

Avian Diversity in the Research Triangle

https://github.com/arm132/McCollum_Schoenecker_ENV872_EDA_FinalProject

Rachel Schoenecker and Aurora McCollum

Contents

1	Rationale and Research Questions	4
2	Dataset Information	5
3	Exploratory Analysis	7
4	Analysis	9
4.1	Question 1: How did the diversity of the Research Triangle’s bird population change during the 2010s?	9
4.2	Question 1a: Is this trend consistent across all three main counties in the region (Durham, Orange, and Wake)?	10
4.3	Question 1b: Was there a clear trend in the presence of species of concern across the decade?	12
4.4	Question 2: How did the popularity of eBird change in the Research Triangle across the 2010s?	12
5	Summary and Conclusions	15
6	References	16

List of Tables

- | | | |
|---|---|---|
| 1 | Descriptions of the 12 variables selected for analysis within the eBird datasets. | 5 |
|---|---|---|

List of Figures

- | | | |
|----|---|----|
| 1 | Scatterplot showing species richness of each county for each month of the 10-year time period. | 7 |
| 2 | Scatterplot showing number of unique sampling events (USEs) in each county for each month of the 10-year time period. | 8 |
| 3 | Map of Durham, Orange, and Wake counties with points symbolizing each sampling location in the eBird dataset. | 8 |
| 4 | Time series decomposition for species richness aggregated across all three counties from 2010 to 2019. | 9 |
| 5 | Time series decomposition for species richness in Durham County from 2010 to 2019. | 10 |
| 6 | Time series decomposition for species richness in Orange County from 2010 to 2019. | 11 |
| 7 | Time series decomposition for species richness in Wake County from 2010 to 2019. | 11 |
| 8 | Number of SGCN observed in each county for each year of the study period. | 12 |
| 9 | Time series decomposition for event count aggregated across all three counties from 2010 to 2019. | 13 |
| 10 | Time series decomposition for event count in Durham County from 2010 to 2019. | 13 |
| 11 | Time series decomposition for event count in Orange County from 2010 to 2019. | 14 |
| 12 | Time series decomposition for event count in Wake County from 2010 to 2019. | 14 |

1 Rationale and Research Questions

According to a 2019 study by Rosenberg et al, North America's bird population has declined by 30% over the last fifty years. Many drivers are causing this steep decline, including the encroachment of human development into natural spaces as well as climate change. While this loss has occurred across habitat guilds and among both rare and common species, it also has important implications for the diversity of North America's bird population as some species near extinction (Rosenberg et al 2019). Additionally, the phenomenon of urbanization itself often supports a less diverse population of generalist species as opposed to habitat specialists (Callaghan et al 2019). This reduction of biodiversity has negative consequences for the resiliency of natural ecosystems, as well as various ecosystem services (Şekercioğlu et al 2004).

The Research Triangle region of central North Carolina is made up of some of the largest cities in the state, including Raleigh, Durham, and Cary. This region is characterized by mixed woodlands that support bird diversity (Minor and Urban 2010), and is currently experiencing rapid development (Doran and Golden 2016). In order to investigate changes in the diversity of the Research Triangle's population during the past decade (2010-2019), we accessed open source eBird data from the three main counties that make up the region (Durham, Orange, and Wake counties) and analyzed them with time series methodology. In this analysis, we hoped to answer the following questions:

1. How did the diversity of the Research Triangle's bird population change over the 2010s?
 - 1a. Is this change consistent across the three main counties in the Research Triangle (Durham, Wake, and Orange)?
 - 1b. Was there a clear trend in the presence of species of concern across the decade?
2. How did the popularity of eBird change in the Research Triangle over the 2010s?

2 Dataset Information

eBird is an open-source citizen science website created by the Cornell University Lab of Ornithology which allows the public to upload their own bird observation data collected from around the world (eBird 2022). The site has become incredibly popular with both amateur birders and scientific researchers since its creation. We downloaded bird observation data from the site for Durham, Orange, and Wake counties collected between January 2010 and December 2019. The data includes many variables including the common and scientific names of each bird species and observation data, time, and geographic coordinates. However, several variables in the dataset were not of use to us, so we selected 12 variables from the original set to create new, tidy datasets for each county (Table 1). This wrangling was performed with v. 4.1 of the R programming language (R Core Team 2021) as well as the tidyverse package (Wickham et al 2019).

Table 1: Descriptions of the 12 variables selected for analysis within the eBird datasets.

