# Prep Exercise (PE04) Functions

### General Instructions

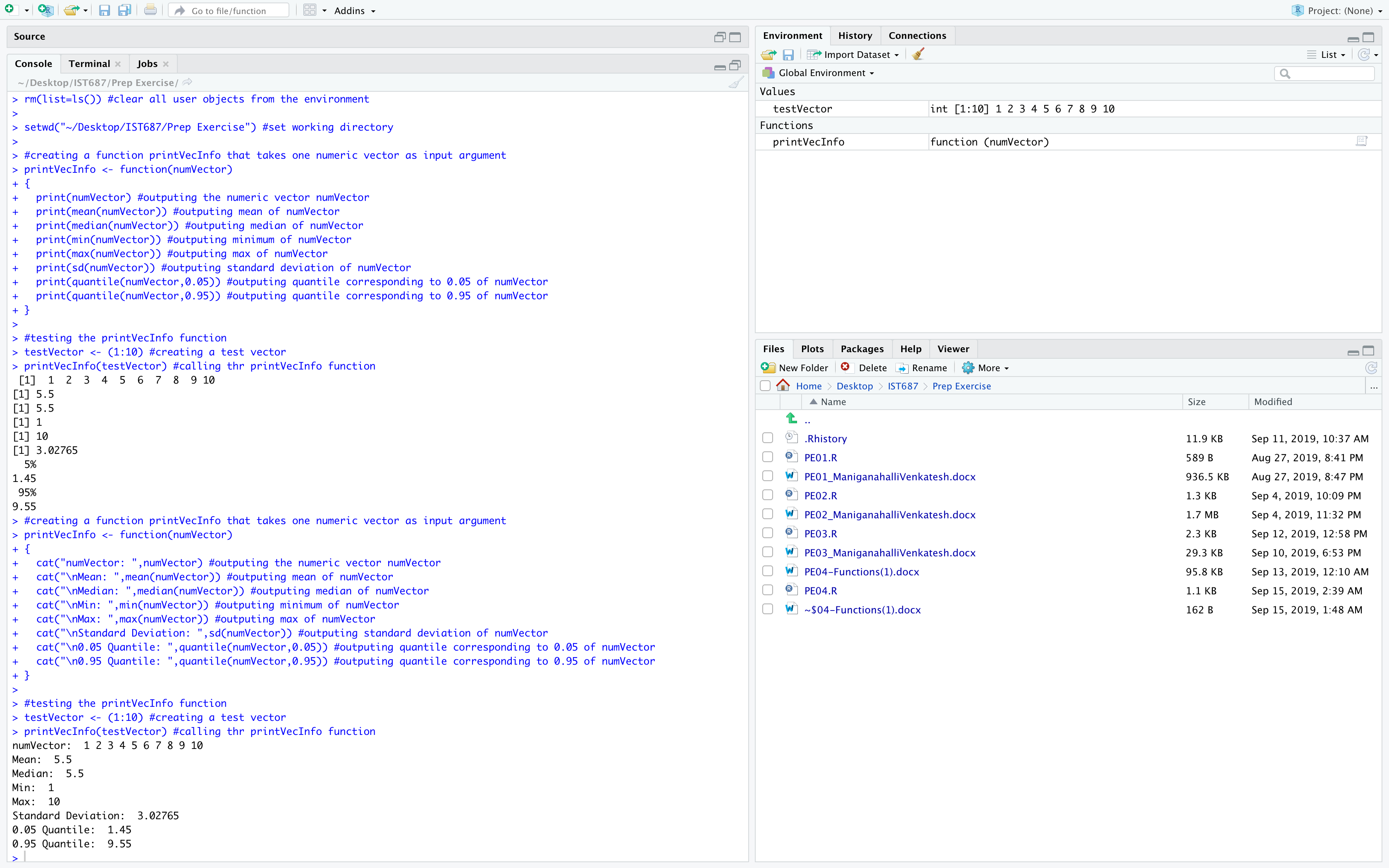
1. For this exercise you will answer all of the questions in this document and turn it in to Blackboard.
2. Before you get started make sure to read Chapters 8-10 of *An Introduction to Data Science* and execute the code throughout the chapters to gain familiarity.
3. Getting Started:
   1. The goal 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.

### 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) )  
}

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.
4. 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.*
5. Repeat step 2.a again. Place a print screen of the results below.



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.
4. 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 dataframe determines the daily air quality measures in New York state between May 1973 and September 1973.**

**Ozone is the mean ozone in parts per billion that is measured between 1pm to 3pm.**

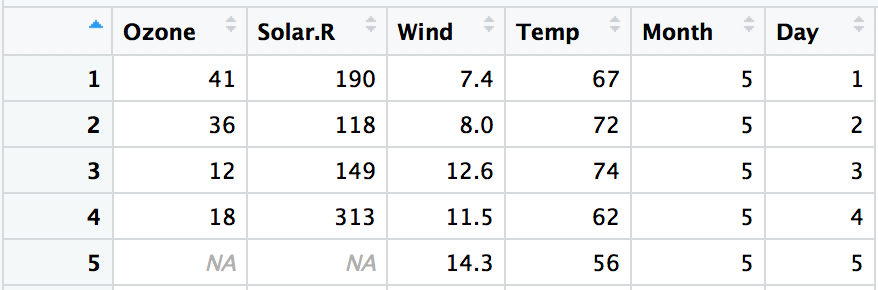
**Solar.R is the solar radiation in Langleys measured from 7am to 10am.**

**Wind is the average wind speed in miles per hour measured at 7am and 10am.**

**Temp is the maximum temperature during the day in degree Fahrenheit.**

**Month is the month number.**

**Day is the day of the month.**

****

**The first row in the dataframe shows the ozone, solar radiation, wind and temperature values recorded on May 1st.**

**On May 2nd ozone value was 41, solar radiation was 190 Langleys, wind speed was 7.4 miles per hour and maximum temperature was 67 degree Fahrenheit.**

**NA denotes the data that is not available or the data that was not measured.**

1. **List any additional resources you used here.**
2. **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.