Midterm_Presentation

Mi Zhang, Peng Liu, Yuanming Leng, Qiannan Shen

11/2/2021

Introduction

Our strawberry 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 Strawberry

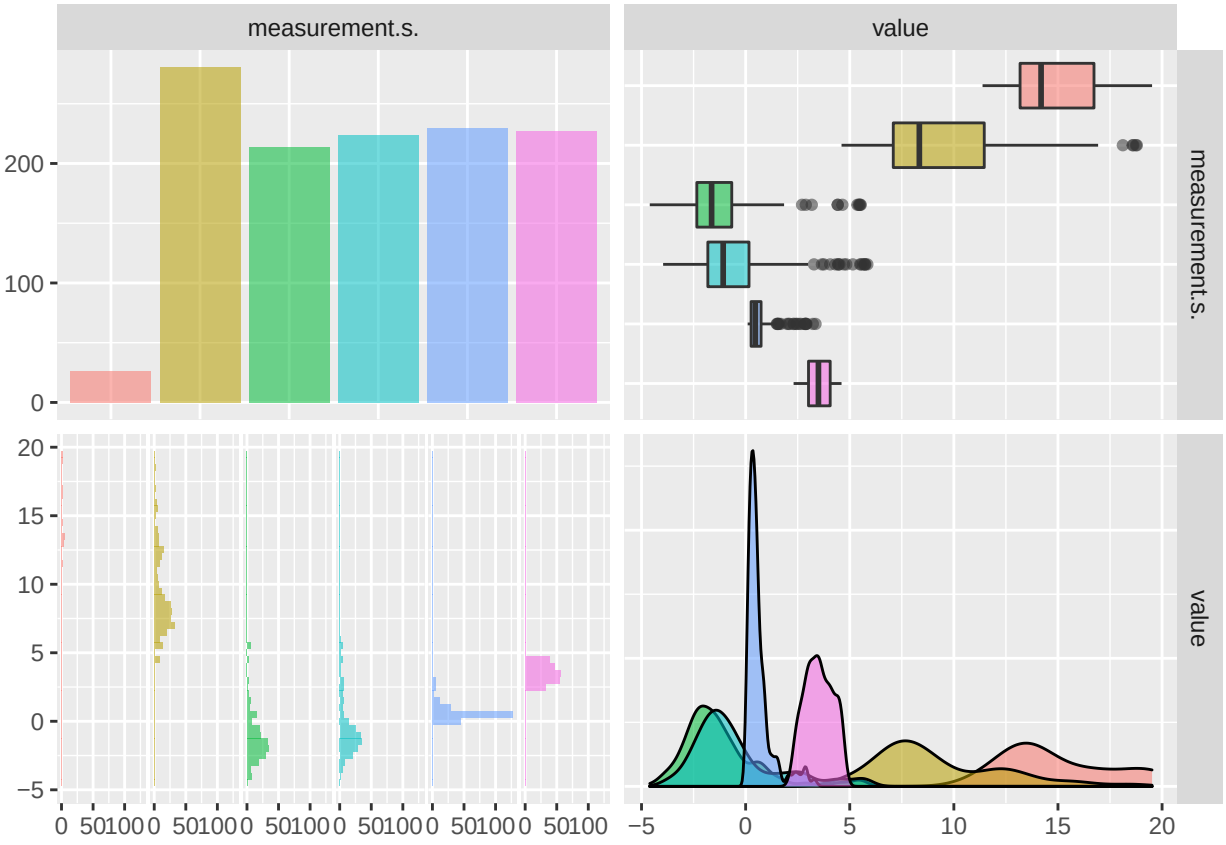
- remove empty/missing values and reduce white space in the cells
- split column with multiple items to separated columns
- redefine "MEASURED IN CWT" by multiplying by 100
- make extreme large value more accessible by using log scale on "value"

