An Analysis Using Regression Models of Urban Solitary Bee Population with Regards to Perennial Gardens Along the Lake Street Corridor in Minneapolis Minnesota

Shana K. Pascal
Department of Resource Analysis, Saint Mary's University of Minnesota, Minneapolis, MN 55408

Keywords: Geographic Information Systems (GIS), Solitary Bees, Pollinator, Urban, Gardens, Global Positioning System (GPS), Ordinary Least Square (OLS), Geographically Weighted Regression (GWR), Hot Spot Analysis, Adjusted R-Squared, ArcMAP™

Abstract

In recent years the decline of the European Honey bee has highlighted the important role bee pollinators have on our ecosystem. In the United States, a focus to conserve bee pollinators has begun. This study focuses on solitary bee populations in urban environments as a conservation effort. The location of perennial gardens and solitary bees was collected along Lake Street in Minneapolis, Minnesota using a handheld global positioning system (GPS) unit. The data was then transferred into ArcGIS software where two different regression models were calculated. The regression models provided insight into which factors play a more important role in solitary bee populations. Results of the regression models show plant variety plays an important role in solitary bee populations in urban environments.

Introduction

Throughout North America there are an estimated 4,000 native bee species not including the honey bee (Berenbaum, 2007). Contrary to popular belief, the honey bee is not a native bee and was introduced to North America by European settlers. In recent years the honey bee has been receiving public attention because of declining numbers due to colony collapse disorder (CCD), a phenomenon scientists cannot fully explain; this disorder leaves managed hives void of nearly all honeybees with a few hives where honeybees inexplicably cannot find their way back to the hive (Gotlieb, Hollender, and Mandelik, 2010). With conservation efforts underway to help honeybees it is increasingly important to remember other bee pollinators. The native bees of North America play an important role in pollinating native flowers, plants, and crops, some of which the honeybee does not prefer or is physiologically unable to pollinate (Ebeling, Klein, Schumacher, Weisser, and Tscharntke, 2008). Roughly 14 million dollars in American food crops alone are pollinated by bees (Xerces Society, 2012), thus making conservation efforts of bees important. One way to help conserve and sustain native bee populations is by converting yards and public green spaces into bee habitats. Major contributing factors to the lack of pollination are loss of habitat and loss of habitat connectivity (Ahrne, Bengtsson, and Elmqvist, 2009). Using residential and business areas to support bee habitats in urban and suburban settings is important (Ascher, Fetridge, and Langellotto, 2008). A pilot project in Seattle, Washington

Pascal, Shana K. 2012. An Analysis Using Regression Models of Urban Solitary Bee Population with Regards to Perennial Gardens Along the Lake Street Corridor in Minneapolis, Minnesota. Volume 14, Papers in Resource Analysis. 15 pp. Saint Mary's University of Minnesota University Central Services Press. Winona, MN. Retrieved (date) http://www.gis.smumn.edu

called Pollinator Pathways utilizes boulevard space for planting perennial gardens for such sustainability. In this study, an area in Minneapolis, Minnesota parallel to East Lake Street was used to document the location of perennial gardens and estimate the quality of habitat for solitary bees. The study was conducted between June 2011 and August 2011. Understanding solitary bees was an important part of conducting this study.

Solitary Bees

According to the Xerces Society (2012), solitary bees are best described as bees that mate, nest, and raise their young independently from one another. Depending on the species and family, mated female solitary bees will lay eggs in the ground or hollowed-out tree branches and plant stems. The female bee then leaves food for the offspring and dies before the next generation hatches the following season. Solitary bees are smaller in size and therefore have a shorter foraging distance. On average, the foraging distance for a solitary bee is between 150-400 meters.

Solitary bees can be divided into two broad groups according to their foraging habits:

- Generalists are the more resilient species that gather nectar and pollen from a variety of flowers. Their resilience allows the bees to live in areas that are weedy or host non-native plants.
- Specialists rely more on the landscape and habitat they are surrounded by. Specialists prefer plants of the same species of groups of plants that are closely related for gathering nectar and pollen.

Methods

Study Area

Data was collected during the summer of 2011 using a Trimble Juno XP handheld global positioning system (GPS). The study was conducted in Minneapolis, Minnesota along a portion of the East Lake Street corridor stretching east from the Mississippi River to 39th Street (Figure 1).

Figure 1. The study area stretching east from the Mississippi River to 39th Street shown as a polygon along Lake Street in Minneapolis, Minnesota.

The study area was comprised of 14 residential streets and 17 alleyways. Observed gardens were located along boulevards because they provided adjacent sidewalk access to gardens without needing property owner permission to access the garden directly in the yard itself. During the study there were times when property owners would provide backyard access for a closer evaluation of garden areas. Data collection occurred between 9 a.m. to 4 p.m. to align data collection with the pollinators' active foraging time (Cunningham, 2011). At the end of the study, 233 perennial gardens were visited and their locations documented (Figure 2).

