Balk

A Geographic Analysis of the Impact of a New Professional Sports Stadium on Residential Real Estate Values in Minneapolis

Anthony M. LaVerde

A thesis submitted to the faculty of Salem State University in partial fulfillment of the requirements for the Master of Science degree in Geo-information Science in the Department of Geography.

May, 2018

Table of Contents

Abstract 1

1.0 Introduction and Background 2

1.1 Overview 2

1.2 Historical Background of Stadium Construction 3

1.3 Modern Stadium Construction History 4

2.0 Critical Assessments of Economic Benefits of Stadiums 7

2.1 Empirical Studies of the Economic Benefits of Stadiums 7

2.2 Empirical Studies of the Economic Benefit of Large Events 9

2.3 Empirical Studies of the Impact of Stadiums on Real Estate Values 10

3.0 Thesis Overview 18

3.1 The Minnesota Twins and Negotiations for a New Stadium 18

3.2 Research Significance 20

4.0 Methodology and Data 22

4.1 Data Description 22

4.2 Introduction to Geographically Weighted Regression (GWR) Analysis 31

4.3 Ordinary Least Squares (OLS) Model and Specifications 34

4.4 Geographically Weighted Regression (GWR) Model and Specifications 38

5.0 Results 40

5.1 Geographically Weighted Regression (GWR) Results – 2016 Real Estate Sales Data 40

5.2 2016 GWR Analysis Coefficient Values and Variable Maps 43

6.0 Analysis of Minneapolis Real Estate Sales Data Prior to Approval of Target Field 55

6.1 Data Description – 2006 Data 55

6.2 2006 OLS Model Specifications and Results 58

6.3 2006 GWR Model Specifications and Results 60

7.0 Comparison of 2016 and 2006 Results 65

7.1 Interpreting Value from Coefficient Results 65

7.2 Overall Real Estate Value Change from 2006 - 2016 66

7.3 Local Patterns in Real Estate Value Change from 2006 – 2016 70

8.0 Conclusion 73

References 75

Maps, Tables, & Appendix

Figure 1 Target Field Location in Minneapolis **22**

Figure 2 Minneapolis U.S. Census Block Groups with centroid points **28**

Figure 3 Thiessen Polygons of 2016 Real Estate Sales Data **30**

Figure 4 2016 GWR Standardized Residual Map **41**

Figure 5 2016 Local R- Squared Values **42**

Figure 6 Distance to Water/ Recreation Coefficient Map **44**

Figure 7 Distance to Light Rail Station Coefficient Map **46**

Figure 8 Percentage of Block Group Population that is Under 18 Years of Age Coefficient Map **48**

Figure 9 Average Living Area of Dwellings in Block Group Coefficient Map **49**

Figure 10 Median Household Income in Block Group Coefficient Map **51**

Figure 11 Average Land Value of Parcels in Block Group Coefficient Map **52**

Figure 12 2016 Distance to Target Field Coefficient Map **54**

Figure 13 2006 GWR Standardized Residual Map **61**

Figure 14 2006 Local R-Squared Values **62**

Figure 15 2006 Distance to Target Field Coefficient Map **64**

Figure 16 2016 Distance to Target Field Coefficient Dollar Value Map **67**

Figure 17 2006 Distance to Target Field Coefficient Dollar Value Map **69**

Figure 18 GWR Distance to Target Field Variable Difference 2006-2016 **71**

Table 1 2016 Residential Real Estate Sales Statistics **25**

Table 2 2016 Residential Real Estate Sales by U.S. Census Block Group **27**

Table 3 2016 OLS Report **34**

Table 4 2016 OLS Coefficient Report **36**

Table 5 2016 GWR Analysis Report **38**

Table 6 2006 Residential Real Estate Sales Statistics **55**

Table 7 2006 Residential Real Estate Sales by U.S. Census Block Group **56**

Table 8 2006 OLS Report **57**

Table 9 2006 OLS Coefficient Report **59**

Table 10 Impact of Target Field Variable on GWR Model in USD **72**

Appendix: Python Script of the Analysis Conducted for Thesis **80**

Abstract

The City of Minneapolis, Minnesota and Hennepin County, Minnesota are among the many government entities that have committed taxpayer funds to finance a professional sports stadium. Both the city and county approved the financing of Target Field in 2007. The new stadium is now the home of the Minnesota Twins, a Major League Baseball franchise. Previous empirical studies have offered differing opinions on whether a new professional sports stadium has an effect on residential real estate values in the surrounding area. This thesis uses a geographically weighted regression (GWR) model to analyze the effect that Target Field had on residential real estate values in the City of Minneapolis, Minnesota. Using GWR analysis of a hedonic real estate price model, this thesis concludes that being close to Target Field had a significantly positive effect on residential real estate values in Minneapolis in 2016. However, when applying the same model to real estate sales data from a year prior to the approval of Target Field, properties in the vicinity of the location of the ballpark may have been more valuable before the ballpark was built then they were in the years that followed its opening.

1.0 Introduction and Background

1.1 Overview

Over the past 25 years, there have been a significant number of new professional sports stadiums constructed in metropolitan areas across North America. Most of the ownership groups of the professional teams that play in these stadiums have secured public funds to finance all or part of the cost of developing them, including the cost of improvements to infrastructure in the vicinity of the stadium. While many academic studies point out that these publically financed stadiums produce a far lower return on investment than promised by professional sports league officials and team owners (Baade, Baumann, and Matheson 2008), there has been some question as to whether a new professional sports stadium will significantly affect residential real estate values in an area. This thesis uses a geographically weighted regression (GWR) model to analyze changes in residential real estate values in the city of Minneapolis, Minnesota between 2006 and 2016, and seeks to determine if there is, or has been, any significant effect on residential real estate values in the city before or after the approval and construction of Target Field in Downtown Minneapolis. Target Field is the home of Major League Baseball's Minnesota Twins franchise. The financing for the stadium was approved in 2007 and the stadium opened in 2010.

1.2 Historical Background of Stadium Construction

The idea of government entities using taxpayer funds to finance sports stadiums does have historical precedent. In fact, the Roman government built stadiums and arenas such as The Colosseum out of public funds to be used by citizens for entertainment purposes. It should be pointed out that the building of these stadiums were evidence of the economic prosperity of the Roman society and were not designed as a means to achieve economic growth or rehabilitation of an area. The rehabilitation of a blighted or undeveloped area of the city is often a main talking point used by stadium advocates while trying to procure public financing for stadium projects in the modern world (Baade and Matheson 2011).

The first modern movement in the United States to build stadiums with partial or complete public funding occurred between 1917 and 1926 (Coates, Humphreys 2008). There were 70 professional sports stadiums in both major and minor leagues built during that time period. There were only 11 professional sports stadiums constructed before 1917 in the U.S. (Coates 2007). The total construction costs of all 70 stadiums built between 1917 and 1926 was about $295 million in 2007 US Dollars. That is approximately $11 million more than the average cost to construct one of the Major League Baseball Stadiums that have been opened since 1990 (Long 2005).

The way professional sports stadiums are used, and the rationale for public subsidies has also changed over time. Early news reports on stadium construction depicted the professional sports stadium as a community gathering place to enjoy many different kinds of spectator activities and not just one sport. This is in stark contrast to how most stadiums are designed today. Early stadium advocates envisioned a public place where the community would gather for events that would serve the broad interests of the public. These could include parades, festivals and concerts, as well as sporting events in multiple sports including track and field, baseball, and football (Coates, Humphreys 2008).

1.3 Modern Stadium Construction History

Most sports stadiums today are designed to be used exclusively for one sport, sometimes with a means to reconfigure the floor to accommodate other income producing events such as concerts. The modern sports arena, even though many of them require public funds to construct them, are usually controlled by the team that operates the arena. The team gets to determine what events are held in the stadium on non-game days, and the team collects all of the revenue from those events (Coates, Humphreys 2008).

As a result of franchise relocation and league expansion, combined with many of the older neighborhood stadiums that are sometimes referred to as “jewel box” stadiums because they were baseball “diamonds” within a city block that looked like a box from above, in urban centers becoming outdated, a new era of stadium construction began in the 1960’s and 1970’s. Many of these stadiums were built to house more than one team, playing different sports. Since the majority of people attending sporting events in the 1960’s and 1970’s travelled to the stadium by automobile (which was not the case in the 1920’s), these new stadiums were built primarily in areas of the city or in a nearby suburban community where a large parking lot could also be built surrounding it. Stadiums constructed in this manner include Veterans’ Stadium in Philadelphia, Qualcomm Stadium in San Diego, The Astrodome in Houston, Atlanta-Fulton County Stadium in Atlanta, Busch Stadium in St. Louis, Riverfront Stadium in Cincinnati and Three Rivers Stadium in Pittsburgh. These stadiums came to be known as “cookie cutter” stadiums because they all looked alike and were all round when observed from above or on a map. While these “cookie cutter” stadiums accomplished the goal of having one facility in an urban area that could play host to a professional football and a professional baseball club, spectator sight lines were negatively affected. To accommodate both sports, the spectator areas had to be designed further back from the playing surface than at the old “jewel box” stadiums.

Franchise owners began to seek opportunities to build new stadiums as a way to revive fan interest and generate greater revenue. Both game attendance and overall interest in the league declined across Major League Baseball in the 1990s. Fans were less interested in part due to the lack of team identity caused by free agency and roster turnover on their favorite teams. Fans were also disillusioned by the high player salaries, and drew a correlation between those salaries and the cost of a middle class family to go to a game. Fans were also tired of team owners, who would threaten to relocate to another city if they were not granted public funding for a new stadium. Many fans also did not enjoy watching baseball in the giant multi-use "cookie-cutter" stadiums that all looked the same. The sight lines were poor for baseball, and fans felt like they were too far away from the action. This fan disinterest and apathy came to a head during the 1994 MLB Strike, when 58% of fans said that they were "disgusted" with Major League Baseball (Dortch, 1996).

Dortch's research did find however that fans in cities with great teams or new ballparks were much happier with their team and league. Dortch found that "It's clear that baseball fans were quicker to forgive a team that had a new ballpark. This hasn't escaped the notice of major league owners" (Dortch, 1996).

As a response to fan disenchantment with the MLB in-stadium experience, clubs around the league began a trend of designing and building ballparks that had the look and feel of stadiums from the "jewel box" era, but contained all the modern amenities that are required to attract the modern fan, and would also help the club attract players through free agency. This "retro-modern" ballpark movement started with the construction of Baltimore's Camden Yards in 1992. This design model continued with the construction of Jacobs Field in Cleveland (1994), The Ballpark at Arlington in the Dallas Ft. Worth Area (1994), and Coors Field in Denver (1995), to name a few. In fact, between 1991 and 2012, 26 Major League Baseball franchises built or significantly renovated stadiums at a cost of approximately $9.4 Billion. Approximately $5.5 Billion of those dollars, or 59% of the total cost of construction was financed by the taxpayers in those communities. Five new MLB stadiums were 100% financed by government entities. Public subsidies for professional stadiums go well beyond baseball. Between 1990 and 2011, 28 of 32 NFL franchises built or renovated their stadiums and took public money to do so. Across all five major sports in North America (NFL, MLB, NBA, NHL, MLS), 125 of 140 clubs played their 2012 season in a stadium built or significantly renovated since 1990 (Baade and Matheson 2011).

2.0 Critical Assessments of Economic Benefits of Stadiums

2.1 Empirical Studies of the Economic Benefits of Stadiums

There is a large volume of empirical research devoted to the issues associated with the cost and potential return on investment of professional sports stadiums. Most of these studies were conducted by economists and focused on topics such as employment (Coates 2007), sales tax revenue (Matheson 2004, Matheson 2006), and overall value to a community in terms of tangible and intangible factors (Johnson et al 2007). In recent years, there have also been several empirical works that tackle the question of how a new professional sports stadium will affect property values in the community in which it will reside.

Proponents of publically financed stadiums have argued that a new sports stadium can have an extremely positive effect on the economy. They typically use *ex ante* projections of how much revenue the new stadium will generate, either through tourism or suburban residents flocking to the area around the new ballpark (Baade and Matheson 2011). They usually use some kind of Input/output analysis with a large multiplier effect (Baade, Baumann, and Matheson 2005). The vast majority of empirical research on public subsidies for professional sports stadiums have revealed that the value to the community almost never comes close to justifying the often exorbitant cost of financing a professional sports stadium (Baade, Baumann, and Matheson 2008).

The type of analysis used by stadium project proponents to justify public subsidies usually neglect three major economic factors. When these factors are taken into account, it becomes doubtful that a new stadium's impact on the local economy is as powerful as stadium proponents would argue that it is (Baade and Matheson 2011).

The first is the factor that Baade and Matheson refer to as the “substitution effect”. The substitution effect is an economic phenomenon that occurs when people spend their disposable income in one place over another. In regards to sports stadiums, if people in the area are spending their money to go to the game at the new stadium, those dollars are not “new” to the economy. In reality, they are disposable dollars that would otherwise have been spent on entertainment in another part of the economy. Suppose a couple on a Saturday night attends the baseball game at the new stadium and eats dinner at a restaurant in the vicinity of that new stadium. They are spending dollars there instead of eating somewhere else and going to a movie at the local theater (Baade and Matheson 2011).

The second major economic factor is referred to as “crowding out”. New stadiums, and especially large events that often accompany new stadiums, affect other types of economic activity in the area. An example of this would be a large event like the Olympics. People who would normally visit the city for other reasons will generally stay away from the city during a large event. On a smaller scale, people who live and work in Salem, Massachusetts see this every year during Halloween. All other activity in Salem stops on October 31st because of the congestion and activity around the Halloween festivities. This is magnified in a large event in a major city (Baade and Matheson 2011).

The third and final economic factor that *ex ante* studies typically do not take into account is that money spent in local economies during games or large events may not stay in the local economy. For example, the athletes that earn high salaries playing professional sports typically do not live in the communities that surround the stadium. This means that that large sum of money most likely will not be recirculated back into the economy (Baade and Matheson 2011).

2.2 Empirical Studies of the Economic Benefit of Large Events

A major incentive that sports team owners and leagues offer to communities to secure funding for stadiums is the promise of bringing large events associated with that sport to the new stadium. Examples of these types of special events are the Super Bowl or the Major League Baseball All-Star Game. The NFL projects that the Super Bowl would have an economic impact to the host city of around $400 million. Several sports organizations such as Major League Baseball and the International Olympic Committee make similar claims of huge dollars flowing into prospective host communities through large events. (Baade, Baumann, and Matheson 2005).

Taxable sales statistics provide a comprehensive way to analyze the impact of an event on the entire economy of an area (Baade, Baumann, and Matheson 2005). An analysis of taxable sales data in Florida over a 25 year period found that on average, large events have been associated with the reduction of taxable sales in the host regions by roughly $34.4 million per event (Baade, Baumann, and Matheson 2005). These taxable sales studies almost universally find that a new professional sports stadium’s contribution to the overall economy of an area is far too insignificant to justify the large subsidies that are required to construct these facilities (Baade and Matheson 2011).

2.3 Empirical Studies of the Impact of Stadiums on Real Estate Values

There have been several recent empirical studies relating to the effects of a new professional sports stadium on real estate values in the surrounding area. These studies differ in several ways, however they all employ a hedonic model. These models apply the same basic approach, where sale price or value is a function of structural, neighborhood (location) and market characteristics. Market characteristics include distance to perceivably desirable amenities such as a professional sports arenas (Tu 2005). According to Tu (2005), this can be expressed in a hedonic equation as:

Price = function of (Structural, Location, Market) or

$$P = f(S,L,M)$$

Tu (2005) conducted an analysis of single family home sales in Prince George’s County Maryland in 2005 to determine if there was any change in property value coinciding with the opening of Fed Ex Field, the current home of the Washington Redskins in Landover, Maryland In 1997. The study found that while homes near the Fed Ex Field site sold at a lower price than homes a distance away from the site, the sale price gap was larger before the completion of the stadium project than it was after the completion of the stadium, indicating a positive effect on property values in the vicinity of the new stadium (Tu 2005).

A study of residential real estate values was conducted in the Dallas – Ft. Worth Area in 2007 to determine if the announcement of a new NFL stadium for the Dallas Cowboys in the Fair Park section of downtown Dallas, the subsequent cancellation of that project, and the eventual announcement of the new stadium location in Arlington, Texas affected residential home sale prices in the area. The study found that average property values declined in Arlington, Texas by approximately 1.5%. This is almost equal to the anticipated household sales tax burden for Arlington residents that was levied in the city to fund the stadium project. This suggests that the amenity effect for Arlington residents to host the new NFL stadium was not statistically different from zero (Dehring et al, 2007).

A land value analysis was also conducted in Berlin, Germany focusing on the areas within 5000 meters of three separate multifunctional sports arenas in lower value neighborhoods in a district of the former East Berlin. The results of the study indicate significant positive impacts within an area of approximately 3000 meters around one of the stadiums (Ahlfeldt and Maennig 2007).

Feng and Humphreys (2012) also used a hedonic price model to analyze real estate values in census block groups in the vicinity of every MLB, NFL, NBA, and NHL stadium in the U.S. Their paper concluded that median home values are higher in block groups that are closer to these facilities then ones that are further away. They surmised that the positive value to the community that finances professional sports stadiums, and possible justification for the expenditure of public funds, may be found in residential real estate values (Feng and Humphreys 2012).

Ahlfeldt and Kavetsos (2014) used a similar analysis method to Tu (2005) in their study of real estate values around two recently constructed stadiums in Central London. These stadiums were the New Wembley Stadium, home of the English National Football Team, and Emirates Stadium, home of the Arsenal Football Club of the English Premier League. This study found a significant increase to residential property values in areas close to the New Wembley Stadium. Properties close to the stadium were valued up to 15% higher than properties that were located away from the stadium. They observed a gradual decrease in values the further the property was away from the stadium. Properties in close vicinity to Emirates Stadium also saw an increase in value. The study found that there was a 1.7% increase in property value for every 10% decrease in distance to Emirates Stadium (Ahlfeldt and Kavetsos 2014).

Lastly, Feng and Humphreys (2018) studied the area around two new professional sports facilities in Columbus Ohio. One was Nationwide Arena, home of the NHL's Columbus Blue Jackets, and the other was Crew Stadium home of the MLS Columbus Crew (Feng and Humphreys 2018). The study concluded that there is a significant positive distance-decaying effect on surrounding home values as a result of the two stadiums being located in that city.

The data used in these studies vary in location, type and sample size. All of these studies used some variation of neighborhood demographic data such as annual income (Tu 2005, Ahlfeldt and Kavetsos 2014), school data (Dehring et al 2007), age statistics (Ahlfeldt and Maennig 2007), or racial/nationality statistics (Tu 2005; Ahlfeldt and Maennig 2007; Feng and Humphreys 2018). Most used sales price of single family homes as a measure of price. As per the mission of all of these studies, the distance from the sold property or statistical block to the stadium was one of the hedonic variables being studied.

Tu (2005) used data from approximately 35,000 single family home real estate transactions in Prince George's County Maryland., a suburban county adjacent to Washington D.C. in the analysis. All of the real estate sales used in this analysis occurred between October 1992 and December 2001. A three mile "Stadium Impact Area" was established around the stadium location.

Dehring et al. (2007) added a temporal component to the data, analyzing 42,351 single family home transactions in Dallas County and 32,061 single family home transactions in Arlington, Texas occurring between January 2004 and March 2005. These real estate transactions occurred in the immediate aftermath of five separate announcements of two prospective sites for the eventually built AT&T Stadium, home of the NFL's Dallas Cowboys. Originally, there was an announcement of a proposed stadium site in the City of Dallas (to replace the existing Cotton Bowl which has stood at that location since 1930). This was followed by an announcement that the project was being abandoned.

That announcement was then followed by three separate announcements of negotiations, city council, and finally voter approval of the eventual site of the new stadium in Arlington, Texas.

Ahlfeldt and Maennig (2007) used land values instead of individual real estate sales as a measure of price. The study examined the areas around three new stadiums on two sites that were constructed as part of Berlin's unsuccessful bid to host the 2000 Summer Olympics. These stadium locations were both situated in the former East Berlin. To accomplish this, they used the assessed values of all 15,937 statistical blocks in 2005. For the structural component, they used the average floor space size for a dwelling in that block. This differs from the works of Tu (2005), Dehring et al (2007), and Feng and Humphreys (2018), who all used individual single family home sales, and with them certain structural amenities such as number of bedrooms, number of bathrooms, or the age of the dwelling.

Feng and Humphreys (2012) used U.S. Census block group data from both the 1990 and 2000 decennial census. Block groups that were within five miles of all MLB, NFL, NBA, and NHL stadiums in the U.S. were used in the analysis.

Ahlfeldt and Kavetsos (2014) used a data set of property transactions from one of the three largest mortgages companies in London for their analysis of real estate values around New Wembley Stadium, and Emirates Stadium. The data set comprised approximately 10% of all properties in London, since this particular company holds the mortgages for that number of properties in the city.

Feng and Humphreys (2018) analyzed a data set of 9,504 single family home transactions in the in Columbus, Ohio. In 2000, after the opening of Nationwide Arena, (opened in 2000), and Crew Stadium (Now Mapfre Stadium, opened in 1999). Nationwide Arena is located in Downtown Columbus, and Crew Stadium is located four miles to the north in more of a suburban area of the city.

The methodology of these studies were all similar in that they all used hedonic models to analyze the impact of professional sports stadiums on real estate values. Where these studies differ is in their approach to adjusting for spatial autocorrelation in the data (Feng and Humphreys 2018).

Spatial autocorrelation occurs when data observations (in this case housing values) are similar to other data points located near them due to geographic proximity (Tobler 1970). It is important to correct for spatial autocorrelation in a hedonic real estate model, since failing to do so can lead to misspecification in the model, and lead to biased estimates (Anselin 1988).

Tu (2005) added a spatial autoregressive term to the hedonic regression that accounted for all sales within 1.8 miles of each individual home sale. Dehring et al. (2007) added a Generalized Least Squares term to account for spatial autocorrelation. This term was based on the elementary school that the children of residents in the neighborhood of the sale would likely attend. Ahlfeldt and Maennig (2007) used an Inverse Distance Weighting interpolator of the 3 closest neighbors to each block to determine the spatial autocorrelation term in their regression analysis .Ahlfeldt and Maennig’s (2007) data was in the form of polygon blocks, as opposed to sale price point data that was used in the other three real estate studies discussed in this thesis. Feng and Humphreys (2012), and Ahlfeldt and Kavetsos (2014) used a similar method to Tu (2005), establishing areas of influence around the stadiums of five miles, and five kilometers respectively (Feng and Humphreys 2012), (Ahlfeldt and Kavetsos 2014). Finally, Feng and Humphreys (2018) used an advanced spatial lag model, which developed a spatial weight matrix based on Queen’s Weight neighbors calculation, known as a spatial two – stage least squares robust estimator (S – 2SLS Robust) (Feng and Humphreys 2018). In order to perform this analysis, they created Thiessen Polygons based on the values of the point data.