Data wrangling for Strawberry and Pesticides

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

Measurement units and value

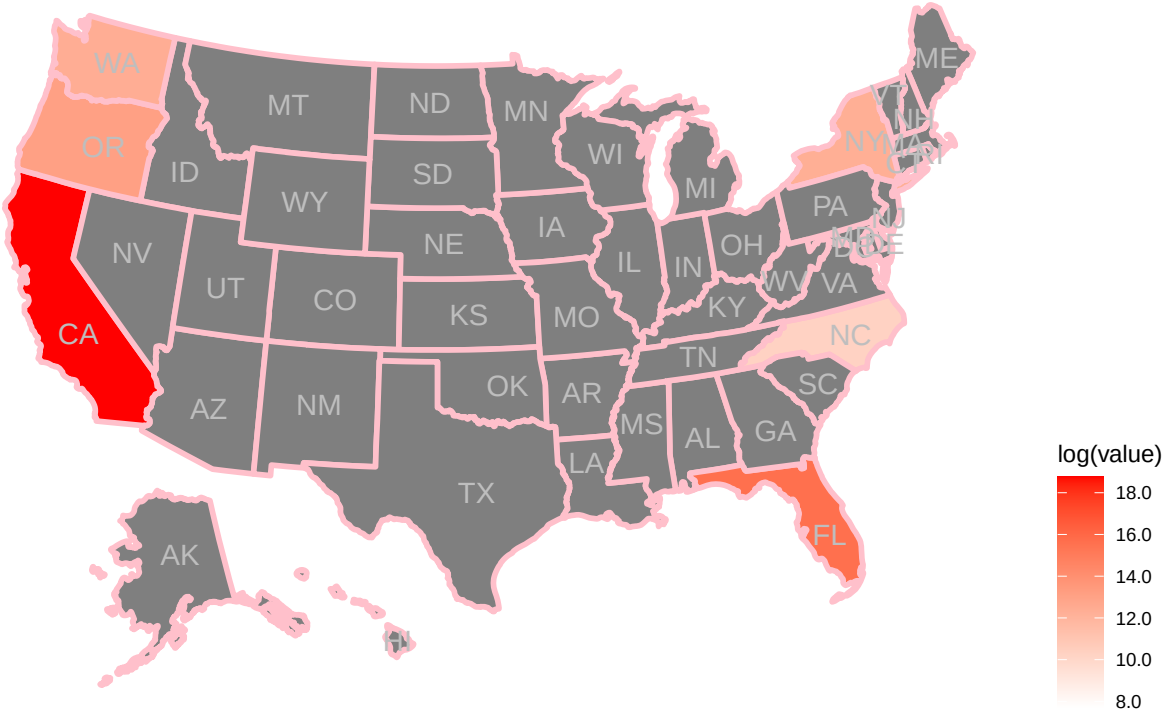
- Our major analysis is base on the measurement units, and our main focus is on “MEASURE IN LB”
- redefine value as production of strawberry
- You can also explore our shiny app here (https://lemony.shinyapps.io/ma615-midterm/)



