Exploratory Analysis on Honey Production in USA

Data Set Link: https://www.kaggle.com/jessicali9530/honey-production

Team Members
Name: Ganesha K S
SRN: PES1201701731
Email ID: toganesha13@gmail.com
Contact No: 8277637836
Name : Arpan Ghoshal
SRN: PES1201701700240
Email ID: arpanghoshal77@gmail.com

Abstract

- This assignment is about an analysis on the honey production in the states of USA.

- The data set contains the information on the production values of Honey on different colonies and the year in which that number of colonies existed

- The analytic code contains the simple box plots which compares the production of honey in various states and we can thus see the variations in the production from 1998 to 2012

Data Set

- Data Set

- numcol: Number of honey producing colonies. Honey producing colonies are the maximum number of colonies from which honey was taken during the year. It is possible to take honey from colonies which did not survive the entire year

- yieldpercol: Honey yield per colony. Unit is pounds

- totalprod: Total production (numcol x yieldpercol). Unit is pounds
.

- stocks: Refers to stocks held by producers. Unit is pounds

- priceperlb: Refers to average price per pound based on expanded sales. Unit is dollars.

- prodvalue: Value of production (totalprod x priceperlb). Unit is dollars.

- Other useful information: Certain states are excluded every year (ex. CT) to avoid disclosing data for individual operations. Due to rounding, total colonies multiplied by total yield may not equal production. Also, summation of states will not equal U.S. level value of production

Questions that this analysis helps to answer:

- How has honey production yield changed from 1998 to 2012?

- Over time, which states produce the most honey? Which produce the least? Which have experienced the most change in honey yield?

- Does the data show any trends in terms of the number of honey producing colonies and yield per colony before 2006, which was when concern over Colony Collapse Disorder spread nationwide?

