Weather Analysis and Findings

Alexis Perumal, 12/28/19
UCSD Data Science Bootcamp, HW#6 Python-API

Summary

Analysis attempting to understand the correlation between latitude and four weather indicators (temperature, Humidity, Cloudiness and Windspeed).

Approach

Analysis was done by hitting the OpenWeather API REST endpoint and extracting real time weather data for 500+ cities randomly selected from around the world.

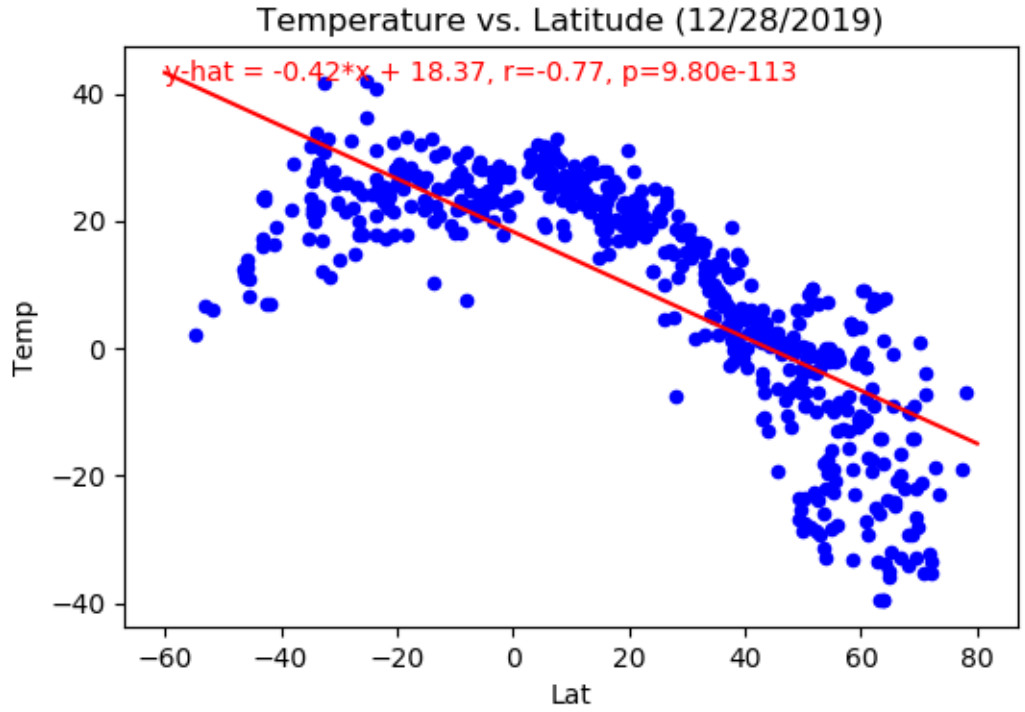
Findings

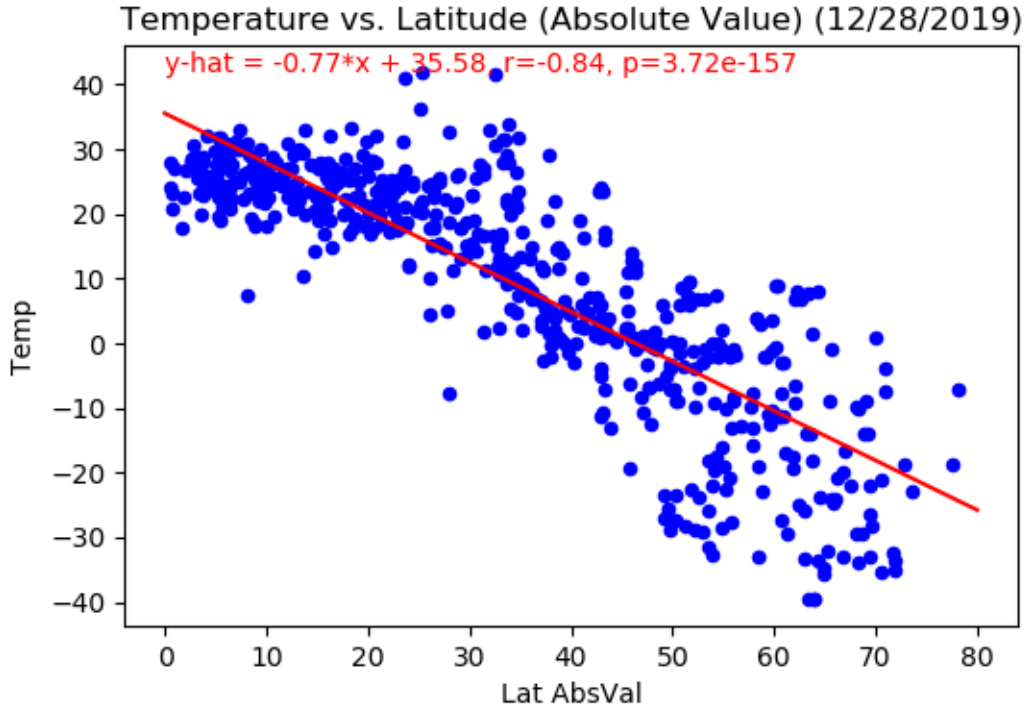
1. **Temperature correlates inversely with latitude with statistical significance.**

