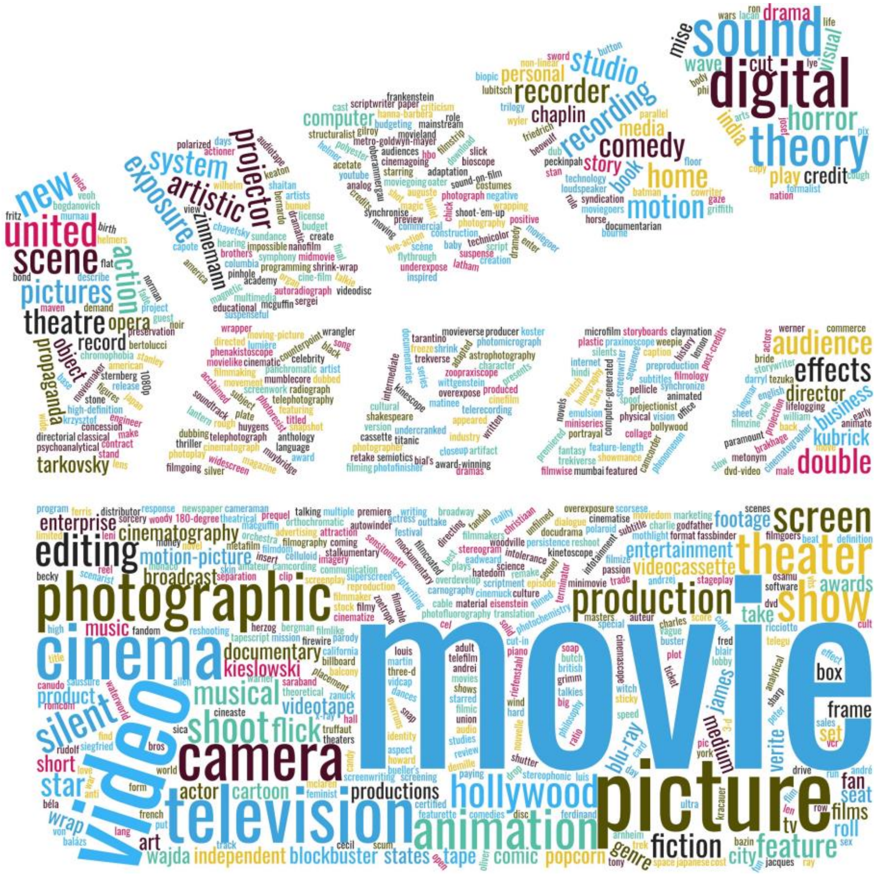


IMDb

Datasets Analysis

Bruno Vaz, Fátima Barros, Maria João Lavoura

Professor: Álvaro Figueira | Data Visualization

January 8, 2021

IMDb Datasets

Internet Movie Database

People **rely on other's ratings** for choosing their next movie or show

Title Akas

Titles translations and regions

Title Basics

Titles categories, genres, year, runtime

Title Crew

Directors and Writers

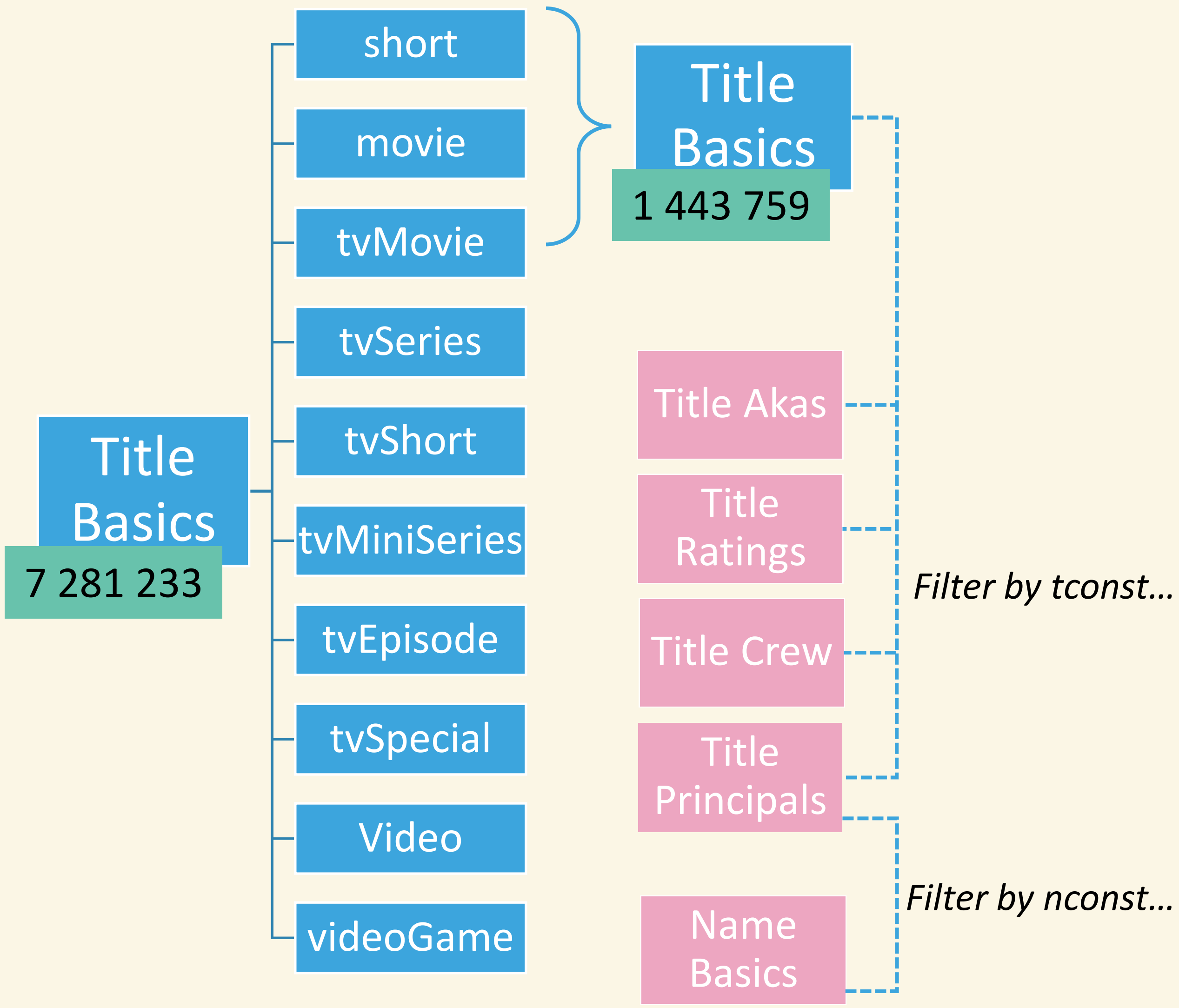
Title Principals

Job of people in the title

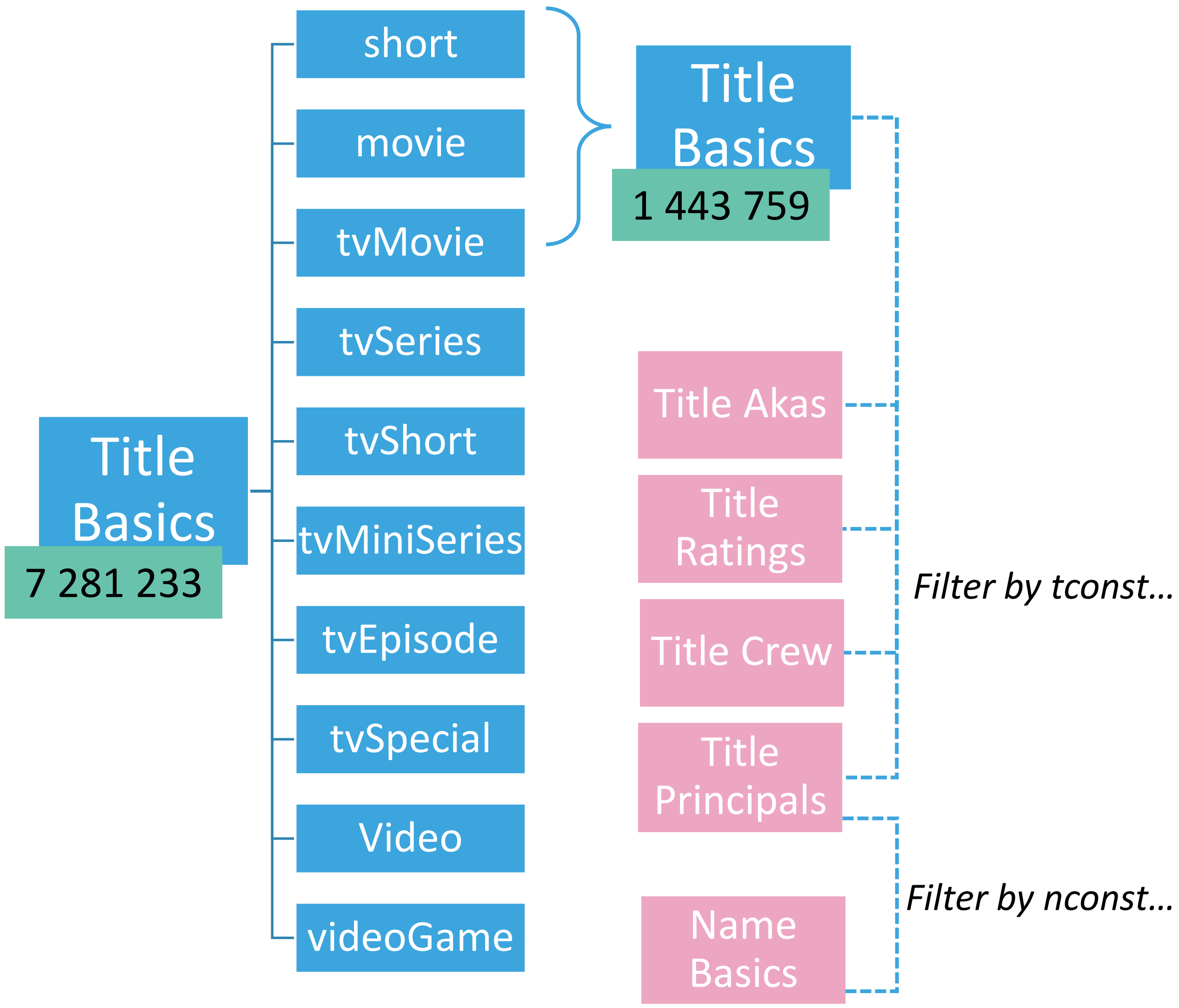
Name Basics

Professions and titles people are known for

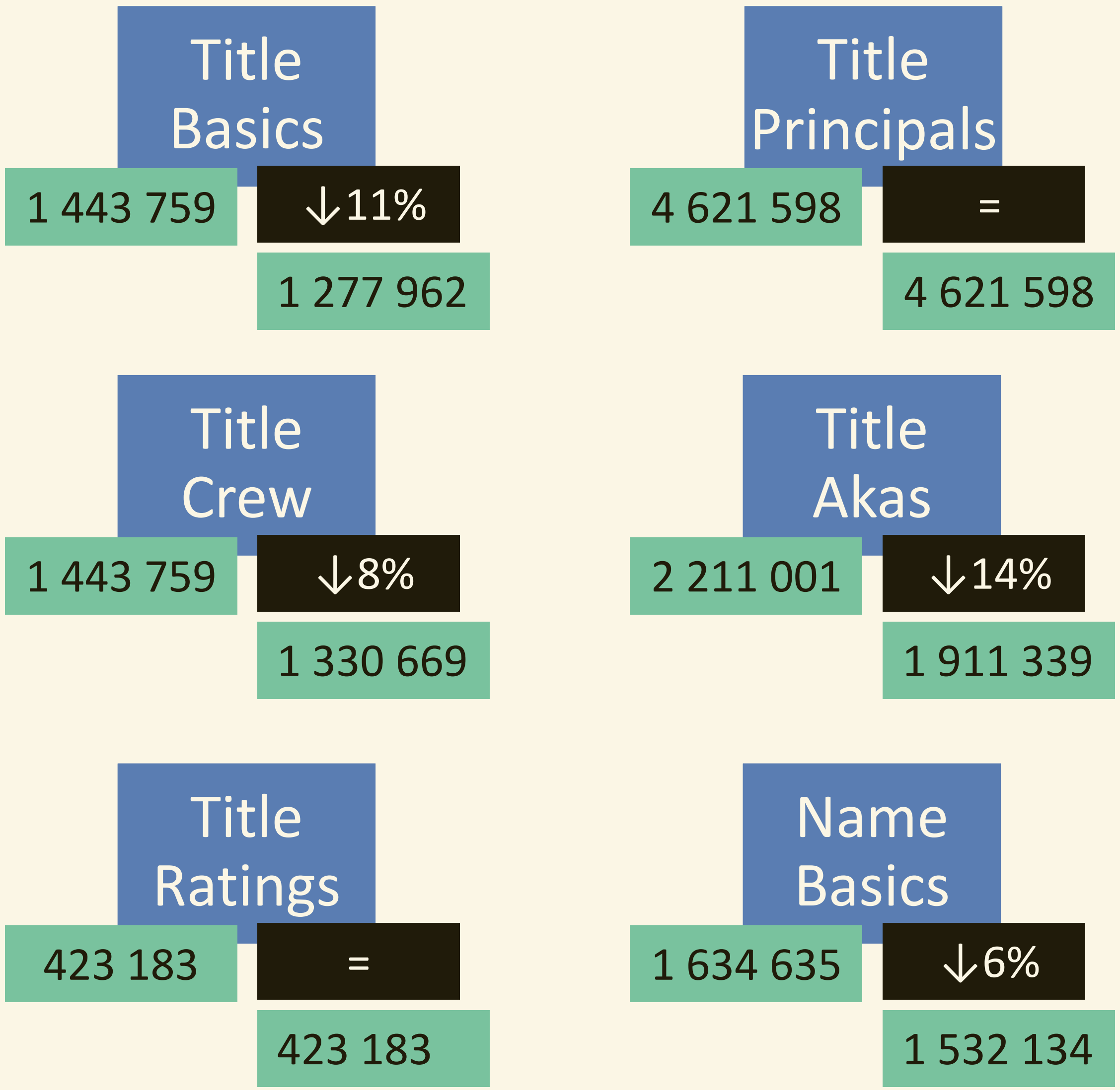
Filtering


Title Basics
7 281 233
short
movie
tvMovie
tvSeries
tvShort
tvMiniSeries
tvEpisode
tvSpecial
Video
videoGame
Title Basics
1 443 759
Title Akas
Title Ratings
Title Crew
Title Principals
Name Basics
Filter by tconst…
Filter by nconst…

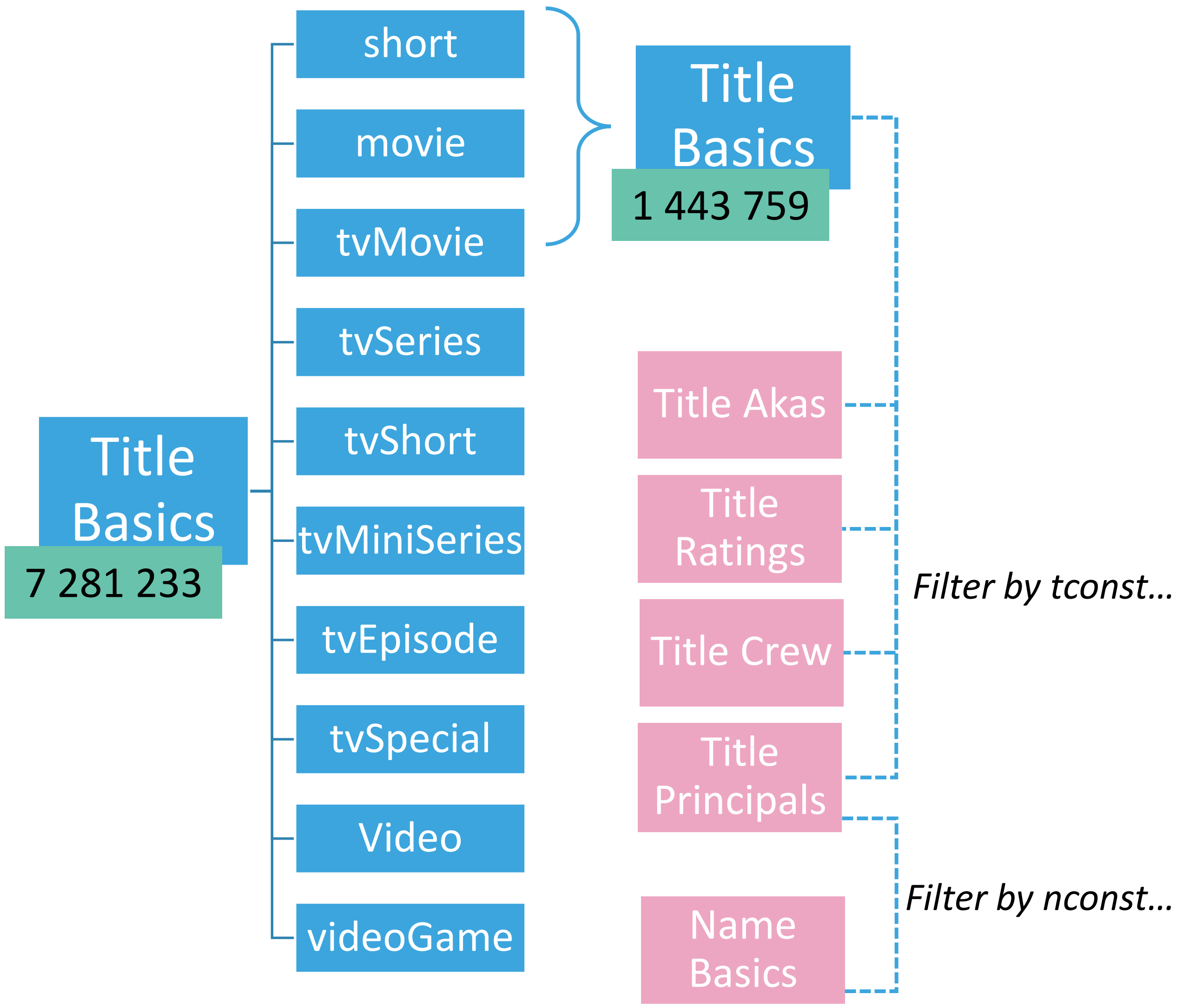
Filtering



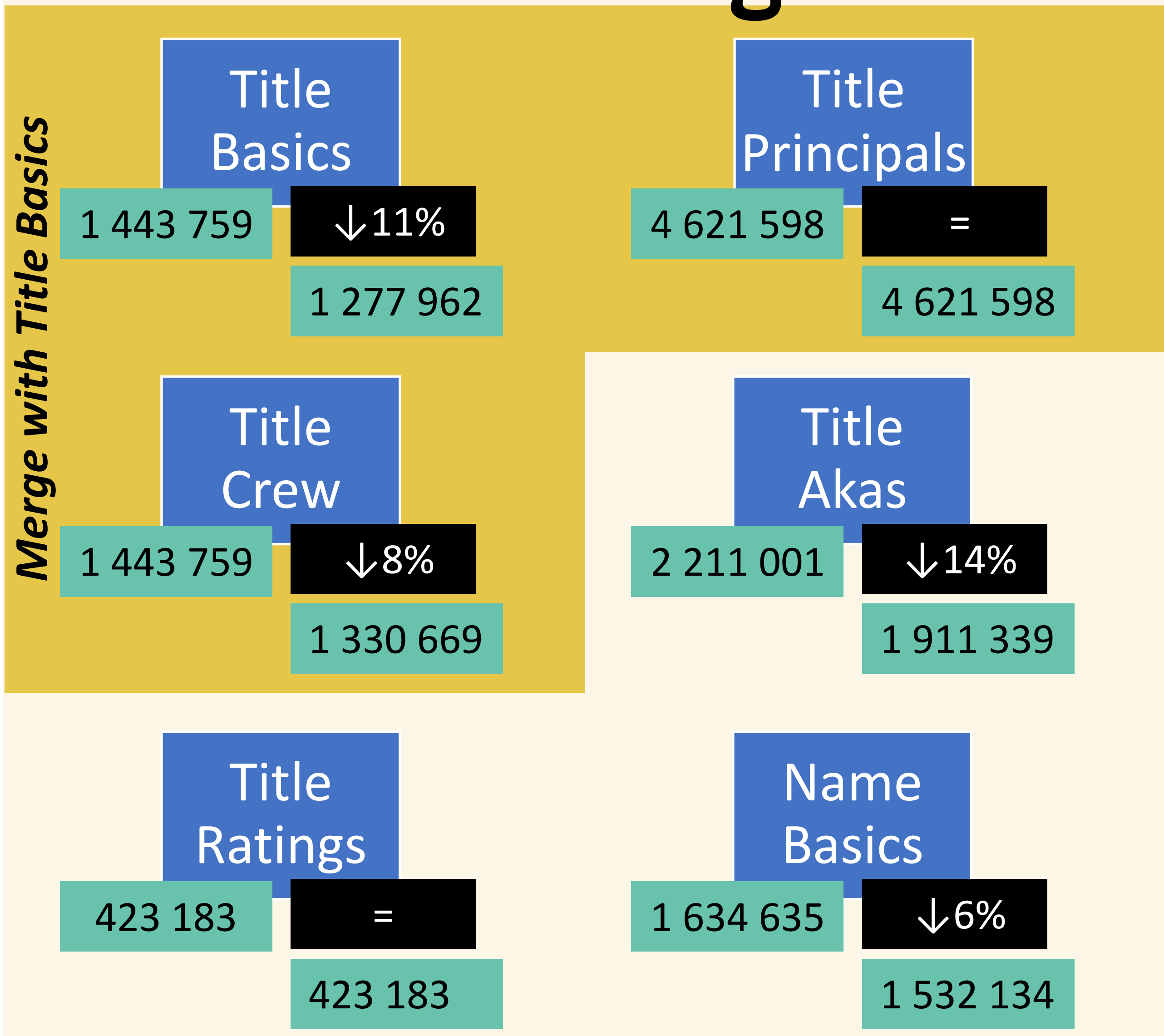
Cleaning



Filtering



Cleaning



Proposed Hypotheses

"*Is the title's success influenced by the number of translations?*"

