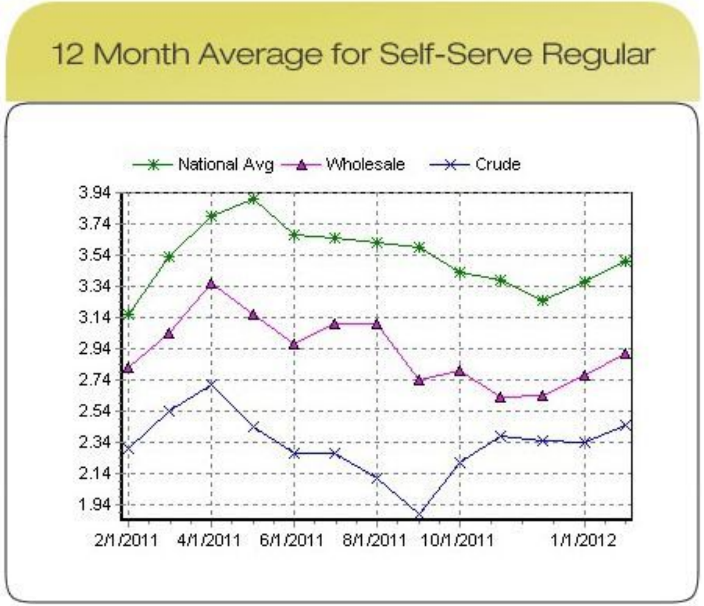
Deceptive Visualizations – Avoiding Pitfalls in Design

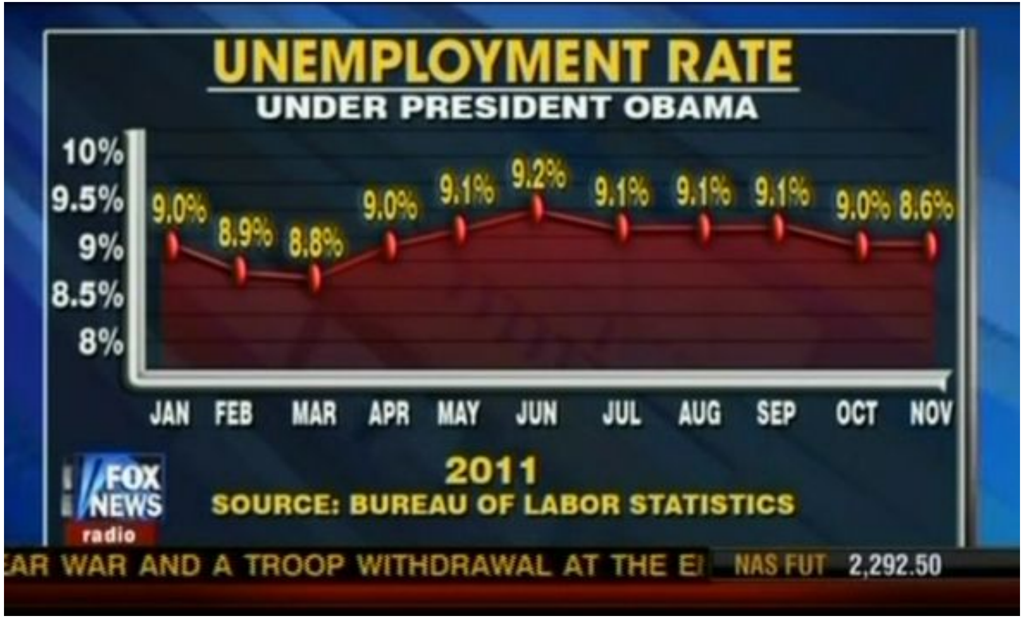
Ross Maciejewski, Arizona State University

Deceptive Visualizations



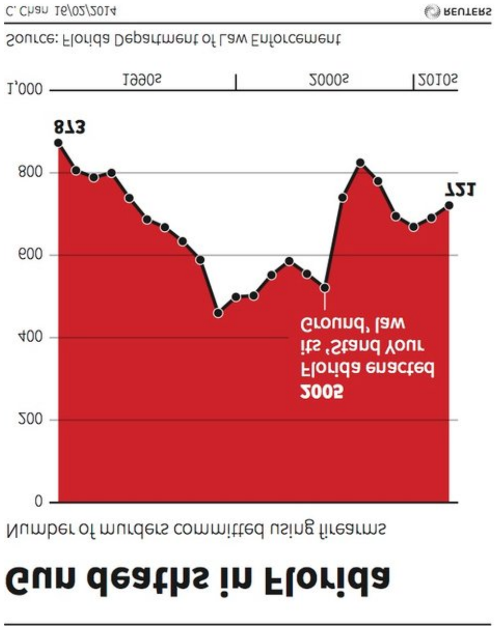
https://www.mediamatters.org/blog/2012/02/21/fox-still-struggling-with-basic-chart-concepts/185049

Deceptive Visualizations



https://www.mediamatters.org/blog/2011/12/12/today-in-dishonest-fox-news-charts/185162

Deceptive Visualizations



Data Types

Nominal
Data whose categories have no implied ordering
Examples include political affiliations of a population

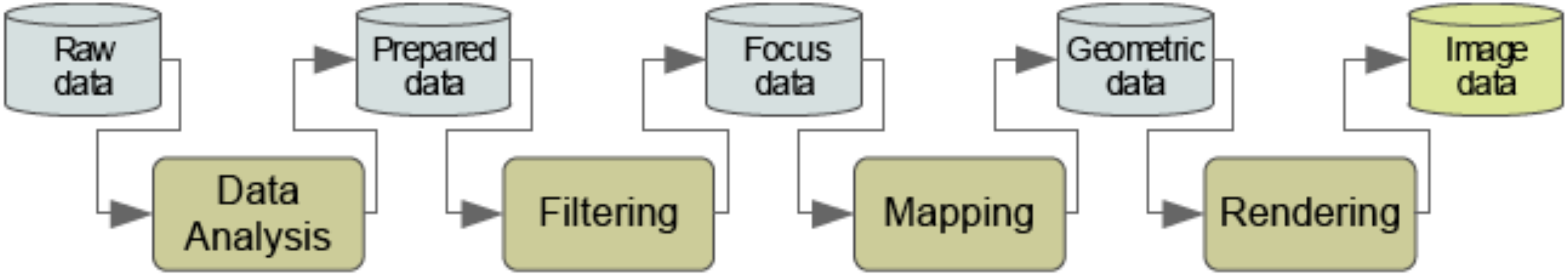
Ordinal
Data that has a specified order, but no specified distance metric
Examples include beverage sizes at McDonalds (Small, medium, large)

Interval
Data that has measurable distances
Examples include periods of time (second, minute, etc.) – the zero point is arbitrary

Ratio
Same as interval, but include a zero point
Example include Celsius scale, height above sea level

1- SS Stevens, “On the Theory of Scales of Measurement,” *Science*, 103(2684):677-680, 1946.

Visualization Pipeline



We want to take these different data types and map them to an appropriate visual representation
Data Analysis – data are prepared for visualization (smooth, interpolate, transform)
Filtering – A subset of the data (usually user defined) is selected for visualization
Mapping – Data are mapped to geometric primitives and their attributes
Rendering – Geometric data are transformed to image data

Image taken from http://www.infovis-wiki.net/index.php/Visualization_Pipeline
Haber, R. & McNabb, D. A. (1990). Visualization idioms: A conceptual model for scientific visualization systems. In Visualization in Scientific Computing.

Mapping Data

We need to know how to assign quantitative dimensions of our data to *aesthetic attributes*[1] of the data

Form	Surface	Motion	Sound	Text

Position
Size
Shape
  polygon
  glyph
  image
Rotation
Resolution
	Color
	  hue
	  brightness
	  saturation
	Texture
	  pattern
	  granularity
	  orientation
	Blur
	Transparency
		Direction
		Speed
		Acceleration
			Tone
			Volume
			Rhythm
			Voice
				Label

L Wilkinson (2005) *The Grammar of Graphics*

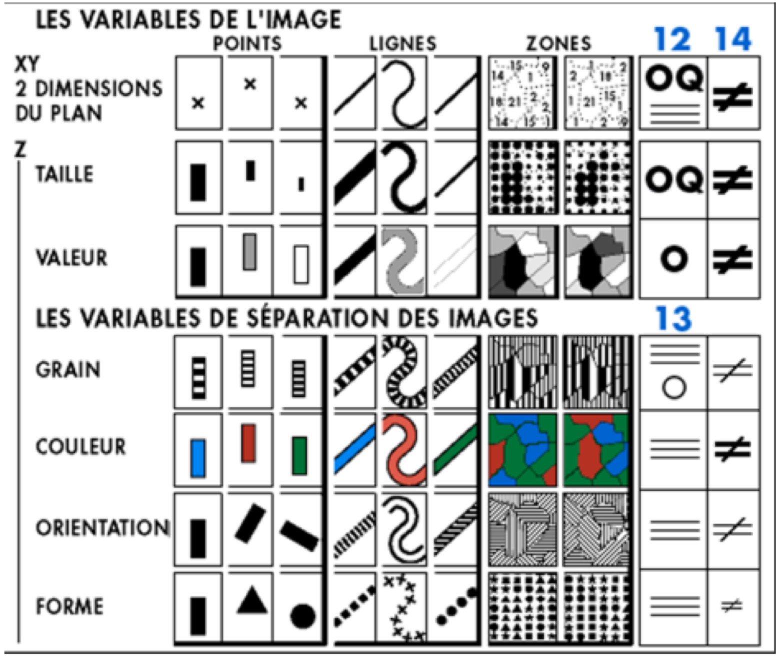
Aesthetic Attributes

- An attribute must be capable of representing both continuous and categorical variables
- When representing a continuous variable, an attribute must vary primarily on **one** psychophysical dimension
- In order to use multidimensional attributes (such as color), we must scale them on a single dimension
- An attribute does not imply a linear perceptual scale
- Much of the skill in graphic design is knowing what combination of attributes should be avoided[1]

