Business Analytics mba6693

Assignment 1

3681282

GS Kumbhare

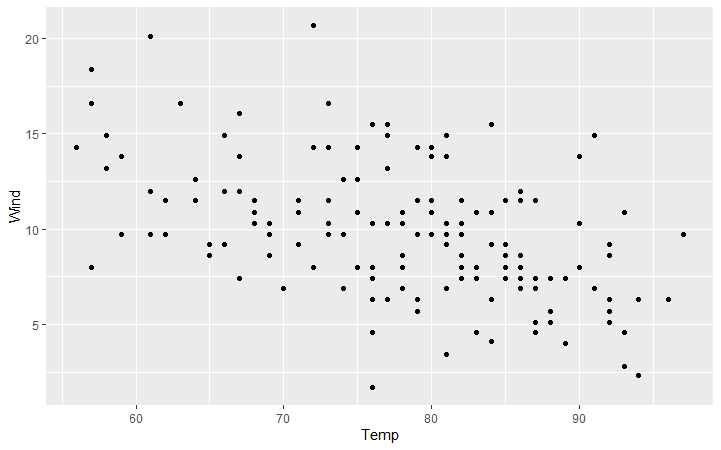
g.kumbhare@unb.ca

Air Quality

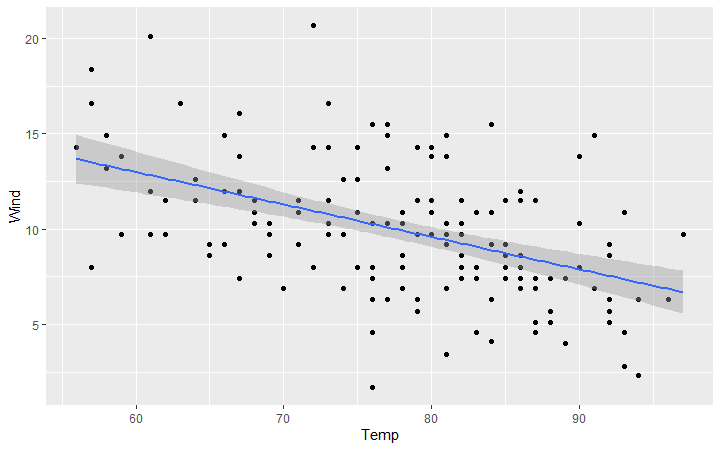
We use dataset from R database for our regression model. The dataset is Air quality data of New York city from the month of May till the end of September of random year. The Variables are Ozone concentration to trap heat and reduce UV rays from sun, solar radiation, windspeed and temperature. All these helps us to understand the weather and air quality of the location.