"*Does the cast/crew of a movie make it successful?*"

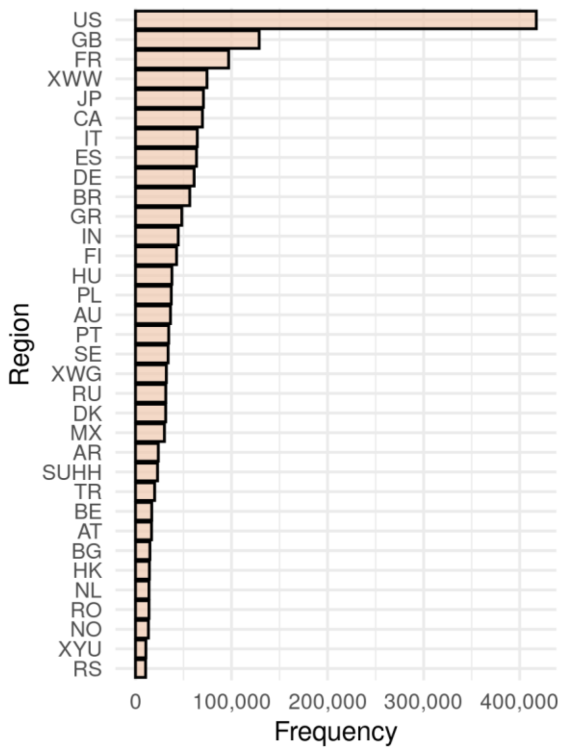
Titles per Region

From dataset: *Title Akas*

↓

Frequency of translated titles per region

(values ≥ 10 000)

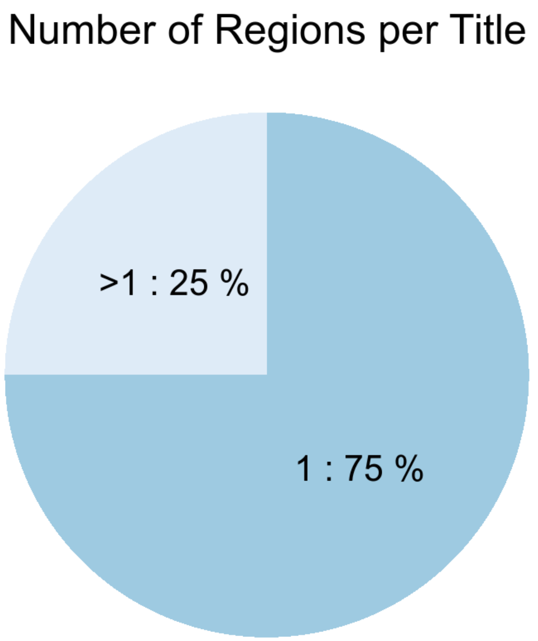


Titles per Region

From dataset: *Title Akas*

↓

Counts or regions per title

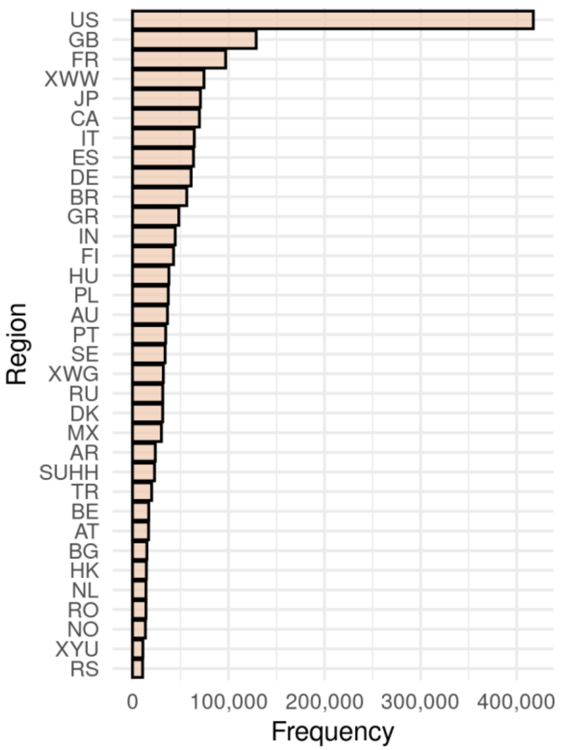


From dataset: *Title Akas*

↓

Frequency of translated titles per region

(values ≥ 10 000)

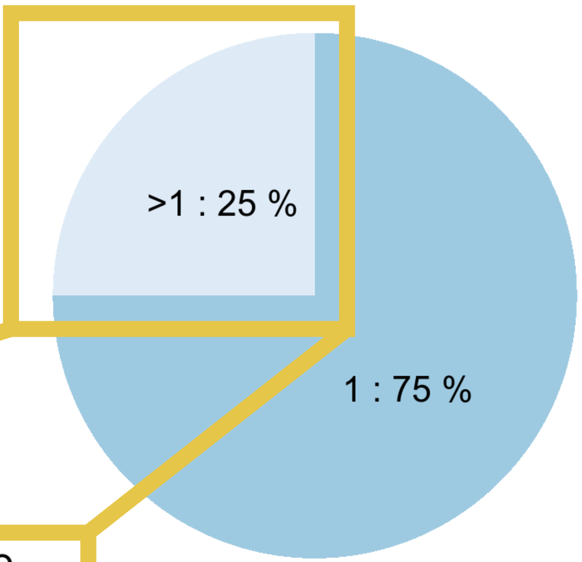


Titles per Region

From dataset: *Title Akas*

↓

Counts or regions per title

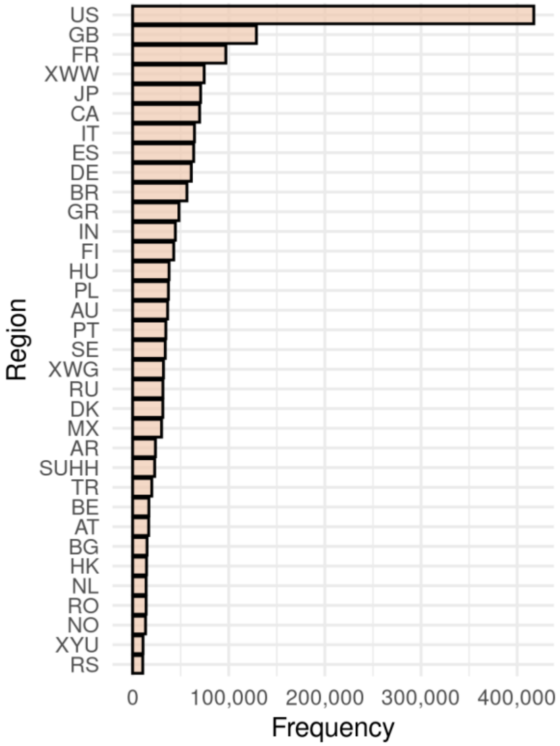


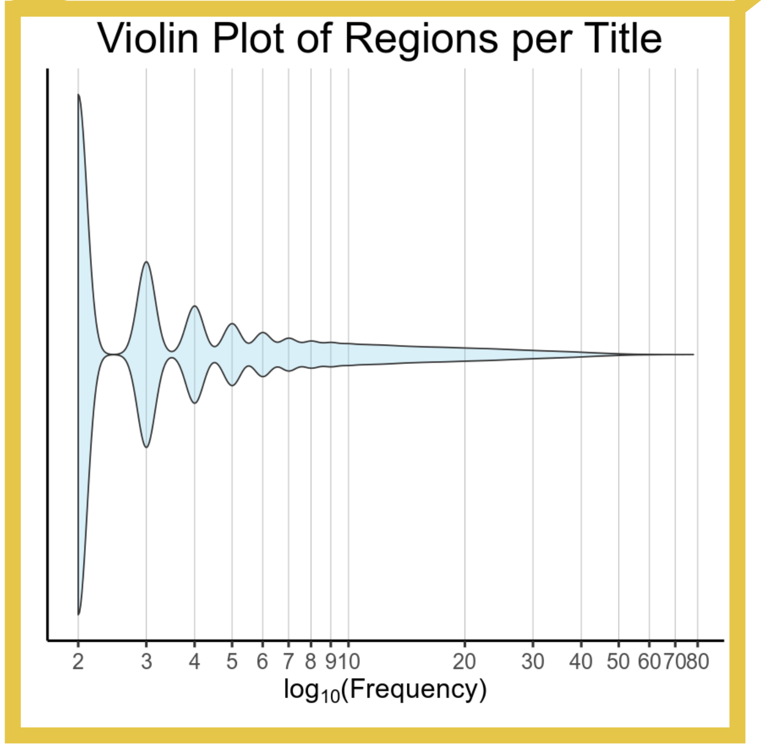
From dataset: *Title Akas*

↓

Frequency of translated titles per region

(values ≥ 10 000)



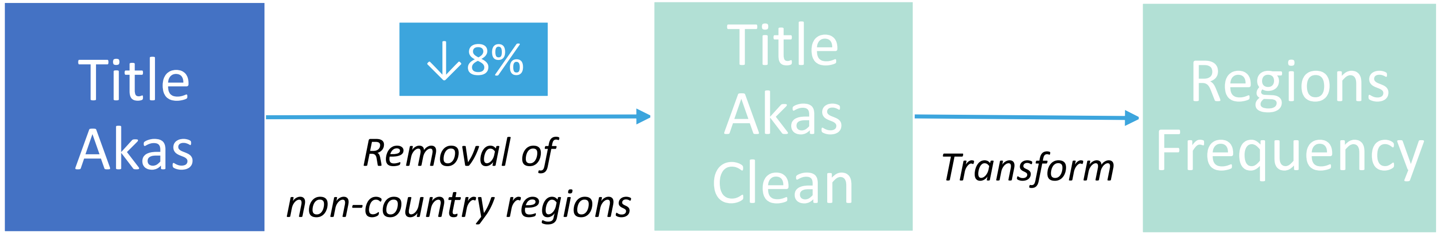


Titles per Region

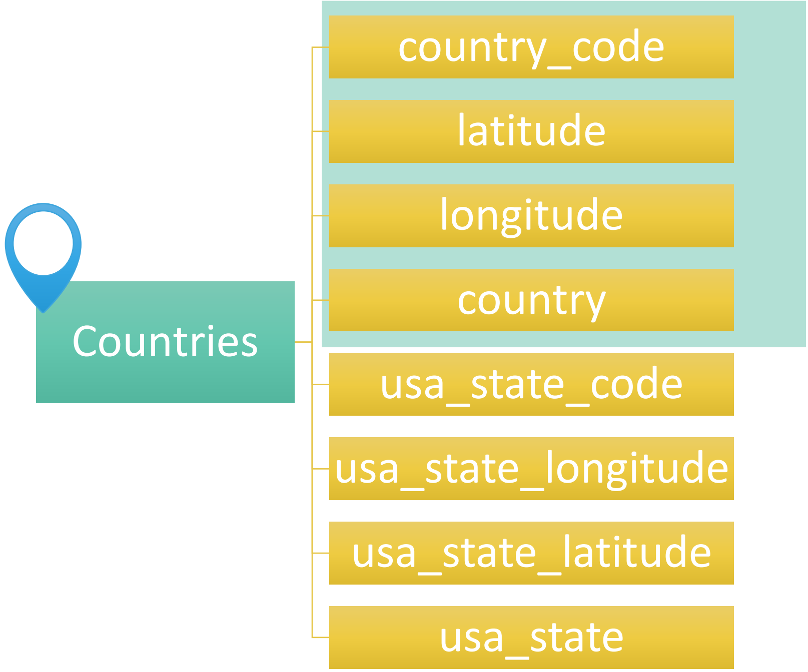
Using an extra dataset...

- Countries
  - country_code
  - latitude
  - longitude
  - country
  - usa_state_code
  - usa_state_longitude
  - usa_state_latitude
  - usa_state

Titles per Region



Using an extra dataset...



Titles per Region

Title Akas → ↓8% Removal of non-country regions → Title Akas Clean → Transform → Regions Frequency

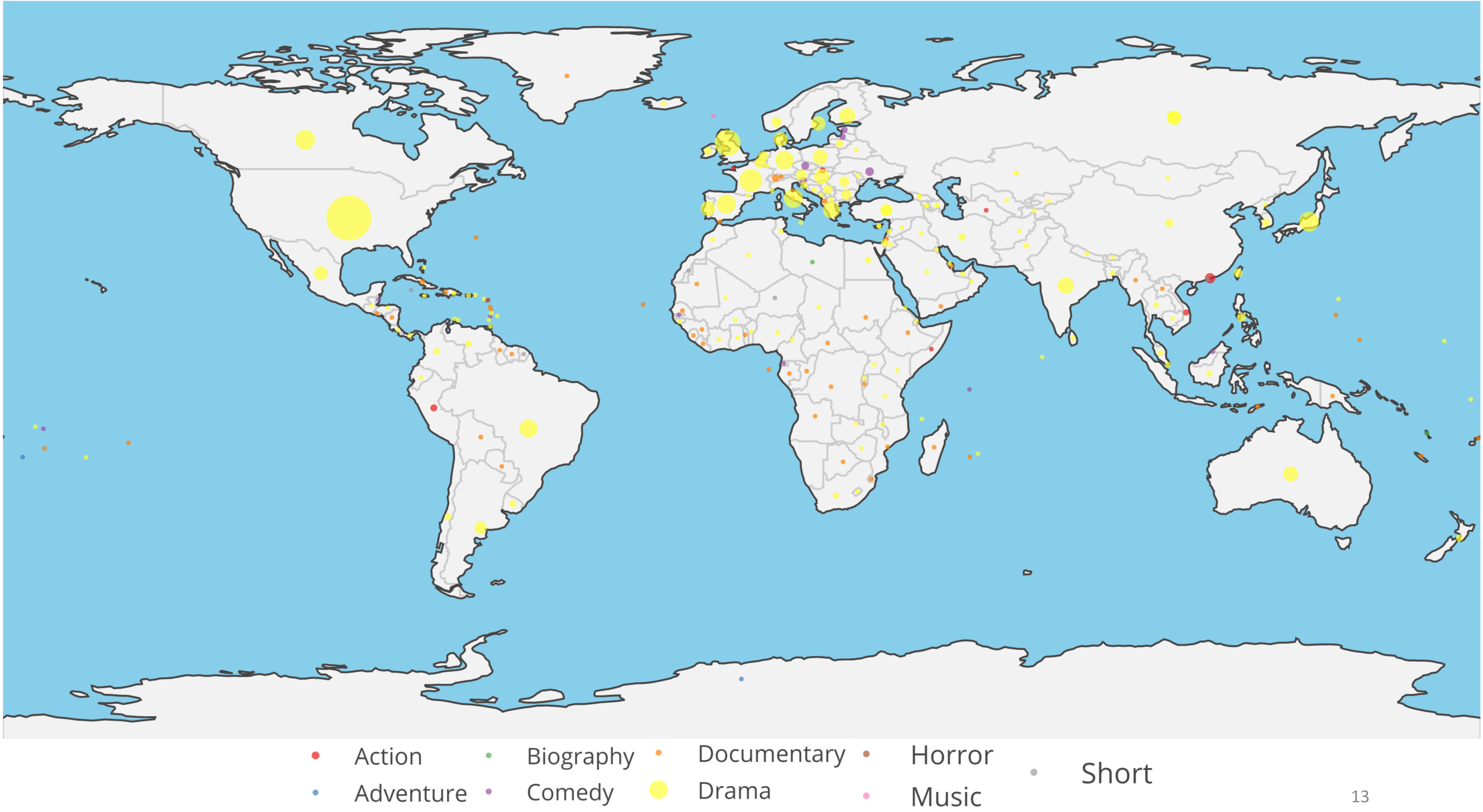
MERGE

Using an extra dataset...

