Introduction to Multivariate Regression & Econometrics

HED 612

Lecture 1

1. Student Introductions
2. About your instructor
3. What is Econometrics?
4. Course logistics
5. 10 MINUTE BREAK
6. What is R
7. R Basics
8. Directories and filepaths
9. Create "R project" and directory structure
10. Next Week

Student Introductions

Student introductions

1. Preferred name
2. Preferred pronouns
3. Academic program (and how far along)
4. GA, RA, TA, and/or job?
5. Why are you interested in this course?

About your instructor

Karina Salazar, instructor

My start in statistical analysis

- Began taking quantitative methods courses as a Master's student
  - Material did not come easy to me! I took intro course twice!
  - Took statistics courses in sociology
  - Causal inference courses at The University of Michigan
- Developed strong data skills
  - Getting data ready for analysis is often very time and labor intensive!
  - Learned data management skills in SPSS and Stata
  - Learned data science skills in Python and R
- Applied what I was learning in class.
  - Worked in assessment and institutional research offices; evaluated retention programs
  - Research assistantship in sociology; worked with national datasets (IPEDS, Survey of Earned Doctorates, Higher Education R&D Survey)
- Sought out teaching opportunities.
  - TA for HED regression and data management courses
  - Delivered practioner-based workshops

Enrollment Management Research Program

- Grew tired of mainstream access inequality policy discourse...
  - Onus on students, families, and K-12 schools
  - Enrollments can only tell us so much
- Do the enrollment management policies and practices of public universities undermine access for underserved student populations?
  - Universities say they care about low-income students, Students of Color
  - But who do they actually recruit? Use data science to collect recruiting events!
- Policy implications
  - Too many policy decisions for increasing access attempt to "fix" student behavior
  - Assumption: doubling low-income, Students of Color applying to a university will double their enrollment

Examples:

- The off-campus recruiting project
- Dissertation Defense

What is Econometrics?

Econometrics (also "Program Evaluation", Causal Inference)

Econometrics is the main methodological strategy used to assess the impact of programs, policies, and interventions.

The goal of econometrics is to determine the **causal effect** of the "program, policy, or treatment" (*the independent variable of interest*) on some educational outcome (*the dependent variable*) by inferring what would have happened in the absence of the "program, policy, or treatment" (*the counterfactual*).

- Do smaller classes improve learning?
- Does offering students financial incentives increase college completion?
- Is online instruction as effective as in-class instruction?

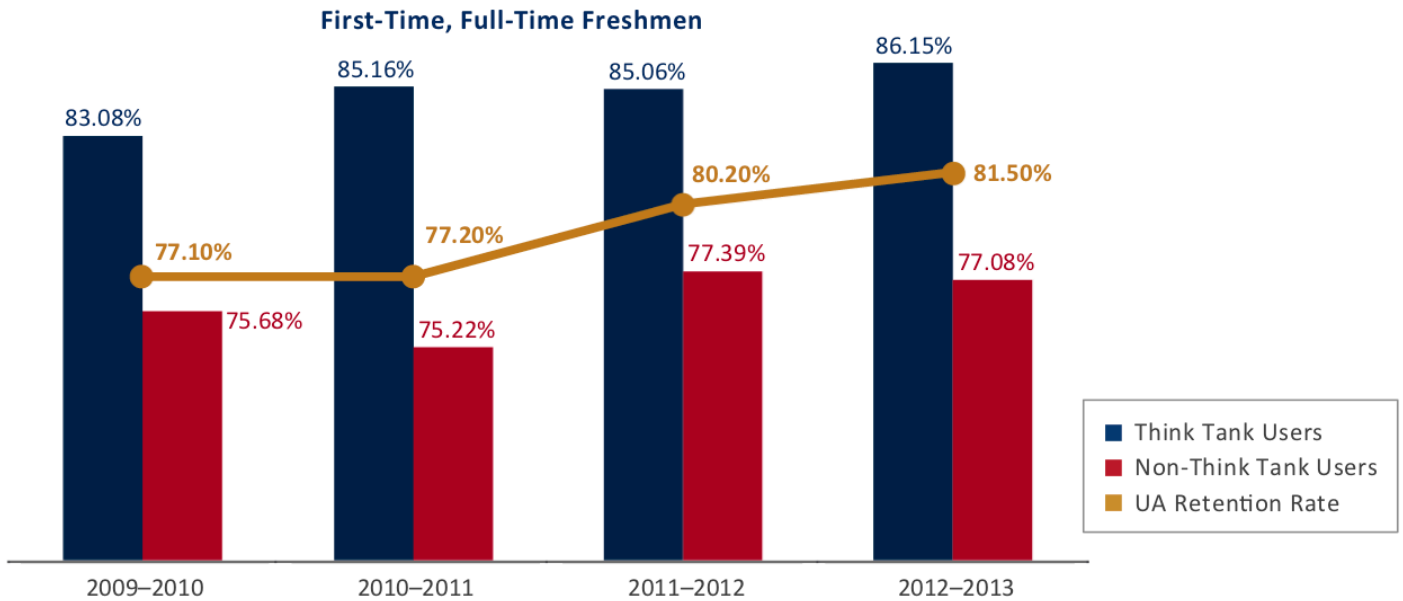
Causal relationships vs. descriptive statistics

