

# Proyecto1

Juan Luis Solórzano (carnet: 201598)

Mica (carnet: XXXXXXXX)

2025-01-20

1. (3 puntos ) Haga una exploración rápida de sus datos, para eso haga un resumen de su conjunto de datos.

```
## Warning: NAs introduced by coercion
```

```
## Warning: NAs introduced by coercion
```

```
##           id           budget           genres           homePage
## Min.      :      5   Min.      :      0   Length:3804   Length:3804
## 1st Qu.: 29807   1st Qu.:      0   Class :character   Class :character
## Median :295644   Median : 3500000   Mode  :character   Mode  :character
## Mean    :291505   Mean    : 24335422
## 3rd Qu.:489998   3rd Qu.: 30000000
## Max.    :922162   Max.    :380000000
```

```
##
## productionCompany productionCompanyCountry productionCountry
## Length:3804       Length:3804             Length:3804
## Class :character   Class :character         Class :character
## Mode  :character   Mode  :character         Mode  :character
```

```
##
```

```
##
```

```
##
```

```
##
```

```
##           revenue           runtime           video           director
## Min.      :0.000e+00   Min.      : 0.0   Mode :logical   Length:3804
## 1st Qu.:0.000e+00   1st Qu.: 91.0   FALSE:3781     Class :character
## Median :3.982e+06   Median :102.0   TRUE :23       Mode  :character
## Mean    :7.859e+07   Mean    :103.3
## 3rd Qu.:7.009e+07   3rd Qu.:116.0
## Max.    :2.847e+09   Max.    :400.0
```

```
##
```

```
##           actors           actorsPopularity           actorsCharacter           originalTitle
## Length:3804       Length:3804             Length:3804             Length:3804
## Class :character   Class :character         Class :character         Class :character
## Mode  :character   Mode  :character         Mode  :character         Mode  :character
```

```
##
```

```
##
```

```
##
```

```
##
```

```
##           title           originalLanguage           popularity           releaseDate
## Length:3804       Length:3804             Min.      :      6.781   Length:3804
## Class :character   Class :character         1st Qu.: 15.774     Class :character
## Mode  :character   Mode  :character         Median : 24.951     Mode  :character
##                                     Mean    : 68.185
```

```

##                               3rd Qu.: 46.337
##                               Max.    :11474.647
##
##      voteAvg      voteCount      genresAmount      productionCoAmount
## Min.   : 1.300   Min.    : 1.0   Min.    : 0.000   Min.    : 0.000
## 1st Qu.: 6.100   1st Qu.: 173.8   1st Qu.: 2.000   1st Qu.: 2.000
## Median : 6.600   Median : 680.5   Median : 3.000   Median : 3.000
## Mean   : 6.616   Mean    : 1871.4   Mean    : 2.664   Mean    : 3.493
## 3rd Qu.: 7.200   3rd Qu.: 2051.0   3rd Qu.: 3.000   3rd Qu.: 5.000
## Max.   :10.000   Max.    :30788.0   Max.    :16.000   Max.    :25.000
##
## productionCountriesAmount  actorsAmount      castWomenAmount  castMenAmount
## Min.   : 0.000             Min.    : 0.0   Min.    : 0   Min.    : 0
## 1st Qu.: 1.000             1st Qu.: 14.0   1st Qu.: 4   1st Qu.: 7
## Median : 1.000             Median : 24.0   Median : 7   Median : 12
## Mean   : 1.708             Mean    : 685.7   Mean    : 3067   Mean    : 11369
## 3rd Qu.: 2.000             3rd Qu.: 41.0   3rd Qu.: 11   3rd Qu.: 21
## Max.   :57.000             Max.    :784594.0   Max.    :922162   Max.    :899405
##                               NA's      :5      NA's      :24

