Radon Mapping Final Report
Geographical and Sustainability Sciences

The University of Iowa
IOWA INITIATIVE FOR SUSTAINABLE COMMUNITIES
iisc.uiowa.edu

Class Led by Sean Young
Course: Health and Environment: GIS Applications

In partnership with Siouxland Interstate Metropolitan Planning Council, the City of Sioux City, and the Siouxland District Health Department

12.2015

simpco IS ACTION!
SIOUX CITY
SIOUXLAND DISTRICT HEALTH DEPARTMENT
The University of Iowa | College of Liberal Arts & Sciences
DEPARTMENT OF GEOGRAPHICAL & SUSTAINABILTY SCIENCES

A PROGRAM OF
The University of Iowa Office of Outreach & Engagement
OUTREACH.UIOWA.EDU
outreach-engagement@uiowa.edu · 319-335-0684

This project was supported by the Iowa Initiative for Sustainable Communities (IISC), a program of the Provost's Office of Outreach and Engagement at the University of Iowa that partners with rural and urban communities across the state to develop projects that university students and faculty complete through research and coursework. Through supporting these projects, the IISC pursues a dual mission of enhancing quality of life in Iowa while transforming teaching and learning at the University of Iowa.

This publication may be available in alternative formats upon request.

Iowa Initiative for Sustainable Communities
Provost's Office of Outreach & Engagement
The University of Iowa
111 Jessup Hall
Iowa City, IA, 52241
Email: iisc@uiowa.edu
Website: http://iisc.uiowa.edu/

Lizzy Mello

GIS Applications: Health and Environment

14 December 2015

Final Paper: **Examining Radon Levels in Woodbury County, Iowa**

Abstract

Radon is an important issues across the global as it is a leading cause of lung cancer [5]. Many people are exposed to radon mainly in their homes, where this radioactive gas can build up to dangerous levels [5]. This is of special interest in Iowa where more than 50% of the homes test above the EPA action level for radon [9]. Working with Siouxland District Health through an IISC partnership, this geospatial analysis looks at radon testing from January 2014 to September 2015 to uncover any areas of high exposure and assess potential population predictors of high radon levels. Using data from Siouxland District Health's radon test kits program and the US Census, regression modelling and hot spot analysis was conducted in ArcMap 10.3. Regression modelling did not reveal any conclusive results, with the only significant predictor variable being ZCTA gender composition. However, distributions showed that ~~than~~ most of the counties tested above the action level for indoor radon and hot spot analysis indicated areas where mean radon levels were especially high. Results also called into question data quality's influence on the results, due to the relationship between data quality measures and the outcome of interest. Recommendations are included to modify the radon test program to improve data quality and the quality of the results.

Introduction

Cancer remains a major source of mortality for the US. Cancer is the second leading cause of death in the US overall and is in the top ten causes of death for all age groups apart from infants [1]. Lung cancer is remains the second most common type of cancer with an estimated 221,200 new cases and 158,040 deaths in 2015 [2], despite decreases in incidence rate over the past decade [3]. The state of Iowa has lung cancer incidence of 61.7 cases per 100,000 residents per year, which is on par with the national incidence rate of 60.4 cases per 100,000 residents [4].

Radon is the second leading cause of lung cancer cases after smoking, causing between 3% and 14% of lung cancer cases worldwide [5]. It is a naturally occurring gas, emitted from decaying uranium in the ground, and is radioactive [5]. Normally, radon is released into the atmosphere and is diluted to a harmless level [5]. However, when radon collects in confined spaces, such as basements and mines, it can rise to a harmful concentration, putting occupants at risk [5].

Though radon was first documented in miners, exposed to high concentrations during their work, there is evidence that even exposure to lower doses, such as in a home setting is dangerous and increases one's risk of developing lung cancer [5]. The US Environment Protection Agency has set the indoor actionable radon level at 4.0 cPi/L, though some argue that any radon exposure is dangerous [6]. The national outdoor radon level is 0.4 pCi/L, which is the target goal, which if achieve who mean that it is no more risky to be in your home than outside it in regards to radon exposure [6]. Risk of lung cancer also has a dose response relationship with radon exposure as with every 2.7 pCi/L increase in radon concentration, an exposed individual's risk of lung cancer increases 16% [6]. Radon also disproportionately affects smokers, increasing their risk more than their nonsmoker counterparts [6, 7].

Radon can enter the home in many ways. If there is radon being emitted from the underlying soil and rocks, then it can leak into a home through cracks, gaps, pipes, or porous materials in the floors and wall, usually in the basements where there is the most contact with the soil [5]. Whether the radon builds up to a dangerous level depends on the construction of the home and the amount of ventilation or air exchange [5]. Because the amount of indoor radon depends on the characteristics of the home, indoor radon levels can vary significantly between nearby homes [5]. For example, a recent study found that home age was not associated with radon levels in home [8].

Radon is an issue of particular interest to the state of Iowa as well. Iowa has been named a Zone 1 state for radon levels, which is a designation given to states where more than 50% of the homes test above the EPA action level of 4.0 pCi/L [9].

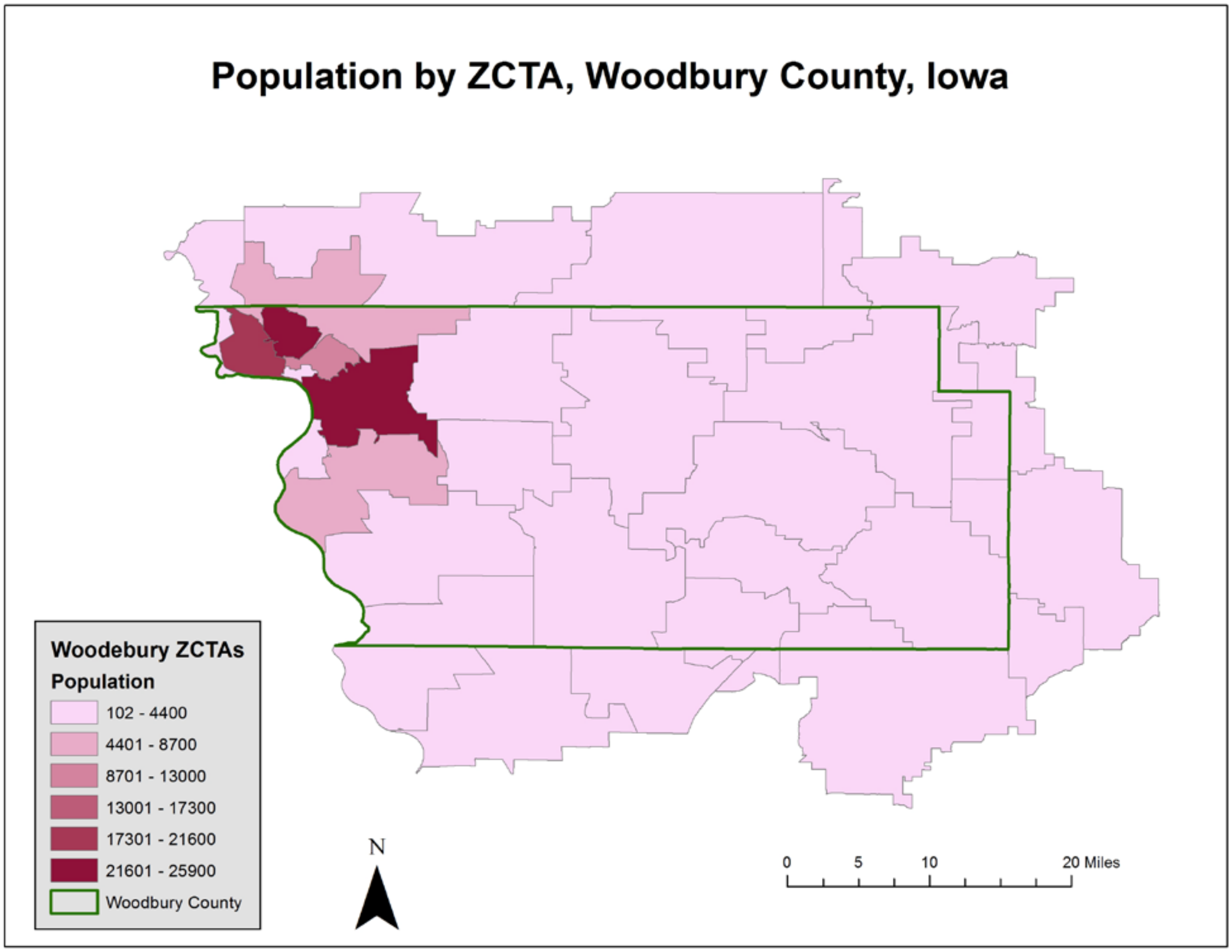
This project was developed in partnership with the Iowa Initiative for Sustainable Communities (IISC). The IISC partners communities throughout Iowa with classes and departments at the University of Iowa to develop projects that are mutually beneficial for students seeking real world experience and communities needing research completed. The community partner on this project was the Siouxland District Health Department of Woodbury County, Iowa. This organization was interested in exploring the relationship between radon levels and lung cancer incidence in their county at a finer scale that had previously been analyzed. When it was discovered that the lung cancer incidence data was unavailable in a timely manner, the focus of this project shifted to focus on the radon data only, assessing the geographic distribution and data quality. The aim is to provide Siouxland District Health with predictors of high radon levels, a known health risk, and recommendations for obtaining better radon level data. The study questions are:

1. What is the geographic distribution of radon at the zip code level Woodbury County, IA?
2. Are there areas where radon is especially high?
3. What other factors influence the radon level and do they vary by zip code?

