

Midterm_Presentation

Mi Zhang, Peng Liu, Yuanming Leng, Qiannan Shen

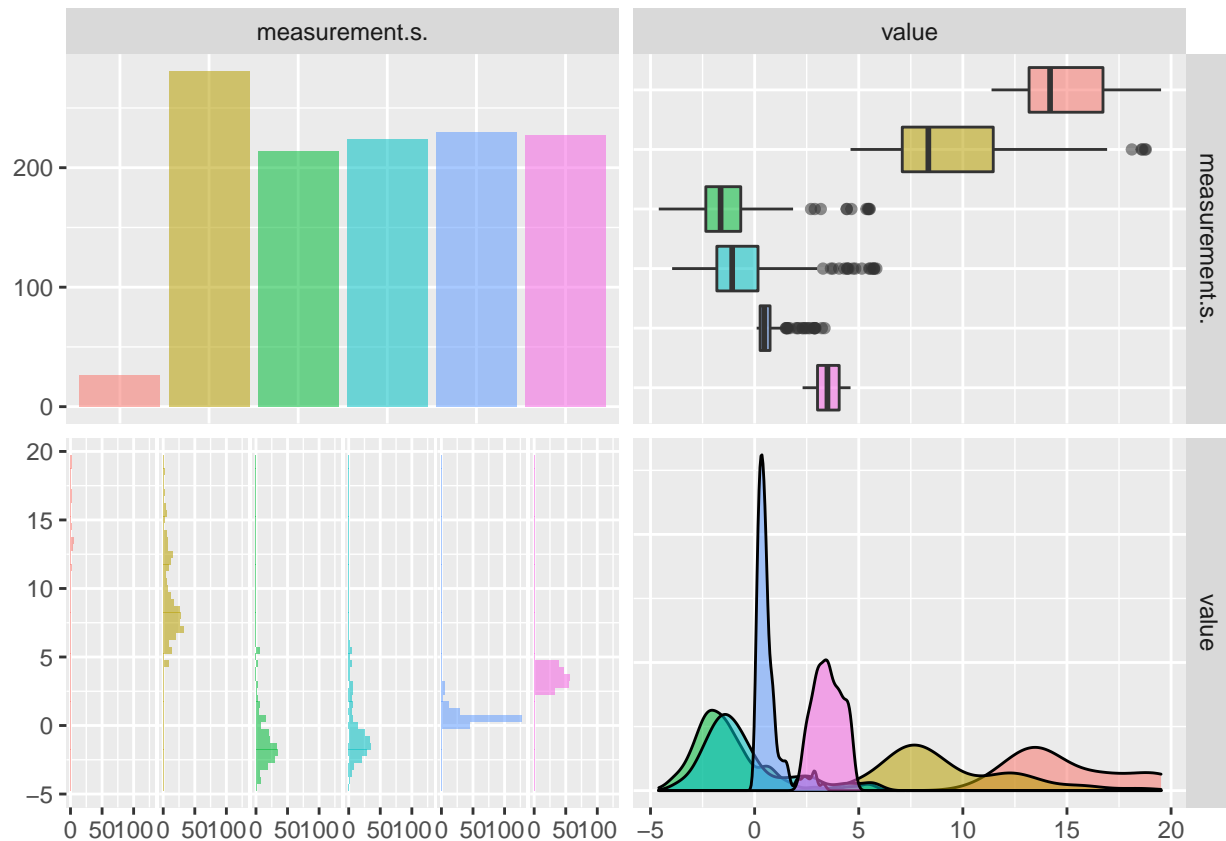
11/2/2021

Introduction

Our strawberries dataset is taking from the United States Department of Agriculture (USDA) and the Pesticide dataset is from Haviland. The strawberry dataset consists the years of 2015, 2016, 2018, 2019 and states including California, Florida, New York, North Carolina, Oregon, and Washington.

