Communicating with Data – Welcome!

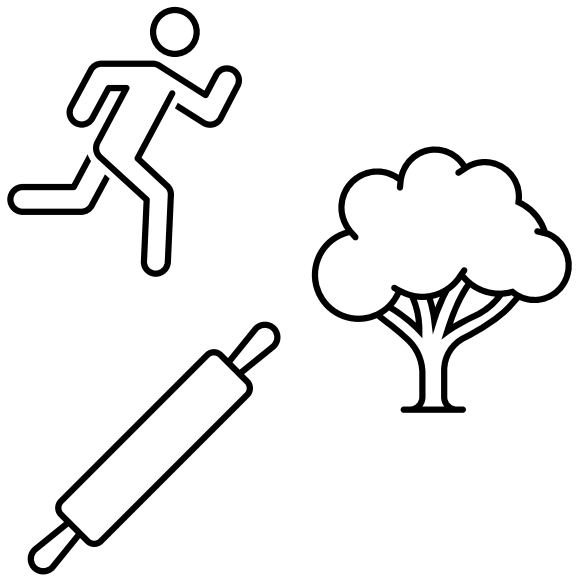
Dr. Ab Mosca (they/them)

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

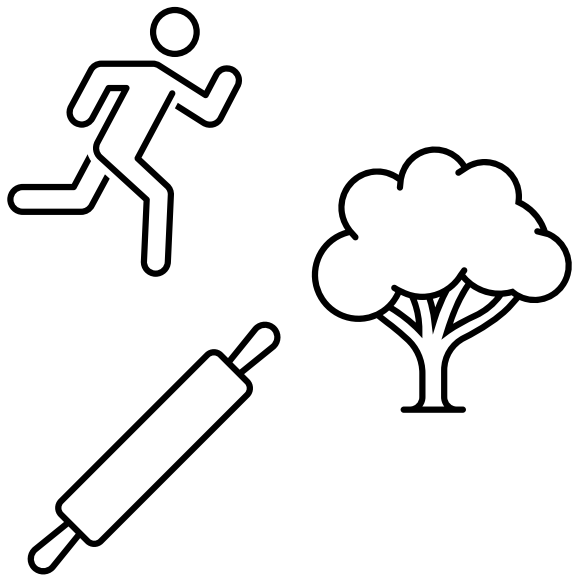
Plan for Today

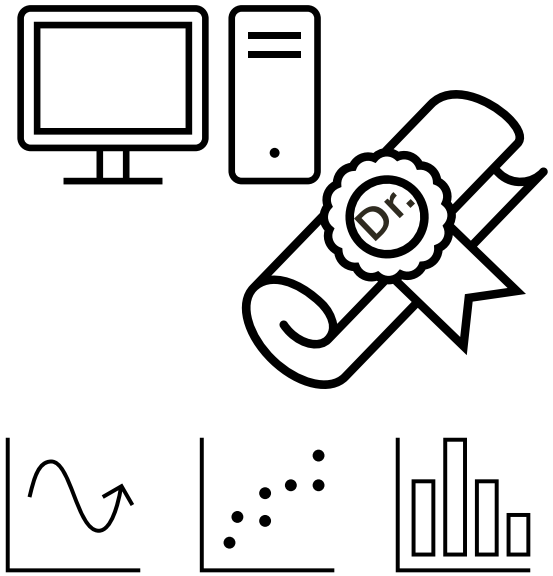
- Intros
- About this course
- How is visualization useful?
- What is data?
- Structure of this course

Who Am I?

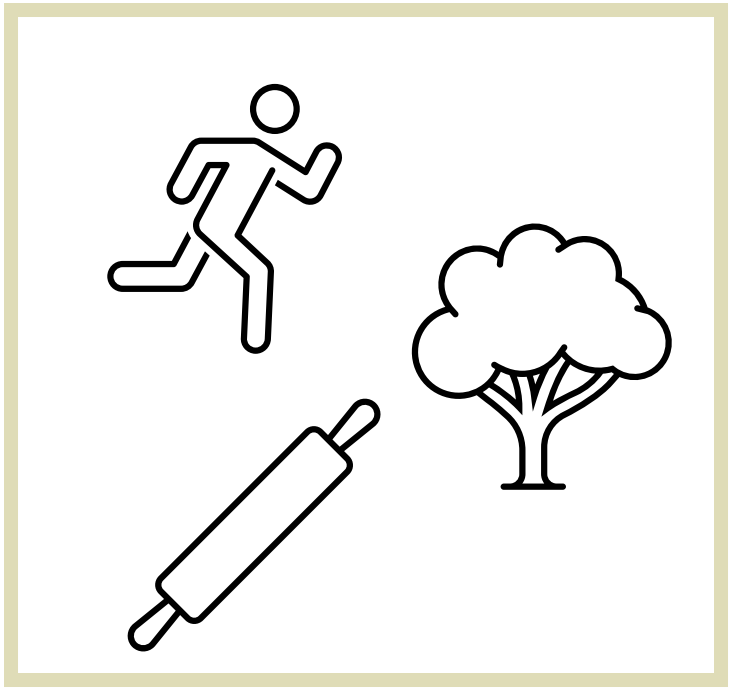


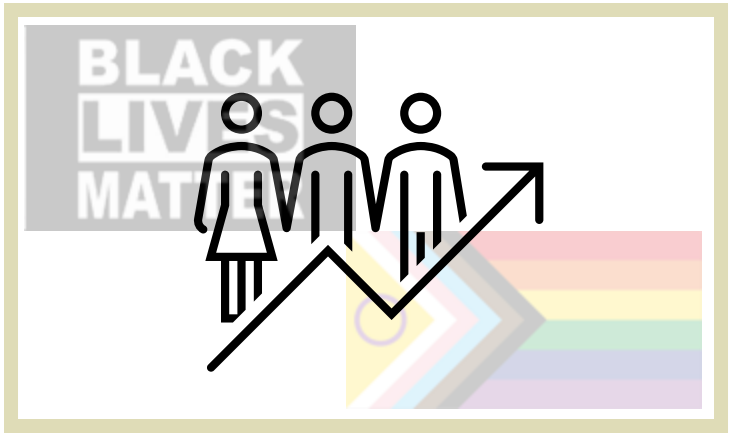
Who Am I?