Materials and Methods

For this analysis, the area of interest was defined as all ZCTAs intersecting the Woodbury County boundary (Map1). This means that the ZCTA could just share a boundary or overlap the county line. This was done because Siouxland District Health requested analysis at the zip code level, however zip codes are not always made to fit within administrative boundaries and this provided a larger sample size. Additionally, the boundaries of zip codes are frequently not released for official publication, because they are just a collection of mail routes, which presumable can change [10]. Instead ZCTAs, or zip code tabulation areas, are the areal representations of those zip codes. Zip codes were verified with their corresponding ZCTA codes twice using the US Census American Fact Finder and the UDS Mapper Zip codes to ZCTA crosswalk [10, 11].

Map 1: Area of Interest



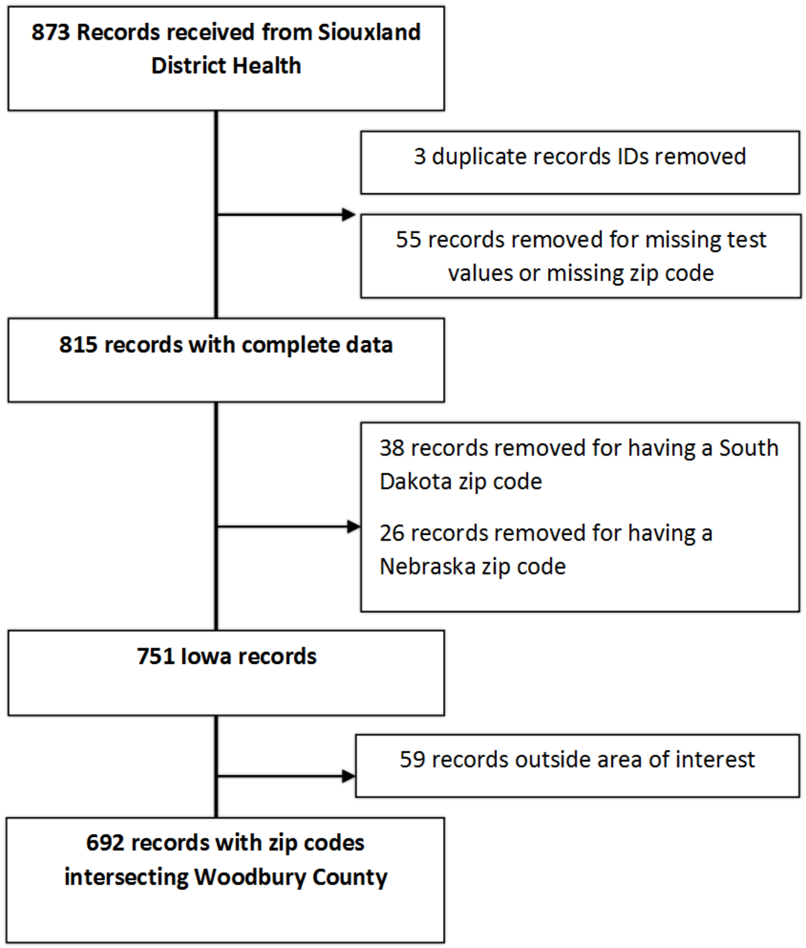
Radon data was provided by Siouxland District Health, which included the test values for all kits collected between January 2014 and September 2015 [12]. These data sets included the date of the test, test value/ outcome, and zip code where the test was collected. The population demographic variables of population size, race/ ethnicity, and sex ratio, were collected from the 2014 estimates for US Census data [13]. Remaining variables were calculated in ArcMap 10.3 [14].

Distance to Siouxland District Health was calculated using the Near function in ArcMap. The location of Siouxland District Health was geocoded and centroids were calculated for ZCTAs since addresses were not provided. A Near function was run from the centroids to the SDH location, calculating the distance.

Mean and median radon levels were calculated for each ZCTA in Excel prior to importing the data to ArcMap. This was done to allow the ZCTAs to be comparable, because a rate is not relevant to the radon level data set, and since point data was not provided.

The original set of radon test values had n=873. Data was removed if it was incomplete or a duplicate record. Additionally, there were 64 records removed for having a zip code in another state and another 59 records for being in Iowa but outside the area of interest. The remaining 692 records inside the area of interest were then aggregated to the ZCTA level. All shapefiles were then projected to UTM Zone 15 projections since it is an ideal projection for small, regional areas.

Figure 1: Sample size diagram of radon tests



Distributions were displayed in both Excel, as charts, and in ArcMap, as maps and histograms. The average radon value for each ZCTA was analyzed for hot and cold spot using the Hot Spot (Getis-Ord Gi*) function. Hot Spots were calculated for both average radon level in each ZCTA (Map 5) and the radon testing rate for each ZCTA (Map 6). Hot spots for both of these variable were calculated using fixed distance method and contiguity with corners methods to examine any differences in results. This was done since the area of interest is so small.

Models were constructed to predict the mean radon level in each ZCTA. Both Ordinary Least Square and Geographically Weighted Regression models were run using various demographic and geographic variables. Some results were confirmed with OpenGeoDa [15] weighted modelling. Models were assessed for spatial autocorrelation using the Moran's I statistic. Geographic variations in the model's predictability was examined by mapping the R^2 value, variable coefficients, and residuals at the ZCTA level.

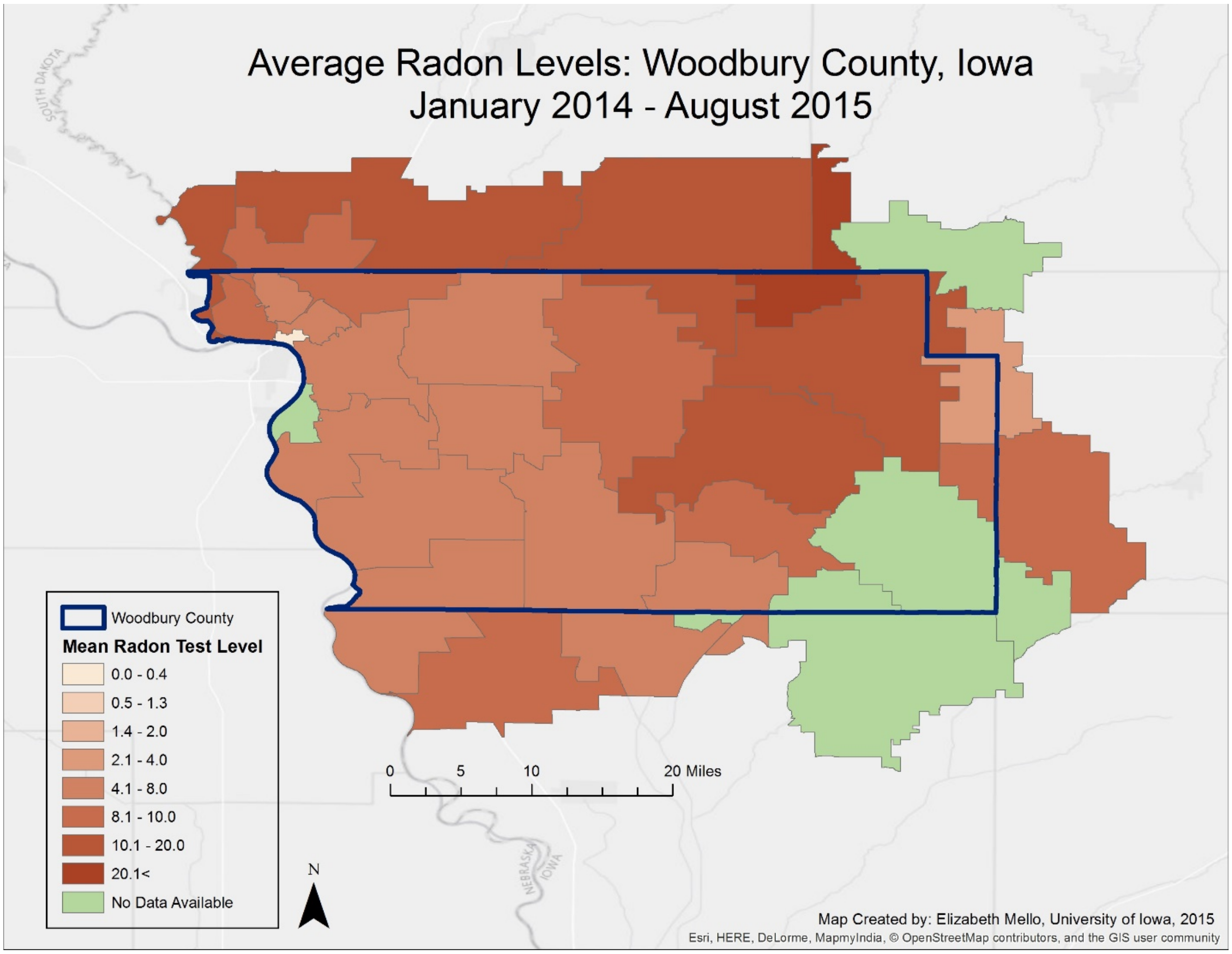
Results

The maps of mean radon level and difference from the mean show the difference in average radon levels across the area of interest (Map 2). There seems to be higher levels in the east part of the county that in the western part, however, there are also a couple ZCTAs that are missing data, mostly in the eastern part of the county. Map 3, shows that many of the ZCTAs are close to the county mean of 7.1, but many are higher. One ZCTA is much higher than the average. Those that are dark green, indicating that they are much lower than the average, appear to be the same ZCTAs that are missing data. This indicates that hot and cold spot analysis might be more indicative of disparities than just the mean.

