

USDA Strawberry Report

Group 3: Rose Determan, Yongrong Chai, Wuji Shan, Tong Sun

11/03/2021



1. Introduction

In this project, we use the data sets strawberries, insecticides, and herbicides-fungicides-other from USDA. We first wrangled these files, finished data cleaning and organizing the final data set; then we start to visualize and explore the data. This report focus on 2 main parts: Data cleaning and EDA.

- Data cleaning: We wrangle the data sets strawberries, insecticides, and herbicides-fungicides-other. Separate some columns and eliminate the redundancy.
- EDA: We plot a pie chart with a proportion of dataset by states, maps of comparison of strawberry yields and pesticide application between states and years, and a dot plot of the relationship between the level of Bee Toxicity and Carcinogen level.

Besides, we deploy a shiny app Shiny_strawberries for data display and exploration.

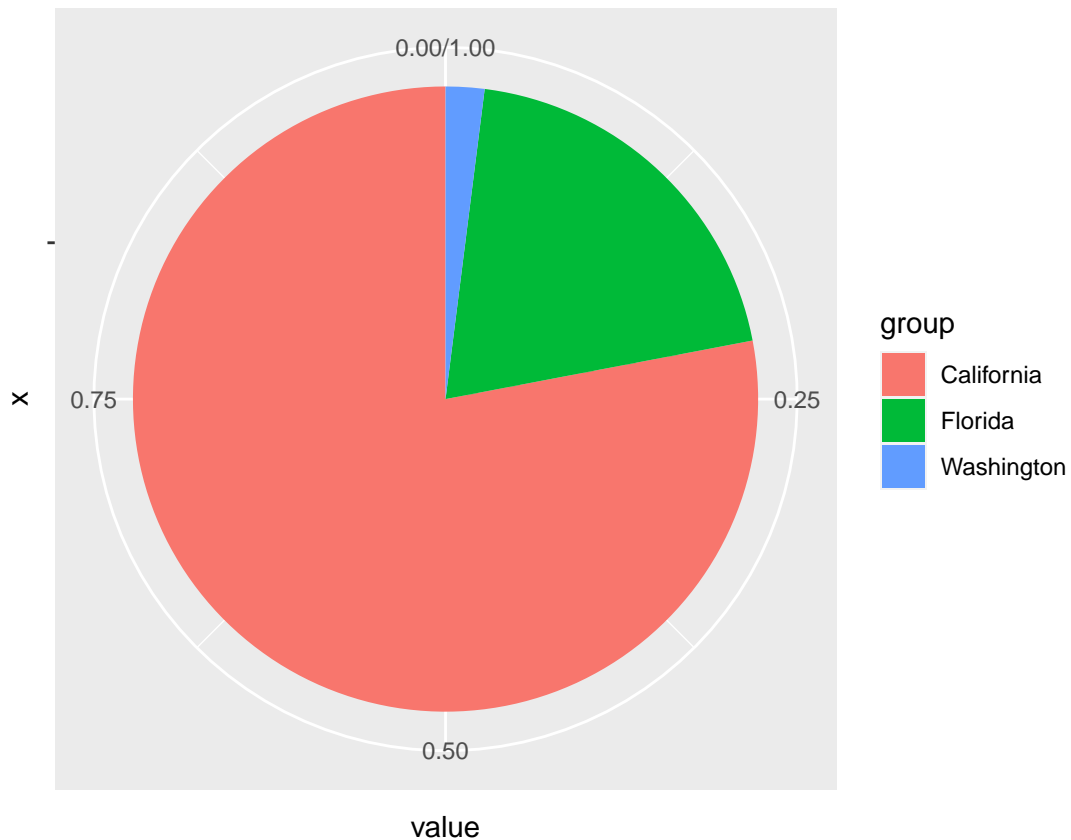
2. Research Questions

- Comparison of strawberry yields between states.
- How does the pesticide application vary between states and years?
- Is there any relationship between Level of Bee Toxicity and Carcinogen Level?

