Creating a Profile of Individuals That Use Election Day Registration in Winona County, Minnesota USA by Using Geographic Information Systems

Marc Loubert
Department of Resource Analysis, Saint Mary's University of Minnesota, Winona, MN 55987

Keywords: GIS, Election Day Registration (EDR), Census Data, Regression, SPSS

Abstract

This study examined Election Day Registration (EDR) usage for the General Election on November 8th, 2016 in Winona County, Minnesota USA. Geographic Information Systems (GIS) was used to create a profile of who used EDR. This information coupled with statistical analysis of census block information was used to create a model of EDR. The purpose of this work is to explain what EDR is and what demographics groups used it in Winona County. The study examines how the data were collected and how the profile for EDR was created. In addition, the study examined 48 demographic variables that were considered important in creating the profile. It also examined the process by which the best prediction model was created using three variables of the 2010 federal census data.

Introduction

History of Election Day Registration

Election Day Registration (EDR) was implemented in 1974 for the State of Minnesota USA. In the beginning, only four states implemented the process. In 2004 much changed due to the increase in the number of states using EDR. As of present, there are fourteen states and the District of Columbia using Election Day Registration. Ten of these states implemented it in the last twelve years. More states are expected to have EDR available for the 2020 election. The reason it was implemented in the first place was the hope it would increase the number of people voting. Some people disliked EDR because processing requirements burden state and local governments. In addition, some argue Election Day Registration encourages ill-informed and ill-prepared voters to participate in elections (Neiheisel and Burde, 2012). In contrast others say it eliminates arbitrary deadlines that cut off registration when voters are most interested, remedies inaccurate voter rolls, and significantly reduces the need for provisional balloting (Neiheisel and Burde). Another reason for the creation of EDR was the cost. EDR is a cheaper option than having State employees working in order to have all registration paperwork filed for the election.

Election Day Registration

There are two main requirements a citizen must meet in order to participate in the democratic process. First, a citizen must register to vote and must vote either "by an absentee ballot or by filling out a ballot at a designated polling place" (Teff, 2005). The primary reasoning behind the creation of EDR was simple: to allow people to register the day of an election and eliminate some of the barriers that kept them from voting, thus increasing voter turnout (Teff, 2005). Some advocates have

---

Loubert, Marc R. 2017. Creating a Profile of Individuals That Use Election Day Registration in Winona County, Minnesota USA by Using Geographic Information Systems. Volume 20, Papers in Resource Analysis. 12 pp. Saint Mary's University of Minnesota University Central Services Press. Winona, MN. Retrieved (date) http://www.gis.smumn.edu

even gone on to argue that EDR eliminates barriers to voting that are associated with registration (Neiheisel and Burde, 2012). To use EDR, one must complete a registration application, make an oath in the form prescribed by the Secretary of State and provide proof of residence" for the last twenty days that they lived in one of the fourteen states using EDR (Teff, 2005). EDR requires only a state issued ID and a single trip for voting and registration as compared to advance registration which requires multiple trips (Brians and Grofman, 2001).

The Secretary of State's office collects data on every voter. They have maps for each county of each year for voter turnout and EDR (Teff, 2005). The data the Secretary of State's office collects is the location of the different districts for voting like Soil and Water Conservation Districts (Teff, 2005). There are twelve districts. The voting methods include: absentee ballot, voting by mail (mail ballot precinct), and voting in person. The last element is the election type. There are nine types of elections. These election types include: Municipal General Election, Municipal Primary, School District Elections, School District Primary, School District, Special Elections, School District Special Primary Special Elections, State General Elections, and State Primary Elections. Presidential voting is every four years. It is under the category of State General Election. In the 2012 general election, Winona County had 25.02 percent of voters using EDR. The state had a 17 percent for EDR.

Minnesota Election Laws

The laws for registering to vote in Minnesota state that a normal voter can register at any time except during the 20 days immediately preceding an election (Teff, 2005). When using EDR, a person has several options. First, a person may present a driver's license or Minnesota identification card. If they do not have a state issued ID, they can register using any document approved by the Secretary of State (Teff, 2005). The third method is for a person to present a current valid student identification card from a post-secondary educational institution in Minnesota" (Teff, 2005). The fourth and the final method "a voter who is registered to vote in the precinct can sign an oath in the presence of the election judge vouching that the voter personally knows that the individual is a resident of the precinct" (Office of the State of Minnesota Secretary of State 2017).

Restriction for Voting

Some limitations make people ineligible to vote such as if they have not finished a felony sentence. A person cannot vote while under guardianship if a judge specifically has revoked a person's right to vote. Another is when a court has ruled an individual legally incompetent (Office of the State of Minnesota Secretary of State, 2017).

