

Physical Location Factors in Voter Participation and Behavior: Undervoting in the 2018 Board of Education Contests in Wake County, North Carolina

Johnny Crawford

May 18, 2020

1. Introduction

Background

Voting is one of the most common collective human activities performed across the world. In one form or another, with varying levels of accuracy and oversight, it is performed in most politically, economically, and culturally stable countries (and many unstable countries) on a routine basis. Many people, especially in the United States, lament the electoral process, whether it be for the less than stellar voter turnout rate often experienced, the ever-increasing hyper-partisan nature of elections, or the disbelief that voting will in the end make a difference to the way in which government runs at the federal, state, or local levels. The ancient Greek statesman, orator, and general Pericles is often cited as having stated, “Just because you do not take an interest in politics doesn't mean politics won't take an interest in you.” (1) It is true that level of indifference one may have toward politics makes absolutely no difference in whether one is affected by the results of politics. Elections are held, officials take office, and the business of government runs, for better or worse. That being stated, the political and electoral process is important to everyone.

There has been a great deal of research on voting, sentiment, political ideal formation, and other concepts associated with the electoral process. One area that has been researched, although possibly not to the extent that it could be researched, is that of physical location factors in determining voting various types of voting behaviors. (2) For example, it was found that shortening the distance of polling locations increased voter turnout, regardless of the level of confusion that change may have caused. (3) Others debate whether the choice of venue, whether it be a private religious institution (such as a church) or a public secular institution (such as a public school) is enough to subtly sway the choices of those casting a ballot. (4) Understanding these concepts are considerations when choosing the placement of a polling place for a particular precinct. This also is a consideration for those who advocate for greater degrees of emphasis being placed on other methods of voting, such as voting by mail or even Internet-based voting systems.

Research Problem

There are many measurable voting behaviors that could be tied to real-world location components. Some of the more unique situations include such things as turnout, under-voting (a ballot where the choices completed are less than the number allowed, whether it be

for top-of-ticket races, lower ballot races, non-partisan positions, specific races), or split-ticket voting (where two or more top-level races are decided for candidates of a different party). Similarly, real-world situations that are measurable and may be associated with particular voting behaviors include the type of venue serving as a particular polling place for a precinct, the prevalence of certain types of venues of a similar nature to the precinct polling place, or the presence of particular types of commercial venues. This project seeks to examine these geo-location factors that may be related to, or exhibited in conjunction with, certain types of voting results or behaviors. This project will utilize real-world election results from the 2018 midterm election for Wake County Board of Education in Wake County, North Carolina – home of Raleigh, the capital of North Carolina. It will also utilize physical location data including precinct polling locations obtained from the Wake County Board of Elections and venue data of locations in the vicinity of the precinct polling locations obtained via the Foursquare API.

Intended Audience

As mentioned previously, everyone has a stake in the performance of government, whether or not there is an explicit desire to be actively participate or participate at a higher level in the process. While there is general applicability of this project, some groups may be more interested in the insights gained from this type of research than others. For example, activists desiring a higher level of engagement among communities they serve or represent may want to utilize the community building and neighborhood commercial aspects of this understanding to channel funding and resources back to the community to create a community that leads to a greater degree of success among its members. Those of charge of elections may wish to increase the level of turnout and achieve their goals of making the best utilization of resources in terms of budgets and locations. Political candidates and campaigns may wish to utilize this knowledge to determine areas where a possible base may be located as well as to better be able to reach those potential voters with whom they are not yet in contact. And finally, the average citizen and taxpayer may realize better representation and less dissatisfaction with elected officials by becoming more knowledge about candidate choices and the conscious or unconscious process behind those decisions.

2. Data

Data sources

Data to be used in this research will include location data and election results data. Location data was obtained through the Wake County website and from Foursquare via the Foursquare API. Location data for precinct polling locations was obtained from the Wake County website. Data includes names, addresses, and latitudes/longitudes for all polling locations. Additional information was obtained through Foursquare via the Foursquare API that included local venues in each precinct area based on closeness to the precinct polling location's latitude and longitude coordinates. Precinct level results for the 2018 midterm Board of Education race in Wake County was obtained from the Wake County Board of Elections website. Full details on the data referenced and locations from where this data was obtained may be obtained in **Section 7. References** at the end of this document.

