informe

Adrián Calzadilla González

12 de diciembre de 2016

Regresión data set Wizmir

Objetivo

Este fichero contiene la información del tiempo de la ciudad de *Esmira* (*Izmir*), que es el segundo mayor puerto de Turquía tras Estambul y la tercera ciudad del país en población con 4.168.415 habitantes en 2015. El fichero posee la información del tiempo correspondiente al periodo *01/01/1994* hasta el *31/12/1997*. A partir de estas características el objetivo es predecir la temperatura media.

Análisis de datos

Hipótesisi inicial

La temperatura media se trata de promedios estadísticos cogidos a lo largo del día, por lo tanto es muy posible que la relación entre la temperatura media y las temperaturas máximas y mínimas sea importante para realizar un buen modelo.

Análisis general

Para analizar el conjunto de datos proporcionado. En primer lugar se lee y exporta el fichero con los datos de *wizmir*, se introducen los nombres de las variables a mano y se comprueba si es un *data.frame*.

```
wizmir <- read.csv2("./wizmir/wizmir.dat", header = F, sep = ",", comment.char = "@", dec = ".")
names(wizmir) <- c("Max_temperature", "Min_temperature", "Dewpoint", "Precipitation", "Sea_level_pressu

# Creamos varias funciones que utilizaremos a lo largo del código

## Funciones para crear una tabla donde poder comparar los resultados

percentage <- function(x){
  x * 100
}

createTable <- function(x){
  df <- data.frame(cbind(x), stringsAsFactors = F)
  porcentaje <- sapply(df[,1], as.numeric)
  porcentaje <- sapply(df[,1], percentage)
  df <- cbind(df, porcentaje)
  df
}

## Función para calcular la moda

getMode <- function(v) {
  uniqv <- unique(v)
  uniqv[which.max(tabulate(match(v, uniqv)))]
}
```

Seguidamente, se analiza el *data.frame* y el tipo de variables que contiene:

```
## Características generales de wizmir
attach(wizmir)
is.data.frame(wizmir)
```

```
## [1] TRUE
```

```
dim(wizmir)
```

```
## [1] 1461   10
```

```
str(wizmir)
```

```
## 'data.frame':    1461 obs. of  10 variables:
##  $ Max_temperature   : num  88.2 88 91.6 64.4 94.1 81.3 62.6 53.6 53.4 91.4 ...
##  $ Min_temperature   : num  57.2 58.6 62.1 42.8 72.3 62.6 53.6 35.6 44.1 59.9 ...
##  $ Dewpoint          : num  53.6 54.9 60.4 37.4 46.8 37.4 51.4 30.3 45.4 54.1 ...
##  $ Precipitation     : num  0 0 0 0.2 0 0 0.2 0 0 0 ...
##  $ Sea_level_pressure: num  30 29.8 29.8 30.1 29.9 ...
##  $ Standard_pressure : num  7.3 7.3 7.2 7.8 7.2 6.8 7.9 7.6 7.7 7.3 ...
##  $ Visibility        : num  9.09 10.7 8.29 21.1 17.2 21.2 17 16.1 9.09 7.37 ...
##  $ Wind_speed        : num  16.1 18.3 18.3 27.5 25.3 27.5 25.3 20.8 16.1 20.8 ...
##  $ Max_wind_speed    : num  34.3 34.3 34.3 34.3 34.3 ...
##  $ Mean_temperature  : num  74.3 75.2 76.1 47.1 83.9 70.8 59.7 44.9 47.8 77.8 ...
```

```
### Existen valores "NA"
table(is.na(wizmir))
```

```
##
## FALSE
## 14610
```

```
### Tipos de datos de las variables
is.double(Max_temperature)
```

```
## [1] TRUE
```

```
is.double(Min_temperature)
```

```
## [1] TRUE
```

```
is.double(Dewpoint)
```

```
## [1] TRUE
```

```
is.double(Precipitation)
```

```
## [1] TRUE
```

```r
is.double(Sea_level_pressure)
```

```
## [1] TRUE
```

```r
is.double(Standard_pressure)
```

```
## [1] TRUE
```

```r
is.double(Visibility)
```

```
## [1] TRUE
```

```r
is.double(Wind_speed)
```

```
## [1] TRUE
```

```r
is.double(Max_wind_speed)
```

```
## [1] TRUE
```

```r
is.double(Mean_temperature)
```

```
## [1] TRUE
```

Resumen del *data.frame* en el que se nos muestra el valor máximo y mínimo registrado, el primer cuartil, la mediana, el tercer cuartil y la media de cada variable.

```r
summary(wizmir)
```

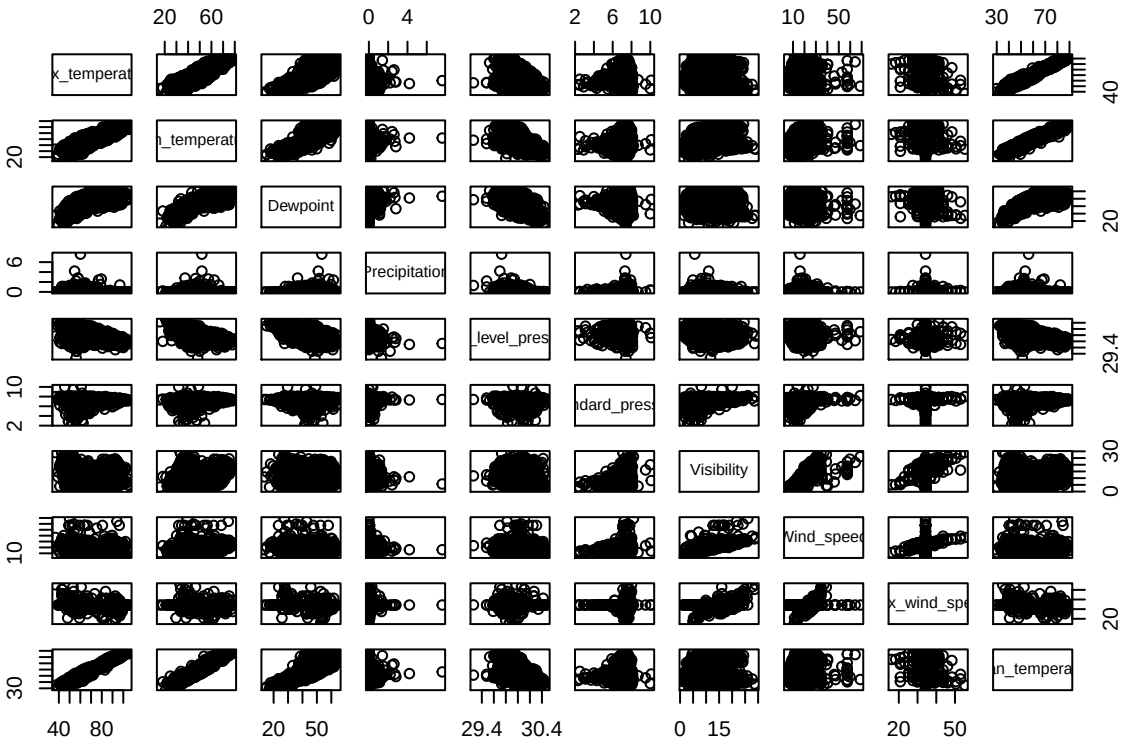
```
##  Max_temperature  Min_temperature    Dewpoint     Precipitation
##  Min.   : 36.70   Min.   :15.80   Min.   :13.60   Min.   :0.00000
##  1st Qu.: 59.00   1st Qu.:40.10   1st Qu.:41.30   1st Qu.:0.00000
##  Median : 70.70   Median :50.00   Median :48.20   Median :0.00000
##  Mean   : 72.22   Mean   :50.74   Mean   :46.62   Mean   :0.09257
##  3rd Qu.: 87.10   3rd Qu.:62.20   3rd Qu.:53.60   3rd Qu.:0.00000
##  Max.   :105.00   Max.   :78.60   Max.   :64.40   Max.   :7.60000
##  Sea_level_pressure Standard_pressure   Visibility      Wind_speed
##  Min.   :29.26      Min.   : 2.300    Min.   : 0.92   Min.   : 4.72
##  1st Qu.:29.85      1st Qu.: 7.100    1st Qu.: 6.56   1st Qu.:16.10
##  Median :29.95      Median : 7.300    Median :10.50   Median :19.81
##  Mean   :29.97      Mean   : 7.197    Mean   :11.16   Mean   :19.81
##  3rd Qu.:30.08      3rd Qu.: 7.600    3rd Qu.:15.40   3rd Qu.:23.00
##  Max.   :30.48      Max.   :10.100    Max.   :29.10   Max.   :68.80
##  Max_wind_speed  Mean_temperature
##  Min.   :16.11   Min.   :29.40
##  1st Qu.:34.28   1st Qu.:49.60
##  Median :34.28   Median :60.00
##  Mean   :34.28   Mean   :61.51
##  3rd Qu.:34.28   3rd Qu.:75.20
##  Max.   :55.24   Max.   :89.90
```

A continuación, se calculan algunas métricas que no salen en *summary*:

```
## Desviación estandar para cada una de las variables
allSd <- apply(wizmir, 2, sd)
## Mediana de cada una de las variables
allMedian <- apply(wizmir, 2, median)
## Rango intercuartílico
allIQR <- apply(wizmir, 2, IQR)
```

Por último, se grafican todas las variables todas con todas.

```
# Gráfica de todos con todos
plot(wizmir)
```



Variables

Conjunto de datos explicado uno a uno.

Max_temperature

Esta variable se utiliza para guardar la temperatura máxima obtenida en cada observación. Su unidad es el grado *Farenheit*.

