

Preguntas_final

Irving, Chuy

2025-01-31

Preguntas

Después de haber concluido la exploración de las variables es momento de empezar a operar con preguntas directas para tener un mayor acercamiento a la base de datos.

1 ¿Cuáles son las 10 películas que contaron con más presupuesto?

##		title	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
## 4954		Justice League	300000000
## 5954		Avengers: Infinity War	300000000
## 608		Superman Returns	270000000
## 3792		Tangled	260000000
## 7135		The Lion King	260000000
## 281		Spider-Man 3	258000000

2. ¿Cuáles son las 10 películas que más ingresos tuvieron?

Table 1: 10 películas con mayores ingresos

title	revenue
Avatar	2847246203
Avengers: Endgame	2797800564
Titanic	2187463944
Star Wars: The Force Awakens	2068223624
Avengers: Infinity War	2046239637
Jurassic World	1671713208
The Lion King	1667635327
Spider-Man: No Way Home	1631853496
The Avengers	1518815515
Furious 7	1515047671

3 ¿Cuál es la película que más votos tuvo?

##	title	voteCount
## 3512	Inception	30788

4. ¿Cuál es la peor película de acuerdo a los votos de todos los usuarios?

No existe una sola película puntuada como la peor, ya que hay 30 películas que comparten la misma puntuación, aún tratando de filtrar por el promedio, la cantidad sigue siendo de 30. A continuación se listan las películas encontradas.

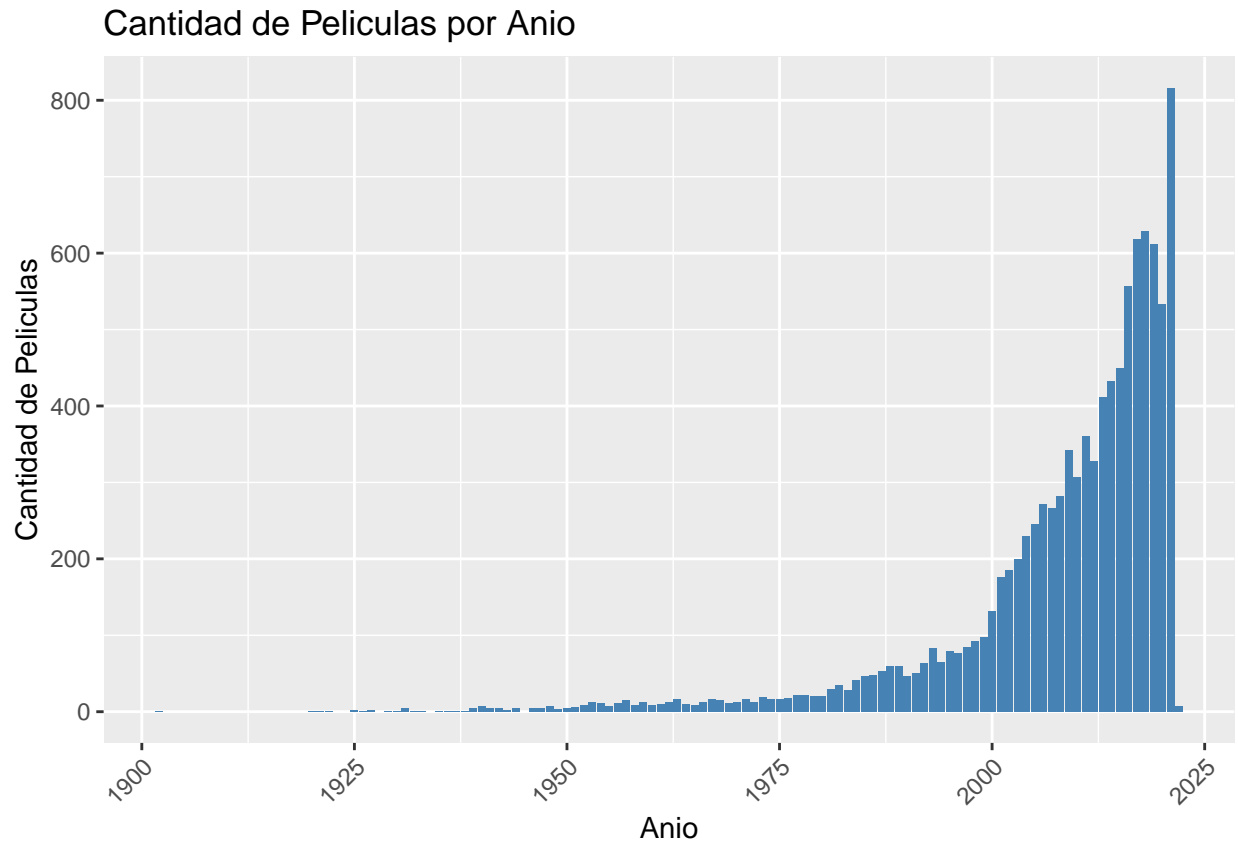
Table 2: Top películas peor

title
El silencio
<U+ADF9><U+C7A5><U+D310> <U+B808><U+C774><U+B514><U+BC84><U+ADF8>: <U+BBF8><U+B
Sing Along
Tar Pits Film
Captain America: Civil War Reenactors
Los cazafantasmas de Petatiux
One Crazy Summer: A Look Back at Gravity Falls
Scream for Summer
La Cantera
Karem, La Posesión
The Real Spider-Man: The Making of The Green Goblin's Last Stand
PorNO
Hot Hair Salon
The Lost Girls
Please Enlighten Me
Five Nights at Freddy's: Home Breach
La Bruja
August Never Ends
Gokuraku sentô: Kyonyû yumomi
The Perfect Wedding
Civil War (Or, Who Do We Think We Are)
Spiderman & Catwoman vs The Komodosarus
Death She Wrote
Kimetsu Orchestra Concert
The Secret Lives of College Freshmen
The Return of La Llorona
El tiempo del no tiempo
Sexo explícito
El sexo me da risa 8
The Next Mega Tsunami

5 ¿Cuántas películas se hicieron en cada año? ¿En qué año se hicieron más películas? Usar un gráfico de barras

```
## # A tibble: 99 x 2
##   Anio Cantidad
##   <dbl>    <int>
## 1  2021      816
## 2  2018      629
## 3  2017      618
## 4  2019      612
## 5  2016      557
```

```
## 6 2020      533
## 7 2015      450
## 8 2014      432
## 9 2013      412
## 10 2011      361
## # i 89 more rows
```



El año en el que se hicieron más películas fue **2021**, con un total de **816** películas.

6. ¿Cuál es el género principal de las 20 películas más recientes?

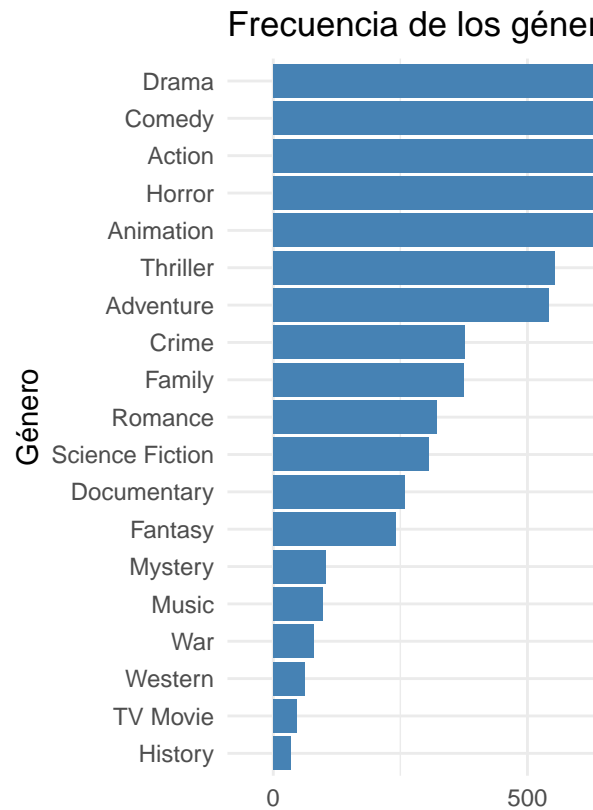
EA continuación se muestra el género principal de cada una de las 20 películas más recientes. Siendo **comedy** el más repetido con 6 incidencias.

Table 3: Películas más recientes con su género principal

title	first_genre
A Shot Through the Wall	Drama
Presque	Comedy
Italian Studies	Drama
American Siege	Action
See for Me	Thriller
WWE Day 1 2022	Action
Harry Potter 20th Anniversary: Return to Hogwarts	Documentary

Gabriel's Rapture: Part One	Romance
Hilda and the Mountain King	Animation
The Kindred	Horror
Death to 2021	Comedy
Lulli	Comedy
American Underdog	Drama
The ExorSIS	Comedy
The Lost Girls	Drama
El Paseo 6	Comedy
Minnal Murali	Action
1000 Miles From Christmas	Comedy

¿Cuál es el género principal que predomina en el conjunto de datos? Representélo usando un gráfico.



El género principal que predomina en todo el conjunto de datos es **drama**.

7 ¿Las películas de qué género principal obtuvieron mayores ganancias?

Es importante mencionar que tomamos como género principal al primer género listado en la columna genres.

```
## # A tibble: 20 x 2
##   genero_principal ganancias_totales
##   <chr>                <dbl>
## 1 Action                140936671043
```

```
## 2 Adventure      86313291491
## 3 Comedy         72990070028
## 4 Drama          66415119599
## 5 Animation      44193668686
## 6 Family         27070466981
## 7 Science Fiction 25767021939
## 8 Horror         23483468549
## 9 Fantasy        22309474780
## 10 Thriller      17111426999
## 11 Crime         13764408000
## 12 Romance       9582546901
## 13 War           5151630856
## 14 Mystery       4640756330
## 15 Music         3493810817
## 16 Western       1843584942
## 17 History       1481950337
## 18 Desconocido   473725872
## 19 Documentary   352161888
## 20 TV Movie      0
```

El genero con mayores ganancias es **Action**, con un total de 1.4093667×10^{11} dólares.

8.¿A qué género principal pertenecen las películas más largas?

Pertenecen a **drama**. NOTA: Se tomaron en cuenta únicamente las primeras 20 películas más largas.

