CRISIS IN A SMALL AMERICAN CITY: A SPATIAL ANALYSIS OF RACE, SUBPRIME LENDING, AND FORECLOSURE IN MUNCIE, IN

A THESIS

SUBMITTED TO THE GRADUATE SCHOOL

IN PARTIAL FULFILLMENT OF THE REQUIREMENTS

FOR THE DEGREE

MASTER OF SCIENCE

BY

KATHRYN J. PREIDT

DR. STEVEN RADIL, ADVISOR

BALL STATE UNIVERSITY

MUNCIE, INDIANA

MAY 2016
Table of Contents
Chapter I. Introduction......................................................................................................................... 4
Chapter II: Literature Review ................................................................................................................ 9
Chapter III: Exploratory Data Analysis ............................................................................................... 22
Chapter IV: Exploratory Spatial Data Analysis ................................................................................... 39
Chapter V: Results and Discussion ..................................................................................................... 54
References............................................................................................................................................. 61
Chapter I. Introduction

The apex of the foreclosure crisis was a taxing period of time for many American homeowners who found themselves at risk of losing their homes (Crump 2008). Aguirre and Reese (2014) cite this crisis as the primary cause in the drop of national homeownership from 69 percent in 2006 to 63 percent in 2013. This period of crisis began in 2006 (Aguirre and Reese 2014) and during this year only 1.2 percent of home mortgage loans had begun the foreclosure process compared to 2010 when that number more than tripled to 4.6 percent (US Census Bureau 2012). Approximately 2.5 million houses had been foreclosed between the years 2007 and 2010 (Henry, Reese, and Torres 2013). This crisis reached its peak in 2010 when about every 1 in 45 houses in the U.S. was in foreclosure (Aguirre and Reese 2014). Although the foreclosure crisis has been described as a national issue, the immediate effects of the crisis have been felt locally and directly impacting people’s lives within cities and neighborhoods. For example, because of foreclosures, crimes increased (stripping of pipes, squatting and arson) (Crump 2013), houses became abandoned, and equity was lost (Crump 2008). Many homeowners also found themselves with negative equity, meaning that they owed more on the house than the house was worth (Schulhofer-Wohl 2012). The Federal Reserve estimates that between the years of 2007 and 2009 over $7 trillion in home equity had been lost (Federal Reserve System 2012). Although the market appears to be improving, the impact of this crisis can still be seen in many cities around the country.

Many critics blame the crisis on predatory subprime lending practices by banks and other financial institutions. This is the lending practice whereby banks provide loans to people looking to purchase property who, sometimes from the very start, are not likely to be able to consistently
make the mortgage payments or cannot qualify for lower interest loans. These loans also often contained complex features that borrowers may not have been fully aware of, such as adjustable rate mortgages where the interest rates increase over time (Crump 2008). These subprime loans were often made to people with little or no credit. This practice was not always a large part of the mortgage origination market. According to a report from the Center for Responsible Lending (2006), subprime loans began to become extremely popular between 1994 and 2005 when the subprime market grew from $35 billion to $665 billion. Lenders defended these business practices, justifying the high interest subprime loans due to the risk associated with lending to people with poor or no credit (Crump 2008). In 2006 the rate of foreclosures for subprime loans was 4.5 percent and increased to 14.5 percent in 2010. This can be compared to the rate of foreclosures for prime or conventional loans that rose from 0.5 percent to 3.5 percent during that same time period (US Census Bureau 2012).

When a homeowner defaults on a loan, he or she then enters the foreclosure process. Once a borrower misses the specific number of payments specified by the loan documents, the lender issues a notice of foreclosure to the borrower. If the borrower is unable to sell the house or renegotiate the terms of the loan with the lender, the house is put up for public sale. If the borrower is able to pay back the funds borrowed before the sale of the property, the sale is then canceled and the foreclosure is dismissed. This process can take as little as one month or up to years between the time of the notice and the time when the house is put up for public sale. Once the house is up for public auction, whoever makes the highest bid at that time receives the deed to the house. If there is no bid placed that is higher than the unpaid balance on the mortgage, the institution that holds the mortgage (whether it be a financial institution or another organization that has been sold a packaged mortgage) retains the deed of the house or property. This is a
general explanation of the process but the timeline can vary by state (Ellen, Madar, and Weselcouch 2013). For example, the timeline for this process in Indiana is typically around 260 days (Renwood Realtytrac LLC 2015a).

Before and after the foreclosure crisis, Indiana’s foreclosure trend tended to differ from the overall trend of the nation. Over the past twenty years this pattern was apparent. The foreclosure rate in Indiana during most of the 1990’s was lower than the national mean being at 0.5 percent in Indiana and 0.9 percent in the United States. The foreclosure rate reached its lowest point in Indiana in the mid-1990’s and then rose steadily alongside the state’s unemployment rate. Beginning in the mid-2000’s Indiana’s foreclosure rate remained higher than the national average while the nation’s average decreased to around one percent (Kinghorn 2011).

The Department of Housing and Urban Development created a document in 2009 containing many factors that led to the foreclosure trends in Indiana. According to this document several factors contributed to the crisis, with most of the blame being laid on the amount of subprime lending. Although many researchers believe that this was indeed a major factor in not only Indiana’s housing crisis but the crisis affecting the entire nation, Indiana has some unique factors that separates itself from the average nationwide foreclosure rate. Indiana was once booming with industrial employment but throughout the 2000’s, some 220,000 manufacturing jobs were lost. This surge in high-paying job loss was followed by a lowered median household income. Indiana also had the slowest housing appreciation increase out of any state during this time (Kinghorn 2011).

Indiana’s homeowners run a higher risk of being given a subprime loan than the average American. The percentage of loans deemed subprime in the United States in 2006 was at 13.5
percent compared to Indiana’s percentage of 15 percent. These numbers dropped after the housing crisis but as recently as 2010, Indiana still has a higher percentage (Kinghorn 2011). This study is not focused on the entire state of Indiana but on the city of Muncie in Delaware County, one of the state’s 16 metropolitan statistical areas. Delaware County ranks the fourth highest in rate of foreclosures in the state (Renwood Realtytrac LLC 2015b). When considering the ratio of foreclosures to land area, Muncie is at 0.26% which is much greater than the national average of 0.09% (Renwood Realtytrac LLC 2015b).

The next chapter begins by investigating the different causes of the foreclosure crisis. Subprime lending is then further analyzed and explored with relation to race and income. Other studies are then cited that have researched this same topic of foreclosure, subprime lending, race and income. This chapter also describes other researchers’ studies using GIS technology to further explore the patterns of this crisis.

Chapter Three discusses a method of analysis used called exploratory data analysis or EDA. This begins by describing EDA and giving examples as to how it is carried out. Then a description of the variables used in this study is given. A detailed description of the EDA used is listed as well as multiple figures displaying the trends of data analyzed. This chapter concludes with a brief discussion of the initial findings.

Chapter Four investigates another form of analysis called exploratory spatial data analysis or ESDA. This chapter is structured similarly to the prior chapter as it first discusses the definition and uses of this form of analysis. This analysis is then broken down into four specific methods used in this study. These methods are defined and described in terms of this project. Figures including maps are an important part of this form of analysis and are included displaying the data.
Chapter Five firstly shares the findings of this study and compares the results to the initial expectations based on previous research. These findings are then explained based on this and other studies. Recommendations for future research are given during this section with a brief explanation of possible missing variables. Also presented in this chapter are the limitations and implications of this study.
Chapter II: Literature Review
The Foreclosure Crisis

Research has shown that there are two main reasons why homeowners default on their mortgage: they are either unable to make the payments on their loan or the amount owed is more than the value of the property (Gilderbloom et al. 2012). Certain “trigger” events are likely to cause borrowers to be unable to make the minimum payments. These include sudden illness, unemployment and divorce (Gilderbloom et al. 2012). Robertson et al. (2008) predicts that medical costs account for half of foreclosure cases. However, these unpredictable events in borrowers’ lives are not the only factors that can lead to foreclosure. Quercia, Stegman, and Davis (2007) fault predatory lending of mortgages with high interest rates as the cause of the surge in foreclosures; they found that predatory loans have a 20 percent better chance of entering foreclosure than other loans.

