DUE: February 12, 2019 at 11:59 PM

ANA1002 – Module 4 Assignment /20

Create and submit an R script which, when run, will print the answers to the following questions and output any graphics. Your R script must include a title with your name and student number and comments for each question number.

1. **(3 marks)** If you try to read in the file *addresses.txt* using the following command

```
addresses <- read.table("addresses.txt", header =
TRUE, sep = ",", quote = "", comment.char="",
stringsAsFactors=FALSE)
```

you will obtain the following error message:

```
Error in scan(file, what, nmax, sep, dec, quote, skip
, nlines, na.strings,  :
  line 1 did not have 6 elements
```

Why is R producing this error message? How would you change the *addresses.txt* file so that the above command properly imports the dataset? (Include your modified dataset)

2. **(6 marks)** Read in the SAS file "*medical.sas7bdat*". Export the data frame to a CSV file which contains only the two most highly correlated variables of the medical dataset. Call your file "*medicalexport.csv*".

3. **(11 marks)** The msleep data set contains the sleep times and weights for various mammals.
    a. Import the csv data from the following website:
    https://raw.githubusercontent.com/genomicsclass/dagdata/master/inst/extdata/msleep_ggplot2.csv
    b. Create a new object called m.subset and use the pipe operator to manipulate the data in the following ways:
        i. Group according to order and genus
        ii. Add a new variable for the proportion of REM sleep as a fraction of the total hours of sleep

iii. Summarize the average sleep_total, average rem_prop, minimum sleep_total, and maximum sleep_total

iv. Filter the data so that the average sleep hours are greater than 5

v. Arrange the data from lowest to highest by average sleep hours

c. Export the original data to a workbook called “mammal.sleep.xlsx” on a sheet named “Original”.

d. Export the m.subset data frame to the SAME excel workbook on a sheet called “Subset”.

Save your R Script as: Last Name, First Name Module 4

Upload your R Script to the “Module 4 Assignment” dropbox on Moodle before February 12, 2019 at 11:59 PM.