

NETFLIX



Conforman el grupo:



José



Nicole



Sofía



Magdalena

CONTEXTO

Netflix fue fundada en 1997 ofreciendo un servicio de alquiler de DVD a través del correo postal. En el año 2000, aún con el alquiler de DVD como único servicio, Netflix introdujo un sistema de recomendación personalizado, basado en la calificación de una a cinco estrellas que sus clientes hacían de los materiales que alquilaban a través de su sitio por Internet. A partir de 2008 se agregan diferentes opciones para acceder al catálogo por streaming. Hoy en dia es una de las plataformas de transmisión de video y medios más populares.



OBJETIVOS DEL ANÁLISIS

El objetivo primario de nuestro trabajo será analizar cómo cambió el catálogo del cine de Netflix a través de los años

INTERROGANTES DEL ANÁLISIS

PAÍSES

- 1 buscaremos encontrar qué países tienen mayor predominancia en Netflix.

PELÍCULAS VS SERIES

- 2 ¿En qué momento se incorporaron las películas a la plataforma? ¿Cómo se fue alterando el tipo de lanzamiento (películas/series)?

GÉNEROS

- 3 ¿Qué grupo de películas (documentales, comedias, etc.) tuvo más influencia en la plataforma a través de los años?

VARIABLES

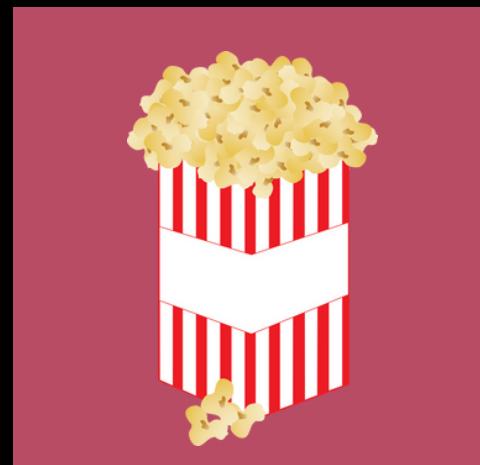


VARIABLES:



show id

Tipo de dato: Carácter
Descripción: Identificación de la película/ serie



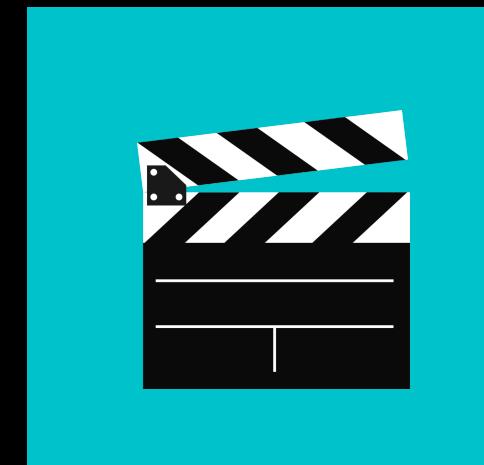
tipo

Tipo de dato: Carácter
Descripción: define si el registro es una película o una serie



título

Tipo de dato: Carácter
Descripción: Título del video

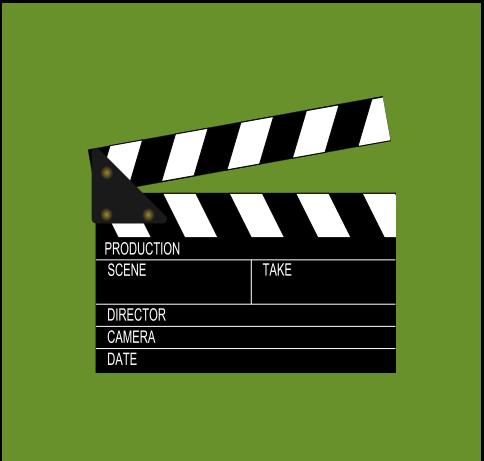


director

Tipo de dato: Carácter
Descripción: Quién o quienes dirigen la película



VARIABLES:



release_year

Tipo de dato: Double
Descripción: Año de lanzamiento



cast

Tipo de dato: caracter
Descripción: Actores y Actrices que participan



country

Tipo de dato: caracter
Descripción: País en el que se produjo



date_added

Tipo de dato: caracter
Descripción: Fecha en la que se agrega a Netflix



VARIABLES:



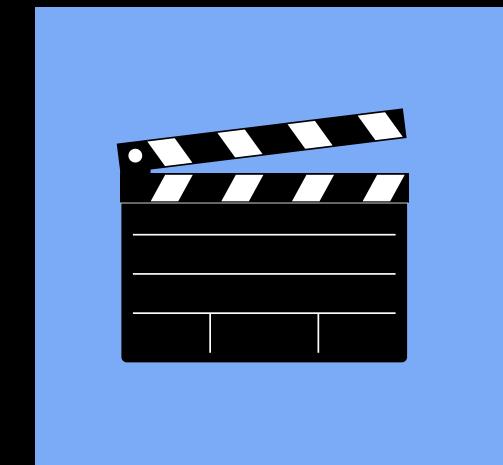
rating

Tipo de dato: carácter
Descripción: Indica quienes pueden verlo (ej +18)



duration

Tipo de dato: carácter
Descripción: Si es una película cuantos minutos, si es una serie cuantas temporadas



listed_in

Tipo de dato: carácter
Descripción: A qué géneros pertenece



description

Tipo de dato: carácter
Descripción: Breve desarrollo sobre la trama.

PROBLEMAS DE LA BASE

MISSINGS



¿QUÉ VARIABLES CONTIENEN
VALORES FALTANTES?

- director (2634 missings)
- cast (825 missings)
- country (831)
- date added (10)
- rating (4)
- duration (3)

MISSINGS

¿ALEATORIOS O NO ALEATORIOS?

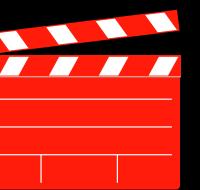
La mayoria de los missings de las variables analizadas resultaron depender de la variable type y de la variable release_year

OUTLIERS

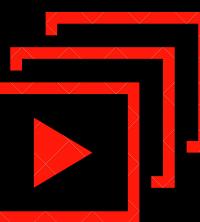
VALOR ATÍPICO: AQUEL QUE SE ENCUENTRA POR FUERA DE LOS “BIGOTES” EN UN GRÁFICO DE TIPO BOXPLOT



Para la variable release_year hay 8088 valores típicos y 719 atípicos (el 8,89%)



Para la variable película hay 5678 valores típicos y 450 atípicos (el 7,93%)



Para la variables serie hay 2417 valores típicos y 259 atípicos (el 10,72%)

¿QUÉ HICIMOS CON LOS OUTLIERS?

Imputamos los años que son inconsistentes con la realidad (son anteriores a 1985 o posteriores a 2022)

Método utilizado: imputación por la mediana.

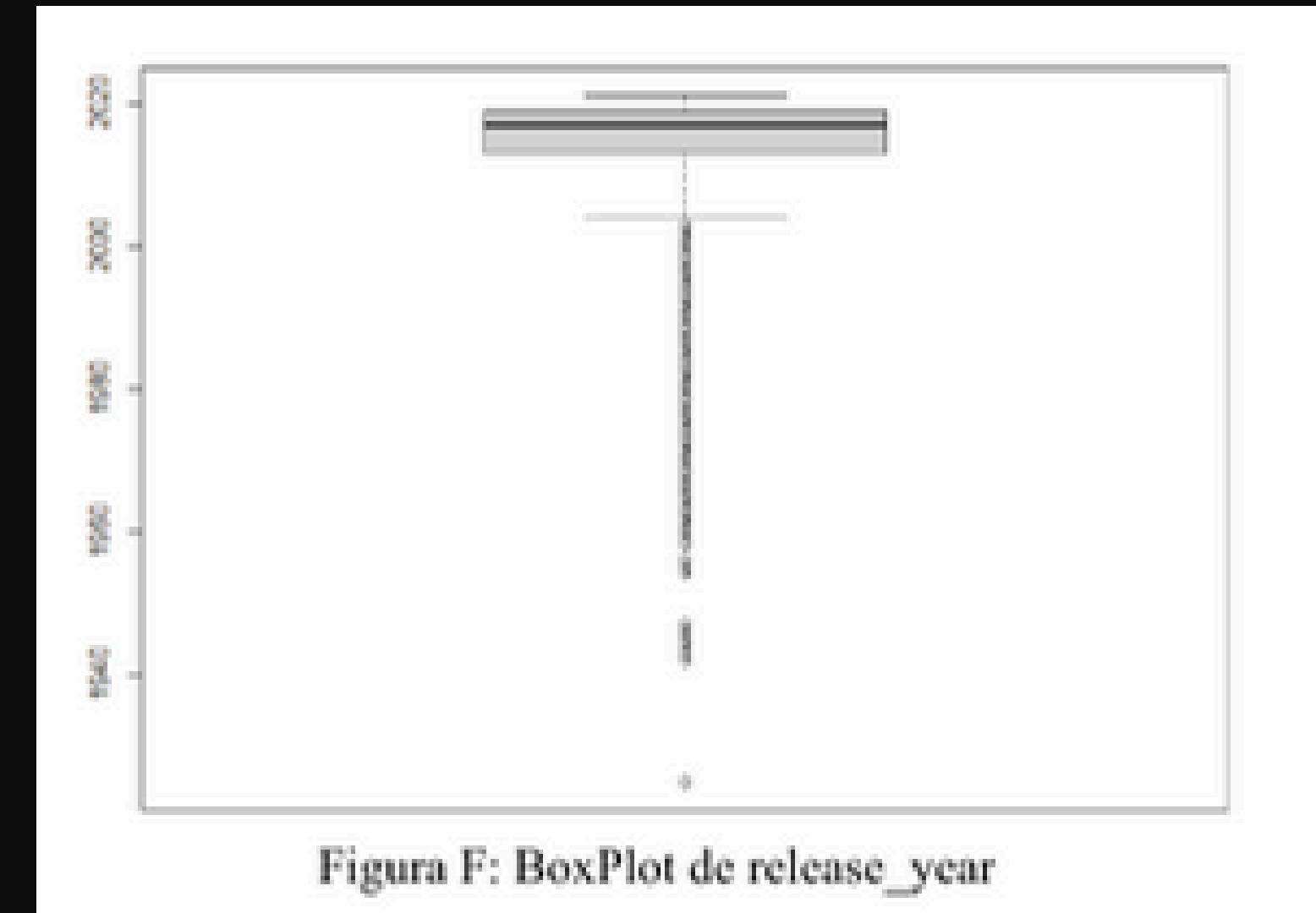


Figura F: BoxPlot de `release_year`

¿QUÉ HICIMOS CON LOS OUTLIERS?

