Visualization Design & Memorable Chart Junk

Announcements

- Lecture videos will be posted to the YouTube playlist listed on the course page
- Slides will be available before lecture
- Paper discussions will require A/V if possible.
- My contact info will soon be updated on the course site

Today

- "Good" Design (30 min)
  - Photography tips
  - Principles of Effective Website Design
  - Principles of Good User Interface Design
  - Examples of Good (Bad) Visualization Design
- Today's Readings
- Worksheet: Time-based Data & Simple Charts (20 min)
- Graph Drawing Preview: Terminology & Goals (15 min)
- Readings for Tuesday: Graph Drawing

www.kodak.com/global/en/corp/top10tips/index.jhtml

Kodak

Top Ten Tips

1. Get down on their level

2. Use a plain background

3. Use flash outdoors

4. Move in close

5. Take some vertical pictures

6. Lock the focus

7. Move it from the middle

8. Know your flash's range

9. Watch the light

10. Be a picture director

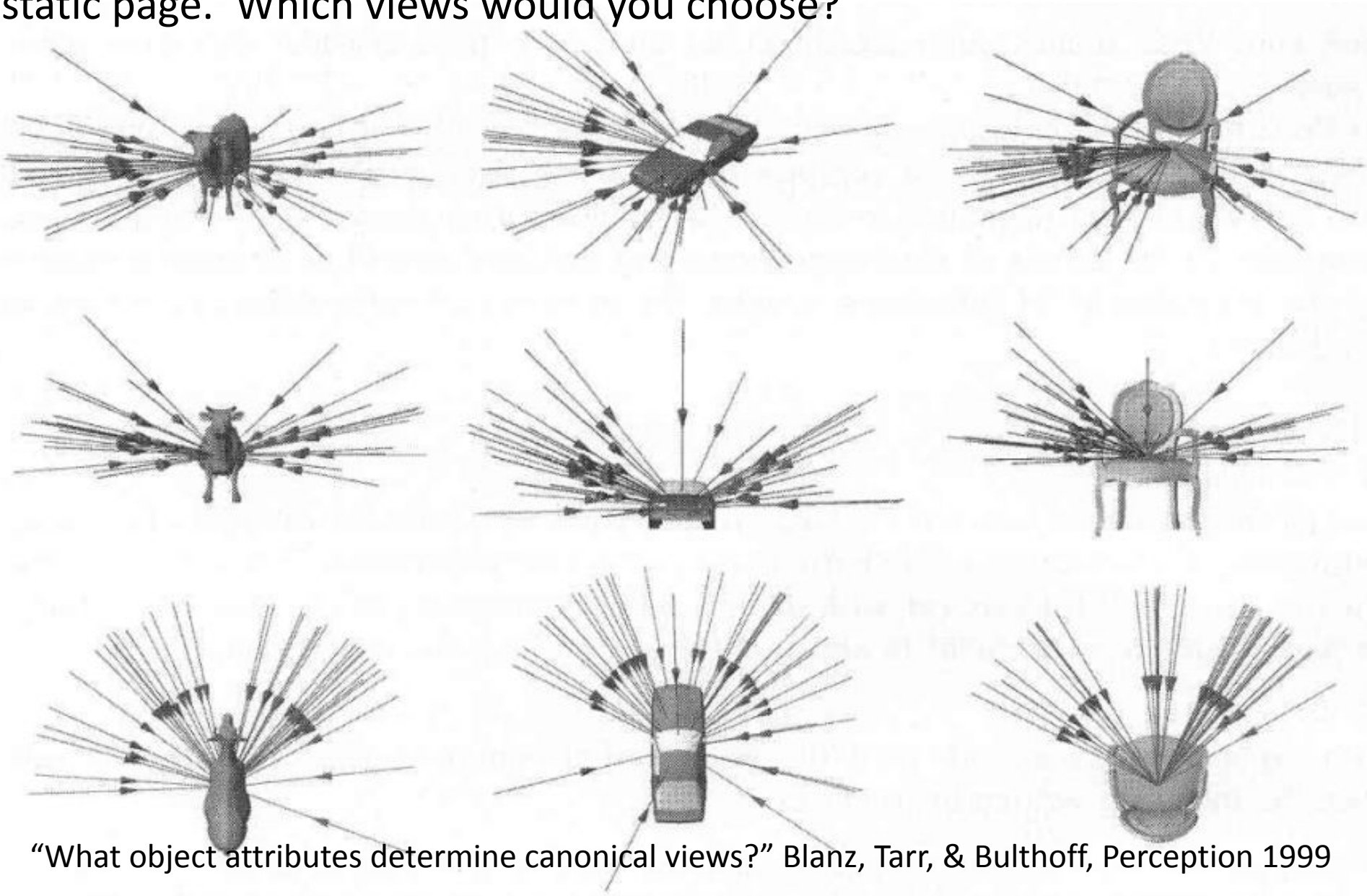
Tips & Projects Center ▶

“Canonical” Viewpoints

- From Dictionary.com:
  - authorized; recognized; accepted
  - the body of rules, principles, or standards accepted as axiomatic and universally binding in a field of study or art: the neoclassical canon
  - a fundamental principle or general rule: the canons of good behavior
  - a standard; criterion: the canons of taste

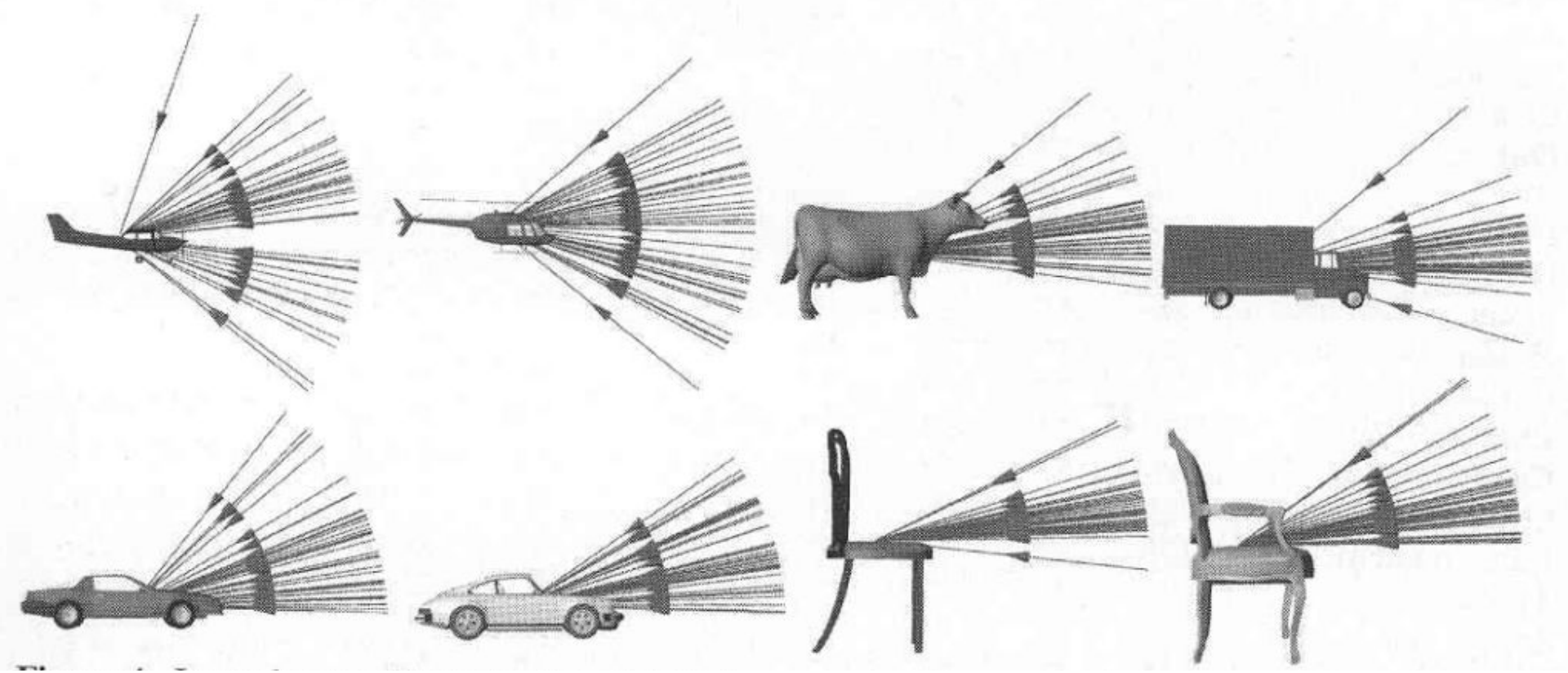
“What object attributes determine canonical views?” Blanz, Tarr, & Bulthoff, Perception 1999

Suppose you were making a brochure and you tried to give your customers the best possible impression of the objects shown on the static page. Which views would you choose?



"What object attributes determine canonical views?" Blanz, Tarr, & Bulthoff, Perception 1999

- Salience and significance of the features
- Stability of viewpoint to small transformations
- Minimize number of occluded features
- Familiarity, Functionality, Aesthetic criteria



"What object attributes determine canonical views?" Blanz, Tarr, & Bulthoff, Perception 1999

1a, 2

1a

1a

1b

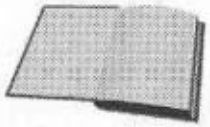
1b

2

1a, 2

1a, 2

1b



1b

2

1b

1a, 2

1b

1a, 2

1b

1b

1b

1b

1a, 2

1a

1b

1b

1a, 2

1b

1b

1a, 2

1b

1a, 2

1a

1b

1a, 1b

Rule of Thirds

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

- align subject with guide lines and intersection points, discourage placement of the subject at the center
- placing the horizon on the top or bottom line, avoid dividing picture in half

Today

- "Good" Design (30 min)
  - Photography tips
  - Principles of Effective Website Design
  - Principles of Good User Interface Design
  - Examples of Good (Bad) Visualization Design
- Today's Readings (30 min)
- Worksheet: Time-based Data & Simple Charts (20 min)
- Graph Drawing Preview: Terminology & Goals (15 min)
- Readings for Tuesday: Graph Drawing

Principles of Effective Website Design

- Guiding the eye (position, color, contrast, size, design elements)
- Spacing, padding, white/empty space, reduce cognitive load
- Navigation/orientation
- Typography (font, size, color, paragraphs)
- Usability/standards/conventions be obvious, "Don't make users think"
- Consistency
- Alignment, polished, simplicity
- Effective writing
- Clarity, sharpness, contrast, exaggeration

http://psd.tutsplus.com/tutorials/designing-tutorials/9-essential-principles-for-good-web-design/
http://uxdesign.smashingmagazine.com/2008/01/31/10-principles-of-effective-web-design/

Principles of Good User Interface Design

- Consistency and standards
  - Match real world: words, phrases and concepts familiar to the user, real-world conventions, natural and logical order, coherency
- Flexibility and efficiency of use: cater/tailor to both inexperienced and experienced users
  - Know your user, user testing, listen to the user
- User control and freedom: a clearly marked "emergency exit" to leave the unwanted, support undo and redo
- Aesthetic and minimalist design: every extra unit of information competes with and diminishes visibility of relevant information
  - System status: keep users informed
- Recognize, diagnose, and recover from errors
  - Error prevention: good error messages, eliminate error-prone conditions, confirmation option
- Help and documentation
  - Recognition rather than recall: information/instructions should be visible or easily retrievable

http://www.sylvantech.com/~talin/projects/ui_design.html

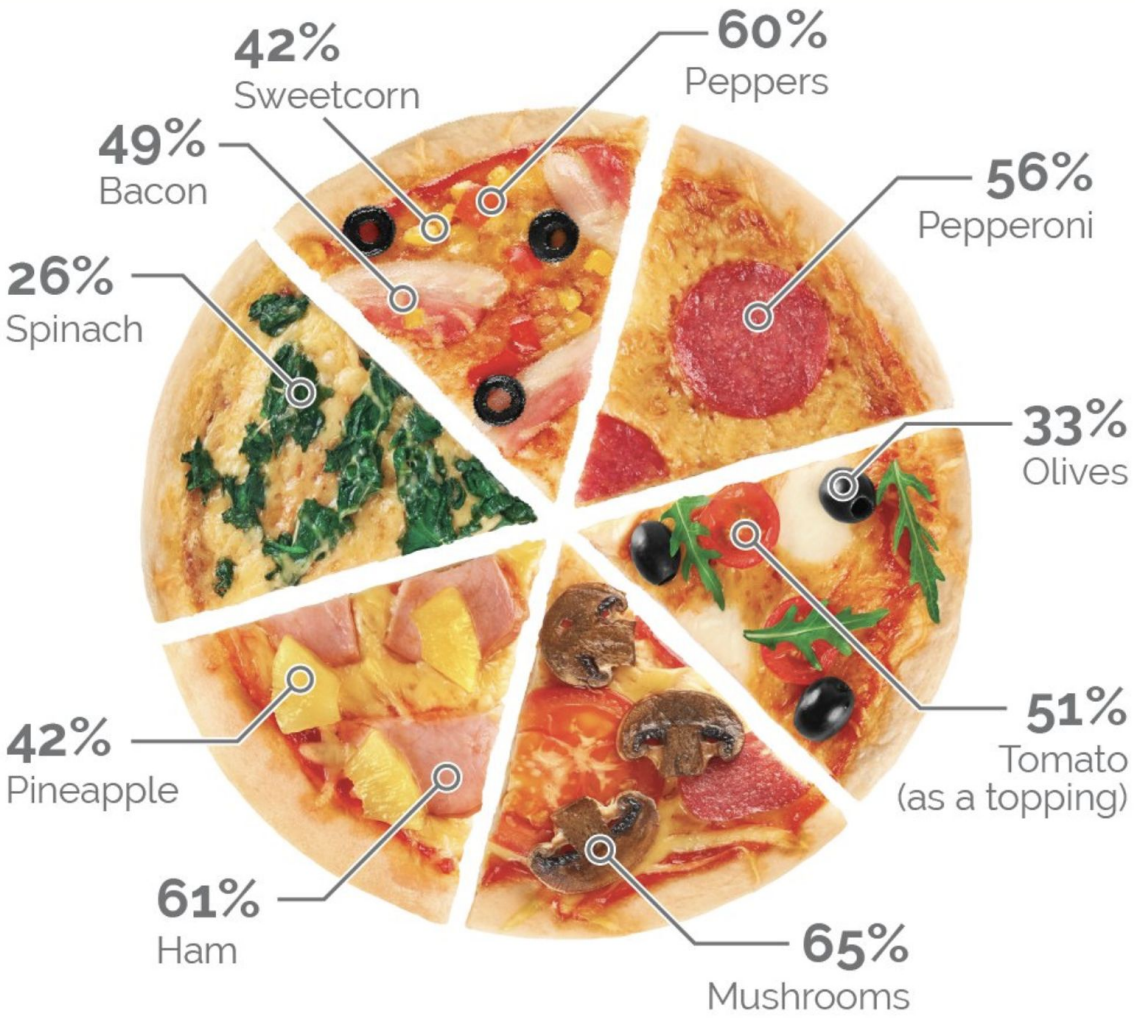
http://www.useit.com/papers/heuristic/heuristic_list.html

Today

- "Good" Design (30 min)
  - Photography tips
  - Principles of Effective Website Design
  - Principles of Good User Interface Design
  - Examples of Good (Bad) Visualization Design
- Today's Reading (30 min): "Useful Junk? The Effects of Visual Embellishment on Comprehension and Memorability of Charts"
- Worksheet: Time-based Data & Simple Charts (20 min)
- Graph Drawing Preview: Terminology & Goals (15 min)
- Readings for Tuesday: Graph Drawing

Mushroom is the UK's most liked pizza topping

Generally speaking, which of the following toppings do you like on a pizza? Select as many as you like



Other items not depicted include: onions (62%), chicken (56%), beef (36%), chillies (31%), jalapeños (30%), pork (25%), tuna (22%), anchovies (18%). 2% of people say they only like Margherita pizzas

