

PREDICTING CULT MOVIE PHENOMENON

OR... USING M*#\$^& F#&%*\$ LINEAR REGRESSION TO PREDICT M#*\$&%# F%#&%# CULT MOVIES!!!



All the *Cult* Movies... Please Stand Up!

- * Definition Subject to a Lot of Debate in Movie Literature
 - * Inclusive Definition: “allows for **major studio** productions, especially **box office bombs**”
 - * Exclusive Definition: “focus more on **obscure, transgressive films shunned by the mainstream**”
- * Trace their origin back to **controversial** and **suppressed** films kept **alive** by **dedicated** fans.
- * Over time, the definition has become more vague and *inclusive* as it drifts away from earlier, stricter views.

T H E
BIG LEBOWSKI



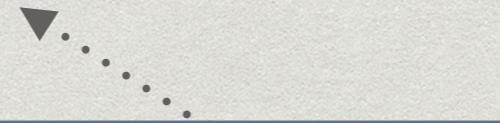
Duuude, Regress or Classify?

- * Began as a Classification Problem
 - * Sparse and/or Contentious Labeling of *Cult* Movies on the inter-web
 - * Wanted to Apply Linear Regression Techniques
- * But Linear Regression Requires a *Continuous* Target Variable!

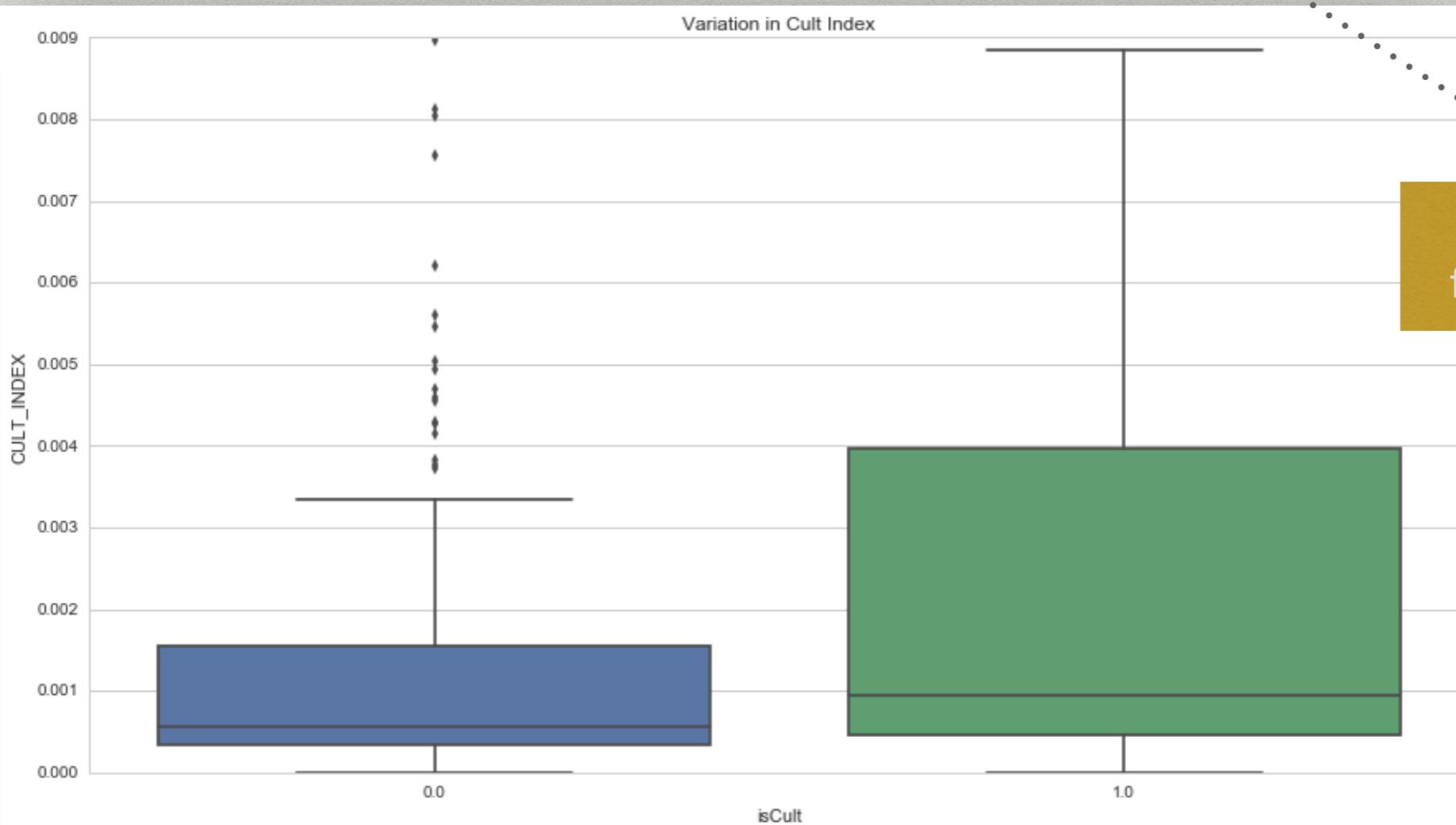
I Believe You Have My... *Cult-Index!*

Adj. Lifetime Gross Revenue * log(Δ Years Since Release)

Δ Years Since Release



Adj. Opening Weekend Revenue * No. of Theaters



rewards older movies
for consistent revenue rate

DATA WRANGLING & EXPLORATION

*"COME ON YOU
SCUZZY DATA... BE IN
THERE!"*

A world inside
the computer
where man
has never been.

Never before now.



TRON

TRON A LISBERGER-KUSHNER PRODUCTION
STARRING JEFF BRIDGES BRUCE BOXLEITNER DAVID WARNER CINDY MORGAN AND BARNARD HUGHES
EXECUTIVE PRODUCER RON MILLER MUSIC BY WENDY CARLOS STORY BY STEVEN LISBERGER AND BONNIE MACBIRD
SCREENPLAY BY STEVEN LISBERGER PRODUCED BY DONALD KUSHNER DIRECTED BY STEVEN LISBERGER
SONGS BY JOURNEY FROM WALT DISNEY PRODUCTIONS TECHNICOLOR®
DOLBY STEREO® Read the Ballantine Book Original motion picture sound track available, CBS Records and Tapes
IN SELECTED THEATRES FILMED IN SUPER PANAVISION® 70
Released by BUENA VISTA DISTRIBUTION CO., INC. © 1982 Walt Disney Productions PG PARENTAL GUIDANCE SUGGESTED SOME MATERIAL MAY NOT BE SUITABLE FOR CHILDREN

Entertainment Finance Macys Philosophy Recoshin Code Repos T.I.M. MyEd Student and S Your Courses | C

Box Office Mojo

Search Site

Search...

Social

Facebook Twitter

Features

News Release Sched. Showtimes at IMDB

Box Office

Summary Weekend Weekly Similar Movies

Total Lifetime Grosses

Domestic: \$7,993,039

Domestic Summary

Opening Weekend: \$918,127 (#14 rank 183 theaters, \$5,017 average) % of Total Gross: 11.5% > View All 3 Weekends Widest Release: 214 theaters

The Players

Director: Richard Linklater
Actors: Matthew McConaughey, Milla Jovovich, Ben Affleck, Parker Posey

Related Products

- Dazed and Confused (Flashback Edition)
- Dazed and Confused - Criterion Collection

Genres

Genre	Rank
Comedy - High School	54

https://en.wikipedia.org/wiki/Category:Lists_of_American_films_by_year

Pages in category "Lists of American films by year"

The following 128 pages are in this category, out of 128 total. This list may not reflect recent changes (learn more).

