CSC 544
Data Visualization

Joshua Levine
josh@arizona.edu

(A quick advertisement) Research Opportunities!

- I am currently looking for students to join my research group.
- Specific topics include:
  - Topological data analysis and visualization
  - Volume visualization with deep learning
  - Visualization to support quantum computing w/ acoustic signals
- Anyone is interested in going deeper on visualization research topics and/or enjoys the course, please reach out to me to discuss

Lecture 01
Introduction

Jan. 11, 2023

Course Material Source Credits

- Miriah Meyer
- Carlos Scheidegger
- Torsten Möller
- Tamara Munzner
- Hanspeter Pfister
- Guoning Chen
- Chuck Hansen
- Valerio Pascucci
- Claudio T. Silva
- Raghu Machiraju
- Ross T. Whitaker
- Many others...

Today's Agenda

- Registration Issues? Please send me a note after class.
- Course webpage:
  - https://jalevine.bitbucket.io/csc544/
- Goals for today: Go over syllabus and introduce the course

Course Syllabus

https://jalevine.bitbucket.io/csc544/syllabus.html

Personnel and Contact

- Lecture: Mon./Wed. 5:00-6:15pm
- Office Hours: Tues. 3:30-4:30pm, Wed. 3:00-4:00pm,
    - Or by appointment. I can also possibly meet over Zoom.

- Piazza is the BEST way to contact me, email only for private issues
- D2L used for grades, distributing videos, and calendar

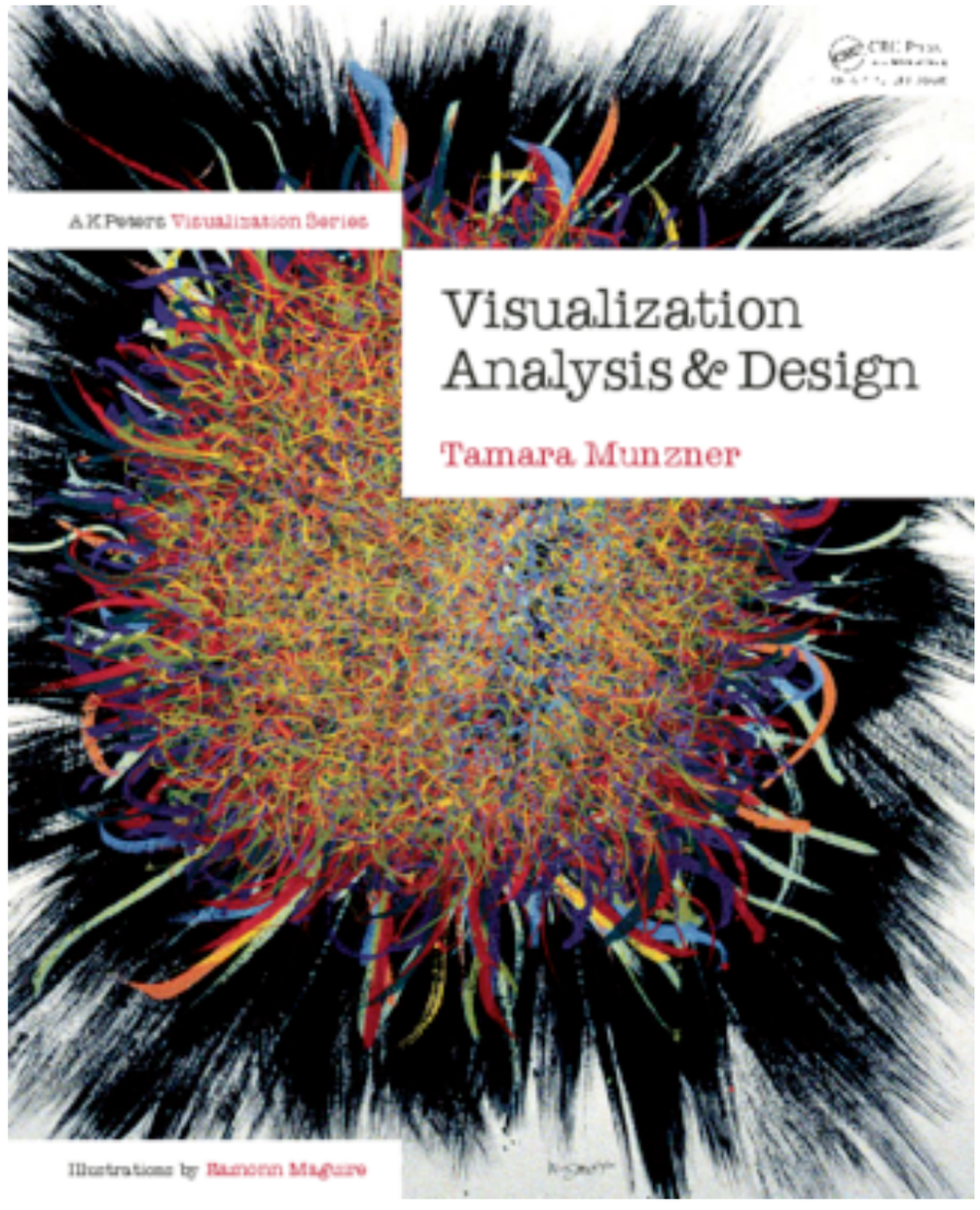
Required Course Materials (Theory)

- Visualization Analysis and Design by Munzner

- ISBN 978-1466508910

- Book is available electronically through the library (but purchasing is recommended!)

