



No Need for Binoculars: Visualization of Avian Citizen Science Data Using JMP® 10

Rosemary Lucas, Development Tester, JMP Division, SAS



Introduction

Citizen science projects, where both scientists and the public can participate in scientific research, are gaining in popularity. These projects contribute to the collection of data in a diversity of fields, sometimes resulting in huge databases of information publicly available for downloading. One of these projects is eBird¹, an online checklist program where amateur and expert birders submit observations on types of birds and their numbers, and information on when and where the birds were seen. With over **40 million records**, this checklist database is a rich resource for exploration. JMP® 10 is an excellent tool for its exploration, providing easy methods for data acquisition, preparation, and display of temporal and spatial distributions of over **800 bird species**.

Data Preparation and Visualization

1. Downloaded and unzipped the Version 3.0 eBird¹ reference data set from Avian Knowledge Network site.
2. Imported the large CSV files using **Text Import**. The largest resulting table was ~5 GB in size.
3. Converted special characters in numeric fields using **Recode**, then converted columns to numeric data type.
4. Used **Compress Selected Columns** to decrease size of tables. The 5 GB table shrank dramatically to under 1 GB.
5. Merged covariate and taxonomic information from related data sets using **Join** and **Update** platforms.
6. Filled missing FIPS(county) values and added zip codes with latitude/longitude to FIPS/zip codes JSL scripts.
7. Shaped data using **Summary**, **Subset**, **Tabulate**, and column formulas.
8. Visualized with **Graph Builder** Bar Chart and Maps. Used **Column Switcher** feature to observe patterns over time.

