Transit Prioritization for an Aging Population

Alachua County, Florida

Wes MacLeod
December 10, 2010

Table of Contents

1. Abstract

Over the next thirty years the portion of the population over the age of 65 is expected to more than double (Rosenbloom, 2003). Although estimates vary in terms of the extent to which this segment of the population will increase, all are certain that increases will be nothing short of significant – whether the number will double, or increase by half is yet to be seen. In addition, by 2030, more than nine percent of the population is estimated to be over the age of 85 (U.S. Census Bureau, 2002). Increasingly these folks will be stranded in the suburban landscape as most have no access to public transit. At present, roughly half the United States population is unable to access public transportation because service is not present (Bailey, 2004).

The retirement population, now able to afford to drive private vehicles, will increasingly be priced out of this option assuming a higher cost of fossil fuels. A greater portion of their fixed retirement income will be put towards private vehicle operation unless they are properly served by public transit.

Alachua County, situated in the heart of North Central Florida and home to the University of Florida, is one of the more progressive counties in the state in terms of planning for rapid transit. While Alachua County does not have as great a population of elderly some as counties in south Florida it still should prioritize transit planning for this demographic. Low-income populations in the county are also subject to the resulting gentrification that may come to fruition with proposed bus rapid transit and transit oriented developments. In the future access to public transit may be viewed as basic human right in the same manner as shelter, but for now most Americans are left to fend for themselves.

For the purposes of this paper future rapid transit corridor locations are analyzed in relation to the projected 65 and older populations. Assumptions of high gas prices will be implicit, and the importance of transit access by the year 2030 will be the driving factor. To test these notions 20 year projections of the elderly (65+) are predicted using a series of regression models.

ESRI's Geographic Information System (GIS) software was used to run spatial statistics models on the projected elderly population. Ordinary Least Squares (OLS) and Geographically Weighted Regression (GWR) models were primarily used to predict the spatial locations of the elderly.

The final iteration resulted in an OLS coefficient of determination value (R^2) of 0.64 with residuals that were neither clustered nor dispersed. Results of the GWR model were an R^2 value of 0.84 – indicating a sufficient goodness of fit, 0 being poor and 1 being ideal – with residuals that may be the result of random chance.

Model results of 2030 elderly populations showed a significant disconnect from proposed rapid transit routes. Analysis shows that proposed routes will serve a significant amount of residential units in the western portion county. Consequently, model outcomes classified this area least likely to house much of the elderly population by 2030. However, further data and variables may be needed to create a more statistically accurate model.

2.Introduction

Defining the Need for Transit Access

Never before has the retirement population been as accustomed to the private vehicle for its primary source of transportation as have the baby boomers. Cheap fossil fuels, the dominance of the automobile and sprawling land use patterns have defined an unsustainable way of life for much of the aging population.

Following the oil embargo crises of the 1970s, much of the developed world took heed to plan for more equitable transportation options – turning their focus away from transit only by the single occupancy vehicle. In the United States however, land and gasoline have remained cheap thus furthering our reliance on the private vehicle, leaving many Americans vulnerable to fossil fuel price hikes and the resulting isolation from retail, community, and health facilities.

Presently, the overwhelming majority of the elderly populations in the United States reside outside of the city center (see figure 2.1). Much of this can be explained by folks "aging in place" or choosing not to leave the residence in which they raised a family or worked for much of their adult life.

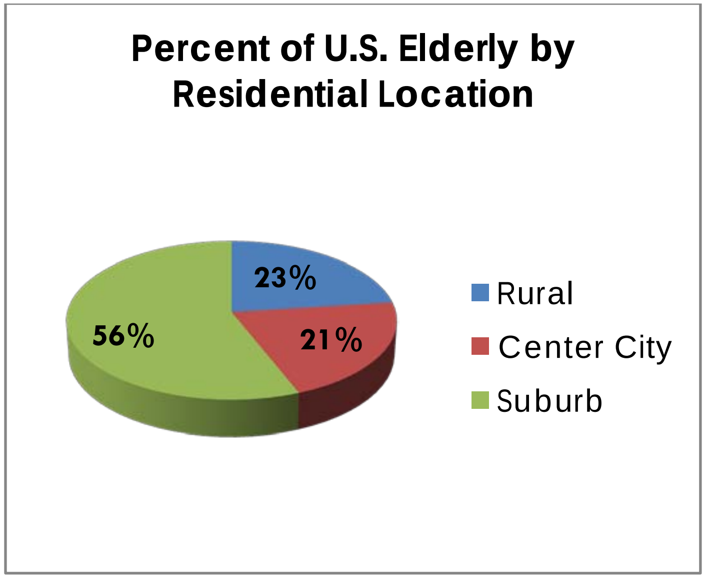


Figure 2.1: Lavada E. DeSalles, "Testimony to U.S. Senate Committee on Banking, Housing and Urban Affairs," July 17, 2002

Elderly habitation patterns create greater reliance on the private vehicle than younger segments of the population. In fact, a greater percent of trips are made by the private vehicle for populations above the age of 65 (John Pucher, 2003). Licensing among the retirement population has also increased in the last decade as a greater portion of women are becoming licensed drivers.

Though much of the elderly population in the U.S. chooses to drive, many only do so as a necessity.

Much of the demographic, if given other transit options, may choose to give up their vehicle. In the U.S. elderly transit ridership is often made up by individuals with no private vehicle or automobile access, yet "in Australia, Europe, and Canada, elderly car drivers make up a meaningful percentage of transit users" (Morris, 1998).

While elderly urban populations have greater access to public transit than do their suburban or rural counterparts many are still unable to take advantage of such service. Most note too great a distance to bus stops and a lack of reliable and consistent service as barriers to using public transit. Public transit access in the suburbs is worse or in many cases non-existent. According to the Community Transit Association of America (CTAA):

> The past two decades have seen many forms of transportation virtually abandon rural areas. Small town residents often travel hundreds of miles just to access the nearest airport; intercity bus service is a shell of its former self; taxi service is scant and expensive; and passenger rail services often only streaks through the countryside in the middle of the night. In 1996, CTAA found that two of five rural counties had no public transit, and another 25 percent had service equal only to one trip per month (Bogren, 1998).

Additionally, isolation is heightened for non-drivers over the age of 65. Studies and research show that more than half this demographic chooses not to leave their residence because transportation options are unavailable or difficult to access. This occurrence is higher for minorities, rural communities or homes with no private vehicle (Bailey, 2004). American culture breeds a sense of personal independence and reliance that often leave older non-drivers reluctant to ask for a ride. Senior citizens note a hesitance to rely on other for rides because they do not want to "impose on others (Stowell Ritter, 2002)."

