

# 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

# R package recommendations 1

Type	Package name	Aim
Data import	haven	Similar to <code>foreign</code> , built for the <code>tidyverse</code>
	readr	For reading in rectangular data files (e.g. CSV), part of the <code>tidyverse</code>
Data cleaning	tidyr	Tidy datasets incl. reshaping to long and wide, part of the <code>tidyverse</code>
Plots	ggpubr	ggplot2-based publication-ready plots
	dotwhisker	Easy coefficient plots in ggplot2
Standard errors	sandwich + lmtest	Robust and clustered standard errors + test coefficients
Modelling	plm	Panel data models (within-between estimator)
	zoo or forecast	Time series tools
	lme4	Multilevel models
	margins or ggeffects	Marginal effects from models
	MASS	Ordered logit models
	mlogit or mnlogit	Multinomial logit models
	AER	Various (Instrumental variable, tobit models)
	rdd	Regression discontinuity design

## R package recommendations 2

Type	Package name	Aim
Bayesian	rstan or rstanarm	Bayesian models via STAN
	rjags	Bayesian models via JAGS
	brms	Bayesian multilevel models via STAN with lme4 syntax
Latent Variable	psych	Exploratory Factor Analysis
	lavaan	CFA/SEM
	mirt	Item Response Theory
	mclust	Mixture modelling
Machine Learning	caret	Streamline the modelling process
	glmnet	lasso, ridge, elastic net regression
	rpart, randomforest	Regression trees, random forests
	e1071	Several misc functions including support vector machine
	neuralnet	Neural networks
Web scraping	keras	R interface to Python 'keras' library
	rvest	Web scraping
	httr	Web scraping, APIs
	RSelenium	Scraping dynamic websites
Text Analysis	quanteda	Quantitative text analysis



## Questions before we finish?