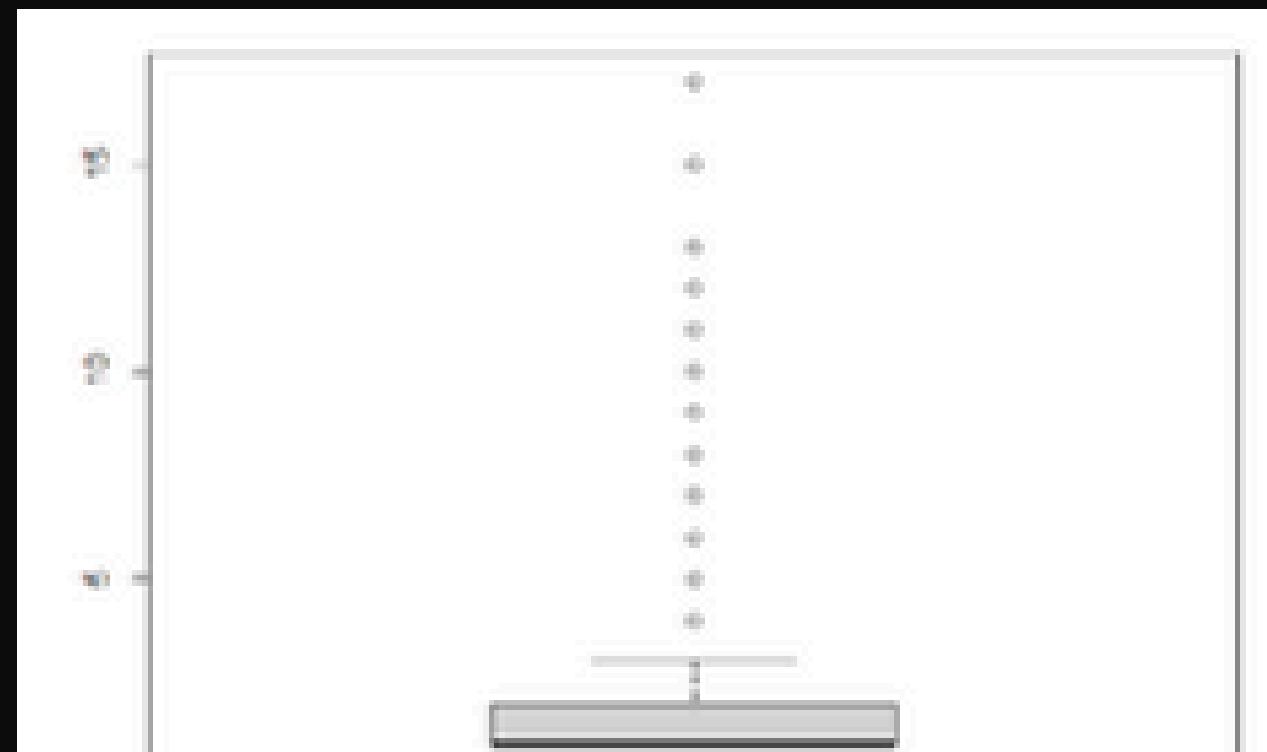


Figura D: BoxPlot de duración de series

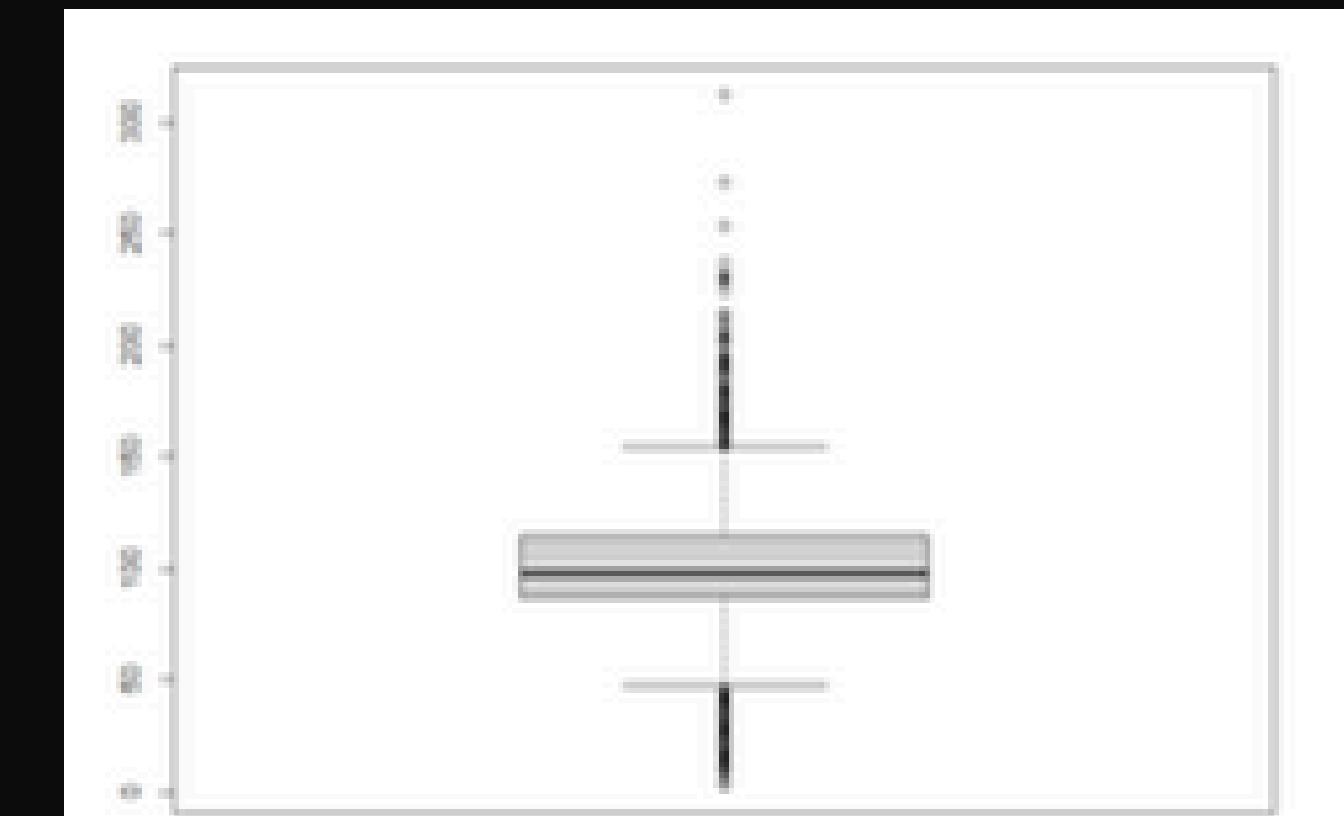
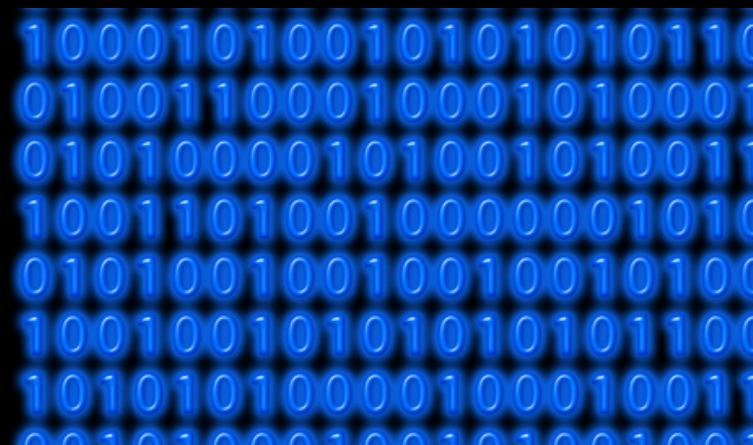


Figura E: BoxPlot de duración de películas

RESPUESTAS

PROCEDIMIENTO

Season 1 ▾



A binary matrix consisting of 10 rows and 15 columns, representing categorical data for each item in the dataset.

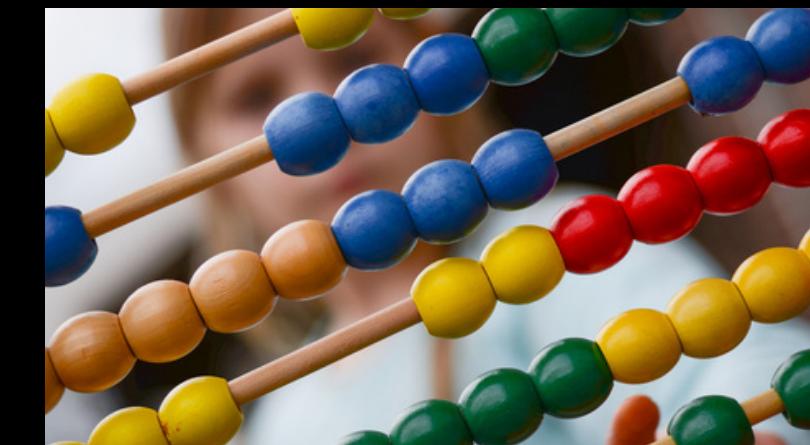
VECTORES BINARIOS

Marcamos con 1 y 0 si cada registro pertenece o no a cada categoría y país del dataset



AGRUPAMOS POR AÑO

nuestro principal interés es analizar el cambio en las películas que se agregan a través del tiempo



CONTAMOS

Cuantas películas por año hay de cada categoría



CALCULAMOS LA PROPORCIÓN

Y analizamos su variación en los años que comprende la base

¿QUE PAÍSES TIENEN MAYOR PREDOMINANCIA EN NETFLIX?

10 PAÍSES CON MÁS PRESENCIA EN NETFLIX

Country	#
United States	390
India	164
United Kingdom	100
Canada	443
France	383
Japan	313
Spain	273
South Korea	231
Germany	224
Australia	199