Medidas de centralidad

```
mean(Max_temperature)
```

```
## [1] 72.22416
```

```r
median(Max_temperature)
```

```
## [1] 70.7
```

```r
getMode(Max_temperature)
```

```
## [1] 64.4
```

Medidas de dispersión

```r
var(Max_temperature)
```

```
## [1] 253.6602
```

```r
sd(Max_temperature)
```

```
## [1] 15.92671
```

```r
max(Max_temperature)
```

```
## [1] 105
```

```r
min(Max_temperature)
```

```
## [1] 36.7
```

```r
range(Max_temperature)
```

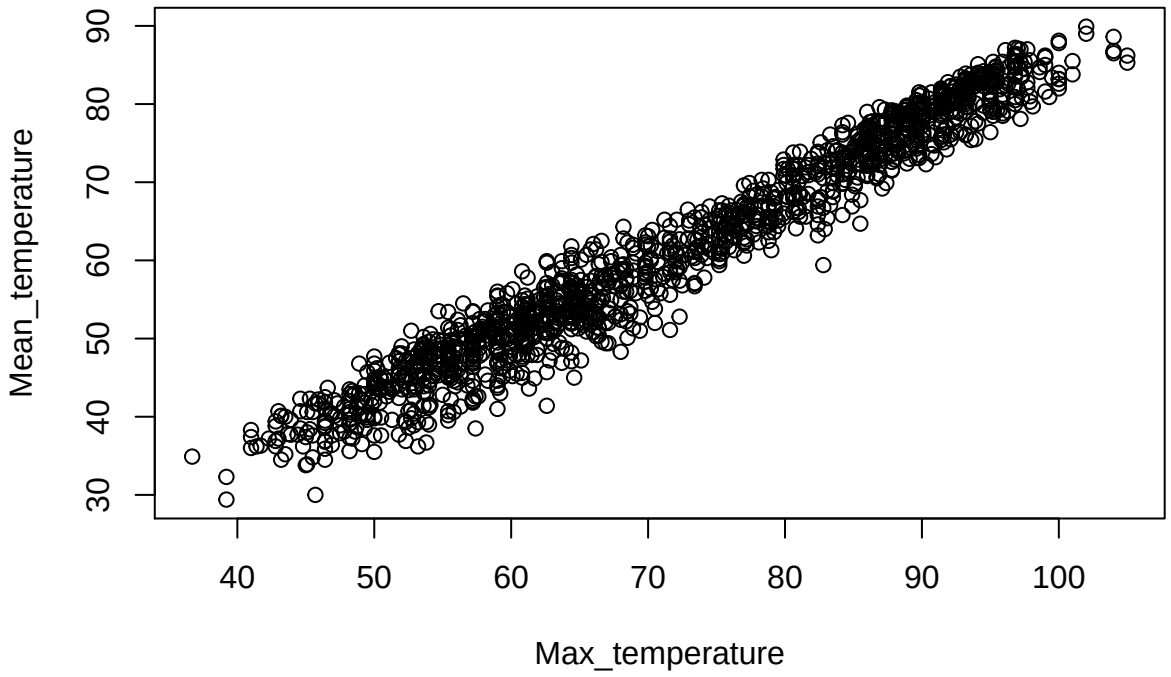
```
## [1]  36.7 105.0
```

```r
quantile(Max_temperature)
```

```
##    0%   25%   50%   75%  100%
##  36.7  59.0  70.7  87.1 105.0
```

Gráfica de Max_temperature respecto a Mean_temperature, la salida.

```r
plot(Max_temperature, Mean_temperature)
```



Min_temperature

Variable que se utiliza para guardar la temperatura mínima registrada a lo largo del día, se mide en grados *Farenheit*.

Medidas de centralidad

```
mean(Min_temperature)
```

```
## [1] 50.74025
```

```
median(Min_temperature)
```

```
## [1] 50
```

```
getMode(Min_temperature)
```

```
## [1] 42.8
```

Medidas de dispersión

```
var(Min_temperature)
```

```
## [1] 174.9292
```

```
sd(Min_temperature)
```

```
## [1] 13.22608
```

```r
max(Min_temperature)
```

```
## [1] 78.6
```

```r
min(Min_temperature)
```

```
## [1] 15.8
```

```r
range(Min_temperature)
```

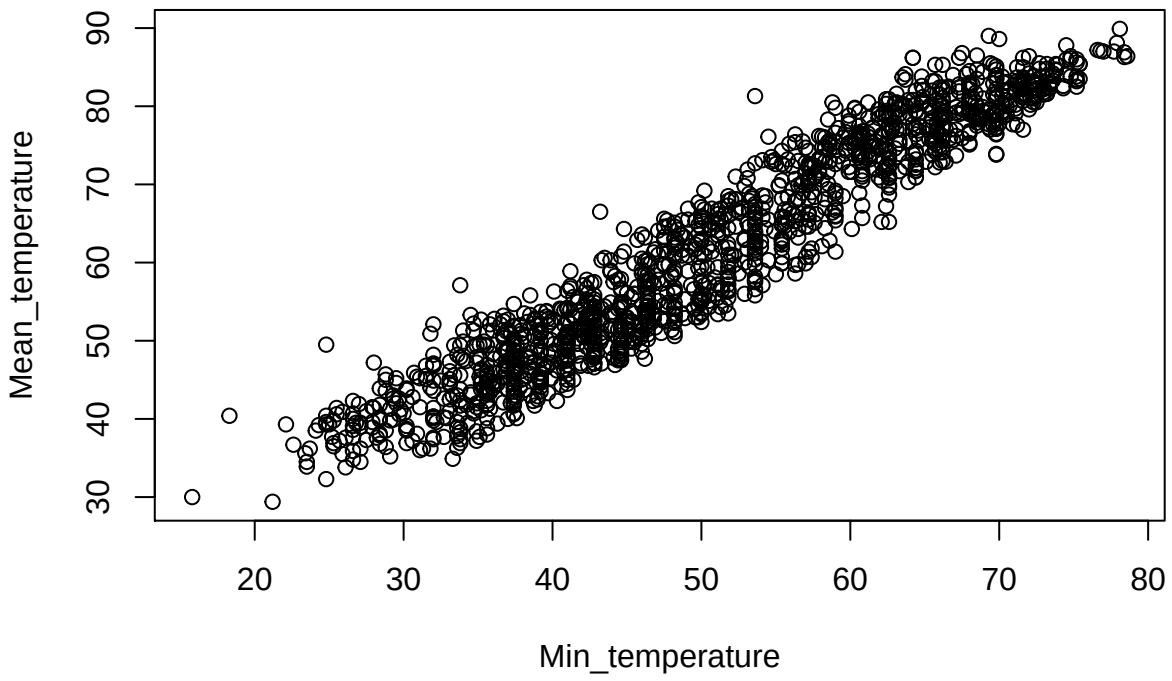
```
## [1] 15.8 78.6
```

```r
quantile(Min_temperature)
```

```
##   0%  25%  50%  75% 100%
## 15.8 40.1 50.0 62.2 78.6
```

Gráfica de Max_temperature respecto a Mean_temperature, la salida.

```r
plot(Min_temperature, Mean_temperature)
```



Dewpoint

El punto de rocío es la temperatura a la que empieza a condensarse el vapor de agua contenido en el aire. La saturación se produce por un aumento de humedad relativa con la misma temperatura, o por un descenso de temperatura con la misma humedad relativa. En este caso su unidad son los grados *Farenheit*.

Medidas de centralidad

```
mean(Dewpoint)
```

```
## [1] 46.62356
```

```
median(Dewpoint)
```

```
## [1] 48.2
```

```
getMode(Dewpoint)
```

```
## [1] 49
```

Medidas de dispersión

```
var(Dewpoint)
```

```
## [1] 87.32051
```

```
sd(Dewpoint)
```

```
## [1] 9.344545
```

```
max(Dewpoint)
```

```
## [1] 64.4
```

```
min(Dewpoint)
```

```
## [1] 13.6
```

```
range(Dewpoint)
```

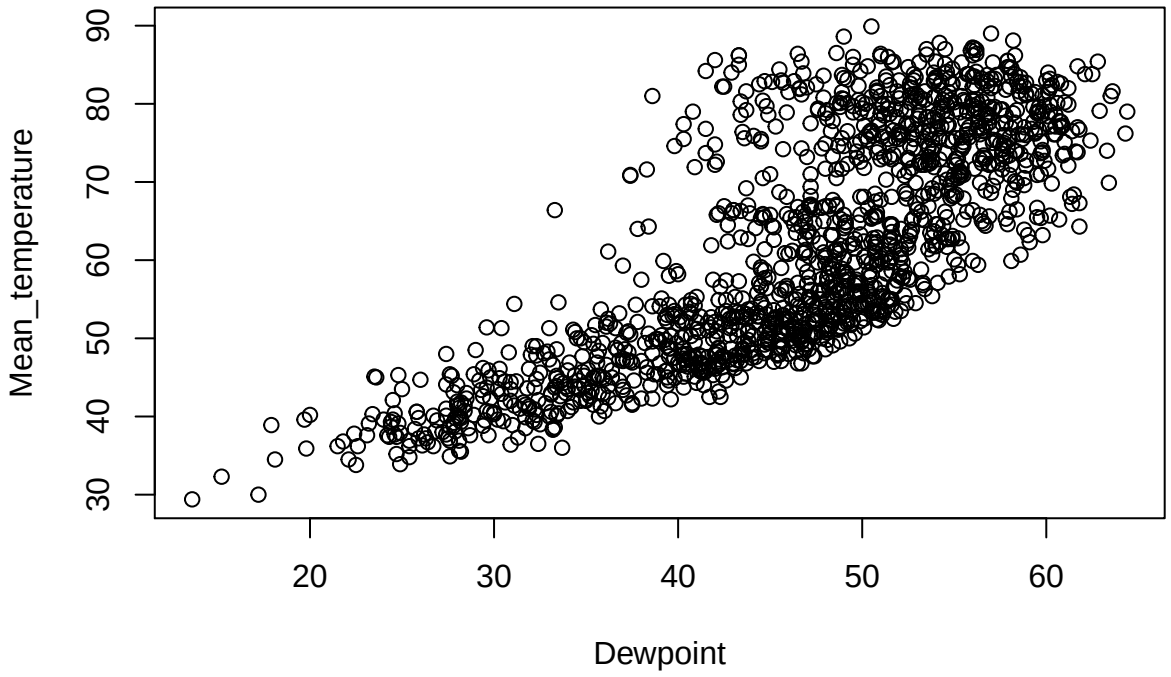
```
## [1] 13.6 64.4
```

```
quantile(Dewpoint)
```

```
##   0%  25%  50%  75% 100%
## 13.6 41.3 48.2 53.6 64.4
```

Gráfica de Dewpoint respecto a Mean_temperature, la salida.

```r
plot(Dewpoint, Mean_temperature)
```



Precipitation

Se considera *precipitación* a cualquier producto de la condensación del vapor de agua que cae bajo el efecto de la gravedad. Las principales formas de precipitación incluyen: lluvia, aguanieve, nieve y granizo.

Este fenómeno ocurre cuando una porción de la atmósfera se satura con vapor de agua, de modo que el agua se condensa y cae.

