Significance of Weather Factors on Wind Speed

Intro:

I have been sailing on Chicago's breeze, for which the city has been dubbed the Windy City, for over 6 years and with that experience have noticed trends in the wind's strength. Using the 2020 hourly normals data set from NOAA or the National Oceanic and Atmospheric Administration, a government agency tasked with reporting and predicting the weather, I hope to confirm or reject my theories and look for new patterns I hadn't considered. My theories going into the project are:

- The wind is strongest in the winter/spring and in the afternoon
- Large temperature differences will cause wind
- More clouds mean more wind
- The wind is strongest from the east and comes most often from the west or north

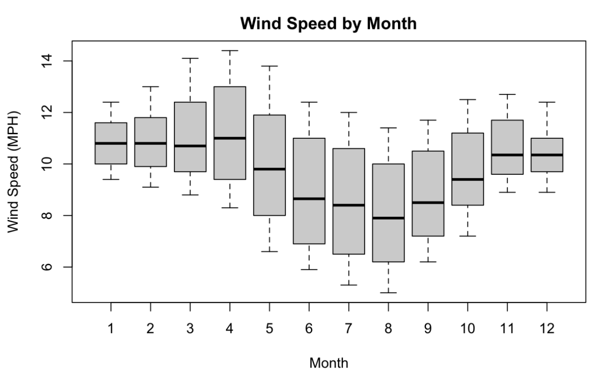
Methods:

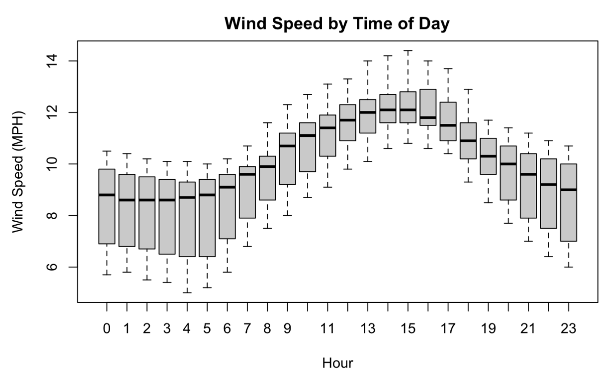
The data set is part of a 30-year record of normals and averages of various metrics taken every hour from Midway Airport (Station USW00014819) and contains 8760 observations of 113 variables. Of the 113, 27 are variables we are interested in with the rest being either flags, which indicate completeness of data, or yearly averages, which we don't need since our set is only one year long. The flags, yearly average, and other constants were removed in the cleaning process. The remaining variables can be divided into time, temperature, humidity, cloud cover, wind, and sea pressure. Unfortunately, the data set didn't contain any precipitation data and sea pressure data was too incomplete for any practical application. In addition to the 27 variables, the cumulative cloud percentage, a conversion of separate percentages of cloud cover into one Okta measurement, and the hourly temperature range or the difference between the 90th% and 10th% temperatures were calculated. Finally, the wind direction was converted from a number to its respective letter representation (ie. 1 becomes N) and new sets were created that pertained to each month and season.

Then using the data set I created a variety of graphs to compare wind speeds based on various key metrics from each category to look for and describe preliminary patterns in the data. I then cross-check these patterns by breaking off the top and bottom 10% of data set by wind speed. If our preliminary conclusions are correct we should see these groups skew on key metrics (ie. the top 10% would contain cloudier measurements). For some metrics, we can also use a P-test and a null hypothesis to compare to our general weather data.

Results:

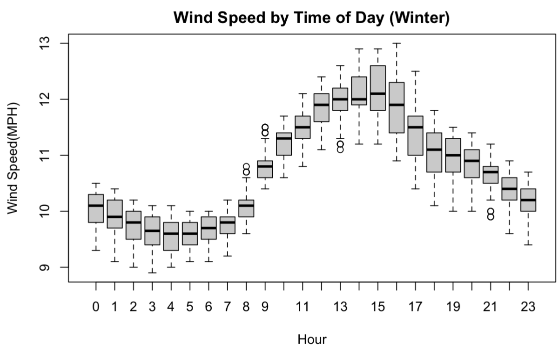
Time:

