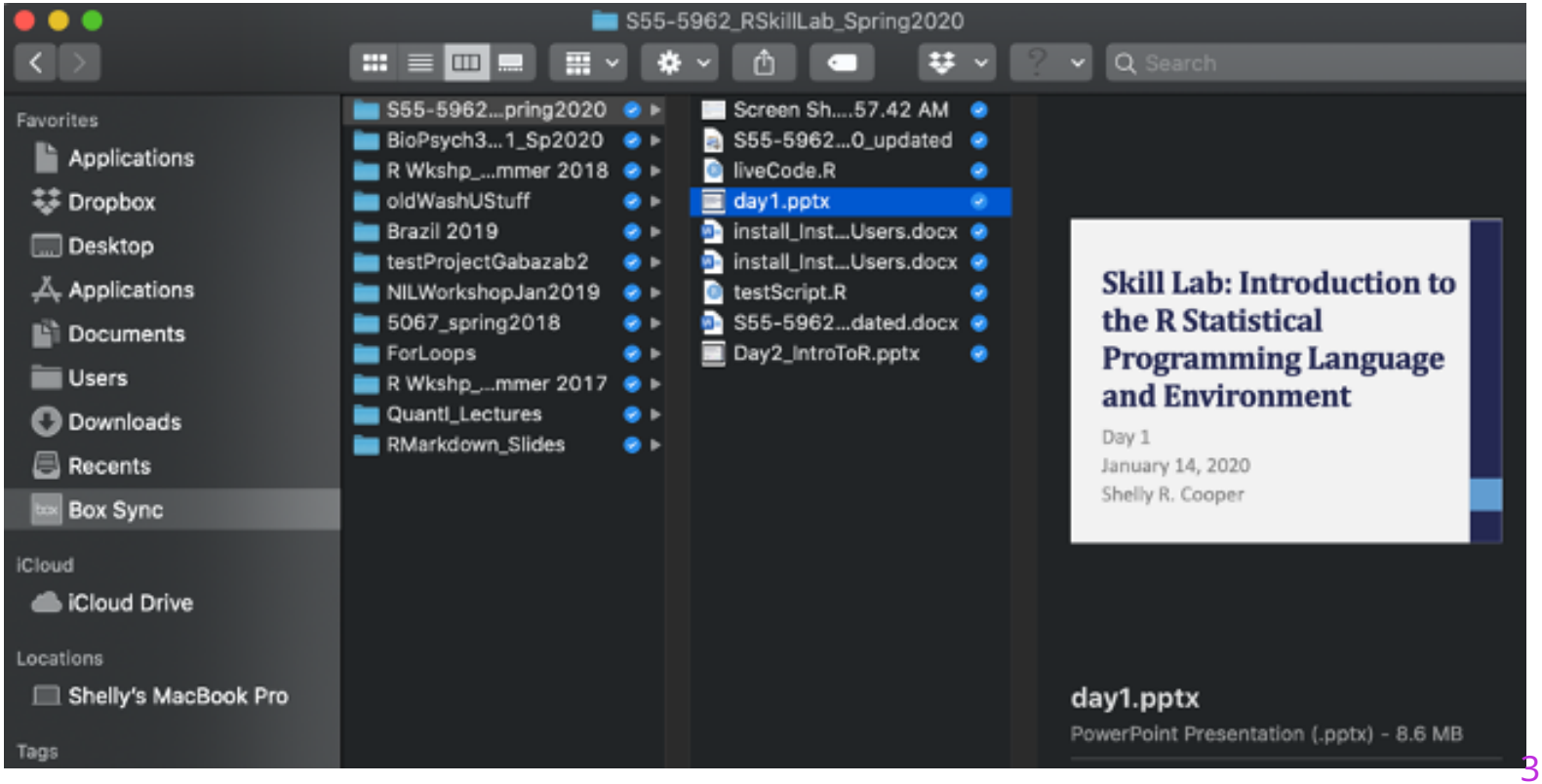
Loading Files

Loading & Organizing

- BE. ORGANIZED.
- It's hard to load in files if you don't know where they are
- We can use packages and tools to our advantage when we are organized
- Less typing
- Being organized will spark joy

