

lab-01

2026-02-06

Laboratorio 1: Análisis Exploratorio (CineVision Studios)

Cargar datos

1. Exploración rápida (resumen del dataset)

Enunciado: Haga una exploración rápida de sus datos, para eso haga un resumen de su conjunto de datos.

##	id	budget	genres	homePage
##	Min. : 5	Min. : 0	Length:19883	Length:19883
##	1st Qu.: 146220	1st Qu.: 0	Class :character	Class :character
##	Median : 869623	Median : 0	Mode :character	Mode :character
##	Mean : 902240	Mean : 9413280		
##	3rd Qu.:1589602	3rd Qu.: 1000000		
##	Max. :1627166	Max. :380000000		
##				
##	productionCompany	productionCompanyCountry	productionCountry	
##	Length:19883	Length:19883	Length:19883	
##	Class :character	Class :character	Class :character	
##	Mode :character	Mode :character	Mode :character	
##				
##				
##				
##	revenue	runtime	video	director
##	Min. :0.000e+00	Min. : 0.00	Mode :logical	Length:19883
##	1st Qu.:0.000e+00	1st Qu.: 10.00	FALSE:19313	Class :character
##	Median :0.000e+00	Median : 86.00	TRUE :84	Mode :character
##	Mean :2.879e+07	Mean : 66.09	NA's :486	
##	3rd Qu.:3.306e+05	3rd Qu.:103.00		
##	Max. :2.847e+09	Max. :750.00		
##				
##	actors	actorsPopularity	actorsCharacter	originalTitle
##	Length:19883	Length:19883	Length:19883	Length:19883
##	Class :character	Class :character	Class :character	Class :character
##	Mode :character	Mode :character	Mode :character	Mode :character
##				
##				
##				
##				
##	title	originalLanguage	popularity	releaseDate
##	Length:19883	Length:19883	Min. :0.000e+00	Length:19883
##	Class :character	Class :character	1st Qu.:5.460e-02	Class :character
##	Mode :character	Mode :character	Median :8.502e+00	Mode :character
##			Mean :2.625e+01	
##			3rd Qu.:2.224e+01	
##			Max. :1.147e+04	
##				
##	voteAvg	voteCount	genresAmount	productionCoAmount
##	Min. : 0.000	Min. : 0.0	Min. : 0.000	Min. : 0.000
##	1st Qu.: 0.000	1st Qu.: 0.0	1st Qu.: 1.000	1st Qu.: 0.000
##	Median : 5.400	Median : 6.0	Median : 2.000	Median : 1.000
##	Mean : 3.837	Mean : 675.9	Mean : 1.949	Mean : 1.973
##	3rd Qu.: 6.800	3rd Qu.: 423.0	3rd Qu.: 3.000	3rd Qu.: 3.000
##	Max. :10.000	Max. :30788.0	Max. :16.000	Max. :89.000
##				
##	productionCountriesAmount	actorsAmount	castWomenAmount	castMenAmount
##	Min. : 0.00	Min. : 0	Min. : 0	Min. : 0
##	1st Qu.: 1.00	1st Qu.: 3	1st Qu.: 0	1st Qu.: 0
##	Median : 1.00	Median : 9	Median : 2	Median : 3
##	Mean : 1.23	Mean : 1082	Mean : 3517	Mean : 8224
##	3rd Qu.: 1.00	3rd Qu.: 21	3rd Qu.: 6	3rd Qu.: 12
##	Max. :155.00	Max. :919590	Max. :922162	Max. :922017
##			NA's :37	NA's :162

```
##      releaseYear
##  Min.      :1902
##  1st Qu.:2013
##  Median :2021
##  Mean    :2017
##  3rd Qu.:2025
##  Max.    :2026
##  NA's    :2
```

2. Tipo de cada variable

Enunciado: Diga el tipo de cada una de las variables (cualitativa ordinal o nominal, cuantitativa continua, cuantitativa discreta).

variable	tipo	descripcion
id	Cualitativa nominal	Identificador; numérico pero no mide magnitud
popularity	Cuantitativa continua	Índice (escala continua)
budget	Cuantitativa continua	Presupuesto (moneda)
revenue	Cuantitativa continua	Ingresos (moneda)
originalTitle	Cualitativa nominal	Texto
originalLanguage	Cualitativa nominal	Código de idioma
title	Cualitativa nominal	Texto (título en inglés)
homePage	Cualitativa nominal	URL (alta cardinalidad, muchos NA)
video	Cualitativa nominal	Binaria (sí/no)
director	Cualitativa nominal	Nombre (alta cardinalidad)
runtime	Cuantitativa continua	Duración (minutos; medida de tiempo)
genres	Cualitativa nominal	Multi-etiqueta (separada por coma o ' ')
genresAmount	Cuantitativa discreta	Conteo de géneros
productionCompany	Cualitativa nominal	Texto / multi-etiqueta (según fuente)
productionCoAmount	Cuantitativa discreta	Conteo de compañías productoras
productionCompanyCountry	Cualitativa nominal	Multi-etiqueta (códigos de país)
productionCountry	Cualitativa nominal	Multi-etiqueta (códigos de país)
productionCountriesAmount	Cuantitativa discreta	Conteo de países
releaseDate	Cualitativa ordinal	Fecha (tiene orden temporal)
voteCount	Cuantitativa discreta	Conteo de votos
voteAvg	Cuantitativa continua	Promedio (0–10)
actors	Cualitativa nominal	Lista de actores (texto, separada por coma o ' ')
actorsPopularity	Cuantitativa continua	Popularidad del elenco (numérica o derivable)

variable	tipo	descripcion
actorsCharacter	Cualitativa nominal	Lista de personajes (texto)
actorsAmount	Cuantitativa discreta	Conteo de actores
castWomenAmount	Cuantitativa discreta	Conteo de actrices
castMenAmount	Cuantitativa discreta	Conteo de actores
releaseYear	Cuantitativa discreta	Año

3. Normalidad en variables cuantitativas y tabla de frecuencias de cualitativas

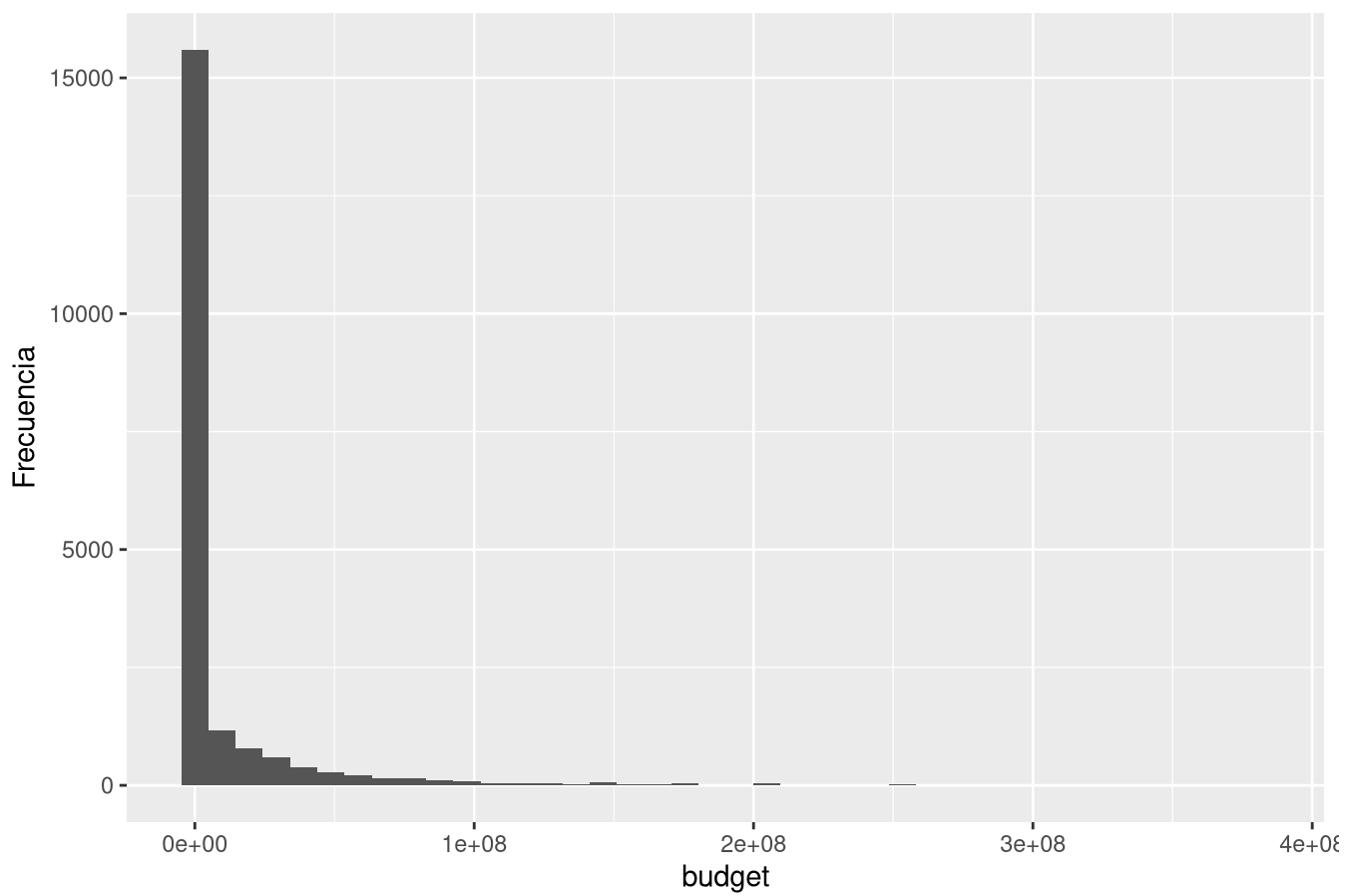
Enunciado: Investigue si las variables cuantitativas siguen una distribución normal y haga una tabla de frecuencias de las variables cualitativas. Explique todos los resultados.

Nota: con muestras muy grandes, pruebas como Shapiro-Wilk tienden a rechazar normalidad por desviaciones pequeñas.
Por eso se evalúa sobre una muestra aleatoria y se apoya con gráficos.

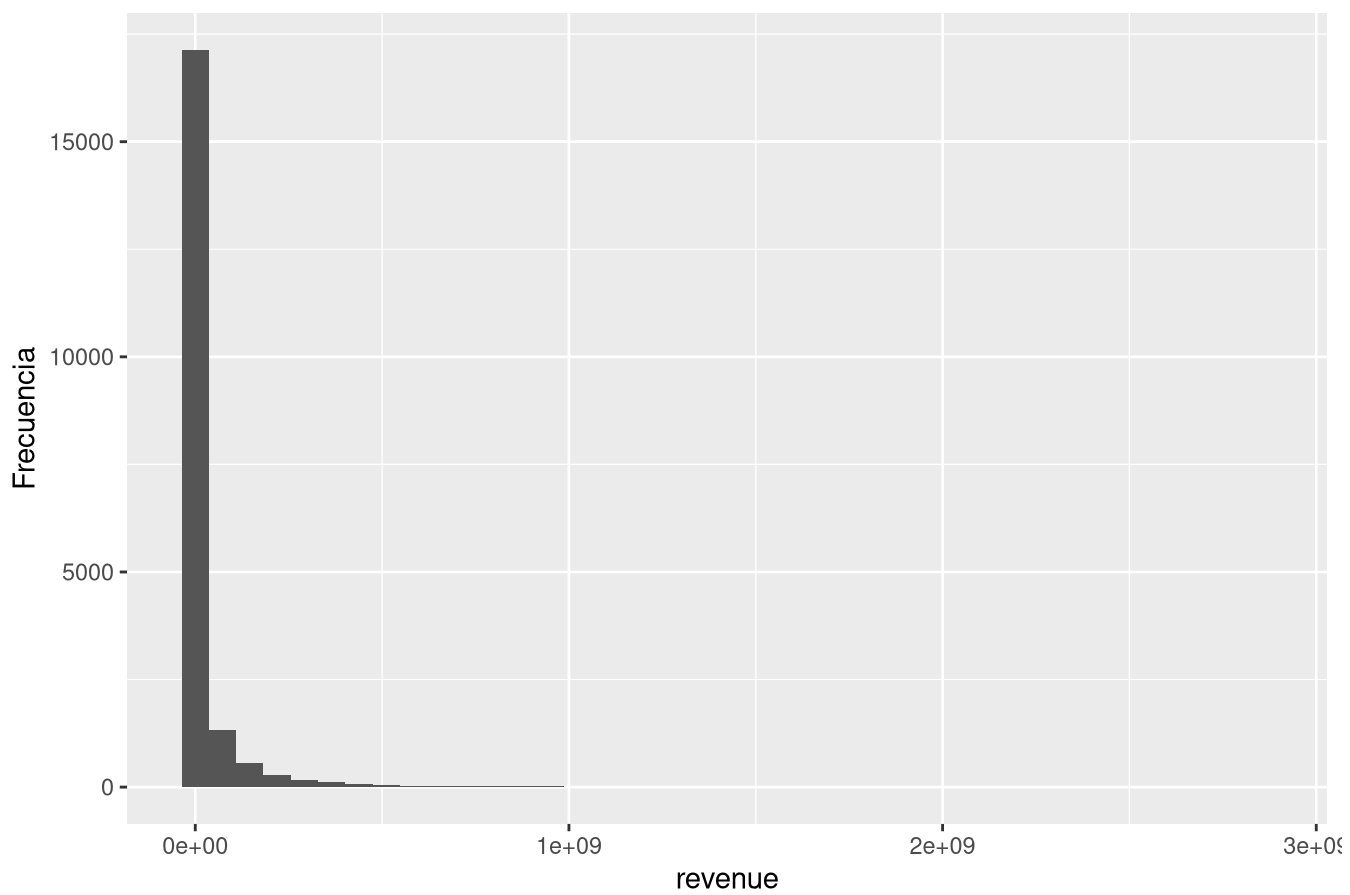
Prueba Shapiro-Wilk (muestra) para variables numéricas

variable	n	p_value
actorsAmount	5000	0
castWomenAmount	4993	0
popularity	5000	0
castMenAmount	4968	0
revenue	5000	0
productionCountriesAmount	5000	0
voteCount	5000	0
budget	5000	0
releaseYear	5000	0
id	5000	0
productionCoAmount	5000	0
voteAvg	5000	0
runtime	5000	0
genresAmount	5000	0

