

Communicating with Data – Mental Models & Data-Visual Mapping

Dr. Ab Mosca (they/them)

Slides based off slides courtesy of Jordan Crouser (<https://jcrouser.github.io/>)

Plan for Today

- A quick history lesson
- Mental models: how we process information
- Visualization building blocks
- Takeaways

Checking in

- Are you on Slack?
 - All communication will be through Slack (see syllabus for details)
- Did you join Gradescope?
- Did you submit lab01 (or Using Gradescope)?

Looking forward

- Hwo1 is released today!
- Find instructions on the course website under the “Homework” tab
- Submit on Gradescope
- Consider working with a partner! Collaboration is encouraged 😊

Looking forward



- Tableau is a drag-and-drop visualization tool suite
- Tableau for Teaching has donated license keys (good for one year) for everyone enrolled in this course
- Instructions for downloading are on the next slide. Download before class on Tuesday 02/11
- Need help? Ping me on Slack, or come to office hours

Preparing Tableau



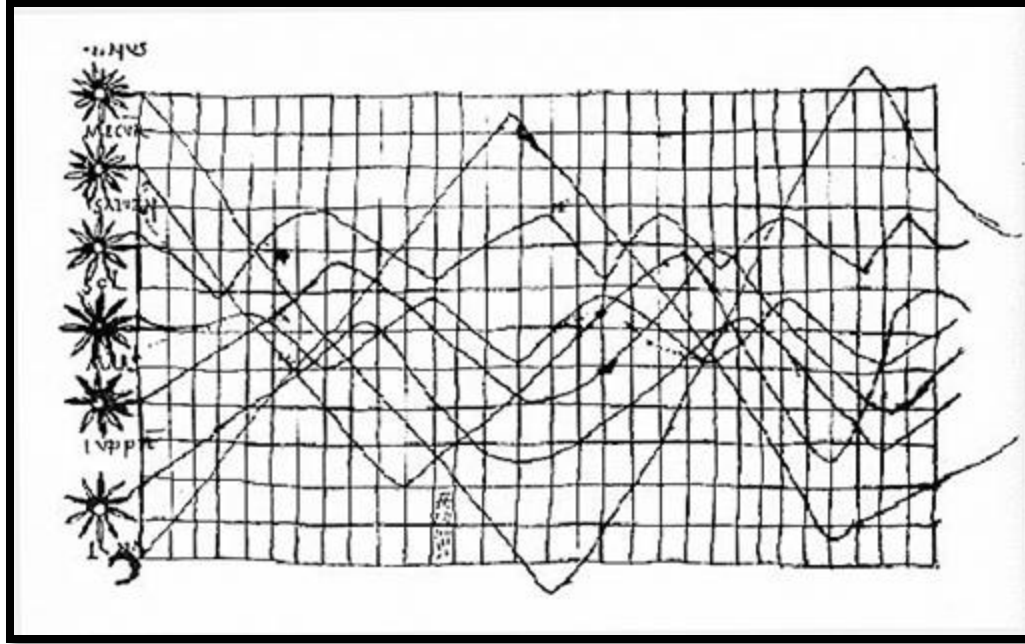
1. Download the latest version of Tableau Desktop and Tableau Prep Builder:
<https://www.tableau.com/tft/activation>
2. Click on the link above and select “Download Tableau Desktop” and “Download Tableau Prep Builder”. On the form, enter your school email address for Business E-mail and enter the name of your school for Organization.
3. Activate with product key: **TCR8-EF41-95Bo-oD5E-FFAE**
4. Already have a copy of Tableau Desktop installed? Update your license in the application: Help menu → Manage Product Keys

(Incomplete)
History of
Visualization:
15,000BC



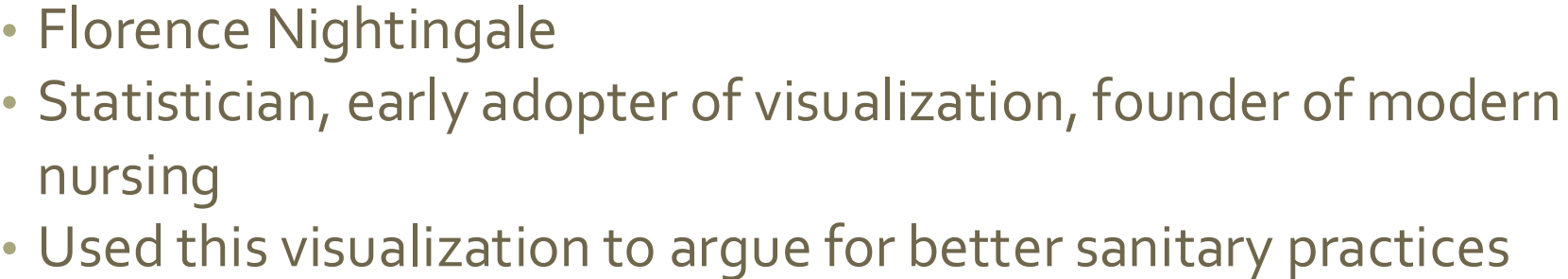
15,000 BC. Laxcaux, France

(Incomplete) History of Visualization: 900s

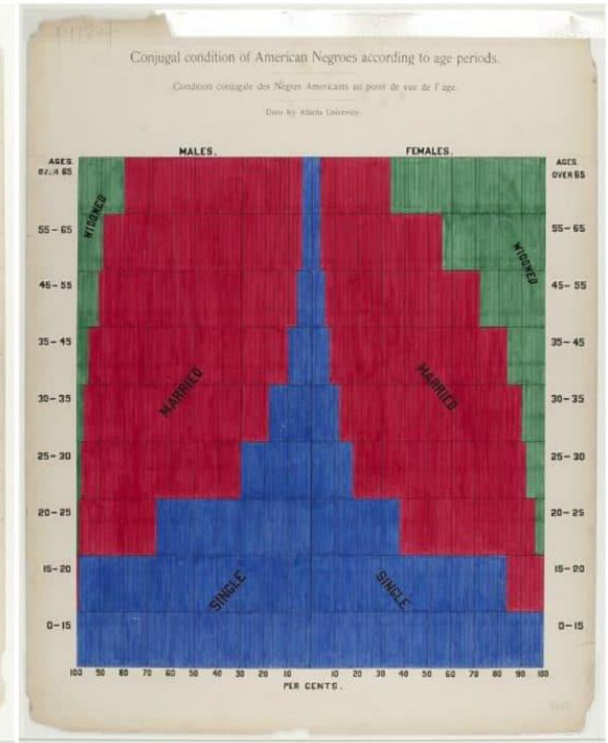
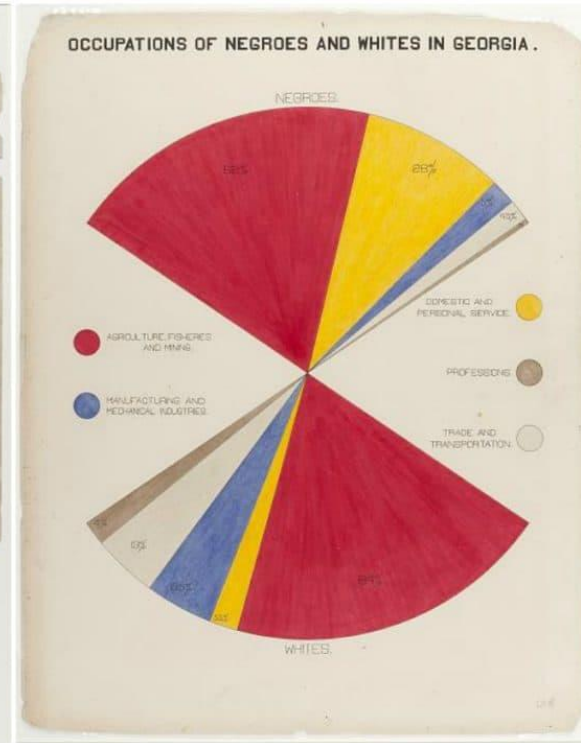
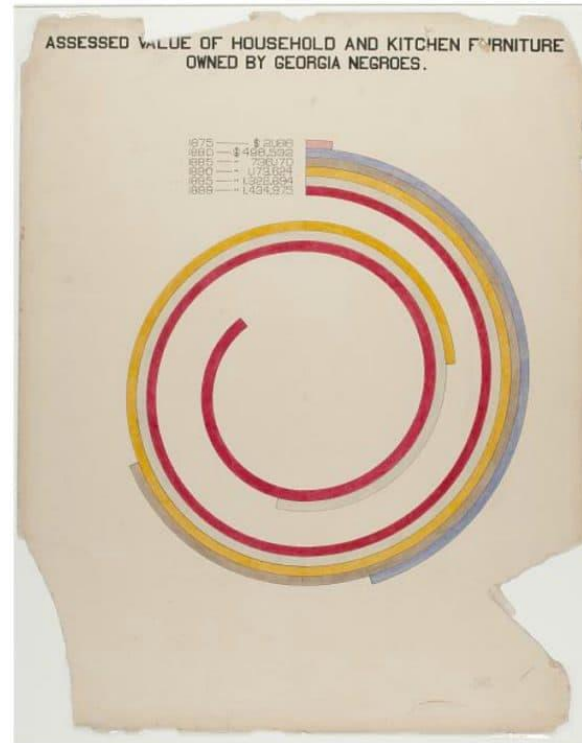


- Oldest known attempt to show changing values graphically
- Inclinations of the planetary orbits over time

(Incomplete)
History of
Visualization:
mid-1800s



(Incomplete) History of Visualization: mid-1800s



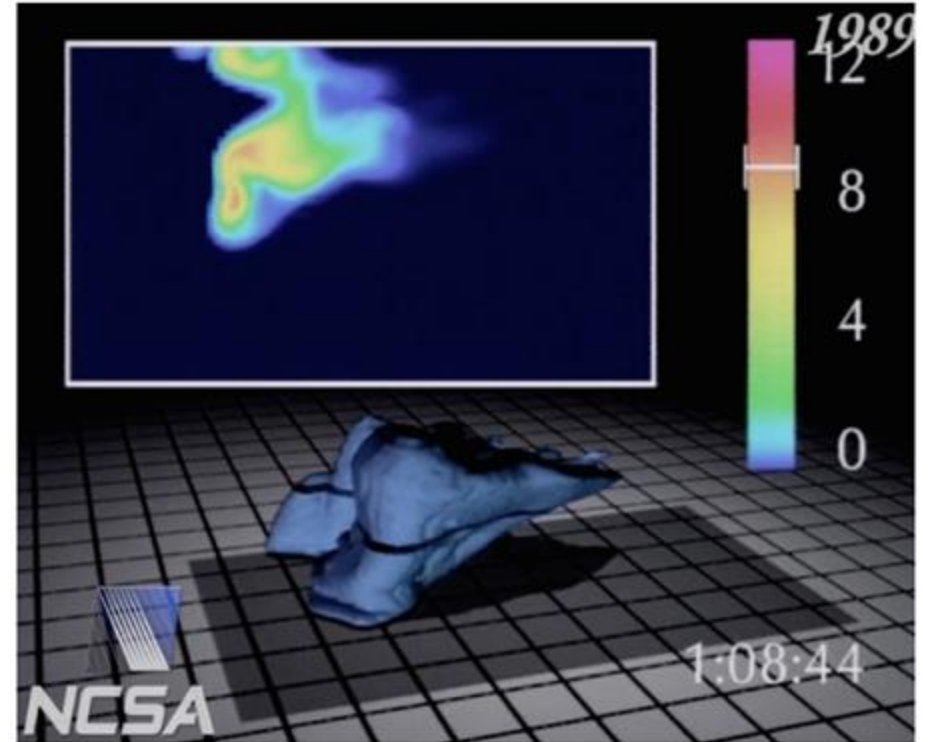
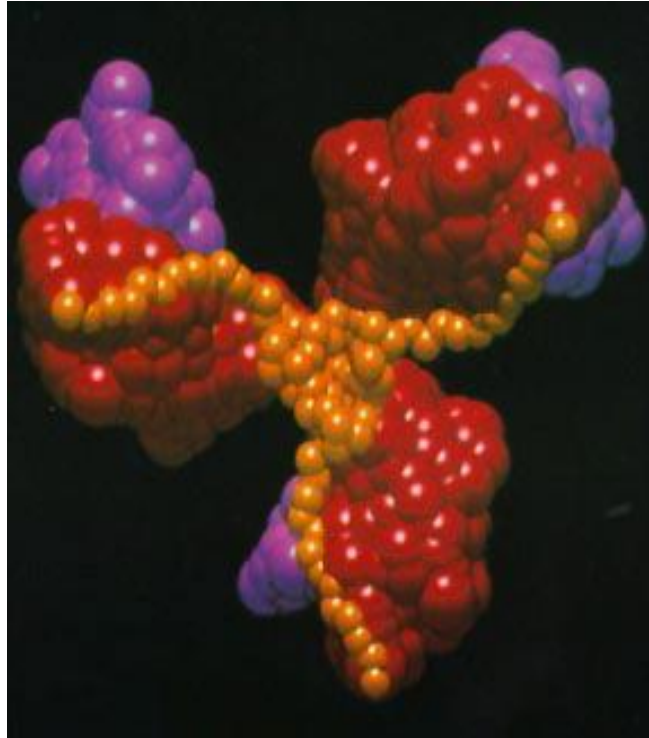
- W.E.B. Du Bois
- Sociologist
- "Data Portraits" were displayed at the Paris Exposition in 1900 to challenge norms and show how Black folks fit into American progress

