CS251 Proj9

CS251 Project 9 - Final Project

- by: Tony Karalekas, Julia Saul, and Steven Parrott
- data file: Red-Breasted Nuthatch Data Set

Abstract:

Every year, thousands of bird species undergo great migrations to escape changing climates. The Red-Breasted Nuthatch stages souther migrations from their northerly breeding grounds in some years. We worked with a data set that measured the count of Red-Breasted Nuthatches on Christmas each year from 1962 to 2014. The data showed the extent of these southern migrations among states and provinces in North America.

With this data, we want to create a data visualization between Longitude, Latitude and Bird Count. Using directional data from each state/province in the data set, we are able to produce a visually appealing display of the birds, their population count and their location. We data-scraped Location data from the internet and added Latitude and Longitude to our .csv Data. This allowed us to create a visualization that appears as an outline of the North American provinces. In addition, we also worked with the data set to create a simple-linear regressions to find trends between Year and Species Count as well Latitude/Longitude and Species Count. We also implemented the ability for the user to select a range of Years to visualize the data over time.

Directions at bottom of page

Problem Statement:

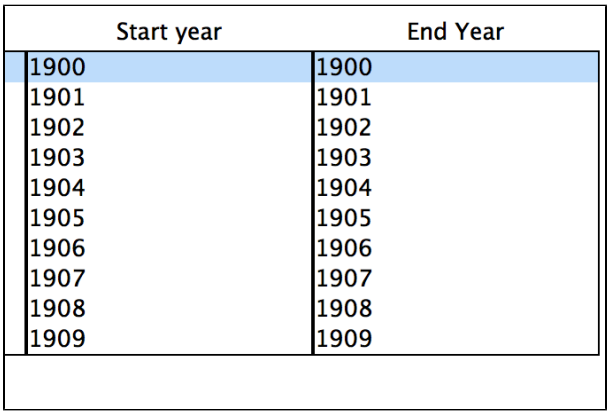
Our data set: "Red-breasted Nuthatch Christmas Count Data. Winter population dynamics of Red-breasted Nuthatches at the level of states and provinces. The data represent state- or province-wide mean abundances (measured as Number of Birds per Party-hour) from 1962 to 2014. Red-breasted Nuthatches stage southern migrations (called irruptions) from their northerly breeding grounds in some years. I am interested in the synchrony and extent of these southern migrations among states and provinces."

This data set is directly from the Colby College Biology Department and Professor Herb Wilson. The features/headers for this data are Region, Count_Yr(the # year completed), Species Number, Year, Reporting Counts(Number of Reports), Reporting Observers(Number of Observers), and we added Longitude and Latitude. As we stated in the abstract, we the problems we developed our program to solve with this data is 1) a visualization by region 2) the option to view by year/time period and 3) a couple simple linear regression.

Methods:

In order to add latitude and longitude, we data scraped from the internet and manually created two new columns in our original data set. Now, if the user selects longitude and latitude to be x and y-axis, the data will appear in coordinate locations of each state in our data set. The visual aspect of our project can be seen by the user by pressing "Choose new axes" where they can select a year range.

In order to allow the user to select ranges of year when they click the button "Choose new axes", we added two new list boxes to the dialog box where the user selects the parameters of the visualization. The first list box allows the user to select a start date and the second list box allows the user to select an end date.



Start year	End Year
1900	1900
1901	1901
1902	1902
1903	1903
1904	1904
1905	1905
1906	1906
1907	1907
1908	1908
1909	1909