1-SM Kosslyn (1994), *The Elements of Graph Design*

Bertin’s Visual Variables

- Visualization is concerned primarily with a mapping to visual form
- [x,y]
- Position
- [z]
- Size (Taille)
- Value (Valeur)
- Color (Couleur)
- Texture (Grain)
- Orientation
- Shape (Forme)



J Bertin (1967), *The Semiology of Graphics*

Position

Position refers to a location in a multi-dimensional space

Bertin restricts his analysis to a piece of paper (or a plane) but in computer graphics, we need not have such a restriction

Continuous variables map to densely distributed locations

Categorical variables map to a lattice

Positions are ordered, but the ordering may or may not have meaning in terms of what is being measured

Sometimes, position is just a way to keep things from overlapping

Position

Cleveland[2] rates position on a common scale as the **best way to represent a quantitative dimension visually**

This reflects research findings that points or line lengths placed adjacent to a common axis enable judgments with the least bias or error

However, this recommendation has a caveat, it depends on how far the graphic primitive (point, line, etc.) is from a reference axis[3]

If a graphic is far from an axis, the multiple steps needed to store and decode the variation can impair judgment

1 - L Wilkinson (2005) *The Grammar of Graphics*
2 – WS Cleveland, *The Elements of Graphing Data*, 1985
3 – D Simkin and R Hastie (1987). An information processing analysis of graph perception. *Journal of the American Statistical Association*, 82, 454-465

Size

Bertin[2] defines size variation in terms of length or area

For three dimensions we have volume

Cleveland[3] ranks area and volume representations among the **worst attributes** to use for graphing data

Some designers assign size to only one dimension of an object

Think of the bar chart where the width of the bar is typically constant, but the height is varied

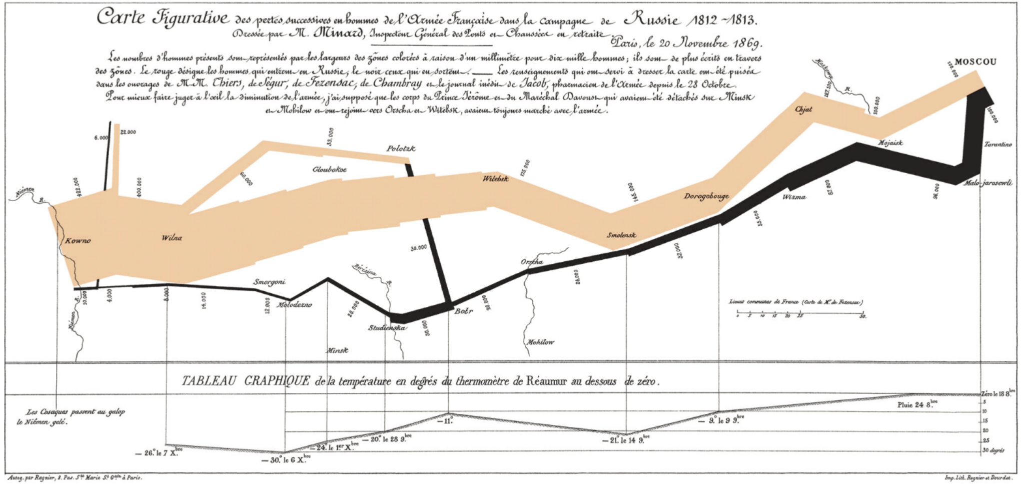
1 - L Wilkinson (2005) *The Grammar of Graphics*
2 – J Bertin (1967), The Semiology of Graphics
3 – WS Cleveland, *The Elements of Graphing Data*, 1985

Size

Size for lines is usually equivalent to thickness

This is less likely to induce perceptual distortion

Size can be used to great effect with path

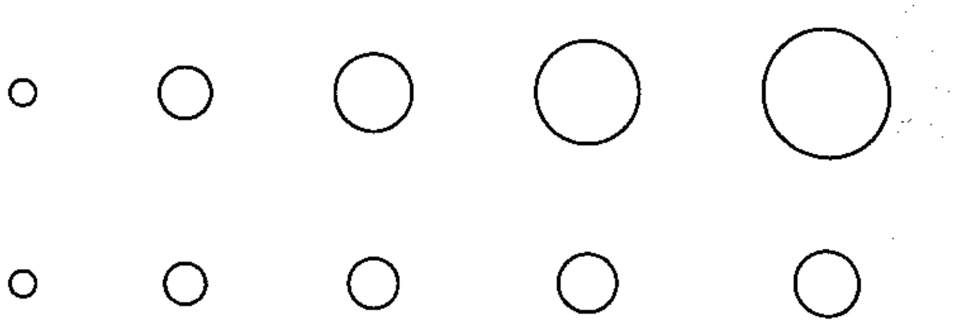


1 – Charles Joseph Minard: Mapping Napoleon’s March, 1861 by John Corbett, Center for Spatially Integrated Social Science

Size

For objects with rotational symmetry, we can map size to the diameter rather than area

Representing data through area or volume should probably be confined to positively skewed data that can benefit from the perceptual equivalent of the square root transformation



Top row changes diameter from 1-5
Bottom row changes area from 1-5

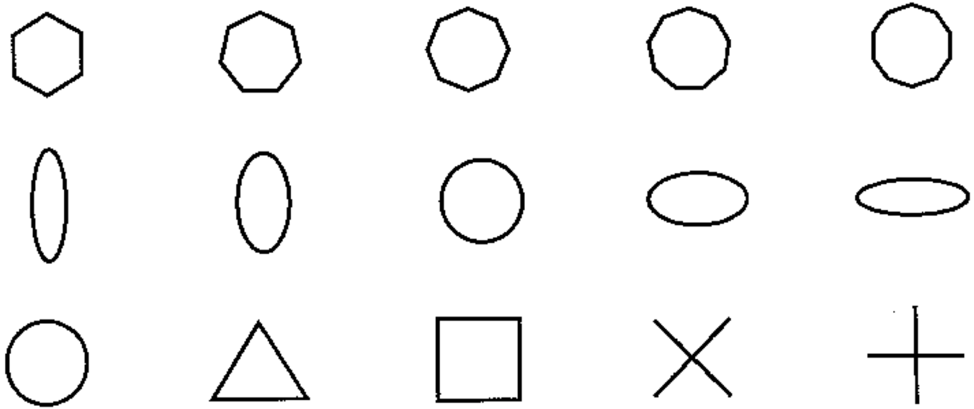
L Wilkinson (2005) *The Grammar of Graphics*

Shape

Shape refers to the shape or boundary of an object

Examples would include map symbols

Shape must vary without affecting size, rotation and other attributes

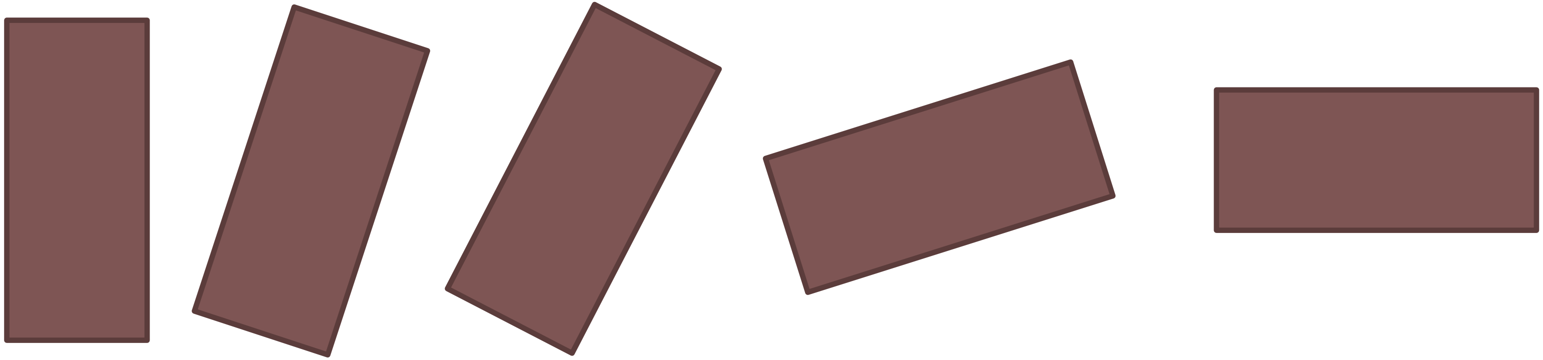


L Wilkinson (2005) *The Grammar of Graphics*

Rotation

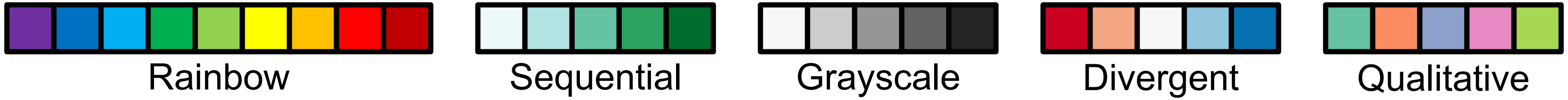
This is the rotational angle of the graphic primitive

