

MassMutual DSDP 2021:

INTRODUCTION TO DATA VISUALIZATION

June 28, 2021
R. Jordan Crouser
Associate Professor of Computer Science
Smith College

Some housekeeping

- Workshop website:
jcrouser.github.io/MassMutual-DataVis
- Rough structure:
 - 6 modules over 3 days (AM and PM)
 - Intro → Walkthrough/Lab → Explore → Share
- Assumptions:
 - R/Rstudio installed
 - Basic proficiency in R

Learning objectives



1. Understand
why data vis works
(and doesn't)

2. Explore some
foundational
methods / tools

3. Opportunities to
get to know
your new team

What I do: analytical tools for messy data



Big idea behind my research

Humans and machines have **complimentary strengths**

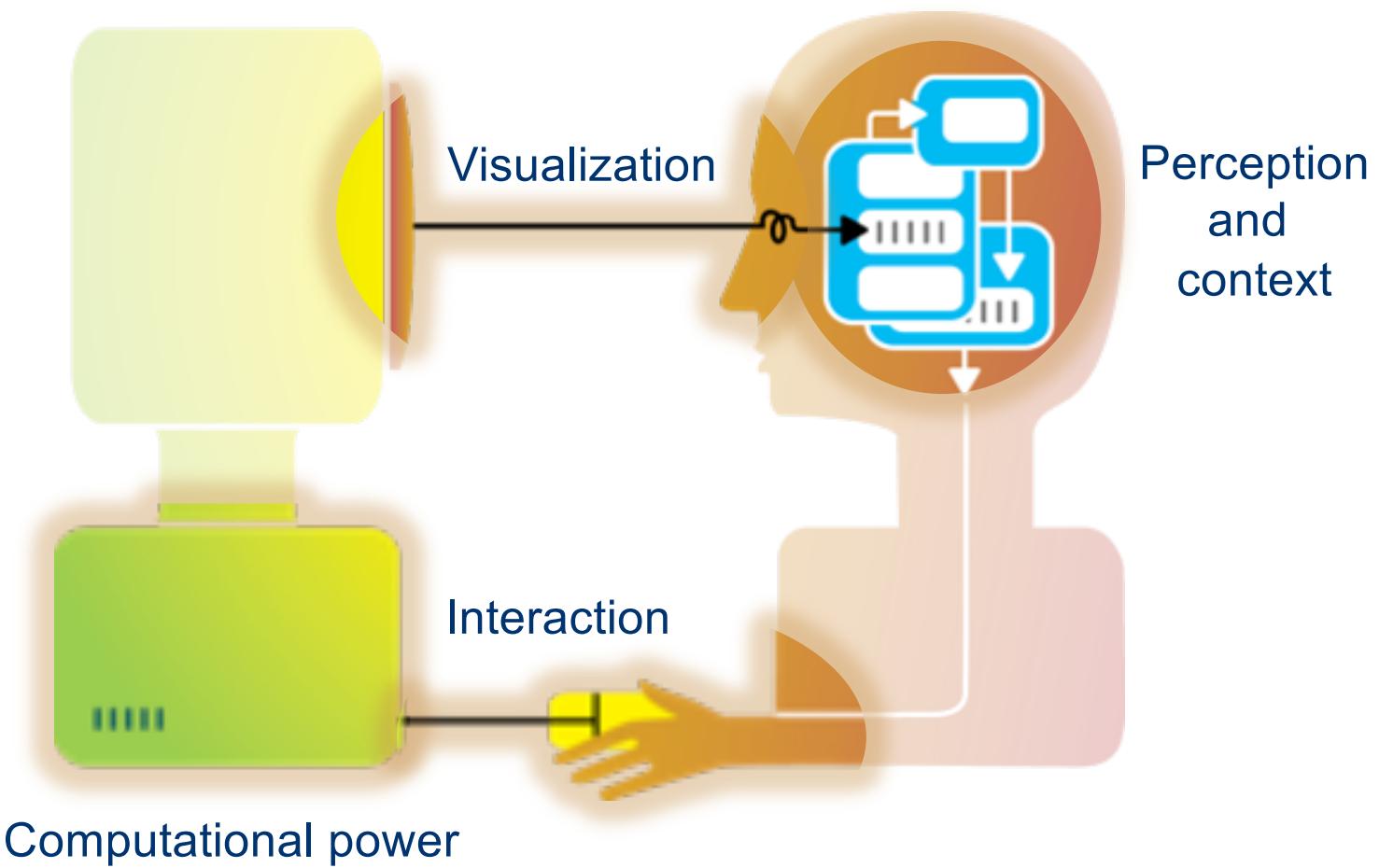
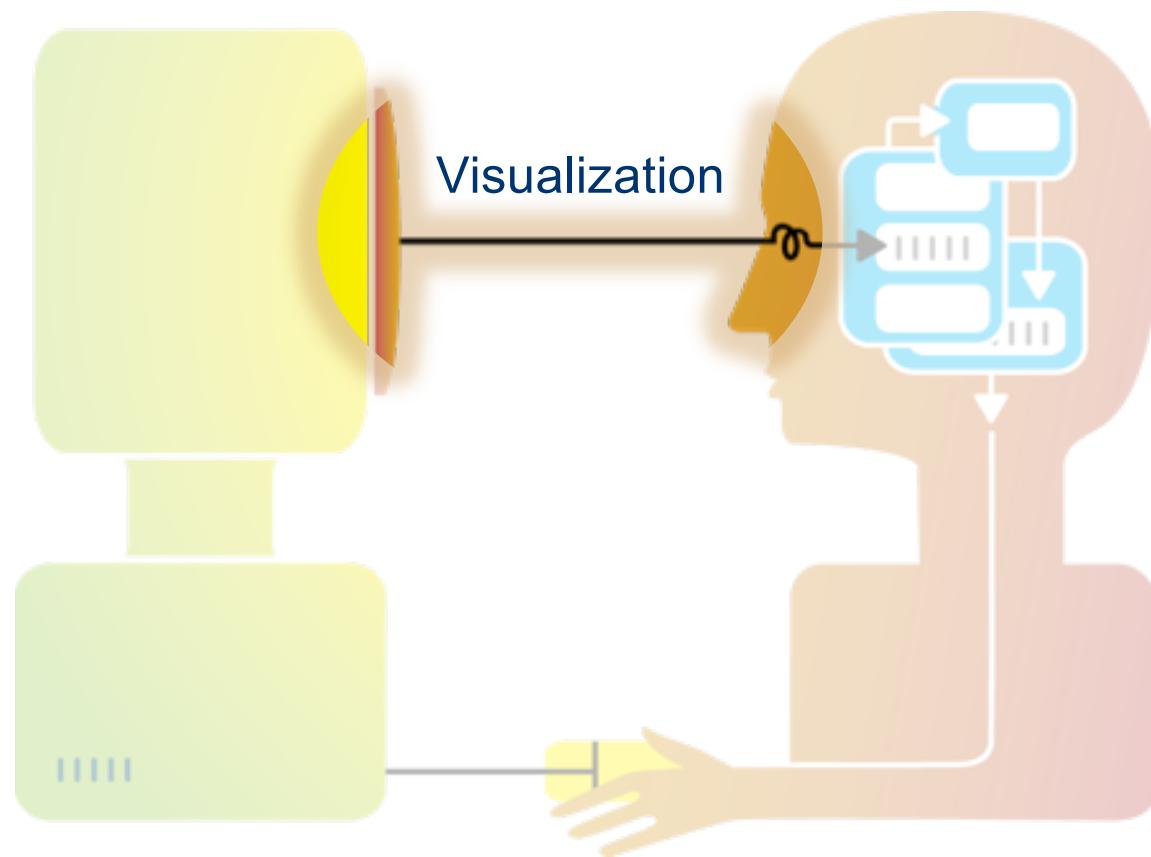


Image credit: Ali Ansari

Focus of this workshop

How do we build **visualizations** that help humans understand patterns in data?



3-minute biographies

About you:

- Your name and pronouns
- Your alma mater
- Your major / area of focus

3 questions:

1. What brought you to **this workshop?**
2. What's one **big thing** you hope to get out of it?
3. What's one thing about you that would probably **surprise us?**

Outline

✓ Introductions

- Visualization overview
 - Flashback to early experiences in data wrangling
 - Visualization (def.)
 - Data (def.)
 - Quick history lesson
- Graphical primitives
- Visual dimensions
- Pre-lunch activity: deconstructing data graphics
- After lunch: ggplot2 crash course

What is visualization?

Google visualization

All Videos Images News Books More Search tools SafeSearch ▾ ⚙

Reading Strategy Data Quotes Sports Creative Techniques

Visualize to create mental pictures with you need. These will always be your tool.

VISUALIZE THE MOST AMAZING LIFE IMAGINABLE TO YOU. Close your eyes and see it clearly. THEN HOLD THE VISION FOR AS LONG AS YOU CAN. NOW PLACE THE VISION IN GOD'S HANDS...AND CONSIDER IT DONE.

VISUALIZATION IS DAYDREAMING WITH A PURPOSE.

SafeSearch ▾ ⚙

The image shows a Google search results page for the query "visualization". The search interface includes a navigation bar with "All", "Videos", "Images" (which is selected), "News", "Books", "More", and "Search tools". There are also "SafeSearch" and "⚙️" buttons. Below the search bar, there are six category labels: "Reading Strategy", "Data", "Quotes", "Sports", "Creative", and "Techniques". Each category has a representative image. The "Quotes" section features a quote from the book "Visualize" by Alan Lakein. The "Techniques" section features a quote from "The Power of Visualization Workbook" by Alan Lakein. The "Sports" section features a man holding a ball. The "Creative" section features a person standing next to a colorful bird. The "Techniques" section features a person's head with a brain diagram. The main search results area displays a grid of 20 images. These images include: three people standing on a grassy field with icons above them; a close-up of a green eye; a colorful, glowing nebula or network visualization; a woman's eye with light rays emanating from it; a woman's face with her hand on her forehead; a colorful, multi-colored eye; a complex network visualization; a woman's face with a wireframe overlay; a silhouette of a person standing next to a bar chart labeled "PRESENT" and "FUTURE"; a black background with the word "visualize" in large, colorful letters surrounded by various positive words like "dreams", "life", "perfect", "clear", etc.; a circular diagram of colored dots connected by lines; a blue eye with the text "THE POWER OF VISUALIZATIONS" overlaid; a silhouette of a person's head with gears inside; a colorful profile of a person's head with the word "visualization"; a silhouette of a person's head with gears inside; and a couple looking at a futuristic interface displaying a globe.

What is visualization?

Google search results for "information visualization".

Search bar: information visualization

User interface elements: Camera icon, Search icon, SafeSearch dropdown, Settings gear, User profile (Jordan)

Navigation tabs: All, News, **Images**, Books, Videos, More ▾, Search tools

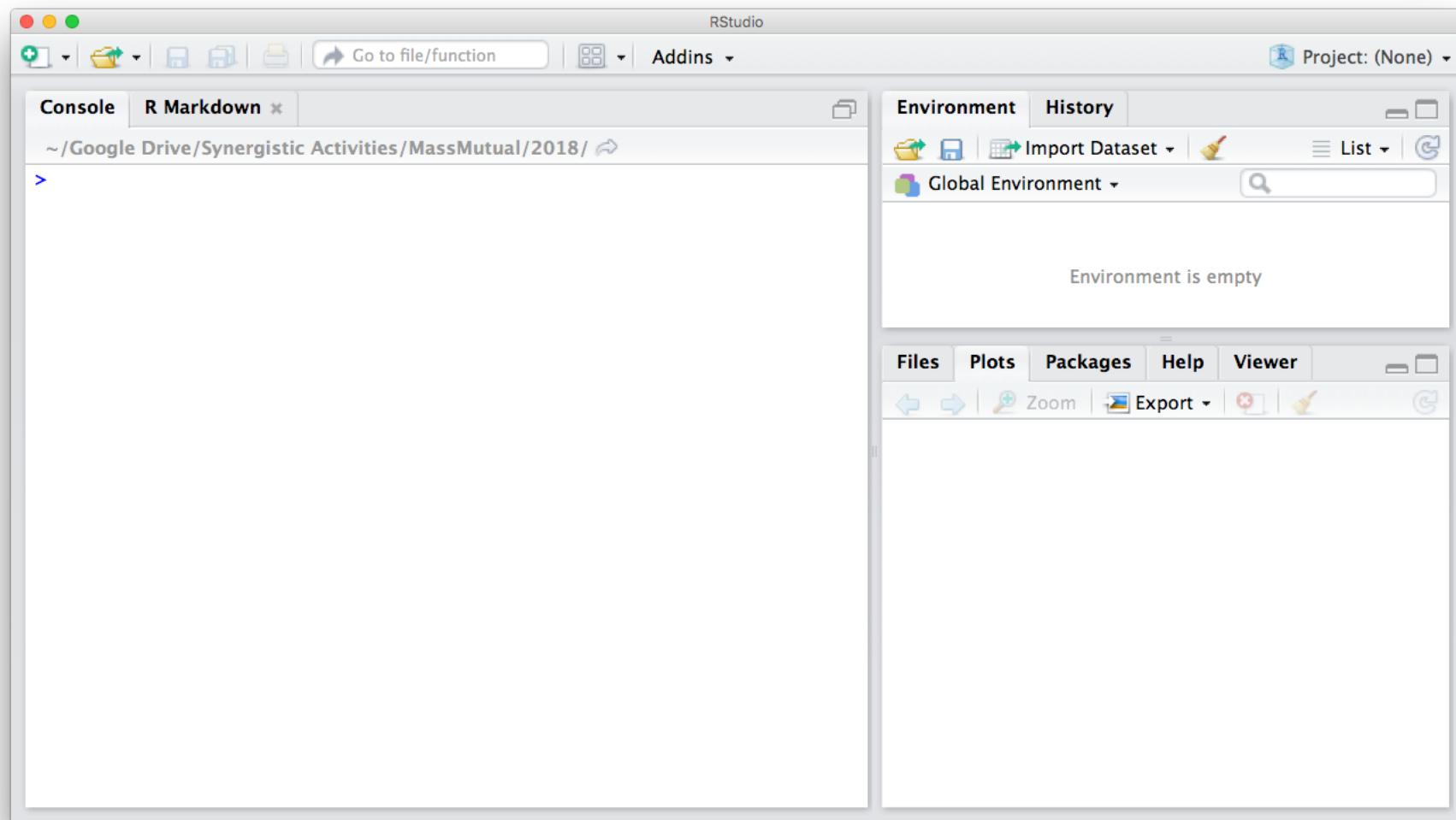
Image categories displayed:

- Examples: A circular sunburst chart and a network graph.
- Design: A diagram showing nodes connected by lines.
- Tools: A collection of various visualization interfaces, including a 3D visualization tool labeled "information visualization tools".
- 3D: Three-dimensional visualization interfaces.
- Data Visualization Map: A world map with data points and connections.
- Data: A circular sunburst chart and a network graph.

