



McCOMBS SCHOOL OF BUSINESS

**Salem Center for Policy**

## PRL Introduction

David Puelz

August 26, 2021



- what class year are you?
- major?
- what are you hoping to get out of this class?

# Warm up: probability is cool



Let's play a game...

- You will flip a coin...
- if heads, you will write down the number 1 if your social security number ends with an even digit, otherwise write down 0.
- if tails, you will write down the number 1 if you ever got a positive COVID test, otherwise 0.

**Question:** What percentage of students have tested positive for COVID?



Law of total probability

$$\begin{aligned}P(1) &= P(1, \text{heads}) + P(1, \text{tails}) \\&= P(1 \mid \text{heads})P(\text{heads}) + P(1 \mid \text{tails})P(\text{tails})\end{aligned}$$

The bar “|” means conditional probability – like fixing a **known** state of the world.

When you're done , tell me whether you wrote down a 1 or 0!

# What is this class all about?



Evaluating policies (economic, governmental, otherwise) using the best available **evidence** + **techniques**.



# How are we doing this?

The **evidence** is data.

- data cleaning and organization
- data summarization

The **techniques** are statistical learning and coding.

- coding
- statistical modeling
- causal inference
- unsupervised learning

# How are we doing this?

The **evidence** is data.

- data cleaning and organization
- data summarization

The **techniques** are statistical learning and coding.

- coding
- statistical modeling
- causal inference
- unsupervised learning

$\underbrace{\text{evidence} + \text{techniques}}_{\text{thoughtful interpretation}} = \text{solid policy analysis!}$



## Foundational topics (weeks 1 through ~10)

- (1) Causality
- (2) Probability
- (3) Prediction (regression)
- (4) Unsupervised learning

## How to do research (weeks 11 through 15)

- (1) Reading academic papers
- (2) Data visualization, avoiding pitfalls
- (3) Resources at UT



# Class structure



## Before class

- readings and coding practice

## During class

- lecture and discussion

## After class

- homeworks (one per week)



## Evaluation

- homeworks (20%)
- in-class midterm (30%)
- research project (40%)
- engagement / participation (10%)



## Evaluation

- homeworks (20%)
- in-class midterm (30%)
- research project (40%)
- engagement / participation (10%)

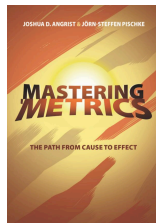
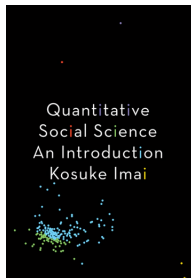
## Research project

- in groups of up to 3
- you choose either
  - (i) replicate existing policy analysis from peer-reviewed research
  - (ii) conduct your own policy analysis, gather data and investigate



- (i) replicate existing policy analysis from peer-reviewed research
- (ii) conduct your own policy analysis, gather data and investigate

These tasks are a bit different. More will be expected from a replication than a new study. Let me know of your group, decision, and research question by [October 29th](#).



- we will use the left book (QSS) for readings and exercises
- we will use the right book (MM) for supplementary reading
- additional readings will be provided as we work our way through the semester



- collaborate with your fellow students
- engage with the readings and in class discussions
- you are research assistants in addition to students, and the workload will be more than a traditional class
- be sure to attend the coding class on Fridays from 3-4p!
- I will be asking a lot of you because I know you're excellent students :)
- keep up with the fast pace and have fun!



- R for data analysis
- RMarkdown for writing reports, homeworks

R: an immensely capable, industrial-strength platform for data analysis.

It's used everywhere:

- **Academic research** (stats, marketing, finance, genetics, engineering)
- **Industry** (Google, Microsoft, eBay, Boring, Citadel, IBM, New York Times)
- **Governments/NGOs** (Rand, DOE, National Labs, US Navy)

R is free and looks the same on all platforms, so you'll always be able to use it.



A huge strength of R is that it is [open-source](#). R has a *core*, to which anyone can add contributed packages.

- 18,042 packages as of last week, as varied as the people who write them
- Some are specific, others general
- Some are great, some decent but unpolished, some are crappy

R has flaws, but so do all options (e.g., Python is great, but the community of stats developers is smaller, interactive data analysis is less slick, and you need to be a more careful and sophisticated programmer.)

Most prefer to use R via an IDE. We'll use [RStudio](#).



- A simple markup language for generating a wide variety of output formats (HTML, PDF, etc) from plain text documents
- Two pillars: (i) a formatting language, (ii) a conversion tool
- Much simpler than, for example, HTML

Rmarkdown allows you to write up data analyses easily within R to make reproducible reports. You can install the package directly in R by running the following command:

```
install.packages("rmarkdown")
```

# To the computer...



Let's go check out the class website, example R and Rmd files, and how Rstudio works ...