- We often infer causality from correlation
- Ice cream consumption causes drownings vs drownings rise when ice cream consumption increases
- Smaller classes cause more learning vs. students in smaller classes have higher test scores

Example

ENGAGING

1 Freshman-to-Sophomore Retention



SUPPORT STRUCTURE: Talent Identity Information Technology Communications

32

Econometrics & Regression

Regression is used to answer both descriptive and causal questions

- One is not better than the other, they're just different types of research questions!
- You can use the same skills to answer both types of ?s via regression

Why learn regression from an econometrics lens?

If you decide to pursue econometrics research:

- Using regression to overcome *selection bias* in *the independent variable of interest* is the first "tool" in a methodologist's causal inference toolkit.
- Provides the foundation for more rigorous tools
  - e.g., Difference-in-Difference, Propensity Score Matching, Regression Discontinuity

If you decide not to pursue econometrics research:

- It's easier to learn the fundamentals of regression by focusing on **one** relationship
- Forces you to be thoughtful and intentional about what variables you include in regression models even when you're focusing on descriptive relationships
  - Avoid the "throwing everything but the kitchen sink" approach to regression modeling

Course logistics

Course logistics

- follow the syllabus

10 MINUTE BREAK

What is R

What is R

According to the Inter-university consortium for political and social research (ICPSR):

> *R is "an alternative to traditional statistical packages such as SPSS, SAS, and Stata such that it is an extensible, open-source language and computing environment for Windows, Macintosh, UNIX, and Linux platforms. Such software allows for the user to freely distribute, study, change, and improve the software under the Free Software Foundation's GNU General Public License."*

- For more info visit R-project.org

Base R vs. R packages

There are "default" packages that come with R. Some of these include:

- `as.character`
- `print`
- `setwd`

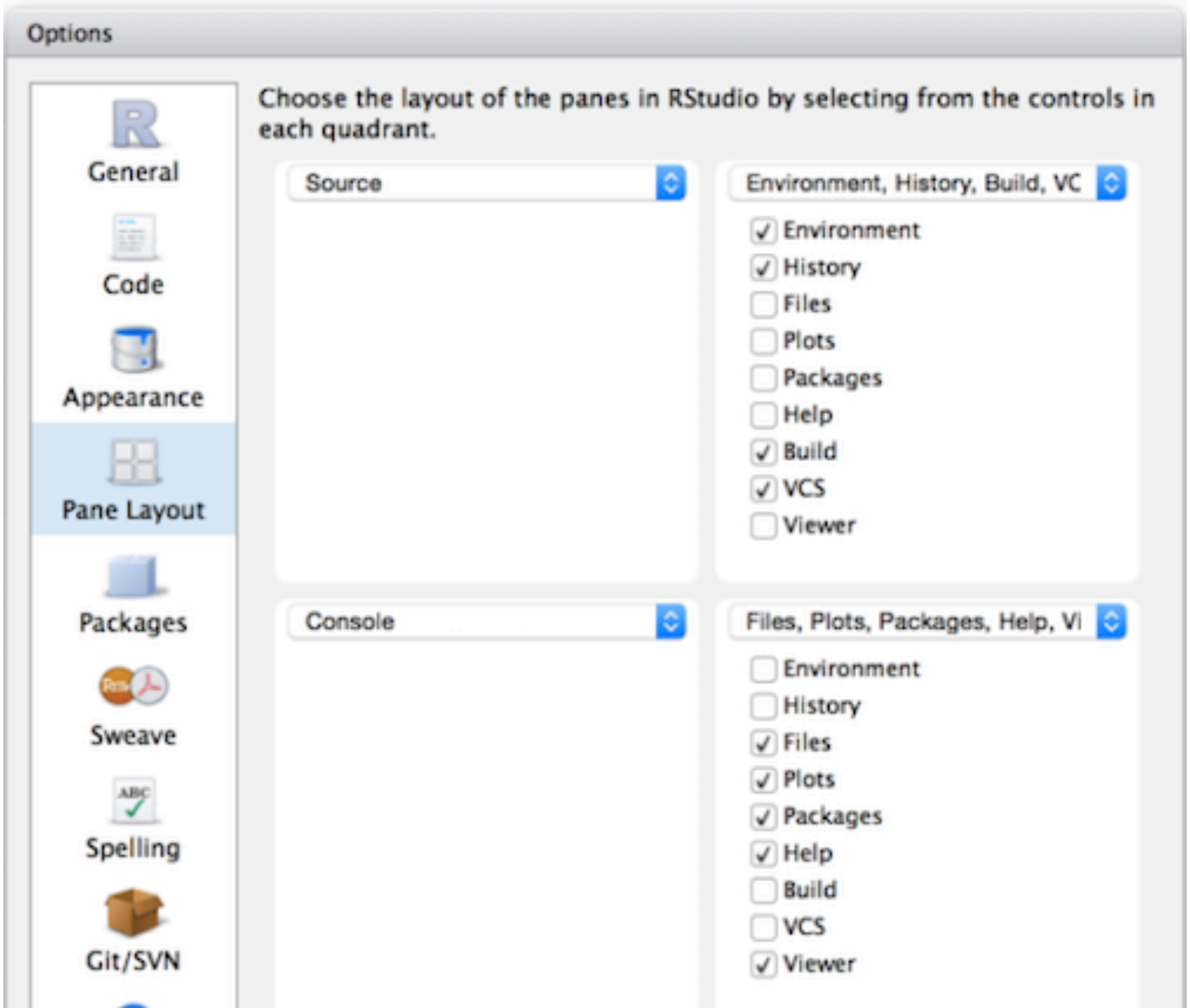
And there are R packages developed and shared by others. Some R packages include:

- `tidyverse`
- `stargazer`
- `equiomatic`

more about these in later weeks...

RStudio

"RStudio is an integrated development environment (IDE) for R. It includes a console, syntax-highlighting editor that supports direct code execution, as well as tools for plotting, history, debugging and workspace management."



R Markdown

R Markdown produces dynamic output formats in html, pdf, MS Word, dashboards, Beamer presentations, etc.

Why R? Capabilities of R

- Modeling
- Graphs
- Presentation
- Websites
- Journals
- Interactive tutorials
- Web apps
- Dashbaords
- Books
- Web scraping
- Maps

For more info visit

Modeling

Run just about any modeling technique! This course will focus on linear regression...

Population Regression Model:

- $Y_i = \beta_0 + \beta_1 X_i + u_1$

OLS Prediction Line:

- $\hat{Y}_i = \hat{\beta}_0 + \hat{\beta}_1 X_i$

What is the effect of engine size (cylinder count) on mpg?

