Correlation between County Political Preference and Big Box Retailers

By Pouya Mohammadi

Duke University

Honey, pick up some milk from the Republican store

Fact: Biden won the presid… with a Whole Foods and 32% of counties with a … Barrel - the widest gap ever.

11:12 AM · Dec 8, 2020 · Twitter Web App

10.4K Like

POLITICS • INTERACTIVE

Are You a J. Crew Democrat or a Pizza Hut Republican?

…als shop at Trader J…

…of Walmart

To Beat Trump, Democrats May Need to Break Out of the 'Whole Foods' Bubble

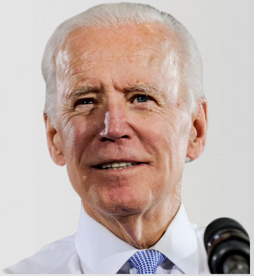
Trader Joe's De… Walmart Repub…

Liberal or Conservative? Where you shop reveals how you vote

Crate & Barrel, L.L.Bean and Sephora: Where you shop during the Christmas holiday season reveals a lot about how you vote.

Why the Hype?

- Computational politics is increasing in popularity
- Increased polarization leads to closer elections
    - Every advantage counts
- Easy, cheap, and targeted way to infer information about voters
    - Could be a future Waffle House Index for politics



Research Question

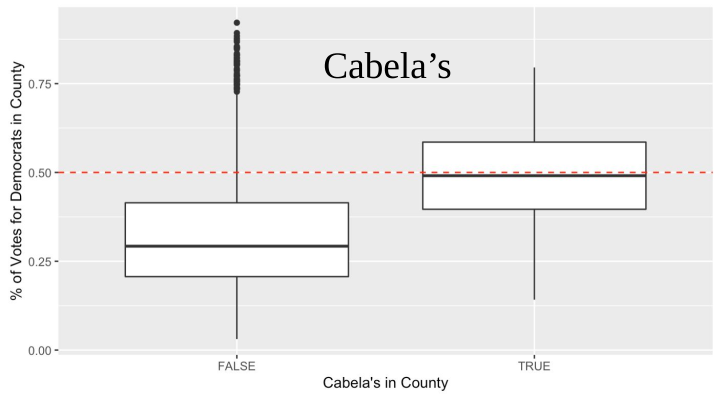
Is there a correlation between the presence of certain big box retailers in a county and that county's political preferences?

