

Environmental and Anthropogenic Influences on the Diversity and Distribution of Native Pollinators in the Order Hymenoptera

Sophia Fan, Natasha Dhamrait, Madison Hoenig

Department of Ecology and Evolutionary Biology
The University of Toronto, EEB313

Abstract:

Native pollinators specifically are incredibly important for local ecosystems due to their essential relationship with agriculture and natural ecosystems. Our group “beeplyr” analyzed an open source pollinator library dataset from the U.S. Geological Survey (USGS) website and tested the relationship between climatic, geographic, and anthropogenic factors on hymenopteran pollinator diversity in the U.S. Within the context of climate change and its ubiquitous effects on all aspects of life, and when accounting for the myriad of ways human action can change the global environment, diversity studies such as this are more important now than ever before. These allow for informed decision making, effective legislation and conservation efforts, in addition to increasing the information available regarding sensitive and ecologically significant species.

When all environmental relationships with diversity were tested and the data was appropriately filtered and transformed, only relative humidity was a significant predictor of pollinator diversity. A model testing the relationship between anthropogenic influences and diversity showed only land use as significant. It was only when combining all environmental and anthropogenic effects that the model had the most predictive results. This led us to conclude that both environmental and anthropogenic predictor variables must be considered to accurately describe their relationship with diversity.

Introduction/Background:

Native pollinators are incredibly important for local ecosystems, a concept that is probably familiar to many. As students, we learn about how important bees are to agriculture and to the co-evolution and persistence of many native, wild plant species. Native hymenopteran pollinators constitute the majority of pollination diversity in North America, but surprisingly not so in agriculture, where the honeybee (*Apis mellifera*), a non-native species, is the pollinator of choice. Several emerging studies show that native pollinators including native bees and wasps are just as effective, if not better at pollinating both our crops and wild plants (see EEB380 - offered by the University of Toronto at the low, low cost of 75% of your waking hours!). Thus, increased research into what factors affect native pollinator abundance and diversity is warranted and will become increasingly important in the context of global change and human disturbance.

Our group “beeplyr” will be working with an open source pollinator library to test hypotheses regarding climatic, geographic, and anthropogenic influences on hymenopteran pollinator diversity in the U.S. Our dataset comes from the Pollinator Library at the U.S. Geological Survey (USGS) website, a collaborative dataset with contributions from various researchers in North Dakota, Texas, and Colorado. Within their respective research sites, they recorded the occurrences of various pollinator species and associated climatic and locational data, and submitted them for compilation at USGS. Data collected includes location (county, state), land use, GPS coordinates, taxon, elevation, air temperature, average wind speed, and relative humidity. We also downloaded human population data from relevant states from a NASA survey for further analysis in conjunction with pollinator data. To calculate diversity, we used Simpson’s diversity index of genera diversity, since certain taxa were not identified down to species level. The Simpson’s diversity index is commonly used to assess biodiversity in a multi-species ecosystem. It measures the observed abundance of a species in relation to total abundance. However, occurrence data alone only allows us to assess the diversity of observations, which may not be reflective of species composition or abundance. We filtered the dataset to include only Hymenoptera and the top sampled families and genera within (the final dataset included Apidae, Halictidae, and Andrenidae, with genera *Apis*, *Bombus*, *Lasioglossum*, *Melissodes*, and *Perdita*), and calculated the diversity index (some

pollinators were not identified to species) at each GPS coordinate. Data was distributed across three separate states, North Dakota, Texas, and Colorado (Figure 1)

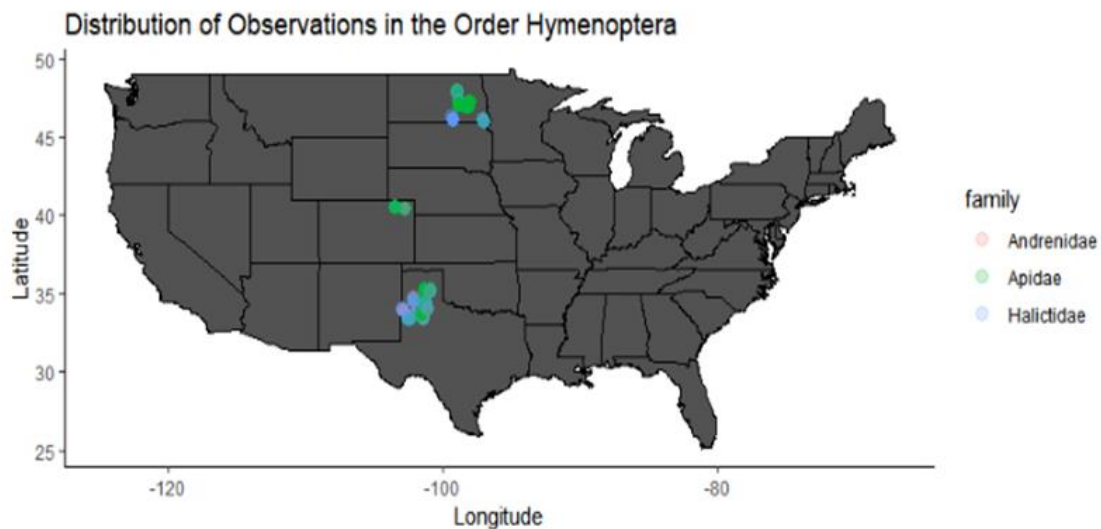


Figure 1: Distribution of Sampled Data Points Separated by Family

Our questions address whether elevation, latitude, air temperature, wind speed, and relative humidity affect diversity, whether different land use management affects diversity, and whether human population density correlates with diversity. We also want to summarize which of these factors are the most important, if at all, in predicting pollinator diversity using simple linear regression and linear mixed effects models. We will be commenting on the limitations of open source, collaborative datasets and the biases that can be associated with them if protocols are not standardized. Our goal is to better describe pollinator ecology outside of simply *Apis mellifera* and in turn, increase understanding of potentially sensitive and important pollinators that are often overlooked. We hope this information will be used towards bettering conservation efforts, inform policies for issues such as land management and environmental regulations, and to encourage further research on the preservation of threatened pollinator species.