By the year 2006, 22 percent of all US home loans were considered subprime (Crump 2008). Lenders received payments in the form of fees with the initial loan. If a homeowner misses payments on the loan, an unscrupulous lender may suggest multiple refinancing, gaining fees for each refinance (Crump 2008). At any point, the financial institution may package and sell the loan to a Wall Street firm for a considerable profit making subprime lending a very lucrative practice (Crump 2008). Subprime lending reached its peak in 2006 when 47 percent of borrowers in ethnically diverse neighborhoods received subprime loans. On the other hand only 22 percent of borrowers in white neighborhoods received these high cost loans (Gilderbloom et al. 2012). By late 2007, the housing market had plummeted and many homeowners owned a house worth less than what that they owed on their mortgage. Within a year from the start of the crisis, over 3.2 million houses had entered the process of foreclosure (Gilderbloom et al. 2012). Although there are many differing opinions on what ultimately caused the housing market to
crash, many experts believe it was caused by the high volume of subprime loans (Calomiris 2008).

Critics of subprime lending have suggested that many lenders engaged in predatory marketing by targeting minority neighborhoods that were considered underserved by other lenders. These neighborhoods were often African American and Hispanic neighborhoods. This made these predatory practices and their effects racially discriminating (Rivera et. al 2008). For example, Crump (2008) researched who was being targeted for subprime lending in Minneapolis and St. Paul, Minnesota and found that African Americans and Hispanics were nearly four times and twice as likely to receive subprime loans, respectively, when compared to the rest of the population. He also randomly selected 40 properties from minority neighborhoods in order to learn the history of the process. Most of the foreclosed properties started with subprime lending and all of them were refinanced multiple times, some as many as six times (Crump 2013). Crump’s research (2008, 2013) showed that subprime lending is tied to race and that the long standing patterns of race-based housing segregation in urban areas created geographies that were underserved, and therefore, targeted by lenders. Before subprime loans became so profitable for lenders, African Americans were more likely to be denied for a loan altogether. However, white Americans were often times denied loans if they sought them in minority neighborhoods. Because of the initial denial of loans, once the subprime market was established, minority neighborhoods, especially African American neighborhoods, were targeted by lenders (Rugh and Massey 2010).

A large body of research has established the spatial character of subprime lending in large cities. Rugh and Massey (2010) examined racial segregation with regards to the foreclosure crisis in the top 100 metropolitan areas according to census data to learn if racially segregated sections
of cities are more likely to be targeted for subprime lending, therefore having more foreclosed properties. They found that segregation is the strongest variable to determine the variation of foreclosures and is unaffected by any other causes of the crisis, whether economic or not. Bocian, Ernst, and Li (2010) reported that even when the variables of income, loan, and neighborhood were controlled, African American and Latinos still received high cost loans while whites received prime loans according to data for the United States in 2004. Laderman and Reid (2008) researched California metropolitan areas and discovered that African Americans were 3.3 times more likely to be in foreclosure than white borrowers. Latino borrowers in these same areas were 2.5 times more likely and Asian borrowers were 1.6 times more likely to be foreclosed.

Geographer Elvin Wyly conducted extensive research in the area of subprime lending discrimination. Wyly et al. (2006) studied subprime lending in the Baltimore area between the years of 1998 and 2002. They chose a multivariate approach in order to evaluate class-monopoly rent, which is defined as being in the position to have access to the exchange value of land creating a class of people that controls the resource of land and another class of people that pays to use this land. This is evaluated in order to understand the division of the city when controlled income and credit risk which the researchers refer to as a demand side model. They also created a supply side model which evaluates the subprime specialization of lenders. Based on these evaluations the researchers were able to conclude that African Americans have a greater chance of being in the subprime market when compared to whites with similar incomes and debt. Through the demand side model they were also able to uncover that lenders’ relationships with those who invest through them greatly determines segmentation. With these results from the
models a clear conclusion can be made that subprime lending segments the mortgage market so that class-monopoly rents can be obtained from subprime loans.

This same research team also created models to evaluate spatial inequality of subprime lending in Baltimore and calculated an estimate of subprime market division using the supply and demand side models for every loan applicant. These values were then averaged for census tracts and an index was constructed. These values represented how much market outcomes can be connected to borrower-lender relationship characteristics. In tracts where the index value is low a conclusion can be drawn that characteristics on the supply side, which are the tactics of lenders and investors, have more weight than those on the demand side, which are the profiles of homebuyers. The opposite is also true. In Baltimore, high index values are clustered in a very exclusively wealthy neighborhood. The lowest index values appear to be spatially clustered in middle and working class neighborhoods that make up smaller outlying cities. It can be concluded that lenders were targeting homebuyers with relatively low risk profiles. Another conclusion can be drawn from this data that is surprising. Low index values do not appear clustered in lower income and predominately African American neighborhoods.

Wyly and Darden (2010) mapped the likelihood of loan applicants to receive subprime loans in the year 2006 for the entire United States using data from the Home Mortgage Disclosure Act by constructing a standard logistic regression model of subprime and prime lending. The maps display the likelihood of a Hispanic or African American homebuyer to receive a subprime loan when compared to a white person of non-Hispanic origin with similar economic circumstances and shows that the area of highest racial disparity between African Americans and whites of non-Hispanic origin is in urban counties of the Midwest such as Detroit and Chicago. Ranking second in terms of this racial disparity of subprime lending are populous
urban counties in the South. These disparities are much lower in the West but it should also be noted that the West is far less populated by African Americans. The maps show that the area of the most severe disparities between those of Hispanic origin and those who are white with non-Hispanic origins is on the East coast. The West has fewer disparities and they are less extreme than other regions. Overall, the disparities are more moderate for those of Hispanic origin compared to African Americans.

Wyly et al. (2007) researched the role of ethnicity and gender in the mortgage market using qualitative methods by distributing a questionnaire to a sample of people participating in a HUD first-timer homebuyer class. This questionnaire asked simple questions about the participants’ gender and race. It asked the question “What do you think this form [questionnaire] is used for?” The responses varied greatly from no response at all to racial profiling. When asked if the form’s racial categories were suitable to identify them, most respondents agreed that they were. These answers do not represent the current amount of nondisclosure of the items as shown in data from the Home Mortgage Disclosure Act. These results seem to implicate lenders, implying that they are not disclosing this information even if those taking out the loans provide this information to the lenders. This is a way for lenders to escape the label of predatory by not disclosing the race or ethnicity of applicants. The researchers also created a model to measure the likelihood of nondisclosure. Their results were that applicants with a bad credit profile are 1.488 times more likely to not disclose this valuable information than those with average or good credit profiles. This is consistent with prior research that people who have bad experiences in the financial market will be less likely to disclose any of this type of information. This lack of disclosure is creating what researchers call “a geography of invisibility” making it difficult to accurately determine how extensive racial or gender discrimination is in the mortgage market.
Wyly and Holloway (2001) published a study on the denial of home mortgage loans in the Atlanta region using data from the Home Mortgage Disclosure Act to sample application characteristics for the purchase of owner occupied property. They created logistic regression models to examine individual and neighborhood-level factors in lending. One model showed that the probabilities for denial were much higher for anyone who was not white. They also found that a single female applicant was less likely to be denied a loan than a single male applicant or a female applicant with a co-applicant. Loan denial was also less likely for those who lived in a high income census tract. The impact that race makes in the denial of loans is also much lower in minority neighborhoods when compared to predominantly white neighborhoods. They found that black applicants with median incomes seeking a median loan amount are more likely to be denied the loan than white applicants with the same income and loan amount. Wyly and Holloway’s study showed that minorities are spatially tied to the same neighborhoods and that loan denial for middle-class and upper-class neighborhoods for African Americans makes it nearly impossible for them to move to a better area.