En este caso su unidad son **

Medidas de centralidad

```r
mean(Precipitation)
```

```
## [1] 0.09256674
```

```r
median(Precipitation)
```

```
## [1] 0
```

```r
getMode(Precipitation)
```

```
## [1] 0
```

Medidas de dispersión

```r
var(Precipitation)
```

```
## [1] 0.1244684
```

```r
sd(Precipitation)
```

```
## [1] 0.3528008
```

```r
max(Precipitation)
```

```
## [1] 7.6
```

```r
min(Precipitation)
```

```
## [1] 0
```

```r
range(Precipitation)
```

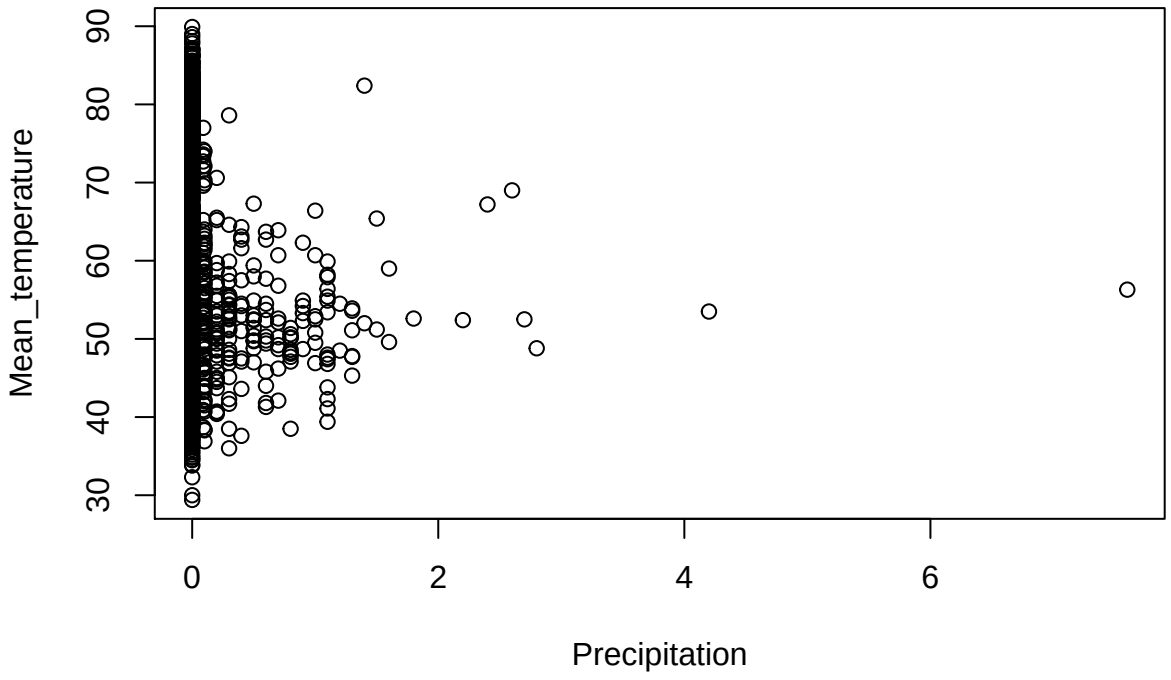
```
## [1] 0.0 7.6
```

```r
quantile(Precipitation)
```

```
##   0%  25%  50%  75% 100%
##  0.0  0.0  0.0  0.0  7.6
```

Gráfica de Precipitation respecto a Mean_temperature, la salida.

```r
plot(Precipitation, Mean_temperature)
```



Sea_level_pressure

La presión atmosférica, a cualquier elevación, se reduce a un valor que se aproxima a la presión al nivel del mar. La unidad en la que se mide esta variable es *inHg*.

Medidas de centralidad

```
mean(Sea_level_pressure)
```

```
## [1] 29.97111
```

```
median(Sea_level_pressure)
```

```
## [1] 29.95
```

```
getMode(Sea_level_pressure)
```

```
## [1] 29.94
```

Medidas de dispersión

```
var(Sea_level_pressure)
```

```
## [1] 0.02811959
```

```
sd(Sea_level_pressure)
```

```
## [1] 0.167689
```

```
max(Sea_level_pressure)
```

```
## [1] 30.48
```

```
min(Sea_level_pressure)
```

```
## [1] 29.26
```

```
range(Sea_level_pressure)
```

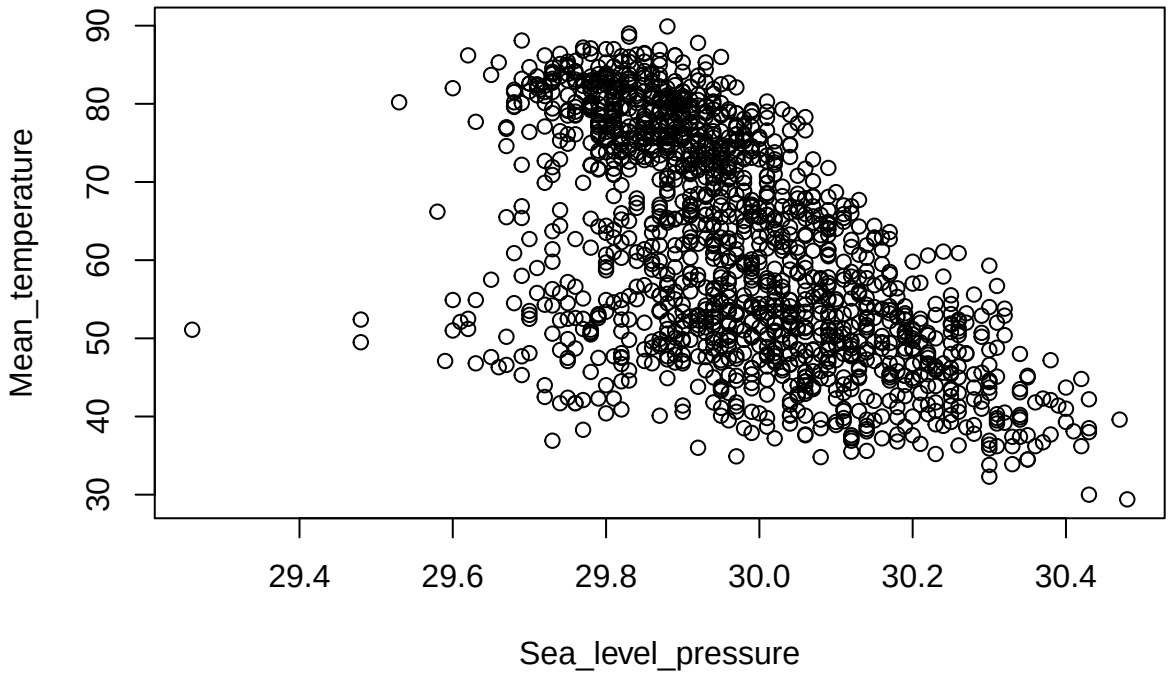
```
## [1] 29.26 30.48
```

```
quantile(Sea_level_pressure)
```

```
##    0%   25%   50%   75%  100%
## 29.26 29.85 29.95 30.08 30.48
```

Gráfica de Sea_level_pressure respecto a Mean_temperature, la salida.

```
plot(Sea_level_pressure, Mean_temperature)
```



Standard_pressure

Medidas de centralidad

```
mean(Standard_pressure)
```

```
## [1] 7.196783
```

```
median(Standard_pressure)
```

```
## [1] 7.3
```

```
getMode(Standard_pressure)
```

```
## [1] 7.2
```

Medidas de dispersión

```
var(Standard_pressure)
```

```
## [1] 0.4691609
```

```r
sd(Standard_pressure)
```

```
## [1] 0.6849532
```

```r
max(Standard_pressure)
```

```
## [1] 10.1
```

```r
min(Standard_pressure)
```

```
## [1] 2.3
```

```r
range(Standard_pressure)
```

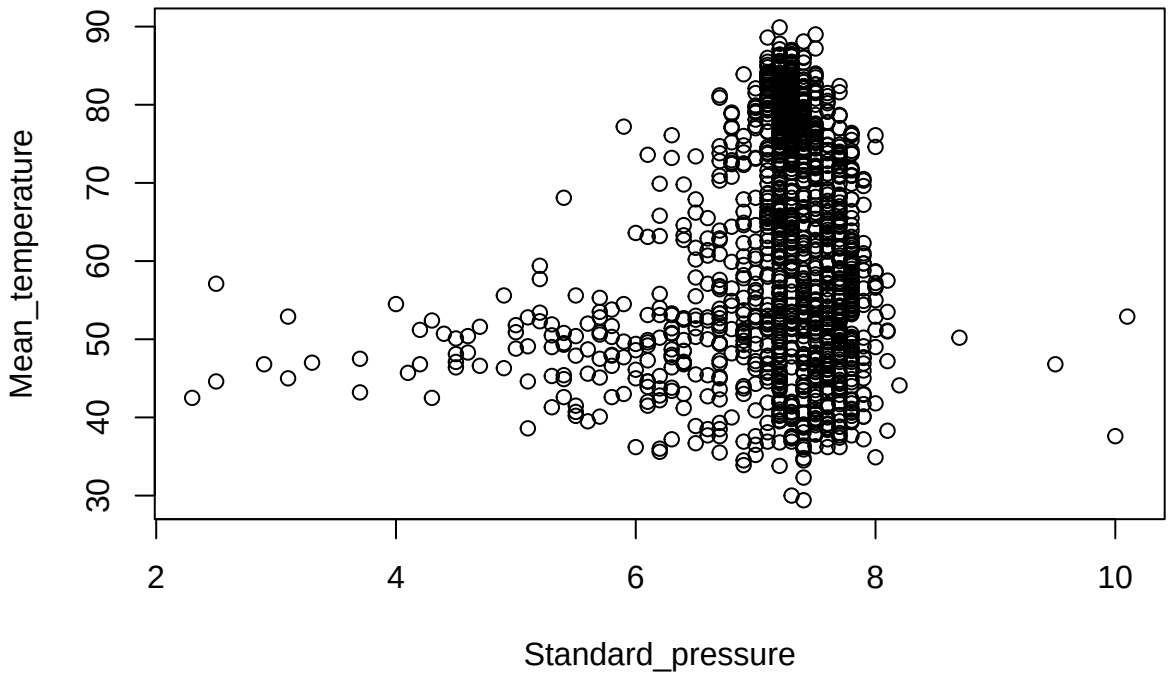
```
## [1]  2.3 10.1
```

```r
quantile(Standard_pressure)
```

```
##   0%  25%  50%  75% 100%
##  2.3  7.1  7.3  7.6 10.1
```

Gráfica de Standard_pressure respecto a Mean_temperature, la salida.

```r
plot(Standard_pressure, Mean_temperature)
```



