

WORKSHEET 3 Using R Studio

Aims and Objectives

This worksheet is designed to give you the opportunity to run an analysis in R Studio after watching and working through worksheets 1 and 2. By now you should have a working version of R Studio and have seen how to read-in data, set up a project, open a new markdown document and add code chunks and text.

To do the following exercise, you can use a dataset of your choice, perhaps in a topic that interests you. If you cannot find any dataset then we added a few to BBL so feel free to use one of those.

Please ensure that you set up a new R project and save the dataset (in .csv) in the folder that the project is set up on.

Some definitions:

Numerical data - can be expressed as a number or can be quantified. E.g. the weight of a person, number of students, shoe size and temperature in a room.

Categorical data - describes qualities or characteristics. E.g. gender, ethnicity and hair colour.

Exercises

- 1) Open R Studio and set up a new project pointing to the folder that contains the data you want to use.
- 2) Read in the .csv file and store it in an R data set (hint: worksheet 2).
- 3) In your data find a numerical column and use an R function to calculate the mean, median and mode.
- 4) In your data find a categorical column and create a frequency table for that column (hint: check the syntax in the RStudio help for table())
- 5) Use plot() to create a graph. Perhaps a scatter plot using two numerical columns. How would you describe this relationship? Explain what you observe below the code chunk.
- 6) Load the “ggplot2” package and try plotting using this library (hint: worksheet 2).
- 7) Add explanations and experiment with the different options for graphs.
- 8) Knit your markdown into an HTML file.