

```
    if len(modifiers) > 0:
        r_mod.use_x = False
        or_mod.use_y = True
        or_mod.use_z = False
        if operation == "MIRROR_X":
            r_mod.use_x = True
            or_mod.use_y = False
            or_mod.use_z = False
        elif operation == "MIRROR_Y":
            r_mod.use_x = False
            or_mod.use_y = True
            or_mod.use_z = False
        elif operation == "MIRROR_Z":
            r_mod.use_x = False
            or_mod.use_y = False
            or_mod.use_z = True
    else:
        bpy.context.scene.objects.active = modifier_ob
        int("Selected" + str(modifier_ob)) # modifier ob is the active
        #mirror_ob.select = 0
        if bpy.context.scene.objects.active != None:
            bpy.context.selected_objects[0]
            one.name.select = 1
            mirror_mod.use_x = True
            if operation == "MIRROR_X":
                mirror_mod.use_y = False
                mirror_mod.use_z = False
            elif operation == "MIRROR_Y":
                mirror_mod.use_x = False
                mirror_mod.use_z = False
            elif operation == "MIRROR_Z":
                mirror_mod.use_y = False
                mirror_mod.use_z = True
            mod.mirror_object = mirror_ob
            if len(modifiers) > 0:
                if operation == "MIRROR_X":  
                    r_mod.use_x = True  
                    or_mod.use_y = False  
                    or_mod.use_z = False  
                elif operation == "MIRROR_Y":  
                    r_mod.use_x = False  
                    or_mod.use_y = True  
                    or_mod.use_z = False  
                elif operation == "MIRROR_Z":  
                    r_mod.use_x = False  
                    or_mod.use_y = False  
                    or_mod.use_z = True
```

Day One: Introduction to Data Science



Data Science with R and Python

À propos de moi

- PhD Student in Sociology and Demography
- I was born in Atlanta and grew up in Myrtle Beach, SC
- First-gen student
- Love reading, biking, lifting, museum exploring, traveling, being active (in general).



Pre-Class Assignment

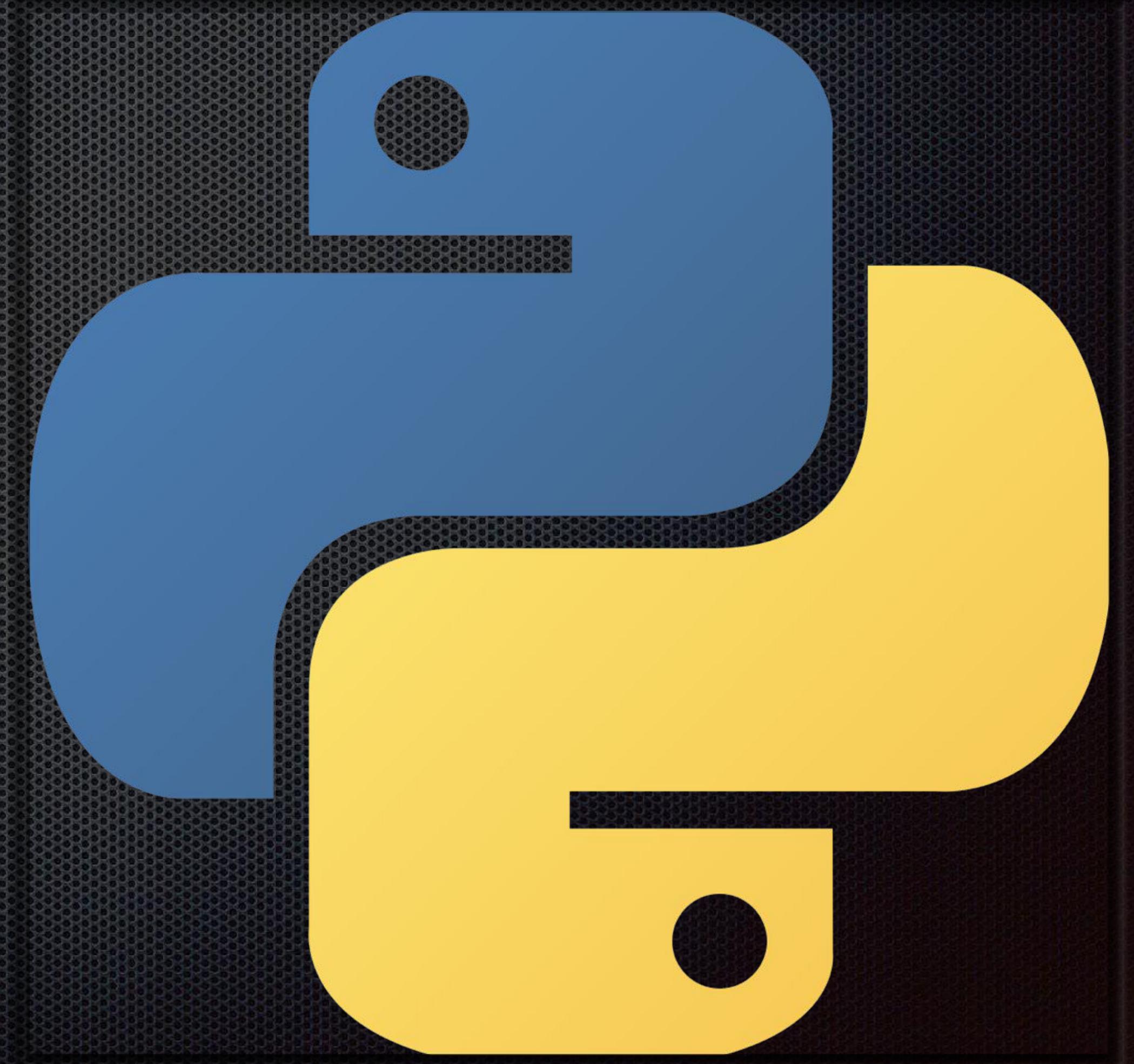
(For instructor) Q: Were you able to install R and R studio?

(For Discussion) Q: What did you find the most interesting from the reading?

Q: What do you wish to get out of the course?



What is R and what is Python?



