Using GIS to Explore the Relationship between Socioeconomic Status and Demographic Variables and Crime in Pittsburgh, Pennsylvania

Stephen E. Mitchell
Department of Resource Analysis, Saint Mary's University of Minnesota, Winona, MN 55987

Keywords: GIS, Crime, Statistical Analysis, Regression, Socioeconomic Variables, Demographic Variables, Hotspot Analysis

Abstract

Many areas across the United States have experienced rising crime rates over the last few decades. The identification of variables related to crime could allow policymakers to develop improved approaches to combating crime. Using multiple regression analysis, relationships between socioeconomic status/demographic variables and crime variables were investigated for Pittsburgh, Pennsylvania in a spatial context using census tracts as the geographic unit of analysis. Using Geographic Information Systems (GIS), choropleth maps were created in order to provide visual representations of the statistically related variables. Hotspot analyses were conducted to identify areas having high or low concentrations of values for crime and demographic variables.

Introduction

Over the last 30 years, crime rates have continued to rise dramatically in most of the United States, in some areas even doubling or tripling (Ackerman and Murray, 2004). The identification of variables related to crime could allow policymakers to develop better strategies to combat some of the causes of crime and could assist law enforcement officials in better anticipating changes in the crime patterns in an area. A wide range of variables are classified as socioeconomic variables. These include household income, educational achievement, employment status, and poverty status. Meanwhile, other demographic variables include age, race, and religion. In the past, the relationship between these variables and types of crime is often unclear, as contradictory findings from empirical studies often exist (Allen, 1996). However, the utilization of newer or improved analytical tools, such as Geographic Information Systems (GIS), has helped to further crime research endeavors.

Over the past two decades, use of GIS in crime analysis has increased dramatically (Vann and Garson, 2001). This powerful tool allows researchers to gain an improved understanding of crime in a spatial context. GIS offers a potent set of analysis tools that can be employed to further enhance the study of the relationships between crime and other variables by allowing researchers to not only map these associations but also conduct higher level spatial analyses. Hot-spot, crime density, proximity, and time-series mapping are a few of the ways crime data can be spatially examined using GIS (Vann and Garson, 2001).

Mitchell, Stephen. 2011. Using GIS to Explore the Relationship between Socioeconomic Status and Demographic Variables and Crime in Pittsburgh, Pennsylvania. Volume 13, Papers in Resource Analysis. 11 pp. Saint Mary's University of Minnesota University Central Services Press. Winona, MN. Retrieved (date) http://www.gis.smumn.edu

The purpose of this study was to gain a better understanding of the relationship between crime and socioeconomic and demographic variables in Pittsburgh, Pennsylvania (Figure 1). Census variables of the most interest are those that can vary over time for an individual, group, or area, such as household income, employment status, housing vacancy and poverty status, as these variables may be influenced by public policy. A description of the variables investigated is provided in Appendix A. Crime variables were limited to the categorization of certain crimes into total property crime and violent crime groups and are later described in more detail. Using multiple linear regression analysis and Geographic Information Systems, relationships among variables were examined using the city's census tracts as the geographic units of analysis (Figure 2). By investigating these relationships, using GIS as a tool of analysis and means of presentation of findings, a deeper understanding of the factors correlating with crime in Pittsburgh was achieved.

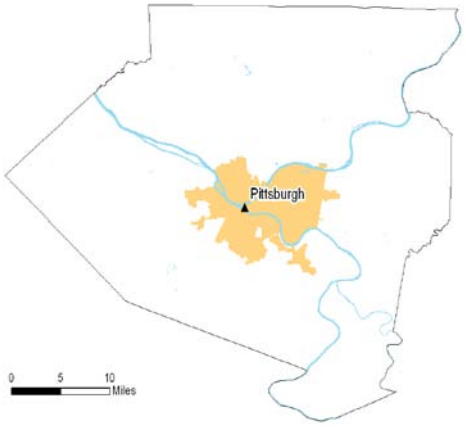

Figure 1. The study area, shown in orange, includes all of the city of Pittsburgh and is situated at the center of Allegheny County, Pennsylvania.

Figure 2. The 141 census tracts of the city of Pittsburgh comprising the study area are displayed having red borders.

Finally, for ethical reasons, it was extremely important that no individuals or specific households be identified in this study. This study only used aggregated crime data, socioeconomic status data and demographic data at the census tract level. This ensured that no individuals or residential locations were uniquely identified in the study and this strictly conforms to the National Archive of Criminal Justice Data standards for data dissemination.

