Advanced Quantitative Methods **R Basics 1**

Instructor: Gregory Eady Office: 18.2.10 Office hours: Fridays 13-15

About me	About you	About the course	What is R?	Latex	Intro to

Today

- 1. About me
- 2. About you
- 3. About the course and its goals
- 4. About R
- 5. Introduction to R

1. Research:

- Methodological research
 - Ideological scaling
 - LLMs in survey research
 - Sensitive beliefs
- Applied research
 - Effects of Russian trolls on voting in the 2016 US election
 - Effects of January 6 insurrection on partisanship
 - Effects of COVID on interest group access
 - Electoral competitiveness and sharing of ideologically extreme content on social media

2. Teaching:

- Advanced Quantitative Methods
- Political Analysis of Social Media Data
- SDS Base Camp
- SDS Advanced Social Data Science II
- MA thesis supervision

About me	About you	About the course	What is R?
000			

Latex

Intro to R 000000000

С

Course website

gregoryeady.com/ResearchMethodsCourse/

What is R? 000000000000 Latex

0

If you need help

- o E-mail
 - . Try to use only for quick questions
 - Copy and paste code, don't screenshot
- o In person
 - · Schedule a meeting on course website
 - Fridays, 13.00-15.00
 - . If no slots left, email to set up another time
 - Can drop by my office if the door is open

About you

.

- Go to the course website
- \circ Weekly readings \rightarrow 1. R Basics 1
- o "Padlet for AQM"
- Make a post describing your background...

Latex

Intro to F 00000000

0

Make a post with the following

- 1. Title: Your full name
- 2. Research methods background
- 3. Statistical software background?
- 4. What would you like to use quantitative research methods for?

Aim of the course

- 1. Identify relevant designs & techniques to solve issues in the study of politics
- 2. Process data in structured and unstructured formats
- 3. Analyze those data
- 4. Reflect on the advantages and disadvantages of different designs and techniques

e What i 00000

000

tex

ntro to R 000000000

Assignment 1 (due October 9)

- Download and load a dataset
- Clean data
- Create figures
- $_{\odot}$ Run and interpret OLS regression models
- o Already available on the course website

Assignment 2 (due December 12)

Consists of three sub-assignments to replicate the main results from top-level political science articles:

- o Explain the research question
- Explain the research design
- Replicate the main regression table(s) & figure(s)
- A small extension for each article to show you understand the design and its potential flaws

Sub-assignments:

- 1. Instrumental variables (hand-in Week 46)
- 2. Regression discontinuity design (hand in Week 48)
- 3. Difference-in-differences (hand in Week 50)

Class structure

- 1. First hour:
 - · Lecture on a particular method and its applications
- 2. 10-15 minute break
- 3. Second hour:
 - Lab to implement the method in ${\sf R}$

What is R? 0000000000000 Latex

ntro to R)0000000000

Lecture schedule

Coding & Preliminaries

Lecture 1	R Basics 1
Lecture 2	R Basics 2
Lecture 3	OLS refresher & OLS in R
Lecture 4	Visualization in R

Experiments

Lecture 5	Credibility Revolution & LATEX
Lecture 6	Experiments
Lecture 7	CACE & Randomization inference

What is R? 000000000000 Latex

About the course

Quasi-experiments

Lecture 8	Instrumental variables
Lecture 9	Panel data
Lecture 10	Regression discontinuity
Lecture 11	Difference-in-differences
Lecture 12	Event studies
Lecture 13	Staggered diff-in-diff

Books & readings

- Mastering 'Metrics: The Path from Cause to Effect Princeton University Press, 2014.
 Joshua D. Angrist and Jörn-Steffen Pischke
- Field Experiments: Design, Analysis, and Interpretation
 W. W. Norton & Company, 2012.
 Alan S. Gerber and Donald P. Green.
- o Articles posted to website

 Causal Inference: The Mixtape (optional) Yale University Press, 2021.
 Scott Cunningham.

What is R? 0000000000000 Latex O ntro to R 0000000000

Why does this course exist?

- 1. Data revolution
- 2. Computational revolution
- 3. Credibility revolution

"Advanced" Quantitative Methods

- The adjective "Advanced" in the course name is <u>not</u> about advanced statistical methods
- $_{\odot}\,$ It is about cleverly thinking about research design
- Successful political science research is rarely about statistical sophistication
- $\,\circ\,$ Indeed, we almost solely rely on OLS

Statistical software

- SPSS (social sciences)
- o Matlab (statistics, computer science, natural sciences)
- C++
- Stata (still dominate in economics)
- \circ Python
- R

Why R?

