



**Tecnológico Nacional de México  
Instituto Tecnológico de Tijuana**

**Departamento de Sistemas y Computación  
Ingeniería en Sistemas Computacionales**

**Semestre:**

Febrero - Junio 2022

**Materia:**

Minería de datos

**Docente:**

JOSE CHRISTIAN ROMERO HERNANDEZ

**Unidad:**

U2

**Título del trabajo:**

Práctica Evaluativa - Unidad 2

**Alumnas:**

Hernandez Pablo Anahi Del Carmen	18210486
Pérez Mora Ana Ivonne	18212074



Exportamos el documento que contiene los datos en formato csv, una vez cargado el documento ejecutamos la librería de ggplot2 para poder realizar la gráfica.

```
DataBase <- read.csv(file.choose())  
  
library(ggplot2)  
  
DataBase
```

Filtramos todos los géneros de películas que se muestran en los datos.

```
filtracion <- (DataBase$Genre == "action") | (DataBase$Genre == "adventure") |  
(DataBase$Genre == "animation") | (DataBase$Genre == "comedy") |  
(DataBase$Genre == "drama")
```

En esta otra filtración se mostrarán solo el nombre de la compañía que realizó la película.

```
filtracion2 <- DataBase$Studio %in% c("Buena Vista Studios", "WB", "Fox",  
                                     "Universal", "Sony", "Paramount Studio")
```

Aquí solo se observan los resultados de ambas filtraciones.

```
filtracion  
filtracion2
```

Fue necesaria la realización de un segundo data frame ya que se hace referencia a los dos filtros creados anteriormente.

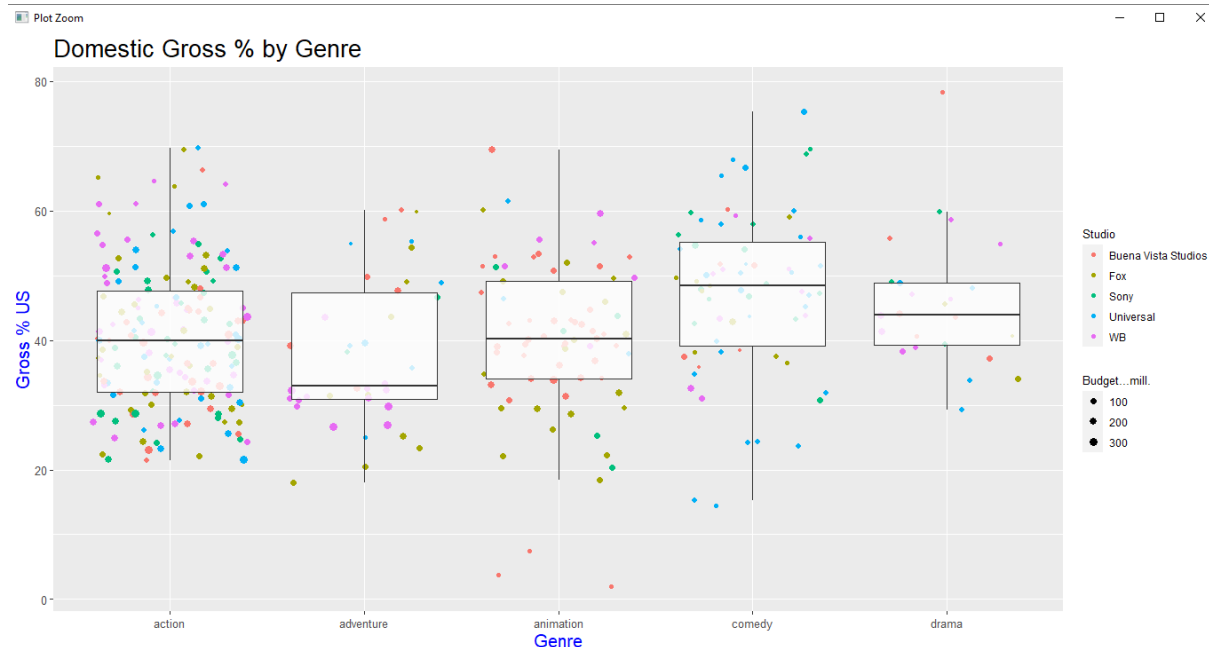
```
DataBase2 <- DataBase[filtracion & filtracion2,]  
  
DataBase2
```

Para la realización de la gráfica solo utilizaremos la variable de DataBase 2, ya en es ahí donde se encuentran todos los datos filtrados.

```
Grafica <- ggplot(data = DataBase2, aes(x= Genre, y= Gross...US))+  
  geom_jitter(aes(size= Budget...mill., colour=Studio))+  
  geom_boxplot(alpha = 0.8, outlier.color = NA)+  
  scale_size_continuous(range = c(1,3))+  
  xlab("Genre")+  
  ylab("Gross % US")+  
  ggtitle("Domestic Gross % by Genre")+  
  theme(  
    axis.title.x = element_text(colour = "Blue", size = 15),  
    axis.title.y = element_text(colour = "Blue", size = 15),  
    axis.text.x = element_text(size = 10),  
    axis.text.y = element_text(size = 10),  
    plot.title = element_text(size = 20),  
    legend.title = element_text(size = 10),  
    text = element_text(family = "Sans"))  
  
Grafica
```



De acuerdo a la gráfica que se nos otorgó, asimilamos que la que se presenta en este documento es idéntica, por lo cual se da por realizado la recreación del código perdido.



Enlace al video:

<https://youtu.be/umuZrkfFq38>