Overview of Data Science

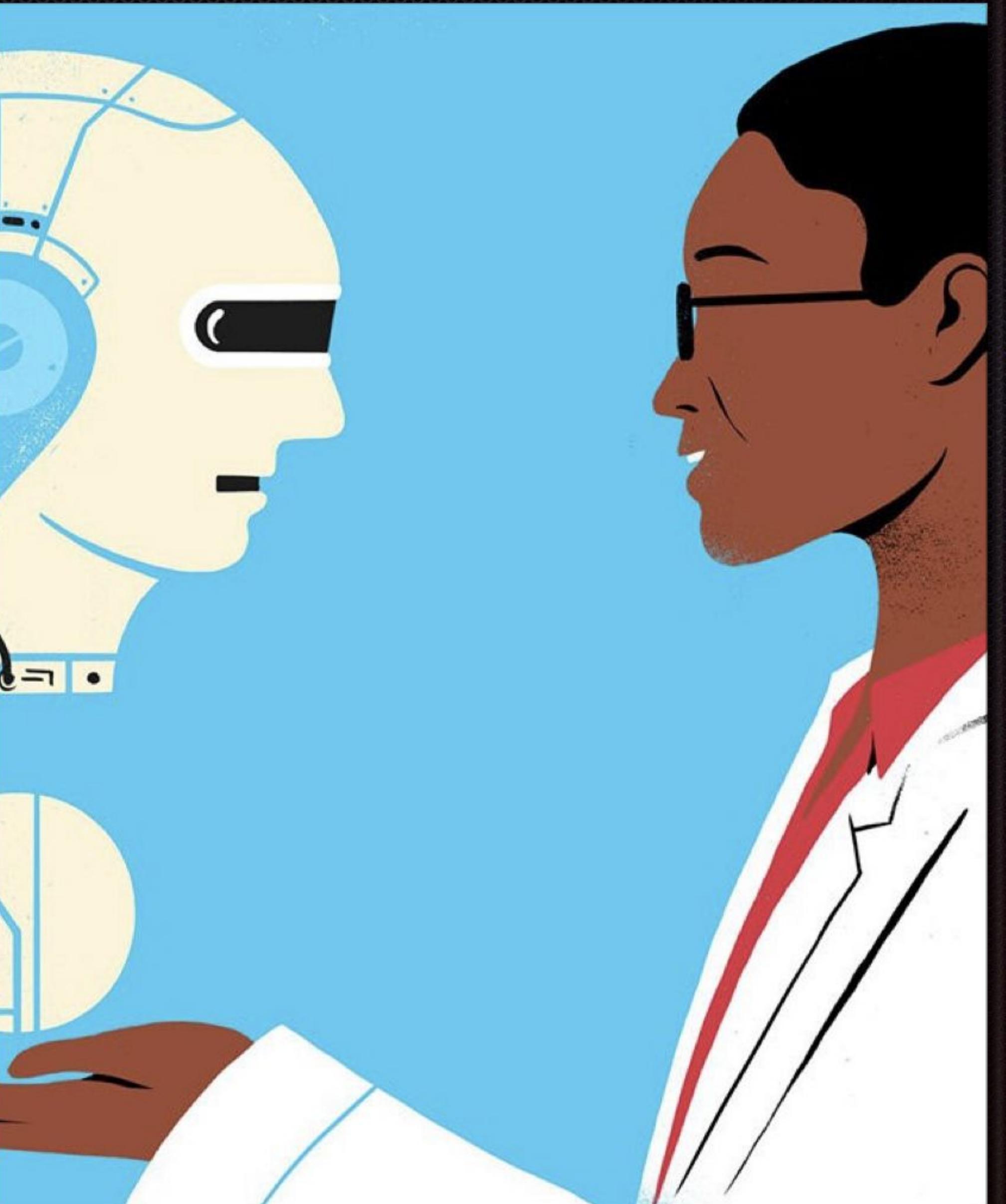
- . **Inference:** A (sometimes) complex procedure for drawing conclusions about large populations based on samples.
- . Q: What is data science? A: It is an interdisciplinary field that combines programming skills, and knowledge of mathematics and statistics to extract meaningful insights from (often messy and unstructured) data.



Picture here: John Tukey

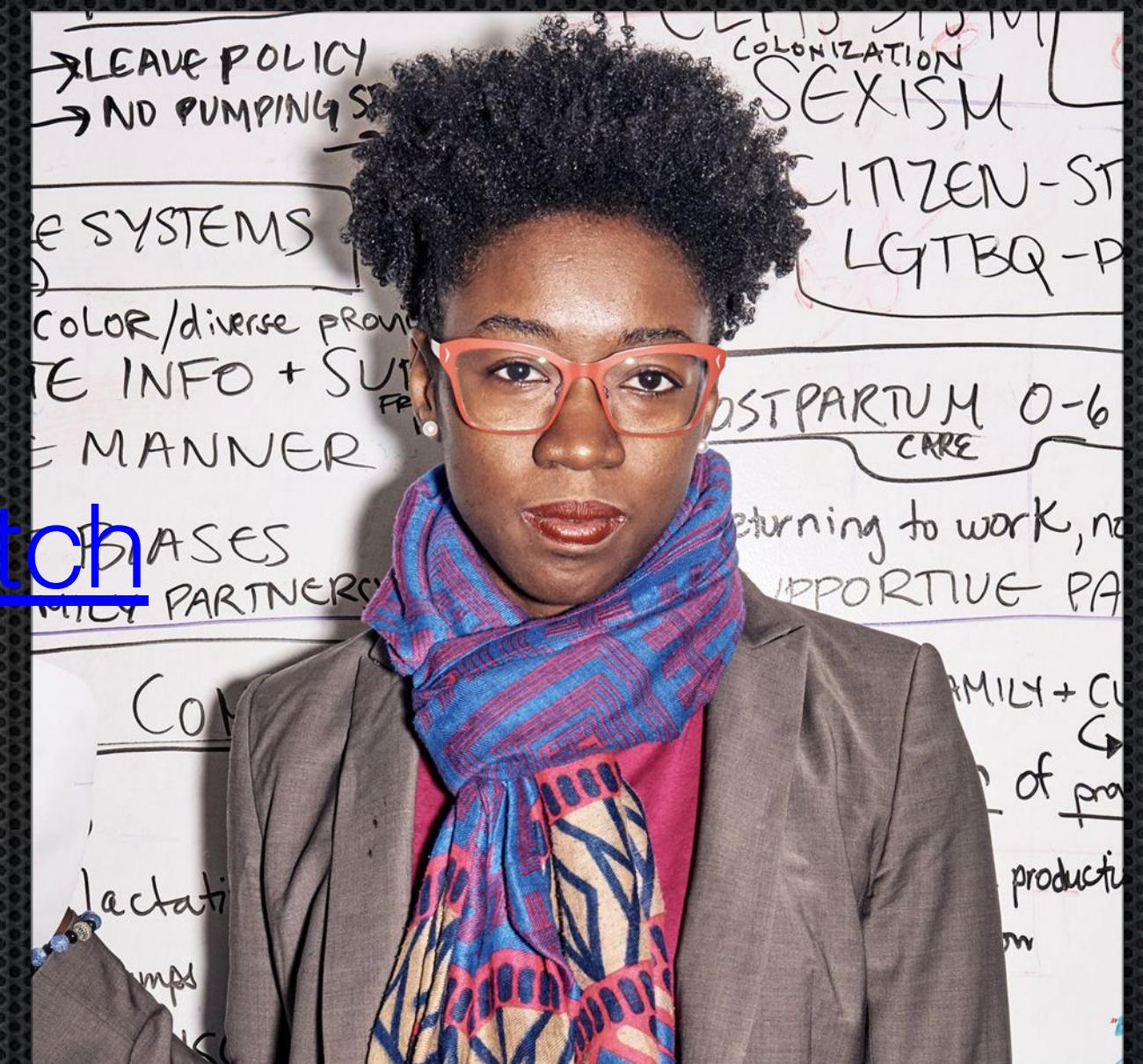
Why is Data Science important?

- Enables us to understand data to make smarter data-driven decisions
- It does not have a moral compass and thus, humans who deal with data should understand the ethical implications of being an authority on data
- It has the power to structure our world in very real ways (e.g., uber, google searches, credit score decisions)
- It is a big part of our lives and we should aim to understand it in both a technical and ethical and social manner.