Table 4: Recuento de géneros y duración total en las películas más largas

first_genre	count
Drama	7
Action	5
Documentary	4
Comedy	1
Crime	1
Music	1
War	1

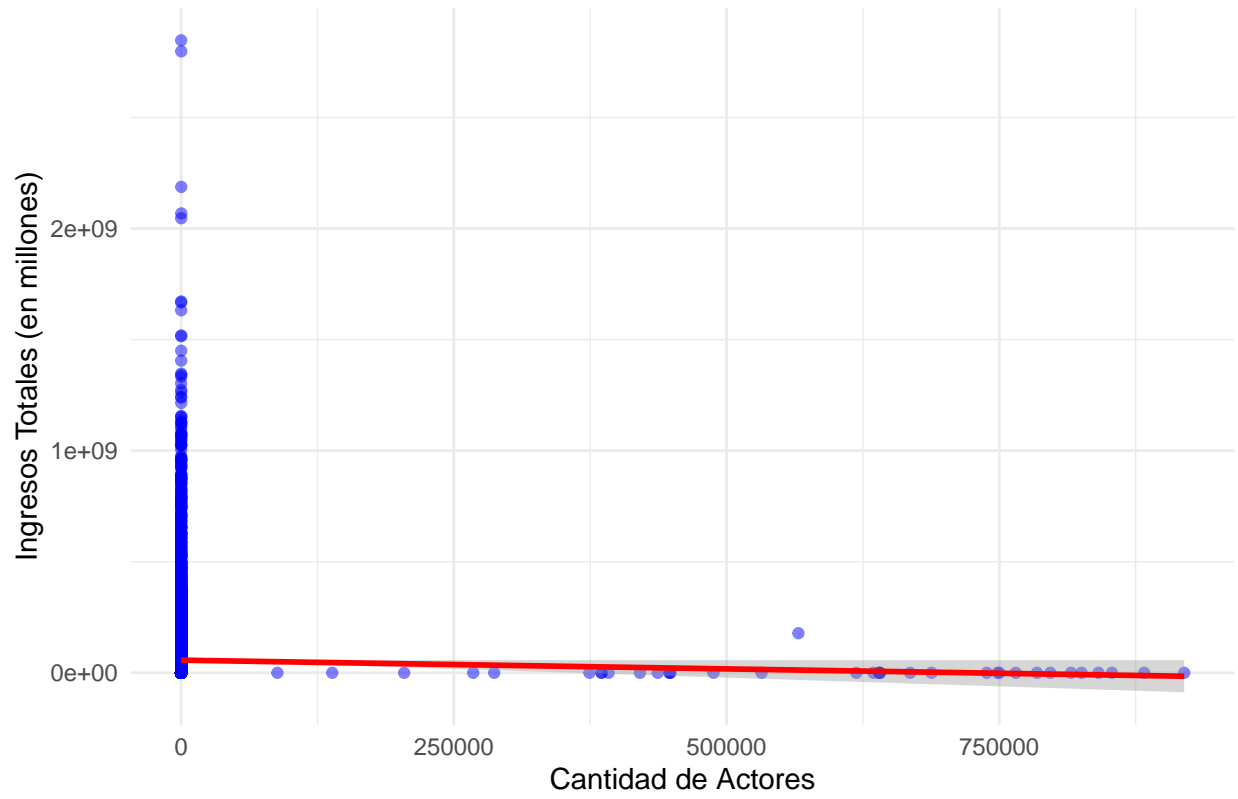
¿La cantidad de actores influye en los ingresos de las películas? ¿se han hecho películas con más actores en los últimos años?

Dada la gráfica de correlación, podemos concluir que la cantidad de actores no influye en los ingresos de las películas.

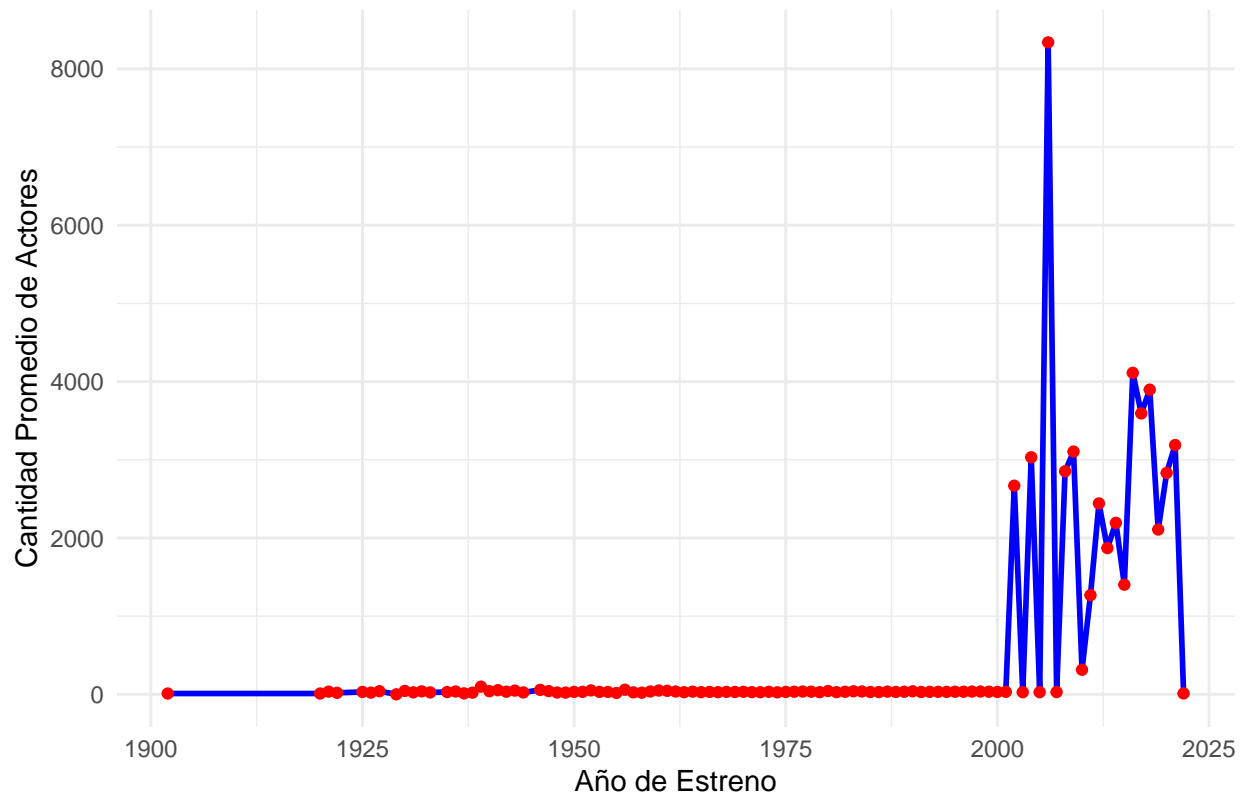
No hay una relación directa que demuestre sí se han hecho películas con más actores en los últimos años, sí que han habido algunas películas con muchos actores, pero no existe una relación que pruebe que en los últimos años se hacen películas con más actores.

```
## 'geom_smooth()' using formula = 'y ~ x'
```

Correlación entre Cantidad de Actores e Ingresos



Tendencia en la Cantidad Promedio de Actores por Año



9 ¿Es posible que la cantidad de hombres y mujeres en el reparto influya en la popularidad y los ingresos de las películas?

Para este análisis decidimos revisarlo de 2 formas distintas. Primero quicimos determinar si a medida que hay mayor porcentaje de hombres o mujeres la popularidad y revenue crece. Pero como se puede observar en las gráficas, el hecho de tener muchos más hombres que mujeres, y viciversa, en una película no implica que haya una relación directa entre el porcentaje del género mayoritario para todos los casos.

##	castMenAmount	castWomenAmount	menRatio	womenRatio	popularity	revenue
## 1	14	0	1	0	32.757	1000000
## 2	2	0	1	0	12.486	0
## 3	2	0	1	0	15.008	0
## 4	10	0	1	0	12.092	45100
## 5	3	0	1	0	80.178	76411819
## 6	26	0	1	0	30.080	138530565
## 7	17	0	1	0	21.489	154856263
## 8	45	0	1	0	33.961	69995385
## 9	15	0	1	0	13.019	35564473
## 10	4	0	1	0	53.496	138241022
## 11	10	0	1	0	12.439	0
## 12	10	0	1	0	12.655	0
## 13	24	0	1	0	22.361	11744471
## 14	14	0	1	0	15.126	0
## 15	20	0	1	0	23.445	212011111
## 16	11	0	1	0	11.643	0

## 17	3	0	1	0	10.953	0
## 18	6	0	1	0	14.245	0
## 19	15	0	1	0	12.514	7177143
## 20	14	0	1	0	11.986	0
## 21	22	0	1	0	13.486	32287044
## 22	10	0	1	0	12.019	148701
## 23	17	0	1	0	7.970	0
## 24	9	0	1	0	9.781	1612259
## 25	11	0	1	0	15.164	0
## 26	30	0	1	0	10.697	5200000
## 27	16	0	1	0	16.149	0
## 28	5	0	1	0	12.268	0
## 29	10	0	1	0	19.691	0
## 30	11	0	1	0	22.066	0

##	castMenAmount	castWomenAmount	menRatio	womenRatio	popularity	revenue
## 1	0	20	0	1	10.612	50007546
## 2	0	3	0	1	17.468	0
## 3	0	19139	0	1	17.882	0
## 4	0	1	0	1	14.849	0
## 5	0	9	0	1	12.835	0
## 6	0	10	0	1	14.664	101860
## 7	0	1	0	1	11.884	0
## 8	0	12	0	1	40.951	0
## 9	0	49131	0	1	13.013	0
## 10	0	50410	0	1	105.507	0
## 11	0	51822	0	1	24.236	0
## 12	0	57505	0	1	14.592	0
## 13	0	63404	0	1	30.499	0
## 14	0	3	0	1	20.937	0
## 15	0	95993	0	1	21.627	0
## 16	0	99942	0	1	36.233	0
## 17	0	104859	0	1	29.386	0
## 18	0	5	0	1	52.480	0
## 19	0	1	0	1	15.594	0
## 20	0	153509	0	1	29.213	0
## 21	0	173197	0	1	9.526	0
## 22	0	1	0	1	15.744	0
## 23	0	212193	0	1	37.689	0
## 24	0	213901	0	1	17.435	0
## 25	0	1	0	1	56.548	0
## 26	0	224903	0	1	38.351	92100
## 27	0	3	0	1	51.453	1000000
## 28	0	251799	0	1	13.573	0
## 29	0	1	0	1	11.748	0
## 30	0	254126	0	1	15.340	0

Sin embargo, si ordenamos por popularidad las películas y observamos el porcentaje de genero del cast, se puede observar que en algunas de las películas populares en efecto hay un mayor porcentaje de hombres. Pero como tal el porcentaje de género no influye directamente en la popularidad y mucho menos en las ganancias según lo que se observa debajo.

