Strawberry, Status, State, Toxicity

Group 1: Will, Zhihui, Shicong, Tao He

11/1/2021

Abstract

In this project, our focus is on strawberry data set that comes from US Department of Agriculture National Agricultural Statistics Service. We combine this data set with the toxicity levels of the chemical they used. After cleaning and wrangling the data, we do the exploratory data analysis to explore the data and address our question - whether states have preferences when applying pesticides on strawberry farming.

Method

The project consists of three parts:

1. Data wrangling and cleaning
2. Exploratory data analysis
3. Shiny App

Data Cleaning and Wrangling

Basically, for data cleaning and wrangling, we divide into three parts. First, we clean the two data sets separately. Then, we do the normalization to make sure that the common column in two data sets is encoding in the same way. Finally, we inner join the two data sets to generate the final data we will use in our further analysis. And below are the details about how we tidy up our data.

Strawberry data set

1. Remove columns without any information.
2. Split columns that contain multiple values
3. Use 0 to fill up the missing value in value column
4. Substitute space in strings with underscore and uppercase the chemical names
5. Rename the columns to make them readable

Pesticide data set

1. Add pesticides that are in the strawberry data set but not in the pesticide data set
2. Substitute space in strings with underscore and uppercase the chemical names
3. Fill up missing value as unknown

Date Frame

column names	description
chemical	different chemical brands
chemical type	different types of chemicals
toxicity-human/bee	whether the chemical has a toxicity to humans or bees
toxicity level	“HIGH”, “MODERATE”, “SLIGHT”, “UNKNOWN”...
state	the states who plant strawberries
year	from 2016 to 2019
measurements	by using different units, like weight, sales,...

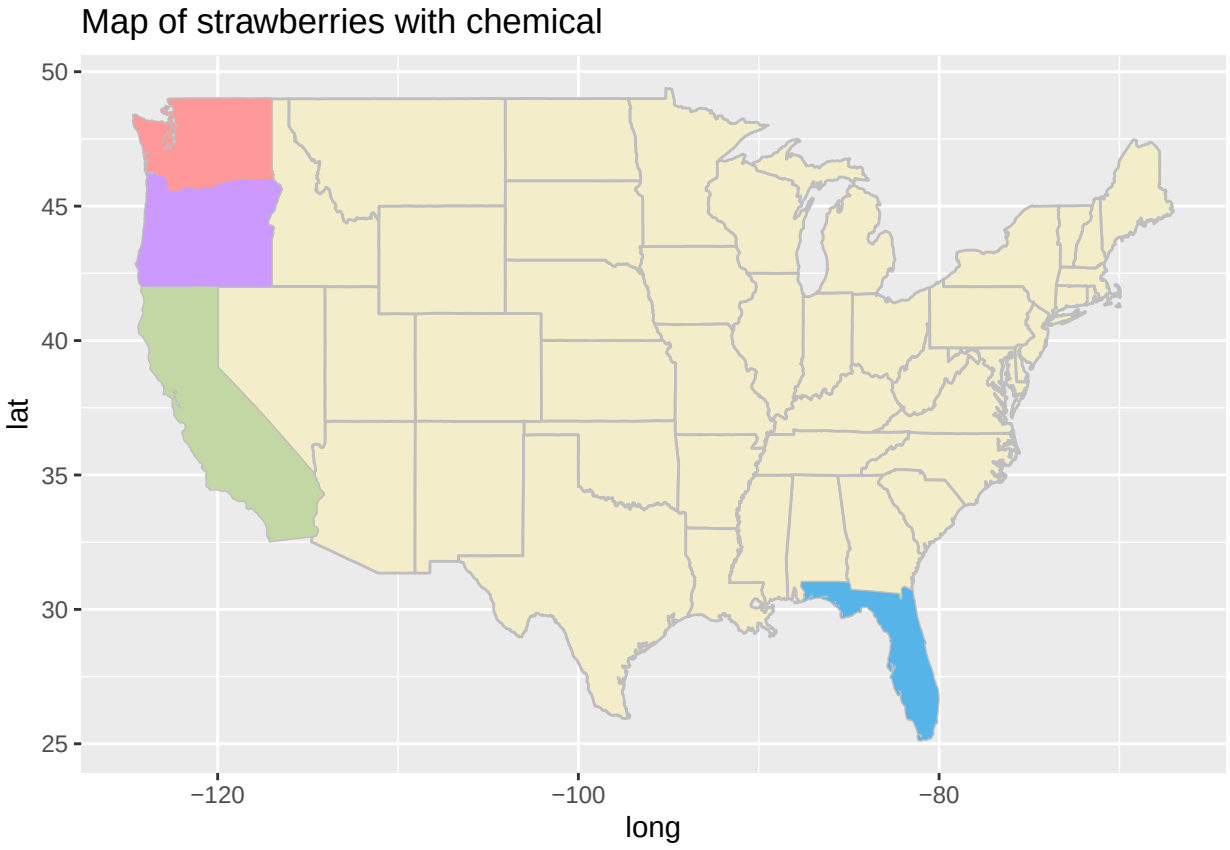
Note that measurements can be TREATED, MEASURED IN PCT OF AREA BEARING; APPLICATIONS, MEASURED IN LB; APPLICATIONS, MEASURED IN LB / ACRE / APPLICATION; APPLICATIONS, MEASURED IN NUMBER or APPLICATIONS, MEASURED IN LB / ACRE / YEAR.