The results of these studies varied as to the effect of a new sports stadium on residential real estate values. However, except for Feng and Humphreys (2012) and Ahlfeldt and Kavetsos (2014), who found possible net positive gains in real estate values for cities that build new professional stadiums, all of these agree that any marginal increase in area real estate value is not sufficient enough to justify the massive public subsidies paid to construct these facilities (Tu 2005; Dehring et al. 2007; Ahlfeldt and Maennig 2007; Feng and Humphreys 2018).

Tu's (2005) results indicated that single family houses inside the established three mile "stadium impact area" sold for 5.07% less than houses outside the impact area after the stadium development. However, using a "quasi- experimental" comparison method know as a "difference in difference" method (Ahlfeld and Kavetsos 2014), (Feng and Humphreys 2018), Tu (2005) determined that there was an increase in comparative property values since a house inside the impact area before the stadium development sold for 13.44% less than a single family house outside the impact area. Tu (2005) determined that because of this, the Fed Ex Field development accounted for an increase of $42 million in local property values. As pointed out by Dehring et al (2007) however, the value of that area may have already been depressed. Tu (2005) does not mention that the site of Fed Ex Field was adjacent to the site of the Capital Center, which lost their professional team tenants in 1997 and was not demolished until 2002. The closing of the former home of the NBA's Washington Wizards and NHL's Washington Capitals, and the sight of the abandoned arena sitting vacant on that site may have already depressed property values in the area (Dehring et al. 2007).

Dehring et al. (2007) concluded that property values dropped in Arlington, Texas as a result of the Cowboys stadium development, which is in contrast to the Tu (2005) study that concluded that real estate values increased in the area around the stadium development in Landover, Maryland. The Dehring et al. (2007) study is also the only one that measured the effect of a stadium project that included a local sales tax increase. Arlington, Texas residents were levied a 1% sales tax as a result of the stadium development, which in and of itself could lower property values. The results also showed that single family home sale prices dropped 1.5% in Arlington, Texas after the approval of the new Cowboys stadium. This is almost the same amount as the projected revenue that would be generated from the sales tax increase.

Similar to the 3 mile “stadium impact area” used by Tu (2005), Ahlfeldt and Maennig (2007) established 1000 meter concentric distance rings around the stadium locations. These “impact zones,” to borrow a phrase from Tu (2005), compared blocks that were within 0-1000m, 1000m-2000m, 2000m-3000m, 3000m-4000m, and 4000m-5000m from each stadium location. The results of the analysis show that were was no statistical difference in property values for blocks that were over 2000m from either location. There was a 3.5% increase in property values within 1000m – 2000m of both stadium sites. There was a 7.5% increase in property values in the vicinity of one of the sites, the area around the swimming and track and field venues known as the Velodrom. The results of the analysis also found that there was no significant statistical impact within 1000m of the other stadium site, known as Max Schmeling Arena. Ahlfeldt and Maennig (2007) attribute this to two factors. First, the parking situation for residents in the area around Max Schmeling Stadium is poor, so the presence of the stadium may make potential buyers leery that during events the parking situation will become intolerable. They state that there is much more parking around the Velodrom. Also, Max Schmeling Arena is used primarily for basketball and handball, and is home to two “nationally important” clubs. This may also affect buyer’s perception of the area around the stadium because they may not want to deal with club supporters in the neighborhood on game days. As a swimming and track and field venue, Velodrom hosts primarily individual sports that often times do not include the rowdy supporter incidents that team sporting events often do.

Feng and Humphreys (2012) set a five mile boundary around each stadium to analyze the impacts that these stadiums had on local real estate. They concluded that the mean increase in residential real estate values due to the presence of these stadiums was $11.2 million (USD) within one mile of the stadium, and that number grows to $277 million in a four mile radius of the stadium. They contend that since the average cost of building one of these stadiums in 2000-2001 was $339 million, the increase in property values, and therefore tax revenues, comes close to justifying the cost in larger cities (Feng and Humphreys 2012).

Ahlfeldt and Kavetsos (2014) also used the same “Difference in Difference” approached used by Tu (2005) in their analysis of the area around the New Wembley, and Emirates Stadium. Also similar to Tu (2005), Ahlfeldt and Kavetsos (2014) set up zones of influence around the two stadiums of 5km. They found that there was a £1.91 Billion increase in residential real estate values around New Wembley Stadium. They also found that there was a £1.04 billion increase in residential real estate values within 5km of Emirates stadium. They also found that there was a £1.41 billion loss due to the closing of the old Arsenal Stadium (Ahlfeldt and Kavetsos 2014).

Feng and Humphreys (2018) were also able to place a monetary value on the increase in residential property value for a new professional sports stadium (Feng and Humphreys 2008). The results of Feng and Humphreys (2018) analysis indicated that average house value in Columbus increased by $2,214 (USD) due to the presence of Nationwide Arena, with an overall dollar impact of $222,500,000 (USD) to single family home values in the city, with a significant distance decaying effect on sale prices the further away the home is from the stadium. The results also show that there is a much smaller effect on sale prices around Crew Stadium. The researchers state that this is probably due to the fact that Nationwide Arena is located downtown, and Crew Stadium is not. It was also observed that distance to Ohio State University’s college football stadium (a dummy variable to account for the presence of the massive campus of Ohio State University) had a larger effect on real estate prices than Nationwide Arena or Crew Stadium.

3.0 Thesis Overview

The subject of this thesis is an analysis of residential real estate values in Minneapolis, Minnesota. The data set used in the analysis is stored in a Geographic Information System (GIS) database that contains approximately 50,000 residential sales occurring between 2006 and 2016. The goal is to determine if the construction of Target Field, the home of Major League Baseball's Minnesota Twins Baseball Club, which was approved and finalized in 2007, had any effect on residential real estate values in the city.

3.1 The Minnesota Twins and Negotiations for a New Stadium

Major League Baseball's Minnesota Twins have called the "Land of 10,000 Lakes" home since their 1961 move from Washington D.C. where they were known as the Senators. The Twins' first Minnesota ballpark was in suburban Bloomington, where they played their home games for 21 seasons. In 1982, The Twins began play in the City of Minneapolis at the Hubert H. Humphrey Metrodome. In 1994, owner Carl Pohlad (who had purchased the team from previous owner Calvin Griffith in 1984), declared the Metrodome "economically obsolete" due to the changing landscape of Major League Baseball (Berg 2010). He argued that the Metrodome lacked modern amenities and luxury suites that are massive revenue streams for MLB owners. The interior of the building also lacked the aesthetic charm that attracted fans to newer style ballparks such as Camden Yards in Baltimore and Jacobs Field in Cleveland. One feature of the ballpark that became a running punchline among fans and the media was the presence of a giant "baggie" in right field that was used so that fly balls would not get lodged in the football seats that needed to be stored when baseball was being played.

Years of negotiations, potential relocations, and rumors of contraction by Major League Baseball ensued. These negotiations included two failed referendums in the Twins Cities to build a ball park on a different site in St. Paul (MPR 1999). Twins ownership actually had an agreement to sell the club to a group that intended to move it to North Carolina. The sale fell apart because the voters in that state rejected a referendum to finance a new stadium (NY Times 1997). Soon after the potential sale to the North Carolina group fell through, Major League Baseball planned to contract the Twins, which was inevitably squashed in 2001 (St. Paul Pioneer Press 2014).

In 2006, the Minnesota Legislature agreed to finance 75% of the new ballpark (approximately $350 million), through a 0.15% sales tax in Hennepin County (HF 2480, 84th MN Legislature). A dispute with land owners over the value of land being taken by eminent domain threatened to cancel the project. However in April 2007, an agreement was reached with the land owners, and the county took control of the land required to start the construction (St. Paul Pioneer Press 2007). The new stadium was named Target Field, due to a naming rights contract with the Minnesota-based retail chain Target. Construction crews broke ground for the new stadium in 2007, and it was completed for the opening of the 2010 Minnesota Twins season on April 12th.

3.2 Research Significance

This thesis builds on the growing list of research investigating the effect of new professional stadiums on residential property values. The geographically weighted regression (GWR) method used in this analysis of Minneapolis real estate has not been employed by any of the previous studies referenced in this thesis.

The subject of this thesis differs from previous research on real estate values surrounding newly constructed professional stadiums in several ways. Unlike previous studies, this thesis analyzes the area around one Major League Baseball stadium in a major U.S. city. MLB stadiums are guaranteed to be active for 81 days per season, due to the scheduling of games. The studies of Fed Ex Field (Tu 2005) and AT&T Stadium (Dehring et al 2007) were conducted in areas surrounding NFL Stadiums. NFL stadiums are only active for game days on 10 days per year. Ahlfeldt and Maennig's (2007) analysis was conducted in areas surrounding three separate venues that could host multiple events. It could be inferred that potential residents in those areas would assume that those venues would be operational all year (Ahlfeldt and Maennig 2007). Feng and Humphreys' (2012) study compared real estate values in census block groups within 5 miles of every NFL, MLB, NBA, and NHL arena in the U.S., to adjacent census block groups outside of the 5 mile area, however they did not build a model to analyze any one city's local variation of property values (Feng and Humphreys 2012). Ahlfeldt and Kavetsos (2014) did look at the local market in London around New Wembley Stadium, and Emirates Stadium, but did not use GWR in the analysis. English Football is also similar to the NFL in regards to how often the stadium is used. For example, English Premier League clubs (Which Arsenal is one of) play 19 home matches per season. Nationwide Arena's main tenant would only be scheduled for 44 home dates. This obviously does not include concerts or other events that could be held at the stadiums (Feng and Humphreys 2018).

The data set used in this research includes all residential housing units such as single family homes, condominium, and town houses. The Landover, Maryland and Columbus, Ohio studies only used single family home sales (Tu 2005, Feng and Humphreys 2018), and it is unclear if all housing unit sales data was utilized for the Dallas – Arlington, Texas (Dehring et al 2007) study. The Berlin study used assessed value of parcels of land in a statistical unit similar to a U.S. Census Block or Block Group (Ahlfeldt and Maennig 2007). Feng and Humphreys (2012) used census block group data on home ownership and value. Ahlfeldt and Kavetsos (2014) used a sample of home values in London based on data supplied by a mortgage company.

One of the issues that Feng and Humphreys (2018) encountered was the difficulty in determining what the effect of Nationwide Arena being located in the downtown area of Columbus, Ohio had on the results of the analysis. The location in London could also explain some of the high real estate values in the vicinity of Wembley and Emirates Stadiums Ahlfeldt and Kavetsos (2014). Similar to the difference in difference approach used by Tu (2005), this thesis compares results from before and after the construction of Target Field, to attempt to determine if the actual stadium project had an effect on real estate values, or if any effect that the data may show is simply a function of the stadium being located in a downtown area.

4.0 Methodology and Data

4.1 Data Description

The residential real estate transactions database used in this thesis contains information and locations of approximately 50,000 property sales in the City of Minneapolis. The analysis can be performed for any year between 2006 and 2016 by utilizing a database query to access the data for a particular calendar year. Analysis for this thesis was conducted using the newest (2016), as well as the oldest (2006) residential sales data available. These two years were chosen since the 2016 data was the most recent data available at the time that this thesis was began, and could be used to find the most current values of residential real estate in the city. The 2006 data was chosen since it is from the year

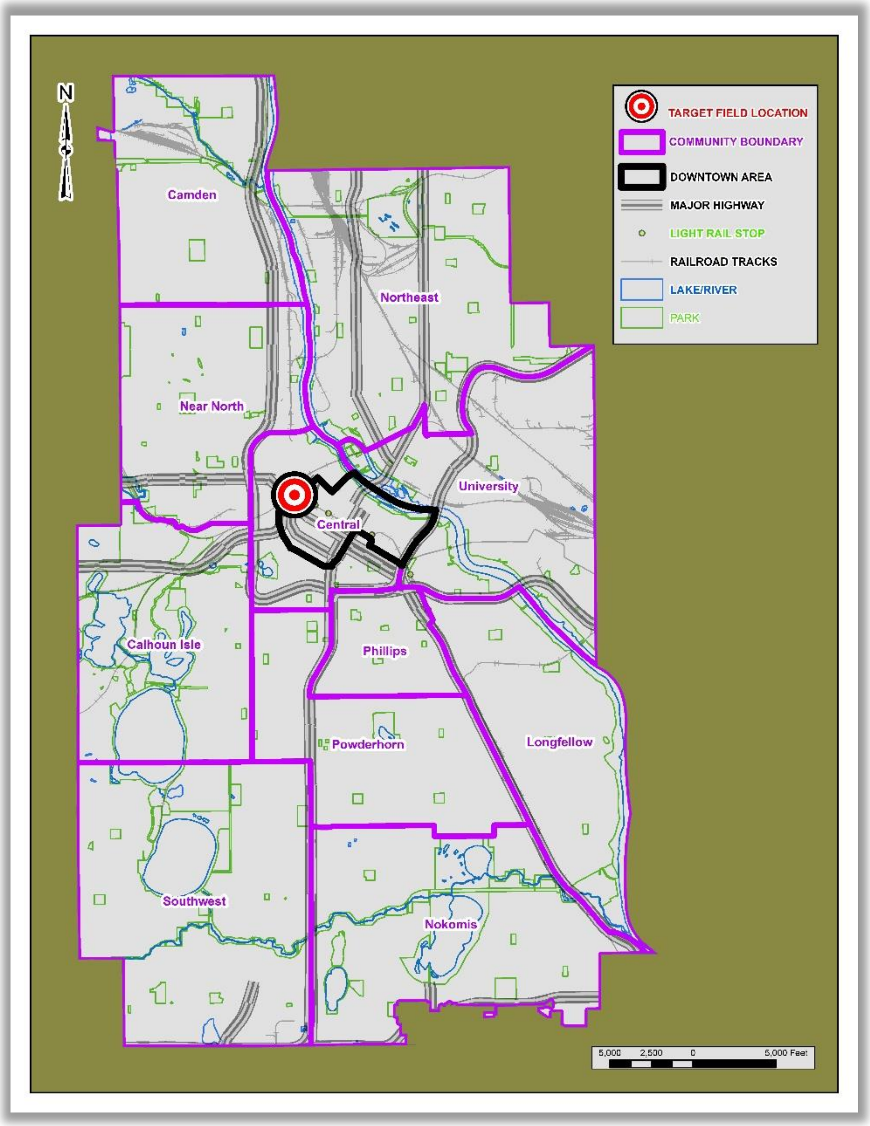


Figure 1- Target Field Location in Minneapolis

before Target Field was finally approved. Home buyers in 2006 were unsure if a new ballpark was going to be built at that location, or if the Minnesota Twins were going to relocate.

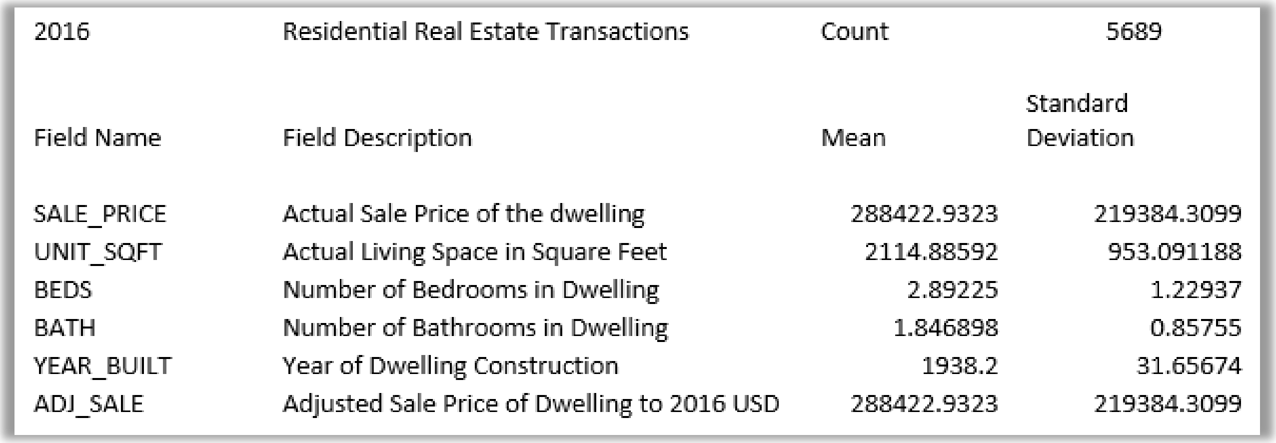
Target Field is located on the edge of the Downtown West Neighborhood, in the adjacent North Loop Neighborhood. As Feng and Humphreys (2018) observed in their analysis of residential real estate values in Columbus, Ohio, any empirical model designed to analyze the effect that Target Field has on property values in Minneapolis will most likely be biased by the stadium's location within Downtown Minneapolis (*Figure 1*). Feng and Humphreys (2018) accounted for this problem of downtown location bias by performing analysis on two separate stadiums, one being in Downtown Columbus, and the other being located approximately 4 miles north of Downtown Columbus. This thesis accounts for bias caused by the Downtown Minneapolis location of Target Field by using the same analysis model on a recent data set of residential real estate sales in the city (2016), and repeating the process using historic data from before the stadium was approved (2006).

The Minneapolis real estate data used in this thesis is available free of charge through the City Assessor's office' website (Minneapolis City Assessor). The data includes information on the location of the property (including street address and neighborhood), the type of housing unit (single family home, condo, or townhouse) sale information (date of sale, and sale price), and building characteristics (number of bedrooms, number of bathrooms, number of stories, square feet of living area, year of construction). The City Assessor's data was then merged with spatial data from the Open Data Minneapolis website (Open Data Minneapolis) to acquire the location of all the residential sales in the city. The resulting data points were then stored in an ESRI file geodatabase.

Only sales from residential dwellings were used in the analysis. All commercial units or industrial buildings were removed from the data set. Also excluded were any residential transactions that were under 50% or over 200% of the assessed value of the property. The low value floor was put in place to remove non-market transactions. It is common practice for families to transfer ownership of properties to an LLC or trust that is controlled by themselves or their family. These sales are obviously not occurring in an open real estate market environment. However, they are still recorded as sales in municipal assessors' offices and at deed registries. These kinds of transactions cause bias in the data and should be removed. The reason for the high value cap is to guard against the sale of large parcels that may have recently been rezoned, and not yet updated in the assessor's database. In cities and towns, there are many instances where a developer buys a parcel of land for a price far above the current assessed value. This could be due to the seller recently secured a zoning change or a variance with the municipality that allows them to sell the parcel to a developer at an increased price, since as a result of that zoning change, the buyer may now be able to build a multi home subdivision on the property. In future years, those individual homes' sale prices would be included in a model such as this one. However, this kind of sale to a developer that has not been updated in the assessor records at the time of the analysis can skew the data. Vacant residential land sales were also excluded from the dataset. It is common for a large parcel to be rezoned and sold to a developer with an existing home already on it. Many times these homes are uninhabited and dilapidated and therefore would bring the assessed value of the property down. To guard against these type of sales biasing the data set, any home that sold for more than 200% assessed value was excluded from the data set. Both the upper value cap and low value floor also guard against data entry mistakes in the city assessor database causing outliers in the data set. A total of 5689 residential real estate transactions were used in the 2016 analysis. The average sale price of a dwelling in Minneapolis in 2016 was $288,422.93. (*Table 1*).

Since the data set consists of some residential sales such as single family homes that are located in a place where they are the only residence sold at that location, and other residential sales such as condominiums and townhouses can possibly share the same exact location with many other residential sales in the same building, aggregating the data around some kind of larger geographic area would make sense to perform this kind of analysis. The previous research on real estate values surrounding new professional stadiums did not have to account for multiple data points at the same location since they used exclusively single-family home sales. Buildings with condominium, town house, and multi – family dwellings can have upwards of several dozen recorded real estate sales at the same geographic location in a calendar year. This can cause problems when using Geographically Weighted Regression (GWR) by incorporating multicollinearity in the results (Fotheringham, Brunsdon, and Charlton 2002). Multicollinearity occurs when two or more variables are highly correlated to each other, and could be explaining the same phenomena, which can cause severe problems when designing regression models (ESRI). In GWR analysis, a bandwidth is calculated to determine the optimum number of neighbors that will have influence on a data observation's coefficient results. In this case, multicollinearity would be caused by the many new construction condominium buildings in downtown locations across the U.S.

Table 1- 2016 Residential Real Estate Sales Statistics

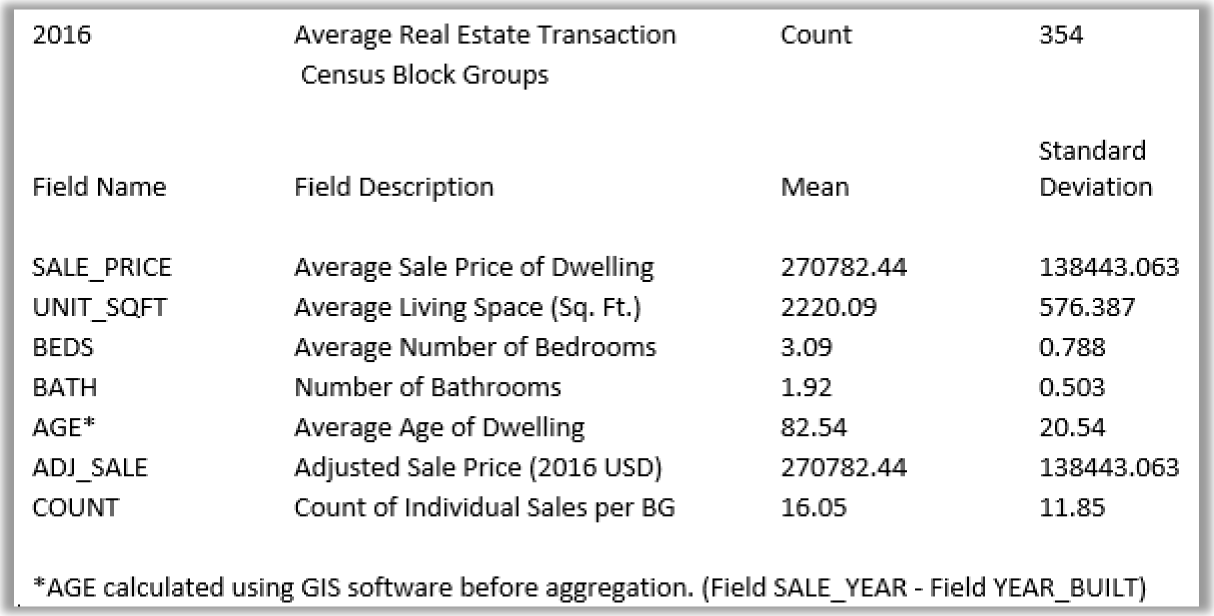


2016	Residential Real Estate Transactions	Count	5689
Field Name	Field Description	Mean	Standard Deviation
SALE_PRICE	Actual Sale Price of the dwelling	288422.9323	219384.3099
UNIT_SQFT	Actual Living Space in Square Feet	2114.88592	953.091188
BEDS	Number of Bedrooms in Dwelling	2.89225	1.22937
BATH	Number of Bathrooms in Dwelling	1.846898	0.85755
YEAR_BUILT	Year of Dwelling Construction	1938.2	31.65674
ADJ_SALE	Adjusted Sale Price of Dwelling to 2016 USD	288422.9323	219384.3099

These building often contain several hundred units in the same building would have the same location coordinates due to the common building address in the assessor database, and could also have been sold in the same year at the same price. This would cause severe model design problems when using GWR (ESRI).

