WARNER BROS. 2020 Q4 PRODUCTION RESEARCH

SUMMARY

- We were given a dataset containing a dozens of variables about thousands of films
- We analyzed the data and tested the relationships between gross income and a variety of other variables
- We summarized our findings and prepared recommendations for future production investment decisions

VARIABLES IN THE DATASET

Movie Production Data Dictionary

Variables Utilized	Data Attributes	Data Type	Description
Color	color	String	movie is color or black&white
Director	director_name	String	name of the director
Number of Critics that Reviewed	num_critic_for_reviews	Integer	number of critics that reviewed the movie
Duration (min)	duration	Integer	duration of the movie
Director FB Likes	director_facebook_likes	Integer	facebook likes for director
Actor 1	actor_1_name	String	name of actor 1
Actor 2	actor_2_name	String	name of actor 2
Actor 3	actor_3_name	String	name of actor 3
Actor 1 FB Likes	actor_1_facebook_likes	Integer	facebook likes for actor 1
Actor 2 FB Likes	actor_2_facebook_likes	Integer	facebook likes for actor 2
Actor 3 FB Likes	actor_3_facebook_likes	Integer	facebook likes for actor 3
Gross Income	gross	String	gross income of the movie
Genres	genres	String	genres of the movie
Movie Title	movie_title	String	title of the movie
Cast Total FB Likes	cast_total_facebook_likes	Integer	facebook likes for total cast
Number of Faces in Poster	facenumber_in_poster	Integer	number of faces in the poster
Number of Users that Reviewed	num_user_for_reviews	Integer	number of users that reviewed the movie
Language	language	String	language of the movie
Country	country	String	country in which the movie belongs
Content Rating	content_rating	String	rating of the movie(PG-13, R- rated and others)
Budget	budget	String	total budget of the movie
Title Year	title_year	Integer	year in which the it was released
IMDb Score	imdb_score	Float	imdb score of the movie
Aspect Ratio	aspect_ratio	Float	aspect ratio in which the movie is filmed.
	~~num_voted_users~~	~~Integer~~	~~number of users that voted for the movie~~
	~~plot_keywords~~	~~String~~	~~keywords of the plot of the movie~~
	~~movie_imdb_link~~	~~String~~	~~imdb link for the movie~~

DATA CLEANING

- Deleted rows containing data for TV and other ambiguous categories:
  - TV-MA
  - TV-Y
  - TV-14
  - TV-PG
  - TV-G
  - Approved
  - Passed
- Renamed columns to increase clarity
- Deleted columns like:
  - IMDb link
  - Plot keywords
  - # of users that voted for the movie
- Excluded outliers and incorrect data
- Missing data was automatically excluded from certain analyses

QUESTIONS

1) Are Facebook engagements significantly correlated with gross income?
2) Can budget be used to predict gross income?
3) Do different genres have different levels of mean gross income?
4) Do different directors have different levels of mean gross income? If so, which directors have movies with the highest mean gross income?
5) Do different countries have different levels of mean gross income? If so, which country generates the highest level of mean gross income?
6) Do all content ratings have the same levels of mean gross income? If not, which content rating category has the highest mean gross income?

HYPOTHESES

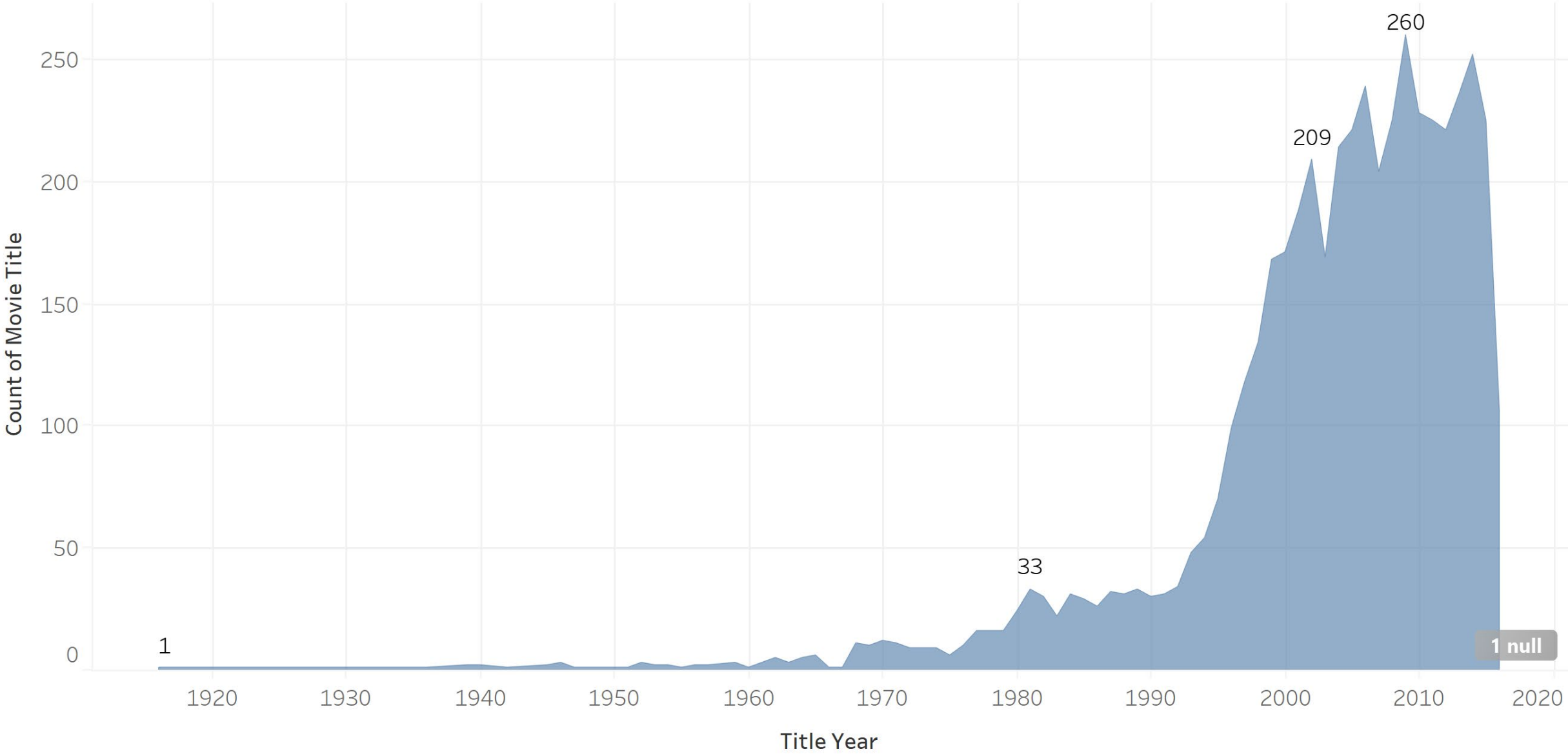
1) H_0 = there is no significant correlation between Facebook engagements and gross income
2) H_0 = there is no significant correlation between budget and gross income
3) H_0 = all genres have the same mean gross income
4) H_0 = all directors have the same mean gross income
5) H_0 = all countries have the same mean gross income
6) H_0 = all content ratings have the same mean gross income

