

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 correlation, if at all, between 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. Finally, the third dataset, created by the Department of the Interior’s US Geological Surveyor Nancy T Baker, looks into the use of pesticides in the United States.

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.

- 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.
- 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.

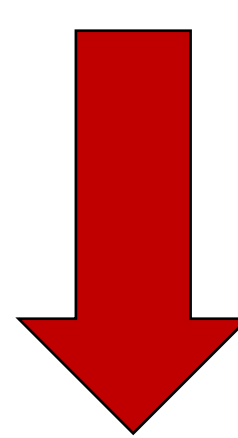
Resources:

Bee colony population dataset: <https://quickstats.nass.usda.gov/>
Honey production sales dataset: https://www.nass.usda.gov/About_NASS/index.php
Pesticide usage sales dataset: <https://catalog.data.gov/dataset/agricultural-pesticide-use-estimates-for-the-usgs-national-water-quality-network-1992-2014-vers>

Tableau: <https://www.tableau.com/>
plotly package in R: <https://plot.ly/r/>
ggplot2 package in R: <https://ggplot2.tidyverse.org/>
Pandas package in Python: <https://pandas.pydata.org/>
PyYAML package in Python: <https://pypi.org/project/PyYAML/>

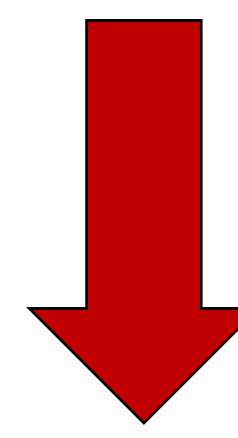
Analyzing Datasets and Deriving Conclusions

Workflow for Data Analysis



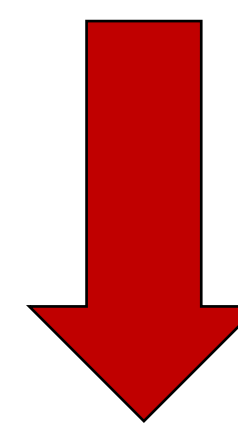
Hypothesize:

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



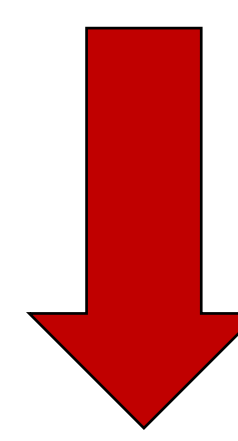
Analyze:

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



Visualize:

- Graphed initial data patterns
- Aggregated and graphed multiple datasets



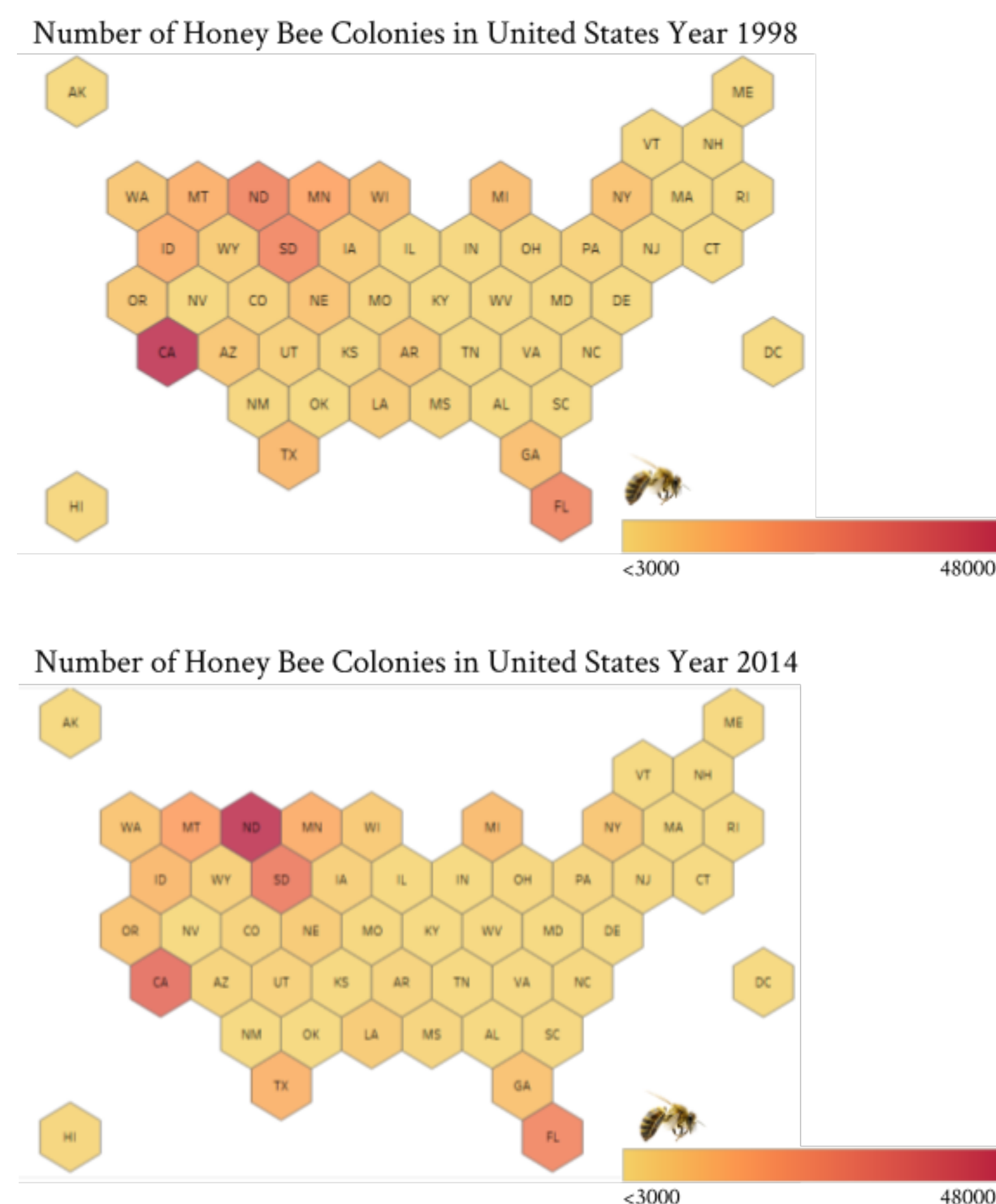
Conclude:

- Analyzed trendlines and r values
- Found data analysis to be inconclusive

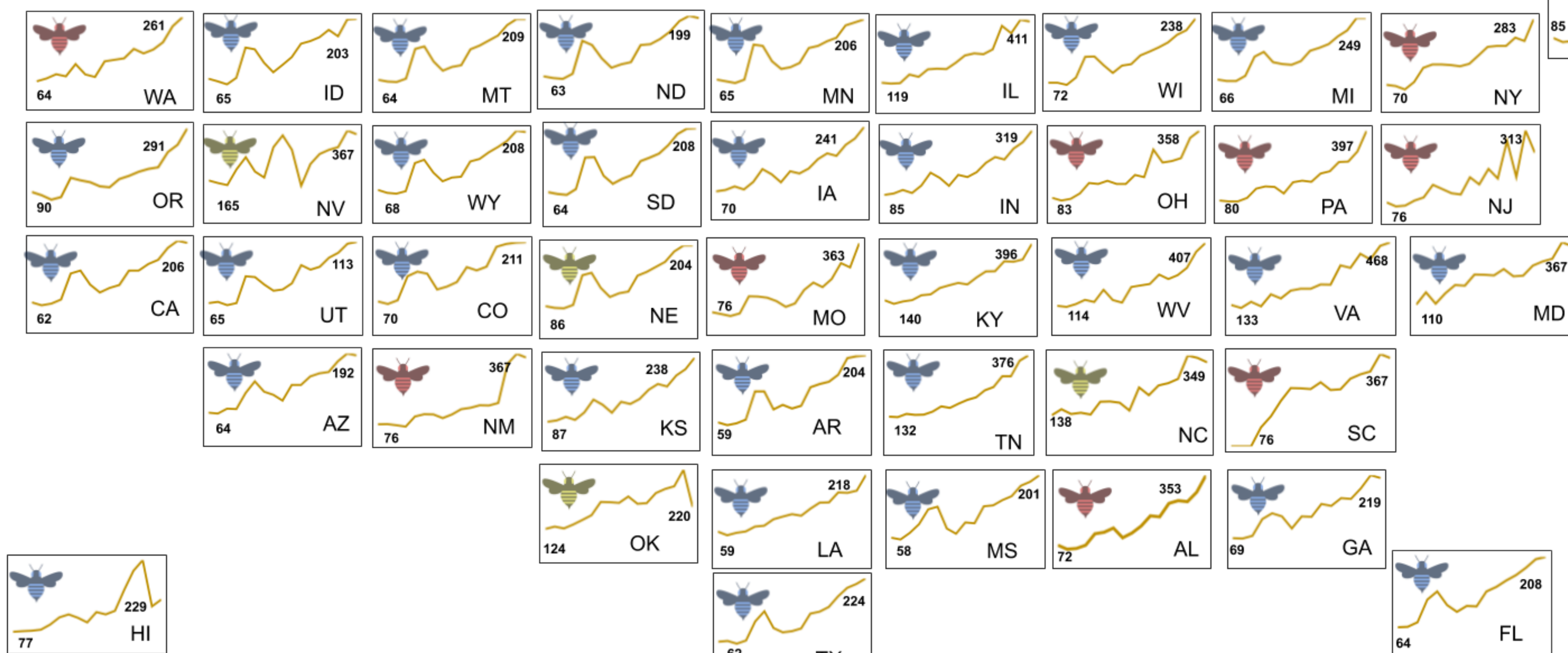
Exploratory Data Analysis

Initial data patterns within the individual datasets are plotted using Tableau and Google Sheets.

The following two graphs reflect the bee colony populations in both 1998 and 2014, which highlights the changes in populations between those years.