   Statistical significance is very strong p=9.8E-113 and the inverse correlation is also strong (r=-0.77, m=-0.42 degrees C per degree of latitude).

   Statistical significance is even stronger, p=3.7E-157 and the inverse correlation stronger (r=-0.84, m=-0.77 degrees C per degree of latitude).

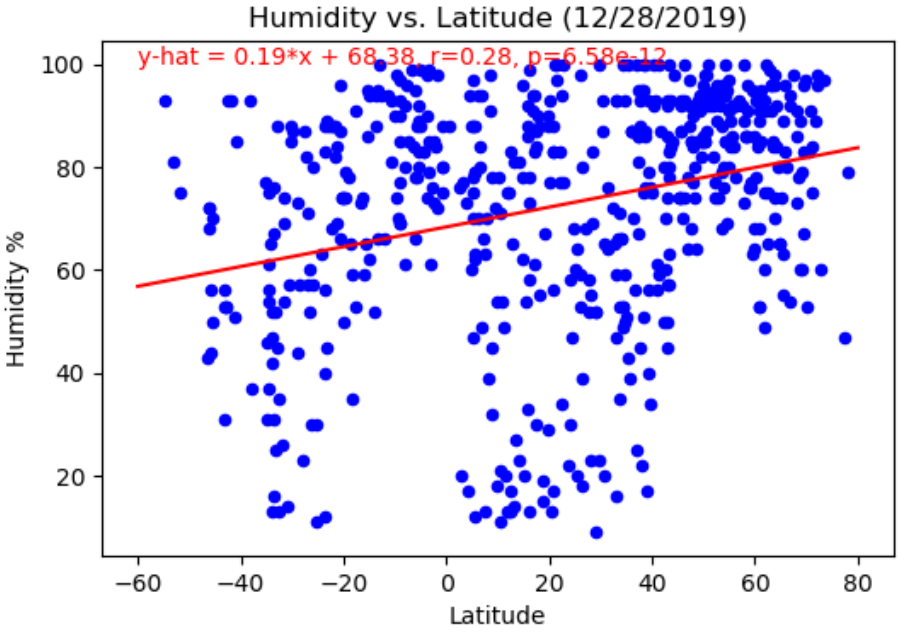
   Using the absolute value of latitude is preferred because we would expect temperatures to be colder closer to the poles. With fewer cities in the southern hemisphere, as evidenced by no datapoints at a latitude < -60, the more numerous norther cities may have enabled a linear regression line to still be statistically significant.

