

ANÁLISIS ESTADÍSTICO DE VIDEOJUEGOS

EQUIPO 7



¿QUÉ ES STEAM?

Es una plataforma de distribución digital de videojuegos lanzada en 2003.

Actualmente la plataforma cuenta con mas de 34,00 juegos y 95 millones de usuarios.

PROBLEMÁTICA

Tras casi 10 mil videojuegos lanzados año con año en la plataforma Steam, surgen las interrogantes que nos permitan definir cuantos de estos videojuegos logran tener éxito, y cuantos de ellos se juegan día con día. Al mismo tiempo, se busca resolver de forma hipotética la interrogante: ¿Qué tan probable es que se tenga éxito si un videojuego es lanzado en la plataforma Steam?

HIPÓTESIS

El desarrollador tiene una fuerte influencia en el
exito del videojuego

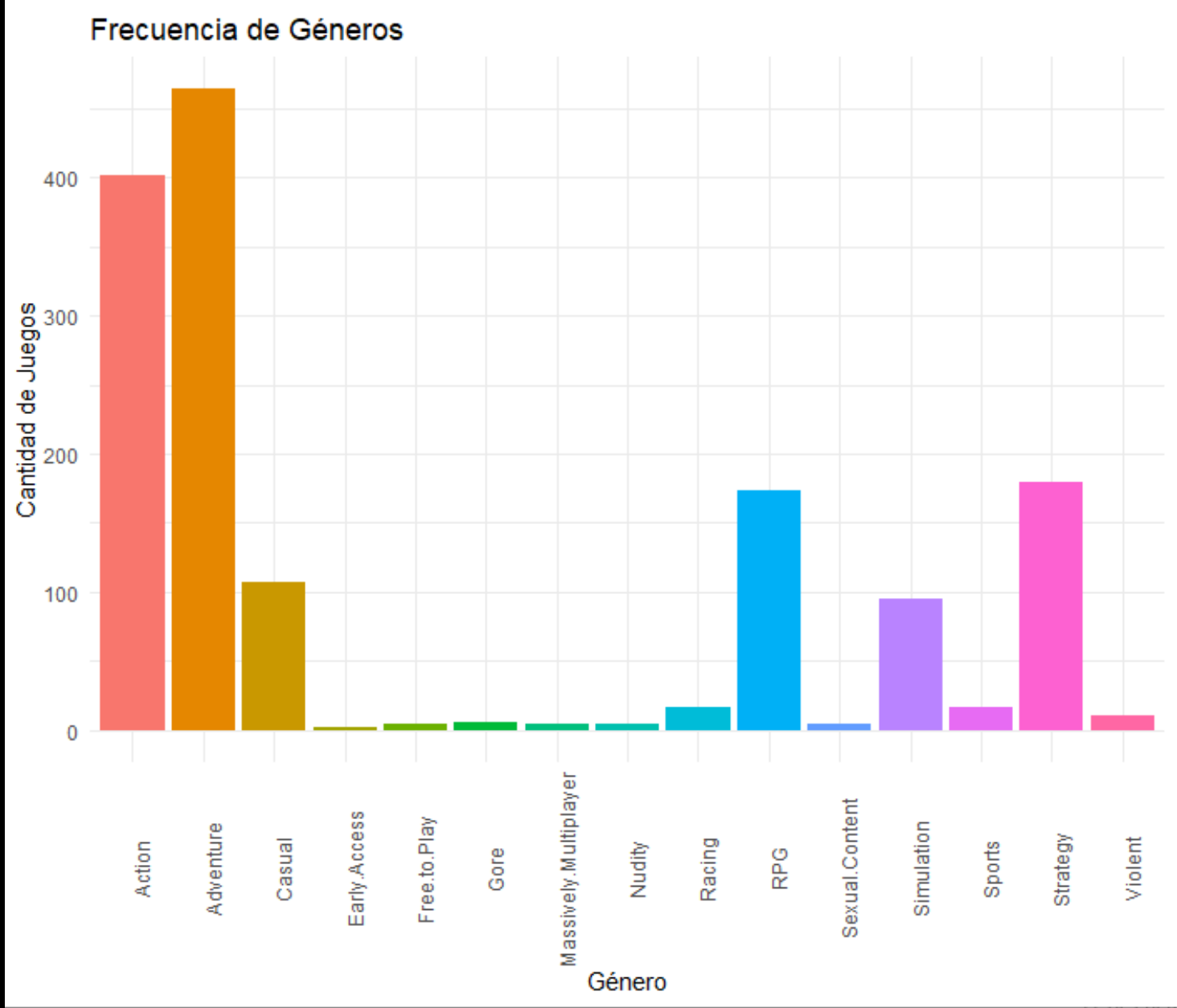
- ¿Cuales son los generos mas implementados?
- ¿Cual es el sistema operativo favorito de los desarrolladores?
- ¿Que videojuego fue el mas/menos jugado?
- ¿Que videojuego fue el mejor/peor valorado?
- ¿Que videojuego tiene el mayor/menor precio?
- ¿El precio influye en su exito?
- ¿Cual ha sido el desarrollador con mas juegos publicados?
- ¿Estos desarrolladores obtienen los mejores puntajes de la critica/usuarios?

INTERROGANTES

EL RETO

- Steam no es una plataforma que tenga abiertos sus datos al público, sin embargo muchos usuarios se encargan de recolectar la información del sitio y publicarlos gratuitamente para uso.
- Nuestro reto fue teniendo tantos datasets disponibles ¿Cual es el dataset adecuado para nuestro análisis?. Ninguno tenía todo lo que queríamos analizar así que tuvimos que trabajar con varios para al final unirlos en uno solo y conservar la mayor parte de información posible

