A Rape Arrest Cluster Analysis for the USA with Demographic Adjustments Based on Poisson Regression

By

Nicole Susanne Richards
B.A. (Mathematics), University of Hawai'i at Mãnoa, 2003
B.A. (Sociology), University of Hawai'i at Mãnoa, 2003

Advisor: Dr. Raid Amin

A Graduate Proseminar
In Partial Fulfillment of the Degree of
Master of Science in Mathematical Sciences
The University of West Florida
November 2010

The Proseminar of Nicole Susanne Richards is approved:

_____ _____

Dr. Raid Amin, Ph.D., Proseminar Advisor Date

_____ _____

Dr. Anthony Okafor, Ph.D., Proseminar Committee Chair Date

Accepted for the Department:

_____ _____

Dr. Kuiyuan Li, Ph.D., Chair Date

TABLE OF CONTENTS

Page

On average, 14,172,384 people are arrested each year in the United States [13]. The FBI compiles arrest data and it is published in the *Uniform Crime Reporting* (UCR) *Guide Handbook* [2], which includes information on twenty-nine different crimes, including rape. Rape affects the lives of thousands of people each year, and the number of reported instances has been steadily increasing over the past decade [14].

Statement of Problem

A cluster analysis was performed in SaTScan™, taking specific demographic variables into account while detecting high risk areas of rape across the United States. Rape [3] is defined by the FBI as "the carnal knowledge of a female forcibly and against her will." This current definition was developed in 1927 [3], and only includes nonconsensual penile penetration of the vagina.

Relevance of Problem

One out of every six American women has been the victim of an attempted or completed rape in her lifetime [11]. The results of this study could provide law enforcement opportunity to allocate resources to specifically target high-risk areas.

Literature Review

SaTScan

This study consisted of spatial and space-time analyses to evaluate the existence of high-risk rape clusters in the United States using a statistical software program called SaTScan™. SaTScan™ [4] uses spatial scan statistics to identify and test for the significance of crime clusters. The crime count (arrest count) in each county are used in both a purely spatial analysis (two dimensions, geographically) and a space-time analysis (three dimensions, geographically across time).

The spatial scan statistic in SaTScan™ creates a circular window that moves over the map, including the centroid of each county. For each window, this spatial scan statistic tests the null hypothesis that there is an equal risk of that crime occurring in all counties, versus the alternative hypothesis that there exists an elevated risk of that crime within the scan window versus the areas outside the window.

For the particular data set, it is assumed the crime arrest counts were rare events, therefore being distributed according to a Poisson model. The likelihood function for the Poisson model can be shown proportional to [4]:

$$\left(\frac{c}{E[c]}\right)^{c}\left(\frac{C-c}{C-E[c]}\right)^{C-c} I()$$

C: *total number of arrests in the United States*
c: *observed number of arrests within the scan window*
E[c]: *expected number arrests within the window under the null*
C-E[c]: *expected number of arrests outside the window*
I(): *indicator function*

If there exists an elevated crime risk when using the one-tailed test, the null is rejected, and the indicator function is equal to one. By a Monte Carlo simulation, a test statistic is created for each random replication that SaTScan™ generates under the null, as well as for the real data set, resulting in a p-value given to each cluster listed in order of significance. For this study, a significant cluster is defined as $p < 0.01$.

Poisson Regression

The Poisson distribution can be applied to data in which the dependent variable is a count and a rare event. The Poisson probability distribution [10] is as follows:

$$Y \sim Poi(\mu)$$

$$Pr(Y = y) = \frac{e^{-\mu}\mu^{y}}{y!}$$

y = 0,1,2,… (number of occurrences)
μ = positive real number, equal to the average number of occurrences during the given time interval
Pr(Y=y) denotes the probability that the outcome is Y

There is a single parameter, μ, which is both the mean and the variance [7]. The distribution is discrete and is used to model the number of events within a given time interval [9]. In this particular model we have:

- *dependent variable*: number of arrests per county = **count**
- *event*: rape = **rare occurrence**

Therefore, the model is based on a Poisson distribution and is analyzed with Poisson regression. Poisson regression [8] is a form of a generalized lineal model (GLM) where the response variable takes on a Poisson distribution. Let's say Y_i is the observed count for the experimental unit *i*. Then we have [10]:

$$Y_i|X_i \sim Poi(\mu_i)$$
$$log(\mu_i) = X_i\beta$$

The log link is most commonly used, and indicates that the covariates influence the mean of the counts (μ) in a multiplicative way. Often we model the count of events within a particular time period, particular region, or particular risk group of people. Therefore, what is of interest is to model the *rate*. Given a specific time period *t*, we model the events occurring in time period *t*. Thus, μ is better described as $\mu = \lambda * t$ where λ is the rate of events [10]:

$$Y_i/X_i \sim Poi(\lambda_i * t_i)$$
$$log(\lambda_i) = X_i\beta$$
$$log(\mu_i/t_i) = log(\mu_i) - log(t_i) = X_i\beta$$
$$log(\mu_i) = log(t_i) + X_i\beta$$

The term $log(t_i)$ is known as the offset variable and it provides the adjustment for the variable risk sets [10], i.e. variable numbers of people at risk in the population. It can be thought of as a predictor variable, but it does not have a parameter in front of it to be estimated, so it is treated different in software programs like SAS.

Limitations

The limitations of the study can manipulate the results obtained from the SaTScan™ analysis. To minimize inappropriate influence, data was eliminated and modified. Lack of rape count data for Florida and Illinois led to the elimination of those states from the study, along with Alaska and Hawaii. Alaska and Hawaii did offer rape counts; however, these states are not part of the continuous forty-eight states. Therefore, they were eliminated because their presence in the study would have skewed the clustering results due to their location. Certain counties were also eliminated from the study because their populations were either never recorded or the population listed was zero: Essex county, Vermont; Yellowstone National Park; Clifton Forge City, Virginia; South Boston City, Virginia; and Issaquena, Mississippi. Counties containing the code '999' due to their population data were each removed as well in the following states: Connecticut, New Jersey, and Vermont.

The *National Archive of Criminal Justice Data* [5] report (NACJD) contains arrest data from 1989-2007. Originally the study focused on rape arrests from 1995-2007. Because of huge areas of missing data from 1995-2002 in Kansas, Montana, New

Hampshire, Vermont, Washington D.C., Wisconsin, Delaware, Kentucky, Mississippi, and South Dakota, the years were narrowed down to 2003-2007. In this study the year 2007 determines current clusters.

Some of the covariate data did not contain information for each individual year. Therefore, data was limited for certain demographic variables. For example, the illiteracy rate covariate was only available for the year 2003. Since only one year was available for use, it was used repeatedly for 2003-2007. The unemployment rate data was incomplete for seven parishes/counties in Louisiana in 2005 and 2006 as a result of hurricane Katrina: Jefferson, Orleans, Plaquemines, St. Bernard, St. Charles, St. John the Baptist, and St. Tammany. The average of the data in 2003, 2004, and 2007 for each parish was determined and filled in for 2005 and 2006. This created a complete covariate data set for SaTScan™ analysis.

According to a statistical average over the past five years, 60% of rapes are not reported to the police [11]. Without these arrest counts included in the analysis, results can be skewed.

Lastly, the FBI's definition of rape is very narrow and only includes nonconsensual penile penetration of the vagina. This problematic definition excludes an extraordinary number of sexual assault acts that would most likely be considered rape by any rational adult.

Study Area and Population

In this study, 2,938 counties were identified in the United States between the years 2003-2007. The population included the entire population of each participating

