

**Technological Institute of Tijuana****Academic Subdirector****Systems and Computing Department****SEMESTER:** August - December 2021**CAREER:** Computer Systems Engineer**MATTER:** Data Mining**JOB NAME:** Unit 2 - Evaluation**STUDENT NAME AND CONTROL NUMBER:**

Castro Cebreros Alejandro - 16211341

Márquez Millán Seashell Vanessa - 17212153

TEACHER NAME: Jose Christian Romero Hernandez**DATE OF DELIVERY:** November 10, 2021**First we select the csv file that we will use**

These code makes you have to choose the file

```
Movies <- read.csv(file.choose())
```

Rename the columns

Only we changed the names on the columns

```
colnames(Movies) <- c("Dayofweek", "Director", "Genre", "MovieTitle",  
"ReleaseDate", "Studio", "AdjustedGrossMill", "Budget",
```

```
"GrossMill", "IMDbRating", "MovieLensRating", "OverseasMill", "Overseas",  
"ProfitMill", "Profit", "RuntimeMin", "UsMill", "GrossUS")
```

We wont use all the genres so we will filter it to select specific ones

Here select the values we need in this case

```
dfgenre <- Movies$Genre %in%  
c("action", "adventure", "animation", "comedy", "drama")  
  
filtermovies <- Movies[dfgenre,]
```

Same as genre, we use the same method with the studios

Here we go to use the equals action than after but with other values, because we want have diferentes types of points

```
dfstudio <- filtermovies$Studio %in% c("Buena Vista  
Studios", "Fox", "Paramount Pictures", "Sony", "Universal", "WB")
```

We make a new dataframe with the other fitered dataframe

In this part we changed to dataframe we go to use

```
Finalmovies <- filtermovies[dfstudio,]
```

We use libraries extrafont for selecting a custom font and ggplot for the graph

We need to have this libtarys because ggplot we help to made the graphic and extrafont is for changed the font type and made taht mor easy

```
library(extrafont)  
library(ggplot2)
```

****Import and load fonts ****

Load the fonts we need

```
font_import()  
loadfonts()
```

We create the structure of the graph

The last for to do is we the others things we did it join in this part, here is when we chose the color and the style we want, and in this case changed the color of the letters, titles, etc.

```
u <- ggplot(Finalmovies, aes(x = Genre, y = GrossUS))
t <- u + geom_jitter(aes(color = Studio, size = Budget)) + geom_boxplot(
  alpha = 0.5)

t +
  xlab("Genre") +
  ylab("Gross % US") +
  ggtitle("Domestic Gross % by Genre") +
  theme(axis.title.x = element_text(color = "Purple", size=30, face =
    "bold"),
        axis.title.y = element_text(color = "Purple", size=30, face =
    "bold"),
        axis.text.x = element_text(size = 20),
        axis.text.y = element_text(size = 20),
        legend.title = element_text(size = 25),
        legend.text = element_text(size = 15),
        legend.justification = c(1,.5),
        text = element_text(family = "Arial"),
        plot.title = element_text(color = "Black",
                                   size = 35,
                                   hjust = 0.5,
                                   face = "bold")))
```

The graph that we end up