Most of the demographic data that could possibly be part of the final hedonic price model is in the form of U.S. Census Block Groups, so it would make sense to aggregate the sale price data into averages based on those areas. GIS data containing 2010 U.S. Census Block Group polygons was obtained from the Minnesota Geospatial Commons (MN Geospatial Commons) website. Also obtained from this site was a polygon layer of all assessor's land parcels in Hennepin County. To obtain the average overall land value of a parcel in each census block group (a potential hedonic variable in the final model), the Hennepin County parcel data was joined to the census block group data using a spatial join. Next, the individual sale points were then joined spatially to the census block polygons and the results produced an average of all fields from the original point data, and an average land value of a parcel in that block group. The average block group consisted of approximately 16 residential real estate sales in 2016, with an average sale price of $270,782.44 (*Table 2*).

Table 2- 2016 Residential Real Estate Sales by U.S. Census Block Group



2016	Average Real Estate Transaction Census Block Groups	Count	354
Field Name	Field Description	Mean	Standard Deviation
SALE_PRICE	Average Sale Price of Dwelling	270782.44	138443.063
UNIT_SQFT	Average Living Space (Sq. Ft.)	2220.09	576.387
BEDS	Average Number of Bedrooms	3.09	0.788
BATH	Number of Bathrooms	1.92	0.503
AGE*	Average Age of Dwelling	82.54	20.54
ADJ_SALE	Adjusted Sale Price (2016 USD)	270782.44	138443.063
COUNT	Count of Individual Sales per BG	16.05	11.85

*AGE calculated using GIS software before aggregation. (Field SALE_YEAR - Field YEAR_BUILT)

After the points were aggregated into census block groups, it became apparent that not all block groups within Minneapolis were represented in the dataset. In fact 24 of the 378 census block groups that are entirely within Minneapolis City Limits did not record a residential real estate sale in 2016. In order to maintain a contiguous data field for the analysis, and avoid zero value gaps in the GWR output maps, centroid points were calculated for all block groups that recorded at least 2 residential real estate sales. A GIS layer was created in the database of these centroid points. The block group centroid points contained all of the data that was included in the block group polygons, namely the averages of residential real estate sale prices, average living area of dwellings sold, and average land values of parcels in each block group (*Figure 2*).

Demographic data was obtained free of charge from the U.S. Census Bureau via the American Fact Finder website (https://factfinder.census.gov/) and added to the GIS database. This data includes tables of 2010 Decennial Census population statistics, as well as 2013 American Community Survey Five-

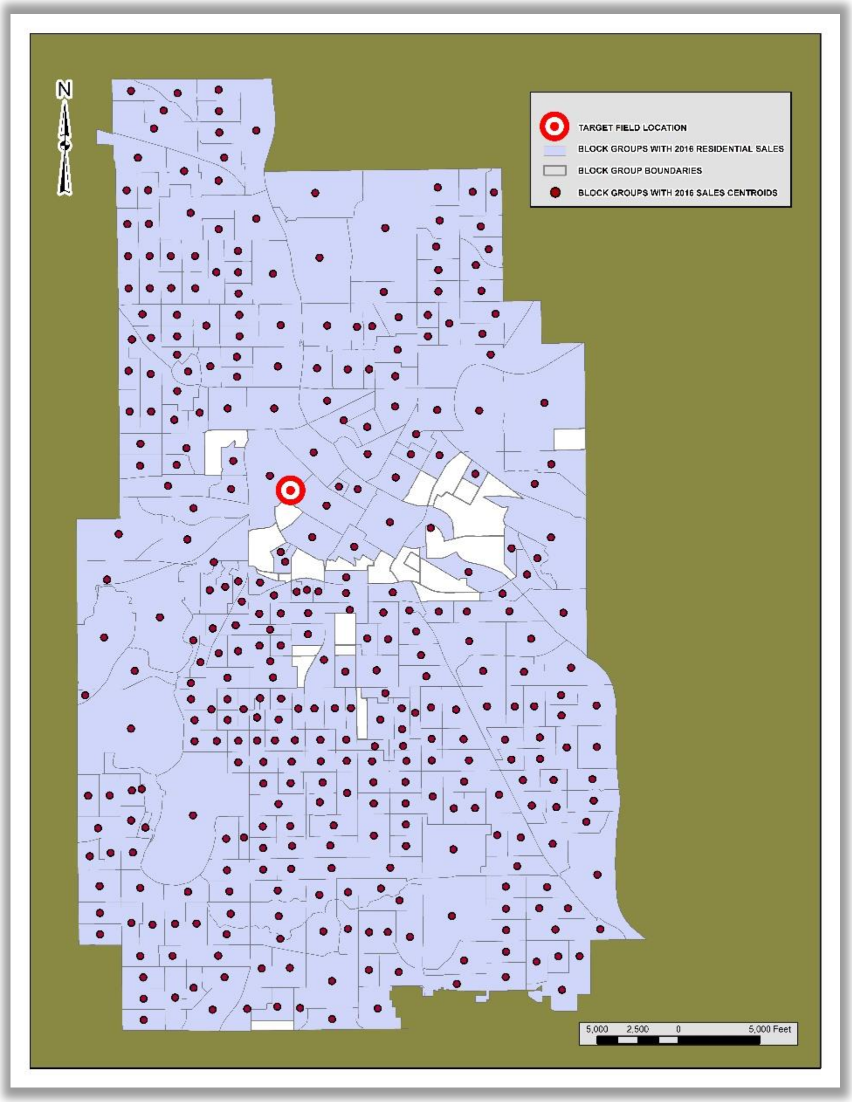


Figure 2- Minneapolis U.S. Census Block Groups with centroid points

Year Estimate income statistics for all census block groups in Hennepin County, Minnesota. These data tables were joined to the block group centroid point data in the GIS.

Calculations were made using GIS tools, to find the distance of the census block group centroid points to several features. These features are local amenities and recreation areas that may affect housing prices in the city, which include distance to the nearest park or recreation area with some kind of water resource (such as a lake or river), distance to the nearest light rail station, and most importantly for this thesis, distance to Target Field.

To conduct their analysis of single family home sales in Columbus, Ohio, Feng and Humphreys (2008) used GIS tools to construct a polygon coverage of their study area known as a Thiessen network. Thiessen network polygons define areas of influence around each point. The value of the Thiessen polygon area is the same as the point value, and the boundary of a Thiessen polygon is half of the distance to its nearest neighboring point in the data set (ESRI).

Since this thesis is using geographically weighted regression (GWR) to analyze data aggregated into census block group centroid points, building a Thiessen polygon network that would allow the coefficient results to be mapped as polygons and displayed as a contiguous data field would aid in the visualization of the results. The analysis of this thesis used Thiessen polygons that were generated from the block group centroid points. Note that when performing GWR using most GIS software, the analyst has the option of producing raster coefficient maps that are interpolated from the point or polygon data observation points. These raster data layers can be quite large and take some time to generate. As a part of this thesis, both methods were tested and analyzed thoroughly. Results showed that the use of Thiessen polygons in the GWR regression improved the model in a small but yet statistically significant manner. In the final results section, the coefficient and residual maps were generated using the Thiessen polygon method. The year to year comparison map was generated using raster mapping based on the Thiessen polygon data. (*Figure 3*).

Figure 3- Thiessen Polygons of 2016 Real Estate Sales Data

4.2 Introduction to Geographically Weighted Regression (GWR) Analysis

Ordinary least squares (OLS) regression is a simple method of linear regression. In an OLS model, the object is to "fit" the data in the model as close to the regression line as possible. In the OLS equation, the dependent variable (the variable whose value in the model is dependent on the independent or explanatory variables), is usually represented in a mathematical equation by ***y***. The independent variables (the variables that explain or predict the dependent variable) are usually represented by ***x*** (Charlton and Fotheringham 2009). A simple OLS linear regression model with one independent *(x*) variable can be described by this equation:

$$y_i = \beta_0 + \beta_1 x_i + \dot{\varepsilon}_i$$

In the above equation, y_i is the dependent variable, x_i is the independent, or explanatory variable, β_0 and β_1 are the coefficients, and $\dot{\varepsilon}_i$ is the error term or residual (Charlton and Fotheringham 2009). This equation can be expanded to account for models with more than one variable. The equation below is an ordinary least squares equation with three independent variables:

$$y_i = \beta_0 + \beta_1 x_{1i} + \beta_2 x_{2i} + \beta_3 x_{3i} + \dot{\varepsilon}_i$$

OLS regression models are powerful explanatory tools for data analysis, however they have limitations when used to model spatial data. OLS regression models assume that the data observations are distributed normally. In the case of a normal, or randomly distributed data set that does not have a geographic location, the residuals would have a normal distribution with a mean of zero (Charlton and Fotheringham 2009). Spatial data (data with a geographic location) is by definition not random, it is spatially correlated. According to Tobler's First Law of Geography "Everything is related to everything else, but near things are more related than distant things" (Tobler 1970). In this vein, an OLS model which looks at the total data set regardless of the location of each observation may not the best way to model spatial data, since the results of the analysis will contain spatial bias (Fotheringham, Brunsdon, and Charlton 2002).

Geographically weighted regression (GWR) models are built to analyze the variations of the interaction of independent variables with a dependent variable in different geographic areas within an overall study area (Fotheringham, Brunson, and Charlton 2002). GWR models account for spatial auto correlation in the data set by running a regression for each data observation that takes into account the value of that observations' nearest geographic neighbors (Mennis 2006). GWR models are based on ordinary least squares (OLS) regression models, but are expanded to include the location of each data observation. As a result, GWR models calculate coefficient results for each observation in the data set (Fotheringham, Brunson, and Charlton 2002).

Since geographically weighted regression (GWR) is a method of analyzing local variation of spatial data in a statistical model, the results of a GWR analysis can be mapped using a geographic information system (GIS) to show the variation in the explanatory variable across a geographic area. This is in contrast to OLS regression, which can be referred to as a global model, since it models the entire data set without taking into account local variation (Fotheringham, Brunsdon, and Charlton 2002). GWR independent variable coefficients are generated by a weighting formula that takes into account the coefficient value of the neighbors to each observation. These spatial weights are generated from a weighting scheme referred to as a kernel (Charlton and Fotheringham 2009).

A typical Geographically Weighted Regression equation would resemble this:

$$y_i = \beta_0(u_{i,}\, v_i) + \sum_{j=1}^{k} \beta_j(u_i v_i) x_{ij} + \dot{\varepsilon}_i$$

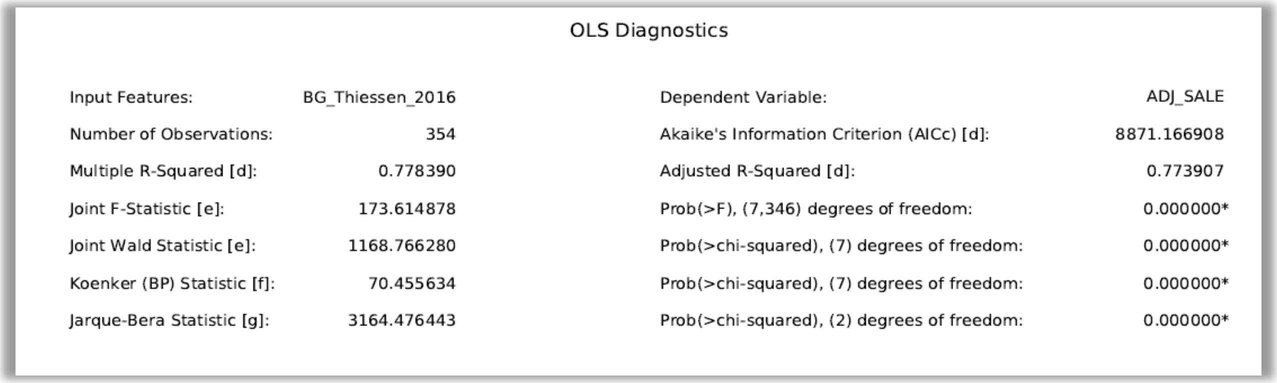
In this equation, ($u_{i,}$ vi) are the coordinate locations of β_0 and $\beta_{j.}$ (Matthews and Yang 2012). By using this formula, GWR models account for spatial correlation by producing separate coefficient estimates for the location of each feature in the data set. Features that are located closer to i are weighed higher in the local coefficient estimate for $\beta_j(u_i v_i)$ than points that are a greater distance away (Matthews and Yang 2012).

4.3 Ordinary Least Squares (OLS) Model and Specifications

An analysis using a geographically weighted regression (GWR) Model starts with a properly specified ordinary least squares (OLS) model (Fotheringham, Brunsdon, and Charlton 2002). As discussed earlier in this thesis, the sale price or value of residential real estate is a function of structural, location, and market characteristics in a hedonic real estate price model. These characteristics are all potential variables in OLS and GWR regression models. As a result of OLS model testing on several dozen potential hedonic variables in an attempt achieve the best possible fit for the data, a final OLS model with seven independent variables was created.

It should be noted that this same method can be used to analyze the effect of hedonic variables on real estate values in any area where value or sales data is available, and is not limited to Minneapolis, or these specific variables. The model can be replicated in a different location to include any number of variables that meet the hedonic model standard that sale price (or property value) is a function of structural, location, and market characteristics. This is a broad spectrum of variables that could be used to analyze any number of factors that could influence real estate values. Any replication of this model in another area could include a professional sports stadium of any kind, or any large facility in general as one of the variables. A python script of this analysis is available in *appendix A*. It can be edited by future

Table 3- 2016 OLS Report



OLS Diagnostics

Input Features:	BG_Thiessen_2016	Dependent Variable:	ADJ_SALE
Number of Observations:	354	Akaike's Information Criterion (AICc) [d]:	8871.166908
Multiple R-Squared [d]:	0.778390	Adjusted R-Squared [d]:	0.773907
Joint F-Statistic [e]:	173.614878	Prob(>F), (7,346) degrees of freedom:	0.000000*
Joint Wald Statistic [e]:	1168.766280	Prob(>chi-squared), (7) degrees of freedom:	0.000000*
Koenker (BP) Statistic [f]:	70.455634	Prob(>chi-squared), (7) degrees of freedom:	0.000000*
Jarque-Bera Statistic [g]:	3164.476443	Prob(>chi-squared), (2) degrees of freedom:	0.000000*

researchers to accommodate different variables that would be used in their analysis, simply by changing the names of the files and tables in the script to correspond with the researcher's data and study area.

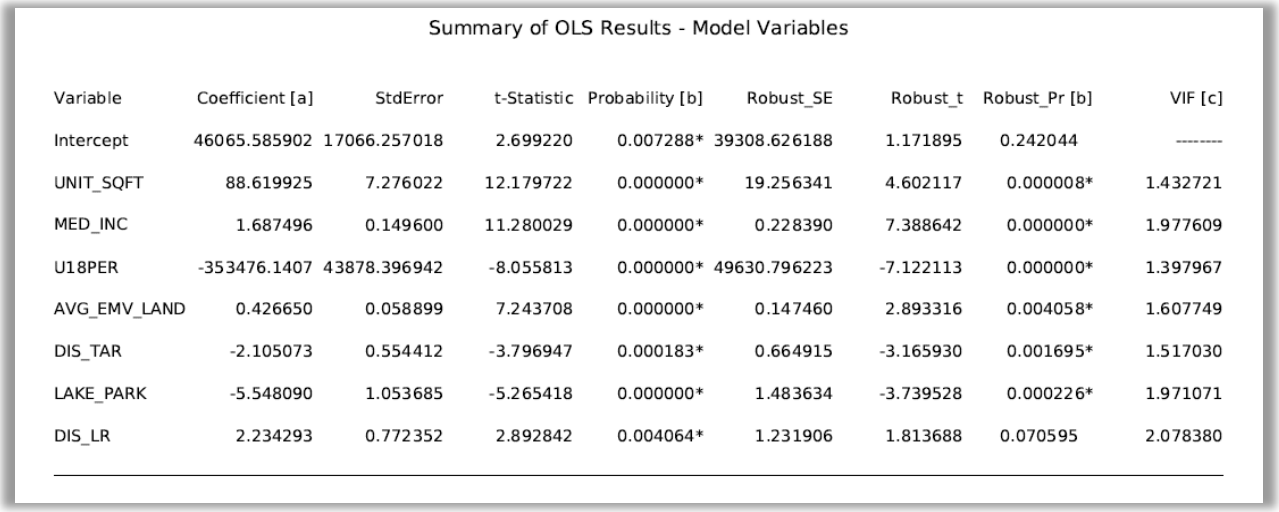
The OLS model for Minneapolis residential real estate sales used average sale price per census block group as the dependent variable ***(y)***, shown as **ADJ_SALE** on the OLS report, and seven independent **(*x*)** variables. The variable GIS database field names, descriptions and sources are listed below.

- **unit_sqft**: Average living area of dwellings in census block group, calculated in US feet^2, data supplied by Open Data Minneapolis
- **dis_tar**: Distance to Target Field, calculated in US Feet from the centroid of the census block group
- **dis_lr**: Distance to light rail station, calculated in US feet from the centroid of the census block group
- **u18per**: Percentage of total population of census block group that is under the age of 18, data supplied by 2010 US Decennial Census
- **med_inc** Median household income of households in census block group, data supplied by 2013 US Census American Community Survey
- **avg_EMV_land**: Average total land value of parcels within census block group, data supplied Hennepin County GIS MetroGIS Parcels layer, updated April 29, 2016
- **lake_park**: Distance to nearest public park with water based recreation available, calculated in US feet from the centroid of census block group

The results of the OLS model analysis of the 2016 data show that all seven of these independent variables were statistically significant in the model. The adjusted R^2 value of the OLS model is .773, meaning that approximately 77% of the variance in the dependent variable is explained by the OLS model (*Table 3*).

The corrected Akaike's Information Criterion (AICc) value, which is another measure of goodness of fit of the model to the data (along with the adjusted R^2 value) is 8871.16 (*See Table 4*). The actual AICc number in and of itself is not important until it is compared to other models of the dependent variable. If there is a difference of more than 4 in the AICc values of two models of the same dependent variable, the model with the lower value is considered a better fit for the data. When the GWR model of the data is run, that AICc number will be compared to the corresponding AICc of the OLS analysis, to determine if the local GWR model improved the goodness of fit for the data when compared to the global OLS model (Charlton and Fotheringham 2009). Low Variance Inflation Factor (VIF) readings for all seven independent variables indicate that there are no issues with redundancy among variables. A VIF reading of greater than 7.5 would indicate multicollinearity and signify variable redundancy (Charlton and Fotheringham).

Table 4- 2016 OLS Coefficient Report



Summary of OLS Results - Model Variables

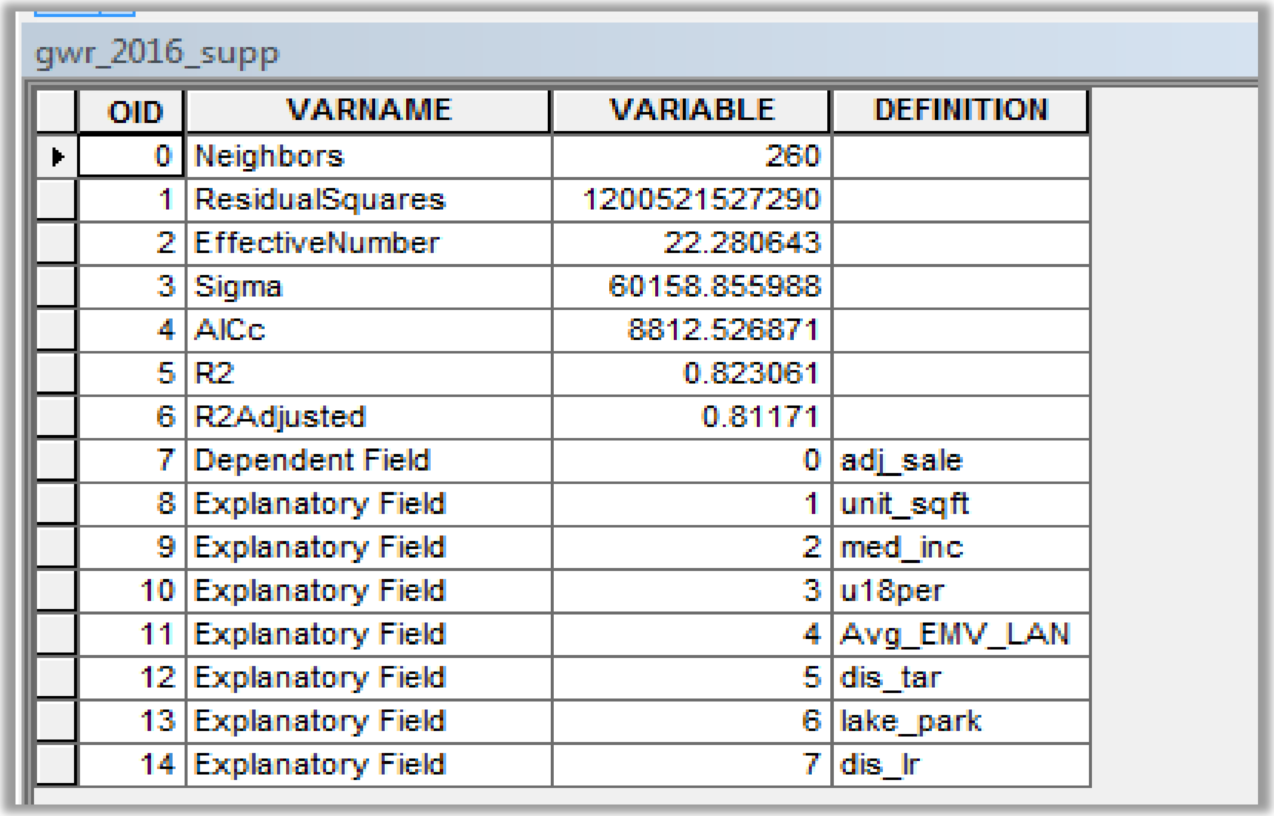
Variable	Coefficient [a]	StdError	t-Statistic	Probability [b]	Robust_SE	Robust_t	Robust_Pr [b]	VIF [c]
Intercept	46065.585902	17066.257018	2.699220	0.007288*	39308.626188	1.171895	0.242044	--------
UNIT_SQFT	88.619925	7.276022	12.179722	0.000000*	19.256341	4.602117	0.000008*	1.432721
MED_INC	1.687496	0.149600	11.280029	0.000000*	0.228390	7.388642	0.000000*	1.977609
U18PER	-353476.1407	43878.396942	-8.055813	0.000000*	49630.796223	-7.122113	0.000000*	1.397967
AVG_EMV_LAND	0.426650	0.058899	7.243708	0.000000*	0.147460	2.893316	0.004058*	1.607749
DIS_TAR	-2.105073	0.554412	-3.796947	0.000183*	0.664915	-3.165930	0.001695*	1.517030
LAKE_PARK	-5.548090	1.053685	-5.265418	0.000000*	1.483634	-3.739528	0.000226*	1.971071
DIS_LR	2.234293	0.772352	2.892842	0.004064*	1.231906	1.813688	0.070595	2.078380

The Jarque- Bera Statistic is significant and indicates bias in the residuals, and that the residuals are not normally distributed (ESRI). This is to be expected as the model is working with spatial data that is not randomly distributed in space, but rather, is spatially correlated. A Moran's I test of the standard residuals reveals that the residuals are indeed clustered spatially, with a Moran's I Z- Score of 2.467106. A Moran's I Z- Score of over 1.65 indicates significant spatial clustering of the standard residual values (ESRI). A model that analyzes each data observation at the local level like GWR should account for the spatial clustering that is present in the data set, as well as increase the Adjusted R^2 value and reduce the AICc value over the global OLS model (Charlton and Fotheringham) (*Table 4*).

