

Obligatorio2: Ejemplo

Eduardo Fernández

Basado en: <https://observablehq.com/@uwdata/a2-example-movies-data>

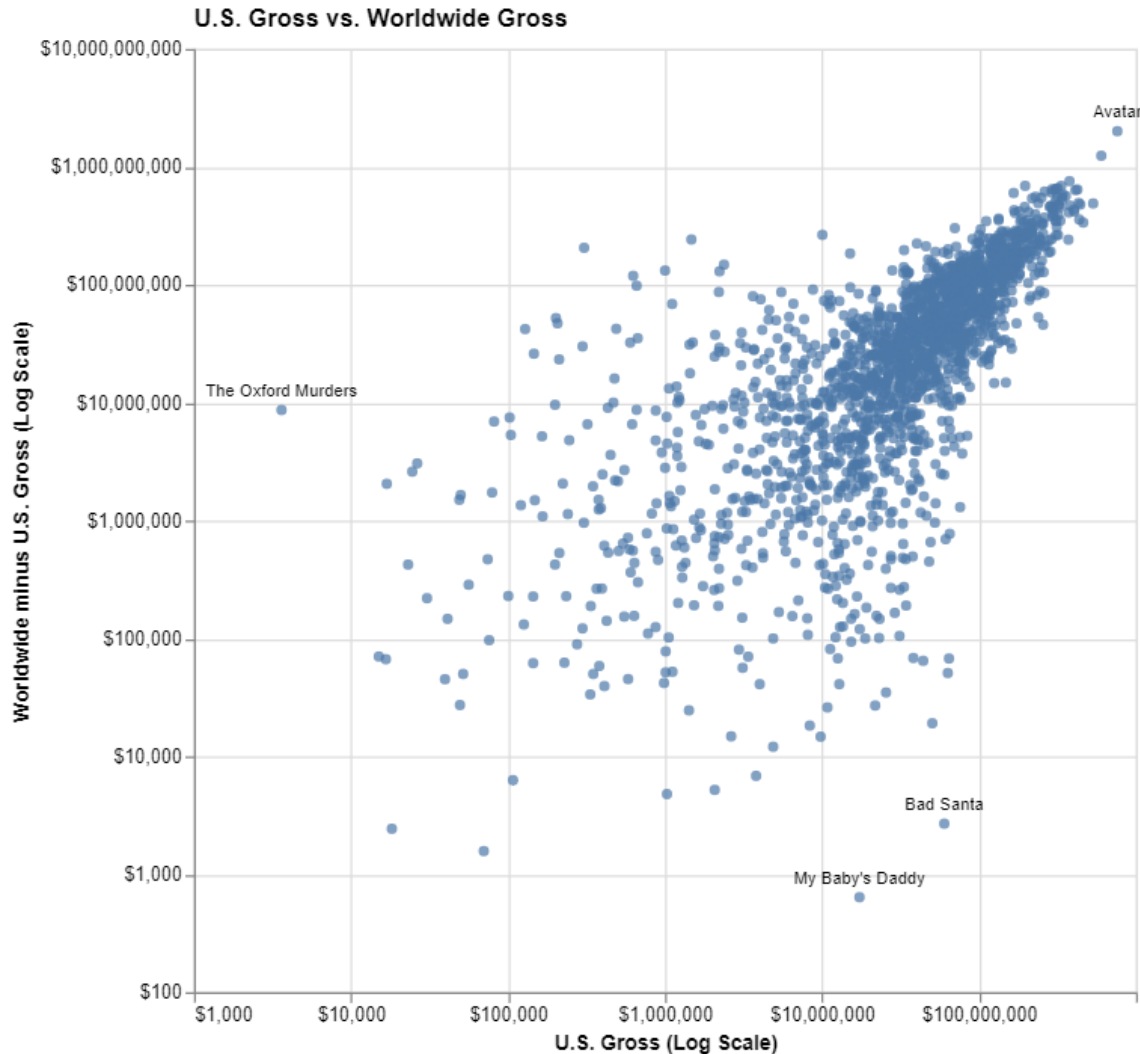
Dataset: Motion Pictures Data

- Colección de 3201 películas (observaciones, filas).
- Datos obtenidos de: <https://cdn.jsdelivr.net/npm/vega-datasets@2.2.0/data/movies.json>
- Se pueden introducir en R haciendo:
 - > `install.packages("jsonlite")`
 - > `library(jsonlite)`
 - > `movies <- fromJSON("C:\\Users\\Usuario\\Downloads\\movies.json")`
 - > `Head(movies)`
- Variables incluidas (columnas):