Safety among aging populations is another important factor when considering public transit prioritization. Drivers over the age of 65 are much more likely to be involved in accidents than younger drivers as a function of total miles driven (see figure 2.2). For drivers over the age of 85 the fatality rate is nine times higher than for drivers aged 25 to 69 (National Highway Traffic Safety Administration, 2000).

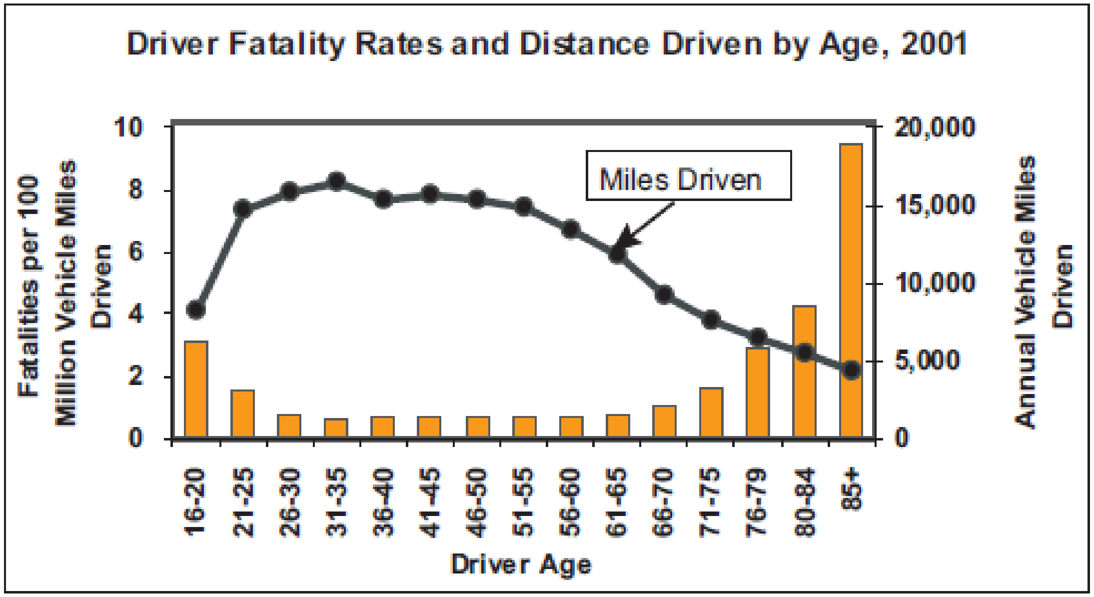


Figure 2.2: Lavada E. DeSalles, "Testimony to U.S. Senate Committee on Banking, Housing and Urban Affairs," July 17, 2002

3.Methodology

Data Collection

In order to determine if the elderly will be served by bus rapid transit in the future one must first predict their habitation patterns. Data layers used for this function are census block group shapefiles for the years 1990 and 2000. Census block groups are created as a result of the underlying population in their boundaries.

Initially, the attempt was made to aggregate 1990 block group data with 2000 blocks as this would create more data points to run the model. However, after exploring and removing the new data it was determined that the spatial inaccuracies were too great. As block groups cover more real estate than census blocks, it would have to be assumed that populations within the 1990 boundaries were evenly distributed. The proportion of the block group coinciding with the year 2000, much smaller census block, would then be calculated as a percentage of the original total. Unfortunately, this yielded extremely high differences and shifts in population change. As a result the decision was made to instead use 2000 census block group data.

First, the year 2000 block group layer was intersected with the 1990 layer. Next a proportional allocation formula is used to assign values to the 1990 blocks that have been intersected with year 2000 boundaries (Jack Baker, 2005). For the purpose of this model and project, it is assumed that population in the overlap areas is evenly distributed.

The newly created shapefile was exploded and new acreages were calculated. Any polygon with less than 7 acres was removed - as they were determined slivers. The following equation is used to repopulate the attributes of 1990 census block file:

newPop1990 = (Intersection shape area/1990Area) *Pop1990

To allow for the equation to function properly all zeroes were removed by adding 1 to all values. The attribute field used for this model is “65_UP.”

Percent change and the number of population increase or decrease is then calculated and attributed to each polygon. Percent change values in the data greater than 500% or less than 500% are removed from the file to lessen error as such great values changes are likely the result of assuming that population is evenly distributed. Outliers were located and removed.

Population Projections

Demographers employ many different strategies to project future population. Some may use a modified compound interest formula; others use historic growth figures to estimate the value. For the purpose of this model – and the realization that I am not a demographer – a simple projection was done to calculate the 65 and up population for the year 2030. Because the data for 1990 and 2000 now had the same spatial location it is possible to take the difference between the two values, multiply the number by three (representative of the next three decades) and add it to the 2000 value.

2030 65 and Up = [(2000_65_Up – 1990_65_Up) * 3] + 2000_65_Up

Essentially, a rudimentary linear formula is applied to determine the population for 2030. This method has merit however; as research has proven that older populations tend to age in place (Bailey, 2004).

Still, to create a proper model other variables are needed to explain the variation in data. 2010 ESRI business analyst data at the block group level were used as explanatory variables. The following layers were used:

- Home Value
- Household Income
- Disposable Income
- Net Worth
- Population By Race

Each layer was converted to points. Centroids of each polygon were created using the calculate geometry function, display X, Y data and export. Each layer was spatially joined to the 2030, 65 and up polygon shapefile. From here the final layer was converted to points using the method described above.

4. Spatial Modeling

Variables, Parameters & Process

ESRI's Geostatistical Wizard is used to create an Inverse Distance Weighted (IDW) model on the 2030 elderly population (see figure 4.1). Initial analysis showed that proposed bus rapid transit routes correlate with the projected figures. However, to get a better understanding of the data and to allow for a more evenly distributed set of prediction points it would be necessary to create a model.