Data Cleaning

Location data was obtained in Excel, and KML format files and parsed in order to provide CSV files for easier analysis. Data obtained via the Foursquare API was added to dataframes that were accessed and stored in the companion Jupyter Notebook that is associated with this report.

Election results data was obtained and referenced for the 2018 Wake County Board of Education races. As this is a non-partisan race where numerous candidates are running in separate districts, but all districts had an election, vote totals from each precinct were aggregated and analyzed by precinct. Vote totals were combined with general precinct-by-precinct statistics that included precinct number, registered voters, ballots cast, and precinct turnout percentage. Undervote was calculated as one minus the number of votes cast for that race over the total number of ballots cast, which yielded a percentage of total ballots cast that represented people who voted but not in a Board of Education race for their district or precinct. Data was combined into single spreadsheets and indexed by precinct number.

Coding was performed to determine and categorize the type of physical location that was being used as a polling place for that precinct based on the name and description of the facility. Common types of facilities where voting was conducted included churches, schools, and community centers.

How data will be used to examine the problem

Statistics such as voter turnout and vote totals were provided by the state and county boards of elections. Statistics associated with under-voting were calculated based on ballots cast and votes cast for a Board of Education contest.

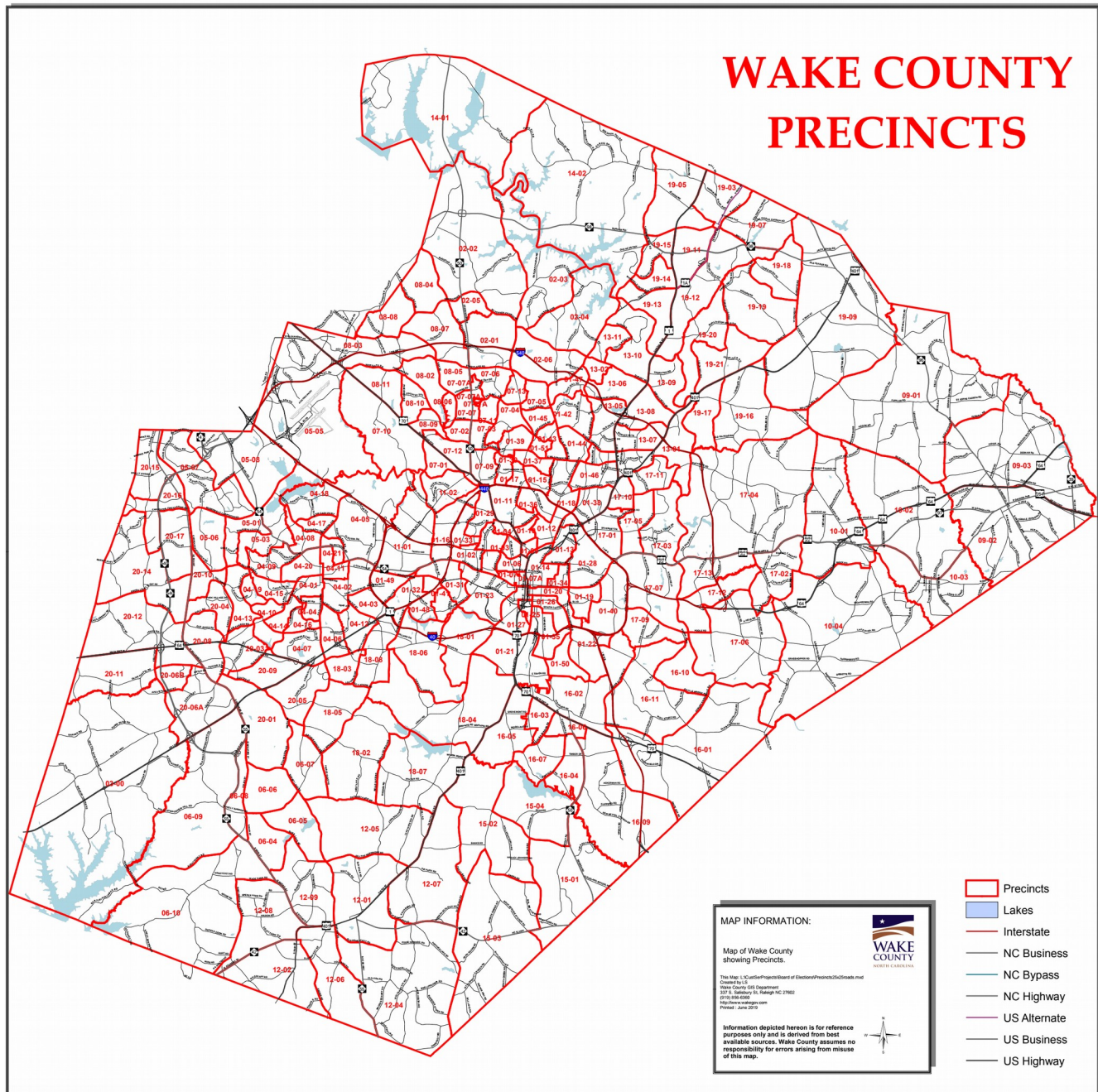
Location data obtained from boards of election and the Foursquare API included type of venue used for voting, precinct polling latitude and longitude, and venues near the precinct polling locations.