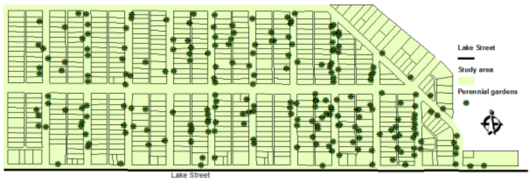

Figure 2. The study area stretches east from the Mississippi River to 39th showing the locations for 233 perennial gardens. Each garden location is represented with a green point symbol.

Data Collection

Prior to data collection, a shapefile and attribute fields were created using ESRI's ArcPad 10 and then transferred to the Trimble Juno GPS unit. The attribute fields were a series of pick lists that allowed for ranges to be selected to help with the speed of collecting data. Attribute fields used for the data collection are summarized in Table 1. All data were collected by walking the streets and alleyways within the study area. Each garden was given an estimated time of 1-5 minutes for observation. In certain instances, some gardens were visited for a longer time. Each garden was recorded as a point on the GPS unit.

Table 1. Attribute fields that were set up prior to field data collection: num_plants, size, b_bees, and s_bees were all set up as pick lists to make the field data collection process more efficient.

ID	num_plants	size	b_bees	s_bees
	1 to 3	10 or less	yes	yes
	4 to 8	10 to 20	no	no
	>9	>20		

Number of Perennial Plants (num_plants)

A count of the individual perennial plant varieties was tallied at each garden and grouped into a range of 1-3, 4-8, and greater than 9.

Garden Size (size)

The size of each garden was visually estimated at each garden and then recorded using the GPS unit. From the pick list garden size was estimate as less than 10 square feet, between 10-20 square feet and greater than 20 square feet. An accurate measurement for each garden was not possible without either disturbing the garden or obtaining permission from property owners.

Pollinator Presence (b_bees and s_bees)

A broad classification of pollinator presence was also documented using the GPS unit. As the attribute fields were set up initially, the simple question of whether or not pollinators were present at each garden was recorded with a yes or no answer from the pick list. A count field was also added allowing the number of each pollinator to be included. This number holds a large margin of error because the pollinators were not trapped for count accuracy.

Data Analysis

Upon completion of data collection, the GPS data was transferred onto a Windows 7 desktop machine using ESRI's ArcPad10. The data was further analyzed using ESRI's ArcMap 10 GIS software. To evaluate the question "does plant diversity and garden size have an effect on solitary bee populations in an urban environment?" a series of statistical analyses including Hot Spot Analysis, Ordinary Least Square Regression Analysis (OLS), Moran's I test and Geographically Weighted Regression

Analysis (GWR) were utilized. Each of the analyses was calculated for both plant variety and garden size.

Hot Spot Analysis

Hot Spot Analysis is a tool that takes into consideration each feature's neighboring feature telling the user if the values are spatially clustered based on high and low z-scores and p-values (ESRI, 2012). Users are prompted to fill in the input feature class, input field, and the output feature class. There are other options that can be addressed by using the drop down menus, but for this study they were left as the defaults (Figure 3). The output is a new feature class with z-scores and p-values included in the attribute table, the output feature class is automatically displayed in the table of contents showing the newly calculated z-scores labeled as Std. Dev in the map legend (ESRI, 2012). Hot spots within the data are z-scores that are statistically significant with larger z-scores clustered together and cold spots are statistically significant z-scores with lower values that are clustered together. Observed data where clustering was seen were further investigated using a regression model.

Regression Analysis

Regression analysis is a statistical analysis that allows the user to better understand their data spatially. There are three main reasons for using regression analysis in a study; they are:

- To better understand major factors
- Predict unknown values
- Hypothesis testing

This study used two types of regression analyses, Ordinary Least Square (OLS) Regression and the Geographically Weighted Regression (GWR) to test the hypothesis. OLS regression evaluates data on a larger scale whereas GWR evaluates data on a smaller scale. Overall, regression analysis works by creating an equation involving the study data as dependent or independent variables. An explanation of the equation can be seen below.

$$Y = \beta_o + \beta_1 X_1 + \beta_2 X_2 + \ldots\ldots\ \beta_n X_n + \varepsilon$$

Y: dependent variable or observed variable
X: independent variable or explanatory variable(s)
β: regression coefficients
ε: random error term

