An Analysis on the Relationship between Population and Biodiversity in New York State

Zeb Moffat

Abstract—This study explores the relationship between human population density and biodiversity across each county in New York state, aiming to understand how human urbanization may affect existing ecosystems. Using two publicly available datasets, one on biodiversity from the NYSDEC and the other from the U.S. Census Bureau, the analysis investigates patterns in species documentation and population in each county. Lasso, linear, ridge, and decision tree regression models were each applied to the data to find potential correlations between population and biodiversity. K-fold cross validation was used to mitigate overfitting due to the small number of counties. The results indicate no significant correlation between population size and number of species documented in each county. Despite these findings, this study suggests further analysis with different datasets to better understand the relationship between biodiversity and population.

Index Terms— Biodiversity, Population, Climate Change, Urbanization.

1 Introduction

Understanding the balance between human population and biodiversity is critical for maintaining healthy ecosystems and supporting sustainable communities. In the context of New York State, this balance holds a particular significance due to its diverse ecosystems and densely populated areas. Exploring these interactions provides insights into how human activities influence ecological systems and informs strategies to promote harmony between societal needs and environmental preservation. This paper explores two key questions: How does population affect biodiversity, how active are larger counties in measuring and monitoring biodiversity, and which biological groups are measured most? Addressing these questions provides valuable insight into the challenges and opportunities for conservation efforts in areas with diverse population densities.

Biodiversity can be defined as the variety of life on Earth. Biodiversity has been significantly declining in recent decades. The loss of biodiversity can be attributed to several factors including climate change, deforestation, urbanization, and invasive species. "Scientists estimate that current extinction rates exceed those of prehistoric mass extinctions. Loss of biodiversity also means loss of genetic diversity and loss of ecosystems" [1].

This analysis attempts to explore how population is directly connected to the loss of biodiversity through population size. While not indicative of the whole globe, looking at the variety of ecosystems and the immense population throughout New York State offers a greater understanding of how human population affects biodiversity. "Biodiversity brings important environmental services to our parks and communities. The variety of plant and animal life that occur naturally in these areas help to clean and protect our environment" [2].

It is widely known that population growth has had a lasting change on our planet and its habitats. "The growing human population has drastically altered Earth's ecosystems, transforming forests, grasslands, and other wilderness areas into farms, pastures, timberlands, mines, and settlements. People are now using around 71% of Earth's habitable land and are indirectly affecting the rest of the globe through pollution and climate change" [3]. This growing population causes issues for our planet, but discouraging human reproduction isn't the solution that everyone can agree on. Instead, emphasizing sustainable practices, making responsible eco choices, and promoting awareness on ways to treat the planet and its ecosystems, are ways to reduce the environmental footprint of all humans.

Understanding biodiversity and how it is affected by the population, even in one state, brings forward an understanding of how to help the planet. By studying this complicated relationship, more steps of action for conservation and sustainability can be identified. A cleaner, healthier planet is beneficial for all inhabitants.

1.1 Related Work

Biodiversity, Biosphere Reserves and the Big Apple: A study of the New York Metropolitan Region focuses on how biodiversity and the urban population interact. It looks at 31 of the New York Counties and how challenging it is to balance urban development and keeping biodiversity [4].

Changes in Frog and Toad Populations over 30 Years in New York State addresses roadside habitat loss, but other wetland populations have been fine. Frogs are a meaningful population in New York and can be a marker for overall biodiversity in the state [5].

Data-intensive Science: A New Paradigm for Biodiversity Studies reviews how important new technologies are in analyzing data for biodiversity. The authors believe in the importance of using data driven science for studying biodiversity [6].

Forecasting Global Biodiversity Threats Associated with Human Population Growth looks at how human population density affects biodiversity, especially mammal and bird populations. Most studies agree that population growth leads to losses in biodiversity, but the extent has been largely unclear [7].

Engaging the Public in Biodiversity Issues goes over the challenges of engaging the public with the growing issue of biodiversity. While a lot of focus nowadays goes towards climate change, it is primarily directed towards global warming, and rarely to other issues like habitat loss, pollution, and biodiversity loss [8].

Biodiversity in the City: Fundamental Questions for Understanding the Ecology of Urban Green Spaces for Biodiversity Conservation highlights just how important studying ecological processes in urban areas is for learning to effectively address biodiversity. With more research given to urban biodiversity, experts can better understand how to address growing biodiversity issues [9].