GENERO MAS IMPLEMENTADO

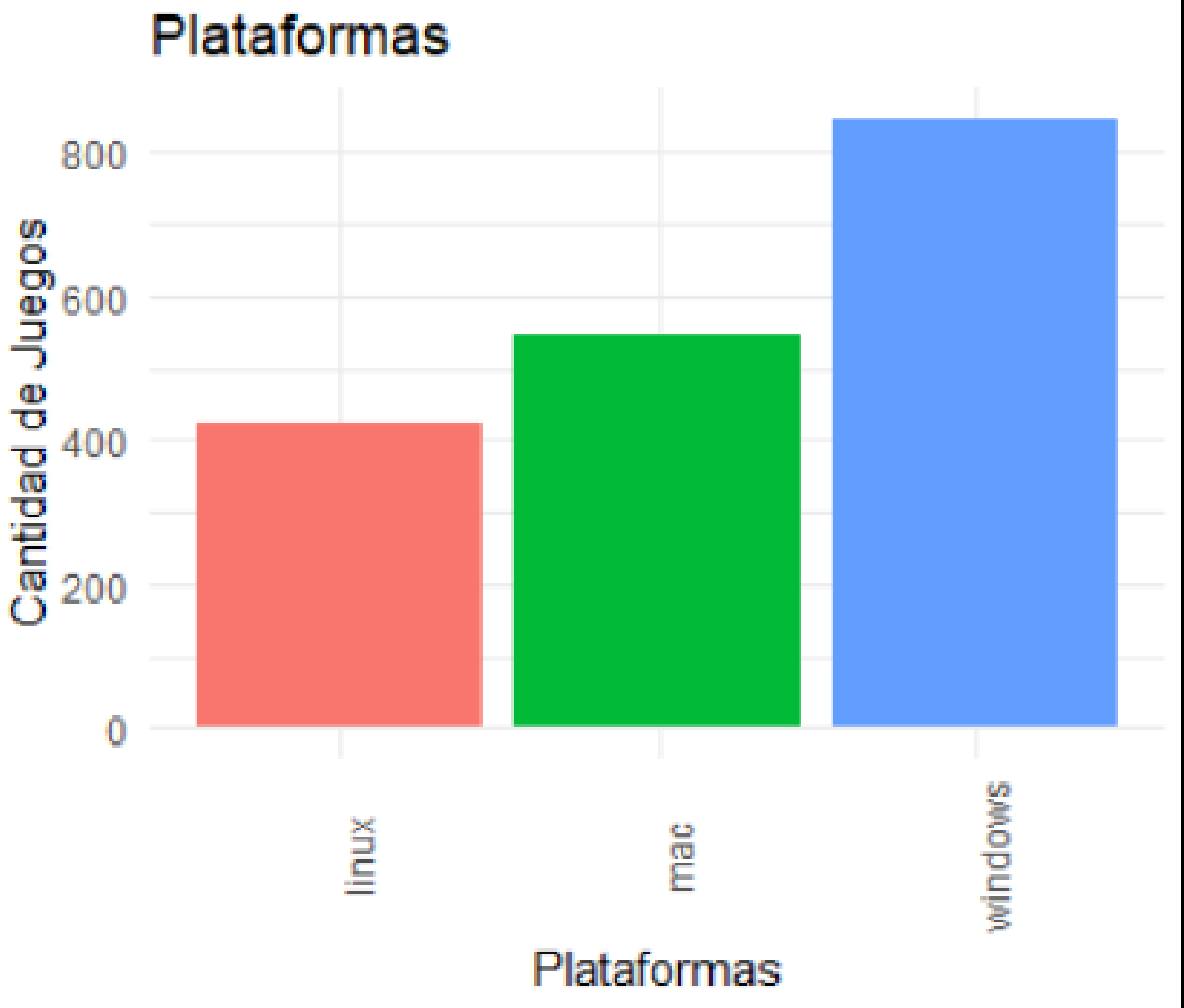


	v1	v2
10	Simulation	95
11	Casual	107
12	RPG	174
13	Strategy	179
14	Action	402
15	Adventure	464

CODIGO

```
> colnames(genres) <- gsub("genre_", "", names(genres))
>
> genres<- as.data.frame(cbind(names(genres), apply(genres, 2, sum)))
> genres<- genres %>%
+   mutate(V2 = as.numeric(V2)) %>%
+   filter(V1!="Indie")
>
> ggplot(genres, aes(V1,V2, fill = V1)) +
+   geom_bar(stat = 'identity') +
+   labs(x = 'Género', y = 'Cantidad de Juegos', title = 'Frecuencia de Géneros') +
+   theme_minimal() +
+   theme(
+     legend.position = "none",
+     axis.text.x = element_text(angle = 90)
+   )
```


SISTEMA FAVORITO DE LOS DESARROLLADORES



	v1	v2
1	linux	422
2	mac	547
3	windows	846

CODIGO

```
> plataforms<- videogames %>% select(starts_with('platform_'))
> colnames(plataforms) <- gsub('platform_', '', names(plataforms))
>
> plataforms<- as.data.frame(cbind(names(plataforms),apply(plataforms, 2, sum)))
> plataforms<- plataforms %>%
+   mutate(V2 = as.numeric(V2)) %>%
+   filter(V2 > 0)
>
> ggplot(plataforms, aes(V1,V2, fill = V1)) +
+   geom_bar(stat = 'identity') +
+   labs(x = 'Plataformas', y = 'Cantidad de Juegos', title = 'Plataformas') +
+   theme_minimal() +
+   theme(
+     legend.position = "none",
+     axis.text.x = element_text(angle = 90)
+   )
> tail(plataforms %>% arrange(V2))
```

