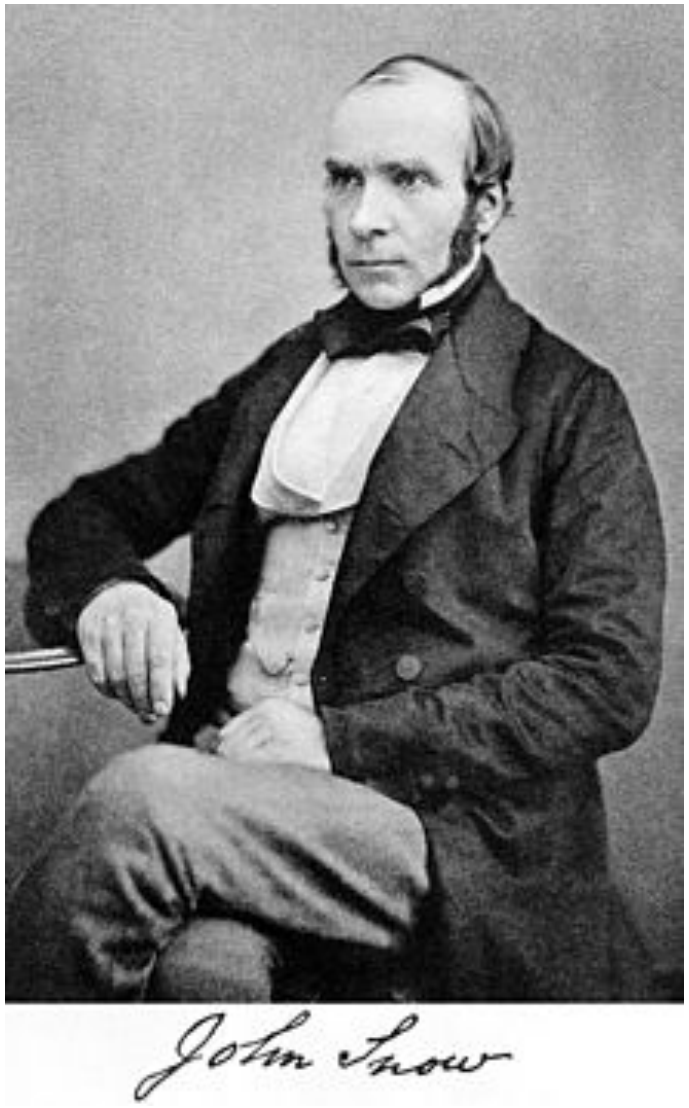
CS639: Data Management for Data Science

Lecture 24: Data Vizualization

[based on slides by John Canny]

Theodoros Rekatsinas

John Snow



Hypothesis at the time:
Diseases such as cholera and bubonic plague are caused by pollution or a noxious form of "bad air”.

John Snow’s Research:
The source of the outbreak was the public water pump [On the Mode of Communication of Cholera]

How did he do it?

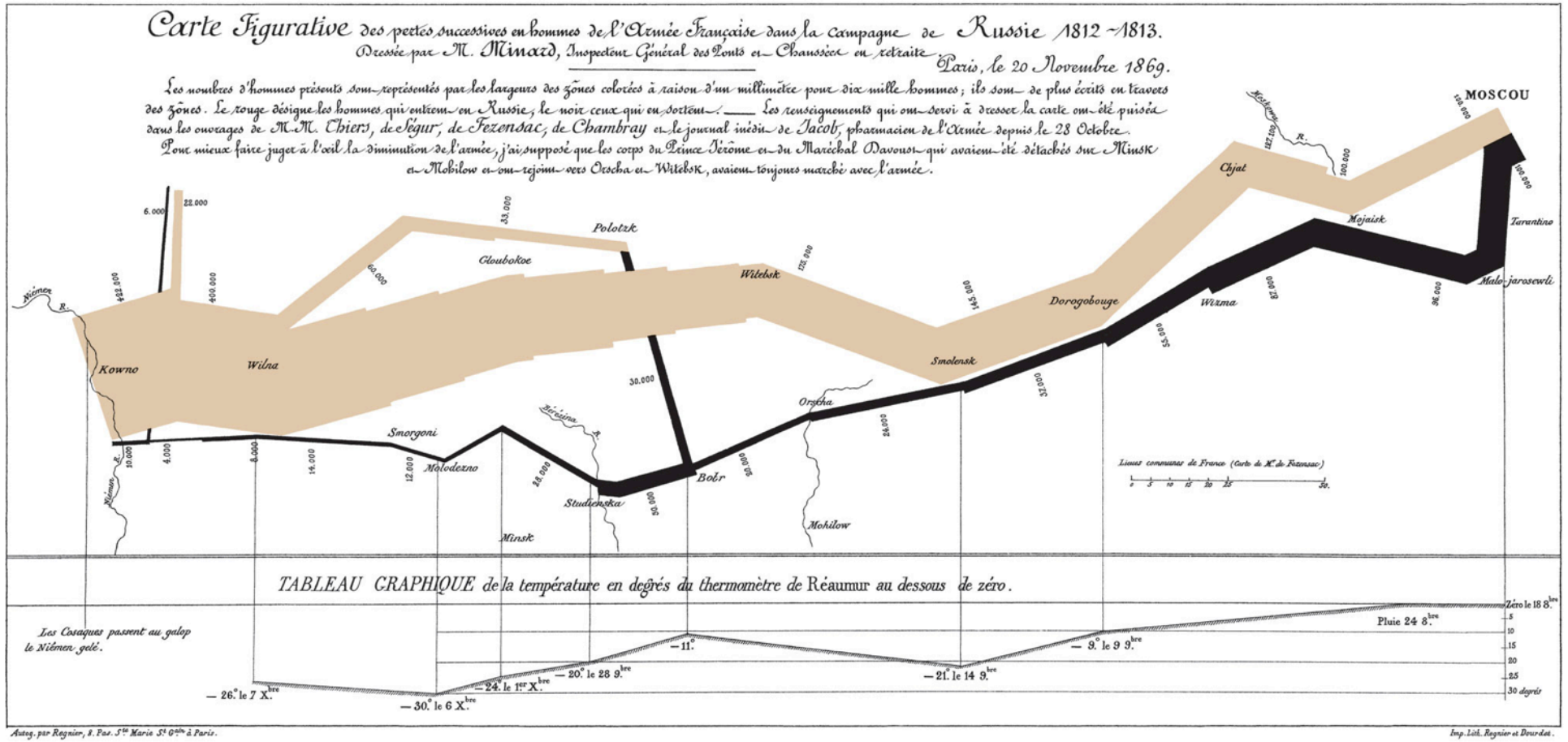
Death/Survival in cholera in 19th century London by district and water company.

St. Savior, Southwark	1	0	19211
St. Savior, Southwark	1	1	406
St. Savior, Southwark	2	0	14129
St. Savior, Southwark	2	1	72
St. Olave, Southwark	1	0	18361
St. Olave, Southwark	1	1	277
St. Olave, Southwark	2	0	0
St. Olave, Southwark	2	1	0
St. George, Southwark	1	0	24651
St. George, Southwark	1	1	388
St. George, Southwark	2	0	23613
St. George, Southwark	2	1	99
Bermondsey	1	0	57063
Bermondsey	1	1	821
Bermondsey	2	0	1785
Bermondsey	2	1	0
Newington	1	0	31482
Newington	1	1	458
Newington	2	0	33473
Newington	2	1	58
Lambeth	1	0	54457
Lambeth	1	1	525
Lambeth	2	0	83648
Lambeth	2	1	138
Wandsworth	1	0	18122
Wandsworth	1	1	268
Wandsworth	2	0	3863
Wandsworth	2	1	7
Campberwell	1	0	23120
Campberwell	1	1	352
Campberwell	2	0	10445
Campberwell	2	1	33
Rotherhithe	1	0	14744
Rotherhithe	1	1	207
Rotherhithe	2	0	0
Rotherhithe	2	1	0

How did he do it?

Snow's **data visualization** study is regarded as the founding event of the science of epidemiology.

Charles Joseph Minard 1869 Napoleon's March



According to Tufte: "It may well be the best statistical graphic ever drawn."
5 variables: Army Size, location, dates, direction, temperature during retreat
https://news.nationalgeographic.com/2017/03/charles-minard-cartography-infographics-history/

Interactivity to Educate

- The famous Gapminder Video, Hans Rosling:
  200 Countries, 200 Years, 4 Minutes
- https://www.youtube.com/watch?feature=player_embedded&v=jbkSRLYSojo

Outline

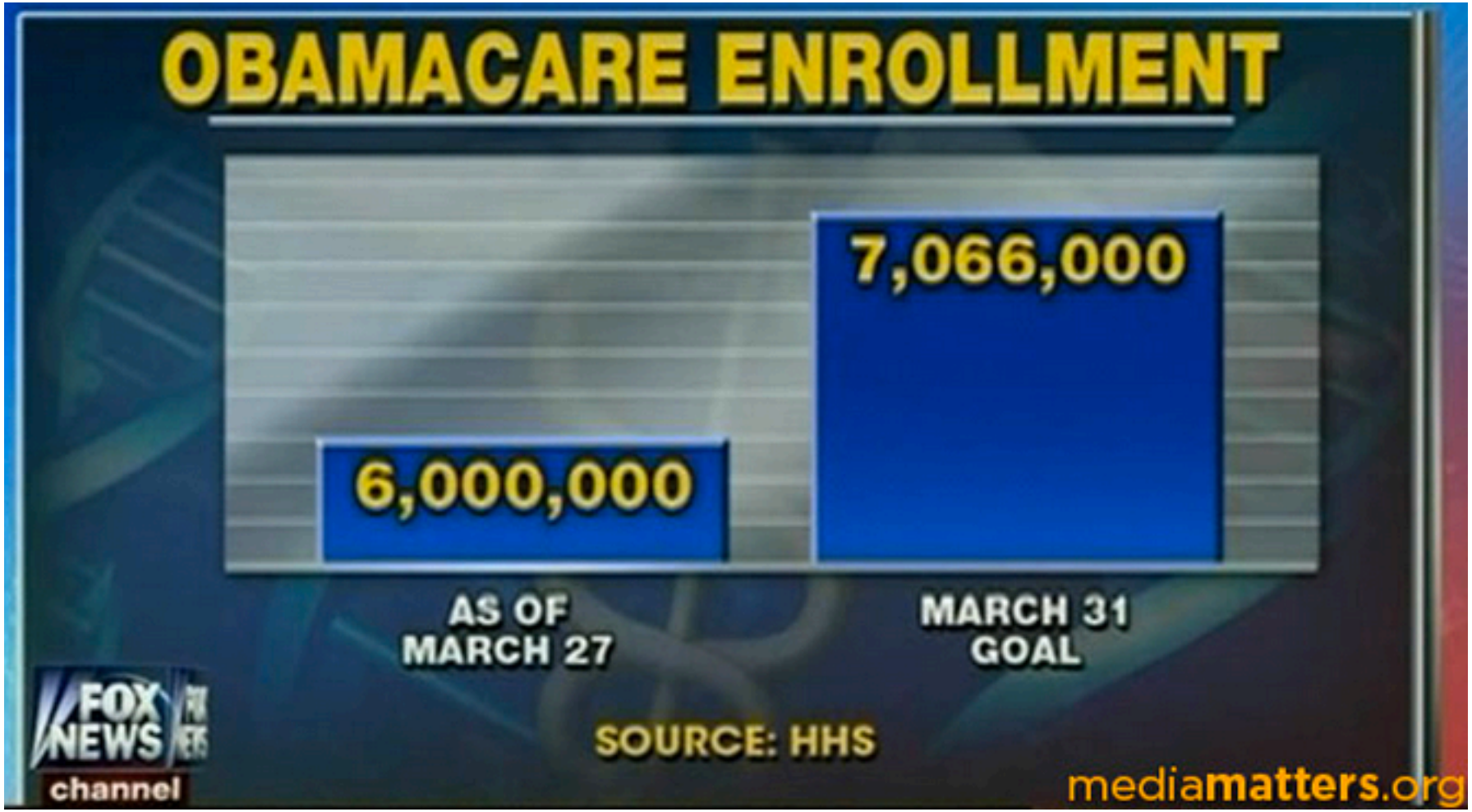
Visualization:

- Some great examples
- Some counter-examples
- Principles for Visualization Design
- Visualization Toolkits preview