Schintler et al. (2009) studied the spatial distribution of foreclosures in the New England region and found that there is a very high degree of spatial association, meaning that many foreclosed homes are near each other. They also discovered that in the year 2000, most foreclosure-laden neighborhoods were inhabited by young large families. Since 2000, they have found that foreclosures have increasingly spread to black and Hispanic neighborhoods. Delgadillo and Pederson (2007) studied loan defaults at the census tract level. They discovered that the higher default rates in non-white census tracts are due to preexisting economic disadvantages. One study examined the theory that foreclosures cause people to relocate to different cities. Case (2011) studied Pinellas County in Florida and found that foreclosures
increase the amount of out migration from the area, which in turn decreases the amount of funding (state and federal) that area receives. This can cause an area like Pinellas County, Florida to be in a perpetual depreciating economic spiral.

Using GIS to Examine Foreclosures

The field of geographic information systems, or GIS, supports the spatial analysis of geographically located data on a variety of topics (Ormsby et al. 2010). The capabilities of GIS make it ideal to spatially analyze foreclosures. For example, Morrow-Jones et al. (2005) researched property transactions using a form of forensic GIS that is used in order to monitor spatial data to find abnormalities that compromise the law or code of ethics. Their specific research did not target violations of unethical practices like predatory lending in the field of foreclosures, but the rapid transaction of high priced properties between small groups of people in order to launder money. This process is called “property flipping.” The procedures to identify where GIS leads them to believe property flipping is occurring was used to help expose practices of unfair or predatory lending as well.

Foreclosures resulting in vacant dwellings are strongly linked to increases in crime. Teasdale et al. (2011) used GIS to evaluate the association between foreclosures and crime using the count feature in different census tracts of Akron, Ohio. These researchers used data from the Akron police department to determine crime counts based on how many calls and arrests were made per census tract. Through spatial analysis it was clear that the tracts with the most crime had higher numbers of foreclosures while the opposite was true for tracts with a low volume of crime. The results show that for every subprime loan foreclosure, the crime count increases by 2-3 percent. The authors conclude that if subprime loans were eradicated, crime counts would fall by 40 percent (Teasdale, Clark, and Hinkle 2011).
Leonard and Murdoch (2009) used GIS to study the relationship between foreclosures and the quality of neighborhoods in order to create a model that reveals the impacts of foreclosures on a local housing market. This is an interesting study because the quality of a neighborhood cannot be quantitatively observed as median income could, for example. This research team used housing data in parcel-based GIS format, and then geocoded the foreclosure data to the parcel file. Race, ethnicity, age and average household were all controlled for. GIS was used to calculate the number of foreclosures within certain distances of homes for sale and was added to a spatial autoregressive model. The researchers concluded that a sale price depreciates if there is a property in foreclosure within 250 feet.

Kobie and Lee (2011) researched the effects of foreclosures on the values of surrounding properties in Cuyahoga County, Ohio. Their two hypotheses were that foreclosures have a negative impact on the sale prices of surrounding homes and that the longer a property is in the foreclosure process the more negative an impact it will have on home prices. The researchers divided up their study area into face blocks which are areas from intersection to intersection and were used to focus on the visibility of a foreclosure to surrounding neighbors. GIS was used to calculate neighborhood variables in order to add them into the regression equations. Three regressions were run on the data for the entire county, the city of Cleveland, and the suburbs of Cleveland. Kobie and Lee found that for the entire county and the suburbs foreclosures do have a negative impact on sale prices. They also found that sales prices are dependent on the amount of time a property has been in foreclosure in the entire county. The city of Cleveland showed no impact on sales price based on the two hypotheses.

Biswas (2012) also studied the impact of foreclosures on housing prices in Worcester, Massachusetts. He hypothesized that the spatial distance of foreclosed properties to properties
for sale negatively affects the sales price but these effects may differ between single family or multifamily types of housing. He also hypothesized that an abundance of foreclosed properties may change the supply and demand curve for an entire community. A data set was used with basic characteristics of homes such as the number of bedrooms and bathrooms as well as the sale price. This data was added to GIS and combined with a data set of crime rates. Biswas chose a hedonic regression model to test for the negative impacts of foreclosures. To include GIS data in this model, two different buffers were drawn with radii of 660 feet and 1320 feet in order to have a count of the foreclosures in the area. He finds that, as expected, a larger lot size and a higher number of bedrooms all increase the sales price of a property. Biswas found that all types of housing under foreclosure negatively impacts the sales price of houses nearby as well as prices in the entire community.

Crump (2013) researched and analyzed subprime lending in the Minneapolis-St. Paul area. He mapped subprime loans and found them to be very unequally dispersed. Subprime loans were highly concentrated in minority communities, specifically African American, Hispanic and Hmong neighborhoods. Crump also mapped the spatial distribution of foreclosures that occurred in 2007. The foreclosure trend of this map was very similar to that of the first map showing the distribution of subprime lending; the foreclosures were highly concentrated in the same African American, Hispanic, and Hmong communities.

It is known to many lenders that geographic location is the best indication of his or her likelihood for credit risk (Crump 2008). Can (1998) states that GIS is a perfect tool for lenders to use to locate regions of high risk. This is of course meant for lenders to make the best decisions for themselves as well as the borrowers by spatially analyzing areas of risk. Although written in 1998 before the burst of the housing bubble and unethical lending practices were brought to the
forefront in the media, this article inadvertently raises the issue of lenders using GIS to better organize predatory lending behavior. GIS is an important tool for analysis of areas based on census data and other demographics. However, this also seems like an ideal instrument for someone in the position of a lender to engage in unscrupulous practices.

**Research Questions**

The literature reviewed points toward a clearly identified finding. This is the targeting of minority neighborhoods by subprime lending programs which, given the spatial segregation of minorities in US cities, lends to a clear expectation about where foreclosures should be expected. However, all prior research on this topic has been conducted in larger cities. Because of this gap in current research, the city of Muncie, Indiana will be studied. By choosing a smaller sized city this study will be able to conclude if this link shown in prior literature remains true regardless of city size. Although Muncie is defined as metropolitan by the Office of Management and Budget (2013) it is still much smaller in size and population than the other cities in which similar studies have been conducted. For a city to be defined as a metropolitan statistical area by the Office of Management and Budget it must only have a population that exceeds 50,000. However, there are four different levels to which a metropolitan area can belong to and Muncie falls in the lowest level which is Level D and includes any city that has a population below than 100,000 inhabitants. The larger cities mentioned in the literature review fall squarely in Level A (Minneapolis-St. Paul) or Level B (Baltimore, Atlanta). So although Muncie is classified as metropolitan, it clearly is not in the same category as larger cities previously mentioned. Do the findings for Level A and B cities also hold for Level D cities?

The main research question focused on by this study is the spatial relationships between subprime lending and minority and low incomes areas of Muncie over the same time period. This
question can help clarify if the early findings described above apply to a different type of city in the urban hierarchy and improve our collective understanding of the dynamics of the crisis in a particular type of city.