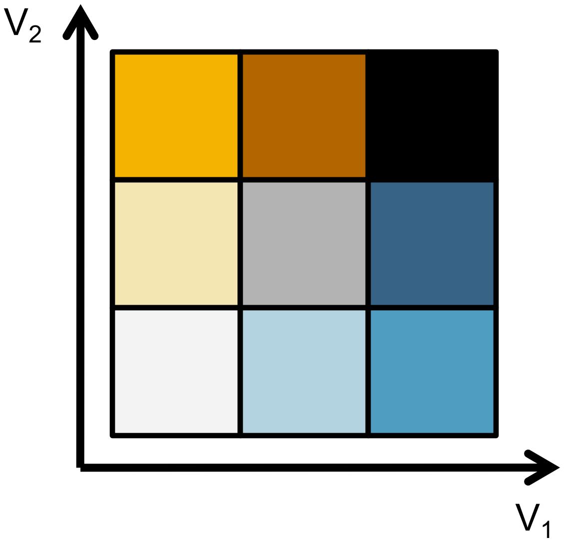
Lines, areas and surfaces can only rotate if they are positionally unconstrained

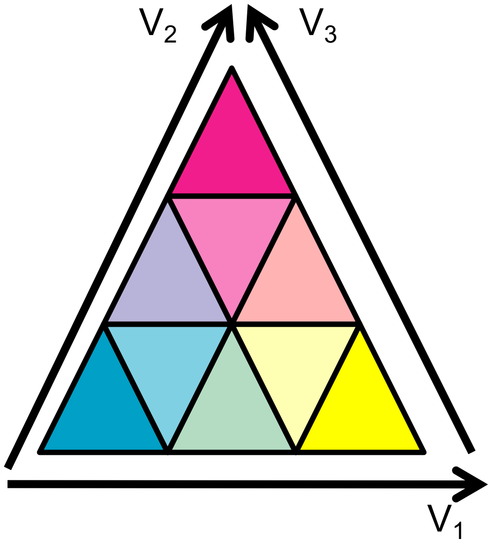


L Wilkinson (2005) *The Grammar of Graphics*

Color


Rainbow
Sequential
Grayscale
Divergent
Qualitative


V2
V1


V2
V3
V1

Texture

Texture alone can be a basis for perception
Two gray areas that have the same overall level of brightness can be discriminated if their texture is different

L Wilkinson (2005) *The Grammar of Graphics*

Texture

Granularity – Changing the size or resolution of patterns in a texture changes the granularity

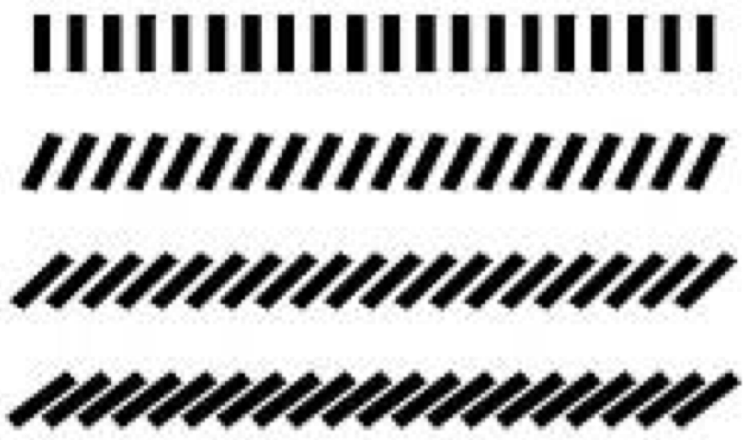
Less grainy patterns (those with low-frequency spatial components) are more difficult to resolve

Pattern – Patterns make use of increasing degrees of randomness to encode data

L Wilkinson (2005) *The Grammar of Graphics*

Texture

Orientation – Variation in texture orientation can introduce visual illusions (making lines not seem parallel) and is typically avoided as an encoding in textures

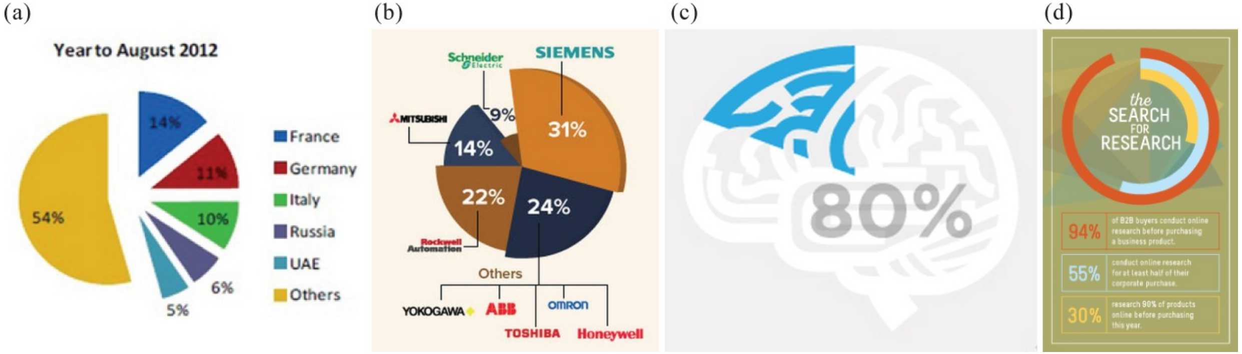


L Wilkinson (2005) *The Grammar of Graphics*

Pie Charts

- Some perceptual theory suggests that pie charts are not a good way of presenting information[1]

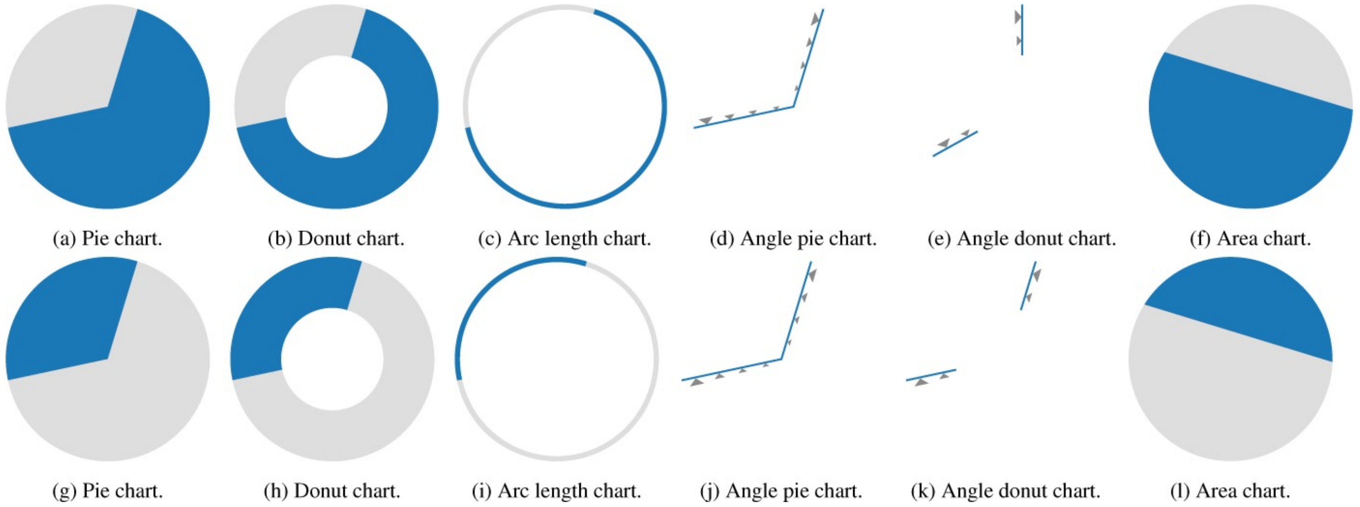
- Often pie charts are presented with the values as part of or close to the pie slice labels (indicating that values may be hard to decode)



Cleveland W.S., McGill R.: Graphical perception: Theory, experimentation, and application to the development of graphical methods. Journal of the American Statistical Association 79, 387 (1984), 531–554

Pie and Donut Charts

- Three different visual cues are present in pie charts:
  - Angle
  - Arc Length
  - Area



D. Skau, R. Kosara. Arcs, Angles, or Areas: Individual Data Encodings in Pie and Donut Charts. Computer Graphics Forum, 2016.

Pie and Donut Charts

- *Donuts are fine*. No adverse effect from removing the pie center.

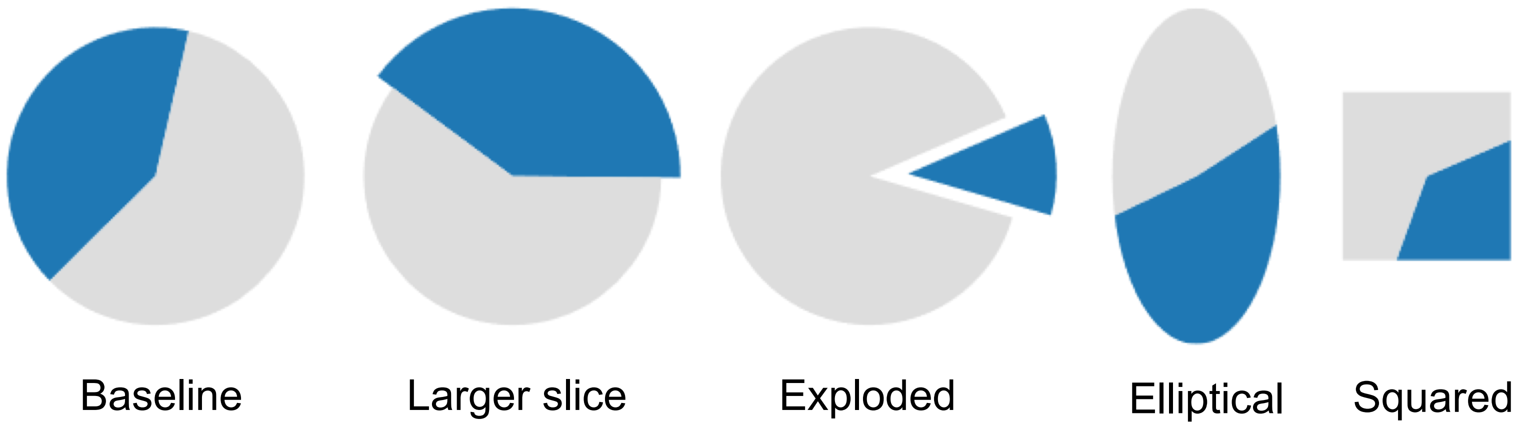
- *Nested donuts are problematic*. Since area and arc length are important, nesting donuts means comparing circles of different radius and area, which is likely problematic.