EL MAS/MENOS JUGADO

Tomando el promedio de tiempo jugado como referencia

Menos Jugado // Praetorians

Más Jugado // Duelyst



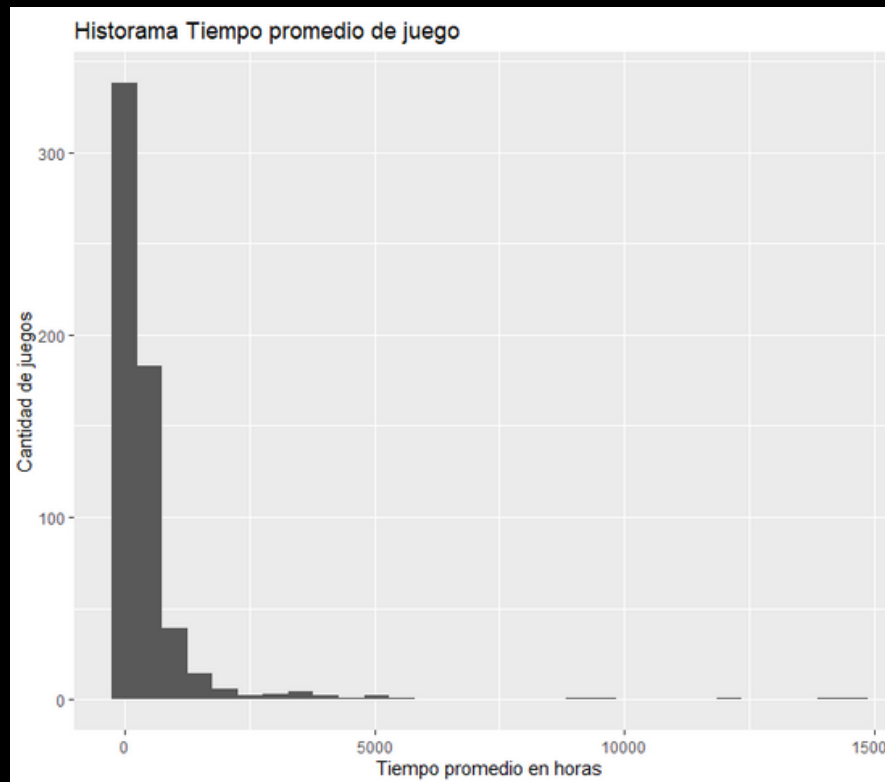
EL MAS/MENOS JUGADO

```
-----Entrada-----
#Menos jugado tomando el promedio
> menos.jugado.a <- videogames2[which.min(videogames2$average_playtime), columns]
> menos.jugado.a

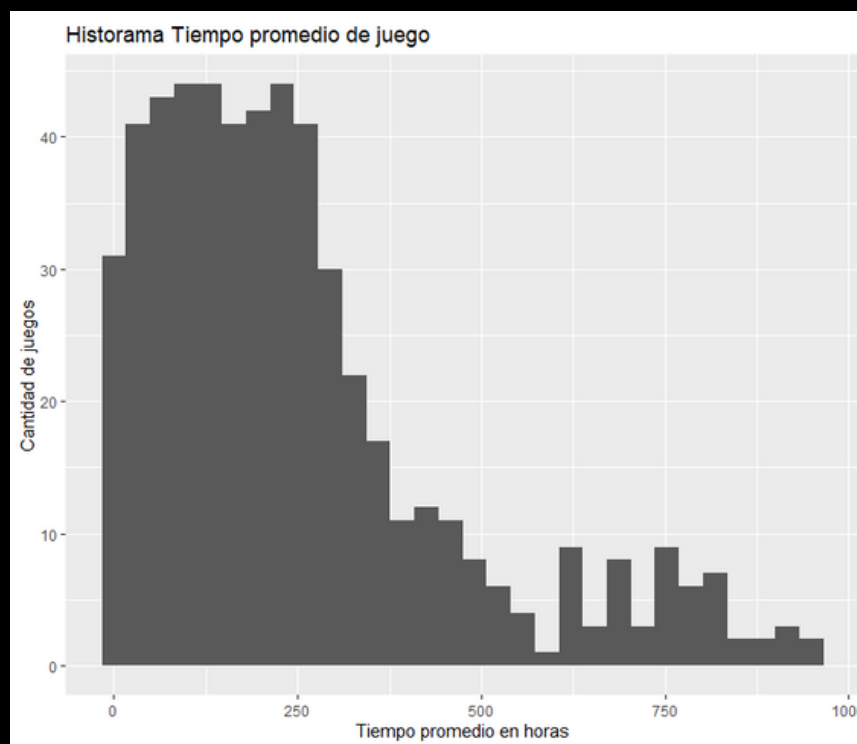
> #Mas jugado tomando el promedio
> mas.jugado.a <- videogames2[which.max(videogames2$average_playtime), columns]
> mas.jugado.a

-----Salida-----
  name release_date average_playtime median_playtime Meta.Score User.Score
228 Praetorians    2014-04-11
  name release_date average_playtime median_playtime Meta.Score User.Score
464 Duelyst      2016-08-23      14620      28847      8.2      7.7
      1              1          7.8          8.2
```