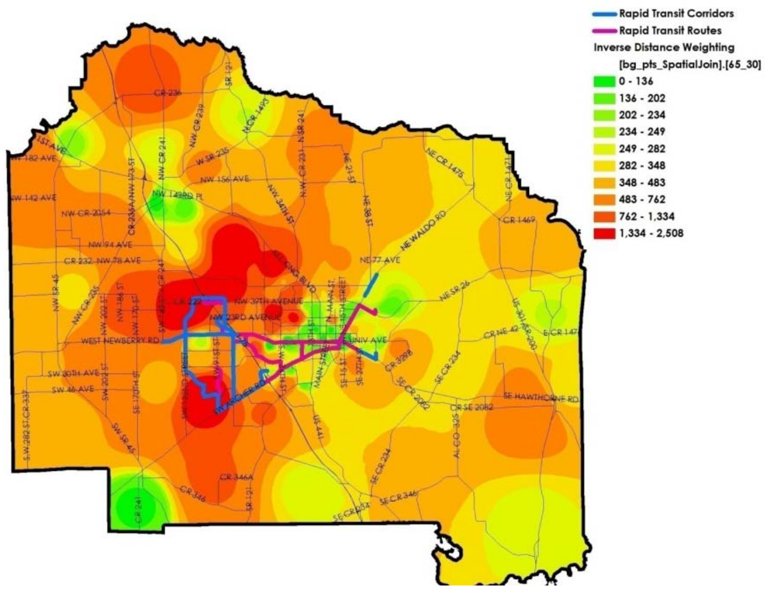


Figure 4.1: Inverse Distance Weighted Model – variable observed 2030 65 and Up Population. Blue and purple lines represent proposed rapid transit locations.

The first step was to use the Ordinary Least Squares (OLS) function to determine the best explanatory variables (dependent variable - 2030 65 and Up). Initial variables used were:

- Disposable Income – median
- Household Income – median
- Home Value – median
- Net Worth – median
- Ethnicity – African American, Hispanic, White

Disposable and household incomes were determined have a direct relationship so the latter was removed. Running the model once more yielded an R^2 value of 0.74. Disposable income and the presence of Hispanics were determined to have a negative correlation with the location of elderly populations. Residuals were used to determine clustering with the Spatial Autocorrelation (Moran's I) function with row standardization (see figure 4.2). A histogram of the residuals showed a slight negative skewness. Even though the residuals were clustered a Geographically Weighted Regression (GWR) model was run.

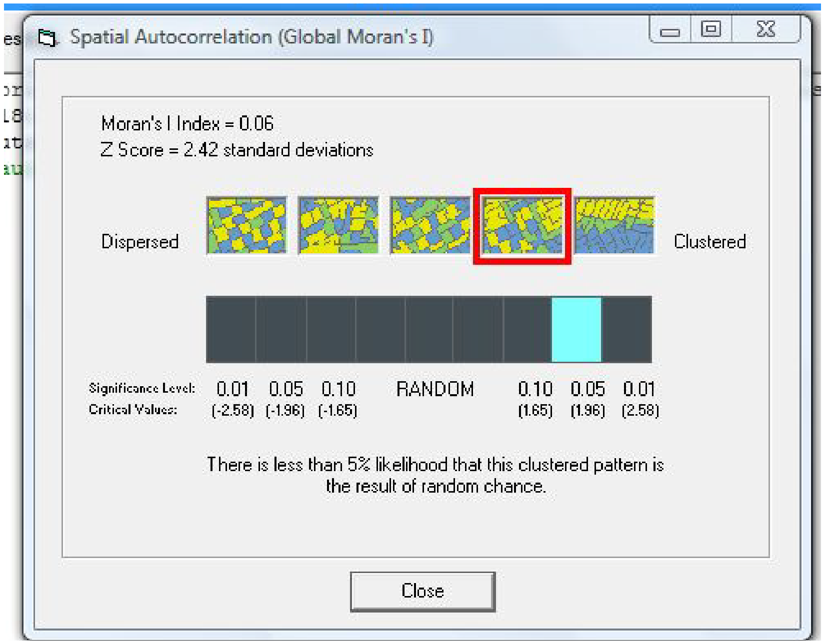


Figure 4.2: Spatial Autocorrelation (Moran' I) results of Ordinary Least Squares (OLS) residuals.

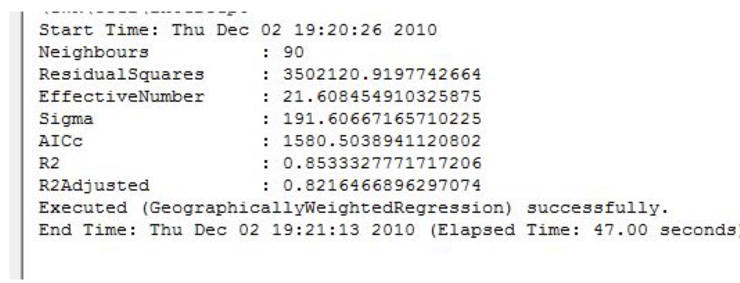
The GWR function was run with the dependent variable: 2030 65 and Up, and the explanatory variables:

- Disposable Income – median
- Home Value – median
- Net Worth – median
- Ethnicity – African American, Hispanic, White

The following parameters were specified to run the model:

- Kernel Type: Adaptive
- Kernel Type: Adaptive
- Bandwidth Method: Bandwidth Parameter
- Number of Neighbors: 90 (6 explanatory variables)

Results yielded an R^2 value of 0.85, greater than with OLS (see figure 4.3); residuals were neither clustered nor dispersed with a Z-score of -0.41.



```
Start Time: Thu Dec 02 19:20:26 2010
Neighbours           : 90
ResidualSquares      : 3502120.9197742664
EffectiveNumber      : 21.608454910325875
Sigma                : 191.60667165710225
AICc                 : 1580.5038941120802
R2                   : 0.8533327771717206
R2Adjusted           : 0.8216466896297074
Executed (GeographicallyWeightedRegression) successfully.
End Time: Thu Dec 02 19:21:13 2010 (Elapsed Time: 47.00 seconds)
```

Figure 4.3: Geographically Weighted Regression (GWR) results.

Predictions

Next, the Public Land Survey System (PLSS) layer was used to create prediction points. An IDW model was created from the predicted values (see figure 4.4). This IDW differed substantially from the initial IDW created from observed points, predicting that the elderly population would be primarily located in the north and south extents of the county.