In addition to descriptive statistics, machine learning techniques will be used to provide further analysis. These will include regression and cluster analysis of vote results as well as venue information obtained from the Foursquare API. Further information on the techniques used have been included in the **Section 3. Methodology**.

3. Methodology

Initial Review

First, I reviewed the precinct map of Wake County precinct locations and boundaries as provided by the Wake County Board of Elections. This map encompassed the total number of precincts spread out over the entire county and includes precincts in cities and towns such as Apex, Clayton, Fuquay-Varina, Garner, Holly Springs, Knightdale, Morrisville, Rolesville, Wendell, Zebulon, Cary, Wake Forest, and Raleigh – the capital of North Carolina.



I then parsed the KML format file obtained from the Wake County website that provided site information for each of Wake County's polling places, one per precinct, to obtain latitude and longitude coordinates. Then, I merged this with the site precinct information file and coded type of building/facility used for each precincts' polling venue. From this information, I obtained a breakdown of where precincts are located by city, town, and municipality as well as an overview of how many types of each type of venues is used across the county as polling places.

city			
APEX	15		
CARY	28		
CREEDMOOR	1		
FUQUAY-VARINA	6		
GARNER	9		
HOLLY SPRINGS	5		
KNIGHTDALE	5		
MORRISVILLE	3		
RALEIGH	114		
ROLESVILLE	3		
WAKE FOREST	9		
WENDELL	4		
WILLOW SPRINGS	1		
ZEBULON	3		

poll_pl_coded			
clubs culture libraries	7		
community centers	24		
government	4		
religious	76		
residential	7		
safety	10		
school	78		

As the above tables show, Raleigh represents most of the precincts in Wake County as the most populous city in the county. Also, the vast majorities of precinct polling locations are either religious buildings, particularly Christian churches of various denomination and belief systems or some type of school. Community centers (which for purposes of this study also includes some parks) are also common polling locations.

General Descriptive Statistics for Ballots Cast

I then reviewed a number of basic voting statistics provided by the Wake County Board of Elections for each election period. The period this study focuses on is the 2018 midterm election. At that time there were 735,952 registered voters in Wake County. 439,153 ballots were cast during this election for a turnout rate of 59.67%. This is opposed to the 2016 general election where there were 705,228 registered voters (with 531,236 ballots cast for a 75.33% turnout rate) and the more comparable 2014 midterm election where there were 670,801 registered voters (with 327,675 ballots casts and a turnout rate of 48.85%). Midterm elections are elections that typically involve races that are elected every two years and do not

typically involve as many higher visibility Federal and State-level elections that drive turnout. However, it was quite surprising to learn of Wake County’s nearly 60% turnout during a non-Presidential, non-gubernatorial election.