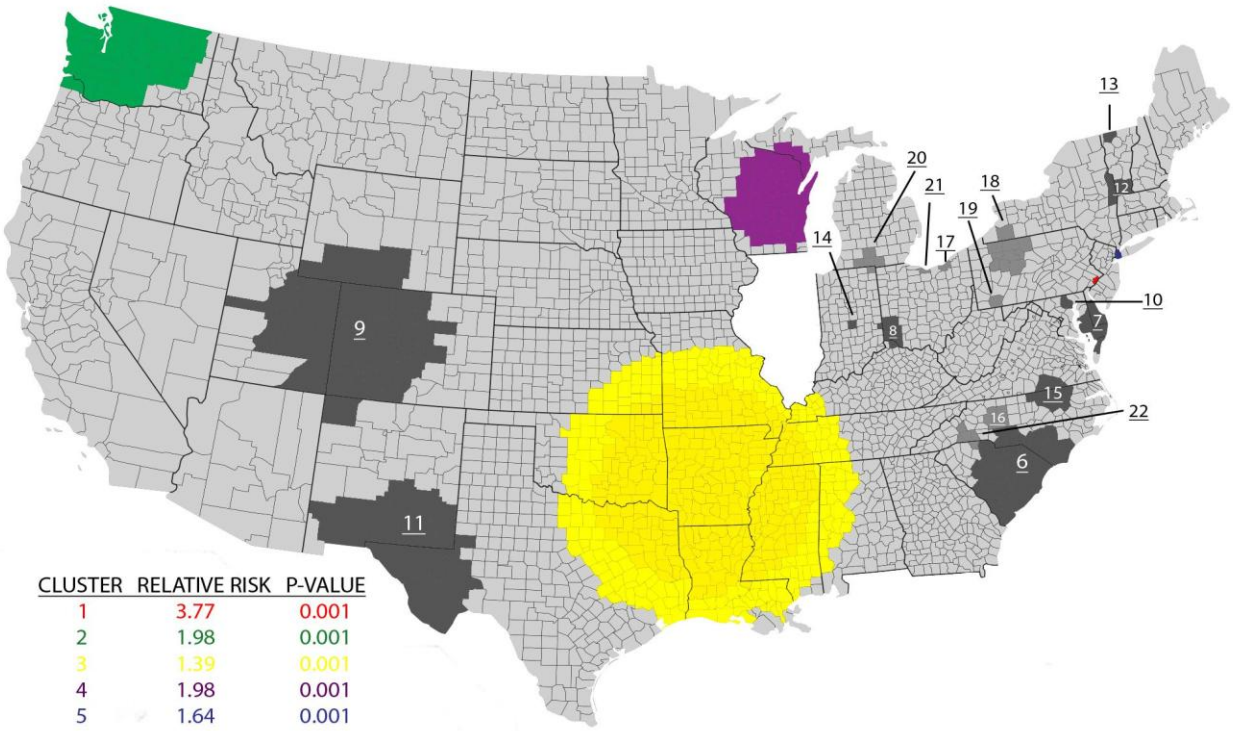
state and county, and during this time period, there were 110,145 arrests documented for rape.

Testing and Analysis

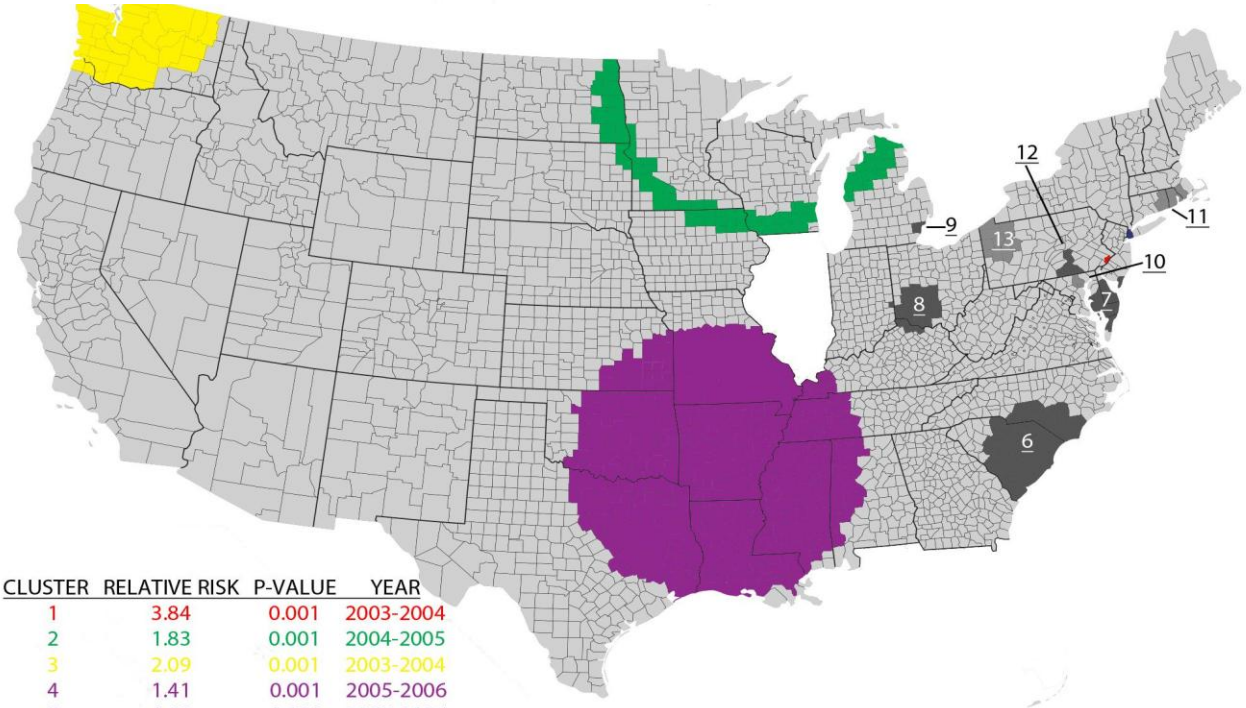
2003-2007, No Covariates

SaTScan™ needs three vital pieces of information to run without error: the rape count, the population of each county, and the latitude/longitude of each county (calculated by the center of each county). A purely spatial analysis (Figure 1) showed that Philadelphia, Pennsylvania was the most likely cluster with a relative risk of 3.77 and a p-value of 0.001.

A space-time analysis (Figure 2) revealed that Philadelphia, Pennsylvania remained the most likely cluster, with the relative risk going up to 3.84. It occurred from 2003-2004, and had a p-value of 0.001. Cluster seven, which included counties in Maryland, Delaware, Virginia, and New Jersey, had a current cluster from 2006-2007, with a relative risk of 2.44 and a p-value of 0.001. However, the most interesting cluster is cluster four, which appeared in the region of New Orleans, Louisiana. It had a relative risk of 1.44, p-value of 0.001, and occurred in 2005-2006…the time of hurricane Katrina. Thornton and Voigt state [12], "…the conditions [after hurricane Katrina] that were conducive for crimes such as rape have not existed to this magnitude in modern times in America." The most likely cluster is extremely important; however, determining if a "Katrina cluster" exists is equally as interesting. Therefore, the focus in this study is on the top three clusters in the spatial analysis.



CLUSTER	RELATIVE RISK	P-VALUE
1	3.77	0.001
2	1.98	0.001
3	1.39	0.001
4	1.98	0.001
5	1.64	0.001



CLUSTER	RELATIVE RISK	P-VALUE	YEAR
1	3.84	0.001	2003-2004
2	1.83	0.001	2004-2005
3	2.09	0.001	2003-2004
4	1.41	0.001	2005-2006
5	1.69	0.001	2005-2006

Regions and Divisions

In order to study the top three clusters from the purely spatial analysis, the United States was broken down into regions and divisions [1], taken from the Census Bureau (Figure 3). Cluster one (the most likely cluster) was taken from the Middle Atlantic Division, cluster two was taken from the Pacific Division, and cluster three (the "Katrina cluster") was taken from the East South Central, West South Central, and West North Central Divisions.

Cluster Analyses

Ten different covariates were selected based on demographics such as socioeconomic status, race, and the density of the county in question. These covariates include: median age, unemployment, illiteracy rate, drop out rate, median household

income, divorce rate, poverty rate, African American rate, male population rate, and density.

Before a cluster analysis could be performed in SaTScan™, several steps needed to be completed to ensure the validity of the results. These steps were performed using Statistical Analysis Software, commonly known as SAS. Four programs were run on each cluster's data set in order to execute the analyses:

1. *Poisson Regression*: All ten covariates were taken along with the rape counts and it was determined whether the covariates were significant or not (Appendix A on disc). All insignificant covariates were discarded.
2. *Model Selection*: All ten covariates were taken along with the rape counts and the best three covariates were determined through a high adjusted R^2, a low MSE, and a low CP (Appendix B on disc).
3. *Multicollinearity*: All ten covariates we taken along with the rape counts and tested for multicollinearity, i.e. one variable being a linear combination of another variable (Appendix C on disc). Any covariate with a variable inflation rate (VIF) greater than 20 was discarded.
4. *Poisson Regression for Predicted Values*: Each cluster had three analyses: one covariate, two covariates, and three covariates. For each analysis, SAS was run to find the predicted values for each combination of covariates. The predicted values take into account the log of the population (the offset variable mentioned earlier), and is a rate that is based on adjusting for the covariate(s). This is used in place of the population file in SaTScan™ (Appendix D on disc).

Cluster 1

Model selection determined poverty rate, African American rate, and male population rate to be the top three covariates for the Middle Atlantic Division.

Poverty Rate

A purely spatial analysis on the Middle Atlantic Division with poverty rate as the covariate (Figure 4) showed twenty-three counties in New Jersey, New York, and Pennsylvania (including Philadelphia) were in the most likely cluster with a relative risk of 2.33 and a p-value of 0.001.

A space-time analysis (Figure 5) showed that the most likely cluster included the same twenty-three counties, with the relative risk going down to 1.88. It still occurred from 2003-2004 with a p-value of 0.001, as it did when the entire US was analyzed with no covariates. There were no current clusters.