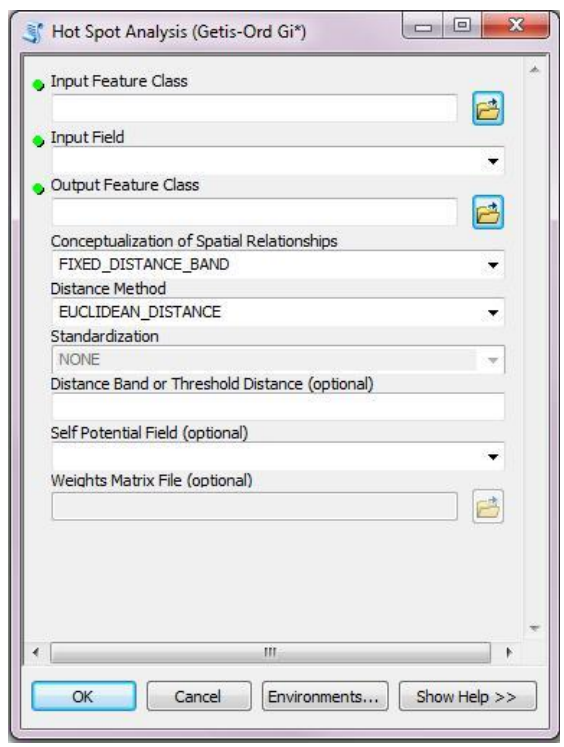

Figure 3. An example of the Hot Spot Analysis Tool form from the ArcGIS 10 Spatial Statistics Toolbox.

Ordinary Least Square Regression Analysis

The OLS regression analysis is the first step in building a regression model. OLS regression helps to find the most important variables for the model. For this study, those variables are the variety of plants (var_num), the size of the garden (size), and number of solitary bees (s_bees). Plant variety and garden size were the independent variables and number of solitary bees was the dependent variable in the equation. The user is prompted to fill in the input feature class, unique ID field, output feature class, dependent variable and then select the explanatory variables (Figure 4). The most important values for consideration are the adjusted R-squared and the Akaike's Information Criterion (AICc) value. These two values measure how well the model performed. Adjusted R-squared values can range from 0 to 100 percent and are represented as decimals in the output summary table; decimals read as a percentage explain the percentage of the Y value used in the equation. The AICc value is a measure of model's performance as well as expresses how well the model stands up to other models; the lower the AICc value, the better the model is. Besides an output table, the OLS regression analysis produces a map of over- and under-predictions displayed as red and blue circles in this study. The red circles indicate areas where the dependent variable is higher than what the model predicted and the blue circles indicate areas where the dependent variables are lower than the model predicted. If clustering is visible on the map there is a possibility that not all the best independent values are included in the equation. Running the Global Moran's I tool provides further information about the clustering of data and whether or not it is statistically significant.

Global Moran's I

Global Moran's I is a tool to help determine whether or not clustering seen in data is statistically significant (Figure 5). Note the input data for Global Moran's I is the residuals feature class from the OLS regression. There are five different ways to evaluate if an OLS model has been run correctly. Those five steps are checked in the summary of outputs and are listed below:

- Positive or negative sign
- Coefficients are statistically significant
- The Adjusted R-squared value is high
- Residuals are normally distributed

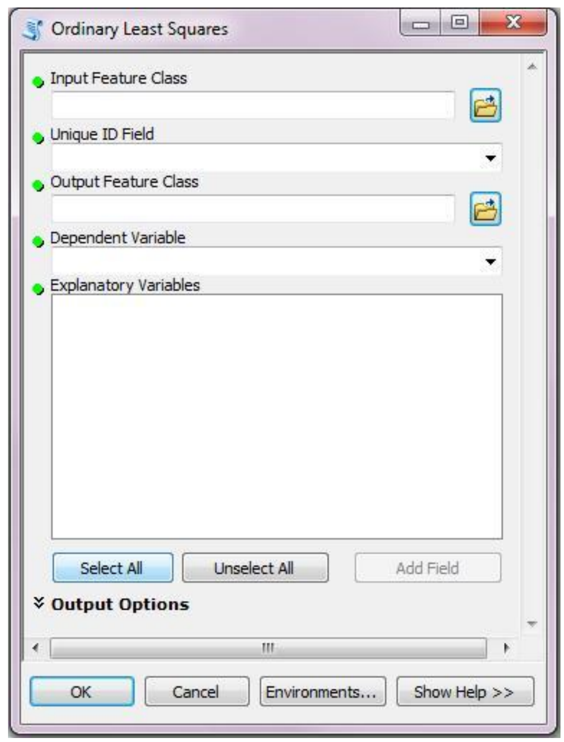

Figure 4. An example of the Ordinary Least Squares Tool form from the ArcGIS 10 Spatial Statistics Toolbox.

When all of these criteria have been met it is safe to assume the OLS regression model is sufficient to use. It is also important to conduct a Geographically Weighted Regression Analysis to make sure the data is being represented the best way.