Some Anti-Examples

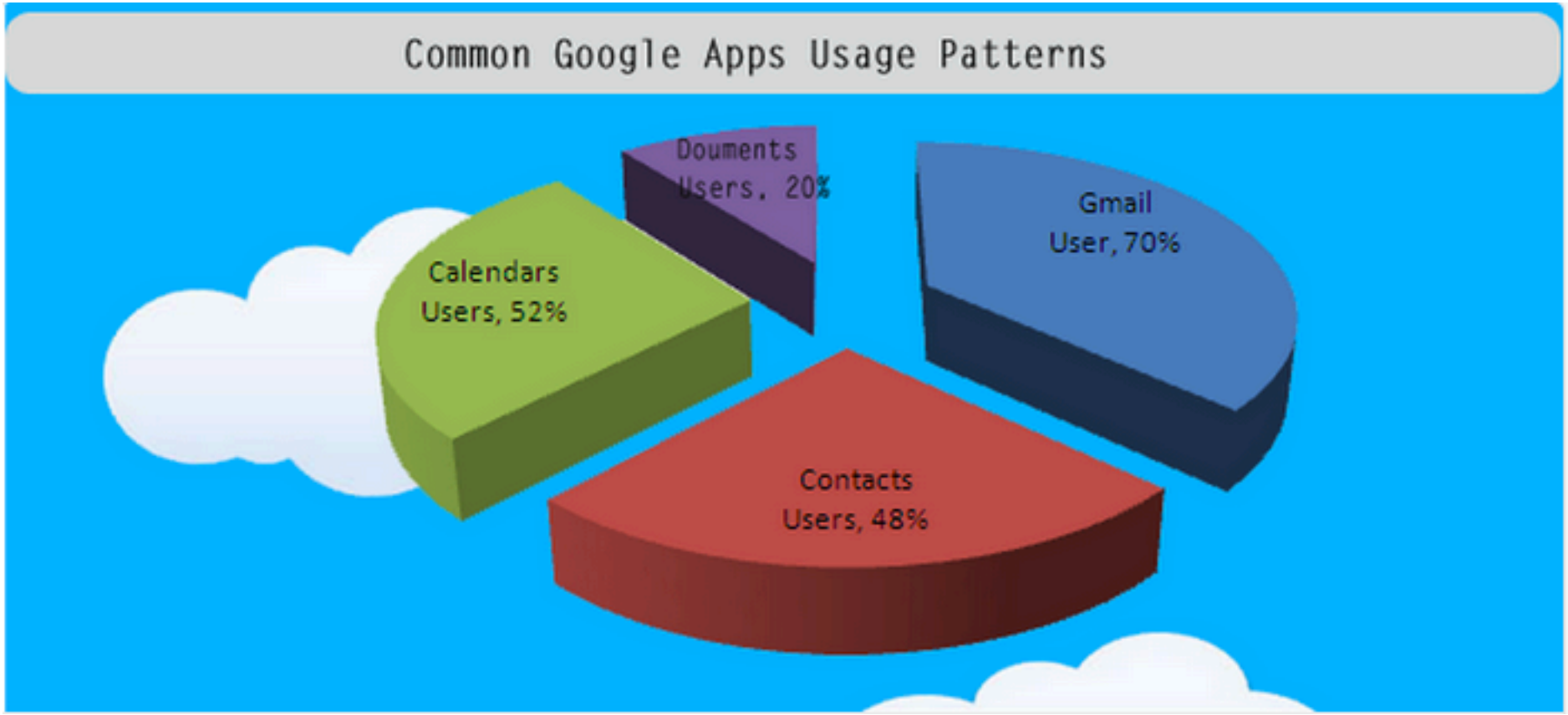
- Courtesy of WTFViz.net

Visualization to Educate?

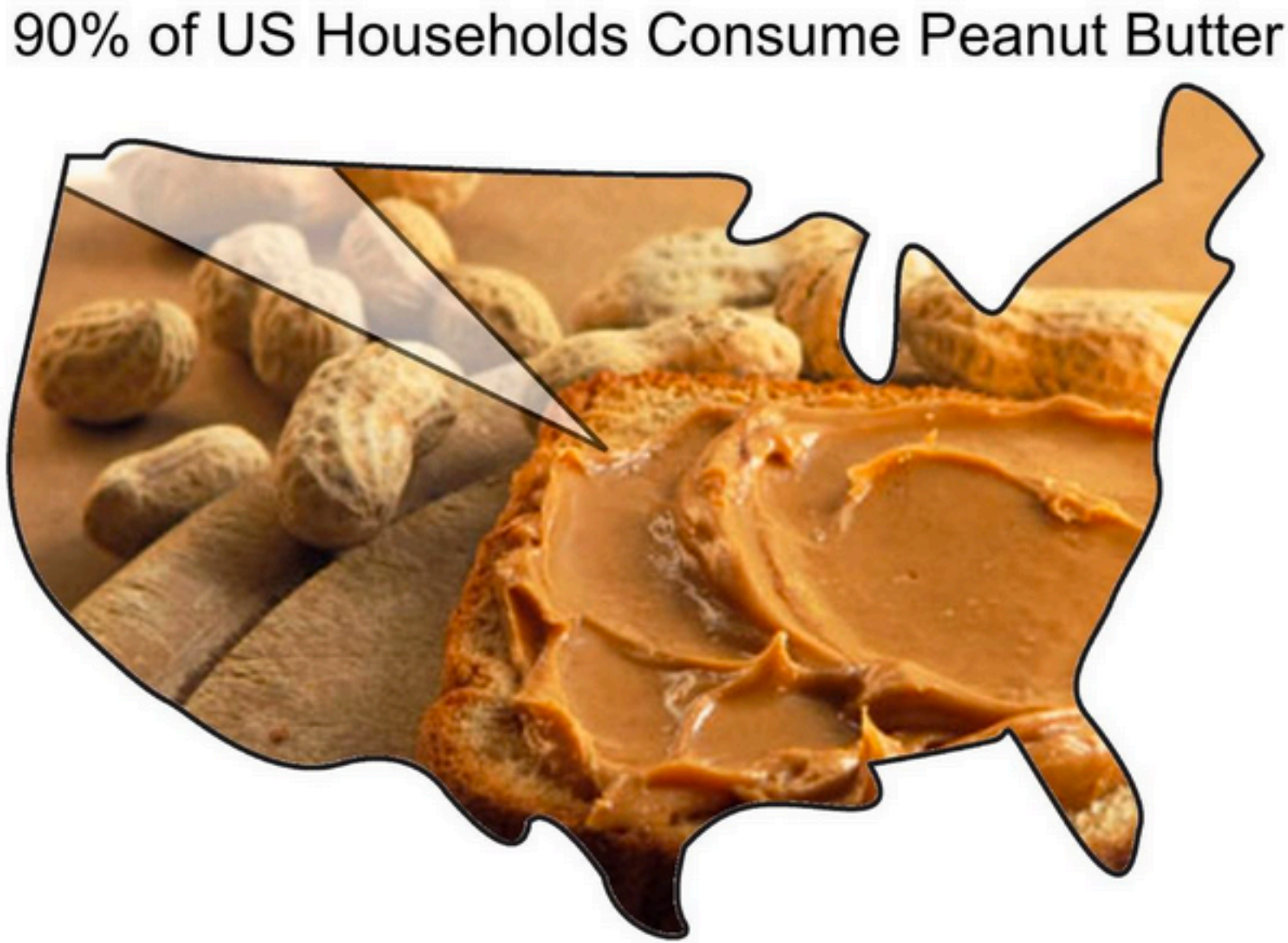


from wtfviz.net

Pie in the Sky?

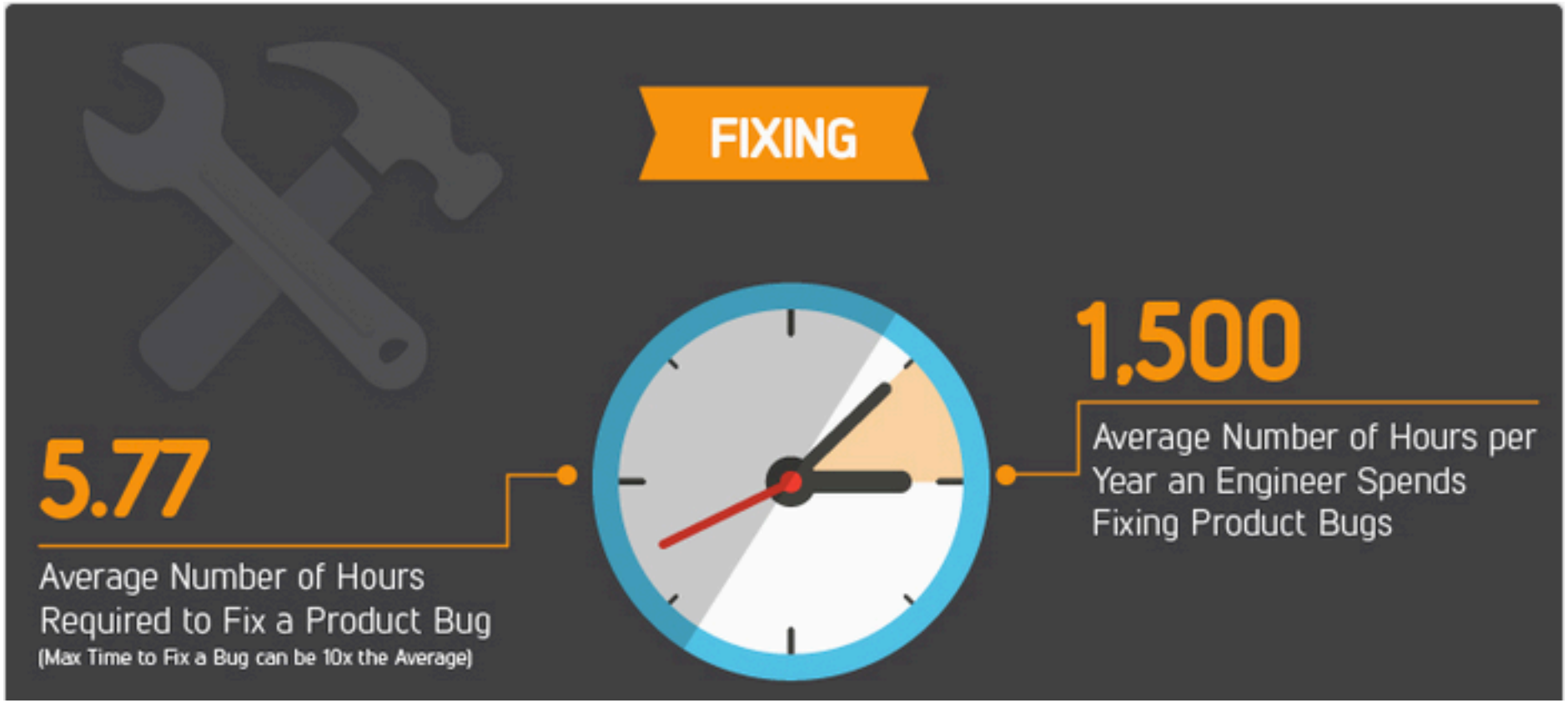


from wtfviz.net



from wtfviz.net

Needs Fixing



from wtfviz.net

Outline

Visualization:

- Some great examples
- Some counter-examples
- Principles for Visualization Design
- Visualization Toolkits preview

Visualization Definitions

- “Transformation of the symbolic into the geometric” [McCormick et al. 1987]

- “... finding the artificial memory that best supports our natural means of perception.” [Bertin 1967]

- “The use of computer-generated, interactive, visual representations of data to amplify cognition.” [Card, Mackinlay, & Shneiderman 1999]

Uses for Data Viz

A: Support reasoning about information (analysis)

- Finding relationships
- Discover structure
- Quantifying values and influences
- Should be part of a query/analyze cycle

B: Inform and persuade others (communication)

- Capture attention, engage
- Tell a story visually
- Focus on certain aspects, and omit others

Uses for Data Viz

Principle 1

- Simplify !