Visibility

Medidas de centralidad

```
mean(Visibility)
```

```
## [1] 11.15803
```

```
median(Visibility)
```

```
## [1] 10.5
```

```
getMode(Visibility)
```

```
## [1] 7.14
```

Medidas de dispersión

```
var(Visibility)
```

```
## [1] 29.23202
```

```
sd(Visibility)
```

```
## [1] 5.406665
```

```
max(Visibility)
```

```
## [1] 29.1
```

```
min(Visibility)
```

```
## [1] 0.92
```

```
range(Visibility)
```

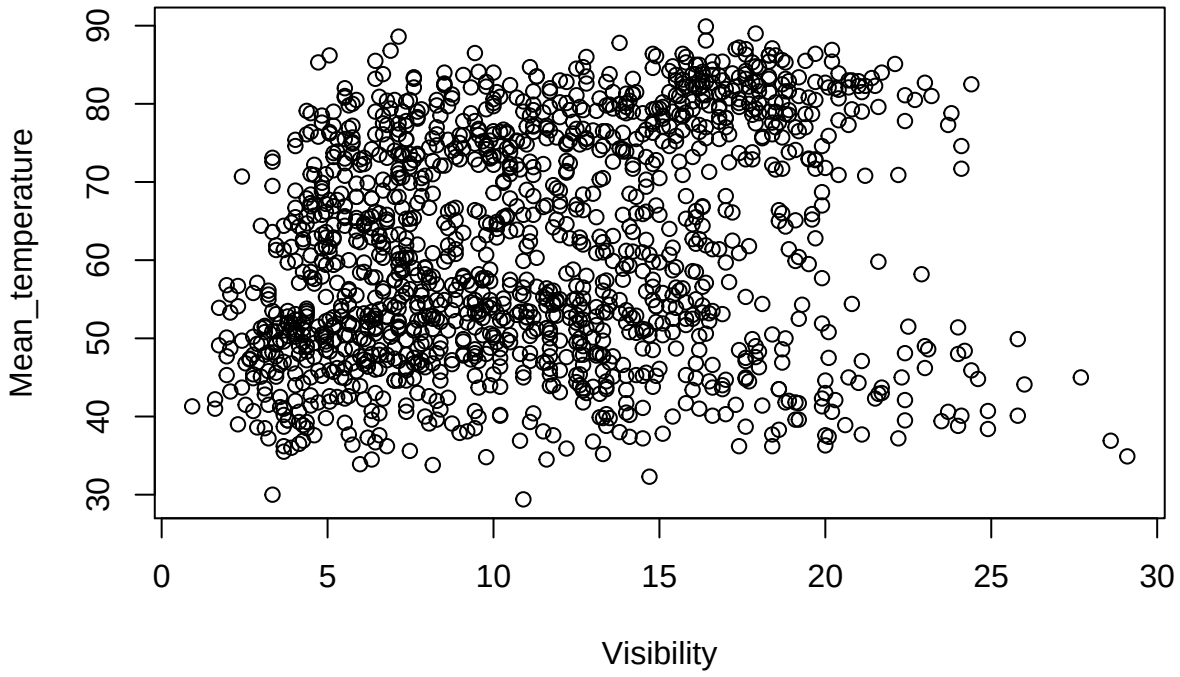
```
## [1]  0.92 29.10
```

```
quantile(Visibility)
```

```
##    0%   25%   50%   75%  100%
##  0.92  6.56 10.50 15.40 29.10
```

Gráfica de Visibility respecto a Mean_temperature, la salida.

```
plot(Visibility, Mean_temperature)
```



Wind_speed

Medidas de centralidad

```
mean(Wind_speed)
```

```
## [1] 19.81176
```

```
median(Wind_speed)
```

```
## [1] 19.81
```

```
getMode(Wind_speed)
```

```
## [1] 20.8
```

Medidas de dispersión

```
var(Wind_speed)
```

```
## [1] 50.9118
```

```r
sd(Wind_speed)
```

```
## [1] 7.13525
```

```r
max(Wind_speed)
```

```
## [1] 68.8
```

```r
min(Wind_speed)
```

```
## [1] 4.72
```

```r
range(Wind_speed)
```

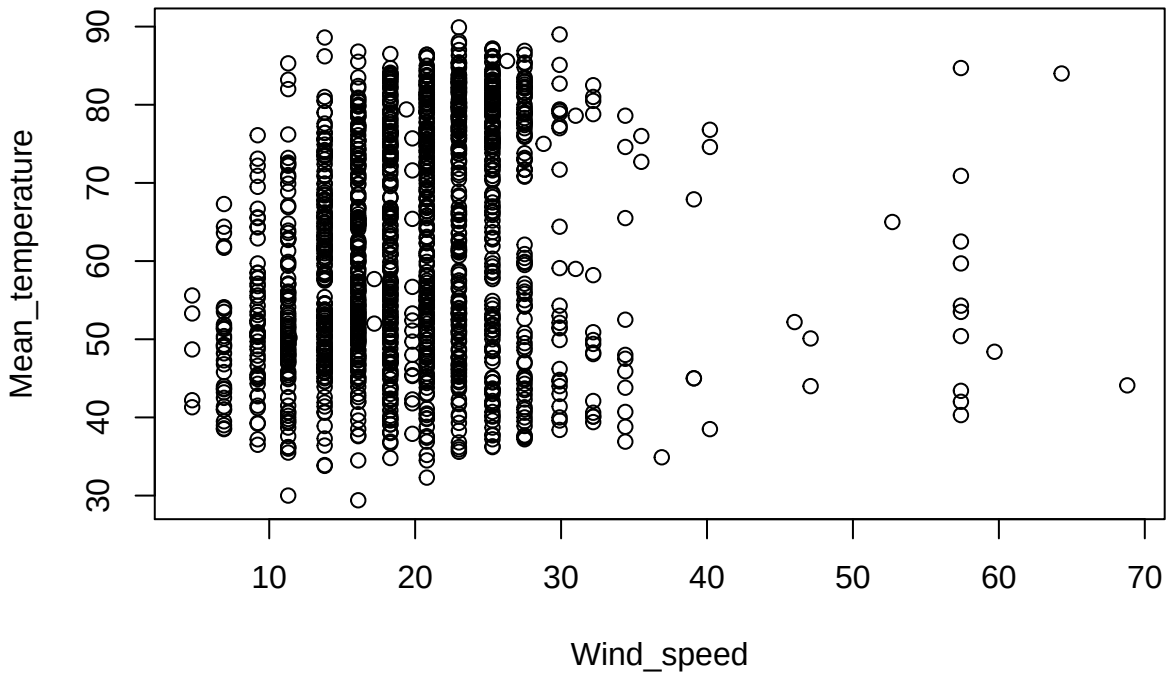
```
## [1]  4.72 68.80
```

```r
quantile(Wind_speed)
```

```
##    0%   25%   50%   75%  100%
##  4.72 16.10 19.81 23.00 68.80
```

Gráfica de Wind_speed respecto a Mean_temperature, la salida.

```r
plot(Wind_speed, Mean_temperature)
```



Max_wind_speed

Medidas de centralidad

```
mean(Max_wind_speed)
```

```
## [1] 34.28037
```

```
median(Max_wind_speed)
```

```
## [1] 34.28
```

```
getMode(Max_wind_speed)
```

```
## [1] 34.28
```

Medidas de dispersión

```
var(Max_wind_speed)
```

```
## [1] 5.962512
```

```
sd(Max_wind_speed)
```

```
## [1] 2.441826
```

```
max(Max_wind_speed)
```

```
## [1] 55.24
```

```
min(Max_wind_speed)
```

```
## [1] 16.11
```

```
range(Max_wind_speed)
```

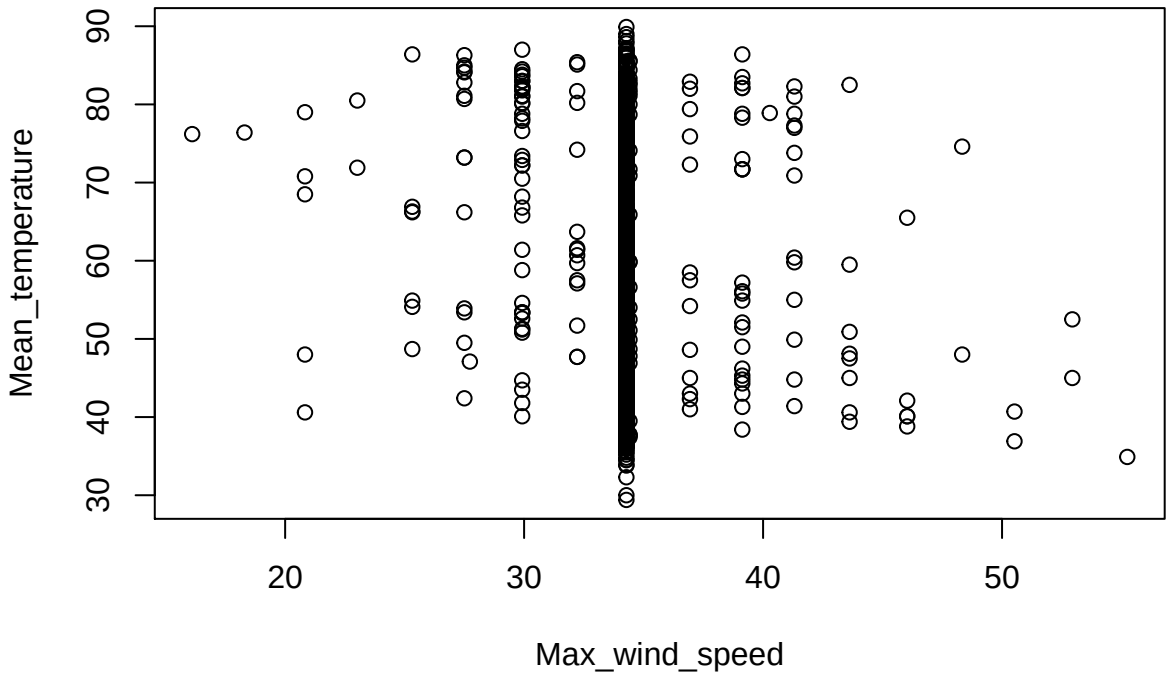
```
## [1] 16.11 55.24
```

```
quantile(Max_wind_speed)
```

```
##    0%   25%   50%   75%  100%
## 16.11 34.28 34.28 34.28 55.24
```

Gráfica de Max_wind_speed respecto a Mean_temperature, la salida.

```r
plot(Max_wind_speed, Mean_temperature)
```



Regresión lineal

Ajuste lineal simple

Después de graficar cada una de las variables contra la salida, se toman como variables candidatas para el ajuste lineal simple a *Max_temperature* y *Min_temperature* como favoritas. Además también se calcula la regresión lineal a las variables: *Dewpoint*, *Sea_level_pressure* y *Standard_pressure*.

