

SUBDIRECCIÓN ACADÉMICA

DEPARTAMENTO DE SISTEMAS Y COMPUTACIÓN

ENERO - JUNIO 2020

INGENIERÍA INFORMÁTICA

MATERIA

DATOS MASIVOS

CATEDRÁTICO:

JOSE CHRISTIAN ROMERO HERNANDEZ

UNIDAD 2

# EVALUATION

ALUMNO

López Valencia Luis Daniel

Fernando Ordaz Zamora

**Dataset is imported and declared as movies**

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

movies

**Dataset analysis**

head(movies)#Muestra los primeros datos

**Filter data by gender**

gen <- movies[movies$Genre %in% c("action", "adventure", "animation", "comedy", "drama"),]

**Filter data by recording studio**

stu <- gen[gen$Studio %in% c("Buena Vista Studios", "WB", "Fox", "Universal", "Paramount Pictures"),]

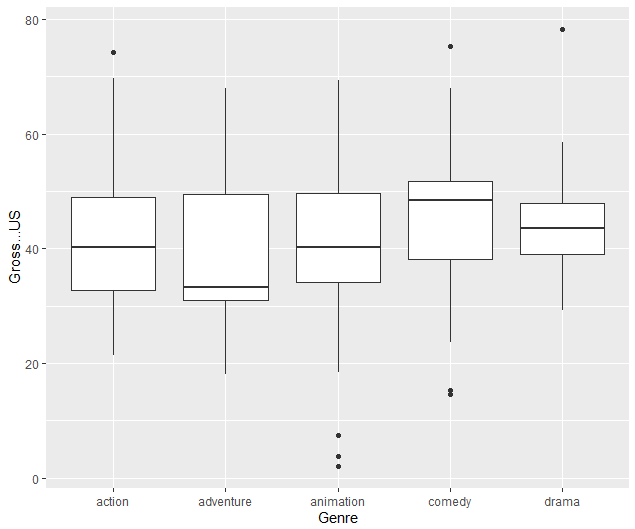
**We use the ggplot2 library to graph**

library(ggplot2)

**We create a variable to save the ggplot structure**

mp <- ggplot(data = stu, aes(x = Genre, y=Gross...US))

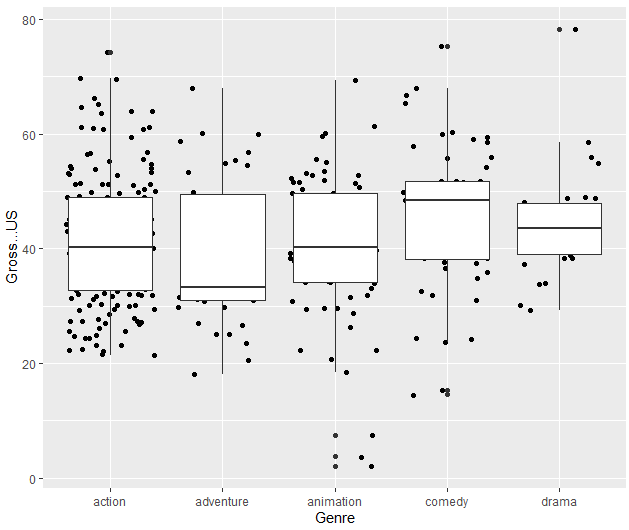
mp +geom\_boxplot()