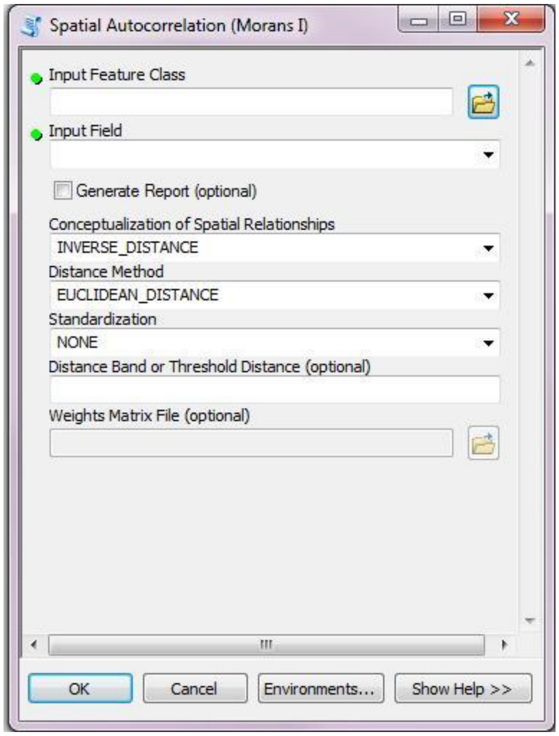

Figure 5. An example of the Global Moran's I Tool form from the ArcGIS 10 Spatial Statistics Toolbox.

Geographically Weighted Regression Analysis

Geographically Weighted Regression Analysis is a tool similar to the OLS tool but takes into consideration feature relationships on a smaller level rather than a larger level. GWR creates an equation for each feature and not just one equation that applies to all features. The benefit of having an equation for each feature allows for there to be change in the features' relationships over a geographic area. There are five other options the user has to choose from, but for this study the defaults were kept (Figure 6). Important numerical data to be considered when running GWR is the adjusted R-squared value and the AICc value.

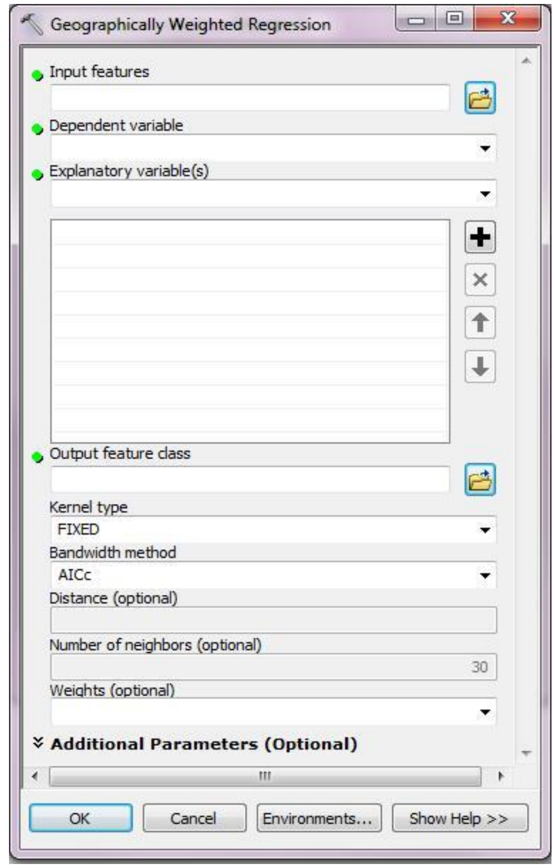

Figure 6. An example of the Geographically Weighted Regression Analysis Tool form from the ArcGIS 10 Spatial Statistics Toolbox.

Results

The results for this study are discussed in sequence of how the regression models were completed. First, the Hot Spot Analysis was conducted, followed by OLS, Global Moran's I, and lastly GWR. Each independent variable plant variety and garden size were then compared to themselves based on the two models to see which model was more suitable for this study.

Hot Spot Analysis for Solitary Bees – Plant Variety

Solitary bees produced a visualization based on the z scores that showed clustering of the data with high values and low values (Figure 7). The figure shows one strong hot spot (red and orange circles) and a few weak cold spots (blue circles). The clustering of high z-scores and scattering of low z-scores required further investigation to determine if the locations of the hot and cold spots were statistically significant or not. Output data summary produced an adjusted R-squared value of 0.65 or 65%. The high adjusted R-squared value suggests 65% of solitary bees present in the study area can be explained by the plant variety found at each garden. The AICc value of 307.57 was later used to compare the OLS regression model with the GWR model. The output map suggested possible clustering of data (Figure 7). To test if these data were statistically significant the Global Moran's I tool was used.

