Levi Moore 850435436 Module 2 Assignment 1

library(GGally)

## Loading required package: ggplot2

## Registered S3 methods overwritten by 'ggplot2':  
## method from   
## [.quosures rlang  
## c.quosures rlang  
## print.quosures rlang

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

library(tidyverse)

## ── Attaching packages ────────────────────────────── tidyverse 1.2.1 ──

## ✔ tibble 2.1.1 ✔ purrr 0.3.2  
## ✔ tidyr 0.8.3 ✔ dplyr 0.8.1  
## ✔ readr 1.3.1 ✔ stringr 1.4.0  
## ✔ tibble 2.1.1 ✔ forcats 0.4.0

## ── Conflicts ───────────────────────────────── tidyverse\_conflicts() ──  
## ✖ dplyr::filter() masks stats::filter()  
## ✖ dplyr::lag() masks stats::lag()

Task 1

air = airquality  
head(air)

## Ozone Solar.R Wind Temp Month Day  
## 1 41 190 7.4 67 5 1  
## 2 36 118 8.0 72 5 2  
## 3 12 149 12.6 74 5 3  
## 4 18 313 11.5 62 5 4  
## 5 NA NA 14.3 56 5 5  
## 6 28 NA 14.9 66 5 6

A. This dataset has figures concerning the ozone, solar, wind, temp, month and day.

B. This dataset has 153 observations and 6 variables.

C. There is mising data in the following variables: ‘Ozone’ and ‘Solar.R’.

D. Ozone is likely the (Y) variable. This variable is dependent on the other variables listed.

Task 2

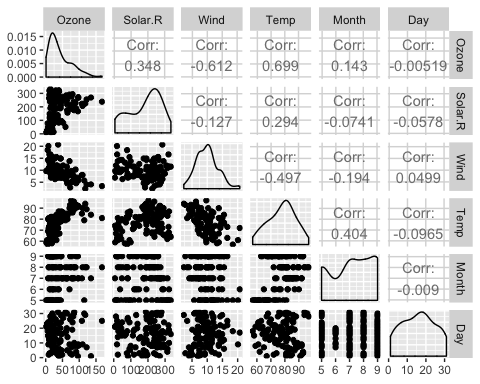
air2 = air %>% filter(!is.na(Ozone)) %>% filter(!is.na(Solar.R))  
head(air2)

## Ozone Solar.R Wind Temp Month Day  
## 1 41 190 7.4 67 5 1  
## 2 36 118 8.0 72 5 2  
## 3 12 149 12.6 74 5 3  
## 4 18 313 11.5 62 5 4  
## 5 23 299 8.6 65 5 7  
## 6 19 99 13.8 59 5 8

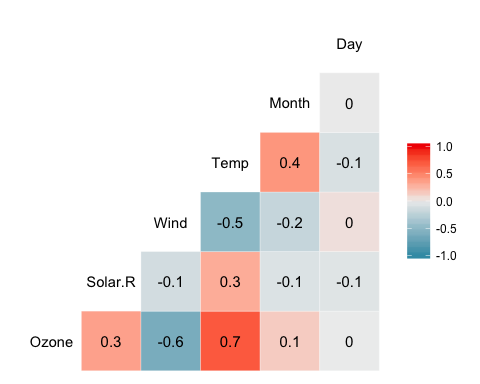
The new dataset, air2, contains 111 rows (observations) and 6 columns (variables)

Task 3

ggpairs(air2)



ggcorr(air2, label = TRUE)

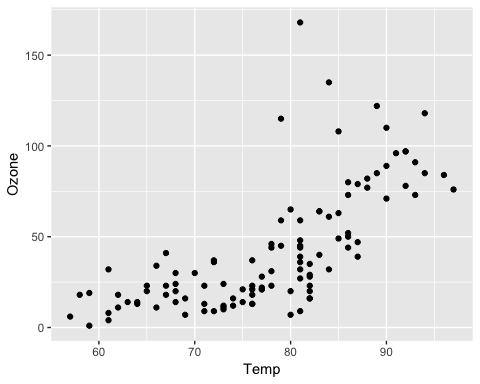


A. The variable that is most strongly correlated with the ‘Ozone’ variable is the ‘Temp’ variable.

B. The variable that is least strongly correlated with the ‘Ozone’ variable is the ‘Day’ variable.

Task 4

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

 There is a slight, positive relationshio between these two variables. As the ‘temp’ increases, the ‘ozone’ figure increases gradually.