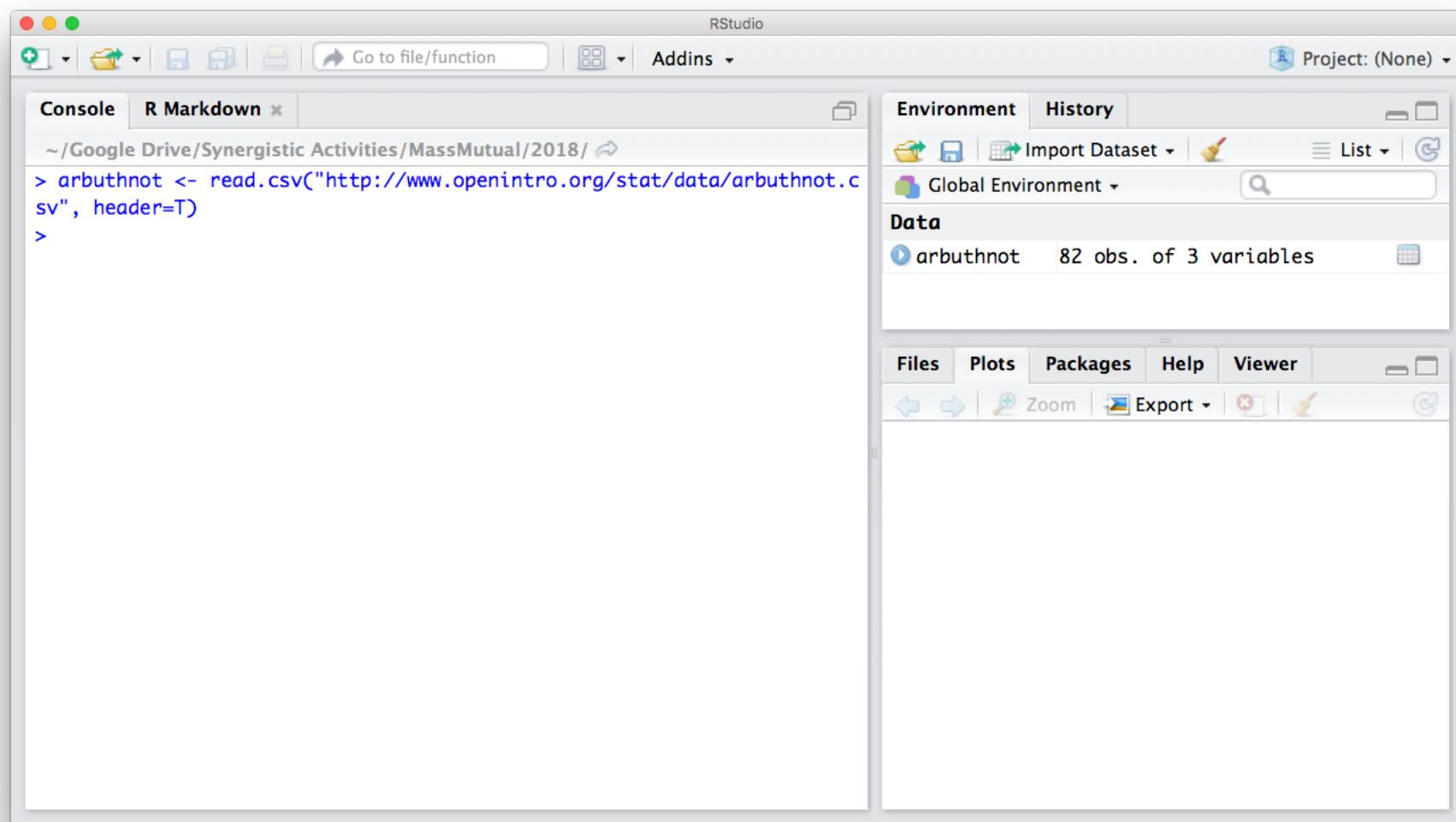
Below these categories, a grid of 20 visualization examples:

- Row 1: Circular sunburst chart, network graph, complex network visualization, conceptual map.
- Row 2: Circular sunburst chart, network graph, complex network visualization, world map.
- Row 3: Circular sunburst chart, network graph, complex network visualization, conceptual map.
- Row 4: Circular sunburst chart, network graph, complex network visualization, conceptual map.
- Row 5: Circular sunburst chart, network graph, complex network visualization, conceptual map.
- Row 6: Circular sunburst chart, network graph, complex network visualization, conceptual map.
- Row 7: Circular sunburst chart, network graph, complex network visualization, conceptual map.
- Row 8: Circular sunburst chart, network graph, complex network visualization, conceptual map.
- Row 9: Circular sunburst chart, network graph, complex network visualization, conceptual map.
- Row 10: Circular sunburst chart, network graph, complex network visualization, conceptual map.

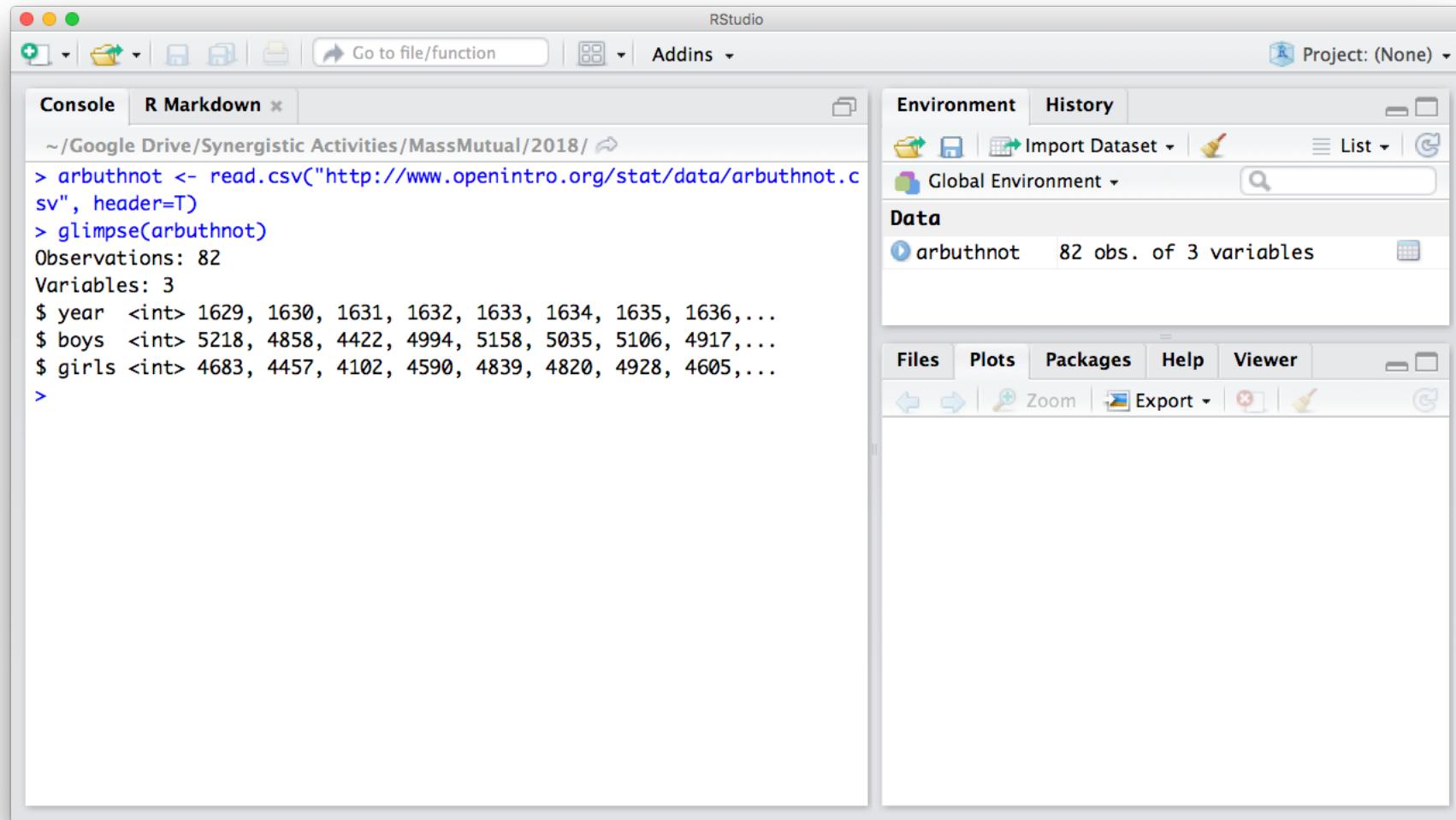
Flashback...



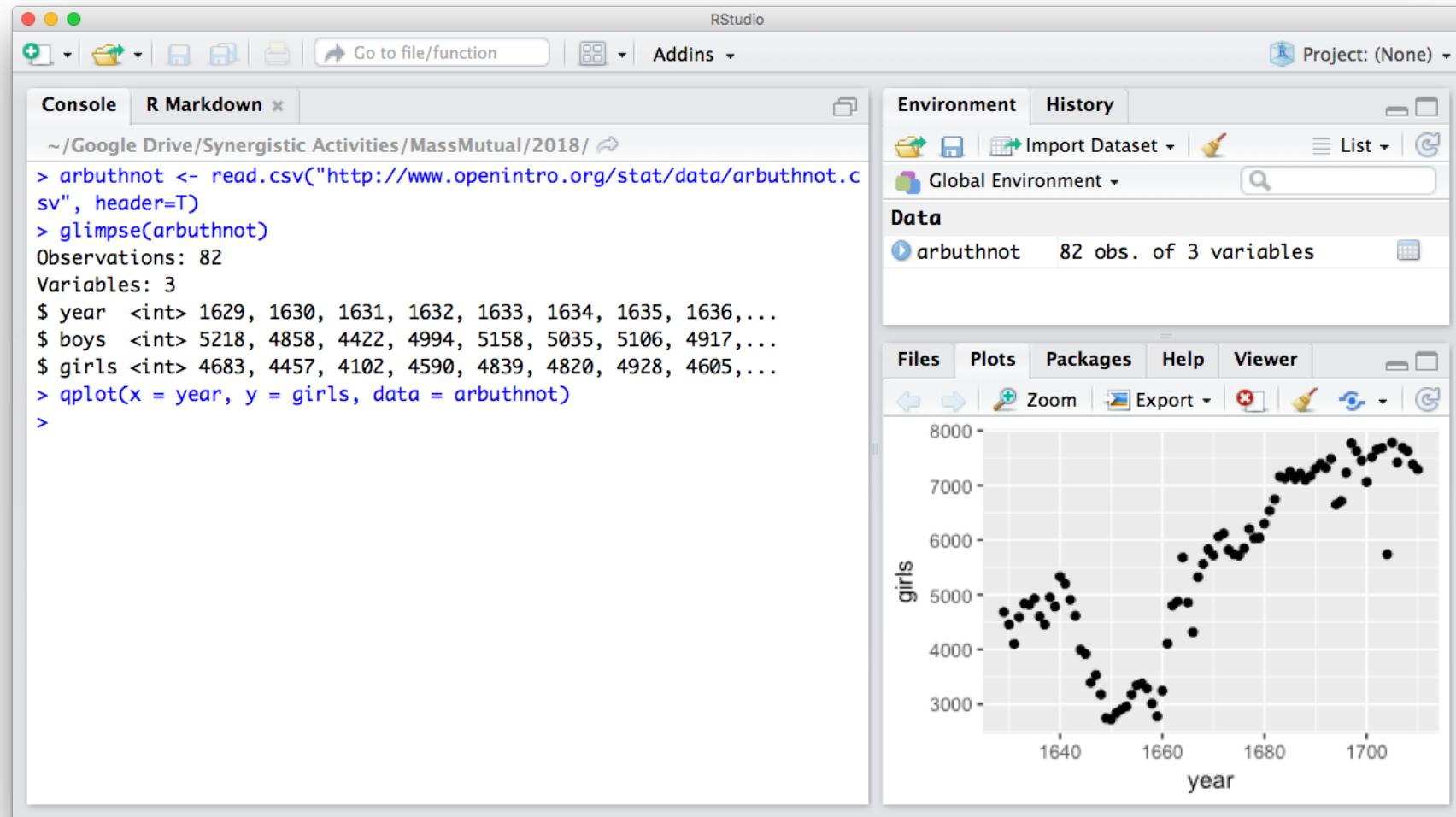
Flashback...