Winona County

Winona County is in the State of Minnesota and is one of 87 counties in Minnesota. Winona County, Minnesota USA is composed of thirteen cities. Figure 1 shows Winona County in Minnesota. Cities in Winona County include: Altura, Dakota, Elba, Goodview, La Crescent, Lewiston, Minneiska, Minnesota City, Rollingstone, St. Charles, Stockton, Utica, and Winona. According to the 2010 census, Winona County had a population of 51,461 people (Data Access and Dissemination Systems (DADS), 2010).

There were 21,379 females over the age of 18 in Winona County (Data Access and Dissemination Systems (DADS), 2010). The numbers of males over the age of 18 was 20,113. Cities with the largest populations were Winona, Goodview, St. Charles, and Lewiston in that order. While La Crescent had a higher population than Goodview, there is only a small portion of the city within Winona County with the remainder in Houston County.

Winona was the most populous city in Winona County in 2010 with about 27,592 people. The median age in the city was 26.7 years. 14.4% of residents were under the age of 18; 33.2% were between the ages of 18 and 24; 18.5% were from 25 to 44; 20.5% were from 45 to 64, and 13.3% were 65 years of age or older. The gender makeup of the city was 47.3% male, and 52.7% female. Winona also has three institutions of higher education. These are Winona State University, Saint Mary's University of Minnesota, and Minnesota State College-Southeast Technical.

When comparing the numbers for the estimated census data from 2015, the County of Winona has decreased in population. Winona County had an estimated population of 51,213 in 2015, which was a decrease of about two hundred people. There was also a decrease for all age groups. Women still made up the majority of individuals in the county with 26,030 people. The city of Winona's population was projected at 27,437 for 2015.

The county has a total area of 642 square miles, of which 626 square miles is land, and 15 square miles is water. Winona County has nine bodies of water. These include: Airport Lake, Bartlet Lake, Bollers Lake, Hunters Lake, Lake Goodview, Lake Winona, Rileys Lake, Mississippi River, and Whitewater River.

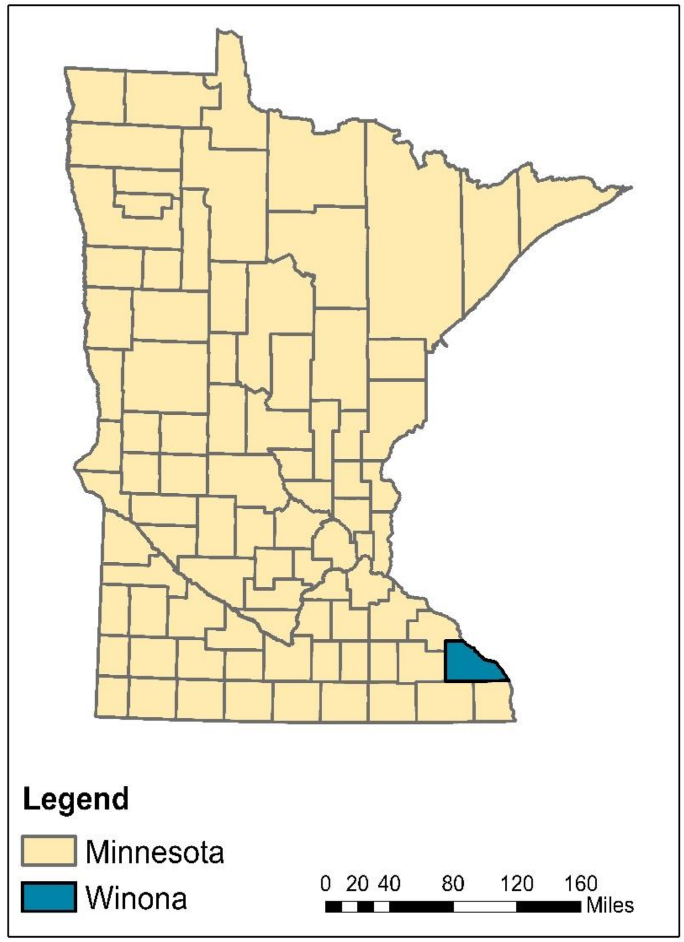

Figure 1. Winona County in the State of Minnesota.

Data Collecting and Processing

The data for the project were divided into four parts. These data consisted of voting, streets, census blocks, and census data.