Methods

Data Collection

Three categories of data were collected for the study. These included data containing socioeconomic status and demographic variables, data for crime occurrences, and GIS data representing the Pittsburgh area tracts and geography. Crime data for the city were acquired from the National Archive of Criminal Justice Data (Cohen and Gorr, 2005). The dataset was comprised of monthly crime occurrences for numerous types of crime for the city of Pittsburgh at the census tract level for the years 1990-2001.

Socioeconomic and demographic data were acquired in the form of the complete year 2000 Summary File 3 for Allegheny County census tracts as distributed by the U.S. Census Bureau. Data at the census tract level were chosen in order to match the geographic unit of analysis of the acquired crime data.

TIGER/Line shapefiles for the census tracts for both 1990 and 2000 for Allegheny County were downloaded from the U.S. Census Bureau website. Allegheny County river polygon and locality points shapefiles were acquired via the Pennsylvania Spatial Data Access website.

Pre-Analysis Examination and Processing of Data

The NAJCD crime data (Cohen and Gorr, 2005) included monthly census tract level crime data for 33 crime types for the period of 1990 through 2001. The data were first reduced to the eight "Part 1" crimes contained in the FBI's Uniform Crime Reporting Crime Index as listed in Table 1 (National*Atlas*, 2011). These crimes are considered to be Part 1 crimes "due to their seriousness and frequency" (National*Atlas*). Then, the monthly crime data for the year 2000, corresponding with the year of the acquired census data, were combined to represent yearly crime occurrences. Finally, crimes were designated as either a property or violent crime and combined to form the two crime groups: total property crime and total violent crime. This served to reduce the data into two distinct, meaningful groups. Table 1 lists the variables for each of the two groups and shows the portion of the total crimes that the variables contribute to their respective categories.

Table 1. The Part 1 crimes of the Uniform Crime Reporting Crime Index distributed into the two crime type categories.

Property Crimes	% of Total
Arson	1.0
Burglary	19.0
Larceny	56.8
Motor Vehicle Theft	23.1

Violent Crimes	% of Total
Aggravated Assault	39.2
Homicide	1.5
Rape	6.2
Robbery	53.0

Choropleth maps were created and examined in order to gain an initial understanding of the distribution of crime across the census tracts (Figure 3 and Figure 4). The number of property crimes ranged from 7 to 1202 occurrences, with a mean of 119.5. Violent crimes varied in number with 5 census tracts having no occurrences to one tract having 140 episodes. The mean number of violent crimes across the census tracts was 16.35. For both crimes, the highest number of crimes occurred in the "arrowhead" shaped census tract that contained the downtown/central business district.

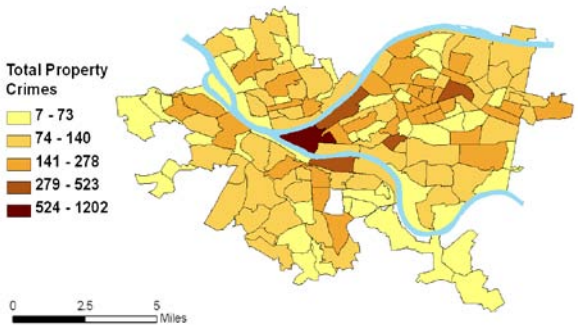

Figure 3. Property crimes across census tracts, categorized using natural breaks.

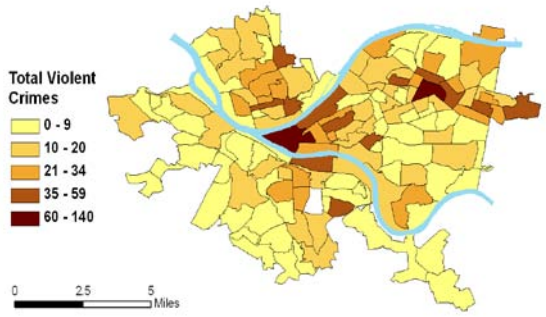

Figure 4. Violent crimes across census tracts, categorized using natural breaks.

When the census tract level crime data were collected in the year 2000, the data were gathered using the 1990 census boundaries, as the new census boundaries had not yet been released. In order to ensure the crime data could be appropriately compared to the 2000 census tract level demographic data, GIS was used determine the extent to which the census tracts had changed from 1990 to 2000. Conveniently, for the purposes of this study, the number of census tracts had decreased from 175 in 1990 to 141 in 2000, assumedly due to a considerable decrease in population in the city in the 1990's. In all cases, the "excess" 1990 census tracts were merely merged to form larger year 2000 census tracts. Using the ArcToolbox Dissolve geoprocessing tool, the 1990 census tracts and their associated crime data were dissolved to correspond to the new 2000 census tract boundaries.