- Lists of American films
- 1
 - List of American films of the 1890s
 - American films of 1891
 - American films of 1892
 - American films of 1893
 - American films of 1894
 - American films of 1895
 - American films of 1896
 - American films of 1897
 - American films of 1898
 - American films of 1899
 - List of American films of 1900
 - List of American films of 1901
 - List of American films of 1902
 - List of American films of 1903
 - List of American films of 1904
 - List of American films of 1905
 - List of American films of 1906
 - List of American films of 1907
 - List of American films of 1908
 - List of American films of 1909
 - List of American films of 1910
 - List of American films of 1911
 - List of American films of 1912
 - List of American films of 1913
 - List of American films of 1914
- 2
 - List of American films of 1915
 - List of American films of 1916
 - List of American films of 1917
 - List of American films of 1918
 - List of American films of 1919
 - List of American films of 1920
 - List of American films of 1921
 - List of American films of 1922
 - List of American films of 1923
 - List of American films of 1924
 - List of American films of 1925
 - List of American films of 1926
 - List of American films of 1927
 - List of American films of 1928
 - List of American films of 1929
 - List of American films of 1930
 - List of American films of 1931
 - List of American films of 1932
 - List of American films of 1933
 - List of American films of 1934
 - List of American films of 1935
 - List of American films of 1936
 - List of American films of 1937
 - List of American films of 1938
 - List of American films of 1939
 - List of American films of 1940
 - List of American films of 1941
 - List of American films of 1942
 - List of American films of 1943
 - List of American films of 1944
 - List of American films of 1945
 - List of American films of 1946
 - List of American films of 1947
 - List of American films of 1948
 - List of American films of 1949
 - List of American films of 1950
 - List of American films of 1951
 - List of American films of 1952
 - List of American films of 1953
 - List of American films of 1954
 - List of American films of 1955
 - List of American films of 1956
 - List of American films of 1957
 - List of American films of 1958
 - List of American films of 1959
 - List of American films of 1960
 - List of American films of 1961
 - List of American films of 1962
 - List of American films of 1963
 - List of American films of 1964
 - List of American films of 1965
 - List of American films of 1966
 - List of American films of 1967
 - List of American films of 1968
 - List of American films of 1969
 - List of American films of 1970
 - List of American films of 1971
 - List of American films of 1972
 - List of American films of 1973
 - List of American films of 1974
 - List of American films of 1975
 - List of American films of 1976
 - List of American films of 1977
 - List of American films of 1978
 - List of American films of 1979
 - List of American films of 1980
 - List of American films of 1981
 - List of American films of 1982
 - List of American films of 1983
 - List of American films of 1984
 - List of American films of 1985
 - List of American films of 1986
 - List of American films of 1987
 - List of American films of 1988
 - List of American films of 1989
 - List of American films of 1990
 - List of American films of 1991
 - List of American films of 1992
 - List of American films of 1993
 - List of American films of 1994
 - List of American films of 1995
 - List of American films of 1996
 - List of American films of 1997
 - List of American films of 1998
 - List of American films of 1999
 - List of American films of 2000
 - List of American films of 2001
 - List of American films of 2002
 - List of American films of 2003

https://en.wikipedia.org/wiki/Category:List_of_American_films

A-B [edit]

Title	Director	Cast	Genre	Note
2010	Peter Hyams	Roy Scheider, Helen Mirren, John Lithgow, Bob Balaban, Dana Elcar	Science fiction	sequel to 2001: A Space Odyssey
The Adventures of Buckaroo Banzai	W. D. Richter	Peter Weller, Ellen Barkin, John Lithgow, Jeff Goldblum	Science fiction	20th Century Fox
Against All Odds	Taylor Hackford	Jeff Bridges, Rachel Ward, James Woods, Richard Widmark, Dorian Harewood, Alex Karras, Jane Greer	Drama	remake of Out of the Past
All of Me	Carl Reiner	Steve Martin, Lily Tomlin, Dana Elcar	Comedy	Columbia
Alley Cat	Victor M. Ordóñez	Karin Mani, Robert Torti	Action	
Amadeus	Miloš Forman	Tom Hulce, F. Murray Abraham, Elizabeth Berridge, Jeffrey Jones	Biography	bio of Wolfgang Amadeus Mozart; won 8 Oscars
The Ambassador	J. Lee Thompson	Robert Mitchum, Ellen Burstyn, Donald Pleasence, Rock Hudson	Thriller	Cannon Group
American Dreamer	Rick Rosenthal	Tom Conti, JoBeth Williams, Giancarlo Giannini	Comedy	Warner Bros.
Angel	Robert Vincent O'Neill	Donna Wilkes, Dick Shawn, Susan Tyrrell	Action	New World Pictures
Bachelor Party	Neal Israel	Tom Hanks, Tawny Kitaen, Adrian Zmed	Comedy	20th Century Fox
Bad Manners	Bobby Houston	Martin Mull, Karen Black	Comedy	New World
Beat Street	Stan Lathan	Rae Dawn Chong, Guy Davis	Drama, Music	
Before Stonewall	Greta Schiller		Documentary	
Best Defense	Willard Huyck	Dudley Moore, Eddie Murphy, Kate Capshaw, George Dzundza	Comedy	Paramount
Beverly Hills Cop	Martin Brest	Eddie Murphy, Judge Reinhold, John Ashton, Lisa Eilbacher, Bronson Pinchot, Ronny Cox	Comedy	1984 top-grossing film; 2 sequels
Birdy	Alan Parker	Matthew Modine, Nicolas Cage	Drama	based on William Wharton novel
Blow	Charles Roven	Michael Caine, Joseph Bologna, Michelle Pfeiffer	Comedy	Paramount

About 8500 Movies

Around 15 Initial Features

Lot of Missing Data

Production Budget ~ 40% Missing

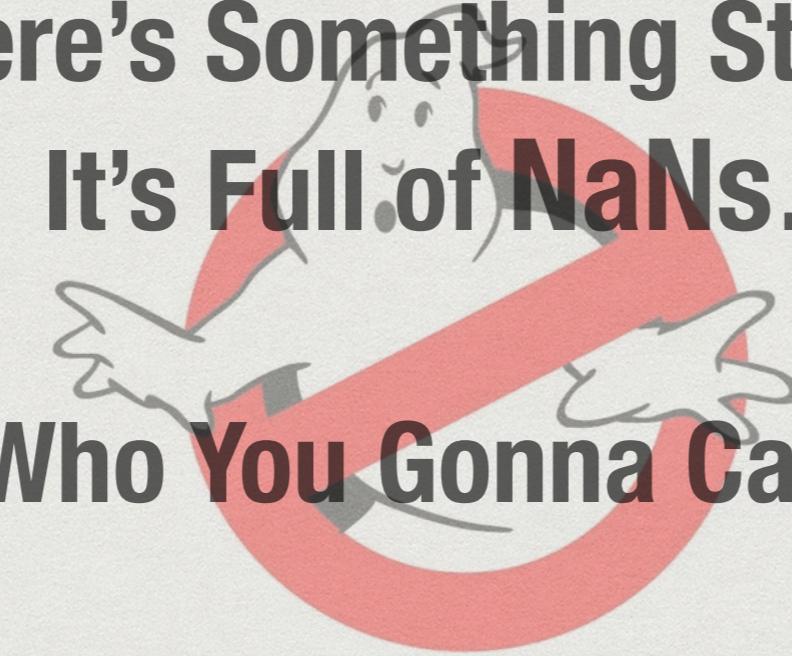
Used C.P.I. for Production Budget Adjustment

Used Bo. Mojo's Inflation Adjuster Formula for Lifetime Gross and Opening Weekend

