

Midterm project

Group 10: Kayla Choi, Kosuke Sasaki, Clare Zike Tan, Franky Yifan Zhang

2021/11/03

1. Introduction

Our research questions about the strawberry production and chemical use in the US are:

- i) Does the type of chemical have a correlation to the amount of strawberries produced (Measured in lb/acre/ application)?
- ii) Do different states use different chemicals?
- iii) How do the amount of strawberries produced differ by state?

To answer these questions, we looked up two data sets, cleaned and merged them into one data set, and created graphs, tables and maps for Exploratory Data Analysis (EDA).

2. Methods

i) Data set

We used two data sets, USDA strawberry data and pesticide data. For the strawberry data set, the provided one in the class does not seem to be enough for our research in terms of information on each state production, so we looked up the data on the USDA web site again and got the data set which include in total 10215 observations of strawberry data in 10 states.

For the pesticide data, we adopted the provided data.

Finally, we combined these two data sets, deleted rows with no information, and selected the variables needed for EDA.

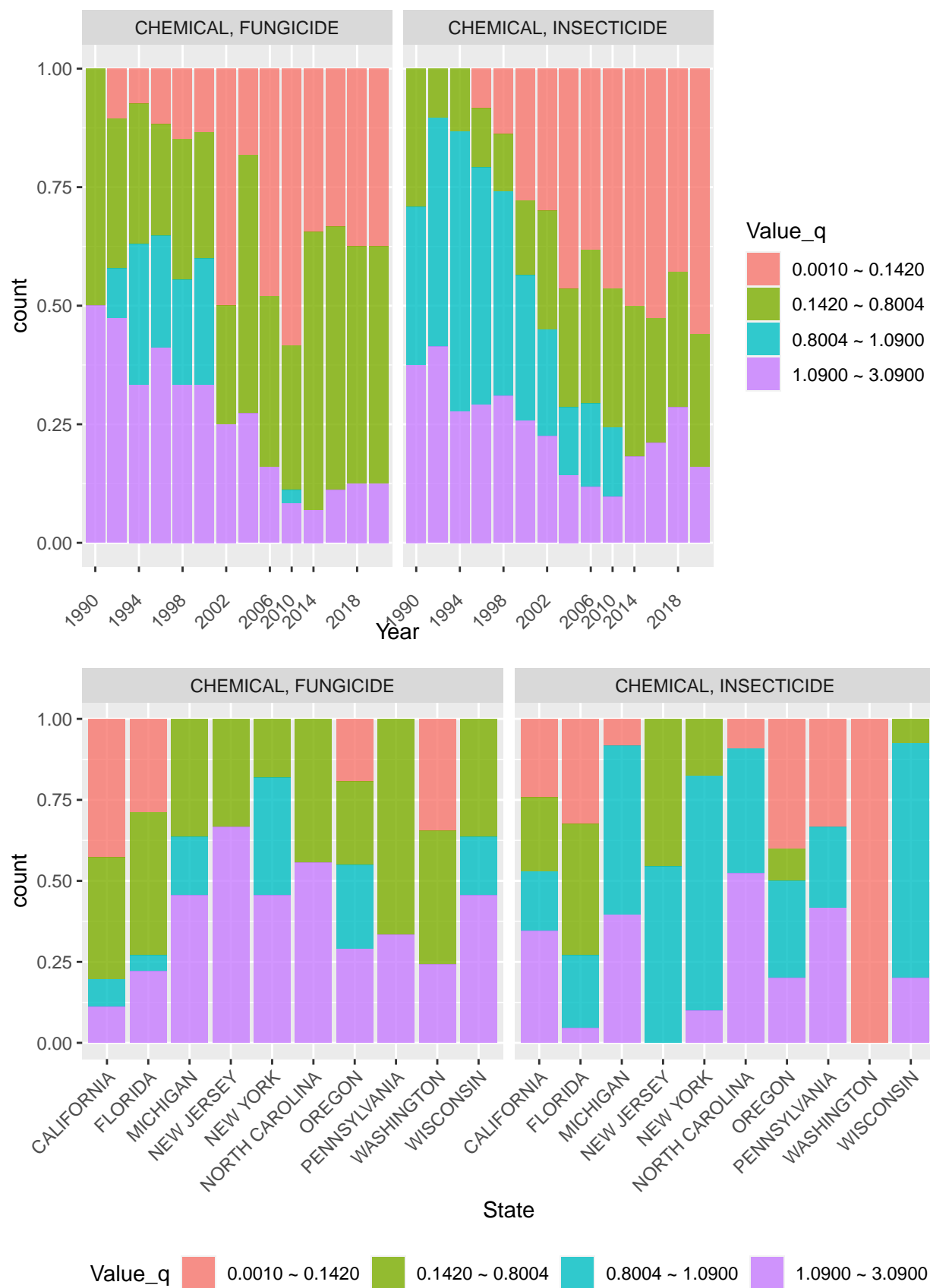
ii) Exploratory Data Analysis

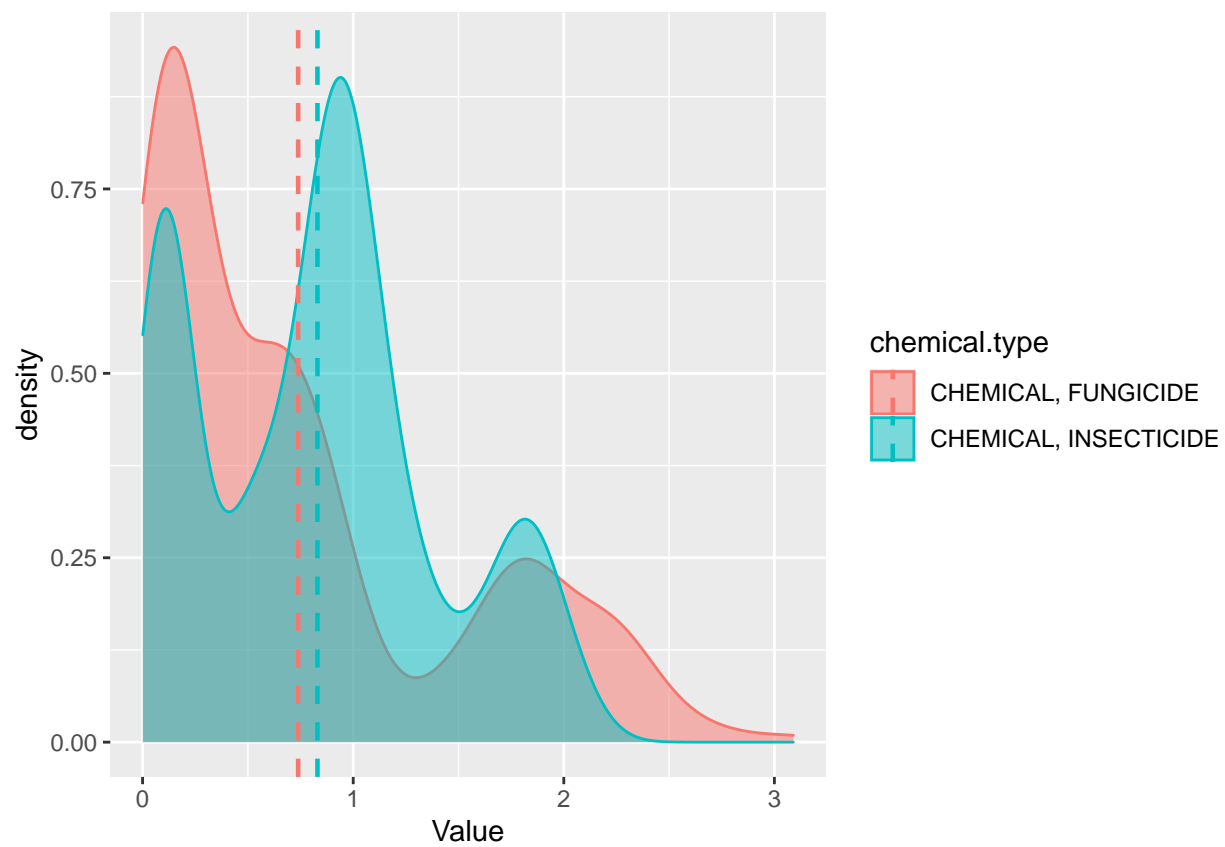
- **EDA by chemical type** First we analyzed the data set making bar chart, violin and box plot, and density graph by chemical type to see if there is any difference in the amount of strawberry production between the two chemical types.