2 Exploratory Data Analysis

There are two datasets used in this analysis. The first dataset *Biodiversity by County - Distribution of Animals, Plants and Natural Communities* was collected by the NYS Department of Environmental Conservation (NYSDEC) and made available to the public on data.gov [10]. The second dataset *Annual Estimates of the Resident Population for Counties in New York: April 1, 2020 to July 1, 2023* was collected b the United States Census Bureau and made available to the public on census.gov [11].

The nice thing about both datasets is that they were both collected and provided by known and reputable institutions with the United States Government. The NYSDEC is known as a trusted authority on biodiversity data in New York. Their data is collected through thorough and scientifically sound methods, making it a reliable source for understanding the biodiversity of that state and its counties. Similarly, the U.S. Census Bureau is a globally respected organization known for the gathering of accurate population data. Both datasets cover every county in New York. This ensures full representation of the state's human population and ecological diversity. County level data allows for detailed analysis of trends in both urban and rural areas for population and biodiversity. This combination of trusted sources, recent data collection, and coverage of each county make these datasets representative of New York state. This data provides an accurate foundation for the analysis of the relationship between human population and biodiversity in New York. By relying on trustworthy datasets, conclusions that are drawn are more likely to be correct and useful in this study.

The following record is from the biodiversity dataset: Albany,Animal,Reptiles,Snakes,Carphophis-amoenus,Eastern-Wormsnake,2009,Special-Concern,not listed,S2,G5,Recently-Confirmed. This record contains information on the Eastern Worm snake in Albany county. Everything here makes sense. This snake is known to be endangered, but only in New York, so the data makes sense.

The following record is from the population dataset: Bronx County, New York; 1,472,653; 1,461,151; 1,424,084; 1,381,808; 1,356,476. This record contains the population data for the Bronx county in the years 2020, 2021, 2022, 2023. The Bronx is known for being a very populated county in New York, so that data makes sense in this case.

Figure 1 provides insight into the frequency each taxonomic group was documented in the state overall (added from each county). A taxonomic group is a category for biological classification, e.g. birds, amphibians, or snails. This is invaluable data because it highlights which groups are documented the most and could prove useful in deciding which groups may need extra attention in conservation efforts.

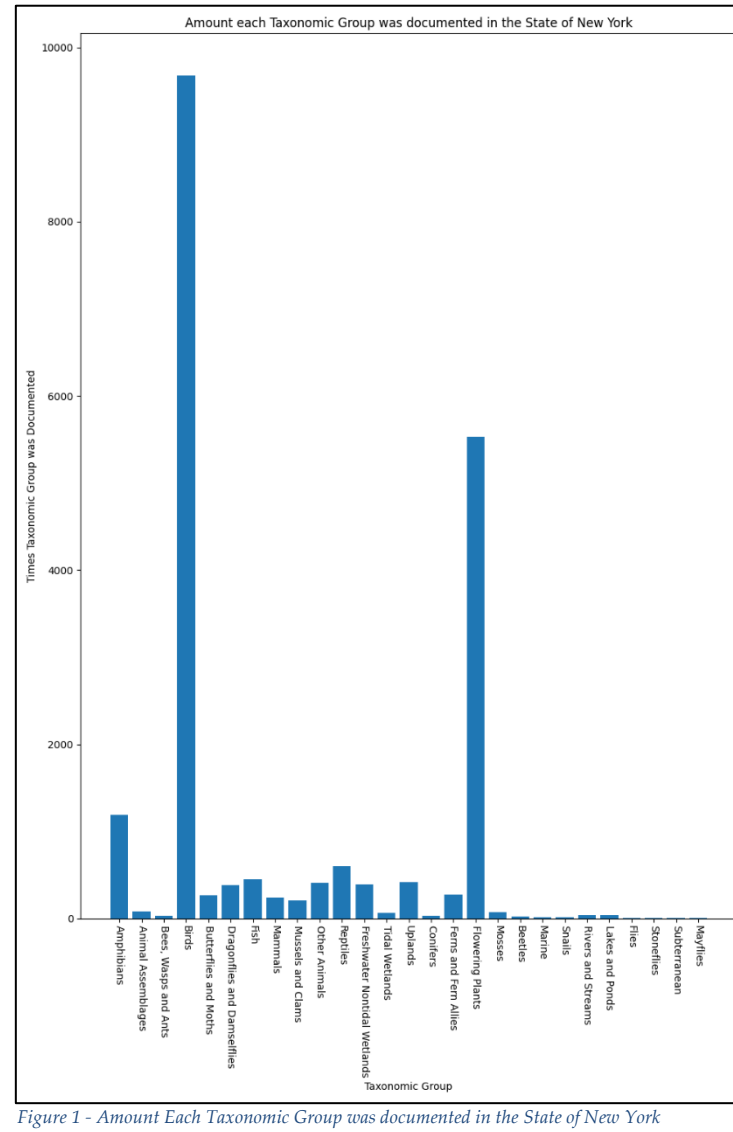

Figure 1 - Amount Each Taxonomic Group was documented in the State of New York

Figure 2 illustrates the relationship between each county population and how many taxonomic groups each documented. This relationship should be indicative of the relationship between how biodiverse each county is. The graph seemingly shows little to no relationship between group documentations and county population.

