Rachamim-Yair Brito

Statistical Ecology

Dr. Brian Maitner

10/13/2025

Midterm

**Introduction:**

Ecology is an important sect of the biological sciences that allows us to understand how biotic and abiotic factors affect various environments and the species within them. A particular aspect of ecology that has had historical significance is avian ecology; with certain journals like The Auk and organizations like The Audubon Society having a historical focus on birds. Accessible via USGS, Ziolkowski *et. al.* (1) had compiled bird data from 1966 to 2021, looking at many different routes, across numerous years, noting what species were seen there and at what frequency alongside other variables like Latitude/Longitude, Temperature, total species count amongst others. Data like this is integral in gaining statistical understanding of how the avian communities at these locations change over time. This is of particular interest since humans are known to be the cause of over 1400 bird species having gone extinct (The Wildlife Society, 2). All of this culminates to the importance of the statistical analyses conducted for this paper. The breadth of data collected by Ziolkowski *et. al.* allows for an abundance of statistically significant questions to be asked. For the sake of this paper, the questions focused on are: How does the total species (TotalSpp) of birds at each route change over time (Year), how does the distribution of avian Orders at each Route change over time, what is the distribution of mean annual temperature in respect to both route and year, and finally what is the distribution of the fraction built (how much development is there in the area) with respect to route and year. The importance of all of these statistical analysis is to create a framework for additional more complex questions to be asked, for instance, how does total species change as fraction built go up. Although these relationships are something important to look at, due to time constraints this paper will strictly focus on providing example codes in R for each question, that can later be reran by changing the appropriate filters for each Route and Year, due to the immense amount of data present, one figure for each question is not appropriate due to visual clutter.

**Materials and Methods:**

The data for this paper originated from Ziolkowski *et. al.* (1)’s paper but was personally obtained from Dr. Brian Maitner’s GitHub (3), where he included additional spatial data such as: Mean annual temperature, Mean annual precipitation, Fraction built, and Human footprint. RStudio (4) was utilized to run all of the statistical analysis in this paper. The following packages within R were also used to help run the analyses and create the figures that will later be shown: ggplot2, bbmle, tidyr, readr, reshape, gplots, plotrix, dplyr, lattice, and e1071 (5-14 respectively). Finally, ChatGPT (15) and the book “Ecological Models and data in R” (16) were used to help understand why certain errors in the code occurred, as well as providing a direction for what code to use to perform the analyses, and how to interpret them. The type of analysis performed here were: a scatterplot looking at Total species in Route 1 across time (Year), a bar plot looking at the distribution of different Orders of birds specifically in Route 1 in 1997, lastly a histogram looking at the distribution of mean annual temperature of the whole data set, as well as another one looking at fraction built. For the plots, filter was used to home in on the specific part of the data that was of interest. For the histograms the unique() function was used since there was only one unique value for each Route + Year combination but was repeated in the dataset multiple times since each combination of Route + Year had multiple rows relating to the different species seen. Summary statistics such as the function summary() and str() were done to gain an understanding of the data structure, and class() was used to check if the variables of interest were the appropriate class, if they were not, they were reclasses using the as.numeric() function for example.

**Results:**

The Florida Breeding Bird Survey data was read into RStudio via this code:



In order to remove the variables that were not of interest a new data frame was made binding the columns that were pertinent via this code:



The first analysis done was a scatterplot looking at how total species (TotalSpp) changed in Route 1 across the years. The code below shows the summary statistics ran and how the plot was coded.

A screenshot of a computer screen

AI-generated content may be incorrect.

This yielded Figure 1.

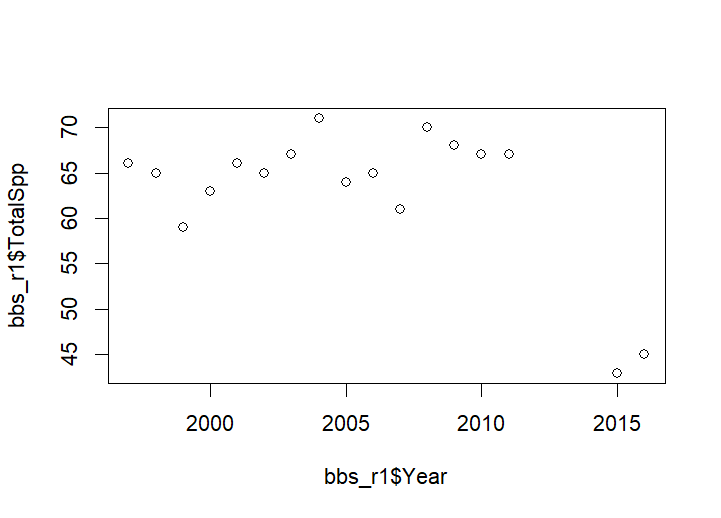


Figure 1: Total Spp of birds in Route 1 between 1996 and 2016

Up until around 2011 there seemed to have been an up trend in Total species count, but then the data sets from 2015 and 2016 showed a notably lower total species count. This same code should be ran for each route to note trends like this in other areas.