- $\hat{\text{mpg}} = \hat{\beta}_0 + \hat{\beta}_1(\text{cyl})$

```
#remotes::install_github("datalorax/equatiomatic")
library(equatiomatic)
mod1 <- lm(mpg ~ cyl, mtcars)
```

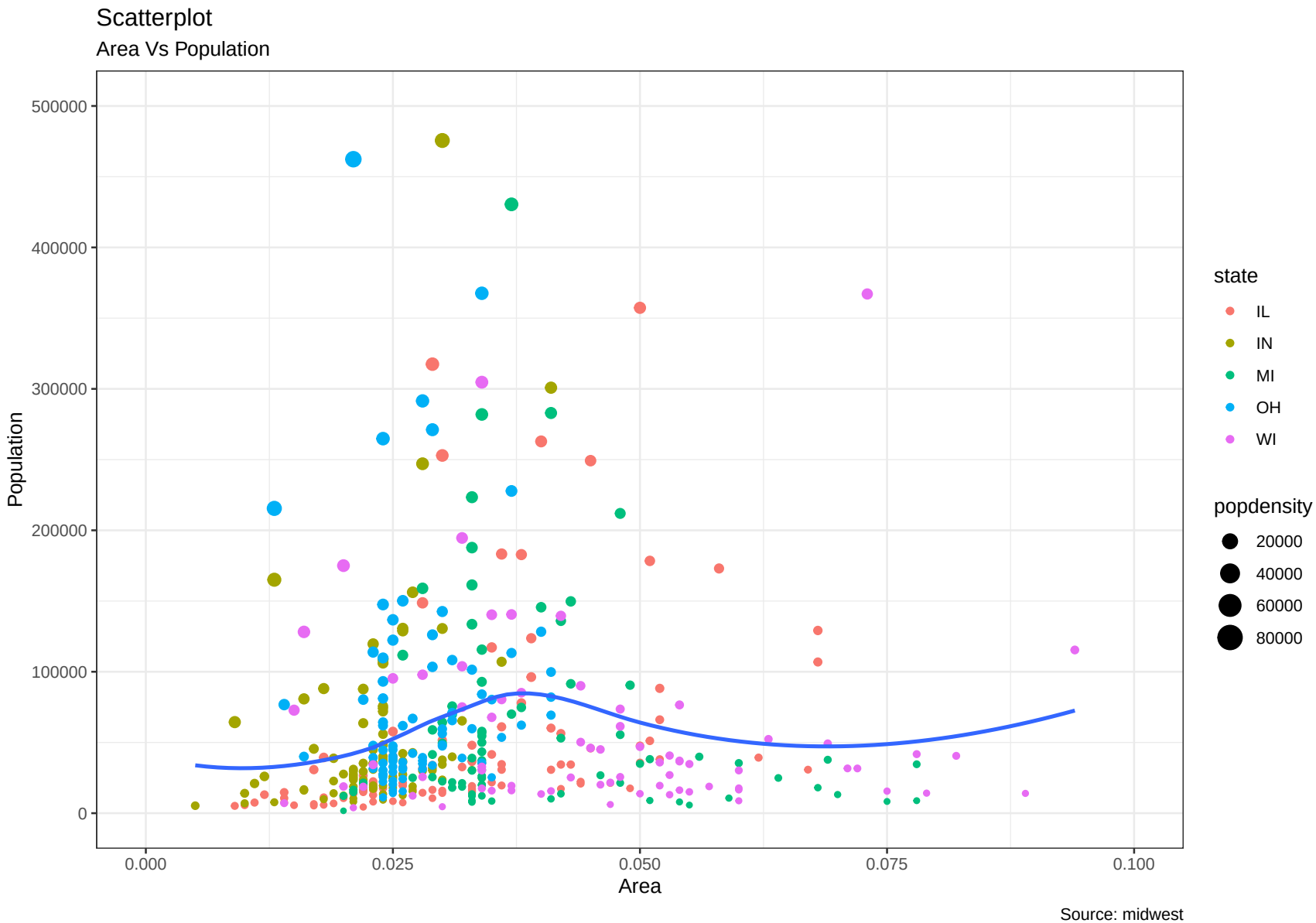
```
extract_eq(mod1, use_coefs = TRUE)
```