Could Not Get to RT Ratings (User:Critic Rating Polarity) as Predictor

Could Not Get to Quote Counts as a Measure of Following

**There's Something Strange,
It's Full of NaNs...**

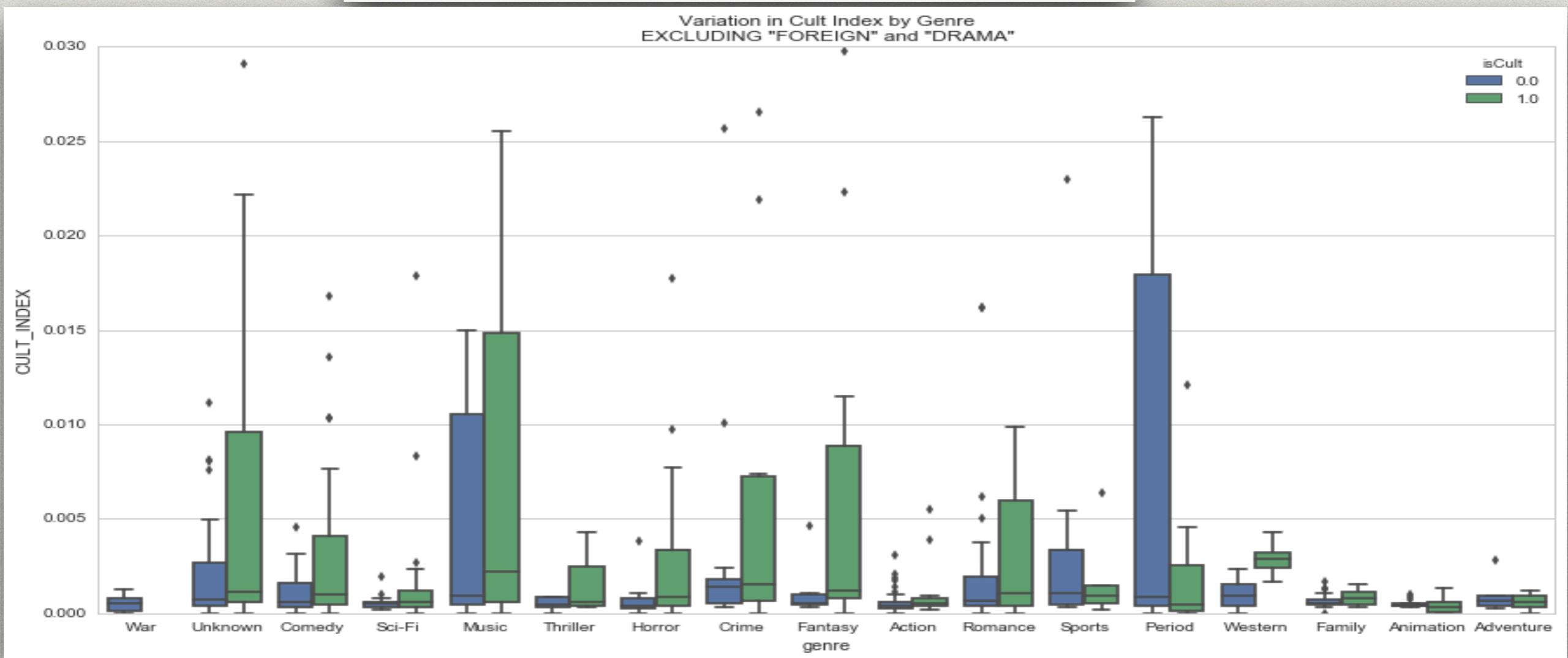
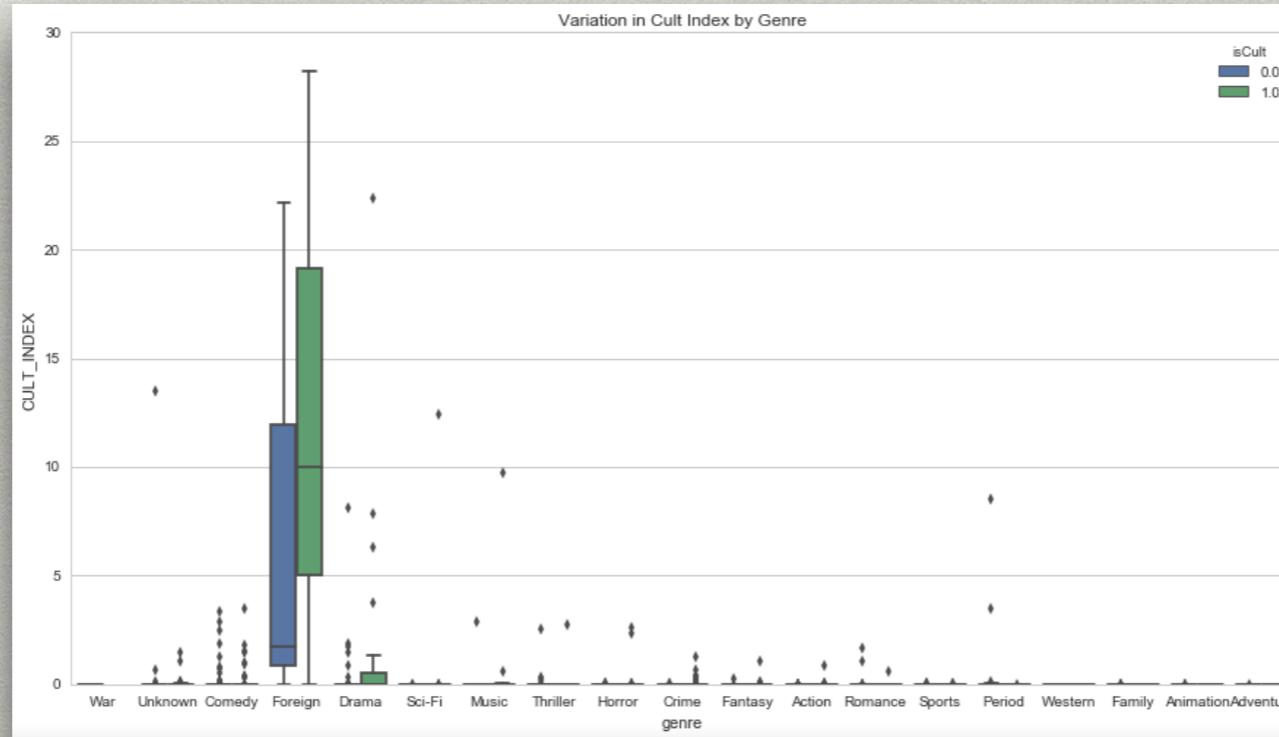


Who You Gonna Call??

