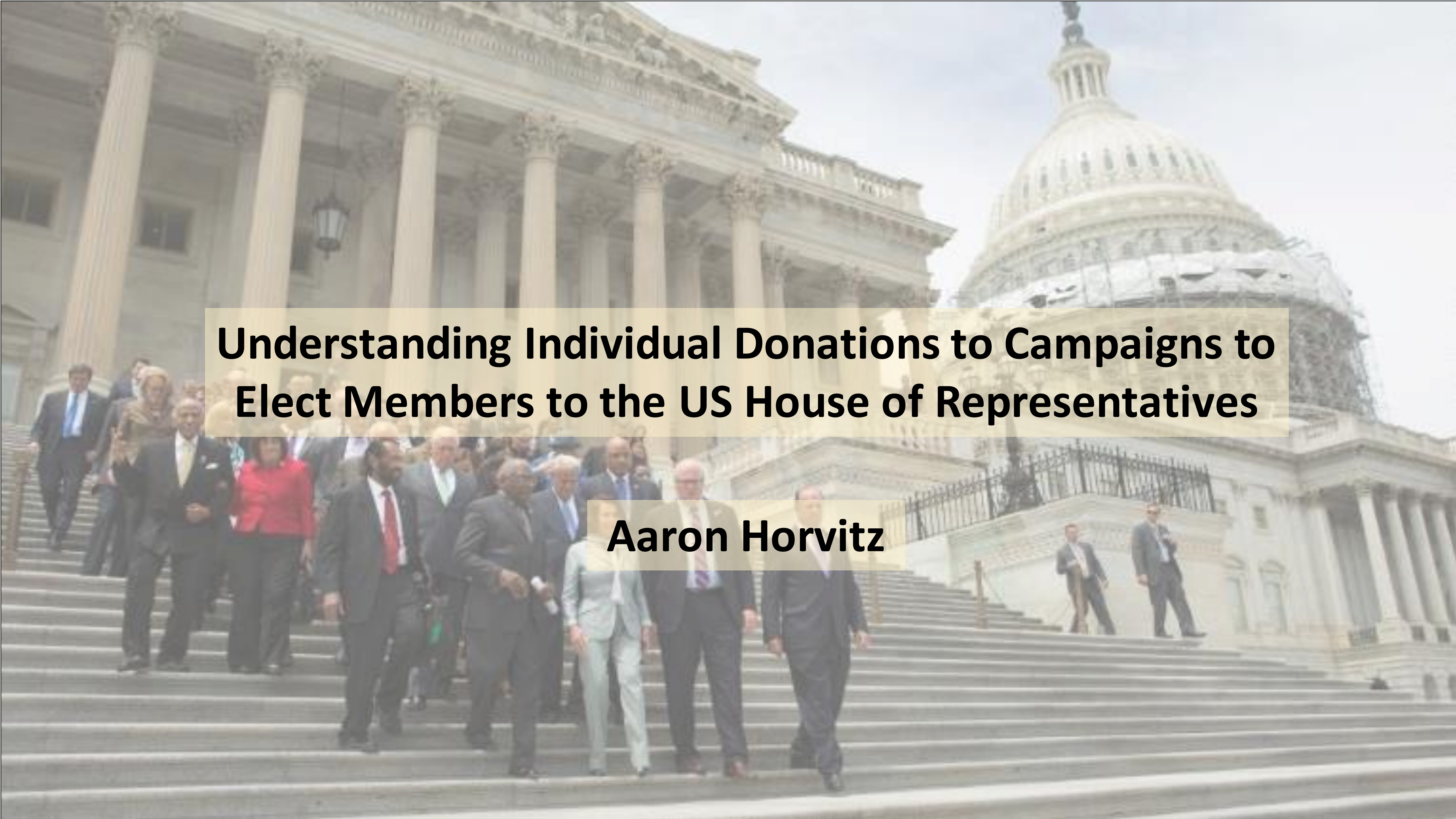
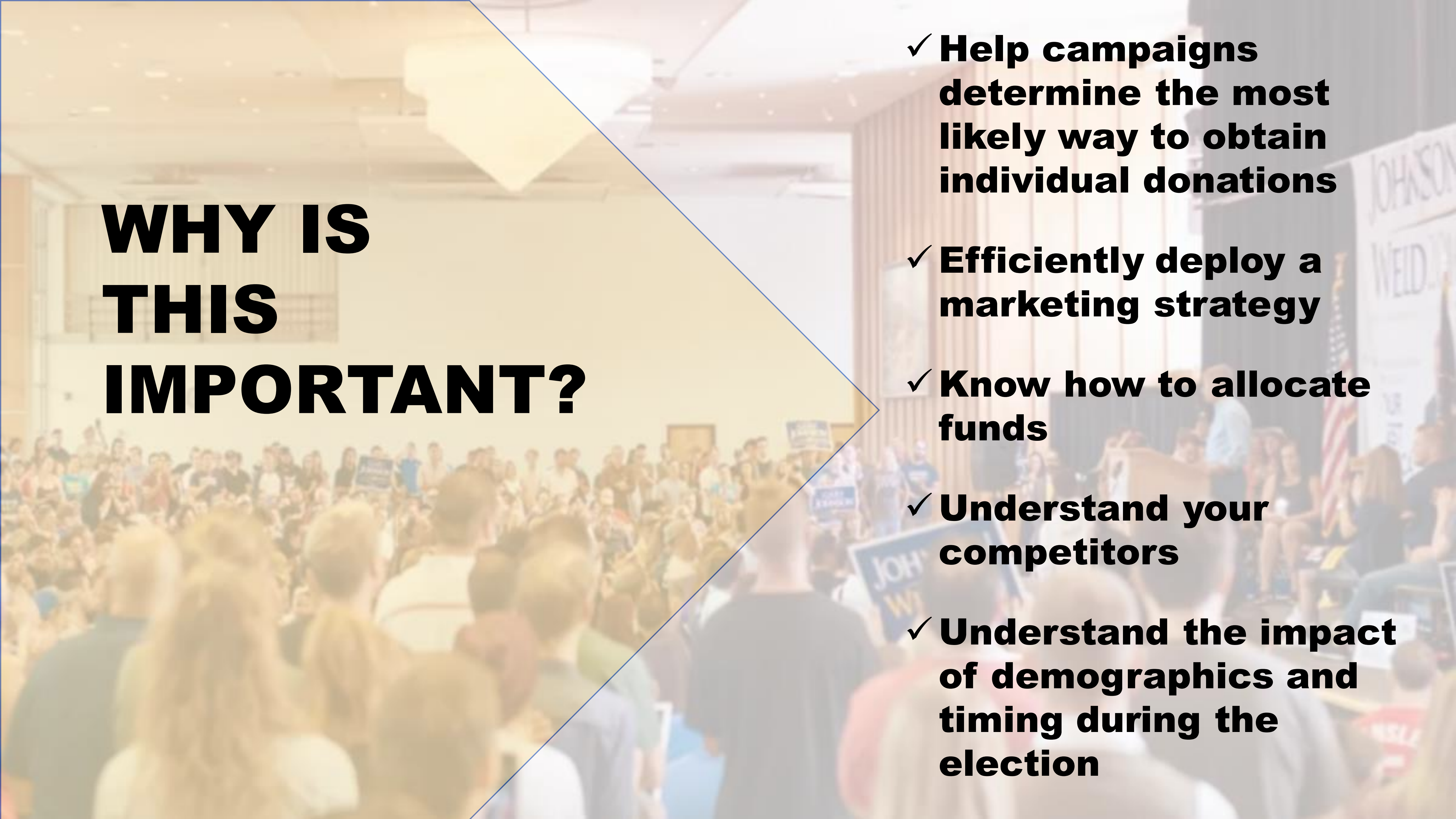


Understanding Individual Donations to Campaigns to Elect Members to the US House of Representatives

Aaron Horvitz



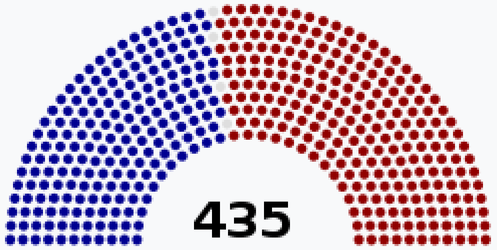
WHY IS THIS IMPORTANT?

- ✓ **Help campaigns determine the most likely way to obtain individual donations**
- ✓ **Efficiently deploy a marketing strategy**
- ✓ **Know how to allocate funds**
- ✓ **Understand your competitors**
- ✓ **Understand the impact of demographics and timing during the election**

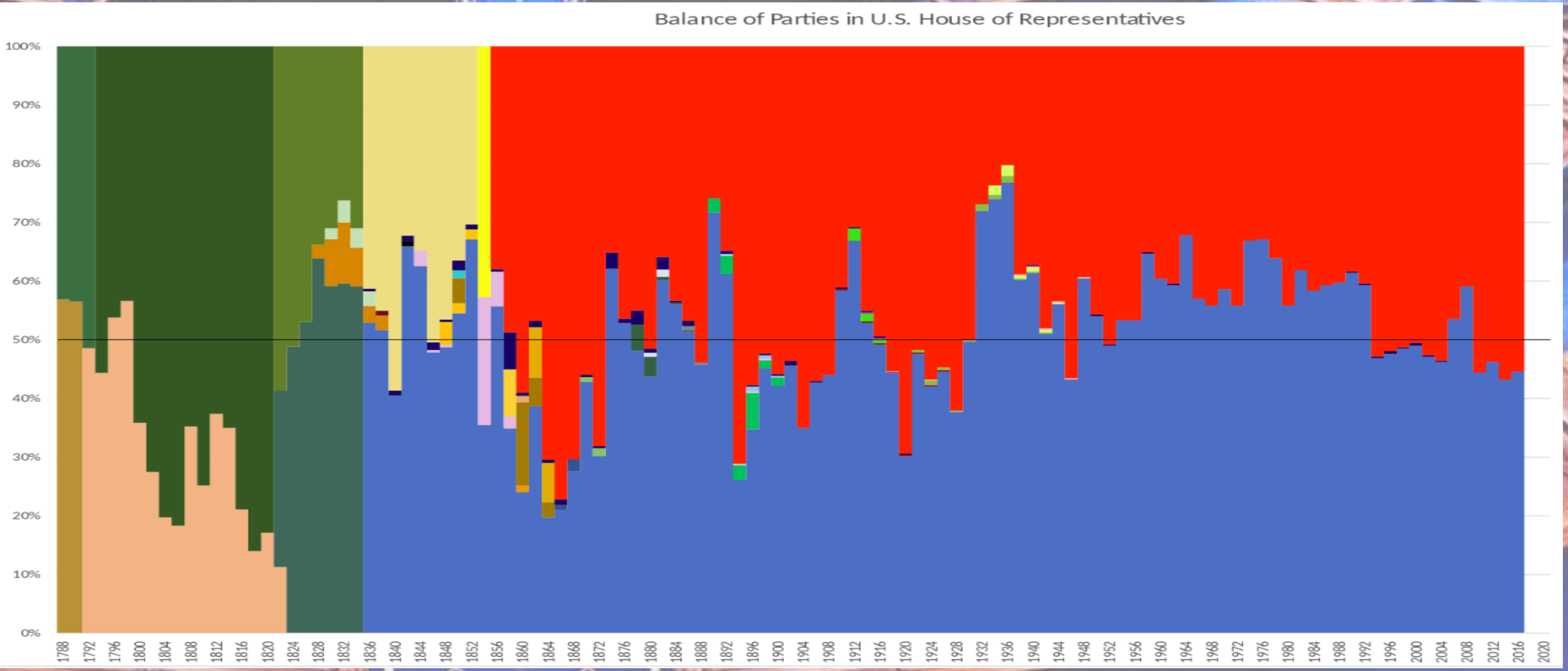
US House of Representatives

1. 2,175 races were analyzed from 2008 to 2016.
2. Elections are held every two years, therefore the model can be tested and updated relatively quickly. The House has been held by the same political parties since 1856.
3. There is extremely low turnover. It's rare for a challenger to beat an incumbent.
4. Each congressional district has roughly the same number of people (771,000).

Seats	435 voting members 6 non-voting members 218 needed for control



Political groups	**Majority** (238) • Republican (238) **Minority** (193) • Democratic (193) **Vacant** (4) • Vacant (4)
Length of term	Two years



At what rate do members of the US House of Representatives hold onto their jobs?

2016 cycle – House reelection rate was 97%
2014 cycle – House reelection rate was 95%
2012 cycle - House reelection rate was 90%
2010 cycle – House reelection rate was 85%
2008 cycle – House reelection rate was 94%
2006 cycle – House reelection rate was 94%
2004 cycle – House reelection rate was 98%
2002 cycle – House reelection rate was 96%
2000 cycle – House reelection rate was 98%
1998 cycle – House reelection rate was 98%

20 Year Average Reelection Rate: 95%

Data Sources

Can I project the return from different marketing expenditures on total individual donations to election campaigns to the House of Representatives?

Federal Election Commission Election Committee Operating Expenditures

2016 Count: 1,747,531
Unique: 99,416
Features: 25

Committee ID
Transaction Amount
Purpose
Entity Type

Multiple Itemized Expenses Reported Per Committee
(Parsed expenses into Marketing Categories)
(Grouped Expenses by Election Committee)

Federal Election Commission Election Committee Master File

2016 Count: 17811
Unique: 17811
Features: 15

Committee ID
Committee Name
Committee Type
Committee Party
Candidate ID

Multiple Committees Per Candidate
(Expenses grouped and summed by Candidate ID)

Federal Election Commission Candidate Summary File for House Races

2016 Count: 2411
Unique: 2411
Features: 50

Candidate ID
Candidate Name
Individual Contributions
Incumbency
District Code

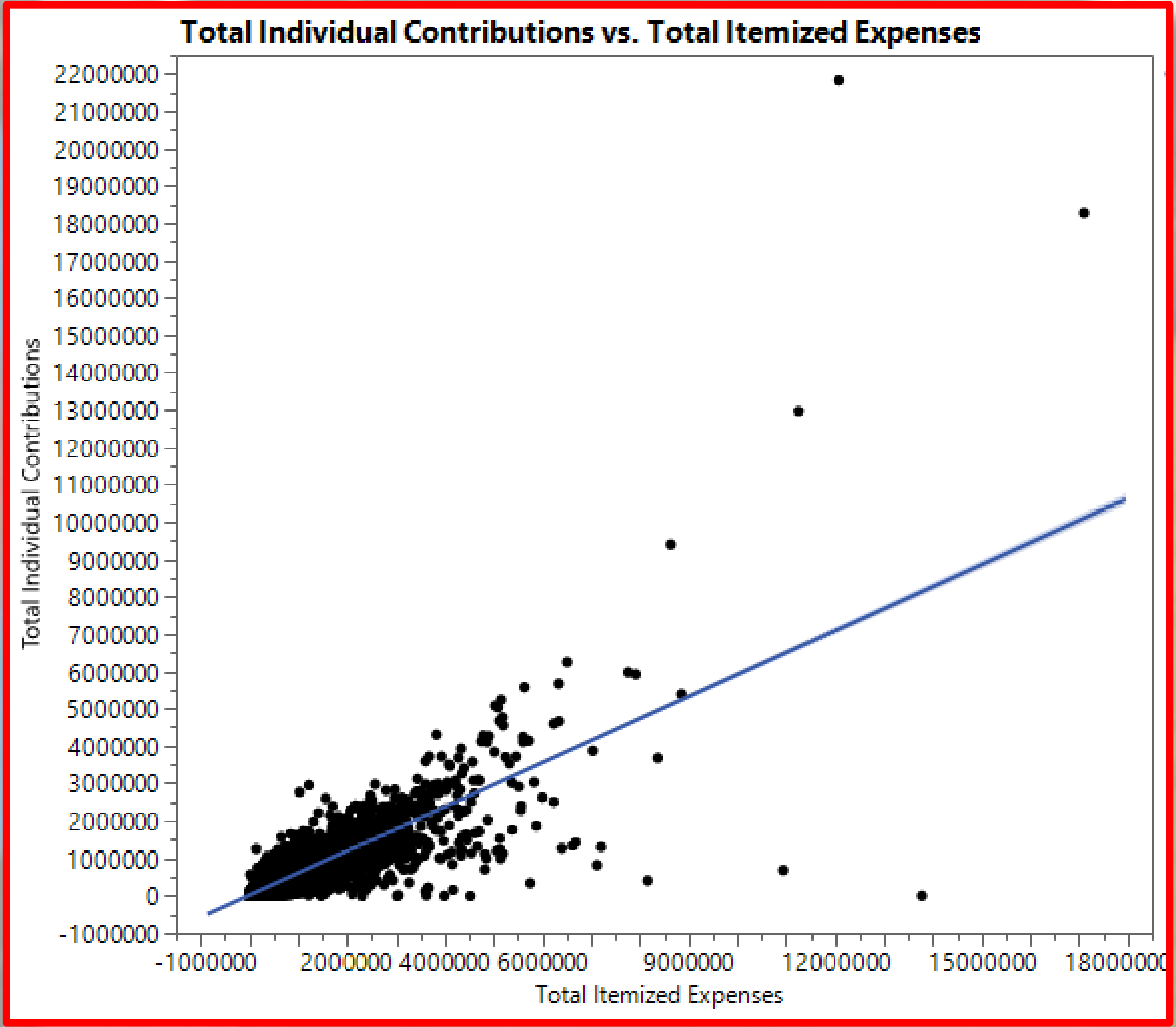
Misreported and Missing Candidate Info
(Only looked at candidates who reported donations and expenses.)

Department of Commerce US Census Bureau Data on Congressional Districts

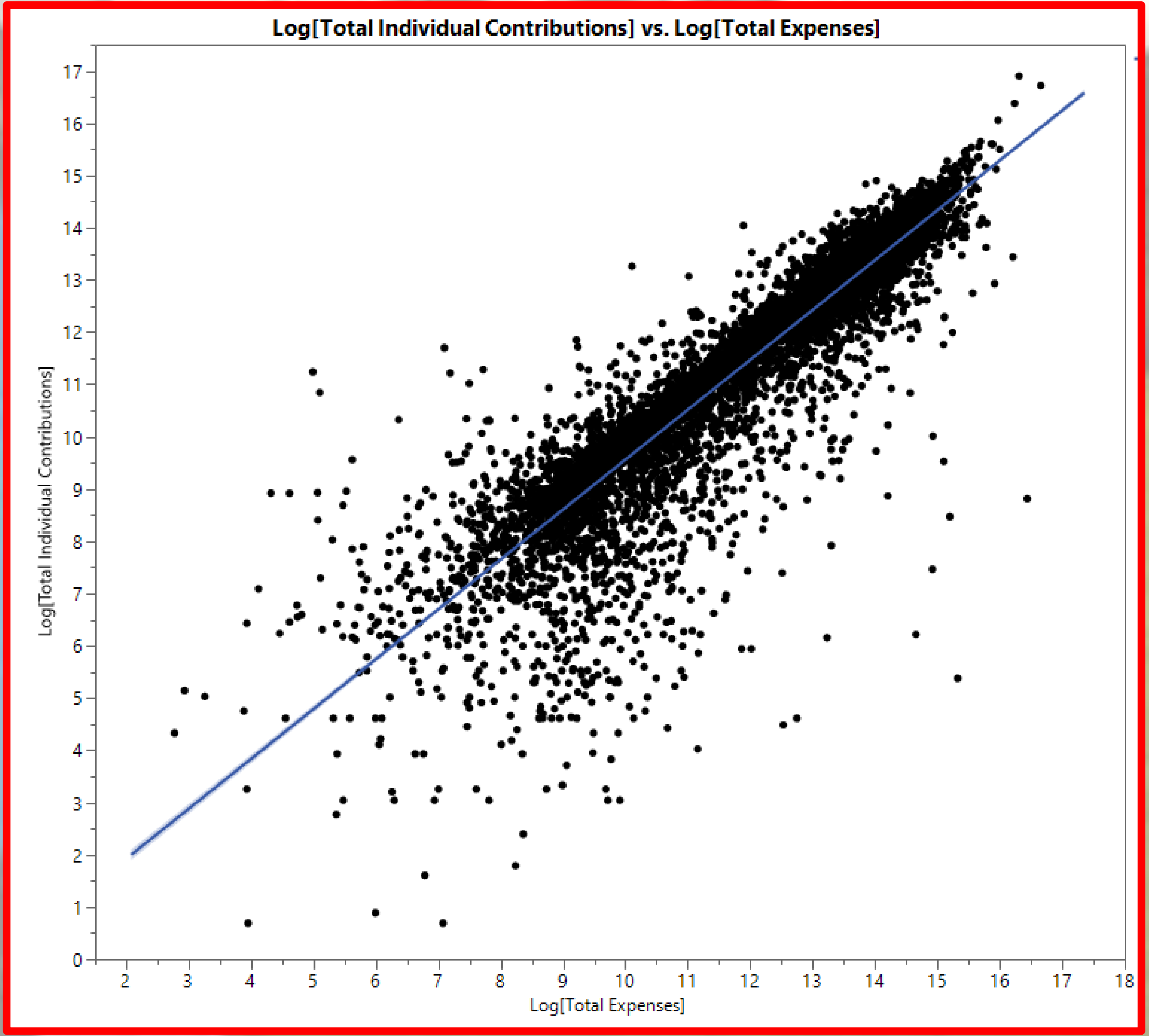
Count: 435
Unique: 435
Features: 15

District Code
Land Area
Percent Urban
Median Income
Total Population

What is our target? Can Expenses Predict or Correlate with Individual Contributions?



