
**Instituto Tecnológico de Tijuana**
**Ingeniería en Informática**

**Subject Name:**

**Data Mining**

**Exercise:**

**Evaluative Practice 1 - Unit 2**

**Teacher:**

**Romero Hernandez Jose Christian**

**Students:**

**Castillo Ramirez Guadalupe 17213043**

**Ramos Rivera Manuel Isai 17212931**

Practice 1

Investigate 3 geometry functions that can be added to the ggplot2 () function

we must first import the data

```
getwd()
setwd("C:\Users\Cylon\Desktop\csv\MineriadeDatos\Unidad 2\Practice 1")
getwd()

movies <- read.csv("P2-Movie-Ratings.csv")
```

we rename the columns to make them look more aesthetic

```
colnames(movies) <- c("Film", "Genre", "CriticRating", "AudienceRating", "BudgetMillions", "Year")
head(movies)
```

Convert numeric data to a factor

```
factor(movies$Year)
movies$Year <- factor(movies$Year)
```

We call the library that we are going to use

```
library(ggplot2)
```

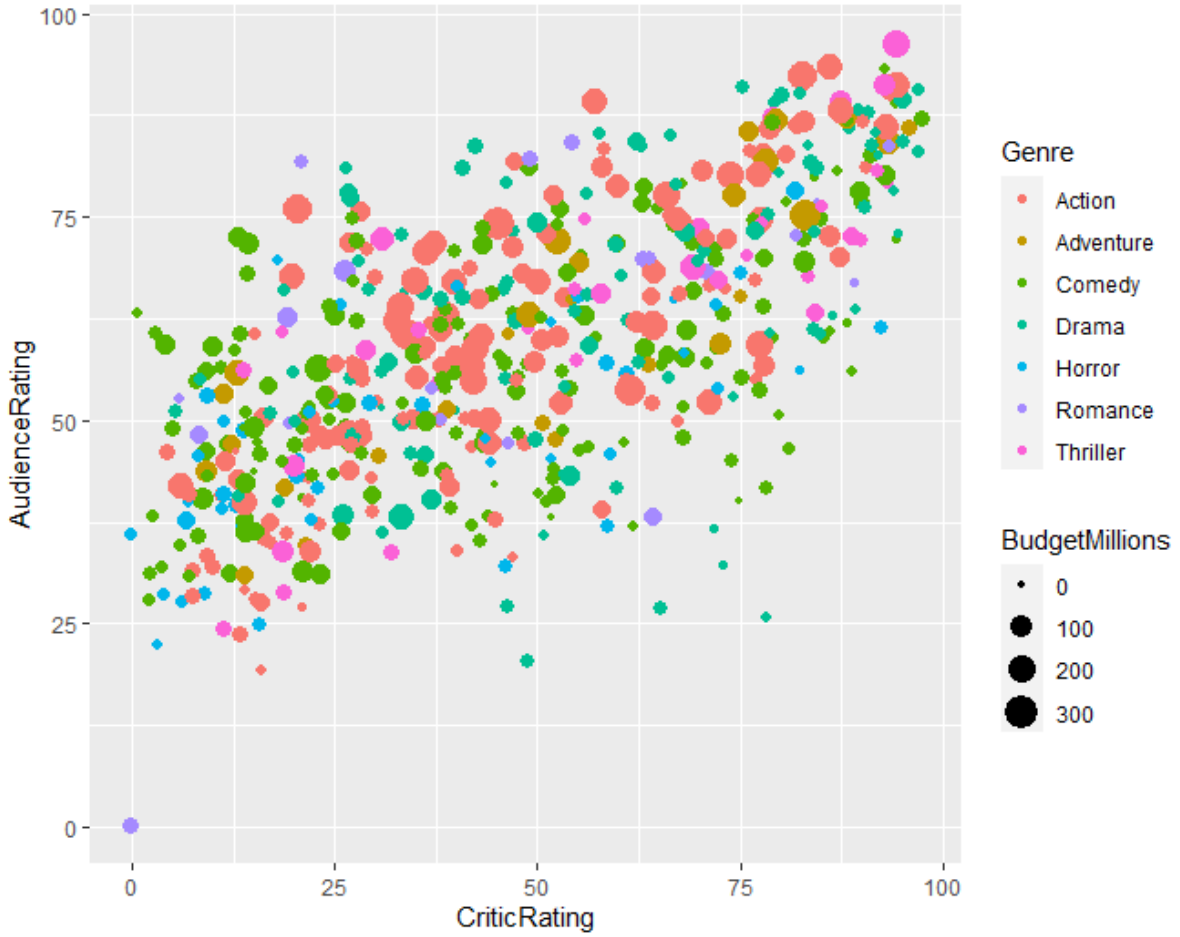
we create our scatter plot with the ggplot function adding color and size

```
ggplot(movies, aes(x=CriticRating, y=AudienceRating,
                   color=Genre, size=BudgetMillions)
```