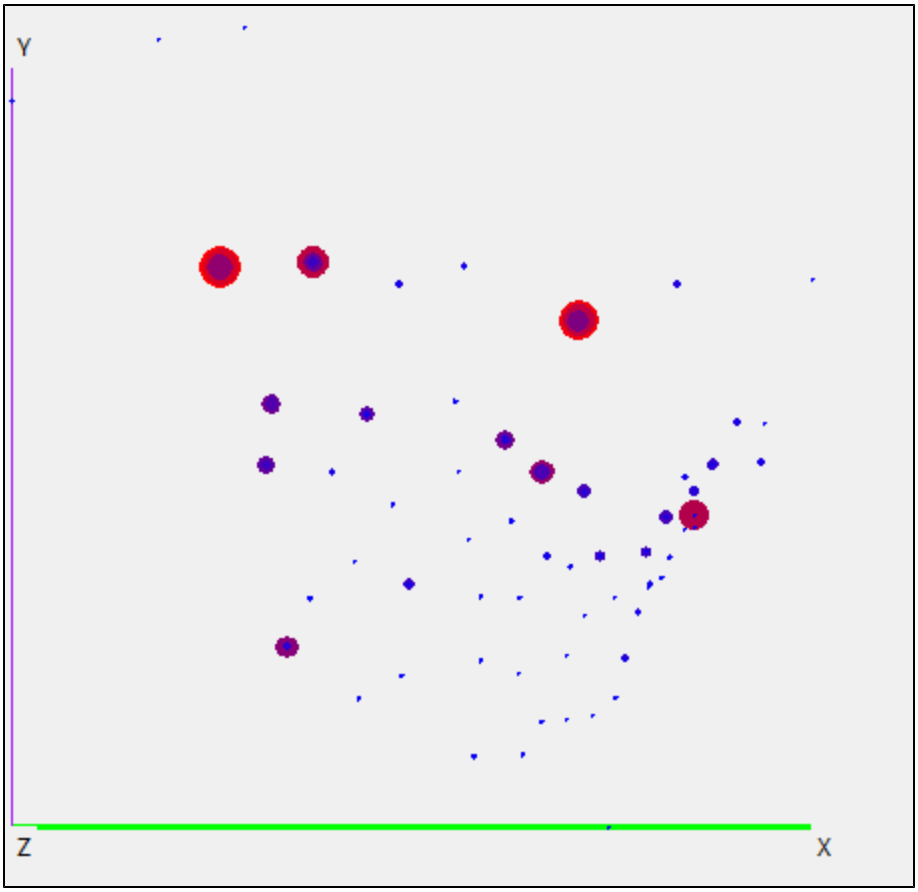
All other list boxes have the header "Years" removed. The years are added to each list box using a for loop. In handleChooseAxes, a new field called selectedyears list is filled with the two years that were selected. A new data object is then initialized to None called new_data. In order for only the data from the selected years to be plotted, a csv file is written. This new csv file is created and called plot.csv. The first row written is with all of the headers and the second row written is with all of the types. Then, using a for loop, every year that falls between the range of the selected years has its entire row added to the new csv file plot.csv. Then, the new data object new_data is created with the new csv file to be read in during buildPoints.

