National Institute of Allergy and Infectious Diseases

Data Visualization Concepts

Data Visualization in Python

NIAID

National Institute of Allergy and Infectious Diseases

R. Burke Squires
NIAID
Bioinformatics and Computational Biosciences Branch (BCBB)

Get the files…

(http://bioinformatics.niaid.nih.gov)
https://github.com/burkesquires/python_biologist

Download “data_viz_files.zip” in 06_data_viz_python folder
(Click on file, then click on download)

Uncompress

Move to desktop

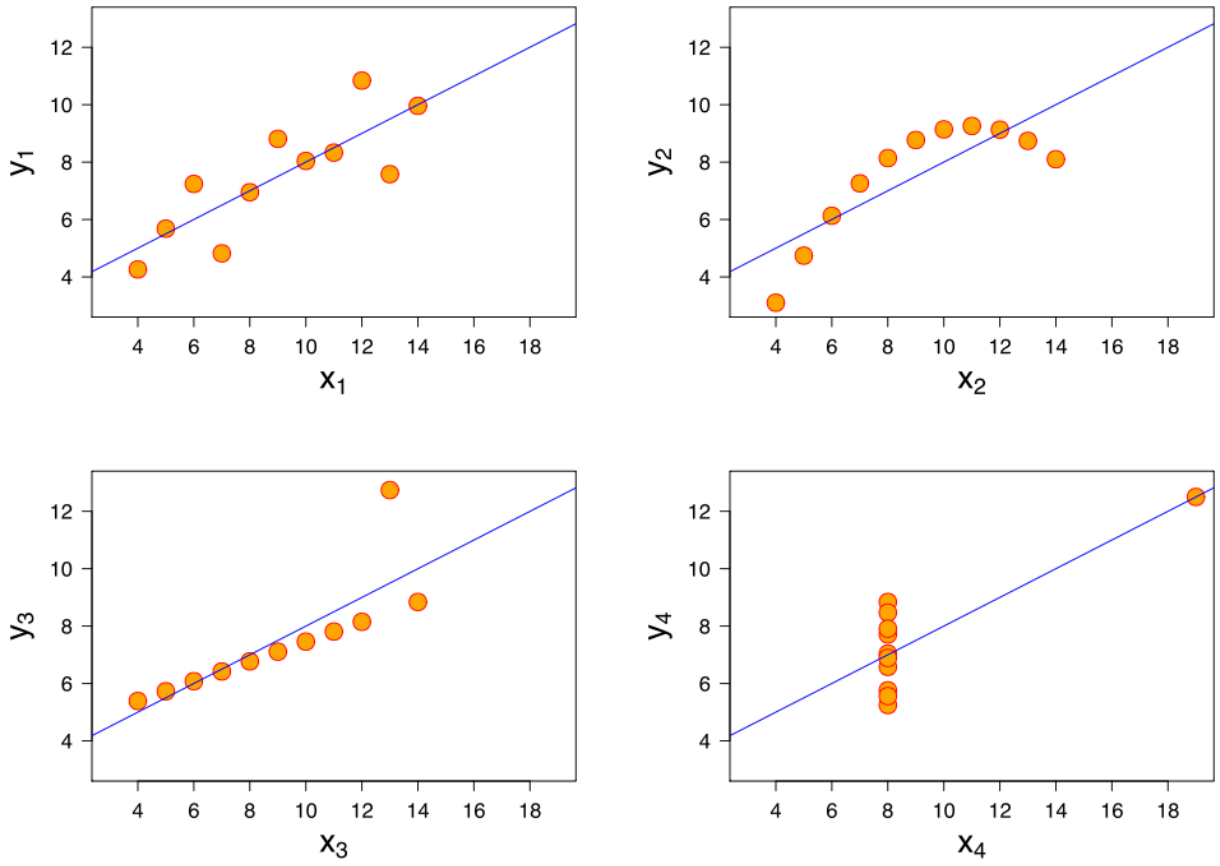
What is Data Visualization?

"The visual analysis of and communication of data"

Anscombe's Quartet

	I		II		III		IV	
	x	y	x	y	x	y	x	y
	10	8,04	10	9,14	10	7,46	8	6,58
	8	6,95	8	8,14	8	6,77	8	5,76
	13	7,58	13	8,74	13	12,74	8	7,71
	9	8,81	9	8,77	9	7,11	8	8,84
	11	8,33	11	9,26	11	7,81	8	8,47
	14	9,96	14	8,1	14	8,84	8	7,04
	6	7,24	6	6,13	6	6,08	8	5,25
	4	4,26	4	3,1	4	5,39	19	12,5
	12	10,84	12	9,13	12	8,15	8	5,56
	7	4,82	7	7,26	7	6,42	8	7,91
	5	5,68	5	4,74	5	5,73	8	6,89
SUM	99,00	82,51	99,00	82,51	99,00	82,50	99,00	82,51
AVG	9,00	7,50	9,00	7,50	9,00	7,50	9,00	7,50
STDEV	3,32	2,03	3,32	2,03	3,32	2,03	3,32	2,03

