

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](https://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

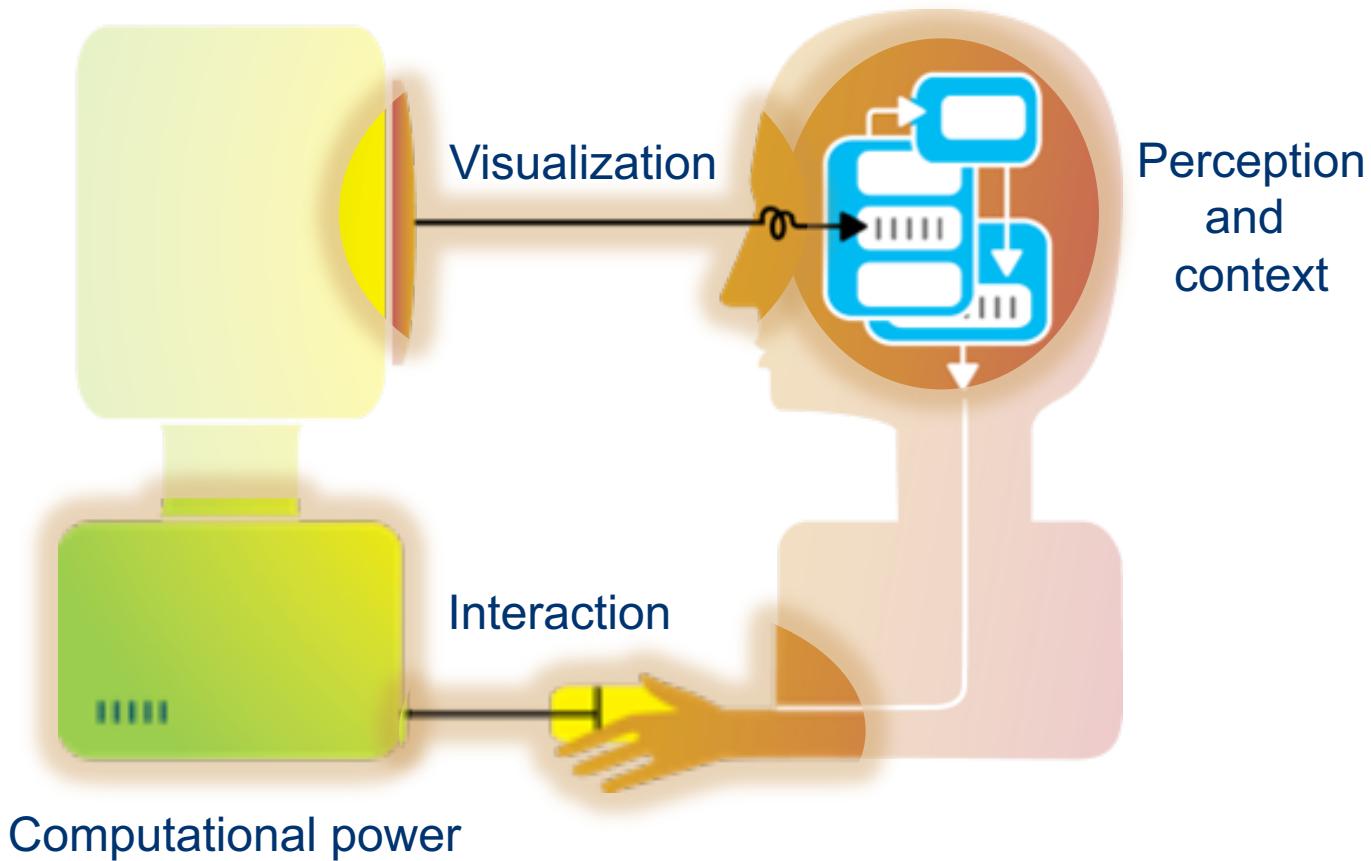
3. Opportunities to  
get to know  
your new team

# What I do: analytical tools for messy data



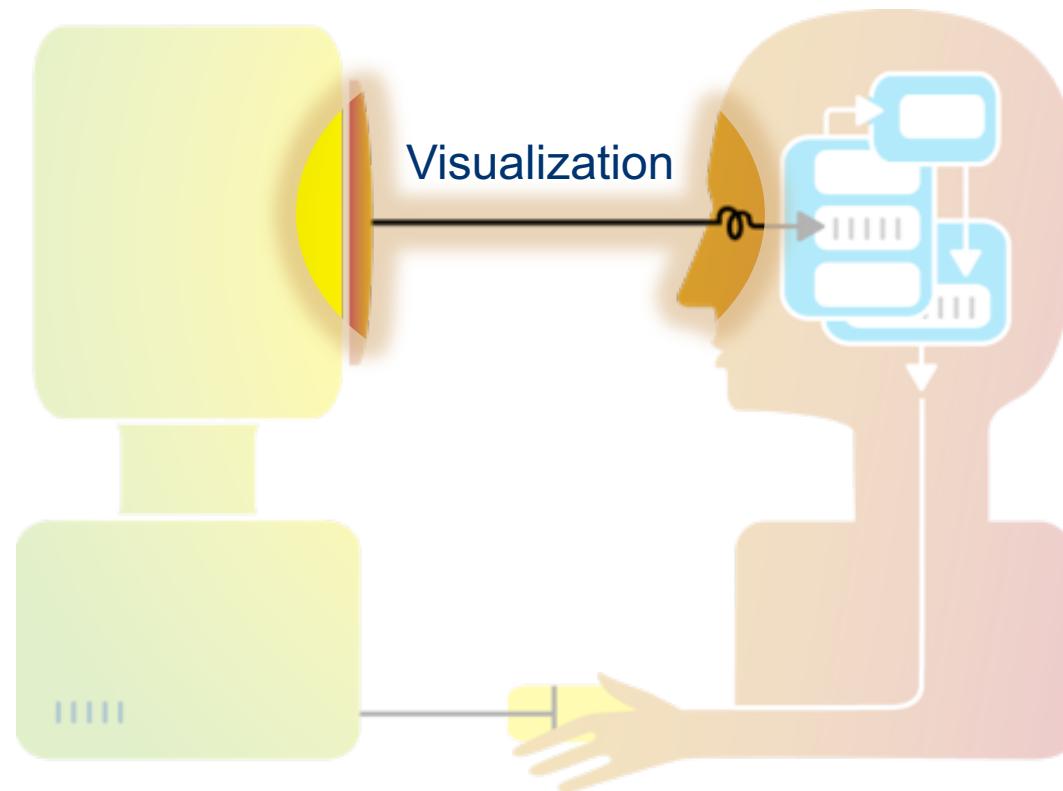
# Big idea behind my research

Humans and machines have **complimentary strengths**



# 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

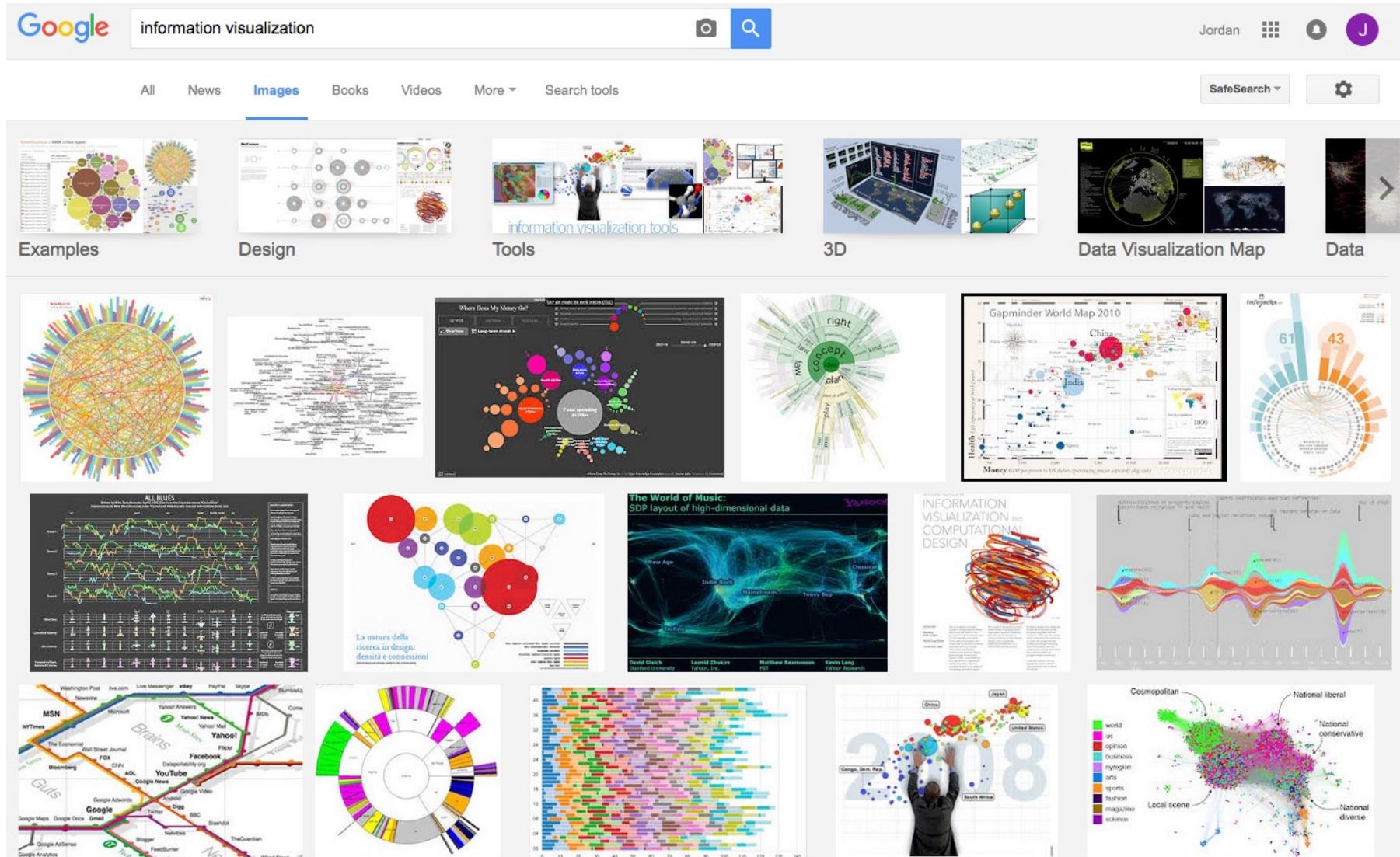
The search results page displays a grid of image thumbnails under several categories:

- Reading Strategy:** Includes a poster titled "Visualize" and a colorful illustration.
- Data:** Shows a complex network graph and a circular diagram.
- Quotes:** Features a quote from the Bible: "VISUALIZATION IS DAYDREAMING WITH A PURPOSE. THE MOST AMAZING LIFE IS THE ONE YOU CAN CLOSE YOUR EYES AND HOLD THE VISION FOR AS LONG AS YOU CAN. THE VISION IN GOD'S HANDS...AND CONSIDER IT DONE."
- Sports:** Displays two images of athletes.
- Creative:** Shows a person standing on a beach with a colorful cloud effect, a book cover, and a person in a suit.
- Techniques:** Includes a person's head filled with icons and a person's face with a hand near their forehead.

