

Lab assignment: Basic R usage and data manipulation

Provide answers to the following questions in [a single r script](#) named 'r_lab_assignment.R' (or something similar to that effect), and upload in the designated assignment folder on Learn. I will assess the assignment as a whole, and then give pass or fail.

1. Type the code below within R to generate a vector called `vec1`

```
vec1 <- c(0, 2, 3, 0, 2, 11, 0, 7, NA)
```

- a) Remove the `NA` value, either by indexing or using the relevant function.
b) Make a logical vector (e.g., either `TRUE` and `FALSE`) indicating the elements equal to zero as `TRUE`, and the remaining elements as `FALSE`.
c) Use the logical vector you have created to pick out the non-zero values and store them in a vector called `no_zero`.
d) Check how many non-zero values you have in `vec1` by taking the length of the vector `no_zero` (using function `length()`)
2. In this exercise you will take data from a table and store the information in a data frame, that you then will export to a plain-text (.csv) file which is easy to work with in R, and any other statistical package.
 - a) Construct a data frame (manually in R) from the table given below, including the three variables which you name in R (remember that a vector in R is a column):
 - `W` (average wage/h, given by the numbers in the table)
 - `YEAR` (including the years for each observation, given in the table)
 - `GENDER` (including characters indicating "Female" or "Male" (hint: `rep()` is useful).The number of rows should be as many as the number of values in the table, e.g., the number of rows in your data frame should be 18!
Please note that the variable `GENDER` is based on the table below, but the character values "Female" and "Male" need be given in R.

- b) Export the data frame as a comma separated (.csv) file.
It is not necessary to supply this .csv file to me as a teacher for grading, but the code should be correct and work!

	2003	2004	2005	2006	2007	2008	2009	2010	2011
Average Wage/h									
Men	120	122	124	130	136	140	143	150	155
Women	109	112	115	121	128	132	135	140	148

3. For this exercise you need to download the data file "Freedman.csv" from Learn.

- a) Import the data to R as `data.frame()` named `Freedman`.
- b) Use `summary()` and `str()` on the imported data.
- c) Sometimes variables that are numeric is read as character or integer, but it is possible to use `as.numeric()` on these variables to define them as numeric. All variables except for the variable `City` should be numeric, make sure they are. (Hint: `Freedman$variable`, and assign new variable)
- d) Using this data set, give the mean values of each column of data (not `City`). Remember that `NA` values must be accounted for in function `mean()`
- e) Retrieve the rows which have a non-white population larger than 30%. (This is what the column `nonwhite` gives us as information, % non-white)

4. For this exercise we will use the `Prestige` data frame in the package `car`.

- a) Install the package, if you haven't done it already, and load the data frame. Read the help file for the data to learn about the variables.
`(?Prestige)`
- b) Select a subset of the data for occupations with more than 50% women and store this data.frame as an object `sub_prestige_women`.
- c) Use this subset and compute the average prestige score.
- d) Now compute the average prestige score for occupations with less than 50% women.
- e) For this final question below, use the complete Prestige data again (e.g., do **not** use `sub_prestige_women`).

Make a for-loop to compute the average (mean) prestige score for the three different types of occupations. Automatically store the three means in a vector.

(Hint: if you want a vector of the professions, there is a function `unique()`)