CLUSTER	RELATIVE RISK	P-VALUE
1	2.33	0.001
2	1.97	0.001
3	1.36	0.001
4	1.59	0.001
5	1.30	0.001

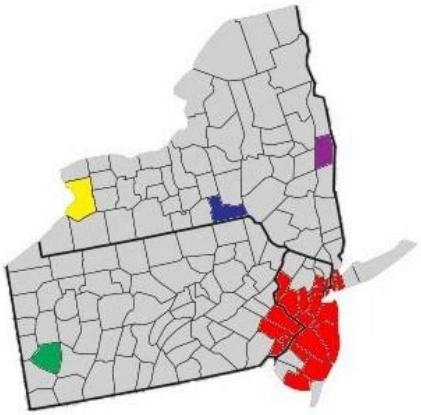


Figure 4: Spatial Analysis-Poverty

CLUSTER	RELATIVE RISK	YEAR	P-VALUE
1	1.88	2003-2004	0.001
2	2.31	2003-2004	0.001
3	1.59	2005-2006	0.001
4	2.10	2003-2003	0.001
5	1.39	2003-2004	0.001

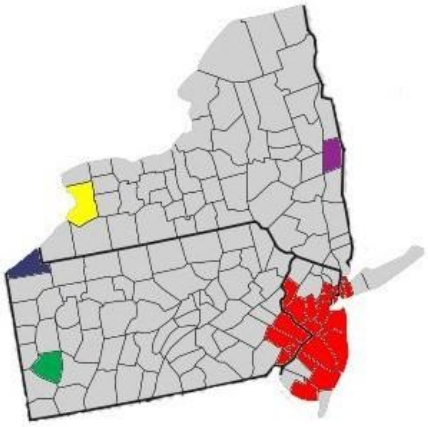


Figure 5: Space-Time Analysis-Poverty

Poverty Rate, African American Rate

When the African American rate covariate was added, a purely spatial analysis showed that only Queens and New York counties in New York remained in the most likely cluster (Figure 6). There was a relative risk of 2.72 and a p-value of 0.001. Philadelphia bumped down to the second most likely cluster, with fifteen other counties in Pennsylvania. Cluster two had a relative risk of 1.35 and a p-value of 0.001.

A space-time analysis (Figure 7) showed that the most likely cluster included the same two counties in New York, with the relative risk going down to 2.69. It occurred from 2005-2006, with a p-value of 0.001. Philadelphia remained in the second most likely cluster; however the number of counties dropped to five. Even though Philadelphia jumped down to cluster two, it still remained in the 2003-2004 time period. The relative risk was 1.48 with a p-value of 0.001. There were no current clusters.

CLUSTER	RELATIVE RISK	P-VALUE
1	2.72	0.001
2	1.35	0.001
3	1.81	0.001
4	1.45	0.001
5	1.94	0.001

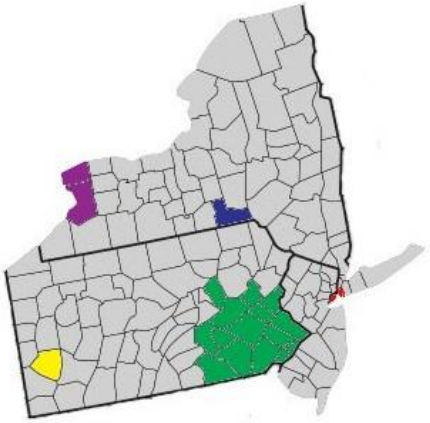


Figure 6: Spatial Analysis-Poverty, African American

CLUSTER	RELATIVE RISK	YEAR	P-VALUE
1	2.69	2005-2006	0.001
2	1.48	2003-2004	0.001
3	1.83	2003-2004	0.001
4	1.65	2005-2006	0.001
5	1.83	2003-2004	0.001

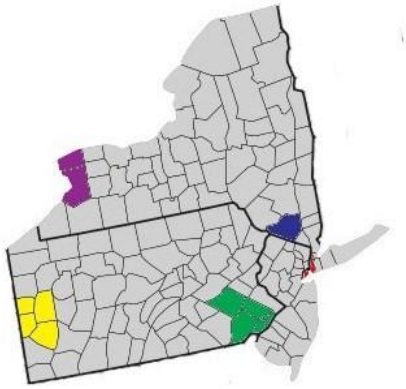


Figure 7: Space-Time Analysis-Poverty, African American

Poverty Rate, African American Rate, Male Population Rate

When the male population rate covariate was added, a purely spatial analysis showed that Queens and New York counties in New York remained as the most likely cluster (Figure 8). There was a relative risk of 2.66 and a p-value of 0.001. Philadelphia stayed in the second most likely cluster, with only one other county in Pennsylvania, Montgomery County. The relative risk resided at 1.35 and the p-value was 0.001.

A space-time analysis (Figure 9) showed Queens and New York counties resided in the most likely cluster again, with the relative risk going slightly down to 2.64. It

appeared from 2005-2006, with a p-value of 0.001. Philadelphia remained in the second most likely cluster with the number of counties increasing to five. Once more, cluster two still remained in the 2003-2004 time period. The relative risk was 1.40 with a p-value of 0.001. There were no current clusters.

CLUSTER	RELATIVE RISK	P-VALUE
1	2.66	0.001
2	1.35	0.001
3	2.16	0.001
4	1.90	0.001
5	1.48	0.001

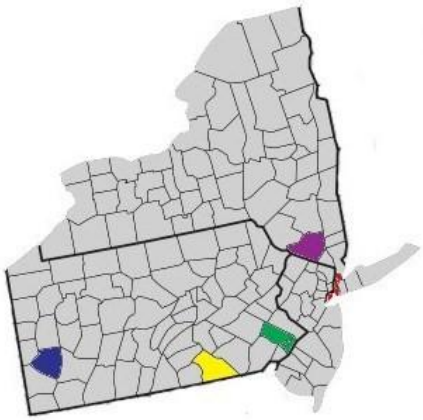


Figure 8: Spatial Analysis-Poverty, African American, Male Population

CLUSTER	RELATIVE RISK	YEAR	P-VALUE
1	2.64	2005-2006	0.001
2	1.40	2003-2004	0.001
3	2.92	2004-2005	0.001
4	1.90	2003-2003	0.001
5	2.30	2003-2004	0.001

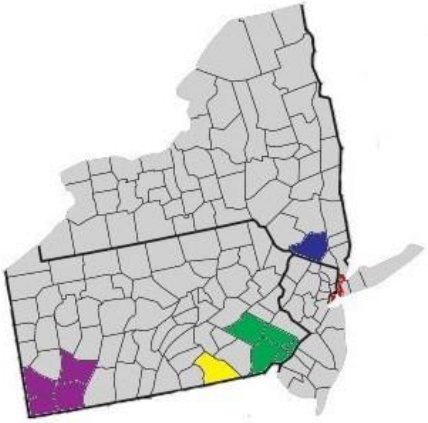


Figure 9: Space-Time Analysis-Poverty, African American, Male Population

Cluster 2

Model selection determined illiteracy rate, African American rate, and male population rate to be the top three covariates for the Pacific Division.

Illiteracy Rate

A purely spatial analysis with illiteracy rate as the covariate (Figure 10) produced three counties in Washington in the most likely cluster: King, Kitsap, and Snohomish. There was a relative risk of 4.07 and a p-value of 0.001. California showed up in the cluster analysis for the first time here, with Los Angles being the sole county in the second most likely cluster. It had a relative risk of 2.12 and a p-value of 0.001.

A space-time analysis (Figure 11) showed that the most likely cluster included the same three counties in Washington, with the relative risk going slightly down to 4.03. It occurred from 2003-2004, with a p-value of 0.001. There were no current clusters.