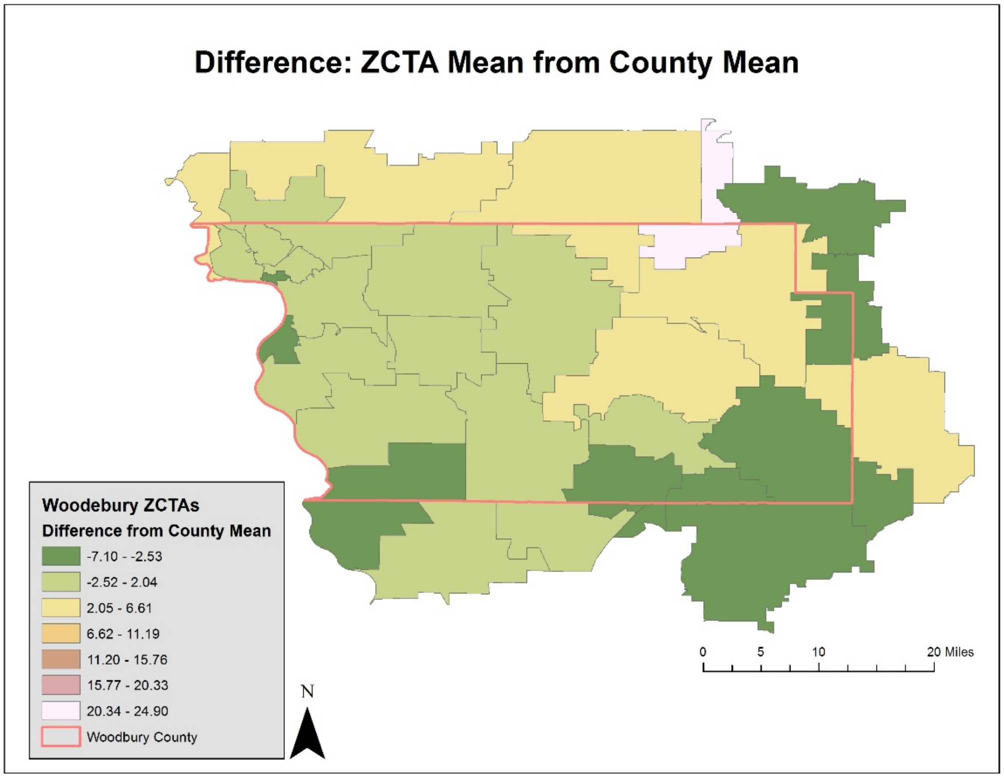
Table 1: County Characteristics for Woodbury County

Number of ZCTAs in area of interest	*29*
Mean radon level (pCi/L)	7.1
Median radon level (pCi/L)	5.5
Maximum detected radon level (pCi/L)	46.2
Minimum detected radon level (pCi/L)	<0.3
Number of ZCTAs with no data	5
Number of ZCTAs with means above action level of 4 pCi/L	22

Map 2: Average ZCTA radon levels (pCi/L) in Woodbury County, IA



Map 3: Difference from county mean radon level (7.1 pCi/L)



The maps of the number of records for each ZCTA also shows an east to west relationship, which seems to be the inverse of spatial pattern of average radon levels (Map 4). Figure 2 shows that there is a large disparity between the numbers of test records for each of the ZCTAs in the area of interest.

<u>Map 4: Comparing mean radon level (pCi/L) and number of test records</u>

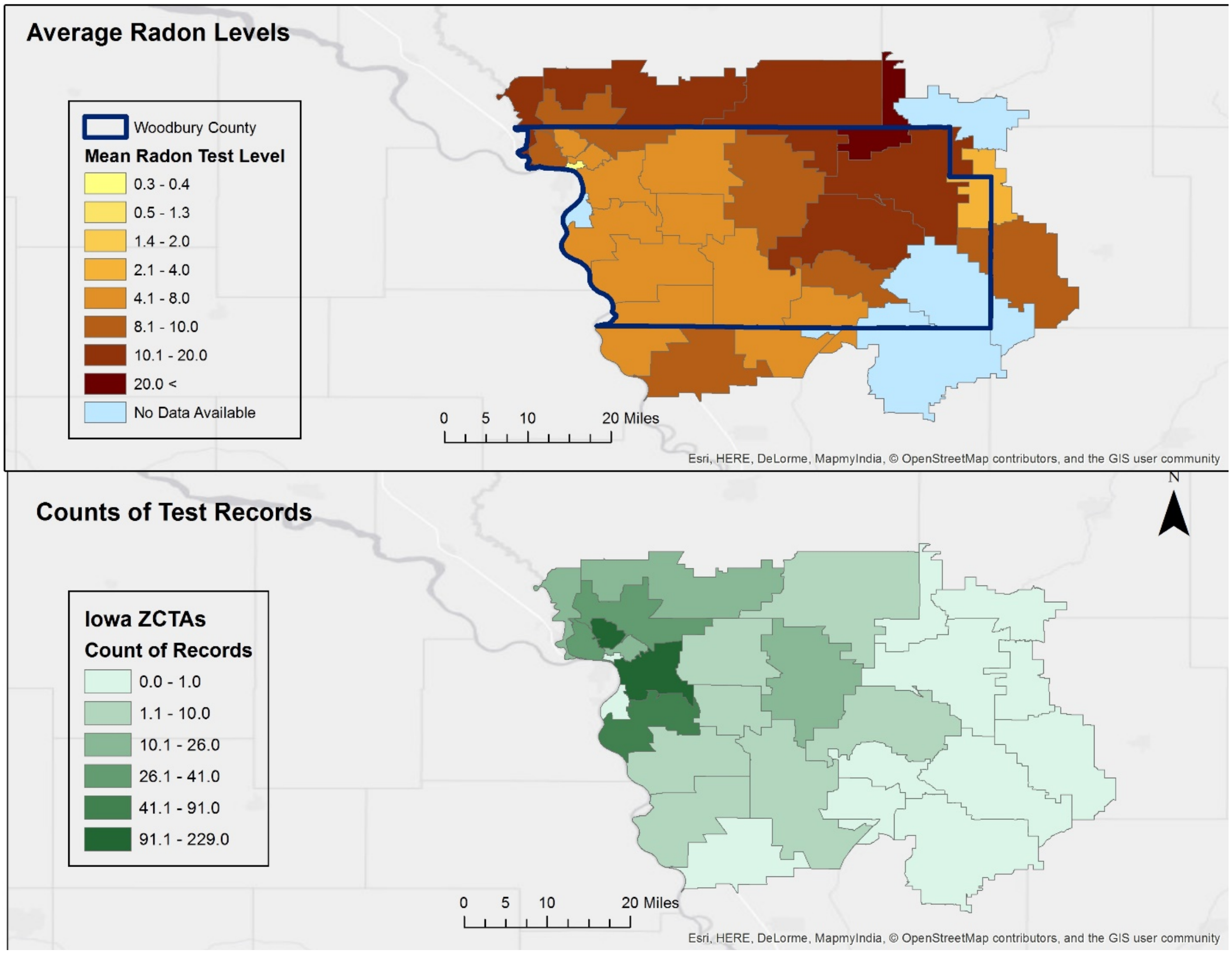
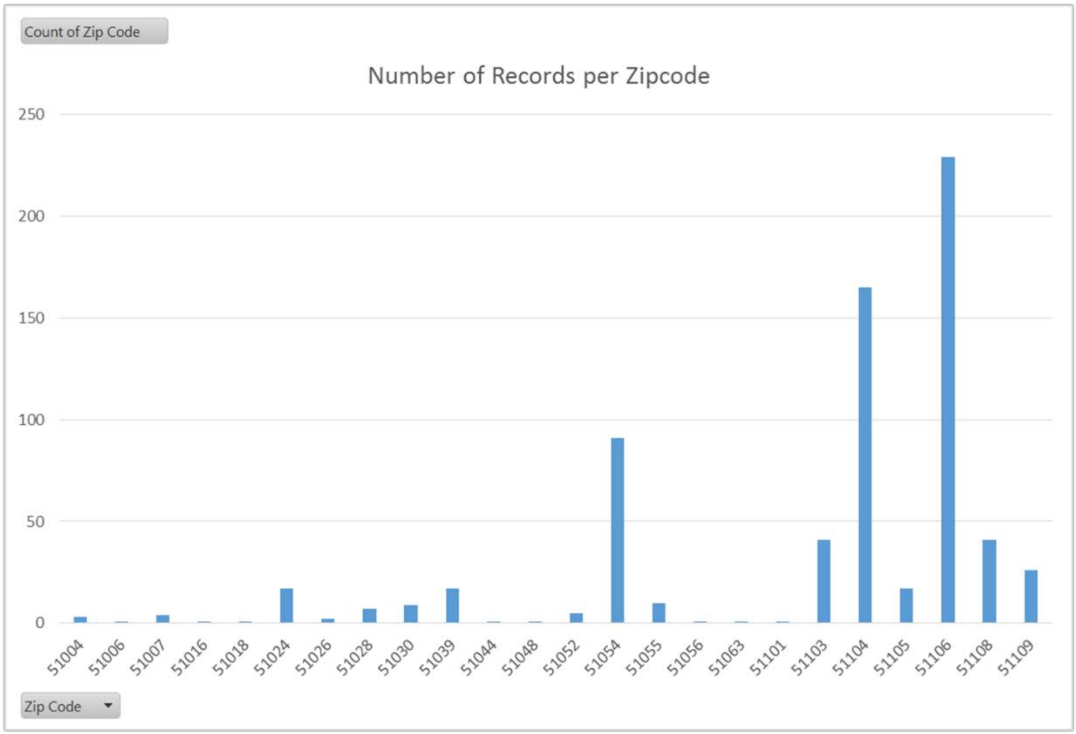
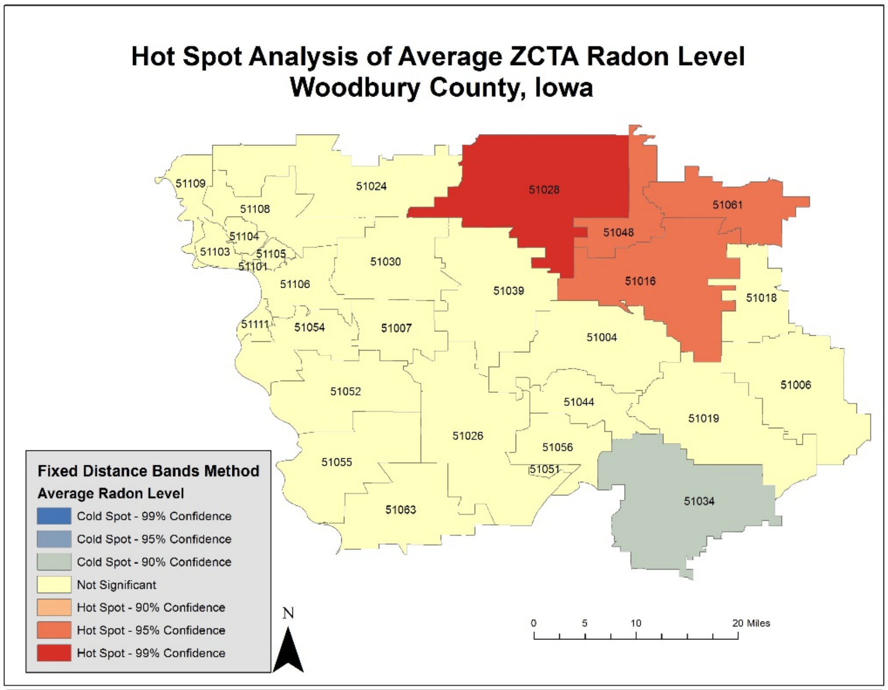