General Descriptive Statistics for Board of Elections Race Undervote

As part of an initial overview of the data, I decided to look at correlation of a possible factor in driving undervote, that of turnout percentage. This data is reported on a precinct-by-precinct basis, so we can easily review this. Sample data from the “undervote” data set looks like this.

	Precinct	RegVot	Cast	Cast-Blank	Turnout	BoE-Cast	Undervote
0	01-01	2331	1332	0	0.5714	990	0.256757
1	01-02	2299	1356	0	0.5898	928	0.315634
2	01-03	2087	1547	0	0.7413	977	0.368455
3	01-04	1297	952	0	0.7340	622	0.346639
4	01-05	1125	861	0	0.7653	551	0.360046
...
199	20-12	3513	2424	0	0.6900	2035	0.160479
200	20-14	6463	4028	0	0.6232	3545	0.119911
201	20-15	4092	2294	0	0.5606	1862	0.188317
202	20-16	3711	2014	0	0.5427	1557	0.226912
203	20-17	3963	2506	0	0.6323	2076	0.171588

A correlation matrix was generated using Pearson correlation with careful attention on the strength of the correlations with the “Undervote” column. Most of the relationships are weakly correlated. This includes turnout ($r = 0.256715$), indicating that a higher turnout might have a slight effect on undervote in Board of Education races. However, turnout is out-of-scope for this research examination.

	RegVot	Cast	Cast-Blank	Turnout	BoE-Cast	Undervote
RegVot	1.000000	0.941927	0.244122	-0.261962	0.899997	-0.243409
Cast	0.941927	1.000000	0.232104	0.036314	0.941248	-0.185581
Cast-Blank	0.244122	0.232104	1.000000	-0.041675	0.226829	-0.070433
Turnout	-0.261962	0.036314	-0.041675	1.000000	-0.019946	0.256715
BoE-Cast	0.899997	0.941248	0.226829	-0.019946	1.000000	-0.481578
Undervote	-0.243409	-0.185581	-0.070433	0.256715	-0.481578	1.000000

Finally, descriptive statistics of undervote, and associated statistics, were examined. Undervote ranged from 11.8% to 48.35% for the Board of Education race in 2018, with an average of 28.0% per precinct.

	RegVot	Cast	Cast-Blank	Turnout	BoE-Cast	Undervote
count	204.000000	204.000000	204.000000	204.000000	204.000000	204.000000
mean	3607.607843	2152.710784	0.303922	0.607501	1570.411765	0.280034
std	1642.337936	951.778729	0.632288	0.090905	797.315611	0.116823
min	336.000000	173.000000	0.000000	0.266200	111.000000	0.118347
25%	2473.500000	1459.000000	0.000000	0.556600	987.750000	0.170929
50%	3263.000000	1999.500000	0.000000	0.616000	1440.500000	0.265193
75%	4589.000000	2694.250000	0.000000	0.671525	1955.000000	0.385662
max	9267.000000	5336.000000	3.000000	0.811800	4462.000000	0.483589

Encoding, Correlation, and Regression of Polling Place Types

I merged polling place types with undervote percentages to perform a Pearson correlation, but none of the polling place locations came back as a strong correlation. While not totally surprising, community centers had the strongest correlation (one of an inverse nature), albeit a weak one.

	Undervote	clubs culture libraries	community centers	government	religious	residential	safety	school
Undervote	1.000000	0.008912	-0.129821	-0.060250	0.091947	0.068271	0.044864	-0.038482

Foursquare API Venue Data

While the polling place location itself does not appear related to undervoting for this particular race, it is possible that venues in the area of the polling place have some influence on undervoting and other electoral behavior in particular races. With that, venue information was obtained using the Foursquare API to determine the most frequent venues near to the polling location. As there are a number of potential variables of interest I used correlation and multiple linear regressions to evaluate the large number of factors and potential venues. As top venues are being categorized for the purpose of regression analysis, the k-means cluster analysis was also performed to see if there are any additional insights that might be provided among the 245 unique venue categories that were determined to exist in Wake County, North Carolina using the Foursquare API.