Column Name	Summary
X	Row number
COMMON.NAME	Primary English common name of the bird identified.
OBSERVATION.COUNT	The number of individuals at the time of observation. An ‘X’ is used when no count was made, to indicate presence.
COUNTY	What county the observation was made in (Durham, Orange, or Wake).
LOCALITY	The location name for the observation. These can be chosen from a list or the observer can name the location.
LOCALITY.ID	Unique alphanumeric code for a location.
LOCALITY.TYPE	Code defining the type of location: plot specific locations on a map (P), choose existing locations from a map (H), or choose to submit data for a town (T), postal code (PC), county (C), or state (S).
LATITUDE	Latitude of the observation in decimal degrees.

Column Name	Summary
LONGITUDE	Longitude of the observation in decimal degrees.
OBSERVATION.DATE	Date of observation
TIME.OBSERVATIONS.STARTED	What time did the observer start their sampling event (24 hour time)
OBSERVER.ID	The unique identification number given to a citizen observer.
SAMPLING.EVENT.IDENTIFIER	The unique number associated with the sampling event; a combination of location, date, observer, and start time.

Using the tidy data, we created tables for each county summarizing the number of unique species observed, or species richness, within each month of the decade. We used this metric as the indicator for bird diversity. We also created summary tables that showed the total number of unique sampling events uploaded for each month of the decade for each county. We used this metric as the indicator for the popularity of eBird.

In order to perform our analysis related to species of concern in the Research Triangle, we also used a dataset provided by the North Carolina Wildlife Resources Commission (NCWRC) detailing Species of Greatest Conservation Need (SGCN) in the state (NCWRC 2020). SGCN are determined by the NCWRC based on any policies concerning threatened or endangered species, risks related to the type of habitat they use, as well as other indicators of concern. The raw data included information on all SGCN identified in North Carolina, but through wrangling we filtered in order to create a vector that only included SGCN that were birds. By cross-referencing this vector with the eBird data, we were able to calculate how many SGCN were observed in each county during each year of the decade.

3 Exploratory Analysis

For our first step in analyzing our data, we created simple plots of species richness (Figure 1) and unique sampling events (Figure 2) across the decade to see if any clear trends could be identified. While all three counties exhibited possible positive trends in species richness, these were not entirely clear and required further analysis. There was a stronger positive trend in unique sample events that was visible for all three counties, but we also wanted to analyze this trend more precisely.

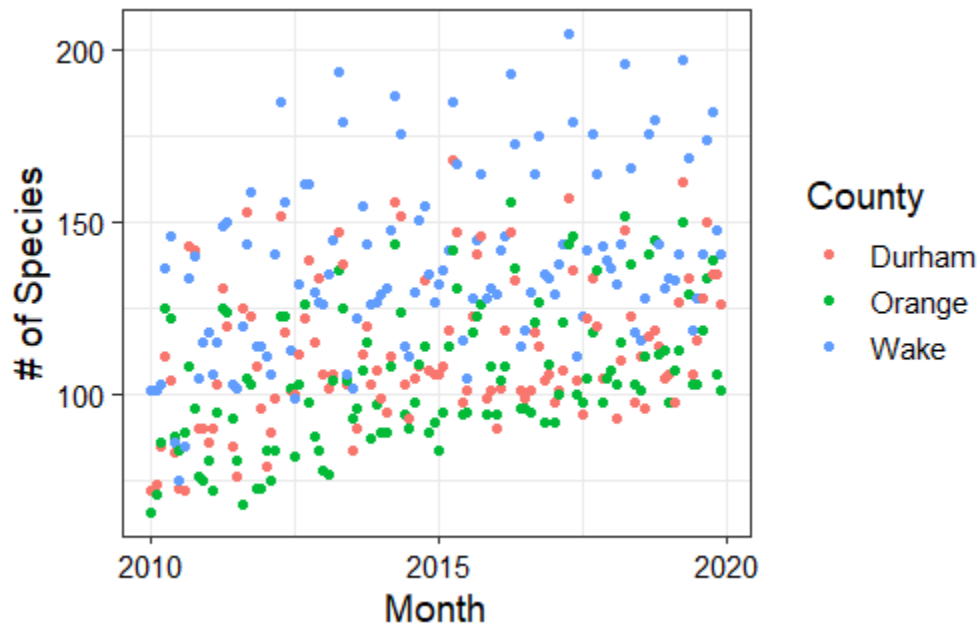


Figure 1: Scatterplot showing species richness of each county for each month of the 10-year time period.

Since the eBird data included latitude and longitude for every observation, the other component of our exploratory analysis included visualizing the location of each observation geospatially (Figure 3). We completed this geospatial analysis in R using the `sf` package (Pebesma 2018). This plot shows that the distribution of sampling locations was not evenly distributed across the region and that observations were concentrated around the urban centers of each county (Durham, Chapel Hill, and Raleigh).