$$\text{mpg} = 37.88 - 2.88(\text{cyl}) + \epsilon$$

Graphs

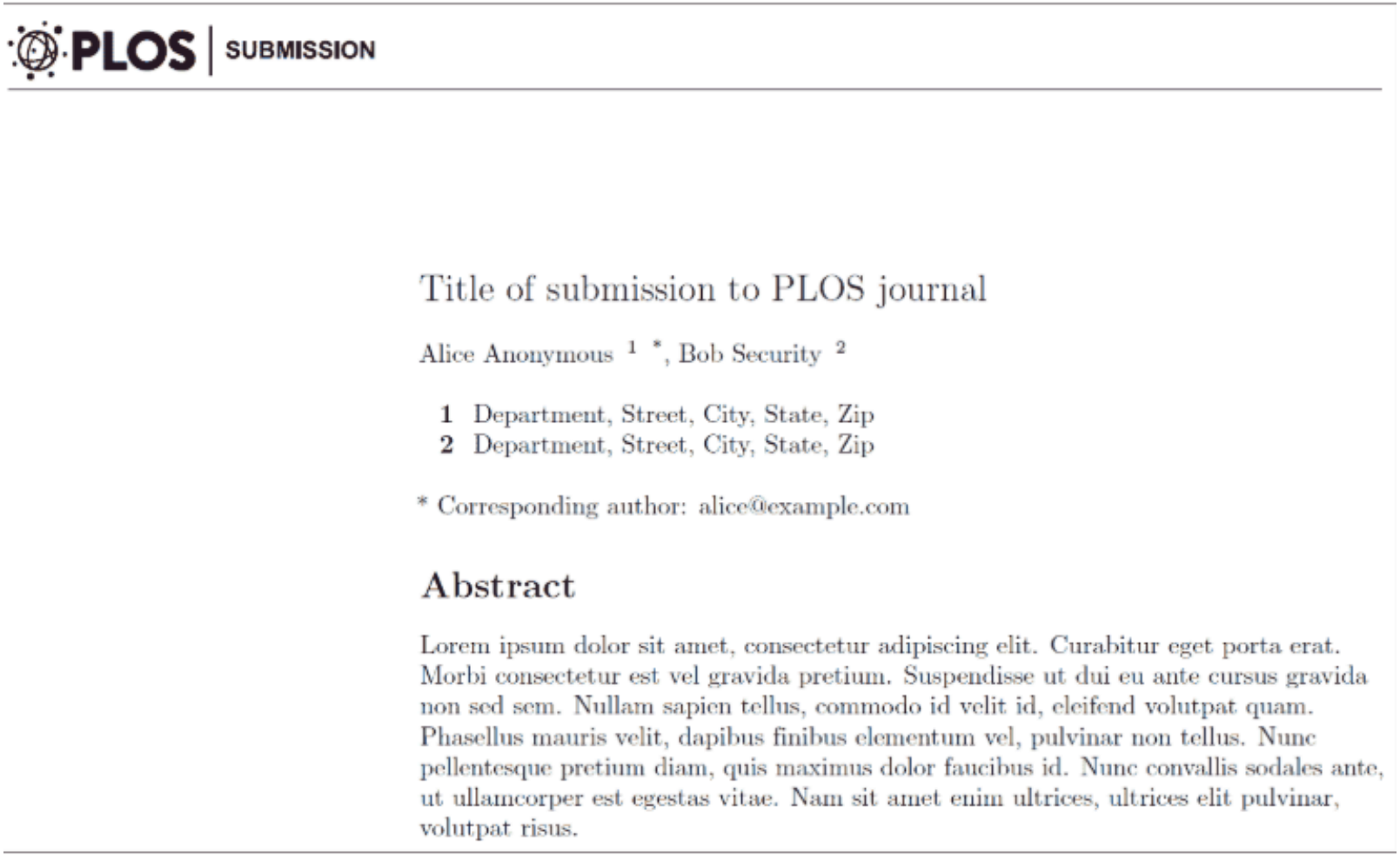
- Create graphs with ggplot2 package

```
#> `geom_smooth()` using formula 'y ~ x'
```