More than 100 socioeconomic and demographic variables were examined in the Summary File 3 data, and ultimately 11 variables were selected. It was believed that these best supported the analysis of this study's goal of focusing on socioeconomic status and demographic variables which are not constant for any given individual or group, such as race, but variable, such as employment status, age and educational attainment (Appendix A). The two variables of "Percent Males of the age of 14 to 29 Years" and "Percent Females of the age of 14 to 29 Years" were created by combining individual age and/or grouped age data covering the age range of 14 to 29 years. It is this age range that is most commonly associated with criminal activity. Furthermore, the "Percent Vacant" housing variable stemmed from the combining of the variables containing data about houses that were vacant due to "for rent," "for sale," or "abandoned" status. Houses vacant due to their owners using the houses for seasonal or recreational use or houses used by seasonal or migrant workers were not included, as these situations indicate the houses still serve an active purpose. Finally, the educational attainment variable of "Percent Bachelors Degree or Higher" was derived by combining the variables of relevant education level.

Analysis

Stepwise Multiple Linear Regression

Using SPSS software, several stepwise multiple linear regression analyses were used to identify which, if any, of the 11 socioeconomic and demographic variables were related to the violent crime and property crime categorical variables. Certain assumptions or limitations always exist when using regression analysis (Berman, 2002). Since multiple regression analysis uses a linear analysis, the assumption is the data are related in a linear fashion (Berman). For example, it was initially assumed as one independent variable (SES/demographic) increases or decreases, the dependent variable would

also increase or decrease, depending on the particular association, in consistent amounts. The variables must also be continuous (non categorical), and in the case of this study, all variables were continuous. Since data for over 100 census tracts were used (141 census tracts), it is confirmed enough samples have been utilized for an appropriate analysis (Berman). The analyses resulted in models showing how much of the variation in a dependent crime variable was explained by the independent variables, individually and cumulatively, and expressed by the Coefficient of Determination (R^2) values. The SES and demographic variables served as independent (predictor) variables, while the property and violent crime variables served as the dependent variables. In the stepwise analysis, one independent variable (SES/demographic) is first added to the regression model and is only retained only if the variable proves to have a significant relationship with the dependent (crime category) variable. Each subsequent independent variable is added to the model and removed from the model if an insignificant relationship is found until all independent variables have been tested for significance.

GIS Hot-Spot Analysis

As applied to crime analysis, "hot-spot mapping is the spatial representation of areas with high concentrations of crime" (Vann and Garson, 2001). In order to move beyond the presentation of attribute values in the form of choropleth maps, hot-spot analyses were conducted for the two categorical total property and total violent crime variables and the statistically significant related socioeconomic status and demographic variables. The limitations of choropleth mapping were that while one could examine distributions of variable values across census tracts, statistically significant relationships found between the independent and dependent variables were not necessarily evident upon visual examination of the maps. The Getis-Ord Gi* spatial statistics method was implemented for hot-spot analyses, which is the method ArcGIS uses for these analyses. Utilizing the Hot Spot Analysis tool in the Spatial Statistics Tools/Mapping Clusters, analyses were conducted for statistically significant variables. Each analysis resulted in the creation of an output feature class representing standardized G scores, and ArcMap automatically symbolized classes of scores to show differences between hot spots, those clusters exceeding 1.65 standard deviations, and in select cases, cold spots, depicting clusters falling below -1.65 standard deviations.

Results

SES and Demographic Variables Related to Property and Violent Crime

Property Crime

The stepwise multiple linear regression analysis was used to compare the independent SES/demographic variables. Total property crime as related to the percent males of the age 14-29 years and the percent unemployed variables were statistically significantly (Table 2). The regression model predicts only 11.8% of the variation in property crime across the census tracts, as indicated by the R^2 value. The strongest predictor of property crime was the percent male age 14 to 29 years variable (PerMale1429) which explained 8.6% of the variation in

the property crime with a positive relationship.

Table 2. For property crime, the multiple linear regression model steps including R-square values, standardized coefficients, and level of significance are shown.