```

2. (5 puntos) Diga el tipo de cada una de las variables (cualitativa ordinal o nominal, cuantitativa continua, cuantitativa discreta)

```

##                               tipo
## id                           cualitativa nominal
## budget                       cuantitativa discreta
## genres                       cualitativa nominal
## homePage                     cualitativa nominal
## productionCompany            cualitativa nominal
## productionCompanyCountry     cualitativa nominal
## productionCountry            cualitativa nominal
## revenue                     cuantitativa discreta
## runtime                     cuantitativa discreta
## video                       cualitativa nominal
## director                   cualitativa nominal
## actors                     cualitativa nominal
## actorsPopularity           cuantitativa continua
## actorsCharacter             cualitativa nominal
## originalTitle               cualitativa nominal
## title                       cualitativa nominal
## originalLanguage            cualitativa nominal
## popularity                  cuantitativa continua
## releaseDate                 cualitativa ordinal
## voteAvg                     cuantitativa continua
## voteCount                   cuantitativa discreta
## genresAmount                cuantitativa discreta
## productionCoAmount          cuantitativa discreta
## productionCountriesAmount   cuantitativa discreta
## actorsAmount                cuantitativa discreta
## castWomenAmount             cuantitativa discreta
## castMenAmount               cuantitativa discreta

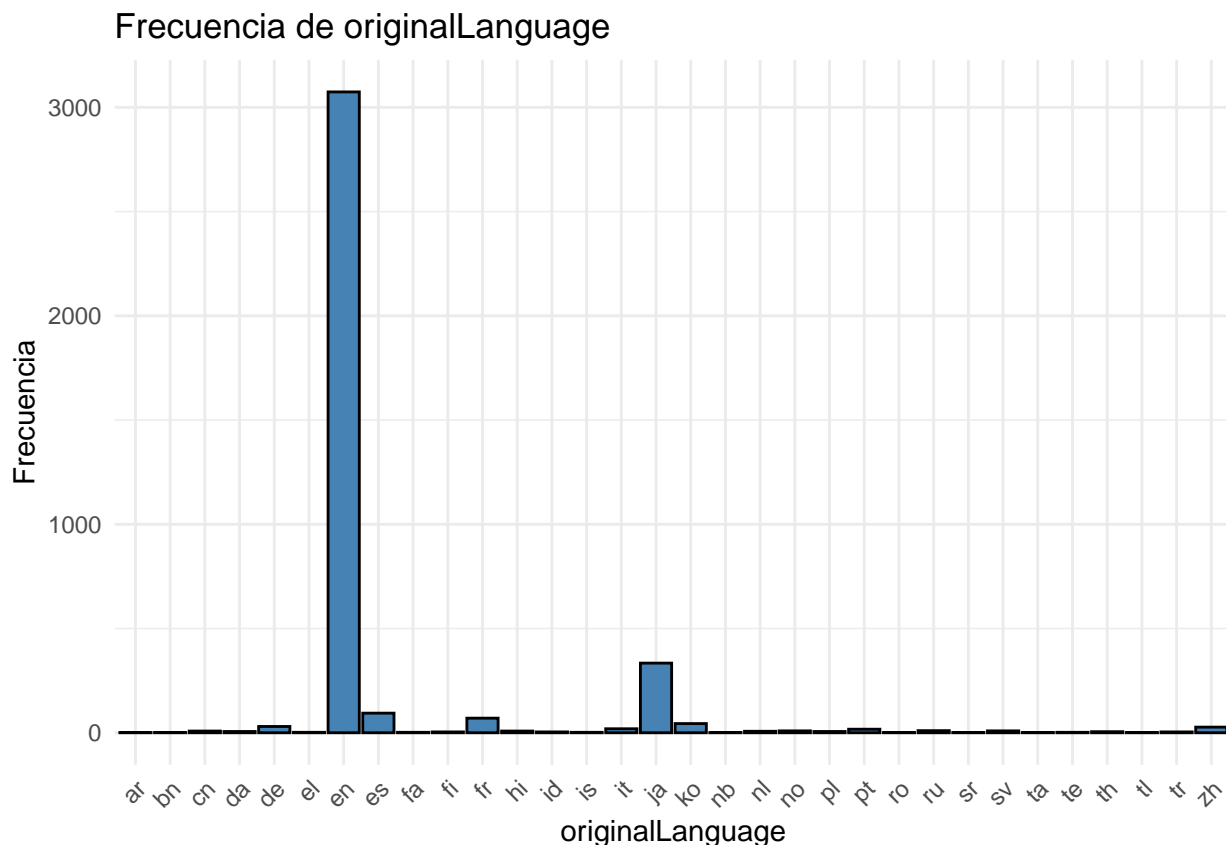
```