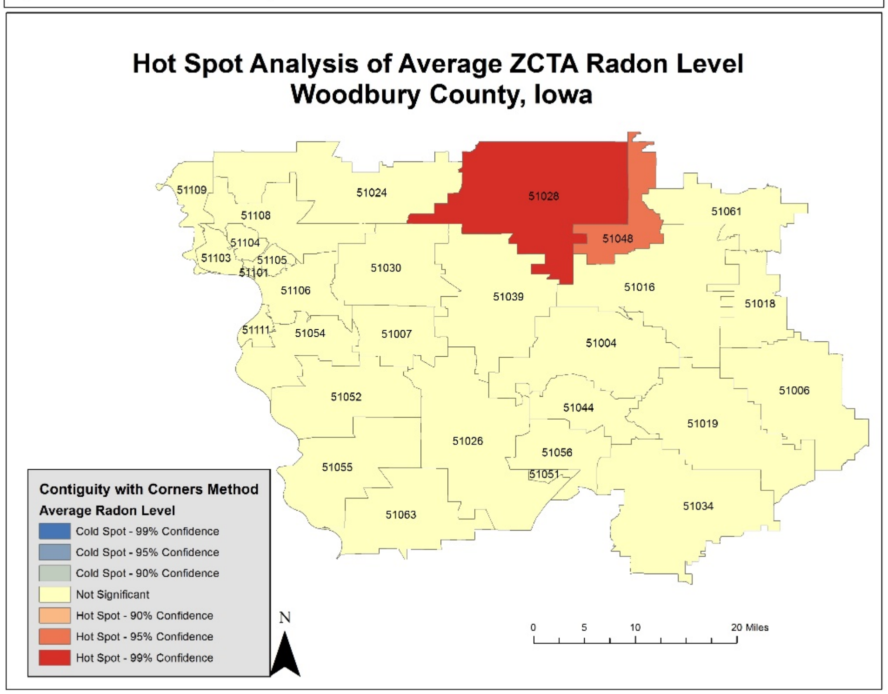
Figure 2: Distribution of records within sampled ZCTAs



The hot spot analysis of mean radon levels showed hot spots in the northeast corner of Woodbury County. The fixed distance hot spot was larger, including four ZCTAS, than the contiguity corners hotspot, but both have the same 99% confidence hotspot of ZCTA 51028. The fixed band method also had one cold spot, but only at 90% confidence.

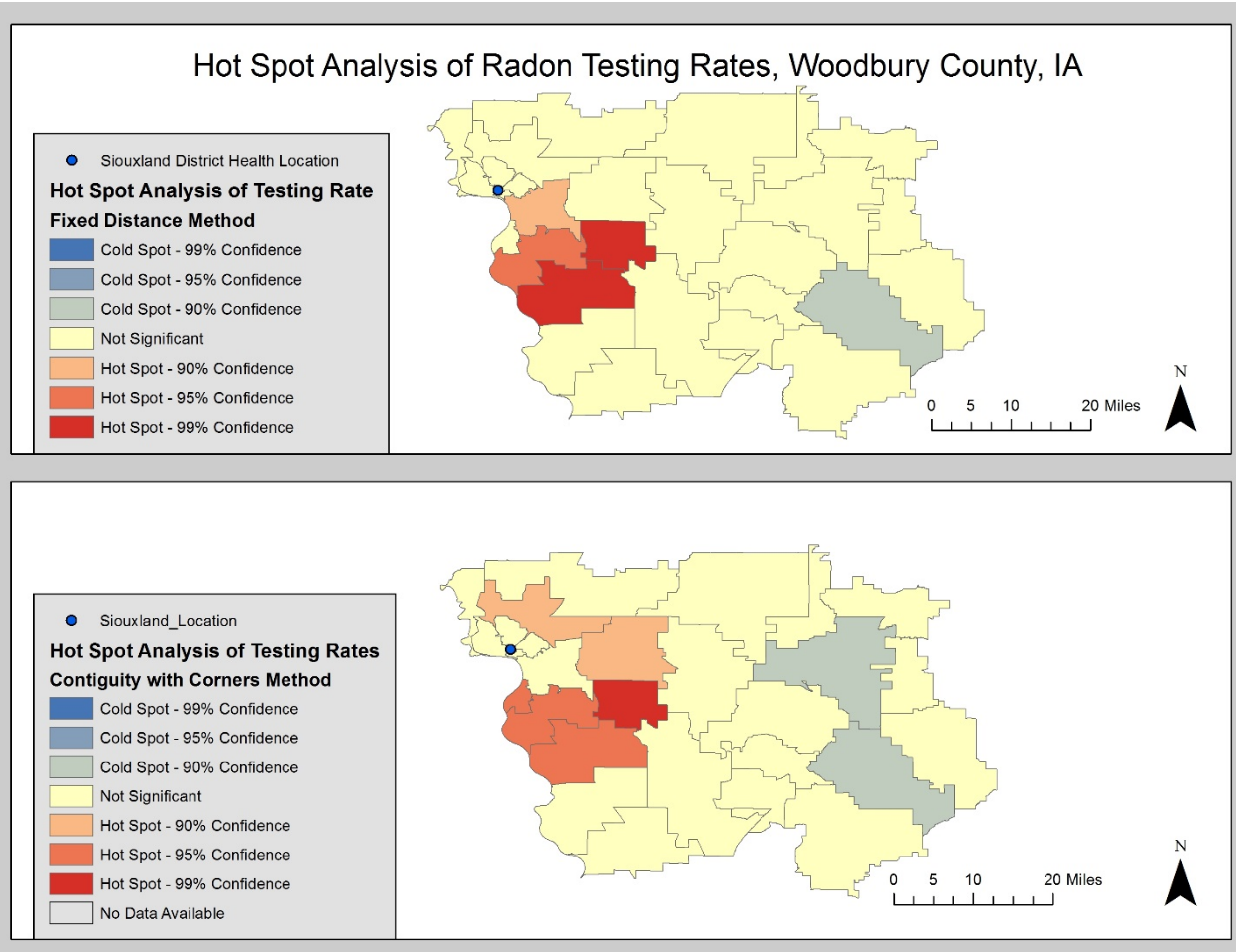
<u>Map 5: Hot spot analysis of mean radon levels</u>





An additional hot spot analysis was run on the testing rates for each ZCTA (number of ZCTA test records/ ZCTA population) to account for the differences in population across Woodbury County, since it includes both urban and rural areas. The results showed hot spots close to Sioux City and cold spots in the eastern part of the county, meaning that there were disparities even when accounting for population size.

Map 6: Hot spot analysis of testing rates



Several models were run to determine the best model to explain the outcome of average radon level for the 29 ZCTAs. First an Ordinary Least Squares (OLS) model was run to assess the global predictability. The original model (Model 1) was not very predictable with a high AIC and very low R^2 values. Variables were removed in attempts to improve the predictability of the model.

