Week 2: Accessing and Viewing Data

Elena (& Willa in spirit)
7/13/2021

Welcome to Week 2!

1. Make sure you have the session 2 practice materials downloaded from the webpage

   https://ucb-psychology-quack.github.io/site/summer_bootcamp/bootcamp

2. Open up `s2_starter_coder.R` in Rstudio and get started with the warm-up

Today's agenda

- Data frames

- Working directory

- Reading in data

- Viewing your data

- Making factors

- Group activity to practice!

What is a "working directory"?

- The current file path that R is using and where it will "look" for files

  - R will assume I want to read in or write out files using this folder

- You can think of file paths as addresses

/Documents

/Downloads

/Music

On "My PC" street....
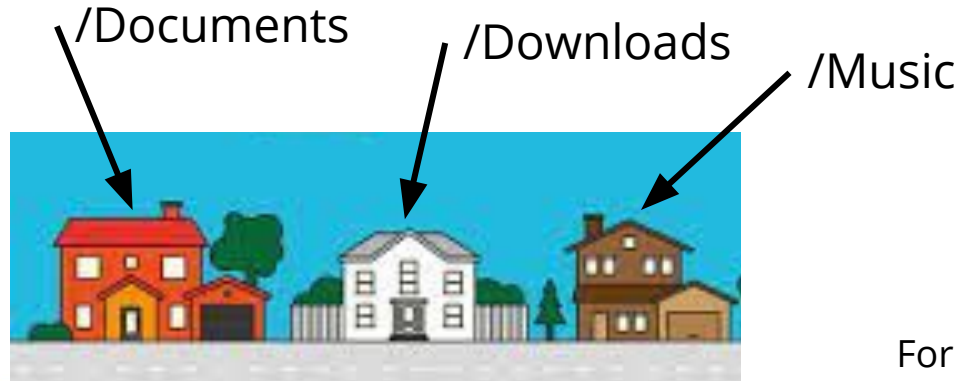
For more info see YaRR: A Pirate's Guide to R chapter

Berkeley
UNIVERSITY OF CALIFORNIA

Neighborhood, streets, and houses have "nested structure"

Street

Neighborhood, streets, and houses have "nested structure"

Street

House 1   House 2

Neighborhood, streets, and houses have "nested structure"

Street

House 1   House 2

Bedroom   Living Room

Neighborhood, streets, and houses have "nested structure"

Street
- House 1
  - Bedroom
  - Living Room
- House 2
  - Bedroom
  - Living Room