Step	Variable	R^2	Beta	Sig
1	PerMale1429	.086	.256	.002
2	PerUnempl	.118	.183	.028

Violent Crime

The stepwise multiple linear regression analysis used to compare the independent SES/demographic variables with the total violent crime variable established that the percent unemployed and the median household income variables were statistically significantly related to total violent crime (Table 3). Both the percent labor force unemployed (PerUnempl) and median household income (MedHouse) variables are nearly equal predictors of violent crime explaining a combined 21.1% of the variation in violent crime across census tracts as indicated by the R^2 value. However, the negative standardized coefficient (Beta) associated with the median household income variable suggests an inverse relationship with violent crime. Therefore, as inferred from the analysis, as median household income increases, violent crime decreases and vice versa.

Table 3. For violent crime, the multiple linear regression model steps including R-square values, standardized coefficients, and level of significance are shown.

Step	Variable	R^2	Beta	Sig
1	PerUnempl	.117	.316	.000
2	MedHouse	.211	-.303	.000

Hotspot Analysis

Property Crime

Having determined the socioeconomic status/demographic variables statistically related to property crime, the strongest predictor variable of property crime, percent males of the age 14-29 years, was mapped in choropleth form (Figure 5). This map was compared to the total property crimes map for census tracts (Figure 6). Although a statistically significant relationship had been found to exist between the two variables, a simple comparison of choropleth maps did not appear to confirm the relationship.

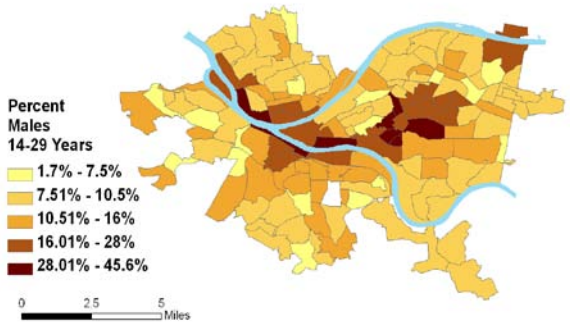

Figure 5. Percent Males age 14-29 years across census tracts, categorized using natural breaks.

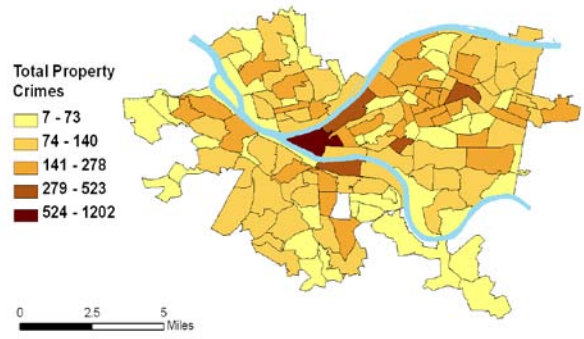

Figure 6. Property crimes across census tracts, categorized using natural breaks.

When maps resulting from the hotspot analyses were compared, it is possible to notice the high or low concentrations of variable values (Figure 7 and Figure 8).

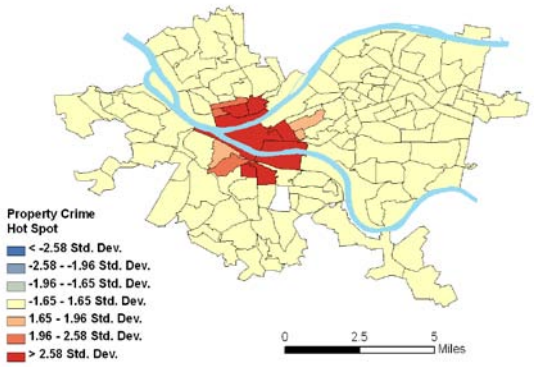

Figure 7. Hotspot, represented in red, for the total property crime variable.

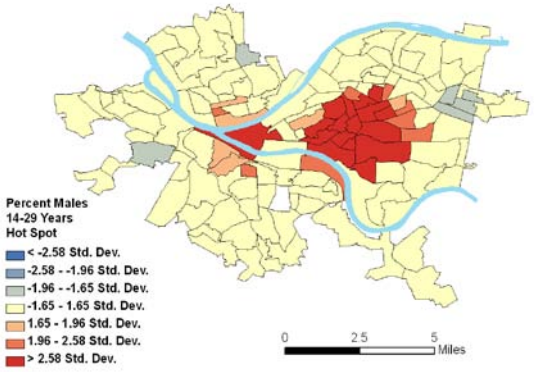

Figure 8. Hotspots for the percent males age 14 to 29 years variable, showing concentrations of this age/gender group.