 - "Title", "US Gross", "Worldwide Gross", "US DVD Sales", "Production Budget", "Release Date", "MPAA Rating", "Running Time min", "Distributor", "Source", "Major Genre", "Creative Type", "Director", "Rotten Tomatoes Rating", "IMDB Rating", "IMDB Votes"

Preguntas iniciales realizadas

- 1.¿Cómo les va comparativamente a las películas en los mercados de EE. UU. y del resto del mundo?
- 2.¿Cómo varían las ganancias según el género?
- 3.¿Cómo se distribuyen las fechas de estreno según género?
- 4.¿Cómo se relacionan los puntajes de los fans y críticos con el éxito financiero?

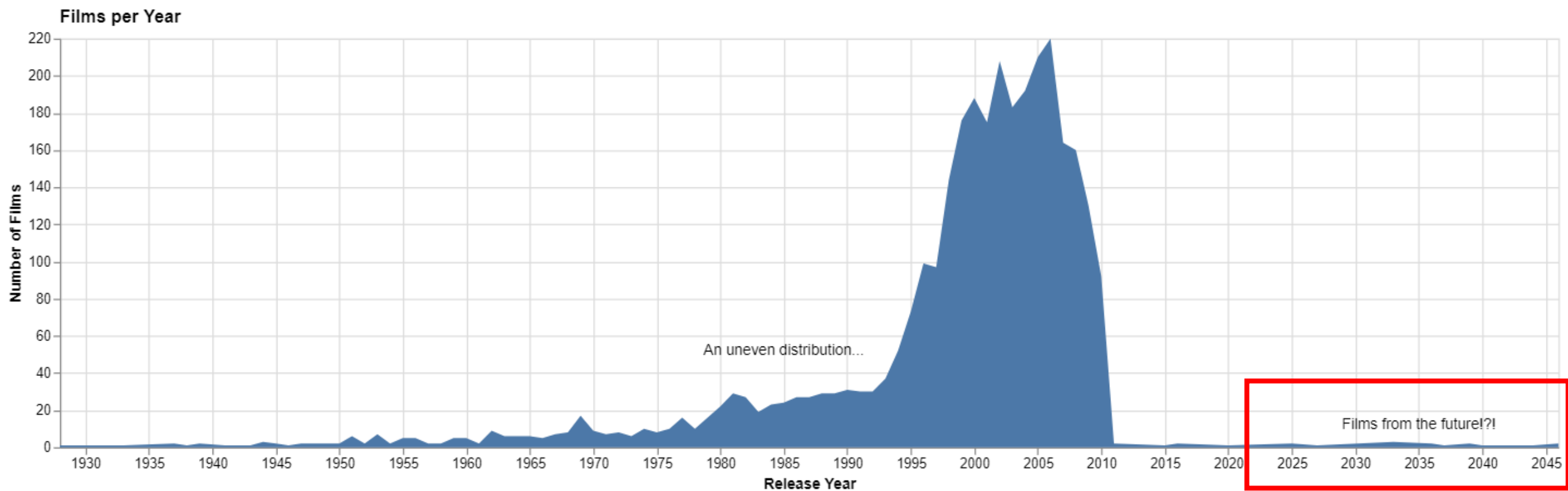
1. Relación entre ingresos en USA y el Mundo



