Spatial Patterns and Demographic Traits of the WORT 89.9 FM Donor Database

By Phoebe Krawczyk, Renee Rollman and Jake Schutt

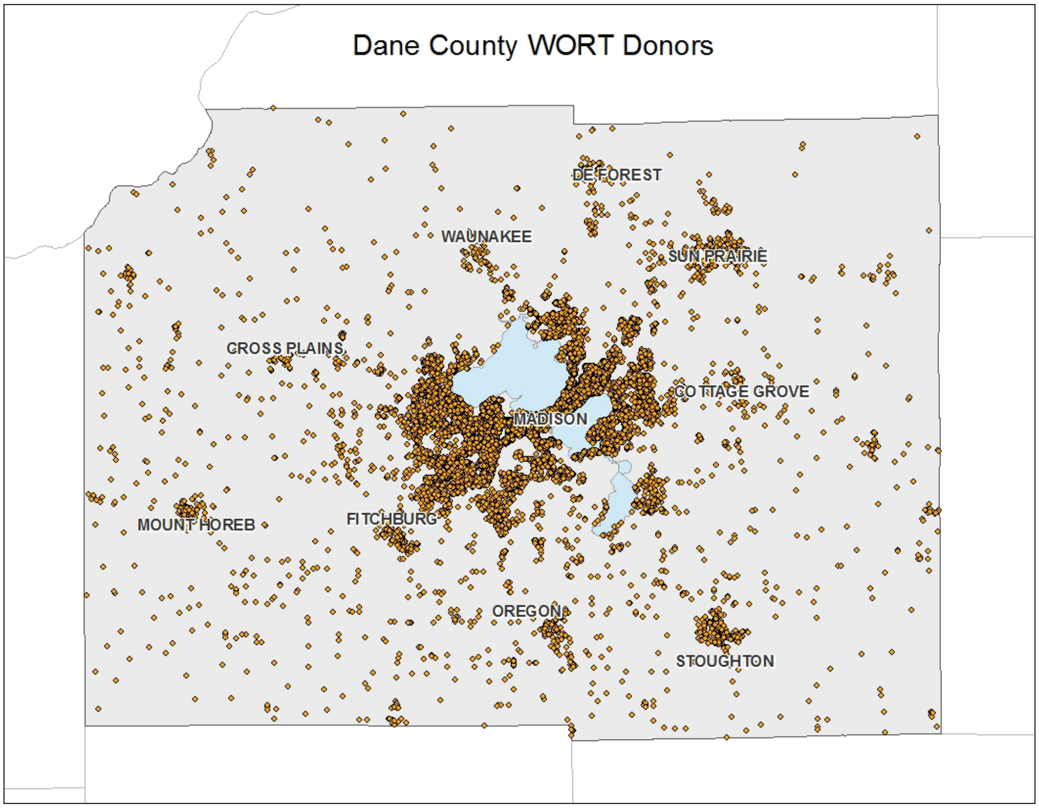
May 17, 2014

WORT-FM is a non-commercial, listener-sponsored, member-controlled community radio station broadcasting to South-Central Wisconsin since 1975. Most of its budget is funded through individual, small, private donations. When the financial pledges are taken the address of the individual pledge is recorded in WORT's donor database. Our project focused on doing spatial and demographic analysis to determine noticeable trends for past and future predictions on the donors in the area of highest concentration, Dane County, which constituted the study area.

Through our research, we aim to answer several questions relating to WORT's donor support. To begin, do any census demographics have a strong correlation with the tracts that are high or low in number of donors. Similarly, does any correlation exist between donor support and political party affiliation. And lastly, which segments of Dane County have the lowest amount of donor support.

This last question could help pinpoint potential areas for target-marketing in hopes of increasing future financial support. In an increasingly complex media landscape, radio as a whole has been steadily losing market share to a vast array of new competing audio streaming choices and podcasts with the advent and adaptation of wireless technology (Mitchell, Rosenstiel & Santhanam, 2012). At the same time there are documented rising tendencies of Americans to spatially sort themselves by similar viewpoints and cultural interests expressed in residential settlement patterns (Bishop, 2009) which adds validity to geographically based niche-marketing strategies.