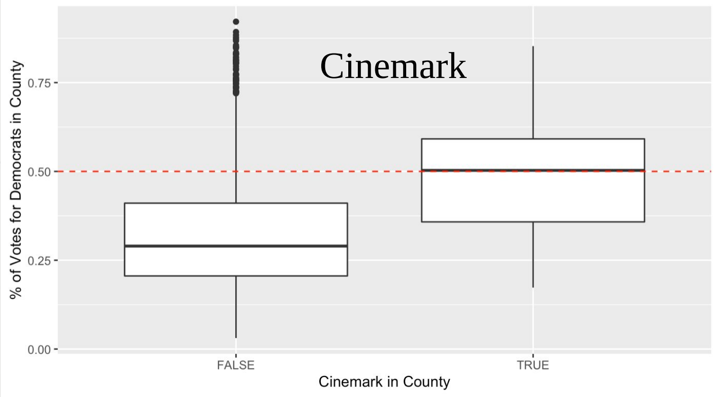
Previous Literature

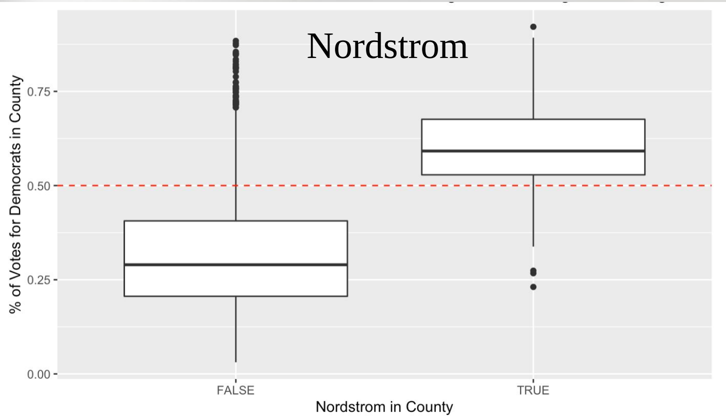
- Most research through investigative journalism
  - T-tests or comparisons of percentages
- One more in-depth analysis online by Aaron Lee
  - Looks at 60 different stores, using a random forest
- Academic journals used to determine covariates
- Our contribution: more rigorous, scientific analysis

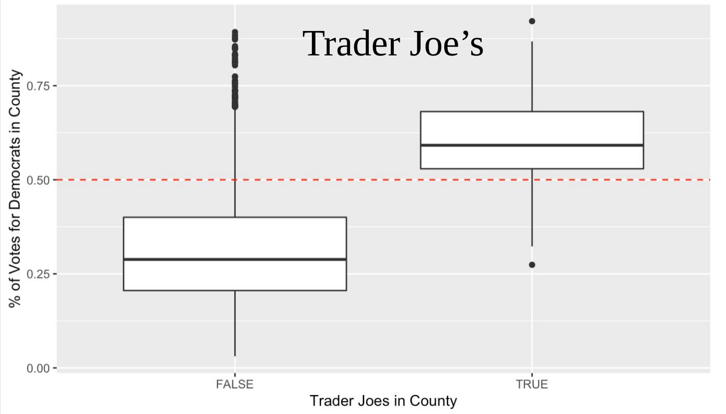
Our data

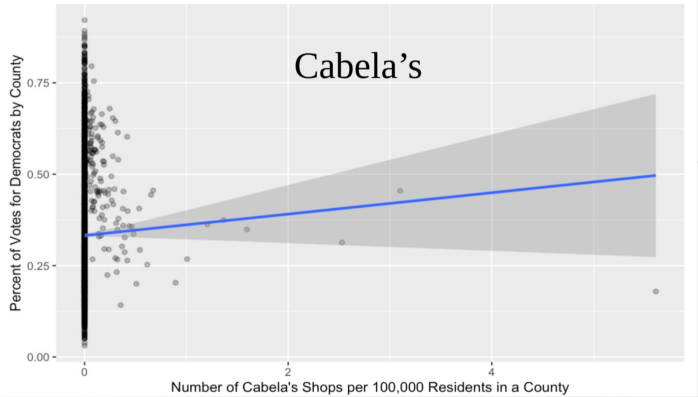
- Response Variable: % of Votes for Democrats in 2020
  - Provided by @tonmcg on Github
- Our stores: Trader Joe’s, Cinemark, Cabela’s, Nordstrom
  - Location data scraped from various online sources
  - Encoded as number of stores per 100k residents in county
- Census data used for other county variables
  - These variables determined from previous literature
- Alaska is removed from our dataset

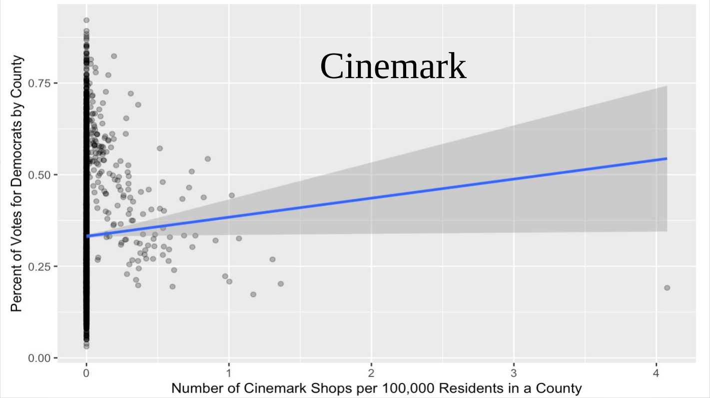

Cabela’s
% of Votes for Democrats in County
0.00
0.25
0.50
0.75
FALSE
TRUE
Cabela's in County