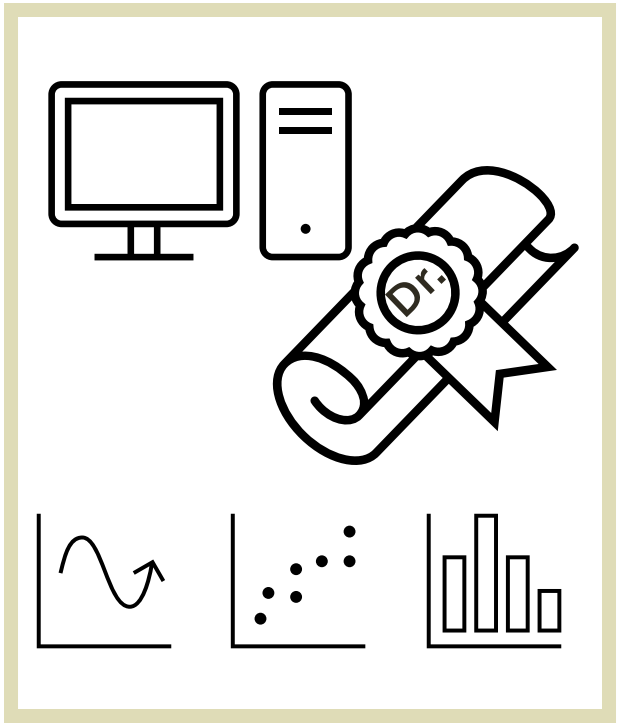




Who Am I?







Who Are You?

- Form groups of 3
- Introduce yourselves (name, pronouns)
- Share:
  - A highlight of your summer break
- Find 1 thing that your entire group has in common (favorite color? hometown? left-handed? Be creative!)
- After about 5 minutes we will go around, introduce ourselves, and share what each group has in common

Who Are You?

- Form **new groups** of 3 (move around!)
- Introduce yourselves (name, pronouns)
- Share:
  - Would you rather live in an estate that has anything you want but you can never leave OR live in a camper van that must move >100 miles every day?
- After about 5 minutes we will go around, introduce ourselves, and share our would you rather answers

Who Are You?

- Form **new new groups** of 3 (move around!)
- Introduce yourselves (name, pronouns)
- Debate:
  - Is a hot dog a sandwich?
- After about 5 minutes we will go around, introduce ourselves, and share our would you rather answers

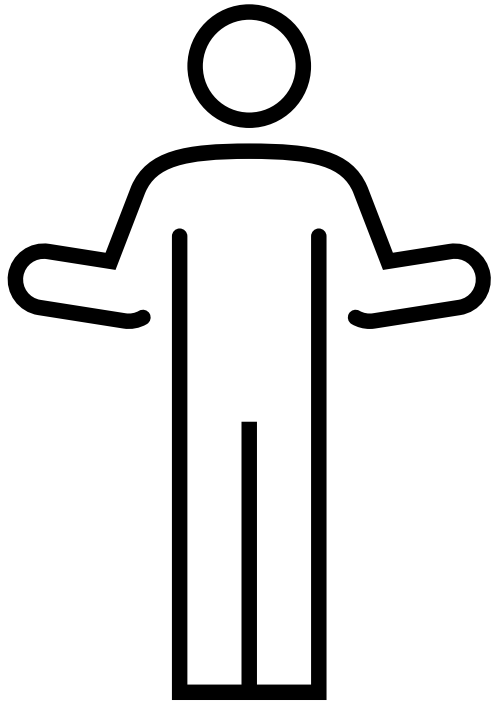
- Name tags!
  - Please make a name tag if you have not already
  - **Write BIG** so I can see your name
  - **Bring your name tag to class with you**

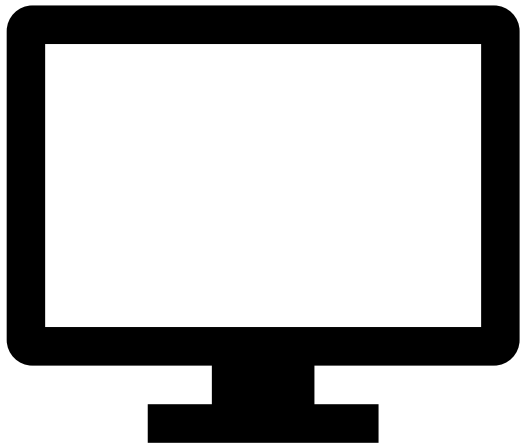
About this course

Communicating with data

About this course

Communicating with data

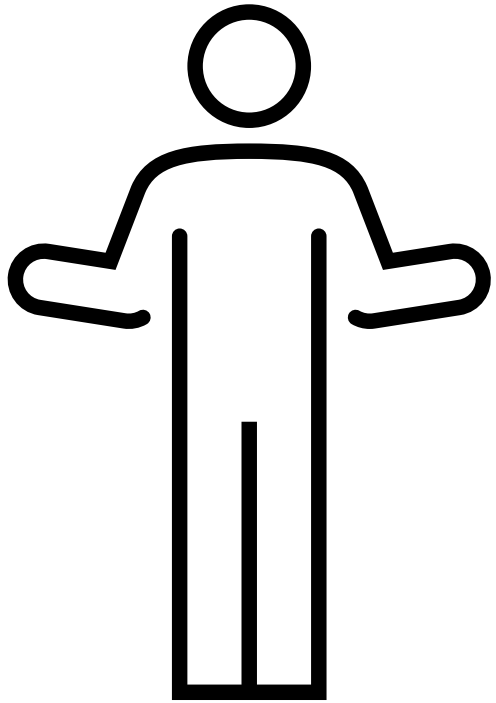


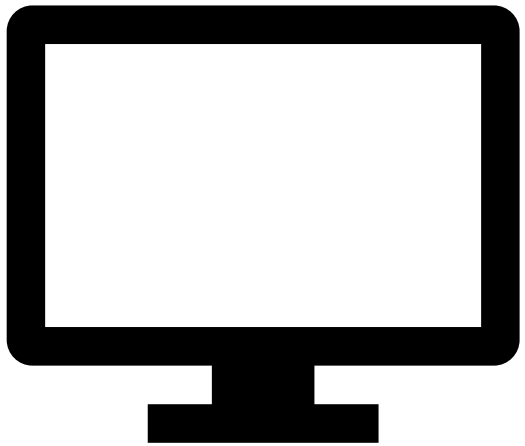


About this course

Communicating with data

What are the strengths of each wrt data and communication?