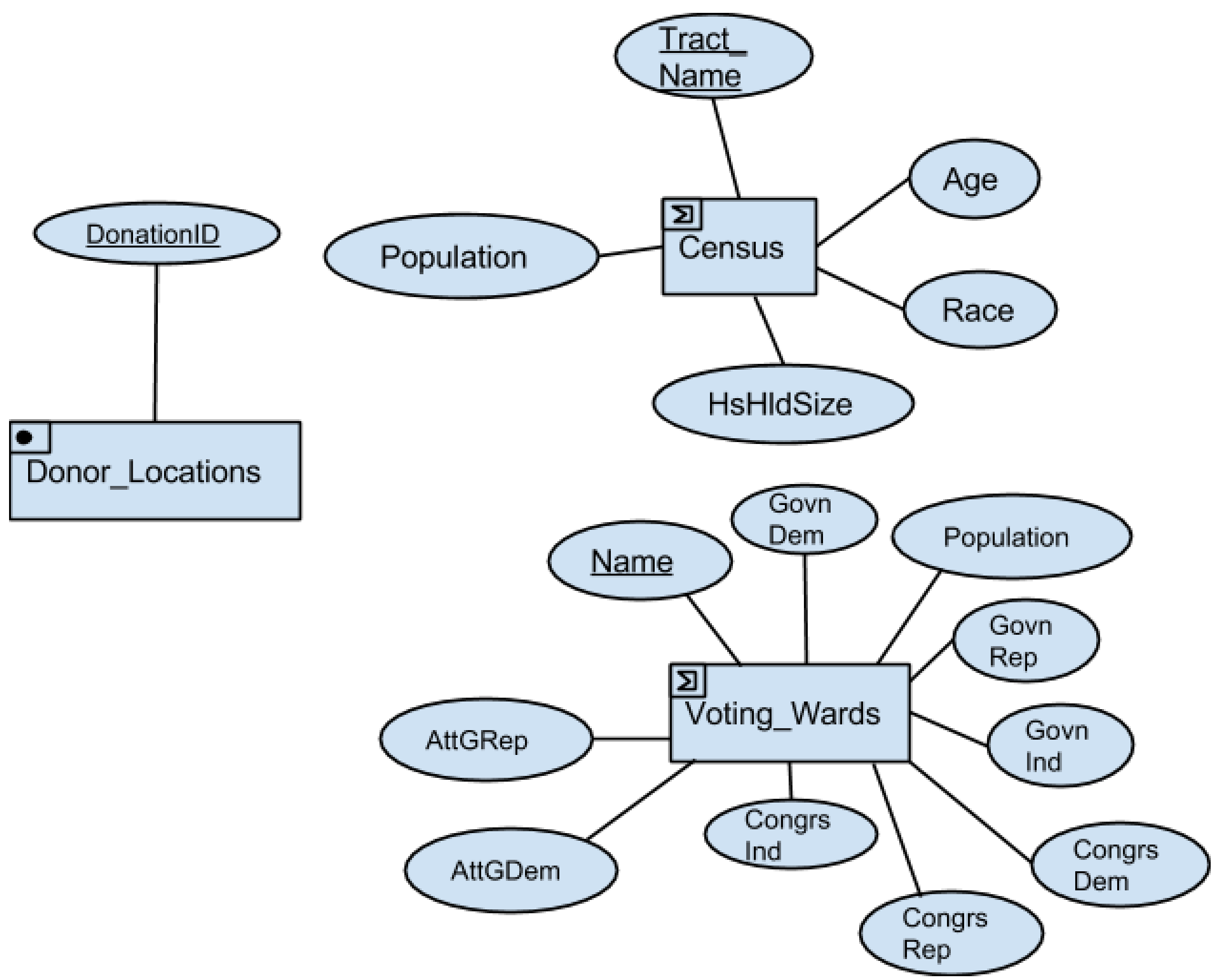
Our beginning base data was a csv document provided by the IT Coordinator at WORT. Care was taken to maintain the databases privacy by not exporting donor names on the list. Instead, each record has a unique donation ID that can be joined back to the original database should the need ever arise. That table was brought into Arc to perform the geocoding algorithm. It managed to match over 85% of the address. For the study we are assuming the records that could not be geocoded were randomly distributed across our study area and our remaining populations were large enough to justify robust statistical analysis.



Once the points were geocoded and could be mapped in space some basic spatial tendencies of variance could be observed. We aimed to take our analysis well beyond that point, however, by comparing the variance of pledge support with a host of demographic population traits. The basic idea was to obtain some useful demographic data that varies across our study area and within each enumeration unit of demographic data. We would then calculate a standardized score of pledge support to use as the independent variable for statistical comparison against the demographics to see if any significant statistical relationships were present.

Database Design

For this database we had three entities: a donor locations entity, a census entity, and a voting ward entity. These entities served as our three relation tables and each had several attributes. The donor locations entity had a donation ID that served as its unique identifier and primary key. The census entity had the following attributes: population, tract name, age, race, and household size with the tract name serving as the unique identifier and primary key. The voting ward entity also had several attributes based off of 2014 election data including: the name of the census ward, the population, and several attributes that included the information on which party was voted for in a certain election. These election attributes were denoted by Govn Dem, Govn Rep, GovnInd, AttGRep, AttGDem, Congrs Inp, Congrs Rep, and Congrs Dem. Such attributes gave us information on how many people voted for democrat in the congress elections, who voted republican in the congress elections, and so on and so forth.



From our ER diagram we were able to construct our logical diagram, displaying the attributes in a table format.

