Day 01: Programming for the Social Sciences

Erin Rossiter

23 January, 2023

Today's plan

Today's plan

- Meta-goals and expectations
- Syllabus
  - Specific course goals
  - How we'll achieve those goals
- Introductions
- Pretest :)
- Break
- A little lecture
- A little lab

The Big Picture (the *really* big picture)

Goals of advanced education

This class is for graduate students, so let's talk about the goals of advanced education:

- Hard skills
- Socialization
- Critical thinking
- Project management
- Communication skills

There's no checklist for original research.

We practice all of this.

How to pursue these goals?

The structure of this class helps us pursue "meta" goals of grad education:

- Learning is messy, need to get dirty
- Process oriented (not outcome oriented)
- There is no way to learn programming through lectures
- I'll help too by:
  - having high expectations
  - creating consequences for putting in the work
  - supporting you along the way!

How to do well in this course (and grad school!)

1. Expect a lot from yourself
2. Consider our time toether scarce and valuable
3. Use the buddy system
4. Stay positive. You're here because you like learning things! (Including R)
5. Use the Internet
6. Read the books.
7. Work *with* me.
8. Homeworks–they are a lot of work.
9. Come with questions.
10. If you don't understand something, ask about it.

My expectations

1. This is a grad *elective*. You don't have to be here, so I expect you want to be here and you'll buying into the process.
2. You're a grad student. Grad school, dissertations, research is all about you taking responsibility and ownership for your own development and work.

   – Example: If a problem set or feedback doesn't make sense, asking about it (in advance of a deadline) demonstrates ownership

Syllabus

Introductions

Introductions

1. Name
2. Program
3. Pick one:
   - A goal for this semester (school related or not)
   - Something you want to learn in this class
4. If applicable, do you know any other programming languages?

Final syllabus notes:

- mental health statement
- remember I'm a mandatory reporter

Final syllabus notes:

- mental health statement
- remember I'm a mandatory reporter

Final syllabus notes:

- mental health statement
- remember I'm a mandatory reporter

A little lecture

Setup

My Jan 17 email walks through having these things installed:

- Slack
  - join class workspace
  - download desktop app
- Github
  - create account
  - download Github Desktop
- R (latest version)
  - note this wipes all library installs
- Rstudio (latest version)

Why R, RStudio, and Rmd files?

Tools of the trade

- **R:** programming language, written in an R script (.R file) – **RStudio:** friendly interface where we edit and execute our R code – **RMarkdown:** (.Rmd files) allow you to use markdown syntax and R and compile into neat .pdf (or .html) file

R – The *language*

- Free, open-source, powerful, and flexible! this means:
  - lots of online resources
  - easier language to learn
  - statistical routines you might want to run have software built in R
- very widely used
- marketable skill
- great data analysis
- **These reasons are why it is widely used in polisci and beyond and why we have an entire semester to learn it!**

R – The *language*

- Free, open-source, powerful, and flexible! this means:
  - lots of online resources
  - easier language to learn
  - statistical routines you might want to run have software built in R
- very widely used
- marketable skill
- great data analysis
- **These reasons are why it is widely used in polisci and beyond and why we have an entire semester to learn it!**

R – The *language*

- Free, open-source, powerful, and flexible! this means:
  - lots of online resources
  - easier language to learn
  - statistical routines you might want to run have software built in R
- very widely used
- marketable skill
- great data analysis
- **These reasons are why it is widely used in polisci and beyond and why we have an entire semester to learn it!**

R – The *language*

- Free, open-source, powerful, and flexible! this means:
  - lots of online resources
  - easier language to learn
  - statistical routines you might want to run have software built in R
- very widely used
- marketable skill
- great data analysis
- **These reasons are why it is widely used in polisci and beyond and why we have an entire semester to learn it!**

R – The *language*

- Free, open-source, powerful, and flexible! this means:
  - lots of online resources
  - easier language to learn
  - statistical routines you might want to run have software built in R
- very widely used
- marketable skill
- great data analysis
- **These reasons are why it is widely used in polisci and beyond and why we have an entire semester to learn it!**

R – The *language*

- Free, open-source, powerful, and flexible! this means:
  - lots of online resources
  - easier language to learn
  - statistical routines you might want to run have software built in R
- very widely used
- marketable skill
- great data analysis
- **These reasons are why it is widely used in polisci and beyond and why we have an entire semester to learn it!**

R – The *language*

- Free, open-source, powerful, and flexible! this means:
  - lots of online resources
  - easier language to learn
  - statistical routines you might want to run have software built in R
- very widely used
- marketable skill
- great data analysis
- **These reasons are why it is widely used in polisci and beyond and why we have an entire semester to learn it!**