**Data**

This study is set within the city of Muncie, Indiana. Reported from the US Census, Muncie is a city of 70,316 people and is located in Delaware County within East-Central Indiana. Muncie was once seen a symbol of prosperous middle-class American society through its association as the site of the ‘Middletown’ sociological studies in the 1920s and 30s (Lynd and Lynd 1929; 1937). Like so many other small cities in Indiana and neighboring states within the so-called ‘rust belt,’ Muncie’s economic fortunes have rapidly declined following capital disinvestment and the departure of once-numerous manufacturing jobs since 1980 (Fraser 2012).

Muncie’s shifting economic fortunes have led to numerous local efforts to revitalize and redevelop areas of the city and to stabilize neighborhoods seen as blighted by the presence of abandoned houses during the foreclosure crisis (Radil and Jiao 2015). Race is another important dynamic within Muncie and the race- and class-based residential patterns first identified by the Lynds continue to divide the city functionally and perceptually (Mitchelson, Alderman, and Popke 2007; Radil and Jiao 2015). The percentage of non-whites in Muncie is 17.2 percent which closely mimics the state’s percentage of minorities at 18.5 percent (United State Census Bureau 2015). However, like most US cities, Muncie has a high degree of residential segregation by race.

The combination of the clear patterns of racial segregation and the broader economic challenges facing the city make this an ideal site to explore the spatial patterns of foreclosures in
a smaller city. Another important factor in site selection is that the crisis has persisted in Muncie. For example, one in every 378 houses in Muncie is currently in foreclosure compared to the state average of one in every 762 (Renwood Realtytrac LLC 2015b). Because of the prevalence of foreclosures even after the housing bust, Muncie is an interesting place to apply GIS to understand trends of foreclosures and subprime lending.

Foreclosure data has been gathered from Delaware County’s Sheriff Department as well as the Doxpop website. These two sources house foreclosure data with the address and financial institution from which the mortgage loan was received. This made it possible to add georeferenced points for every foreclosure as well as flag certain institutions as subprime based on the subprime lender list developed by the U.S. Department of Housing and Urban Development. The demographic data of the percent of non-white residents as well as median income was found through the U.S. Census Bureau on the American FactFinder webpage. These data sources are further explained in the next chapter.

Methods

This project employed several interrelated methods. First, data from the public foreclosure records was georeferenced in a GIS using address matching to facilitate an exploratory spatial data analysis, or ESDA (Anselin 1999). Apart from the location of foreclosures, spatial patterns of key variables such as prime or subprime status and income and race measures were explored and visualized. Second, the exploratory data analysis tests of statistical relationships between quantitative variables where possible; for example, scatterplot and correlation analysis were performed on measures of the proportion of subprime foreclosures and demographic variables. Third, four different techniques of ESDA were conducted. The first
ESDA technique used was global Moran’s I. This test explores a variable’s spatial dependence and determines if values are clustered, random, or dispersed. The next technique used was local Moran’s I. This test examines where clusters and outliers are located and creates several cartographical outputs describing these clusters or outliers. The third technique was a geographically weighted regression which creates a regression model for each feature predicting the dependent variable based on the independent variable(s). The final technique was the location quotient. This test determines the spatial concentration for each census block group. Through these methods the relationships between subprime lending, income, and minority neighborhoods were extensively evaluated.
Chapter III: Exploratory Data Analysis

Exploratory data analysis (EDA) can be defined as the first stage of data analysis used in order to begin to understand a dataset, explore relationships between variables, and develop ideas for future investigation of the data (Shelly 1996; Behrens 1997; Cox and Jones 1981). However, EDA can also be described as not just a method to analyze data but also an attitude toward data (Tukey 1977). This attitude or philosophy of EDA is an openness towards the story that the dataset can tell instead of making assumptions of a type of statistical technique that should be employed (Behrens 1997). Tukey’s (1977, 1-3) words describe the importance of EDA best, “Exploratory data analysis can never be the whole story but nothing else can serve as the foundation stone- as the first step.”

EDA largely involves the use of descriptive statistics and graphical visualizations to answer questions about trends of variables (Behrens 1997). EDA is highly dependent on graphs, such as histograms, in order to quickly visualize the peaks or shape of each variable in a dataset. Graphical analysis is important because it is allows for quick visualization by displaying different variable observations at once (Behrens 1997) and can be used to check prior assumptions about the data (Cox and Jones 1981). Technology is taking an important role in modern EDA. When Tukey first explained EDA in 1977, he suggested using colored pencils and graph paper. With the advancement of modern technology, researchers now have powerful computer programs that can process a large amount of data and produce high resolution displays of this data making EDA more accessible (Shelly 1996).

According to Myatt (2007) exploratory data analysis involves four main parts, the first of which is defining the problem in sufficient detail. This involves not only the problem that is to be
solved but also projecting the deliverables of the project as well as a plan for analysis. In order to provide correct expectations from the project contributors and the stakeholders, these deliverables should be defined early on. This step is crucial because it will help focus and expedite the analysis process. A few important parts of this step are to develop objectives, create a timeline, and discuss expectations for those working on the project as well as any others who hold stock in the project. A broad goal or problem should be selected and then further broken down into manageable pieces or objectives. This step should also include any positive or negative implications of the project. This was an important step of this research project because these steps assist in developing and the creation of a plan for the execution of research and analysis.

The second part of an EDA is preparing the data. This is the collecting, cleaning, and transforming of the dataset(s). In many cases, this is the most time consuming part of the project. The data must be collected from trusted sources and then cleaned up in order to prepare it for analysis. The cleaning and collecting of data involved finding a historic list of foreclosed properties and transforming that into an organized excel file. A key way to visualize data at this point is to create a frequency distribution as a histogram. This data can either have a normal or skewed distribution reflecting whether parametric or nonparametric statistical tests are appropriate. Sometimes a dataset must be transformed in order to meet the distributional assumptions of certain statistical tests. For instance, normalization refers to a mathematical operation that changes the original values of a variable to a new range and is a common type of transformation used when a dataset must have a normal or near-normal distribution in order to meet the assumption of many types of parametric statistical tests.
The third part of an EDA involves choosing the appropriate form of analysis. First, a data summary is required, which can be explained as the initial story that the data tells without any in-depth analysis. The next step is the uncovering of hidden relationships that are not obvious from a general summary of the data. This can involve graphs or descriptive statistics. Data tables are a common way to display data in an easy to access form. Scatterplots are important tools to use in this step to find a relationship between variables. After these steps, three types of statistics should be used on the data set: descriptive as a way to quantify a summary of the data; inferential as a way to make confident statements about a large population; and comparative as a way to realize relationships between variables. The final piece of exploratory data analysis is the deployment of the results in either written or graphical form. This is an important piece of the project because a written report or other deliverable is how all of the work from the project will be translated for others to learn from.

One disadvantage of exploratory data analysis is that it sometimes can allow a researcher to perceive a pattern that does not truly exist, one that Shelly (1996) likens to a person looking at an inkblot. Shelly suggests that a supplement to EDA should always be confirmatory data analysis (CDA) in order to check the significance and confirm that a perceived pattern is accurate. Tukey (1980) also explains that neither CDA nor EDA is adequate alone. Smith and Behrens (1996) make the comparison that EDA is similar to investigative work because it is explored with hunches deducted from it. These intuitions should then be put to the test in order to verify them using CDA. This is different than EDA because CDA is executed with a goal in mind, a null hypothesis to accept or reject (Tukey 1980). In the case of this study, the null hypothesis is that there is no relationship that ties race and income to subprime lending.
Variable Description

Foreclosure data was primarily gathered from the Delaware County Sheriff’s Department. These foreclosures were collected through the Sheriff’s property sale information. A Sheriff’s sale is a monthly auction of properties that the court has determined are foreclosed and any person or company that plans to bid on one of these properties must be able to pay in full that day. This successful bidder also becomes responsible for any outstanding taxes or liens on the newly purchased property. Records from the Sheriff’s sales included the street address of the property, the names of the plaintiff and defendant, the amount of the judgment against the defendant, and the bid amount for the property. These records ranged in date from August 2009 to December 2014 (Liberman Technologies 2015).