Log Transformation of the Expense and Total Individual Contributions Variables



Target Variable
Log (Total Individual Contributions)

ind_ite_con	Individual Itemized Contribution	Currency	Sum of itemized contributions from individuals	Contributions must be itemized (i.e. listing specific information about the donor and the contribution) when the total given to the campaign from a single individual exceeds $200.
ind_uni_con	Individual Unitemized Contribution	Currency	Sum of unitemized contributions from individuals	Total given to the campaign where specific information about the donors is not required because they have not given more than $200
ind_con	Individual Contribution	Currency	Total contributions from individuals	Sum of the previous two fields (i.e. ind_ite_con and ind_uni_con)

Coded Dummy Variables

Presidential Cycle vs. Midterm Election Cycle

1 Presidential Cycle

0 Mid Term Cycle

Open Seat vs. Not Open

1 Open Seat

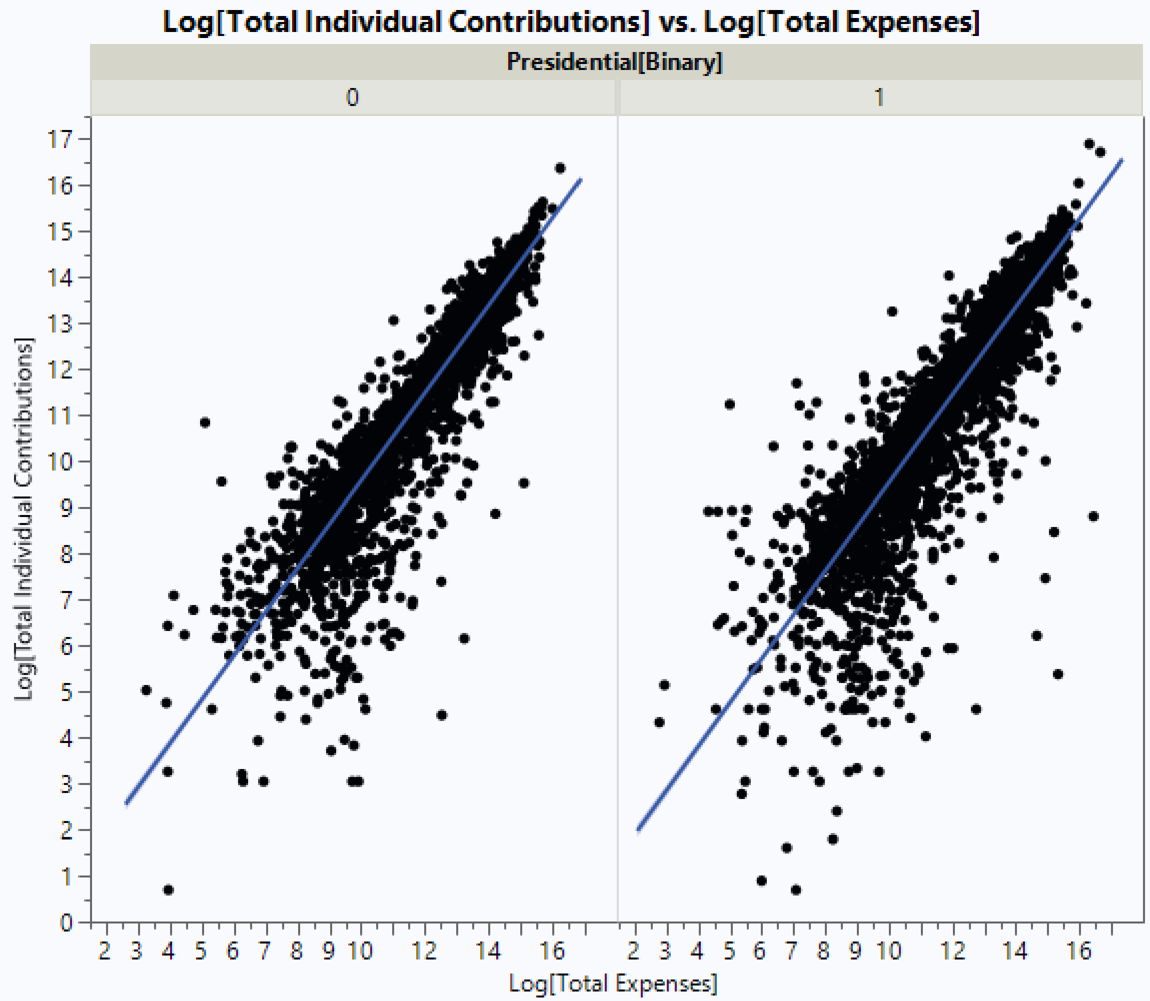
0 Not Open

Incumbent vs. Challenger

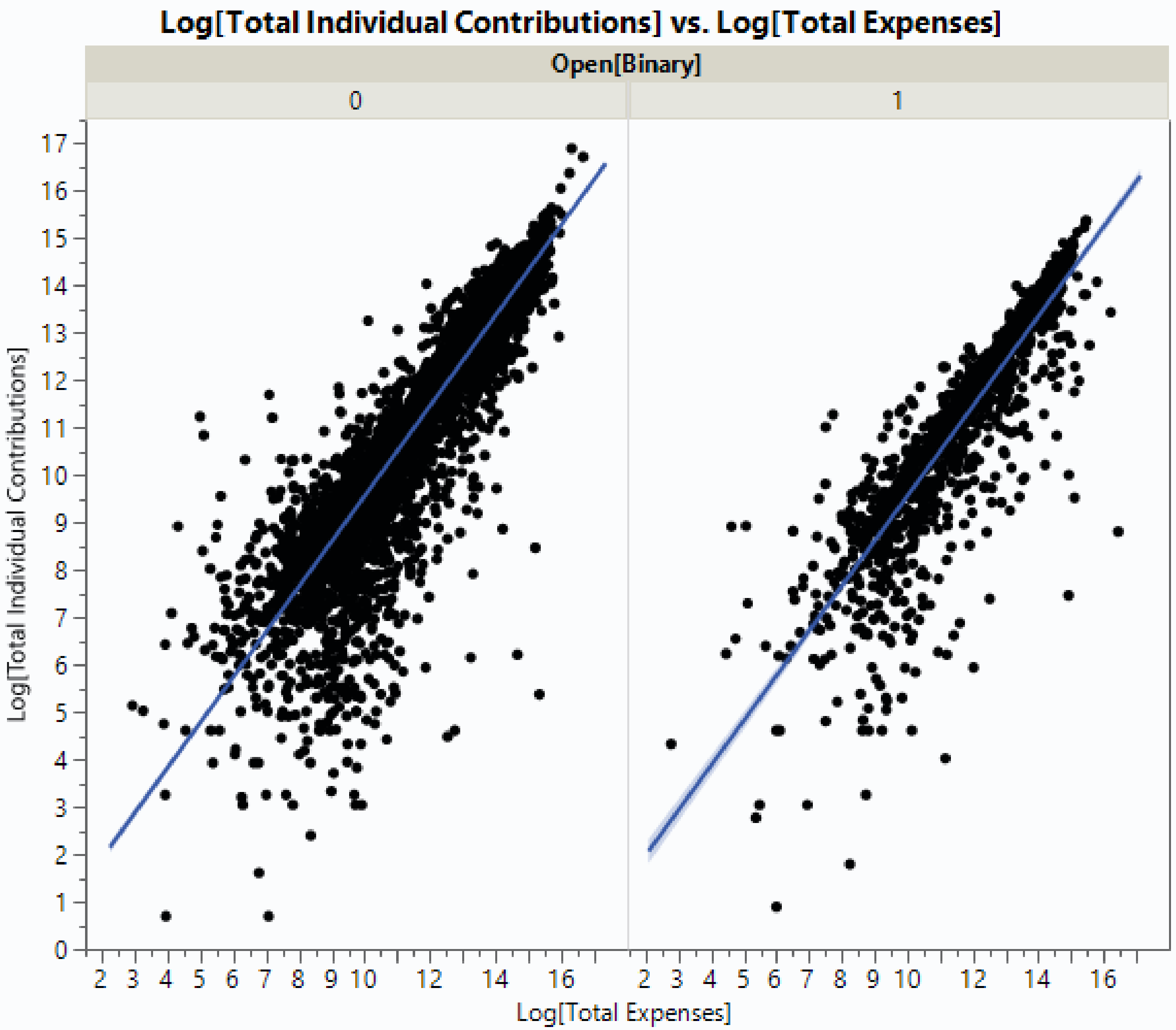
1 Incumbent

0 Challenger

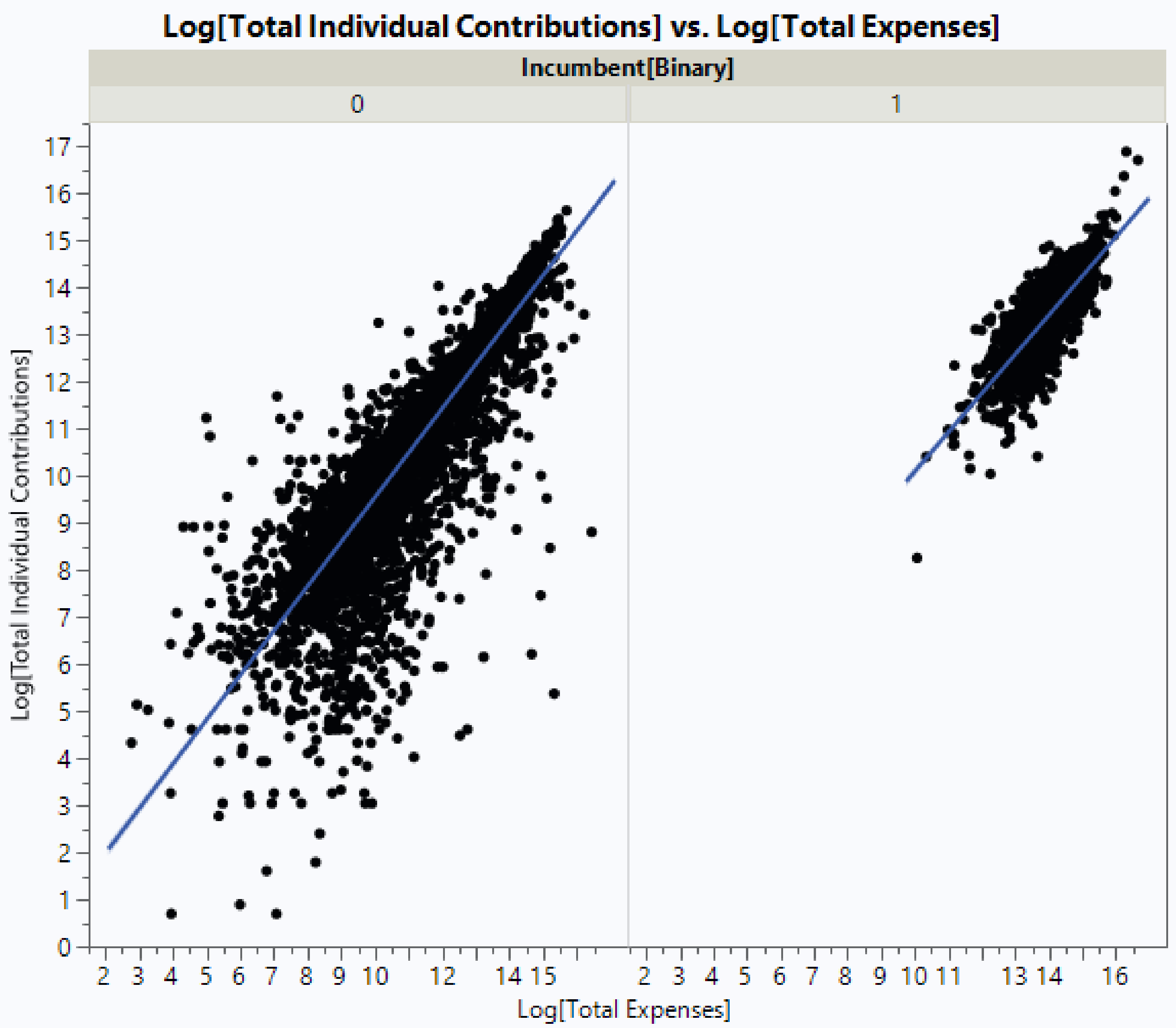
Presidential Cycle vs. Midterm Election Cycle (2008 to 2016)



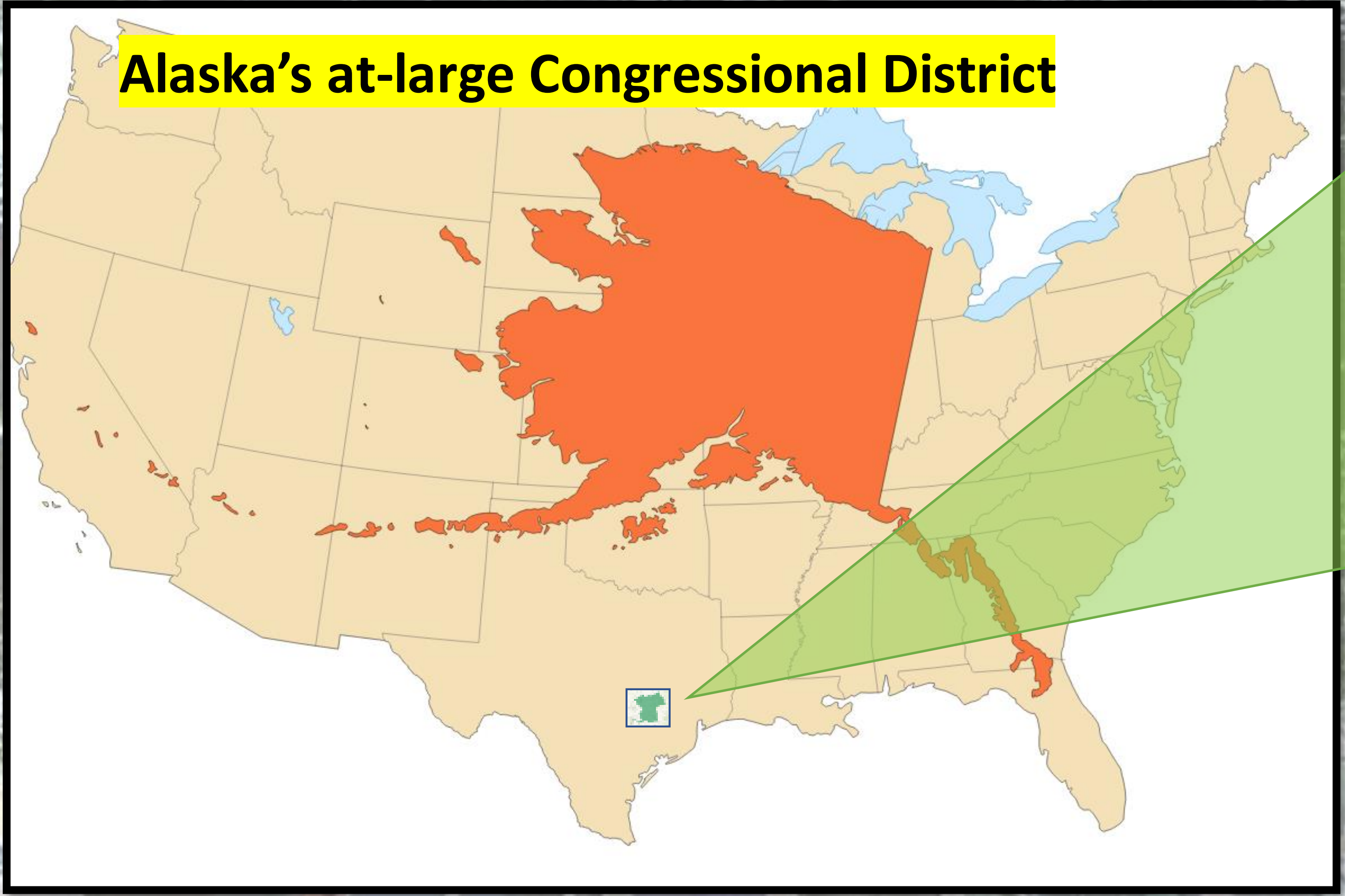
Open Seat vs. Not Open Seat (2008 to 2016)

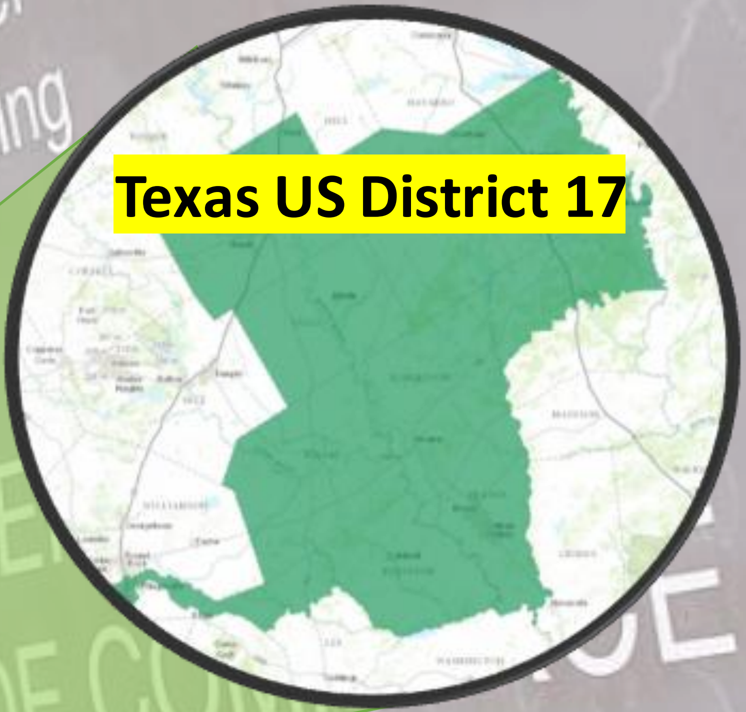


Incumbent vs. Challenger (2008 – 2016)



Demographic Variables and Illustration of Differences Between Districts





<u>Demographic Variables</u>

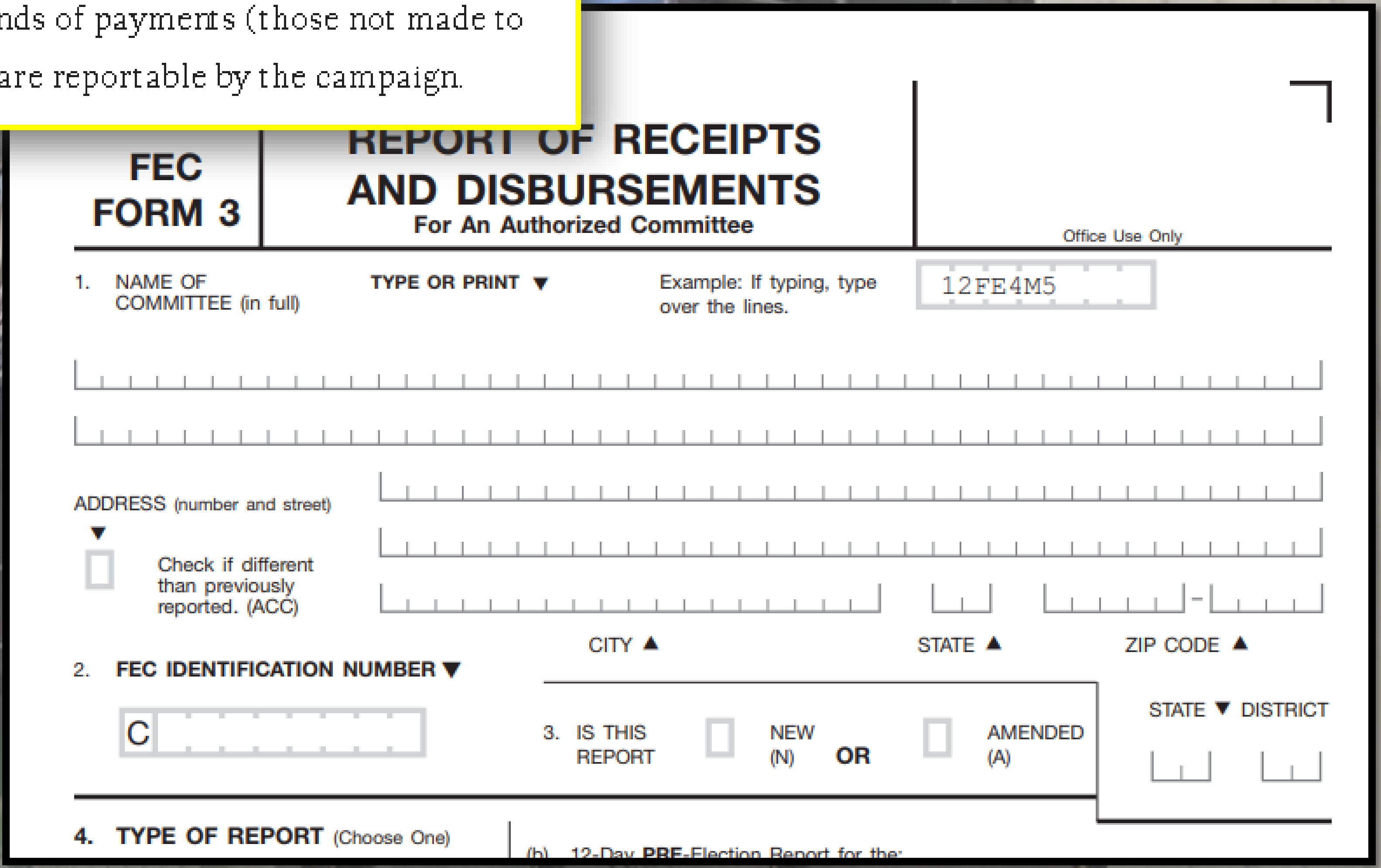
- Land Area
- Percent Urban
- Median Income
- Total Population

Marketing Expense Variables

- Travel
- Food and Beverage
- Marketing Expense Direct Response
- Marketing Expense for Analytics
- Marketing Expense for Bulk Email
- Marketing Expense for Canvassing
- Marketing Expense for Collateral and Print
- Marketing Expense for CRM
- Marketing Expense for Digital Advertising
- Marketing Expense for Digital Assets
- Marketing Expense for Events
- Marketing Expense for Fax Blasting
- Marketing Expense for Gifts
- Marketing Expense for Imagery
- Marketing Expense for Mail
- Marketing Expense for Outdoor Advertising
- Marketing Expense for Phone Banking
- Marketing Expense for Promotional Items
- Marketing Expense for Public Relations
- Marketing Expense for Social Media
- Marketing Expense for Text Blasting
- Marketing Expense for TV and Radio
- Marketing Expense Fundraising
- Marketing Expense Other

Reporting Expenses to the Federal Election Commission

An expenditure is a purchase, payment, distribution, loan, advance, deposit or gift of money or anything of value to influence a federal election. "Disbursement" is a broader term that covers both expenditures and other kinds of payments (those not made to influence a federal election). All disbursements are reportable by the campaign.



