Trends in the Production of Honey Across the United States of America Between 1998 and 2022

"Analysis of Honey Production Trends in the United States Using Data Visualization Techniques"

Jovanni Garcia
California State University Long Beach

Jeff Kim
California State University Long Beach

Johnerson Li
California State University Long Beach

Ren Yamasaki
California State University Long Beach

***Abstract*— For our project, we aim to find reasons and solutions to the decrease in honey production. We collected certain data on honey production such as total honey produced per year, honey produced within each state, total number of colonies, etc. Using the collected data, we plotted line, bar, box charts for the appropriate data to find any trends and correlations between the decline in honey production and other factors. Through the analysis of our data and graphs, some findings included a constantly high honey production rate of North Dakota, a constant number of honey producing colonies within a long term, and a decreasing number of yield per colony over the years. This results that one cause of the decreasing rate of honey production is not the number of honeybee colonies, but the decrease in honey yield per colony. Overall, the production of honey is decreasing by the years due to the decrease in yield per honeybee colony and is predicted to drop even lower at this rate.**

I. Introduction

A dataset on honey production using data sourced from the National Agricultural Statistics Service, USDA. This dataset can be found at the United States Department of Agriculture's publication releases and was created by the Economics, Statistics, and Market Information System and revised by Jovanni Garcia, Johnerson Li, Jeff Kim, and Ren Yamasaki.. This dataset was chosen due to its biological relevance to the current concerns regarding the international decline of insects with >40% of species being threatened with extinction (Sanchez-Bayo and Wyckhuys, 2018). Since 2006, commercial beekeepers in the United States have reported honey bee colony loss rates averaging 30 percent each winter — startling, when compared to historical loss rates of 10 to 15 percent (Gemma Bush, Ohio State University, 2020). For much of the past 10 years, beekeepers, primarily in the United States and Europe, have been reporting annual hive losses of 30 percent or higher, substantially more than is considered normal or sustainable. But this winter, many U.S. beekeepers experienced losses of 40 to 50 percent or more (Elizabeth Grossman, Yale, 2013). This declining trend caught the attention of many people such as beekeepers and scientists as a problem to one of the biggest ecosystems to nature. This project will study interactions between factors measuring successful honey production to better understand which areas may be contributing to the downfall of honeybees and possibly find solutions and strategies to mitigate the factors leading to these downfalls.

II. Methods

The necessary modules were imported into the PyCharm, followed by the data in .csv format. We opted for Python as our programming language for crafting visualizations. In our toolkit, we harnessed the power of various libraries: Pandas for structuring our data into dataframes, Plotly to craft visually engaging graphs, NumPy for mathematical operations, and Scikit-Learn for conducting linear regression analysis. We found data sets from the United States Department of Agriculture, known as USDA, showing an annual report of total honey production, average price, price by color class, yield per colony for each state (with the exception of the states that refused to reveal data on honey production and those that did not meet the minimum expectations for honey production) as shown in the picture below.

[Colonies which produced honey in more than one State were counted in each State]

State	Honey producing colonies [1]	Yield per colony	Production	Stocks December 15 [2]	Average price per pound [3]	Value of production [4]
	(1,000)	(pounds)	(1,000 pounds)	(1,000 pounds)	(dollars)	(1,000 dollars)
Alabama	7	39	273	66	5.52	1,507
Arizona	25	36	900	360	2.38	2,142
Arkansas	20	49	980	176	1.85	1,813
California	320	43	13,760	2,752	1.95	26,832
Colorado	30	41	1,230	455	2.39	2,940
Florida	192	46	8,832	883	2.24	19,784
Georgia	101	34	3,434	412	2.52	8,654
Hawaii	15	105	1,575	79	1.70	2,678
Idaho	107	35	3,745	637	1.76	6,591
Illinois	10	52	520	156	5.65	2,938
Indiana	9	55	495	149	3.91	1,935
Iowa	35	58	2,030	1,259	2.54	5,156
Kansas	8	62	496	164	3.25	1,612
Kentucky	7	33	231	58	5.76	1,331
Louisiana	33	69	2,277	228	2.46	5,601
Maine	10	30	300	78	3.36	1,008
Michigan	95	47	4,465	1,384	2.95	13,172
Minnesota	108	55	5,940	2,495	1.80	10,692
Mississippi	25	73	1,825	146	1.99	3,632
Missouri	9	41	369	100	3.59	1,325
Montana	110	81	8,910	3,208	1.61	14,345
Nebraska	37	52	1,924	250	1.69	3,252
New Jersey	14	31	434	91	4.32	1,875
New York	58	56	3,248	844	3.13	10,166
North Carolina	12	38	456	123	5.50	2,508
North Dakota	495	78	38,610	8,108	1.59	61,390
Ohio	16	75	1,200	576	3.50	4,200
Oregon	95	29	2,755	1,102	2.40	6,612
Pennsylvania	19	48	912	392	4.12	3,757
South Carolina	16	46	736	66	3.44	2,532
South Dakota	245	61	14,945	8,668	1.77	26,453
Tennessee	7	51	357	54	4.23	1,510
Texas	157	57	8,949	1,253	2.00	17,898
Utah	28	34	952	171	2.02	1,923
Vermont	6	47	282	96	3.94	1,111
Virginia	5	40	200	54	6.03	1,206
Washington	98	37	3,626	798	2.51	9,101
West Virginia	6	46	276	58	3.81	1,052
Wisconsin	45	50	2,250	855	3.11	6,998
Wyoming	38	40	1,520	608	1.71	2,599
Other States [5 6]	33	42	1,375	303	4.68	6,435
United States [6 7]	2,706	54.5	147,594	39,715	2.10	309,947