Before Getting Started with Data: Fighting Bias in Algorithms

- Joy Buolamwini (TED)
https://www.youtube.com/watch?v=UG_X_7q63rY



Where does data come from?

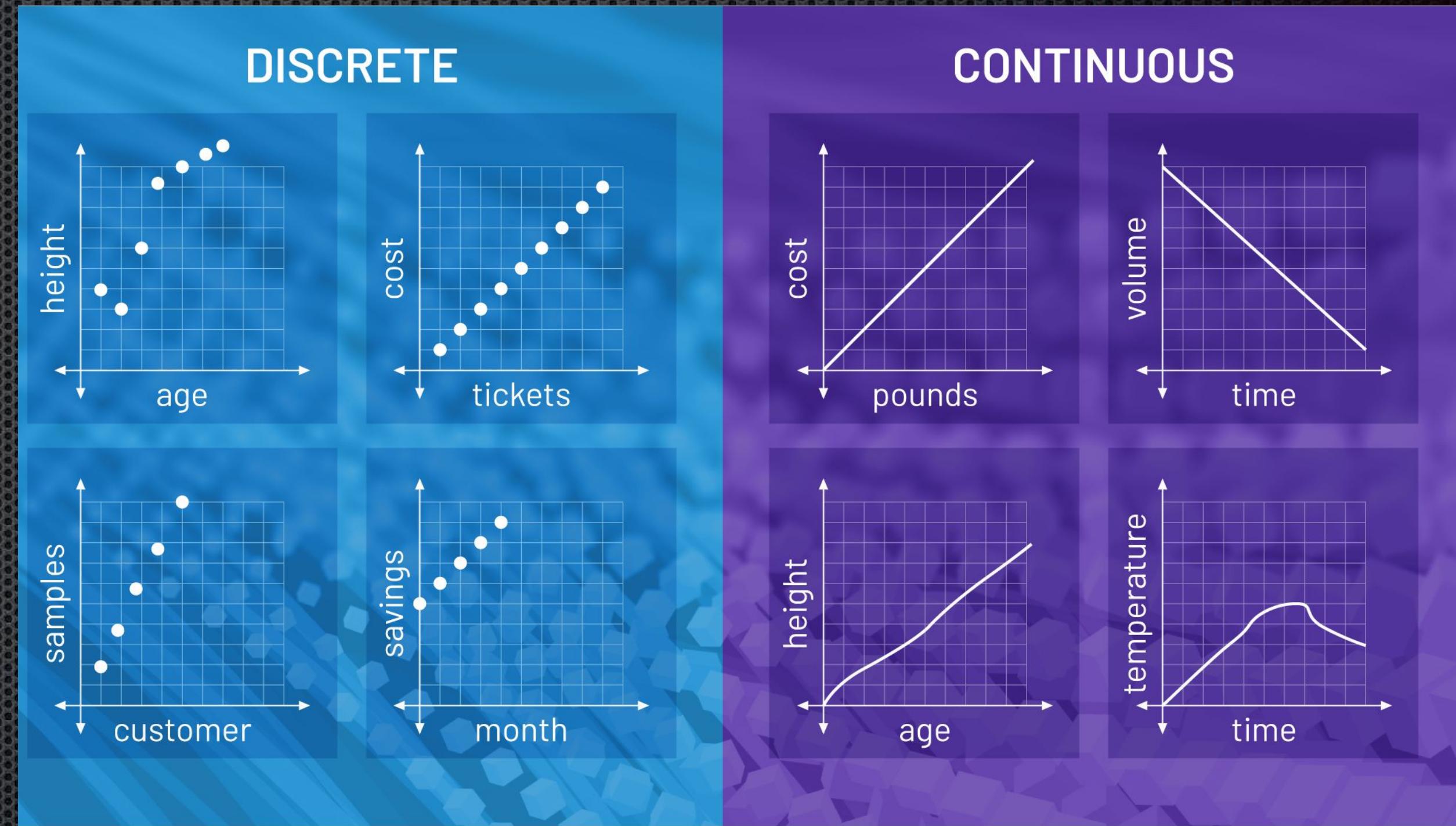
Q: What kinds of things can we obtain data from?