3. EDA – We added mouse over into our plots.

3.1 Show a pie chart with proportion in number of three states: CA, FL, WA

```
##      group value
## 1 California 0.78
## 2  Florida 0.20
## 3 Washington 0.02
```



The proportions of strawberries numbers have a significant difference between states. California has the largest share, and Washington has the smallest one. (CA > FL > WA)

3.2 Show a map of Strawberries Yield measured in CWT/Acre and Pesticide Application

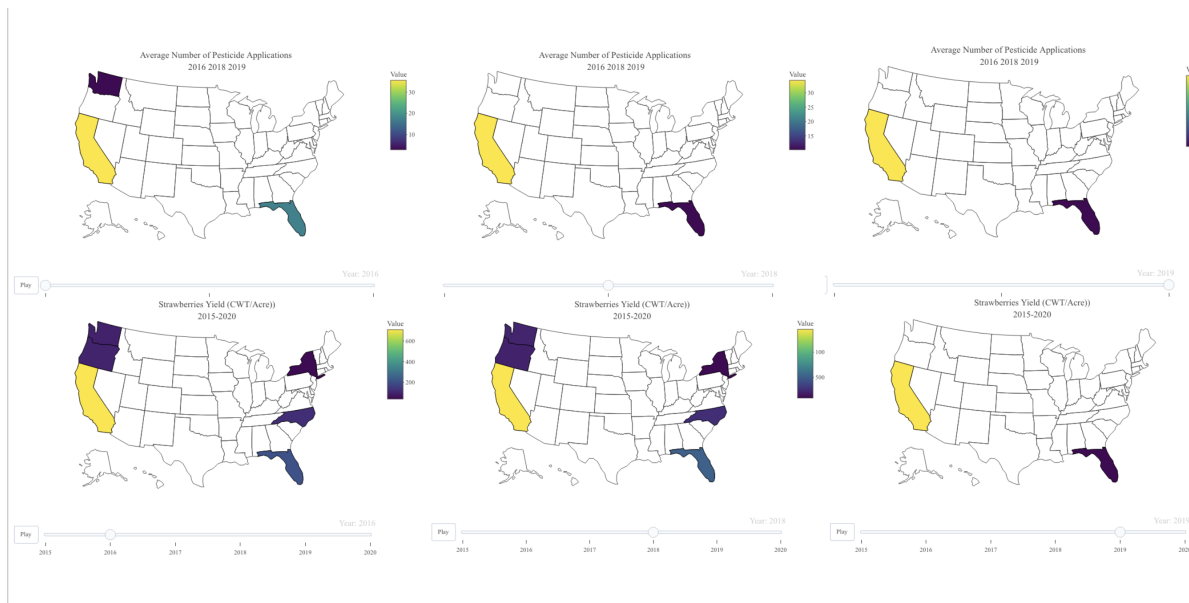
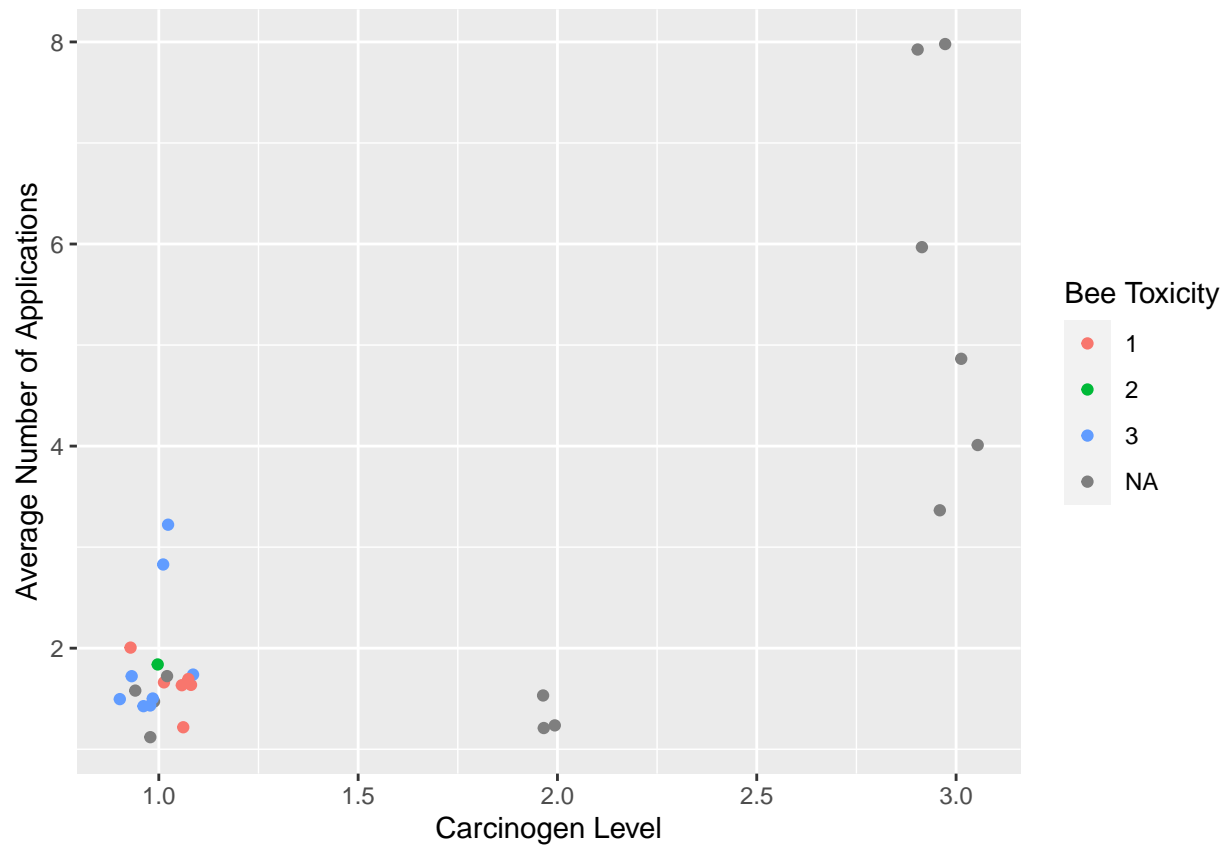


