

CasoHuracanes-Equipo3.R

r2623423

2024-05-10

```
# TC1002S.201 - Equipo 3
# Rero: Caso Huracanes
# Jesús Ángel Guzmán Ortega A01799257
# Julio Cesar Vivas Medina A01749879

# Instalación de librerías -----

# Librerías
library(dplyr)

##
## Attaching package: 'dplyr'
## The following objects are masked from 'package:stats':
##
##   filter, lag
## The following objects are masked from 'package:base':
##
##   intersect, setdiff, setequal, union

library(ggplot2)
library(corrplot)

## corrplot 0.92 loaded

library(factoextra)

## Welcome! Want to learn more? See two factoextra-related books at https://goo.gl/ve3WBa

#install.packages("ggplot2")
#install.packages("ggmap")
#install.packages("maps")
#install.packages("mapdata")

# Cargamos los datos
#setwd("/cloud/project/retoAnalitica/codigo")
huracanes <- read.csv("CasoHuracanesCSV.csv")

# Filtrar los datos del Océano Pacífico Norte
huracanes_pacifico <- huracanes[huracanes$Ocean == "Pacific", ]

# 3.3.1 Eliminar los datos que no son reales (-999 en Wind y Pressure)
```

```

huracanes_pacifico <- huracanes_pacifico[!(huracanes_pacifico$Wind <= 0 | huracanes_pacifico$Pressure <= 0)]

# 3. Exploración de los datos -----
# 3.1. Identificar el tipo de dato de tres columnas diferentes
Name_class <- class(huracanes$Name)
Name_class

## [1] "character"

Year_class <- class(huracanes$year)
Year_class

## [1] "integer"

Co2_class <- class(huracanes$CO2)
Co2_class

## [1] "numeric"

# 3.2. Valores de los cuartiles de las variables : WIND, PRESSURE, CO2
cuartiles_wind <- quantile(huracanes_pacifico$Wind, probs = c(0.25, 0.5, 0.75))
cuartiles_wind

## 25% 50% 75%
## 30 45 70

cuartiles_pressure <- quantile(huracanes_pacifico$Pressure, probs = c(0.25, 0.5, 0.75))
cuartiles_pressure

## 25% 50% 75%
## 982 999 1007

cuartiles_co2 <- quantile(huracanes_pacifico$CO2, probs = c(0.25, 0.5, 0.75))
cuartiles_co2

## 25% 50% 75%
## 22154.34 22181.81 22429.38

# Preparación de datos -----
# Convertir la columna "Fecha" al tipo de dato de fecha
huracanes_pacifico$Fecha <- as.Date(huracanes_pacifico$Fecha)

# 3.3.1 Eliminar la columna "Population" ya que todos los valores son 0
huracanes_pacifico <- huracanes_pacifico[, -grep("Population", names(huracanes_pacifico))]
huracanes_pacifico <- huracanes_pacifico[, !(names(huracanes_pacifico) == "Name")]

# 3.3.2 Eliminar los posibles NA
huracanes_pacifico <- na.omit(huracanes_pacifico)

# Filtrar huracanes_pacifico para incluir solo los registros donde el estatus sea "HU"
huracanes_pacifico_hu <- filter(huracanes_pacifico, Status == "HU")

# Realizar técnicas estadísticas -----

# 4.1 Muestra la MEDIA y el PROMEDIO de las variables : WIND y PRESSURE, CO2
media_wind_hu <- mean(huracanes_pacifico_hu$Wind)
media_wind_hu