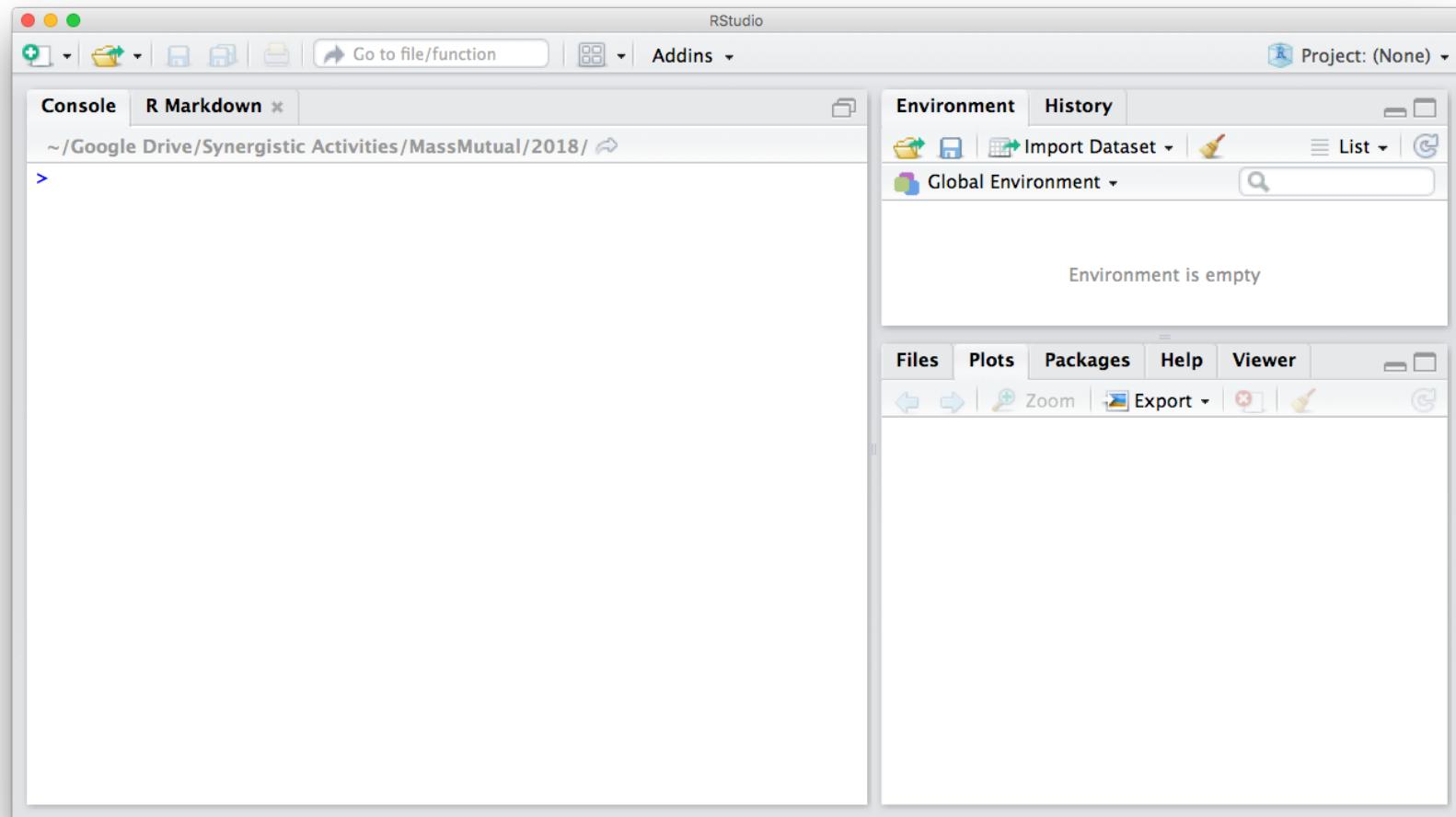
Below these categories are several more image thumbnails:

- A group of people standing on a grassy field with icons above them.
- A close-up of a human eye with a green iris.
- A vibrant, abstract visualization of a starry sky or neural network.
- A close-up of a human eye looking forward.
- A woman's face with her hand resting against her forehead.
- A close-up of a human eye with a rainbow-colored iris.
- A complex network graph with many nodes and connections.
- A woman's face with a wireframe overlay.
- A bar chart with "PRESENT" and "FUTURE" labels.
- A collage of words related to visualization: "future", "see", "perfect", "life", "dreams", "want", "exactly", "detail", "always", "create", "imagination", "power", "now", "person", "already", "envision", "things", "positive", "picture", "color", "vision", "goals", "imagine", "mind", "crystal", "visualization", "mental", "clarity", "use".
- A network graph with text at the bottom: "exciting!", "beautiful!", "technical!", "BUT", "[what does it mean?]".
- A close-up of a human eye with the text "THE POWER OF VISUALIZATIONS" overlaid.
- A silhouette of a human head with gears inside.
- A colorful visualization of a human head with the word "visualization" inside.
- A silhouette of a human head with gears inside.
- A couple looking at a futuristic interface showing a globe and data.
- A globe visualization on a screen.

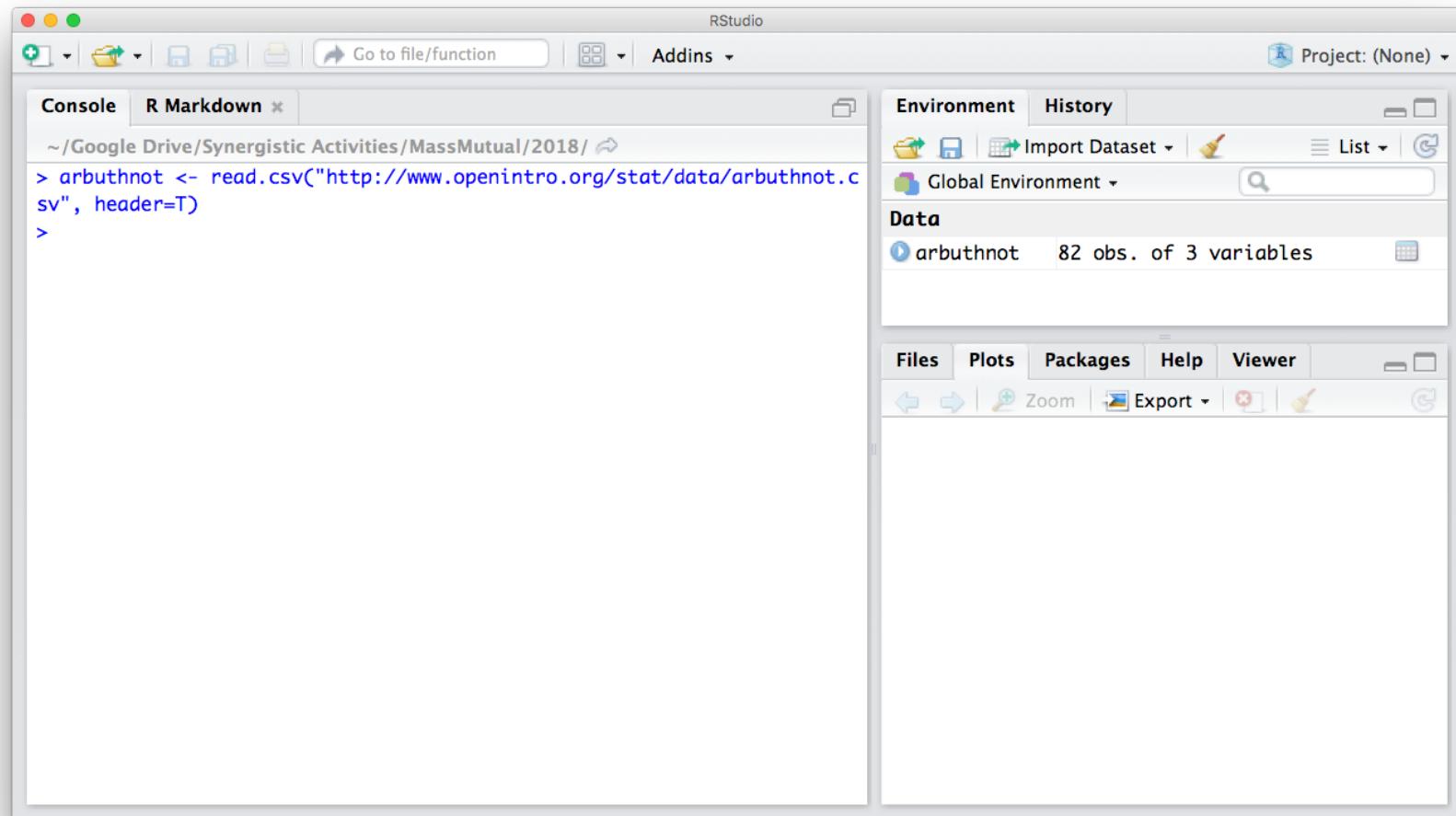
# What is visualization?



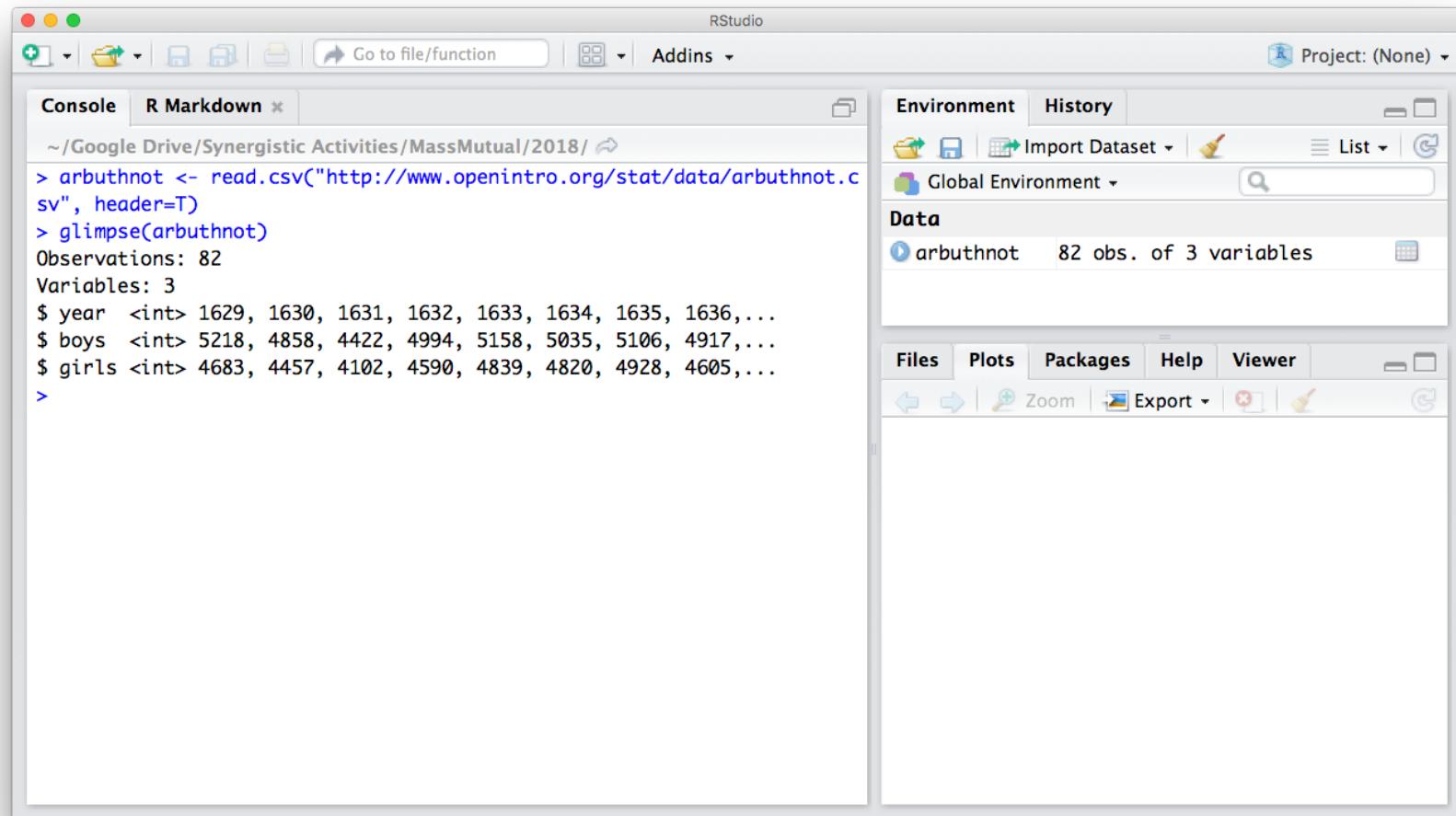
# Flashback...



