Understanding the Relationship between Tree Canopy and Crime in Minneapolis, Minnesota using Geographically Weighted Regression

Andrew W. Eckerson
Department of Resource Analysis, Saint Mary's University of Minnesota, Winona, MN 55987

Keywords: Tree Canopy, Crime, GIS, LiDAR, Ordinary Least Squares Regression, Geographically Weighted Regression

Abstract

Influenced by reoccurring findings in the literature suggesting a negative relationship between tree canopy and crime rates, Esri ArcGIS spatial statistic tools were used to execute a series of regression analyses examining that relationship in the city of Minneapolis, Minnesota. Crime rates, by neighborhood, were modeled for with ordinary least squares and geographically weighted regression tools using tree canopy coverage and a number of other demographic data as independent variables. A spatially adjusted geography weighted regression (GWR) model indicated a statistically significant negative correlation existed in 64 of the 85 Minneapolis neighborhood observed in the study (r^2 = .82). These findings support the conclusions of previous literature that the relationship between tree canopy and crime is inverse.

Introduction

The value of abundant tree canopy in municipalities is well established. Studies have concluded that tree canopy reduces the costs of cooling a home, slows the formation of urban smog (Akbari, Pomerantz, and Taha, 2001), and promotes a general sense of wellness and tranquility (Hartig, Mang, and Evans, 1991). It also sequesters atmospheric carbon (Nowak, 1992) and can reduce feelings of danger (Herzog and Chernick, 2000).

After interviewing the residents of apartment complexes in inner-city Chicago about social behavior and interactions in their neighborhoods, researchers Frances Kuo and William Sullivan reported persons living in spaces with more tree canopy experienced a stronger social fabric with less fear, conflict, and incivility between residents (Kuo and Sullivan, 2001).

Consistent with the popular contention that positive social environments mitigate many of the psychological precursors to crime, Kuo and Sullivan (2001) also discovered lower crime rates in and around the apartment complexes where residents had more nearby trees and, by correlation, tended to know and get-along with their neighbors better.

More recent additions to the literature, with the use of remote sensing and GIS to observe larger sample populations with more quantitative methods on the relationship between tree canopy and crime, have drawn conclusions similar to those of Kuo and Sullivan (2001).

The belief that tree canopy correlates negatively with crime has been affirmed to varying degrees in studies covering Austin, Texas (Snelgrove, Michael, Waliczek, and Zajicek, 2004), Baltimore County, Maryland (Troy, Grove, and O'Neil-Dunne, 2012),

Eckerson, Andrew. 2013. Understanding the Relationship between Tree Canopy and Crime in Minneapolis, Minnesota using Geographically Weighted Regression. Volume 15, Papers in Resource Analysis. 9 pp. Saint Mary's University of Minnesota Central Services Press. Winona, MN. Retrieved (date) http://gis.smumn.edu.

Philadelphia, Pennsylvania (Wolfe and Mennis, 2012) and Portland, Oregon (Donovan and Prestemon, 2012). The goal of this study was to further explore the relationship within the geography of the city of Minneapolis, Minnesota.

Methods

Study Area

The area and units of observation used in this study were 85 of the 87 neighborhoods that make up the city of Minneapolis. The northwestern neighborhoods of Camden Industrial Park and Humbolt Industrial Park were excluded from the study area due to their lack of socio-demographic data as they contain no resident population (Figure 1).

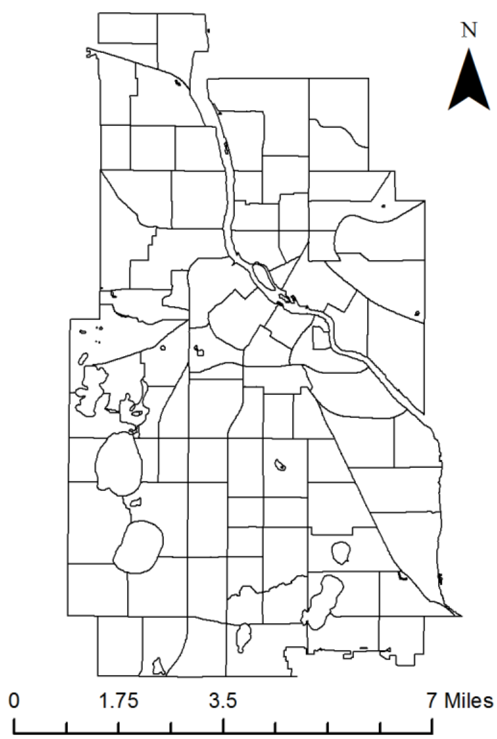

Figure 1. Minneapolis neighborhood boundaries with lakes, rivers, and Camden and Humbolt industrial parks removed.

