ESCI upf.
International Business

P2 | Basic tools for data visualization

ggplot2

Marta Coronado Zamora
29 September 2023

Keep in touch

marta.coronado@prof.esci.upf.edu
@geneticament
@marta-coronado
Institut de Biologia Evolutiva (IBE-UPF-CSIC)

Session content

- Solving doubts
- Exercise: reproducing `ggplot2` graphics (`P2_exercises.Rmd`)
- Group project: part B

Practice

Reproducing `ggplot2` plots

- Open the document `P2_exercises.Rmd` in RStudio and complete the exercises.
- Upload the completed document to Aul@-ESCI at the end of the session.

Project

Group project

Parts

- **Part A** | Understand the origin of our data set and the meaning of the variables
- **Part B** | Visually describe our data set
- **Part C** | ❓

Project

Group project

Part A

- Describe your data set:
  - Where and why was the information collected?
  - Which is the meaning of each variable?
  - Do the variables have unit? Which one?
  - Does the data set have a long format?

Project

Group project

Part B

- Write the code to:
  - Read it into R
  - Reshape the data if necessary into long format
  - Check the variable classes and update them if necessary

Project

Group project

Part B

- Write the code to:
  - Read it into R
  - Reshape the data if necessary into long format
  - Check the variable classes and update them if necessary
- Explore your data using `ggplot2` graphics
  - Represent the **distribution of the variables**: pick one continuous variable and one discrete variable and use histograms or bar graphs to show their distribution
  - **Summarize the data**: use one geom to summarize data (e.g.: `geom_smooth`, boxplots, ...) of two variables
- Explain your data with graphics and text
  - Choose the **three graphics** that better describe your data
  - **Customize** and **annotate** them
  - Accompany the figures with your **hypothesis** and/or **interpretation**

Add everything (**tidy**) to the initial `R Markdown` document and **submit it before the next practical session** (one per group).