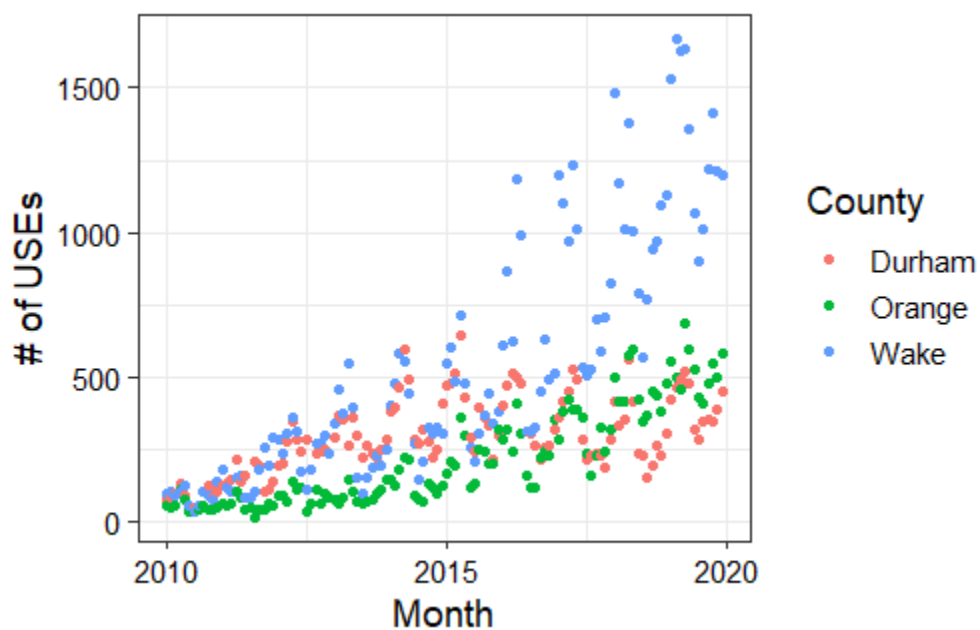


Figure 2: Scatterplot showing number of unique sampling events (USEs) in each county for each month of the 10-year time period.

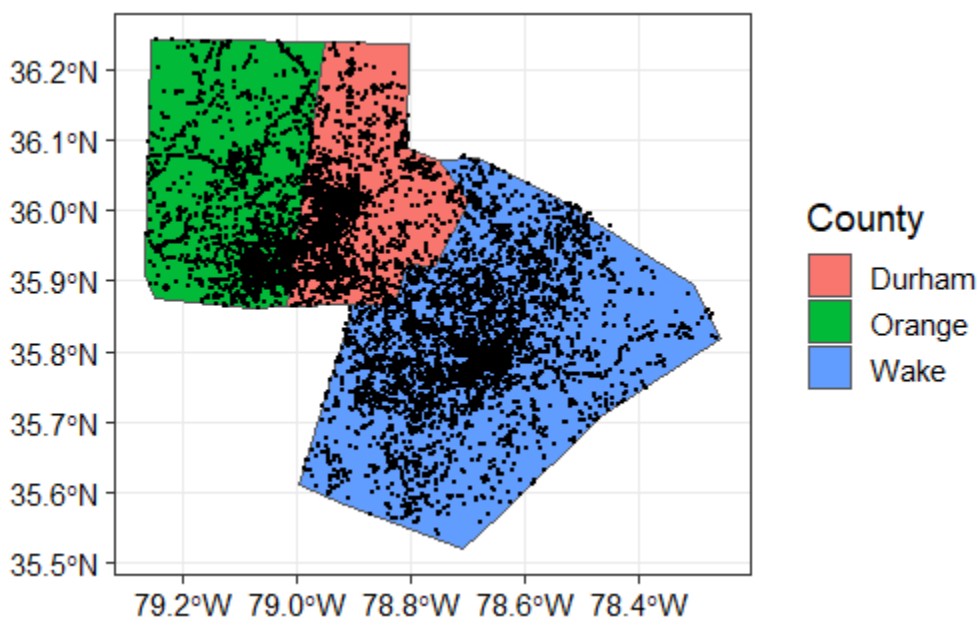


Figure 3: Map of Durham, Orange, and Wake counties with points symbolizing each sampling location in the eBird dataset.

4 Analysis

4.1 Question 1: How did the diversity of the Research Triangle's bird population change during the 2010s?

A time series analysis of species richness across all three counties combined showed both a seasonal trend as well as a strong positive monotonic trend (Figure 4). The seasonal trend likely correlates with the spring and fall migration seasons when more species would be present in the area. A seasonal Mann-Kendall test supported our identification of a significant positive trend in species richness across the time period (score = 195, tau = 0.366, $p < 0.001$). In order to complete this and all subsequent time series analysis in R, we utilized the zoo (Zeileis and Grothendieck 2005) and tseries (Trapletti and Hornik 2021) packages to construct the time series objects and the trend (Pohlert 2020) and Kendall (McLeod 2011) packages to perform the seasonal Mann-Kendall tests.