FEC FORM 3

REPORT OF RECEIPTS AND DISBURSEMENTS

For An Authorized Committee

Office Use Only

1. NAME OF COMMITTEE (in full) TYPE OR PRINT ▼ Example: If typing, type over the lines. 12FE4M5

ADDRESS (number and street) ▼

Check if different than previously reported. (ACC)

CITY ▲ STATE ▲ ZIP CODE ▲

2. FEC IDENTIFICATION NUMBER ▼

C

3. IS THIS REPORT NEW (N) OR AMENDED (A)

STATE ▼ DISTRICT

4. TYPE OF REPORT (Choose One)

(b) 12-Day PRE-Election Report for the:

Parsing Itemized Expenses Into Categories

Original Expense Data Provided in ASCII-28 delimited format


```
C00004606|N|2015|M4|15951124870|21B|F3X|SB|DIRECT MAIL SYSTEMS, INC.|CLEARWATER|FL|
, RENT, POSTAGE|ORG|4041320151241802314|1002280|37970601||C00008169|N|2015|Q1|15970
5|Q1|15951129812|21B|F3X|SB|WELLS FARGO|BOULDER|CO|80303|03/11/2015|20.18|G|BANK FEE
|N|2015|M4|15970340878|21B|F3X|SB|WELLS FARGO|PORTLAND|OR|97228|03/31/2015|5||BANK F
30487|21B|F3X|SB|FIFTH THIRD BANK|ELMHURST|IL|60062|03/11/2015|85.76|O|SERVICE CHARG
RG|4041420151241820362|1002717|SB21B.5147||C00417295|N|2015|Q1|15951130848|21B|F3X|
001|Administrative/Salary/Overhead Expenses |||ORG|4031820151240826105|998141|48E92
61956|N|2015|M3|15950876123|21B|F3X|SB|OVERTON PARK|ATLANTA|GA|303393361|02/27/2015|
G|4031820151240826113|998142|B10197B234B8A403CA01||C00361956|N|2015|M3|15950876122|
240866101|998163|D621091||C00108613|N|2015|M3|15970304031|21B|F3X|SB|BAILEY, STEFAN|
S|IN|462401468|02/05/2015|212.63|P|PAYROLL PROCESSING FEE||||ORG|40319201512408661
|15950876774|21B|F3X|SB|PRECISION MARKETING, INC.|ARLINGTON|VA|222070670|02/10/2015|
986|02/13/2015|5322.65||PAYROLL TAXES|||PAYROLL TAXES|ORG|4031920151240866261|99818
.E48107||C00038505|N|2015|M3|15950876754|21B|F3X|SB|ADP EASYPAY|CHARLOTTE|NC|2821739
3.08||EMPLOYEE MEDICAL INSURANCE|||EMPLOYEE MEDICAL INSURANCE|ORG|40319201512408662
240866277|998184|50213.E47977||C00038505|N|2015|M3|15950876758|21B|F3X|SB|COMPUTER F
LEEN HODGMAN|RALEIGH|NC|27605|02/13/2015|1145.62||PAYROLL|||PAYROLL|ORG|403192015124
15950876762|21B|F3X|SB|MDI IMAGING & MAIL|DULLES|VA|20166|02/10/2015|4653.89||DIRECT
|CHICAGO|IL|606731258|02/13/2015|100||POSTAGE FOR POSTAGE METER||||POSTAGE FOR POSTA
866299|998184|50316.E48141||C00038505|N|2015|M3|15950876766|21B|F3X|SB|PIRYX, INC.|
|F3X|SB|PIRYX, INC.|SAN FRANCISCO|CA|94105|02/09/2015|5.81||ONLINE CREDIT CARD FEES|
INE CREDIT CARD FEES|||ONLINE CREDIT CARD FEES|ORG|4031920151240866310|998184|50316
.E48099||C00038505|N|2015|M3|
98184|50316.E48121||C00064766|
|SIMI VALLEY|CA|93062|02/27/20
A395||C00170258|N|2015|M3|1595
03764|N|2015|M3|15970303144|2
20151240825072|998064|SB21B.41
```


Parser in Python – Almost 3,000 lines of code

```python

def overlooked(expenses_df):

    print("")
    print("--------Parsing Overlooked Items--------")
    print("")

    local_df=expenses_df

    local_df['CATEGORIES'] =   np.where(local_df['CATEGORIES'
                                np.where(local_df['CATEGORIES'].str.
                                np.where(local_df['CATEGORIES'].str.
                                np.where(local_df['CATEGORIES'].str.
                                np.where(local_df['CATEGORIES'].str.
                                np.where(local_df['CATEGORIES'].str.
                                np.where(local_df['CATEGORIES'].str.
                                np.where(local_df['CATEGORIES'].str.
                                            ))))))))

    parsing_report(local_df)      #Print out parsing report
    return local_df

```

```python
# Read headers into a dataframe

header_df=pd.read_csv(header_filepath)
local_df = pd.read_csv(fec_data_filepath,encoding = "ISO-8859-1",sep='|',names = header_df,index_col=False)

#Print dataframe with headers
local_df.head(600)
```

Parsing Itemized Expenses Into Categories

Count: 1,747,531 Expense Items In 2016

C00004606	03/27/2015	982.80	RENT
C00010470	03/31/2015	269.37	BANK SERVICE FEE
C00465211	03/01/2015	750.00	PAC ADMINISTRATIVE EXPENSE, SUPPLIES
C00008169	03/05/2015	310.00	TAX PREPARATION
C00008169	02/24/2015	550.00	TAX PAYMENT
C00382028	04/09/2015	88.78	TELEPHONE
C00445593	01/12/2015	14.75	BANK FEE
C00445593	02/11/2015	20.29	BANK FEE
C00445593	03/11/2015	20.18	BANK FEE
C00570226	03/30/2015	1750.00	LEGAL CONSULTING
C00570226	03/30/2015	1750.00	POLITICAL CONSULTING
C00570226	03/30/2015	1227.43	SOCIAL MEDIA SERVICES
C00570226	03/30/2015	1750.00	SOCIAL MEDIA SERVICES
C00393769	03/11/2015	124.95	CREDIT CARD PROCESSING FEES
C00393769	03/31/2015	5.00	BANK FEE
C00043992	03/02/2015	89.94	CREDIT CARD PROCESSING FEES
C00473918	03/19/2015	100000.00	POLLING/SURVEYS
C00473918	03/03/2015	80.00	CREDIT CARD SERVICE CHARGES
C00473918	03/03/2015	25.00	CREDIT CARD SERVICE CHARGES
C00473918	03/06/2015	25000.00	PUBLICATION & DUES
C00040253	03/11/2015	85.76	SERVICE CHARGE
C00553271	01/02/2015	2540.00	SOCIAL MEDIA CONSULTING
C00417295	02/03/2015	350.00	RENT/UTILITIES FOR MASS PAC OCAMPAI
C00417295	03/02/2015	350.00	RENT/UTILITIES FOR MASS PAC OCAMPAI
C00417295	03/31/2015	350.00	RENT/UTILITIES FOR MASS PAC OCAMPAI
C00417295	03/31/2015	566.26	CAMPAIGN CREDIT CARD CHARGES
C00417295	03/31/2015	566.26	CAMPAIGN MEALS CHARGED TO CREDIT C
C00417295	03/20/2015	6000.00	MASS PAC CAMPAIGN EVENT - FOOD/FACI
C00002261	02/02/2015	136.18	BANK FEE
C00002261	02/09/2015	3.00	BANK FEE

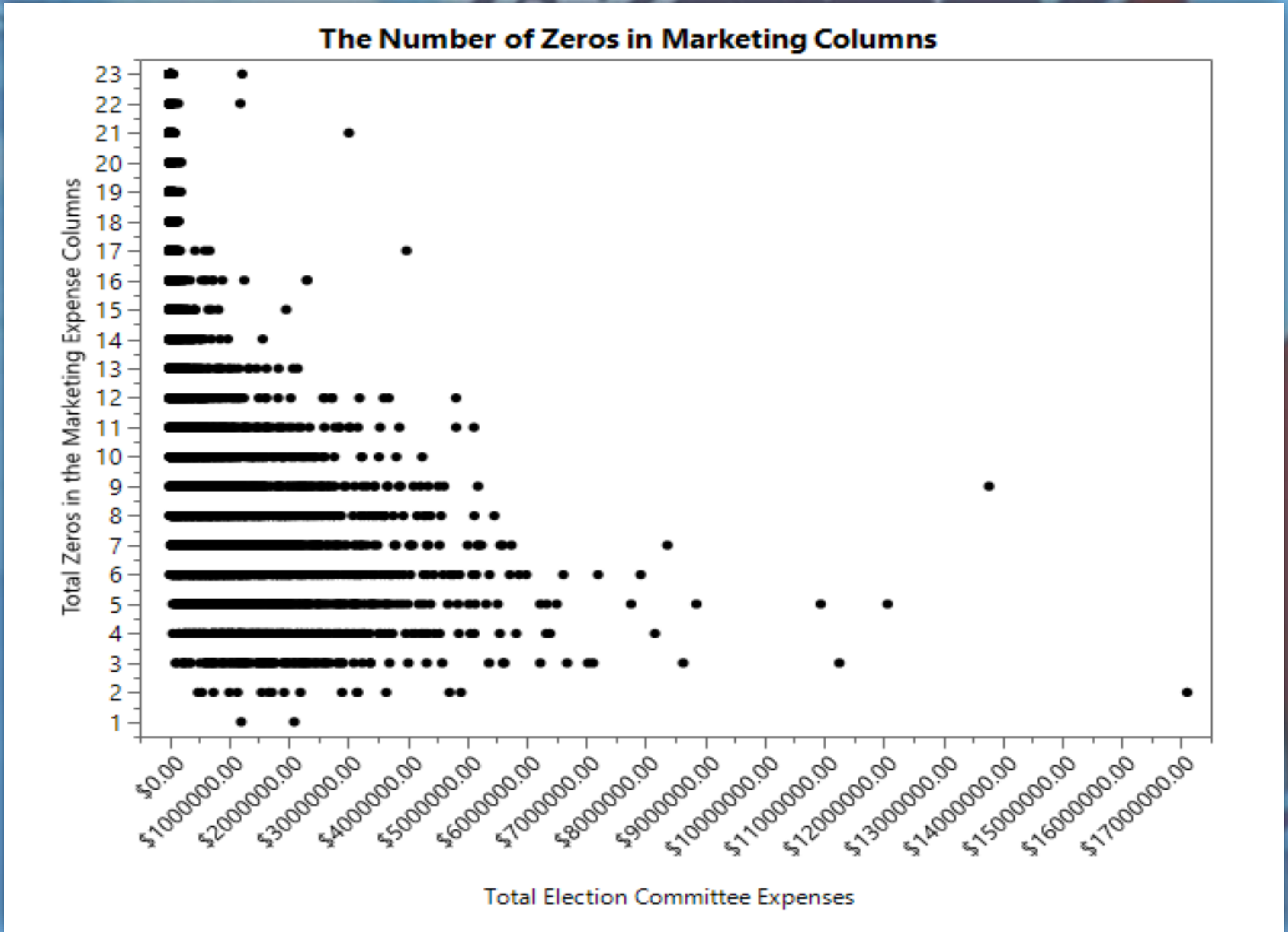
PARSER

52 Total Categories
24 Marketing Related

Accounting Expense
Association Dues ar
Bank Fees
Cell Phones
Compensation as Benefits
Compensation as Bonuses and Commission
Compensation for Salary and Hourly
Compensation to Consultants
Compensation to Contractors
Compensation to Staff
Credit Card PayPal and Transaction Fees
Donation Expense
Food and Beverages
In Kind Contributions
Legal Expense
Marketing Expense Direct Response
Marketing Expense for Analytics
Marketing Expense for Bulk Email
Marketing Expense for Canvassing
Marketing Expense for Collateral and Print
Marketing Expense for CRM
Marketing Expense for Digital Advertising
Marketing Expense for Digital Assets
Marketing Expense for Events
arketing Expense for Fax Blasting
eting Expense for Gifts
eting Expense for Imagery
arketing Expense for Mail
Marketing Expense for Outdoor Advertising
Marketing Expense for Phone Banking
Marketing Expense for Promotional Items
Marketing Expense for Public Relations
Marketing Expense for Social Media
Marketing Expense for Text Blasting
Marketing Expense for TV and Radio
Marketing Expense Fundraising
Marketing Expense Other
Travel Expense
Food and Beverages
Obscurely Marked Items
Office Expense For Furniture and Equipment
Office Expense for IT Services
Office Expense for Network and Computer Equipment
Office Expense for Office Services
Office Expense for Rent and Utilities
Office Expense for Software
Office Expense for Supplies
Other Fees and Interest
Petty Cash
Reimbursements
Tax Expense
Uncategorized Expenses

Addressing Complications by Removing Low Spenders

(Number of zeros in the marketing expense columns)



Addressing Complications by Removing Low Spenders

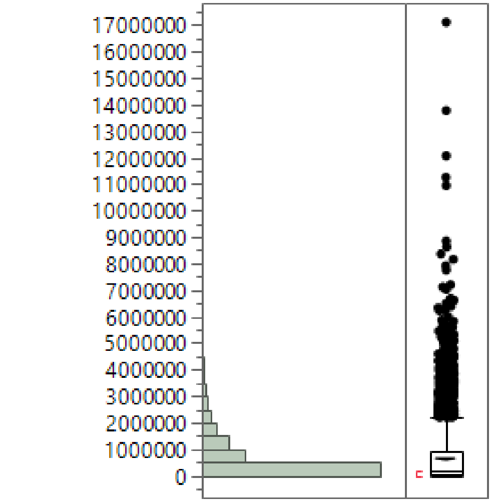
(Cutting out unrealistic and paper candidates)

It’s important to point out that the average winner in the House spent $1.3 M. There are many independents and third party candidate who are not serious. Additionally, many candidates run, who are considered by their political parties as, “paper candidates.” For instance, a Republican or Democrat may run in a district, where they have no hope of winning, but their candidacy maintains the party’s presence on the ballot. Third parties, like the Green and Libertarian parties, frequently have these “paper candidates” in national races. These type of candidates’ election committees make up the bulk of those reporting less than $250,000 in total itemized expenses to the Federal Election Commission. They tend not to report enough marketing expenses or income to be meaningful in the model. Since their total budgets are minuscule, many of their marketing expenses show up as zero. Including them in the model, results in higher variance between actual and predicted outcomes, among election committees with low total expenses, not just marketing. Therefore, they were excluded.

All

Distributions

Total Itemized Expenses



Quantiles

100.0%	maximum	17102459.28
99.5%		5345063.9002
97.5%		3245397.5205
90.0%		1823221.125
75.0%	quartile	902892.9825
50.0%	median	205559.035
25.0%	quartile	26842.35
10.0%		7290.822
2.5%		1716.878
0.5%		281.777
0.0%	minimum	15

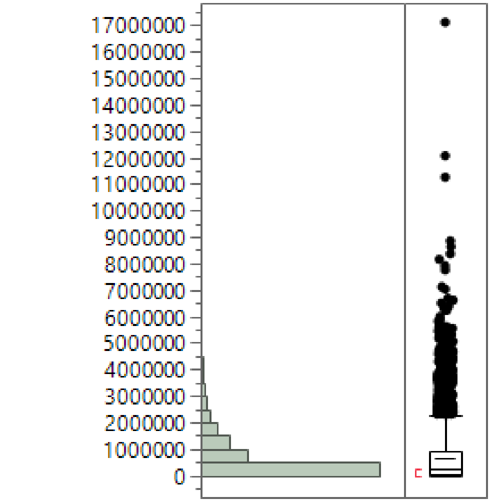
Summary Statistics

Mean	638396.22
Std Dev	991481.82
Std Err Mean	11598.064
Upper 95% Mean	661131.77
Lower 95% Mean	615660.67
N	7308

Challenger

Distributions

Total Itemized Expenses



Quantiles

100.0%	maximum	17102459.28
99.5%		5509216.2288
97.5%		3428007.2307
90.0%		1867854.189
75.0%	quartile	938928.44
50.0%	median	221853.18
25.0%	quartile	26062.09
10.0%		7111.528
2.5%		1609.0705
0.5%		292.177
0.0%	minimum	17.79

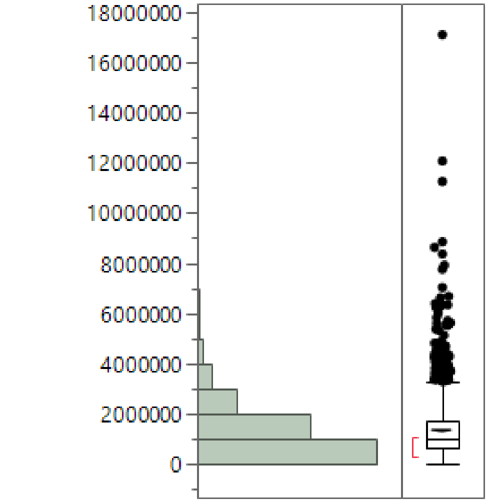
Summary Statistics

Mean	662968.2
Std Dev	1010166.1
Std Err Mean	13241.321
Upper 95% Mean	688926.11
Lower 95% Mean	637010.29
N	5820

Incumbent

Distributions

Total Itemized Expenses



Quantiles

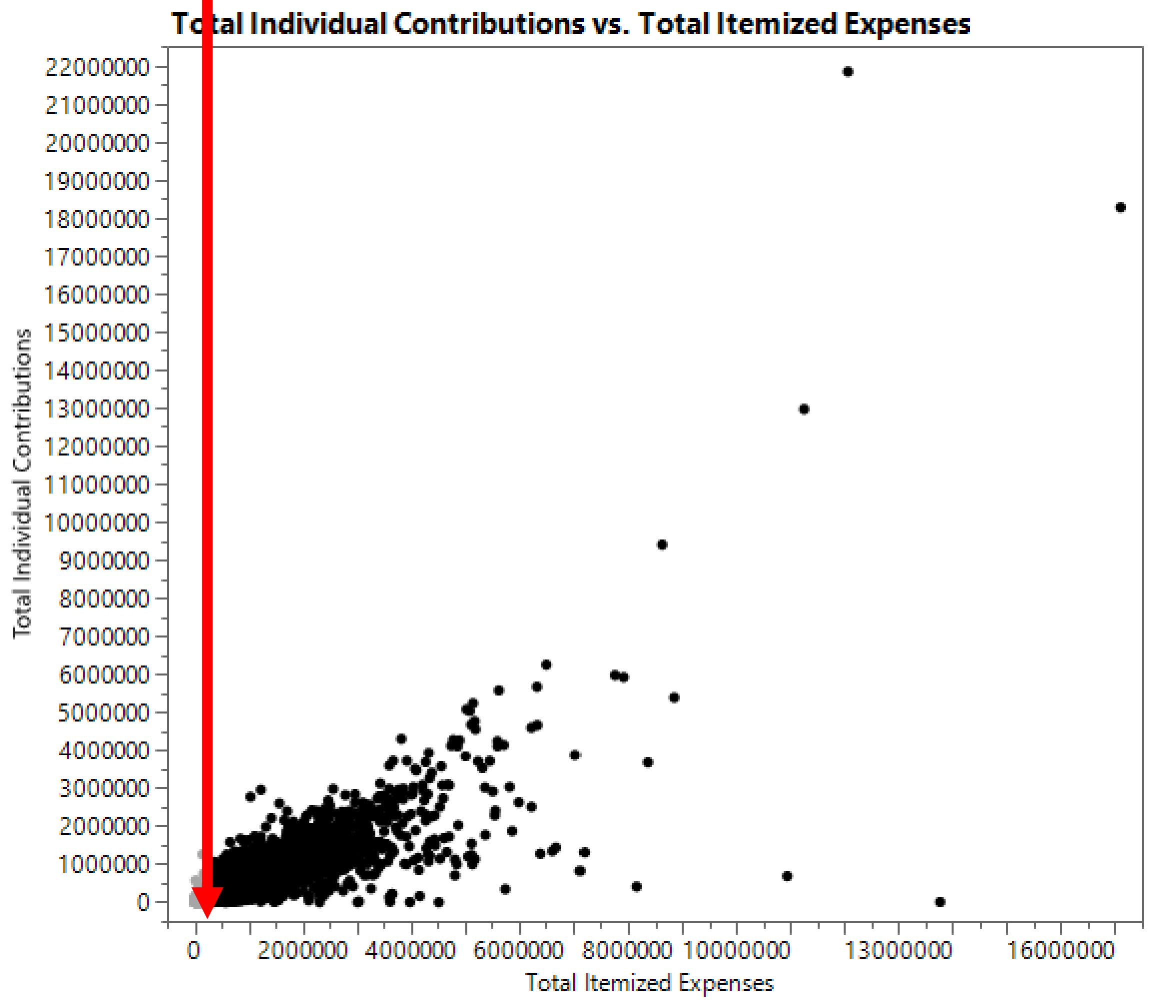
100.0%	maximum	17102459.28
99.5%		6697058.9208
97.5%		4231302.124
90.0%		2624087.928
75.0%	quartile	1691754.29
50.0%	median	996739.79
25.0%	quartile	628001.02
10.0%		396451.306
2.5%		225923.44
0.5%		93965.5344
0.0%	minimum	23662.99

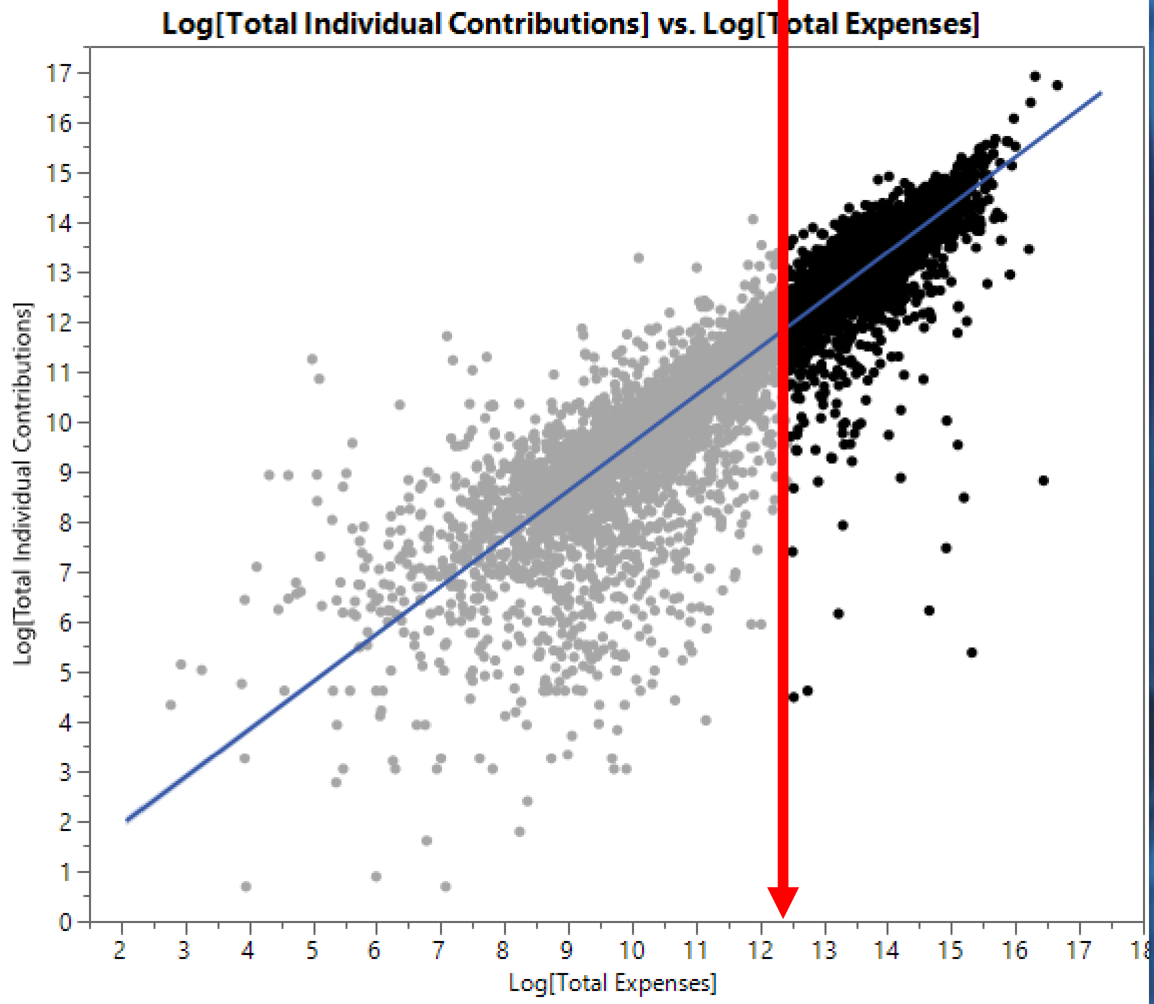
Summary Statistics

Mean	1330028.5
Std Dev	1156600.9
Std Err Mean	25920.769
Upper 95% Mean	1380863.2
Lower 95% Mean	1279193.8
N	1991