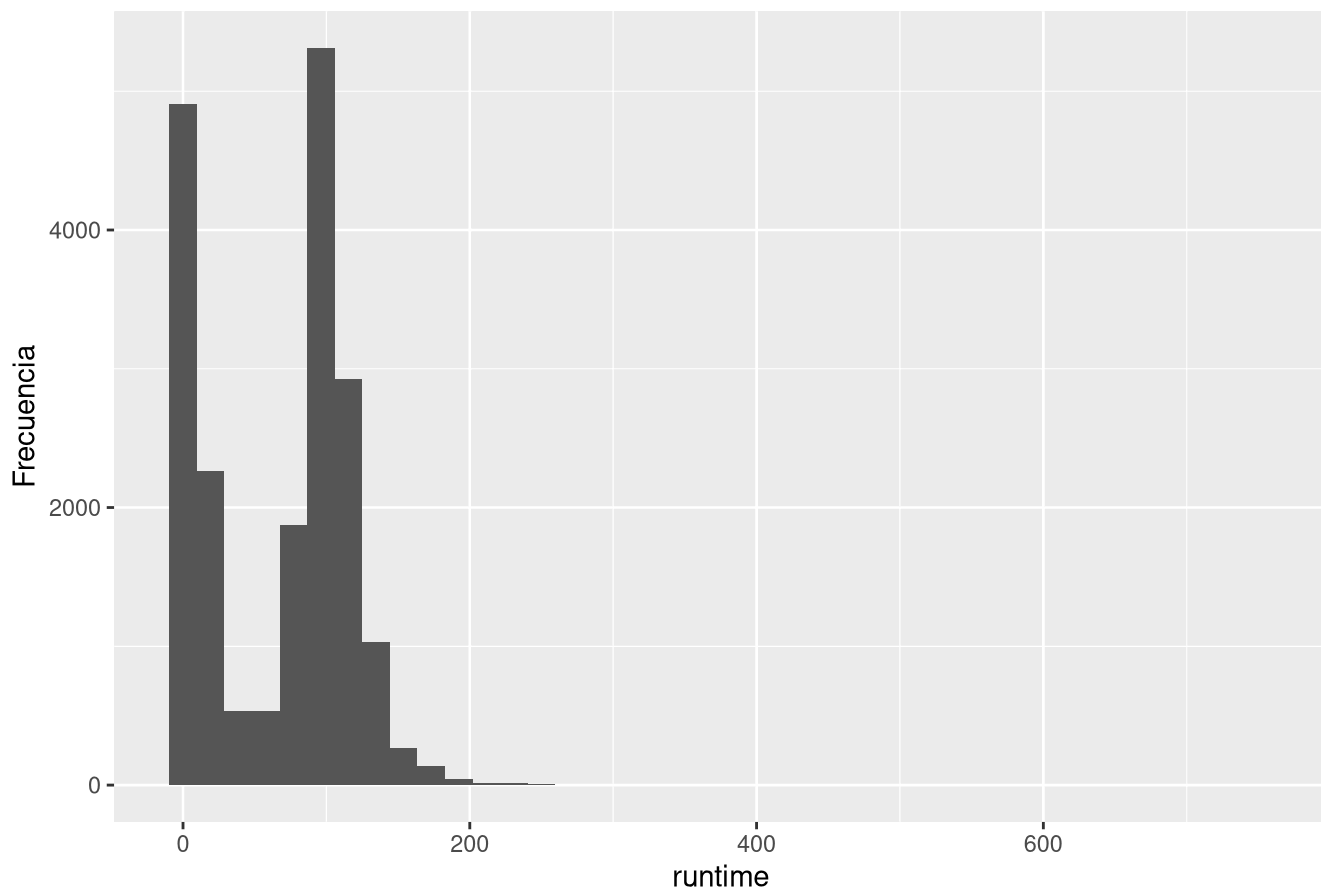
Histograma: budget



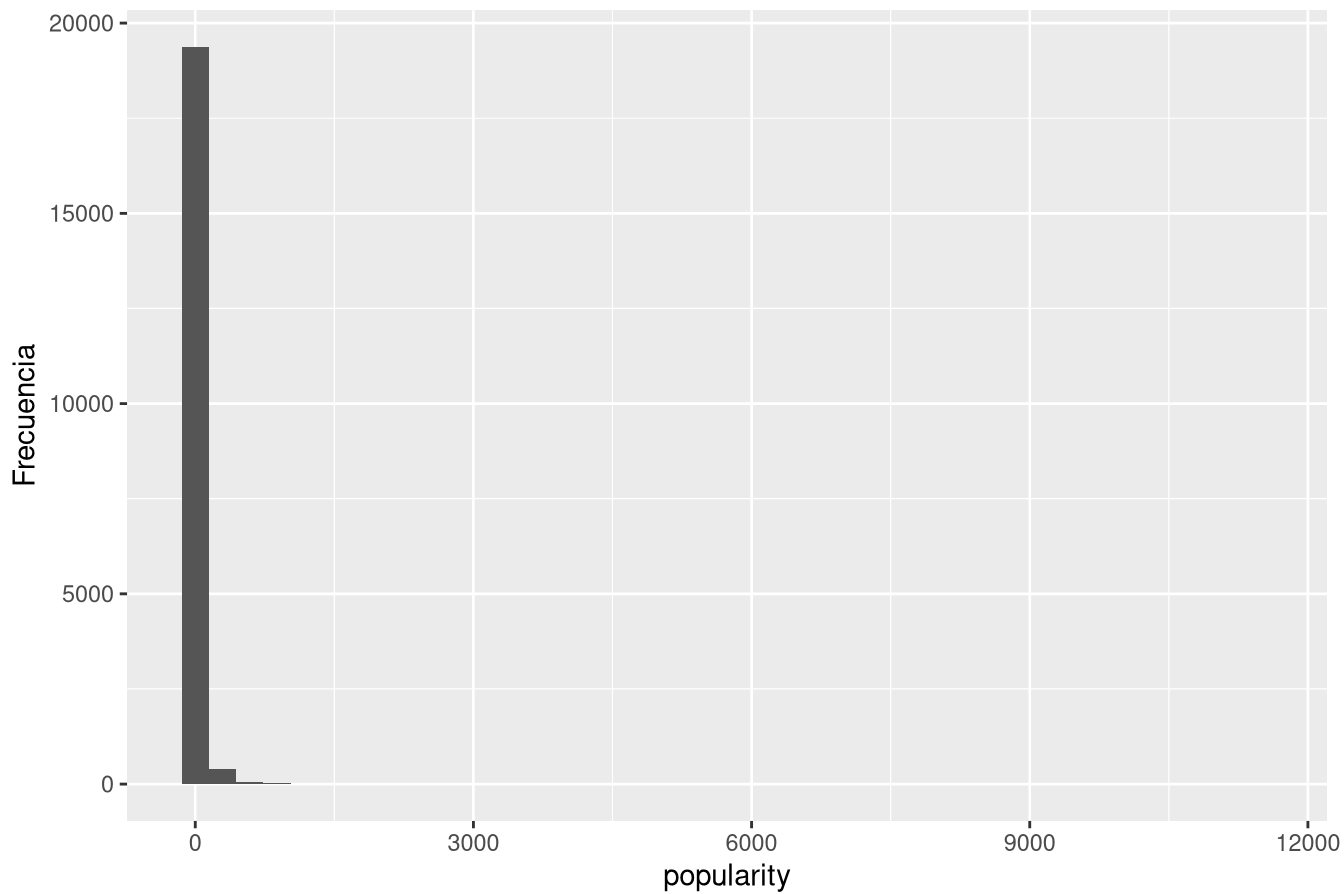
Histograma: revenue



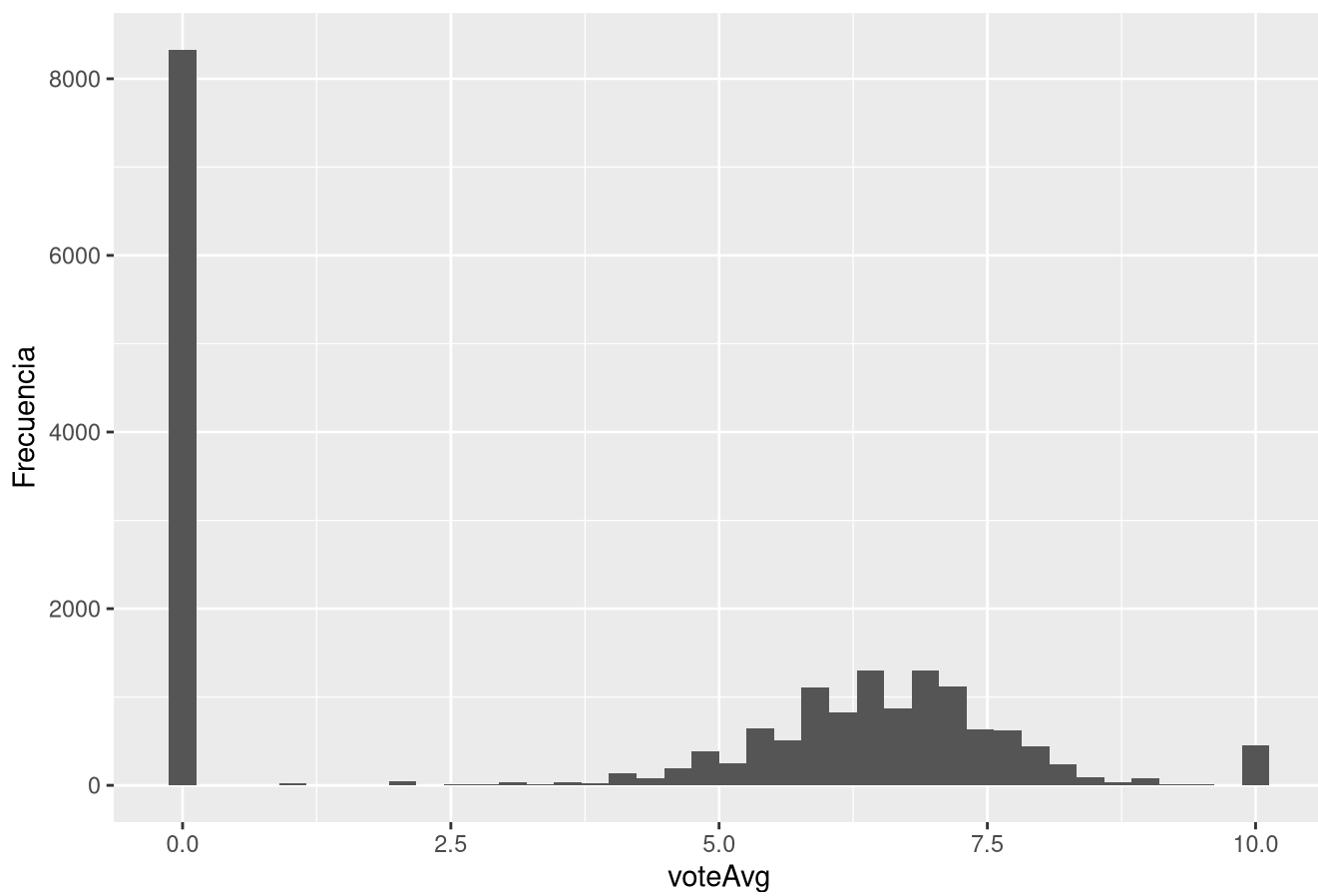
Histograma: runtime



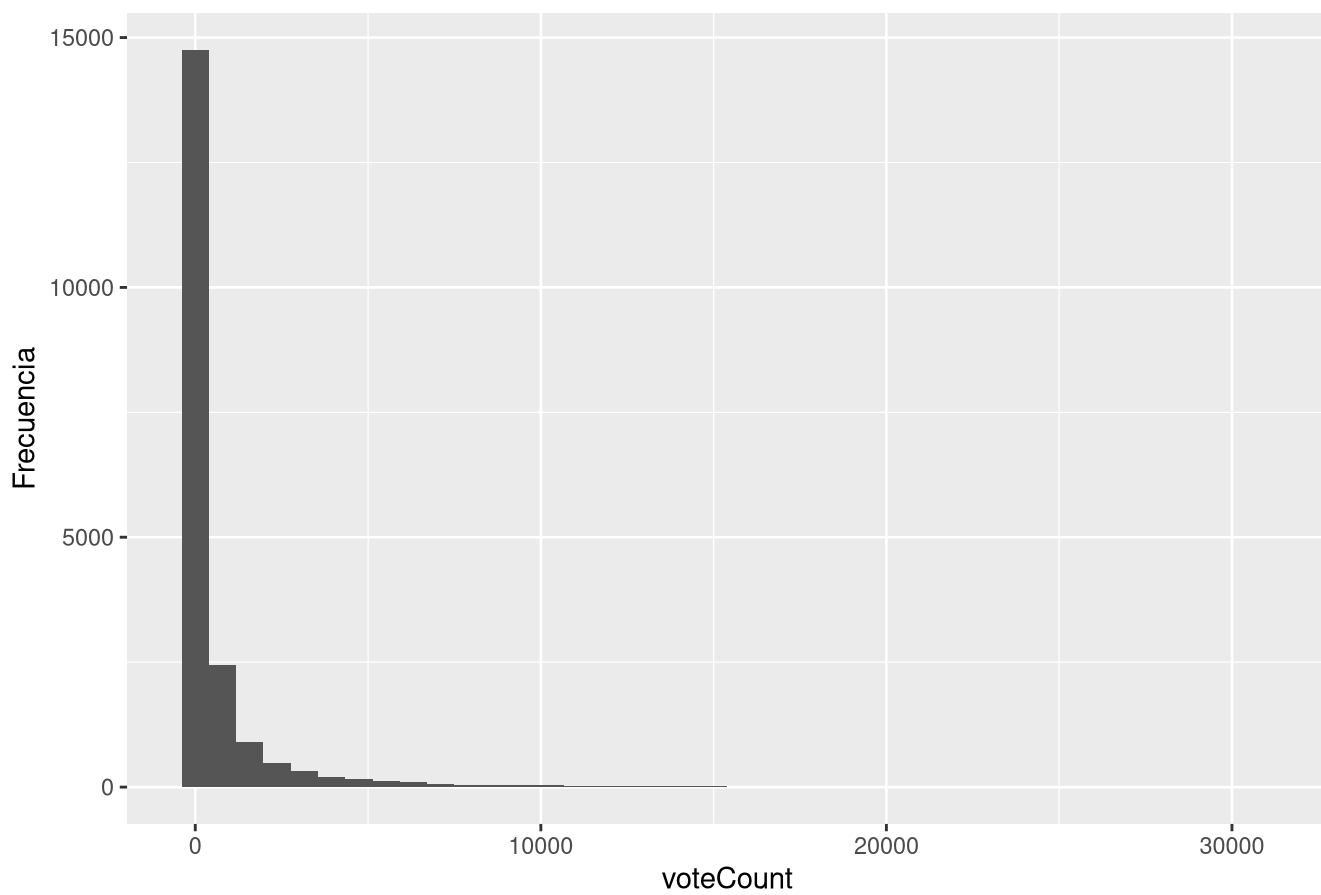
Histograma: popularity

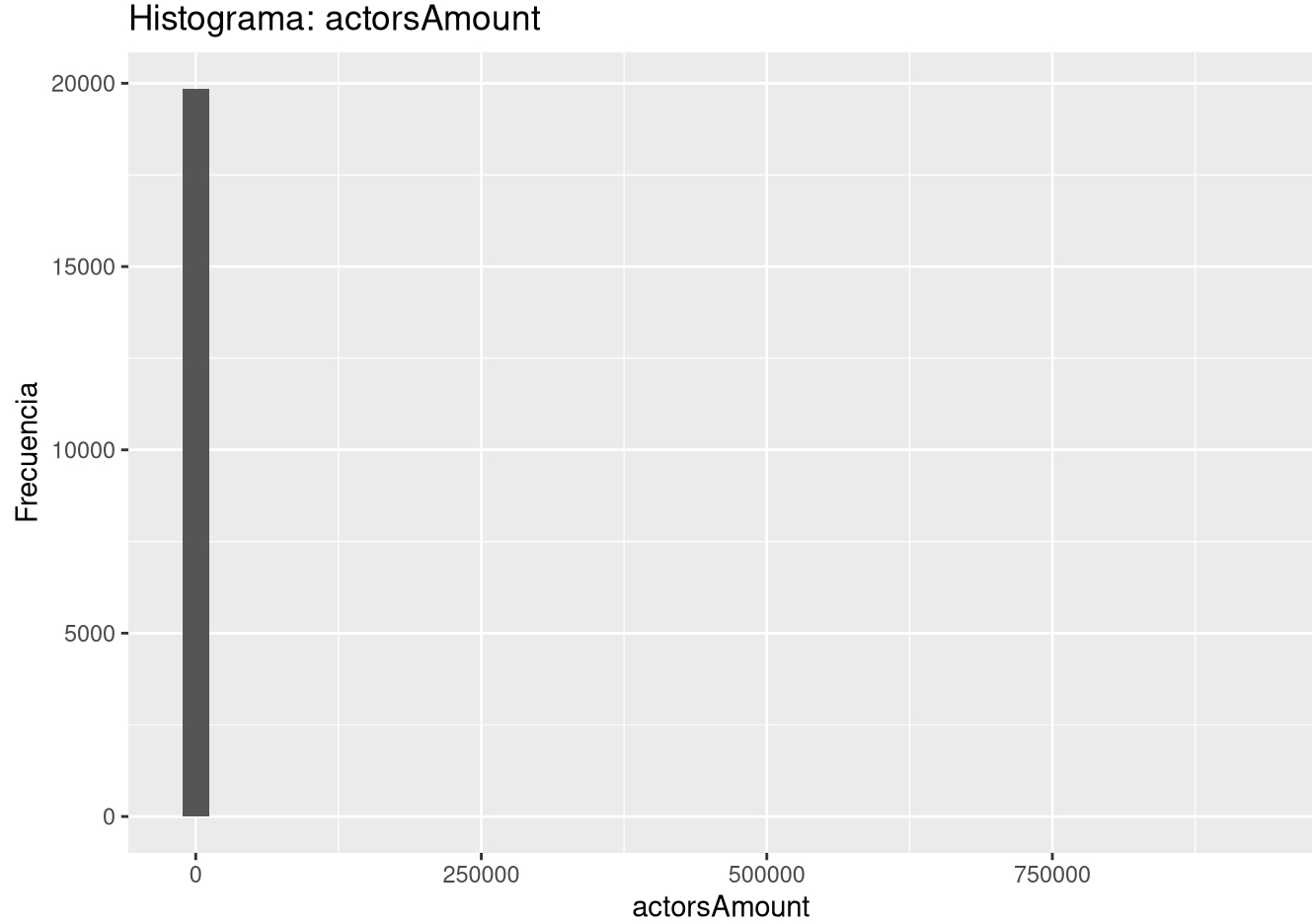


Histograma: voteAvg



Histograma: voteCount





Top 15 frecuencias por variable cualitativa (incluye NA)

categoria	frecuencia	variable
	1965	genres
Documentary	1847	genres
Drama	1705	genres
Comedy	1016	genres
Horror	502	genres
Animation	318	genres
Drama Romance	301	genres
Horror Thriller	282	genres
Comedy Drama	260	genres
Comedy Romance	250	genres
Thriller	201	genres
Music	199	genres
Drama Comedy	187	genres
Drama Thriller	171	genres
Romance Drama	161	genres
	8476	homePage

categoria	frecuencia	variable
NA	5800	homePage
https://watch.njpwworld.com/details/60769 (https://watch.njpwworld.com/details/60769)	16	homePage
http://www.naruto-movie.com/ (http://www.naruto-movie.com/)	6	homePage
https://tractorted.com/ (https://tractorted.com/)	6	homePage
http://www.avikomfilm.com (http://www.avikomfilm.com)	4	homePage
http://www.thehungergames.movie/ (http://www.thehungergames.movie/)	4	homePage
http://initiald-movie.com/ (http://initiald-movie.com/)	3	homePage
http://kizumonogatari-usa.com/ (http://kizumonogatari-usa.com/)	3	homePage
http://postbellek.com (http://postbellek.com)	3	homePage
http://www.512red.com (http://www.512red.com)	3	homePage
http://www.americanreunionmovie.com/ (http://www.americanreunionmovie.com/)	3	homePage
http://www.kungfupanda.com (http://www.kungfupanda.com)	3	homePage
http://www.kungfupanda.com/ (http://www.kungfupanda.com/)	3	homePage
http://www.missionimpossible.com/ (http://www.missionimpossible.com/)	3	homePage
	5656	productionCompany
Paramount	55	productionCompany
Universal Pictures	50	productionCompany
Warner Bros. Pictures	37	productionCompany
Toei Animation	36	productionCompany
DreamWorks Animation	33	productionCompany
The Asylum	33	productionCompany
Columbia Pictures	32	productionCompany
Marvel Studios	31	productionCompany
Walt Disney Pictures	31	productionCompany
Warner Bros. Animation	29	productionCompany
20th Century Fox	25	productionCompany
CRAV - Unisinos	25	productionCompany
Metro-Goldwyn-Mayer	25	productionCompany
Disney Television Animation	24	productionCompany
	8163	productionCompanyCountry
US	1186	productionCompanyCountry
US US	833	productionCompanyCountry