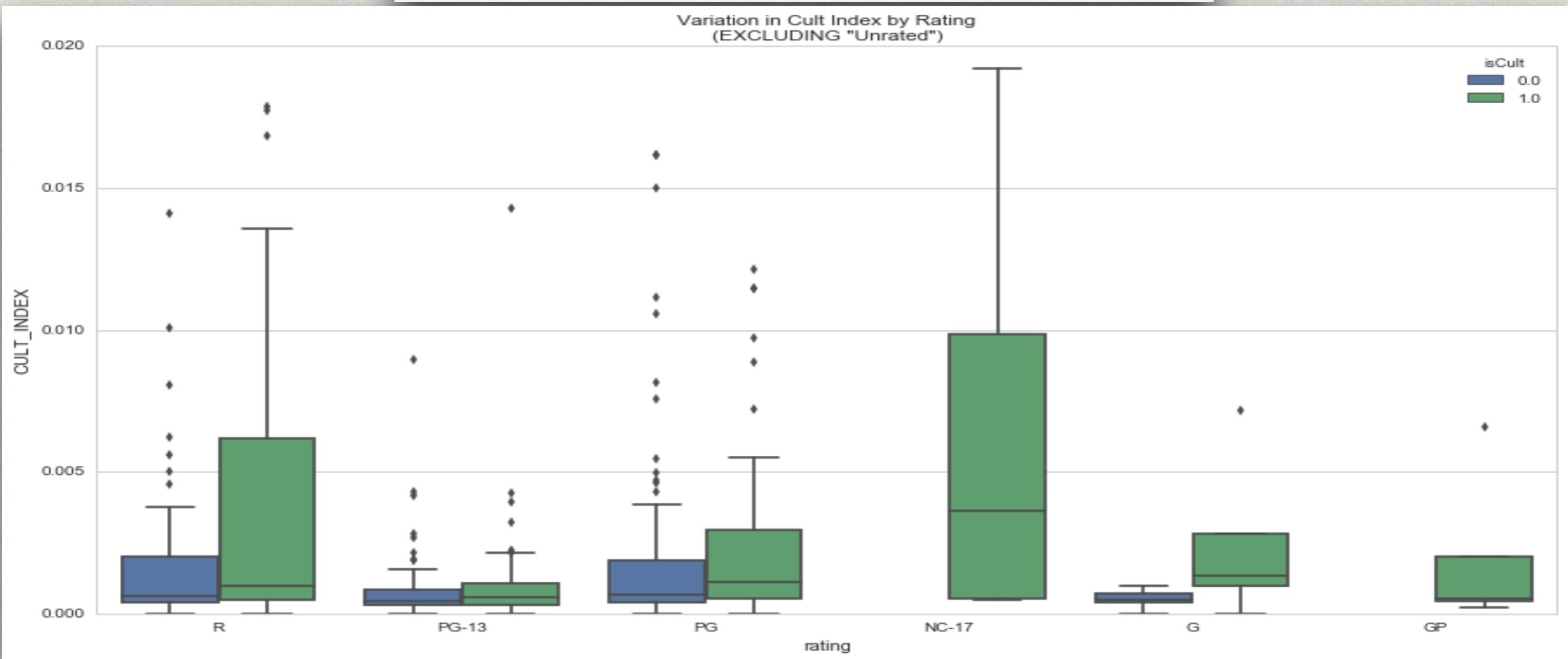
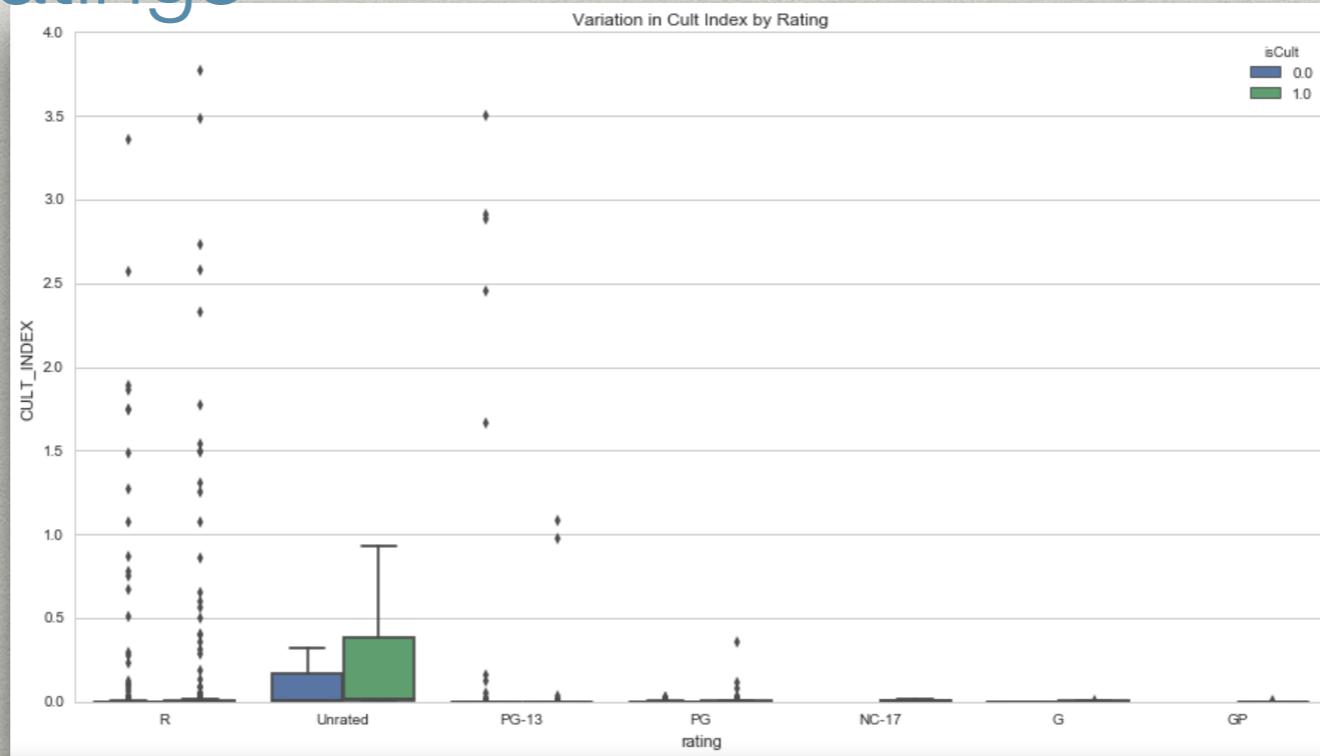
Impute Strategy for Production Budget:

1. Median Production Budget For Same Year,
Same Studio, Same Genre;
2. Median Production Budget For Same Studio, Same Genre?
3. Linear Regression:
Train on $B1 * \text{LifetimeGross} + B2 * \text{OpeningRevenue} = \text{ProductionBudgets}$;
Predict on Missing ProductionBudgets