```

```
## [1] 88.57887
media_pressure_hu <- mean(huracanes_pacifico_hu$Pressure)
media_pressure_hu

## [1] 968.8112
media_co2_hu <- mean(huracanes_pacifico_hu$CO2)
media_co2_hu

## [1] 21884.01
# 4.2 Indica los valores de los cuartiles de las variables : WIND, PRESSURE, CO2
cuartiles_wind_hu <- quantile(huracanes_pacifico_hu$Wind, probs = c(0.25, 0.5, 0.75))
cuartiles_wind_hu

## 25% 50% 75%
## 70 85 100
cuartiles_pressure_hu <- quantile(huracanes_pacifico_hu$Pressure, probs = c(0.25, 0.5, 0.75))
cuartiles_pressure_hu

## 25% 50% 75%
## 959 972 981
cuartiles_co2_hu <- quantile(huracanes_pacifico_hu$CO2, probs = c(0.25, 0.5, 0.75))
cuartiles_co2_hu

##      25%      50%      75%
## 22153.14 22181.81 22429.38

# 4.3 Emite conclusiones
# Revisión de datos
str(huracanes_pacifico)

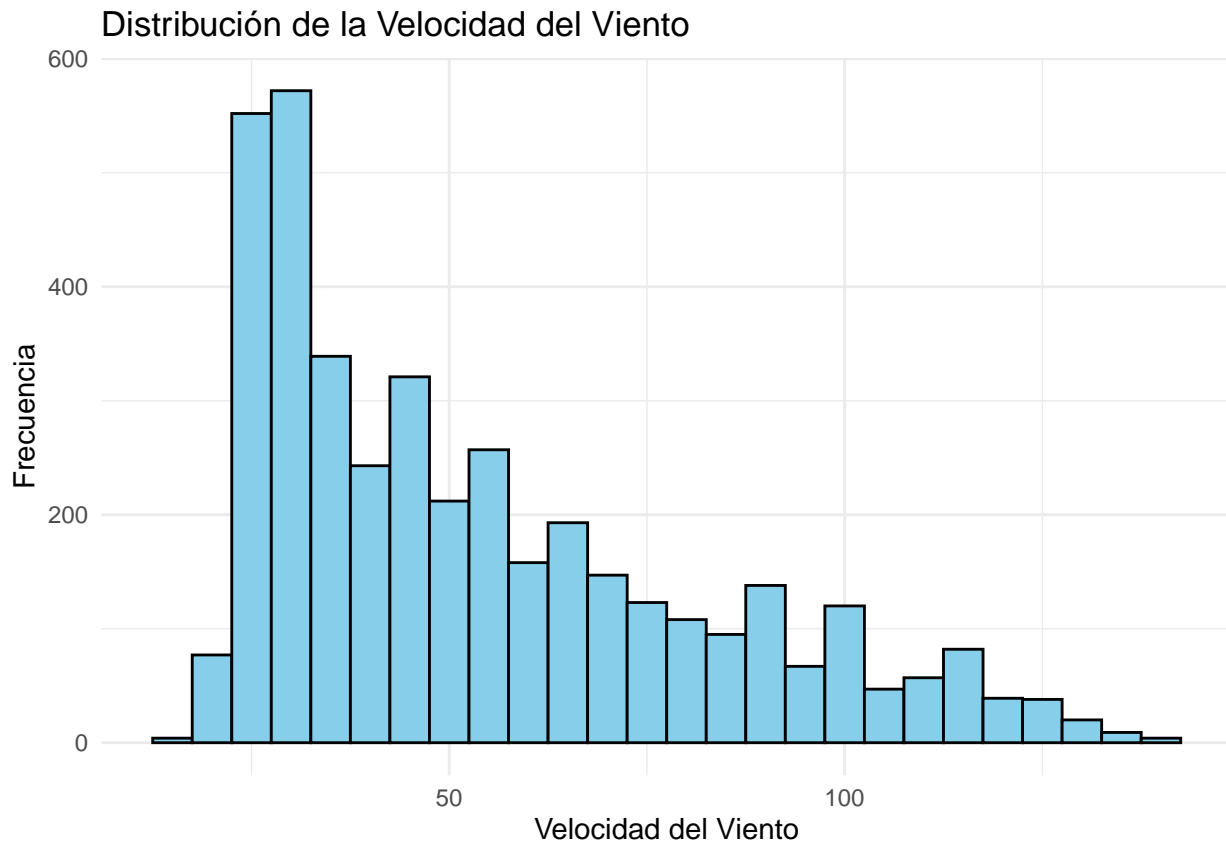
## 'data.frame': 4022 obs. of 13 variables:
## $ Clave : int 49610 49937 49945 49957 50170 50425 50533 50575 50630 52010 ...
## $ ID : chr "EP071954" "EP021957" "EP021957" "EP021957" ...