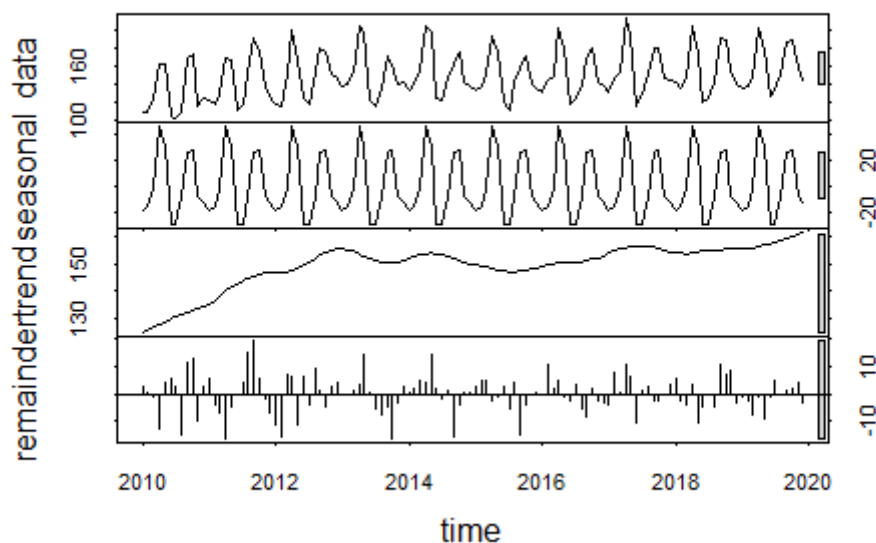


Figure 4: Time series decomposition for species richness aggregated across all three counties from 2010 to 2019.

4.2 Question 1a: Is this trend consistent across all three main counties in the region (Durham, Orange, and Wake)?

Individual time series analysis for each county also exhibited seasonality as well as positive monotonic trends (Figures 5-7). The pattern and rate of this positive trend was slightly different among the counties, however. While species richness saw a steeper increase earlier in the decade in Wake County, there was a more gradual increase throughout the period in Orange County, and Durham County saw a steeper increase in more recent years. Seasonal Mann-Kendall tests produced significant results for Durham (score = 172, $\tau = 0.322$, $p < 0.001$), Orange (score = 381, $\tau = 0.710$, $p < 0.001$), and Wake counties (score = 345, $\tau = 0.650$, $p < 0.001$).

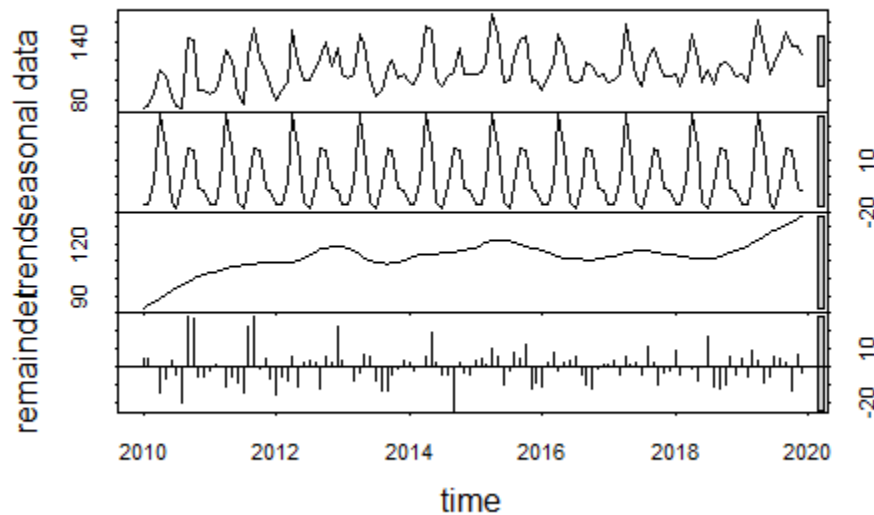


Figure 5: Time series decomposition for species richness in Durham County from 2010 to 2019.