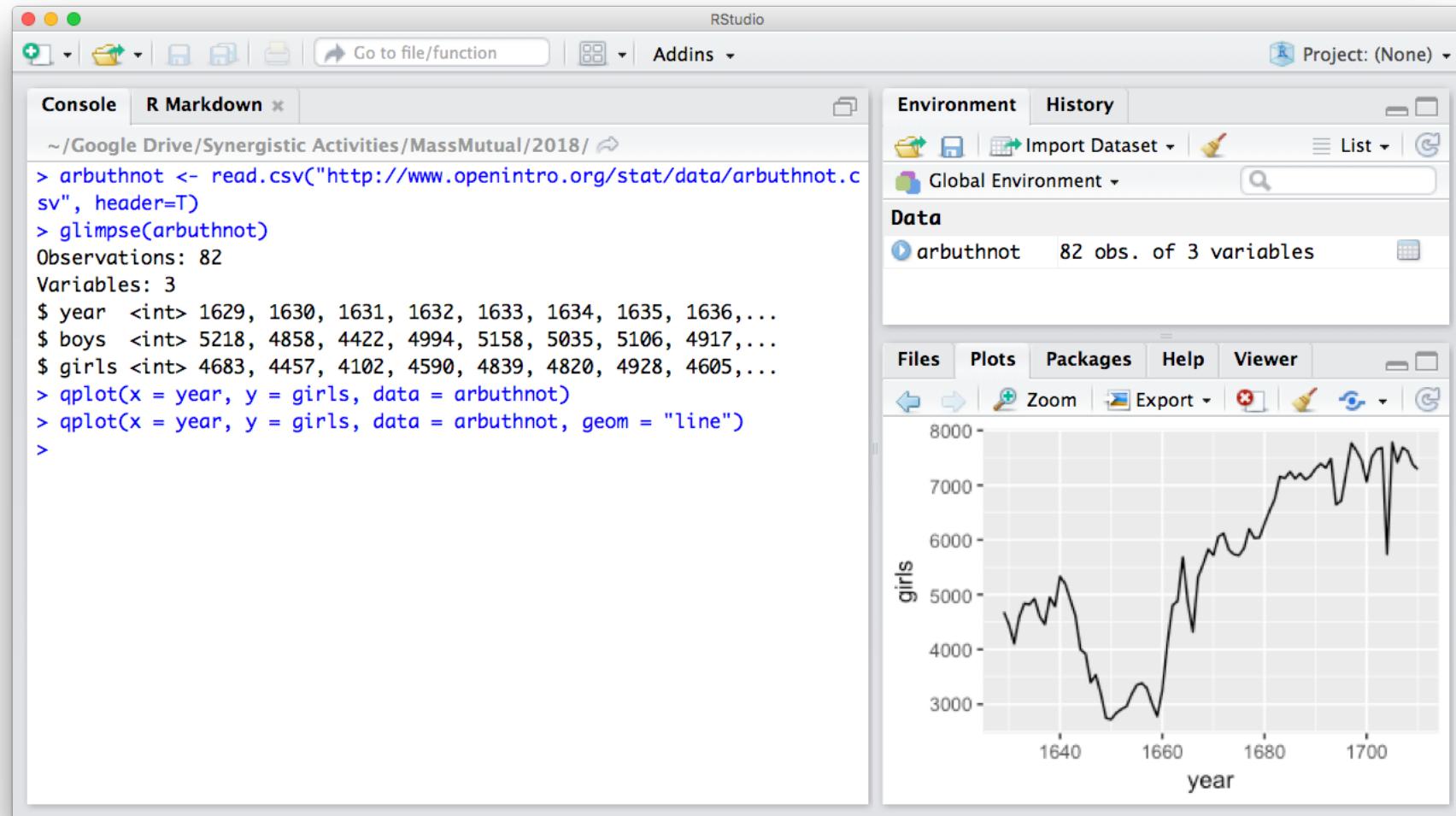
Last week...



Last week...



Last week...



Question

What makes these
“visualizations” **useful?**

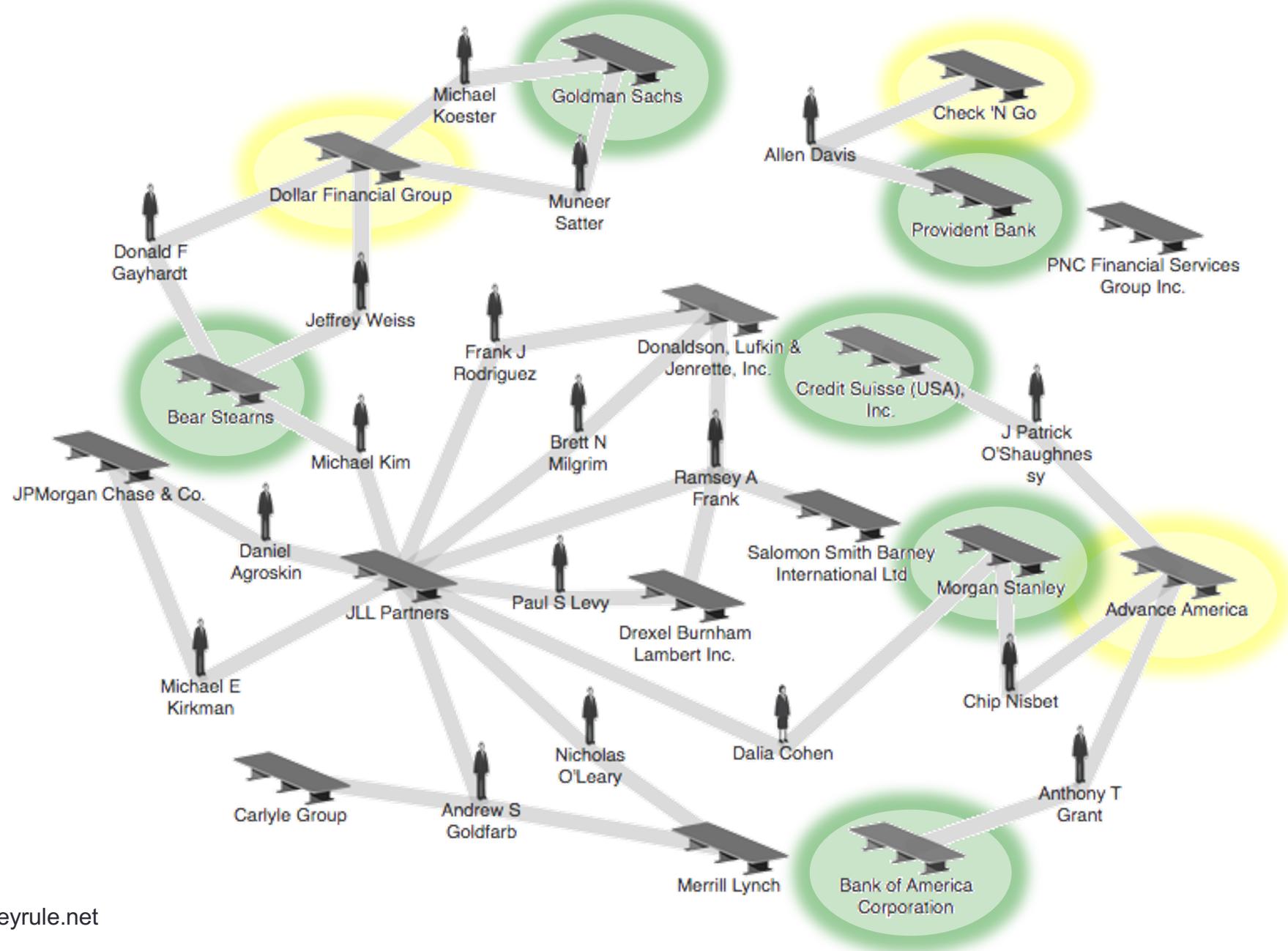


Do they help you spot trends?

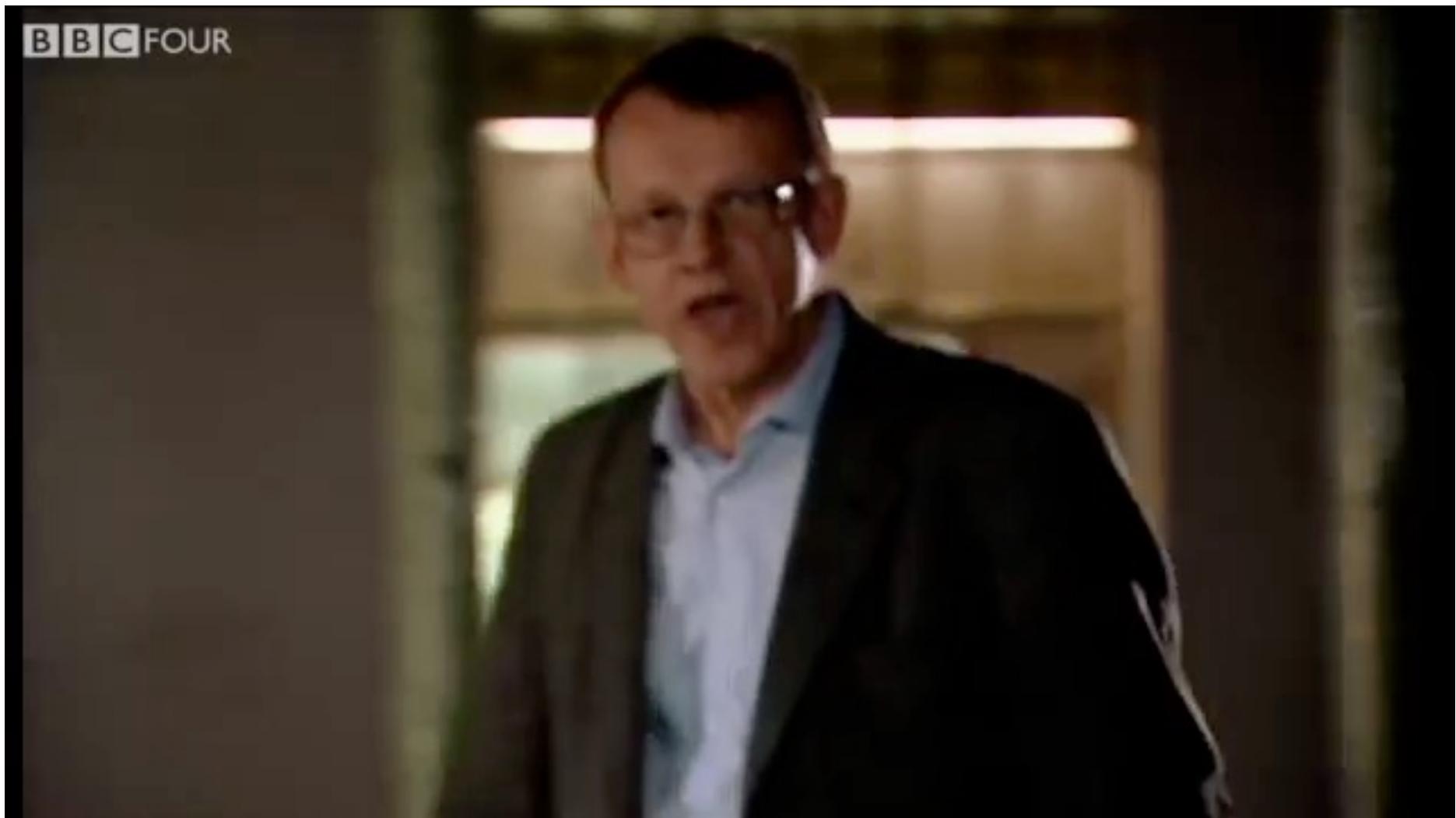


More info here: http://en.wikipedia.org/wiki/1854_Broad_Street_cholera_outbreak

Do they help you explore?



Do they tell a story?



Hans Rosling's 200 Countries, 200 Years, 4 Minutes – The Joy of Stats – BBC Four
<https://www.youtube.com/watch?v=jbkSRLYSojo>

Visualization (def.)