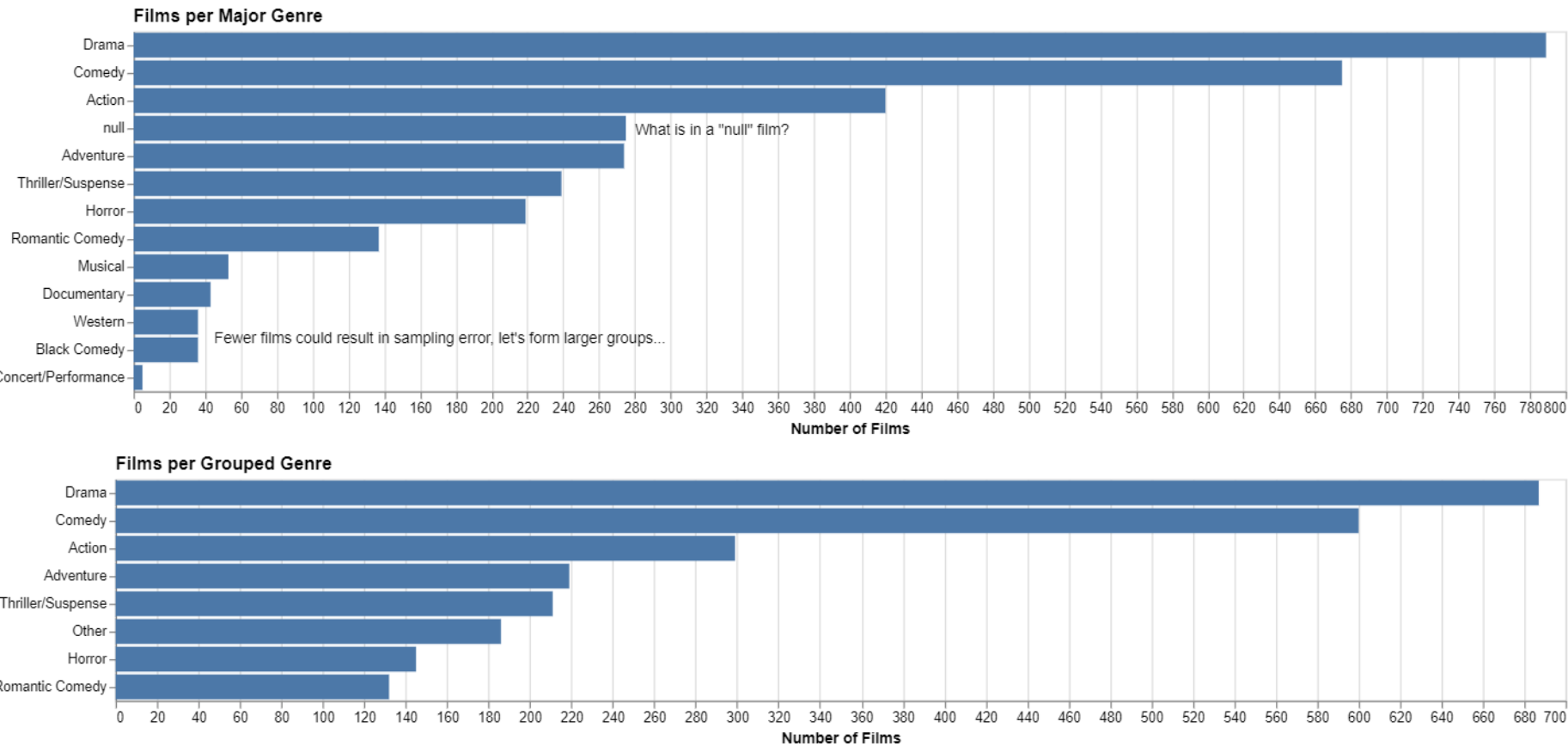
1. Una asociación muy importante.
2. Algunos outliers explicables (?).
3. Quizá convenga seguir el análisis considerando únicamente las ganancias mundiales.