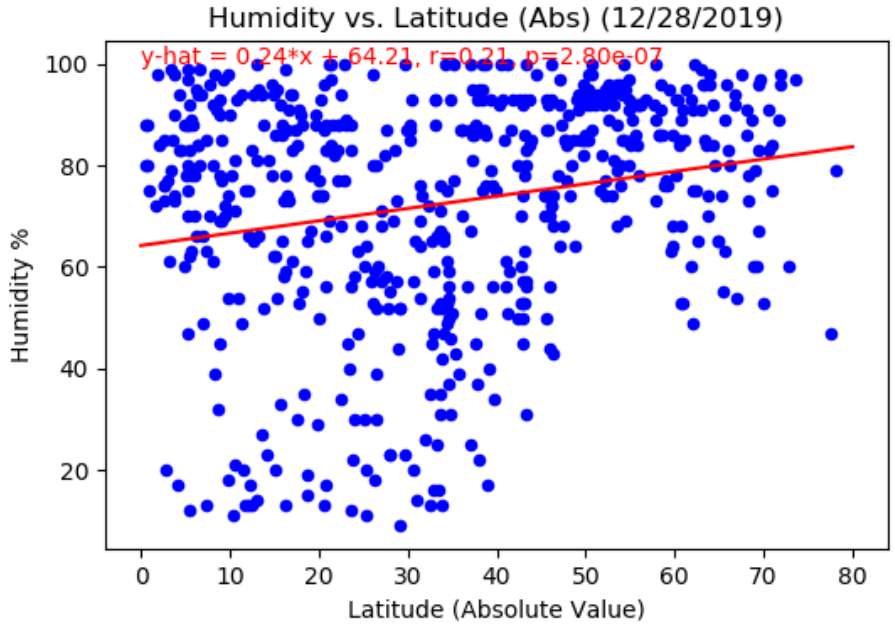

Temperature vs. Latitude (12/28/2019)
y-hat = -0.42*x + 18.37, r=-0.77, p=9.80e-113
Temp
Lat
40
20
0
−20
−40
−60
−40
−20
0
20
40
60
80

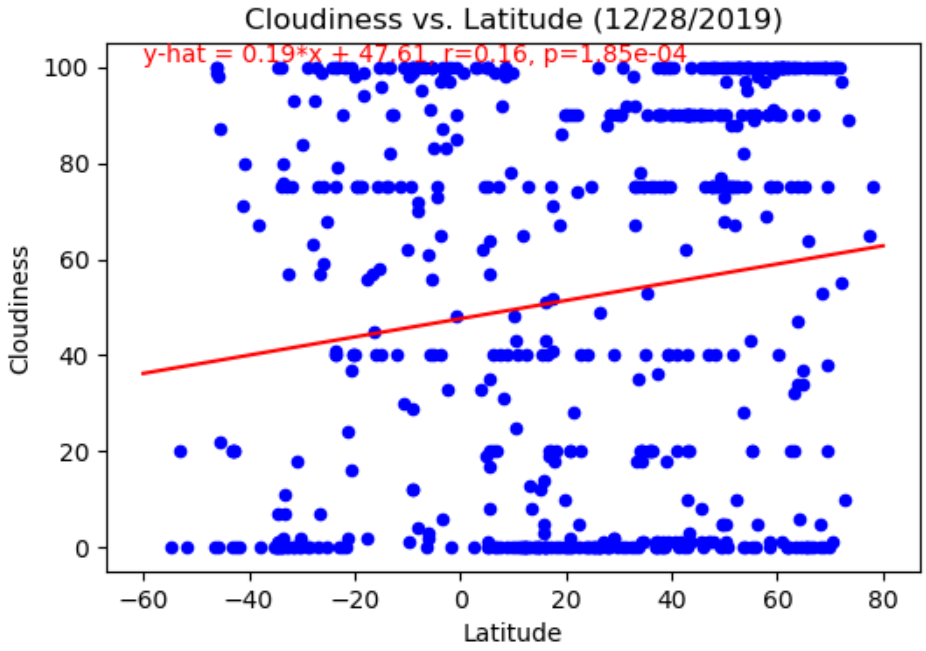

Temperature vs. Latitude (Absolute Value) (12/28/2019)
y-hat = -0.77*x + 35.58, r=-0.84, p=3.72e-157
Temp
Lat AbsVal
40
30
20
10
0
−10
−20
−30
−40
0
10
20
30
40
50
60
70
80