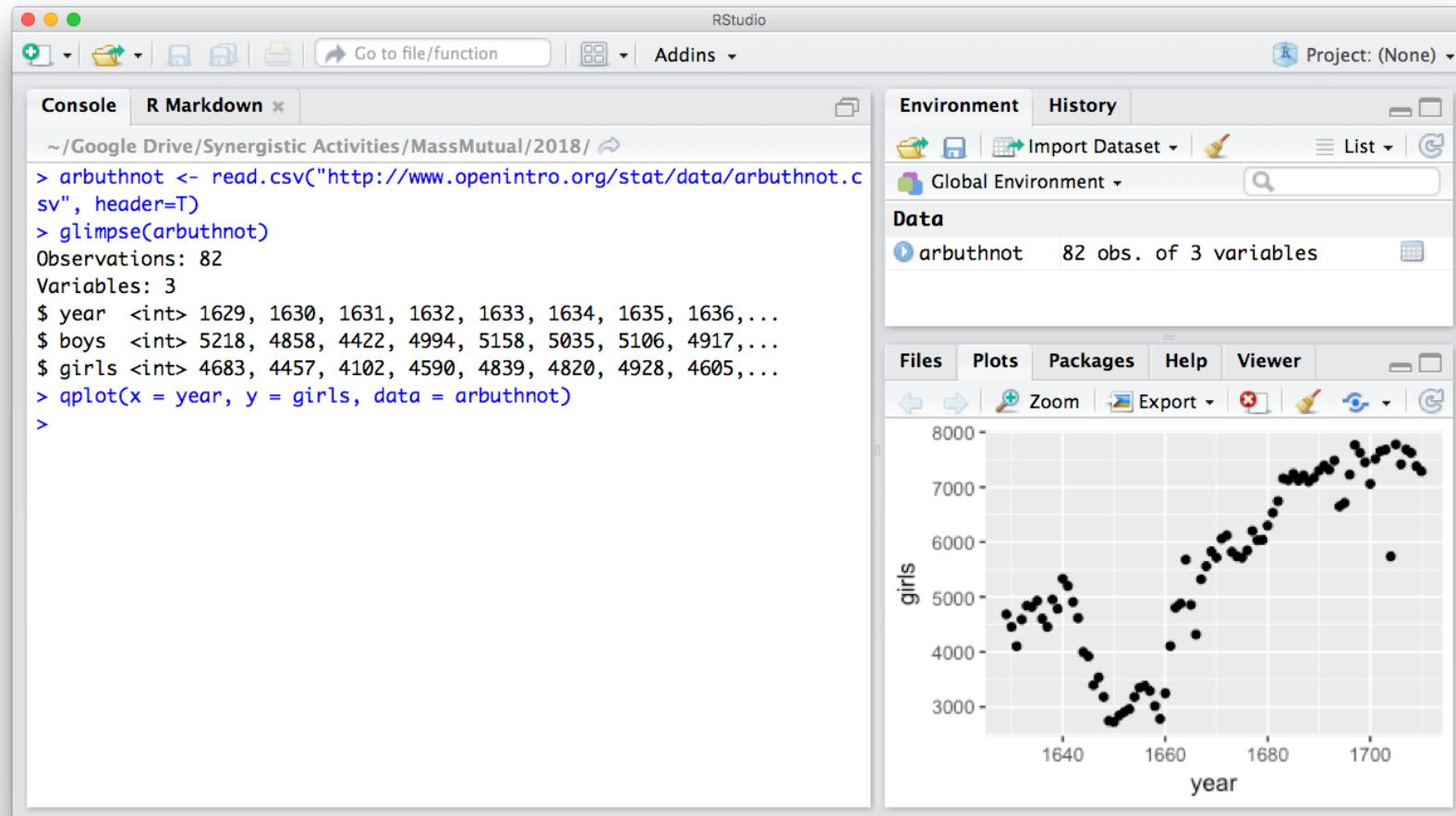
# Flashback...



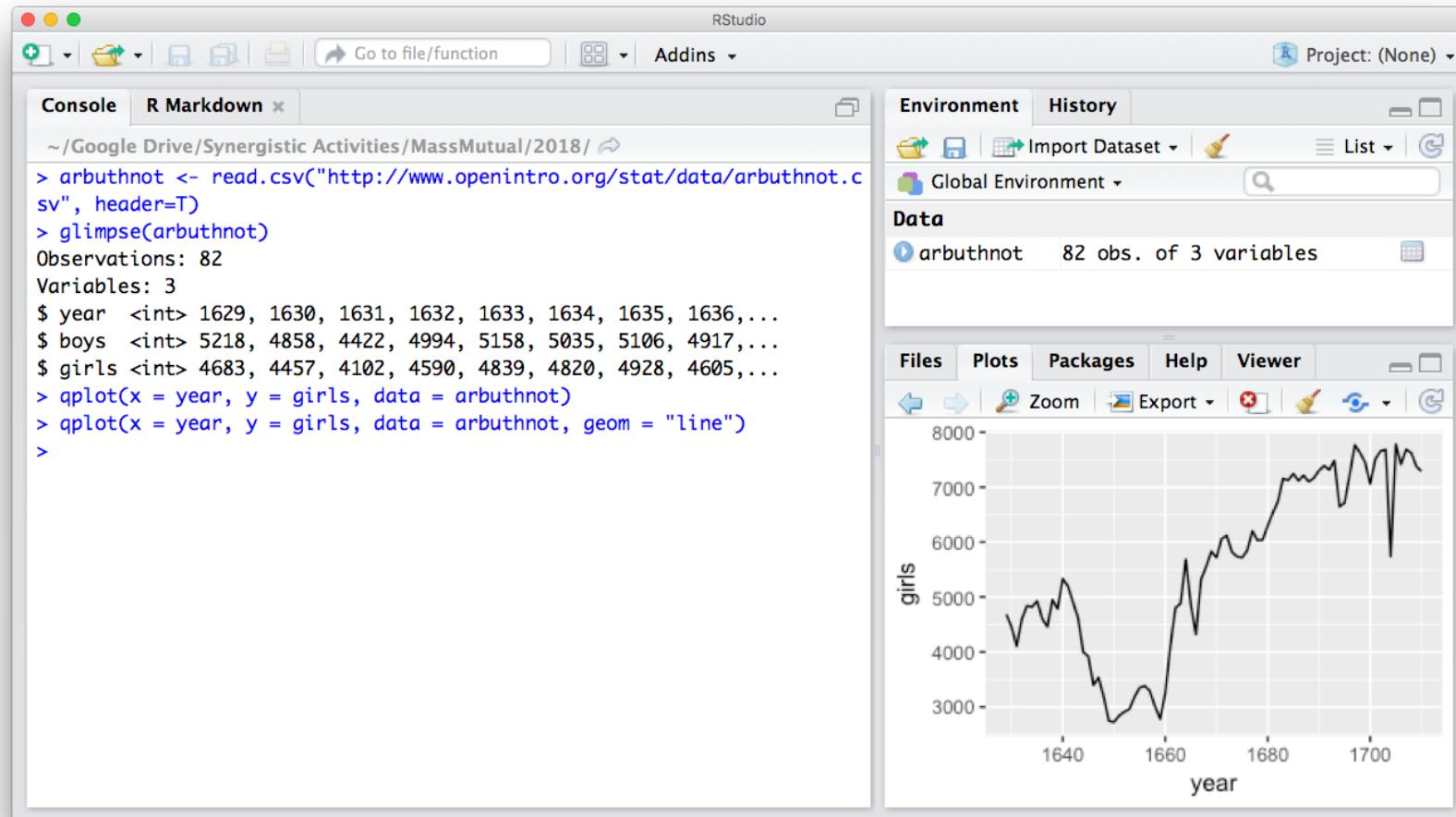
# Last week...



# Last week...



# Last week...



# Question

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

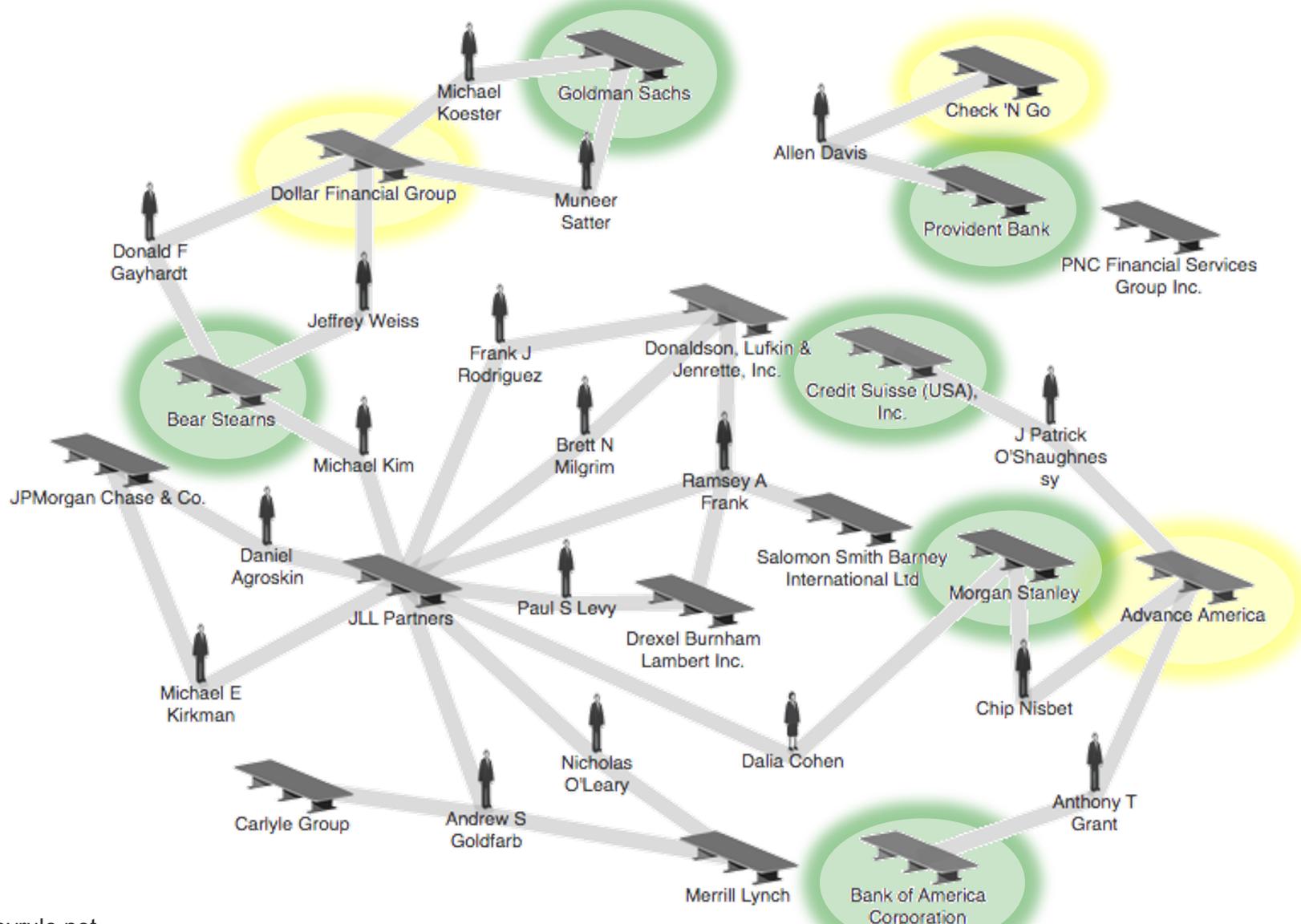


# Do they help you spot trends?



More info here: [http://en.wikipedia.org/wiki/1854\\_Broad\\_Street\\_cholera\\_outbreak](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.)