Least Ordinary Square Regression and Global Moran's I for Solitary Bees – Plant Variety

The OLS for plant variety used var_num as the independent variable and the solitary bee count (num_s_bee) as the dependent variable. The summary of all the OLS results for this model can be seen in Table 2 with the adjusted R-squared value and the AICc values represented in bold. The output map shows one hot spot area and a cold spot on the western edge of the study area (Figure 8).

Global Moran's I for Solitary Bees – Plant Variety

The results of the Global Moran's I indicate data were statistically significantly clustered, and a less than 1% chance the data was randomly clustered (Figure 9). Based on information provided by Global Moran's I values in the OLS summary, outputs were double checked using the five previously state validating criteria to see if the model was a strong as

Table 2. A summary of the OLS diagnostics from the regression model. The dependent variable, number of solitary bees is shown as num_s_bees.

OLS Diagnostics	
Independent Variable	Num_plants
Dependent Variable	num_s_bees
Number of Observation	85
Akaike's Information Criterion (AICc)	307.56
Multiple R-Squared	0.65
Adjusted R-Squared	0.65
Joint F-Statistic	158.99
Prob(>F), (1,83) degrees of freedom	.000000*
Joint Wald Statistic	191.4
Prob(>chi-squared), (1) degree of freedom	.000000*
Koenker(BP) Statistic	0.002517
Prob(>chi-squared), (1) degree of freedom	0.95
Jarque-Bera Statistic	0.81
Prob(>chi-squared), (2) degree of freedom	0.66

Table 3 shows the OLS model results to be. The values show the OLS for solitary bees and plant variety is not strong. The independent variable used did not show statistical significance in the summary making the model weak. The next step in the regression model was to run a

Figure 7. The figure shows the output from running the Hot Spot Analysis Tool using plant variety. From this figure it is possible to see one area that stands out as strong hot spot.

Figure 8 . The figure shows the output from running the Ordinary Least Square Tool using plant variety. From this figure it is possible to see more defined areas of cold spots on the most western edge of the study area and more hot spots than there were in the Hot Spot analysis.

Geographically Weighted Regression Analysis.

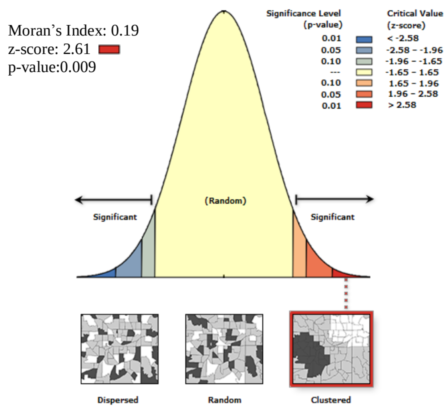

Figure 9. Output from ESRI's ArcMap Global Moran's I tool. Given the z-score of 2.61, there is a less than 1% likelihood that this clustered pattern could be the result of random chance.

Table 3. The table shows the verified OLS results. Results were verified by previously stated criteria.

OLS Analysis Performance Check	
Variable have expected positive or negative values	Yes
Variables are statistically significant	Yes
High adjusted R squared value	Yes, .65
Residuals are normally distributed	Yes, .66
Koenker value is statistically significant	Yes, .95

Geographically Weighted Regression for Solitary Bees – Plant Variety

To test and see which model performs better, a GWR analysis was conducted after determining the data were not best represented by the OLS regression analysis (Table 4). The focus of the numerical data from the GWR was the AICc value and the adjusted R-squared value. The adjusted R-squared value is .69 or 69%. This means the numbers of bees present at in the study area can be explained by the plant variety. The GWR map produced by the analysis showed very limited clustering except for in one location where both the Hot Spot Analysis and the OLS showed clustering (Figure 10).

Table 4. Numerical output data from the ArcGIS GWR analysis.

GWR Results for Solitary Bees and Plant Variety	
Bandwidth	310.69
Residual Squares	145.26
Effective Number	6.97
Sigma	1.36
AICc	**300.36**
Multiple R -Squared	0.71
Adjusted R- Squared	**0.69**

Hot Spot Analysis for Solitary Bees – Garden Size

The Hot Spot Analysis for solitary bees and garden size produced a figure that showed two hot spot and two cold spots (Figure 11). Strong clustering of both high z-scores and low z-scores required further investigation to determine if the locations of the hot and cold spots were statistically significant.

Ordinary Least Square Regression for Solitary Bees – Garden Size

The OLS for solitary bees and garden size used (size) as the independent variable and the solitary bee count (num_s_bee) as the dependant variable. The summary of all the OLS results for this model can be seen in Table 5. The numerical values that were given the most attention for this study were the AICc and the adjusted R-squared

Figure 10. The figure shows the output from running the Geographically Weighted Regression Model using plant variety. From this figure it is possible to see the data is less clustered than in both the hot spot analysis and the OLS model.