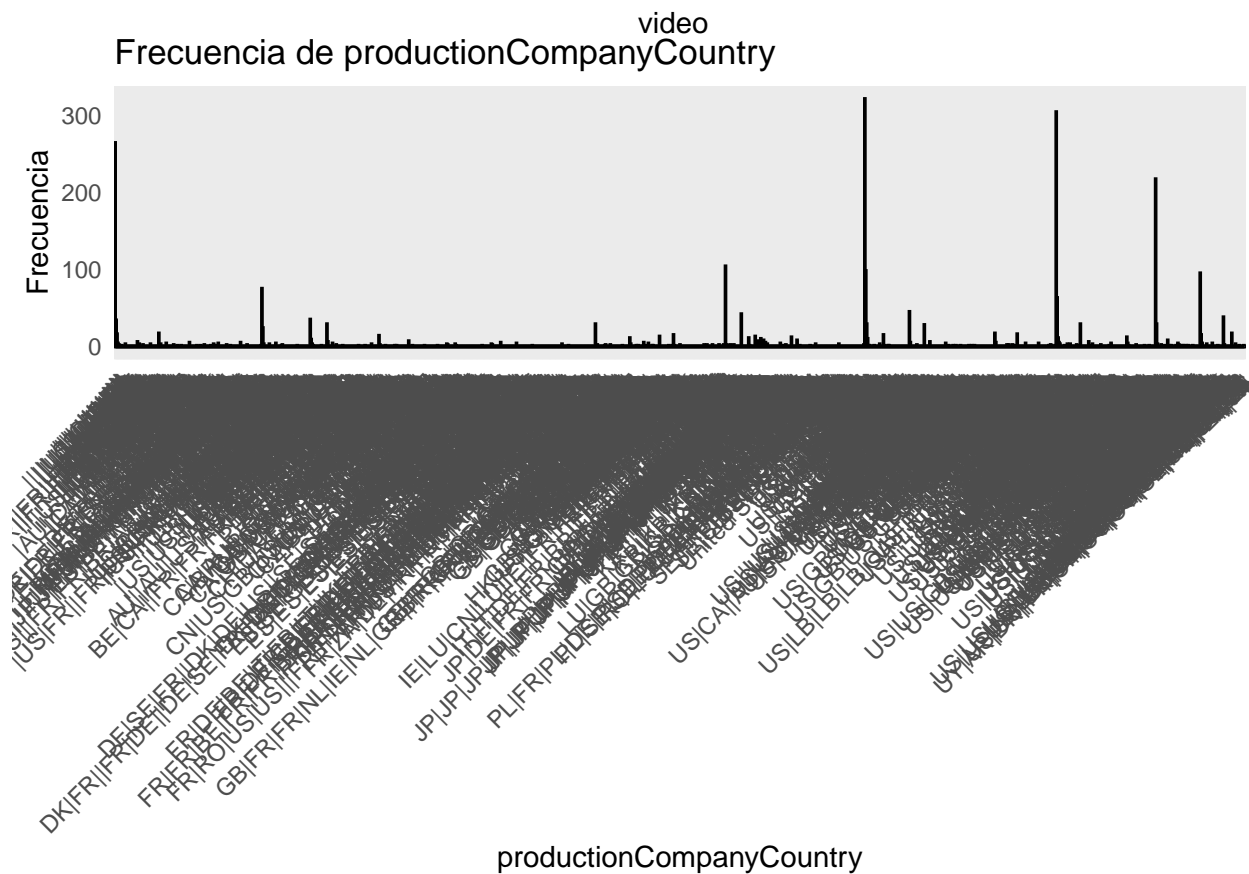
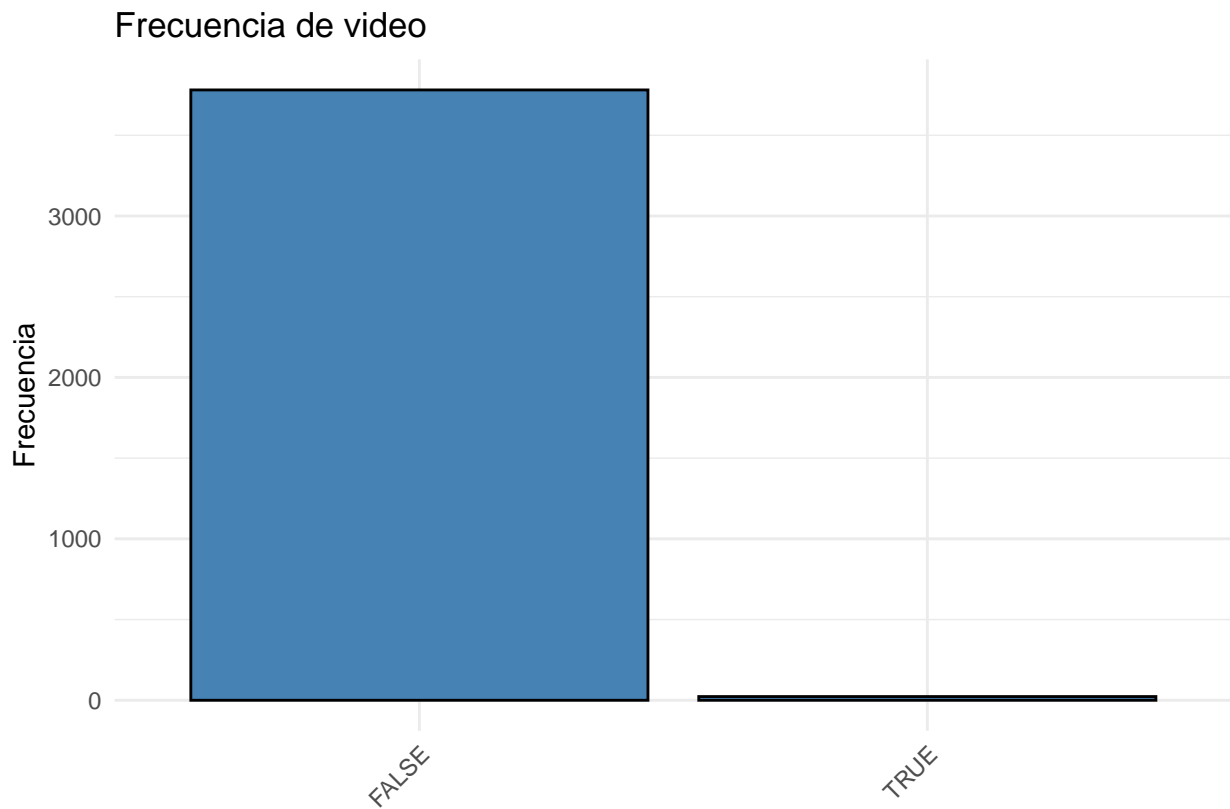
3. (6 puntos) Investigue si las variables cuantitativas siguen una distribución normal y haga una tabla de frecuencias de las variables cualitativas. Explique todos los resultados.