categoria	frecuencia	variable
US US US	474	productionCompanyCountry
US	314	productionCompanyCountry
	299	productionCompanyCountry
US	292	productionCompanyCountry
JP	283	productionCompanyCountry
NA	247	productionCompanyCountry
US US US US	183	productionCompanyCountry
	182	productionCompanyCountry
GB	141	productionCompanyCountry
US US	139	productionCompanyCountry
US US	135	productionCompanyCountry
US US	131	productionCompanyCountry
United States of America	4968	productionCountry
	3874	productionCountry
US	850	productionCountry
Japan	613	productionCountry
FR	428	productionCountry
GB	357	productionCountry
United Kingdom United States of America	339	productionCountry
BR	327	productionCountry
IN	315	productionCountry
DE	307	productionCountry
United Kingdom	294	productionCountry
Canada United States of America	222	productionCountry
CA	202	productionCountry
ES	178	productionCountry
AR	168	productionCountry
FALSE	19313	video
NA	486	video
TRUE	84	video
	976	director
Steven Spielberg	29	director
Clint Eastwood	28	director

categoria	frecuencia	variable
Ridley Scott	23	director
Kunihiko Yuyama	19	director
Martin Scorsese	19	director
Steven Soderbergh	19	director
Robert Zemeckis	18	director
Ron Howard	18	director
Woody Allen	17	director
Sam Liu	16	director
Tim Burton	16	director
Alfred Hitchcock	15	director
Robert Rodriguez	15	director
Antoine Fuqua	14	director
	2601	actors
FALSE	480	actors
Donovan Haessy	6	actors
TRUE	6	actors
Dylan Walker	4	actors
Gwenaëlle Bernal Jérémie Garret	3	actors
Hisayasu Satō	3	actors
Kim Nam-joon Kim Seok-jin Min Yoon-gi Jung Ho-seok Park Ji-min Kim Tae-hyung Jeon Jung-kook	3	actors
Travis Page	3	actors
Wilfrid George Kirby Clarke Ewan Ross Hastie Andrew Carstairs	3	actors
0.6	2	actors
André Rieu	2	actors
Anouck Hautbois Benjamin Bollen	2	actors
Barnaby Moore	2	actors
Bart De Pauw	2	actors
	2608	actorsPopularity
0	311	actorsPopularity
0 0	263	actorsPopularity
0 0 0	142	actorsPopularity
0.0071	133	actorsPopularity

categoria	frecuencia	variable
0 0 0 0	129	actorsPopularity
0.0143	79	actorsPopularity
0 0 0 0 0	76	actorsPopularity
0.0071 0.0071	51	actorsPopularity
0.0071 0	49	actorsPopularity
0 0.0071	47	actorsPopularity
0 0 0 0 0 0	33	actorsPopularity
0 0 0 0 0 0 0 0 0 0	31	actorsPopularity
0.0143 0.0071	26	actorsPopularity
0 0 0 0 0 0 0	25	actorsPopularity
	2835	actorsCharacter
	233	actorsCharacter
Self	197	actorsCharacter
	191	actorsCharacter
	189	actorsCharacter
	188	actorsCharacter
	165	actorsCharacter
	126	actorsCharacter
	106	actorsCharacter
Self Self Self Self Self Self Self Self Self Self	76	actorsCharacter
	63	actorsCharacter
Himself	59	actorsCharacter
Self Self	55	actorsCharacter
	43	actorsCharacter
Self Self Self	39	actorsCharacter
Cinderella	4	originalTitle
Pinocchio	4	originalTitle
<U+30C9><U+30E9><U+3048><U+3082><U+3093> <U+306E><U+3073> <U+592A><U+3068><U+96F2><U+306E><U+738B><U+56FD>	3	originalTitle
A Christmas Carol	3	originalTitle
Abduction	3	originalTitle
Aftermath	3	originalTitle
Aladdin	3	originalTitle

categoria	frecuencia	variable
Alice in Wonderland	3	originalTitle
Altitude	3	originalTitle
Arena	3	originalTitle
Awake	3	originalTitle
Carrie	3	originalTitle
Crash	3	originalTitle
Dracula	3	originalTitle
Fallen	3	originalTitle
Cinderella	5	title
Pinocchio	4	title
A Christmas Carol	3	title
Abduction	3	title
Aftermath	3	title
Aladdin	3	title
Alice in Wonderland	3	title
Altitude	3	title
Arena	3	title
Awake	3	title
Beauty and the Beast	3	title
Carrie	3	title
Crash	3	title
Desire	3	title
Dracula	3	title
en	11961	originalLanguage
es	1238	originalLanguage
fr	1094	originalLanguage
ja	868	originalLanguage
pt	628	originalLanguage
de	461	originalLanguage
zh	365	originalLanguage
ko	336	originalLanguage
it	302	originalLanguage
nl	192	originalLanguage

categoria	frecuencia	variable
ru	190	originalLanguage
id	173	originalLanguage
sv	133	originalLanguage
tl	110	originalLanguage
ar	106	originalLanguage
2026-01-30	263	releaseDate
2025-11-07	258	releaseDate
2025-11-08	239	releaseDate
2025-12-05	231	releaseDate
2025-11-20	225	releaseDate
2026-01-01	212	releaseDate
2025-11-21	207	releaseDate
2025-11-14	204	releaseDate
2025-12-12	203	releaseDate
2025-11-28	189	releaseDate
2025-12-31	183	releaseDate
2025-12-01	181	releaseDate
2025-11-13	175	releaseDate
2025-11-15	160	releaseDate
2025-11-06	156	releaseDate

4. Preguntas de negocio

4.1 ¿Cuáles son las 10 películas que contaron con más

presupuesto?

##		title	budget	releaseYear
## 1	Pirates of the Caribbean: On Stranger Tides	3.80e+08	2011	
## 2	Avengers: Age of Ultron	3.65e+08	2015	
## 3	Avengers: Endgame	3.56e+08	2019	
## 4	Avatar: Fire and Ash	3.50e+08	2025	
## 5	Pirates of the Caribbean: At World's End	3.00e+08	2007	
## 6	Justice League	3.00e+08	2017	
## 7	Avengers: Infinity War	3.00e+08	2018	
## 8	Superman Returns	2.70e+08	2006	
## 9	Tangled	2.60e+08	2010	
## 10	The Lion King	2.60e+08	2019	
##	director			
## 1	Rob Marshall			
## 2	Joss Whedon			
## 3	Anthony Russo Joe Russo			
## 4	James Cameron			
## 5	Gore Verbinski			
## 6	Zack Snyder			
## 7	Anthony Russo Joe Russo			
## 8	Bryan Singer			
## 9	Byron Howard Nathan Greno			
## 10	Jon Favreau			

4.2 ¿Cuáles son las 10 películas que más ingresos tuvieron?

##		title	revenue	budget	releaseYear
## 1		Avatar	2847246203	2.37e+08	2009
## 2		Avengers: Endgame	2797800564	3.56e+08	2019
## 3		Titanic	2187463944	2.00e+08	1997
## 4	Star Wars: The Force Awakens	2068223624	2.45e+08		2015
## 5	Avengers: Infinity War	2046239637	3.00e+08		2018
## 6	Zootopia 2	1744338246	1.50e+08		2025
## 7	Jurassic World	1671713208	1.50e+08		2015
## 8	The Lion King	1667635327	2.60e+08		2019
## 9	Spider-Man: No Way Home	1631853496	2.00e+08		2021
## 10	The Avengers	1518815515	2.20e+08		2012
##	director				
## 1	James Cameron				
## 2	Anthony Russo Joe Russo				
## 3	James Cameron				
## 4	J.J. Abrams				
## 5	Anthony Russo Joe Russo				
## 6	Jared Bush Byron Howard				
## 7	Colin Trevorrow				
## 8	Jon Favreau				
## 9	Jon Watts				
## 10	Joss Whedon				

4.3 ¿Cuál es la película que más votos tuvo?

```
##           title voteCount voteAvg
## 13402 Inception      30788      8.4
```

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

Se interpreta como la película con menor **voteAvg**.

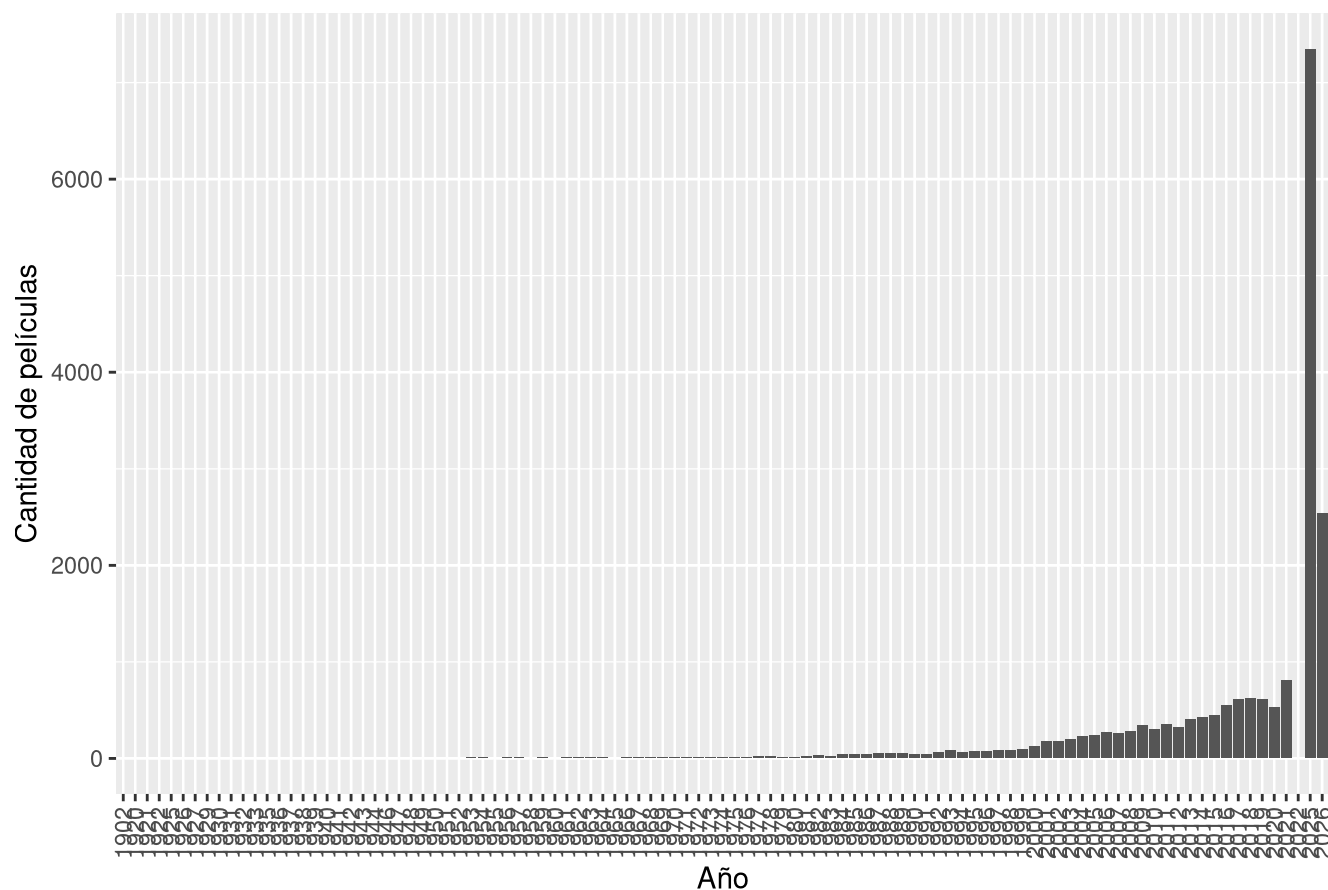
Para evitar casos con muy pocos votos (p.ej. 1 voto), se muestra también el peor caso considerando un umbral de votos.

```
## $peor_sin_filtro
##           title voteAvg voteCount
## 1 The Halloween Harvest          0          1
##
## $peor_con_umbral
##           title voteAvg voteCount
## 1 Death Note      4.2      3388
```

4.5 ¿Cuántas películas se hicieron en cada año? ¿En qué año se hicieron más películas? (gráfico de barras)

```
##      releaseYear  peliculas
## 1          2025      7351
## 2          2026      2537
## 3          2021       814
## 4          2018       628
## 5          2017       617
## 6          2019       611
## 7          2016       557
## 8          2020       531
## 9          2015       450
## 10         2014       432
```


Películas por año (máximo en 2025 con 7351 películas)



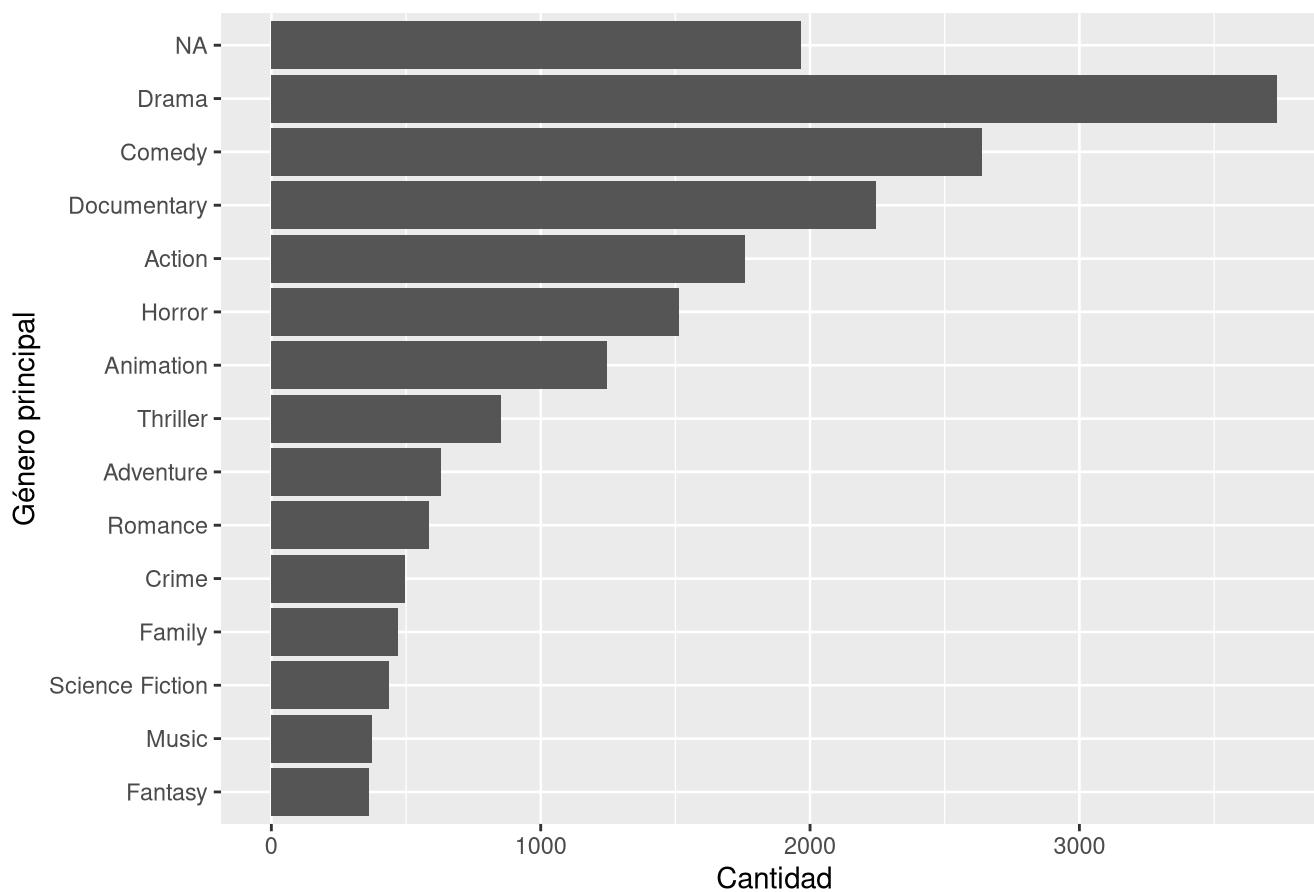
4.6 Género principal de las 20 películas más recientes, género predominante y género de películas más largas

##	title	releaseDate	main_genre
## 1	A Fading Man	2026-05-07	Drama
## 2	Elon Musk Unveiled - The Tesla Experiment	2026-03-12	Documentary
## 3	Skunk	2026-02-25	<NA>
## 4	Anastasia	2026-02-25	<NA>
## 5	Nikki hako no koi	2026-02-06	<NA>
## 6	Immersed	2026-02-01	Drama
## 7	Cinderella	2026-02-01	Animation
## 8	Aladdin	2026-02-01	Animation
## 9	THE RING AND THE DECK	2026-02-01	Thriller
## 10	Crimson High 3	2026-02-01	Animation
## 11	Conversations with Rasparagus Asparagus Baragus	2026-02-01	Comedy
## 12	Highway To Hell	2026-02-01	Comedy
## 13	Pari's daughter	2026-02-01	Drama
## 14	Escort	2026-02-01	Action
## 15	Dream	2026-02-01	Drama
## 16	Lively	2026-02-01	Drama
## 17	01-02-2026 فایق سواری در تهران	2026-02-01	Romance
## 18	Midnight	2026-02-01	War
## 19	Emir - Posljednji dalmatinski težak	2026-02-01	<NA>
## 20	Our Dead Husband	2026-02-01	Thriller

```
##      main_genre n
## 1      Drama 5
## 2      <NA> 4
## 3    Animation 3
## 4      Comedy 2
## 5    Thriller 2
## 6      Action 1
## 7 Documentary 1
## 8      Romance 1
## 9        War 1
```

```
##      main_genre    n
## 1      Drama 3734
## 2      Comedy 2640
## 3    Documentary 2245
## 4      <NA> 1965
## 5      Action 1760
## 6      Horror 1513
## 7    Animation 1247
## 8      Thriller  854
## 9    Adventure  629
## 10     Romance  585
## 11     Crime   497
## 12     Family  469
## 13 Science Fiction 437
## 14      Music   373
## 15     Fantasy  364
```