- **1.** Extremely flexible
- 2. Workflow is replicable
- 3. R versus Python
- 4. Free
- 5. Open-source and free libraries
- **6.** Massive R community online
- **7.** Cutting-edge methods are often implemented in R first
- 8. Data visualization

About me

lbout you

About the course

What is R? 00000000000 Latex

Intro to R 0000000000



Slide 19 of 40

What is R? 00000000000 Latex O Intro to R 000000000

Hadley Wickham created 2 important R libraries:

o tidyverse

o ggplot2

me About you	About the course	What is R?	Latex	Intro t
OO		000000000000	O	000000

INFLUENZA-LIKE ILLNESS ACTIVITY LEVEL MINIMAL LOW MODERATE HIGH NO DATA 1 2 3 4 5 8 7 8 9 OCT. 1-7, 2017 OCT. 8-14 OCT. 15-21 OCT. 22-28 OCT. 29-NOV. 4 NOV. 5-11 NOV. 12-18 NOV. 19-25 NOV. 26- DEC. 2 DEC. 3-9 DEC. 10-16 DEC. 17-23 DEC. 31-JAN. 6, 2018 JAN. 7-13 DEC. 24-30 JAN. 14-20 JAN, 21-27 JAN. 28-FEB. 3 FEB. 4-10 FEB. 11-17

SOURCE: CENTERS FOR DISEASE CONTROL AND PREVENTION INFLUENZA DIVISION

About me	About you	About the course	What is R?	Latex	Intro to R
			00000000000		

Supreme Court departures and replacements

Supreme Court justices from 1955-2016 by their ideological leanings, based on their Judicial Common Space scores

DEPARTING JUSTICE O ENTERING JUSTICE



Years refer to Supreme Court terms, which run from October to September.

SOURCES: LEE EPSTEIN, CHAD WESTERLAND, THE JOURNAL OF LAW, ECONOMICS, AND ORGANIZATION

About me 000	About you OO	About the course	What is R? 0000000000000	Latex O	Intro to R 0000000000
Partiern	ies Danmarkskort ved	Folketingsvalget 2019			
Af @fghjo A	orth med inspiration fra @h . Socialdemokratiet 25.9%	unrkindbrg V. Venstre, Danmarks Liberale Parti 23.4%	O. Dansk Folkeparti 8.7%	B. Radikale \	'enstre 8.6%
	-	-		-	



Slide 23 of 40

About me	About you	About the course	What is R?	Latex	Intro to R
			00000000000		

Solar panels have become steadily cheaper

When accounting for a solar panel's generative capacity (using cost-per-watt), solar panel costs have decreased and become more consistent over the last 20 years.



Data sources: NREL (U.S. Dept of Energy)



Large panels-per-capita in the Northeast, despite low annual solar radiation

While the Northeast region receives the least solar irradiance in the US, it also has counties with disproportionately high solar panel installation rates.



About me About you About the course 000 00 00000000	What is R?	Latex	Intro to R
	000000000000000	O	0000000000

Drawbacks to R

- No graphical interface
- $_{\odot}\,$ Can be a bit intimidating at first
- Cryptic error messages
 - object of type 'closure' is not subsettable

You will code every week in this course

- $_{\odot}\,$ If you have not coded before, it will be frustrating at first
- $_{\odot}\,$ Everyone who finishes this course has learned how to code in R
- For those who already know R (perhaps many of you), you will probably need to learn more to complete the assignments
- We will use tidyverse throughout
 - You can use Base R functions as well, if you really want to

About me	About you	About the course	What is R?	Latex	Intro to R
000	OO		00000000000	O	000000000



VICTORIA FINE • AUG 31, 2015 • 4:59 AM

I Was Intimidated by Coding Until I Learned This Secret Strategy: Googling

- Searching on Google and asking an LLM like ChatGPT is a skill
- $_{\odot}$ You will develop this through experience
- It is extremely rare that one cannot find a solution by using Google or ChatGPT

0

Latex

- $_{\odot}$ Typesetting system for scientific writing
- $_{\odot}$ As easy as or easier to learn than R
- $_{\odot}$ Modern way to use $\ensuremath{\text{PTEX}}$ is through Overleaf
 - Overleaf is an online ${\ensuremath{\mathbb A}} ^{T} E^X$ interface for collaboration
- o Free
- \circ Does your bibliography for you
- $_{\odot}$ Easy formatting of regression tables from R
- o 'Easy' to do mathematical formulas
- Can work simultaneously on the same document (like Google Docs)

Let's learn some R

Install R:

https://cran.r-project.org/mirrors.html

Install RStudio:

https://posit.co/download/rstudio-desktop/

These links are on the course website under the first "Weekly Readings"

About me	About you	About the course	What is R?	Latex	Intro to R
000	OO		00000000000	O	0000000000

What is RStudio?