##	Variable	P_value
## voteAvg	voteAvg	1.494879e-16
## voteCount	voteCount	3.700000e-24
## genresAmount	genresAmount	3.700000e-24
## productionCoAmount	productionCoAmount	3.700000e-24
## productionCountriesAmount	productionCountriesAmount	3.700000e-24
## actorsAmount	actorsAmount	3.700000e-24
## castWomenAmount	castWomenAmount	3.700000e-24
## castMenAmount	castMenAmount	3.700000e-24

Como se puede ver en la tabla de resumen de la prueba de Anderson-Darlin como el  $p - valor < 0.01$  para cada variable se puede concluir que ninguna variable sigue una distribución normal de estos datos.

```
## Warning: `aes_string()` was deprecated in ggplot2 3.0.0.
## i Please use tidy evaluation idioms with `aes()`.
## i See also `vignette("ggplot2-in-packages")` for more information.
## This warning is displayed once every 8 hours.
## Call `lifecycle::last_lifecycle_warnings()` to see where this warning was
## generated.
```





Las variables cualitativas cullas tablas no se muestran se debe a que son nombres de actores o cualquier otra

que tiene muchos posibles valores, al punto de ser únicas para cada película. Dicho eso podemos notar que el idioma original mas frecuente en estos datos es el ingles, seguido muy de lejos por el japones y el francés.