Journal articles

- Journal articles with rticles package


PLOS | SUBMISSION

Title of submission to PLOS journal

Alice Anonymous [1] *, Bob Security [2]

1 Department, Street, City, State, Zip
2 Department, Street, City, State, Zip

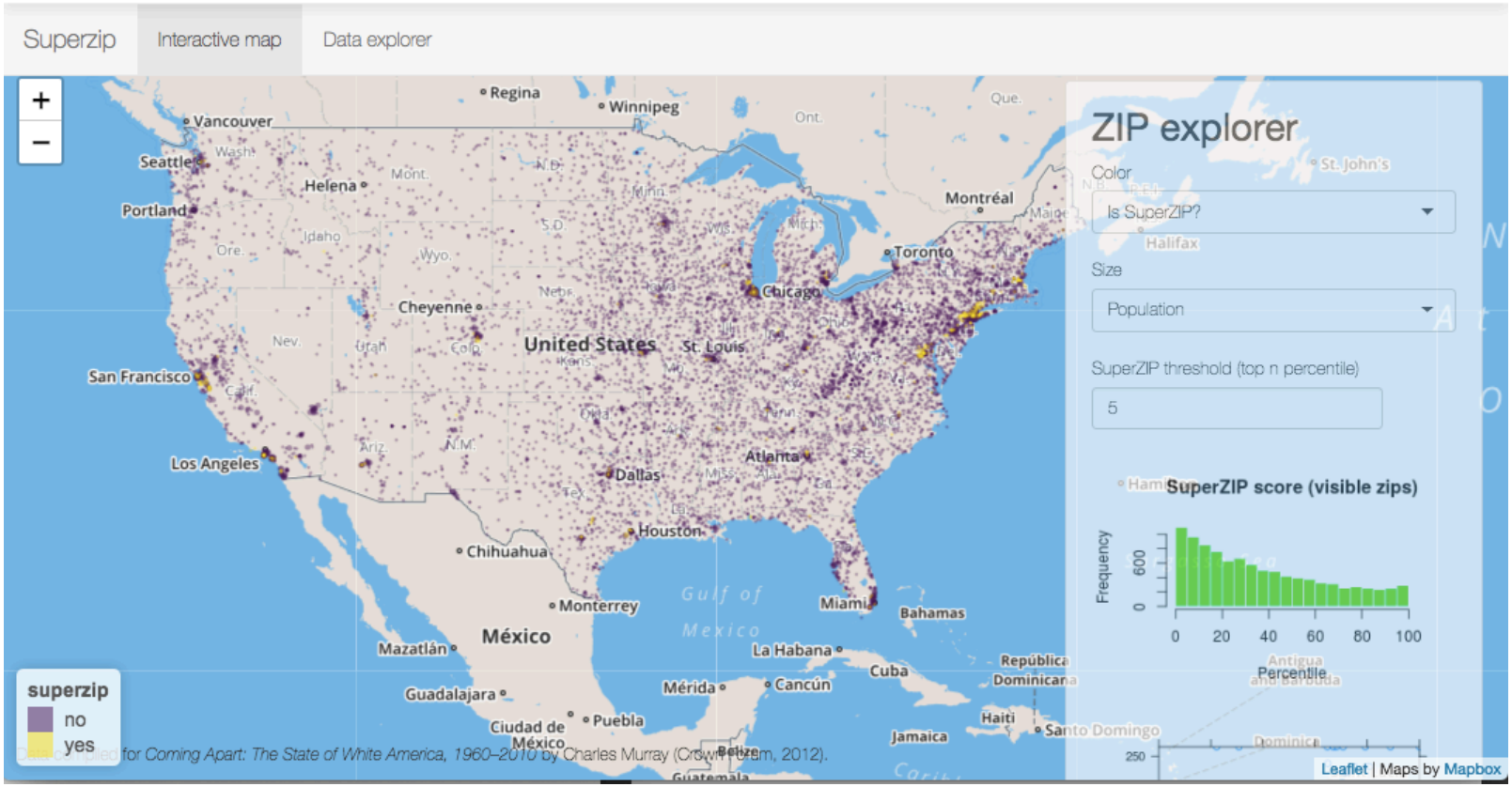
* Corresponding author: alice@example.com