Berkeley
UNIVERSITY OF CALIFORNIA

Neighborhood, streets, and houses have "nested structure"

```
                              Street
              ┌─────────────────┴─────────────────┐
           House 1                              House 2
        ┌─────┴─────┐                        ┌─────┴─────┐
     Bedroom    Living Room              Bedroom    Living Room
              ┌─────┴─────┐
           Closet      Drawers
```
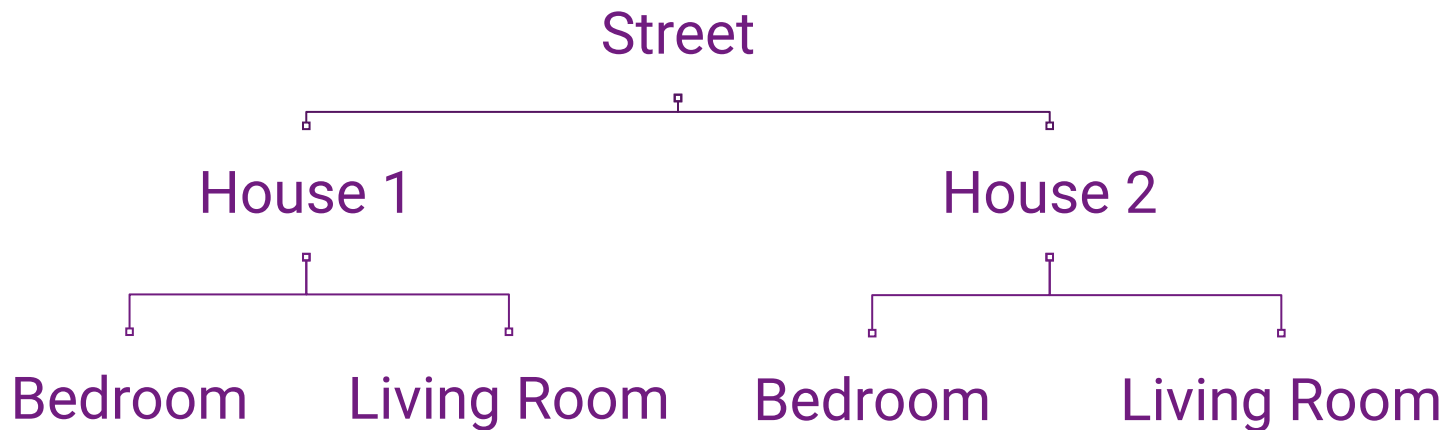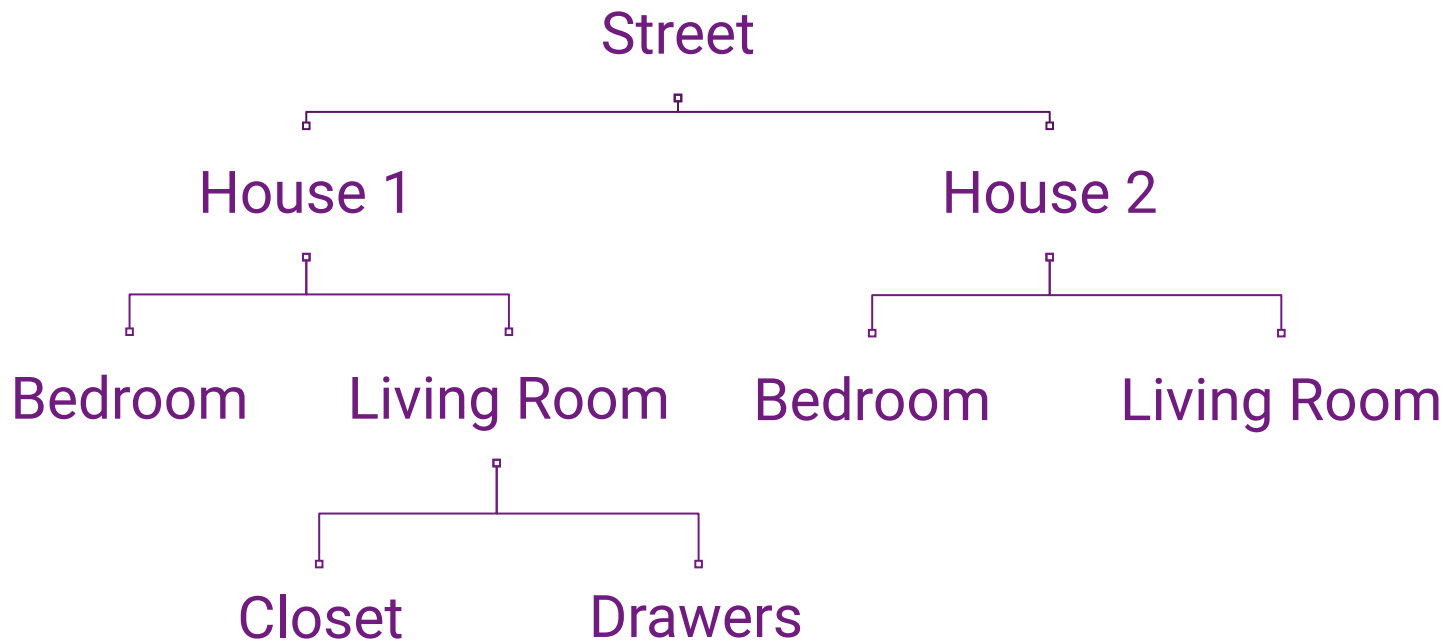
So do folders!

Box

So do folders!

Box

QuACK   Project1

So do folders!

Box

QuACK   Project1

2020_Fall   2021_Summer

So do folders!

