



TECNOLÓGICO NACIONAL DE MÉXICO



**INSTITUTO TECNOLÓGICO DE TIJUANA**

**CARRERA**

**INGENIERÍA EN SISTEMAS COMPUTACIONALES**

**MATERIA**

**MINERÍA DE DATOS**

**TAREA**

**PRÁCTICA EVALUATORIA, UNIDAD #2**

**FECHA ENTREGA**

**04/05/2022**

**ALUMNO(A)**

**HOWARD HERRERA ERWIN #18210716**

**PÉREZ LÓPEZ ALICIA GUADALUPE #18210514**

**DOCENTE**

**JOSE CHRISTIAN ROMERO HERNANDEZ**



**TECNOLÓGICO NACIONAL DE MÉXICO**

**INSTITUTO TECNOLÓGICO DE TIJUANA**

**SUBDIRECCIÓN ACADÉMICA**

Departamento de Sistemas y Computación

**EXAMEN**

Carrera: Ingeniería En Sistemas Computacionales/ Tecnologías de la información/ Informática Período: **Febrero-Junio 2022** Materia: Minería de datos Grupo: Salón: Unidad (es) a evaluar: Unidad 2 Tipo de examen: Práctico Fecha: Catedrático: Jose Christian Romero Hernandez Firma del maestro: Calificación:

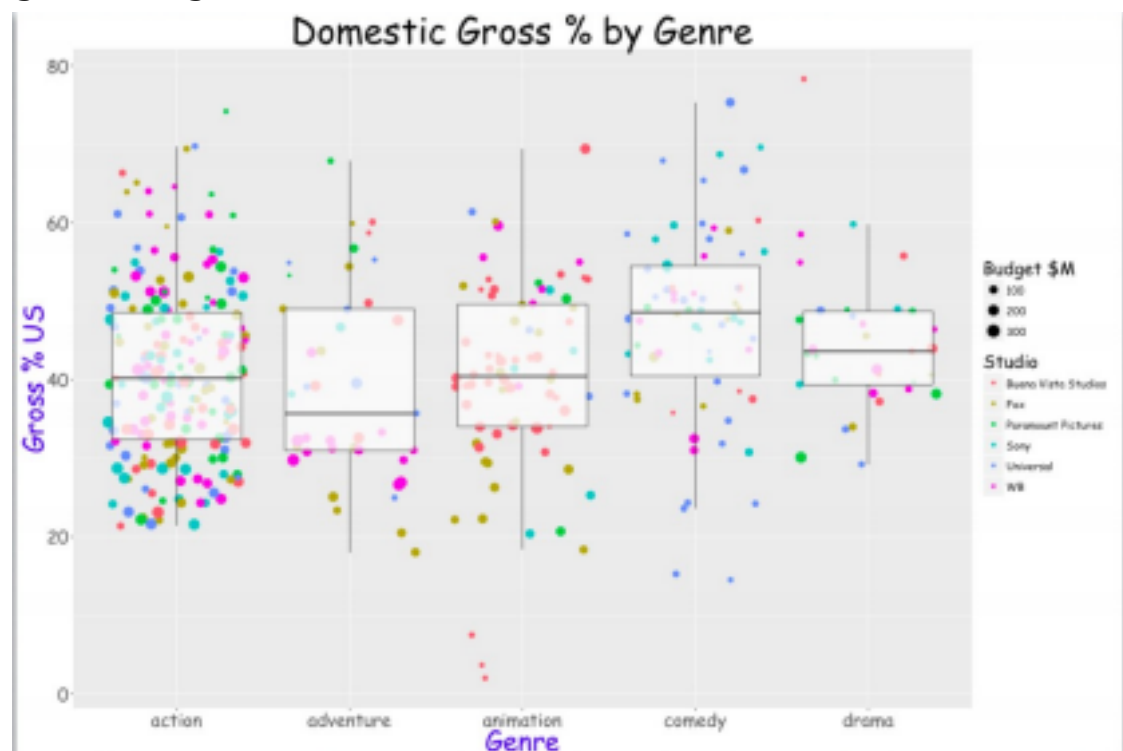
Alumno: \_\_\_\_\_ No. Control: \_\_\_\_\_

**Instrucciones**

**Desarrolle el siguiente problema con R y RStudio para la extracción de conocimiento que el problema requiere.**

Los directivos del sitio web de reseñas de películas están muy contentos con su entrega anterior y ahora tienen una nueva requerimiento para usted.

