1X: Introduction to R

Essex Summer School in Social Science Data Analysis – University of Essex

Lorenzo Crippa

11 July, 2021

University of Essex – Department of Government

Today's session

Welcome! My name is Lorenzo, I am going to be your instructor for today. We also have Phil today, who will help us.

Today we'll introduce R.

Today's session

Welcome! My name is Lorenzo, I am going to be your instructor for today. We also have Phil today, who will help us.

Today we'll introduce R.
The goals of today's session are to:

Today's session

Welcome! My name is Lorenzo, I am going to be your instructor for today. We also have Phil today, who will help us.

Today we'll introduce R.
The goals of today's session are to:

0. Introduce what R and RStudio are

Today's session

Welcome! My name is Lorenzo, I am going to be your instructor for today. We also have Phil today, who will help us.

Today we'll introduce R.
The goals of today's session are to:

0. Introduce what R and RStudio are
1. Learn about data types, R objects, and how to create them

Today's session

Welcome! My name is Lorenzo, I am going to be your instructor for today. We also have Phil today, who will help us.

Today we'll introduce R.
The goals of today's session are to:

0. Introduce what R and RStudio are
1. Learn about data types, R objects, and how to create them
2. Learn how to import and manage datasets

Today's session

Welcome! My name is Lorenzo, I am going to be your instructor for today. We also have Phil today, who will help us.

Today we'll introduce R.
The goals of today's session are to:

0. Introduce what R and RStudio are
1. Learn about data types, R objects, and how to create them
2. Learn how to import and manage datasets
3. Learn how to generate basic plots and summary statistics

Today's session

Welcome! My name is Lorenzo, I am going to be your instructor for today. We also have Phil today, who will help us.

Today we'll introduce R.
The goals of today's session are to:

0. Introduce what R and RStudio are
1. Learn about data types, R objects, and how to create them
2. Learn how to import and manage datasets
3. Learn how to generate basic plots and summary statistics
4. Learn how to run basic statistical analyses

Session structure

Today's class is structured in two main sessions:

Session structure

Today's class is structured in two main sessions:

1. Data types, R objects, data manipulation (3h)

Session structure

Today's class is structured in two main sessions:

1. Data types, R objects, data manipulation (3h)
    1.1 Hands-on explanation
    1.2 Exercises

Session structure

Today's class is structured in two main sessions:

1. Data types, R objects, data manipulation (3h)
    1.1 Hands-on explanation
    1.2 Exercises
2. Plotting and basic summary statistics (2:30h)

Session structure

Today's class is structured in two main sessions:

1. Data types, R objects, data manipulation (3h)
    1.1 Hands-on explanation
    1.2 Exercises
2. Plotting and basic summary statistics (2:30h)
    2.1 Hands-on explanation
    2.2 Exercises

Session structure

Today's class is structured in two main sessions:

1. Data types, R objects, data manipulation (3h)
    1.1 Hands-on explanation
    1.2 Exercises
2. Plotting and basic summary statistics (2:30h)
    2.1 Hands-on explanation
    2.2 Exercises

A break between sessions one and two

Session structure

Today's class is structured in two main sessions:

1. Data types, R objects, data manipulation (3h)
    1.1 Hands-on explanation
    1.2 Exercises
2. Plotting and basic summary statistics (2:30h)
    2.1 Hands-on explanation
    2.2 Exercises

A break between sessions one and two

!! **Always** feel free to interrupt me during the presentation for questions or clarifications by raising your virtual hand. I'll try to check on them as often as I can.

Session structure

Today's class is structured in two main sessions:

1. Data types, R objects, data manipulation (3h)
    1.1 Hands-on explanation
    1.2 Exercises
2. Plotting and basic summary statistics (2:30h)
    2.1 Hands-on explanation
    2.2 Exercises

A break between sessions one and two

!! **Always** feel free to interrupt me during the presentation for questions or clarifications by raising your virtual hand. I'll try to check on them as often as I can. Worst case scenario: feel free to interrupt me while I speak.

0. Introduction to R and RStudio

What is R?

- A **programming language** developed specifically for **professional** data analysis (based on S).

What is R?

- A **programming language** developed specifically for **professional** data analysis (based on S). No point and click

What is R?

- A **programming language** developed specifically for **professional** data analysis (based on S). No point and click
- It is a powerful language. It is suited to two types of works:

What is R?

- A **programming language** developed specifically for **professional** data analysis (based on S). No point and click
- It is a powerful language. It is suited to two types of works:
  1. An interactive work; users employ packages and functions that were already written (today's session)

What is R?

- A **programming language** developed specifically for **professional** data analysis (based on S). No point and click
- It is a powerful language. It is suited to two types of works:
  1. An interactive work; users employ packages and functions that were already written (today's session)
  2. A programming work, more advanced; users write new functions and contribute to the language

What is R?

- A **programming language** developed specifically for **professional** data analysis (based on S). No point and click
- It is a powerful language. It is suited to two types of works:
  1. An interactive work; users employ packages and functions that were already written (today's session)
  2. A programming work, more advanced; users write new functions and contribute to the language
- R is free and open source, and there is an incredible community that uses the language and contributes to it

What is R?

- A **programming language** developed specifically for **professional** data analysis (based on S). No point and click
- It is a powerful language. It is suited to two types of works:
  1. An interactive work; users employ packages and functions that were already written (today's session)
  2. A programming work, more advanced; users write new functions and contribute to the language
- R is free and open source, and there is an incredible community that uses the language and contributes to it
- We need to have R installed on our computer to make it run

What is RStudio?

We run R in RStudio. What is RStudio?

What is RStudio?

We run R in RStudio. What is RStudio?

- It is an integrated development environment (IDE)

What is RStudio?

We run R in RStudio. What is RStudio?

- It is an integrated development environment (IDE)
- It automatically flags syntax errors and assigns different colors to different chunks of code

What is RStudio?

We run R in RStudio. What is RStudio?

- It is an integrated development environment (IDE)
- It automatically flags syntax errors and assigns different colors to different chunks of code
- The basic version (the one you'll need 99.999% of the times) is free

What is RStudio?

We run R in RStudio. What is RStudio?

- It is an integrated development environment (IDE)
- It automatically flags syntax errors and assigns different colors to different chunks of code
- The basic version (the one you'll need 99.999% of the times) is free
- We still need to have R installed on our computer in order for RStudio to properly run

1. Data types, R objects, data manipulation

2. Plotting and basic summary statistics

Extra topics

Useful R packages

Some R packages we haven't covered but that might be useful to some of you:

Useful R packages

Some R packages we haven't covered but that might be useful to some of you:

Type	Package name	Aim
Data import	`foreign`	Imports `.dta` files when `haven` doesn't
	`readr`	Imports files from various formats
Data cleaning	`reshape`	Turning long datasets into wide and viceversa
	`tidyr`	Tidy datasets
Plots	`ggplot2`	Plots based on Grammar of Graphics
	`lattice`	Plots, less versatile than `ggplot2`
Standard errors	`sandwich` + `lmtest`	Robust and clustered standard errors + test coefficients
	`estimatr`	Alternative to `sandwich` and `lmtest`
Modelling	`plm`	Panel data models (within-between estimator)
	`ggeffects`	Marginal effects from models
	`AER`	Various (Instrumental variable, tobit models)
	`MASS`	Various (ordered logit and probit)
	`mlogit` or `mnlogit`	Multinomial logit models
	`lme4`	Multilevel models
	`rdd`	Regression discontinuity design
	`zoo` or `forecast`	Time series tools

That's all

Thank you for the attention!

Lorenzo Crippa

l.crippa@essex.ac.uk