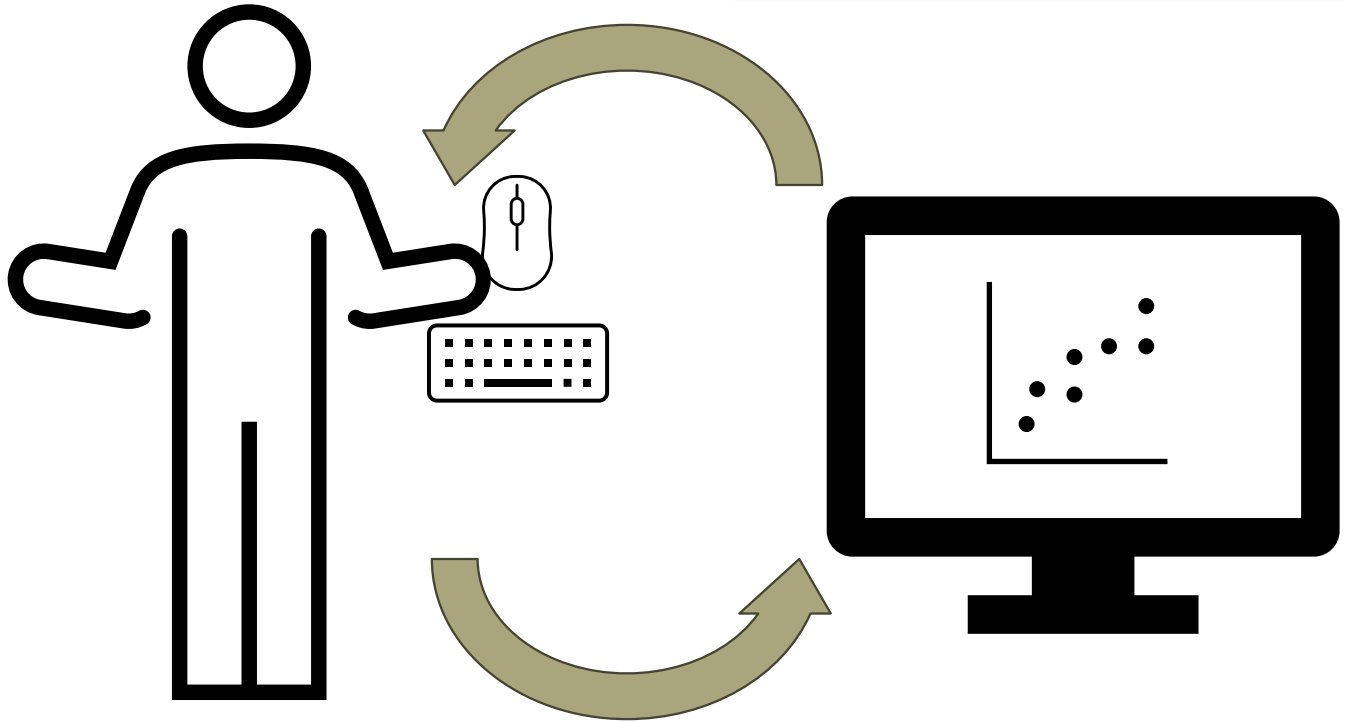




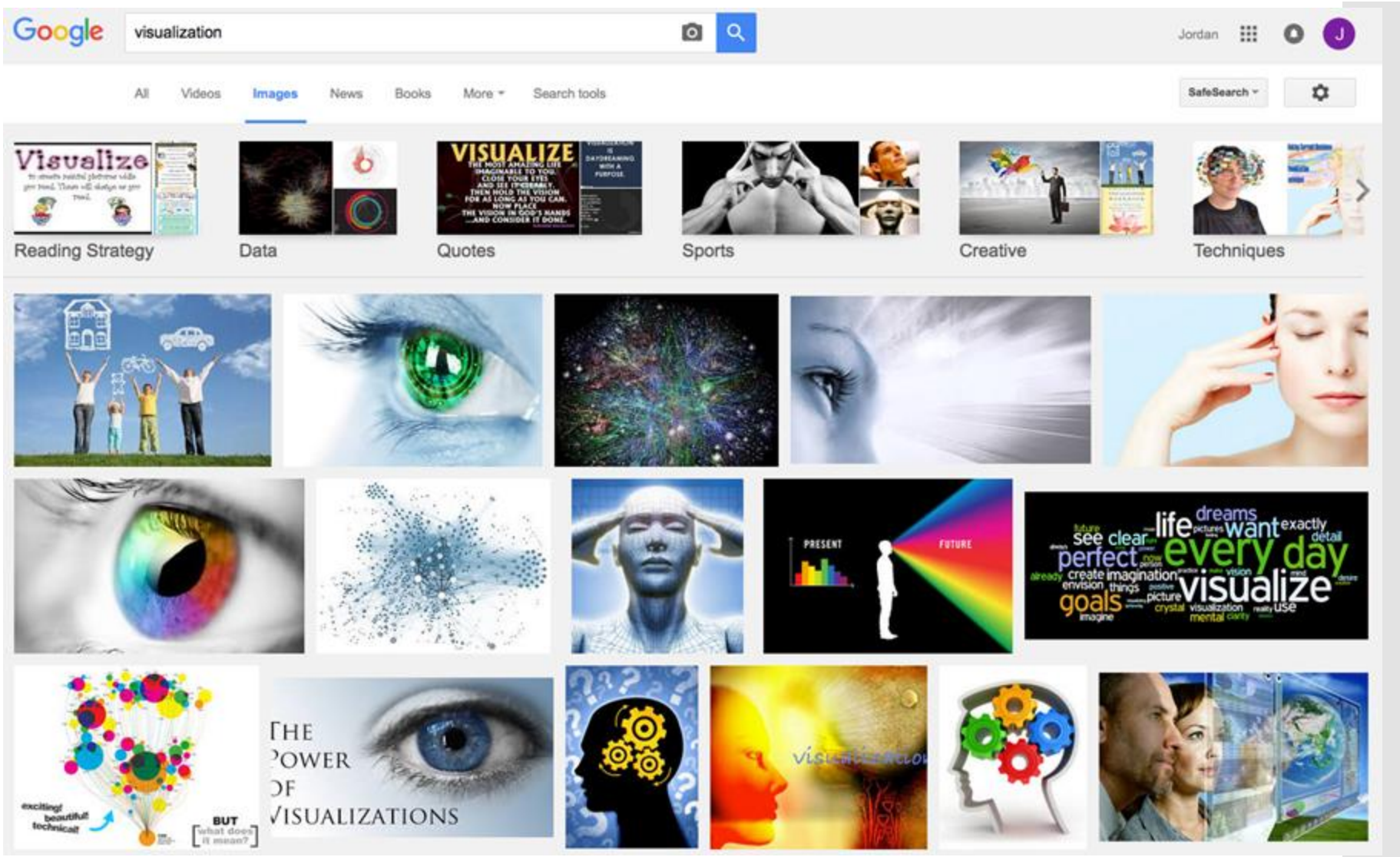
About this course

Communicating with data

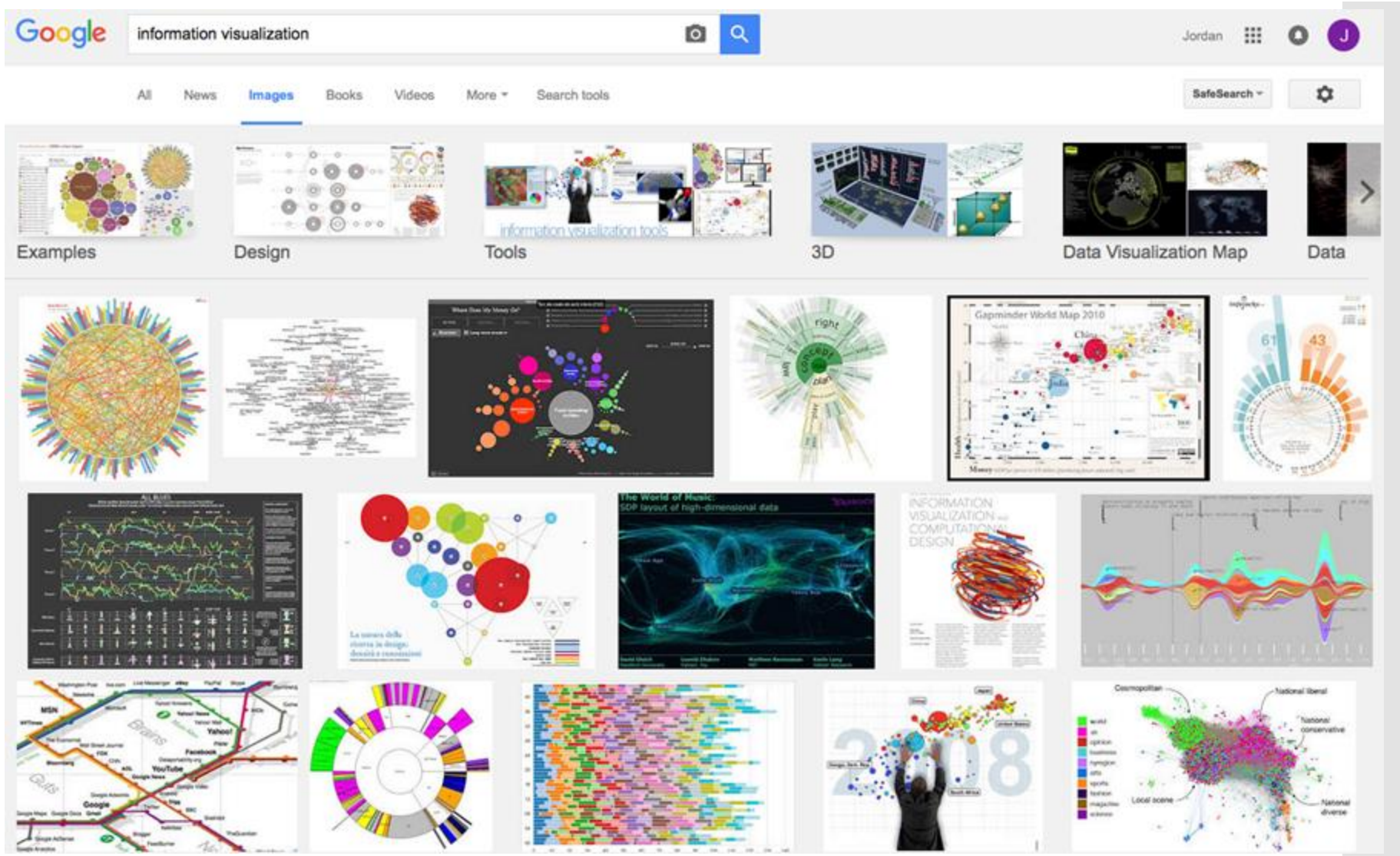
What are the strengths of each wrt data and communication?



What is visualization?



What is visualization?



Perhaps a more helpful question:

What are some ways
a “visualization” can be **useful**?

Does it help you spot trends?

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

Does it help you explore?