(Incomplete) History of Visualization: 1970s



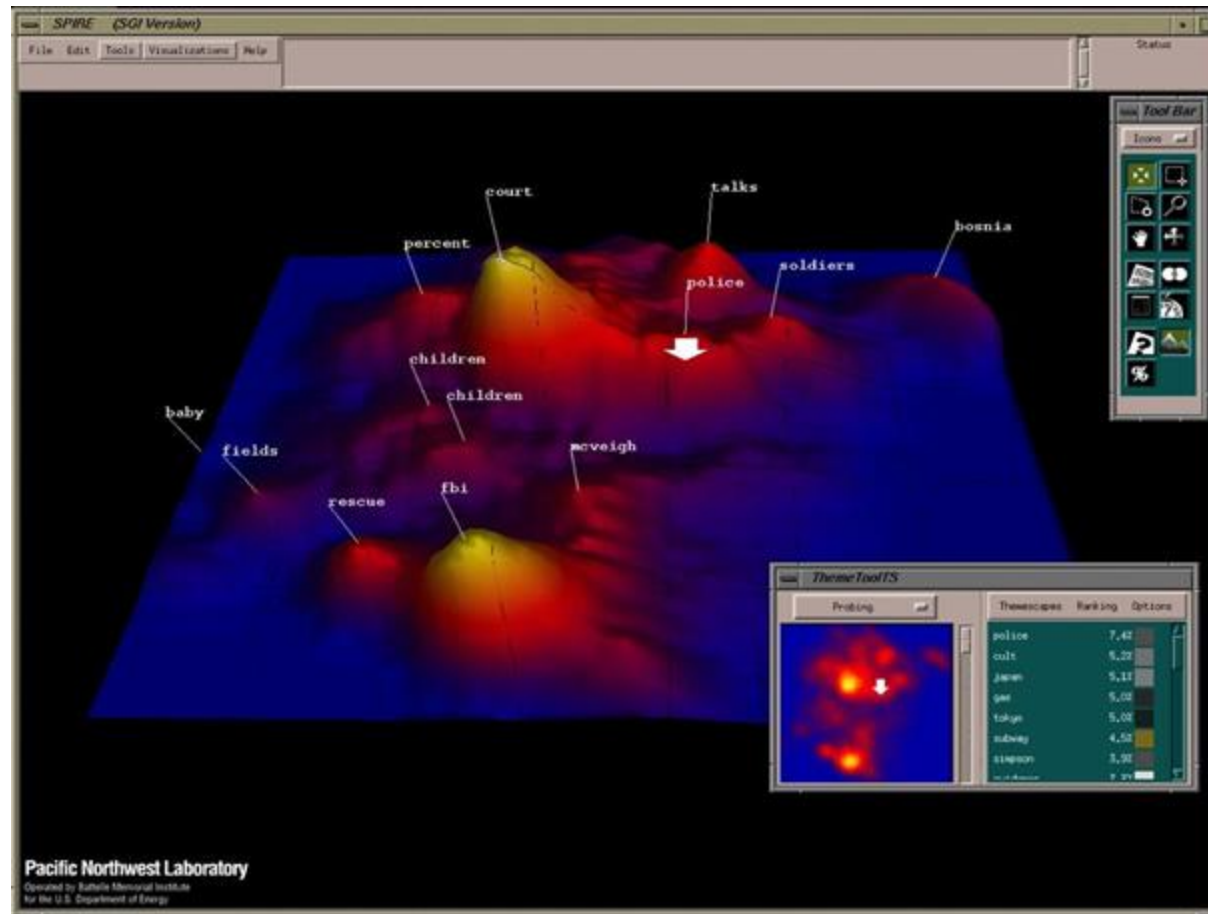
- CAD/CAM, building cars, planes, chips
- Starting to think about: 3D, animation, edu, medicine

(Incomplete) History of Visualization: 1980s



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

(Incomplete) History of Visualization: 1990s



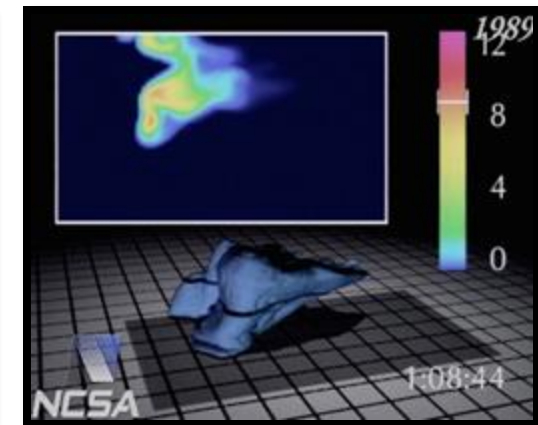
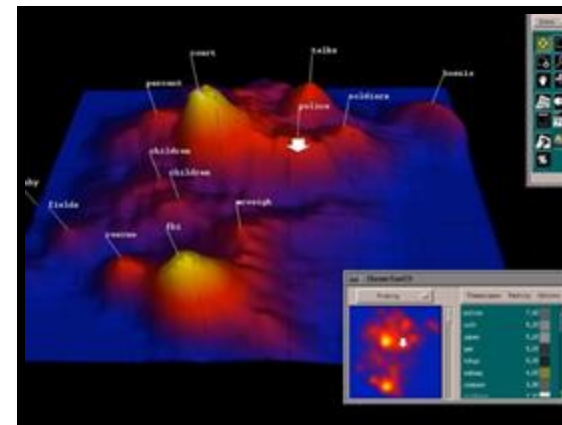
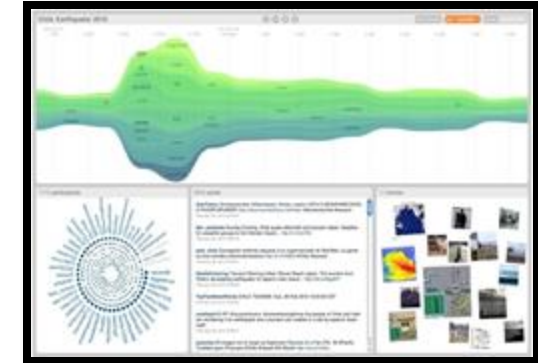
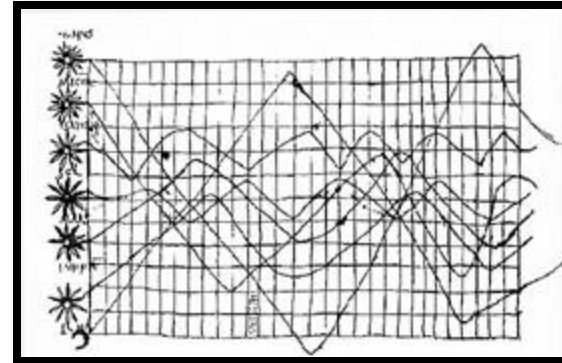
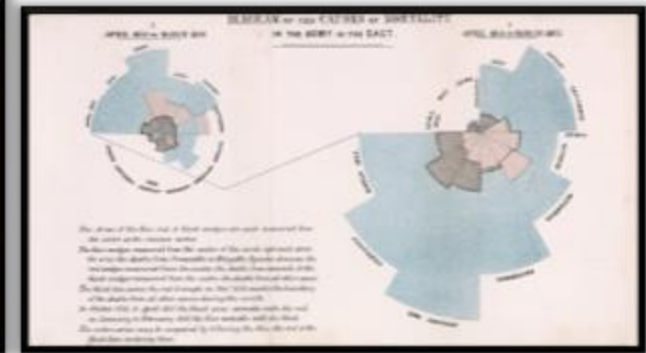
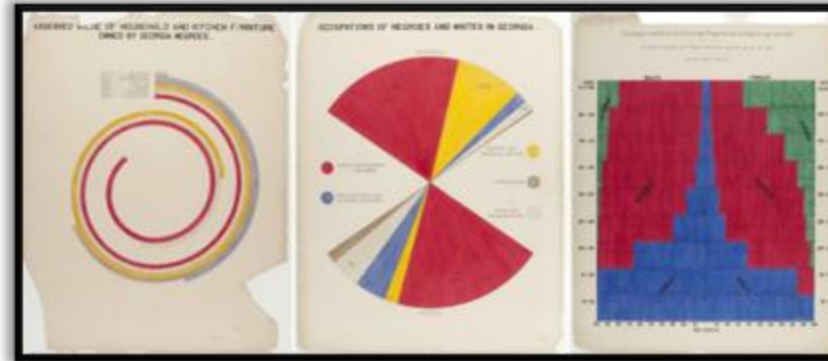
- Information visualization, storytelling
- Starting to think about: human cognition, interaction