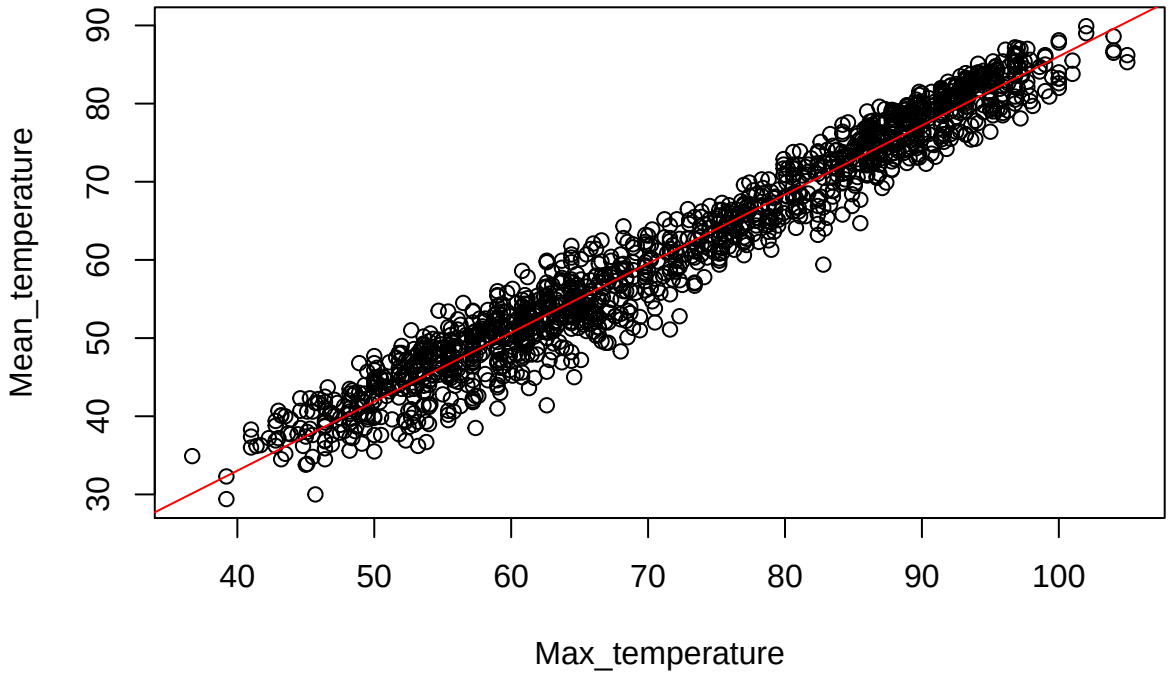
Max_temperature

```r
# wizmirTrain <- read.csv2("./wizmir/wizmir-5-1tra.dat", header = F, sep = ",", dec = ".", comment.char
lm1 <- lm(Mean_temperature~Max_temperature, data = wizmir)
summary(lm1)
```

```
##
## Call:
## lm(formula = Mean_temperature ~ Max_temperature, data = wizmir)
##
## Residuals:
##     Min      1Q  Median      3Q     Max
## -11.606  -1.901   0.352   2.140   7.472
##
## Coefficients:
##                  Estimate Std. Error t value Pr(>|t|)
## (Intercept)     -2.290034   0.359675  -6.367 2.58e-10 ***
## Max_temperature  0.883328   0.004863 181.634  < 2e-16 ***
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
```

```
##
## Residual standard error: 2.96 on 1459 degrees of freedom
## Multiple R-squared:  0.9576, Adjusted R-squared:  0.9576
## F-statistic: 3.299e+04 on 1 and 1459 DF,  p-value: < 2.2e-16
```

```
plot(Mean_temperature~Max_temperature, wizmir)
abline(lm1, col="red")
```




```
confint(lm1)
```

```
##                       2.5 %     97.5 %
## (Intercept)     -2.9955695 -1.5844977
## Max_temperature  0.8737884  0.8928678
```

```
# Lista para ir introduciendo los resultados del r2 ajustado.
adj.r.squared <- list(lm1=round(summary(lm1)$adj.r.squared, digits = 4))
```

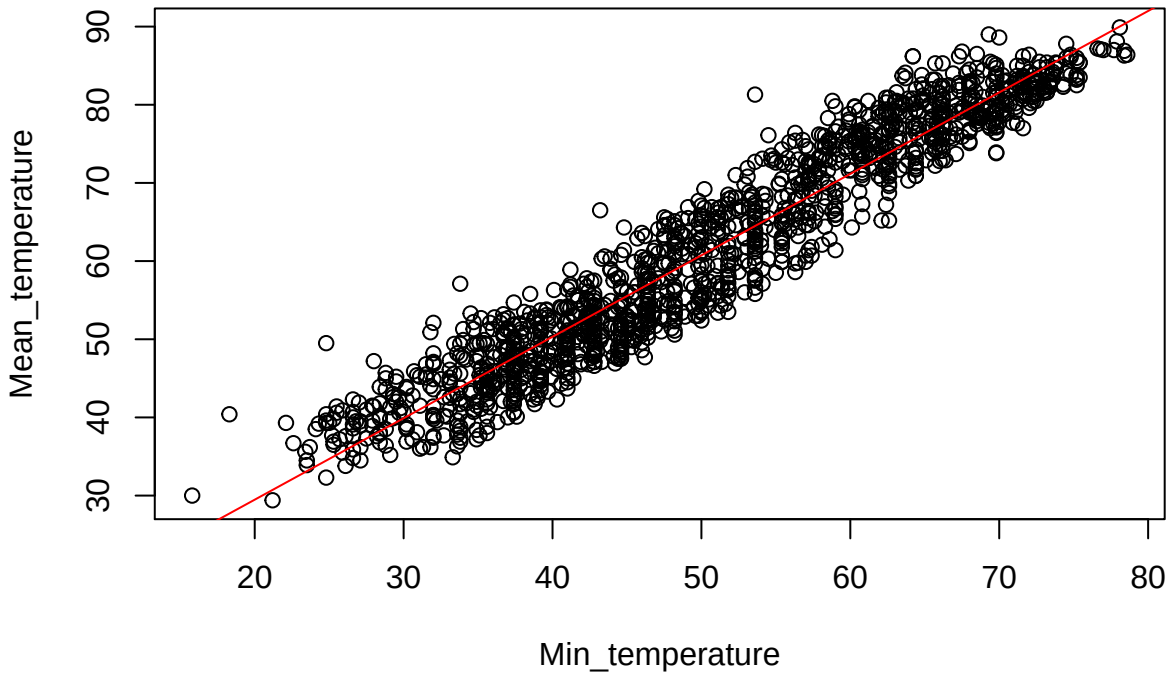
Min_temperature

```
lm2 <- lm(Mean_temperature~Min_temperature, data = wizmir)
summary(lm2)
```

```
##
## Call:
## lm(formula = Mean_temperature ~ Min_temperature, data = wizmir)
##
## Residuals:
##     Min      1Q  Median      3Q     Max
## -9.1119 -3.0412 -0.2556  3.1368 16.8125
```

```
##
## Coefficients:
##                  Estimate Std. Error t value Pr(>|t|)
## (Intercept)     8.635132   0.424606   20.34   <2e-16 ***
## Min_temperature 1.042022   0.008098  128.68   <2e-16 ***
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Residual standard error: 4.092 on 1459 degrees of freedom
## Multiple R-squared:  0.919,  Adjusted R-squared:  0.919
## F-statistic: 1.656e+04 on 1 and 1459 DF,  p-value: < 2.2e-16
```

```
plot(Mean_temperature~Min_temperature, wizmir)
abline(lm2, col="red")
```




```
confint(lm2)
```

```
##                    2.5 %   97.5 %
## (Intercept)     7.802228 9.468035
## Min_temperature 1.026138 1.057907
```

```
adj.r.squared <- c(adj.r.squared, lm2=round(summary(lm2)$adj.r.squared, digits = 4))
```

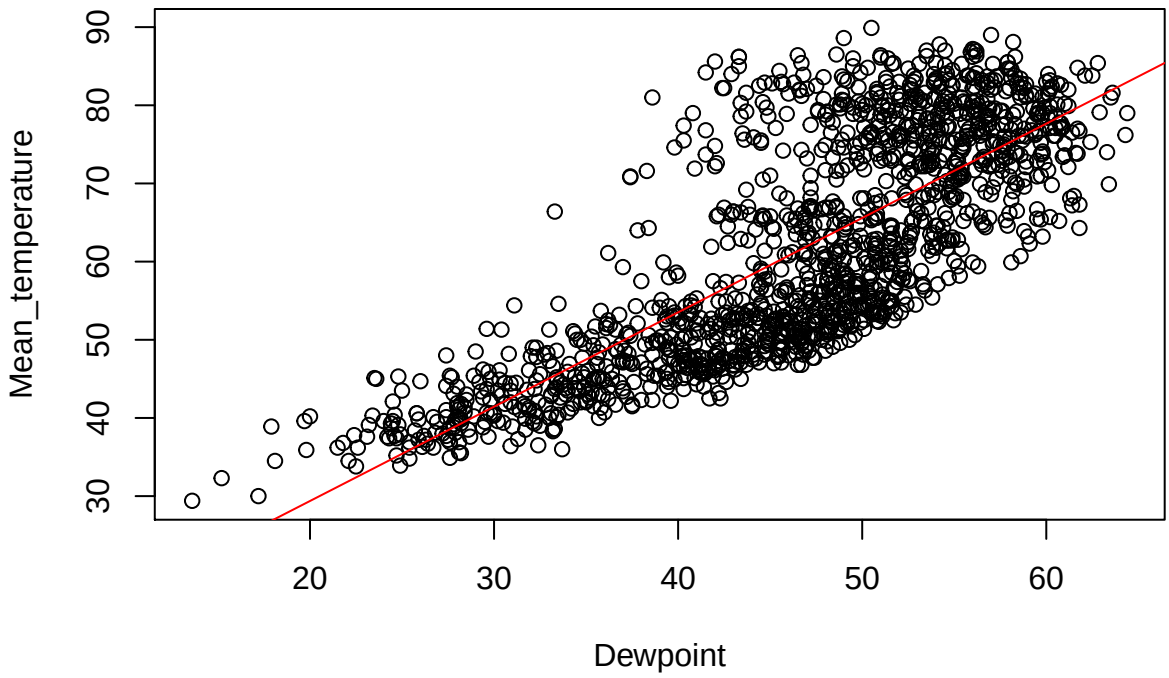
Dewpoint

```
lm3 <- lm(Mean_temperature~Dewpoint, data = wizmir)
summary(lm3)
```

```
##
```

```
## Call:
## lm(formula = Mean_temperature ~ Dewpoint, data = wizmir)
##
## Residuals:
##     Min      1Q  Median      3Q     Max
## -15.525  -6.976  -1.101   5.441  29.673
##
## Coefficients:
##             Estimate Std. Error t value Pr(>|t|)
## (Intercept)  5.23422    1.18759   4.407 1.12e-05 ***
## Dewpoint     1.20697    0.02498  48.326  < 2e-16 ***
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Residual standard error: 8.918 on 1459 degrees of freedom
## Multiple R-squared:  0.6155, Adjusted R-squared:  0.6152
## F-statistic:  2335 on 1 and 1459 DF,  p-value: < 2.2e-16
```

```
plot(Mean_temperature~Dewpoint, wizmir)
abline(lm3, col= "red")
```




```
confint(lm3)
```

```
##                 2.5 %   97.5 %
## (Intercept) 2.904660 7.563774
## Dewpoint    1.157982 1.255965
```