Sanity check

- Distribución irregular



2. ¿Cómo varían las ganancias según el género? Sanity check previo



Se Hizo:

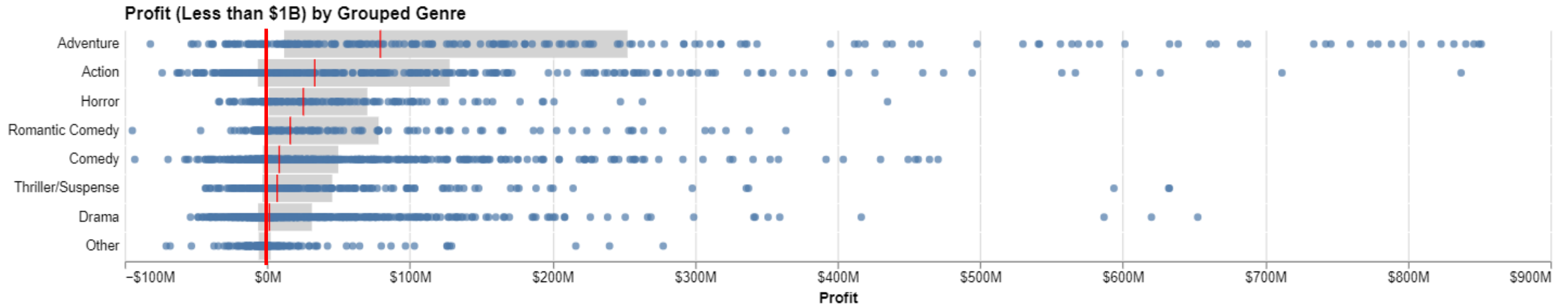
- Géneros con pocas películas reagrupados.
 - Black Comedy -> Comedy
 - Todos los géneros con menos de 100 películas se reagrupan en "Other"
- Ingresos Mundiales > 0
- 2010 >= Año Estreno >= 1995
- Ganancias = Ingresos - Costos

Ahora todos los géneros tienen más de 100 películas.

2. ¿Cómo varían las ganancias según el género?

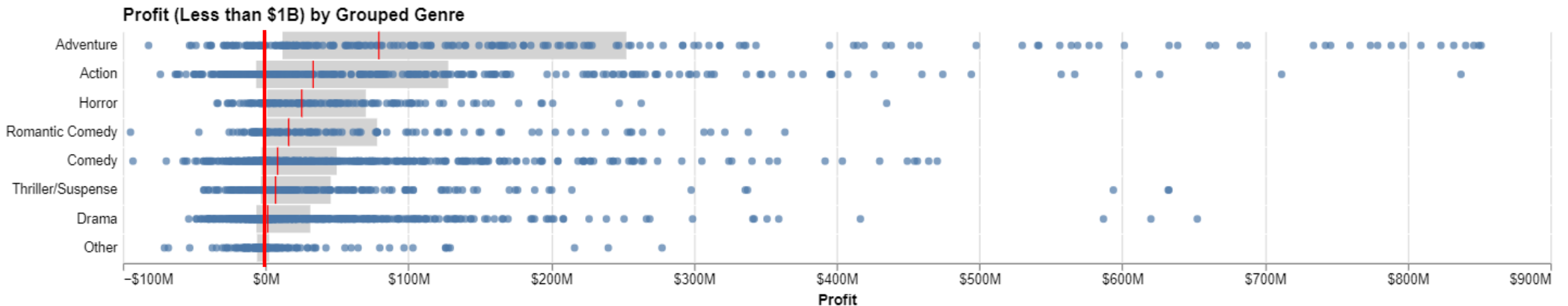


Se quitan los outliers de más de 1.000.000.000 de ganancia.



