

DSCP Project

Letterboxd Analysis

Jonathan Dietrich, Matthew Giadla, Kaelin Silas, Julie Walsh, Sijia Wu

Introduction



Our group investigated the Letterboxd dataset provided by Kaggle. Letterboxd is an app with over 17 million users where users can rate and favorite movies. This dataset (24.71GB) provided information such as studio, actors, release date, rating, and more for over 950,000 movies.

Throughout this project we sought to answer the following question: **What defines a successful movie studio?**

actors.csv: id, name, role

countries.csv: id, country

crew.csv: id, role, name

genres.csv: id, genre

languages.csv: id, type, language

movies.csv: id, name, date, tagline

posters.csv: id, link

releases.csv: id, country, date, type, rating

studios.csv: id, studio

themes.csv: id, theme

Introduction - What defines a “successful studio”?

We investigated this question in two different ways:

Cumulative Methods Approach:

- Combining data across various variables in many csvs: movies, studios, crew, cast, release dates, rating, etc
- Built a success score based on multiple factors: longevity, production volume, global reach, and average movie ratings (ratings weighted at 25%).
- Analyzed similarities across these top 10 studios

User Ratings approach:

- Focused only on overall user rating totals to reflect both productivity and audience reception.
- Tracked how studio's yearly average ratings
- Compared trends across studios to identify differences in consistency, growth, and audience approval over time.

Most successful studios? Method 1 - Cumulative

Statistical Methods (Ran with HTCondor)

Statistical Method Step 1: Data Aggregations: means, counts, totals

Join (by studio)

- avg_rating: average rating across movies
- studio_movies_per_year: total movies, years active (first year - last year), average movies per year
- studio_country_reach: average number of countries released per movie
- actors_per_movie: avg # actors per movie by studio (measure cast size)
- crew_info: total # unique crew roles, total # crew members (measures production depth per movie)

Statistical Step 2: Normalization (rescaling) to standardize metrics

Rescale each metric (0-1)

Compute success_score using weighted sum:

- 25% rating
- 10% movies per year
- 10% total movies
- 15% years active (including re-releases)
- 15% country reach
- 10% crew roles
- 15% crew total

Statistical Method Step 3: Success Score as a weighted sum

Success score is computed as a weighted sum of normalized metrics, which is a form of linear combination:

$$\sum_i w_i \cdot x_i$$

each w_i are the weights (25% for rating, and x_i are the normalized metrics

Number of Jobs, Memory, Disk Space

1. Four jobs with 1 CPU, 4 GB Memory, 2GB Disk Space

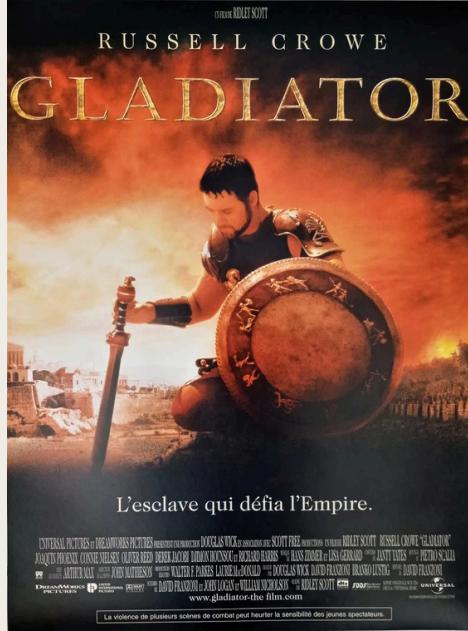
$$x_{\text{rescaled}} = \frac{x - \min(x)}{\max(x) - \min(x)}$$

Sort descending by success

Method 1 results (+outside research for validity)

Studio

1. Mill Film
2. Casino Royale Productions
3. Orion-Nova Productions
4. Warner Bros. Pictures
5. 16:14 Entertainment
6. Torridon Films
7. Société Westi
8. P of A Productions Limited
9. Patron Inc.
10. 8:38 Productions