add geometry

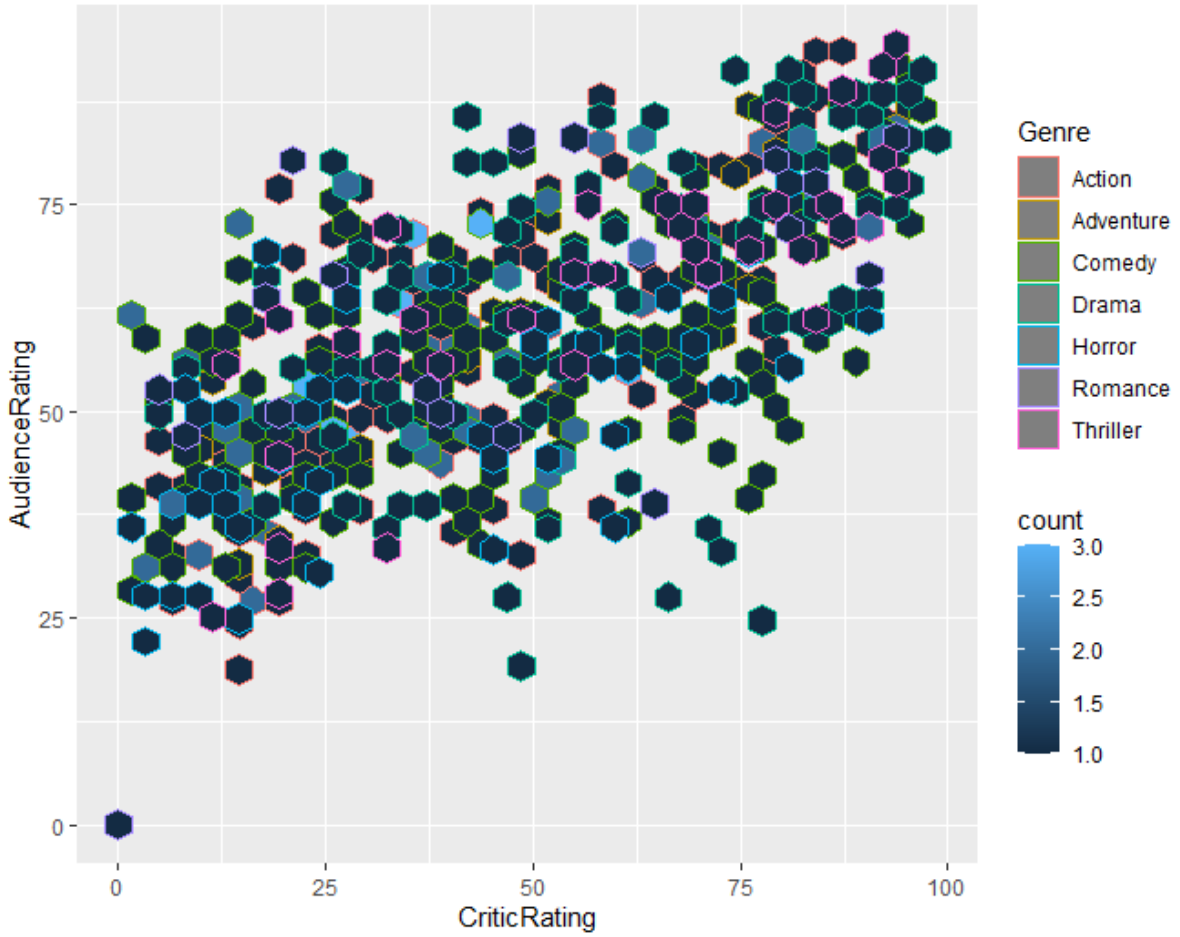
1) geom_Jitter()

```
ggplot(movies, aes(x=CriticRating, y=AudienceRating,
                   color=Genre, size=BudgetMillions)) +
  geom_jitter()
```