## $ Ocean : chr "Pacific" "Pacific" "Pacific" "Pacific" ...
## $ Fecha : Date, format: "19-09-19" "8-08-19" ...
## $ year : int 1954 1957 1957 1957 1958 1959 1959 1959 1960 1967 ...
## $ Month : chr "September" "August" "August" "August" ...
## $ Time : int 0 0 0 1800 0 1800 1200 1800 1200 ...
## $ Status : chr " TS" " TS" " TS" " TS" ...
## $ Latitude: num 21.1 21.1 28 34.8 13.1 15.3 22.1 19.1 20.2 19.3 ...
## $ Longitud: num -116 -136 -148 -162 -102 -145 -122 -104 -107 166 ...
## $ Wind : int 45 60 45 55 45 130 70 120 50 130 ...
## $ Pressure: int 981 994 1004 987 992 952 967 958 987 933 ...
## $ CO2 : num 6629 7965 7965 7965 8215 ...

# View(huracanes_pacifico)

#5.- Análisis por gráficas

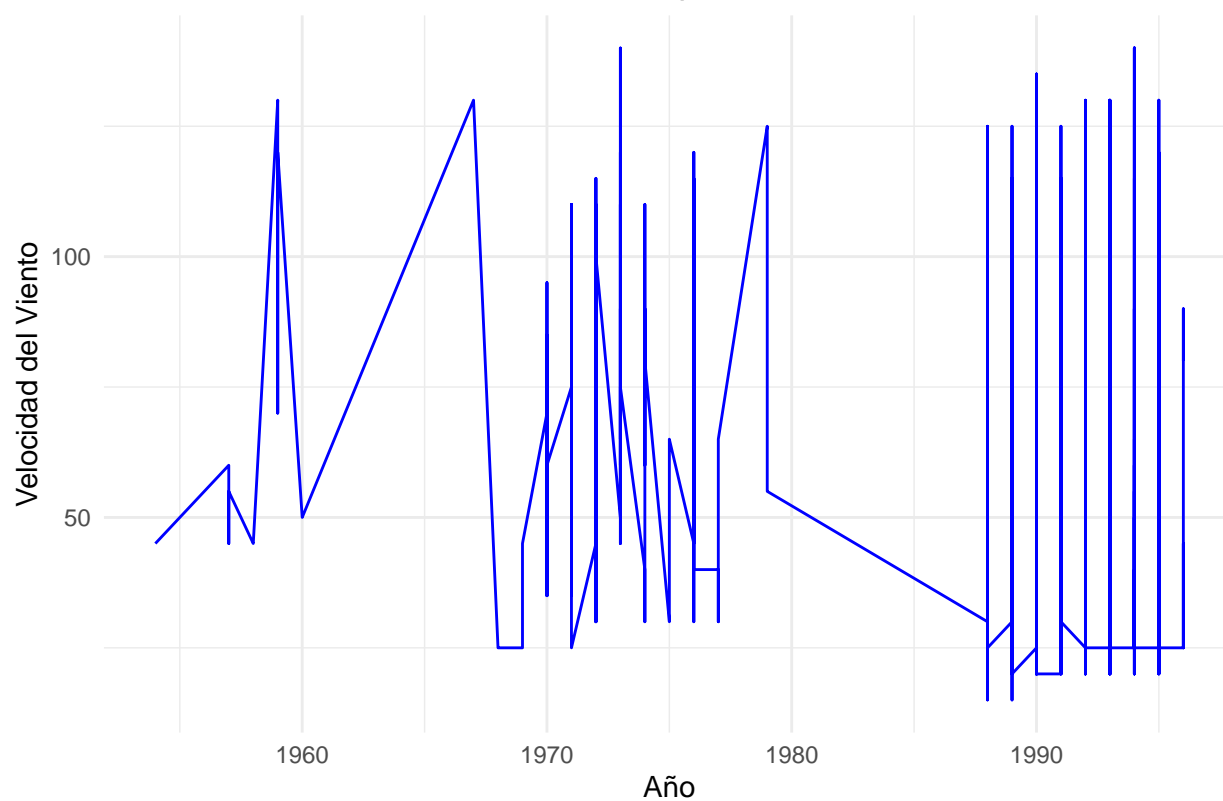
# 5.1.1 Gráfica la variable WIND
ggplot(huracanes_pacifico, aes(x = Wind)) +
  geom_histogram(binwidth = 5, fill = "skyblue", color = "black") +
  labs(title = "Distribución de la Velocidad del Viento",
       x = "Velocidad del Viento",
```

```
y = "Frecuencia") +  
theme_minimal()
```