Box
- QuACK
  - 2020_Fall
  - 2021_Summer
- Project1
  - 1-Design
  - 2-Data

Berkeley
UNIVERSITY OF CALIFORNIA
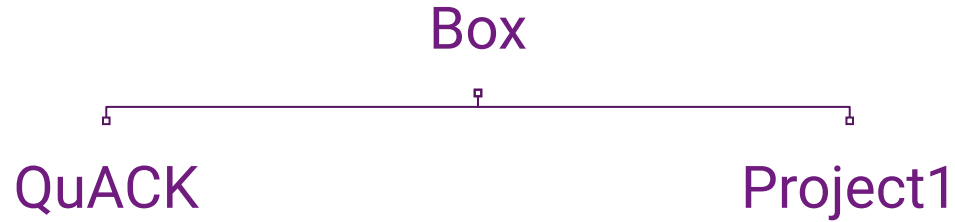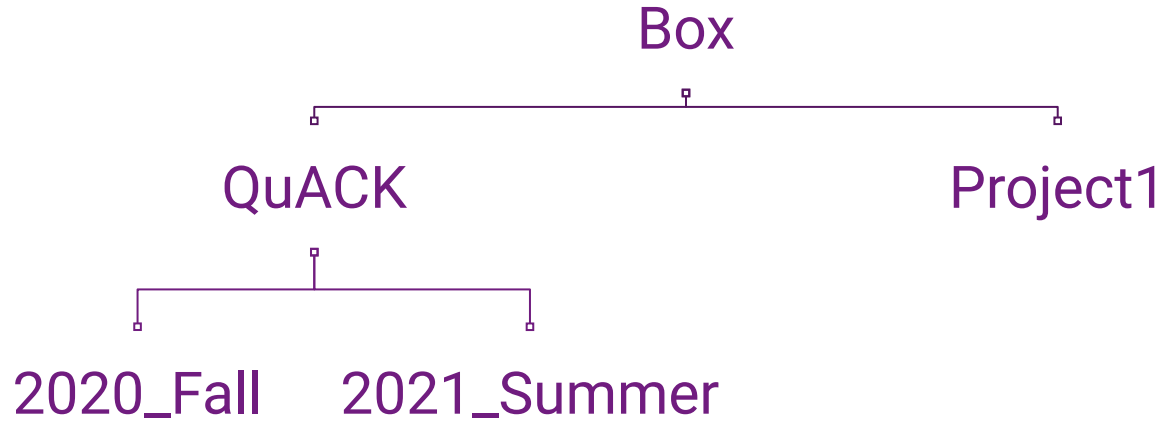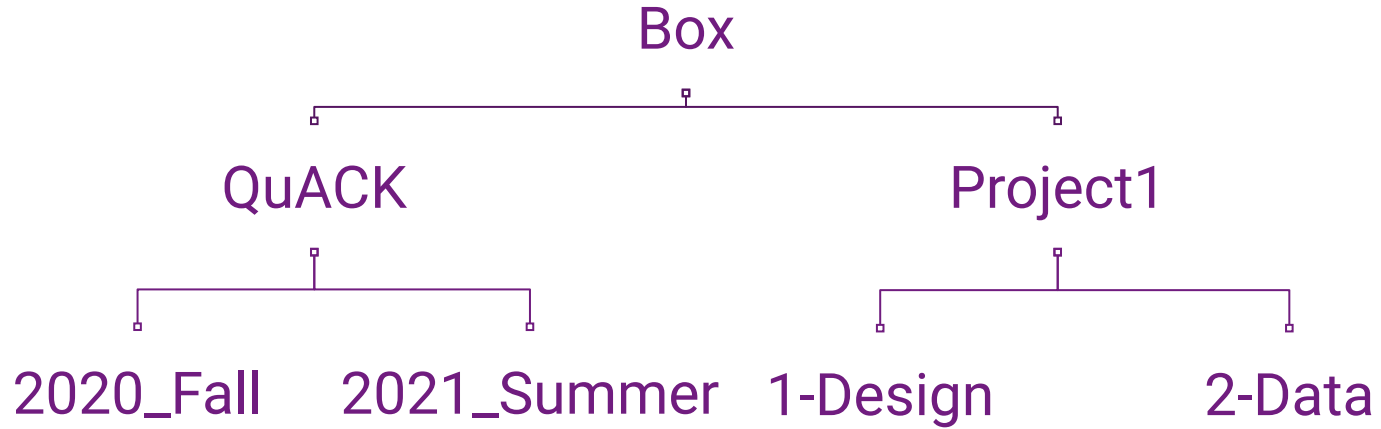
So do folders!
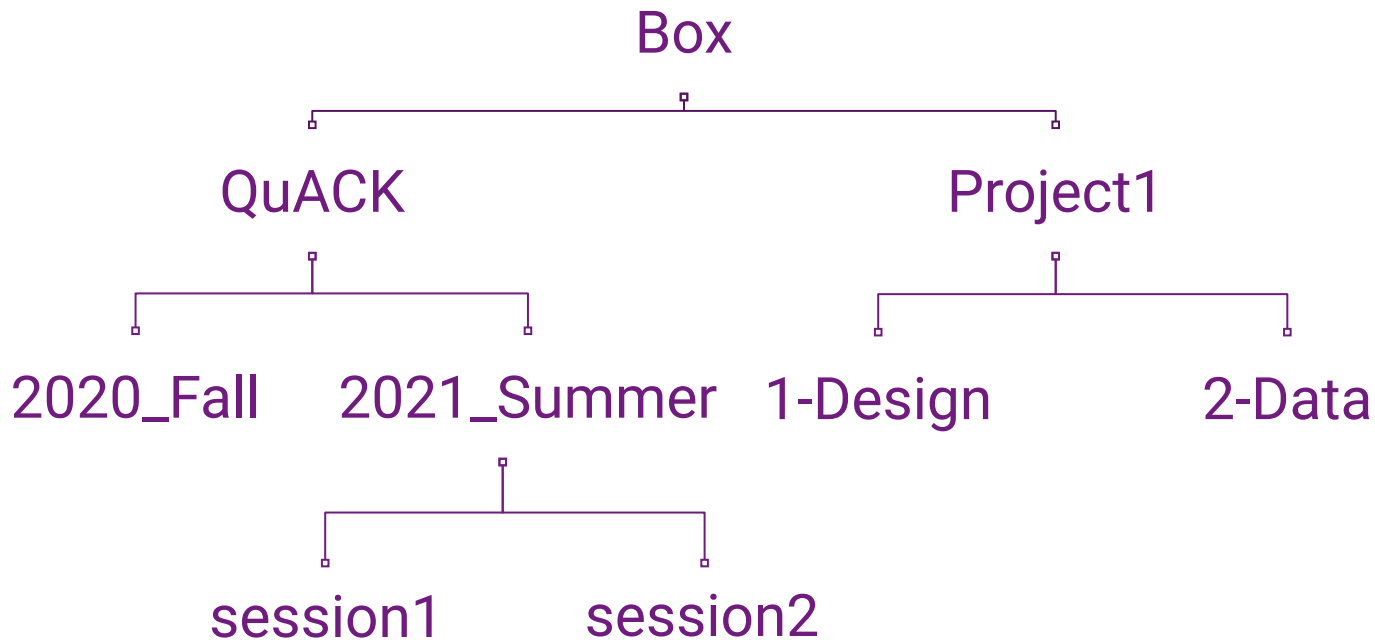
Let's look at my file structure more closely