##	castMenAmount	castWomenAmount	menRatio	womenRatio	popularity	revenue
----	---------------	-----------------	----------	------------	------------	---------

## 1	25	11	0.6944444	0.3055556	11474.647	401842256
## 2	33	13	0.7173913	0.2826087	8443.740	1631853496
## 3	13	8	0.6190476	0.3809524	6055.643	215000000
## 4	10	9	0.5263158	0.4736842	5887.379	31000000
## 5	9	9	0.5000000	0.5000000	5804.441	215000000
## 6	18	12	0.6000000	0.4000000	5051.222	191000000
## 7	2	3	0.4000000	0.6000000	4789.705	0
## 8	34	15	0.6938776	0.3061224	3828.374	148000000
## 9	26	11	0.7027027	0.2972973	3062.764	500000000
## 10	12	8	0.6000000	0.4000000	2466.985	0
## 11	12	3	0.8000000	0.2000000	2216.355	0
## 12	19	2	0.9047619	0.0952381	2179.912	178143
## 13	22	13	0.6285714	0.3714286	2066.867	430238384
## 14	10	4	0.7142857	0.2857143	1699.701	0
## 15	9	5	0.6428571	0.3571429	1553.397	18312201
## 16	14	9	0.6086957	0.3913043	1477.724	67000000
## 17	13	4	0.7647059	0.2352941	1473.287	0
## 18	20	14	0.5882353	0.4117647	1372.990	503063688
## 19	13	9	0.5909091	0.4090909	1340.953	5370393
## 20	14	12	0.5384615	0.4615385	1176.353	0
## 21	23	6	0.7931034	0.2068966	1162.670	27000000
## 22	35	16	0.6862745	0.3137255	1130.609	757930663
## 23	3	2	0.6000000	0.4000000	1085.594	0
## 24	7	5	0.5833333	0.4166667	1055.010	0
## 25	26	9	0.7428571	0.2571429	980.267	771000000
## 26	5	9	0.3571429	0.6428571	973.571	515596
## 27	23	12	0.6571429	0.3428571	968.579	396000000
## 28	32	18	0.6400000	0.3600000	943.722	331096766
## 29	30	15	0.6666667	0.3333333	922.978	167381210
## 30	8	4	0.6666667	0.3333333	915.636	128000000

10. ¿Quiénes son los directores que hicieron las 20 películas mejor calificadas?

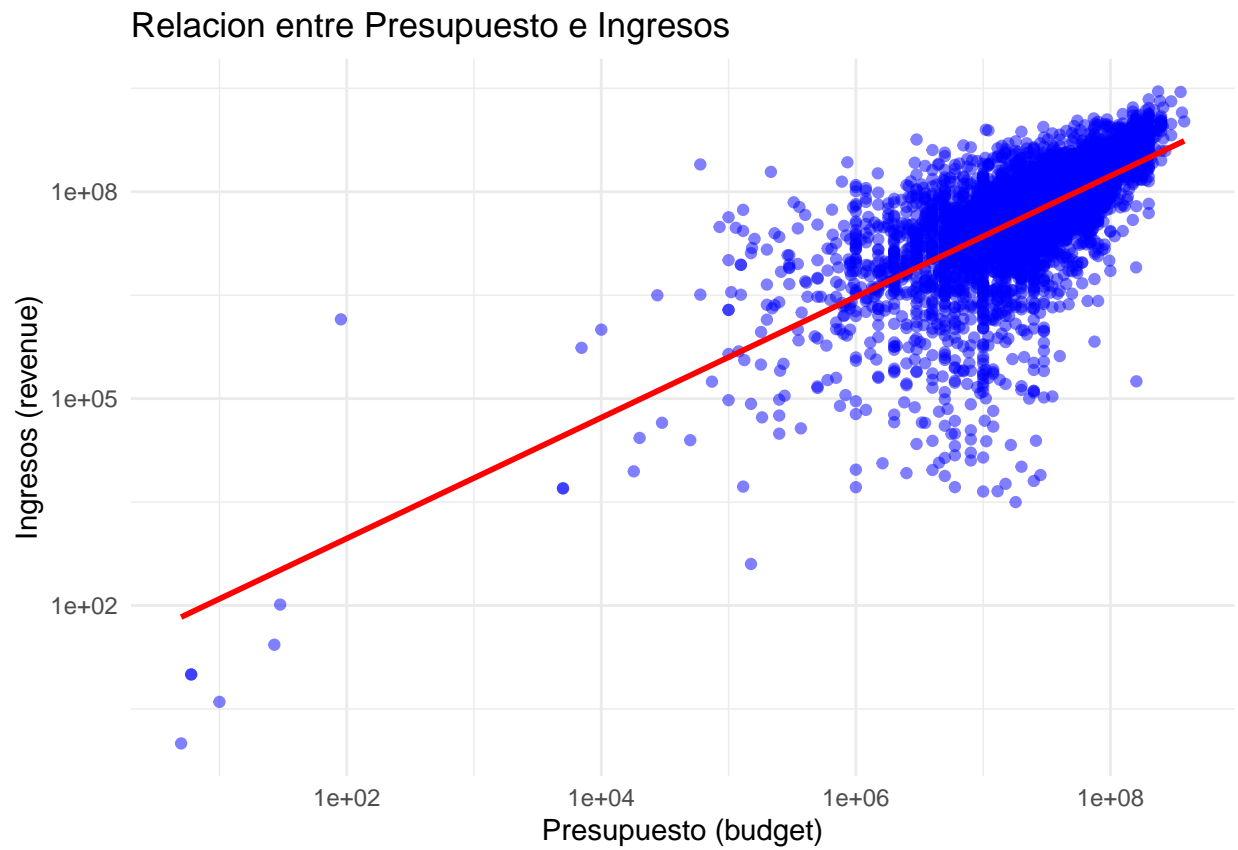
Table 5: Directores de las 20 películas mejor calificadas

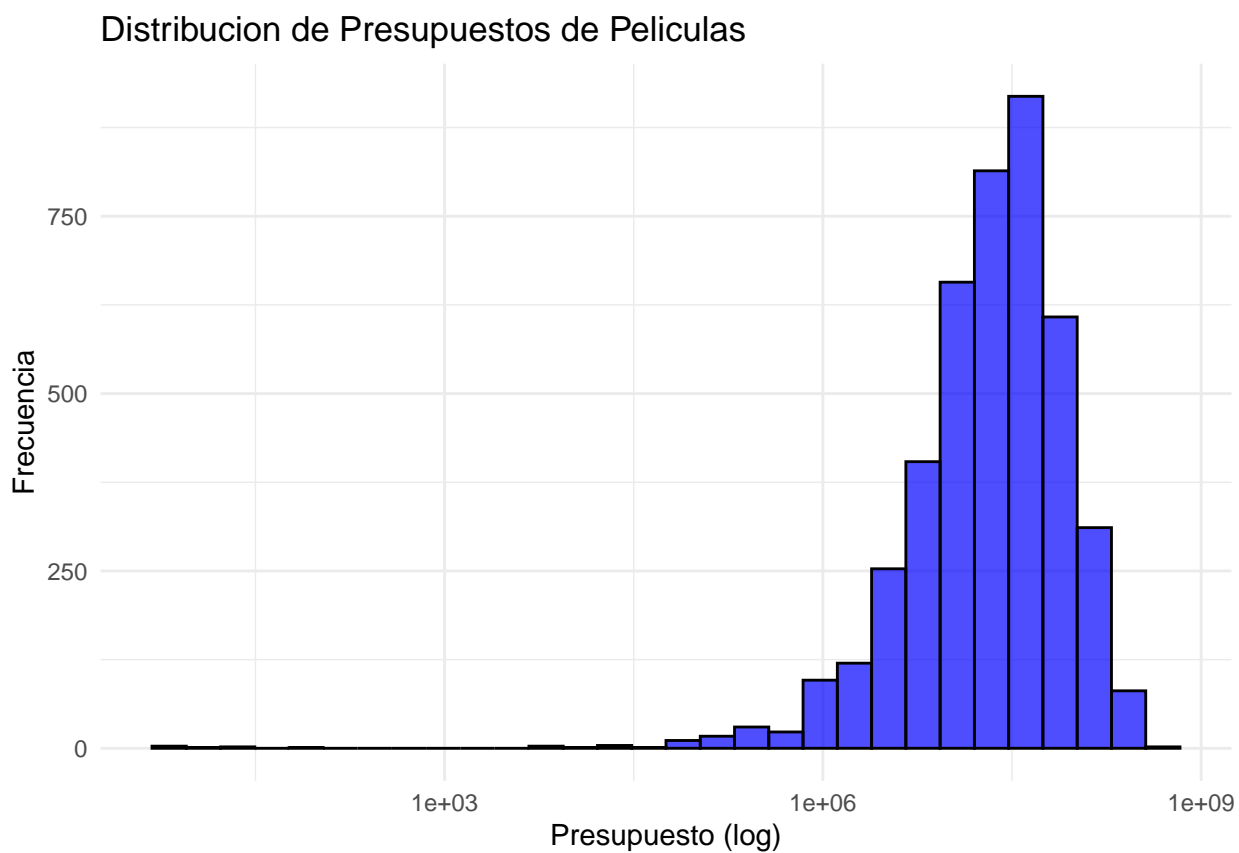
director	voteAvg
Thomas Coven	10.0
Víctor Barba, Juan Olivares	10.0
Rebecca Sugar	10.0
Laurent Bouzereau	10.0
Kaku Arakawa	10.0
Christin Baker	10.0
Miguel Angel Zavala	10.0
Dave Bullock, Troy Adomitis, Victor Cook	9.6
Samuel Leong	9.5
Won Myeong-jun	9.5
Selena Quintanilla	9.4
Haruo Sotozaki	9.3
Haruo Sotozaki	9.3
Ulises Valencia	9.2
Kim Nam-joon, Jeon Jung-kook, Kim Tae-hyung, Park Ji-min, Jung Ho-seok, Kim Seok-jin, Min Yoon-gi	9.2