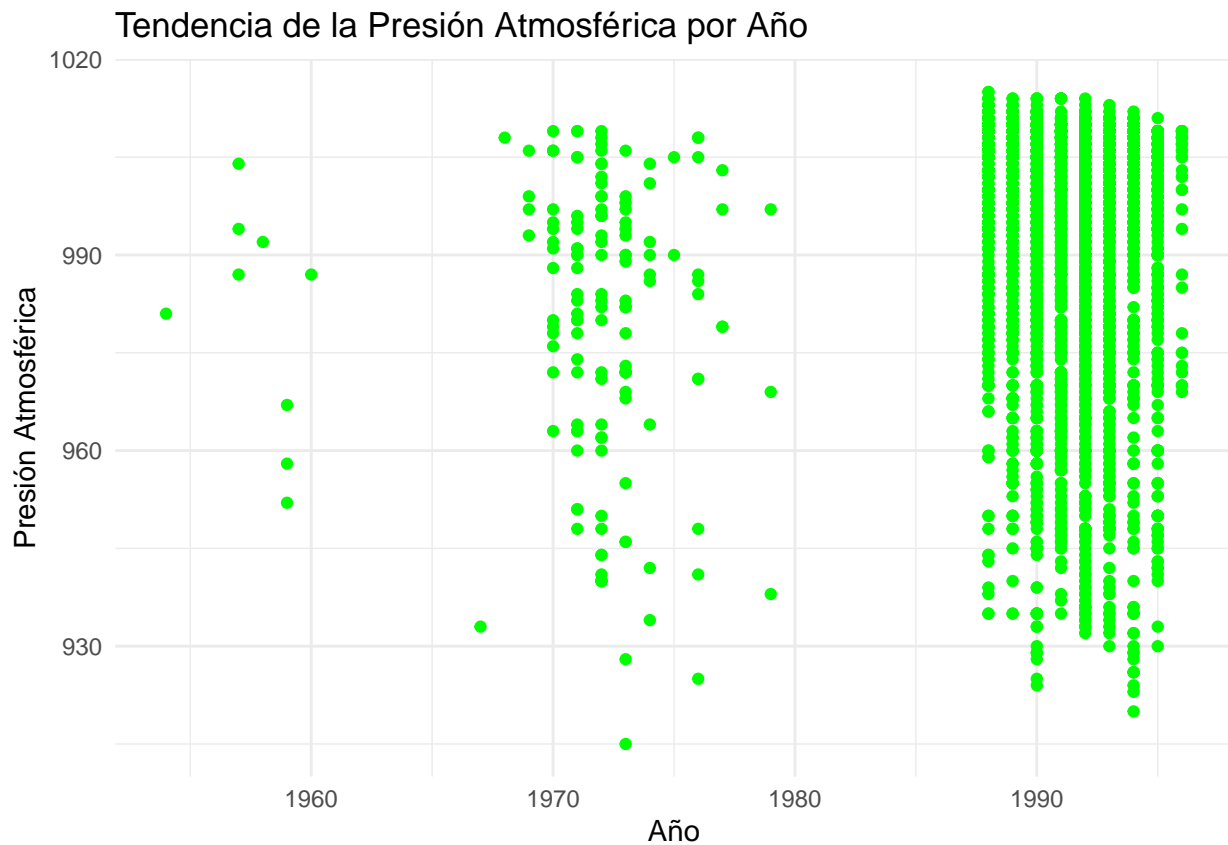


```
# 5.1.2 Gráfica años con la variable WIND  
ggplot(huracanes_pacifico, aes(x = year, y = Wind)) +  
  geom_line(color = "blue") + # Utiliza líneas  
  labs(title = "Tendencia de la Velocidad del Viento por Año",  
        x = "Año",  
        y = "Velocidad del Viento") +  
  theme_minimal()
```