- *Keep cues consistent*. Since more cues seem to make for better judgment, providing conflicting cues is counterproductive.

D. Skau, R. Kosara. Arcs, Angles, or Areas: Individual Data Encodings in Pie and Donut Charts. Computer Graphics Forum, 2016.

Judgement Error in Pie Charts

How do variations in pie charts perform?



R. Kosara and D. Skau. Judgment Error in Pie Chart Variations. Eurographics Conference on Visualization – Short Paper, 2016.

Judgement Error in Pie Charts

- The larger slice leads to systematic overestimation over almost the entire range of values.

- The exploded pie chart shows higher error. Perhaps the gap between the two slices adds a level of distraction that causes higher error.

- The ellipse yields much higher error than the circle.

- The square produces larger error, and the ellipse actually leads to more error than the square.

R. Kosara and D. Skau. Judgment Error in Pie Chart Variations. Eurographics Conference on Visualization – Short Paper, 2016.

When to Use a Pie Chart?

- Useful for displaying data classified into nominal or ordinal categories, however, too many slices can make these difficult to read (7-9 slices may be a good rule of thumb)

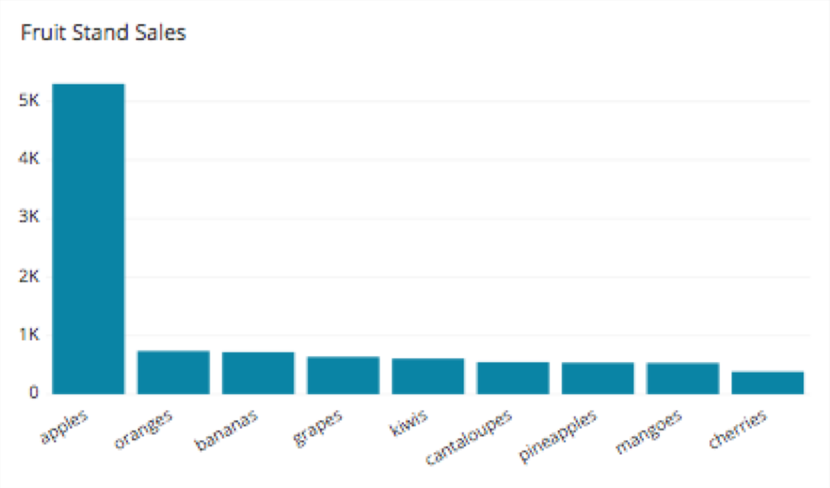
- Generally used to show percentage or proportional data

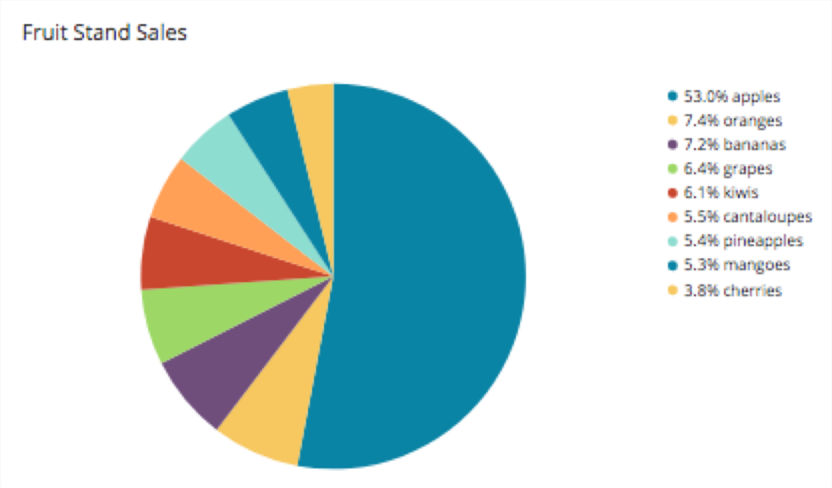
- Zero and negative values cannot be well represented

- Temporal data is generally not well associated with pie charts

When to Use a Pie Chart?

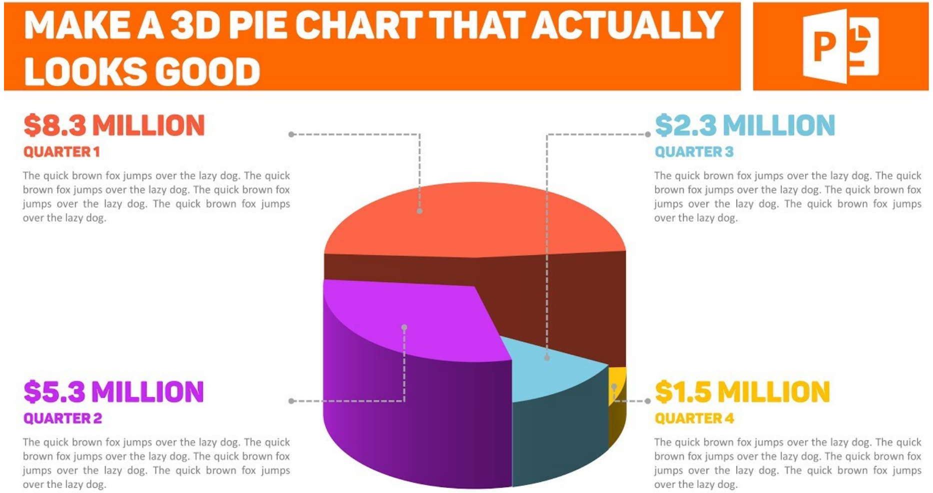
- Often bar charts may be better for accuracy, but many people appreciate the circular aesthetic of a Pie

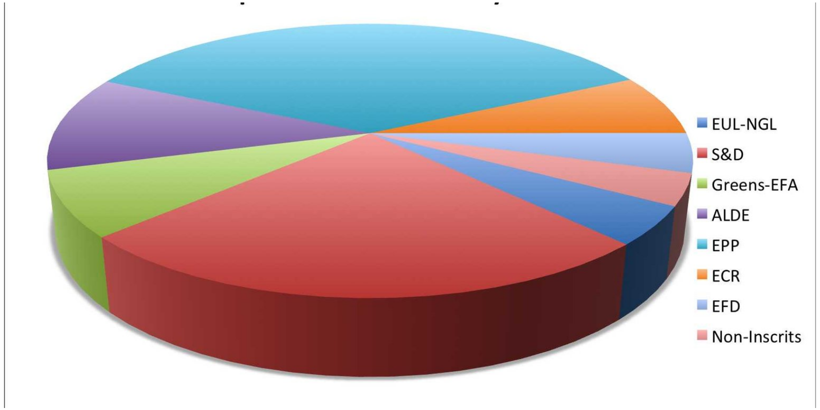




3D and Pie Charts

Please, do NOT create 3D Pie Charts



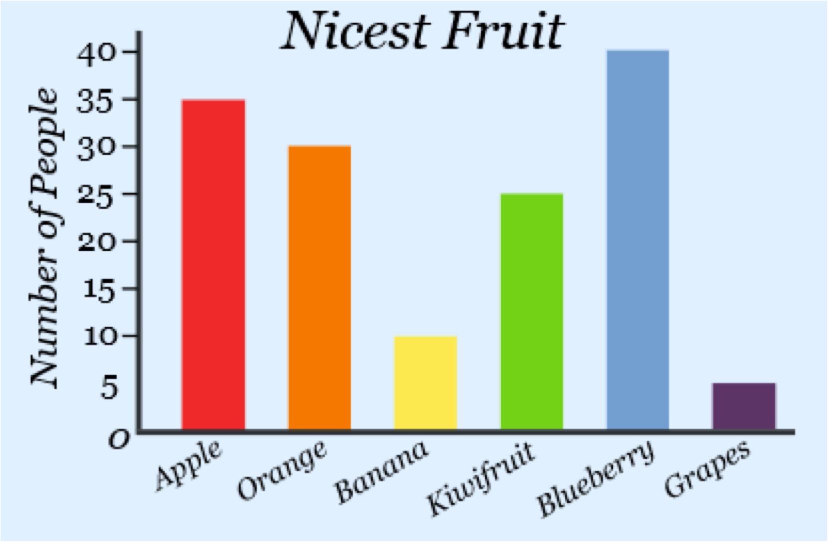


Bar Charts

- Bar charts provide a useful alternative of presenting a set of values associated with the levels of a factor

- Can be used for values other than proportions or comparing between different groups

