

Statistical Computing: descriptive statistics with R

Andy Buerki

Overview

Descriptive statistics aim to describe a data set, typically a sample of a larger population of interest, by summarising and visualising selected trends and features. Deriving descriptive statistics can be the goal of an analysis or, more often, a vital step in understanding the structure of data before the application of inferential statistics. The clear and accurate presentation of descriptive statistics is also of key importance in write-ups of research that features quantitative aspects.

This workshop focuses on using R to produce descriptive statistics from data sets, including high-quality plots and graphs that are of the right quality for submission to journals. The advantage of using R is that once the basics are mastered, it is very quick and easy to produce an array of high quality measures and graphs. R is today the tool of choice for quantitative linguists due to its power, flexibility and expandability. In this workshop we are going to use R through an interface called R Studio which facilitates an enhanced user experience.

Aims

By the end of the workshop,

- participants will have produced a set of key descriptive statistics measures and visualisations for an example data set provided.
- Using guidance on handouts provided, participants should be able to apply similar techniques to produce descriptive statistics for their own data sets.
- Participants will be in a position to judge the potential of R and therefore whether they wish to invest in learning more about using R for descriptive statistics.

Topics

2.10 - 3.00	Basics	<ul style="list-style-type: none">• Elements of the R Studio interface• Importing and exporting data into and out of R• Data manipulation in R: displaying, partially displaying, copying and creating data objects
	Descriptive statistics I	<ul style="list-style-type: none">• Data summarisation functions• central tendency and dispersion• checking distributions• frequency tables• layout, size formatting and file formats for graphs
3.10 – 4.00	Descriptive statistics II	Figures: <ul style="list-style-type: none">• scatterplots, bar plots, histograms, line graphs, pie charts, boxplots, interaction plots, etc.• adjusting scales, adding axis labels and legends, titles, regression lines, dot shapes and colours, etc.

Prerequisites

No prior knowledge of R is assumed, but good computer skills and previous knowledge of basic descriptive statistics will be very helpful.

Software installation

For the workshop, university laptops with R and R Studio pre-installed will be supplied. *To install R and R Studio on other university-owned computers running Windows*, there is an installer in Cardiff Apps > Cardiff Apps > School Applications > ENCAP . To install the software on any other computer, download and install, in this order, R (<http://www.stats.bris.ac.uk/R/>) and R Studio (<http://www.rstudio.com/>). Both R and R Studio are free.

Reading List

No preparation is required for the workshop, but for keen participants, I would recommend chapters 2, 3 & 4 of Levshina (2015) as an introduction before the workshop, and Chang (2013) as a follow-up and resource for further learning.

Levshina, N. (2015). *How to do Linguistics with R: Data exploration and statistical analysis*, Amsterdam: Benjamins.

Chang, W. (2013). *R Graphics Cookbook*, Farnham: O'Reilly.