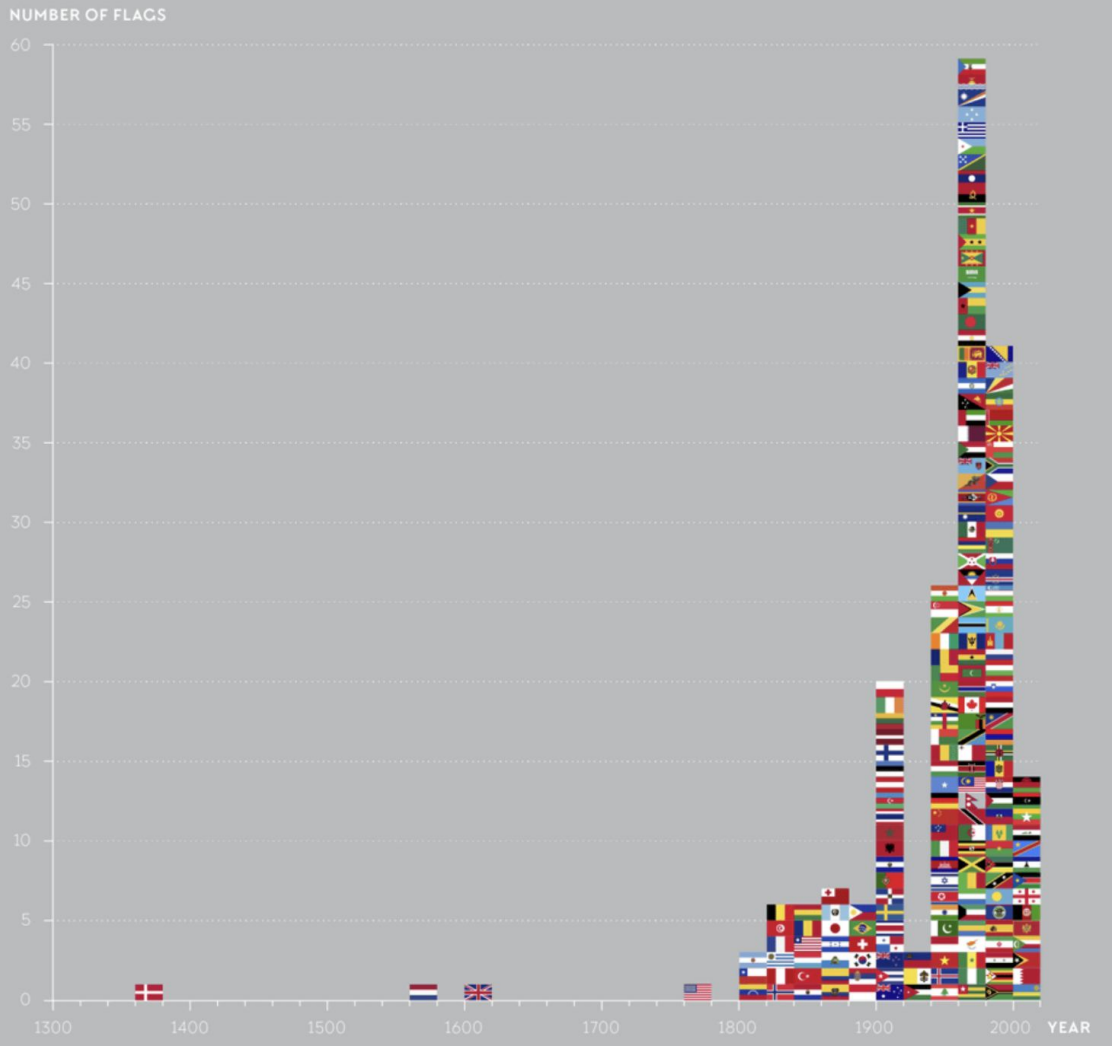
YouGov | yougov.com

February 26-28, 2017

Worst pie chart ever?

How old are national flags?

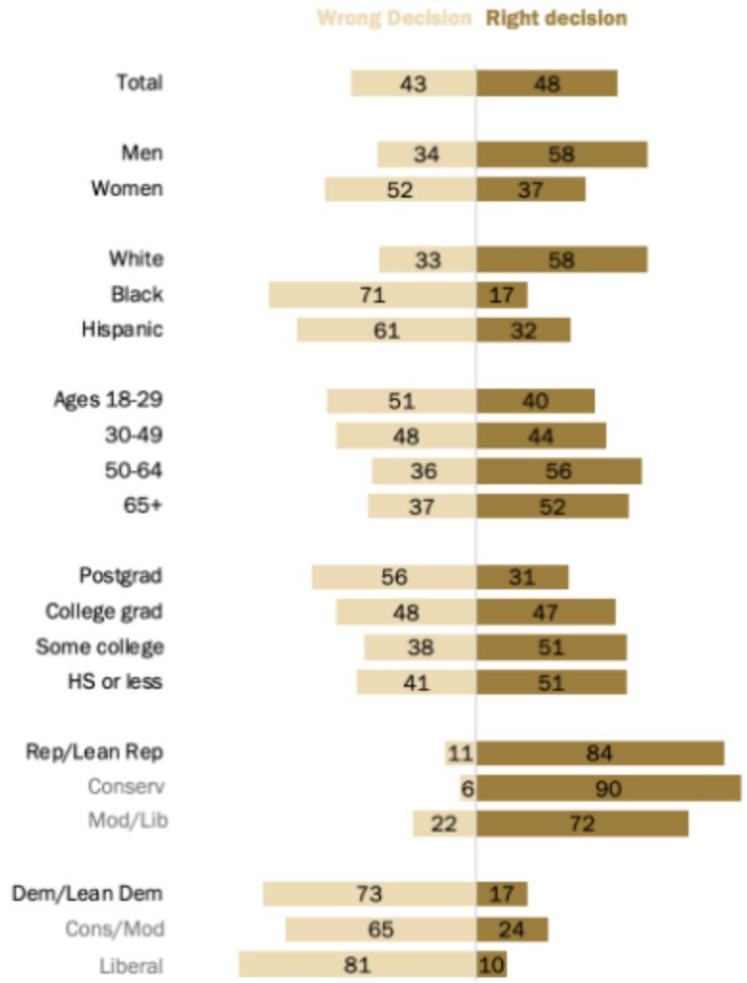
This timeline is based on the date in which each nation adopted their current national flag and is sorted by a 20 year interval. As you can see only four of the current national flags are dated before the 19th century.



Excellent use-case
For on-chart imagery

Wide partisan, demographic differences in views of U.S. decision to conduct airstrike that killed Soleimani

% who say U.S. decision to conduct the airstrike that killed Iranian Gen. Soleimani was the ...

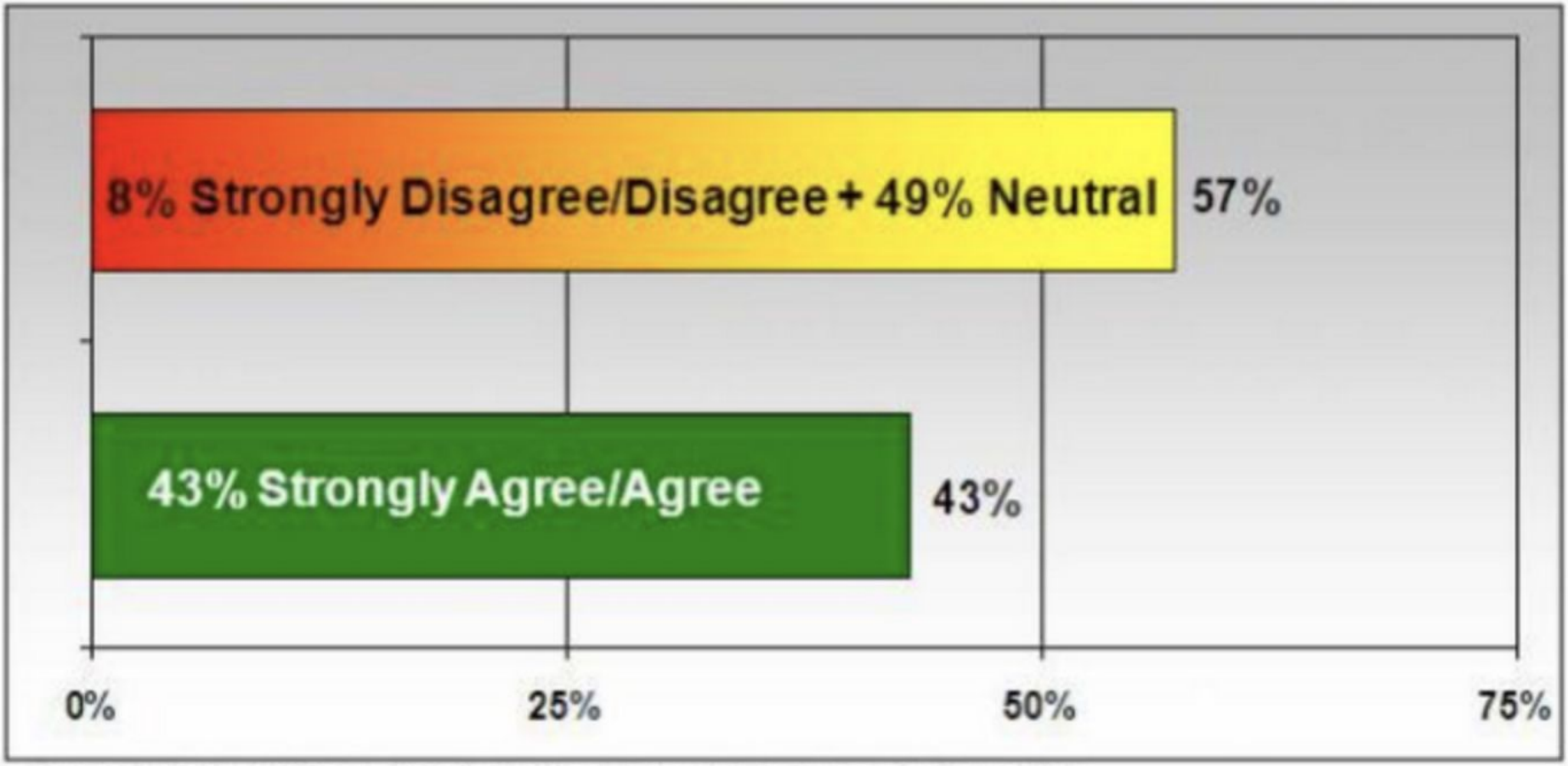


Notes: Don't know responses not shown. Whites and blacks include only those who are not Hispanic; Hispanics are of any race.
Source: Survey of U.S. adults conducted Jan. 8-13, 2020.

PEW RESEARCH CENTER

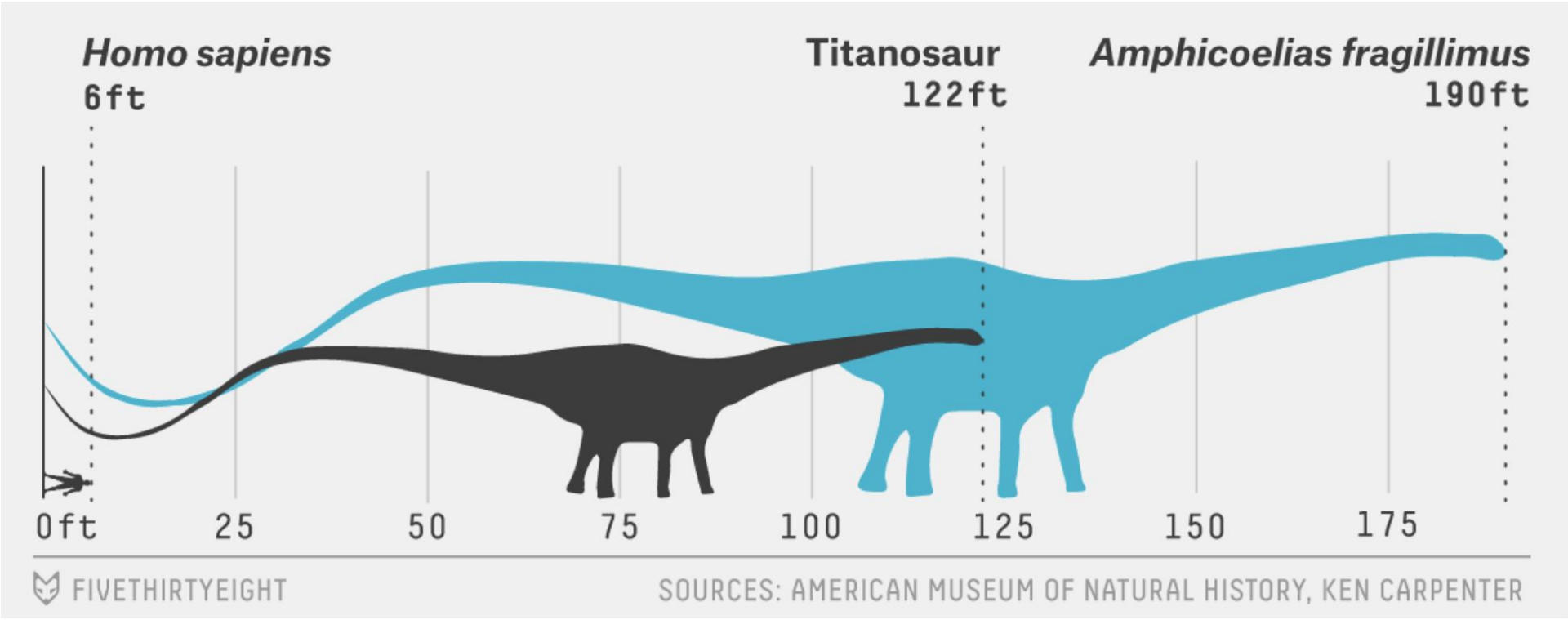
Unbiased reporting
Good data-ink ratio

Figure 2. Parent Response to *"New vaccines are safe for my child(ren)."*



Source: C.S. Mott Children's Hospital National Poll on Children's Health, March 2007

Over-simplication?
Manipulative visualization?
Confusing / unneccessary colors/gradient

