DATA

First Dataset from IMDB

userId	movieId	rating	timestamp
2	1	5.0	859046895
5	1	4.0	1303501039
8	1	5.0	858610933
11	1	4.0	850815810
14	1	4.0	851766286
17	1	5.0	1350206819
28	1	3.0	884008350

Data set 1 contained movie ratings from

668 unique IMDB users

Dimensions: 105339   4

Computed average rating by MovieID

movieId	mean(rating)
1	3.907328
2	3.353261
3	3.189655
4	2.818182
5	3.250000
6	4.073913
7	3.381818

group_by(imdb_temp, movieId) %>% summarise(mean(rating))

DATA

Second Dataset from IMDB

movieId	title	genres
1	Toy Story (1995)	Adventure|Animation|Children|Comedy|Fantasy
2	Jumanji (1995)	Adventure|Children|Fantasy
3	Grumpier Old Men (1995)	Comedy|Romance
4	Waiting to Exhale (1995)	Comedy|Drama|Romance
5	Father of the Bride Part II (1995)	Comedy
6	Heat (1995)	Action|Crime|Thriller
7	Sabrina (1995)	Comedy|Romance

Dataset 2 contained the movie titles for those movieID's

Merged with first IMDB dataset

movieId	mean(rating)	title	year	genre
1	3.907328	Toy Story	1995	Adventure
2	3.353261	Jumanji	1995	Adventure
3	3.189655	Grumpier Old Men	1995	Comedy
4	2.818182	Waiting to Exhale	1995	Comedy
5	3.250000	Father of the Bride Part II	1995	Comedy
6	4.073913	Heat	1995	Action
7	3.381818	Sabrina	1995	Comedy

Year and main Genre Extracted
* year was important because duplicate movie titles
imdb <- left_join(b,a, by="movieId")

DATA

Movie Finance Dataset from BoxOfficeMojo

X	year	title	budget	domestic	worldwide
1	December 18, 2009	*Avatar*	425000000	760507625	2783918982
2	May 24, 2007	*Pirates of the Caribbean: At World's End*	300000000	309420425	963420425
3	November 6, 2015	*Spectre*	300000000	199044149	875844149
4	July 20, 2012	*The Dark Knight Rises*	275000000	448139099	1084439099
5	July 2, 2013	*The Lone Ranger*	275000000	89289910	259989910
6	March 9, 2012	*John Carter*	275000000	73058679	282778100
7	November 24, 2010	*Tangled*	260000000	200821936	586581936

Cleaned Years and Titles

X	year	title	budget	domestic	worldwide
1	2009	Avatar	425000000	760507625	2783918982
2	2007	Pirates of the Caribbean: At World's End	300000000	309420425	963420425
3	2015	Spectre	300000000	199044149	875844149
6	2012	John Carter	275000000	73058679	282778100
4	2012	The Dark Knight Rises	275000000	448139099	1084439099
5	2013	The Lone Ranger	275000000	89289910	259989910
7	2010	Tangled	260000000	200821936	586581936

```
mojo$title <- gsub("[*]", "", mojo$title) ; mojo$year <- str_sub(mojo$year, -5,-1)
```

DATA

movieId	mean(rating)	title	year	genre
1	3.907328	Toy Story	1995	Adventure
2	3.353261	Jumanji	1995	Adventure
3	3.189655	Grumpier Old Men	1995	Comedy
4	2.818182	Waiting to Exhale	1995	Comedy

imdb

X	year	title	budget	domestic	worldwide
1	2009	Avatar	425000000	760507625	2783918982
2	2007	Pirates of the Caribbean: At World's End	300000000	309420425	963420425
3	2015	Spectre	300000000	199044149	875844149
6	2012	John Carter	275000000	73058679	282778100
4	2012	The Dark Knight Rises	275000000	448139099	1084439099

mojo

movies <- left_join(mojo, imdb, by = c('title','year'))

| X	year	title	budget	domestic	worldwide	movieId	rating	genre	overseas
1	2009	Avatar	425000000	760507625	2783918982	72998	3.856061	Action	2023411357
2	2007	Pirates of the Caribbean: At World's End	300000000	309420425	963420425	53125	3.569444	Action	654000000
3	2015	Spectre	300000000	199044149	875844149	136020	3.428571	Action	676800000
6	2012	John Carter	275000000	73058679	282778100	93363	3.250000	Action	209719421
7	2010	Tangled	260000000	200821936	586581936	81847	3.058333	Animation	385760000