Directories

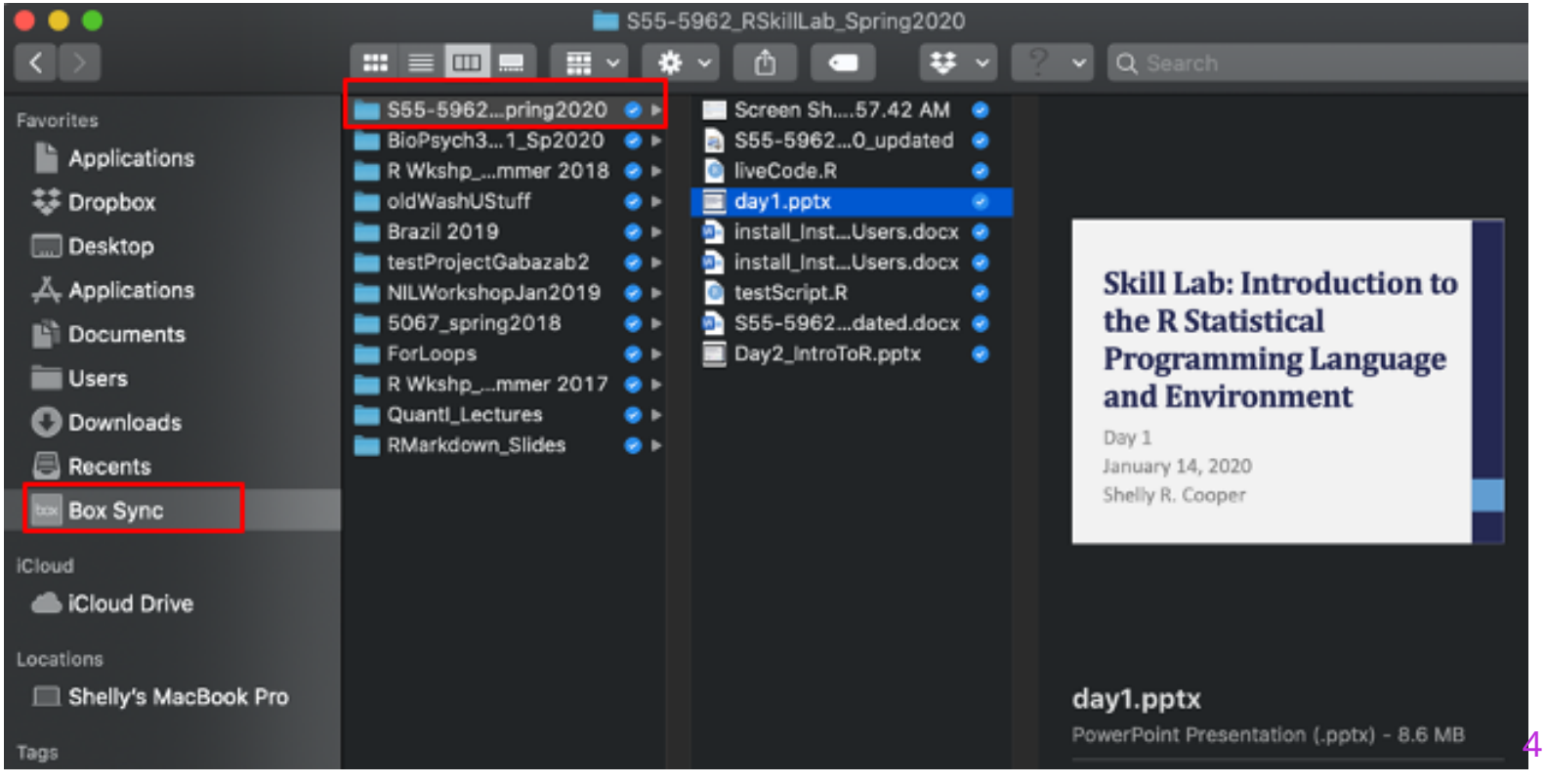
Your computer is made up of a series of folders



File paths

These are the instructions that tell the computer where to find your file. What series of folders does the computer need to look to find your stuff?