Exploratory Data Analysis

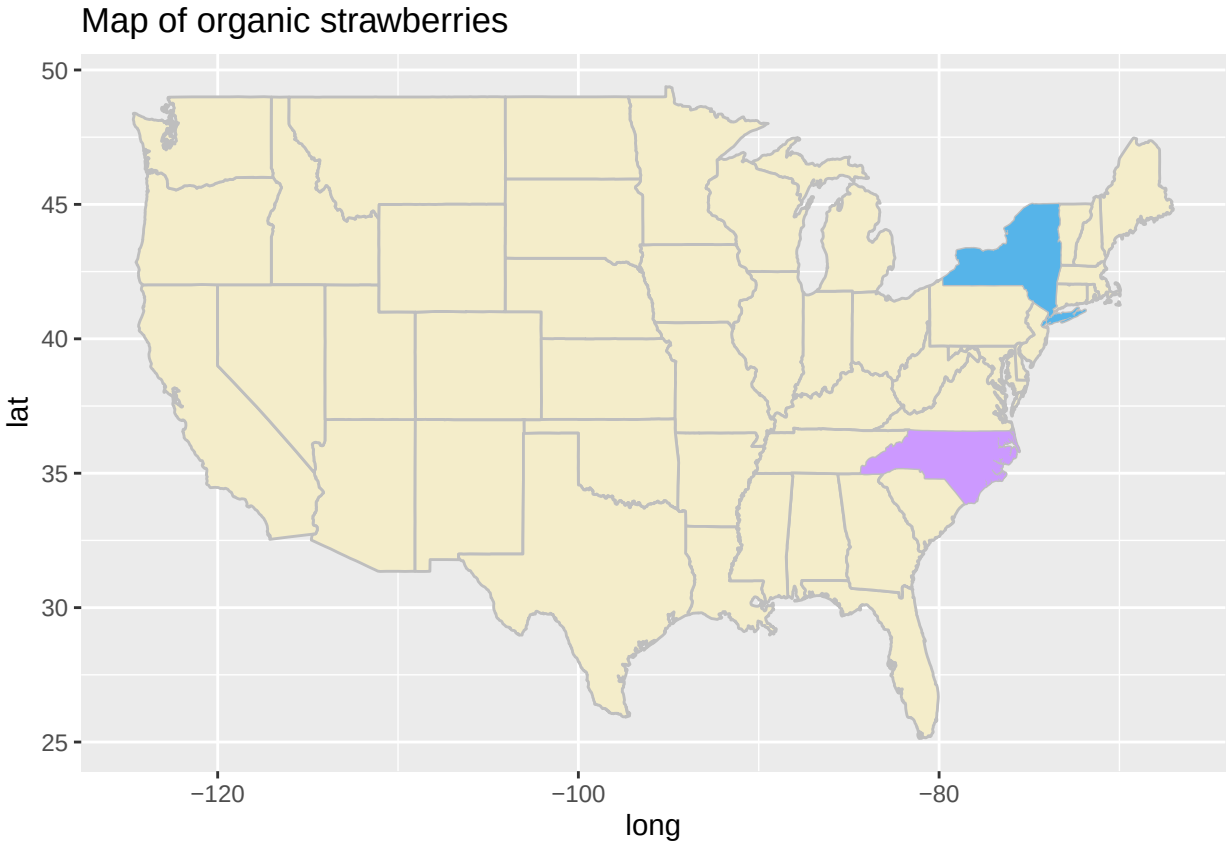
Organic and Chemical

First, we look at the organic strawberries versus strawberries with chemicals. We find that four states used chemicals in strawberry cultivation, while two states did not use them. We put our emphasis on the strawberries with chemicals in our further analysis.