Data Analysis

Environmental Effects:

We hypothesize that all climatic variables will have an effect on pollinator diversity. Specifically, we predict that:

1. The lower the latitude, the greater the diversity.
2. The lower the elevation, the greater the diversity.
3. Higher humidity generates more diversity
4. Higher temperatures generate more diversity
5. Lower wind speed generates more diversity

To address whether climatic variables - air temperature, relative humidity, and wind speed - affect pollinator diversity, we initially used simple linear regression models both with and without interactions between all predictor variables. Then we realized that each GPS coordinate is nested within a study site or county, meaning GPS coordinates that are closer to each other are more similar (tested with Moran's I). Thus, we used a linear mixed effects models to account for the nestedness of "county" and used AICc

model selection to pick the best model (See TABLE). Predictor variables were standardized since factors like elevation and humidity were on wildly different scales.

Assumptions to do with normality, homoscedasticity, and multicollinearity were checked using QQ plots and histograms, residual plots, and VIF tests, respectively (SEE FIGURE AND TABLE). The data appeared normal in the histograms but the Q-Q plot was somewhat skewed. However, it didn't seem bad enough to warrant transformation beyond initial scaling. The residual plot was heavily right-skewed, but attempts at transforming the data to different powers and roots did not yield noticeably different results. This is probably due to the fact that our data was not sampled at random, on a transect, and instead was a conglomeration of various scientists who happened to be studying in certain areas. The VIF test revealed that elevation and latitude were multicollinear, which makes sense since each location has an associated elevation, and is just a feature of the data collected. Thus, the final models only incorporated latitude, which was a better predictor than elevation. Intuitively and because physics, most of these climatic variables will be correlated in some way, but our data does not reflect this, and this is probably also caused by how the data was collected and compiled.

Land Use Effects:

We were also interested in how land use affect species diversity. We hypothesize that land usage is a significant predictor of pollinator diversity. It is predicted that protected land areas such as Wildlife Refuges and Conservation Reserves will serve as predictors for increased pollinator diversity. We created a boxplot to visualize the spread of land use observation and tests for normality were conducted using a histogram and the Shapiro-Wilks' test. The calculated data points for Simpson's Diversity Index were transformed to the cube root, which is a fairly strong transformation having effects on the distribution shape (Cox, 2005). This was done to address the skew of land use observations. We used Levene's test to measure the homogeneity of variance and ran one-way analysis of variance to determine if land use was a significant predictor of pollinator diversity. Finally, we ran two models using the transformed diversity index and the untransformed diversity index to determine which land use types were most significant. We used the AIC as a basis for model selection and selected the model with the lowest AIC score. Inherent biases in our data exist, due to non-random sampling and uneven distribution of observations per land use site. Additionally, the Simpson's diversity index indicates diversity trends but masks important population measurements such as composition and relative abundance, which may be more informative of ecosystem dynamics.

Human Population Effects:

This study is also interested in assessing how human population density affects pollinator diversity. It is hypothesized that hymenopteran communities sampled in geographic areas with higher human population density would have lower diversity than those sampled in geographic areas with lower human populations. It is predicted that areas with greater human populations would exert greater anthropogenic influences on the local environments, such as increased pollution, habitat destruction, or environmental degradation which would consequently result in lower pollinator diversity. The GPS coordinates in the dataset were first converted to spatial points and projected using a sinusoidal transformation [CRS("+proj=sinu=ellps:7271")]. A sinusoidal transformation was chosen since the GPS coordinates in this dataset spanned the continental United States of America, and therefore requires larger maps for proper mapping. The transformed spatial points were then plotted on a map of the United States, while separating by hymenopteran families, to observe the distribution of data.

In order to assess how human population density influences pollinator diversity, a human population.adf file was downloaded and converted to a raster file (United States Geological Survey, 2019). The human population density raster points were extracted from the dataset and bound to the total dataset for analysis. Additionally, the raster points were converted to a dataframe so that human population density and pollinator diversity could be plotted on ggplot. To plot on a ggplot, the raster points needed to be filtered so that the large population density of large cities, such as New York City, did not skew the scale of the pollinator data.

Generalized least squares models were run on the data to include spatial variation as a random effect in linear models. Assumptions of normal distribution and equal variance were tested. All different projection GLS models were compared using AIC tests.

Additionally, Moran's I tests were run on both human population and pollinator diversity tests to assess if points closer together are more or less similar to each other compared to points further away.

As established, there are many inherent biases in this dataset, and this also includes spatial biases. It is assumed that data was sampled across a uniform gradient of human population densities. However, this is not an accurate assumption and will skew our data towards human populations that were easier for researchers to sample, such as at research sites where population sizes are small.

Results:

Environmental Effects and Diversity:

In checking assumptions of linear regression, we started with a VIF test (SEE TABLE). As shown, elevation and latitude had values of more than 10, which indicates significant multicollinearity. As discussed in data analysis, this is due to the non-random nature of data sampling to create the dataset. All other factors have $VIF < 5$, which indicates little multicollinearity. However, again, these factors *should* be correlated, but this dataset does not show it.

Table 1:

<i>Elevation</i>	<i>Latitude</i>	<i>Wind speed</i>	<i>Relative humidity</i>	<i>Air temperature</i>
10.8049	11.2999	1.3257	1.3646	1.3606

Upon inspection, air temperature and relative humidity seem more or less normally distributed, with elevation and latitude predictably somewhat skewed and bimodal. The QQ plot is most likely skewed due to the non-normality of elevation and latitude in this dataset. As mentioned, any transformation of data, especially latitude and elevation, didn't yield any noticeably different results.

