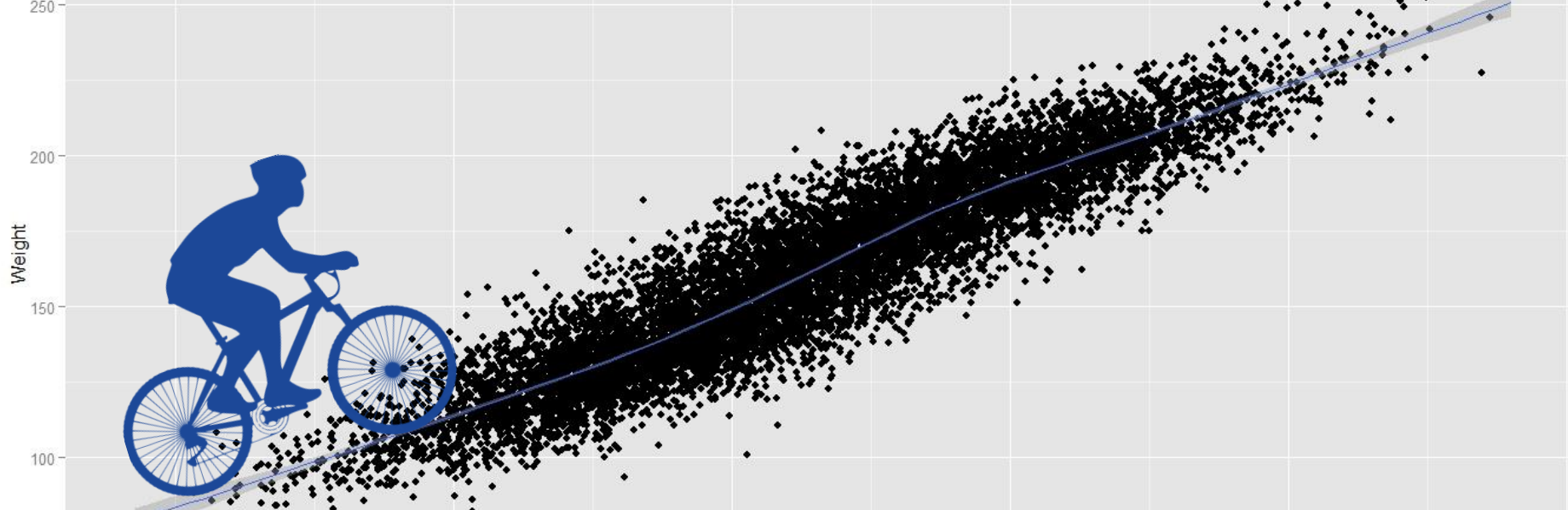




**Demanda del servicio  
público de renta de  
bicicletas en Seúl:**  
Interacción de la demanda  
con las condiciones del clima





## DATASET

Información de condiciones del clima durante un año en la ciudad de Seúl, Corea del Sur y la demanda diaria del servicio de renta de bicicletas:

dataset descargado de:

<https://www.kaggle.com/saurabhshahane/seoul-bike-sharing-demand-prediction>



# Análisis

```
12 str(RAWDATA) #1a
13 dim(RAWDATA) #e1
14 tail(RAWDATA)
15 summary(RAWDATA)
```

el dataset cuenta con 8760 observaciones en la muestra y 14 variables

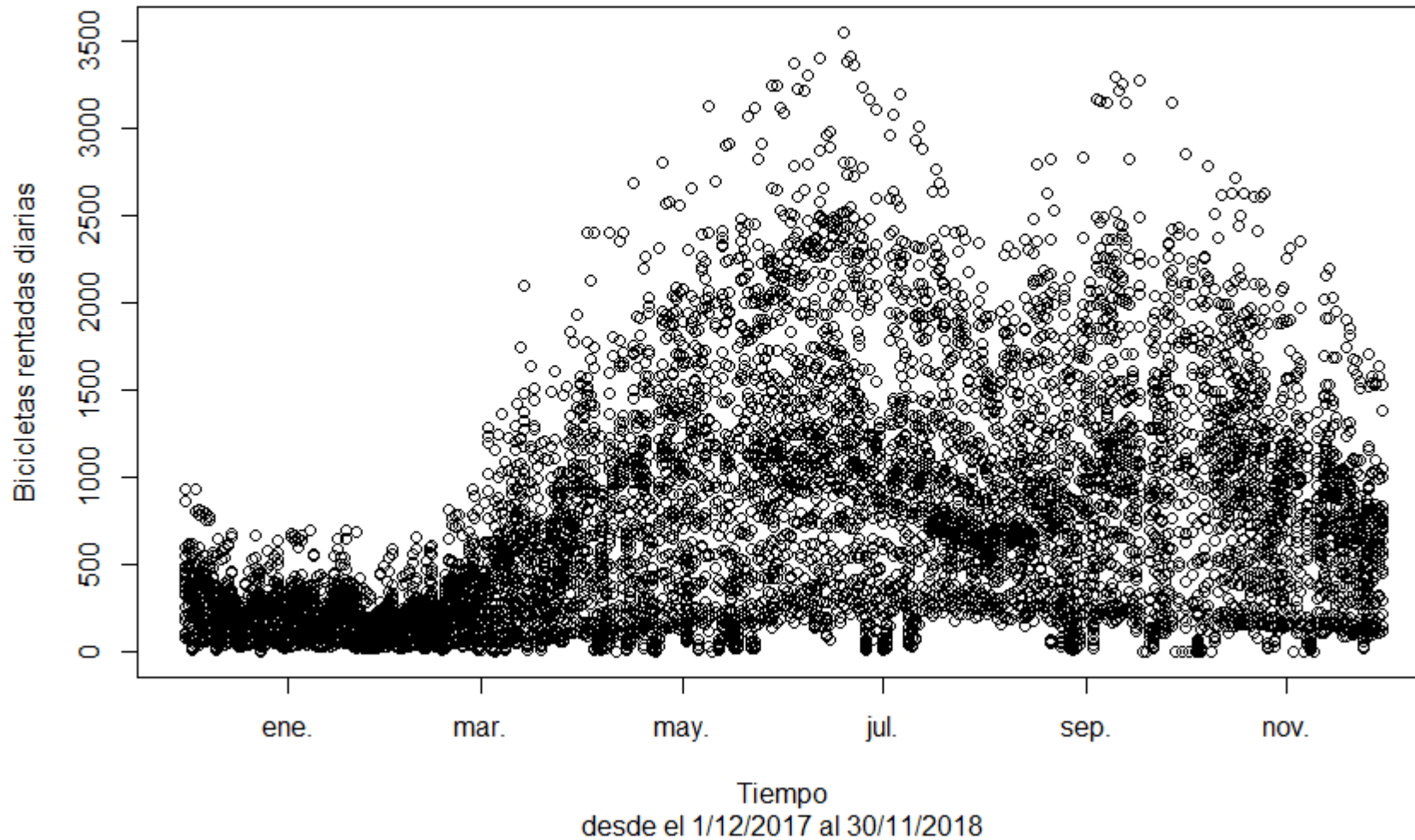
```
*data.frame*: 8760 obs. of 14 variables:
 $ Date           : chr  "01/12/2017" "01/12/2017" "01/12/2017" "01/12/2017" ...
 $ Rented.Bike.Count : int  254 204 173 107 78 100 181 460 930 490 ...
 $ Hour           : int   0 1 2 3 4 5 6 7 8 9 ...
 $ Temperature..C. : num  -5.2 -5.5 -6 -6.2 -6 -6.4 -6.6 -7.4 -7.6 -6.5 ...
 $ Humidity...     : int   37 38 39 40 36 37 35 38 37 27 ...
 $ Wind.speed..m.s. : num   2.2 0.8 1 0.9 2.3 1.5 1.3 0.9 1.1 0.5 ...
 $ Visibility..10m. : int  2000 2000 2000 2000 2000 2000 2000 2000 2000 1928 ...
 $ Dew.point.temperature..C.: num -17.6 -17.6 -17.7 -17.6 -18.6 -18.7 -19.5 -19.3 -19.8 -22.4 ..
 $ Solar.Radiation..MJ.m2. : num   0 0 0 0 0 0 0 0 0.01 0.23 ...
 $ Rainfall..mm.    : num   0 0 0 0 0 0 0 0 0 0 ...
 $ Snowfall..cm.    : num   0 0 0 0 0 0 0 0 0 0 ...
 $ Seasons          : chr   "winter" "winter" "winter" "winter" ...
 $ Holiday          : chr   "No Holiday" "No Holiday" "No Holiday" "No Holiday" ...
 $ Functioning.Day  : chr   "Yes" "Yes" "Yes" "Yes" ...
```

Se aplica 'mutate' para cambiar formato

```
20
21 DATA <- mutate(RAWDATA, Date = as.Date(Date, format = "%d/%m/%Y"))
22 str(DATA)
```

ahora el summary nos indica que este dataset se generó desde el 1 de dic del 2017 hasta el 30 de Noviembre del 2018

## Cantidad de bicicletas rentadas en el tiempo



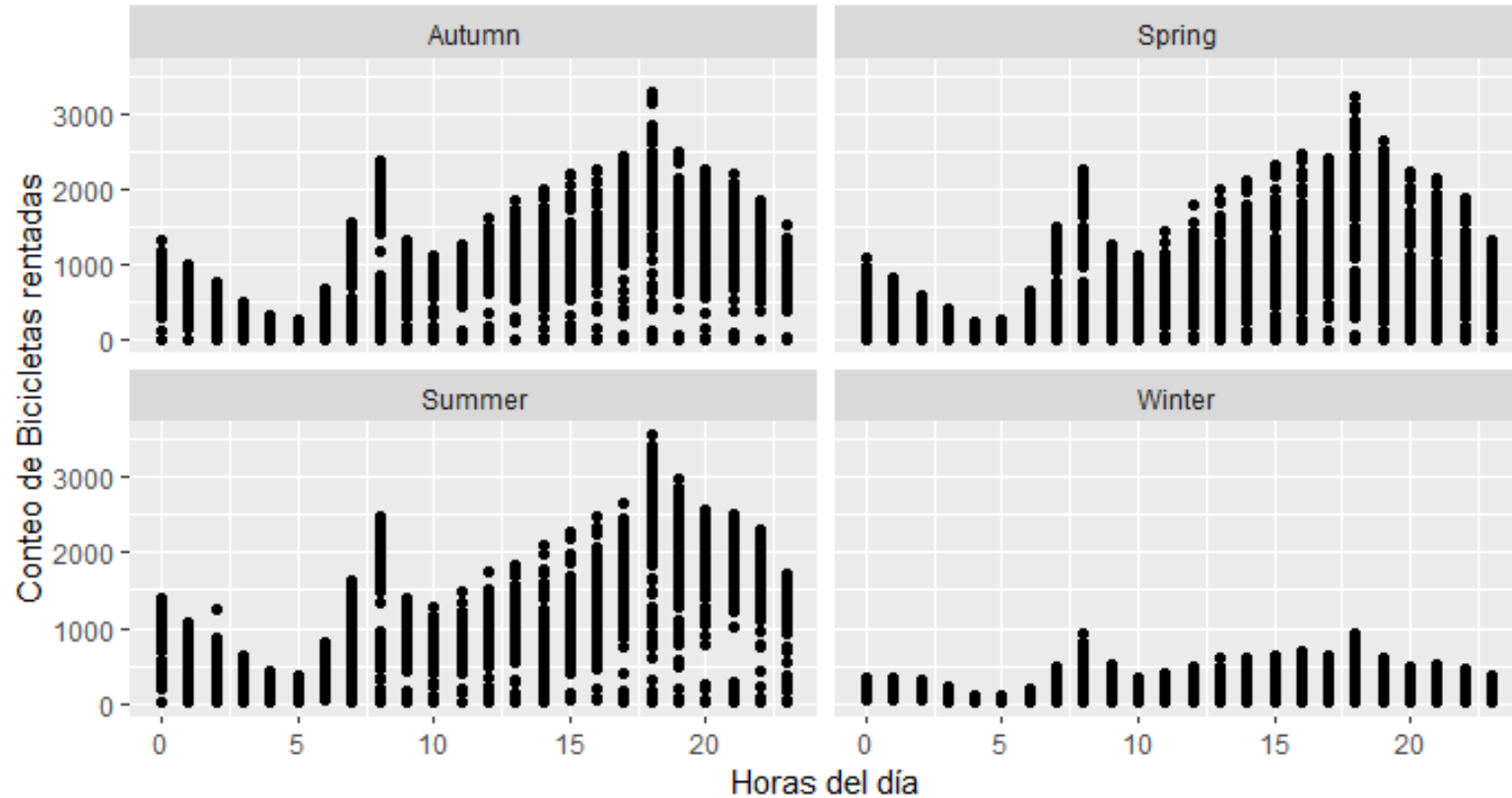
En esta gráfica está la información completa del año en usos del servicio, vemos que tiene cierto comportamiento estacional



```
25  
26 #a continuación podemos ver el desarrollo de rentas de bicicletas en el tiempo  
27 plot(x=DATA$Date, y = DATA$Rented.Bike.Count, ylab = "Bicicletas rentadas diarias",  
28      xlab = "Tiempo", main = "Cantidad de bicicletas rentadas en el tiempo",  
29      sub = "desde el 1/12/2017 al 30/11/2018")  
30
```