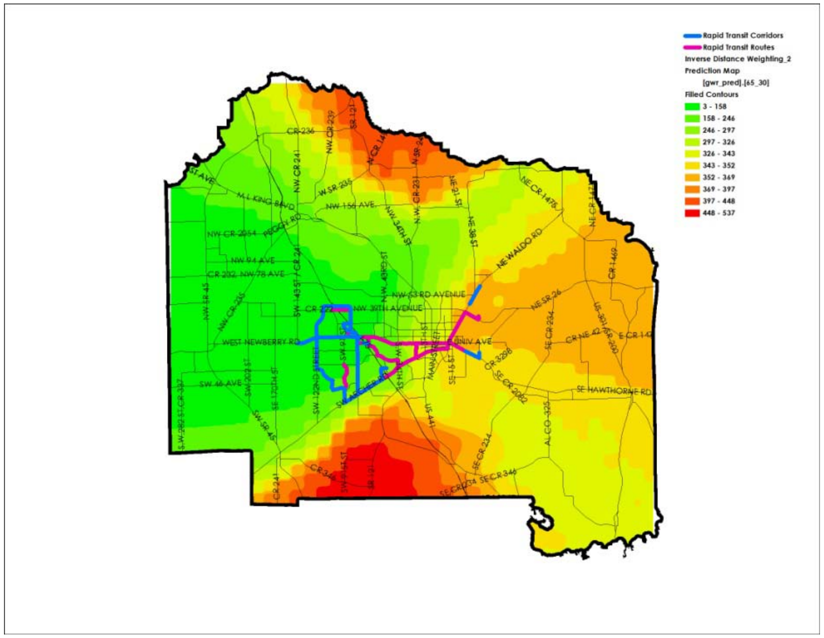


Figure 4.4: IDW model – variable GWR predicted point values. Blue and purple lines represent proposed rapid transit locations.

Because the initial run through OLS produced an outcome with clustered residuals the function was revisited. First, the disposable income variable was removed as much of the elderly population can be classified as no longer earning an income.

Results yielded a similar R_2 value similar with clustered residuals; however the Z-score was smaller. The coefficient output table confirmed standard error was much higher for the Hispanic variable – nearly four times higher than remaining explanatory variables. The model was run once more, this time with only four explanatory variables:

- Home Value – median
- Net Worth – median
- Ethnicity – African-American, White

The results yielded a lower R^2 value, 0.64, compared to 0.74; however, the Moran's I function determined the residual pattern may be a function of random chance (see figure 4.5).

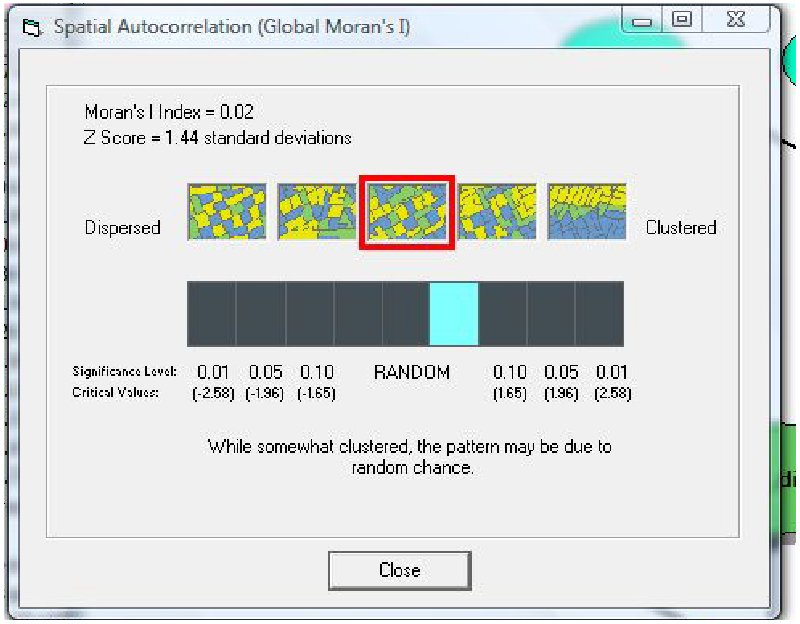


Figure 4.5: Spatial Autocorrelation (Moran's I) results of OLS residuals from four remaining explanatory variables.

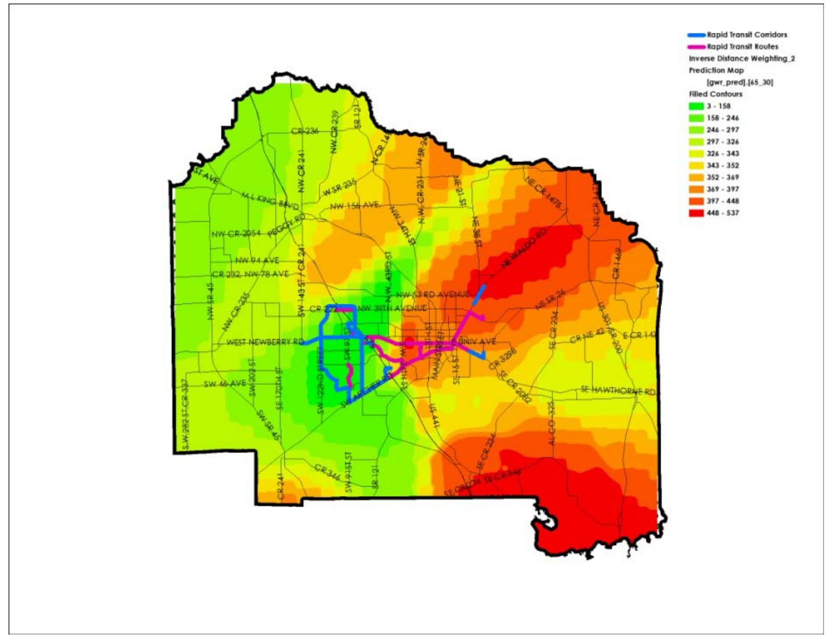


Figure 4.6: IDW model – variable GWR predicted point values with four explanatory variables. Blue and purple lines represent proposed rapid transit locations.

Next, GWR was run using the same four variables and parameters as specified above, but with only 60 neighbours. Results yielded and R^2 value of 0.84 with residuals that may be a result of random chance. Predicted points were used to create an IDW model. The surface showed a concentration of high values along NE Waldo Road and in the south-east portion of the county (see figure 4.6).

Examining "local R^2" attributes showed values above 0.7 were concentrated in the north west portion of the county (see figure 4.7). The presence of Paynes Prairie and other significant public lands may be the cause of such a significant difference in R^2 values. As a result, a raster mask was created using features from the Public Lands shapefile provided by the Florida Geographic Data Library.

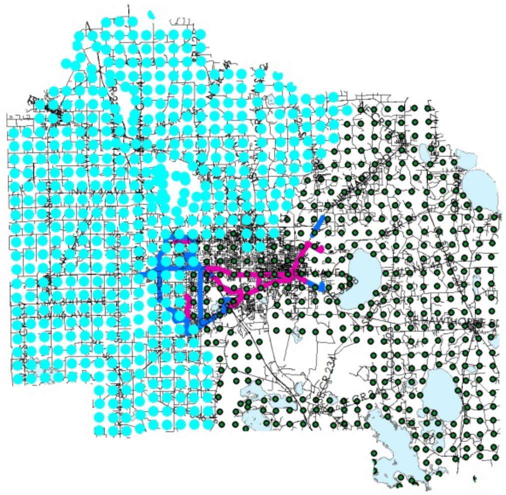


Figure 4.7: Prediction points with local R^2 values greater than 0.7 (shown in bright blue).