(1) Movie: Gladiator (2000)
Winner: Best Visual Effects Oscar 2001



(2) Movie: Casino Royale (2006)
ft Daniel Craig as James Bond



(3) Movie: 12 Angry Men (1957)
3 Oscars wins

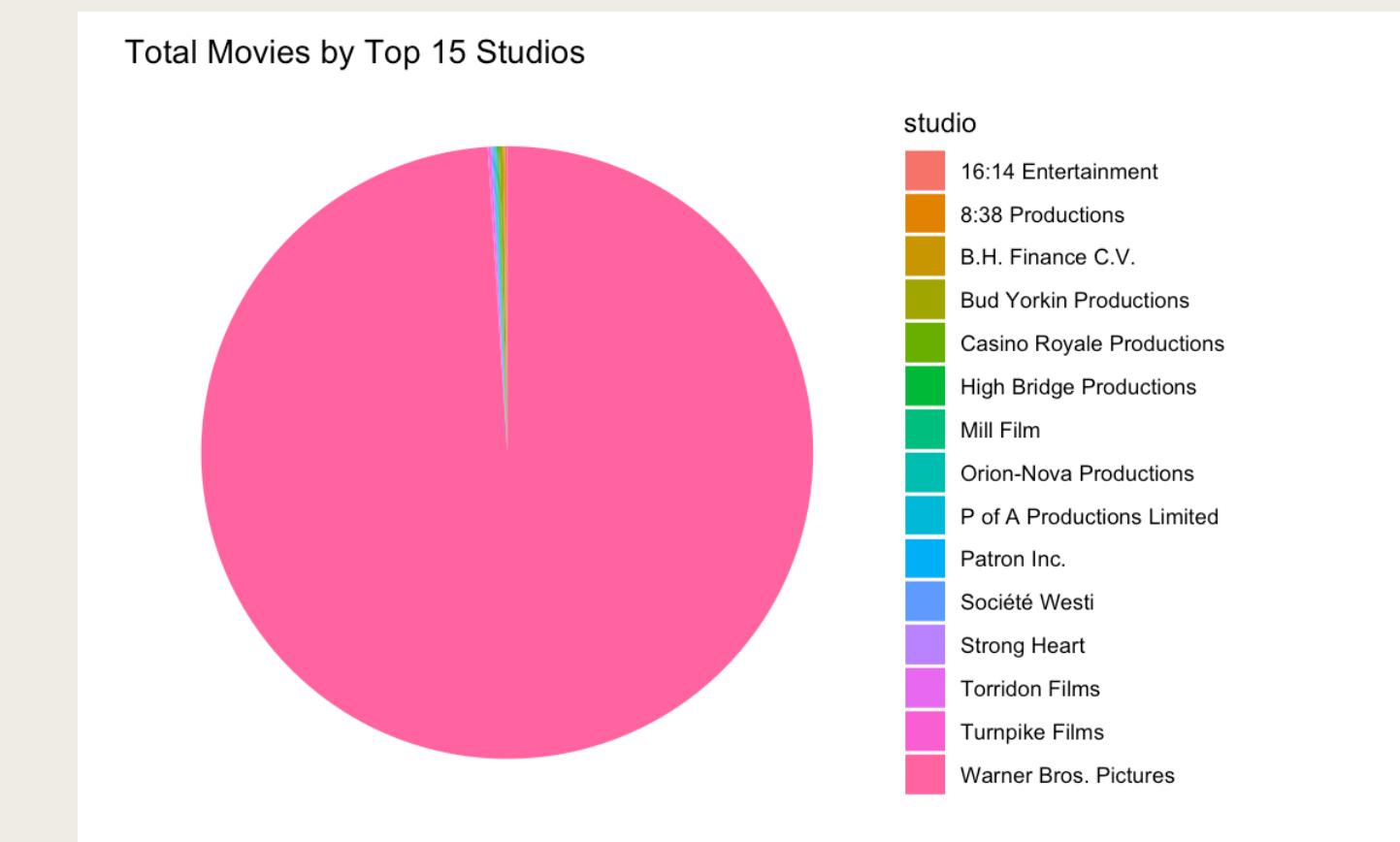


(10) 8:38 Productions -
Prisoners with
Hugh Jackman, Jake Gyllenhall

studio <chr>	success_score <dbl>	avg_rating <dbl>	total_movies <int>	years_active <dbl>	avg_movies_per_year <dbl>	avg_countries_per_movie <dbl>	movies <chr>
Mill Film	0.4273659	4.110000	1	22	0.04545455	57.000000	Gladiator
Casino Royale Productions	0.3963809	4.010000	1	15	0.06666667	76.000000	Casino Royale
Orion–Nova Productions	0.3922749	4.620000	1	67	0.01492537	33.000000	12 Angry Men
Warner Bros. Pictures	0.3870366	3.211314	1745	101	17.27722772	9.414327	Barbie, Joker, The Batman, The
16:14 Entertainment	0.3865346	3.837101	2	8	0.25000000	55.500000	Blade Runner 2049, 12 Strong
Torridon Films	0.3865346	3.837101	2	8	0.25000000	55.500000	Blade Runner 2049, 12 Strong
Société Westi	0.3858257	4.230000	1	90	0.01111111	2.000000	Napoleon
P of A Productions Limited	0.3761549	4.090000	1	20	0.05000000	58.000000	Harry Potter and the Prisoner o
Patron Inc.	0.3721815	4.380000	1	68	0.01470588	41.000000	Rear Window
8:38 Productions	0.3702959	4.290000	1	3	0.33333333	59.000000	Prisoners

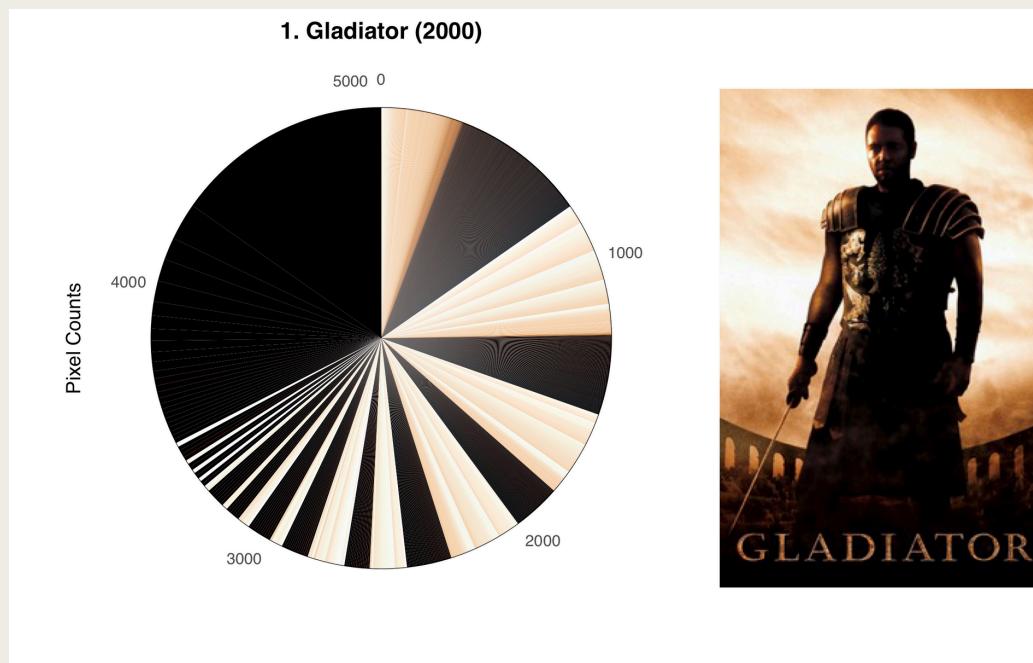
Results

1. Types of movies found
2. Small studio domination
3. Visualization - Blockbusters/critically acclaimed movies have switched to darker posters