Map: Straberry with chemical

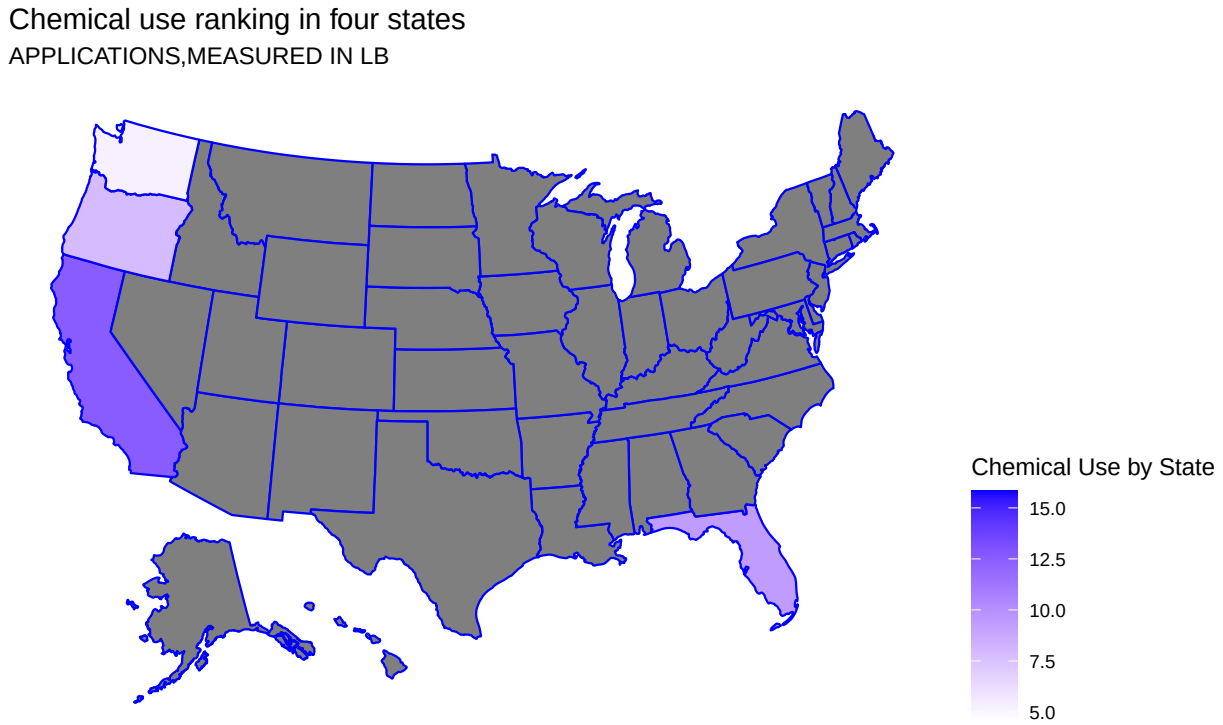


Organic Strawberry

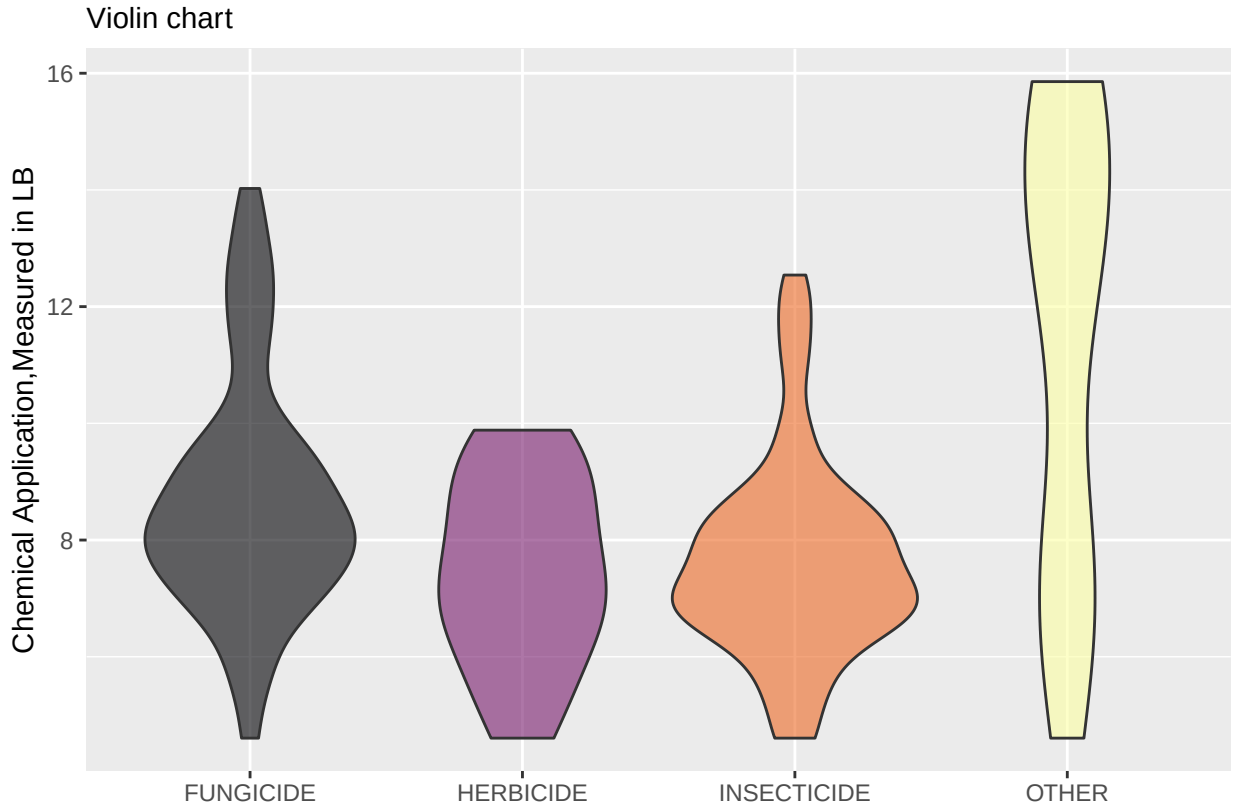


The Usage of Chemical

We measure the usage of chemistry in two dimensions: the difference of the dosage across states, and the difference of the dosage among chemical types. We choose the chemical application measured in LB in the data set. For the comparison of chemical dosages between different states, we compare them by the shade of color. The darker color represents more chemical usage.