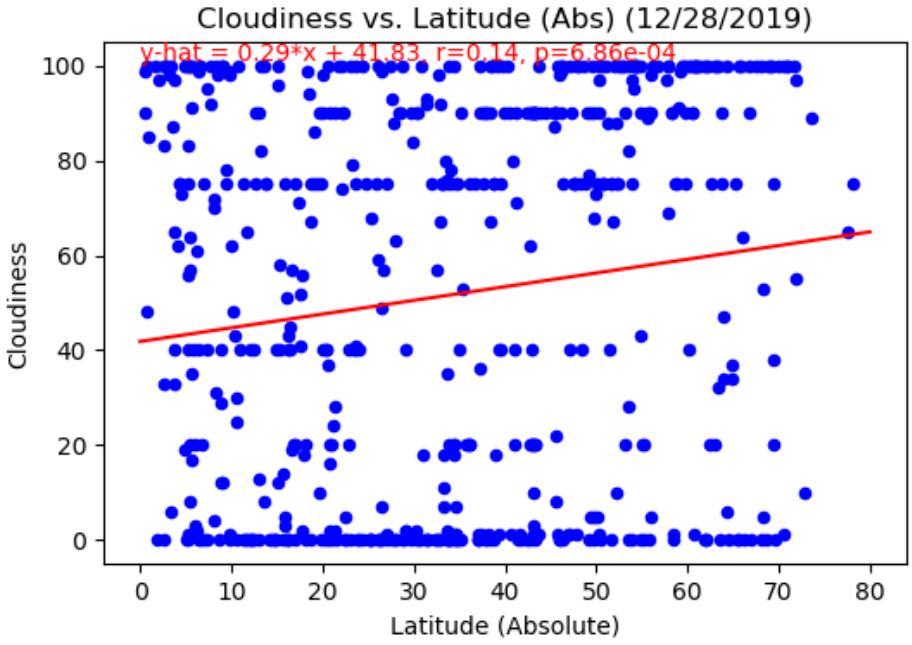
2. **Humidity and Cloudiness correlate positively with latitude, with statistical significance for 12/28/19.**

   This finding wasn't quite as obvious and although this was demonstrated for the sample date presented, 12/28/19, it did not hold true for all dates explored. To have greater confidence in this finding would require datasets across several dates spread across the year. Also, notice that although this was statistically significant, the impact was slight for both weather indicators (absolute value latitudes and non-absolute value latitudes).