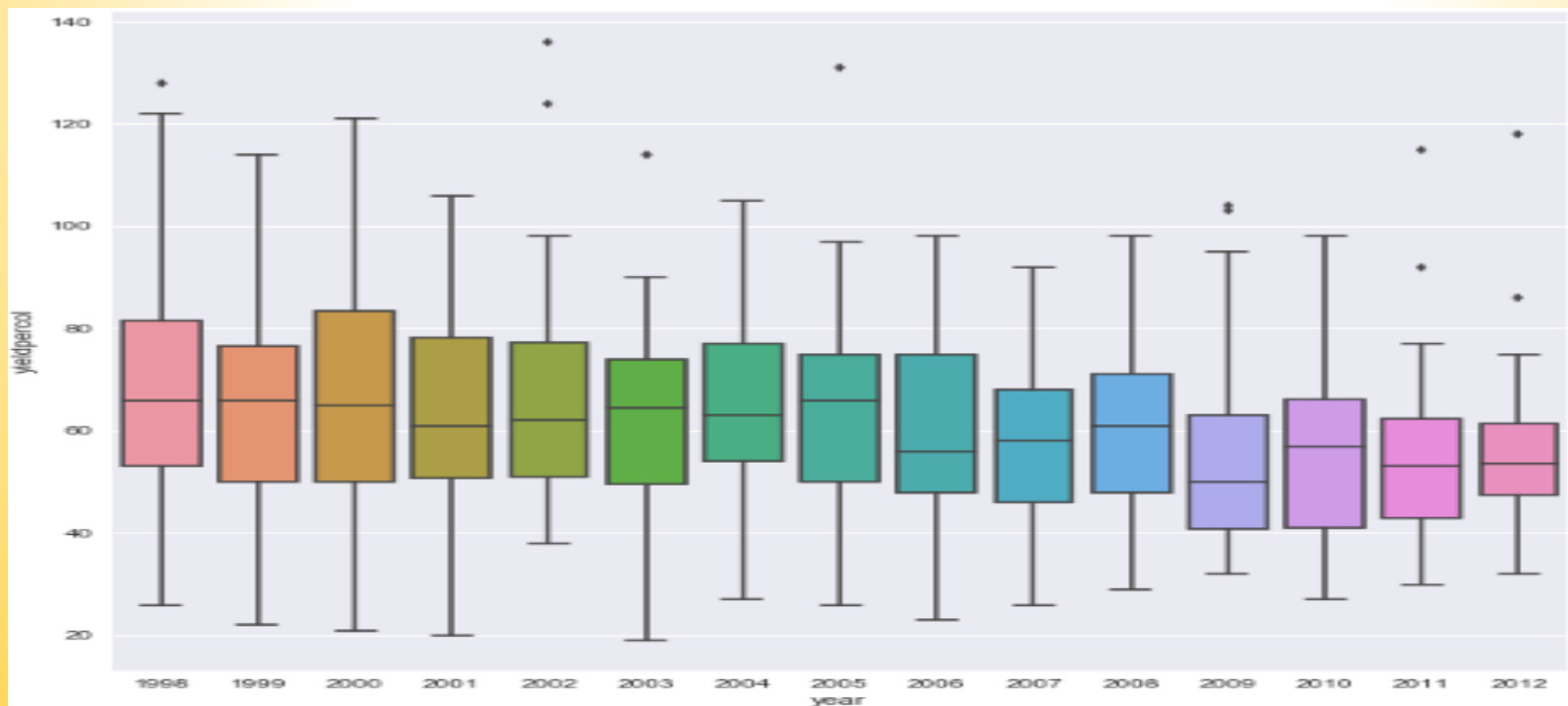# Exploratory Analysis OR Insights Drawn

- The boxplots that are plotted in the kernel tells us about the production values, price per pound, yield per colony.
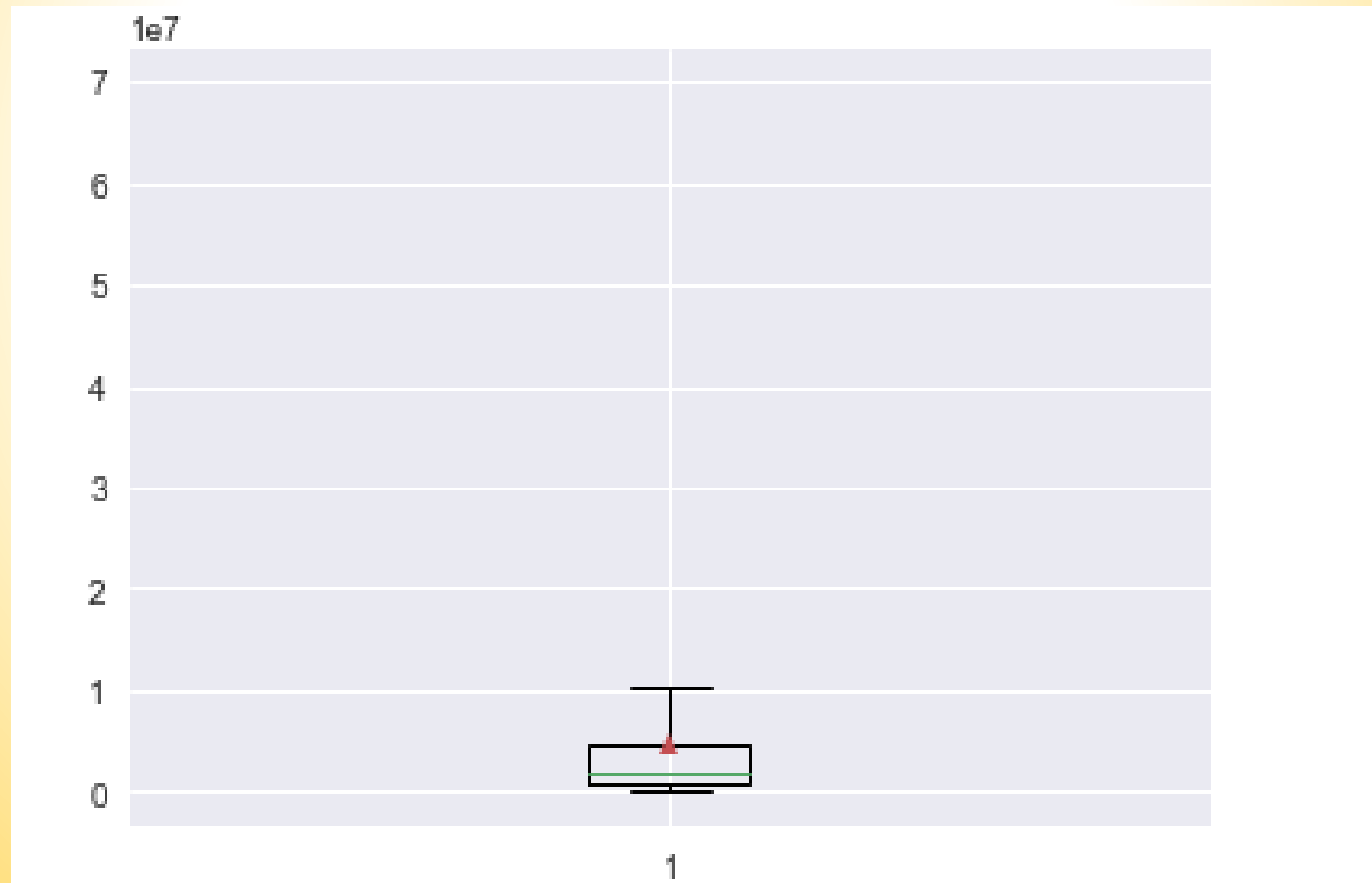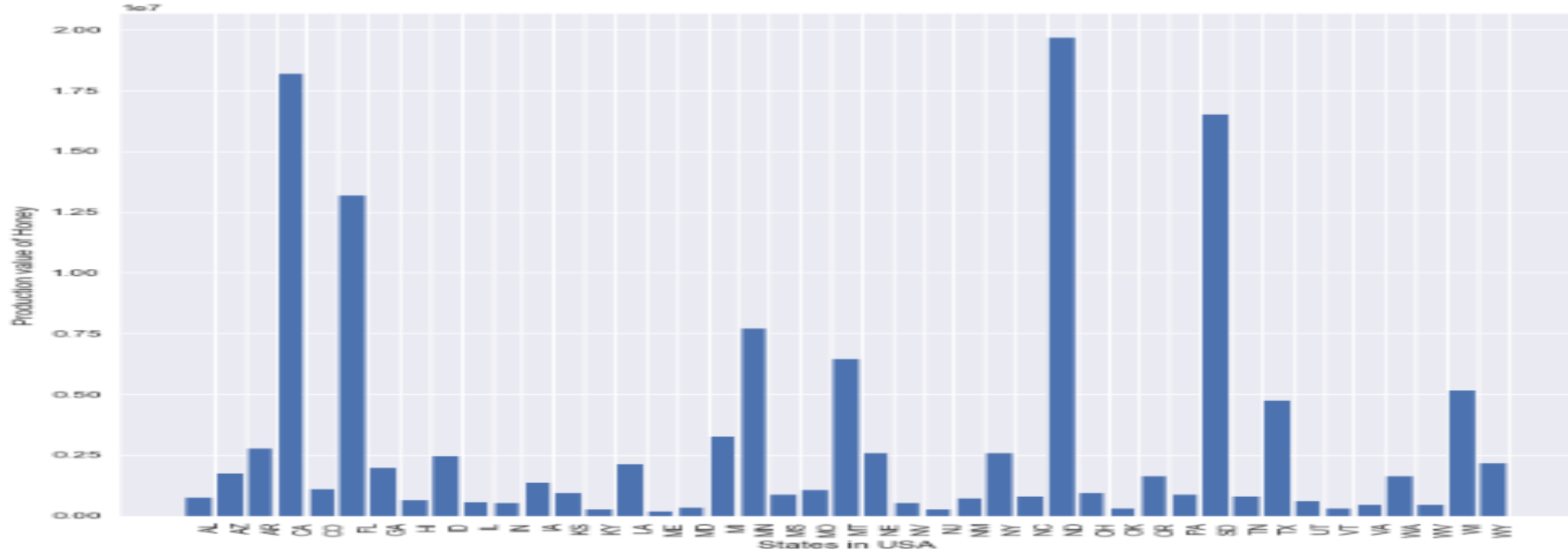
Exploratory Analysis OR Insights Drawn