The 85 observed neighborhoods occupy an area of 53.46 square miles and are home to a population of 382,578 people with a density of 7,156 persons per square mile.

Tree Canopy Delineation

Past studies have used varying methodologies for the delineation of tree canopy including field observations (Donovan and Prestemon, 2012) and remote sensing approaches such as using Light Detection and Ranging (LiDAR) (Troy *et al*., 2012). Normalized Difference Vegetation Indexes (NDVI) derived from multispectral satellite imagery have also been used to include public green spaces along with urban trees (Snelgrove *et al*., 2004; Wolfe and Mennis, 2012). A LiDAR delineation was selected for this study in order to isolate tree canopy.

LiDAR data of the city of Minneapolis, collected under off-leaf conditions in the fall of 2011, was acquired from the Minnesota Geospatial Information Office and used to produce three intermediate rasters: a digital elevation model (DEM), a digital surface model (DSM), and a LiDAR classification raster based on the mean high return of the individual LiDAR point's American Society of Photogrammetry and Remote Sensing (ASPRS) classification. All of these intermediate rasters were created at a 1 meter cell size.

To derive tree canopy the DEM was subtracted from the DSM to transform the values of the DSM from elevation to height. Using the classification raster all non-vegetation cells were then removed from the transformed DSM, and finally all cells less than 3 meters in height were removed. The minimum height of 3 meters was selected to enhance classification accuracy by defining canopy height as slightly taller than a ranch style home, not counting the roof.

The gaps in the evident tree crowns in the resulting tree canopy raster, which are due to the nature of the off-leaf conditions in which the LiDAR data was collected, were "filled" with the use of the majority filter and boundary clean geoprocessing tools in Esri ArcGIS.

Percentages of canopy cover were calculated for the land area of the neighborhoods with a mean coverage of 18.29% at a standard deviation of 6.87% (Figure 2).

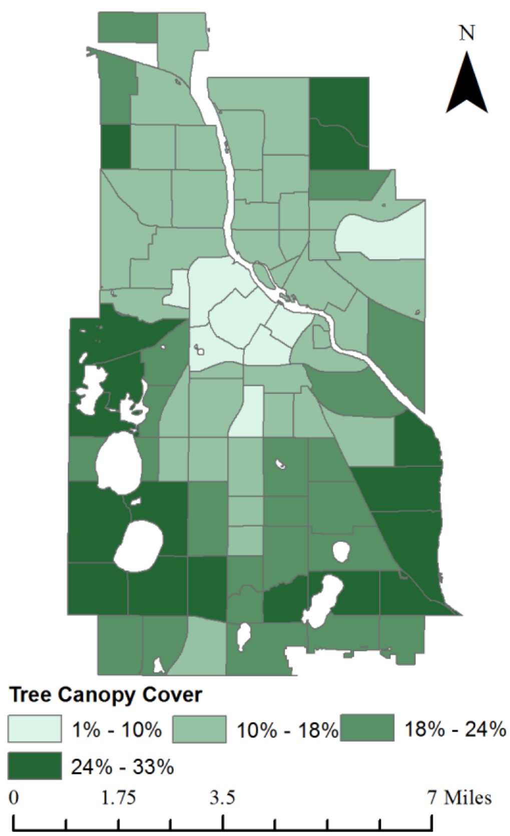

Figure 2. Tree canopy coverage in Minneapolis neighborhoods by percentage.

Accuracy Assessment

Using the methods described by Congalton and Green (2009) for thematic map accuracy assessment and sample size selection, the final canopy coverage percentage raster was assessed for classification accuracy against a 1 meter cell sized 2010 National Agriculture Imagery Program (NAIP) orthoimage of the study area.

Following the recommendations of Congalton and Green (2009) for study areas less than one million acres in size, 100 samples were randomly generated as point objects to begin the evaluation.

Random point samples were converted to 1-by-1 meter raster cells expanded into 3-by-3 homogenous raster cell clusters for evaluation. Three meter raster cells were used as the final accuracy assessment samples to account for the potential positional error that may occur when using single point sampling (Congalton and Green, 2009).