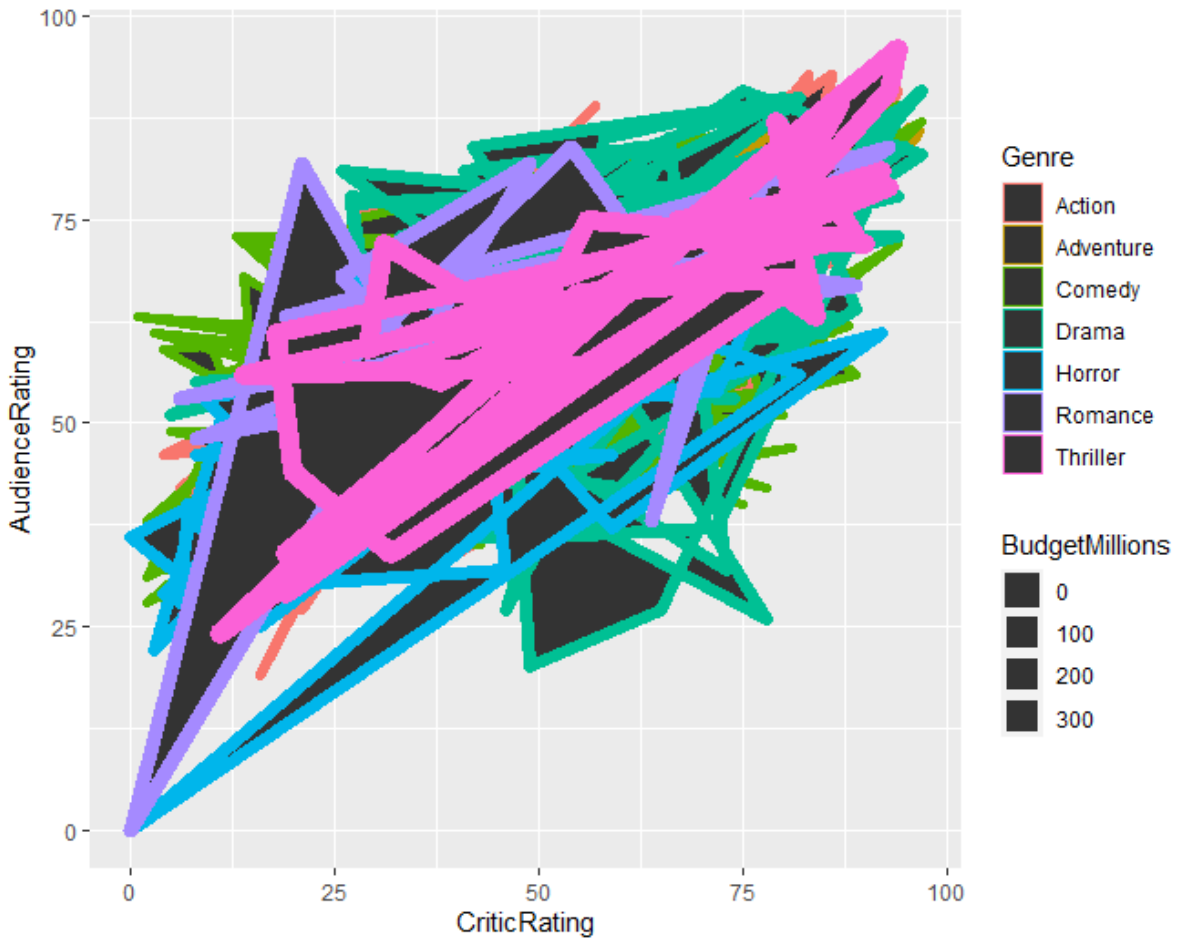

2) geom_hex()

```
ggplot(movies, aes(x=CriticRating, y=AudienceRating,
                   color=Genre, size=BudgetMillions)) +
  geom_hex()
```




3) geom_polygon()

```
ggplot(movies, aes(x=CriticRating, y=AudienceRating,
                   color=Genre, size=BudgetMillions)) +
  geom_polygon()
```


100
75
50
25
0
AudienceRating
0
25
50
75
100
CriticRating
Genre
Action
Adventure
Comedy
Drama
Horror
Romance
Thriller
BudgetMillions
0
100
200
300