Chart Design: Simplifying

- Example from Tim Bray

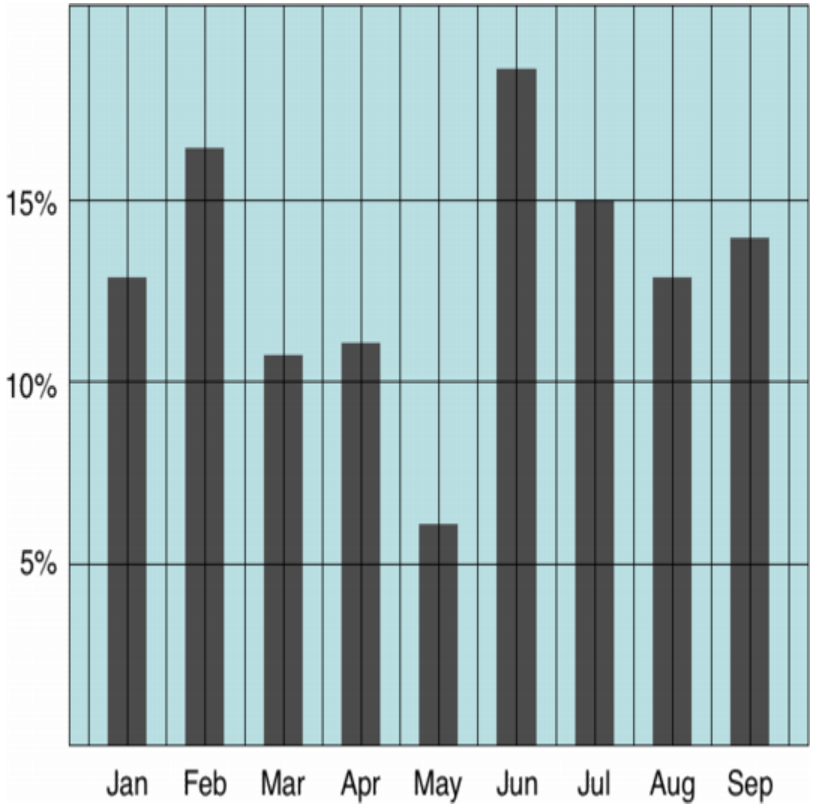


Chart Design: Simplifying

- Example from Tim Bray

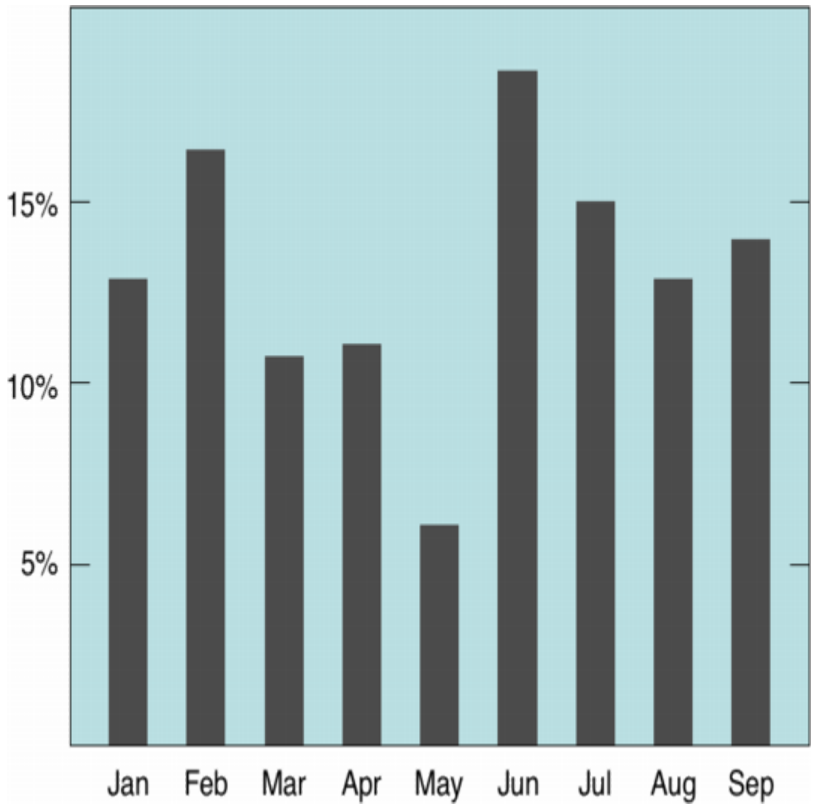


Chart Design: Simplifying

- Example from Tim Bray

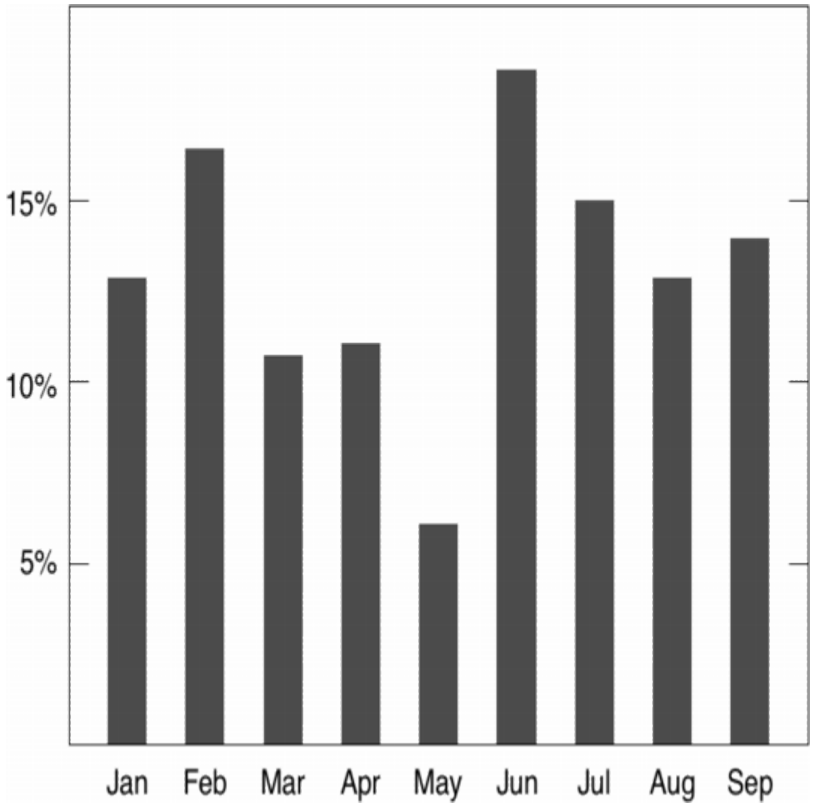


Chart Design: Simplifying

- Example from Tim Bray

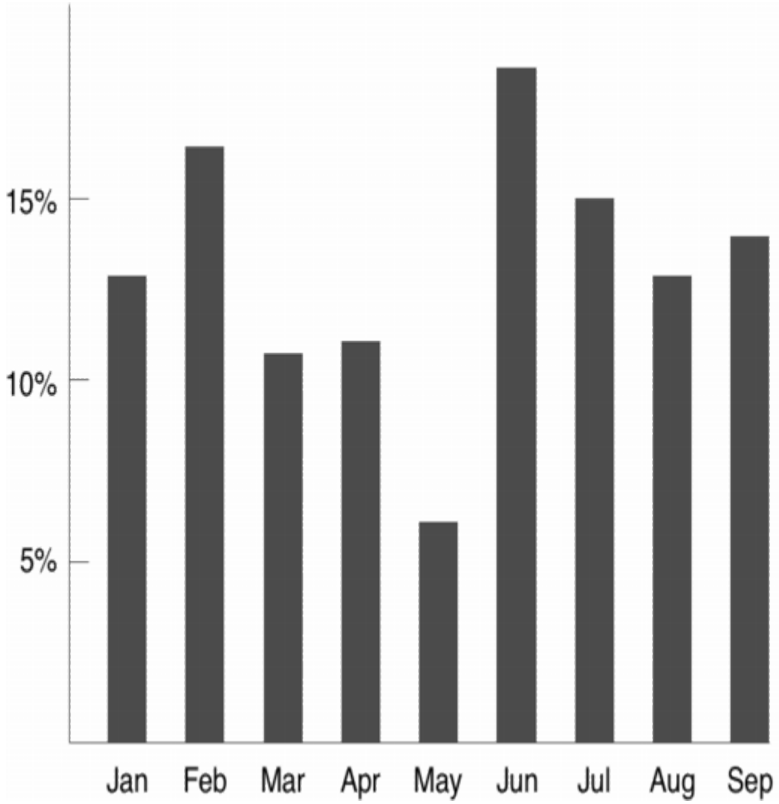


Chart Design: Simplifying

- Example from Tim Bray

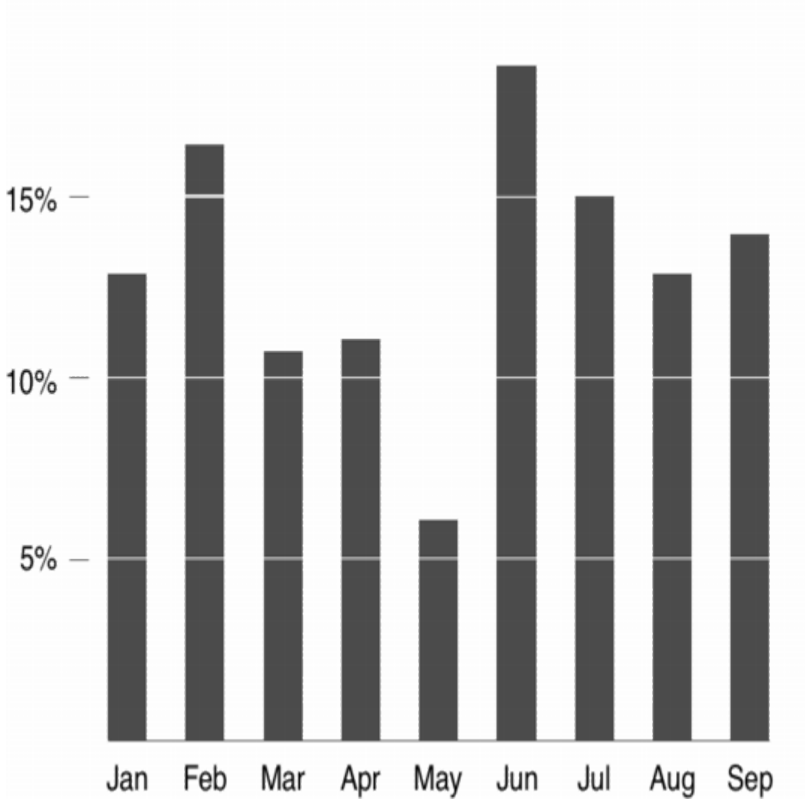
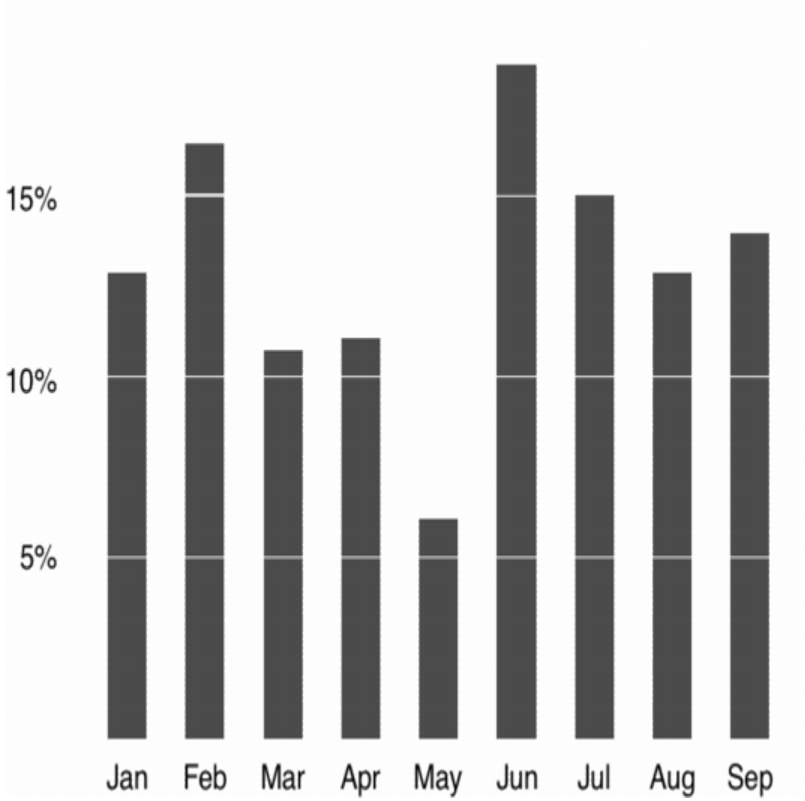