- Many other materials distributed through the webpage

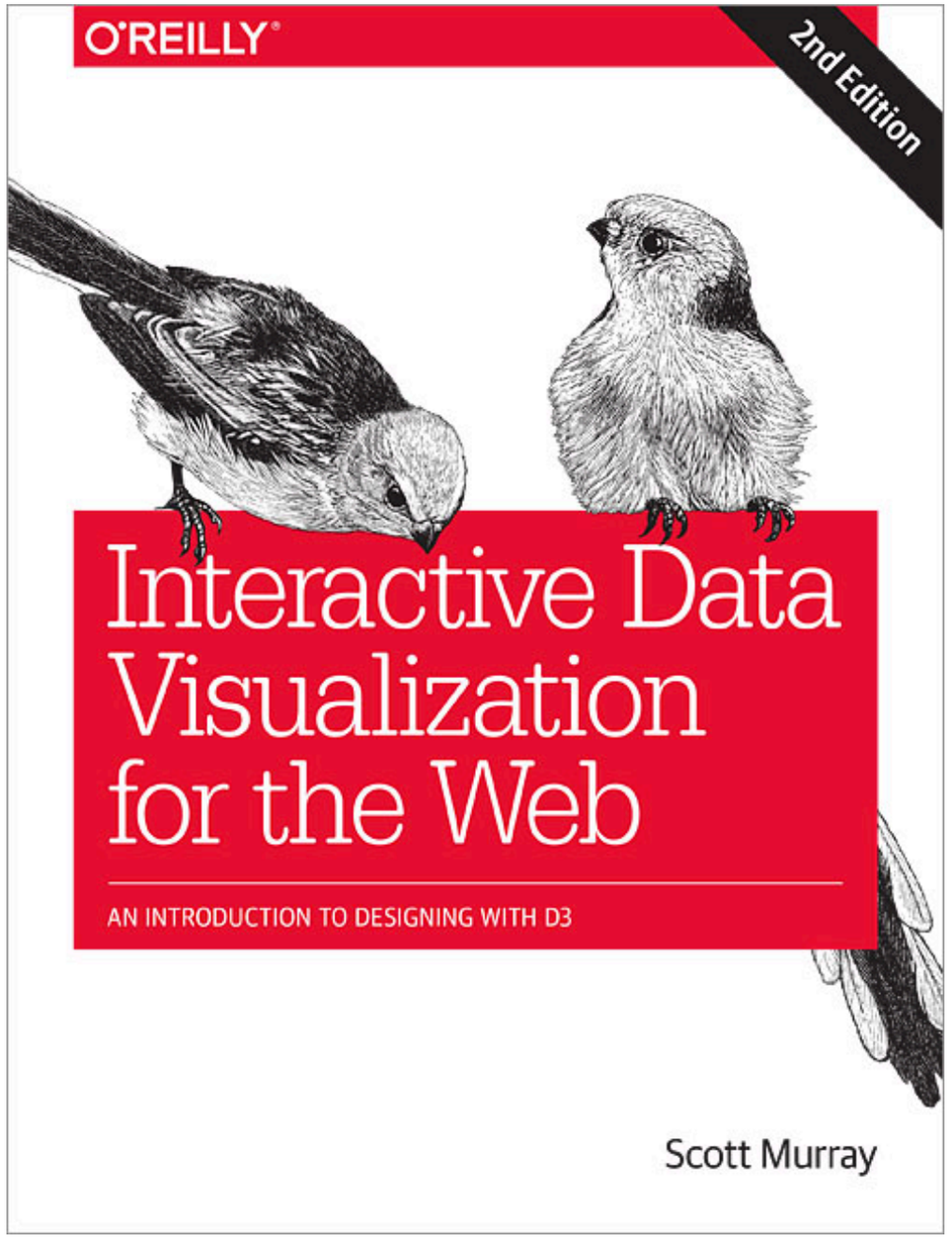


Required Course Materials (Practice)

- Interactive Visualization for the Web by Murray

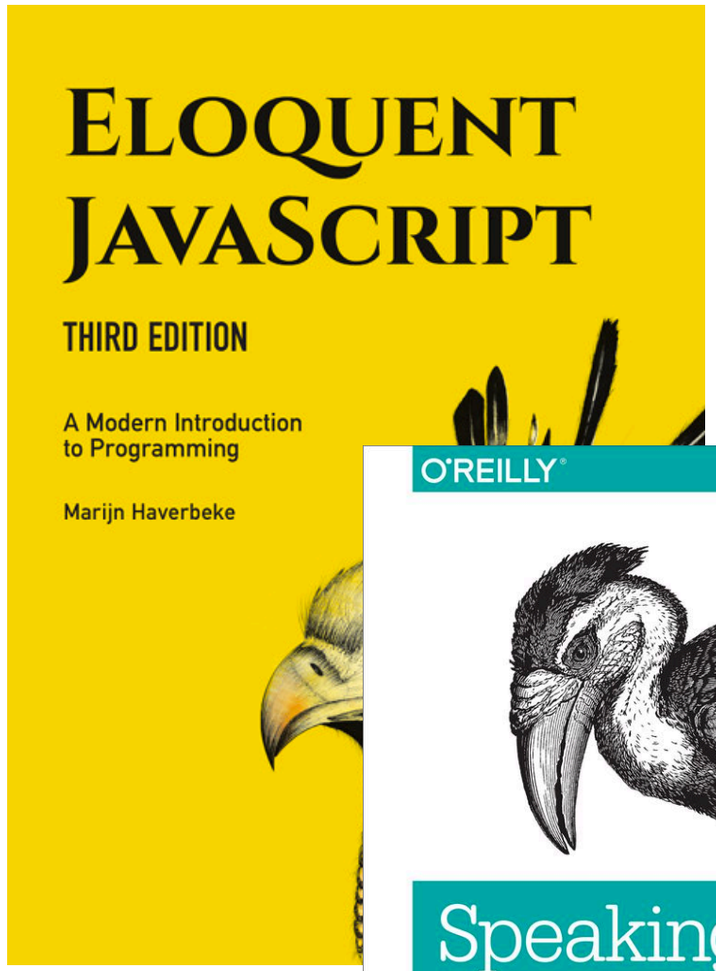
- ISBN 978-1466508910

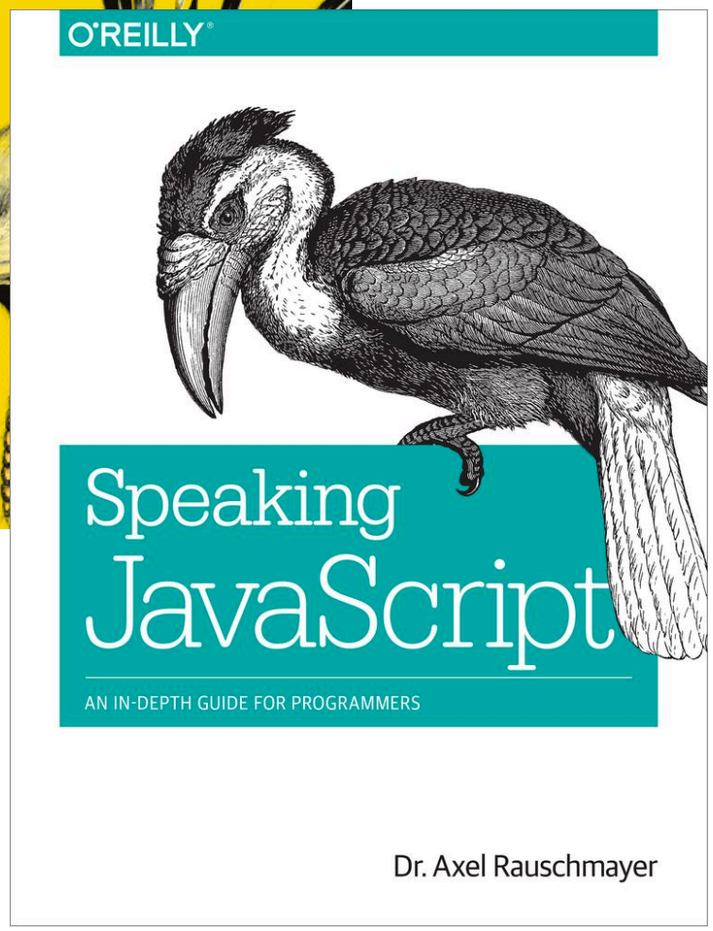
- Book is available electronically through the library (but purchasing is recommended!)



Additional Recommended Course Materials

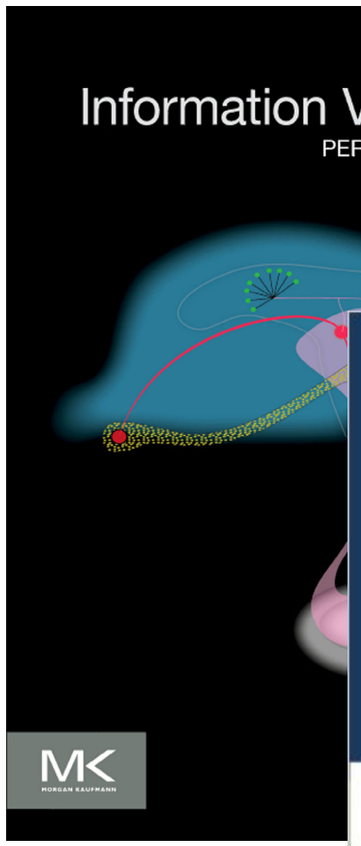
- If you do not know Javascript, you will need a reference!