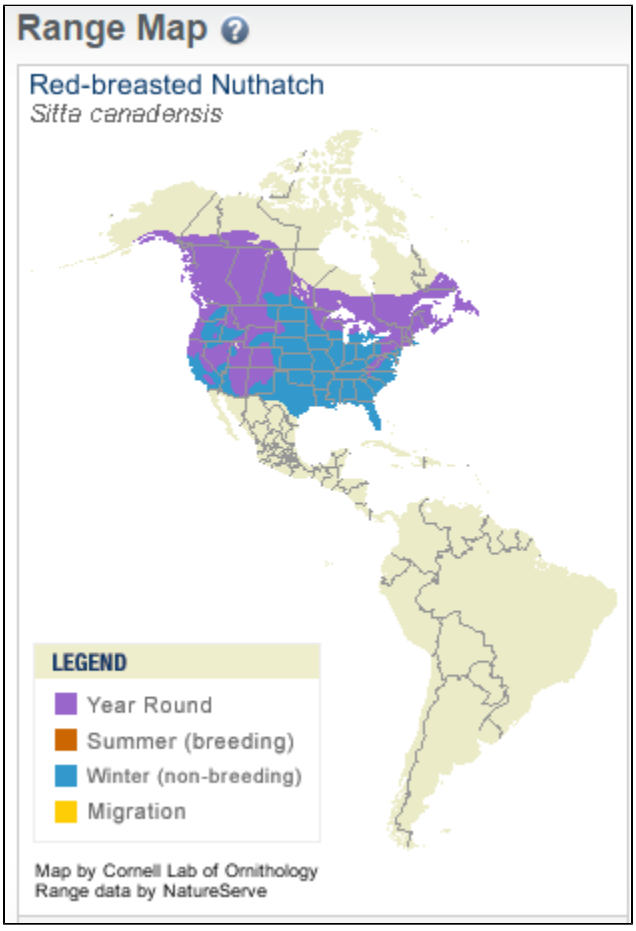
```
self.new_data = None

with open("plot.csv", "wb") as csvfile:
        writer = csv.writer(csvfile, delimiter= ',', quotechar='|',
                            quoting= csv.QUOTE_MINIMAL)

        writer.writerow(self.data.get_raw_headers())
        writer.writerow(self.data.get_raw_type())


        for i in range(self.data.get_raw_num_rows()):

            if self.data.get_value( i, 'Year') >= self.selectedyears[0]
                and self.data.get_value( i, 'Year') <= self.selectedyears[1]:
                writer.writerow(self.data.get_raw_row(i))

self.new_data = data.Data("plot.csv")

self.buildPoints()
```

In buildPoints, instead of passing in the original data object to be normalized, the new data object created from plot.csv is read in. This way, whatever years the user selects will decide the only data points that will be visualized. It was important to create a new data file so the original data file won't be overwritten and unable to be used later.

Results:

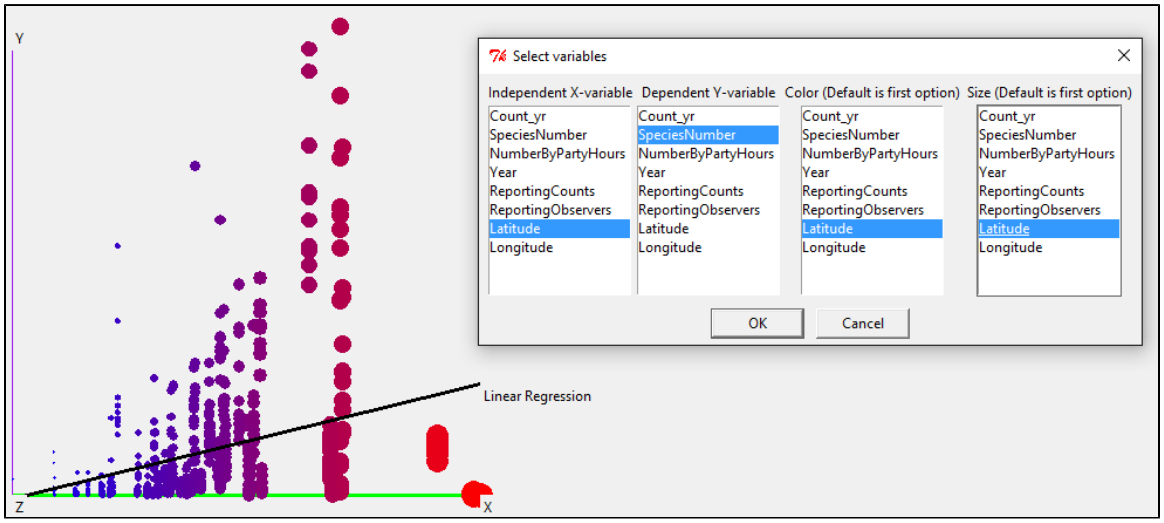
This is what our data looks like when the year range 2004-2013 is plotted by clicking "Choose new axes". This corresponds with a map of the species range included below our plotted data.