Then the barplot was coded to look at the the distribution of different Orders of birds in Route 1 in 1997. the filter function was used to filter for only the data important for yielding the results looked at here. The code the was used was as follows:

A computer screen shot of a computer code

AI-generated content may be incorrect.



Finally the code for the plot was this;

A computer screen shot of a code

AI-generated content may be incorrect.

This yielded Figure 2, which showed that the order Passeriformes was far more abundant (1002 sightings) than all of the rest of the Orders combined (second place went to Columbiformes at 56 sightings). This makes sense considering Passeriformes has the most species out of all avian Orders.

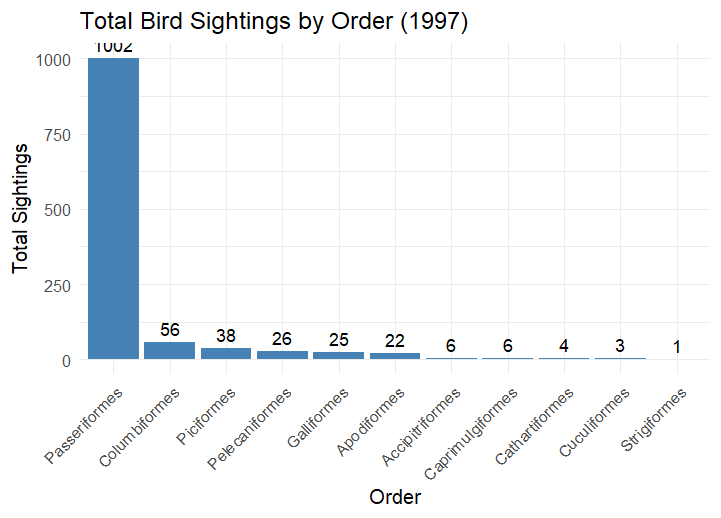


Figure 2 Bar graph looking at the total sightings of each Order of bird in Route 1 in 1997.

The Histograms for mean annual temperature and fraction built were coded as follows, and yielded Figures 3 and 4.

A close up of text

AI-generated content may be incorrect.

Breaks = 6

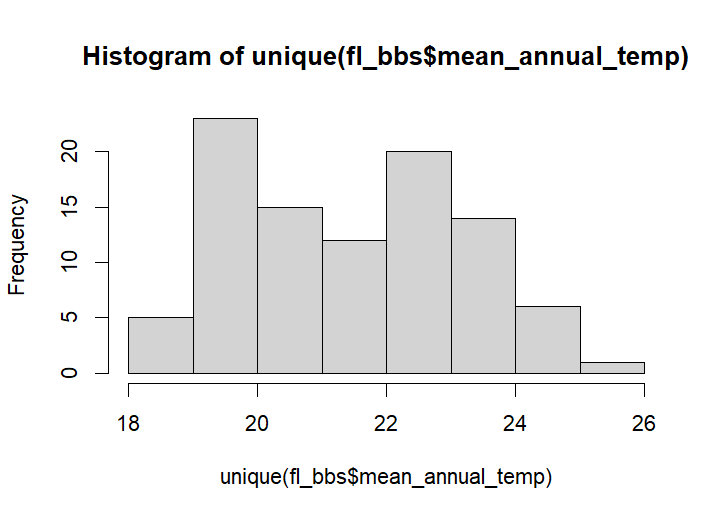
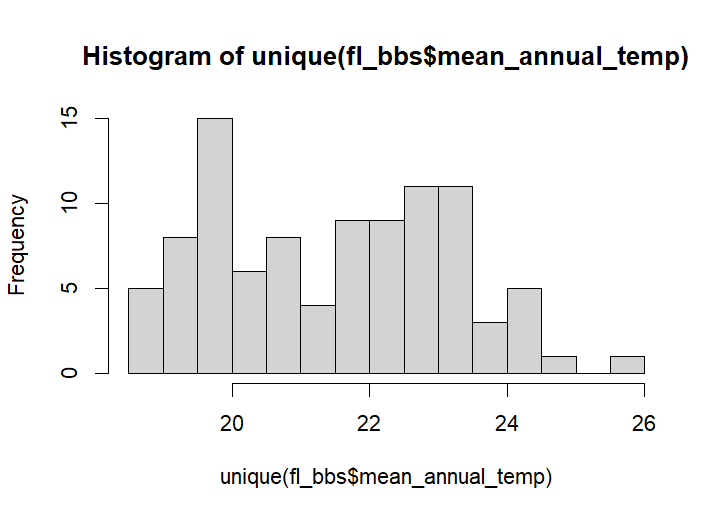


