# Alex Fields

## BAN 502

### Module 2 - Assignment 1

#### Task 1 - Q&A

Describe this dataset. How many variables and observations are there? **There are 153 observations (rows) and 6 Variables (columns)** Is there any missing data? **Yes, there are some NA values in the dataset** Which variable is likely to be the response (Y) variable? **Ozone would be the best predictor variable**

#### Task 2 - Q&A

How many rows and columns remain in this new (air2) data frame? **There are 111 Observations and 6 Values when droping NA**

#### Task 3 - Q&A

Which variable is most strongly correlated with the “Ozone” variable? **Temp is highly correlated with Ozone (0.7)** Which variable is least strongly correlated with the “Ozone” variable? **Wind is least correlated with Ozone (-0.6)**

#### Task 4 - Q&A

Describe the relationship between “Temp” and “Ozone”. **Using Scatter plot function, we can see that Temp and Ozone are positively correlated. When Temp increases, Ozone also increases.**

#### Task 5 - Q&A

*a*. Discuss the quality of this model (mention the R square value and significance of the predictor variable). **The model is relatively good but not the best. For every 1 unit of Temp increase you see -147 decrease of Ozone. The R^2 value for this is equal to 0.488. This is showing partial correlation.** *b*. Use the code “confint(model1)” to generate 95% confidence intervals for the coefficients. In what range does the slope coefficient likely fall? **The range for Ozone and Temp, resepectively fall between [-184.818372, -110.473773] [1.964787, 2.913433].**

#### Task 8 - Q&A

*Residuals* - We can see with our Residual plot that the SRR is not too noisy (give or take a few data points) giving hope to a good model. *R^2* - We can see that the R^2 value is not that great, showing a value of 0.48. This shows mild correaltion in the model. *Hypothesis Test* - We can see that our linear model shows a p-value of less than .05 meaning Temp was a significant predictor of Ozone. *Common Sense* - Overall using common sense, I would assume that the higher the temperature the more likely Ozone is to either go up or down (depending on how Ozone is measured). We can see with the scatterplot that these are relatively correlated, so to me this makes a decent model, statistics aside.

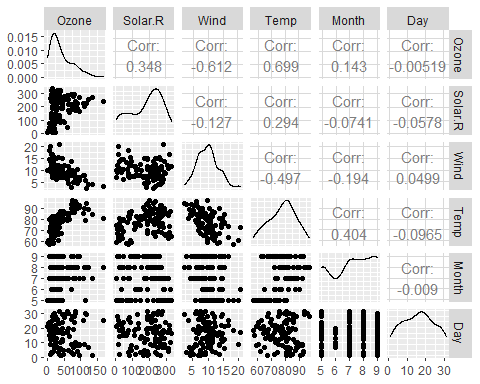
#### Task 9 - Q&A

How might the model that you constructed in Task 5 be used? **We would use the confidence interval to determine where a certain parameter falls into a certain proportion. Meaning we can “predict” that Ozone or Temp will fall into a certain range.** Are there any cautions or concerns that you would have when recommending the model for use? **The model is not perfect and you would want to fine tune the model before publishing or using for actual research. If you can show a higher R^2 value I would be ok with using this model. If you were able to collect a larger pool of data, I will feel confident that this data would align to a “good” model.**