Donor_Locations

DonationID

Census

Tract_Name	Age	Race	HsHldSz	Population

Voting_Ward

Name	GovDem	GovRep	GovInd	ConDem	ConRep	ConInd	AttGDem	AttGRep	Population

We were able to perform queries across the three entities by use of creating new columns through sql queries. The first query written created a new column that read through the donor location points and assigned them to their respective census tract. An equivalent query was performed to read through the donor location data and designate the points to their respective voting ward.

```
ALTER TABLE danedonationloc
ADD census_id VARCHAR(50);
UPDATE danedonationloc
SET census_id=tractname FROM danecensus
WHERE ST_Within (danedonationloc.geom,danecensus.geom)
```

```
ALTER TABLE danedonationloc
ADD voting_id VARCHAR(45);
UPDATE danedonationloc
SET voting_id=name FROM danevotingward
WHERE ST_Within (danedonationloc.geom,danevotingward.geom)
```

Two additional queries were written in order to count the number of donor points in each tract and ward. The first query below created a new column and populated it with the number of donation points per census tract. The second query performs the same function, only in relation to the voting wards.

```
ALTER TABLE danedonationloc
ADD donorsum VARCHAR(50);
UPDATE danedonationloc
SET donorsum =
(select count(*)
FROM danedonationloc as d, danecensus as c
WHERE ST_Within(d.geom, c.geom))
GROUP BY tractname
```

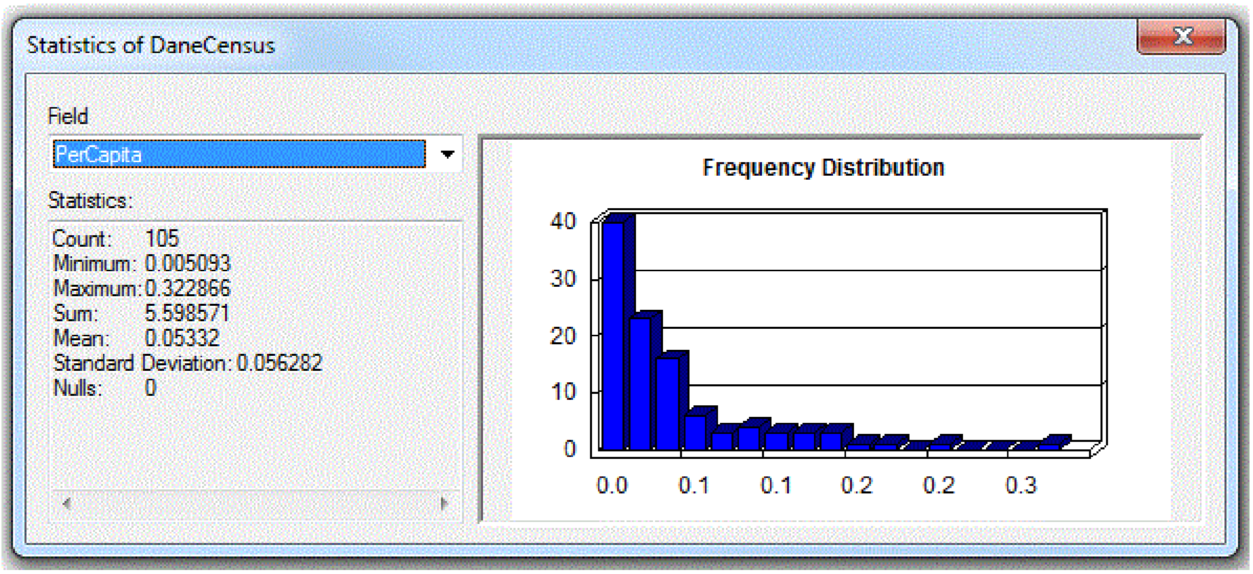
```
ALTER TABLE danedonationloc
ADD donorsum_ward VARCHAR(50);
UPDATE danedonationloc
SET donorsum_ward =
(select count(*)
FROM danedonationloc as d, danevotingward as v
WHERE ST_Within(d.geom, v.geom))
GROUP BY name
```