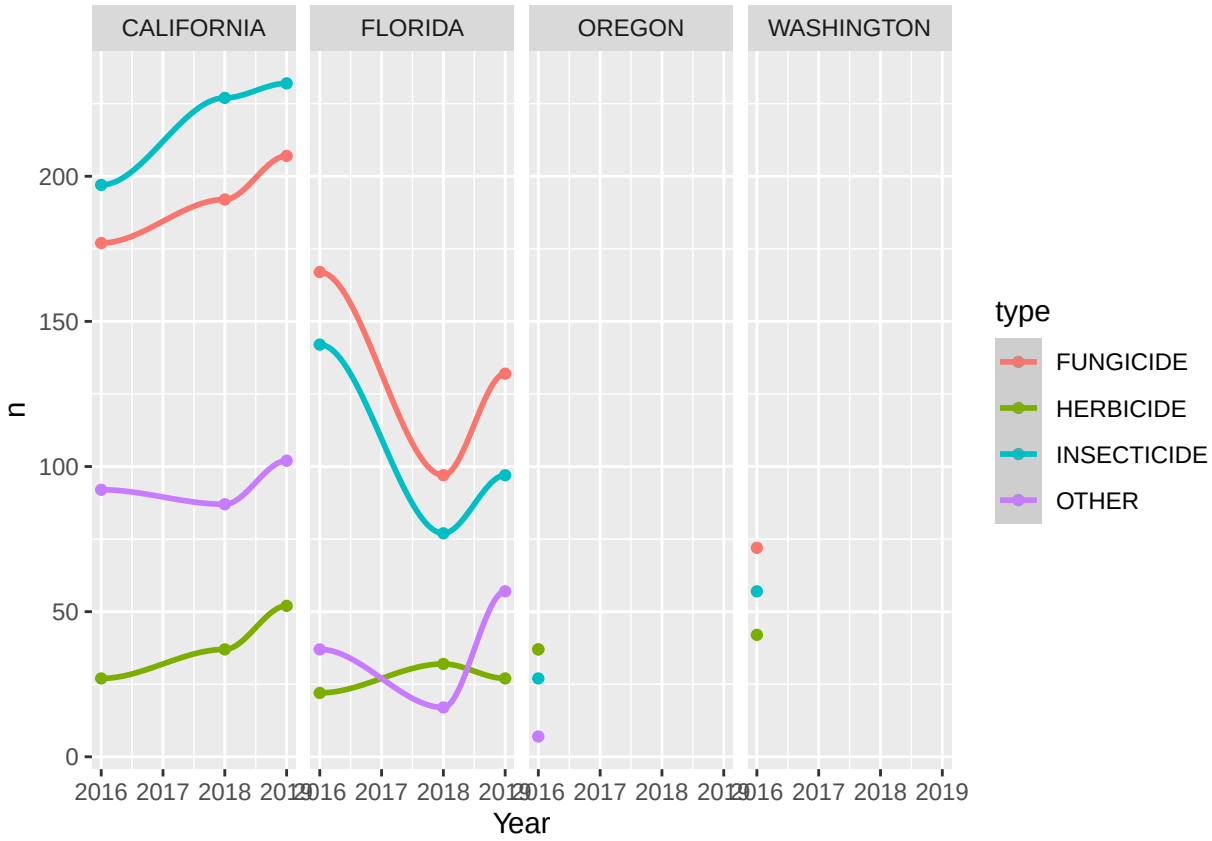


Chemical Usage By Chemical Types



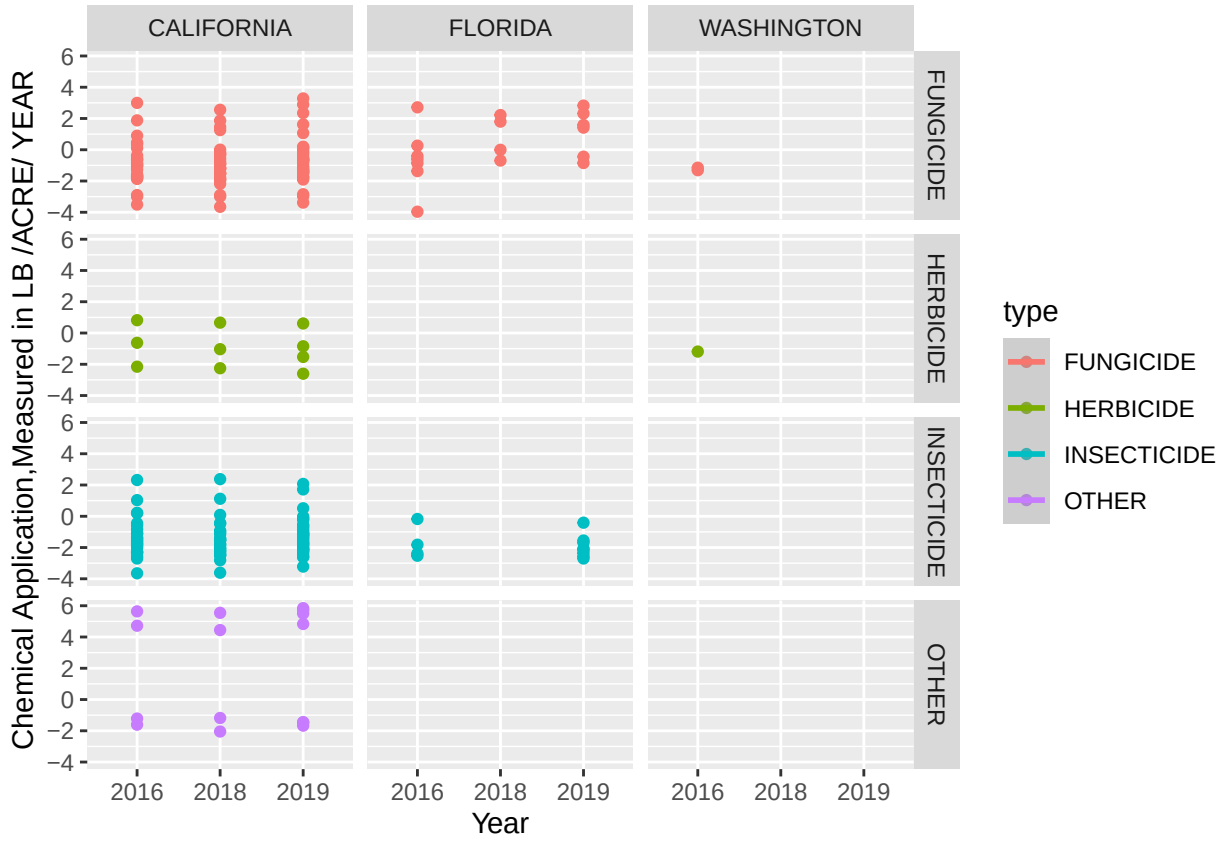
We use violin plots to witness the distribution of usage amounts when comparing different chemical types. The wider shape shows more points are distributed in this range. And the higher shape in the plot represents the range of the data points is larger.