4. Results

As mentioned in **Section 3. Methodology**, there was no indication that the polling location for a precinct changed the likelihood or magnitude of undervoting for the Board of Education races being examined. Thus, I turned my attention to location data obtained via the Foursquare API in order to determine if there was any influence placed on undervoting behavior based on venues that were in close proximity to the polling location.

Correlation and linear regression models did not yield any strong correlations, either in a positive or negative relationship with the variable of undervoting. There were a number of weak correlations that were found in the process however that might be a jumping point for further research in this area. For example, the table below demonstrates those weak correlations that indicate less undervoting, less of a magnitude for undervoting, and a possible measure of greater voter engagement and participation.

Undervote

Bookstore	-0.124002
Candy Store	-0.101513
Dog Run	-0.109004
Frozen Yogurt Shop	-0.108342
Juice Bar	-0.120230
Mediterranean Restaurant	-0.103864
Music Store	-0.121777
Park	-0.146749
Shopping Mall	-0.141959
Sporting Goods Shop	-0.121676
Thai Restaurant	-0.117153
Theater	-0.106923

It is unclear what part shopping malls ($r = -0.141959$) or parks ($r = -0.141959$) might play in influencing the decision to not participate in undervoting behavior. All-in-all, those factors that are indicated in this list seem to indicate characteristics that may be thought of as more high-brow, sophisticated, or associated with an affluent society. These include theater, diverse dining options, and malls. Many of these venues also involve experiencing the outdoors (such as dog runs and sporting good shops) or (possibly minus the candy store) healthier personal pursuits.

In a similar fashion, a larger number of venues in close proximity correlated positively with undervoting, again in a mostly weak fashion. Most notably, the strongest positive correlation was with toy/game stores ($r = 0.210559$). Other entertainment pursuits seemed weakly correlated, such as video game stores, arcades, comic shops, and circuses. However a number of outdoor pursuits, such as garden centers ($r = 0.177410$) and farms ($r = 0.162914$) made an appearance on this list.

Undervote			
Undervote	1.000000	Lingerie Store	0.111608
Arcade	0.123019	Liquor Store	0.133831
Breakfast Spot	0.116209	Martial Arts Dojo	0.105704
Burger Joint	0.128821	Medical Supply Store	0.103022
Chinese Restaurant	0.104399	Mobile Phone Shop	0.137258
Circus	0.100004	Mongolian Restaurant	0.109866
Comic Shop	0.122370	Outlet Store	0.120649
Doctor's Office	0.114715	Pet Store	0.144493
Electronics Store	0.106840	Pool Hall	0.109866
Farm	0.162914	Pub	0.134448
Fish Market	0.115508	Restaurant	0.126215
Flower Shop	0.131104	Sandwich Place	0.118882
Food	0.120003	Skating Rink	0.109866
Garden Center	0.177410	South Indian Restaurant	0.115508
Himalayan Restaurant	0.115508	Sports Bar	0.118663
Hotel	0.107470	Supermarket	0.149811
Indian Restaurant	0.137885	Toy / Game Store	0.210559
Irish Pub	0.128860	Trail	0.104139
Lawyer	0.136243	Video Game Store	0.148693

Luckily, as encoding of the data was done in order to perform this part of the analysis, cluster analysis was also conducted by way of k-means cluster analysis that yielded additional insights into these groups.

Based on multiple iterations of k-means cluster analysis, a decision was made to evaluate six clusters based on venues nearest each precinct. Sample data are found below, and additional data is available in the companion Jupyter Notebook for this project.