Anscombe's Quartet

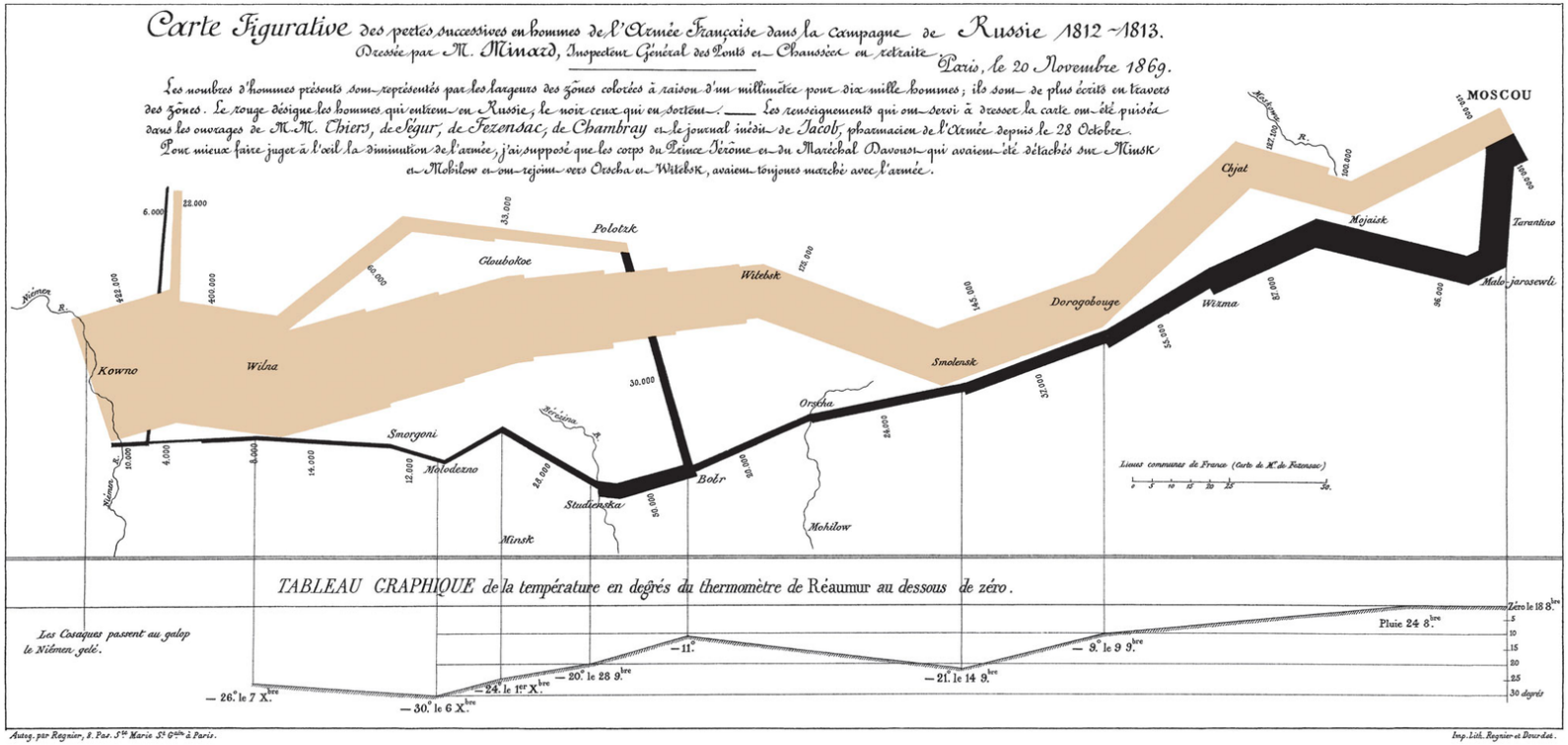


Objectives

- Data Analysis
  - Understand the data
  - Derive information from data

- Communication
  - Of information
  - Of trends
  - Involves simplification

Napoleon's March by Charles Minard



National Institute of Allergy and Infectious Diseases

Minard's Greatest Hit, c.1869

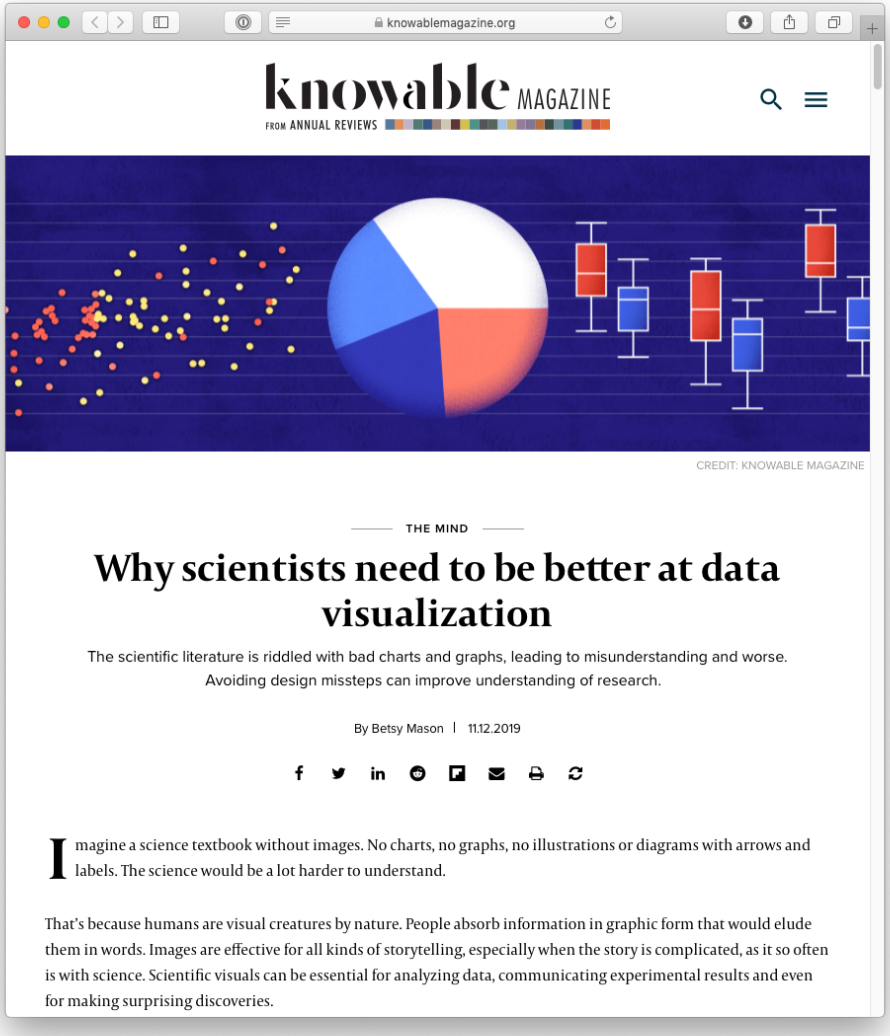
NIAID

Why scientists need to be better at data visualization

The scientific literature is riddled with bad charts and graphs, leading to misunderstanding and worse. Avoiding design missteps can improve understanding of research.

By Betsy Mason

Knowable Magazine
11.12.2019



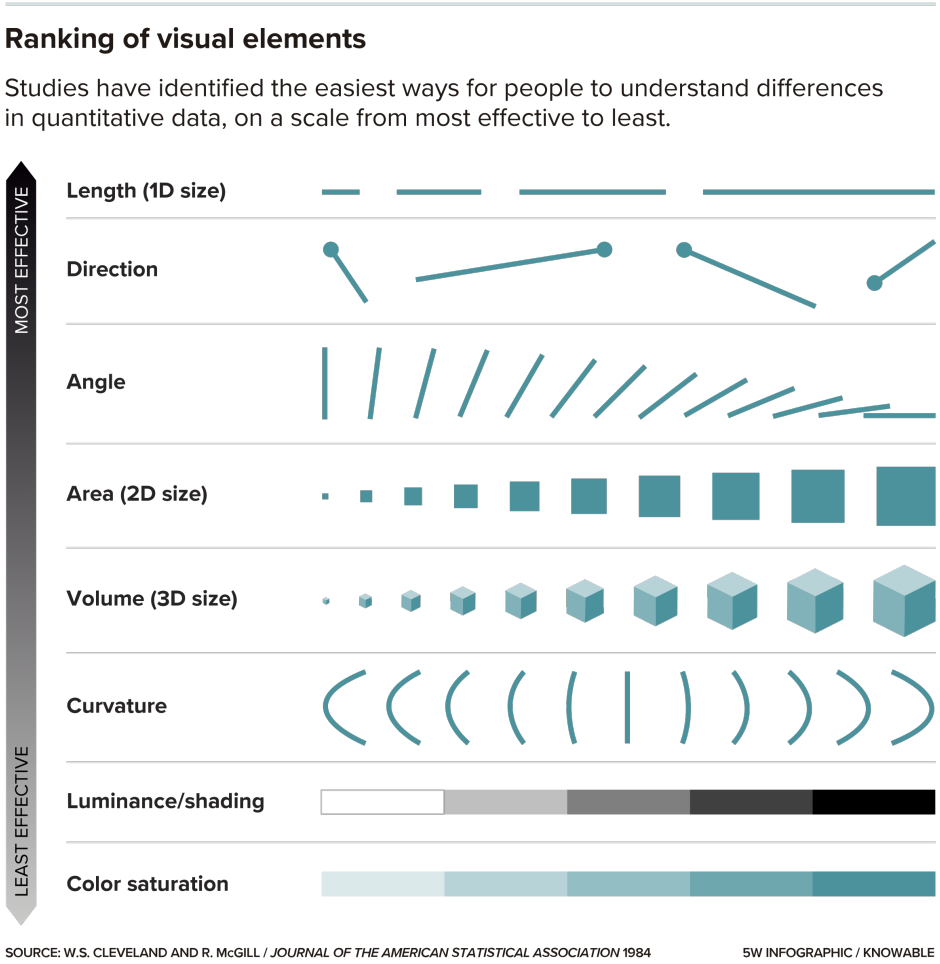
https://www.knowablemagazine.org/article/mind/2019/science-data-visualization

NIH National Institute of Allergy and Infectious Diseases

NIAID

Ranking Visual Elements

People are better at discerning subtleties in some types of visuals than others — the length of two lines, for example, or the direction of a line are easier to tell apart than shades of gray or the intensity of a color. Studies show that graphs using visual elements high on this list are easier to read and more effective than those near the bottom.