A: Sensor measurements, events, text, images, and videos



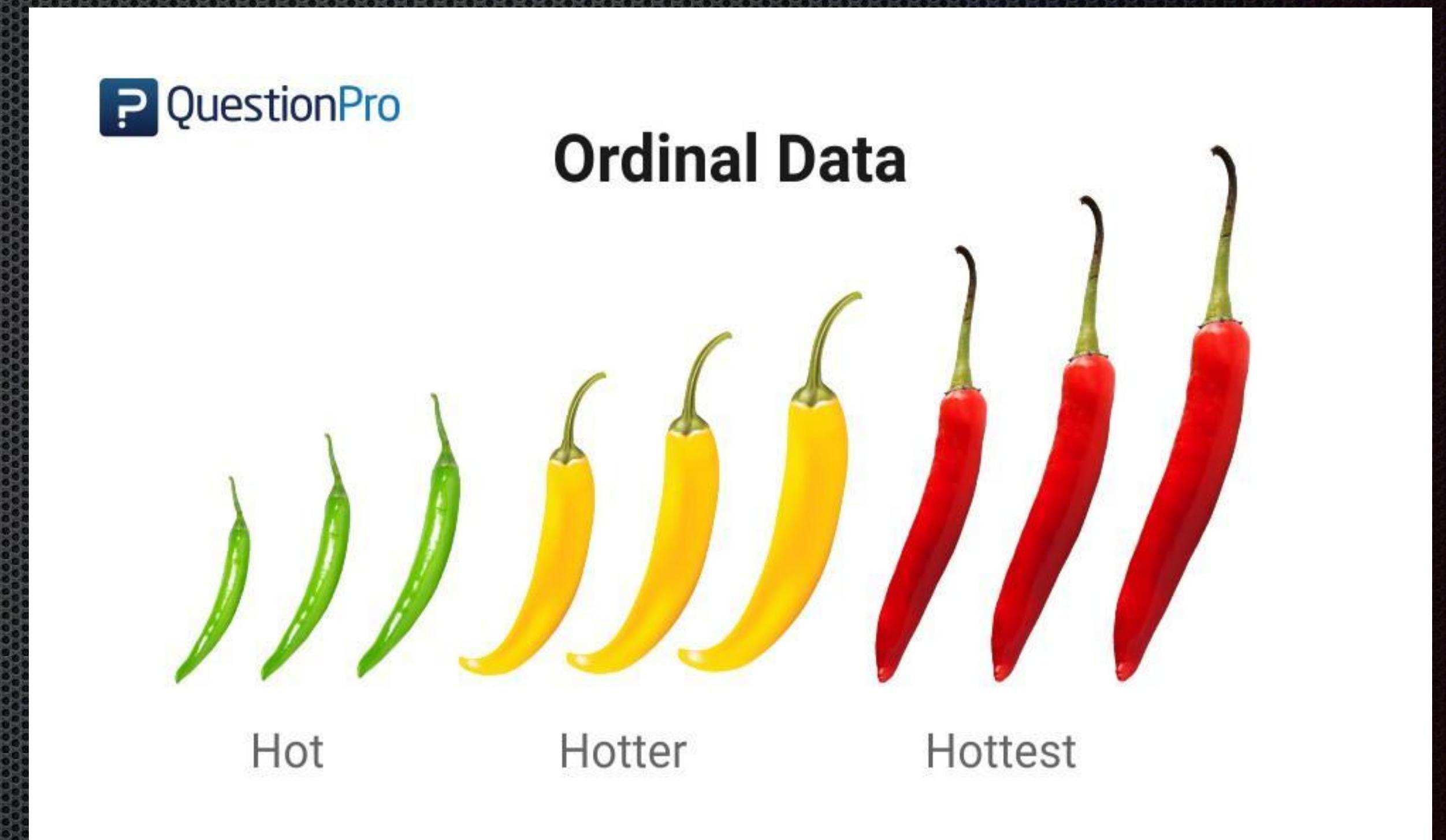
Types of data

- . Q: What are different kinds of data types that exist?
- . A: Continuous: : data that can take on any value in an interval
- . A: Discrete: data that can only take on integer values, such as counts



Types of data

- . Q: What are different kinds of data types that exist?
- . A: Binary: a special case of categorical data with just two categories of values (0/1, true/false)
- . A: Ordinal : categorical data that has an explicating ordering



Typical Data Format

Data Science with R and Python

Data Frame: a rectangular data (like a spreadsheet) that is the basic structure for statistical and machine learning models

Table 1-1. A typical data format

Rectangular Data

- . Rectangular data: a two-dimensional matrix with rows indicating observations and columns indicating features (variables)
 - . Q: Does data always come in this format? What must happen to data that does not look this way?

