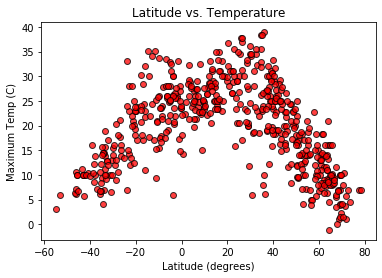
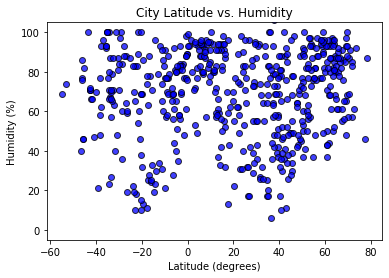
A multi-scatterplot analysis of weather conditions was performed on over 500 cities using a random number generator from the Numpy library to randomly pick geo-coordinates across the earth. CitiPy was then used to find the nearest city to the randomly generated coordinates, thus creating a list of cities.

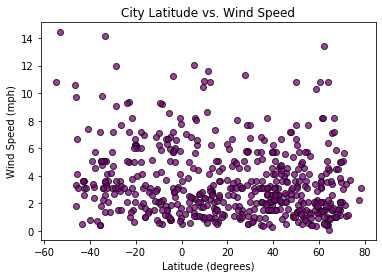
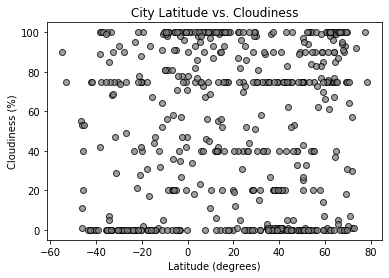
The city list is then fed into a looped URL query using the Openweather API, where weather station data was collected for all (most) cities. The data was then written into a dataframe where Matplotlib was then used to plot scatter charts of humidity, cloudiness, windspeed and maximum temp by latitude.

During the data collection & display, several key observations on data cleanliness were made.

1. Erroneous station reporting is possible, particularly when reporting % humidity or cloudiness. Specifically, a problem was encountered on the humidity side, whereby a reading of 326% was passed back. The false reading distorted the output plot of humidity vs. latitude, by compressing most of the data into a narrow range, while an outlier was visible at > 300%. A preventive measure was put in place on both humidity and cloudiness scatter plots by defining a fixed YLIM range.
2. Key and Index errors appeared repeatedly, whereby CityPy could not accurately assign a nearby city to the randomly selected coordinates. This situation required a Try / Except pair of actions. The number of exceptions limited the number of observable city weather data points, but was still above the threshold of 500 in most cases.

The key scatterplots are shown below.



The maximum temperature plot follows common sense, showing maximum temperatures being higher near the equator (0 degrees) and falling off approaching +/- 90 degrees.

Humidity vs. longitude is a more nuanced observation. The higher humidity is more concentrated in the northern latitudes, while the lower humidity spots are less concentrated, particularly in the southern hemisphere. The most likely explanation is that a higher concentration of cities occurs in more humid locations (often correlating to large body-of-water proximity or access), whereas the lower humidity deserts are sparsely populated, particularly those in the southern hemisphere (latitude < 0).

Cloudiness presented a different challenge. There appears to be minimal correlation to latitude. However, when secondarily comparing with the Humidity chart, it suggestes the more humid and cloudy areas might be more correlated in the northern hemisphere. To directly check this relationship, another scatterplot was generated to view Cloudiness vs. Humidity. As expected, the higher rate of cloudiness correlates well to the higher level of humidity.

Similarly, Wind Speed appears to concentrate lower windspeeds with higher latitude, and it’s tempting to speculate that Windspeed and humidity are inversely correlated. Sure enough, a dedicated scatterplot confirms this hypothesis.