ESTADOS UNIDOS VS INDIA

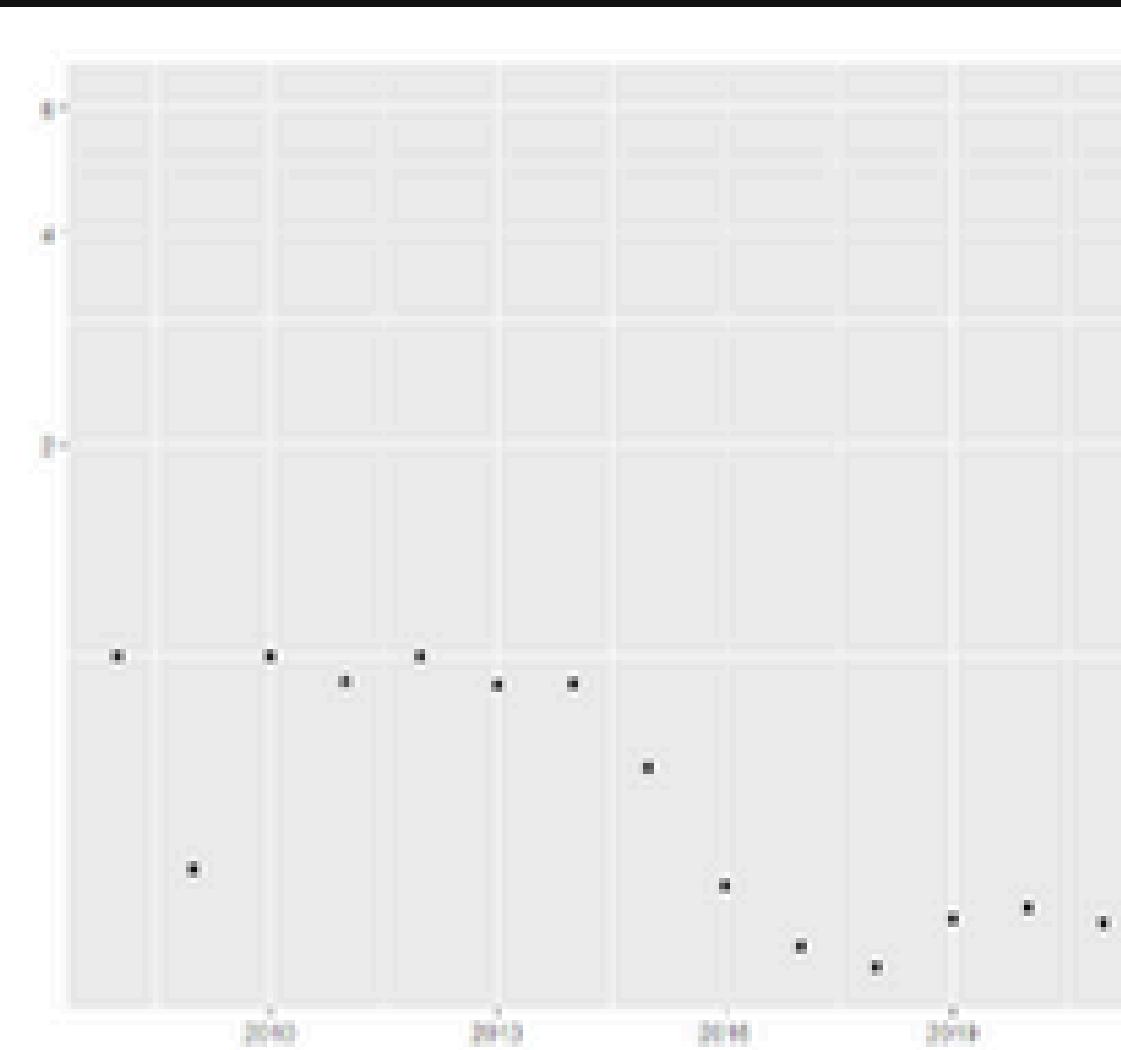


Figura 1E: evolución de Estados Unidos en la plataforma

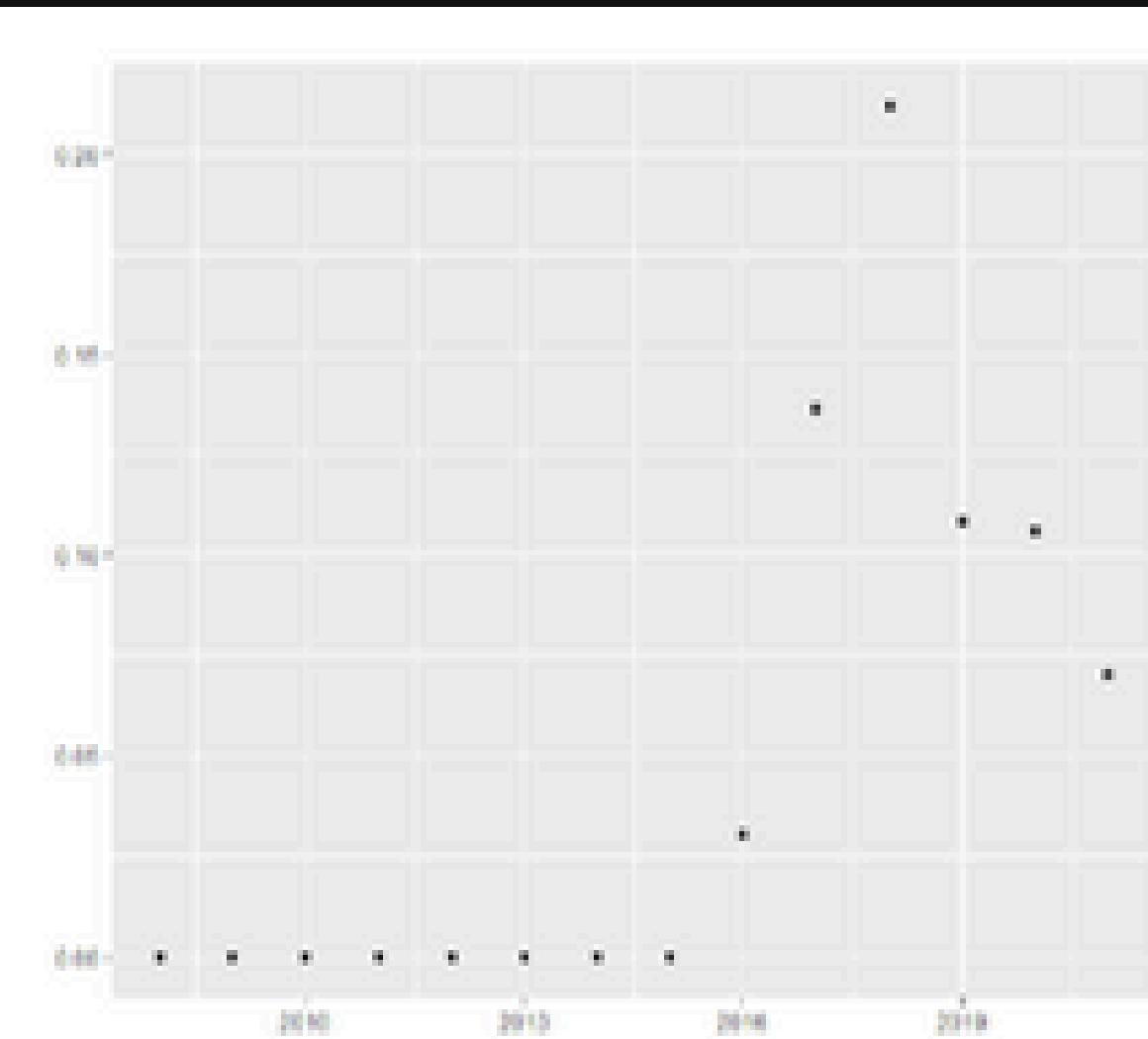


Figura 2E: evolución de India en la plataforma



ÍNDICES DE CORRELACIONES

CORRELACIÓN PROPORCIÓN LANZAMIENTOS DE
EEUU CON AÑO DE LANZAMIENTO

```
sample estimates:  
cor  
-0.7542993
```

CORRELACIÓN PROPORCIÓN LANZAMIENTOS DE
INDIA CON AÑO DE LANZAMIENTO

```
sample estimates:  
cor  
0.7131164
```

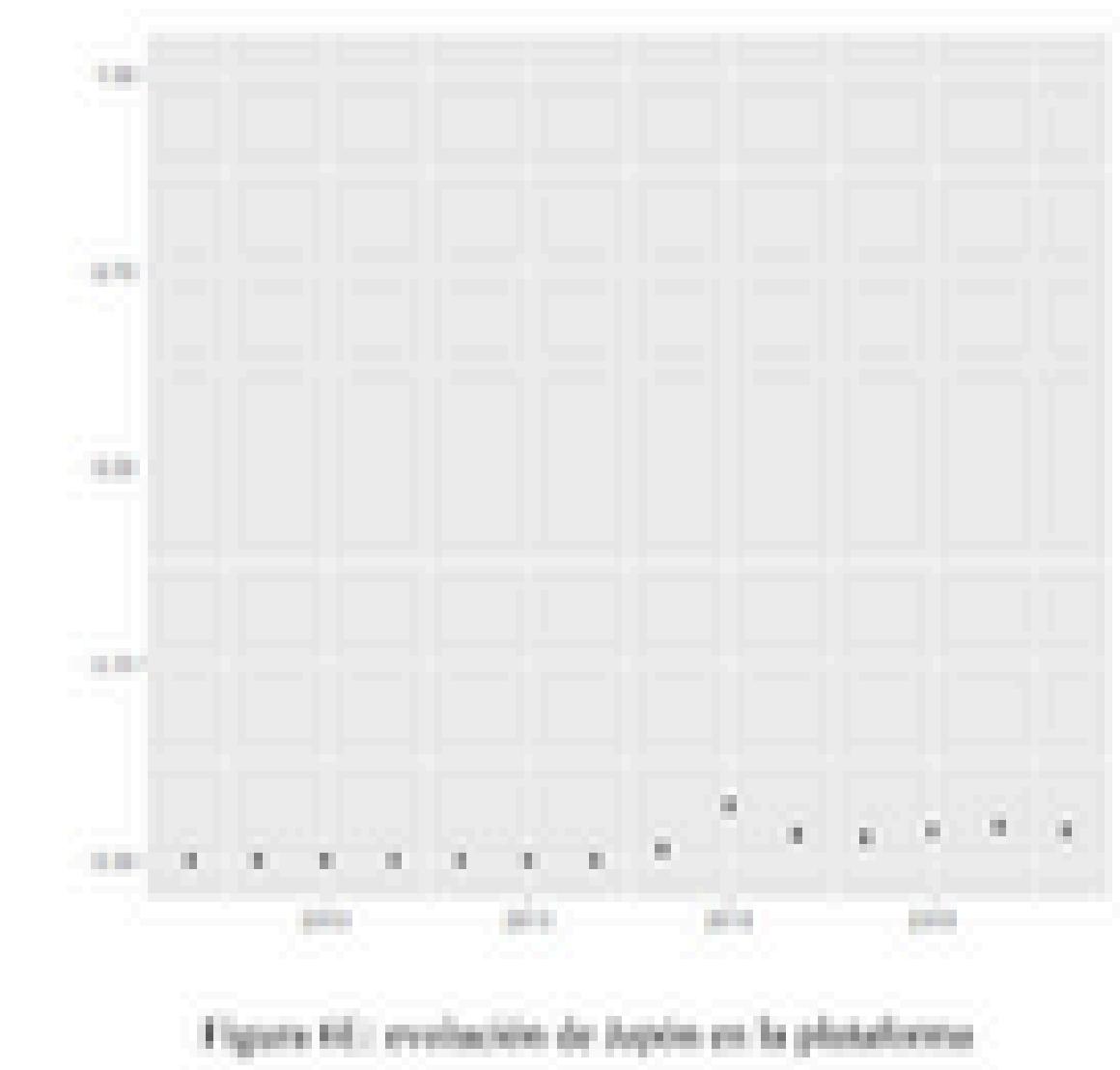
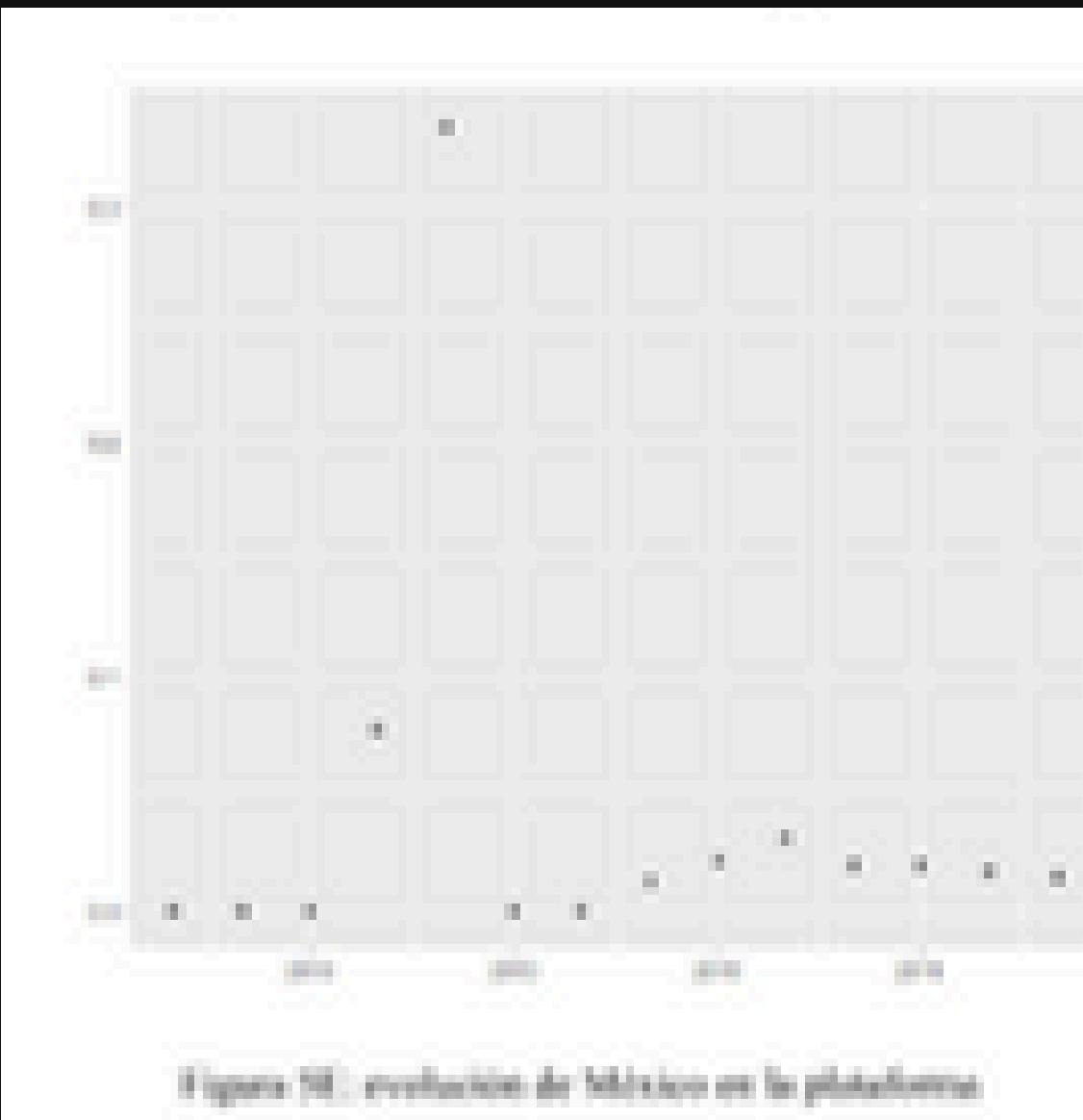
ÍNDICE DE CORRELACIÓN