Table 1-1. A typical data format

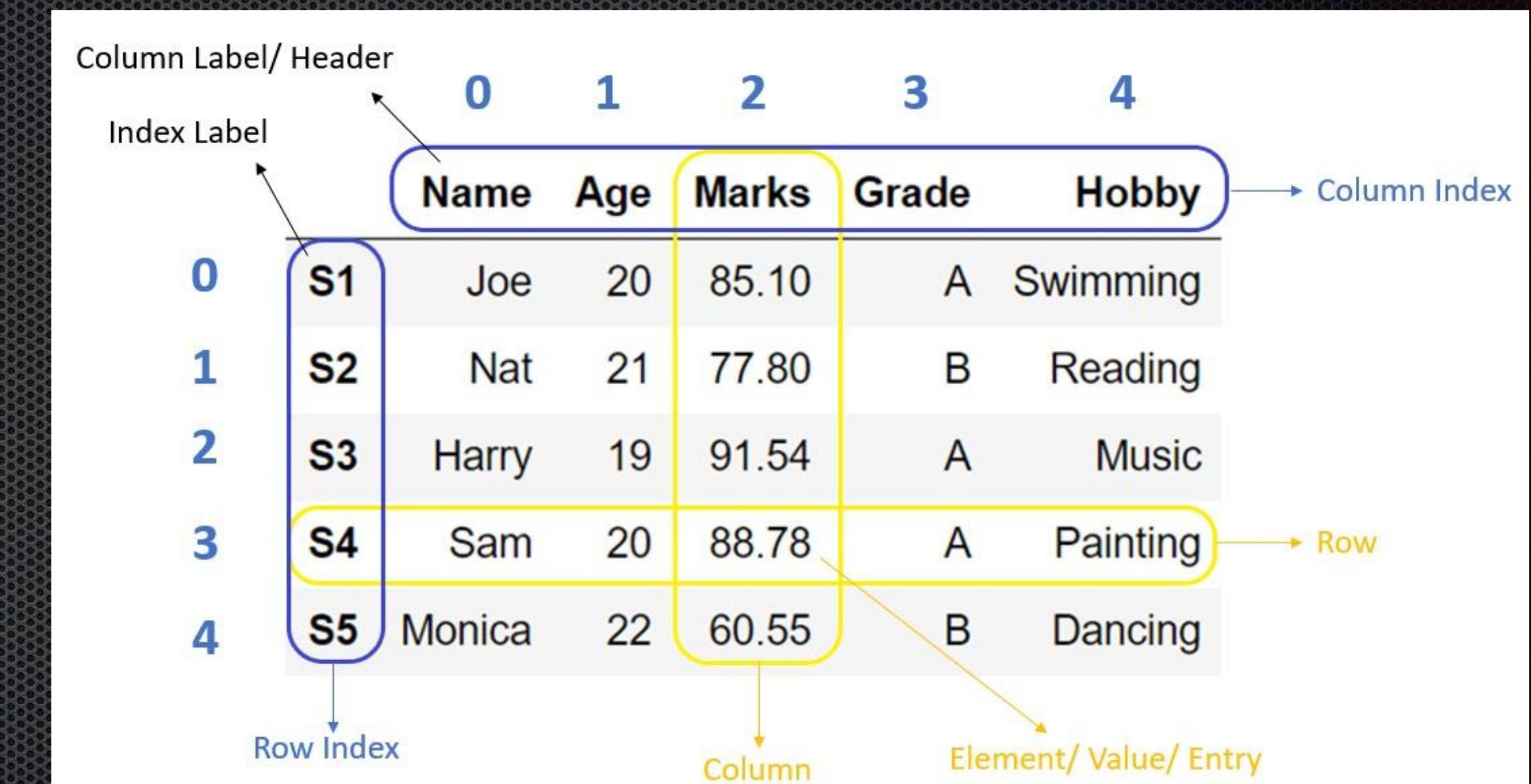
Data Frames in R vs. Python

R Intro-to-R.R x new_metadata x

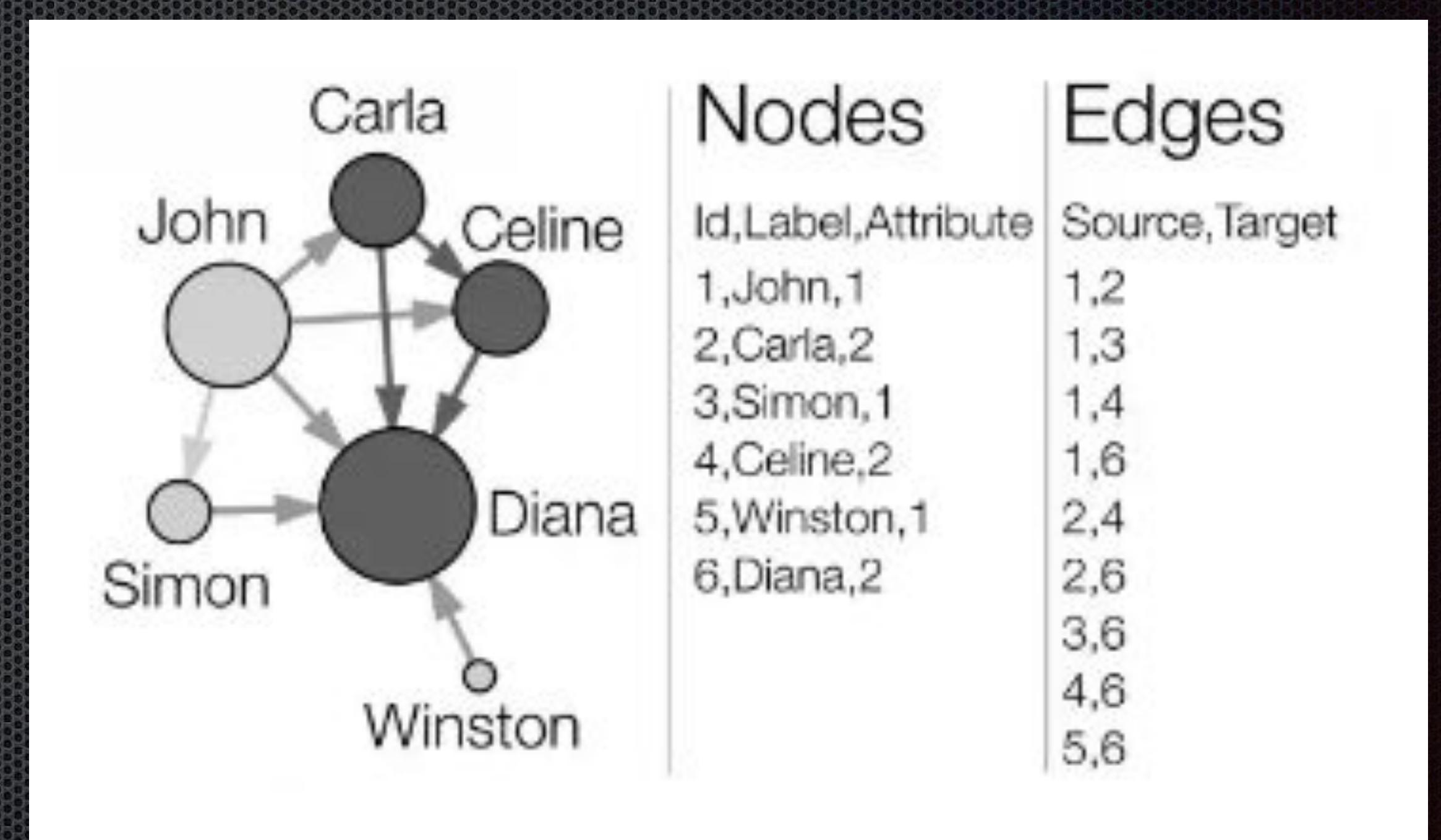
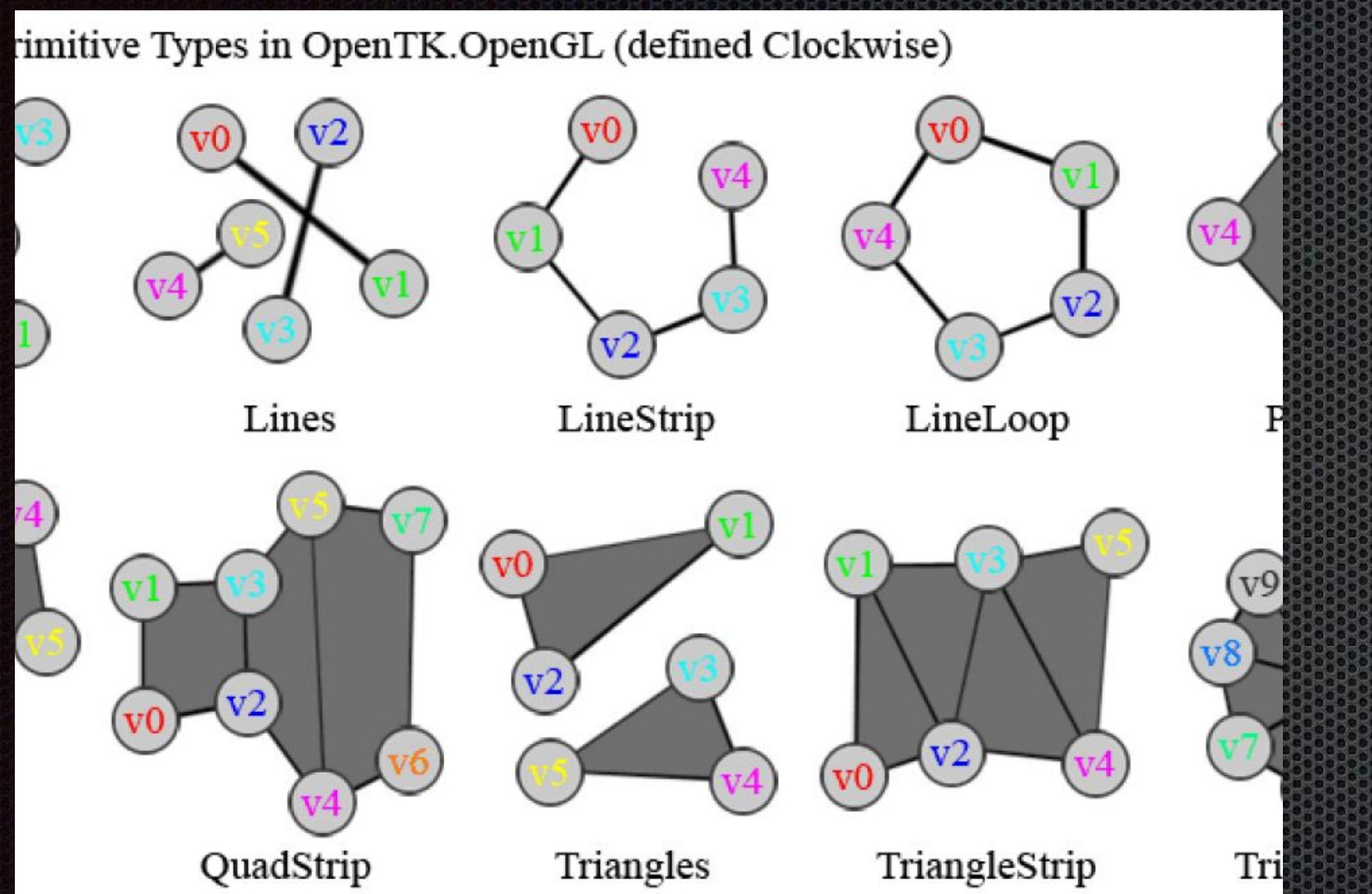
Filter

	genotype	celltype	replicate	samplemeans	age_in_days
sample1	Wt	typeA	1	10.266102	40
sample2	Wt	typeA	2	10.849759	32
sample3	Wt	typeA	3	9.452517	38
sample4	KO	typeA	1	15.833872	35
sample5	KO	typeA	2	15.590184	41
sample6	KO	typeA	3	15.551529	32
sample7	Wt	typeB	1	15.522219	34
sample8	Wt	typeB	2	13.808281	26
sample9	Wt	typeB	3	14.108399	28
sample10	KO	typeB	1	10.743292	28
sample11	KO	typeB	2	10.778318	30
sample12	KO	typeB	3	9.754733	32