[1] Honey producing colonies are the maximum number of colonies from which honey was harvested during the year. It is possible to harvest honey from colonies which did not survive the entire year.

Table A

Some of the data above could show a minimal inaccuracy due to the fact that the most of the data is rounded by thousands, but would not show any problems in finding trends and making comparisons. We first examined the dataset to uncover the relationships between various factors. Subsequently, we performed data cleaning and null value checks. Our analysis encompassed data from 1998 to 2022, during which we transformed the information into CSV files, to enable the creation of the graphs depicted in the image below.

```
"state","numcol","yieldpercol","totalprod","stocks","priceperlb","prodvalue","year"
"AL",7000,52,364000,55000,2.85,1037000,2013
"AZ",29000,36,1044000,251000,1.96,2046000,2013
"AR",22000,60,1320000,66000,2.02,2666000,2013
"CA",330000,33,10890000,2505000,2.11,22978000,2013
"CO",26000,43,1118000,324000,2.10,2348000,2013
"FL",220000,61,13420000,1074000,2.03,27243000,2013
"GA",67000,50,3350000,637000,2.26,7571000,2013
"HI",13000,83,1079000,65000,1.97,2126000,2013
"ID",83000,32,2656000,1036000,2.02,5365000,2013
"IL",7000,48,336000,101000,4.19,1408000,2013
"IN",6000,47,282000,82000,2.77,781000,2013
"IA",39000,48,1872000,1217000,2.45,4586000,2013
"KS",6000,46,276000,39000,2.50,690000,2013
"KY",3000,41,123000,17000,3.25,400000,2013
"LA",50000,98,4900000,490000,1.89,9261000,2013
"ME",7000,43,301000,27000,3.14,945000,2013
"MI",85000,55,4675000,982000,2.16,10098000,2013
"MN",130000,58,7540000,1282000,1.99,15005000,2013
"MS",17000,116,1972000,39000,1.86,3668000,2013
"MO",10000,47,470000,85000,2.62,1231000,2013
"MT",159000,94,14946000231000,2.09,31237000,2013
"NE",46000,60,2760000,1628000,2.07,5713000,2013
"NJ",11000,44,484000,34000,4.19,2028000,2013
"NY",55000,48,2640000,1030000,2.12,5597000,2013
"NC",10000,38,380000,84000,3.67,1395000,2013
"ND",480000,69,33120000,6955000,2.04,67565000,2013
"OH",17000,45,765000,390000,3.29,2517000,2013
"OR",62000,35,2170000,456000,2.39,5186000,2013
"PA",13000,45,585000,257000,3.03,1773000,2013
"SD",265000,56,14840,6381000,2.07,30719000,2013
"TN",7000,45,315000,63000,3.55,1118000,2013
"TX",106000,59,6254000,1689000,2.10,13133000,2013
"UT",30000,34,1020000,92000,2.09,2132000,2013
"VT",3000,51,153000,46000,3.89,595000,2013
"VA",5000,35,175000,42000,4.50,788000,2013
"WA",69000,39,2691000,1023000,2.30,6189000,2013
```

CSV file made by the used data set

Example of csv file of honey production in 2013.