#### 4. Responda las siguientes preguntas:

##### 4.1. (3 puntos) ¿Cuáles son las 10 películas que contaron con más presupuesto?

Las 10 peliculas con mayor presupuesto fueron:

##	originalTitle	budget
## 717	Pirates of the Caribbean: On Stranger Tides	380000000
## 4711	Avengers: Age of Ultron	365000000
## 5953	Avengers: Endgame	356000000
## 164	Pirates of the Caribbean: At World's End	300000000
## 5954	Avengers: Infinity War	300000000
## 608	Superman Returns	270000000
## 7135	The Lion King	260000000
## 281	Spider-Man 3	258000000
## 412	Harry Potter and the Half-Blood Prince	250000000
## 2509	Harry Potter and the Deathly Hallows: Part 1	250000000

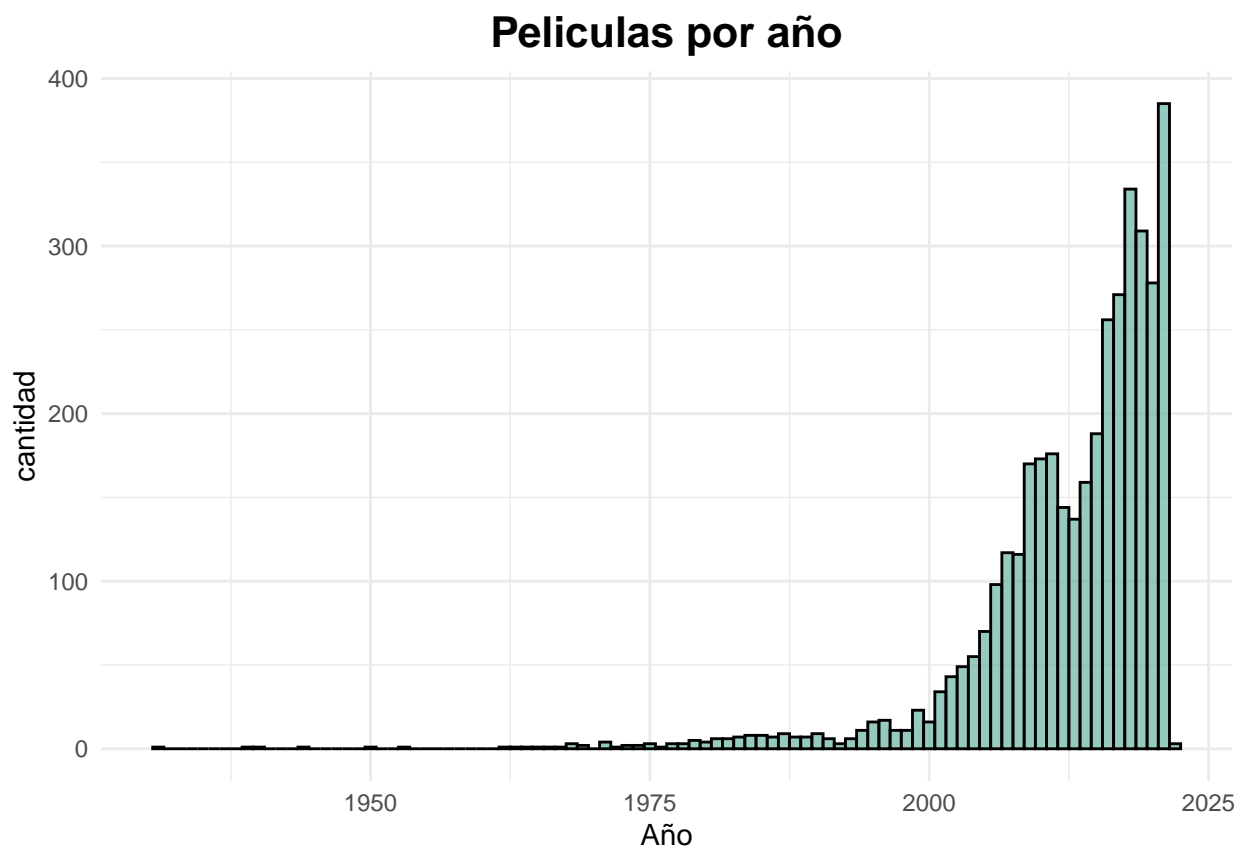
##### 4.2. (3 puntos) ¿Cuáles son las 10 películas que más ingresos tuvieron?

##### 4.3. (3 puntos) ¿Cuál es la película que más votos tuvo?

La película que más votos tuvo fue Inception.

4.4. (3 puntos) ¿Cuál es la peor película de acuerdo a los votos de todos los usuarios?

4.5. (8 puntos) ¿Cuántas películas se hicieron en cada año? ¿En qué año se hicieron más películas? Haga un gráfico de barras.