DATA VISUALIZATION

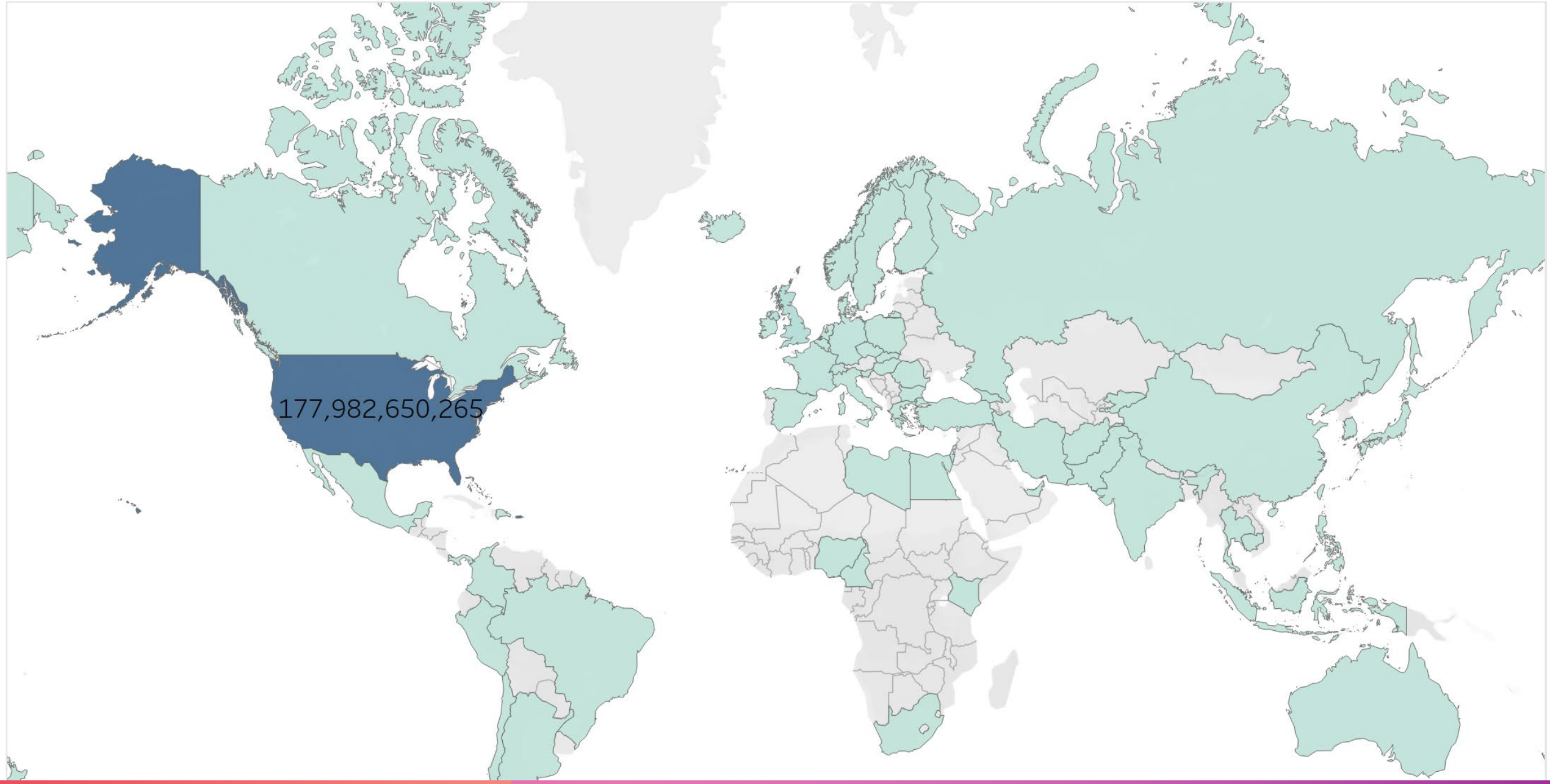
MAIN GRAPHICS

- Number of movies released each year
- Median gross income by title year
- Mean gross income by country of origin
- Sum of gross income by country
- Mean gross income by content rating
- Top 25 directors by mean gross income
- Top 10 directors by mean gross income
- Top 10 directors by IMDb score

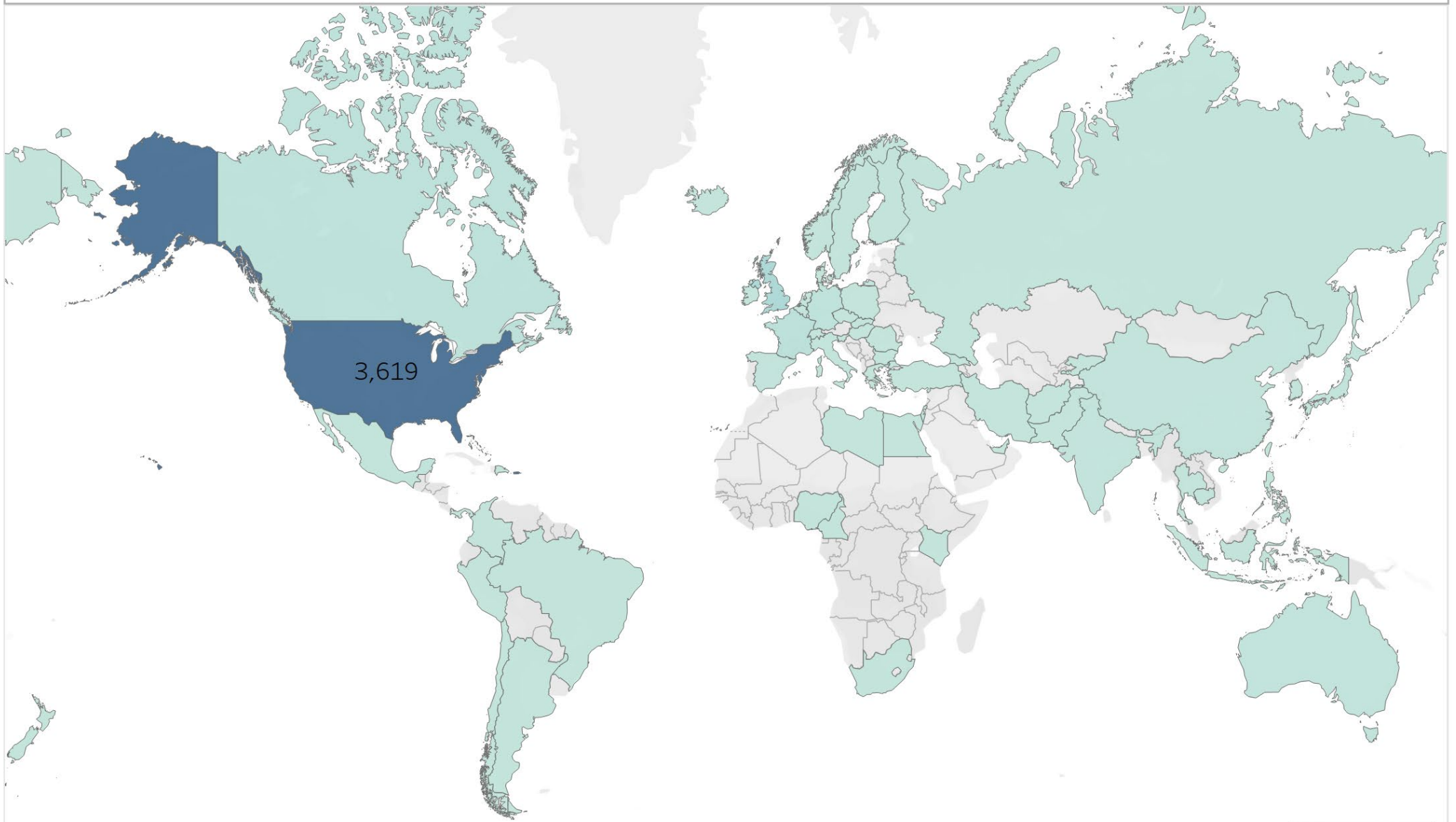
Movies Released per Year



Sum of Gross Income by Country of Origin



Count of Movies Produced by Country

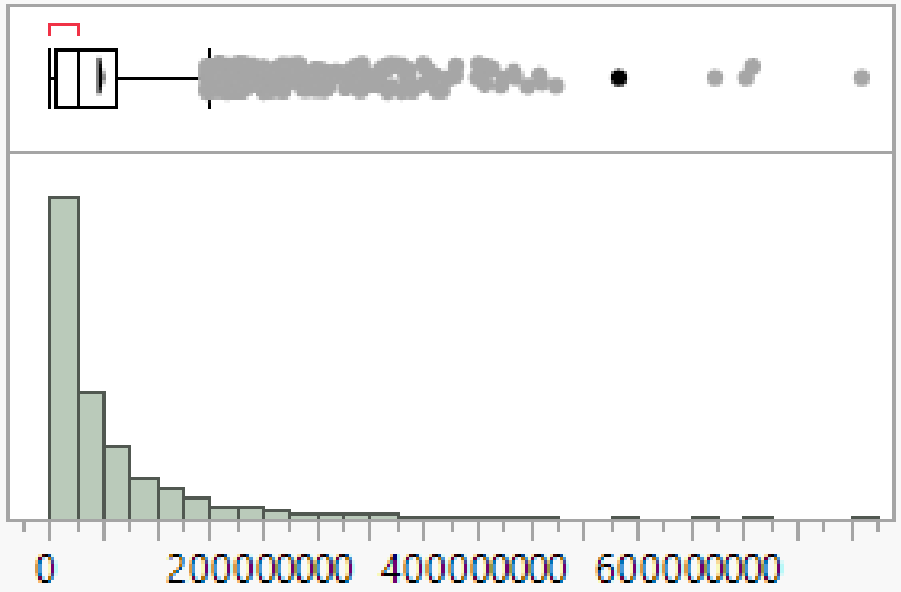


CNTD(Movie Title)