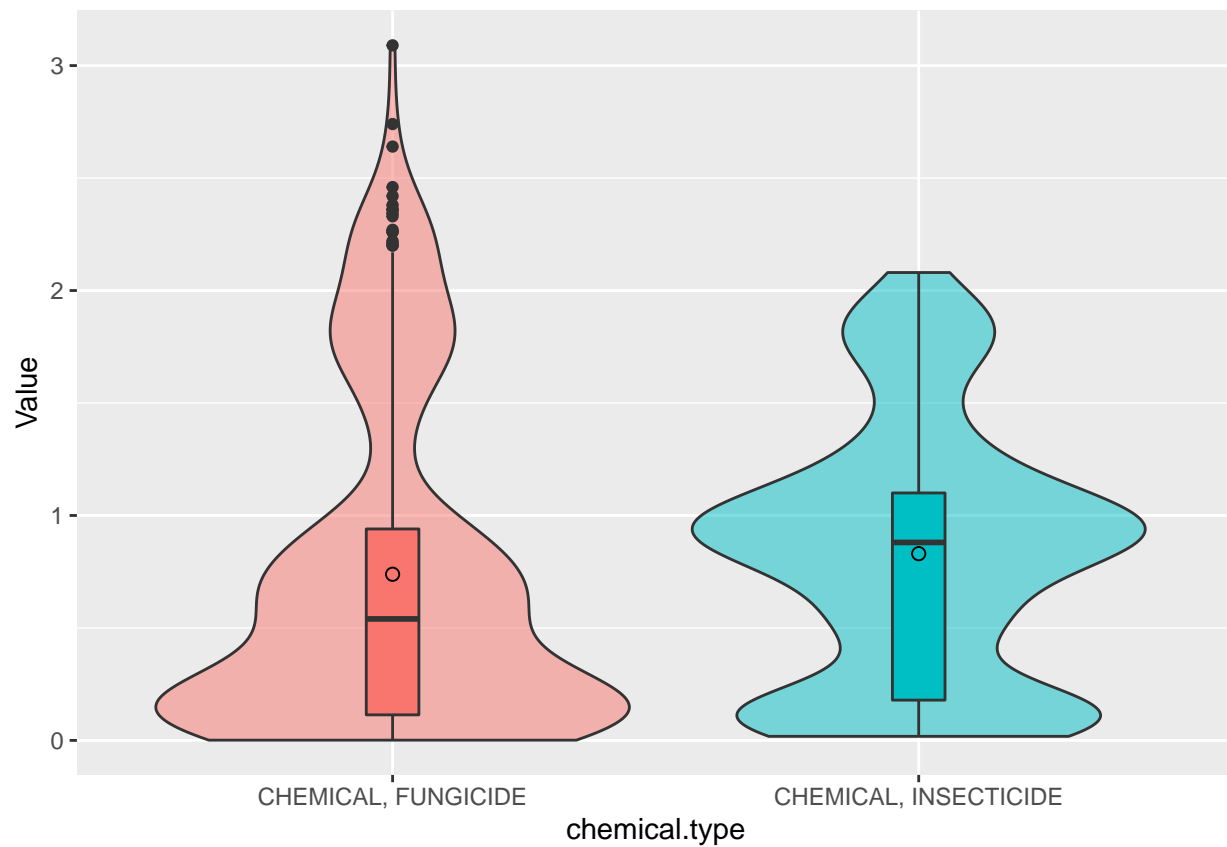
- **EDA by year and state** We also analyzed the data in terms of the amount of strawberry production by year and state where we make use of shiny application which allows us to explore strawberry production trends in each state and year more easily.