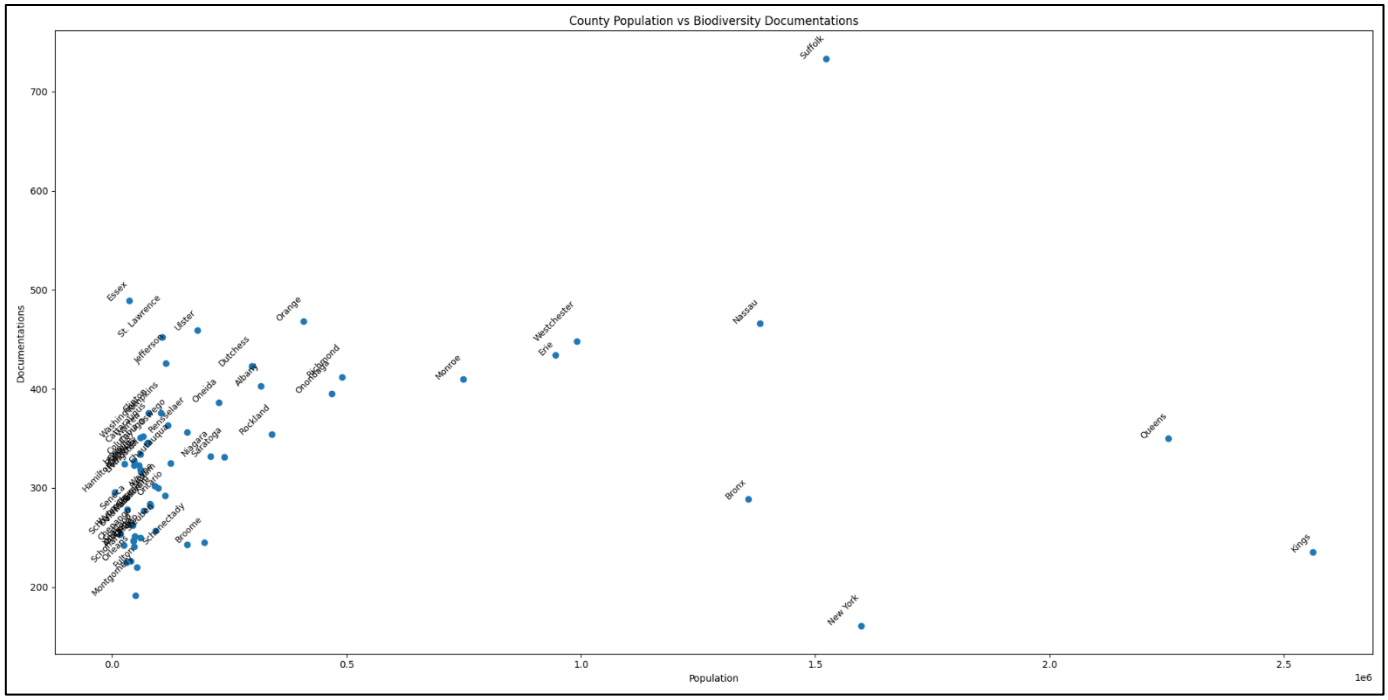

Figure 2 – County Population vs Biodiversity Documentations

3 Methodology

Due to the quantitative nature of this data, regression was picked to analyze the relationship between County Population and Biodiversity documentations to hopefully find some type of correlation. Four regression models were trained using the sklearn Python library. One issue encountered in training these models was that only 62 records could be used in training because there are only 62 counties in New York. To address this problem, k-fold cross validation was used. The data was divided into five equal subsets and the model was trained on four of these subsets, then tested on the remaining fold. This was then repeated four more times for a total of five. This technique helps mitigate overfitting in small datasets because it ensures each point is used in both the training and testing of the model. To evaluate each model, mean squared error and mean average error were used. Mean squared error is useful in finding large errors from the model. It takes large errors and emphasizes them, making them easy to find. It is especially good at finding outliers. Mean average error is useful in finding all errors. It is more balanced and doesn't point out outliers as much. The lower the MAE, the better overall the model is. These metrics are effective in finding a good regression model. The four models use lasso regression, linear regression, ridge regression and decision tree regression.