(Incomplete) History of Visualization: 2000s

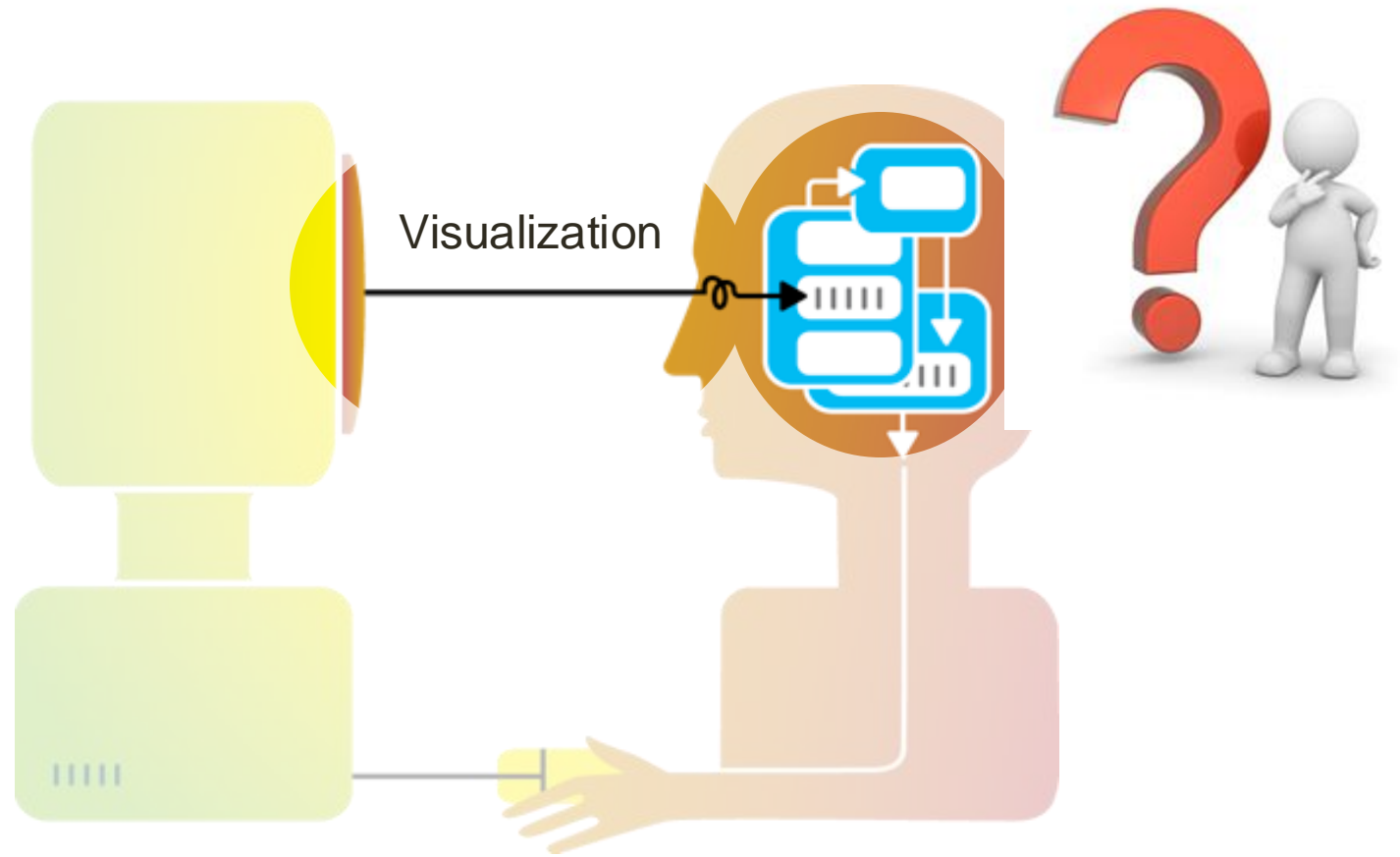


- Coordination across multiple views, interaction
- Starting to think about: sensemaking, provenance

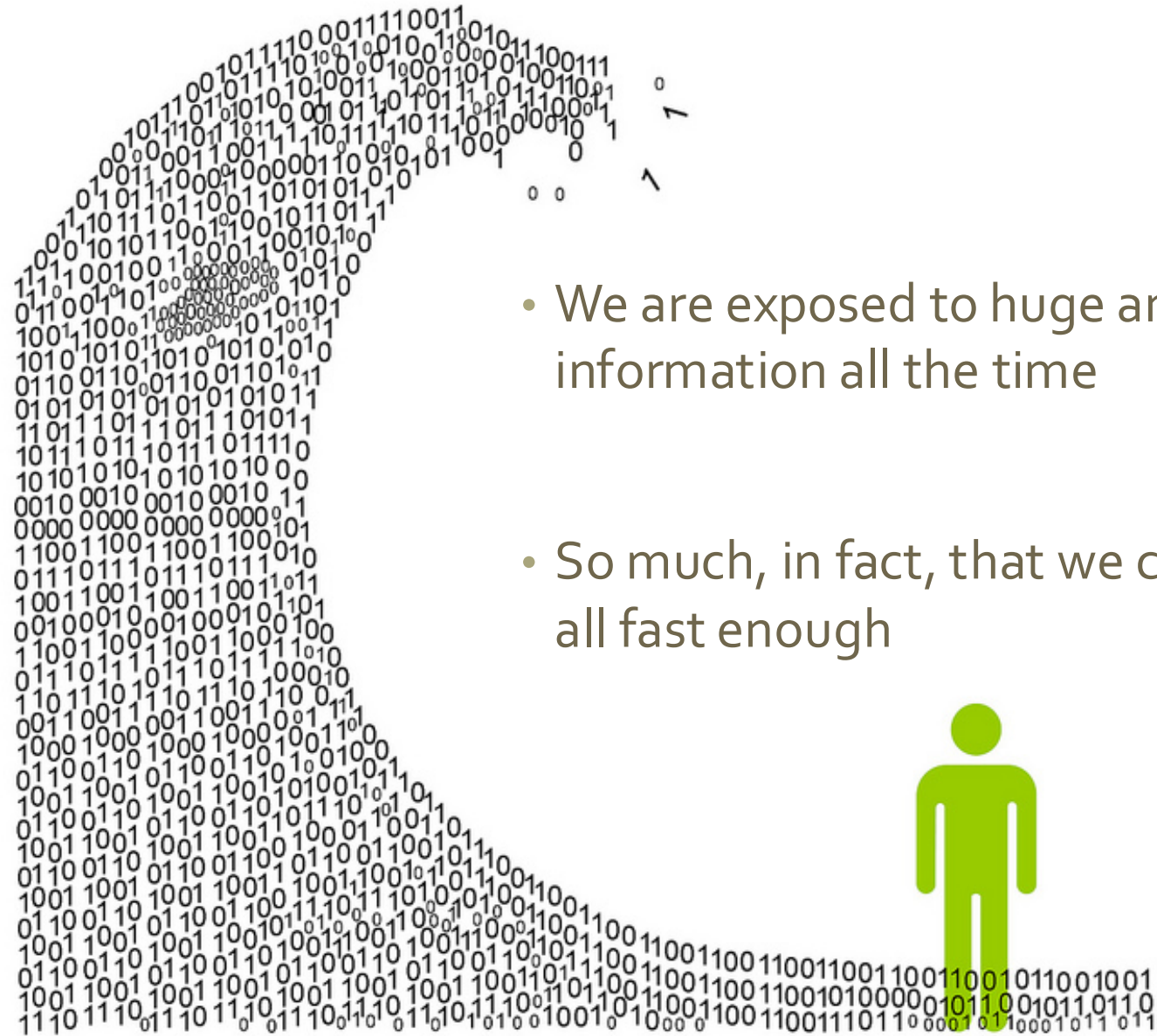
Discussion:
what are they
all trying to do?