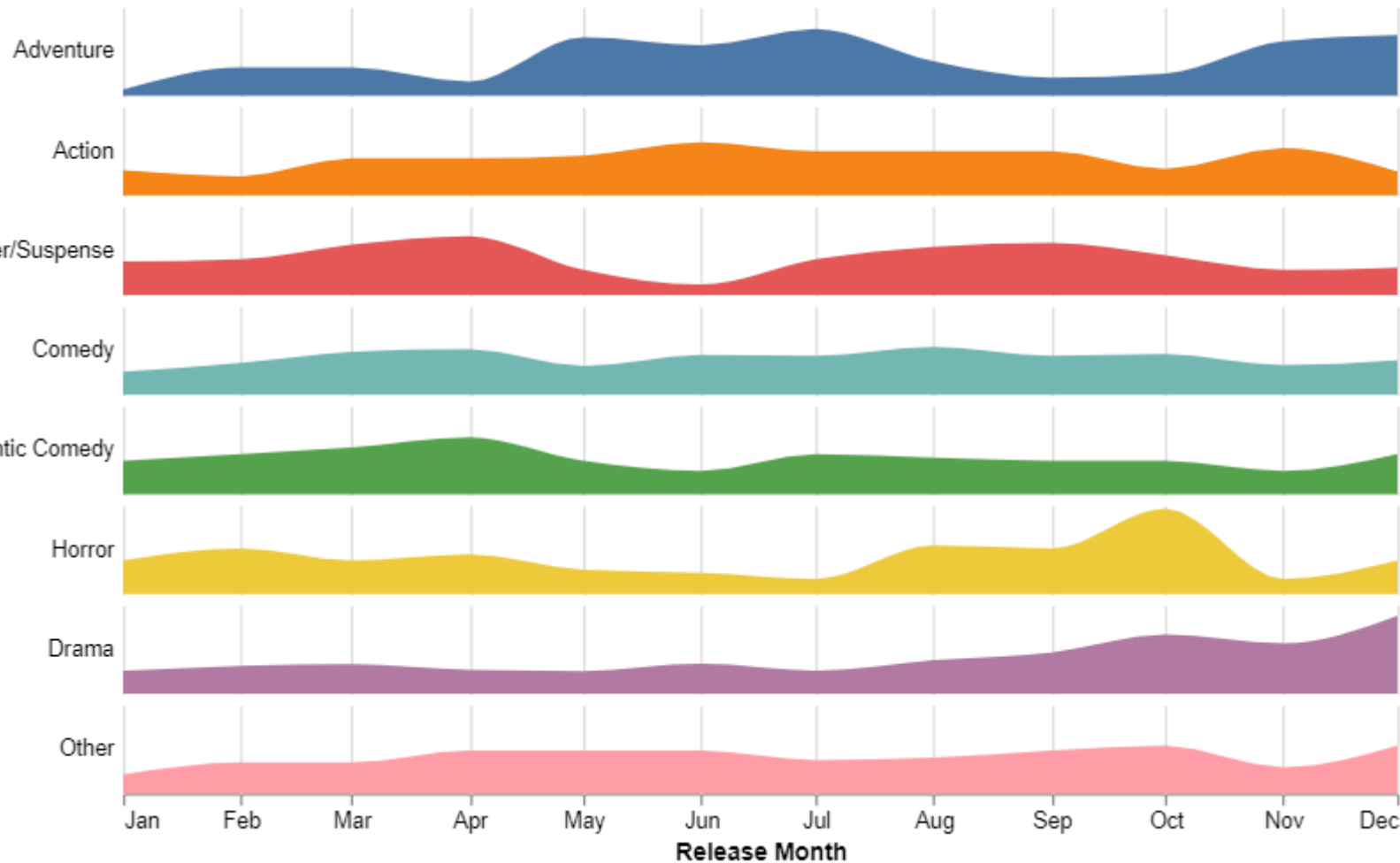
2. ¿Cómo varían las ganancias según el género?

- El 75% de las películas dan alguna ganancia.
- El 25% de las películas de acción y aventuras ganan más de \$100M.