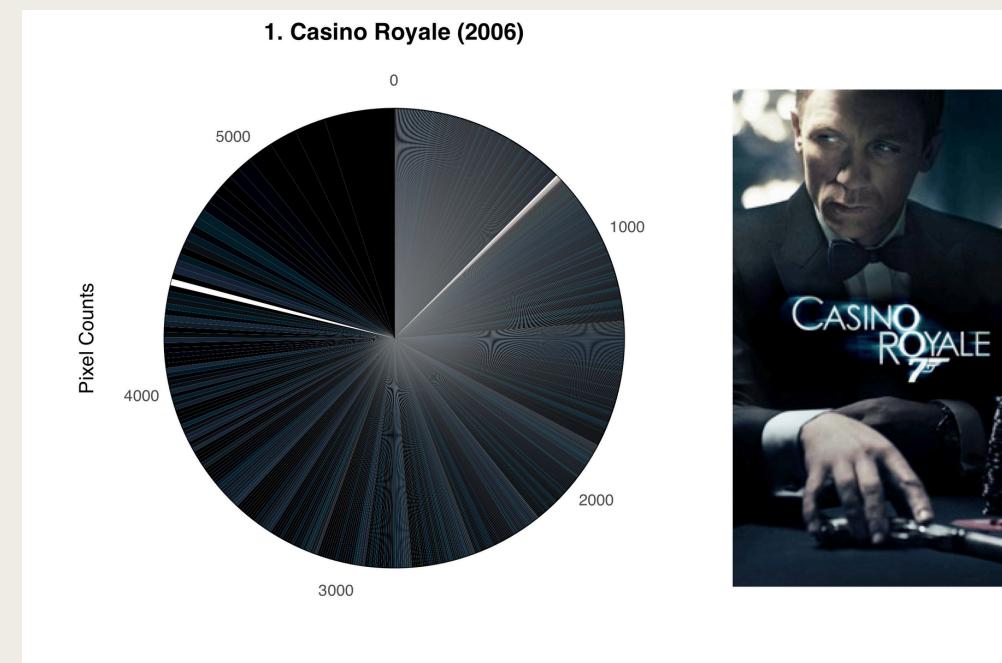


Top Studios (Cumulative) Movie Posters

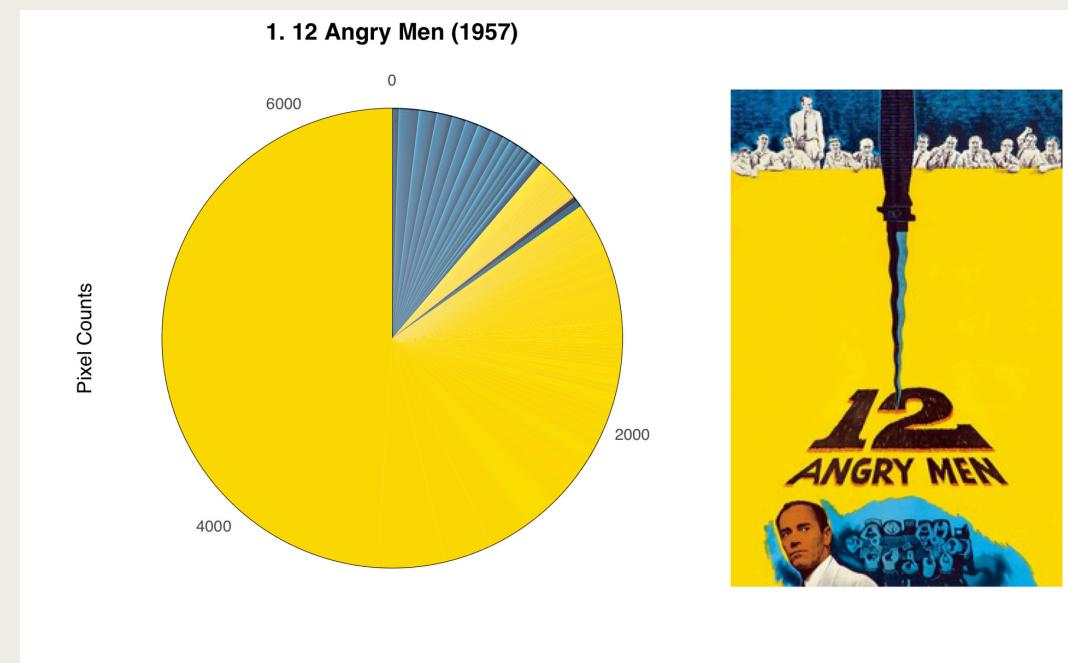
Mill Film



Casino Royale Productions



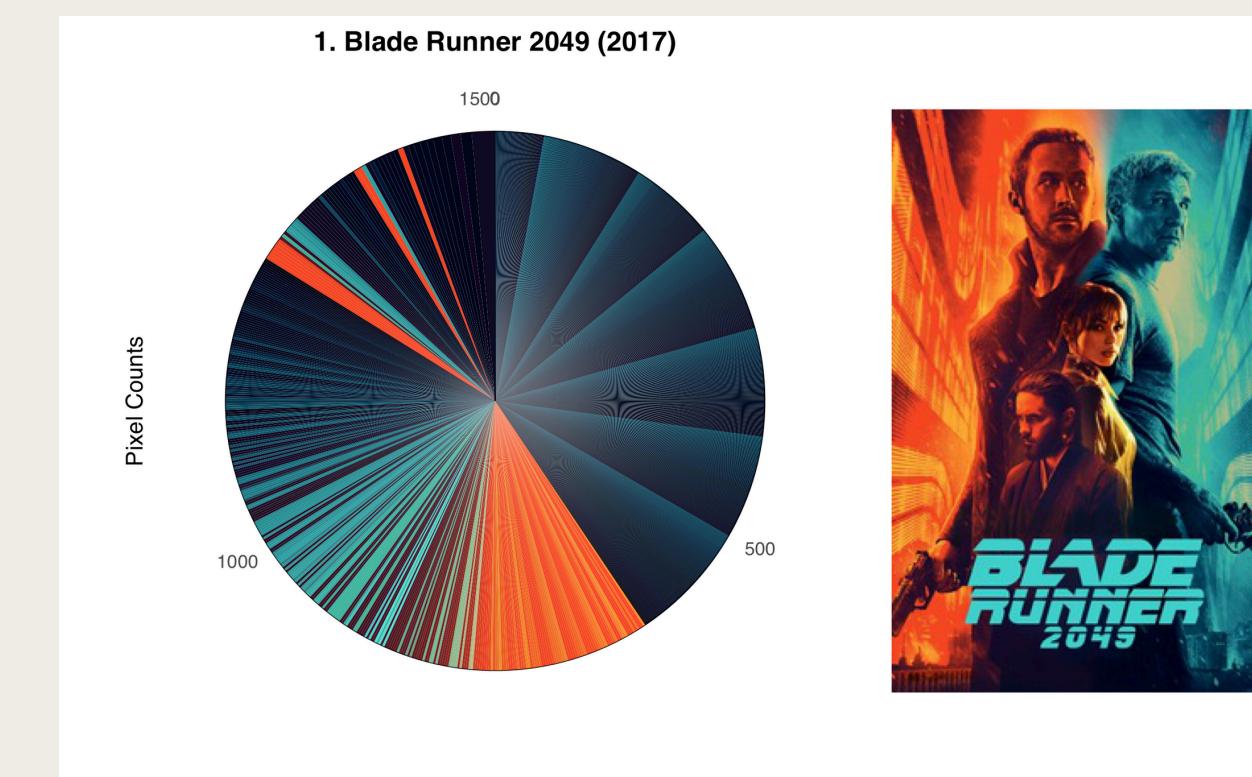
Orion-Nova Productions



Warner Bros. Pictures