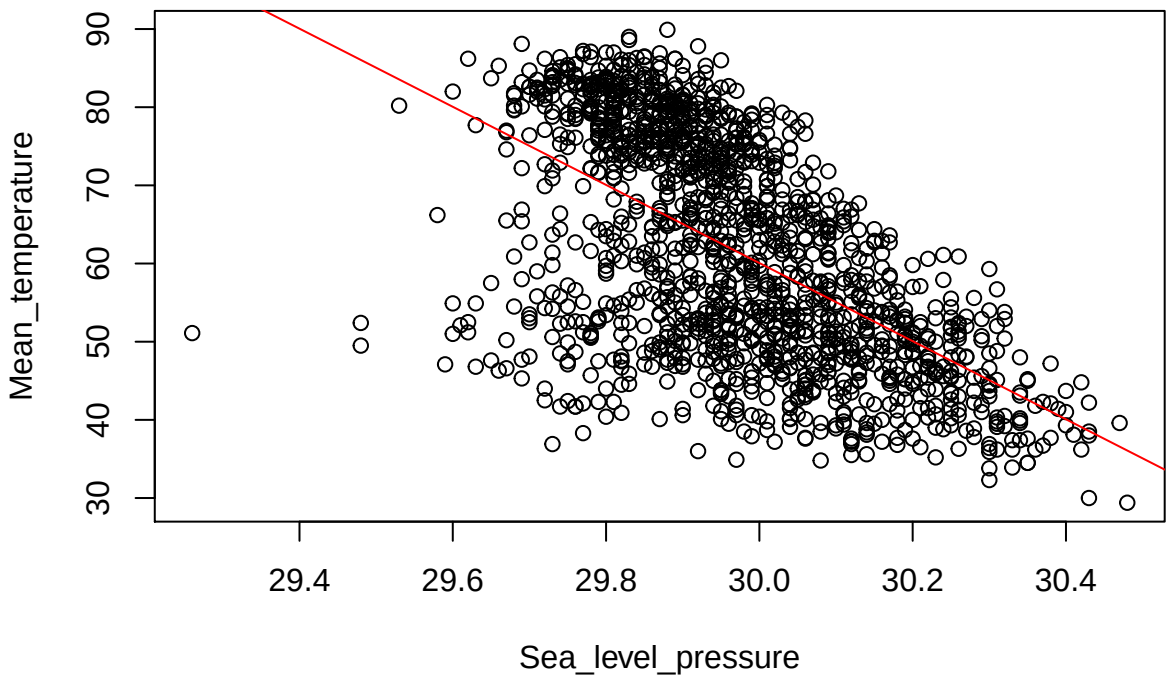
```
adj.r.squared <- c(adj.r.squared, lm3=round(summary(lm3)$adj.r.squared, digits = 4))
```

Sea_level_pressure

```r
lm4 <- lm(Mean_temperature~Sea_level_pressure, data = wizmir)
summary(lm4)
```

```
##
## Call:
## lm(formula = Mean_temperature ~ Sea_level_pressure, data = wizmir)
##
## Residuals:
##     Min      1Q  Median      3Q     Max
## -45.946  -8.260   1.138   9.441  23.839
##
## Coefficients:
##                    Estimate Std. Error t value Pr(>|t|)
## (Intercept)        1559.337     54.658   28.53   <2e-16 ***
## Sea_level_pressure  -49.976      1.824  -27.40   <2e-16 ***
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Residual standard error: 11.68 on 1459 degrees of freedom
## Multiple R-squared:  0.3398, Adjusted R-squared:  0.3394
## F-statistic:   751 on 1 and 1459 DF,  p-value: < 2.2e-16
```

```r
plot(Mean_temperature~Sea_level_pressure, wizmir)
abline(lm4, col = "red")
```




```r
confint(lm4)
```

```
##                        2.5 %     97.5 %
## (Intercept)        1452.11939 1666.55410
## Sea_level_pressure  -53.55307  -46.39847
```

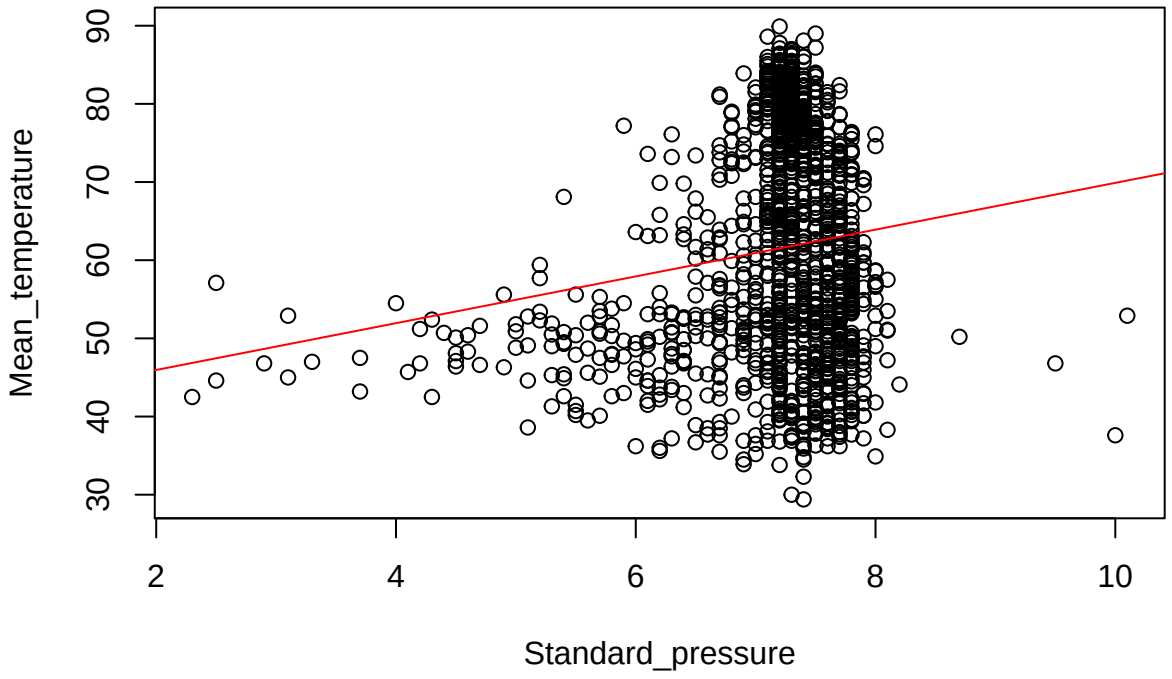
```
adj.r.squared <- c(adj.r.squared, lm4=round(summary(lm4)$adj.r.squared, digits = 4))
```

Standard_pressure

```
lm5 <- lm(Mean_temperature~Standard_pressure, data = wizmir)
summary(lm5)
```

```
##
## Call:
## lm(formula = Mean_temperature ~ Standard_pressure, data = wizmir)
##
## Residuals:
##     Min      1Q  Median      3Q     Max
## -32.716 -11.117  -1.413  13.280  28.383
##
## Coefficients:
##                   Estimate Std. Error t value Pr(>|t|)
## (Intercept)        39.9750     3.9318  10.167  < 2e-16 ***
## Standard_pressure   2.9920     0.5439   5.501 4.45e-08 ***
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Residual standard error: 14.23 on 1459 degrees of freedom
## Multiple R-squared:  0.02032,    Adjusted R-squared:  0.01965
## F-statistic: 30.26 on 1 and 1459 DF,  p-value: 4.448e-08
```

```
plot(Mean_temperature~Standard_pressure, data = wizmir)
abline(lm5, col = "red")
```



```
confint(lm5)
```

```
##                       2.5 %    97.5 %
## (Intercept)       32.262400 47.687647
## Standard_pressure  1.925113  4.058831
```

```
adj.r.squared <- c(adj.r.squared, lm5=round(summary(lm5)$adj.r.squared, digits = 4))
```

Conclusiones

Para ver que ajuste lineal es mejor se va a utlizar el *R2 ajustado* ya que, *R2* aumenta automáticamente a medida que se agrega nuevas variables independientes a una ecuación de regresión. *R2 ajustado* aumenta sólo cuando se agrega nuevas variables independientes que hacen aumentar el poder explicativo de la ecuación de regresión.

Al finalizar el ajuste lineal simple de las variables seleccionadas, se ha comprobado que tanto *Max_temperature* como *Min_temperature* tienen un valor de *R ajustado* muy bueno superior al 91%. Las otras variables quedan descartadas debido a que su *R ajustado* es muy bajo girando en torno al 61% la mejor de ellas, tal como se muestra en la siguiente tabla:

```
createTable(adj.r.squared)
```

```
##          x porcentaje
## lm1 0.9576      95.76
## lm2  0.919      91.90
## lm3 0.6152      61.52
## lm4 0.3394      33.94
## lm5 0.0196       1.96
```

Modelo lineal múltiple

En primer lugar calculamos el modelo lineal para las dos mejores valores de *R ajustado* conocidos hasta ahora: *Max_temperature* y *Min_temperature*.