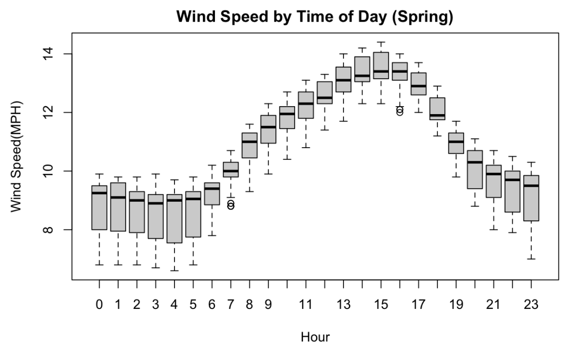


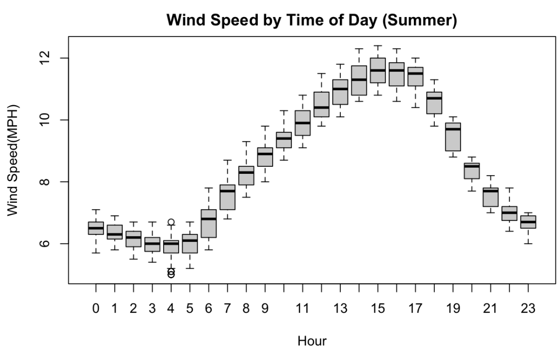


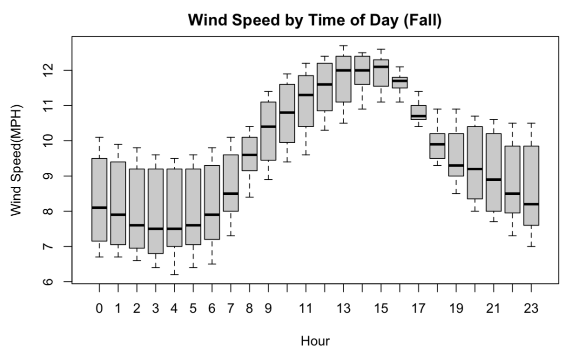
In the monthly data we can see higher median wind speeds from November to April when compared with summer months with the highest speeds being in April and the lowest speeds in August. Additionally, the Spring and Summer months see a wider range of wind speeds relative to the winter.

For the time of day, we can see an increase in the wind speed in the afternoon peaking at 3 pm with a notable tightening of the middle 50% of data sets in the afternoon as well. This trend persists regardless of the season and is strongest in the Summer while weakest in the fall

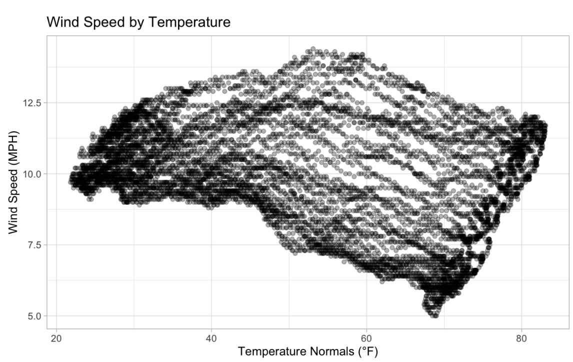






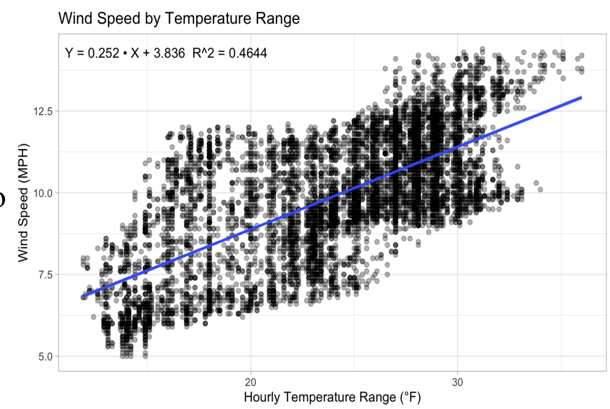


Temperature:

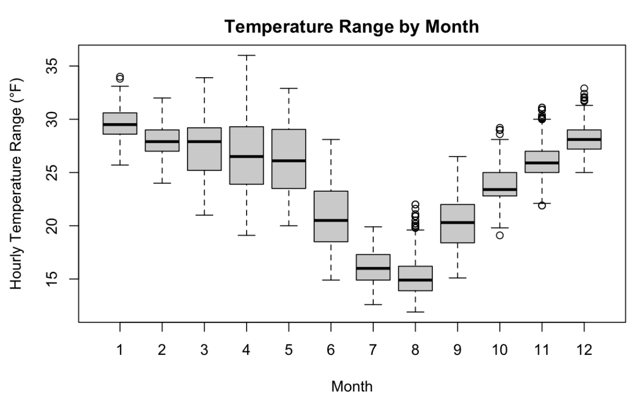


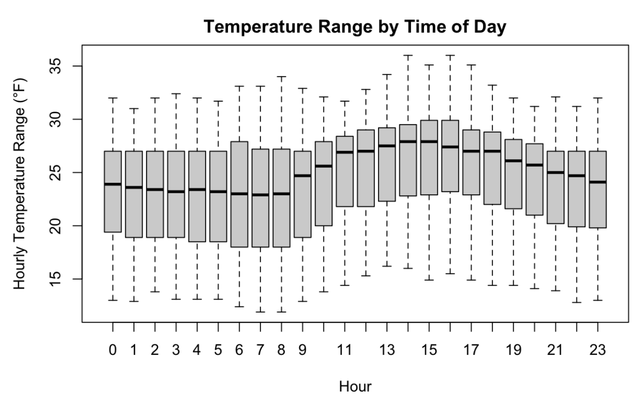
The normal temperature appears to bear little direct relationship with wind speed however we do see some interesting trends such as the converging of wind speeds at colder temperatures towards around 10 MPH and a distinct sloping cut-off on the right-hand side for which I have no explanation.

Perhaps more interesting and pertinent is the hourly temperature range. The higher this range value is the more the air is changing temperature because we see greater variation in temperatures in an hour. We can see in the graph that as this range increases so too does wind speed. The R^2 of 0.4644, though somewhat low, when considering the amount of variability in weather patterns, highlights a clear relationship. This trend also makes sense because wind is formed from pressure differences caused by different temperatures in the air.

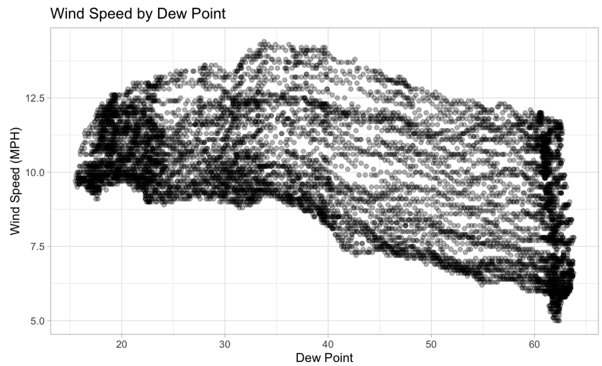


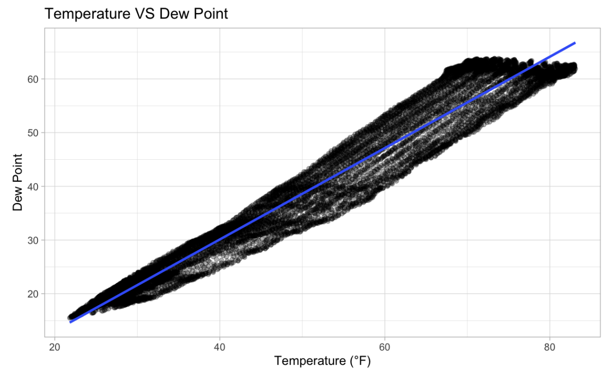
This relationship can also help us explain some of the trends we saw earlier with the time of the year and time of day. The hourly temperature range is a lot lower in the summer and higher in the winter matching what we see of wind speed in these same months. Less pronounced but similar is the bump in the median hourly temperature range around the afternoon. Therefore, I can draw the correlation that to some extent the reason the winter/spring and afternoon see greater wind speed is because the temperature is changing more rapidly. I in fact had a coach point out this trend to me one summer with the afternoon sun heating the cooler morning air and leading to wind.