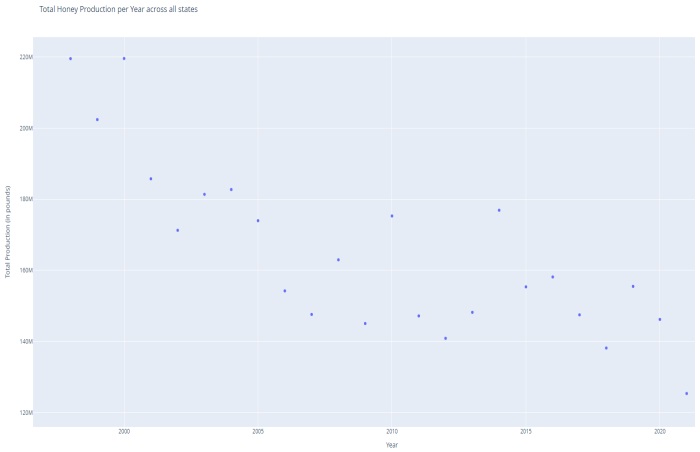
During this process, we discovered that some of the states were not consistent on the chart. There were states that showed up one year, but was not on the other years due to avoiding disclosing data. The csv contained variables such as Numcol, yieldPerCol, totalProd, stocks, pricePerLb, prodValue, year.Numcol (Number of honey producing colonies) represents the measurement of maximum number of bee colonies that were used for honey production during the year, and this includes colonies from which honey was harvested, even if they did not survive the entire year. yieldPerCol represents the average amount of honey, in pounds, produced by each colony. totalProd represents total production in pounds and it is calculated as the number of colonies multiplied by yieldPercol. Stocks represent the quantity of honey in pounds that is held in the inventory by producers. PricepPerLb represents the average market price per pound of honey and it is calculated based on expanded sales. This data is important because it reflects the market value of honey. ProdValue (Value of production) represents the total monetary value of honey production, calculated as totalProd multiplied by pricePerLb. This data shows the total revenue generated from honey production. Year simply shows which year the data set was extracted from.

```
"""
    Calculating the average (mean) total honey production per year across all states
"""
# average production of honey per year using .groupby() method provided by pandas to get the mean per year
prod_per_year = df.groupby('year').totalprod.mean().reset_index()

# columnOfInterest = df['columnName'] then reshape it to get it into the right format
X = prod_per_year['year']
X = X.values.reshape(-1, 1)

# total product per year
y = prod_per_year['totalprod']
"""
    ScatterPlot Using Plotly
    plot y vs X as a scatterplot.
"""
# Create a scatter plot
scatter = go.Scatter(x=X.ravel(), y=y, mode='markers', name='Total Production')

# Create a figure
fig = go.Figure(data=scatter)

# Update layout of the figure
fig.update_layout(
    title='Average Honey Production per Year across all states',
    #title='Total Honey Production per Year across all states',
    xaxis_title='Year',
    yaxis_title='Average Production (in pounds)',
    #yaxis_title='Average Production (in pounds)',
    hovermode='closest'
)

# Show the figure
#fig.show()
```

Code Snippet Used for Graphing

The sample snippet code shown above exemplifies how we used the csv files we created and implemented a visualization of them using Python. In the example above, the scatterplot featured below is created with the x-axis being “Honey Production Per Year” and the y-axis being “Total Product Per Year”. The code is then used for other data sets with the appropriate type of graph to be used.

Trends in production were visualized by grouping the data by XX and plotting a line graph. The dataset was then ranked by production to compare data points on a linear scale. This was done by grouping, transforming, and sorting the data. Changes in honey production ranking were quantified by grouping the dataset by state and aggregating to get mean, min, and max ranking values over time. The total and average honey production was displayed in a scatterplot.



Total Honey Production per Year.

After creating the scatterplot, we made a line of best fit indicating the average on the next scatter plot. The regression line shows the overall decrease in total honey production over the years.