	Lasso	Linear	Ridge	Decision Tree
MSE	8325.0993	8325.0993	8325.0993	5962.5323
MAE	67.7205	67.7205	67.7204	17.4677

Table 1 – MSE and MAE for each Regression Model

Table 1 shows how each model performed in MSE and MAE. The issue is that the lasso, linear, and ridge models are performed basically the same. By a landslide Decision Tree regression is proven to be the best model for predicting biodiversity in the state of New York.

4 Results and Discussion

Plotting taxonomic groups highlights just how many birds and flowering plants have been documented in each county. By a landslide they are much higher than any other group. While this may be indicative that less measured groups deserve more attention in conservation, it may just be that there are not many of those groups within the state.

The result from plotting the biodiversity and population data seems to show that biodiversity and population do not have any strong correlation. The counties are all over the place in how many groups they measured, regardless of population.

For machine learning, three of the four models performed nearly identically. The fourth model performed significantly better, but due to the nature of a small dataset this may just be because of overfitting. K-fold cross validation was used to help mitigate this risk, but overfitting can still happen.

5 Conclusion

All data and analysis done in this analysis seems to point to low or no correlation between biodiversity and population in New York. I believe this is due to the biodiversity dataset being only documentations of species but not their counts for each county. My metric of the number of species documented in each county may not be truly indicative of how biodiverse each county is. While my analysis seems to show no correlation, I believe that there is one. I believe this analysis didn't go about it in a way that is truly indicative of how biodiversity and human population are related.

References

[1] "Sustaining New York's Animals, Plants and Ecosystems," *New York State Department of Environmental Conservation,* https://dec.ny.gov/nature/animals-fish-plants/biodiversity-species-conservation (2024).

[2] "Biodiversity," *New York State Parks, Recreation and Historic Preservation*, https://parks.ny.gov/environment/biodiversity.aspx (2024).

[3] "Natural Resources," *Population Connection*, Firefly Partners, https://populationconnection.org/why-population/natural-resources/ (2024).

[4] W.D. Solecki and C. Rosenzweig, "Biodiversity, Biosphere Reserves, and the Big Apple: A study of the New York Metropolitan Region," *The New York Academy of Sciences, https://nyaspubs.onlinelibrary.wiley.com/doi/abs/10.1196/annals.1319.004 (2006)*

[5] James P. Gibbs, K. Kristian Whiteleather, Frederick W. Schueler, "CHANGES IN FROG AND TOAD POPULATIONS OVER 30 YEARS IN NEW YORK STATE", https://esajournals.onlinelibrary.wiley.com/doi/abs/10.1890/03-5408 (2005)

[6] Steve Kelling, Wesley M. Hochachka, Daniel Fink, Mirek Riedewald, Rich Caruana, Grant Ballard, Giles Hooker, "Data-intensive Science: A New Paradigm for Biodiversity Studies", https://academic.oup.com/bioscience/article/59/7/613/334937 (2009)

[7] Jeffrey K. McKee, Paul W. Sciull, David Fooce, Thomas A. Waite, "Forecasting global biodiversity threats associated with human population growth", https://www.sciencedirect.com/science/article/abs/pii/S0006320703000995

[8] Michael J. Novacek, "Engaging the public in biodiversity issues", https://www.pnas.org/doi/full/10.1073/pnas.0802599105 (2008)

[9] Christopher A. Lepczyk, Myla F. J. Aronson, Karl L. Evans, Mark A. Goddard, Susannah B. Lerman, J. Scott MacIvor, "Biodiversity in the City: Fundamental Questions for Understanding the Ecology of Urban Green Spaces for Biodiversity Conservation", https://academic.oup.com/bioscience/article/67/9/799/4056044 (2017)

[10] NYSDEC, "Biodiversity by County - Distribution of Animals, Plants and Natural Communities", https://catalog.data.gov/dataset/biodiversity-by-county-distribution-of-animals-plants-and-natural-communities (2022)

[11] United States Census Bureau, "Annual Estimates of the Resident Population for Counties in New York: April 1, 2020 to July 1, 2023", https://www.census.gov/data/tables/time-series/demo/popest/2020s-counties-total.html (2023)