



STA 235 - Wrapping up

Spring 2021

McCombs School of Business, UT Austin

We have seen a lot of topics this semester

Regression

Regression

Potential Outcomes framework

Regression

Potential Outcomes framework

Randomized Controlled Trials

Regression

Potential Outcomes framework

Randomized Controlled Trials

Observational studies

Regression

Model selection

Potential Outcomes framework

Randomized Controlled Trials

Observational studies

Regression

Model selection

Potential Outcomes framework

Regularization

Randomized Controlled Trials

Observational studies

Regression

Model selection

Potential Outcomes framework

Randomized Controlled Trials

Regularization

Observational studies

Prediction

**How do we bring everything
together?**

Case Study

The use of shared bikes

- Q1: How to predict demand?
- Q2: How to incentivize use?



Q1: How to predict demand?

- What type of problem is it?
- How would you approach this problem?
- What is your outcome variable?
- What data would you ask for?
 - Think about granularity (level), time scope, variables, other data sources.

Let's look at the data

```
bikedc <- read.csv("https://raw.githubusercontent.com/maibennett/sta235/main/exampleSite/content/Cla  
head(bikedc)
```

```
##           datetime season holiday workingday weather temp  atemp humidity  
## 1 2011-01-01 00:00:00      1      0           0      1 9.84 14.395      81  
## 2 2011-01-01 01:00:00      1      0           0      1 9.02 13.635      80  
## 3 2011-01-01 02:00:00      1      0           0      1 9.02 13.635      80  
## 4 2011-01-01 03:00:00      1      0           0      1 9.84 14.395      75  
## 5 2011-01-01 04:00:00      1      0           0      1 9.84 14.395      75  
## 6 2011-01-01 05:00:00      1      0           0      2 9.84 12.880      75  
##    windspeed casual registered count  
## 1      0.0000      3         13     16  
## 2      0.0000      8         32     40  
## 3      0.0000      5         27     32  
## 4      0.0000      3         10     13  
## 5      0.0000      0          1      1  
## 6      6.0032      0          1      1
```

Let's look at the data

- Would you **transform** some of this data?
- What **prediction method** would you use?

```
head(bikedc)
```

```
##           datetime season holiday workingday weather temp  atemp humidity
## 1 2011-01-01 00:00:00      1      0          0      1 9.84 14.395      81
## 2 2011-01-01 01:00:00      1      0          0      1 9.02 13.635      80
## 3 2011-01-01 02:00:00      1      0          0      1 9.02 13.635      80
## 4 2011-01-01 03:00:00      1      0          0      1 9.84 14.395      75
## 5 2011-01-01 04:00:00      1      0          0      1 9.84 14.395      75
## 6 2011-01-01 05:00:00      1      0          0      2 9.84 12.880      75
##  windspeed casual registered count
## 1      0.0000      3         13     16
## 2      0.0000      8         32     40
## 3      0.0000      5         27     32
## 4      0.0000      3         10     13
## 5      0.0000      0          1      1
## 6      6.0032      0          1      1
```

Q2: How to incentivize use?

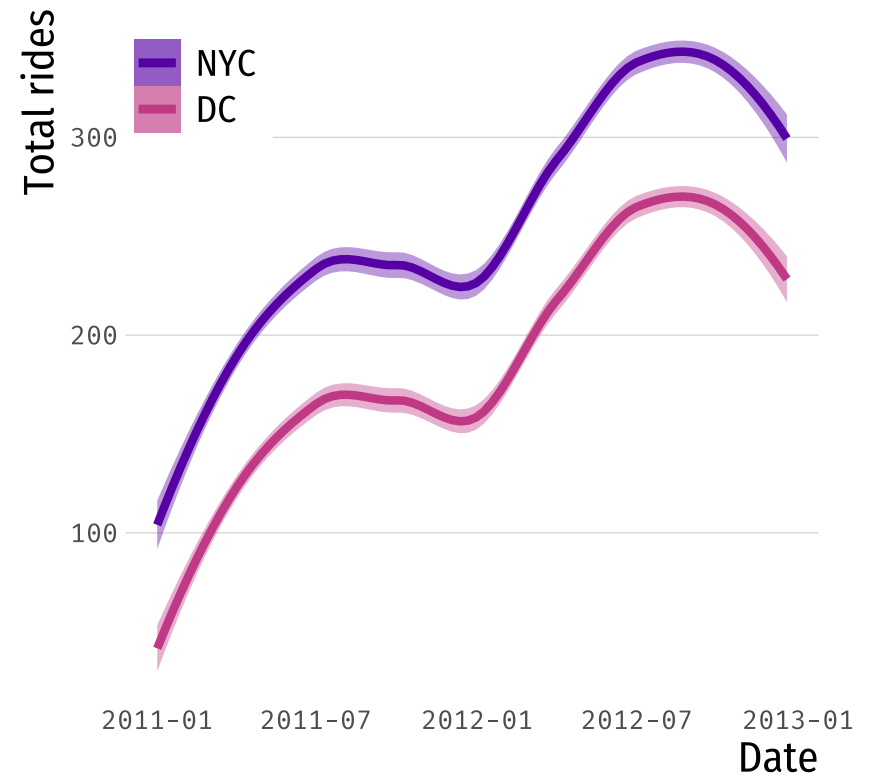
- What type of problem is it?
- How would you approach this problem?
- What is your outcome variable?
- What data would you ask for?
 - Think about granularity (level), time scope, variables, other data sources.