El consultor anterior había creado una gráfica para ellos que se ilustra en la siguiente imagen.



Sin embargo, el código R utilizado para crear la gráfica se ha perdido y no puede ser recuperado.

Su tarea es crear el código que volverá a crear la misma gráfica haciendo que se vea lo más cerca posible del original.

Se le proporcionará con un nuevo conjunto de datos el cual puedes



encontrar en este link:

<https://github.com/jcromerohdz/DataMining/blob/master/Datasets/Project-Data.csv> O si ya tienen clonado mi repositorio entonces se encuentre en la carpeta: DataMining/Datasets/Project-Data.csv

### **Pista**

- Tenga en cuenta que no todo los Géneros (Genre) y estudios (Studio) son usados. - Necesitarás filtrar tu dataframe después de importar el archivo csv.

### **Instrucciones de evaluación**

- Tiempo de entrega 4 días  
- Al terminar poner el código y la documentación con su explicación en el branch correspondiente de su github, así mismo realizar su explicación de la solución en su google drive en documento de google (Portada, Introducción, Desarrollo, etc). - Finalmente defender su desarrollo en un video de 6-8 min explicando su solución y observaciones, este servirá para dar su calificación de esta práctica evaluatoria, este video debe subirse a youtube para ser compartido por un link público (Utilicen algún software de captura de video con las cámaras encendidas y graben su defensa para elaborar el video).

Happy Coding :)



## Práctica

We assign the graphics library, and open the file that we will analyze called Project data.

```
library(ggplot2)
```

```
#Read doc "Project-Data"
```

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

```
movies
```

RStudio interface showing the console output of the 'movies' data frame. The console displays the following data:

	Day.of.Week	Director	Genre	Movie.Title
1	Friday	Brad Bird	action	Tomorrowland
2	Friday	Scott Waugh	action	Need for Speed
3	Friday	Patrick Hughes	action	The Expendables 3
4	Friday	Phil Lord, Chris Miller	comedy	21 Jump Street
5	Friday	Roland Emmerich	action	White House Down
6	Friday	David Ayer	action	Fury
7	Thursday	Rob Marshall	adventure	Into the Woods
8	Friday	Daniel Espinosa	action	Safe House
9	Friday	Gary Shore	action	Dracula Untold
10	Friday	Henry Joost, Ariel Schulman	horror	Paranormal Activity 3
11	Friday	Eric Brevig	animation	Yogi Bear
12	Friday	Morten Tyldum	biography	The Imitation Game
13	Thursday	Bryan Singer	biography	Valkyrie
14	Friday	Seth Gordon	comedy	Horrible Bosses
15	Friday	Ryan Murphy	drama	Eat Pray Love
16	Friday	Paul Greengrass	drama	Captain Phillips
17	Friday	Hayao Miyazaki	animation	Ponyo
18	Friday	Jaume Collet-Serra	action	Non-Stop
19	Friday	Jonathan Liebesman	action	Battle: Los Angeles

## Filters

we filter the information (data) that we want to obtain, in this case the genre and the production studios.

```
> movies<-movies[movies$Studio=="Buena Vista Studios"|movies$Studio=="Fox"|movies$Studio=="Paramount Pictures"|mo
vies$Studio=="Sony"|movies$Studio=="Universal"|movies$Studio=="WB",]
>
> movies<-movies[movies$Genre=="action"|movies$Genre=="adventure"|movies$Genre=="animation"|movies$Genre=="comed
y"|movies$Genre=="drama",]
> |
```



## Graphis

We designate the value x the gender data.