Park Jun-soo	9.2
Igor Kopylov	9.2
Park Jun-soo	9.1
Michael Goguen	9.0
Filip Ghiorghi	9.0

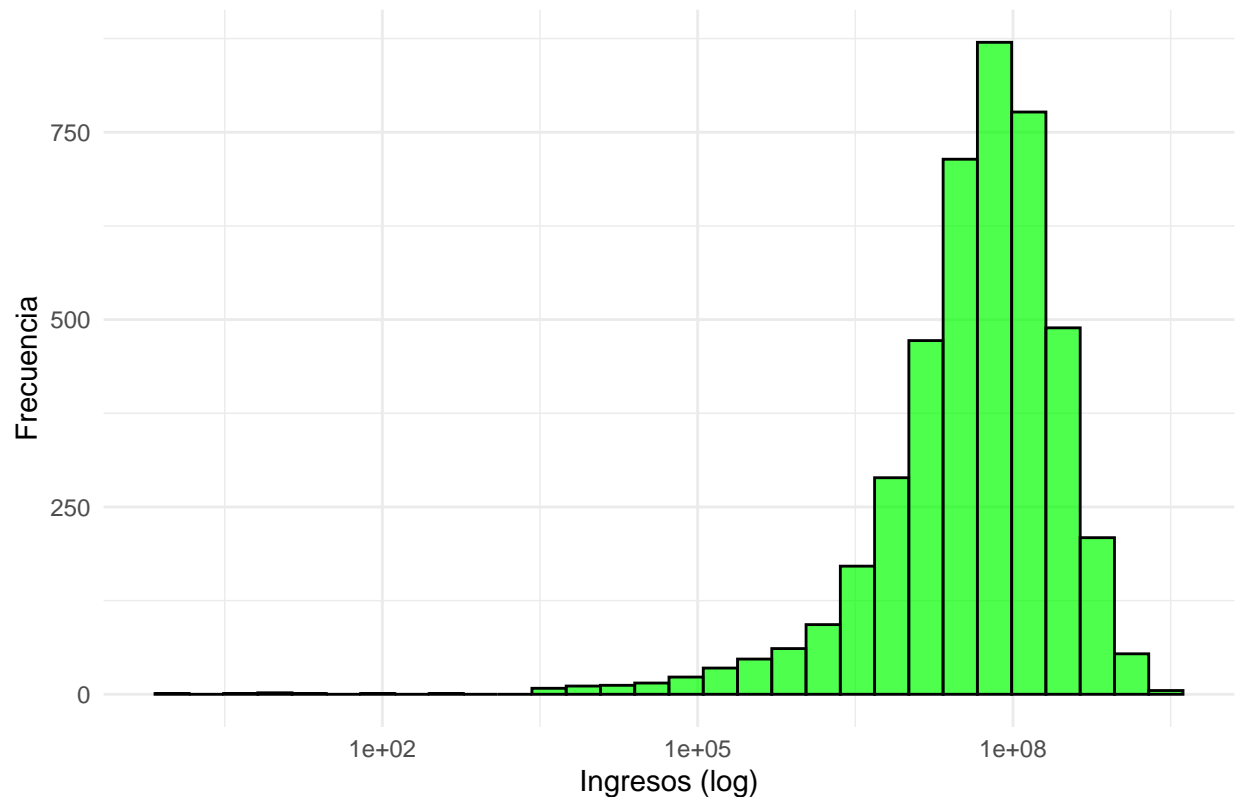
11 ¿Cómo se correlacionan los presupuestos con los ingresos? ¿Los altos presupuestos significan altos ingresos? Haga los gráficos que necesite, histograma, diagrama de dispersión

```
## 'geom_smooth()' using formula = 'y ~ x'
```





Distribucion de Ingresos de Peliculas



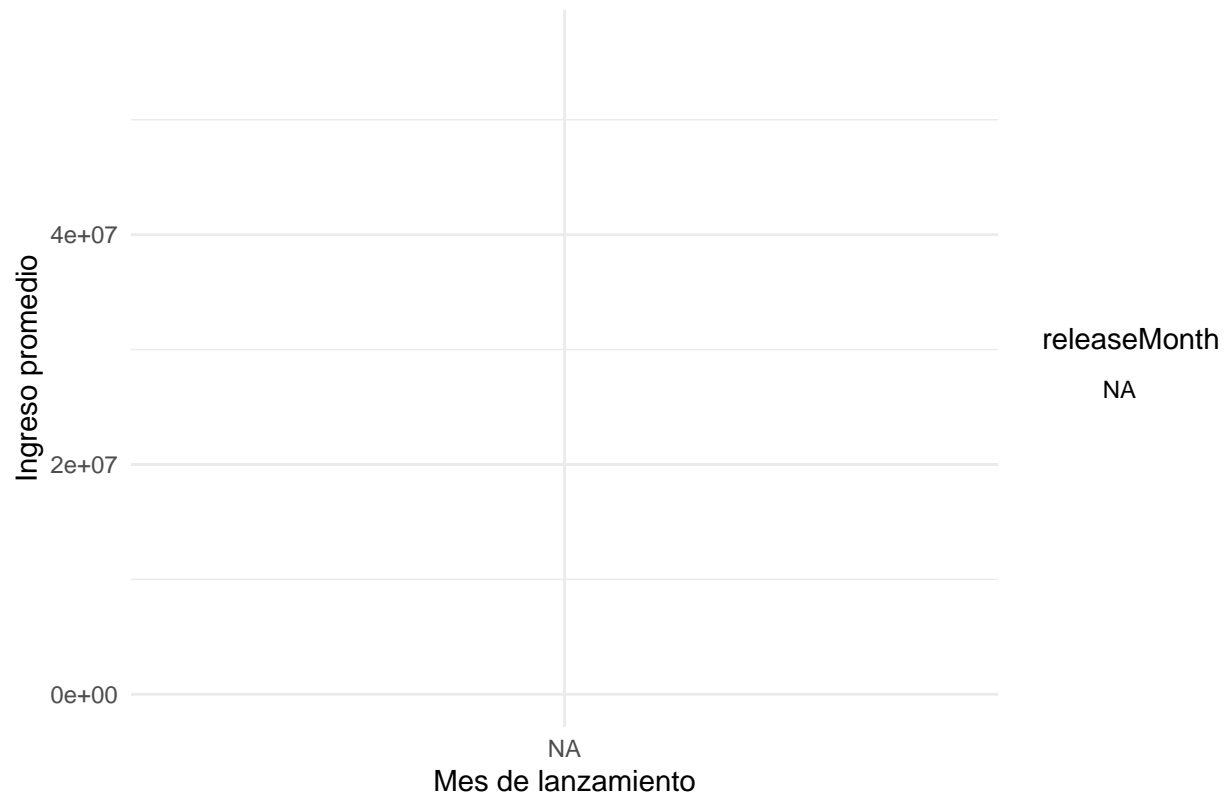
Como se puede observar, si hay una tendencia y relación clara entre el budget y el revenue. Aunque un budget alto no represente siempre un pexito, es claro que mientras mayor presupuesto tenga la película, más oportunidad de generar ganancias tiene. A medida que incremente el presupuesto también lo hacen las ganancias.

12. ¿Se asocian ciertos meses de lanzamiento con mejores ingresos?

Al calcular el promedio de ingresos de las películas agrupadas por mes, podemos observar como las películas estrenadas en junio son la mejor recibidas, seguidas de mayo y julio, el próximo mes en que reciben nien a las películas es en diciembre.

```
## Warning: There was 1 warning in 'mutate()'.  
## i In argument: 'releaseDate = dmy(releaseDate)'.  
## Caused by warning:  
## ! All formats failed to parse. No formats found.
```

Promedio de ingresos por mes de lanzamiento



13 ¿En qué meses se han visto los lanzamientos con mejores ingresos? ¿cuántas películas, en promedio, se han lanzado por mes?

##	mes_texto	revenue	title
## 3211	December	2847246203	Avatar
## 5953	April	2797800564	Avengers: Endgame
## 308	November	2187463944	Titanic
## 4948	December	2068223624	Star Wars: The Force Awakens
## 5954	April	2046239637	Avengers: Infinity War
## 4915	June	1671713208	Jurassic World
## 7135	July	1667635327	The Lion King
## 9050	December	1631853496	Spider-Man: No Way Home
## 3398	April	1518815515	The Avengers
## 5088	April	1515047671	Furious 7
## 6181	November	1450026933	Frozen II
## 4711	April	1405403694	Avengers: Age of Ultron
## 5799	February	1346739107	Black Panther
## 2510	July	1341511219	Harry Potter and the Deathly Hallows: Part 2
## 5149	December	1332698830	Star Wars: The Last Jedi
## 6429	June	1303459585	Jurassic World: Fallen Kingdom
## 4766	November	1274219009	Frozen
## 6109	March	1263521126	Beauty and the Beast
## 5626	June	1242805359	Incredibles 2
## 6272	April	1238764765	The Fate of the Furious
## 4339	April	1214811252	Iron Man 3