Results

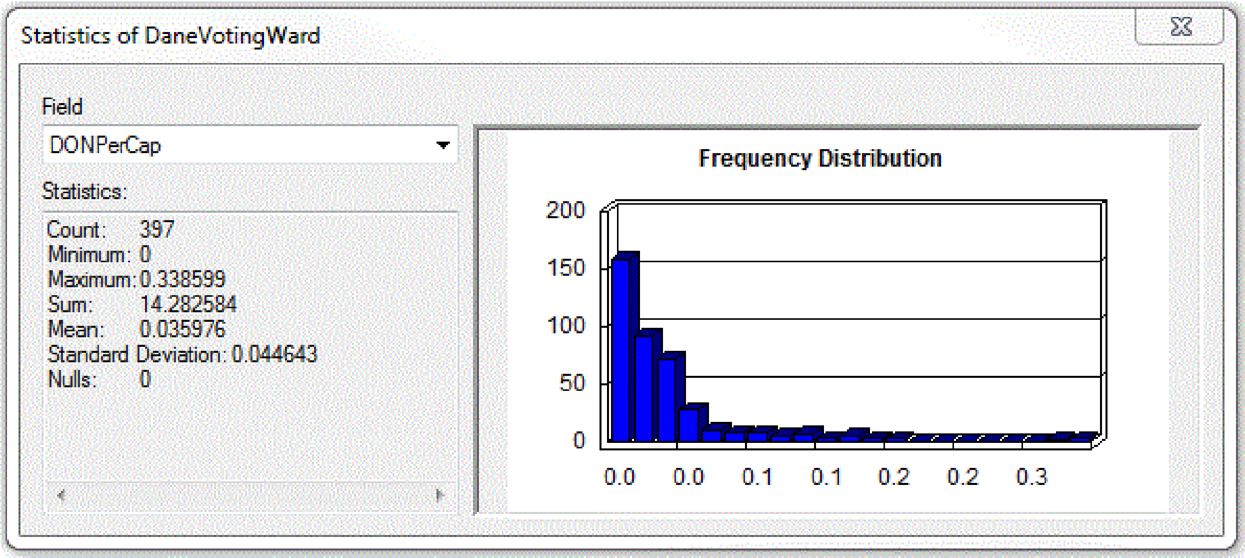
Once our queries had a populated column in our census tract and voting ward data that totaled the number within each polygonal enumeration unit, the data was normalized dividing the number of pledges by population. This per capita measure of pledge support gave an equal measure of support in areas with varying population densities. With these columns created, the voting ward and census tract tables were exported from PgAdmin into shapefiles so we could use GIS programs for spatial analysis.

The census shapefile consisted of 105 polygons covering our study area and 397 polygons that made up the voting ward shapefiles to provide somewhat finer level of analysis. The population size variance was much higher in the voting wards where population with enumeration units ran from as low as 6 to 3,864. The 105 census districts were more similar to one another in range of population per enumeration unit varying from 1,268 to 10,625.

However, despite the size of the varying enumeration units between census and voting data, some clear spatial patterns were apparent and present in both datasets when per capita pledge support was mapped. For starters in both cases the distribution was decidedly non-normal with a skew that had a few areas in our county with very strong support, but many more with a low level of support.

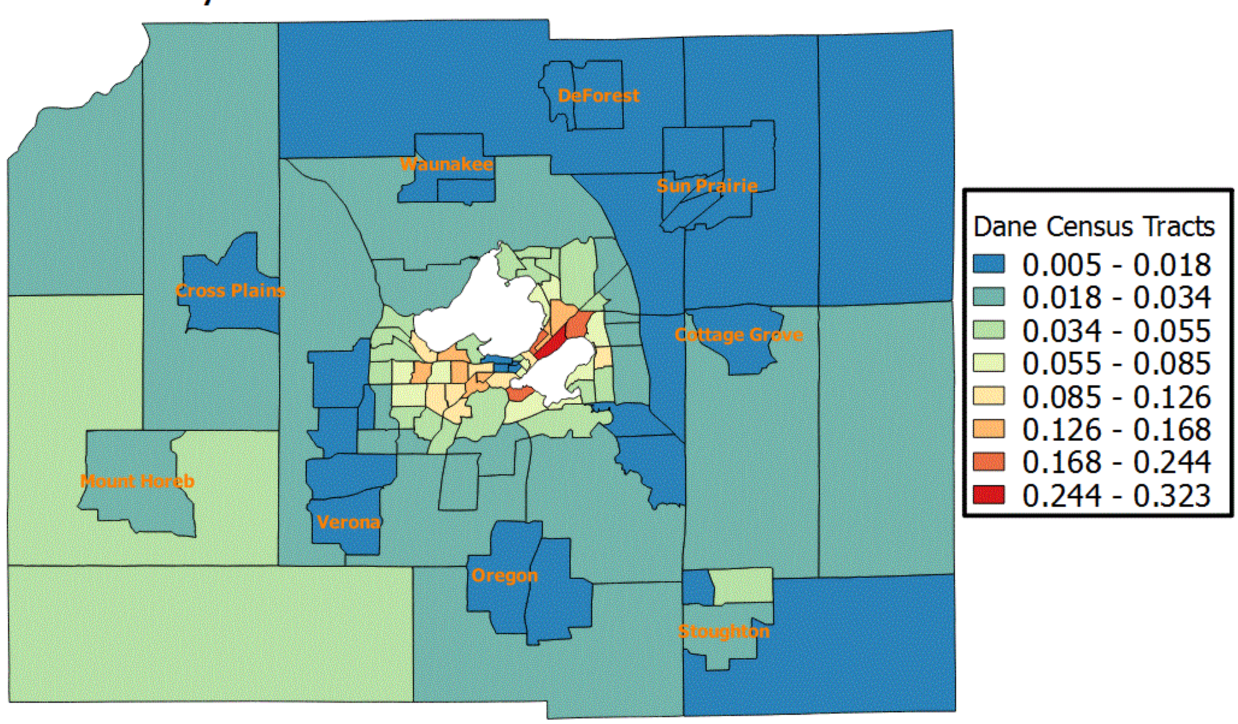


In the case of per capita support in our census data, this skew is readily apparent by the frequency distribution and how the resulting mean is much closer to the lowest per capita support values than the highest.