Impact of Cutting Off Spenders Below $250,000

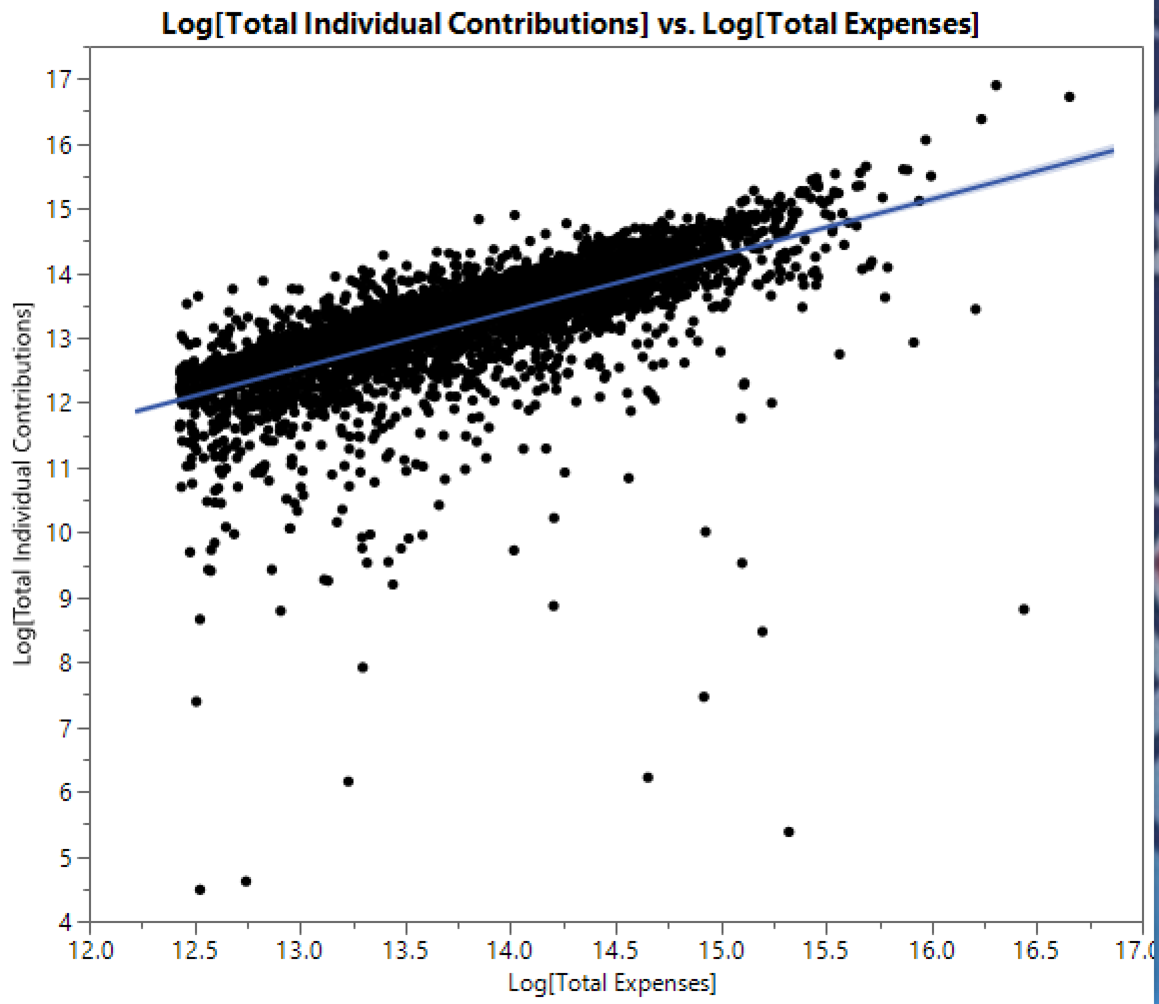
$250,000 Cutoff



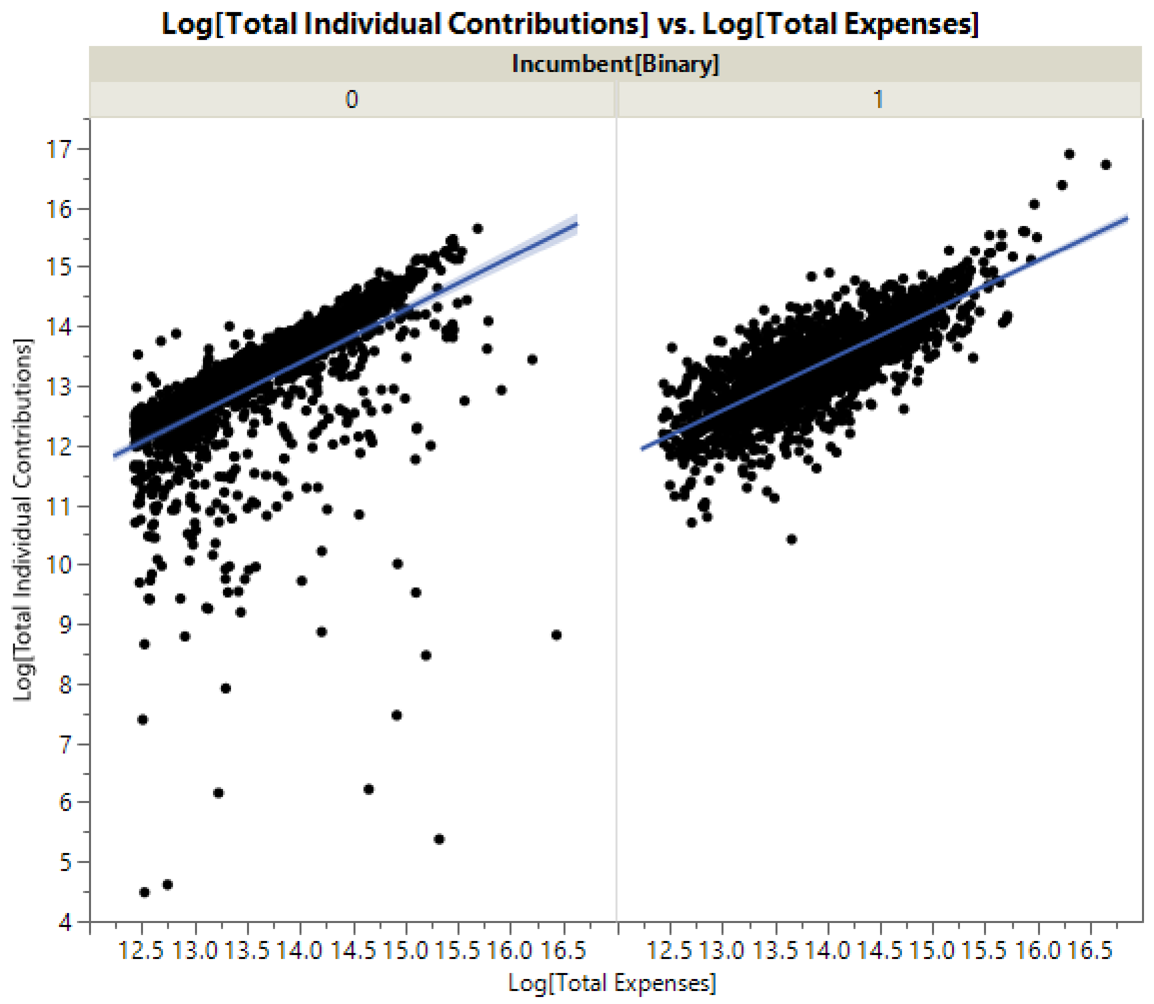


Impact of Cutting Off Spenders Below $250,000

Log [Total Individual Contributions] vs. Log [Total Spending] After $250,000 Cut Off

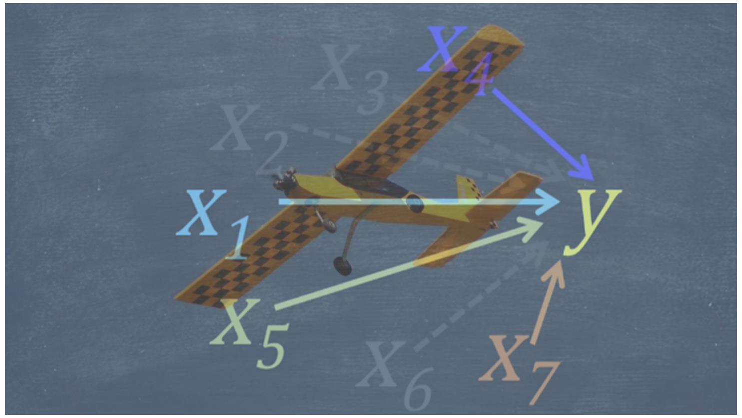


Log [Total Individual Contributions] vs. Log [Total Spending] After $250,000 Cut Off



Parameter Estimates

Term	Estimate	Std Error	t Ratio	Prob>|t|	Lower 95%	Upper 95%	VIF
Intercept	8.7536182	0.142478	61.44	<.0001*	8.4741974	9.033039	.
Log[Food and Beverage]	0.0866791	0.010548	8.22	<.0001*	0.0659922	0.1073659	1.4902084
Log[Events]	0.0846836	0.012017	7.05	<.0001*	0.0611156	0.1082516	1.4771977
Log[TV and Radio	0.0562073	0.00591	9.51	<.0001*	0.0446163	0.0677983	1.2840112
Log[Promotional Items]	0.0508741	0.007035	7.23	<.0001*	0.0370768	0.0646713	1.108474
Log[Travel]	0.0466189	0.007433	6.27	<.0001*	0.0320421	0.0611958	1.4708739
Log[Collateral]	0.0385243	0.009133	4.22	<.0001*	0.020614	0.0564346	1.1995673
Log[Analytics]	0.0280557	0.004949	5.67	<.0001*	0.0183493	0.0377621	1.2401431
Log[Text Blasting]	0.0279273	0.004458	6.26	<.0001*	0.0191838	0.0366708	1.1017282
Log[Public Relations]	0.0270696	0.004166	6.50	<.0001*	0.0189001	0.0352391	1.0899787
Log[Direct Response]	0.0182515	0.003013	6.06	<.0001*	0.0123425	0.0241604	1.0806747
Log[Digital Assets]	0.0150689	0.004774	3.16	0.0016*	0.0057074	0.0244305	1.1323418
Median Income	4.7313e-6	9.595e-7	4.93	<.0001*	2.8496e-6	6.613e-6	1.0691533
Log[Gifts]	-0.010066	0.00454	-2.22	0.0267*	-0.01897	-0.001162	1.2230298
Log[CRM]	-0.017672	0.008536	-2.07	0.0385*	-0.034412	-0.000932	1.0254116
Log[Outdoor Advertising]	-0.018304	0.003959	-4.62	<.0001*	-0.026068	-0.01054	1.1259863



Effect Summary

Source	LogWorth	PValue
Log[TV and Radio	20.280	0.00000
Log[Food and Beverage]	15.433	0.00000
Log[Promotional Items]	12.169	0.00000
Log[Events]	11.601	0.00000
Log[Public Relations]	9.990	0.00000
Log[Travel]	9.361	0.00000
Log[Text Blasting]	9.339	0.00000
Log[Direct Response]	8.784	0.00000
Log[Analytics]	7.783	0.00000
Median Income	6.053	0.00000
Log[Outdoor Advertising]	5.396	0.00000
Log[Collateral]	4.590	0.00003
Log[Digital Assets]	2.791	0.00162
Log[Gifts]	1.573	0.02673
Log[CRM]	1.414	0.03855

Summary of Fit

RSquare	0.420507
RSquare Adj	0.416173
Root Mean Square Error	0.688891
Mean of Response	13.22044
Observations (or Sum Wgts)	2022

AICc	BIC
4249.341	4344.437

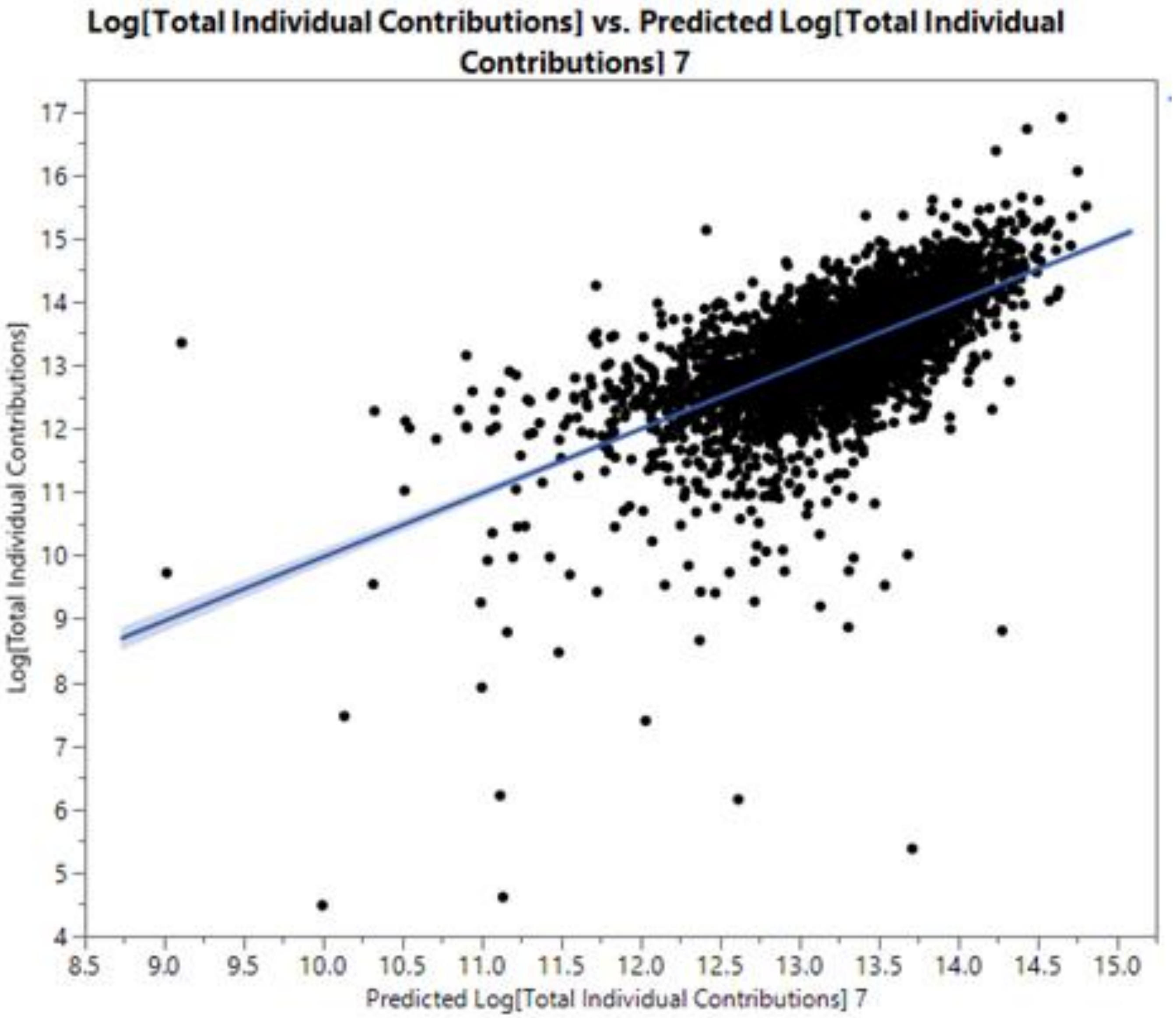
Analysis of Variance

Source	DF	Sum of Squares	Mean Square	F Ratio
Model	15	690.8056	46.0537	97.0429
Error	2006	951.9884	0.4746	Prob > F
C. Total	2021	1642.7940		<.0001*

Final Model

Final Model

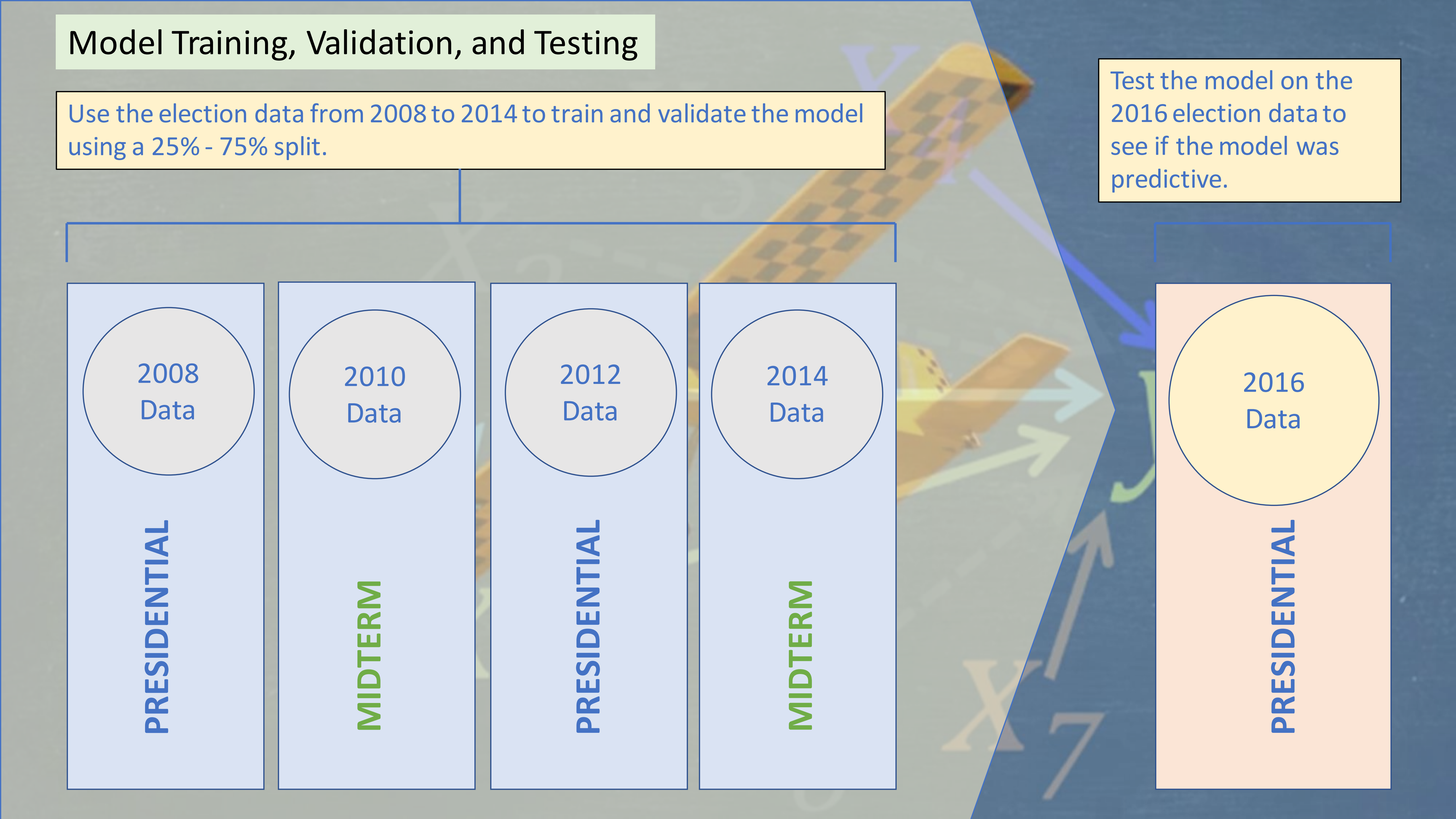
Predicted Values for 2016 vs. the Actual Values in 2016



Recommendations to Management

- ✓ The two most impactful marketing activities on individual donations to House races are events, and food and beverages. The presence of these two items at the top of the list may come as no surprise to election campaign managers, who often organize dinner events where the admission charge is a few hundred dollars a plate.
- ✓ Following that is TV and radio advertisement, marketing collateral, and promotional items. These go hand in hand with enhancing the personal brand of the candidate.
- ✓ Text blasting had a more significant impact on individual donations than door to door advertising, direct response marketing, bulk email, and mail. An opt-in text message option could keep donors involved in the campaign and help funnel additional revenue.
- ✓ Money spent on gifts to donors hurts donations. It's possible that donors view their donations as an investment in the candidate. For example, if a significant donor donates $1000 and receives a $100 gift, they may only give $900 next time. Therefore, if one of your more notable donors has a death in the family or you want to express your appreciation, it may often be better to send a card instead of a gift.
- ✓ Outdoor advertising does not yield a return on investment. It's better to sell signs and other outdoor advertisements, rather than give them away.
- ✓ Although a CRM (Contact Record Management) system is necessary, increased spending did not correlate to higher donations.
- ✓ When it comes to races for the US House of Representatives, spending money on social media advertising like Facebook had no impact on donations. It may have had an impact on voter behavior, but that was not analyzed at all. This effect may not be the same for Presidential candidates, who are generally well known.

BUILDING THE MODEL


Model Training, Validation, and Testing
Use the election data from 2008 to 2014 to train and validate the model using a 25% - 75% split.
Test the model on the 2016 election data to see if the model was predictive.
2008 Data
PRESIDENTIAL
2010 Data
MIDTERM
2012 Data
PRESIDENTIAL
2014 Data
MIDTERM
2016 Data
PRESIDENTIAL

Model Assumptions

Zero Mean

Constant Variance

Independence

Normality

Feature Selection

The initial variables were selected based on subject matter expertise and available data.

Adaptive LASSO was initially utilized for feature selection to slim down the list of variables in the model.

Untransformed variables invalidated the model, because of issues of non-constant variance.

An attempt was made to add binary variables as interaction terms, however it reduced the effectiveness of the model. The binary interaction terms had to be removed because they were either statistically insignificant or they had high variance inflation factor scores

Evaluated Two Models

Failed Model Assumptions

Model 1
(All Variables Model)
All binary, demographic, log transformed and non-transformed data variables were introduced to Adaptive LASSO for initial feature selection.

Model 2
(Only Log Transformed Variables Model)
Only binary, demographic and log transformed data were introduced to Adaptive LASSO for initial feature selection.

Model Fit
- ❑ Regression Standard Error
- ❑ Coefficient of Determination
- ❑ Analysis of Variance

Model Assumptions
- ❑ Zero Mean
- ❑ Constant Variance
- ❑ Independence
- ❑ Normality

Valid or Not Model
- ❑ Studentized Residuals
- ❑ Marginal Model Plots
- ❑ Outliers and Leverage Points
  - ❑ Cook's Distance
- ❑ Multicollinearity – VIF's

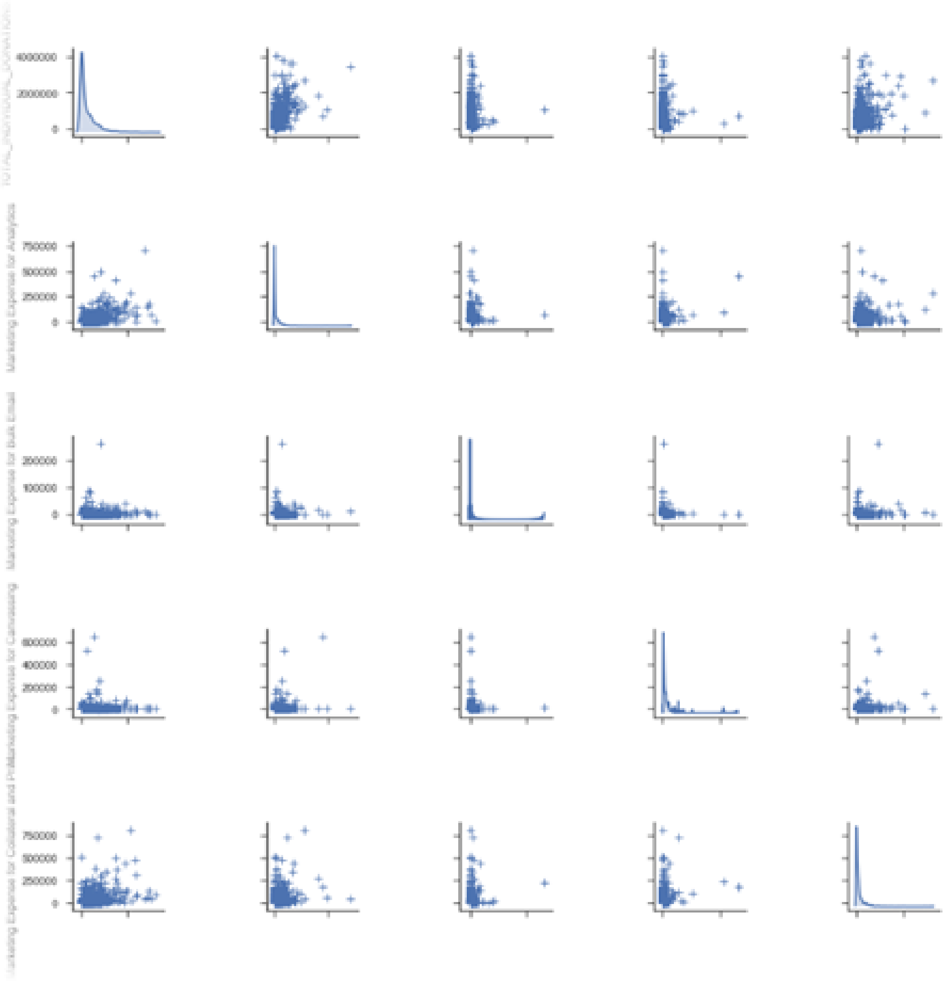
Model Building Steps – Details on the performance of each step follow the conclusions

Step 1	Draw scatter plots of the data. Decide whether transformations of Y or any of the x's needed.
Step 2	Run Adaptive LASSO. BIC Validation Method
Step 3	Analyze t-test and VIF's. Remove scores higher than 0.05. Assess the extent of collinearity. Remove terms that have high VIF scores above 10.
Step 4	Check Model Assumptions: Is the functional form of the model correct? (Asses the Zero Mean, Constant Variance, Independence, and Normality Assumptions)
Step 5	Assess the adequacy of the model: Evaluate the model fit numerically
Step 6	Attempt to add any binary or interaction variables based that might enhance the model.
Step 7	Check for outliers and leverage points. (Cook's Distance and Studentized Residuals)
Step 8	Assess the adequacy of the model: How do the marginal model plots look?