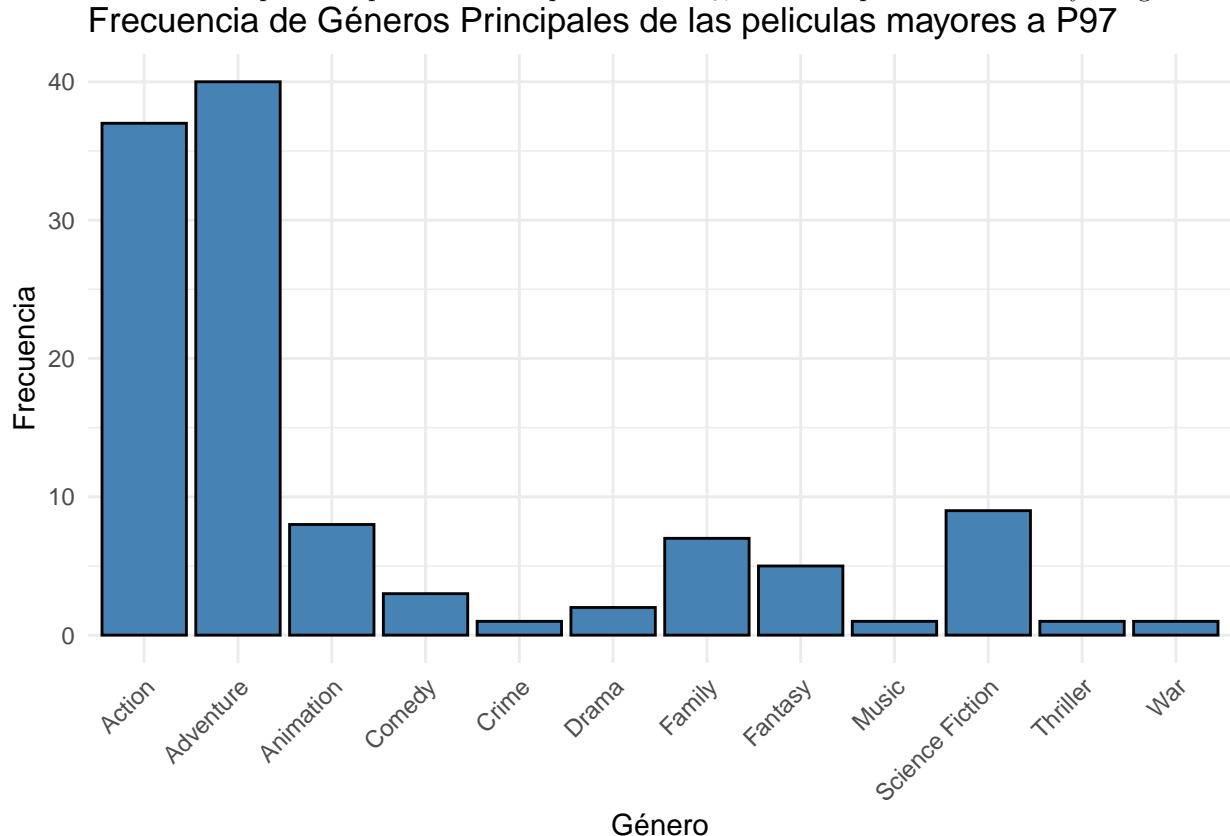


El año que se rodaron más películas fue el 2021

4.6. (9 puntos) ¿Cuál es el género principal de las 20 películas más recientes? ¿Cuál es el género principal que predomina en el conjunto de datos? Representelo usando un gráfico. ¿A qué género principal pertenecen las películas más largas?

4.7. (8 puntos) ¿Las películas de qué género principal obtuvieron mayores ganancias?