- **EDA by toxicity level** Finally we analyzed the data in terms of the amount of strawberry production by toxicity level.

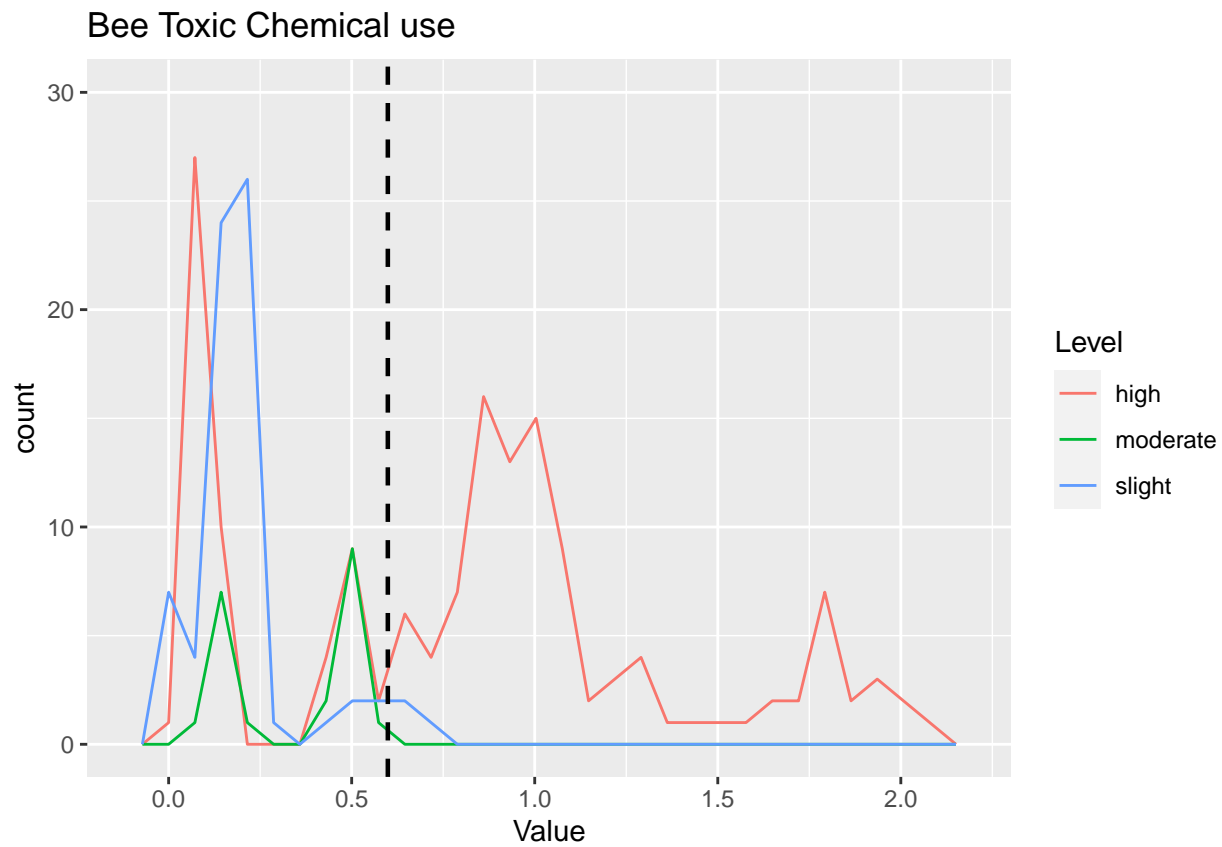
3. Results





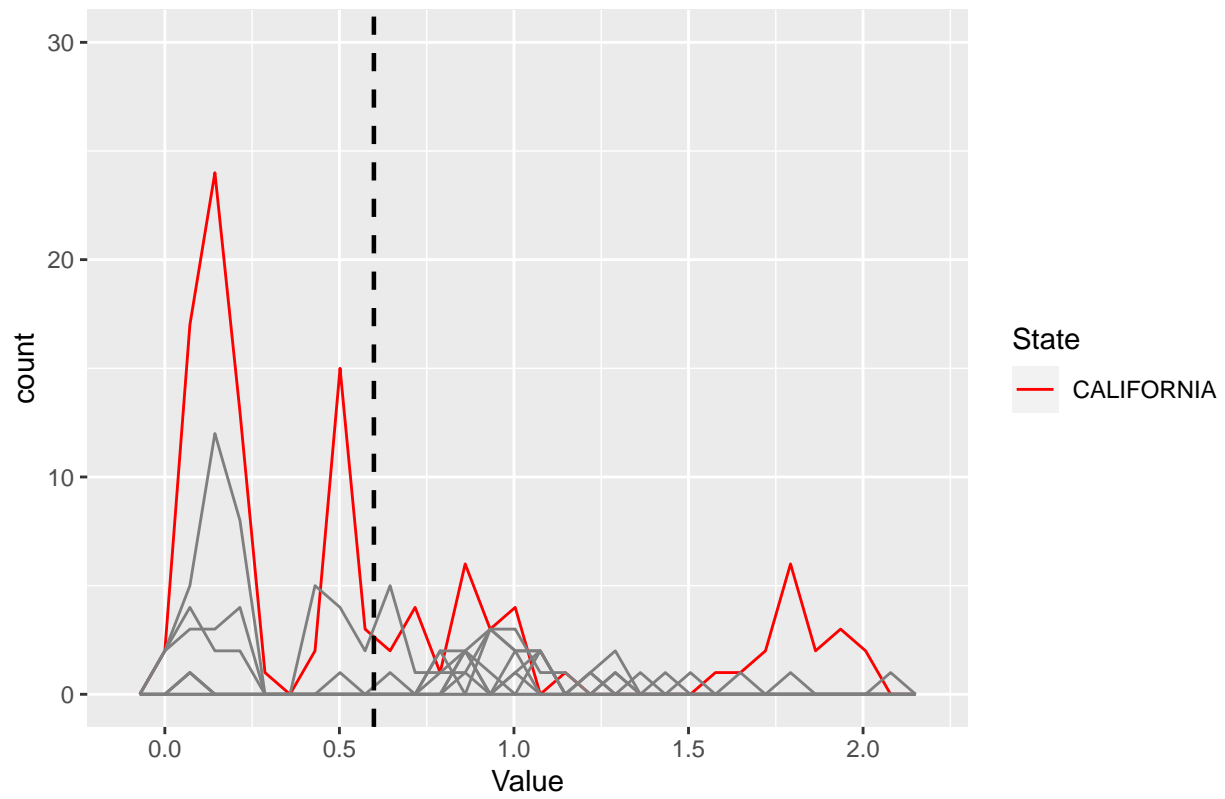


```
## PhantomJS not found. You can install it with webshot::install_phantomjs(). If it is installed, please
## 'stat_bin()' using 'bins = 30'. Pick better value with 'binwidth'.
```