Abstract

Lorem ipsum dolor sit amet, consectetur adipiscing elit. Curabitur eget porta erat. Morbi consectetur est vel gravida pretium. Suspendisse ut dui eu ante cursus gravida non sed sem. Nullam sapien tellus, commodo id velit id, eleifend volutpat quam. Phasellus mauris velit, dapibus finibus elementum vel, pulvinar non tellus. Nunc pellentesque pretium diam, quis maximus dolor faucibus id. Nunc convallis sodales ante, ut ullamcorper est egestas vitae. Nam sit amet enim ultrices, ultrices elit pulvinar, volutpat risus.

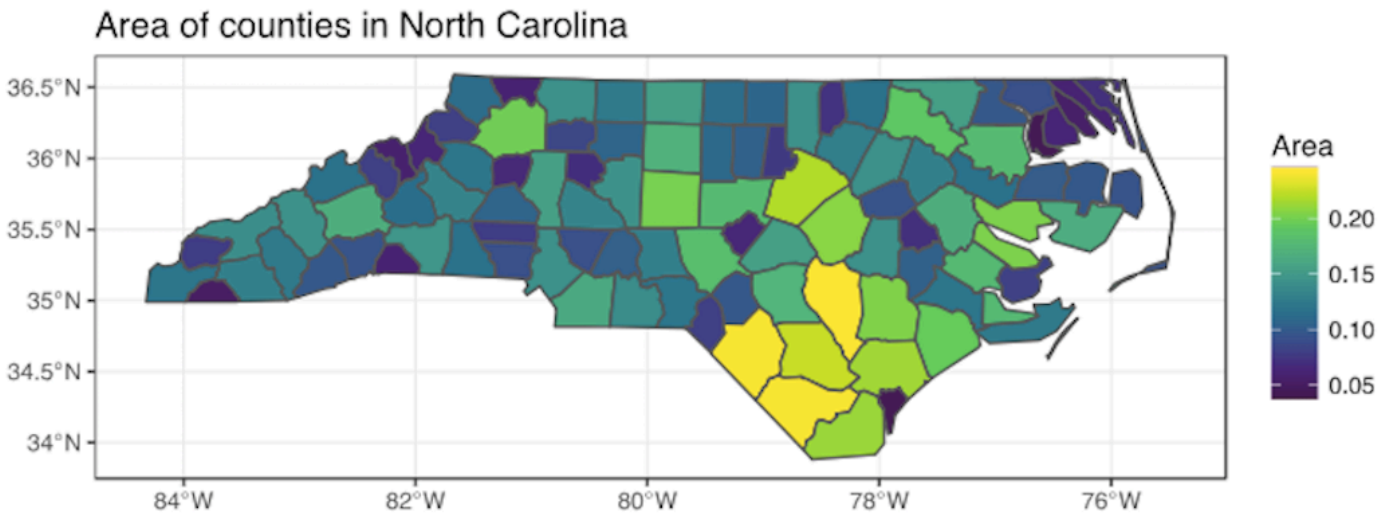
Interactive web apps

- Interactive web apps with shiny package



Mapping

- Mapping with sf package & ggplot



R Basics

R as a calculator

```
5
#> [1] 5
5+2
#> [1] 7
10*3
#> [1] 30
```

Executing commands in R

```
5
#> [1] 5
5+2
#> [1] 7
10*3
#> [1] 30
```

Three ways to execute commands in R

1. Type/copy commands directly into the "console"
2. 'code chunks' in RMarkdown (.Rmd files)
   - Can execute one command at a time, one chunk at a time, or "knit" the entire document
3. R scripts (.R files)
   - This is just a text file full of R commands
   - Can execute one command at a time, several commands at a time, or the entire script