Additionally, the hotspot analysis result for the percent unemployed variable is provided (Figure 9). Notice that all three variables have hotspots, or high concentrations of values, located near the downtown/ central area. It is important to note that because the SES and demographic variable values were converted to percentages, the hotspot concentrations were not a result of the fact that certain census tracts had sometimes markedly higher populations than others.

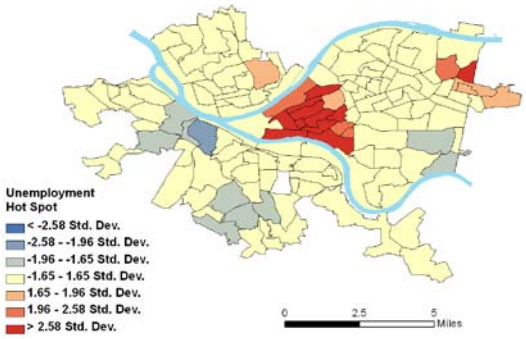

Figure 9. Hotspots (red) and coldspots (blue) presented for the percent unemployed variable. The coldspots represent census tract clusters having low concentrations of unemployed individuals.

Violent Crime

Similarly to the results of the examination of choropleth maps depicting the distribution of property crime and its related variables, the investigation of choropleth maps for violent crime in its related variables did not prove to be an adequate method to visualize the relationships between the variables. Using hotspot analysis, the statistically significant relationship between violent crime and median household income becomes more apparent (Figure 10 and Figure 11). Notice while a violent crime hotspot exists in the central/downtown area, a median household income coldspot is located in the same area. This is to be expected since the relationship between violent crime and income were found to be an inverse relationship in that a decrease in household income was associated with an increase in violent crime. Therefore, in the downtown area,

the cluster having high concentrations of violent crime lies in close proximity to the cluster having concentrations of low income households.

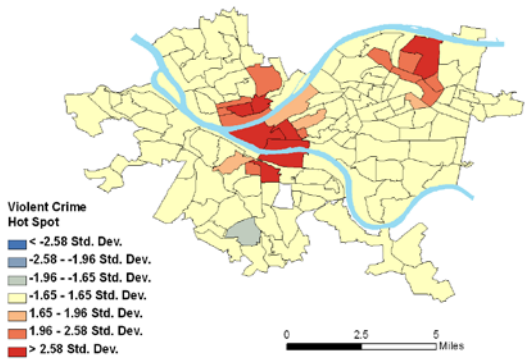

Figure 10. Hotspots for the total violent crime variable.

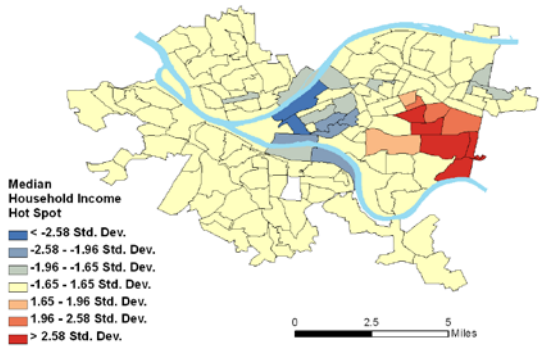

Figure 11. Coldspots and Hotspots for the Median Household Income variable.

Discussion

Significant Variables

Of the 11 variables investigated, 3 variables were found to have statistically significant relationships with crime: percent males 14-29 years (property crime), median household income (violent crime) and percent unemployed (both crime categories). It is widely known that males of the years 14-29 commit a disproportionately large portion of crimes, and for property crimes, this is confirmed in this study. The corresponding female age group served as a control, since it was an initial concern that the high concentrations of both males and females age 14-29 near the major universities downtown and in the Oakland area to the east would somehow skew results. It is unknown whether this similarity in spatial concentrations for both gender groups impacted the violent crime model.

Of particular interest is the fact it was established that unemployment has a significant positive relationship with both violent and property crime. One of the primary goals of this study was to investigate SES/demographic variables that can be changed and positively impacted through public policy, and unemployment is a good example of this type of variable. The results suggest successful job creation policies should carry the added benefit of a reduction in property and violent crime in the areas where unemployment is reduced.