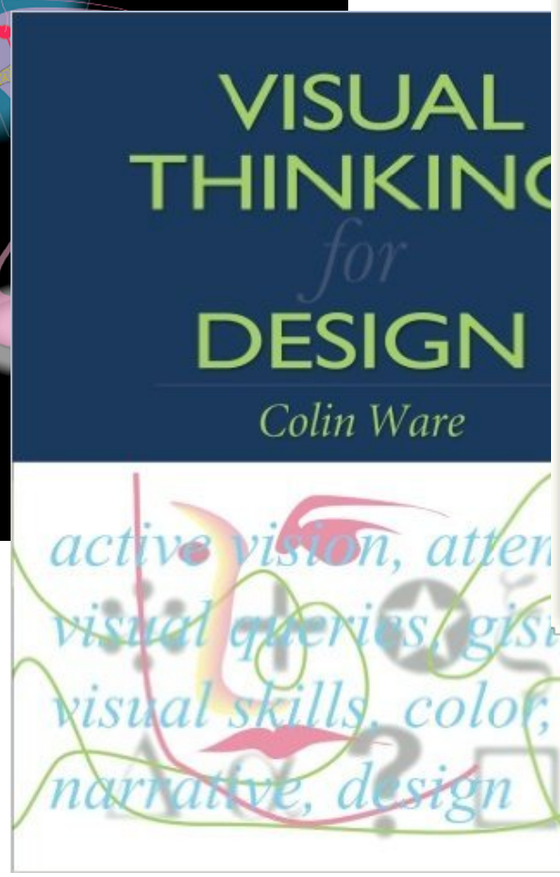


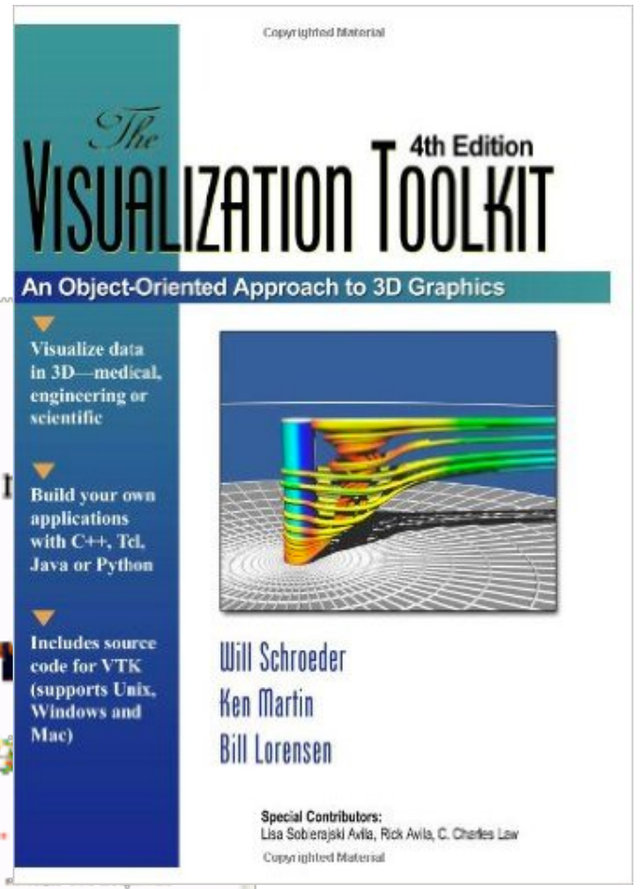


- Two electronic references:
  - https://eloquentjavascript.net/
  - http://speakingjs.com/

Additional References







- Plus relevant research papers from time-to-time

Prerequisites

- No specific courses, and non-majors are welcome.
- Programming skills are necessary! We will develop assignments in Javascript
- Remedial linear algebra will be used, but we will review many of the concepts throughout.
- Previous experience in HCI, graphics, or visualization very useful, but not required.

Course Policies

- **Do**:
  - Come to class prepared and ready to participate
  - Write clean, correct, documented, and tested code
  - Contact me if you have any special needs
  - Be considerate of others and respectful of them and their time.
  - Discuss problems (but not implementations) with classmates before coding.
- **Don't**:
  - Steal code, share answers (from both others in the class and outside of it) without attribution
  - Use copyrighted materials without attribution
  - Be dishonest
  - Criticize people. Instead, critique ideas
  - Violate course/department/ university policies (academic integrity, title IX, etc.)

Coding Policies

- You **must write your own code**. Collaborating with or sharing code with other people in this course is plagiarism and if caught will be reported to academic dishonesty.
    - You are allowed to **discuss** assignments with other students at the **conceptual** level.
    - You must **cite sources** for code snippets or ideas taken from **external sources**.
- OK to: get ideas from the book, from the web
- NOT OK to: share or distribute code, use ideas without attribution.

Course Grading

1. Assignments (~6 total): 55%
    - Expect them to take 2-5 hrs/week
    - A late penalty of 10% per day will apply to assignments. Maximum 5 days late.
2. Design Critiques (rolling deadlines): 5%
3. Class Project (multiple milestones): 35%
4. Class Participation: 5%

Class Participation

- Attendance is **not** mandatory, and will not be collected but…

- I encourage you to come to class, to make your best effort to be here on time, **and** to be a participating member of the audience.

- Discussions on Piazza will also factor into your participation grade

- Both are meant to measure how much you improved the classroom experience for your peers.

Course Recordings

- This class is offered in person and lectures will not be recorded.
- I can make available recordings from a previous offering of this course (Fall 2020) in exceptional circumstances
    - While the material from last semester is similar, I make no guarantee that watching the videos replaces the in person lecture experience.

Data Visualization

How / Why / What?

Vis is Both Ubiquitous and Subtle

- Frank Anscombe,
  “Graphs in Statistical Analysis”

- Let’s try to make sense of this data!

- DISCUSS: what would you first do to understand this dataset?

I		II		III		IV	
x	y	x	y	x	y	x	y
10.0	8.04	10.0	9.14	10.0	7.46	8.0	6.58
8.0	6.95	8.0	8.14	8.0	6.77	8.0	5.76
13.0	7.58	13.0	8.74	13.0	12.74	8.0	7.71
9.0	8.81	9.0	8.77	9.0	7.11	8.0	8.84
11.0	8.33	11.0	9.26	11.0	7.81	8.0	8.47
14.0	9.96	14.0	8.10	14.0	8.84	8.0	7.04
6.0	7.24	6.0	6.13	6.0	6.08	8.0	5.25
4.0	4.26	4.0	3.10	4.0	5.39	19.0	12.50
12.0	10.84	12.0	9.13	12.0	8.15	8.0	5.56
7.0	4.82	7.0	7.26	7.0	6.42	8.0	7.91
5.0	5.68	5.0	4.74	5.0	5.73	8.0	6.89

What did we try?