Data Cleaning for Strawberries

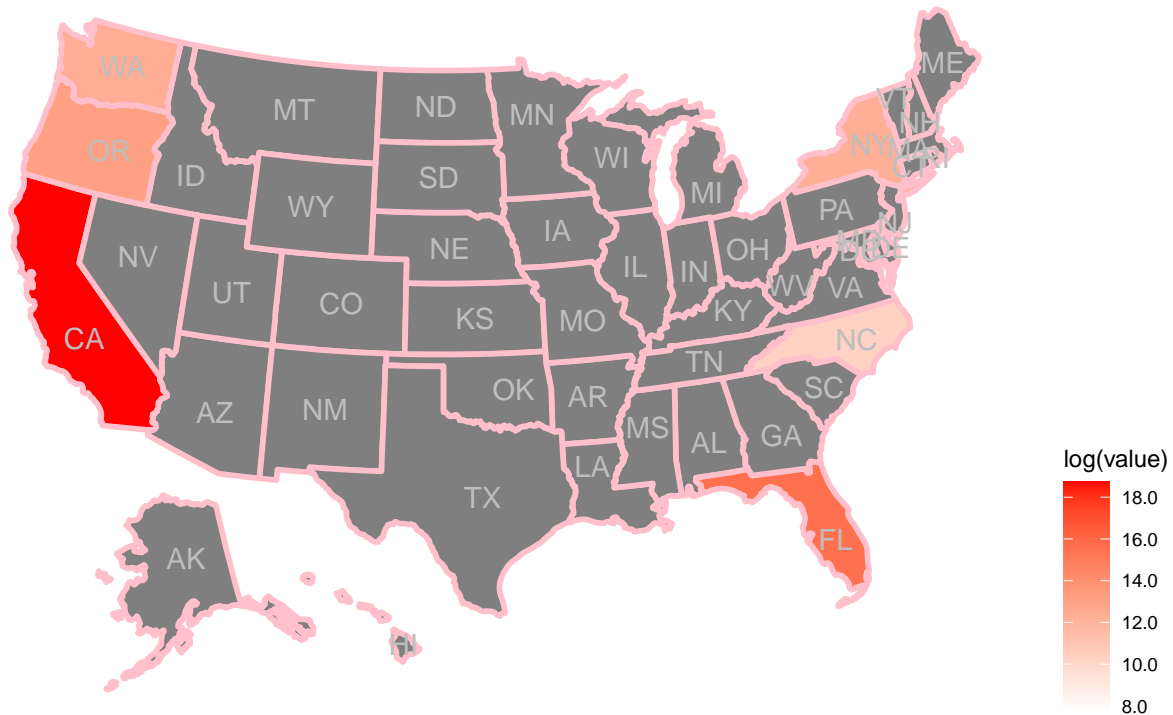
- remove empty/missing values and reduce white space in the cells
- split column with multiple items to separated columns
- redefine “MEASURED IN CWT” by multiplying by 100
- make extreme large value more accessible by using log scale on “value”
- redefine value as production of strawberries



- The following map is an example showing that both California and Florida have higher amount of annually strawberry production in pounds than other states.

```
map(years = "2019",
     chemical = "NOP USDA CERTIFIED", measurement = "MEASURED IN LB")
```