Max_temperature y Min_temperature

```
lm6 <- lm(Mean_temperature~Max_temperature+Min_temperature, data = wizmir)
summary(lm6)
```

```
##
## Call:
## lm(formula = Mean_temperature ~ Max_temperature + Min_temperature,
##     data = wizmir)
##
## Residuals:
##     Min      1Q  Median      3Q     Max
## -8.3783 -0.7617 -0.0479  0.7314  7.3927
##
## Coefficients:
##                  Estimate Std. Error t value Pr(>|t|)
## (Intercept)     -0.934295   0.161645   -5.78 9.13e-09 ***
```

```
## Max_temperature  0.547739   0.004895  111.90  < 2e-16 ***
## Min_temperature  0.450962   0.005894   76.51  < 2e-16 ***
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Residual standard error: 1.322 on 1458 degrees of freedom
## Multiple R-squared:  0.9916, Adjusted R-squared:  0.9915
## F-statistic: 8.559e+04 on 2 and 1458 DF,  p-value: < 2.2e-16
```

```
adj.r.squared <- c(adj.r.squared, lm6=round(summary(lm6)$adj.r.squared, digits = 4))
```

Modelo con todas las variables

```
lm7 <- lm(Mean_temperature~., data = wizmir)
summary(lm7)
```

```
##
## Call:
## lm(formula = Mean_temperature ~ ., data = wizmir)
##
## Residuals:
##     Min      1Q  Median      3Q     Max
## -8.2364 -0.7082 -0.0368  0.7096  6.1974
##
## Coefficients:
##                     Estimate Std. Error t value Pr(>|t|)
## (Intercept)        57.317459   8.225439   6.968 4.85e-12 ***
## Max_temperature     0.575097   0.005759  99.866  < 2e-16 ***
## Min_temperature     0.367019   0.009271  39.589  < 2e-16 ***
## Dewpoint            0.044901   0.007304   6.148 1.01e-09 ***
## Precipitation      -0.061384   0.102212  -0.601 0.548227
## Sea_level_pressure -1.931807   0.268079  -7.206 9.22e-13 ***
## Standard_pressure   0.062776   0.051095   1.229 0.219412
## Visibility          0.086336   0.012295   7.022 3.34e-12 ***
## Wind_speed          0.002370   0.007237   0.327 0.743339
## Max_wind_speed     -0.047252   0.014177  -3.333 0.000881 ***
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Residual standard error: 1.256 on 1451 degrees of freedom
## Multiple R-squared:  0.9924, Adjusted R-squared:  0.9924
## F-statistic: 2.109e+04 on 9 and 1451 DF,  p-value: < 2.2e-16
```

```
adj.r.squared <- c(adj.r.squared, lm7=round(summary(lm7)$adj.r.squared, digits = 4))
```

Eliminación variables no relevantes

```
lm8 <- lm(Mean_temperature~.-Precipitation-Wind_speed, data = wizmir)
summary(lm8)
```

```
##
## Call:
## lm(formula = Mean_temperature ~ . - Precipitation - Wind_speed,
##     data = wizmir)
##
## Residuals:
##     Min      1Q  Median      3Q     Max
## -8.2411 -0.7057 -0.0322  0.7086  6.2103
##
## Coefficients:
##                     Estimate Std. Error t value Pr(>|t|)
## (Intercept)        56.669473   8.021645   7.065 2.49e-12 ***
## Max_temperature     0.576301   0.005463 105.484  < 2e-16 ***
## Min_temperature     0.366055   0.009152  39.995  < 2e-16 ***
## Dewpoint            0.044627   0.007287   6.125 1.17e-09 ***
## Sea_level_pressure -1.910630   0.261468  -7.307 4.48e-13 ***
## Standard_pressure   0.061106   0.050909   1.200 0.230215
## Visibility          0.089494   0.009618   9.305  < 2e-16 ***
## Max_wind_speed     -0.047075   0.014146  -3.328 0.000897 ***
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Residual standard error: 1.255 on 1453 degrees of freedom
## Multiple R-squared:  0.9924, Adjusted R-squared:  0.9924
## F-statistic: 2.714e+04 on 7 and 1453 DF,  p-value: < 2.2e-16
```

```
adj.r.squared <- c(adj.r.squared, lm8=round(summary(lm8)$adj.r.squared, digits = 4))
```

```
lm9 <- lm(Mean_temperature~.-Precipitation-Wind_speed-Standard_pressure-Max_wind_speed, data = wizmir)
summary(lm9)
```

```
##
## Call:
## lm(formula = Mean_temperature ~ . - Precipitation - Wind_speed -
##     Standard_pressure - Max_wind_speed, data = wizmir)
##
## Residuals:
##     Min      1Q  Median      3Q     Max
## -8.2343 -0.7145 -0.0398  0.7017  5.9728
##
## Coefficients:
##                     Estimate Std. Error t value Pr(>|t|)
## (Intercept)        56.422809   7.945011   7.102 1.92e-12 ***
## Max_temperature     0.576421   0.005483 105.122  < 2e-16 ***
## Min_temperature     0.369002   0.009147  40.342  < 2e-16 ***
## Dewpoint            0.042461   0.007288   5.826 6.97e-09 ***
## Sea_level_pressure -1.941377   0.260934  -7.440 1.71e-13 ***
## Visibility          0.083852   0.009211   9.103  < 2e-16 ***
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Residual standard error: 1.26 on 1455 degrees of freedom
## Multiple R-squared:  0.9923, Adjusted R-squared:  0.9923
## F-statistic: 3.772e+04 on 5 and 1455 DF,  p-value: < 2.2e-16
```

```
adj.r.squared <- c(adj.r.squared, lm9=round(summary(lm9)$adj.r.squared, digits = 4))
```

```
lm10 <- lm(Mean_temperature~.-Precipitation-Wind_speed-Standard_pressure-Max_wind_speed-Dewpoint, data =
summary(lm10)
```

```
##
## Call:
## lm(formula = Mean_temperature ~ . - Precipitation - Wind_speed -
##     Standard_pressure - Max_wind_speed - Dewpoint, data = wizmir)
##
## Residuals:
##     Min      1Q  Median      3Q     Max
## -8.2321 -0.7288 -0.0668  0.7313  6.4240
##
## Coefficients:
##                     Estimate Std. Error t value Pr(>|t|)
## (Intercept)        67.031221   7.820552   8.571  < 2e-16 ***
## Max_temperature     0.572009   0.005492 104.155  < 2e-16 ***
## Min_temperature     0.399317   0.007607  52.490  < 2e-16 ***
## Sea_level_pressure -2.258427   0.258067  -8.751  < 2e-16 ***
## Visibility          0.052847   0.007603   6.951 5.45e-12 ***
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Residual standard error: 1.274 on 1456 degrees of freedom
## Multiple R-squared:  0.9922, Adjusted R-squared:  0.9921
## F-statistic: 4.609e+04 on 4 and 1456 DF,  p-value: < 2.2e-16
```

```
adj.r.squared <- c(adj.r.squared, lm10=round(summary(lm10)$adj.r.squared, digits = 4))
```

Conclusiones

La unión de las variables *Max_temperature* y *Min_temperature* nos da un *R cuadrado ajustado* de 0.9915 mientras que las de la unión de todas las variables da como resultado 0.9924. Si se quitan las variables apenas hay diferencias respecto al resultado por lo que para una sencilla interpretación de la salida es mejor utilizar el *lm6* por su simplicidad a la hora de elegir variables.

```
createTable(adj.r.squared)
```

```
##           x porcentaje
## lm1  0.9576      95.76
## lm2   0.919      91.90
## lm3  0.6152      61.52
## lm4  0.3394      33.94
## lm5  0.0196       1.96
## lm6  0.9915      99.15
## lm7  0.9924      99.24
## lm8  0.9924      99.24
## lm9  0.9923      99.23
## lm10 0.9921      99.21
```