GRÁFICAS



Gráficar sin ajustar



Grafica ajustada

Promedio de horas jugadas por juego: 250 horas

EL MEJOR/PEOR VALORADO

Mejor valorado por la critica

Undertale// 9.2

UNDERTALE™

Peor valorado por la crítica

Zombeer// 3.2



Mejor Valorado por el público

To The Moon// 9



Peor Valorado por el público

Corpse of Discovery // 1.3



CODIGO

```
-----Entrada-----
#Juego peor valorado por la crítica
> juego.peor.v.c<-videogames[which.min(filter(videogames,Meta.Score>0)$Meta.Score), columns]
> juego.peor.v.c

> #Juego peor valorado por los usuarios
> juego.peor.v.u<-videogames[which.min(filter(videogames,User.Score>0)$User.Score), columns]
> juego.peor.v.u

> #Juego mejor valorado por la crítica
> juego.mejor.v.c<-videogames[which.max(filter(videogames,Meta.Score>0)$Meta.Score), columns]
> juego.mejor.v.c

> #Mejor valorado por los usuarios
> juego.mejor.v.u<-videogames[which.max(filter(videogames,User.Score>0)$User.Score), columns]
> juego.mejor.v.u

-----Salida-----

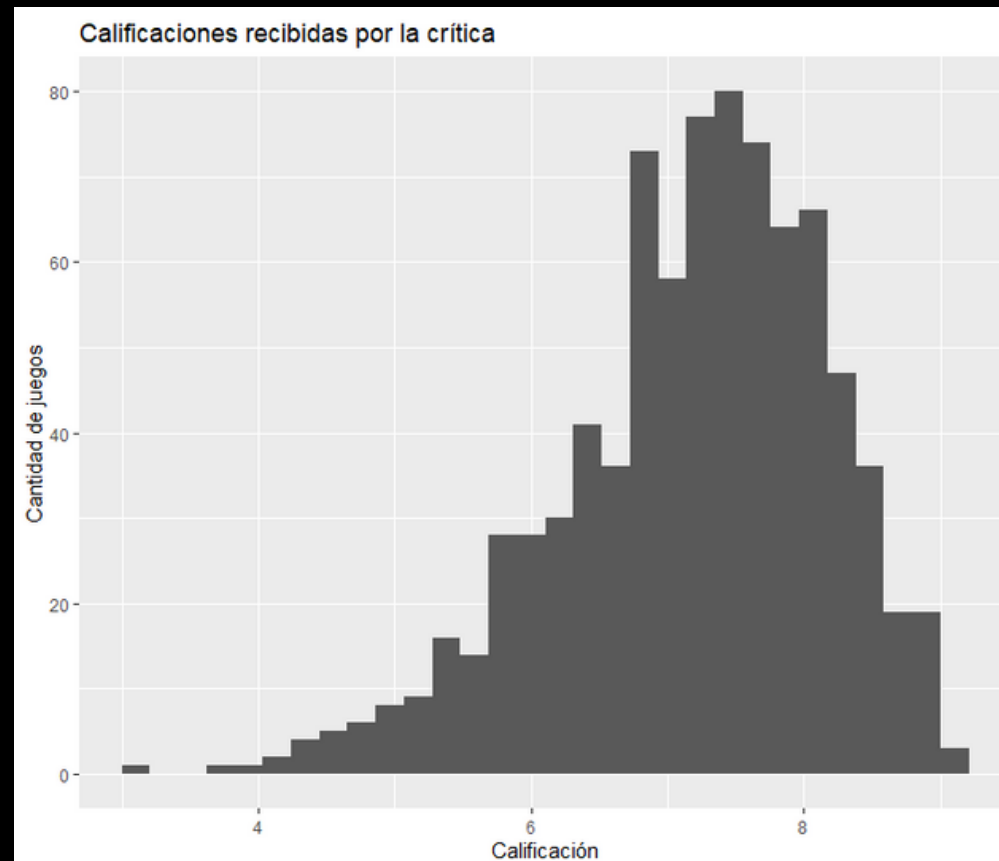
  name Meta.Score User.Score release_date
369 Zombeer      3.2         5    2015-01-30

  name Meta.Score User.Score release_date
471 Corpse of Discovery 4.9         1.3    2015-08-25

  name Meta.Score User.Score release_date
483 Undertale      9.2         8.3    2015-09-15

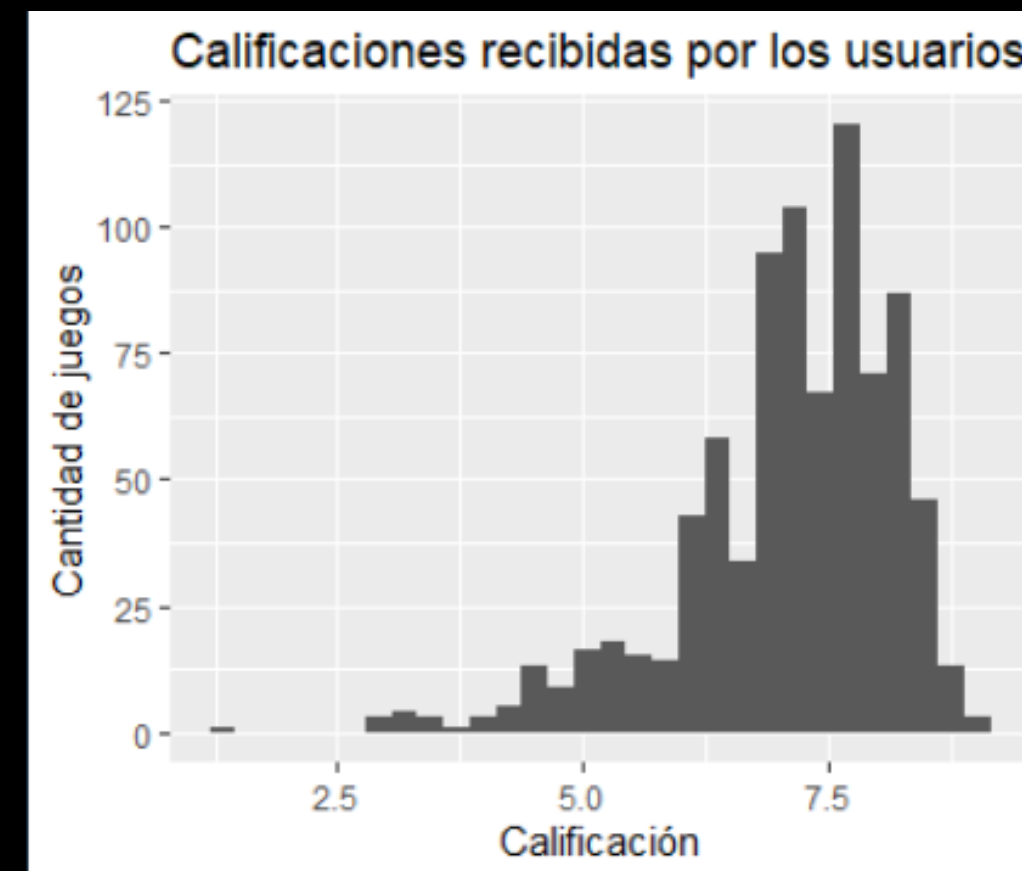
  name Meta.Score User.Score release_date
112 To the Moon      8.1         9     2012-09-07
```