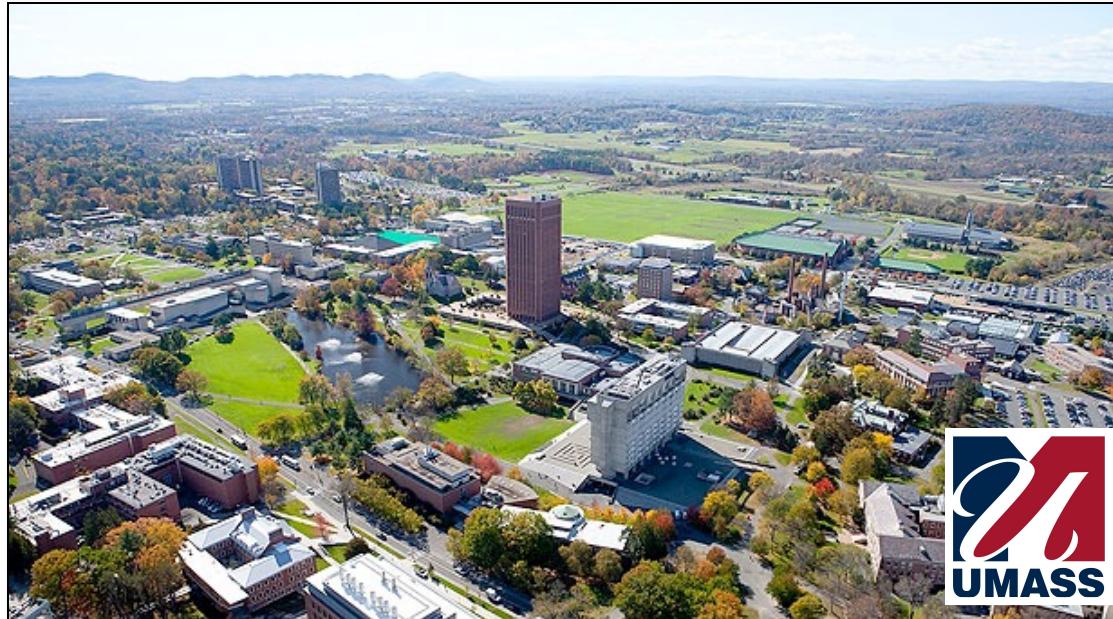
and a corresponding set of *observations* (a.k.a. *records*) over these variables. For example:



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

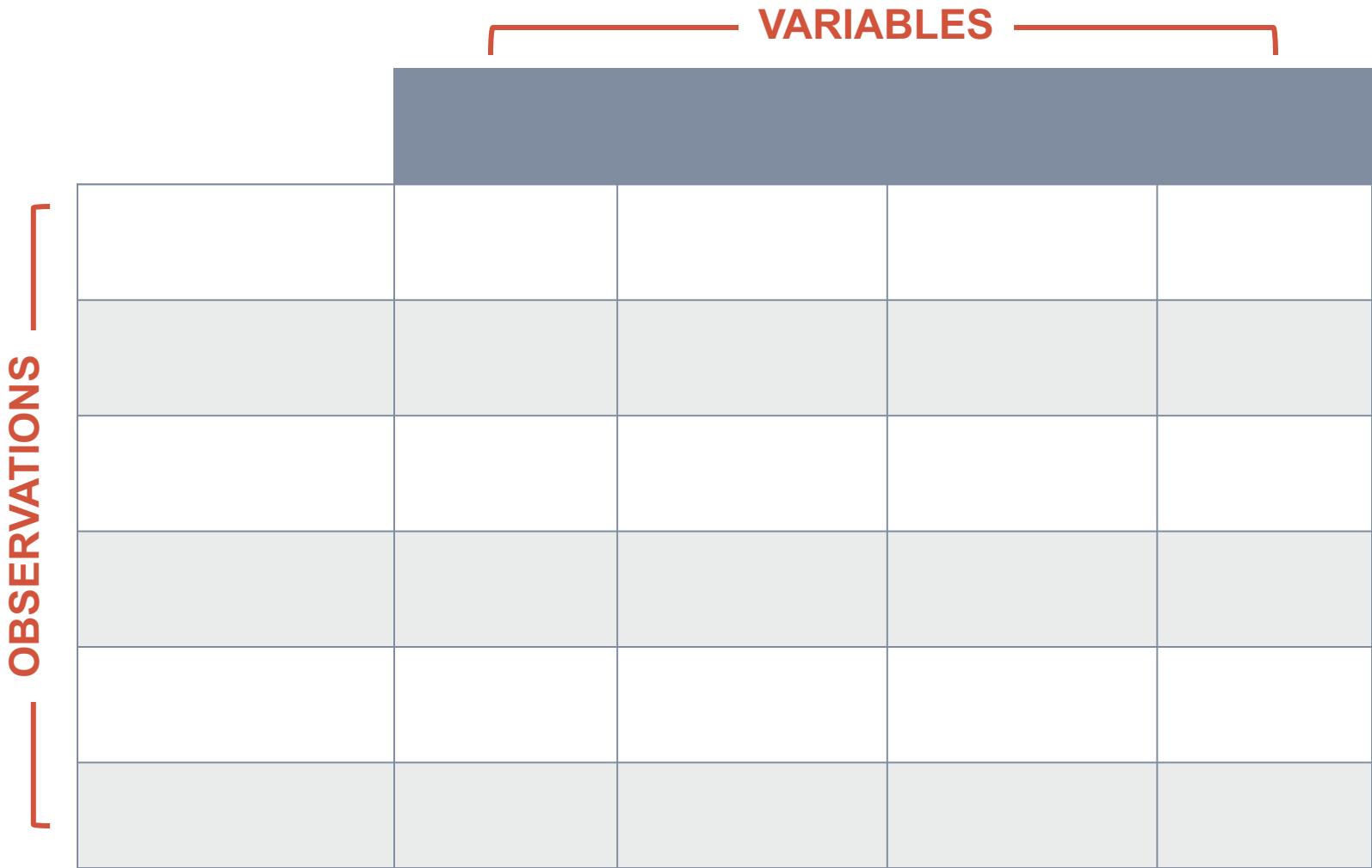
# Data (def.)

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:



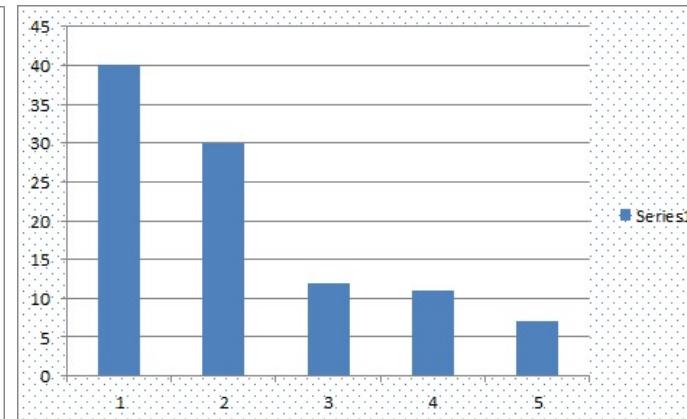
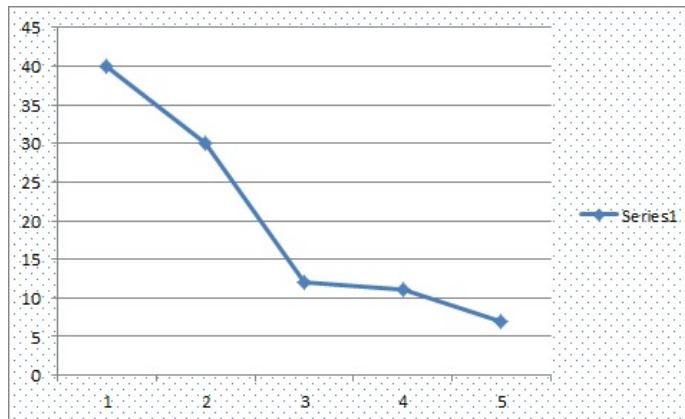
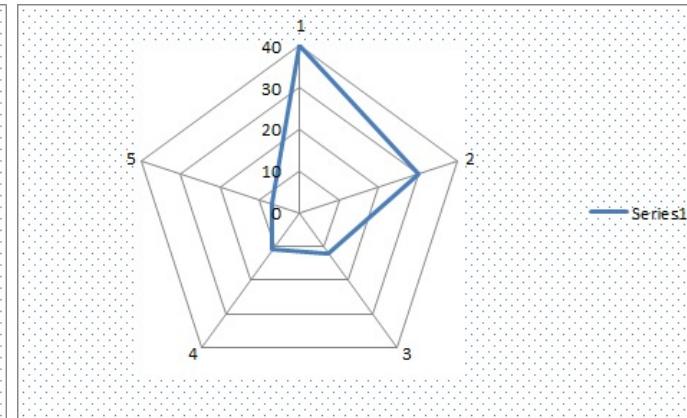
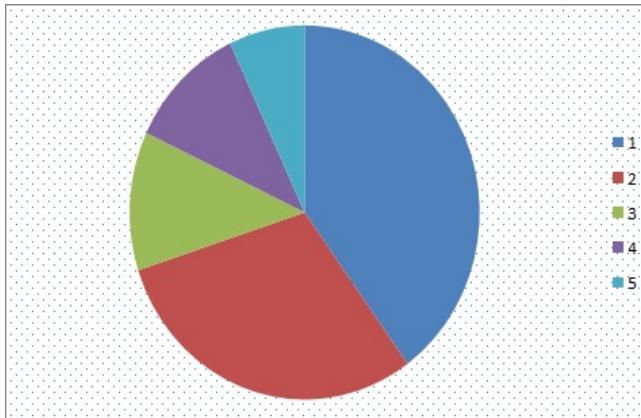
# Why is this important?

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

# Key question for this workshop

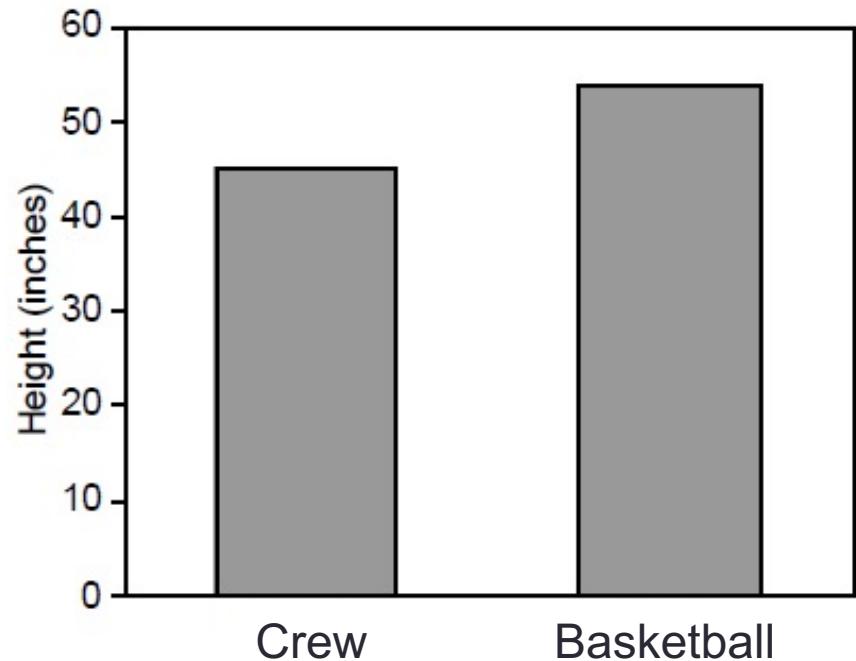
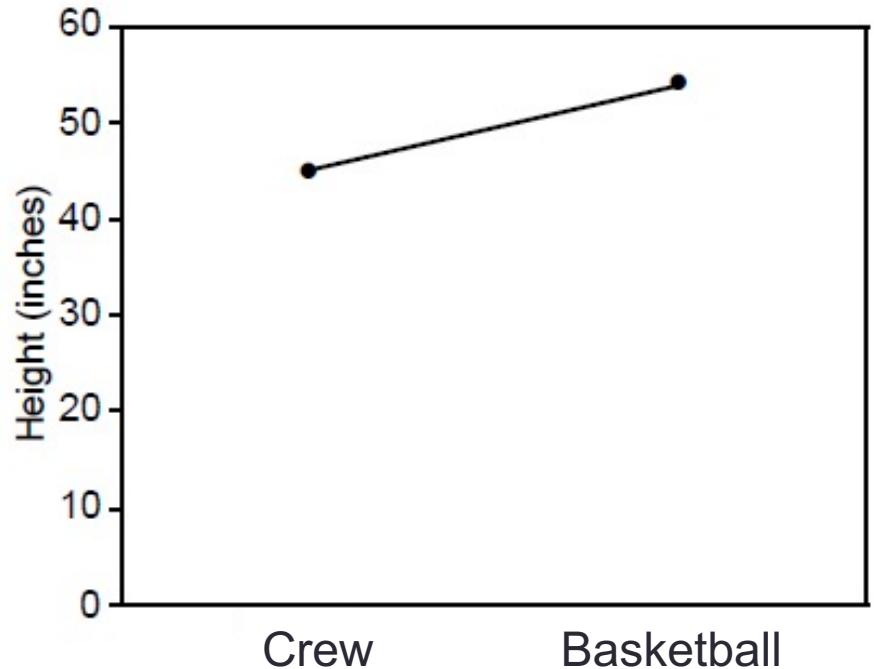
Which **data dimension** should be mapped