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

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

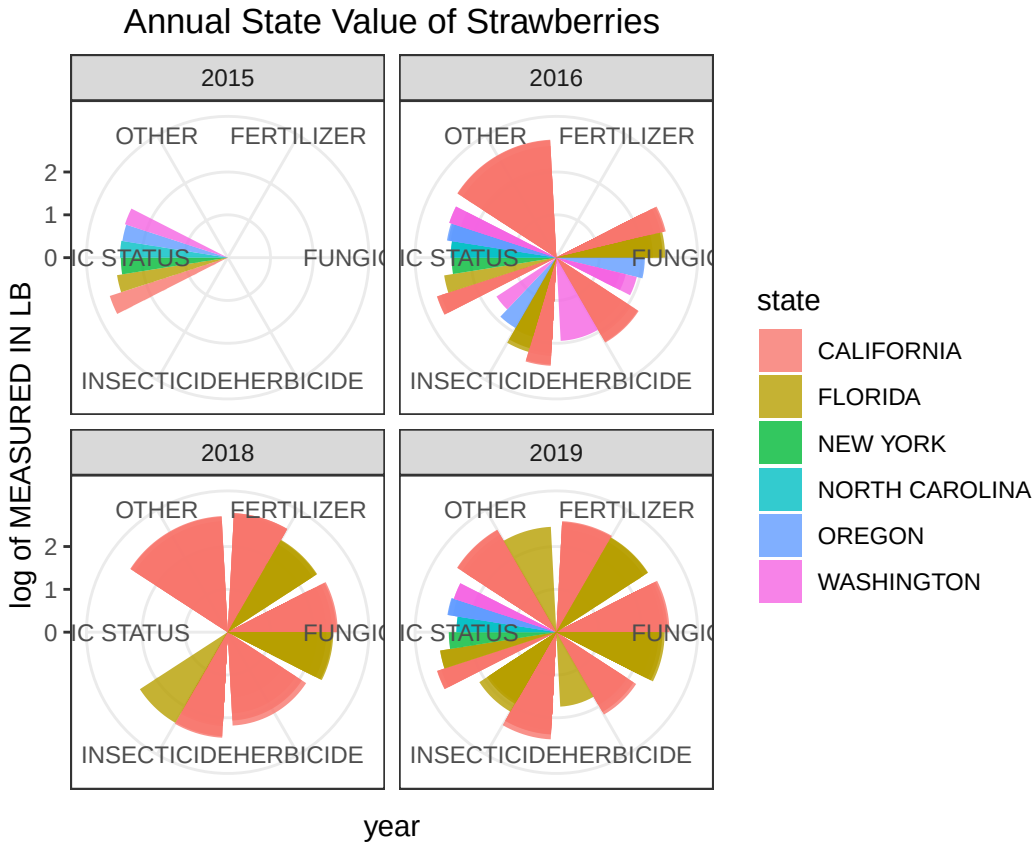
U.S. State Map of Strawberry Value



Annual state value of strawberry

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")
```



Questions

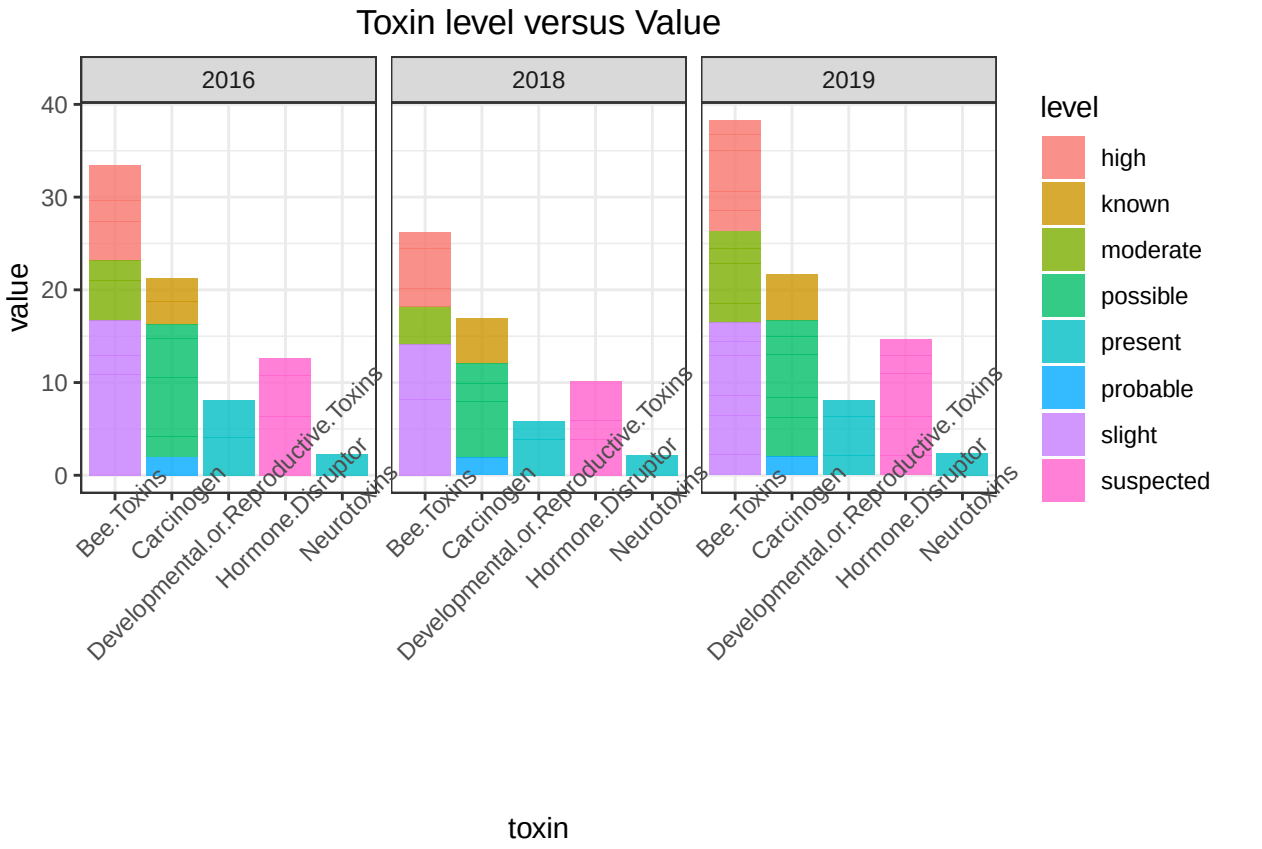
After reviewing the summary of our dataset, we came up with following questions:

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

Strawberry and chemical Toxins

Here is our first approach of our question. After wrangled USDA strawberry dataset with Pesticide dataset, we tried to figure out how chemicals can impact the production or sales of strawberry. The plot below shows the bee toxins related chemicals are proportioned to larger strawberry 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 pestide toxins which are California, Florida and Washington. Since both California and Florida have more data avaible, we will focus on those two states only.

Therefore, the below plots shows how each chemical type are proportioned to the strawberry production/sale value on log scale. In the total 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")
```