Q2: How to incentivize use?

- If you only had access to the previous data structure, **what intervention could you design?**
- Imagine now that bike share is present in **two cities:** DC and NYC. Would this change your approach?
- Now imagine that in 2012, **NYC modified their prices for casual and registered riders**. What approach could you use now?

Review Material

```
bikenyc <- read.csv("https://raw.githubusercontent.com")  
  
ggplot(bikenyc, aes(x = date, y = count)) +  
  geom_smooth(method = "loess", color = "blue") +  
  geom_smooth(data = bikedc, aes(x = date, y = count),  
             method = "loess", color = "red")
```



Let's wrap things up

Objectives of this course

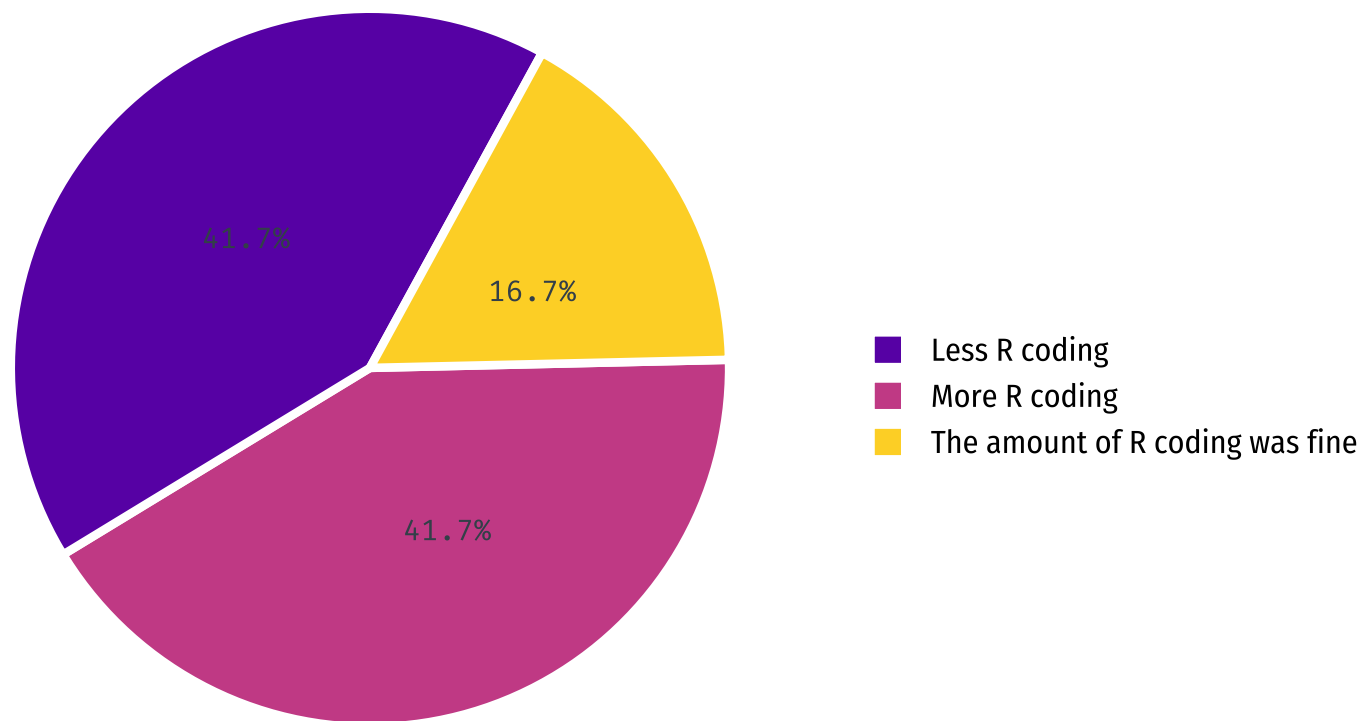
- **Learn to think about causality:**
 - Is this correlation or causation?
 - What assumptions need to hold?
 - How can I answer a causal question?
- **Learn to think about prediction:**
 - What do we care when thinking about a good prediction model?
 - What type of data do I need?
 - How can I use prediction to answer some pressing questions?