Voting data

Voting data were obtained from the Minnesota Secretary of State's office. The data used for this study were registered voters in the Winona County. The data used consisted of people's names, house number, street numbers, zip codes, and registration dates. There were other data not used. These data included districts for location in Winona County. The data used were downloaded into Environmental

Systems Research Institute (ESRI's) ArcMap so that it could be filtered from an excel file and used as needed.

Data for the study were for the date of November 8th, 2016. This data allowed selection of attributes and were exported into a table called EDR for understanding purposes of this study.

The EDR table provided a list of people who used EDR. House numbers and street names were separate and were combined. The new address field contained each of the cardinal directions of north, south, east, and west. The cardinal direction was required, so house numbers were listed first in address fields.

Streets Data

Street data was used for geocoding voter data. The 2016 roads data were used, and it was obtained from the Winona County Planning and Zoning department.

Census Block Data

Census block data were obtained from the census.gov website. Block groups, which range from 800 to 3000 people (Teff, 2005) were the geographic unit used in this study as it was the smallest unit of data available. It is data of the population and divided into different types of categorical information in the block groups (Teff, 2005). Block groups were chosen to minimize statistical inaccuracies to the extent possible (Teff, 2005). For Winona County, the 2010 census data contained 38 blocks groups, and the data was obtained in the form of a shapefile.

Census Data

Census data was obtained from the American Fact Finder. It was obtained as Excel files from summary file 1 (SF1) data for the 2010 census. The data collected for this project contained information about sex, age, and housing. The information was used for creating and testing the model of Winona County that would have a good correlation with the EDR.

SF1 data were utilized to group individuals as homeowners, renters, gender, age, and race. All were used to determine the best variables for the study. Some of the categories from the SF1 data that were utilized in the study were males, females, and total population for people 18 and older. These were subdivided into sub-categories, which were a combination of age and gender. Age was divided into four-year intervals for both genders. These data were used because the minimum age for voting is 18.

Methods

The methods used for the study contained different stages; each stage needed to be completed before going onto the next stage.

Geocoding

The first step consisted of geocoding voter data. This was done by creating an address locator. The address locator that was used was a dual range geocoder. The reference data used were the Winona roads layer and the information in this table along with links to the address locator fields. Table 1 indicates field and address locator information used to complete the process for identifying the location of each EDR user. After creating the address locator, the geocoding tool used the newly created address locator and the EDR table to identify the address for each person using EDR. The input fields for the geocoding

Table1. Fields for the address locator joined to Winona roads. Map names indicate fields in the address locator, and Road fields indicate Winona roads field names and their connections in the address locator.

Map Names	**Road Fields**
From Left	FROMLEFT
From Right	FROMRIGHT
To Left	TOLEFT
To Right	TORIGHT
Prefix Direction	PREDIR
Street Name	FULLNAME
Left ZIP Code	LEFTZIP
Right ZIP Code	RIGHTZIP

tools were the address and zip code. Figure 2 shows the geocoding result for locations of each person that used EDR. Table 2 shows the number of EDR users for Winona County. It shows the mean, maximum number of users in a census block group, the minimum number, maximum, and standard deviation. Figure 3 shows the locations of the census block groups for the maximum and minimum numbers.

Table 2. The descriptive statistics for EDR for the 2016 election. EDR is the total number of EDR. Summarized are the mean, maximum, minimum, and standard users per block group.

EDR	**Mean**	**Max**	**Min**	**Std.**
3836	100.9	440	39	69.7

Select by Location and Select by Attribute

The census block group shapefile and EDR points were combined to determine how many points occurred within each census block group. This was accomplished by using the "select by an attribute" on the census blocks shapefile and choosing each of the census blocks

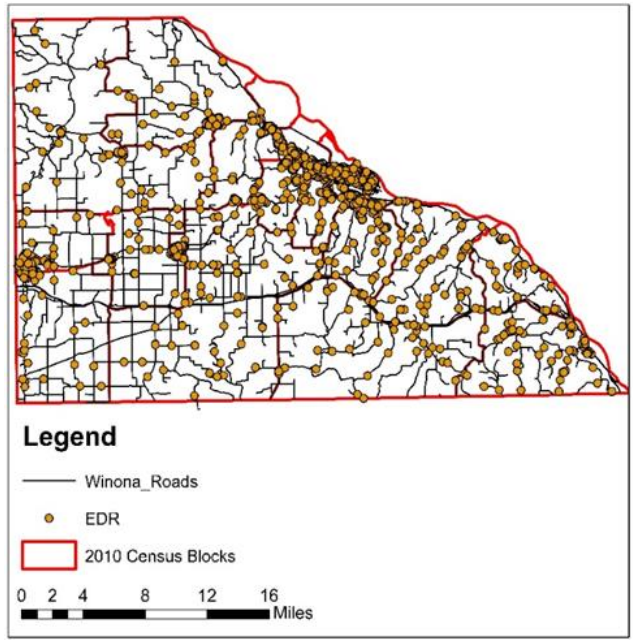

Figure 2. EDR participants geocoded for Winona County along with the roads and census blocks.

groups, one at a time for all 38 block groups of the county. The number of points in each census block group were then selected. Figure 4 shows each of the Winona County census block groups with the number of EDR users in that census block group.