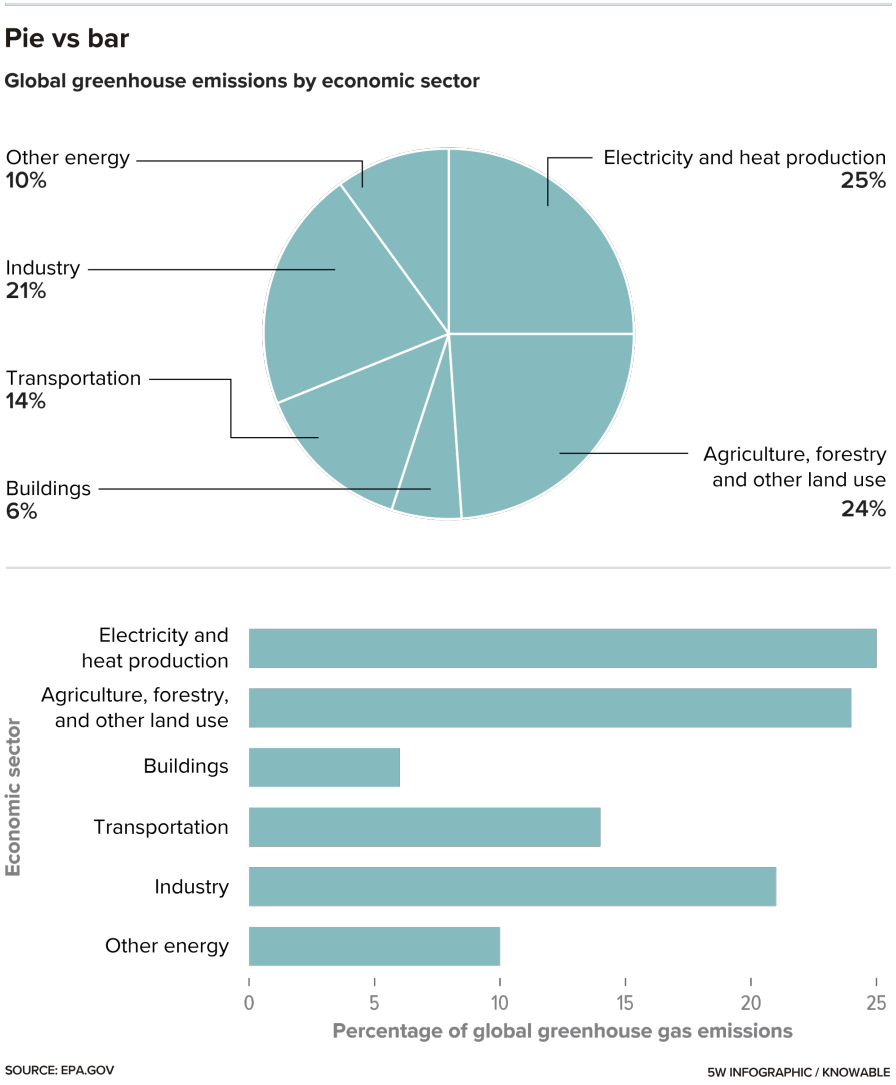


NIH National Institute of Allergy and Infectious Diseases

NIAID

Pie vs Bar Charts

Pie charts are best used to show the relationship of the pieces to the whole. In this graphic, the pie chart effectively shows how big or small each economic sector's contribution is to total greenhouse gas emissions, but it's difficult to compare sectors with each other. The bar chart allows for easy comparisons between sectors but doesn't convey how each one relates to the total.



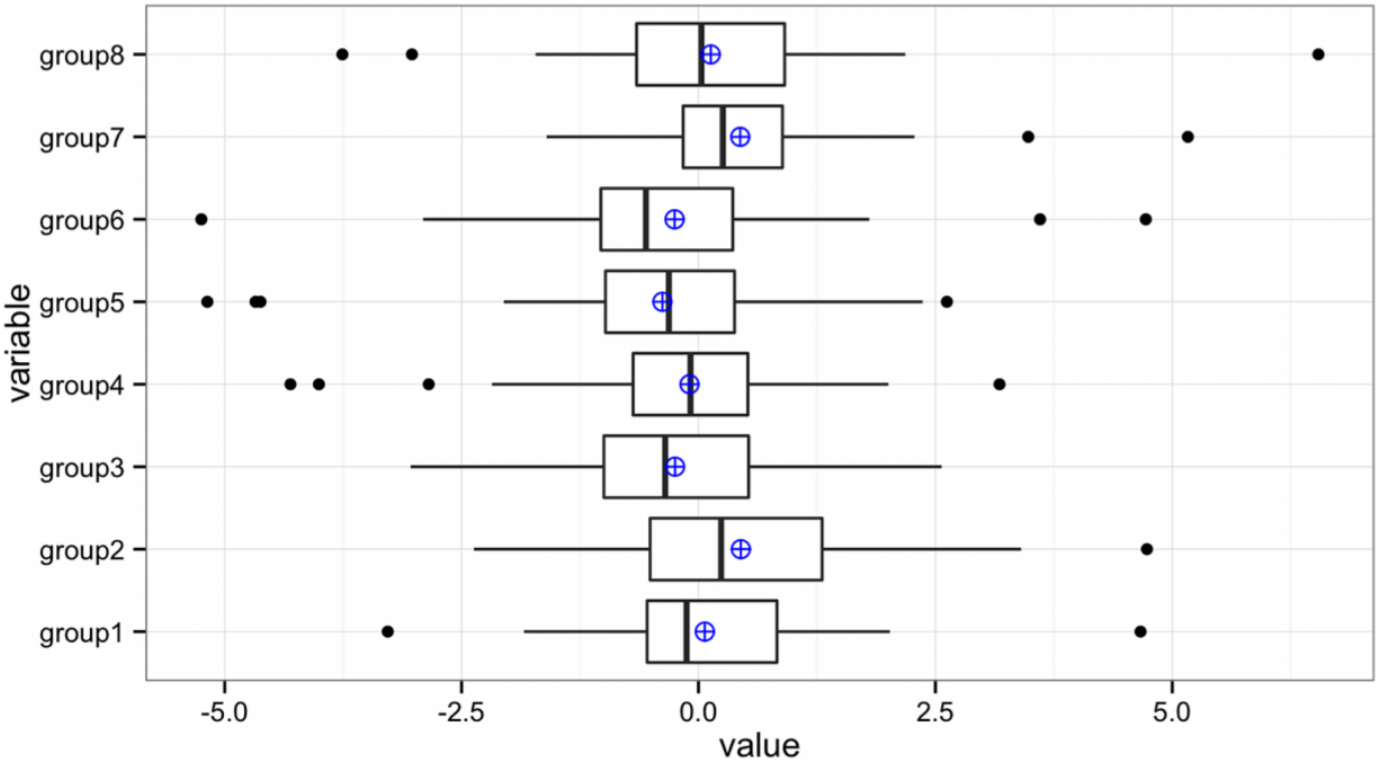
Ten-ish Simple Rules / Guidelines

- Rule 1: Know Your Audience
- Rule 2: Identify Your Message
- Rule 3: Adapt the Figure to the Support Medium
- Rule 4: Captions Are Not Optional
- Rule 5: Do Not Trust the Defaults
- Rule 6: Use Color Effectively
- Rule 7: Do Not Mislead the Reader
- Rule 8: Avoid “Chartjunk”
- Rule 9: Message Trumps Beauty
- Rule 10: Get the Right Tool

http://journals.plos.org/ploscompbiol/article?id=10.1371/journal.pcbi.1003833

Ten-ish Simple Rules / Guidelines

- Know your audience



National Institute of Allergy and Infectious Diseases

http://journals.plos.org/ploscompbiol/article?id=10.1371/journal.pcbi.1003833

NIAID

Ten-ish Simple Rules / Guidelines

- Identify your message

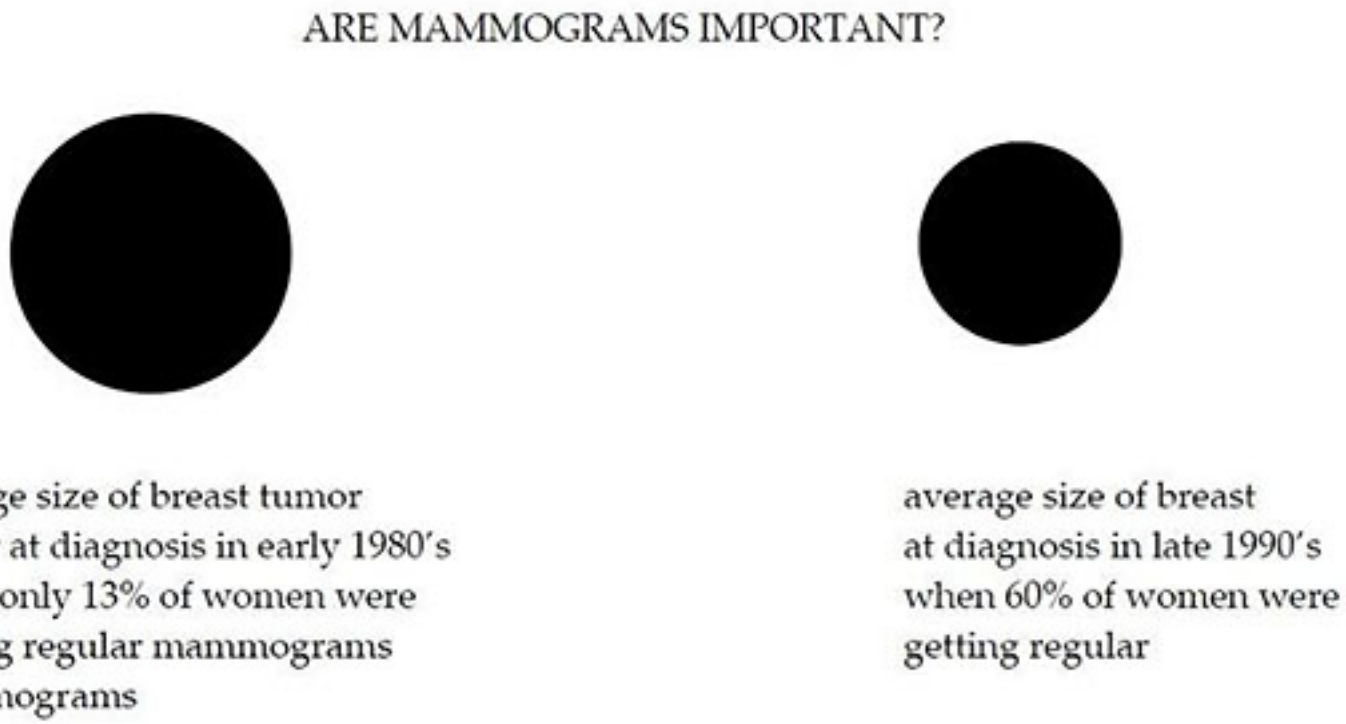
National Institute of Allergy and Infectious Diseases

http://upload.wikimedia.org/wikipedia/commons/9/9b/Social_Network_Analysis_Visualization.png

