MET CS544 A1 Assignment 1

REQUIREMENTS:

You must work on your assignments individually. Sharing code or results with others is not allowed.

Please submit a document (word or PDF) containing a summary of your results and the R code used to generate your results. In your R code, insert comments when appropriate to increase readability. At bare minimum, you should insert question numbers before the code block that you use to answer the question.

Homework submission filenames should take on the form: LASTNAME_Homework #.doc(x)/.pdf and LASTNAME_Homework #.R.

Q1. Use the below R command to randomly sample 30 data points from beaver2$temp and assign them to a vector, **temp**. (60 points – 6 points for each sub-question)

```
temp <- sample(beaver2$temp, 30)
```

A. Display the vector, **temp**. Since values in **temp** is randomly sampled, it is necessary that you report the data values in **temp** so that we can evaluate your answers for the downstream questions.
B. Compute the mean and median of **temp**.
C. Convert the data in **temp** into integers and save them into a vector, **temp_int**. Compute the mode of **temp_int**.
D. Compute the variance and the standard deviation of **temp**.
E. Compute the five number summary, the interquartile range, and outliers, if any in **temp**.
F. Compute the standardized version (z-scores) of **temp**.
G. Identify values that are less than the first quartile of the data.
H. Using indexing, show the expression for accessing the first item and the last one of **temp**. The code should work for a vector of any size (no hard coding).
I. Create a matrix of size 5 x 6 using the data in **temp**. The first six values belong to the first row of the matrix. Assign the result to the variable, **temp.matrix**, and display the result.
J. Show the code for displaying the matrix's first and last columns. The code should work for matrices of any size (no hard coding).
K. Assign row names for the **temp.matrix** as Row1, Row2, Row3, and so on. Assign column names as Day1, Day2, …. The code should work for matrices of any size.

Q2. Using function data() to load R data set **airquality**. (40 points – 8 points for each sub-question)

© 2018 Trustees of Boston University. Materials contained within this course are subject to copyright protection.

A. How many variables are there? How many observations are there?

B. Are there missing values in the data set? If so, clean up the data by removing those observations with missing values and save the cleaned data set. Report the number of observations in the cleaned data set.

C. Show the mean, median, mode, and 1st and 3rd quantiles of variables **Temp** and **Wind** in the cleaned data set.

D. In the cleaned data set, how many days with a **Temp** higher than 78, and how many days with a **Wind** lower than 9?

E. The numbers in column **Temp** represent the temperature in Fahrenheit. Modify the cleaned data set by adding a new column, **TempC**, showing temperature in Celsius. Display the new resulting data frame. The equation to convert Fahrenheit to Celsius: $T(°C) = (T(°F) - 32) \times 5/9$

© 2018 Trustees of Boston University. Materials contained within this course are subject to copyright protection.