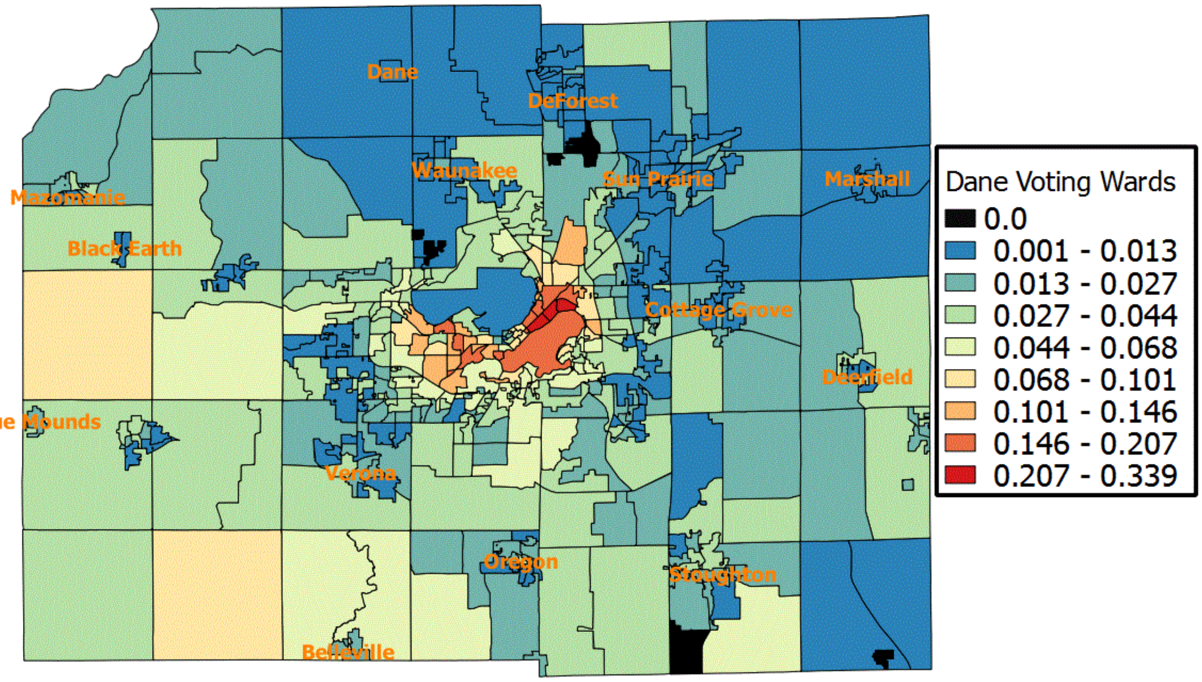


This was also very apparent in the frequency distribution and computed mean for per capita support in voting wards. With the extreme outliers in the non-normal distribution being a few remarkably high areas of per capita support.

Number of WORT Donations Per Capita Across 2010 Dane County Census Tracts-Natural Breaks Classification



Number of WORT Donatons Per Capita Across 2014 Dane County Voting Wards-Natural Breaks Classification