SPSS

A regression analysis was conducted to determine the census demographic groups having the best relationship with EDR. Regression analysis is a statistical process for estimating the relationships among variables with the dependent variable. Here the dependent variable was the number of EDR for each census block group. There are different regression techniques for modeling and analyzing variables when the focus is on the relationship between a dependent variable and one or more independent variables or predictors (Gimpel and Schuknecht, 2003).

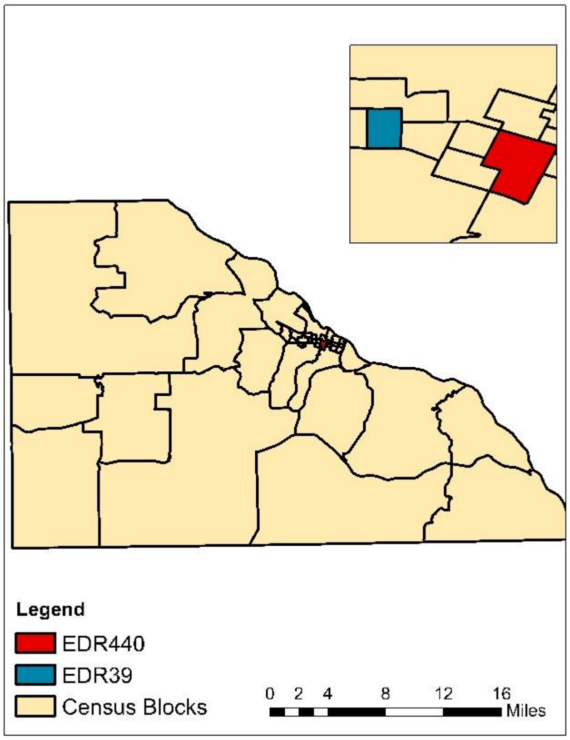

Figure 3. The areas in Winona County that had 39 EDR users, which was the minimum number of users for a census block group, and 440 which was the maximum number.

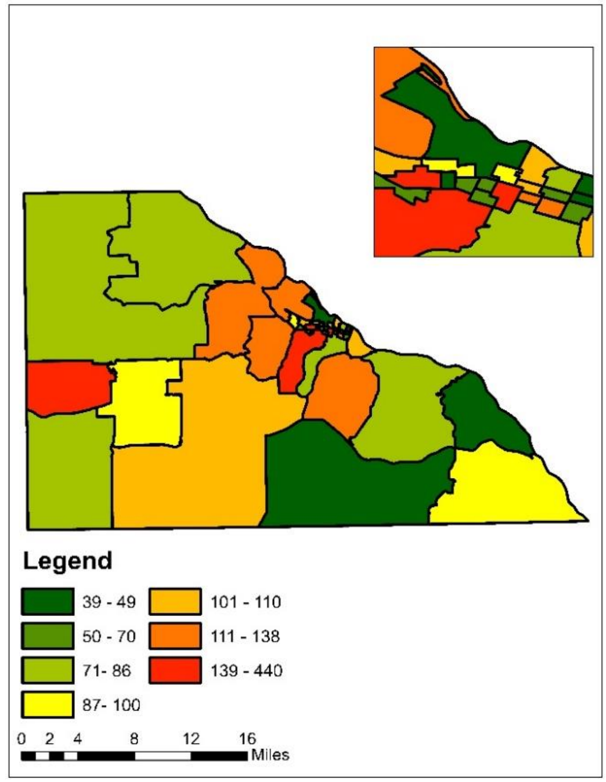

Figure 4. Number of EDR participants in each census block group shown in categories differentiated by seven colors where the darker green is the lowest and red is the highest number of EDR registrants. The right upper corner insert of the city of Winona.