FINAL DATA SET! ✨

HEADLINES
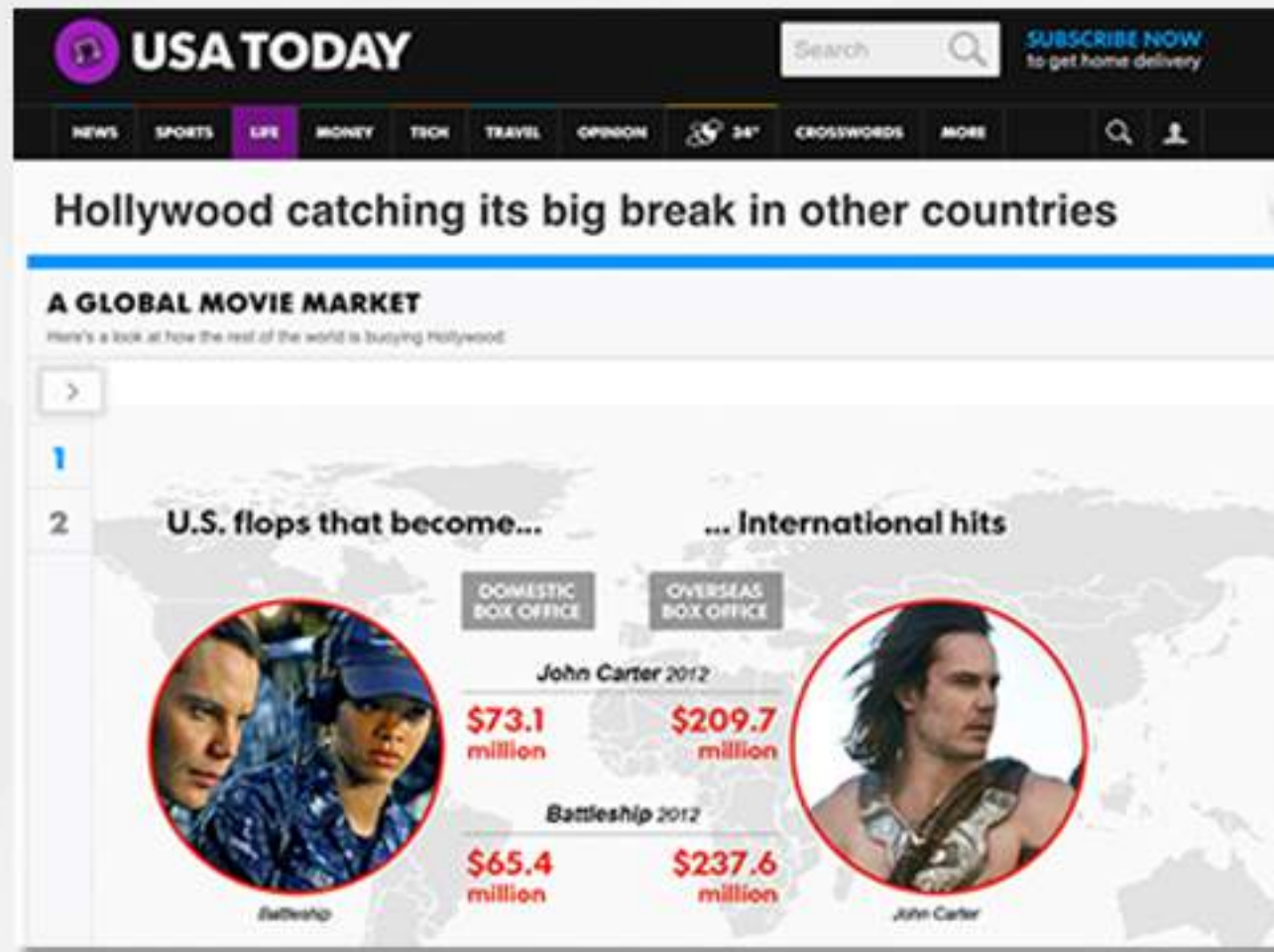
"Where Americans once were the only game in town for Hollywood, U.S. audiences are taking a back seat to moviegoers across the globe – particularly in Asia."