1 3,619

Distributions

Gross Income



Quantiles

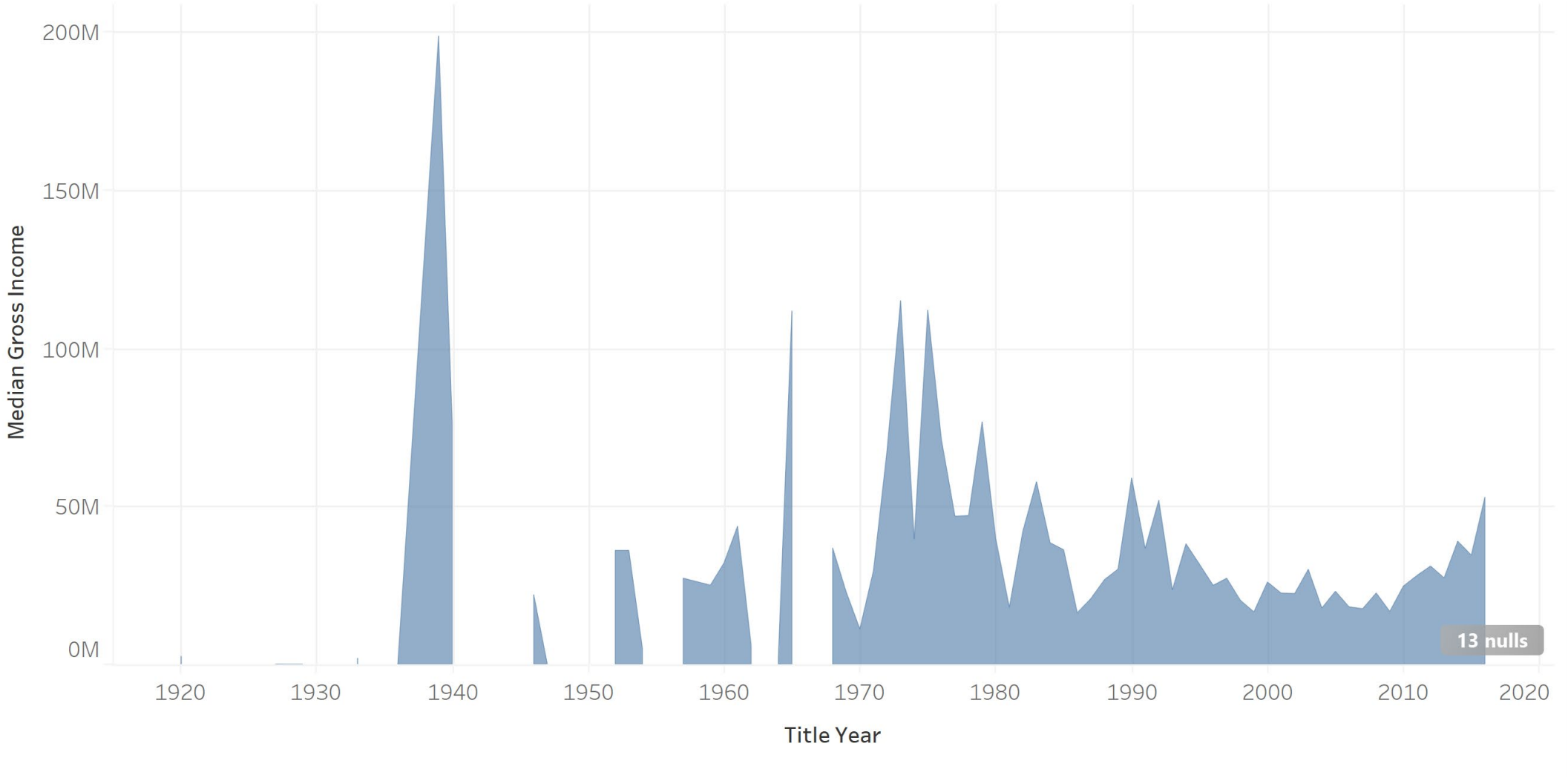
100.0%	maximum	760505847
99.5%		40259040.09
97.5%		24484360.45
90.0%		12530988.24
75.0%	quartile	6231415.625
50.0%	median	2552968.5
25.0%	quartile	535437.25
10.0%		38072.3
2.5%		2388.4325
0.5%		398.9605
0.0%	minimum	162

Summary Statistics

Mean	4852020.7
Std Dev	6855535.8
Std Err Mean	106598.51
Upper 95% Mean	5061011.1
Lower 95% Mean	4643030.3
N	4136

DISTRIBUTION OF MOVIES BY GROSS INCOME

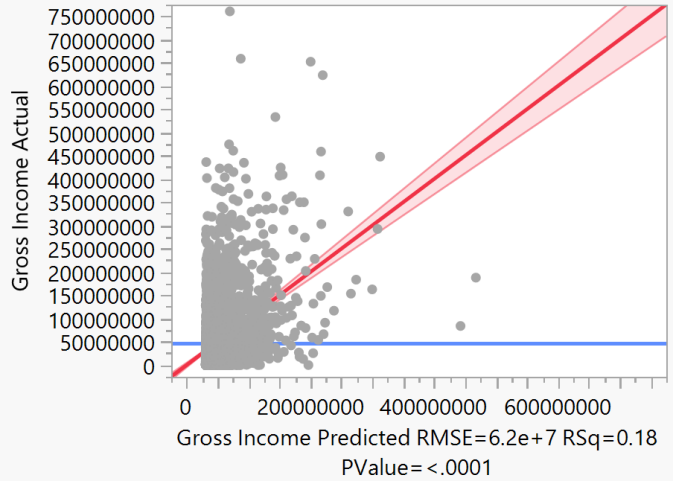
MEDIAN GROSS INCOME BY TITLE YEAR



HYPOTHESIS 1

Facebook engagements have a significant relationship with **gross income**

- **Multiple linear regression** using:
  - Movie FB likes
  - Cast total FB likes
  - Director FB likes
- **P-value** < 0.001* for all the variables meaning they all have significant relationships with gross income
- **R-square adjusted** (coefficient of determination) ~ 0.178
  - Weak but significant
- **No multicollinearity** – no VIF above 5



Summary of Fit

RSquare	0.177784
RSquare Adj	0.177187
Root Mean Square Error	62193787
Mean of Response	48543113
Observations (or Sum Wgts)	4134

Analysis of Variance

Source	DF	Sum of Squares	Mean Square	F Ratio
Model	3	3.4542e+18	1.151e+18	297.6699
Error	4130	1.5975e+19	3.868e+15	Prob > F
C. Total	4133	1.9429e+19		<.0001*

Effect Summary

Source	LogWorth	PValue
Movie FB Likes	107.365	0.00000
Cast Total FB Likes	30.405	0.00000
Director FB Likes	6.037	0.00000

Remove Add Edit ☐ FDR

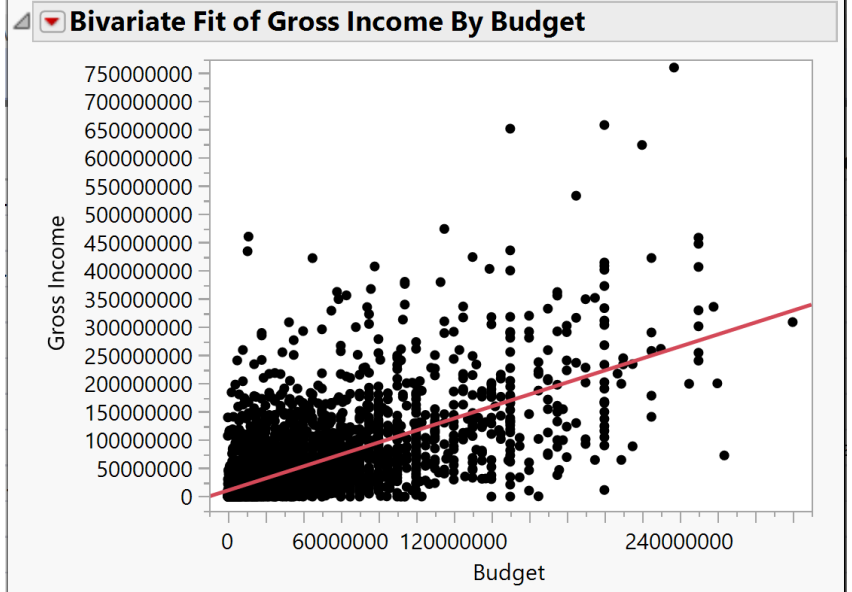
Parameter Estimates

Term	Estimate	Std Error	t Ratio	Prob>|t|	VIF
Intercept	30892759	1168892	26.43	<.0001*	.
Director FB Likes	1638.4788	333.297	4.92	<.0001*	1.0356108
Cast Total FB Likes	626.27258	53.52815	11.70	<.0001*	1.0544321
Movie FB Likes	1091.1926	47.95222	22.76	<.0001*	1.0671486

HYPOTHESIS 2

Budget is significantly correlated with **gross income**

- Simple linear regression
- **R-square** (coefficient of determination): ~0.426
- **Gross income** = 11,023,963 + 1.062(budget)
- **Prob > F:** <0.0001*
  - Significant because p-value is less than 0.05



Linear Fit

Linear Fit

Gross Income = 11027963 + 1.0622611*Budget

Summary of Fit

RSquare	0.425511
RSquare Adj	0.425363
Root Mean Square Error	52962837
Mean of Response	51229640
Observations (or Sum Wgts)	3877

