Check in questions:
How have your first few weeks been?
Do you think ketchup is a smoothie?

Warm-up

5:00-5:10

1. Make a new R script
   a. Save it to your computer and name it “week2_warmup” Add a
   b. comment with the title of the script, the date, and your name
2. Create three different vectors, each with 5 items
   a. Vector called “pid” with 5 participant id numbers Vector called “ages” with 5
   b. ages of participants Vector called “condition” with the condition of each
   c. participant, either cond1, cond2, or control.
3. Make the condition vector into a factor
4. Run length(pid). What does it tell you?

Extra time? Download this week’s materials from the course site!

Berkeley
UNIVERSITY OF CALIFORNIA

Week 2: Directories and Data

Sophie & Lou

9/16/2024

Today's agenda

1. Warm-up
2. Check-in + any questions about the program?
3. Go over warm-up + data frames (!)
4. Working directories & setting up your project folder
5. Reading in data
6. Viewing and summarizing data

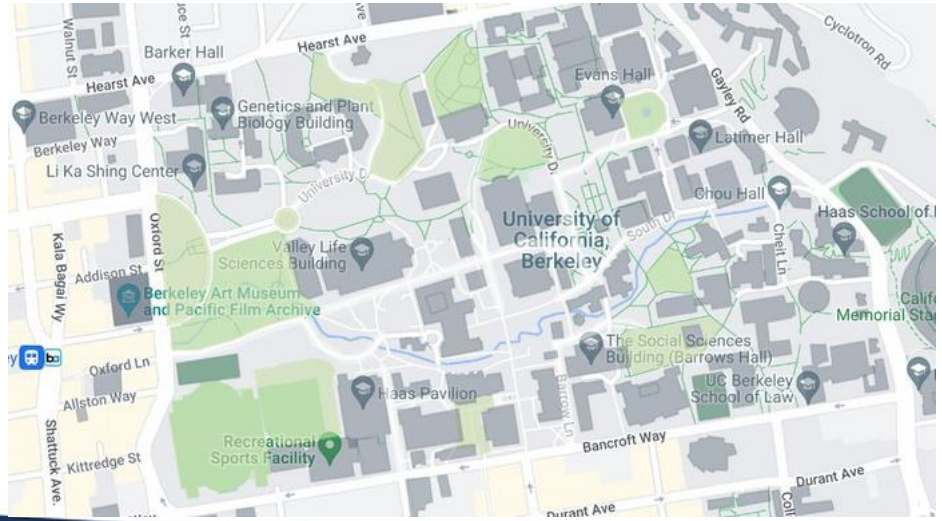
What is a “working directory”?

- The current file path that R is using and where it will “look” for files
  - R will assume you want to read in or write out files using this folder
- You can think of file paths as addresses
  - Each file has one!
- Our files are in a **nested structure**

For more info see YaRR: A Pirate’s Guide to R chapter

For instance....

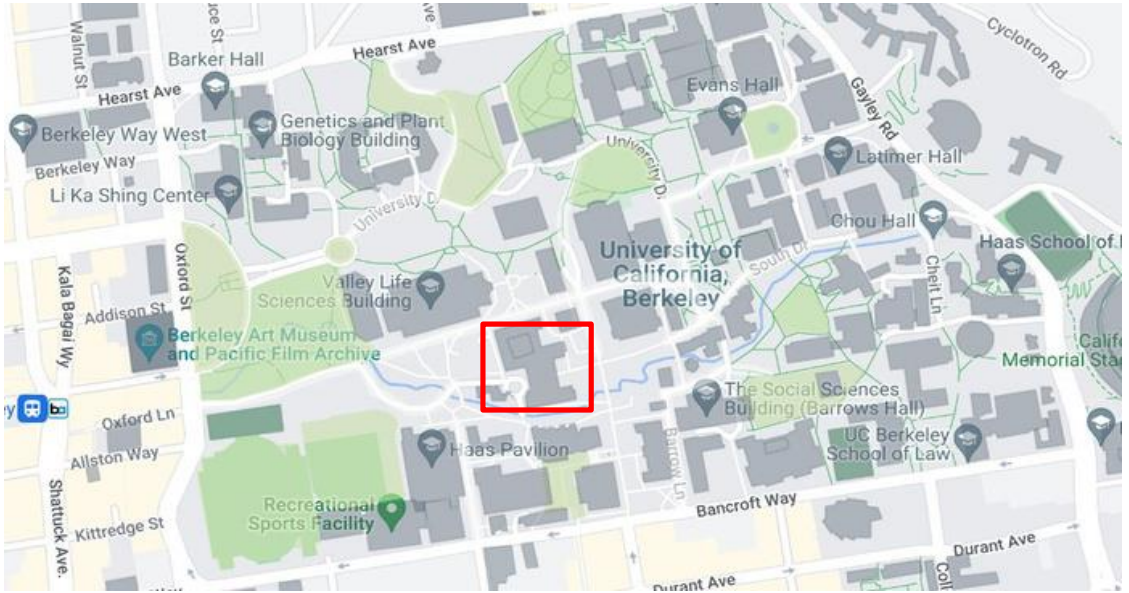
- Let's say you have a class! But where are you going? Where is it being held? File path: "Berkeley/..."
-