- Use position on a common scale to encode their values

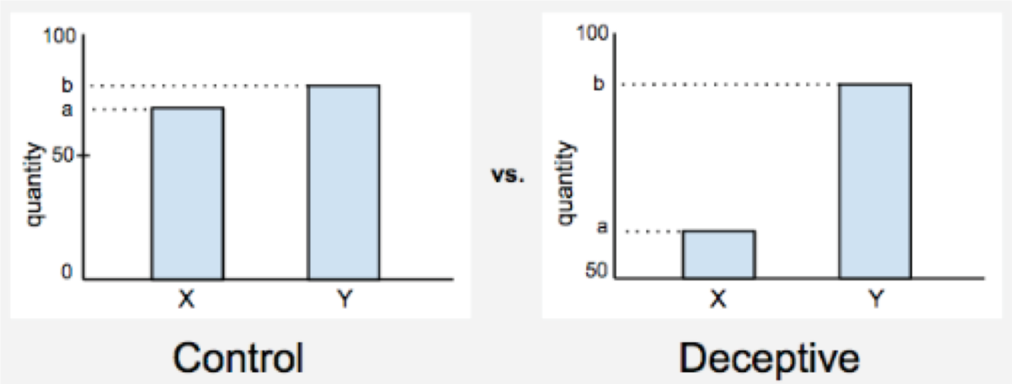


https://www.mathsisfun.com/data/bar-graphs.html

Deceptions in Bar Charts

Truncated Axis distortion – leads to message exaggeration or understatement

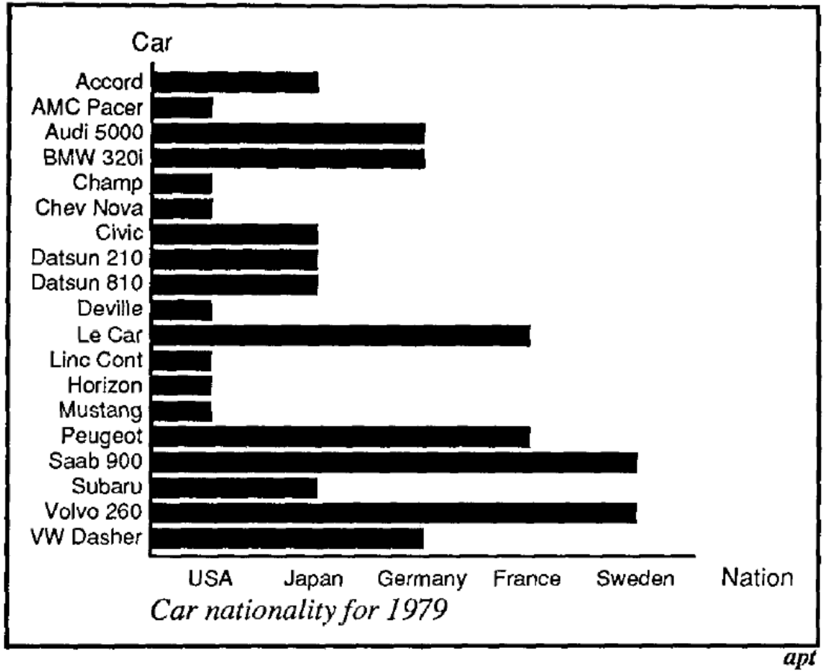
- Impacts “how much” type of questions



A. Pandey, K. Rall, M. L. Satterthwaite, O. Nov, E. Bertini, “How Deceptive are Deceptive Visualizations?: An Empirical Analysis of Common Distortion Techniques,” *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems, 2015.*

Deceptions in Bar Charts

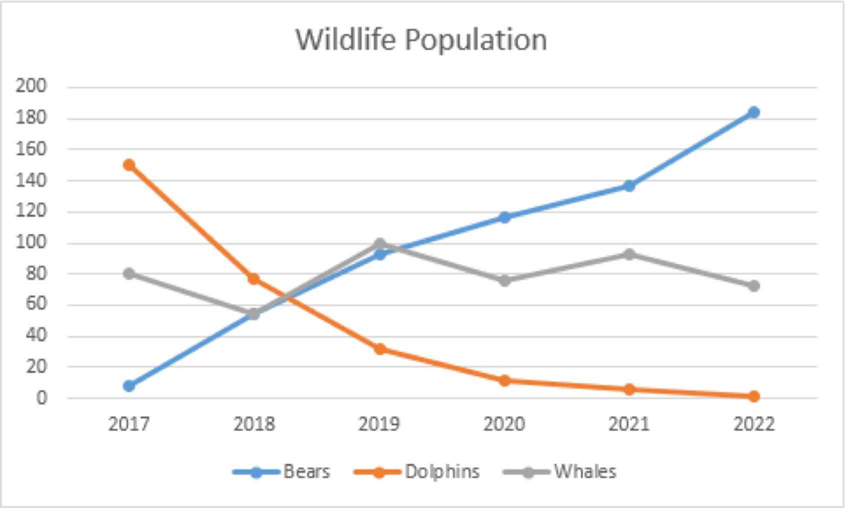
Encoding nominal variables on two axes can misrepresent data



J. Mackinlay, Automating the Design of Graphical Presentations of Relational Information, ACM Transactions on Graphics, 5(2): 110-141, 1986.

Line Charts

- Often used to compare changes over time
- Used in identifying trends, patterns and anomalies
- Typically time is on the x-axis, and a measure on the y-axis
- Can overwhelm if too many time series are shown

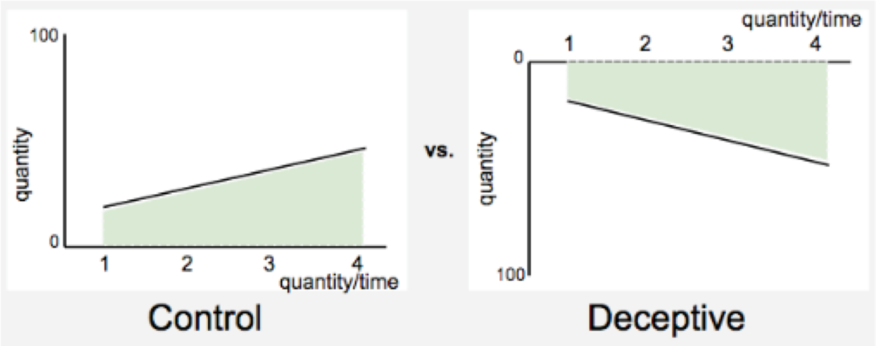


https://www.excel-easy.com/examples/line-chart.html

Deceptions in Line Charts

Inverted Axis – humans tend to relate directions with trends: upwards – increase, downwards – decrease, right – front, left – back

- Impacts "what" questions – what can you say about the trend?

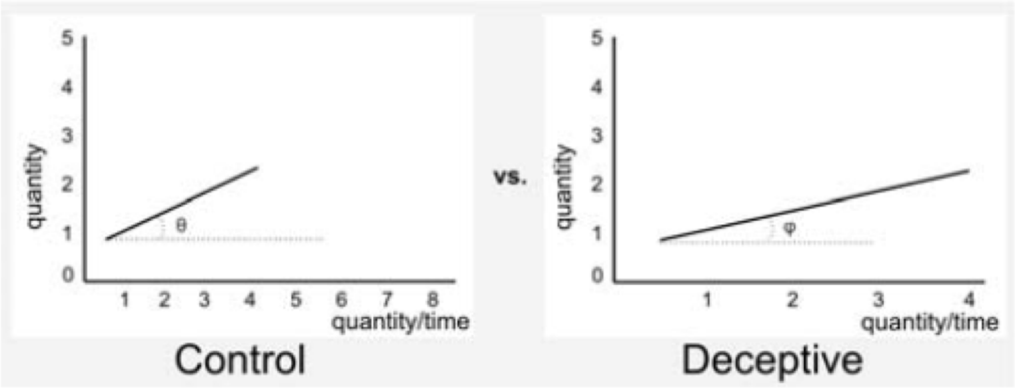


A. Pandey, K. Rall, M. L. Satterthwaite, O. Nov, E. Bertini, "How Deceptive are Deceptive Visualizations?: An Empirical Analysis of Common Distortion Techniques," *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems, 2015.*

Deceptions in Line Charts

Aspect Ratio– Primarily affects line charts making the rate of increase/decrease appear to change

- Impacts “how much” questions – How much bigger is A than B?



A. Pandey, K. Rall, M. L. Satterthwaite, O. Nov, E. Bertini, “How Deceptive are Deceptive Visualizations?: An Empirical Analysis of Common Distortion Techniques,” *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems, 2015.*

Aspect Ratio Guidelines

Our ability to perceive trends and patterns in a given display is heavily influenced by the aspect ratio

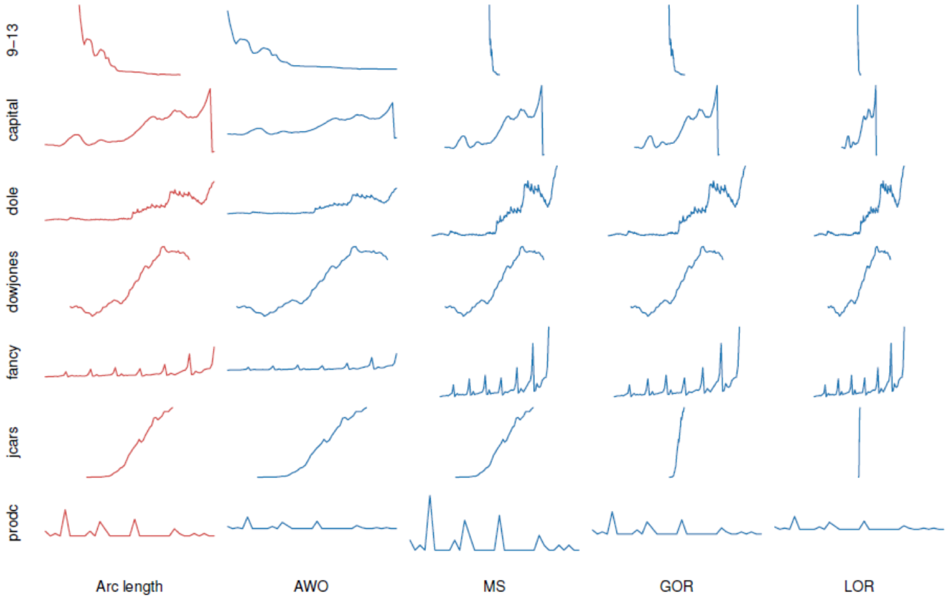
Aspect ratio affects densities, relative distances and orientations

Several methods have been proposed for automatically selecting the aspect ratio

Aspect ratio: a = pixel width of x axis / pixel width of y axis

Jeffrey Heer, Maneesh Agrawala. Multi-Scale Banking to 45 Degrees. IEEE Trans. Visualization & Comp. Graphics (Proc. InfoVis), 12(5), 701–708, 2006.

