



EDUCACIÓN
SECRETARÍA DE EDUCACIÓN PÚBLICA



TECNOLÓGICO
NACIONAL DE MÉXICO



**TECNOLÓGICO NACIONAL DE MÉXICO INSTITUTO
TECNOLÓGICO DE TIJUANA**

SUBDIRECCIÓN ACADÉMICA

DEPARTAMENTO DE SISTEMAS Y COMPUTACIÓN

SEMESTRE FEBRERO-JUNIO 2022

CARRERA

Ingeniería en informática

MATERIA

Minería de datos

TÍTULO

Práctica evaluatoria 2

Integrantes:

Munguía silva Edgar Geovanny #17212344

Lopez Higuera Saul Alfredo #18210493

NOMBRE DEL MAESTRO

Jose Christian Romero Hernandez

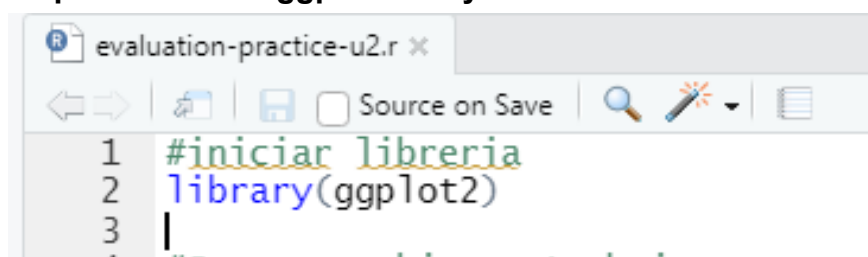
Tijuana Baja California 03 de mayo del 2022

Introduction.

In unit 2 we learned more about the origin of the graphs, about their grammar among other things, as well as the types of probability distributions and statistics, without forgetting that we learned to generate graphs in the R studio of point dispersion, graph in facets and statistical graphs.

We continue with some things that we saw in unit 1 such as what it is to load csv files to use them as databases and also to install more packages.

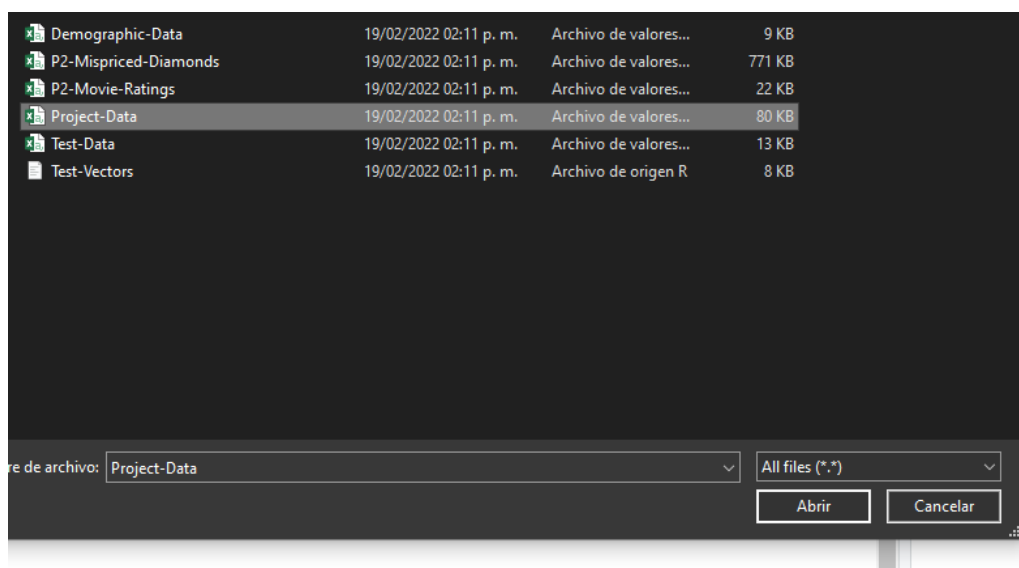
Step #1: Start the ggplot library to be able to create and visualize our plot.



```
> library(ggplot2)
Warning message:
package 'ggplot2' was built under R version 4.1.3
> |
```

