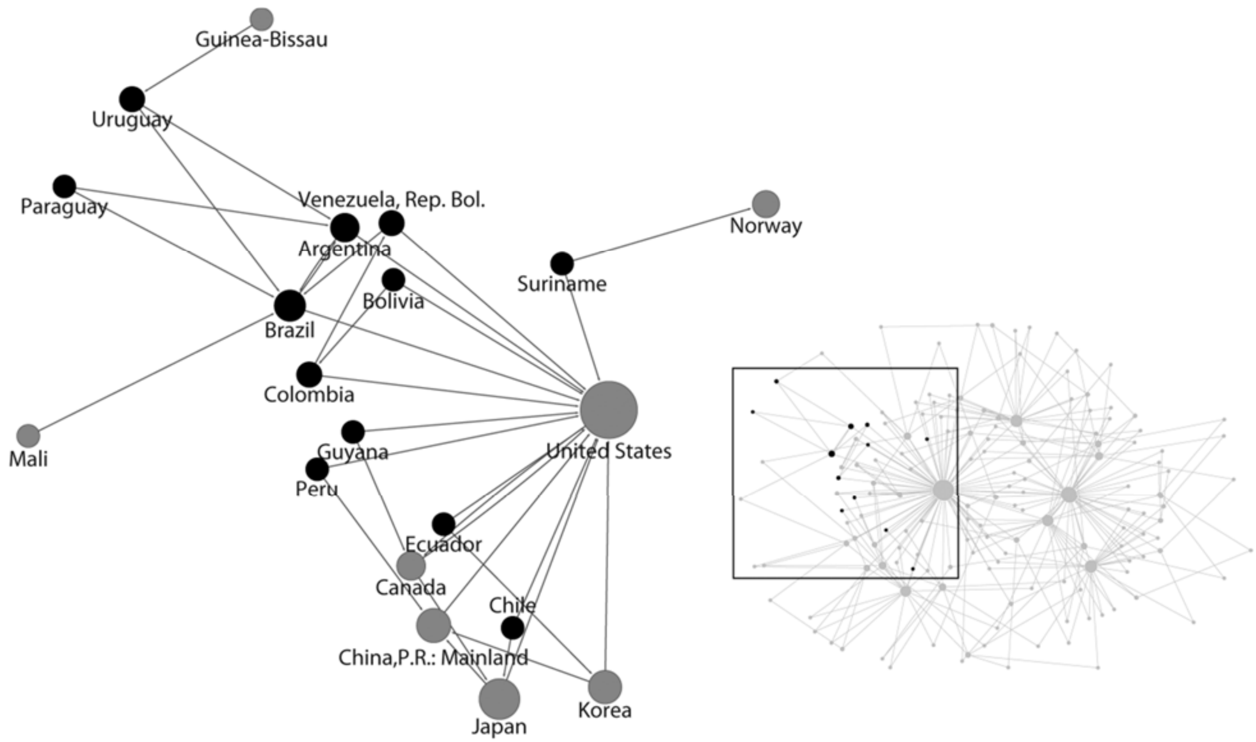
Merging Nodes

- Aggregated level representation
- Taking a group of nodes and unite them into a new node
- Connection within groups: Loops, node color



Micro/Macro Reading

- Focus on interesting zones
- Show the context
- Sub-network of the South American countries and their connections:



Printing Quality?

JPG vs. PDF

- Why are JPG pictures great for photos but not for network visualizations?
- Why are PDF drawings of network visualizations so much better?
- What does 300dpi mean?

JPG vs. PDF

PDF 400 %

Vector Graphics
- Geometrical Objects
- Scales infinitely

JPG 400 %

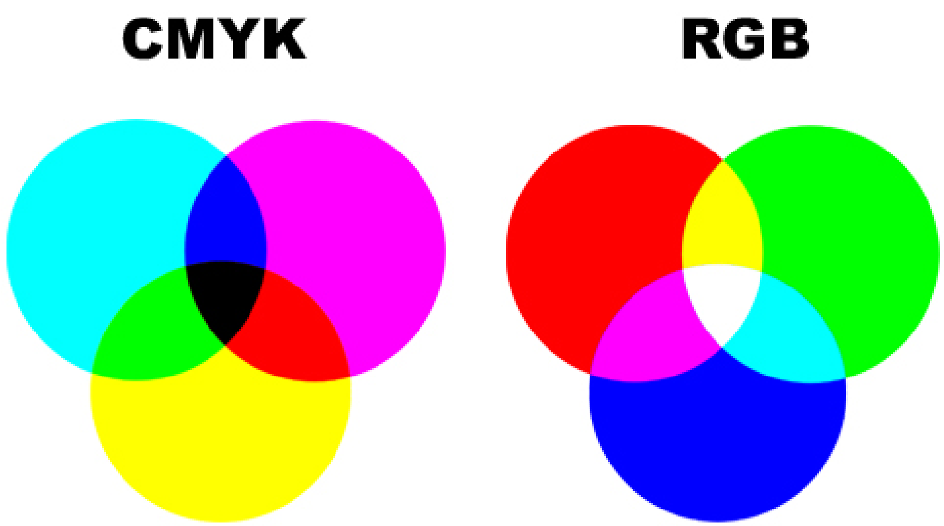
Raster Graphics
- Pixels
- Scales infinitely