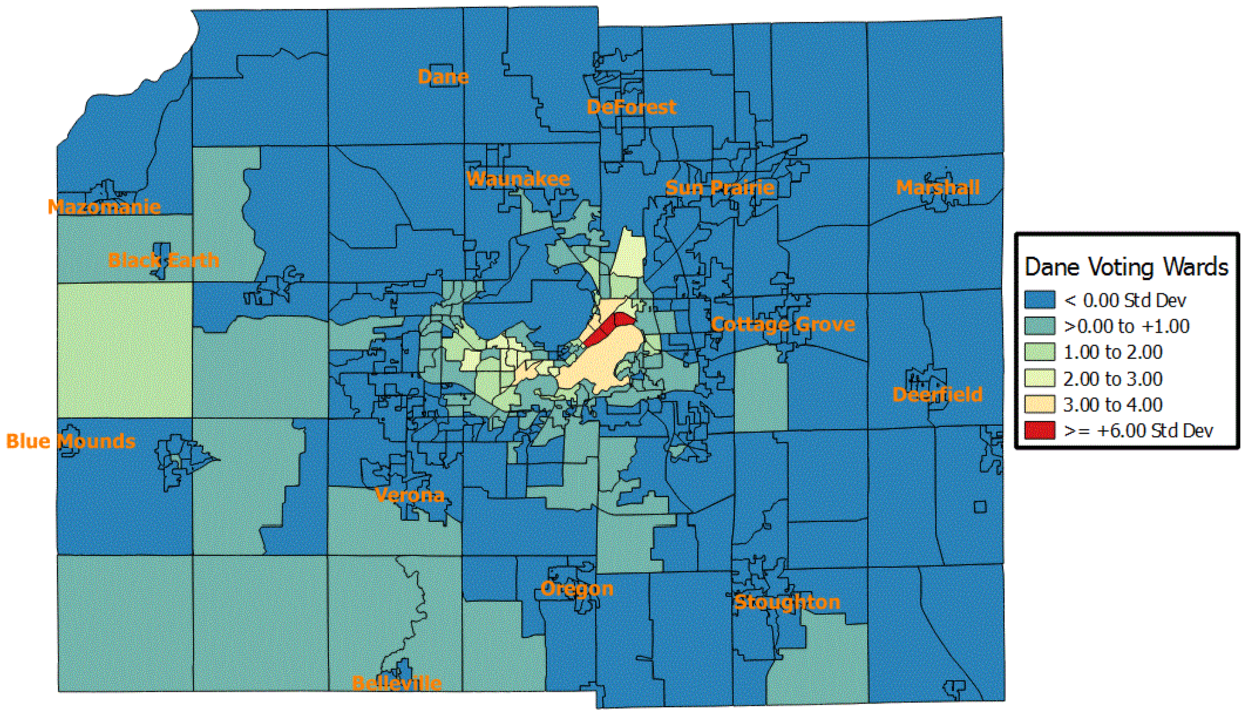


Once we mapped the per capita support across census tracts and voting ward several major spatial trends were apparent. The big outliers of exceptionally high per capita support were located in the Near East Side neighborhoods south of East Washington Avenue in the Willy/Atwood neighborhood. As a whole, per capita support was stronger inside Madison city limits, than outside. Within our city the areas of weakest support were in the University of Wisconsin campus area. Possible reasons for this could be a more transient population less rooted in our community with a low disposable income; but that is simple speculation that would require further study to actually determine.

Outside of Madison city limits, throughout Dane County, support was lowest in the northeast region of the county. The census tract with lowest per capita pledge support is in the northeast corner of Dane county where there was 1 pledge to every 195 people, whereas the Near East census tract with highest per capita support has 1 pledge for every 3 to 4 citizens.

When mapped in units of standard deviation the strength of donorship within the Madison city limits and strong positive outlier areas in the Willy/Atwood neighborhood are readily apparent.

Number of WORT Donations Per Capita Across 2014 Dane County Voting Wards-Standard Deviation Classification



With a normalized per capita level of pledge support score for all census tracts and voting wards, we could conduct statistical analysis to inquire if pledge support had any significant statistical relationships with the associated demographics contained in the census tract and voting ward tables. The first step towards this analysis was to normalize the data to account the varying population sizes within the individual enumeration districts. This was done by dividing the total number within each district that exhibit a trait by all members of a district to get a percentage of each population that exhibits a demographic we wished to explore.

A crucial point about this assumption is that we were not analyzing the demographics directly associated with the individual who made the pledge. But rather generalizing all donors within each enumeration unit as being essentially representative of the entire population in that particular area.