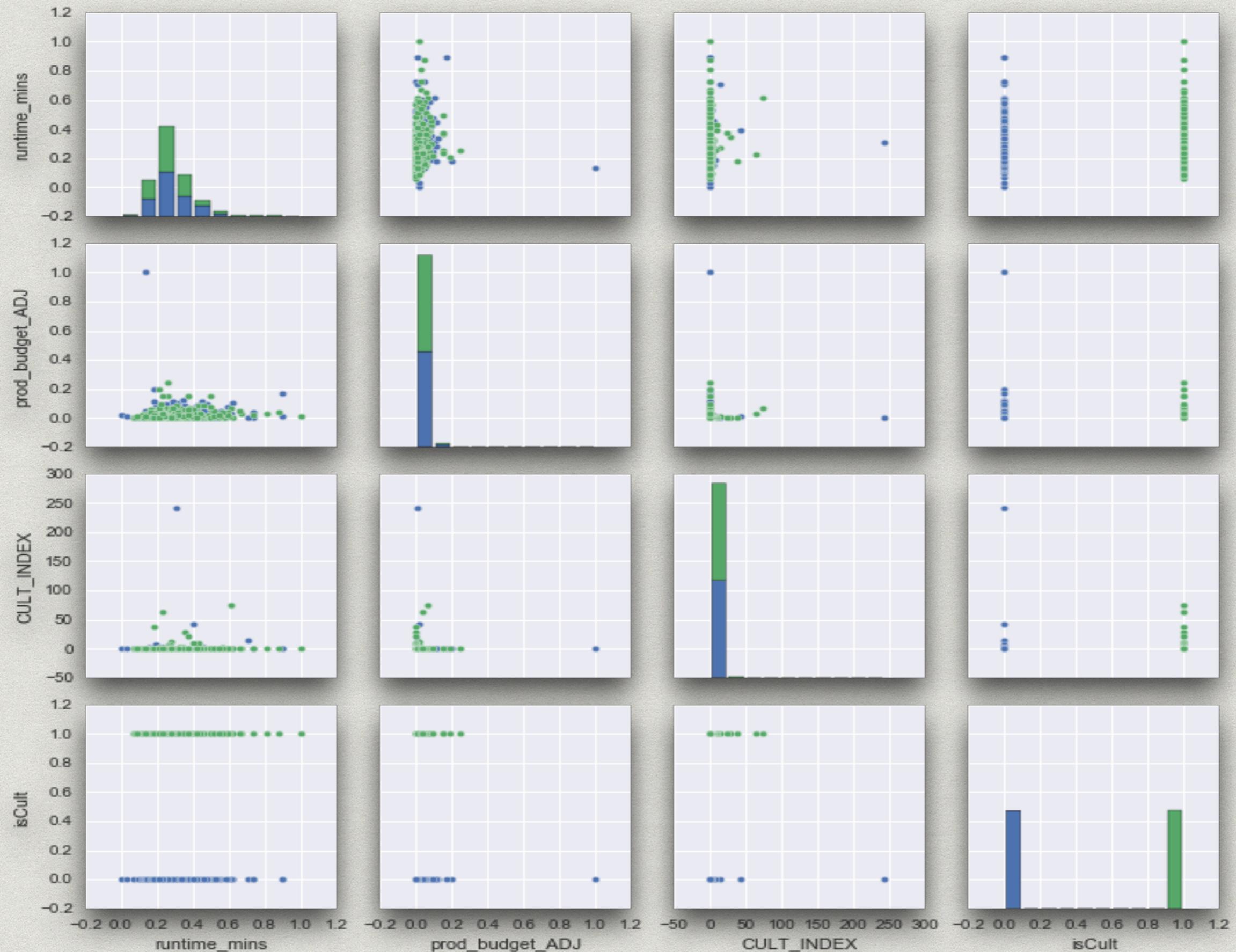
Cult Index By Genre



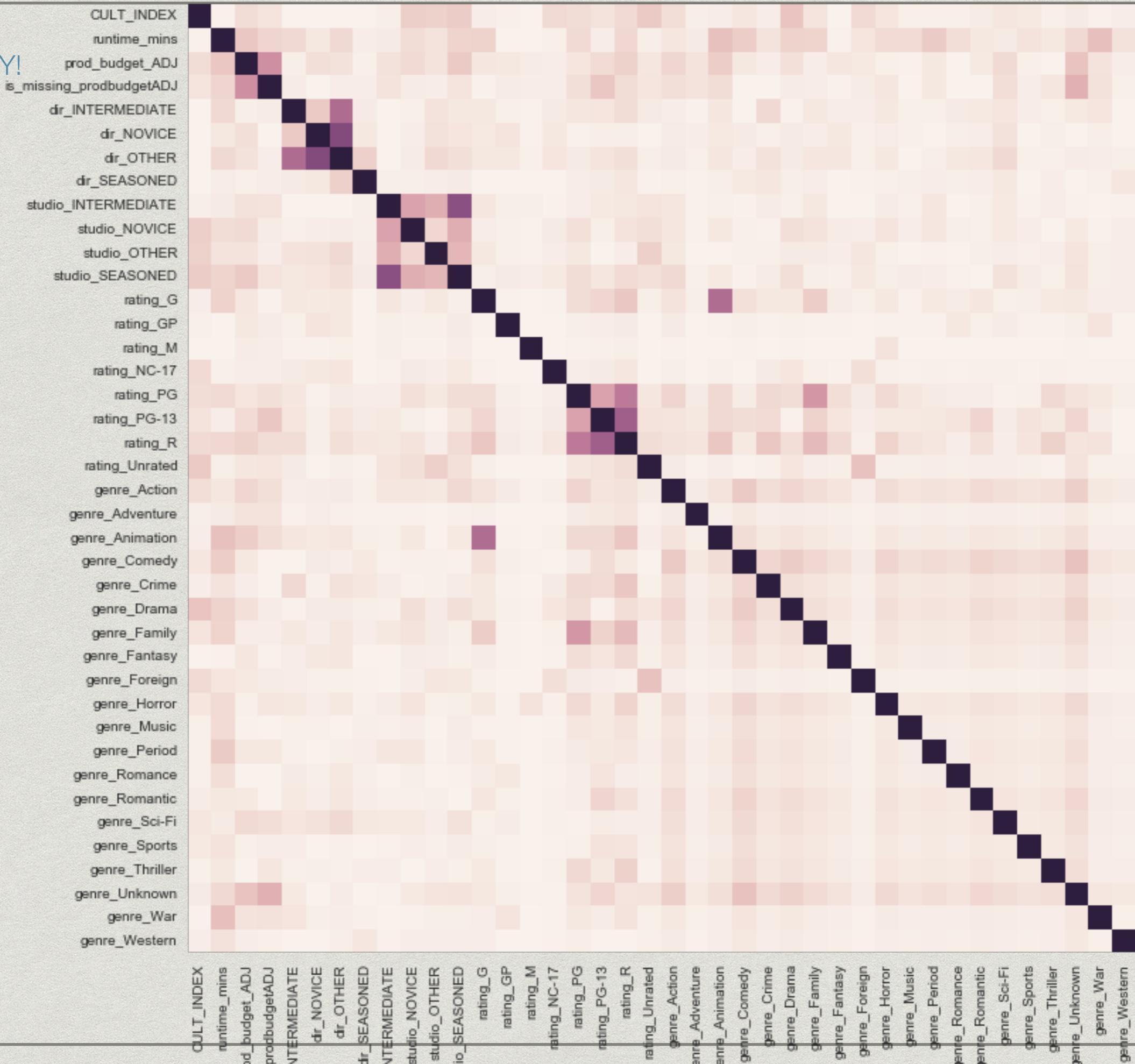
Cult Index By Ratings



Weak Numerical Predictors!

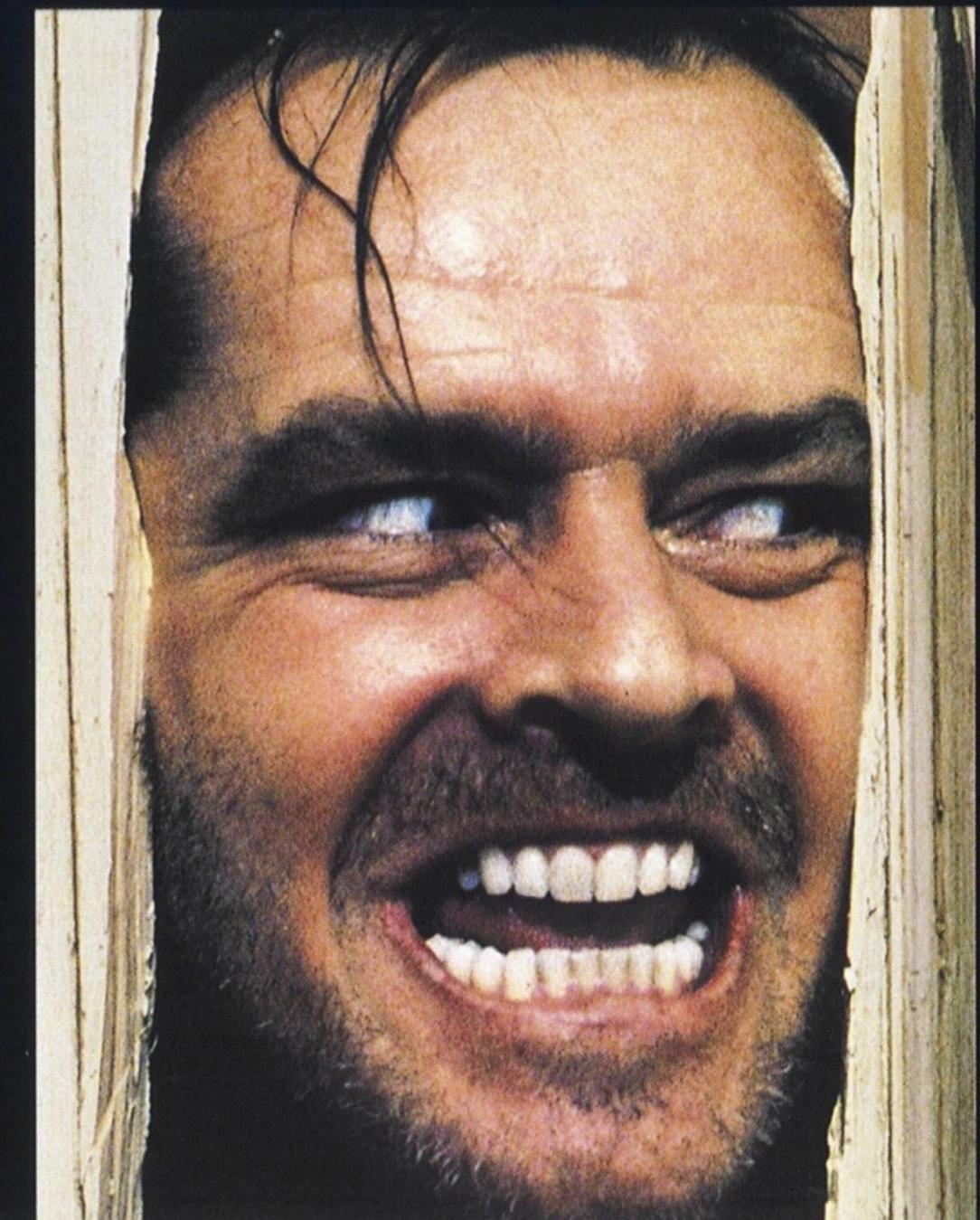


MULTI-COLLINEARITY!