Professional Drawings

- Actual printing quality for publications
  - Vector graphics: EPS/PDF
  - Dot graphics: 300 dpi, PNG is better than JPG, make large Figures
  - Screenshots have 72dpi (!), use big screen (iPad2 = 2048 x 1536 pixel)
  - E.g. 1280 x 800 $\rightarrow$ 4.3 x 2.7 inch = 10.8 x 7.8 cm
- Don’t use tools that cannot export as PDF or EPS
- Post-processing of figures:
  - Raster: Photoshop, Gimp
  - Vector: Illustrator, AutoCAD,

Difference Between CMYK and RGB

- Newton (1672): White light as the sum of spectral colors
- Helmholtz (1852): Additive and subtractive color mixing
  $\rightarrow$ 3 colors create „all“ colors: RGB, CMYK
  - Adding wavelengths of light
  - subtracting (absorbing) wavelengths of light



Summary: Check-Lists

Visualizing networks is craft rather than art

Graphical Excellence

Graphical excellence [Tufte 2001]…

- is a matter of substance, statistics, and design
- consists of complex ideas communicated with clarity, precision, and efficiency
- give the viewer the greatest number of ideas in shortest time with the least ink in the smallest space
- is nearly always multivariate
- requires telling the truth about the data
- induce the viewer to think about the substance rather than the methodology

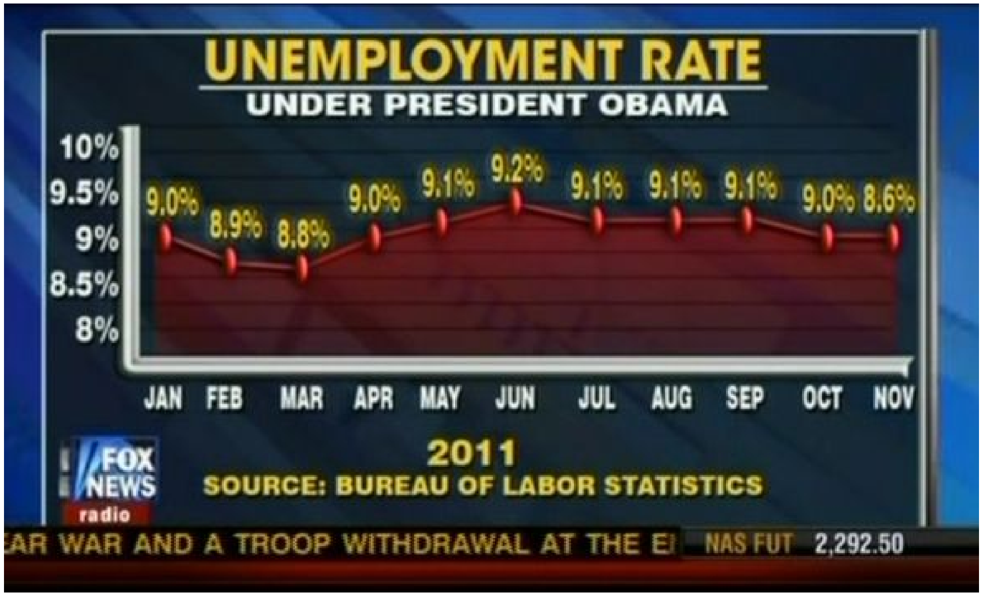
→ Above all **show the data**

Graphical Integrity

„As to the propriety and justness of representations sums of money, and time, by parts of space, tho‘ very readily agreed to by most men, yet a few seem to apprehend that there may possibly be some deceptions in it, of which they are not aware…“