NIAID

Ten-ish Simple Rules / Guidelines

- Captions are not optional

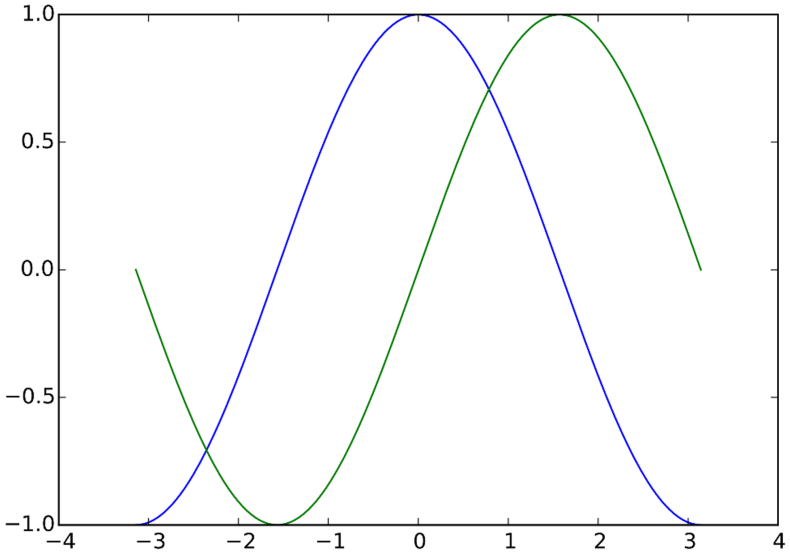


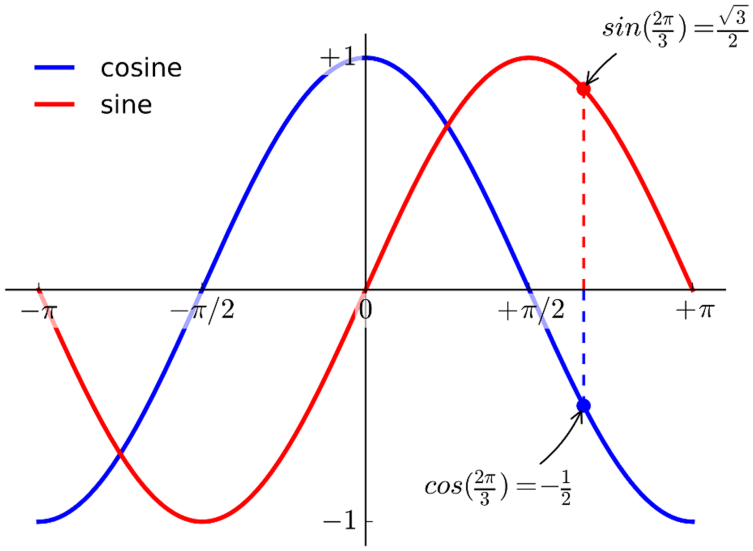
National Institute of Allergy and Infectious Diseases

http://ksrowell.com/blog-visualizing-data/2010/12/23/data-visualization-double-take/

Ten-ish Simple Rules / Guidelines

- Do not use the defaults





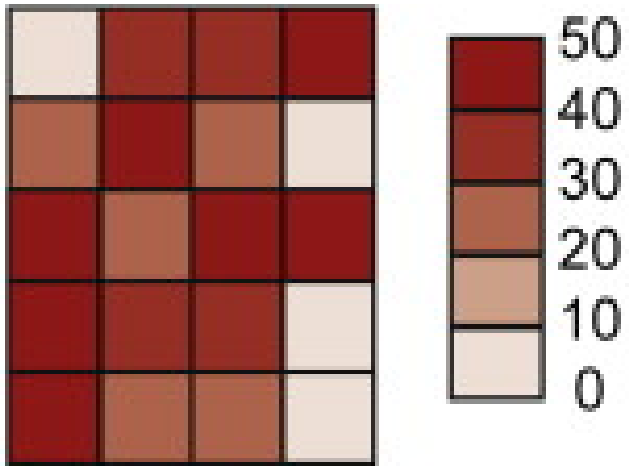
National Institute of Allergy and Infectious Diseases

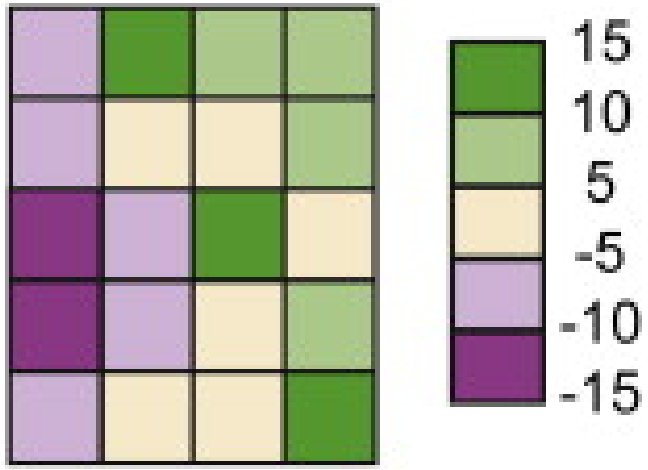
http://journals.plos.org/ploscompbiol/article?id=10.1371/journal.pcbi.1003833

NIAID

Ten-ish Simple Rules / Guidelines

- Use color effectively

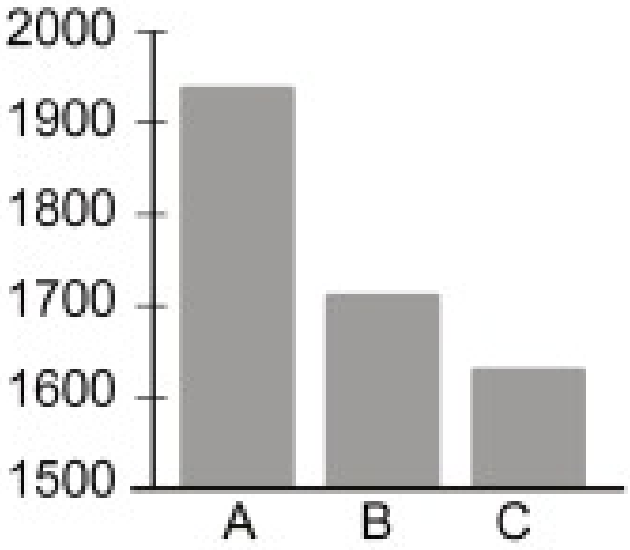


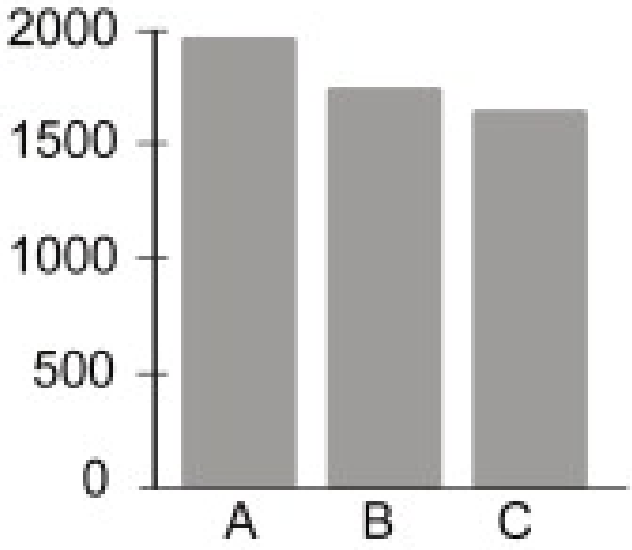


http://journals.plos.org/ploscompbiol/article?id=10.1371/journal.pcbi.1003833

Ten-ish Simple Rules / Guidelines

- Do not mislead the reader





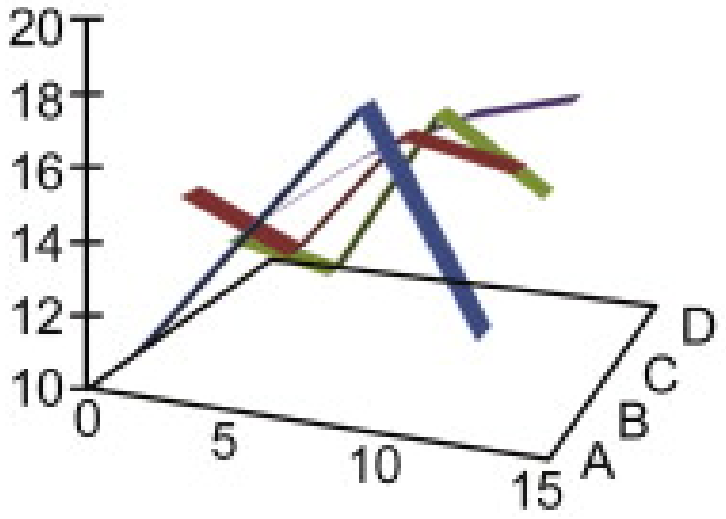
National Institute of Allergy and Infectious Diseases