- https://cscheid.net/courses/fall-2019/csc444/lectures/lecture1/anscombe/

Something interesting just happened: isn't this a visualization too?!

I		II		III		IV	
x	y	x	y	x	y	x	y
10.0	8.04	10.0	9.14	10.0	7.46	8.0	6.58
8.0	6.95	8.0	8.14	8.0	6.77	8.0	5.76
13.0	7.58	13.0	8.74	13.0	12.74	8.0	7.71
9.0	8.81	9.0	8.77	9.0	7.11	8.0	8.84
11.0	8.33	11.0	9.26	11.0	7.81	8.0	8.47
14.0	9.96	14.0	8.10	14.0	8.84	8.0	7.04
6.0	7.24	6.0	6.13	6.0	6.08	8.0	5.25
4.0	4.26	4.0	3.10	4.0	5.39	19.0	12.50
12.0	10.84	12.0	9.13	12.0	8.15	8.0	5.56
7.0	4.82	7.0	7.26	7.0	6.42	8.0	7.91
5.0	5.68	5.0	4.74	5.0	5.73	8.0	6.89

The way in which data is presented changes how we consume it, **drastically**.

A Bit of History

- In 1987, NSF started “Visualization in scientific computing” as a new discipline, and a panel of the ACM coined the term “scientific visualization”
  - *Scientific* visualization, briefly defined: The use of computer graphics for the analysis and presentation of computed or measured scientific data.

- Today: visualization refers to **Techniques** to enable **Users** to find insight into **Data**

The ability to take data—to be able to **understand** it, to **process** it, to **extract value** from it, to **visualize** it, to **communicate** it—that's going to be a hugely important skill in the next decades...

Because now we really do have essentially free and ubiquitous data. So the complimentary scarce factor is the **ability to understand that data and extract value from it.**

Hal Varian, Google's Chief Economist
The McKinsey Quarterly, Jan 2009

Definition: Visualization

- From Merriam-Webster: **vi·su·al·i·za·tion**

    1. formation of mental visual images
    2. the act or process of interpreting in visual terms or of putting into visible form

- From Munzner: “Computer-based **visualization** systems provide visual representations of datasets designed to help people carry out tasks more effectively.”

What's Different about 544 from 444

- Primarily, we will cover a very similar sequence of topics.
- But there will be some **important differences**:
  - I will provide a different take on this material that focuses on emerging research in visualization
  - I will accelerate through some of the non-research topics as a result
  - My expectations for you are also different: I will expect you are capable of conducting visualization research by the end of the semester.

Main Themes for CSC544

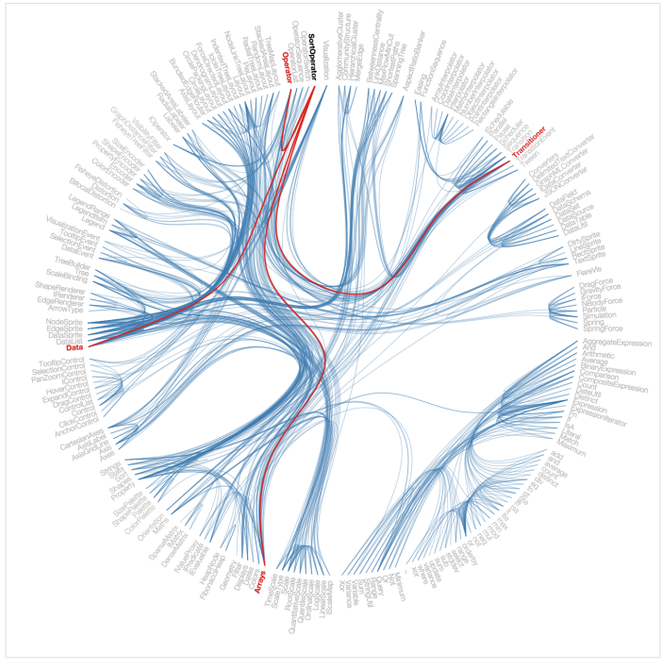
- **Mechanics**: how do I build a visualization?
  - Javascript, CSS, HTML, d3.js, vtk.js
- **Principles**: why should I build it in this way?
  - mathematical and perceptual arguments
- **Techniques**: what do I use to turn principles and mechanics in an actual visualization?
  - algorithms, software libraries

Mechanics

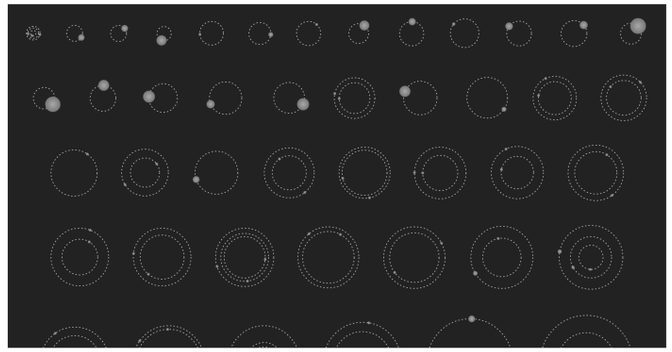
http://bit.ly/1swfb5p

Stick with it, though!