**Add jitter or noise or fluctuation in the data to scatter the data**

p <- mp + geom\_jitter() + geom\_boxplot()

p

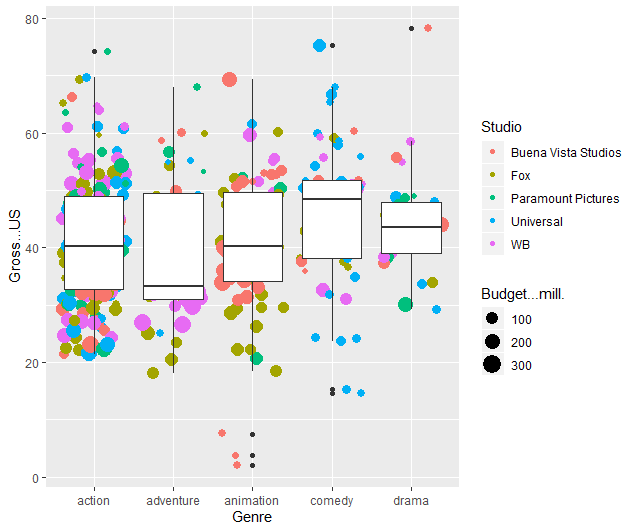


**Adjust the size and color to the study genre**

p <- mp +geom\_jitter(aes(size = Budget...mill., color=Studio)) +

geom\_boxplot()

p

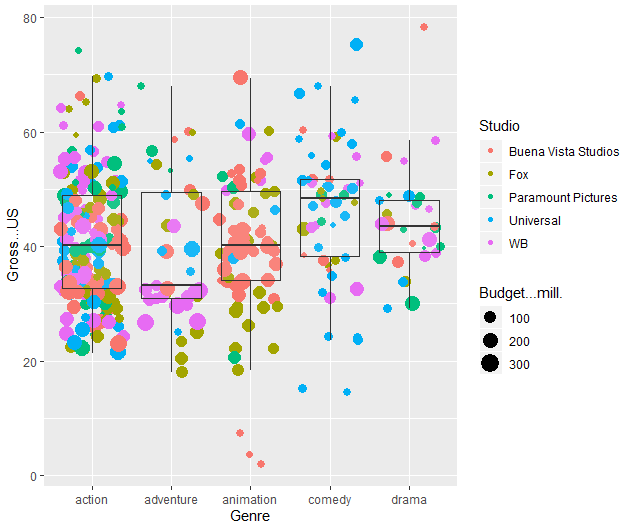


**Black spots removed and boxplot transparent**

p <- mp +geom\_jitter(aes(size = Budget...mill., color=Studio)) +

geom\_boxplot(alpha=0.10, outlier.colour =NA)

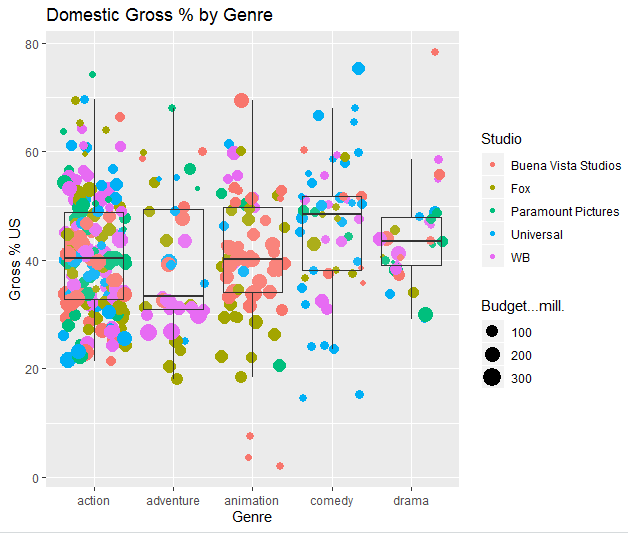
p



**Title added to x, y axes to plot**

p <- p +xlab("Genre") + ylab("Gross % US") + ggtitle("Domestic Gross % by Genre")

p



**Color, size and font are changed**

p <- p + theme(axis.title.x = element\_text(color = "Blue", size = 18) ,

axis.title.y = element\_text(color = "Blue", size = 18) ,

axis.text.x = element\_text(size = 12),

axis.text.y = element\_text(size = 12),

plot.title = element\_text(size = 18),

legend.title = element\_text(size = 18),

text = element\_text(family = "Bodoni 72 Smallcaps"))

p