PROPORCIÓN DE PELÍCULAS Y AÑO DE LANZAMIENTO

Se realizó un test de correlación entre el año de lanzamiento y la cantidad de películas lanzadas.
Estados Unidos --> corr = -0.754
India --> corr = 0.713

Después de analizar los 10 países con más presencia, pudimos ver que en su mayoría tendían a tener un índice de correlación positivo y alto.

MEXICO Y JAPÓN



Estos países fueron los únicos, aparte de Estados Unidos, que no aumentaron su presencia en netflix de manera abrupta a partir del 2016.

¿QUÉ GÉNEROS DE PELÍCULAS TUVIERON MÁS INFLUENCIA EN LA PLATAFORMA A TRAVÉS DE LOS AÑOS?

10 GÉNEROS CON MÁS PRESENCIA EN NETFLIX

Listed_in	n
International Movies	2752
Dramas	2427
Comedies	1674
International TV Shows	1351
Documentaries	869
Action & Adventure	859
TV Dramas	763
Independent Movies	756
Children & Family Movies	641
Romantic Movies	616



ANÁLISIS GRÁFICO

▶ Reproducir

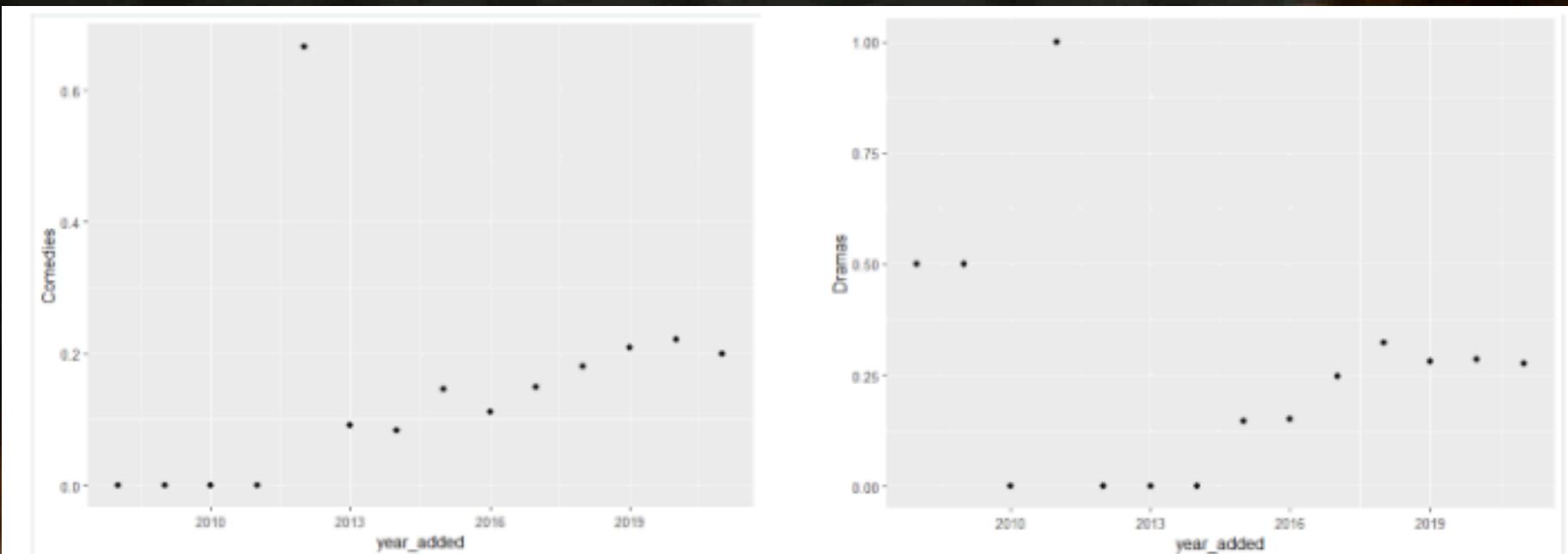


Figura 1F: Comedies and Dramas

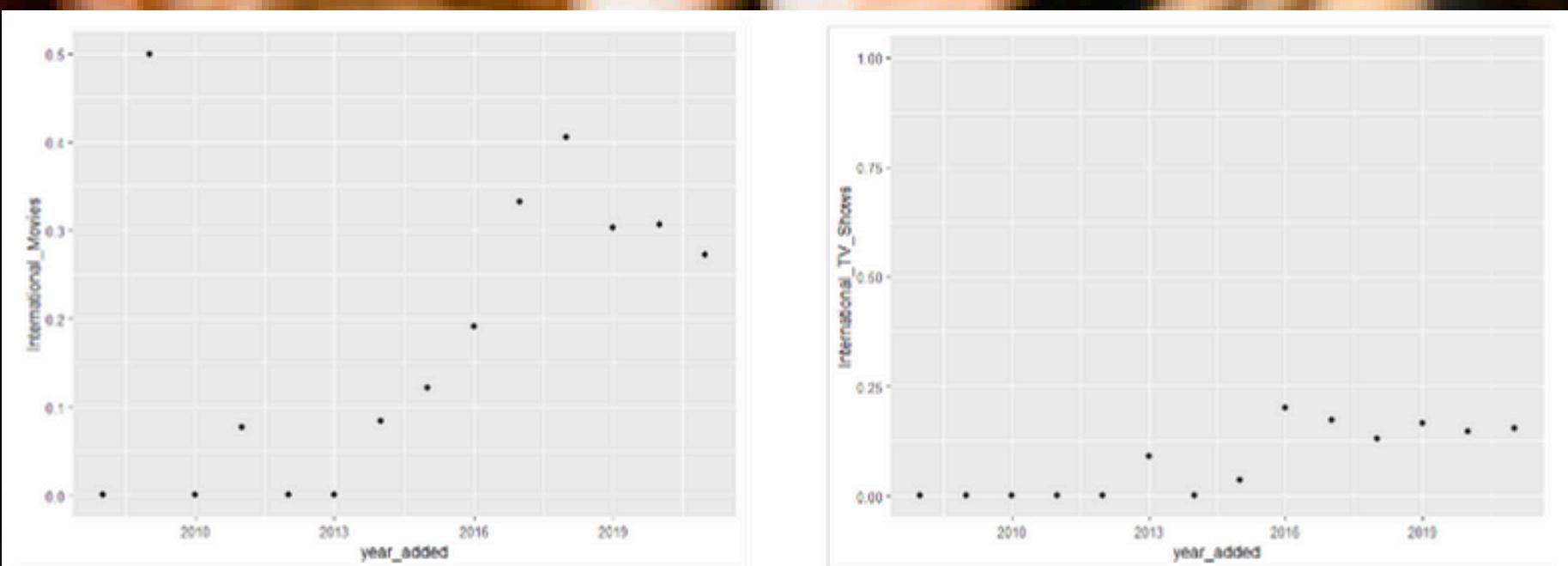


Figura 2F: International Movies y Internacional TV Shows



ANÁLISIS GRÁFICO

▶ Reproducir

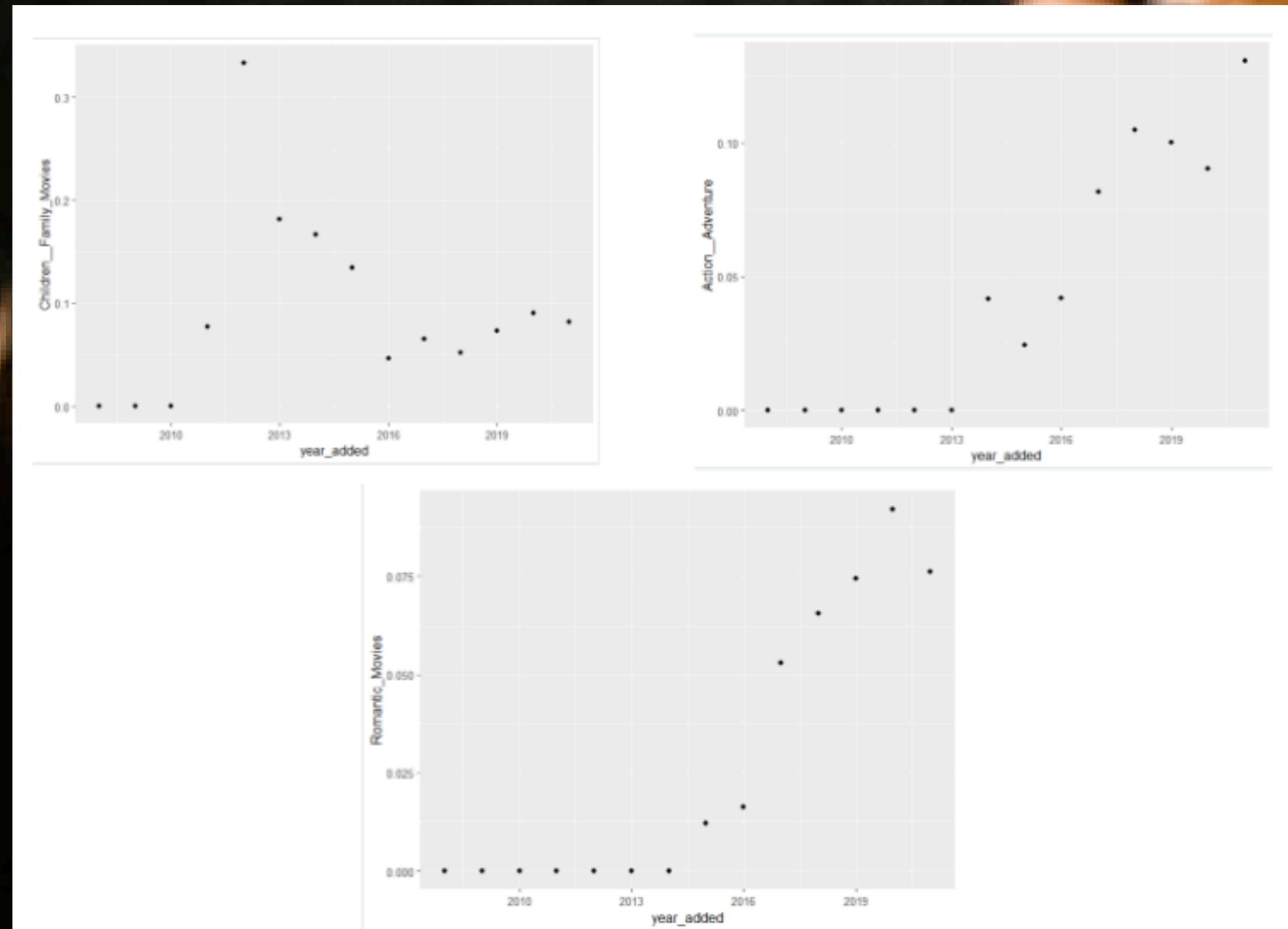


Figura 3F: Children and Family Movies, Romantic Movies and Action Movies



CONCLUSIONES GÉNEROS

Dramas a través del tiempo

H0) la proporción de Dramas en los últimos años (2015-2021) no es mayor a la de Dramas en los primeros años (2008-2014).