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



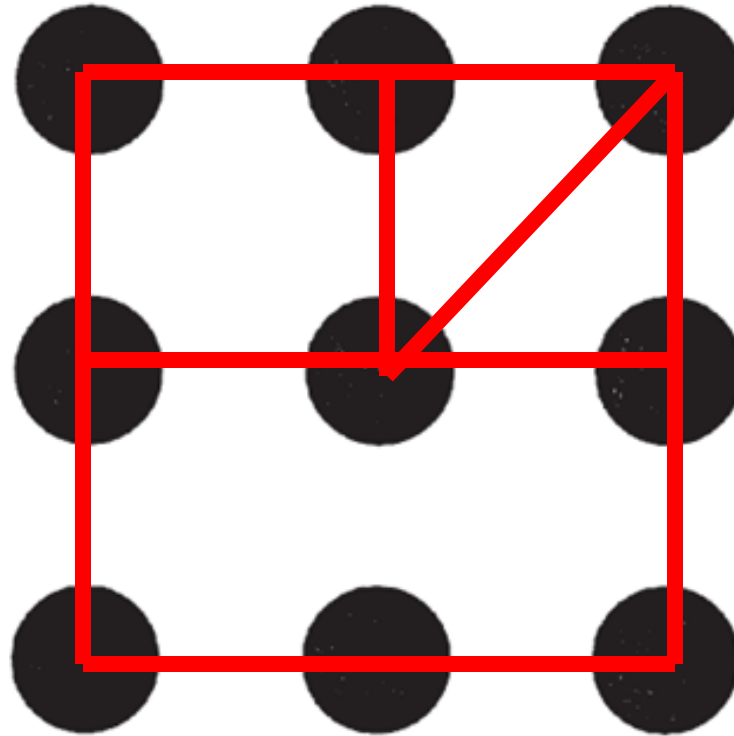
Mental models



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

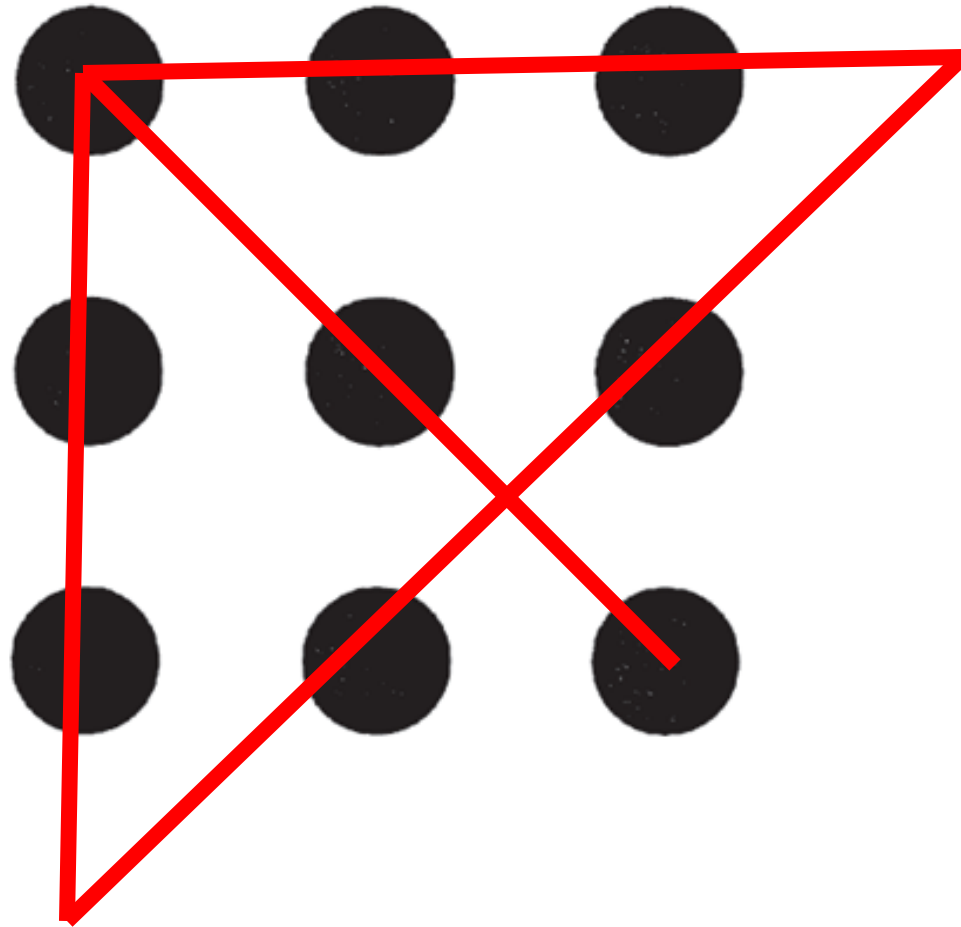
The 9-dot problem

Task 1: Connect all 9 dots using only straight lines



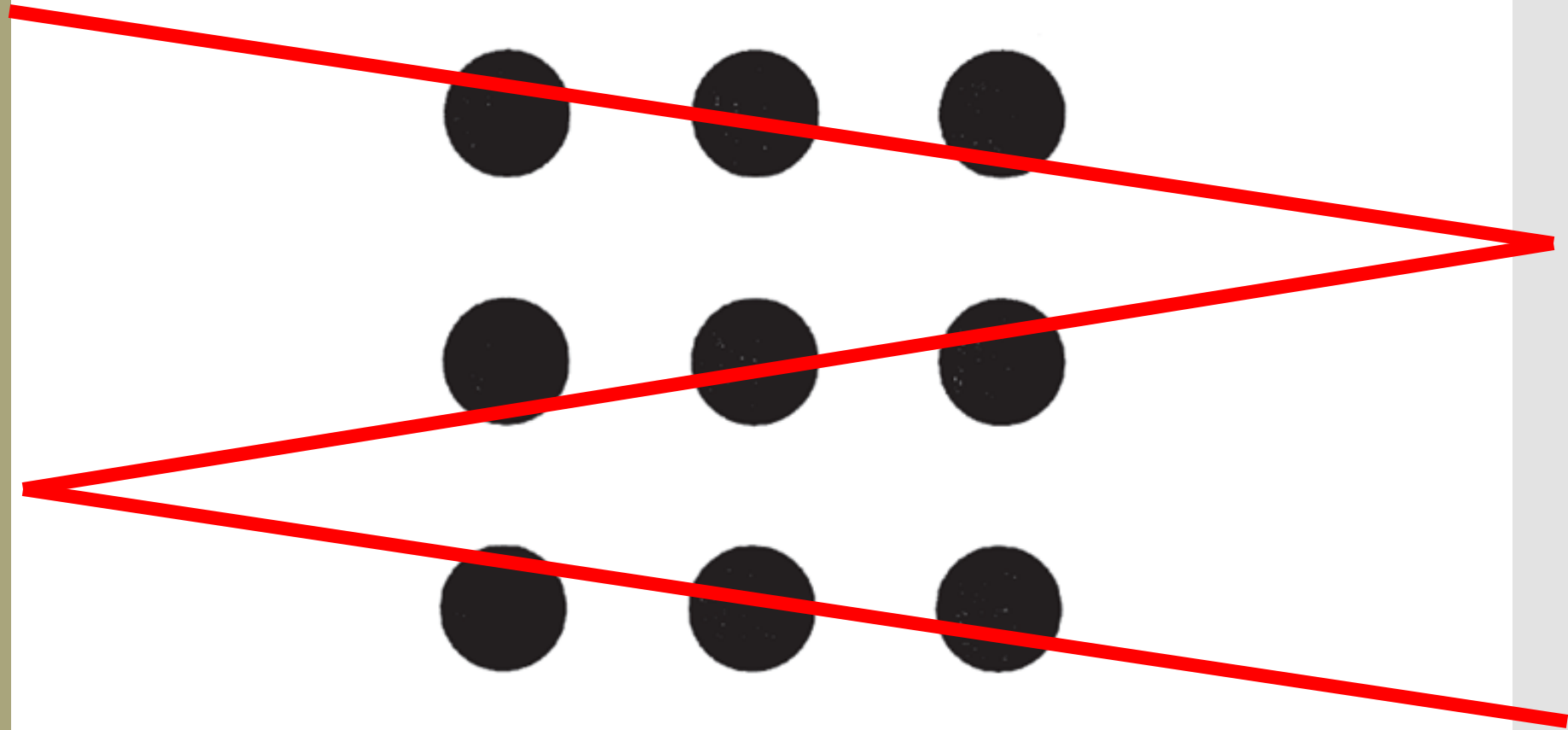
The 9-dot problem

Task 2: Connect all 9 dots using 4 straight lines



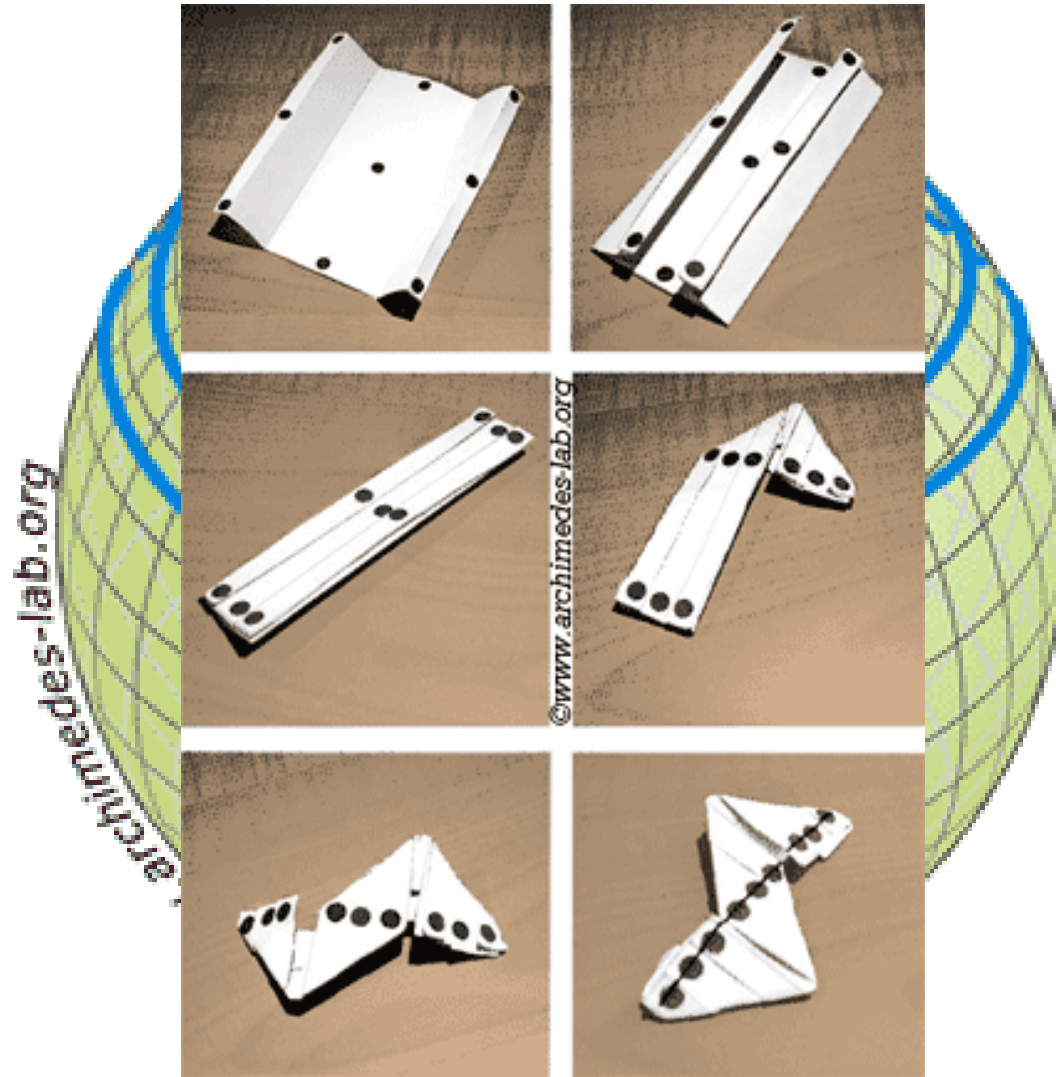
The 9-dot problem

Task 3: Connect all 9 dots using 3 straight lines



The 9-dot problem

Task 4: Connect all 9 dots using 1 straight line



Mental Models: a Sketch



Mental Models

1. We tend to see what we expect to see

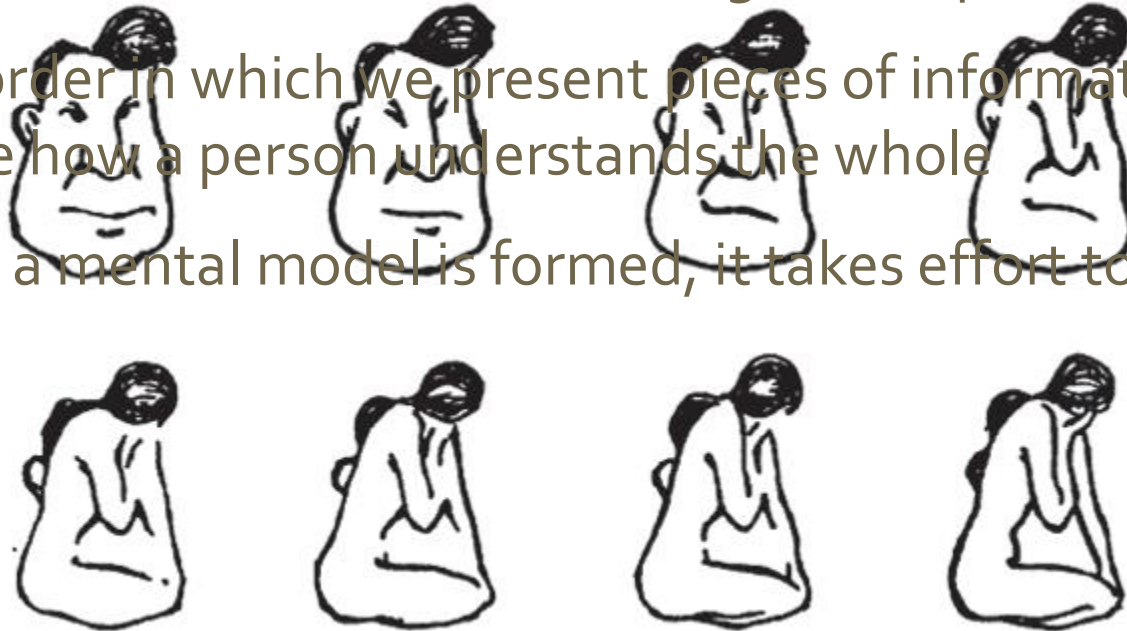
- Mental models are built from prior experience
- We expect new input to “fit” the existing model
- Updates are **expensive**: given input that almost fits, we’ll distort information to avoid re-fitting the model
- **Expectation** is at least as strong as perception



Mental Models

2. Mental models form quickly, & update slowly

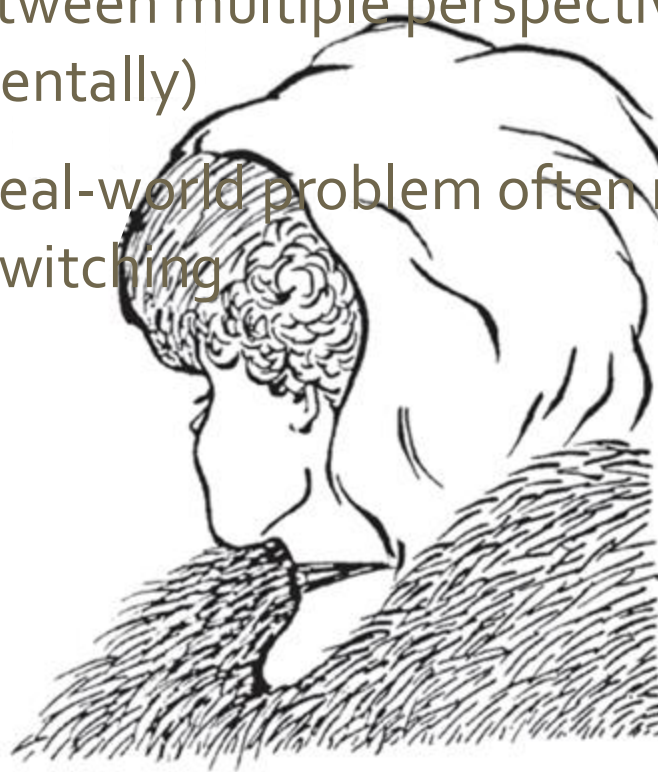
- “First impressions matter”
- Early information can have the highest impact
- The order in which we present pieces of information can shape how a person understands the whole
- Once a mental model is formed, it takes effort to alter it



Mental Models

3. New information gets incorporated into the existing model

- Integrating competing perspectives is challenging
- Switching between multiple perspectives is also difficult (visually or mentally)
- **Tricky part:** real-world problem often require such perspective switching



Mental Models

4. Initial exposure interferes with accurate perception



Blur size

128px
64px
32px
16px
8px
None

Mental Models

4. Initial exposure interferes with accurate perception

- Longer exposure to ambiguous data makes people **more confident** in their initial model
- This is true even if new data presents strong evidence that their model is **wrong**!
- Important: need to be intentional when we design, because incremental information can be **misleading**