Showing 1 to 12 of 12 entries, 5 total columns



Non-rectangular data



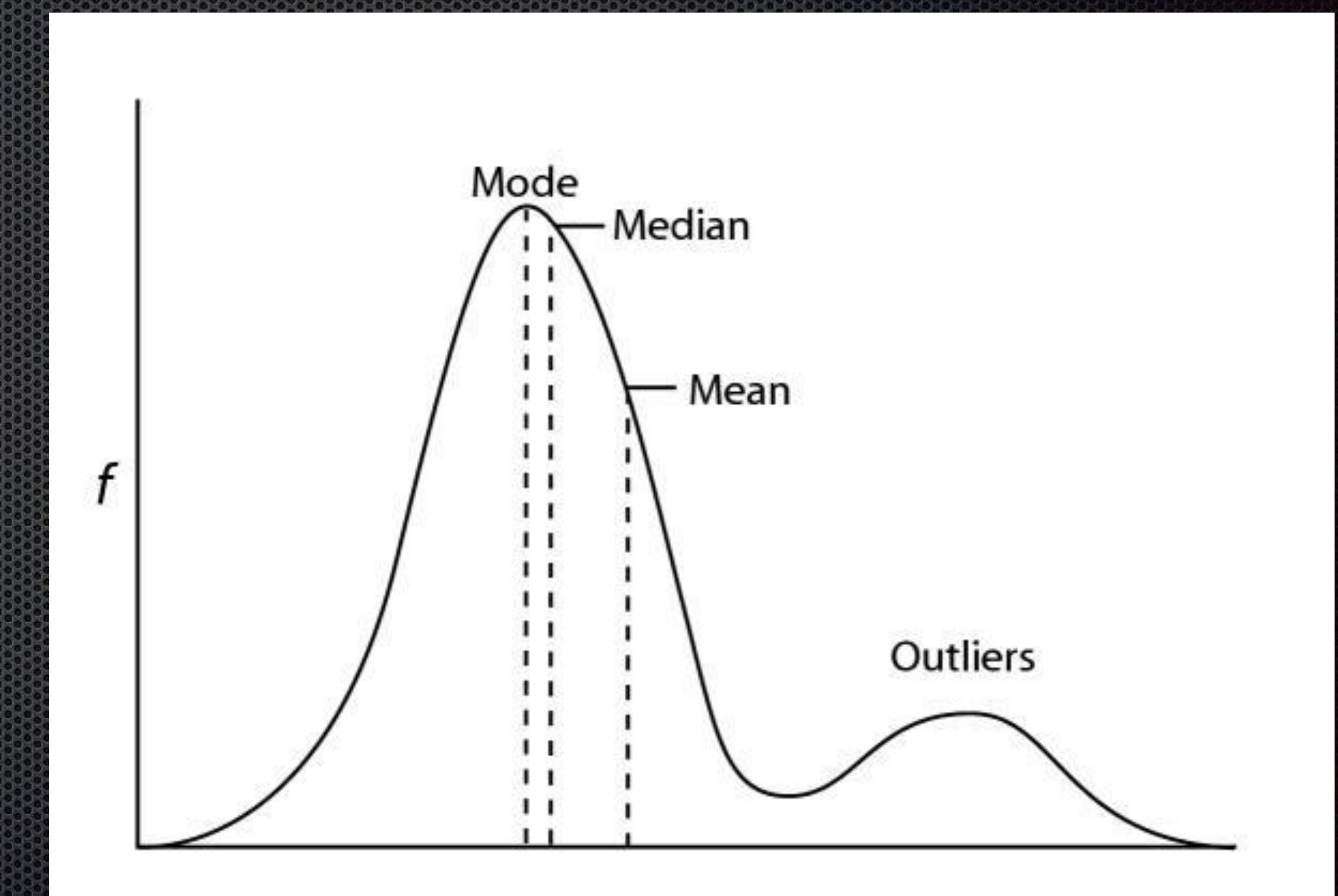
Breakout rooms: Key Estimates of Location (15 minutes)

1. Mean:
2. Weighted mean:
3. Median:
4. Weighted median:
5. Trimmed mean:
6. Robust:
7. Outlier:



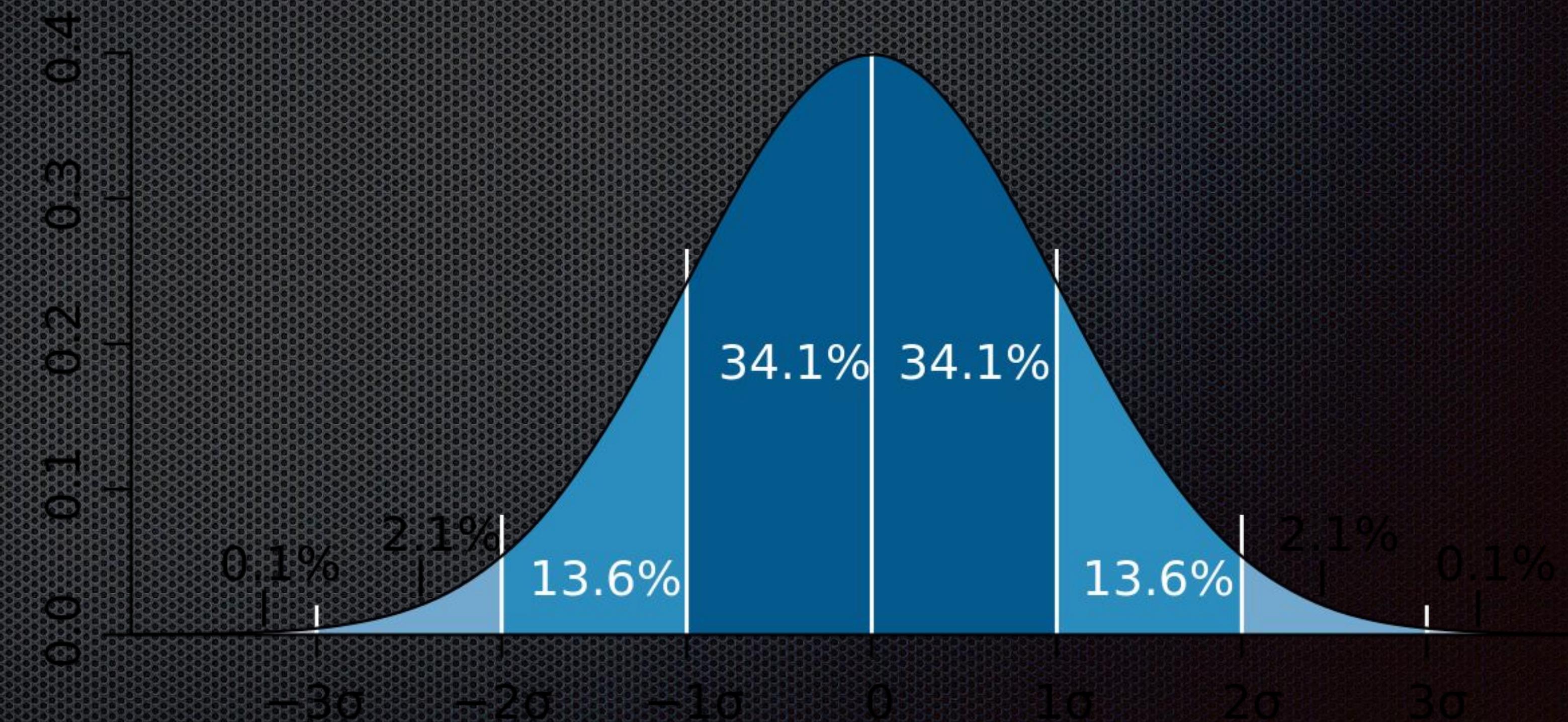
Estimates of Location (Key Terms)