Cluster 0:

	precinct	poll_pl_coded	Cluster Labels	1st Most Common Venue	2nd Most Common Venue	3rd Most Common Venue
0	01-01	religious	0.0	Ski Area	Garden	Playground
2	01-03	religious	0.0	Fast Food Restaurant	Home Service	Fish Market
16	01-17	religious	0.0	Bank	Basketball Court	Fast Food Restaurant
20	01-21	community centers	0.0	Dog Run	Playground	Park
32	01-34	community centers	0.0	Fast Food Restaurant	Hardware Store	Mattress Store
43	01-45	religious	0.0	Italian Restaurant	Fast Food Restaurant	Harbor / Marina
44	01-46	community centers	0.0	Indie Movie Theater	Playground	Park
49	01-51	residential	0.0	Massage Studio	Disc Golf	Playground
75	04-19	religious	0.0	Intersection	Arcade	Flower Shop
80	05-05	community centers	0.0	Gas Station	Pool	Gym

Cluster 1:

	precinct	poll_pl_coded	Cluster Labels	1st Most Common Venue	2nd Most Common Venue	3rd Most Common Venue
27	01-29	school	1.0	Hunting Supply	Pool	Yoga Studio
31	01-33	school	1.0	Pool	Campground	Trail
40	01-42	community centers	1.0	Pool	Park	Tennis Court
85	06-05	school	1.0	Pool	Athletics & Sports	Soccer Field
109	08-07	clubs culture libraries	1.0	Pool	Yoga Studio	Flower Shop
137	13-09	school	1.0	Pool	Playground	Yoga Studio
145	15-04	school	1.0	ATM	Pool	Farm
187	19-18	school	1.0	Pool	Yoga Studio	Flower Shop
205	20-17	school	1.0	Pool	Park	Yoga Studio

Cluster 2:

	precinct	poll_pl_coded	Cluster Labels	1st Most Common Venue	2nd Most Common Venue	3rd Most Common Venue	
	1	01-02	religious	2.0	Dry Cleaner	Pizza Place	Bagel Shop
	3	01-04	school	2.0	Dessert Shop	Pizza Place	Gym
	4	01-05	religious	2.0	Southern / Soul Food Restaurant	Pizza Place	Salon / Barbershop
	5	01-06	clubs culture libraries	2.0	Boutique	Furniture / Home Store	Jewelry Store
	6	01-07	school	2.0	Sushi Restaurant	Bar	Arts & Crafts Store

	197	20-08	religious	2.0	Music Store	Food Truck	Coffee Shop
	198	20-09	religious	2.0	Mexican Restaurant	Sandwich Place	Supermarket
	199	20-10	safety	2.0	Pizza Place	Ice Cream Shop	Coffee Shop
	200	20-11	religious	2.0	Gift Shop	Sporting Goods Shop	Yoga Studio
	203	20-15	school	2.0	Gym	Convenience Store	IT Services

Cluster 3:

	precinct	poll_pl_coded	Cluster Labels	1st Most Common Venue	2nd Most Common Venue	3rd Most Common Venue	
	61	04-05	school	3.0	Garden Center	Athletics & Sports	Yoga Studio
	188	19-19	school	3.0	Gym Pool	Athletics & Sports	Yoga Studio
	189	19-20	community centers	3.0	Athletics & Sports	Yoga Studio	Flower Shop

Cluster 4:

	precinct	poll_pl_coded	Cluster Labels	1st Most Common Venue	2nd Most Common Venue	3rd Most Common Venue
12	01-13	religious	4.0	Baseball Field	Park	Bus Stop
30	01-32	community centers	4.0	Baseball Field	Yoga Studio	Food
119	10-03	school	4.0	Park	American Restaurant	Baseball Field
123	12-01	school	4.0	Pool	Farm	Baseball Field
124	12-02	community centers	4.0	Baseball Field	Playground	Discount Store
147	16-02	school	4.0	Baseball Field	Yoga Studio	Food
178	19-07	community centers	4.0	Dog Run	Baseball Field	Yoga Studio
182	19-13	school	4.0	Football Stadium	Baseball Field	Yoga Studio
193	20-04	school	4.0	Baseball Field	Martial Arts Dojo	Construction & Landscaping
204	20-16	school	4.0	Soccer Field	Baseball Field	Yoga Studio

