Honeybee Production and Colony Loss

Adam Vajdak, Lessa Fleming, Elizabeth Martin

Summary of Findings and Recommendations

Our primary finding from analyzing and combining 2 datasets that go over honeybee production and colony loss was that overall number of colonies, production of colonies, and honeybee keepers are all on a steady decline. Since these bees are an important part of the ecosystem, our recommendation would be to provide funding to incentivize beekeeping and other bee-saving initiatives in the interest of maintaining a healthy environment for the future. Improving the quantity and quality of beekeeping should have a positive effect on the number of colonies and the amount of honey production. Since there is a correlation between the amount of production and its price, this incentive would also have a benefit in the form of decreased costs for honey in the future.

Costs are up, colonies are down, save the bees.

Specification

Problem

This project goes over a few different problems related to honeybee honey production, the loss of colonies of honeybees, and the price of the resulting honey. Here are our main research questions:

- Is there a trend in colony loss over the years? Is there a loss in beekeepers over the years as well?
- How much honey is produced each year, and is there a loss of production over the years?
- What is the average trend in price over time, and do colony numbers affect that?
- Can we predict by state what the trend will be in the future for total honey production, number of colonies, and colony loss?

Hypothesis

Our hypotheses for our research questions are as follows:

- The overall trend in colony loss will have a positive slope (increasing the amount of loss over time). Beekeeper counts will vary with negative correlation to colony loss (as colony loss increases, we expect number of beekeepers to decrease).
- The overall trend of production of honey over the years will have a negative slope (decreasing the amount of honey produced over time).
- The overall trend in price will increase as production and colony count decrease (as supply of honey decreases, price will increase).
- Trends will be unique for different states, which have different cultures and climates that affect honeybees. We will be able to predict this with ARIMA time series analysis.

Data for Analysis

Our datasets are retrieved from Kaggle.

The first dataset, Bee Colony Loss, contains 365 rows of 9 columns. These columns are index, Year, Season, State, Total Annual Loss, Beekeepers, Beekeepers Exclusive to State, Colonies, and Colonies Exclusive to State. Index and Season are not valuable fields, as index is only a count from 0 to 364, and season is indicated as Annual across the board, with no variation in this field. State is a text field indicating which state the observation refers to, as a full-text version of the state name (with District of Columbia). Total annual loss is expressed as a percentage of bees lost during that year timeframe. Year in this dataset is expressed as the changeover from year A to year B (i.e. 2016/17). The years for this dataset range from 2010 to 2016. Beekeepers and Colonies are expressed as a total count of beekeepers and colonies respectively recorded for that state and year. Beekeepers exclusive to state and colonies exclusive to state are percentages of the total number of beekeepers or colonies which only count towards that individual state and do not cross state lines.

The second dataset, US Honey Dataset, contains 1115 rows of 9 columns. These columns are index, state, colonies_number, yield_per_colony, production, stocks, average_price, value_of_production, and year. Index is not a valuable field, as it is only a count from 0 to 1114. State in this dataset is listed as the full-text name of the state, without spaces in the case of 2 or more word states. The field colonies_number is the number of recorded colonies of bees recorded in the year period. The field yield_per_colony appears to be a calculated field representing the average production yield for each colony, dividing the production field by the colonies_number field. Production is the amount of honey produced by all honey-producing colonies in pounds. The field average_price refers to the average price per pound of honey in that state and year in USD. Year is the year in which the data is tabulated. The year range for this dataset goes from 1995 to 2021. Not all states have data for each year.

Observation

For average production of honey per year, there was a dramatic decrease in production across nearly all states between the years of 2009 and 2010. The rates have had dips and increases since then, but for the most part has stayed relatively even since then, but with a slight downward trend.

When looking at the data by state, there are several states that have slight increases, however the vast majority are seeing downward overall trends in honey production as well as beekeepers. The difference between production and colonies does have a regional distinction. There is significantly less honey production in the northeast part of the US versus the West, both of which have somewhat proportionally balanced production and colonies. Colonies in the South region of the US tend to have less production per colony, and those in the Midwest region have more production per colony.

Colony loss has been on an increase for the most part from 2011 to 2017, with a dip in loss in 2012 then resuming the increase after that. The average annual loss is roughly 40%, which is a very high amount of colony loss.

When looking at number of beekeepers by state, it can be observed that there are 2 states that stand out above the rest: Pennsylvania and Virginia. These states have an exceptionally high number of beekeepers, hovering around an average between 400 and 500, while most other states have averages between 100 and 200.