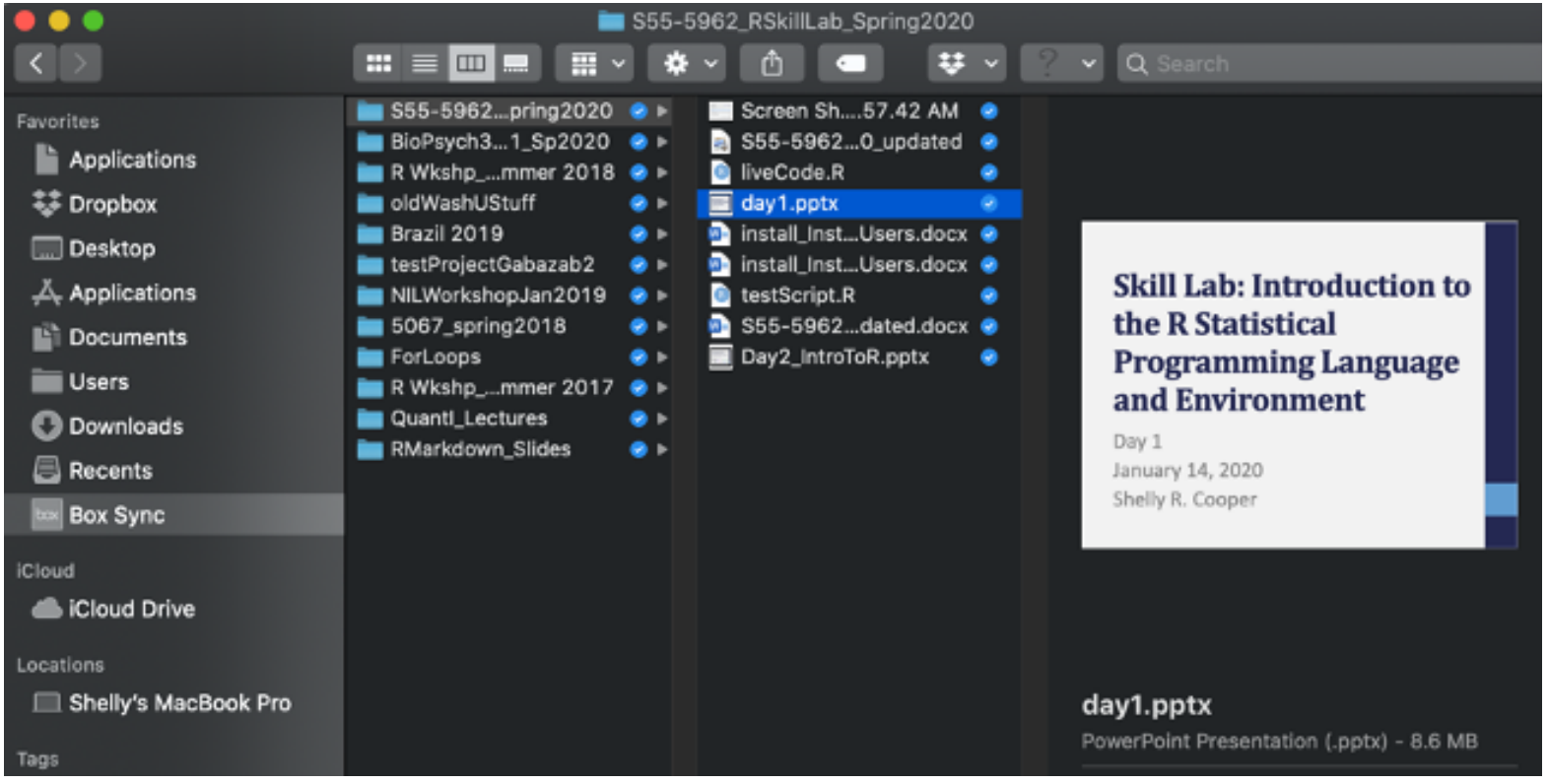
`/Users/Coop/Box Sync/S55-5962_RSkillLab_Spring2020`



File paths

In order to get the actual file, include the name in the file path

`/Users/Coop/Box Sync/S55-5962_RSkillLab_Spring2020/day1.pptx`



R is lazy!