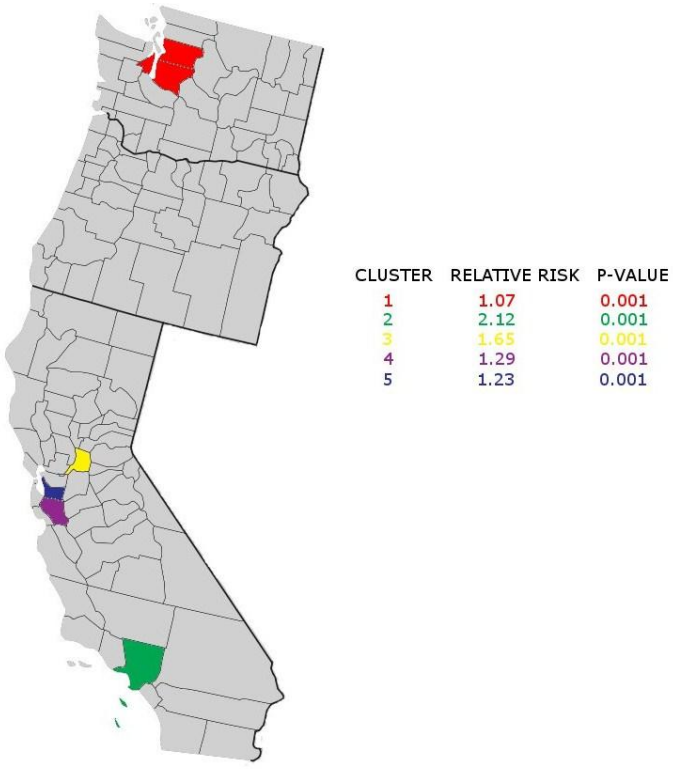


Figure 10: Spatial Analysis-Illiteracy

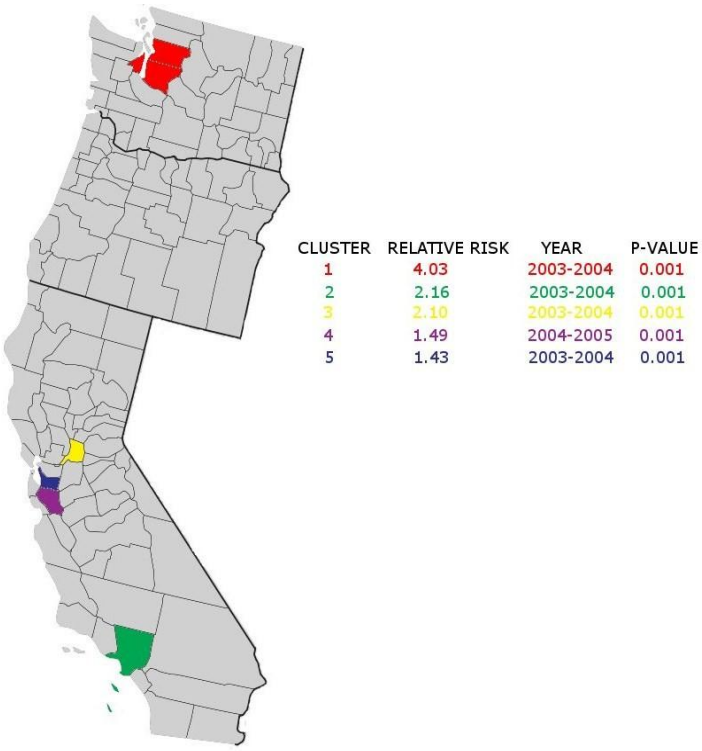


Figure 11: Space-Time Analysis-Illiteracy

Illiteracy Rate, African American Rate

When the African American rate covariate was added, a purely spatial analysis determined that three additional counties in Washington were in the most likely cluster, six counties in all (Figure 12). The relative risk went down to 3.31 and there was a p-value of 0.001. Los Angeles stayed in cluster two; however, two more counties in California appeared: Orange and San Bernardino counties. The relative risk went down to 1.57 and there was a p-value of 0.001.

A space-time analysis (Figure 13) showed that the most likely cluster included the same six counties in Washington, with the relative risk going down to 3.37. It occurred from 2003-2004, with a p-value of 0.001. There were no current clusters.

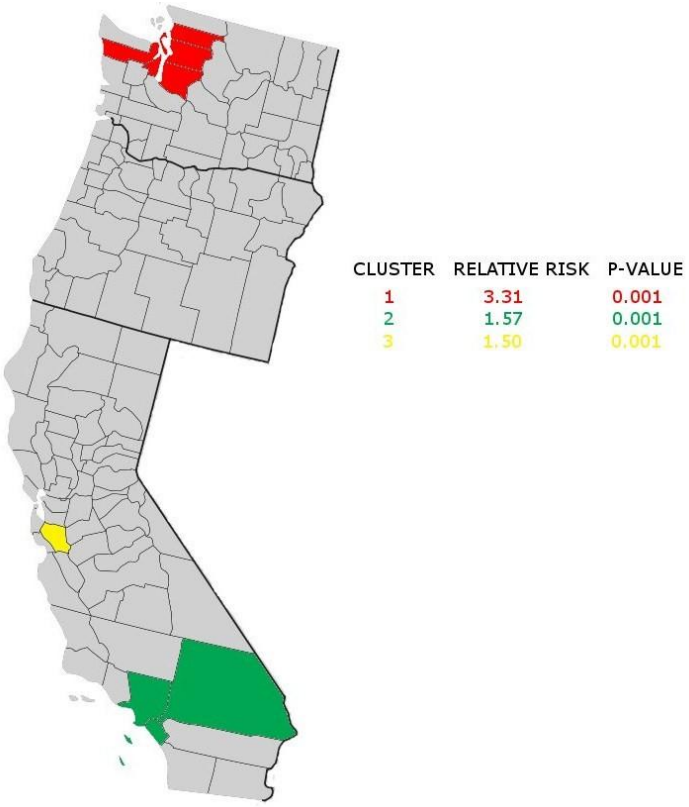


Figure 12: Spatial Analysis-Illiteracy, African American

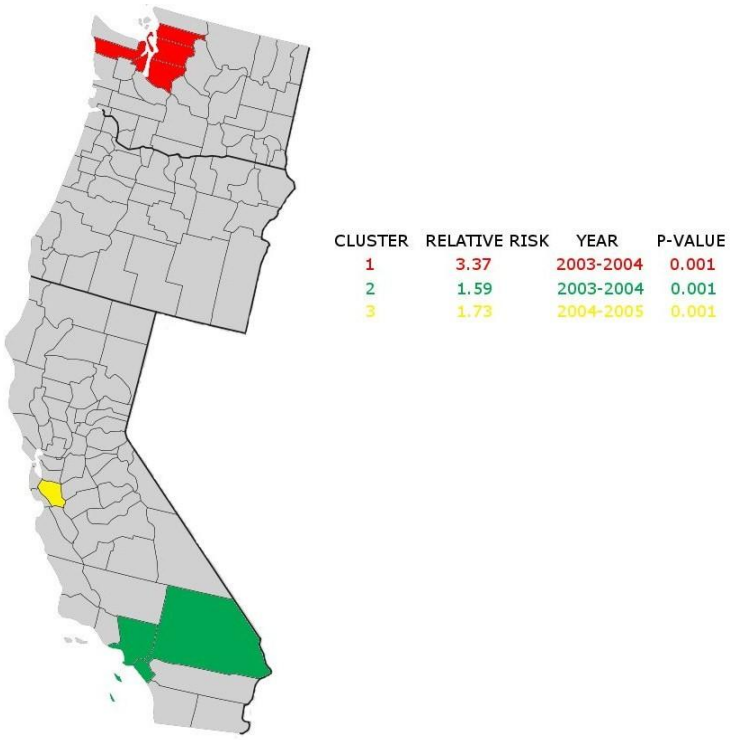


Figure 13: Space-Time Analysis-Illiteracy, African American

Illiteracy Rate, African American Rate, Male Population Rate

When the male population rate covariate was added, a purely spatial analysis determined that the three additional counties in Washington that appeared in the two-covariate analysis dropped back out of the most likely cluster (Figure 14). The relative risk went back up slightly to 3.73 and the p-value was 0.001. Los Angeles stayed in cluster two, along with Orange and San Bernardino Counties. The relative risk went up to 1.74 and it had a p-value of 0.001.

A space-time analysis (Figure 15) showed that the most likely cluster went *back* to the same six counties in Washington from the two-covariate analysis, with the relative risk going slightly up to 3.41. It still occurred from 2003-2004, with a p-value of 0.001. There were no current clusters.