"And foreign markets are getting the industry's highest-profile films first. Battleship opened in Asia and Europe more than a month before it reached the USA last May.

http://usat.ly/ZuKGrn

VISUALIZATION

Domestic vs Foreign Revenue
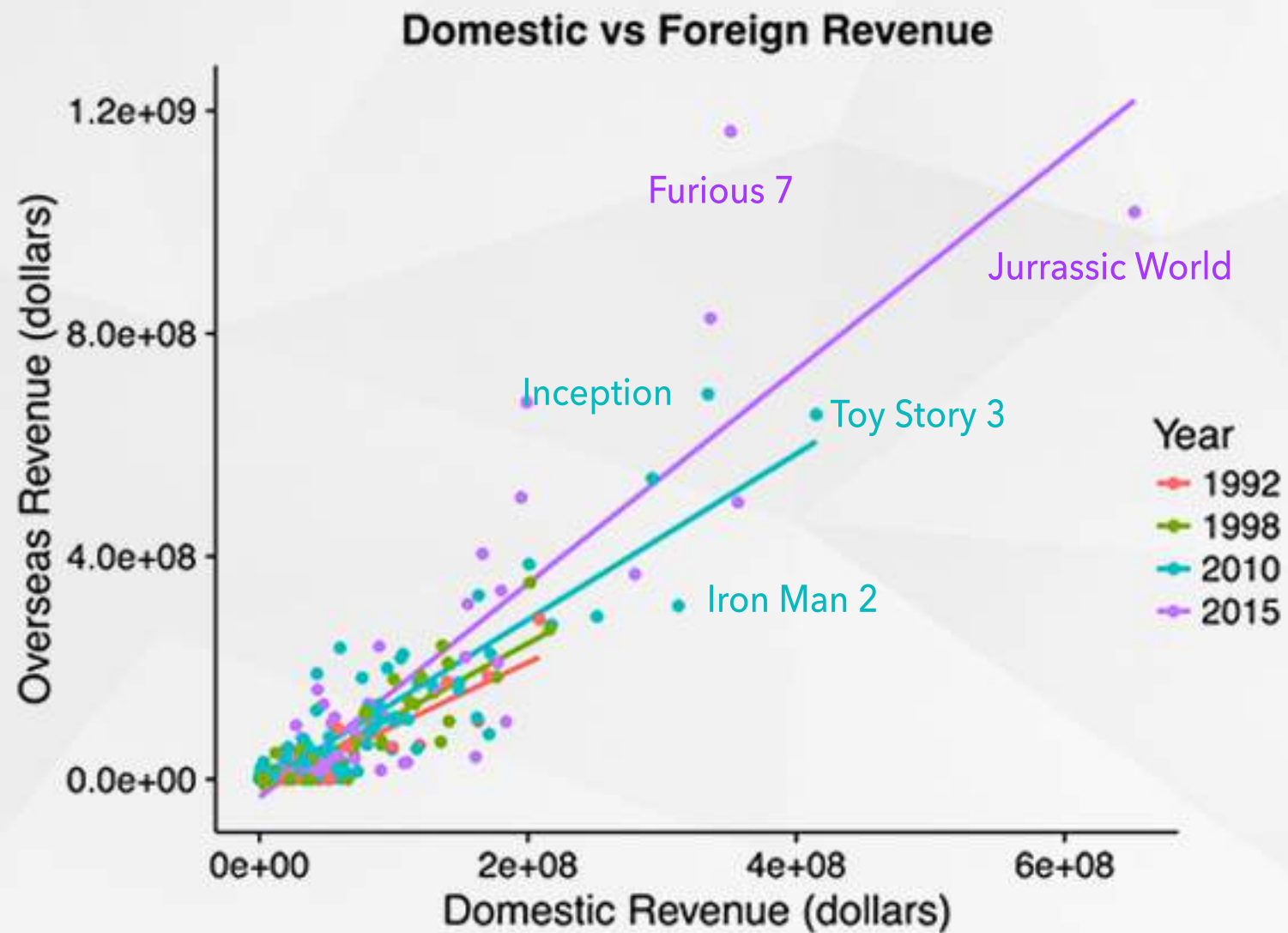
WHAT THIS SHOWS:

Overseas revenue has accounted for an increasing percentage of total revenue every year since 1992.

