library(tidyverse)

## ── Attaching core tidyverse packages ──────────────────────── tidyverse 2.0.0 ──  
## ✔ dplyr 1.1.4 ✔ readr 2.1.5  
## ✔ forcats 1.0.0 ✔ stringr 1.5.1  
## ✔ ggplot2 3.5.1 ✔ tibble 3.2.1  
## ✔ lubridate 1.9.3 ✔ tidyr 1.3.1  
## ✔ purrr 1.0.2   
## ── Conflicts ────────────────────────────────────────── tidyverse\_conflicts() ──  
## ✖ dplyr::filter() masks stats::filter()  
## ✖ dplyr::lag() masks stats::lag()  
## ℹ Use the conflicted package (<http://conflicted.r-lib.org/>) to force all conflicts to become errors

library(tidymodels)

## Warning: 程序包'tidymodels'是用R版本4.4.2 来建造的

## ── Attaching packages ────────────────────────────────────── tidymodels 1.2.0 ──  
## ✔ broom 1.0.7 ✔ rsample 1.2.1  
## ✔ dials 1.3.0 ✔ tune 1.2.1  
## ✔ infer 1.0.7 ✔ workflows 1.1.4  
## ✔ modeldata 1.4.0 ✔ workflowsets 1.1.0  
## ✔ parsnip 1.2.1 ✔ yardstick 1.3.1  
## ✔ recipes 1.1.0

## Warning: 程序包'dials'是用R版本4.4.2 来建造的

## Warning: 程序包'infer'是用R版本4.4.2 来建造的

## Warning: 程序包'modeldata'是用R版本4.4.2 来建造的

## Warning: 程序包'parsnip'是用R版本4.4.2 来建造的

## Warning: 程序包'recipes'是用R版本4.4.2 来建造的

## Warning: 程序包'rsample'是用R版本4.4.2 来建造的

## Warning: 程序包'tune'是用R版本4.4.2 来建造的

## Warning: 程序包'workflows'是用R版本4.4.2 来建造的

## Warning: 程序包'workflowsets'是用R版本4.4.2 来建造的

## Warning: 程序包'yardstick'是用R版本4.4.2 来建造的

## ── Conflicts ───────────────────────────────────────── tidymodels\_conflicts() ──  
## ✖ scales::discard() masks purrr::discard()  
## ✖ dplyr::filter() masks stats::filter()  
## ✖ recipes::fixed() masks stringr::fixed()  
## ✖ dplyr::lag() masks stats::lag()  
## ✖ yardstick::spec() masks readr::spec()  
## ✖ recipes::step() masks stats::step()  
## • Learn how to get started at https://www.tidymodels.org/start/

library(GGally)

## Warning: 程序包'GGally'是用R版本4.4.2 来建造的

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

library(lmtest)

## Warning: 程序包'lmtest'是用R版本4.4.2 来建造的

## 载入需要的程序包：zoo

## Warning: 程序包'zoo'是用R版本4.4.2 来建造的

##   
## 载入程序包：'zoo'  
##   
## The following objects are masked from 'package:base':  
##   
## as.Date, as.Date.numeric

air = airquality

nrow(air)

## [1] 153

ncol(air)

## [1] 6

any(is.na(air$Ozone)) # check if any row in Column Ozone has empty data

## [1] TRUE

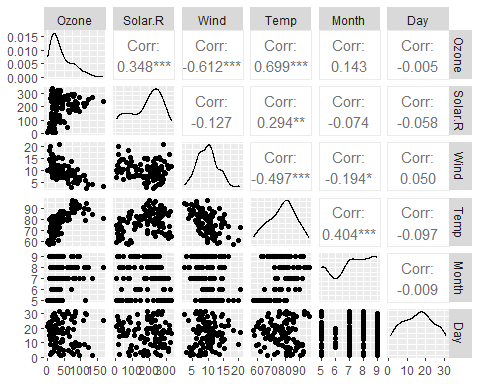
air2 = air %>% drop\_na() #drop all rows with empty data  
nrow(air2)

## [1] 111

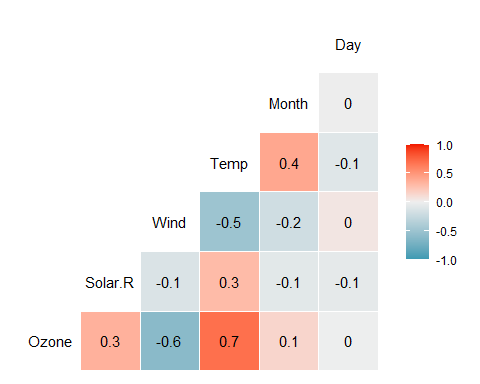
ncol(air2)

## [1] 6

#Check correlations between variables in the dataset  
ggpairs(air2)

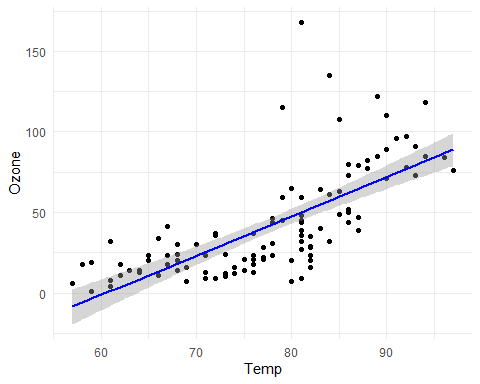


ggcorr(air2, label = TRUE)



# Plot “Temp” (x-axis) vs “Ozone” (y-axis). Describe the relationship.  
ggplot(air2, aes(Temp, Ozone)) +   
 geom\_point() +  
 geom\_smooth(method = "lm", color = "blue") +  
 theme\_minimal()

## `geom\_smooth()` using formula = 'y ~ x'



# Build the Model Using Tidymodels  
  
# Step 1: Recipe  
ozone\_recipe = recipe(Ozone ~ Temp, air2)  
  
#Step 2: Model  
ozone\_mod = set\_engine(linear\_reg(), "lm")  
  
#Step 3: Workflow  
ozone\_workflow =   
 workflow() %>%   
 add\_model(ozone\_mod) %>%  
 add\_recipe(ozone\_recipe)  
  
#Step 4: Fit the Model - Fit the workflow to the dataset  
ozone\_fit = fit(ozone\_workflow, air2)  
  
#Examine the Model  
summary(ozone\_fit$fit$fit$fit)

##   
## Call:  
## stats::lm(formula = ..y ~ ., data = data)  
##   
## 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

ozone\_fit$fit$fit$fit$coefficients #Slope

## (Intercept) Temp   
## -147.64607 2.43911

summary(ozone\_fit$fit$fit$fit)$r.squared #R Square

## [1] 0.4879601

confint(ozone\_fit$fit$fit$fit, level = 0.95) # 95% Confidence Intervals

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

predict(ozone\_fit, new\_data = tibble(Temp = 80)) #predicted “Ozone” value when “Temp” is 80

## # A tibble: 1 × 1  
## .pred  
## <dbl>  
## 1 47.5

dwtest(ozone\_fit$fit$fit$fit)

##   
## Durbin-Watson test  
##   
## data: ozone\_fit$fit$fit$fit  
## DW = 1.8644, p-value = 0.2123  
## alternative hypothesis: true autocorrelation is greater than 0

#p-value = 0.2123: Since the p-value is greater than 0.05, there is insufficient evidence to reject the null hypothesis of no autocorrelation. This implies that there is no significant autocorrelation present in your residuals.