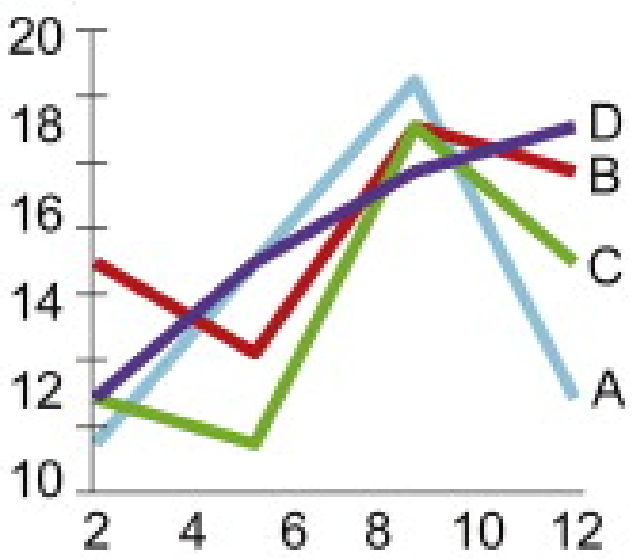
http://journals.plos.org/ploscompbiol/article?id=10.1371/journal.pcbi.1003833

NIAID

Ten-ish Simple Rules / Guidelines

- Create the simplest graph





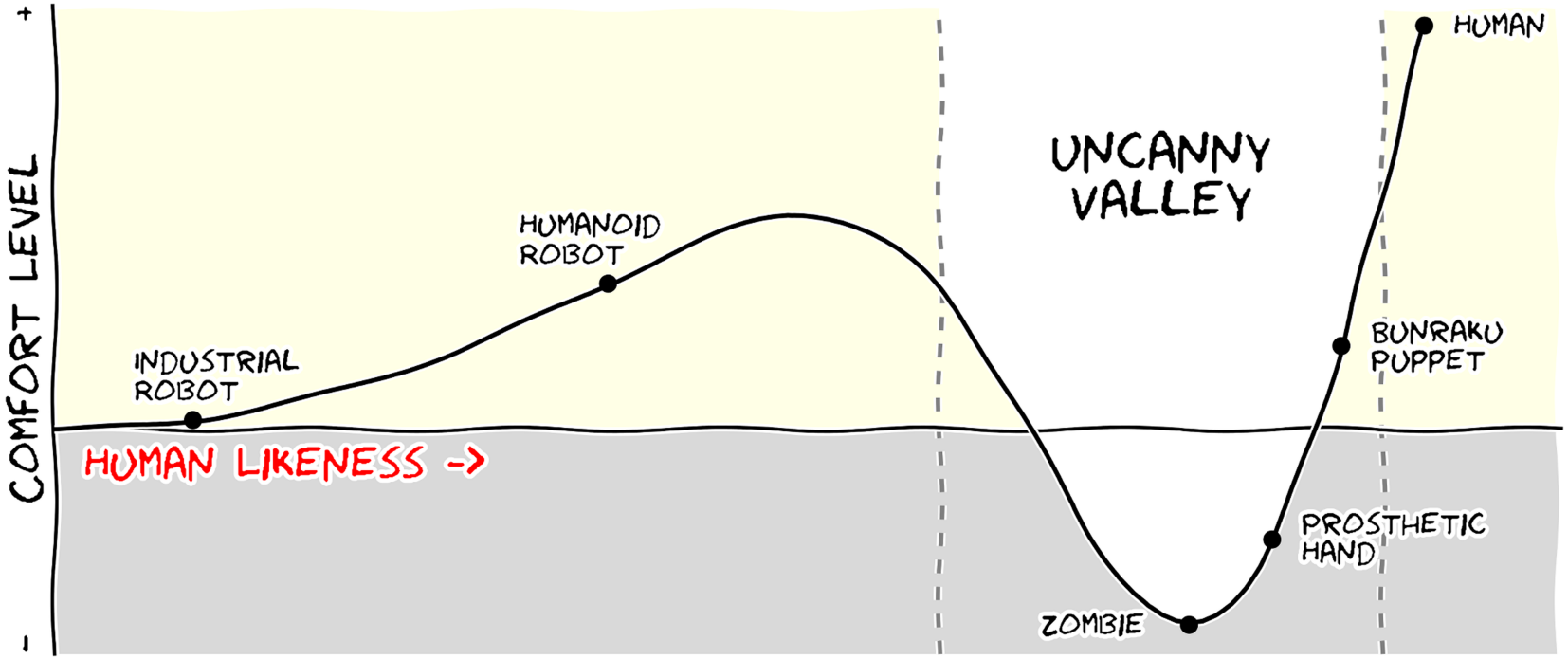
National Institute of Allergy and Infectious Diseases

http://journals.plos.org/ploscompbiol/article?id=10.1371/journal.pcbi.1003833

NIAID

Ten-ish Simple Rules / Guidelines

- Message trumps beauty



National Institute of Allergy and Infectious Diseases

http://journals.plos.org/ploscompbiol/article?id=10.1371/journal.pcbi.1003833

NIAID

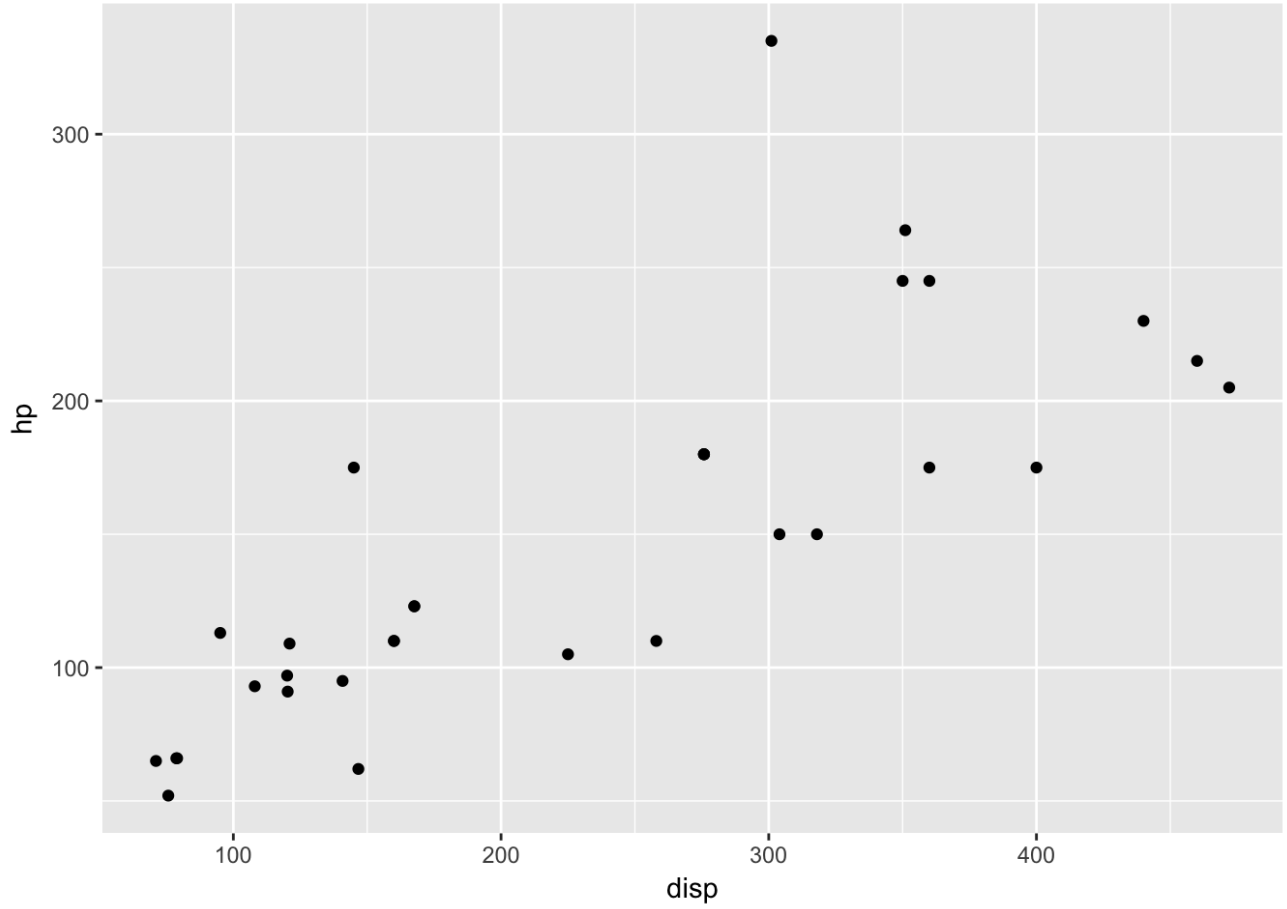
Ten-ish Simple Rules / Guidelines

- Get the right tool…

http://journals.plos.org/ploscompbiol/article?id=10.1371/journal.pcbi.1003833

How Much Data Can You Put In A Chart?