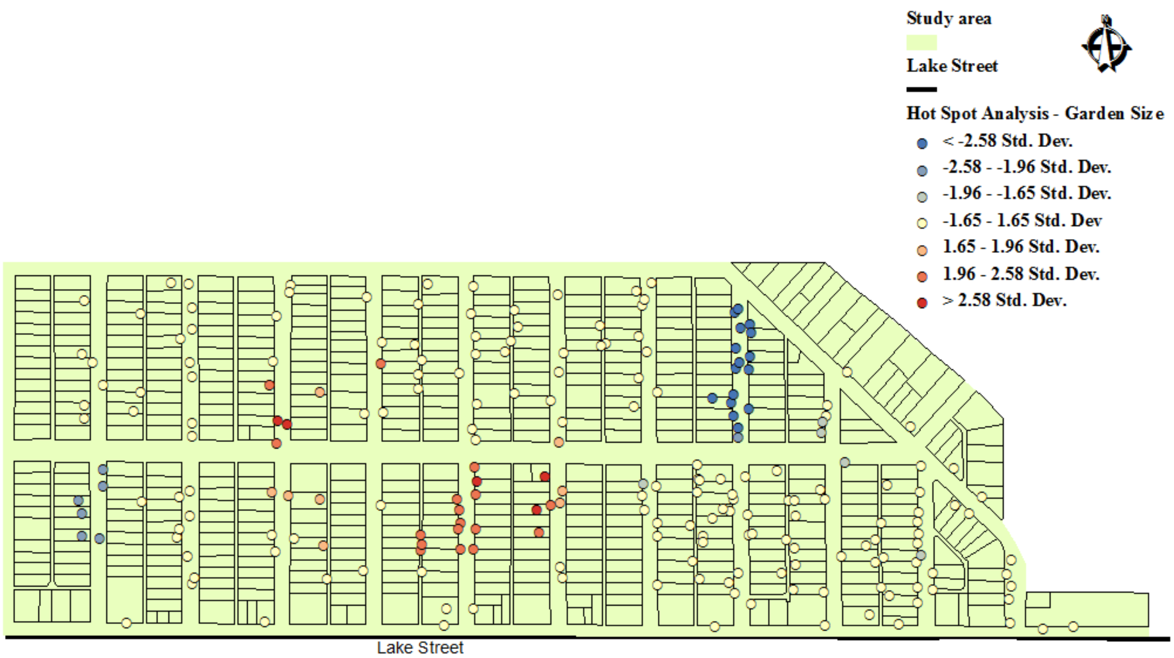

Figure 11 .The figure shows the output from running the Hot Spot Analysis Tool using plant variety. From this figure it is possible to see two defined areas of cold and hot spots. The clustering of the data in these areas indicates a need for further investigation.

value are seen in bold in Table 5. The output figure does not show any strong clustering among the high or low z-scores (Figure 12). To understand whether the OLS model was a good representation of the data, the Global Moran's I was conducted.

Global Moran's I for Solitary Bees – Garden Size

The output for the Global Moran's I showed areas where data were statistically significant based on z-scores, either positive or negative and also showed whether the data were statistically significant and clustered (Figure 13). Based on the Global Moran's I values, the OLS summary was double checked using the previous criteria (Table 6).

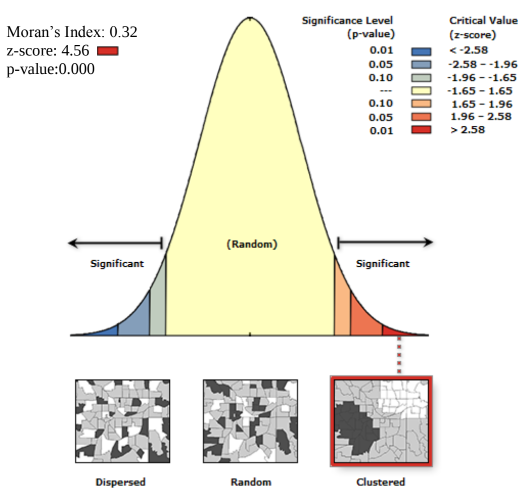

Figure 13. Output from ESRI's ArcMap Global Moran's I tool. Given the z-score of 4.56, there is a less than 1% likelihood that this clustered pattern could be the result of random chance.

The values show the OLS for solitary bees and garden size is not as strong. The low adjusted R-squared value indicated a weak model.

Geographically Weighted Regression for Solitary Bees – Garden Size

Table 5. Summary of OLS diagnostics from the regression model for solitary bees and garden size.