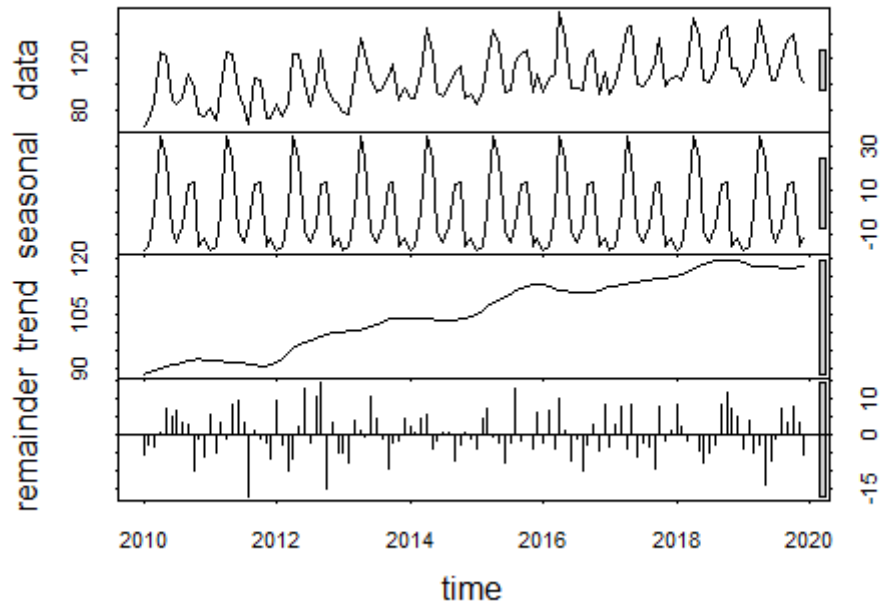


Figure 6: Time series decomposition for species richness in Orange County from 2010 to 2019.

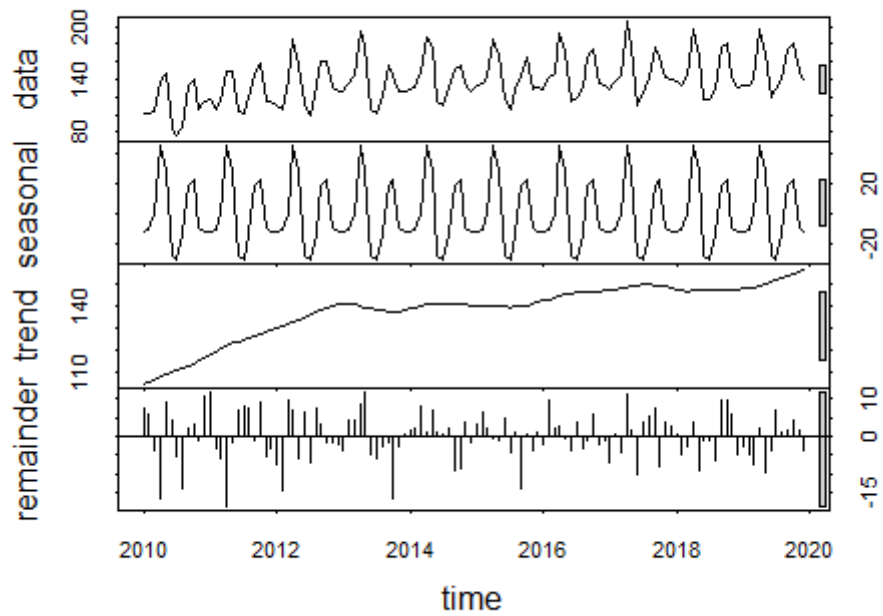


Figure 7: Time series decomposition for species richness in Wake County from 2010 to 2019.

4.3 Question 1b: Was there a clear trend in the presence of species of concern across the decade?

For each of the three counties, the number of SGCN observed in the final year of the period (2019) was noticeably higher than the number observed in the first year of the period (2010; Figure 8). However, none of the counties exhibited a constant increase in SGCN count across the entire decade, suggesting that further analysis is necessary to fully understand the trend in these data.

