**Denise Dodd**

**DSC530**

**Data Exploration and Analysis Final**

***Statistical/Hypothetical Question***

The hypothesis posed is: Elements of weather are all closely related and impact one another. The variables of temperature, humidity, windspeed, visibility, and air pressure were reviewed.

***Outcome of your EDA***

Through several statistical tests, I determined that there are some weak relationships between the variables, but overall they seem to act independently of each other.

* The PMF showed that there is a weak positive relationship between humidity and temperature.
* The CDF showed that there is a positive relationship between air pressure and temperature.
* The normal probability plot showed that windspeed is independent of temperature.
* The scatterplots and corresponding covariance and correlation stats showed that air pressure and humidity have a very weak negative relationship. An additional scatterplot and corresponding calculations show there is no relationship between windspeed and visibility.
* The CorrelationPermute Hypothesis Test determined that any correlation between wind speed and visibility is not statistically significant. We can accept the null hypothesis that there is no connection between these variables.
* The Regression Analysis showed that humidity and windspeed do have a relationship with air pressure as these variables contribute to 19.8% of the variability in air pressure.

Overall, the hypothesis that all variables are impacted by other variables was false. The strongest relationships were the positive relationships between temperature/humidity and temperature/air pressure. Temperature/windspeed and visibility/windspeed act independently of one another and have no apparent relationship.

***What do you feel was missed during the analysis?***

I did not consider the time of year in the analysis. It would be interesting to see if the relationships or the strength of the relationships between the variables changed with the seasons.

***Were there any variables you felt could have helped in the analysis?***

The dataset provided a categorical variable of “Weather” which displayed if it was rainy, cloudy, sunny, etc. It would be helpful if there were quantitative variables describing what percentage of could cover there was, how many centimeters of rain or snow there was, etc. These elements could have a strong effect on the variables that were studied.

***Were there any*** ***assumptions made you felt were incorrect?***

There were some assumptions I was making in what the outcome would be that proved incorrect. I assumed windspeed would affect visibility as leaves and dust would blow and obstruct visibility. Additionally, I thought that air pressure and humidity would have a much stronger correlation since humidity tends to make the air feel “heavy.” These were assumptions I had that were contradicted by the data.

***What challenges did you face, what did you not fully understand?***

The largest challenge was finding a way to subset the data and incorporate each of the identified variables in the dataset. Temperature seemed to be the most prevalent weather measurement, so I decided to subset by grouping data related to temperatures above and below the mean temperature. I also had a challenging time interpreting the Regression Analysis as both explanatory variables had p-values of 0.