Tendencia de la Velocidad del Viento por Año



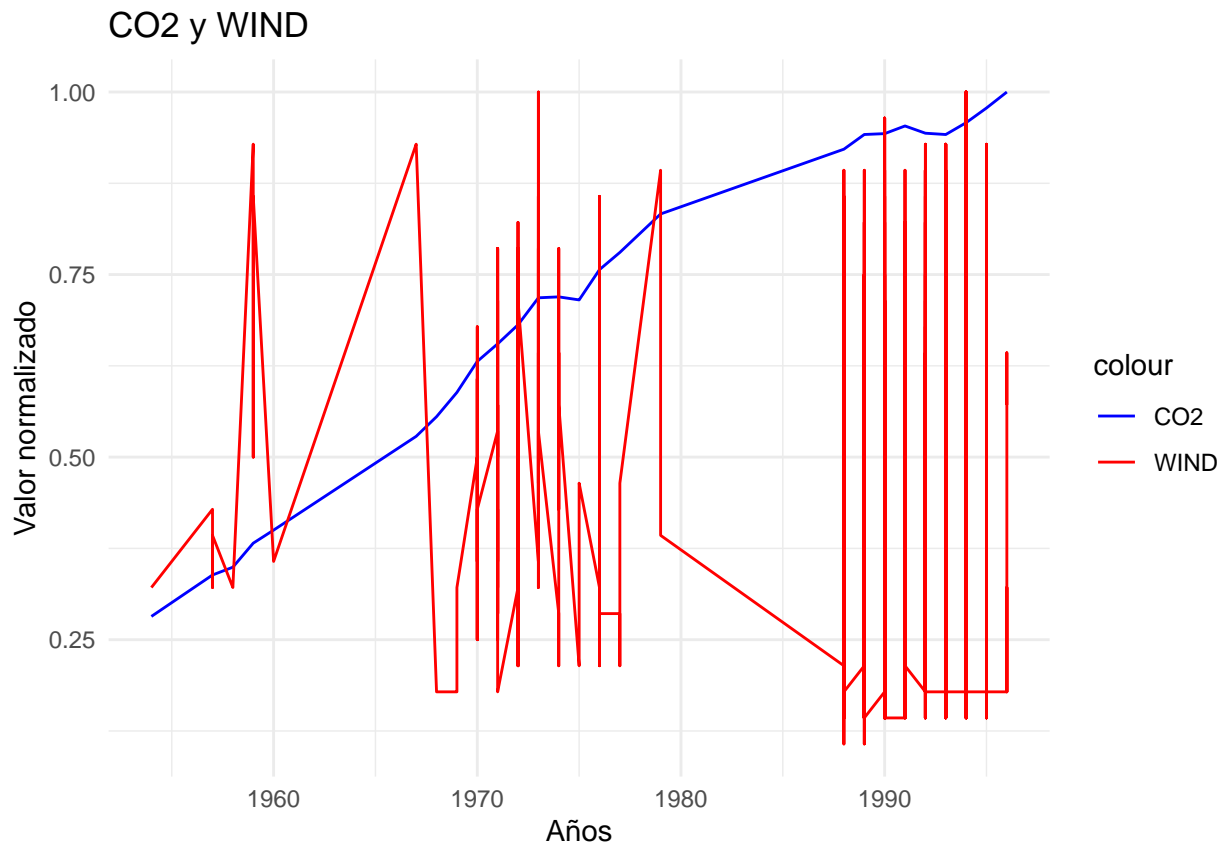
```
# 5.1.3 Gráfica años con la variable Presión
ggplot(huracanes_pacifico, aes(x = year, y = Pressure)) +
  geom_point(color = "green") +
  labs(title = "Tendencia de la Presión Atmosférica por Año",
        x = "Año",
        y = "Presión Atmosférica") +
  theme_minimal()
```



```
# 5.1.4 Elabora gráficas sobre los datos CO2 y WIND
#Popukation en el oceano es 0 por lo que se omite

# Calcula los valores normalizados para CO2 y Wind
huracanes_pacifico$CO2_norm <- huracanes_pacifico$CO2 / max(huracanes_pacifico$CO2)
huracanes_pacifico$Wind_norm <- huracanes_pacifico$Wind / max(huracanes_pacifico$Wind)

# Gráfica usando ggplot2 con datos normalizados
ggplot(huracanes_pacifico, aes(x = year)) +
  geom_line(aes(y = CO2_norm, color = "CO2")) +
  geom_line(aes(y = Wind_norm, color = "WIND")) +
  labs(x = "Años", y = "Valor normalizado", title = "CO2 y WIND") +
  scale_color_manual(values = c("blue", "red"), labels = c("CO2", "WIND")) +
  theme_minimal()
```



```
# 5.2.1 Generar los cuartiles de la variable CO2
cuartiles_CO2 <- quantile(huracanes_pacifico$CO2, probs = c(0.25, 0.5, 0.75))
print("Cuartiles de la variable CO2:")

## [1] "Cuartiles de la variable CO2:"
print(cuartiles_CO2)

##      25%      50%      75%
## 22154.34 22181.81 22429.38

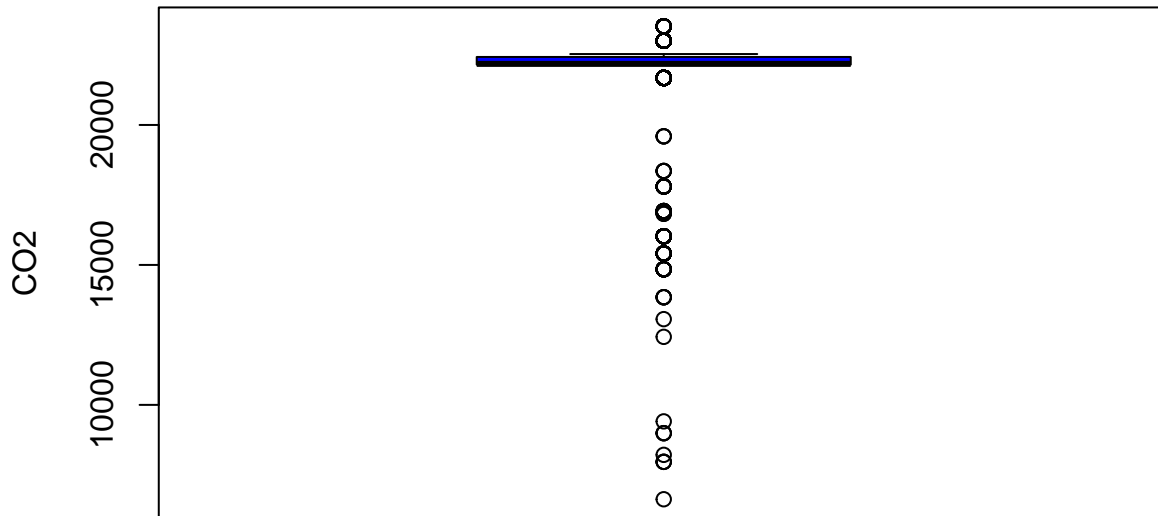
# 5.2.2 Generar los cuartiles de la variable Ocean & CO2
cuartiles_Ocean_CO2 <- quantile(huracanes_pacifico$`Ocean & CO2`, probs = c(0.25, 0.5, 0.75))
print("Cuartiles de la variable Ocean & CO2:")

## [1] "Cuartiles de la variable Ocean & CO2:"
print(cuartiles_Ocean_CO2)

## 25% 50% 75%
## NA NA NA

# Graficar los cuartiles de la variable CO2
boxplot(huracanes_pacifico$CO2, main = "Diagrama de Caja - CO2", ylab = "CO2", col = "blue")
```