Ordinary Least Squares was run once more, this time with the raster mask. Surprisingly, the outcome resulted in residuals that were neither clustered nor dispersed (see figure 4.8) and an R^2 value of 0.64 (see figure 4.9). GWR resulted in an R^2 of 0.84 and residuals that may be the result of chance. Prediction points were used to create a simple Kriging model to analyze the surface (figure 4.10). Still, local R^2 values greater than 0.7 remained concentrated in the north and west portion of the county.

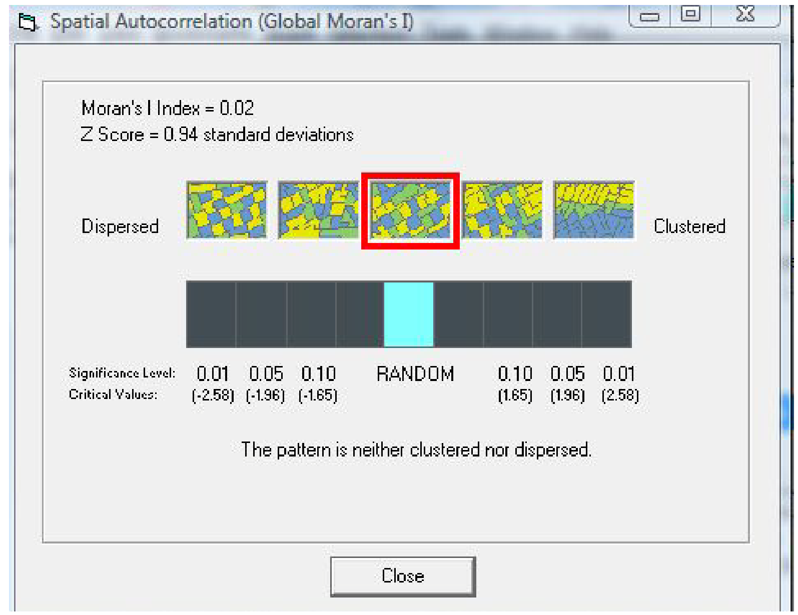


Figure 4.8: Spatial Autocorrelation (Moran's I) results of OLS residuals from four remaining explanatory variables with raster mask applied.

OID	Field1	Variable	Coef	StdError	t_Stat	Prob	Robust_SE	Robust_t	Robust_Pr
0	0	Intercept	102.178935	73.750093	1.385475	0.168382	79.045312	1.292663	0.198516
1	0	WHITE_12	0.22346	0.017255	12.950502	0	0.029733	7.515461	0
2	0	BLACK_12	0.07696	0.058634	1.312548	0.191745	0.056462	1.363044	0.175326
3	0	VAL_MEDIA	-0.000259	0.000561	-0.461639	0.645151	0.00071	-0.364346	0.716224
4	0	NW_MEDIAN	0.000723	0.00028	2.583502	0.010925	0.000333	2.174161	0.031569

Attributes of ols3_diag

OID	Field1	Diag_Name	Diag_Value	
0	0	AIC	1822.827387	Akaike's Information Criterion: A relative measure of performance used to compare mod
1	0	R2	0.647405	R-Squared, Coefficient of Determination: The proportion of variation in the dependent va
2	0	AdjR2	0.636122	Adjusted R-Squared: R-Squared adjusted for model complexity (number of variables) a
3	0	F-Stat	57.378571	Joint F-Statistic Value: Used to assess overall model significance.
4	0	F-Prob	0	Joint F-Statistic Probability (p-value): The probability that none of the explanatory variab
5	0	Wald	109.040606	Wald Statistic: Used to assess overall robust model significance.
6	0	Wald-Prob	0	Wald Statistic Probability (p-value): The computed probability, using robust standard err
7	0	K(BP)	36.699147	Koenker's studentized Breusch-Pagan Statistic: Used to test the reliability of standard e
8	0	K(BP)-Prob	0	Koenker (BP) Statistic Probability (p-value): The probability that heteroskedasticity (non-
9	0	JB	61.596619	Jarque-Bera Statistic: Used to determine whether the residuals deviate from a normal d
10	0	JB-Prob	0	Jarque-Bera Probability (p-value): The probability that the residuals are normally distribu
11	0	Sigma2	69297.88729	Sigma-Squared: OLS estimate of the variance of the error term.

Figure 4.9: OLS coefficient and diagnostic output table.