Y percentage

```
# Graphs
```

```
Ggraph <- ggplot(movies, aes(x=Genre, y=Gross...US))
```

```
Ggraph + geom_jitter(aes(size=Budget...mill.,color=Studio)) + geom_boxplot(size=0.4, alpha=0.8)
```

```
GGgraph <- Ggraph + geom_jitter(aes(size=Budget...mill.,color=Studio)) + geom_boxplot(size=0.4, alpha=0.8)
```

```
GGgraph
```

```
> Ggraph <- ggplot(movies, aes(x=Genre, y=Gross...US))
```

```
>
```

```
> Ggraph + geom_jitter(aes(size=Budget...mill.,color=Studio)) + geom_boxplot(size=0.4, alpha=0.8)
```

```
>
```

```
> GGgraph <- Ggraph + geom_jitter(aes(size=Budget...mill.,color=Studio)) + geom_boxplot(size=0.4, alpha=0.8)
```

```
> GGgraph
```

## Graphic customization

We define the color of the points and the size.

```
> GGgraph +
```

```
+ xlab("Genre") +
```

```
+ ylab("Groos % US") +
```

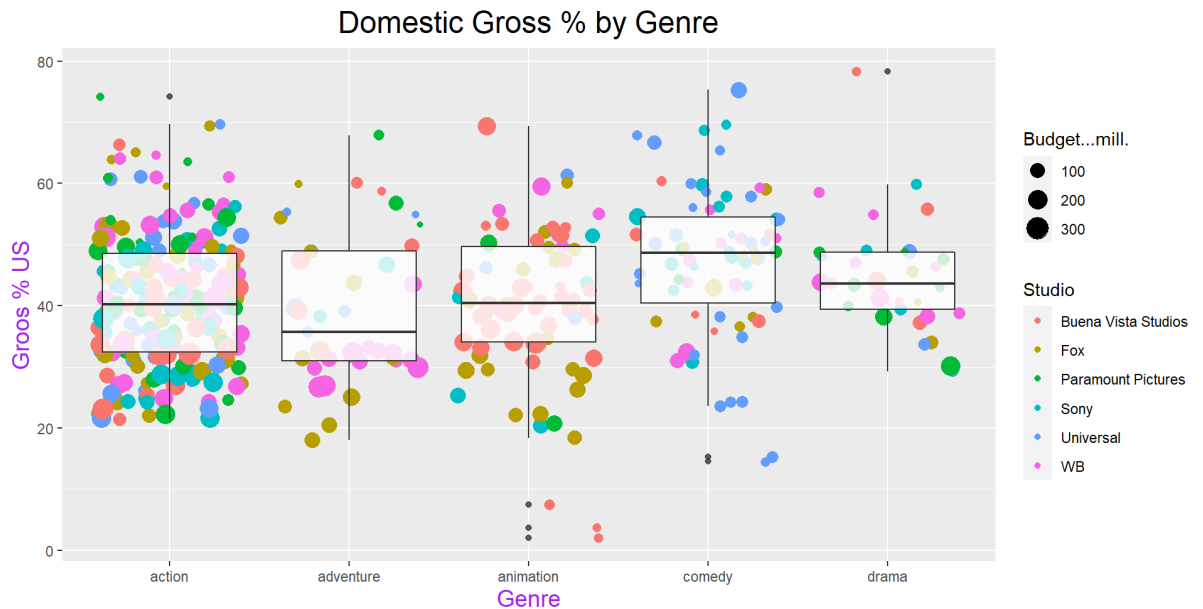
```
+ ggtitle("Domestic Gross % by Genre")+
```

```
+ theme(axis.title.x = element_text(color = "Purple", size=14),
```

```
+ axis.title.y = element_text(color = "Purple", size=14),
```

```
+ plot.title = element_text(color = "Black", size=18,hjust=0.5))
```

## Graph



## Conclusions

**Alicia Pérez:** We had a bit of difficulty knowing which was the correct filter to obtain the graph, as similar as possible to the one we were asked to match, after investigating, we managed to have a similar one, in particular, it is interesting to me as a data can make the graph look very different, it makes a big difference.

**Erwin Howard:** It was really interesting to see how the graph changes when trying to resemble the one that was asked of us in the evaluation. In the end we succeeded and reinforced teamwork and communication, seeing my partner's points of view.

In this practice reinforce knowledge learned in the last unit.

Video:

[https://www.youtube.com/watch?v=uaQcs3C\\_VFc](https://www.youtube.com/watch?v=uaQcs3C_VFc)

GitHub:

<https://github.com/Aliciap26/MINERIA-DATOS/tree/Unit-2/Evaluativ%20practice>