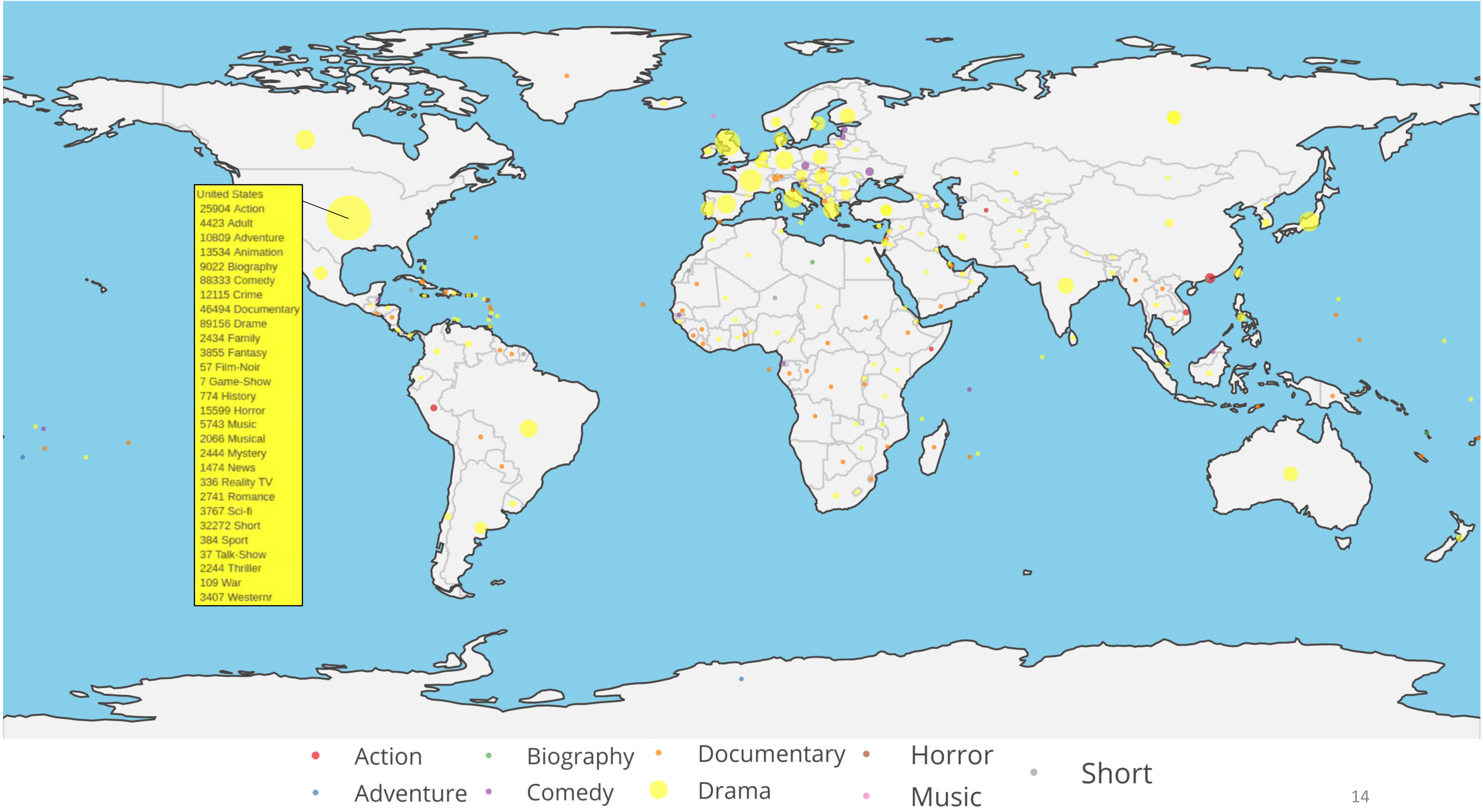
Countries
- country_code
- latitude
- longitude
- country
- usa_state_code
- usa_state_longitude
- usa_state_latitude
- usa_state

A data.frame: 223 × 34

region	Action	Adult	Adventure	Animation	Biography	Comedy	Crime	Documentary	Drama	⋯	Sport	Talk-Show	Thriller	War	Western	Freq	latitude	longitude	country	most_frequent_genre
<chr>	<int>	<int>	<int>	<int>	<int>	<int>	<int>	<int>	<int>	⋯	<int>	<int>	<int>	<int>	<int>	<int>	<dbl>	<dbl>	<chr>	<chr>
AD	0	0	0	0	0	1	0	3	5	⋯	0	0	0	0	0	22	42.546245	1.601554	Andorra	Drama
AE	24	0	18	15	18	69	18	103	212	⋯	0	0	2	1	0	711	23.424076	53.847818	United Arab Emirates	Drama
AF	6	0	3	2	2	1	3	17	38	⋯	0	0	0	0	0	98	33.939110	67.709953	Afghanistan	Drama
AG	1	0	0	0	0	0	1	1	2	⋯	0	0	0	0	0	12	17.060816	-61.796428	Antigua and Barbuda	Horror
AL	80	1	52	41	5	89	27	583	339	⋯	0	0	1	3	1	1335	41.153332	20.168331	Albania	Documentary
AM	10	0	9	7	5	39	4	74	86	⋯	0	0	0	2	0	277	40.069099	45.038189	Armenia	Drama
AM	10	0	9	7	5	39	4	74	86	⋯	0	0	0	2	0	277	40.069099	45.038189	Armenia	Drama
AN	1	0	0	0	0	2	0	4	7	⋯	0	0	0	0	0	19	12.226079	-69.060087	Netherlands Antilles	Drama
AO	0	0	1	0	0	3	0	17	7	⋯	0	0	0	0	0	48	-11.202692	17.873887	Angola	Documentary
AQ	0	0	3	0	0	0	0	1	0	⋯	0	0	0	0	0	4	-75.250973	-0.071389	Antarctica	Adventure
AR	2367	15	1290	731	739	4814	1341	1738	6113	⋯	3	0	157	3	184	23743	-38.416097	-63.616672	Argentina	Drama
AS	0	0	0	0	0	2	0	1	0	⋯	0	0	0	0	0	4	-14.270972	-170.132217	American Samoa	Comedy
AT	1305	13	1014	136	505	3389	1011	1360	3872	⋯	7	0	52	10	422	16741	47.516231	14.550072	Austria	Drama
AU	3556	46	1816	908	1216	6852	1431	4459	8089	⋯	10	0	238	21	143	36444	-25.274398	133.775136	Australia	Drama
AW	6	0	0	0	2	1	0	1	7	⋯	0	0	0	0	0	18	12.521110	-69.968338	Aruba	Drama
AZ	59	0	26	6	19	53	21	27	119	⋯	0	0	0	2	0	377	40.143105	47.576927	Azerbaijan	Drama
AZ	59	0	26	6	19	53	21	27	119	⋯	0	0	0	2	0	377	40.143105	47.576927	Azerbaijan	Drama
BA	71	0	38	12	29	124	25	120	221	⋯	0	0	2	1	0	809	43.915886	17.679076	Bosnia and Herzegovina	Drama



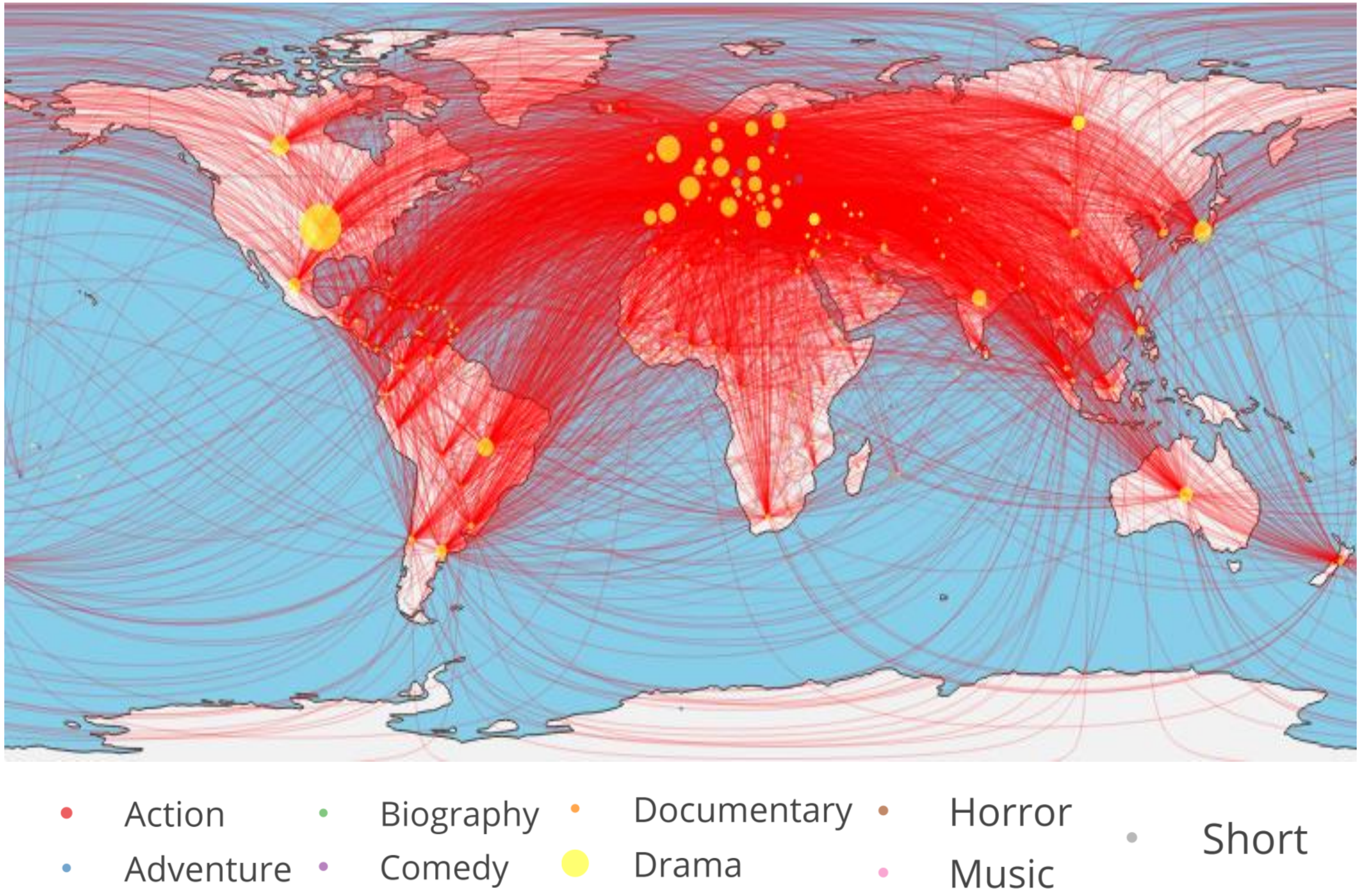
https://github.com/fsbarros98/IMDb_analysis/blob/main/graphic_results/regions_images/worldmap.html



https://github.com/fsbarros98/IMDb_analysis/blob/main/graphic_results/regions_images/worldmap.html

Titles per Region

Links between the regions



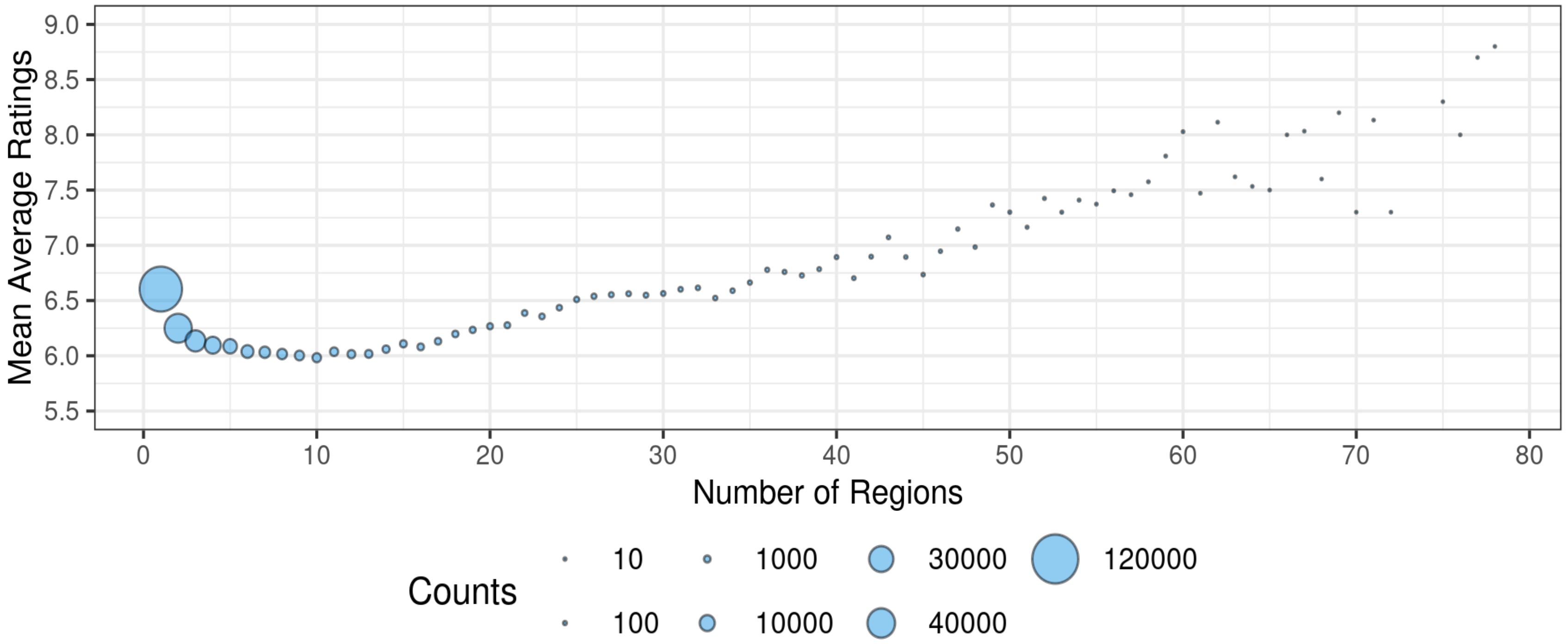
European continent has a lot of connections with the remaining countries

titles that are translated to more than one region are typically translated to some European country

https://github.com/fsbarros98/IMDb_analysis/blob/main/graphic_results/regions_images/worldmap.html

Titles per Region

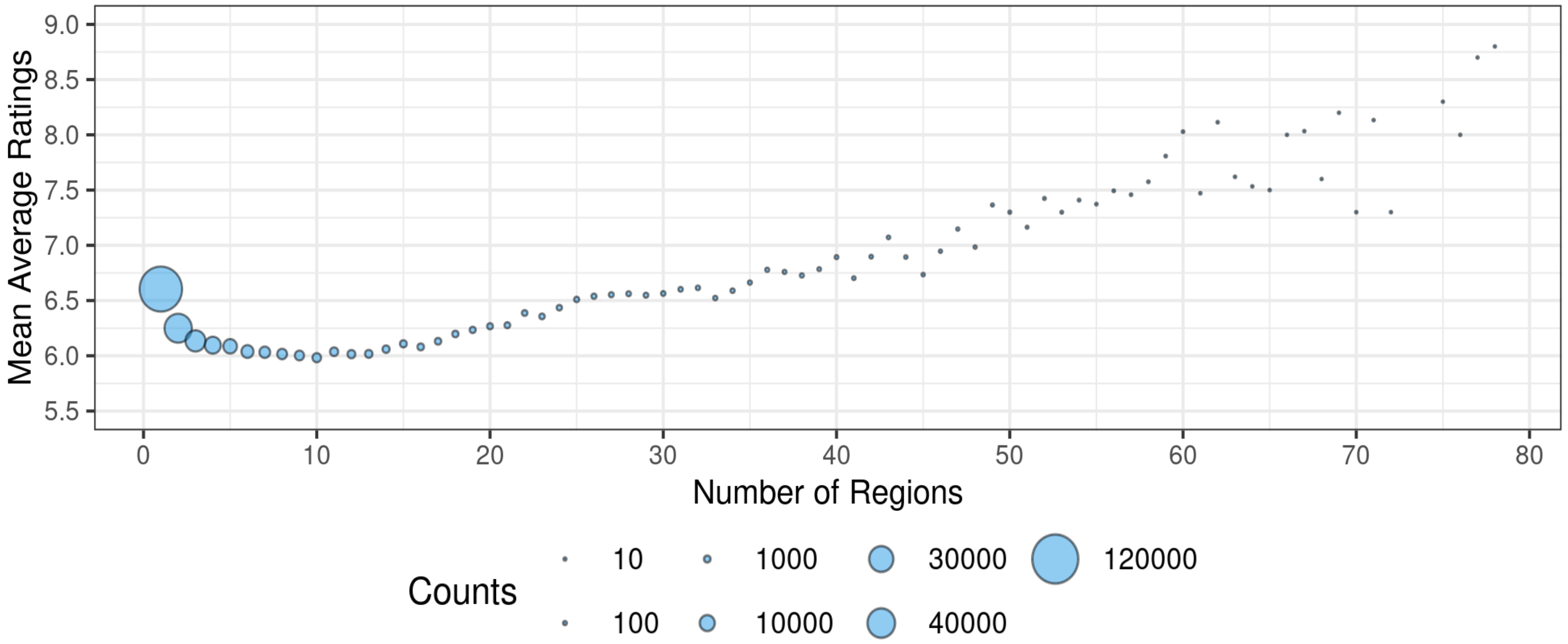
Number of Regions vs. Mean Average Ratings



area of the circles represents the number of titles that were translated to a certain number of region

Titles per Region

Number of Regions vs. Mean Average Ratings



area of the circles represents the number of titles that were translated to a certain number of region