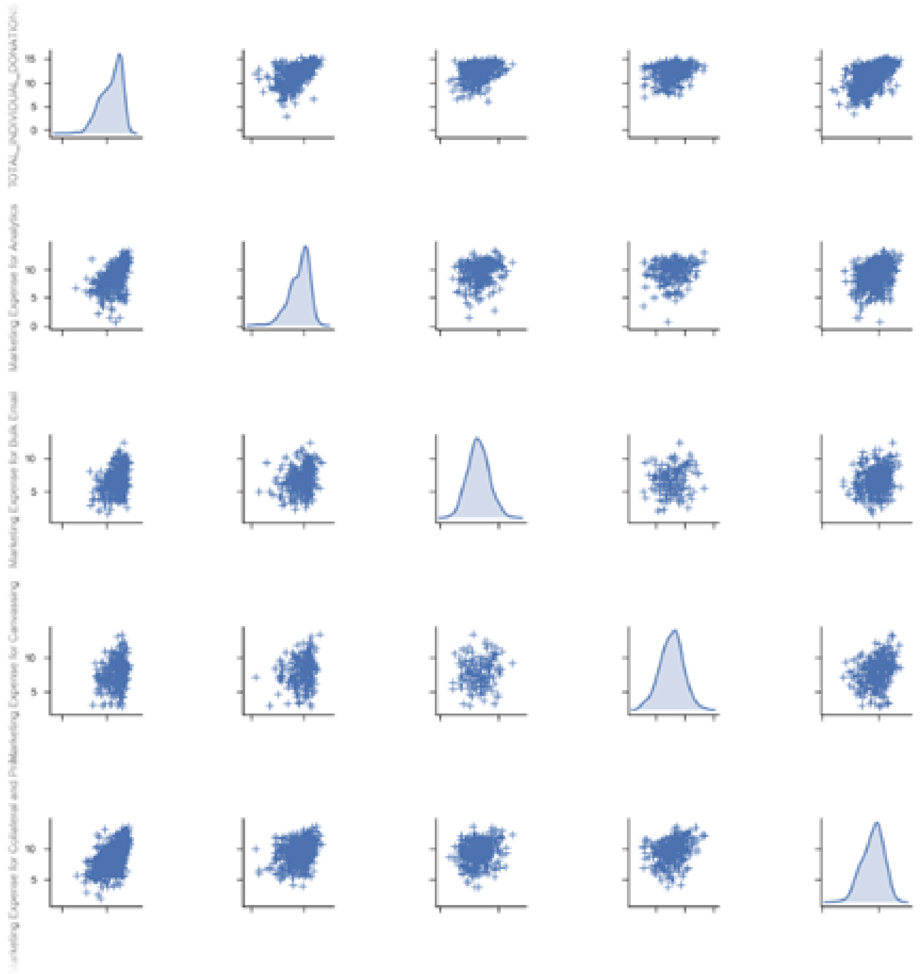
Step 1 **Draw scatter plots of the data. Decide whether transformations of Y or any of the x's needed.**

Log(x+1) transformation used to account for multiple 0 values.

Before Log Transformation



After Log Transformation



Step 2 **Feature Selection: Run Adaptive LASSO. BIC Validation Method**

Adaptive LASSO

Term	Estimate	Std Error	Prob > ChiSqua
TV and Radio	3.77E-07	3.86E-08	<.0001
Scale	0.644786	0.0103809	<.0001
Log[Food and Beverage]	0.0720556	0.0116194	<.0001
Log[Text Blasting]	0.0283648	0.0046338	<.0001
Promotional Items	6.56E-07	1.07E-07	<.0001
Log[Travel]	0.0458186	0.0076598	<.0001
Events	9.97E-07	2.07E-07	<.0001
Direct Response	4.60E-07	1.02E-07	<.0001
Log[Public Relations]	0.0173021	0.0043147	<.0001
Public Relations	2.26E-06	5.86E-07	0.0001
Log[Direct Response]	0.0116233	0.0031202	0.0002
Median Income	3.71E-06	1.02E-06	0.0003
Log[Collateral]	0.0332116	0.0100825	0.001
Log[Promotional Items]	0.0230995	0.0073274	0.0016
Log[Analytics]	0.0159027	0.0053229	0.0028
Log[Digital Assets]	0.0140056	0.0049043	0.0043
Log[Outdoor Advertising]	-0.012513	0.0046673	0.0073
Log[Bulk Email]	0.0112637	0.0046065	0.0145
Log[TV and Radio	0.0161934	0.0067261	0.0161
Log[Canvassing]	-0.008676	0.0043362	0.0454
Fundraising	4.40E-07	2.58E-07	0.0882
Log[Fundraising]	0.0059179	0.0045616	0.1945
Food and Beverage	6.52E-07	5.08E-07	0.1995
Log[Mail]	0.005723	0.0046533	0.2187
Social Media	1.05E-06	9.62E-07	0.2751
Bulk Email	-1.98E-06	1.94E-06	0.3084
Analytics	5.18E-07	5.12E-07	0.3116
Imagery	3.72E-07	3.99E-07	0.3514
Presidential[Binary][0-1]	0.0252497	0.0293099	0.389
Log[Social Media]	-0.004244	0.0053303	0.426
Fax Blasting	0.0000153	2.03E-05	0.4515
Log[Gifts]	-0.003737	0.005123	0.4657
Open[Binary][0-1]	-0.030189	0.0442252	0.4948
CRM	-8.18E-06	1.32E-05	0.5353
Land Area	6.71E-07	1.16E-06	0.5622
Intercept	39.828635	72.418957	0.5823
Log[CRM]	-0.006666	0.0123645	0.5898
Collateral	1.12E-07	2.27E-07	0.6233
Log[Fax Blasting]	0.0036396	0.008517	0.6691
Log[Imagery]	-0.002457	0.0057782	0.6706
Urban Population Percentage	-29.6371	72.374053	0.6822
Rural Population Percentage	-29.625	72.397596	0.6824
Digital Assets	2.65E-07	6.66E-07	0.6915
Outdoor Advertising	6.22E-07	1.60E-06	0.6982
Total Population	1.47E-07	4.17E-07	0.7236
Mail	6.21E-08	1.92E-07	0.7463
Canvassing	2.94E-07	9.14E-07	0.7479
Digital Advertising	4.26E-07	1.59E-06	0.789
Log[Phone Banking]	-0.001994	0.0074662	0.7894
Log[Digital Advertising]	-0.002825	0.0127953	0.8252
Incumbent[Binary][0-1]	-0.008712	0.0424122	0.8372
Phone Banking	5.78E-08	3.97E-07	0.8843
Text Blasting	-8.28E-08	6.91E-07	0.9047
Gifts	1.09E-07	2.74E-06	0.9684
Travel	2.44E-08	6.77E-07	0.9712

Parameter Estimates

<table>
<tr><th>Term</th><th>Estimate</th><th>Std Error</th><th>t Ratio</th><th>Prob>|t|</th><th>Lower 95%</th><th>Upper 95%</th><th>VIF</th></tr>
<tr><td>Intercept</td><td>10.196362</td><td>0.1421</td><td>71.75</td><td><.0001*</td><td>9.917681</td><td>10.475043</td><td>.</td></tr>
<tr><td>TV and Radio</td><td>4.0068e-7</td><td>3.402e-8</td><td>11.78</td><td><.0001*</td><td>3.3397e-7</td><td>4.674e-7</td><td>1.7831252</td></tr>
<tr><td>Log[Food and Beverage]</td><td>0.0778401</td><td>0.010006</td><td>7.78</td><td><.0001*</td><td>0.0582159</td><td>0.0974644</td><td>1.5320666</td></tr>
<tr><td>Promotional Items</td><td>7.1241e-7</td><td>1.001e-7</td><td>7.12</td><td><.0001*</td><td>5.1605e-7</td><td>9.0876e-7</td><td>1.23187</td></tr>
<tr><td>Log[Travel]</td><td>0.0475617</td><td>0.006955</td><td>6.84</td><td><.0001*</td><td>0.0339226</td><td>0.0612008</td><td>1.474513</td></tr>
<tr><td>Direct Response</td><td>5.7337e-7</td><td>8.425e-8</td><td>6.81</td><td><.0001*</td><td>4.0815e-7</td><td>7.386e-7</td><td>1.2082565</td></tr>
<tr><td>Log[Text Blasting]</td><td>0.0273063</td><td>0.004181</td><td>6.53</td><td><.0001*</td><td>0.0191059</td><td>0.0355066</td><td>1.1061324</td></tr>
<tr><td>Events</td><td>1.1335e-6</td><td>1.88e-7</td><td>6.03</td><td><.0001*</td><td>7.6488e-7</td><td>1.5022e-6</td><td>1.1511698</td></tr>
<tr><td>Median Income</td><td>3.7721e-6</td><td>9.043e-7</td><td>4.17</td><td><.0001*</td><td>1.9986e-6</td><td>5.5456e-6</td><td>1.0823828</td></tr>
<tr><td>Public Relations</td><td>2.365e-6</td><td>5.766e-7</td><td>4.10</td><td><.0001*</td><td>1.2343e-6</td><td>3.4957e-6</td><td>1.27637</td></tr>
<tr><td>Log[Collateral]</td><td>0.0343902</td><td>0.008631</td><td>3.98</td><td><.0001*</td><td>0.0174627</td><td>0.0513177</td><td>1.2276864</td></tr>
<tr><td>Log[Analytics]</td><td>0.0186068</td><td>0.004749</td><td>3.92</td><td><.0001*</td><td>0.009293</td><td>0.0279207</td><td>1.30396</td></tr>
<tr><td>Log[Public Relations]</td><td>0.0162487</td><td>0.004251</td><td>3.82</td><td>0.0001*</td><td>0.0079112</td><td>0.0245862</td><td>1.3009769</td></tr>
<tr><td>Log[Digital Assets]</td><td>0.0157345</td><td>0.004459</td><td>3.53</td><td>0.0004*</td><td>0.00699</td><td>0.024479</td><td>1.1306413</td></tr>
<tr><td>Log[Direct Response]</td><td>0.0106016</td><td>0.003026</td><td>3.50</td><td>0.0005*</td><td>0.0046665</td><td>0.0165366</td><td>1.2473968</td></tr>
<tr><td>Log[Outdoor Advertising]</td><td>-0.012304</td><td>0.003824</td><td>-3.22</td><td>0.0013*</td><td>-0.019804</td><td>-0.004804</td><td>1.2033316</td></tr>
<tr><td>Log[Promotional Items]</td><td>0.0218684</td><td>0.007179</td><td>3.05</td><td>0.0023*</td><td>0.0077903</td><td>0.0359466</td><td>1.323011</td></tr>
<tr><td>Log[TV and Radio</td><td>0.0174883</td><td>0.006591</td><td>2.65</td><td>0.0080*</td><td>0.0045621</td><td>0.0304145</td><td>1.8273766</td></tr>
<tr><td>Log[Fundraising]</td><td>0.0093283</td><td>0.003887</td><td>2.40</td><td>0.0165*</td><td>0.0017053</td><td>0.0169512</td><td>1.1721337</td></tr>
<tr><td>Log[Bulk Email]</td><td>0.0095517</td><td>0.004037</td><td>2.37</td><td>0.0181*</td><td>0.0016338</td><td>0.0174697</td><td>1.0942186</td></tr>
<tr><td>Log[Social Media]</td><td>-0.003895</td><td>0.005066</td><td>-0.77</td><td>0.4421</td><td>-0.01383</td><td>0.00604</td><td>1.0865758</td></tr>
<tr><td>Incumbent[Binary][0]</td><td>-0.012696</td><td>0.017331</td><td>-0.73</td><td>0.4639</td><td>-0.046684</td><td>0.021293</td><td>1.455803</td></tr>
<tr><td>Log[Phone Banking]</td><td>-0.001803</td><td>0.007</td><td>-0.26</td><td>0.7968</td><td>-0.015532</td><td>0.0119259</td><td>1.2643979</td></tr>
</table>

Remove:
Log[Bulk Email]
Incumbent[Binary]
Log[Social Media]

Model shrinkage from 54 to 22 terms

Adaptive Lasso, using BIC as the validation method, is the best variable shrinkage method to create an informative parametric model that errs on the side of under fitting.

Step 3

Analyze t-test and VIF's. Remove scores higher than 0.05. Assess the extent of collinearity. Remove terms that have high VIF scores above 10.

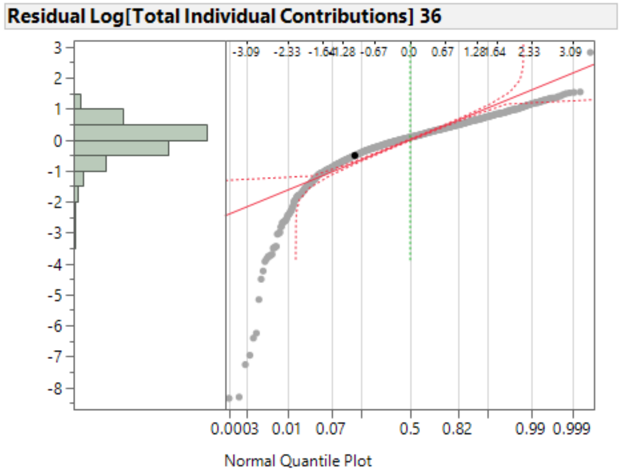
All VIF's are below 10, thus showing no signs of multicollinearity.

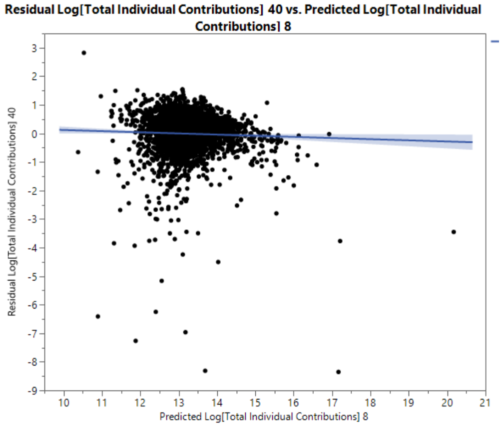
Parameter Estimates

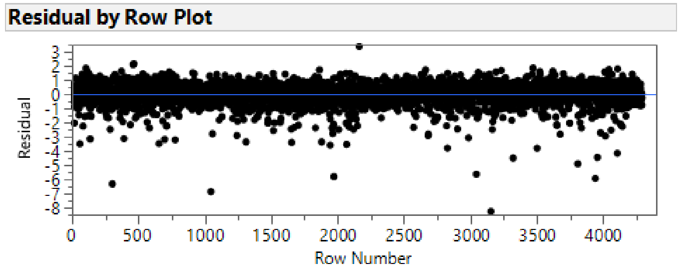
Term	Estimate	Std Error	t Ratio	Prob>|t|	Lower 95%	Upper 95%	VIF
Intercept	10.169299	0.138733	73.30	<.0001*	9.8972226	10.441376	.
Median Income	3.7802e-6	8.996e-7	4.20	<.0001*	2.0159e-6	5.5444e-6	1.0792239
Direct Response	5.6981e-7	8.405e-8	6.78	<.0001*	4.0499e-7	7.3464e-7	1.2049973
Events	1.1482e-6	1.861e-7	6.17	<.0001*	7.8315e-7	1.5132e-6	1.1313483
Promotional Items	7.1028e-7	9.993e-8	7.11	<.0001*	5.143e-7	9.0625e-7	1.2297088
Public Relations	2.3711e-6	5.754e-7	4.12	<.0001*	1.2427e-6	0.0000035	1.2750993
TV and Radio	0.0000004	3.37e-8	11.87	<.0001*	3.3389e-7	4.6605e-7	1.7780426
Log[Travel]	0.0482245	0.006778	7.12	<.0001*	0.0349325	0.0615166	1.4046729
Log[Food and Beverage]	0.079178	0.009694	8.17	<.0001*	0.0601656	0.0981903	1.4442774
Log[Direct Response]	0.0106786	0.003017	3.54	0.0004*	0.0047625	0.0165948	1.2442996
Log[Fundraising]	0.0097541	0.003731	2.61	0.0090*	0.0024367	0.0170716	1.0845115
Log[Analytics]	0.0181439	0.004696	3.86	0.0001*	0.0089337	0.0273541	1.2792853
Log[Bulk Email]	0.0090564	0.004014	2.26	0.0242*	0.001184	0.0169288	1.0853428
Log[Collateral]	0.0342751	0.008549	4.01	<.0001*	0.0175096	0.0510405	1.2083589
Log[Digital Assets]	0.01558	0.004452	3.50	0.0005*	0.0068487	0.0243112	1.1299959
Log[Outdoor Advertising]	-0.013257	0.003711	-3.57	0.0004*	-0.020535	-0.005979	1.1359663
Log[Promotional Items]	0.0225572	0.007068	3.19	0.0014*	0.0086959	0.0364185	1.2859394
Log[Public Relations]	0.0158653	0.004226	3.75	0.0002*	0.007577	0.0241536	1.2897744
Log[TV and Radio]	0.016783	0.006501	2.58	0.0099*	0.0040327	0.0295333	1.7850742
Log[Text Blasting]	0.0273838	0.004174	6.56	<.0001*	0.019197	0.0355706	1.1061856

Step 4

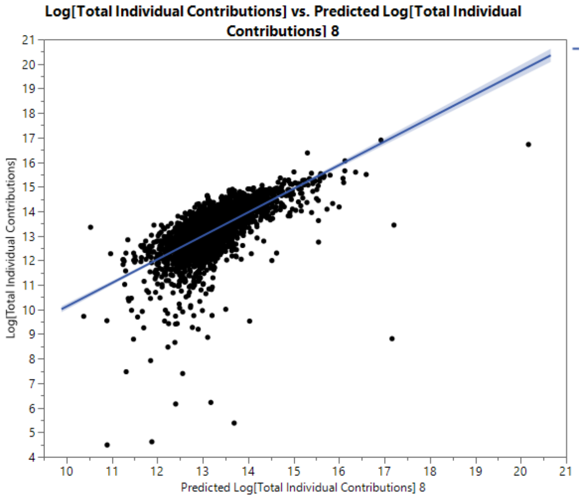
Check Model Assumptions: Is the functional form of the model correct? (Asses the Zero Mean, Constant Variance, Independence, and Normality Assumptions)







This is not a valid model!



- **Check Zero Mean** – The zero mean assumption does not hold. The residuals do not average out to zero as we move across the plot from left to right.
- **Check Constant Variance** – There is evidence of non-constant variance. The variation of the residuals is not similar as we move across the residual plot from left to right.
- **Check Independence** – The independence assumption does not hold – The residuals do not look random.
- **Check Normality** – An analysis of the QQ-plot shows the normality assumption does hold. The histogram is approximately bell shaped and symmetric. The QQ plots are relatively close to the line.

SOLUTION: With non-constant variance, it’s necessary to use transformed data. In this case, it’s important to use Adaptive LASSO for feature selection of only the log transformed data, binary and demographic data in a new model.

Step 2 **Feature Selection: Run Adaptive LASSO. BIC Validation Method**

Adaptive LASSO

Term	Estimate	Std Error	Prob > ChiSquare
Log[TV and Radio]	0.0557369	0.0061268	<.0001
Scale	0.6892142	0.0110025	<.0001
Log[Food and Beverage]	0.0869921	0.0108062	<.0001
Log[Promotional Items]	0.0497646	0.0071669	<.0001
Log[Events]	0.0810085	0.0125343	<.0001
Log[Public Relations]	0.0260219	0.0042453	<.0001
Log[Travel]	0.0452762	0.0075809	<.0001
Log[Direct Response]	0.0180501	0.003065	<.0001
Log[Text Blasting]	0.0266615	0.0045282	<.0001
Log[Analytics]	0.0262991	0.0050723	<.0001
Log[Outdoor Advertising]	-0.020451	0.0041955	<.0001
Median Income	4.75E-06	1.07E-06	<.0001
Log[Collateral]	0.0398376	0.0093644	<.0001
Log[Digital Assets]	0.0135796	0.0048386	0.005
Log[Gifts]	-0.01156	0.0049285	0.019
Log[CRM]	-0.017549	0.0086314	0.042
Log[Fundraising]	0.0080966	0.0042845	0.0588
Log[Bulk Email]	0.0082804	0.004382	0.0588
Log[Mail]	0.0081252	0.0044488	0.0678
Log[Fax Blasting]	0.0100788	0.0079617	0.2055
Presidential[Binary][0-1]	0.0384673	0.0311472	0.2168
Log[Digital Advertising]	0.0123213	0.0119833	0.3038
Log[Social Media]	-0.004981	0.005488	0.3641
Log[Canvassing]	-0.00361	0.0041334	0.3825
Open[Binary][0-1]	-0.039745	0.0471179	0.3989
Intercept	38.745922	76.528892	0.6127
Land Area	5.21E-07	1.22E-06	0.6695
Urban Population Percentage	-30.01865	76.478332	0.6947
Rural Population Percentage	-29.97396	76.502825	0.6952
Incumbent[Binary][0-1]	0.0169617	0.0447545	0.7047
Log[Imagery]	-0.001797	0.0057695	0.7554
Total Population	-1.52E-08	4.42E-07	0.9726
Log[Phone Banking]	0.0001238	0.0076303	0.9871

Parameter Estimates

Term	Estimate	Std Error	t Ratio	Prob>|t|	Lower 95%	Upper 95%	VIF
Intercept	8.7536182	0.142478	61.44	<.0001*	8.4741974	9.033039	.
Median Income	4.7313e-6	9.595e-7	4.93	<.0001*	2.8496e-6	6.613e-6	1.0691533
Log[Travel]	0.0466189	0.007433	6.27	<.0001*	0.0320421	0.0611958	1.4708739
Log[Food and Beverage]	0.0866791	0.010548	8.22	<.0001*	0.0659922	0.1073659	1.4902084
Log[Direct Response]	0.0182515	0.003013	6.06	<.0001*	0.0123425	0.0241604	1.0806747
Log[Analytics]	0.0280557	0.004949	5.67	<.0001*	0.0183493	0.0377621	1.2401431
Log[CRM]	-0.017672	0.008536	-2.07	0.0385*	-0.034412	-0.000932	1.0254116
Log[Collateral]	0.0385243	0.009133	4.22	<.0001*	0.020614	0.0564346	1.1995673
Log[Digital Assets]	0.0150689	0.004774	3.16	0.0016*	0.0057074	0.0244305	1.1323418
Log[Events]	0.0846836	0.012017	7.05	<.0001*	0.0611156	0.1082516	1.4771977
Log[Gifts]	-0.010066	0.00454	-2.22	0.0267*	-0.01897	-0.001162	1.2230298
Log[Outdoor Advertising]	-0.018304	0.003959	-4.62	<.0001*	-0.026068	-0.01054	1.1259863
Log[Promotional Items]	0.0508741	0.007035	7.23	<.0001*	0.0370768	0.0646713	1.108474
Log[Public Relations]	0.0270696	0.004166	6.50	<.0001*	0.0189001	0.0352391	1.0899787
Log[TV and Radio]	0.0562073	0.00591	9.51	<.0001*	0.0446163	0.0677983	1.2840112
Log[Text Blasting]	0.0279273	0.004458	6.26	<.0001*	0.0191838	0.0366708	1.1017282

Model shrinkage from 31 to 16 terms

Adaptive Lasso, using BIC as the validation method, is the best variable shrinkage method to create an informative parametric model that errs on the side of under fitting.