Thank you Emily Rosenthal for these slides!

It’s in Dwinelle Hall

- File path: “Berkeley/Dwinelle/…”



It’s in the basement! In room 75!

- File path: “Berkeley/Dwinelle/Basement/Room75”

Example:
Users/SophieRegan/Documents/Experiments/BrillaintStudy/Data

Working directory and loading data

- Organization is up to you -- ask around for how people organize their data and files
  - We have some suggestions, too!
- Keep this structure in mind when loading your data!
- R has to know *specifically* where to look for the data file (its address), or it won't be able to read it in

Most common errors when loading data

- Wrong file path (i.e., wrong working directory so R is looking in the wrong place and cannot find the file)
- Forgot to put quotes around the name of the file and/or file path! read.csv(“penguins.csv”)

Other tips:

- Don’t use spaces in your file names or folder names!

What now?

Commands to Inspect the Data

1

1. head()
   - Displays the first 6 rows of your data.
   - Example: head(data)

2

2. tail()
   - Shows the last 6 rows of your data.
   - Example: tail(data)

3

3. dim()
   - Returns the dimensions of the dataset.
   - Example: dim(data)

Summary and Structure of Data

1. str()
   - Displays the structure of your dataset.
   - Example: str(data)

2. summary()
   - Provides summary statistics for each column.
   - Example: summary(data)

3. names()
   - Lists all column names in the dataset.
   - Example: names(data)

Checking for Missing or NA Values

1. is.na()
   - Identifies missing values in the dataset.
   - Example: is.na(data)

2. sum(is.na())
   - Summarizes the total number of missing values.
   - Example: sum(is.na(data))

3. complete.cases()
   - Finds rows without missing values.
   - Example: complete.cases(data)

Initial Visualization

1. plot()
- Basic plots of data.
- Example: plot(data$variable1, data$variable2)

2. hist()
- Histogram of a specific variable.
- Example: hist(data$variable)

3. boxplot()
- Boxplot to visualize the distribution of data.
- Example: boxplot(data$variable)

Common Data Types in R

1. Numeric
   - Example: x <- 10.5
   - Double precision numbers.

2. Integer
   - Example: y <- as.integer(5)
   - Whole numbers.

3. Character
   - Example: name <- "John"
   - Strings of text.

4. Logical
   - Example: flag <- TRUE
   - Boolean values (TRUE/FALSE).

5. Factor
   - Example: factor_var <- factor(c("Male", "Female"))
   - Categorical data.

Complex Data Types in R

1. Vectors
   - Example: c(1, 2, 3, 4)
   - Homogeneous data.

2. Lists
   - Example: list(name = "John", age = 25)
   - Heterogeneous data.

3. Data Frames
   - Example: data.frame()
   - Similar to tables (rows and columns).

4. Matrices
   - Example: matrix(1:9, nrow = 3)
   - 2D homogeneous arrays.

Comparison of Data Types

Data Type	R	Python	JavaScript
Numeric	Numeric (double)	int, float	Numbers (double)
Character	Character	String	String
Logical/Boolean	Logical	Boolean	Boolean
Factor	Factor	N/A	N/A
Data Frame/Array	Data Frame	Pandas DataFrame	Arrays/Objects