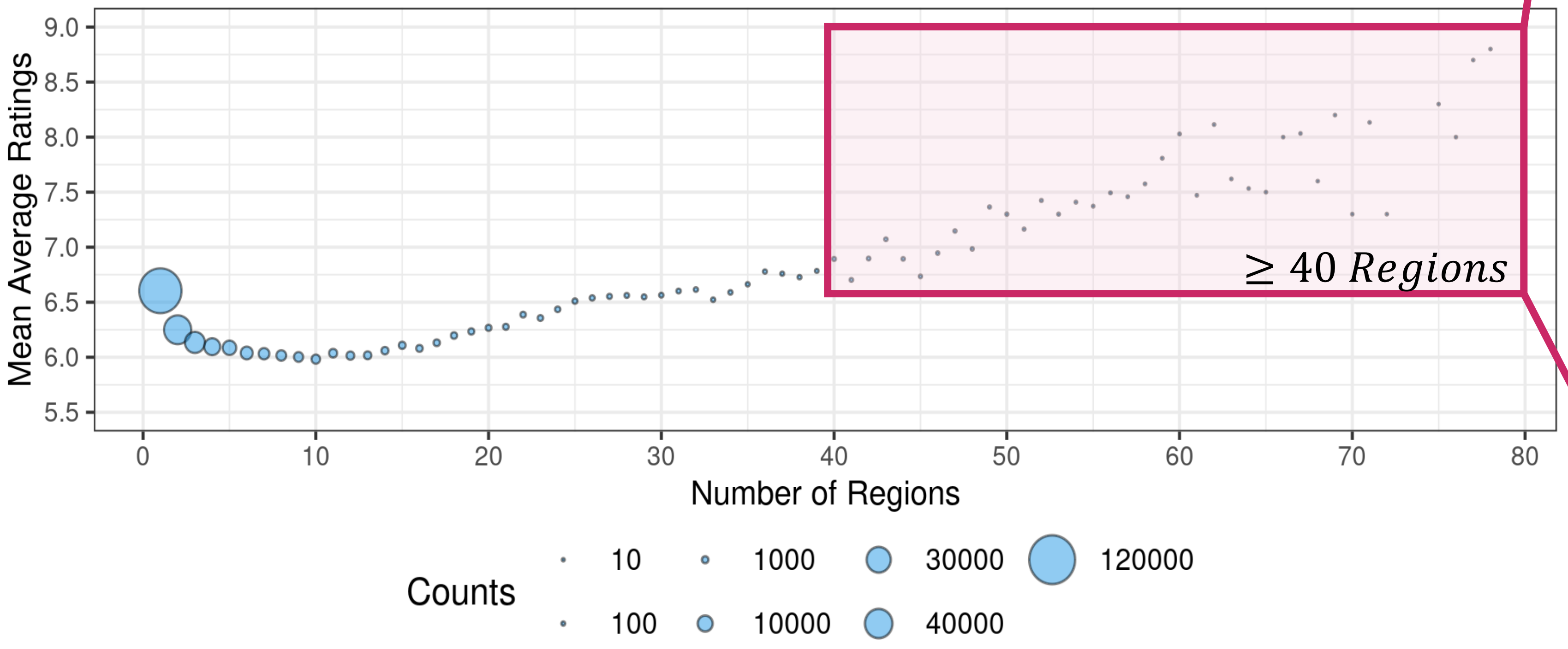
Number of regions

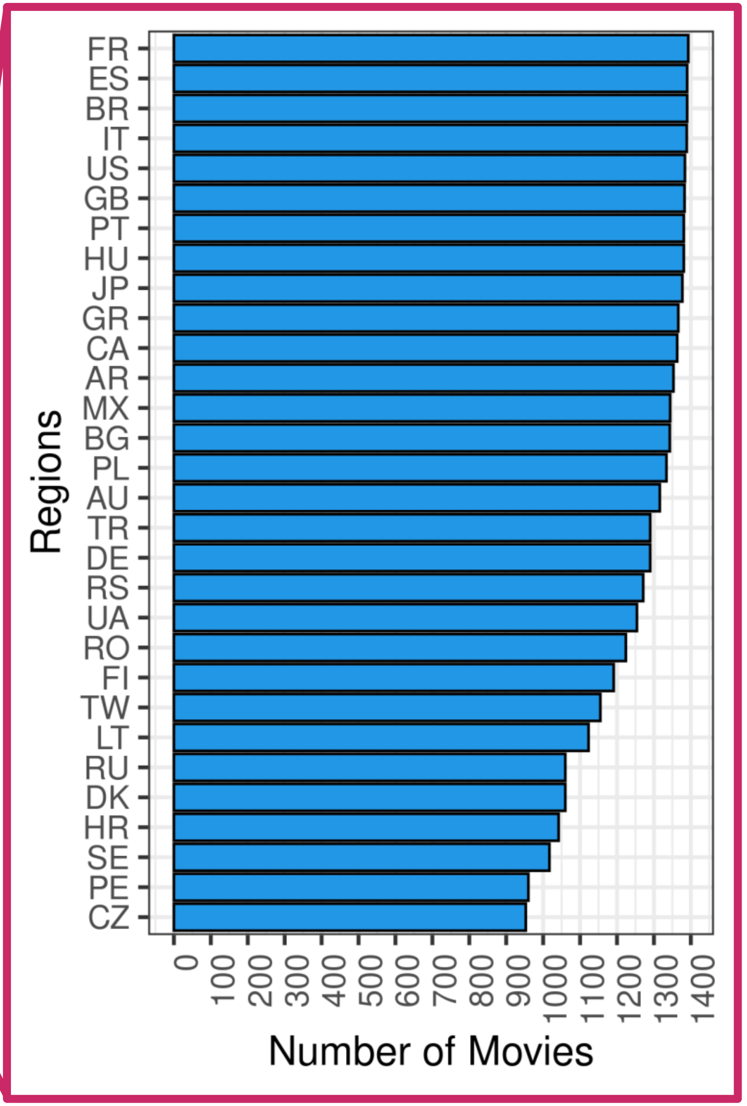
Counts

Mean Average Ratings

Titles per Region

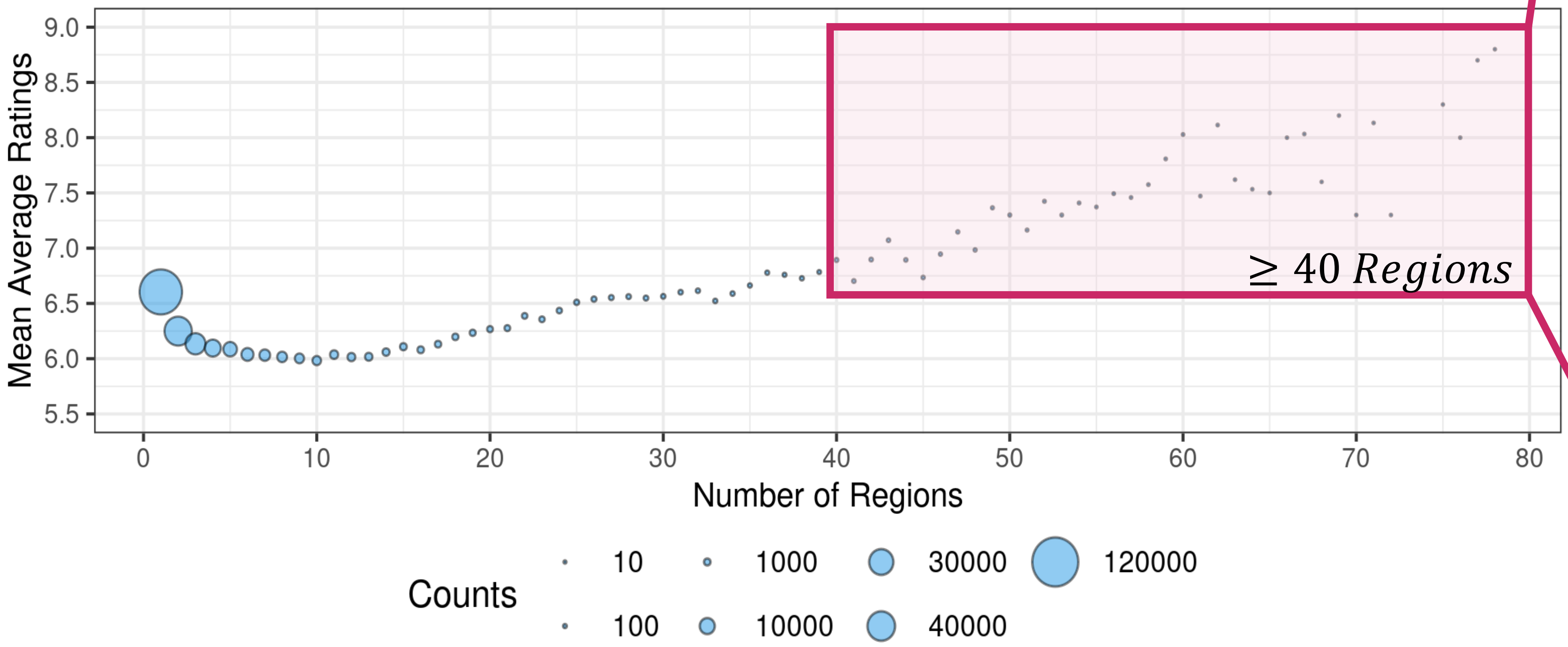
Number of Regions vs. Mean Average Ratings

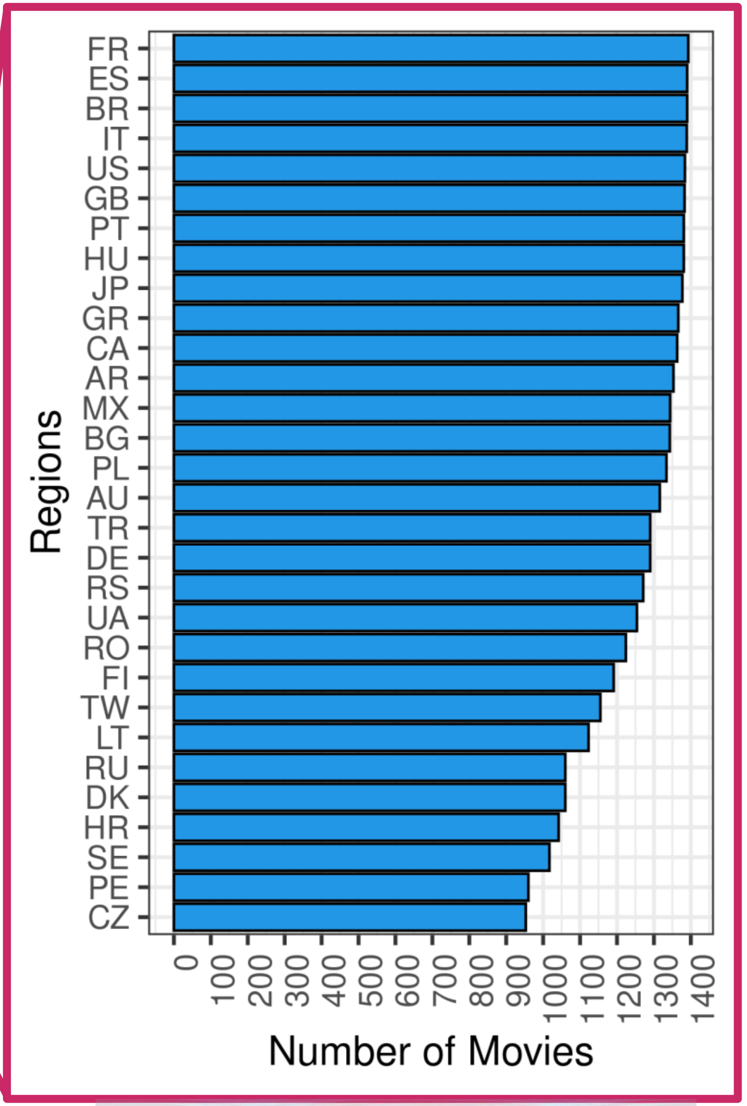




Titles per Region

Number of Regions vs. Mean Average Ratings

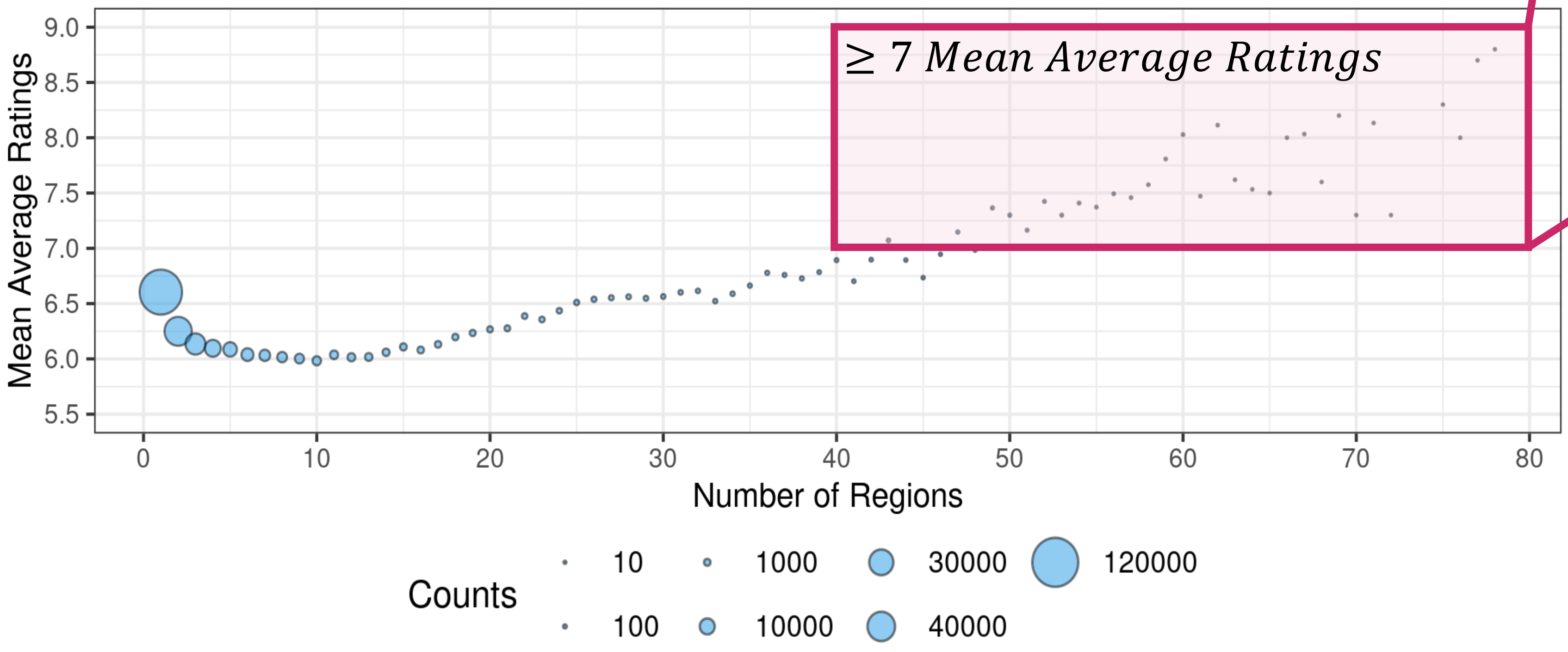


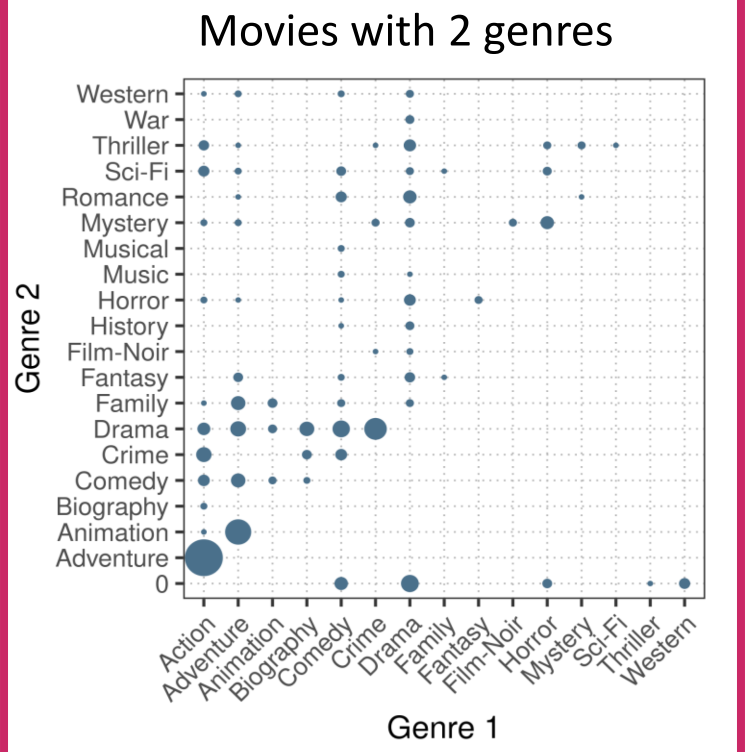


Most of these countries are European

Titles per Region

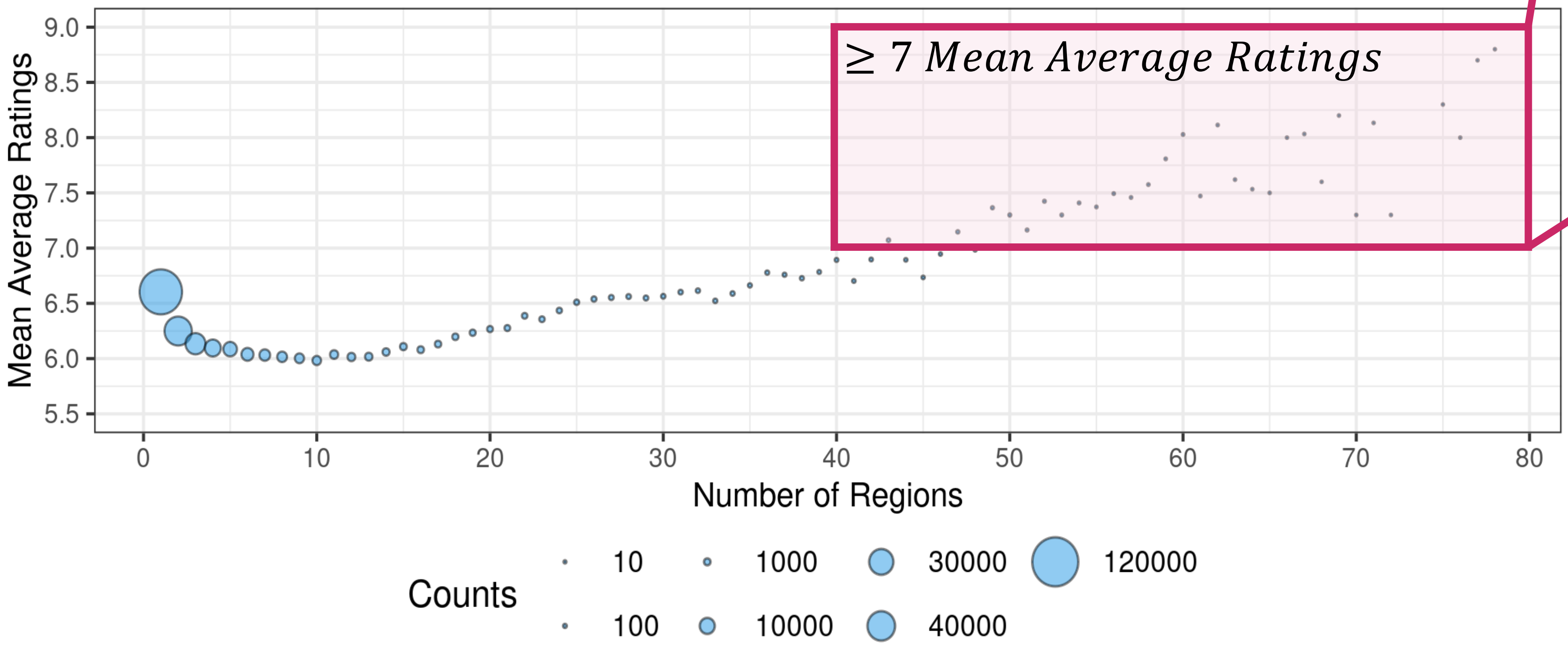
Number of Regions vs. Mean Average Ratings



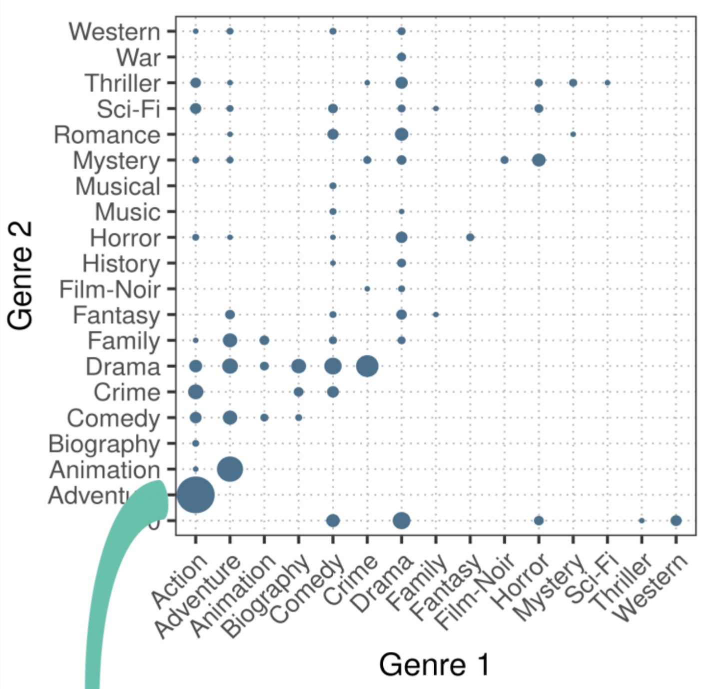


Titles per Region

Number of Regions vs. Mean Average Ratings



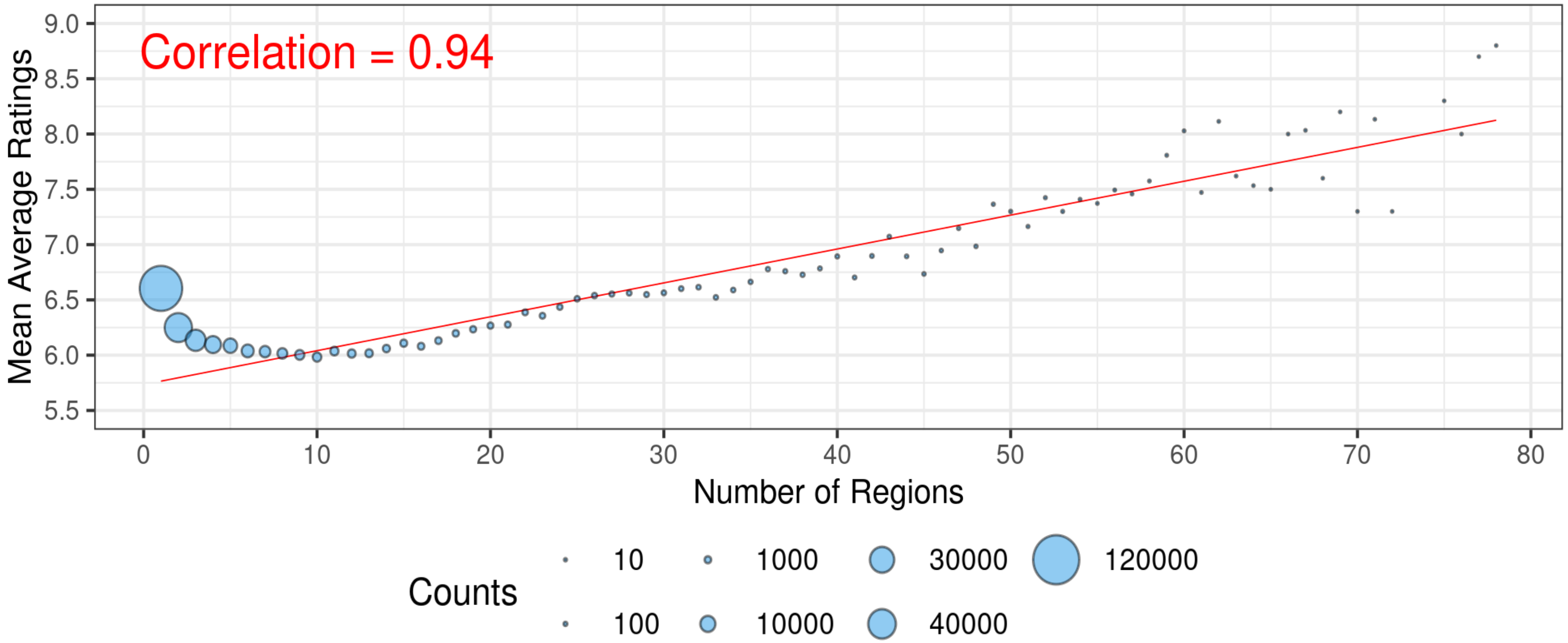
Movies with 2 genres



Action and Adventure are the most popular genres of these titles

Titles per Region

Number of Regions vs. Mean Average Ratings



The success of a movie is higher when translated to other languages

Proposed Hypotheses

Titles per Region

“Is the title’s success influenced by the number of translations?”

Titles’ Success

Cast and Crew Influence

Title’s Awards

“Does the cast/crew of a movie make it successful?”