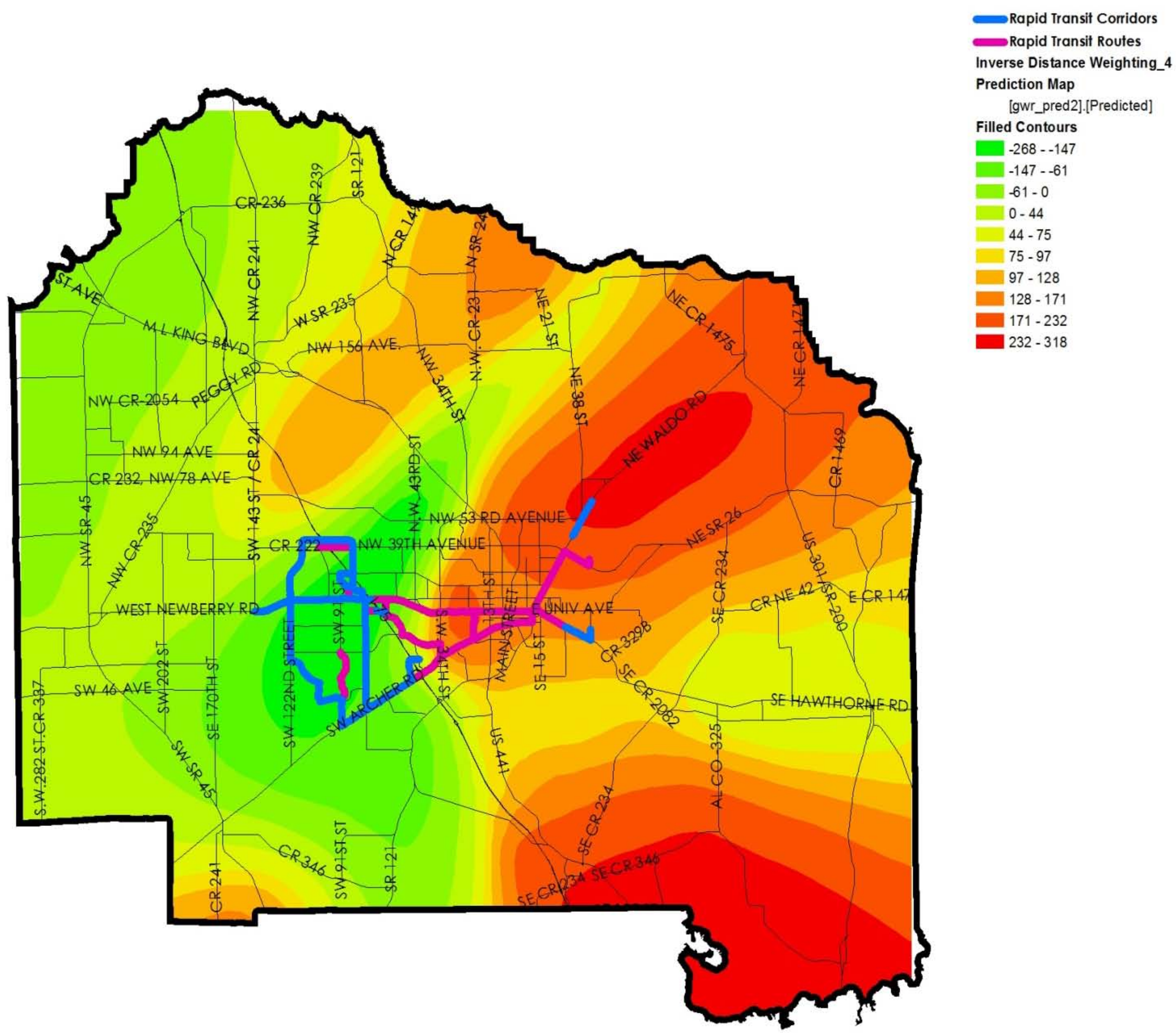


Figure 4.10: Simple Kriging model – variable GWR predicted point values with four explanatory variables and applied raster mask. Blue and purple lines represent proposed rapid transit locations.

5. Discussion

Based on the model results, it may only be feasible to make assumptions about data in the north and west areas of the county where local R^2 values are greater (0.7 or higher).

That said, according to model results, proposed rapid transit lines are located in areas where the elderly may not be located in the future – namely just west of I-75 and north of Archer Road (see figure 4.11).

Assuming, the model is correct more emphasis and study is needed just north east of proposed lines in the Alachua County Urban cluster area. Additionally, if in fact growth does shift to the east side of Gainesville more lines and transit will be necessary. Current transit options are sparse and head ways are unappealing for riders.

Clearly future rapid transit routes have been proposed to serve a large segment of the residential units in the city and county (see figure 4.11 & 12), unfortunately many of these folks may travel by private vehicle regardless of available transit options. Those less fortunate, such as the elderly or low-income may be left without such a choice in the future, if transit is not provided in close proximity to their geographic locations.

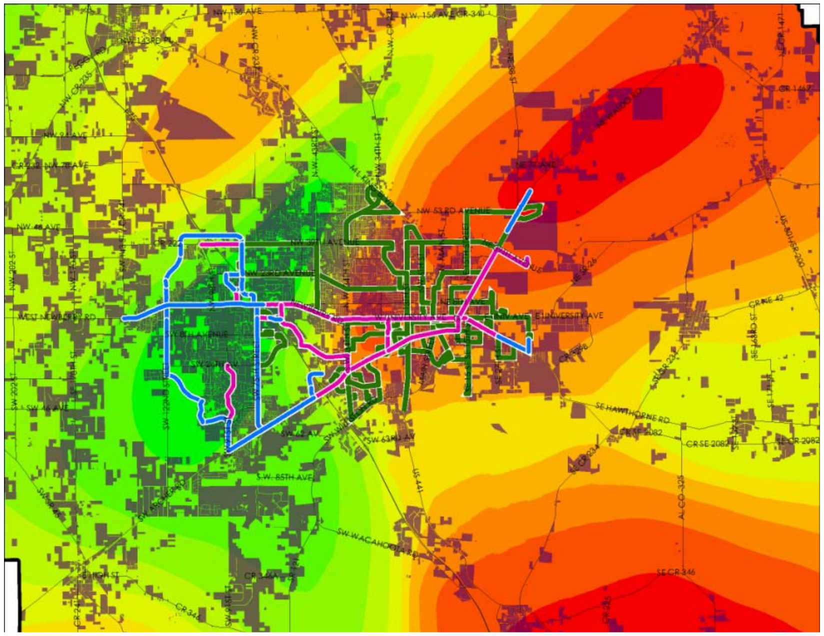


Figure 4.11: Simple Kriging model – variable GWR predicted point values with four explanatory variables and applied raster mask. Vacant and existing residential and commercial land use layers are shown in dark purple. Blue and purple lines represent proposed rapid transit locations. Green lines represent existing RTS routes.

