Introduction to R - practise exercises

This sheet contains your exercises for the first lesson. For this homework please complete all tasks below and copy and paste all of your answers into a word document or similar alternative. The ideal format for the document to hand-in is pdf.

Question 1.

Perform some simple mathematical calculations as listed below.

$$\sqrt{6*2}$$

4+3-2
1046*934

Question 2.

For this question we will utilise some existing data which comes with the R package MASS. This data contains the average heights and weights for a group of American women

(a) First you will need to load the data

data(women)

(b) Use the nrow() function to count the number of rows in the women dataset

(c) Use the colnames() function to display the column headers.

(d) You now need to add another column to the data, this column will include the women's ages. You can give each of the samples any age you like and you must label the column "ages". Print the new data frame to the terminal.

(e) Using the functions shown in session 1, what are the mean and sum of each of the columns?

An easier way to find some basic information about the data within a frame is this:

summary(women)

```
##
        height
                        weight
##
   Min.
           :58.0
                           :115.0
                    Min.
##
   1st Qu.:61.5
                    1st Qu.:124.5
   Median :65.0
                    Median :135.0
##
           :65.0
##
   Mean
                    Mean
                           :136.7
##
    3rd Qu.:68.5
                    3rd Qu.:148.0
##
   Max.
           :72.0
                    Max.
                           :164.0
```

Question 3.

For this question you will need to load a new data frame, perform some edits and re-save the data frame in a different file. For this you will need to download some data from the internet. Bioinformatics and data scientists often use free data from the internet, this is common practice. (a) download the following dataset to your computer and load it in R into variable cohort.

http://tbb.bio.uu.nl/BDA/fig4.tsv

cohort <- read.table("fig4.tsv", sep="\t", header=TRUE, stringsAsFactors=FALSE)</pre>

Your data frame should now look like this:

##		Name	First_Name	Last_Name	Age	Weight	Gender	Married
##	1	Patient01	Adriana	Mattos	35	64.5	F	TRUE
##	2	Patient02	Matthias	Uyttenhaghe	56	78.5	М	FALSE
##	3	Patient03	Jan	Aerntsz	34	50.3	М	TRUE
##	4	Patient04	Amin	Abboud	67	89.6	М	TRUE
##	5	Patient05	Janet	Thomlinson	31	87.5	F	FALSE
##	6	Patient06	Frederique	Vos	73	69.4	F	TRUE

(b) Make a subset containing only women (Gender == 'F'), add this to a new variable called "Females" and show your code.

(c) save data frame *females* into file 'fig4-females.csv' in a comma seperated table, show your code.

Question 4.

cohort

Now we know a little more about R and how to use a data frame we can look at other data types.

(a) What other types of data can R handle?

Try this website for some extra help with this: https://www.statmethods.net/input/datatypes.html

(b) What do these vectors look like when printed to the terminal and how would you select the first and third element from vector three?

a <- c(1,2,5.3,6,-2,4) # numeric vector b <- c("one","two","three") # character vector c <- c(TRUE,TRUE,TRUE,FALSE,TRUE,FALSE) #logical vector</pre>

Matrices are another form of data which R can handle.

(c) Make a matrix with four columns and four rows. Your matrix should have both column headers and row names. what does the matrix look like? Explain how the dimnames = argument works by searching the help files for the matrix() function.

```
# generates 4 x 4 numeric matrix
y <- matrix(1:20, nrow = 4, ncol = 4)
rnames <- c("R1", "R2", "R3", "R4")
cnames <- c("C1", "C2", "C3", "C4")
mymatrix <- matrix(y, nrow = 4, ncol = 4, byrow = TRUE, dimnames = list(rnames,cnames))
mymatrix</pre>
```

We have already used data frames in the above examples, both variables *women* and *cohort* were data frames.

(d) Use R code to make your own data frame, it should have at least three rows and each row should have at least three columns. Show the code and final result.

(e) Now change the coloumn headers to something else, show the code and result.

Question 5.

The final learning objective for today is to understand the concept of an R package and how to install one. Alongside the tidyverse set of packages R offers users the ability to write and publish their own packages. As such a multitude of packages can be found online.

One good website for exploring packages is the METACRAN site.

https://r-pkg.org

Download and install a package from this site and tell us what it does...