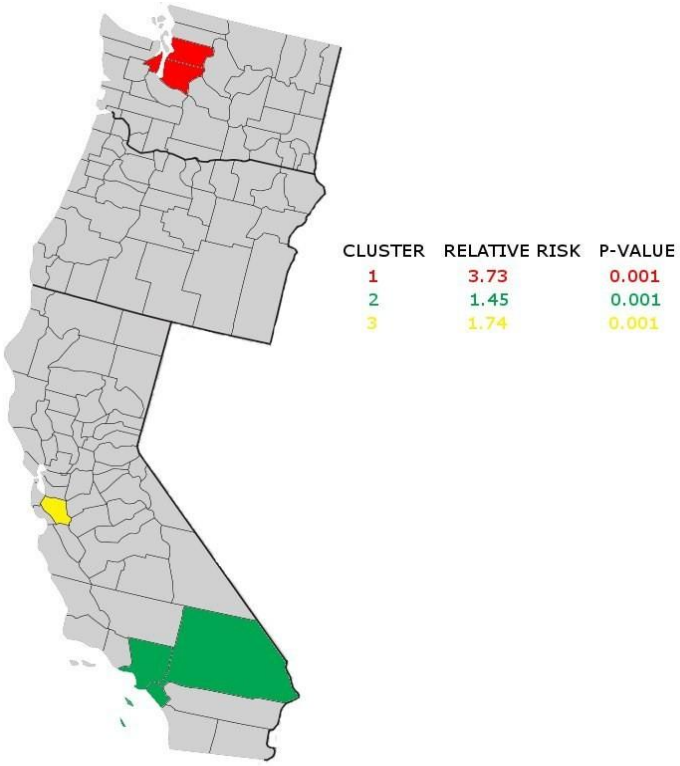


Figure 14: Spatial Analysis-Illiteracy, African American, Male

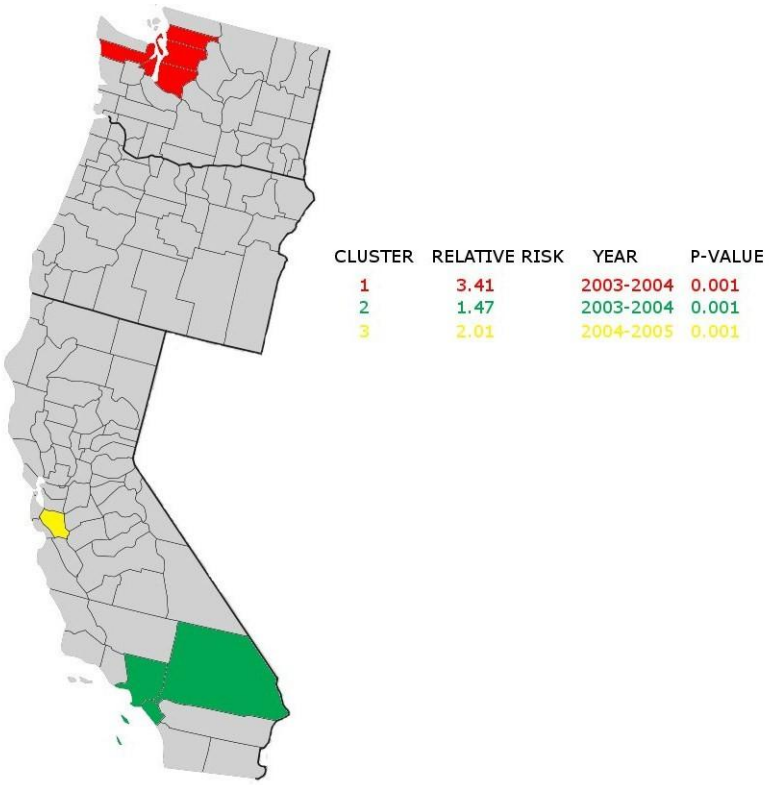


Figure 15: Space-Time Analysis-Illiteracy, African American, Male

Cluster 3

Model selection determined median age, median household income, and African American rate to be the top three covariates for the East South Central, West South Central, and West North Central Divisions.

Median Age

A purely spatial analysis with median age as the covariate (Figure 16) determined fourteen counties in Texas appeared in the most likely cluster. There was a relative risk of 4.72 and a p-value of 0.001. Orleans Parish, where New Orleans is located (direct impact from hurricane Katrina), and Jefferson Parish, the parish directly west of Orleans Parish, showed up in cluster five with numerous counties in Alabama, Mississippi, and Louisiana. There was a relative risk of 1.49 and a p-value of 0.001.

A space-time analysis (Figure 17) showed that the most likely cluster included the same fourteen counties in Texas, with the relative risk going slightly down to 4.61. It occurred from 2005-2006, with a p-value of 0.001. Orleans and Jefferson Parishes stayed in cluster five and occurred in 2005-2006 as well, during the aftermath of Katrina. There was a relative risk that rose to 1.51, and it had a p-value of 0.001. There were *five* current clusters here: 1) cluster three, consisting of only Saint Louis City, Missouri: relative risk of *8.61*, and a p-value of 0.001, 2) cluster seven, consisting of four counties in Nebraska: relative risk of 3.51, and a p-value of 0.001, 3) cluster eight, consisting of Crittenden, Arkansas and Shelby, Tennessee: relative risk of 2.92, and a p-value of 0.001, 4) cluster nine, consisting of only Sedgwick, Kansas: relative risk of 3.35, and a p-value of 0.001, and 5) cluster seventeen, consisting of only Sumner, Tennessee: relative risk of 2.54, and a p-value of 0.008. All current clusters occurred from 2006-2007.