```
Welch Two Sample t-test
```

```
data: proporciones_genero$Dramas[8:14] and proporciones_genero$Dramas[1:7]
t = -0.27827, df = 6.3626, p-value = 0.6052
```

Con un valor p mayor a 5% no se rechaza la hipótesis nula. No hubo un aumento significativo de los dramas en los últimos años.

Comedias a través del tiempo

H0) la proporción de comedias en los últimos años (2015-2021) no es mayor a la de comedias en los primeros años (2008-2014).

```
data: proporciones_genero$Comedies[8:14] and proporciones_genero$Comedies[1:7]
t = 0.57345, df = 6.3132, p-value = 0.2931
```

Con un valor p mayor a 5% no se rechaza la hipótesis nula. No hubo un aumento significativo de las comedias en los últimos años.

Evolución de las películas internacionales a través del tiempo

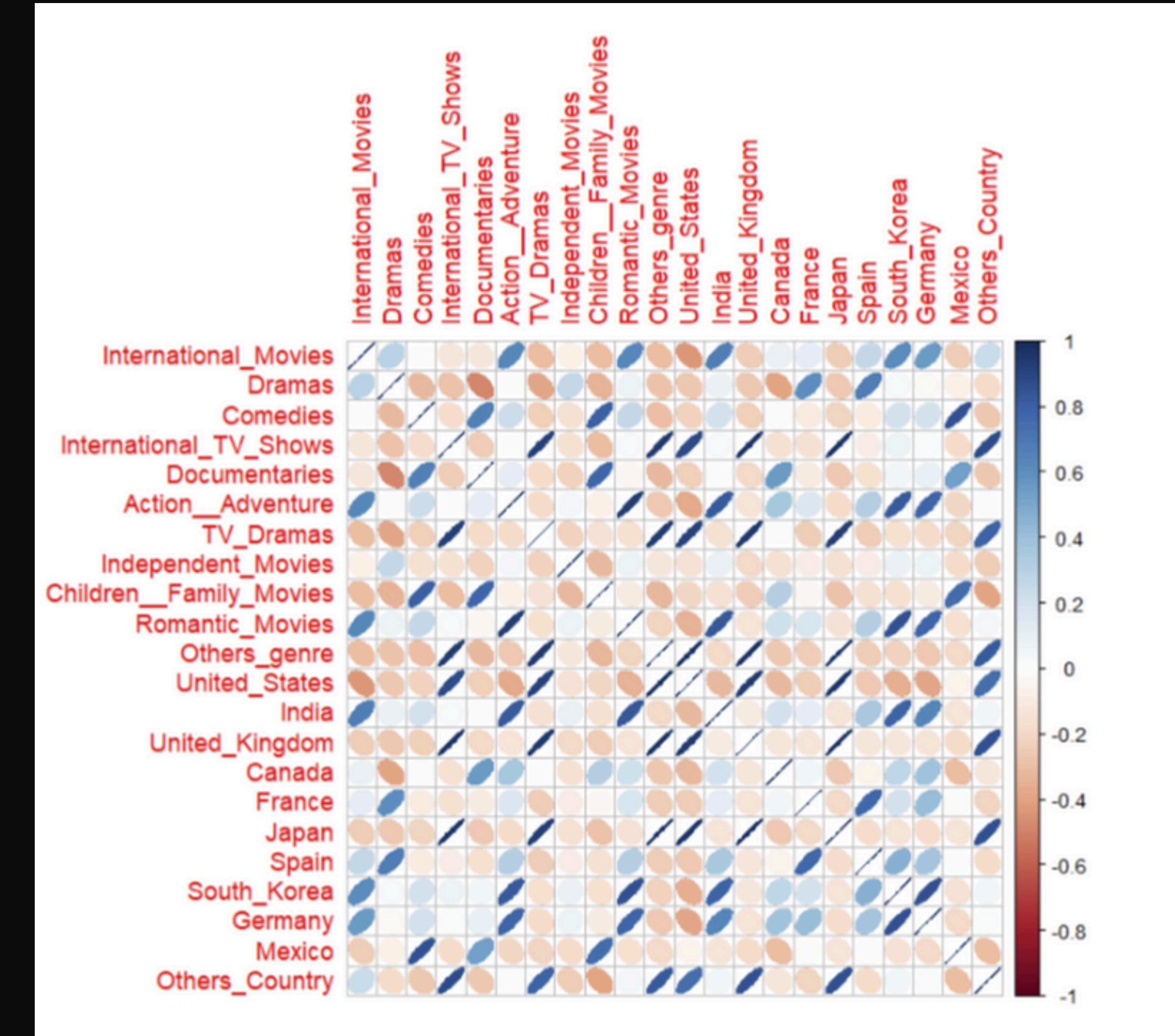
H0) la proporción de películas internacionales en los últimos años (2015-2021) no es mayor a la de películas internacionales en los primeros años (2008-2014).

```
data: proporciones_genero$International_Movies[8:14] and proporciones_genero$International_Movies[1:7]
t = 2.3402, df = 8.935, p-value = 0.0221
```

Con un valor p menor a 5% se rechaza la hipótesis nula. Hubo un aumento significativo de las películas internacionales en los últimos años.