OLS Diagnostics – Garden Size	
Independent Variable	s_bee2
Dependent Variable	num_s_bees
Number of Observation	85
Akaike's Information Criterion (AICc)	392.80
Multiple R-Squared	0.65
Adjusted R-Squared	0.05
Joint F-Statistic	5.77
Prob(>F), (1,83) degrees of freedom	.02*
Joint Wald Statistic	6.87
Prob(>chi-squared), (1) degree of freedom	.01*
Koenker(BP) Statistic	0.12
Prob(>chi-squared), (1) degree of freedom	0.73
Jarque-Bera Statistic	1.66
Prob(>chi-squared), (2) degree of freedom	0.44

To test and see which model was better, the GWR was run after determining the data were not best represented by the OLS regression analysis. The summary of outputs for the GWR can be seen in Table 7. The numerical data most important from the GWR were the AICc value and the adjusted R-squared value. The adjusted R-squared value is .09 or 9%. The number of bees present at in the study area cannot be explained by the size of the garden. The GWR figure produced by the analysis does not show any clustering of data (Figure 14).

Table 6. The results of answering five question that can be answered to better understand if an analysis is good or not.

OLS Analysis Performance Check	
Variable have expected positive or negative values	Yes
Variables are statistically significant	Yes
High adjusted R squared value	No, .05
Residuals are normally distributed	Yes, .44
Koenker value is statistically significant	Yes, .73

Table 7. Numerical data output for the GWR analysis.

GWR Results for Solitary Bees and Garden Size	
Bandwidth	436.98
Residual Squares	435.45
Effective Number	4.85
Sigma	2.33
AICc	**390.34**
R2	0.13
Adjusted R2	**0.09**

Discussion

This study was undertaken to determine if there were relationships between solitary bee populations and perennial gardens within an urban setting. A comparison of the two regression models for each variable was compared and then based on the AICc and the adjusted R-squared value an overall inference was made.

Comparison of summary results using plant variety

To determine which model best represented the data for solitary bees and plant variety, AICc values and adjusted R-squared values needed to be compared. The AICc value that was lowest indicated which model was better for the data. In this study, the AICc value for the OLS model was 307.57 and the AICc value for the GWR was 300.36. The AICc value of 300.36 indicated data from this study can best be explained on a smaller scale than on a larger scale by using the GWR model. Another number that was tracked during the study was the adjusted R-squared value; in this instance the adjusted R-squared values also increased from 65% to 69%. Both models had high adjusted R-squared values making the variety of plants in a garden significant on the pollinator population.

Comparison of summary of results using garden size

To determine which model best represented the data for solitary bees and garden, AICc values were compared. By evaluating the OLS and GWR summary of results, a comparison showed the AICc value for OLS was 392.80 and the AICc value for GWR was 390.34 indicating the GWR model was a better model to be used to determine if pollinator population was affected by garden size. Along with the increased AICc value for the GWR model the adjusted R-squared value also increased. In the OLS, the adjusted R-squared value was .05, and for the GWR the adjusted R-squared value was .09. The low adjusted R-squared values indicated