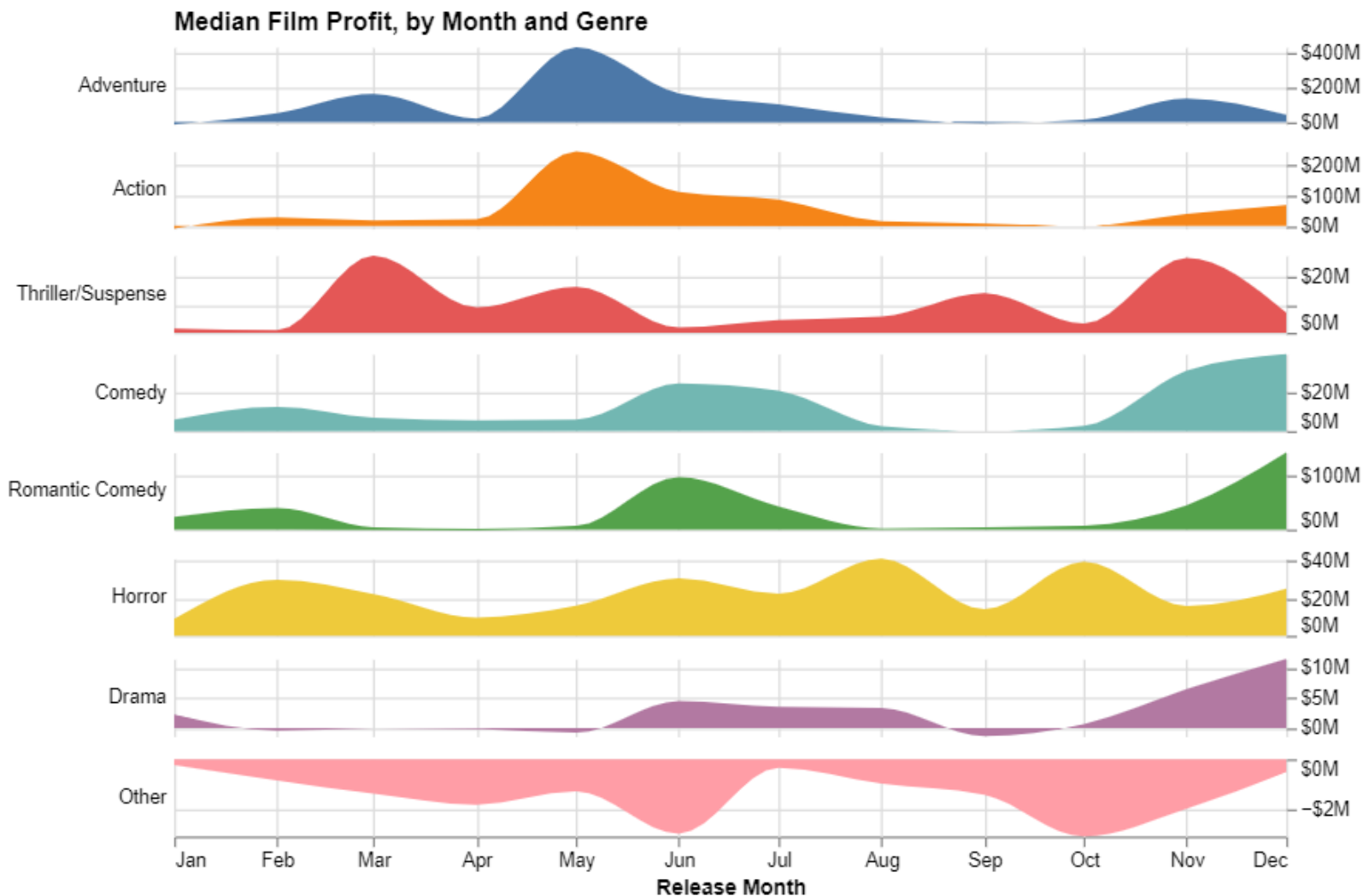
3. Distribución de fechas de estreno según el género.

Percentage of Releases per Month by Genre



- Aventuras y Acción alcanzan su pico en verano
- Thrillers y Comedias tienen su punto máximo en primavera
- Las películas de terror en torno a Halloween
- Los dramas tienen su pico antes de la temporada de los Oscar

Pregunta extra: ¿Las ganancias se alinean con las fechas de estreno?