Vamos a tomar a las películas por encima del percentil 97  $P_{97}$  como las que obtuvieron mayores ganancias\

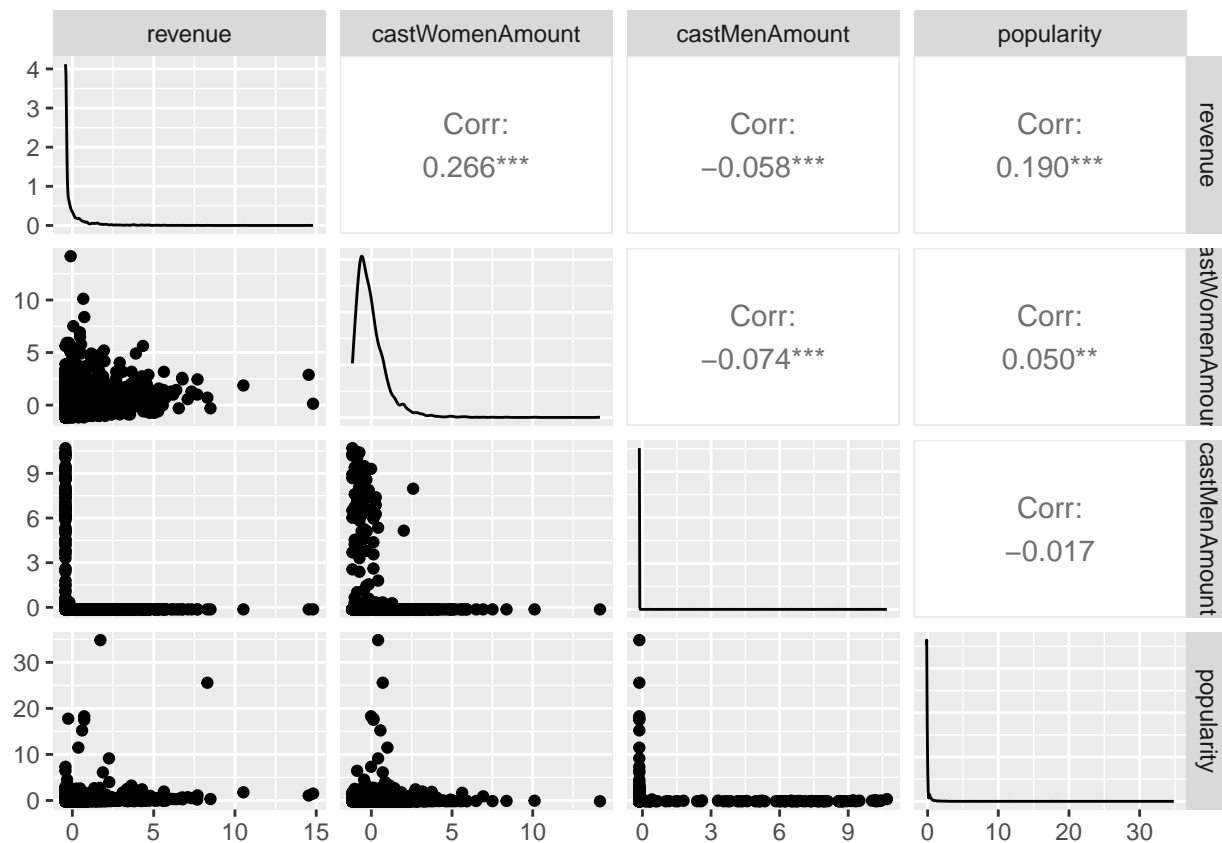


Como se puede apreciar en la gráfica de frecuencias, el genero principal de las películas con mayor ganancia es aventura y acción.

4.8. (3 puntos) ¿La cantidad de actores influye en los ingresos de las películas? ¿Se han hecho películas con más actores en los últimos años?

4.9. (3 puntos) ¿Es posible que la cantidad de hombres y mujeres en el reparto influya en la popularidad y los ingresos de las películas?

```
## 'data.frame':   3780 obs. of  4 variables:
## $ revenue      : num  14.8 14.54 10.52 8.49 8.3 ...
## $ castWomenAmount: num  0.13 2.88 1.867 -0.304 0.709 ...
## $ castMenAmount  : num  -0.137 -0.136 -0.136 -0.137 -0.136 ...
## $ popularity     : num  1.498 1.08 1.773 0.322 25.575 ...
```