Chart Design: Simplifying

- Example from Tim Bray



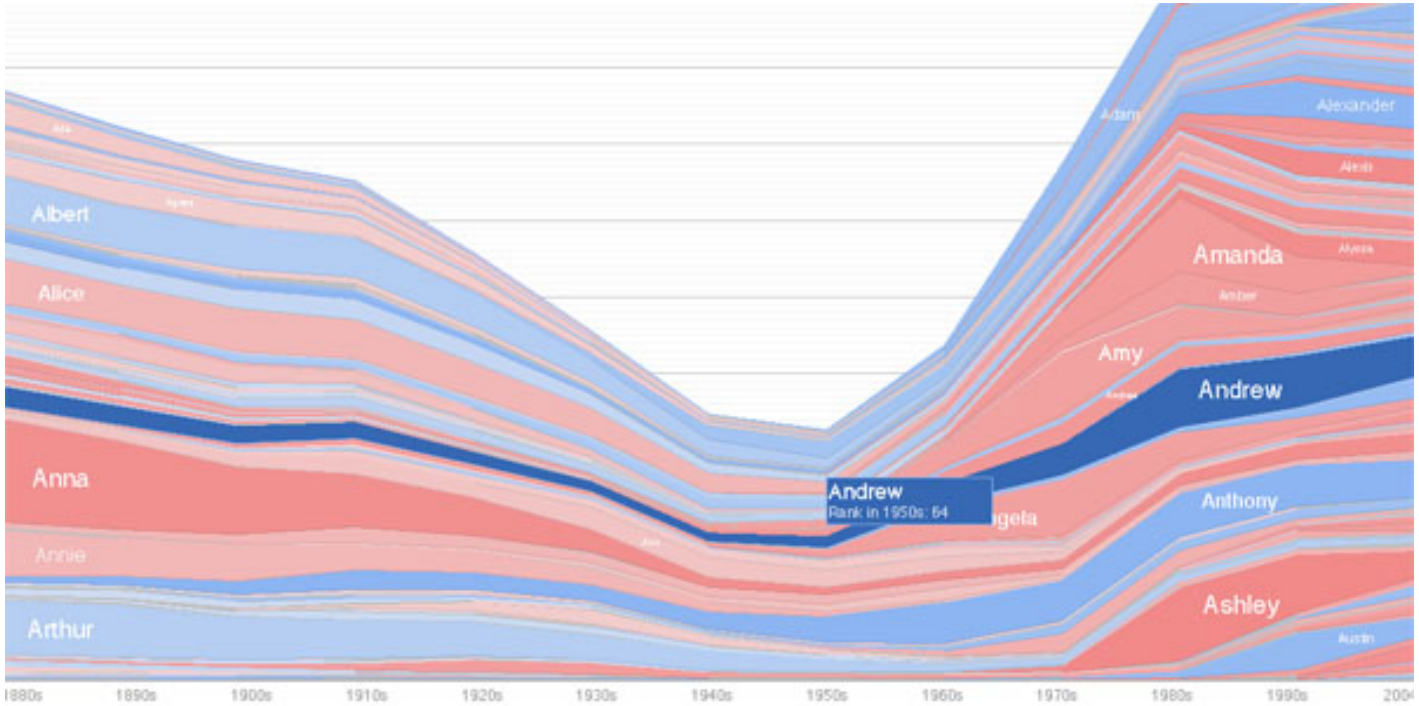
Principle 1: Simplify

- Tables and charts
  - Reduce chartjunk/tablejunk; increase data-ink ratio
  - Lessons from perception: Limit the number of objects displayed at once

- Beware:
  - Gratuitous 3D
  - Shadows
  - Gratuitous animation
- How do you tell if a feature is gratuitous?
  Ask whether using it reveals more information.

Interactive Chart Design: Simplifying

- With interactive charts you can keep things very simple by **hiding** and **dynamically revealing** important structure.
- On an interactive chart, you reveal the information most useful for **navigating** the chart.



Principle 2: Understand Magnitudes

Which is brighter?

Principle 2: Understand Magnitudes

(128, 128, 128) **(144, 144, 144)**

Which is brighter?

Just Noticeable Difference

- JND (Weber's Law)

$$\Delta S = k\frac{\Delta I}{I}$$

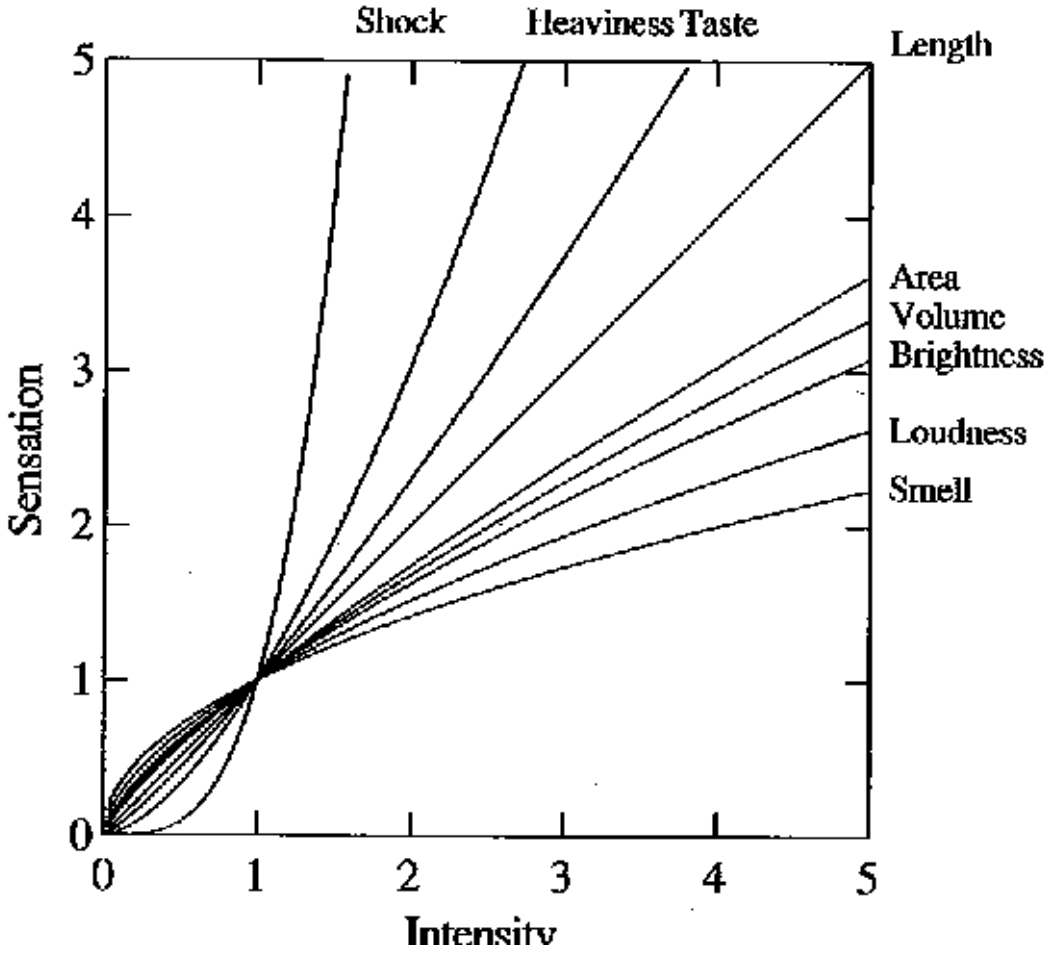
- Ratios more important than magnitude
- Most continuous variations in stimuli are perceived in discrete steps

Steven's Power law

$$S = I^p$$

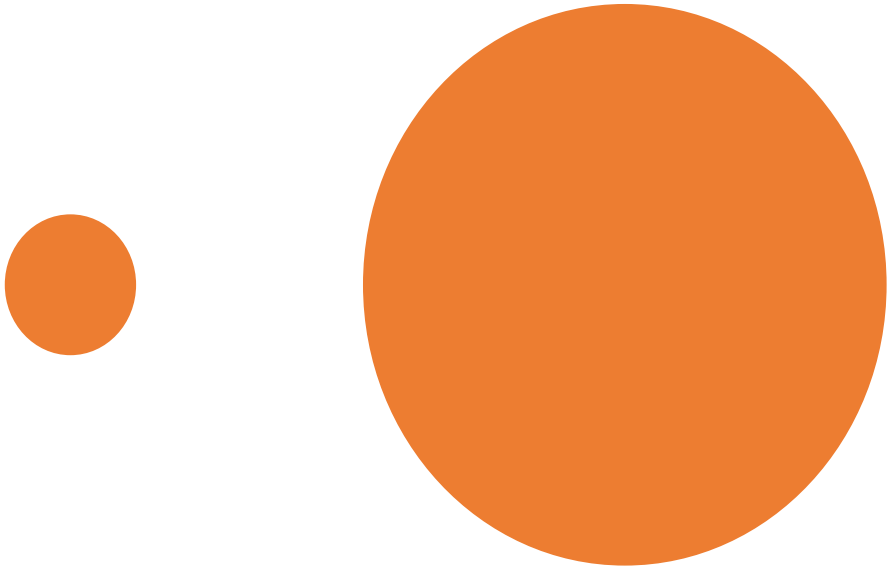
S = sensation
I = intensity

p < 1 : underestimate
p > 1 : overestimate

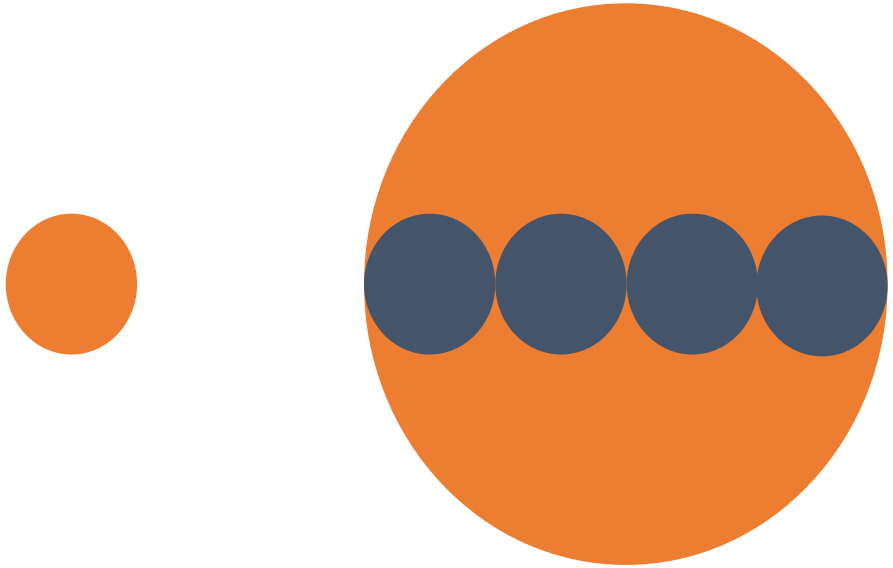


[graph from Wilkinson 99, based on Stevens 61]

[alternate graph : http://www.undergrad.ahs.uwaterloo.ca/~wchedder/stevenspowerlaw.htm]