Lack Of Fit

Analysis of Variance

Source	DF	Sum of Squares	Mean Square	F Ratio
Model	1	8.0509e+18	8.051e+18	2870.124
Error	3875	1.087e+19	2.805e+15	Prob > F
C. Total	3876	1.892e+19		<.0001*

Parameter Estimates

Term	Estimate	Std Error	t Ratio	Prob>|t|
Intercept	11027963	1134291	9.72	<.0001*
Budget	1.0622611	0.019828	53.57	<.0001*

HYPOTHESIS 3

Movies of different **genres** will have significantly different mean **gross incomes**

- **Prob > F:** <0.0001*

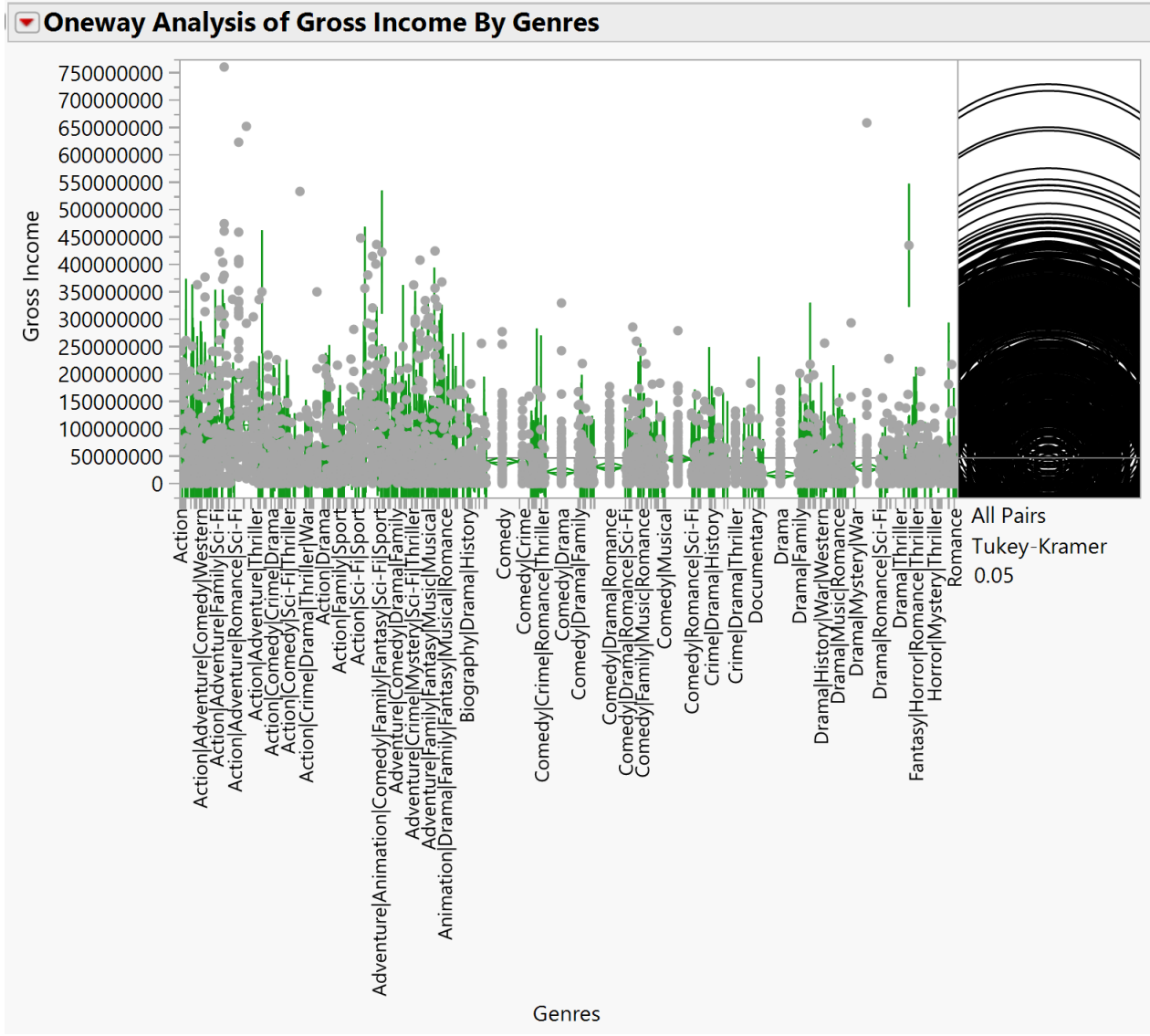
Oneway Anova

Summary of Fit

Rsquare	0.432125
Adj Rsquare	0.299644
Root Mean Square Error	57375887
Mean of Response	48531409
Observations (or Sum Wgts)	4135

Analysis of Variance

Source	DF	Sum of Squares	Mean Square	F Ratio	Prob > F
Genres	782	8.3969e+18	1.074e+16	3.2618	<.0001*
Error	3352	1.1035e+19	3.292e+15		
C. Total	4134	1.9432e+19			



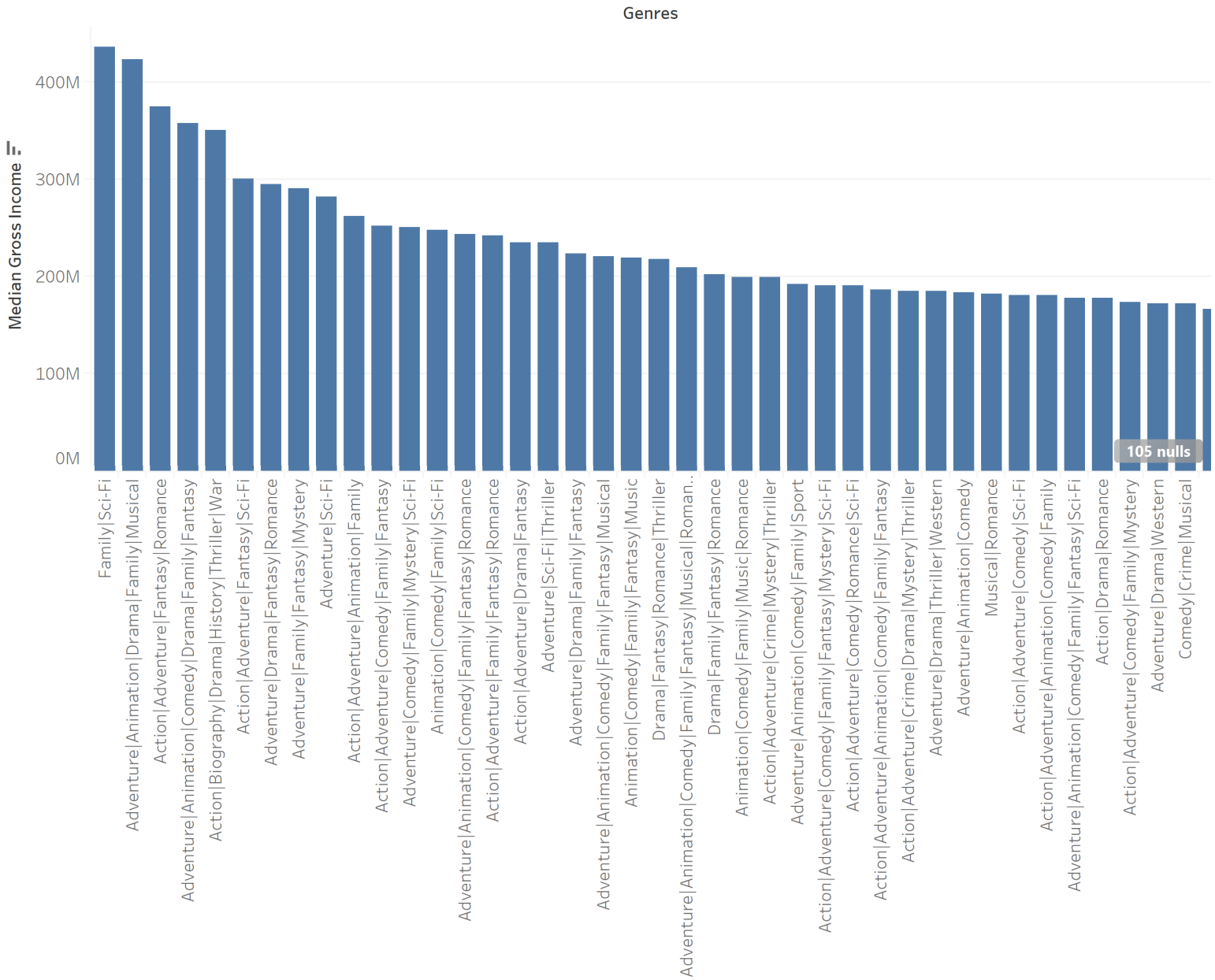
HYPOTHESIS 3: CONNECTING LETTERS REPORT

Connecting Letters Report

Level	Mean
Family|Sci-Fi	434949459
Adventure|Animation|Drama|Family|Musical	422783777
Adventure|Animation|Comedy|Drama|Family|Fantasy	356454367
Action|Biography|Drama|History|Thriller|War	350123553
Action|Adventure|Fantasy|Sci-Fi	296684758
Adventure|Drama|Fantasy|Romance	295436148
Action|Adventure|Fantasy|Romance	289279970
Adventure|Family|Fantasy	126460533
Adventure|Animation|Comedy|Family	126298501
Action|Adventure|Comedy|Crime|Thriller	125305545
Action|Biography|Drama|Thriller|War	125069696
Action|Adventure	124983128