HEADLINES

The Economist | World politics | Business & finance | Econo

Hollywood goes global
Bigger abroad

"Big noisy spectacle travels best. Jason Statham, the close-cropped star of many a mindlessly violent film, is a particular Russian favourite. Films based on well-known literature (including cartoon books) and myths may also fare well."

"Comedy travels badly: Will Ferrell and Adam Sandler provoke guffaws at home but incomprehension abroad

http://econ.st/KAk82t

VISUALIZATION
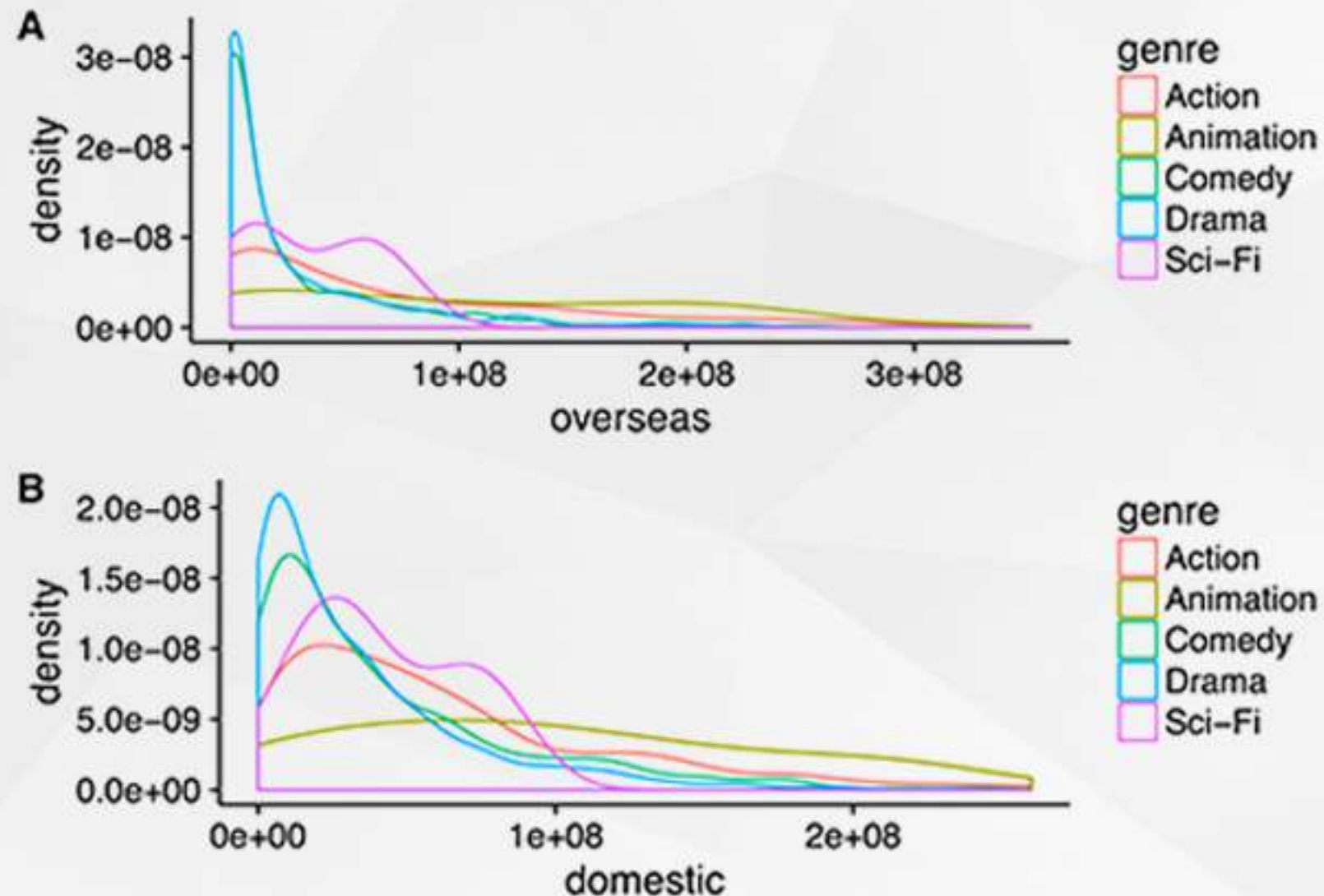
These Density plots confirm Drama and Comedy genres perform worse Overseas vs Domestically. Majority of Drama/Comedy movies generate less than 50 mil