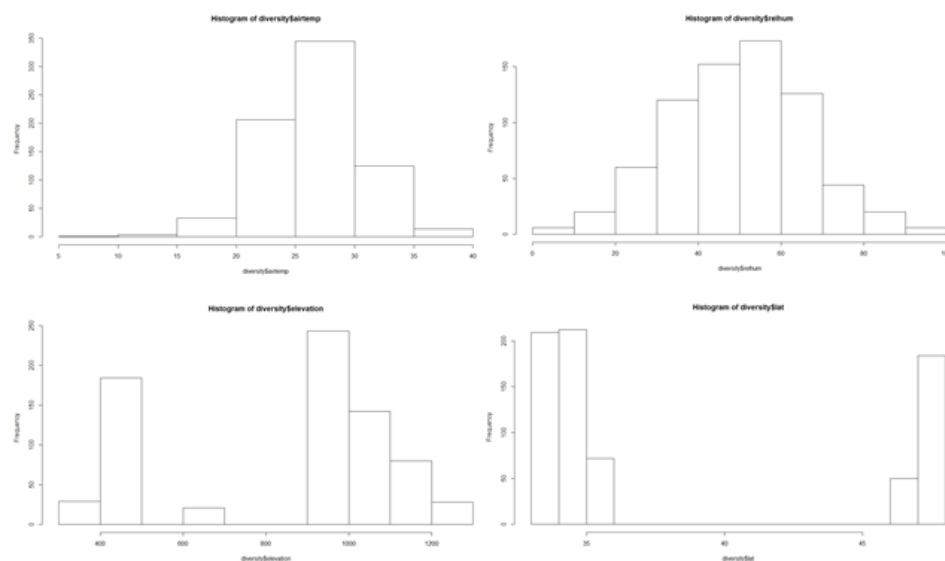


Figure 2: Distribution of Environmental Predictor Variables

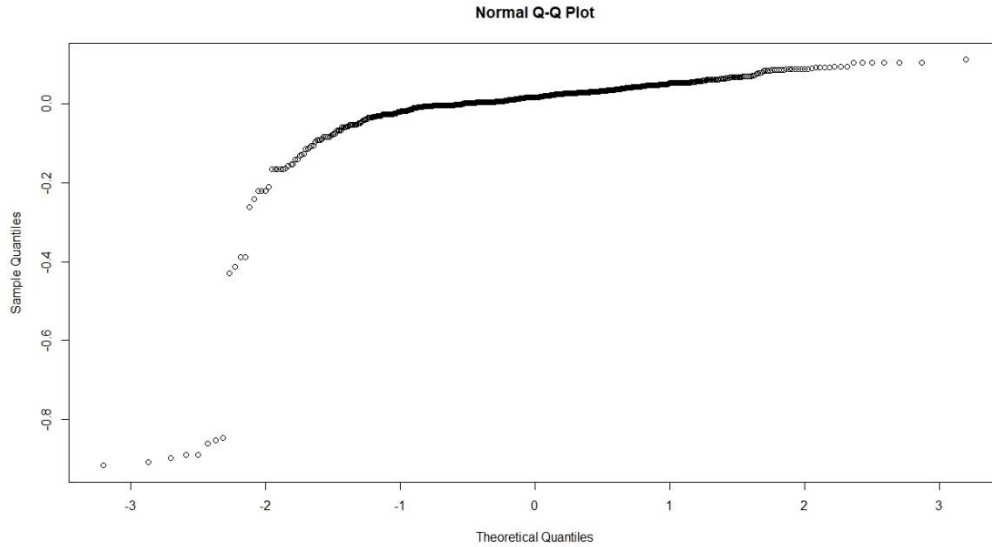


Figure 3: QQplot of Predictor Variables

The residuals vs. fitted plot showed a heavily clumped, right skewed pattern. Again, transformations of the data to various powers and roots beyond initial scaling didn't yield any conceivable effects, so we chose to ignore this violation in homoscedasticity.

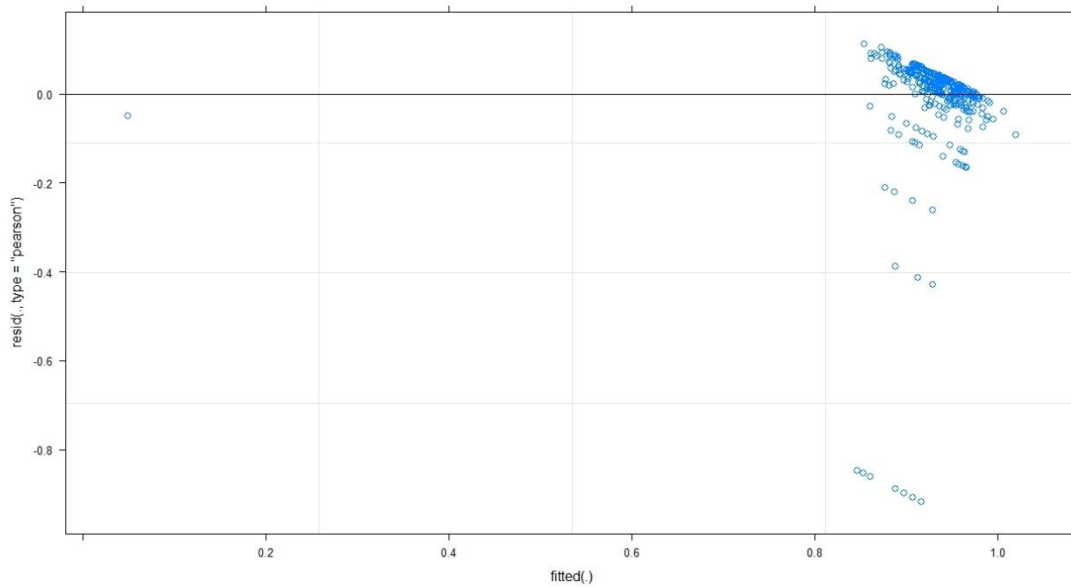


Figure 4: Model Outputs

In our linear regressions and linear mixed effects models, we had an $n = 727$ and chose significance at $\alpha = 0.05$ throughout. The model using non-interaction between relative humidity, air temperature, and wind speed yielded an adjusted $R^2 = 0.017$ with a $p\text{-value} = 0.0014$, and the only significant predictor within this model was relative humidity with $p\text{-value} = 9.02e-05$.

Table 2:

<i>Df</i>	<i>F value</i>	<i>Residual SE</i>	<i>Adjusted R²</i>	<i>p-value</i>	<i>Predictor</i>	<i>Estimate</i>	<i>Standard error</i>	<i>p-value</i>
723	5.221	0.1225	0.01714	0.001432 ***	Relative humidity	1.241e-03	3.153e-04	9.02e-05 ***
					Wind speed	4.874e-05	5.556e-04	0.930
					Air Temperature	1.989e-03	1.220e-03	0.103

In optimizing the AICc model, we first optimized random effects - the nestedness of individual GPS points within counties resulted in the need to account for it as a random effect. We tested fixing the intercepts against letting the intercepts vary with latitude. The AICc values showed that the saturated model with varying slopes was better.

Table 3:

<i>Model</i>	<i>df</i>	<i>AICc</i>
Fixed intercepts	10	-993.6226
Intercepts varying	12	-1024.8539

After inspecting the initial linear regression, we decided to use latitude, relative humidity, and air temperature in our linear mixed effects model selection. The model with full effects and interactions had the lowest AICc score.