A secondary source of foreclosure information was from public records on the Doxpop website that contains a variety of public records for most of Indiana. The purpose of this website is to help the public easily obtain court records, although some subscriptions for the website include a fee. The Delaware County Clerk’s office enters their records into this database and recommends it to people conducting research. The foreclosure court records are very similar to the sheriff sale records and include the lender, the borrower, the borrowers address, and the filing date. This source was used to access data missing from the sheriff’s department’s records from January 2008 to August 2009. This time frame is an important variable because of the timing of the foreclosure crisis which many experts agree began in 2006 (Aguirre and Reese 2014).

In order to make the determination whether the loan of the foreclosed properties was subprime or prime, information from the U.S. Department of Housing and Urban Development (HUD) was used. The goal of HUD is to “create strong, sustainable, inclusive communities.” To adhere to this goal HUD has created a list of lenders that have specialized in subprime loans for
over a ten year time period. This product is created by accessing information lending institutions provide through the Home Mortgage Disclosure Act (HMDA). This act was passed in 1975 and requires financial institutions to publically disclose mortgage information to monitor the integrity of said institutions. The HMDA indicators that HUD used in order to create this list were the origination rates of loans (subprime lenders typically have lower origination rates), shares of home refinance loans (subprime lenders typically have higher shares of these loans), and the percentage of the lender’s portfolio sold to a Government Sponsored Enterprise (most subprime lenders do not sell a high percentage of their portfolios to GSEs).

From the Sheriff’s sale and other records of foreclosures in Muncie, the foreclosures that were mortgaged through subprime institutions based on HUD’s list were flagged. ArcGIS software was used in order to categorize both prime and subprime foreclosures into their proper census block group in order to compare them with demographic data. Because the records of foreclosures for Muncie include addresses, they were able to be geocoded as points (Figures 1 and 2). After the geocoding process these points were assigned a new attribute of block group. The counts of subprime foreclosures and total foreclosures were taken for each block group in order to create new variables to analyze. This is an important step to move towards the spatial understanding of foreclosure and subprime lending relationships.
Figure 1: A map of Muncie’s census block groups displaying all 1,411 foreclosures from 2008 to 2014. These foreclosures are widely dispersed throughout the city with a slightly higher concentration in the southern portion.
Figure 2: A map of Muncie's census block groups displaying all 267 foreclosures that began with subprime loans from 2008 to 2014. Like the previous map these foreclosures are also distributed widely throughout the city with a slight concentration in the Industrial and Southside neighborhood. From this map it is obvious that there are few subprime loans.
Demographic data was gathered from American FactFinder, a website that houses data that is collected through the U.S. Census Bureau. The demographic data needed for this study is collected at the census block group level for three variables: median household income, percent African American, and percent Hispanic (Table 1, Figure 3). The variable of median household income was chosen because the dataset is not symmetrical, meaning that the data does not create a perfect bell curve when graphed. Because of this, mean household income would be an inappropriate choice considering the distribution of the data; a few extreme outliers could skew the results of the mean value. This information is given in 2013 inflation adjusted dollars from the American Community Survey and the percent African American and percent Hispanic origins are also from the American Community survey. These numbers are added together in order to evaluate the percent of residents categorized as non-white.

<table>
<thead>
<tr>
<th>Source</th>
<th>Median Household Income</th>
<th>Percent Non-white</th>
<th>Total Count of Subprime Loans</th>
<th>Percent of Subprime Loans</th>
<th>Subprime Ratio</th>
</tr>
</thead>
<tbody>
<tr>
<td>Mean</td>
<td>31348.74</td>
<td>17.68</td>
<td>5.06</td>
<td>0.19</td>
<td>0.01</td>
</tr>
<tr>
<td>Median</td>
<td>28255</td>
<td>12.3</td>
<td>4</td>
<td>0.19</td>
<td>0.01</td>
</tr>
<tr>
<td>Variance</td>
<td>240909143</td>
<td>342.78</td>
<td>17.85</td>
<td>0.03</td>
<td>0</td>
</tr>
<tr>
<td>Std. Dev.</td>
<td>15521.25</td>
<td>18.514</td>
<td>4.225</td>
<td>0.174</td>
<td>0.011</td>
</tr>
<tr>
<td>Skewness</td>
<td>0.96</td>
<td>2.004</td>
<td>6.22</td>
<td>2.823</td>
<td>1.028</td>
</tr>
</tbody>
</table>

Table 1: A summary of the variables, their sources, and some descriptive statistics for each from the 69 observations.
Figure 3: (A-E) Frequency histograms displaying the distribution of each variable.
Data Analysis

Through the beginnings of using the principles of EDA to analyze the data, the relationships between subprime lending, minority neighborhoods, and median household income become clearer. To begin the analysis many scatterplots were created comparing different data variables. Scatter plots are a widely used tool for representation of the relationship between two variables. The strength and direction of any association between data points are the two main pieces of information gathered from scatterplots. The direction of the association can be positive, negative, or have no association and the strength can be strong, moderate, or weak (McGrew and Monroe 2000).

First, the number of subprime loans and the percent of non-white inhabitants and median income per census block group were examined. Figure 4a shows almost no relationship between percent non-white and count of subprime foreclosures. Although figure 4b shows a slightly stronger relationship than the previous figure, there is still almost no relationship between these variables. As discussed in Chapter 1, prior research indicates that the variables of percent non-white and subprime counts should be more strongly related. Also, based on some of the previous studies, the relationship between the number of subprime loans and median income should have a negative association showing that the lower the median household income is, the more likely it is to see a concentration of subprime loans.
Figure 4: Total count of subprime foreclosures per census block group compared to (A) percent of non-white residents per census block group and (B) median household income in 2013 inflation adjusted dollars per census block group.
Instead of using the total count of subprime loans, similar scatterplots were created using the percent of subprime loans that led to foreclosure. This was calculated by taking the count of subprime loans that led to foreclosure per census block group out of the total number of foreclosures compared to only the total count of subprime loans that led to foreclosure used before. The first scatter plot comparing this variable to the percent of non-white inhabitants is similar to Figure 4a because it is also has almost no relationship. The plot comparing the median household income to the percent subprime variable is also very similar to Figure 4b showing a very weak and positive association, and therefore, is still not in agreement with prior research.

This disagreement between this data and the previous research conducted on similar data led to the idea to remove the zero values from this dataset. Not only did this not display the relationship expected for median income as with all of the former plots so far, it also changed the association from positive to negative with the percent non-white variable, meaning that the more
white a neighborhood is, the more likely it is to find a concentration of subprime loans in that area. A different variable was then created to compare the median income and percent non-white to a measure of subprime lending. This new variable, referred to as the subprime ratio, is the count of subprime loans divided by the total count of households in a census block group. This new variable yielded the expected results based on similar studies. Figure 5a shows the subprime ratio variable plotted with median household income. This displays almost no relationship whatsoever. Figure 5b also shows the expected relationship with the ratio and percent of non-white inhabitants. This new plot displays a positive and but still weak association. This weakly positive association shows that subprime lenders may be targeting areas of minority more so than areas with less diversity.
Figure 5: Ratio of the number of subprime foreclosures out of the total number of households per census block group compared to (A) median household income in 2013 inflation adjusted dollars per census block group and (B) percent of non-white residents per census block group.
Another variable was chosen to analyze. This variable is the number of block groups that had no subprime foreclosures. Based on prior research this lack of subprime foreclosures should be clustered in high income and mostly white areas. The values for the percent non-white variable and median income variable were divided into quartiles to be evaluated. Table 2 displays the results of this analysis. There were a higher number and percentage of block groups with no subprime loans in the first quartile of percent non-white values. This quartile ranged from 0%-2.8% non-white. This reflects previous research because it shows that in Muncie block groups with no subprime foreclosures are more common in whiter areas. When comparing block groups with no subprime foreclosures to median income the number of block groups without subprime foreclosures are almost distributed evenly throughout the quartiles. This shows that median income may not play as big of a role in determining subprime lending patterns in Muncie as it did in other cities.