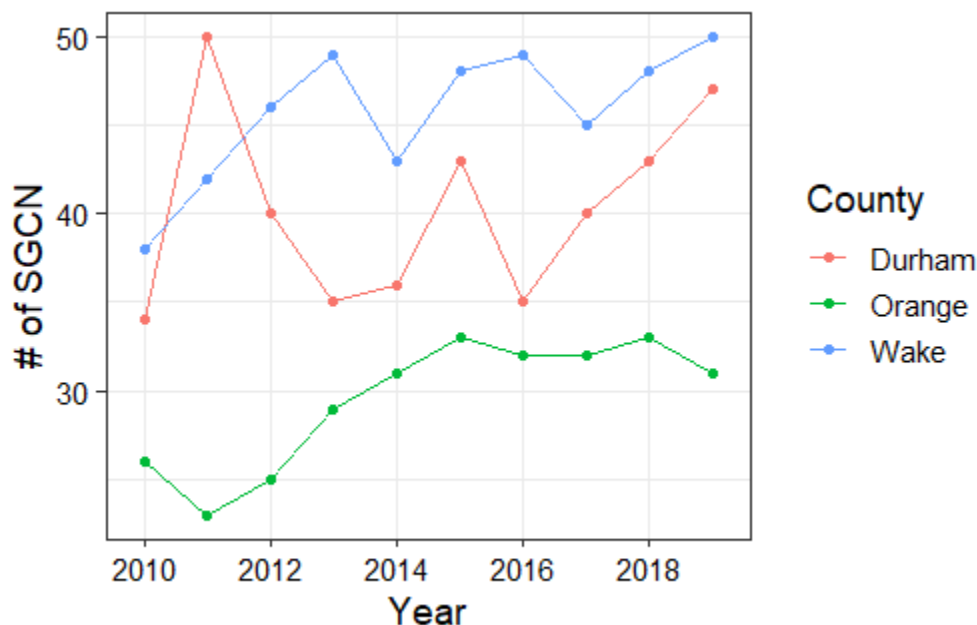


Figure 8: Number of SGCN observed in each county for each year of the study period.

4.4 Question 2: How did the popularity of eBird change in the Research Triangle across the 2010s?

A time series analysis of unique sampling events across all three counties combined showed both a seasonal trend as well as a strong positive monotonic trend (Figure 9). The seasonal trend in this time series is likely related to a higher frequency in birding activity during the spring migration and summer breeding season when weather conditions are pleasant. Our visual assessment of the decomposition was supported by a seasonal Mann-Kendall test that showed a significant positive trend in species richness across the period (score = 520, tau = 0.963, $p < 0.001$). This trend was also consistent within each county (Figures 10-12). Additionally, seasonal Mann-Kendall tests showed that this monotonic trend was significant for Durham (score = 283, tau = 0.525, $p < 0.001$), Orange (score = 495, tau = 0.918, $p < 0.001$), and Wake counties (score = 514, tau = 0.952, $p < 0.001$).

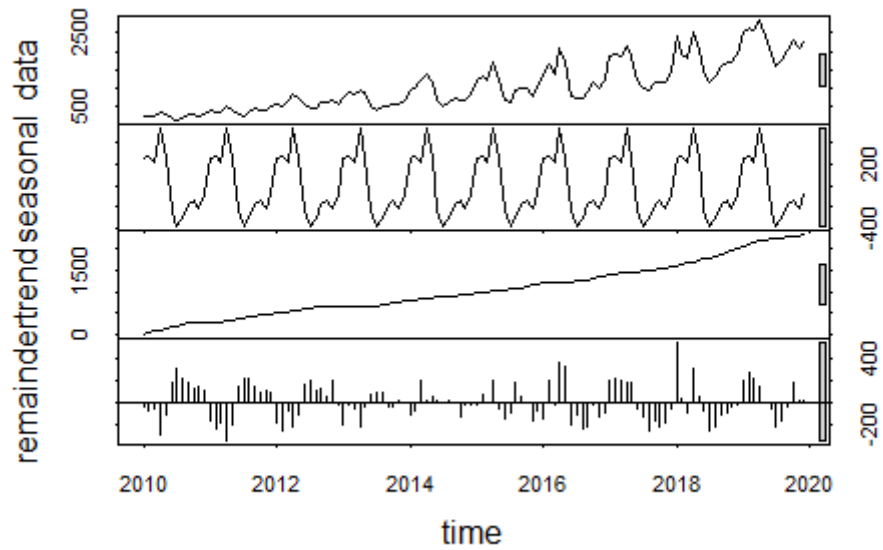


Figure 9: Time series decomposition for event count aggregated across all three counties from 2010 to 2019.

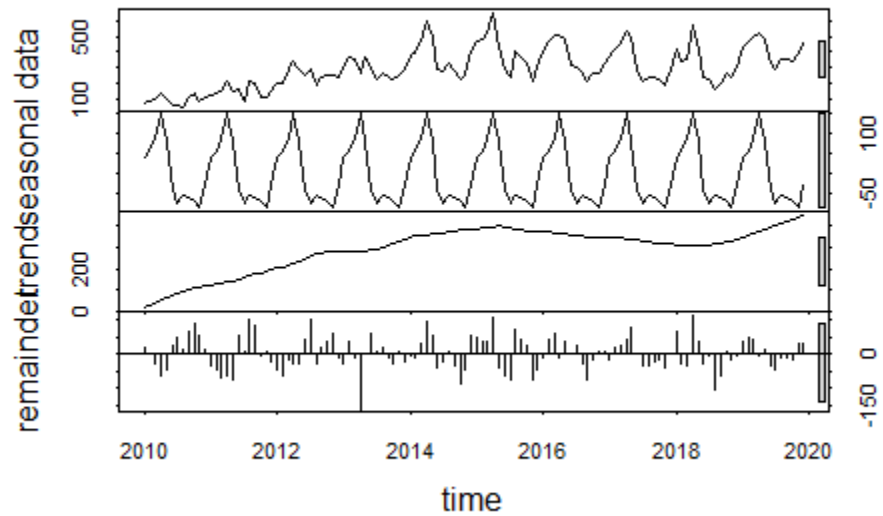


Figure 10: Time series decomposition for event count in Durham County from 2010 to 2019.

