Instituto Tecnológico de Tijuana

Subdirección Académica

Departamento de sistemas y computación

SEMESTRE:
Agosto-Diciembre 2021

Nombre de la Materia:

Minería de Datos

Actividad:

Práctica Evaluativa Unidad 2

Profesor:

ROMERO HERNANDEZ JOSE CHRISTIAN

Alumno(s):

Castillo Ramirez Guadalupe   17213043

Ramos Rivera Manue Isai   17212931

# specification our workplace our workplace
getwd()
setwd(("/home/guadalupe/Escritorio/PRacticaEvaluatoria2/PracticaEvaluatoria2")
getwd()

# importing the data with which we are going to work
movies <- read.csv("Project-Data.csv")

str(movies)
summary(movies)

filtering the data, because not all the data present in the document is used

#filterin Genre
filtGenre <- movies$Genre %in%
c("action","adventure","animation","comedy","drama")

# creating the dataframe with the filtered genres
filtermovies <- movies[filtGenre,]

#filtering the studies of our new data frame
filterStudio <- filtermovies$Studio %in% c("Buena Vista Studios","Fox","Paramount
Pictures","Sony","Universal","WB")

# creating the new data frame with all the data we need filtered
Resultmovies <- filtermovies[filterStudio,]

#rename columns

colnames(Resultmovies) <- c("Dayofweek", "Director", "Genre", "MovieTitle",
"RealseDate", "Studio","AdjustedGroosMill","BudgetMill", "GrossMill","IMDbRating",
"MovieLensRating", "OverseasMill",
                    "Overseas", "ProfitMill", "Profit", "RuntimeMin", "UsMill",
"GrossUS")

#Create transparent color

mycol <- rgb(0, 0, 255, max = 255, alpha = 125, names = "blue50")

installing and importing the library ggplot2

```
install.packages('ggplot2')

library(ggplot2)
```
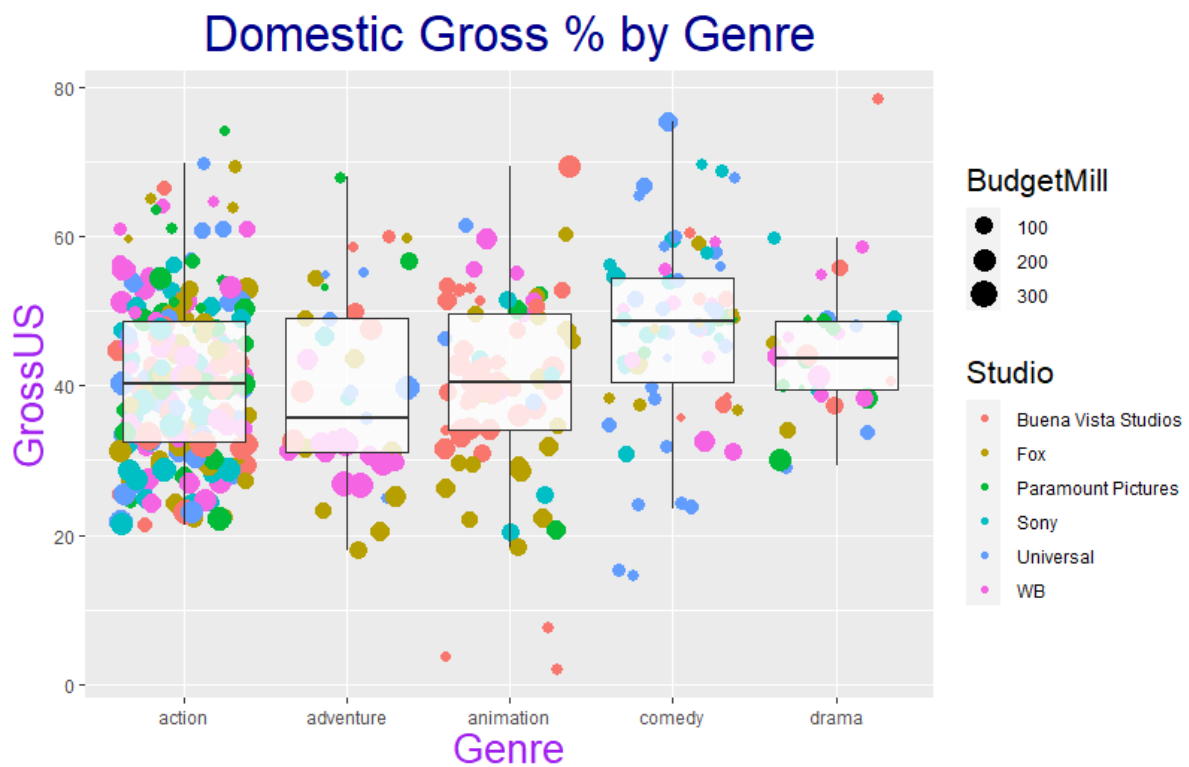
#creating a scatter plot

```
Graphic <- ggplot(data = Resultmovies,
          mapping = aes(x=Genre, y=GrossUS)
          ) +
      labs(title='Domestic Gross % by Genre') +
      theme(axis.title.x = element_text(color = "Purple", size=20),
          axis.title.y = element_text(color = "Purple", size=20),
          legend.title = element_text(color="Black",size=16 ),
          plot.title =  element_text(color = "DarkBlue", size = 25, hjust = 0.5 )


          )

Graphic + geom_boxplot()

Graphic + geom_jitter(aes( size = BudgetMill, color=Studio )) +
geom_boxplot(aes(),alpha=0.8, outlier.colour = NA )
```

Video
https://youtu.be/Zu6iQghpa2g
Repositorio
https://github.com/isarari99/Data-Mining-Ramos

Conclusión

Nos resultó sencilla la práctica ya que al filtrar solo los datos necesarios se construyó una gráfica pequeña nos ahorró la carga de todos los datos que contiene el archivo , es así que la a; crear gráficas nos beneficia para filtrar datos necesarios , sin necesidad de representar todo.