Visual
representations
of data that
reinforce human
cognition



Data (def.)

a set of *variables* that capture various aspects of the world:



*Tuition rates, enrollment numbers,
public vs. private, etc.*

Data (def.)

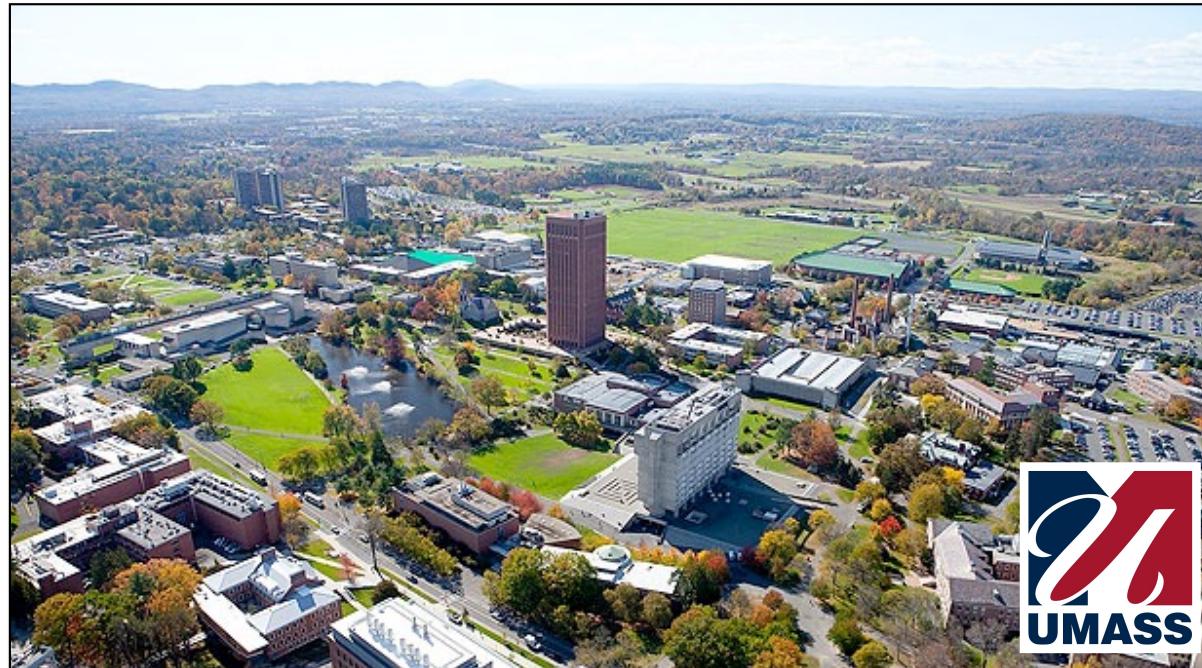
and a corresponding set of *observations* (a.k.a. *records*) over these variables. For example:



*tuition = \$46,288, enrollment = 2,563,
private, etc.*

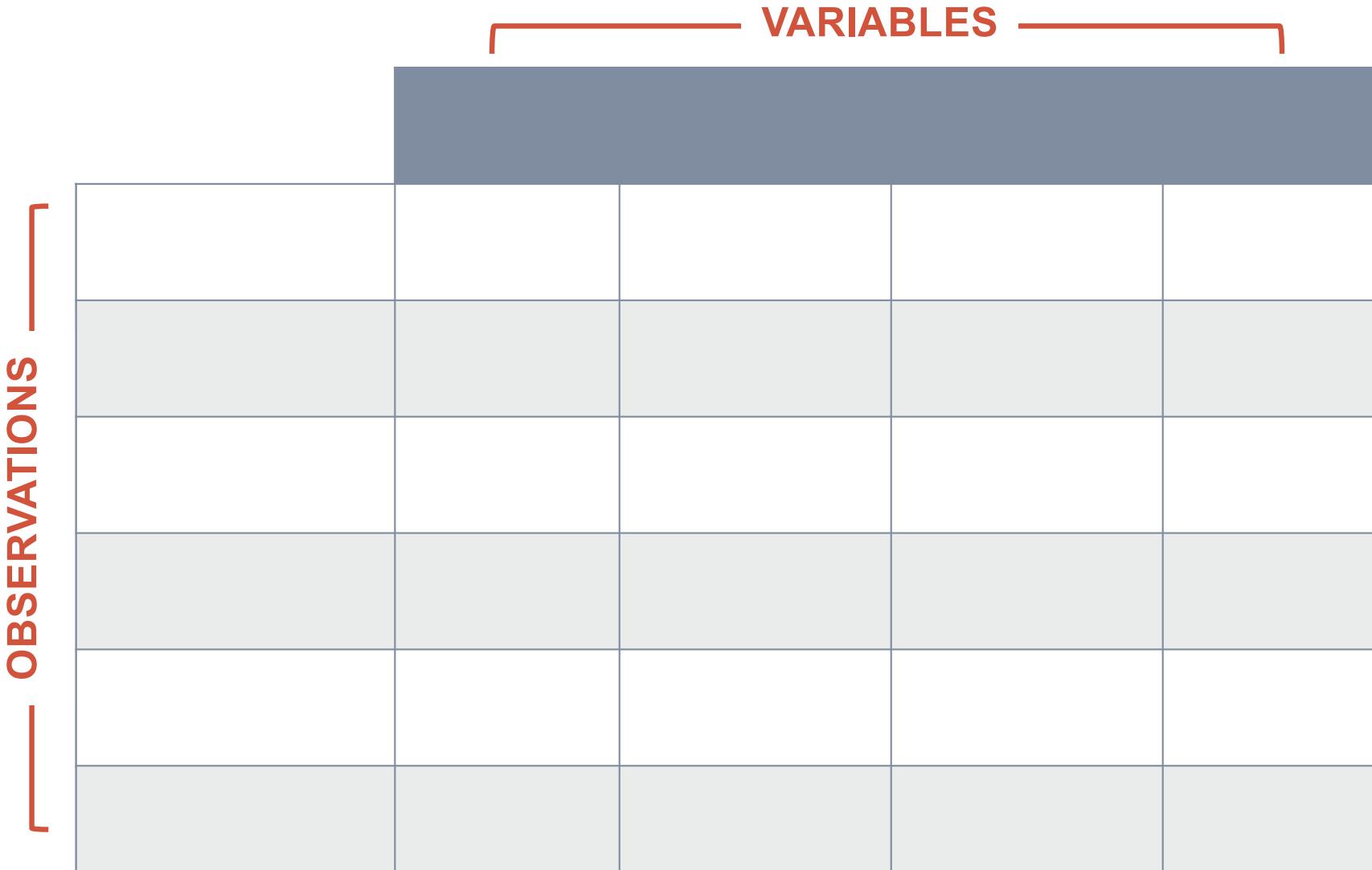
Data (def.)

and a corresponding set of *observations* (a.k.a. *records*) over these variables. For example:



*tuition = \$16,115, enrollment = 28,635,
public, etc.*

One way to think about this:



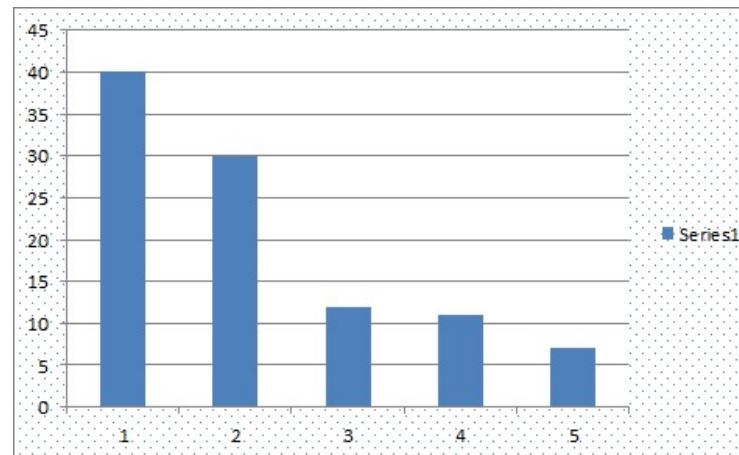
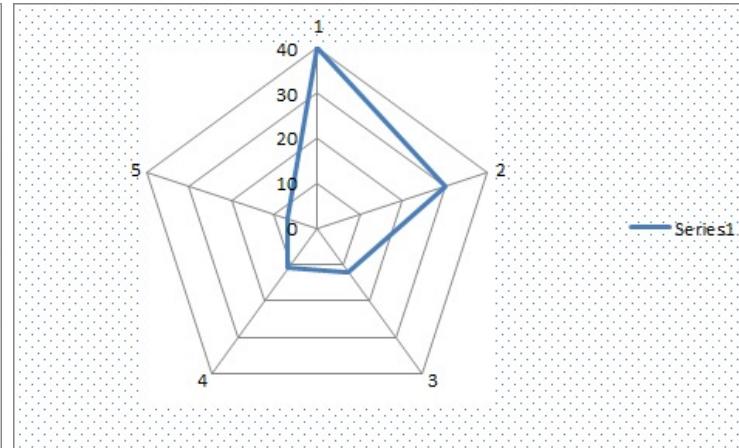
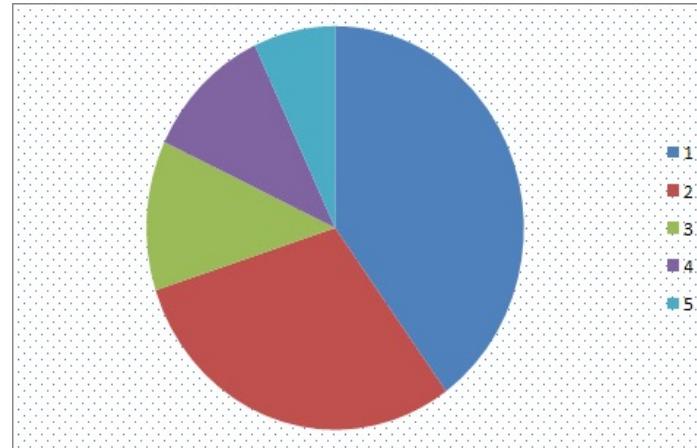
Why is this important?

- Data have dimensions
- Visualizations have dimensions, too
- To build visualizations, we need to **map** data dimensions to visual dimensions

Key question for this workshop

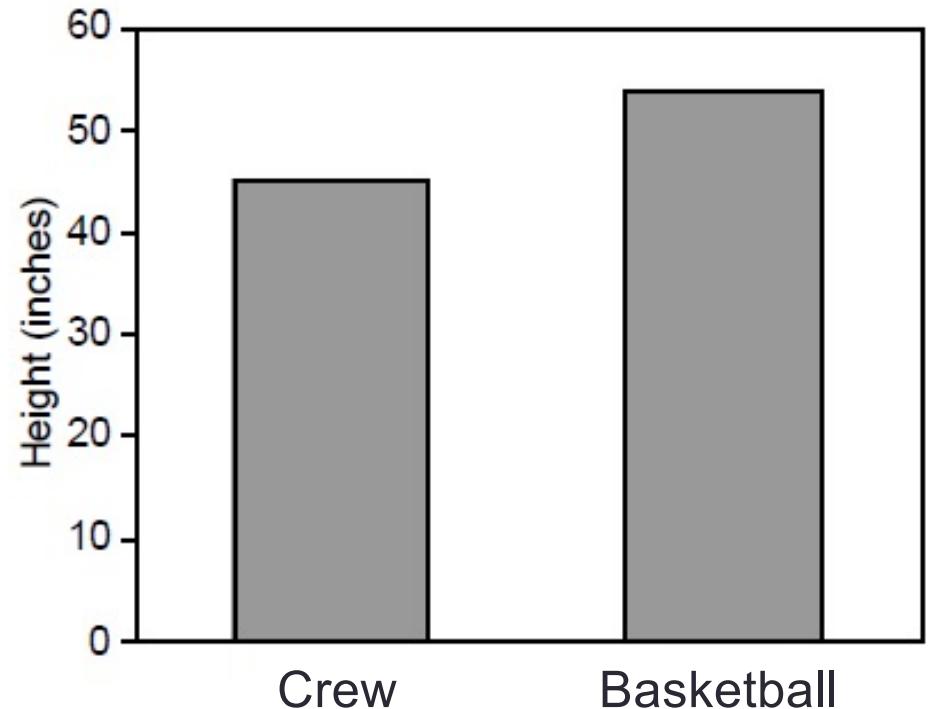
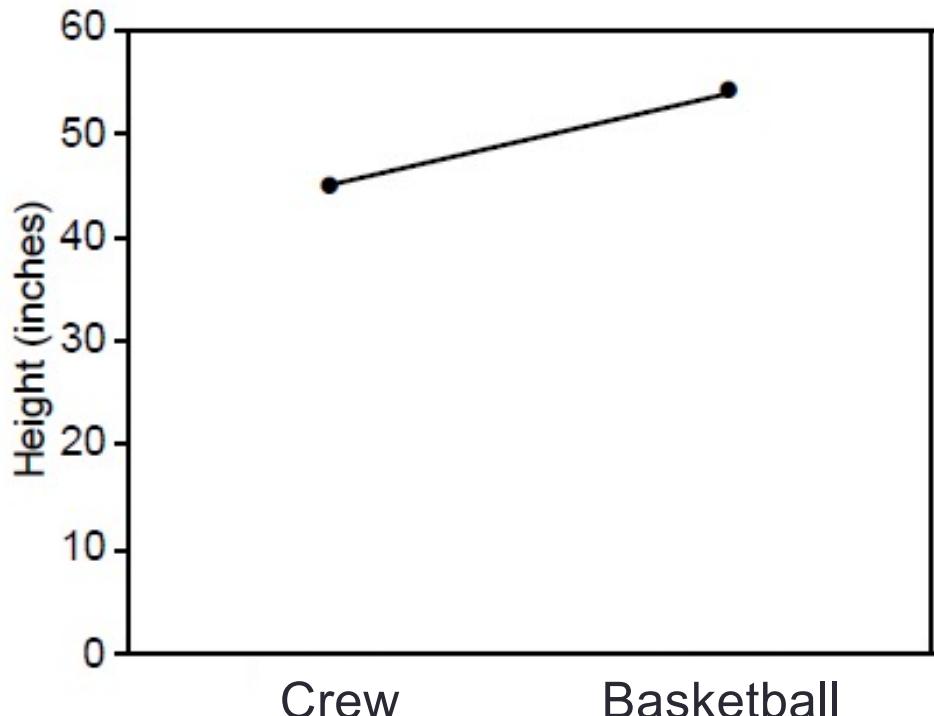
Which **data dimension** should be mapped