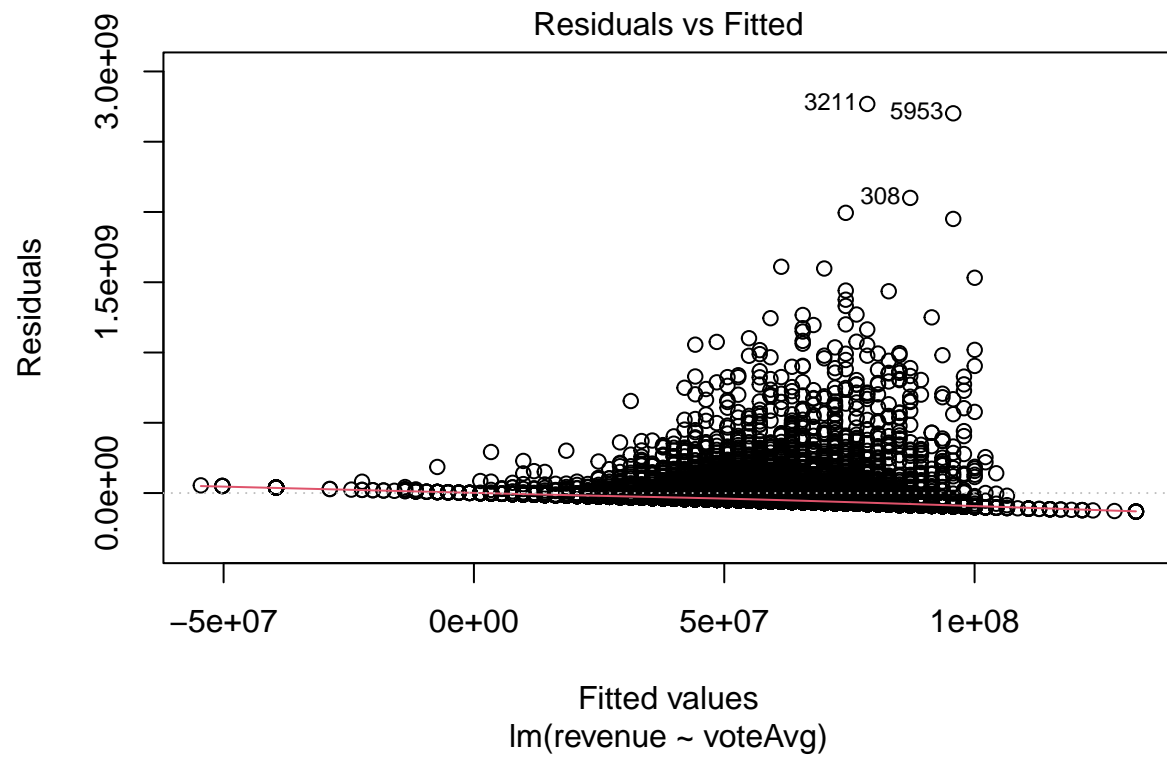
## 5302	June	1156730962	Minions
## 5702	April	1153296293	Captain America: Civil War
## 5940	July	1148461807	Aquaman
## 7240	June	1131927996	Spider-Man: Far From Home
## 5955	March	1128276090	Captain Marvel
## 3772	June	1123794079	Transformers: Dark of the Moon
## 69	December	1118888979	The Lord of the Rings: The Return of the King
## 3736	October	1108561013	Skyfall
## 4666	June	1100000000	Transformers: Age of Extinction

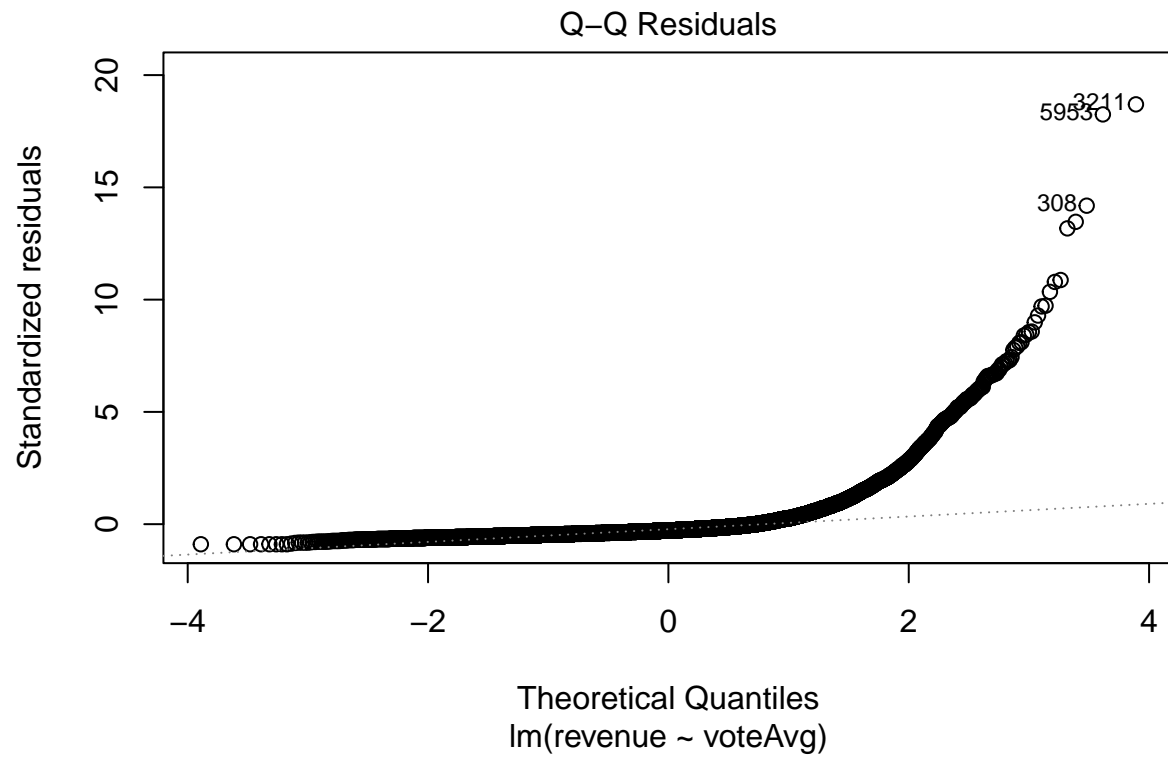
El promedio de películas por mes en general es **833.3333333**. A continuación se muestra como se distribuyen las películas en los meses del año.

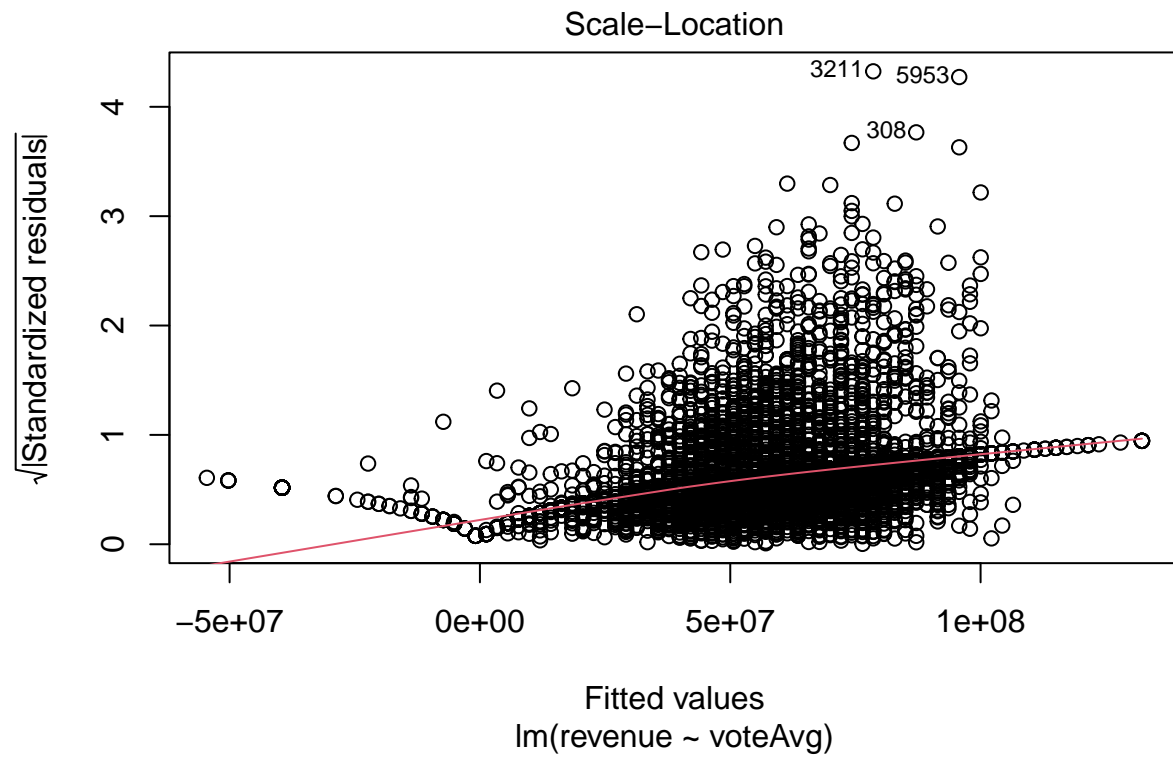
```
## # A tibble: 12 x 2
##   mes_texto cantidad_peliculas
##   <chr>          <int>
## 1 April             696
## 2 August            913
## 3 December         935
## 4 February         706
## 5 January          652
## 6 July             812
## 7 June             819
## 8 March            815
## 9 May             698
## 10 November        807
## 11 October        1068
## 12 September       1079
```