Box

QuACK

2020_Fall   2021_Summer

session1   session2
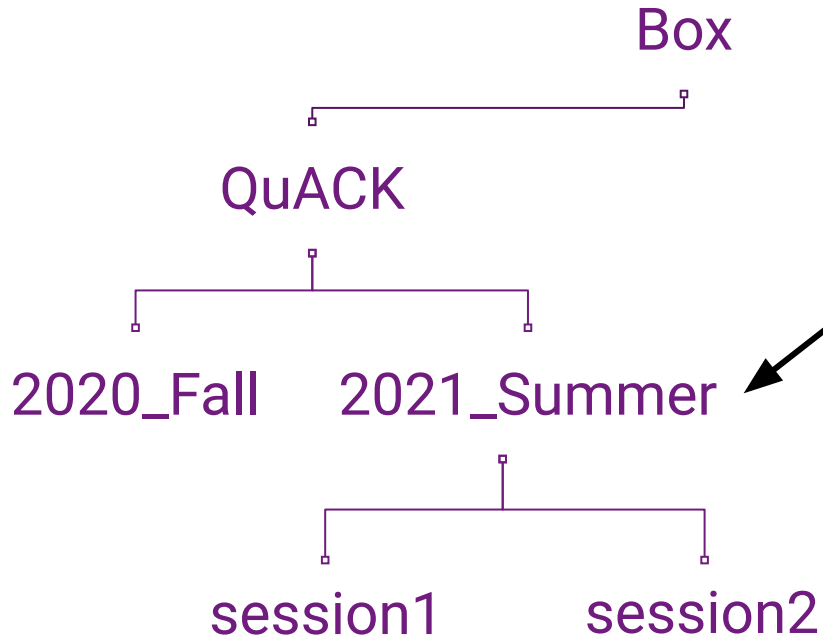
I've chosen to keep all my data files for the bootcamp (i.e., penguins.csv and covid_attitudes.csv) at this level

} I have a folder for each week to organize the scripts for that week

Berkeley
UNIVERSITY OF CALIFORNIA

Different ways to organize your files

What I did:
- 2021_Summer (folder)
  - *penguin.csv*
  - *covid_attitudes.csv*
  - session1 (folder)
    - *scripts*
  - session2 (folder)
    - *scripts*

Another option:
- 2021_Summer
  - *penguin.csv*
  - *Covid_attitudes.csv*
  - *session1_script.R*
  - *session2_script.R*

^^ No separate folders for each session

Working directory and loading data

- Organization is up to you (and ask your grad student for how they do it!)

- Whatever you do though, you need to understand it when you're loading your data!

- R has to know where to look for the data file, or it won't be able to read it in

Group Activity

Groups:

https://docs.google.com/spreadsheets/d/1QrJg_aJkbehIQXNdAX10GCc2s0cmly SKZMYx4UL7vjA/edit?usp=sharing

Instructions (also editable in downloaded materials):

https://docs.google.com/document/d/1wv6UzVqeqHJceIpOtxMmCyWmnP0zyT 7w2NSLD7NPSOQ/edit?usp=sharing

Help doc:

https://docs.google.com/document/d/1IesSkLxwvE3uUYDp6Qd3LZcYvGlwUwe GFQLhM7_YIlY/edit?usp=sharing

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!