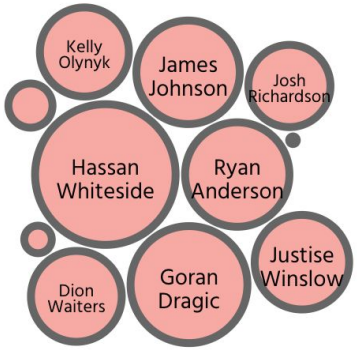


Simple, intuitive, effective
Educational for all ages

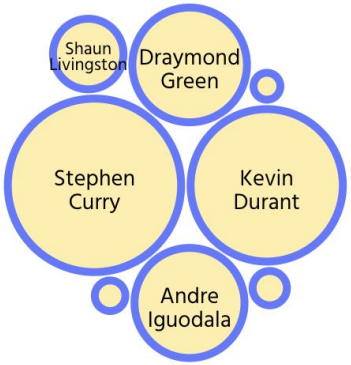
Casey Haber

'18-'19 '19-'20 '20-'21 '21-'22 '22-'23 '23-'24

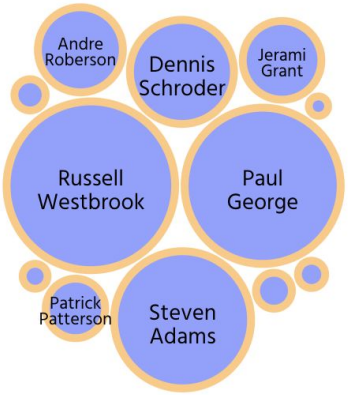
Miami Heat - $129.6M



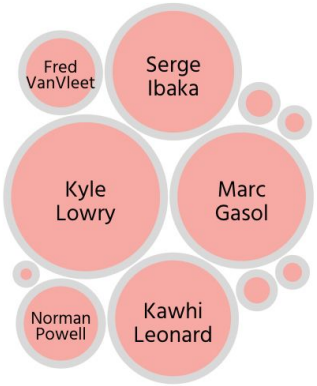
Golden State Warriors - $121.0M



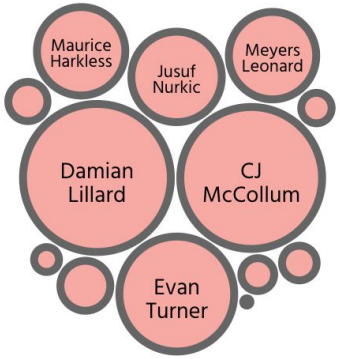
Oklahoma City Thunder - $146.8M



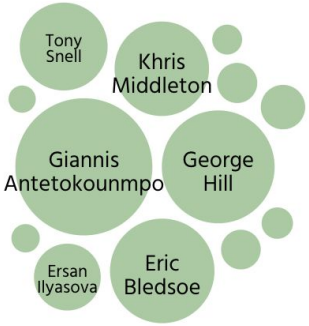
Toronto Raptors - $131.4M



Portland Trail Blazers - $126.0M

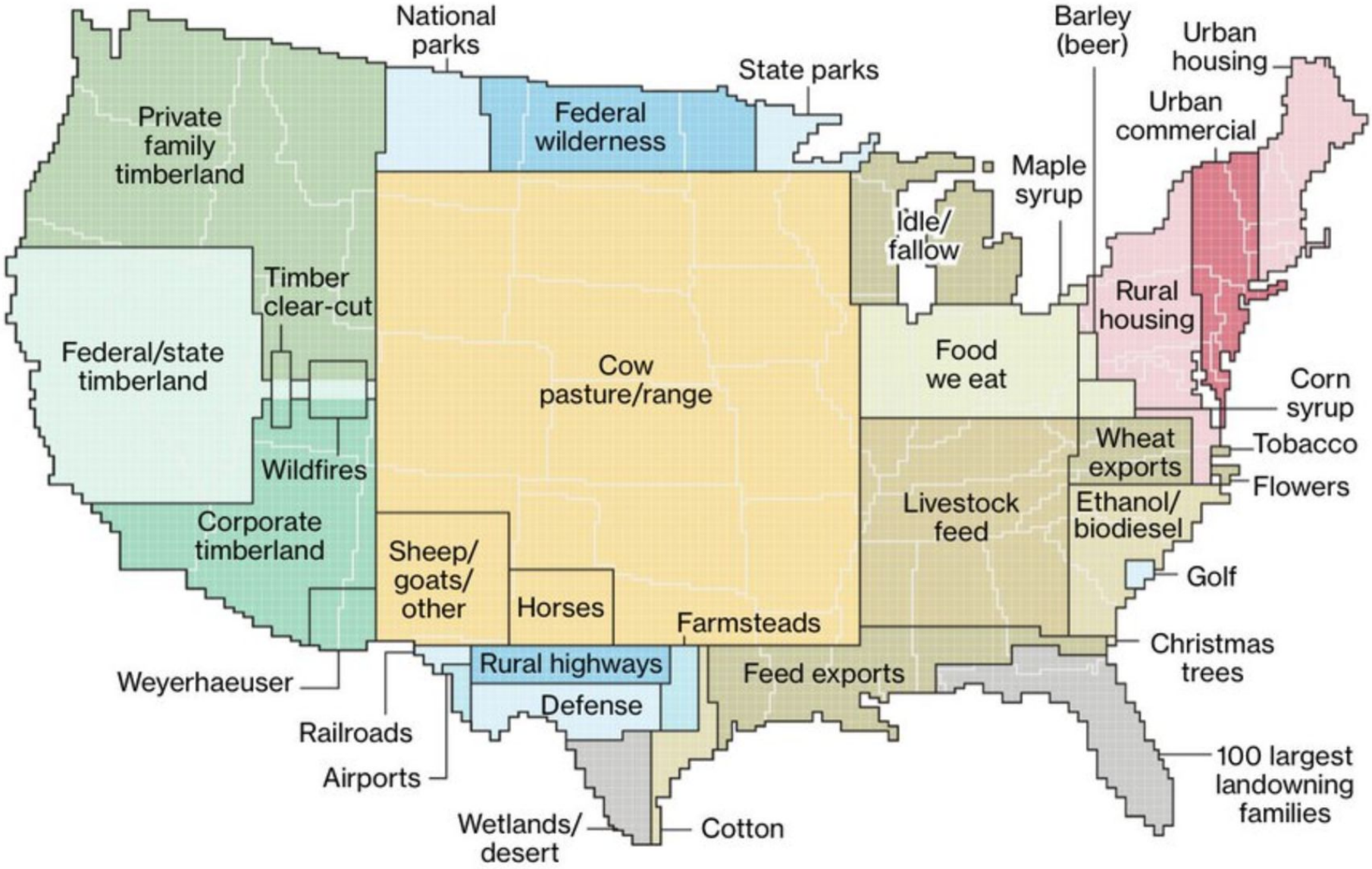


Milwaukee Bucks - $107.4M

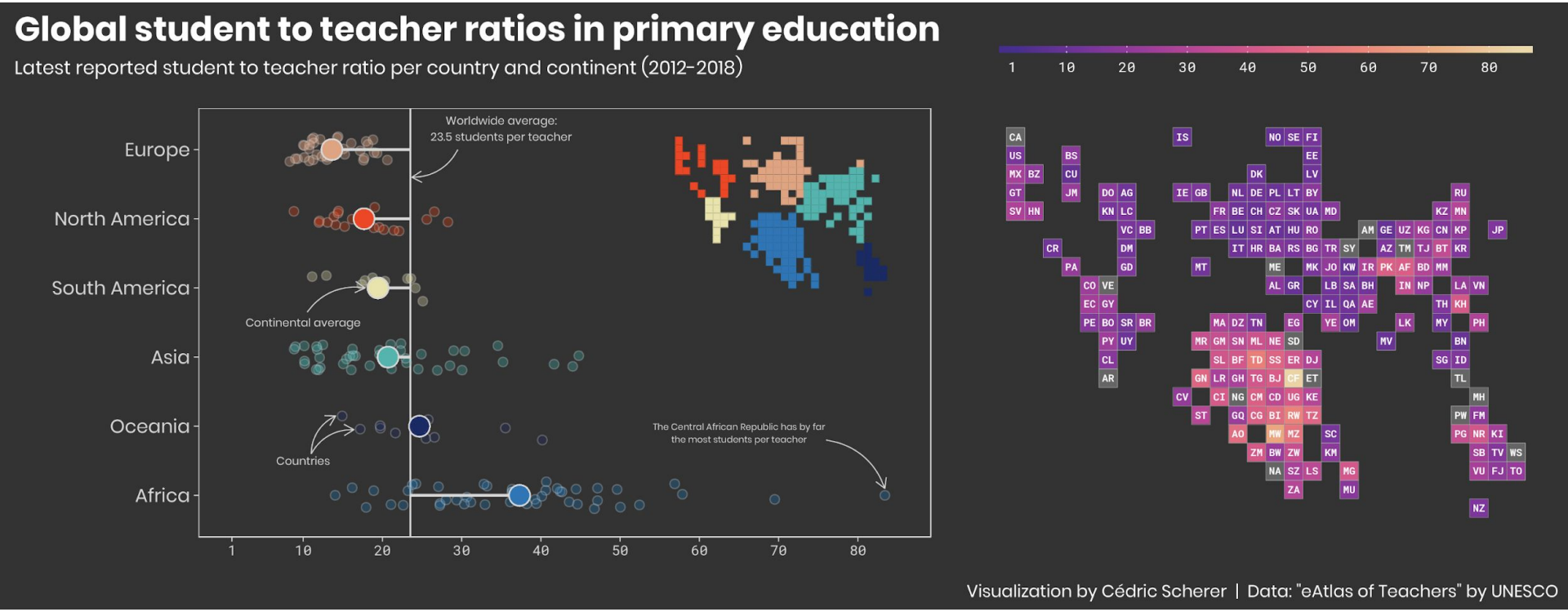


http://cahaber.me/#/basketball

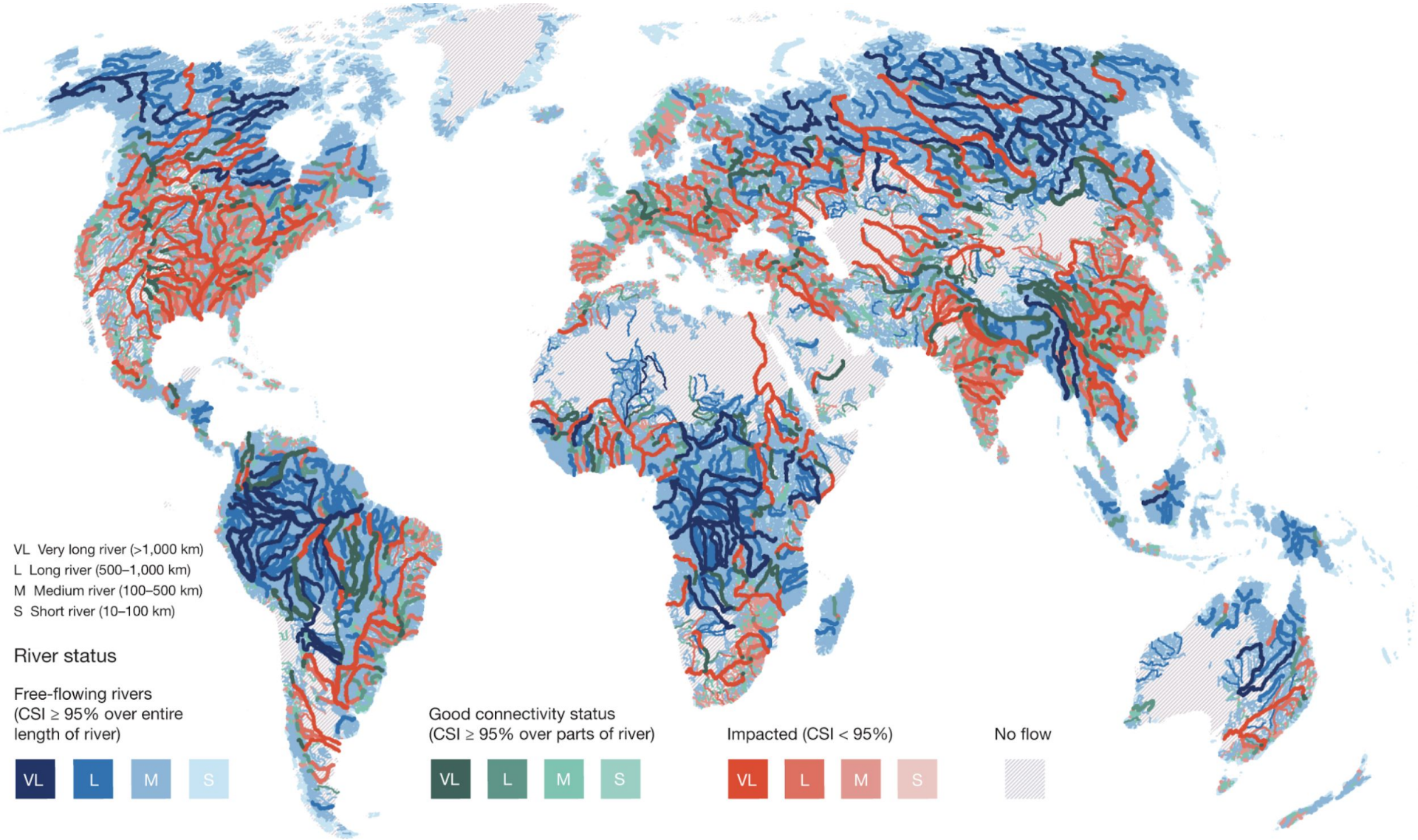
Compare players within team
Compare players on different teams
Different team choices/strategies



Effective non-pie shaped pie chart!



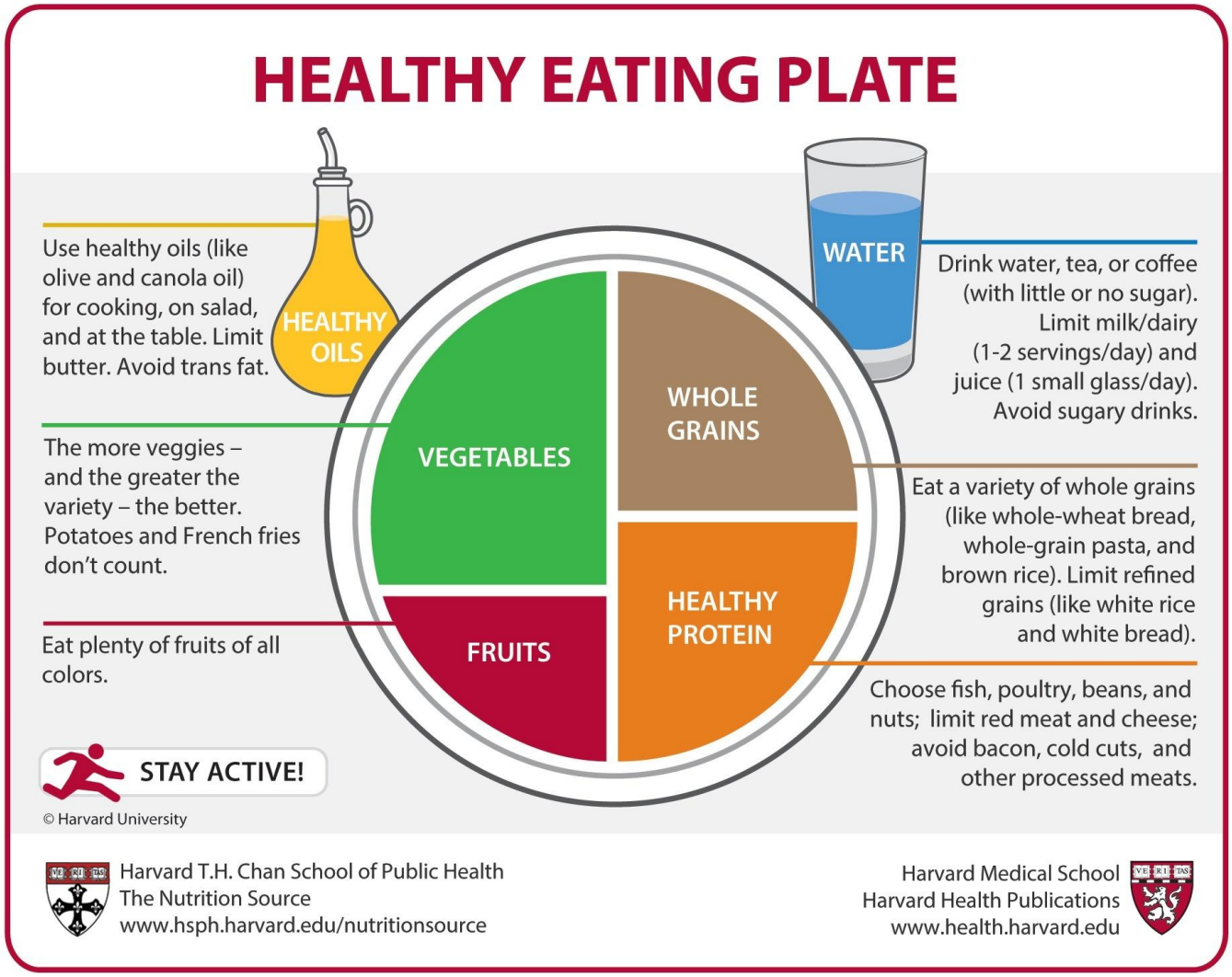
Are all datasets relevant to plot on a world map?



