CE 5315 Assignment # 1
Basic Statistics and R Programming
Due June 27th 2022

Download Air Quality data from CAMS 15 (Channelview, TX). You can obtain this data from the Texas Commission on Environmental Quality - https://www.tceq.texas.gov/cgi-bin/compliance/monops/monthly_summary.pl?cams=15

You need to download Data from Oct 1., 2020 – Sep 30., 2021. You will need to obtain the following information (on a daily basis).

1. Ozone
2. Resultant Wind Speed
3. Resultant Wind Direction
4. Outdoor temperature (F)
5. Dew Point Temperature
6. Relative Humidity
7. Solar Radiation
8. Oxides of Nitrogen

Compile the data into a single csv file. (Note you can download data on a monthly basis from the web site). Replace missing data as NA. You can either use R or a spreadsheet to make this file.

Use the data to obtain the following:

1. Plot an empirical CDF of maximum daily Ozone using Beard Plotting position formula.
2. Plot a histogram of mean daily temperature using Freedman-Diaconis binning formula.
3. Write a function to compute that takes a vector of daily data for a variable and returns the 1^{st} ,5^{th}, 10^{th}, 25^{th}, 50^{th}, 75^{th}, 90^{th}, 95^{th}, $99^{th.}$ Use the function to compute these summaries for Daily maximum values of the oxides of nitrogen. Compare it with the results you obtain by manual calculations.
4. Obtain correlations between Daily max Ozone and Daily min Dew Point Temperature and Maximum daily solar radiation and maximum daily wind speed and daily maximum concentration of oxides of nitrogen and create a correlation plot. Use Spearman Rank Correlation Coefficients.
5. Use 90^{o} Quads of Wind Direction and compute the probability of Ozone being less than or equal to 10 ppm or higher. Create a contingency table showing joint, marginal probabilities. Calculate the probability P(O3 > 10 | WD = N (0 - 90])

Notes:
You can use either R GUI, R studio or Jupyter-Notebook for your analysis.

Your submission should have 3 parts – 1) The data (csv file that you created); 2) A summary document having all the necessary Figures and results (as a Table). Please ensure your tables and figures are labeled appropriately. 3) A listing of R codes which are neatly and completely documented. I should be able to take the R code and data provided by you to obtain your results.

Submit via Blackboard as a zipped file that is named using the following format “lastname-Assignment1”.

Scoring Rubric: **Data Compilation 20 pts**; **Q1 – Q5 80 points (16 points each).**