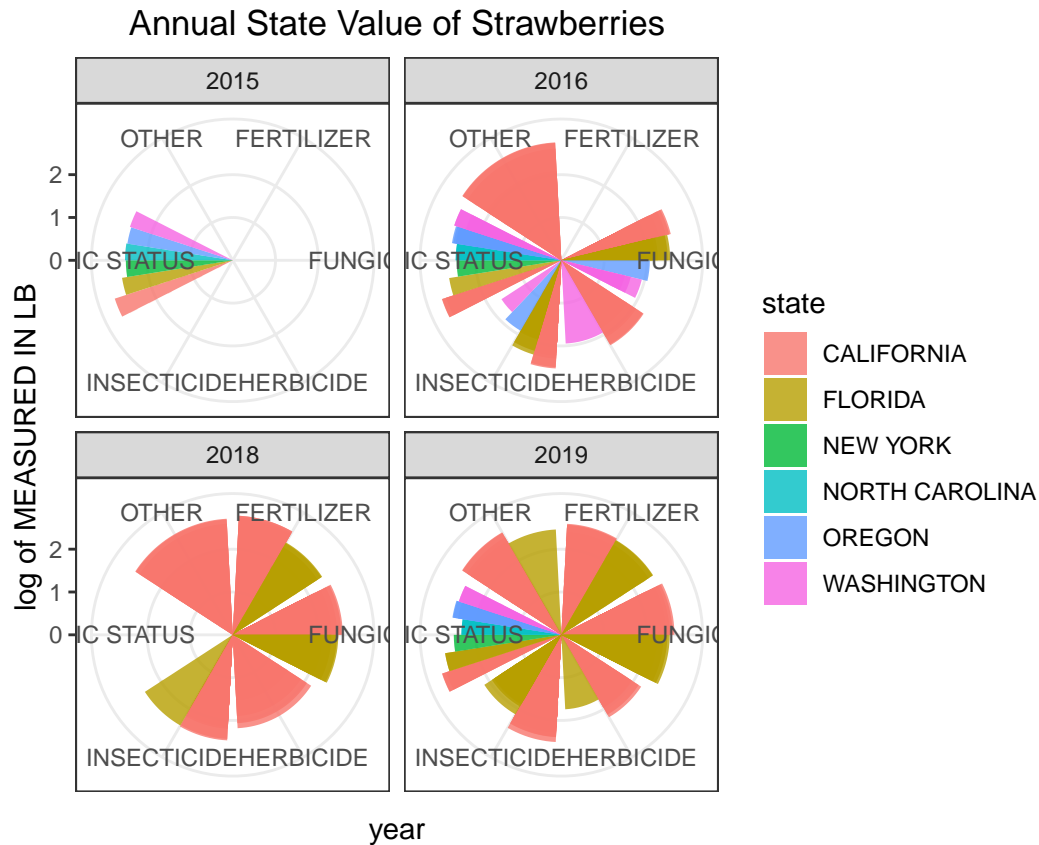
U.S. State Map of Strawberry Value



Annual state value of strawberries

As you can see in the plot below, California and Florida are the two states where USDA collected large amount of data in each type of chemicals. We also noticed that California and Florida increasingly used all kinds of chemicals in recent years.

```
plot1("MEASURED IN LB")
```



Data wrangling for Strawberries and Pesticides

- drop empty rows/columns, remove white space
- rename colname of Pesticide to chemical in order to match the colname in strawberries data
- use toupper() to capitalize all chemical names
- use pivot_longer() to make all toxins and levels into one column
- use inner_join() to wrangle Pesticide and Strawberry dataset

Questions

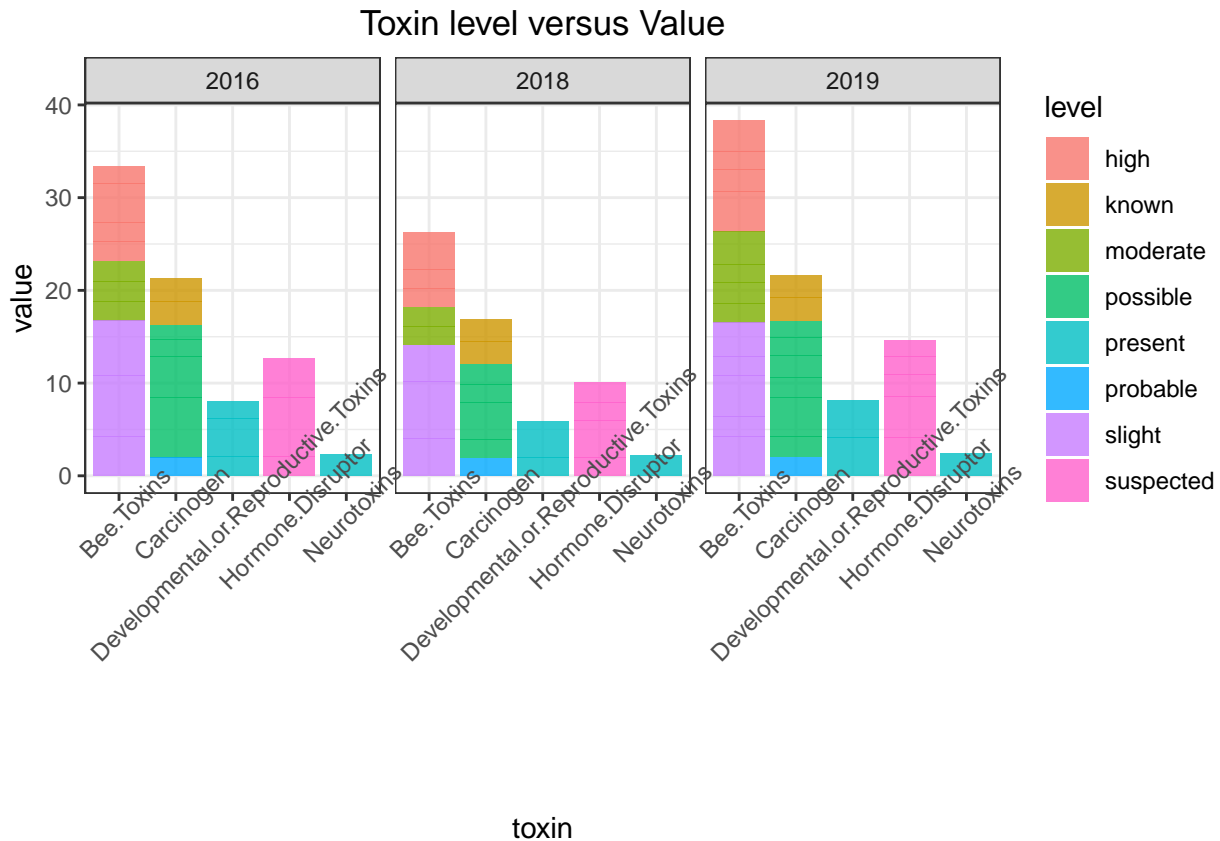
After reviewing the summary of our dataset, we came up with following questions:

- Which toxin has higher strawberry production/sale value?
- Which type of chemical is commonly related to toxicity?

Strawberries and Chemical Toxins

Here is our first approach of our question. After wrangled USDA Strawberries dataset with Pesticide dataset, we tried to figure out how chemicals can impact the production or sales of strawberries. The plot below shows the bee toxins related chemicals are proportioned to larger strawberries production values which directly answered our second question.

```
plot3("MEASURED IN LB")
```

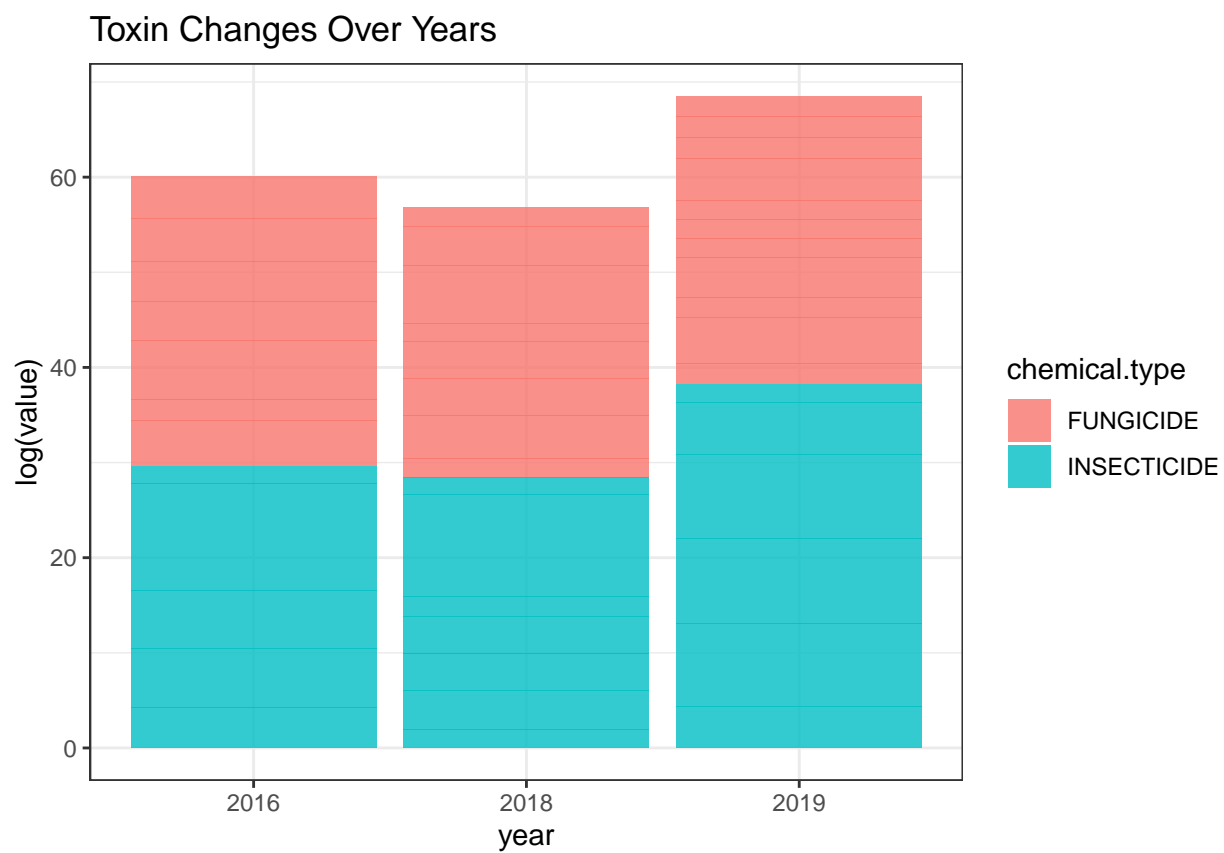


Comparing fungicide and insecticide related values

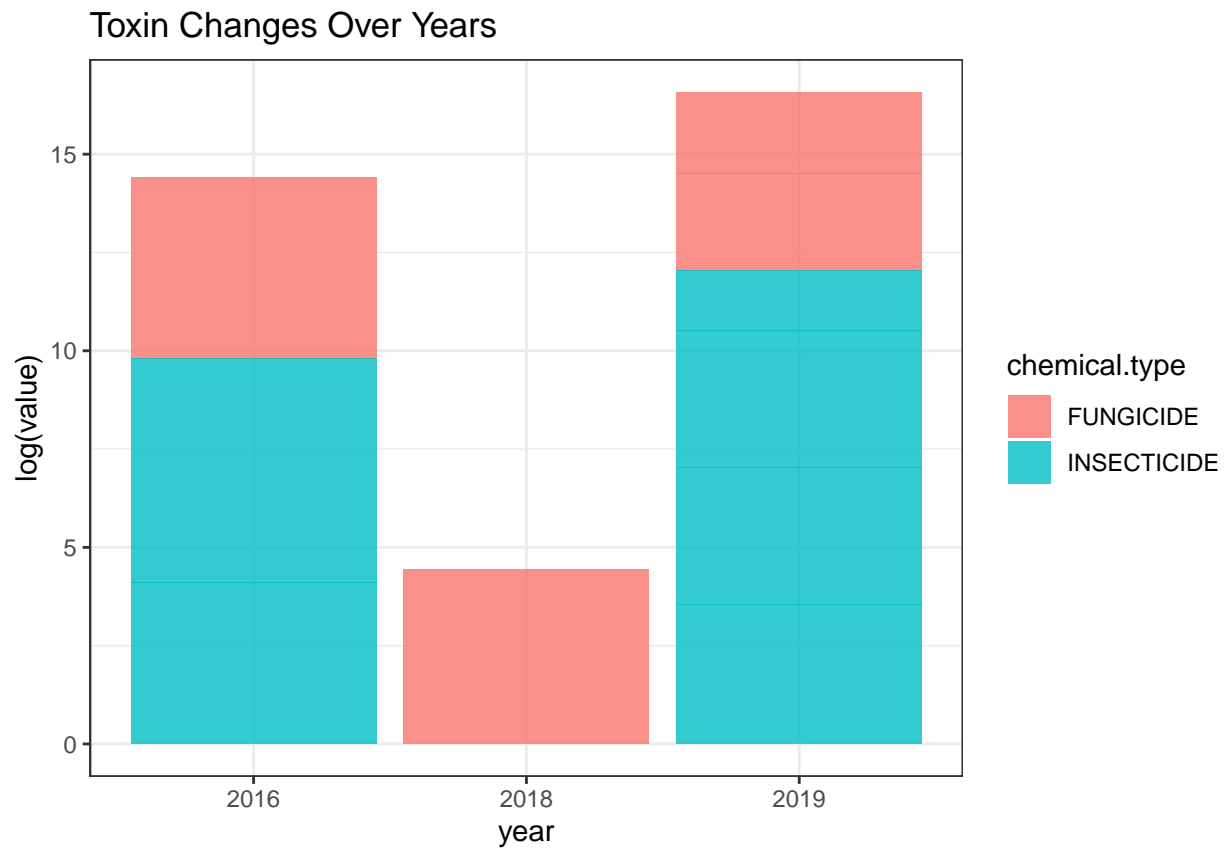
After looking carefully at our data, we realized that only insecticide and fungicide are related to toxins, and there are only three states related to pesticide toxins which are California, Florida and Washington. Since both California and Florida have more data available, we will focus on those two states only.