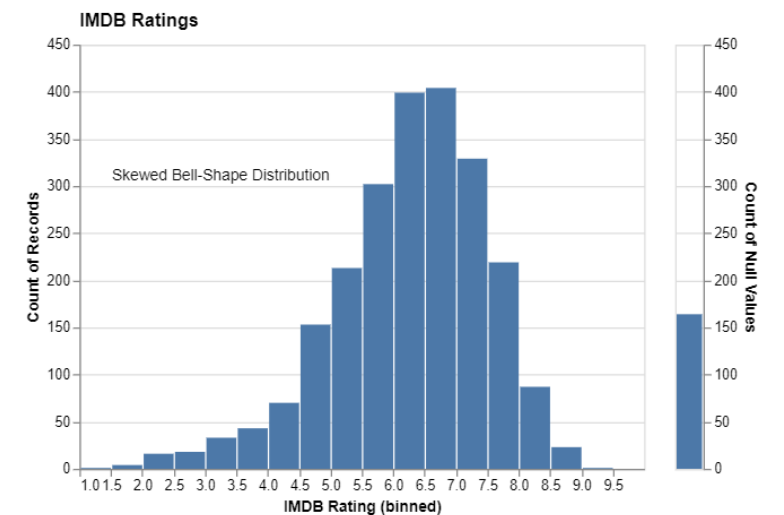
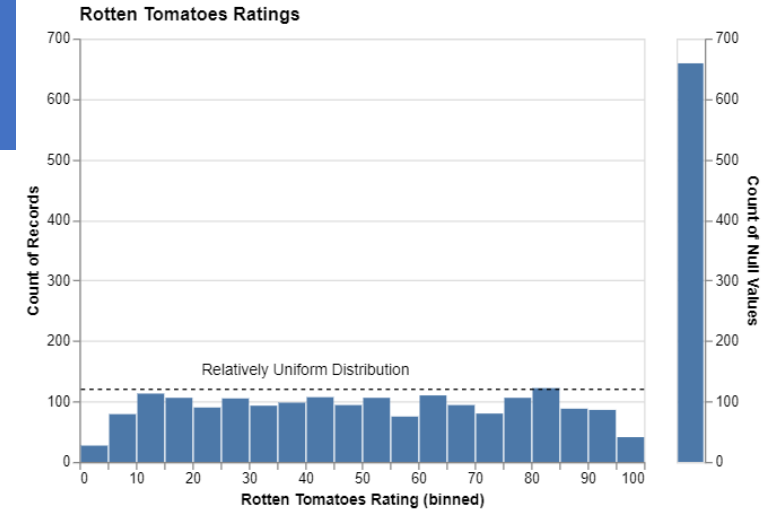
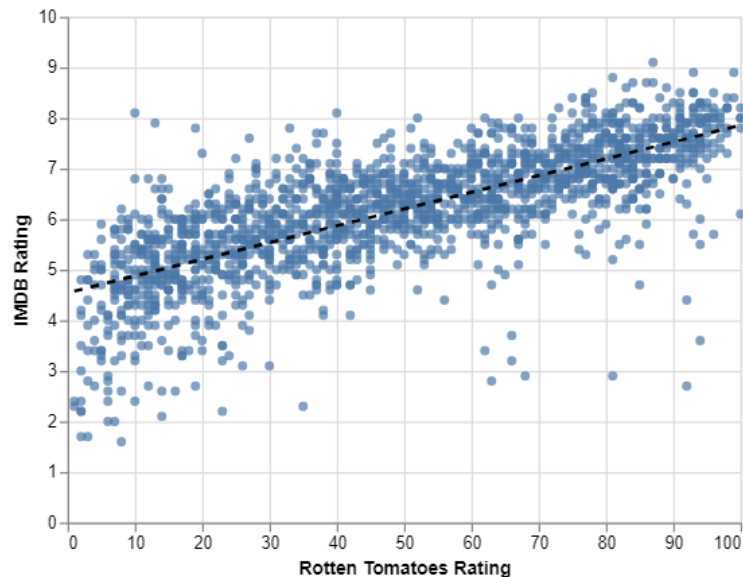
Algo:

- La aventura y la acción son más rentables en mayo
- las comedias en verano o fin de año.
- Las ganancias del terror son robustos
- mientras que a Otros no les va bien.