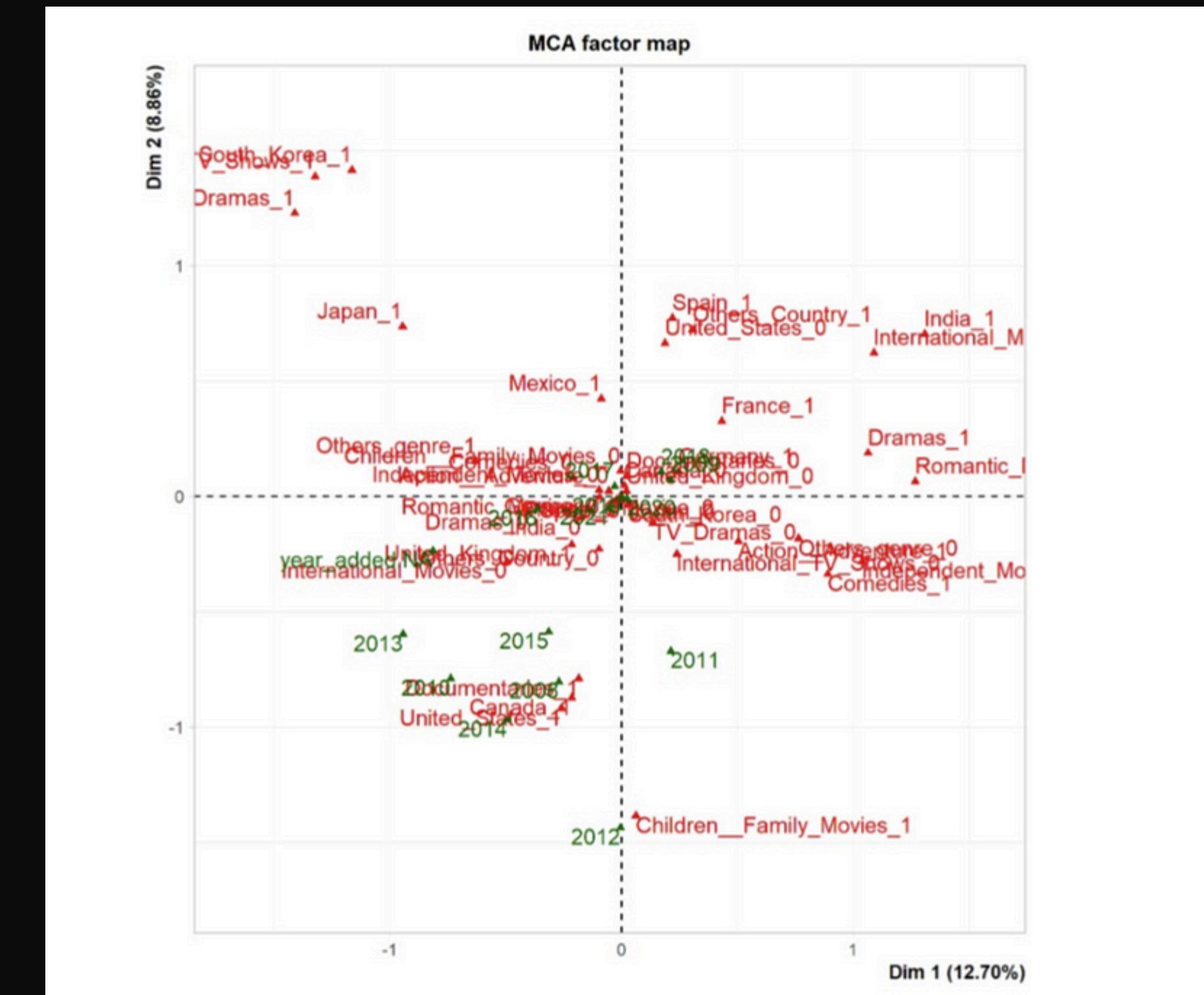
COMPLEMENTO PREGUNTAS 1 Y 2

CORRELACIÓN ENTRE LAS VARIABLES

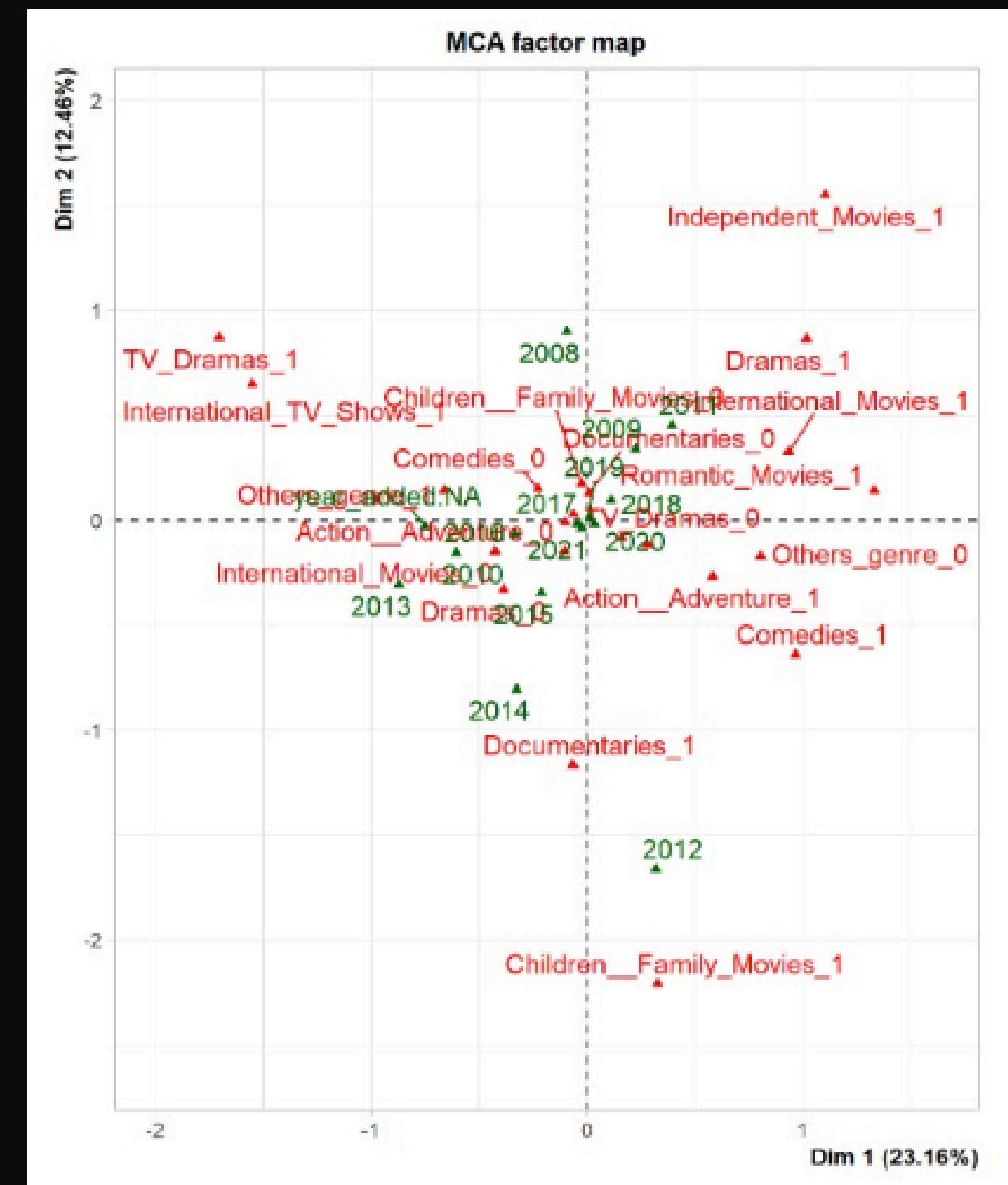


RELACIÓN A TRAVÉS DE LOS AÑOS

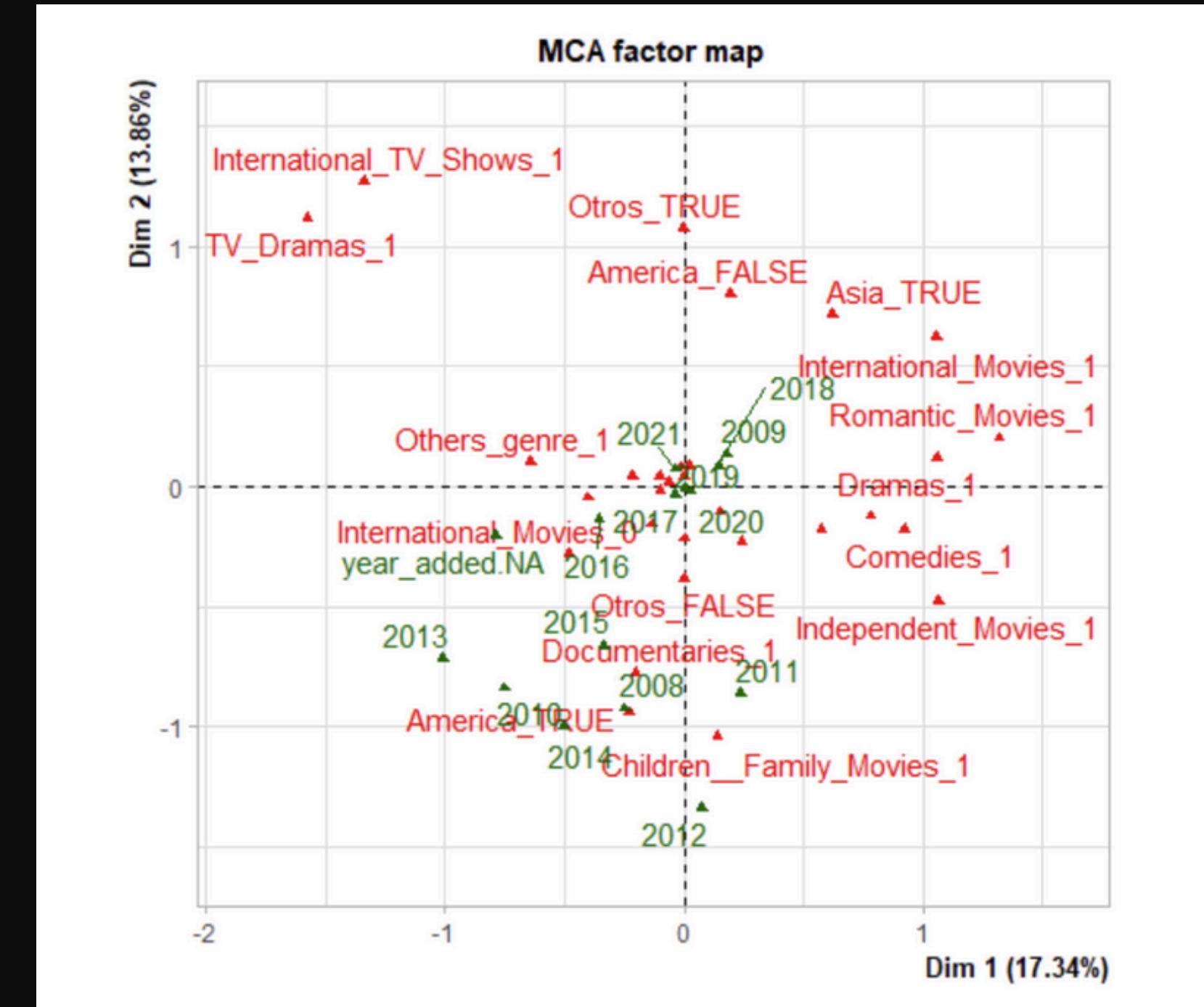
Realizamos un análisis MCA en el cuál detallamos con distintos criterios, para comprender qué preponderancia tiene cada género o país en los años que comprende la base.



GÉNEROS A TRAVÉS DE LOS AÑOS



GÉNEROS Y CONTINENTES A TRAVÉS DE LOS AÑOS

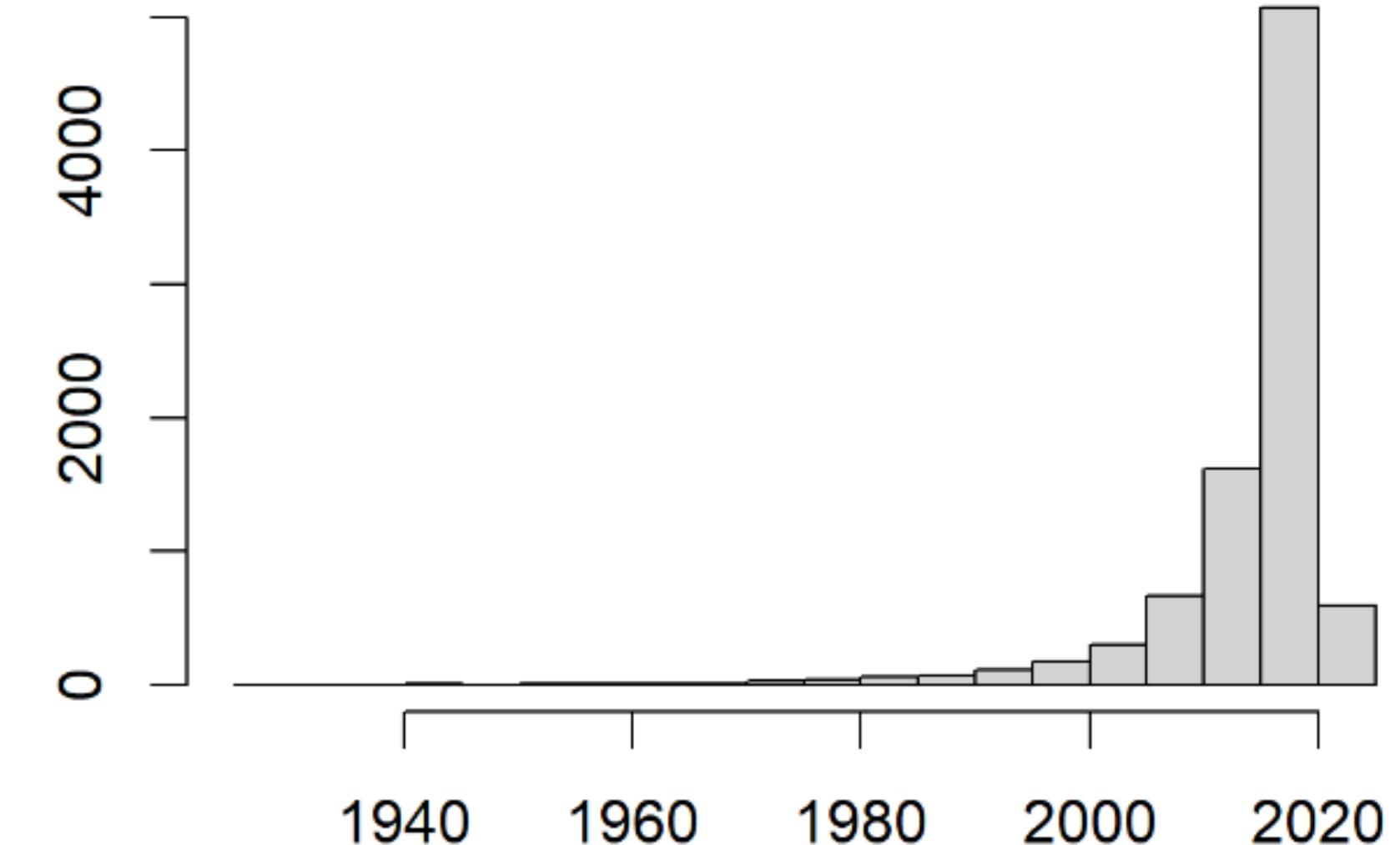


¿EN QUÉ MOMENTO SE INCORPORARON LAS PELÍCULAS A LA PLATAFORMA?