overseas, while Action and Animation genres show a much wider distribution of revenues.
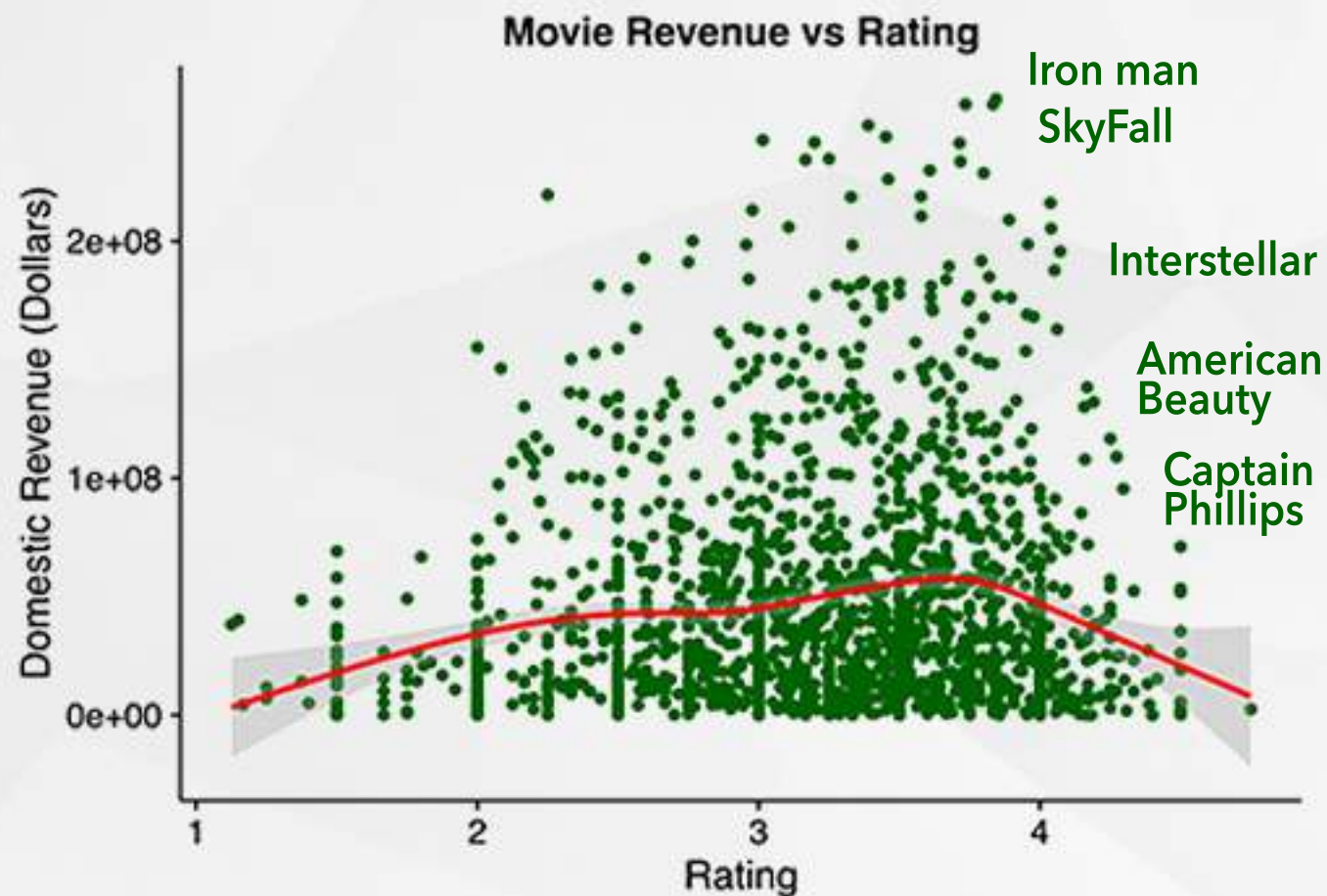
HEADLINES
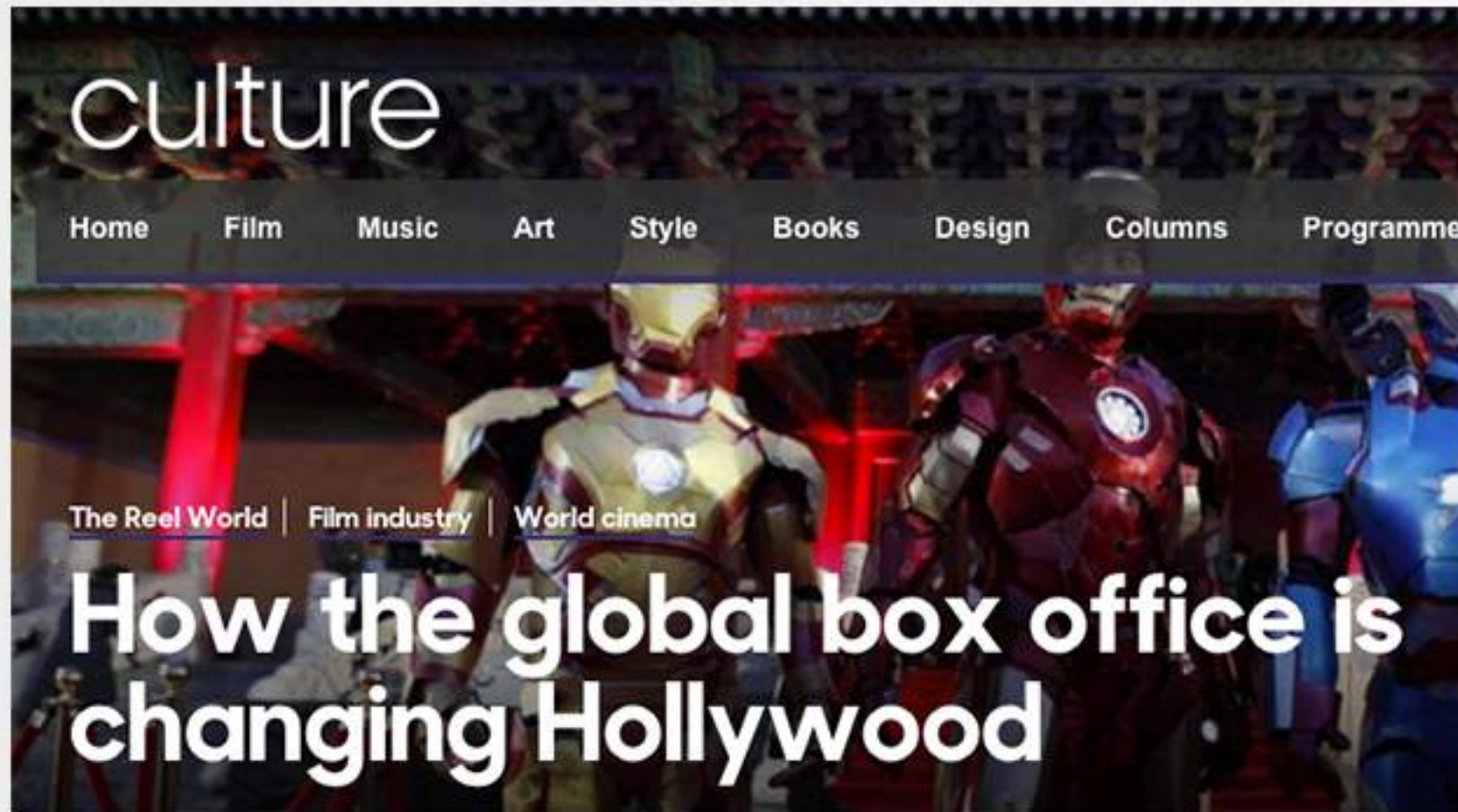
"...little effort is being made to deliver sophisticated storytelling ... movies are crafted mainly to provoke visceral - as opposed to intellectual response"

