Data sets in R

Loading a data set included in R

```
data()  # Provides a list of the data sets available in R via the 'datasets' package
```

A list with all the available data sets in the R package 'datasets' and descriptions of the individual data sets also be found by locating the 'datasets' package in the **Package** tab in the bottom right pane of the RStudio window and clicking on it or via the following link: http://stat.ethz.ch/R-manual/R-patched/library/datasets/html/00Index.html

Note: Not all data sets might be suitable for every type of analysis.

```
cars  # Prints the data set 'cars' from the 'datasets' package in the Console
cars <- cars  # Stores the 'cars' data set in an object called 'cars'

Titanic <- Titanic  # Stores the 'Titanic' data set in an object called 'Titanic'
mydataR <- as.data.frame(Titanic)  # Transforms the 'Titanic' data set from a table into a data frame
```

Downloading a data file

The data set used for this tutorial as well as a number of other ones can be found here: http://personal.strath.ac.uk/david.young/SQA/

Note: The file format for the data sets under the given link is .XLSX and the files need to be converted to .csv before they can be used. Alternatively, any other CSV file where the data is separated by commas can be used to try the steps in this tutorial.

Checking the file type

The file type for any file in Windows can be checked by right clicking on the file and opening "Properties". The file type and extension can be seen under "Type of file". In this case, the file type is the Excel spreadsheet format .XLSX. However, to import this Excel file into R it needs to be in the Excel format .csv.

Using Excel the file can be re-saved in the correct format as follows:

1. Double click on the data file to open it in Excel.
2. Click on the **File** tab and select **Save as**.
3. When the dialog window opens, select **CSV (Comma delimited)**
4. Click **Save**.

Without Excel, files can also be converted online. For example via the following website: https://convertio.co/xls-csv/

Importing CSV files (.csv) into R

This tutorial focuses on how to import data files in the Excel format CSV into R. However, the basic ideas can be applied to other data files as well and more information on importing those other formats along with the required code can be found here: http://www.r-tutor.com/r-introduction/data-frame/data-import

```r
# Read CSV file
mydataCSV <- read.table(file.choose(), header=TRUE, sep=",")
```

Note: An alternative command is read.csv(), which does the same thing as read.table() but has header=TRUE and sep="," as default settings, so that they do not need to be typed out in the code.

Checking a data frame in R

```r
mydataCSV
# Prints the data frame 'mydataCSV' in the Console

View(mydataCSV)
# Opens the data frame 'mydataCSV'as spreadsheet in seperate tab

names(mydataCSV)
# Prints a list of the variables in the data frame called 'mydataCSV'

sum(cars$speed)
# Calculates the sum of the variable 'speed' in the data set 'cars'
```