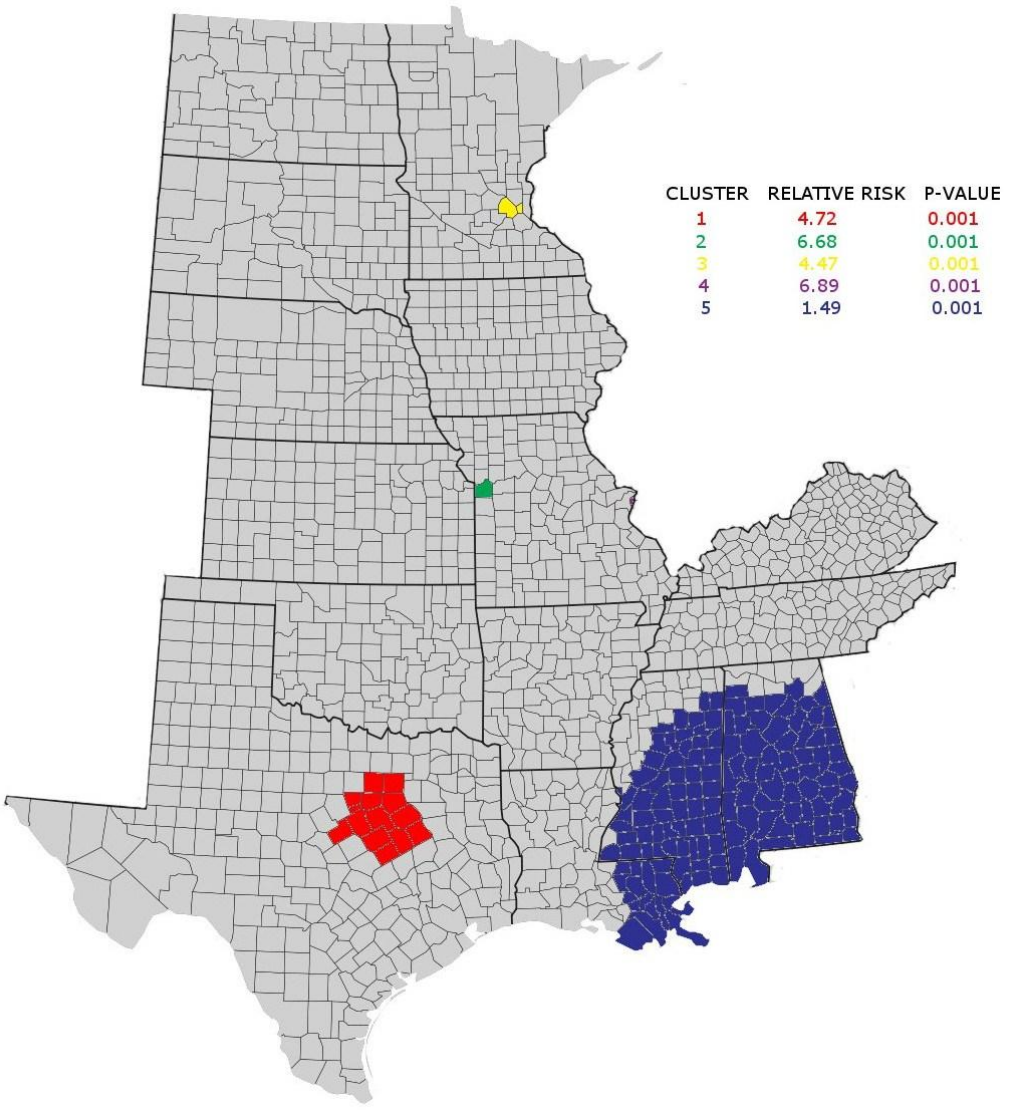


Figure 16: Spatial Analysis-Median Age

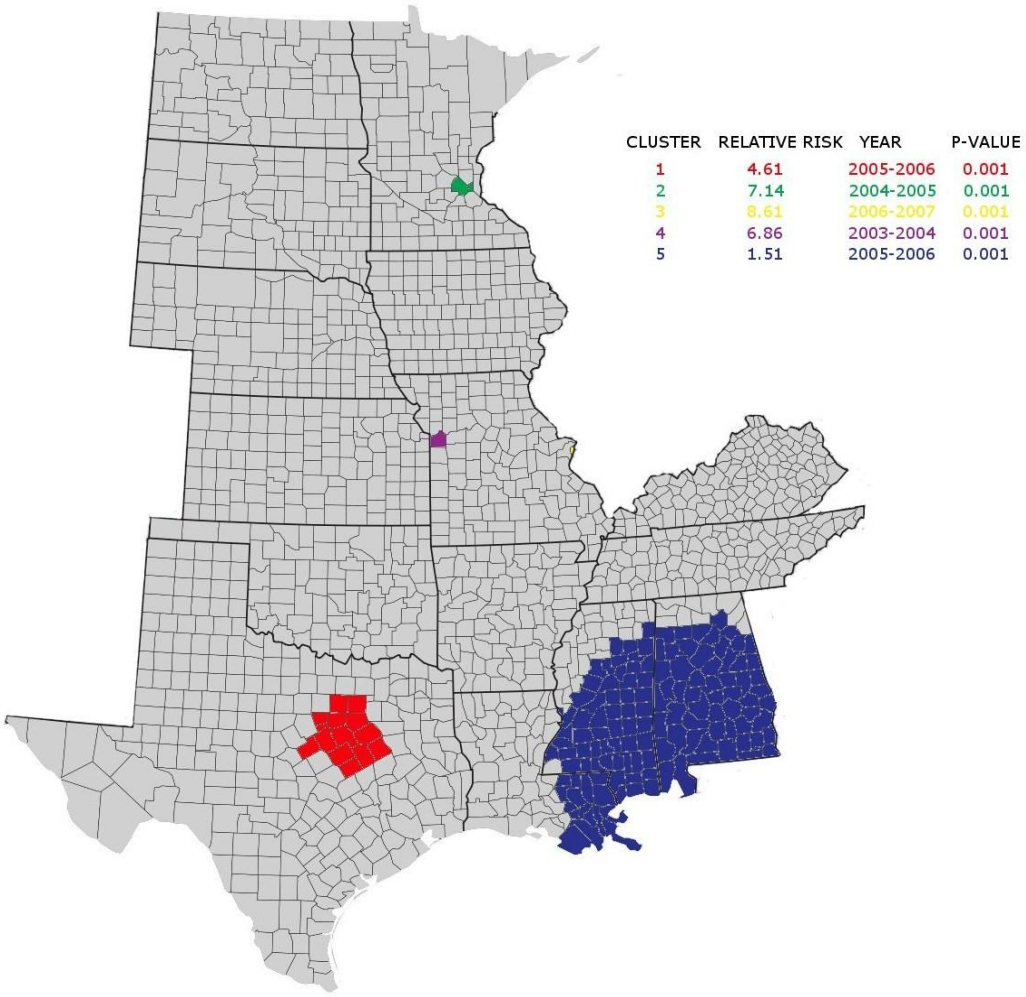


Figure 17: Space-Time Analysis-Median Age

Median Age, Median Household Income

When the median household income covariate was added (Figure 18), the fourteen counties in Texas stayed in the most likely cluster. The relative risk went down to 4.16 and the p-value was 0.001. Orleans and Jefferson Parish appeared in cluster five again with numerous counties in Alabama, Mississippi, Tennessee, and Louisiana as well. The relative risk went up to 1.61 and the p-value was 0.001.

A space-time analysis (Figure 19) showed that the most likely cluster included the identical fourteen counties in Texas, with the relative risk lowered to 4.18. The years the cluster appeared changed from 2005-2006 to 2003-2004, and there was a p-value of

0.001. Orleans and Jefferson Parishes did not show up in this analysis *at all*. There are *four* current clusters here, with Sumner, Tennessee dropping out: 1) cluster three, again consisting of only Saint Louis City, Missouri: relative risk that skyrocketed to *10.42*, and a p-value of 0.001, 2) cluster six, consisting of four counties in Nebraska: relative risk that went down to 2.80, and a p-value of 0.001, 3) cluster seven, consisting of only Sedgwick, Kansas: relative risk that lowered to 2.89, and a p-value of 0.001, and 4) cluster nine, consisting of Crittenden, Arkansas, ten other counties in Arkansas, and two counties in Tennessee: relative risk of 1.93, and a p-value of 0.001. All current clusters appeared from 2006-2007.

