Reproducible Workflow

How to stop hard-coding your working directory!

Katherine Hoffman, M.S.
Biostatistics Computing Club

June 4, 2019

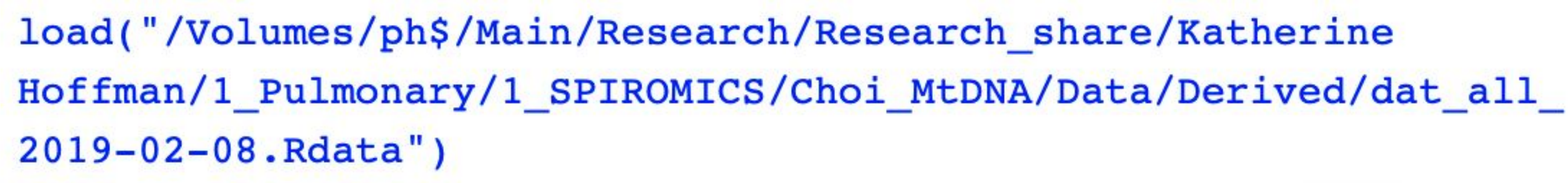
Weill Cornell Medicine

First things first:

My code examples will be structured using the folder hierarchy Liz presented in our first computing club session

- Project_Name/
    - Code/
    - Data/
        - Raw/
        - Derived/
    - Output/
    - Reports/
    - Notes/

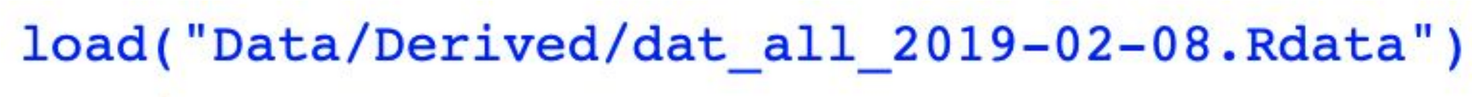
What’s a “hard-coded” working directory?



```
load("/Volumes/ph$/Main/Research/Research_share/Katherine
Hoffman/1_Pulmonary/1_SPIROMICS/Choi_MtDNA/Data/Derived/dat_all_
2019-02-08.Rdata")
```

🤯

What’s a “non hard-coded” working directory?

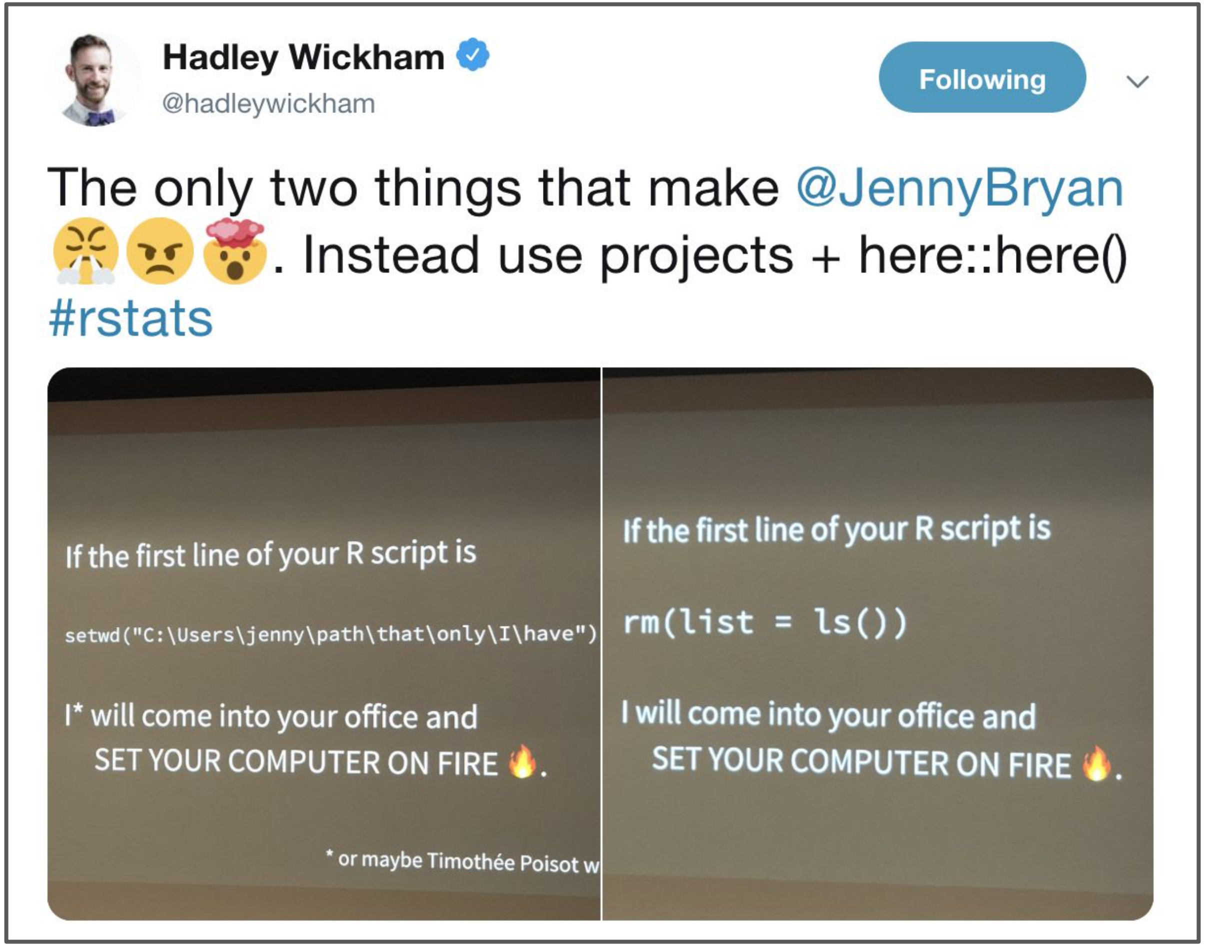


```
load("Data/Derived/dat_all_2019-02-08.Rdata")
```

