Analytics Report: Honey Production In USA(1998-2012)

Name: Akshay Thorave Roll No: MDS202247

2022-10-30

INTRODUCTION

In 2006, global concern was raised over the rapid decline in the honeybee population, an integral component of American honey agriculture. Large number of hives were lost to Colony Collapse disorder, a phenomenon of disappearing worker bees causing the remaining hive colony to collapse. Later, some industries observed recovery, but the American honey industry is still struggling. The U.S. used to locally produce over half the honey it consumes yearly. Now, honey mostly comes from overseas.This dataset provides insight into honey production in America by state from 1998 to 2012. We will try to visualize this data using different packages in R like 'ggplot' and 'ggplot2'.

ABOUT THE DATA SET

The dataset gives information about 'Honey Production' in the states of America from 1998 to 2012. In this dataset, there are a total of 8 variables out of which 7 are 'Numeric' variables and 1 'Categorical' variable:

WORKING VARIABLES

VariableName	VariableType	Description
state	Categoric(Nominal)	State In USA
numcol	Numeric(Discrete)	No. of honey producing colonies
yieldpercol	Numeric(Discrete)	Honey Yield Per Colony(In Pounds)
totalprod	Numeric(Discrete)	Total Honey Production(numcol * yieldpercol)(In Pounds)
stocks	Numeric(Discrete)	Stocks held by producers(In Pounds)
priceperlb	Numeric(Continuous)	Average price per Pound(In Dollars)
prodvalue	Numeric(Continuous)	Value of Production(totalprod * priceperlb)(In Dollars)
year	Numeric(Discrete)	Year

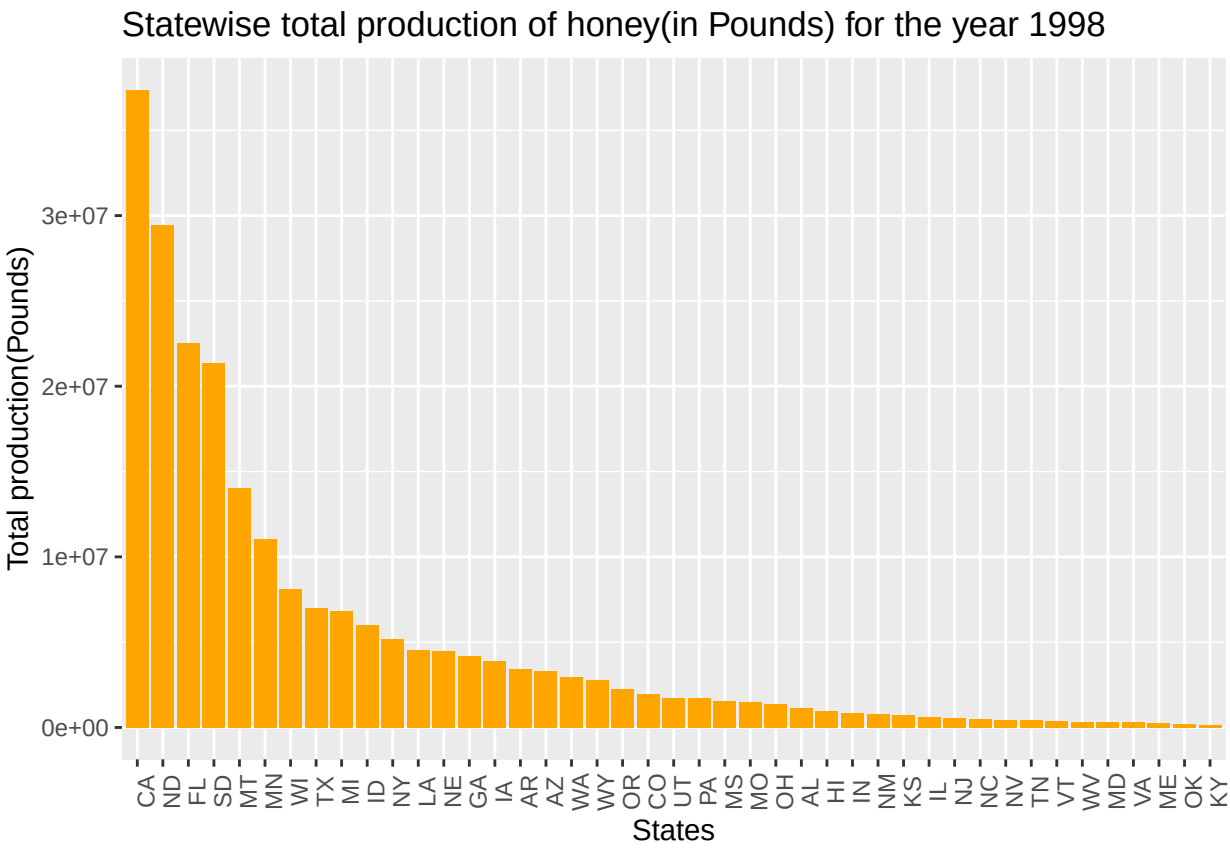
OBJECTIVE OF THE PROJECT

The main objective of this project is to 'Develop An Analytics Dashboard' using the R-Shiny application for the above dataset.