Allows us to clearly see significant variations in distributions of rivers
AND use (overuse?) of technology to control nature

http://www3.gehealthcare.com/en/Products/Categories/Healthcare_IT/Quality_Management#tabs/tab1900328377C74CAC8AD7E8D4A2072591

Stereotypical Infographic



https://cdn1.sph.harvard.edu/wp-content/uploads/sites/30/2012/09/HEPJan2015.jpg

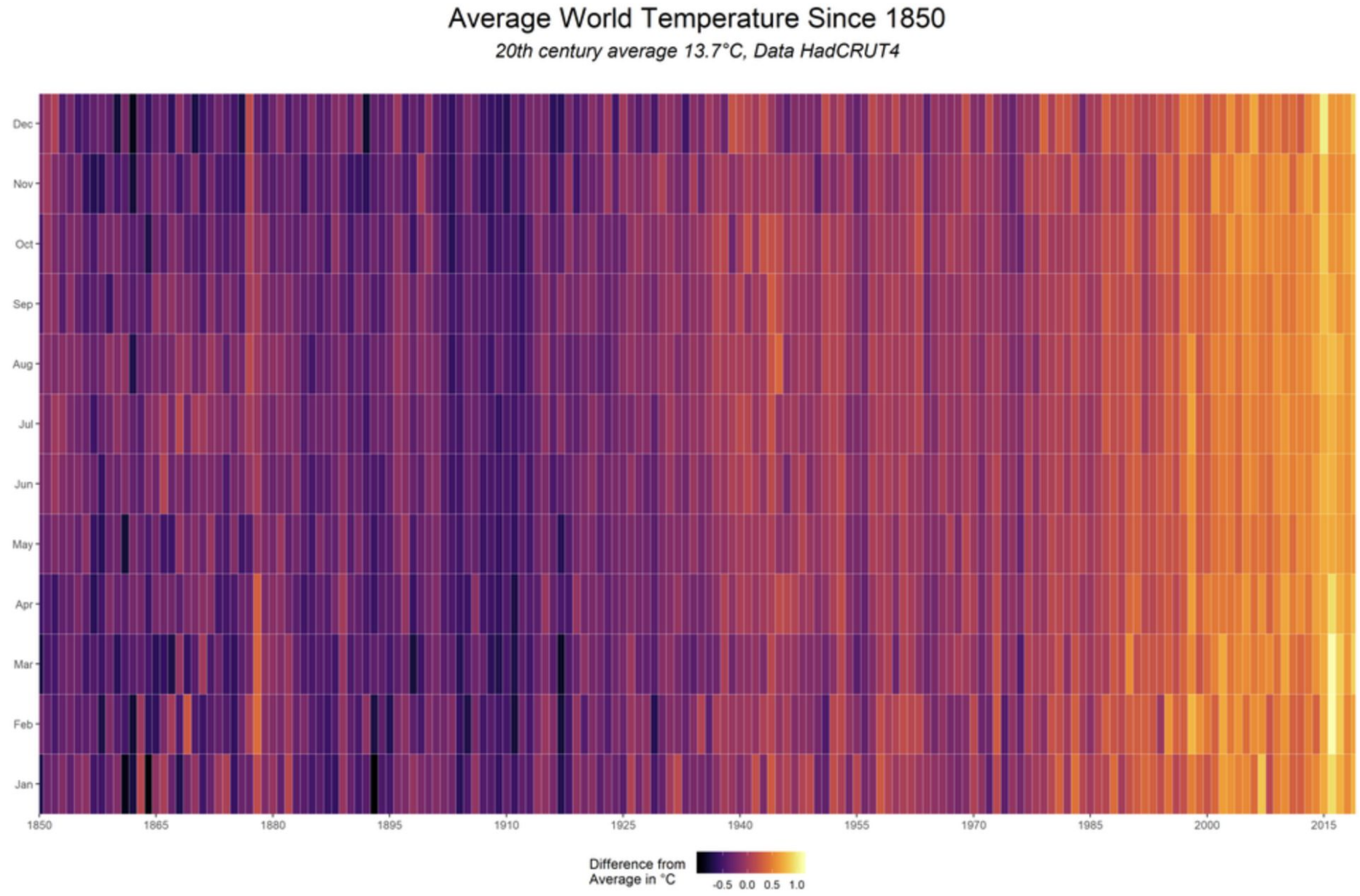
Vast improvement on original food pyramid

- Scientific Visualization (SciVis)
  - really large quantities of data
  - data usually has inherent structure
  - often has a spatial and/or temporal component (coordinate system)
  - often appropriate to use of 3D visualization techniques
  - such as medical, hurricane, computational fluid dynamics (CFD) data
- Information Visualization (InfoVis)
  - smaller datasets
  - data that does not have an inherent structure (may not have coordinate system)
  - financial stock market data, demographic census data, genetic data, etc.
- Visual Analytics
  - involves a cycle of rapidly creating visualizations to answer questions and generate new questions about a dataset
- Infographics are typically in the realm of InfoVis, and often they show the results of the visual analytics process, but SciVis is not really a part of most infographics.
- Annual IEEE Visualization and IEEE InfoVis (Information Visualization) conferences are two separate entities. The set of people organizing, attending and involved one conference is almost disjoint from the other set.

http://blog.visual.ly/the-beautiful-world-of-scivis/

https://visualizeit.wordpress.com/2007/06/07/the-great-infovis-and-scivis-divide/

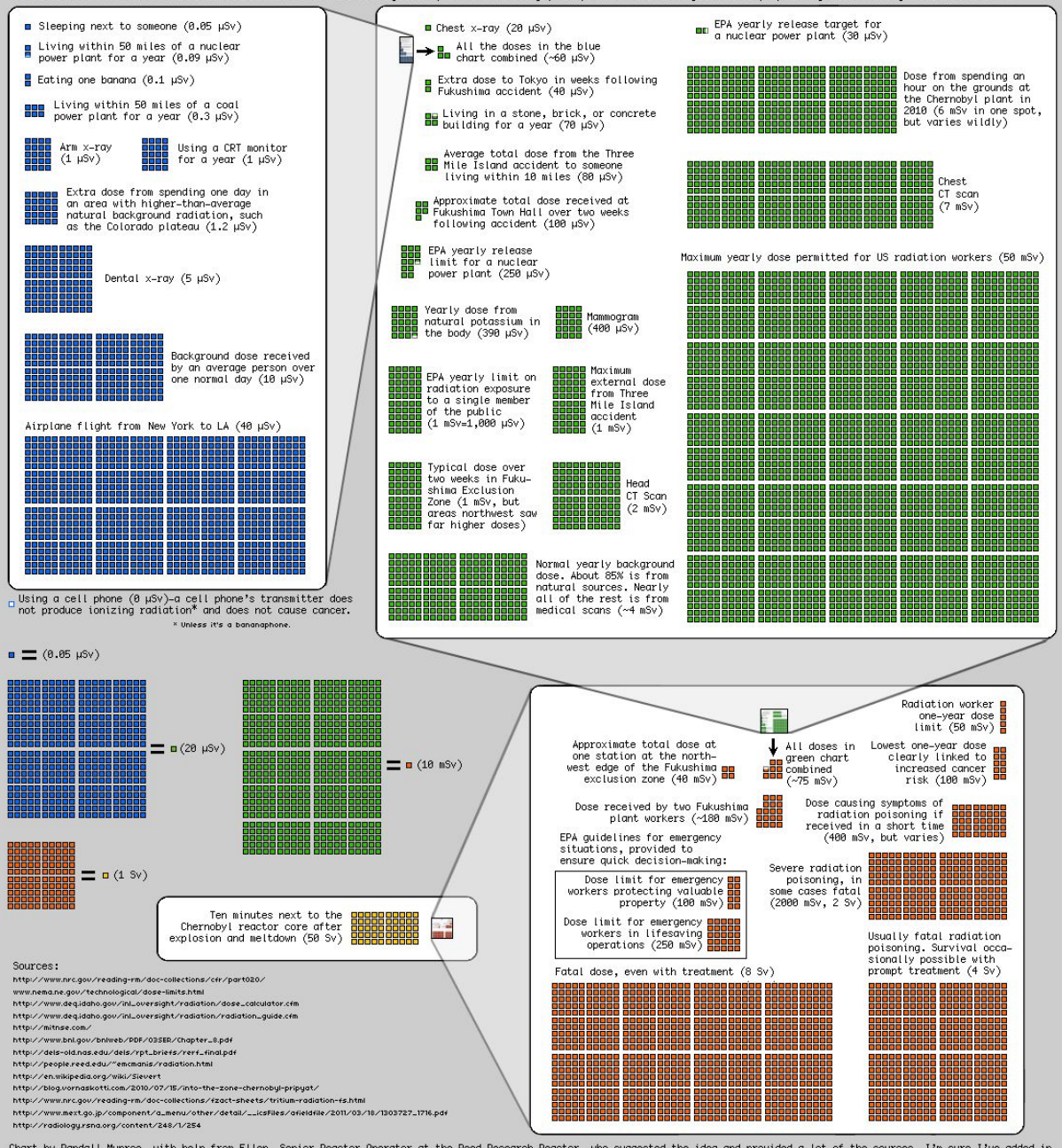
An Interactive Visualization!
(interactive data collection)



Visualization Challenge: very large datasets

Radiation Dose Chart

This is a chart of the ionizing radiation dose a person can absorb from various sources. The unit for absorbed dose is "sievert" (Sv), and measures the effect a dose of radiation will have on the cells of the body. One sievert (all at once) will make you sick, and too many more will kill you, but we safely absorb small amounts of natural radiation daily.
Note: The same number of sieverts absorbed in a shorter time will generally cause more damage, but your cumulative long-term dose plays a big role in things like cancer risk.



Sources:

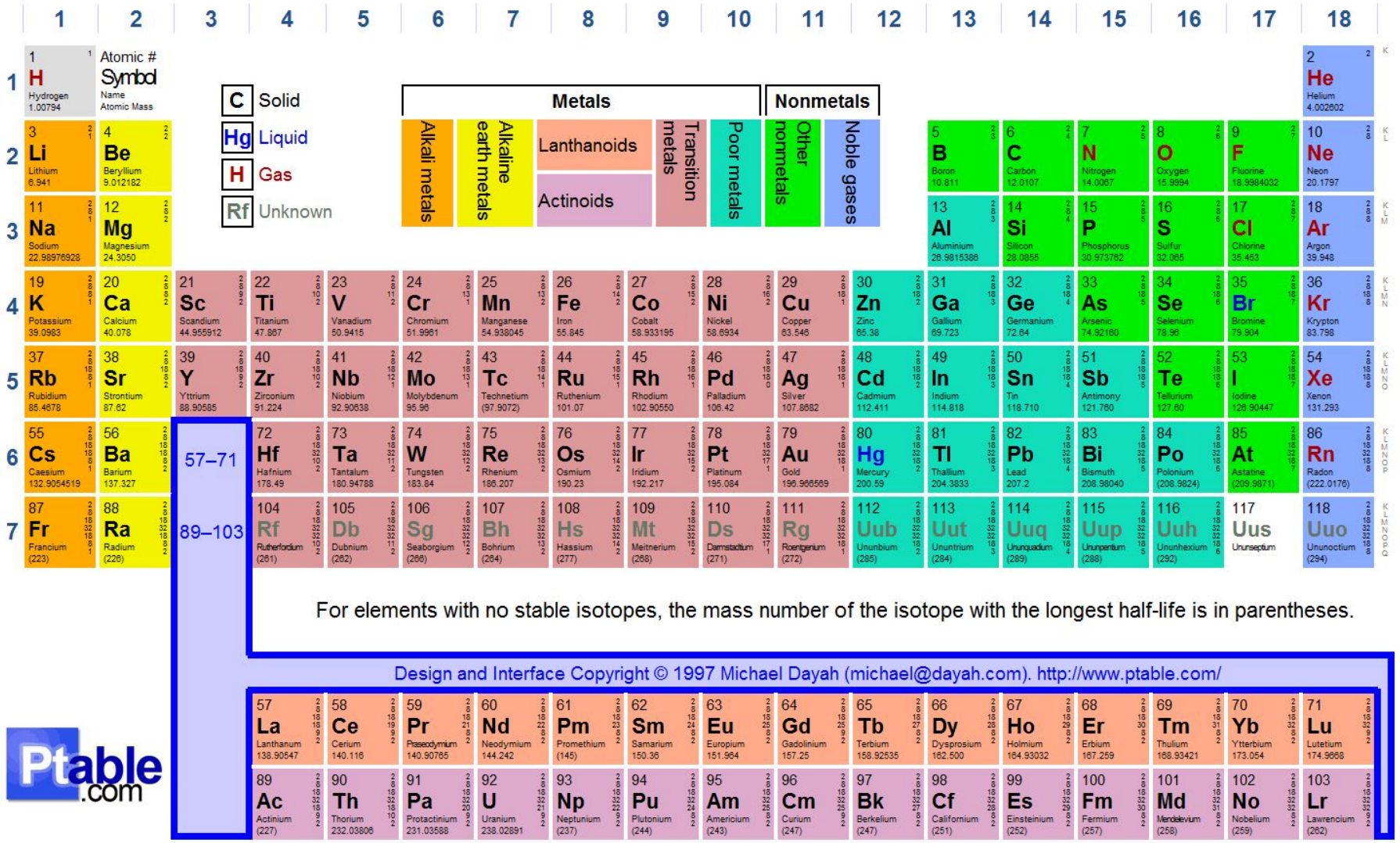
http://www.nrc.gov/reading-rm/doc-collections/cfr/part020/
www.nema.ne.gov/technological/dose-limits.html
http://www.deq.idaho.gov/inl_oversight/radiation/dose_calculator.cfm
http://www.deq.idaho.gov/inl_oversight/radiation/radiation_guide.cfm
http://mitnse.com/
http://www.bnl.gov/bnlweb/PDF/03SER/Chapter_8.pdf
http://dels-old.nas.edu/dels/rpt_briefs/rerf_final.pdf
http://people.reed.edu/~emcmanis/radiation.html
http://en.wikipedia.org/wiki/Sievert
http://blog.vornaskotti.com/2010/07/15/into-the-zone-chernobyl-pripyat/
http://www.nrc.gov/reading-rm/doc-collections/fact-sheets/tritium-radiation-fs.html
http://www.mext.go.jp/component/a_menu/other/detail/__icsFiles/afieldfile/2011/03/18/1303727_1716.pdf
http://radiology.rsna.org/content/248/1/254

Chart by Randall Munroe, with help from Ellen, Senior Reactor Operator at the Reed Research Reactor, who suggested the idea and provided a lot of the sources. I'm sure I've added in lots of mistakes; it's for general education only. If you're basing radiation safety procedures on an internet PNG image and things go wrong, you have no one to blame but yourself.