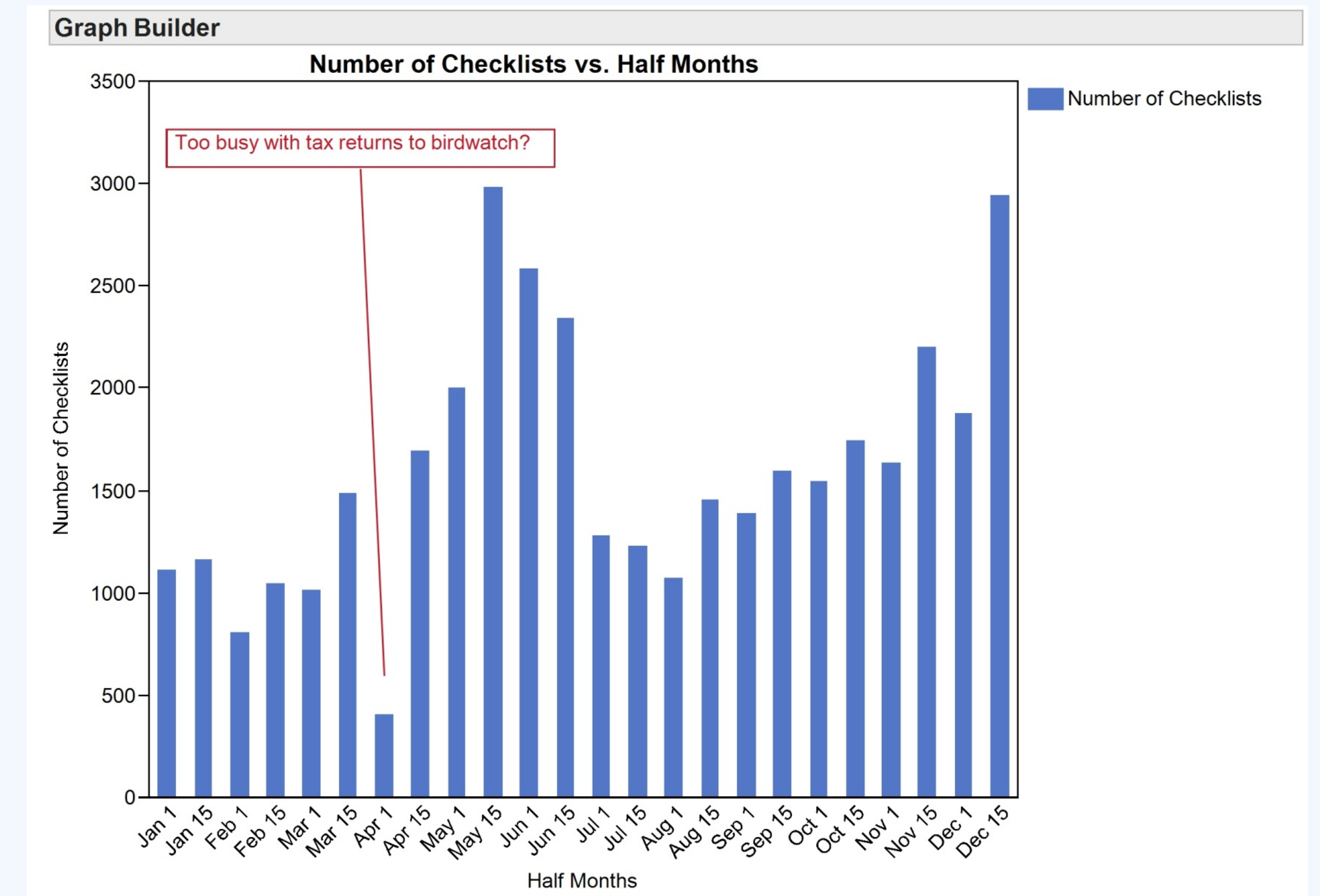


Figure 1. 2008 Checklist totals for half-months, using Bars in Graph Builder. Note the dip in checklists submitted in early April.

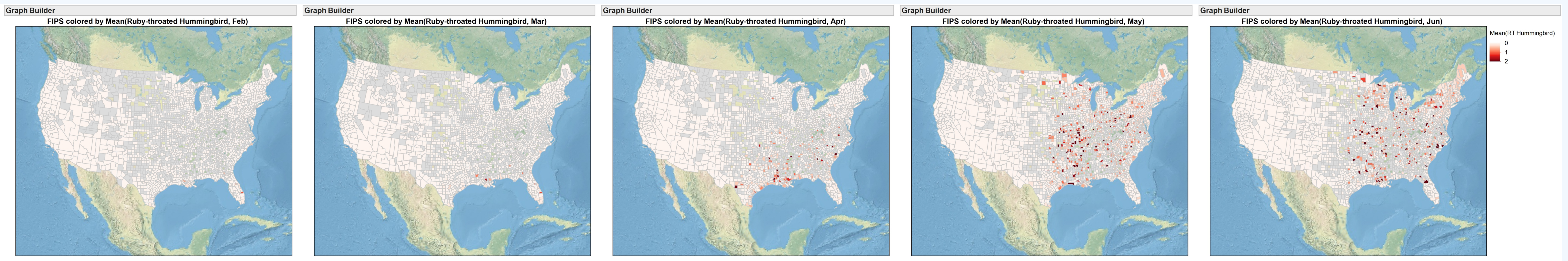


Figure 2. Average Ruby-throated hummingbird counts by FIPS(counties) for the year 2010, months February through June. Observe the influx from the south over time. Patterns like these are found easily using Graph Builder and the Column Switcher.