Step 3

Analyze t-test and VIF's. Remove scores higher than 0.05. Assess the extent of collinearity. Remove terms that have high VIF scores above 10.

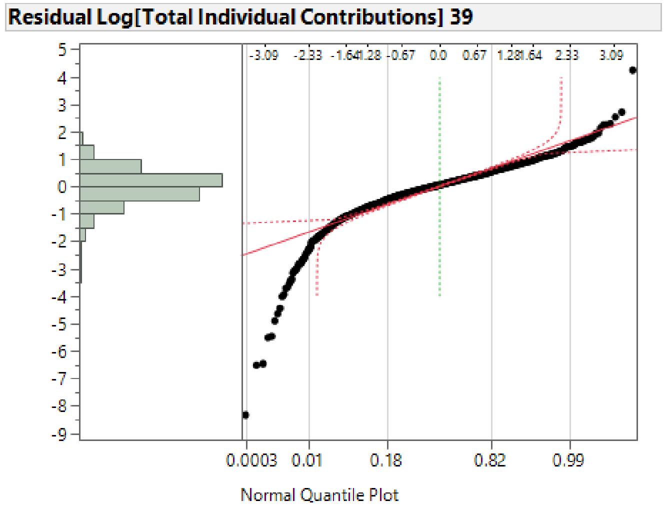
All VIF's are below 10, thus showing no signs of multicollinearity.

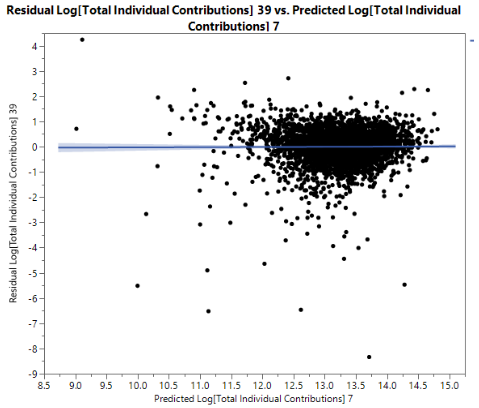
Parameter Estimates

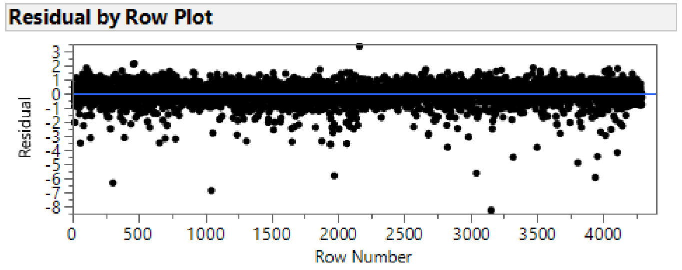
Term	Estimate	Std Error	t Ratio	Prob>|t|	Lower 95%	Upper 95%	VIF
Intercept	8.7536182	0.142478	61.44	<.0001*	8.4741974	9.033039	.
Log[Food and Beverage]	0.0866791	0.010548	8.22	<.0001*	0.0659922	0.1073659	1.4902084
Log[Events]	0.0846836	0.012017	7.05	<.0001*	0.0611156	0.1082516	1.4771977
Log[TV and Radio	0.0562073	0.00591	9.51	<.0001*	0.0446163	0.0677983	1.2840112
Log[Promotional Items]	0.0508741	0.007035	7.23	<.0001*	0.0370768	0.0646713	1.108474
Log[Travel]	0.0466189	0.007433	6.27	<.0001*	0.0320421	0.0611958	1.4708739
Log[Collateral]	0.0385243	0.009133	4.22	<.0001*	0.020614	0.0564346	1.1995673
Log[Analytics]	0.0280557	0.004949	5.67	<.0001*	0.0183493	0.0377621	1.2401431
Log[Text Blasting]	0.0279273	0.004458	6.26	<.0001*	0.0191838	0.0366708	1.1017282
Log[Public Relations]	0.0270696	0.004166	6.50	<.0001*	0.0189001	0.0352391	1.0899787
Log[Direct Response]	0.0182515	0.003013	6.06	<.0001*	0.0123425	0.0241604	1.0806747
Log[Digital Assets]	0.0150689	0.004774	3.16	0.0016*	0.0057074	0.0244305	1.1323418
Median Income	4.7313e-6	9.595e-7	4.93	<.0001*	2.8496e-6	6.613e-6	1.0691533
Log[Gifts]	-0.010066	0.00454	-2.22	0.0267*	-0.01897	-0.001162	1.2230298
Log[CRM]	-0.017672	0.008536	-2.07	0.0385*	-0.034412	-0.000932	1.0254116
Log[Outdoor Advertising]	-0.018304	0.003959	-4.62	<.0001*	-0.026068	-0.01054	1.1259863

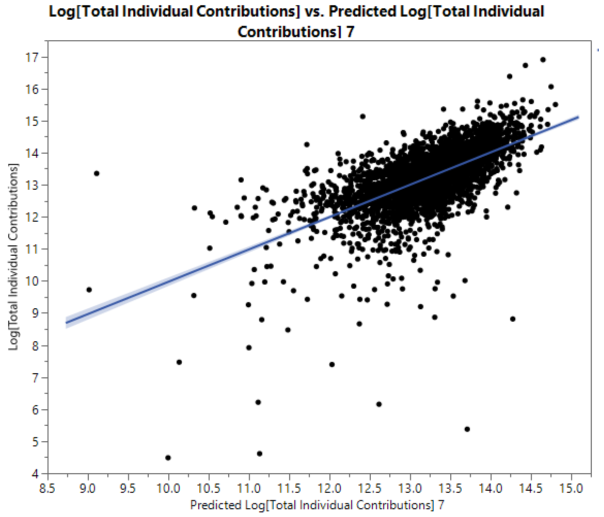
Step 4

Check Model Assumptions: Is the functional form of the model correct?
(Asses the Zero Mean, Constant Variance, Independence, and Normality Assumptions)









Assumptions Hold

- **Check Zero Mean** – The zero mean assumption does hold. The residuals average out to zero as we move across the plot from left to right.
- **Check Constant Variance** – There is little evidence of non constant variance. The variation of the residuals looks similar as we move across the plot from left to right.
- **Check Independence** – The residuals look random.
- **Check Normality** – This histogram looks approximately bells shaped, and symmetric, and the QQ-plot points lie close to the line.

Step 5 Attempt to add any binary or interaction variables based that might enhance the model.

Attempts to introduce binary terms to account for incumbent and challenger reduced the performance of the model in terms of AIC, BIC, and Adjusted R squared.

Summary of Fit	
RSquare	0.433383
RSquare Adj	0.425134
Root Mean Square Error	0.683584
Mean of Response	13.22044
Observations (or Sum Wgts)	2022

AICc	BIC
4232.599	4405.569

Summary of Fit	
RSquare	0.421204
RSquare Adj	0.416585
Root Mean Square Error	0.688648
Mean of Response	13.22044
Observations (or Sum Wgts)	2022

AICc	BIC
4248.943	4349.615

Parameter Estimates

Term	Estimate	Std Error	t Ratio	Prob>|t|	Lower 95%	Upper 95%	VIF
Intercept	8.8007058	0.152479	57.72	<.0001*	8.50		
Median Income	4.739e-6	9.583e-7	4.95	<.0001*	2.85		
Log[Travel]	0.0423209	0.008043	5.26	<.0001*	0.02		
Log[Food and Beverage]	0.0892753	0.010836	8.24	<.0001*	0.06		
Log[Direct Response]	0.0169182	0.003014	5.61	<.0001*	0.01		
Log[Analytics]	0.0236314	0.005248	4.50	<.0001*	0.01		
Log[CRM]	-0.016139	0.008489	-1.90	0.0574	-0.0		
Log[Collateral]	0.0349488	0.009157	3.82	0.0001*	0.01		
Log[Digital Assets]	0.0159251	0.004788	3.33	0.0009*	0.00		
Log[Events]	0.085563	0.012516	6.84	<.0001*	0.06		
Log[Gifts]	-0.005813	0.004975	-1.17	0.2428	-0.0		
Log[Outdoor Advertising]	-0.020536	0.00402	-5.11	<.0001*	-0.0		
Log[Promotional Items]	0.0554823	0.007202	7.70	<.0001*	0.04		
Log[Public Relations]	0.0264617	0.004146	6.38	<.0001*	0.01		
Log[TV and Radio	0.0586662	0.006169	9.51	<.0001*	0.04		
Log[Text Blasting]	0.0262246	0.004466	5.87	<.0001*	0.01		
Log[Travel]*Incumbent[Binary][0]	0.0133374	0.008026	1.66	0.0967	-0.0		
Log[Food and Beverage]*Incumbent[Binary][0]	0.0065123	0.01039	0.63	0.5309	-0.0		
Log[Direct Response]*Incumbent[Binary][0]	0.0016999	0.003014	0.56	0.5728	-0.0		
Log[Analytics]*Incumbent[Binary][0]	-0.009134	0.005255	-1.74	0.0823	-0.0		
Log[CRM]*Incumbent[Binary][0]	-0.008399	0.008486	-0.99	0.3224	-0.0		
Log[Collateral]*Incumbent[Binary][0]	-0.011234	0.008849	-1.27	0.2044	-0.0		
Log[Digital Assets]*Incumbent[Binary][0]	-0.003033	0.00478	-0.63	0.5259	-0.0		
Log[Events]*Incumbent[Binary][0]	0.0131621	0.011351	1.16	0.2464	[illegible]	[illegible]	[illegible]
Log[Gifts]*Incumbent[Binary][0]	0.0064756	0.004947	1.31	0.1907	-0.003226	0.0161773	2.609499
Log[Outdoor Advertising]*Incumbent[Binary][0]	-0.011275	0.004007	-2.81	0.0049*	-0.019134	-0.003417	3.706749
Log[Promotional Items]*Incumbent[Binary][0]	-0.027084	0.006939	-3.90	<.0001*	-0.040692	-0.013475	17.80658
Log[Public Relations]*Incumbent[Binary][0]	0.0089861	0.004139	2.17	0.0301*	0.000868	0.0171041	3.994102
Log[TV and Radio*Incumbent[Binary][0]	0.0080626	0.006058	1.33	0.1834	-0.003818	0.0199432	21.2266

Parameter Estimates

Term	Estimate	Std Error	t Ratio	Prob>|t|	Lower 95%	Upper 95%	VIF
Intercept	8.7266134	0.143484	60.82	<.0001*	8.44522	9.0080067	.
Median Income	4.671e-6	9.599e-7	4.87	<.0001*	2.7884e-6	6.5536e-6	1.0709011
Log[Travel]	0.0473226	0.007444	6.36	<.0001*	0.0327239	0.0619213	1.4763353
Log[Food and Beverage]	0.0883288	0.010598	8.33	<.0001*	0.0675448	0.1091129	1.5053106
Log[Direct Response]	0.0181548	0.003013	6.03	<.0001*	0.0122467	0.024063	1.0811353
Log[Analytics]	0.0275725	0.004957	5.56	<.0001*	0.0178504	0.0372946	1.2450389
Log[CRM]	-0.017707	0.008533	-2.08	0.0381*	-0.034441	-0.000973	1.0254187
Log[Collateral]	0.0379498	0.009137	4.15	<.0001*	0.0200312	0.0558685	1.2015338
Log[Digital Assets]	0.014769	0.004776	3.09	0.0020*	0.0054031	0.0241349	1.1341936
Log[Events]	0.0870305	0.012108	7.19	<.0001*	0.0632854	0.1107757	1.5005394
Log[Gifts]	-0.008067	0.004717	-1.71	0.0874	-0.017318	0.0011846	1.3212471
Log[Outdoor Advertising]	-0.019011	0.003984	-4.77	<.0001*	-0.026823	-0.011199	1.140861
Log[Promotional Items]	0.0518545	0.007061	7.34	<.0001*	0.0380067	0.0657023	1.1173929
Log[Public Relations]	0.0270377	0.004164	6.49	<.0001*	0.018871	0.0352044	1.0900051
Log[TV and Radio	0.0551982	0.005944	9.29	<.0001*	0.0435415	0.0668549	1.2995187
Log[Text Blasting]	0.0277542	0.004458	6.23	<.0001*	0.0190111	0.0364974	1.1024164
Log[Public Relations]*Incumbent[Binary][0]	0.0037003	0.002381	1.55	0.1203	-0.000969	0.0083696	1.3020204

Step 6

Assess the adequacy of the model: Evaluate the model fit numerically

Crossvalidation

Source	RSquare	RASE	Freq
Training Set	0.4205	0.68616	2022
Validation Set	0.4099	0.74078	701
Test Set	0.4200	0.79333	638

75% Training

25% Validation

Test

2008 Data | 2010 Data | 2012 Data | 2014 Data

2016 Data

Summary of Fit

RSquare	0.420507
RSquare Adj	0.416173
Root Mean Square Error	0.688891
Mean of Response	13.22044
Observations (or Sum Wgts)	2022

AICc	BIC
4249.341	4344.437

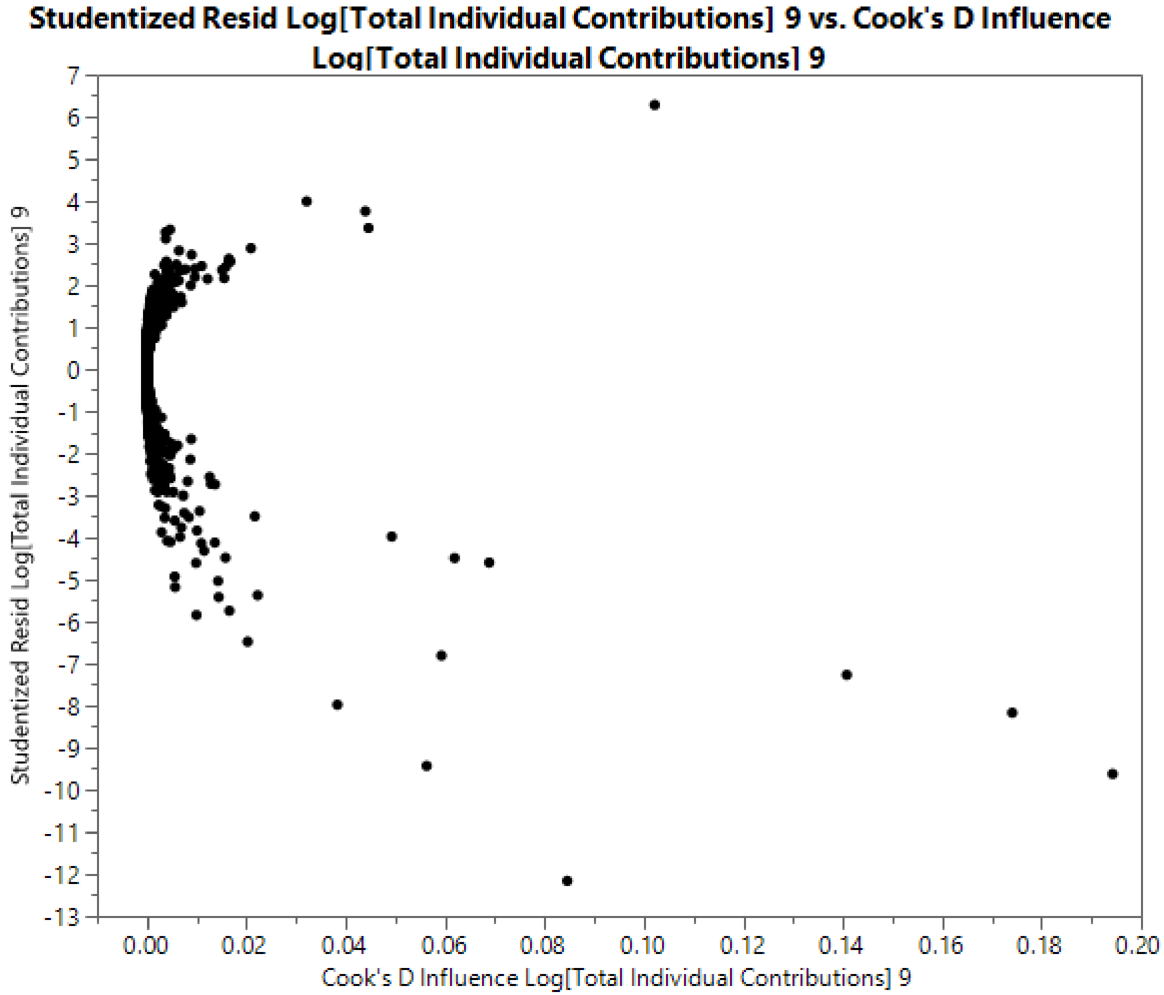
Analysis of Variance

Source	DF	Sum of Squares	Mean Square	F Ratio
Model	15	690.8056	46.0537	97.0429
Error	2006	951.9884	0.4746	Prob > F
C. Total	2021	1642.7940		<.0001*

Effect Summary

Source	LogWorth	PValue
Log[TV and Radio	20.280	0.00000
Log[Food and Beverage]	15.433	0.00000
Log[Promotional Items]	12.169	0.00000
Log[Events]	11.601	0.00000
Log[Public Relations]	9.990	0.00000
Log[Travel]	9.361	0.00000
Log[Text Blasting]	9.339	0.00000
Log[Direct Response]	8.784	0.00000
Log[Analytics]	7.783	0.00000
Median Income	6.053	0.00000
Log[Outdoor Advertising]	5.396	0.00000
Log[Collateral]	4.590	0.00003
Log[Digital Assets]	2.791	0.00162
Log[Gifts]	1.573	0.02673
Log[CRM]	1.414	0.03855

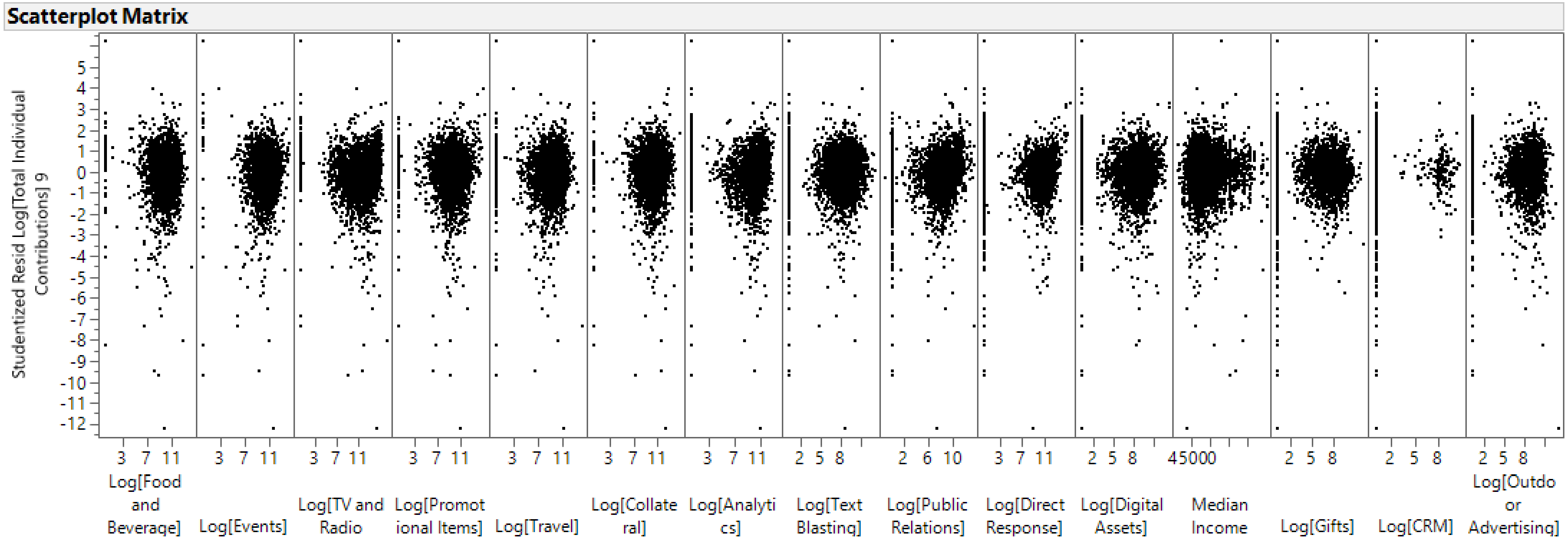
Step 7 **Check for outliers and leverage points. (Cook's Distance and Studentized Residuals)**



There are clearly outliers, however since none of them have a cook's distance greater than .5, none of them are leverage points. Therefore they should have a minimal impact on the model.