Y
Z
X

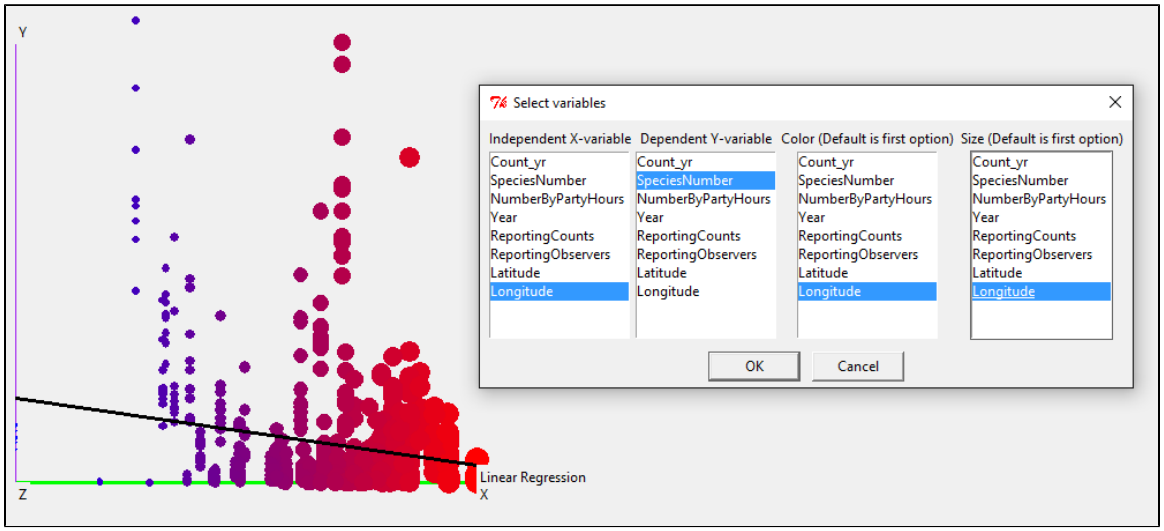

Range Map
Red-breasted Nuthatch
Sitta canadensis
LEGEND
Year Round
Summer (breeding)
Winter (non-breeding)
Migration
Map by Cornell Lab of Ornithology
Range data by NatureServe

We then ran various data analyses on our data.

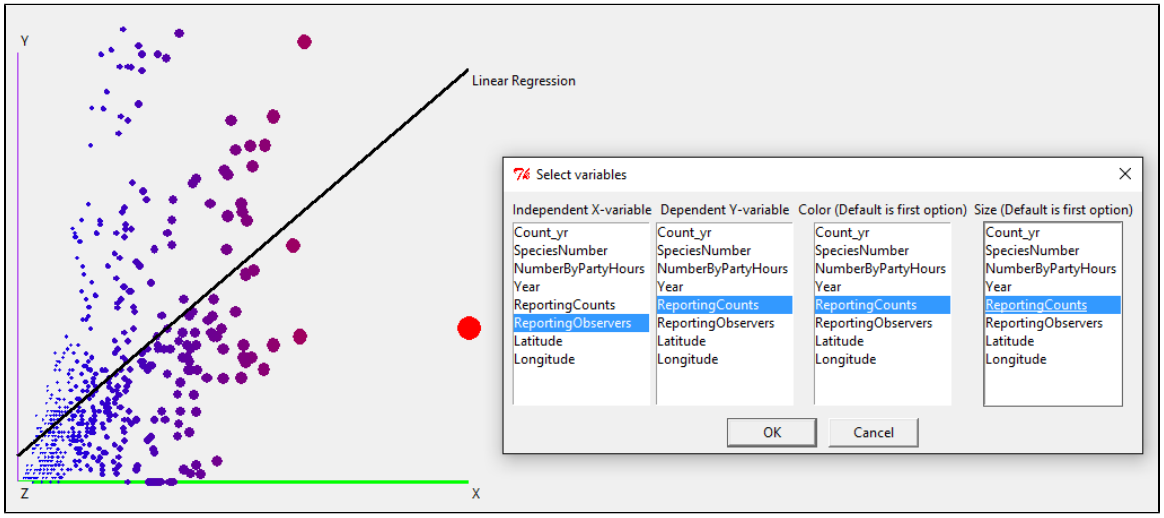
Data analysis 1: Latitude vs species number. This shows that the farther North you are, the more birds there are.



