Episode 8 Homework

High levels of ground ozone are considered harmful for humans and also exert negative effects on vegetation and ecosystems. Research groups are attempting to define models that correlate various environmental factors (temperature, wind speed, etc.) with ground ozone levels, with the future goal of creating forecasting models that predict periods of critical ground ozone levels based on relevant easily measurable environmental factors.

Download the data files for this episode. There are 2 data files (.txt) and one Jupyter Notebook file, pandas_demo.ipynb. The data files contain daily measurements of various environmental factors over a period of several years in the Houston, Texas area.

Launch Jupyter Notebook from the Anaconda Navigator. The Notebook will open as an application running in your web browser. Go to **File > Open…** and open the downloaded pandas_demo.ipynb file.

This notebook is written for Python 2.7. The notebook is organized in a sequence of cells, where each cell either contains documentation or a short Python code snippet plus its generated output.

- **Code cells** are marked by **In [some number]:** and **Out [some number]:** on the left margin.
- **Documentation** cells do not have any left margin label but often begins with bold face heading.

Code cells can be run all together or individually. To run all cells (in sequence from top to bottom), go to **Cell > Run All**. To run individual code cells, select as single or multiple cells and click on **Cell > Run Cells** in the menu.

Note: Some cells may rely on having cells above to be run before, e.g. in order to execute required import statements or populating variables, i.e. data. **It is best to either run all cells together first, or execute the cells individually from top to bottom.**

1. Run the Jupyter Notebook and explore how changes to the code affect selection, grouping and plotting of the data.

2. Open Spyder and write a Python script that includes:

    a. A function that calculates and returns the number of ozone days (column ‘Ozone day’) for each year.

b. A function that calculates and returns the number of ozone days for the entire observation period grouped by calendar month.
c. A function that groups the data by days with daily precipitation >0.5 and <=0.5, and whether or not the day was an ozone day. The function should return all column values for these 4 groups.
d. Create functions that neatly output the return values of the 1a, 1b functions.
e. Create a function that creates and prints a summary statistic with min, max, mean, std, median and values for the wind speed columns (WSR0 to WSR9) of the grouped data returned by function 1c.

3. Add the following plotting functionality to your script
    a. Add a function that creates a bar histogram for the output of 2a and 2b. The plot should include appropriate title, legend and axes labels.
    b. Create a single scatter plot that plots the wind speed (column 'WSR0') against the daily temperature ('T0') for all groups, where each group is plotted in a different color. The plot should show the appropriate labels for x and y axes.