## Bicicletas rentadas en diferentes estaciones del año

renta de bicicletas por hora



fuelle: <https://www.kaggle.com/hardikjain10/seoul-bike-rented>

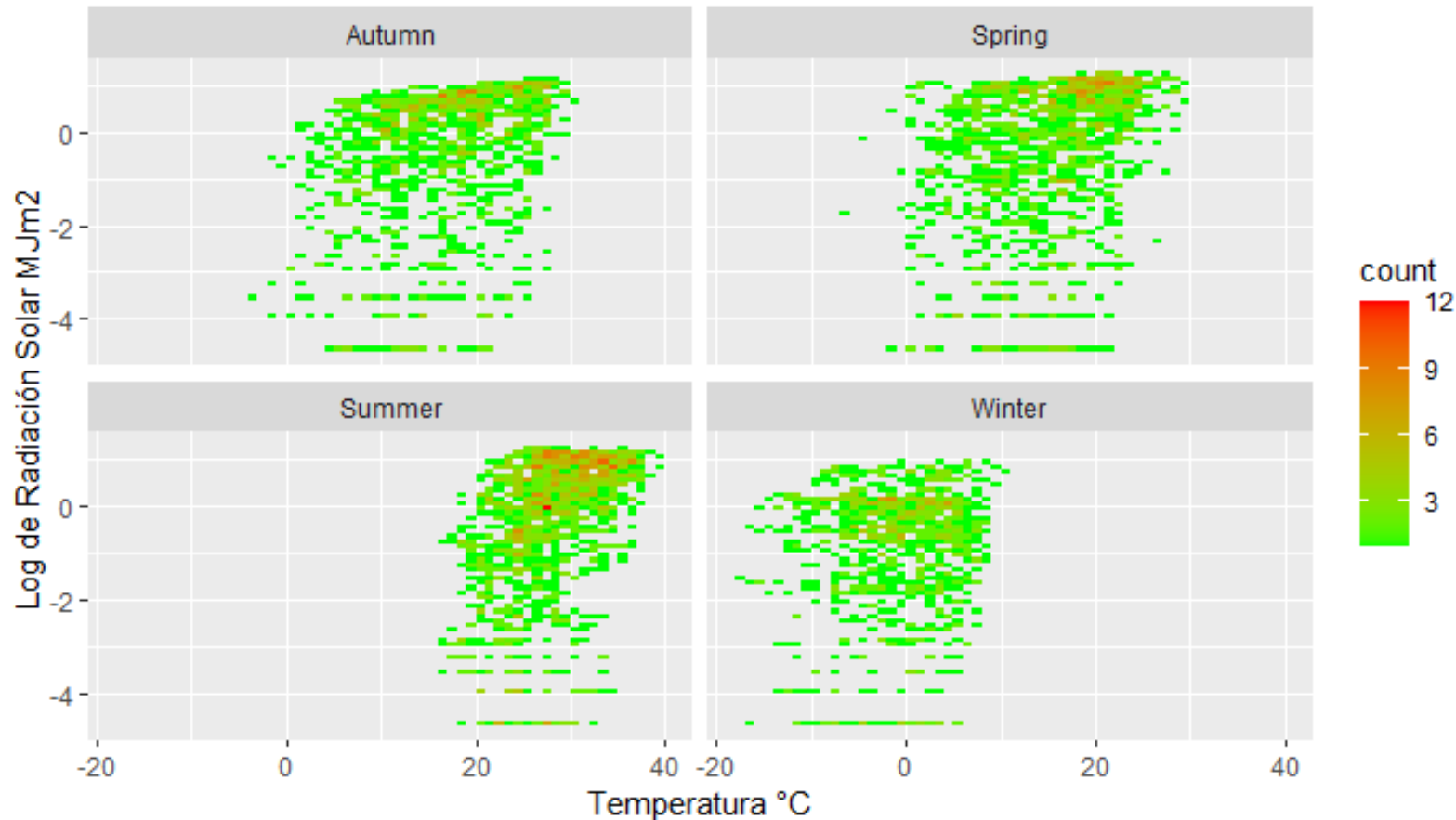
Se puede apreciar mejor la diferencia en la demanda de cada estación



```
37 ggplot(DATA, aes(x=Hour, y = Rented.Bike.Count)) +  
38   labs(x = "Horas del día", y = "Conteo de Bicicletas rentadas",  
39     title = "Bicicletas rentadas en diferentes estaciones del año",  
40     subtitle = "renta de bicicletas por hora",  
41     caption = "fuente: https://www.kaggle.com/hardikjain10/seoul-bike-rented",  
42     alt = "Add alt text to the plot") +  
43   geom_point() +  
44   theme_gray() +  
45   facet_wrap("Seasons")
```