Mental Models

The good:

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

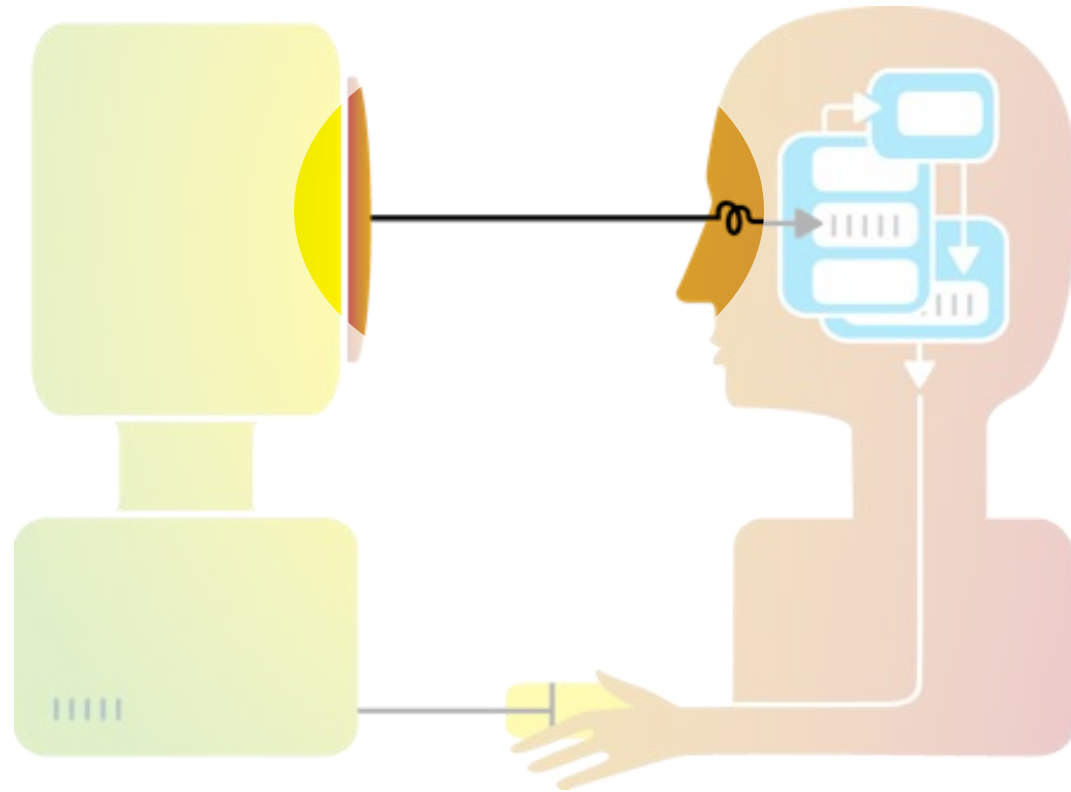
The bad:

- 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?



Data

- Remember...

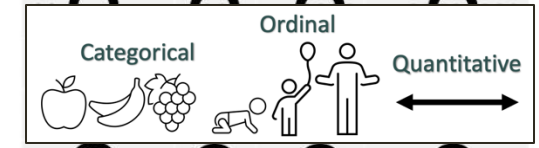
| country | year | cases | population |
|-------------|------|--------|------------|
| Afghanistan | 1999 | 1725 | 19987071 |
| Afghanistan | 2000 | 1666 | 20095360 |
| Brazil | 1999 | 31737 | 17206362 |
| Brazil | 2000 | 80488 | 17404898 |
| China | 1999 | 212258 | 1272915272 |
| China | 2000 | 216766 | 128012583 |

variables

| country | year | cases | population |
|-------------|------|--------|------------|
| Afghanistan | 1999 | 1725 | 19987071 |
| Afghanistan | 2000 | 1666 | 20095360 |
| Brazil | 1999 | 31737 | 17206362 |
| Brazil | 2000 | 80488 | 17404898 |
| China | 1999 | 212258 | 1272915272 |
| China | 2000 | 216766 | 128012583 |

observations

| country | year | cases | population |
|-------------|------|--------|------------|
| Afghanistan | 1999 | 1725 | 19987071 |
| Afghanistan | 2000 | 1666 | 20095360 |
| Brazil | 1999 | 31737 | 17206362 |
| Brazil | 2000 | 80488 | 17404898 |
| China | 1999 | 212258 | 1272915272 |
| China | 2000 | 216766 | 128012583 |



values

Data → Visuals

- Remember...

| country | year | cases | population |
|-------------|------|--------|------------|
| Afghanistan | 1999 | 216745 | 19987071 |
| Afghanistan | 2000 | 21666 | 20095360 |
| Brazil | 1999 | 31737 | 17206362 |
| Brazil | 2000 | 80488 | 17404898 |
| China | 1999 | 212258 | 127215272 |
| China | 2000 | 216766 | 128023583 |

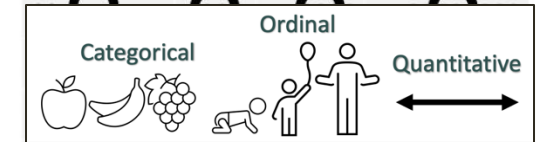
variables

| country | year | cases | population |
|-------------|------|--------|------------|
| Afghanistan | 1999 | 216745 | 19987071 |
| Afghanistan | 2000 | 21666 | 20095360 |
| Brazil | 1999 | 31737 | 17206362 |
| Brazil | 2000 | 80488 | 17404898 |
| China | 1999 | 212258 | 127215272 |
| China | 2000 | 216766 | 128023583 |

observations

| country | year | cases | population |
|-------------|------|--------|------------|
| Afghanistan | 1999 | 216745 | 19987071 |
| Afghanistan | 2000 | 21666 | 20095360 |
| Brazil | 1999 | 31737 | 17206362 |
| Brazil | 2000 | 80488 | 17404898 |
| China | 1999 | 212258 | 127215272 |
| China | 2000 | 216766 | 128023583 |

values



- **Big idea behind visualization**

- Data have dimensions
- Visualizations have dimensions, too
- To build good visualizations, we need to **map data dimensions to visual dimensions** in a principled way

Data → Visuals

- Remember...

| country | year | cases | population |
|-------------|------|--------|------------|
| Afghanistan | 1999 | 216745 | 19987071 |
| Afghanistan | 2000 | 21666 | 20095360 |
| Brazil | 1999 | 31737 | 17206362 |
| Brazil | 2000 | 80488 | 17404898 |
| China | 1999 | 212258 | 127215272 |
| China | 2000 | 216766 | 128023583 |

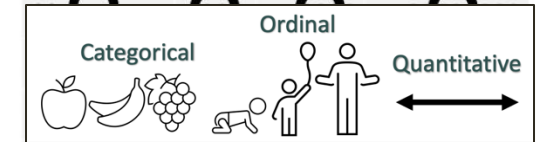
variables

| country | year | cases | population |
|-------------|------|--------|------------|
| Afghanistan | 1999 | 216745 | 19987071 |
| Afghanistan | 2000 | 21666 | 20095360 |
| Brazil | 1999 | 31737 | 17206362 |
| Brazil | 2000 | 80488 | 17404898 |
| China | 1999 | 212258 | 127215272 |
| China | 2000 | 216766 | 128023583 |

observations

| country | year | cases | population |
|-------------|------|--------|------------|
| Afghanistan | 1999 | 216745 | 19987071 |
| Afghanistan | 2000 | 21666 | 20095360 |
| Brazil | 1999 | 31737 | 17206362 |
| Brazil | 2000 | 80488 | 17404898 |
| China | 1999 | 212258 | 127215272 |
| China | 2000 | 216766 | 128023583 |

values



- Big idea behind visualization

- Data have dimensions
- Visualizations have dimensions, too
- To build good visualizations, we need to

map data dimensions to visual dimensions in a principled way

Data → Visuals

Data

| country | year | cases | population |
|-------------|------|--------|------------|
| Afghanistan | 1999 | 1866 | 19987071 |
| Afghanistan | 2000 | 1866 | 20095360 |
| Brazil | 1999 | 30737 | 17206362 |
| Brazil | 2000 | 80488 | 17404898 |
| China | 1999 | 210258 | 1272015272 |
| China | 2000 | 210766 | 128602583 |

variables

| country | year | cases | population |
|-------------|------|--------|------------|
| Afghanistan | 1999 | 1866 | 19987071 |
| Afghanistan | 2000 | 1866 | 20095360 |
| Brazil | 1999 | 30737 | 17206362 |
| Brazil | 2000 | 80488 | 17404898 |
| China | 1999 | 210258 | 1272015272 |
| China | 2000 | 210766 | 128602583 |

observations

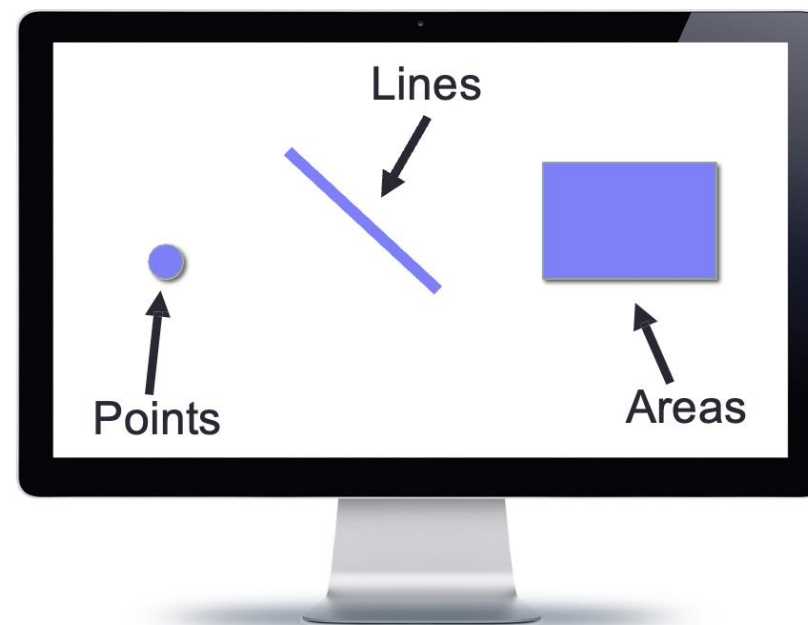
| country | year | cases | population |
|-------------|------|--------|------------|
| Afghanistan | 1999 | 1866 | 19987071 |
| Afghanistan | 2000 | 1866 | 20095360 |
| Brazil | 1999 | 30737 | 17206362 |
| Brazil | 2000 | 80488 | 17404898 |
| China | 1999 | 210258 | 1272015272 |
| China | 2000 | 210766 | 128602583 |

values



Visuals

- **Marks**
- The “ink”