Figure 1: Strawberries

The maps above shows that there was a large increasing from 2018 to 2019, and based on the research, with the development of newer varieties, strawberry season has expanded in both California and Florida. That's the reason of yield increasing. The other thing is that California produces strawberries year-round, and that's why strawberries are primarily grown in California.

For the comparison about average number of pesticide applications, there is an interesting founding here, there is a negative relationship between strawberry yield and pesticide applications from 2018 to 2019.

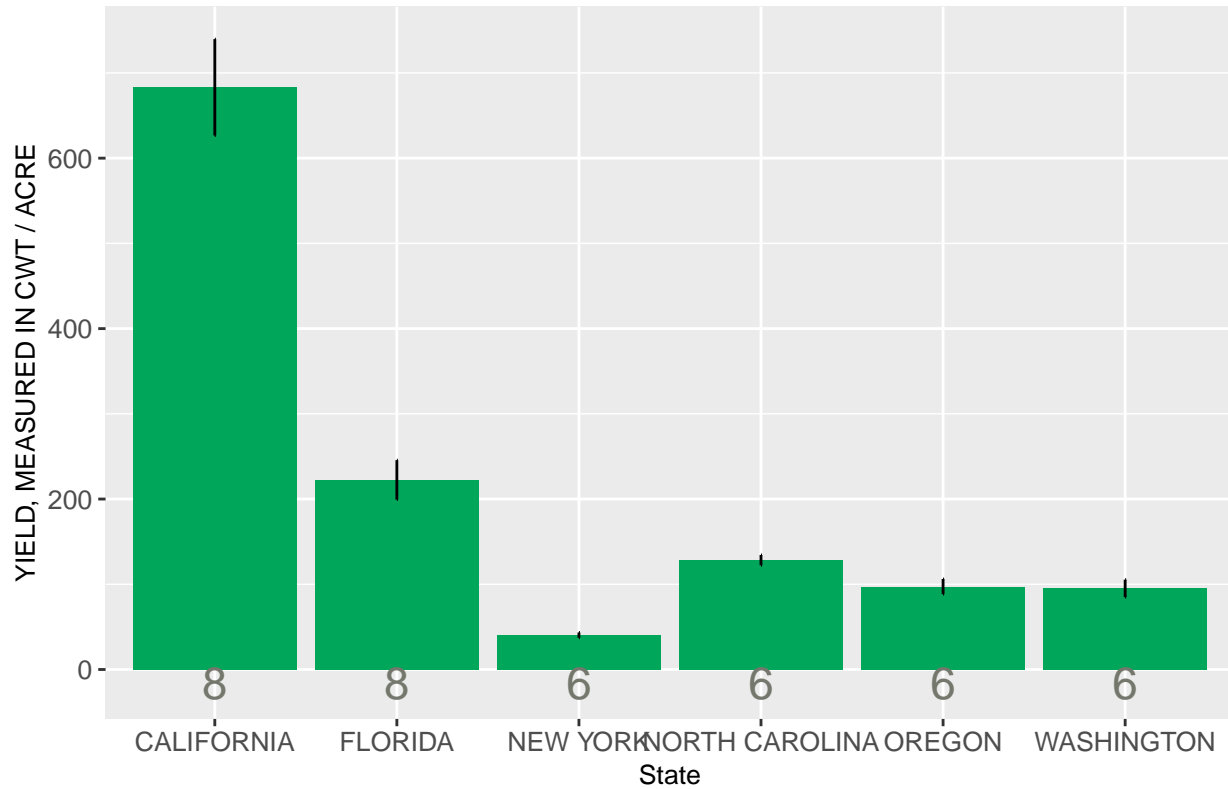
3.3 Show the correlation of average number of Bee Toxicity measured in applications and carcinogen level



3.4 Plot the mean yields by state

Strawberry yield in CWT/acre with standard deviation error bars. The number below each bar indicates the sample size represented in the bar. California clearly produces the most strawberries, followed by Florida and North Carolina. As the standard deviation error bars show, each state reliably produces similar quantities of strawberries across the years for which we have data.