- Another boxplot plotted gives us an idea about the yield per colony of the honey from different states throughout the year. This data set also includes the outliers to give an idea about the practicalities in the spread of the data
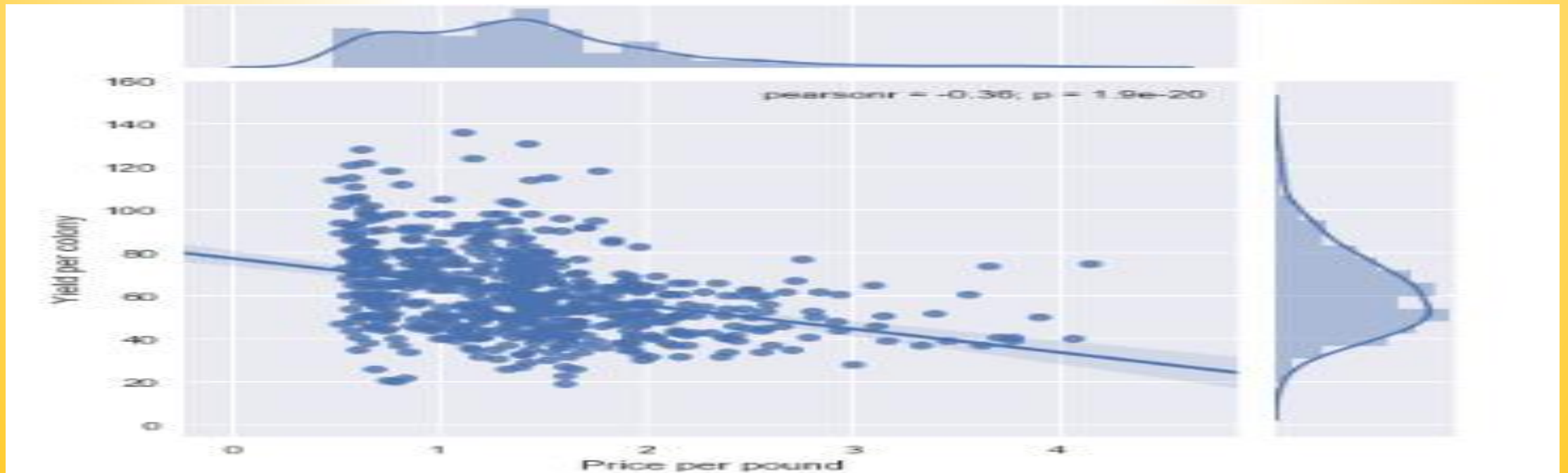
Exploratory Analysis OR Insights Drawn