1. Mean: sum of all values divided by the number of values
2. Weighted mean: sum of all values divided by the sum of weights
3. Median: value such that one-half of the data lies above and below (50th percentile)
4. Weighted median: value such that one-half of the sum of weights lies above and below sorted data
5. Trimmed mean: average of all values after dropping a fixed number of extreme values
6. Robust: not sensitive to extreme values
7. Outlier: a data value that is very different from most



Breakout rooms: Key Estimates of Variability (15-20 minutes)

- . 1. What is variability?
- . Define:
- . Deviations:
- . Variance:
- . Standard deviation:
- . Range:
- . Percentile:



Key Estimates of Variability

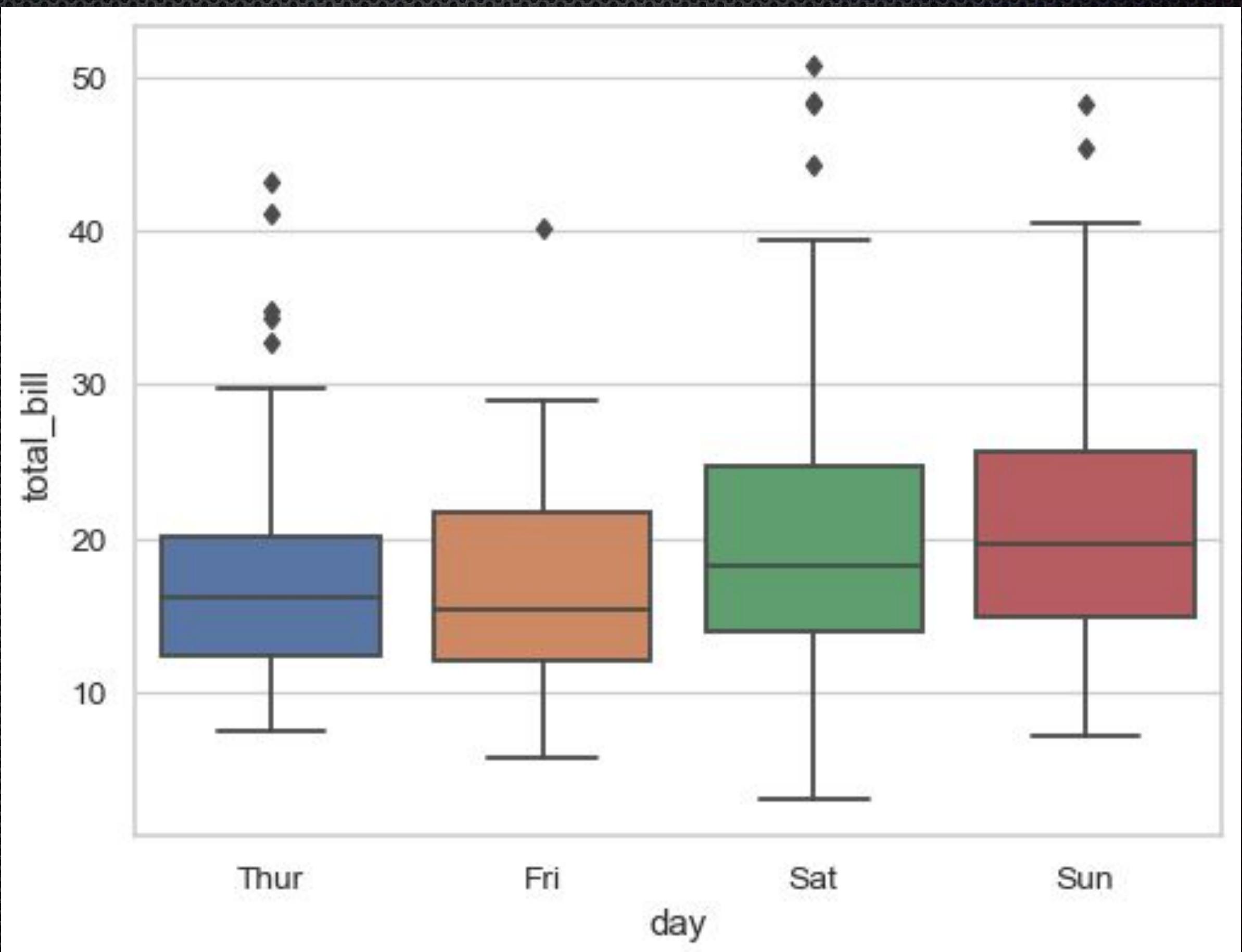
- . 1. What is variability?: measures whether the data values are tightly clustered or spread out.
- . Define:
 - . Deviation: Difference between observed values and estimate of location
 - . Variance: Sum of squared deviations from the mean divided by $n-1$ where n is number of data values.
 - . Standard deviation: Square root of variance
 - . Range: difference between largest and smallest value in a data set
 - . Percentile: the value such that P percent of the values take on this value or less and $(100-P)$ percent take on this value or more
 - . VIDEO to reiterate: https://www.youtube.com/watch?v=wDAd_QHKoOg