😊

What’s the big deal?

- **A personal problem:** this working directory will change if you...
    - Switch back and forth from PC to Mac
    - Are unable to work on the Research Share drive due to poor internet
    - Move work from your personal drive to the shared drive
    - Ever want to change your project organization/folder hierarchy
- **A collaboration problem:** If you give someone code to review or pass off an entire project, they’ll have to change all your working directories
- **A scary problem:** If you ever meet Jenny Bryan in real life...😳



Hadley Wickham

@hadleywickham

The only two things that make @JennyBryan 😤😠🤯. Instead use projects + here::here() #rstats

If the first line of your R script is

```
setwd("C:\Users\jenny\path\that\only\I\have")
```

I* will come into your office and
SET YOUR COMPUTER ON FIRE 🔥.

* or maybe Timothée Poisot w

If the first line of your R script is

```
rm(list = ls())
```

I will come into your office and
SET YOUR COMPUTER ON FIRE 🔥.

What are we going to do about it?

We’ll do what Jenny and Hadley recommend, of course!

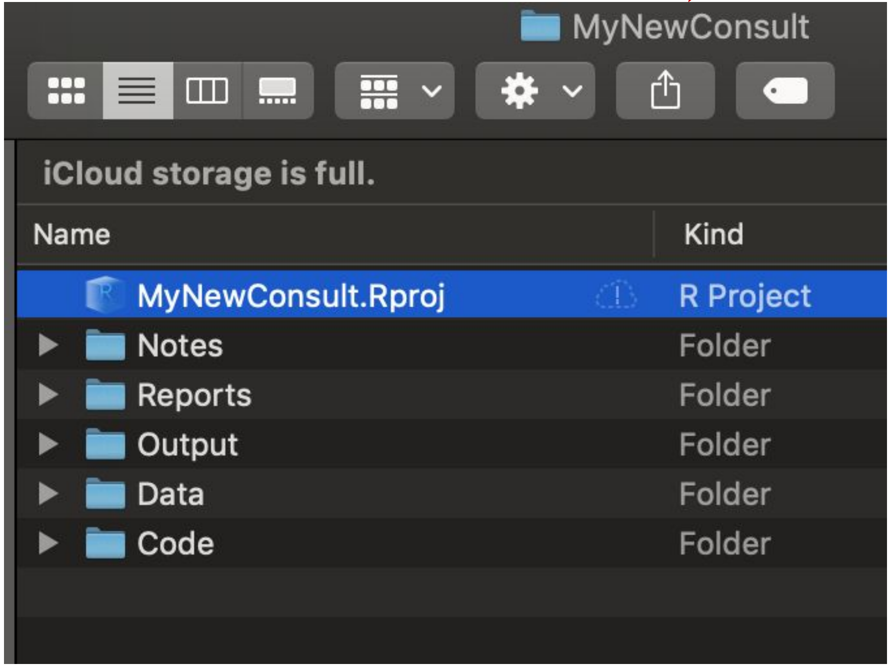
1. Implement R Projects

2. Utilize the `here` package

What the heck is an R Project?