Chemical Usage By Types In States



To address our question - whether it exists preference on the type of chemical among states, we combine the two dimensions of state and type in the plot blow. As the plot shows, the preference does exist.

Chemical Usage in Three Dimensions

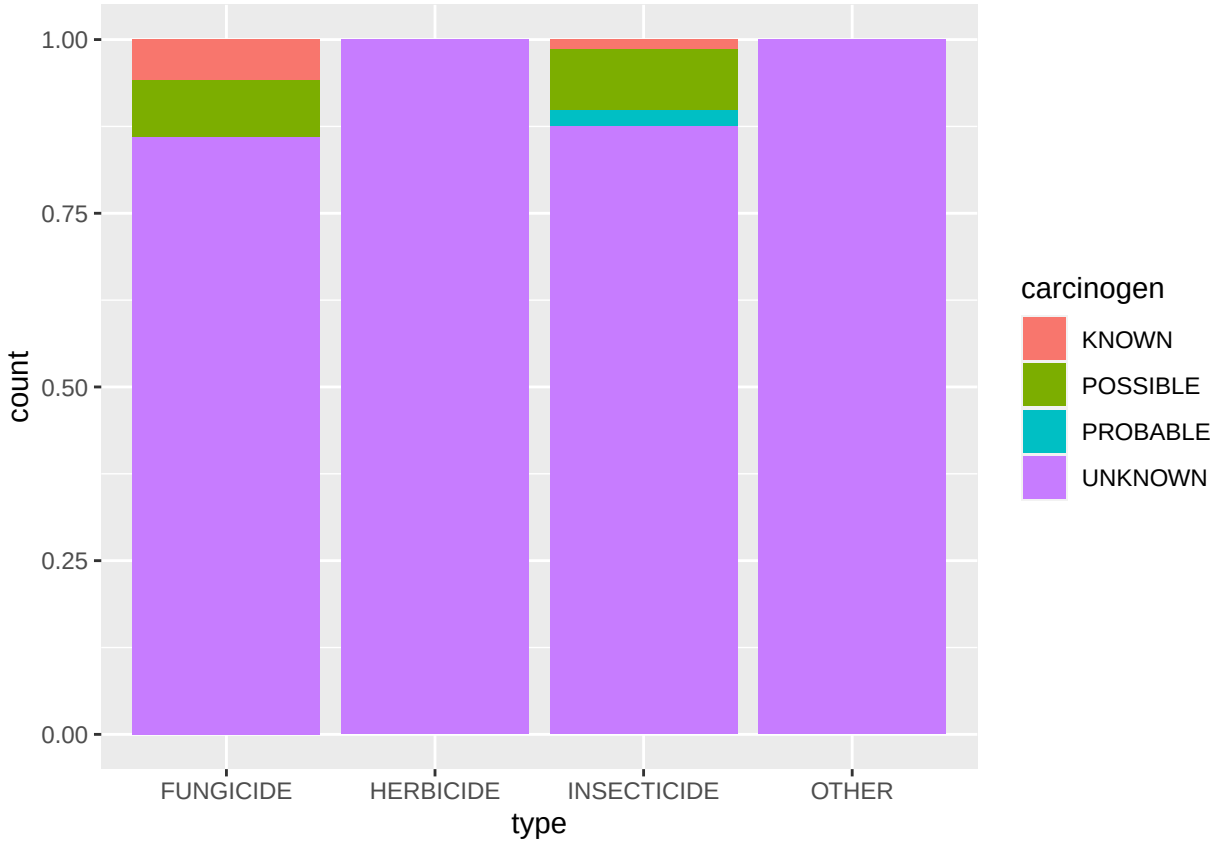


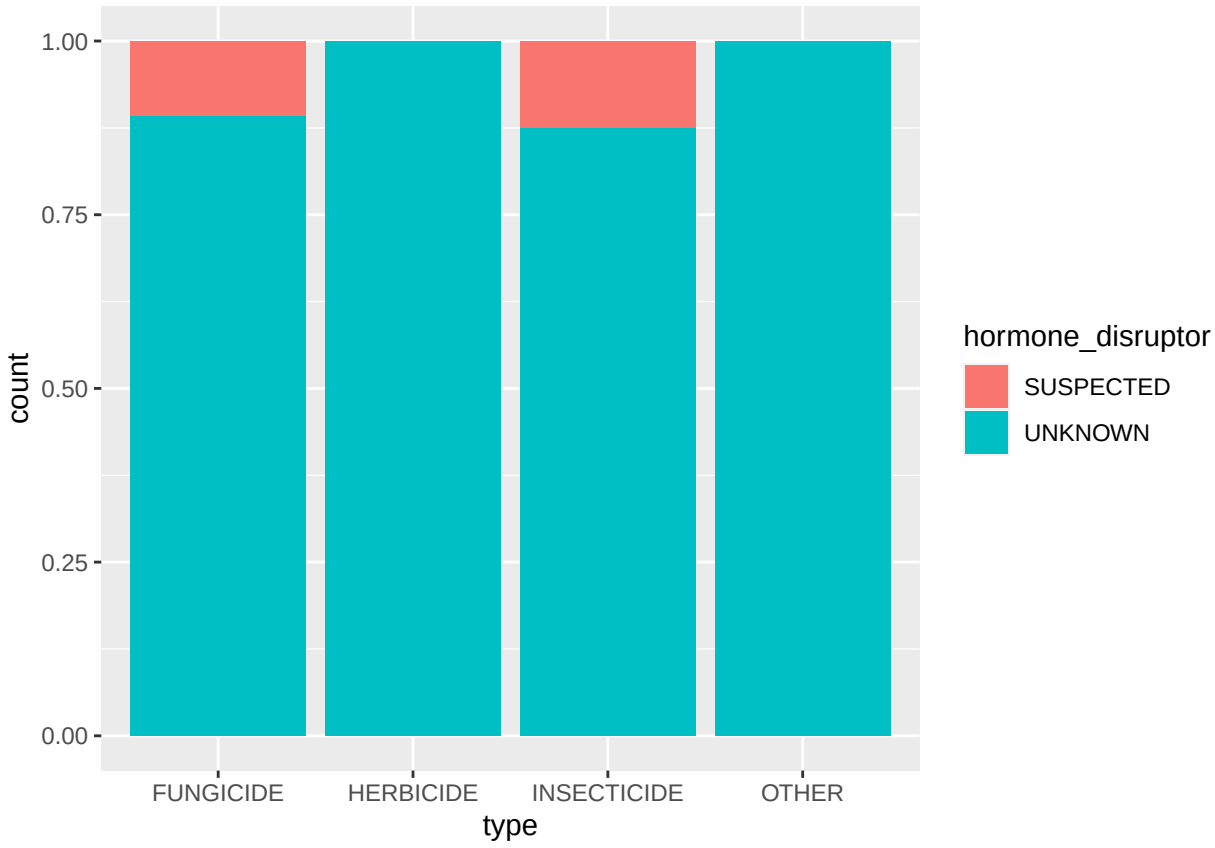
Then we combine three dimensions(time, states, chemical types) into one plot. When we compare horizontally, we can see the chemical usage between different states.When comparing vertically, we can see different chemical dosages in the same state. We also find that the annual chemical usage in each state remains stable based on the data we have. Besides,to show the plot in a more intuitively way, we rescale the y axis by log10.