to which **visual dimension?**

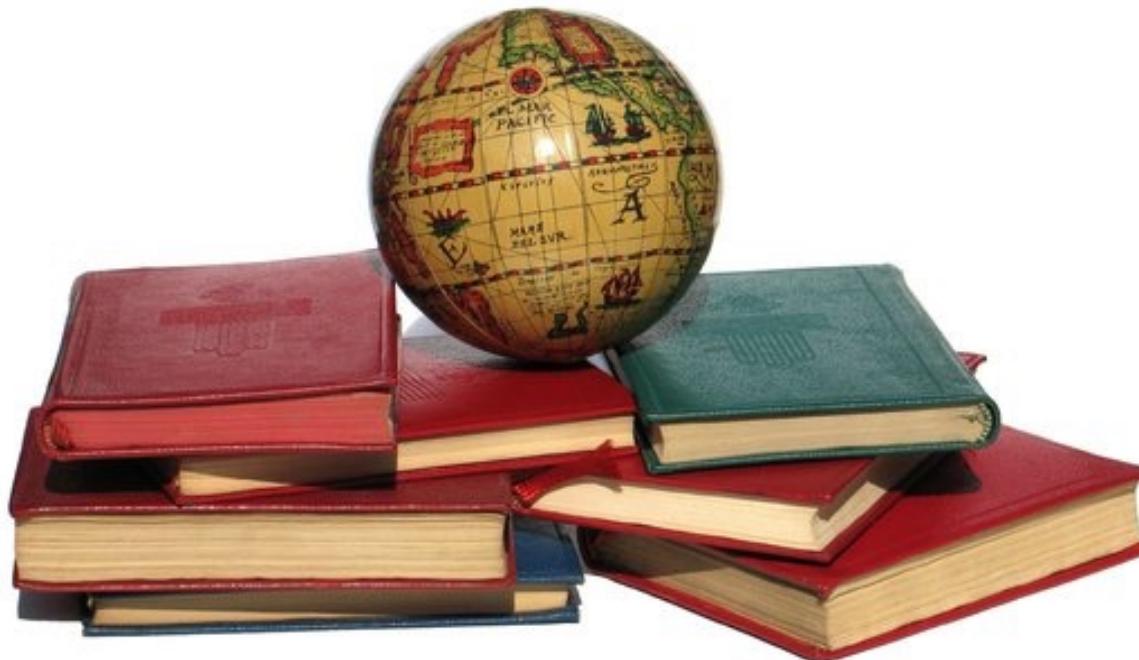


# Answer: it depends

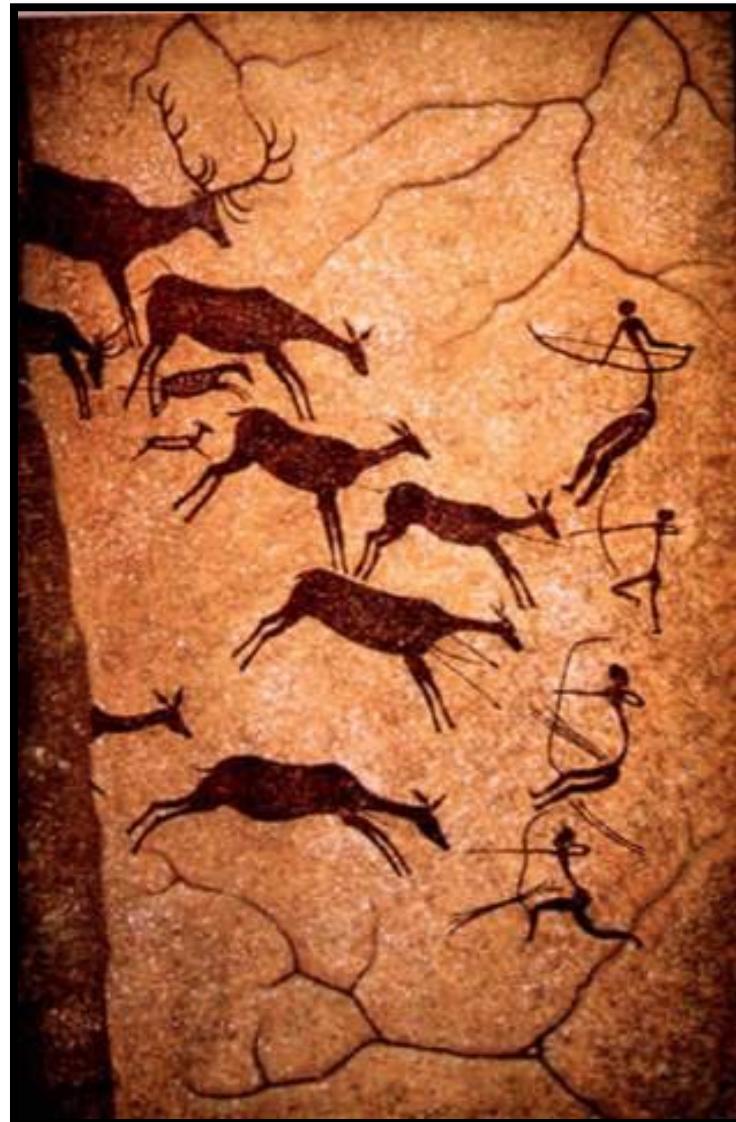
Average Height for Youth Sports Participants