- An R Project is an object that you put in the “home directory” for one individual project
- When clicked, it opens a new RStudio window with the working directory set to wherever the .RProj object is
- When working in that R Project, your RStudio session has its own environment so you don't have to worry about files getting crossed over from one project to another
- You don't have to reference your entire working directory every time you want to read something in/export out!

Some folder that contains your code, data, etc. for one project



MyNewConsult

iCloud storage is full.

Name	Kind
MyNewConsult.Rproj	R Project
Notes	Folder
Reports	Folder
Output	Folder
Data	Folder
Code	Folder

Keep your .Rproj at the top of your folder hierarchy

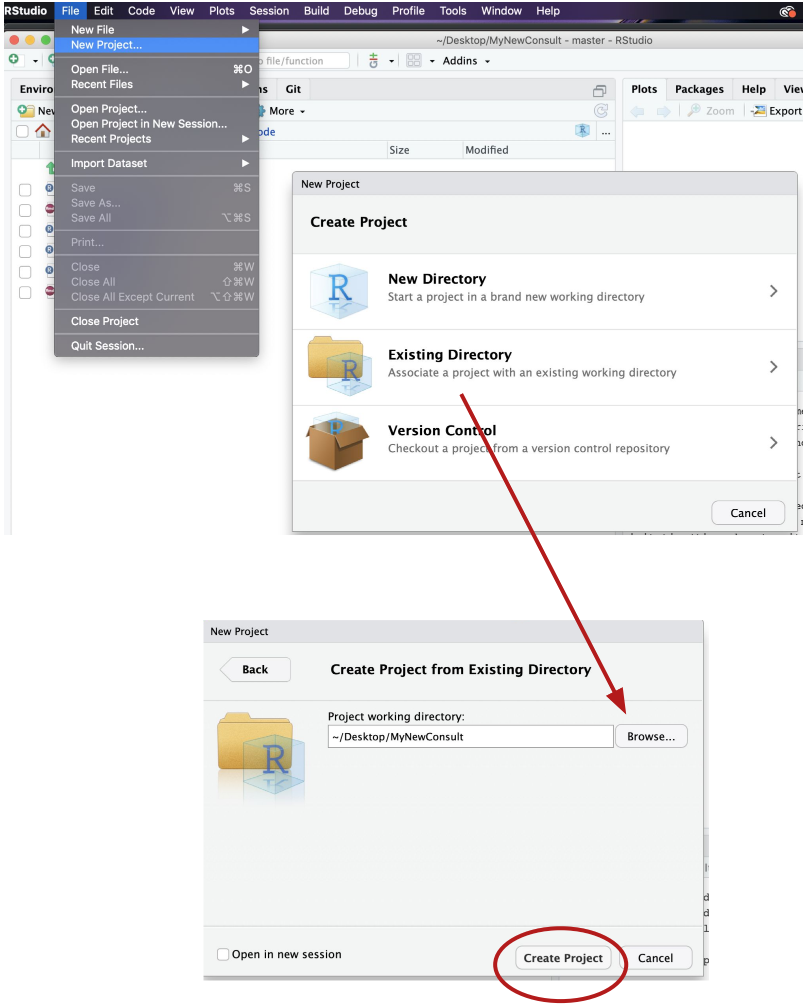
Let’s make our first R Project!

At this point it will be easiest to follow along if you download this folder:

https://github.com/hoffmakl/Rproj-here-tutorial

Let’s make our first R Project!