Lumos: Increasing Awareness of Analytic Behavior during Visual Data Analysis

Arpit Narechania, Adam Coscia, Emily Wall, Alex Endert

https://lumos-webapp-4aeadb3bf30d.herokuapp.com/

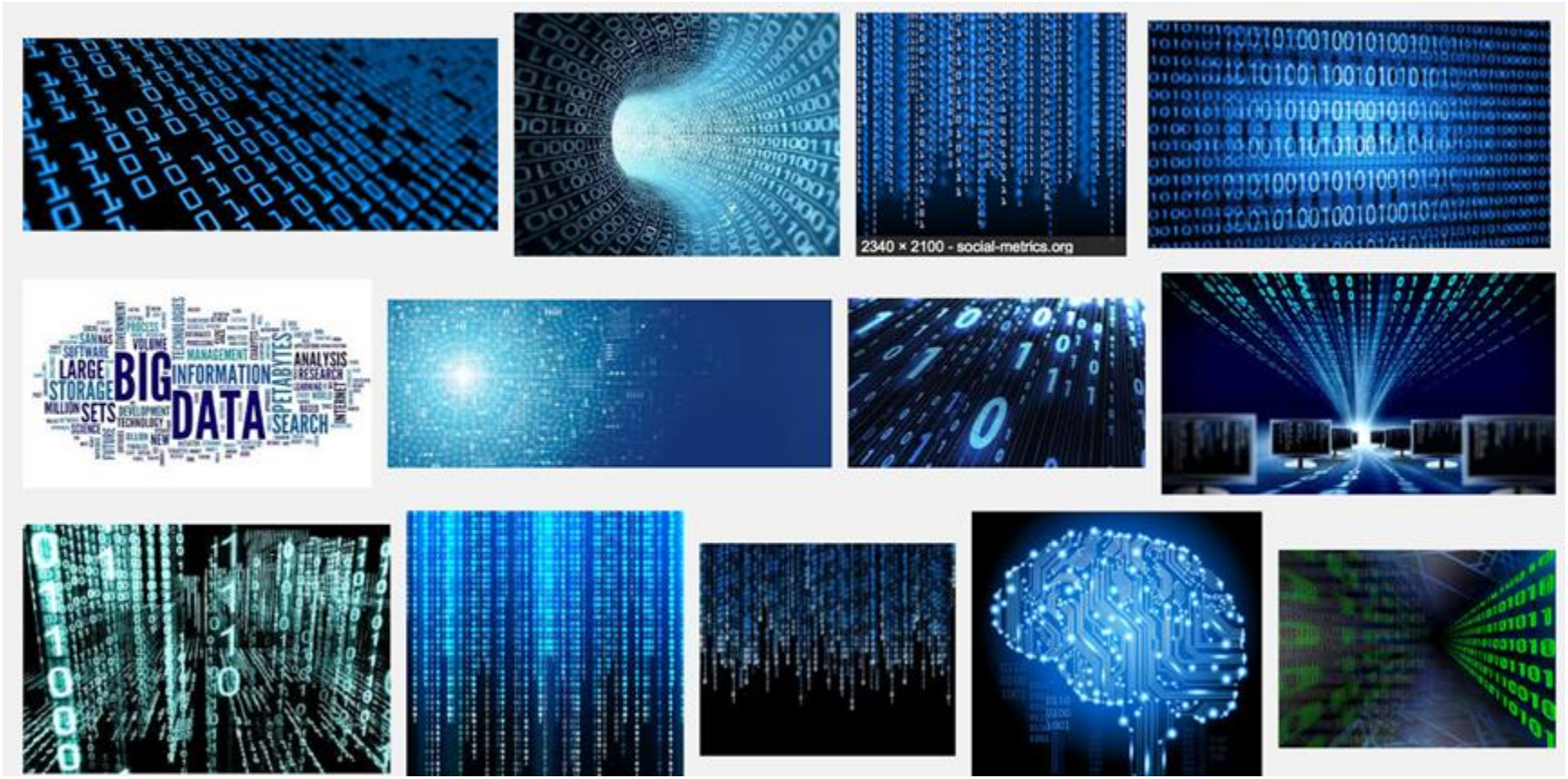
Does it tell a story?

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**

Wait… what is “data”?



Data: a definition

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

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

Data: a definition

A dataset also contains a set of *observations* (also called *records)* over these variables. For example:

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

Data: a definition

A dataset also contains a set of *observations* (also called *records)* over these variables. For example:

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

One way to think about this:

