Introduction to R for Data Analysis in the Health Sciences
BIOST 509
Homework Assignment #1
Due: 1:00 PM on October 4, 2017

Instructions

Submit your answers to the below questions in a R Script (.R), Word (.doc or .docx) or pdf file to Canvas. If you are waiting for a spot to open up in the course and do not have access to Canvas, save your responses and you can submit them when you enroll.

Include the R commands that you used to obtain your answers as well as the answers themselves.

Questions

All questions for this homework are related to an airquality dataset with daily air quality measurements from New York between May and September 1973. The following variables are included:

- Ozone: Mean ozone in parts per billion from 1300 to 1500 hours at Roosevelt Island
- Solar.R: Solar radiation in Langleys in the frequency band 4000–7700 Angstroms from 0800 to 1200 hours at Central Park
- Wind: Average wind speed in miles per hour at 0700 and 1000 hours at LaGuardia Airport
- Temp: Maximum daily temperature in degrees Fahrenheit at La Guardia Airport.
- Month: Number of month
- Day: Number of day during month

The dataset is available on the course canvas site (airquality.csv), and on the web at https://github.com/adw96/biost509/tree/master/datasets

First, download the file from canvas to your computer. Read in the dataset and store it as an object, using a sensible name.

1. How many rows and columns does the data set have?

2. What is the largest daily temperature? (Don't find it by looking through the data - use an R function!)

3. On which month and day did the highest temperature occur? (Don't find it by looking through the data - use an R function!)

4. What is the R command you used to identify the month and day of the maximum temperature?

5. Compute the following descriptive statistics for temperature.

Mean temperature:
Standard deviation of temperature:
Median temperature:

6. The ozone variable has missing values. What value do you get when you try to compute the mean of the ozone variable using R's default mean function? Why do you think R returned this value for the mean?
(We will see how to fix this problem later.)