Top 15 géneros principales (dataset completo)



##		title	runtime
## 1	How Ponyo was Born ~Hayao Miyazaki's Thought Process~		750
## 2		Before the End	675
## 3	League of Legends Worlds25 - Finals in Cinema		420
## 4	Crystal Lake Memories: The Complete History of Friday the 13th		400
## 5		In Search of Darkness: 1995-1999	384
## 6		Vu 2025 (l'année du zapping)	366
## 7		Napoleon	333
## 8		White Gardenia: The King James Bible	330
## 9		NJPW Wrestle Kingdom 20	326
## 10		1900	317
## 11		The Idolmaster Gakuen: Story of Re;IRIS	284
## 12		Shpongles - Live at Mission Ballroom	263
## 13		Indie Beat	261
## 14		2025 Rock & Roll Hall of Fame Induction Ceremony	261
## 15		Marigold First Dream 2026	253
## 16	Andrea Bocelli - The Celebration: 30th Anniversary		253
## 17		The Metropolitan Opera: Arabella	252
## 18		Cleopatra	248
## 19		Kill Bill: The Whole Bloody Affair	247
## 20		Hamlet	242

##	main_genre	releaseYear
## 1	Documentary	2009
## 2	Action	2025
## 3	<NA>	2025
## 4	Documentary	2013
## 5	Documentary	2025
## 6	<NA>	2025
## 7	Drama	1927
## 8	Horror	2026
## 9	<NA>	2026
## 10	Drama	1976
## 11	Animation	2026
## 12	<NA>	2025
## 13	Music	2025
## 14	Music	2025
## 15	Action	2026
## 16	Documentary	2025
## 17	<NA>	2025
## 18	Drama	1963
## 19	Action	2011
## 20	Drama	1996

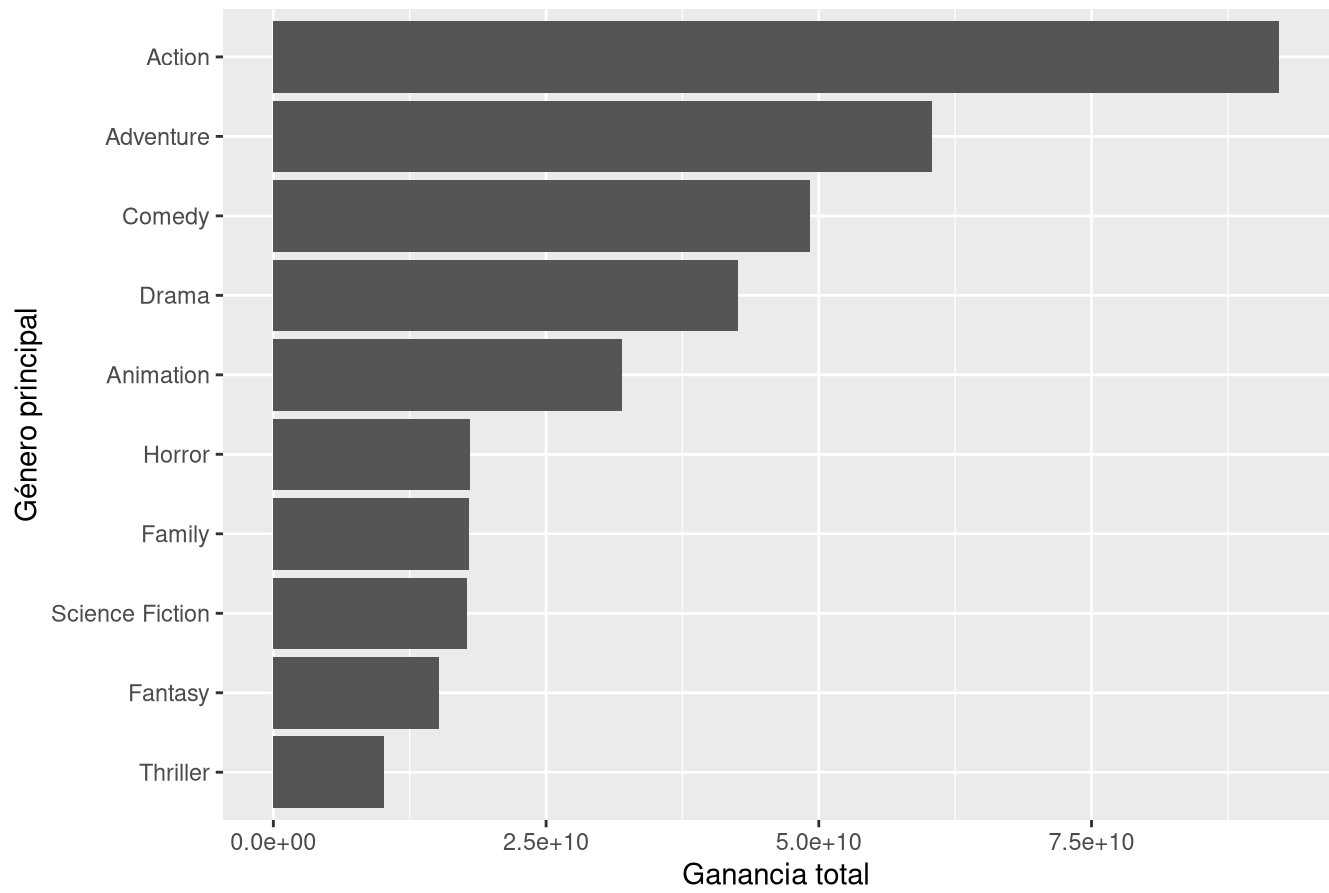
##	main_genre	n
## 1	<NA>	5
## 2	Documentary	4
## 3	Drama	4
## 4	Action	3
## 5	Music	2
## 6	Animation	1
## 7	Horror	1

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

Se calcula **ganancia** como `revenue - budget` para películas con ambos valores positivos.

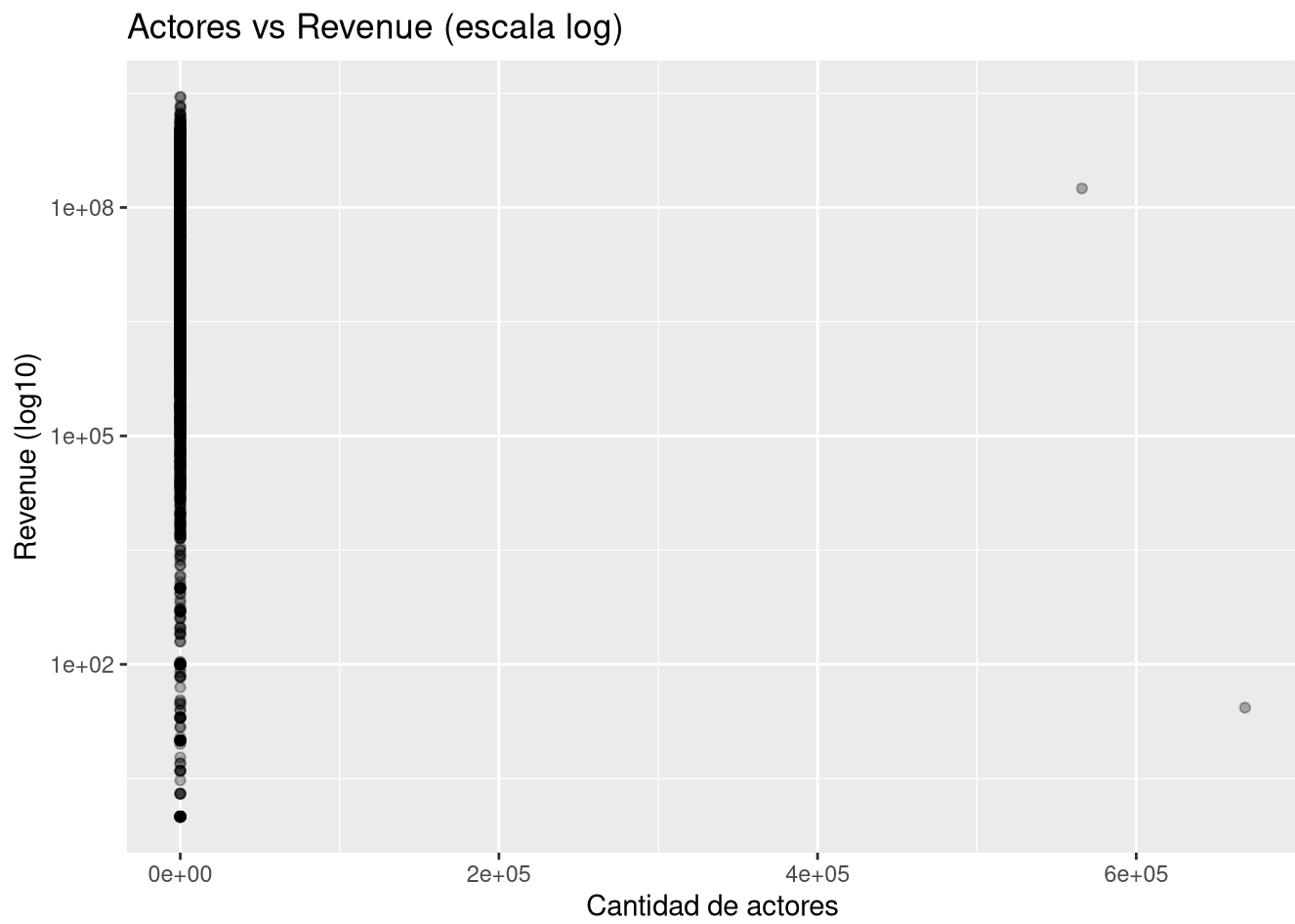
```
## # A tibble: 15 × 4
##   main_genre      peliculas ganancia_total ganancia_promedio
##   <chr>          <int>         <dbl>          <dbl>
## 1 Action           779      92195613009      118351236.
## 2 Adventure        351      60354070276      171948918.
## 3 Comedy           805      49220834911       61143894.
## 4 Drama            860      42588500938       49521513.
## 5 Animation        181      32005682028      176826973.
## 6 Horror           373      18026440092       48328258.
## 7 Family           132      17978264869      136198976.
## 8 Science Fiction  145      17757646253      122466526.
## 9 Fantasy           137      15146197026      110556183.
## 10 Thriller        201      10151243101       50503697.
## 11 Crime            211       8254157389       39119229.
## 12 Romance          106       6325555484       59675052.
## 13 War               41       3379441894       82425412.
## 14 Mystery           60       3378308280       56305138.
## 15 Music             35       2174083494       62116671.
```

Top 10 géneros por ganancia total (revenue - budget)



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

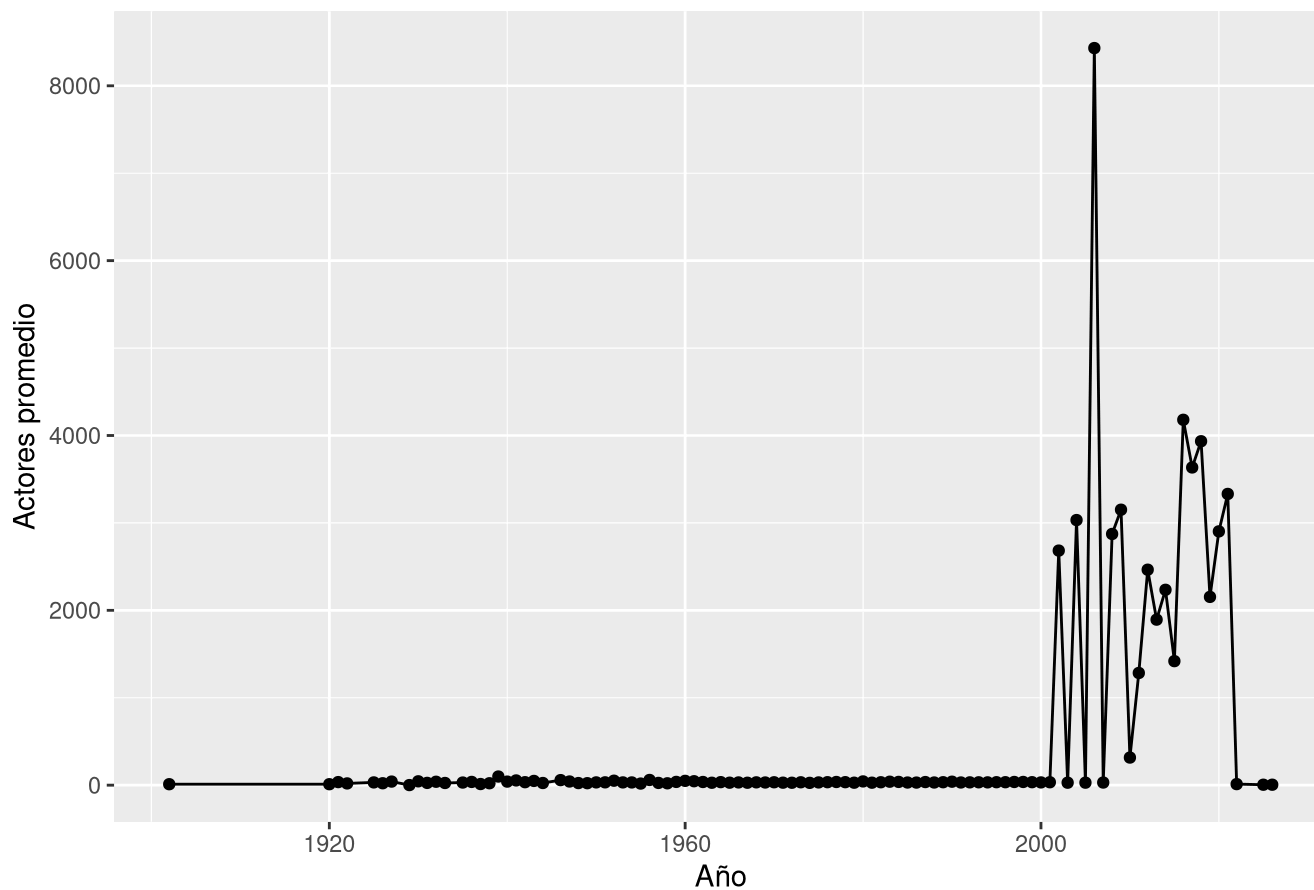
```
## [1] 0.4469277
```



```
## # A tibble: 10 × 3
```

	releaseYear	actores_prom	peliculas
	<int>	<dbl>	<int>
## 1	2015	1419.	445
## 2	2016	4179.	548
## 3	2017	3634.	611
## 4	2018	3934.	623
## 5	2019	2154.	599
## 6	2020	2903.	520
## 7	2021	3331.	781
## 8	2022	12.1	7
## 9	2025	5.10	5412
## 10	2026	5.22	1955