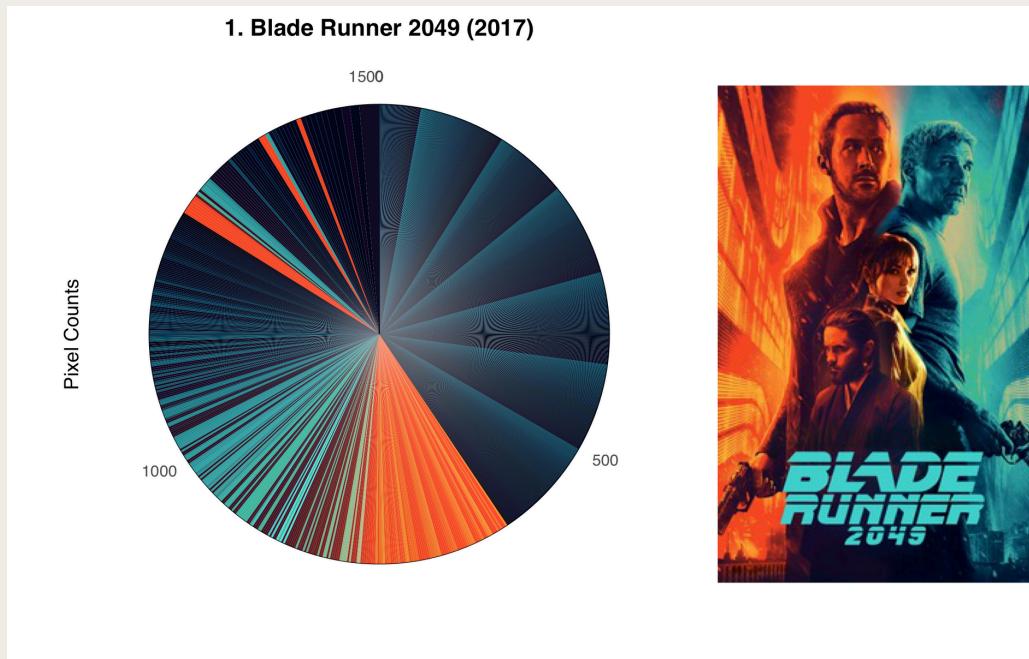


16:14 Entertainment

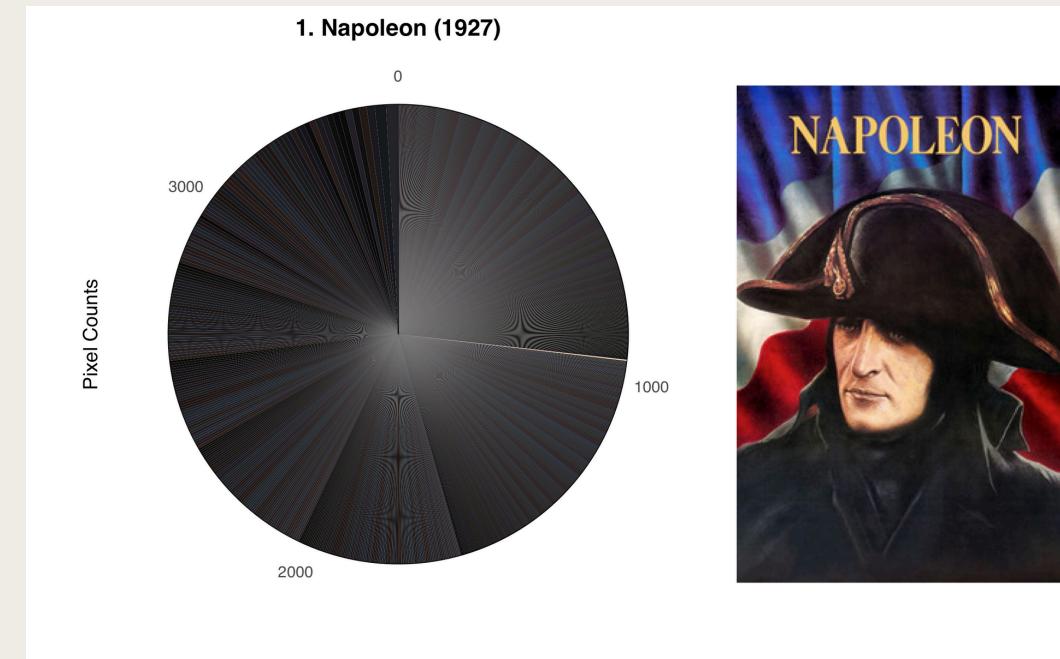


Top Studios (Cumulative) Movie Posters

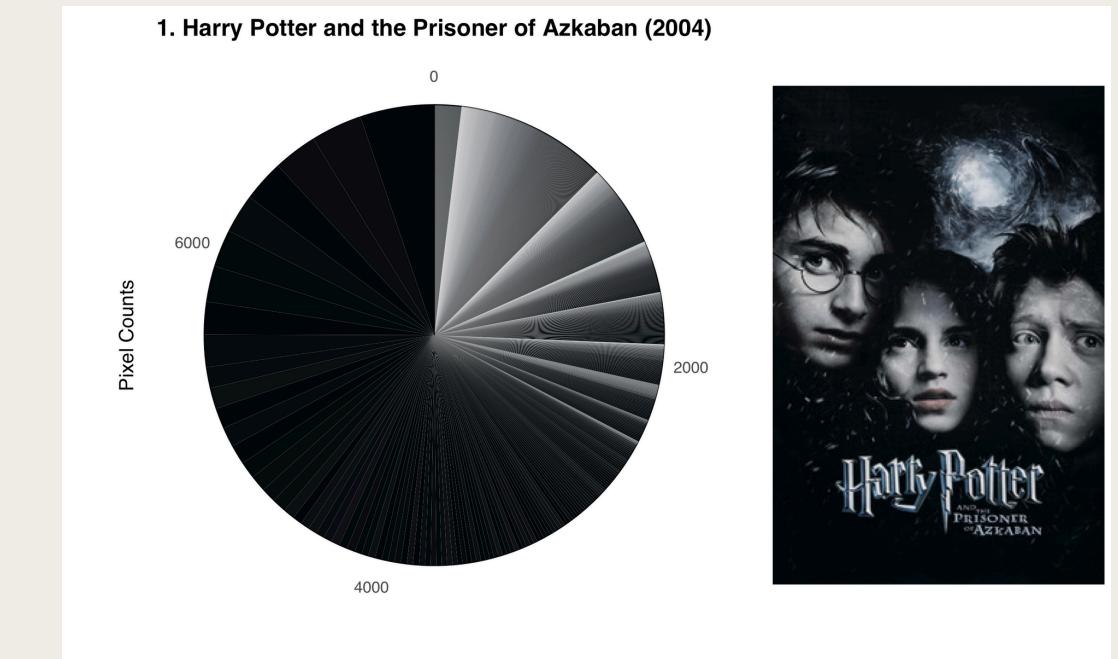
Torridon Films



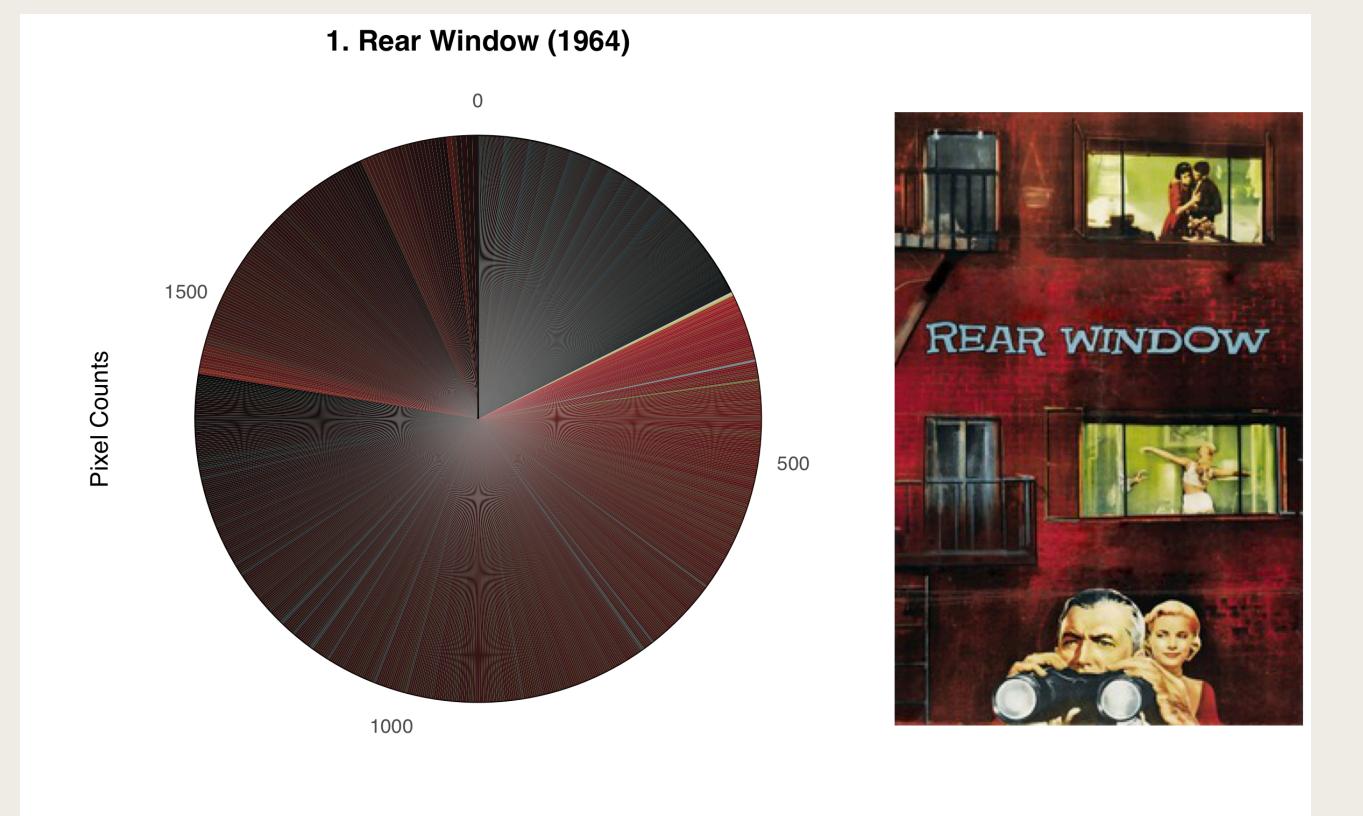
Société Westi