Objectives of this course

Even if you don't do these studies, learn how to be **CRITICAL CONSUMER** of these products/analysis

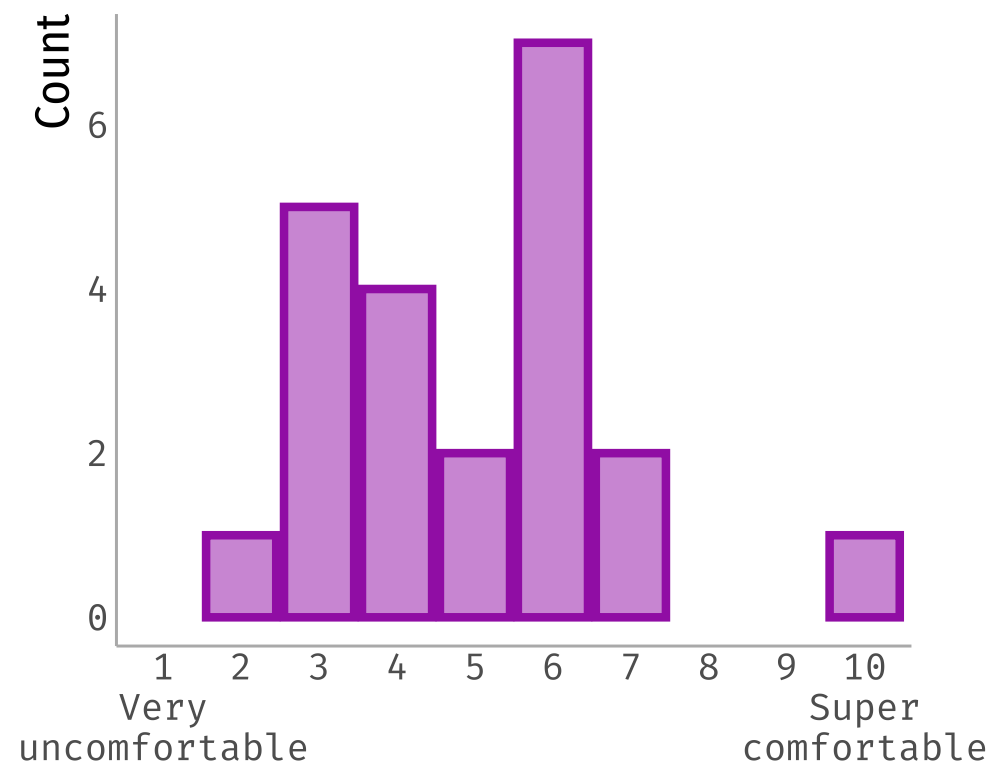
Same with coding!

Do you wish we would have done more or less R coding?