!NOISSERGER

STANLEY KUBRICK'S
THE SHiNiNG



Regression Strategy

1. Model 1: Base Model, with Significant Predictors
2. Model 2: Add Interaction Effects
3. Regularize Both
4. More Outlier Analysis

5,070 data points, 39 Features incl. Interactions:

Testing the Theory that Certain Combination of Pre-Release Features Can Predict Potential Cult Phenomena

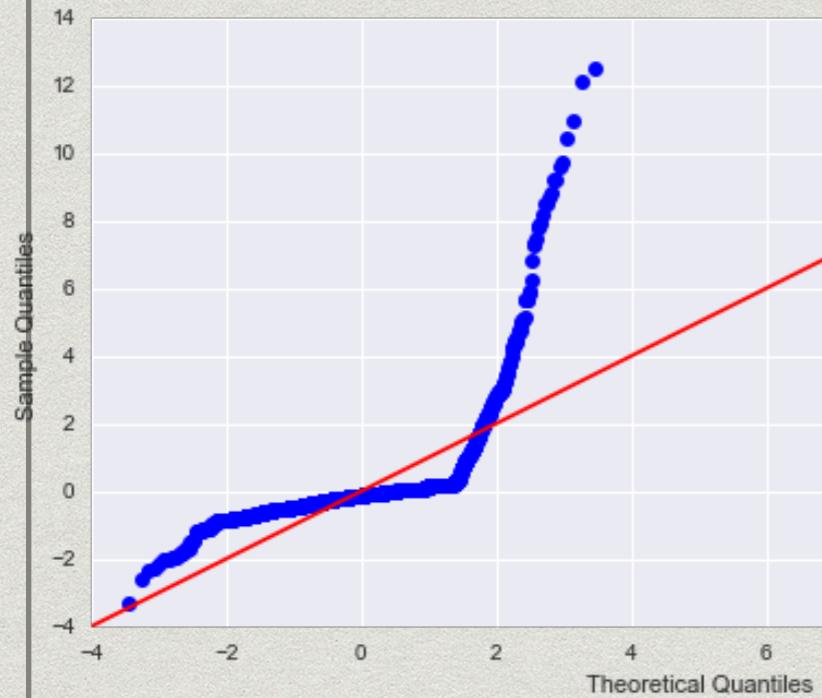
$\hat{y} = \beta_0 + \beta_1 * \text{adj_productionBudget} +$
 $\beta_2 * \text{runtime_mins} +$
 $\beta_3 * \text{is_missing_adj_productionBudget} +$
 $(\beta_4 \dots \beta_6) * (\text{director group dummies}) +$
 $(\beta_7 \dots \beta_9) * (\text{studio group dummies}) +$
 $(\beta_{10} \dots \beta_{16}) * (\text{MPAA ratings group dummies}) +$
 $(\beta_{17} \dots \beta_{30}) * (\text{genre group dummies}) +$
interactions

Regression Strategy

1. Model 1: Base, with Significant Predictors
2. Model 2: Add Interaction Effects
3. Regularize Both
4. More Outlier Analysis