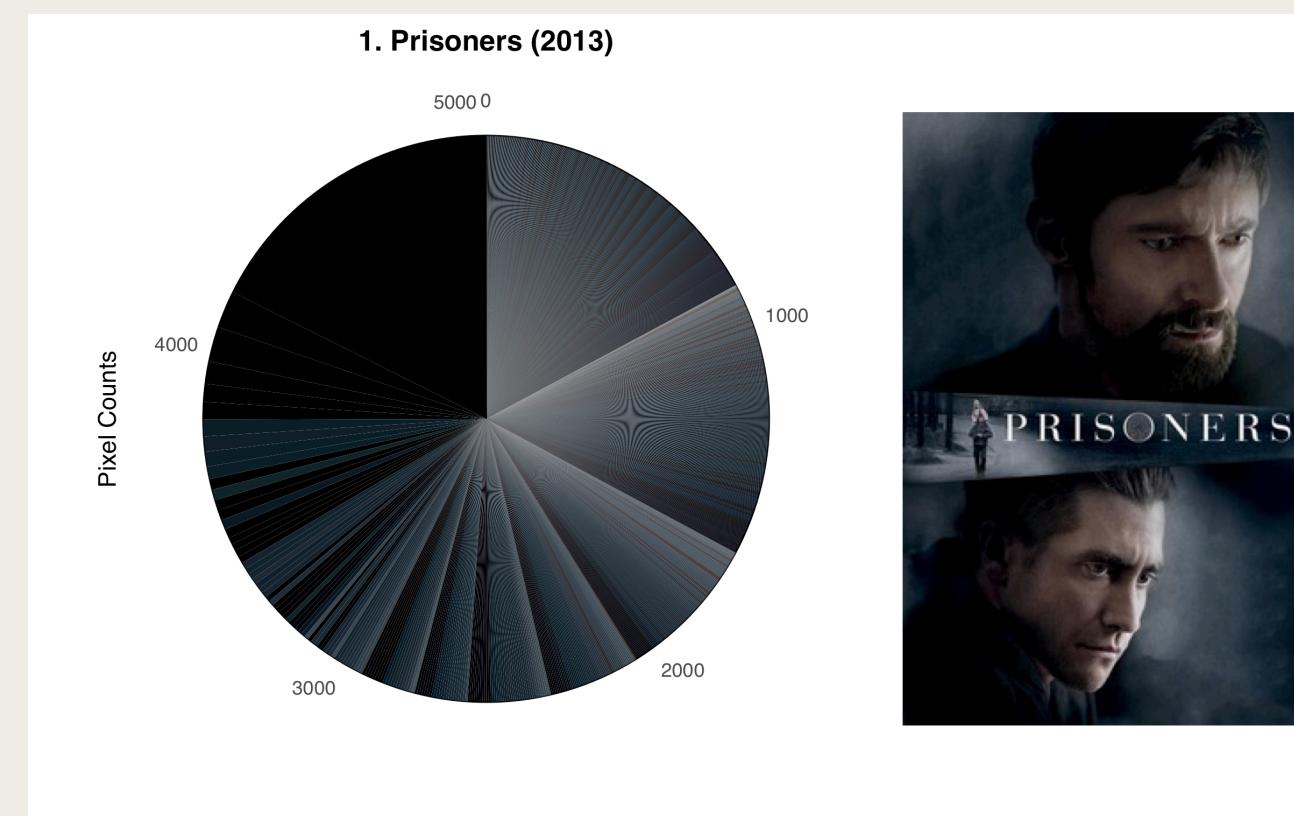
P of A Productions Limited



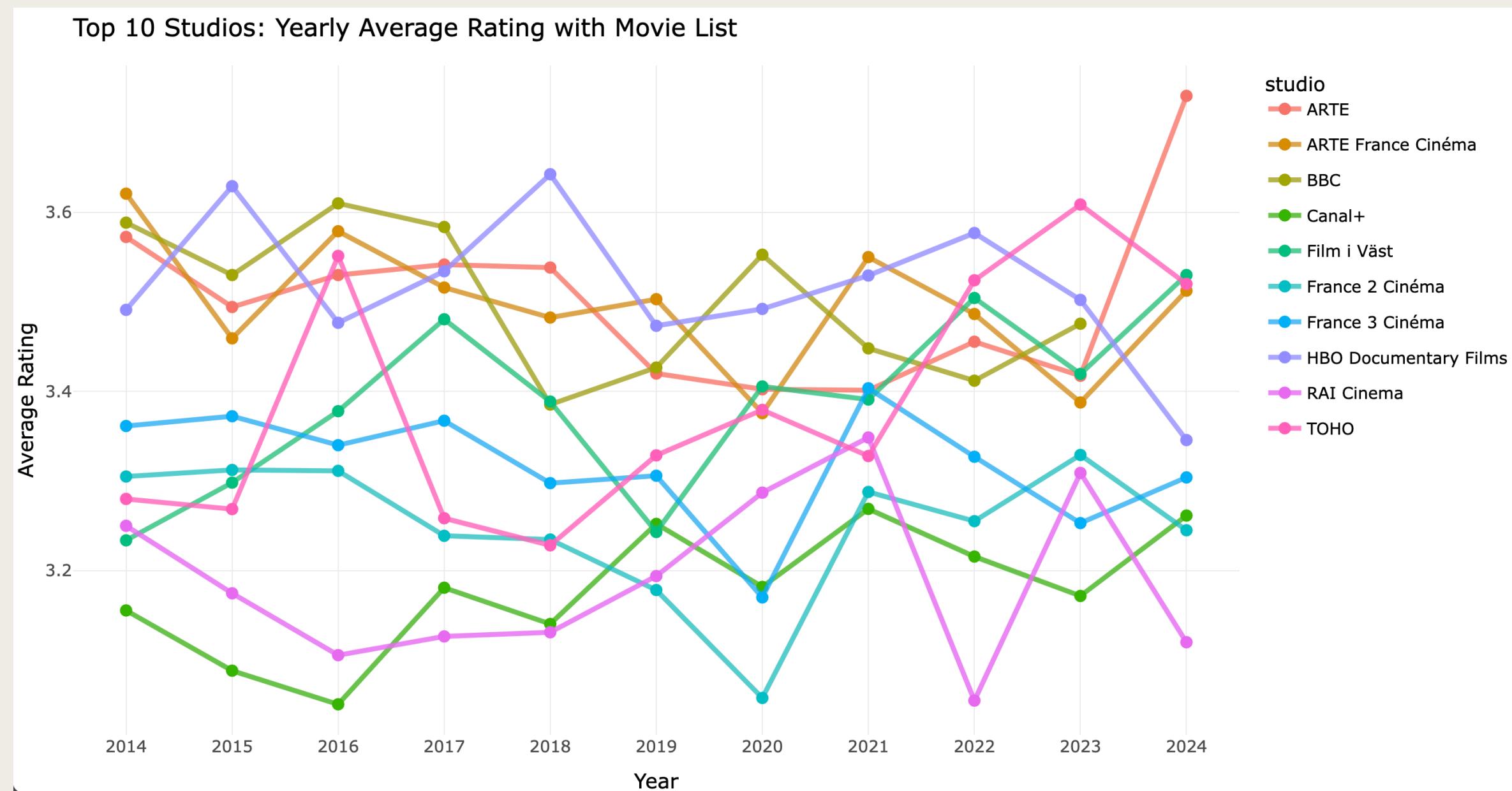
Patron Inc.



8:38 Productions



Method 2 Results



The top 10 most influential studios, in terms of total viewer-rated impact, over the past decade, identify long-term trends in studio reputation and consistency.

Grouped summary statistics (sum of ratings by studio)

Parallel jobs (10 jobs) on CHTC to compute **yearly average ratings for each of the top 10 studios**. These jobs requested 512 MB of memory and 4096 MB of disk space.