Analysis

Our primary goal in this analysis was to perform time-series analysis and fit an ARIMA model to our data to show how the trends in production, colony count, and price would be forecasted to fare based on historical data, however our attempts at these models did not yield good results. Our forecasting models were not stable and had a very wide range of error for the majority of states.

However, we were still able to see an overall trend with our analysis. We found that the current trend of colony loss is likely to continue, which will have a huge impact on honey production and costs. This does not account for the other side-effects of the declining bee population.

Recommendation

Since these bees are an important part of the ecosystem, our recommendation would be to provide funding to incentivize beekeeping and other bee-saving initiatives in the interest of maintaining a healthy environment for the future. Improving the quantity and quality of beekeeping should have a positive effect on the number of colonies and the amount of honey production. Since there is a correlation between the amount of production and its price, this incentive would also have a benefit in the form of decreased costs for honey in the future.

References

https://www.kaggle.com/datasets/mohitpoudel/us-honey-production-19952021

https://www.kaggle.com/datasets/thedevastator/beekeepers-in-the-united-states-report-colony-lo

Appendices

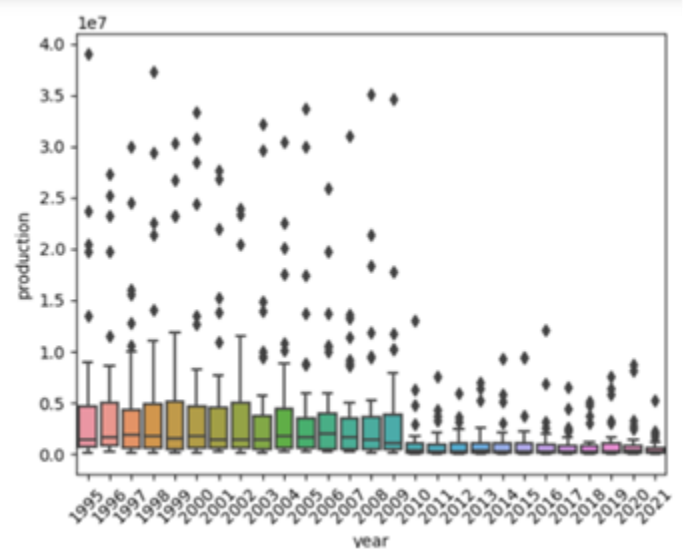

Figure 1. *Box and whisker chart showing the range of values across observations of each state's honey production from 1995 to 2021. This plot indicates a steady production from 1995 to 2009, but then indicates a dramatic drop-off in the year 2010.*