Honey Price in Cents between 1998-2014

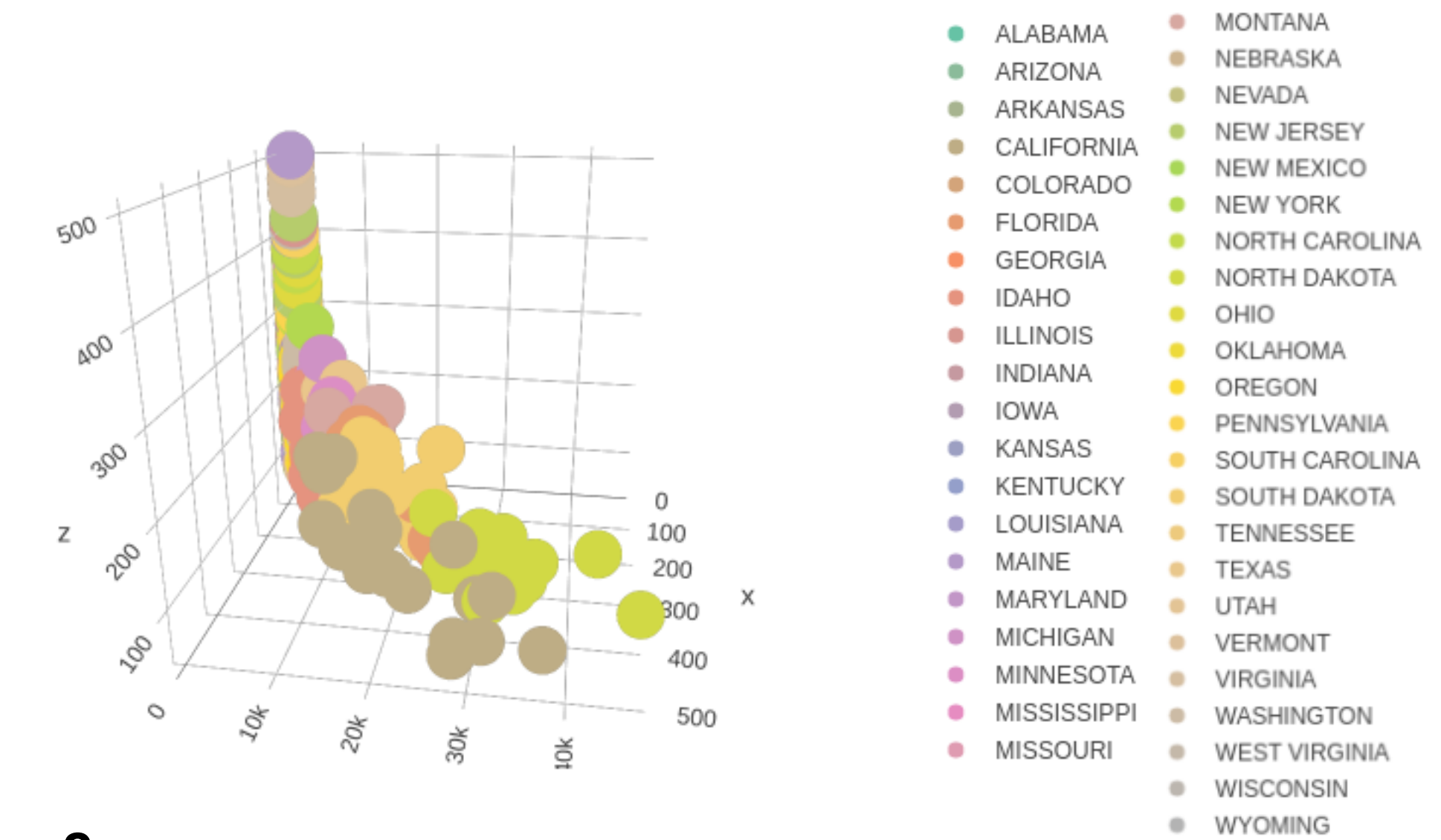
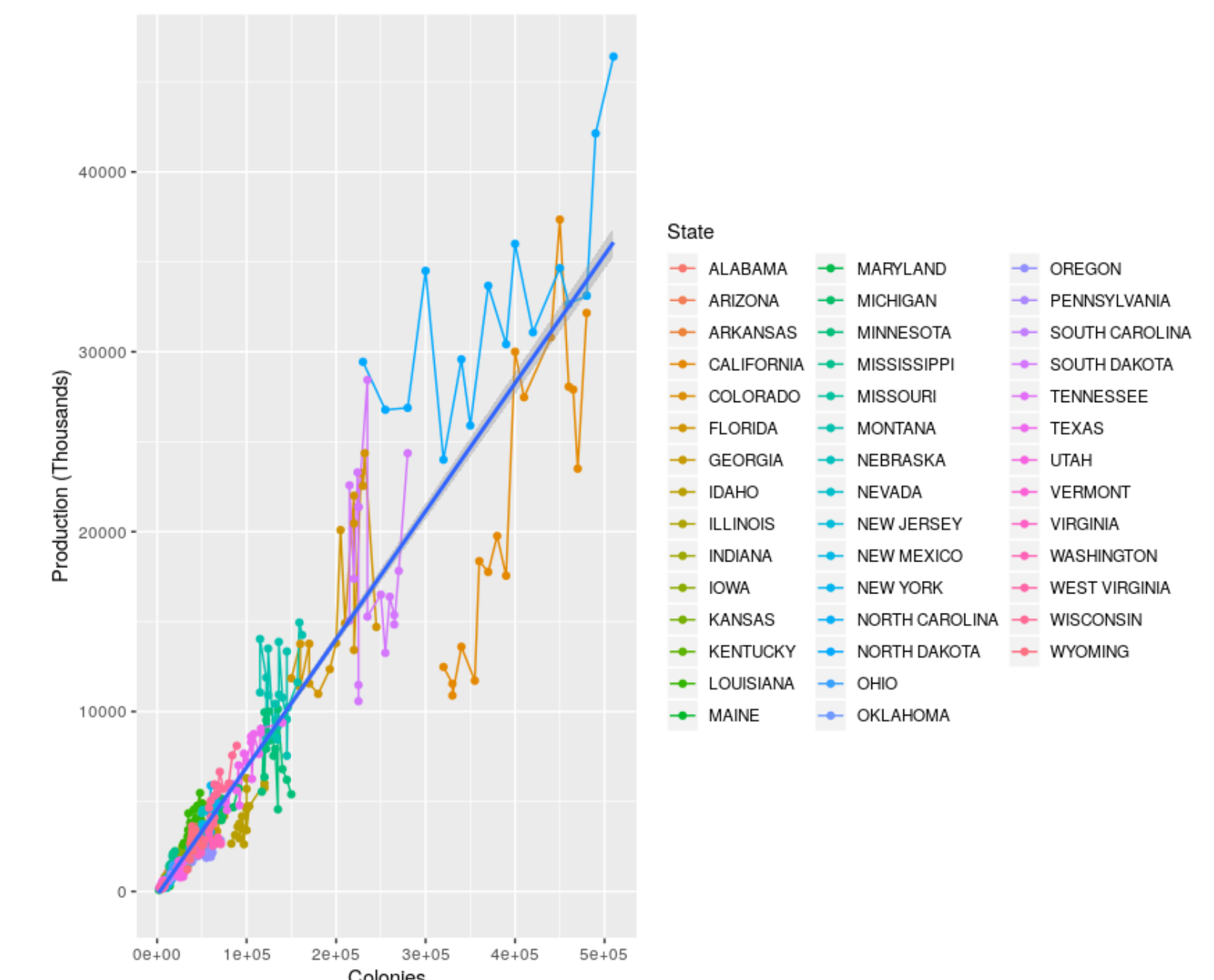


Model Application

Graphs were created in R to inspect and aggregate the datasets using ggplot2 and plotly.