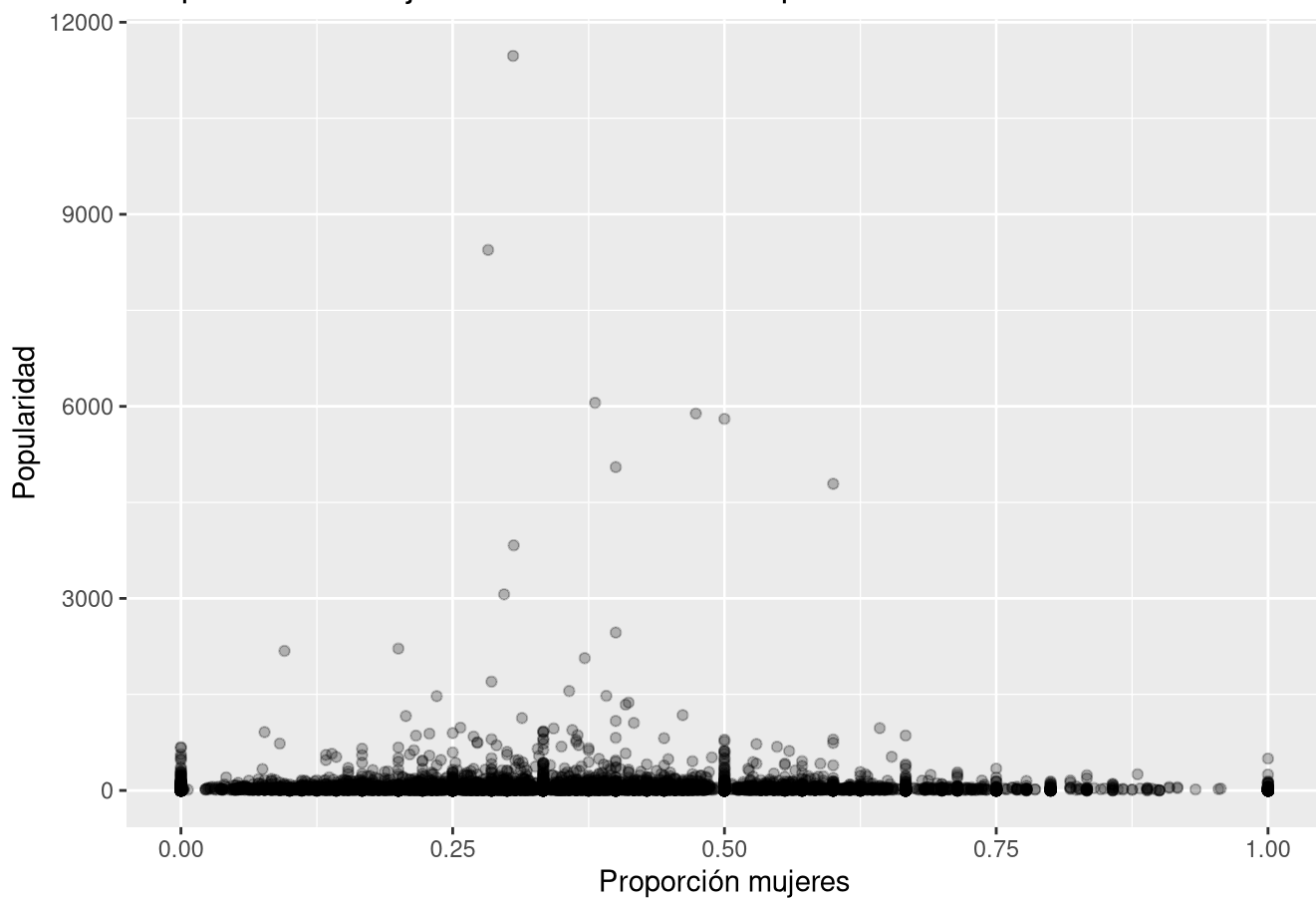
Promedio de actores por año



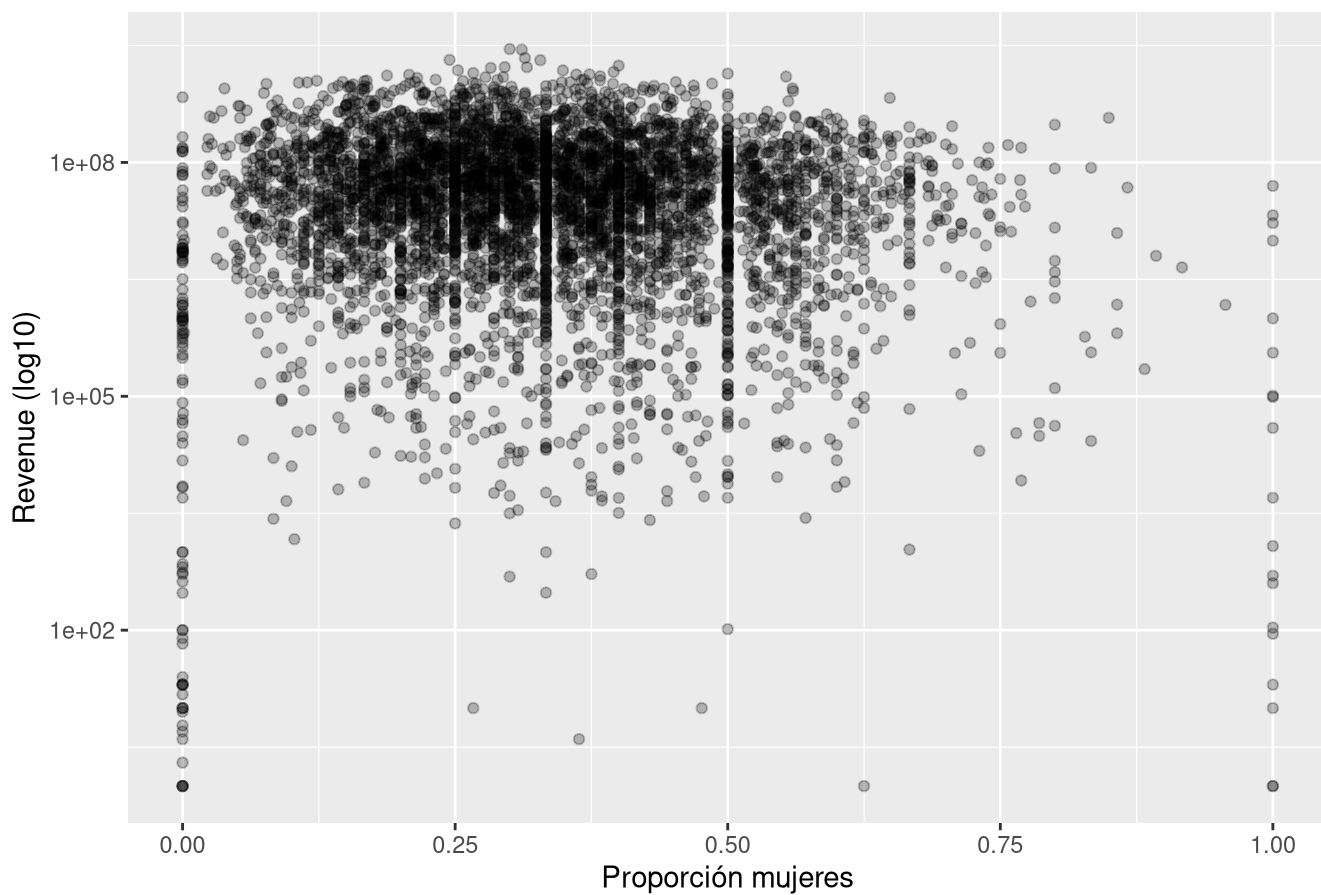
4.9 ¿La cantidad de hombres y mujeres en el reparto influye en la popularidad y los ingresos?

```
## $corr_prop_mujeres_popularity
## [1] 0.09198082
##
## $corr_prop_mujeres_revenue
## [1] -0.004383732
```

Proporción de mujeres en el elenco vs Popularidad



Proporción de mujeres en el elenco vs Revenue (log)



4.10 ¿Quiénes son los directores que hicieron las 20

películas mejor calificadas?

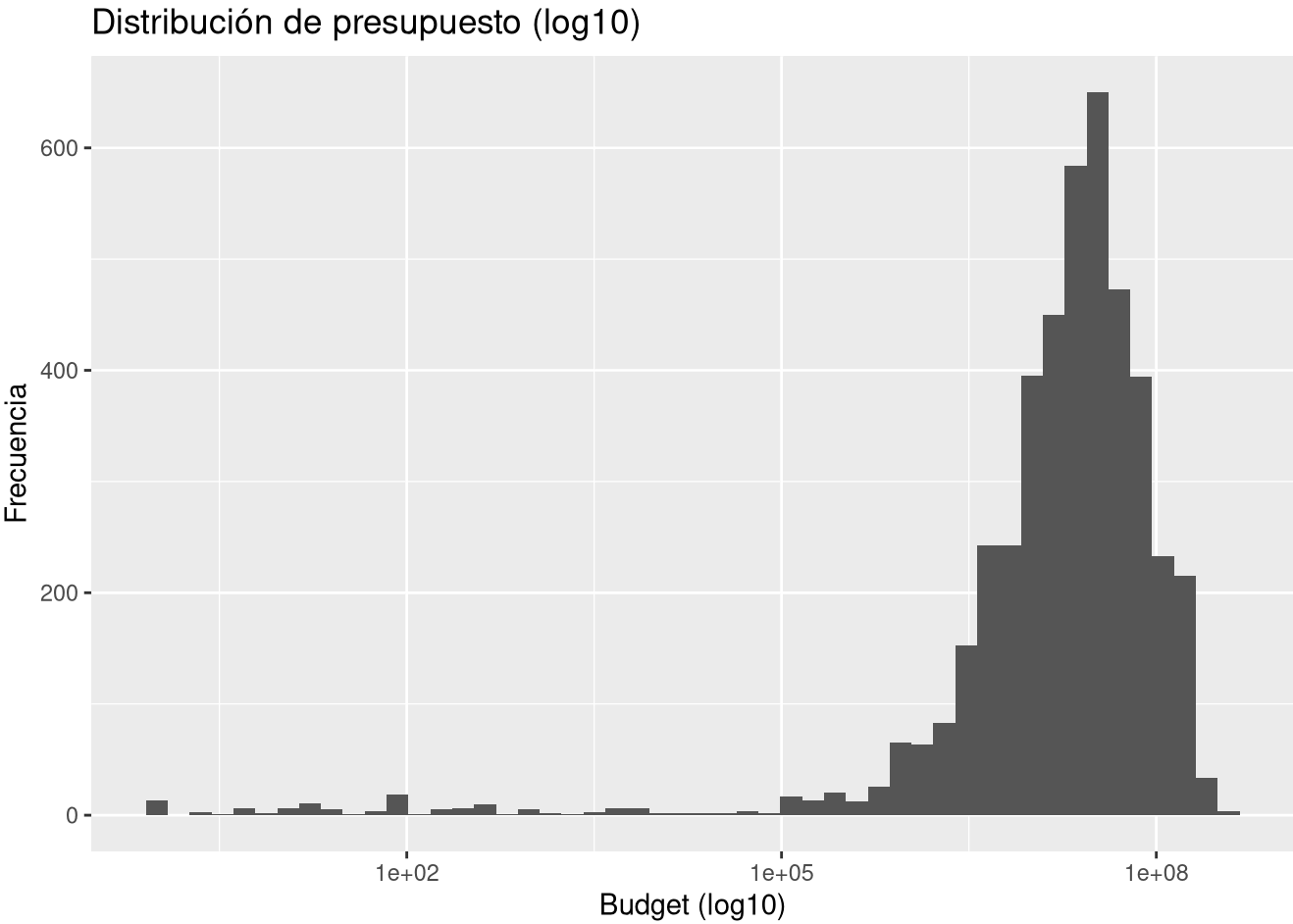
Para hacerlo más robusto, se filtra al 10% de películas con más votos y luego se ordena por **voteAvg** y **voteCount**.

##	director	title
## 10050	Frank Darabont	The Shawshank Redemption
## 10026	Francis Ford Coppola	The Godfather
## 13070	Aditya Chopra	Dilwale Dulhania Le Jayenge
## 10108	Steven Spielberg	Schindler's List
## 10028	Francis Ford Coppola	The Godfather: Part II
## 16495	Makoto Shinkai	Your Name.
## 19218	Tosca Musk	Gabriel's Inferno
## 9978	Christopher Nolan	The Dark Knight
## 10265	Quentin Tarantino	Pulp Fiction
## 9895	Robert Zemeckis	Forrest Gump
## 9959	Peter Jackson	The Lord of the Rings: The Return of the King
## 10140	Frank Darabont	The Green Mile
## 17808	Bong Joon-ho	Parasite
## 9962	Hayao Miyazaki	Spirited Away
## 10232	Roberto Benigni	Life Is Beautiful
## 10304	Martin Scorsese	GoodFellas
## 10112	Sergio Leone	The Good, the Bad and the Ugly
## 10090	Sidney Lumet	12 Angry Men
## 18938	Jon Watts	Spider-Man: No Way Home
## 12087	Giuseppe Tornatore	Cinema Paradiso
##	voteCount	voteAvg
## 10050	20598	8.7
## 10026	15380	8.7
## 13070	3372	8.7
## 10108	12282	8.6
## 10028	9266	8.6
## 16495	8274	8.6
## 19218	2188	8.6
## 9978	26690	8.5
## 10265	22501	8.5
## 9895	22045	8.5
## 9959	18952	8.5
## 10140	13380	8.5
## 17808	12979	8.5
## 9962	12339	8.5
## 10232	10781	8.5
## 10304	9741	8.5
## 10112	6385	8.5
## 10090	6127	8.5
## 18938	5630	8.5
## 12087	3210	8.5

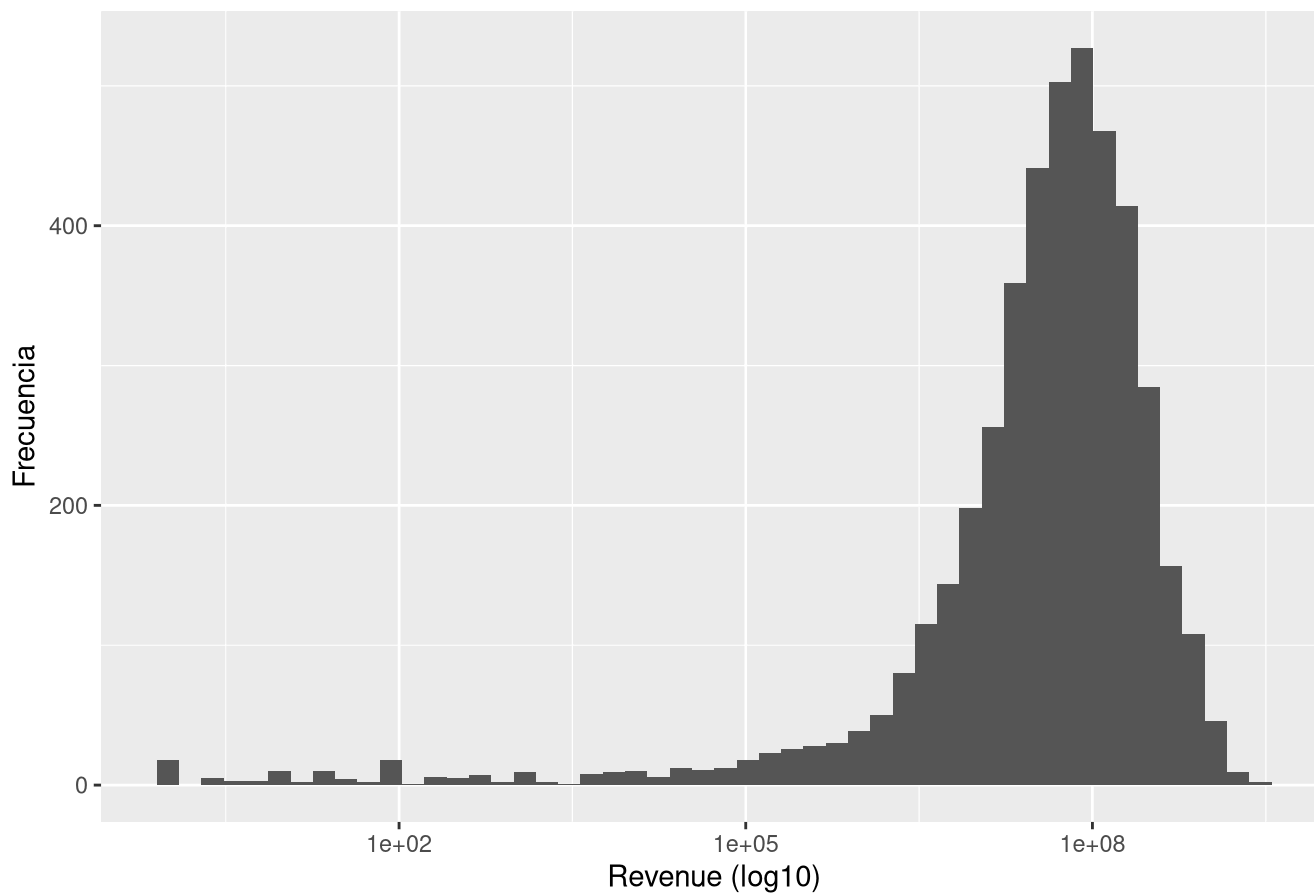
4.11 ¿Cómo se correlacionan los presupuestos con los

ingresos? (histograma y dispersión)

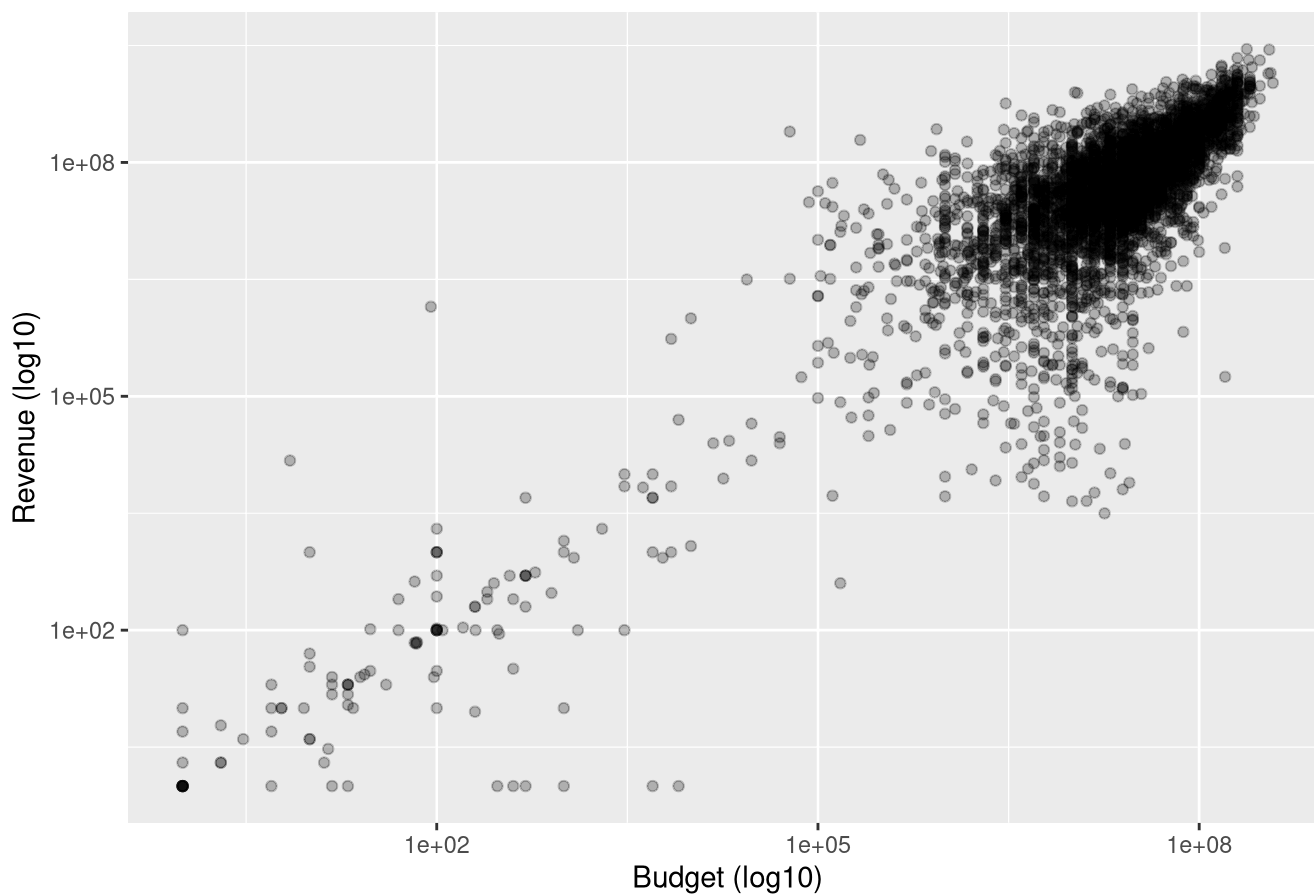
[1] 0.7057162



Distribución de ingresos (log10)