Data → Visuals

Data

| country | year | cases | population |
|-------------|------|--------|------------|
| Afghanistan | 1999 | 1845 | 19987071 |
| Afghanistan | 2000 | 1866 | 20095360 |
| Brazil | 1999 | 30737 | 17206362 |
| Brazil | 2000 | 80488 | 17404898 |
| China | 1999 | 210258 | 1272015272 |
| China | 2000 | 210766 | 128602583 |

variables

| country | year | cases | population |
|-------------|------|--------|------------|
| Afghanistan | 1999 | 1845 | 19987071 |
| Afghanistan | 2000 | 1866 | 20095360 |
| Brazil | 1999 | 30737 | 17206362 |
| Brazil | 2000 | 80488 | 17404898 |
| China | 1999 | 210258 | 1272015272 |
| China | 2000 | 210766 | 128602583 |

observations

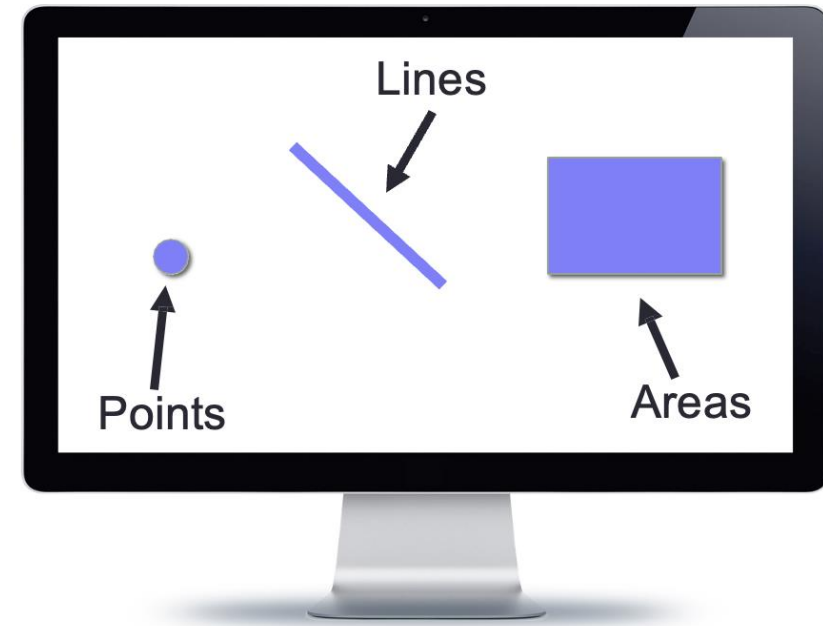
| country | year | cases | population |
|-------------|------|--------|------------|
| Afghanistan | 1999 | 1845 | 19987071 |
| Afghanistan | 2000 | 1866 | 20095360 |
| Brazil | 1999 | 30737 | 17206362 |
| Brazil | 2000 | 80488 | 17404898 |
| China | 1999 | 210258 | 1272015272 |
| China | 2000 | 210766 | 128602583 |

values



Visuals

- **Marks**
 - The “ink”

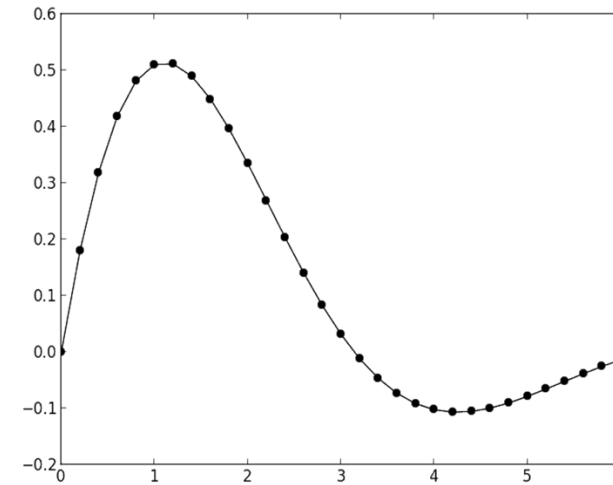
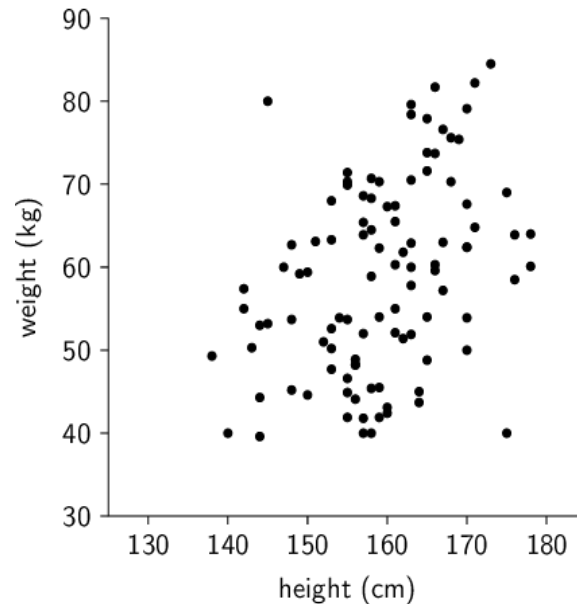


- **Channels or dimensions**
 - *How the marks show up on the page*

Visual Channels / Dimensions

Position

- Encode information using *where* mark is drawn
- Ex.



along a common scale



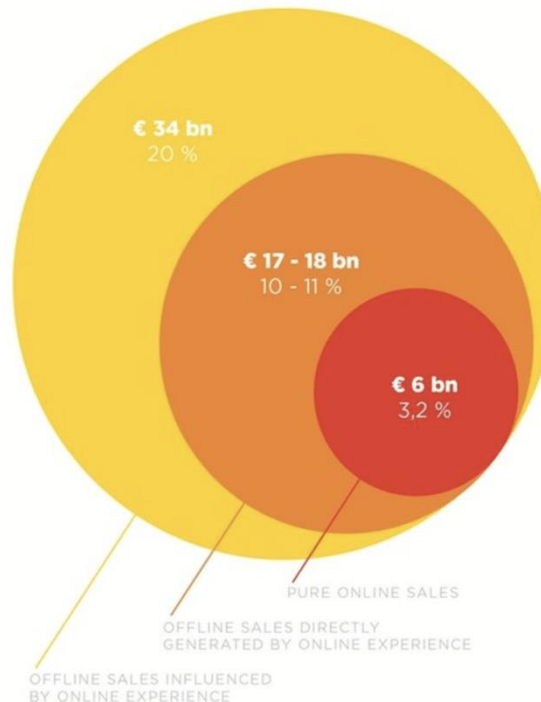
spatial region

Visual Channels / Dimensions

Size / Area

- Encode information using *how big* mark is drawn
- Ex. **HOW DIGITAL MARKETING INFLUENCES GLOBAL LUXURY SALES**

2011 €bn and
% of market



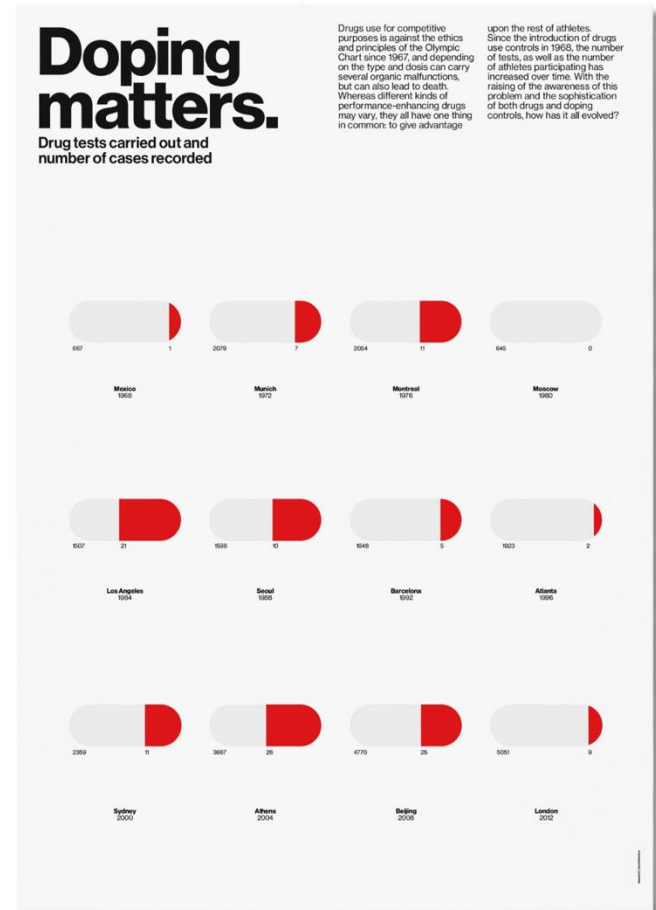
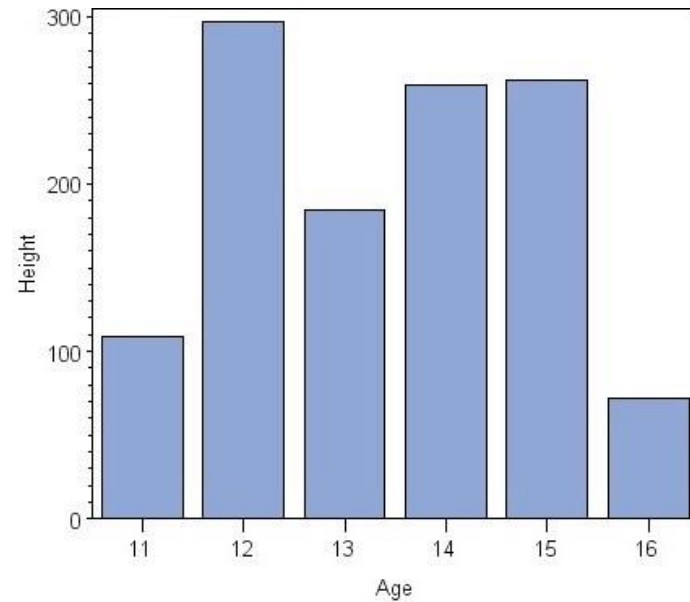
"The Force Awakens" could produce **\$9.6 billion** in revenue from worldwide ticket sales, merchandise, and home entertainment in roughly the first year of release.

| | | |
|-----------------------|--|-----------------------------------|
| \$5.0B Merchandise | \$500M Disney's Infinity 3.0 Videogame | \$214M TV licensing (intl.) |
| | \$780M Star Wars Battlefront Videogame | \$235M TV licensing (domestic) |
| | | \$458M DVDs/downloads |
| | \$1.65B Box office sales (international) | |

Visual Channels / Dimensions

Length

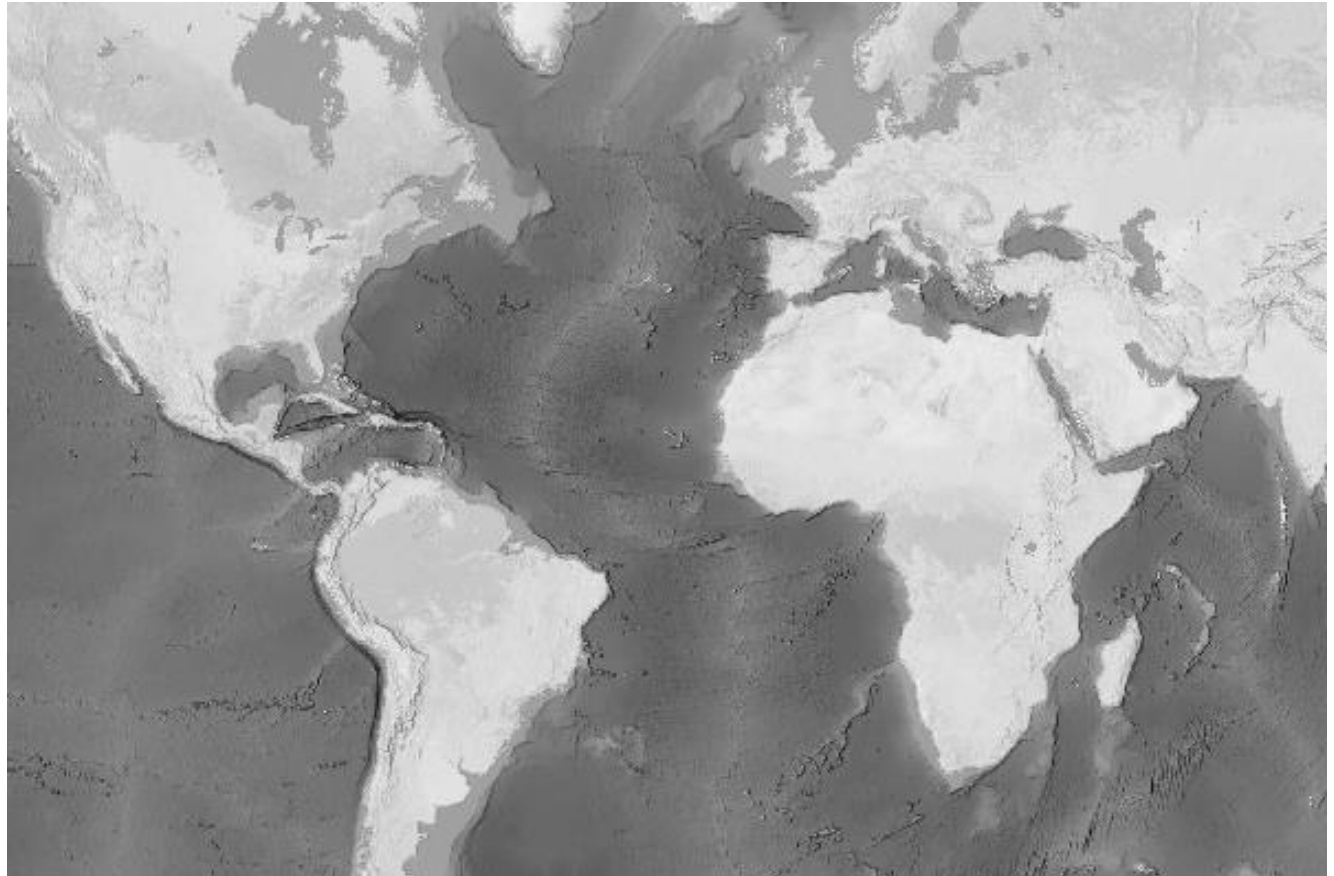
- Encode information using *how long* mark is drawn
- Ex.



