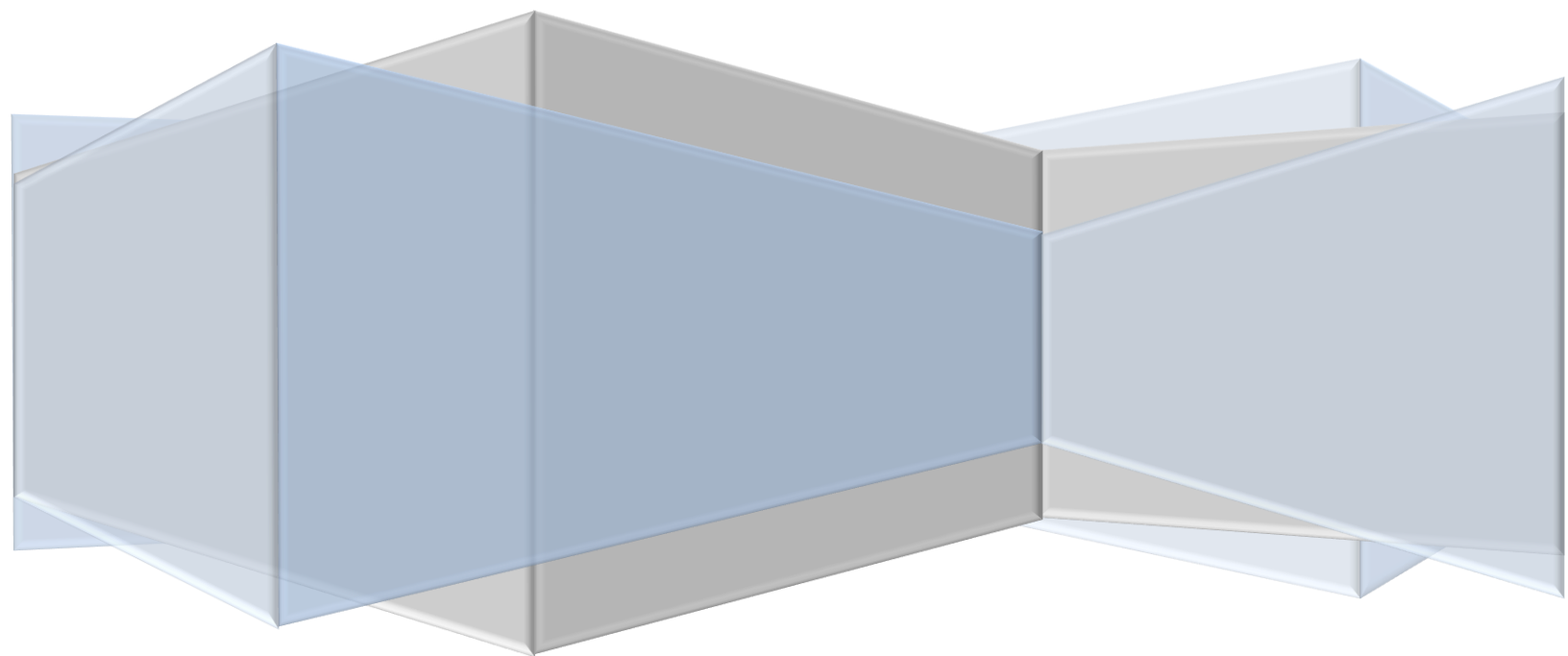


Rutgers Data Science Boot Camp

Python-APIs

Homework Assignment 06

Mark Visco



Assignment

Using the dependencies below, randomly selected over 500 unique cities based on latitude & longitude and then built a series of scatter plots to showcase the following parameters compared to latitude: Temperature, Humidity, Cloudiness and Wind Speed.

Dependencies

- Jupyter Notebooks
- Matplotlib
- Pandas
- Citipy
- API URL
- API Key

Methodology

In order to complete this homework assignment, I first had to access the OpenWeatherMap.org website. This website provides current weather, forecasts and historical data to developers of web services and mobile applications. In order to access this data I will need an api (application programming interface) key, which basically unlocks the door to the website's data.

Once I entered the website I was able to request an api key by following the instructions found in the "API" section. From the available options, I decided to go with the Current weather data. This made the most sense since it provided a snapshot in time of the weather spanning the entire world. To get my unique sample of over 500 cities I utilized Citipy. Instead of scrolling through the OpenWeatherMap data and getting my sample, I used Citipy which only contains the city names with geo_coordinates.

To ensure my coding would produce over 500 city names, I set my size parameter to 1,500 which ended up returning 597 unique city names. To prevent the code from stopping due to an error with the city names, I utilized Try/Except/Continue in my for loop statement. By doing this my final unique sample count was 542 unique cities.