bbc.com/culture/story/20130620-is-china-hollywoods-future

VISUALIZATION

The rating sweet spot that generates the most revenue is between

3.5 and 3.8. Movies that score greater than 4 show a sharp decline in revenues.

This could be due to the fact that the average movie goer more easily appreciates

an average movie (cough cough ** Michael Bay movies).

VISUALIZATION

Using Top 100 movies 2015
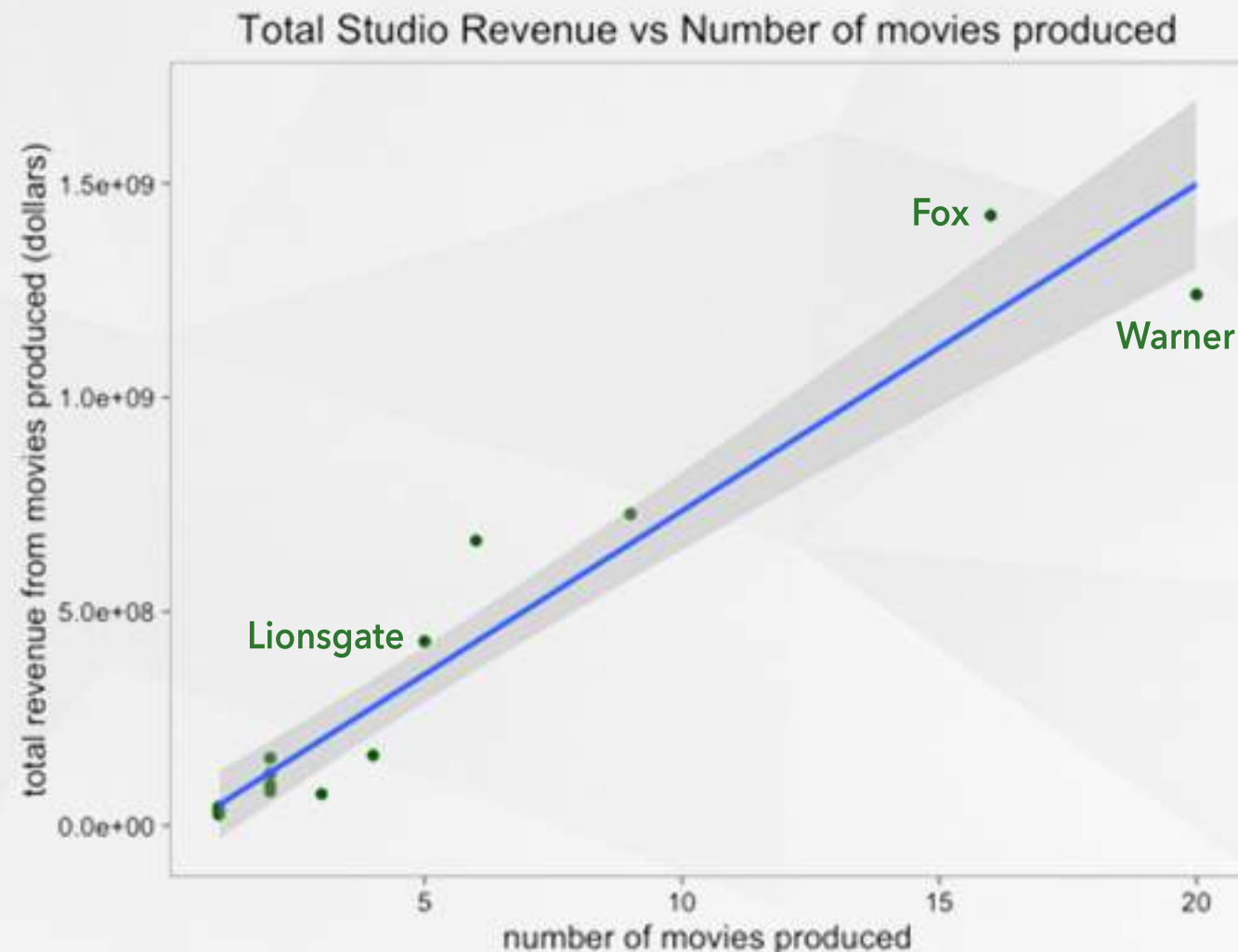
There seems to be a linear trend between the number of movies a studio produces and its total domestic revenue. This doesn't have to be the case; for example, Lionsgate could have produced 10 of the top 100 movies in 2015 and generated 100mil revenue. This means movie studios do equally well selecting which movies to produce.