## Temperatura y radiación solar en las estaciones del año



fuelle: <https://www.kaggle.com/hardikjain10/seoul-bike-rented>

Para verificar la calidad de la información, comparando la radiación solar y la temperatura diaria en las diferentes estaciones anuales para demostrar que los datos son coherentes

Como es de esperarse, las temperaturas más altas se encuentran en verano y las más bajas en invierno, otoño y primavera tienen temperaturas similares

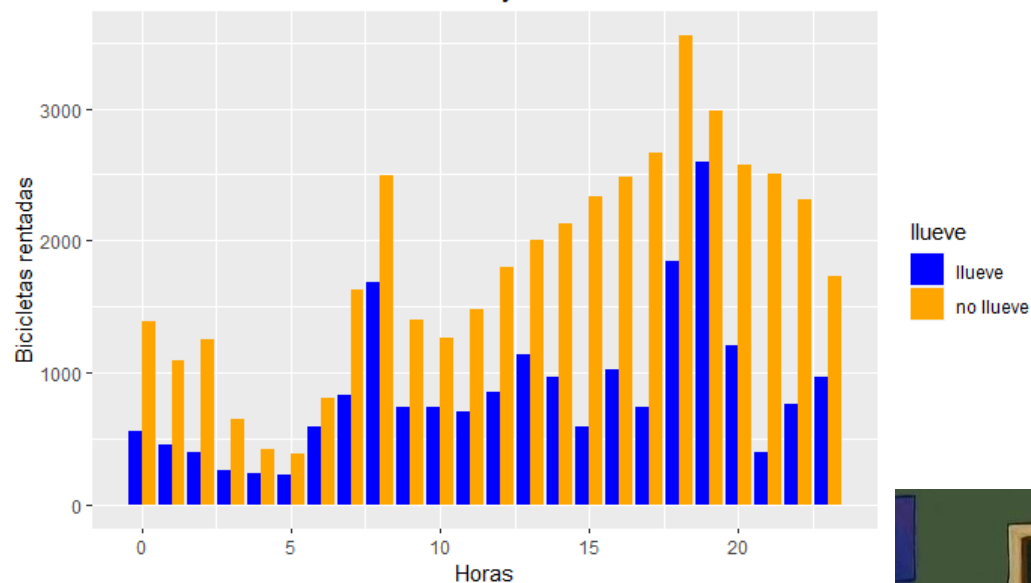
también podemos observar cierta relación de la variable  $\log(\text{Solar.Radiation..MJ.m2.})$  con la temperatura

```
49 ggplot(DATA, aes(x=Temperature..C., y = log(Solar.Radiation..MJ.m2.))) + #se agregó
50   labs(x = "Temperatura °C", y = "Log de Radiación Solar MJm2",
51         title = "Temperatura y radiación solar en las estaciones del año",
52         caption = "fuente: https://www.kaggle.com/hardikjain10/seoul-bike-rented",
53         alt = "Add alt text to the plot") +
54   geom_bin2d(binwidth = c(1, 0.1)) +
55   scale_fill_gradient(low="green", high = "red") +
56   theme_gray() +
57   facet_wrap("Seasons")
--
```

# Se guardan nuevas variables discretas para seguir analizando a través de gráficos de ggplot

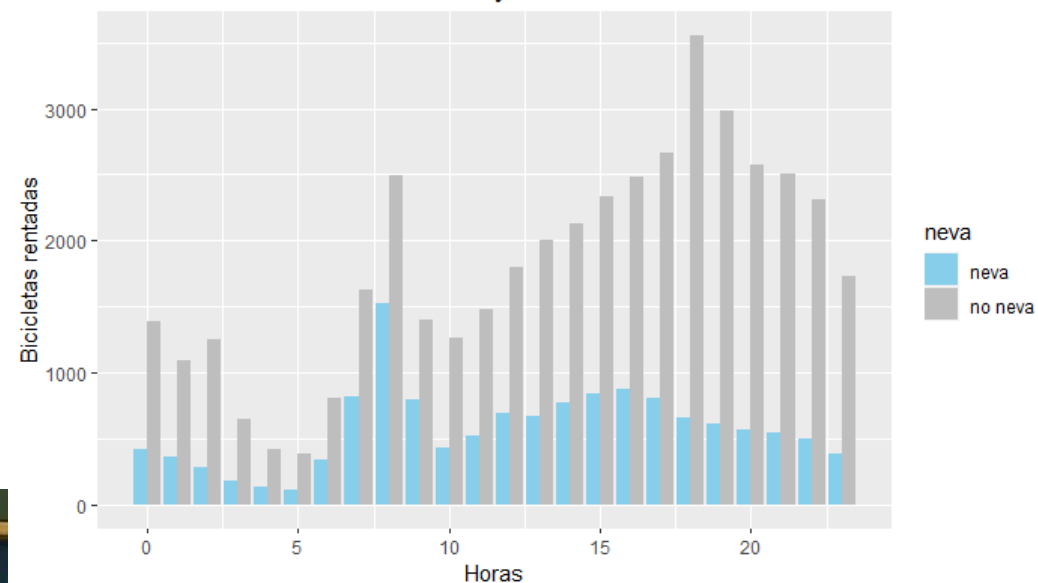
```
for(i in 1:nrow(DATA) ) {  
  DATA1$llueve[i] <- if(DATA$Rainfall.mm.[i]>0){"llueve"}else{"no llueve"} #si llueve = 1, si no, 0  
  DATA1$neva[i] <- if(DATA$ Snowfall..cm.[i]>0){"neva"}else {"no neva"}#si neva = 1, si no, 0  
  DATA1$radiación[i] <- if(DATA$ Solar.Radiation..MJ.m2.[i]>0.3){"Alta radiación"}else{"Baja  
radiación"} #si hay radiación solar = 1, si no, 0  
  DATA1$temp[i] <- if(DATA$ Temperature..C.[i]>20){"Mayor a 20°C"}else{if(DATA$  
Temperature..C.[i]>5){"De 5°C a 20°C"}else{"Menor a 5°C"}}  
}  
DATA1 <- mutate(DATA1, llueve = as.factor(llueve))  
DATA1 <- mutate(DATA1, neva = as.factor(neva))  
DATA1 <- mutate(DATA1, radiación = as.factor(radiación))  
DATA1 <- mutate(DATA1, temp = as.factor(temp))
```