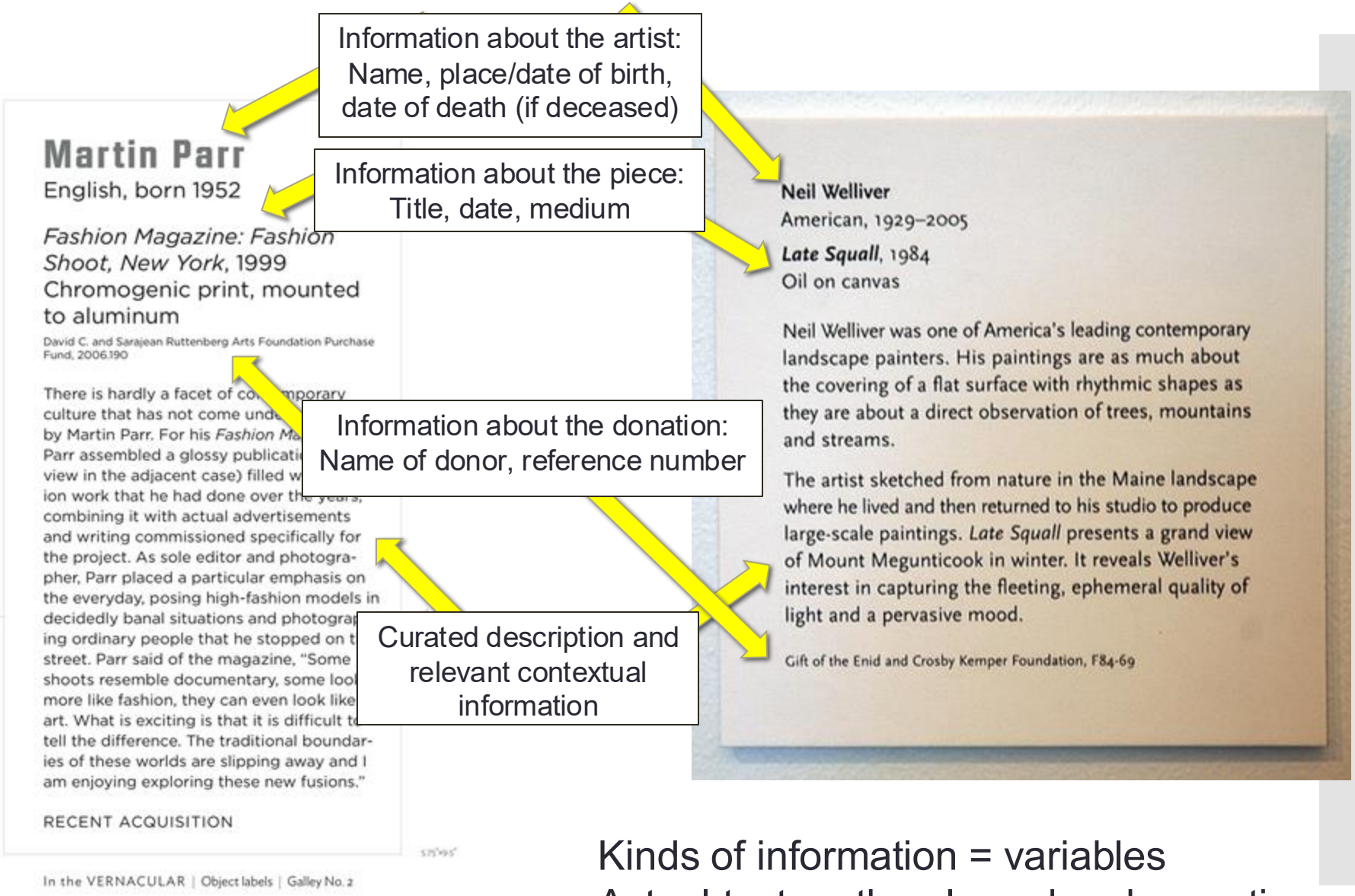
Also called: Attributes, Dimensions

VARIABLES

OBSERVATIONS

	Tuition	Enrollment	Public vs. Private	…
Smith College	$46,288	2,563	private	
UMass Amherst	$16,115	28,635	public	
Hampshire College	$48,065	1,400	private	
Mount Holyoke College	$43,886	2,189	private	
Amherst College	$50,562	1,792	private	
⋮				

Another way to think about this



Kinds of information = variables
Actual text on the placard = observations

Data: a definition

Each variable may be either *independent* or *dependent*:

- An *independent variable* is not controlled or affected by another variable (e.g., time in a time-series dataset)
- A *dependent variable* is affected by a variation in one or more associated independent variables (e.g., temperature in a region)

Basic data types

- Nominal
- Ordinal
- Scale / Quantitative
  - Ratio
  - Interval

An unordered set {…}
of non-numeric values

For example:

- Categorical (finite) data
  {apple, orange, pear}
  {red, green, blue}
- Arbitrary (infinite) data
  {“12 Main St. Boston MA”, “45 Wall St. New York NY”, …}
  {“John Smith”, “Jane Doe”, …}

Basic data types

- Nominal
- Ordinal
- Scale / Quantitative
  - Ratio
  - Interval

An ordered set <…>
(also known as a tuple)

For example:

- Numeric: <2, 4, 6, 8>
- Binary: <0, 1>
- Non-numeric:
  <G, PG, PG-13, R>

Basic data types

- Nominal
- Ordinal
- Scale / Quantitative
  - Ratio
  - Interval

A numeric range […]

Ratios

- Distance from “absolute zero”
- Can be compared mathematically using division
- For example: height, weight

Intervals

- Ordered numeric elements that can be mathematically manipulated, but cannot be compared as ratios
- E.g.: date, current time

Converting between basic data types

- Q → O [0, 100] → <F, D, C, B, A>
- O → N <F, D, C, B, A> → {C, B, F, D, A}
- N → O (??)
  - {John, Mike, Bob} → <Bob, John, Mike> ??
  - {red, green, blue} → <blue, green, red> ??
- N → Q (??)
  - Bob + John = ??

Readings in Information Visualization: Using Vision To Think. Card, Mackinglay, Schneiderman, 1999

Basic operations

- Nominal (N)
  - Equality: = and ≠
  - Frequency: how often does *x* appear?
- Ordinal (O)
  - Relation to other points: >, <, ≥, ≤
  - Distribution: relative frequency
- Quantitative (Q)
  - Other mathematical operations: (+, -, *, /, etc.)
  - Descriptive statistics: average, standard deviation, etc.

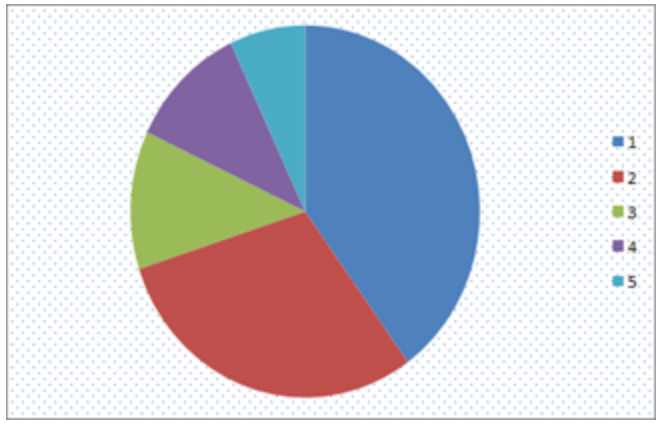
Why is this important?

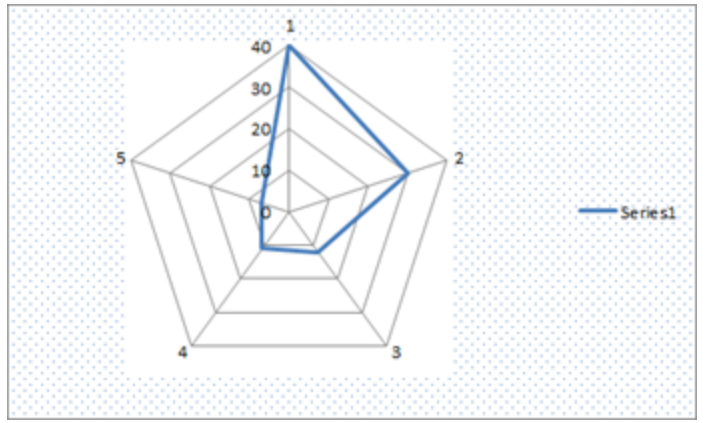
- Data have variables
- Visualizations have variables, too
- To build visualizations, we need to **map** data variables to visual variables