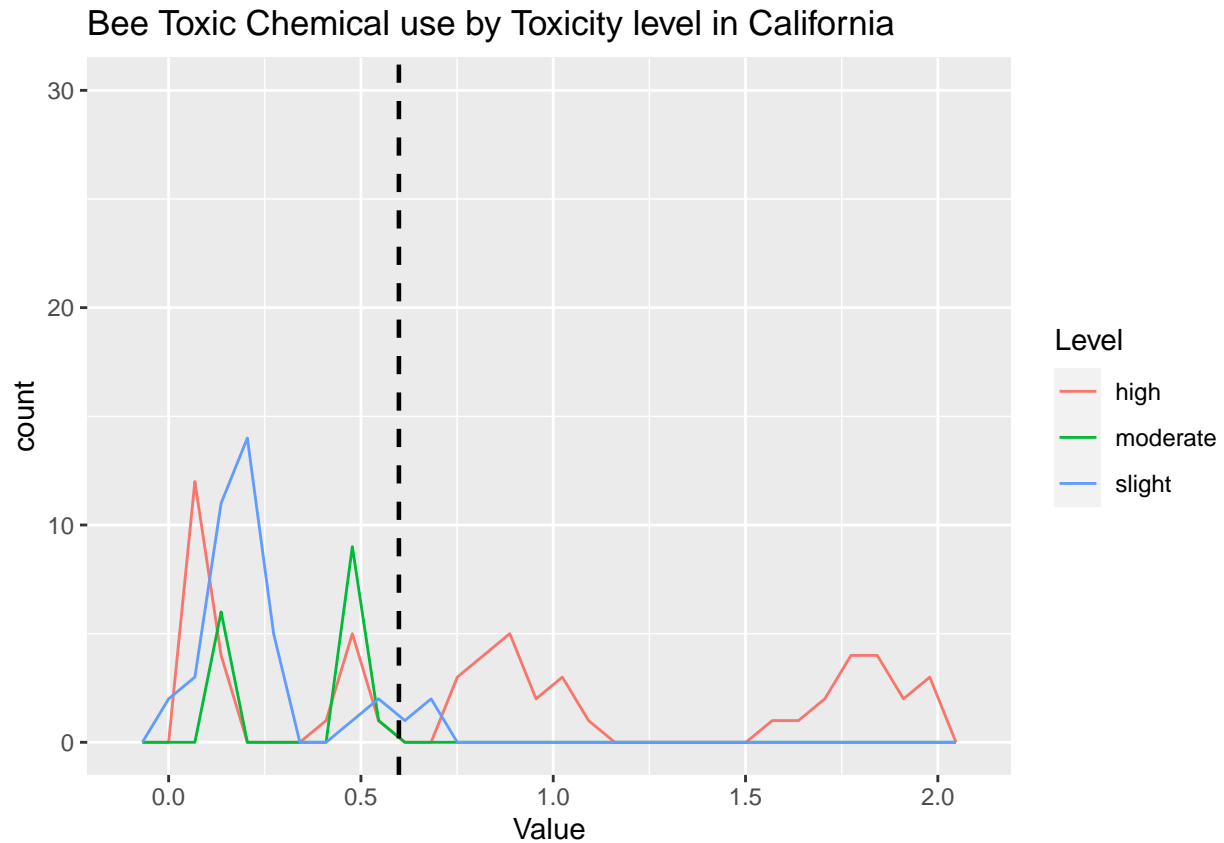


```
## 'stat_bin()' using 'bins = 30'. Pick better value with 'binwidth'.
```

Bee Toxic Chemical use in California compared to other states



```
## 'stat_bin()' using 'bins = 30'. Pick better value with 'binwidth'.
```



interactive plot of mean value

'summarise()' has grouped output by 'State', 'chemical.type'. You can override using the '.groups' a

'summarise()' has grouped output by 'State', 'chemical.type'. You can override using the '.groups' a

4. Conclusion

- Strawberry produced by acre by application has decreased over time.
- California used to be the top producer
- Florida is overtaking California
- High toxic chemical for bee seemed to be correlated with larger strawberry production.