<table>
<thead>
<tr>
<th>Quartile</th>
<th>Number of Zero BGs</th>
<th>% of Zero BGs</th>
<th>% of BGs with Zeros</th>
</tr>
</thead>
<tbody>
<tr>
<td>1st quartile</td>
<td>11</td>
<td>47.83%</td>
<td>42.31%</td>
</tr>
<tr>
<td>2nd quartile</td>
<td>4</td>
<td>17.39%</td>
<td>16.00%</td>
</tr>
<tr>
<td>3rd quartile</td>
<td>5</td>
<td>21.74%</td>
<td>20.00%</td>
</tr>
<tr>
<td>4th quartile</td>
<td>3</td>
<td>13.04%</td>
<td>12.00%</td>
</tr>
</tbody>
</table>

Table 2: Census block groups that have zero subprime foreclosures are compared to percent of non-white residents per census block group and median household income in 2013 inflation adjusted dollars per census block group.
DA was an informative step of this research project as these scatterplots were able to confirm and deny some preconceived ideas about the relationships of these variables. All of the relationships between variables were far too low to be considered significant. Some of the relationships were an opposite direction of what was predicted based on the previous research. Although this step was useful, it is impossible to make the determination whether to accept or reject the null hypothesis of this study, being that there is no relationship between race, income, and subprime lending. This process did lead to the finding of the subprime ratio and that variable’s relationship with percent non-white and median income. Although this variable also showed only a slightly negative relationship with median income, it was the only variable with this direction of association. This variable also showed a slightly stronger positive relationship with percent non-white. Although these findings are not remarkable by any means, they are pointing toward using this variable for further research. This step ultimately led to the decision to try spatial analysis instead of a standard regression analysis in order to accept or reject the null hypothesis.
Chapter IV: Exploratory Spatial Data Analysis

Although EDA is an important step in any analysis of data, in this case it did not provide the expected relationships between variables. Because each variable is linked to a census block group, a spatial analysis approach is a better fit for this project. Spatial analysis is a powerful and unique way to view and analyze this dataset. Although mapping this data is important for a spatial understanding and visualization, mapping alone will not tell the entire story of the data. Exploratory Spatial Data Analysis (ESDA) is a branch of EDA focusing specifically on how variables relate spatially and is required to quantitatively explore the spatial relationships of data. The ESDA approach to analyze data is important to detect spatial patterns and processes that would otherwise remain unseen if only using EDA (Haining, Wise and Ma 1998).

ESDA techniques calculate important values to assess spatial heterogeneity and spatial dependence as well as produce new maps in order to visually analyze data (Anselin 1999). These two spatial concepts can be challenging to differentiate. The concept of spatial dependence can be tied to the First Law of Geography identified by Tobler (1970) stating that the closer things are in space, the more closely related they are. Spatial dependence is the concept that values will cluster in space because of a specific process. An excellent example of this concept is how housing prices are similar in the same neighborhood. Anselin and Getis (1992) reference that the lesser size of the spatial unit is inversely proportional to the spatial dependence of the variables stating that lesser the size of the unit, the greater the chance that neighboring units are spatially dependent. These researchers also suggest that longer and narrower spatial units are more likely to have higher spatial dependence with neighboring units than those units that are more compact.
Spatial heterogeneity is the concept that a process is not distributed evenly across space. Unlike spatial dependence this concept is not the clustering of values of one process but the dispersion of these values in a given area. Spatial heterogeneity can be considered a special case of spatial dependence when the clustering of spatially dependent variables are not clustered evenly throughout space thus making it possible for spatial heterogeneity to be a result of spatial dependence (Anselin and Getis 1992). Relating back to the example of high housing prices being near and dependent on other high housing prices as an example of spatial dependence, if these clusters of high prices are not distributed evenly throughout a city, this is an example of spatial heterogeneity.

ESDA can be divided into two categories of analysis. The first of these is global or whole map statistics which calculate properties of spatial data based on the entire geographic extent of this data. The second category involves focused or local statistics which analyzes spatially defined subsets of the data (Haining, Wise and Ma 1998). Local statistics are an important function of ESDA because some trends may be marginalized at a global scale when at a local scale they are prominent (Josselin 2003). Furthermore, ESDA is made up of many different techniques that fit into each category. Four different ESDA techniques were used in order to calculate spatial relationships between variables. Using these ESDA techniques is an important form of analysis for this dataset because of the spatial ties each variable has.

The first ESDA technique used to detect possible patterns was a measure of global spatial autocorrelation. This technique measures the amount of spatial dependence variables have. The most popular measure of global spatial autocorrelation is Moran’s I (Anselin 1999). This inferential statistic tool measures the point feature’s value as well as location and indicates whether the values of the points are clustered, dispersed, or random as a Moran’s I index value.
that ranges from negative one to positive one (ESRI 2013). A positive value indicates that areas close together exhibit similar values while negative values indicate areas close together exhibit dissimilar values. Variables with an autocorrelation value of zero indicate a completely random pattern and can be described as not possessing spatial autocorrelation (McGrew and Monroe 2000).

The formula for Moran’s I is \( I = \frac{n}{S_o} \frac{\sum_{i=1}^{n} \sum_{j=1}^{n} w_{i,j} z_i z_j}{\sum_{i=1}^{n} z_i^2} \) where \( n \) is equal to the total number of features, or in the case of this study census block groups, \( w_{i,j} \) is the spatial weight between feature \( i \) and feature \( j \), \( z_i \) is the deviation of the attribute of feature \( i \) from its mean, \( z_j \) is the deviation of the attribute of feature \( j \) from its mean, and \( S_o \) is the aggregate of all the spatial weights. This tool also calculates the p-value and z-score in order to assess the statistical significance of the outcome (ESRI 2013).

A Moran’s I test was performed on each subprime variable (count of subprime foreclosures, percent of subprime foreclosures, and subprime ratio). Each variables’ Moran’s I result was evaluated for six of the different conceptualizations of spatial relationships to check for robustness of results. This process was executed because how a neighborhood is defined may impact the results of this test. The goal was to find which variables remained highly clustered no matter how a neighborhood was defined. The first neighborhood conceptualization was the inverse distance. This conceptualization means that neighbors impact the value of other neighbors but the closer neighbors are, the more impact they have. The next conceptualization was the inverse distance squared. This is similar to the previous conceptualization but the slope is much sharper so closer neighbors have a much more significant impact than neighbors further away. The next conceptualization was a fixed distance band. This means that there is a specific
distance where neighbors have an impact on the value. Outside of this band, there is no influence. Next was the zone of indifference conceptualization. Like the previous conceptualization this one has a specified distance where the neighbors have an impact but outside of that zone the impacts drops off quickly. The last two conceptualizations are very similar. These are polygon contiguity conceptualizations and calculate the impact of neighbors that either share an edge or an edge along with a corner. These are the only neighbors that impact the value.