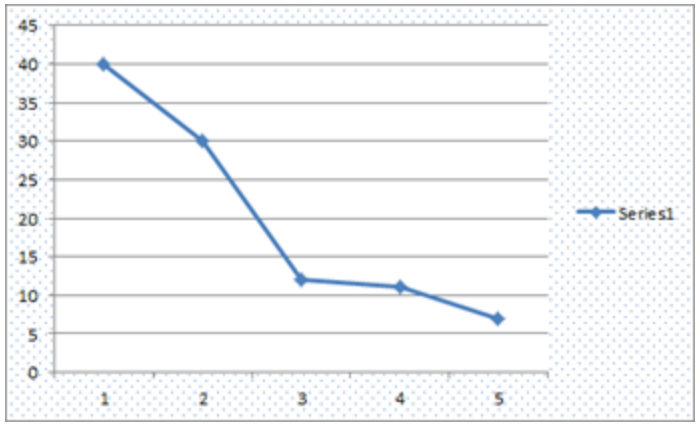
Key question for this course

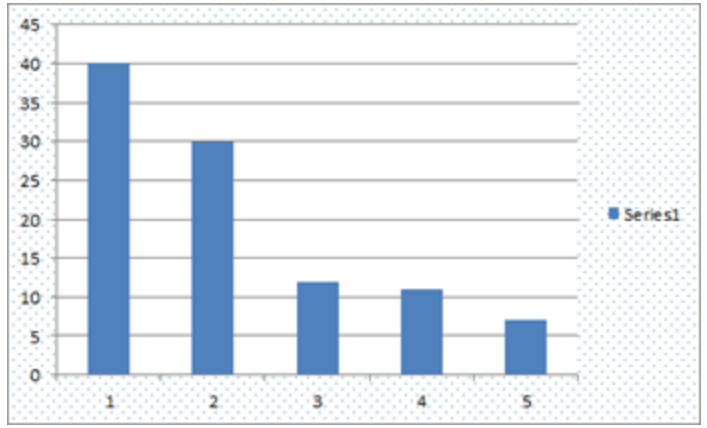
Which **data dimension** should be mapped

to which **visual dimension?**

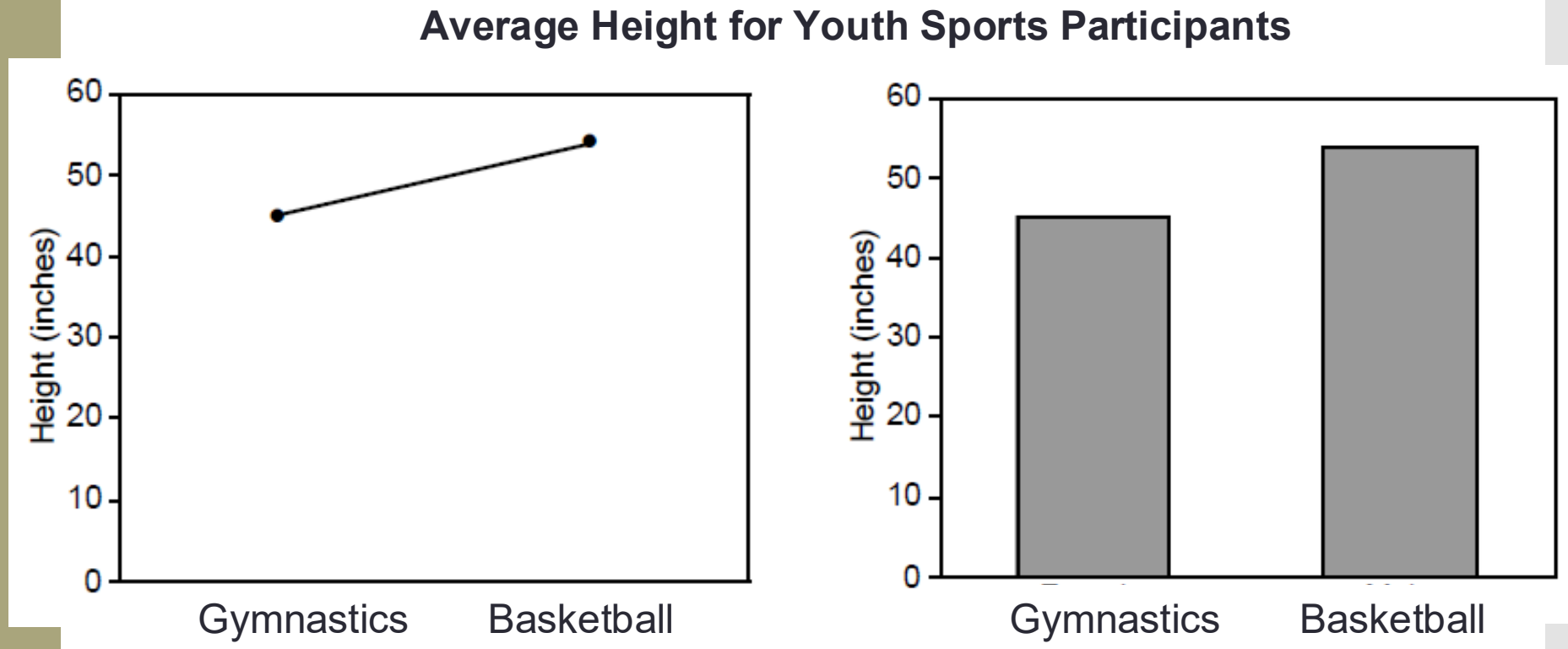








Answer: it depends



What we'll cover in this class

- Introduction to data
- Data-visual mapping
- Introduction to perception
- Tableau
- User-centered design

What we will **not** cover in this class

- Introduction to coding
- Introduction to statistics
- Data wrangling
- Advanced visualization theory (take 235 for this!)

