



**National Technology of Mexico
Technological Institute of Tijuana**

ACADEMIC SUBDIRECTION
Systems and Computing Department

SEMESTER
February – June 2021

CAREER
Information and Communication Technologies Engineer

SUBJECT AND KEY:
Data Mining BDD-1703TI9A

STUDENT'S NAME AND REGISTRATION:
Camacho Manabe Juan Daniel 17210534
Velázquez Farrera César Alejandro 17212937

NAME OF THE JOB:
Practice #1

UNIT TO BE EVALUATED
Unit II

TEACHER'S NAME:
MC. José Christian Romero Hernández

Practice #1

VideoGames

```
Games <- read.csv("vgsales.csv")
Games <- read.csv(file.choose())

#Overview of the nature of the data
head(Games)
colnames(Games)
head(Games)
tail(Games)
str(Games)
summary(Games)

install.packages("ggplot2")
library("ggplot2")

install.packages("dplyr")
library(dplyr)

#1. Scatterplot inclusion
#Global video games sales by platform
ScatPlot <- ggplot(Games, aes(x=Global_Sales, y=Platform,
                             color=Genre))+geom_point()
ScatPlot

#Videogames produced per year
Scatplot2 <- ggplot(Games, aes(x=Platform, y=Year,
                              color=Genre))+geom_point()
Scatplot2

#2. Facets inclusion
#Facets
v <- ggplot(Games, aes(x=Global_Sales))
v + geom_histogram(binwidth = 10, aes(fill=Genre),
                  color="Black")

v + geom_histogram(binwidth = 10, aes(fill=Genre),
                  color="Black") + facet_grid(Genre~., scales="free")

#3. Distribution Graphs
geo<- filter(Games, Genre == "Action")
geo

#Sales per year in the US
```

```

ggplot(data=Games, aes(x=Year, y=EU_Sales)) +
  geom_point()

#Sales per year in the US and we add color