How Much Data Can You Present In A Single Graph? Two Dimensions?

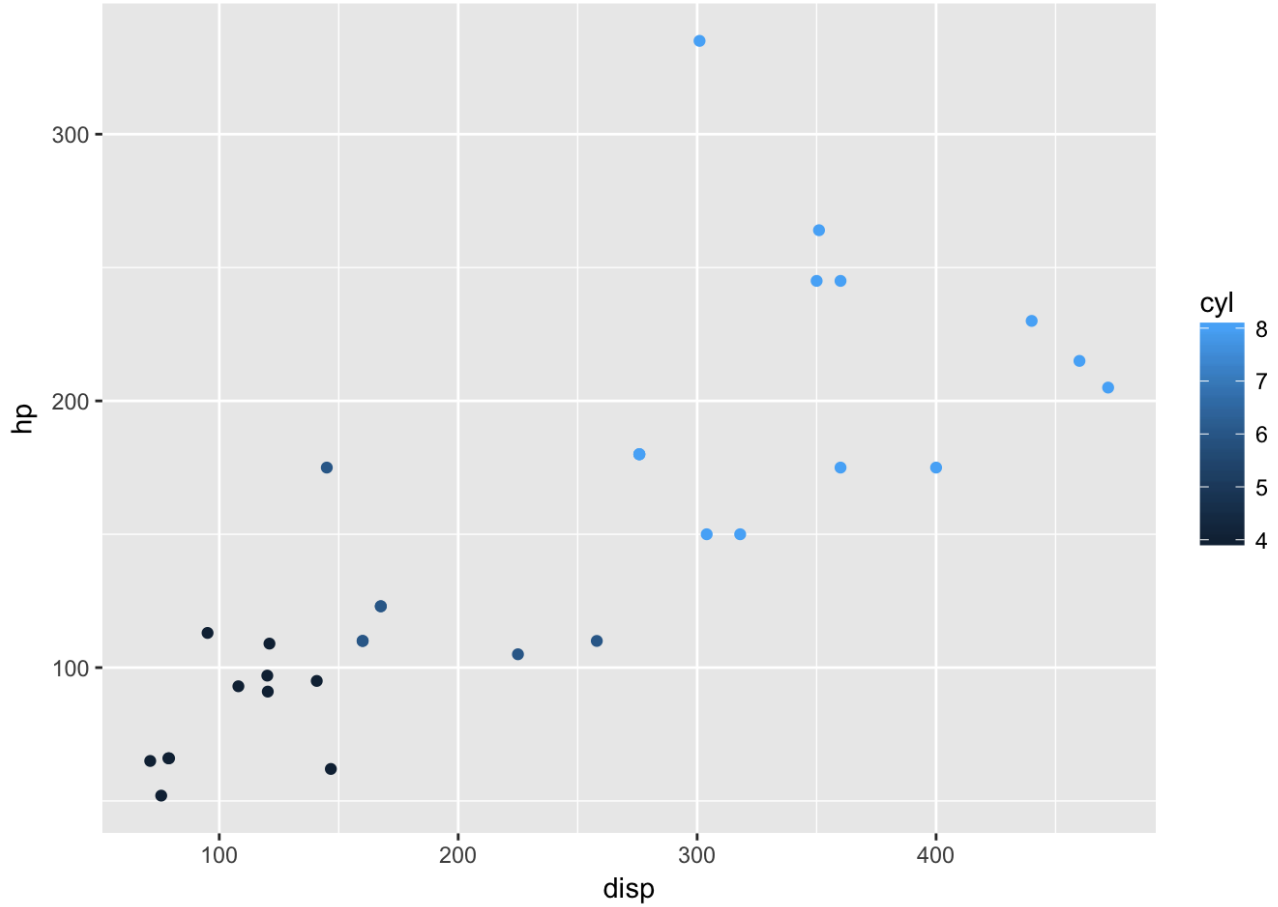


National Institute of Allergy and Infectious Diseases

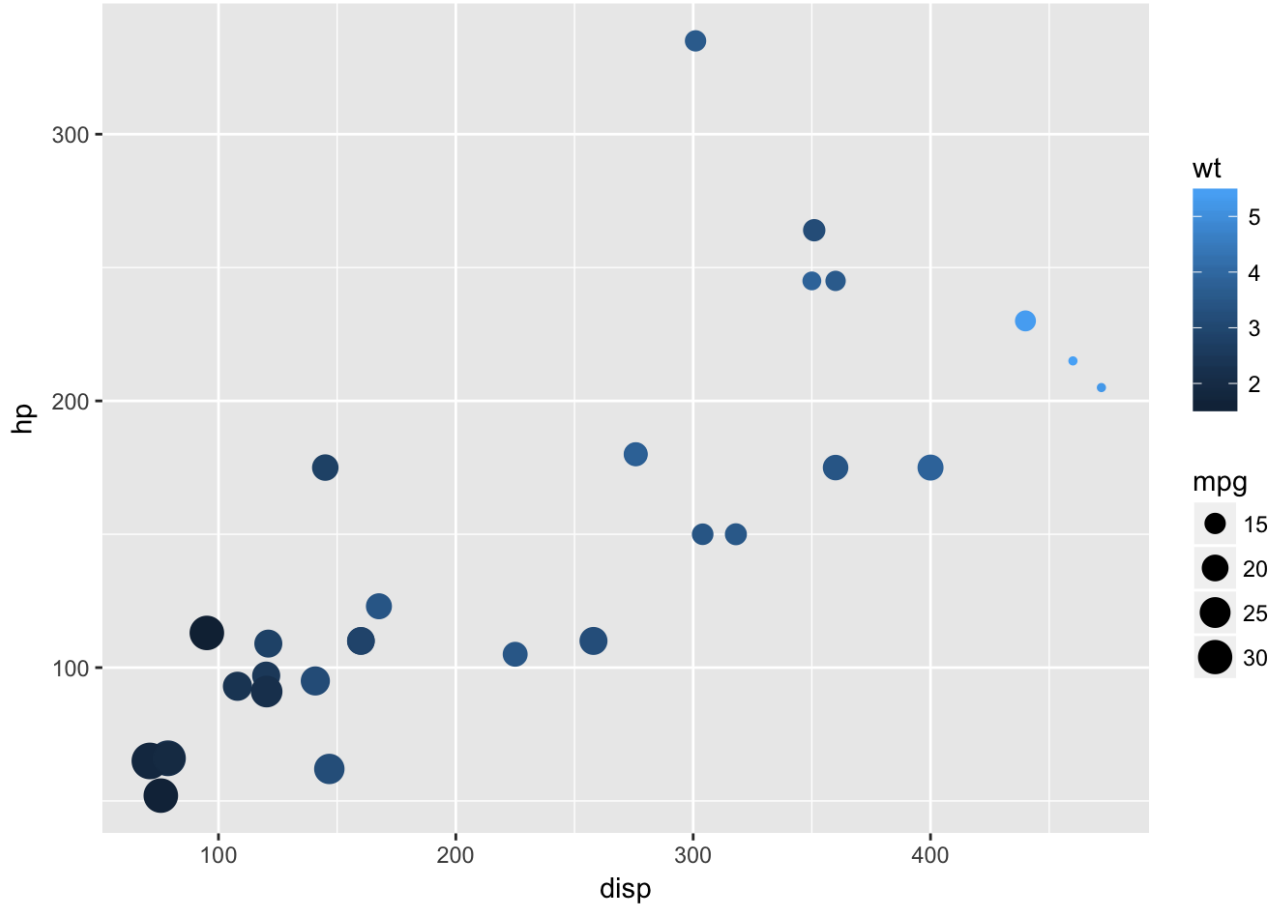
NIAID

Source: Sean Davis, NCI - https://github.com/seandavi/hour_of_code; NOTE: These are "R" visualizations