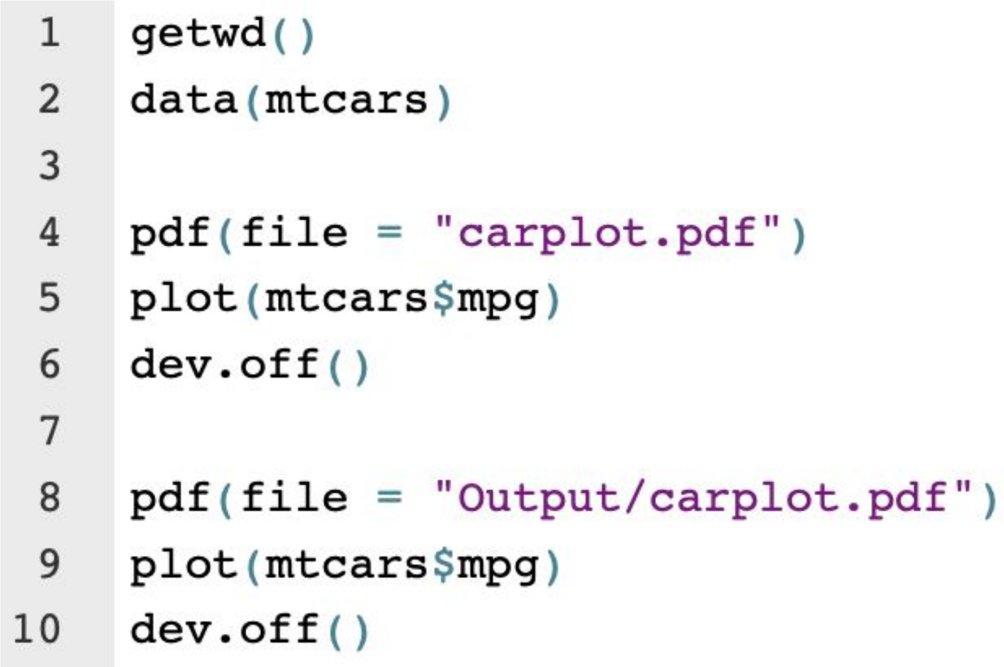
1. Open RStudio
2. Click **File --> New Project...**
3. Choose **Existing Directory**
4. Click **Browse...** to find a folder containing your code/data for one project (such as a consult)
5. **Open** that folder, and choose **Create Project**
6. A new RStudio window will open
7. You can look at the top of that RStudio window to see that your home directory is now pointing to wherever you placed that Project 👌


RStudio File Edit Code View Plots Session Build Debug Profile Tools Window Help
~/Desktop/MyNewConsult - master - RStudio
New File
New Project...
Open File... ⌘O
Recent Files
Open Project...
Open Project in New Session...
Recent Projects
Import Dataset
Save ⌘S
Save As...
Save All ⌥⌘S
Print...
Close ⌘W
Close All ⇧⌘W
Close All Except Current ⌥⇧⌘W
Close Project
Quit Session...
Git
More
Size
Modified
Plots Packages Help Viewer
Zoom
Export
Addins
New Project
Create Project
New Directory
Start a project in a brand new working directory
Existing Directory
Associate a project with an existing working directory
Version Control
Checkout a project from a version control repository
Cancel
New Project
Back
Create Project from Existing Directory
Browse...
□ Open in new session
Create Project
Cancel

Project working directory: ~/Desktop/MyNewConsult

Let’s test our new R Project out...

1. Check out your working directory with **`getwd()`**
2. Make a plot and save it, then try saving it in an “Output” folder
3. Move the entire folder containing your R Project to a new location and try again

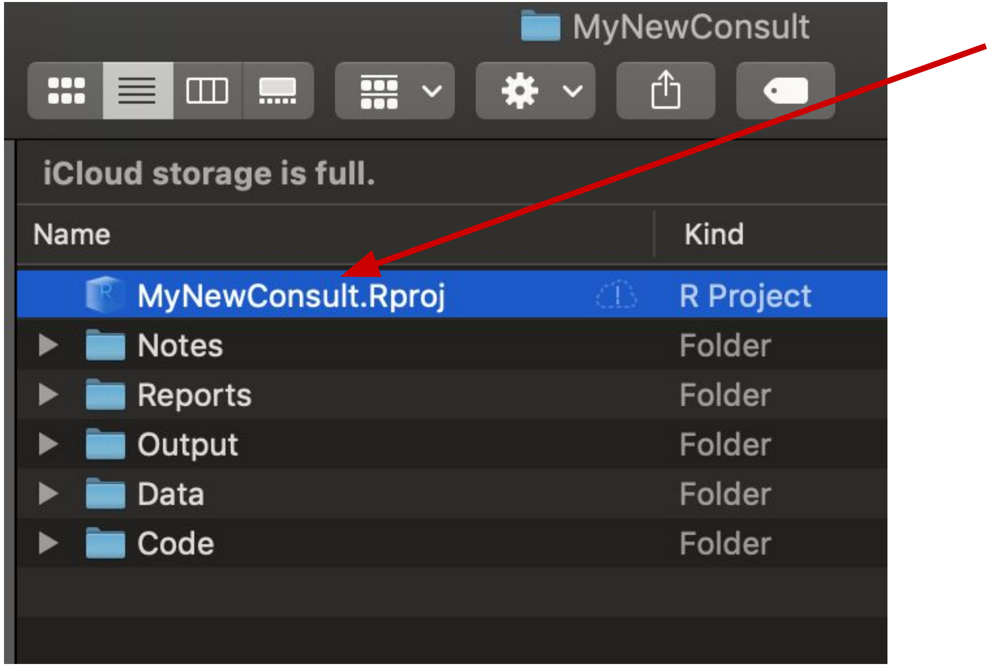


```r
getwd()
data(mtcars)

pdf(file = "carplot.pdf")
plot(mtcars$mpg)
dev.off()

pdf(file = "Output/carplot.pdf")
plot(mtcars$mpg)
dev.off()
```