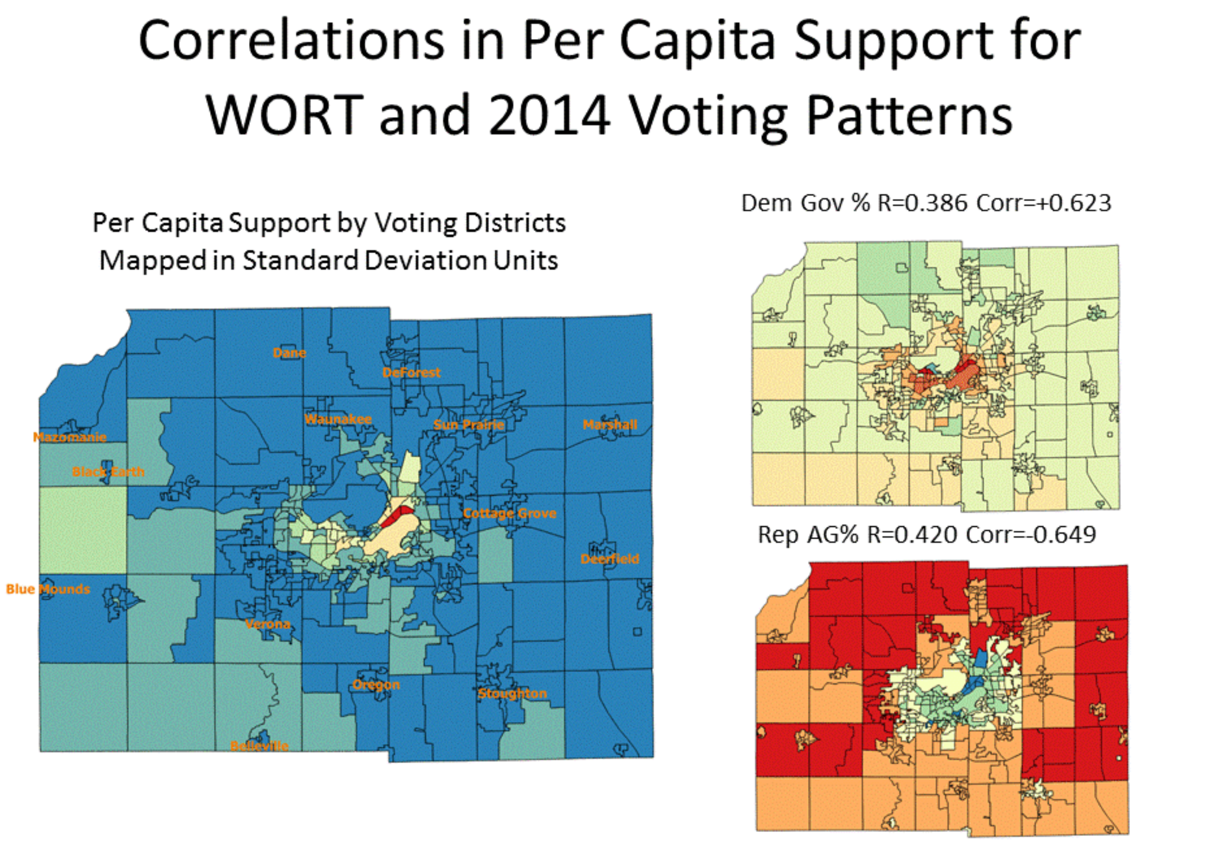
Then comparing the per capita support across all the enumeration units against our many demographic variables to see if statistical relationships could be proven to account for the variance.

In the census data, we normalized 23 demographic variables to control for population in our 105 inhabited census tracts in Dane County. We ran linear regression analysis on all 23 with per capita level of support as a dependent variable. The 23 demographic variables we investigated using 2010 census tract data explored various aspects of age, race, residential and social details within all these populations.

The results of analyzing that dataset showed weak to no statistical relationships between these census demographics and per capita pledge support coming from Dane County census tracts. The only two census variables with positive correlation strong enough to produce an adjusted squared multiple R score above 0.1 was percentage of husband and wife families and percentage of people living in owner occupied homes. The one census variable with an adjusted squared multiple R score above 0.1 and negative correlation was percentage of a population whose age is under 15. So while one must temper how much significance to attribute to weak statistical relationships, there is some indication that one could predict pledges are somewhat more likely in areas with husband and wife families are living in their own homes, but with a lower than average number of children in the house.

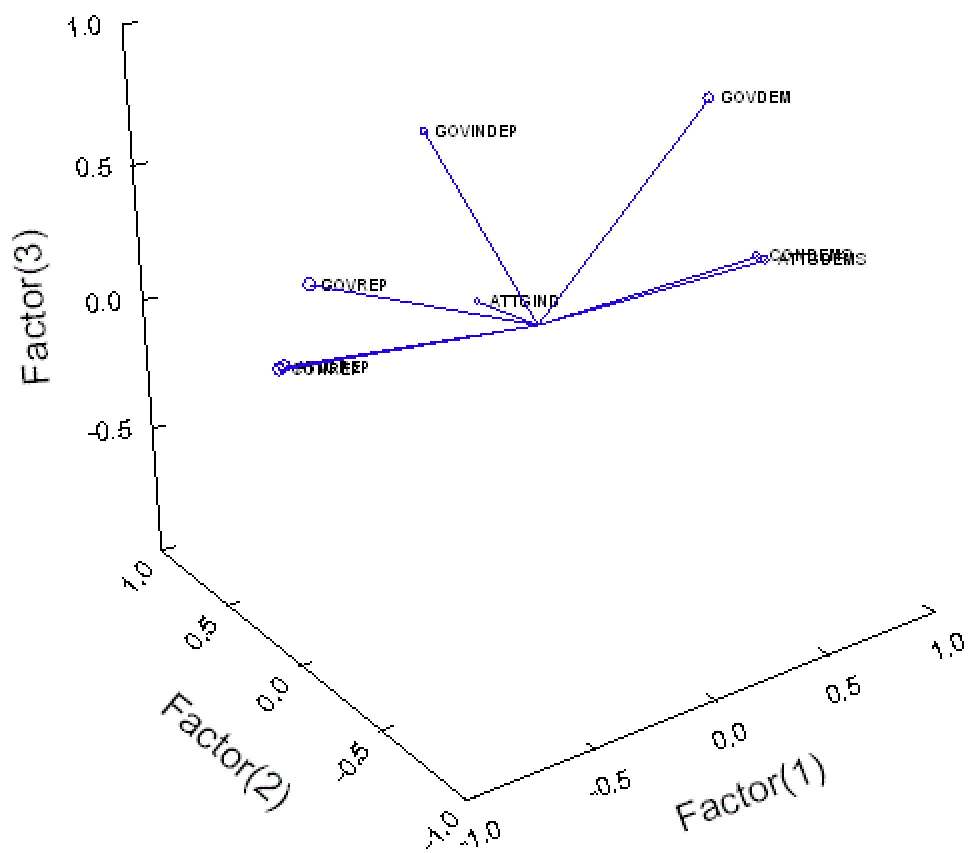
Analysis of the 2014 voting ward data set seemed to have much stronger statistical relationships with per capita level of support. For this dataset we explored eight demographic variables. All population percentages from voting wards for those voting Republican, Democrat and independent for governor and attorney general and those who voted Republican or Democrat for congress in 2014 state wide races. Four of these variables had significant adjusted squared multiple R scores, AG Dem at 0.438, AG Rep at 0.420, Gov Dem at 0.386 and Con Dem at 0.382. Three of the four, AG Dem, Gov Dem and Con Dem, had strong positive correlations and AG Rep has a clear and significant negative correlation. The other four variables we ran linear regression on, Con Rep, Gov

