1X: Introduction to R

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

Lorenzo Crippa

12 July, 2020

University of Essex – Department of Government

Today's session

Welcome! Today we'll have an introduction to the R programming language.

Today's session

Welcome! Today we'll have an introduction to the R programming language.

The goals of today's session are to:

Today's session

Welcome! Today we'll have an introduction to the R programming language.
The goals of today's session are to:

0. Introduce what R and RStudio are

Today's session

Welcome! Today we'll have an introduction to the R programming language.
The goals of today's session are to:

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

Today's session

Welcome! Today we'll have an introduction to the R programming language.
The goals of today's session are to:

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

Today's session

Welcome! Today we'll have an introduction to the R programming language.
The goals of today's session are to:

0. Introduce what R and RStudio are
1. Learn about 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! Today we'll have an introduction to the R programming language.
The goals of today's session are to:

0. Introduce what R and RStudio are
1. Learn about 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 this way:

Session structure

Today's class is structured this way:

1. A hands-on presentation of the content material (1 hour long)

Session structure

Today's class is structured this way:

1. A hands-on presentation of the content material (1 hour long)
2. A series of exercises (1 hour long)

Session structure

Today's class is structured this way:

1. A hands-on presentation of the content material (1 hour long)
2. A series of exercises (1 hour long)
3. A review of the exercises run (2 hours long)

Session structure

Today's class is structured this way:

1. A hands-on presentation of the content material (1 hour long)
2. A series of exercises (1 hour long)
3. A review of the exercises run (2 hours long)
4. + An introduction to more advanced packages if time allows it

Session structure

Today's class is structured this way:

1. A hands-on presentation of the content material (1 hour long)
2. A series of exercises (1 hour long)
3. A review of the exercises run (2 hours long)
4. + An introduction to more advanced packages if time allows it

Break: We can decide together on a break when we are roughly mid-way through the class.

Session structure

Today's class is structured this way:

1. A hands-on presentation of the content material (1 hour long)
2. A series of exercises (1 hour long)
3. A review of the exercises run (2 hours long)
4. + An introduction to more advanced packages if time allows it

Break: We can decide together on a break when we are roughly mid-way through the class.

!! **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 this way:

1. A hands-on presentation of the content material (1 hour long)
2. A series of exercises (1 hour long)
3. A review of the exercises run (2 hours long)
4. + An introduction to more advanced packages if time allows it

Break: We can decide together on a break when we are roughly mid-way through the class.

!! **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 drag and drop

What is R?

- A **programming language** developed specifically for **professional** data analysis (based on S). No drag and drop
- 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 drag and drop
- 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 drag and drop
- 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 drag and drop
- 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 has a rather steep learning curve, but the passage from 1. to 2. comes quicker than expected

What is R?

- A **programming language** developed specifically for **professional** data analysis (based on S). No drag and drop
- 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 has a rather steep learning curve, but the passage from 1. to 2. comes quicker than expected
- 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 drag and drop
- 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 has a rather steep learning curve, but the passage from 1. to 2. comes quicker than expected
- 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 don't run R directly in it. We run it in RStudio.

What is RStudio?

We don't run R directly in it. We run it in RStudio. What is RStudio?

What is RStudio?

We don't run R directly in it. We run it in RStudio. What is RStudio?

- It is an integrated development environment (IDE)

What is RStudio?

We don't run R directly in it. We run it 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 don't run R directly in it. We run it 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 don't run R directly in it. We run it 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

What is RStudio?

We don't run R directly in it. We run it 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

Always run the code in RStudio **from a script** (.R extension).

What is RStudio?

We don't run R directly in it. We run it 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

Always run the code in RStudio **from a script** (.R extension). Don't run it from the console

1. Topic presentation (in RStudio)

2. Exercises

Instructions

- Now we'll run some exercises directly in RStudio to apply what we've seen before

Instructions

- Now we'll run some exercises directly in RStudio to apply what we've seen before
- You'll work individually to complete the script `intro_to_R_ex_blank.R` that I have provided you with

Instructions

- Now we'll run some exercises directly in RStudio to apply what we've seen before
- You'll work individually to complete the script `intro_to_R_ex_blank.R` that I have provided you with
- In case you have questions, please raise your virtual hand and I will create a separate breakout room

Instructions

- Now we'll run some exercises directly in RStudio to apply what we've seen before
- You'll work individually to complete the script `intro_to_R_ex_blank.R` that I have provided you with
- In case you have questions, please raise your virtual hand and I will create a separate breakout room
- I'll try to keep a list of raised hands in order not to miss anyone

Instructions

- Now we'll run some exercises directly in RStudio to apply what we've seen before
- You'll work individually to complete the script `intro_to_R_ex_blank.R` that I have provided you with
- In case you have questions, please raise your virtual hand and I will create a separate breakout room
- I'll try to keep a list of raised hands in order not to miss anyone
- If questions are relevant to all, I'll answer them in the classroom

Instructions

- Now we'll run some exercises directly in RStudio to apply what we've seen before
- You'll work individually to complete the script `intro_to_R_ex_blank.R` that I have provided you with
- In case you have questions, please raise your virtual hand and I will create a separate breakout room
- I'll try to keep a list of raised hands in order not to miss anyone
- If questions are relevant to all, I'll answer them in the classroom
- **Remember:** there is more than one way of doing what exercises ask!

3. Review of exercises

4. 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