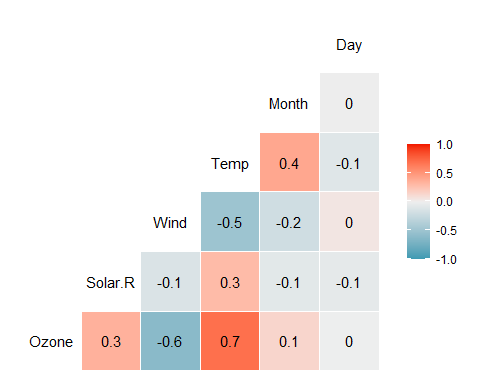
## Ozone Solar.R Wind Temp   
## Min. : 1.00 Min. : 7.0 Min. : 1.700 Min. :56.00   
## 1st Qu.: 18.00 1st Qu.:115.8 1st Qu.: 7.400 1st Qu.:72.00   
## Median : 31.50 Median :205.0 Median : 9.700 Median :79.00   
## Mean : 42.13 Mean :185.9 Mean : 9.958 Mean :77.88   
## 3rd Qu.: 63.25 3rd Qu.:258.8 3rd Qu.:11.500 3rd Qu.:85.00   
## Max. :168.00 Max. :334.0 Max. :20.700 Max. :97.00   
## NA's :37 NA's :7   
## Month Day   
## Min. :5.000 Min. : 1.0   
## 1st Qu.:6.000 1st Qu.: 8.0   
## Median :7.000 Median :16.0   
## Mean :6.993 Mean :15.8   
## 3rd Qu.:8.000 3rd Qu.:23.0   
## Max. :9.000 Max. :31.0   
##

air2 = air %>% drop\_na()

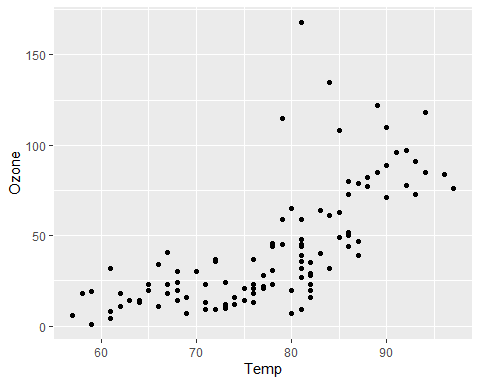
ggpairs(air2)



ggcorr(air2, label = TRUE)



ggplot(air2, aes(x=Temp, y=Ozone)) + geom\_point()



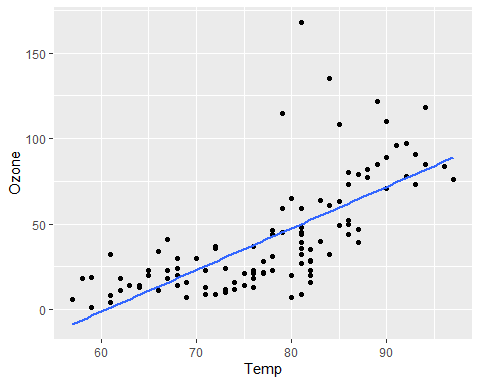
X = air2$Temp  
Y = air2$Ozone  
  
model1 = lm(Y ~ X, air2)  
summary(model1)

##   
## Call:  
## lm(formula = Y ~ X, data = air2)  
##   
## Residuals:  
## Min 1Q Median 3Q Max   
## -40.922 -17.459 -0.874 10.444 118.078   
##   
## Coefficients:  
## Estimate Std. Error t value Pr(>|t|)   
## (Intercept) -147.6461 18.7553 -7.872 2.76e-12 \*\*\*  
## X 2.4391 0.2393 10.192 < 2e-16 \*\*\*  
## ---  
## Signif. codes: 0 '\*\*\*' 0.001 '\*\*' 0.01 '\*' 0.05 '.' 0.1 ' ' 1  
##   
## Residual standard error: 23.92 on 109 degrees of freedom  
## Multiple R-squared: 0.488, Adjusted R-squared: 0.4833   
## F-statistic: 103.9 on 1 and 109 DF, p-value: < 2.2e-16

confint(model1, level = 0.95)

## 2.5 % 97.5 %  
## (Intercept) -184.818372 -110.473773  
## X 1.964787 2.913433

ggplot(air2, aes(x=Temp, y=Ozone)) + geom\_point() + geom\_smooth(method="lm", se = FALSE)



p1 = predict(model1,data.frame("X"=80))  
p1

## 1   
## 47.48272

plot(model1)