This study used the stepwise multiple regression analysis method. This was conducted to determine the relationship between the census data variables and EDR. The equation used for this analysis was as follows (Zar, 2010):

$Y = B_0 + B_1 + B_n X_n$

Y = Predicted variable; EDR here

X = Used to model or predict the dependent variable

B = Coefficients computed by the regression analysis which represents the strength and type of relationship that value X has to Y

N= The other coefficients that represent the variables for the model

Exploratory Regression

Environmental Systems Research Institute (ESRI) offers a method called Exploratory Regression which evaluates all possible combinations of variables to determine which variables have a best chance of increasing the probability for a model to solve the problem. In addition, the Exploratory Regression tool creates summary reports which compare all passing models. Reports show the best passing models with the census group demographics combinations used. The best passing model was decided by evaluating adjusted R-squared values. The Exploratory Regression tool is a data mining tool that will try all possible combinations of explanatory variables to determine which models pass all the basic Ordinary Least Squares diagnostics (OLS) (Mitchell, 2005). When there are multiple potential explanatory variables, the analysis determined which combination of the independent variable does the best job of explaining the variability in the dependent data.

While Exploratory Regression is like the Stepwise Regression method, Exploratory Regression is found in multiple statistical analysis software packages. While other regression analyses only look for models with the highest adjusted R-squared value, Exploratory Regression looks for models that meet all the requirements and assumptions of the OLS method (Mitchell, 2005).

Ordinary Least Squares

Ordinary Least Squares is the best known of all regression techniques and is a good starting point for spatial regression analysis (Mitchell, 2005). It provides a global model of the variable or process to understand or predict; it creates a single regression equation to represent that process (Mitchell). OLS allows linear regression to generate predictions or to model dependent variables regarding their relationships with a set of explanatory variables (Mitchell).

Ordinary Least Squares is a method for estimating the unknown parameters in a linear regression model, with the goal of minimizing the sum of the squares of the differences between the observed response values of the variable being predicted and those predicted by a linear function of a set of explanatory variables (Mitchell, 2005). Visually this is the sum of the squared vertical distances between each data point in the set and the corresponding point on the regression line (Mitchell). The smaller the differences, the better the model fit of the data (Mitchell). A simple formula can express the resulting estimator, especially in the case of a single independent variable (Mitchell).

The OLS estimator is consistent when the independent variables are exogenous, and optimal in the class of linear unbiased estimators when the errors are homoscedastic and serially uncorrelated (Mitchell, 2005). Under these conditions, the method of OLS provides minimum-variance mean-unbiased estimation when the errors have finite variances (Mitchell). Under the additional assumption, errors are normally distributed, OLS is the maximum likelihood estimator (Mitchell).

Geographically Weighted Regression

Geographically weighted regression (GWR) is one of several spatial regression techniques increasingly used in geography and other disciplines (Bivand, 2017). GWR provides a local model of the variable or process people are trying to understand or predict by fitting a regression equation to every feature in the dataset (Bivand). When used correctly, these methods provide robust and reliable statistics for examining and estimating linear relationships (Bivand).

GWR is an exploratory technique mainly intended to indicate where non-stationarity is taking place in the dataset, that is where locally weighted regression coefficients move away from their global values (Bivand, 2017). The only concern is that there are proper coefficient values of a global model. With using this method, the fitted points for all the data may not represent specific local variations in the data local regression implementations (Bivand). It differs, however in that it eliminates the local variation in data space and by moving a weighted window over the data. It thus estimates one set of coefficient values at every chosen relevant point (Bivand). The fit points are often the points where observations were made, but they are not necessary. If the local coefficients vary in space, they can be taken as an indication of non-stationarity (Bivand).

Results

The results were compiled for two census years. These years were 2010 and 2015.

2010 Census Results

The first step was to employ SPSS to create a model. The best model created used three predictor variables. These variables (Table 3) produced a R-square value of .954. These three variables explained 95.4 percent of the relationship for EDR in Winona County. The remaining 4.6 percent was due to unknown factors.

Table 3. Variable with the R-square and the three variables influence the relationships between EDR, and the three variables. Model 1 is females18 to 21 years of age. Model 2 is females18 to 21 and total number of owners and renters. Model 3 is females 18-21, total number of owners and renters and the total number of males 21 years of age. The last field is the Sig. F, which shows each variable for the model with the significant value that is under .05.

Models	**Variables**	**R SQ.**	**Sig. F**
1	F 18-21	82.7%	0.000
2	Total number	95.0%	0.000
3	Males 21	95.7%	0.021

Table 4 lists the names of the variables of the three-explanatory variable and their contribution to the R-square value. The second variable, total number of owners and renters, was a variable that has a strong correlation more for older people than that of younger people.

Table 4. Variables for the Regression along with the change in the R squared value associated with each variable.