to which **visual dimension?**

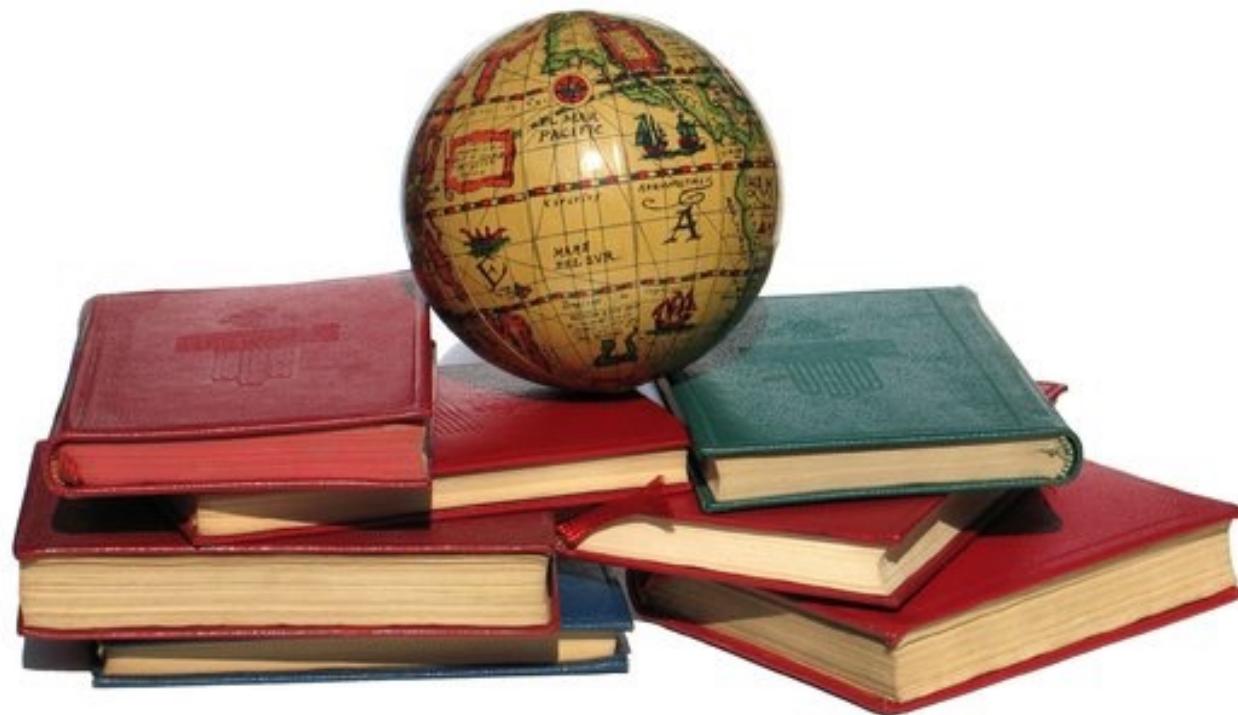


Answer: it depends

Average Height for Youth Sports Participants



A quick history lesson...