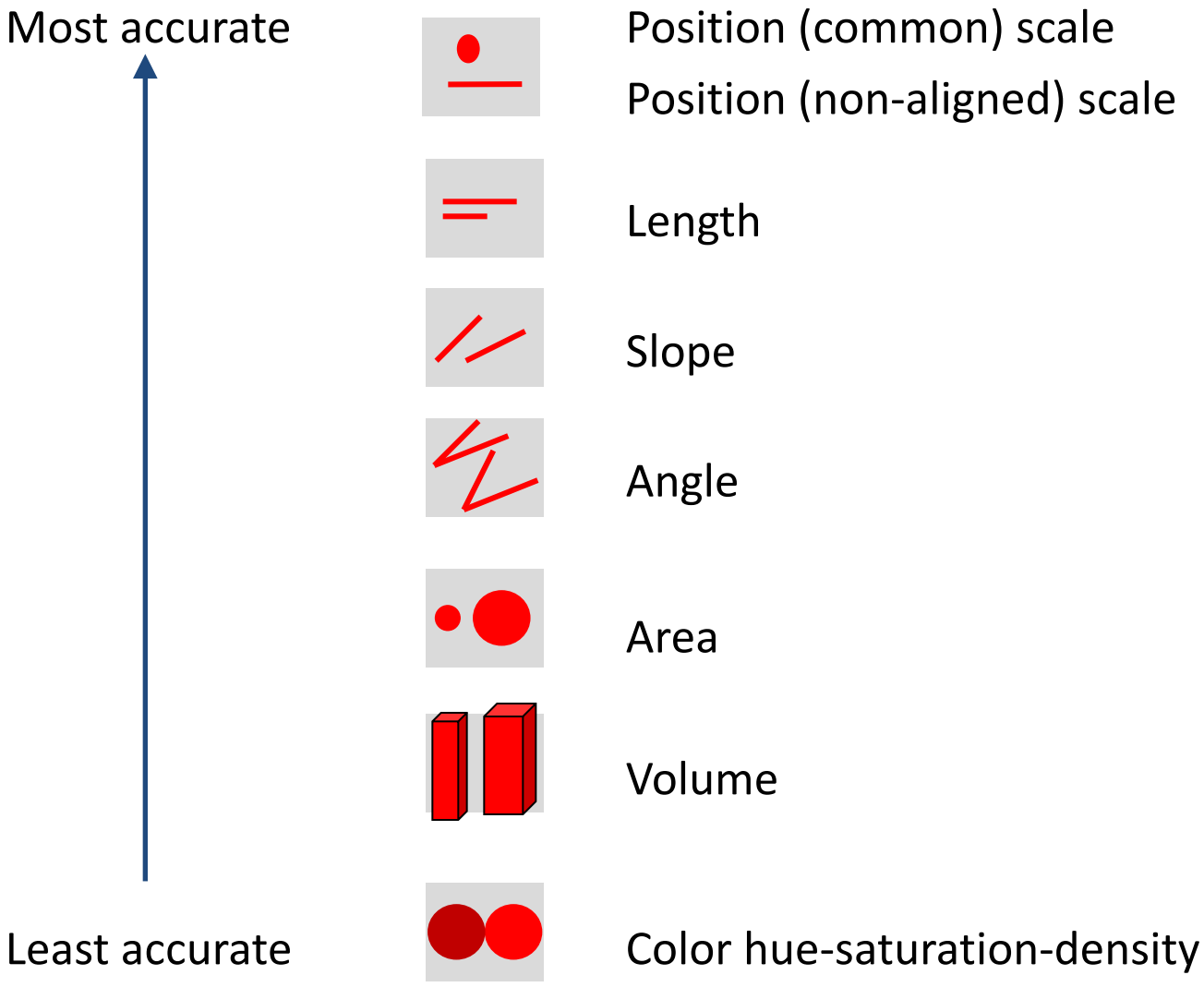


Compare area of circles



Compare area of circles

Principle 2: Understand Magnitudes



Principle 3: Use Color

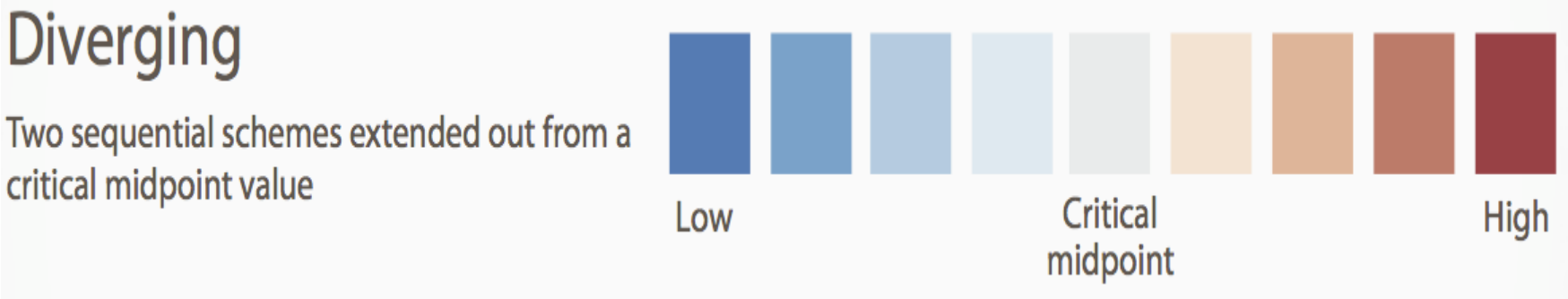
- Color
  - Choose colors based on the information you want to convey
    - Sequential
    - Diverging
    - Categorical
  - Use online resources to discover and record your color schemes
    - Color Brewer
    - Kuler
    - Colour Lovers
  - Where possible, use your organization's palette