4. Relación entre ratings y el éxito financiero

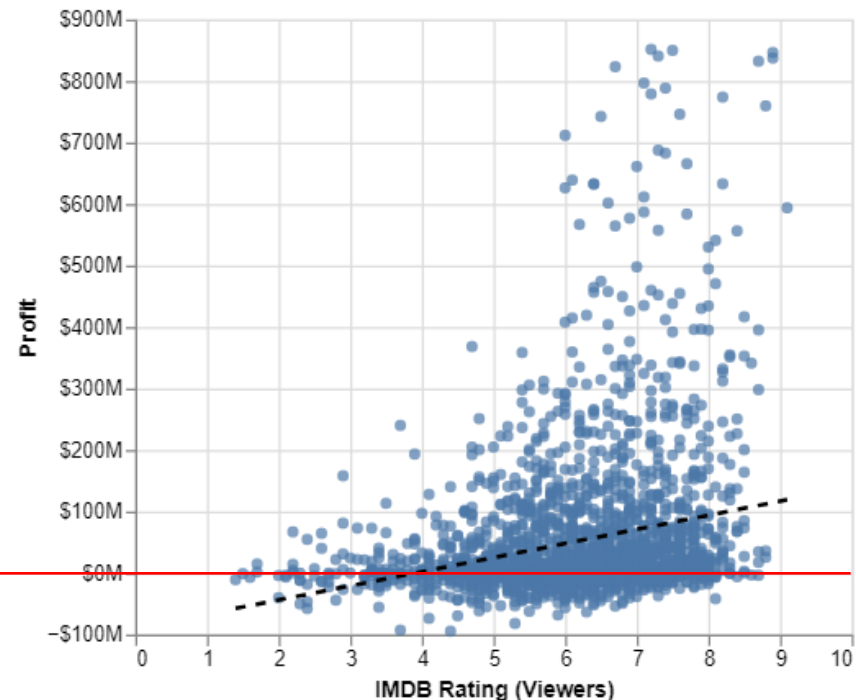
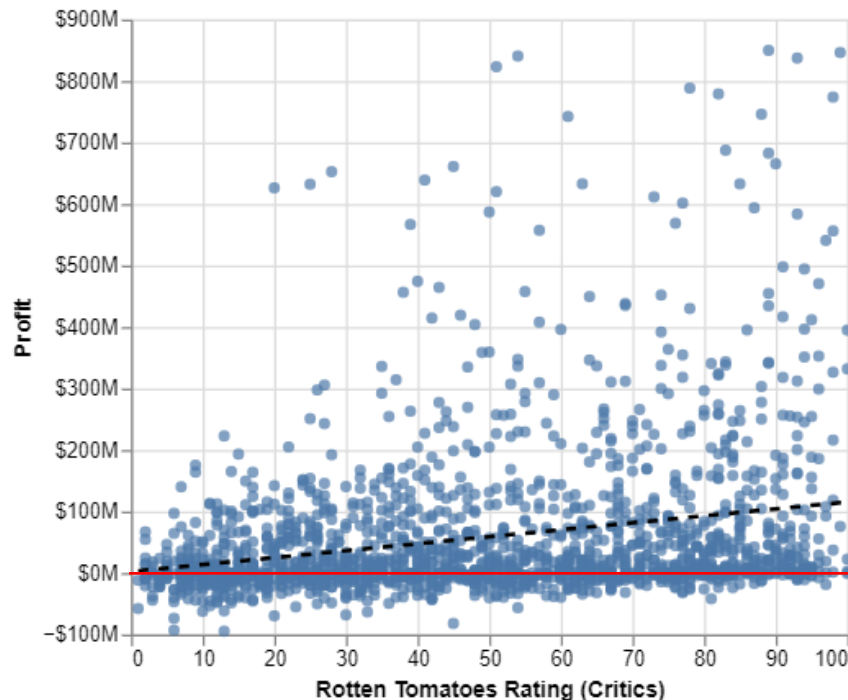
Sanity check previo

- Distribuciones diferentes.
- Ambos exhiben valores nulos.
- Los ratings están bastante correlacionados.



4. Relación entre ratings y el éxito financiero

- Asociación positiva entre ganancias y ratings.
- La regresión lineal lo muestra aunque no demasiado bien.
- Se aprecia que los más lucrativos tienen mejores críticas en IMDB.
- Muchos films con muy buenas críticas no hacen dinero.



4. Relación entre ratings y el éxito financiero (ganancias < \$200M)

- En Rotten Tomatoes se puede hacer dinero aun si no les gustan las películas.
- Mejores críticas implican mejor dinero, pero no hay garantías.
- **Conclusión?: Si además de buscar buenas críticas, se buscase el género y la fecha de estreno adecuados, podrían mejorarse las ganancias.**