Variables	**C R Sq.**
Females from 18-21 years of Age	82.7%
Total number of owners and renters	12.4%
Males 21 years of Age	0.7%
Total R square	95.7%

Using these three variables in ArcMap and by using the OLS to create the SPSS model areas, where there was a positive or negative standard deviation for that location, were identified. This model does not include spatial analysis relationships between the variables. Therefore, GWR was used to provide the result for the final model. Figure 5 is a depiction of the OLS model. It shows which census blocks have a low or high standard deviation where blue color is the lowest and red is the highest.

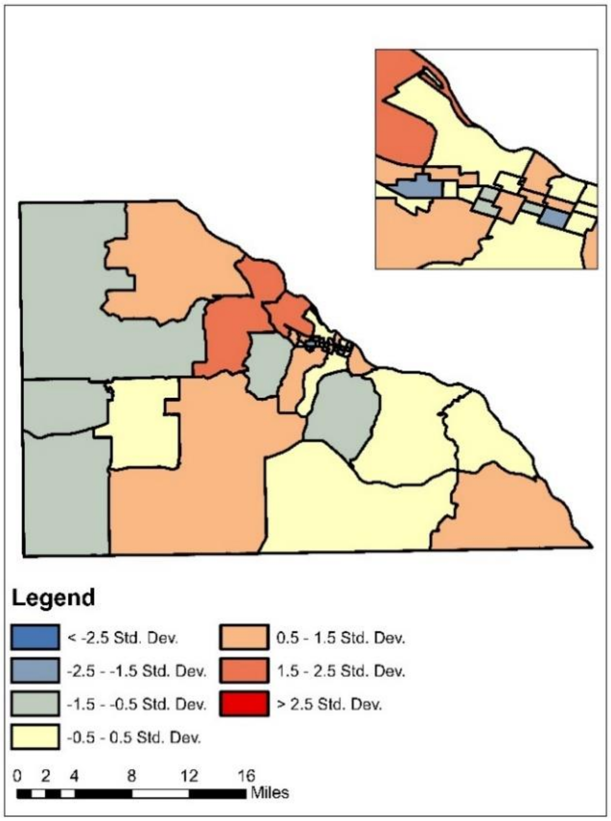

Figure 5. OLS model for the three variables for the 2010 data that used the methods SPSS and Exploratory Regression.

Figure 6 does not have any dark red, so the standard deviation of these variables does not have a positive 2.5 standard deviation or higher. This means the standard deviation indicates data points tend to be close to the mean.

The next analysis conducted used Geographically Weighted Regression (GWR). This analysis showed changes in the areas of the spatial relationships and did not modify the R-squared value. Figure 6 shows the areas in Winona County along with these changes and the effect that GWR has on the standard deviation in a positive or negative way.

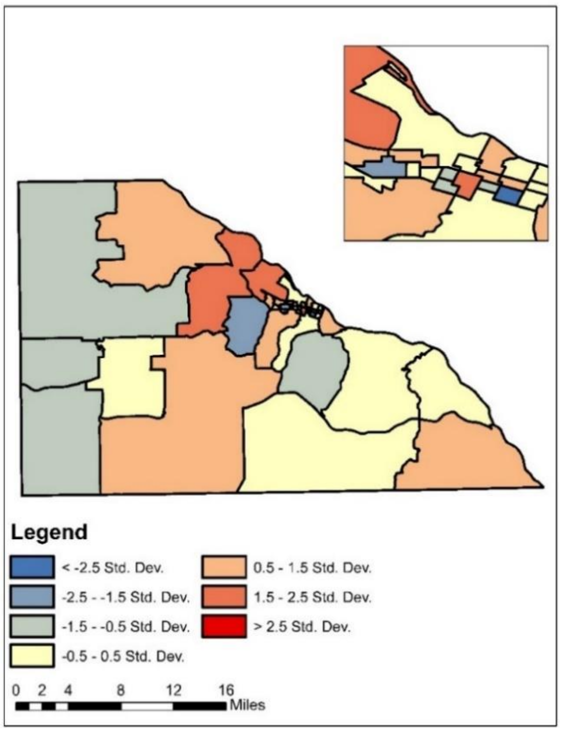

Figure 6. GWR model for the three variables for the 2010 data that used the methods of SPSS and Exploratory Regression.