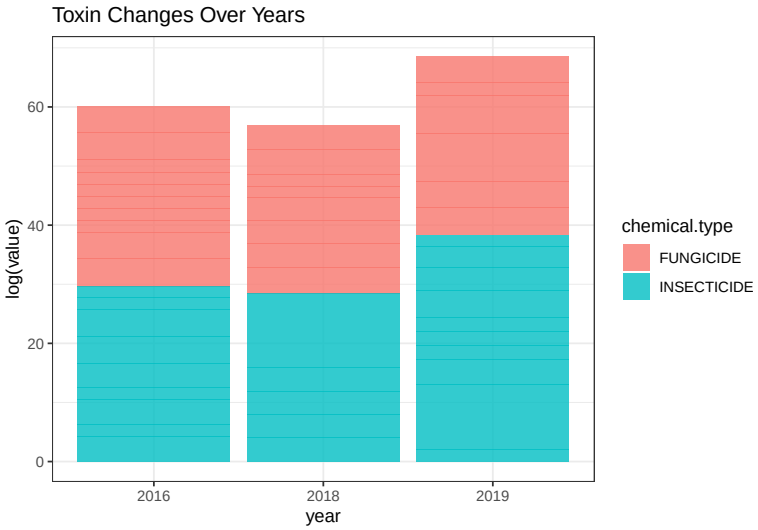


Figure 1: California

For the Florida plot, we can clarify find that the insecticide is much higher than fungicide, though we don't have date of insecticide in 2018.