Table 4:

<i>Model</i>	<i>df</i>	<i>AICc</i>
Latitude * Rel. hum. * Air temp.	12	-1090.921 ***
Latitude	6	-1065.624
Relative humidity	6	-1082.151
Air temperature	6	-1063.843
Latitude * Rel. Hum.	8	-1080.783
Latitude * Air temp.	8	-1070.548
Rel. hum. * Air temp.	8	-1079.313
Latitude + Rel. hum.	7	-1082.807
Latitude + Air temp.	7	-1064.728
Rel. hum. + Air temp.	7	-1081.218
No fixed effects	5	-1064.531

The full model showed that the only significant predictors were, once again, relative humidity, as well as the interaction between latitude and air temperature. The latter makes intuitive sense, since air temperature does vary with latitude, but following discussion above, we decided to discard the significance of any factor to do with latitude simply because the way the dataset was constructed doesn't allow for any robust conclusions to be made about latitude and its influence on diversity.

Random effects

Table 5: Full model

<i>Group</i>	<i>Name</i>	<i>Variance</i>	<i>Standard Deviation</i>	<i>Corr</i>
County	(Intercept)	0.01024	0.1012	
	Latitude	0.02274	0.1508	1.00
Residual		0.01209	0.1100	

Fixed effects

Table 6:

Predictor	df	Estimate	Standard error	p-value
Latitude	12.3	-0.1043	0.0355	0.0691
Relative humidity	726.1	0.0254	0.0524	7.82e-07 ***
Air temperature	725.5	0.0074	0.0050	0.1394
Lat : Rel hum	721.5	-0.0072	0.0054	0.1816
Lat : Air temp	721.8	-0.0201	0.0048	3.81e-05 ***
Rel hum : Air temp	717.0	0.0055	0.0044	0.2092
Lat : Rel hum : Air temp	722.2	0.0019	0.0053	0.7135

Despite its significance in both our simple linear regression and the linear mixed effects models, the estimates for relative humidity remain quite low (linear regression: 1.241×10^{-3} , linear mixed effects: 0.0254), meaning that relative humidity explains and influences relatively little of the variation in diversity. Nonetheless, one can see a clear increasing relationship between relative humidity and diversity, as shown below. We sorted the points by county to give a better idea of what is occurring locally, and extracted the counties that had steeper slopes.

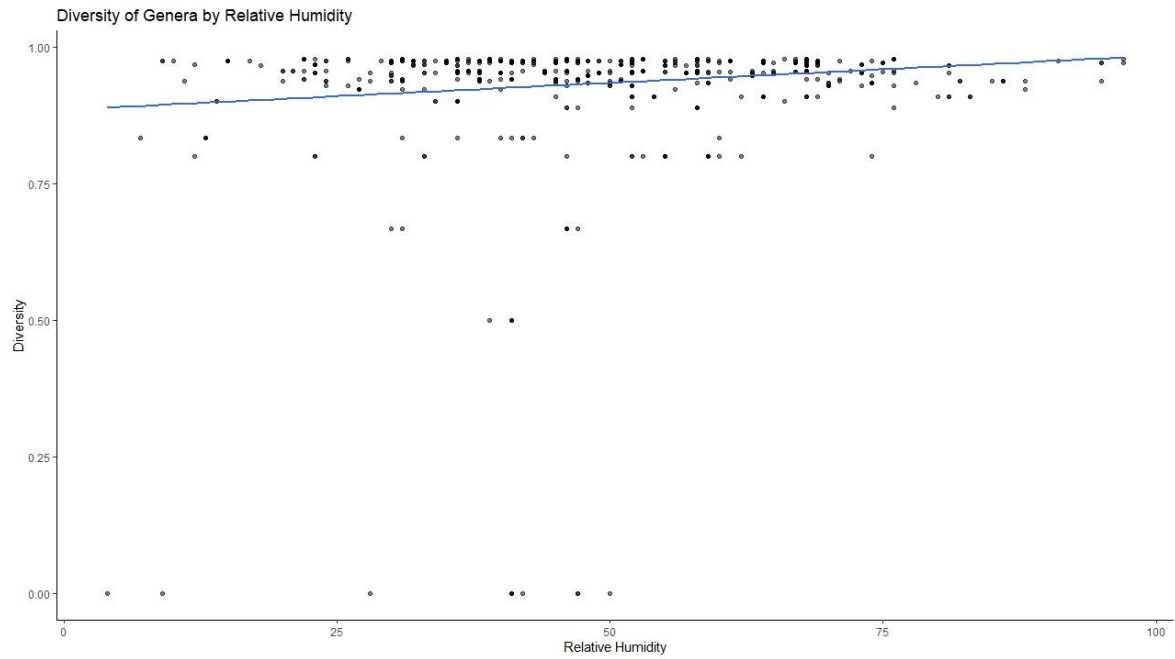


Figure 5:

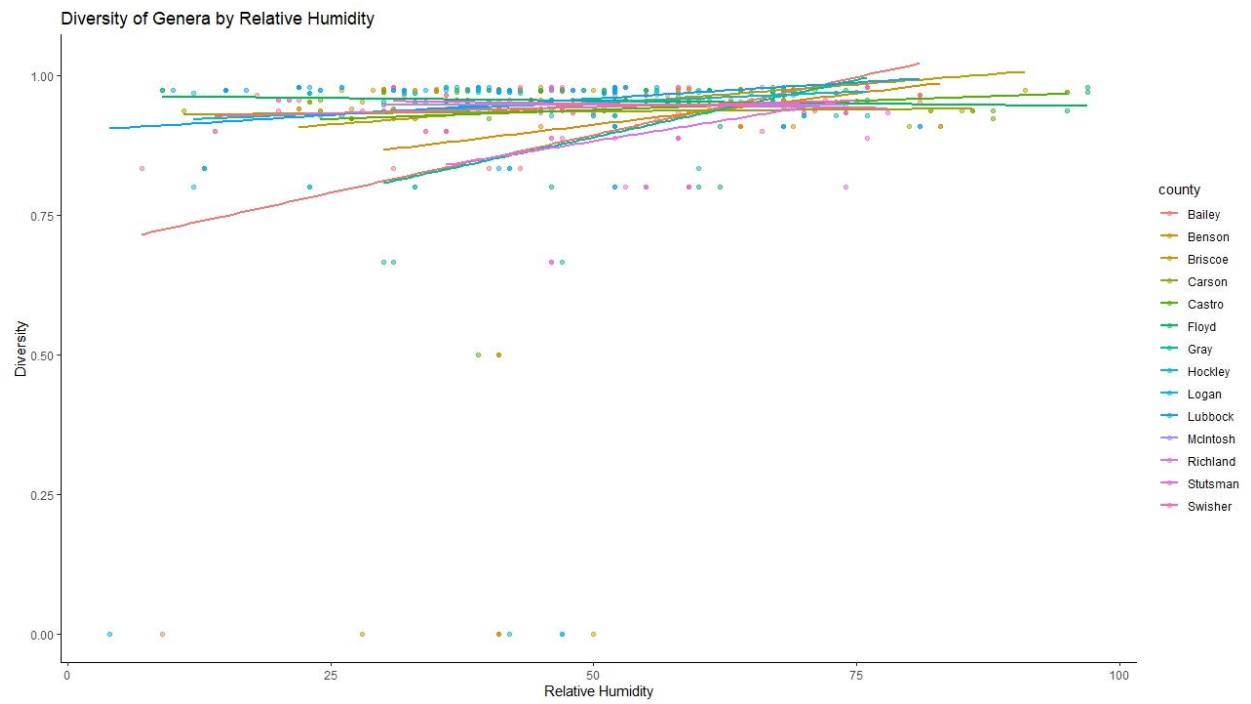


Figure 6:

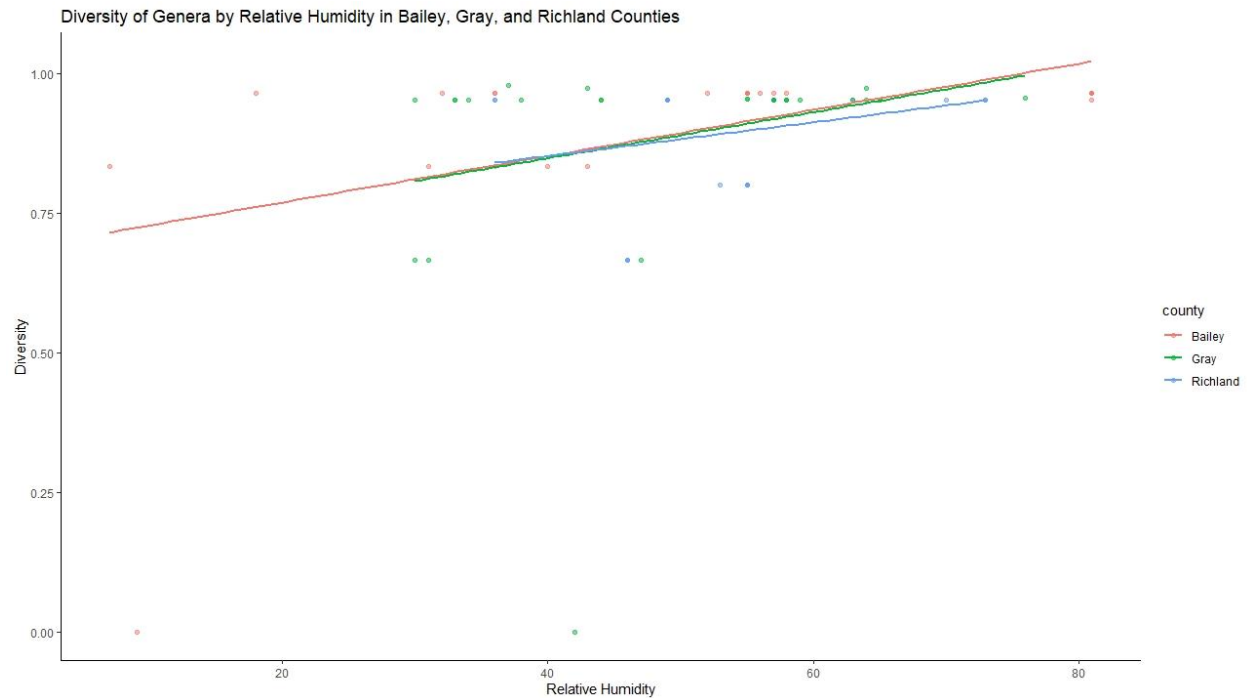


Figure 7:

Land use and Diversity

To get a visual overview of the spread of our data, we plotted families by land use type over diversity index.

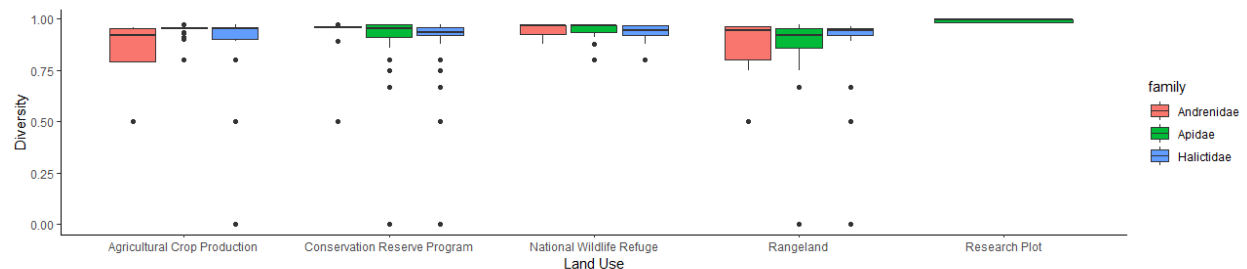


Figure 8:

We then wanted to determine if diversity indices were significantly different among land use types using a one way analysis of variance. First we tested our assumptions. We used a histogram and the Shapiro-Wilk's test to assess the normality of our data. These tests indicated that the data was not normally distributed and was somewhat right-skewed.

In order to correct for this violation of ANOVA assumptions, the diversity index was transformed to the cube root. To measure the homogeneity of variances, we used Levene's test as the data was not normally distributed. The results demonstrated a p value = 0.0027, which indicated that the variances were not equal. Despite these violations, we went ahead and ran the ANOVA.

Table 7: ANOVA Table, p-value = 4.0464e-5 indicated a significant difference between land use plots.