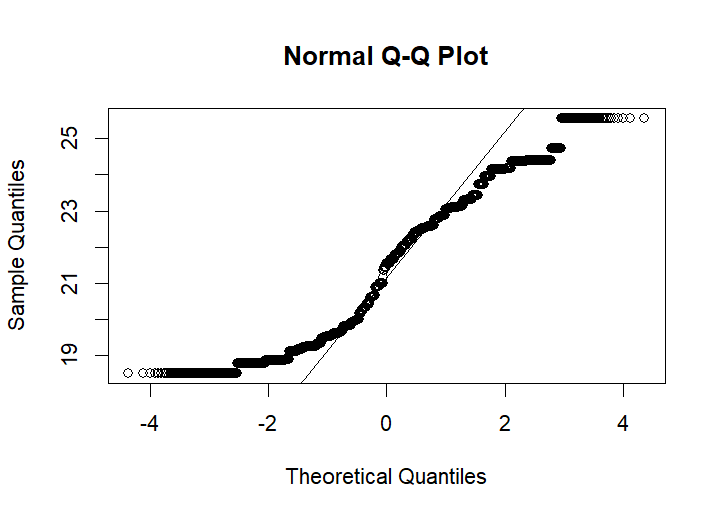
Figure 3 Histogram looking at the distribution of mean annual temperature throughout the whole data set

Breaks = 20



Mean annual temperature displayed a relatively normal distribution. A histogram with a break of 6 and 20 were performed to see if the distribution changes with higher resolution. They did not.

Q-Q plot:



Skewness = 0.129671, this indicates that the data is mostly Normally diostributed (not really any skewing).

The code for the fraction built histogram went as follows.

A close up of a text

AI-generated content may be incorrect.

Break = 6

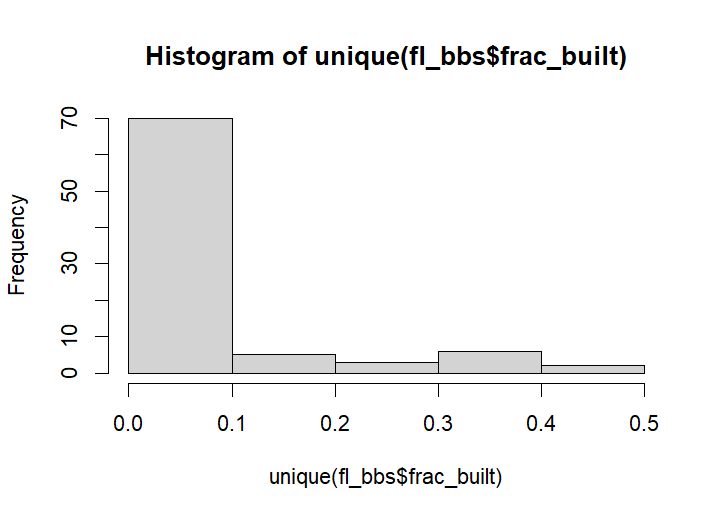
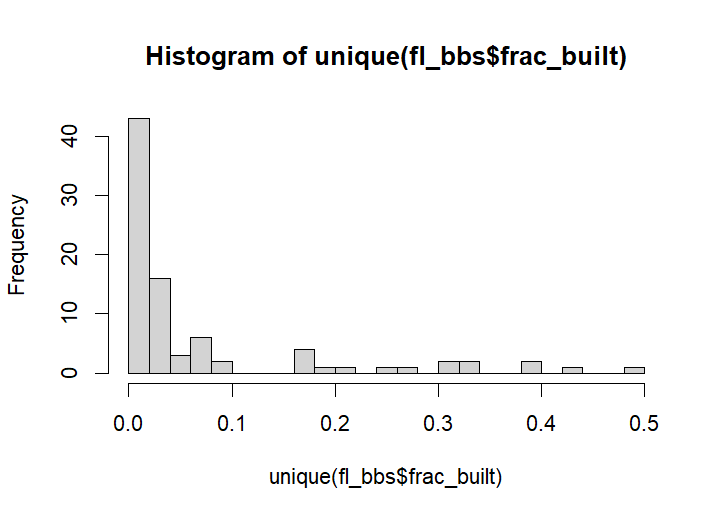


Figure 4 Histogram looking at the distribution of fraction built of the routes for the whole dataset.