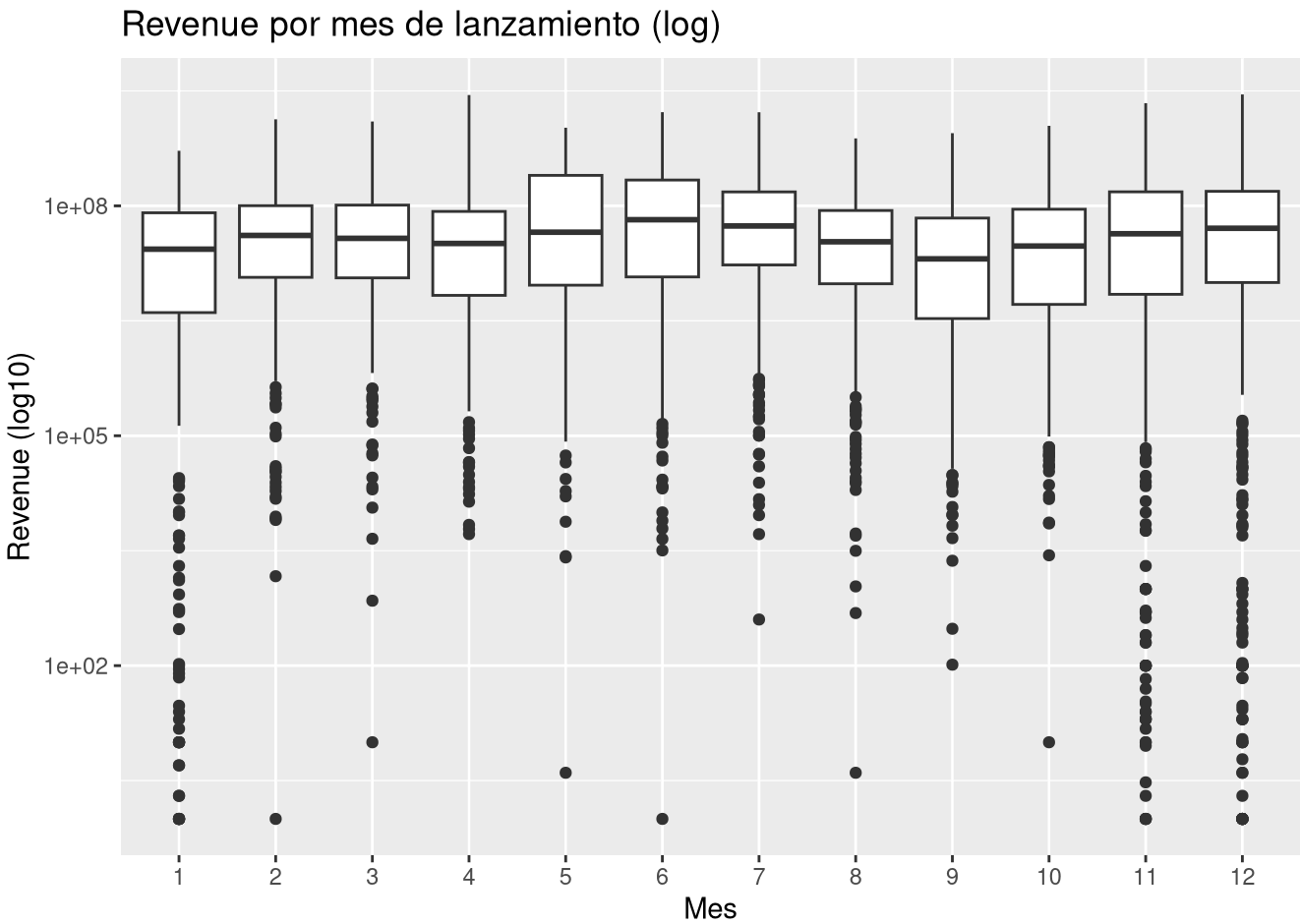


Budget vs Revenue (log-log)



4.12 ¿Se asocian ciertos meses de lanzamiento con

mejores ingresos?

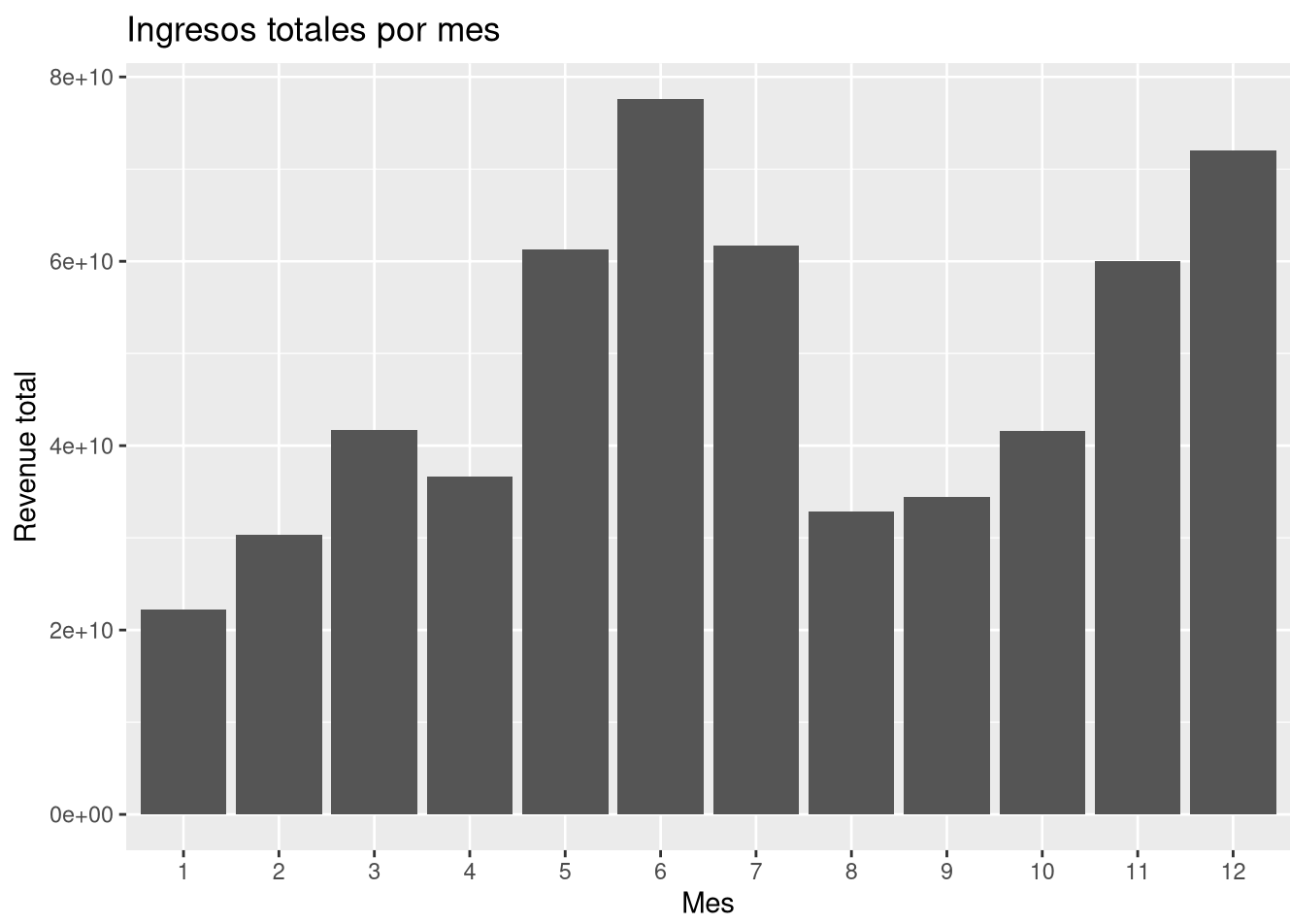


```
## # A tibble: 12 × 4
##   releaseMonth películas revenue_prom revenue_mediana
##       <int>      <int>      <dbl>         <dbl>
## 1         6        468  165807439.    66001002
## 2         5        371  165272557.    45361000
## 3         7        465  132764089.    54682547
## 4        11        470  127811630.    43278503
## 5        12        579  124411468.    51053787
## 6         4        358  102252891.    32339075
## 7         3        429   97108375.    37713879
## 8        10        489   85148627.    29918745
## 9         2        366   82768572.    41146058
## 10        8        464   70777333.    34032922.
## 11        1        368   60407604.    27233270
## 12        9        571   60335029.    20350754
```

4.13 ¿En qué meses se han visto los lanzamientos con mejores ingresos? ¿Cuántas películas, en promedio, se

han lanzado por mes?

```
## # A tibble: 12 × 4
##   Mes revenue_total revenue_prom peliculas
##   <int>         <dbl>         <dbl>     <int>
## 1     6  77597881637    94747108.       819
## 2    12  72034239751    16517826.     4361
## 3     7  61735301475    76028696.       812
## 4     5  61316118519    87719769.       699
## 5    11  60071466034    12700099.     4730
## 6     3  41659492701    51053300.       816
## 7    10  41637678840    39133157.     1064
## 8     4  36606534881    52671273.       695
## 9     9  34451301833    31958536.     1078
## 10     8  32840682550    35930725.       914
## 11     2  30293297436    39138627.       774
## 12     1  22229998382     7127284.     3119
```

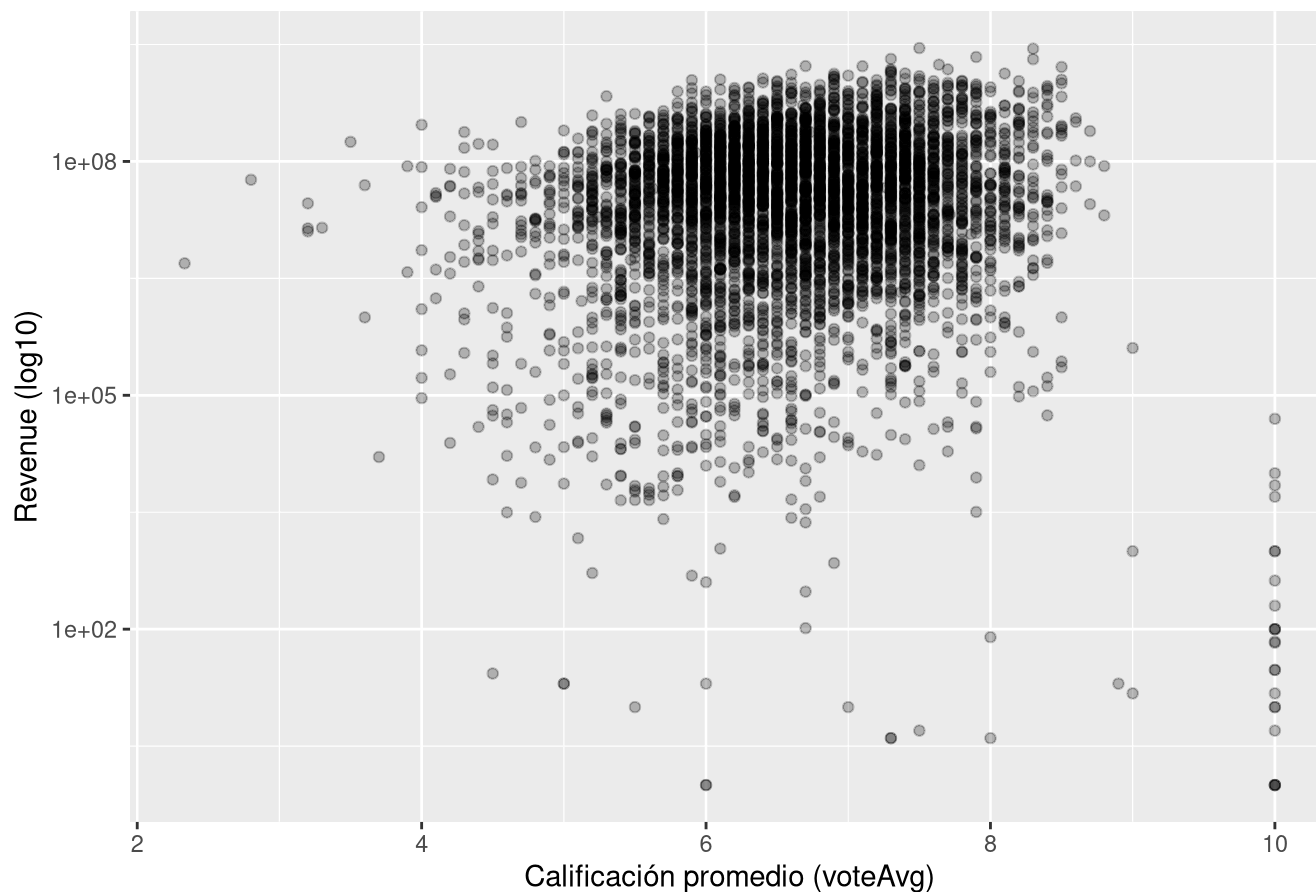


```
## [1] 1656.75
```

4.14 ¿Cómo se correlacionan las calificaciones con el éxito comercial?

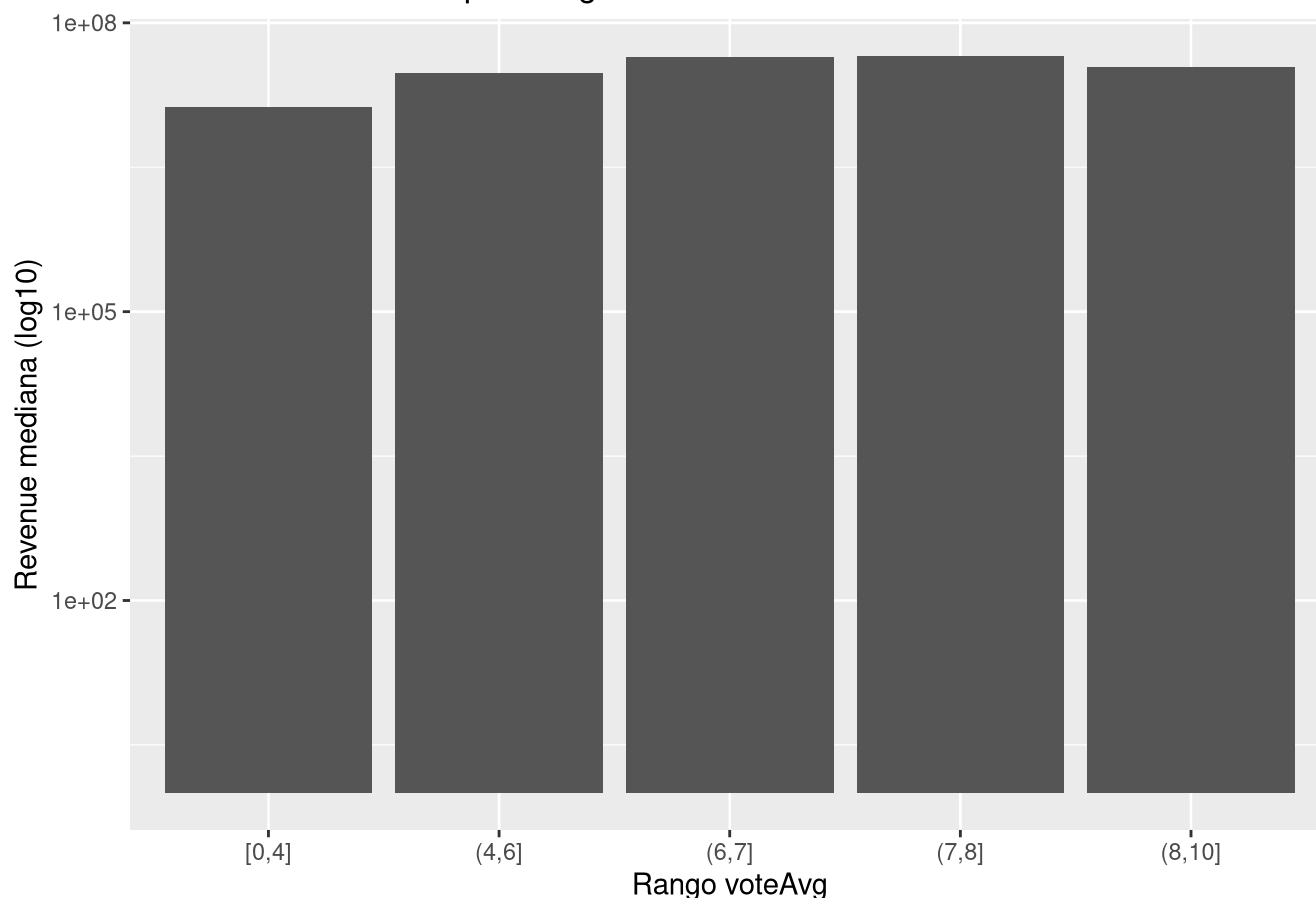
```
## [1] 0.1063246
```

VoteAvg vs Revenue (log)



```
## # A tibble: 5 × 3
##   bin    revenue_mediana películas
##   <fct>      <dbl>      <int>
## 1 [0,4]    13255913         20
## 2 (4,6]    30068418        1280
## 3 (6,7]    44400000        2335
## 4 (7,8]    45150000        1452
## 5 (8,10]   34934009         209
```