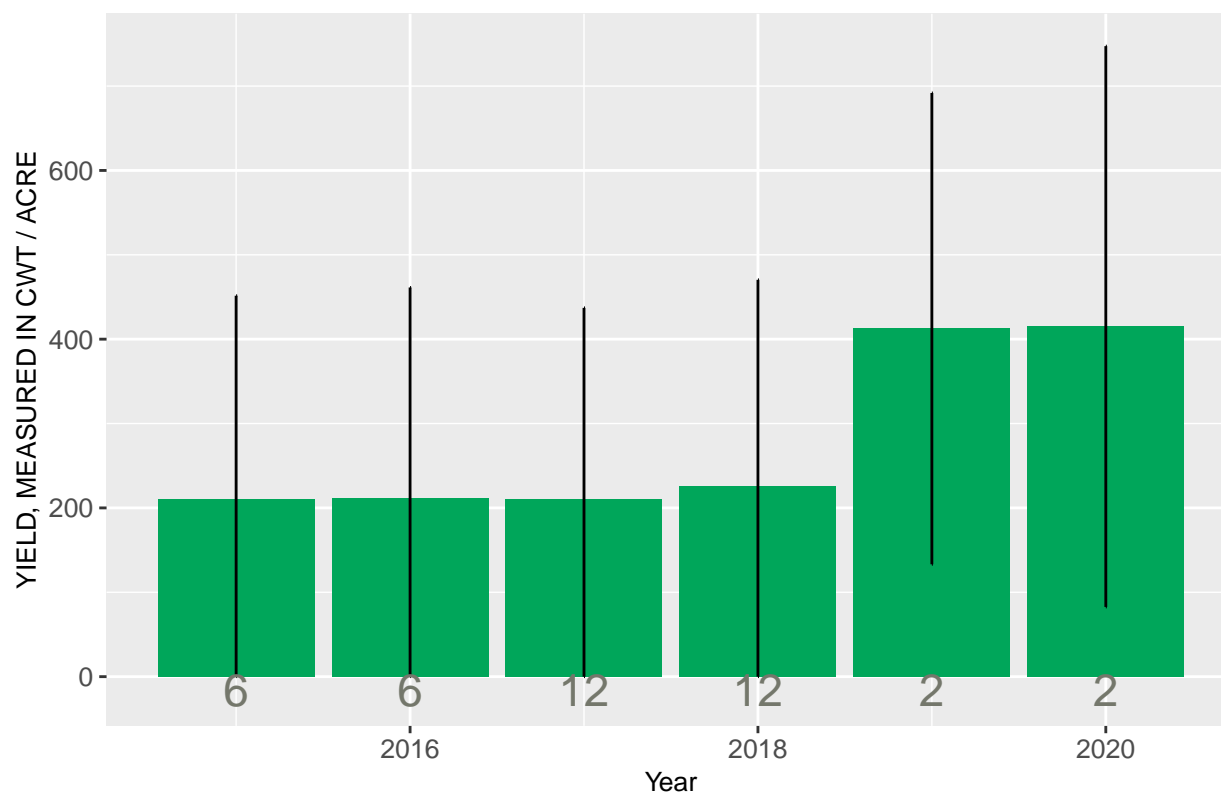
Mean of Strawberry Yield with St.Dev Errors and Count



3.5 Plot the mean yields by year

Strawberry yield in CWT/acre with standard deviation error bars. The number below each bar indicates the sample size represented in the bar. We can see that in 2019 and 2020 there are very small sample sizes, and, as a result, large standard deviations. Each year also has a large standard deviation, since the selected states produce very different quantities of strawberries.

Mean of Strawberry Yield with St.Dev Errors and Count



4. Reference

<https://quickstats.nass.usda.gov/#56E84525-1350-34A1-9ED7-27363BD5A7D3>

<https://www.dailymail.co.uk/health/article-4710590/Why-buy-organic-strawberries.html>