The next analysis used the Exploratory Regression tool to find the highest R-square model. This model had three variables. The result was the same as with SPSS, and the R-squared was the same at a value of .954. Figures 5 and 6 show the effects of spatial relationships and the final product for the outcome of the analysis. When both methods were completed in Geographically Weighted Regression, they provided the outcome with the same result of all areas. The regression model for EDR in Winona County was:

Y = 5.726 + 270(F18¬_21) + .107(TNUHOR) + .406(M21_21)

2015 Census Results

The four variables in Table 5 produced an R-square number of .885 for 2015. Therefore, these four variables explained 88.5 percent of the relationship for EDR in Winona County. The remaining 11.5 percent were unknown. Table 5 shows the analysis for census data that are estimated for 2015 and how these results impacted the regression analysis. Table 5 did not have housing data for the analysis because it was not completed for that year. Figure 7 is a depiction of the OLS model for the 2015 data, and areas are positive standard deviation. Figure 8 is a depiction of the GWR model for 2015 data with changes small changes. The regression model for EDR in Winona County for 2015 was:

Y = 39.722+.540(F18_19) + 1.232(F35_39) + .431(M21_21) + -.396(M18_19).

Table 5. Summary of the impact of exploratory analysis of variables and EDR. Model 1 is females 18 to 19 years of Age. Model 2 is females 18 to 19 and females 35 to 39 years of age. Model 3 is females 18 to 19, females 35 to 39 and the total number of males 21. Model 4 is females 18 to 19, females 3 to 39, males aged 21 and males 18 to 19 years of age. The other field is the Sig. F, which shows each variable for the model with the significant value that is under .05.

Models	**Variables**	**R SQ.**	**Sig. F**
1	F18_19	74.8%	.000
2	F35_39	81.8%	.001
3	M21_21	85.3%	.007
4	M18_19	88.5%	.005

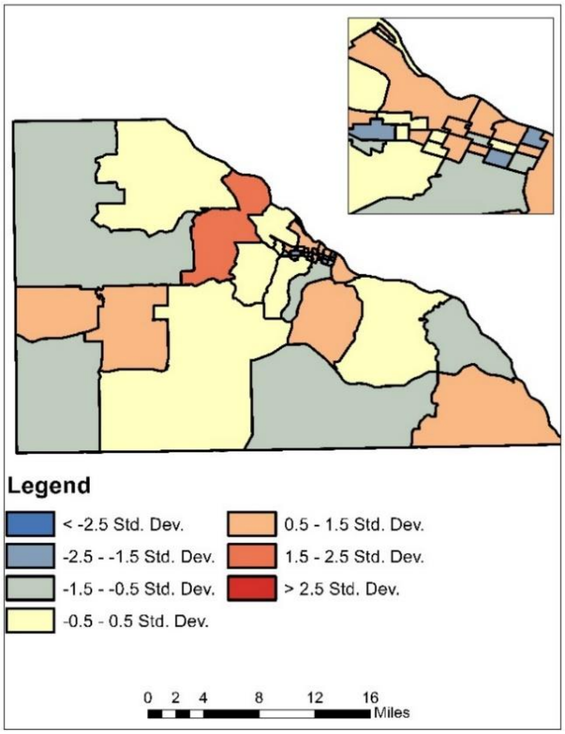

Figure 7. OLS model for the four variables for the 2015 data that used the methods of SPSS and Exploratory Regression.

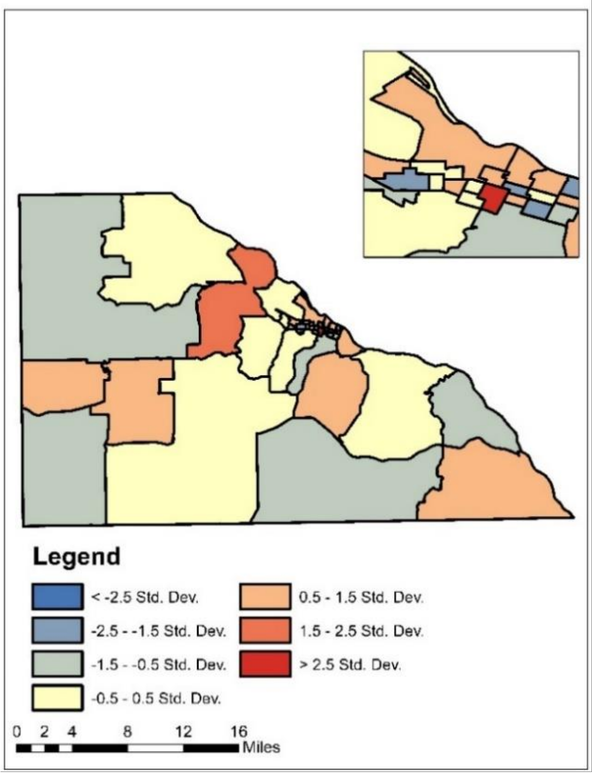

Figure 8. GWR model for the four variables for the 2015 data that used the methods SPSS and Exploratory Regression.