Some extra info...



MyNewConsult

iCloud storage is full.

Name	Kind
MyNewConsult.Rproj	R Project
Notes	Folder
Reports	Folder
Output	Folder
Data	Folder
Code	Folder

The actual name of your .Rproj doesn’t matter at all

If you delete your .Rproj, nothing will happen to your data/code/etc.

What else!

- Version control with Git by automatically starting a Git repository when you make a new R Project
- Set R Project options (like never save R history)
- Use **`packrat`** to keep track of which package versions you are using

How could projects possibly get any better?

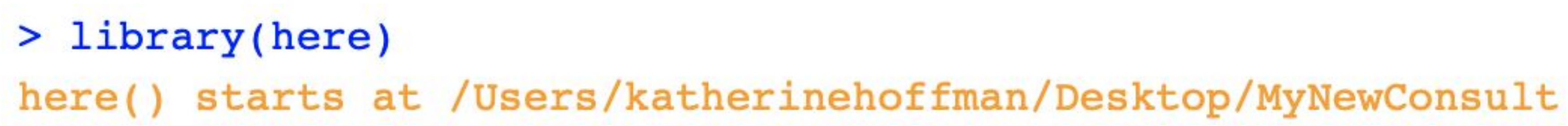
The `here` package!

Animation by Allison Horst
@allison_horst

What the heck is the **`here`** package?

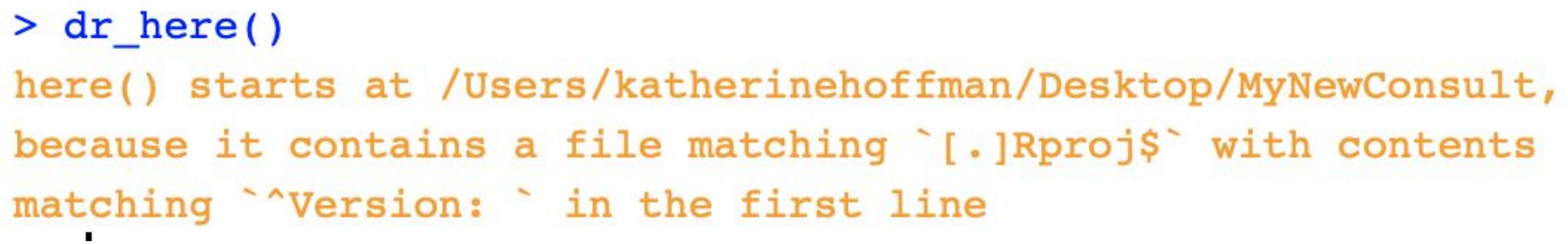
- A package with one main function: **`here()`**
- **`here()`** is a wrapper function that helps establish file paths by searching for a file path that *should* be your home or “root” directory
    - See the full heuristics for that decision here
- All we need to know for now… **If you have an R Project, it will always locate that R Project as the root directory**
- This is super helpful for RMarkdown reports, because an .Rmd file’s root directory is wherever that .Rmd file is located

`here::here()` in action



```
> library(here)
here() starts at /Users/katherinehoffman/Desktop/MyNewConsult
```

Loading the **`here`** package returns a message letting you know what your working directory has been recognized to be (if you've started an R Project, that's where it defaults!) You can double check that with...



```
> dr_here()
here() starts at /Users/katherinehoffman/Desktop/MyNewConsult,
because it contains a file matching `[.]Rproj$` with contents
matching `^Version: ` in the first line
```

`here::here()` in action

Let’s look at examples 1, 2, and 3 in Code/

Resources

https://www.tidyverse.org/articles/2017/12/workflow-vs-script/

https://github.com/jennybc/here_here/blob/master/README.md