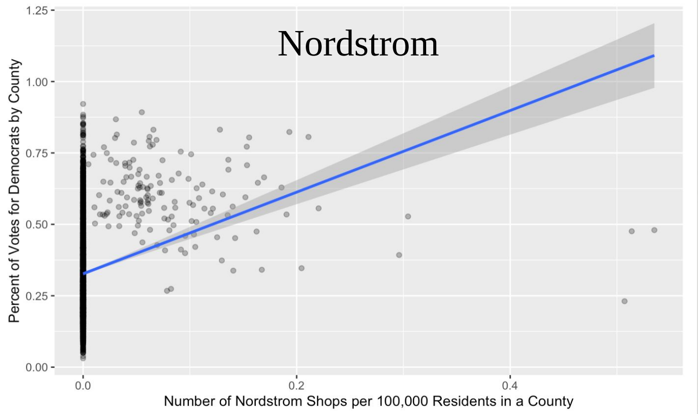

Cinemark
% of Votes for Democrats in County
0.00
0.25
0.50
0.75
FALSE
TRUE
Cinemark in County

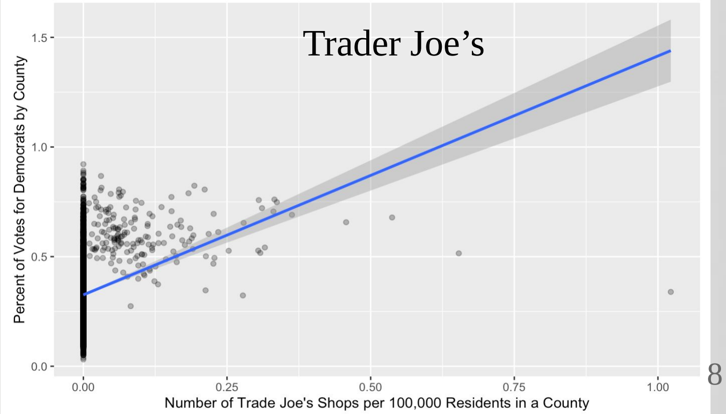

Nordstrom
% of Votes for Democrats in County
0.00
0.25
0.50
0.75
FALSE
TRUE
Nordstrom in County


Trader Joe’s
% of Votes for Democrats in County
0.00
0.25
0.50
0.75
FALSE
TRUE
Trader Joes in County


Cabela’s
Percent of Votes for Democrats by County
0.00
0.25
0.50
0.75
0
2
4
Number of Cabela's Shops per 100,000 Residents in a County


Cinemark
Percent of Votes for Democrats by County
0.00
0.25
0.50
0.75
0
1
2
3
4
Number of Cinemark Shops per 100,000 Residents in a County


Nordstrom
Percent of Votes for Democrats by County
0.00
0.25
0.50
0.75
1.00
1.25
0.0
0.2
0.4
Number of Nordstrom Shops per 100,000 Residents in a County