TOP GENRES BY MEDIAN GROSS INCOME

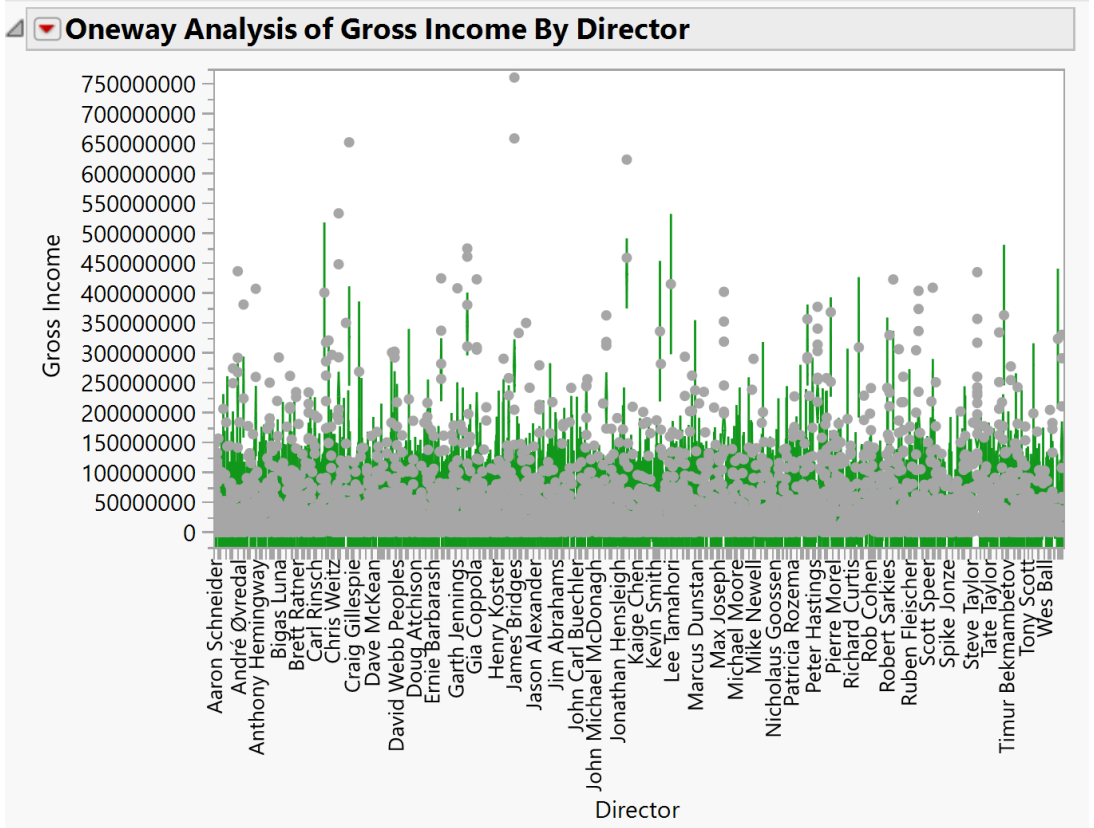
- Column chart showing the top genres by mean gross income
- A large proportion of the films fit into multiple genre categories
- **Top genres** by gross income:
  1) Family | Sci-Fi
  2) Adventure | Animation | Drama | Family Musical
  3) Action | Adventure | Fantasy | Romance



HYPOTHESIS 4

Different **directors** have movies with different levels of mean **gross income**

- ANOVA
- P-value < 0.0001* meaning different directors have significantly different levels of mean gross income



Oneway Anova

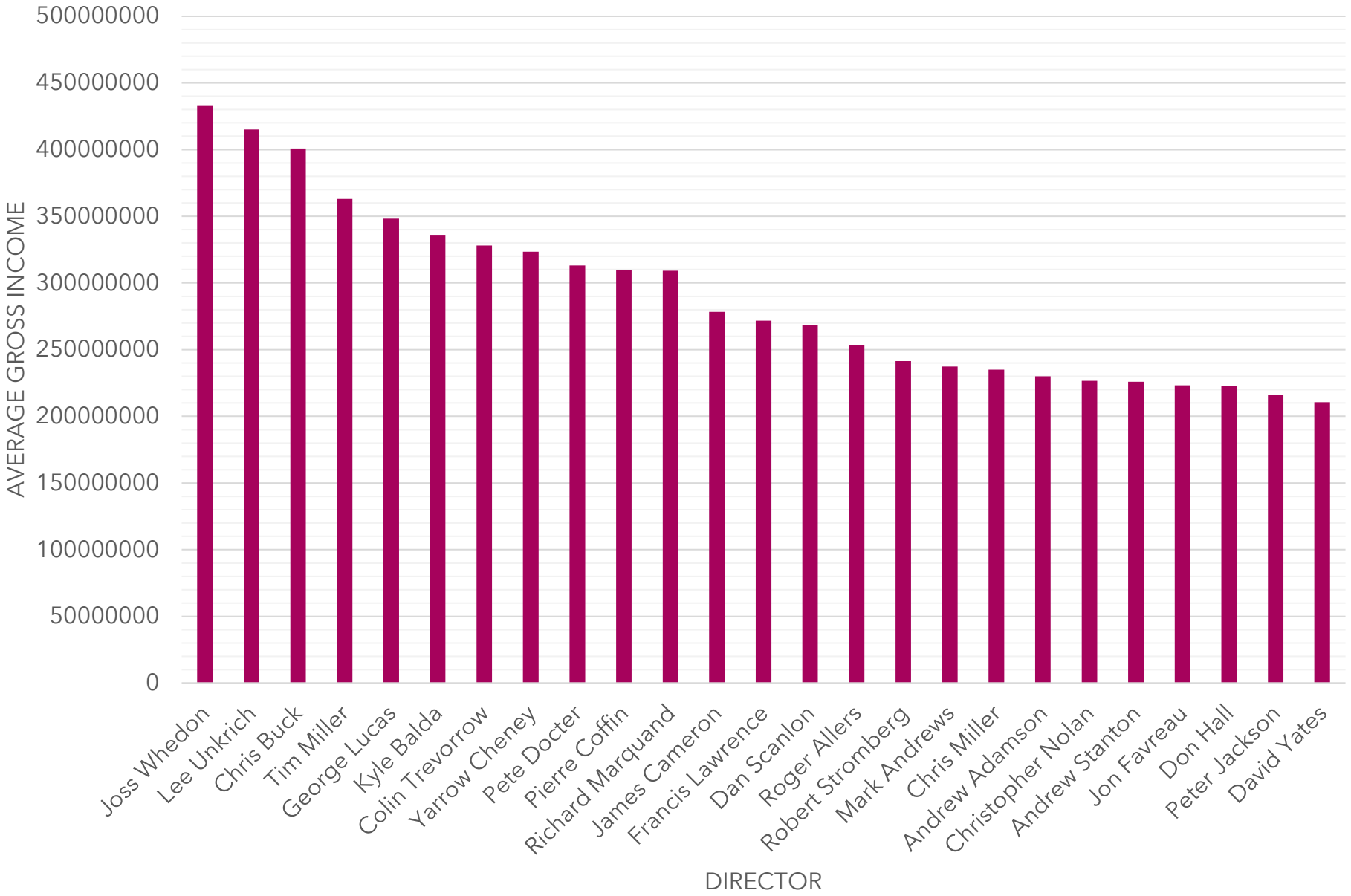
Summary of Fit

Rsquare	0.582062
Adj Rsquare	0.238723
Root Mean Square Error	59822934
Mean of Response	48543113
Observations (or Sum Wgts)	4134