- The values of the production of honey is appended to a list and the name of the states to another list which is then used to plot a bar graph which helps us to identify which state in the USA produces honey which has the highest production value.
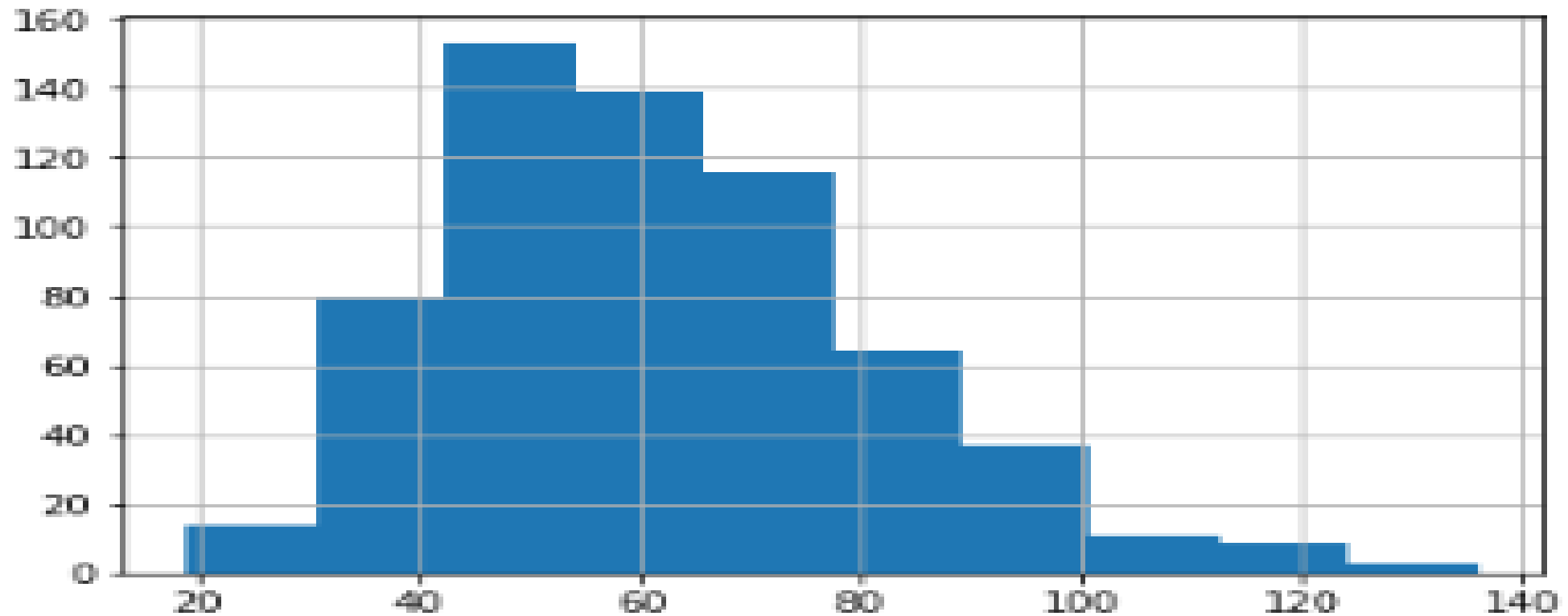
Exploratory Analysis OR Insights Drawn

- The values of the price of honey per pound and the yield per colony are used to plot a scatter plot which identifies the amount of honey produced which has varying values of pounds (which tells us the which honey is worth buying).