Aspect Selection Methods



J. Talbot, J. Gerth, P. Hanrahan, “Arc Length-based Aspect Ratio Selection,” IEEE Transactions on Visualization and Computer Graphics 17(12): 2276-2282, 2011

Jeffrey Heer, Maneesh Agrawala. Multi-Scale Banking to 45 Degrees. IEEE Trans. Visualization & Comp. Graphics (Proc. InfoVis), 12(5), 701–708, 2006.

R Packages can help with this
https://www.rdocumentation.org/packages/ggthemes/versions/3.5.0/topics/bank_slopes

Other Chart Components

- Axes and legends can often be as important as the data themselves
- Poor axis choices and label choices can lead to confusion
- Axis tick labels provide cognitive context for most basic plot types
- They support estimation and contribute to the overall appearance of the graphic
- Cleveland suggests choosing the scales so that the data rectangle fills up as much of the scale-line rectangle as possible

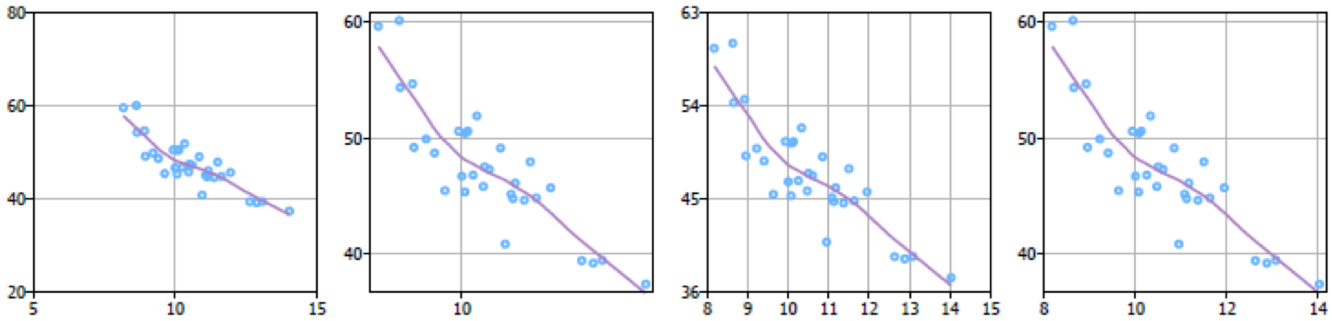
W. S. Cleveland. The Elements of Graphing Data. Wadsworth Publ. Co., Belmont, CA, USA, 1985.

Other Chart Components

- It is desirable to label the x and y axes with “nice” number
- Nice number are simple decimal numbers
- What numbers would you pick if the data range was from 105 to 543?
- What is the data range is 2.03 to 2.17
- Humans are good at picking “nice” numbers, but simplistic algorithms are not
- A primary observation is that the “nicest” numbers in decimal are 1, 2 and 5

P. Heckbert. Nice numbers for graph labels. In A. Glassner, editor, Graphics Gems, pages 61–63 657–659. Academic Press, Boston, 1990.

Tic Labeling Methods



(a) Heckbert (b) R's pretty (c) Wilkinson (d) Extended

1 – J. Talbot, S. Lin, P. Hanrahan, "**An Extension of Wilkinson's Algorithm for Positioning Tick Labels on Axes**," IEEE Transactions on Visualization and Computer Graphics 16(6): 1036-1043, 2010

2 - L. Wilkinson. The Grammar of Graphics (Statistics and Computing). Springer-Verlag New York, Inc., Secaucus, NJ, USA, 2005.

R Packages can help with this -
https://www.rdocumentation.org/packages/base/versions/3.5.1/topics/pretty

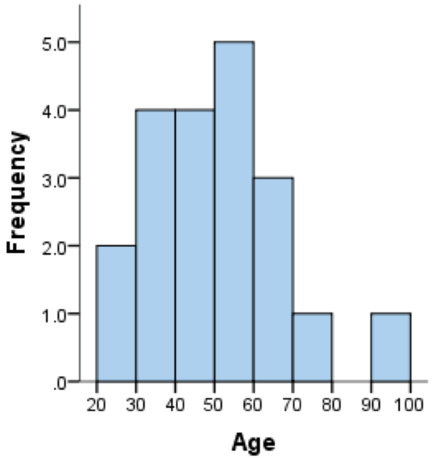
Histograms

Typical first look visualization method

- Shows the shape of the data distribution

The choice of the histogram bin width greatly impacts the resultant visualization

There is no “best” number of bins, instead, different bin sizes can reveal different features of the data



36	25	38	46	55	68	72	55	36	38
67	45	22	48	91	46	52	61	58	55

https://statistics.laerd.com/statistical-guides/understanding-histograms.php

Histograms

Number of bins (k) can be user specified or chosen from a suggested bin width (h) such that

$$k = \left\lceil \frac{\max x - \min x}{h} \right\rceil$$

Common choices for k include the square-root choice where $k = \sqrt{N}$

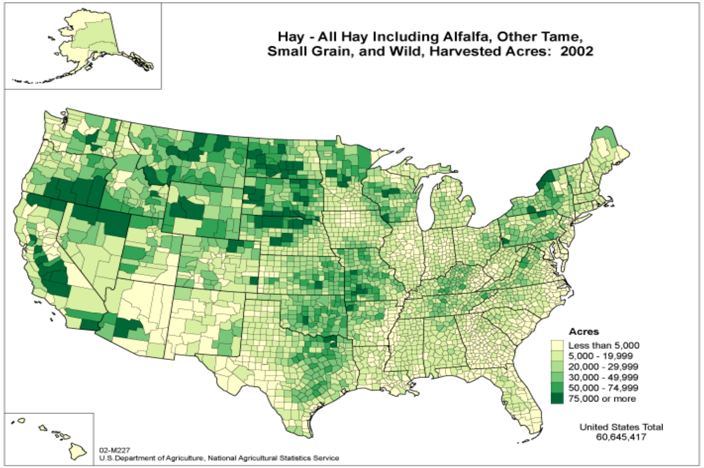
1 – H. A. Sturges. The choice of a class interval. *Journal of the American Statistical Association, p. 65-66, 1926*
2 - D. W. Scott. On optimal and data-based histograms. *Biometrika, 66(3):605-610, 1979.*
3 - Freedman, David; Diaconis, Persi, "On the histogram as a density estimator: L_2 theory". *Probability Theory and Related Fields* (Heidelberg: Springer Berlin) **57** (4): 453–476, 1981.

Choropleth Maps

Areas of the map are shaded in proportion to a measured variable

Coloring is based on a classification (histogram binning) of the distribution of the measured variable

http://www.nass.usda.gov



J. K. Wright. *Problems in Population Mapping*. Notes on Statistical Mapping, with Reference to the Mapping of Population Phenomena. AGS and Population Association of America, p. 1-18, 1938.

Map Classification

The visual representation of the choropleth map is highly influenced by the class interval selection

Popular choices for class interval selection include

- Equal interval selection
- Jenks' Natural Breaks[1]
- Minimum boundary error[2]

These choices for optimizing the class interval selection are highly dependent on the underlying data distribution

1– G. F. Jenks. The data model concept in statistical mapping. *International Yearbook of Cartography, 26:186-190, 1967.*

2 – E. K. Cromley and R. G. Cromley. An analysis of alternative classifications schemes for medical atlas mapping. *European Journal of Cancer. Series B (Methodological*), *26(2):211-252, 1964.*

Map Classification

The visual representation of the choropleth map is highly influenced by the class interval selection

Popular choices for class interval selection include

- Equal interval selection
- Jenks' Natural Breaks[1]
- Minimum boundary error[2]

These choices for optimizing the class interval selection are highly dependent on the underlying data distribution

1– G. F. Jenks. The data model concept in statistical mapping. *International Yearbook of Cartography, 26:186-190, 1967.*

2 – E. K. Cromley and R. G. Cromley. An analysis of alternative classifications schemes for medical atlas mapping. *European Journal of Cancer. Series B (Methodological*), *26(2):211-252, 1964.*

Map Classification

Classifies (bins) data such that each class occupies an equal interval along the number line