Method 2 Results

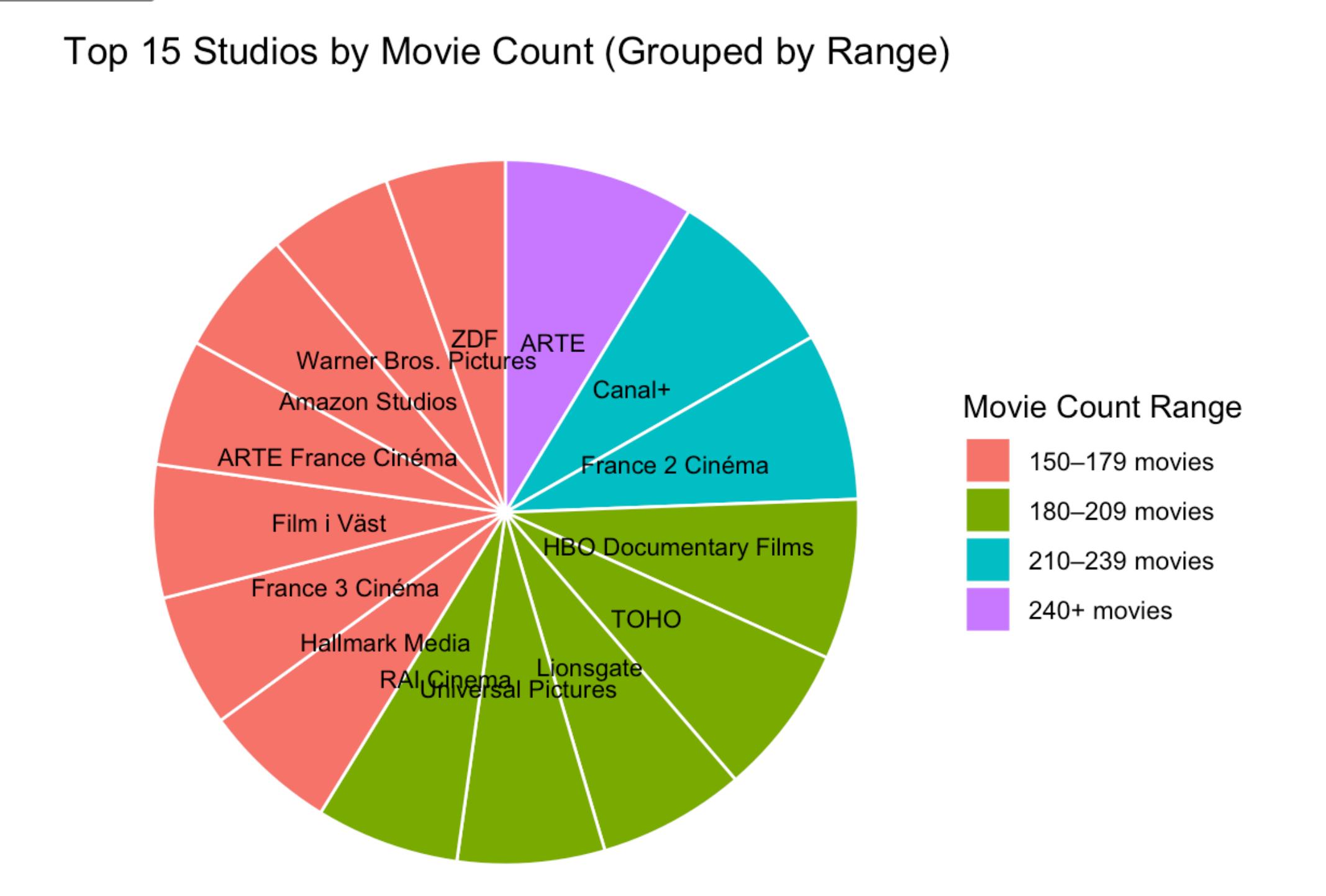
	Df	Sum Sq	Mean Sq	F value	Pr(>F)	
studio	9	29.36	3.262	34.153	< 2e-16	***
date	1	0.03	0.030	0.312	0.576786	
studio:date	9	3.03	0.336	3.522	0.000243	***
Residuals	1885	180.03	0.096			