4.4 Geographically Weighted Regression (GWR) Model and Specifications

Using the model parameters that were established by fitting the OLS model to the data, the GWR model was executed and the results were compared. Using ArcGIS 10.5 software, the GWR model used the nearest 260 neighbors of each data location to calculate the model. (*Table 5*).

Table 5- 2016 GWR Analysis Report



gwr_2016_supp

OID	VARNAME	VARIABLE	DEFINITION
0	Neighbors	260	
1	ResidualSquares	1200521527290	
2	EffectiveNumber	22.280643	
3	Sigma	60158.855988	
4	AICc	8812.526871	
5	R2	0.823061	
6	R2Adjusted	0.81171	
7	Dependent Field	0	adj_sale
8	Explanatory Field	1	unit_sqft
9	Explanatory Field	2	med_inc
10	Explanatory Field	3	u18per
11	Explanatory Field	4	Avg_EMV_LAN
12	Explanatory Field	5	dis_tar
13	Explanatory Field	6	lake_park
14	Explanatory Field	7	dis_lr

The GWR model performed better than the OLS model in terms of R^2 value, adjusted R^2 value, and in the AICc goodness of fit measure. The R^2 value for the GWR (local) model was .823 compared to .778 in the OLS (global) model. The GWR model also had a higher adjusted R^2 value than the OLS model, .811 (GWR), .773 (OLS). This means that the local model explained approximately 4% more of the variance in the dependent variable than the global model. The AICc goodness of fit statistic values were 8812 in the GWR model, and 8871 in the OLS model. The 59 point drop in AICc indicates that the GWR model fits data better statistically than the OLS model (Charlton and Fotheringham 2009)**.**

One of the goals of GWR analysis is to remove spatial correlation and bias from the residual results the model. To test if spatial correlation exists in the results of the GWR model, a Moran's I test was performed on the standard residual values. The standard residual values were found to be randomly distributed (z-score 0.428337), indicating no statistically significant spatial correlation in the residual values.

5.0 Results

5.1 Geographically Weighted Regression (GWR) Results – 2016 Real Estate Sales Data

When analyzing the results of a GWR model, mapping the residual values can provide the analyst with an initial idea of whether the model performed well in all locations within the study area, or if there are certain areas were the model dramatically under predicted or over predicted the dependent variable. The residuals are the values of the dependent variable that were not explained through the independent variables in the model. (Charlton and Fotheringham 2009).

The actual residual values are computed by taking the predicted dependent variable values produced by the model and subtracting the actual values of the dependent variable observed by the model. Standardized residual values are a ratio value calculated by dividing the residual number by its estimated standard deviation. Since GWR regression calculates a value of all variables for every data observation, every data observation will have its own standard residual value (ESRI).

Upon mapping the standard residuals for this data set using a choropleth design (*Figure 4*), it is apparent that most of the model falls within -2.5 to 2.5 standard deviations. The outliers are mostly greater than 2.5 standard deviations and are located on the border of the Central Community in Downtown. The only less than 2.5 standard deviations are immediately adjacent to Target Field. According to the map of the standard residuals, the local model performs well in all parts of the city expect for several places that are immediately adjacent to the downtown area, or in the shadow of the major highway junction of Interstate 394 and Interstate 94. These are primarily non-residential districts of the city, and the high deviation in the residuals could be due to the clustering of residential sales in one or two locations in each block group, as well as real estate values being impacted by the presence of the elevated highway.

GWR also calculates R^2 values for each data observation. These are known as Local R^2 values. Mapping the Local R^2 values can show local variation of how much of the dependent variable was explained by the model (*Figure 5*). The 2016 Local R^2 Map shows a trend of higher values in the Southwest of Minneapolis, gradually decreasing as it moves north and east.

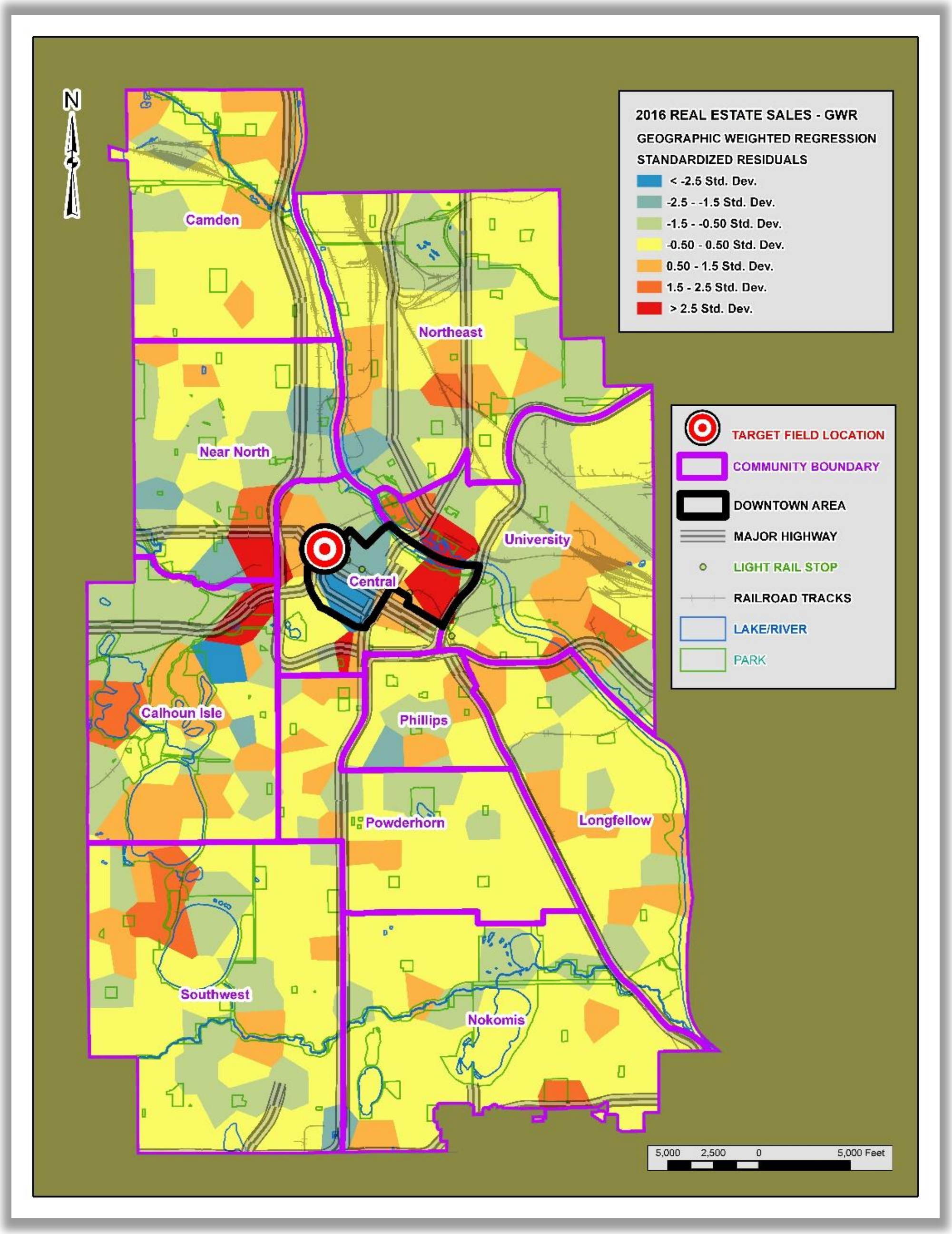


Figure 4 - 2016 GWR Standardized Residual Map

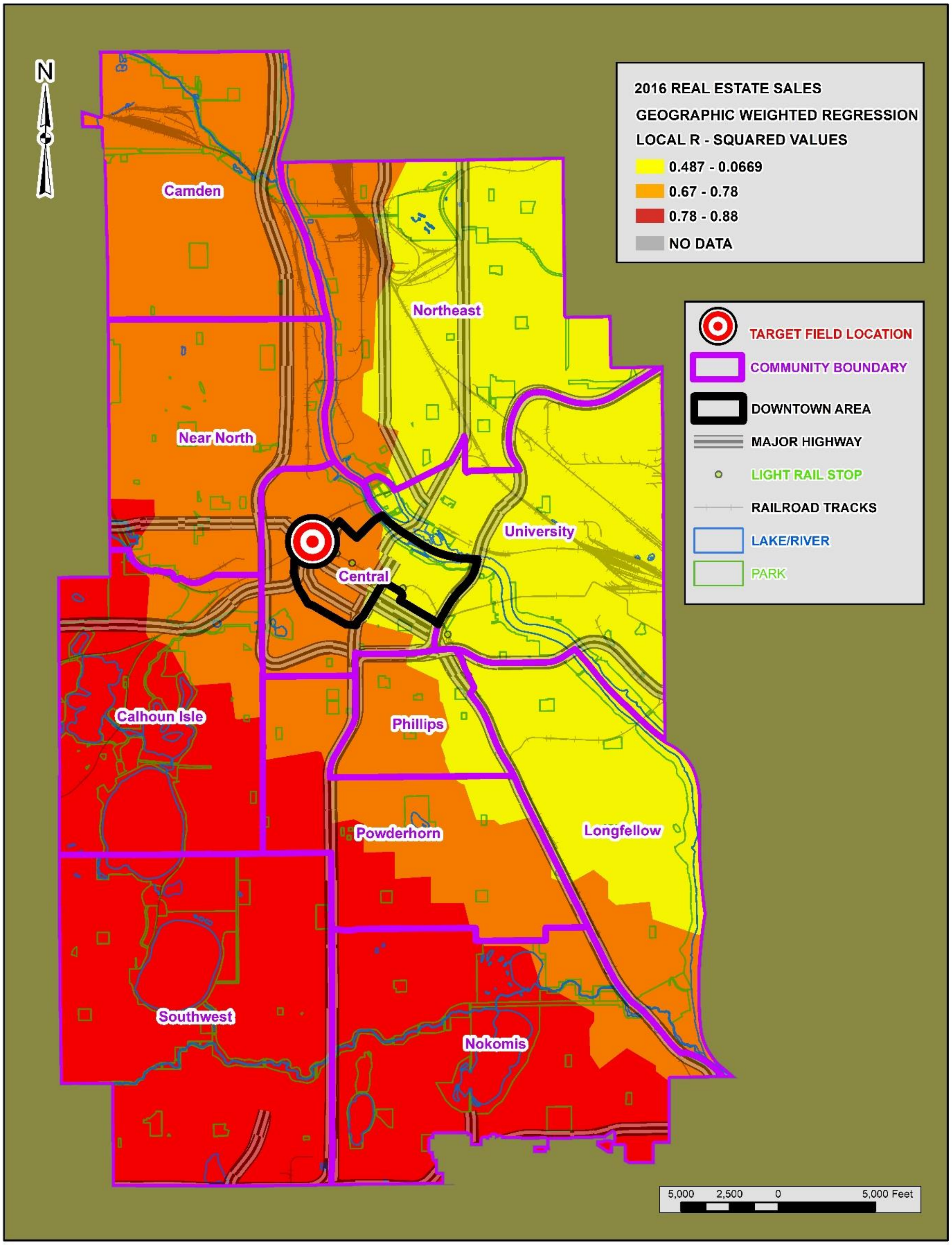


Figure 5- 2016 Local R - Squared Values

5.2 2016 GWR Analysis Coefficient Values and Variable Maps

The coefficient results of the GWR regression indicate changes in the value of the dependent variable due to the presence of the independent variable within the model. What that means is, if an independent (**x**) variable has a positive coefficient value, then as the value of that independent (**x**) variable increases, it is expected that the value of the dependent (**y**) variable will increase. Conversely, if the coefficient value of the independent (**x**) variable is negative, then it is expected that an increase in value of that independent (**x**) variable will lead to a decrease in the dependent (**y**) variable (Charlton and Fotheringham 2009).

Of the seven independent variables included in the GWR model, four of them produced positive coefficients values, and three of the variables produced negative coefficient values. Negative coefficient results occurred for the distance to Target Field (**dis_tar**) and distance to water/recreation (**lake_park**) variables. Since the variables representing distance to these local amenities are calculated fields that increase the further away that the block group centroid point is from the amenity, the coefficient results are expected to be negative based on the hypothesis that being close to such amenities would have a positive effect on real estate sale prices.

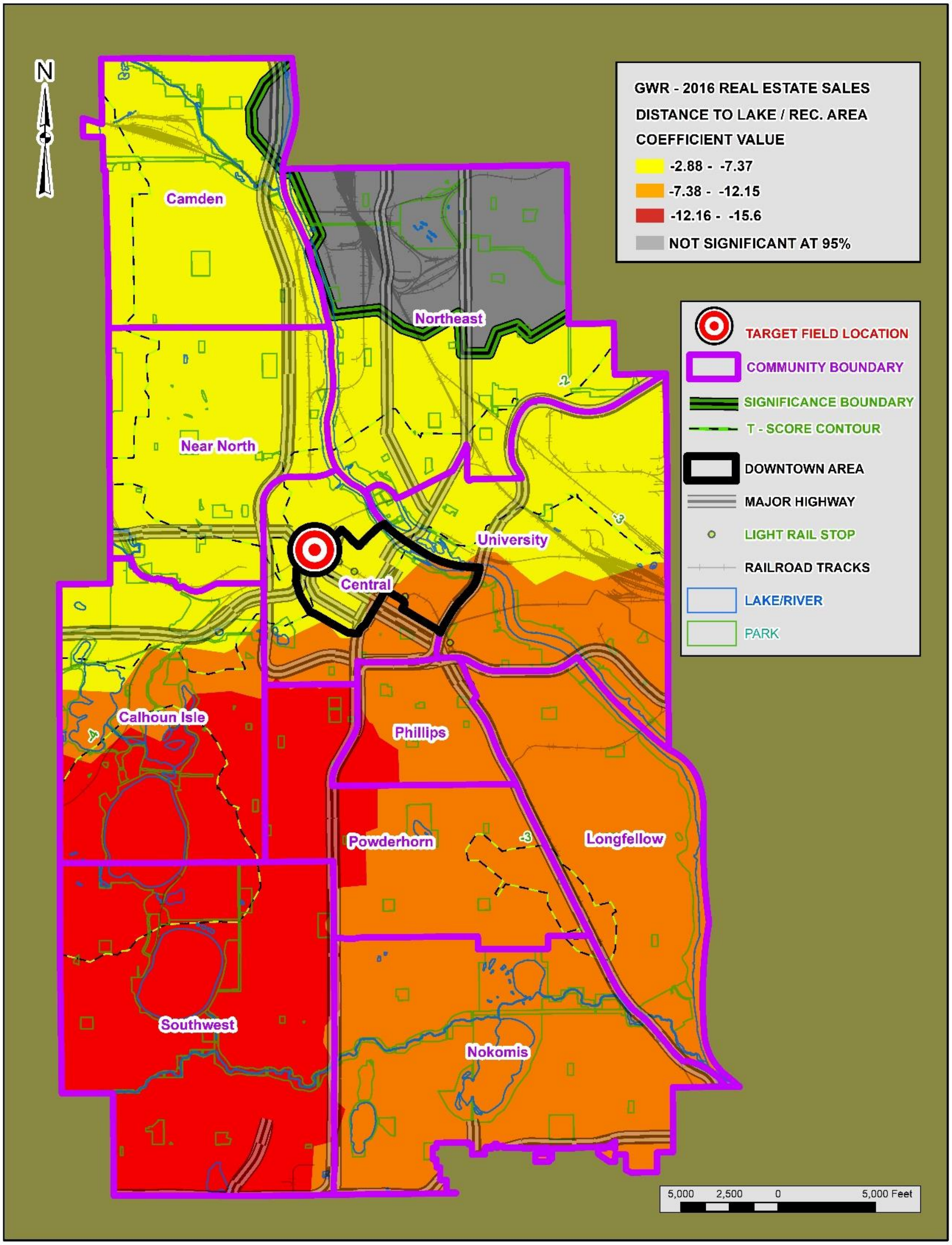


Figure 6- Distance to Water/ Recreation Coefficient Map

The distance to water/ recreation (**lake_park**) variable produced negative coefficient values (*Figure 6*). All data observations produced a negative value in the range of -2.88 - -15.6. As distance to water/recreation increases, residential real estate values decrease, and vice versa. The strong negative coefficients indicate a strong value added to real estate in those areas nearest water. This variable was statistically significant at 95% confidence in all areas of Minneapolis with the exception of the Northeast Community.

When the coefficient results are mapped, the data indicates that property being closely located to water/recreation locations were more of a factor in increased real estate sales prices in the Southwest, Calhoun Isle and Powderhorn Communities. This is to be expected since Southwest and Calhoun Isle are the most affluent areas of the city based on median income and overall land value, and also home to several lakes with public parks and recreation sites. It can be inferred from the results that people pay a premium to live in these areas of the city. Powderhorn abuts these two communities, so it is reasonable to assume that this geographic relationship has an effect on real estate values in that community.

The coefficient values for the distance to a Metro Light Rail station (**dis_lr**) variable produced positive results (*Figure 7*). However, this variable was only statistically significant at 95% confidence in the southern communities of the city. It was expected that being close to a light rail station would be seen as a positive among perspective home buyers, therefore a larger distance to a light station would produce negative coefficient values. This was not the case in Minneapolis in 2016. In the southern area of the city, which was the only area where this variable was statistically significant, the communities of Southwest and Nokomis show that a greater distance from a light rail station produced a strong positive coefficient value, indicating that a shorter distance to a light rail station had a negative effect on property values in those communities.

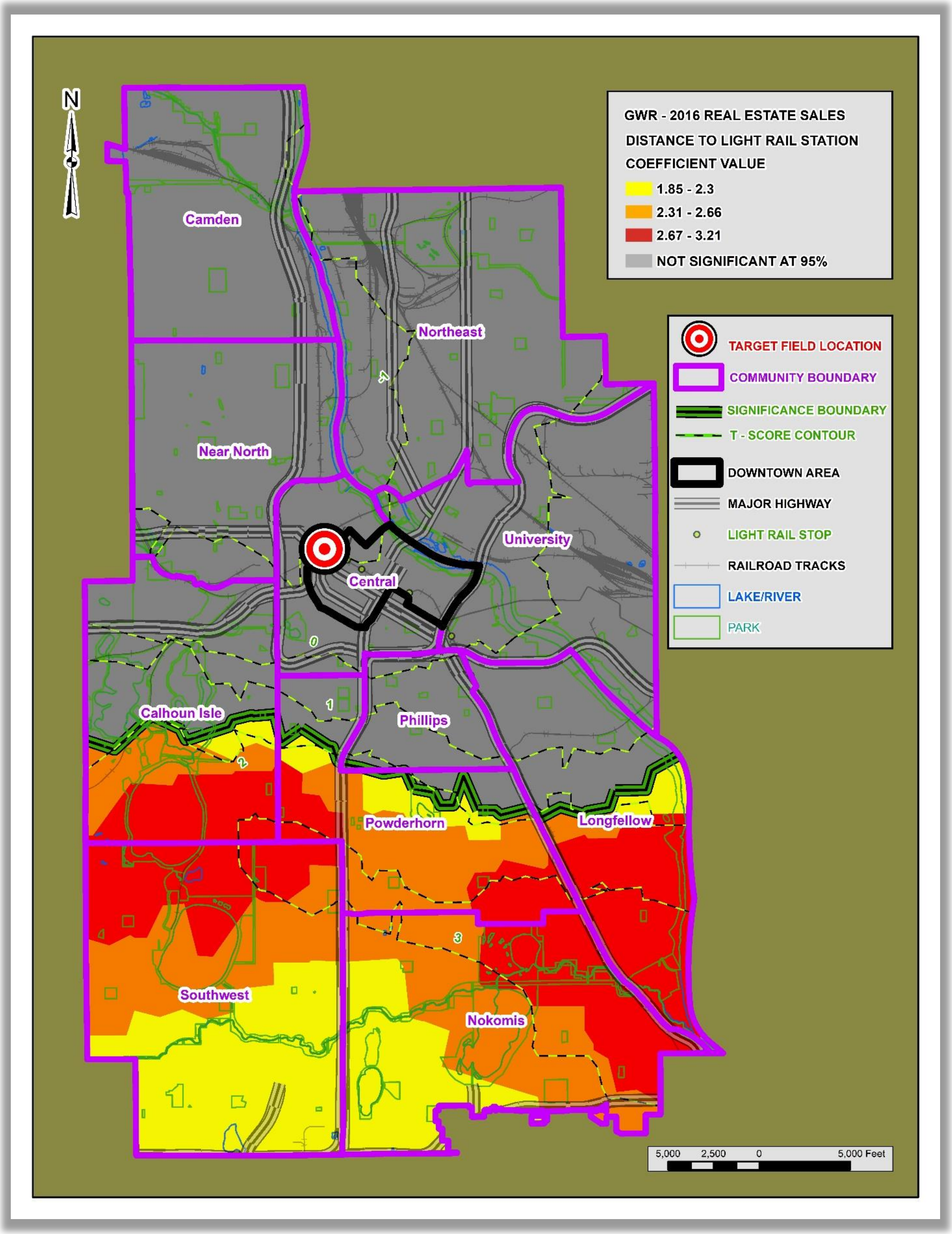


Figure 7- Distance to Light Rail Station Coefficient Map

The variable in the model that had the largest effect size of all the independent variables in the model was percentage of the population of the block group that is under the age of 18 (**u18per)**. This variable produced a high negative coefficient value. The coefficient values for this variable showed a significant trend for lower negative values, and therefore the greatest impact, in the lower income northern communities than the higher income southern communities. This variable explains more of the real estate value in Minneapolis that any other variable in the model (*Figure 8*). The reason for this relationship is beyond the scope of this thesis.