The 3-by-3 sample cluster size was deemed as a sufficient area as the intermediate LiDAR data was noted to have met or exceeded horizontal accuracy standards of a < 0.6 m root-mean-square error (RMSE).

Overall classification accuracy was determined to be 94% with both the user and producer error amounting to 86% for tree canopy and 96% for non-tree canopy (Table 1).

Table 1. Classification accuracy assessment error matrix, Tree Canopy (TC) and Non-Tree Canopy (NTC).

		Reference		
		TC	**NTC**	**Total**
Classified	**TC**	19	3	22
	NTC	3	75	78
	Total	22	78	100

Overall Accuracy = (19 + 75)/100 = **94%**

Producer's Accuracy
TC = 19/22 = **86%**
NTC = 75/78 = **96%**

User's Accuracy
TC = 19/22 = **86%**
NTC = 75/78 = **96%**

Whereas tree canopy was defined as vegetation exceeding 3 meters in height, it is important to note that while off-leaf

LiDAR derived canopy height models have been shown to be as accurate as their on-leaf counterparts, the heights of deciduous compound leaf trees tend to be slightly underestimated when rendered from off-leaf LiDAR data (Wasser, Day, Chasmer, and Taylor, 2013).

Crime Data

Data on crime occurrences in Minneapolis neighborhoods were obtained from Part I of the Minneapolis Police Department's monthly Uniform Crime Report (UCR) statistics from the years of 2010, 2011, and 2012.

Part I of UCR statistics count the number of committed indicator crimes, a group of crimes designated by the U.S. Department of Justice to best reflect the overall criminality of an area. Part I UCR crimes are auto thefts, robberies, homicides, burglaries, aggravated assaults, larcenies, rapes, and arsons.

While the effectiveness of using Part I UCR statistics to describe criminality is debated in the literature, sufficient evidence exists to defend their utility as a criminality metric (Gove, Hughes, and Geerken, 1985). Based on such evidence this study gives an equal weight of 1 per occurrence for each reported UCR crime.

In instances where multiple crimes are committed in the same incident report, it is the practice of the Minneapolis Police Department to record only the most serious charge for UCR reporting. For example, if a homicide were committed during a robbery only the homicide would be counted for UCR reporting. The total number of UCR Part I crimes from 2010 to 2012 was calculated per 1,000 residents for each neighborhood and produced a mean rate of 184.15 crimes per 1,000 residents and a standard deviation of 172.41 (Figure 3).

Statistical Analysis

Following the work of Troy *et al.*'s (2012) study of Baltimore County, the Esri ArcGIS Exploratory Regression, Ordinary Least Squares Regression (OLS) and Geographically Weighted Regression (GWR) tools were used to discern the relationship between UCR crime statistics and tree canopy coverage.

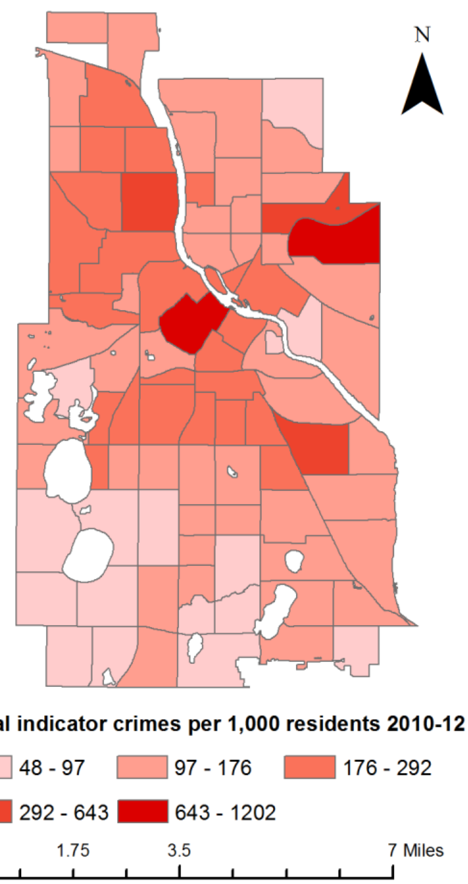

Figure 3. Total of URC indicator crimes per 1,000 residents, from 2010 to 2012.

Modeling for crime as the dependent variable, Exploratory Regression produced a reliable model from the independent social demographic and land cover variables (Table 2).