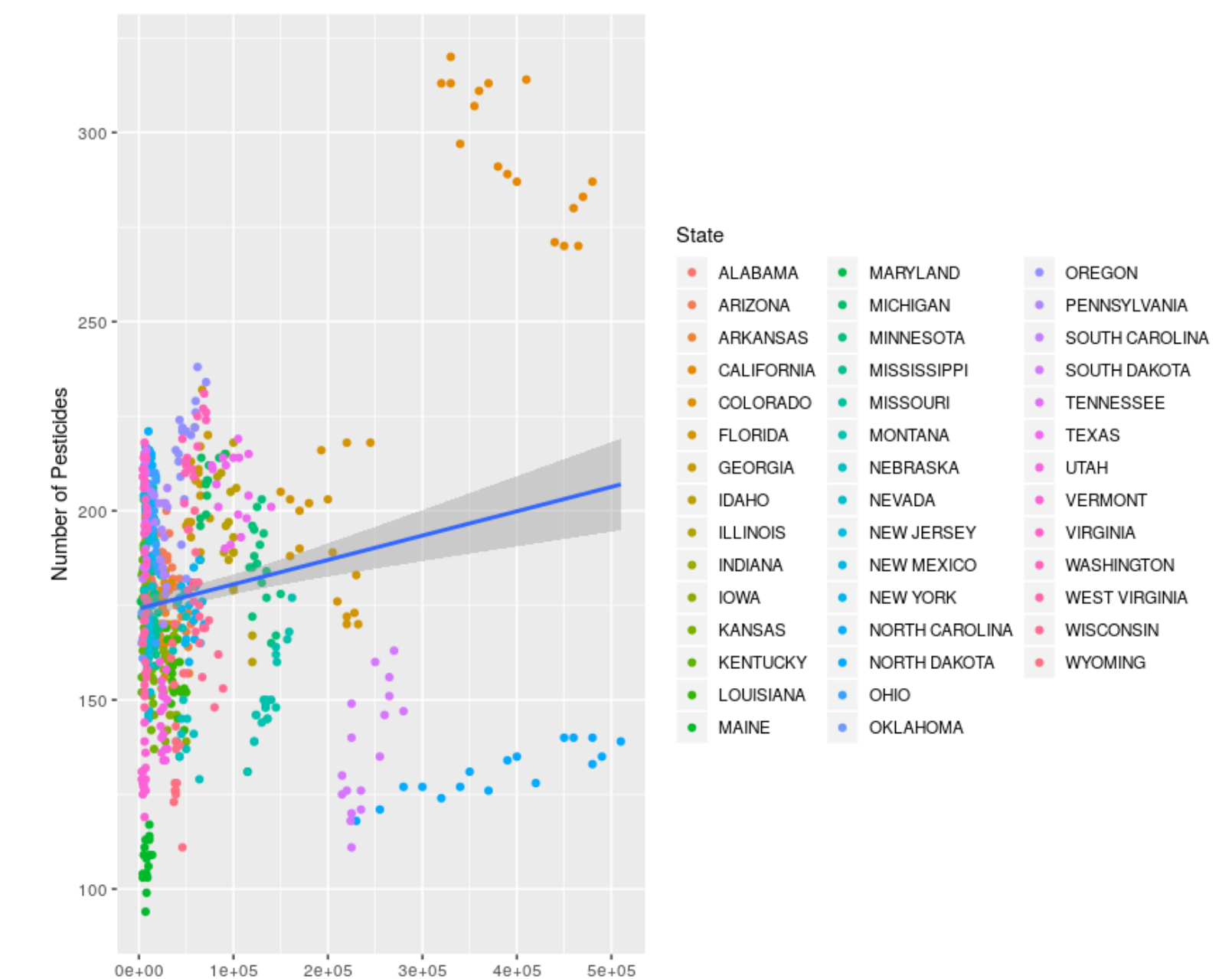
Question 1

The first graph demonstrates the correlation between honey production and bee colony population for all states from 1998 to 2014. However, the second graph shows that there is no correlation when honey price data is accounted for.

