STAT 435

Name: William Franzen

<u>Lab Assignment #6 (*15 points*)</u>

<u>Notes:</u> ~~~~~~

- If you need help on an R function/command, type *?functionname* or *?commandname* and help for this function/command will appear in the Help window.
- Create a new folder on your desktop and name it **LA6_*your name***. Set this as your working directory in RStudio.
- In RStudio, open a blank source file (R Script) to work in, and make sure all History entries are cleared before you start your work on the following questions.
- In the R Script, add <u>*Lab Assignment-6 (by **your name**)*</u> as a comment line, and clearly label your answer to each question and part.

~~~~~~

1) Retrieve the file **`HCC.txt`** from **Data, Script, and other R Files** folder under **Course Documents** in the Moodle Course page and save it into your **LA6_*your name*** folder.

   a) Open the file and take a good look at it.

   b) Using **`read.table()`** function read the dataset into an R data frame, naming it **`HCC`**.

   c) Check the structure of your data frame.

   d) Convert Reason and Month variables into factors <u>using relevant labels</u>. See the **`HCC.txt`** file for the information needed.

   e) Check the structure of your data frame again.

   f) Use the **`head()`** command to look at the first five rows of the data frame. Use the **`tail()`** command to look at the last seven rows of <u>only</u> the first and third variables in the data frame.

   g) Write and run a command that will give you the average length of stay for the month of April. Report the result here: 13.64286

   h) Write and run a command that will give you the <u>longest</u> stay before a normal discharge in the month of January. Report the result here: 25
~~~~~~

2) Install and load the package **`dslabs`**.

   a) The **`murders`** is one of the datasets in **`dslabs`** package. Type **`?murders`** and take a good look at the documentation on this dataset.

   b) Using **`sink()`** function create a file named **sumstats.txt** in your **LA6*_your name*** folder.

   c) Inside R, add a line of text "Summary Statistics on No. of Gun Murders in the South" to the file you created in part b. Then, calculate summary statistics on **`total`** (Number of gun murders) variable for all the states in the South region.

   d) After leaving a blank line, add another line of text "Summary Statistics on No. of Gun Murders in the West" to the file. Then calculate summary statistics on **`total`** (Number of gun murders) variable for all the states in the West region.

   e) Use **`sink()`** function to close the **sumstats.txt** file. Open this file and make sure that the summary statistics have been directed to the file.

   f) Using **`write.table()`** function write the murders dataset into a new file called **murdersUS.txt** in your **LA6*_your name*** folder. Do not quote anything, do not name the rows, and use tab as delimiter. Open the murdersUS.txt file, look at it, and make sure the dataset has been written out properly.

3) Retrieve the file **`stu_surveys.pdf`** from **Data, Script, and other R Files** folder under **Course Documents** in the Moodle Course page and save it into your **LA6*_your name*** folder. In the file, you will find scanned copies of five student surveys.

   a) Create an empty data frame, naming it ***survey***.

   b) Using the **`edit()`** function, invoke the editor in **R** and enter the student survey data from the file **`stu_surveys.pdf`** into the ***survey*** data frame. Don't forget to name the variables as you are entering the data.

   c) Print the data frame ***survey*** and make sure you have 5 observations and 18 variables in it.

   d) Write the data into a CSV file naming it **survey.csv**. Check the file to make sure everything is in order. If not, fix it in R using the relevant arguments.

   e) Read the dataset into an R data frame, naming it **`survey_in`**.

   f) Display the structure of your data frame to make sure that you got everything right and that you have 5 observations and 18 variables.

g) Calculate the summary statistics on the age variable (in years). Report the results below:

Minimum is 20.08 and maximum is 32 years.

First quartile is 20.42 and third quartile is 24.17 years.

Median is 21.25 years.

Mean is 23.58 years.

h) Find the *percentage* of students who slept at least seven hours last night and report the result here: 60%

4) From the *Journal of Statistics Education (JSE)* of the *American Statistical Association* at http://jse.amstat.org/jse_data_archive.htm , find and take a close look at the data file **`BeefDemand.txt`** and its documentation file **`BeefDemandDoc.txt`** .

a) Read the data into a data frame, naming it **`beef.demand.`**

b) Only with data from years 1990 – 2000, create a new data frame naming it **`beef.demand.sub`**. Then, print it.

Save this lab assignment file with the blanks filled in, naming it **LA6_*your name***.

Save your RScript naming it **RScript_*your name***.

Email the professor the following five files at sgazioglu@mtech.edu :
LA6_*your name*, **RScript_*your name***, **sumstats.txt**, **murdersUS.txt**, and **survey.csv**

Have '**Stat435 – LA6**' in the subject line of the e-mail.