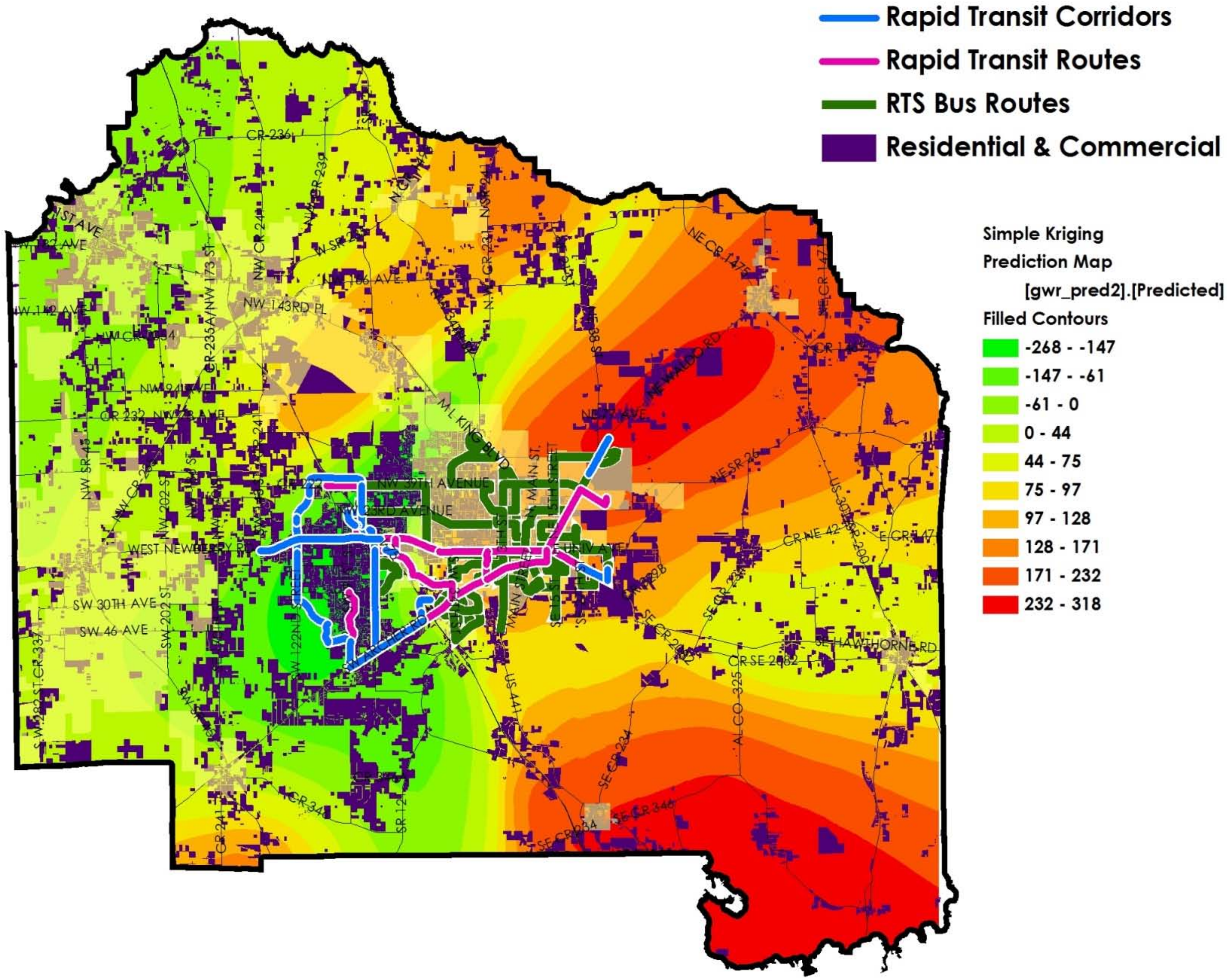


Figure 4.12: Simple Kriging model – variable GWR predicted point values with four explanatory variables and applied raster mask. Vacant and existing residential and commercial land use layers are shown in dark purple. Blue and purple lines represent proposed rapid transit locations. Green lines represent existing RTS routes. Yellow transparency represents the incorporated areas of Alachua County.

6. Conclusion

To generate a stronger statistical model additional variables and data is necessary. Model results in the south and east areas of Alachua County yielded undesirable local R^2 values. Additional variables are needed to explain variation in this area of the County. Furthermore, only 130 points were used to run spatial statistics functions – roughly half the preferred number of values need to create a statistically sound model.

At present the Gainesville Metropolitan Transportation Planning Organization (MTPO), has included peak oil assumptions in there transportation modeling efforts. This, in combination with spatial predictions of low-income and elderly populations may help prioritize public transit planning in Alachua County.

Additionally, comprehensive transportation and land use planning efforts should place greater emphasis on spatial habitation patterns. Less focus should be placed on predicting (or dictating) land use and more initiative should be taken to examine demographic trends – whether the segment elderly, low-income, obese, adolescent, etc. Effective spatial modeling of such patterns will allow for a more proper allocation of future resources and should result in more effective planning outcomes on the ground.

7. Works Cited

Bailey, L. (2004). *Aging Americans; Stranded Without Options.* Surface Transportation Policy Project.

Bogren, S. (1998). You Can Get There From Here. *Community Transportation* .

Jack Baker, X. R. (2005). *GIS Applications for Demographic Estimation and Projection.* BEBR - University of New Mexico.

John Pucher, J. R. (2003). Socioeconomics of Urban Travel; Evidence from the 2001 NHTS. *Transportation Quarterly* .

Morris, S. R. (1998). The Travel Patterns of Older Australians in an International Context:Policy Implications and Option. *Transportation Research Record* , 189 - 193.

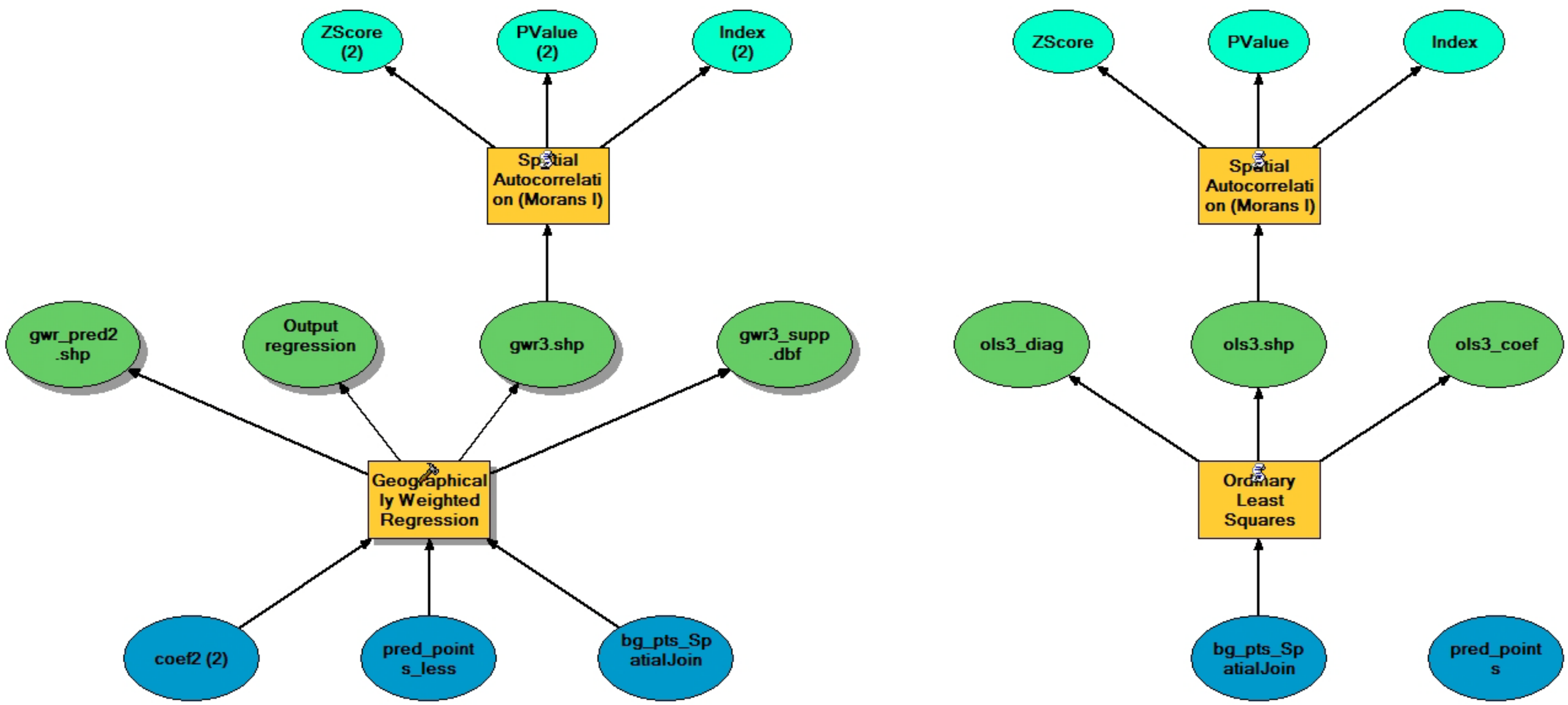
National Highway Traffic Safety Administration. (2000). *Traffic Safety Facts 2000.* U.S. Department of Transportation.