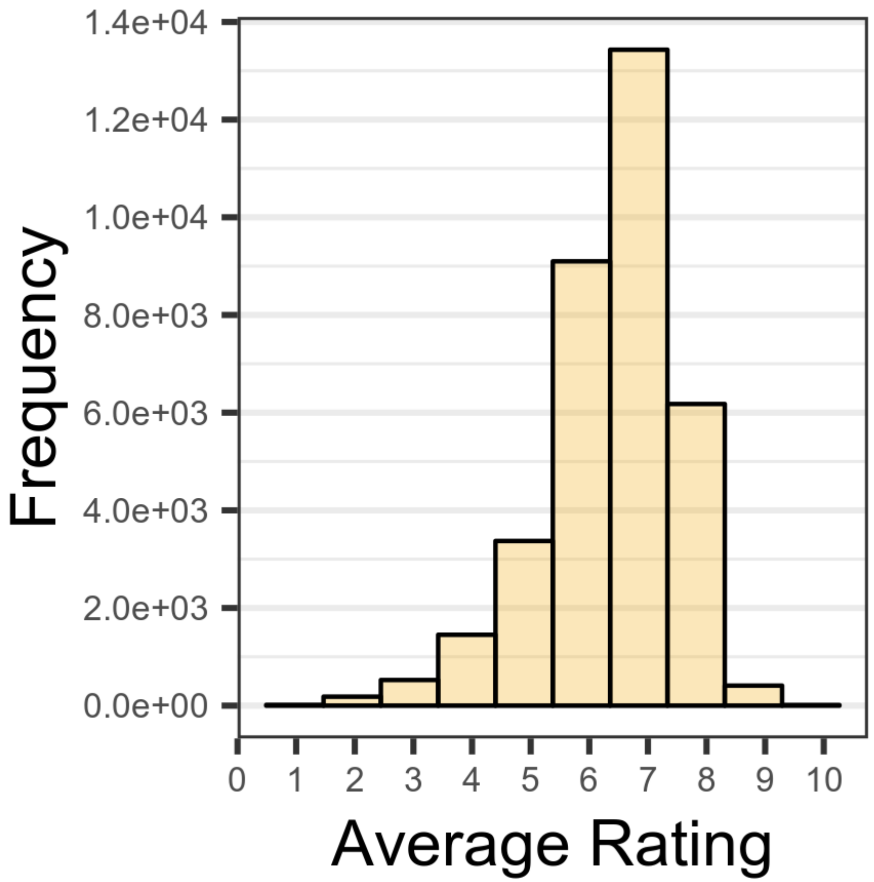
Titles' Success

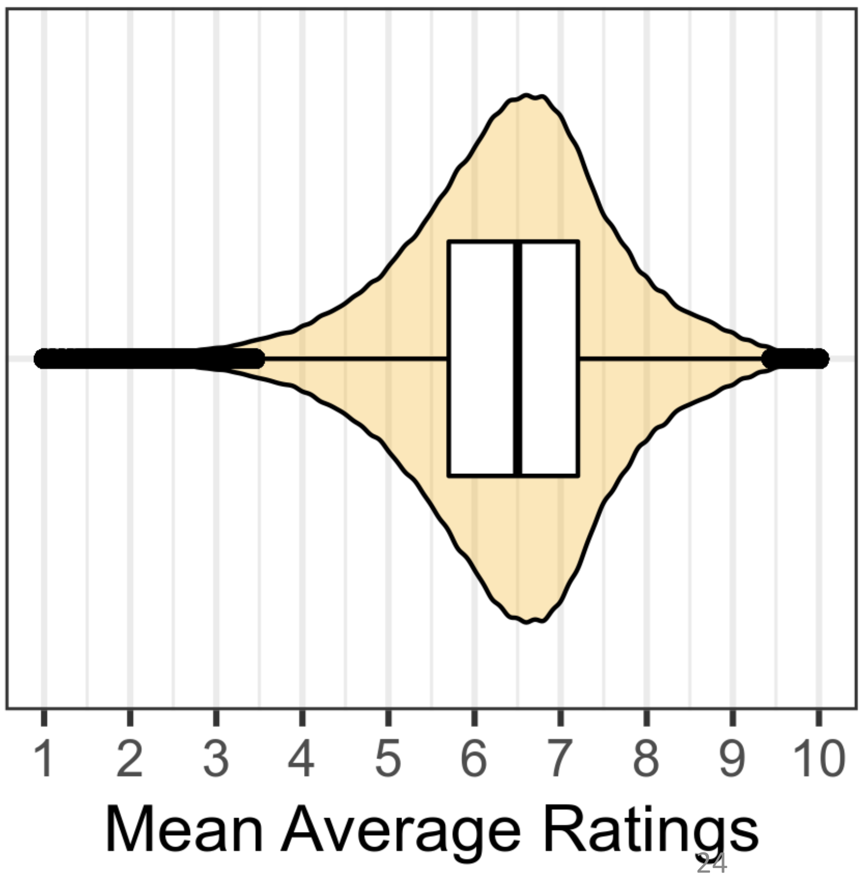
Cast and Crew Influence

From dataset: *Title Ratings*
↓
Average Rating Histogram

From dataset: *Title Ratings & Name Basics*
↓
Violin & Box Plot

Most ratings are considered *average*





Titles' Success

Cast and Crew Influence

Merged Basics Ratings Crew

$numVotes \geq 1000$
$averageRating \geq 7$ *or*
$averageRating \leq 3$

Merged Basics Ratings Crew Transformed

Titles’ Success

Cast and Crew Influence

Merged Basics Ratings Crew → Merged Basics Ratings Crew Transformed

numVotes $\geq$ *1000*
averageRating $\geq$ *7 or*
averageRating $\leq$ *3*

Name Basics → Name Basics filtered

Filter profession by

Actors/Actresses

Directors

Writers

People being known for a single movie goes from 1 to 3087

Titles' Success

Cast and Crew Influence

Merged Basics Ratings Crew

numVotes ≥ 1000
averageRating ≥ 7 or
averageRating ≤ 3

Merged Basics Ratings Crew Transformed

Name Basics

Filter profession by

Actors/Actresses

Directors

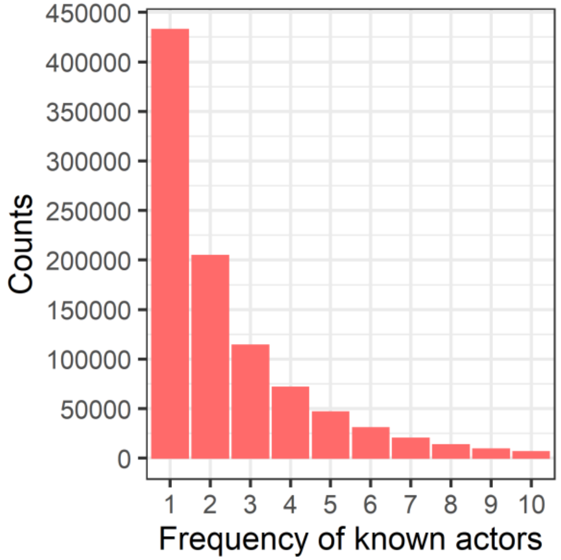
Writers

Name Basics filtered

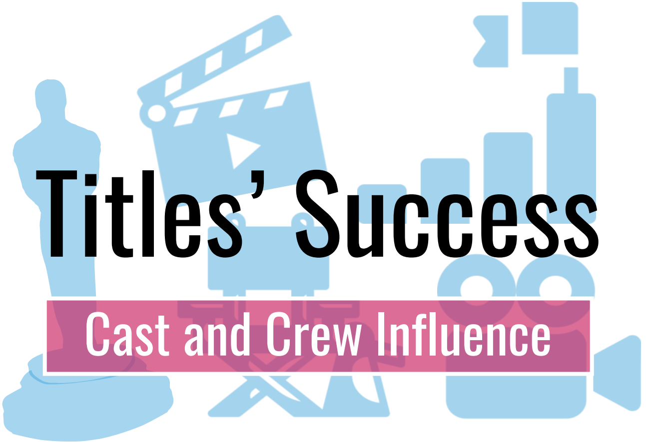
People being known for a single movie goes from 1 to 3087

Filter data by

Keep only the titles with a maximum of 10 people they're known for



Top 10 frequencies of the original one, that goes up to 3087 people known for a movie.



Titles' Success

Cast and Crew Influence

features of this graph

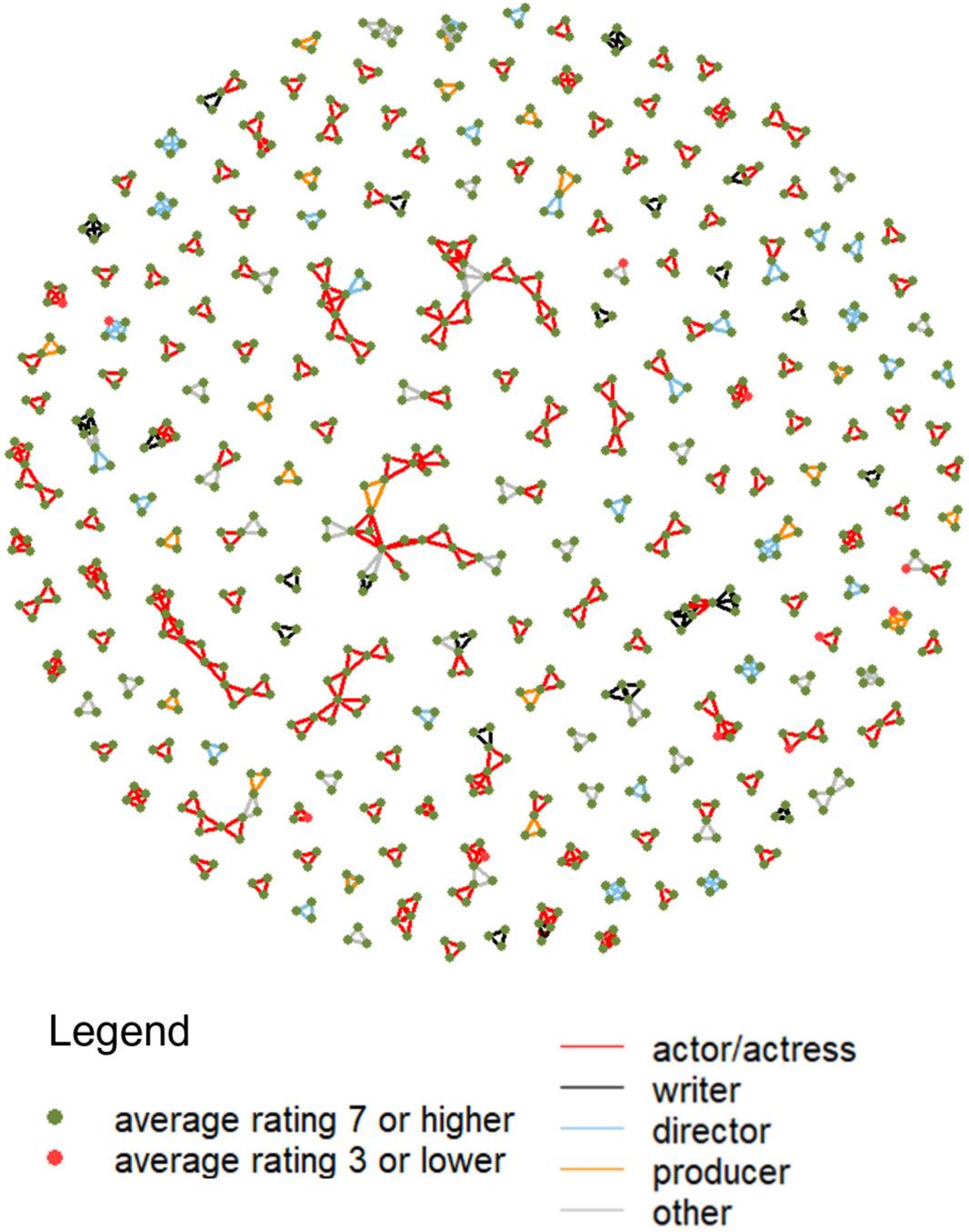
each node is a movie and **each link is a casting person** in common

red nodes represent unsuccessful titles

green nodes represent successful titles

links are colored based on the **primary profession** of each person

a person has to be known for 3 or 4 titles



Titles' Success

Cast and Crew Influence

features of this graph

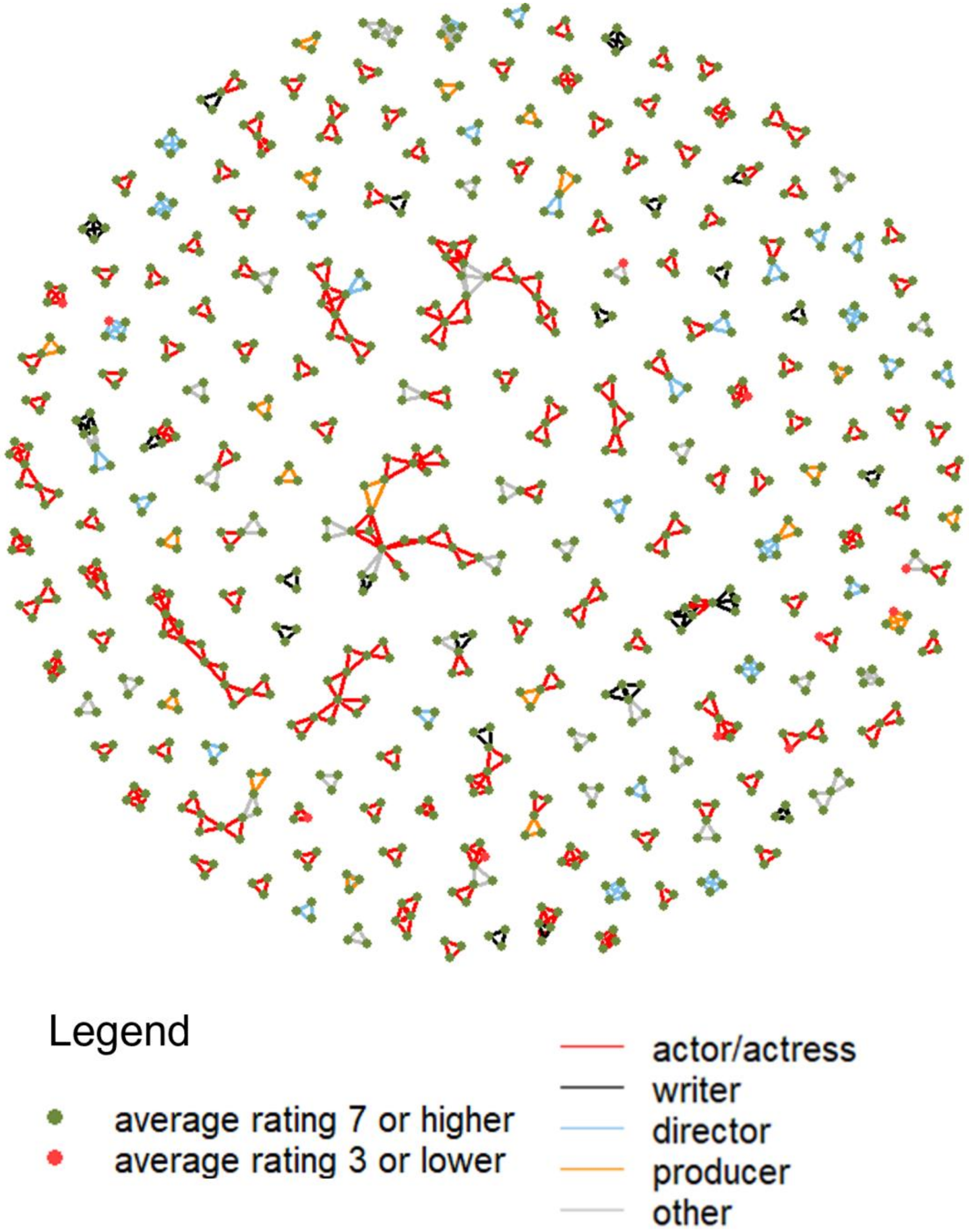
each node is a movie and **each link is a casting person** in common

red nodes represent unsuccessful titles

green nodes represent successful titles

links are colored based on the **primary profession** of each person

a person has to be known for 3 or 4 titles



only 11 unsuccessful titles against 753 successful ones

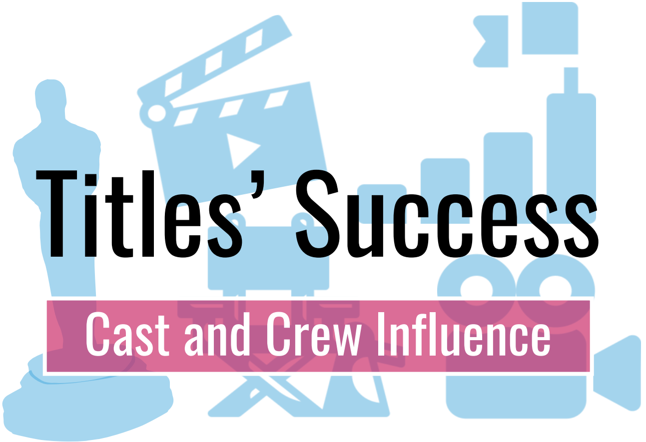
unsuccessful titles are present in small size chains

successful titles are present in all sizes of chains

successful movie tends to have more common people with other successful titles, than with unsuccessful ones

Titles' Success

Cast and Crew Influence



features of this graph

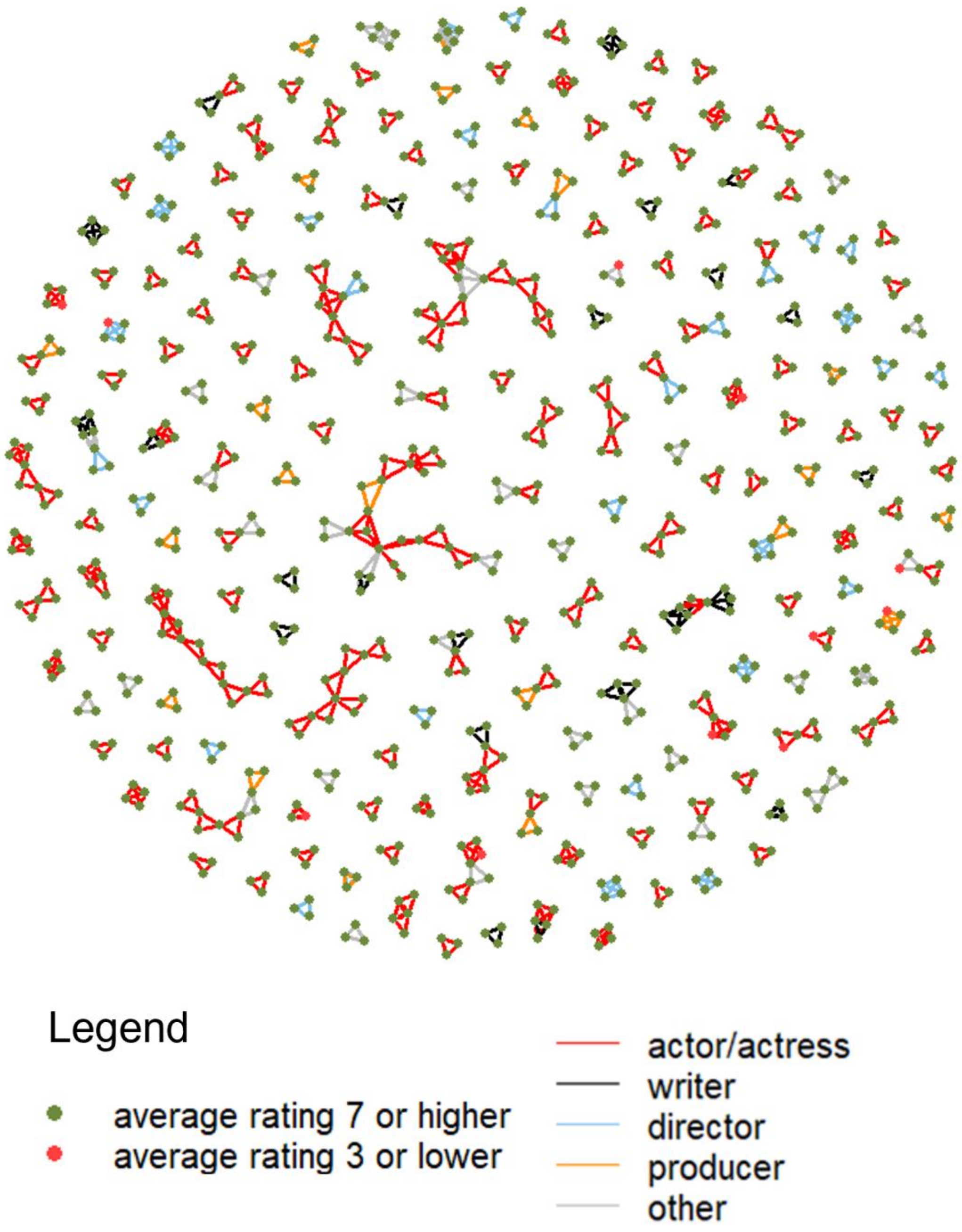
each node is a movie and **each link is a casting person** in common

