

1X: Introduction to R

Essex Summer School for Social Science Data Analysis

Phil Swatton

10/07/2022

Welcome to 1X Intro to R!

I'm Phil, a PhD student in the Department of Government at the University of Essex.

Today, we'll be learning how to use R (but you knew that already).

Today's session

- Getting you ready to use R in the wild
- Some statistical knowledge helpful but not necessary
- Plenty of code that you can refer back to
- Some good coding practices
- Assuming no previous experience with R

Class structure

1. Using RStudio, data types, data structures
2. Importing datasets, manipulating data
3. Descriptive statistics, plots
4. Statistical tests and models

Each section will be roughly an hour to an hour and a half long.
Each section will include a set of exercises so you can practice what we've learned.

I'm not necessarily expecting you to finish the exercises today - it's okay if you don't! I will be uploading a solutions sheet to the Box folder later and as before will be available for contact throughout the second session of ESS.

What is R? What is RStudio?

- R is:
 - Both a program and a programming language
 - Excellent for statistics & data analysis
 - Open source
- RStudio is:
 - An IDE (Integrated Development Environment)
 - Purpose-built for writing R code
 - Also open source
 - Let's take a look. . .

Some useful resources

- Google (no joke)
- RStudio Learn
- R for Data Science
- R bloggers
- RStudio Cheatsheets

Some R package recommendations

Type	Package name	Aim
Data import	haven	Similar to foreign
	readr	For reading in rectangular data files (e.g. CSV), part of the tidyverse
Data cleaning	tidyr	Tidy datasets incl. reshaping to long and wide, part of the tidyverse
Plots	ggpubr	ggplot2-based publication-ready plots
Standard errors	sandwich + lmtest	Robust and clustered standard errors + test coefficients
Modelling	plm	Panel data models (within-between estimator)
	margins orggeffects	Marginal effects from models
	AER	Various (Instrumental variable, tobit models)
	mlogit or mnlogit	Multinomial logit models
	lme4	Multilevel models
	rdd	Regression discontinuity design
Latent Variable	zoo or forecast	Time series tools
	psych	Exploratory Factor Analysis
	lavaan	CFA/SEM
Web scraping	rvest	Web scraping
	httr	Web scraping, APIs
	RSelenium	Scraping dynamic websites
Text Analysis	quanteda	Quantitative text analysis

Questions before we finish?