Table 2: Model Components

	Model 1	*Model 2*	*Model 3*
Type	OLS	OLS	GWR
Variables included	Distance, number of records, sex ratio, % white, median age	Distance, number of records, sex ratio	Distance, number of records, sex ratio
AIC	202.21	195.59	187.11
Adjusted R^2	-0.017	0.061	0.446
F-statistic	0.90	1.60	-
Jarque-Bera Statistic	70.13*	68.47*	-
Moran's I Index	0.045	0.045	-0.050
Moran's I Zscore	0.669	0.704	-0.120
Moran's I p-value	0.503	0.481	0.912

Table 3: Model 1 OLS summary

Summary of OLS Results - Model Variables

Variable	Coefficient [a]	StdError	t-Statistic	Probability [b]	Robust_SE	Robust_t	Robust_Pr [b]	VIF [c]
Intercept	23.569589	16.528934	1.425959	0.167311	13.246350	1.779327	0.088404	--------
COUNT_OF_R	-0.013763	0.027606	-0.498576	0.622812	0.015672	-0.878194	0.388922	1.443346
NEAR_DIST	-0.000022	0.000078	-0.276085	0.784948	0.000067	-0.322106	0.750281	1.851605
PERCWHITE	-4.273982	17.666983	-0.241919	0.810987	13.286451	-0.321680	0.750600	1.865473
MEDIANAGE	0.058300	0.274606	0.212304	0.833740	0.182228	0.319929	0.751910	1.783924
SEXRATIO	-0.133919	0.064242	-2.084620	0.048396*	0.037797	-3.543099	0.001735*	1.163230

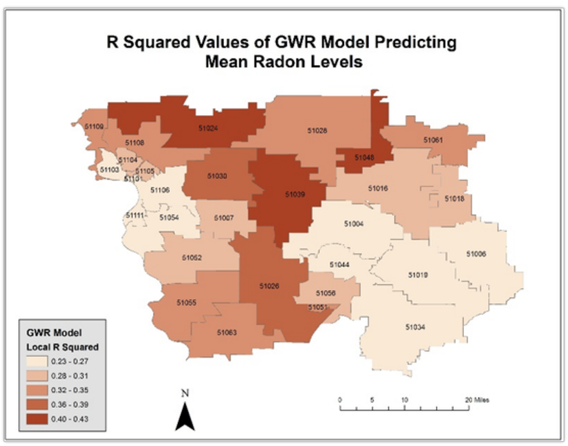
Table 4: Model 2 OLS diagnostics

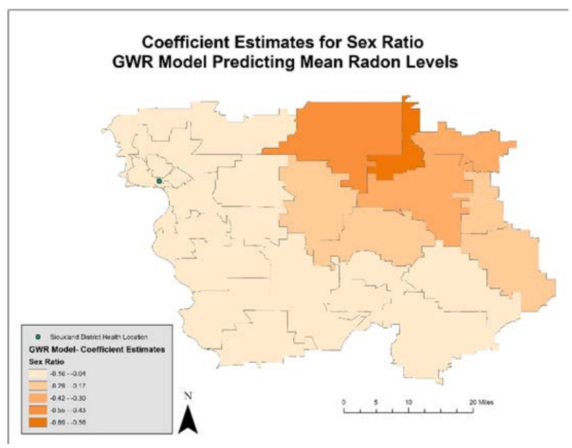
Summary of OLS Results - Model Variables

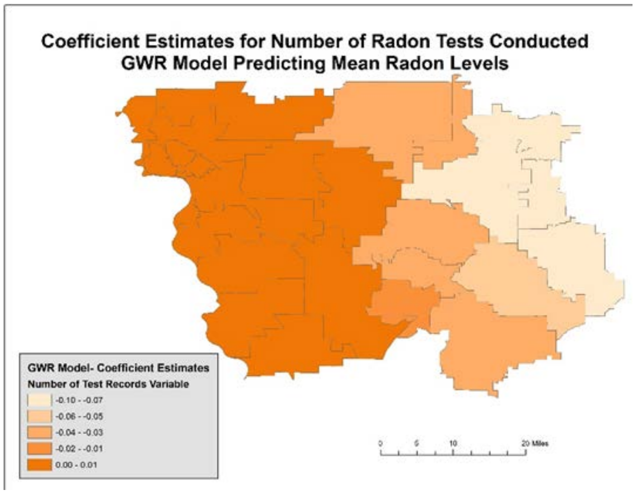
Variable	Coefficient [a]	StdError	t-Statistic	Probability [b]	Robust_SE	Robust_t	Robust_Pr [b]	VIF [c]
Intercept	21.576188	6.962496	3.098916	0.004756*	5.549556	3.887912	0.000659*	--------
COUNT_OF_R	-0.014806	0.025900	-0.571657	0.572658	0.012966	-1.141855	0.264334	1.376568
NEAR_DIST	-0.000026	0.000064	-0.404659	0.689168	0.000052	-0.492298	0.626803	1.331678
SEXRATIO	-0.128803	0.058783	-2.191152	0.037976*	0.041707	-3.088290	0.004880*	1.055285

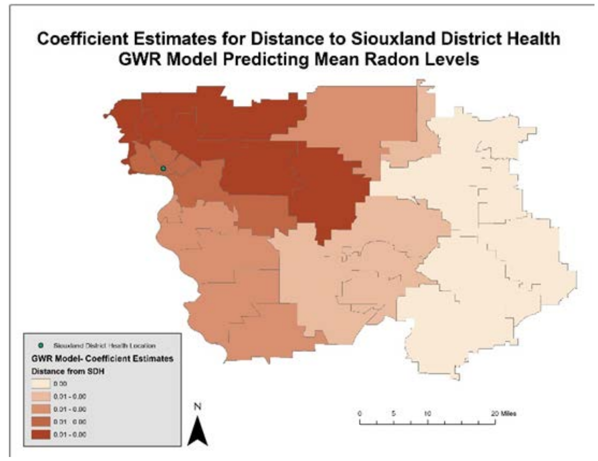
Since the five variable model was not able to be run in a GWR model due to multicollinearity, alternate approaches were used. A three variable model was analyzed as an OLS model (Model 2) and a GWR (Model 3), both of which were an improvement on the original model, especially the GWR. The GWR model was run with an adaptive kernel and AICc bandwidth. The GWR model had the lowest AIC and the highest R^2. It also showed variation in variable influence and model predictability across space with the varying spatial distributions of the local R^2 values and the coefficient estimates for predictor variables (Map 7). Moran's I statistic tests for all three models were not significant (p-value > 0.05), indicating no spatial autocorrelation.

Map 7: Differences in GWR model predictability and variable coefficients within area of interest









The five variable model was rerun using OpenGeoDa in an attempt to address the multicollinearity, however, the results were similar to the original OLS model (Supplemental File 5). The weighted spatial lag model did show some improvement in the model diagnostics, but not as much as the GWR three variable model. Therefore, the selected model was the GWR model.

Discussion

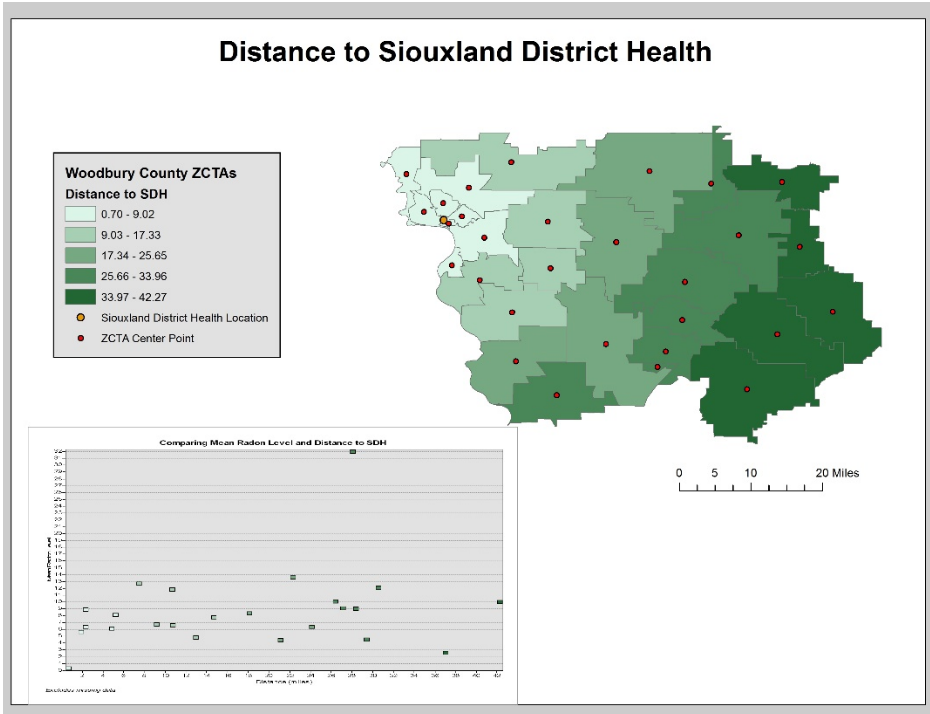
The most significant result from this analysis is that the majority of ZCTAs had mean radon values above the EPA actionable level of 4 cPi/L. While this could be due the convenience sampling method, since homeowners who are concerned about radon or have a history of high radon levels might be more likely to participate, it is still a concerning finding. Many of the participating individuals should be counseled on how best to fix their homes to reduce their radon exposure.

While the overall level of radon in Woodbury County is high, there is still a geography disparity in average radon levels. Some areas are worse than others. There are many ZCTAs where the mean radon level is not far above the actionable level and some targeted action could reduce the levels below the recommended threshold. However, there are also areas where the levels are very high, which may warrant immediate action. The hot spot analysis confirms this by comparing the mean radon level of a ZCTA to the areas around it, flagging ZCTAs that are significantly higher than the other areas. The northeast corner of Woodbury County was flagged as being significantly high using both hot spot methods, and is an area of concern.