GRÁFICAS

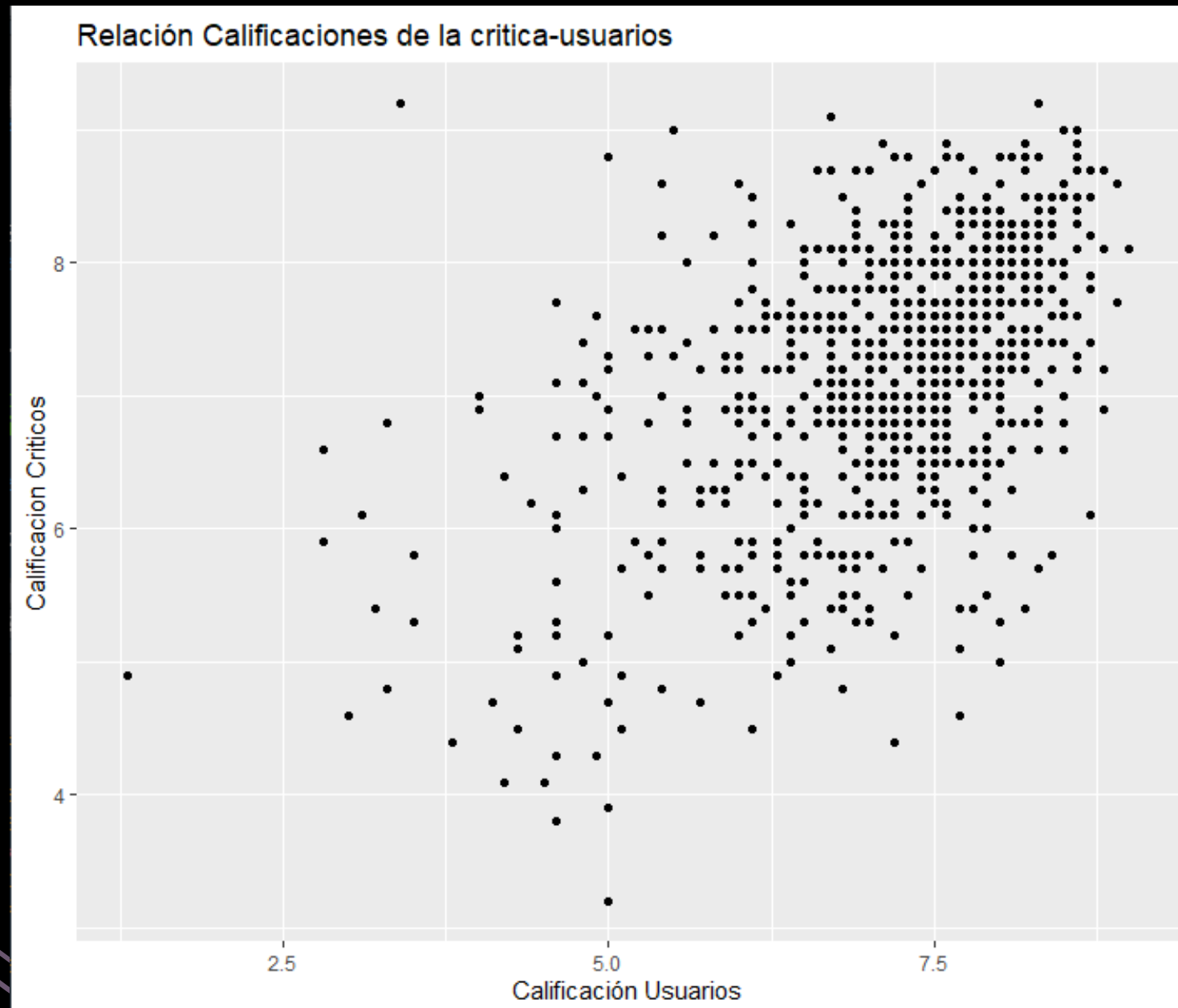


Promedio de calificaciones recibidas: 7.19

Promedio de calificaciones recibidas: 7.17



RELACIÓN CRÍTICAS USUARIO-CRÍTICA



Coeficiente de Correlación: 0.5189

VIDEOJUEGO MAS CARO/BARATO

Mas barato

Waveform// 0.79 USD



Mas caro

ARK: Survival Evolved// 44.99 USD



CODIGO

```
-----Entrada-----
> # # ¿Cual es el videojuego mas caro/barato?
> columns <- c('name', 'release_date', 'price', 'publisher', 'Meta.Score', 'User.Score')
> videogames3<-videogames%>%filter(price>0)