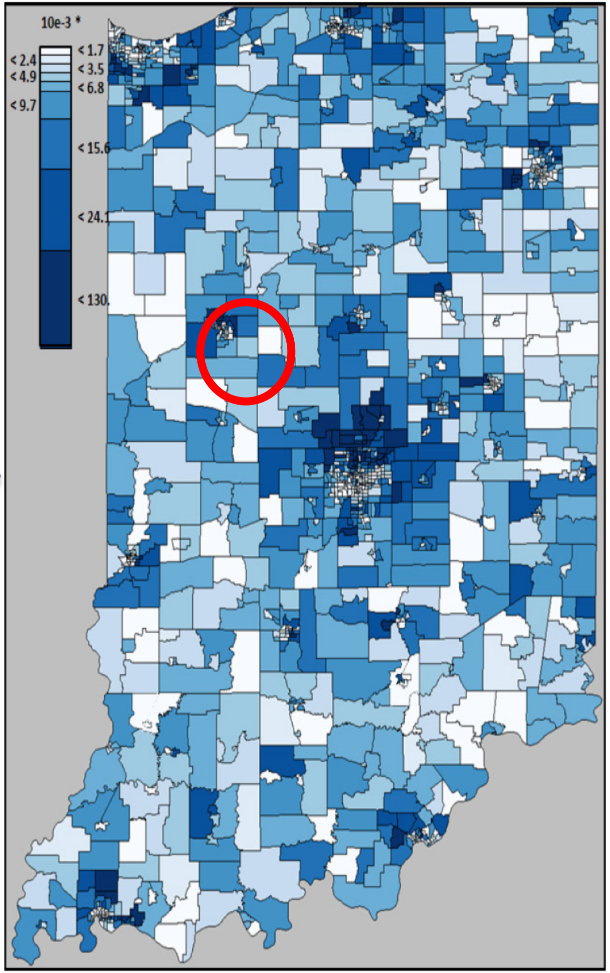
$$\frac{range}{NumClasses} = \frac{High - Low}{NumClasses}$$

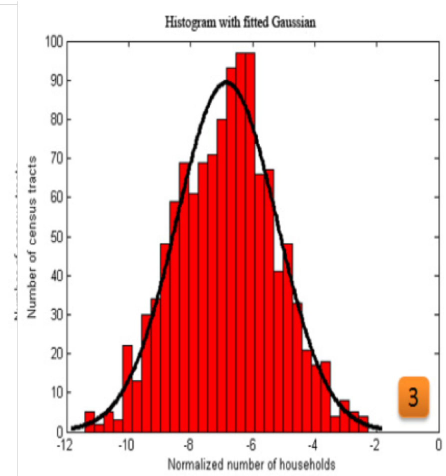
So all we do is find the maximum and minimum values of our data and divide our data set such that each color bin holds the same amount of data

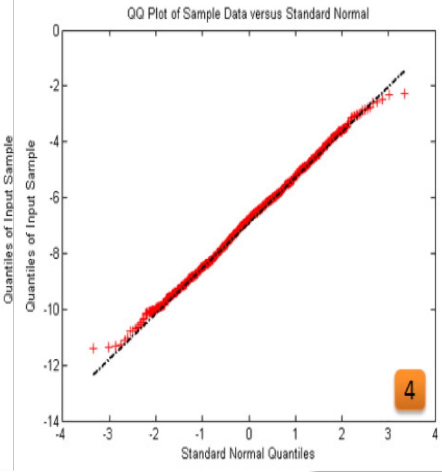
Advantage – easy to compute

Disadvantage - this fails to consider how data are distributed

TA Slocum, RB McMaster, FC Kessler and HH Howard, "Thematic Cartography and Geovisualization," Prentice Hall Series, 2009.







R. Maciejewski, A. Pattath, S. Ko, R. Hafen, W. S. Cleveland and D. S. Ebert, "Automated Box-Cox Transformations for Improved Visual Encoding," in *IEEE Transactions on Visualization and Computer Graphics*, vol. 19, no. 1, pp. 130-140, Jan. 2013.

Map Shading

- The number of colors depends on the number of classes (bins)
- Too many classes can overwhelm the user and distract them from seeing trends
- Too many classes can compromise legibility as colors become difficult to distinguish
- Typical cartographic rule of thumb is 5-7 classes
- Typical coloring schemes[1] include sequential, divergent and qualitative
- For more details on mapping as a visual representation, see [2]

1 – C. A. Brewer. http://colorbrewer.org
2 – A. MacEachren. *How Maps Work*. Guilford Press, 1995.

Picking Colors
Maps and Beyond

Sequential Color Schemes

- Suited for ordered data
- Lightness steps dominate the look of the scheme
- Light values are low data values, dark are high
- Good for Ordinal, interval and ratio data types

Diverging Color Schemes

- Puts an emphasis on critical midrange values
- Color change represents deviation from a meaningful midrange critical value
- Good for ratio data types where looking at data above and below a ‘zero’ point

Qualitative Color Schemes

- Does not imply magnitude difference
- Used to show differences between classes
- Good for Nominal data types

C. A. Brewer. http://colorbrewer.org

Handy Color References

- D. Borland and R. M. Taylor II, "Rainbow Color Map (Still) Considered Harmful," in *IEEE Computer Graphics and Applications*, vol. 27, no. 2, pp. 14-17, March-April 2007.
- D. A. Szafir, “Modeling color difference for visualization design,” in *IEEE transactions on visualization and computer graphics,* 24 (1), 392-401
- L. Bartram, A. Patra, M. Stone, “Affective Color in Visualization,” in *Proceedings of the 2017 CHI Conference on Human Factors in Computing Systems*, 1364-1374, 2019
- S. Lin, J. Fortuna, C. Kulkarni, M. Stone, J. Heer, “Selecting Semantically-Resonant Colors for Data Visualization,” in *Computer Graphics Forum*, 32:401-410, 2013
- M. Stone, D. A. Szafir, V. Setlur, “An engineering model for color difference as a function of size,” *Color and Imaging Conference*, 253-258, 2014
- M. Harrower, C. A. Brewer, “ColorBrewer.org: An online tool for selecting colour schemes for maps,” *The Cartographic Journal*, 40(1):27-37, 2003

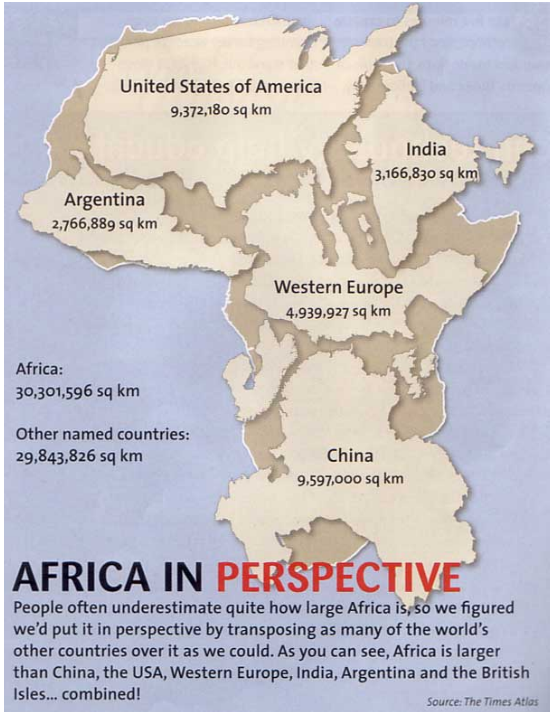
Map Projections

A map projection is a portrayal of the earth's surface onto a flat surface

There is no universally perfect projection

Projection classifications:

- Equivalent (Equal Area) – Area is similar on globe and flat map, but shape is not
- Conformal – Shape (scale) is similar on globe and flat map, but the area is not
- Cylindrical – projection of sphere onto a cylinder
- Conic – projection of sphere onto a cone
- Azimuthal – projection of a sphere onto a plane



Map Projections

https://source.opennews.org/articles/choosing-right-map-projection/

Choropleth Challenges

Need to standardize the statistics

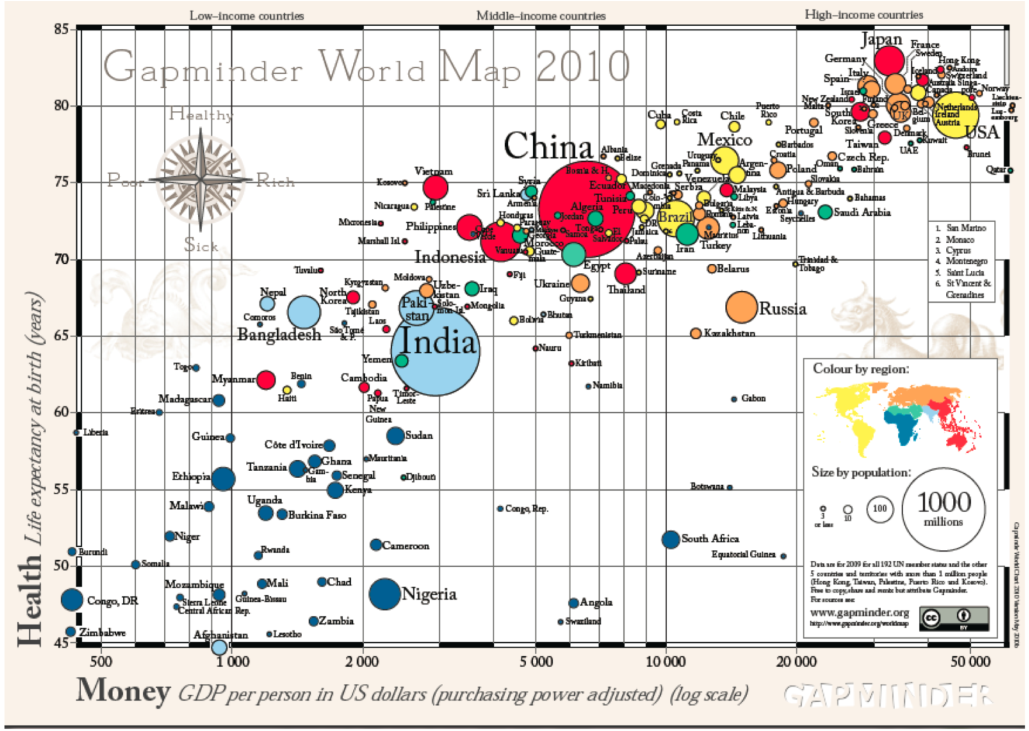
- Divide by population
- Divide by area

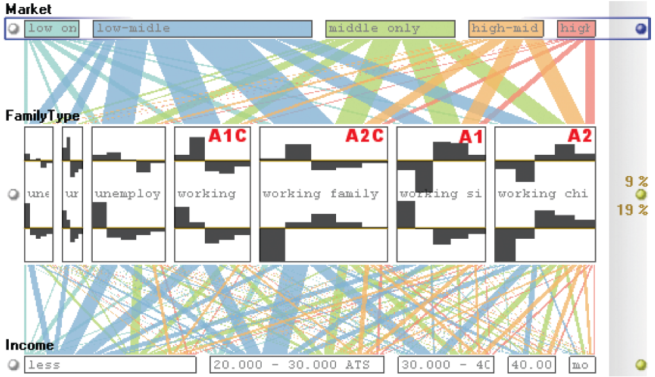
Need to be aware that size of areas can cause visual perception issues