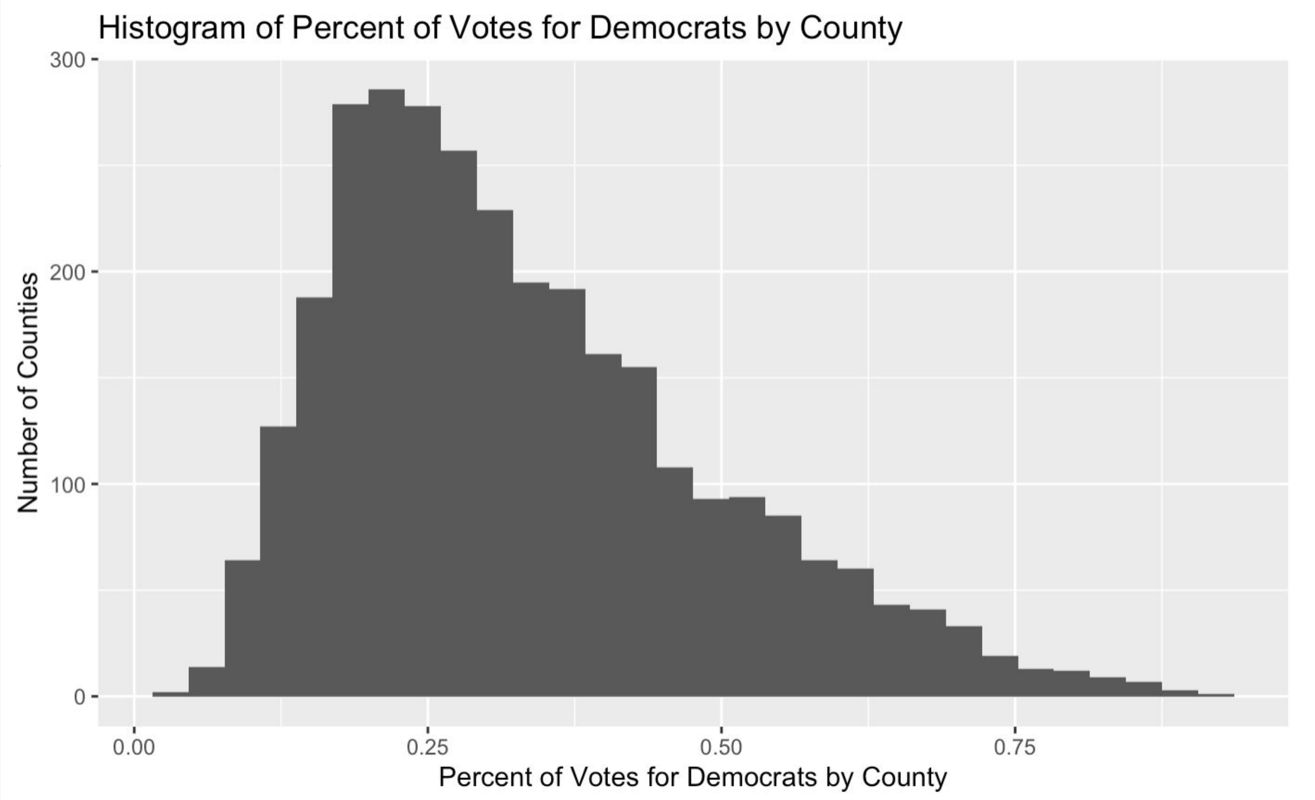

Trader Joe’s
Percent of Votes for Democrats by County
0.0
0.5
1.0
1.5
0.00
0.25
0.50
0.75
1.00
Number of Trade Joe's Shops per 100,000 Residents in a County

Model Choice

- Beta Regression
  - Response variable is % of votes for Democrats
- Other options
  - Count based models with offset (NB, Poisson)
    - Modeling % of votes for either party
  - Logistic Regression
    - Response variable is the party winner of the county
  - OLS model

Beta Regression

- Assumptions
  - Response follows a Beta distribution
  - Independence between observations
  - Linearity between predictors and response
- The response variable is an observation-specific mean for a Beta distribution, centered at that observation
  - Use Maximum Likelihood Estimation to optimize this and precision parameter


Histogram of Percent of Votes for Democrats by County
300
200
100
0
Number of Counties
0.00
0.25
0.50
0.75
Percent of Votes for Democrats by County