• An Integrated Development Environment (IDE)

About me	About you	About the course	What is R?	Latex	Intro to R
					000000000

	RStudio		
🔍 = 🥸 💣 = 🔒 🔂 🍈 🍺 Go to file/function 🛛 🗟 = Addins =			Project: (None) •
• Untitled1 ×		Environment History Connections	
🗇 🗇 🔄 🔄 Source on Save 🔍 🎢 📲 🗌	👄 Run 👌 📤 Source 🔹 🗟	😅 🕞 📑 Import Dataset + 🥑	🗏 List • 🎯 •
1		Global Environment +	٩,
		Environment is empty	
		Files Plots Packages Help Viewer	- - -
1:1 (Top Level) +	R Script ‡		
Console Terminal × Jobs ×			
R version 3.6.0 (2019-04-20) "Planting of a Tree" Copyright (C) 2019 The R Foundation for Statistical Computing Platform: x86.4 complex-dwarfs.6 (64-bit) R is free software and comes with ASSULTELY ND MARSNAT. You are malcome to redistribute it under certain conditions. Type 'licenseC)' or 'licenseC)' for distribution details. Natural Longuage support but running in an English locale R is a collaborative project with many contributors. Type 'demo(') for some information and 'clatistica') for some damos, 'help()' for on-line help, or 'help.star(C)' for an HIM browser interface to help. Type 'demo(') to quit R.			

RStudio basics

- o Comment and uncomment: Ctrl + Shift + c
- Run line of code: Ctrl + Enter
- Run code in console: Enter
- Clear console: Ctrl + 1
- \circ Indent: Tab
- o Outdent: Shift + Tab

R basics

- $_{\odot}$ Calculate: (2 + 5) / 7
- Create a variable:
 - var_1 <- c(1, 4, 5, 7)
- Create another variable:
 - var_2 <- c(1, 15, 25, -1)
- Elements of a vector: var_1[1:3]
- o Arithmetic with variables:
 - var_1 * var_2
 - var_1 * 2
- o Create a data.frame
 - D <- data.frame(var_1, x = c(1, 2, 3, 4))
 - D <- data.frame(var_1, x = c(1, 2, 3))

R basics

- o Variables in data: names(D)
- o Index variable within a data.frame:
 - D\$var1[2:4]
- Create new variable within existing data.frame:
 - D\$new_var <- c(100, 200, 300, 400)
- Basic sampling functions:
 - sample(1:6, 10)
 - rnorm(10, mean = 5, sd = 0.25)

R basics

- o Install a package: install.packages("tidyverse")
- Load a package: library(tidyverse)
- Load data from a package: data(iris)
- What variables are in the data?: names(iris)
- Histogram of a variable:
 - ggplot(iris, aes(x = Sepal.Length)) +
 geom_histogram()

About me	About you	About the course	What is R?	Latex	Intro to R
000	OO		000000000000	O	000000000000

A "for" loop

```
x <- NA
for (i in 1:10) {
    print(i)
    x[i] <- i
}
```

About	Abo

About the 00000000

Latex

Intro to R 0000000000

Functions

```
get_mean <- function(x) {
  the_mean <- sum(x) / length(x)
  return(the_mean)
}</pre>
```

Intro to R

Loading data

- $_{\odot}$ Download GM.csv from the course website
- $_{\odot}$ Then load that dataset into R by doing the following:

Find the directory and file name of the file
Set your working _directory_
setwd("/Users/mkf215/Downloads/")

Load the dataset GM into a data.frame object # that we'll call "D" D <- read_csv("GM.csv")</pre>

What is R? 000000000000 Latex O

.

What's next

- $_{\odot}$ Introduction to tidyverse by Hadley Wickham
- Data processing with tidyverse
- $\circ\,$ To do:
 - Complete the next DataCamp exercises before next class