First, we understand simple linear regression

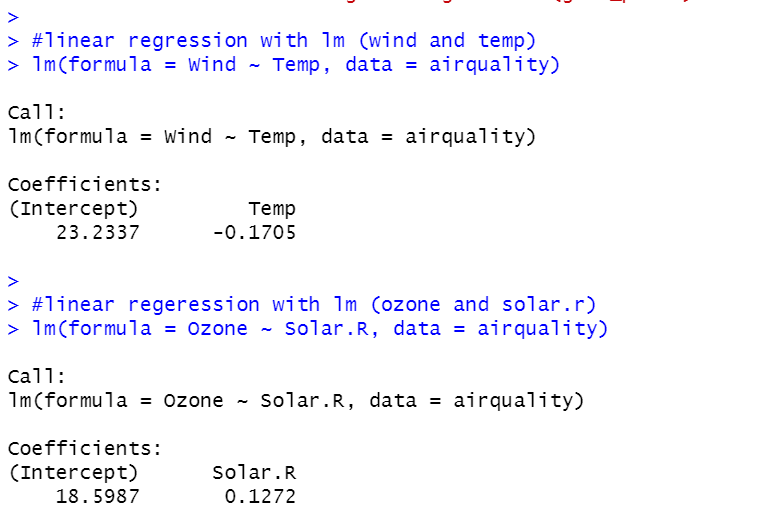
1. Wind Vs Temp

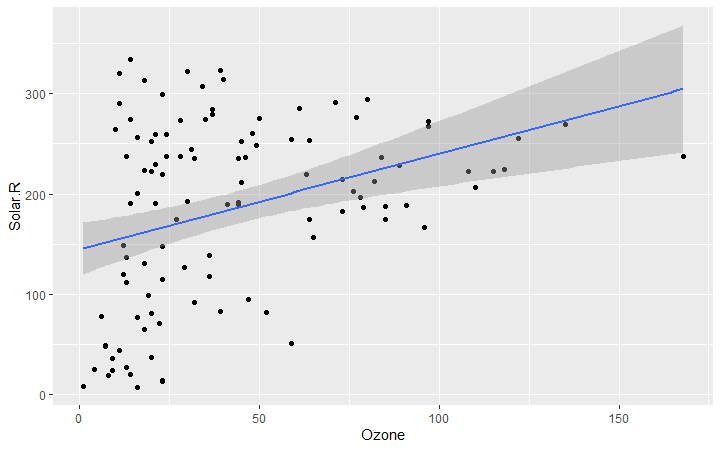


The above is Temp vs wind Scatter plot.



The above figure is Temp vs Wind trendline. The trendline appears linear. It can be understood from this situation that, as the temperature increases, the wind speed decreases. The slope is linear. Further while analysing, the intercept for wind is 23.233 and temperature coefficient is -0.1705. Keeping this in mind, it can be found that at 0 temp the windspeed is at 23.23. Another finding is that every unit change in windspeed, temperature will change by -0.1705 units.

1. Ozone vs Solar radiation:

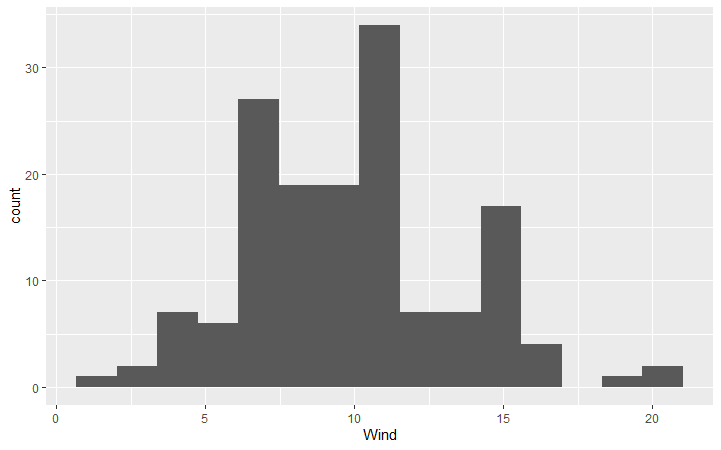


In ozone vs Solar radiation analysis, it is observed that as solar radiation increases, ozone layer increases simultaneously. The increase is linear as shown on the trendline. Further by comparing the data of this group, its intercepts and solar radiation coefficient is found. It is understood from the data that at 0 radiation from sun the ozone is at 18.598 units. Every unit change in ozone the radiation will change by 0.127 units.

Histogram

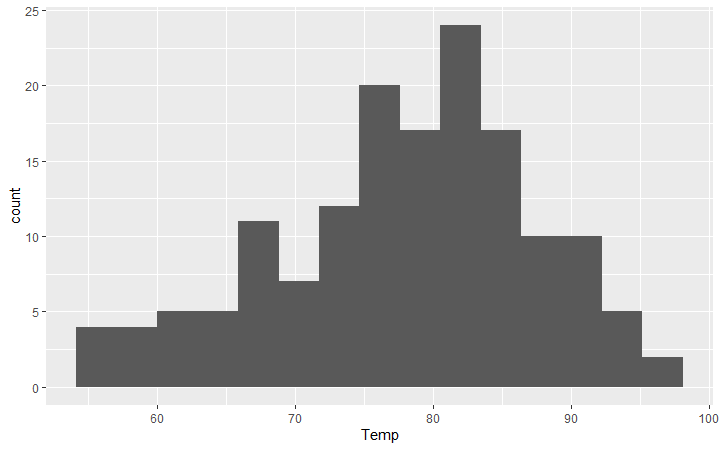
With histogram we can understand the concentration of the data.