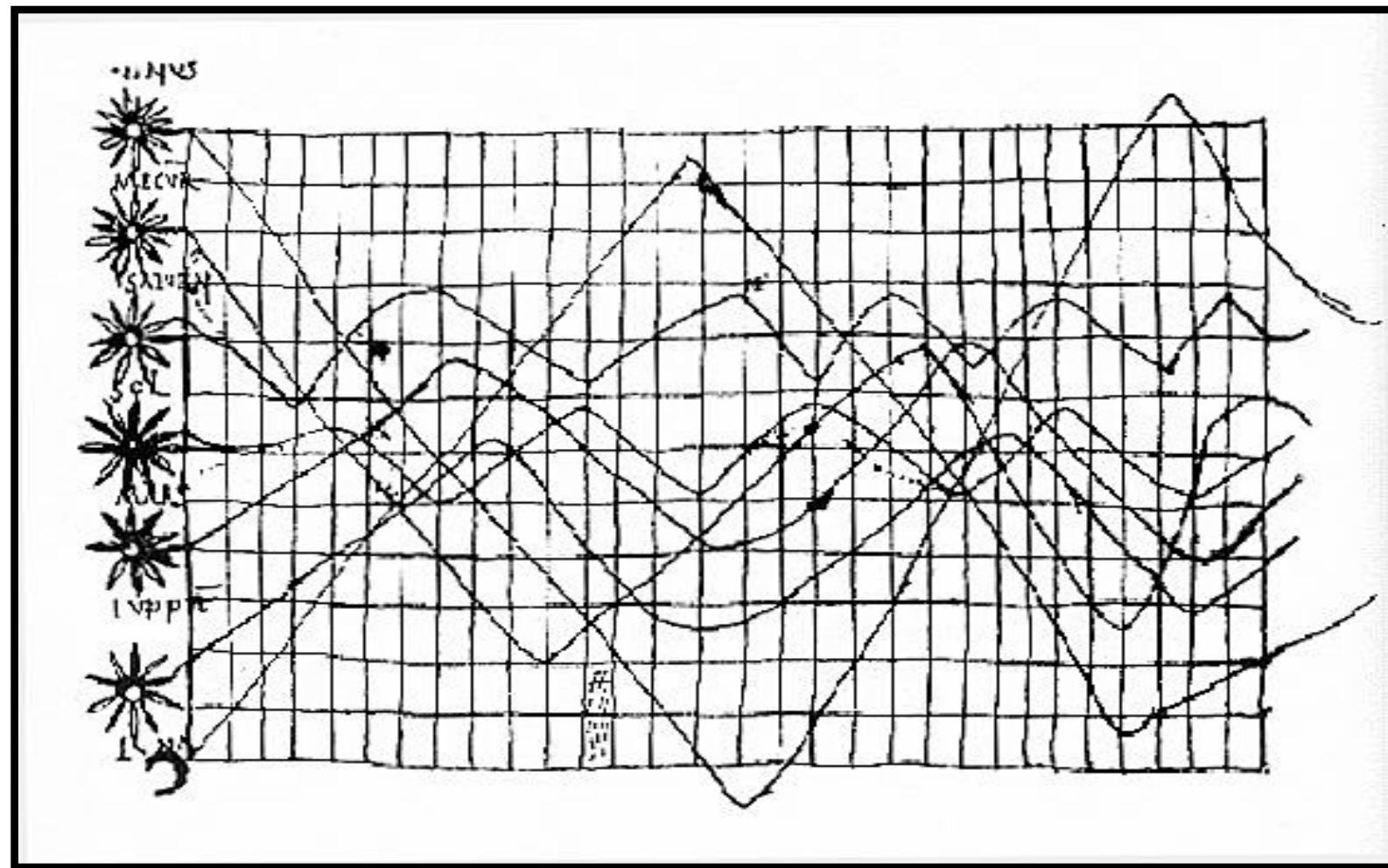


(Incomplete) History of Visualization: 15,000BC



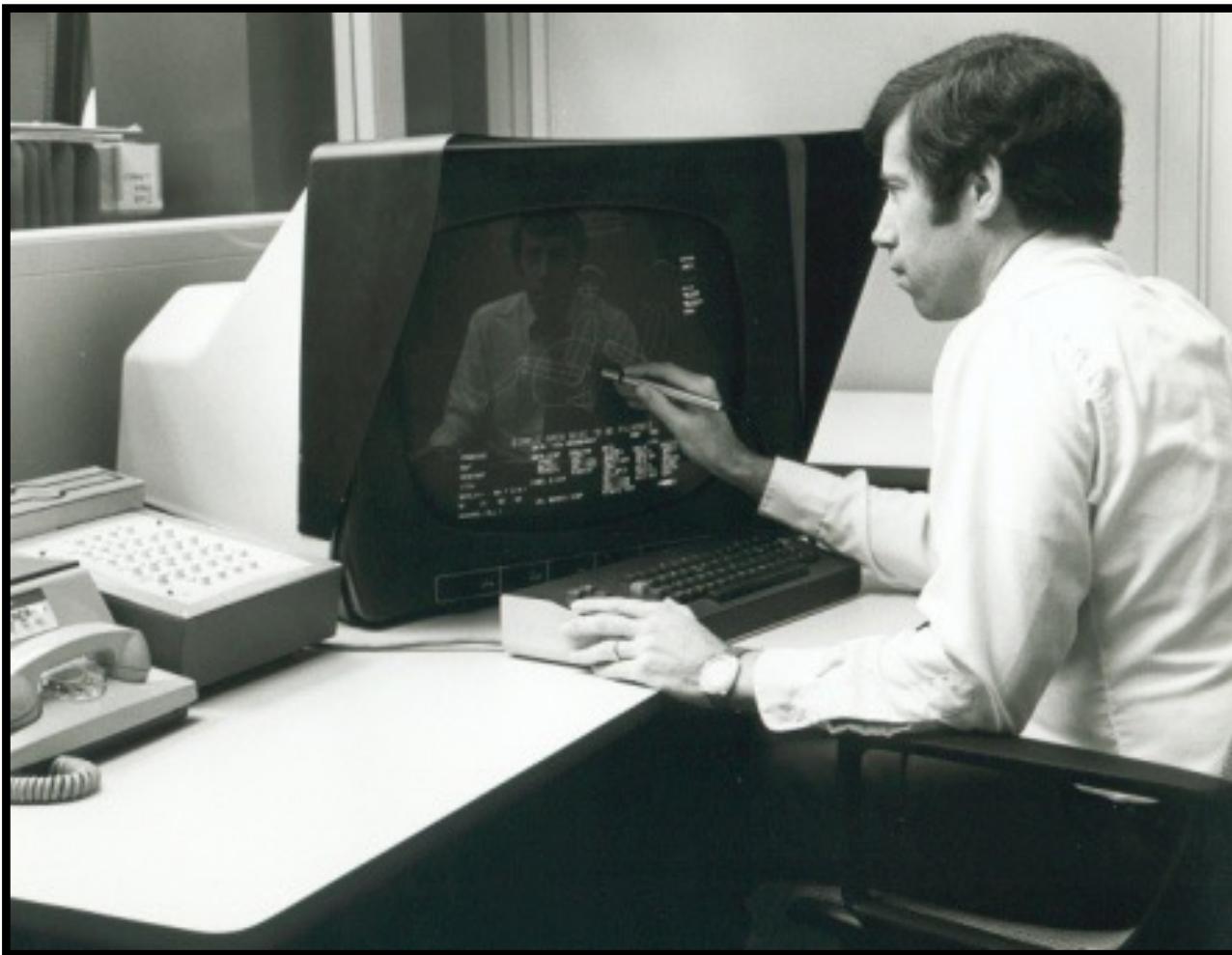
15,000 BC. Laxcaux, France

(Incomplete) History of Visualization: 900s



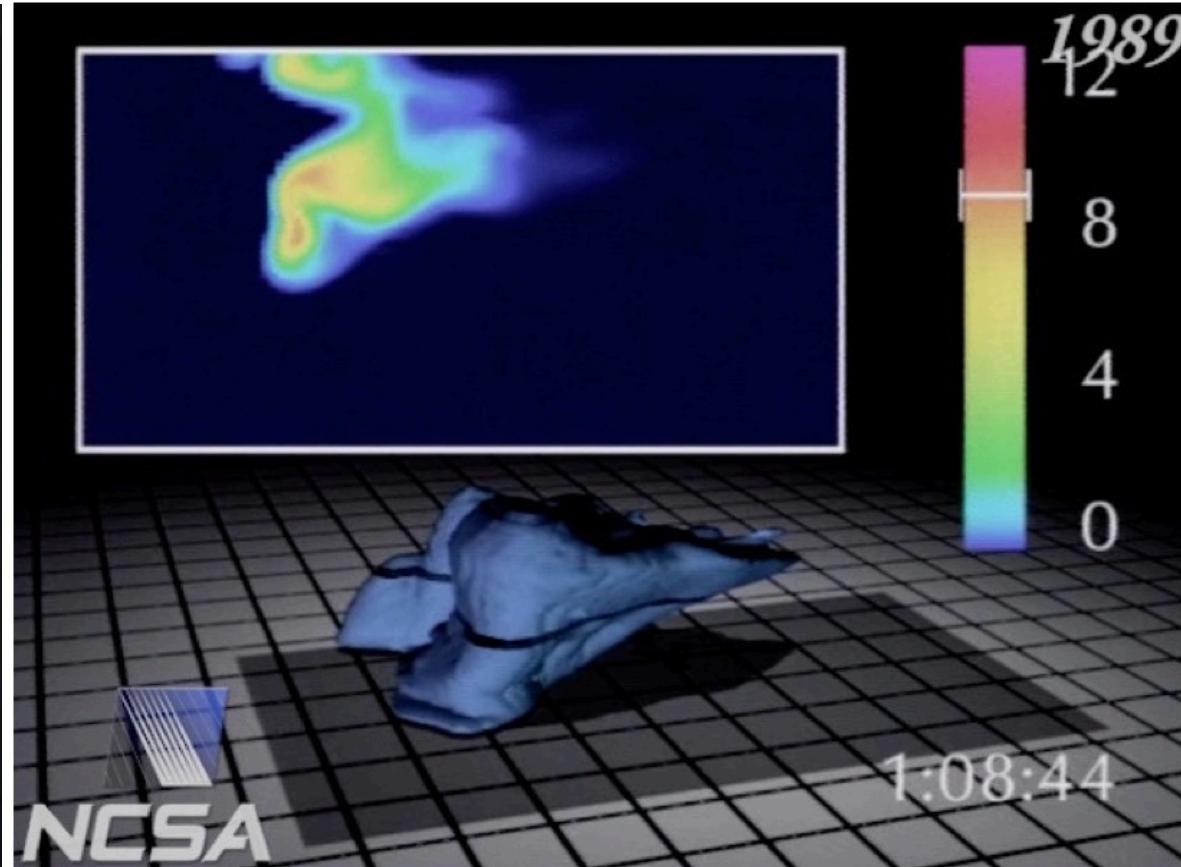
"De cursu per zodiacum", illustrator unknown

(Incomplete) History of Visualization: 1970s



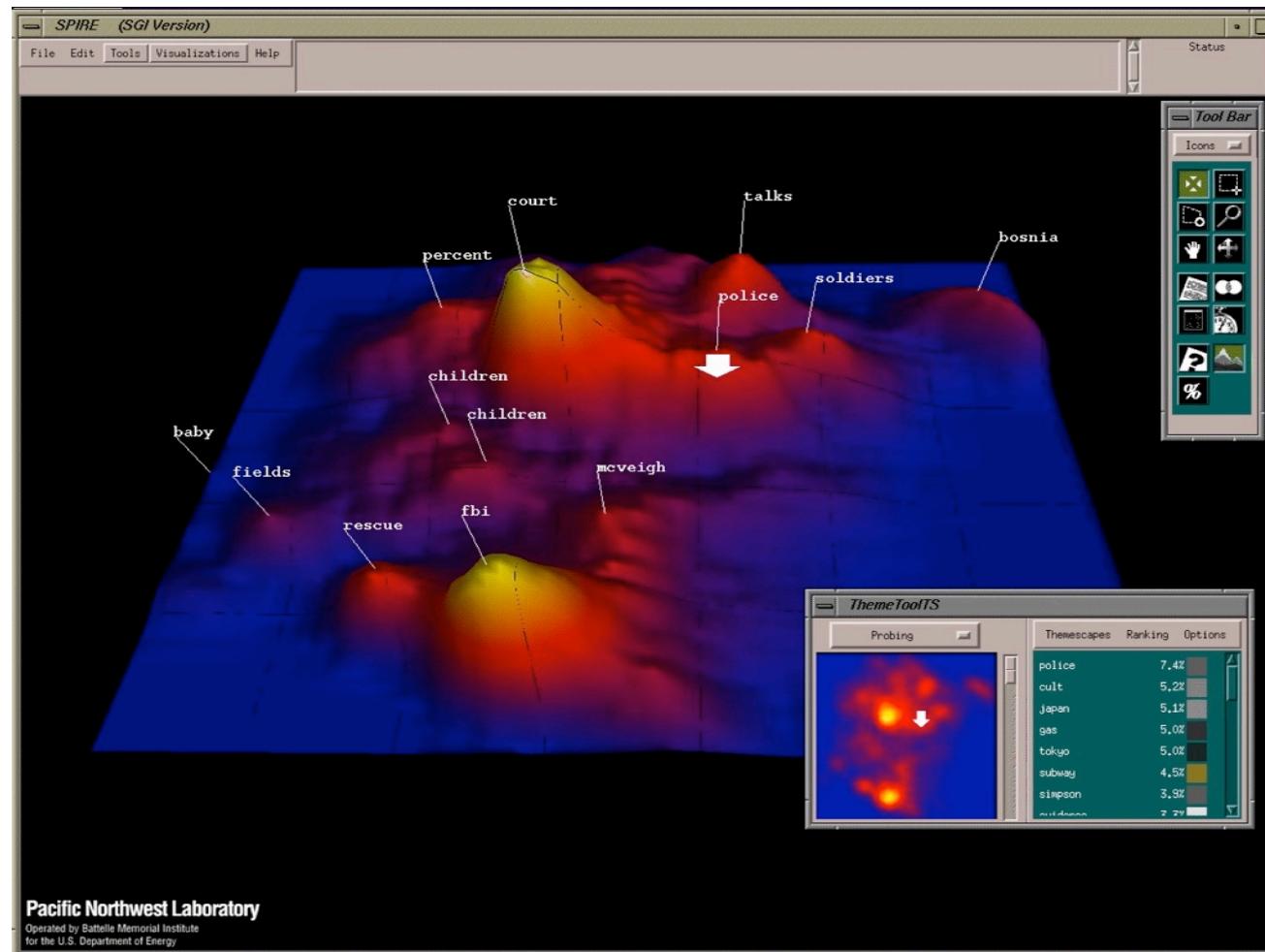
- CAD/CAM, building cars, planes, chips
- Starting to think about: 3D, animation, edu, medicine

(Incomplete) History of Visualization: 1980s



- Scientific visualization, physical phenomena
- Starting to think about: photorealism, entertainment

(Incomplete) History of Visualization: 1990s



- Information visualization, storytelling
- Starting to think about: online spaces, interaction

(Incomplete) History of Visualization: 2000s



- Coordination across multiple views, interaction
- Starting to think about: sensemaking, provenance

(Incomplete) History of Visualization: 2010s

Visual First Amendment

Interactive data visualizations that present a new and deeper understanding of the Supreme Court

Issues ▾

Cases & Courts ▾

Justices ▾

- Justice Voting Ideology
- Justice Agreement Network

About
Data Sources
Research
News
Contact

Pratt
School of Information and Library Science

Justices by Ideology

The Supreme Court Database codes most decisions as either conservative or liberal ([see definitions](#)). Using that information and justices' votes, we computed an ideological direction for each justice's votes on First Amendment cases:

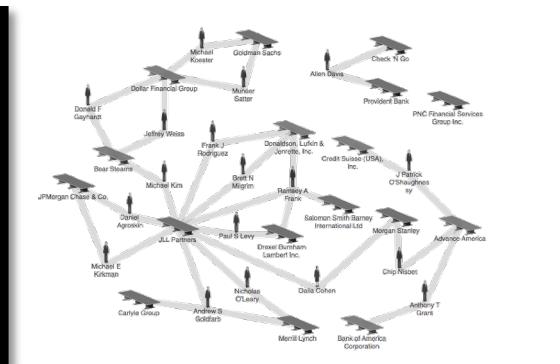
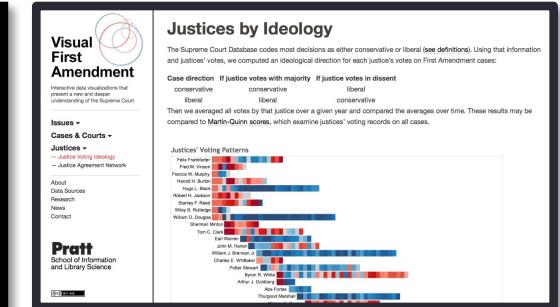
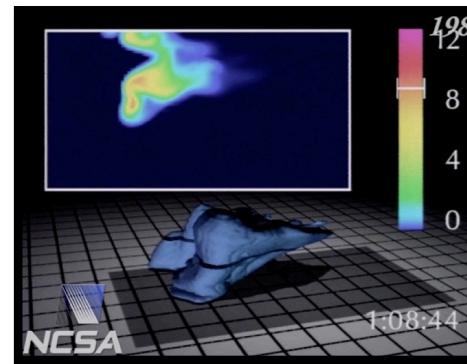
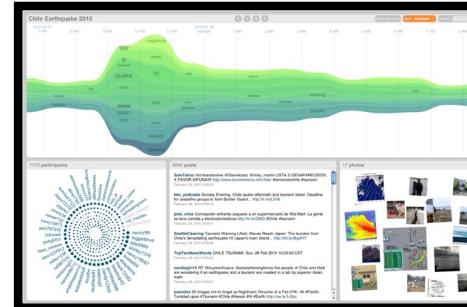
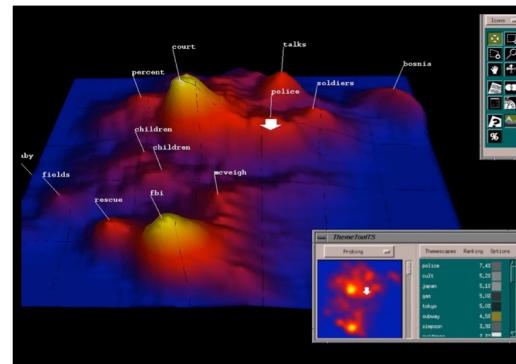
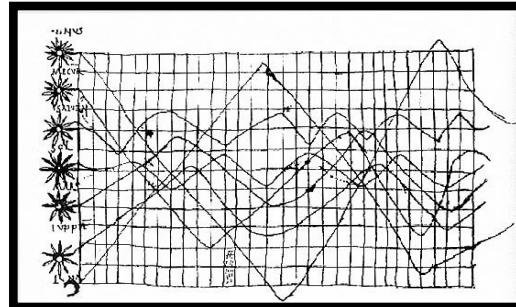
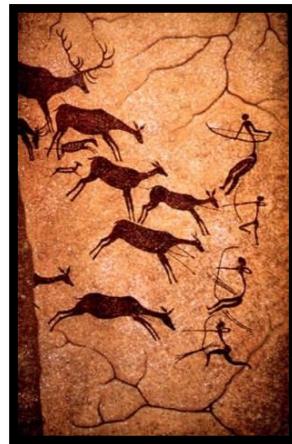
Case direction	If justice votes with majority	If justice votes in dissent
conservative	conservative	liberal
liberal	liberal	conservative

Then we averaged all votes by that justice over a given year and compared the averages over time. These results may be compared to [Martin-Quinn scores](#), which examine justices' voting records on all cases.

