Practical 1: Data manipulation in R

1 Getting started

Log-in to a PC in the usual way. Now open Rstudio – if you've not done this before, you mustn't have worked through the introductory practical sheet! Naughty! Go back and do that one first, it's on the course webpage/BB! Then come back here in half an hour's time...

Click on the "Start" button in the bottom left-hand-corner and search for "rstudio". Once you've opened Rstudio you should see four panels – if there is no text editor open in the top-left panel, click on the green button with a white plus symbol and open a new `R Script`. Remember – to execute a line of code in the text editor, hit `Ctrl-Enter`.

In the questions below, the important part is **understanding** what's going on rather than just typing the **R** commands. Make sure you answer the questions under each piece of **R** code. If you can't answer the question, then ask! Where possible, write your answer down on the worksheet. You should create the folder 'practical 1' in a directory called (e.g.) 'mas1802-computing' – this is where you can save your **R** session and any other files associated with this practical.

Work through *all* of the questions – the practice questions **and** the questions in the assignment! You will really struggle with the assignment questions if you don't work through the practice questions first. **And read through everything very carefully!**

2 Practice questions

You should try these questions before attempting the assignment.

1. In listing **??**, what type of variables are `x` and `y`? Why does the code produce an error message?

   Listing 1: Simple error message

   ```
   x = "TRUE"
   y = 2
   x*y
   ```

2. The **seq** command:

   Listing 2: The sequence command

   ```
   x = seq(0, 10)
   y = seq(0, 1001, by=0.5)
   z = seq(1000, 0, by= -exp(1))
   ```

   a) For each line in listing **??**, describe what **R** is doing.

   b) What is the length of each vector?

   c) Generate the vector: -1, -0.5,0.0, . . . , 10.

   d) Generate the vector: 100, 102, . . . , 206.

3. Vectors:

Listing 3: Vectors

```
#You will find out more about the runif command in a few weeks.
x = runif(5000, 1, 8)
```

Carefully type the code in listing 3 into **R** and then answer the following questions.

a) What is the length of `x`?

b) How many values in your vector `x` are below 2?

c) How many values in your vector `x` are above 7?

d) How many values in your vector `x` are below 3 or above 8?

e) How many times does the value 2 occur?

f) What are the $3400^{\text{th}} - 3402^{\text{nd}}$ elements in your vector?

g) What is the smallest value in `x`?

h) What is the largest value in `x`?

i) Create a new vector called `y` which contains all values of `x` that are between 2 and 6 (inclusive).

j) **Tricky:** Which value in `x` is the closest to 5, i.e.

$$\min(|5 - x_i|)?$$

You can use the **abs** to work out absolute values.

4. Data frames:

Listing 4: Dataframes

```
blob = data.frame(c1=runif(50), c2=rnorm(50), c3=runif(50))
```

Hint: Go over Section 2.5 of your lecture notes again before trying these questions!

a) What does the code in listing 4 do?

b) How many rows in `blob` are there where `c1` > 0.2?

c) How many rows in `blob` are there where `c1` > 0.2 and `c2` > 0.2?

d) How many rows in `blob` are there where `c1` > 0.2 and `c2` > 0.2 and `c3` > 0.2?

e) How many rows in `blob` are there where `c1` > 0.2 or `c3` < 0.5?

3 Assignment

You will complete the assessed work associated with this practical sheet within Numbas.

1. Open the Numbas application on any University cluster machine (or through RAS), as you do for other CBAs
2. Choose the MAS1802 module from the list of courses
3. Open the item 'MAS1802 - R Practical 1 - assessed'

This should open the Numbas test for this practical which you should complete by no later than **3pm, Friday 13th March**.

Unlike other Numbas tests, you can try this one as many times as you like! Only your last attempt submitted before the deadline will be marked! However, each time you completely re-start the test you must make sure you download (and save to an appropriate folder) the new datasets generated for each question (see points 4⟶6 below).

Some notes:

1. There are four questions to complete, each with multiple parts.
2. Each question requires you to download your own personal dataset. **Follow the instructions about how to do this, given within the Numbas test itself, really carefully!**
3. In the first lecture, you were told to make sure you keep your files organised – for example, you should create a directory called (e.g.) 'mas1802-computing', and then have folders for each practical (e.g. 'practical 1')
4. So, once you've clicked on the link to download your personal dataset for a particular question, you should save it in a convenient place (e.g. save the file in your 'practical 1' folder)
5. Then, you should change the current working directory of your Rstudio terminal to match the folder containing your saved files – by clicking on `Session`⟶`Set Working Directory`⟶`Choose Directory`, and then selecting the folder where you have stored your personal datasets (e.g. 'practical 1'; once you have chosen the folder, click `Select Folder`)
6. Once your working directory is the same as the downloaded file then you will be able to run the commands given in Numbas to load the file into **R**
7. To complete the assignment, you will use RStudio to answer the questions given in the Numbas test, so you will need to have both RStudio and the Numbas test open simultaneously.
8. As with any Numbas test, you can save your work, pause and log-out of the system, then resume the test at any time. But don't forget – all questions must be completed by **3pm, Friday 13th March**.
9. There is **no practice test**! But you can try the Numbas test as many times as you like... see above.
10. Deadline: **3pm, Friday 13th March 2020** :-)