Exploratory Analysis OR Insights Drawn

- We also specify the use of a Histogram in the dataset by plotting the histogram for the same column i.e., yield per colony.

Exploratory Analysis OR Insights Drawn

- Another relationship can be drawn between stocks held by the producers of honey and the total production (yield per colony X number of colonies) by drawing a joint plot which turns out to be an approximately linear graph from which we can conclude that the more the honey was produced the more financially successful the producers were and hence they contain more stocks.

Exploratory Analysis OR Insights Drawn
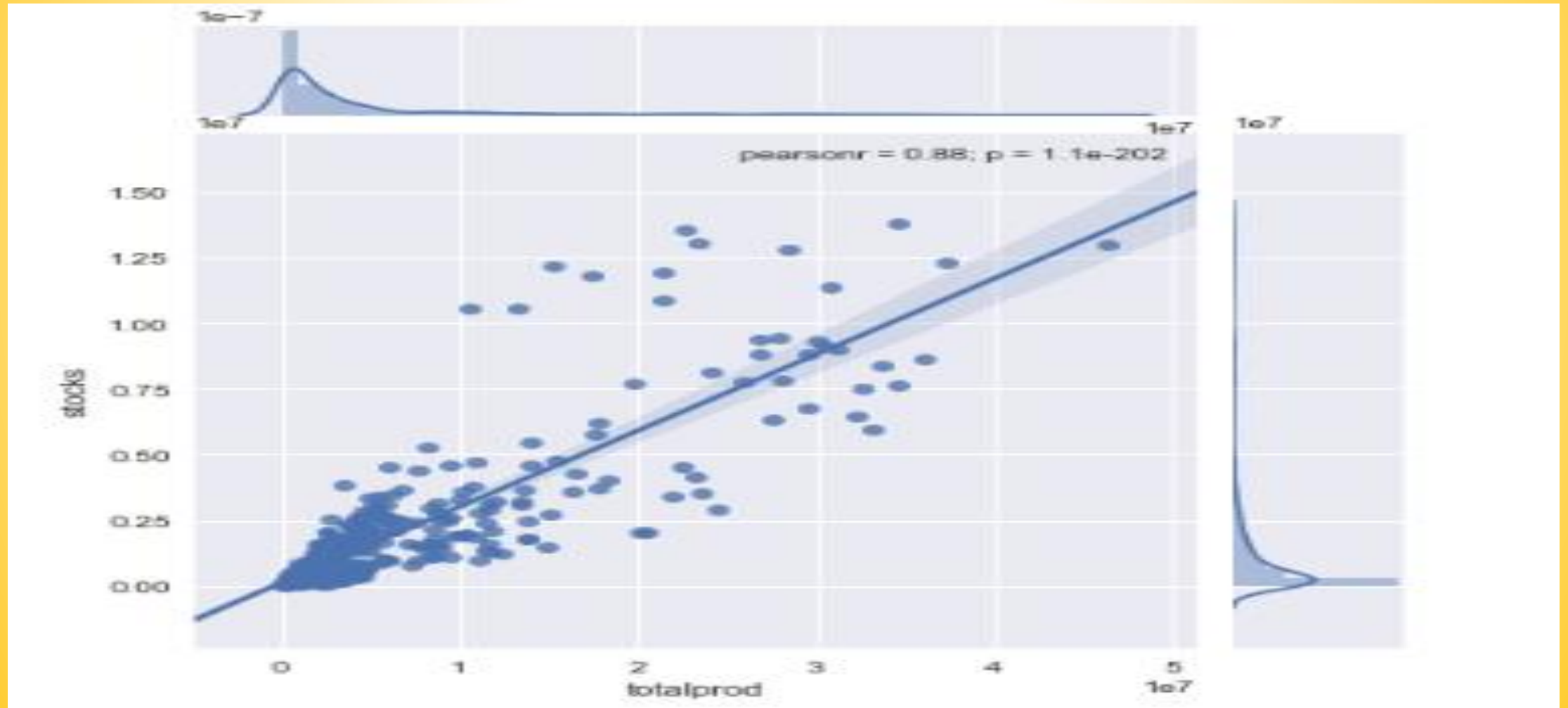
Conclusion

- The production per colony is decreasing in most of the states
- Top 5 states with positive correlation with respect to year: Mississippi, South Carolina, Maine, Oklahoma, Virginia
- Top 5 states with negative correlation with respect to year: Florida, Oregon, Minnesota, Indiana, Michigan
- The total production is decreasing in most of the states
- Top 5 states with positive correlation with respect to year: Mississippi, North Dakota, North Carolina, Kentucky, New Jersey

Thank You!!