Rosenbloom, S. (2003). *The Mobility Needs of Older Americans: Implications for Transportation Reauthorization.* The Brookings Institution.

Stowell Ritter, A. S. (2002). *Understanding Senior Transportation: A Report and Analysis of a Survey of Consumers Age 50+.* AARP.

U.S. Census Bureau. (2002). *Projections of the Population by Age, Sex, Race, and Hispanic Origin for the United States, 1999 to 2100, Middle Series".* Department of Commerce.

8. Appendix



8.1 ModelBuilder Diagram – GWR (right) and OLS (left).

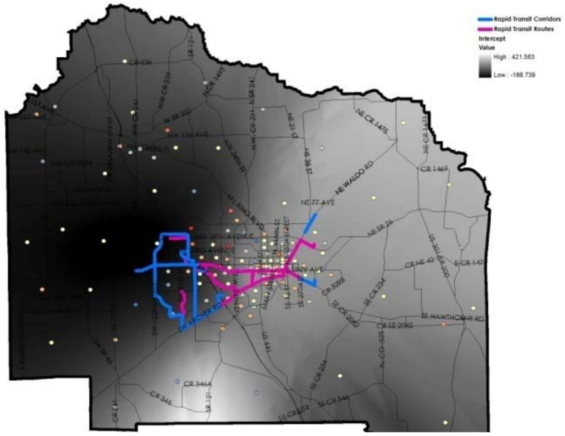
Attributes of ols3_coef

OID	Field1	Variable	Coef	StdError	t_Stat	Prob	Robust_SE	Robust_t	Robust_Pr
0	0	Intercept	102.178935	73.750093	1.385475	0.168382	79.045312	1.292663	0.198516
1	0	WHITE_12	0.22346	0.017255	12.950502	0	0.029733	7.515461	0
2	0	BLACK_12	0.07696	0.058634	1.312548	0.191745	0.056462	1.363044	0.175326
3	0	VAL_MEDIA	-0.000259	0.000561	-0.461639	0.645151	0.00071	-0.364346	0.716224
4	0	NW_MEDIAN	0.000723	0.00028	2.583502	0.010925	0.000333	2.174161	0.031569

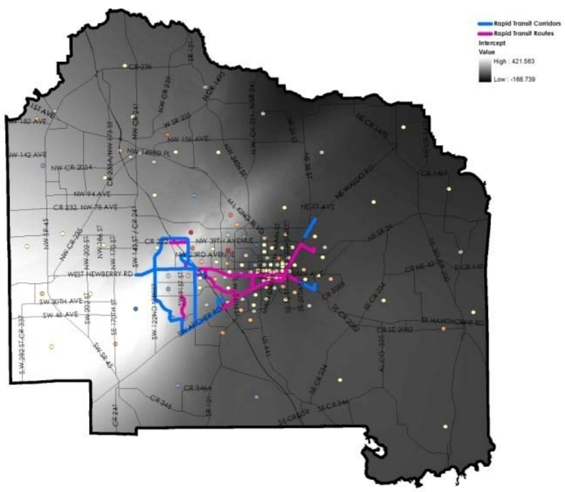
Attributes of ols3_diag

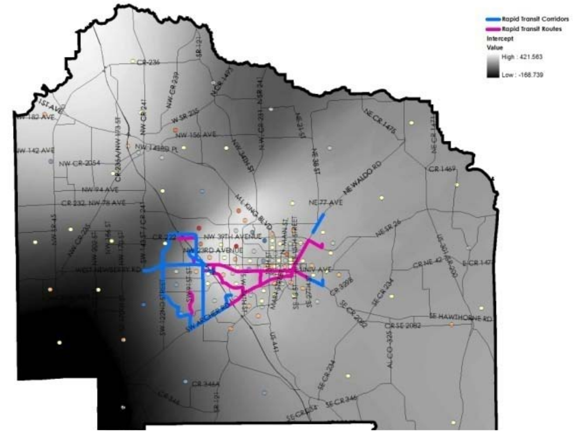
OID	Field1	Diag_Name	Diag_Value	
0	0	AIC	1822.827387	Akaike's Information Criterion: A relative measure of performance used to compare moc
1	0	R2	0.647405	R-Squared, Coefficient of Determination: The proportion of variation in the dependent va
2	0	AdjR2	0.636122	Adjusted R-Squared: R-Squared adjusted for model complexity (number of variables) a
3	0	F-Stat	57.378571	Joint F-Statistic Value: Used to assess overall model significance.
4	0	F-Prob	0	Joint F-Statistic Probability (p-value): The probability that none of the explanatory variab
5	0	Wald	109.040606	Wald Statistic: Used to assess overall robust model significance.
6	0	Wald-Prob	0	Wald Statistic Probability (p-value): The computed probability, using robust standard err
7	0	K(BP)	36.699147	Koenker's studentized Breusch-Pagan Statistic: Used to test the reliability of standard e
8	0	K(BP)-Prob	0	Koenker (BP) Statistic Probability (p-value): The probability that heteroskedasticity (non-
9	0	JB	61.596619	Jarque-Bera Statistic: Used to determine whether the residuals deviate from a normal d
10	0	JB-Prob	0	Jarque-Bera Probability (p-value): The probability that the residuals are normally distribu
11	0	Sigma2	69297.88729	Sigma-Squared: OLS estimate of the variance of the error term.

8.2: Final OLS results – coefficient and diagnostic table.



8.3 &4:Coefficient rasters – net worth (above) and home value (below).





8.5 & 6:Coefficient rasters – African-American (above) and white (below).