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

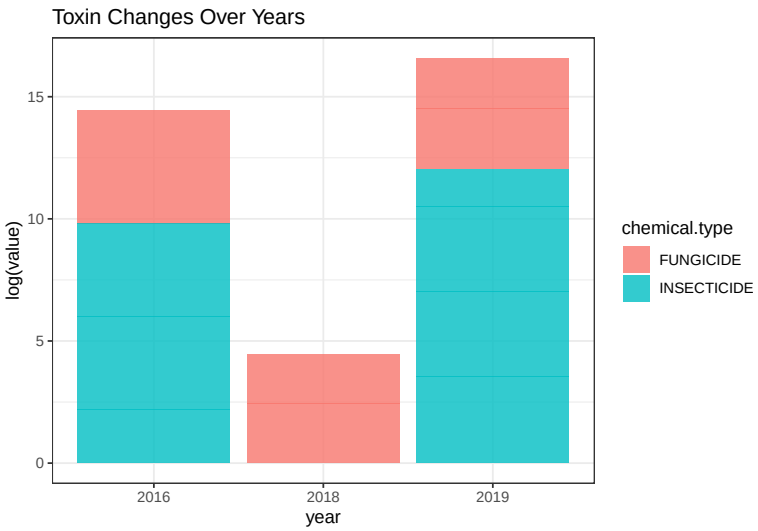


Figure 2: Florida

Bee toxins

When we focus on bee toxin, we can find that the chemical type is different than the total toxin which we see above in California. Insecticide is more higher than fungicide.

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

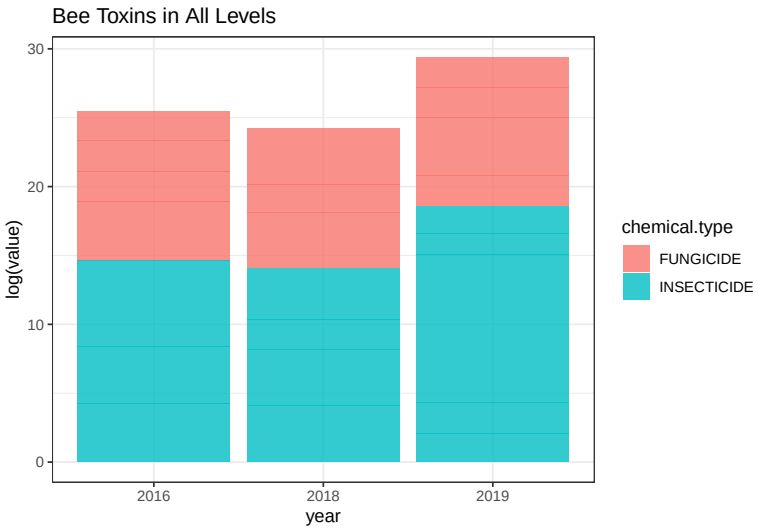


Figure 3: California

But for Florida, the bee toxin shows same result as total toxin.

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

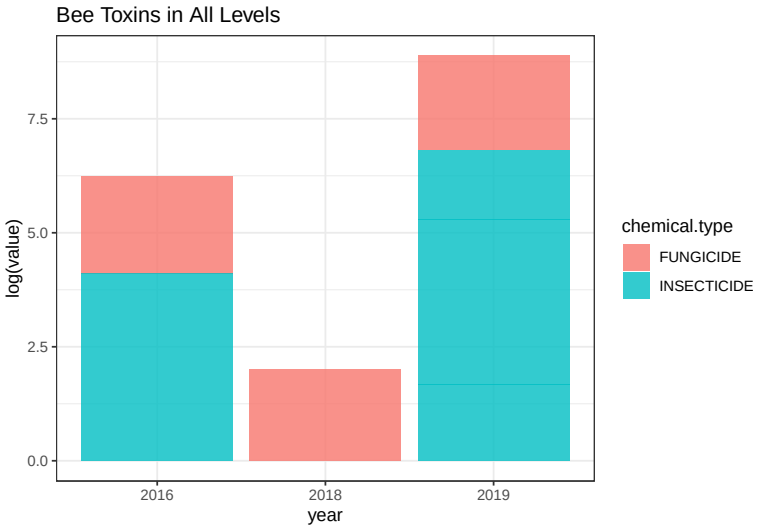


Figure 4: Florida

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

Limitations

- Missing values are the most common thing when we deal with data cleaning, processing, and wrangling;
- Data size shrank after wrangling (not able to match all chemicals);
- We don't know how chemical usages are related to strawberry production, are they harmful or beneficial to strawberry production.
- We did not characterize toxin levels into numeric levels because they are in natural language and people have different perspectives defining them.

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 other types of toxins.
- Insecticide is more commonly related to toxicity than fungicide on strawberry production.

Thanks

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

Citations

- Haviland's lectures
- https://www.r-graph-gallery.com
- https://r-lang.com
- https://stackoverflow.com
- https://shiny.rstudio.com/tutorial