	Degrees of Freedom	Sum of Squares	Mean Squares	F Value	Pr(>F)
Land Use	4	0.2879	0.071967	6.4525	4.0464e-5
Residuals	887	9.8930	0.011153		

Table 8: Transformed Diversity Model, AIC: -1471.983

Coefficients	Estimate	Standard Error	T value	Pr(> t)
Intercept	0.960210	0.011011	87.208	2e-16
Conservation Reserve Program	-0.005558	0.012370	-0.449	0.65333
National Wildlife Refuge	0.019881	0.014215	1.399	0.16228
Rangeland	-0.012181	0.014532	-0.838	0.40212
Research Plot	0.036208	0.036208	2.6932	0.00723

We constructed a simple linear regression model using the transformed diversity index to determine which land use types were most significant contributors of pollinator diversity. The research plot was the most significant.

We also investigated the model without the transformed data for comparison.

Table 9: Diversity Model, AIC: -1277.676

Coefficients	Estimate	Standard Error	T value	Pr(> t)
Intercepts	0.909214	0.012277	74.055	<2e-16
Conservation Reserve Program	-0.001095	0.013793	-0.079	0.9368
National Wildlife Refuge	0.032815	0.015850	2.070	0.0387
Rangeland	-0.025322	0.016204	-1.563	0.1185
Research Plot	0.80091	0.014997	5.341	1.18e-7

Using a simple linear regression model to compare land use types with the calculated Simpson's Diversity Index. This model indicated that the intercept and the Research Plot were the most significant. Comparing the AIC score for both models, the model with transformed data yielded a lower value, thus indicating a better fit.

Human Population and Diversity:

Before we could run models including spatial variation, the assumptions of the model were tested. The human population density was not normally distributed, with most data points skewed towards lower population densities.

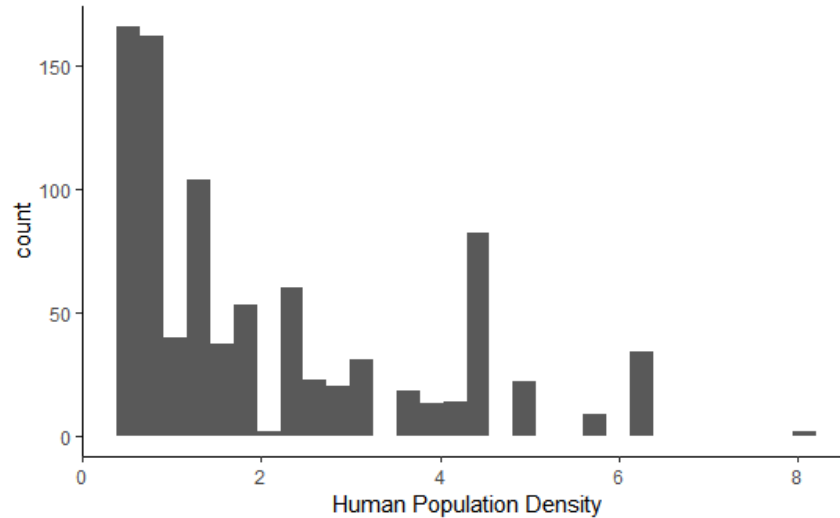
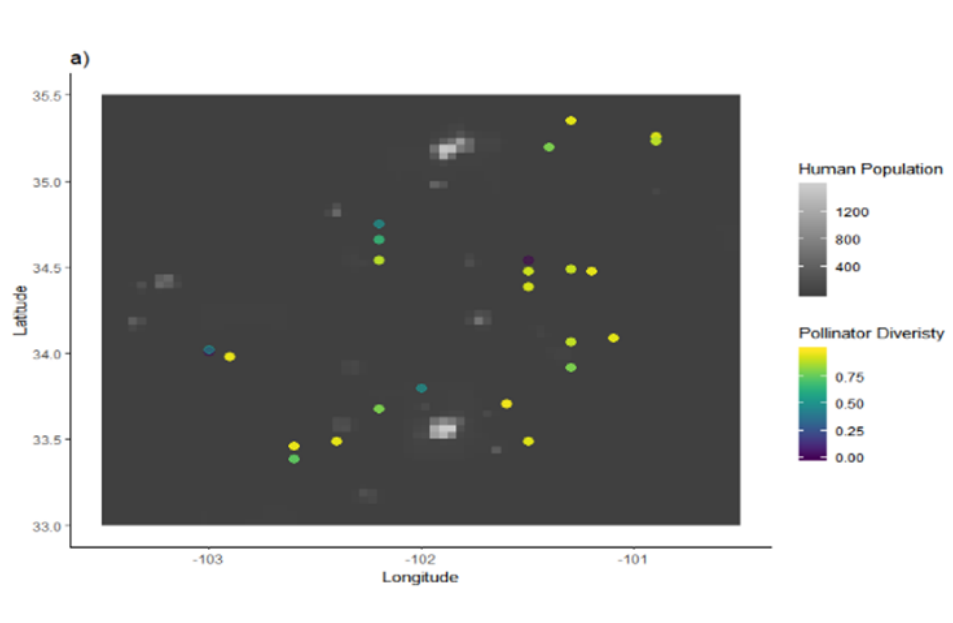


FIGURE 9: Distribution of Human Population Values.

Represented graphically, human population density does not appear to be normally distributed. This conclusion is supported by the results of a Shapiro-Wilk's test ($p\text{-value} < 2.2e-16$). When graphically plotted, there does not appear to be any relationship between human population size and pollinator diversity in any of the three states (Figure 10).



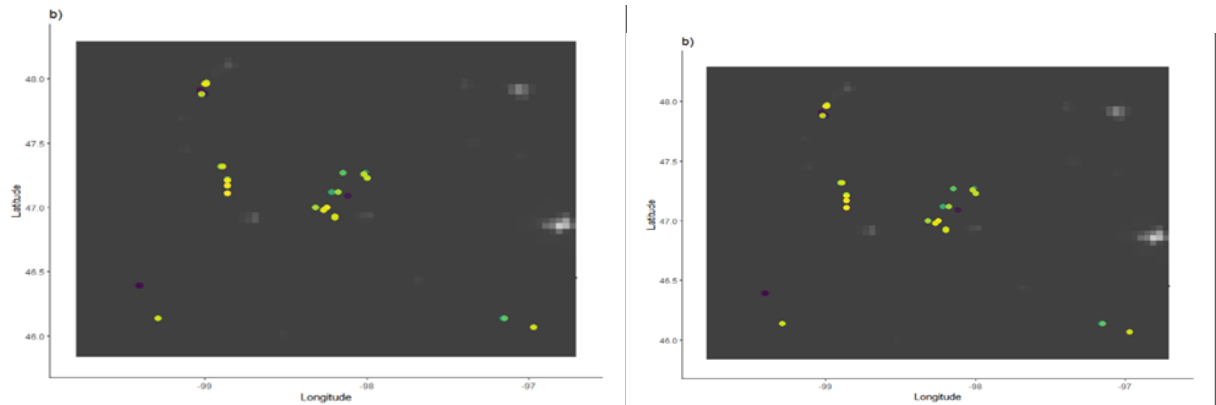


Figure 10: Distribution of Hymenoptera diversity plotted over human population density per state a) Texas, b) North Dakota, c) Colorado. The black to white scale represents human population density, with the lighter colors showing greater density. The colored scale represents Hymenoptera diversity, with lighter colors representing higher diversity.

