

Macfarlane Lab

Gregory Macfarlane

1/17/23

Table of contents

Welcome!	3
1 Onboarding	4
1.1 Working hours	4
1.2 Communication	5
1.3 Wages	5
2 Project Workflow	6
2.1 Motivation	6
2.2 R Projects	8
2.3 Git and GitHub	9
2.4 Targets	9
3 Summary	10
References	11

Welcome!

You have been hired into my lab, or are about to be hired in my lab. This website is a collection of resources that my students use to get trained and learn how to do things. It is also a helpful thing for me to remember all the things I have to do.

1 Onboarding

The process to onboard in my lab is:

1. I decide to hire you, and you agree to work for me.
2. I send a message to Jolene Johnson in the department office to begin the process to hire you.
3. Join the [BYU Transportation Lab Slack workspace](#)
4. You complete any necessary university hiring forms; Jolene will email you with these instructions.
5. You complete Y-time and desk worker safety training, and any other safety trainings that may be necessary.
6. I send a message to Julie Brinton in the department office to give you after-hours access to the building, the transportation lab, and your desk as appropriate.
 - I will tell you when your access expires; please mark this date down and ask me to extend it if you are still working for me
7. You begin working

1.1 Working hours

When I hire you I will state expectations for a weekly commitment. This is typically:

- Undergraduates: 5-10 hours per week
- Graduate student academic year: 10-20 hours per week
- Graduate student Spring/Summer: 20-30 hours per week

You may work whatever hours you choose, with the exception of any weekly project team meetings or project sponsor calls. That said, I strongly recommend you schedule a regular weekly time to dedicate to the research projects you are working on. Other commitments might easily crowd out the time for your research if you do not prioritize the time. Also, I want to pay you! But I can't if you don't put in time.

I strongly recommend that students who work for me as a research assistant quit any other research assistance, internship, or other external employment.

1.2 Communication

I use Slack for almost all lab-related communication. Please install the Slack app on your phone and your computer. I expect you to respond to messages within a reasonable amount of time, but I never expect you to respond while you are with your families or other times when you are not working.

If you would like to reach me with a message, use slack.

If you would like to meet with me in person about a brief issue, please schedule an [office hour appointment using Calendly](#). If you would like to schedule a longer meeting with me, please schedule a [research meeting appointment using Calendly](#).

1.3 Wages

Students in my lab have their wages determined by skill and seniority, following the basic outlines of the [BYU hourly pay scale](#). Each semester I will evaluate student work and give raises as appropriate.

If you believe you merit a raise or there are needs in your family that might necessitate a higher wage, please speak to me. I'll do what I can.

2 Project Workflow

Most of my academic work uses R and R markdown (or more recently, Quarto) as a backbone.

2.1 Motivation

The goal of this workflow is to allow you to work effectively in my lab, and to ensure that I and future students (including yourself!) know where things are and can rebuild them if necessary. For example, a peer reviewer may ask for changes to the analysis many months after you are gone. If you follow this workflow, the chances of me being able to productively use your work increase.

A workflow that you have probably used in the past consists of:

- Keep data in a spreadsheet (Excel)
- Do your analysis in the same spreadsheet
- Copy a table or numbers from a spreadsheet into a document (Word)
- Write about the analysis in Word

This workflow has its benefits, mainly simplicity. Excel definitely has its place. But there are *lots* of drawbacks.

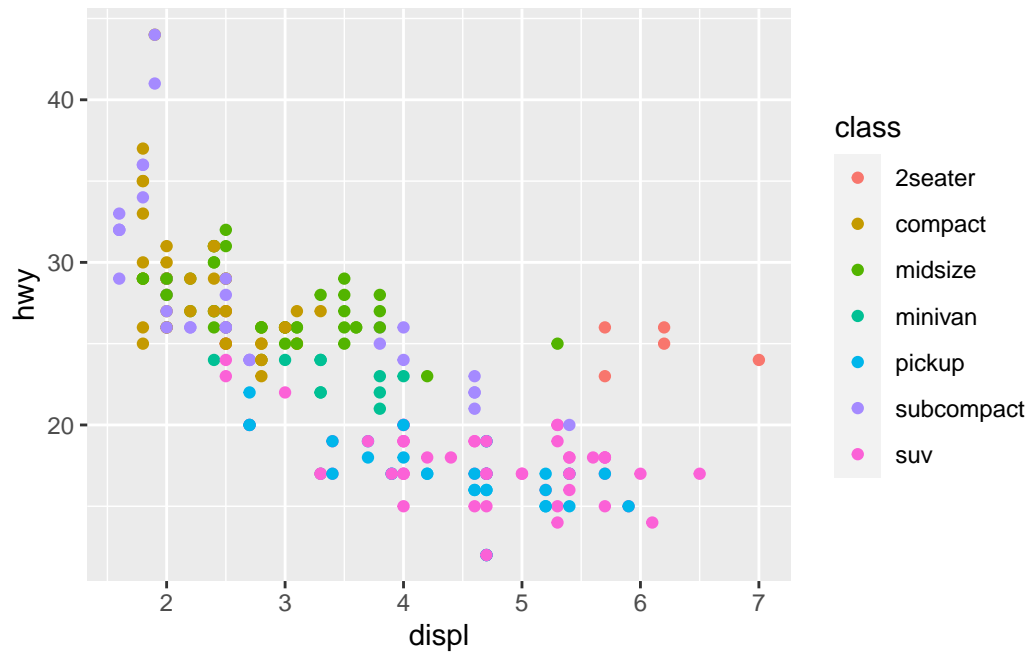
First, spreadsheets tend to hide formulae or make them very difficult to read. For example, [JP Morgan Chase bank lost almost \\$6 Billion when a trader believed a formula in an excel spreadsheet](#) that he couldn't see the error in, because the formula wasn't shown. It would be better to have clearly written methods and analysis, than to just show the numbers that go in that analysis.