# A quick history lesson...

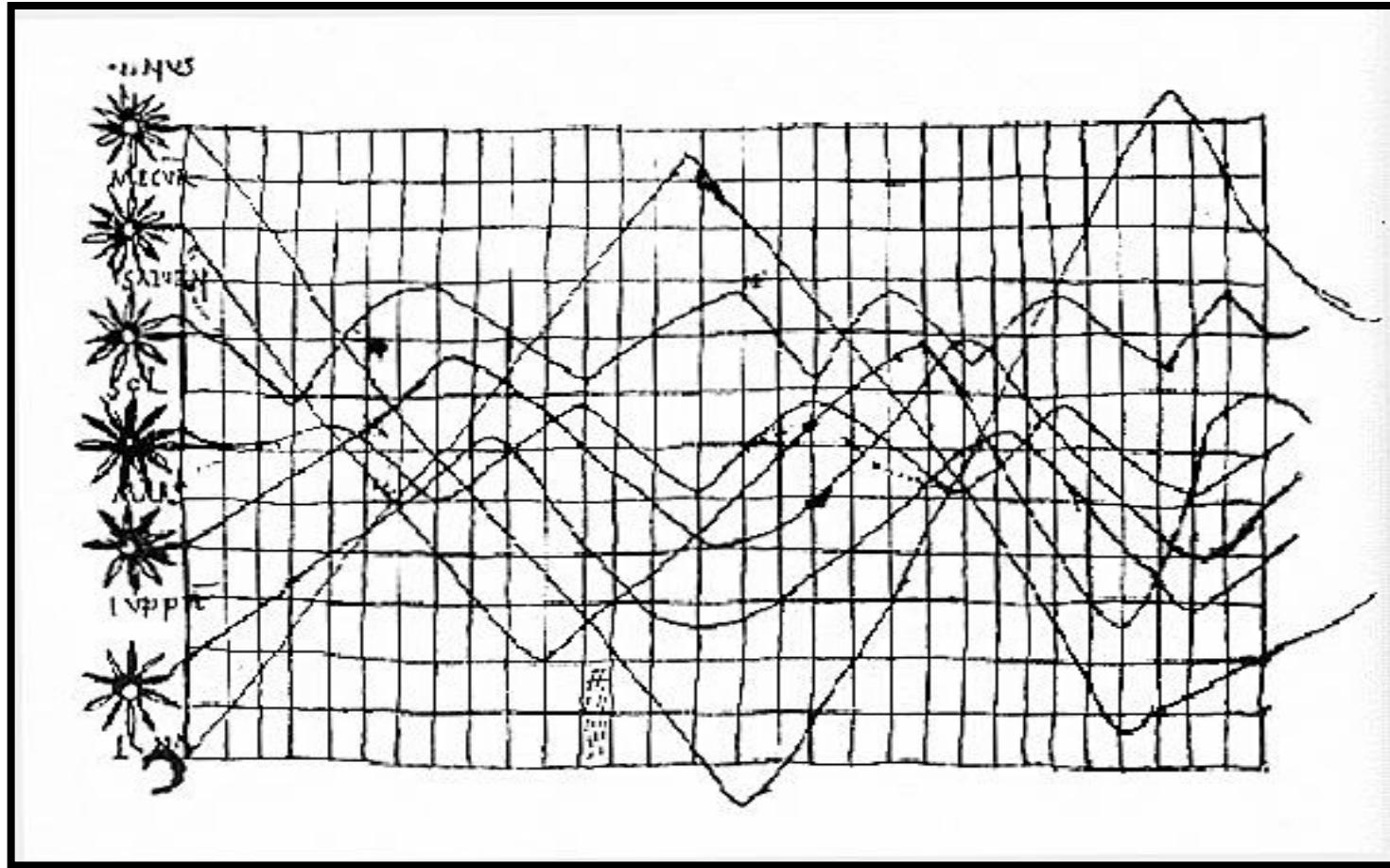


# (Incomplete) History of Visualization: 15,000BC



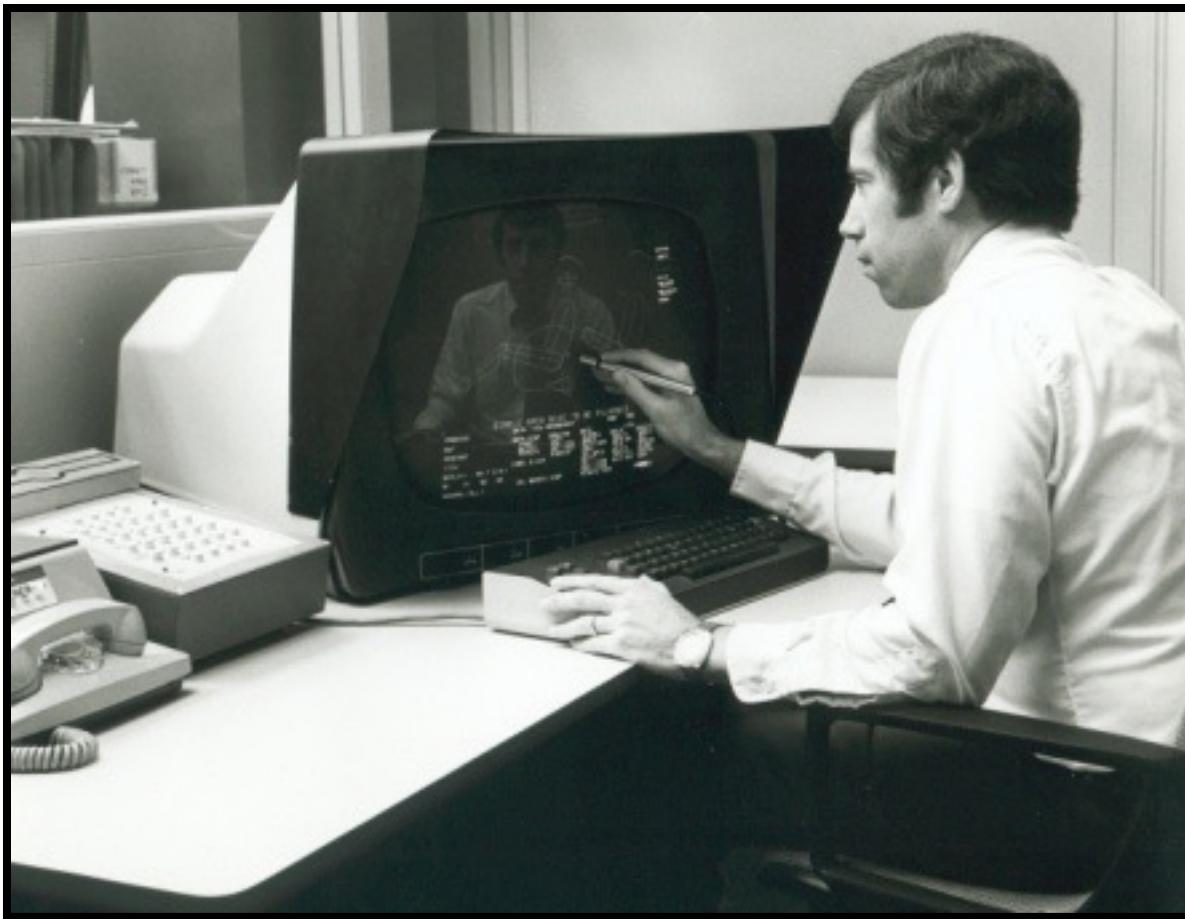
15,000 BC. Laxcaux, France

# (Incomplete) History of Visualization: 900s



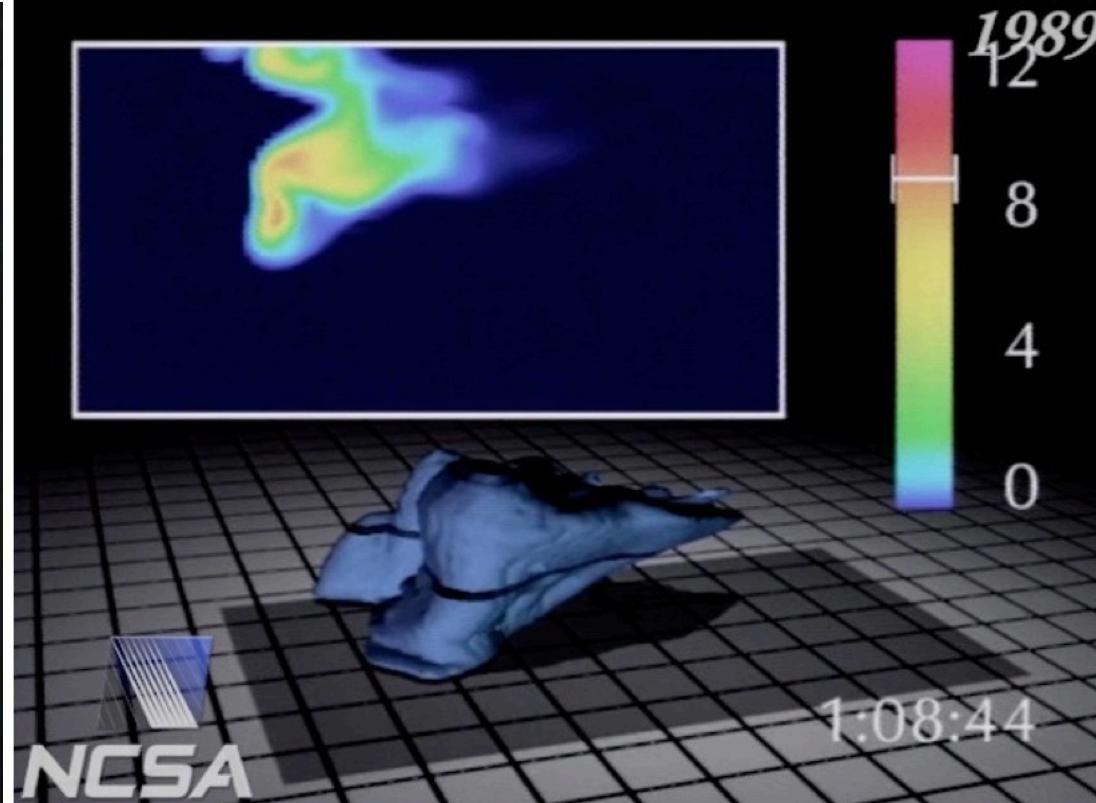
"De cursu per zodiacum", illustrator unknown

# (Incomplete) History of Visualization: 1970s



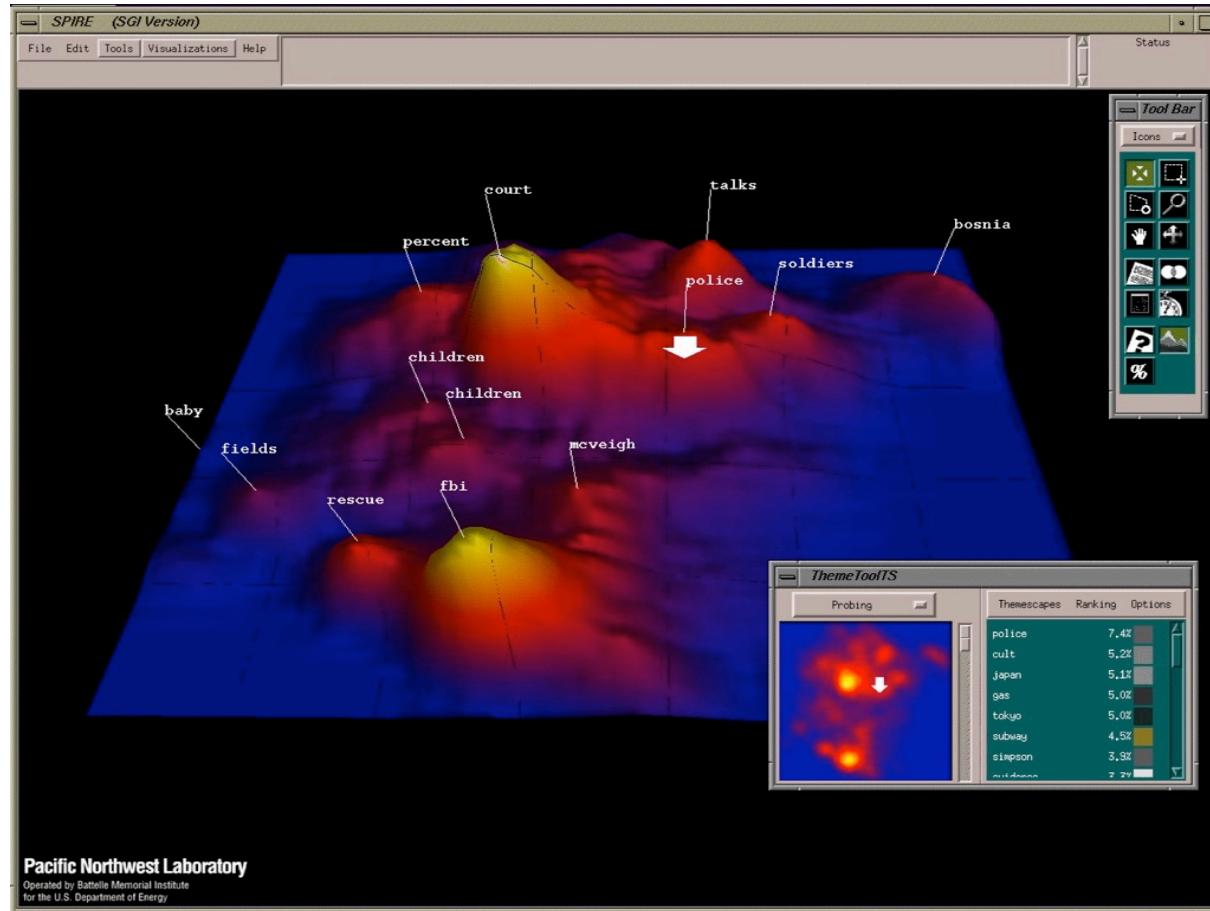
- CAD/CAM, building cars, planes, chips
- Starting to think about: 3D, animation, edu, medicine

# (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

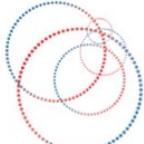
# (Incomplete) History of Visualization: 2000s



- 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



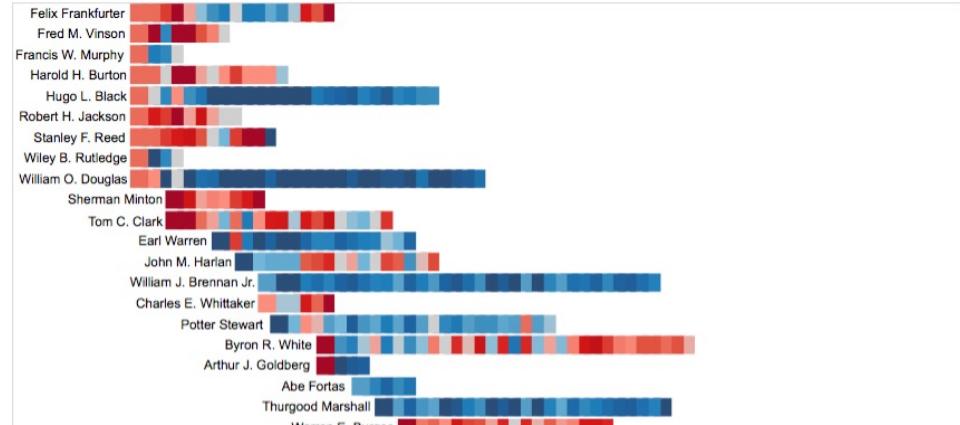
## Justices by Ideology

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

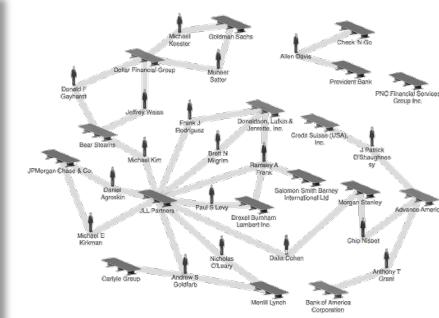
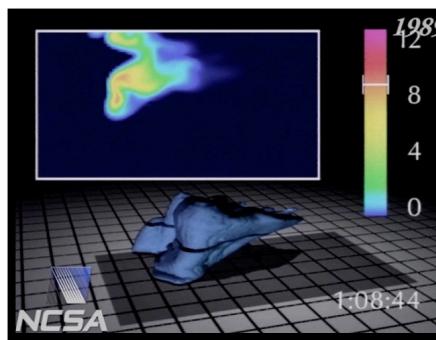
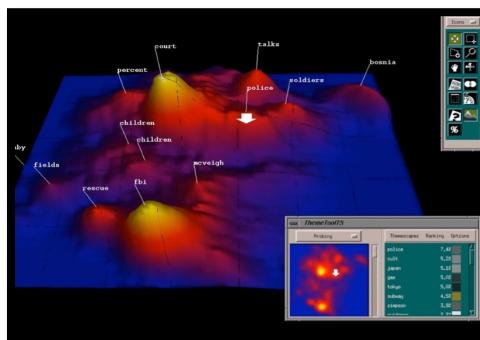
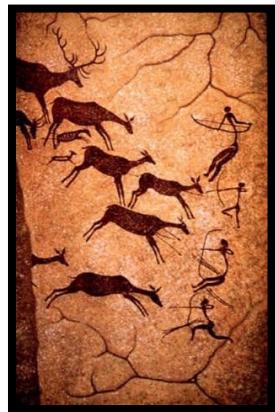
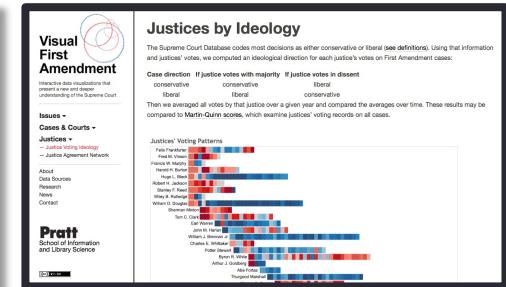
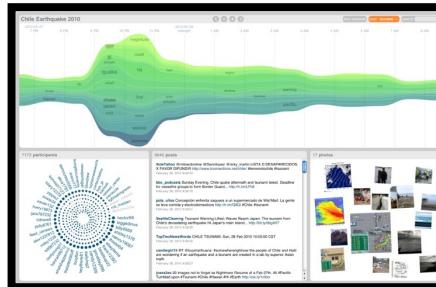
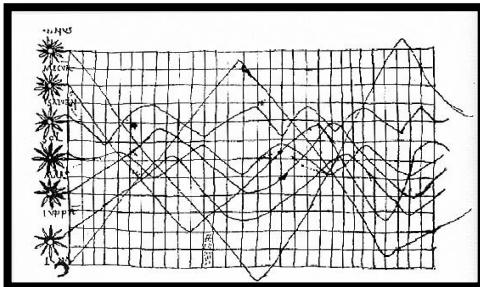
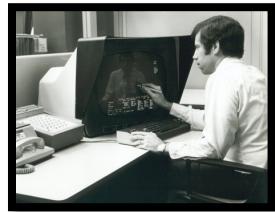
Then we averaged all votes by that justice over a given year and compared the averages over time. These results may be compared to [Martin-Quinn scores](#), which examine justices' voting records on all cases.