VISUALIZATION

Domestic Gross vs Opening Weekend
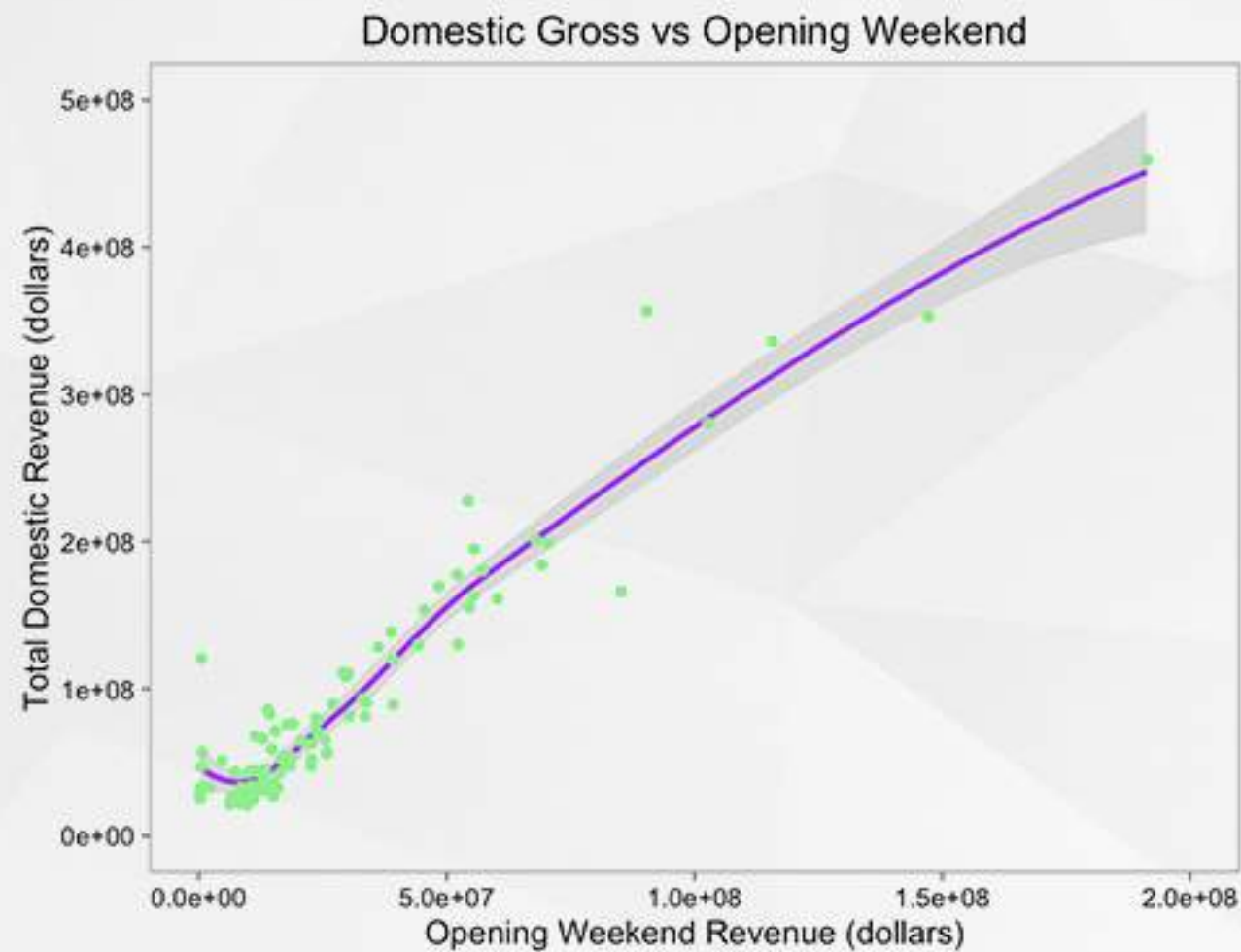
There seems to be a linear trend between how much revenue a movie made on its opening weekend and its lifetime domestic revenue.

NEXT STEPS

Possibly use a data scrapping method to extract my own data from sites like BoxOfficeMojo and the-Numbers so I can get more detailed info such as: day-by-day revenue per movie, and revenue by Country ~ genre.

See if there is any relationship between movie ratings/revenue by script word count or vocabulary.