# Prep Exercise (PE04) Functions

### General Instructions

1. For this exercise you will answer all of the questions in this document and turn it in to Blackboard.
2. Before you get started make sure to read Chapters 8-10 of *An Introduction to Data Science* and execute the code throughout the chapters to gain familiarity.
3. Getting Started:
   1. The goal of this prep exercise is to practice writing functions. The functions will then be used in this week’s homework to sample from a data set several times and explore the meaning of the results.

# IST 687, Standard Homework Heading

#

# Student name: Thadhani Hitesh Chandrakumar

# Homework number: PE04

# Date due: 18th Sep 2019 11:59PM

#

# Attribution statement: (choose the statements that are true)

# 1. I did this work by myself, with help from the book and the professor

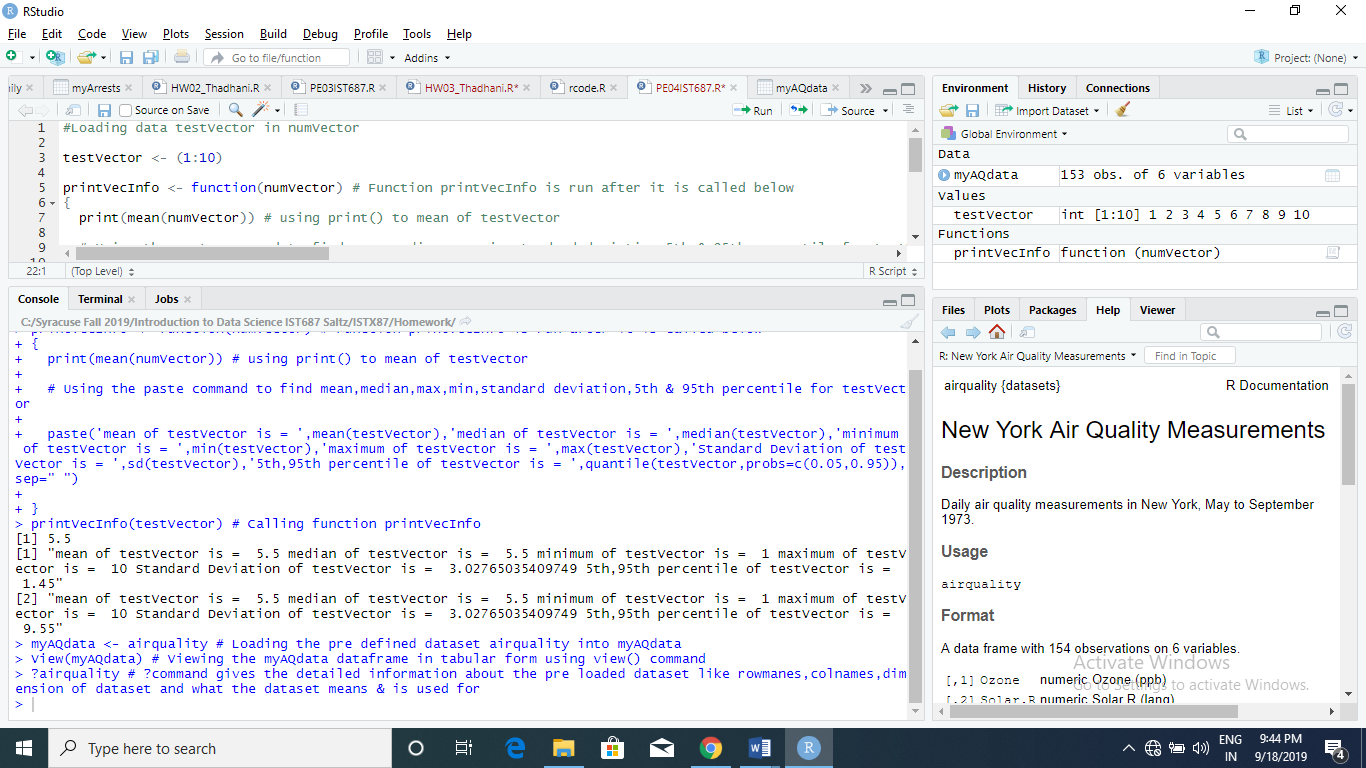
# 2. I did this homework with help from the book and the professor and these Internet sources: <provide the urls>

# 3. I did this homework with coaching from <Name of another student> but did not cut and paste any code

### Prep Exercise

1. **Create a Function**
2. Create a new function ‘printVecInfo’ and have it take one numeric vector as its input argument. Here’s a shell of the function:

printVecInfo <- function(numVector)  
{  
 # Use print( ) inside of a function when you want some output  
 # to appear on the console.  
 print(mean(numVector) )  
}