red nodes represent unsuccessful titles

green nodes represent successful titles

links are colored based on the **primary profession** of each person

a person has to be known for 3 or 4 titles



clear majority of actors and actresses

bigger and medium chains are dominated by actors/actresses

most relevant profession in these successful titles is the actor/actress

Titles' Success

Cast and Crew Influence

Title Crew

Title Ratings

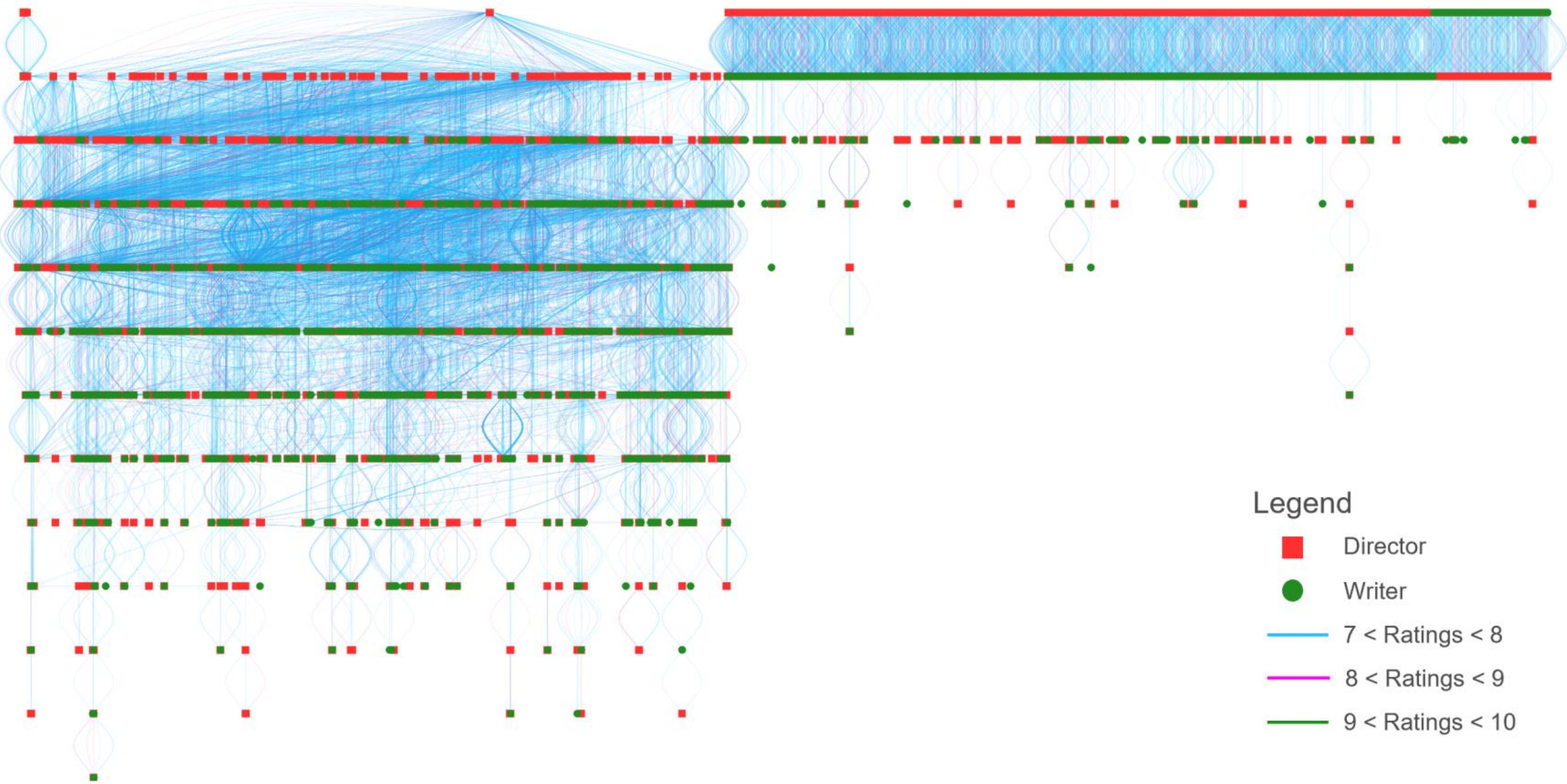
numVotes $\geq$ *1000*
averageRating $\geq$ *7*

Filtered Merged Title Crew and Ratings

each node is a director or writer

each link is a title separated by rating in color

not display isolates and only display connected elements



Title Crew

Title Ratings

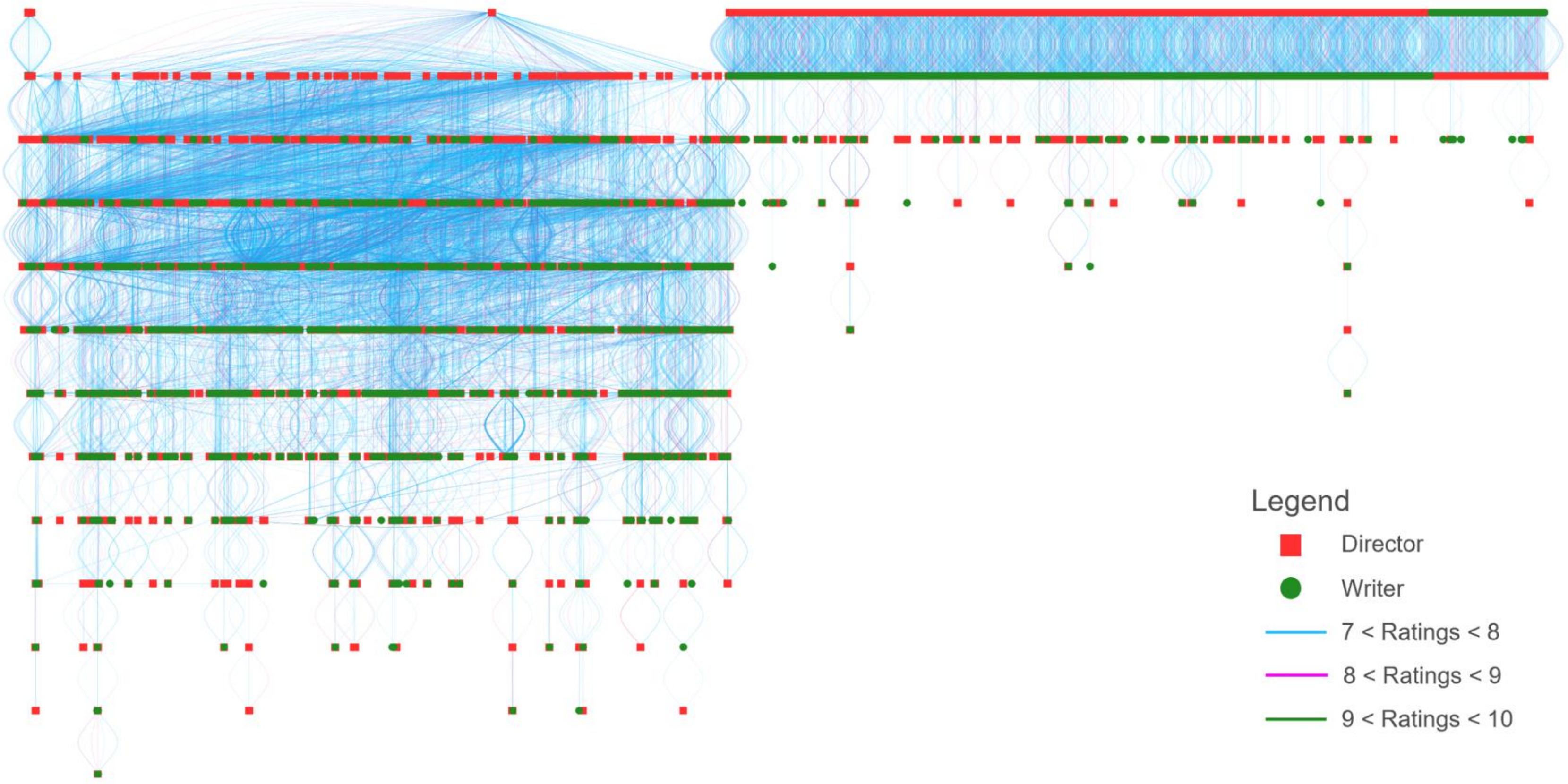
numVotes ≥ *1000*
averageRating ≥ *7*

Filtered Merged Title Crew and Ratings

each node is a director or writer

each link is a title separated by rating in color

not display isolates and only display connected elements



clear division amongst the data

Titles' Success

Cast and Crew Influence

Title Crew

Title Ratings

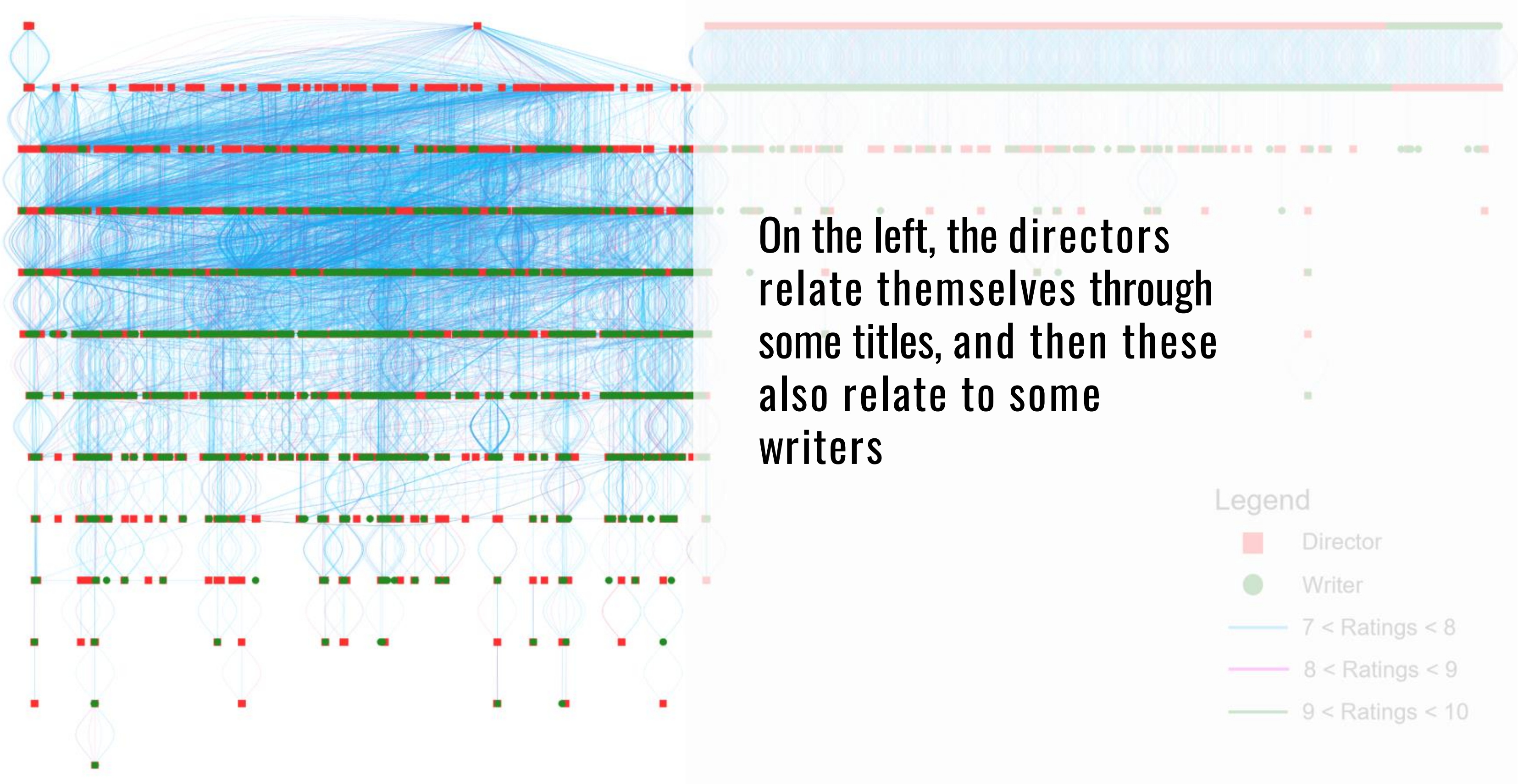
numVotes ≥ 1000
averageRating ≥ 7

Filtered Merged Title Crew and Ratings

each node is a director or writer

each link is a title separated by rating in color

not display isolates and only display connected elements



On the left, the directors relate themselves through some titles, and then these also relate to some writers

clear division amongst the data

Titles' Success

Cast and Crew Influence

Title Crew

Title Ratings

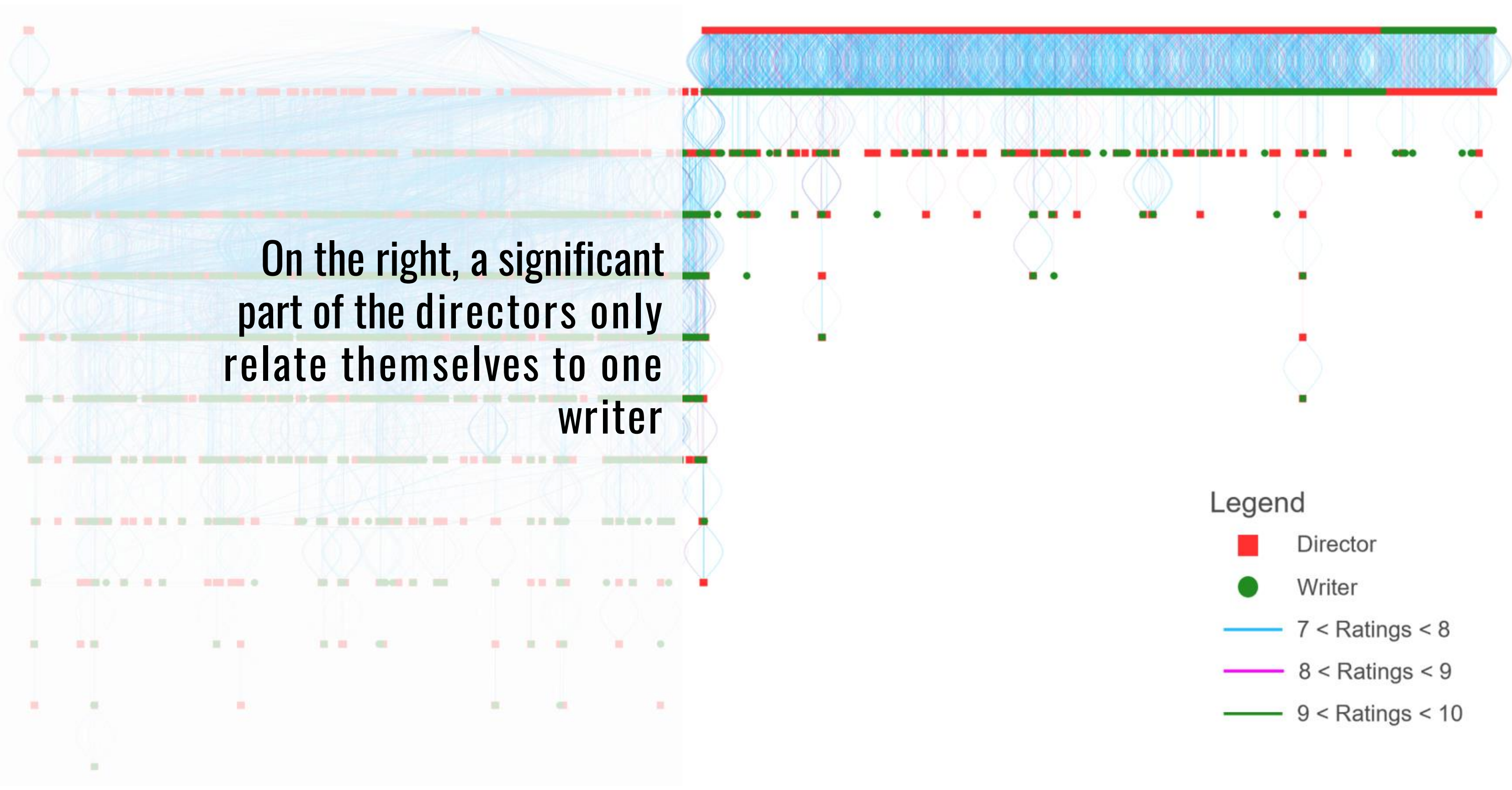
numVotes ≥ 1000
averageRating ≥ 7

Filtered Merged Title Crew and Ratings

each node is a director or writer

each link is a title separated by rating in color

not display isolates and only display connected elements



On the right, a significant part of the directors only relate themselves to one writer

clear division amongst the data

Titles' Success

Cast and Crew Influence

Title Crew

Title Ratings

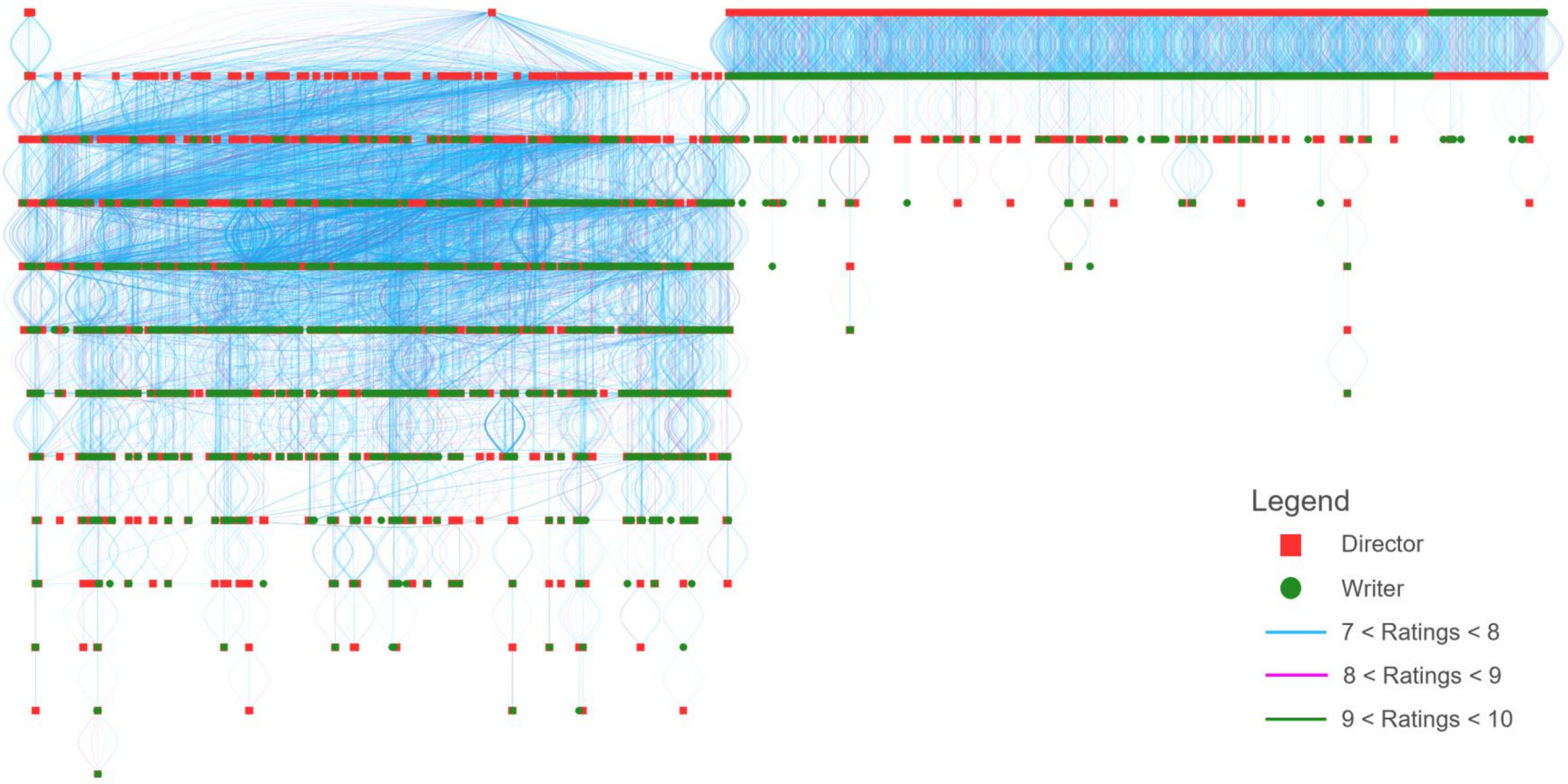
numVotes ≥ *1000*
averageRating ≥ *7*

Filtered Merged Title Crew and Ratings

each node is a director or writer

each link is a title separated by rating in color

not display isolates and only display connected elements



clear division amongst the data

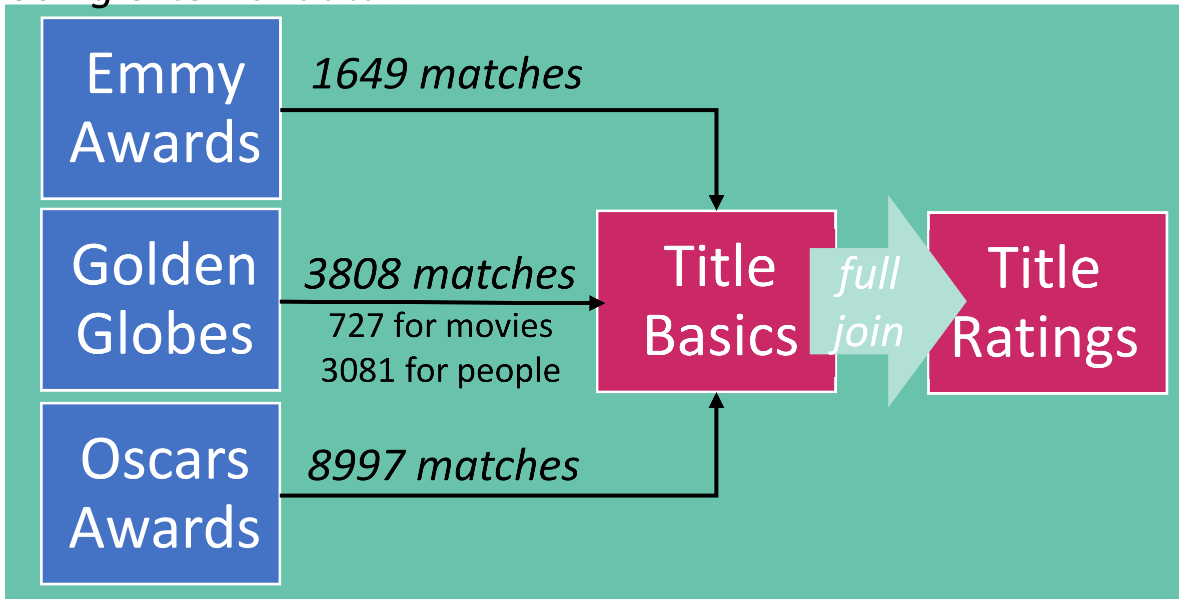
no specific kin between the number of members of a crew and the title's average rating above 7