Visual Channels / Dimensions

Color: Luminance

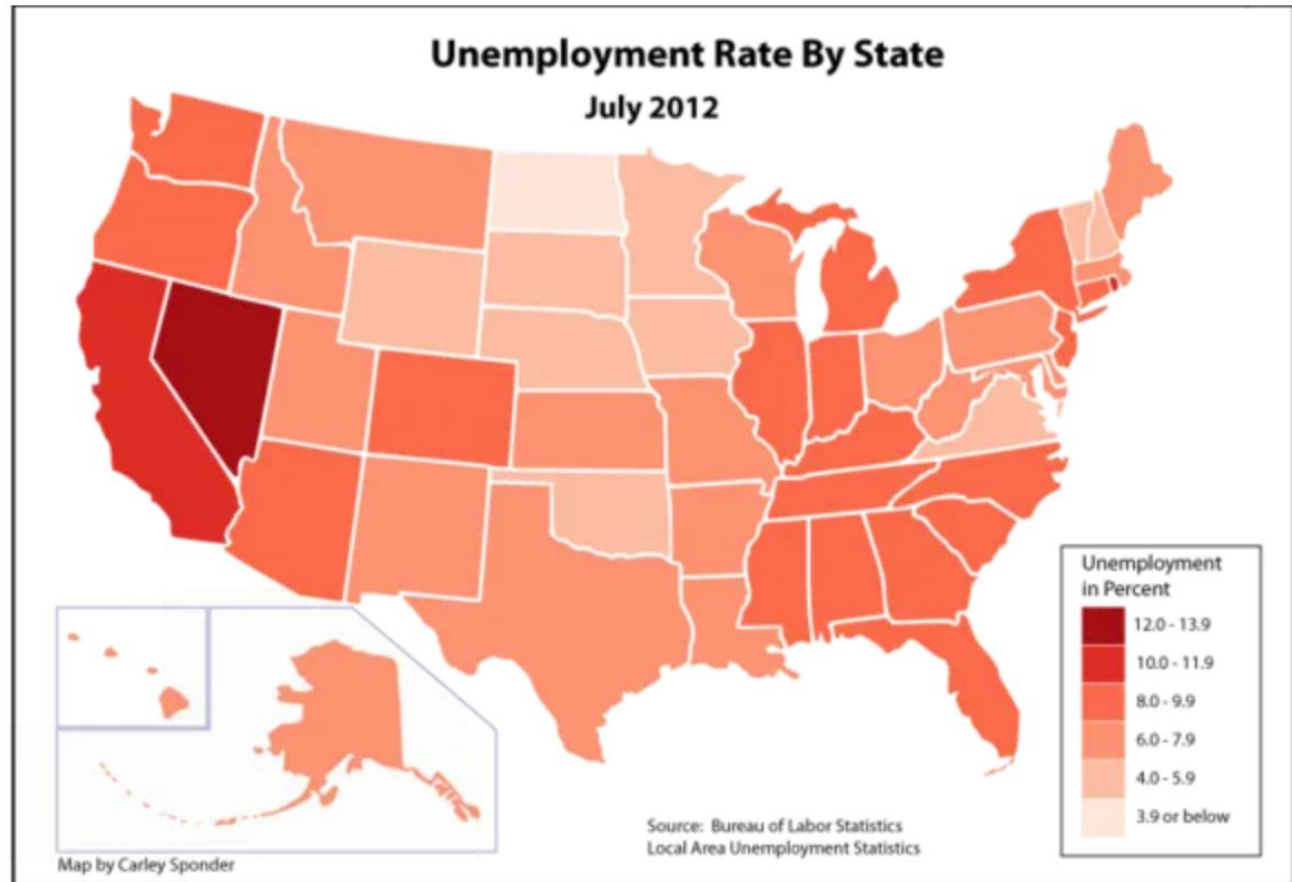
- Encode information using *how dark* mark is drawn
- Ex.



Visual Channels / Dimensions

Color: Saturation

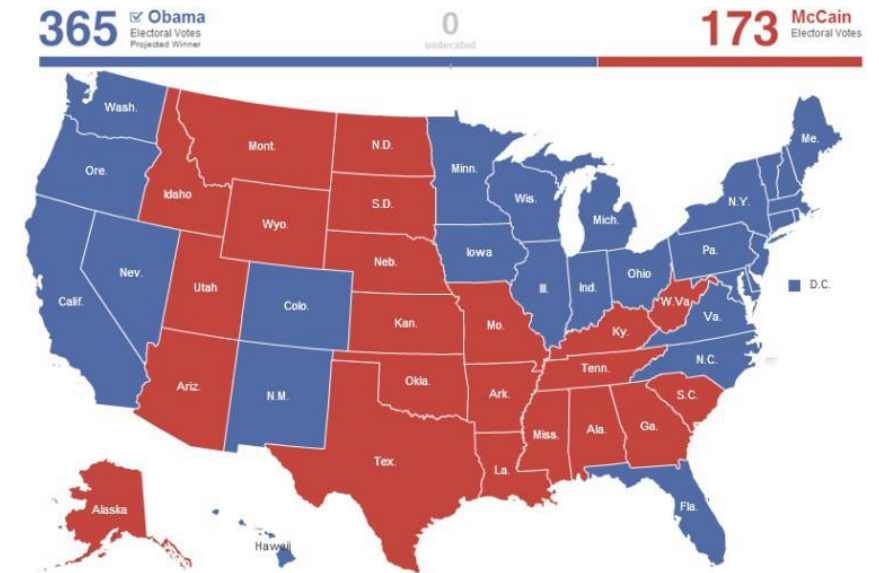
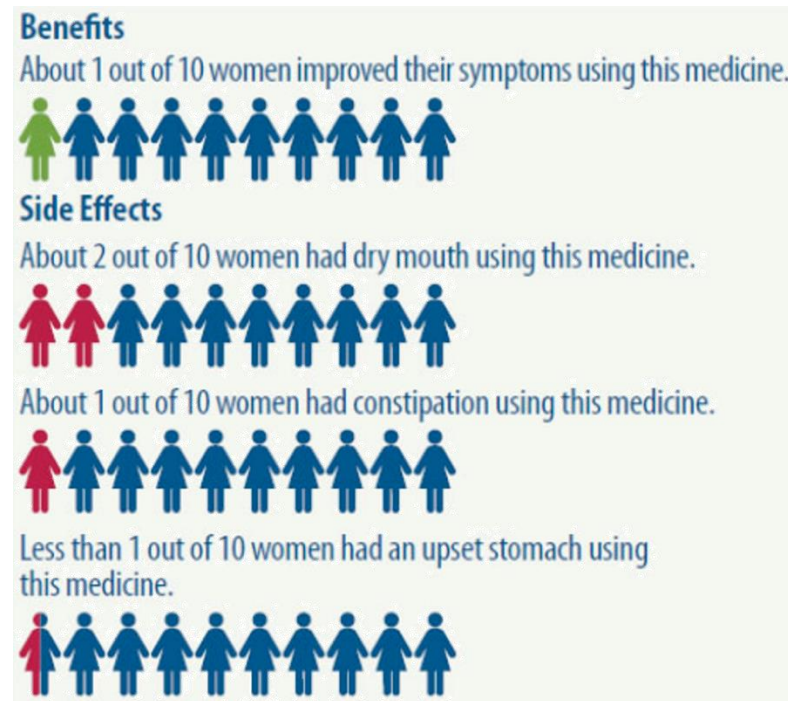
- Encode information using *how much color* mark has
- Ex.



Visual Channels / Dimensions

Color: Hue

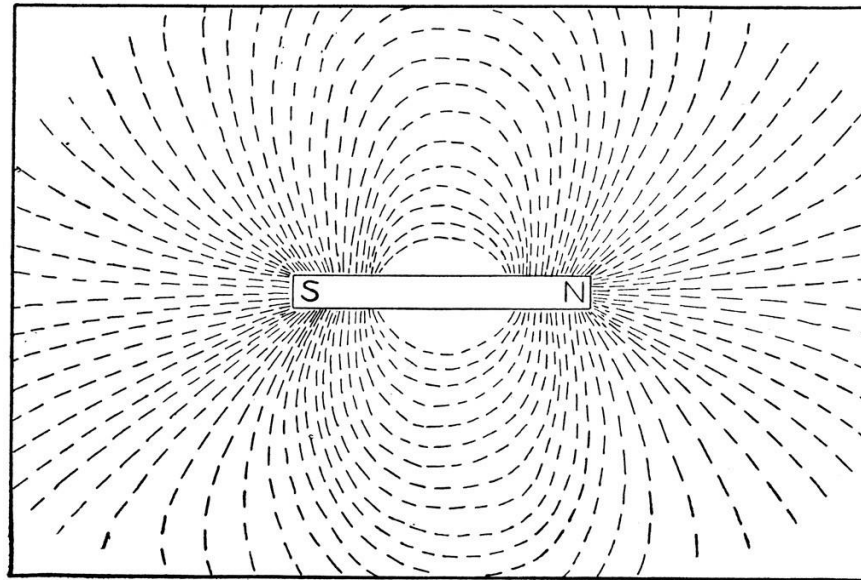
- Encode information using *hue* of mark
- Ex.



Visual Channels / Dimensions

Orientation / Tilt / Angle

- Encode information using how mark is *rotated*
- Ex.



Visual Channels / Dimensions

Shape

- Encode information using how mark is *shaped*
- Ex.