Diagrama de Caja – CO2



#Análisis de datos

#6.1 Prepara los datos

```
data_for_cor <- huracanes_pacifico[, c("Wind", "Pressure", "CO2", "year", "Time", "Longitud", "Latitude")]
correlation_matrix <- cor(data_for_cor)
correlation_matrix
```

```
##           Wind      Pressure      CO2      year      Time
## Wind      1.000000000 -0.979061130 -0.08080223 -0.05240347  0.007147203
## Pressure -0.979061130  1.000000000  0.09693325  0.06643609 -0.010379891
## CO2      -0.080802227  0.096933249  1.000000000  0.93844176 -0.124223261
## year     -0.052403475  0.066436087  0.93844176  1.000000000 -0.120708602
## Time      0.007147203 -0.010379891 -0.12422326 -0.12070860  1.000000000
## Longitud  0.001058286 -0.003866773 -0.09476855 -0.12331976  0.014739549
## Latitude -0.022581683  0.026449371 -0.03145183 -0.02393488 -0.000844307
##           Longitud      Latitude
## Wind      0.001058286 -0.022581683
## Pressure -0.003866773  0.026449371
## CO2      -0.094768552 -0.031451828
## year     -0.123319757 -0.023934880
## Time      0.014739549 -0.000844307
## Longitud  1.000000000 -0.230618943
## Latitude -0.230618943  1.000000000
```

```
dataK <- huracanes_pacifico_hu[, c("Wind", "Pressure", "CO2", "year")]
```