1) In the dashboard, we will represent different variables using different plots.
2) We will try to observe the trend of different variables over the period.
3) We will try to analyze dependent and independent variables and the relation between them by using different multivariate graphs.

Here are some examples of univariate and bivariate graphs.

1) Bar graph for state-wise Honey Production For The Year 1998.



From the above graph we can observe that;

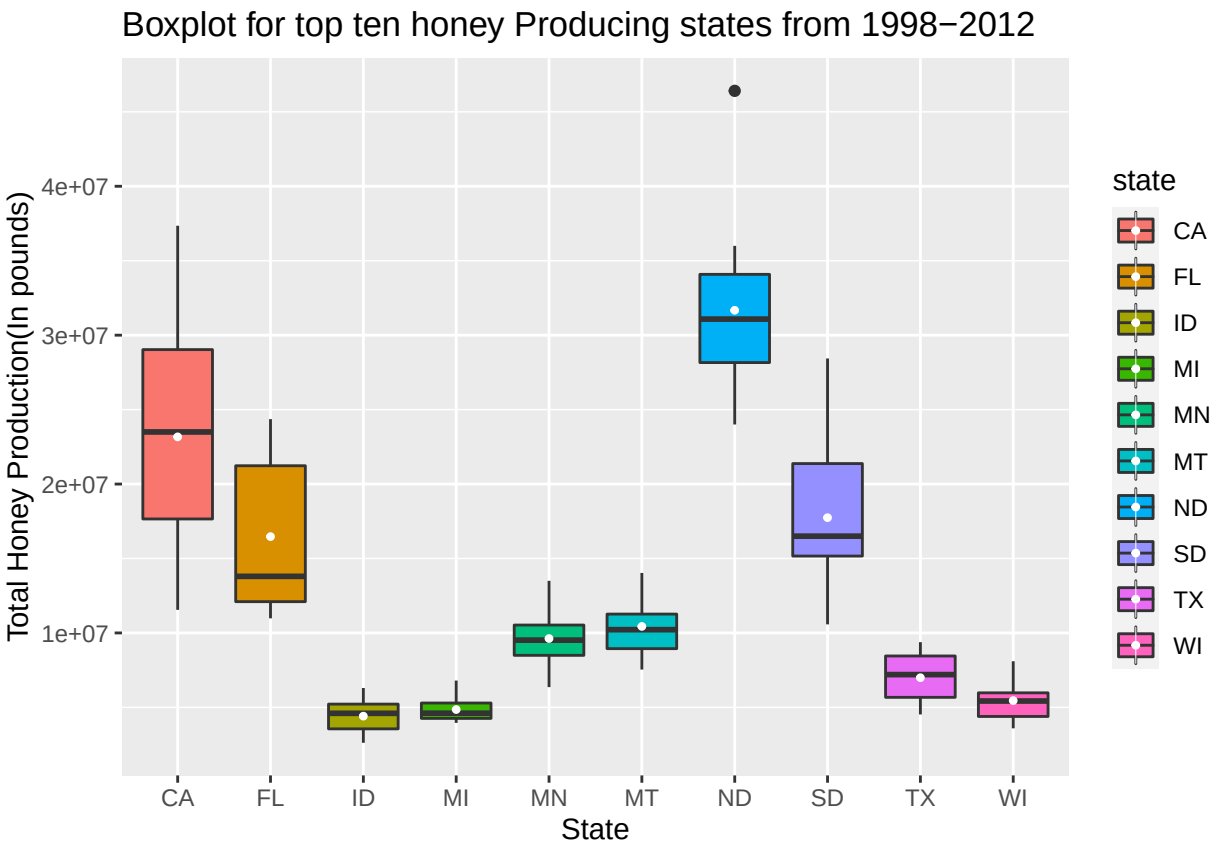
1) California, North Dakota, Florida, and South Dakota are the top 4 honey producing states for the year 1998. They are approximately producing half of the total honey production of the USA for the year 1998.

2) On the other hand, Kentucky and Oklahoma are the least honey-producing states for the year 1998.

3) Most of the top honey-producing states are located in the central part of the USA. It means honey production is more in the central part of the USA as compared to others.