Out[5]:

	City	Cloud Cover	Country	Date	Humidity	Latitude	Longitude	Max Temp	Temperature	Wind Speed
0	Sawankhalok	40	TH	1561524886	74	17.31	99.83	97.00	92.28	5.82
1	Atuona	0	PF	1561524547	77	-9.80	-139.03	80.87	80.87	19.01
2	Punta Arenas	90	CL	1561524548	96	-53.16	-70.91	32.00	32.00	3.36
3	Sibolga	48	ID	1561524665	71	1.74	98.78	82.85	82.85	2.77
4	Shitanjing	100	CN	1561524886	81	39.23	106.34	60.01	60.01	5.99
5	Ushuaia	75	AR	1561524510	92	-54.81	-68.31	30.20	30.20	5.82
6	Hobart	0	AU	1561524337	50	-42.88	147.33	57.20	56.23	8.05
7	Busselton	46	AU	1561524886	82	-33.64	115.35	63.00	63.00	21.00
8	Bathsheba	20	BB	1561524886	74	13.22	-59.52	80.60	80.60	19.46
9	Hilo	40	US	1561524886	58	19.71	-155.08	82.40	76.06	12.75
10	Arraial do Cabo	0	BR	1561524887	76	-22.97	-42.02	73.31	73.31	10.98

Note: I initially set the size parameter to 10 while I was developing my code and once I was successful in getting 10 returns, I increased the number to 1,500.

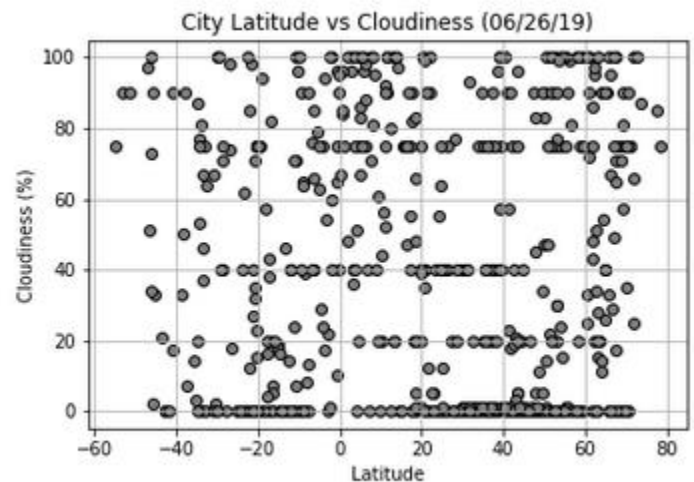
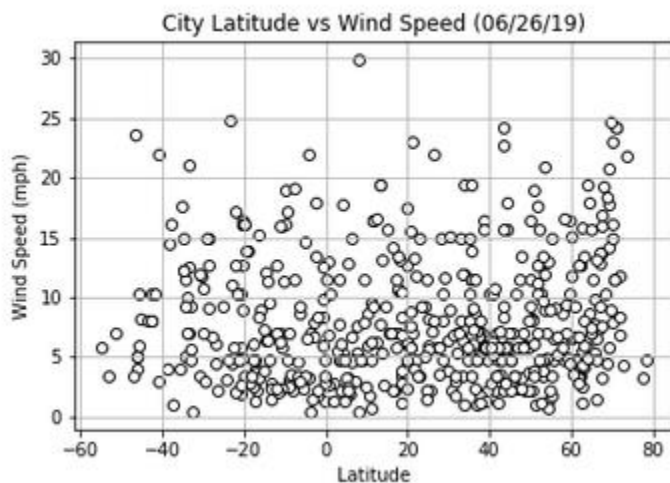
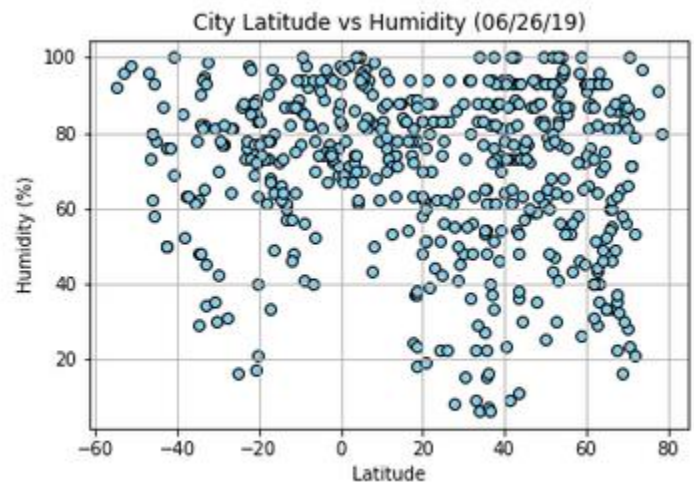
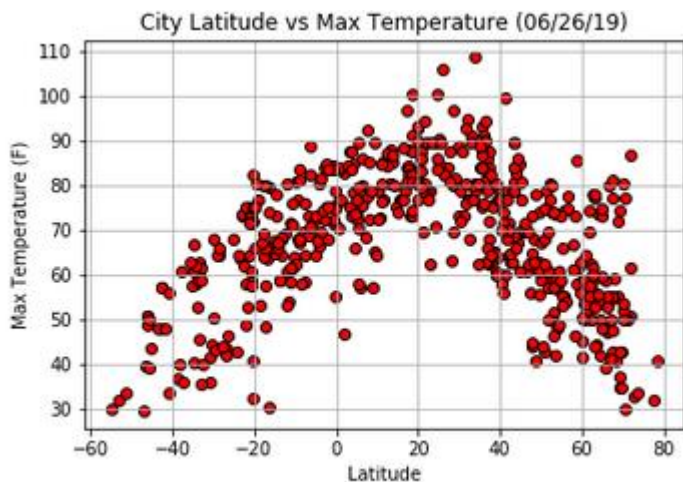
Plotting the Data