General information

Course website:

- https://amosca01.github.io/SDS-CS109/

- Syllabus
- Slides
- Schedule
- Assignments
- Links

General information

Time Management:

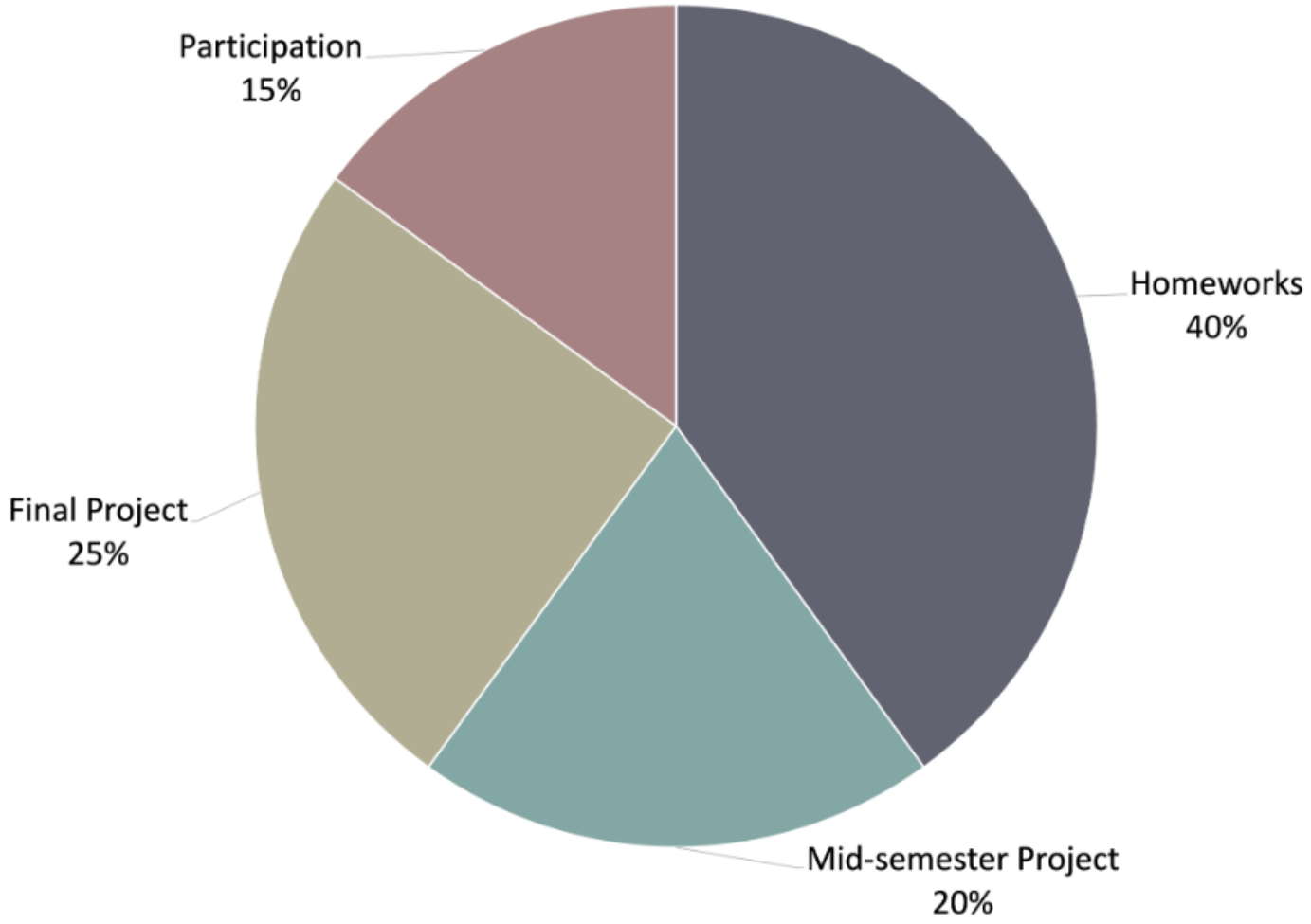
Smith college guidelines suggest 3 hours of work per credit per week. This is a 4 credit course, so you should expect to spend 12 hours on it per week:

- 2.5 hours in class
- 1-2 hours preparing for class
- **7.5-8.5 hours on homework and outside study**

The purpose of class time is to introduce and explore new topics. To complete assignments, you will often need to look up, learn, and practice material that extends beyond that covered during class meeting times.

General information

Assessment:



General information

Late Policy:

We develop skills through practice, and assignments build on one another **AND** life happens.

- You have unlimited extensions in this class
- Extensions must be requested before assignment due date
- For an extension, submit a file with:
  - Your name (and group members' names)
  - Assignment number and original due date
  - Duration of extension and new due date
  - Review of any prior extension requests
  - Ex.

```
Ab Mosca and Jordan Crouser
HW09 originally due 10/01/24
2 day extension, now due 10/03/2024
Previous extentions: HW02 (1 day), HW7 (2 days)
```

- Late work receives lowest priority for grading

General information

Revise and Resubmit:

We develop skills through repeated practice and learn from mistakes.

- You have unlimited revise and resubmits in this class
- You cannot revise and resubmit an assignment on which you earned a o
- **Your resubmission must include a changelog**
- Highest grade will prevail
- Resubmitted work receives lowest priority for grading

General information

Comfy Classroom:

- Need to stand up and leave for a minutes? Do it.
- Want to sit somewhere other than a chair? Go for it.
- Have a concentration aid? Use it.
- Hungry? Thirsty? Eat and drink (be careful of spills!).
- Have kids and no childcare? Bring your kids.

- Do what you need to do to learn, just be respectful of other learners.

What you'll get

By the end of this course, you will:

- Know how to use visualization to **communicate ideas**
- Know the **foundational methods** and tools available
- Be familiar with **ongoing research** in visualization
- Have (marketable!) experience developing useful visualizations to solve **real problems**

What I expect from you

- You like challenging problems, and you're excited about "figuring stuff out"
- You're willing to get comfortable asking questions
- You're interested in the perspectives of people with very different backgrounds from your own
- You turn things in on time, show up to class, or let me know in advance

What you can expect from me

- I'm flexible w.r.t. the topics we cover:
  - This course is a collaboration
  - If there's something you want to learn that's not on the agenda, speak up!
  - If I'm doing something that doesn't work for you (Font too small on presentations? Speaking too quickly? Using a marker or color you can't see?), please let me know!

I'm here to help you succeed, and I believe you all have the ability to succeed

Questions?