Bicicletas rentadas cuando llueve y cuando no llueve



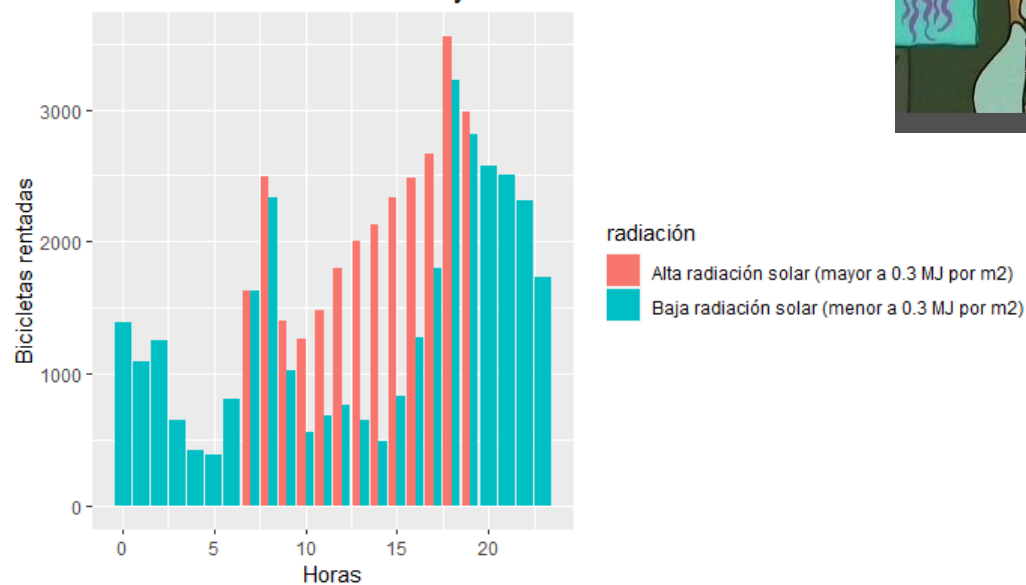
fuelle: <https://www.kaggle.com/hardikjain10/seoul-bike-rented>

Bicicletas rentadas cuando neva y cuando no neva



fuelle: <https://www.kaggle.com/hardikjain10/seoul-bike-rented>

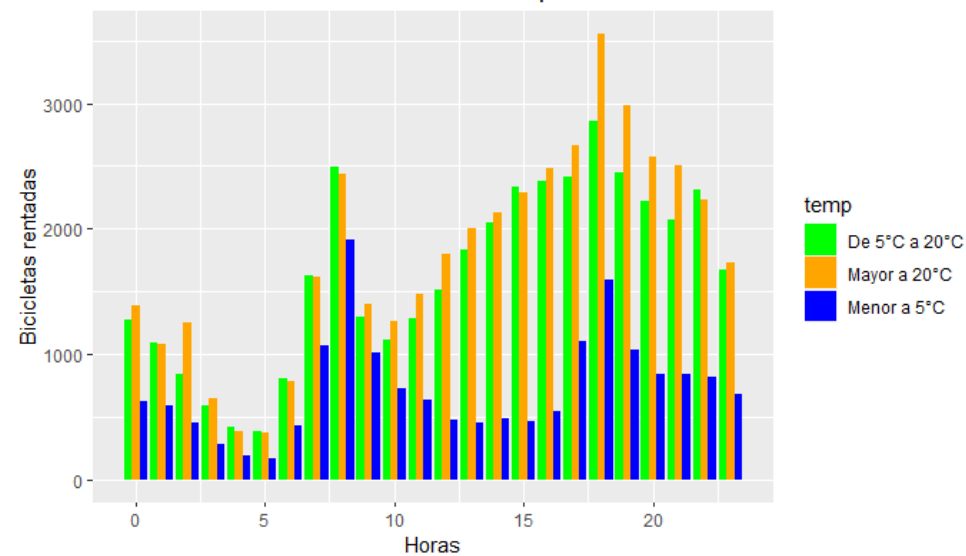
Bicicletas rentadas cuando hay o no radiación solar