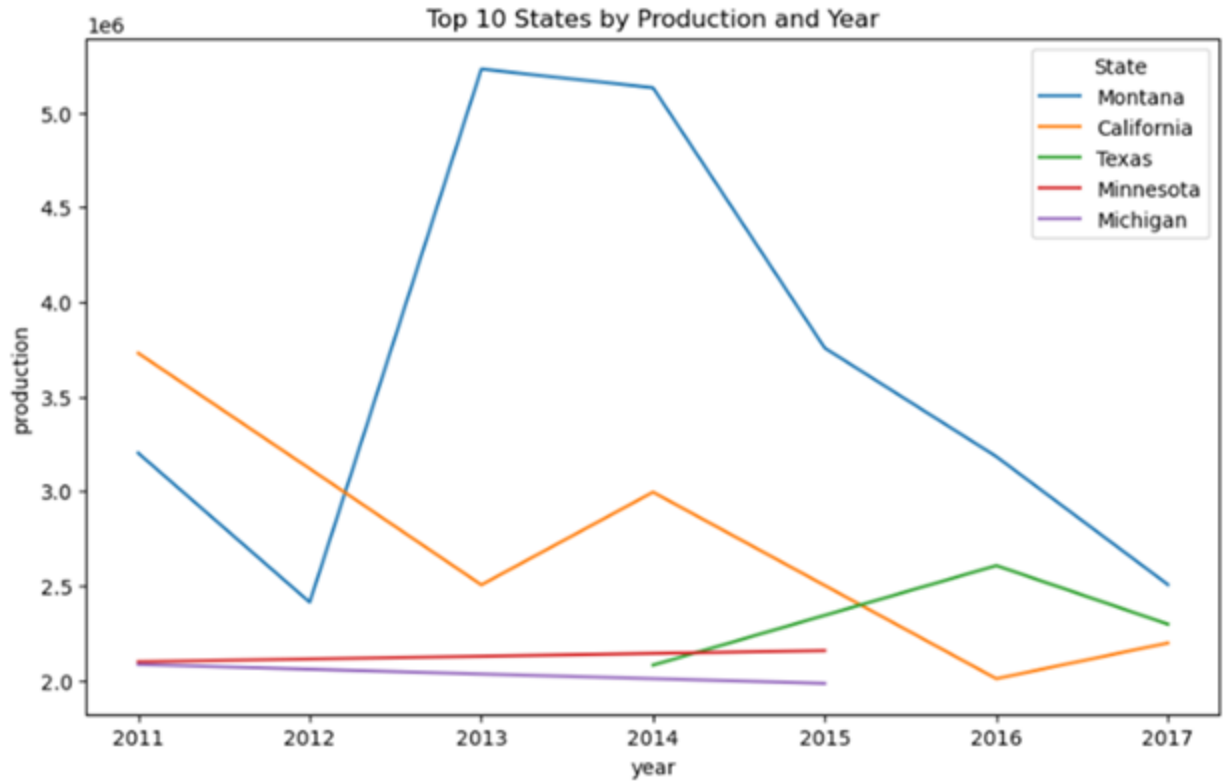

Figure 2. *Line chart showing production trends for the top 5 producing states. While there are some increases, the overall trend remains negative, and these increases do not show long-term stability.*

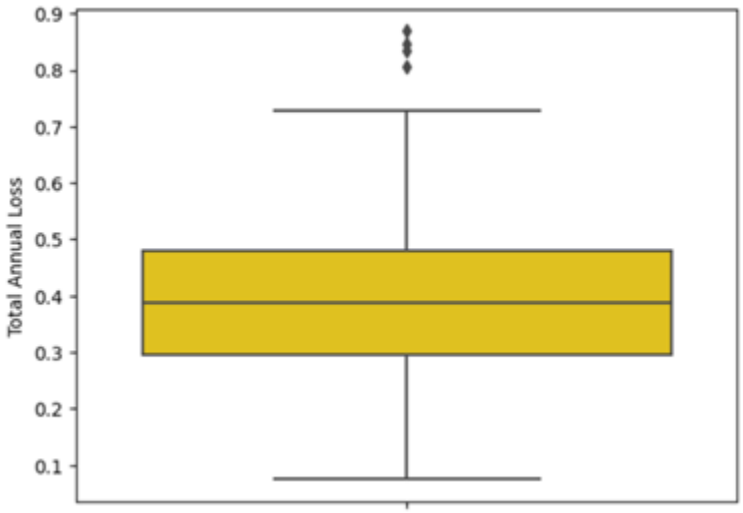

Figure 3. *Box plot of overall annual loss percentage. There is an average of 40% loss of colonies each year.*