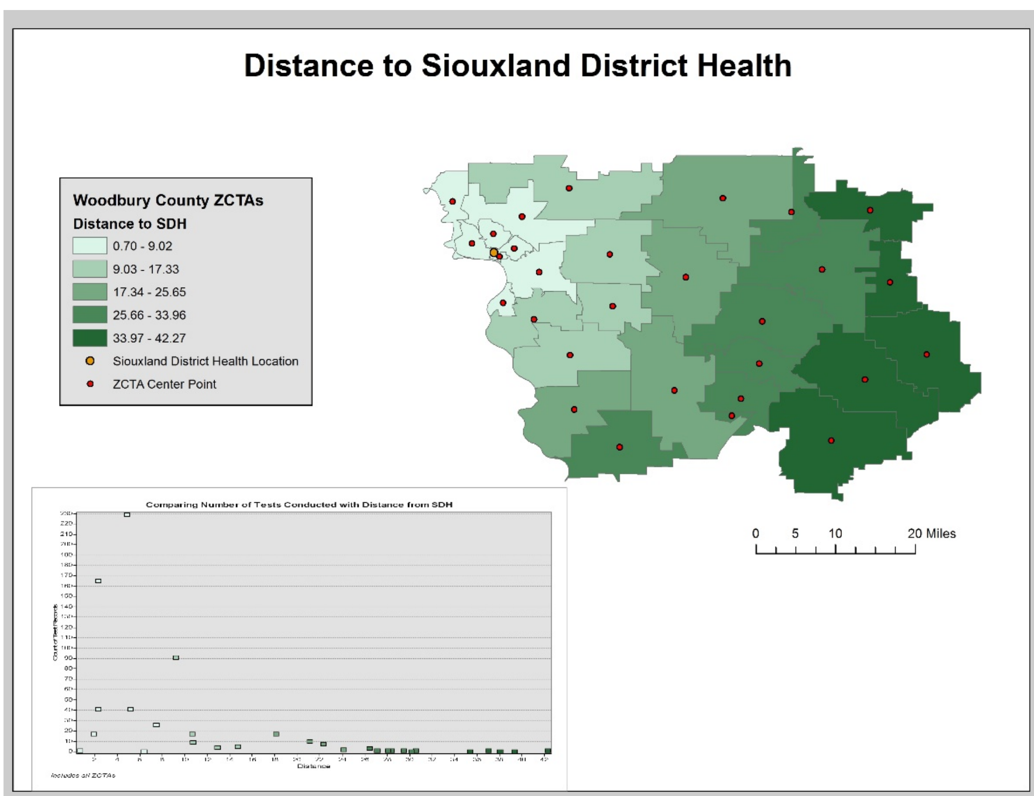
All the models only had one significant variable, the gender ratio for the population, which means that ZCTAs that had more women compared to men had higher mean radon levels. While women are likely not the cause of higher radon levels there could be a confounding variable that is not accounted for in this analysis such as motivation for participating in an opt-in program like the radon test kits. Either way, it was a significant variable that should continue to be examined.

The other variables in the final GWR model were the number of test records per ZCTA and the distance to Siouxland District Health. This indicates that data quality could have been a major influencing factor in this analysis, since both of these variables are related to the sampling and distribution of tests. This is further confirmed by the hot spot analysis of testing rates, showing that there was a clear disparity in testing across the area of interest.

Map 8: Distance to Siouxland District Health



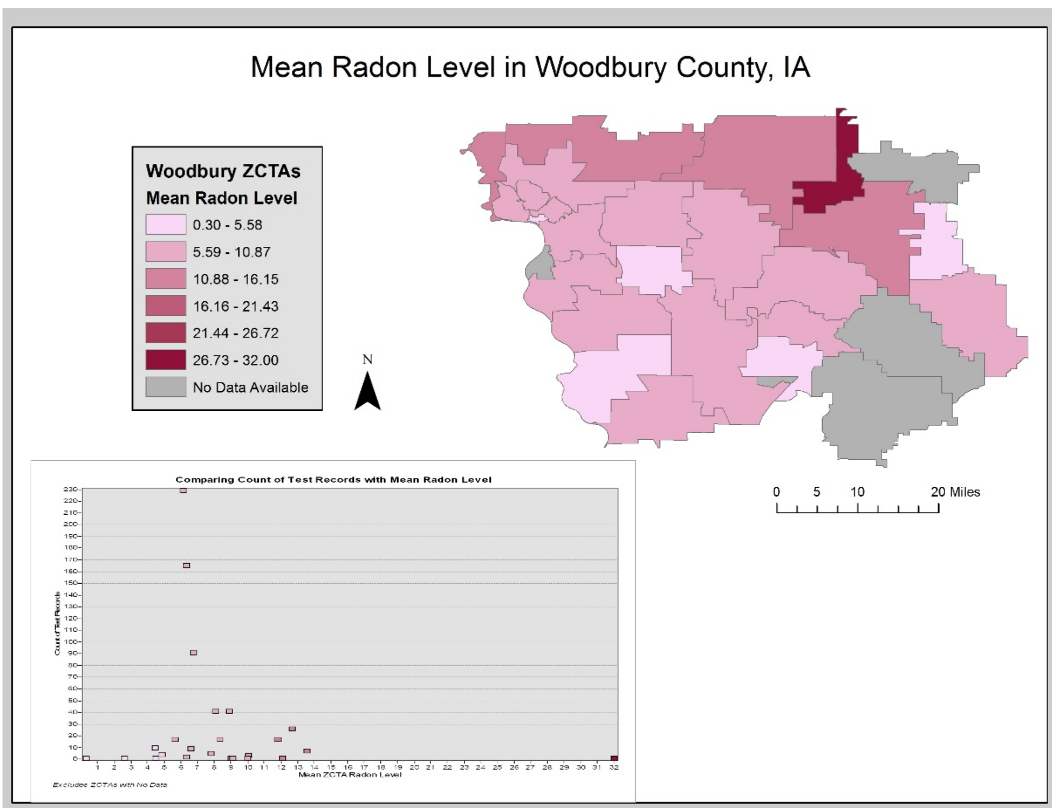
Map 9: Relationship between distance to Siouxland District Health and number of tests



Limitations

The most significant limitation in this analysis is the disparity in sample size between ZCTAs. Some ZCTAs had over two hundred test records while many had zero or only one. Regardless of the population size and testing rate, which does show disparities (Map 6), having so few records makes the ZCTA mean radon level heavily influenced by outliers. That is confirmed by looking at a histogram of mean radon levels and the number of records (Map 10) which shows that those ZCTAs with very high or very low mean radon levels had the fewest records. ZCTAs with many test records had mean radon closer to the center of the range.

<u>Map 10: Comparing mean radon levels to number of radon tests</u>



Low ZCTA sample size and missing data also has implications for the generalizability of these results. Since the aggregated results are so prone to being influenced by outliers, it is unlikely that they are representative of the population. It also has some implications for conclusions drawn between ZCTAs. There is low comparability between the well sampled and the poorly sampled ZCTAs, meaning any results really should be taken with caution.

The nature of the sampling technique also influences the results. This data was collected from a convenience sample of an opt-in program. This sampling method did not have an organized sampling plan to ensure that it sampled a group that is representative of the overall population in Woodbury County. Since the radon testing was an opt-in program, it is safe to assume that participants are significantly different that the general population. This could be due to a variety of factors such as being more concerned about their health, being more concerned about radon, or being more health educated. Since no information was provided about the sample population, we were unable to test if the sample population is representative of the general population, thus the results are only generalizable to the people who opt into this program.

Finally, no addressed where provided so there was no point data for the test records. The highest geographic level reported was zip code. The nearly 700 records had to be aggregated to 29 ZCTAs, which is much resolution lost. This also limits the analysis options and limits the effectiveness of the models since they were run on 29 records instead of 700.

Recommendations

The analysis of this data has given way to many insights into how data quality can influence the distribution of radon level data. Not to focus on the negatives, there are many constructive steps that Siouxland District Health can take to ensure more representative and informative radon data in the future, ensuring that a future collaboration can look at the original research questions.

First, radon test kits need to be made more available to areas of Woodbury County that are far from Siouxland District Health. ZCTAs that were farther from the Siouxland District Health headquarters were more likely to be outliers and have a small number of test records. Since the radon test kits are only available for purchase at the front desk, it makes sense that those people who live far from Siouxland District Health are far less likely to travel to buy one and those who do travel to buy one have very high radon levels that they were presumably concerned about.

This could be done by offering shipping of test kits, partnering with local organizations, and advertising. Another way this could be done, with less effort from Siouxland District Health, is to develop a data use agreement with other organizations that collect similar data. Partnering with other organizations like this would be mutually beneficial since both parties would receive more comprehensive data for little or no cost.

We did not include income or the price of the radon test kit as a factor because it only costs $5 to purchase a kit. However, it is possible that this does select for households with a slightly higher disposable income. This could be addressed by lowering the cost or distributing vouchers for kits to low income households.

We also advise collecting the addresses of the radon test sites. This could allow for even finer detailed analysis, especially if some ZCTAS continue to have small numbers. Including the address means that points of the test locations can be mapped and then the radon levels can be interpolated across the county, providing more meaningful values than the ZCTA aggregated values. Any concerns about protecting the privacy of homeowners can be addressed in analysis. Additionally, it appears that users of the radon test kits need to provide their addresses to get the results returned to them, meaning that adding this resolution to the data will not come at a cost, just ensuring that it is reported to the person completing the analysis. Points do not have to be reported to the public at a fine resolution, but still contribute immensely to the strength of the analysis.

Collecting some basic demographic information about those who participate in the radon testing could add more resolution to the analysis. This would be useful because calculations could be done to see how similar the sample population is to the general population and for stratification by population variables to investigate any disparities between populations. After the initial analysis, this information could be used for surveillance, ensuring that high groups are receiving interventions.