Model Equation

$$
\begin{aligned}
logit(\mu_i) = &\beta_0 + \beta_1 * (\#.of.Trader.Joes.per.100k.residents_i) + \beta_2 * (\#.of.Cabelas.per.100k.residents_i) + \\
&\beta_3 * (\#.of.Cinemark.per.100k.residents_i) + \beta_4 * (\#.of.Nordstrom.per.100k.residents_i) + \\
&\beta_5 * (Percent.White_i) + \beta_6 * (Percent.Black_.Population_i) + \beta_7 * (Percent.Hispanic_i) + \\
&\beta_8 * (Percent.Asian_{)} + \beta_9 * (Percent.Male_i) + \beta_{11} * (Percent.with.College.Education_i) + \\
&\beta_{12} * (Percent.in.Poverty_i) + \beta_{13} * (Percent.Unemployed_i) + \beta_{14} * (Urban.County_i = True)
\end{aligned}
$$

Where μ_i is the observation specific mean for a Beta distribution modeling the percentage of Democratic votes in a county and ϕ_i is estimated using $\phi_i = \log(\phi_i)$

Our Dataset

Independent Variables

- Number of Big Box Retailers per 100k Residents *
  - Trader Joe’s
  - Cabela’s
  - Cinemark
  - Nordstrom
- Rural vs. Urban **
- % Male
- % Poverty
- % Education
- Population % by race
  - Black
  - White
  - Asian
  - Hispanic
- % Unemployment

Dependent Variable

- Percent Votes for Joe Biden in 2020

* Variables of interest for which we want to draw inferences.
** Binary Variables. The rest of the variables are numeric

Results

	Estimates	Exp(Estimates)	Standard Errors	P-Values
Intercept	-0.953	0.385	0.264	**<0.001*****
Trader Joe's per 100k	1.285	3.614	0.223	**<0.001*****
Cabela's per 100k	0.002	1.002	0.059	0.979
Cinemarks per 100k	-0.090	0.914	0.079	0.253
Nordstroms per 100k	0.772	2.165	0.329	**0.019***
% White	-1.453	0.234	0.129	**<0.001*****
% Black	1.013	2.753	0.126	**<0.001*****
% Hispanic	-0.522	0.593	0.136	**<0.001*****
% Asian	4.253	70.341	0.418	**<0.001*****
% Male	-1.426	0.240	0.406	**<0.001*****
% Educated	0.028	1.028	0.001	**<0.001*****
% Poverty	-0.001	0.999	0.002	0.567
% Unemployed	0.110	1.117	0.007	**<0.001*****
Urban	0.132	1.141	0.020	**<0.001*****