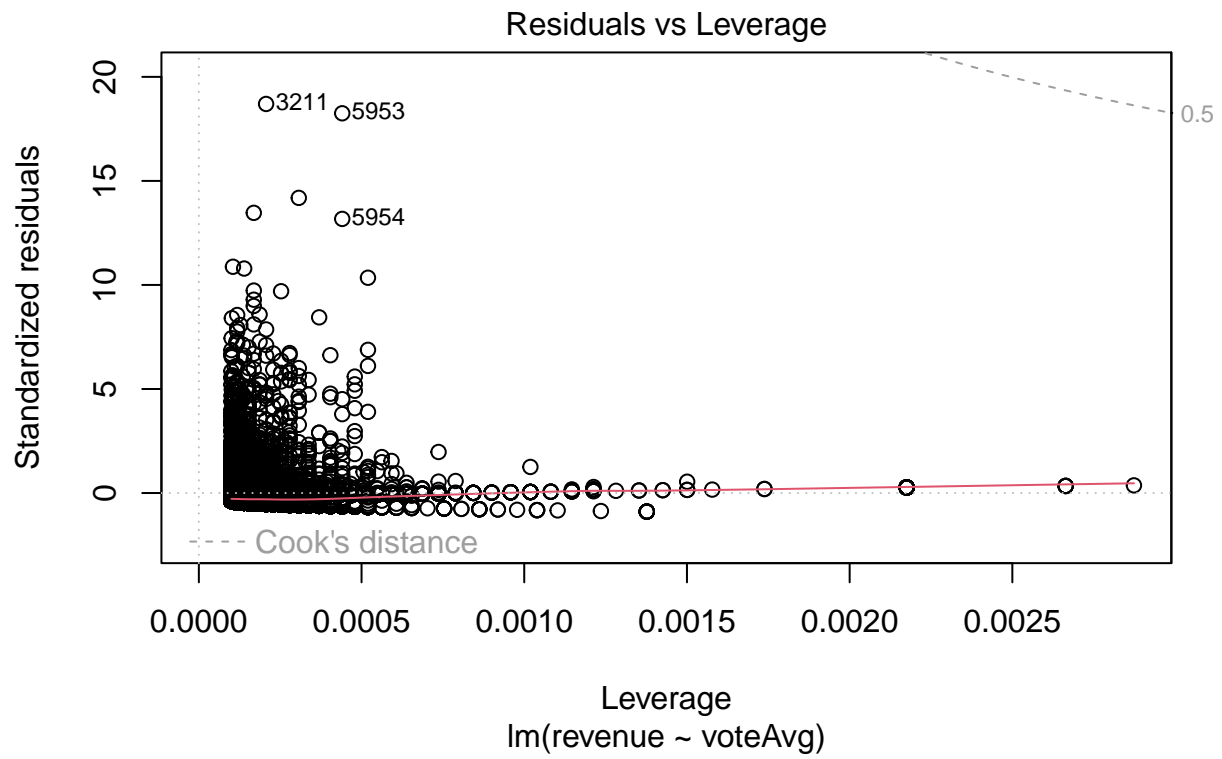
14.¿Cómo se correlacionan las calificaciones con el éxito comercial?

Existe una relación significativa entre las calificaciones y sus ingresos, pero esta relación no es definitiva ya que el modelo solo aplica un 2% de los ingresos, lo que nos dice que existen otros factores que definen el éxito comercial.





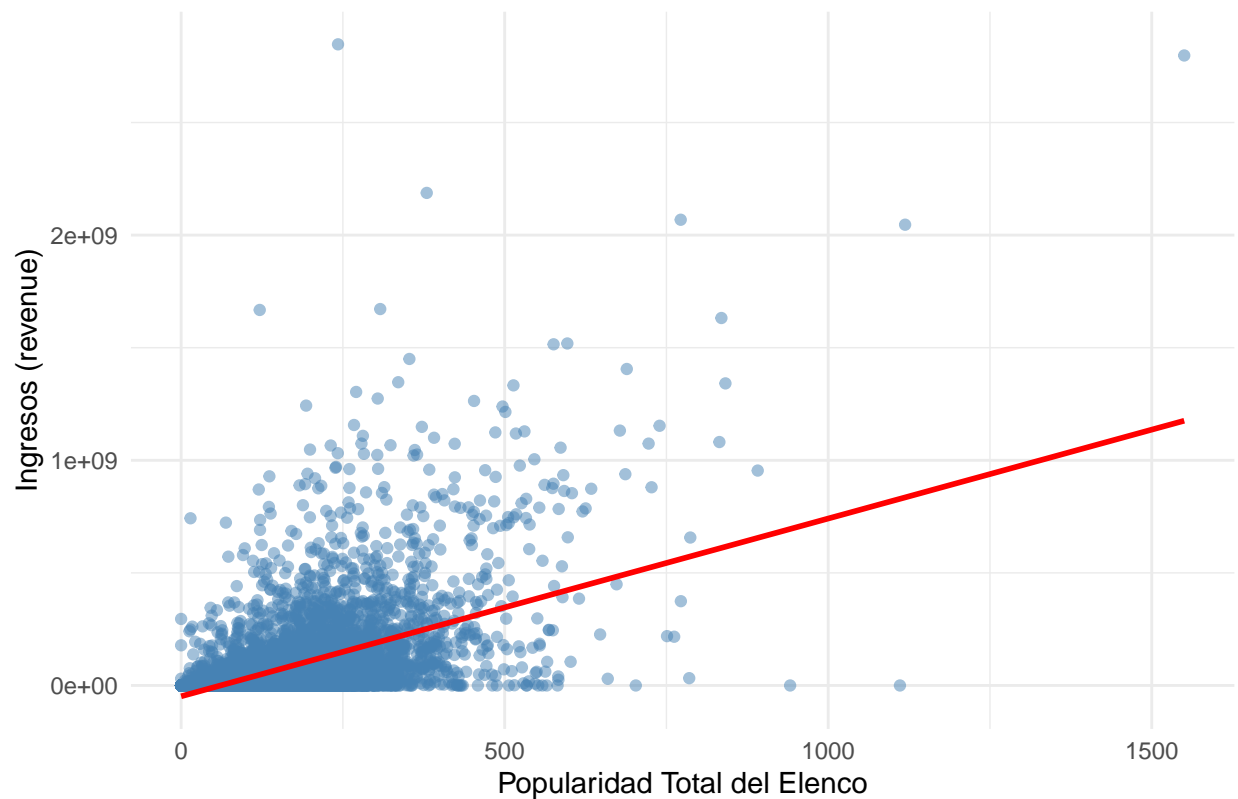




15 ¿La popularidad del elenco está directamente correlacionada con el éxito de taquilla?

'geom_smooth()' using formula = 'y ~ x'

Relacion entre la Popularidad del Elenco y los Ingresos

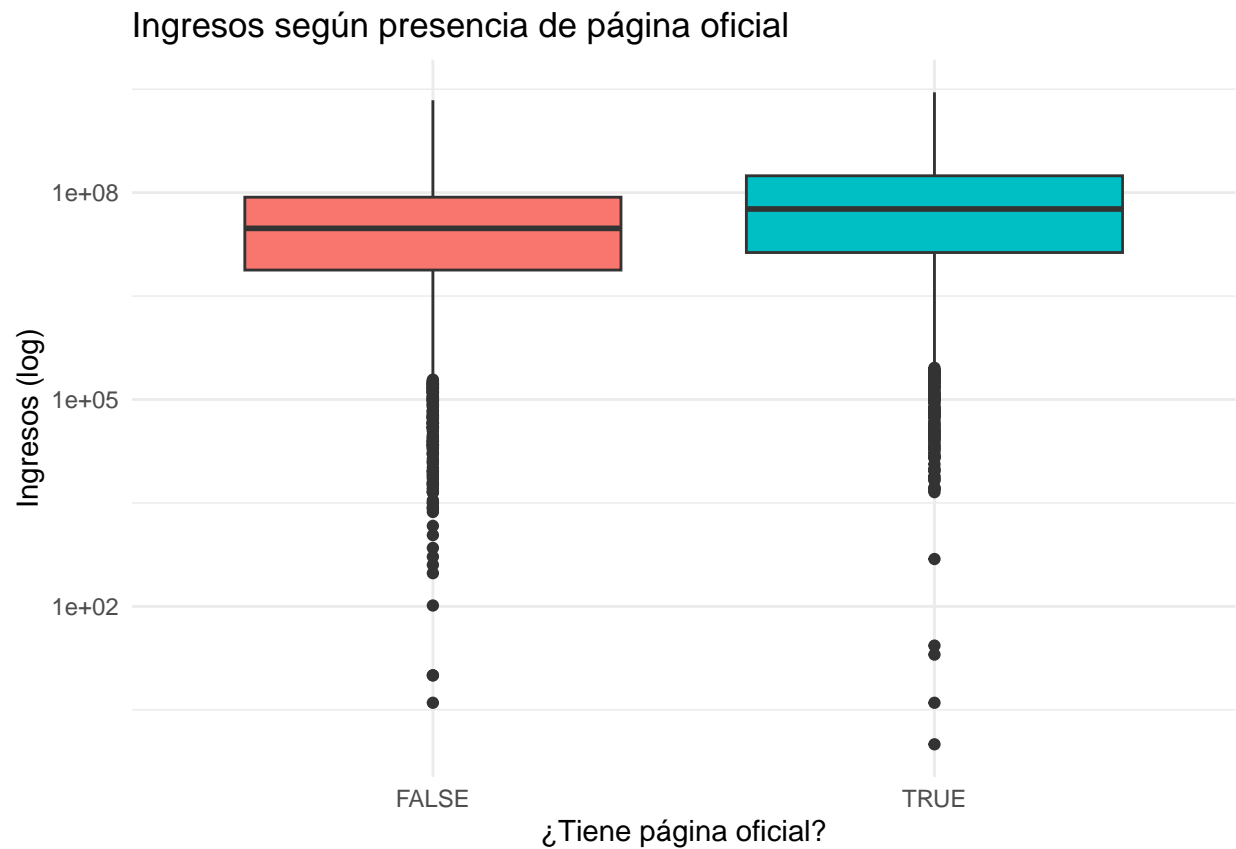


Al final se tiene un coeficiente de **0.5492487**. Lo cual indica que hay una correlacion entre la popularidad del elenco y las ganancias y exito en taquilla. Aunque no se tan alto si existen una relacion entre ambas variables.

16. ¿Qué estrategias de marketing, como videos promocionales o páginas oficiales, generan mejores resultados?

Se observa que tener una **página oficial** está fuertemente asociado con mayores ingresos de las películas mientras que los **videos promocionales** no tienen un impacto significativo. Es decir que las películas con una página oficial tienden a ganar más dinero mientras que la presencia de un video promocional no garantiza mayores ingresos.