Therefore, the below plots show how each chemical type is proportioned to the strawberry production/sale value on log scale. In both human and bee toxins, insecticide and fungicide seem to have even proportion in strawberry production values in California, but we need further analysis.

```
plot4("MEASURED IN LB", "CALIFORNIA")
```



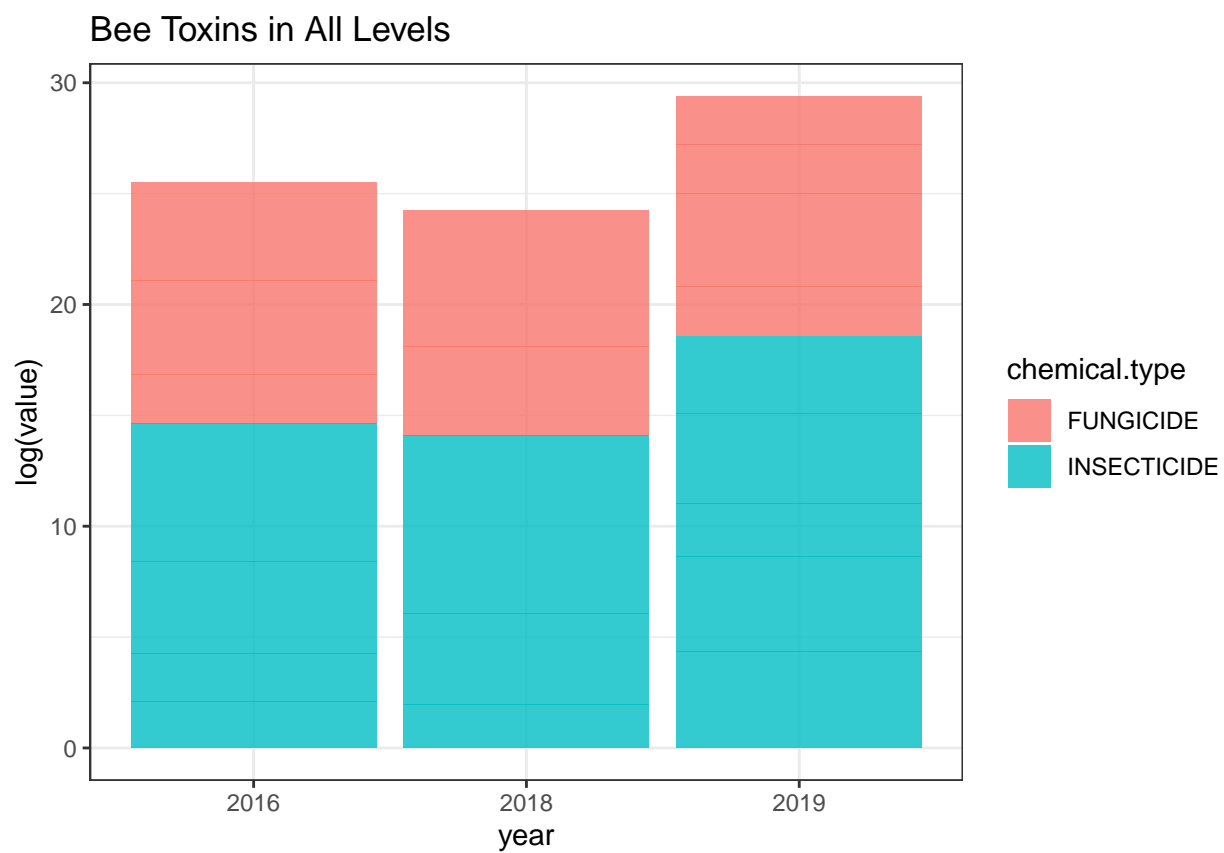
```
plot4("MEASURED IN LB","FLORIDA")
```



Bee toxins

- But looking solely at bee toxins, insecticide chemicals have higher proportion in both California and Florida. Therefore, we can answer our third question that insecticide is more commonly related to toxicity.

```
plot5("MEASURED IN LB","Bee.Toxins","CALIFORNIA")
```



```
plot5("MEASURED IN LB","Bee.Toxins","FLORIDA")
```



Conclusion

- California and Florida have lower production but more data are collected from those two states according to shiny display.
- Bee toxins are related to higher strawberry production values than human toxins.
- Insecticide is more commonly related to toxicity than fungicide.

Thanks

- Professor Haviland
- TA Bruce
- Our lovely MA-615 Classmates
- Our teammates

Citations

- <https://www.r-graph-gallery.com>
- <https://r-lang.com>
- <https://stackoverflow.com>
- <https://shiny.rstudio.com/tutorial>