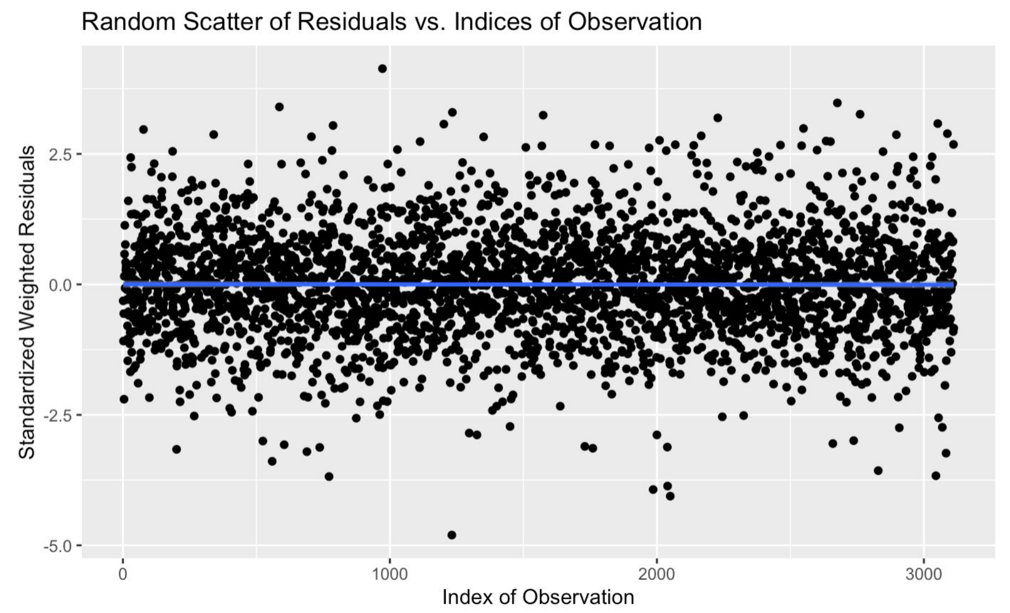
Discussion

- There is promise that these markers could be more efficient and effective tools in politics
- Strengths:
  - Big advancement in under-studied area of research
  - Accessible to all and does not rely on paid data
- Weaknesses:
  - Assumes 2 party system
  - Data collection process is cumbersome for adding stores

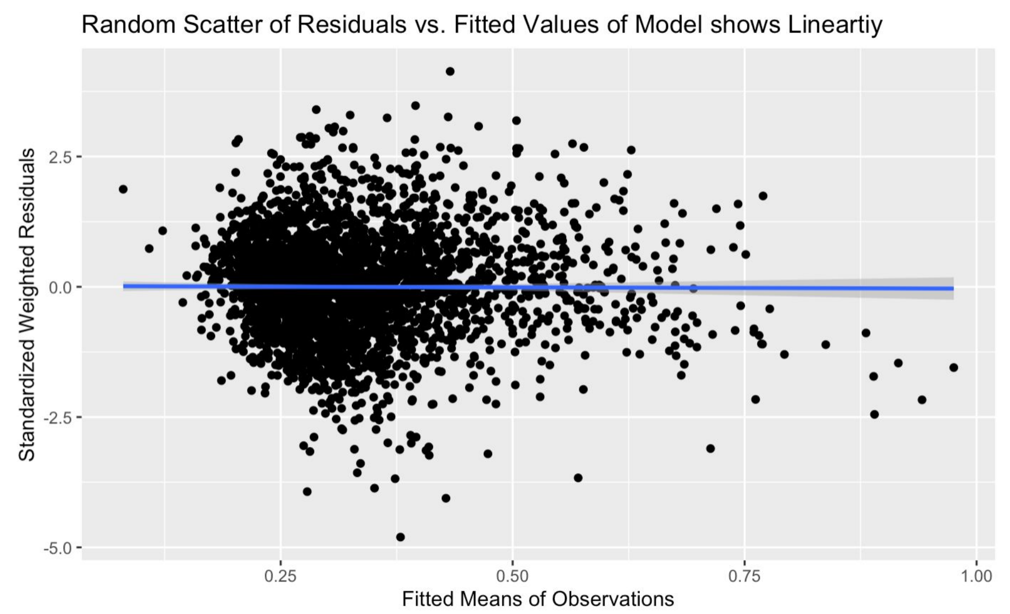
Sources

- http://clipart-library.com/clipart/gTeoreELc.htm
- https://pngimg.com/uploads/donald_trump/donald_trump_PNG54.png
- https://www.politico.com/interactives/uploads/image-service/2021/2/2/a3feafb2e8.png

