

SESSION 1

INTRODUCTION TO R

R FOR SOCIAL DATA SCIENCE

JEFFREY ZIEGLER, PHD

ASSISTANT PROFESSOR IN POLITICAL SCIENCE & DATA SCIENCE
TRINITY COLLEGE DUBLIN

FALL 2022

ROAD MAP FOR TODAY

- About me, and structure of class
- Prerequisites and software
 - ▶ Get familiar with R, Rstudio, GitHub, and GitHub desktop

GENERAL INFO ABOUT COURSE

Instructor Jeffrey Ziegler, PhD

Email zieglerj@tcd.ie

In-Person Sessions 18:00 - 19:30 T/Th

Zoom <https://tcd-ie.zoom.us/j/6159065248>

DATA SCIENCE IS SOCIAL SCIENCE

- Quantitative social science: Using quantitative data to learn about the social, economic, and political world
 - ▶ Industry (Facebook, Google, Netflix, etc.)
 - ▶ Nonprofits and governments (Give Directly, OxFam, local services)
- This class will give you the hands-on **tools** of quantitative social scientists a.k.a "data scientists"

MODULE OBJECTIVES

- Introduce fundamentals of computer programming
- Get familiar with R programming language
- Develop understanding of core software design principles
- Learn crucial data science techniques (data manipulation, visualization)
- Practice these concepts using social science examples

APPROACH TOWARD LEARNING

Preparation + synthesis + practice = learning

- Individual preparedness: Reading & slides before class
- In class:
 - ▶ Discussion and Q&A on important concepts
- "Tutorial":
 - ▶ Group work on problems in R
- Exercises: Individual "homework" assignments

MATERIALS

- Matloff, Norman. 2011. The Art of R Programming: A Tour of Statistical Software Design. San Francisco, CA: No Starch Press.
- Roger D. Peng. 2016. **R Programming for Data Science**. Leanpub.
- Wickham, Hadley, and Garrett Golemund. 2017. **R for Data Science: Import, Tidy, Transform, Visualize, and Model Data**. Sebastopol, CA: O'Reilly Media.
- Wickham, Hadley. 2019. **Advanced R**. 2nd ed. Boca Raton, FL: Chapman and Hall/CRC.

ADDITIONAL ONLINE MATERIALS

- [Git Book](#)
- [R Documentation](#)
- [R Inferno](#)

PREREQUISITES AND SOFTWARE

- Introductory module - no formal prerequisites
- Computer with Windows/Mac/Linux OS (no Chrome books)
- Required software:
 - ▶ **R** (version 4+) - statistical programming language
 - ▶ **RStudio** - integrated development environment for R
 - ▶ **Git** - version control system
 - ▶ **GitHub Desktop** - integrated application for GitHub version control

TOOLS: REASONS TO LEARN R & GITHUB

R is...

- a tool to perform statistical analysis
- free
- cross-platform
- open-source
- can produce high level graphics
- can integrate with document publishing

GitHub is how we'll *share our work with each other*

R and GitHub are widely used in academics and industry,
put it on your resume!

"ASSESSMENT"

1. Code exists
2. Code runs and does what it has to do
3. Code is legible (meaningful naming, comments)
4. Code is modular (no redundancies, use of abstractions)
5. Code is optimized (no needless loops, runs fast)

MODULE OUTLINE

Sessions 1 - 2	Introduction and Computation
Sessions 3 - 4	R Basics
Sessions 5 - 6	Control Flow
Reading Week	
Sessions 7 - 8	Functions
Sessions 9 - 10	Debugging, Testing, Performance and Complexity
Sessions 11 - 12	Data Wrangling
Sessions 13 - 14	Visualisation
Sessions 15 - 16	Gathering electronic data

"TUTORIAL": DOWNLOADING R AND RSTUDIO

- There is a number of integrated development environments (**IDEs**) available for:
 - ▶ R (RStudio)
 - ▶ Python (Spyder, PyCharm)
- As well as text editors with R/Python-specific extensions (Visual Studio Code, Atom, Sublime Text, Vim)
- Try different ones and choose what works best for you!

LINKS FOR DOWNLOADING R AND RSTUDIO

- R: <https://cran.r-project.org/>
- RStudio: <https://www.rstudio.com/products/rstudio/>

CLASS BUSINESS

Today, we...

- Downloaded R and RStudio

Next time, we'll go over...

- Computers, computational thinking, algorithms
- Programming languages and computer programs
- Debugging
- Command-line interfaces (CLI)
- Version controlling with Git/GitHub