https://xkcd.com/radiation/

Visualization Challenge: representing different scales

Periodic Table of Elements

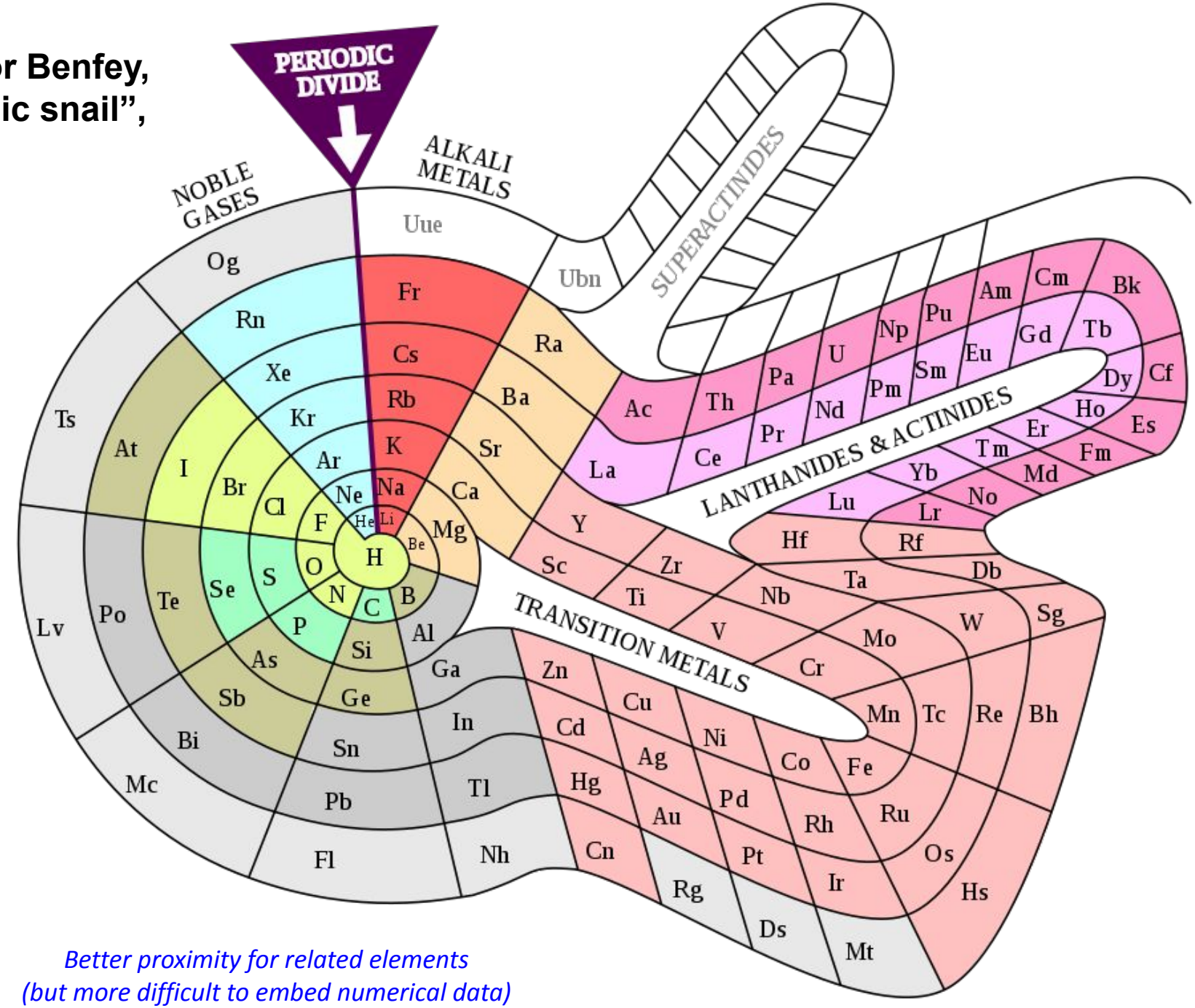


For elements with no stable isotopes, the mass number of the isotope with the longest half-life is in parentheses.

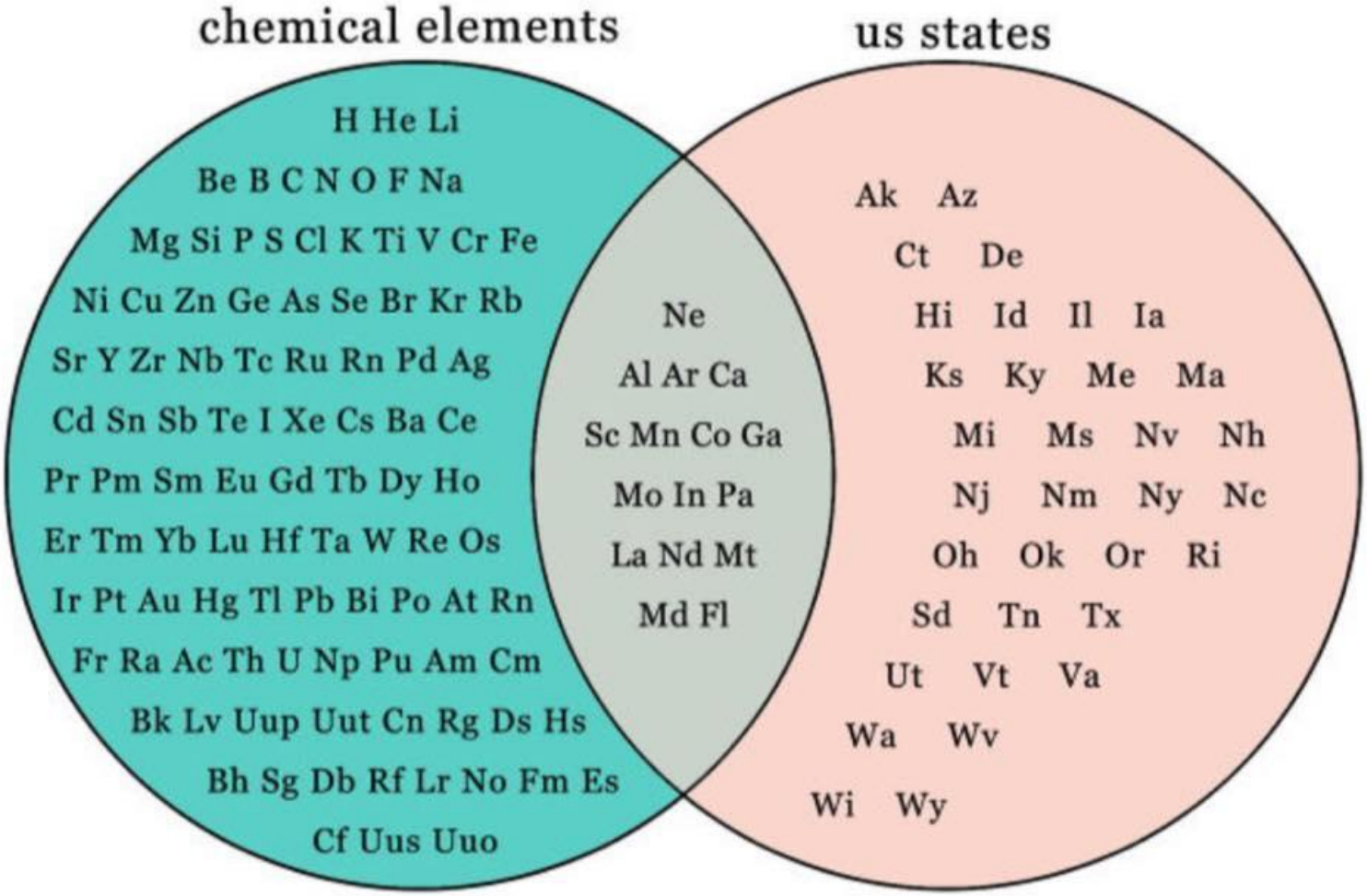
Ptable.com

http://www.ptable.com/

Theodor Benfey, "periodic snail", 1964

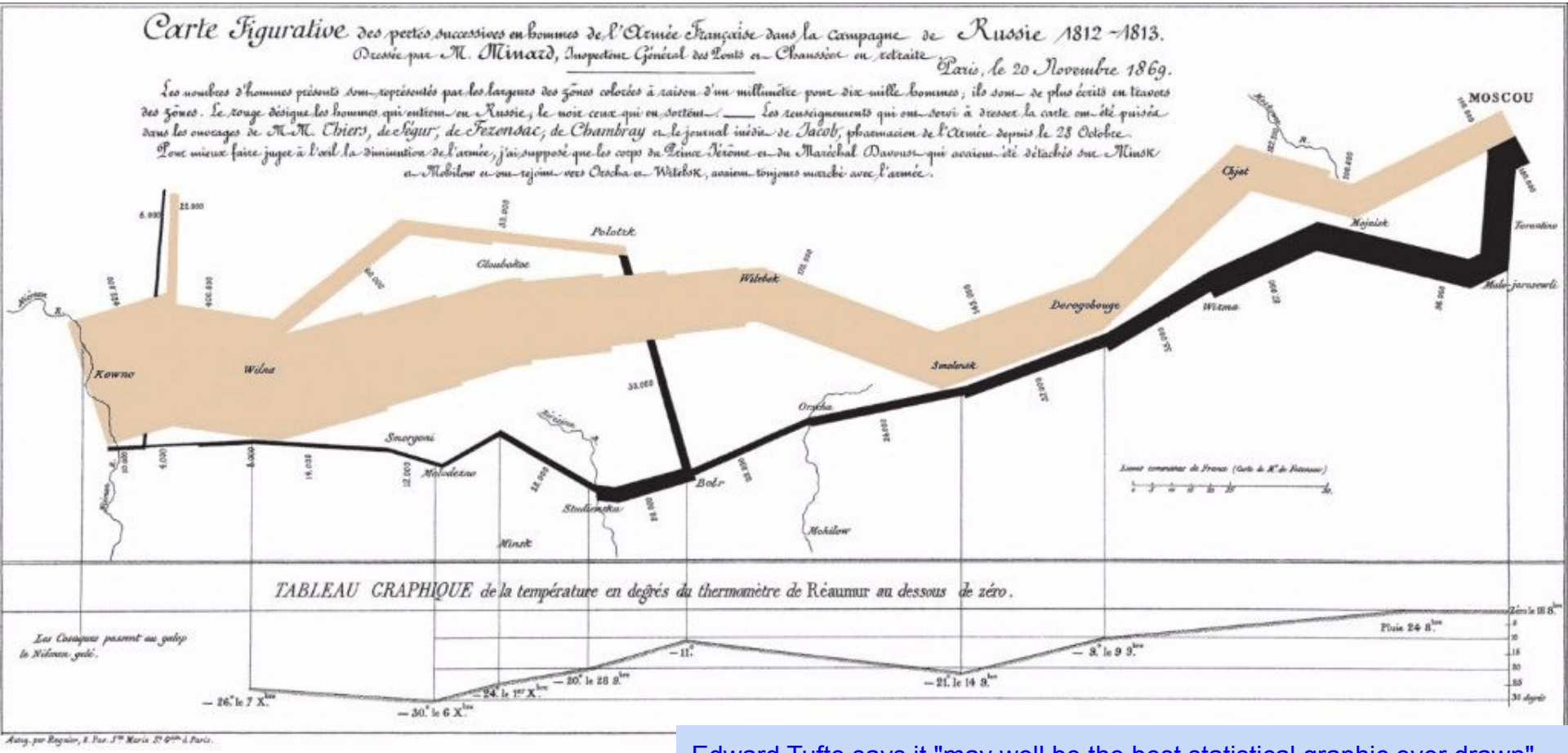


Better proximity for related elements
(but more difficult to embed numerical data)



chicanaspice

this information is so satisfying but idk what to do with it



Edward Tufte says it "may well be the best statistical graphic ever drawn"

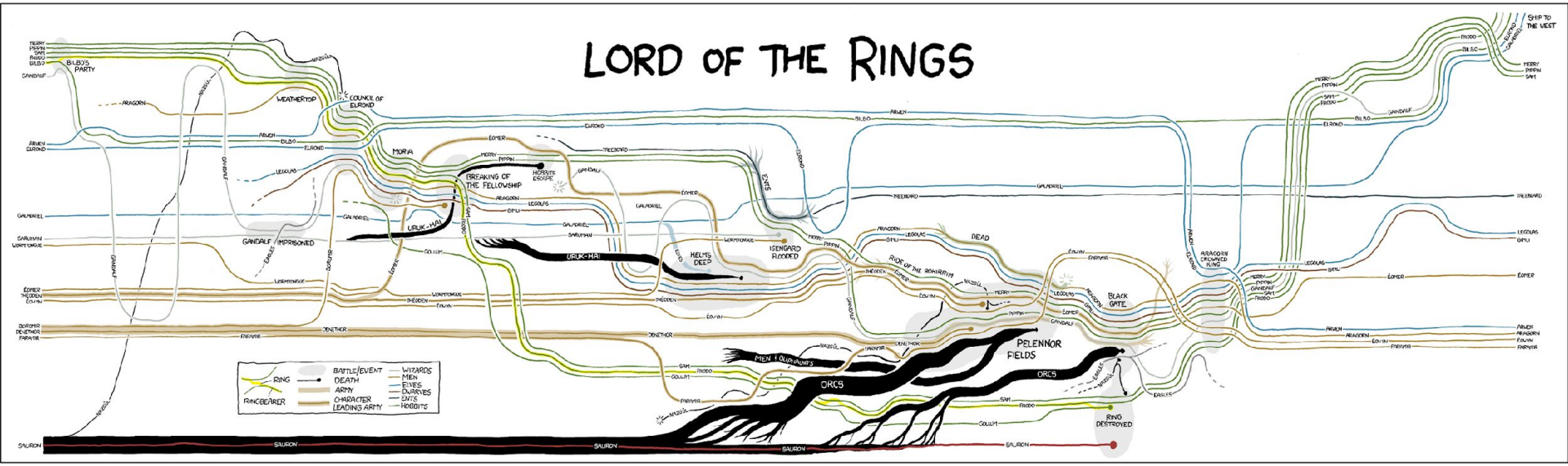
Figurative Map of the successive losses in men of the French Army in the Russian campaign 1812-1813.

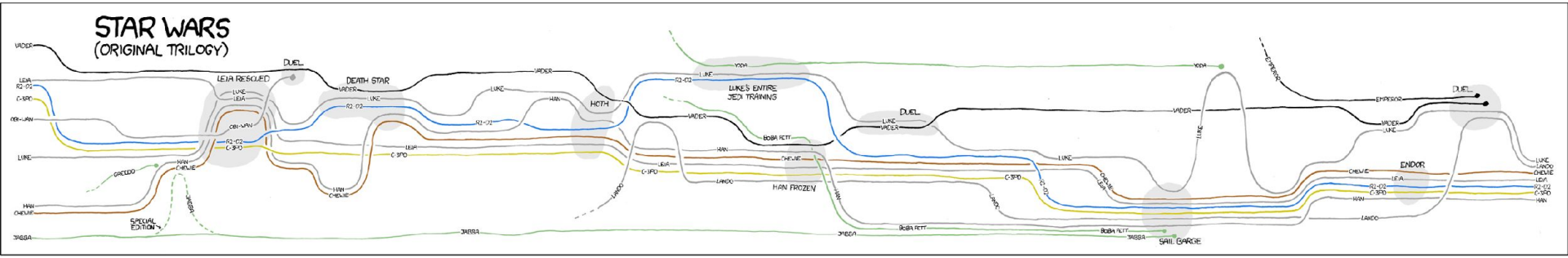
Charles Joseph Minard, *1869.*

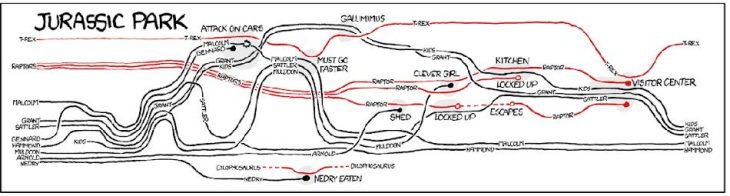
The numbers of men present are represented by the widths of the colored zones at a rate of one millimeter for every ten thousand men; they are further written across the zones. The red designates the men who enter Russia, the black those who leave it. — The information which has served to draw up the map has been extracted from the works of M.M. Thiers, de Ségur, de Fezensac, de Chambray and the unpublished diary of Jacob, the pharmacist of the Army since October 28th.

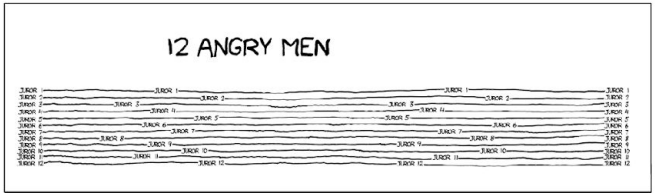
In order to better judge with the eye the diminution of the army, I have assumed that the troops of Prince Jérôme and of Marshal Davout, who had been detached at Minsk and Mogilev and have rejoined near Orsha and Vitebsk, had always marched with the army.