> #Videojuego mas barato
> videogames3[which.min(videogames3$price), columns]

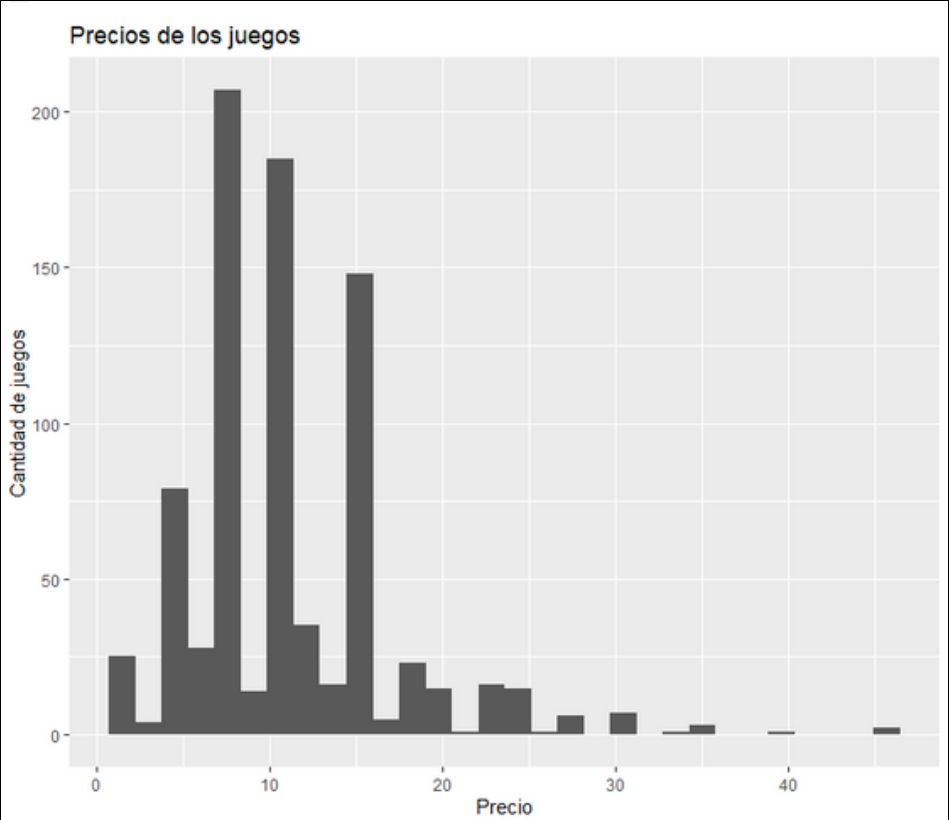
> #videojuego mas caro
> videogames[which.max(videogames$price), columns]

-----Salida-----

  name release_date price publisher Meta.Score User.Score
88 Waveform  2012-03-20  0.79 Eden Industries      8.1      7.5

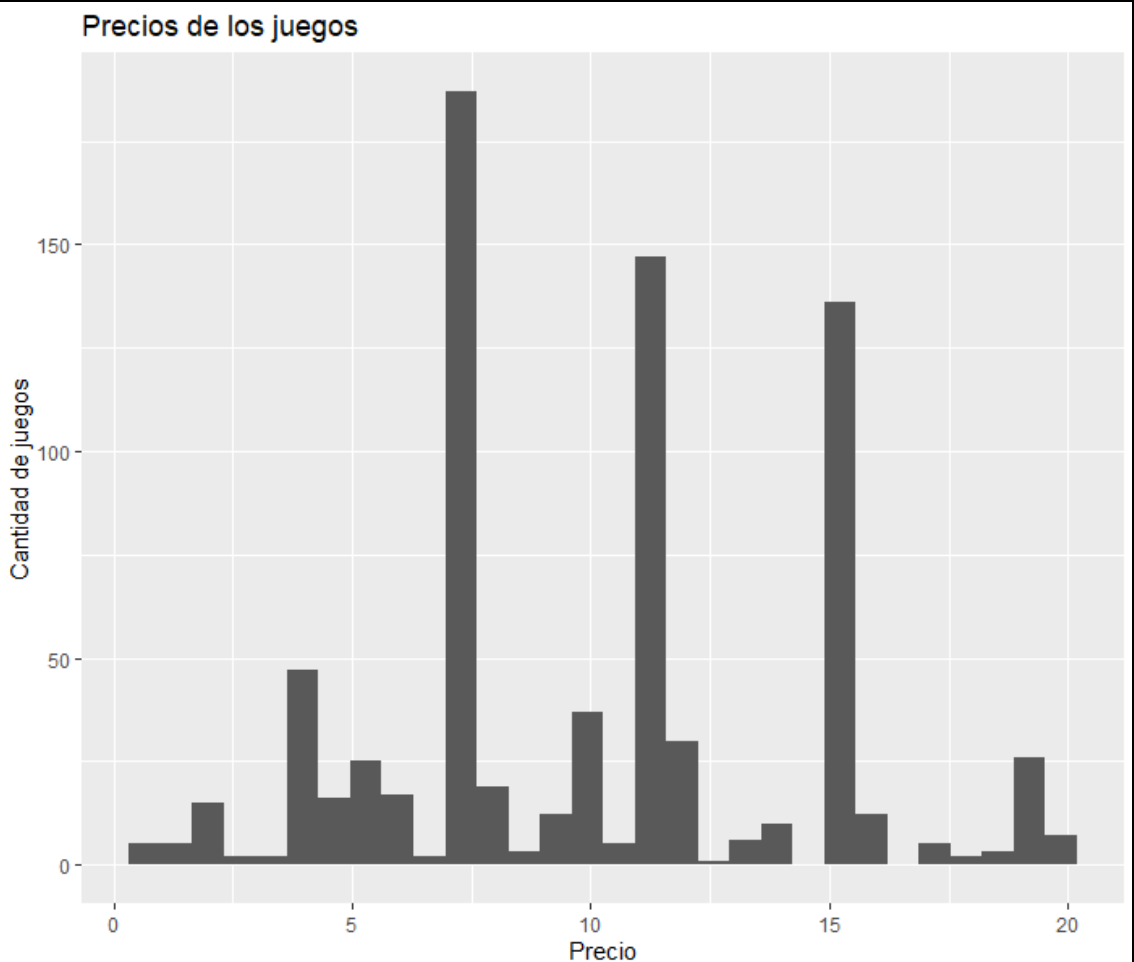
  name release_date price publisher Meta.Score User.Score
746 ARK: Survival Evolved  2017-08-27 44.99 Studio Wildcard      7      4.9
```