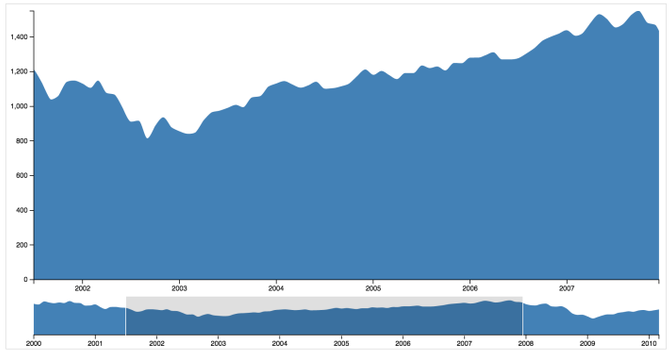
Hierarchical Edge Bundling



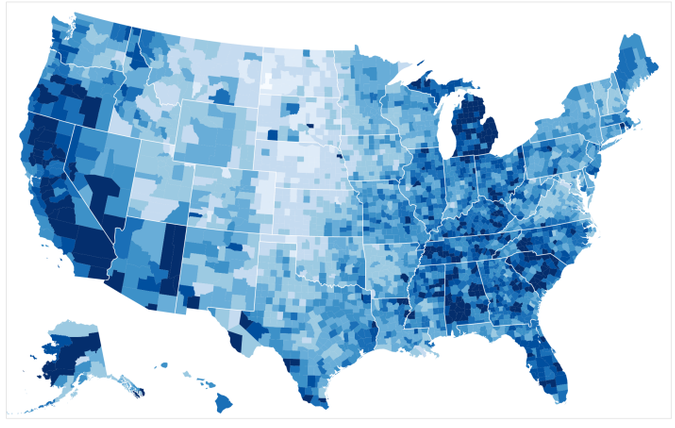
Kepler's Tally



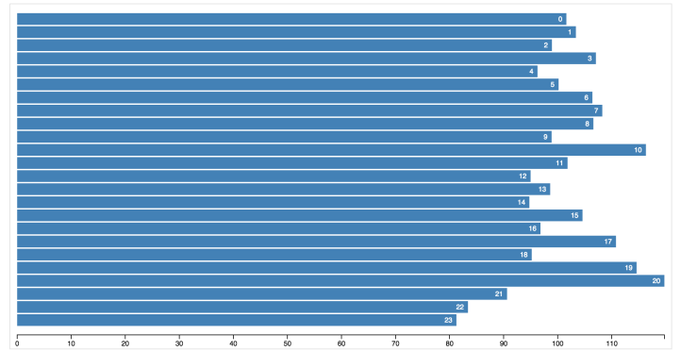
Focus+Context via Brushing



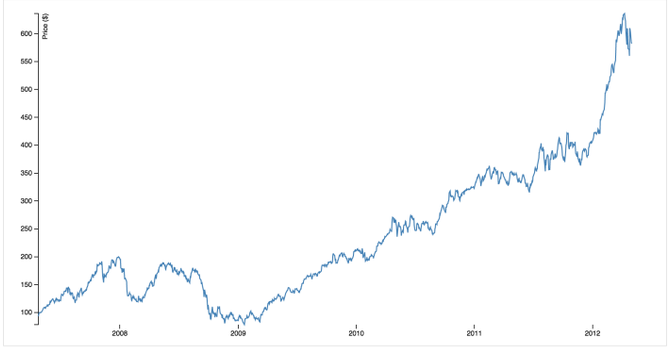
Choropleth



Sortable Bar Chart



X-Value Mouseover



http://bl.ocks.org/mbostock

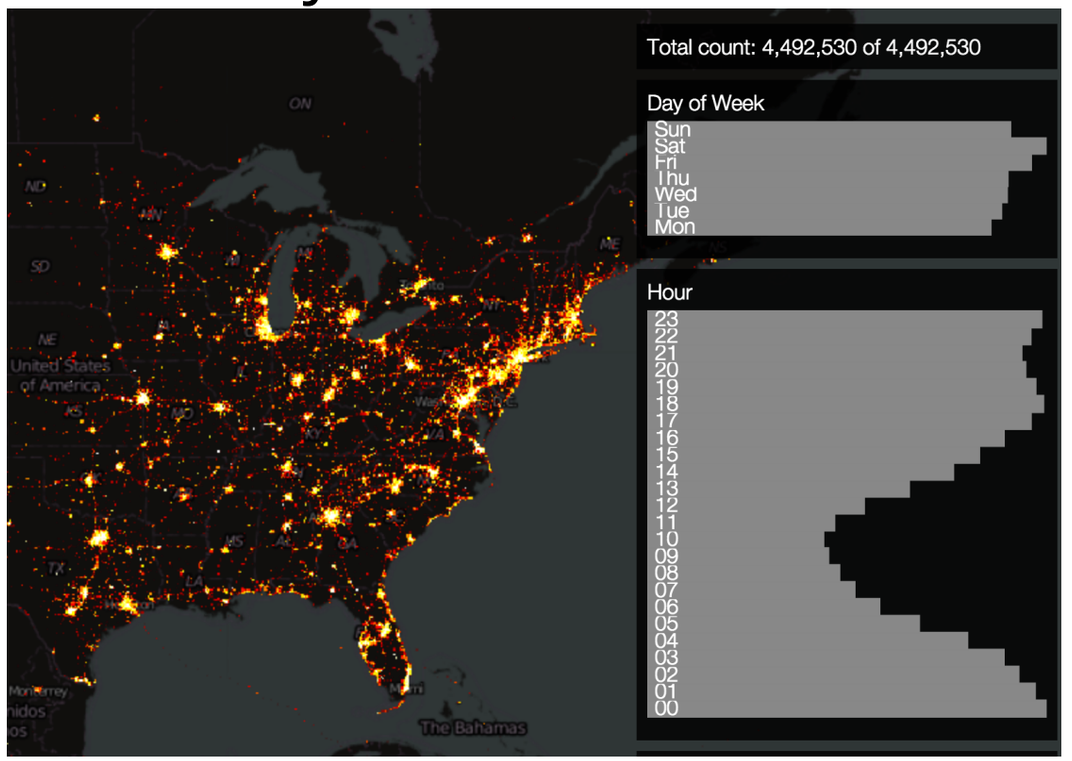
Good reasons to choose the web stack:

It’s ubiquitous

apple.com

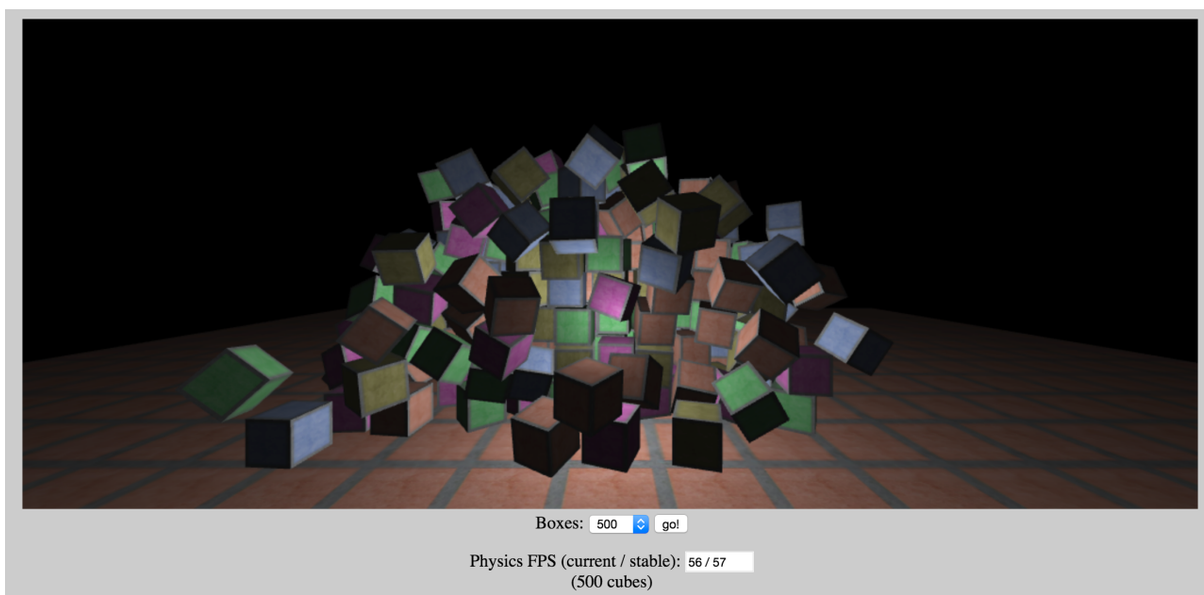
arstechnica.com

It’s easy to talk to a server



nanocubes.net

It’s fast!



Principles

Key issue:

We have limited cognition and memory

youtube.com/watch?v=FWSxSQsspiQ

YouTube Search

The "Door" Study

2,353,335 views • Mar 13, 2010

11K 119 SHARE SAVE ...