One of the variables that produced entirely positive coefficient values as a result of the GWR model was the average living area of a dwelling in the census block group (**unit_sqft**), measured in U.S. Feet^2 (*Figure 9*). The coefficient values for this particular variable, while all positive and statistically significant, decreased from west to east across the city. This could be due to the fact that living space is more a part of the value of a dwelling in primarily single family home neighborhoods away from downtown, which is largely dominated by condominium and apartment units. Living area of a dwelling had a greater effect on 2016 home sales in Calhoun Isle and Southwest (again the most affluent area of the city, containing primarily single family homes) and less in the Longfellow and University areas of the city.

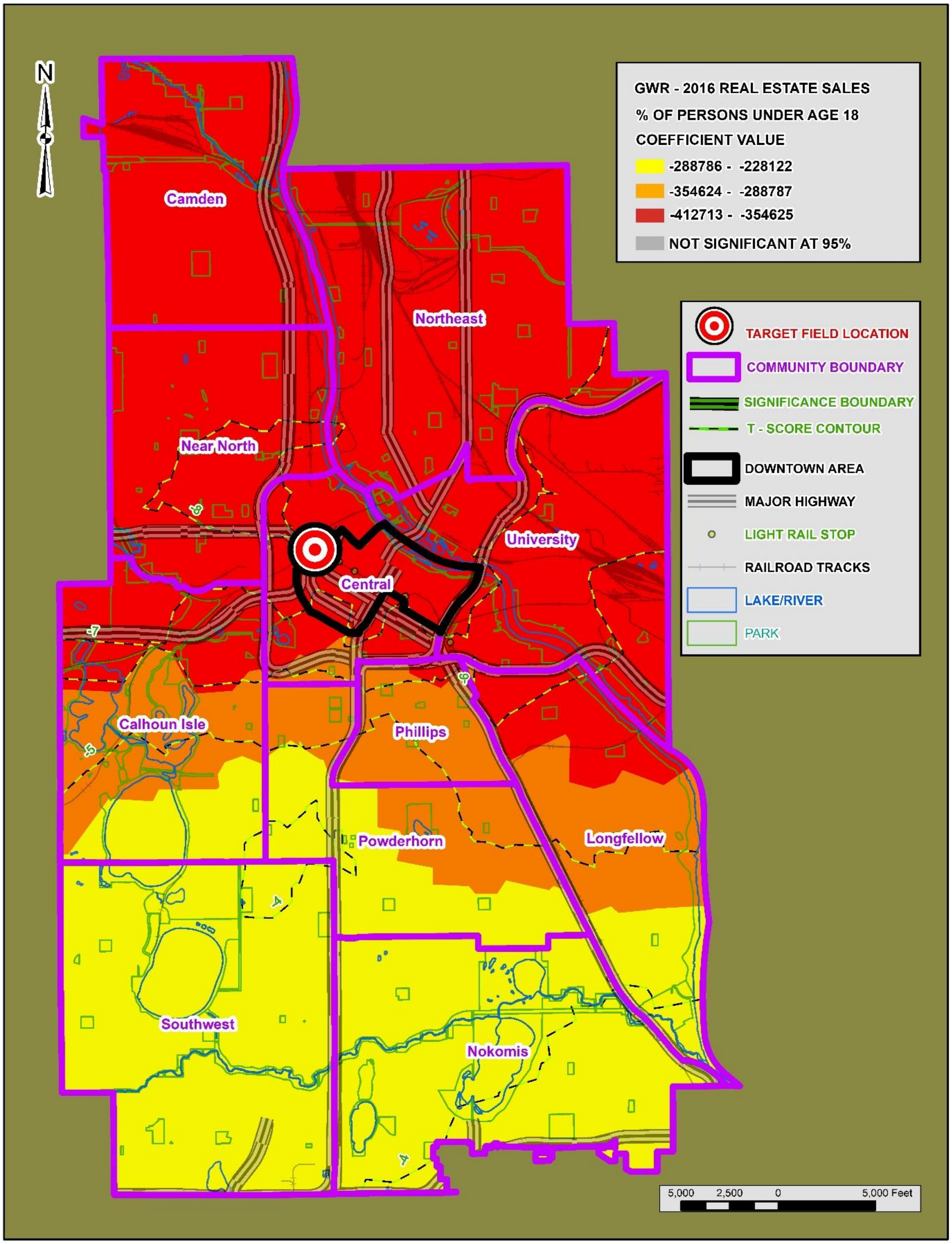


Figure 8- Percentage of Block Group Population that is Under 18 Years of Age Coefficient Map

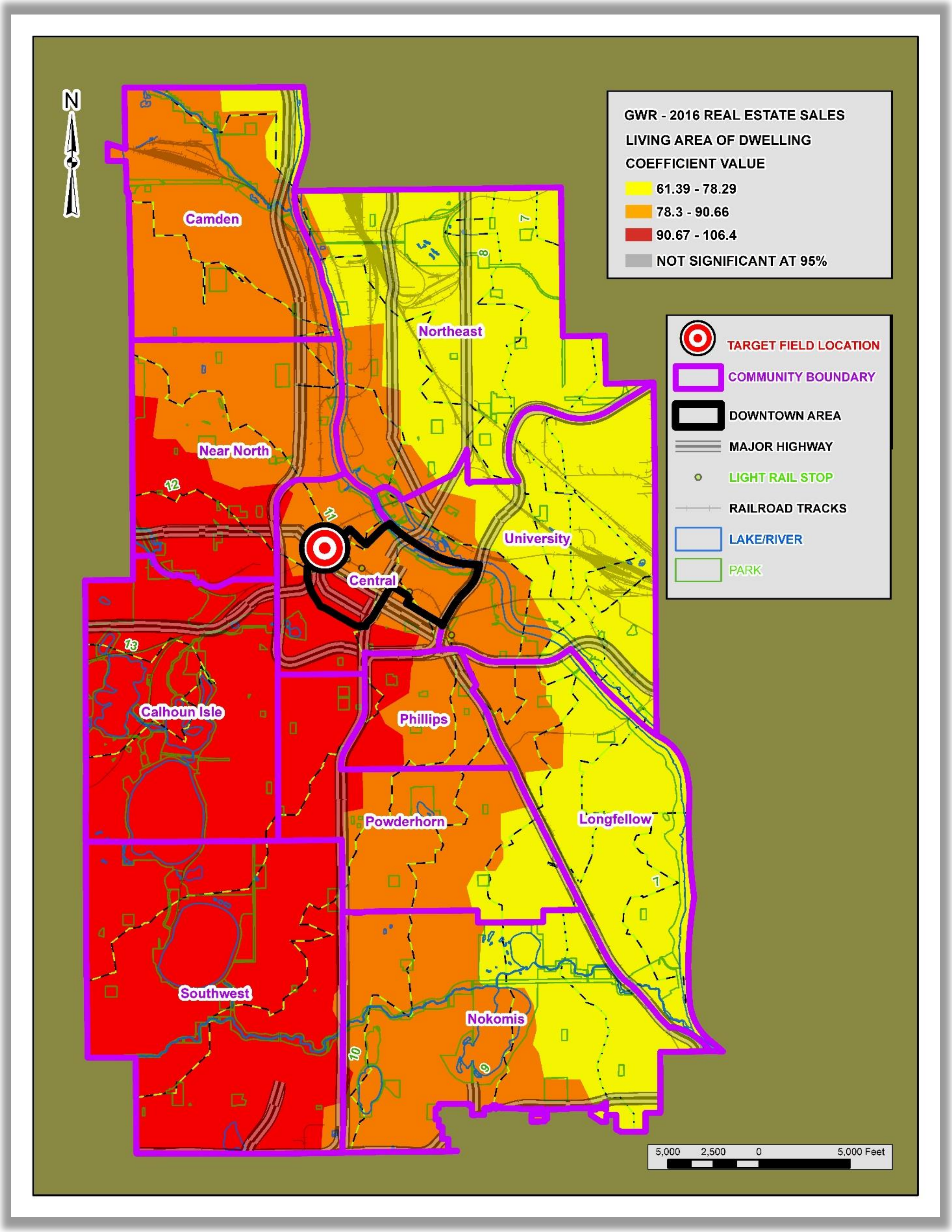


Figure 9- Average Living Area of Dwellings in Block Group Coefficient Map

Another variable that showed entirely positive coefficient values was median annual income for households in the census block group (**med_inc)**. There was very little variance in the coefficient results for this variable, however, the results were statistically significant for a large majority of the city. The exception to this was a large section of the University Community. This could be due to relatively low income being reported by students living in the area around the University of Minnesota. The coefficient values did trend higher in the downtown and North Minneapolis areas. This indicates that median household income was a greater factor in sale price in those areas, as opposed to other parts of the city. (*Figure 10*).

The average overall land value for parcels in the census block group (**AVG_EMV_LAND**) variable also produced positive coefficient values. Again, coefficient values were higher in the southern communities of the city, Southwest, Nokomis, and Calhoun Isle. (*Figure 11*).

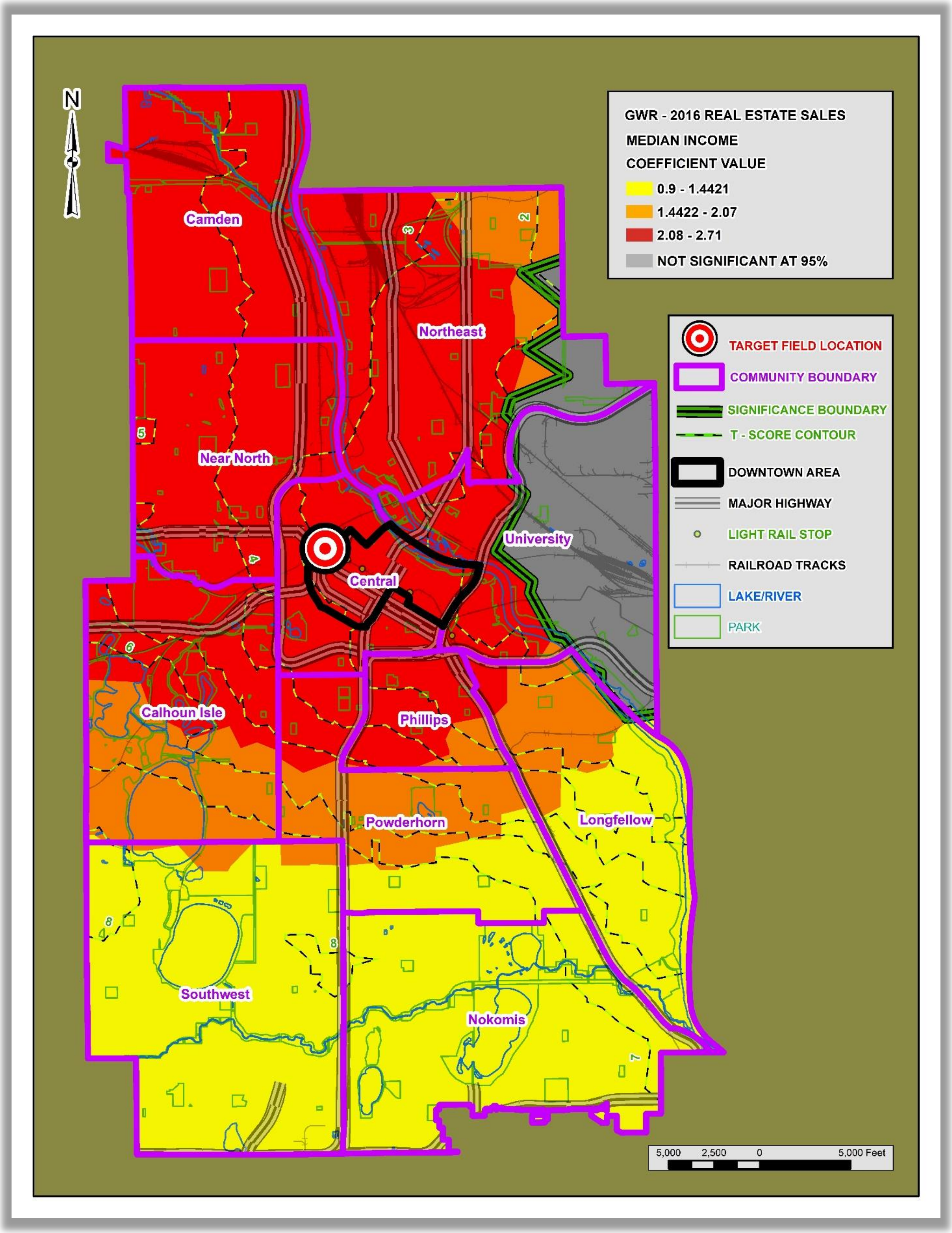


Figure 10- Median Household Income in Block Group Coefficient Map

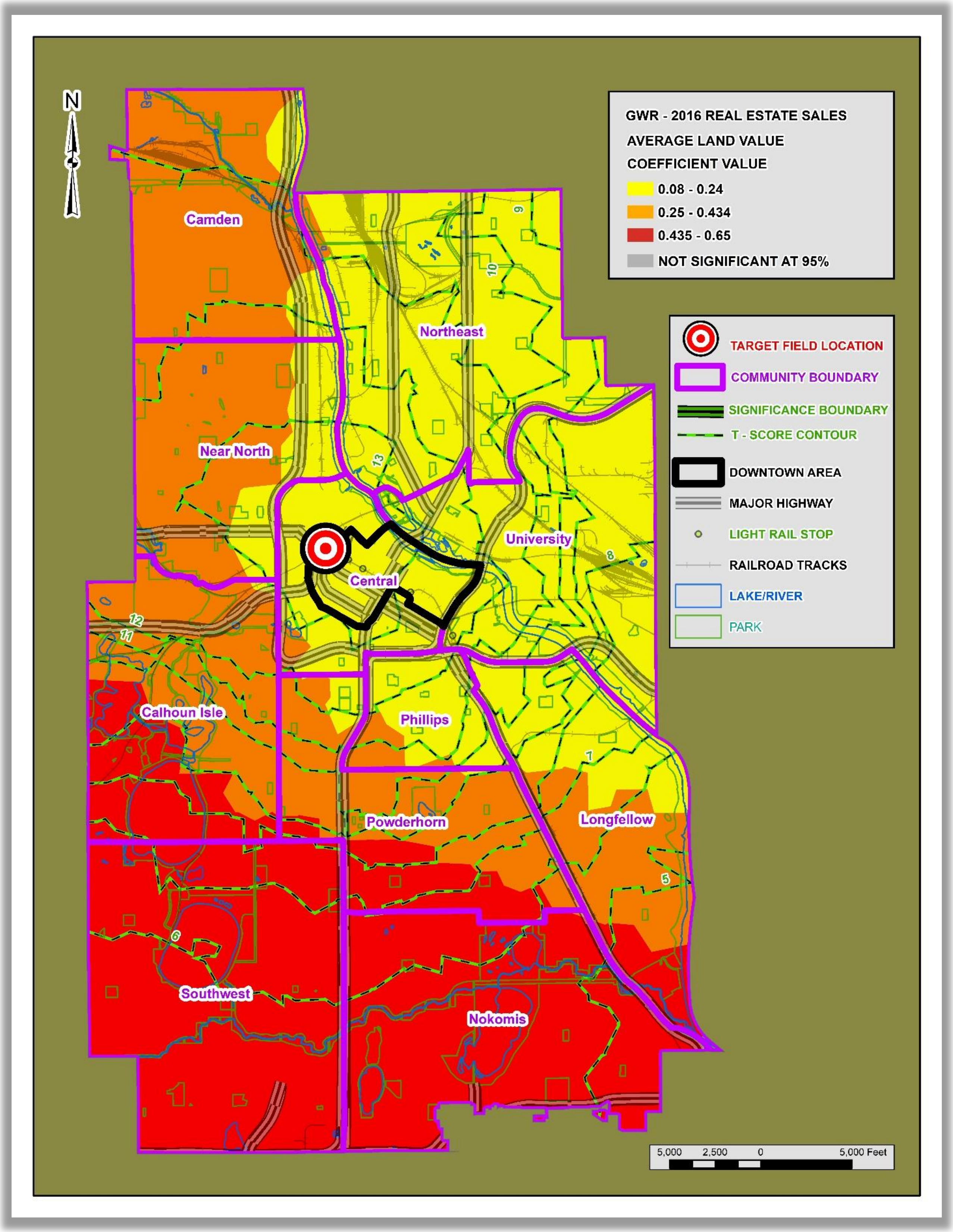


Figure 11- Average Land Value of Parcels in Block Group Coefficient Map

The final variable in the GWR model is the subject of the thesis, distance to Target Field (**dis_tar**). Since one of the assumptions of this thesis is that being in close proximity to a Major League Baseball stadium is considered an amenity that potential home buyers would consider a positive benefit when purchasing a home in the area, a higher distance from the ballpark would create a negative coefficient value. The results of the GWR analysis show a statistically significant negative coefficient for a majority of the census block groups in Minneapolis in 2016. The value of these coefficients show the same distance decaying effects that Feng and Humphreys (2018) witnessed in their results (*Figure 12*). However, the distance decay in this model is centered to the south east of the ballpark, closer to the geographic center of downtown. This could indicate that the value added to real estate in the city is not the existence of the ballpark itself, but in the properties' proximity to Downtown Minneapolis. Feng and Humphreys (2018) theorized that this may be the case when they concluded in their analysis that Nationwide Arena in Downtown Columbus added $222.5 million to real estate values in that city, and Crew Stadium located 4 miles away from the downtown area only added $35 million to real estate values. They questioned whether Nationwide Arena really was a contributing factor to the rise in real estate values, or if it was its location and the proximity to Downtown Columbus that was the real driving factor in the added value to their model.

The next section of this thesis explores if real estate values are affected by Target Field, or if real estate values are affected by the proximity of those properties to Downtown Minneapolis, and would be that valuable given their location if the ballpark were there or not. The analysis was conducted by placing a dollar value on the ballpark's influence on real estate in Minneapolis based on the 2016 coefficient values.

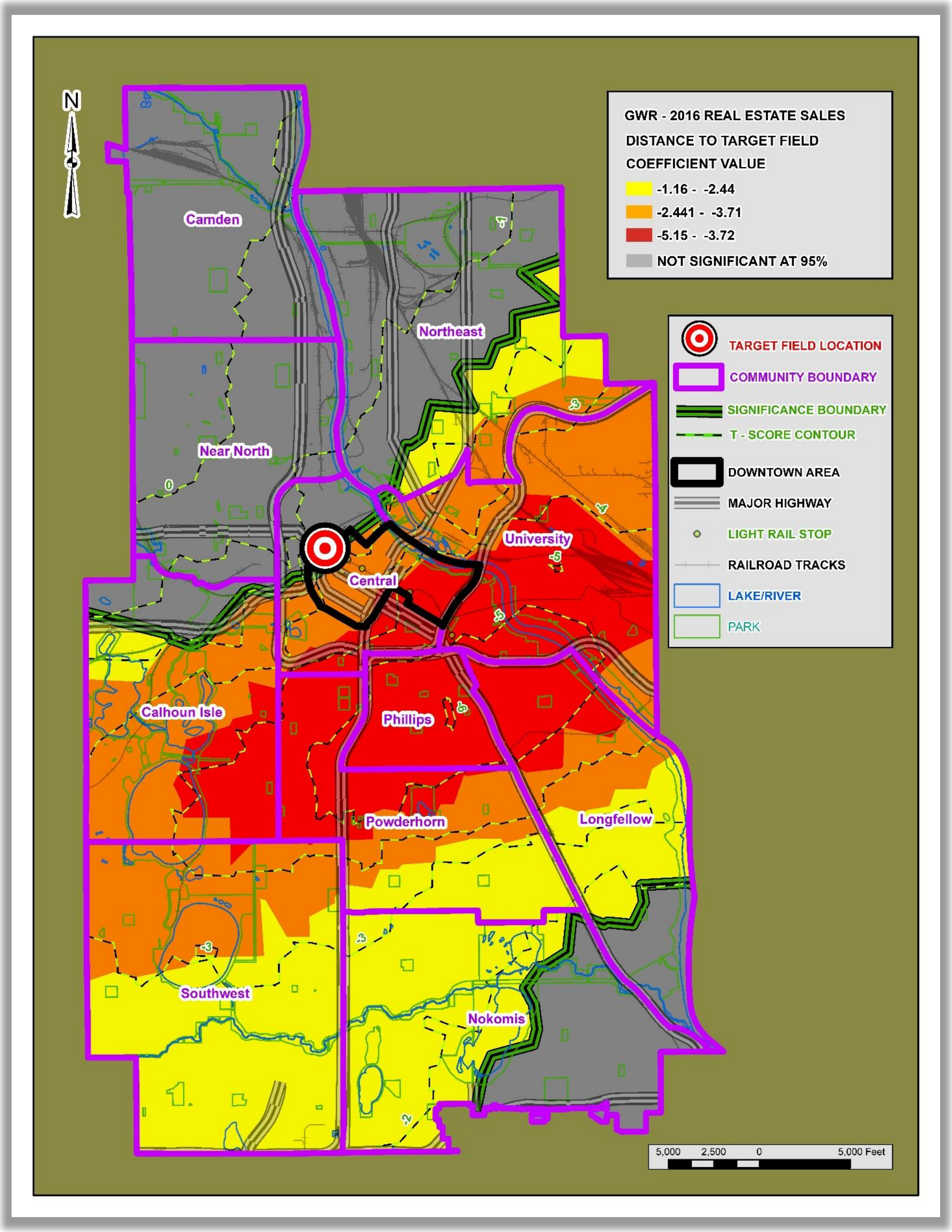


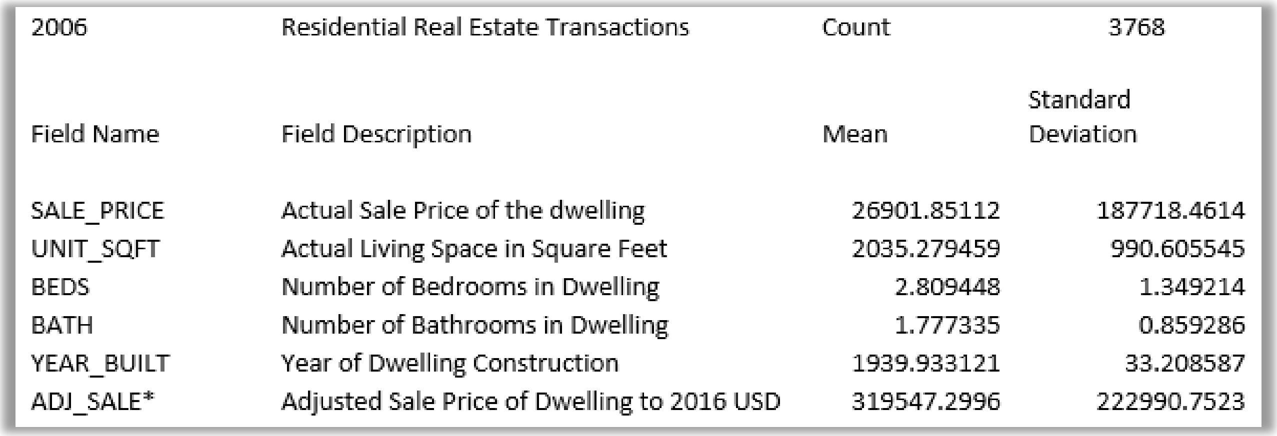
Figure 12- Distance to Target Field Coefficient Map

6.0 Analysis of Minneapolis Real Estate Sales Data Prior to Approval of Target Field

6.1 Data Description – 2006 Data

The same OLS and GWR analysis was performed on real estate sales data from 2006, which was the year before Target Field was finally approved, to see what the impact of proximity to that then vacant location in Downtown Minneapolis had on real estate values in the city. The data from those two years will then be compared to determine a final figure on the impact of Target Field to residential real estate values in Minneapolis.