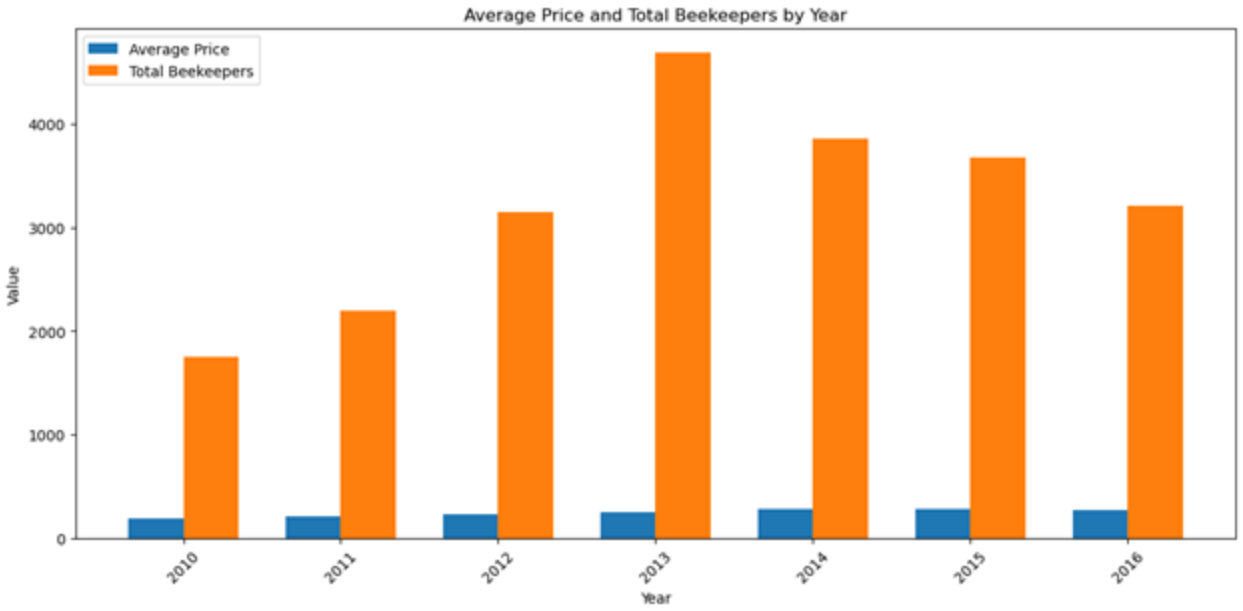

Figure 4. *Over time, we can see that there was a steady increase in beekeeping from 2010 to 2013, however that trend reversed and has since been on a decline.*

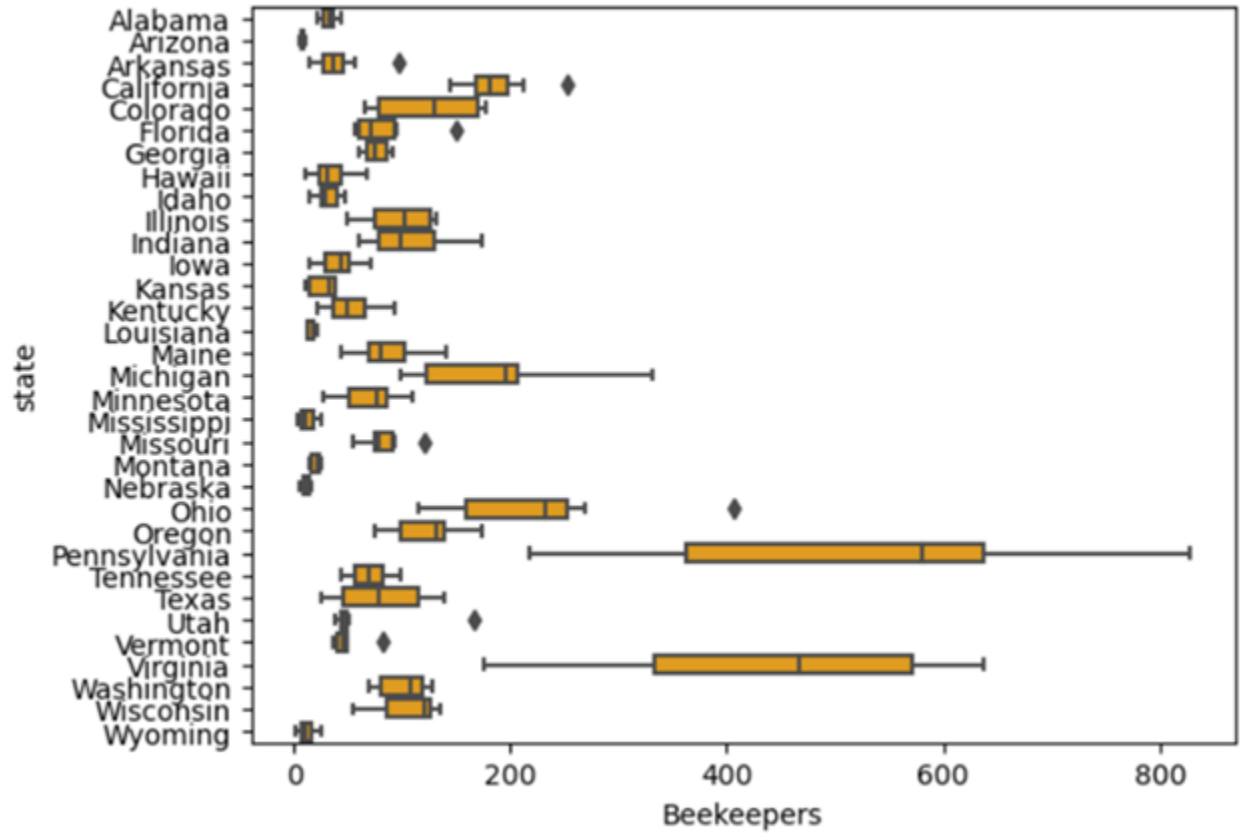

Figure 5. *Box plot showing range of number of beekeepers by state across the years with complete data (2011 through 2017), (n=7). The states with the widest range of beekeepers, as well as the highest*

median beekeepers are Virgina and Pennsylvania. Some states have a higher number of beekeepers, which is keeping the average higher than it otherwise would. Excluding the outlier states, the average number of beekeepers hovers between 100 and 200 recorded beekeepers for the state.