Second, this method can rely on lots of manual tasks: push this button, copy this figure, update this number. Doing it once is easy, but what if I ask you to go back and change something? Will you remember every table that depends on that calculation? Or what if you leave and another student has to figure out what you did? It would be better to have a reproducible document that re-builds itself whenever the analysis changes.

This is the problem that R and R markdown solves. For example, I can embed R code into this document. I can load a library, read a dataset, and create a figure all at once.

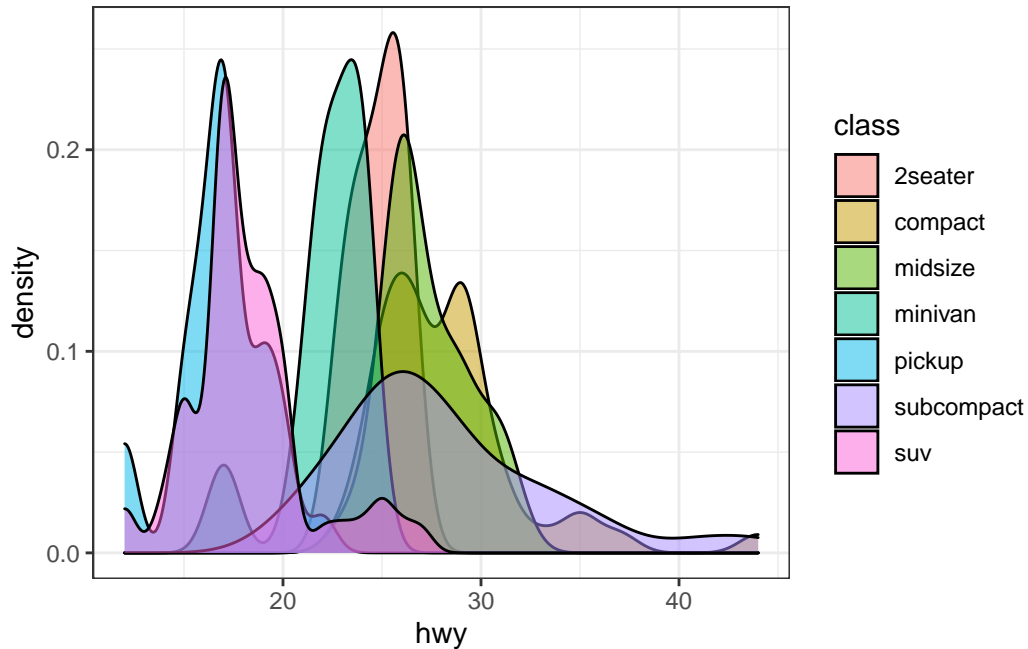
```
library(ggplot2)
```

```
ggplot(mpg, aes(displ, hwy, colour = class)) +  
  geom_point()
```



And if I tell you that I'd rather see a different figure, you can just change the code and it will print something different

```
ggplot(mpg, aes(hwy, fill = class)) +  
  geom_density(alpha = 0.5) + theme_bw()
```



So, how do you make this happen?

2.2 R Projects

The basis of every project — a paper, thesis, or report — should be an R project. An R project is a folder associated with an `.Rproj` file that can be created and opened in R Studio. For an introduction to installing R and Rstudio, there is a [tutorial at DataCamp](#).

The R project folder should be organized into these subfolders as follows:

```
<project>/  
- README.md  
- <project.Rproj>  
- data/  
- R/  
- py/  
- ....
```

- `README.md` is a markdown file that describes what the project is about. What is the central question, who is the project sponsor, and what other data resources might need to be referenced in the project.

- `data/` is the folder where you place all of your input or intermediate data. There may be subfolders in this folder. Preference is for plain text (csv or json) data over data stored in spreadsheets or binary formats. If you get data from the internet, it might be better to write code to download the data instead of storing the downloaded data. That would be more reproducible.
- `R/` is a folder where you write your R scripts. Each script should be clearly named, e.g., `data_cleaning.R`.
- `py/` is a folder where you write your python scripts. Each script should be clearly named, e.g., `data_cleaning.py`.

Other folders might be useful, but should be carefully named and organized.¹ Additionally, it might be a good idea to put `README` file in some of the sub-folders if the information in them might need more explanation.

The LinkedIn Learning Course [R: Essential Training: Wrangling and Visualizing Data](#) shows how to implement this

For an introduction to installing R and R Studio, there is a [tutorial at DataCamp](#). Follow along with the tutorial if you are new to R and R Studio.

If you are curious about projects, [there is a helpful R Studio resource page](#) with instructions on how they work.

2.3 Git and GitHub

2.4 Targets

¹A more detailed and robust description of an R project folder is [here](#).

3 Summary

In summary, this book has no content whatsoever.

1 + **1**

[1] 2

References