Assignment

Assignment means creating a variable – or more generally, an "object" – and assigning values to it

- `<-` is the assignment operator
  - in other languages `=` is the assignment operator
- good practice to put a space before and after assignment operator

```
# Create an object and assign value
a <- 5
a
#> [1] 5

b <- "yay!"
b
#> [1] "yay!"
```

Directories and filepaths

Working directory

(Current) Working directory

- the folder/directory in which you are currently working
- this is where R looks for files
- Files located in your current working directory can be accessed without specifying a filepath because R automatically looks in this folder

Function `getwd()` shows current working directory

```
getwd()
#> [1] "/Users/karinasalazar/Dropbox/HED612 Linear Regression/lectures/lecture1"
```

Command `list.files()` lists all files located in working directory

```
getwd()
#> [1] "/Users/karinasalazar/Dropbox/HED612 Linear Regression/lectures/lecture1"
list.files()
#>  [1] "abor.png"                  "data-structures-overview.png"
#>  [3] "education.jpg"             "example1.html"
#>  [5] "example1.Rmd"              "HW1.pdf"
#>  [7] "HW1.Rmd"                   "lecture1_files"
#>  [9] "lecture1.1_ua_files"       "lecture1.aux"
#> [11] "lecture1.out"              "lecture1.pdf"
#> [13] "lecture1.Rmd"              "lecture1.tex"
#> [15] "lecture1.vrb"              "normaldist.jpg"
#> [17] "normaldist.png"            "pane_layout.png"
#> [19] "professors.png"            "rticles.png"
#> [21] "sf.png"                    "shifts.jpg"
#> [23] "shifts.png"                "shiny.png"
```

Absolute vs. relative filepath

Absolute file path: The absolute file path is the complete list of directories needed to locate a file or folder.

```
setwd("Users/Karina/rclass/lectures/lecture2")
```

Relative file path: The relative file path is the path relative to your current location/directory. Assuming your current working directory is in the "lecture2" folder and you want to change your directory to the data folder, your relative file path would look something like this:

```
setwd("../../data")
```

`File path shortcuts`

Key	Description
~	tilde is a shortcut for user's home directory (mine is my name)
../	moves up a level
../../	moves up two level

Create “R project” and directory structure

What is an R project? Why are you doing this?

What is an "R project"?

- helps you keep all files for a project in one place
- When you open an R project, the file-path of your current working directory is automatically set to the file-path of your R-project

Why am I asking you to create R project and download a specific directory structure?

- I want you to be able to run the .R scripts for each lecture on your own computer
- Sometimes these .R scripts point to certain sub-folders in order to open data
- If you create the R project and create directory structure I recommend, you will be able to run .R scripts from your own computer without making any changes to file-paths!

Follow these steps to create "R project" and directory structure

1. Download the zip folder on D2L:
   - Unzip the folder: this is a shell of the file directory you should use for this class
   - Move it to your preferred location (e.g, documents, desktop, dropbox, etc)
2. In RStudio, click on "File" » "New Project" » "New Directory" » "New Project"
   - In "Directory name", type "hed612_project" as the title of the Rproject for the course
   - In "Create project as subdirectory of", click browse and:
     - save the R Project within the hed612 folder (same general folder as data and lectures)

After you follow these steps

- You can add any additional sub-folders you want to the "HED612" folder
  - e.g., "syllabus", "homework"
- You can add any additional files you want to the sub-directory folders you unzipped
  - e.g., in "HED612/lectures/lecture1" you might add an additional document of notes you took

Next Week

Next Week

Next week will focus on reviewing intro statistics

Reading

- Heiss, pages 1-24 (stop at import/export of text files)
- Stock & Watson Chapter 3
  - Read/skim if you feel a bit rusty on some of the foundational concepts (t-tests, correlations, confidence intervals)

Problem Set #1

- Due next Wednesday 1/20/2021 at 4:15pm
- Just giving you some practice with working within an R script and changing "working directories" via relative/absolute filepath