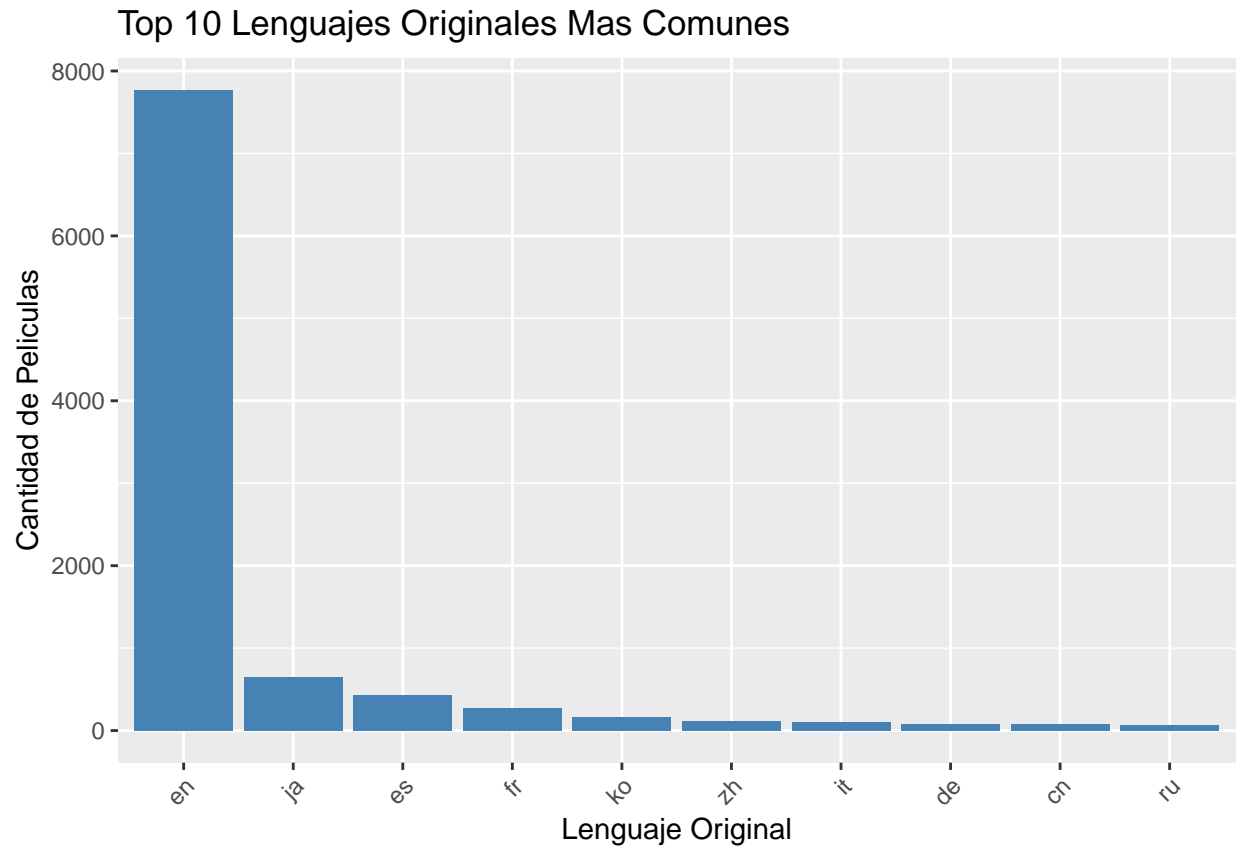




Extras

¿Cuales son los 10 lenguajes originales mas comunes entre todas las peliculas?

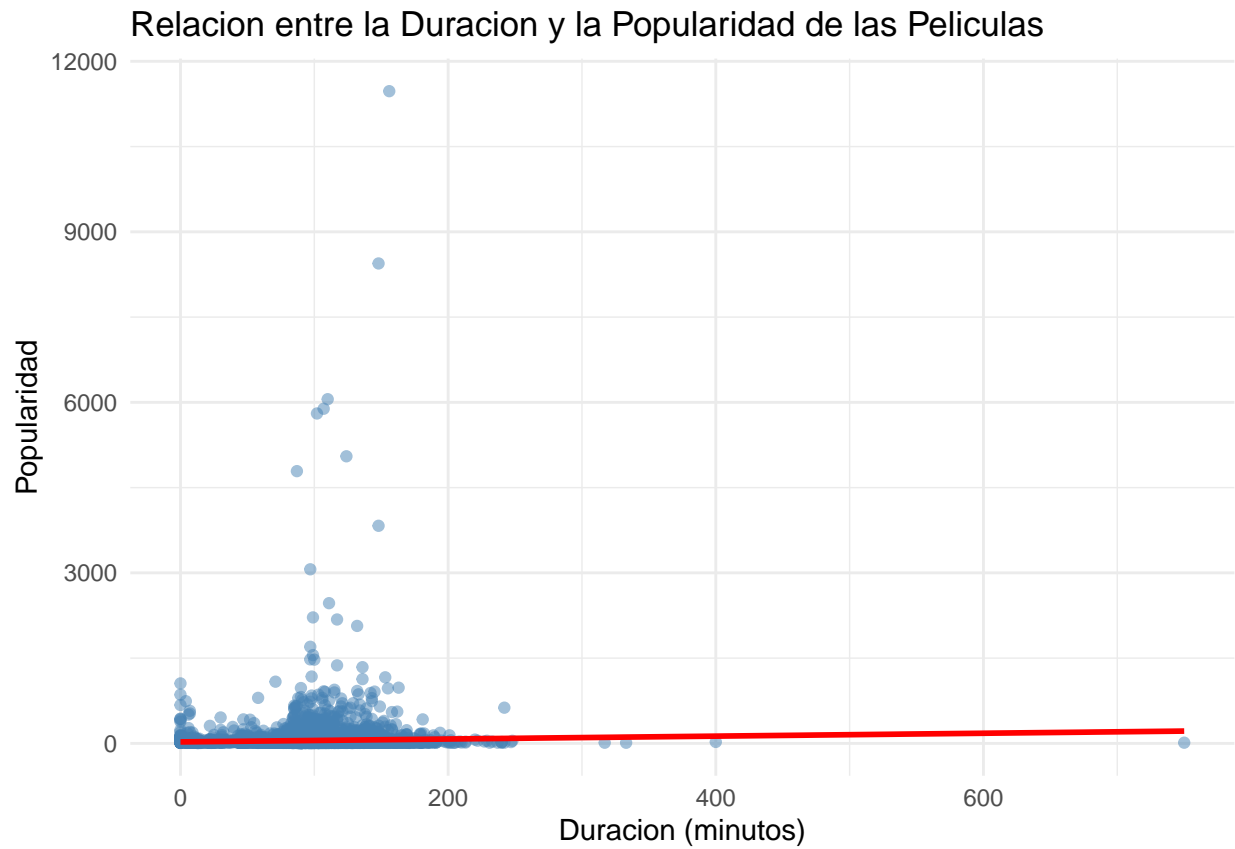
```
## # A tibble: 10 x 2
##   originalLanguage cantidad_len
##   <chr>                <int>
## 1 en                    7772
## 2 ja                     644
## 3 es                     425
## 4 fr                     271
## 5 ko                     167
## 6 zh                     119
## 7 it                     100
## 8 de                      84
## 9 cn                      80
## 10 ru                     67
```



¿Hay alguna correlación entre el largo de la película y su popularidad? Son populares las películas largas?

##	runtime	popularity
## 9348	750	12.190
## 5359	400	25.716
## 3886	333	11.224
## 963	317	11.607
## 1264	248	45.072
## 7066	247	23.168
## 1949	242	12.449
## 9687	242	628.372
## 3741	240	16.700
## 5593	240	24.845
## 6160	240	15.204
## 2949	237	12.369
## 415	233	32.151
## 9220	231	12.326
## 177	229	45.994
## 517	227	33.961
## 365	222	47.378
## 1169	220	66.514
## 4278	213	30.499
## 9769	213	14.382

```
## 'geom_smooth()' using formula = 'y ~ x'
```



```
## [1] "El coeficiente de correlacion es: 0.03"
```

Al final se puede observar que realmente no hay una alta correlación entre la duración de la película y su popularidad. Si una película es larga o corta no debería de influir en la popularidad de la película.

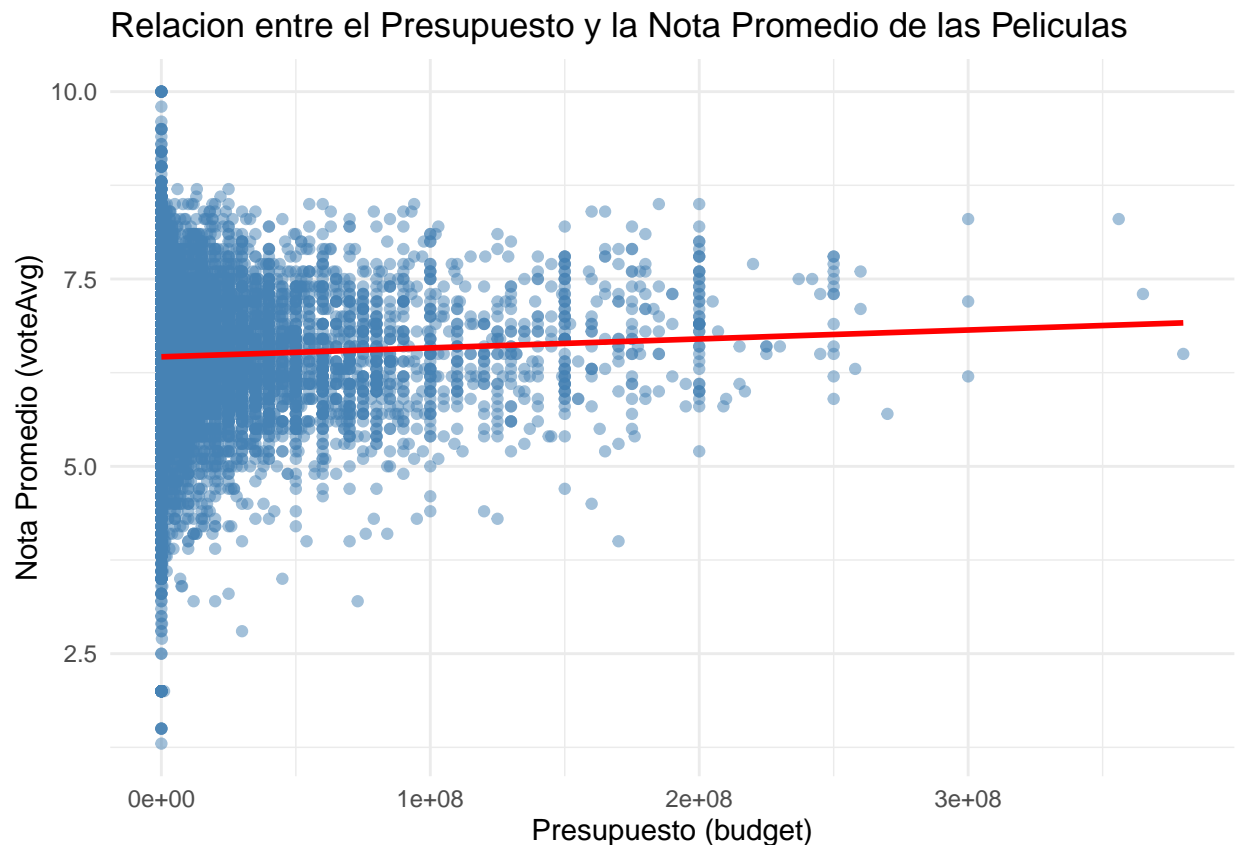
El coeficiente de correlación entre la duracion de la pelicula y su popularidaa es **0.03**