Principle 3: Use Color

- Color

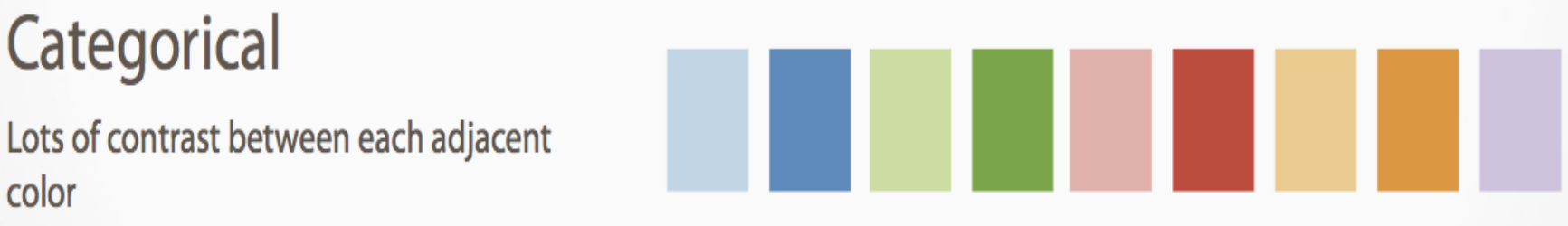
Principle 3: Use Color

- Color



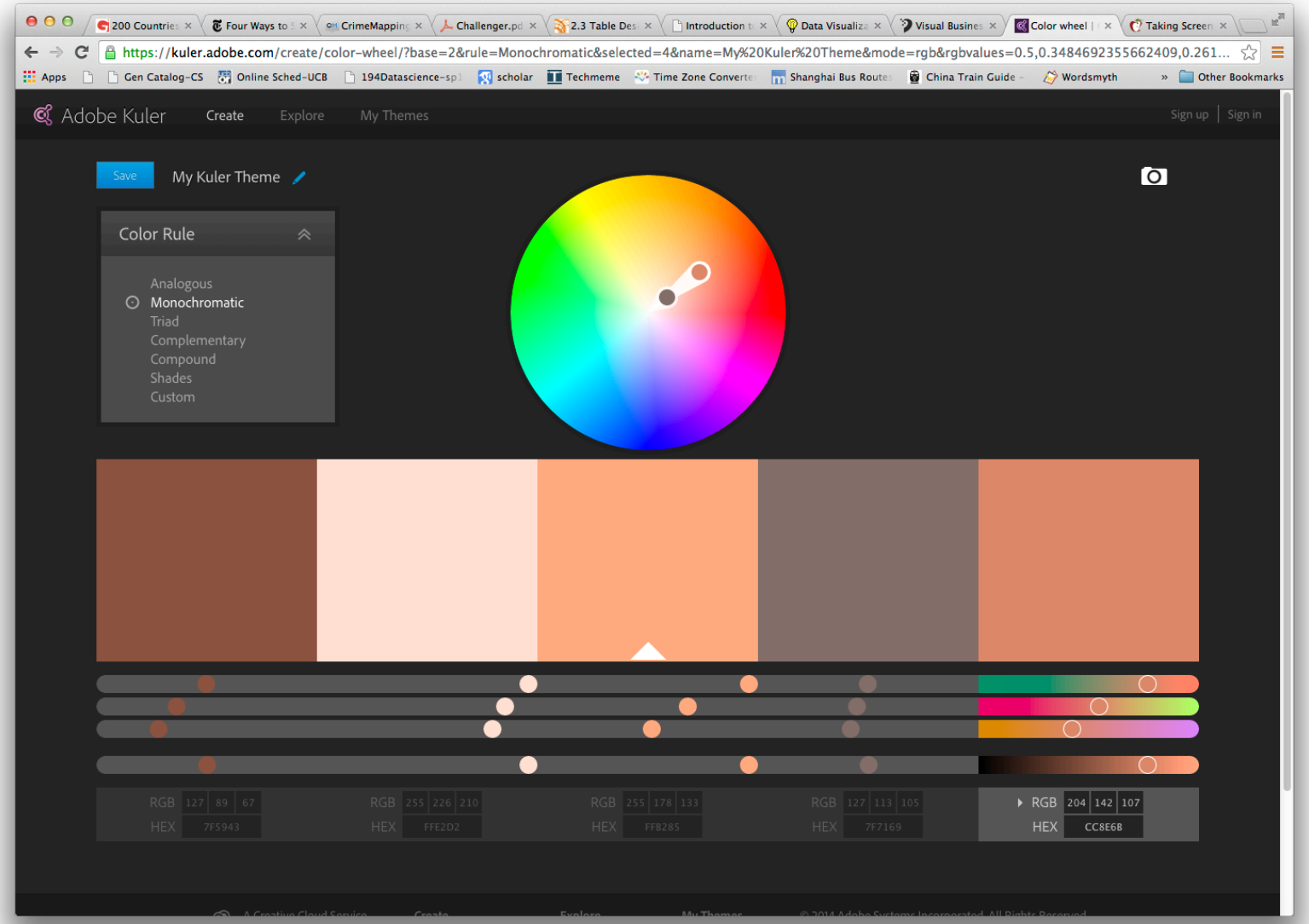
Principle 3: Use Color

- Color



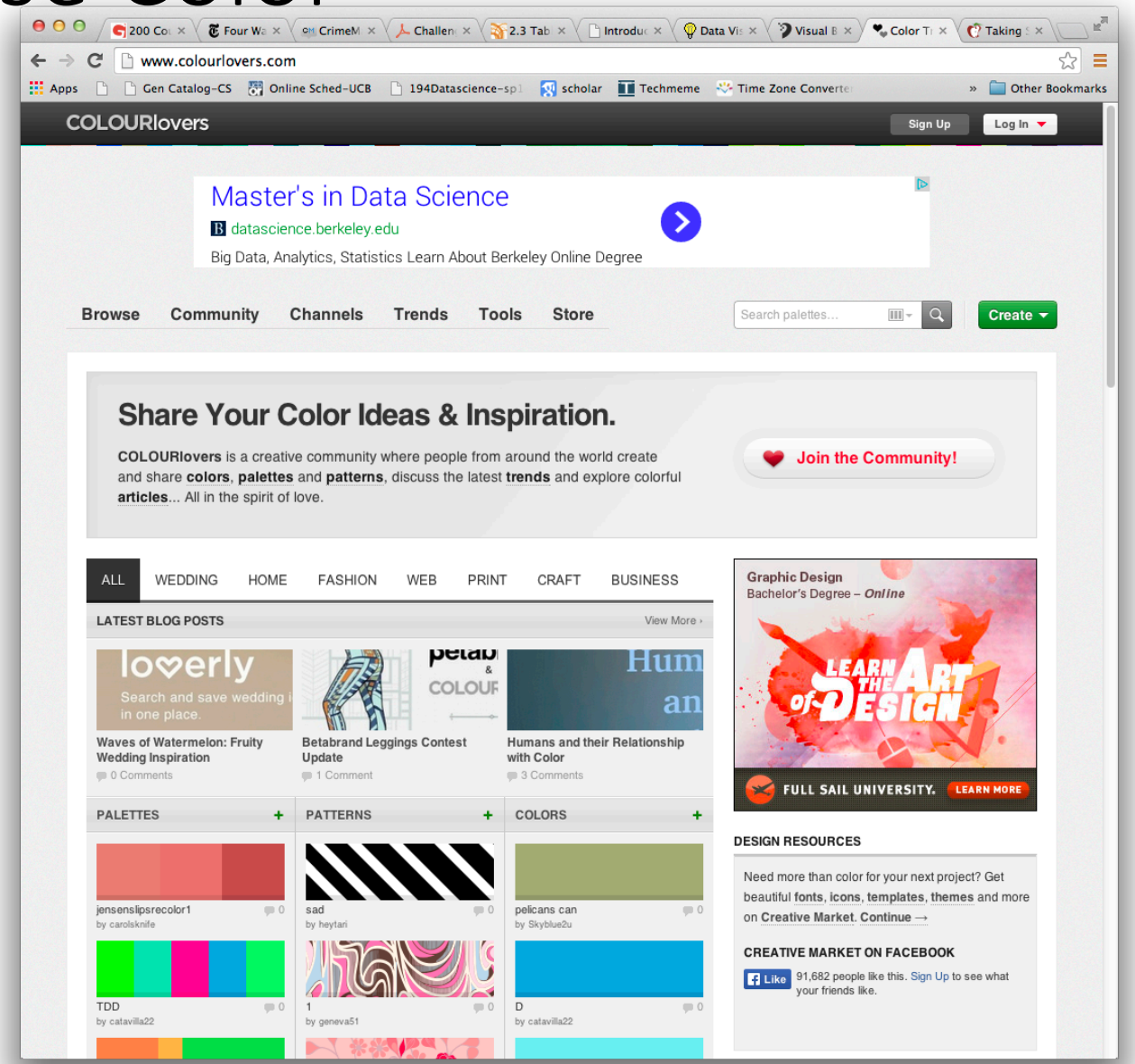
Principle 3: Use Color

- Color



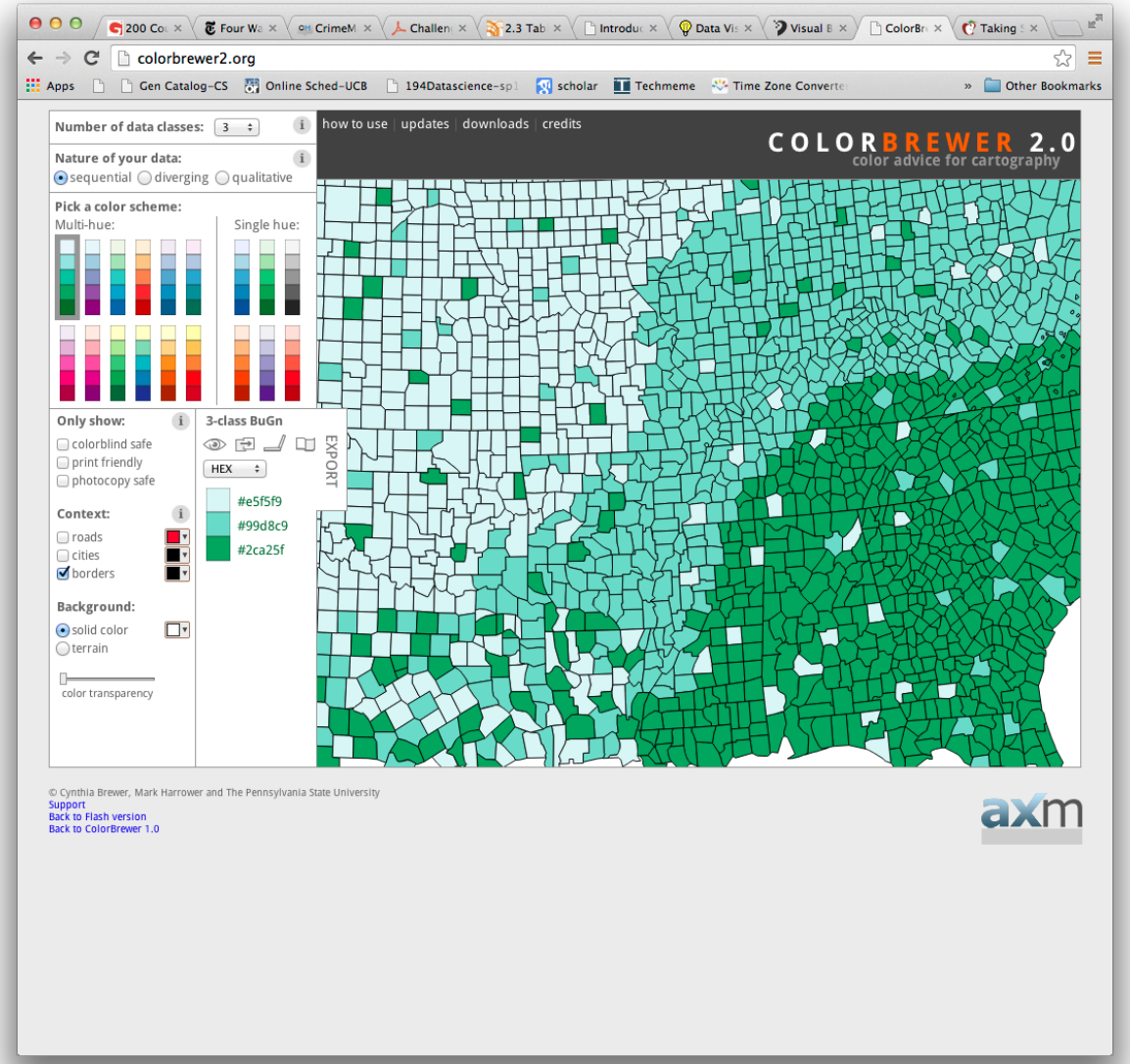
Principle 3: Use Color

- Color



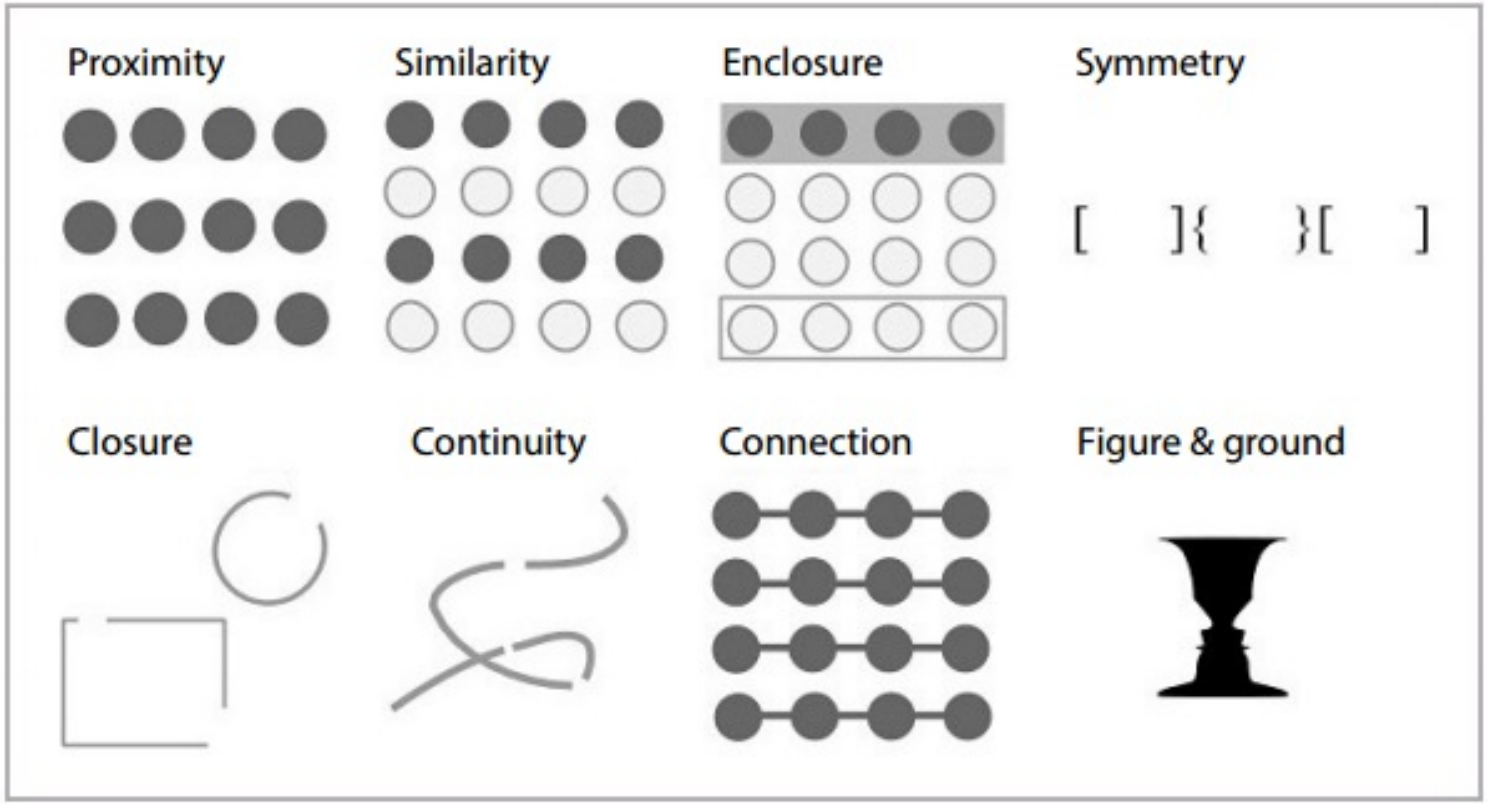
Principle 3: Use Color

- Color



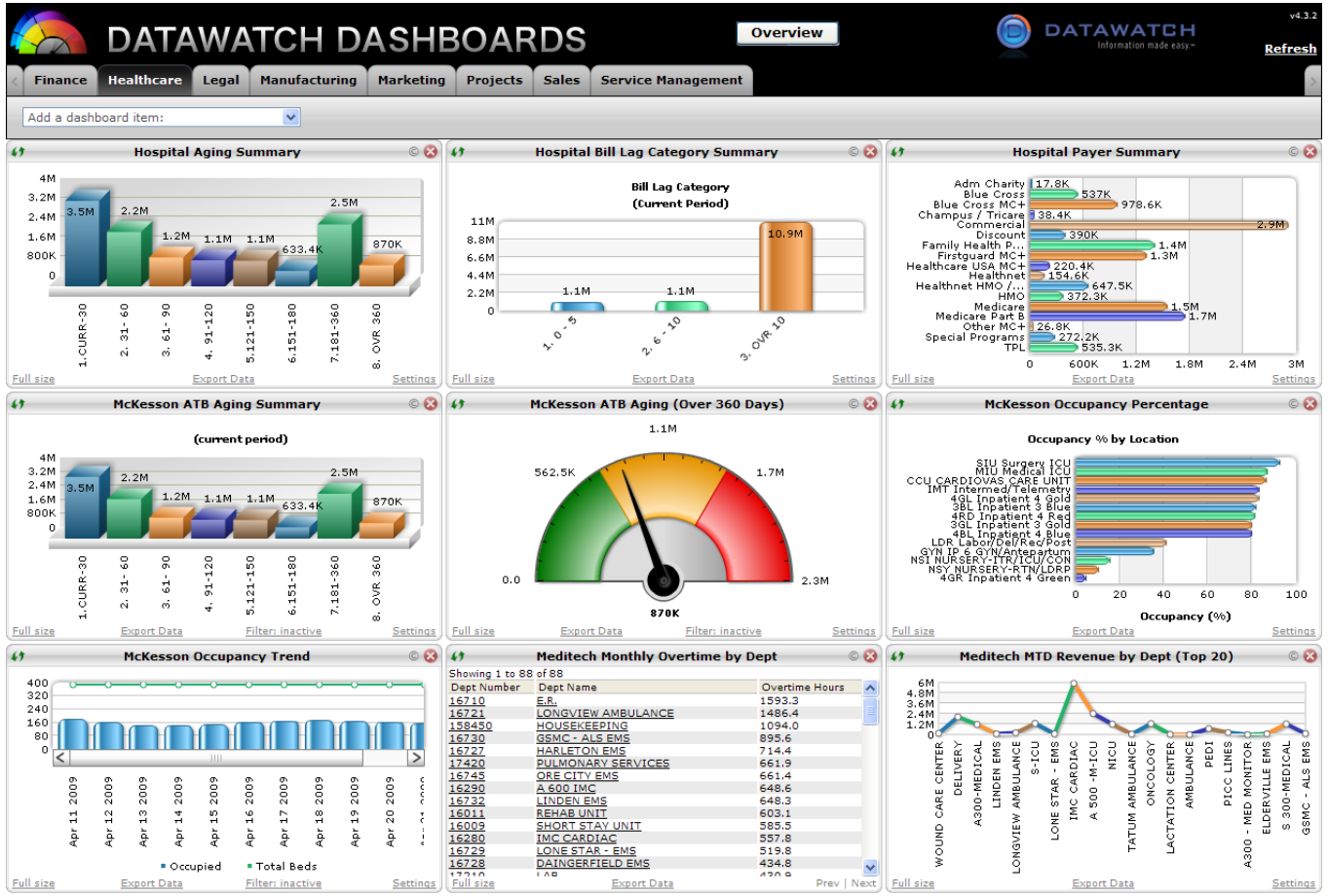
Principle 4: Use Structure

- Gestalt Psychology principles (1912):



Source http://blog.fusioncharts.com/2014/03/how-to-use-the-gestalt-principles-for-visual-storytelling-podv/

Principle 4: Use Structure (but not like this)