Working Directory

- Where R is going to *look for* files
- Where R is going to *save* files

Working directory

How do you know your working directory?

- `getwd()`

How do you change your working directory?

- `setwd("/your/path/goes/here")`
- Note the quotes!
- *HINT: press tab within the quotes and see what happens!*

An Alternative: RProjects

Getting and setting your working directory can be a pain in the @$\$

- What happens if you reorganize your computer and you want to move the files?

`RProjects` provides a nice alternative with several added benefits

1. It syncs to Github. Excellent for version control and open science!
2. Your project is it's own contained ecosystem. If you move it on your computer, it doesn't matter. No need to get/set your working directory.
3. Easy to look for files within that project (rather than the entire computer)

We are going to make this together next Wednesday!

.R files

- Aka **scripts**
- Text files
- Contain the code that you've written
- (Equivalent to syntax files in SPSS)

Why use them?

- Keep track of what functions you use
- Save only the commands/functions/progress that is useful
- Make notes to yourself!
  - `# Updated code for R class!`
  - `# reliability estimates for depression scale`
  - `# scatter plot for BMI predicting diabetes diagnosis`
- Share your analyses with collaborators and readers

Other file types

- `.RProject` -- not where you would write any form of code -- more to come!
- `.Rmd` -- aka "RMarkdown" or "RNotebook"; will talk about for reproducibility!
- **Your data!**

Your data

Original data files

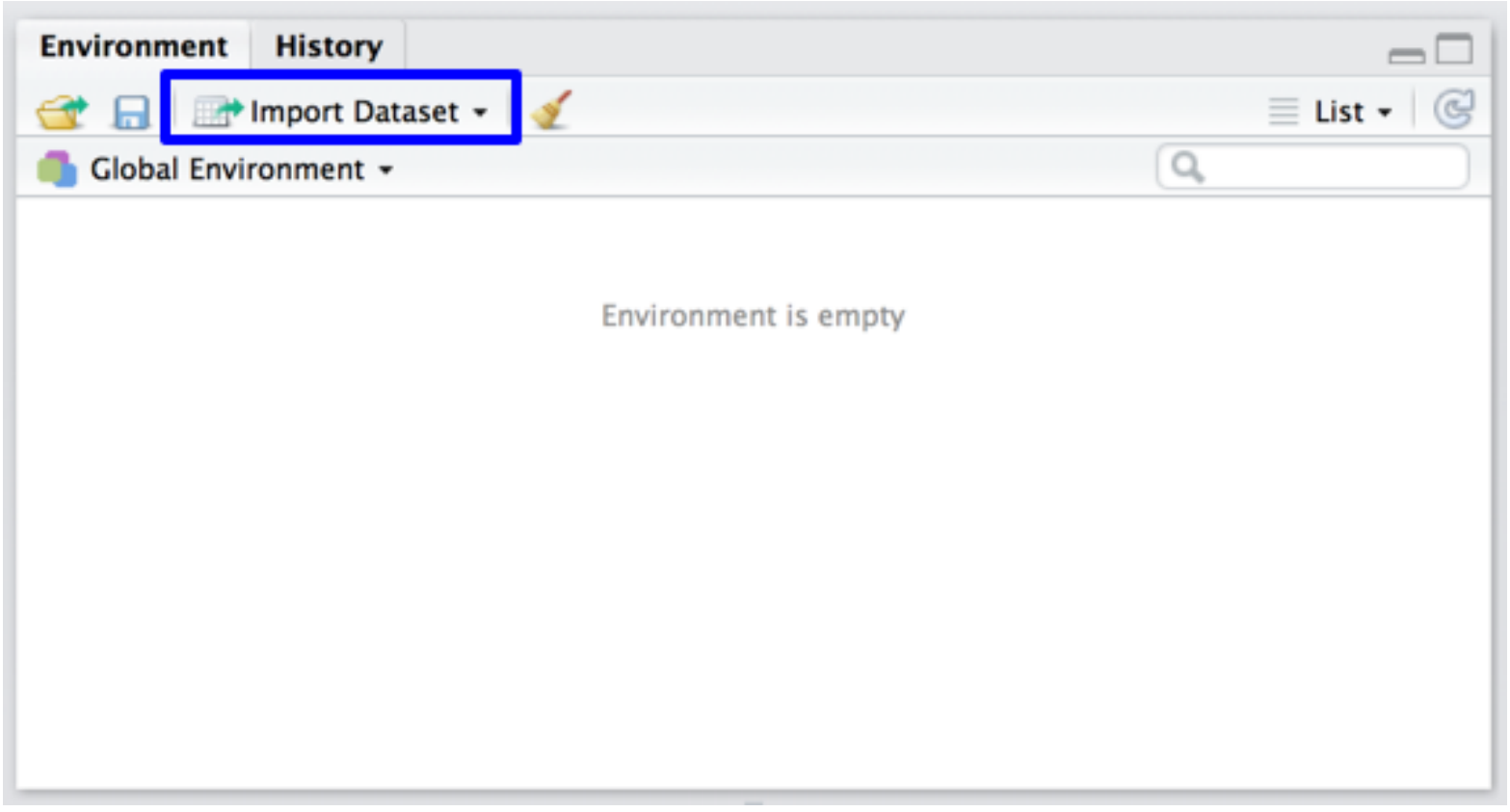
- Most of the time, these will either be `.csv` or `.txt`, depending on how you collect the data