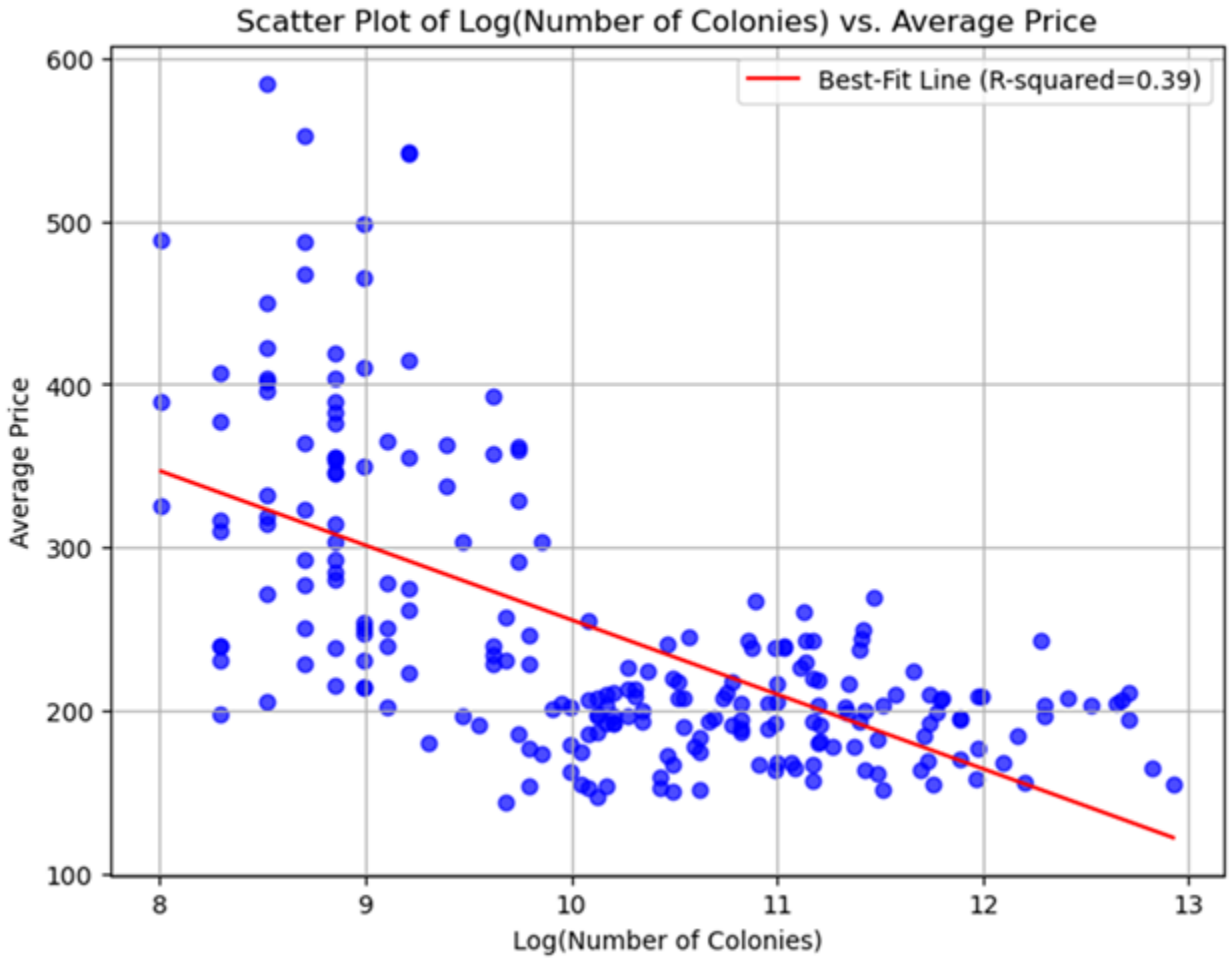

Figure 6. *Scatterplot showing relationship between average price of honey by pound and a log of the number of colonies (n=231). Simple linear regression was used to add a line of best fit to this data, showing a strong downward trend, as the log of the number of colonies of honeybees increases, the mean price of the honey decreases (R-squared = 0.39).*