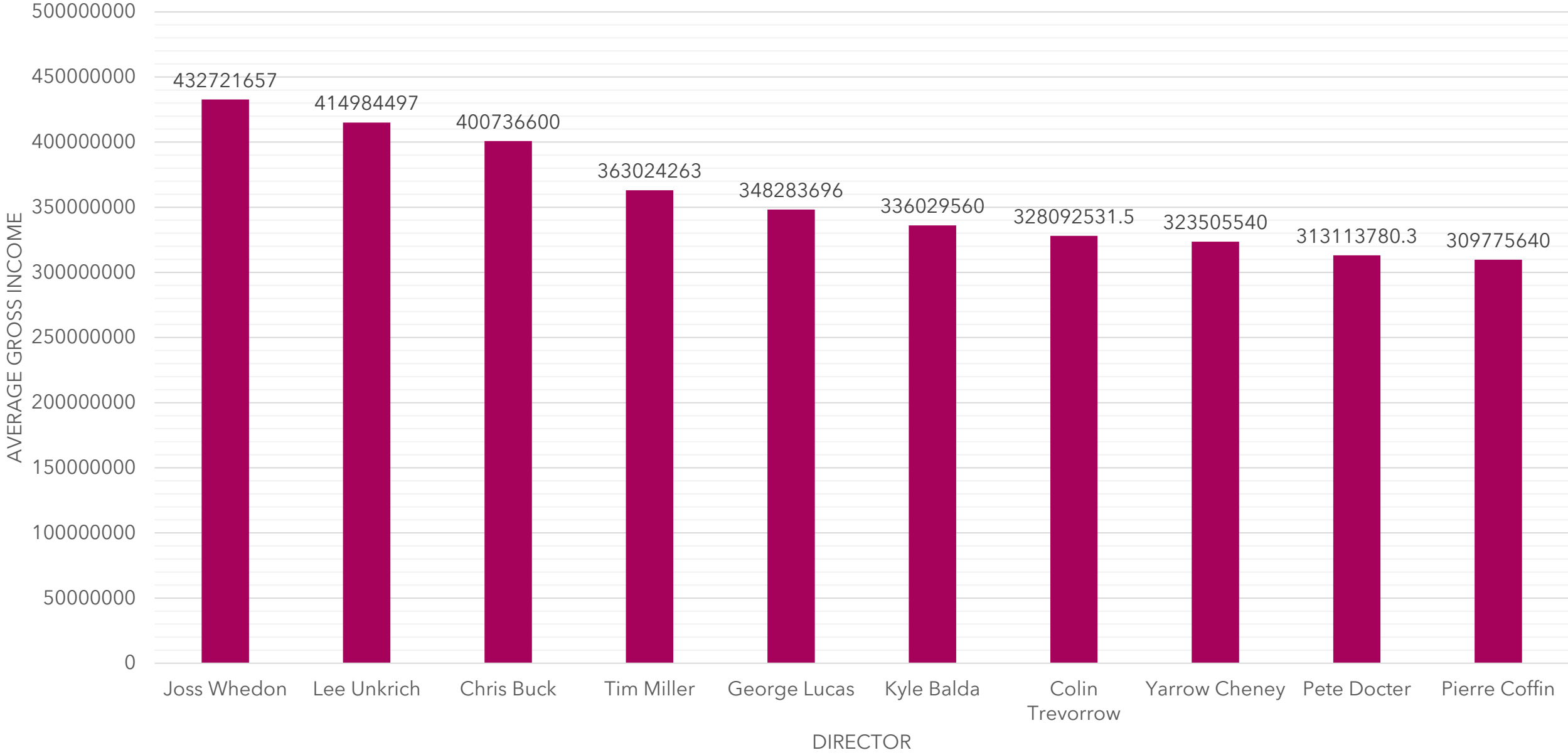
Analysis of Variance

Source	DF	Sum of Squares	Mean Square	F Ratio	Prob > F
Director	1864	1.1309e+19	6.067e+15	1.6953	<.0001*
Error	2269	8.1203e+18	3.579e+15		
C. Total	4133	1.9429e+19			

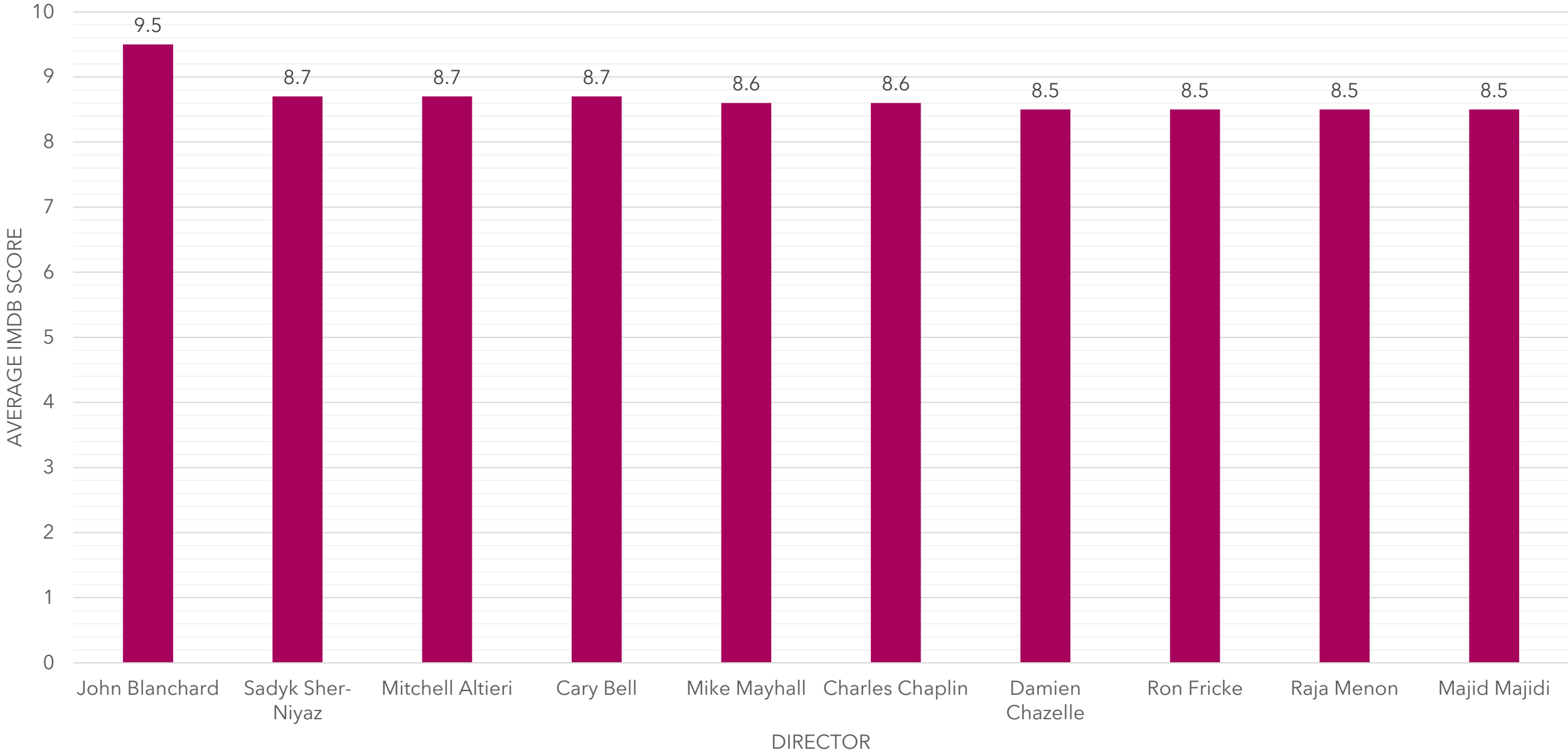
Top 25 Directors by Mean Gross Income



Top 10 Directors by Average Gross Income



Top 10 Directors by Average IMDb Score

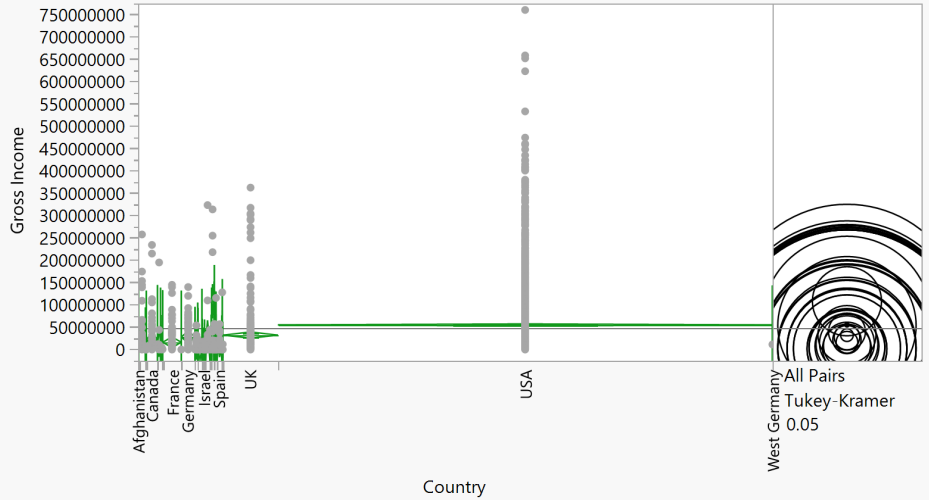


HYPOTHESIS 5

Countries have different levels of mean **gross income**

- ANOVA + Tukey

Oneway Analysis of Gross Income By Country



Oneway Anova

Summary of Fit

Rsquare	0.044397
Adj Rsquare	0.033407
Root Mean Square Error	67404939
Mean of Response	48531409
Observations (or Sum Wgts)	4135