The percent of subprime foreclosures remained completely random in every conceptualization while the count of subprime foreclosures varied from clustered to random. Only the subprime ratio variable remained highly clustered for each separate conceptualization, so this was the variable chosen to investigate further. Figure 6 shows the output of the Moran’s I test for the inverse distance conceptualization for the subprime ratio. This figure displays that there is only a very small chance that the clustering pattern could result from random chance. In other words, the test indicates that, as a global property of the subprime mortgage ratio variable, similar values tend to cluster together in geographic space in the study area which suggests that areas of clustering are subject to similar processes. However, the test does not identify the specific areas where such clustering occurs. For that, a separate test is needed.
The next ESDA technique performed was a localized cluster and outlier analysis or local Moran’s I test. This is also a test of spatial autocorrelation and is based on the global Moran’s I statistic. As the name suggests this tool identifies clusters as well as outliers in a dataset. Unlike global Moran’s I, this test is performed at a local level and is therefore capable of finding areas of clusters that are not described with a global analysis (ESRI 2013). The formula for local Moran’s I test is $I_i = \frac{x_i - \bar{x}}{S_i^2} \sum_{j=1, j\neq i}^n w_{i,j} (x_j - \bar{x})$ where $x_i$ is an attribute’s value for feature $i$, $\bar{x}$ is the mean
of that attribute, $w_{i,j}$ is the spatial weight between those features, and 
\[ S_i^2 = \frac{\sum_{j=1,j\neq i}^{n} (x_j - \bar{x})^2}{n-1} - \bar{Y}^2 \]
where $n$ is the number of features and $x_j$ is an attribute for the feature $j$ (ESRI 2013).

Figure 7 shows the mapped results of this analysis. The clusters can show spatially significant high-high clusters, meaning that high values are close to other high values, low-low clusters, meaning that low values are close to other low values, high-low outliers, meaning that high values are surrounded by low values, and low-high outliers, meaning that low values are surrounded by high values (ESRI 2013). Like the global Moran’s I technique, this technique too has different ways to define a neighborhood that may change the results. Each neighborhood conceptualization was tested and each time gave the same clusters, meaning that in these areas there are clusters of similar values no matter how a neighborhood is defined. Figure 7 shows that high values of subprime loan foreclosures were clustered in the southeast portion of Muncie while low values are clustered in the northwest portion of the city.

This tool also provides measures of statistical significance with the z-score and p-value for each census unit which are mapped in Figures 8 and 9 respectively. The p-value map mirrors the cluster/outlier map showing that these clusters also have a high statistical significance. Figure 8 is showing that these red areas have high z-scores and represent the tails of a normal distribution (ESRI 2013). The census block groups that are clustered and have high statistical significance also have high z-scores. Knowing this, it can be determined that these are hot spot regions unlikely to be the result of chance or a random outcome. This localized analysis made it clear where the clusters of subprime lending are and that these areas are not there by random chance. The next step is to see how the variables of median income and percent non-white affect this spatial distribution.
Figure 7: A map displaying the different clusters and outliers of the subprime ratio in Muncie. This map displays high values clustering with other high values in the south central part of the city and low values clustering with other low values in the north central part of the city.
Figure 8: A map displaying z-scores of the subprime ratio after the local Moran’s I test. The red and blue colors indicate the block groups that represent the tails of a normal distribution.
Figure 9: A map displaying the p-values of the subprime ratio after the local Moran’s I test. The darker colors show extreme statistical significance while the blue color shows no significance.
The next ESDA technique used was geographically weighted regression (GWR). This tool is also considered local analysis similar to the previous technique in order to evaluate this data on a local scale. This is a method that predicts the value of a dependent variable based on one or more independent variables by calculating a regression model for every feature in this case (ESRI 2013). This technique predicted the value of the subprime ratio based on the independent variables of percent non-white and median income and then compared this prediction with the actual values. The formula for GWR is:

\[ Y_i = X_i^t \beta (u_i, v_i) + \varepsilon_i = \beta_0 (u_i, v_i) + \sum_{k=1}^{p} X_{ik} \beta_k (u_i, v_i) + \varepsilon_i \]

where \((u_i, v_i)\) is the coordinates of feature \(i\) and \(\beta (u_i, v_i)\) is the vector for the location parameter estimates.

Two maps were created from the outputs of this tool. Figure 10 displays the standard deviation of the regression residual of the tool. This shows where the model over predicted (red) and under-predicted (blue) the subprime ratio variable. Figure 11 displays the \(r^2\) values for each census block. This value ranges from zero to one and ultimately shows that a variable not listed in the independent variables list is more responsible for variation than the ones that are listed. The highest \(r^2\) value is where only 19% of the variance is explained by median household income and percent non-white at best. This means that 81% of the variance is explained by other variables that are not listed. This \(r^2\) value of 19% is only for one block group. The second highest percent is only 11%. Very low \(r^2\) values are located in the southeast portions of Muncie. This indicates that this is the area most unexplained by these independent variables. This was an important step in the analyses of this study in order to understand if the explanatory variables are significant. This also allows for speculation of what the missing explanatory variables might be.
Regression Residuals for Subprime Ratio Variable from GWR Evaluation

Figure 10: A map displaying the regression residuals of the subprime ratio variable after the GWR. The darker colors indicate where the model did not closely predict the actual value.
Figure 11: A map displaying the $r^2$ values for the subprime ratio variable as an output of the GWR. The map shows the percentage of variation that can be explained by the variables of median income and percent non-white. Although the dark areas indicate higher values, these values only go as high as 19% which is still surprisingly low.
The location quotient (LQ) was the final ESDA technique to be used. Unlike the other techniques, there is not a tool in ArcMap to perform this analysis. LQ is a local measure of spatial concentration or dispersion appropriate for spatially aggregated data (Burt, Barber, and Rigby 2009). Although primarily used in the fields of planning and economics as a way to analyze the economic strength of a particular region relative to a larger geography, it has also been applied to a diverse set of topics (Burt, Barber, and Rigby 2009). If the LQ for a particular area is greater than one, this area has a higher than average concentration of the variable. The opposite is true for an area with an LQ value less than one (Burt, Barber, and Rigby 2009).

For this data the count of subprime foreclosures was divided by the total amount of foreclosures for every census block group. These differing values represent the spatial concentration of each census block group. These values were then divided by the value of the city’s spatial concentration or the total number of subprime foreclosures in the city divided by the total number of foreclosures in the city. The formula for the LQ is 
\[
LQ = \frac{X/Y}{X'/Y'}
\]
where X is the value of an attribute, Y is the total value of that attribute for that feature, X’ is the entire regions value of an attribute, and Y’ is the entire regions total value of that attribute. This then created a value for every census block group comparing that block group’s spatial concentration of subprime loans to that of the entire city. The output of this process is displayed as a map in Figure 12. This shows that 21 different block groups have a higher than average spatial concentration of subprime loans. All but one of the block groups that are clustered with subprime lending based on the local Moran’s I are included in that spatially concentrated group. The next map, Figure 13, shows this same data but with more data classes in order to visualize how concentrated or not concentrated the block group is. Through this method, 18 of the block groups are still considered concentrated.
Figure 12: A map displaying the LQ values for each block group. The spatial concentration of subprime loans leading to foreclosure are found in the central and western portions of the city.
Figure 13: A second map displaying the LQ values for each block group. This shows the variation of spatial concentration to display how concentrated or not concentrated a block group is by displaying more classes of data.
Chapter V: Results and Discussion

The goal of this study was to find the spatial relationship between subprime lending and minority and low income areas of a small city. This is a gap left by past academic research since all previous studies linking these variables were conducted in large, urban areas. These studies found that subprime lending was most common in minority and low income neighborhoods. This specific study is important in order to determine if these relationships hold true regardless of location. Hopefully this study will open doors for others to research these same relationships in other smaller scale urban regions in order to learn where these relationships are prevalent. This chapter will summarize the statistical findings of the last two chapters, provide implications, recommendations, and limitations of this study.