Independent variables were derived

from the Minnesota Land Cover Classification System and a report compiled by Minnesota Compass including information from the 2010 U.S. Census, the 2005-2009 American Communities Survey, and OnTheMap's 2009 Local Employment Dynamics data.

Table 2. Variables selected for use in OLS and GWR regression models.

Variable	Description	Mean	STDV
PER_CANOPY	% canopy cover	18.29	6.87
RTCRIME	Annual rate of crimes per 1,000 residents	184.15	172.41
PER_VACANT	% vacant homes	8.39	4.39
POP_DEN	Persons per square mile	7868.94	3936.59
PER_PO2000	% homes built after 2000	7.10	13.61
P25TO34	% population age 25 to 34	21.33	7.48
PO25NOEDU	% population with no education, secondary or higher	12.45	10.08
COMTORES	Ratio of commercial to residential land cover	1.45	5.44

Variable selection for exploratory regression was guided by known crime correlates from previous literature and texts (Troy *et al*., 2012; Lee, Beaver, and Wright, 2009).

Results

Ordinary Least Square Regression

The model produced from the ArcGIS Exploratory Regression tool was processed in the OLS regression tool to verify its validity (Figure 4).

OLS regression found a statically significant negative relationship ($p < 0.01$) between crime and tree canopy. The Adjusted R-squared value, a measure of model performance, was 0.72, meaning the model explains 72% of the occurrence of the dependent variable of crime.

$$Y = \beta_{0+} \beta_1 X_{1+} \beta_2 X_{2 + \ldots + } \beta_n X_{n +} \varepsilon$$

Figure 4. The OLS regression equation where Y equals the dependent variable, β are the regression coefficients, X are the independent variables and ε is the random error term.

All model coefficients were statistically significant and returned Variance Inflation Factor values of < 7.5, indicating that independent variable distributions were not redundant. The model's Jarque-Bera statistic of 2.00, > .05, suggested the residuals follow a normal distribution and the results of the model are unbiased (Table 3).

Table 3. Coefficients and significance and Variance Inflation Factors (VIF) for OLS regression model with crime as the dependent variable.

Variable	Coefficient	VIF
Intercept	179.05**	-----
PER_CANOPY	-7.68***	2.23
PER_VACANT	-14.47***	1.57
POP_DEN	-0.01***	1.25
PER_PO2000	-5.46***	1.62
P25TO34	4.64**	1.41
PO25NOEDU	1.69**	1.03
COMTORES	18.25***	1.40
Adj R-Squared	**0.72**	
Jarque-Bera	**2.00**	

*** Significant at 99%
** Significant at 95%
* Significant at 90%

Results of the OLS model were also evaluated for spatial autocorrelation using the ArcGIS Spatial Autocorrelation tool. Spatial Autocorrelation reported a z-score of 1.11 and reported the model's resulting pattern to be significantly random, concluding that key independent variables were present in the model. Nonrandom patterns in Spatial Autocorrelation are markers of spatially redundant information in the independent variables (Griffith, 1987).

Geographically Weighed Regression

Having established that an acceptable model was reached with the use of OLS regression, the acceptable set of variables was processed with the ArcGIS GWR tool.

The purpose of the GWR is to account for spatial variation in relationships over distances and improve model performance in circumstances where regression variables are non-stationary and geographic in nature (Brunsdon, Fotheringham, and Charlton, 1998). The GWR model predicted a negative relationship between crime and tree canopy with a model performance R-squared value of 0.82, a 10% improvement over the OLS model.

In the application of the GWR over an urban-rural gradient by Troy *et al.* (2012) a fixed kernel distance was established to account for significant urban-rural dissimilarities between observed areas. Due to the relative similarity of Minneapolis neighborhoods, in the context of population density and urbanization, kernel distances in this study's GWR model were calculated automatically by an adaptive kernel type with extent distances based on the Akaike Information Criterion (AICc) statistic.

Results of the GWR model's conditional statistics per coefficient were all < 30, indicating no significant presence of multicollinearity between the independent variables; additionally, coefficient standard errors were small in relation to the coefficient values for independent variables, a similar marker of model stability.

The overall AICc statistic for the GWR model was 999.89, less than the 1020.16 value reported in the OLS model. A lower AICc statistic indicates the GWR model is better than the OLS model at explaining the occurrence of the dependent variable.