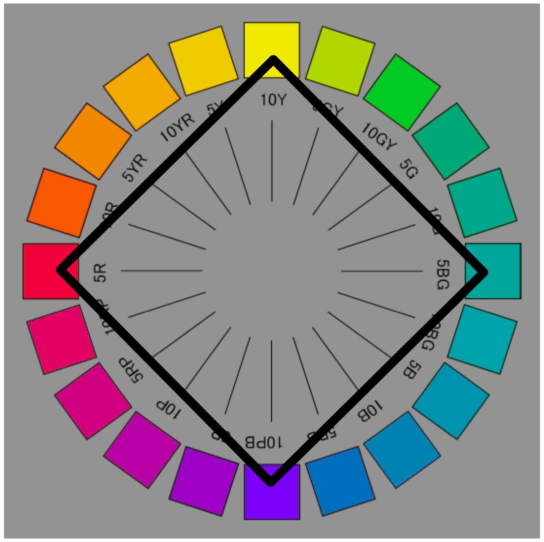
William Playfair

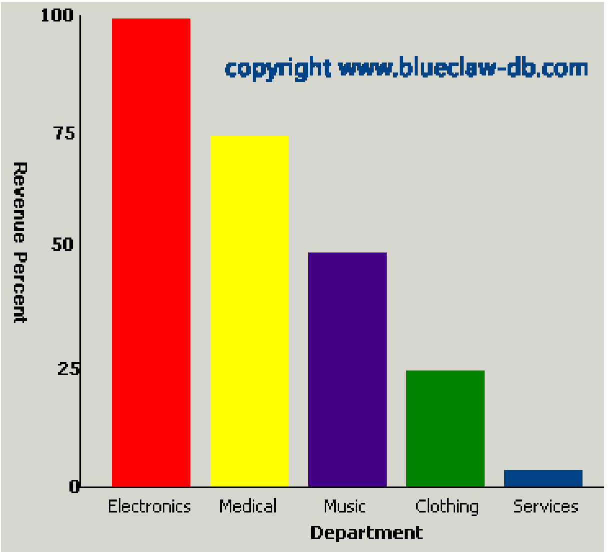
The Commercial and Political Atlas (1786) [Tufte 2001]



Smart Use of Colors

- Just use colors when they carry additional information
- Color hue and saturation are used for different data
- Colors often have meaning
- But, learn to visualize without colors



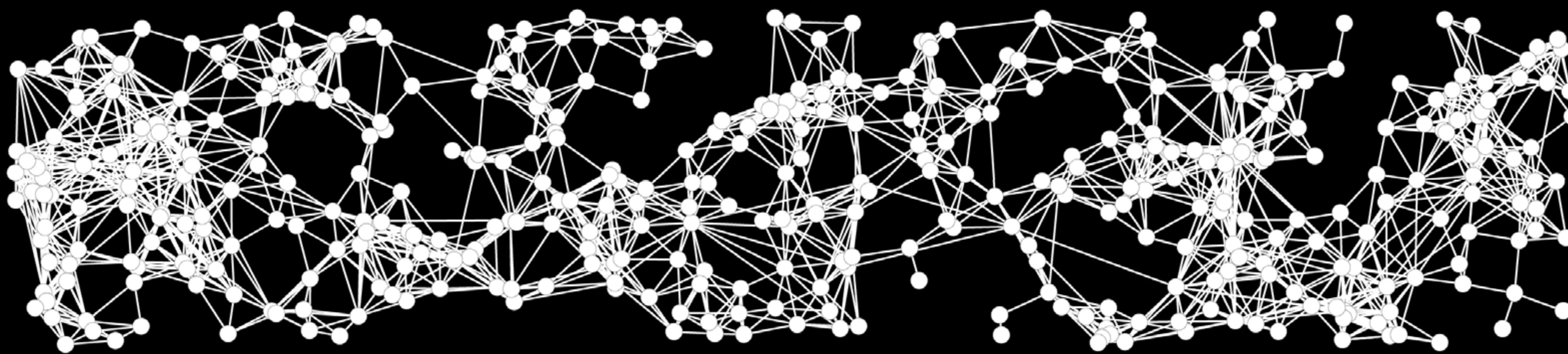


And Finally…

- Explain what you did
  - Describe mapping of data to visual elements
  - Use a legend or caption
- Be consistent across visualizations
  - Easier to memorize and recognize repeated designs
  - Find your style – a good one
- It is all about the story
  - Narrative quality of the visualization

Main References

- E. R. Tufte, Visual Display of Quantitative Information, Second Edition, 2001.
- J. Mackinlay, Automating the design of graphical presentations of relational information, in: ACM Trans. Graph. 5, 2 (Apr. 1986), 110-141, 1986.
- S.S. Stevens, Psychophysics: Introduction to Its Perceptual, Neural, and Social Prospects, Transaction Publishers, 1975.
- J. Bertin, Semiology of Graphics: Diagrams, Networks, Maps, University of Wisconsin Press, 1984.
- L. Krempel, Visualisierung komplexer Strukturen – Grundlagen der Darstellung mehrdimensionaler Netzwerke, Campus Verlag, 2005.
- M. Lodge, Magnitude scaling, quantitative measurement of opinions, Sage Publications, 1981.
- **J. Pfeffer, Jürgen. Fundamentals of Visualizing Communication Networks. IEEE China Communications 10 (3), 82-90. 2013**



jpfeffer@cs.cmu.edu
www.pfeffer.at
@JurgenPfeffer