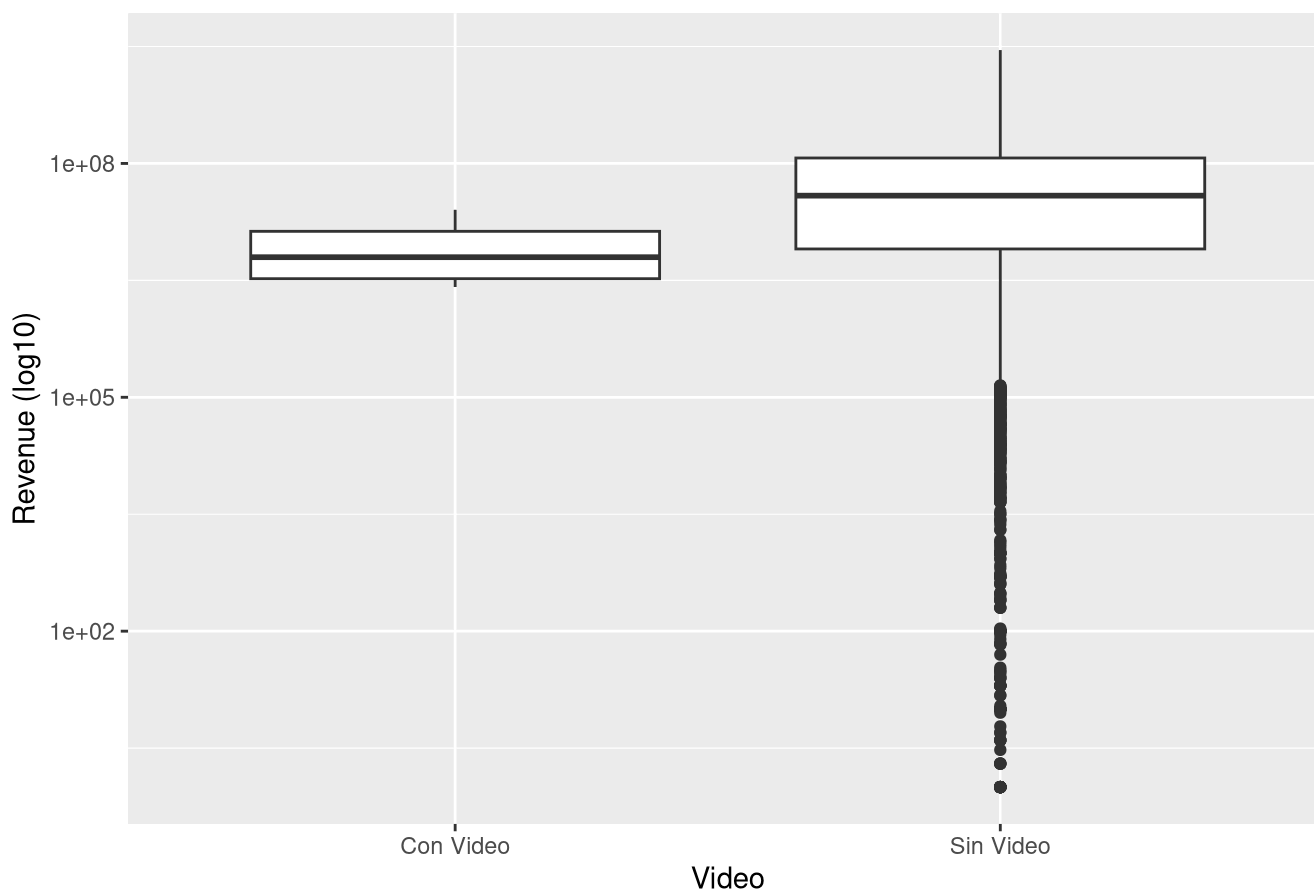
Mediana de revenue por rango de calificación



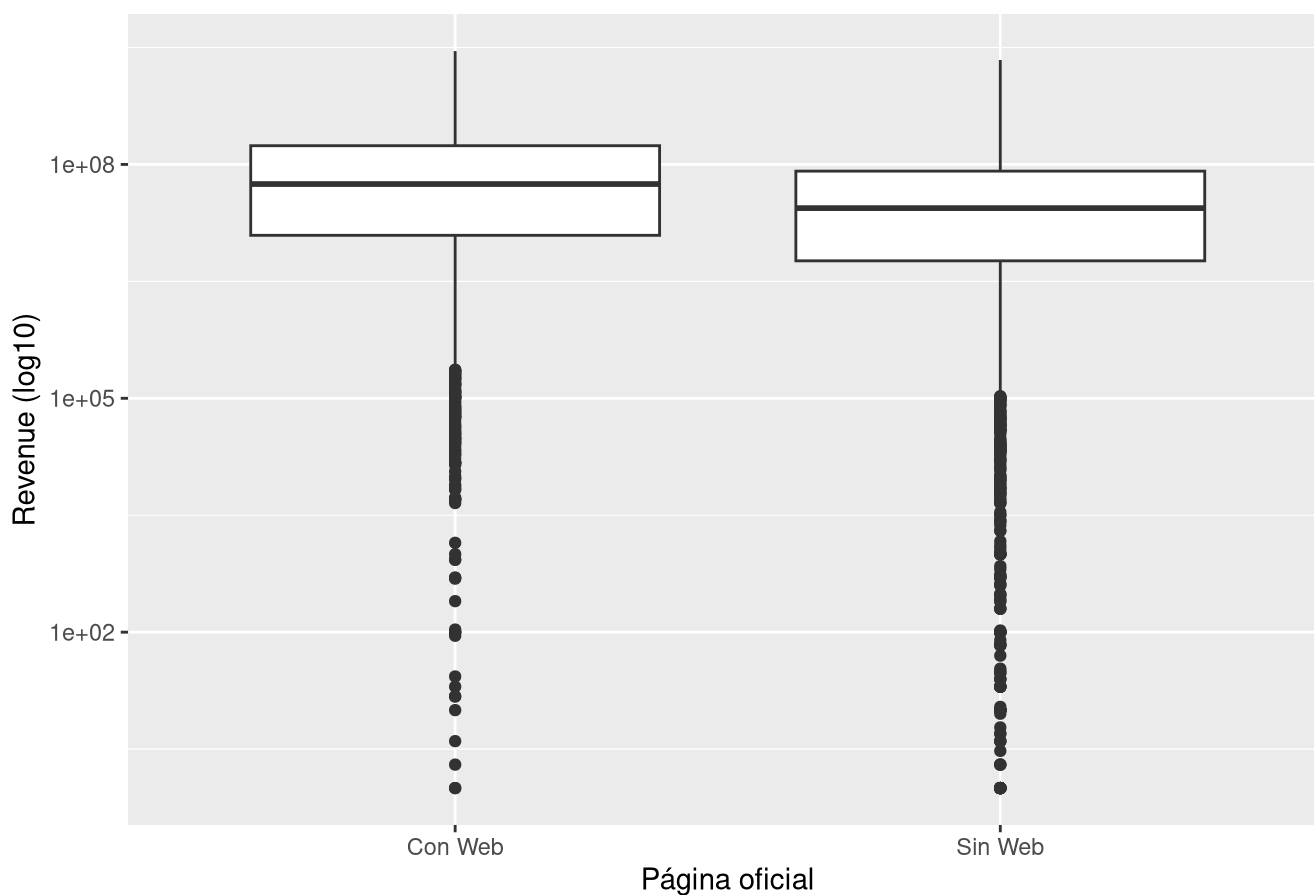
4.15 ¿Qué estrategias de marketing, como videos promocionales o páginas oficiales, generan mejores resultados?

```
## # A tibble: 4 × 5
## # Groups:   has_homepage [2]
##   has_homepage has_video peliculas revenue_prom popularity_prom
##   <chr>         <chr>      <int>      <dbl>      <dbl>
## 1 Con Web      Sin Video    2537    150830255.    85.2
## 2 Sin Web      Sin Video    2857     66424609.    31.0
## 3 Sin Web      Con Video      1     25422088     15.5
## 4 Con Web      Con Video      3      5702633      18.3
```

Revenue según presencia de video (log)



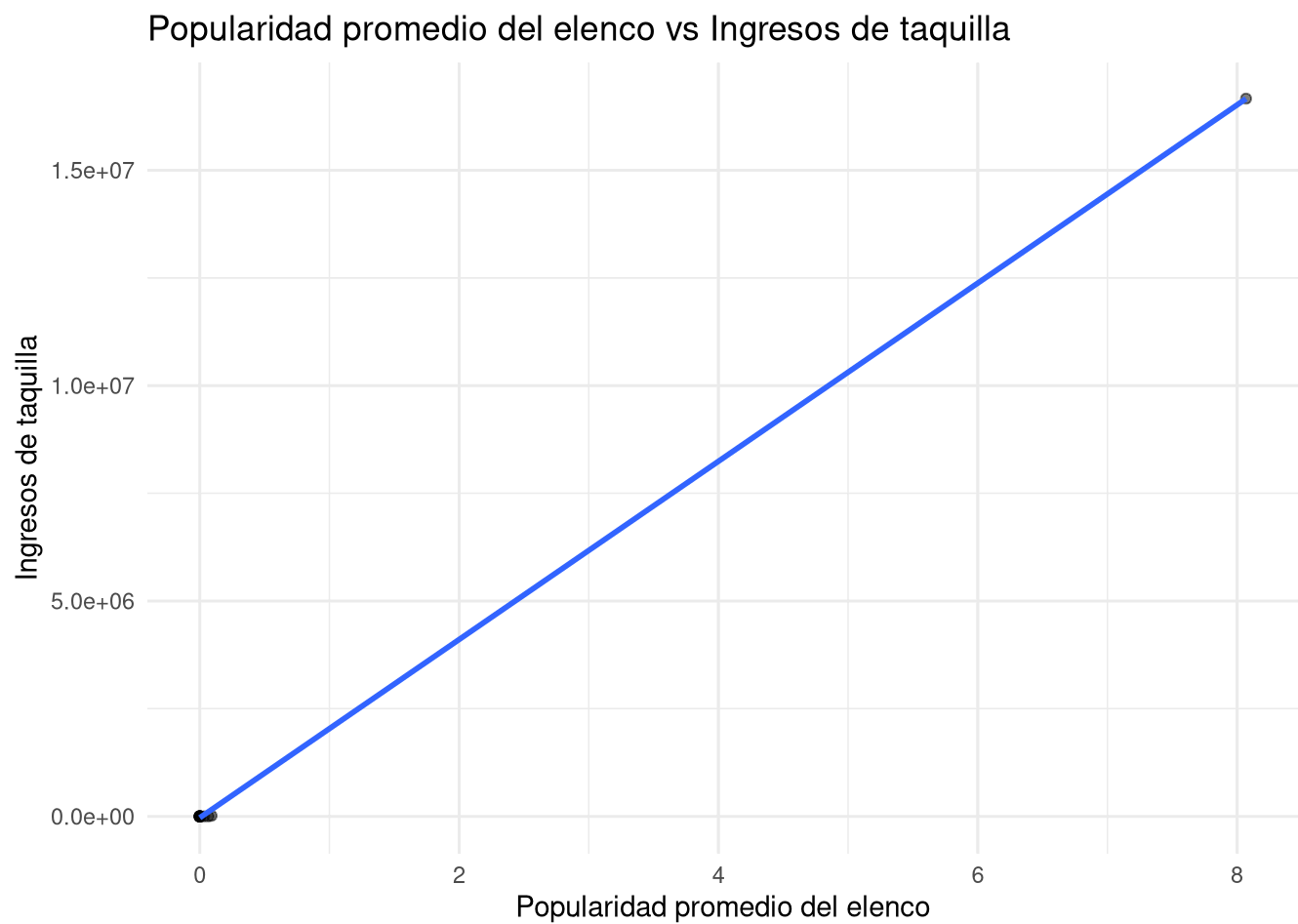
Revenue según presencia de página oficial (log)



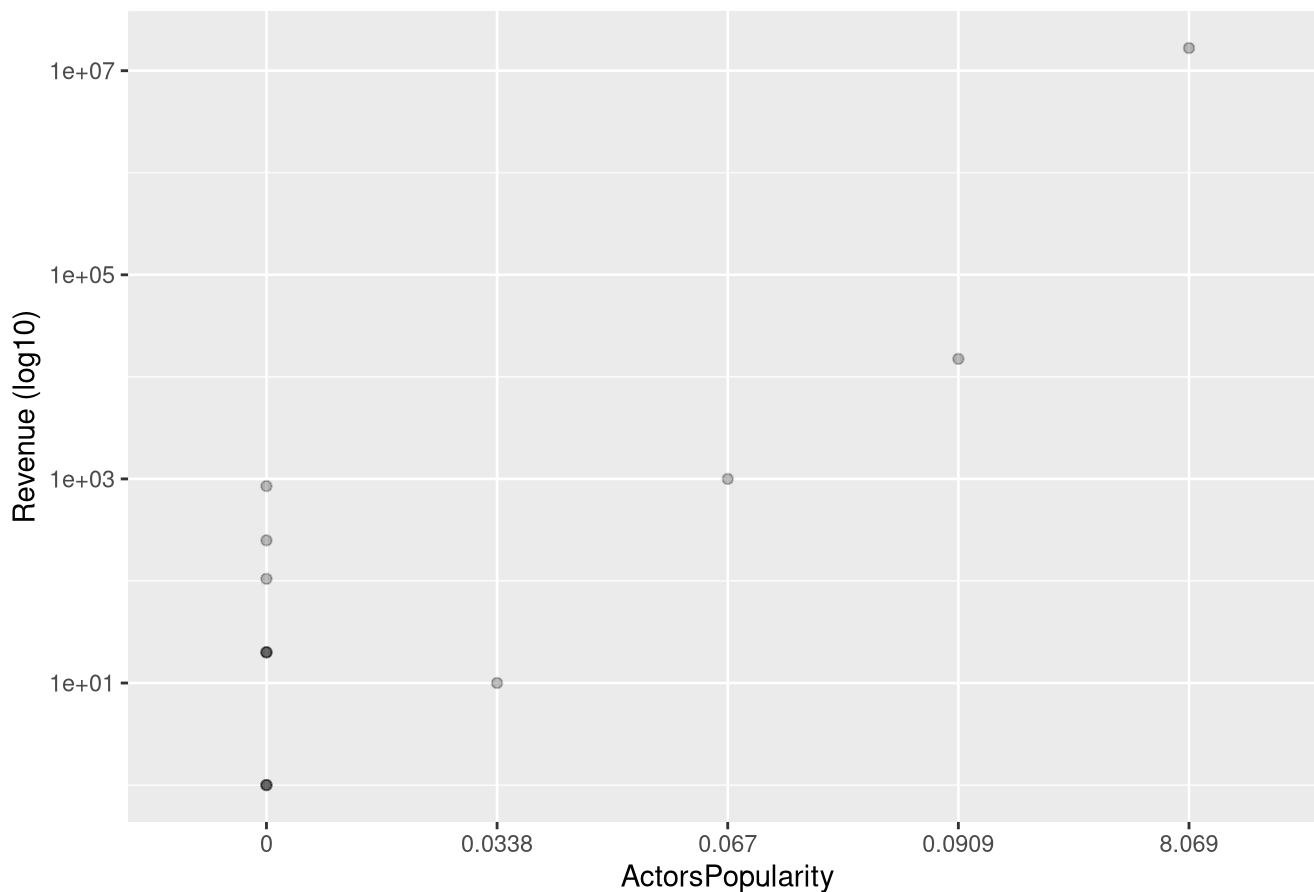
4.16 ¿La popularidad del elenco está directamente

correlacionada con el éxito de taquilla?