$$\text{Variance} = s^2 = \frac{\sum (x - \bar{x})^2}{n - 1}$$

$$\text{Standard deviation} = s = \sqrt{\text{Variance}}$$

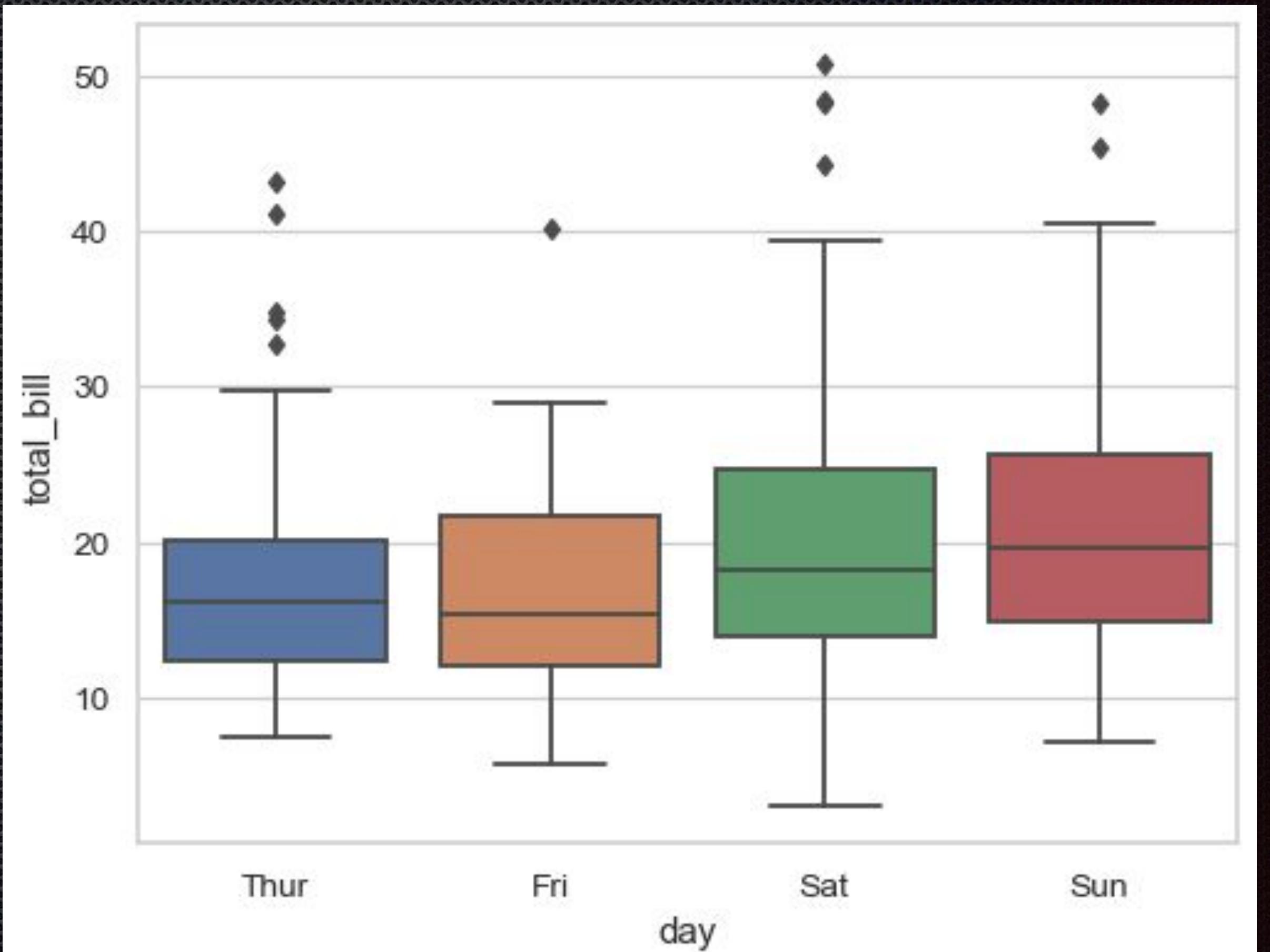
Exploring Data Distributions (Key terms)

- . Define:
- . Boxplot:
- . Frequency table:
- . Histogram:
- . i.i.d.



Exploring Data Distributions (Key terms)

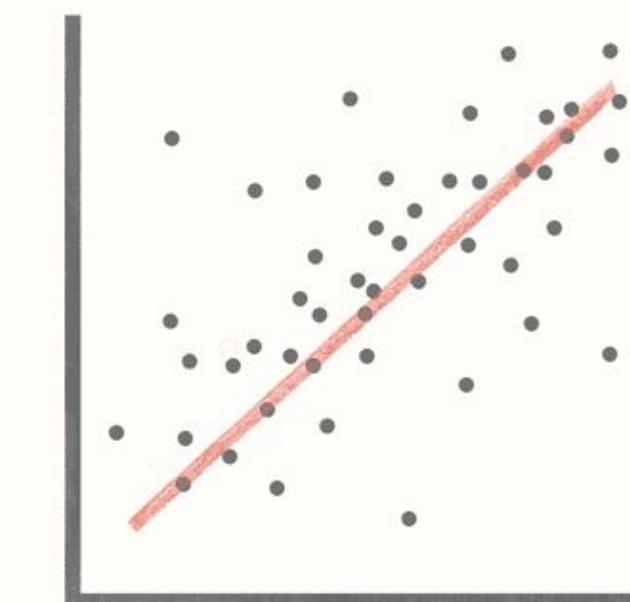
- . Define:
 - . Boxplot:introduced by Tukey as a quick way to visualize distribution of data
 - . Frequency table: a tally of the count of numeric data values that fall into a set of intervals (bins)
 - . Histogram: plot of the frequency table with the bins on the x-axis and the count (or proportion) on the y-axis).
 - . I.i.d. :A collection of random variables is independent and identically distributed if each random variable has the same probability distribution as the others and all are mutually independent. ([Video](#))



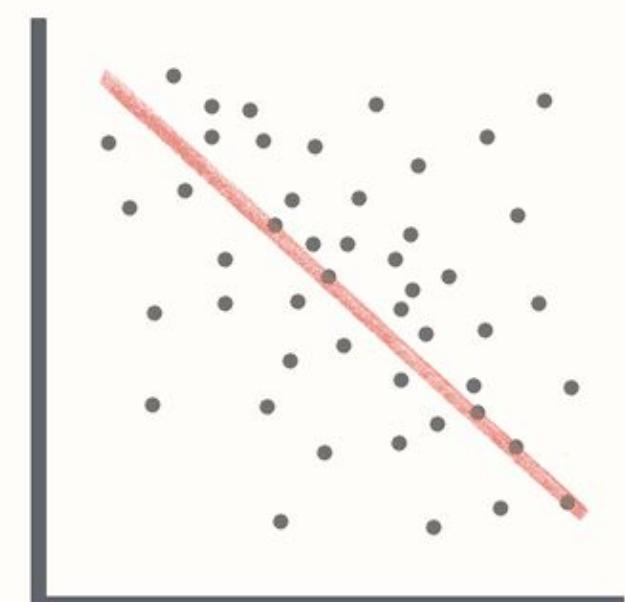
Correlation (explained by Cory Baldwin)

<https://www.youtube.com/watch?v=dsyTQNUvqH0>

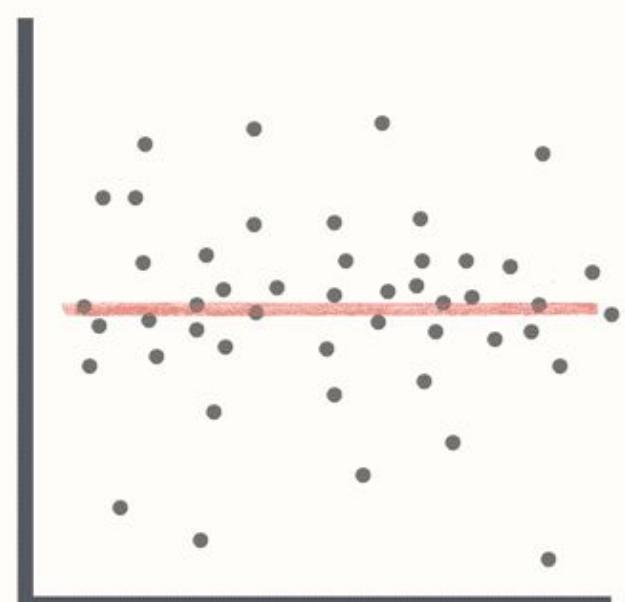
Correlation Coefficient



Positive Correlation



Negative Correlation



No Correlation

Ethics: The Age of Surveillance Capitalism

- Explanation of Shoshana Zuboff:
<https://www.youtube.com/watch?v=FcADchWhwUk>