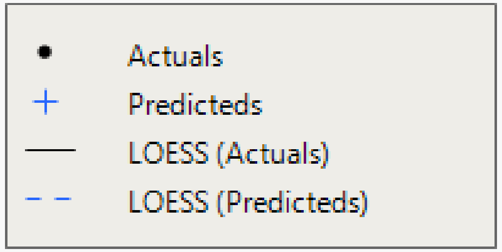
Step 7 **Check for outliers and leverage points. (Cook's Distance and Studentized Residuals)**

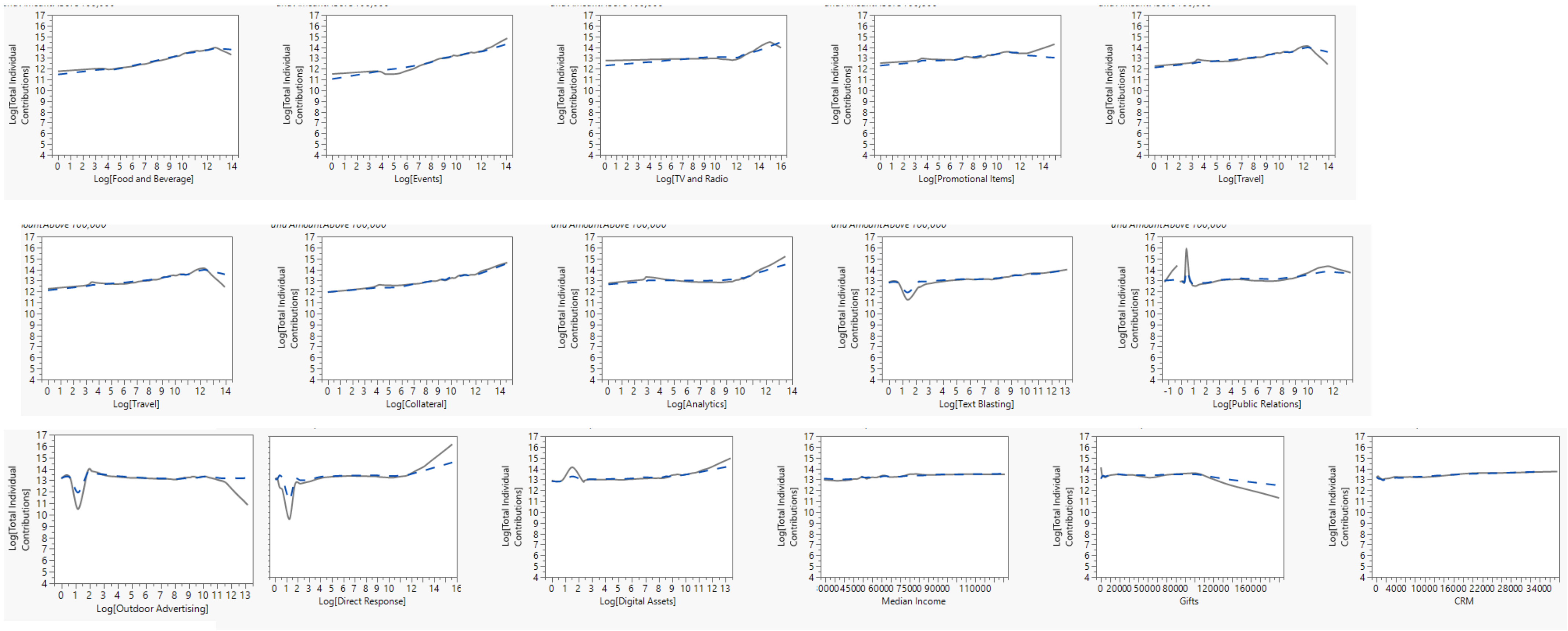
A quick check of the studentized residuals against each predictor shows a fairly random distribution, except for zero values. That's expected, since there are a lot in this dataset.



Step 8 Assess the adequacy of the model: How do the marginal model plots look?

All of the marginal model plots look good. The predicted overlay the actual in each one.





Statistical Conclusions

A LOOK AT THE MODEL

- $b_0 = 8.7536182$
- $b_1 = 0.0280557$
- $b_2 = 0.0385243$
- $b_3 = -0.017672$
- $b_4 = 0.0150689$
- $b_5 = 0.0182515$
- $b_6 = 0.0846836$
- $b_7 = 0.0866791$
- $b_8 = -0.010066$
- $b_9 = -0.018304$
- $b_{10} = 0.0508741$
- $b_{11} = 0.0270696$
- $b_{12} = 0.0279273$
- $b_{13} = 0.0466189$
- $b_{14} = 0.0562073$
- $b_{16} = 0.0000047313$

Log[Total Individual Donations] = b_0 +
+ b_1 x Log[Analytics]
+ b_2 x Log[Collateral]
+ b_3 x Log[CRM]
+ b_4 x Log[Digital Assets]
+ b_5 x Log[Direct Response]
+ b_6 x Log[Events]
+ b_7 x Log[Food and Beverage]
+ b_8 x Log[Gifts]
+ b_9 x Log[Outdoor Advertising]
+ b_{10} x Log[Promotional Items]
+ b_{911} x Log[Public Relations]
+ b_{12} x Log[Text Blasting]
+ b_{13} x Log[Travel]
+ b_{14} x Log[TV and Radio]
+ b_{15} x Median Income

Statistical Conclusions Continued

Impact of Individual Donations

Log(Total Individual Donations) = (b_n) (Log (Marketing Item))
A 1% increase in Marketing Item is predicted to produce a (1.01^{B_n} - 1) x 100 % change in Total Individual Donations

Impact of Median Income in a District

Ln(Total Individual Donations) = 0.0000047313 Median Income
= (d/d(Total Individual Donations) = d (Median Income) x (b_{16})
= 100 (d/d(Total Individual Donations) = 100 d (Median Income) x (b_{16})
= (Percentage Change of Total Individual Donations) = 100 d (Median Income) x (b_{16})
Therefore if Median Income were increased by one unit, I expect Total Individual Donations to increase by (0.0000047313 x 100) %.

- A 1% change in Food & Beverage spending is predicted to produce a 0.0863% change in total individual donations.
- A 1% change in Event spending is predicted to produce a 0.084% change in total individual donations.
- A 1% change in TV and Radio spending is predicted to produce a 0.056% change in total individual donations.
- A 1% change in Promo Item spending is predicted to produce a 0.0506% change in total individual donations.
- A 1% change in Travel spending is predicted to produce a change of 0.046% in total individual donations.
- A 1% change in Collateral spending is predicted to produce a change of 0.039% in total individual donations.
- A 1% change in Analytics spending is predicted to produce a change of 0.028% in total individual donations.
- A 1% change in Text Blasting spending is predicted to produce a change of 0.028% in total individual donations.
- A 1% change in Public Relations spending is predicted to produce a change of 0.0269% in total individual donations.
- A 1% change in Direct Response spending is predicted to produce a change of 0.018% in total individual donations.
- A 1% change in Digital Asset spending is predicted to produce a change of 0.015% in total individual donations.
- A 1% change in Gift spending is predicted to produce a change of -0.011% in total individual donations.
- A 1% change in CRM spending is predicted to produce a change of -0.02% in total individual donations.
- A 1% change in Outdoor Advertising is predicted to produce a change of -0.0182% in total individual donations.
- An increase of $1 dollar in median income in a congressional district is expected to increase total donations by (0.0000047313 x 100)%.

Additional Conclusions

Interesting Points from the Data

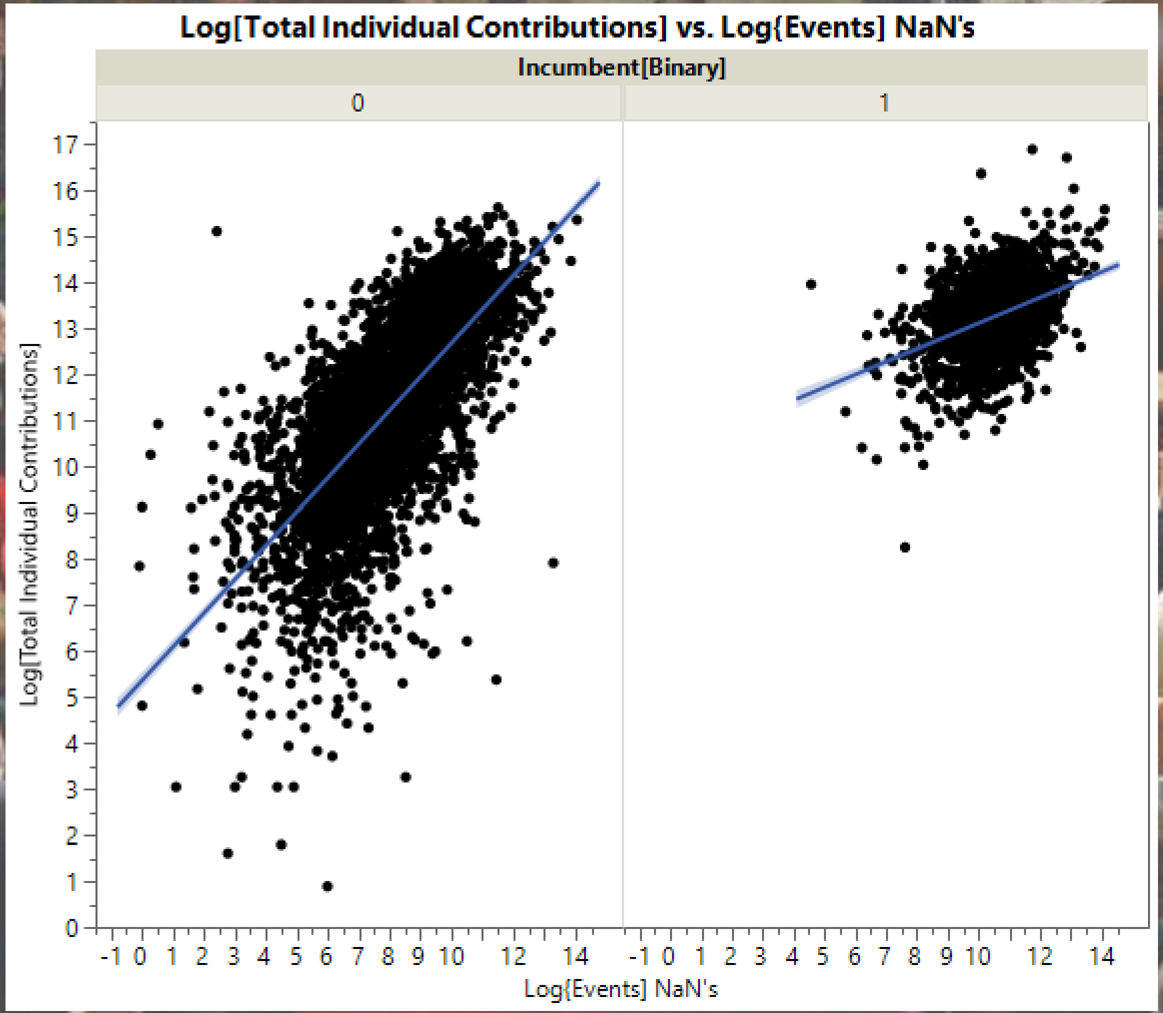
- Giving gifts to donors tends to reduce donations.
- Midterm and Presidential Cycles did not have as much of an impact as Incumbent vs. Challenger on individual donations to House races.
- Incumbents receive more money from individual donations than challengers until both are spending equal amounts on each marketing activity.
- Challengers have a larger capacity to earn more from marketing activities than incumbents.

Looking Forward

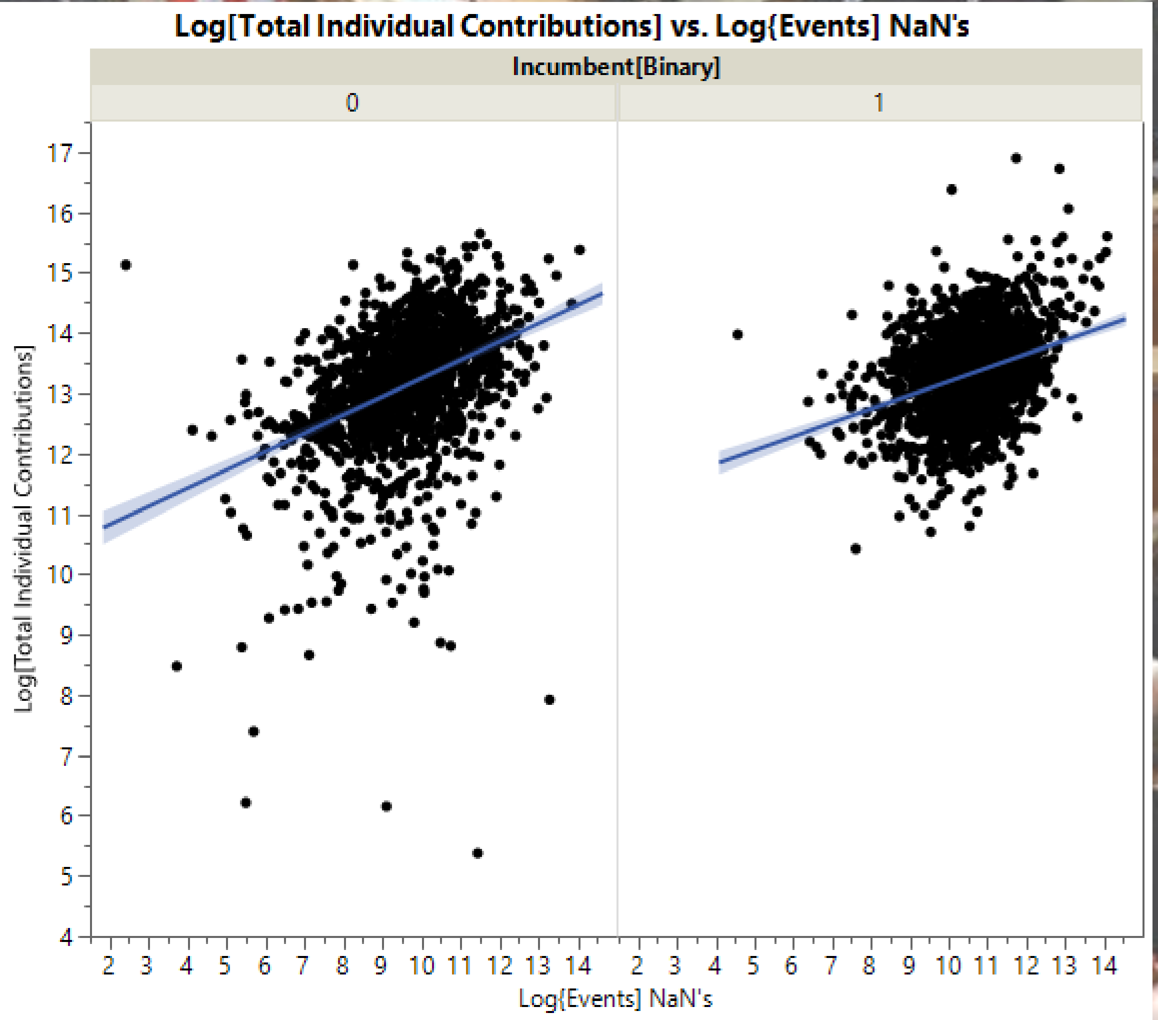
- The model may be improved by looking at more interaction variables with incumbents vs. challengers after figuring out a way to tighten up the accuracy of the parser.
- Tightening up the parser, to reduce white noise is essential to making the model more accurate.
- Advertisement tends to also compete with gubernatorial elections. In the future, a dummy variable could be added to reflect that.
- Analyzing sentiment from social media feeds could potentially add another element that would be predictive.
- Poll data could also add another element that would be predictive.

Additional Conclusions

Log [Events] and Log [Total Individual Contributions] Before $250,000 Cut Off

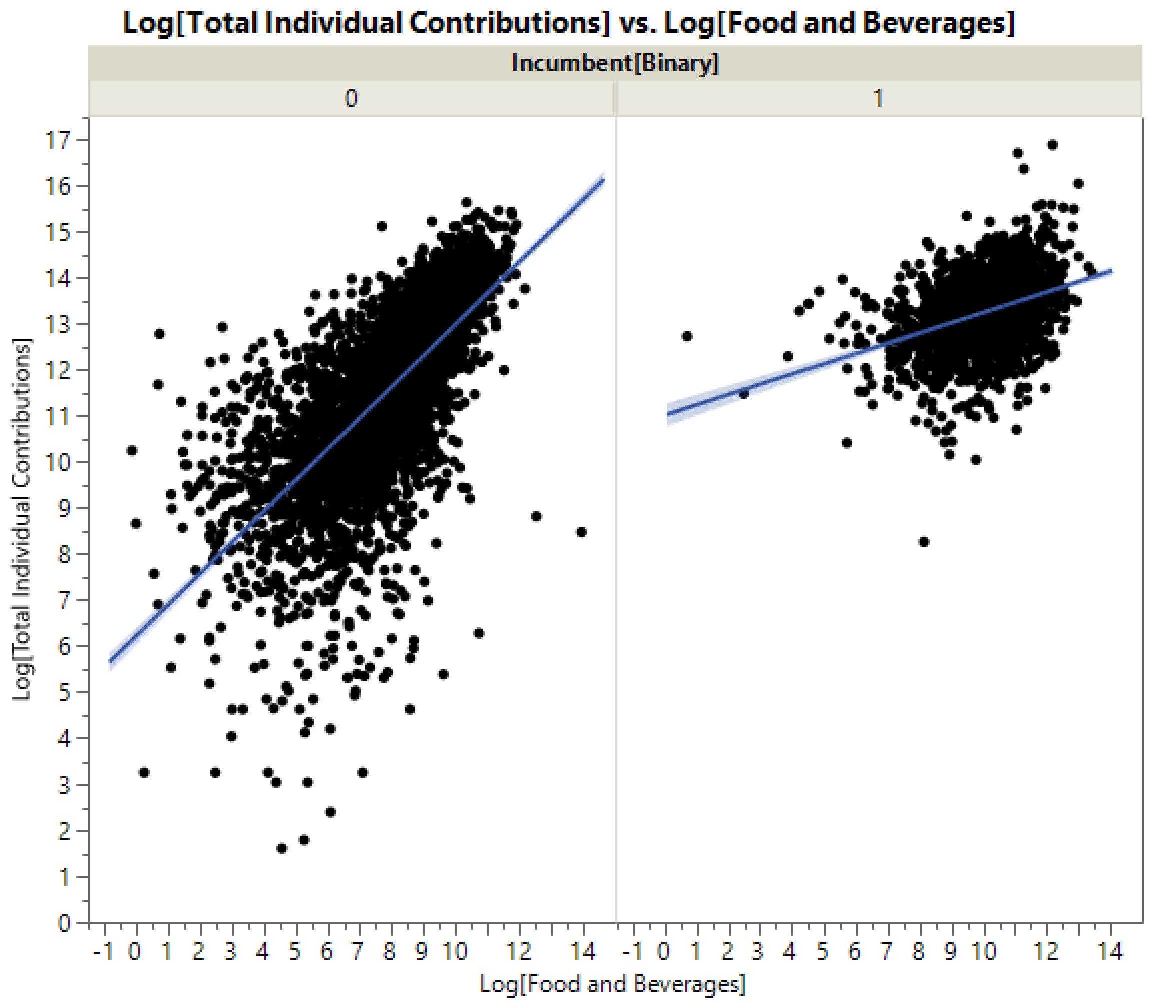


Log [Events] and Log [Total Individual Contributions] After $250,000 Cut Off

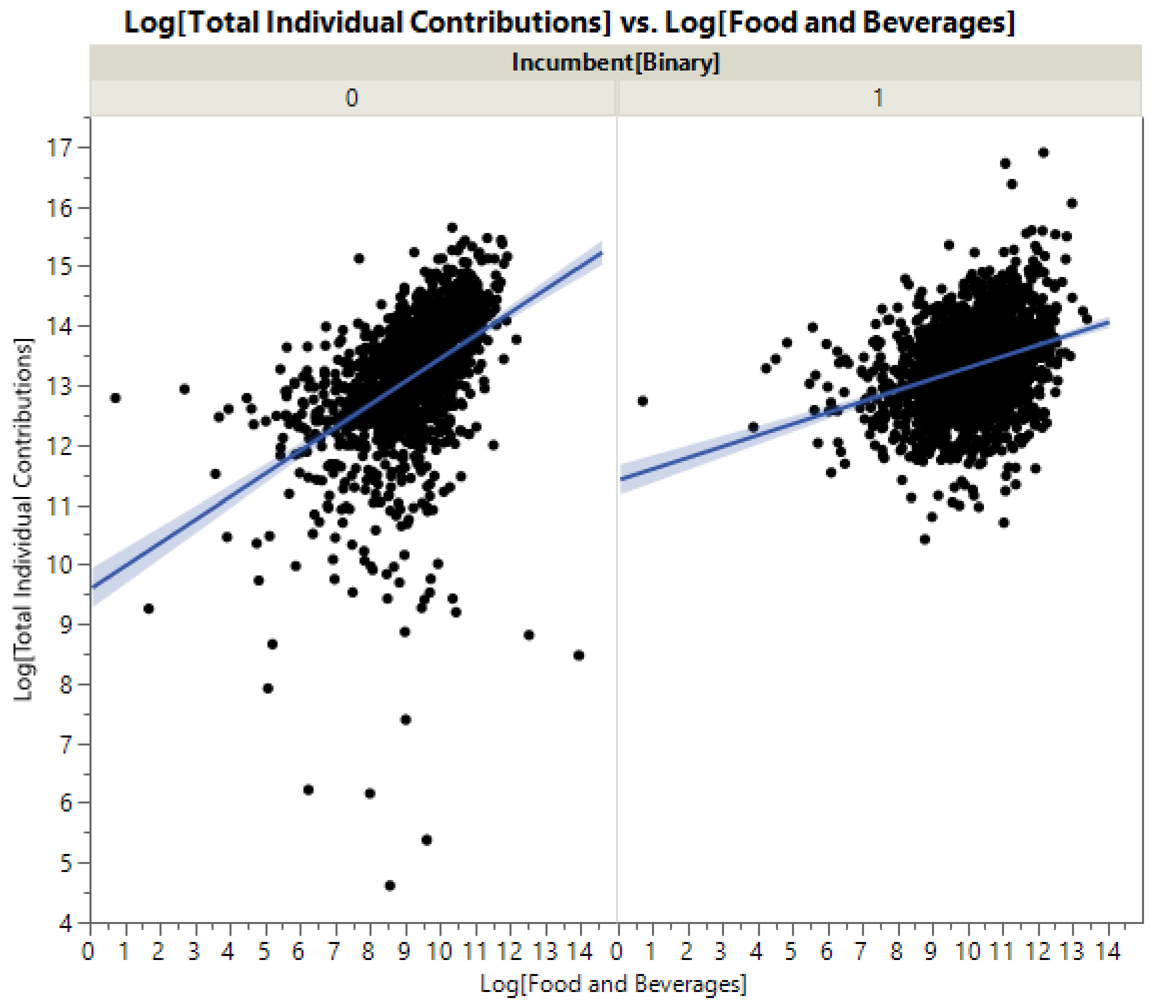


Additional Conclusions

Log [Food and Beverages] and Log [Total Individual Contributions] Before $250,000 Cut Off