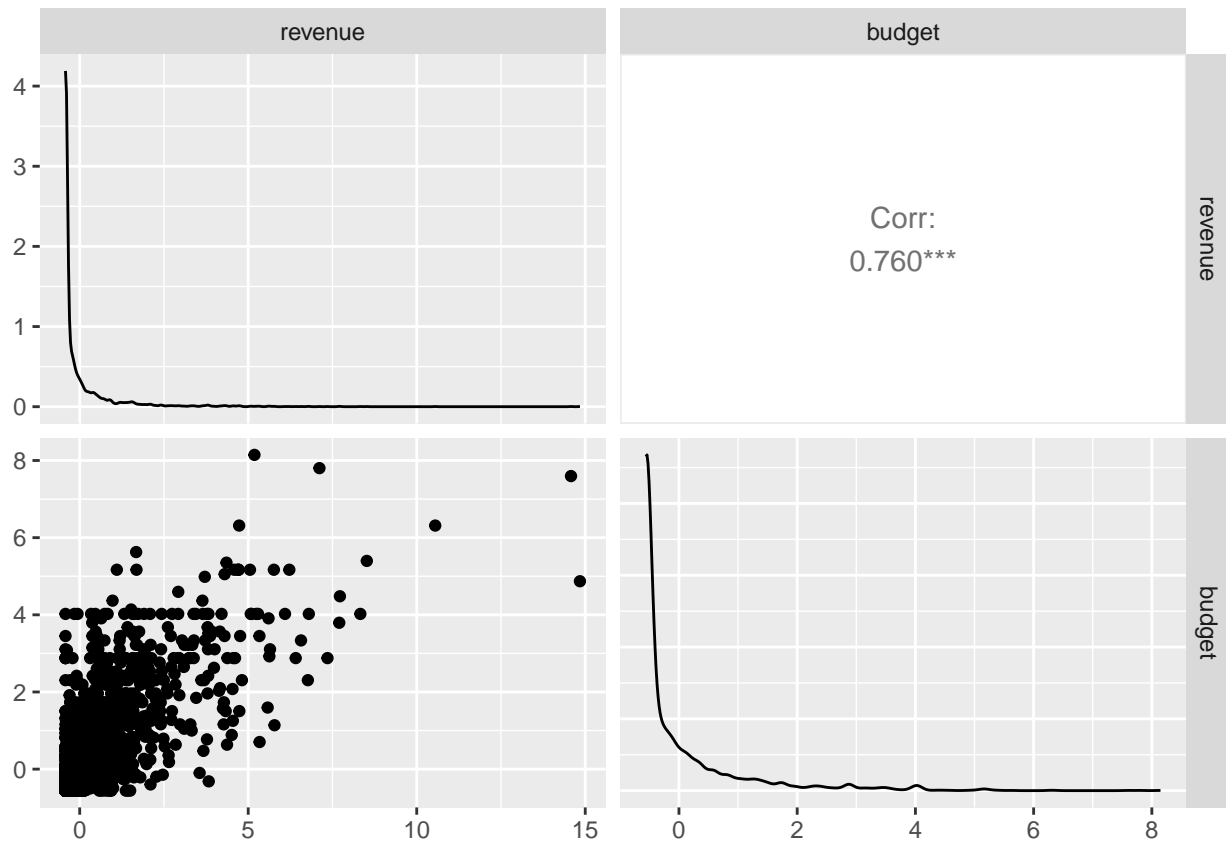
Como el coeficiente d

4.10. (8 puntos) ¿Quiénes son los directores que hicieron las 20 películas mejor calificadas?

4.11. (8 puntos) ¿Cómo se correlacionan los presupuestos con los ingresos? ¿Los altos presupuestos significan altos ingresos? Haga los gráficos que necesite, histograma, diagrama de dispersión.

```
## 'data.frame': 3804 obs. of 2 variables:
## $ revenue: num 14.84 14.58 10.55 8.52 8.33 ...
## $ budget : num 4.87 7.6 6.31 5.4 4.02 ...
```





- 4.12. (5 puntos) ¿Se asocian ciertos meses de lanzamiento con mejores ingresos?
- 4.13. (6 puntos) ¿En qué meses se han visto los lanzamientos con mejores ingresos? ¿Cuántas películas, en promedio, se han lanzado por mes?
- 4.14. (7 puntos) ¿Cómo se correlacionan las calificaciones con el éxito comercial?
- 4.15. (5 puntos) ¿Qué estrategias de marketing, como videos promocionales o páginas oficiales, generan mejores resultados?
- 4.16. (4 puntos) ¿La popularidad del elenco está directamente correlacionada con el éxito de taquilla?
5. (¡10 puntos extras!) Genere usted otras seis preguntas que le parezcan interesantes porque le permitan realizar otras exploraciones y respóndalas. No puede repetir ninguna de las instrucciones anteriores.