Sergiu Buciumas

Submit through D2L’s dropbox area by 6PM, Monday, September 5th.

STAT 8030, Homework 1: All of these questions must be done using R (after all, this is an R programming course). Make sure to upload your **COMMENTED** code in D2L, and also provide *written answers* in this Word document for parts preceded with an asterisk (\*).

1) A) Use R to read in the Excel file available at <http://datadryad.org/bitstream/handle/10255/dryad.316/LA_data_Hereford.xls?sequence=1> (Note: I was unable to read this by specifying the URL; I had to save a copy of the .xls file and then read in the .xls file locally. You may do the same, but make sure the file is an .xls file—don’t resave it as a csv.



\*B) Use R to determine the number of observations and the number of variables in the data set. How many observations are there? How many variables?



In our data set are 1367 observations and 37 variables. We ca use dim() function to can find dimension of the dataset, in our case is a matrix 1367x37.

\*C) Order the data by publication year (pub\_year), then by author, then by *descending* site mean (site\_mean). With the data thus ordered, what is the value of site\_mean for the 800th observation? *[Hint: to have one of the variables in descending order while the others are in ascending order, don’t use the decreasing=option. Instead, look at the “Examples” section of ?order.]*

*a. order the data by publication year(pub\_year)*

**

*b. by author*

**

*c.* by *descending* site mean (site\_mean).

**

d. of site\_mean for the 800th observation



\*D) Report the minimum, median, and maximum of the site\_mean variable.



In our case min of site\_mean is = 0

Median is 0.895 and max value is 25741.67

E) Create a new variable named that is defined as follows:

it should equal 1 if site\_mean is less than 10, it should equal 2 if site\_mean is at least 10 but more than 100, it should equal 3 if site\_mean is at least 100 but less than 1000, it should equal 4 if site mean is at least 1000 but less than 10000, and it should equal 5 if site\_mean is at least 10000 but less than 100000.

(It is possible to do this in one line of code; for an extra challenge see if you can accomplish this. But it is not necessary for you to do this with only one line.)



\*F) Determine the distinct values of the site variable. How many distinct values of site are there?



and now the sum of distinct values:



\*G) Determine how many missing values of life\_cycle there are.

Life\_cycle column doesn’t have any missing data.



H) Create the data set with the following properties:

It should include only observations for which there are more than two authors (that is, the Author field contains \_et\_al or etal).

It should only contain the variables Author, rel\_fit, and sm\_magnitude

It should be sorted by Author, rel\_fit, and sm\_magnitude (each in ascending order).



1. Use R to create a .csv file of the data set created for part H. Also upload this .csv as part of your homework submission.



\*J) Read in the .csv file you just created as an R data set. Print out the first several rows, and compare visually with the first several rows of the data set from part H. (They should look the same.)



**Additional questions for which no R code is necessary:**

\*2) Explain why the result of as.numeric(as.factor(c(4.5,3.2,7))) is 2, 1, 3 rather than 4.5, 3.2, 7.

Because as.numeric will return the underlying numeric (integer) representation, which is often meaningless as it may not correspond to the factor levels.



Basically if we want to get output 4.5, 3.2, 7 that instead of as.factor will use as.character.



\*3) Explain why length(mtcars$mpg) is 32 but length(mtcars) is 11.



screenshot provided does explain differences, so basically length(mtcars$mpg) provide how many observations are in mpg variable/column. length(mtcars) provide how may variable does have mtcars.

\*4) Suppose I want to learn how to use the xtable function, but I have never installed the xtable package that contains this function. So I install the package and try to bring up the help page for the xtable function (see below). It doesn’t come up. What went wrong?

*>* ***install.packages("xtable")***

*trying URL 'https://cran.revolutionanalytics.com/bin/windows/contrib/3.2/xtable\_1.7-4.zip'*

*Content type 'application/zip' length 382663 bytes (373 KB)*

*downloaded 373 KB*

*package ‘xtable’ successfully unpacked and MD5 sums checked*

*The downloaded binary packages are in*

*C:\Users\bbarney2\AppData\Local\Temp\Rtmp6dOiV7\downloaded\_packages*

*>* ***?xtable***

*No documentation for ‘xtable’ in specified packages and libraries:*

*you could try ‘??xtable’*

***Answer: Because does require : library("xtable") before looking for help on the package.***

Bellow screenshot does peovide output when we are using library(“ ”)

**

**