Visualization

1. Uses perception to point out interesting things.

MTHIVLWYADCEQGHKILKMTWYN
ARDCAIREQGHLVKMFPSTWYARN
GFPSVCEILQGKMFPSNDRCEQDIFP
SGHLMFHKMVPSTWYACEQTWRN

MTHIVLWYADCEQGHKILKMTWYN
ARDCAIREQGHLVKMFPSTWYARN
GFPSVCEILQGKMFPSNDRCEQDIFP
SGHLMFHKMVPSTWYACEQTWRN

Visualization

1. Uses perception to point out interesting things.

2. Uses pictures (external aids) to enhance working memory.

An Exercise

Compute

$$\begin{array}{r} 37 \\ \underline{\times\ 51} \end{array}$$

An Exercise

Compute

37
x 51

Use a Mental Buffer:

An Exercise

Compute

37
x 51

Use a Mental Buffer:

An Exercise

Compute

$$\begin{array}{r} 37 \\ \times\ 51 \\ \hline \end{array}$$

Use a Mental Buffer:

$$\begin{array}{r} 37 \\ \times\ 51 \\ \hline \end{array}$$

An Exercise

Compute

$$\begin{array}{r} 37 \\ \times\ 51 \\ \hline \end{array}$$

Use a Mental Buffer:

$$\begin{array}{rr} & 37 \\ \times & 51 \\ \hline 1 * 7 = & 7 \end{array}$$

An Exercise

Compute

$$\begin{array}{r} 37 \\ \times\ 51 \\ \hline \end{array}$$

Use a Mental Buffer:

$$\begin{array}{r} 37 \\ \times\ 51 \\ \hline \end{array}$$

$1 * 7 = 7$

$1 * 3 * 10 = 3\_$

An Exercise

Compute

37
x 51

Use a Mental Buffer:

37
x 51

1 * 7 = 7

1 * 3 * 10 = 3_

5 * 10 * 7 = 35_

An Exercise

Compute

37
x 51

Use a Mental Buffer:

37
x 51

1 * 7 = 7

1 * 3 * 10 = 3_

5 * 10 * 7 = 35_

5 * 10 * 3 * 10 = 15__

An Exercise

Compute

37
x 51

Use a Mental Buffer:

37
x 51

1 * 7 = 7

1 * 3 * 10 = 3_

5 * 10 * 7 = 35_

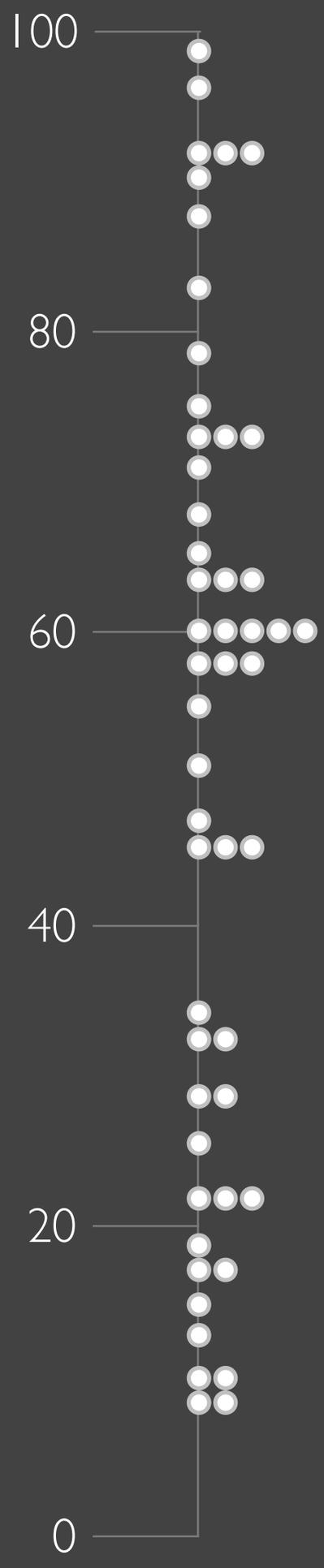
5 * 10 * 3 * 10 = 15__

= 7 + 30 + 350 + 1500 = 1887

15 19 60
33 11 75
57 34 79
18 51 92
73 22 13
71 60 22
17 10 68
73 18 55
65 46 29
60 73 22
46 92 97
10 58 46
57 17 83
26 99 33
88 92 60
91 29 57
96 12 47

given these 50 numbers . . .

. . . what number appears most often?



given these 50 numbers . . .

. . . what number appears most often?

Visualization

1. Uses perception to point out interesting things.

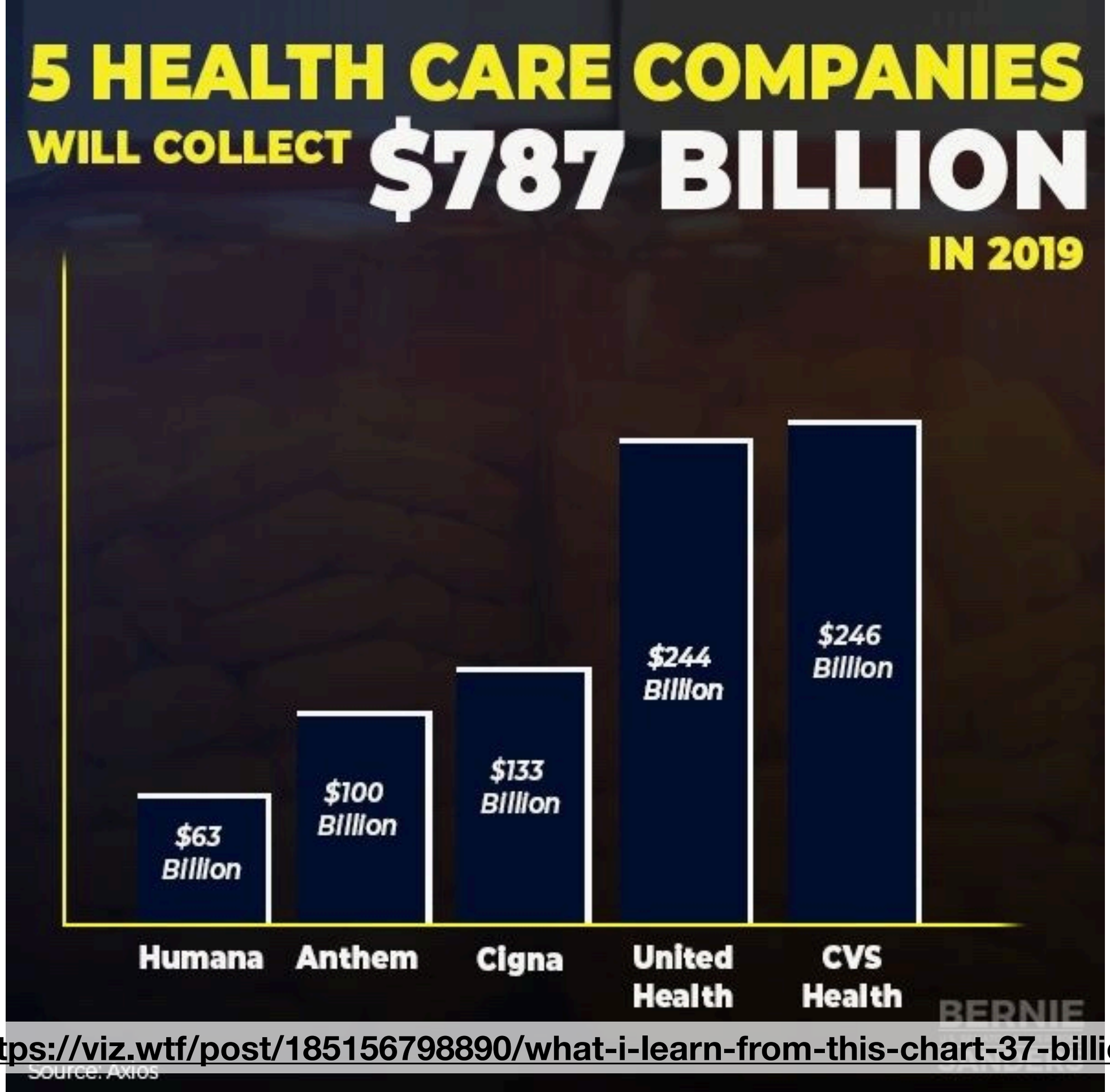
2. Uses pictures (external aids) to enhance working memory.