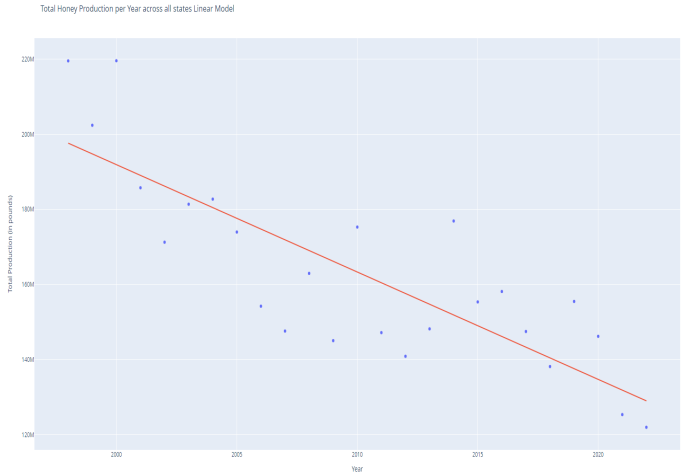
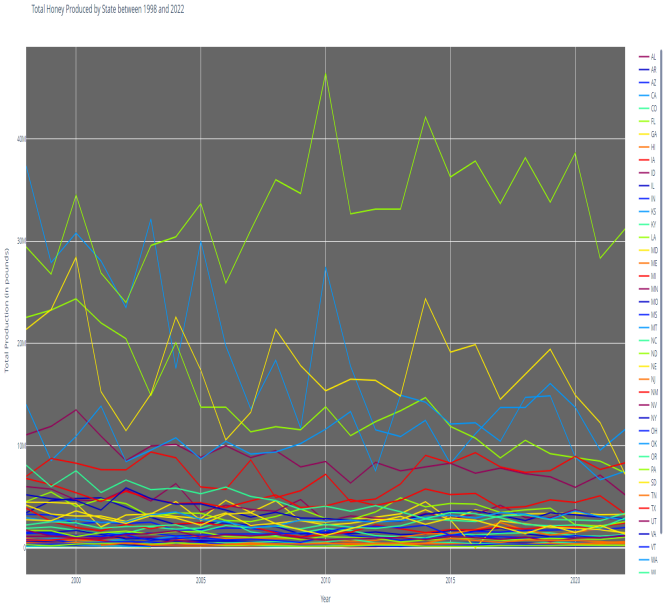


Table showing annual report of honey production.

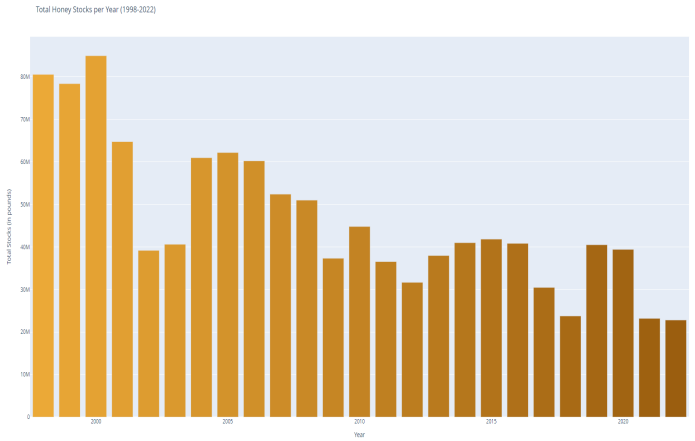
III. Results

Through analysis of the various graphs made by the data set, there were some discoveries made about the honey production within the United States. We found that honey production could vary greatly from state to state. The graph shows there are some states that have a much greater honey production rate including North Dakota consistently at the top, followed by California, South Dakota, Florida, Montana, Minnesota and Texas. North Dakota had a climate and ecosystem that was well-suited to honey bees and thus, is where a great majority of where the nation's honey production comes from. Colder climates make beekeeping a much more difficult endeavor, but it is not impossible. States like Virginia and West Virginia, though not known for their honey production, still house around 5-6,000 honey-producing bee colonies each. As evidenced in the graphic below, the lime green line that hovers way above all the other lines is North Dakota. Some states have had their honey production decline much more drastically than others, but North Dakota continues to lead the states in honey production. Other factors for increase in honey production include abundant flowering plants and diverse landscapes. These factors along with good climate conditions allow states such as North Dakota, South Dakota, California, and Florida to stand above the other states in greater production rate of high quality honey.