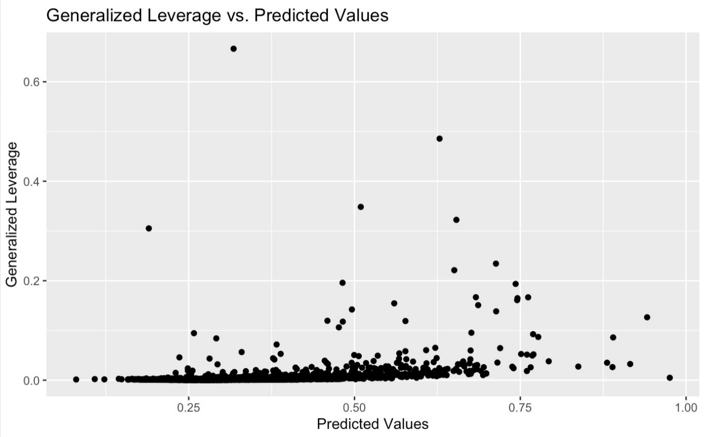
Assumption Checks

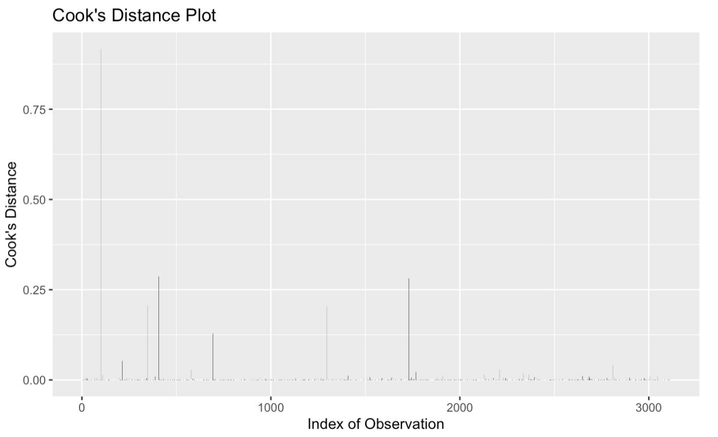


Assumption Checks



Influential Points/Outlier Checks





Model with Non-Constant Phi

```
Phi coefficients (precision model with log link):
                      Estimate Std. Error z value Pr(>|z|)    
(Intercept)           5.740051   0.775532   7.401 1.35e-13 ***
trader_joes_100k      0.018062   0.671902   0.027 0.978554    
cabelas_100k         -0.111478   0.171890  -0.649 0.516632    
cinemarks_100k        0.567020   0.219957   2.578 0.009941 ** 
nordstroms_100k       0.572770   1.030838   0.556 0.578460    
perc_white            1.322797   0.389809   3.393 0.000690 ***
perc_black            1.644291   0.386354   4.256 2.08e-05 ***
perc_hispanic        -0.748993   0.412262  -1.817 0.069250 .  
perc_asian            2.804502   1.170219   2.397 0.016550 *  
perc_male            -2.692661   1.180939  -2.280 0.022602 *  
percent_educated     -0.047305   0.003206 -14.754  < 2e-16 ***
percent_poverty      -0.020413   0.007100  -2.875 0.004040 ** 
Unemployment_rate_2019  0.074787   0.021477   3.482 0.000497 ***
Metro_2013            0.398169   0.059363   6.707 1.98e-11 ***
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
```

Model with Only Store Counties

Coefficients (mean model with logit link):

	Estimate	Std. Error	z value	Pr(>|z|)	
(Intercept)	-1.496100	1.546428	-0.967	0.333316	
trader_joes_100k	0.829847	0.237901	3.488	0.000486	***
cabelas_100k	-0.164059	0.068719	-2.387	0.016968	*
cinemarks_100k	-0.504093	0.112123	-4.496	6.93e-06	***
nordstroms_100k	0.538056	0.318971	1.687	0.091632	.
perc_white	1.026308	1.009459	1.017	0.309301	
perc_black	2.585381	0.993342	2.603	0.009249	**
perc_hispanic	2.120932	0.988343	2.146	0.031877	*
perc_asian	4.754683	1.254945	3.789	0.000151	***
perc_male	-4.806984	2.191560	-2.193	0.028278	*
percent_educated	0.030973	0.003682	8.412	< 2e-16	***
percent_poverty	0.022772	0.007042	3.234	0.001221	**
Unemployment_rate_2019	0.054965	0.021105	2.604	0.009204	**
Metro_2013	-0.072051	0.101359	-0.711	0.477179	