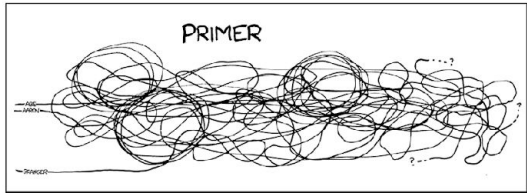
THESE CHARTS SHOW MOVIE CHARACTER INTERACTIONS.
THE HORIZONTAL AXIS IS TIME. THE VERTICAL GROUPING OF THE
LINES INDICATES WHICH CHARACTERS ARE TOGETHER AT A GIVEN TIME.











http://imgs.xkcd.com/comics/movie_narrative_charts_large.png

Today

- "Good" Design (30 min)
- Today's Readings:
  - How To
  - "Eenie, Meenie, Minie, Moe: Selecting the Right Graph for Your Message", Stephen Few, Intelligent Enterprise, 2004"
  - "Useful Junk? The Effects of Visual Embellishment on Comprehension and Memorability of Charts"
- Worksheet: Time-based Data & Simple Charts (20 min)
- Graph Drawing Preview: Terminology & Goals (15 min)
- Readings for Tuesday: Graph Drawing

Today

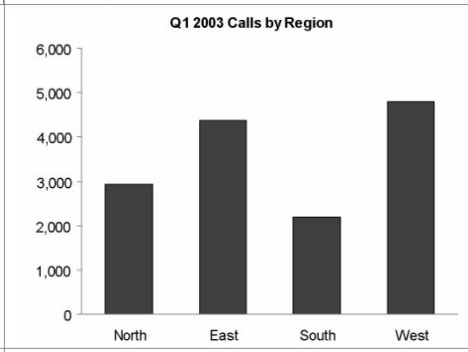
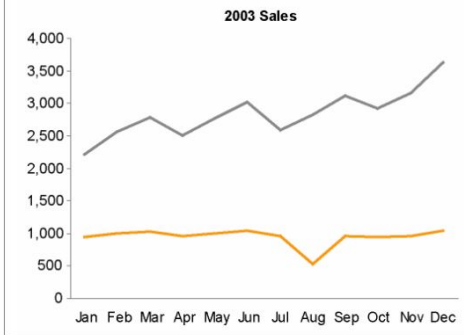
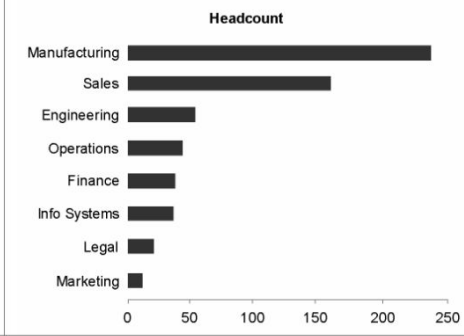
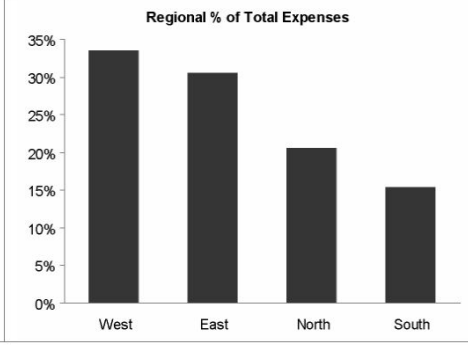
- "Good" Design (30 min)
- Today's Readings:
  - How To
  - "Eenie, Meenie, Minie, Moe: Selecting the Right Graph for Your Message", Stephen Few, Intelligent Enterprise, 2004"
  - "Useful Junk? The Effects of Visual Embellishment on Comprehension and Memorability of Charts"
- Worksheet: Time-based Data & Simple Charts (20 min)
- Graph Drawing Preview: Terminology & Goals (15 min)
- Readings for Tuesday: Graph Drawing

Today

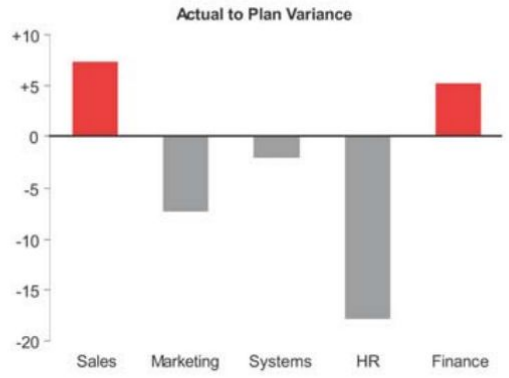
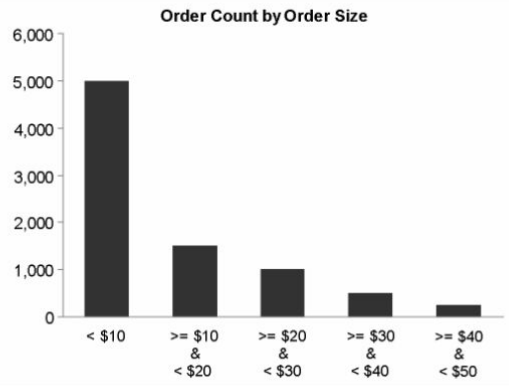
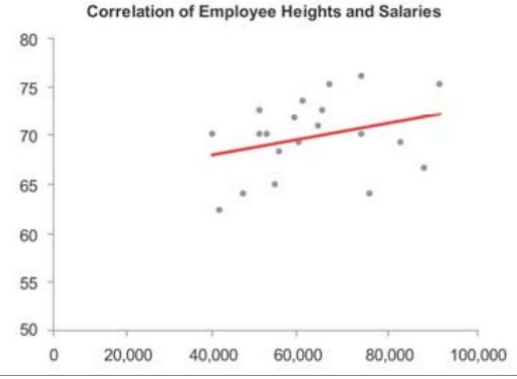
- "Good" Design (30 min)
- Today's Readings:
  - How To
  - "Eenie, Meenie, Minie, Moe: Selecting the Right Graph for Your Message", Stephen Few, Intelligent Enterprise, 2004"
  - "Useful Junk? The Effects of Visual Embellishment on Comprehension and Memorability of Charts"
- Worksheet: Time-based Data & Simple Charts (20 min)
- Graph Drawing Preview: Terminology & Goals (15 min)
- Readings for Tuesday: Graph Drawing

"Eenie, Meenie, Minie, Moe: Selecting the Right Graph for Your Message", Stephen Few, Intelligent Enterprise, 2004

- Learn conventions/patterns, applies to almost all business data (not necessarily all scientific data)
- Avoid viewer confusion / mis-information
- Definition: Categorical vs. quantitative
- Definition: Nominal (order free), ordinal (ordered), interval (e.g. histogram)
- Available visual attributes for encoding data: location, size, shape, orientation, color
  choose wisely!!

Type/Description	Encoding Methods	Example
Nominal Comparison A simple comparison of the categorical subdivisions of one or more measures in no particular order	• Bars only (horizontal or vertical)	
Time Series Multiple instances of one or more measures taken at equidistant points in time	• Lines to emphasize overall pattern • Bars to emphasize individual values • Points connected by lines to slightly emphasize individual values while still highlighting the overall pattern • Always place time on the horizontal axis	
Ranking Categorical subdivisions of a measure ordered by size (either descending or ascending)	• Bars only (horizontal or vertical) • To highlight high values, sort in descending order • To highlight low values, sort in ascending order	
Part-to-Whole Measures of individual categorical subdivisions as ratios to the whole	• Bars only (horizontal or vertical) • Use stacked bars only when you must display measures of the whole as well as the parts	

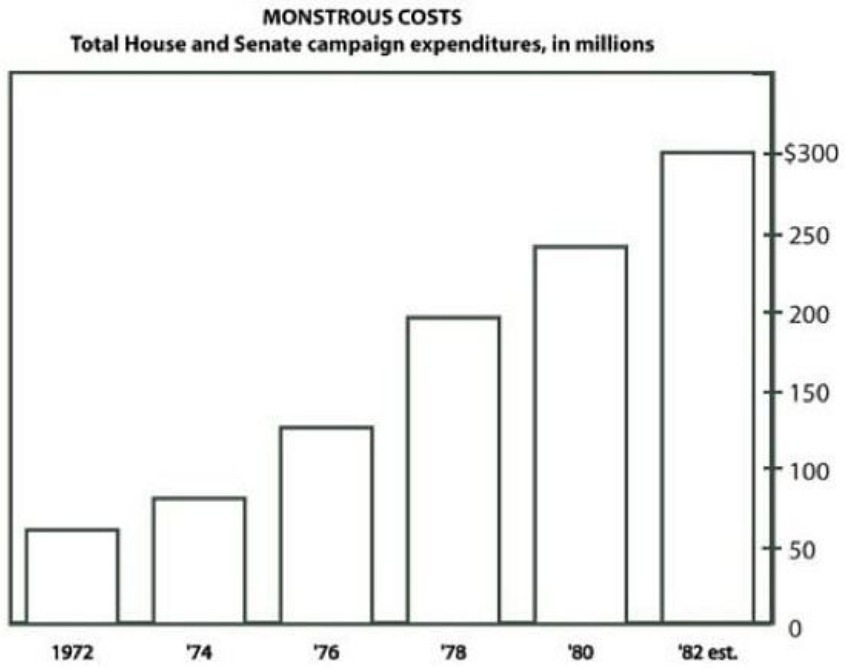
"Eenie, Meenie, Minie, Moe: Selecting the Right Graph for Your Message", Stephen Few, Intelligent Enterprise, 2004

Deviation Categorical subdivisions of a measure compared to a reference measure, expressed as the differences between them	• Lines to emphasize the overall pattern only when displaying deviation and time-series relationships together • Points connected by lines to slightly emphasize individual data points while also highlighting the overall pattern when displaying deviation and time-series relationships together • Bars to emphasize individual values, but limit to vertical bars when a time-series relationship is included • Always include a reference line to compare the measures of deviation against	
Frequency Distribution Counts of something per categorical subdivisions (intervals) of a quantitative range	• Vertical bars to emphasize individual values (called a *histogram*) • Lines to emphasize the overall pattern (called a *frequency polygon*)	
Correlation Comparisons of two paired sets of measures to determine if as one set goes up the other set goes either up or down in a corresponding manner, and if so, how strongly	• Points and a trend line in the form of a scatter plot • Bars may be used, arranged as a *paired bar graph* or a *correlation bar graph*, if scatter plots are unfamiliar • (Note: For descriptions of these graphs, see my book *Show Me the Numbers*.)	

"Eenie, Meenie, Minie, Moe: Selecting the Right Graph for Your Message", Stephen Few, Intelligent Enterprise, 2004

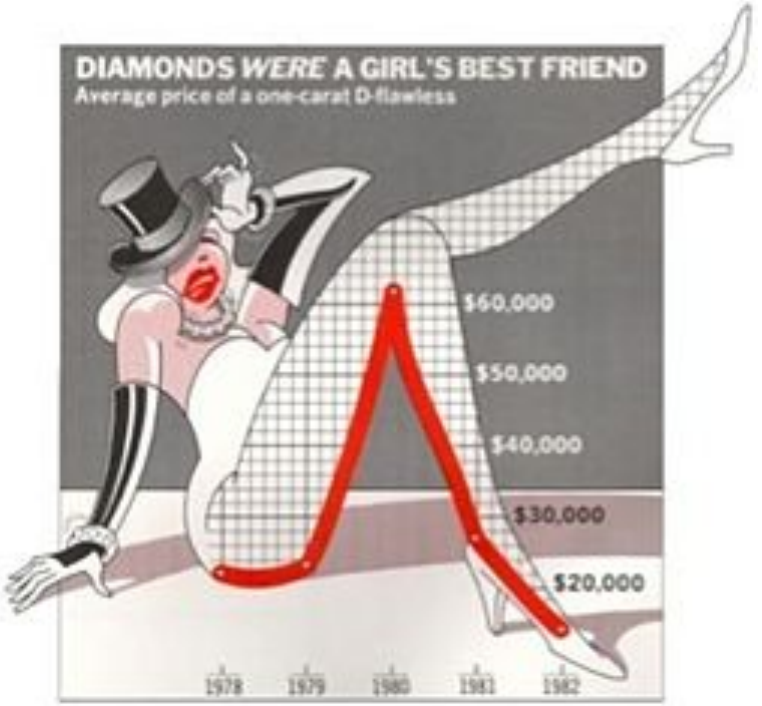
Reading for Today

- "Useful Junk? The Effects of Visual Embellishment on Comprehension and Memorability of Charts" Bateman et al., CHI 2010.



- Article discussed here: http://eagereyes.org/criticism/chart-junk-considered-useful-after-all

What is “Chart Junk”?



Nigel Holmes

- Extraneous elements in a chart or visualization
- Does not represent data
- Data-to-ink ratio (aim to convey more data with less ink)
- According to Edward Tufte:
  It’s not just unnecessary, it’s harmful (distracting)
- According to Nigel Holmes:
  Visualization should engage the reader’s interest

Study Design

"Useful Junk? The Effects of Visual Embellishment on Comprehension and Memorability of Charts" Bateman, Mandryk, Gutwin, Genest, McDine, & Brooks, CHI 2010

- Compare embellished charts to plain ones
- Measured:
  - interpretation accuracy
    was no worse for embellished charts
  - long-term recall (2-3 weeks later)
    was better for embellished charts,
    topic & details of the chart were more memorable
- Prior work:
  - Higher data-to-ink ➔ faster response & greater accuracy [Gilan & Richman]
  - Other work shows a somewhat weak correlation between data-to-ink and interpretability or aesthetics
- Author's caution:
  - Not an endorsement of chart junk
  - Embellishments can lead to bias!

- Relatively small sample pool
  - 10 tested with ~5 min recall
  - 10 tested with 2-3 week recall
- Found no difference in time to read & describe embellished vs. non-embellished
- Participants preferred the embellished charts and found them more attractive

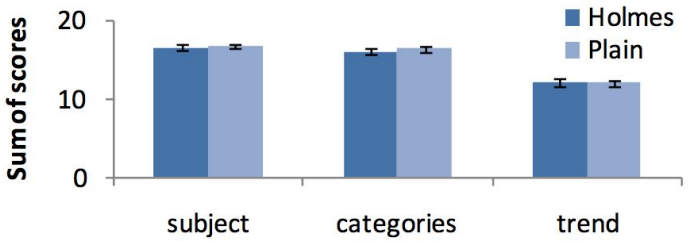


Figure 4. Means ± SE for description scores.

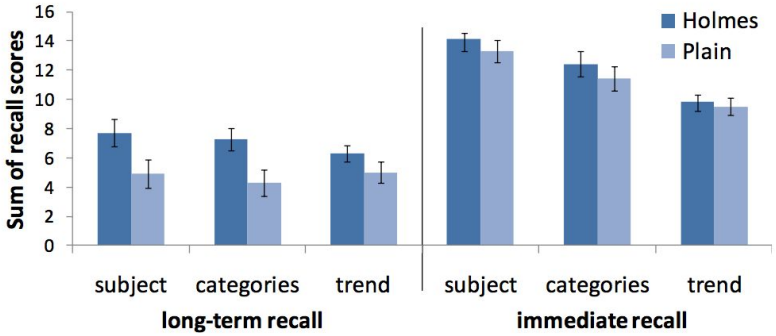
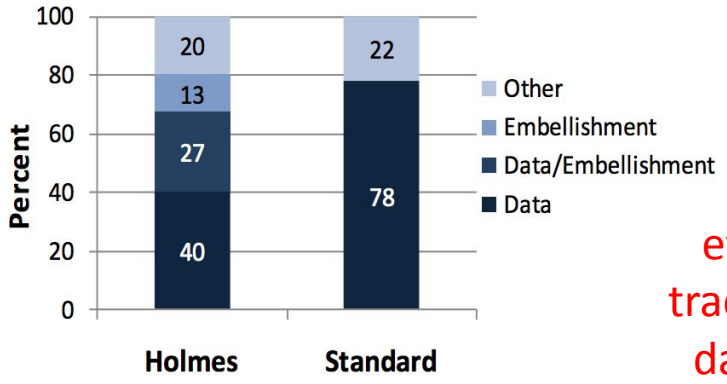


Figure 5. Means ±S E for recall scores for long-term and immediate recall.



eye tracker data

Figure 9. Percentage of on-screen time spent looking at different chart elements for Holmes and Plain charts.