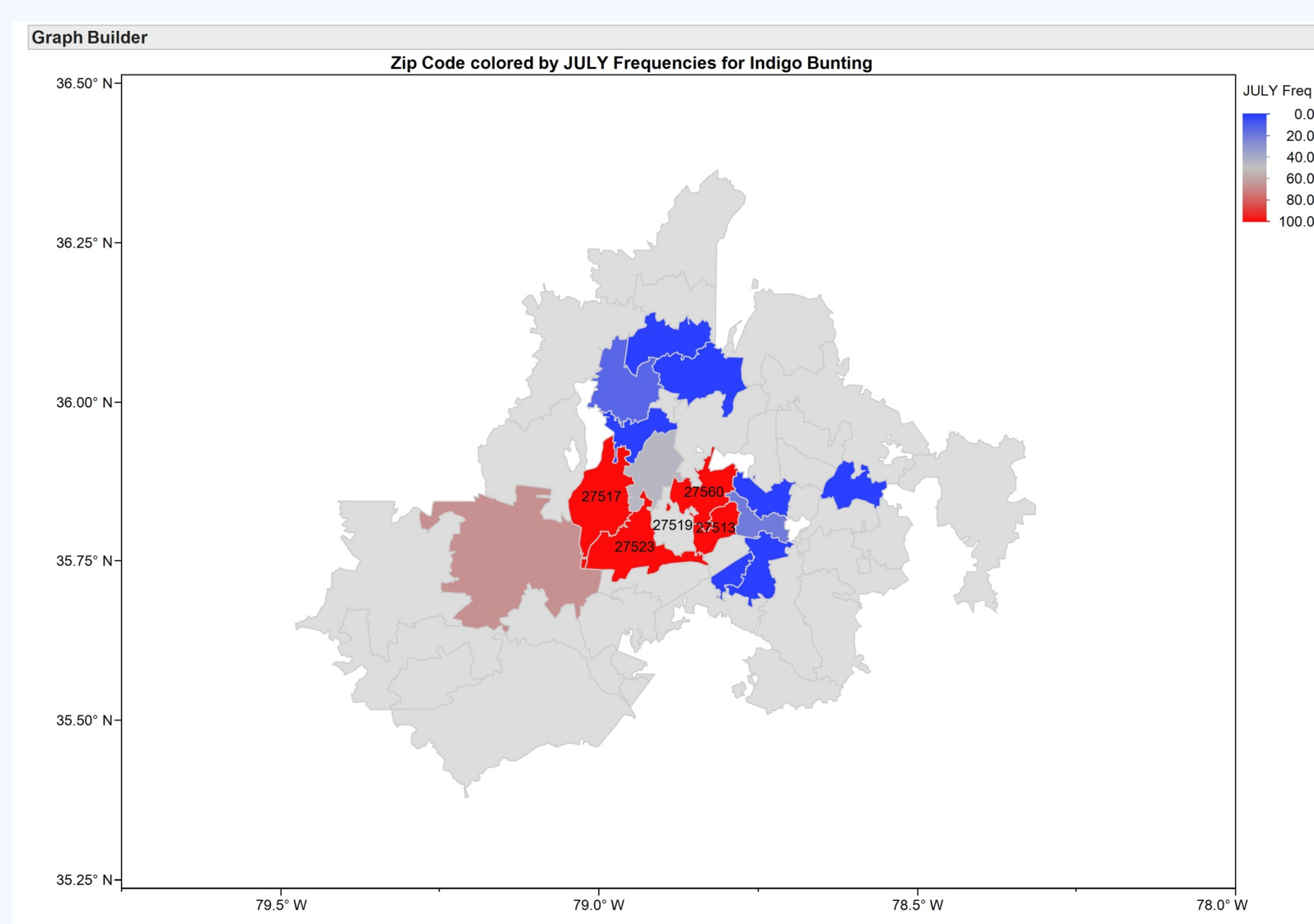


Figure 3. July, 2010, Frequency of checklists with Indigo Bunting (pictured in the upper right corner), centered on my zip code, 27519. 100% of the submitters for the red-colored zip codes saw or heard the bird. This indicates July is a good time to observe Indigo Buntings in my extended area.