Interacciones

Para realizar las *interacciones* se usarán las cinco variables escogidas anteriormente para el ajuste lineal simple.

```
lm11 <- lm(Mean_temperature~Max_temperature*Min_temperature, data = wizmir)
summary(lm11)
```

```
##
## Call:
## lm(formula = Mean_temperature ~ Max_temperature * Min_temperature,
##     data = wizmir)
##
## Residuals:
##     Min      1Q  Median      3Q     Max
## -8.3212 -0.7603 -0.0658  0.7496  7.2755
##
## Coefficients:
##                                  Estimate Std. Error t value Pr(>|t|)
## (Intercept)                     0.1869122  0.7085095   0.264    0.792
## Max_temperature                 0.5315974  0.0110708  48.018   <2e-16 ***
## Min_temperature                 0.4277485  0.0154499  27.686   <2e-16 ***
## Max_temperature:Min_temperature 0.0003172  0.0001952   1.625    0.104
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Residual standard error: 1.321 on 1457 degrees of freedom
## Multiple R-squared:  0.9916, Adjusted R-squared:  0.9916
## F-statistic: 5.712e+04 on 3 and 1457 DF,  p-value: < 2.2e-16
```

```
adj.r.squared <- c(adj.r.squared, lm11=round(summary(lm11)$adj.r.squared, digits = 4))
```

```
lm12 <- lm(Mean_temperature~Max_temperature*Dewpoint, data = wizmir)
summary(lm12)
```

```
##
## Call:
## lm(formula = Mean_temperature ~ Max_temperature * Dewpoint, data = wizmir)
##
## Residuals:
##      Min       1Q   Median       3Q      Max
## -11.2905  -1.9209   0.3356   1.9794   6.0783
##
## Coefficients:
##                            Estimate Std. Error t value Pr(>|t|)
## (Intercept)              -0.6561922  1.6386485  -0.400  0.68889
## Max_temperature           0.7331318  0.0272498  26.904  < 2e-16 ***
## Dewpoint                  0.0892242  0.0354511   2.517  0.01195 *
## Max_temperature:Dewpoint  0.0014528  0.0005384   2.699  0.00705 **
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Residual standard error: 2.738 on 1457 degrees of freedom
```

```
## Multiple R-squared:  0.9638, Adjusted R-squared:  0.9637
## F-statistic: 1.293e+04 on 3 and 1457 DF,  p-value: < 2.2e-16
```

```
adj.r.squared <- c(adj.r.squared, lm12=round(summary(lm12)$adj.r.squared, digits = 4))
```

```
lm13 <- lm(Mean_temperature~Max_temperature*Sea_level_pressure, data = wizmir)
summary(lm13)
```

```
##
## Call:
## lm(formula = Mean_temperature ~ Max_temperature * Sea_level_pressure,
##     data = wizmir)
##
## Residuals:
##      Min       1Q   Median       3Q      Max
## -10.6433  -1.6697   0.2502   1.8987   7.0152
##
## Coefficients:
##                                    Estimate Std. Error t value Pr(>|t|)
## (Intercept)                       137.59207   56.52928   2.434 0.015053
## Max_temperature                     3.22597    0.85722   3.763 0.000174
## Sea_level_pressure                 -4.53465    1.88721  -2.403 0.016393
## Max_temperature:Sea_level_pressure -0.08005    0.02866  -2.793 0.005285
##
## (Intercept)                       *
## Max_temperature                   ***
## Sea_level_pressure                *
## Max_temperature:Sea_level_pressure **
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Residual standard error: 2.608 on 1457 degrees of freedom
## Multiple R-squared:  0.9672, Adjusted R-squared:  0.9671
## F-statistic: 1.431e+04 on 3 and 1457 DF,  p-value: < 2.2e-16
```

```
adj.r.squared <- c(adj.r.squared, lm13=round(summary(lm13)$adj.r.squared, digits = 4))
```

```
lm14 <- lm(Mean_temperature~Max_temperature*Standard_pressure, data = wizmir)
summary(lm14)
```

```
##
## Call:
## lm(formula = Mean_temperature ~ Max_temperature * Standard_pressure,
##     data = wizmir)
##
## Residuals:
##      Min       1Q   Median       3Q      Max
## -12.8282  -1.7938   0.3241   2.1088   9.4965
##
## Coefficients:
##                                  Estimate Std. Error t value Pr(>|t|)
## (Intercept)                      21.51147    5.05884   4.252 2.25e-05 ***
```

```
## Max_temperature                    0.37079    0.08367   4.432 1.00e-05 ***
## Standard_pressure                 -3.22944    0.69603  -4.640 3.80e-06 ***
## Max_temperature:Standard_pressure  0.06999    0.01150   6.087 1.47e-09 ***
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Residual standard error: 2.853 on 1457 degrees of freedom
## Multiple R-squared:  0.9607, Adjusted R-squared:  0.9606
## F-statistic: 1.187e+04 on 3 and 1457 DF,  p-value: < 2.2e-16
```

```
adj.r.squared <- c(adj.r.squared, lm14=round(summary(lm14)$adj.r.squared, digits = 4))
```

```
lm15 <- lm(Mean_temperature~Min_temperature*Dewpoint, data = wizmir)
summary(lm15)
```

```
##
## Call:
## lm(formula = Mean_temperature ~ Min_temperature * Dewpoint, data = wizmir)
##
## Residuals:
##     Min      1Q  Median      3Q     Max
## -9.9144 -3.0288 -0.0883  2.9532 15.4280
##
## Coefficients:
##                           Estimate Std. Error t value Pr(>|t|)
## (Intercept)              1.038e+01  1.865e+00   5.567 3.08e-08 ***
## Min_temperature          8.751e-01  4.538e-02  19.286  < 2e-16 ***
## Dewpoint                 3.737e-02  4.242e-02   0.881   0.3784
## Min_temperature:Dewpoint 2.021e-03  8.958e-04   2.256   0.0242 *
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Residual standard error: 4.025 on 1457 degrees of freedom
## Multiple R-squared:  0.9218, Adjusted R-squared:  0.9216
## F-statistic:  5722 on 3 and 1457 DF,  p-value: < 2.2e-16
```

```
adj.r.squared <- c(adj.r.squared, lm15=round(summary(lm15)$adj.r.squared, digits = 4))
```

```
lm16 <- lm(Mean_temperature~Min_temperature*Sea_level_pressure, data = wizmir)
summary(lm16)
```

```
##
## Call:
## lm(formula = Mean_temperature ~ Min_temperature * Sea_level_pressure,
##     data = wizmir)
##
## Residuals:
##     Min      1Q  Median      3Q     Max
## -8.9267 -3.0450 -0.3169  3.0384 17.0698
##
## Coefficients:
##                                    Estimate Std. Error t value Pr(>|t|)
```

```
## (Intercept)                         -271.40127   72.37093  -3.750 0.000184
## Min_temperature                        6.03703    1.54635   3.904 9.89e-05
## Sea_level_pressure                     9.32088    2.41219   3.864 0.000116
## Min_temperature:Sea_level_pressure    -0.16637    0.05167  -3.220 0.001310
##
## (Intercept)                         ***
## Min_temperature                     ***
## Sea_level_pressure                  ***
## Min_temperature:Sea_level_pressure **
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Residual standard error: 4.072 on 1457 degrees of freedom
## Multiple R-squared:  0.9199, Adjusted R-squared:  0.9198
## F-statistic:  5580 on 3 and 1457 DF,  p-value: < 2.2e-16
```

```
adj.r.squared <- c(adj.r.squared, lm16=round(summary(lm16)$adj.r.squared, digits = 4))
```

```
lm17 <- lm(Mean_temperature~Min_temperature*Standard_pressure, data = wizmir)
summary(lm17)
```

```
##
## Call:
## lm(formula = Mean_temperature ~ Min_temperature * Standard_pressure,
##     data = wizmir)
##
## Residuals:
##      Min       1Q   Median       3Q      Max
## -11.7784  -2.9398  -0.2942   3.0066  16.9358
##
## Coefficients:
##                                    Estimate Std. Error t value Pr(>|t|)
## (Intercept)                         8.90549    5.76084   1.546    0.122
## Min_temperature                     1.14620    0.13599   8.429   <2e-16 ***
## Standard_pressure                  -0.08059    0.79960  -0.101    0.920
## Min_temperature:Standard_pressure -0.01357    0.01876  -0.723    0.470
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Residual standard error: 4.071 on 1457 degrees of freedom
## Multiple R-squared:   0.92,  Adjusted R-squared:  0.9198
## F-statistic:  5583 on 3 and 1457 DF,  p-value: < 2.2e-16
```

```
adj.r.squared <- c(adj.r.squared, lm17=round(summary(lm17)$adj.r.squared, digits = 4))
```

```
lm18 <- lm(Mean_temperature~Dewpoint*Sea_level_pressure, data = wizmir)
summary(lm18)
```

```
##
## Call:
## lm(formula = Mean_temperature ~ Dewpoint * Sea_level_pressure,
##     data = wizmir)
```

```
##
## Residuals:
##      Min       1Q   Median       3Q      Max
## -26.1480  -5.6672  -0.9189   4.9531  28.5753
##
## Coefficients:
##                               Estimate Std. Error t value Pr(>|t|)
## (Intercept)                 -765.6889   182.2390  -4.202 2.81e-05 ***
## Dewpoint                      29.4409     3.9574   7.439 1.72e-13 ***
## Sea_level_pressure            25.8840     6.0596   4.272 2.07e-05 ***
## Dewpoint:Sea_level_pressure   -0.9461     0.1318  -7.177 1.13e-12 ***
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Residual standard error: 8.506 on 1457 degrees of freedom
## Multiple R-squared:  0.6506, Adjusted R-squared:  0.6499
## F-statistic: 904.4 on 3 and 1457 DF,  p-value: < 2.2e-16
```

```
adj.r.squared <- c(adj.r.squared, lm18=round(summary(lm18)$adj.r.squared, digits = 4))
```

```
lm19 <- lm(Mean_temperature~Dewpoint*Standard_pressure, data = wizmir)
summary(lm19)
```

```
##
## Call:
## lm(formula = Mean_temperature ~ Dewpoint * Standard_pressure,
##     data = wizmir)
##
## Residuals:
##     Min      1Q  Median      3Q     Max
## -17.363  -6.637  -1.163   5.172  29.642
##
## Coefficients:
##                              Estimate Std. Error t value Pr(>|t|)
## (Intercept)                -33.27015   16.79871  -1.981   0.0478 *
## Dewpoint                     1.70125    0.37167   4.577 5.11e-06 ***
## Standard_pressure            5.35797    2.30073   2.329   0.0200 *
## Dewpoint:Standard_pressure  -0.06879    0.05086  -1.352   0.1764
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Residual standard error: 8.78 on 1457 degrees of freedom
## Multiple R-squared:  0.6277, Adjusted R-squared:  0.627
## F-statistic:   819 on 3 and 1457 DF,  p-value: < 2.2e-16
```

```
adj.r.squared <- c(adj.r.squared, lm19=round(summary(lm19)$adj.r.squared, digits = 4))
```

```
lm20 <- lm(Mean_temperature~Sea_level_pressure*Standard_pressure, data = wizmir)
summary(lm20)
```

```
##
## Call:
```

```
## lm(formula = Mean_temperature ~ Sea_level_pressure * Standard_pressure,
##     data = wizmir)
##
## Residuals:
##     Min      1Q  Median      3Q     Max
## -46.468  -8.178   1.000   9.418  23.973
##
## Coefficients:
##                                     Estimate Std. Error t value Pr(>|t|)
## (Intercept)                          498.081    561.755   0.887   0.3754
## Sea_level_pressure                   -14.826     18.702  -0.793   0.4281
## Standard_pressure                    145.750     78.044   1.868   0.0620
## Sea_level_pressure:Standard_pressure   -4.827      2.599  -1.858   0.0634
##
## (Intercept)
## Sea_level_pressure
## Standard_pressure                    .
## Sea_level_pressure:Standard_pressure .
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Residual standard error: 11.67 on 1457 degrees of freedom
## Multiple R-squared:  0.3427, Adjusted R-squared:  0.3413
## F-statistic: 253.2 on 3 and 1457 DF,  p-value: < 2.2e-16
```

```
adj.r.squared <- c(adj.r.squared, lm20=round(summary(lm20)$adj.r.squared, digits = 4))
```

Tabla de resultados

```
createTable(adj.r.squared)
```

```
##           x porcentaje
## lm1  0.9576      95.76
## lm2   0.919      91.90
## lm3  0.6152      61.52
## lm4  0.3394      33.94
## lm5  0.0196       1.96
## lm6  0.9915      99.15
## lm7  0.9924      99.24
## lm8  0.9924      99.24
## lm9  0.9923      99.23
## lm10 0.9921      99.21
## lm11 0.9916      99.16
## lm12 0.9637      96.37
## lm13 0.9671      96.71
## lm14 0.9606      96.06
## lm15 0.9216      92.16
## lm16 0.9198      91.98
## lm17 0.9198      91.98
## lm18 0.6499      64.99
## lm19  0.627      62.70
## lm20 0.3413      34.13
```

Como se puede observar la mejor aproximación que se tiene al realizar las interacciones es la regresión lineal *lm13*, que interacciona *Max_temperature* con *Sea_level_pressure*. Aún así su valor es inferior a *lm6*, tal como se muestra en la siguiente tabla.