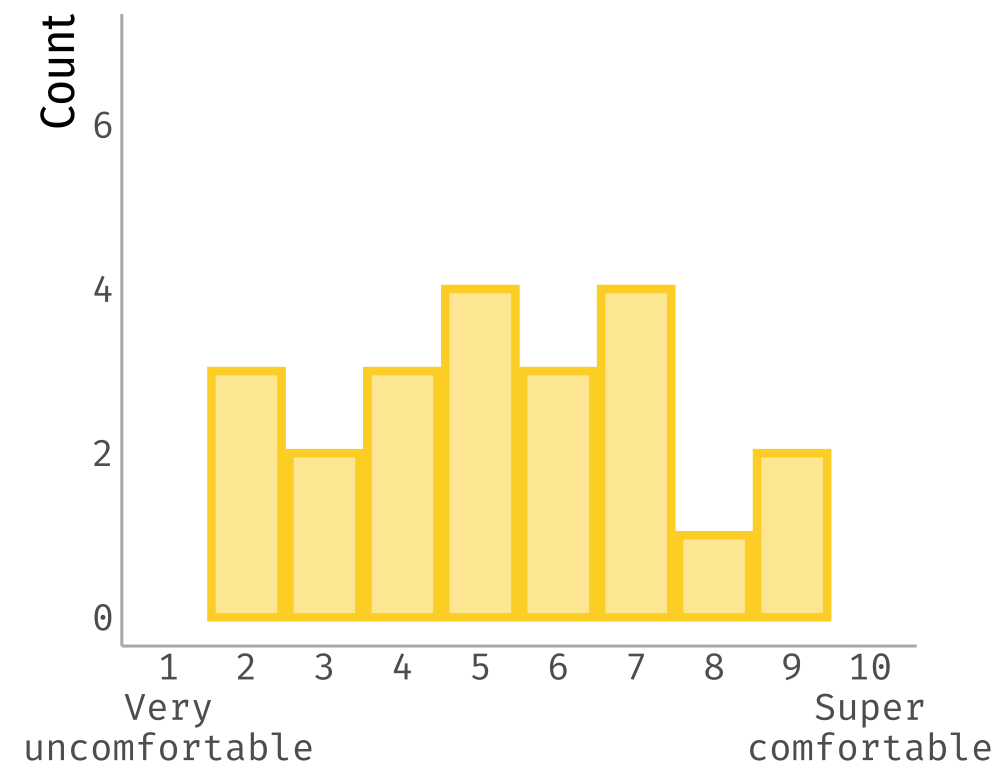


How comfortable are you with R?

Start of the semester

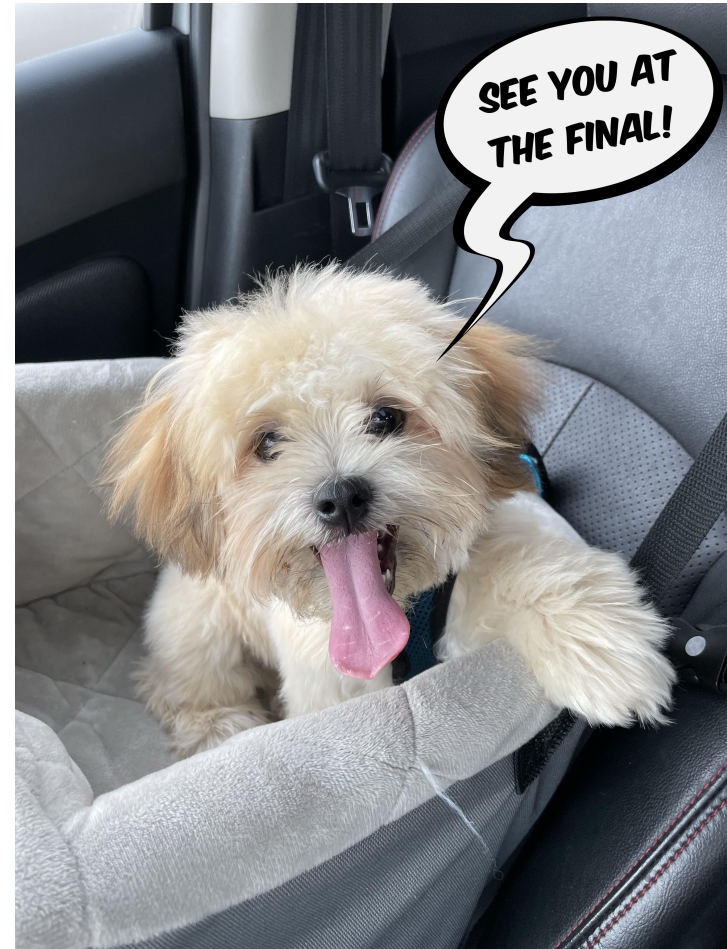


End of the semester



Remember for the final exam...

- You will need a **webcam**.
- Try to have a **stable internet connection**.
- The final exam will cover everything we have seen in the class (with an emphasis on prediction).
- **Attend office hours if you have questions!**



**We're very close to the finish line!
Good luck!**