Rep, Gov Ind and AG Ind, had much weaker R scores showing little to no relationship when single variable linear regression was performed with per capita support as the dependent variable.



To provide some multivariate exploration, principal component analysis was run on the eight voting demographic variables to see how they interacted and to account for the demonstrated statistical variance. As the plot below demonstrates, two of the variables, Dem AG and Dem Con, were strongly positively correlated along factor 1, which accounted for over 62% of the variance. The second factor had a negative correlation and was chiefly composed of Rep AG and Rep Con.

Factor Loadings Plot



Variance Explained by Components

1	2	3
5.022	1.435	1.004

Percent of Total Variance Explained

1	2	3
62.777	17.937	12.555

Summary Conclusions

While no significant correlation appears to exist between the census demographics and the number of donors, we did find that correlations were present between political party affiliation and donor support. Donor support appeared to have a correlation with areas that voted democratic in the 2014 state-wide races. Likewise, a significant negative correlation was also evident, demonstrating the inverse relationship between the likelihood to donate to WORT and voting for Republican in the 2014 Attorney General race. Finally, our research pinpointed segments of low donor support in Dane County; namely the area surrounding the Madison campus, as well pockets of NNE Dane County, such as Waunakee.

Potential Weaknesses and Future Directions

This study and project was a useful learning exercise to the three of us and we feel our findings have some useful results to share. At the same time it is worth acknowledging potential weaknesses and suggesting future improvements to this database and similar studies. As previously mentioned, our study did not explore the demographics of the individuals within the donor database to find statistical relationships. Rather it assumed all pledges received from any enumeration district, were homogenous representatives of homogeneous areas lacking spatial variation within each district and the individuals who composed each population group within our polygons. This is a limitation of the resolution of the demographic data we had available to use and perhaps other data sources could be obtained with more targeted demographics.

In addition while the study area of Dane County is not a bad choice, we could have made different decision that could have shaped our analysis in different ways and had the potential to reveal other information. We had over 35,000 geocoded points from the entire United States, but we needed to keep close to Madison to allow for enough donation density to provide coverage throughout our study area. But potentially, we could have either enlarged the area into adjacent counties to approximate the terrestrial listening range of the station's FM signal or perhaps restricted our study area to Madison city limits for a more honest comparison.

Lastly, for the sake of simplicity, we only got a list of donations and addresses which is of limited use. In terms of dollar amounts, there is a large variance in amounts for individual financial pledges, and these factors of magnitude should be considered when looking to maximize fundraising abilities. In addition, all pledges from the database from 2001-2014 have no temporal attributes. With our dynamic and competitive local and global media landscape seeing if there are temporal changes in spatial patterns of pledging is worth examining, but we are unable to query this with our database design. This study is going to continue as part of Jake Schutt's Capstone Project this summer and more complex data will be added for further the analysis. Specifically adding the pledge amounts and pledge year, as well as considering more demographic data tables to try to find significant statistical relationships.

References

Mitchell, A, Rosenstiel, T. & Santhanam, L.H. (2012). Audio: How Far Will Digital Go?. Retrieved from http://www.stateofthemedia.org/2012/audio-how-far-will-digital-go/

Bishop, B. (2009) The Big Sort: Why the Clustering of Like-Minded America is Tearing Us Apart. Boston, MA: Mariner Books.