Cluster 5:

	precinct	poll_pl_coded	Cluster Labels	1st Most Common Venue	2nd Most Common Venue	3rd Most Common Venue
14	01-15	school	5.0	Night Market	Park	Yoga Studio
15	01-16	community centers	5.0	Moving Target	Park	Yoga Studio
24	01-26	community centers	5.0	Park	Food Service	Health & Beauty Service
36	01-38	school	5.0	Park	Yoga Studio	Gastropub
38	01-40	community centers	5.0	Insurance Office	Park	Yoga Studio
50	02-01	school	5.0	Park	Yoga Studio	Gastropub
73	04-17	safety	5.0	Bus Station	Park	Yoga Studio
94	07-04	school	5.0	Miscellaneous Shop	Park	Yoga Studio
157	17-02	school	5.0	Boxing Gym	Park	Yoga Studio

5. Discussion

While a lot of analysis was conducted in order to determine if geo-location and physical factors associated with polling locations and nearby venues were conducted and yielding only mild, weak indications that this might be the case, I am positive that this research has merit and was well worth the time. Elections and voting, like many applications that data scientists investigate, yield lots of data and leads to even more questions.

One example of a question that comes to mind is whether the effect on physical voting location would be stronger for a different political race, perhaps one at a higher level of government, or a similar local position, or even one where the candidates have partisan labels and party affiliations on the ballot tied to their candidacy. Many such questions were raised through this research.

Recommendations

My most immediate recommendation would be to complete additional research into this area, which I intend to pursue at some point in the future. There is research existing in this area already, but with the importance of the political process and how it touches on everyone's lives and livelihood, more research could be needed.

While some factors were determined to be weakly related to the particular behavior I examined, that of undervoting, perhaps those places where a weak relationship was noted could be explored further, for example, for the intent of the behavior. For example, some

aspects of undervoting may be related to factors such as ballot design – the voter didn't realize that they was a back side of the ballot. Other common insights might be that those voters with venues that contributed to undervoting might be uncomfortable with the level of knowledge necessary for making those choices. Insight into this being the reason and not that say, for instance, people just didn't like the candidates on the ballot for that given position and could use different choices for representation. It is unclear of that level of intent, but now that we have more insight that people frequenting video game stores, farms, or garden shops may fall into one of these mentioned categories, it is well worth it to go there to seek insight. This is true for candidates who might be truly concerned with what the gamer, the farmer, or the gardener truly wants them to do while in office or the advocate for those in the community who believe that not only should be people be involved but to also be more informed in the choices that they are making.

6. Conclusion

While this idea seemed a bit far-fetched when I proposed it to myself, I was pleasantly surprised to see at least some weak results into a phenomenon that most people are unaware of that has real-world implications for many people. Now that you are aware of how voting may be influenced by your surroundings, what other aspects of your life may also be influenced by such things? It is an important area for additional research, whether in the realm of politics and voting, but in all walks of life and avenues of public discourse.

7. References

Citations

1. Pericles on politics

https://www.brainyquote.com/quotes/pericles_387116

2. Wikipedia – Polling Place

https://en.wikipedia.org/wiki/Polling_place

3. How Polling Places Can Affect Your Vote

<https://psmag.com/news/how-polling-places-can-affect-your-vote-20318>

4. The Polling Place Priming (PPP) Effect: Is Voting in Churches (or Anywhere Else) Unconstitutional?

<http://www.bu.edu/law/journals-archive/bulr/documents/blumenthalandturnipseed.pdf>

Data sources

Wake County Voting Results (by Precinct)

<http://www.wakegov.com/elections/data/Pages/registrationstatistics.aspx>

Wake County Precincts

<https://data.wakegov.com/datasets/precincts/data>

Polling Places

<https://data.wakegov.com/datasets/polling-places/data>