We also advise requesting cancer incidence data several months in advance to completing analysis. After waiting over a month for the Iowa Cancer Registry to answer our data request, we finally had to proceed with a study aim that did not include the registry data to complete the project within the give time frame. We recommend requesting cancer incidence data for the radon test collection period as well as ten years or other period of time prior since cancer takes a long time to develop.

Finally, we recommend continuing to collect the radon test kit data and amassing several years' worth of data for the next analysis. Looking that the test results over time can reveal if the predictors or hot spots have changed over time and attempt to explain why.

Conclusions

Moving forward, changes in data quality should be put into place to ensure that the analysis is not measuring access to Siouxland District Health, but instead has a representative sample of the actual burden of radon in Woodbury County. Improving data quality will make the results more generalizable and representative of the Woodbury County population, therefore yielding more relevant and accurate results.

However, there is still concern about the levels of radon in Woodbury County since many ZCTAs tested above the actionable level and there was a large disparity across the county. Hot spot analysis revealed areas of potential concern. While the model results were inconclusive, the authors recommend further research into identifying characteristics of high risk and target populations in Woodbury County.

References

[1] Ten Leading Causes of Death and Injury. (2015, March 31). Centers for Disease Control and Prevention. Retrieved October 1, 2015. http://www.cdc.gov/injury/wisqars/leadingcauses.html

[2] Common Cancer Types. (2015, January 26). National Cancer Institute Retrieved October 1, 2015. http://www.cancer.gov/types/common-cancers

[3] Lung Cancer Trends. (2015, May 5). Centers for Disease Control and Prevention. Retrieved October 1, 2015. http://www.cdc.gov/cancer/lung/statistics/trends.htm

[4] U.S. Cancer Statistics Working Group. *United States Cancer Statistics: 1999–2012 Incidence and Mortality Web-based Report*. Atlanta: U.S. Department of Health and Human Services, Centers for Disease Control and Prevention and National Cancer Institute; 2015. Available at: https://nccd.cdc.gov/uscs/cancersbystateandregion.aspx

[5] Radon and health. (2014, October 1). Retrieved October 1, 2015, from World Health Organization http://www.who.int/mediacentre/factsheets/fs291/en/

[6] What Is a Safe and Acceptable Level of Radon Gas?" *Radon.com*. Air Check, Inc., 2009. Web. 8 Dec. 2015. http://www.radon.com/radon/radon_levels.html

[7] "Health Risk of Radon." *EPA*. Environmental Protection Agency, 22 Oct. 2015. Web. 08 Dec. 2015. http://www.epa.gov/radon/health-risk-radon#never

[8] "Majority of NW Iowa Homes Tested for Radon Have Dangerous Levels, According to UI Study." *The Loop: UI Healthcare*. University of Iowa Healthcare, 7 Oct. 2015. Web. 1 Dec. 2015. https://medcom.uiowa.edu/theloop/news/majority-of-nw-iowa-homes-tested-for-radon-have-dangerous-levels-according-to-ui-study

[9] Radon Data. (2015). Retrieved October 1, 2015, from Iowa Department of Health. https://www.idph.state.ia.us/radon/Data.aspx

[10] "How ZCTAs Are Created." *Geography: Zip Code Tabulation Areas (ZCTAs)*. United States Census Bureau, 9 Feb. 2015. Web. 4 Oct. 2015. https://www.census.gov/geo/reference/zctas.html

[11] "ZIP Code to ZCTA Crosswalk." *UDS Mapper*. American Academy of Family Physicans, 2015. Web. 08 Dec. 2015. http://www.udsmapper.org/zcta-crosswalk.cfm

[12] Radon Test Kit Data. 5 Oct. 2015. Raw data. Siouxland District Health, Sioux CIty, IA.

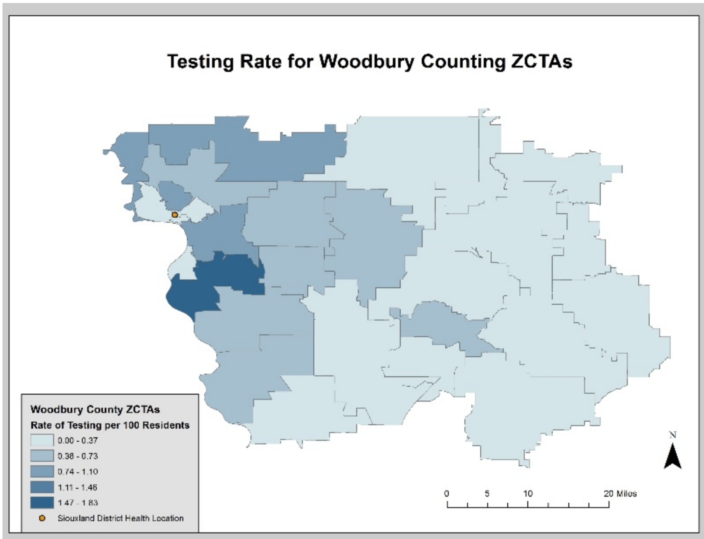
[13] "American Fact Finder." *United States Census Bureau*. US Department of Commerce, n.d. Web. 31 Oct. 2015. <http://factfinder.census.gov/faces/nav/jsf/pages/index.xhtml>.

[14] ESRI 2014. ArcGIS Desktop: Release 10.3. Redlands, CA: Environmental Systems Research Institute.

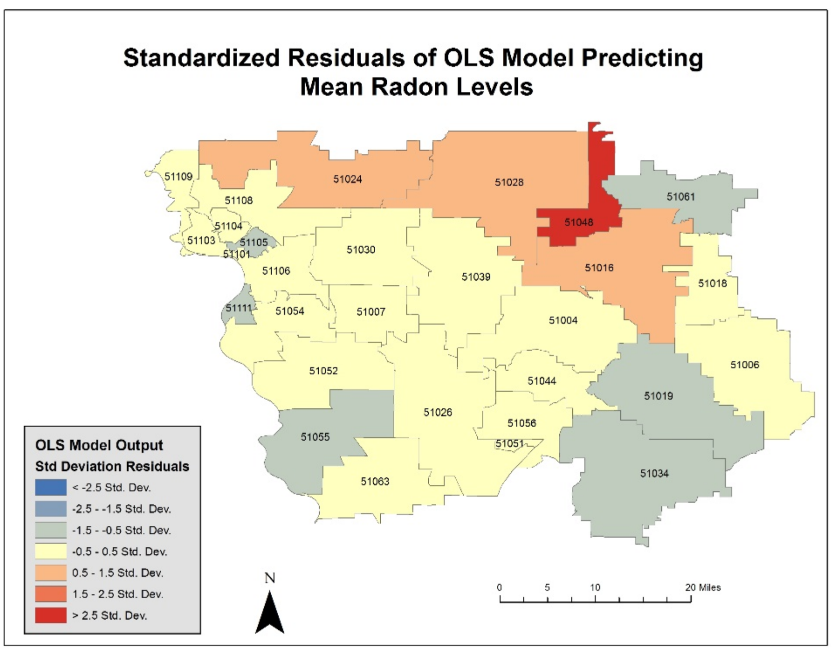
[15] Anselin, Luc, Ibnu Syabri and Youngihn Kho (2006). GeoDa: An Introduction to Spatial Data Analysis. Geographical Analysis 38 (1), 5-22.

Supplemental Figures

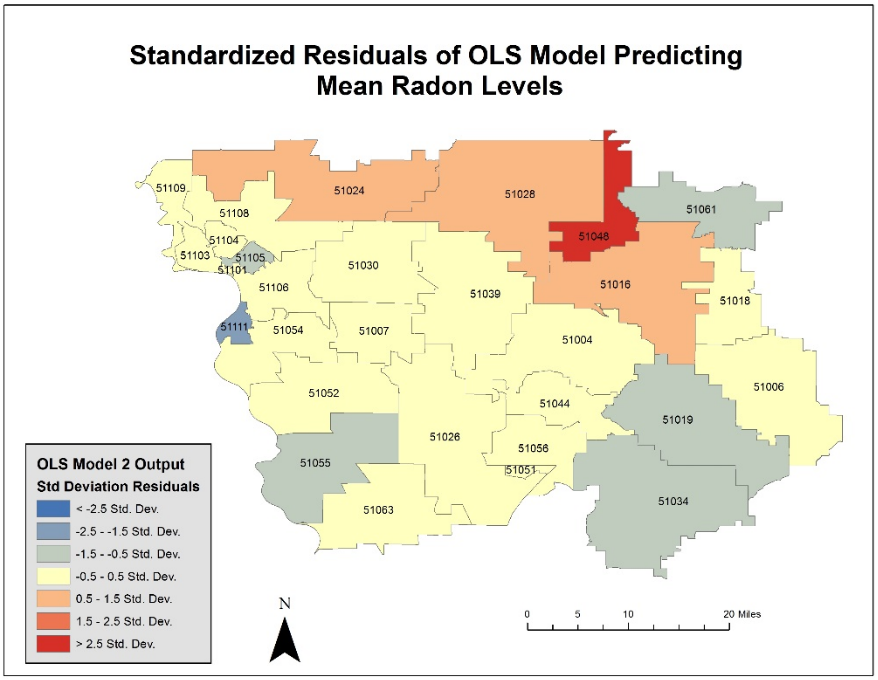
<u>Supplemental Figure 1: Radon testing rates in Woodbury County</u>