All spatially adjusted coefficient values for tree canopy for each neighborhood were negative in the GWR model. The impact, or significance, of these coefficients on the dependent variable were tested by calculating t-statistics by dividing these coefficients by their standard errors. Using a threshold for 95% significance employed by Troy *et al.* (2012) t-statistics less than -1.65 merited a significant negative coefficient relationship to the independent variable. T-statistics more than 1.65 merited a significant positive relationship, and t-statistics falling between -1.65 and 1.65 indicated non-significant relationships.

The results of the t-statistics calculation found that 61 Minneapolis neighborhoods had a significant negative relationship between crime rates and tree canopy. Twenty-four neighborhoods were not statistically significant, thus indicative of a neutral relationship (Figure 5).

Conclusion/Discussion

While the results of the GWR model indicate an affirmation of the literatures findings, this conclusion is ultimately an estimation dependent on the acceptance of the adjusted R-Squared value of 0.82 as high enough to merit a significant model. The fact that the model meets the statistical tests in OLS means the model can be trusted, but the 0.82 value means it can only explain 82% of the variations in reported crimes.

At the very least the fact that both the OLS and GWR models meet statistical reliable testing implies that they can safely be accepted as being indicative of some level of an overall negative relationship between tree canopy and crime in Minneapolis.

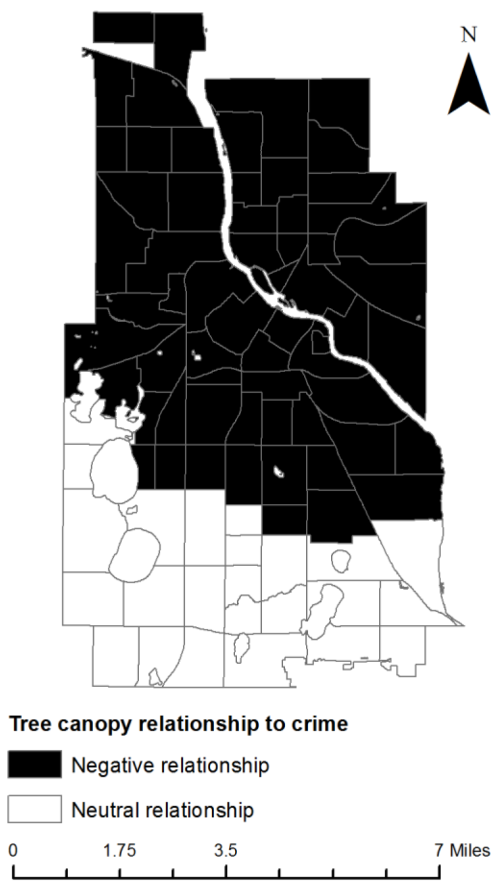

Figure 5. Distribution of the relationship between tree canopy and crime, via the t statistic.

Potential Sources of Error

Not described in the bulk of this paper were difficulties in finding a reliable OLS model via exploratory regression. Excessive correlation in many of the candidates for independent variables was common across some of the stronger crime correlates in the socio-demographic data.

Over correlation in the distribution of variables related to vacant homes, land cover, and employment opportunities required the computation of new variables to reach an acceptable OLS model.

Conflicts were ultimately resolved when two variables were removed, recomputed, and reintroduced into exploratory regression. The candidate variable number of jobs in a neighborhood was replaced with number of jobs per person ages 18 and older in a neighborhood, and the candidate variable percentage of commercial land cover in a neighborhood was replaced with a ratio of commercial land cover to residential land cover in a neighborhood.

While the study was able to reach a reliable OLS model, it is the author's suspicion that if this study were carried out analyzing a larger group of U.S. Census Block Groups rather than neighborhoods, the larger sample size would inherently produce more dissimilar socio-demographic independent variables. Such changes in the division of the study population would intuitively result in increased OLS and GWR model performance. Neighborhoods were selected as units of study due to the ease of availability of UCR data reported at the neighborhood level from the Minneapolis Police Department.

It is also worth noting that canopy delineations could be executed more accurately with the use of more advanced image analysis software, such as Trimble's eCognition. The use of such software was beyond the resources of this study.

Suggestion for Future Research

A natural suggestion for further research on this topic regarding the city of Minneapolis would be to investigate the role that height and ownership of all urban vegetation plays in criminality.

In the work of Donovan and Prestemon (2012) it was observed that tall trees of public ownership exerted a negative influence on crime, while smaller vegetation of private ownership exerted a positive influence. Troy *et al.* (2012) also observed that trees in public space have a stronger negative influence on crime than those in private property.