3. Must respect the math in the data

Not everything you can do with data makes sense

https://imgur.com/gNefvUG/

5 HEALTH CARE COMPANIES WILL COLLECT $787 BILLION IN 2019



https://viz.wtf/post/185156798890/what-i-learn-from-this-chart-37-billion

Source: Axios

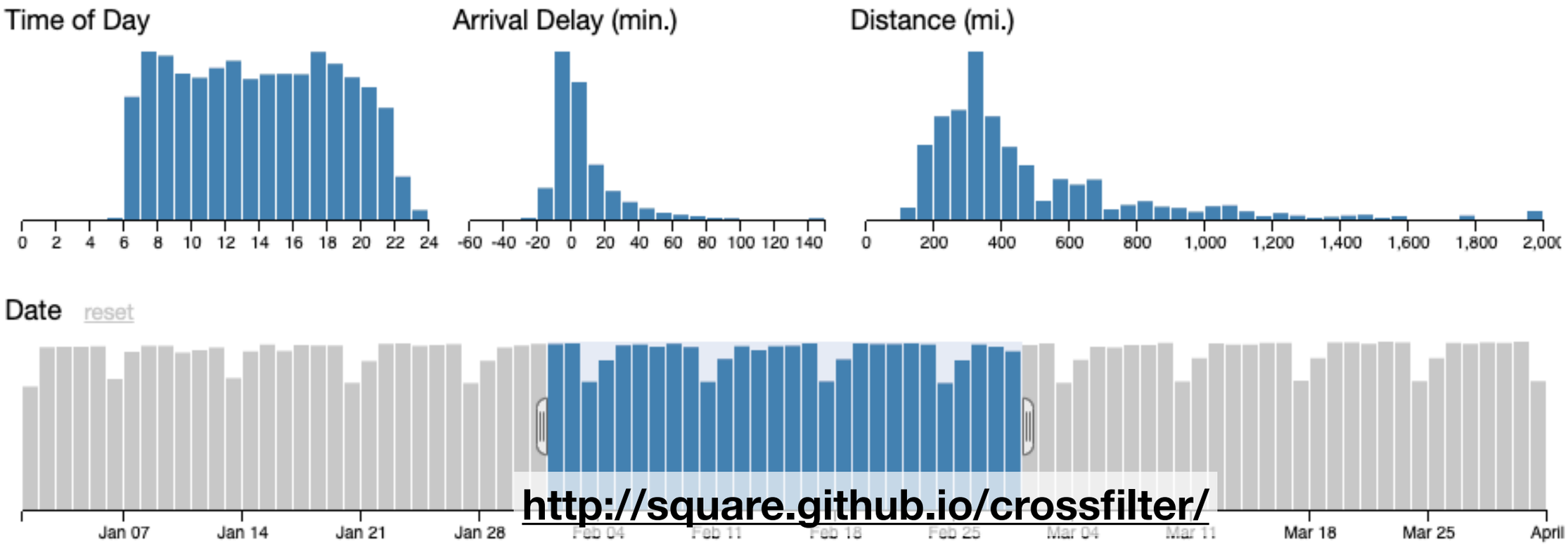
Techniques

Linked Views

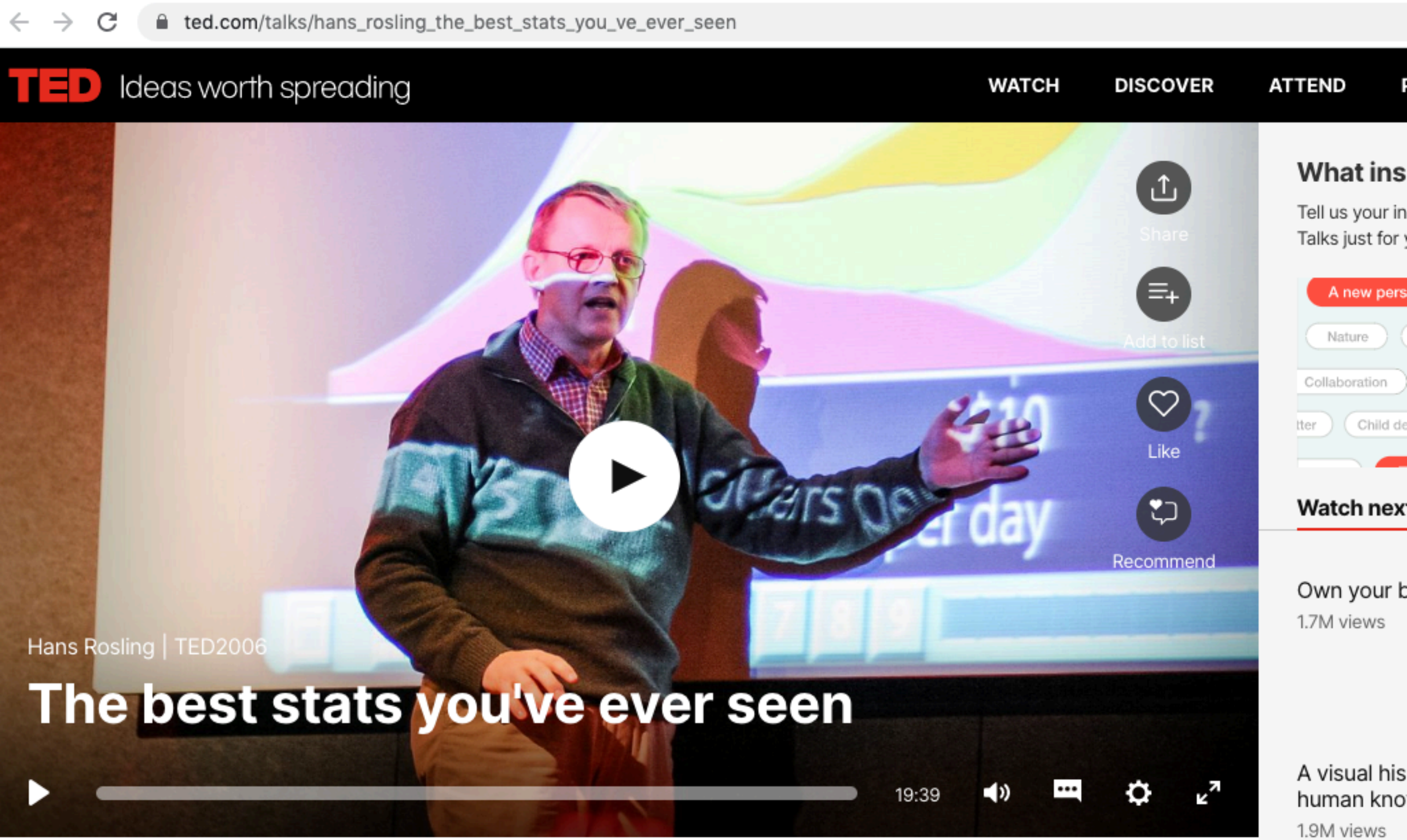
Example: Airline on-time performance

The coordinated visualizations below (built with D3) show nearly a quarter-million flights from early 2001: part of the ASA Data Expo dataset. The dataset is 5.3MB, so it might take a few seconds to download. Click and drag on any chart to filter by the associated dimension. The table beneath shows the eighty most recent flights that match the current filters; these are the *details on demand*, anecdotal evidence you can use to weigh different hypotheses.