Table 6- 2006 Residential Real Estate Sales Statistics



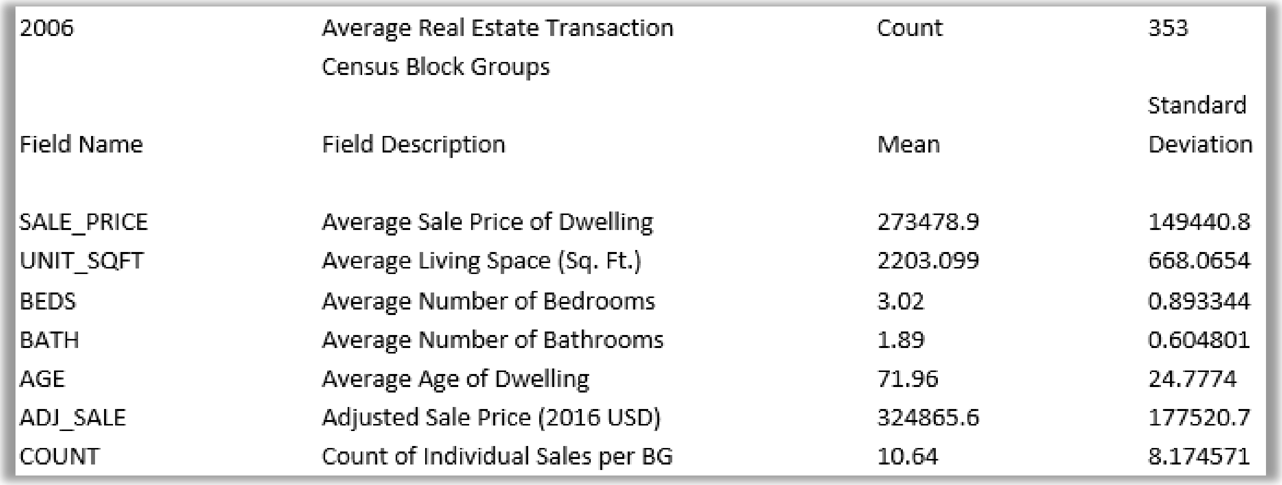
2006	Residential Real Estate Transactions	Count	3768
Field Name	Field Description	Mean	Standard Deviation
SALE_PRICE	Actual Sale Price of the dwelling	26901.85112	187718.4614
UNIT_SQFT	Actual Living Space in Square Feet	2035.279459	990.605545
BEDS	Number of Bedrooms in Dwelling	2.809448	1.349214
BATH	Number of Bathrooms in Dwelling	1.777335	0.859286
YEAR_BUILT	Year of Dwelling Construction	1939.933121	33.208587
ADJ_SALE*	Adjusted Sale Price of Dwelling to 2016 USD	319547.2996	222990.7523

The 2006 real estate point data set was comprised of 3768 individual sales. As was the case in the 2016 analysis, all of the data points were recorded sales of residential housing units of all types (single family homes, condominiums, town houses, etc.). The 3768 home sales in 2006 were approximately two thirds of the number of home sales that occurred in 2016 (5689). It should be noted that there may be sample size bias when comparing the two data sets (*Table 6*).

When an analysis compares financial transaction data from two data sets that occurred a number of years apart, it is important to correct for inflation over that time period. According to the US Bureau of Labor Statistics, $1000 in June of 2006 had the same buying power as $1187.87 in June of 2016 (Bureau of Labor Statistics 2017). To account for this inflation rate, the actual sale prices for all properties in 2006 were increased by 18.78 %. The inflation adjusted sale prices are located in the **adj_sale** field in the database.

Using the same methodology as the 2016 analysis, the 2006 real estate sale points were averaged by the 2010 US Decennial Census Block Group that they occurred in. The block groups with the adjusted average real estate sale prices for 2006 were then joined with 2010 US Decennial Census population data, 2009-2013 US Census American Community Survey 5 – year estimate median income data, and average parcel value data from the Minneapolis City Assessor's Office. From there, centroid values for those block groups were calculated using GIS. In all, 353 of the 413 US Census Block Groups in Minneapolis contained at least two residential real estate sales in 2006 (*Table 7*). The centroid values of the block groups with recorded real estate sales for 2006 were calculated. From these block group

Table 7- 2006 Residential Real Estate Sales by U.S. Census Block Group

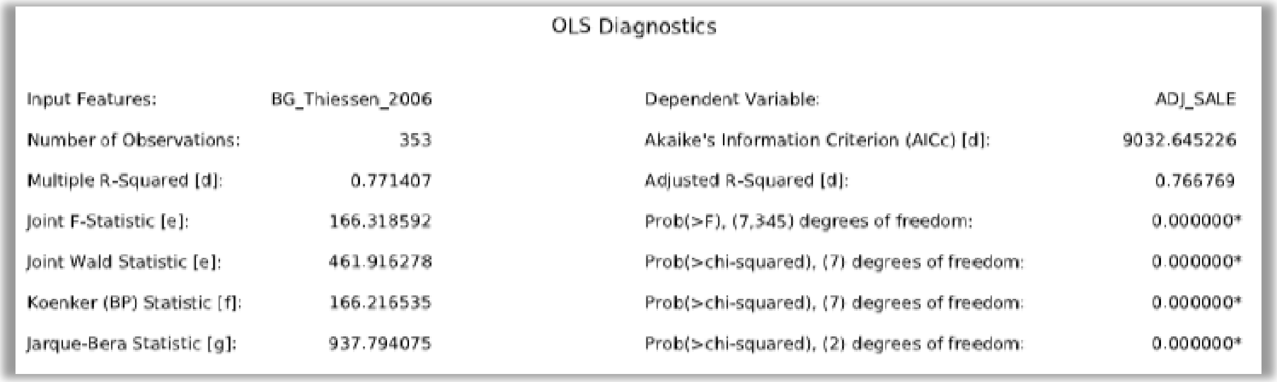


2006	Average Real Estate Transaction Census Block Groups	Count	353
Field Name	Field Description	Mean	Standard Deviation
SALE_PRICE	Average Sale Price of Dwelling	273478.9	149440.8
UNIT_SQFT	Average Living Space (Sq. Ft.)	2203.099	668.0654
BEDS	Average Number of Bedrooms	3.02	0.893344
BATH	Average Number of Bathrooms	1.89	0.604801
AGE	Average Age of Dwelling	71.96	24.7774
ADJ_SALE	Adjusted Sale Price (2016 USD)	324865.6	177520.7
COUNT	Count of Individual Sales per BG	10.64	8.174571

centroid points, distances to Minneapolis amenities such as Target Field, the Metro Light Rail and recreational areas on a lake or river were calculated for each point. Finally Thiessen polygons were created from the 353 centroid points to create a contiguous field for use in both the OLS and GWR models.

2010 US Census Blocks Groups were used as the base polygons in all analysis years in order to maintain a consistent methodology for every year in the data set. This is the same rationale for using 2010 Census Population data and 2013 ACS five year estimate median income for all analysis years as well. Parcel value information was only available for 2016 at the time of this analysis, so that data layer was used for both analysis years. A copy of the Python script that was written to process this data from the original individual sale points through the GWR analysis is available in *Appendix A* of this thesis.

Table 8- 2006 OLS Report



OLS Diagnostics

Input Features:	BG_Thiessen_2006	Dependent Variable:	ADJ_SALE
Number of Observations:	353	Akaike's Information Criterion (AICc) [d]:	9032.645226
Multiple R-Squared [d]:	0.771407	Adjusted R-Squared [d]:	0.766769
Joint F-Statistic [e]:	166.318592	Prob(>F), (7,345) degrees of freedom:	0.000000*
Joint Wald Statistic [e]:	461.916278	Prob(>chi-squared), (7) degrees of freedom:	0.000000*
Koenker (BP) Statistic [f]:	166.216535	Prob(>chi-squared), (7) degrees of freedom:	0.000000*
Jarque-Bera Statistic [g]:	937.794075	Prob(>chi-squared), (2) degrees of freedom:	0.000000*

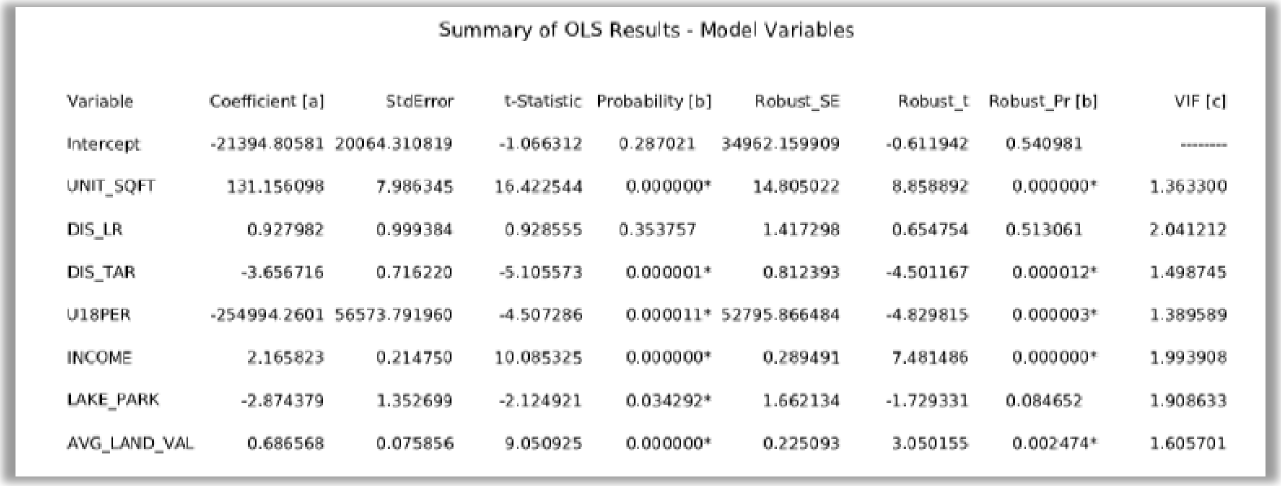
6.2 2006 OLS Model Specifications and Results

The OLS model for the 2006 data was executed with the same seven independent variables as the 2016 OLS model, however it produced very different results. Of the seven variables only six were statistically significant (*Table 8*). The variable that was not statistically significant was the distance to a light rail station (**dis_lr**)).

The distance to light rail being statistically insignificant is expected. The Metro Blue Line light rail service had just began operating in June 2004. It is possible that this new service had not affected real estate sale prices in the first two years of service.

The study variable that is the focus of this thesis however, the distance to Target Field, was again among the statistically significant variables in 2006. This is surprising since the ballpark had not been developed yet. In fact, the site of the ballpark, or even if The Minnesota Twins were going to relocate to either St. Paul or another market entirely, was still being debated until the approval of funding by Hennepin county and the settlement of the land taking issues in 2007(St. Paul Pioneer Press 2007). The fact that the distance to the ballpark was significant in the 2006 OLS model may indicate that what Feng and Humphreys (2018) speculated in their study, namely that the value added to real estate prices in Columbus, Ohio by the presence of Nationwide Arena, may be a product of the arena's downtown location, and not the arena itself, could possibly be happening with Minneapolis real estate values in regards to Target Field and its downtown location (*Table 9*).

Table 9- 2006 OLS Coefficient Report



Summary of OLS Results - Model Variables

Variable	Coefficient [a]	StdError	t-Statistic	Probability [b]	Robust_SE	Robust_t	Robust_Pr [b]	VIF [c]
Intercept	-21394.80581	20064.310819	-1.066312	0.287021	34962.159909	-0.611942	0.540981	--------
UNIT_SQFT	131.156098	7.986345	16.422544	0.000000*	14.805022	8.858892	0.000000*	1.363300
DIS_LR	0.927982	0.999384	0.928555	0.353757	1.417298	0.654754	0.513061	2.041212
DIS_TAR	-3.656716	0.716220	-5.105573	0.000001*	0.812393	-4.501167	0.000012*	1.498745
U18PER	-254994.2601	56573.791960	-4.507286	0.000011*	52795.866484	-4.829815	0.000003*	1.389589
INCOME	2.165823	0.214750	10.085325	0.000000*	0.289491	7.481486	0.000000*	1.993908
LAKE_PARK	-2.874379	1.352699	-2.124921	0.034292*	1.662134	-1.729331	0.084652	1.908633
AVG_LAND_VAL	0.686568	0.075856	9.050925	0.000000*	0.225093	3.050155	0.002474*	1.605701

6.3 2006 GWR Model Specifications and Results

When the same GWR model that was used to analyze the 2016 data was applied to the 2006 data, the model produced higher coefficient values for the distance to Target Field (**dis_tar**) variable across the city when compared to 2016. The distance to Target Field (**dis_tar)** variable was statistically significant at 95% in 86% of block groups in 2006, as opposed to 68% of the block groups being statistically significant at 95% in 2016. In other words, the area of statistical significance for the variable shrank 21% in Minneapolis from 2006 to 2016. The mean coefficient value in 2006 was -5.02 for the distance to Target Field variable. In 2016, the maximum or highest coefficient value observed in any census block group was -5.1. Mapping the standardized residual values showed several census block groups to the west of downtown producing both high (>2.5 Standard Deviation) and low outlier values (<-2.5 Standard Deviation), which was consistent with the 2016 standardized residual mapping of the GWR results, however they were located slightly to the west. (*Figure 13*). The map of the 2006 Local R^2 values shows a similar trend of a higher explained variance in the southwest communities of the city, gradually decreasing moving to the north and east (*Figure 14*).

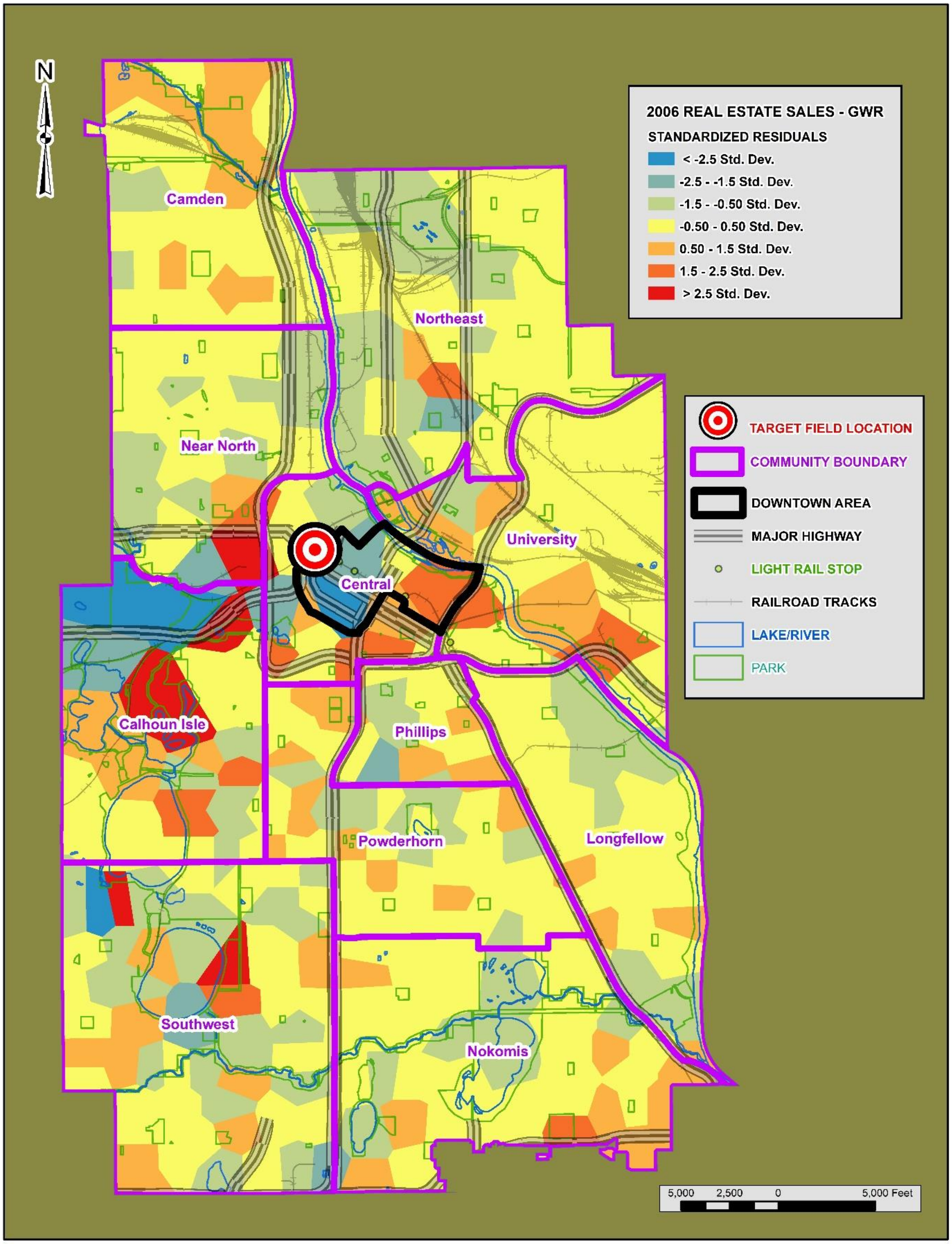


Figure 13- 2006 GWR Standardized Residual Map

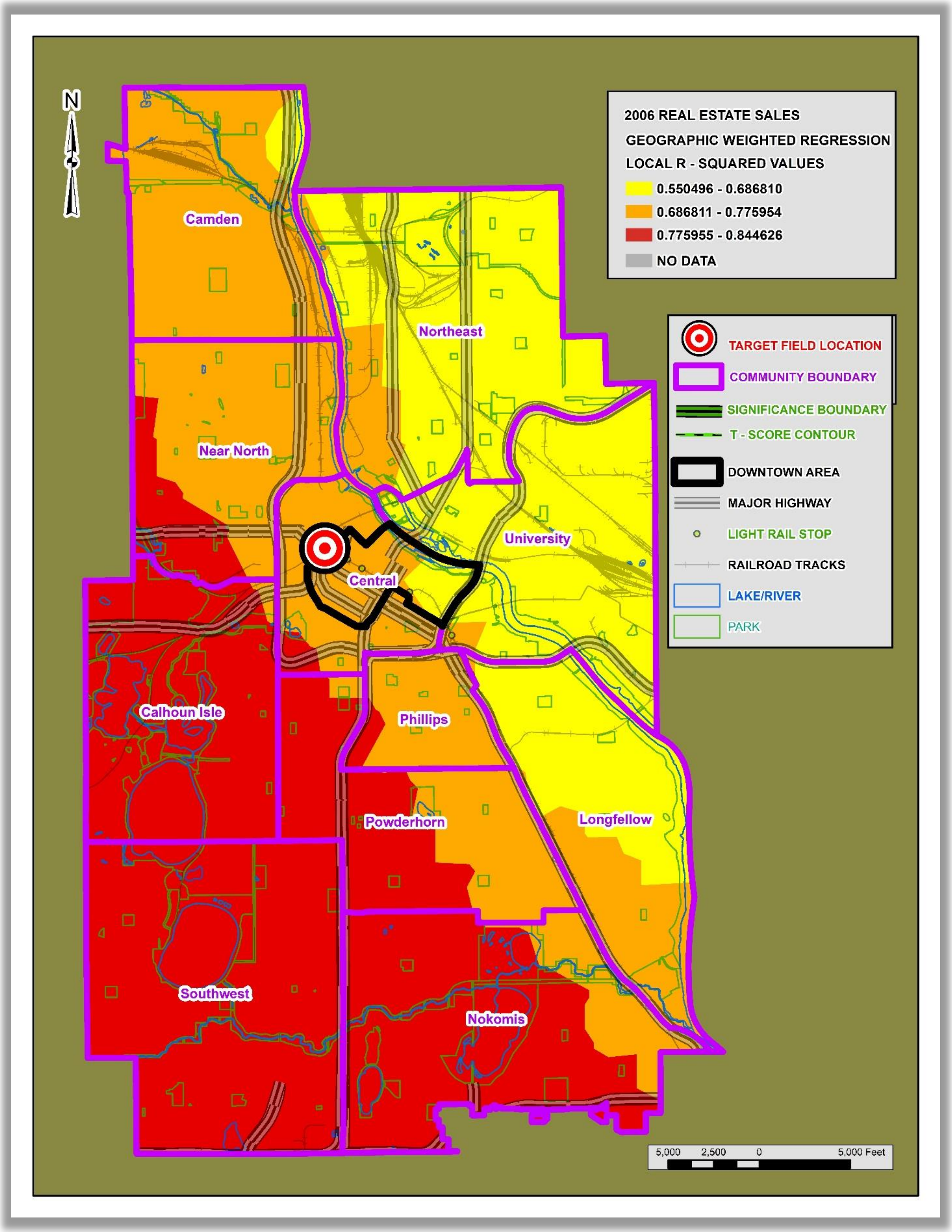


Figure 14- 2006 Local R- Squared Values

The GWR results also showed a distinct geographic pattern that became apparent when the distance to Target Field (**dis_tar**) variable was mapped. Mapping the results of the distance to Target Field **(dis_tar**) coefficient values showed a large concentration of high coefficient values in the Calhoun Isle Community. As discussed earlier in this thesis, this is the most affluent and expensive area of the city. It has the highest median income and is a primarily single family home community with close proximity to both recreational areas and downtown. The coefficient values starting at a point and reducing the further away from that point known as “Distance Decay” (Feng Humphreys 2018) is occurring in the Calhoun Isle community in 2006, instead of in the downtown area to the southeast of the ballpark as it was occurring in 2016 (*Figure 15*).