These are *NOT* altered by R! *(different from SPSS)*

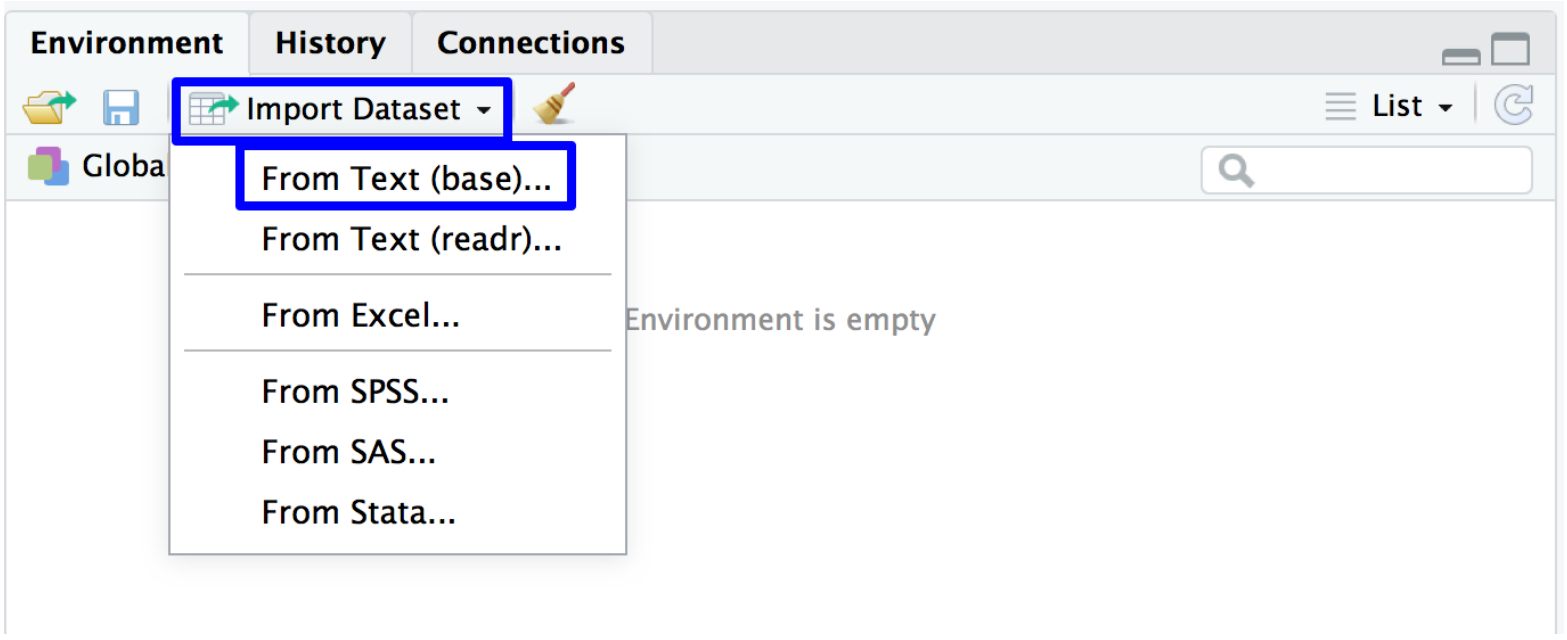
If your data is not one of these two formats, don't worry! R can do a lot of stuff!