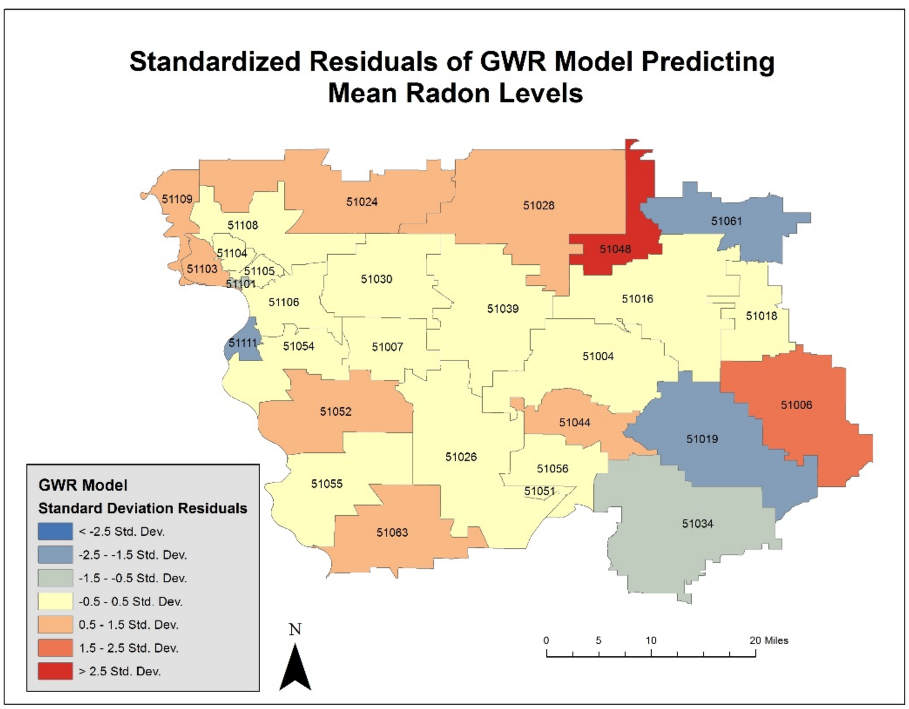
<u>Supplemental Figure 2: Standardized Residuals for Model 1</u>



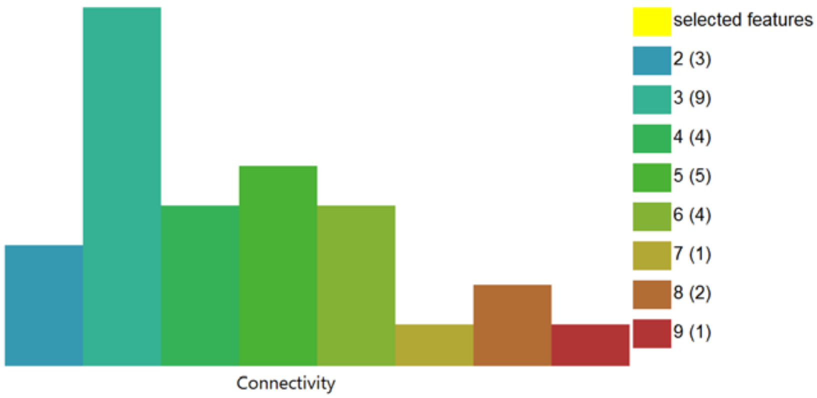
<u>Supplemental Figure 3: Standardized Residuals for Model 2</u>



<u>Supplemental Figure 4: Standardized Residuals for Model 3</u>



Supplemental Figure 5: OpenGeoDa modelling output and weights histogram



Weighted on Queen Contiguity

Regression

SUMMARY OF OUTPUT: SPATIAL LAG MODEL - MAXIMUM LIKELIHOOD ESTIMATION

Data set : Woodbury_ZCTA_all

Spatial Weight : Queen_weights.gal

Dependent Variable : MEAN_2 Number of Observations: 29

Mean dependent var : 7.17464 Number of Variables : 7

S.D. dependent var : 6.1958 Degrees of Freedom : 22

Lag coeff. (Rho) : 0.1301

R-squared : 0.173401 Log likelihood : -91.3358

Sq. Correlation : - Akaike info criterion : 196.672

Sigma-square : 31.7315 Schwarz criterion : 206.243

S.E of regression : 5.63307

Variable	Coefficient	Std.Error	z-value	Probability
W_MEAN_2	0.1301	0.2509049	0.5185233	0.6040932
CONSTANT	22.73	14.67	1.54942	0.1212807
COUNT_OF_R	-0.01231455	0.02444747	-0.5037146	0.6144619

SEXRATIO -0.130792 0.05690414 -2.298462 0.0215354

MEDIANAGE 0.05714516 0.2435677 0.2346172 0.8145060

PERCWHITE -4.829743 15.71575 -0.3073187 0.7586009

NEAR_DIST -1.861512e-005 6.908819e-005 -0.26944 0.7875912

REGRESSION DIAGNOSTICS

DIAGNOSTICS FOR HETEROSKEDASTICITY

RANDOM COEFFICIENTS

TEST	DF	VALUE	PROB
Breusch-Pagan test	5	10.17405	0.0704511

DIAGNOSTICS FOR SPATIAL DEPENDENCE

SPATIAL LAG DEPENDENCE FOR WEIGHT MATRIX : Queen_weights.gal

TEST	DF	VALUE	PROB
Likelihood Ratio Test	1	0.2139006	0.6437272

======================== END OF REPORT =============================

Regression

SUMMARY OF OUTPUT: ORDINARY LEAST SQUARES ESTIMATION

Data set : Woodbury_ZCTA_all

Dependent Variable : MEAN_2 Number of Observations: 29

Mean dependent var : 7.17464 Number of Variables : 6

S.D. dependent var : 6.1958 Degrees of Freedom : 23

R-squared : 0.164082 F-statistic : 0.902929

Adjusted R-squared : -0.017640 Prob(F-statistic) : 0.496019

Sum squared residual: 930.587 Log likelihood : -91.4428

Sigma-square : 40.4603 Akaike info criterion : 194.886

S.E. of regression : 6.36084 Schwarz criterion : 203.089

Sigma-square ML : 32.0892

S.E of regression ML: 5.66473

Variable	Coefficient	Std.Error	t-Statistic	Probability
CONSTANT	23.56959	16.52893	1.425959	0.1673125
COUNT_OF_R	-0.01376345	0.02760555	-0.4985756	0.6228129
SEXRATIO	-0.1339192	0.06424152	-2.08462	0.0483977
MEDIANAGE	0.05829995	0.2746061	0.2123039	0.8337399
PERCWHITE	-4.273982	17.66698	-0.2419192	0.8109874
NEAR_DIST	-2.152783e-005	7.797548e-005	-0.2760846	0.7849478

REGRESSION DIAGNOSTICS

MULTICOLLINEARITY CONDITION NUMBER 41.643024

TEST ON NORMALITY OF ERRORS

TEST	DF	VALUE	PROB
Jarque-Bera	2	58.48345	0.0000000

DIAGNOSTICS FOR HETEROSKEDASTICITY

RANDOM COEFFICIENTS

TEST	DF	VALUE	PROB
Breusch-Pagan test	5	9.783018	0.0816220
Koenker-Bassett test	5	2.265553	0.8113126

SPECIFICATION ROBUST TEST

TEST	DF	VALUE	PROB
White	20	14.12394	0.8241568

DIAGNOSTICS FOR SPATIAL DEPENDENCE

FOR WEIGHT MATRIX : Queen_weights.gal

(row-standardized weights)

TEST MI/DF VALUE PROB

Moran's I (error) 0.016456 N/A N/A

Lagrange Multiplier (lag) 1 0.1831796 0.6686548

Robust LM (lag) 1 3.5440778 0.0597584

Lagrange Multiplier (error) 1 0.0162170 0.8986665

Robust LM (error) 1 3.3771152 0.0661077

Lagrange Multiplier (SARMA) 2 3.5602948 0.1686133

======================== END OF REPORT =============================

Regression

SUMMARY OF OUTPUT: SPATIAL ERROR MODEL - MAXIMUM LIKELIHOOD ESTIMATION

Data set : Woodbury_ZCTA_all

Spatial Weight : Queen_weights.gal

Dependent Variable : MEAN_2 Number of Observations: 29

Mean dependent var : 7.174640 Number of Variables : 6

S.D. dependent var : 6.195803 Degrees of Freedom : 23

Lag coeff. (Lambda) : 0.056827

R-squared : 0.165470 R-squared (BUSE) : -

Sq. Correlation : - Log likelihood : -91.429075

Sigma-square : 32.0359 Akaike info criterion : 194.858

S.E of regression : 5.66003 Schwarz criterion : 203.062

Variable	Coefficient	Std.Error	z-value	Probability
CONSTANT	23.215	14.74665	1.574256	0.1154283
COUNT_OF_R	-0.01179937	0.02418976	-0.4877836	0.6257032
SEXRATIO	-0.1305571	0.056852	-2.296438	0.0216508
MEDIANAGE	0.04914624	0.2415401	0.2034703	0.8387675
PERCWHITE	-3.970772	15.96084	-0.2487821	0.8035295
NEAR_DIST	-1.977744e-005	7.10708e-005	-0.278278	0.7807991

LAMBDA 0.05682737 0.2617184 0.2171317 0.8281057

REGRESSION DIAGNOSTICS

DIAGNOSTICS FOR HETEROSKEDASTICITY

RANDOM COEFFICIENTS

TEST	DF	VALUE	PROB
Breusch-Pagan test	5	9.919437	0.0775504

DIAGNOSTICS FOR SPATIAL DEPENDENCE

SPATIAL ERROR DEPENDENCE FOR WEIGHT MATRIX : Queen_weights.gal

TEST	DF	VALUE	PROB
Likelihood Ratio Test	1	0.02736405	0.8686128

======================= END OF REPORT ============================