Week8_AdamsKimberly

2022-07-30

Final Project Step 1

Introduction

Birdwatching is a very popular hobby in the United States and has a significant impact to the economy. Birdwatchers (or birders) spend money not only on their birdwatching gear, but often also travel to see specific species and thus promote ecotourism. With global warming, weather patterns are shifting with seasonality changes and temperature increases along with increased severity of weather events such as hurricanes. This changes could both impact bird populations and the eagerness of birders to go birdwatching.

I would like to look at how different weather aspects such as precipitation, temperature, and hurricanes effect bird observations both in terms of the number of checklists that are being submitted by birders and the diversity of species being observed across the state of Florida.

Data science allows us to bring together different information sources and datasets to look for patterns and hopefully provide insight into the larger picture than any one dataset alone. It also provides the rigor of statistical analysis to help us determine whether the patterns we perceive are merely coincidences or something more organized.

Research Questions

Temperature

1) What are the overall temperature trends?
2) Does temperature effect on the number of checklists submitted to eBird by birders?
3) Does temperature effect on the number of species observed by birders?

Precipitation

4) What are the overall precipitation trends?
5) Does precipitation effect on the number of checklists submitted to eBird by birders?
6) Does precipitation effect on the number of species observed by birders?

Hurricanes

7) What are the overall trends for number of hurricanes in a given year?
8) What are the overall trends for hurricanes intensity through the years?
9) Does the number of hurricanes in a given year effect the number of checklists submitted to eBird by birders?
10) Does the number of hurricanes in a given year effect on the number of species observed by birders?
11) Does average hurricane intensity in a given year effect on the number of species observed by birders?

Approach

My idea is to attempt to plot a series of linear regression models to try to ascertain trends within each of the variables and within variable interactions and see if any of the interactions are statistically significant. I anticipate that by using yearly averages of each variable that I will be able to avoid seasonal variances in the data both for the weather data and for the bird spring and fall migrations spiking bird species in the area.

Data

Bird Data

My primary dataset is from eBird.

eBird Basic Dataset. Version: EBD_relMay-2022. Cornell Lab of Ornithology, Ithaca, New York. May 2022. https://ebird.org/data/download

eBird is an online citizen science project started in 2002 by the Cornell Lab of Ornithology promoting conservation, wildlife education, and science-based studies. Birders from all over the world submit their checklists of sightings stating what species they saw and where and when they saw them. The goal of eBird data is to help scientists better track bird population trends both spatially and temporally. By involving the public, eBird is able to not only inspire a love of nature and sense of involvement, but at the same time collect an immense amount of data to see global bird population trends. Data entry from the public is reviewed by regional volunteers to promote consistency and accuracy.

The original data set has 49 columns. I requested all data from Florida only, but the dataset can be customized for a particular location and time period. Much of the variables are blank due to lack of input from the original birder during submission. These variables will not be useful for study, but certain fields like the date, species, and location among others are required fields and thus present. In the number observed column, it is possible to encounter an X where the birder submitted that the species was present, but the birder did not count how many individuals of that species there were. The data also includes a unique sampling event identifier which enables us to group observations by checklists.

Weather Data

My other two datasets are both come from the Florida Climate Center. The Florida Climate Center provides historical weather data and related information to any interested party and collaborates with the National Climatic Data Center. The datasets I am using both show the respective weather statistic (precipitation and temperature) for each month and the yearly average ranging from as far back as January 1895 to June 2022.

Temperature: https://climatecenter.fsu.edu/products-services/data/statewide-averages/temperature

Precipitation: https://climatecenter.fsu.edu/products-services/data/statewide-averages/precipitation

Since the eBird data only goes back to 2002, I won't technically need the weather data before that date, but It would be helpful when looking separately at the weather trends.

Required Packages

I will need/want:

- readxl – for reading excel files
- ggplot2 – for graphing
- auk – for working with eBird data

- ppcor – for partial correlations between variables

And also to wrangle the data:

- formattable
- plyr
- dplyr
- purrr

And probably others that I will discover I need as I go along and try to do something.

Plots and Table Needs

At very least I will want to graph each variable over time to see the general trends. Then I will also want to see if each variable has a normal distribution of values. I will also want to plot the interactions of each of the variables to see if a linear regression would fit the data well or not.

Questions for future steps

Since I haven't yet started plotting the data, I am not sure exactly what I will need. I could foresee the data not fitting a linear trend and therefore might need to learn how to work with non-linear modeling.

The eBird data is in a .txt file that is larger than excel can handle, so I am going to have to figure out how to get R to read that directly, but I think that should be similar to reading a csv file.

There are definitely things I need to get stronger on, but honestly I still feel like I am at the point where I don't' know what I don't know. I feel like I will find out as I go to do something and then realize I don't yet know how to do it.