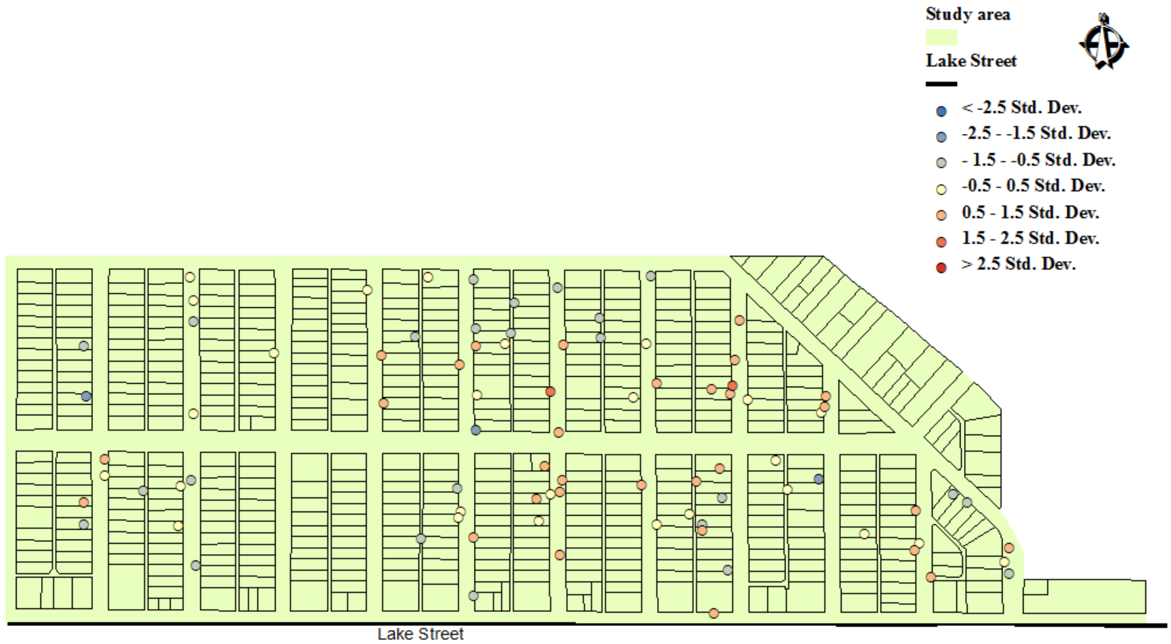

Figure 12. The figure shows the output from running the Ordinary Least Squares Model using garden size. From this figure it is possible to see the lack of clustering amongst the data.

Figure 14. The figure shows the output from running the Geographically Weighted Regression Model using garden size. From this figure it is possible to see the data does not change from the OLS Model to the GWR Model.

there was no statistically significant importance of size of gardens on the number of pollinators present.

Comparing regression models for both plant variety and garden size

By comparing outcomes of the two regression models with one another, plant variety had a greater effect on the solitary bee population in the study area. This was supported by plant variety having a lower AICc value (300.36) and a higher adjusted R-squared value (.69).

Data Limitations

This study was impacted by many limitations. The design of the study would benefit from the inclusion of more independent variables. Some independent variables that would have helped would have been time of day, temperature, sun vs. shaded gardens, wind condition, and distance to nesting locations.

Continued Work

As colonies of commercial honeybees and certain species of native bumblebees continue to decline, studying populations of urban bee pollinators is important to understanding how people can improve conservation efforts through plantings of flowering perennial plants. Having a future study that includes more independent variables would greatly improve the outcome of the regression models and would allow for suggestive solutions in urban areas. Also, having a longer study and including more data would likely be more meaningful to bee specialists who might be able to make stronger conclusions from larger data sets taken over a longer period of time.

Conclusion

A full regression model was conducted using Hot Spot Analysis, Least Ordered Squares Regression Analysis, Global Moran's I, and Geographically Weighted Regression Analysis to analyze data collected along a portion of the East Lake Street corridor in Minneapolis, Minnesota to evaluate the question "Does the variety of plants and the size of the garden have an effect on pollinator population?" After calculating the two regression models for the two variables, the outcome suggested the independent variable that had the most impact on solitary bee numbers at the gardens was the plant variety. This result makes it possible to infer higher numbers of perennial plants positively impact the number and presence of bees. The analysis also showed there is no real statistical significance in the size of the perennial flower garden. Conducting a study on a larger scale may provide different results as larger, more established gardens may provide native bees with a stable nectar and pollen resources and specific habitat requirements female bees need to optimize their reproduction capacity over time.

Acknowledgements

I would like to thank the Professors at Saint Mary's University of Minnesota in particular John Ebert who oversaw my education from start to finish. I would also like to thank Neil Cunningham for having a passion for all things small for if it weren't for him I would have done a project on water. Thank you for all your time that you helped me look at pollinators during the Minnesota state government shutdown. I would also like to thank my family and friends who supported me though this whole process, your words of encouragement, persistent reminders of a

fast approaching deadline and tireless hours of babysitting helped bring this project to a close.

References

Ahrne, K., Bengtsson, J., and Elmqvist, T. 2009. Bumble Bess (Bombus ssp) along a Gradient of Increasing Urbanization. *PLoS ONE* 4(5): e5574.

Ascher, J. S., Fetridge, E. D., and Langellotto, G. A. 2008. The Bee Fauna of Frsidential Gardens in a Suburb of New York City (Hymenoptera:Apoidea) *Entomological Society of America, 101(6)*, 1067-1077.

Berenbaum, M. 2007. The Birds and the Bees – How Pollinators Help Maintain Healthy Ecosystems.

Cunningham, N. 2011. Green Noise LLC. Personal Inteview. June 12, 2011.

Ebeling, A., Klein, A., Schumacher, J., Weisser, W. W., and Tscharntke, T. 2008 How does plant richness affect pollinator richness and temporal stability of flower visits? *Oikos, 117*, 1808-1815.

ESRI. 2012. How Hot Spot Analysis (Getis-Ord Gi*) works. Retrieved January 4, 2012 from http://help.arcgis.com/en/arcgisdesktop/10.0/help/index.html#//005p0000001100000.

Gotlieb, A., Hollender, Y., and Mandelik, Y. 2010. Gardening in the desert changes bee communities and pollination network characteristics. *Basic and Applied Ecolocy, 12*, 310-320.

Xerces Society. 2012. Retrieved July 10, 2011 from http://www.xerces.org/ pollinator-conservation/.