Michael Fanelli

Dr. Hill

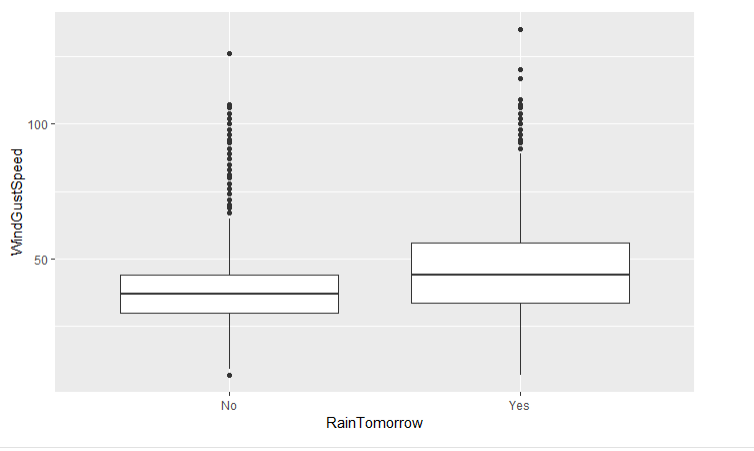
BAN 502

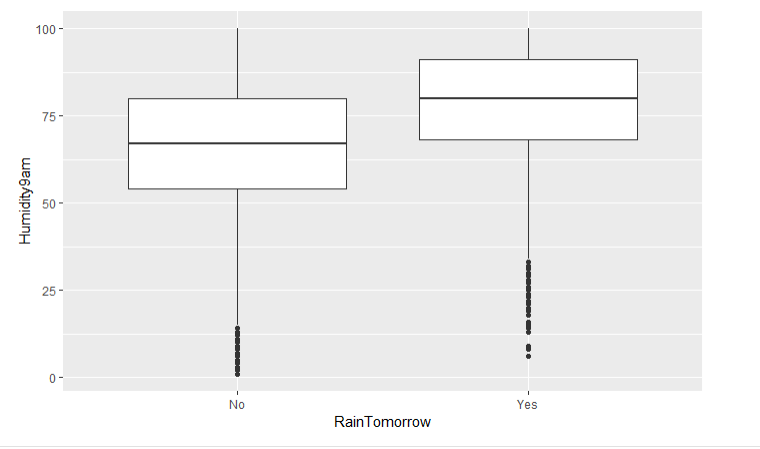
02/25/19

Course Project Part 1

I was given a data set with information about rain patterns given certain information. This information included date, location, MinTemp, MaxTemp, Rainfall, WindGustDir, WindGustSpeed, WindDir, WindSpeed, Humidity, Pressure, Clouds, Temp, and weather it rained that day or the following. Using these variables and the information provided with them I created tables and charts to find a common predictor for our response variable which was whether it would “RainTomorrow”. Upon getting the data I mutated RainToday, RainTomorrow, WindGustDir, WindDir9am, and WindDir3pm form characters into factors for R coding calculations to work. Furthermore, I used vim\_plot to view the “missingness” among the variables. All variables had missing information with cloud coverage at 9am and 3pm missing about 40% of its information. Based on looking at the missingness, I decided to imputate the missing information based on data in each variable.

After configuring the data to work with R coding I created boxplots, Histograms, and tables. When looking at the tables and charts my thought process was to find aspects of variables that would have a high probability of it raining the following day. While examining the charts and tables I came to the conclusion that rainfall, wind gust speed, humidity and pressures have certain aspects to predict if it rains one day it will have a high percent chance of rain the following. Example being rainfall. With Rain fall, if it is over 35.6mm of rain that fell that day then the next day has well over 50% chance of raining. Similar goes with wind gust speed, humidity, and pressures. When wind gusts speed reach 74 km/h it has over a 50% chance of rain the next day. When it comes to humidity, if humidity reaches over 79% at 3pm it has a 51%-90% chance of raining the next day. Lastly, pressures at both 9am and 3pm. If pressure is a low, about 1006.2 and down, it has a 50%-77% chance of rain.





Based on how I approached this data, I have come to the conclusion that the best variables for predicting rain for the following day would be the variables listed previously. These variables have shown that once a certain point has been reached that the likelihood of it raining the following day is over 50%. All other variables were to inconsistent on the likelihood of it raining based on certain parameters or limits.