### Justices' Voting Patterns

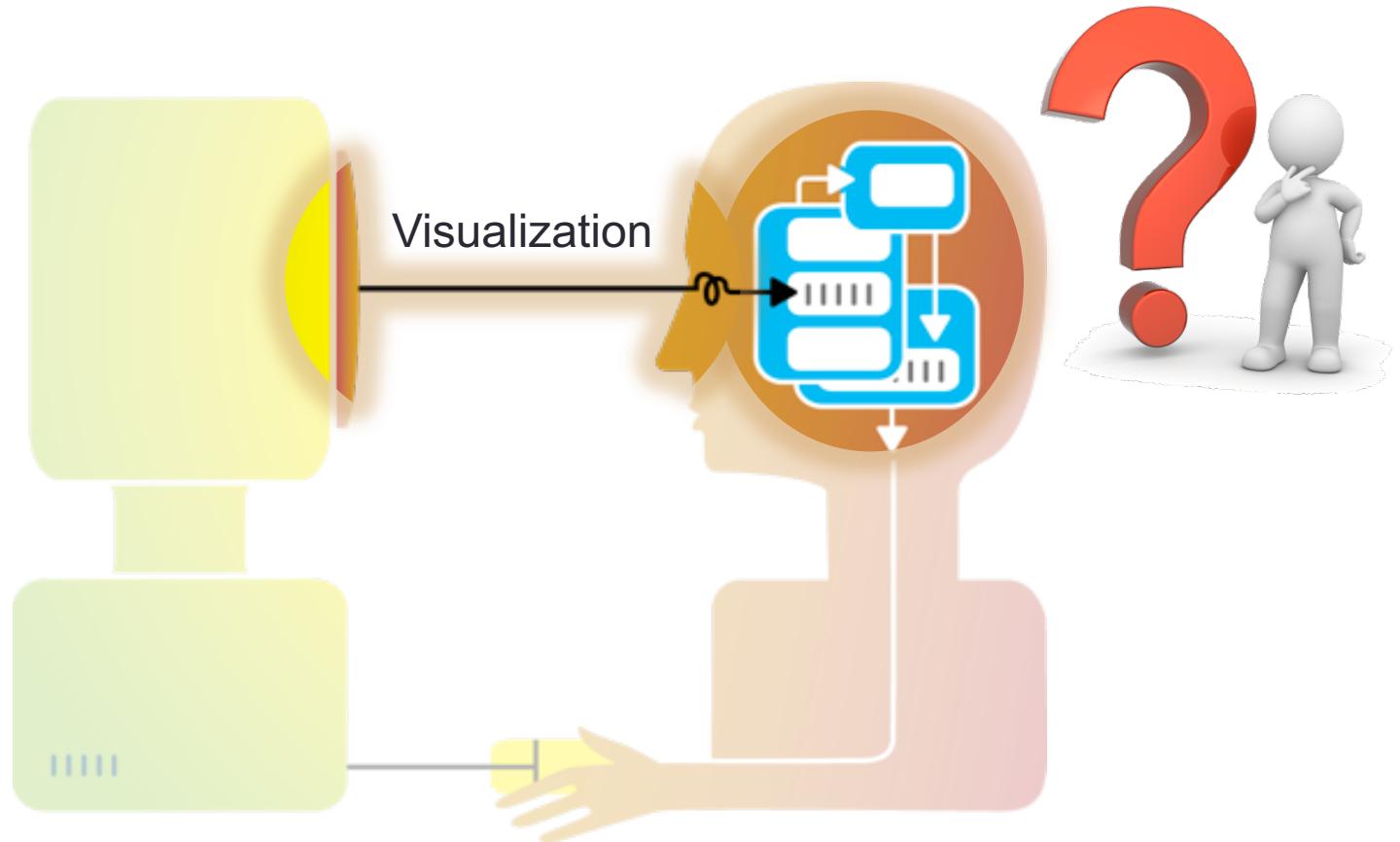


- Human-machine collaboration, machine learning
- Starting to think about: evaluation, new media, DH

# 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

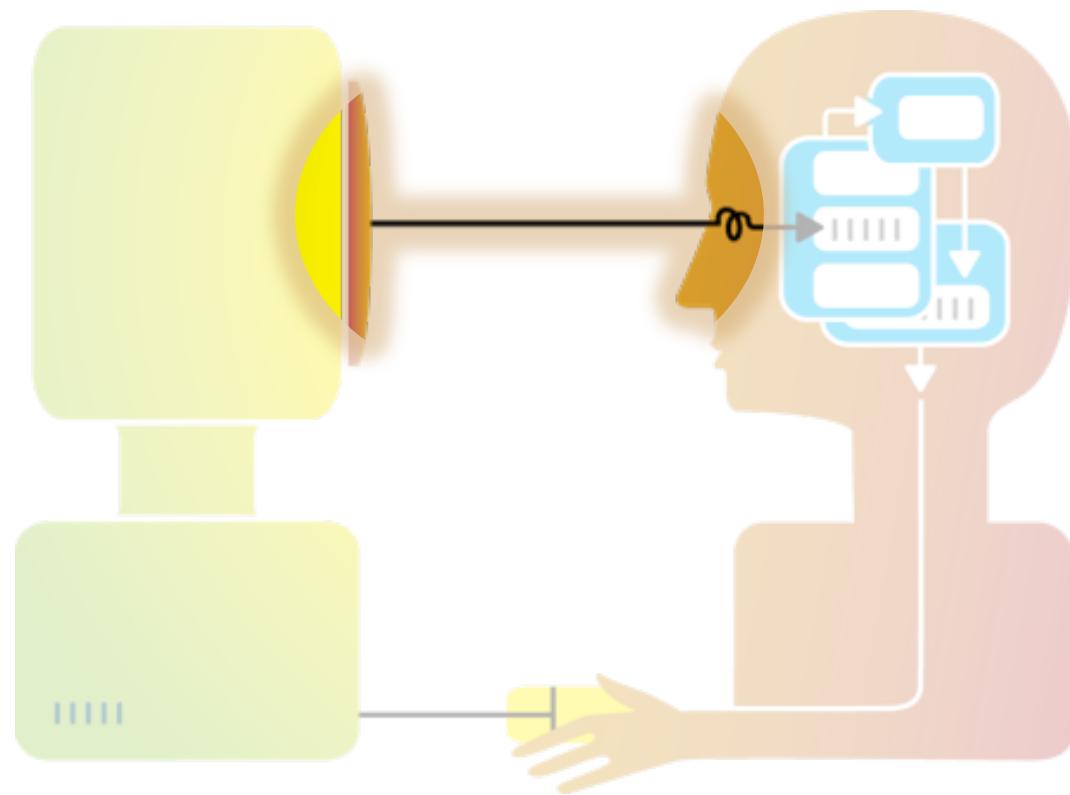
## The bad:

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

## The ugly:

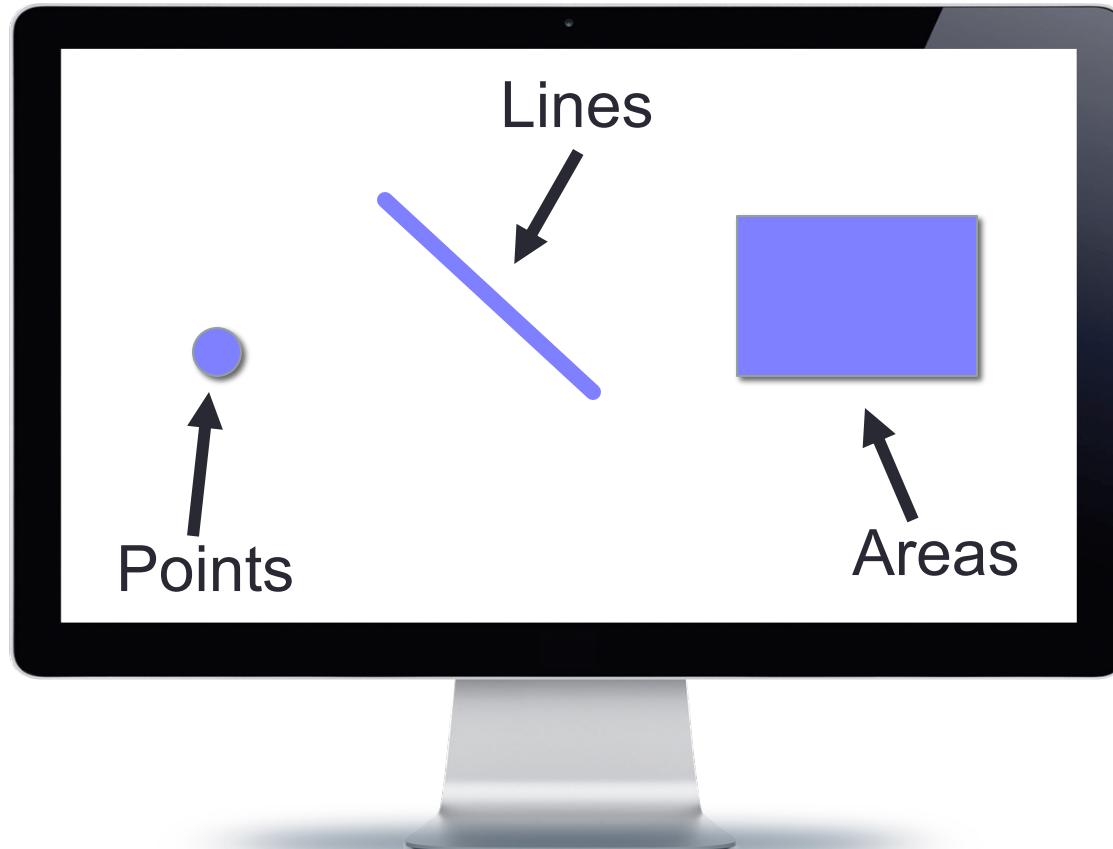
- Mental models are unavoidable: everyone has them, and they’re all different
- **Key:** be aware of how mental models form, how they shape perception, and how to support (or challenge) them

# So what do we have to work with?



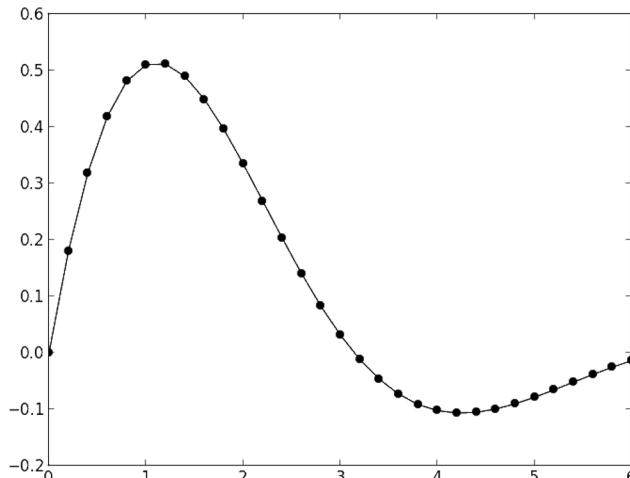
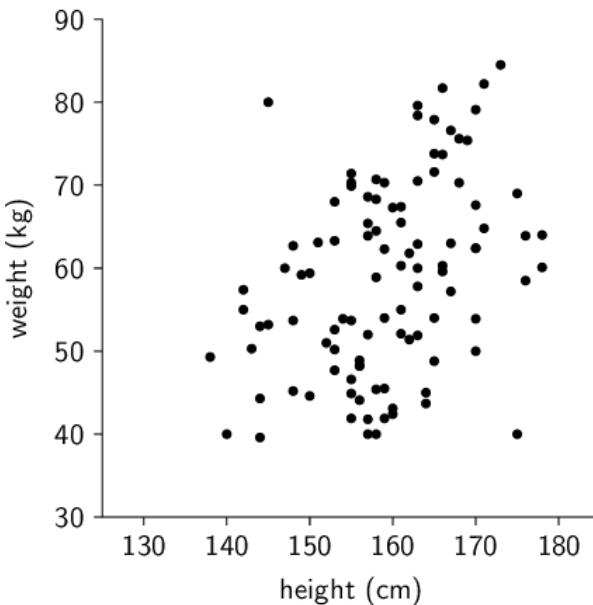
# Graphical primitives