The results of the exploratory data analysis did not provide strong relationships between variables that were initially expected. Both variables of the count of subprime related foreclosures as well as the percent of subprime foreclosures did not have the relationship previous research had found. The only variable that showed the directionally expected outcome, meaning that there was a positive relationship with percent non-white and a negative relationship with median income, was the subprime ratio variable, although this relationship was hardly negative at all with almost no relationship. Even though this variable did show this directional relationship, it still did not show a very strong association with either one of the explanatory variables. The lack of strong relationships in these tests led to the spatial investigation of these variables.

The results of the exploratory spatial data analysis were much more telling than the EDA. The Moran’s I test indicated that spatial clustering was evident. In order to examine the subprime
ratio variable more closely, three local tests were performed to evaluate each census block group separately. The local Moran’s I showed areas with clusters of high values in the southeast part of the city and clusters of low values in the northwest part of the city. The geographically weighted regression showed very low $r^2$ values indicating that there is one or more missing variable(s) that would account for the clustering of subprime foreclosures. Finally the location quotient showed that 21 of the census block groups have a higher than average spatial concentration of subprime foreclosures. Included in these block groups were the block groups that showed high-high clusters in the local Moran’s I analysis. With all of these tests in mind it is clear that clustering and concentration of subprime foreclosures exists in Muncie, but it is unclear what variables led completely to these foreclosures.

The results of the numerous and different analyses were not quite what was expected. The relationships between subprime lending, minority neighborhoods, and income did exist, but not as strongly as was originally anticipated. As mentioned, one important factor that makes this study different than those performed before is that this study was conducted in a small city. Based on this study alone, there may be other important variables that affect the distribution of subprime loans in small cities that do not have the same effects in large metropolitan areas. In Muncie, lenders are not targeting those with low income or those that live in minority neighborhoods as much as in big cities, but does this leave room for lenders to target people based on a different attribute? Because of the size of the city, local lenders may play a larger role than large corporations. Local lenders more in touch with the city’s history may be choosing other factors on which to exploit homeowners if subprime targeting is taking place in Muncie.

Based on the cluster map produced by the local Moran’s I test, the division of Muncie is clear. In one portion of the city there is a cluster of an extremely low amount of subprime
lending while in the other side of the city there is a cluster of an extremely high amount of subprime lending. This socioeconomic division is historical in nature and was not first noticed in this research. During the Lynds’ 1929 Middletown study, they too noticed this division of Muncie. According to Fraser (2012), the socioeconomic division at this time had professionals in the north and the middle class in the south divided by the white River. This division became less and less through the following years because of the rise of industry in the South of Muncie bridging this gap. This boom in industry did not last long for Muncie as between 1979 and 2009 many large employers left town leaving Muncie with less than half the manufacturing jobs it once had, mostly affecting the southern half of the city (Fraser 2012). This made the division clear once more of the economically prosperous north side of Muncie with big employers like Ball State University and Ball Memorial Hospital and the south side of Muncie that was left devastated by deindustrialization. Through this research it is clear that the north side which is already benefitting from those large employers is also benefitting from a lower amount of clustering of subprime lending while the already underserved south section of the city is dealing with a heavy amount of clustering of subprime lending.

Drawing from the definitions of place presented by Massey (1994) and Agnew (1987) also gives reason for the socioeconomic division of Muncie. Agnew divides his definition of place into three interrelated concepts of location, locale, and sense of place. Location is described by Agnew as the economic function of a place. As discussed earlier the southern portion of the city was once the site of higher paying industrial jobs but is now left with mostly lower paying retail positions while the northern portion of the city is dominated by higher paying jobs in higher education and health care. Agnew defines locale as the mix of key institutions that organize social, economic, and political activities within a place. Ball State University is an
example of an important institution that not only provides higher paying employment opportunities within the northern portion of the city but also organizes many student-driven and open to the public activities. Agnew’s final concept, *sense of place*, deals with the unique character or identities that develop within a place which help differentiate it from other places. Taken together, the distinct *location* and *locale* of the northern portion of the city provides a different *sense of place* than is present in the south. From this perspective, northern Muncie is quite distinct from the south.

Massey’s definition of place compliments this understanding of two Muncies that is apparent through Agnew’s definition. Massey describes a place as being a product of human activity. Again the activity of the northern side of the city is centered on the university. Massey also defines a place as dynamic which is very true in the case of Muncie. The northern portion of the city is prospering around the university’s growth and other employment opportunities while the southern half changed in a more negative way with a population decline and loss of jobs. Through these two definitions of place it is clear that Muncie can be understood as two separate places: one that is now centered on the university and overall requires a more highly educated workforce and one that has changed from an economically sound place of industry to a place of declining population and lower paying jobs.

Based on this division of the city, several recommendations for future research and future variables come to mind, the first being a type of employment variable. The residents of Muncie that were once employed by big industry may not have the proper training or education to find another good paying job in a city that has shifted away from industry. One variable to explore would be the type of employment of those who possessed subprime loans. In the case of Muncie, lenders may be targeting the areas that have been hit hardest by the decline of jobs and
population. This type of targeting may be specific to towns that have lost a great amount of industry.

Another suggestion would be to divide the city into a northern and southern portion and then run the previous analysis to find different and more telling results. Because of the differing economic and cultural histories of each half of Muncie, different variables may play altered roles when each section is explored individually. In most of these methods of analysis, it is possible for areas to become marginalized when viewed as part of a much bigger whole. That is why this exploration could have different results entirely because of Muncie’s diverse history.

A final suggestion would be to explore an education variable. This variable is readily available on the American FactFinder website. An evaluation of this variable has been conducted. Figure 14 shows the $r^2$ map from a geographically weighted regression performed with an education variable. This education variable is the percent of higher education within a census block group calculated from information provided on the American FactFinder website. Although the original $r^2$ map in the ESDA chapter showed the highest value being 19%, that value was only in a single block group. Adding the variable of education shows a much more widespread area with a higher $r^2$ value. This evaluation gives some clues that education is an important factor in the city of Muncie, but it also shows areas with very low $r^2$ values, meaning that there are even more unexplored variables that may have an impact on this analysis.
Figure 14: An $r^2$ value map after the education variable was added to the list of explanatory variables in the geographically weighted regression. This shows the highest value affects multiple block groups unlike that of the previous analysis. This means that this education variable is an important missing link but other variables are still missing.
This study has a few limitations. The research is dependent on the data found from the sheriff’s sale records as well as from the Doxpop website. Missing foreclosures not listed in these sources may change the results. Another limitation is the subprime lending list provided by the U.S. Department of Housing and Urban Development. This list is created by HUD using indicators that suggest which financial institutions are more likely to issue subprime loans. This means that there is a chance that not every loan flagged as subprime was a subprime loan or that more subprime loans may exist that were not issued by one of these financial institutions flagged as subprime. If time allowed, more extensive paperwork may have been found through the city for each and every mortgage. This would be more helpful in determining for certain which loans were subprime. Unfortunately, there is not a guarantee that if time was put into finding this paperwork, it would even be available.

Through this research the spatial and non-spatial relationships between subprime lending, median income, and minority were explored in a small city. Previous research in larger cities showed that minorities and people living in low income areas were more likely to receive a subprime loan. This research indicates that there are clusters of subprime loans leading to foreclosure in this small city. This study also shows that there are spatial and non-spatial relationships between these variables for Muncie. However, there are also missing variables that would help further explain the clustering of these subprime loans.
References


Fraser, M. 2012. Down and Out in the New Middletowns. *Dissent Magazine*. 


Indiana Real Estate Trends & Market Info.


