Marci B Copeland BAN502 Module 2 Assignment 1

Task 1: Air contains 6 variables and 153 observations. The dataset contains missing data. Ozone is likely to be the Y variable.

Task 2: Air2 contains 6 variables and 116 observations.

Task 3: Which variable is most strongly correlated with the “Ozone” variable? Day Which variable is least strongly correlated with the “Ozone” variable? Temp

Task 4: Relationship between Ozone and Temp based on the chart indicates a weak correlation between the two variables.

Task 5: The R squared value is weak and not good being 0.48 since it is not close to 1. However the p value is good for this model.This model provides a good baseline.

Task 8: This does not meet the four linear regression model assumptions.

Task 9: I think this model coud be used in looking at the correlation between Ozone and Temp but not sure there is enough information that informative decisions can be made. I think more models are needed to provide a better correlation between all variables.

install.packages(“GGally”) install.packages(“car”) install.packages(“lmtest”)

library(tidyverse)

## -- Attaching packages ---------------------------------------------------------- tidyverse 1.2.1 --

## v ggplot2 3.2.1 v purrr 0.3.3  
## v tibble 2.1.3 v dplyr 0.8.3  
## v tidyr 1.0.0 v stringr 1.4.0  
## v readr 1.3.1 v forcats 0.4.0

## -- Conflicts ------------------------------------------------------------- tidyverse\_conflicts() --  
## x dplyr::filter() masks stats::filter()  
## x dplyr::lag() masks stats::lag()

library(GGally)

## Warning: package 'GGally' was built under R version 3.6.2

## Registered S3 method overwritten by 'GGally':  
## method from   
## +.gg ggplot2

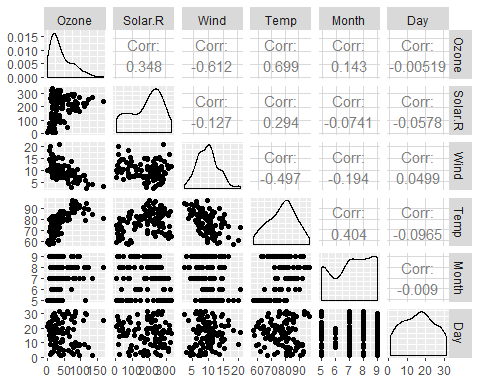
##   
## Attaching package: 'GGally'

## The following object is masked from 'package:dplyr':  
##   
## nasa

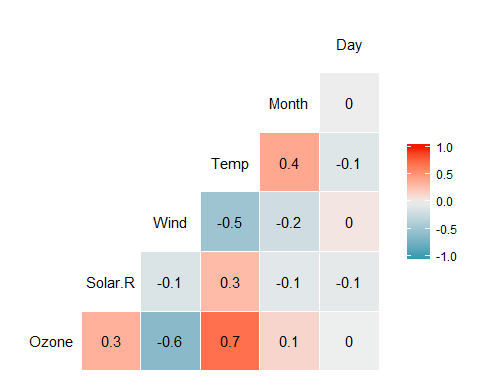
air = airquality

air2<-drop\_na(air,Ozone,Solar.R)

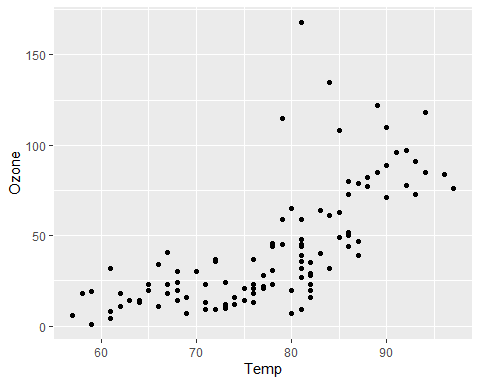
ggpairs(air2)



ggcorr(air2, label = TRUE)



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



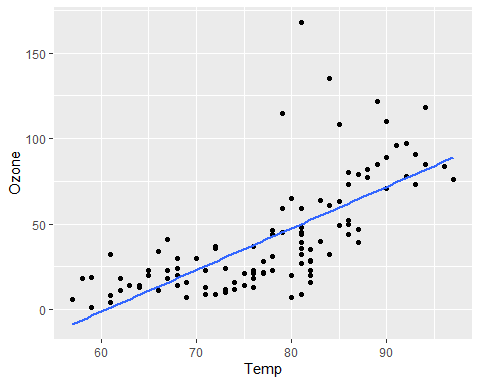
model1<-lm(Ozone ~ Temp,air2)  
summary(model1)

##   
## Call:  
## lm(formula = Ozone ~ Temp, 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 \*\*\*  
## Temp 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)

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

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



model2<-lm(Ozone ~ Temp==80,air2)  
summary(model2)

##   
## Call:  
## lm(formula = Ozone ~ Temp == 80, data = air2)  
##   
## Residuals:  
## Min 1Q Median 3Q Max   
## -41.42 -24.42 -10.67 19.58 125.58   
##   
## Coefficients:  
## Estimate Std. Error t value Pr(>|t|)   
## (Intercept) 42.417 3.211 13.209 <2e-16 \*\*\*  
## Temp == 80TRUE -11.750 19.534 -0.602 0.549   
## ---  
## Signif. codes: 0 '\*\*\*' 0.001 '\*\*' 0.01 '\*' 0.05 '.' 0.1 ' ' 1  
##   
## Residual standard error: 33.37 on 109 degrees of freedom  
## Multiple R-squared: 0.003309, Adjusted R-squared: -0.005835   
## F-statistic: 0.3618 on 1 and 109 DF, p-value: 0.5487

confint(model2)

## 2.5 % 97.5 %  
## (Intercept) 36.05195 48.78138  
## Temp == 80TRUE -50.46504 26.96504