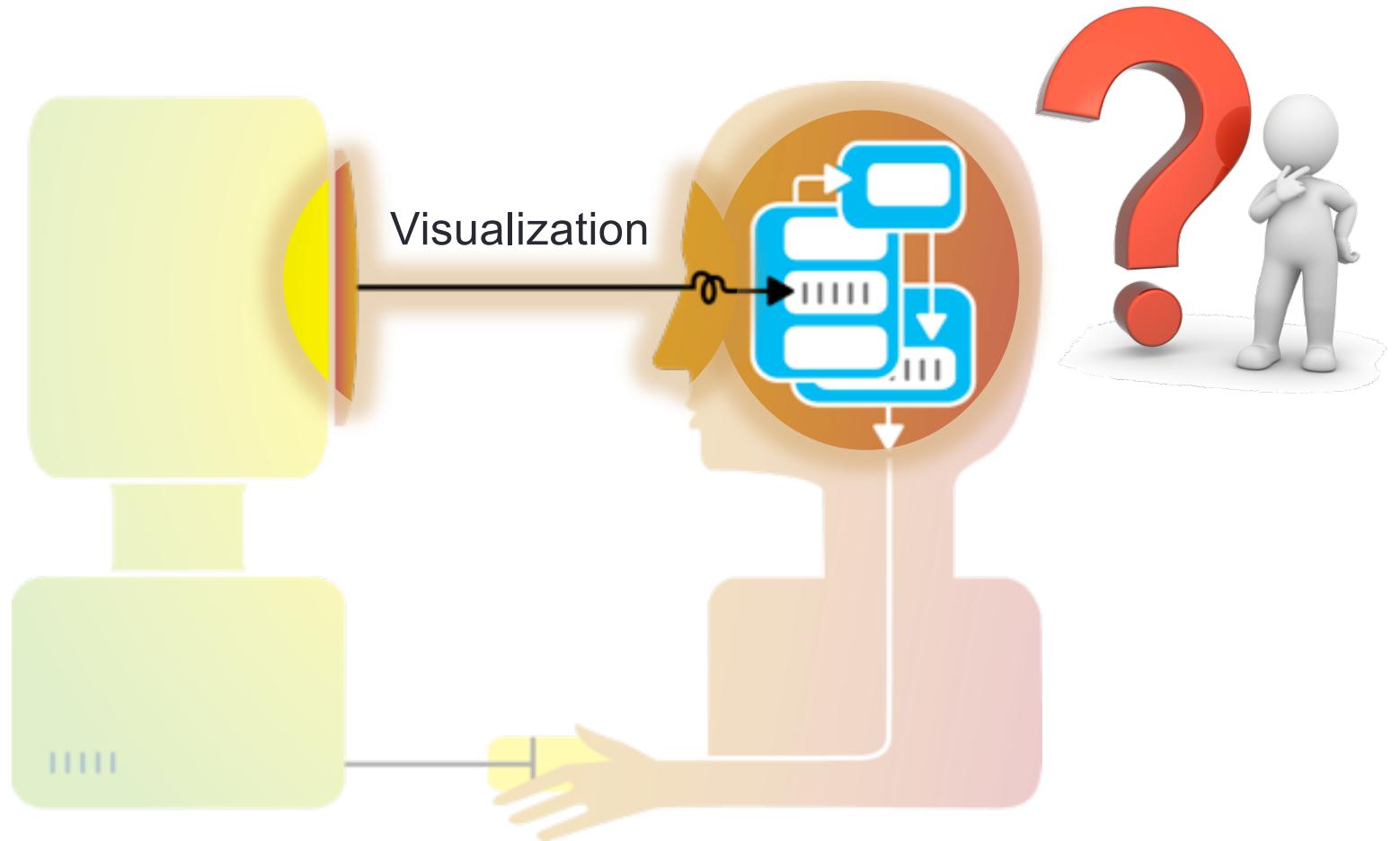
Justices' Voting Patterns

- Human-machine collaboration, machine learning
- Starting to think about: evaluation, new media, DH

Discussion: what are they all trying to do?



Visualization helps shape *mental models*



Information overload

- We are exposed to huge amounts of information all the time
- So much, in fact, that we can't process it all fast enough



Mental models

To cope, we construct **mental models**: abstracted, simplified versions of the world that are more manageable



Mental Models: a Sketch



1. We tend to see what we expect to see



2. Mental models form quickly, & update slowly



3. New information gets incorporated into the existing model



4. Initial exposure interferes with accurate perception



Blur size

- 128px
- 64px
- 32px
- 16px
- 8px
- None

The good, the bad, and the ugly...

The good:

- Well-tuned mental models let us process information quickly
- Frees up more processing power to synthesize information

The bad:

- People (esp. experts) tend not to notice information that contradicts their mental model
- A “fresh pair of eyes” can be beneficial

The ugly:

- Mental models are unavoidable: everyone has them, and they’re all different
- **Key:** be aware of how mental models form, how they shape perception, and how to support (or challenge) them

So what do we have to work with?

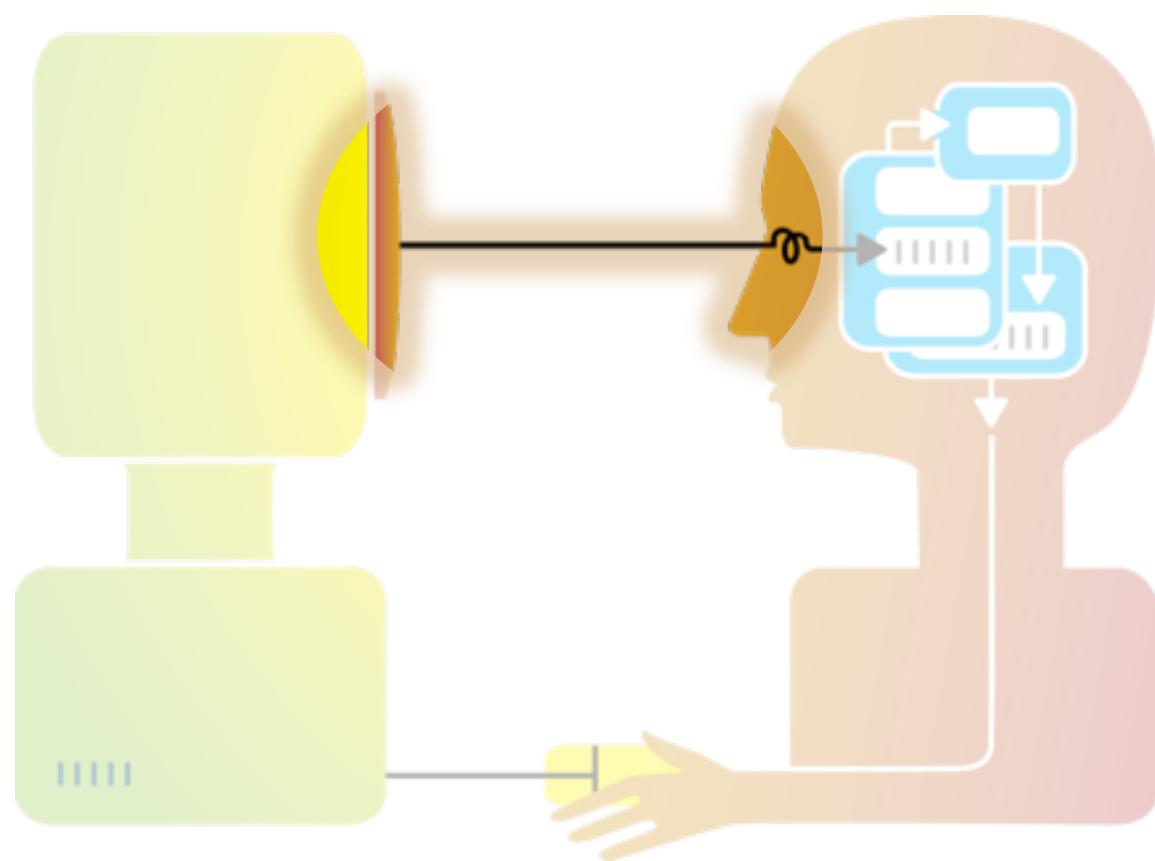
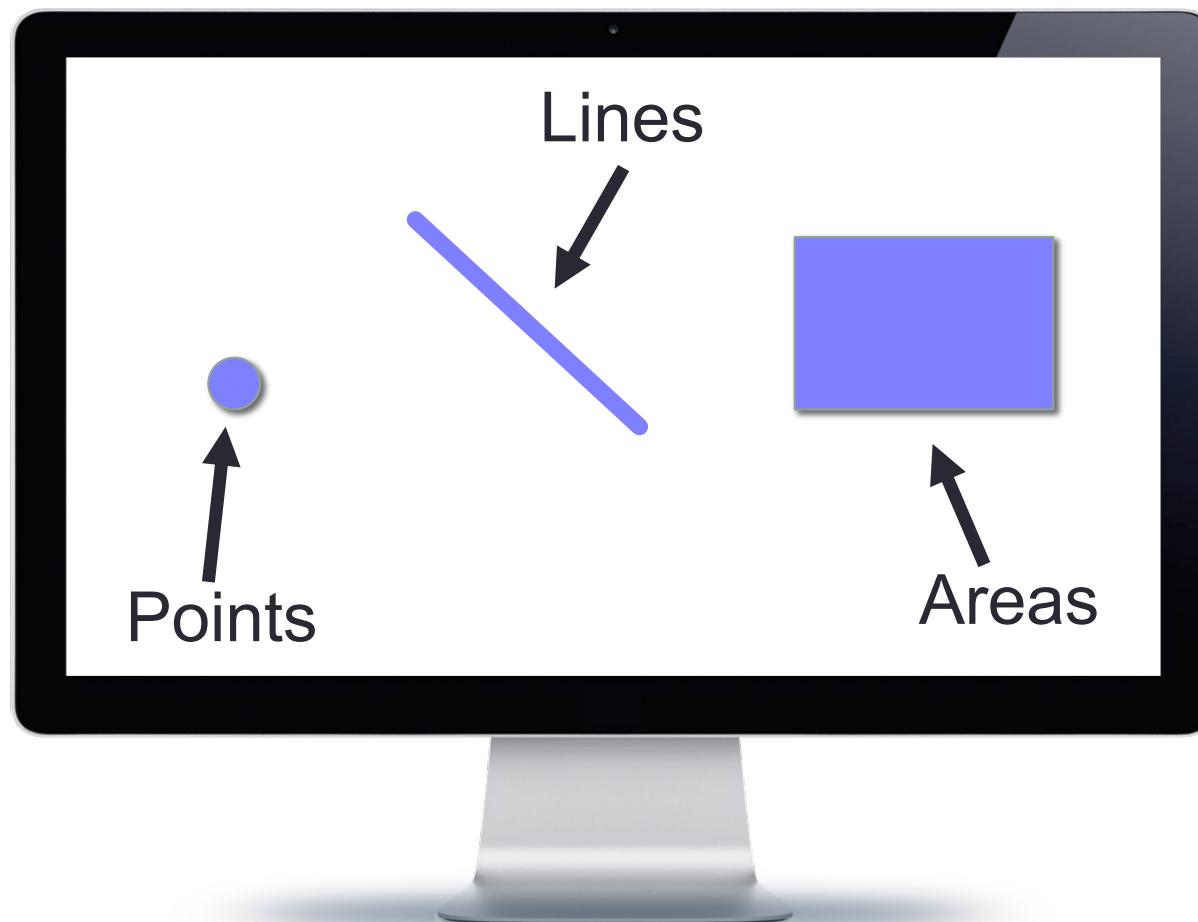


Image credit: Ali Ansari

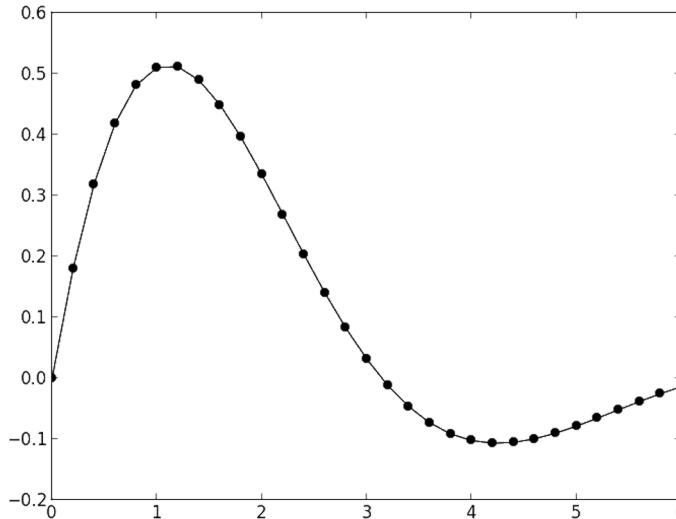
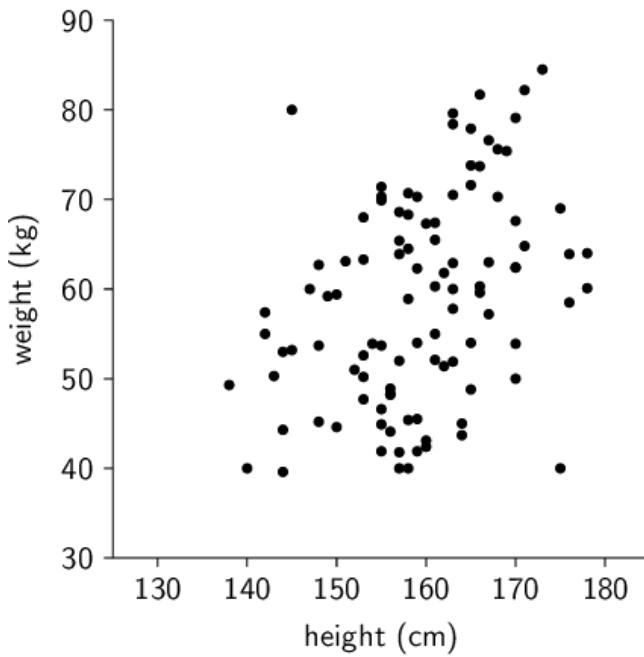
Graphical primitives