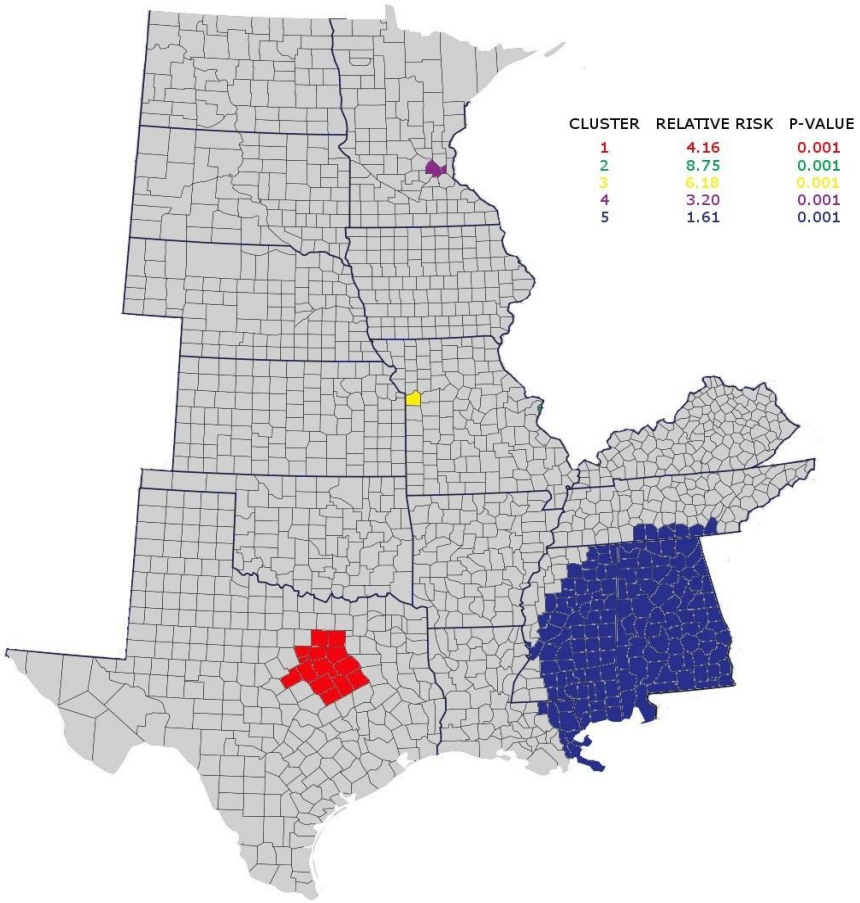


Figure 18: Spatial Analysis-Median Age, Median Household Income

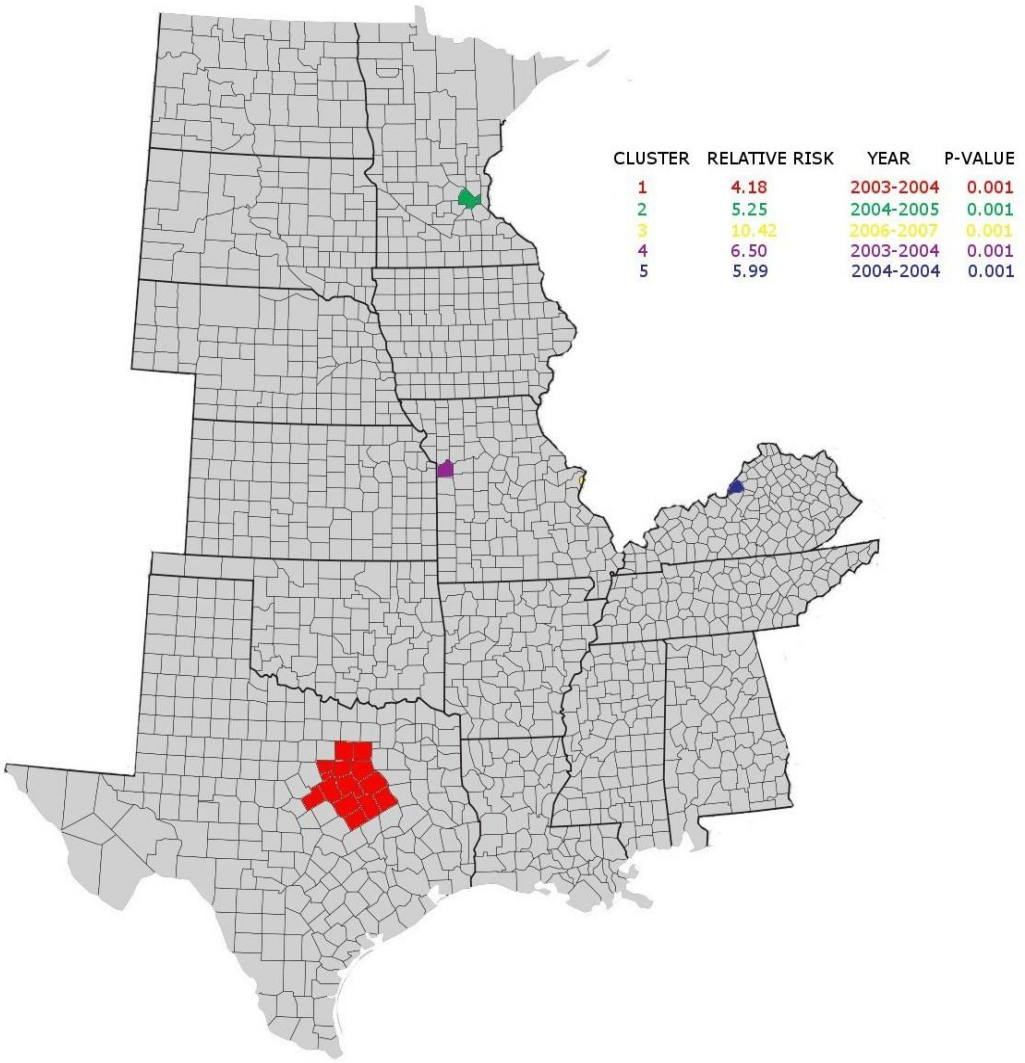


Figure 19: Space-Time Analysis-Median Age, Median Household Income

Median Age, Median Household Income, African American Rate

When the African American rate covariate was added (Figure 20), the most likely cluster consisted of the same fourteen counties in Texas. The relative risk went down to 3.84, and there was a p-value of 0.001. Orleans Parish dropped out of the analysis, and Jefferson Parish appeared in cluster six as a single county. The relative risk went up to *4.03*, and the p-value was 0.001.

A space-time analysis (Figure 21) included the equivalent fourteen counties in Texas as the most likely cluster, with a relative risk dropping to 3.97. It occurred from

2003-2004, with a p-value of 0.001. Jefferson Parish popped back up here in cluster four, unaccompanied by any other county, occurring in *2006*. The relative risk is an elevated *8.03*, with a p-value of 0.001. There is only one current cluster here, cluster five, consisting of simply Saint Louis City, Missouri. It had a current cluster from 2006-2007, a relative risk that lowered to 4.63, and a p-value of 0.001. Saint Louis City was the only county that remained a current cluster throughout cluster three's analyses.

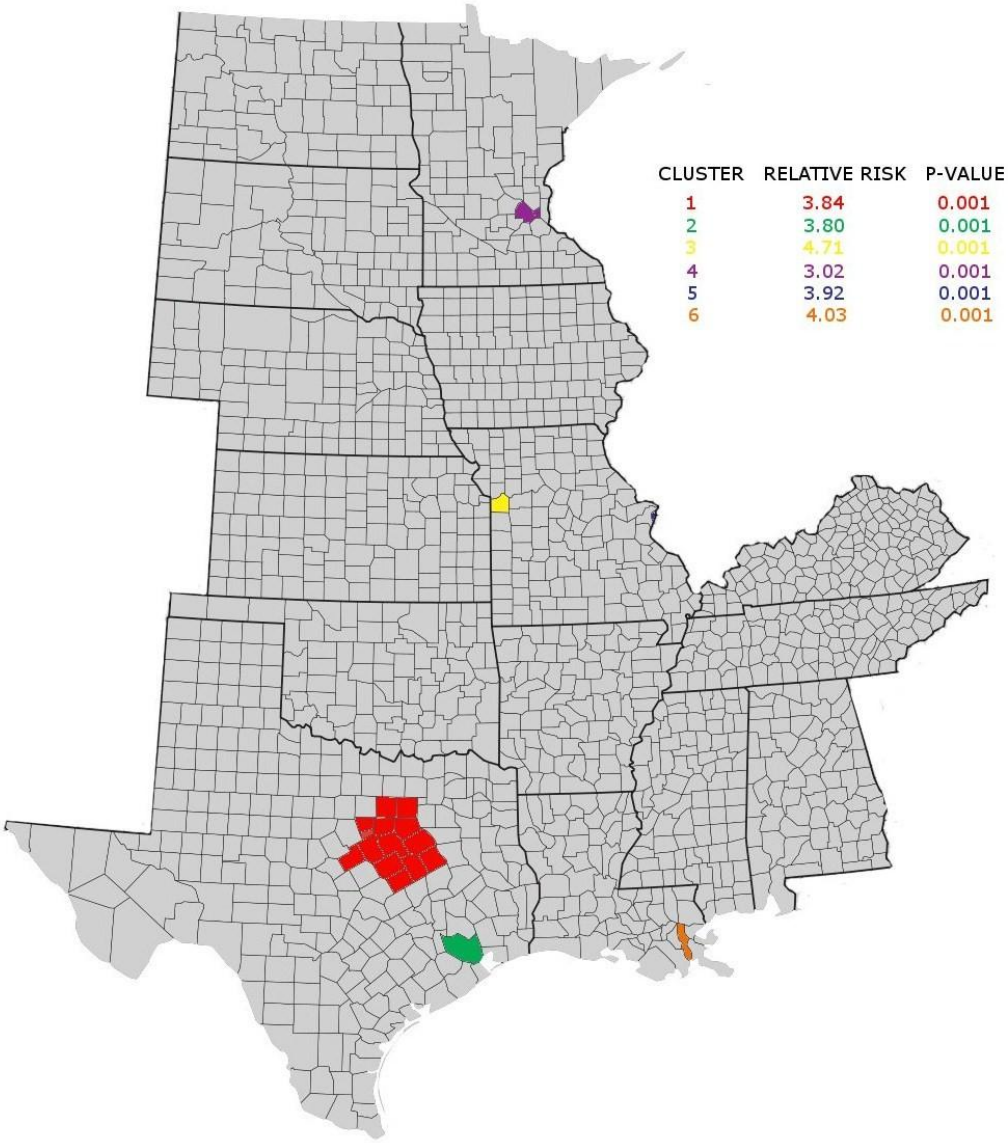


Figure 20: Spatial Analysis-Median Age, Median Household Income, African American Rate

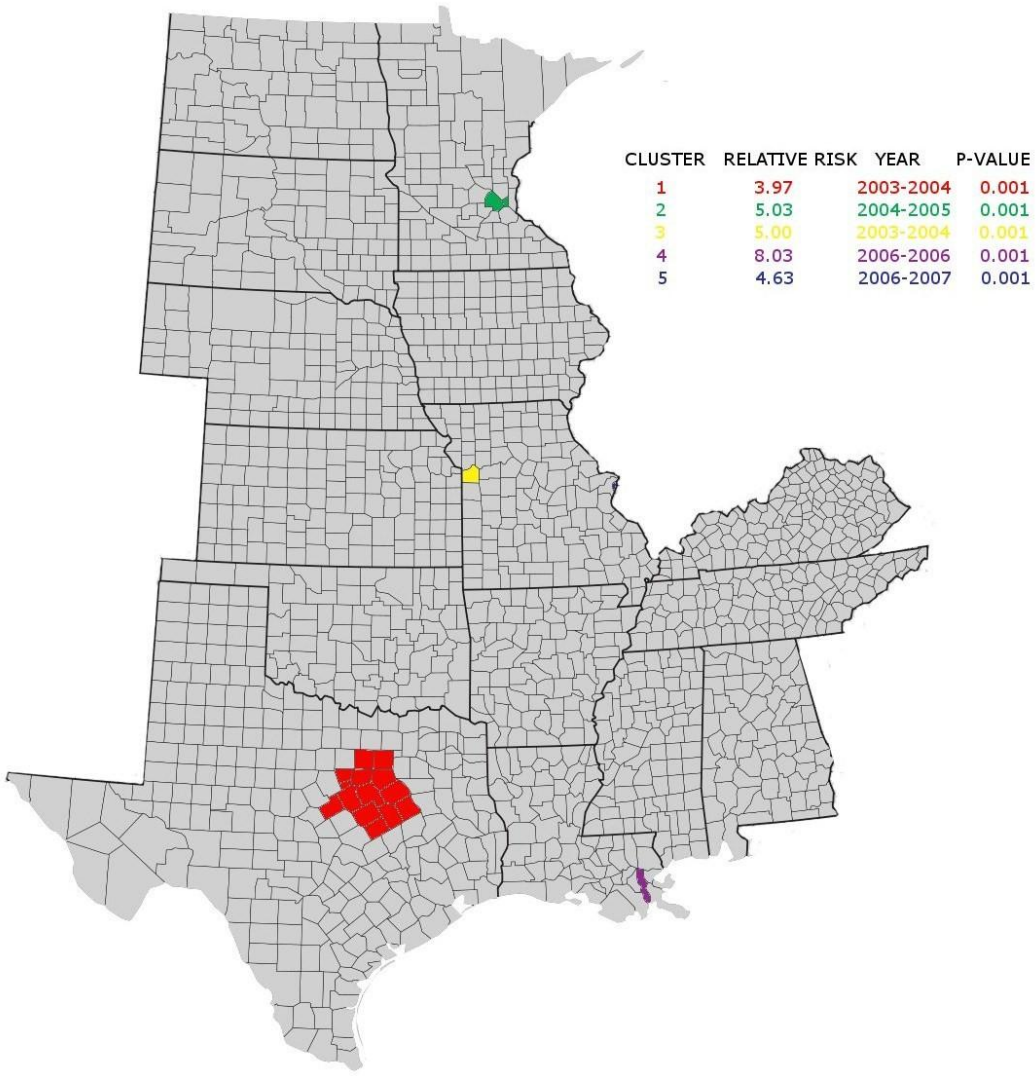


Figure 21: Space-Time Analysis-Median Age, Median Household Income, African American Rate