The images we draw are composed of marks: like ink



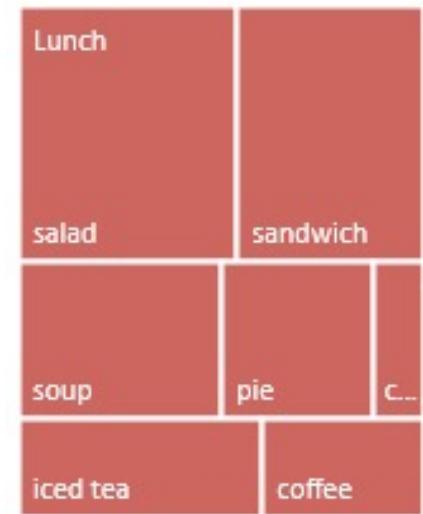
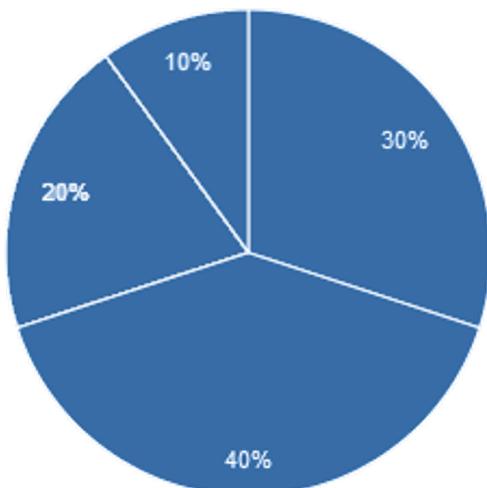
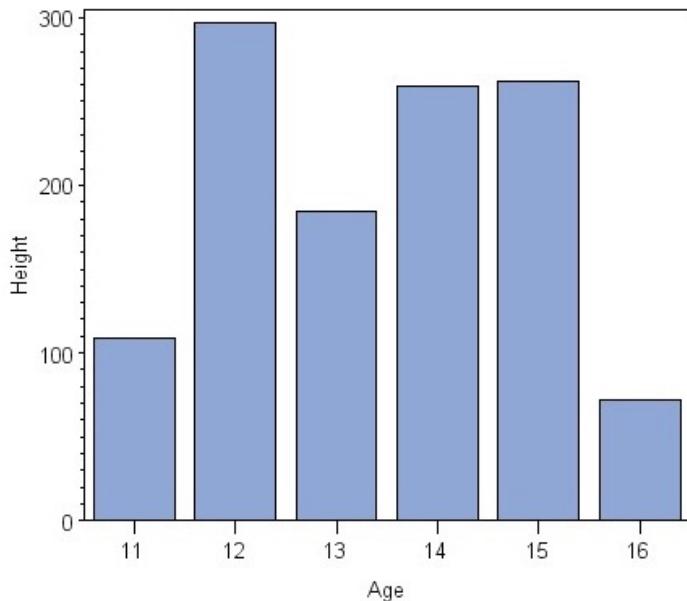
# Visual dimension: position

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



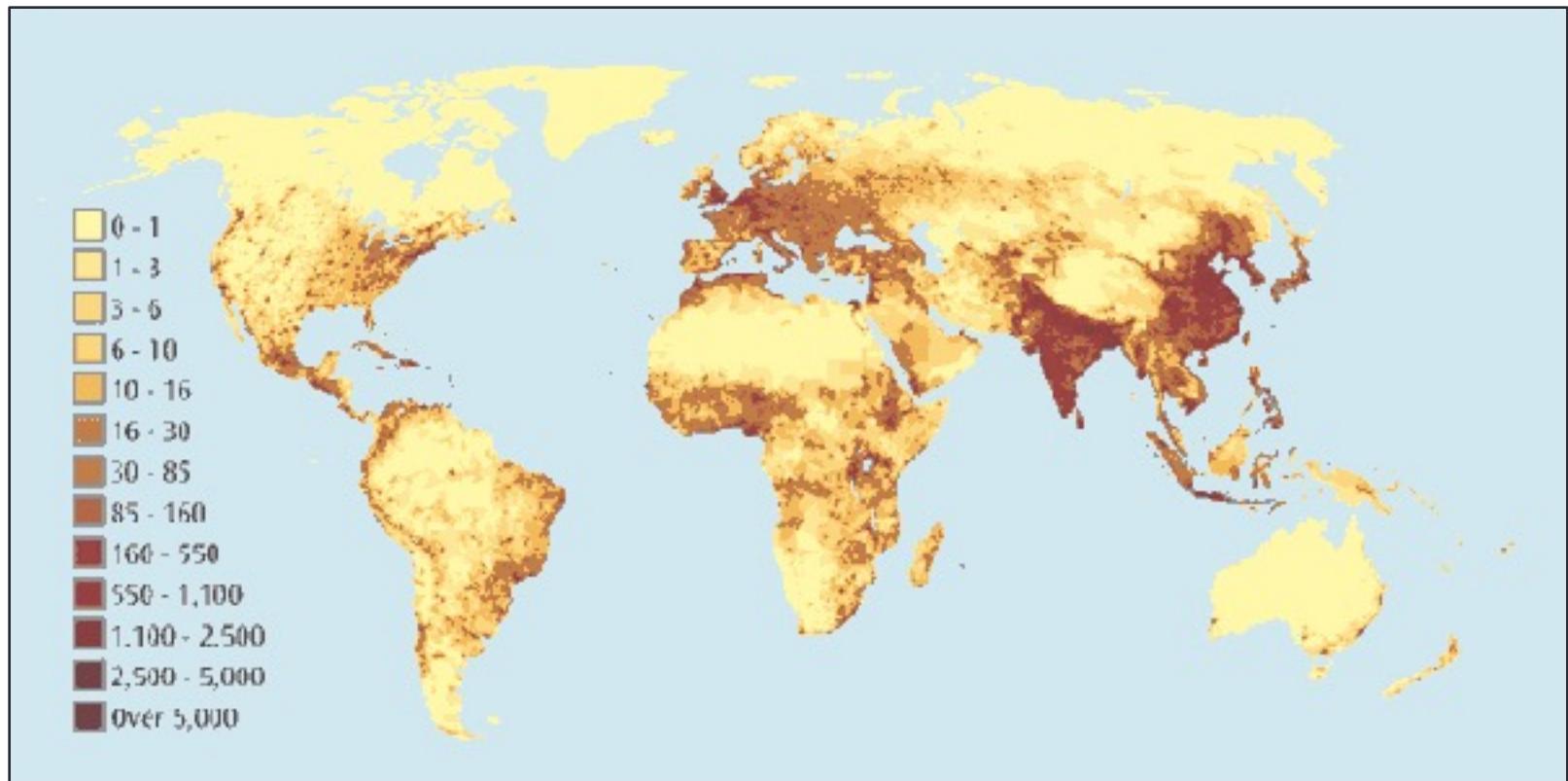
# Visual dimension: size

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



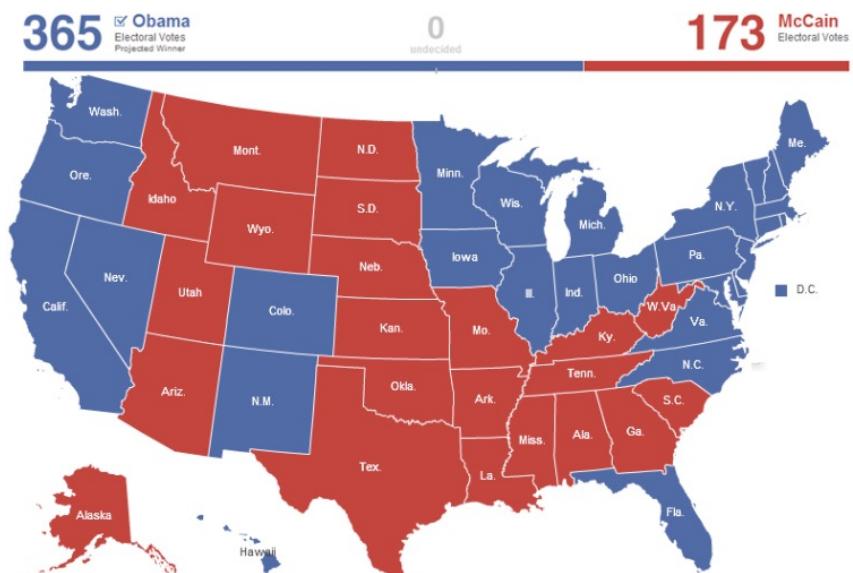
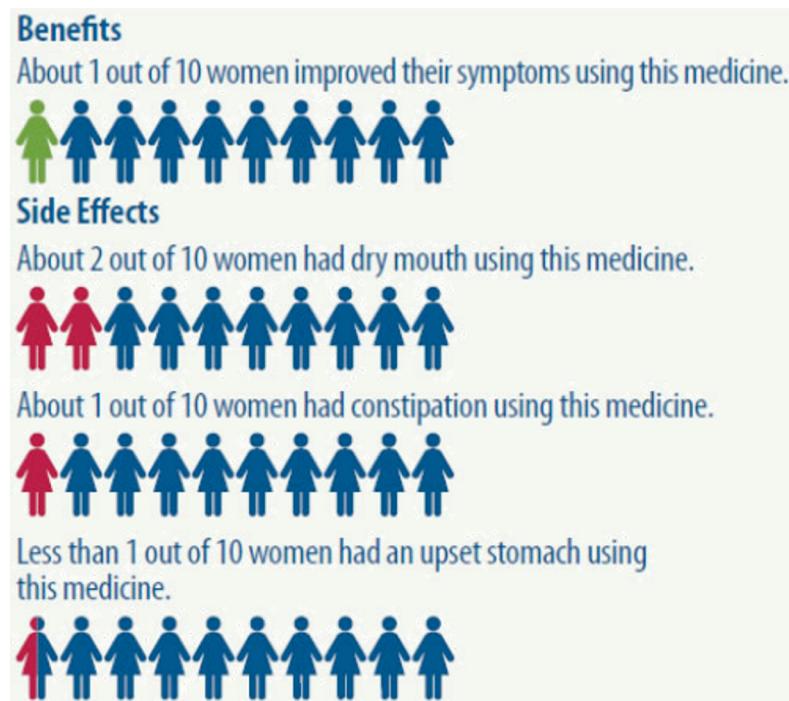
# Visual dimension: value

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



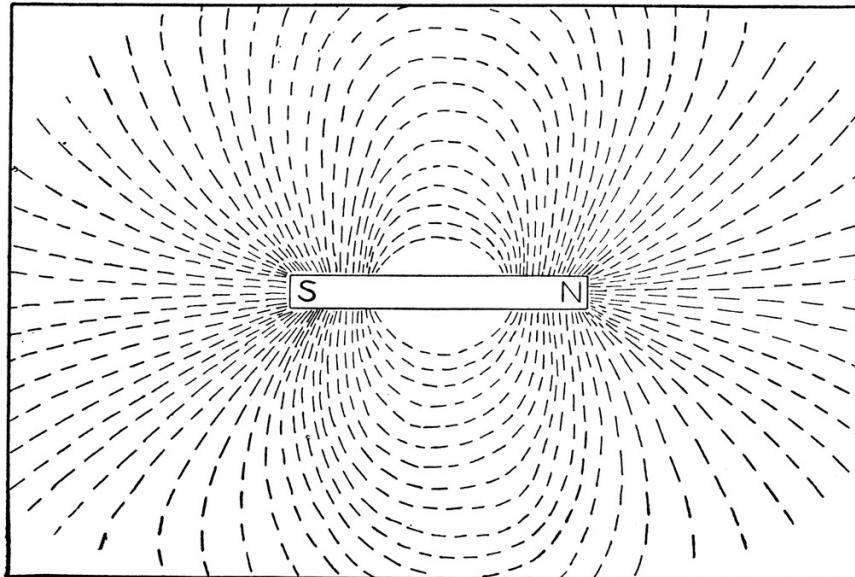
# Visual dimension: color

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



# 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

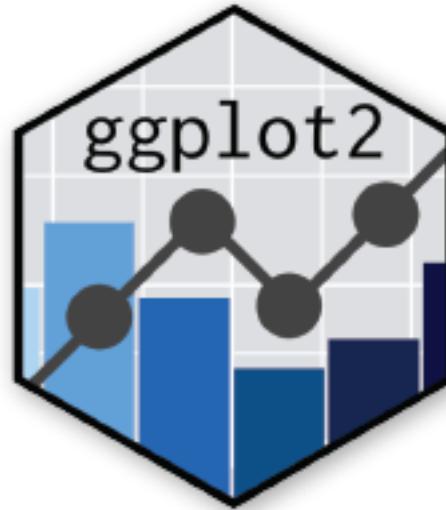
	Categorical	Numerical
POSITION	x x x	
SIZE	■ ■ ■	
VALUE	■ ■ ■	
COLOR	■ ■ ■	
ORIENTATION	■ ■ ■	
SHAPE	■ ▲ ●	



# 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')`