The last part of the assignment was to plot the data. Since the website refreshed the data every 10 minutes, I ran my list of unique city names against the website data and then created a CSV file which I saved and then used to create my four scatter plot diagrams. This way I was ensuring I was comparing the same exact cities and results in all four graphs.

I used different colors in each of the scatter diagrams to give a visual representation of the results:

- Red = Maximum Temperature
- Gray = Cloudiness
- White = Wind Speed
- Blue = Humidity (not really sure which color represents being sweaty/uncomfortable)

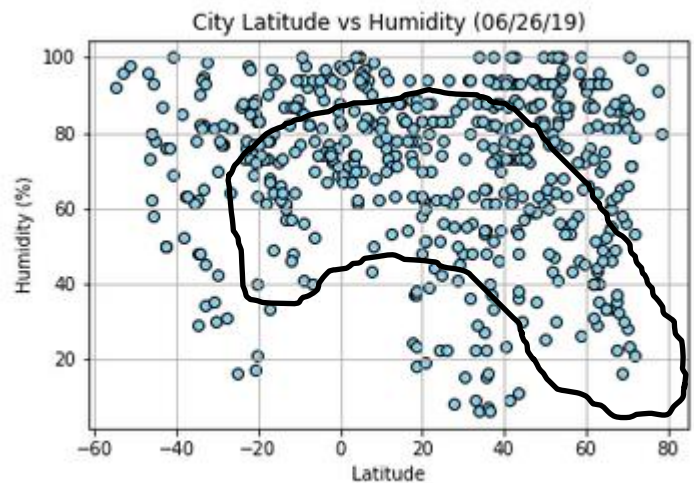
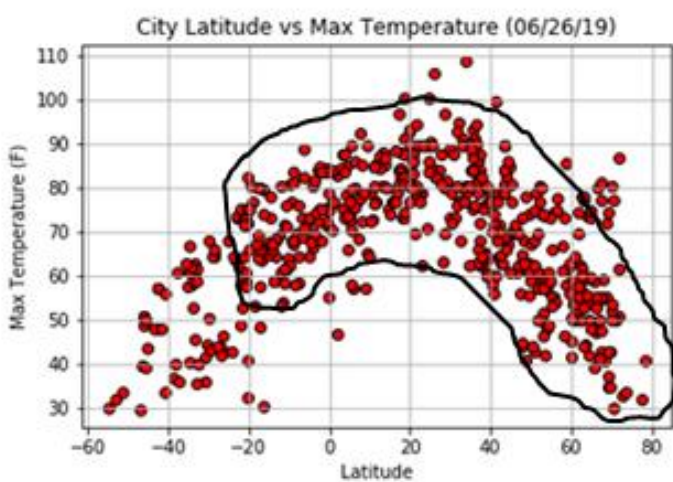
Output



Observations

Upon looking at the four graphs the first thing that stood out was the Humidity and the Wind Speed scatter plot diagrams. The majority of the cities were experiencing humidity levels greater than 40-50%. On very humid days the air outside is very stagnant and you hardly feel any type of breeze. Looking at the Wind Speed scatter plot the majority of the cities were experiencing Wind Speeds between 0 to 10 mph.

With regard to temperature, if I capture the majority of the cities and circle them and then superimpose that over the humidity scatter plot you can see a strong relationship of the high temperature and the higher levels of humidity



The final observation is noted on all four graphs. When I generated my Cities list using the CitiPy, I specifically made sure to include the Latitude Range of -90 to 90 and the Longitude Range from -180 to 180. However, if you notice on the graphs, the cities only range from -60 to 80 because, after 50 degrees south, there aren't any cities but rather stations or research centers. Interestingly enough, McMurdo Station (the farthest station in the south) is operated by the United States and supports 1,258 residents.