Step #2: Import the csv file

```
#Cargar archivo a trabajar
csv<-read.csv(file.choose())
```



Step #3: We filter all the data, gender and name of the study.

```
#Filtrar los datos (Género y nombre del estudio)
filtrogenero<-(csv$Genre=="action")|(csv$Genre=="adventure")|(csv$Genre=="animation")|(csv$Genre=="comedy")|(csv$Genre=="drama")

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

filtrogenero

filtroestudios
```

Step #4: Create a second dataframe based on filtered data.

```
#Crear segundo dataframe en base a los datos filtrados
csv2<-csv[filtrogenero & filtroestudios,]
csv2
```

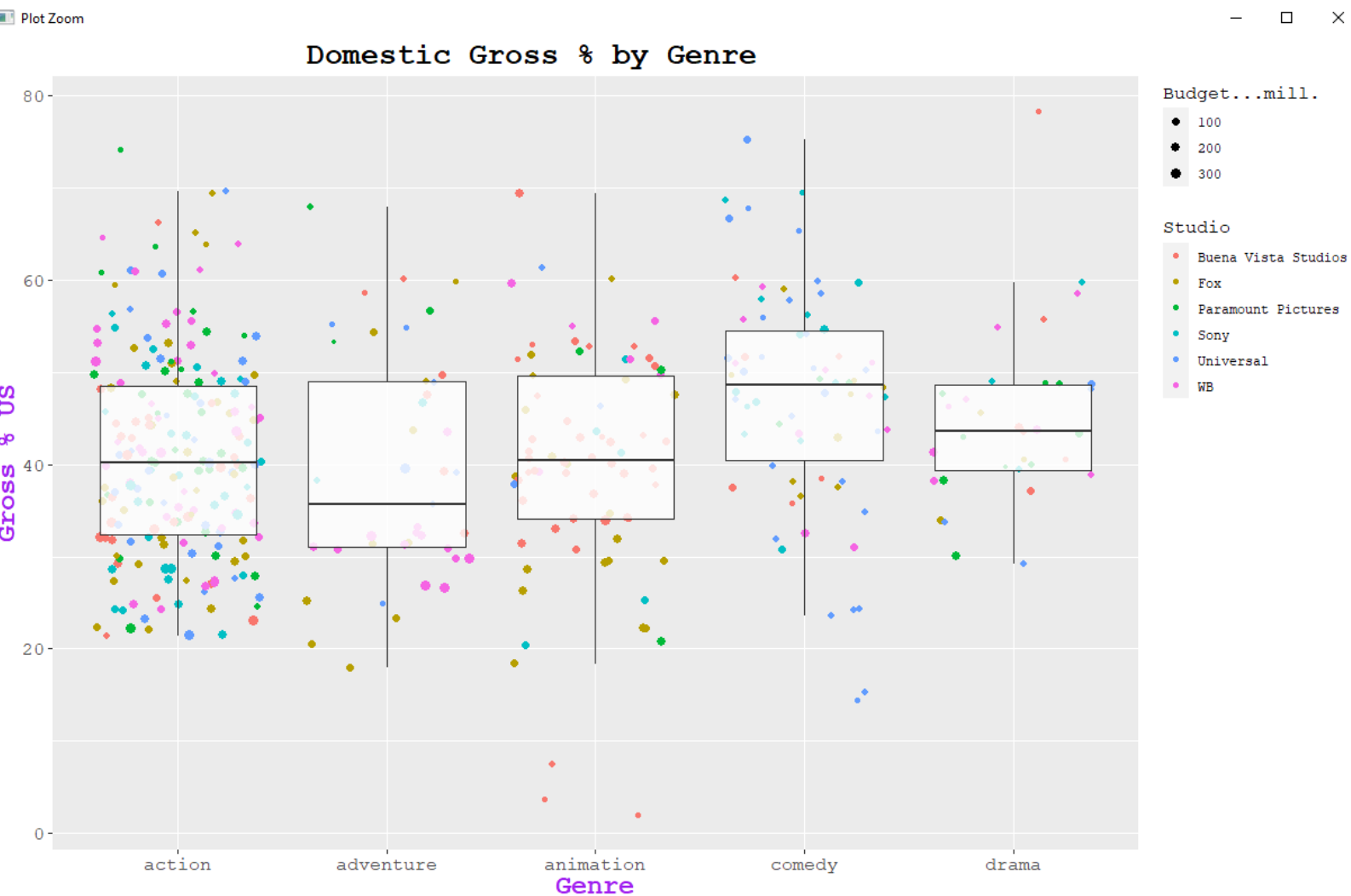
Console Terminal Jobs			
R 4.1.2 · D:/Descargas/			
> csv2<-csv[filtrogenero & filtroestudios,]			
> csv2			
Day.of.Week	Director	Genre	Movie.Title Release.Date Studio
1 Friday	Brad Bird	action	Tomorrowland 22/05/2015 Buena Vista Studios
2 Friday	Scott Waugh	action	Need for Speed 14/03/2014 Buena Vista Studios
4 Friday	Phil Lord, Chris Miller	comedy	21 Jump Street 16/03/2012 Sony
5 Friday	Roland Emmerich	action	White House Down 28/06/2013 Sony
6 Friday	David Ayer	action	Fury 17/10/2014 Sony
7 Thursday	Rob Marshall	adventure	Into the Woods 25/12/2014 Buena Vista Studios
8 Friday	Daniel Espinosa	action	Safe House 10/02/2012 Universal
9 Friday	Gary Shore	action	Dracula Untold 10/10/2014 Universal
11 Friday	Eric Brevig	animation	Yogi Bear 17/12/2010 WB
15 Friday	Ryan Murphy	drama	Eat Pray Love 13/08/2010 Sony
16 Friday	Paul Greengrass	drama	Captain Phillips 11/10/2013 Sony
17 Friday	Hayao Miyazaki	animation	Ponyo 14/08/2009 Buena Vista Studios
18 Friday	Jaume Collet-Serra	action	Non-Stop 28/02/2014 Universal
19 Friday	Jonathan Liebesman	action	Battle: Los Angeles 11/03/2011 Sony
20 Friday	Christopher McQuarrie	action	Jack Reacher 21/12/2012 Paramount Pictures
21 Friday	Dennis Dugan	comedy	Just Go With It 11/02/2011 Sony
22 Friday	Robert Luketic	comedy	The Ugly Truth 24/07/2009 Sony
23 Friday	Jake Kasdan	comedy	Bad Teacher 24/06/2011 Sony