We will work with `.csv` to keep things simple.

Loading .csv files



Loading .csv files



Loading .csv files

Import Dataset

Name: midus

Encoding: Automatic
Heading: Yes / No
Row names: Automatic
Separator: Comma
Decimal: Period
Quote: Double quote (")
Comment: None
na.strings: NA
Strings as factors

Input File

```
"ID","sex","age","BMI","physical_health_self","mental_health
10001,"Male",61,26.263,2,4,42,7.75,5.5,"No","No"
10002,"Male",69,24.077,5,5,34,8.25,6,"No","Yes"
10005,"Female",80,NA,4,4,49,9.333,4,"No","No"
10006,"Female",60,NA,3,3,NA,NA,NA,"No","Yes"
10010,"Male",55,NA,4,3,28,8.25,8,"No","Yes"
10011,"Female",52,25.991,5,4,41,7,5.5,"No","No"
10014,"Male",57,NA,3,3,NA,NA,NA,"No","No"
10015,"Female",53,32.121,3,3,31,7.375,6,"No","Yes"
10017,"Male",46,NA,3,4,NA,NA,NA,"No","No"
10018,"Male",49,22.499,4,4,41,8.5,6,"No","No"
10019,"Male",51,29.987,4,5,38,7.625,4.5,"No","No"
10020,"Female",56,NA,3,3,NA,NA,NA,"Yes","Yes"
```

Data Frame

ID	sex	age	BMI	physical_health_self	mental_l
10001	Male	61	26.263	2	4
10002	Male	69	24.077	5	5
10005	Female	80	NA	4	4
10006	Female	60	NA	3	3
10010	Male	55	NA	4	3
10011	Female	52	25.991	5	4
10014	Male	57	NA	3	3
10015	Female	53	32.121	3	3
10017	Male	46	NA	3	4

Loading .csv files

The following line of code will appear in your console:

```
midus <- read.csv("~/Desktop/rSkillLab/midus.csv")
```

If you do *not* use a RProject, then you MUST copy/paste this line of code into your script (.R) file!

A note about Excel

- Excel is owned by Microsoft
- Microsoft is not a non-profit; they make a stupid amount of money
- When programming languages were developed, Microsoft didn't play nicely with things that weren't also owned by Microsoft
- What does this mean for you?
  - If you are working with a data file, it is better to save it as a `.csv` or `.txt` rather than `.xls` or `.xlsx`
  - It *used* to be really hard to get your data into `R` if it was a `.xls` or `.xlsx` format. It's now a lot easier. So it doesn't matter much, but if you have a say in the matter, change it.
  - Color coding/highlighting in Excel doesn't translate to `R`. Stop using it for data. Story time.

Typical workflow in R

1. Open a script (new or existing)

2. Prepare to run analyses:

    - Set your working directory (if not using RProject)
    - Load your data
    - Load any packages you might want to use in the analyses

3. Write code/run analyses

4. Save your script!

    - Make sure that this includes the code to open your `.csv` from your **Dropbox/Box/Github** etc.
    - Again, note: `R` does't change the original data file!

Typical format of .R file