Introduction to R and RStudio part1 - homework

Anna Lorenc anna.lorenc@kcl.ac.uk

Challenge 1

1. Read the content of the file `mouse_small.csv` into an object mouse (function: `read.csv()`).

2. Check with `str()` what type of the data is in each column. We would like to avoid factors, so read the file in again, adding the parameter 'as.is=TRUE'. Check again what type of the data is in each column.

3. Check names of columns with `colnames()`. Why some of them contain dots? Compare with the column names in the csv file.

4. There is one funny column "X" containing only NAs. This is an artifact of how the spreadsheet program saved this data and we would like to get rid if it. The easiest way is to set this column to NULL, like: `dataFrame$column <- NULL`. Please do it.

5. To comply with publisher requirements you need to create a new mouse ID, something like MouseNumber, for example Mouse13. To automatically create such identifiers you can use the function `paste()`. Below are some examples how it works.

```
paste("My mouse is", "very sick")
```

```
## [1] "My mouse is very sick"
```
```
paste("My mouse is", "very sick", sep="")
```

```
## [1] "My mouse isvery sick"
```
```
paste("giraffe",c(1:4))
```

```
## [1] "giraffe 1" "giraffe 2" "giraffe 3" "giraffe 4"
```
```
paste(LETTERS[1:16], 1:2, sep="")
```

```
##  [1] "A1" "B2" "C1" "D2" "E1" "F2" "G1" "H2" "I1" "J2" "K1" "L2" "M1" "N2"
## [15] "O1" "P2"
```

To ad a new column you can use assignment like: `dataFrame$newColumn <- valueOfnewColumn`

6. Compute mean, variance and standard deviation of weights.

7. You need an additional column, showing mouse age at weighting, `age`. To obtain its values, it should be enough to subtract Birth.Date from Weighing.Date. Why it does not not work as `mouse$Weighing.Date-mouse$Birth.Date`?

To turn character data into date format, you can use the function 'as.Date()'. It requires specification of data format, as in examples below:

```
as.Date(c("2-June-16") , format="%d-%B-%y")
as.Date(c("2-June-16") , format="%d-%B-%y")
as.Date(c("2-4-1986") , format="%d-%m-%Y")
as.Date(c("2/4/1986") , format="%d/%m/%Y")
as.Date(c("2-Jun-16") , format="%d-%b-%y")
```

Turn the columns `Birth.Date` and `Weighing.Date` into date format. Add 'age'column with their difference. Bonus: add age_weeks column with their age in weeks. Hint: after division round the values with the function `round()`

8. Which rows have Btnl1-B12/Btnl1-B12 Min:+ve mice? Extract only rows with these. Save `mouse` as an R object and write out a csv file. Hint: function `which()` returns indexes of a logical vector which are TRUE, for example:

```
c(13, 4, 9, 5, 8) > 5
```

```
## [1]  TRUE FALSE  TRUE FALSE  TRUE
```

```
which(c(13, 4, 9, 5, 8) > 5)
```

```
## [1] 1 3 5
```

```
which(c("n","a", "d")=="a")
```

```
## [1] 2
```

Challenge 2

1. Read in the file `APCmin_Natalie_data` and add age column, as in Challenge 1.
2. Order mice in the table by names. Hint: use function `order()`.

```
order(c(23,14,5,6))
```

```
## [1] 3 4 2 1
```

```
c(23,14,5,6)[order(c(23,14,5,6))]
```

```
## [1]  5  6 14 23
```

```
order("spider","flea","butterfly","other_insect")
```

```
## [1] 1
```

```
c("spider","flea","butterfly","other_insect")[order(c("spider","flea","butterfly","other_insect"))]
```

```
## [1] "butterfly"    "flea"         "other_insect" "spider"
```

3. Save it as an R file.