Source https://www.vocalabs.com/blog/my-dashboard-pet-peeve

Principle 4: Use Structure

Source https://www.vocalabs.com/blog/my-dashboard-pet-peeve

Chart Selection – Andrew Abela

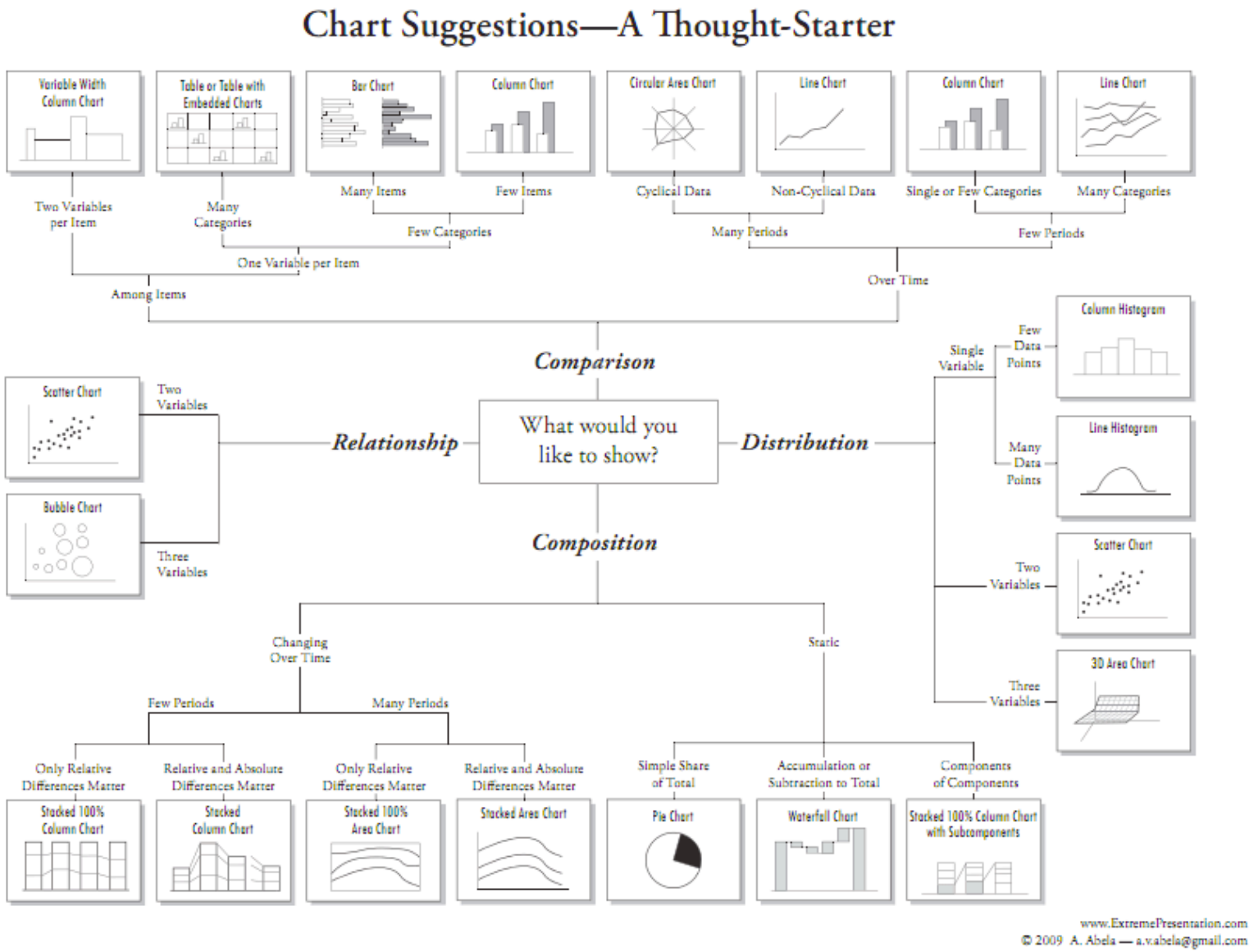
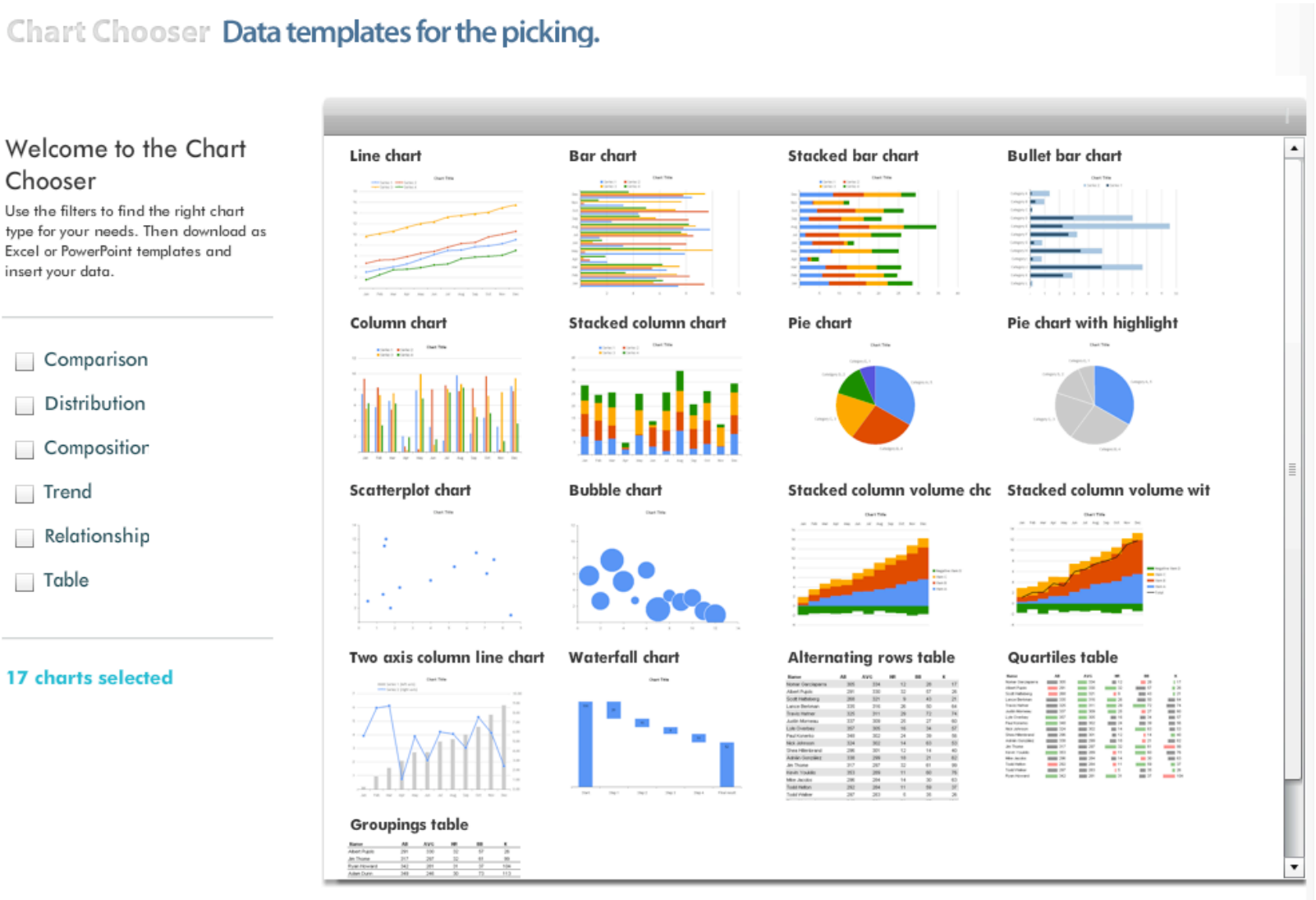
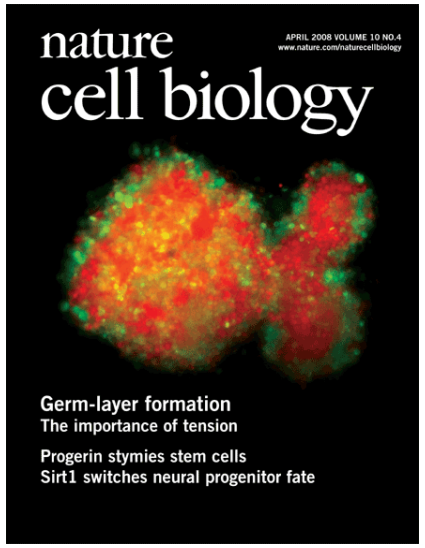
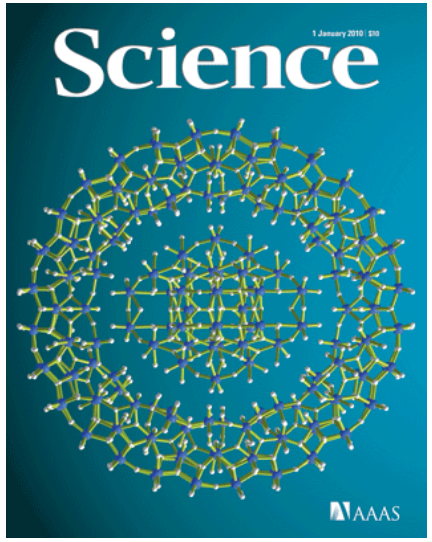


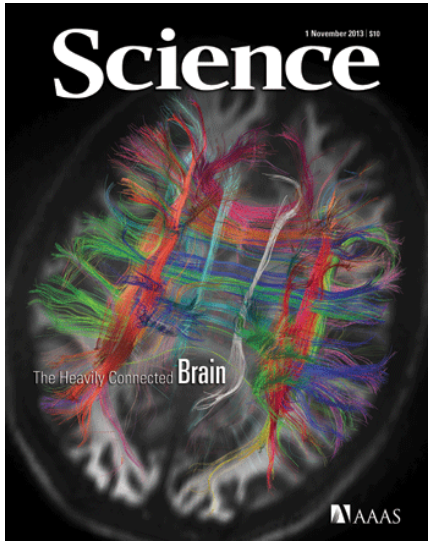
Chart Selection – Juice Analytics

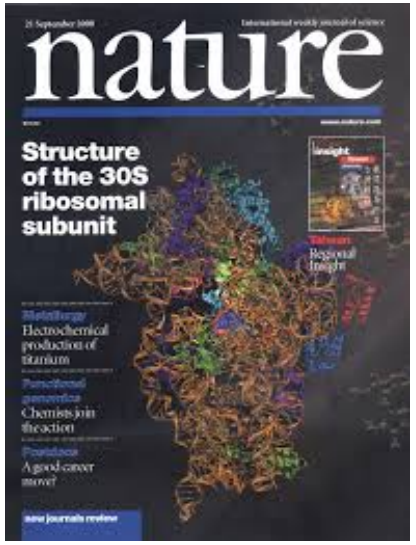


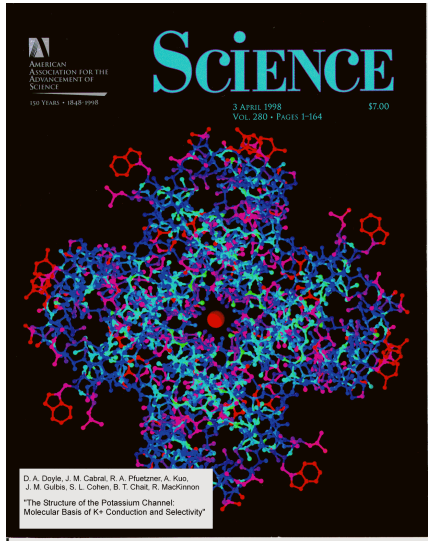
Data Viz in the Sciences











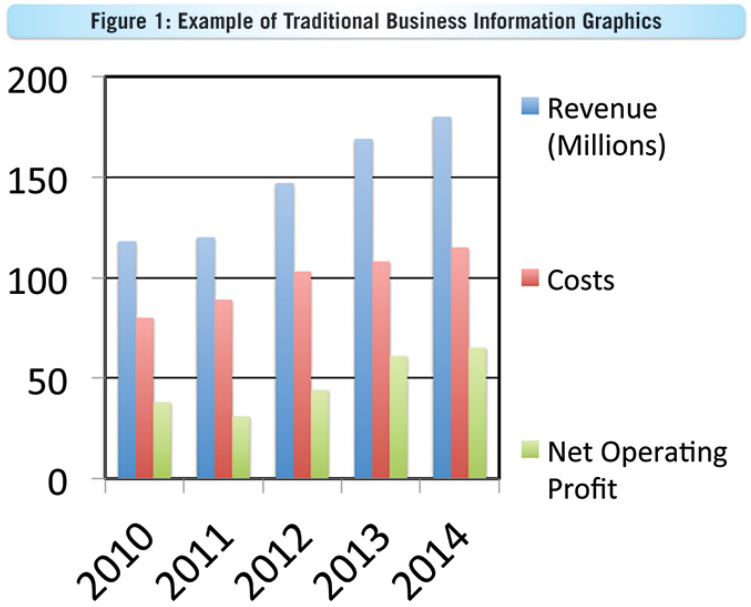
A case for Ugly visualizations

People instinctively gravitate to attractive visualizations, and they have a better chance of getting on the cover of a journal.

But does this conflict with the goals of visualization?:

- Rapid exploration
- Focus on most important details
- Easy and fast to develop and customize

e.g. Powerpoint vs Keynote



Outline

Visualization:

- Some great examples
- Some counter-examples
- Principles for Visualization Design
- Visualization Toolkits preview

Interactive Toolkits: D3

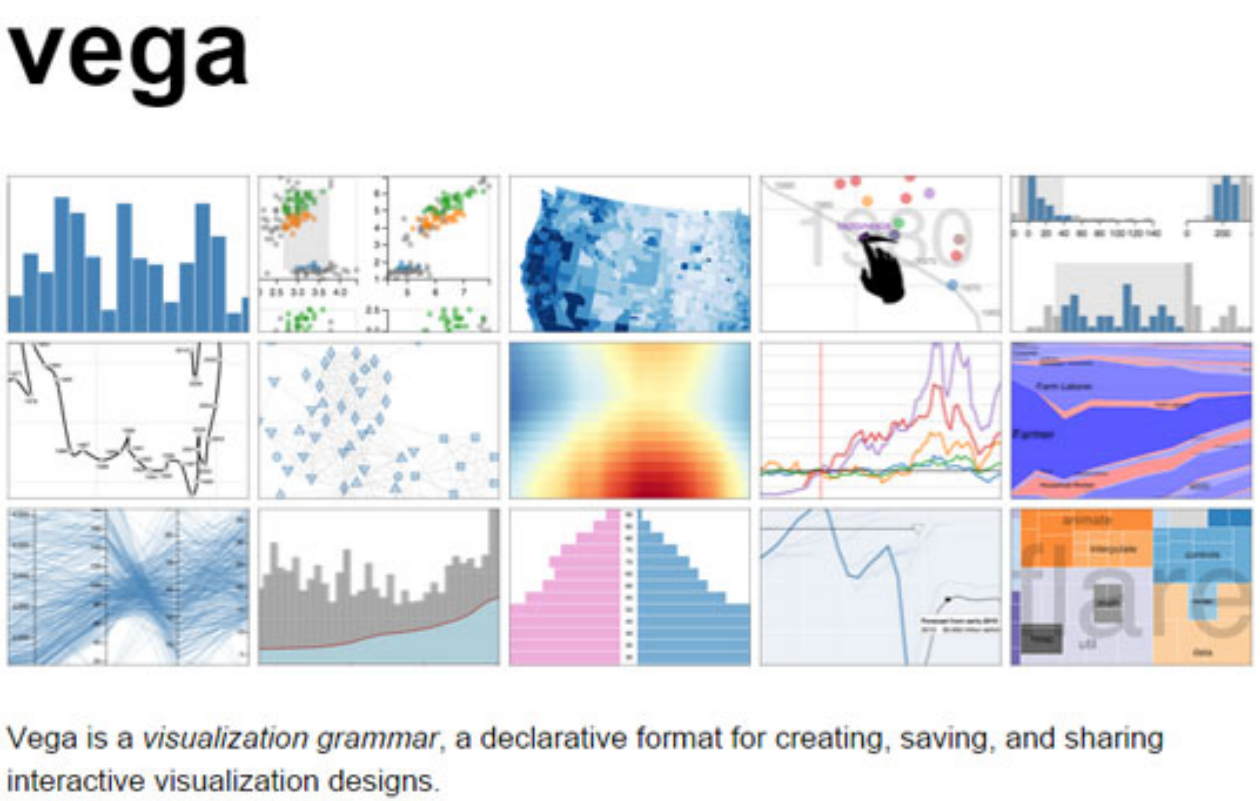
Without Doubt, the most widely used interactive visualization framework is ***D3***, developed around 2011 by Jeff Heer, Mike Bostock and Vadim Ogievetsky.