Analysis of Variance

Source	DF	Sum of Squares	Mean Square	F Ratio	Prob > F
Country	47	8.627e+17	1.836e+16	4.0400	<.0001*
Error	4087	1.8569e+19	4.543e+15		
C. Total	4134	1.9432e+19			

Connecting Letters Report

Level					Mean
Chile	A	B	C	D	12188642
Hong Kong	A	B	C	D	12011449
Hungary	A	B	C	D	11687596
West Germany	A	B	C	D	11433134
Aruba	A	B	C	D	10076136
Romania	A	B	C	D	9100987
Spain			C	D	8836518
Mexico	A	B	C	D	7472590
Argentina	A	B	C	D	7230936
Ireland	A	B	C	D	7033630
New Line	A	B	C	D	6712451
Colombia	A	B	C	D	6517198
Indonesia	A	B	C	D	4105123
Russia	A	B	C	D	3739952
Italy	A	B	C	D	3179708
Brazil	A	B	C	D	2712574
Thailand	A	B	C	D	2602325
Iran	A	B	C	D	2424419
Poland	A	B	C	D	2402682
South Korea	A	B	C	D	2235213
India			C	D	2117136
Greece	A	B	C	D	1908996
Netherlands	A	B	C	D	1884888
Denmark	A	B	C	D	1418469
Czech Republic	A	B	C	D	1294772
Israel	A	B	C	D	1220488
Afghanistan	A	B	C	D	1127331
Belgium	A	B	C	D	680567
Finland	A	B	C	D	611709
Norway	A	B	C	D	451137
Sweden	A	B	C	D	126812
Philippines	A	B	C	D	70071
Cameroon	A	B	C	D	32631
Georgia	A	B	C	D	17149
Iceland	A	B	C	D	15897

Levels not connected by same letter are significantly different.

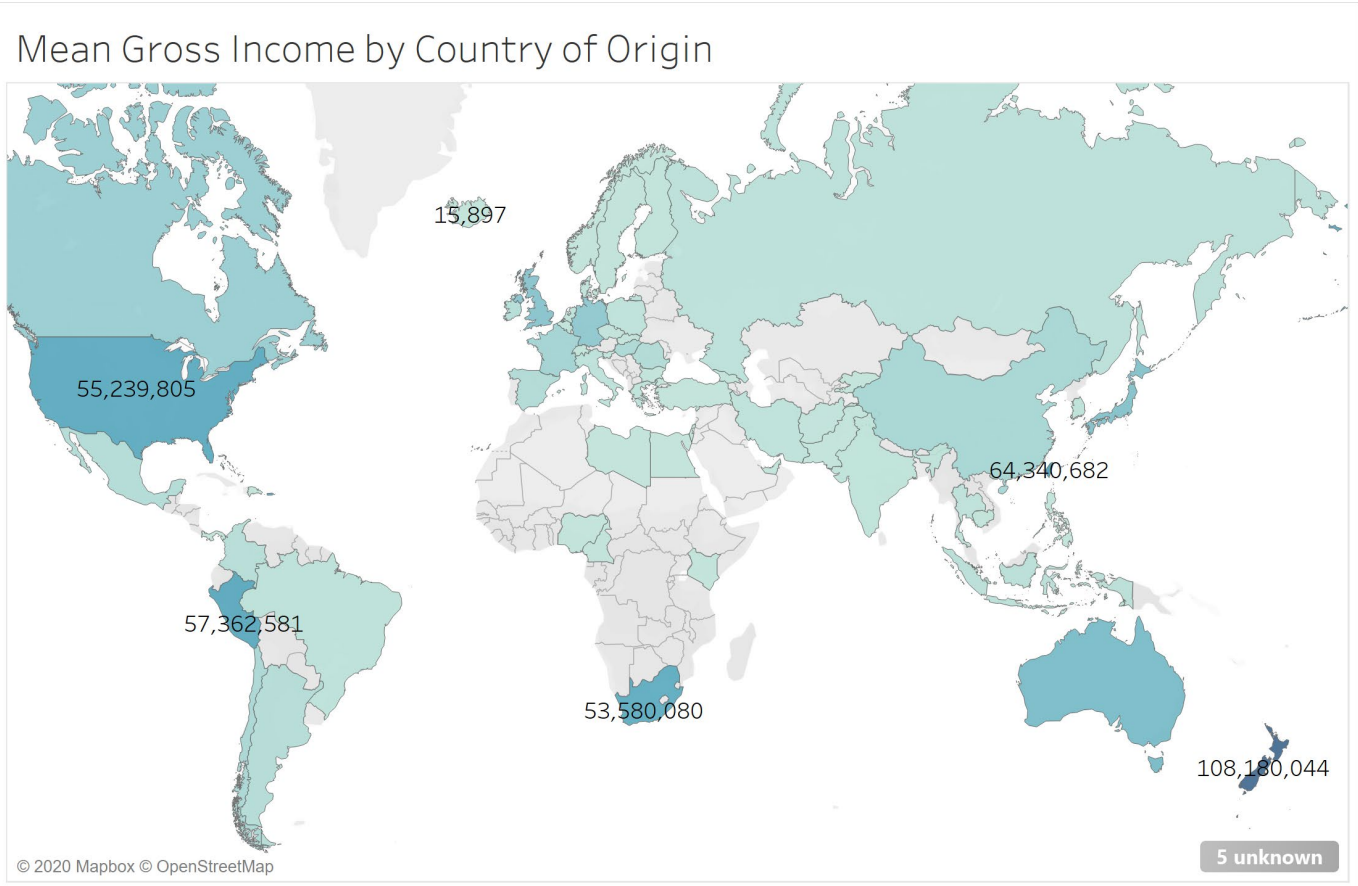
MEAN GROSS INCOME BY COUNTRY OF ORIGIN

Top countries:

1) New Zealand ~ $108 million
2) Taiwan ~ $64 million
3) Peru ~ $57 million
4) U.S.A. ~ $55 million
5) South Africa ~ $54 million

Means for Oneway Anova

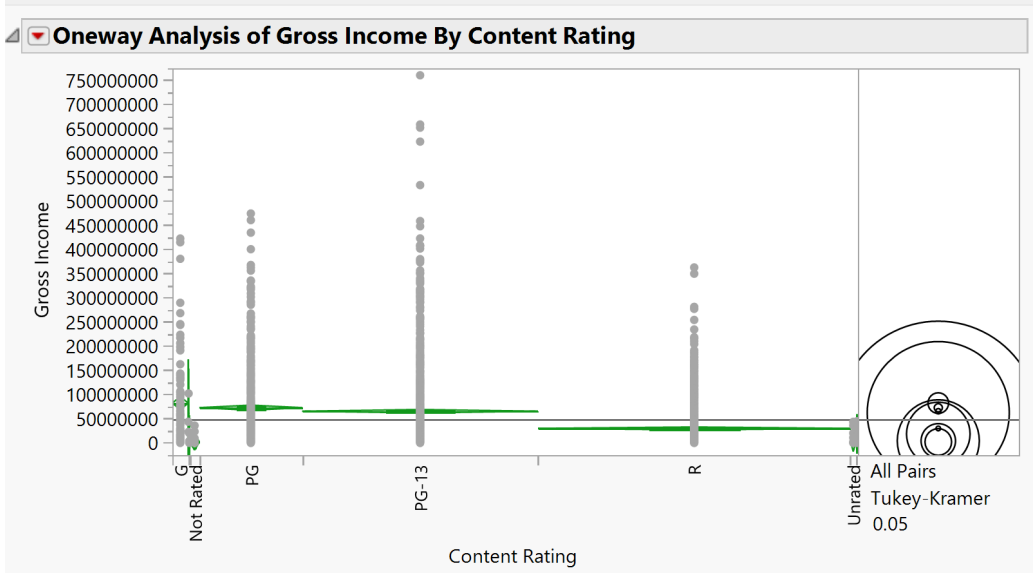
Level	Number	Mean	Std Error	Lower 95%	Upper 95%
New Zealand	12	108180044	19458130	70031512	146328575
Taiwan	2	64340682	47662489	-29103754	157785118
Peru	1	57362581	67404939	-74787808	189512970
USA	3222	55239805	1187487.2	52911683	57567926
South Africa	3	53580080	38916260	-22716982	129877143



HYPOTHESIS 6

Different **content ratings** will have different mean **gross income** and G-rated movies will have the highest mean gross income

- ANOVA + Tukey



Connecting Letters Report

Level			Mean
G	A		82455164
PG	A		72952731
PG-13	A		65648540
M	A	B	62554450
GP	A	B	43800000
R		B	29998625
X	A	B	18658812
NC-17	A	B	4476870
Unrated		B	4302599
Not Rated		B	2090037

Levels not connected by same letter are significantly different.