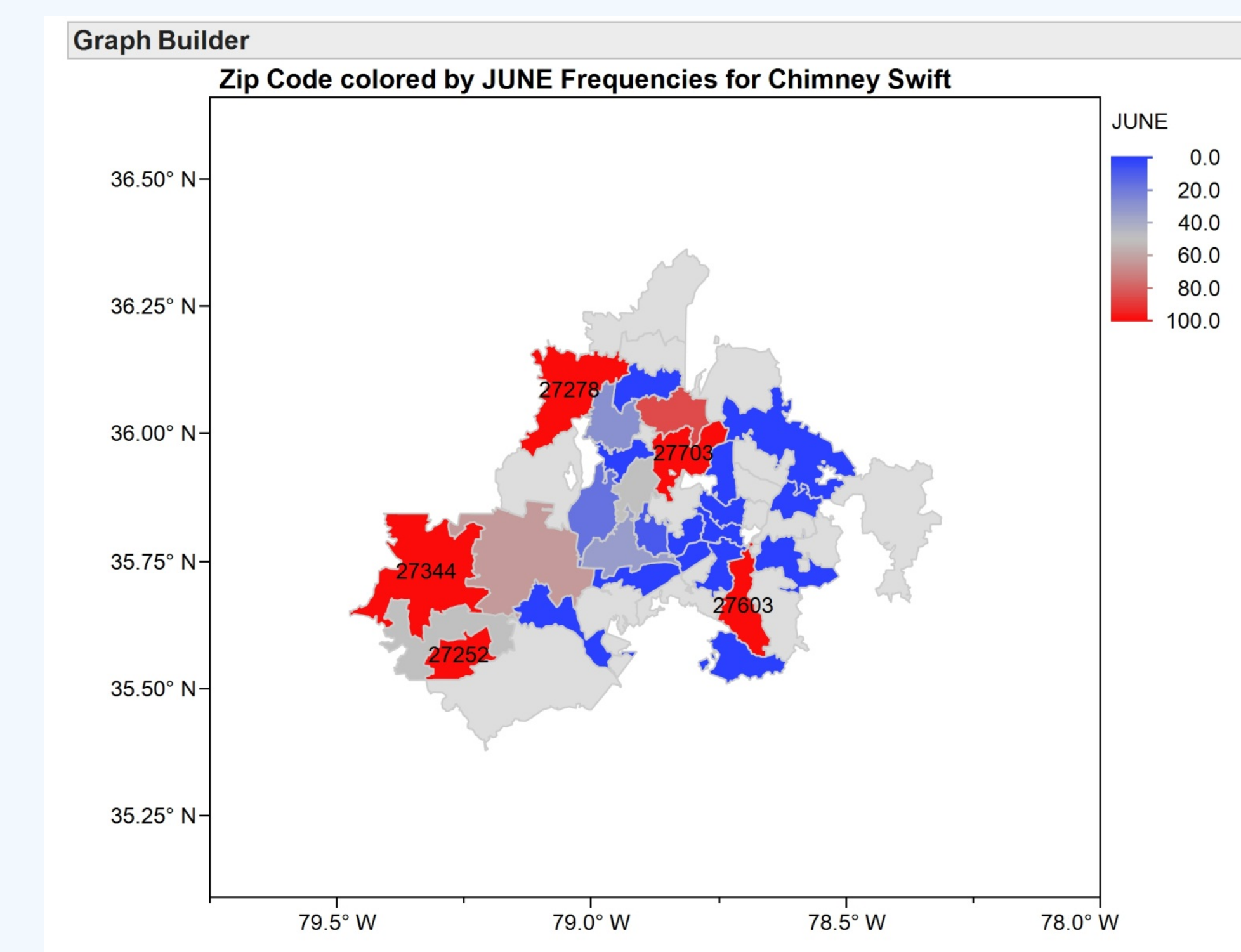


Figure 4. June, 2010, Frequency of checklists with Chimney Swifts near my zip code, 27519. eBirders in my immediate area did not observe that species as frequently as those farther away.

References

¹M. Arthur Munson, Kevin Webb, Daniel Sheldon, Daniel Fink, Wesley M. Hochachka, Marshall Iliff, Mirek Riedewald, Daria Sorokina, Brian Sullivan, Christopher Wood, and Steve Kelling. *The eBird Reference Dataset, Version 3.0*. Cornell Lab of Ornithology and National Audubon Society, Ithaca, NY, December 2011.

eBird, <http://eBird.org>

Avian Knowledge Network, <http://www.AvianKnowledge.net>

Acknowledgements

Cornell Lab of Ornithology and National Audubon Society, for making the data publicly available.

John Ponte, Principal Software Developer, JSL script for Lat/Lon to FIPS/zip code conversions.

Kyoko Keener, Development Tester, photo credit.