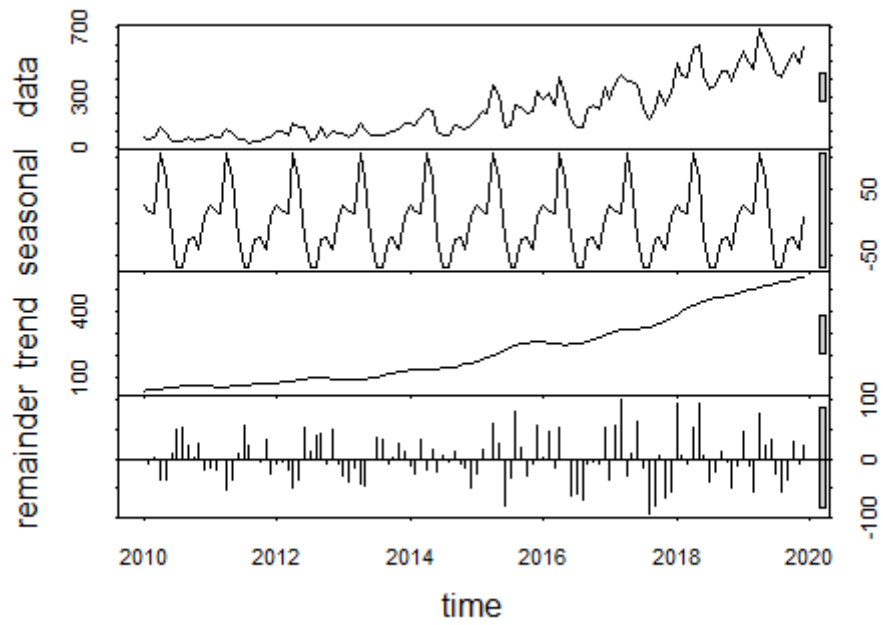


Figure 11: Time series decomposition for event count in Orange County from 2010 to 2019.

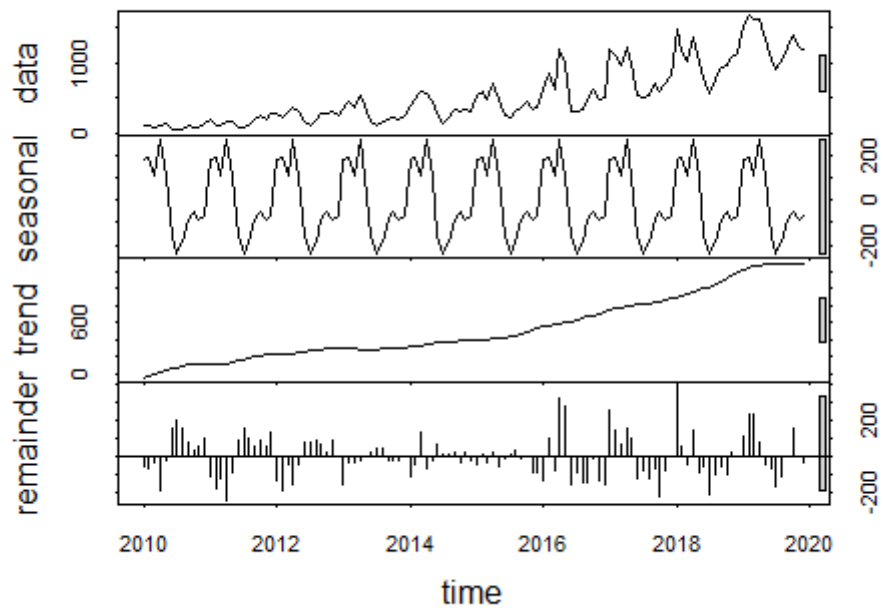


Figure 12: Time series decomposition for event count in Wake County from 2010 to 2019.

5 Summary and Conclusions

According to our analysis of eBird data, species richness and diversity increased in the Research Triangle between 2010 and 2019. This conclusion does not support our initial hypothesis that diversity of the Research Triangle’s bird population would have decreased due to human development. However, our other analyses invite the possibility that this increase in species richness may be a product of confounding variables.