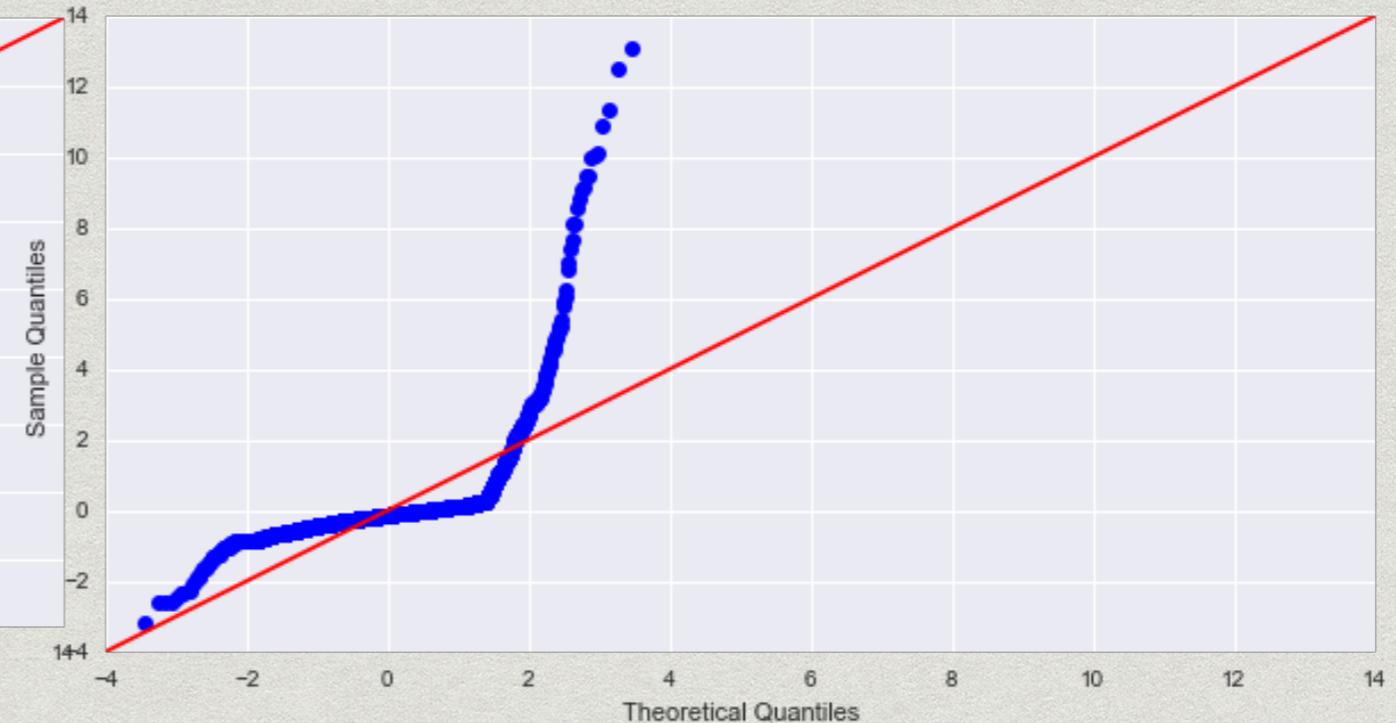
Base Model Residual Q-Q Plot

Adj. R² = 0.109

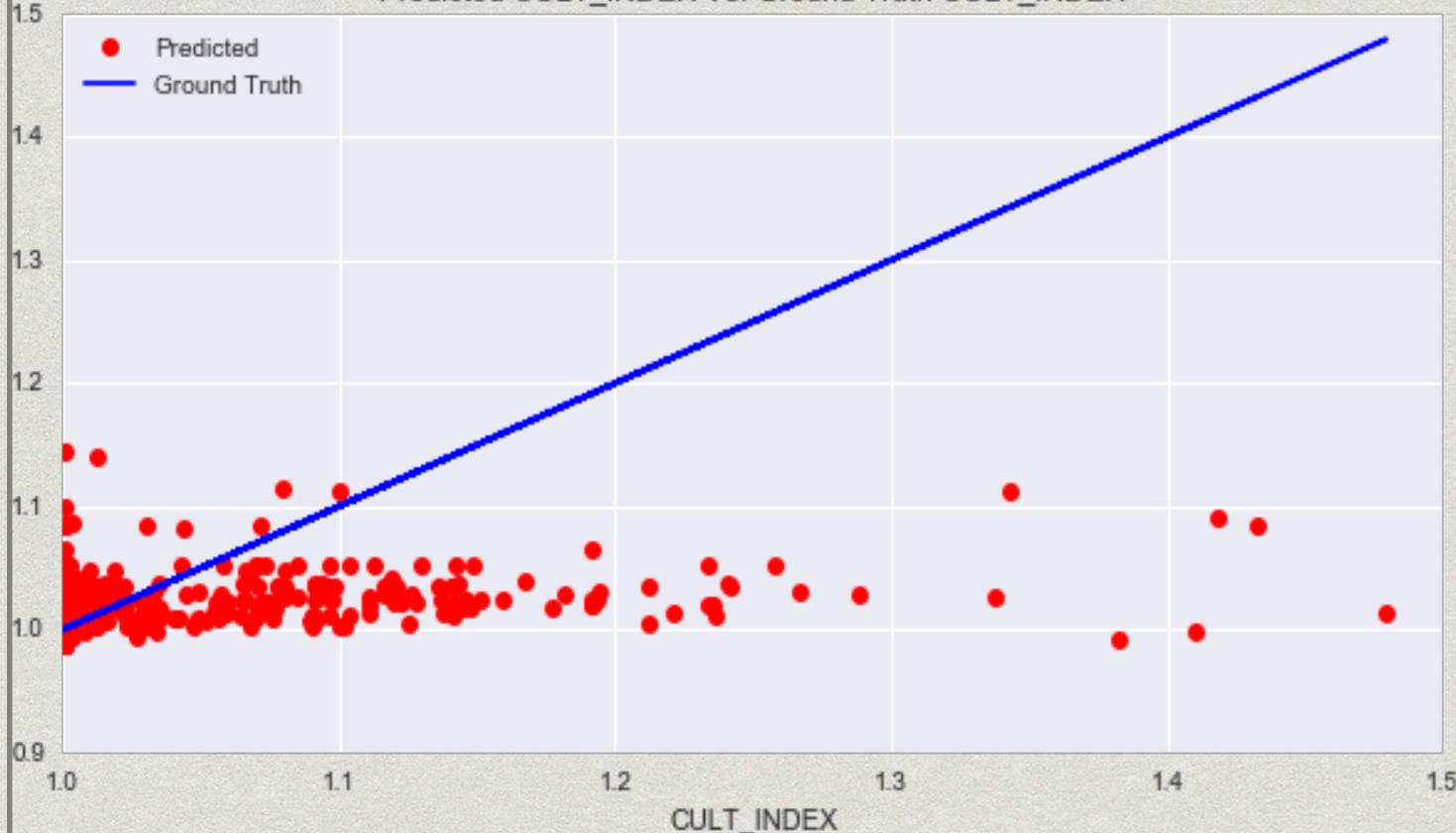


Interaction Model Residual Q-Q Plot

Adj. R² = 0.116



Predicted CULT_INDEX Vs. Ground Truth CULT_INDEX

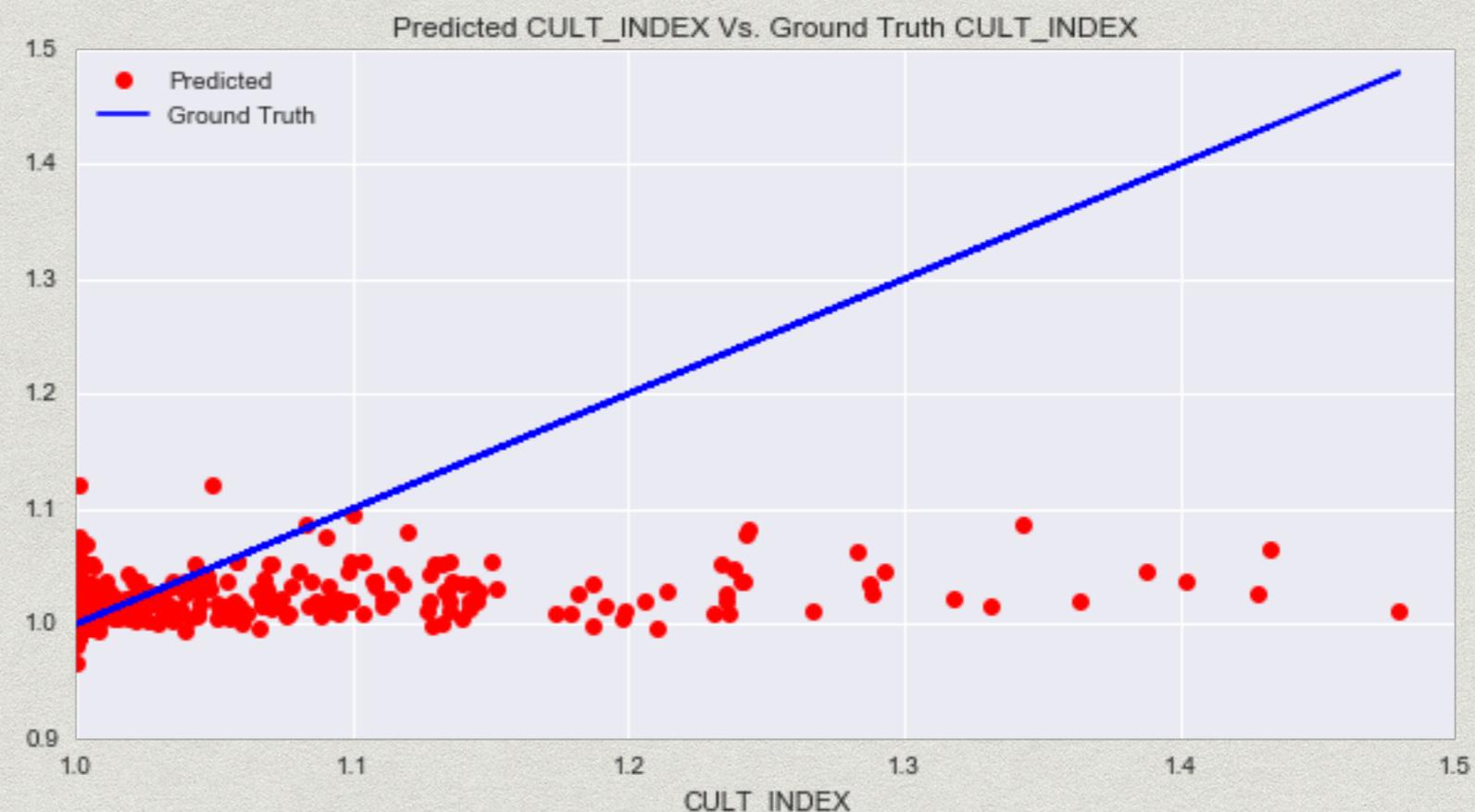


Regression Strategy

1. Model 1: Base, with Significant Predictors
2. Model 2: Add Interaction Effects
3. Regularize Both
4. More Outlier Analysis

Base Model Predicted Vs. Ground Truth

Interaction Model Predicted Vs. Ground Truth



#*(%^#% Outlier Influence!



“Hat” values help determine the “Cook’s Distance”

Cook's distance measures the effect of deleting a given observation. Points with a large Cook's distance are considered to merit closer examination in the analysis.

“Influence” and “Leverage” realize deviations from normalized residual errors.