#6.2 Gráfica la correlación de los datos

```
corrplot(correlation_matrix, method = "ellipse")
```



```

## [408] 6 6 6 6 6 6 6 6 6 6 6 6 6 6 6 6 6 6 6 6 6 6 6 6 6 6 6 6 6 6 6 6
## [445] 6 6 6 6 6 6 6 6 6 6 6 6 6 6 6 6 6 6 6 6 6 6 6 6 6 6 6 6 6 6 6 6
## [482] 6 6 6 6 6 6 6 6 6 6 6 6 6 6 6 6 6 6 6 6 6 6 6 6 6 6 6 6 6 6 6 6
## [519] 2 2 2 2 2 2 2 2 2 2 2 2 2 2 2 2 2 2 2 2 2 2 2 2 2 2 2 2 2 2 2 2
## [556] 2 2 2 2 2 2 2 2 2 2 2 2 2 2 2 2 2 2 2 2 2 2 2 2 2 2 2 2 2 2 2 2
## [593] 2 2 2 2 2 2 2 2 2 2 2 2 2 2 2 2 2 2 2 2 2 2 2 2 2 2 2 2 2 2 2 2
## [630] 2 2 2 2 2 2 2 2 2 2 2 2 2 2 2 2 2 2 2 2 2 2 2 2 2 2 2 2 2 2 2 2
## [667] 2 2 2 2 2 2 2 2 6 6 6 6 6 6 6 6 6 6 6 6 6 6 6 6 6 6 6 6 6 6 6 6
## [704] 6 6 6 6 6 6 6 6 6 6 6 6 6 6 6 6 6 6 6 6 6 6 6 6 6 6 6 6 6 6 6 6
## [741] 6 6 6 6 6 6 6 6 6 6 6 6 6 6 6 6 6 6 6 6 6 6 6 6 6 6 6 6 6 6 6 6
## [778] 6 6 6 6 6 6 6 6 6 6 6 6 6 6 6 6 6 6 6 6 6 6 6 6 6 6 6 6 6 6 6 6
## [815] 6 6 6 6 6 6 6 6 6 6 6 6 6 6 6 6 6 6 6 6 6 6 6 6 6 6 6 6 6 6 6 6
## [852] 6 6 6 6 6 6 6 6 6 6 6 6 6 6 6 6 6 6 6 6 6 6 6 6 6 6 6 6 6 6 6 6
## [889] 6 6 6 6 6 6 6 6 6 6 6 6 6 6 6 6 6 6 6 6 6 6 6 6 6 6 6 6 6 6 6 6
## [926] 6 6 6 6 6 6 6 6 6 6 6 6 6 6 6 6 6 6 6 6 6 6 6 6 6 6 6 6 6 6 6 6
## [963] 6 6 6 6 6 6 6 6 6 6 6 6 6 6 6 6 6 6 6 6 6 6 6 6 6 6 6 6 6 6 6 6
## [1000] 6 6 6 6 6 6 6 6 6 6 6 6 6 6 6 6 6 6 6 6 6 6 6 6 6 6 6 6 6 6 6 6
## [1037] 6 6 6 6 6 6 6 6 6 6 6 6 6 6 6 6 6 6 6 6 6 6 6 6 6 6 6 6 6 6 6 6
## [1074] 6 6 6 6 6 6 6 6 6 6 6 6 6 6 6 6 6 6 6 6 6 6 6 6 6 6 6 6 6 6 6 6
## [1111] 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1
## [1148] 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1
## [1185] 1 1 1 3 3 3 3 3 3 3 3 3 3 3 3 3 3 3 3 3 3 3 3 3 3 3 3 3 3 3 3 3
## [1222] 3 3 3 3 3 3 3 3 3 3 3 3 3 3 3 3 3 3 3 3 3 3 3 3 3 3 3 3 3 3 3 3
## [1259] 3 3 3 3 3 3 3 3 3 3 3 3 3 3 3 3 3 3 3 3 3 3 3 3 3 3 3 3 3 3 3 3
##
## Within cluster sum of squares by cluster:
## [1] 79293.82 81319.72 2449143.45 231873639.30 8551901.82
## [6] 688867.03
## (between_SS / total_SS = 92.6 %)
##
## Available components:
##
## [1] "cluster" "centers" "totss" "withinss" "tot.withinss"
## [6] "betweenss" "size" "iter" "ifault"

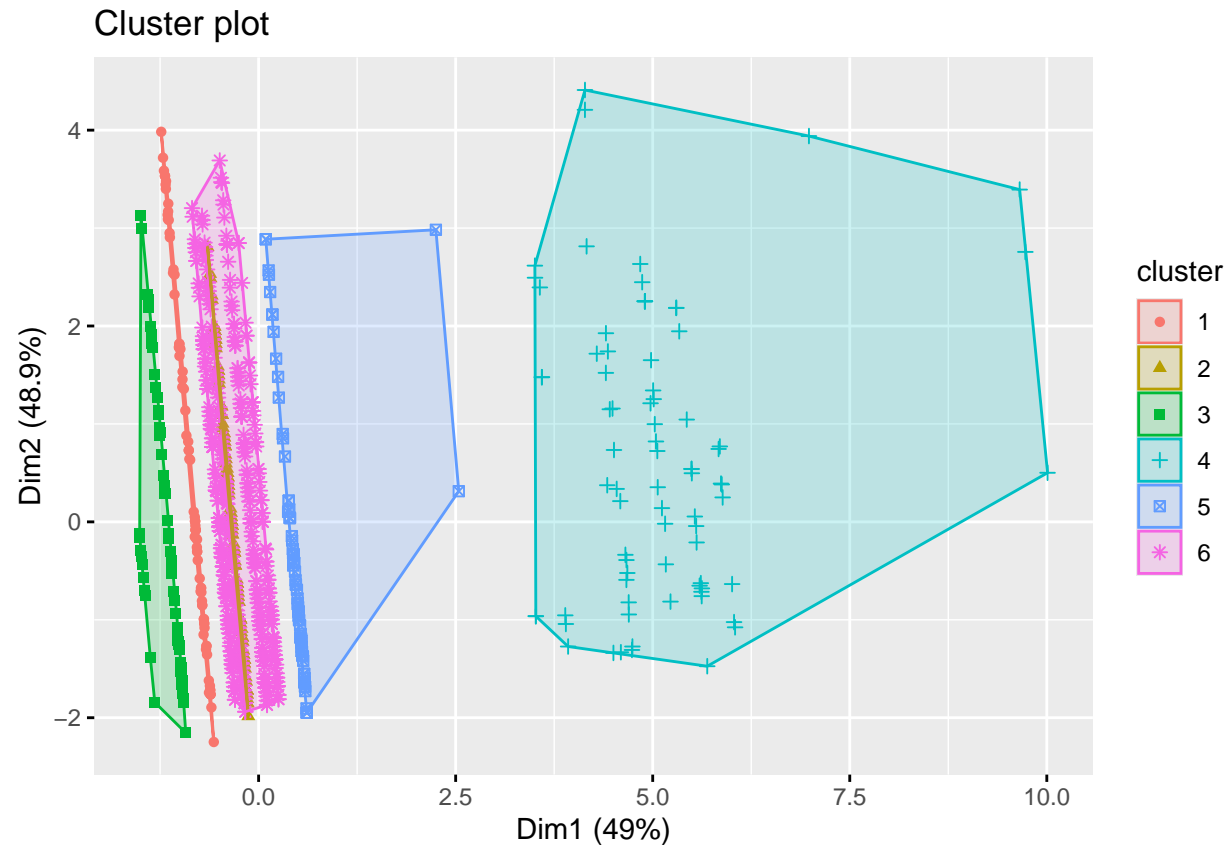
```