GRÁFICAS

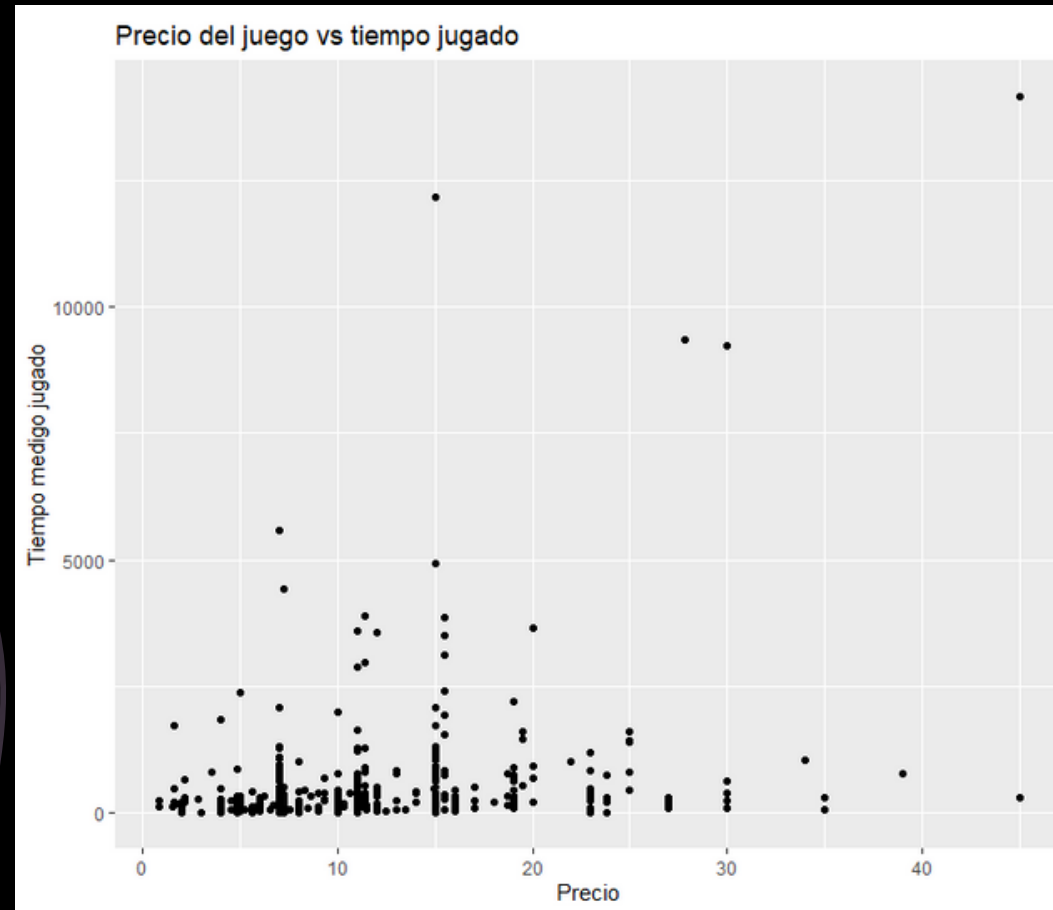


Sin ajustar

Ajustada

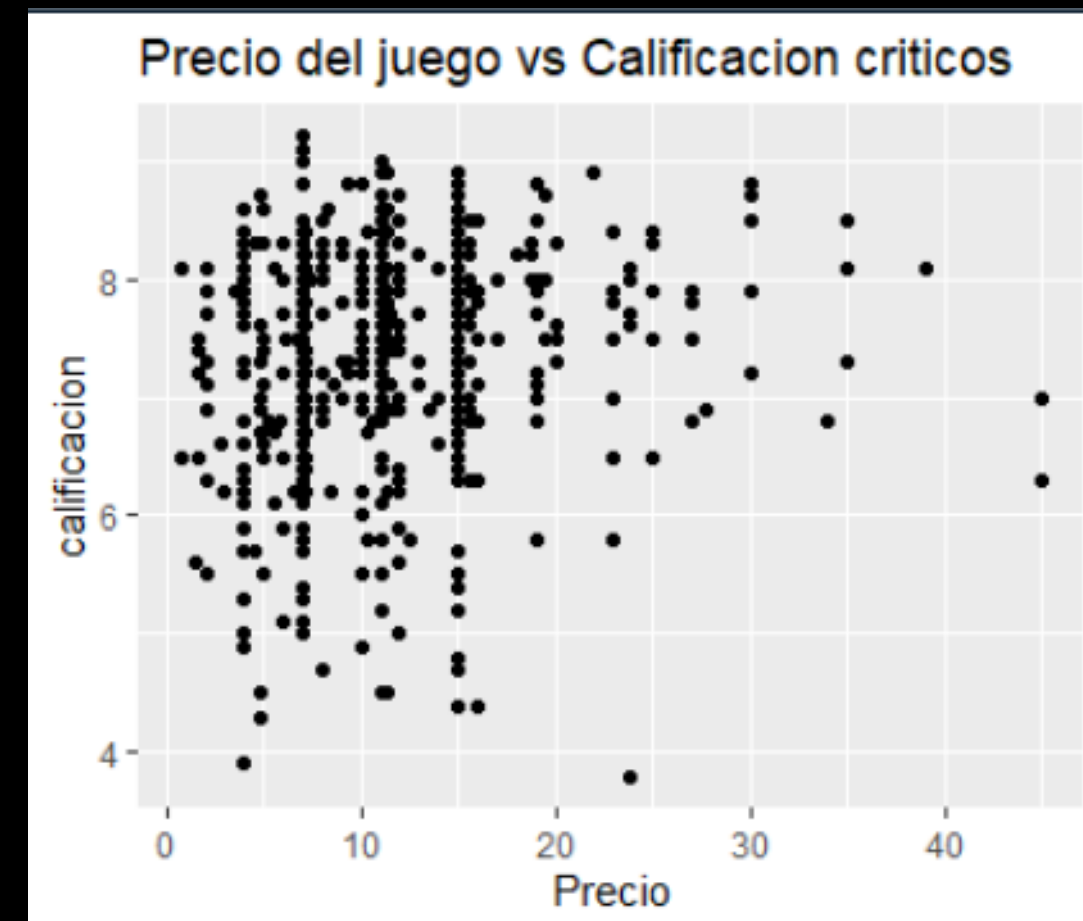


GRÁFICAS



Correlacion: 0.2458

Correlacion:0.1675



DESARROLLADOR CON MAS JUEGOS

Daedalic Entertainment

16 videojuegos publicados



NecoreGames

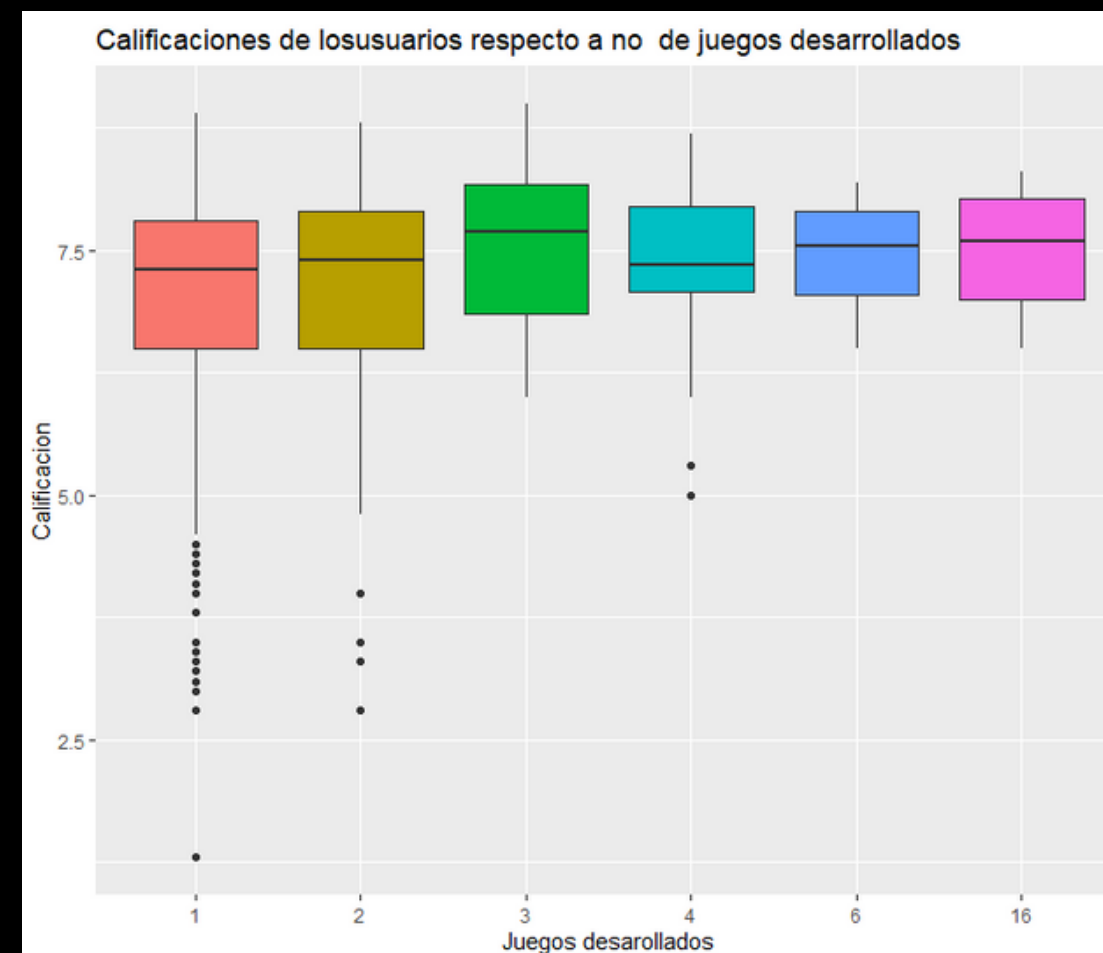
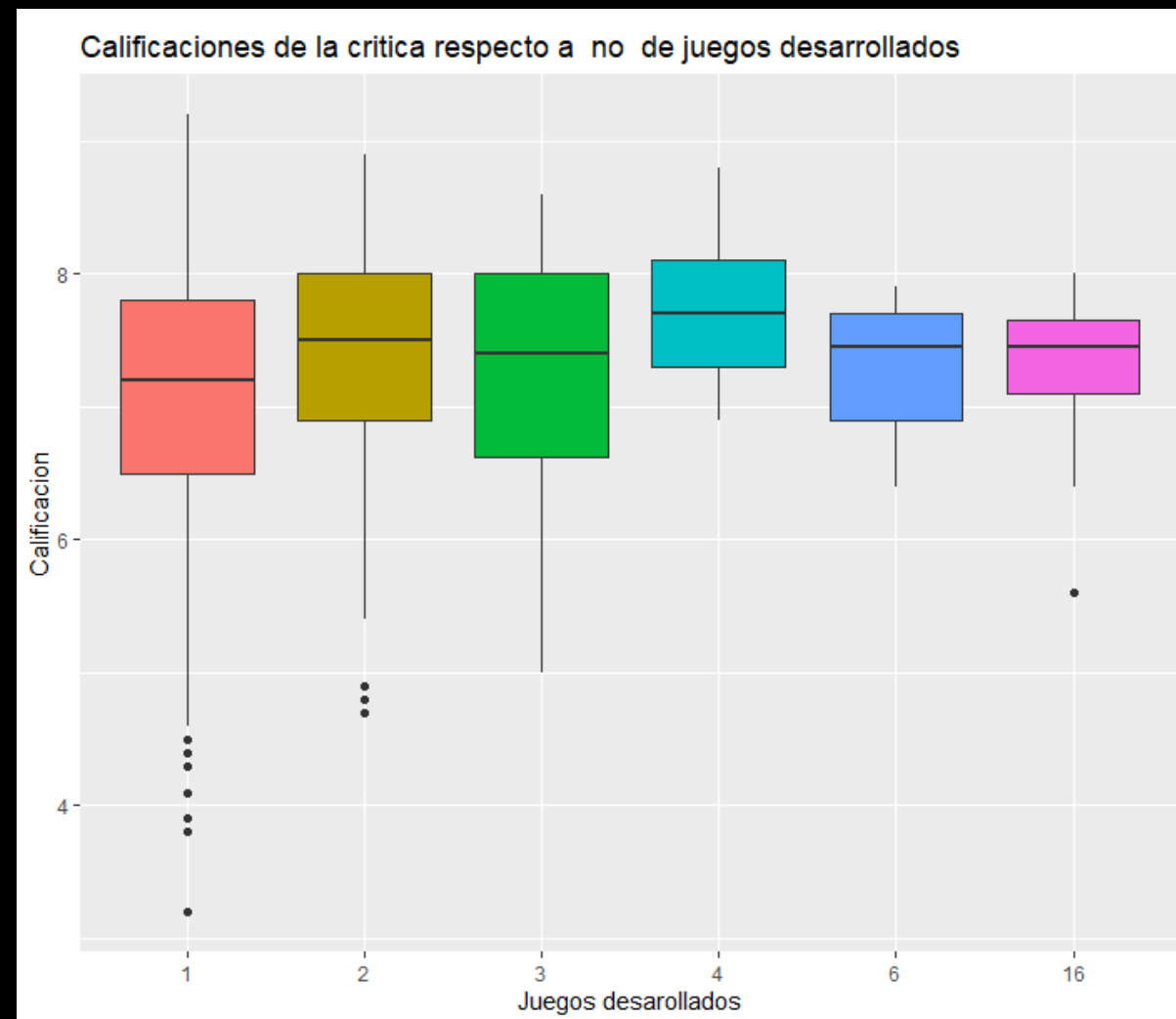
6 juegos publicados



CÓDIGO

```
-----Entrada-----  
> # Desarrolladores con más juegos publicados  
> desarrolladores<-as.data.frame(table(videogames$developer))  
> names(desarrolladores)<-c('developer','frecuencia')  
> #summary(desarrolladores)  
> head(desarrolladores%>%arrange(desc(frecuencia)))  
-----Salida-----  
      developer frecuencia  
1 Daedalic Entertainment      16  
2      NeocoreGames          6  
3      11 bit studios          4  
4          ACE Team           4  
5      Amanita Design          4  
6      Blendo Games           4
```

GRÁFICAS



HALLAZGOS

- Los generos favoritos son la acción y la aventura, a estos generos pertenecen los juegos mejor valorados
- La mayoría de los juegos se desarrollan para windows
- En promedio este tipo de juegos son jugados por alrededor de 250 horas
- La critica y los usuarios no le otorgan una calificación arriba de 8 a la gran mayoría de estos juegos
- La calificacion de los críticos y los usuarios esta muy relacionada, cosa contraria con el tiempo medio de juego donde no parece haber correlación

HALLAZGOS

- La mayoría de los vieojuegos independientes cuesta 8, 12 o 15 dolares
- Existe una ligera tendencia por jugar mas los juegos mas caros
- En general, los estudios que tienen mas juegos desarrollados parecen tener calificaciones mas bajas que los estudios que han desarrollado pocos juegos

¡GRACIAS!

Jesús Ismael Pineda Hernández
Luis Felipe Trujillo Jiménez
Enrique Pérez Rivera