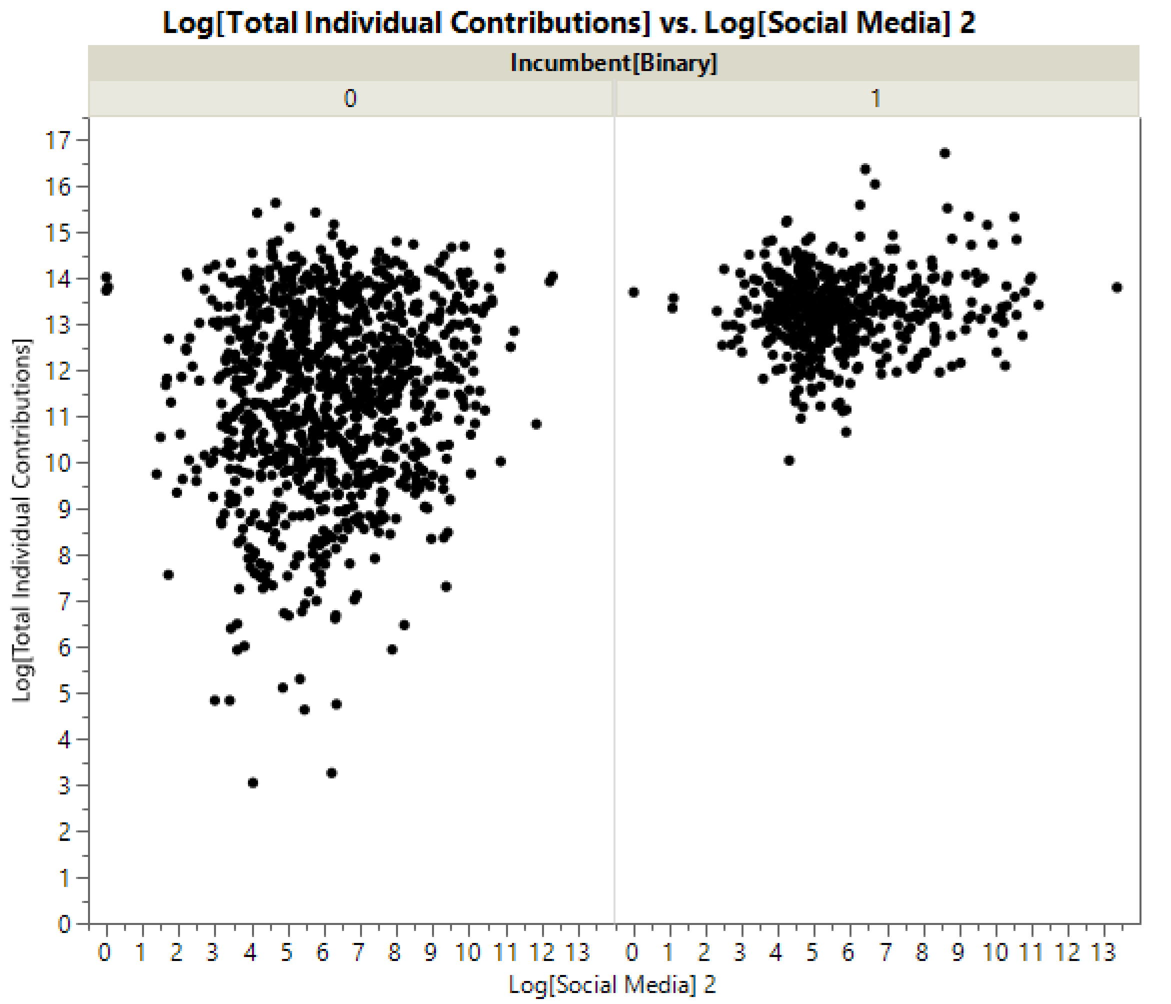


Log [Food and Beverages] and Log [Total Individual Contributions] After $250,000 Cut Off

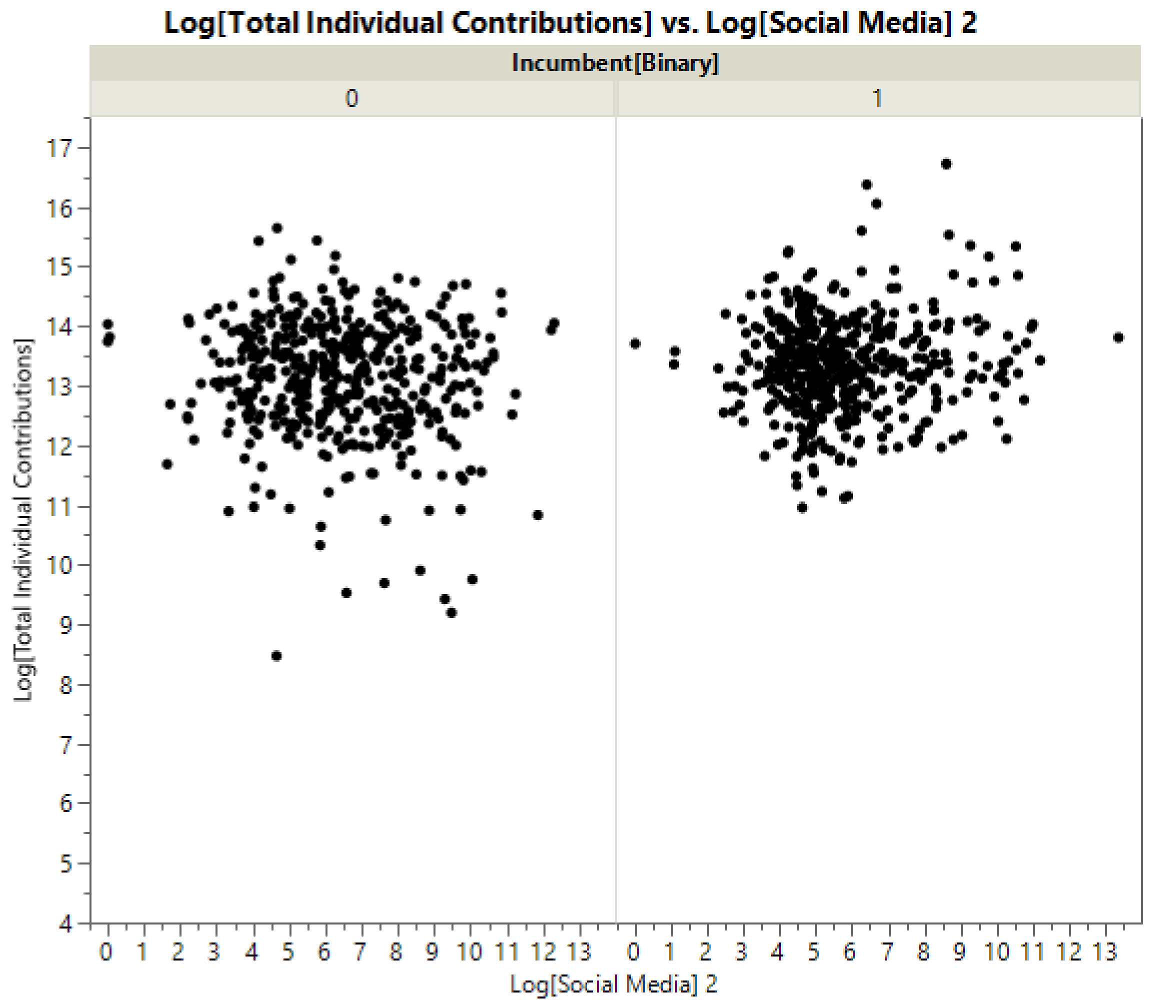


Additional Conclusions

**Log [Social Media]
and Log [Total Individual Contributions]
Before $250,000 Cut Off**

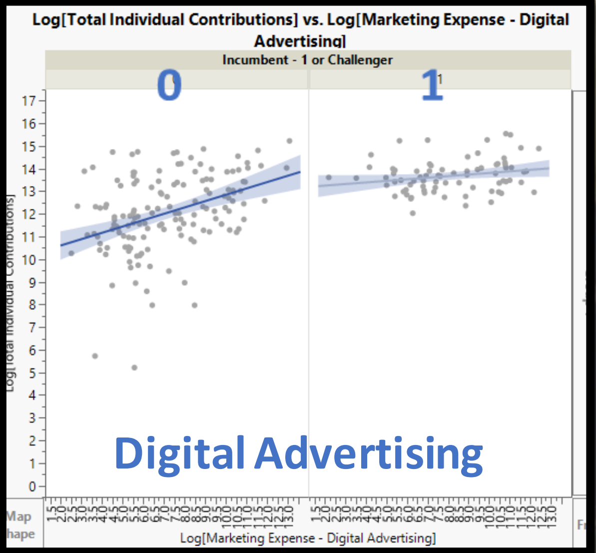


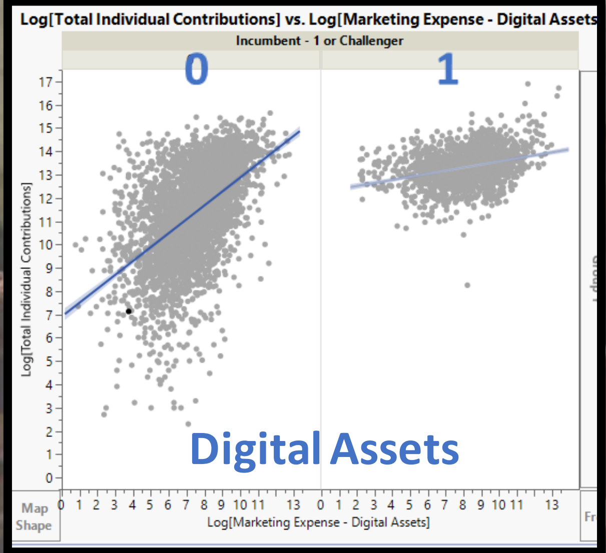
**Log [Social Media]
and Log [Total Individual Contributions]
After $250,000 Cut Off**

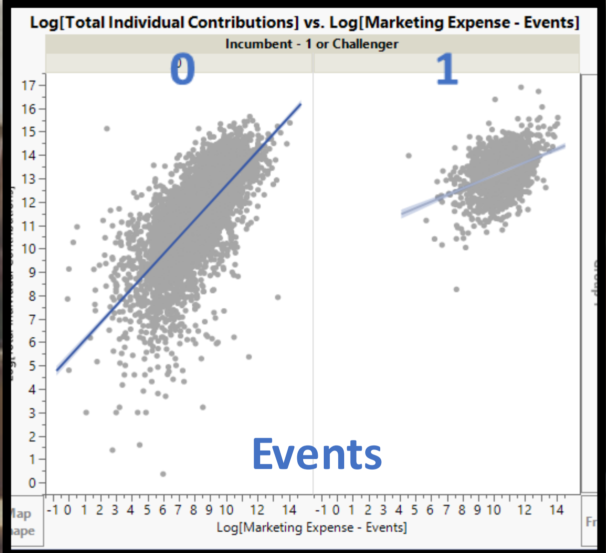


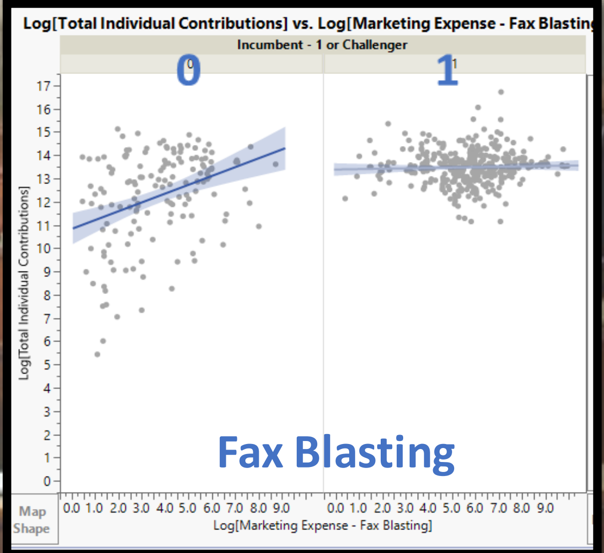
Challenger (0) vs. Incumbent (1) Effects on Marketing Expenses

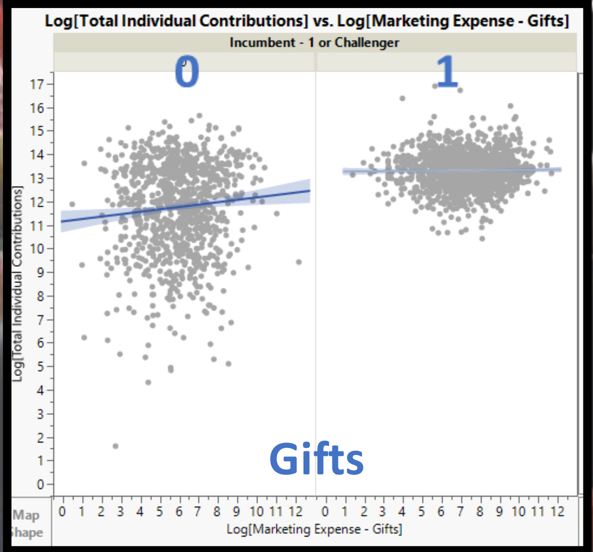
(All data – no cutoff)

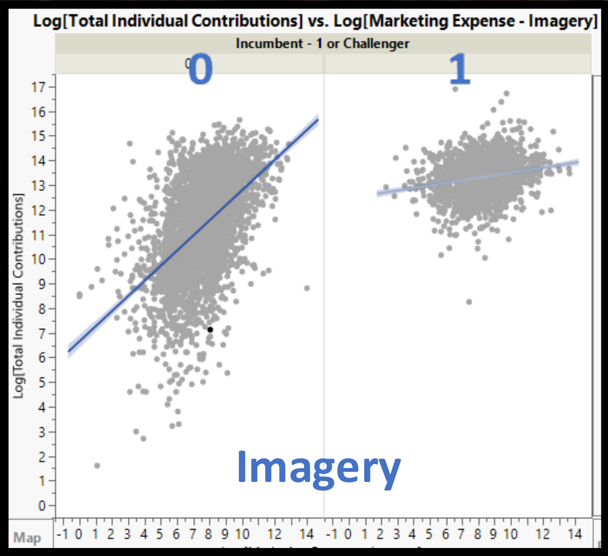


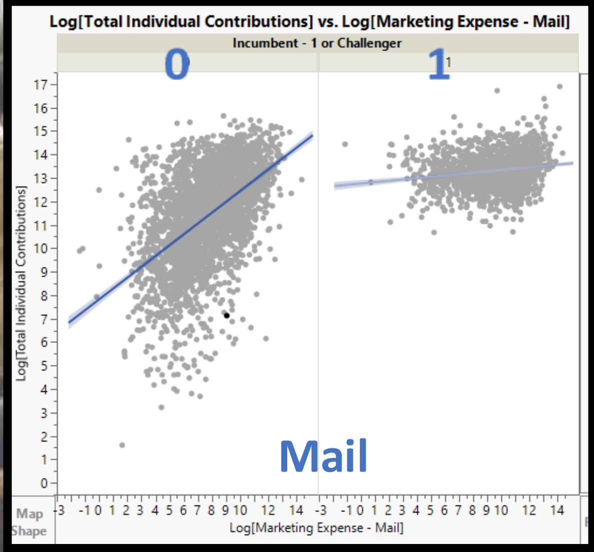


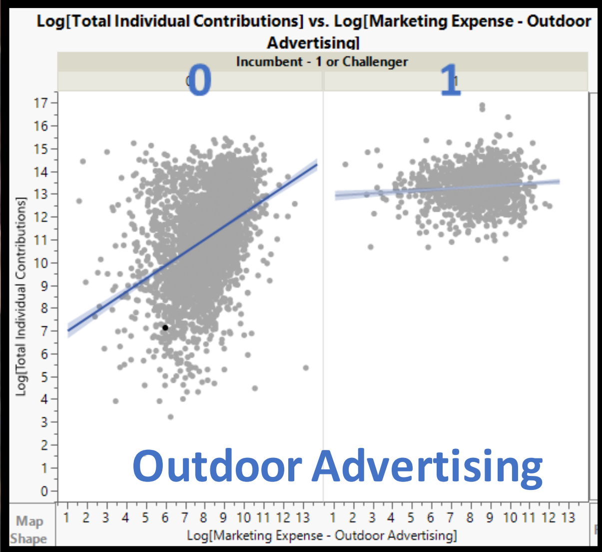






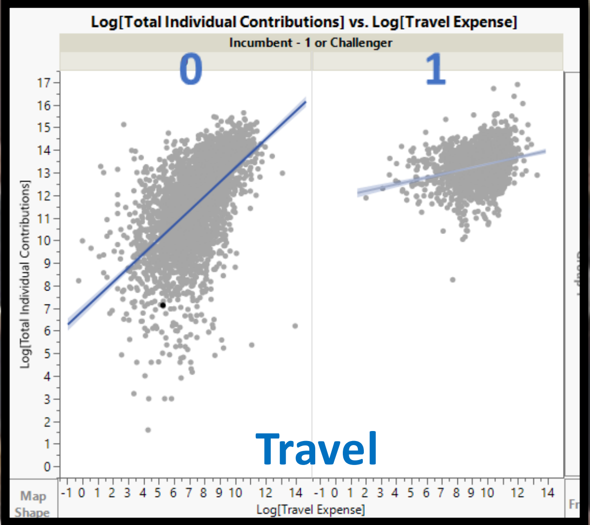


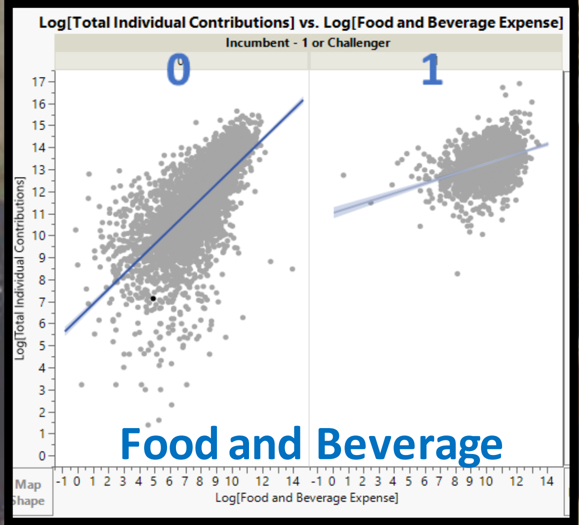


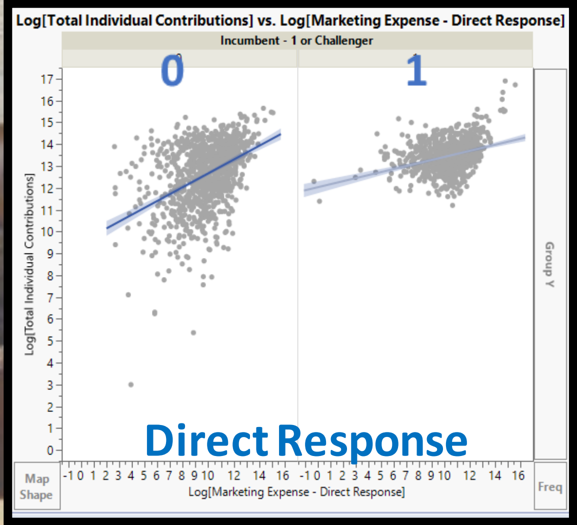


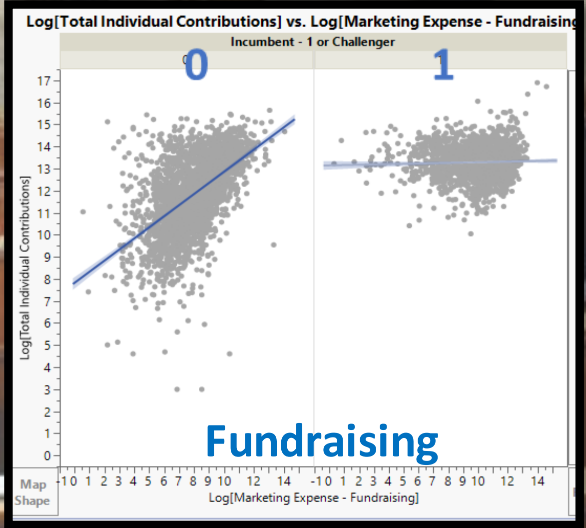
Challenger (0) vs. Incumbent (1) Effects on Marketing Expenses

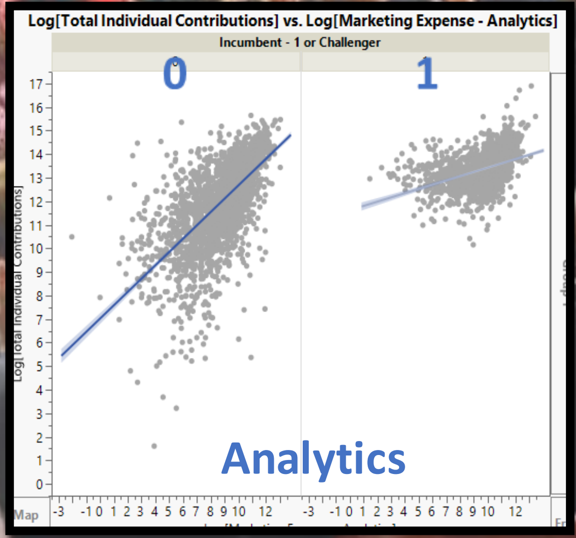
(All data – no cutoff)

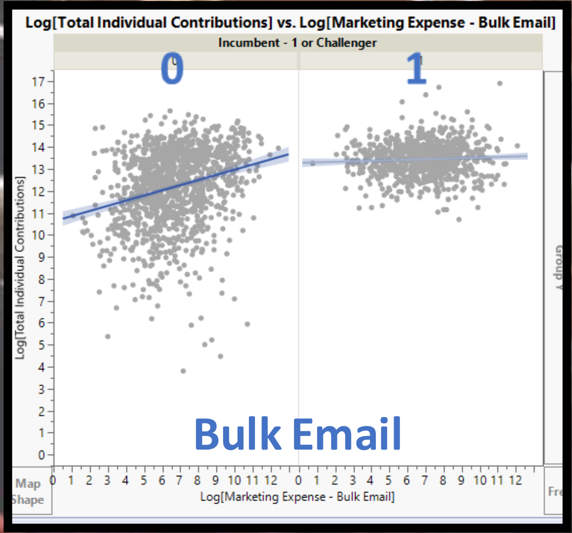


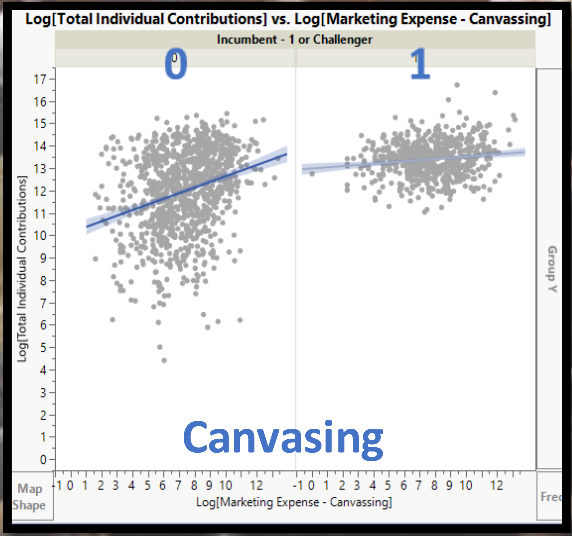


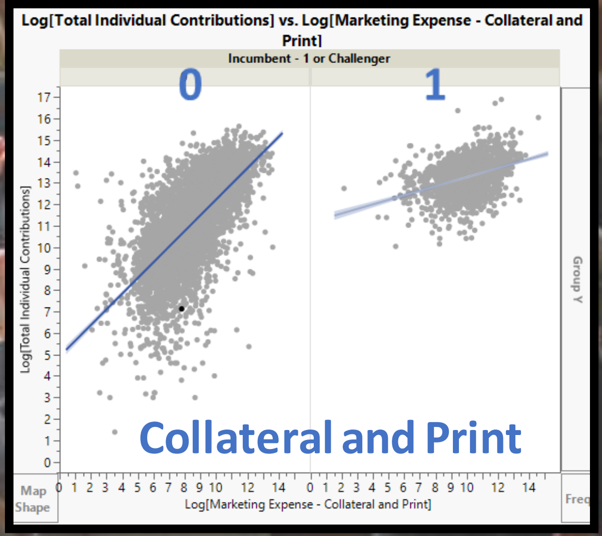












Challenger (0) vs. Incumbent (1) Effects on Marketing Expenses
(All data – no cutoff)