Break = 20



Skewness = 2.001979 meaning heavily skewed to the right.

Most of the fraction built were on the lower end sub 0.1.

**Citation:**

1. Ziolkowski Jr., D.J., Lutmerding, M., Aponte, V.I., and Hudson, M-A.R., 2022, North American Breeding Bird Survey Dataset 1966 - 2021: U.S. Geological Survey data release, <https://doi.org/10.5066/P97WAZE5>.
2. The Wildlife Society. (2024, January 8). *Humans caused bird extinction throughout history.* The Wildlife Society. [https://wildlife.org/humans-caused-bird-extinction-throughout-history/](https://wildlife.org/humans-caused-bird-extinction-throughout-history/?utm_source=chatgpt.com)
3. B. Maitner. (n.d.). *FL\_BBS.csv* [Data set]. In *Statistical\_ecology\_course* (repository). GitHub. Retrieved October 14, 2025, from [https://github.com/bmaitner/Statistical\_ecology\_course/blob/main/data/BBS/FL\_BBS.csv](https://github.com/bmaitner/Statistical_ecology_course/blob/main/data/BBS/FL_BBS.csv?utm_source=chatgpt.com)
4. **RStudio Team. (2023). *RStudio: Integrated Development Environment for R* (Version 2023.09.0+463) [Computer software]. Posit Software, PBC.** [**https://posit.co/**](https://posit.co/)
5. H. Wickham. ggplot2: Elegant Graphics for Data Analysis. Springer-Verlag New York, 2016.
6. Bolker B, R Development Core Team (2023). \_bbmle: Tools for General Maximum

Likelihood Estimation\_. doi:10.32614/CRAN.package.bbmle <https://doi.org/10.32614/CRAN.package.bbmle>, R package version 1.0.25.1, <https://CRAN.R-project.org/package=bbmle>.

1. Wickham H, Vaughan D, Girlich M (2024). \_tidyr: Tidy Messy Data\_.

doi:10.32614/CRAN.package.tidyr <https://doi.org/10.32614/CRAN.package.tidyr>, R

package version 1.3.1, <https://CRAN.R-project.org/package=tidyr>.

1. Wickham H, Hester J, Bryan J (2024). \_readr: Read Rectangular Text Data\_.

doi:10.32614/CRAN.package.readr <https://doi.org/10.32614/CRAN.package.readr>, R

package version 2.1.5, <https://CRAN.R-project.org/package=readr>.

1. Wickham H (2007). “Reshaping data with the reshape package.” \_Journal of Statistical

Software\_, \*21\*(12). <https://www.jstatsoft.org/v21/i12/>.

1. Warnes G, Bolker B, Bonebakker L, Gentleman R, Huber W, Liaw A, Lumley T, Maechler

M, Magnusson A, Moeller S, Schwartz M, Venables B, Galili T (2024). \_gplots: Various

R Programming Tools for Plotting Data\_. doi:10.32614/CRAN.package.gplots

<https://doi.org/10.32614/CRAN.package.gplots>, R package version 3.2.0,

<https://CRAN.R-project.org/package=gplots>.

1. Lemon, J. (2006) Plotrix: a package in the red light district of R. R-News, 6(4): 8-12.
2. Wickham H, François R, Henry L, Müller K, Vaughan D (2023). \_dplyr: A Grammar of

Data Manipulation\_. doi:10.32614/CRAN.package.dplyr

<https://doi.org/10.32614/CRAN.package.dplyr>, R package version 1.1.4,

<https://CRAN.R-project.org/package=dplyr>.

1. Sarkar D (2008). \_Lattice: Multivariate Data Visualization with R\_. Springer, New

York. ISBN 978-0-387-75968-5, <http://lmdvr.r-forge.r-project.org>.

1. Meyer D, Dimitriadou E, Hornik K, Weingessel A, Leisch F (2024). \_e1071: Misc

Functions of the Department of Statistics, Probability Theory Group (Formerly:

E1071), TU Wien\_. doi:10.32614/CRAN.package.e1071

<https://doi.org/10.32614/CRAN.package.e1071>, R package version 1.7-16,

<https://CRAN.R-project.org/package=e1071>.

1. OpenAI. (2025). *ChatGPT (GPT-5)* [Large language model]. OpenAI. <https://chat.openai.com/>
2. Bolker, Benjamin M.. Ecological Models and Data in R, Princeton University Press, 2008. ProQuest Ebook Central,

http://ebookcentral.proquest.com/lib/usf/detail.action?docID=765289.

Created from usf on 2025-09-03 19:29:09.