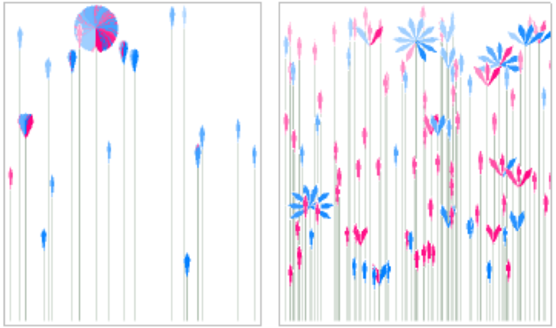
Modifiable areal unit problem – source of statistical bias occurring when data is aggregated into districts

Ecological Fallacy – inferences about the nature of individuals are based solely upon aggregate statistics collected for the group to which those individuals belong

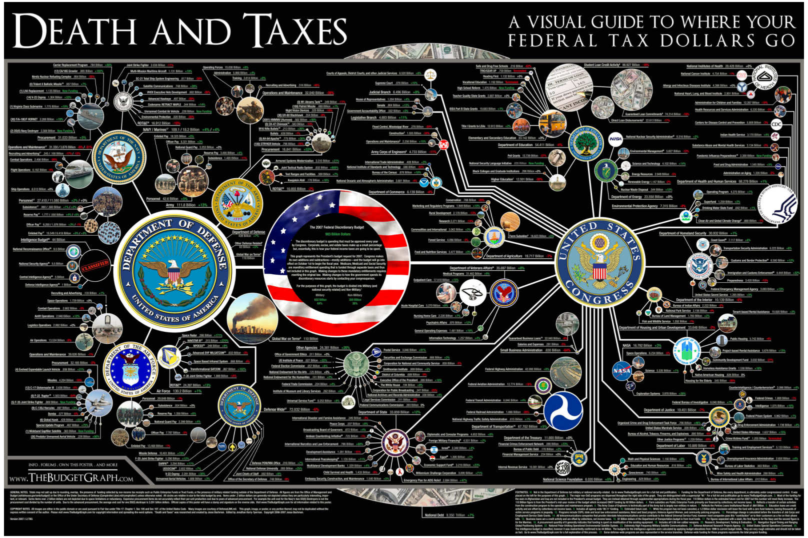
Further Afield

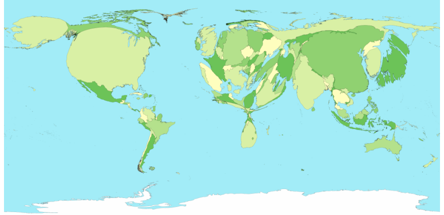


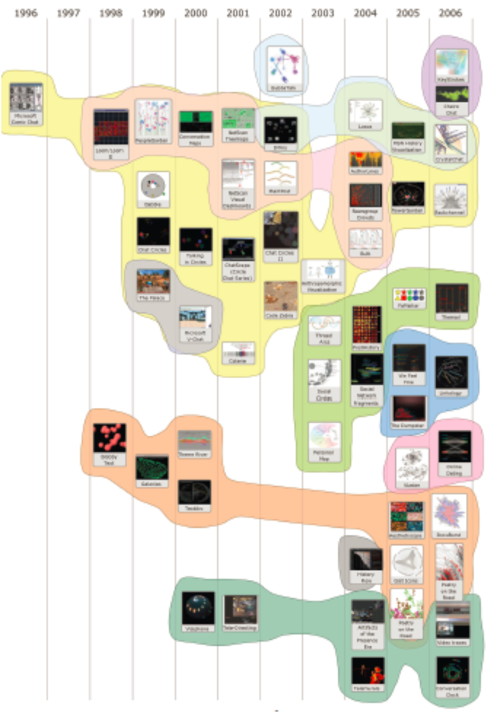


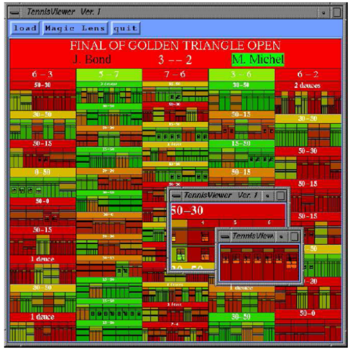


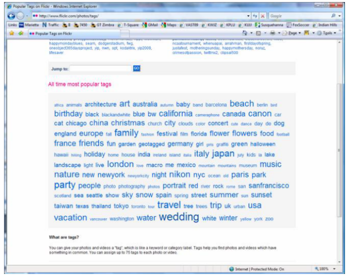
Further Afield











What is Visualization

- “The use of computer-supported, interactive visual representations of data to amplify cognition.”[1]
- This is not simply the process of making a graphic or an image, the goal is to create insight, not pretty pictures
- We want to help people form a mental image of something and internalize their own understanding
- We want to promote discovery, decision making and explanations

1-Readings in Information Visualization: Using Vision to Think, SK Card, J Mackinlay and B.Shneiderman, 1999

Why is Visualization Helpful?

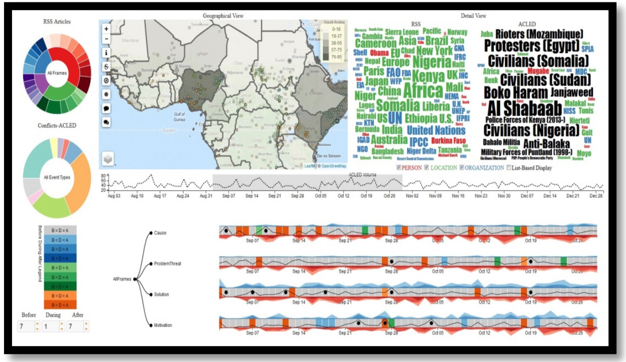
- It utilizes the high bandwidth of the human visual systems
- The human mind is fast and parallel
- Humans are great at visual pattern recognition
- We have pre-attentive visual phenomena
- Visual tools can extend memory and cognitive capacity
- We think visually!
- “A picture is worth a thousand words.” – *Printers Ink,* pp. 96-97 December 8, 1921

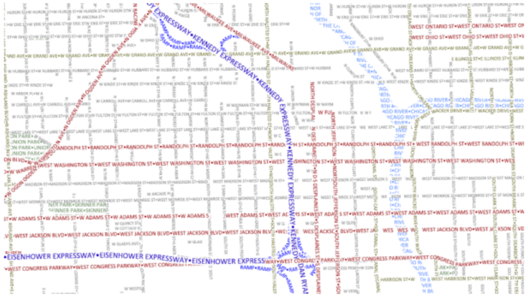
The Power of Visualization

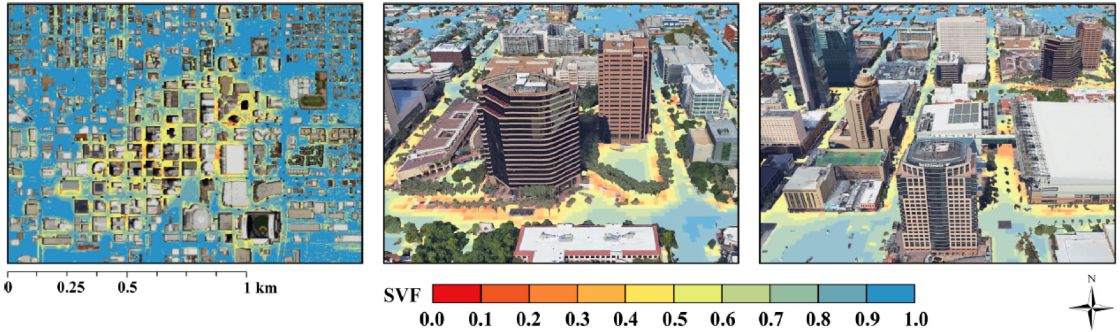
"Contained within the data of any investigation is information that can yield conclusions to questions not even originally asked. That is, there can be surprises in the data … To regularly miss surprises by failing to probe thoroughly with visualization tools is terribly inefficient because the cost of intensive data analysis is typically very small compared with the cost of data collection."

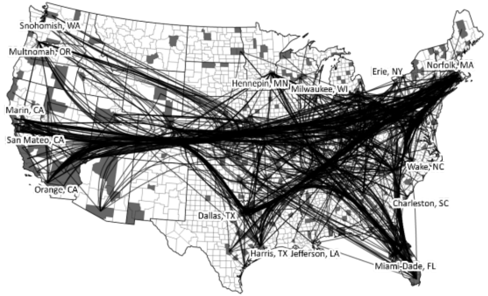
W. S. Cleveland
The Elements of Graphing Data

Visual Analytics and Data Exploration Research Lab









Vader.lab.asu.edu