The images we draw are composed of marks: like ink



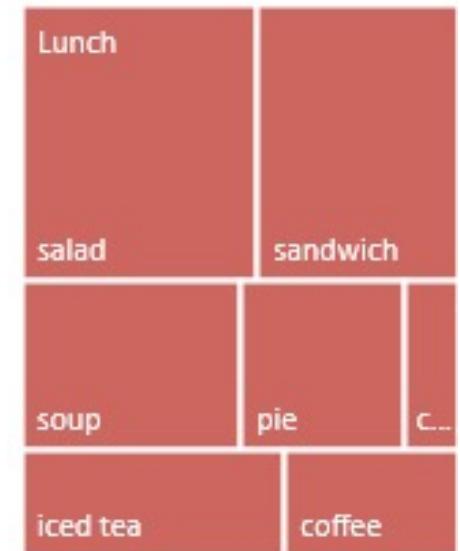
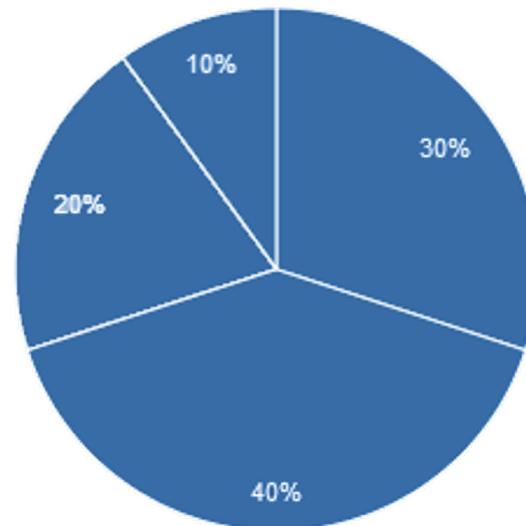
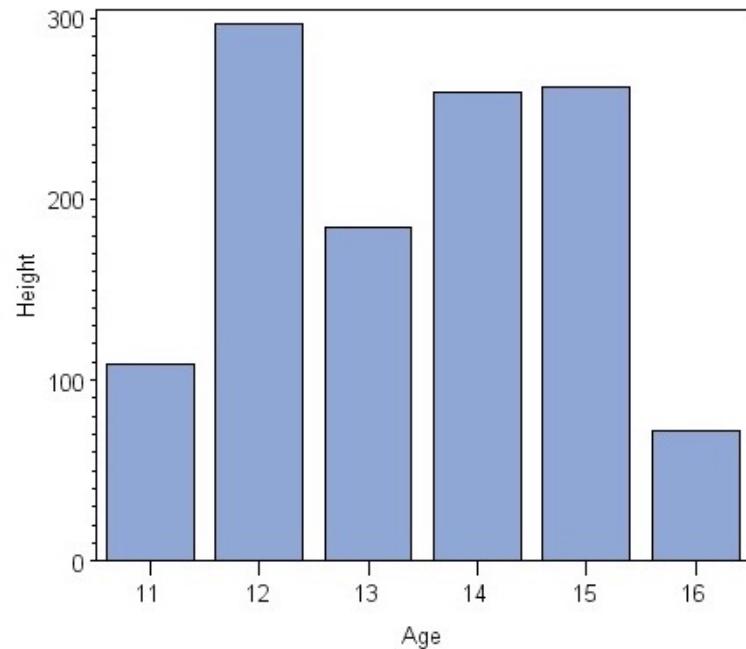
Visual dimension: position

- Encode information using **where** the mark is drawn
- Some examples:



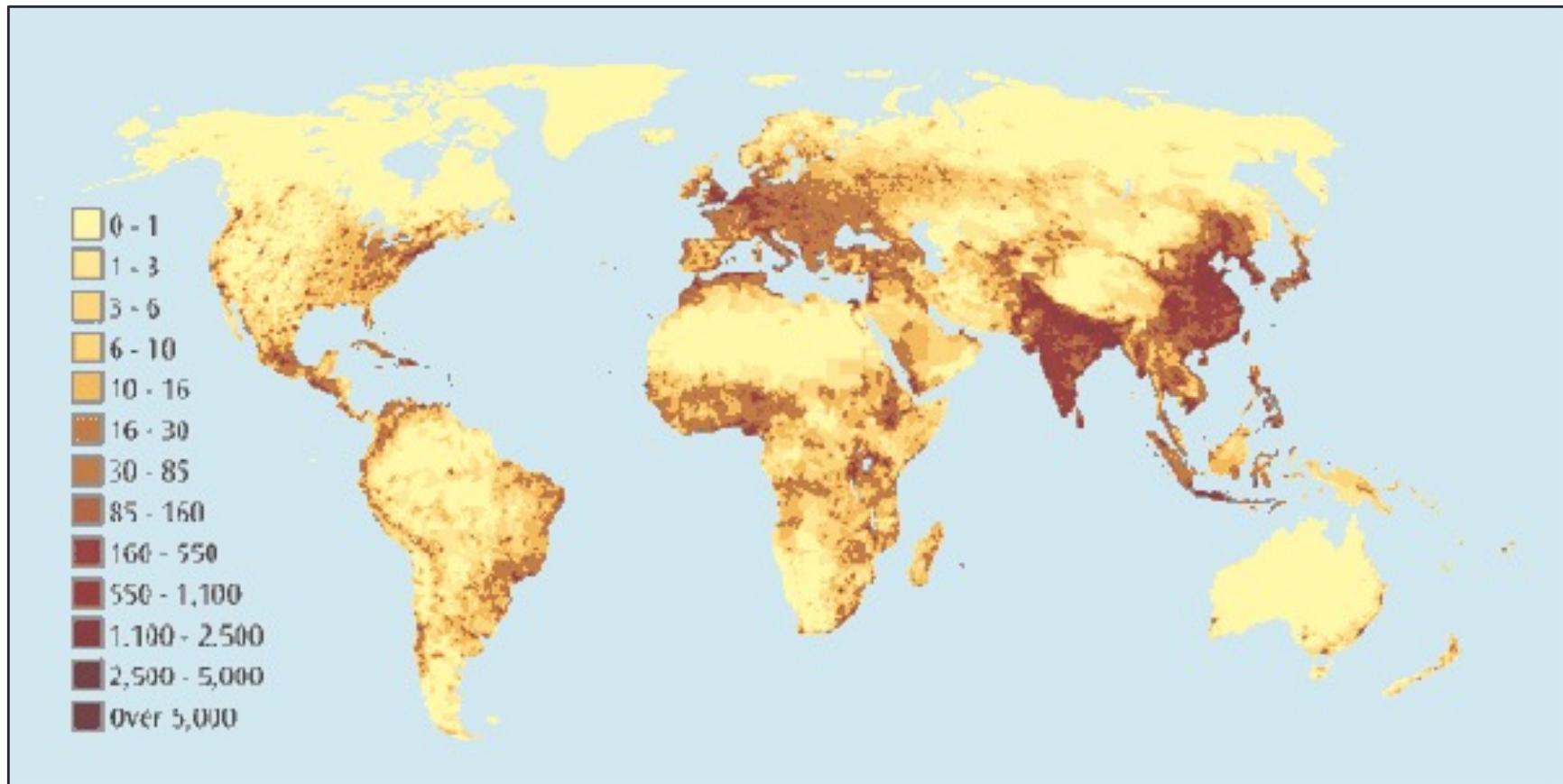
Visual dimension: size

- Encode information using **how big** the mark is drawn
- Examples:



Visual dimension: value

- Encode information using **how dark** the mark is drawn
- Example:



Visual dimension: color

- Encode information using the **hue** of the mark
- Examples:

Benefits

About 1 out of 10 women improved their symptoms using this medicine.



Side Effects

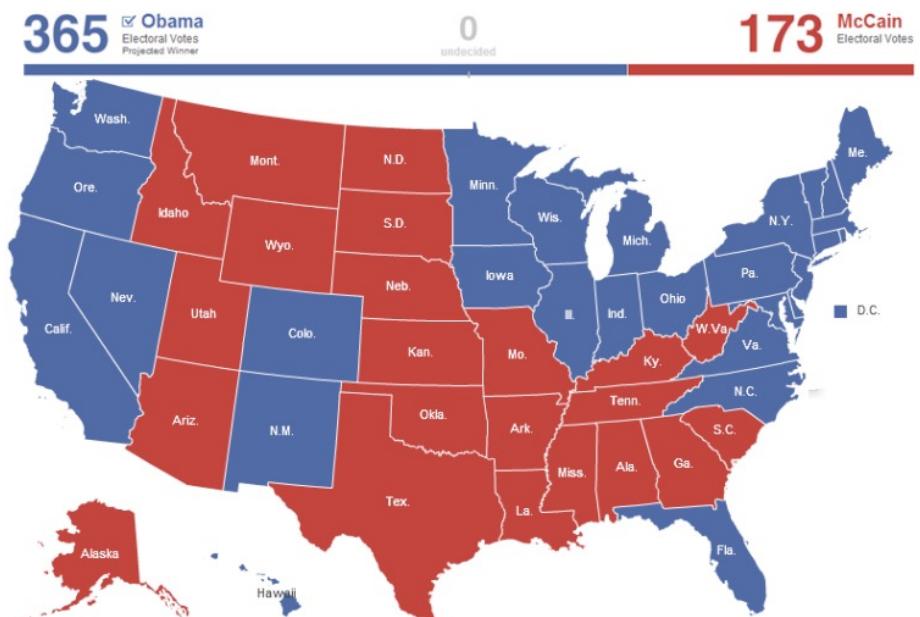
About 2 out of 10 women had dry mouth using this medicine.



About 1 out of 10 women had constipation using this medicine.

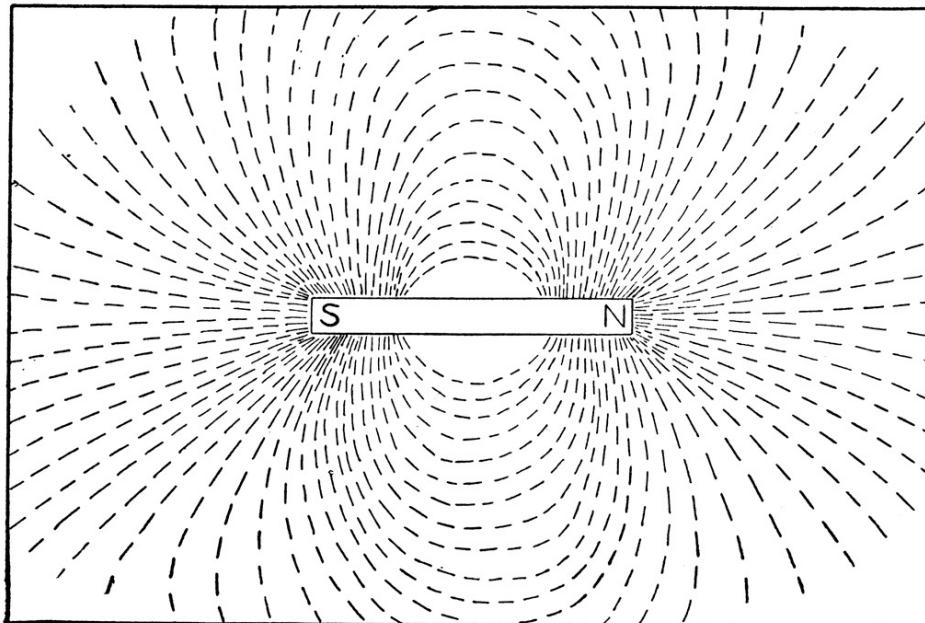


Less than 1 out of 10 women had an upset stomach using this medicine.



Visual dimension: orientation

- Encode information using how the mark is **rotated**
- Examples:



Visual dimension: shape

- Encode information using how the mark is **shaped**
- Examples:



Discussion: visual dimensions & data type

	Categorical	Numerical
POSITION		
SIZE		
VALUE		
COLOR		
ORIENTATION		
SHAPE		



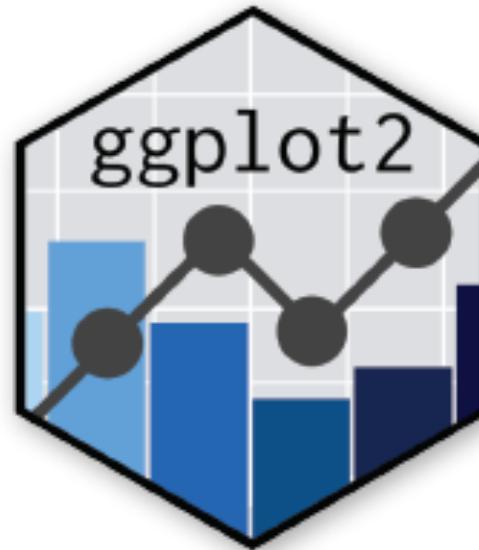
Jacques Bertin, *Semioleogie Graphique*
(Semiology of Graphics), 1967.

Think-pair-share: deconstructing graphics

1. Find a data visualization you think is interesting
 - Some ideas: NYTimes, VisualisingData.com, Visual.ly
 - Remember to cite your source!

2. Identify the following:
 - What is the **data** that's being visualized? Where did it come from?
 - Which **data dimensions** are mapped to which **visual dimensions**?
 - How does this **shape your understanding** of the data?
 - If you **liked** the visualization: what is it doing **well**?
 - If you **disliked** the visualization: what would you **change**?

After lunch



- Mini-lecture and lab: building data graphics with ggplot2
- **TODO** (if you haven't already):
 > `install.packages('ggplot2')`