1. Wind speed



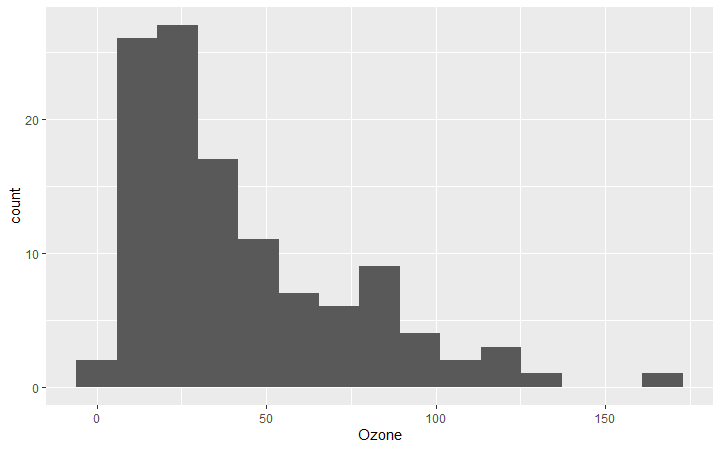
Wind speed is highest concentrated between 10- 11.25 units. With almost 40 counts in 150 days.

1. Temperature Histogram



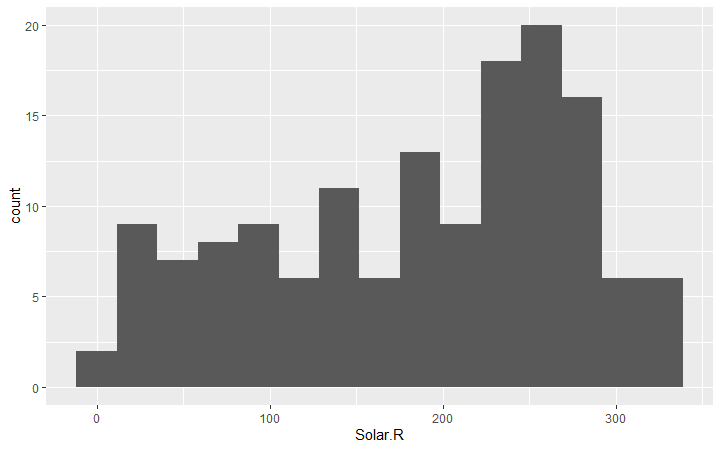
The temperature is concentrated between 81- 82 units for the given five months of data collection.

1. Ozone histogram



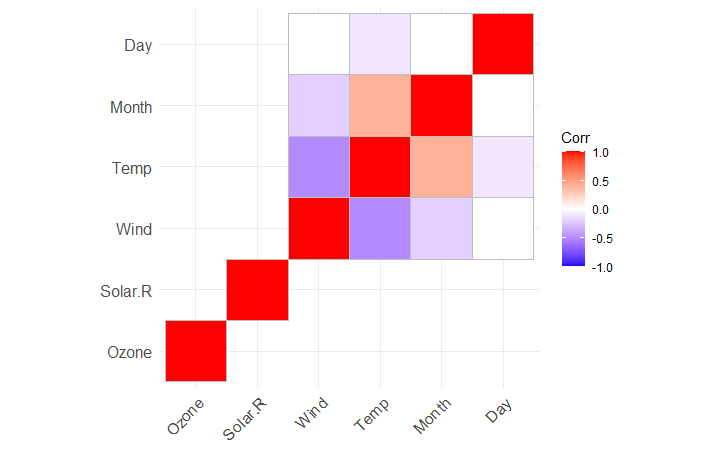
Ozone is concentrated between 20-30 units.

1. Solar histogram



The solar radiation is highly concentrated between 250- 260 units.

Correlation



Conclusion

With all the study done here we understood how regression modelling is done in R. We understood linearity and how correlation works. We further studied dependability of the variable on each other.