Regression Model Limitations

For each of the two crime categories, two SES/demographic variables were found to have a significant relationship with the category. For property crime, the explaining variables only account for, as indicated by the R^2 value for the model, 11.8% of the variation in property crime across the census tracts. This means the remaining 88.2% of the variation is unexplained – a substantial amount – and confirm this regression model used is limited. The violent crime model accounted for 21.1% of the variation in violent crime across the census tracts, leaving 78.9% of the variation unexplained. The addition of more demographic variables, such as

race, additional age groups, national origin, or religion could possibly serve to create a stronger model, or perhaps SES/demographic variables are simply limited in their potential as predictors of crime. Another consideration is the use of the census tract as the geographic unit of analysis in crime research has limitations.

Hotspots

A comparison of Figure 7 and Figure 10 show hotspots for both violent crimes and property crimes exist in the downtown area and neighboring census tracts. Since each day, large numbers of people move into the area to work or for recreation, it is possible that the daily influx of commuters and visitors to the area contributes to crime totals. This possibility reveals one limitation in spatial crime analysis. People who live within a certain geographic boundary, such as a neighborhood or census tract, do not commit crimes exclusively within their home areas. However, an additional violent crime hotspot exists to the north and east of the downtown area, and this occurrence may not be explained by dramatic daily shifts in population concentrations that are experienced downtown. Examination of the data also did not reveal why no property crime hotspots were also present in the northeast area of the city, and further analyses of the data in the northeast area using smaller geographic units of analysis, such as block groups, could provide an explanation.

Acknowledgements

I wish to thank Dr. Dave McConville, Patrick Thorsell and John Ebert of the Department of Resource Analysis for implementing an excellent program and for their instruction and guidance. I would also like to thank my fellow students in the GIS program who greatly contributed to the learning experience.

References

Ackerman, W. V., and Murray, A. T. 2004. Assessing spatial patterns of crime in Lima, Ohio. *Cities, 21(5)*, 423-437. Retrieved January 20, 2009 from http://search.ebscohost.com.xxproxy.smumn.edu/login.aspx?direct=true&db=keh&AN=15425607&site=ehost-live.

Allen, R. C. 1996. Socioeconomic Conditions and Property Crime: A Comprehensive Review and Test of the Professional Literature. *American Journal of Economics and Sociology, 55(3)*, 293-305. Retrieved February 3, 2009, from http://search.ebscohost.com.xxproxy.smumn.edu/login.aspx?direct=true&db=keh&AN=9609246085&site=ehost-live.

Berman, E. M. 2002. *Essential Statistics for Public Managers and Policy Analysts*. Washington, D.C. Congressional Quarterly, Inc.

Cohen, J., and Gorr, W. L. 2005. Development of Crime Forecasting and Mapping Systems for Use by Police in Pittsburgh, Pennsylvania, and Rochester, New York, 1990-2001 [Data File]. Retrieved February 10, 2009 from the National Archive of Criminal Justice Data (ICPSR) distributor site http://www.icpsr.umich.edu/icpsrweb/NACJD/studies/4545?q=NACJD.

National*Atlas*. 2011. Summary of the Uniform Crime Reporting Program. Retrieved January 22, 2011 from http://www.nationalatlas.gov/articles/people/a_crimereport.htm.

Vann, I. B., and Garson, G. D. 2001. Crime Mapping and Its Extension to

Social Science Analysis. *Social Science Computer Review, 2001, 19(4)*, 471-479.

Appendix A. Socioeconomic Status and Demographic Variables of the United States Census Bureau's Summary File 3 used in the Stepwise Multiple Linear Regression Model.

1) PerMale1429: Percent males of the total population of the age of 14 to 29 years. Calculated from the single and categorical age data for the age range.

2) PerFem1429: Percent females of the total population of the age of 14 to 29 years. Calculated from the single and categorical age data for the age range.

3) PerUnemp: Percent of all those participating in the labor force who are unemployed.

4) PerLessHigh: Percent of those age 25+ with no high school diploma or GED equivalent.

5) PerBachPlus: Percent of those above the age of 25 holding a Bachelors degree or higher.

6) MedHouse: Median household income.

7) PerPubAst: Percent of households receiving public assistance income.

8) PerPoverty: Percent of households with an income level at or below poverty level.

9) PerVacant: Percent of residences vacant due to a for-rent, for-sale, or abandonment status. Residences vacant due to the fact that owners were overseas or that the residence is primarily used by migrant workers were not included.

10) PerFore10: Percent of those foreign born and having arrived to the U.S. within the last 10 years. Even non-U.S. citizens are counted in the census.

11) PerFamFem: Percent of households having a female head-of-household and with no male present.