Task 5

model1 = lm(formula = Ozone~Temp, data = 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)

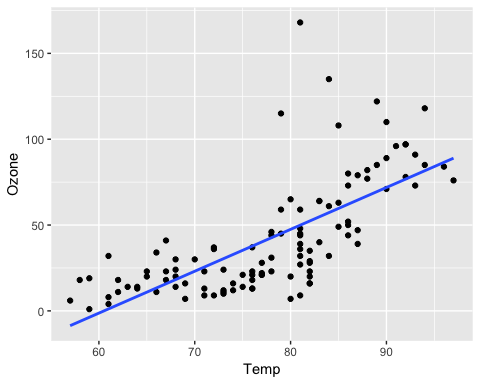
## 2.5 % 97.5 %  
## (Intercept) -184.818372 -110.473773  
## Temp 1.964787 2.913433

A. The R squared value is relatively, at 0.488. Since the coefficient is low (2.91) there isn’t a large correalation between these two.

B. The slope conefficient will likely fall between 1.96 and 2.91, looking at a 95% confidence interval.

Task 6

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

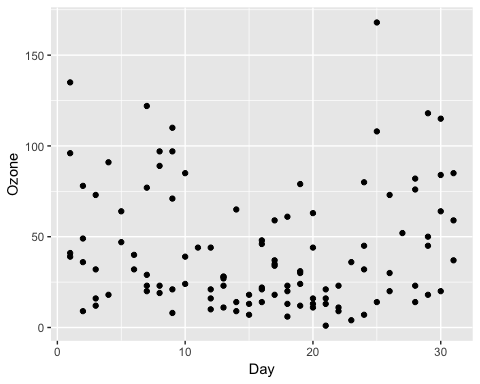


Task 7

#Manual  
#-147.65+2.44\*80=  
#Prediction 47.55

Task 8

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

 There is no relationship between the two variables “Day” and “Ozone”. This shows a 0.0 correlation.

Task 9

model2=lm(formula = Ozone~Day, data = air2)  
summary(model2)

##   
## Call:  
## lm(formula = Ozone ~ Day, data = air2)  
##   
## Residuals:  
## Min 1Q Median 3Q Max   
## -41.00 -24.23 -11.04 19.96 126.08   
##   
## Coefficients:  
## Estimate Std. Error t value Pr(>|t|)   
## (Intercept) 42.41536 6.64353 6.384 4.32e-09 \*\*\*  
## Day -0.01983 0.36604 -0.054 0.957   
## ---  
## Signif. codes: 0 '\*\*\*' 0.001 '\*\*' 0.01 '\*' 0.05 '.' 0.1 ' ' 1  
##   
## Residual standard error: 33.43 on 109 degrees of freedom  
## Multiple R-squared: 2.693e-05, Adjusted R-squared: -0.009147   
## F-statistic: 0.002936 on 1 and 109 DF, p-value: 0.9569

confint(model2)

## 2.5 % 97.5 %  
## (Intercept) 29.248109 55.5826192  
## Day -0.745321 0.7056539

A. The quality of this model is not great. The correalation itself is is basically 0, the R squared value is larger than that of model1, and the intercept is lower, but the data doesnt lend itself to telling a story of a relationship between“Day” and “Ozone”.

1. The slope coefficient will likely fall between -0.75 and 0.71, in a 95% confidence interval.

Task 10

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