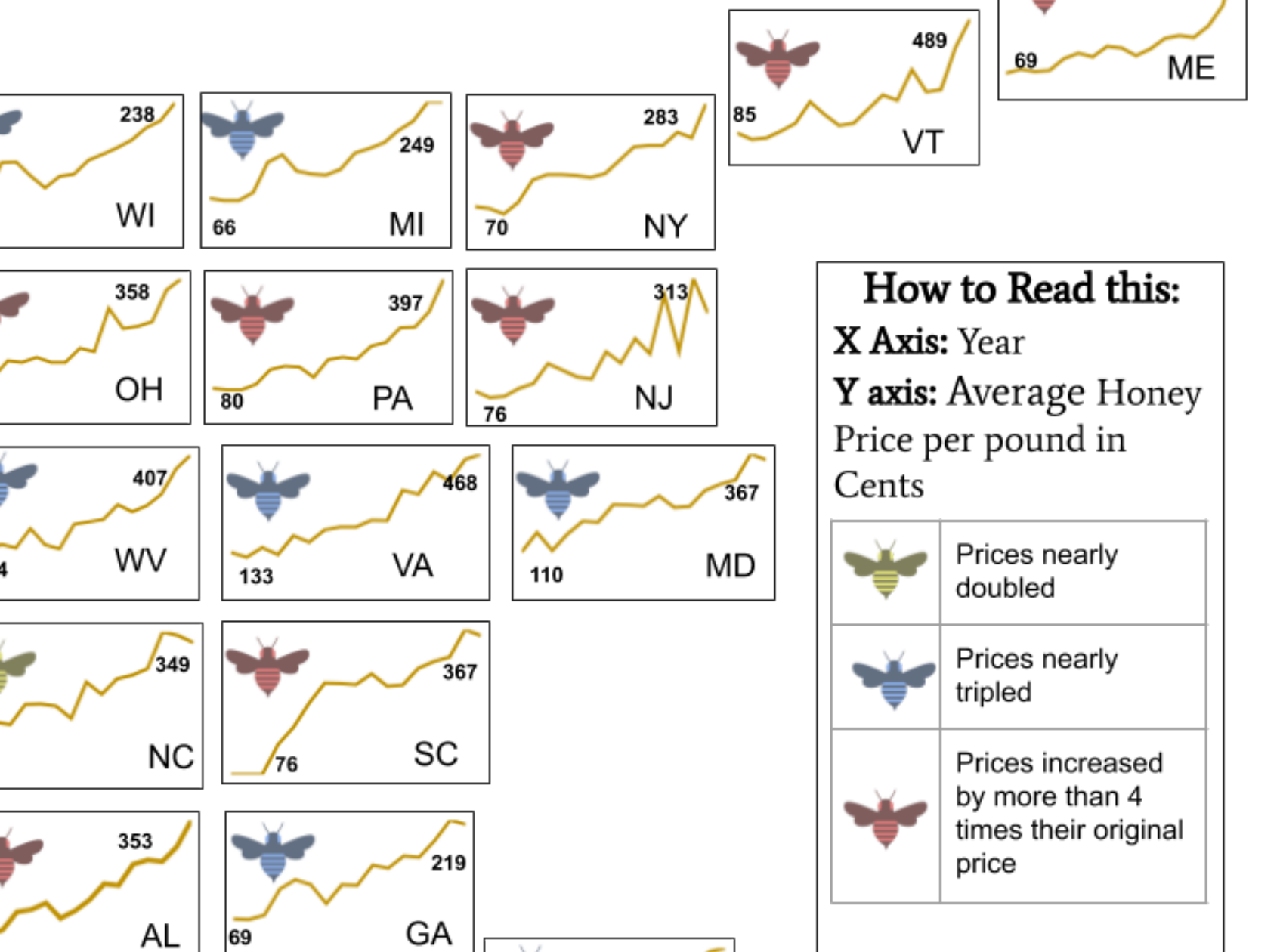


Question 2

The following graph shows that there is not a correlation between bee colony population and pesticide usage for all states from 1998 to 2014. Our model is the blue line.



The graph below demonstrates the change in the price of honey from 1998 to 2014 for each state.



Conclusion

We found the data analysis to be inconclusive. There is a relationship between bee colony population and honey production, however, there is no direct linear relationship with the price of honey. Additionally, the number of unique types of pesticides used within a given state does not correlate with the bee colony population within that state.