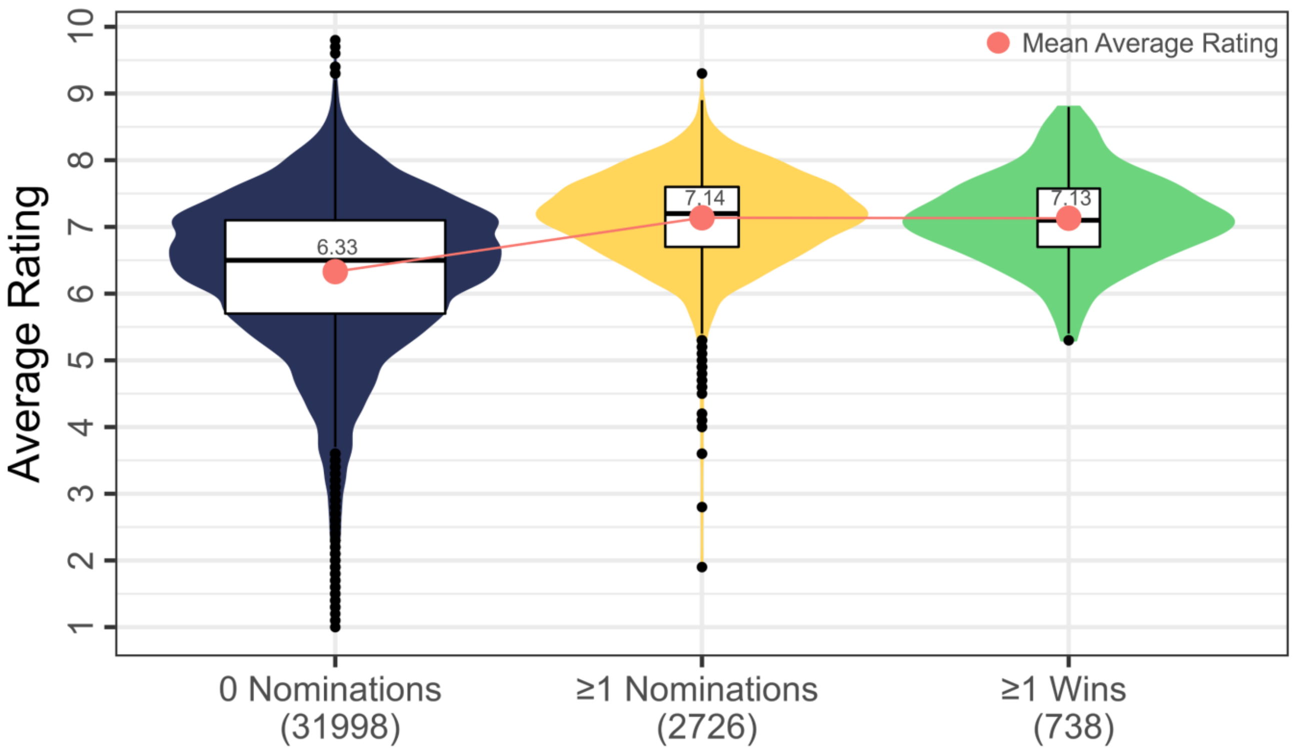
one specific director is heavily related toother directors, so making a movie with this director may lead to success

Titles' Success

Awards

Using external data...

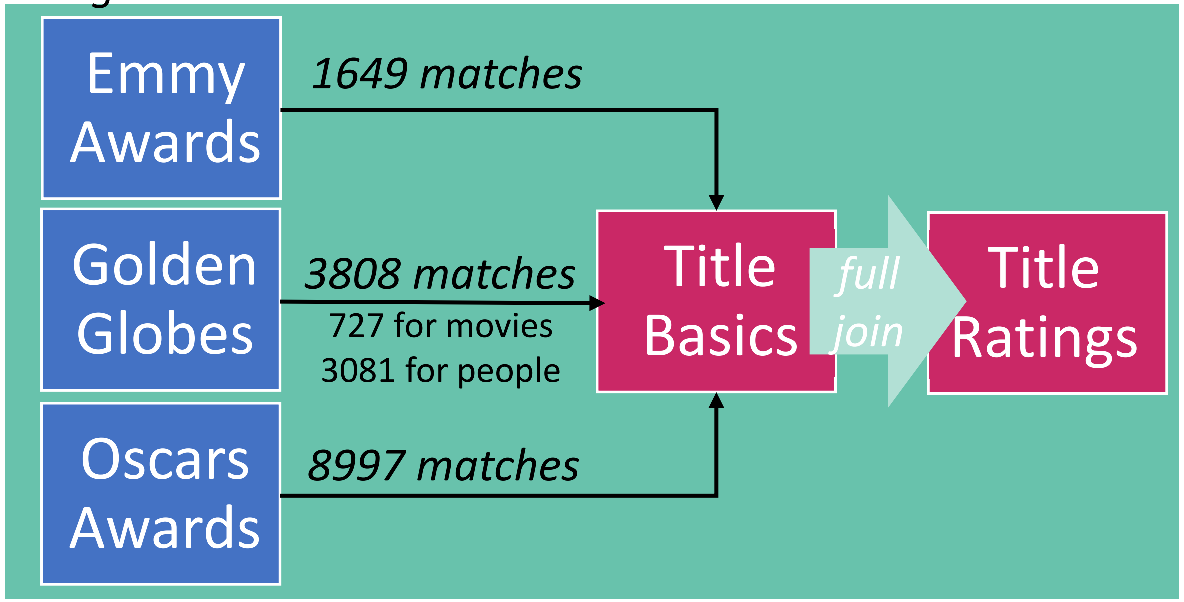


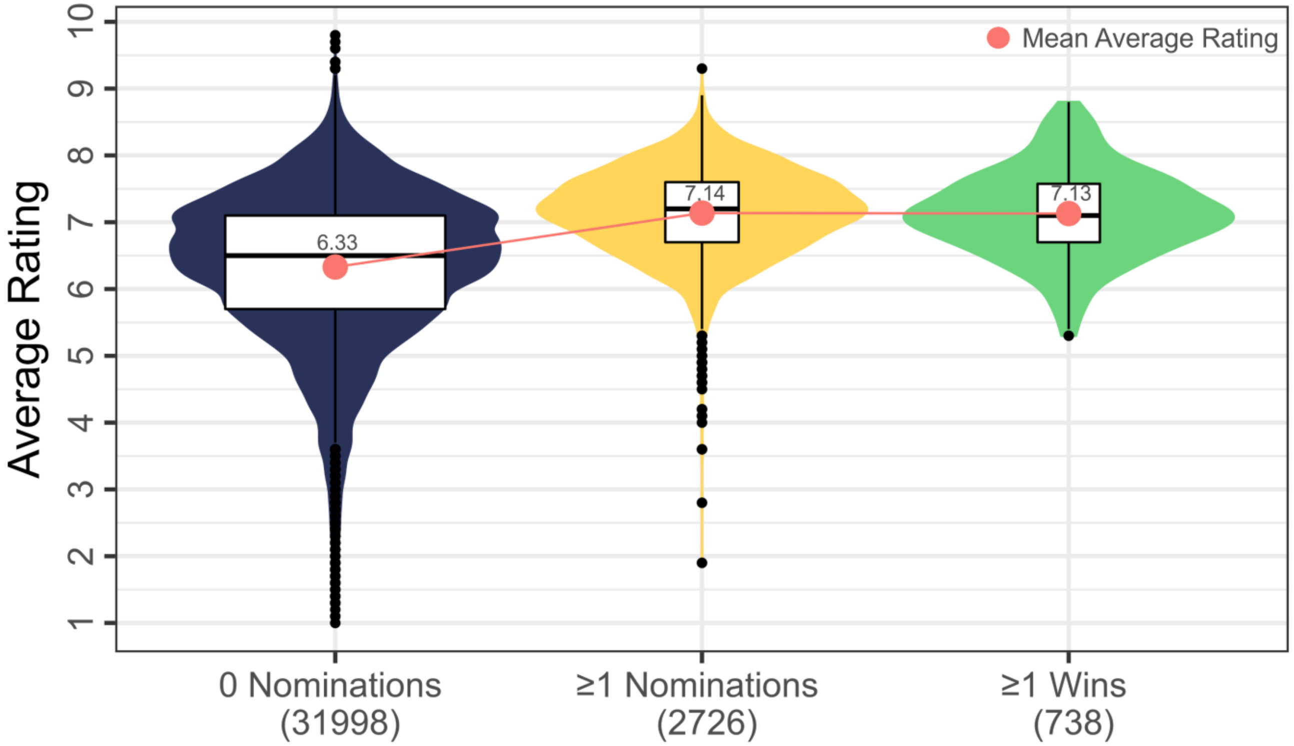


Titles' Success

Title's Awards

Using external data...





titles with no nominations have a distribution with a wider range than those which had nominations

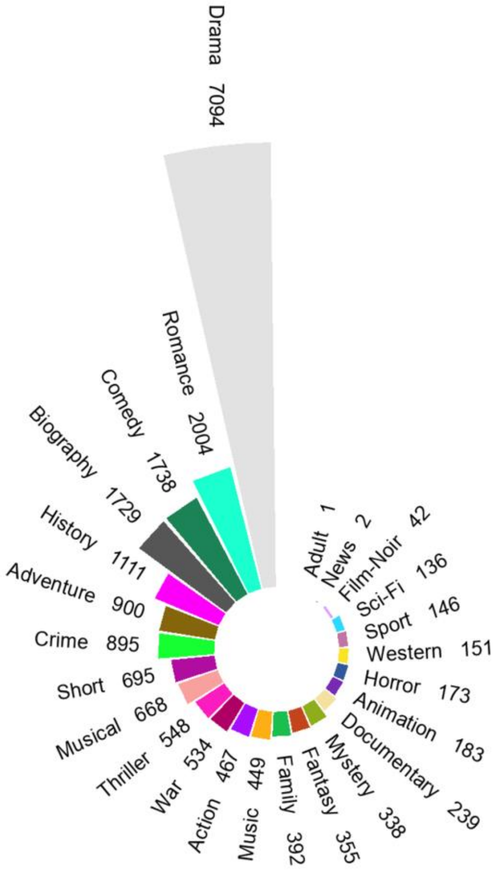
most titles with nominations and wins have their IMDb average ratings above 5

distribution of titles with nominations and with winnings is very similar

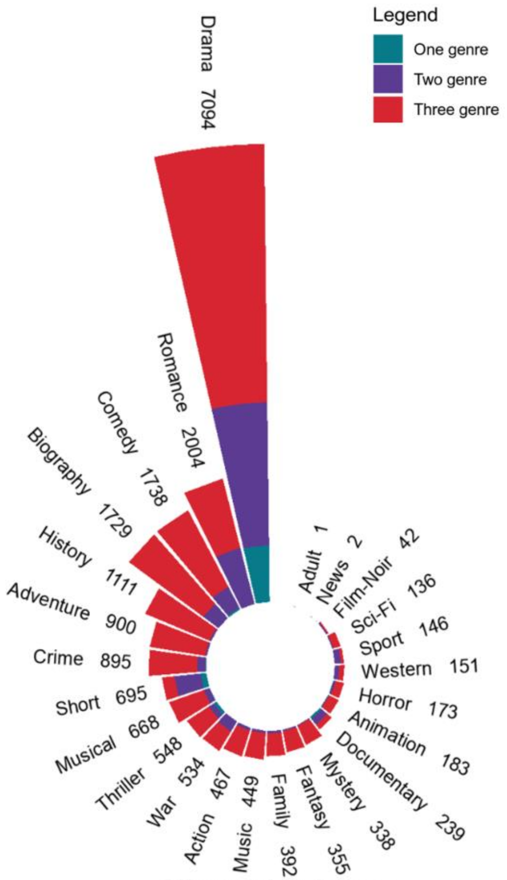
Titles' Success

Awards

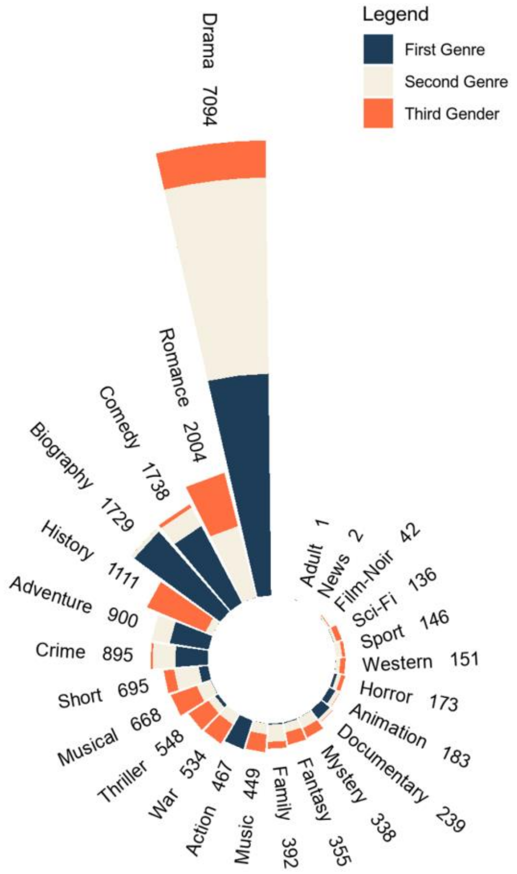
Considering only titles that actually won an award



All awards winners
All genres combined



All awards winners
Division by number of genres



All awards winners
Division by position of genre

Titles’ Success

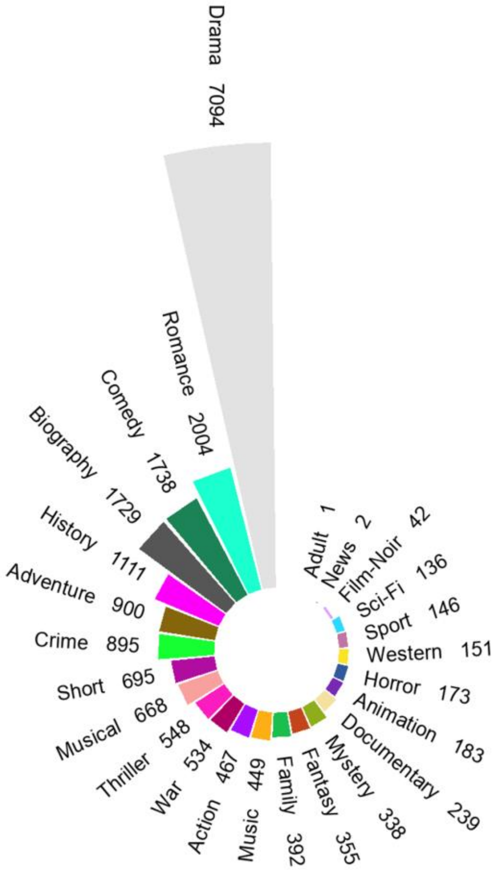
Awards

Considering only titles that actually won an award

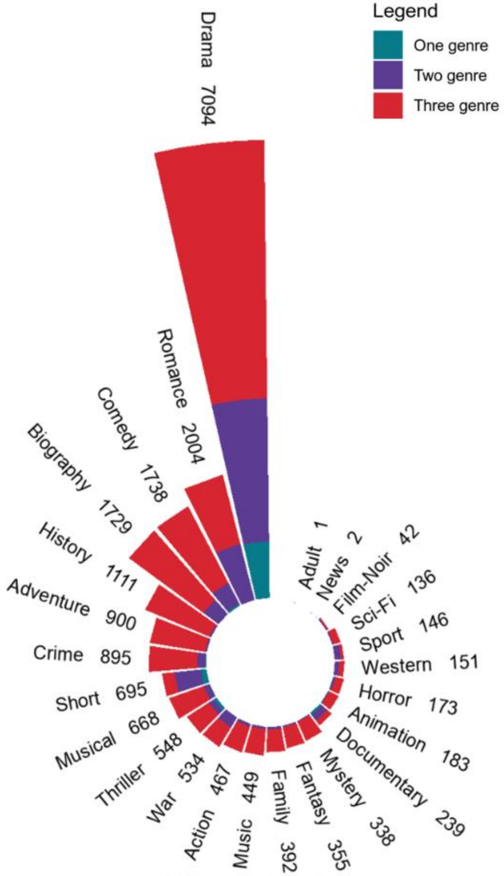
drama is the most popular genre for titles that won any type of award

most of the biography titles have this genre as their first genre

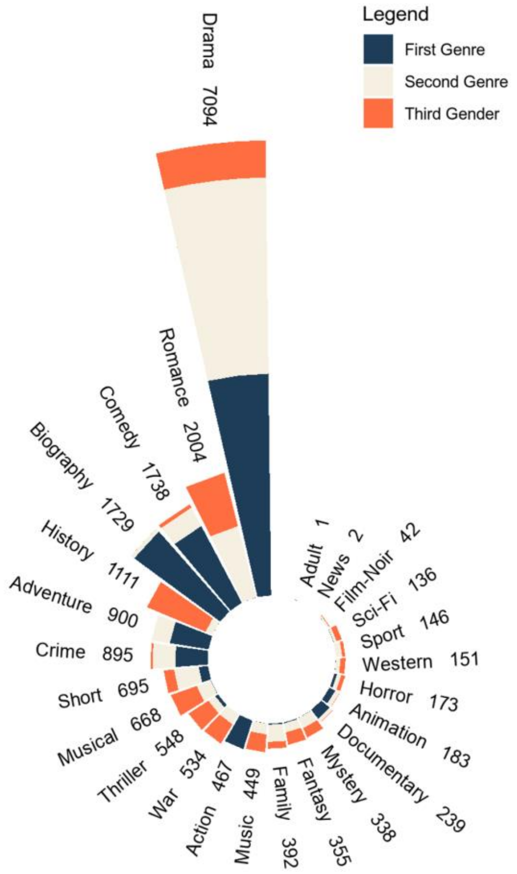
position of the genre is important for defining the type of the movie.