Similarly, as above, we will plot bar graphs for the remaining years in the dashboard and try to find the topmost and least honey-producing state in each subsequent year.

2) Boxplot for total production of top ten honey Producing states from 1998-2012.



The above plot is a box plot representation of the total production of the top ten honey-producing states from 1998 to 2012. We can observe that,

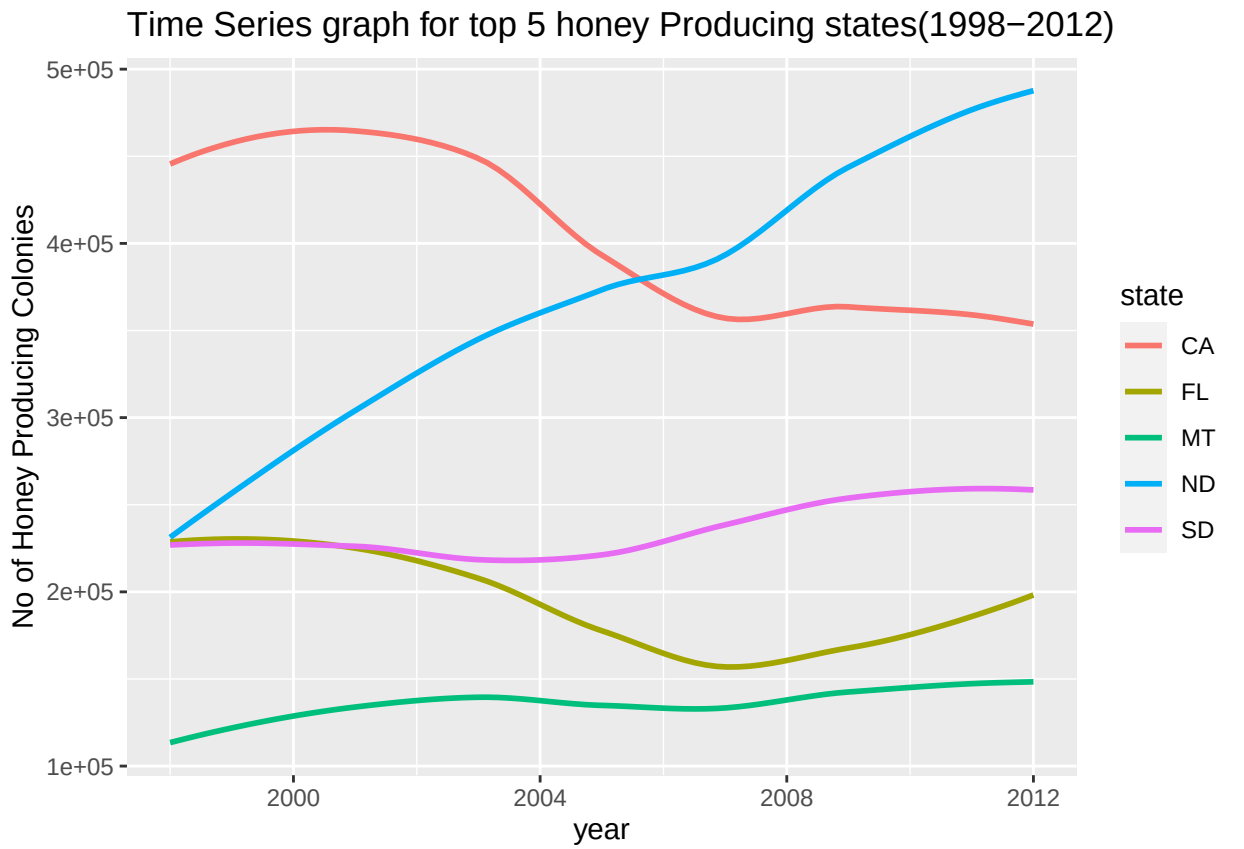
1) Although the total honey production in California is varying over a wide range, it is quite symmetric about the median. Like California, total honey production in Texas and Idaho is also quite symmetric about the median.

2) The highest median honey production is in North Dakota state.

3) In Wisconsin and Minnesota, median honey production and mean honey production is almost same.

In the dashboard, we will plot some more boxplots for region-wise located states and try to draw observations about a range of production, highest mean production, highest median production, symmetry about mean/median production, etc.

3) Number of colonies in top 5 honey Producing states from 1998-2012.