This expected relationship is supported by the lack of significance in the results of a linear model. This may be due to inherent biases in the data since both our diversity and human population data were abnormally distributed, and most of the sampled sites were in rural areas and had smaller human population densities.

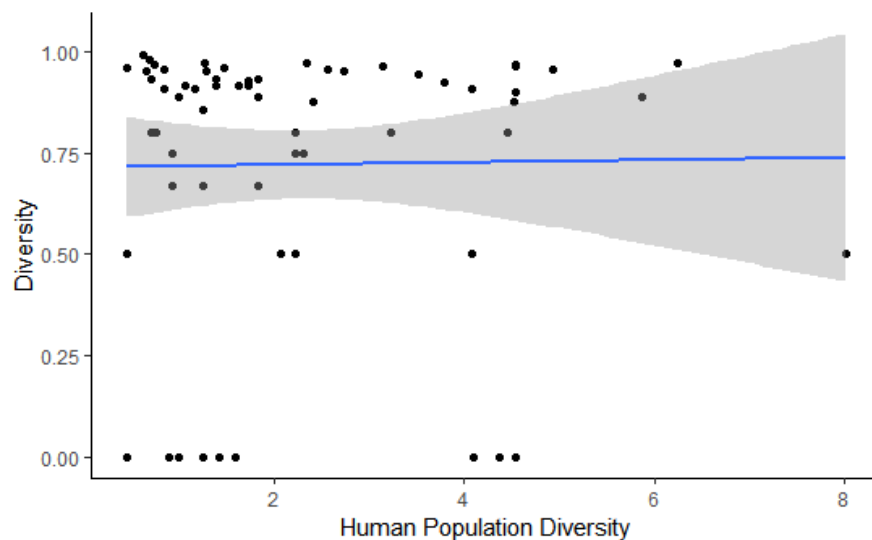


FIGURE 11: Linear Model showing the relationship between Diversity and Human Population Density. (Call = diversity~human population, df = 63, Residual SE = 0.3385, P value = 0.907, $R^2 = 0.0002183$, Adjusted $R^2 = -0.01565$, F Statistic = 0.01375 on 1, AIC = -1206.189)

The results show that human population density by itself is not a significant predictor of pollinator diversity. This is an interesting result since it directly contradicts our hypothesis. However, the lack of significant results may be due to increased variance in the model due to increased spatial variance. As seen by TABLE 10, When Moran's I was calculated, both human population density and pollinator diversity had significant results, meaning sites that are closer together are more similar than sites further away.

TABLE 10: Moran's I for Human Population Density and Hymenoptera Diversity

Variable	Observed	Expected	sd	p Value
Diversity	0.1971406	-0.015625	0.07928985	0.00728798
Human Population	0.4337529	-0.015625	0.07892905	1.245042e-08

To account for this, spatial variation was included as a random effect in the model through the use of a general least squares model. Different transformations of the data were tested against a null model without variation to assess which were the most significant. As seen in Table 12, the model that did not account for spatial variation was the most significant of all tested models, despite not improving on the original model. This means that spatial variation as a random effect is not a significant predictor of diversity.

TABLE 12: AIC table showing the AIC values of different models accounting for spatial variation. As shown by the results, the null model, which did not account for spatial variation was the most significant.

Model	df	AIC
Null Model	3	47.61988
Model with Exponential Transformation	4	53.38110
Model with Gaussian Transformation	4	54.92737
Model with Linear Transformation	4	54.40799
Model with Rational Transformation	4	53.85956
Model with Spatial Transformation	4	54.27076

Additionally, the relationship between diversity and human population density was tested per family to see if any of the families were more or less susceptible to human population density. As shown by TABLE #, this was not the case. Family-level diversity did not have a significant relationship with human population density.

TABLE 13: Family Diversity and Human Population Density. No family diversity was significantly predicted by the human population density. (call = diversity(per family)~human population)

Family	Term	estimate	Std.error	P.value	AIC value
Andrenidae	Intercept	0.934231990	0.018335628	1.237223e-67	-168.9334
Andrenidae	Human Population	-0.003156720	0.006589548	6.330783e-01	-168.9334
Apidae	Intercept	0.949548375	0.005748568	0.000000e+00	-1106.2355
Apidae	Human Population	-0.003350961	0.002356910	1.556505e-01	-1106.2355
Halictidae	Intercept	0.891840597	0.020742549	2.402847e-11	-138.4437

Halictidae	Human Population	-0.001357949	0.006925144	8.447094e-01	-138.4437
------------	------------------	--------------	-------------	--------------	-----------

Environmental and Anthropogenic Effects:

Since both environmental and Anthropogenic effects were significant in predicting pollinator diversity, we ran a comprehensive model that included all significant predictor variables. When compared to models that included anthropogenic effects or environmental effects only, the comprehensive model had the highest significance (TABLE #). This indicates that both anthropogenic and environmental effects are required to accurately predict pollinator diversity.

TABLE 14: Model Comparison between Environmental and Anthropogenic Effects. The model that included both anthropogenic and environmental effects was the most significant. It had the lowest Residual SE, p-value, AIC value in addition to resulting in the greatest R² and adjusted R² values.