All awards winners
All genres combined



All awards winners
Division by number of genres

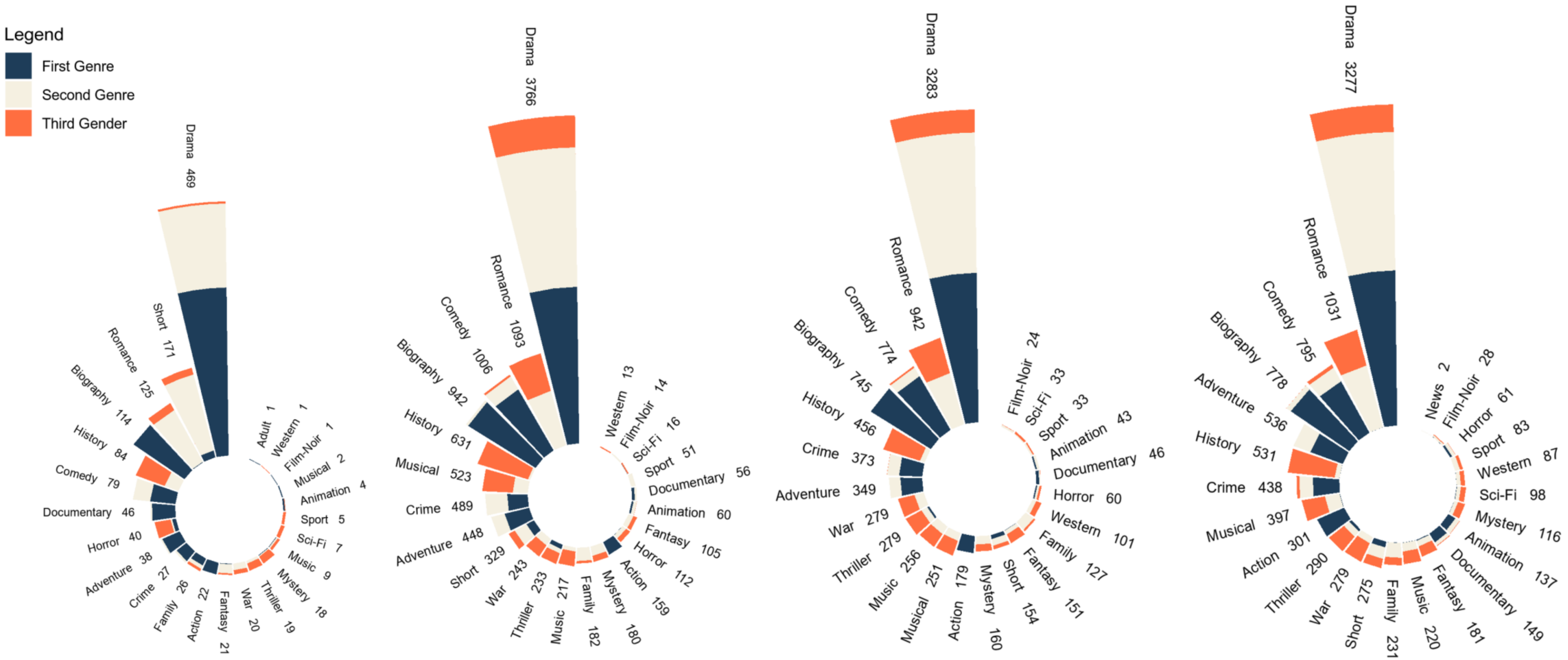


All awards winners
Division by position of genre

Titles' Success

Awards

Considering only titles that actually won an award



Emmy winners

Golden Globe winners (for movies)

Golden Globe winners (for people)

Oscar winners

not considering drama, a movie is more likely to get a Golden Globe for the people if it is a romance and for the movie, if it is a comedy

Titles' Success

Awards

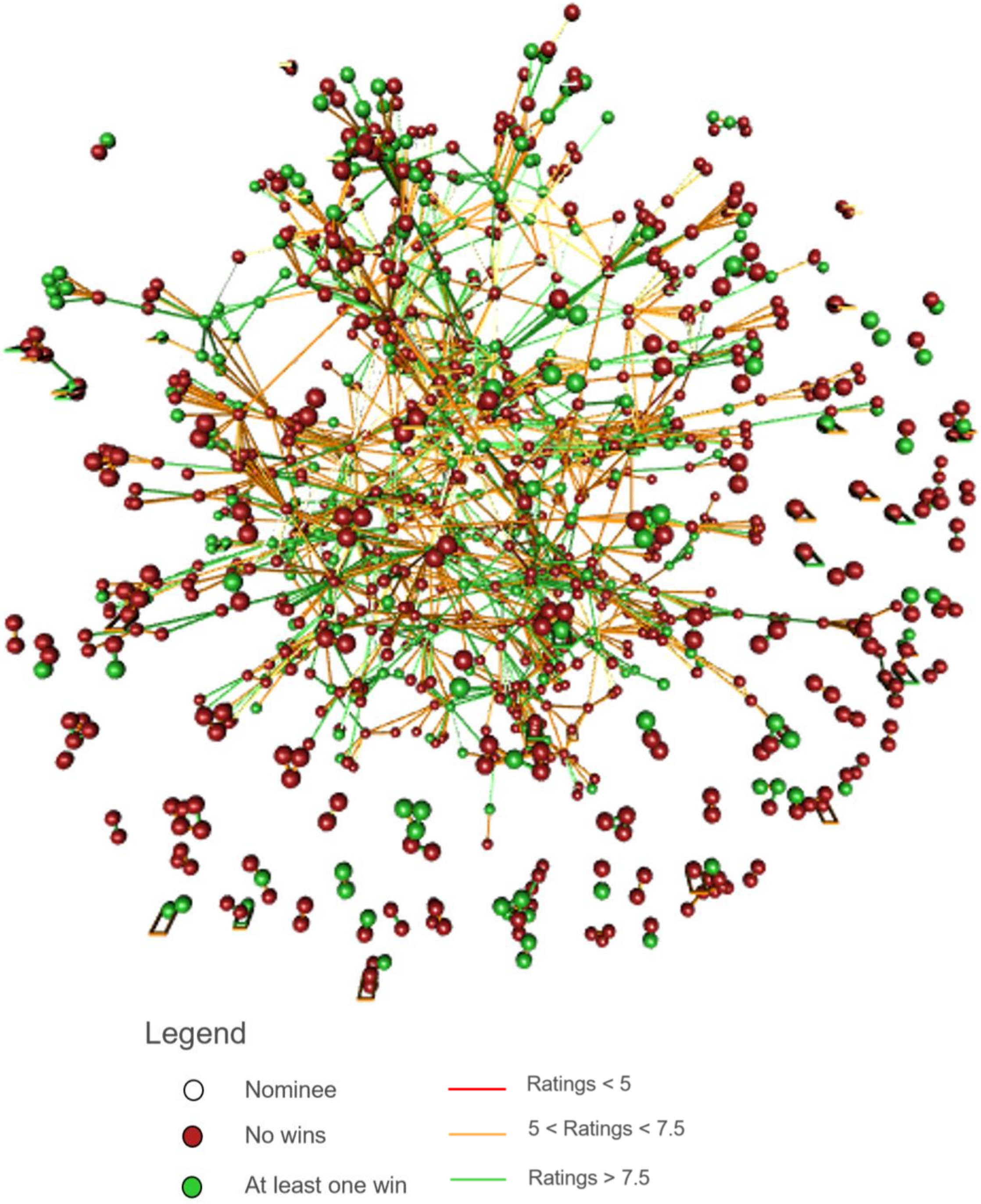
features of this graph

each node is a nominee, and **each link is a movie** in common

red nodes represent nominated people for the **Golden Globes** without wins

green nodes represent nominated people for the **Golden Globes** that won at least 1 award

links are colored based on the **ratings** of each title



Titles' Success

Awards

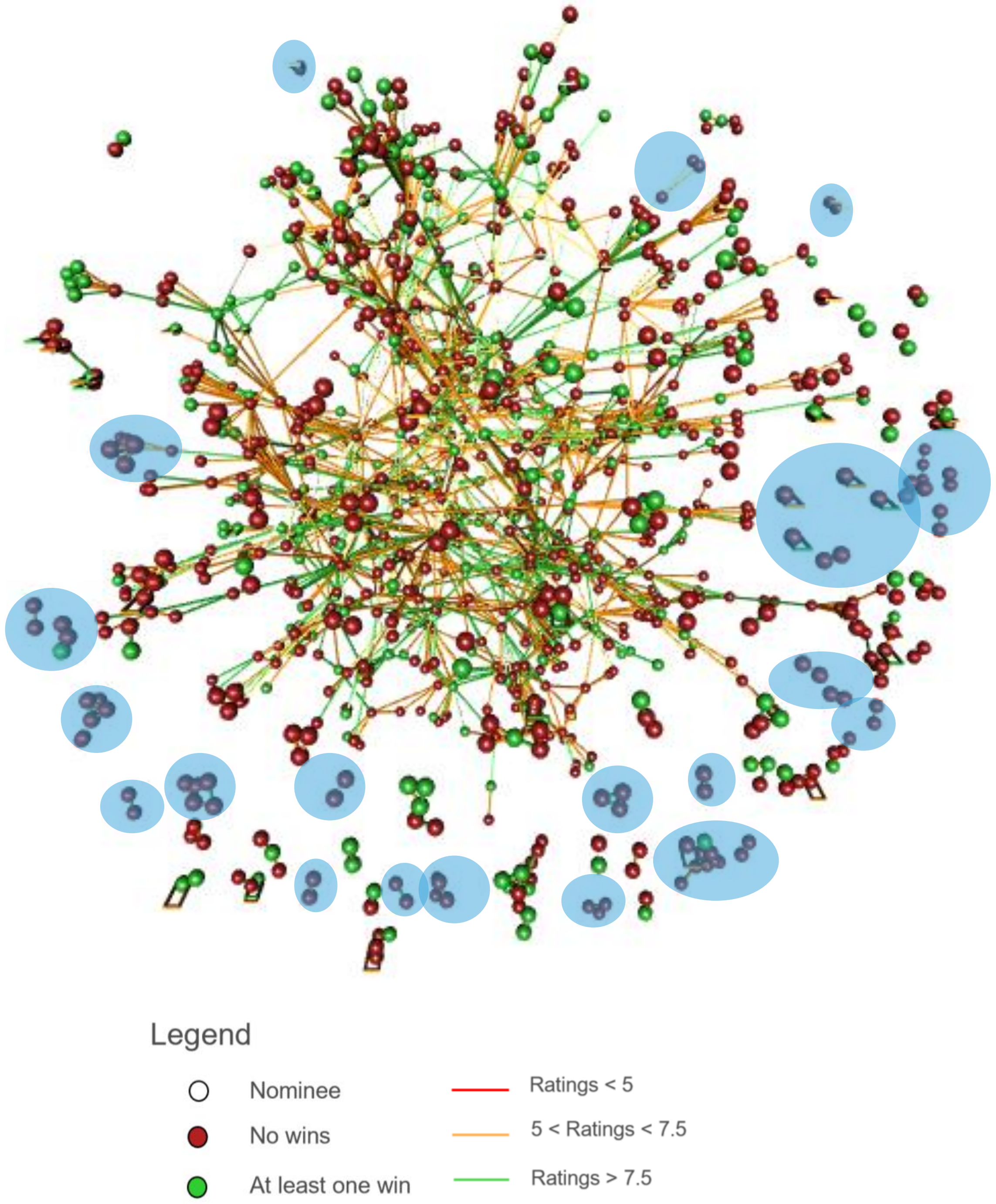
features of this graph

each node is a nominee, and **each link is a movie** in common

red nodes represent nominated people for the **Golden Globes** without wins

green nodes represent nominated people for the **Golden Globes** that won at least 1 award

links are colored based on the **ratings** of each title



nominees that did not win the award are more likely to be connected to other nominees that also did not win the award, regardless of the movie ratings

no relevant connection to the rating and the winning, all nominees have medium-high ratings

Titles' Success

Awards

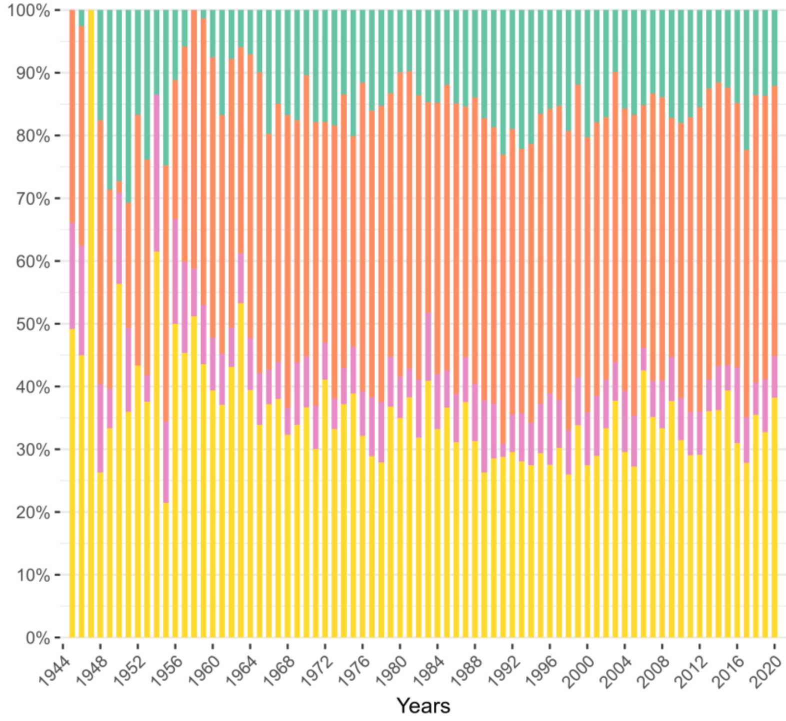
Categories of the awards

Corresponding data

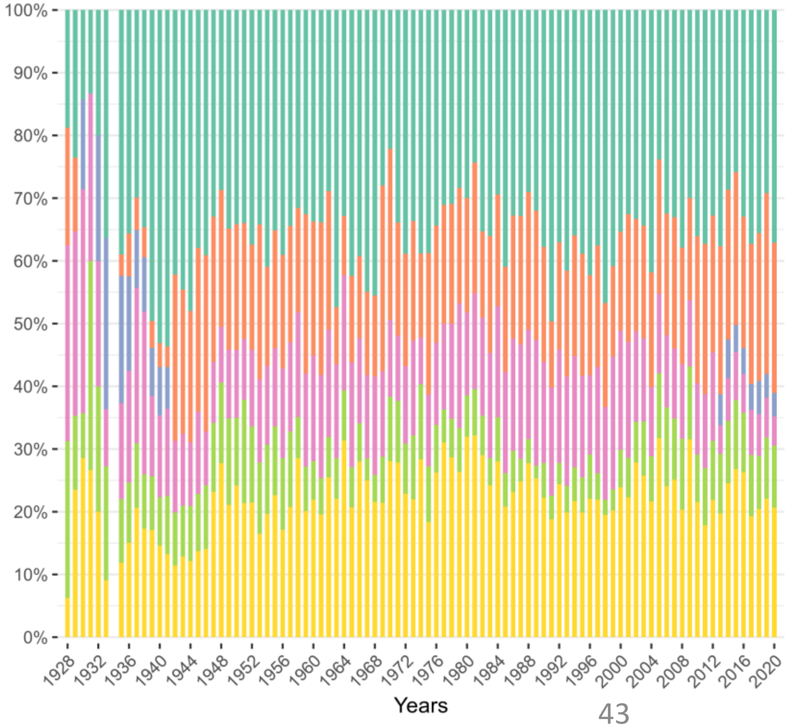
Others | Producer | Writer | Film | Director | Actor

Emmys Awards

Golden Globes



Oscars Awards



Titles' Success

Awards

Categories of the awards

Corresponding data

Others | Producer | Writer | Film | Director | Actor

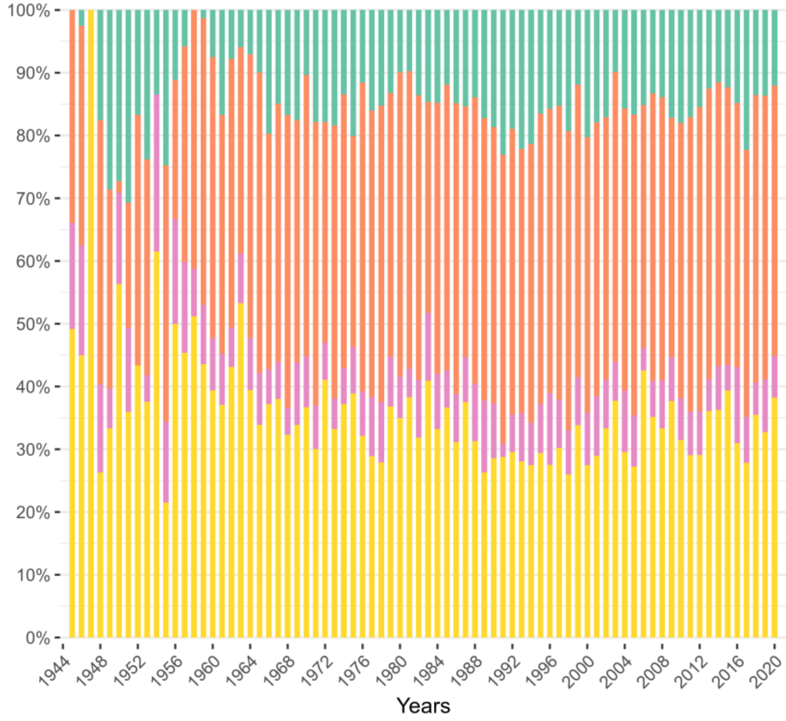
very few producer dedicated awards

Golden Globes dedicates most awards to actors and film

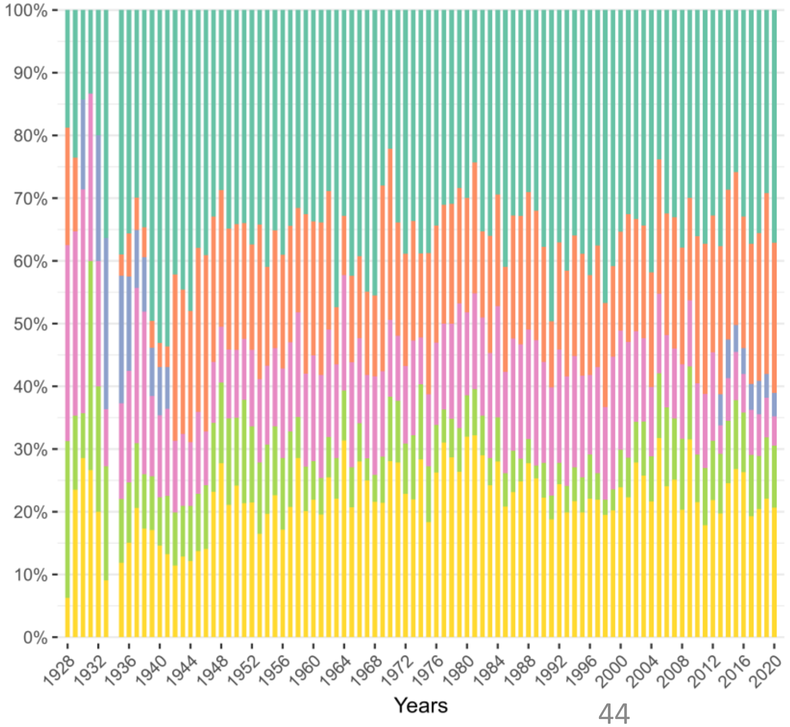
Oscars have more well distributed categories than others

Emmys Awards

Golden Globes



Oscars Awards



Main Conclusions

US is the country to which most titles are translated.

Titles that are translated to more than one region are typically translated to a European country

As the number of regions increases, the average ratings get higher and the frequency of translated titles diminishes.

Actors are the most relevant people for characterizing the title's success, that is, the average ratings.

Despite this, there may be some specific members of the crew, more exactly directors, that can be related to a title's success.

Titles that win any type of the awards explored have higher ratings.

Any questions?

IMDb

Datasets Analysis