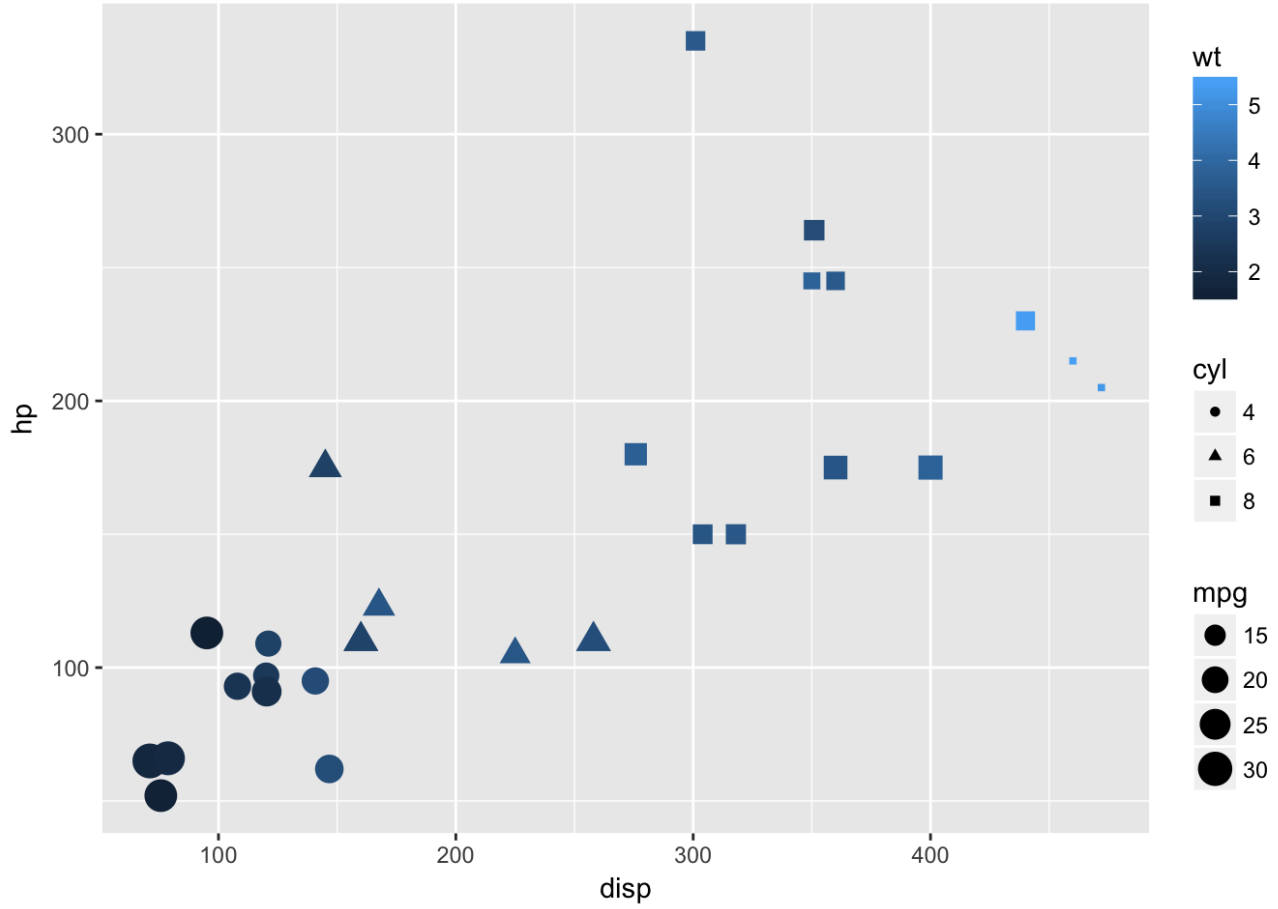
How Much Data Can You Present In A Single Graph? Three Dimensions?



How Much Data Can You Present In A Single Graph? Four Dimensions?

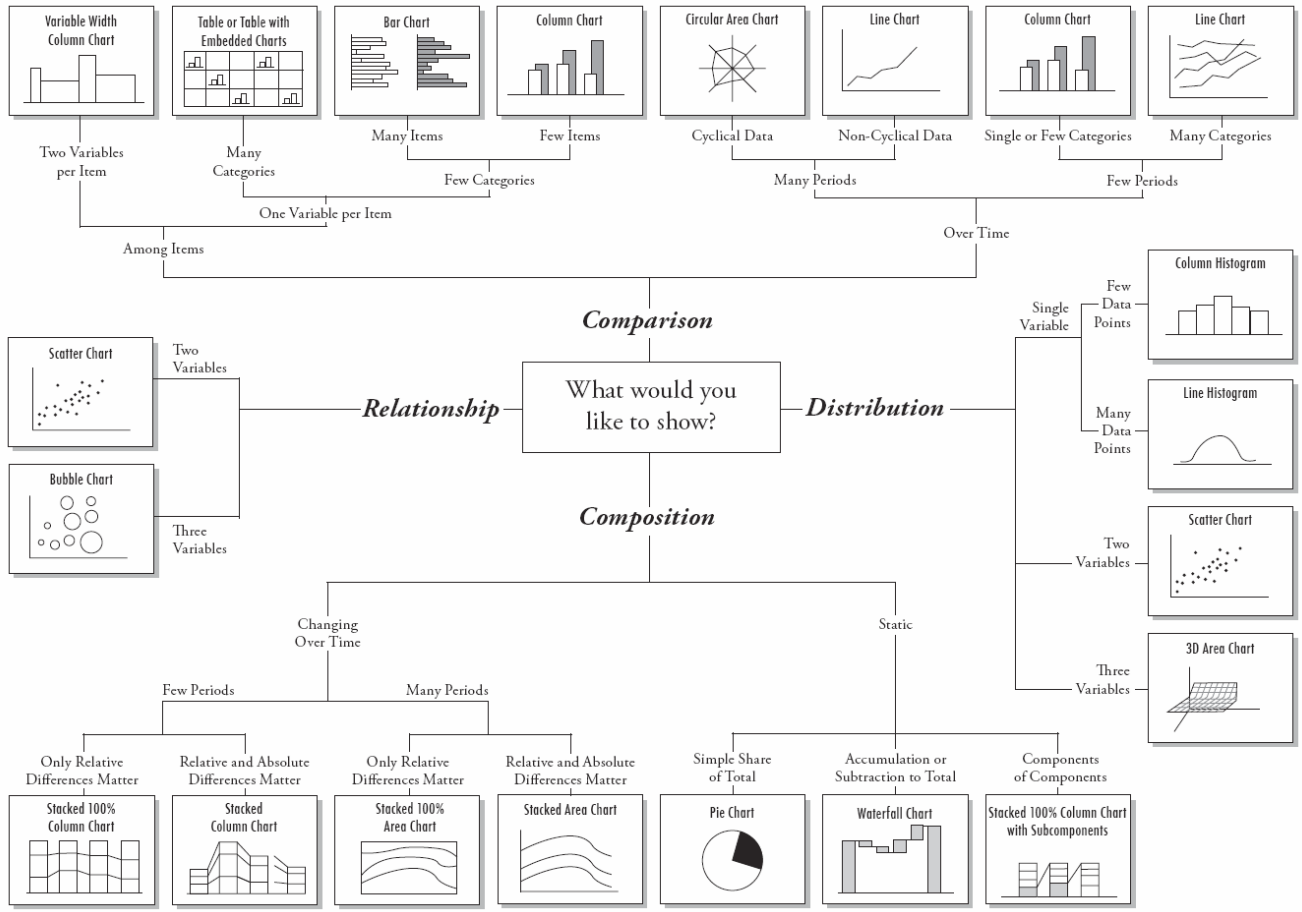


How Much Data Can You Present In A Single Graph? Five Dimensions?



How Do I Know Which Chart To Use?