A future research question could explore the strength and sign of coefficients when the height and ownership of trees are divided into separate independent variables.

Acknowledgements

Many thanks and gratitude goes to Mr. John Ebert, Ms. Greta Bernatz, and Dr. David McConville for their instruction, support, and advice during my time as a graduate student in the Department of Resource Analysis at Saint Mary's University of Minnesota.

References

Akbari, H., Pomerantz, M., and Taha, H. 2001. Cool Surfaces and Shade Trees to Reduce Energy Use and Improve Air Quality in Urban Areas. *Solar Energy, Vol. 70 No. 3, 2001*. pp 295-310. Retrieved February 22, 2013 from Science Direct.

Brunsdon, C., Fotheringham, S., and Charlton, M. 1998. Geographically Weighted Regression. *Journal of the Royal Statistical Society: Series D (The Statistician), Vol. 47 No. 3.* pp 431-443. Retrieved November 18, 2013 from JSTOR.

Congalton, R., and Green, K. 2009. *Assessing the Accuracy of Remotely Sensed Data: Principles and Practices.* Boca Raton, FL: CRC Press.

Donovan, G., and Prestemon, J. 2012. The Effect of Trees on Crime in Portland, Oregon. *Environment and Behavior, Vol. 44 No. 1, 2012.* pp 3-30. Retrieved August 7, 2012 from Sage Publications.

Gove, W.R., Hughes, M., and Geerken, M. 1985. Are Uniform Crime Reports A Valid Indicator Of The Index Crimes? An Affirmative Answer With Minor Qualifications. *Criminology, Vol. 23 No. 3, 1985.* pp 451-502. Retrieved November 8, 2013 from EBSCO Host.

Griffith, D.A. 1987. Spatial Autocorrelation. *Washington, DC: Association of American Geographers.* pp 4. Retrieved November 18, 2013 from Google Scholar.

Hartig, T., Mang, M., and Evans, G.W. 1991. Restorative Effects of Natural Environment Experiences. *Environment and Behavior, Vol. 23 No. 1, January 1991.* pp 3-26. Retrieved August 7, 2013 from Sage Publications.

Herzog, T., and Chernick, K. 2000. Tranquility and Danger in Urban and Natural Settings. *Environmental Psychology Vol. 20, 2000.* pp 29-39. Retrieved August 7, 2013 from Science Direct.

Kuo, S., and Sullivan, W. 2001. Environment and Crime in the Inner City: Does Vegetation Reduce Crime? *Environment and Behavior, Vol. 33 No. 3, May 2001*. pp 343-367. Retrieved July 19, 2013 from Science Direct.

Lee, E., Beaver, K., and Wright, J. 2009, *Handbook of Crime Correlates.* pp 233-243. Retrieved November 6, 2013 from Google Books.

Nowak, D. 1992. Atmospheric Carbon Reduction by Urban Trees. *Environmental Management, Vol. 37, 1993.* pp 207-217. Retrieved February 22, 2013 from Science Direct.

Snelgrove, A.G., Michael, J.H., Waliczek, T.M., and Zajicek, J.M. 2004. Urban Greening and Criminal Behavior: A Geographic Information System Perspective. *HortTechnology, Vol. 14 No. 1, January 2004.* pp 48-51. Retrieved August 7, 2013 from Science

Direct.

Troy, A., Grove, M.J., and O'Neil-Dunne, J. 2012. The Relationship between Tree Canopy and Crime Rates Across an Urban-Rural Gradient in the Greater Baltimore Region. *Landscape and Urban Planning, Vol. 106, 2012.* pp 262-270. Retrieved February 22, 2013 from Science Direct.

Wasser, L., Day, R., Chasmer, L., and Taylor, A. 2013. Influence of Vegetation Structure on Lidar-derived Canopy Height and Fractional Cover in Forested Riparian Buffers During Leaf-Off and Leaf-On Conditions. *PloS one, Jan. 31, 2013*. Retrieved November 8, 2013 from Google Scholar.

Wolfe, M., and Mennis, J. 2012. Does Vegetation Encourage or Suppress Urban Crime? Evidence from Philadelphia, PA. *Landscape and Urban Planning, Vol. 108, September, 2012.* pp 112-122. Retrieved February 22, 2013 from Science Direct.