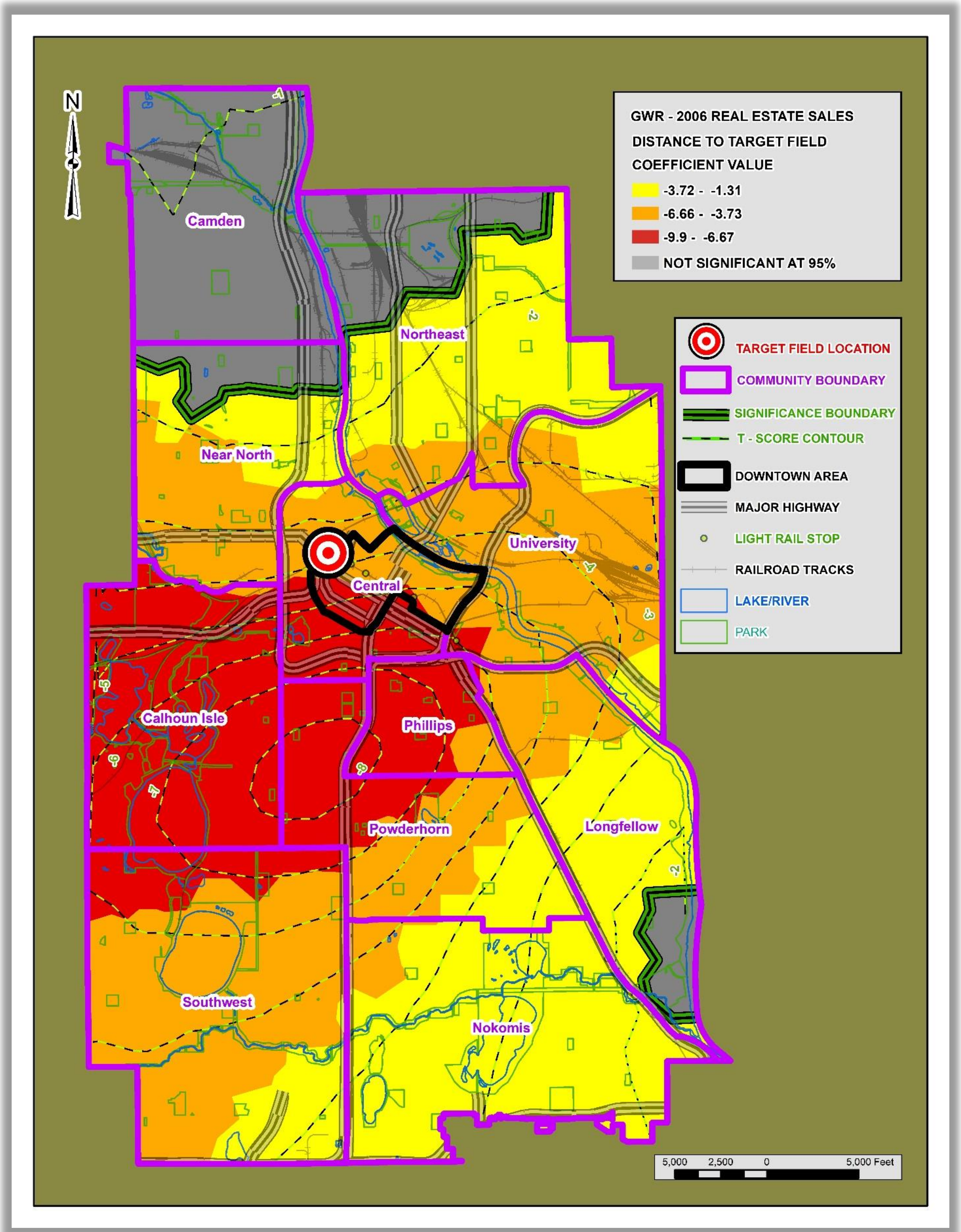


Figure 15- 2006 Distance to Target Field Coefficient Map

7.0 Comparison of 2016 and 2006 Results

7.1 Interpreting Value from Coefficient Results

In GWR analysis, the coefficient values for the independent variable are expressed in the units of the dependent variable (Charlton and Fotheringham 2009). In this case, the dependent variable, adjusted average sale price per block group (**adj_sale**), is measured in US dollars (USD). The independent variable that this thesis focuses on, the distance to Target Field (**dis_tar)** variable is measured in US feet.

The distance to Target Field (**dis_tar**) independent variable is part of a seven variable model. The coefficient values, and the dollar figures derived from them, would not be the same if analyzed as a one variable model. The numbers expressed by this coefficient are for the variable as a part of the overall GWR model and cannot be interpolated independently of it.

The coefficient value indicates that for that census block group, for every foot away from target field, the dollar value of adjusted average sale price goes down by that coefficient number. All of the significant coefficient values are negative because the value in the “Distance to Target Field” (**dis_tar**) variable are higher the further that the block group is from Target Field. In other words, low values for (**dis_tar**) equal higher negative coefficients. The actual dollar value of the Target Field impact can be measured by multiplying the absolute value of the coefficients by the actual distance to Target variable.

7.2 Overall Real Estate Value Change from 2006 - 2016

When both the 2016 and 2006 GWR distance to Target Field (**dis_tar**) coefficient results are mapped the 2006 coefficient numbers are higher in almost all areas of the city. The 2006 distance to Target Field coefficient map indicates especially high coefficient values in the community of Calhoun Isle and parts of abutting communities.

When the coefficient results are calculated into actual dollar values, the results show a much higher dollar value for the distance to Target Field variable (**dis_tar**) in the 2006 GWR model as compared to the 2016 GWR model. According to the results of the 2016 GWR model, the distance to Target Field variable contributed a mean value of $47,078.00 (USD) per residential real estate sale in areas of Minneapolis where the variable was statistically significant that year. This is a mean value for the 240 census block groups that recorded at least two residential real estate sale in 2016, and produced a coefficient value in the distance to Target Field (**dis_tar**) variable that was statistically significant at 95%. Since GWR is a "local" model, and coefficient results are produced for each observation in the data set (Fotheringham, Brunsdon, and Charlton 2002). A dollar value for an average residential real estate sale was calculated for each of the 240 census block groups. These values were then multiplied by the number of actual residential real estate transaction in that block group in that year. These dollar value statistics were then added together to produce a total dollar figure impact for the distance to Target Field variable in the 2016 GWR model, which was $178,281,907.00 (USD) (*Figure 16*).

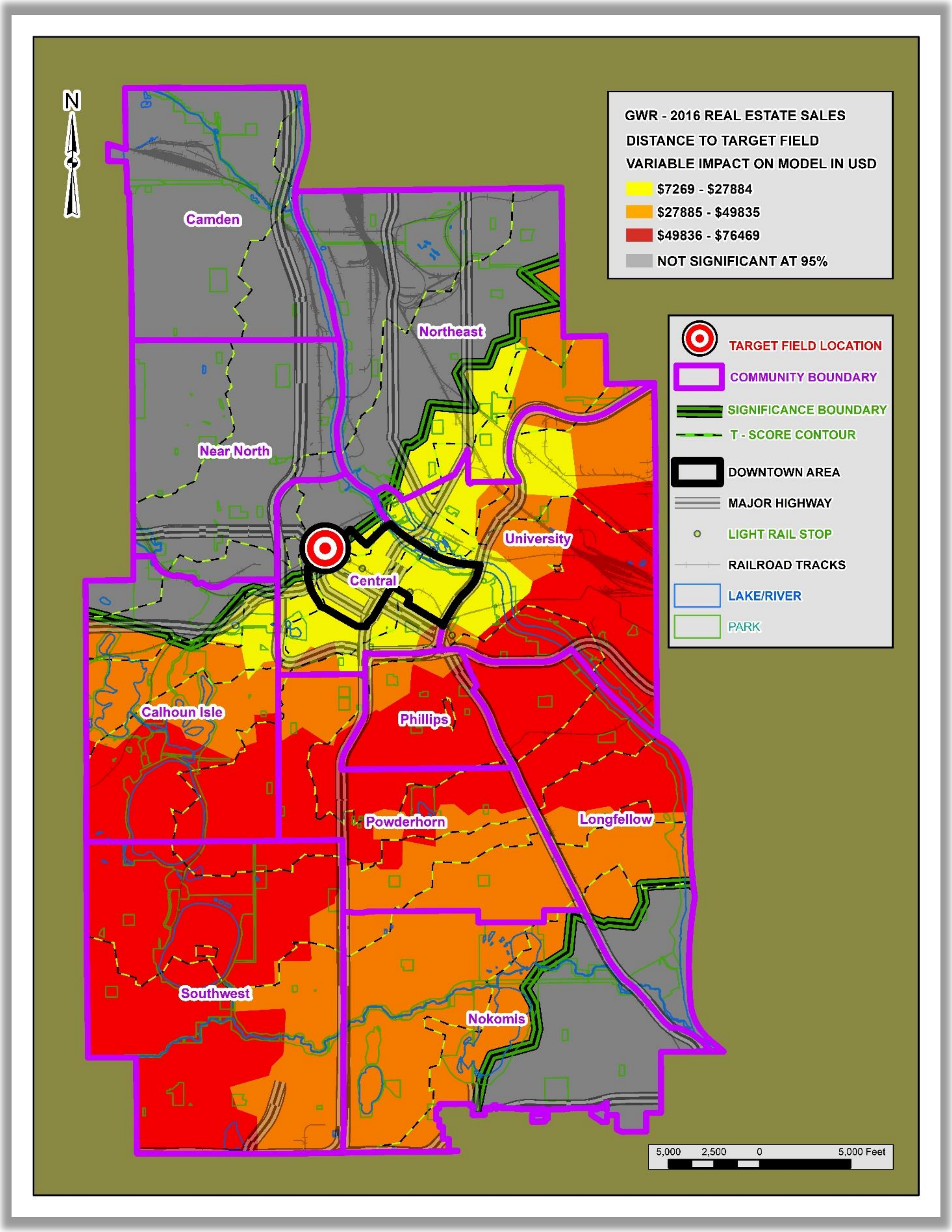


Figure 16- 2016 Distance to Target Field Coefficient Dollar Value Map

At face value, the results of the 2016 GWR analysis indicate that the presence of Target Field may indeed contribute a substantial sum of money to residential real estate values in Minneapolis. However, when the same analysis is repeated on the results of the 2006 GWR analysis, the dollar values calculated from the distance to Target Field variable were considerably higher in that model than in the 2016 model. The mean value that the distance to target Field (**dis_tar**) variable produced as a result of the 2006 model was $78,036.00 (USD). This dollar figure is inflation adjusted to 2016 USD. The actual value in 2006 USD was $63,815.00. This is still considerably higher than the $47.078.00 (USD) mean impact that the distance to Target Field (**dis_tar**) variable had on residential real estate prices as part of the 2016 GWR model. As was the case with the 2016 GWR model, the mean dollar impact figures were calculated from the 2006 GWR coefficient results, which resulted in a mean financial impact figure for each of the 305 census block groups that contained at least two residential real estate sales that year, and produced a coefficient value in the distance to Target Field (**dis_tar**) variable that was statistically significant at 95% in 2006. When added together, the total dollar figure impact for the distance to Target Field (**dis_tar**) variable was $227,316,856 (USD) (*Figure 17*). Again this figure is adjusted for inflation to 2016 USD. In 2006 USD, the total distance to Target Field variable impact figure is $185,424,683. The GWR models indicate that the distance to Target Field (**dis_tar**) variable is worth approximately $49 million less in 2016 than it was in 2006, which was the year before the ballpark was actually approved.

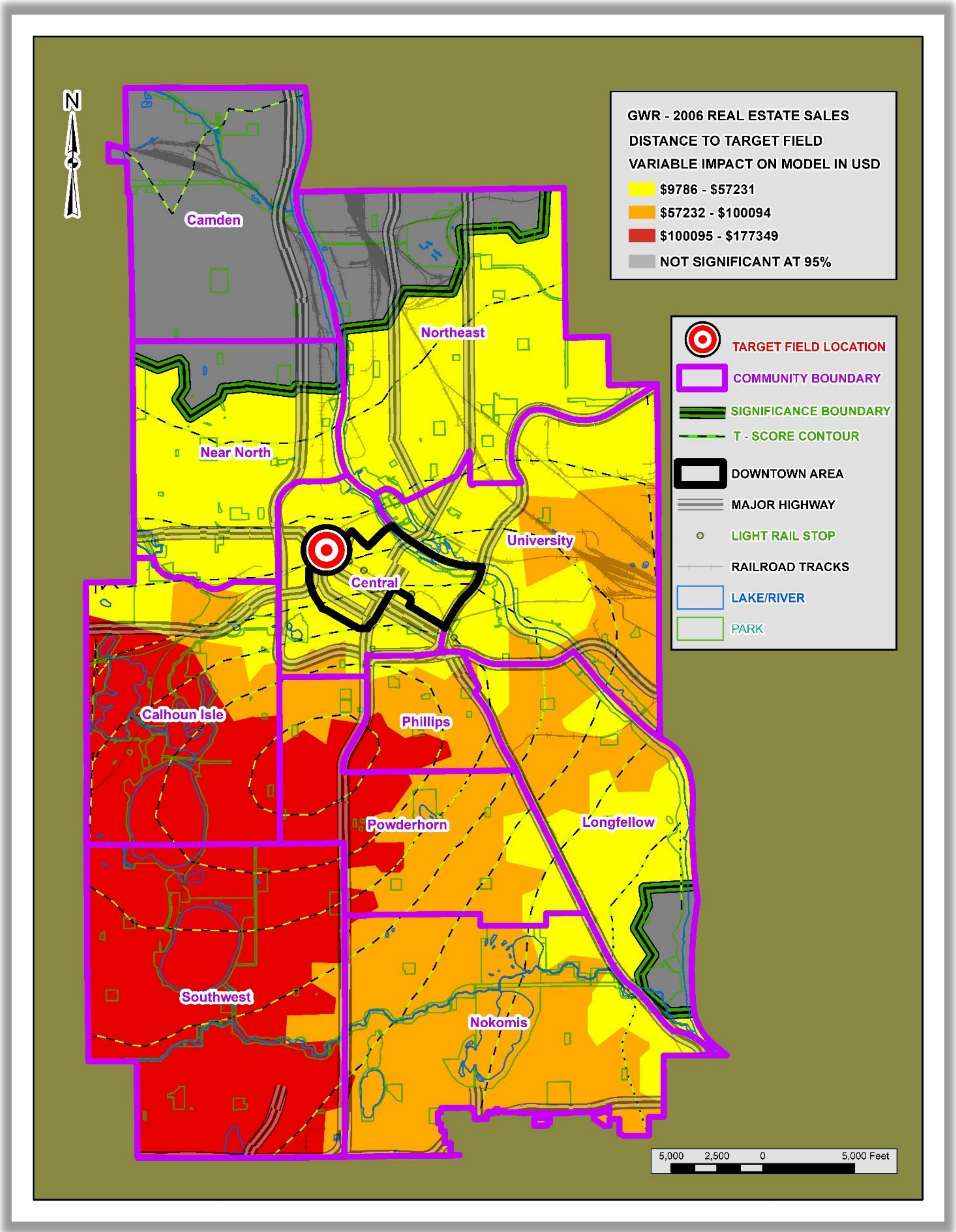


Figure 17- 2006 Distance to Target Field Coefficient Dollar Value Map

7.3 Local Patterns in Real Estate Value Change from 2006 – 2016

GWR models are built to analyze local patterns. In this case, it is found that residential real estate values in the City of Minneapolis may have gone down over a ten year period from 2006 to 2016. The question is where in the city has the difference been most pronounced, and what are some of the reasons for it? Comparing coefficient maps to identify areas of change is crucial when performing geographic analysis.

When comparing the distance to Target Field coefficient value maps using GIS, what stands out to the naked eye is that the coefficient numbers are considerably higher in the Calhoun Isle Community in 2006 than in any other community in either year. These high coefficient values seem to extend into adjacent block groups in neighboring communities in 2006. In 2016, Calhoun Isle still has higher coefficient values for the distance to Target Field variable (**dis_tar**) than most other areas of the city, but they are less than half of what they were in 2006. The highest coefficient values for the distance to Target Field (**dis_tar**) variable in 2016 are located just to the south and west of Target Field, surrounding the Phillips Community.

Using raster data of the coefficient values that were produced as part of the GWR process, the distance to Target Field variable coefficient value raster from the 2016 analysis was subtracted from the 2006 distance to Target Field variable coefficient value raster to detect local differences between the two coefficient results. The results of this analysis produced a map that shows which areas of the city had the largest coefficient change from 2006 to 2016 (*Figure 18*).

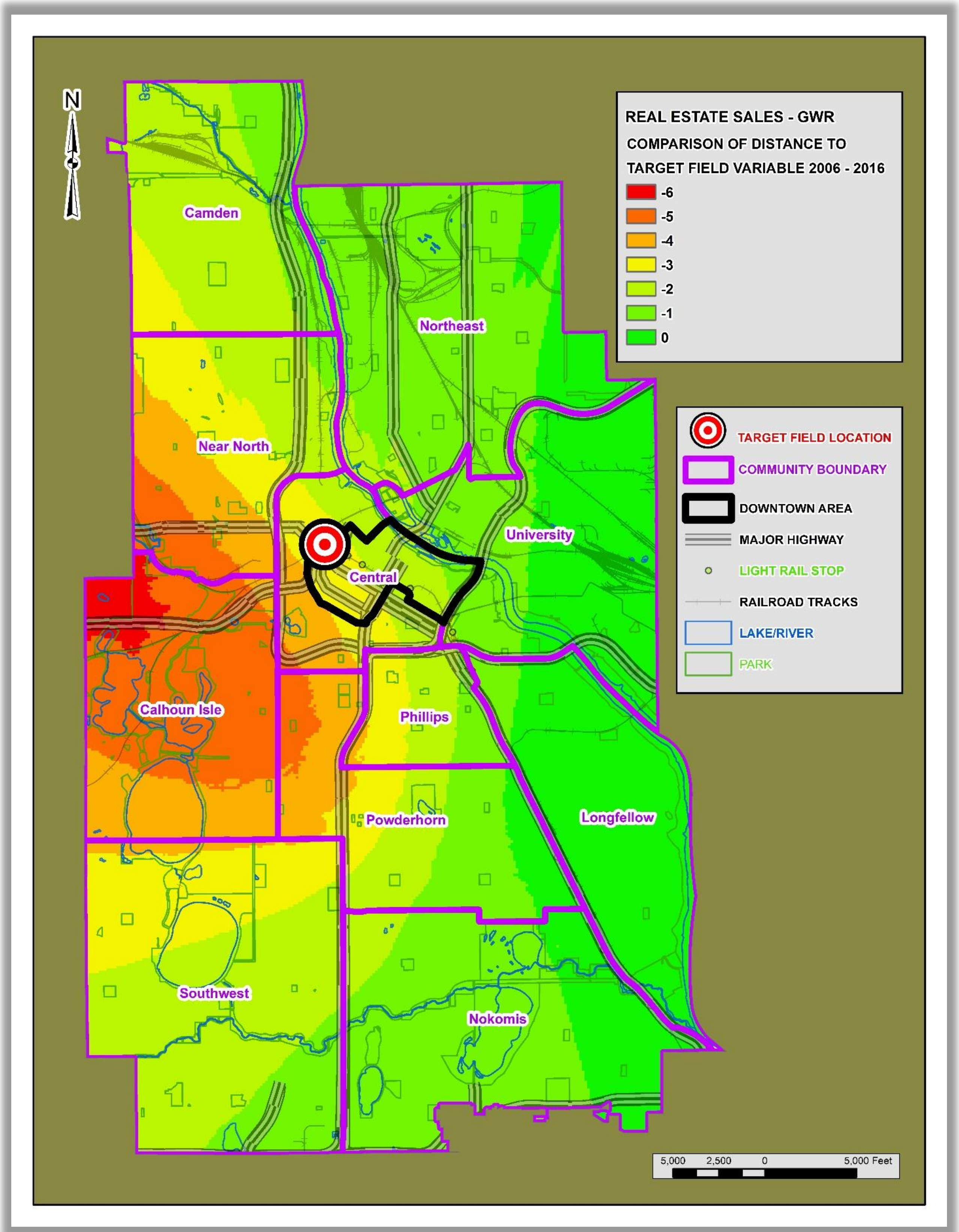


Figure 18- GWR Distance to Target Field Variable Difference 2006-2016

With the exception of a couple of small areas that showed slightly above zero improvement, all areas of the city had a lower coefficient value for the distance to Target Field (**dis_tar**) in 2016 than in 2006. The area that lost the most coefficient value was the Calhoun Isle Community. The comparison map (*Figure 18*) shows that the distance to Target Field (**dis_tar**) coefficient value in Calhoun Isle went down by a value of 5-6 between 2006 and 2016. The results also show that the loss of value for this coefficient had a distance decaying effect from the low point in Calhoun Isle, meaning the further in distance away from Calhoun Isle, the less the coefficient value loss was.

Table 10- Impact of Target Field Variable on GWR Model in USD

	2016 USD	2006(adjusted to 2016 USD)	2006 USD
Count	240	304	304
Mean Impact per Sale	$47,078	$78,205	$63,815
Total Impact	$178,287,907	$227,236,132	$185,424,683

A potential reason for this loss of value could be that real estate in that community has not completely rebounded from the housing crisis of last decade (U.S. Bureau of Labor Statistics). Calhoun Isle, given its location close to downtown, prominence of single family dwellings, and featuring amenities such as access to lakes and recreation areas, may have been an overvalued area before the housing crash and may not yet have fully rebounded. However, the cause of this possible drop in residential real estate values in Minneapolis from 2006 to 2016 is beyond the scope of this thesis.