o <- ggplot(Games, aes(x=Year, y=EU_Sales))
o
h <- o + geom_point(color="Blue")
h
#adding captions and decorating with a theme
h +
  xlab("Years") +
  ylab("Million units") +
  ggtitle("Videogames sales in the EU") +
  theme(axis.title.x = element_text(color = "DarkGreen", size=20),
        axis.title.y = element_text(color = "Red", size=20),
        axis.text.x = element_text(color="Black",size = 5),
        axis.text.y = element_text(color="Black",size = 5),
        legend.title = element_text(size = 20),
        legend.text = element_text(size = 20),
        legend.position = c(1,1),
        legend.justification = c(1,1),
        plot.title = element_text(color = "DarkGreen",size = 25))

# X-axis (grid) values
# Number or vector representing the mean / s
# Number or vector representing the standard deviation / s log = FALSE)
# If TRUE, probabilities are returned as logarithms

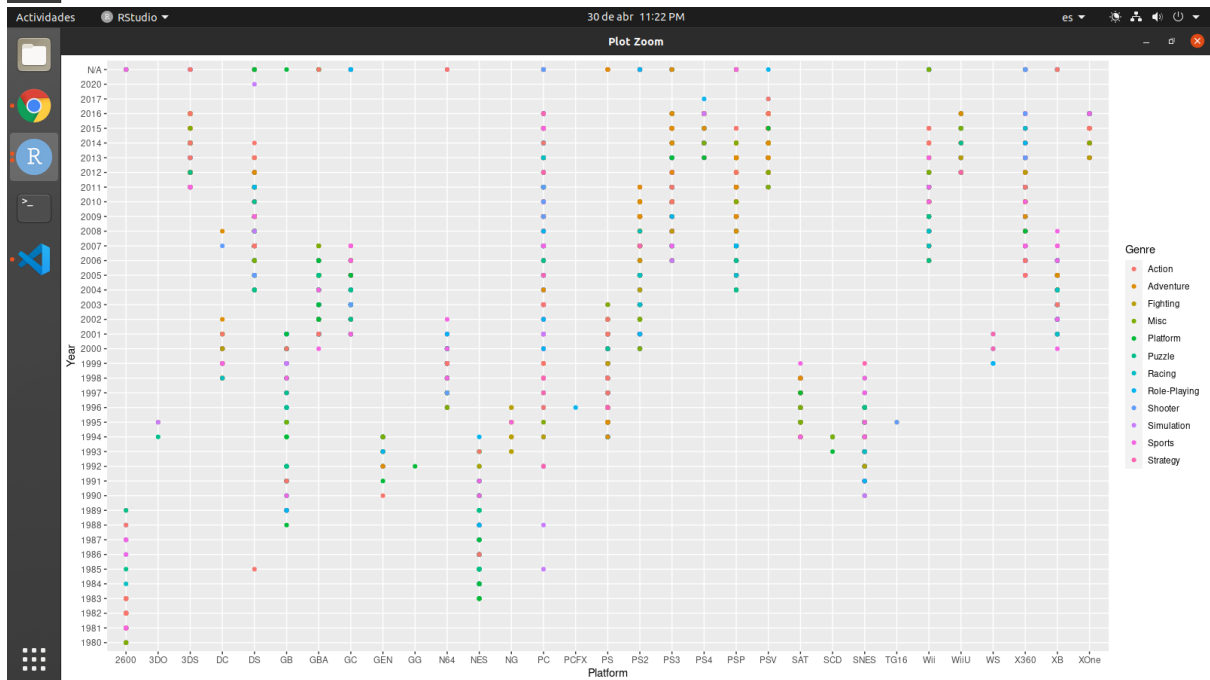