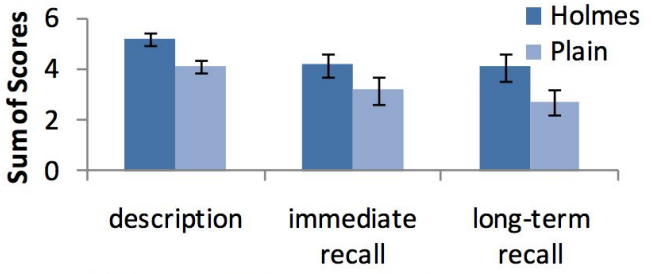


Figure 6. Means ± SE for sum of value message scores.

- Viewing time was unlimited for this study
  - Participants ended up spending the ~same amount of time on embellished vs. non-embellished
  - Effect of limiting time *not measured*
- Chart junk for these examples was tightly coupled with subject & details of chart
  - Quote from Holmes: "I think [Tufte] missed the point of much that I was trying to do: TIME magazine charts were aimed at lay readers, not unintelligent ones, but busy ones. I knew they'd get the point quicker if they were somehow attracted to the graphic."
- What about charts from paper on last slide? What was their point? How good is your recall? Will your recall them in 2-3 weeks? Why didn't the authors use embellishment?

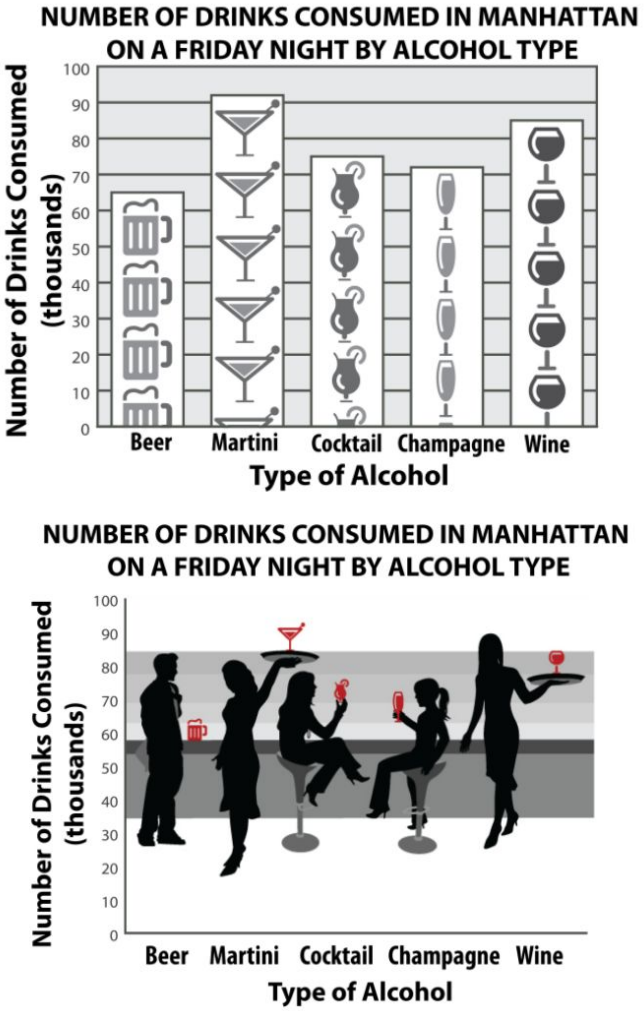


Figure 10. Less extreme visual imagery in charts.

http://nigelholmes.com/graphic/data-dump/

Homework Assignment 2: Time-Based Datasets

- Team of 2 or 3
- Obtain an interesting time-based dataset
  - Should be collectable* from online sources, and
  - Require a modest effort to prepare*

  ** = you'll submit your scripts/code to document*
- Use Excel or Google Sheets
  - Create a variety (one of each?!) of the charts following the guidelines from "Eenie, Meenie, Minie, Moe: Selecting the Right Graph for Your Message"
  - Excellent labels and captions for each.
- Upload your assignment to Submitty by 11:59pm on Thursday. And post one of the images on the forum...

Pair Worksheet (~20 minutes)

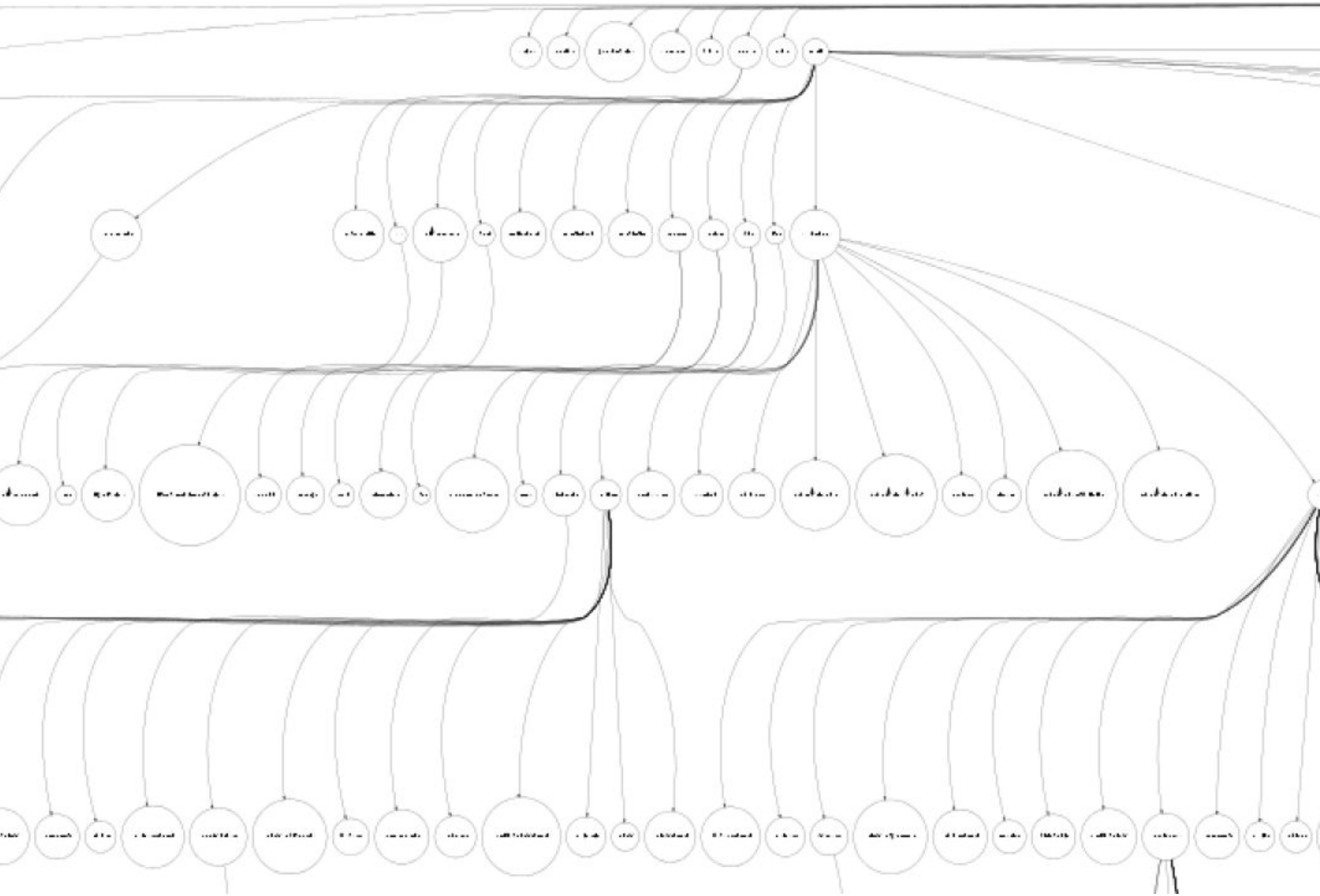
- Meet new people
  - *Work with someone you did not know before this class*
  - *Work with a different partner every time*
- 1 worksheet per team of 3
  - *Sketch & brainstorm on the page*
  - Submit on Submitty
- Use color! Be creative!

Tools for scraping data from the web

- copy-paste
- wget
- grep / sed / awk / sort / uniq
- Favorite programming language to parse/strip out unnecessary html formatting
- Save as .csv (comma separated value) files to upload to Excel / Google Sheets
- Python has lots of packages for parsing (e.g., json format)
- Selenium for automated browsing of websites

Today

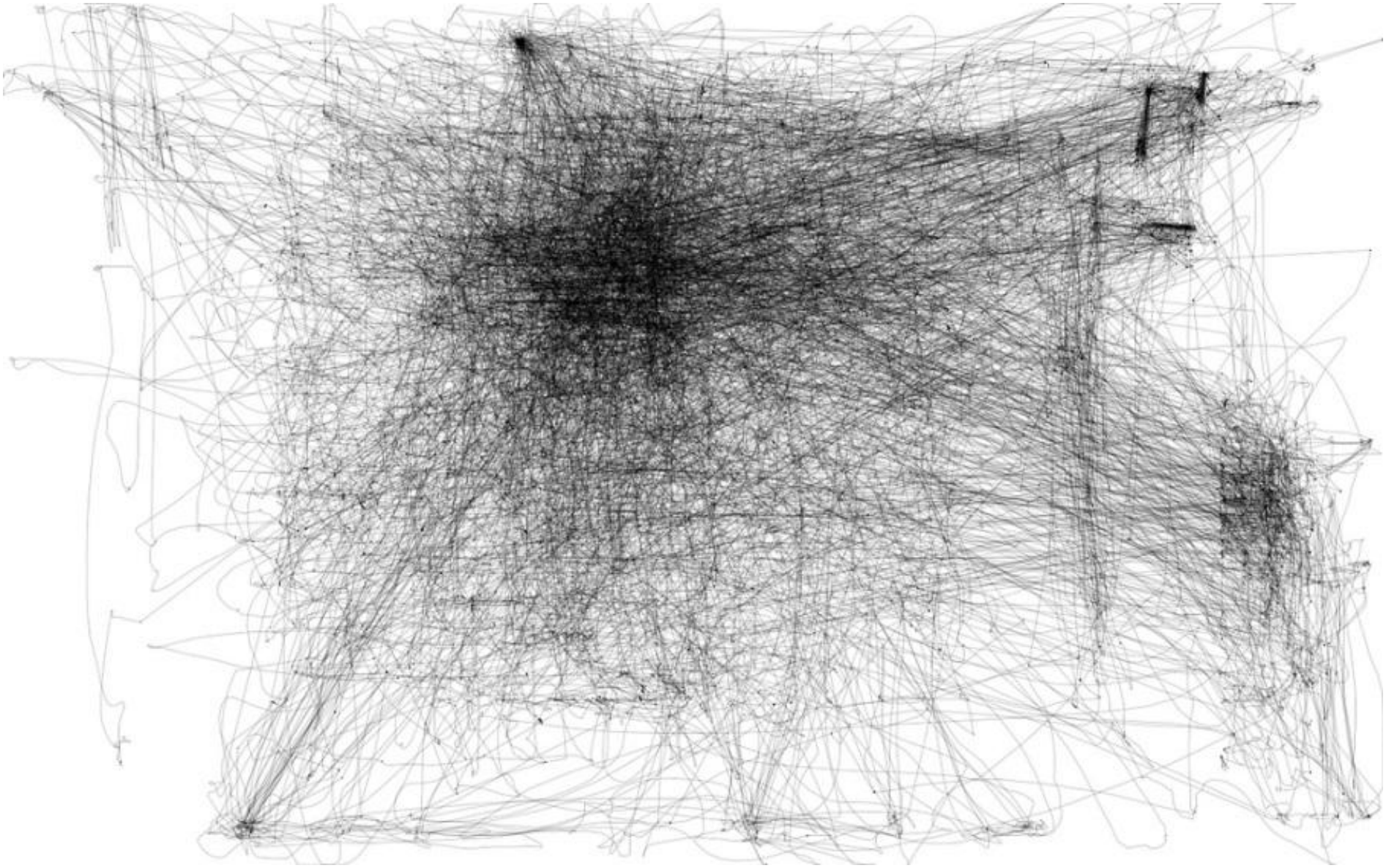
- "Good" Design (30 min)
  - Photography tips
  - Principles of Effective Website Design
  - Principles of Good User Interface Design
  - Examples of Good (Bad) Visualization Design
- Today's Reading (30 min): "Useful Junk? The Effects of Visual Embellishment on Comprehension and Memorability of Charts"
- Worksheet: Time-based Data & Simple Charts (20 min)
- Graph Drawing Preview: Terminology & Goals (15 min)
- Readings for Tuesday: Graph Drawing



Justin J. (from Spring ‘12)

Graph Terminology I

- Directed / Undirected Edges
- Tree (no cycles) vs. Graph (cycles allowed)
  - Cycle: A path along edges through the graph where only starting & ending vertices are repeated.
  - Walk: A sequence of vertices and edges
    - closed walk, simple cycle, directed cycle, …
- Valence (a.k.a. Degree) of a Vertex:
  # of edges incident on the vertex
- Regular: Each vertex has same valence,
  a 3-regular graph is also called cubic

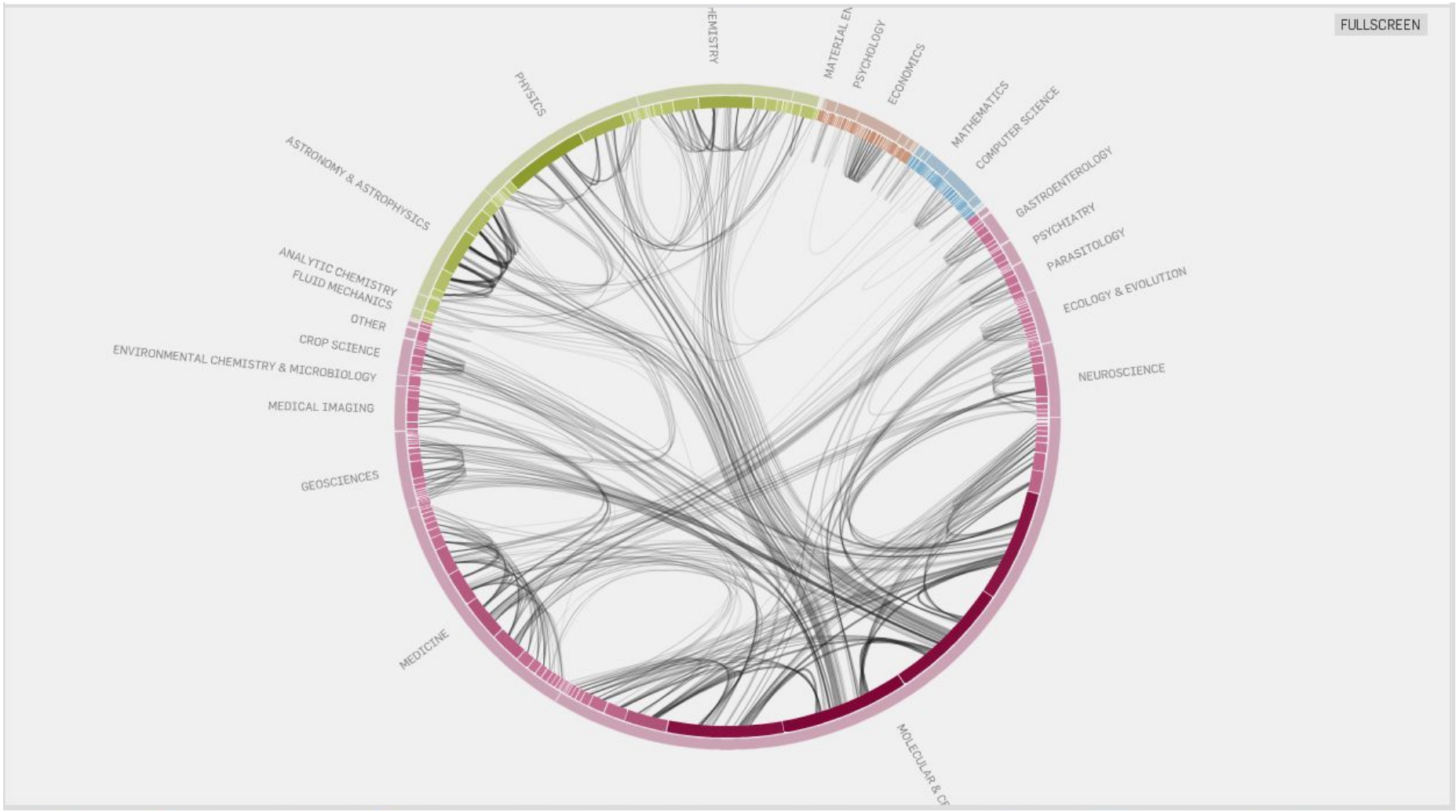


IOGraphica http://iographica.com

Graph Terminology II

- Polygon: 2D flat or on a sphere, with straight or great circle edges
- Polyhedron: 3D solid formed by flat faces
- Polytope: flat sides in any dimension