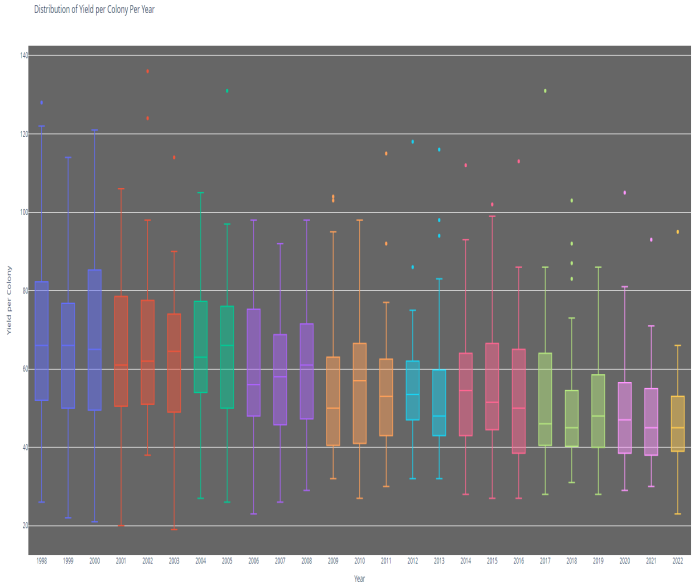


Total Number of Honey Produced by Each State

Our approach for analyzing the total honey production within each state between 1998 and 2022 has given us a closer look to factors of greater quantity and higher quality honey production. States like North Dakota and South Dakota with favorable climate conditions and landscapes would benefit the rate of honey production, while states like California, Montana, and Texas keep a high honey production rate compared to others due to their agriculture and vast variety of nectar sources.

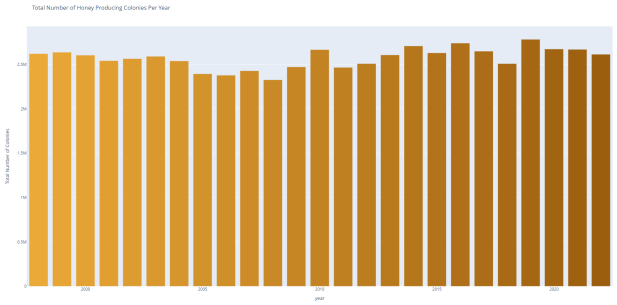


Total Stocks of Honey per Year

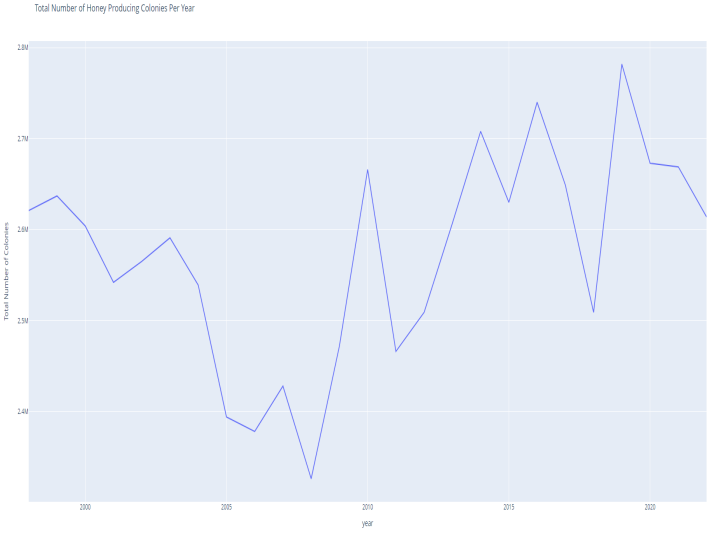


Number of yield per colony

Finally, the effect of the colony number was studied... The annual decline in yield per colony was plotted with a box plot. In a box plot, also known as a box-and-whisker plot, the "box" represents 50% of the data. The specifics of what is included in this box are as follows: Lower Quartile (Q1), Upper Quartile (Q3), Interquartile Range (IQR), Median, Whiskers, and Outliers. On average, with respect, its interquartiles, we found that the distribution of yield per colony per year from the years 1998 to 2022, had gradually declined through the years, as well as total honey stocks. However, we also observed that the total number of honey producing colonies per year stayed stable and had even increased. A steady rate of total colonies for over a time period of 24 years shows that the number of colonies was not a factor for the decrease in honey production in the United States.