RStudio – The application

- RStudio is the application where we use *R*
- We could run R *without* RStudio (if time I'll show)
- Metaphor:
  - You could check your email *through a browser* on your phone, but we have an app to make it easier

RStudio – The application

- RStudio is the application where we use *R*
- We could run R *without* RStudio (if time I'll show)
- Metaphor:
  - You could check your email *through a browser* on your phone, but we have an app to make it easier

RStudio – The application

- RStudio is the application where we use *R*
- We could run R *without* RStudio (if time I'll show)
- Metaphor:
  - You could check your email *through a browser* on your phone, but we have an app to make it easier

RStudio – The application

- RStudio is the application where we use *R*
- We could run R *without* RStudio (if time I'll show)
- Metaphor:
  - You could check your email *through a browser* on your phone, but we have an app to make it easier

RMarkdown – The file format

- Just like making a .docx document in Word, we can make a .Rmd document in RStudio
- when the file extension is .R it is an R script
- when the file extension is .Rmd, you can insert R code into the document, but you can also easily write text around it and create a readable .pdf

RMarkdown – The file format

- Just like making a .docx document in Word, we can make a .Rmd document in RStudio
- when the file extension is .R it is an R script
- when the file extension is .Rmd, you can insert R code into the document, but you can also easily write text around it and create a readable .pdf

RMarkdown – The file format

- Just like making a .docx document in Word, we can make a .Rmd document in RStudio
- when the file extension is .R it is an R script
- when the file extension is .Rmd, you can insert R code into the document, but you can also easily write text around it and create a readable .pdf

Why RMarkdown?

- **Easiest** way for us to clearly communicate about the statistics and R programming we're learning this semester i.e., homework and slides
- Commonly used in political science when people share datasets and code, so good to be fluent even if you don't use it in your own work
- Might find it useful for replication archives some day
- Some people even write papers in RMarkdown
- Resources:
  - Website with basics
  - Shorter cheat sheet
  - Longer cheat sheet
  - Full book!

Why RMarkdown?

- **Easiest** way for us to clearly communicate about the statistics and R programming we're learning this semester i.e., homework and slides
- Commonly used in political science when people share datasets and code, so good to be fluent even if you don't use it in your own work
- Might find it useful for replication archives some day
- Some people even write papers in RMarkdown
- Resources:
  - Website with basics
  - Shorter cheat sheet
  - Longer cheat sheet
  - Full book!

Why RMarkdown?

- **Easiest** way for us to clearly communicate about the statistics and R programming we're learning this semester i.e., homework and slides
- Commonly used in political science when people share datasets and code, so good to be fluent even if you don't use it in your own work
- Might find it useful for replication archives some day
- Some people even write papers in RMarkdown
- Resources:
  - Website with basics
  - Shorter cheat sheet
  - Longer cheat sheet
  - Full book!

Why RMarkdown?

- **Easiest** way for us to clearly communicate about the statistics and R programming we're learning this semester i.e., homework and slides
- Commonly used in political science when people share datasets and code, so good to be fluent even if you don't use it in your own work
- Might find it useful for replication archives some day
- Some people even write papers in RMarkdown
- Resources:
  - Website with basics
  - Shorter cheat sheet
  - Longer cheat sheet
  - Full book!

Why RMarkdown?

- **Easiest** way for us to clearly communicate about the statistics and R programming we're learning this semester i.e., homework and slides
- Commonly used in political science when people share datasets and code, so good to be fluent even if you don't use it in your own work
- Might find it useful for replication archives some day
- Some people even write papers in RMarkdown
- Resources:
  - Website with basics
  - Shorter cheat sheet
  - Longer cheat sheet
  - Full book!

Expectations about tools of the trade

- I clarified common confusion between R, RStudio, and Rmd
- Moving forward, I assume familiarity with these tools

Why Github?

Github will be used for

- In-class labs
- Problem sets
- Final project

Why?

- Allows easy, clean, transparent code sharing
- Good to be comfortable with it for access to cutting-edge methods, future collaboration, etc.

Lab

Practicing Github

- I've invited you to the course repository (if you sent username)
- Accept invite, then
  - "fork" repository
  - "clone" to your desktop
  - copy and rename `lab.R` to make your own
  - open renamed file
    - make sure it opens in RStudio

Up Next

Up Next

Due ASAP (latest by next class)

- Get set up according to the email I sent on Jan 17

Due next class

- In class activity

Distributed tomorrow by 4pm

- PS1