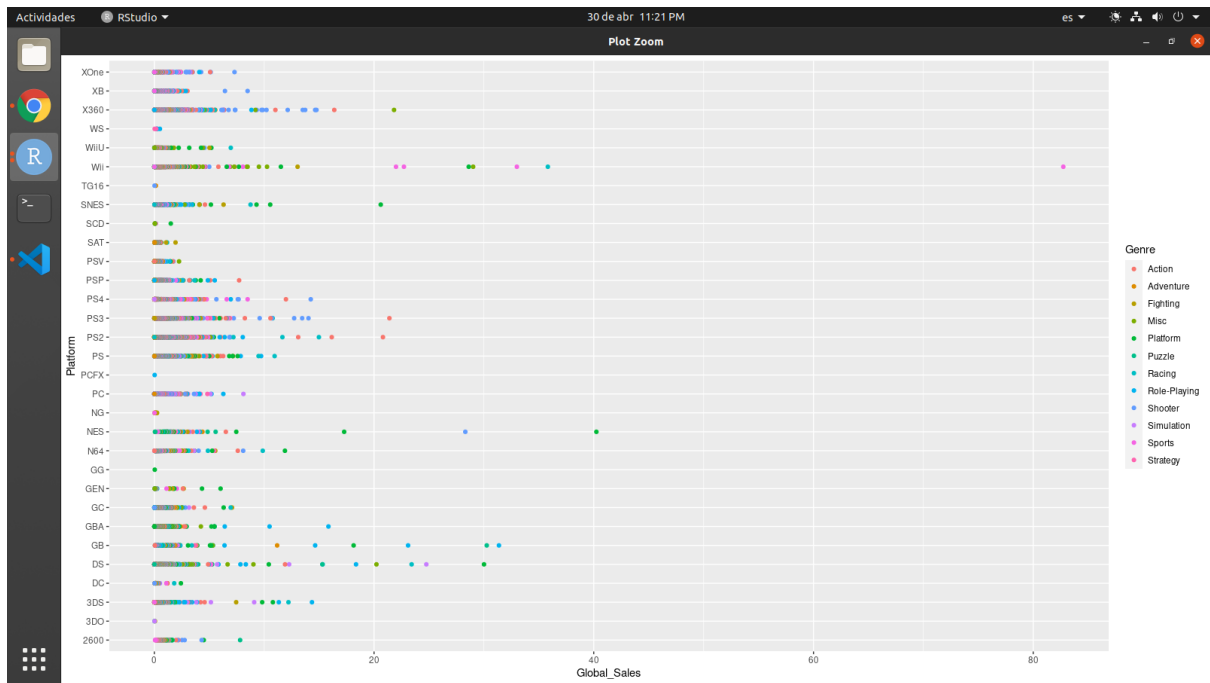
#We separate the variable that we want to analyze
x<-Games$EU_Sales
distribution<-dnorm(x,
                    mean = 0.1467,
                    sd = 1,
                    log = FALSE)

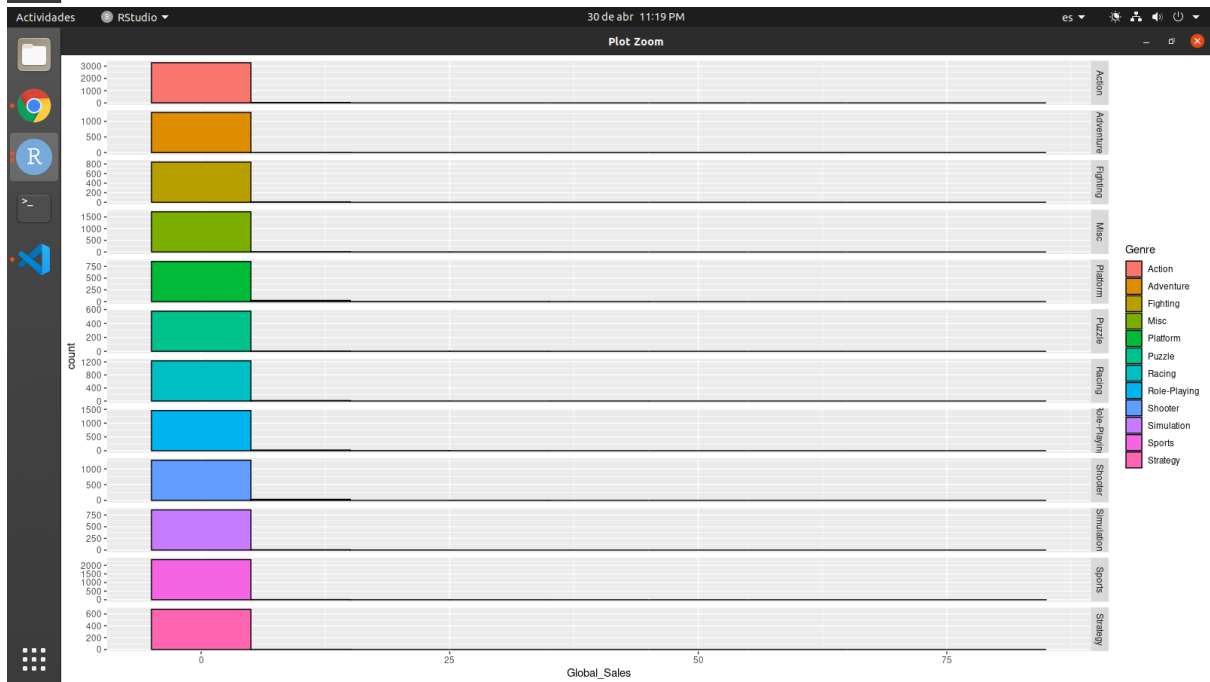
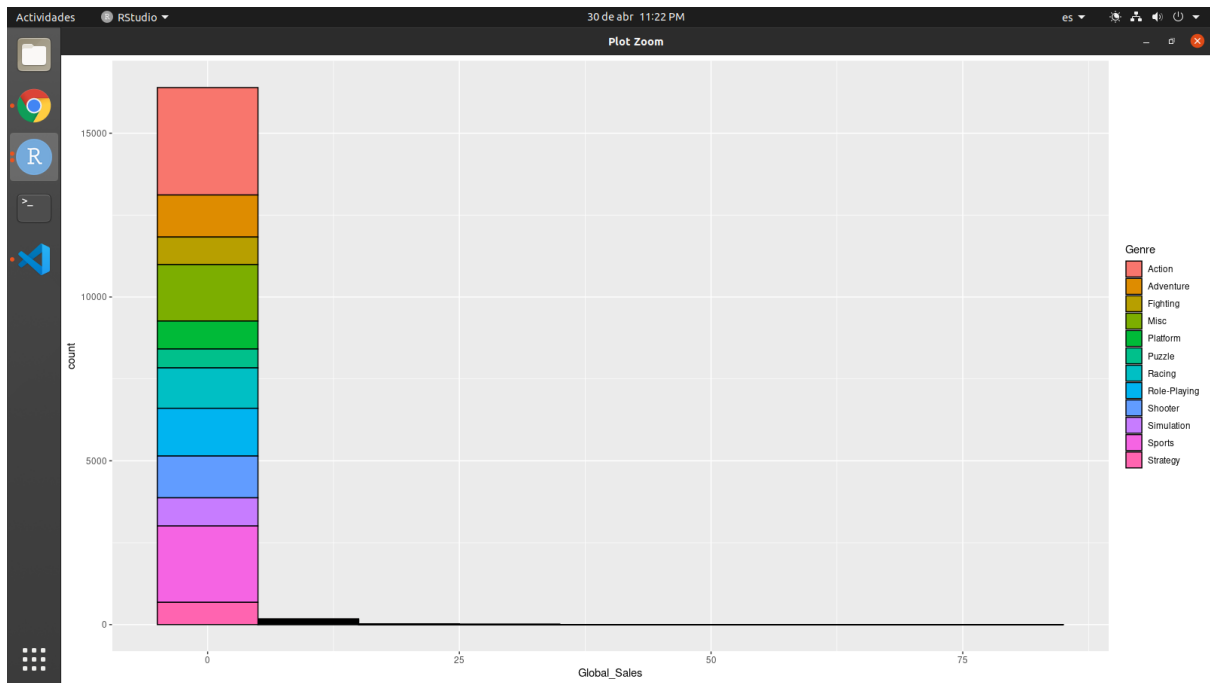
#We separate the variable that we want to analyze
x<-Games$EU_Sales

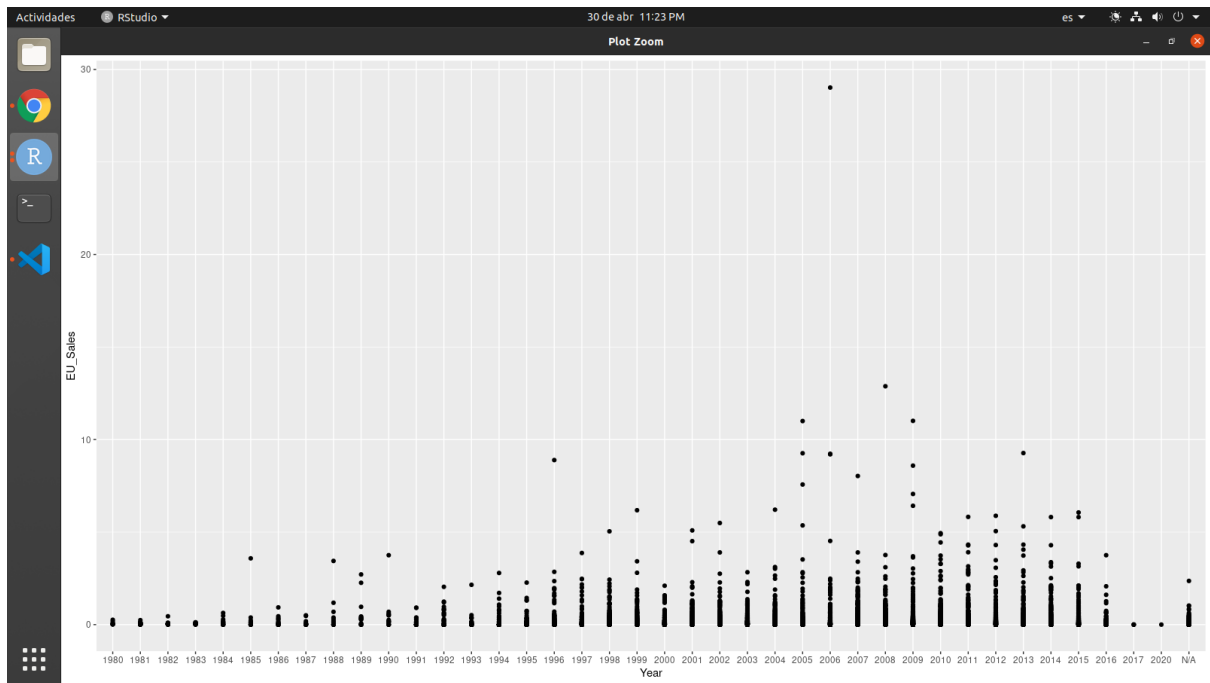
#Graph the normal distribution
plot(x, dnorm(x, mean = 0.1467, sd = 1), type = "l",
     ylab = "", lwd = 1, col = "red")

#Graph the normal distribution
plot(x, dnorm(x, mean = 0.1467, sd = 1),
     ylab = "", lwd = 1, col = "red")

```







Videogames sales in the EU

