Analyzing Honey Bees in the United States

Charly Huang, Josephine Lyons, Pragati Pant, Meenu Ravi

Rensselaer Polytechnic Institute, Troy, NY, United States

Abstract

Honey bees are a significant contributors to ecology and agriculture within the United States. Since the early 2000s, the slogan “Save the Bees” has been spreading rapidly across America, and it has been evident that bees are dying at a swift pace. Many of these campaigns focus on harmful pesticides that kill much of the bee populations. There is also a huge astigmatism against honey as it is their main production, and bees can be harmed to produce honey. This project aims to determine the effectiveness of the various “Save the Bees” campaigns and how they correlate with, if at all, the population of bees in the United States, the use of bee-killing pesticides, and the sale of honey.

To make an informed analysis, multiple datasets have been derived from different sources. The first dataset looks at the number of bee colonies in different states, and was found on the website for the U.S. Department of Agriculture National Agricultural Statistics Service Quick Stats Dataset. The second dataset looks at the pounds of honey and the prices of honey for different states, and was retrieved from the website of the National Agricultural Statistics Service (NASS) of the U.S. Department of Agriculture, which is developed and supported by the Albert R. Mann Library at Cornell University. For the purposes of this project, only data provided for the years 1998 to 2017 is accounted for. Finally, the third dataset looks into the use of pesticides in the United States. Created by the Department of the Interior’s US Geological Surveyor Nancy T Baker, this dataset has data from 1992 to 2014 for pesticide use in the US watershed. Altogether, these datasets allow for enough analysis to provide conclusions to the given hypotheses.

This poster focuses on the various aspects of the project itself. We begin by stating our hypotheses about the given topics of bees population, honey production sales, and pesticide usage sales. We, then, highlight our process for analyzing the datasets for potential patterns as well as initial analysis. Furthermore, a deeper analysis is provided through given code and data visualization. Finally, our conclusion summarizes the findings of this project as well as whether our hypotheses were supported or not.

Questions and Hypotheses

The following states the two questions that drove this project.

1. **What is the relationship between the national honey production and the honey bee colony population in the United States between the years of 1998 and 2014 and how does this relationship affect the average honey production prices?** We hypothesize that the declining bee population has lowered honey production sales.

2. **How has the use of harmful bee killing pesticides correlated with the number of honey bee colonies in the United States between the years 1998 and 2014?** We hypothesize that the number of honey bee colonies will increase as the amount of harmful bee killing pesticides decreases and vice versa.

Analyzing Datasets and Deriving Conclusions

Workflow for data analysis

Hypothesize:
- Retrieved datasets from sources
- Stated project questions
- Formed our own hypotheses

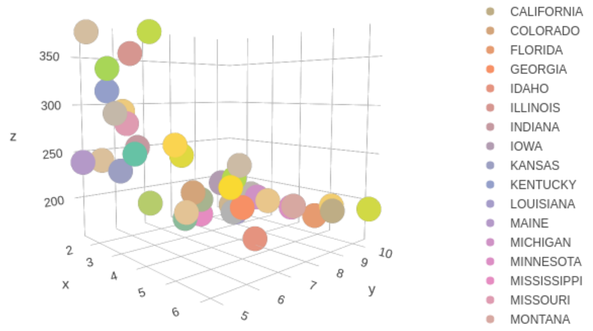
Analyze:
- Cleaned datasets
- Found initial data patterns
- Performed thorough data analysis

Visualize:
- Created data patterns graphs
- And somehin
- And something

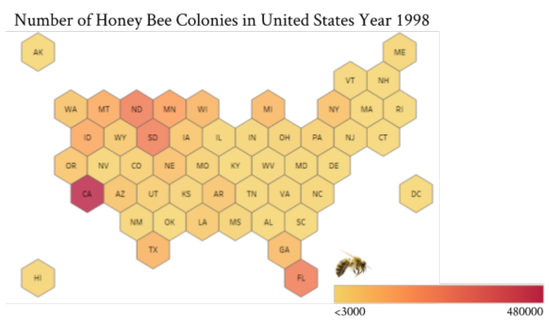
Conclude:
- Did somehing
- And somehin
- And something

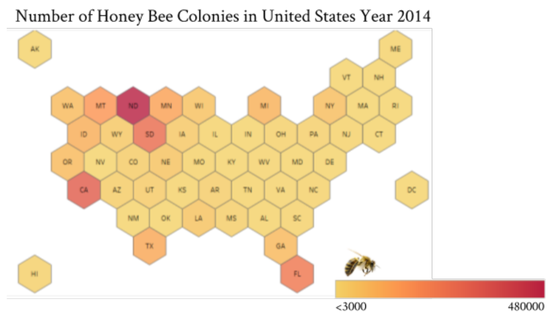
Visualize data

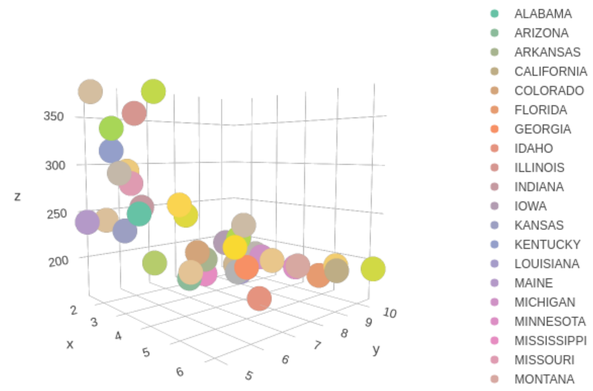
Several methods are used to inspect the dataset.

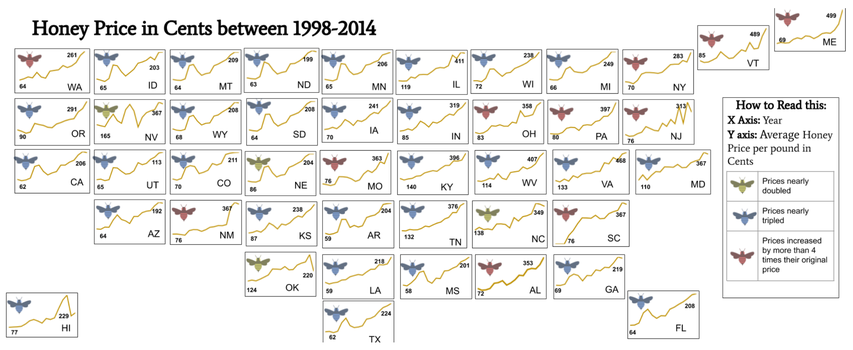


Analyze datasets









Conclusion

Need conclusion

Resources:

Beautifulsoup package in python: https://www.crummy.com/software/BeautifulSoup/
Python pandas visualization: https://pandas.pydata.org/pandas-docs/stable/visualization.html#visualization-hist
Python clean dataset: http://www.developintelligence.com/blog/2017/08/data-cleaning-pandas-python/
ScrapeR Package in R: https://cran.r-project.org/web/packages/scrapeR/scrapeR.pdf
Read excel files in R: https://www.datacamp.com/community/tutorials/r-tutorial-read-excel-into-r
R deal with missing data: https://www.statmethods.net/input/missingdata.html
R visualization: https://www.analyticsvidhya.com/blog/2015/07/guide-data-visualization-r/

Sponsors:
TetherlessWorld
IDEA
Rensselaer Institute for Data Exploration and Applications
DSRC