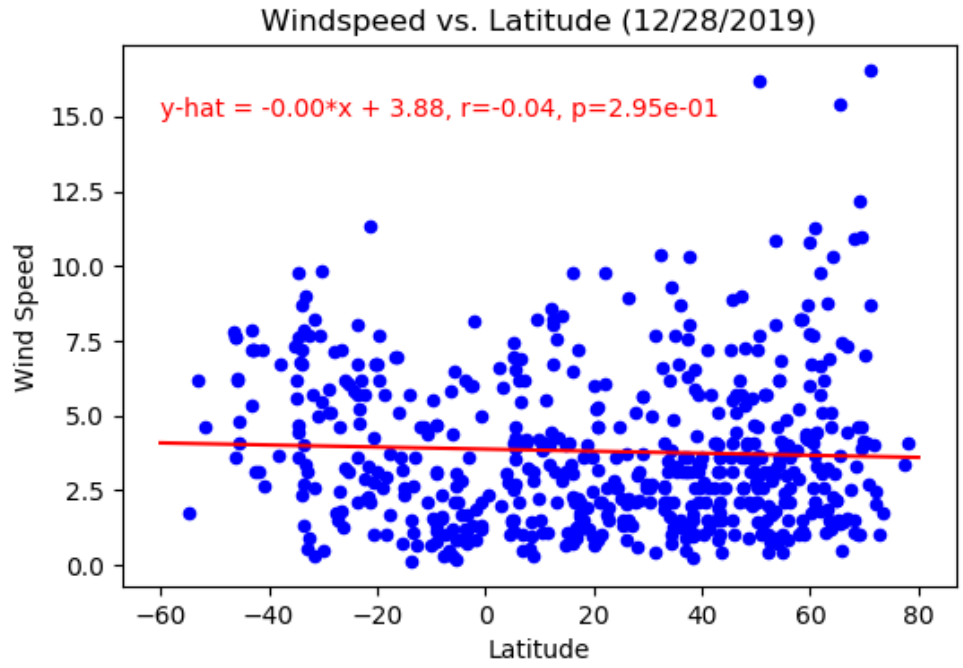


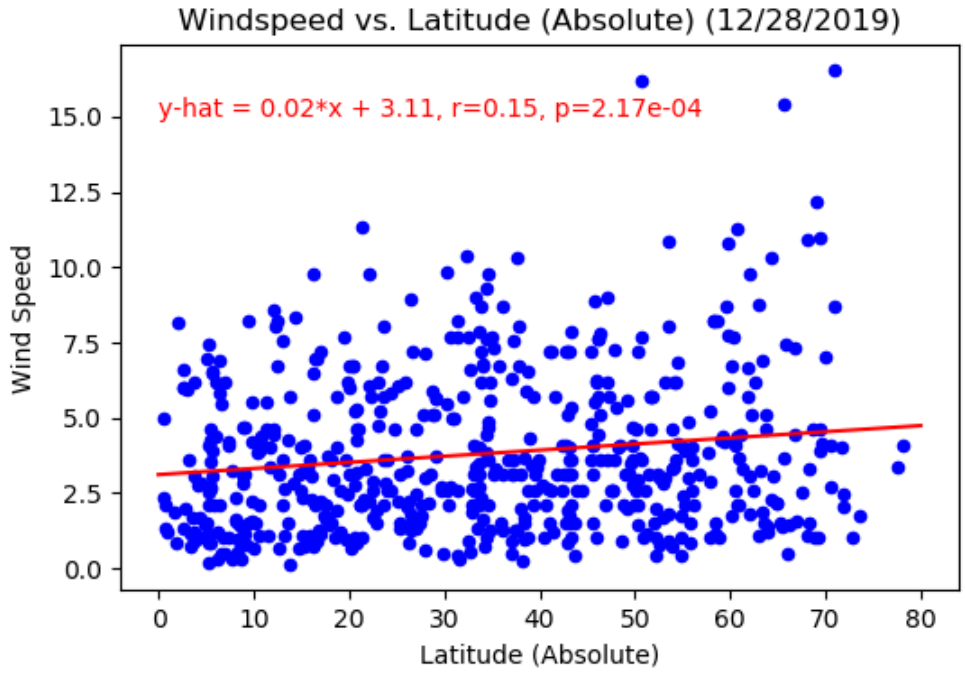

Cloudiness vs. Latitude (12/28/2019)
y-hat = 0.19*x + 47.61, r=0.16, p=1.85e-04
Cloudiness
Latitude
100
80
60
40
20
0
−60
−40
−20
0
20
40
60
80


Cloudiness vs. Latitude (Abs) (12/28/2019)
y-hat = 0.29*x + 41.83, r=0.14, p=6.86e-04
Cloudiness
Latitude (Absolute)
100
80
60
40
20
0
0
10
20
30
40
50
60
70
80

3. **Wind speed had statistically significant correlation with latitude (absolute value) on 12/28/19.**

   No statistical significance was seen with raw latitude, but when applied to latitude absolute values, p=2.2E-04. It does seem intuitive that there would be more extreme weather, hence greater wind, toward the poles.





Conclusion and Recommendations

1. **To gain a robust understanding requires a more detailed analysis: accounting for curvature of the earth with the application of transforms, looking at sample dates across the year, separate analysis for the northern and southern hemispheres.**

2. **Recommend creation of a regression model that evaluates for other factors including time of day, hemisphere (northern or southern), distance to the coast, etc.**

See the Jupyter Notebook for detailed analysis and dataset tables.