The above graph is showing, the number of honey-producing colonies in the top 5 honey-producing states from 1998 to 2012. From this graph, we can observe that,

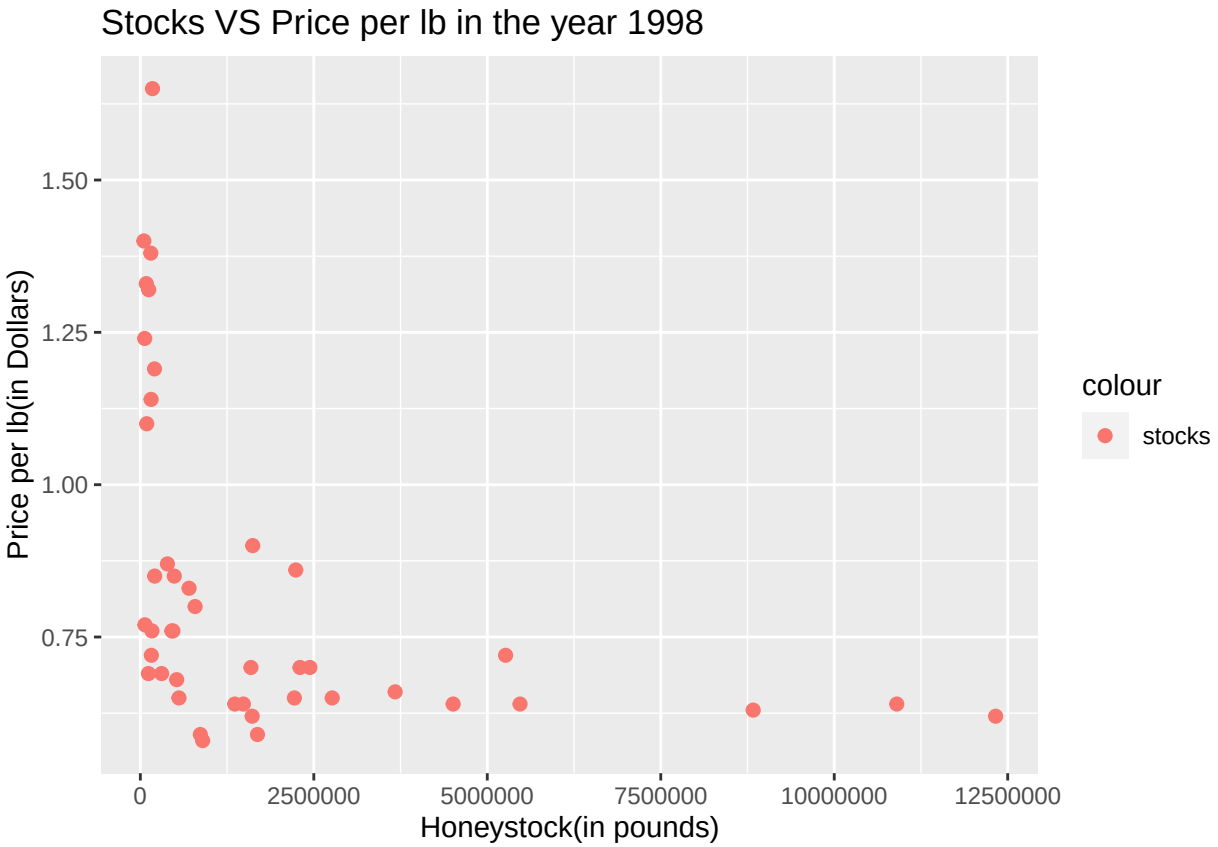
1) The number of honey-producing colonies in North Dakota state is rapidly increased over the period whereas, the number of honey-producing colonies is decreasing in California state after 2003.

2) The number of colonies in Montana state is almost the same every year. The number of colonies in Florida decreased after 2003 and again increased after 2007.

3) In South Dakota state, the number of colonies is slightly increasing over the period.

To analyze the trend, we will plot a 'Time Series Graph' for the remaining states in the dashboard. Also, we will try to find out the regions where No. of honey-producing colonies increasing/decreasing.

4) Point plot:Stocks vs Price per Pound of Honey for the year 1998.



From the above plot it is easy to observe that,

1) In 1998, Honeystock in most of the states is less than 2500000 pounds.
2) The price of honey per pound is decreasing with an increase in honey stock. It means, the price per pound of honey is more when honey stock is less and vice versa. Thus, above two are dependent variables.
3) In the year 1998, the price per pound of honey is almost less than 1.5 Dollars irrespective of the honey stock.

In the dashboard, we will plot more point plots for subsequent years and also for different variables. Using that we will try to analyze how variables depend on each other.

Conclusion:

1) Total honey production is more in the central part of the USA as compared to other parts.
2) In some states, total honey production is varying over a wide range whereas it is almost the same for some states throughout the period.
3) Price per pound of honey and Honey stock are dependent variables. The price per pound of honey is increasing with a decrease in honey stock.
4) The number of honey-producing colonies is increasing in some states and is decreasing in other states. For some states, it is almost the same throughout the period.