Model	Call	df	Residual SE	P-value	R ²	Adjusted R ²	F statistic	AIC
Environmental Effects only	<u>Fixed:</u> Latitude * Relative Humidity * Air Temp <u>Random:</u> (1 + Latitude County)	/	/	/	/	/	/	- 1206.189
Anthropogenic Effects only	Human pop * land use	882	0.1168	<2.2e- 16	0.1043	0.09519	11.42 on 9	- 1286.879
Anthropogenic and Environmental	Latitude * Relative Humidity * Air Temp + Human Population * County * Land Use	711	0.06196	<2.2e- 16	0.7804	0.7586	25.6 on 71	- 2062.828

There are inherent biases associated with running large models such as this full model. Statistics such as R² tend to favour bigger models simply because there are more predictor values to explain the variation in the dependent variable. Therefore, we must run AIC tests to account for and penalize large models with many predictors.

Discussion

Environmental Effects and Diversity:

Following all analyses of environmental factors and their influence on diversity, we emerged on the other side of this long, dark tunnel with only one safely significant predictor - relative humidity. A higher relative humidity seems to be conducive to increased pollinator diversity, which does make intuitive sense as higher relative humidity is good for insects in general, as it lessens their desiccation stress. However, it is important to note again the biases present in our data as a result of its construction. Intuitively, we can already tell that humidity, wind speed, and air temperature should all be correlated, especially across a latitudinal and elevational gradient. However, the majority of our data comes from study sites and research plots from researchers that just happened to be working in certain areas. This dataset is by no means a transect study with a robust sampling design. We are therefore unable to draw as many conclusions as we'd like about latitudinal and elevational patterns, and even the conclusions we can draw about relative humidity, wind speed, and air temperature are to be taken doubtfully, at best. Further research is needed, especially transect studies (latitudinally and elevation-wise), to be able to shed light on the real relationship between all these geographical and climatic factors and pollinator diversity.

In addition, this dataset contains only occurrence data and provides no estimates or proxies for pollinator abundance or species composition. We would have liked to investigate further to see how climatic factors can affect individual species or genera in terms of their relative abundances, but that required a much better sampling design from the dataset.

Land Use and Diversity:

Our analysis of land use as an anthropogenic factor and its influence on pollinator diversity indicated that land use was a significant predictor of pollinator diversity in observed samples. Note, however, that the dataset violated both assumptions of normal distribution and homogeneity due to non-random sampling and the limitations of occurrence data. One of the most significant contributors to the non-normality of the land use data is uneven sampling; the only contributors to this dataset worked only in 3 states, in which researchers were asked to simply sample where they were working. This leads to a bias in land use types where researchers would likely be located. This is why we see significance in research plots in both our models. Land use categorization is also arbitrary as many datasets, including our own, fail to specifically define land use categories and what kind of land matrices are contained in these areas. Further research should involve more robust definitions of land use categories.

The habitat affinity of most pollinator species is unknown (Winfree et al. 2011). With increasing anthropogenic land use, it is important to understand the effects of land use on pollinator diversity. Further studies should include long-term monitoring of relative abundance and population structure in pollinators over a greater spread of land use types, including disturbed and undisturbed areas.

Human Population Density and Diversity:

Human population density by itself was not significant in predicting pollinator diversity. It is important to note the biases of our dataset could skew this data since data collected did not range across a gradient of human population densities and samples were biased to rural areas with low human population densities. More information about pollinator diversity in dense urban areas would allow for a more robust analysis of the relationship between human population and hymenopteran diversity.

Environmental and Anthropogenic Effects:

As seen in the results, both environmental and anthropogenic influences are necessary to accurately predict pollinator diversity. This has implications for legislation and conservation efforts. In order to effectively draft legislation, promote conservation efforts, and direct personal decisions that addresses pollinator preservation, comprehensive studies into the interplay between human action and environmental influences must be conducted.

Conclusion

Native pollinators specifically are incredibly important for local ecosystems due to their essential relationship with agriculture and natural ecosystems. Our group “beeplyr” analyzed an open source pollinator library dataset from the U.S. Geological Survey (USGS) website and tested the relationship between climatic, geographic, and anthropogenic factors on hymenopteran pollinator diversity in the U.S. Within the context of climate change and its ubiquitous effects on all aspects of life, and when accounting for the myriad of ways human action can change the global environment, diversity studies such as this are more important now than ever before. These allow for informed decision making, effective legislation and conservation efforts, in addition to increasing the information available regarding sensitive and ecologically significant species.

When all environmental relationships with diversity were tested and the data was appropriately filtered and transformed, only relative humidity was a significant predictor of pollinator diversity. A model testing the relationship between anthropogenic influences and diversity showed only land use as significant. It was only when combining all environmental and anthropogenic effects that the model had the most predictive results. This led us to conclude that both environmental and anthropogenic predictor variables must be considered to accurately describe their relationship with diversity.

Further studies of threats to and global shifts in pollinator diversity and abundance are crucial to long-term monitoring and conservation. Open source collaborative data sets are promising in their utilization in conservation research. We ran into limitations due to the incompleteness and biases of the data set. However, higher quality collaborative data collection can be achieved through the use of establishing more robust sampling protocols and regulating sampling sites. If given a more comprehensive dataset and more resources to do an in-depth analysis of the data, future steps include: analyzing the diversity and distribution of pollinated plants to assess greater ecological relationships, more inclusive dataset with less violations of assumptions to report results with more significance, pollinator data over longer periods of time to show temporal change, as well as more information regarding population structure and composition to illuminate community trends and relationships. We hope this information will be used towards encouraging more and better pollinator research, bettering conservation efforts, and informing policies for issues such as land management and environmental regulations in preserving threatened pollinator species.

References:

Data from: United States Geological Survey, Pollinator Library.

<https://www.npwrc.usgs.gov/pollinator/Downloads>. Accessed 11/07/2019.

Center for International Earth Science Information Network - CIESIN - Columbia University, and Centro Internacional de Agricultura Tropical - CIAT. 2005. Gridded Population of the World, Version 3 (GPWv3): Population Density Grid. Palisades, NY: NASA Socioeconomic Data and Applications Center (SEDAC). <https://doi.org/10.7927/H4XK8CG2>. Accessed 08/11/2019

Cox, N. J. 2005, November 5. Transformations: An Introduction.

<http://fmwww.bc.edu/repec/bocode/t/transint.html>.

Simpson's Diversity Index. (n.d.). . <https://geographyfieldwork.com/Simpson'sDiversityIndex.htm>.

Winfree, R., Bartomeus, I., and Cariveau, D. P. 2002. Native Pollinators in Anthropogenic Habitats. *Annual Review of Ecology, Evolution, and Systematics*. 42.1: 1–22.