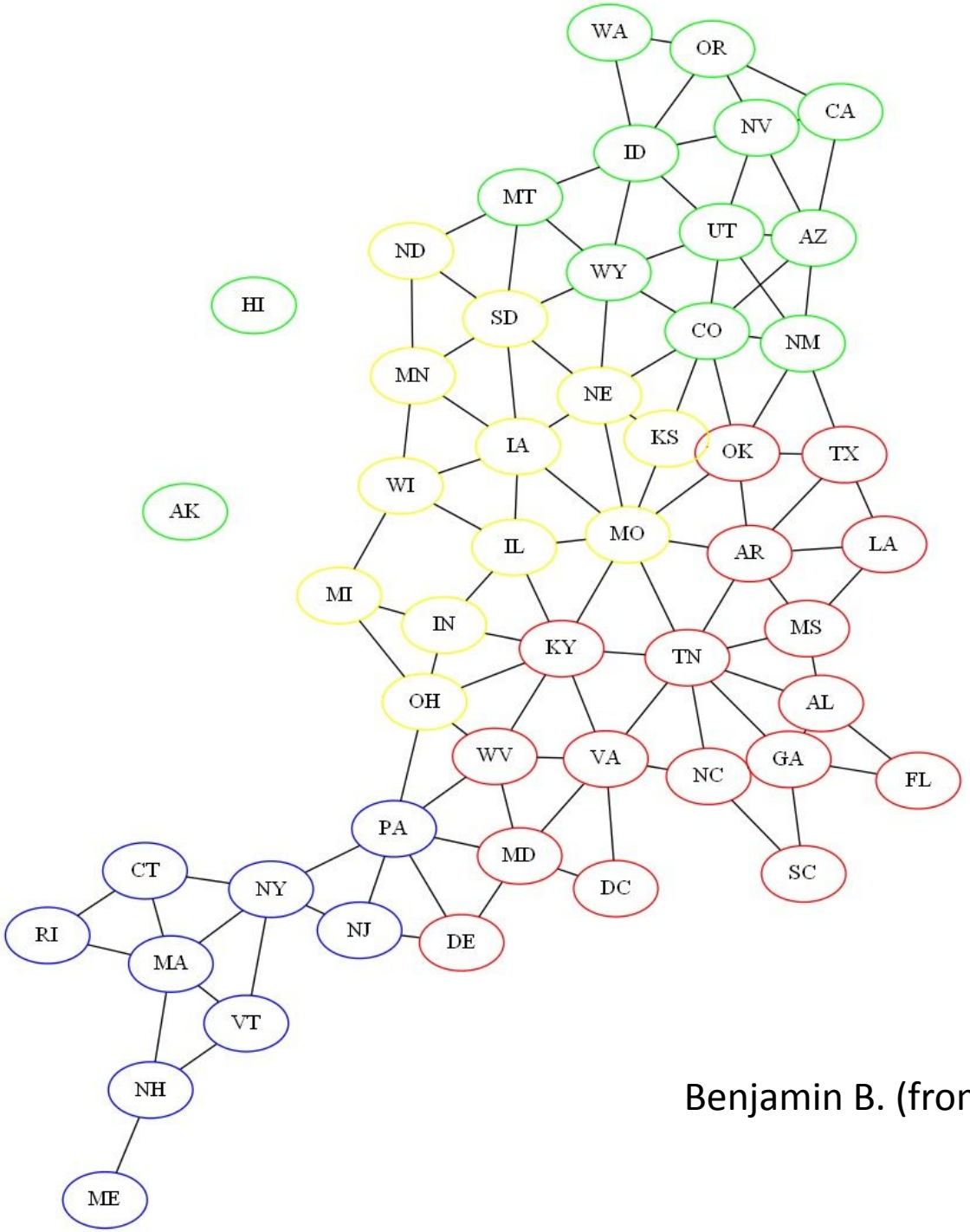
- Bipartite: vertices can be split into two groups, A & B. No edge connects a vertex in A to another vertex in A. Same for B.
- Clique – subset of vertices in an undirected graph with an edge connecting every pair of vertices in the subset.



Stefaner, Moritz. "Citation Patterns." *Wellformed Eigenfactor*. N.p., 2008. Web.

Graph Terminology III

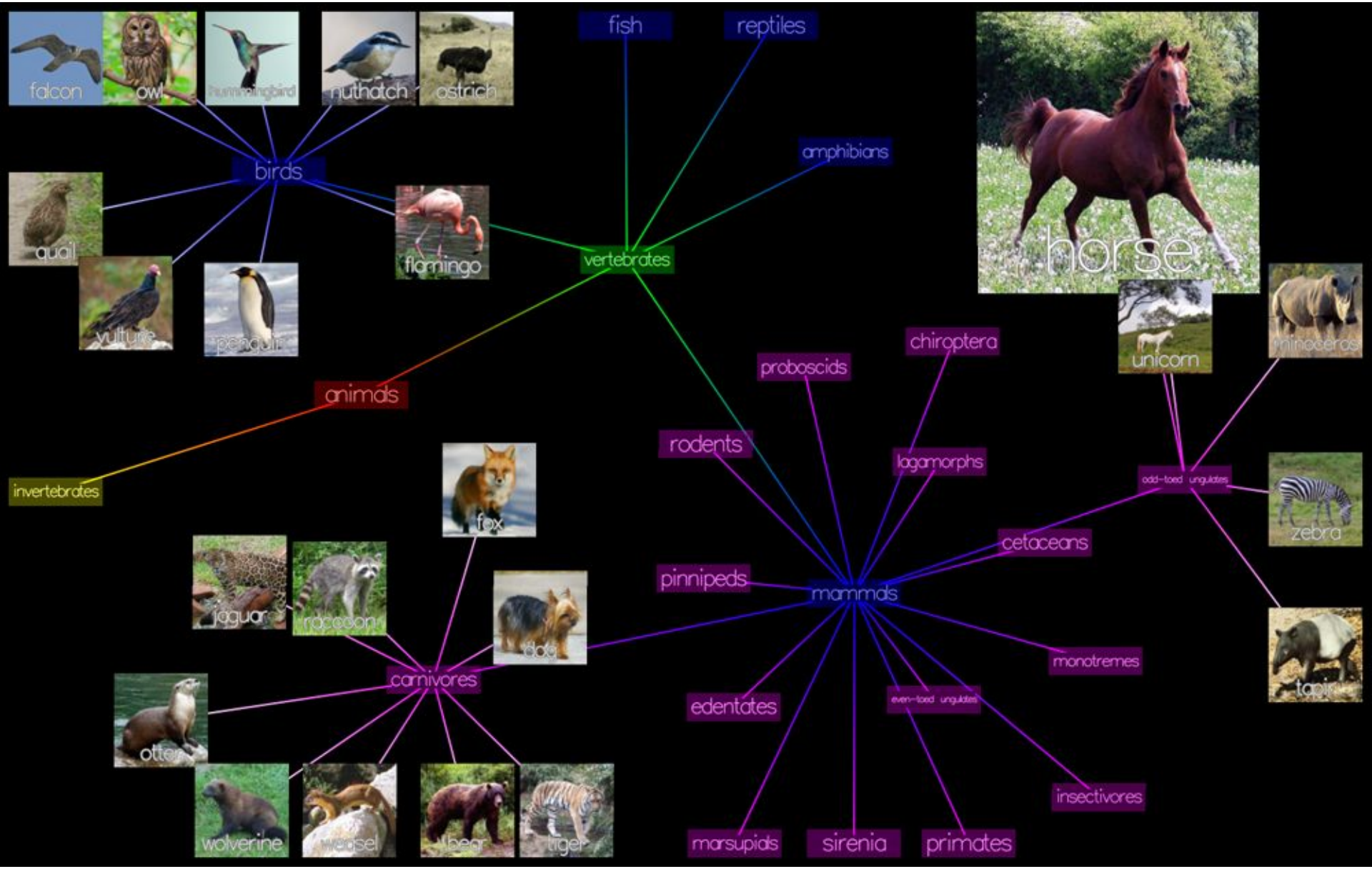
- Upward Drawing (of a tree) – no child is drawn with vertically above (higher y value than) its parent.
- Plane Graph – A 2D drawing of the graph where no edges cross (touching at the endpoint vertices they share is ok)
- Planar Graph – A graph for which a Plane Graph exists.
- Euler's Theorem for planar graphs:
  For a plane graph with n vertices, m edges and f faces, we have $n - m + f = 2$.



Benjamin B. (from Spring ‘12)

Graph Drawing Goals

- Automated!
- Can read all of the labels
- Can follow the line and see exactly which 2 vertices it connects
- Aesthetically pleasing
- Layout should display as much symmetry as possible
- Crossing free or minimal-crossing layout
- All edge lengths are approximately equal
- Even vertex distribution
- Distance between nodes in final layout should be as close as possible to “graph distance” (# of edges on shortest path between those nodes)


falcon
owl
hummingbird
nuthatch
ostrich
fish
reptiles
amphibians
birds
quail
vulture
penguin
flamingo
vertebrates
horse
unicorn
rhinoceros
animals
invertebrates
chiroptera
proboscids
rodents
lagamorphs
odd-toed ungulates
zebra
fox
cetaceans
pinnipeds
mammals
jaguar
raccoon
dog
monotremes
tapir
carnivores
otter
even-toed ungulates
edentates
insectivores
wolverine
weasel
bear
tiger
marsupials
sirenia
primates

Graph Drawing Questions

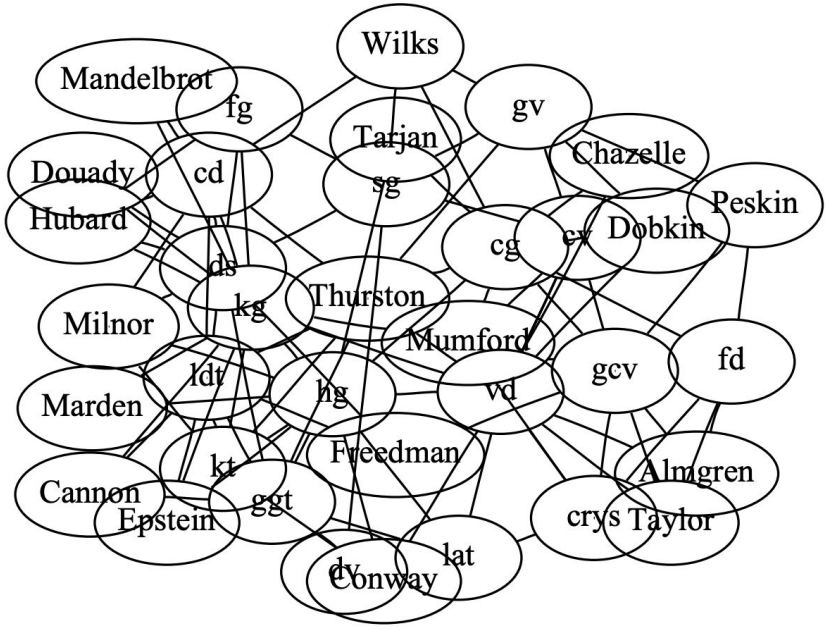
- What is the metric of success for each of our goals?
- Can we guarantee to find a solution? The optimal or best solution?
- Can we use randomness? Does it help?
- How expensive/slow are the different algorithms to draw graphs?
- How does it scale with more nodes/edges?
  - Does it lose effectiveness in meeting our goals?
  - How is the running time affected?
- How do we label the nodes/edges with color/words/images?

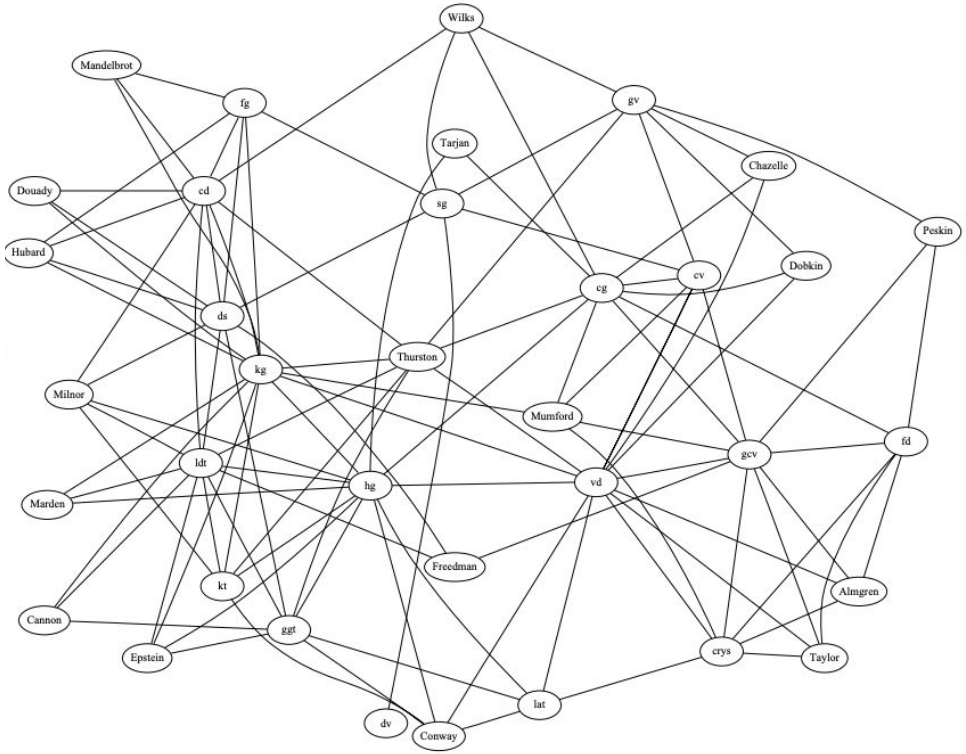
Today

- “Good” Design (30 min)
  - Photography tips
  - Principles of Effective Website Design
  - Principles of Good User Interface Design
  - Examples of Good (Bad) Visualization Design
- Today’s Reading (30 min): “Useful Junk? The Effects of Visual Embellishment on Comprehension and Memorability of Charts”
- Worksheet: Time-based Data & Simple Charts (20 min)
- Graph Drawing Preview: Terminology & Goals (15 min)
- Readings for Tuesday: Graph Drawing

Reading for Tuesday *(pick one)*

"Improved force-directed layouts", Gansner and North, Graph Drawing, 1999.





Reading for Tuesday *(pick one)*

"A Technique for Drawing Directed Graphs" Gansner, Koutsofios, North, & Vo, IEEE Trans. on Software Engineering, 1993.