Summary

The identification of high risk areas in the United States for rape was successful in determining that when covariates were not taken into account, Philadelphia, Pennsylvania was the most likely cluster, and occurred from 2003-2004. When poverty was introduced, Philadelphia stayed in the most likely cluster, but twenty-three other counties were included as well. The relative risk went down, and it remained in the 2003-2004 time frame. When the African American covariate was added, Philadelphia bumped

down to the second most likely cluster, and was replaced by Queens and New York counties in New York. Philadelphia's cluster continued to occur from 2003-2004. Lastly the male covariate was introduced, and the two clusters did not change in either the space or space-time analysis. Although Queens and New York counties replaced Philadelphia as the most likely cluster when covariates were added, Philadelphia remained a prominent county in the second most likely cluster, and occurred from 2003-2004 each time.

In the Pacific division, King, Kitsap, and Snohomish counties in Washington appeared as the most likely cluster in each analysis. An additional three counties in Washington appeared when the African American rate covariate was added to the data, but then dropped back out when the male covariate was included. Each analysis occurred from 2003-2004. California showed up on the map for the first time here, with Los Angeles existing in the second most likely cluster each time.

Lastly, the "Katrina cluster" determined fourteen counties in Texas as the most likely cluster, with a high relative risk throughout. Orleans and Jefferson parishes appeared in cluster five, and both the most likely cluster and cluster five occurred in 2005-2006, during Hurricane Katrina. When median household income was introduced, both clusters remained the same for the spatial analysis. However, when a space-time analysis was run, the most likely cluster in Texas occurred in 2003-2004, and Orleans and Jefferson parishes disappeared from the study entirely. Lastly the African American covariate was introduced. Texas still continued to appear in the most likely cluster and remained in the 2003-2004 time period. Orleans parish dropped out of the spatial analysis and Jefferson parish dropped to cluster six; however Jefferson parish was the

only parish in that cluster. The relative risk elevated to 4.03. Jefferson parish increased to cluster four in the space-time analysis and occurred in *2006*, with a relative risk of *8.03*. Although Orleans and Jefferson parishes did not remain in the most likely cluster, they remained a staple in the analysis, with Jefferson parish showing a significant presence when all three covariates were analyzed. Saint Louis City was the only county that remained a current cluster throughout this analysis.

Suggestions for Further Study

One subject that remains to be explored is arrest data for Illinois and Florida. If there is data that exists that has been recorded in the same way the FBI compiled the UCR, adding these two states would be crucial and could predict an entirely different outcome. If this data does not exist, the NACJD did provide sporadic data for both Florida and Illinois throughout the years. A separate analysis on each state individually with the particular years of data that have been provided would be interesting to analyze and see where high risk clusters appear.

Philadelphia occurred in 2003-2004 for each analysis of the Middle Atlantic division. Further study in this area could be done to determine what caused clusters to appear during this time. A breakdown of Philadelphia into zip codes or demographic areas could be useful in establishing a specific area as the driving force for the cluster.

Saint Louis City remained the only current cluster throughout the entire analysis. Once the data for 2008-2009 is published, it would be interesting to see if it still remains a current cluster in 2009.

Lastly, Jefferson parish remained significant in cluster three's analysis. Although it was not in the most likely cluster, it appeared in five of the six analyses, and always

occurred in 2005-2006, the year of hurricane Katrina and its aftermath. Performing a cluster analysis solely on Louisiana, first by counties and then by zip codes, might help clarify if such clusters exist.

References

[1] Census Bureau. (2010). Census Regions and Divisions of the United States. Retrieved October 2010 from the Census Bureau website: http://www.census.gov/geo/www/us_regdiv.pdf

[2] U.S. Department of Justice, Federal Bureau of Investigation. *Uniform Crime Reporting Data [United States]: County-Level Detailed Arrest and Offense Data*, 1995 [Computer file]. 2nd ICPSR ed. Ann Arbor, MI: Inter-university Consortium For Political and Social Research [producer and distributor], 2001.

[3] MtJoy, R. (2010). The FBI's Shockingly Narrow Definition of Rape. Retrieved October 2010 from the Women's Rights website: http://womensrights.change.org/blog/view/the_fbis_shockingly_narrow_definition_of_rape

[4] Kulldorff, M. *SaTScan™ User Guide*. Cambridge, MA: 2009.

[5] Bynum, T., Maxwell, C. (2009). Survey Documentation and Analysis. Retrieved June 2010 from National Archive of Criminal Justice Data website: http://search.icpsr.umich.edu/NACJD/query.html?nh=50&rf=3&col=series&tx0=Uniform+Crime+Reporting+Program+Data+[United+States]%3A+County-Level+Detailed+Arrest+and+Offense+Data&fl0=title%3A&op0=%2B&ty0=w&col=abstract&tx1=NACJD&op1=%2B&fl1=archive%3A&ty1=w&t

[6] Allison, J.A., Wrightsman, L.S. *Rape: The Misunderstood Crime*. Newbury Park, CA: Sage Publications, Inc., 1993.

[7] Kutner, M.H., Nachtsheim, C.J., Neter, J., Li, W. *Applied Linear Statistical Models*. New York, NY: McGraw-Hill/Irwin Press, 2005.

[8] Larget, B. (2007). Poisson Regression. Retrieved November 2010 from the Department of Botany and Statistics, University of Wisconsin-Madison website: http://www.stat.wisc.edu/courses/st572-larget/Spring2007/handouts24-2.pdf

[9] NIST/SEMATECH e-Handbook of Statistical Methods. (2010). Engineering Statistics Handbook. Retrieved November 2010 from the U.S. Commerce Department: NIST Agency Website: http://www.itl.nist.gov/div898/handbook/eda/section3/eda366j.htm

[10] Wall, M.M. (2010). Count Outcomes – Poisson Regression Generalized Linear Models Part 3. Retrieved November 2010 from the Division of Biostatistics, University of Minnesota website: http://www.biostat.umn.edu/~melanie/PH7402/2010/countoutcome.pdf

[11] Bureau of Justice Statistics. (2000). Rape and Sexual Assault: Reporting to Police

and Medical Attention. Retrieved November 2010 from the Rape, Abuse, and Incest National Network website: http://www.rainn.org/get-information/statistics/reporting-rates

[12] Thornton, W.E., Voigt, L. Disaster Rape: Vulnerability of Women to Sexual Assaults During Hurricane Katrina. *Journal of Public Management & Social Policy* 2007; 2: 23-49.

[13] FBI UCR Program. (2010). Number of Arrests per Year. Retrieved November 2010 from the NumberOf.net website: http://www.numberof.net/number-of-arrests-per-year/

[14] Carter, D.L., Prentky, R.A., Burgess, A.W. *Rape and Sexual Assault*. New York and London: Garland Publishing Inc, 1988.

[15] Browse State Civilian Labor Force Data from the BLS. Economic Time Series Page. Retrieved August 2010 from the Economagic website: www.economagic.com/blssta.htm

[16] National Center for Education Statistics. CCD - Build a Table. Retrieved August 2010 from the U.S. Department of Education Institute of Education Sciences website: http://nces.ed.gov/ccd/bat/

[17] Census Bureau Geography. Census 2000 Centers of Population by County. Retrieved June 2010 from the U.S. Census Bureau website: http://www.census.gov/geo/www/cenpop/county/ctyctrpg.html

[18] SOI Tax Stats. Free County Income Data Downloads. Retrieved August 2010 from the Internal Revenue Service website: http://www.irs.gov/taxstats/article/0,,id=217542,00.html

[19] National Center for Education Statistics. State and County Estimates of Low Literacy. Retrieved August 2010 from the U.S. Department of Education Institute of Education Sciences website: http://nces.ed.gov/naal/estimates/StateEstimates.aspx

[20] State and County Quick Facts. Retrieved August 2010 from the U.S. Census Bureau website: http://quickfacts.census.gov/qfd/index.html

[21] CDC Wonder. Bridged-Race Population Estimates Request. Retrieved August 2010 from the CDC website: http://wonder.cdc.gov/Bridged-Race-v2008.html