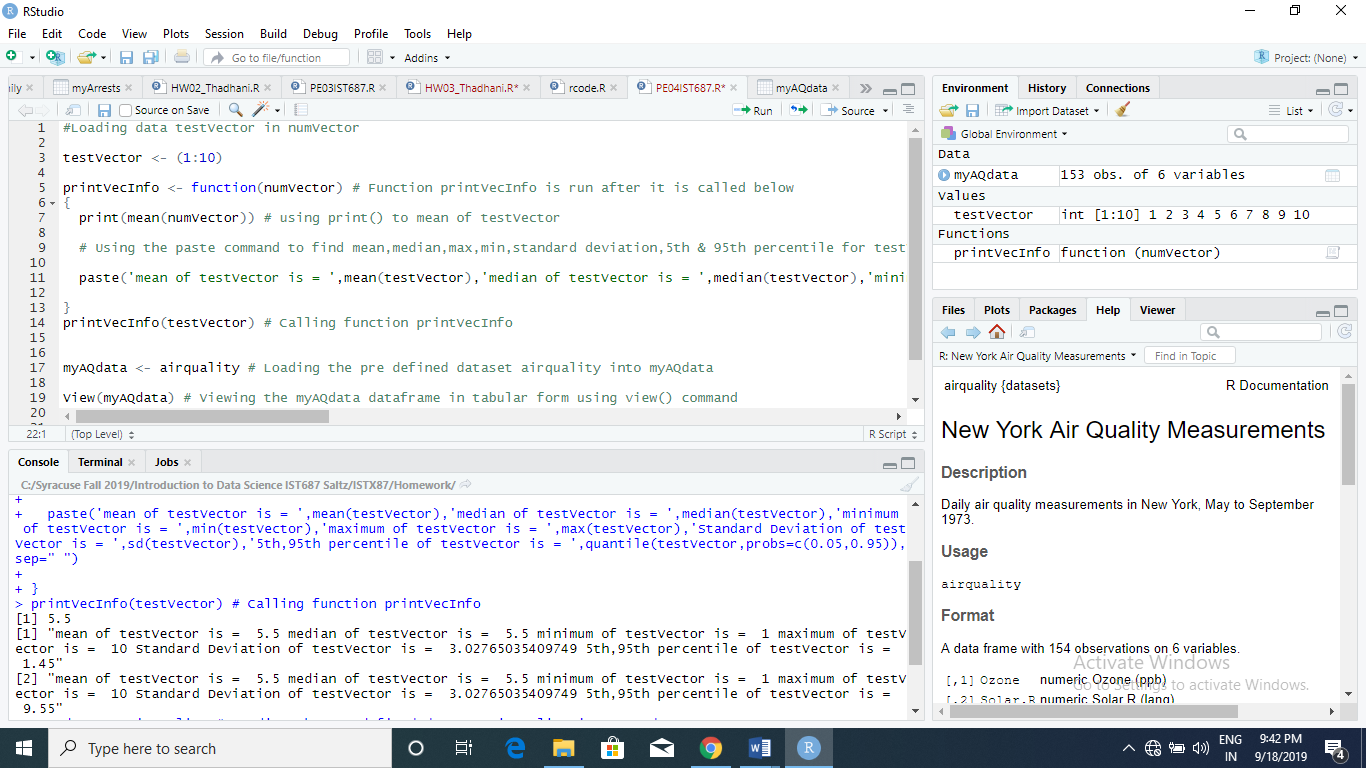


Function is defined in screenshot above

1. Make the function print all of the following information for the vector supplied in the argument:
   1. Mean (hey great that is already in the sample above!)
   2. Median
   3. Min and Max
   4. Standard Deviation
   5. 0.05 and .95 quantiles *Hint: use the quantile( ) function.*
2. **Let’s Test the Function and Add some Helpful Labels**
3. Test the function with this vector. testVector <- (1:10).
   1. Did you get 5.5, 10, 1, 3.02765, 5% 1.45 95% 9.55? If not, check your work.

Got the output as requested. Screenshot is attached. Used single paste command to get it rather than individual print command.