Some questions to consider: How does time-of-day correlate with arrival delay? Are longer or shorter flights more likely to arrive early? What happened on January 12? How do flight patterns differ between weekends and weekdays, or mornings and nights? Fork this example and try your own data!

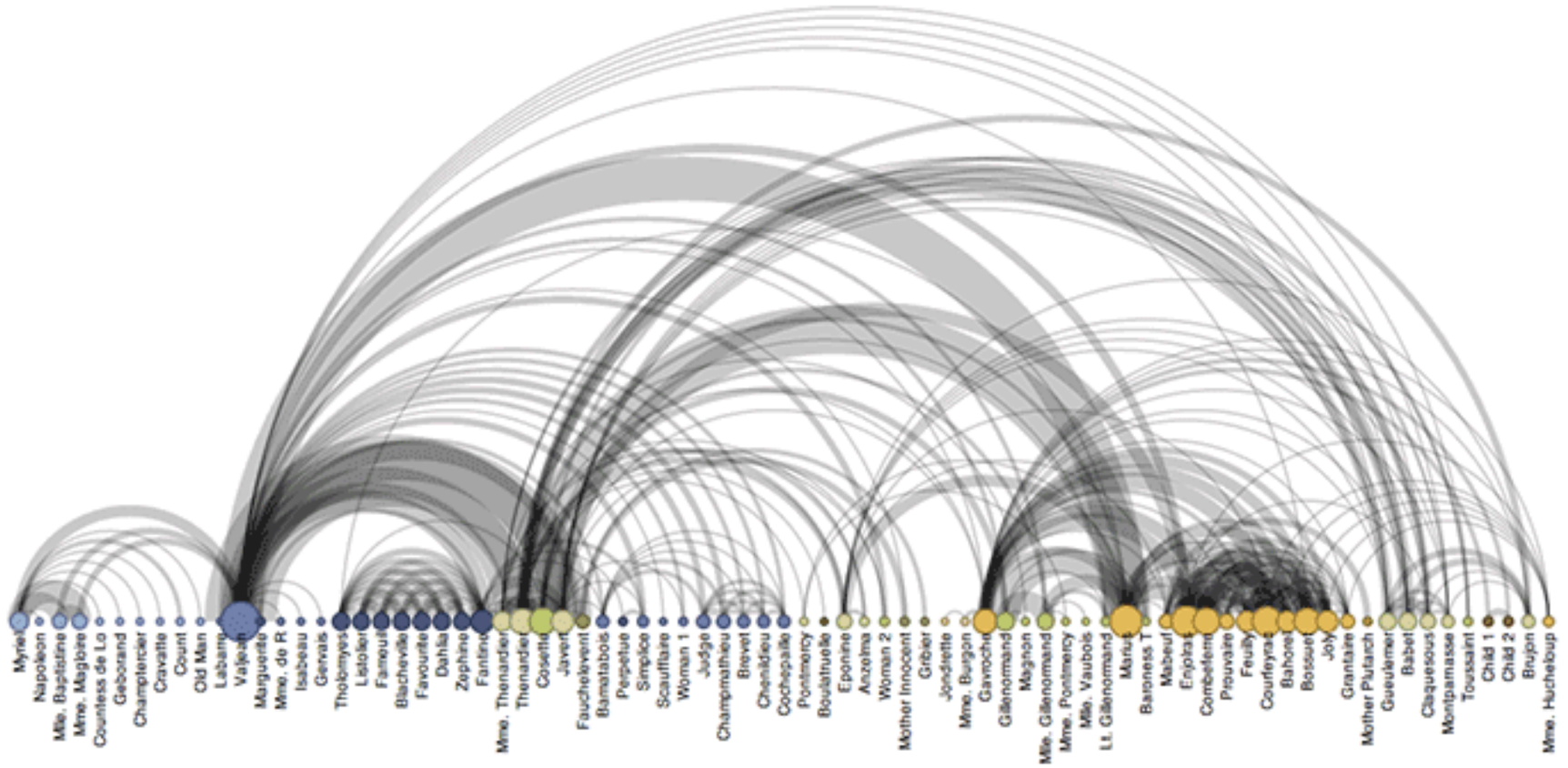


Inspire / Tell a Story



A Tour through the Visualization Zoo

Jeffrey Heer, Michael Bostock, Vadim Ogievetsky



Arc diagram of character co-occurrences in chapters of Victor Hugo's Les Miserables.

ABSTRACT

A survey of powerful visualization techniques, from the obvious to the obscure.

MATERIALS

PDF | BibTeX

CITATION

A Tour through the Visualization Zoo
Jeffrey Heer, Michael Bostock, Vadim Ogievetsky
Communications of the ACM, 53(6), pp. 59-67, 2010.

https://idl.cs.washington.edu/papers/visualization-zoo/

A TOUR OF VISUALIZATIONS, GOOD, BAD, AND UGLY

By the end of this course, you will have the skills to create many of these visualizations yourself, to tell whether they are a good or a bad design, and **why**.

THE GOOD

- The Periodic Table. You might not think of the Periodic Table as a good visualization, or even as a vis at all! Still, it's a really great one. The spatial arrangement of the elements make your eyes think for you: electron affinity up and to the right; metallic character down and to the left; etc.
- Gapminder.
- The Upshot's Rent vs. Own calculator.
- The Upshot's per-precinct election maps.
- The Wind map.
- Stop-and-Frisk decline in New York City.
- America's Cup course in San Francisco.
- Gun deaths in the US.

THE BAD, AND THE UGLY

- Smartphone buyers.
- A hairball. We'll learn in the Principles section why this is a terrible visualization, and in the Techniques section how to create something better.
- Bad colormaps.
- More bad colormaps (and how to fix them).
- A compendium of bad infographics and visualizations.
- A non-periodic periodic table.
- Obvious chart is obvious.

https://cscheid.net/courses/fall-2019/csc444/lectures/lecture1.html#a-tour-of-visualizations-good-bad-and-ugly

Lec02 Reading

- The basics of the web stack and javascript:
    - https://cscheid.net/courses/fall-2019/csc444/lectures/lecture2.html
    - https://cscheid.net/courses/fall-2019/csc444/lectures/lecture3.html
- Murray, Chapter 3
- See also (recommended)
    - Books on javascript
    - Exercises from CSC 444 Lecture 3

Assignment 00

Assigned: Wednesday, January 11
Due: Monday, January 23, 4:59:59 pm