fuelle: <https://www.kaggle.com/hardikjain10/seoul-bike-rented>



Bicicletas rentadas en relación a la temperatura del día

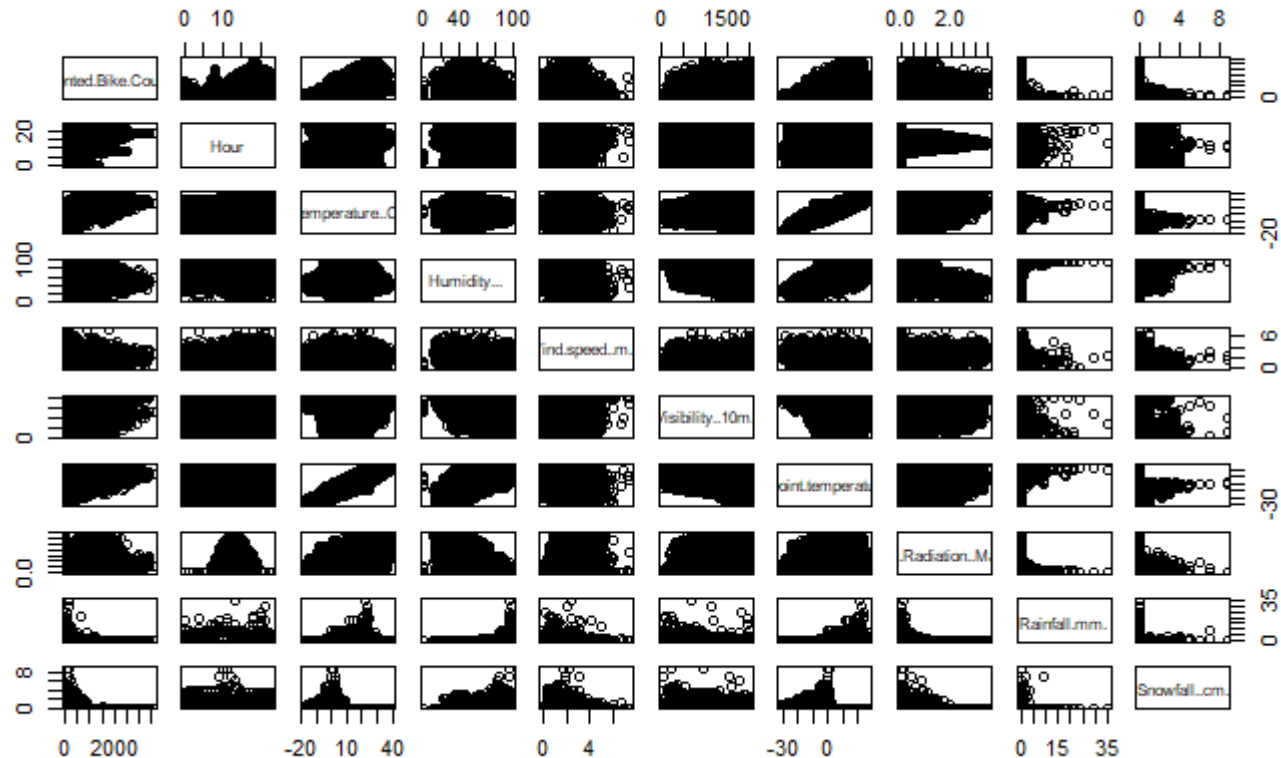


fuelle: <https://www.kaggle.com/hardikjain10/seoul-bike-rented>



# Realizamos un análisis de correlaciones

A simple vista no se puede apreciar ninguna variable con una correlación visible, salvo por Temperature..C. y Dew.point.temperature..C. (punto de condensación)



```
DATAnumeric <- select(DATA, -Date, -Seasons, -Holiday, -Functioning.Day)
attach(DATAnumeric)
pairs(DATAnumeric) # por lo visto no se puede apreciar ninguna variable co
CORTable <- cor(DATAnumeric)
view(CORTable)
```

# Prueba simple de hipótesis:

diferencia de medias para muestras grandes

$$H_0 : X_1 = X_2$$

$$H_1 : X_1 \neq X_2$$

La hipótesis nula ( $H_0$ ) se formula:

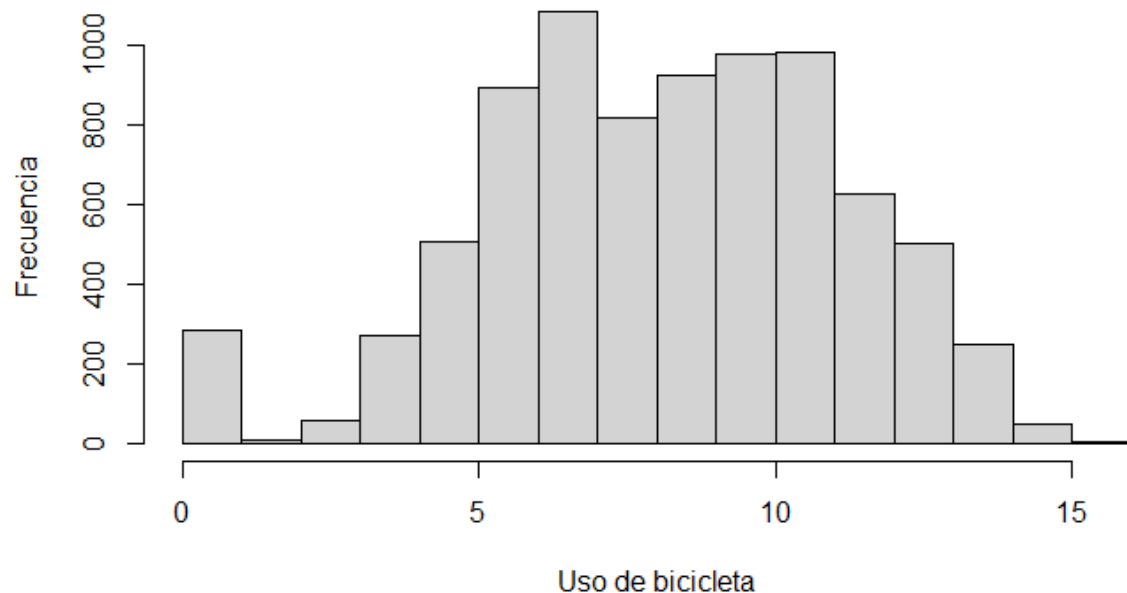
No existe diferencia en la demanda de bicicletas en días normales y días lluviosos

La hipótesis alternativa ( $H_1$ ) dice:

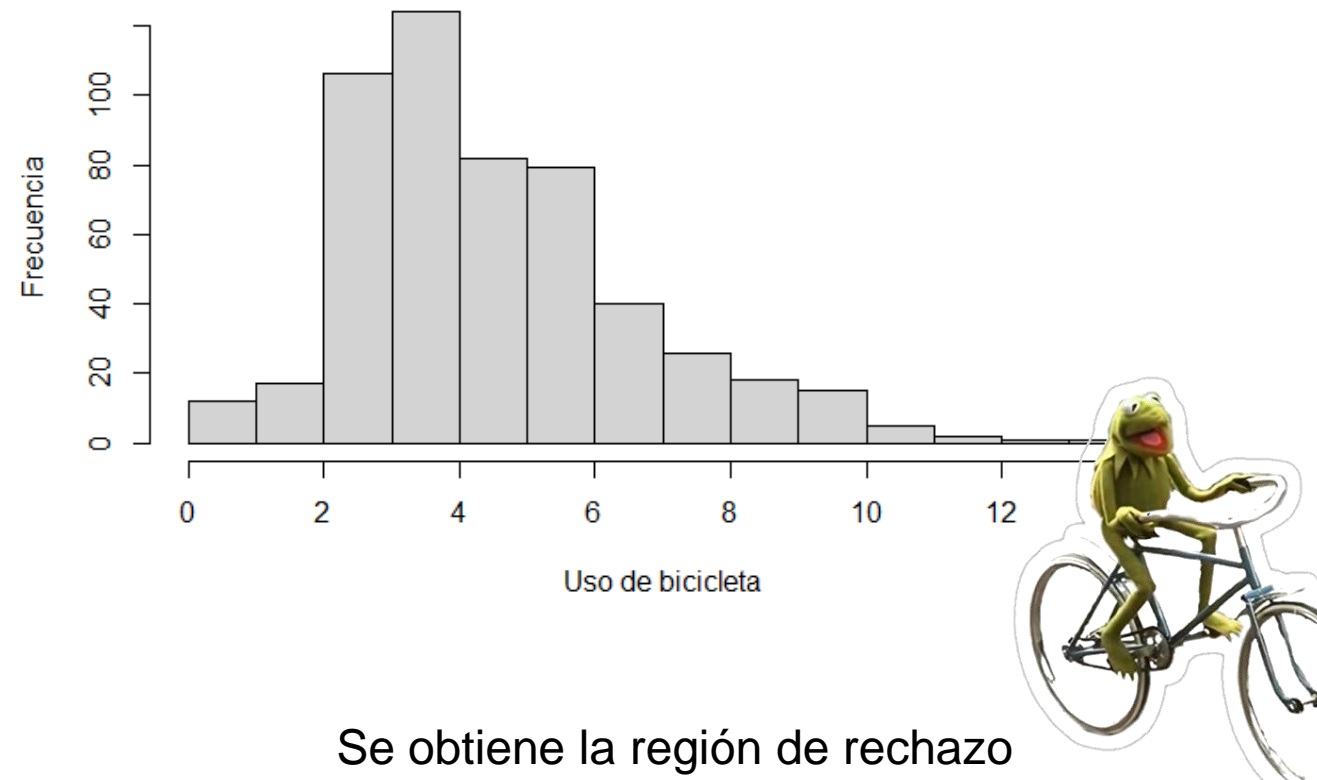
Existe diferencia en las medias en la demanda del servicio en días lluviosos y días regulares



Histograma de uso de bicicleta<sup>1/3</sup> en días normales



Histograma de uso de bicicleta<sup>1/3</sup> en días lluviosos



Se establecen las variables y el estimador z

```
160 #se establecen las muestras para la prueba de hipótesis:
161 m1 <- DATA1nollueve$Rented.Bike.Count
162 n1 <- length(DATA1nollueve$Rented.Bike.Count)
163 m2 <- DATA1sillueve$Rented.Bike.Count
164 n2 <- length(DATA1sillueve$Rented.Bike.Count)
165
166 #obtenemos la función del estadístico z
167 z0 <- (mean(m1)-mean(m2))/sqrt(var(m1)/n1 + var(m2)/n2)
168 z0
```

Se obtiene la región de rechazo

```
170 #Se obtiene la región de rechazo
171 (z.025 <- qnorm(p = 0.025, lower.tail = FALSE))
172 z.025
173
174 (z0 < -z.025) | (z0 > z.025)
175 ~~~
```







```
> (z0 < -z.025) | (z0 > z.025)  
[1] TRUE
```

La prueba nos hace Rechazar la hipótesis nula( $H_0$ ) y no tener elementos para rechazar la hipótesis alternativa( $H_1$ )

Existe diferencia clara entre las dos muestras, por lo tanto se infiere que la demanda de renta de bicicletas es mayor en días soleados que en días lluviosos

# Regresión lineal

Se realizaron 4 regresiones lineales para proponer un modelo de predicción

```
Call:
lm(formula = Rented.Bike.Count ~ Hour + Temperature..C. + Humidity... +
  Wind.speed..m.s. + Visibility..10m. + Dew.point.temperature..C. +
  Solar.Radiation..MJ.m2. + Rainfall.mm. + Snowfall..cm. +
  Seasons + Holiday)
```

```
Residuals:
    Min       1Q   Median       3Q      Max
-1426.72  -276.33   -42.17   220.85  2279.61
```

```
Coefficients:
              Estimate Std. Error t value Pr(>|t|)
(Intercept)  7.013e+02  1.025e+02   6.840 8.42e-12 ***
Hour         2.811e+01  7.842e-01  35.841 < 2e-16 ***
Temperature..C. 1.766e+01  3.909e+00   4.517 6.36e-06 ***
Humidity...   -1.017e+01  1.099e+00  -9.251 < 2e-16 ***
Wind.speed..m.s. 1.333e+01  5.437e+00   2.452 0.014220 *
Visibility..10m.  5.245e-03  1.055e-02   0.497 0.618982
Dew.point.temperature..C. 7.233e+00  4.092e+00   1.768 0.077173 .
Solar.Radiation..MJ.m2. -7.741e+01  8.106e+00  -9.549 < 2e-16 ***
Rainfall.mm.   -5.850e+01  4.559e+00 -12.833 < 2e-16 ***
Snowfall..cm.   3.891e+01  1.196e+01   3.253 0.001148 **
SeasonsSpring   -5.293e+01  1.460e+01  -3.625 0.000291 ***
SeasonsSummer   -2.020e+01  1.791e+01  -1.128 0.259394
SeasonsWinter   -2.973e+02  2.094e+01 -14.200 < 2e-16 ***
HolidayNo Holiday  1.327e+02  2.306e+01   5.754 9.03e-09 ***
```

```
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
```

```
Residual standard error: 462.1 on 8746 degrees of freedom
Multiple R-squared:  0.4876,    Adjusted R-squared:  0.4868
F-statistic: 640.1 on 13 and 8746 DF, p-value: < 2.2e-16
```

```
Call:
lm(formula = Rented.Bike.Count ~ Hour + Temperature..C. + Humidity... +
  Solar.Radiation..MJ.m2. + Rainfall.mm. + Snowfall..cm. +
  Holiday)
```

```
Residuals:
    Min       1Q   Median       3Q      Max
-1404.01  -287.82   -40.25   220.30  2291.73
```

```
Coefficients:
              Estimate Std. Error t value Pr(>|t|)
(Intercept)  351.7266    31.2532  11.254 < 2e-16 ***
Hour         27.3493     0.7601  35.981 < 2e-16 ***
Temperature..C. 31.7196     0.5127  61.871 < 2e-16 ***
Humidity...   -8.0453     0.3250 -24.757 < 2e-16 ***
Solar.Radiation..MJ.m2. -83.8085     7.5052 -11.167 < 2e-16 ***
Rainfall.mm.  -59.6807     4.5824 -13.024 < 2e-16 ***
Snowfall..cm.   18.2082    11.9678   1.521  0.128
HolidayNo Holiday 161.3258    23.1762   6.961 3.63e-12 ***
```

```
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
```

```
Residual standard error: 468.4 on 8752 degrees of freedom
Multiple R-squared:  0.473,    Adjusted R-squared:  0.4726
F-statistic: 1122 on 7 and 8752 DF, p-value: < 2.2e-16
```