Signif. codes:	0	'***'	0.001	'**'	0.01	'*'
					0.05	'. '
					0.1	' 1

Successful studios based on average

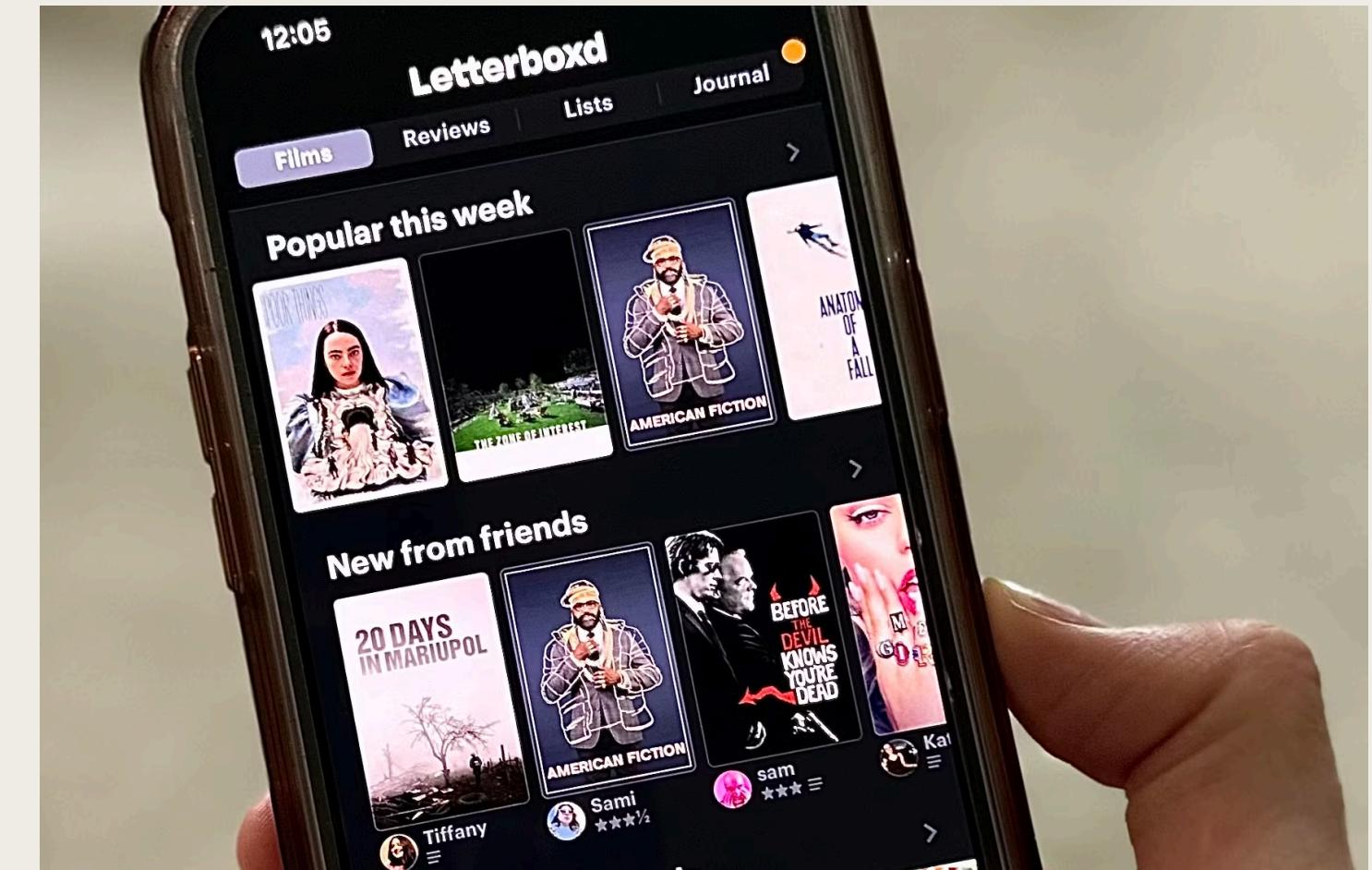
	diff	lwr	upr	p adj
ARTE France Cinéma-ARTE	0.037336667	-0.06229999	0.136973324	0.9743948
BBC-ARTE	0.045000687	-0.05659461	0.146595988	0.9265468
Canal+-ARTE	-0.285308960	-0.37637010	-0.194247823	0.0000000
Film i Väst-ARTE	-0.093076432	-0.19128913	0.005136265	0.0806599
France 2 Cinéma-ARTE	-0.205216333	-0.29727341	-0.113159254	0.0000000
France 3 Cinéma-ARTE	-0.155742983	-0.25378574	-0.057700226	0.0000234
HBO Documentary Films-ARTE	0.053511919	-0.03961688	0.146640715	0.7225183
RAI Cinema-ARTE	-0.272159272	-0.36828808	-0.176030465	0.0000000
TOHO-ARTE	-0.090712958	-0.18540476	0.003978847	0.0736463
BBC-ARTE France Cinéma	0.007664020	-0.10322881	0.118556851	1.0000000
Canal+-ARTE France Cinéma	-0.322645627	-0.42397551	-0.221315743	0.0000000
Film i Väst-ARTE France Cinéma	-0.130413099	-0.23821546	-0.022610740	0.0051247
France 2 Cinéma-ARTE France Cinéma	-0.242552999	-0.34477883	-0.140327169	0.0000000
France 3 Cinéma-ARTE France Cinéma	-0.193079649	-0.30072721	-0.085432091	0.0000007
HBO Documentary Films-ARTE France Cinéma	0.016175253	-0.08701674	0.119367244	0.9999729
RAI Cinema-ARTE France Cinéma	-0.309495939	-0.41540327	-0.203588611	0.0000000
TOHO-ARTE France Cinéma	-0.128049625	-0.23265437	-0.023444880	0.0042978
Canal+-BBC	-0.330309647	-0.43356606	-0.227053233	0.0000000
Film i Väst-BBC	-0.138077119	-0.24769231	-0.028461927	0.0027548
France 2 Cinéma-BBC	-0.250217020	-0.35435281	-0.146081233	0.0000000
France 3 Cinéma-BBC	-0.200743670	-0.31020662	-0.091280715	0.0000003
HBO Documentary Films-BBC	0.008511232	-0.09657316	0.113595621	0.9999999
RAI Cinema-BBC	-0.317159959	-0.42491201	-0.209407911	0.0000000
TOHO-BBC	-0.135713645	-0.24218568	-0.029241610	0.0022736
Film i Väst-Canal+	0.192232528	0.09230247	0.292162583	0.0000001
France 2 Cinéma-Canal+	0.080092628	-0.01379447	0.173979728	0.1733742
France 3 Cinéma-Canal+	0.129565978	0.02980294	0.229329017	0.0016610
HBO Documentary Films-Canal+	0.338820880	0.24388272	0.433759040	0.0000000
RAI Cinema-Canal+	0.013149688	-0.08473304	0.111032421	0.9999928
TOHO-Canal+	0.194596002	0.09812414	0.291067863	0.0000000
France 2 Cinéma-Film i Väst	-0.112139901	-0.21297834	-0.011301462	0.0158561
France 3 Cinéma-Film i Väst	-0.062666551	-0.16899748	0.043664381	0.6916882
HBO Documentary Films-Film i Väst	0.146588351	0.04477059	0.248406117	0.0002332
RAI Cinema-Film i Väst	-0.179082840	-0.28365163	-0.074514047	0.0000029
TOHO-Film i Väst	0.002363474	-0.10088585	0.105612798	1.0000000
France 3 Cinéma-France 2 Cinéma	0.049473350	-0.05119958	0.150146280	0.8686453
HBO Documentary Films-France 2 Cinéma	0.258728252	0.16283441	0.354622095	0.0000000
RAI Cinema-France 2 Cinéma	-0.066942940	-0.16575288	0.031867001	0.4949745
TOHO-France 2 Cinéma	0.114503374	0.01709088	0.211915872	0.0077196
HBO Documentary Films-France 3 Cinéma	0.209254902	0.10760105	0.310908754	0.0000000
RAI Cinema-France 3 Cinéma	-0.116416290	-0.22082549	-0.012007091	0.0153546
TOHO-France 3 Cinéma	0.065030025	-0.03805766	0.168117711	0.6004655
RAI Cinema-HBO Documentary Films	-0.325671192	-0.42548036	-0.225862018	0.0000000
TOHO-HBO Documentary Films	-0.144224877	-0.24265080	-0.045798958	0.0001596
TOHO-RAI Cinema	0.181446314	0.08017718	0.282715444	0.0000007

Method 2 Results



Weaknesses and Future Work

- Young age range of reviewers- confounding variable contributing to ratings
 - Future work could include aggregating movie ratings from critics
- Arbitrary weighting system - metrics being weighed differently on importance in our cumulative model could have impacted our results
 - Future work could include an analysis of how important each metric is, and weighting them on those results
- Flawed data - our dataset contained studios that were involved in the production of only one movie, and/or a handful of related films.
 - Future work: filtering out studios that produced a small number of films



Conclusion

- Using a cumulative approach (Method 1) revealed that many successful studios are smaller but work on large blockbusters
- Analyzing movie ratings alone (Method 2) concluded that successful studios produced movies that tell stories with depth and hold the features of independent studio production
- While there is not one specific recipe that leads to a successful production studio, there are significant findings that reveal themselves depending on metrics one chooses to focus on.