[1] 0.6243066



Popularidad del elenco vs Revenue (log)

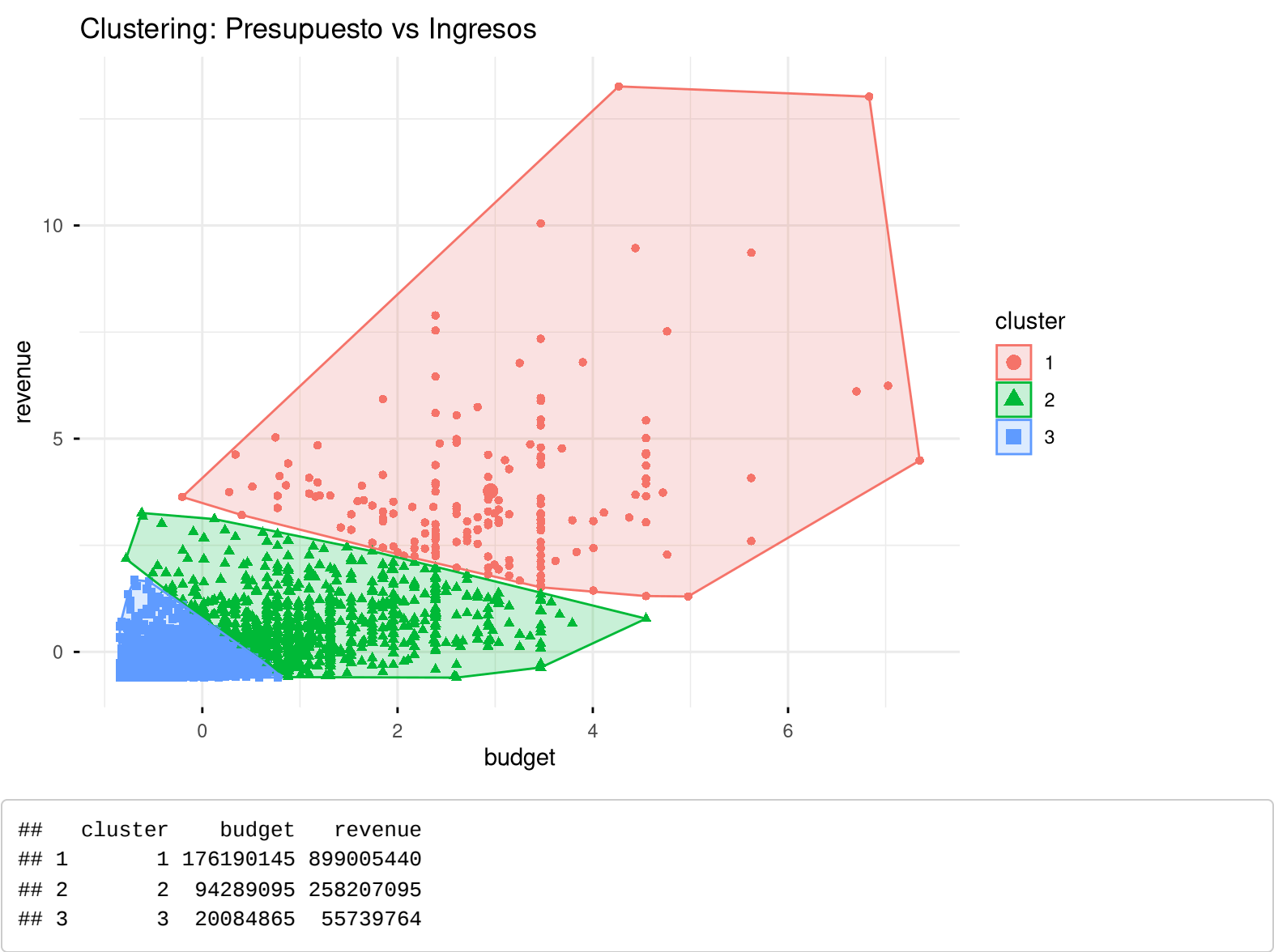


5. Preguntas extra (10 puntos)

Se proponen seis preguntas adicionales (distintas a las anteriores) y se responden con análisis y gráficos.

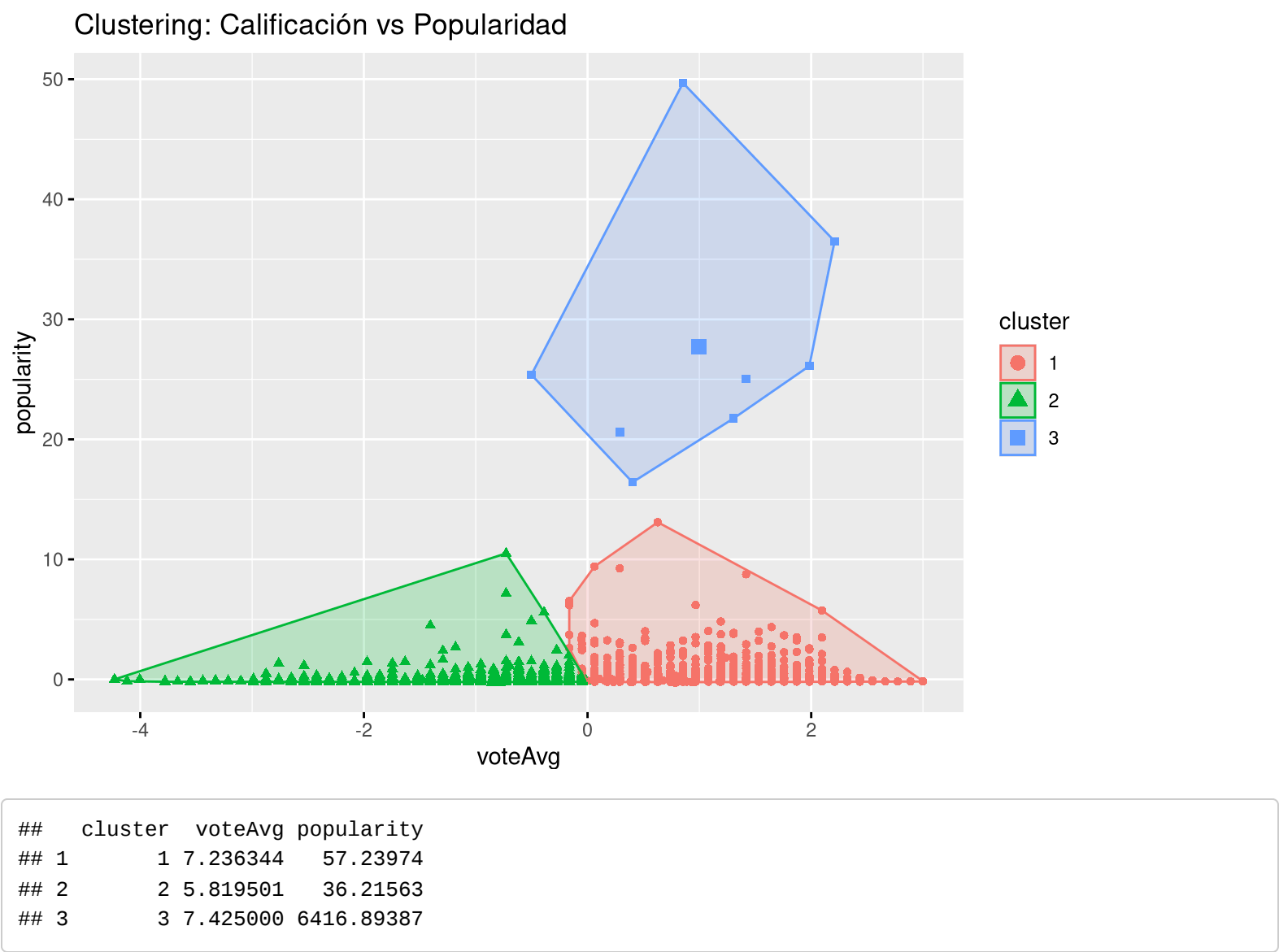
Extra 1: ¿Es posible agrupar las películas en categorías

financieras basándose en presupuesto e ingresos?



Extra 2: ¿Existen grupos de películas “de culto” (alta

calificación, baja popularidad) vs “virales”?



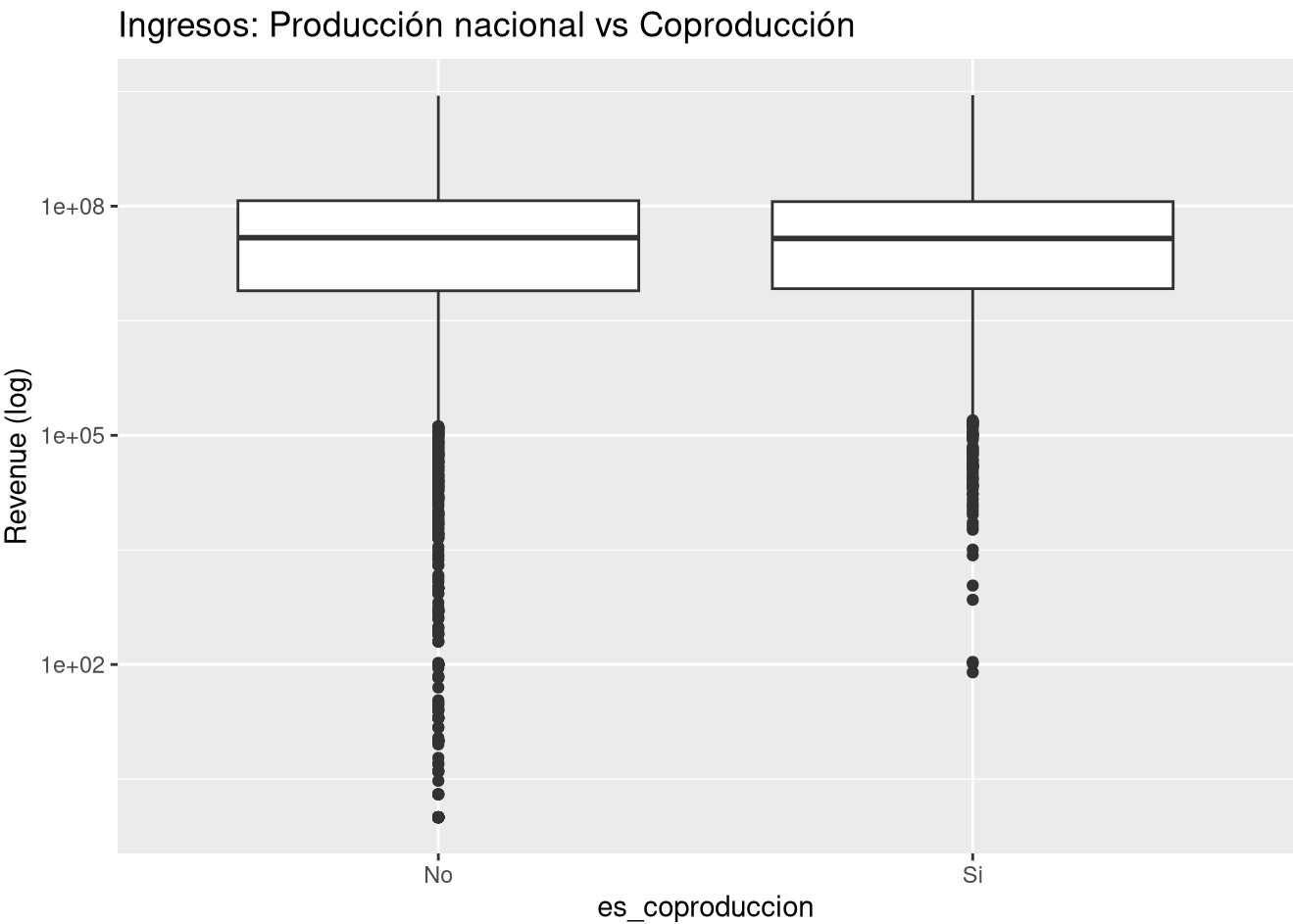
Extra 3: ¿Cuáles son las películas más rentables en términos relativos (ROI)?

##	title	budget	revenue	ROI
## 12334	Alice in Wonderland	3000000	572000000	189.66667
## 10100	Snow White and the Seven Dwarfs	1488423	184925486	123.24256
## 10305	Gone with the Wind	4000000	402352579	99.58814
## 13592	The Rocky Horror Picture Show	1200000	112892319	93.07693
## 12091	Cinderella	2900000	263591415	89.89359
## 9992	Saw	1200000	103911669	85.59306
## 10202	E.T. the Extra-Terrestrial	10500000	792965500	74.52052
## 11178	My Big Fat Greek Wedding	5000000	368744044	72.74881
## 9893	Star Wars	11000000	775398007	69.49073
## 12019	Saturday Night Fever	3500000	237113184	66.74662

Extra 4: ¿Las coproducciones (varios países) generan más

ingresos promedio?

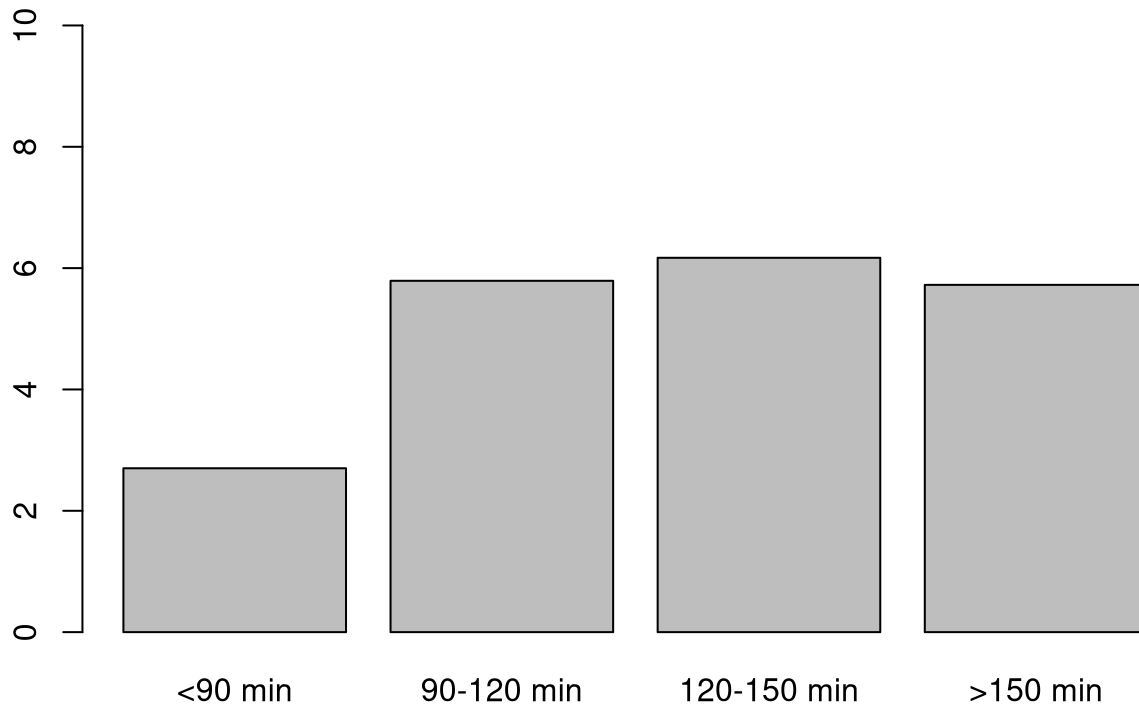
```
## es_coproduccion revenue
## 1 No 106977342
## 2 Si 103856831
```



Extra 5: ¿Existe una duración (runtime) “ideal” que maximice la calificación promedio?

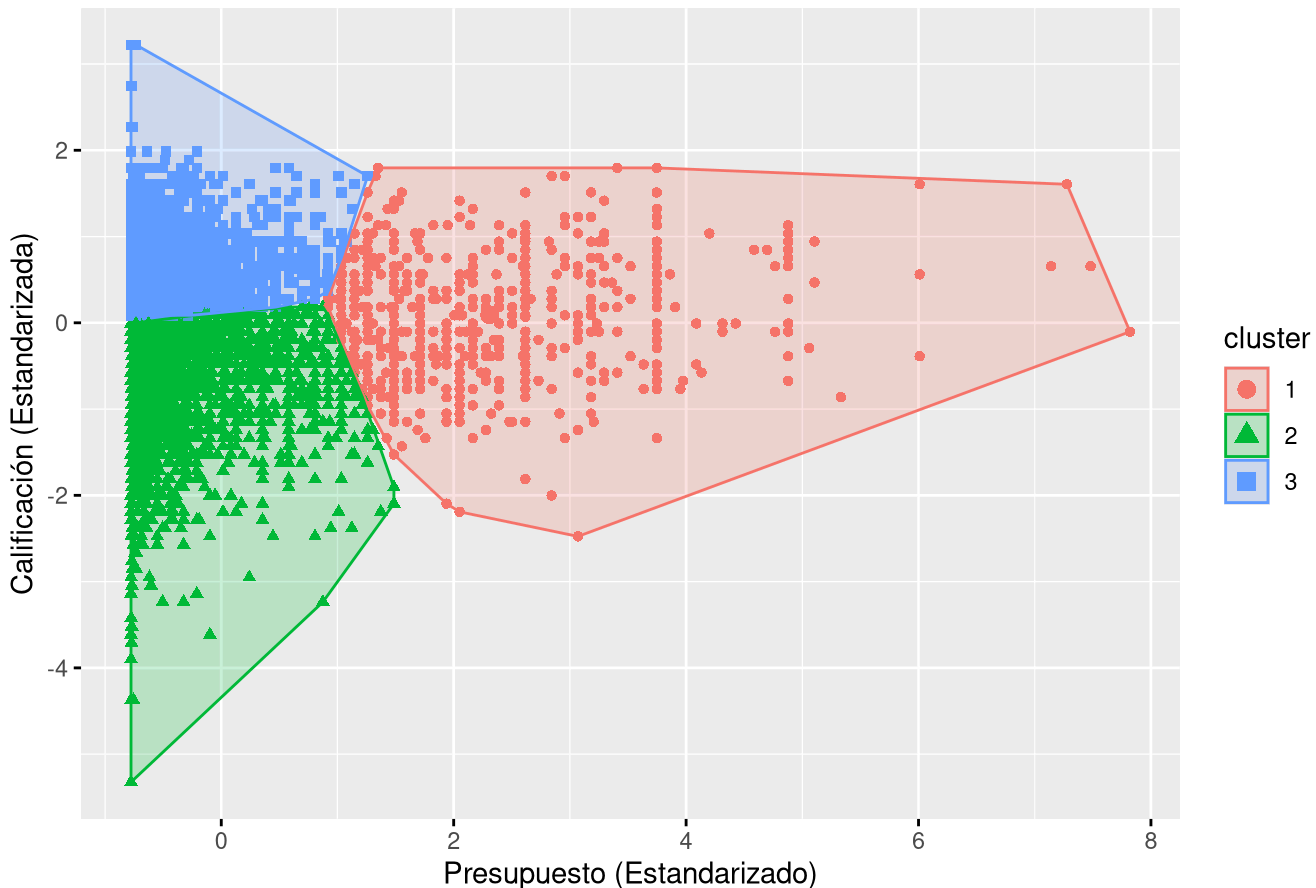
```
## rango_duracion voteAvg
## 1 <90 min 2.701265
## 2 90-120 min 5.790743
## 3 120-150 min 6.170026
## 4 >150 min 5.724146
```

Calificación promedio según duración



Extra 6: ¿Se pueden identificar grupos según presupuesto y calificación? (¿El dinero compra buenas críticas?)

Clustering: Presupuesto vs Calificación



##	cluster	budget	voteAvg
## 1	1	136429551	6.705811
## 2	2	25081252	5.873428
## 3	3	16939485	7.510606