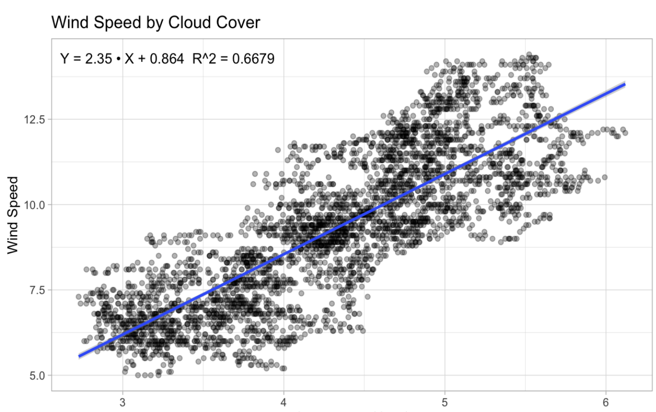
Humidity:





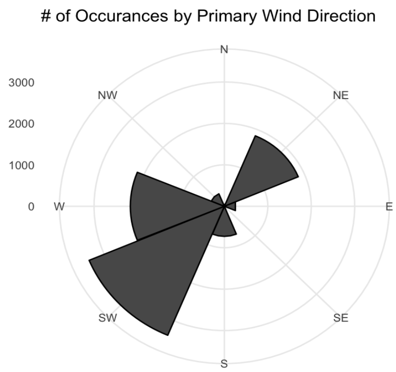
Humidity is measured in Dew point and is the temperature the air needs to be cooled by to start condensation. Therefore, dew point is highly correlated to temperature and this is reflected in our graphs. Similar to temperature then it has seemingly little direct correlation to wind speed and even the graph looks similar to the temperature graph except for the lack of the right-hand slopping cut-off trend.

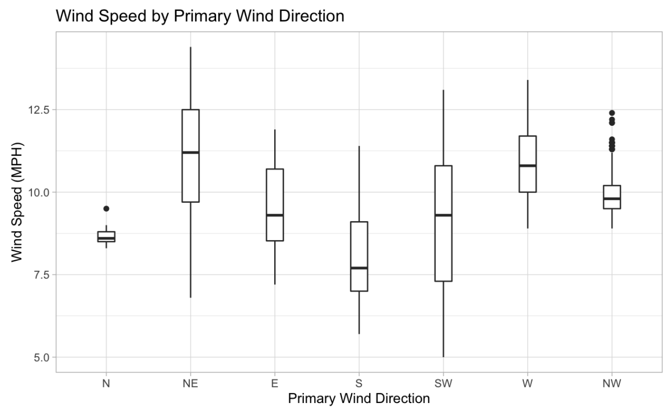
Cloud Cover:



Cloud cover is measured in Oktas and uses a 0-8 scale to describe cloud cover with 0 meaning no cloud cover and 8 being full coverage. When we graph cloud coverage to wind speed we see that the greater the cloud cover the greater the wind speed with a strong R^2 of 0.6679. This trend is likely a demonstration of the correlation between low-pressure areas, wind, and clouds/storms. In fact, when sailing and looking for wind dark spots caused by cloud cover can often be a good sign that there may be higher wind in that area.

Direction:

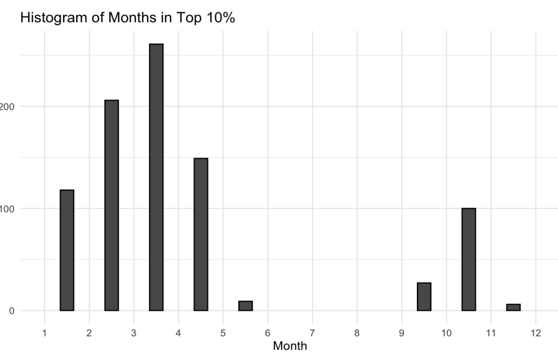


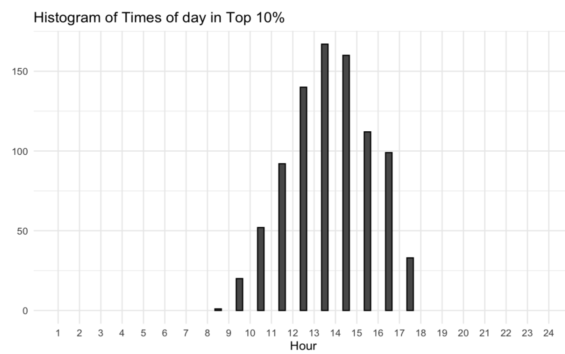


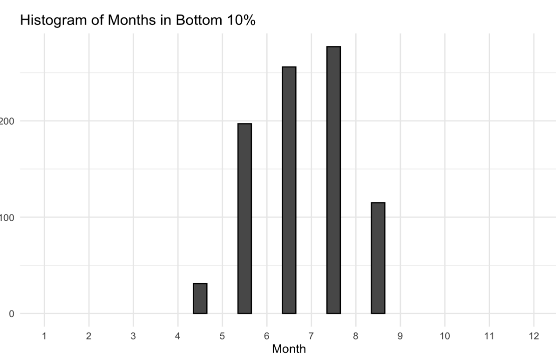
To measure the wind direction I used the primary wind direction which simply describes the wind each hour as mostly coming from one of the eight cardinal directions. The wind came most frequently from the southwest, west, and northeast and in fact came from the southeast zero times throughout the whole year. This is quite different from my initial assumptions however this may have to do with the difference in location between midway and the lake where I sail. All the directions demonstrated large ranges of possible wind speeds with the northeast being on average the strongest, however, the range here also varied substantially and so it is hard to make a concrete conclusion.

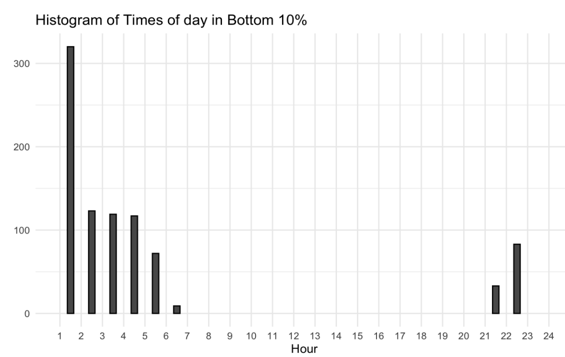
Cross-checking:

Of the conclusions found above, we will check the three with the strongest found effect, namely time, cloud cover, and the hourly temperature range. For time we should expect the top 10% of wind speeds to come mostly from the spring and winter months as well as from the afternoon while the slowest 10% of speeds should come from mostly the summer and late at night or very early in the morning.









Looking at the histograms we do in fact see these trends, with the lack of almost any overlap between the bottom and top 10% being striking. In fact when considering that in our full data set all of these should occur roughly the same number of times the trend seems to be somewhat confirmed

For cloud cover, the top 10% should have a higher mean cloud cover than the full data set and the bottom 10% should have a mean cloud cover lower than the full data set. Hourly temperature range is a similar story where one would expect the top 10% to have an above-average range and the bottom 10% to be below average. For both, we can test this hypothesis by creating their respective null hypotheses where Subset μ = Full Set μ. By then constructing our alternative hypothesis to match what we expect to occur, if we can reject our null hypothesis we can make a case that our preliminary conclusion was correct. To calculate our P-value we will be using a t-test in both cases as we are attempting to use a limited sample to make a conclusion about heavier or lighter wind more generally.

Cloud cover:

Null hypothesis: $\mu_0 = 4.318501$

Top 10%

Alternative hypothesis: $\mu > 4.318501$

Using the T-test we get a P-Value of

$2.158145 \times 10^{-203}$

Bottom 10%

Alternative hypothesis: $\mu > 4.318501$

Using the T-test we get a P-Value of

$2.158145 \times 10^{-203}$

Hourly Temperature Range:

Null hypothesis: $\mu_0 = 24.06062$

Top 10%

Alternative hypothesis: $\mu > 24.06062$

Using the T-test we get a P-Value of 0

Bottom 10%

Alternative hypothesis: $\mu > 24.06062$

Using the T-test we get a P-Value of 0

In all four instances, we have a P-value less than the standard threshold of 0.05 and we may therefore reject our null hypothesis in favor of our alternative.

Conclusion:

Overall my experience was relatively successful at recognizing the most important weather factors however wind direction didn't quite match my expectations. Additionally, the lack of pressure and precipitation data was disappointing. NOAA almost certainly has other data sets that contain these factors so it could be interesting to look at their correlation with wind speed. The data could also have been extended to a longer time period to see if these trends hold up over the years or if wind speeds may have even changed over time. Any of these questions could make for interesting next steps. The project has also generally helped me see the power of data visualization in summarizing a data set, as without the graphs I made I would have been unable to even begin to understand the sheer amount of data in my data set.