Note from the authors: *D3 is intentionally a low-level system. During the early design of D3, we even referred to it as a "visualization kernel" rather than a "toolkit" or "framework"*

Interactive Toolkits: Vega

Vega is a “visualization grammar” developed on top of d3.js

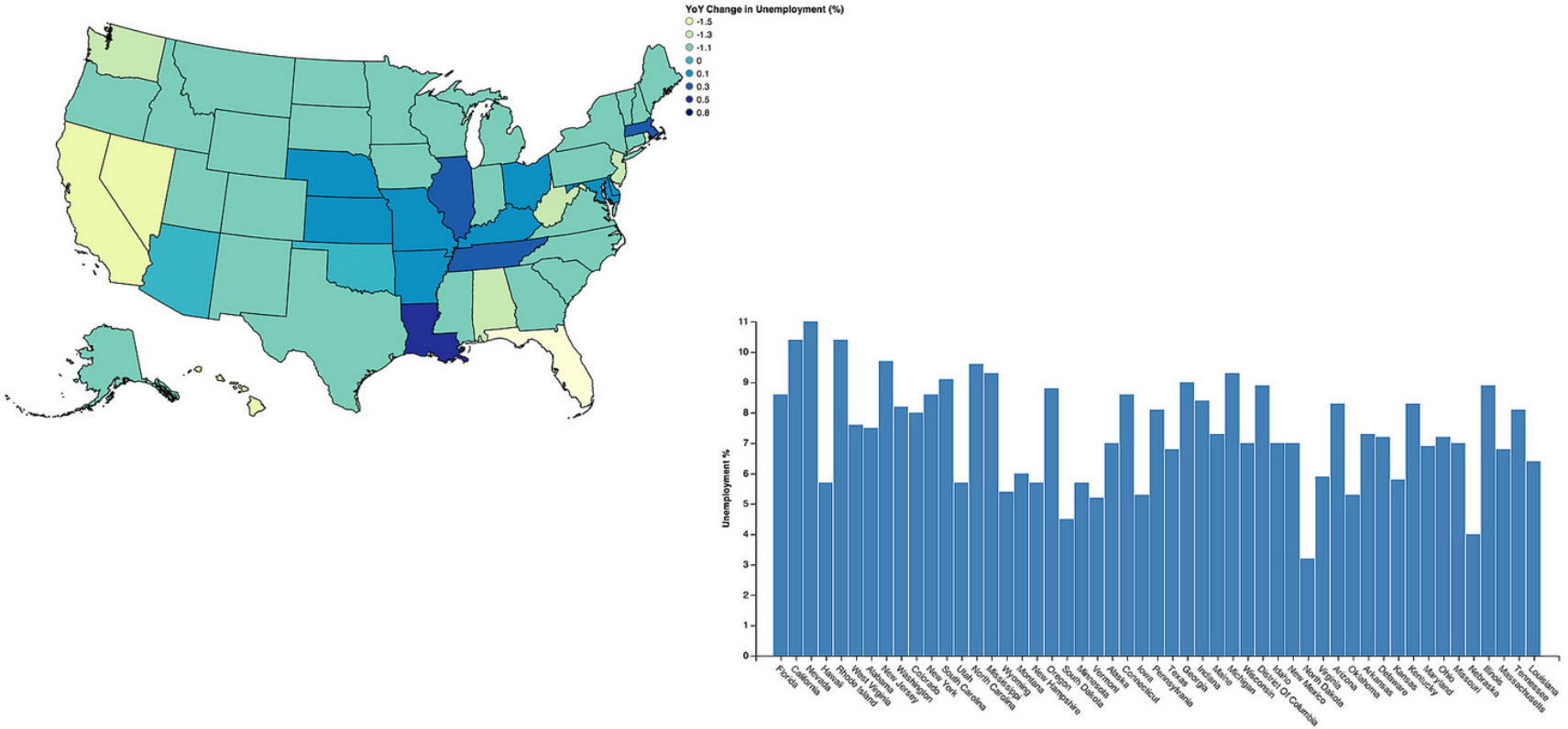
It specifies graphics in JSON format.



Interactive Toolkits: Vincent

Vincent is a Python-to-Vega translator.

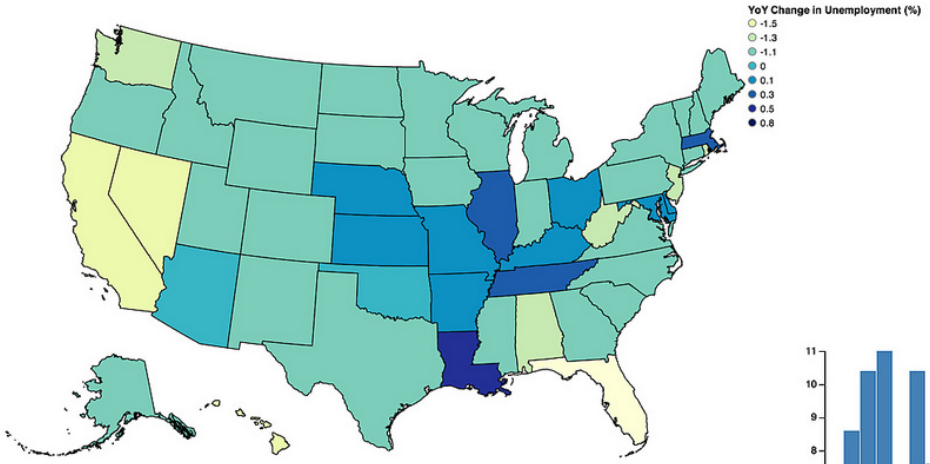
Trivia question: why is it called Vincent? Hint: Vincent+Vega= ?

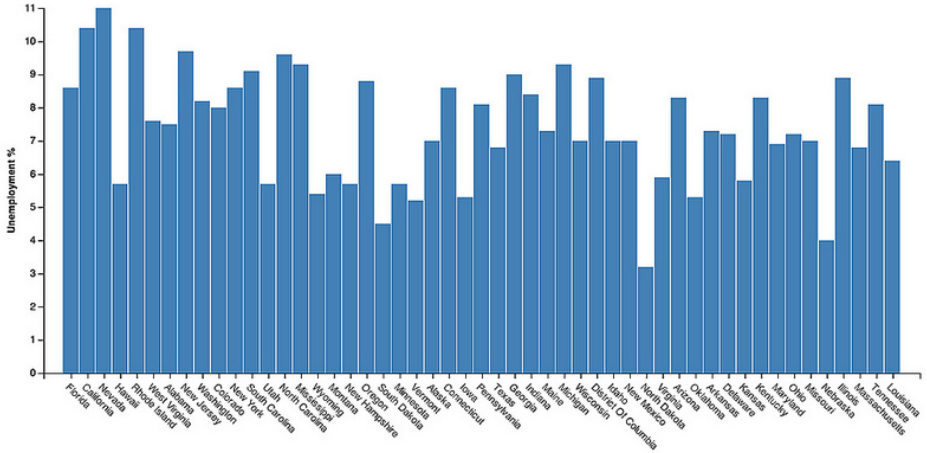


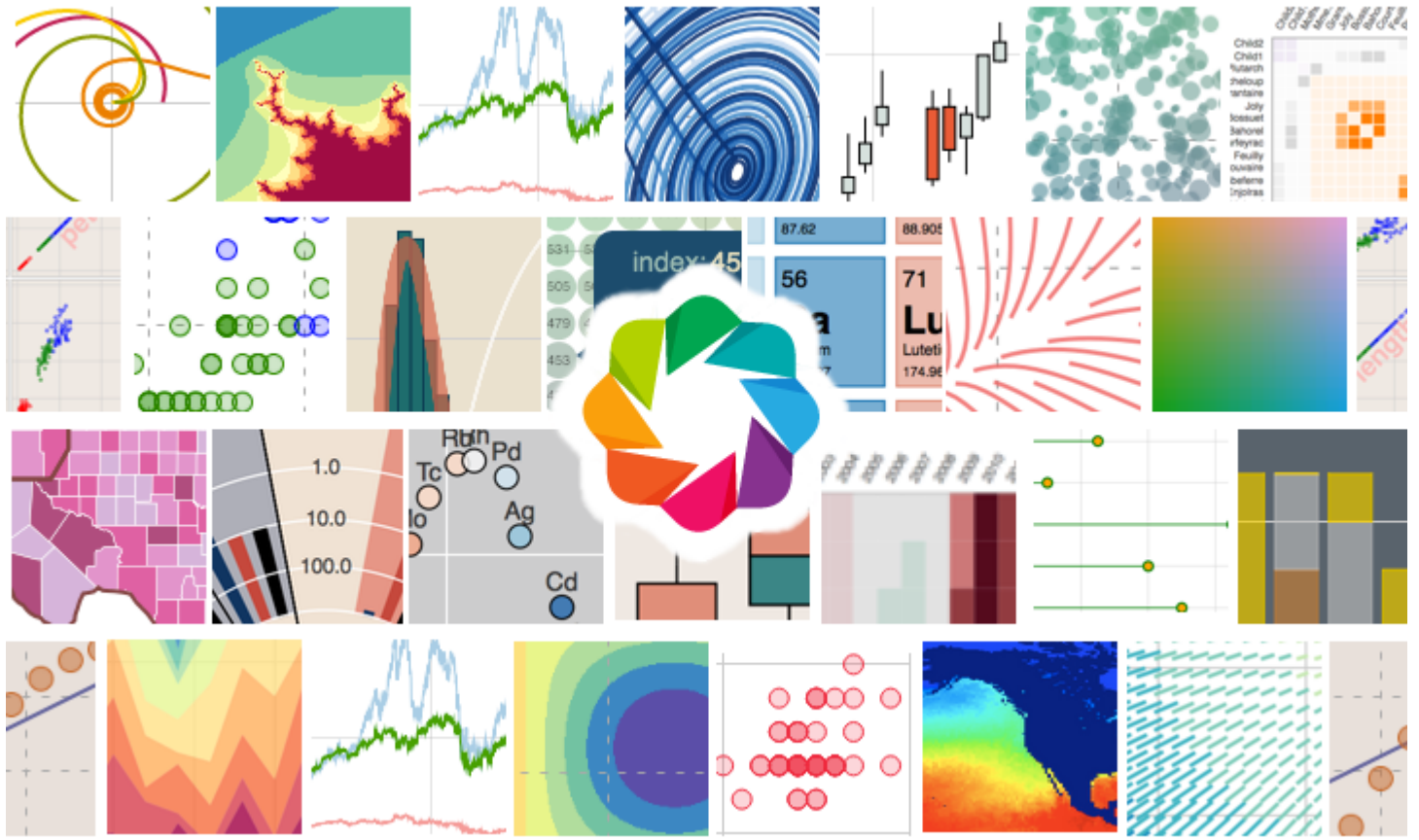
Interactive Toolkits: Vincent

Vincent is a Python-to-Vega translator.

Trivia question: why is it called Vincent? Hint: Vincent+Vega= ?





Bokeh: Another Interactive Viz Library

Bokeh is an independent Viz library focused more heavily on big data visualization. Has both Python and Scala bindings.