Oneway Anova

Summary of Fit

Rsquare	0.089668
Adj Rsquare	0.087651
Root Mean Square Error	65748940
Mean of Response	49277952
Observations (or Sum Wgts)	4071

Analysis of Variance

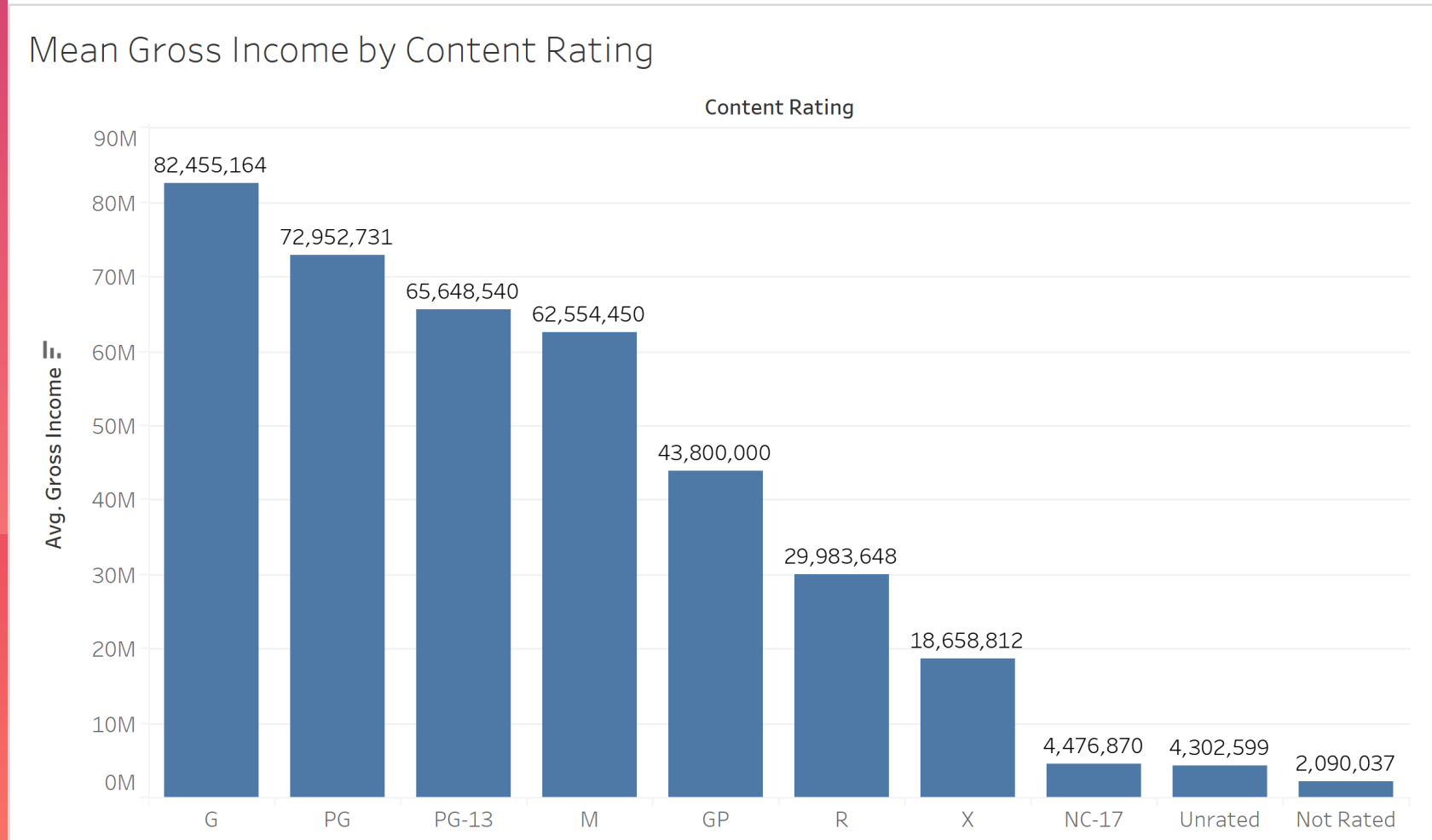
Source	DF	Sum of Squares	Mean Square	F Ratio	Prob > F
Content Rating	9	1.7292e+18	1.921e+17	44.4456	<.0001*
Error	4061	1.7555e+19	4.323e+15		
C. Total	4070	1.9285e+19			

Means for Oneway Anova

Level	Number	Mean	Std Error	Lower 95%	Upper 95%
G	95	82455164	6745698.9	69229896	95680433
GP	1	43800000	65748940	-85103974	172703974
M	2	62554450	46491521	-28594424	153703324
NC-17	6	4476870.3	26841892	-48147956	57101697
Not Rated	57	2090037.5	8708659.8	-14983711	19163786
PG	611	72952731	2659917.4	67737835	78167628
PG-13	1400	65648540	1757214.3	62203436	69093644
R	1855	29998625	1526570.3	27005710	32991539
Unrated	34	4302599.1	11275850	-17804250	26409448
X	10	18658812	20791640	-22104204	59421827

Std Error uses a pooled estimate of error variance

MEAN GROSS INCOME BY CONTENT RATING



GROSS INCOME PREDICTION FORMULA

- **Most significant relationships** with gross income:
  1) # of critics that reviewed the movie
  2) # of users that reviewed the movie
  3) Genre(s)
- **R-square adjusted** (coefficient of determination): ~ **0.612**

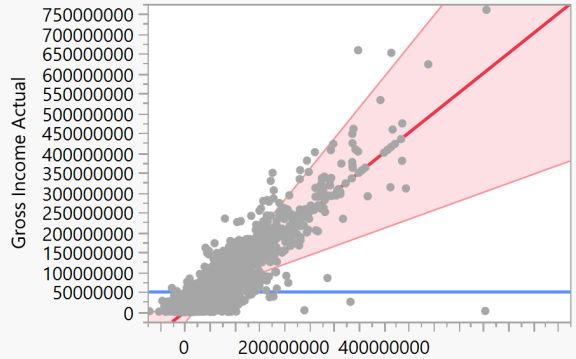
Summary of Fit

RSquare	0.846579
RSquare Adj	0.612166
Root Mean Square Error	43849754
Mean of Response	52523076
Observations (or Sum Wgts)	3760

Analysis of Variance

Source	DF	Sum of Squares	Mean Square	F Ratio
Model	2272	1.5777e+19	6.944e+15	3.6115
Error	1487	2.8592e+18	1.923e+15	Prob > F
C. Total	3759	1.8636e+19		<.0001*

```
-110519881.4
+ 150205.67065 • # Critics that Reviewed
+ 49921.923706 • # users that reviewed the movie
+ -19886718.48 • Aspect Ratio
+ 0.0558532351 • Budget
+ 505449.40547 • Duration (min)
+ 11384574.871 • IMDb Score
+ Match(Color, " Black and White" ⇒ -6844279.355,
                "Color" ⇒ 6844279.3547,
                else ⇒ .)
+ Match(Content Rating, "G" ⇒ 27591364.625,
                        "GP" ⇒ 47626157.958,
                        "M" ⇒ -11889305.45,
                        "NC-17" ⇒ -21396994.72,
                        "Not Rated" ⇒ -10360893.52,
                        "PG" ⇒ 15515009.112,
                        "PG-13" ⇒ 10241656.297,
                        "R" ⇒ -10330005.81)
```



Effect Summary

Source	LogWorth	PValue
# users that reviewed the movie	37.346	0.00000
# Critics that Reviewed	24.850	0.00000
Genres	17.486	0.00000
IMDb Score	10.018	0.00000
Duration (min)	9.805	0.00000
Content Rating	7.702	0.00000
Budget	5.100	0.00001
Director	4.755	0.00002
Aspect Ratio	3.322	0.00048
Color	1.329	0.04687

FINDINGS

- **# of customers that reviewed** has a significant relationship with gross income
- **Facebook engagements** are positively correlated with gross income
- **Budget** is positively correlated with gross income
- **Best countries to produce in:** New Zealand, Taiwan, Peru, U.S.A., South Africa
- **Top directors:**
  - **By mean gross income:** Joss Whedon, Lee Unkrich, Chris Buck
  - **By average IMDb score:** John Blanchard, Sadyk Sher-Niyaz, Mitchell Altieri
- **Best content ratings:** G, PG, PG-13
  - As the expected age of the viewer increases, gross income decreases
- Most movies are produced in the United States

RECOMMENDATIONS

1) **Get the audience to review** the movie on IMDb
2) **Set up Facebook pages** for the movie, cast, and director
3) **Use Facebook** to promote the movie
4) **Increase the production budget**
5) Produce movies in **New Zealand, Taiwan, Peru, U.S.A., or South Africa**
    a) If you want access to a robust movie supply chain, produce movies in the United States
6) Hire **Joss Whedon** (top gross income), **John Blanchard** (top IMDb score), Lee Unkrich, or Chris Buck as a director
7) Produce movies for the **whole family** – rated G, PG, & PG-13
8) Make **Family | Sci-Fi** movies

QUESTIONS?

About our data visualization…

About our process…

About our findings…

About our recommendations…