Step #5: creation of the graph.

```
#Creacion de grafica
graf<-ggplot(data=csv2,aes(x=Genre,y=Gross...US))+geom_jitter(aes(size=Budget...mill.,colour=Studio))+geom_boxplot(alpha=0.8,outlier.colour=
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=17),axis.title.y=element_text(colour="Blue",size=17),axis.text.x=element_text(size=12),axis.text.y=element_text(size=1
2),plot.title=element_text(size=22),legend.title=element_text(size=12),text=element_text(family="mono"))
> |
```

Step #6: Show the graph.

```
#Mostrar datos de la grafica
graf+ xlab("Genre") + ylab("Gross % US") + ggtitle("Domestic Gross % by Genre")+theme(axis.title.x=element_text(family="mono",
color="Purple", size=18,face="bold"),axis.title.y = element_text(family = "mono",color = "Purple", size=18,face="bold"),
legend.justification = c(1,1),plot.title = element_text(family = "mono",
color="Black", size = 20,hjust=0.4,face="bold"))
```



Conclusions.

Edgar Munguia: Working with a lot of data may be complicated, because most of the times, you will have to work with specific data, i mean, you will need to filter the data in order to analyze and interpret the respective data, that was what we learned in this practice, to filter, create and show specific data in a graphical interpretation, in this case, we completely ignored some data to focus in other data that we needed to analyze.

Saul Higuera: What we can do with data mining is impressive, in this unit we have to learn about more types of graphs, everything we have learned from rstudio seems very interesting to me, besides the way in which it can be graphed seems very useful to me with a tool as powerful as RStudio.

Github repository link:

<https://github.com/Saul12344/mineria-de-datos>

Evidence video link (Youtube):

<https://www.youtube.com/watch?v=WiNdIBujUA4>