#6.3.1 Gráfica los grupos obtenidos mediante el algoritmo de k-means

```

fviz_cluster(km5, data = dataK, repel = TRUE, geom = "point", show.clust.cent = FALSE)

```



#6.4 Obtén la regresión lineal para 3 pares de variables a tu elección

```
lm_model1 <- lm(Wind ~ Pressure, data = huracanes_pacifico)
summary(lm_model1)
```

```
##
## Call:
## lm(formula = Wind ~ Pressure, data = huracanes_pacifico)
##
## Residuals:
##      Min       1Q   Median       3Q      Max
## -59.240  -3.250   0.167   3.171  37.541
##
## Coefficients:
##              Estimate Std. Error t value Pr(>|t|)
## (Intercept) 1463.55207    4.62354   316.5  <2e-16 ***
## Pressure    -1.42088    0.00466  -304.9  <2e-16 ***
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Residual standard error: 5.644 on 4020 degrees of freedom
## Multiple R-squared:  0.9586, Adjusted R-squared:  0.9586
## F-statistic: 9.299e+04 on 1 and 4020 DF, p-value: < 2.2e-16
```

```
lm_model2 <- lm(CO2 ~ Pressure, data = huracanes_pacifico)
summary(lm_model2)
```

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

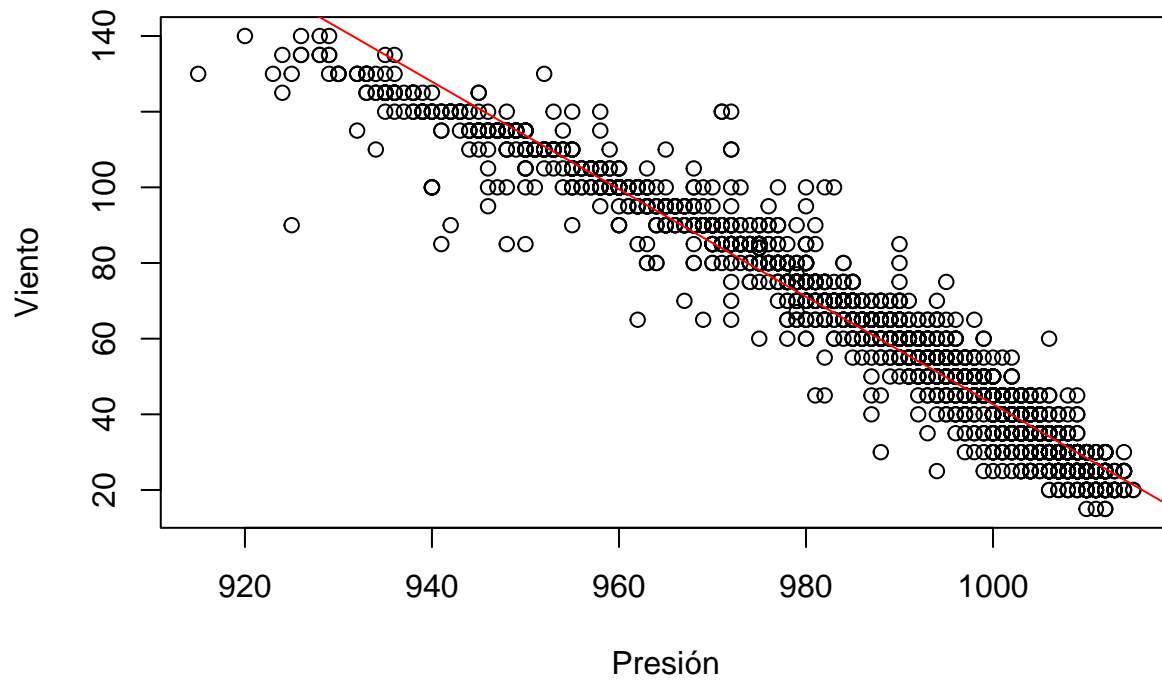
```
## lm(formula = C02 ~ Pressure, data = huracanes_pacifico)
##
## Residuals:
##      Min       1Q   Median       3Q      Max
## -15322.1    46.8    161.1    392.3   1653.3
##
## Coefficients:
##              Estimate Std. Error t value Pr(>|t|)
## (Intercept) 15283.577   1092.203   13.993 < 2e-16 ***
## Pressure      6.797      1.101    6.175 7.27e-10 ***
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Residual standard error: 1333 on 4020 degrees of freedom
## Multiple R-squared:  0.009396, Adjusted R-squared:  0.00915
## F-statistic: 38.13 on 1 and 4020 DF, p-value: 7.273e-10

lm_model3 <- lm(C02 ~ Wind, data = huracanes_pacifico)
summary(lm_model3)

##
## Call:
## lm(formula = C02 ~ Wind, data = huracanes_pacifico)
##
## Residuals:
##      Min       1Q   Median       3Q      Max
## -15432.2    55.6    172.7    393.7   1637.2
##
## Coefficients:
##              Estimate Std. Error t value Pr(>|t|)
## (Intercept) 22237.1061    46.0396  483.00 < 2e-16 ***
## Wind         -3.9040     0.7595   -5.14 2.88e-07 ***
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Residual standard error: 1335 on 4020 degrees of freedom
## Multiple R-squared:  0.006529, Adjusted R-squared:  0.006282
## F-statistic: 26.42 on 1 and 4020 DF, p-value: 2.879e-07

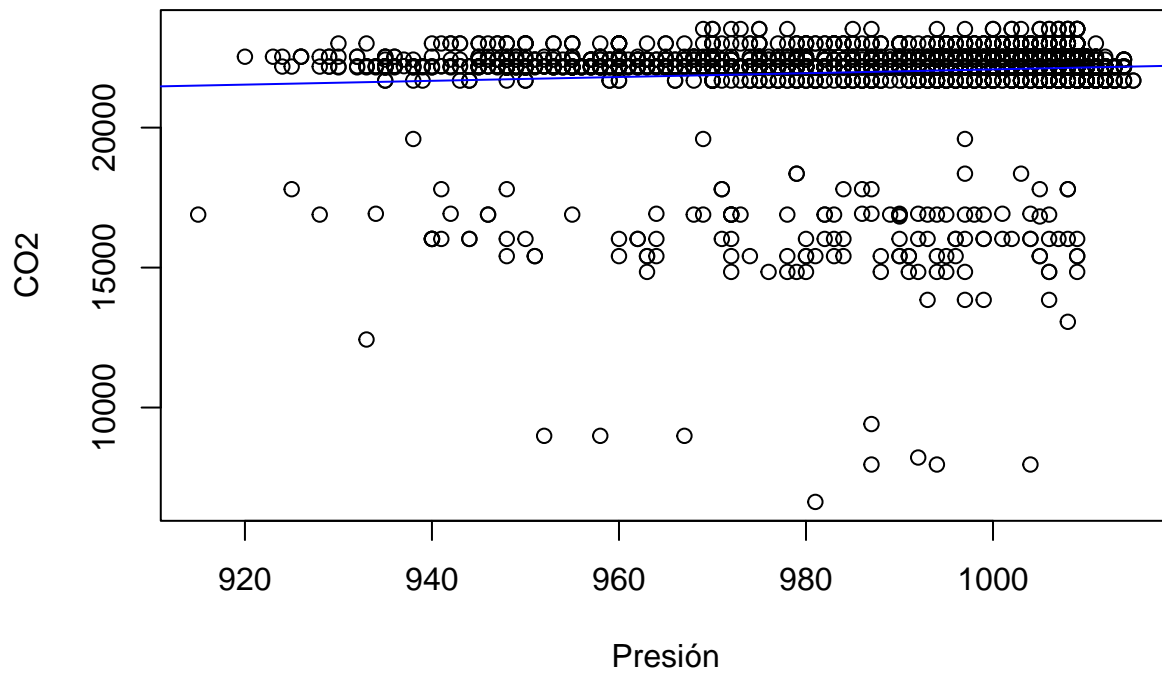
# Gráfica 1: Wind ~ Pressure
plot(huracanes_pacifico$Pressure, huracanes_pacifico$Wind,
     main = "Regresión Lineal: Wind ~ Pressure",
     xlab = "Presión",
     ylab = "Viento")
abline(lm_model1, col = "red")
```

Regresión Lineal: Wind ~ Pressure



```
# Gráfica 2: C02 ~ Pressure
plot(huracanes_pacifico$Pressure, huracanes_pacifico$C02,
     main = "Regresión Lineal: C02 ~ Pressure",
     xlab = "Presión",
     ylab = "C02")
abline(lm_model2, col = "blue")
```

Regresión Lineal: CO2 ~ Pressure



```
# Gráfica 3: CO2 ~ Wind
plot(huracanes_pacifico$Wind, huracanes_pacifico$CO2,
     main = "Regresión Lineal: CO2 ~ Wind",
     xlab = "Viento",
     ylab = "CO2")
abline(lm_model3, col = "green")
```

Regresión Lineal: CO2 ~ Wind