Data analysis 2: Longitude vs species number. This shows that the farther west you are, the more birds there are.

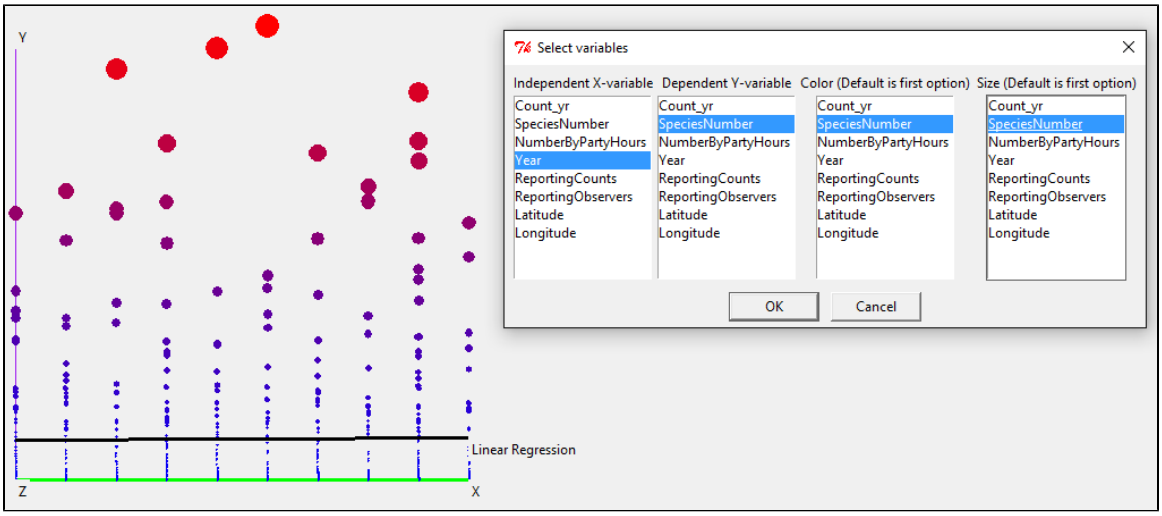


Data analysis 3: Reporting observers vs reporting counts. This shows that with more people observing birds, there will be more bird spottings.



Data analysis 4: Species number vs. Year. Each "column" represents a different year. This shows that the species population has not changed

much from 2004 to 2013.



Summary/Conclusion:

Overall, through our linear regressions we learned that the Red Breasted Nuthatch populations are greatest in the north-west region of the United States and Canada. In addition we learned that the more people observing birds, the more birds we are going to see (which is rather obvious). In addition, we were able to successfully implement a function that prompts the user for a selected date range in which they wanted to view the data. This function then excluded all the data that was not selected and saved the data we wanted to a new csv file. This new csv file is called "plot.csv" cand can then be read into our display application to perform regressions.

Note to Grader:

We received some help/guidance with our idea outline from Bruce. In addition, the final project code and working directory is turned in withing Julia Saul's Courses Folder. That is the code we all worked on and would like to turn in for grading. Finally, obviously, this is the writeup that we would like to be used for our grading process; the three of us collaborated on every piece of this project --both code and writeup.

Directions: Open Data Set > Choose New Axes > Select Any Data and Select Years You Want > Data is then Plotted > Then Open New CSV File to perform analysis.

Note: The new CSV file is overwritten every time Choose Axes is selected.