INCORPORACIÓN

CONTEXTO DE LA BASE

- Naturalmente en la plataforma se incluyeron películas más recientes por avance tecnológico. Es por esto que en la imagen de al lado se ve mucha más frecuencia de películas modernas
- Netflix comenzó a funcionar 100% en streaming en el año 2010; a partir de ese momento dejó de depender de cualquier soporte físico para distribuir contenidos. En 2007
- En 2013, Netflix apostó por un contenido para seducir al cliente extranjero, lanzando así su primera serie de producción propia "House of Cards". A partir de ese momento, Netflix logró doblar la cifra de sus suscriptores en apenas tres años: pasó de 60 millones en 2014 a los 103 millones en 2017.



INCORPORACIÓN

POR FECHA DE LANZAMIENTO

La base de datos cuenta con una variable llamada “date_added”, en la que se indica el día, mes y año en que se realizó el lanzamiento en la plataforma.

Creamos una nueva columna llamada year_added, que contenga solo el año de lanzamiento.

Luego, agrupamos los años y contamos la cantidad de lanzamientos que hubo en cada año.

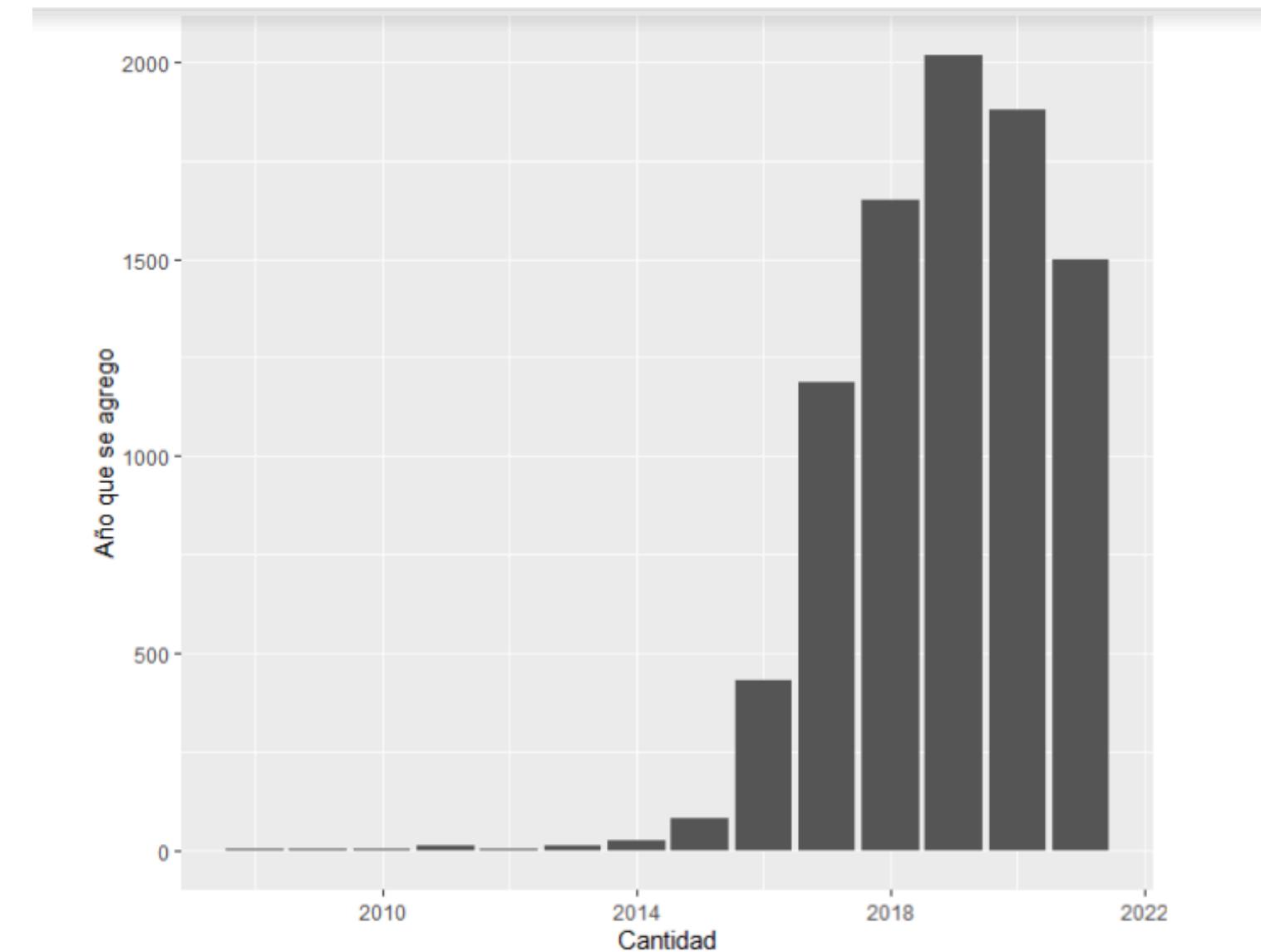


Figura 1H: lanzamientos por año

INCORPORACIÓN

La explicación por la baja en la tendencia se puede comprender desde la perspectiva que ya que en los últimos años aparecieron otras plataformas de video como Amazon Prime Video, HBO, Disney+ y más que comenzaron a hacerle competencia.. Estas nuevas plataformas empezaron, no solo a sacarle parte de la clientela, sino también parte del contenido que anteriormente era exclusivo de Netflix. Esto llevó a que los lanzamientos por año disminuyeran.



¿CÓMO SE FUE ALTERANDO EL TIPO DE LANZAMIENTO (PELÍCULAS/SERIES)?

CANTIDAD SERIE/ PELÍCULA

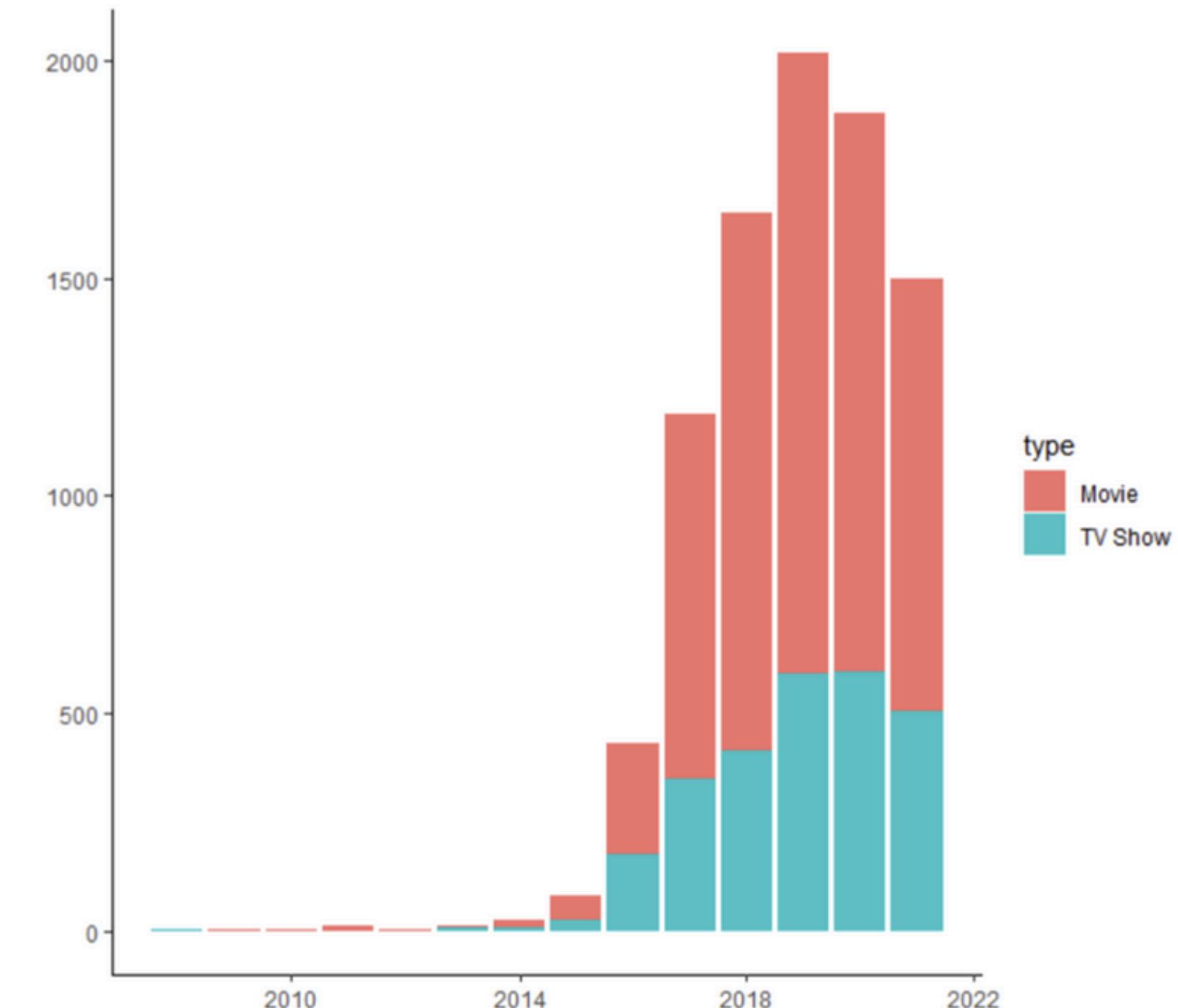
POR FECHA DE LANZAMIENTO

Prueba de Chi Cuadrado:

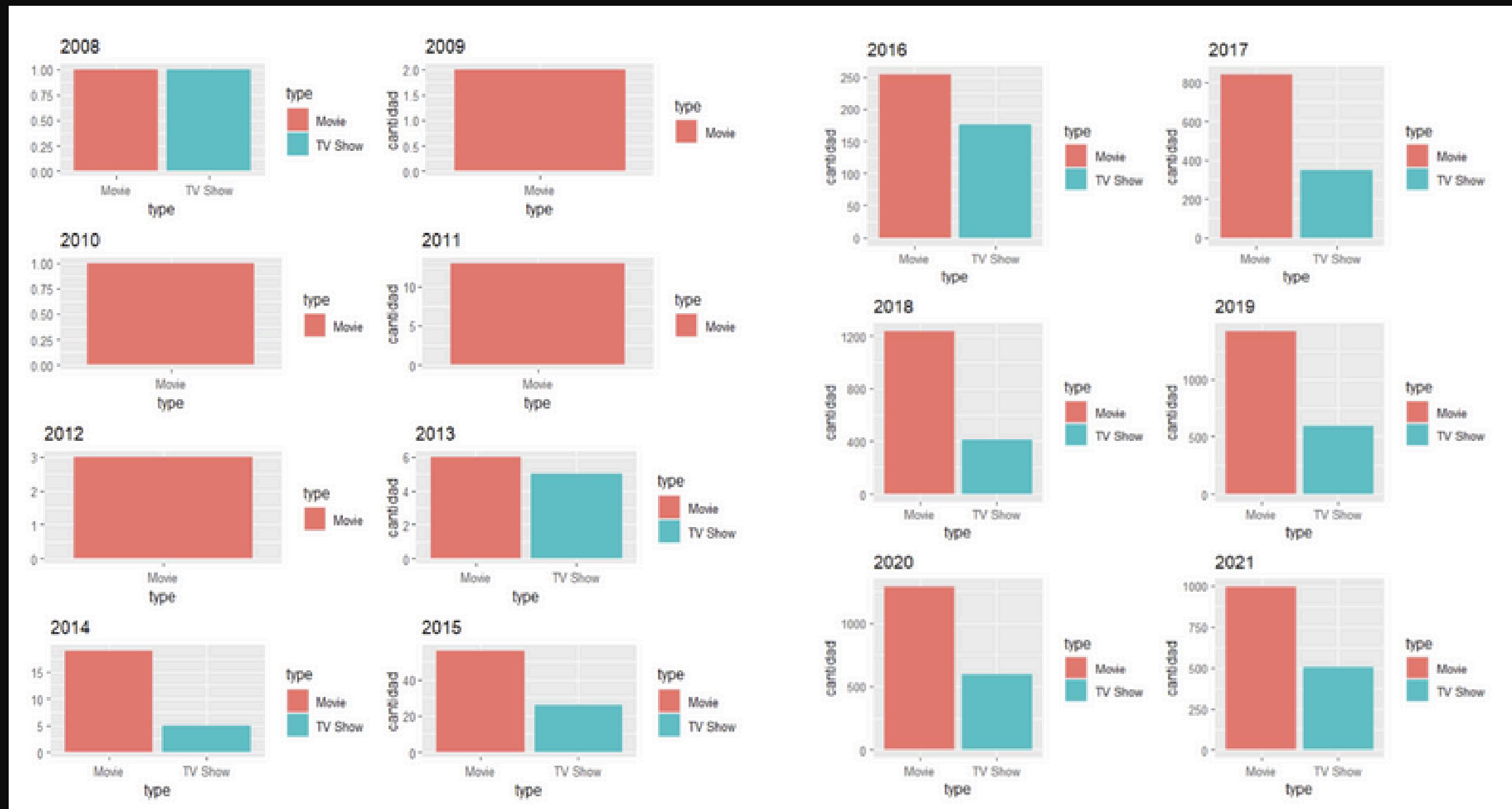
H0) las variables año y tipo de lanzamiento son independientes.

Se obtuvo un valor p menor a 5%

La hipótesis nula se rechaza y se concluye con un nivel de confianza del 95% que existe dependencia entre años y tipo de lanzamiento.



CANTIDAD SERIE/ PELÍCULA A TRAVEZ DE LOS AÑOS



CANTIDAD SERIE/ PELÍCULA A TRAVEZ DE LOS AÑOS

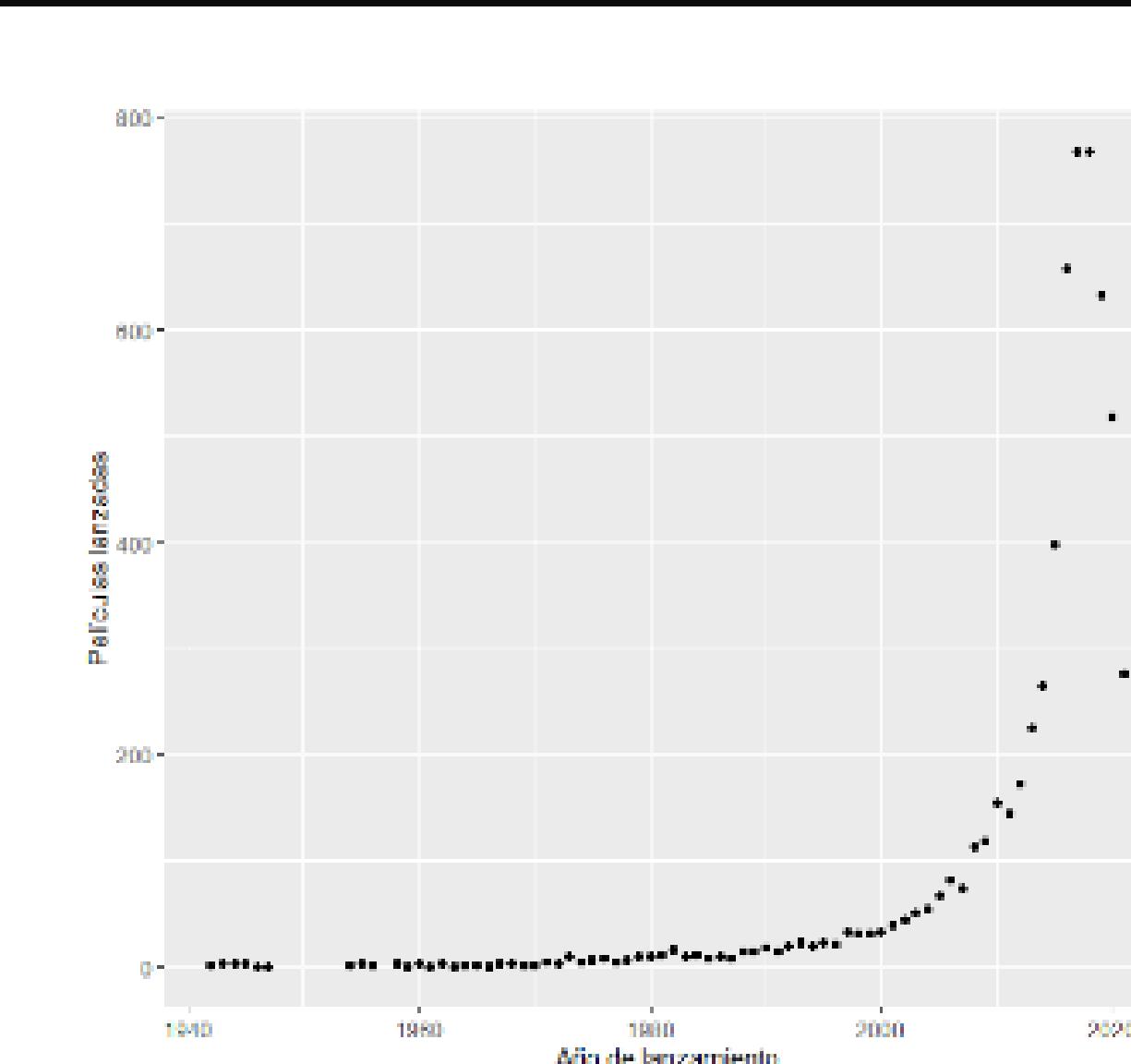


Figura 4I: películas a través del tiempo

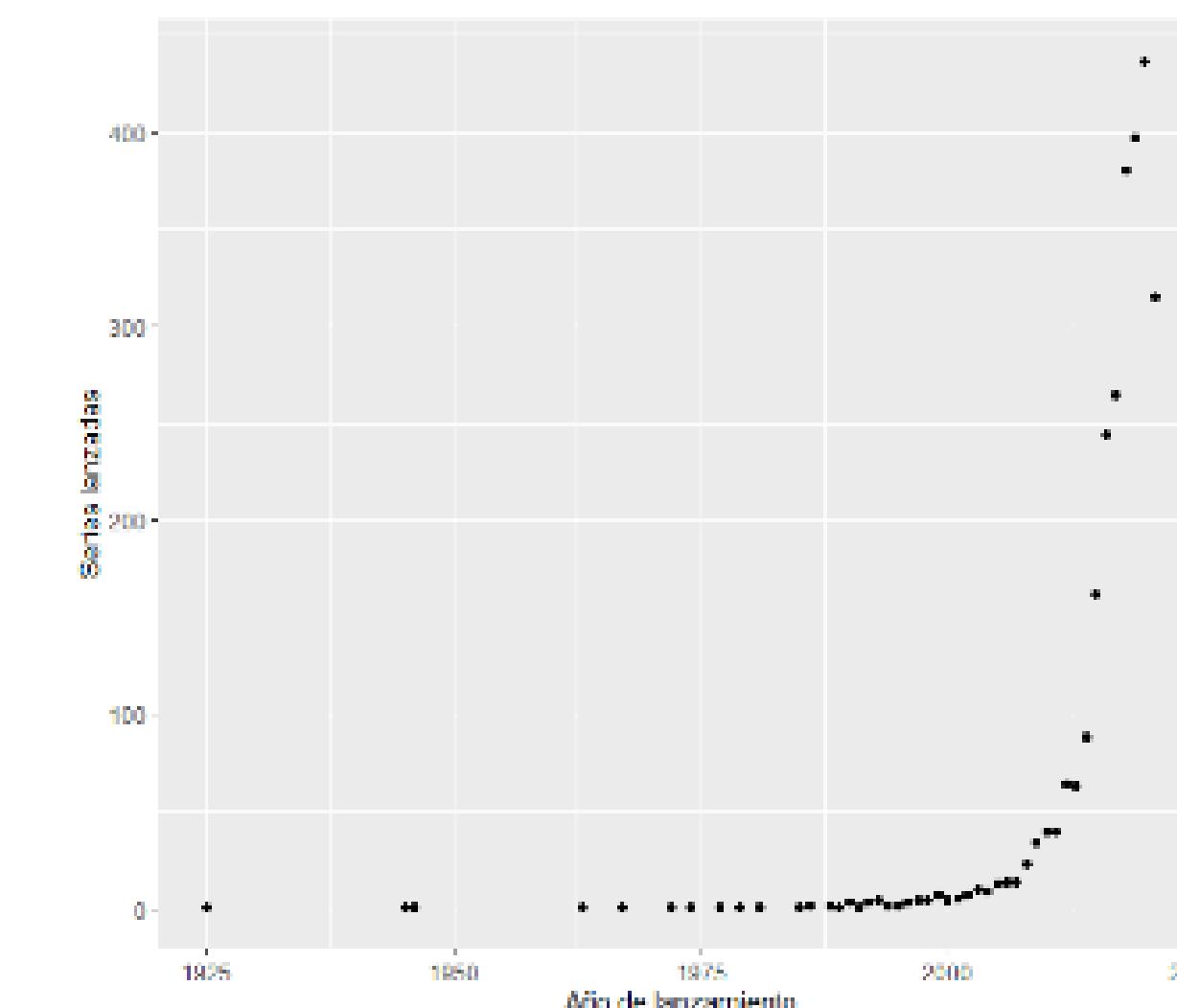
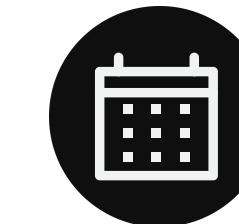


Figura 5I: Series a través del tiempo

CONCLUSIÓN

Pudimos observar el contexto de Netflix y su historia en los datos, entendiendo en por qué de los sucesos que fuimos encontrando en el análisis. Logramos contestar las preguntas planteadas en un principio conectandolas entre si y entendiendo cómo se relacionaban.

ANALÍTICA DESCRIPTIVA



2do cuatrimestre - 2022
Trabajo final



Integrantes

- Eppens, Magdalena
- Gonzalez del Solar, Sofía
- Reiman, Nicole
- Salinas Aguilar, Jose Rubén

MUCHAS GRACIAS