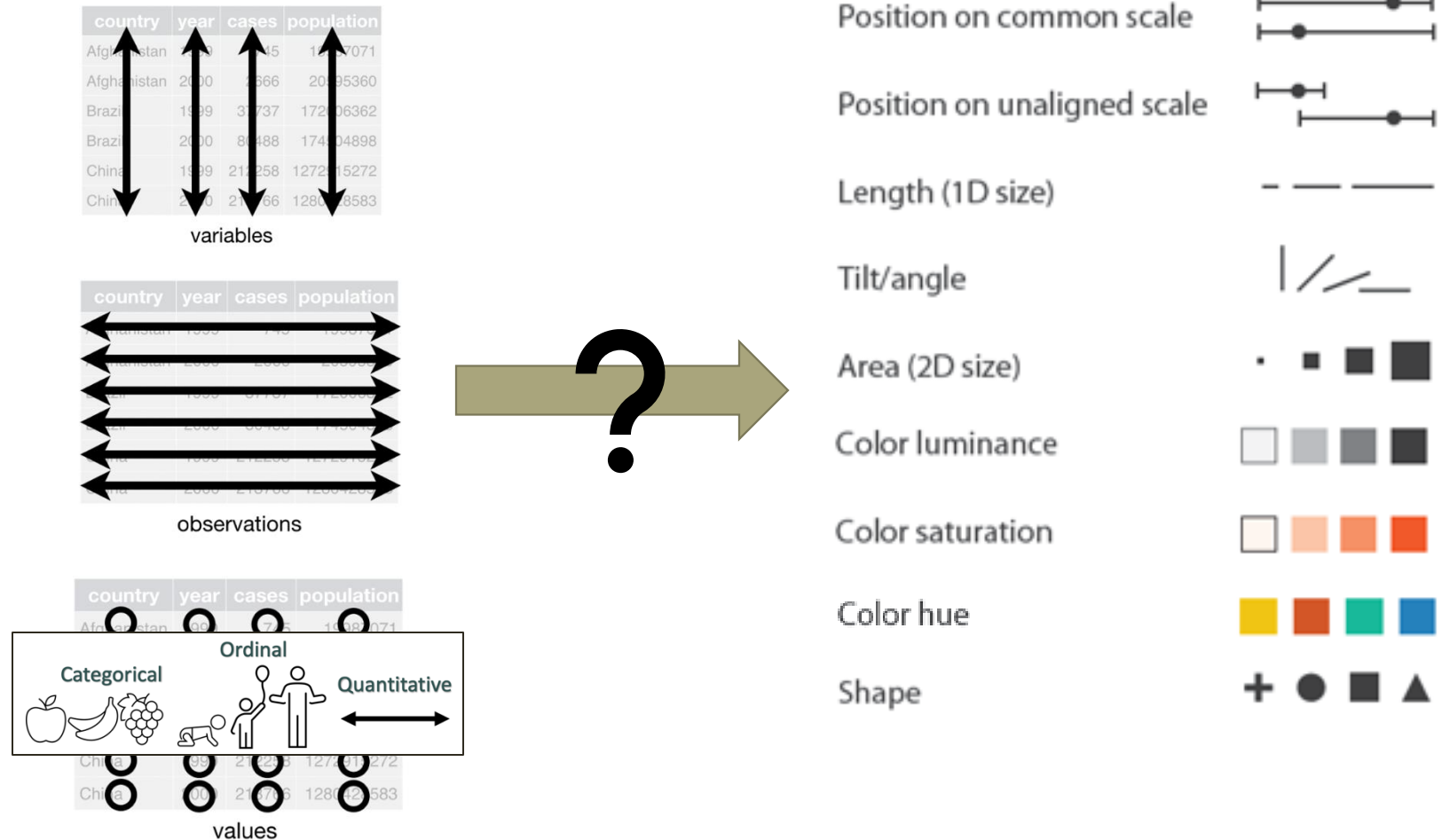


Data → Visuals

- Remember... **Big idea behind visualization**
 - Map data dimensions to visual dimensions in a principled way

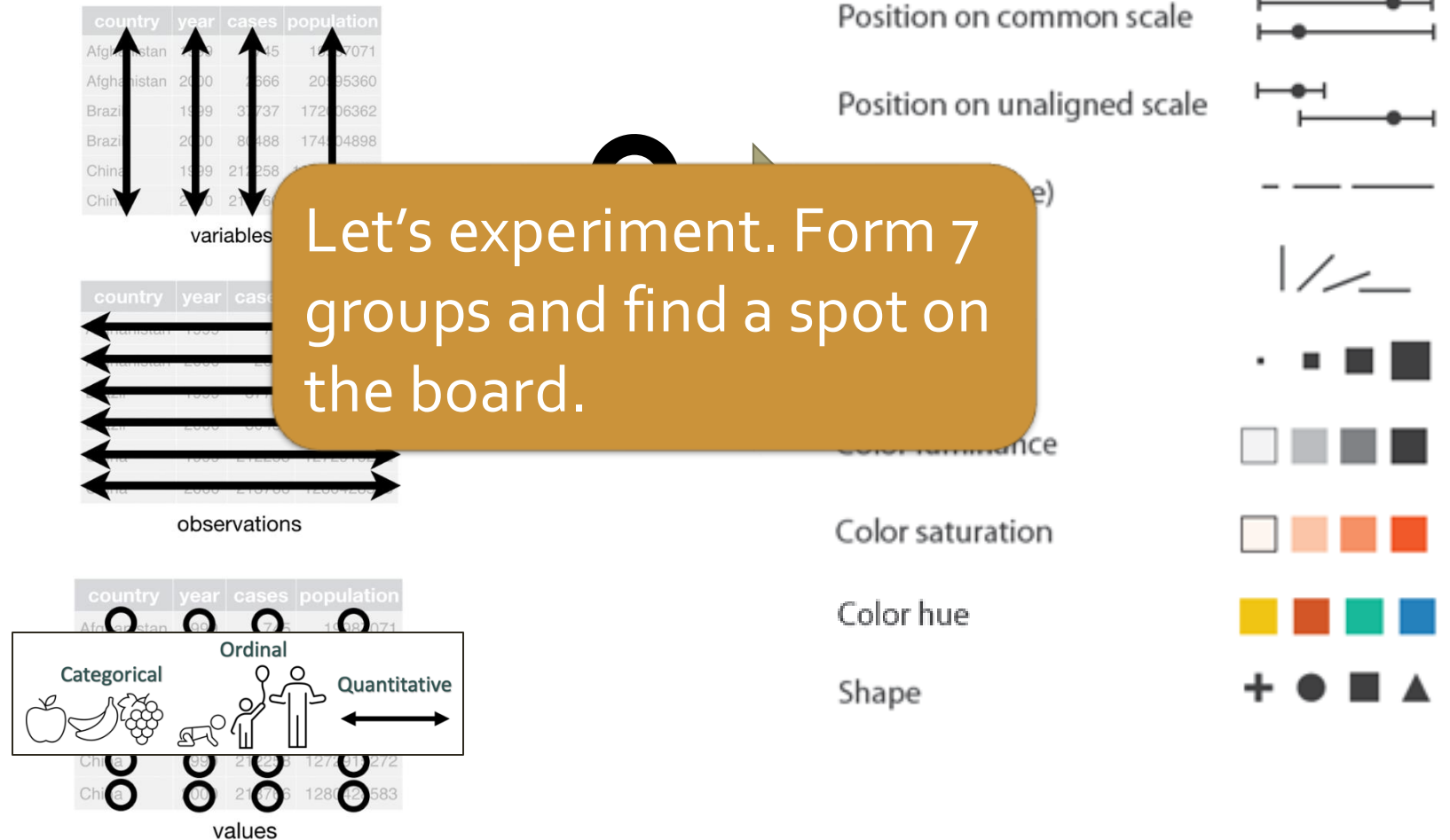
Data → Visuals

- Remember... **Big idea behind visualization**
 - Map data dimensions to visual dimensions in a principled way



Data → Visuals

- Remember... **Big idea behind visualization**
 - Map data dimensions to visual dimensions in a principled way












Work with your group to represent each observation in this dataset as a point (mark) styled using the visual channel you were assigned.

Be prepared to share with the class.

Data → Visuals

| Name |
|-----------|
| blueberry |
| asparagus |
| pumpkin |










| | |
|-----------------------------|---|
| Position on common scale |  |
| Position on unaligned scale |  |
| Length (1D size) |  |
| Tilt/angle |  |
| Area (2D size) |  |
| Color luminance |  |
| Color saturation |  |
| Color hue |  |
| Shape |  |

Work with your group to represent each observation in this dataset as a point (mark) styled using the visual channel you were assigned.

Be prepared to share with the class.

Data → Visuals

| Season |
|--------|
| summer |
| spring |
| fall |








| | |
|-----------------------------|---|
| Position on common scale |  |
| Position on unaligned scale |  |
| Length (1D size) |  |
| Tilt/angle |  |
| Area (2D size) |  |
| Color luminance |  |
| Color saturation |  |
| Color hue |  |
| Shape |  |

Work with your group to represent each observation in this dataset as a point (mark) styled using the visual channel you were assigned.

Be prepared to share with the class.

Data → Visuals










| Price (per lb) |
|----------------|
| 6.45 |
| 4.99 |
| 0.24 |

| | |
|-----------------------------|---|
| Position on common scale |  |
| Position on unaligned scale |  |
| Length (1D size) |  |
| Tilt/angle |  |
| Area (2D size) |  |
| Color luminance |  |
| Color saturation |  |
| Color hue |  |
| Shape |  |

Data → Visuals

What type of variable is Name?








| Name |
|-----------|
| blueberry |
| asparagus |
| pumpkin |

| | |
|-----------------------------|---|
| Position on common scale |  |
| Position on unaligned scale |  |
| Length (1D size) |  |
| Tilt/angle |  |
| Area (2D size) |  |
| Color luminance |  |
| Color saturation |  |
| Color hue |  |
| Shape |  |

Data → Visuals

What type of variable is Season?







| Season |
|--------|
| summer |
| spring |
| fall |

| | |
|-----------------------------|---|
| Position on common scale |  |
| Position on unaligned scale |  |
| Length (1D size) |  |
| Tilt/angle |  |
| Area (2D size) |  |
| Color luminance |  |
| Color saturation |  |
| Color hue |  |
| Shape |  |

Data → Visuals


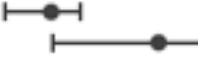







What type of variable is Price?




| Price (per lb) |
|----------------|
| 6.45 |
| 4.99 |
| 0.24 |

| | |
|-----------------------------|---|
| Position on common scale |  |
| Position on unaligned scale |  |
| Length (1D size) |  |
| Tilt/angle |  |
| Area (2D size) |  |
| Color luminance |  |
| Color saturation |  |
| Color hue |  |
| Shape |  |

Data → Visuals

- Remember... **Big idea behind visualization**
 - Map data dimensions to visual dimensions in a principled way
 - Insight 1: Not all visual dimensions can represent all data types

| | |
|-----------------------------|---|
| Position on common scale |  |
| Position on unaligned scale |  |
| Length (1D size) |  |
| Tilt/angle |  |
| Area (2D size) |  |
| Color luminance |  |
| Color saturation |  |
| Color hue |  |
| Shape |  |


| Categorical | Ordinal | Quantitative |
|---|---|---|
|  |  |  |
| ✓ | ✓ | ✓ |
| ✓ | ✓ | ✓ |
| | ✓ | ✓ |
| | ✓ | ✓ |
| | ✓ | ✓ |
| | ✓ | ✓ |
| ✓ | | |
| ✓ | | |

Data → Visuals

- Remember... **Big idea behind visualization**
 - Map data dimensions to visual dimensions in a principled way
 - Insight 1: Not all visual dimensions can represent all data types

➔ **Magnitude** Channels: **Ordered** Attributes

Position on common scale 

Position on unaligned scale 

Length (1D size) 

Tilt/angle 

Area (2D size) 

Depth (3D position) 

Color luminance 

Color saturation 

Curvature 

Volume (3D size) 

➔ **Identity** Channels: **Categorical** Attributes

Spatial region 

Color hue 

Motion 

Shape 

Try it out!

- Work with 2 other people. Be prepared to share your work with the class.
- Find a data visualization you think is interesting
 - Some ideas for where to look: New sites, government sites, Tableau Viz Gallery, massvis.mit.edu
- Identify the following:
 - What is the data that's being visualized?
 - Is the data source included?
 - What marks and channels are used in the visualization?
 - What data do each mark and channel represent?
 - What do you learn from looking at the visualization?
 - What additional questions do you have?