¿Las peliculas con mejor budget tienen mejores notas promedio?

```
##      voteAvg    budget
## 717      6.5 380000000
## 4711     7.3 365000000
## 5953     8.3 356000000
## 164      7.2 300000000
## 4954     6.2 300000000
## 5954     8.3 300000000
## 608      5.7 270000000
## 3792     7.6 260000000
## 7135     7.1 260000000
## 281      6.3 258000000
## 412      7.7 250000000
## 2509     7.8 250000000
## 4035     7.8 250000000
```

```
## 4039      7.3 250000000
## 4052      6.2 250000000
## 4179      7.6 250000000
## 4856      7.3 250000000
## 4881      7.5 250000000
## 5150      6.5 250000000
## 5280      5.9 250000000
```

```
## 'geom_smooth()' using formula = 'y ~ x'
```



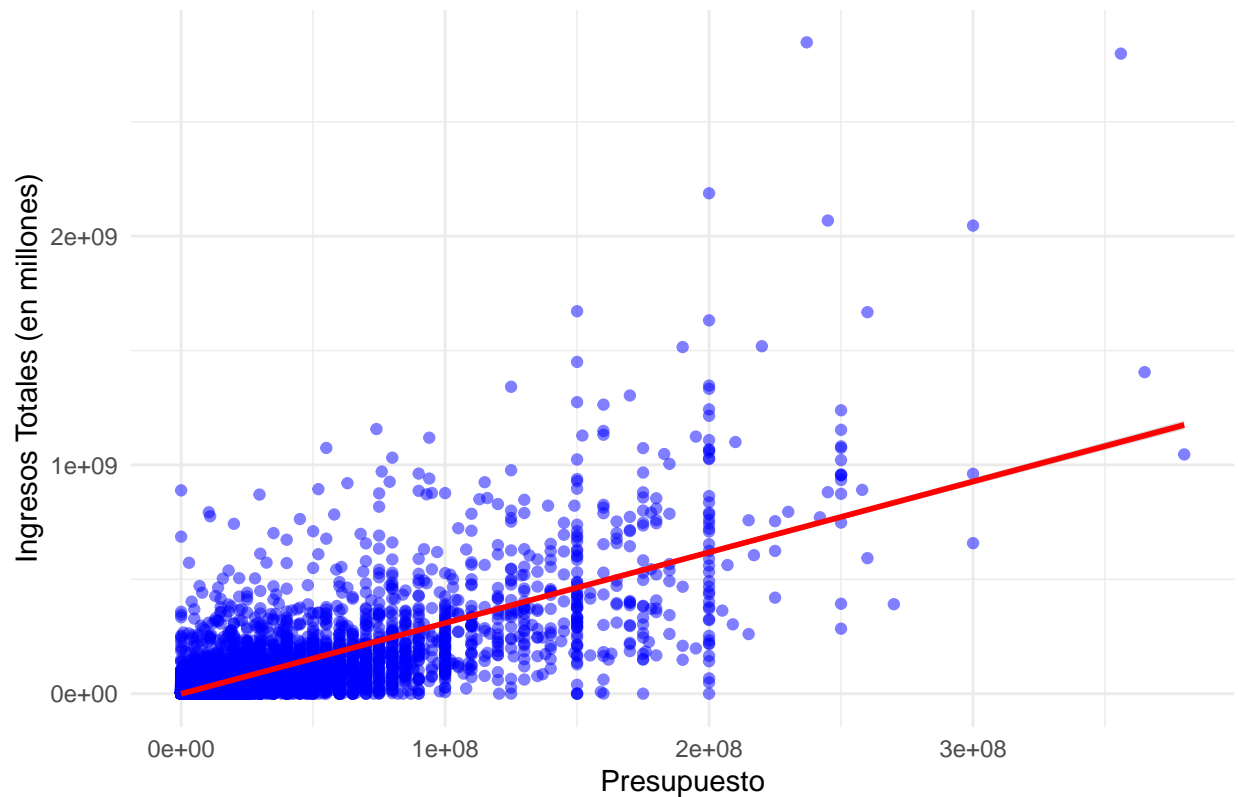
Se obtuvo un coeficiente de **0.04**. Lo cual indica que el presupuesto de una película no influye para nada en la nota promedio de la película. un buen budget no implica buenas criticas.

¿Existe relación entre el presupuesto de la película y sus ingresos?

Los resultados indican una **relación positiva y moderadamente fuerte** entre el presupuesto de una película y sus ingresos. El **coeficiente de correlación de Pearson (0.757)** sugiere una asociación positiva considerable entre ambas variables. Además el **modelo de regresión lineal** muestra que en promedio por cada unidad adicional en el presupuesto los ingresos aumentan en **3.09 veces dicho valor**. Sin embargo, el **$R^2 = 0.5737$** indica que aproximadamente el **57.4% de la variabilidad en los ingresos** puede explicarse por el presupuesto, lo que sugiere que otros factores también influyen en el éxito de una película. En general el presupuesto es un buen predictor de los ingresos, pero no el único determinante del éxito financiero.

```
## 'geom_smooth()' using formula = 'y ~ x'
```


Correlación entre Presupuesto e Ingresos



```
## [1] 0.7574542
```

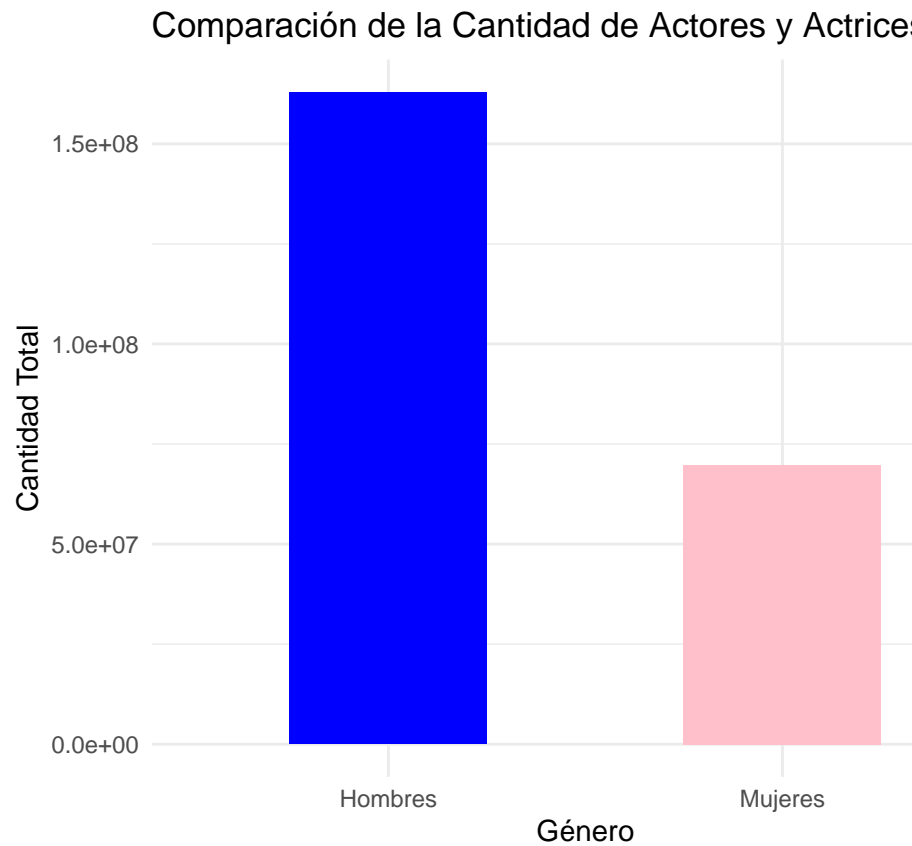
```
##
## Call:
## lm(formula = revenue ~ budget, data = movies)
##
## Residuals:
##      Min       1Q   Median       3Q      Max
## -618034787 -13405091   650592   755710 2114754621
##
## Coefficients:
##              Estimate Std. Error t value Pr(>|t|)
## (Intercept) -6.506e+05  1.095e+06  -0.594   0.552
## budget       3.093e+00  2.667e-02 116.004 <2e-16 ***
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Residual standard error: 97670000 on 9998 degrees of freedom
## Multiple R-squared:  0.5737, Adjusted R-squared:  0.5737
## F-statistic: 1.346e+04 on 1 and 9998 DF, p-value: < 2.2e-16
```

¿Cuál era el índice promedio de popularidad de los actores de las 10 películas más taquilleras?

Table 6: 10 películas con mayores ingresos

title	avgpopularity	revenue
Avatar	6.55	2847246203
Avengers: Endgame	15.05	2797800564
Titanic	3.27	2187463944
Star Wars: The Force Awakens	4.11	2068223624
Avengers: Infinity War	17.21	2046239637
Jurassic World	6.41	1671713208
The Lion King	6.07	1667635327
Spider-Man: No Way Home	13.92	1631853496
The Avengers	5.28	1518815515
Furious 7	10.10	1515047671

En todo el dataset, ¿la cantidad de actores y actrices es igual?



No, el género que más abunda es el masculino.