Total number of honeybee colonies



The observations made from the box plots – shorter whiskers, reduced size of the box quartiles, and a falling median – indicate a decreasing yield per colony over the years. Removing the factor that perhaps a decline in the yield was due to factors that reduced the number of colonies, we can conclude that perhaps other factors were to be observed pertaining to the decline of honey production yield over the past few decades. This trend suggests a decline in honey production efficiency or effectiveness per colony.

IV. Problems of Dataset

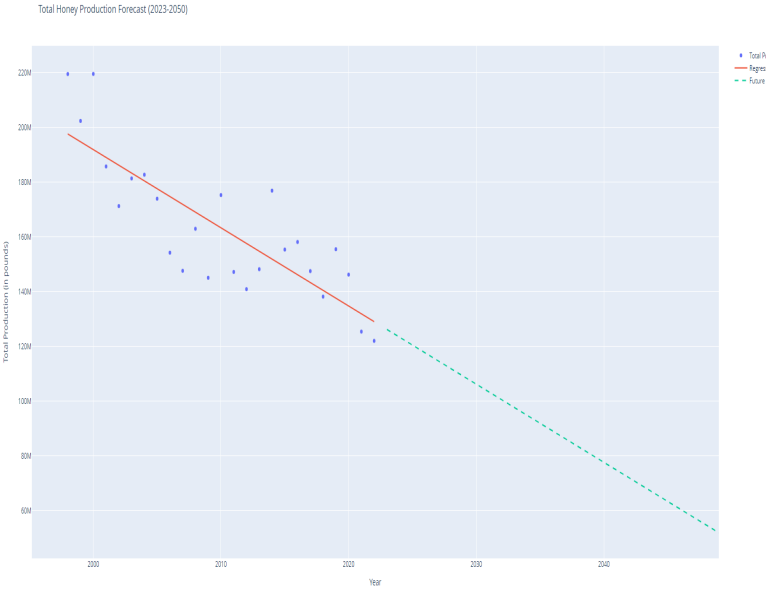
We were able to observe initially that not all 50 states of the United States were individually observed at every single calendar year that honey production statistics were recorded in this dataset. Some states either had laws that prevented them from publicly showing their production statistics, while others may not have had honey that met the standards for this dataset. Some states such as Hawaii are included in some years, and not in others. This inconsistency of data availability made it less accurate to have general honey production criteria since

data was sometimes included and then excluded without warning.

Another problem we encountered while working with the dataset was not with the dataset itself, but rather the limitations of our research due to not isolating a specific factor for honey production decline. The dataset only sought to give an objective view of the numbers of colonies and honey stock across the United States. With further inquiry, we believe we can try to isolate the main problem that is affecting honey bee colony efficiency in order to figure out conventional and effective solutions to the decreased honey production.

V. Conclusion

The overall production of honey in the US is drastically decreasing. Individual bees/colonies are becoming less efficient, with reduced yield. Thus, something may be detrimentally impacting the health of honeybees in America thereby reducing their honey production. More information is needed on factors affecting honeybee health and honey production efficiency before a comprehensive conclusion can be reached, for example, information on Environmental Changes: Changes in climate, habitat loss, or increased use of pesticides can adversely affect bee health and productivity. Bee Health Issues: Issues such as diseases, parasites (like Varroa mites), and colony collapse disorder can significantly reduce colony productivity.Agricultural Practices: The availability and variety of flowering plants, due to shifts in agricultural practices, can affect honey yields. Policy and regulatory changes: changes in agricultural or environmental policies might impact beekeeping practices such as regulations that control what pesticides can and cannot be used, as well as funding towards beekeepers. Change in market demand and consumer preferences, such as shifts in consumer preferences and market demand for honey can influence production practices as well, further impacting funding for research and welfare of bee colonies. Establishing the underlying problem will give us something to target and hopefully begin to stop and potentially reverse the harm to the honeybee ecosystem and may even give us insight into the larger issue of insect death and extinction. In the projection graph below, based on current trends, it is projected by 2030 and beyond that the total honey production forecast is expected to fall below 50 million pounds by 2050.



Forecast of Total Honey Produced from 2030-2050

Acknowledgment

We would like to thank the honeybees of America for their untiring, relentless work for the global ecosystem, for which without their work, this paper would not be possible. We would also like to thank the creator and accumulator of the data source, for making it possible for us to conduct our research.

References

[1] https://www.kaggle.com/datasets/jessicali9530/honey-production

[2] https://usda.library.cornell.edu/concern/publications/hd76s004z?locale=en&page=2#release-items