# Regresión lineal

Se realizaron 4 regresiones lineales para proponer un modelo de predicción

```
Call:
lm(formula = Rented.Bike.Count ~ Hour + Temperature..C. + Humidity... +
  Wind.speed..m.s. + Visibility..10m. + Dew.point.temperature..C. +
  Solar.Radiation..MJ.m2. + Rainfall.mm. + Snowfall..cm. +
  Seasons + Holiday)
```

```
Residuals:
    Min       1Q   Median       3Q      Max
-10.5803  -0.9544   0.1906   1.3402  11.8094
```

```
Coefficients:
              Estimate Std. Error t value Pr(>|t|)
(Intercept)   9.340e+00  5.154e-01  18.123 < 2e-16 ***
Hour          1.186e-01  3.942e-03  30.089 < 2e-16 ***
Temperature..C.  8.903e-03  1.965e-02   0.453  0.6505
Humidity...    -6.343e-02  5.525e-03 -11.481 < 2e-16 ***
Wind.speed..m.s. -4.161e-02  2.733e-02  -1.522  0.1279
Visibility..10m. -6.109e-05  5.302e-05  -1.152  0.2492
Dew.point.temperature..C.  8.533e-02  2.057e-02   4.148 3.38e-05 ***
Solar.Radiation..MJ.m2. -9.976e-02  4.075e-02  -2.448  0.0144 *
Rainfall.mm.    -3.998e-01  2.292e-02 -17.447 < 2e-16 ***
Snowfall..cm.   8.424e-02  6.013e-02   1.401  0.1612
SeasonsSpring   6.286e-02  7.341e-02   0.856  0.3918
SeasonsSummer   5.837e-01  9.003e-02   6.484 9.44e-11 ***
SeasonsWinter  -1.296e+00  1.052e-01 -12.310 < 2e-16 ***
HolidayNo Holiday  8.621e-01  1.159e-01   7.437 1.13e-13 ***
```

---  
Signif. codes: 0 '\*\*\*' 0.001 '\*\*' 0.01 '\*' 0.05 '.' 0.1 ' ' 1

Residual standard error: 2.323 on 8746 degrees of freedom  
Multiple R-squared: 0.4406, Adjusted R-squared: 0.4398  
F-statistic: 530 on 13 and 8746 DF, p-value: < 2.2e-16

```
Call:
lm(formula = Rented.Bike.Count ~ Hour + Humidity... + Dew.point.temperature..C. +
  Solar.Radiation..MJ.m2. + Rainfall.mm. + Seasons + Holiday)
```

```
Residuals:
    Min       1Q   Median       3Q      Max
-10.5914  -0.9521   0.1960   1.3325  11.7625
```

```
Coefficients:
              Estimate Std. Error t value Pr(>|t|)
(Intercept)   9.264374  0.179636  51.573 < 2e-16 ***
Hour          0.117873  0.003801  31.010 < 2e-16 ***
Humidity...   -0.063481  0.002109 -30.096 < 2e-16 ***
Dew.point.temperature..C.  0.093974  0.004828  19.466 < 2e-16 ***
Solar.Radiation..MJ.m2. -0.092578  0.036670  -2.525  0.0116 *
Rainfall.mm.   -0.399926  0.022741 -17.586 < 2e-16 ***
SeasonsSpring   0.062808  0.070798   0.887  0.3750
SeasonsSummer   0.572544  0.089601   6.390 1.74e-10 ***
SeasonsWinter  -1.282298  0.102117 -12.557 < 2e-16 ***
HolidayNo Holiday  0.871464  0.115852   7.522 5.92e-14 ***
```

---  
Signif. codes: 0 '\*\*\*' 0.001 '\*\*' 0.01 '\*' 0.05 '.' 0.1 ' ' 1

Residual standard error: 2.323 on 8750 degrees of freedom  
Multiple R-squared: 0.4402, Adjusted R-squared: 0.4397  
F-statistic: 764.6 on 9 and 8750 DF, p-value: < 2.2e-16



- En los 4 modelos el coeficiente de determinación arrojó un valor menor del 0.5, por lo que este modelo solo explica la mitad de la variabilidad de los datos de respuesta en torno a su media, a pesar de tener *pvalues* aceptables

```
[1,] [,1] [,2] [,3]
[1,] "Regresión 1" "Regresión 2" "Regresión 3"
[2,] "dependiente original, completa" "dependiente original, reducida" "dependiente^(1/3), completa"
[3,] "0.487558907479703" "0.472987192690839" "0.440646398749741"

[1,] [,4]
[1,] "Regresión 4"
[2,] "dependiente^(1/3), reducida"
[3,] "0.440240870503299"
```