Our time series analysis of unique sampling events across the decade showed that eBird popularity drastically increased from 2010 to 2019. If more people survey birds and upload their data, logically, this will lead to more species being observed. Therefore, based on our analyses it is not possible to say whether species richness increased, or how much of this increase was caused by changes in eBird usage. The conclusion that species richness increased is also not supported by the results of our SGCN analysis, which did not show any clear trend over time. If species diversity were increasing, it would be likely that the number of SGCN observed in the Research Triangle was also increasing. Because of this, we cannot make any definitive statements answering the question of how bird diversity in the Research Triangle changed during the 2010s.

Despite our lack of conclusive results, researching changes in bird populations across North America is incredibly important considering past literature that has shown the risks they face. In terms of future work related to this study, using a more consistent data source such as the US Geological Survey Breeding Bird Survey (USGS BBS), which is collected along pre-specified 25mi transects once a year, could help illustrate the changes we were looking for without the confounding variables brought on by the nature of open-source data. That being said, citizen science and datahubs such as eBird are still very valuable for conservation research and bolstering interest in environmental protection among the wider public.

6 References

- Callaghan, CT, Major, RE, Wilshire, JH, Martin, JM, Kingsford, RT, Cornwell, WK. 2019. Generalists are the most urban-tolerant of birds: A phylogenetically controlled analysis of ecological and life history traits using a novel continuous measure of bird responses to urbanization, *Oikos*. 128, 845-858. <https://doi.org/10.1111/oik.06158>
- Doran, EMB and Golden, JS. 2016. Climate & Sustainability Implications of Land Use Alterations in an Urbanizing Region: Raleigh-Durham, North Carolina, *Journal of Environmental Protection*. 7, 1072-1088. <http://dx.doi.org/10.4236/jep.2016.77096>
- eBird Basic Dataset. Version: EBD_relFeb-2022. Cornell Lab of Ornithology, Ithaca, New York. Feb 2022.
- McLeod, AI. 2011. Kendall: Kendall rank correlation and Mann-Kendall trend test. R package version 2.2. <https://CRAN.R-project.org/package=Kendall>
- Minor, E, Urban, D. 2010. Forest bird communities across a gradient of urban development, *Urban Ecosystems*. 13, 51-71. <https://doi.org/10.1007/s11252-009-0103-1>
- North Carolina Wildlife Resources Commission. 2020. Addendum 1 to the Wildlife Action Plan. Raleigh, NC. <https://www.ncwildlife.org/plan>
- Pebesma, E. 2018. Simple Features for R: Standardized Support for Spatial Vector Data. *The R Journal* 10 (1), 439-446, <https://doi.org/10.32614/RJ-2018-009>
- Pohlert, T. 2020. trend: Non-Parametric Trend Tests and Change-Point Detection. R package version 1.1.4. <https://CRAN.R-project.org/package=trend>
- R Core Team. 2021. R: A language and environment for statistical computing. R Foundation for Statistical Computing, Vienna, Austria. <https://www.R-project.org/>
- Rosenberg, KV, Dokter, AM, Blancher, PJ, Sauer JR, Smith, AC, Smith, PA, Stanton, JC, Panjabi, A, Helft, L, Parr, M, and Marra, PP. 2019. Decline of the North American avifauna, *Science*. 366, 120-124. <https://doi.org/10.1126/science.aaw1313>
- Şekerciöğlu, ÇH, Daily, GC, Ehrlich, PR. 2004. Ecosystem consequences of bird declines, *Proceedings of the National Academy of Sciences of the United States of America*. 101, 18042-18047. <https://doi.org/10.1073/pnas.0408049101>
- Trapletti, A and Hornik, K. 2021. tseries: Time Series Analysis and Computational Finance. R package version 0.10-49.
- Wickham, H, Averick, M, Bryan, J, Chang, W, D’Agostino McGowan, L, François, R, Grolemund, G, Hayes, A, Henry, L, Hester, J, Kuhn, M, Pedersen, TL, Miller, E, Bache, SM, Müller, K, Ooms, J, Robinson, D, Seidel, DP, Spinu, V, Takahashi, K, Vaughan, D, Wilke, C, Woo, K, and Hiroaki Yutani, H. 2019. Welcome to the tidyverse. *Journal of Open Source Software*. 4, 1686. <https://doi.org/10.21105/joss.01686>
- Zeileis, A and Grothendieck, G. 2005. zoo: S3 Infrastructure for Regular and Irregular Time Series. *Journal of Statistical Software*. 14, 1-27. doi:10.18637/jss.v014.i06