8.0 Conclusion

This thesis concludes is that, based on a geographically weighted regression (GWR) model using seven hedonic independent variables, with average real residential real estate sale price per census block group as the dependent variable, the distance to Target Field in Downtown Minneapolis has a positive and statistically significant effect on those values. It can be inferred from the analysis that the impact of the distance to Target Field could possibly contribute approximately $178,281,907 to residential real estate values in the city in 2016 as part of the overall model. This is consistent with the findings of Tu (2005), Ahlfeldt and Maennig (2007), and particularly Feng and Humphreys (2018). However this thesis sought to answer the question posed by Feng and Humphreys (2018) as to whether a downtown stadium actually contributes those dollars to local real estate values, or if it is just merely the stadiums' downtown location that actually adds the value. Using the same GWR model on real estate transaction data in Minneapolis, Minnesota, recorded in the year prior to the final approval of Target Field (2006), the results of the model show that the value of that variable was approximately $49 million higher in the year before the final approval of the stadium.

This thesis found no evidence of a significant increase in residential real estate values over a ten year time period that could be contributed to the construction of Target Field. The thesis also contends that there is no empirical evidence that indicates that any value has been added to local residential real estate in Minneapolis, Minnesota due to the presence of Target Field. It certainly does not find any evidence to suggest that the difference in residential real estate values justify the massive public expenditures used to build this major league sports stadium. Most likely, any increase or decline in residential real estate values has more to do with the value of being close to a downtown area, and the overall state of the local real estate market, and would probably not be any different if the city had not agreed to recently finance and construct a new sports stadium.

This thesis is unique in that it is the only hedonic analysis of the impact of a professional sports stadium on residential real estate values that utilizes the powerful tool known geographically weighted regression (GWR). Therefore, it is the only study of its kind that analyzes local variation of all variables in the study area, to assess their impact in those areas. This method of using GWR to analyze a hedonic real estate model is easily replicated, and can be adapted to any geographic area where real estate sales data is available. The number of independent variables used in the model, and the specific variables themselves are all adaptable, and can be defined by the individual researcher through variable testing. This model can used to analyze the impact of any stadium in any sport. It can also be used to analyze the impact of any large facility that may Impact property values in a positive of negative way, including facilities that may impact the surrounding environment such as factories, large farms, prisons, or virtually any other variable that may affect the value of real estate in an area.

References

Ahlfeldt, G and W.Maennig. 2007. Impact of Sports Arenas on Land Values: Evidence from Berlin, International Association of Sports Economists, Working Paper No. 07-03.

Ahlfeldt, G and G.Kavetsos. 2014. Form or Function? The effect of new sports stadia on property prices In London. Journal of the Royal Statistical Society: Series A (Statistics in Society), 177, 169-190.

Anselin, L. 1988. Spatial Econometrics: Methods and Models. Dordrecht, The Netherlands: Kluwer Academic.

Baade, R.A and V.A. Matheson. 2011. Financing Professional Sports Facilities. College of the Holy Cross, Department of Economics, Faculty Research Series, Paper No. 11-02.

Berg, Steve. 2010. Target Field. Minneapolis, MN: MVP Books. pp. 49–71. ISBN 978-0-7603-3842-1.

Charlton, M.E. and A.S. Fotheringham. 2009. Geographically Weighted Regression White Paper. National Centre for Geocomputation, National University of Ireland Maynooth; March 3, 2009.

Charlton, M.E. and A.S. Fotheringham. 2009. Geographically Weighted Regression; A Tutorial on using GWR in ArcGIS 9.3. National Centre for Geocomputation, National University of Ireland Maynooth.

Coates, D. 2007. Stadiums and Arena’s: Economic Development or Economic Redistribution.

Contemporary Economic Policy; February 2007; p. 565-577.

Coates, D. and B.R. Humphreys. 2008. Do Economists Reach a Conclusion on Subsidies for Sports Franchises, Stadiums, and Mega-Events? North American Association of Sports Economists Working Paper Series, Paper No. 08-18.

Dehring, C.A., C.A. Depken, and M.R. Ward. 2007. The Impact of Stadium Announcements on Residential Property Values: Evidence from a Natural Experiment in Dallas-Fort Worth, Contemporary Economic Policy 25(4): pp. 627-638.

Dortch, S. 1996. The Future of Baseball. American Demographics; Apr 1996, Vol. 18 Issue 4, p. 22-30

ESRI. Interpreting OLS Results. ArcGIS Pro Tool Reference. http://pro.arcgis.com/en/pro-app/tool-reference/spatial-statistics/interpreting-ols-results.html

Feng, X. and B.R. Humphreys. 2008. Assessing the Economic Impact of Sports Facilities on Residential Property Values: A Spatial Hedonic Approach. North American Assoc. of Sports Economists Working Paper Series, Paper No. 08-12.

Feng, X. and B.R. Humphreys. 2012. The Impact of Professional Sports Facilities on Housing Values: Evidence from Census Block Group Data. City, Culture and Society, 3, 189-200.

Feng, X. and B.R. Humphreys. 2018. Assessing the Economic Impact of Sports Facilities on Residential Property Values: A Spatial Hedonic Approach. Journal of Sports Economics, 19 (2), 188-210.

Fotheringham, A.S., C. Brunsdon, and M.E. Charlton. 2002. Geographically Weighted Regression: The Analysis of Spatially Varying Relationships, Wiley.

Hennepin County (MN) GIS. https://www.hennepin.us/your-government/open-government/gis-open-

HF 2480 Status in the House for the 84th Legislature. 2005. www.revisor.mn.gov. Retrieved April 21, 2017.

Johnson, B.K, M.J. Mondello and J.C. Whitehead. 2005. The Value of Public Goods Generated by a National Football League Team. Journal of Sport Management, 2007, 21, 123-136.

Long, J.G. 2005. Full Count: The Real Cost of Public Funding for Major League Sports Facilities. Journal of Sports Economics, 6.2, 119-143.

Matheson V.A. 2004. Economic Multipliers and Mega-Event Analysis. College of the Holy Cross Working Paper Series, 04-02.

Matheson V.A. 2006. Mega-Events: The effect of the world’s biggest sporting events on local, regional, and national economies. College of the Holy Cross, Department of Economics Faculty Research Series, Paper No. 06-10.

Mathews, S.A., T.C. Yang. 2012. Mapping the Results of Local Statistics: Using Geographically Weighted Regression. Demographic Research, 26(6), 151-166.

Mennis J. 2006, Mapping the Results of Geographically Weighted Regression, The Cartographic Journal, 43(2), 171-179.

Minnesota Geospatial Commons. https://gisdata.mn.gov/

New York Times "BASEBALL PLAYOFFS; NOTES". The New York Times. October 4, 1997. ISSN 0362-4331. Retrieved April 19, 2017. https://www.nytimes.com/1997/10/04/sports/baseball-playoffs-notes.html

Radio, Minnesota Public. "MPR: St. Paul Takes Lead in Stadium Push". news.minnesota.publicradio.org. Retrieved April 19, 2017.

St. Paul Pioneer Press "Contraction nearly stole Twins baseball from Minnesota – Twin Cities". July 10, 2014. Retrieved April 21, 2017. https://www.twincities.com/2014/07/10/contraction-nearly-

St. Paul Pioneer Press “County takes title of land, puts Twins ballpark closer to reality”. May 1, 2007. Retrieved February 7, 2018. https://www.twincities.com/2007/05/01/county-takes-title-of-land-

St. Paul Pioneer Press “The day a bridge collapsed in Minneapolis and lives changed forever”. July 29, 2017. Retrieved February 10, 2018. https://www.twincities.com/2017/07/29/the-day-a-bridge-

Tobler, W.R. 1970. A computer movie simulating urban growth in the Detroit region. Economic Geography, 46(2), 234-240.

Tu, C.C. 2005. How Does a New Sports Stadium Affect Housing Values? The Case of FedEx Field. Land Economics, Vol. 81, No. 3 (Aug., 2005), pp. 379-395.

U.S. Census Bureau. American Fact Finder. https://factfinder.census.gov/

U.S. Bureau of Labor Statistics. Inflation Calculator. https://data.bls.gov/cgi-bin/cpicalc.pl

U.S. Bureau of Labor Statistics. https://www.bls.gov/spotlight/2014/housing/home.htm

Appendix

```
#  BALK - A Geographic Analysis of the Impact of a New Professional Sports Stadium on Residential Real Estate Values in a Major U.S. City
#
#  Written By; Anthony M. LaVerde
#
#  A thesis submitted to the faculty of Salem State University in partial fulfillment of the requirements for the
#  Master of Science degree in Geo-information Science in the Department of Geography at Salem State University
#  May 2018
#
#  Python script of the analysis used in this thesis
#
#  INSTRUCTIONS: Install database on C: drive, path should read "C:\\Thesis_Research_Database\\minneapolis_stadium.gdb"
#
#  This script requires ArcGIS 10.5 Advanced or Personal Use License



import arcpy
from arcpy import env
import os


# ASKS USER FOR ANALYSIS YEAR


user_prompt_year = "year"
arcpy.env.workspace = "C:\\Thesis_Research_Database\\minneapolis_stadium.gdb"
ind_points_layer = "sale_points"
ind_points_select = "sale_points_select"
block_groups = "Census_BG_2010"


for year in user_prompt_year:
        user_prompt_year = raw_input("Please select a year between 2006 and 2016:  ")
        if user_prompt_year > "2016":
                print("Data set is for years 2006 through 2016, Please select a year in that range")
        elif user_prompt_year < "2006":
                print("Data set is for years 2006 through 2016, Please select a year in that range")
        else:
                print("Continue on")
                break

print user_prompt_year + " Selected"
agg_poly = "Block_Groups_Sales_" + user_prompt_year
print agg_poly + " " + "File Created"

# SELECTS USER DEFINED ANALYSIS YEAR FROM DATABASE AND CREATES FILE OF ONLY THAT YEARS REAL ESTATE SALES

where_clause = '"SALE_YEAR" = ' + "'" + user_prompt_year + "'"

selection = arcpy.Select_analysis(ind_points_layer, ind_points_select, where_clause)

print user_prompt_year + " Feature Class Retrieved"



### END OF PART 1 SELECTING YEAR
### BEGIN PART 2 AGGRAGATING TO BLOCK GROUP
```

```
# Spatial Join Field Mapping Code
# Averaging Inflation Adjusted Sale Price to 2016 USD (adj_sale), Actual Sale Price (SALE_PRICE),
# Dwelling Living Area in US Square Feet (UNIT_SQFT), and year of sale (SALE_YEAR), which should be constant after selecting a year.

fieldmappings = arcpy.FieldMappings()
fieldmappings.addTable(block_groups)
fieldmappings.addTable(ind_points_select)

avg_sale_index = fieldmappings.findFieldMapIndex("adj_sale")
sale_index = fieldmappings.findFieldMapIndex("SALE_PRICE")
unit_sqft_index = fieldmappings.findFieldMapIndex("UNIT_SQFT")

fieldmap = fieldmappings.getFieldMap(avg_sale_index)
fieldmap2 = fieldmappings.getFieldMap(sale_index)
fieldmap3 = fieldmappings.getFieldMap(unit_sqft_index)

field = fieldmap.outputField
field2 = fieldmap2.outputField
field3 = fieldmap3.outputField


field.name = "adj_sale"
field2.name = "sale_price"
field3.name = "unit_sqft"

field.aliasName = "adj_sale"
field2.aliasName = "sale_price"
field3.alaiasName = "unit_sqft"

fieldmap.outputField = field
fieldmap2.outputField = field2
fieldmap3.outputField = field3

fieldmap.mergeRule = "mean"
fieldmap2.mergeRule = "mean"
fieldmap3.mergeRule = "mean"

fieldmappings.replaceFieldMap(avg_sale_index, fieldmap)
fieldmappings.replaceFieldMap(sale_index, fieldmap2)
fieldmappings.replaceFieldMap(unit_sqft_index, fieldmap3)

# Deleting unneeded and redundant fields

bed = fieldmappings.findFieldMapIndex("BEDS")
fieldmappings.removeFieldMap(bed)
bath = fieldmappings.findFieldMapIndex("BATH")
fieldmappings.removeFieldMap(bath)
state = fieldmappings.findFieldMapIndex("STATE")
fieldmappings.removeFieldMap(state)
county = fieldmappings.findFieldMapIndex("CO")
fieldmappings.removeFieldMap(county)
tract = fieldmappings.findFieldMapIndex("TRACTCE10")
fieldmappings.removeFieldMap(tract)
land = fieldmappings.findFieldMapIndex("ALAND10")
fieldmappings.removeFieldMap(land)
water = fieldmappings.findFieldMapIndex("AWATER10")
fieldmappings.removeFieldMap(water)
acres = fieldmappings.findFieldMapIndex("Acres")
fieldmappings.removeFieldMap(acres)
built = fieldmappings.findFieldMapIndex("YEAR_BUILT")
fieldmappings.removeFieldMap(built)
year = fieldmappings.findFieldMapIndex("sale_yr_long")
fieldmappings.removeFieldMap(year)
```

```
# Executing Spatial Join Individual sale points aggregated to 2010 US Census Block Group polygons

agg_join = arcpy.SpatialJoin_analysis(block_groups, ind_points_select, agg_poly, "#", "#", fieldmappings)


print "aggregated block groups complete"

# Removing Null Values
bg_poly_clean = "bg_poly_clean" + user_prompt_year
rem_null_where = '"SALE_YEAR" = ' + "'" + user_prompt_year + "'"

rem_null = arcpy.Select_analysis(agg_poly, bg_poly_clean, rem_null_where)

print "Null Values Deleted"

# Creating Centroid Points


bg_centroid_points = "bg_centroid_points_" + user_prompt_year
bg_centroid_calc = arcpy.FeatureToPoint_management(bg_poly_clean, bg_centroid_points, "CENTROID")

print "Block Group Centroid Point Feature Class Created"



# DELETING UNNEEDED FILES
print "Deleting Unneeded Files"


arcpy.Delete_management(ind_points_select)
arcpy.Delete_management(agg_poly)
arcpy.Delete_management(bg_poly_clean)

print "Joining Tables"
# JOINING CENSUS TABLES TO CENTROIDS

# Table Variables

centroid_points = bg_centroid_points
bg_id_field = "geoid"
acs_table = "Med_Income_2013"
inc_field = "med_inc"

print "Join ACS 2013 5 year estimate table for median income per census block group data"
# Join ACS Table with Median Income data

inc_table_join = arcpy.JoinField_management(centroid_points, bg_id_field, acs_table, bg_id_field, inc_field)

cen_table = "Population_2010"
pop_field = "u18per"

land_table = "Land_Value_2016"
land_field = "Avg_EMV_LAND"
# Join Census 2010 Table with Under 18 Population Percentage Data

print "Joining Population data from US Census 2010 for percentage of persons under 18 years of age in block group"

pop_table_join = arcpy.JoinField_management(centroid_points, bg_id_field, cen_table, bg_id_field, pop_field)

print "Joining Minneapolis Land Value Table from City Assessor's Office"

land_table_join = arcpy.JoinField_management(centroid_points, bg_id_field, land_table, bg_id_field, land_field)

# CALULATE DISTANCES TO AMENTITY VARIABLES

print "Table Joins Successfully Executed"
```

```
target_field = "Target_Field"
lake_park = "Lake_Parks"
light_rail = "Light_Rail_Stations"
tar_rename_field = "dis_tar"
near_fid_field = "NEAR_FID"
lake_park_rename_field = "lake_park"

near_field = "NEAR_DIST"
lr_rename_field = "dis_lr"

field_spec = "LONG"
# DISTANCE CALCULATION PROCESSING

# Query Expressions
near_exp = "!NEAR_DIST!"


print "CALCULATING CENTROID POINT DISTANCE TO TARGET FIELD"

target_distance = arcpy.Near_analysis(centroid_points, target_field)

target_near_fid_delete = arcpy.DeleteField_management(centroid_points, near_fid_field)
long_tar_field = arcpy.AddField_management(centroid_points, tar_rename_field, field_spec)
long_tar_calc = arcpy.CalculateField_management(centroid_points, tar_rename_field, near_exp, "PYTHON")
double_tar_delete = arcpy.DeleteField_management(centroid_points, near_field)


print "CALCULATING CENTROID POINT DISTANCE TO NEAREST OUTDOOR RECREATION AREA FEATURING LAKE OR RIVER ACCESS"


lake_park_distance = arcpy.Near_analysis(centroid_points, lake_park)

lake_park_near_fid_delete = arcpy.DeleteField_management(centroid_points, near_fid_field)
long_lake_field = arcpy.AddField_management(centroid_points, lake_park_rename_field, field_spec)
long_lake_calc = arcpy.CalculateField_management(centroid_points, lake_park_rename_field, near_exp, "PYTHON")
double_lake_delete = arcpy.DeleteField_management(centroid_points, near_field)

print "CALCULATING DISTANCE TO NEAREST LIGHT RAIL STATION"


light_rail_distance = arcpy.Near_analysis(centroid_points, light_rail)

light_rail_near_fid_delete = arcpy.DeleteField_management(centroid_points, near_fid_field)
long_lr_field = arcpy.AddField_management(centroid_points, lr_rename_field, field_spec)
long_lr_calc = arcpy.CalculateField_management(centroid_points, lr_rename_field, near_exp, "PYTHON")
double_lr_delete = arcpy.DeleteField_management(centroid_points,near_field)


final_centroid_points = "Analysis_Points_" + user_prompt_year
where_2plus = '"Join_Count" >= 2'

centroid_converted = arcpy.Select_analysis(centroid_points, final_centroid_points, where_2plus)

centroid_point_delete = arcpy.Delete_management(centroid_points)
print "CENTROID POINT PROCESSING COMPLETED"

#generating thiessen polygons
print "GENERATING THIESSEN POLYGONS BASED ON BLOCK GROUP CENTROID POINTS"

thiessen_poly = "BG_Thiessen_" + user_prompt_year
thiessen_field_spec = "ALL"


thiessen_poly_generate = arcpy.CreateThiessenPolygons_analysis(final_centroid_points, thiessen_poly, thiessen_field_spec)
```

```
print "THIESSEN POLYGON PROCESSING COMPLETE"

# OLS REGRESSION

print "DATA PROCESSED AND READY FOR ORDINARY LEAST SQUARES REGRESSION"

print "INITIATING OLS REGRESSION PROCESS"

ols_output = "OLS_Results" + user_prompt_year
coeff_table = "OLS_Coeff" + user_prompt_year
diag_table =  "OLS_Diag" + user_prompt_year

ols = arcpy.OrdinaryLeastSquares_stats(thiessen_poly, "OBJECTID_1", ols_output, "adj_sale",
"unit_sqft;med_inc;u18per;Avg_EMV_LAND;dis_tar;dis_lr;lake_park", coeff_table, diag_table)

print "OLS Regression Successful"

gwr_output = "GWR_Results_" + user_prompt_year


print "INITIATING GWR REGRESSION PROCESS"

arcpy.GeographicallyWeightedRegression_stats(thiessen_poly, "adj_sale",
"unit_sqft;med_inc;u18per;Avg_EMV_LAND;dis_tar;dis_lr;lake_park", gwr_output, "ADAPTIVE", "AICc","#", "#",
"#","C:\\Thesis_Research_Database\\minneapolis_stadium.gdb", "#", "#", "#", "#")

print "GWR Regression Successful"

dis_tar_raster = "dis_tar"
dis_tar_rename = "dis_tar" + user_prompt_year
dis_lr_raster = "dis_lr"
dis_lr_rename = "dis_lr" + user_prompt_year
lake_park_raster = "lake_park"
lake_park_rename = "lake_park" + user_prompt_year
med_inc_raster = "med_inc"
med_inc_rename = "med_inc" + user_prompt_year
u18per_raster = "u18per"
u18per_rename = "u18per" + user_prompt_year
unit_sqft_raster = "unit_sqft"
unit_sqft_rename = "unit_sqft" + user_prompt_year
Avg_EMV_LAND_raster = "Avg_EMV_LAND"
Avg_EMV_LAND_rename = "Avg_EMV_LAND" + user_prompt_year
intercept_raster = "Intercept"
intercept_rename = "Intercept" + user_prompt_year

arcpy.Rename_management(dis_tar_raster, dis_tar_rename)
arcpy.Rename_management(dis_lr_raster, dis_lr_rename)
arcpy.Rename_management(lake_park_raster, lake_park_rename)
arcpy.Rename_management(med_inc_raster, med_inc_rename)
arcpy.Rename_management(u18per_raster, u18per_rename)
arcpy.Rename_management(unit_sqft_raster, unit_sqft_rename)
arcpy.Rename_management(Avg_EMV_LAND_raster, Avg_EMV_LAND_rename)
arcpy.Rename_management(intercept_raster, intercept_rename)

arcpy.Delete_management(dis_tar_raster)
arcpy.Delete_management(dis_lr_raster)
arcpy.Delete_management(lake_park_raster)
arcpy.Delete_management(med_inc_raster)
arcpy.Delete_management(u18per_raster)
arcpy.Delete_management(unit_sqft_raster)
arcpy.Delete_management(Avg_EMV_LAND_raster)
arcpy.Delete_management(intercept_raster)


# CALCULATING IMPACT OF DISTANCE TO TARGET FIELD VARIABLE IN THE GWR MODEL
```

```
impact_calc_poly = "Target_Impact_" + user_prompt_year
target_impact = "tar_imp"

target_impact = "tar_imp"
impact_calc_exp = "abs(" + "!C5_dis_tar!" + "*" "!dis_tar!" + ")"
total_target_impact = "tot_tar_im"
tot_tar_exp = "!Join_Count!" + "*" + "!tar_imp!"

# CALCULATING IMPACT OF TARGET FIELD FOR AVERAGE REAL ESTATE SALE IN BLOCK GROUP
print "CALCULATING IMPACT OF TARGET FIELD FOR AVERAGE REAL ESTATE SALE IN BLOCK GROUP"

gwr_thiessen_join = arcpy.SpatialJoin_analysis(gwr_output, thiessen_poly, impact_calc_poly)

impact_calc_field = arcpy.AddField_management(impact_calc_poly, target_impact, field_spec)
impact_calc = arcpy.CalculateField_management(impact_calc_poly, target_impact, impact_calc_exp, "PYTHON")

#CALCULATING TOTAL IMPACT OF TARGET FIELD IN BLOCK GROUP FOR YEAR, BY MULTIPLYING IMPACT DOLLARS BY NUMBER OF
RESIDENTIAL SALES
print "CALCULATING TOTAL IMPACT OF TARGET FIELD IN BLOCK GROUP FOR YEAR, BY MULTIPLYING IMPACT DOLLARS BY NUMBER OF
RESIDENTIAL SALES"

total_impact_calc_field = arcpy.AddField_management(impact_calc_poly, total_target_impact, field_spec)
total_impact_calc = arcpy.CalculateField_management(impact_calc_poly, total_target_impact, tot_tar_exp, "PYTHON")

print "ANALYSIS COMPLETE, PLEASE ADD TARGET_IMPACT FEATURE CLASS TO ARCMAP 10.5 OR LATER TO VIEW DATA"
```