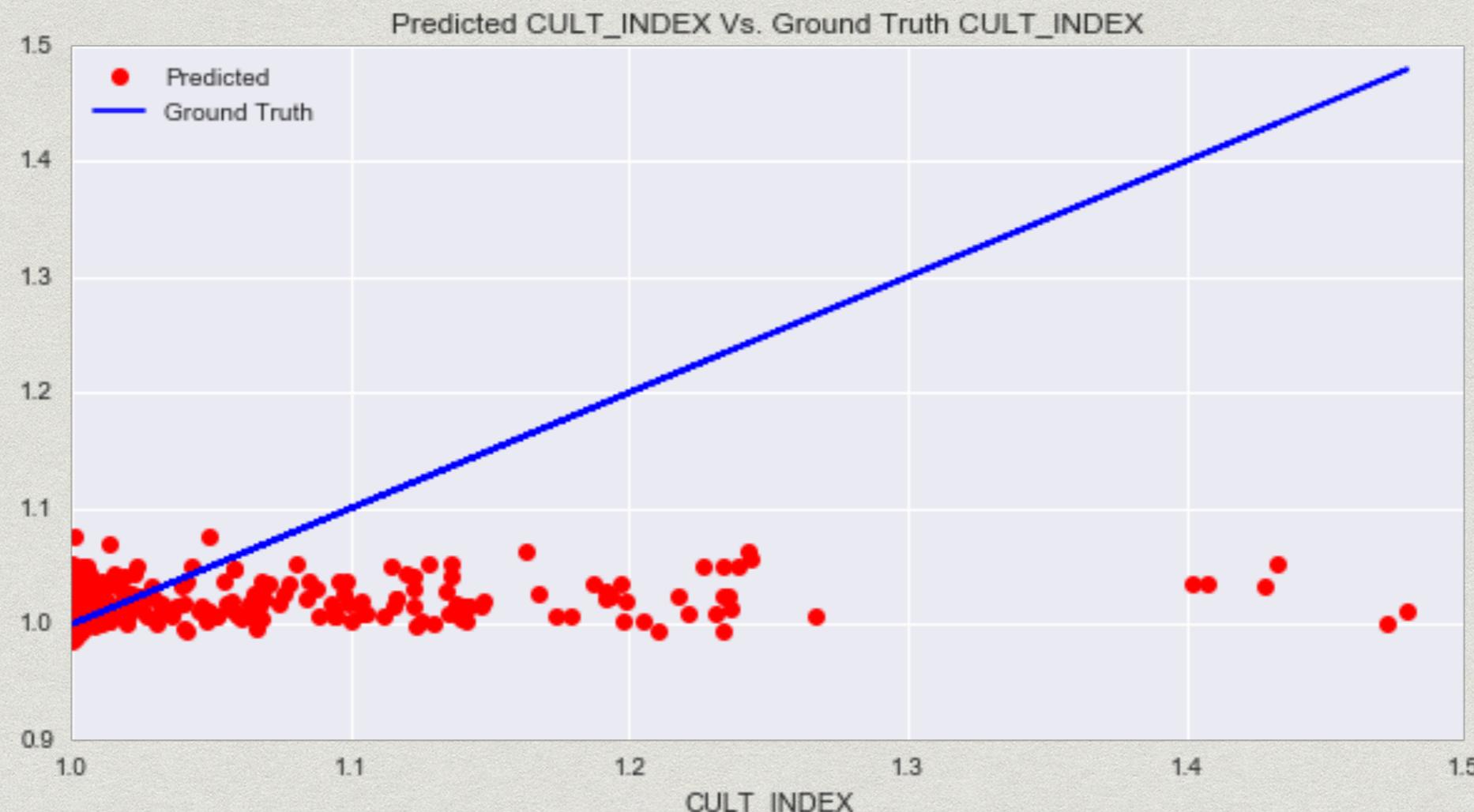
Regression Strategy

1. Model 1: Base, with Significant Predictors
2. Model 2: Add Interaction Effects
3. Regularize Both
4. More Outlier Analysis

OLS Regression Results					
Dep. Variable:	CULT_INDEX	R-squared:	0.150		
Model:	OLS	Adj. R-squared:	0.141		
Method:	Least Squares	F-statistic:	16.68		
Date:	Fri, 29 Apr 2016	Prob (F-statistic):	6.59e-97		
Time:	12:26:20	Log-Likelihood:	7029.5		
No. Observations:	3534	AIC:	-1.398e+04		
Df Residuals:	3496	BIC:	-1.375e+04		
Df Model:	37				
Covariance Type:	nonrobust				

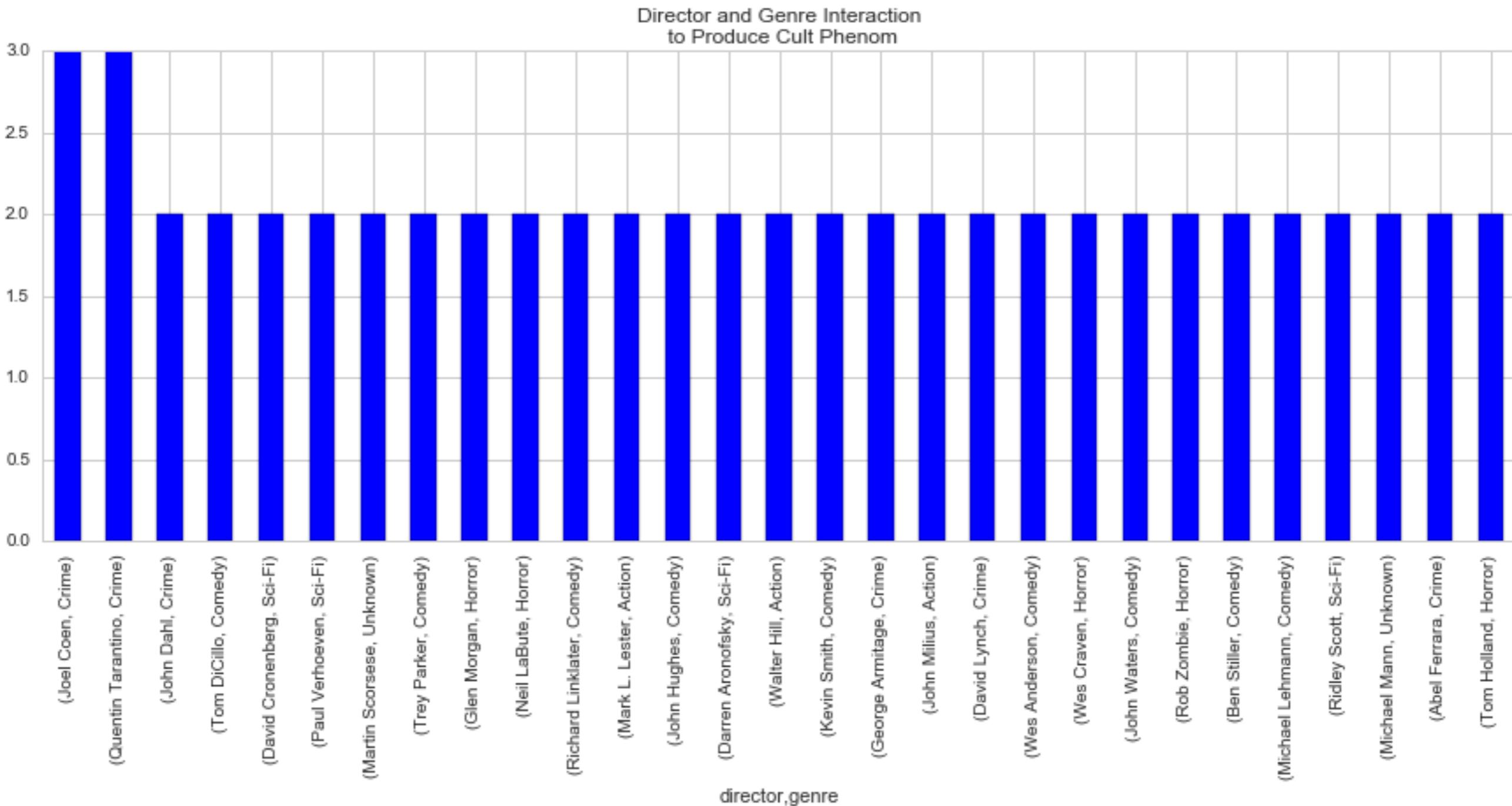
Removing Outliers Starts to Improve Adj. R²

Third Run: Interaction Model (-Some Outliers) Predicted Vs. Ground Truth



Results

- * Significant Interactions Observed Between:
 - * “Seasoned Directors” and “Intermediate” Studios, and “Novice” Studios;
 - * Genres by themselves seem to have significant (albeit, small) positive correlations with cult index;
 - * Studios by themselves also seem to have significant (albeit, small) positive correlations with cult index, BUT “seasoned” studios do worse (probably because they are involved in big movies too)



Lessons Learned, To-Dos

- * Be Ruthless with Extreme Outliers
- * Explore More Interactions
- * Get More Features:
 - * # of Quoted Dialogs
 - * Fan Vs. Critic Review Polarity
- * **Finally, ASK THE THREE QUESTIONS BEFORE BEGINNING A PREDICTION PROBLEM:**
 - * **What is the problem?**
 - * **Can a human predict this problem?**
 - * **How well? (why should a machine do better?)**