**\* include your name and section (4677/6001) on your assignment.**

**Assigned Wed., Oct. 24; due Wed, Oct. 31**

INSTRUCTIONS:

You will be using the assigned dataset “**airquality.csv**” to “wrangle” some data (subsetting and indexing), and perform some descriptive statistics, and then submit a few graphs. **Ensure any graphs included in your homework have axes labeled, and ideally a title saying what the graph is.**

While you are free to do this in any software/language you feel comfortable with (e.g., Matlab), *I would highly recommend using Rstudio* (which is an application for running R) or Python (libraries: pandas, matplotlib), as both of these are free to use and have excellent documentation. If you are not familiar with any of the above languages, I would recommend you use Rstudio, since I am also providing a free ebook (see assignment page) that not only covers all the tasks we have to do, but actually is where the datasets are drawn from. (I.e., if you use R you will have more explicit instructions on how exactly to get the necessary answers to these questions using the data provided.)

For those of you who are curious, this data set is built into R, and details about what it is can be found here: <https://www.rdocumentation.org/packages/datasets/versions/3.5.1/topics/airquality>

**Submission types:** You may simply copy and paste text and images into a docx (or other plain text) file, or use LaTex to produce a pdf. You may alternatively submit a Jupyter or R notebook.

1. Import the airquality.csv file and save it as a data frame. (Unless you are using R or Rstudio, in which case you can simply type “airquality” and you will automatically have this data as a dataframe. To double-check, type “data.class(airquality)” and you should see the output as “data.frame”) **Next, tell me what language/software you have used for this assignment:**
2. Calculate the mean, median, and mode for the wind speed and temperature variables. Write your results into a table (however you want – e.g., you can just use MS word) like so:

|  |  |  |
| --- | --- | --- |
|  | **Wind speed** | **Temperature** |
| **Mean** |  |  |
| **Median** |  |  |
| **Mode** |  |  |

1. Make a plot (bar or dots) of the solar radiation values, and a plot of the Temperature values. Include your plots in this assignment. Which one is closest to resembling a normal distribution?
2. Calculate the mean, the range (i.e., maximum and minimum values), and the standard deviation of the ozone values.
3. Does 95% of the data fit within 2 standard deviations of the mean? Show your work.
4. What was the lowest temperature in May? (Show the code you used to find the answer).
5. Calculate the mean for wind speed and temperature as a function of month. Plot these results as either a bar plot or line plot (ideally on one single plot. Two plots are accepted if you are really struggling to figure out how to do that.)