Choose the Right Chart Type for your Data



National Institute of Allergy and Infectious Diseases

Source: http://img.labnol.org/di/choosing_a_good_chart2.pdf

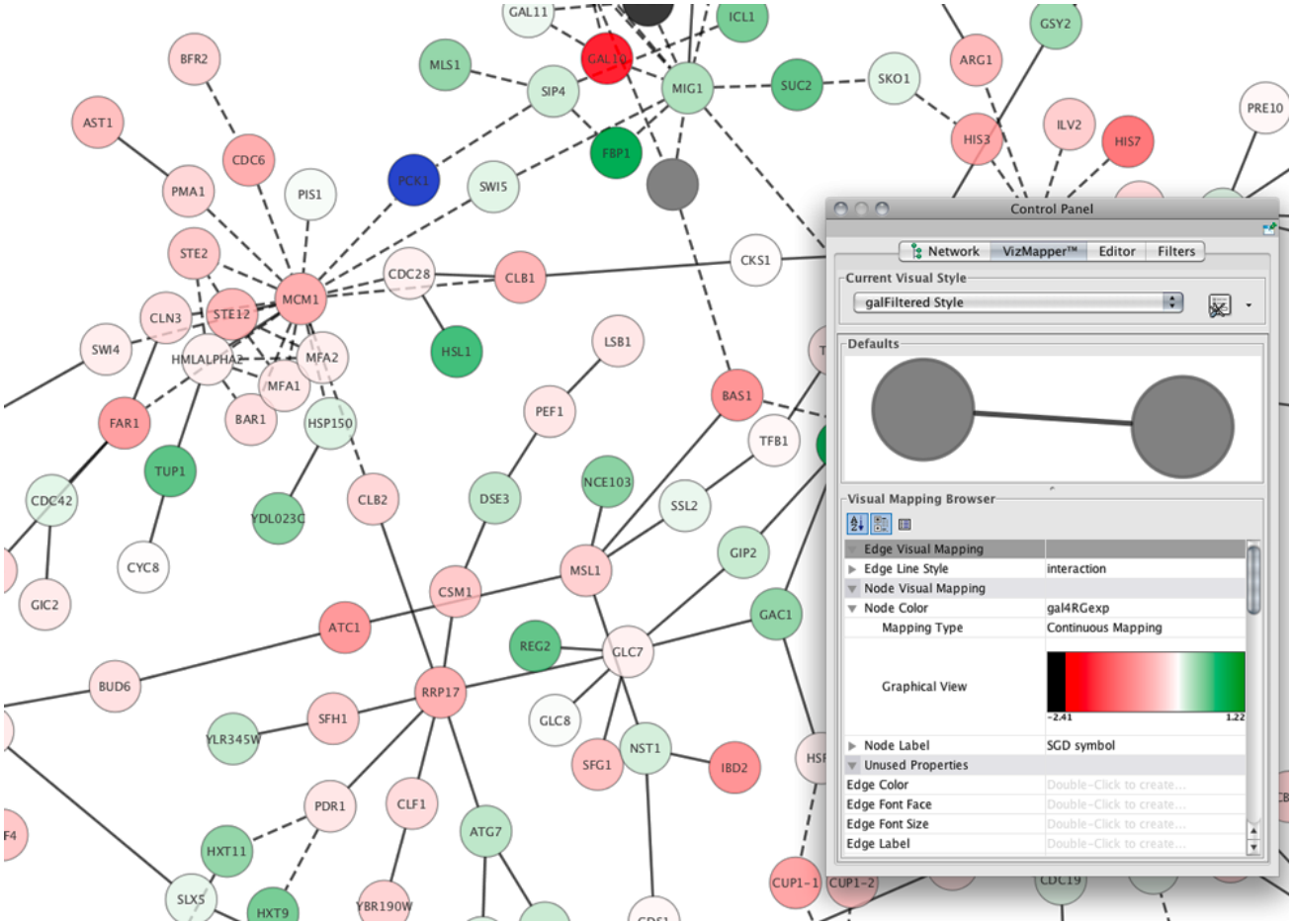
NIAID

Data Visualization

- Static data visualization
  - matplotlib, seaborn, ggpy (ggplot)
- Dynamic visualization
  - Bokeh
- Declarative Visualization
  - Altair
- Specialized visualizations
  - Mapping of geographical data, timelines
  - 3D visualization – Vispy

Additional (non-python) Data Visualization Techniques:

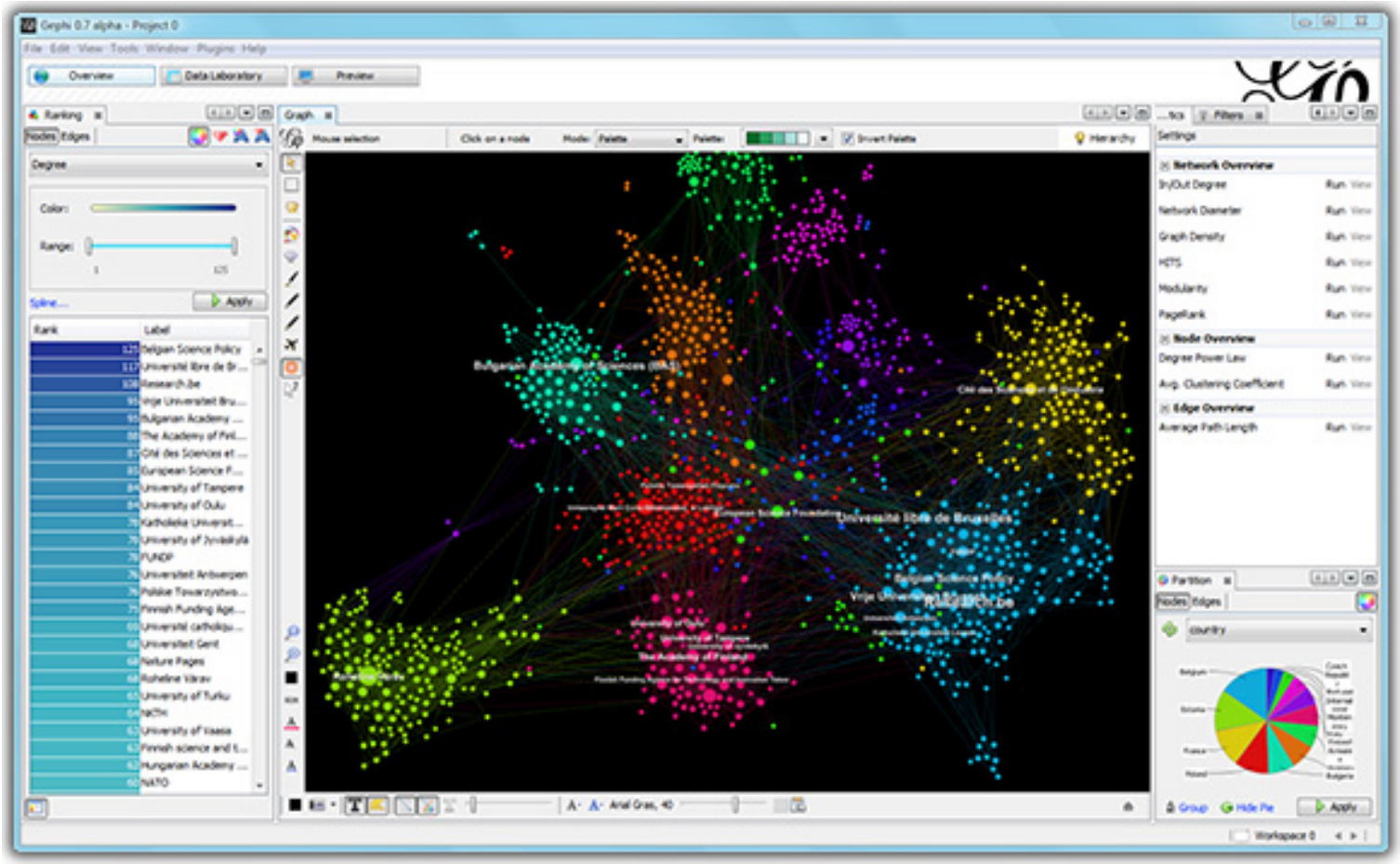
Network Visualization: Cytoscape



National Institute of Allergy and Infectious Diseases

NIAID

Network Visualization: Gelphi



National Institute of Allergy and Infectious Diseases

NIAID

References

- **Ten Simple Rules for Better Figures**. Nicolas P. Rougier , Michael Droettboom, Philip E. Bourne. PLOS Computational Biology. September 11, 2014

Resources

- one-dataset-visualized-25-ways
  - http://flowingdata.com/2017/01/24/one-dataset-visualized-25-ways/

Resources

- Nature Methods: Point of View column
  - http://blogs.nature.com/methagora/2013/07/data-visualization-points-of-view.html
  - http://mkweb.bcgsc.ca/pointsofview/
- Publications
  - https://www.knowablemagazine.org/article/mind/2019/science-data-visualization
  - https://www.annualreviews.org/doi/pdf/10.1146/annurev-biodatasci-080917-013424

- http://www.vanseodesign.com/web-design/gestalt-principles-of-perception/
- Course
  - http://enrico.bertini.io/teaching/
- Chart Chooser
  - http://labs.juiceanalytics.com/chartchooser/