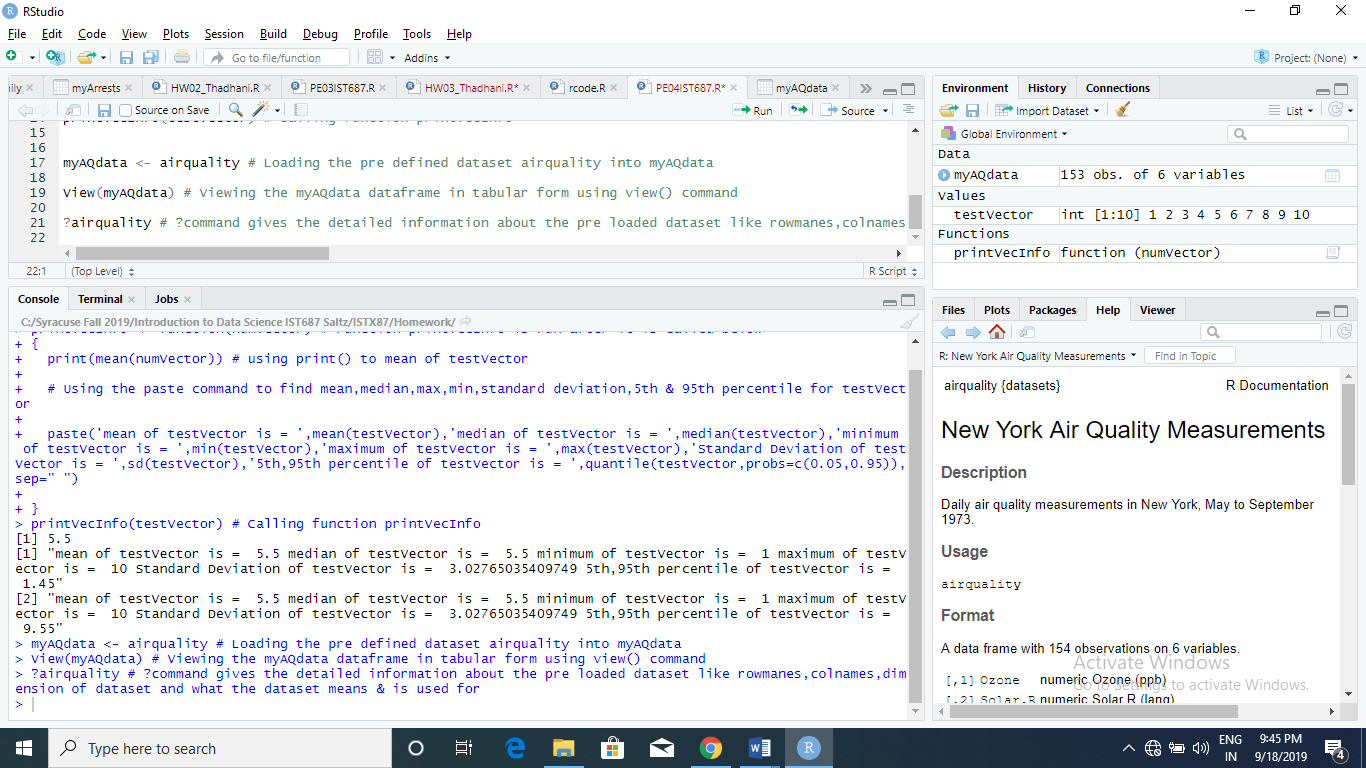
1. Now let’s make that output easier to read. Add labels to each element of the function’s output. *Hint: You might want to use paste0 in your function or switch from the print() command to the cat() command in your function.*
2. Repeat step 2.a again. Place a print screen of the results below.



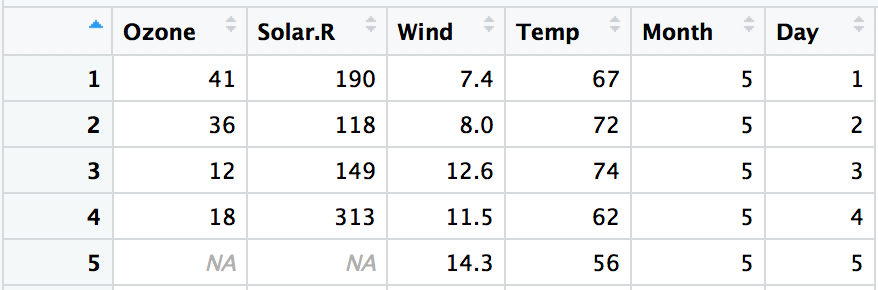
Loading the testVector to test the descriptive statistics values needed to be calculated with paste command

1. **Explore the dataframe!**
2. For this week’s homework we’ll practice sampling using a built-in data frame called airquality. Copy the dataframe into a new dataframe called myAQdata.
3. Using the VIEW command, explore the data. You may also want to use the ?command. You may even have to do additional research to help you understand the elements of the dataset.

? command is used to get more detailed information about the dataset pre-loaded into R Studio like number of observation & variables, column names, type of data and its features.



1. Please use your own words to describe the different elements of the dataset and their values based on the first 5 rows of the dataframe.

****The dataset is about the Air Quality of New York measured from May to September 1973. There are 154 observations and 6 variables such as Ozone, Wind, Temp, Solar for each day of the month. 1st 5 rows describe the measurements of Ozone, Wind, Temp and Solar radiation for May 1 to May 5 1973. All these variables values is taken to measure the Air Quality with Wind speed measured at LaGuardia Airport, Solar Radiation at Central Park and Ozone at Roosevelt Island. For the 1st row May 1 1973, Wind Speed is the lowest among the 5 days and Ozone in parts per billion is the highest. So by looking at the data for the 1st five days of May 1973 we can combine the 4 variables to see the Air Quality measured taking different air quality values into consideration at different times of the day.

1. **List any additional resources you used here.**

No additional resources used other than book Intro to Data Science by Saltz & Stanton.

1. **Be sure to save your R file as this will become the starting code for your homework.**

***You must submit all Prep Exercises to blackboard prior to the deadline specified for each assignment.*** PE assignments are due on the evening prior to the lecture class. Late PE assignments will not be accepted for credit.