Discussion

Several types of EDR studies exist. For example, one type compares results from one area with another similar area in another county or state. Another uses a different county that has a larger population. The last incorporated change detection on the county, and evaluated if people in the same state for different years had an increase or decrease using EDR.

This study would have been improved if data could have been obtained from Summary File 3 tables. Summary File 3 data are more detailed and could have helped refine findings.

Results reveal the three variables in Table 3 with Figure 6 show there are eleven orange or higher colored areas. These areas suggest greatest use of EDR for the 2016 election. The orange areas in Figure 6 are areas containing at least 111 EDR users. The dark orange color is located around or in the city of Winona. There were two areas that had more than 139 EDR users and had a low standard deviation for the model. The first area was the city of Saint Charles. A possible explanation might be that this is an older aged community. The second area was in the city of Winona. There were three areas in Winona that had less than -2.5 standard deviation. All three locations had at least 100 individuals use EDR.

Figure 8 represents the 2015 census data. This shows twelve areas orange. One of the cities in orange is the city of Saint Charles. There is one area that has dark orange. There is one area that has red. The area that is red is located by Winona State University. The area that is dark orange is just outside the city of Winona. Here, three areas had 101 to 110 individuals using EDR, and those areas had a -.5 to .5 standard deviation. All three areas are in or near the city of Winona.

One census block had 139 to 440 EDR users and had a -.5 to .5 standard deviation. This was also located in the city of Winona.

There are few reasons why EDR is being used in Winona County at a rate higher than the state average. The first relates to the number of females 18 to 21. Winona County has a large number of people that are in this age range. The second is the city of Winona has three institutions of higher education which means more young people are moving into the county.

Conclusions

The purpose of the study was to create a profile for individuals that used EDR in Winona County. A model was created to identify areas where individuals used EDR. Using 2010 census data, three variables were identified as important predictors. These variables included: total number of owners and renters, females 18 to 21, and 21-year-old males. These three variables produced an R-square value of .957 or more clearly stated; they explained 95.7% of the variability in the EDR data. 95.7 percent for predicting EDR relationships for Winona County.

Using 2015 census data, four variables were identified as important predictors. These were: females 18 to 19, females 35 to 39, 21-year-old males and males 18 to 19. These four variables generated an R-square value of .885 or more clearly stated, they explained 88.5% of the variability in the EDR data, and 88.5 percent for predicting EDR relationships for Winona County. An important limitation impacting this work was that data from summary file 1 was used instead of the more complete summary file 3.

Acknowledgements

A special thank you and acknowledgment should go to the Minnesota Secretary of State staff who proved the data. In addition, a statement of thanks is shared with and to the Department of Resource Analysis staff at Saint Mary’s University of Minnesota for their guidance during this study.

References

Bivand, R. 2017. Regression, Geographically Weighted. https://cran.rproject.org/web/packages/spgwr/vignettes/GWR.pdf.

Brians, C. L., and Grofman, B. 2001. Election Day Registration's Effect on U.S. Voter Turnout. Social Science Quarterly,82(1), 170-183. doi:10.1111/0038-4941.00015.

Data Access and Dissemination Systems (DADS). 2010, October 05. American FactFinder. Retrieved July 03, 2017, from http://factfinder.census. gov/faces/nav/jsf/pages/index.xhtml

Gimpel, J., and Schuknecht, J. 2003. Political participation and the accessibility of the ballot box. Political Geography, 22(5), 471-488. doi:10.1016/s0962-6298(03)00029-5.

Mitchell, A. 2005. Esri Press. Retrieved June 02, 2017, from http://esripress .esri.com/display/index.cfm?fuseaction =display&websiteID=86&moduleID=0

Neiheisel, J. R., and Burden, B. C. 2012. The Impact of Election Day Registration on Voter Turnout and Election Outcomes. American Politics Research,40(4), 636-664. doi:10.1177/1532673x11432470.

Office of the State of Minnesota Secretary of State 2017. Retrieved February 27, 2017 from http://www.sos.state.mn.us.

Teff, C. P. 2005. Who Uses Election Day Registration? A Case Study of the 2000

General Election in Anoka County, Minnesota. Saint Mary's University of Minnesota University Central Services Press. Winona, MN. 1-14.

Zar, J. H. 2010. Biostatistical Analysis. 5th Edition. Prentice Hall, Inc., Upper Saddle River, NJ. 1-960.