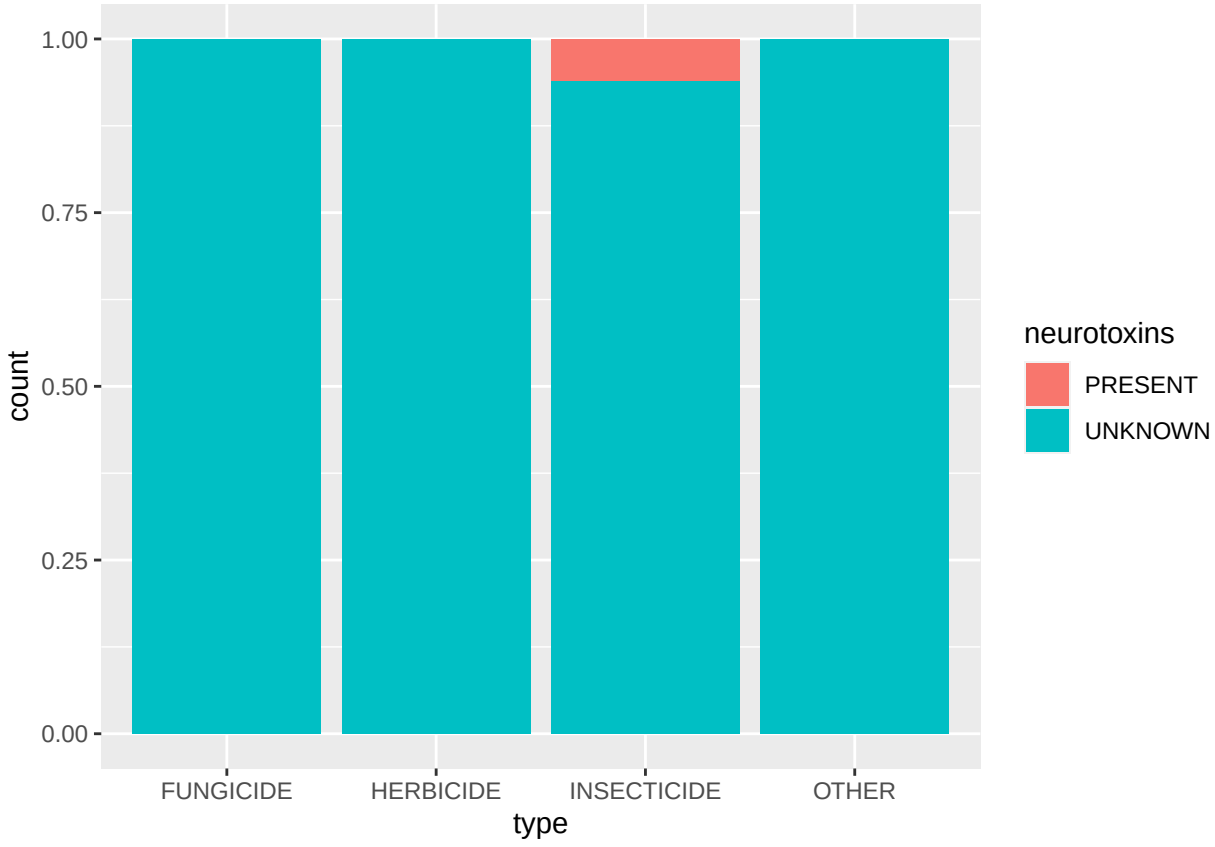
Toxicity

After we observed the preference for each state, we ask ourselves what are factors might influence the choice of each state? We used five bar charts to explore whether people prefer to use the more toxic chemicals or the less toxic chemicals on different types of chemicals? Besides, we measured the toxicity level into two parts, one is for humans, the other is for bees. The toxicity to humans can be measured by four dimensions, which is Carcinogen, Hormone Disruption, Neurotoxins, Developmental Or Reproductive Toxins, respectively.

The Toxicity to Humans




1.00
0.75
0.50
0.25
0.00
count
FUNGICIDE
HERBICIDE
INSECTICIDE
OTHER
type
hormone_disruptor
SUSPECTED
UNKNOWN


1.00
0.75
0.50
0.25
0.00
count
FUNGICIDE
HERBICIDE
INSECTICIDE
OTHER
type
neurotoxins
PRESENT
UNKNOWN

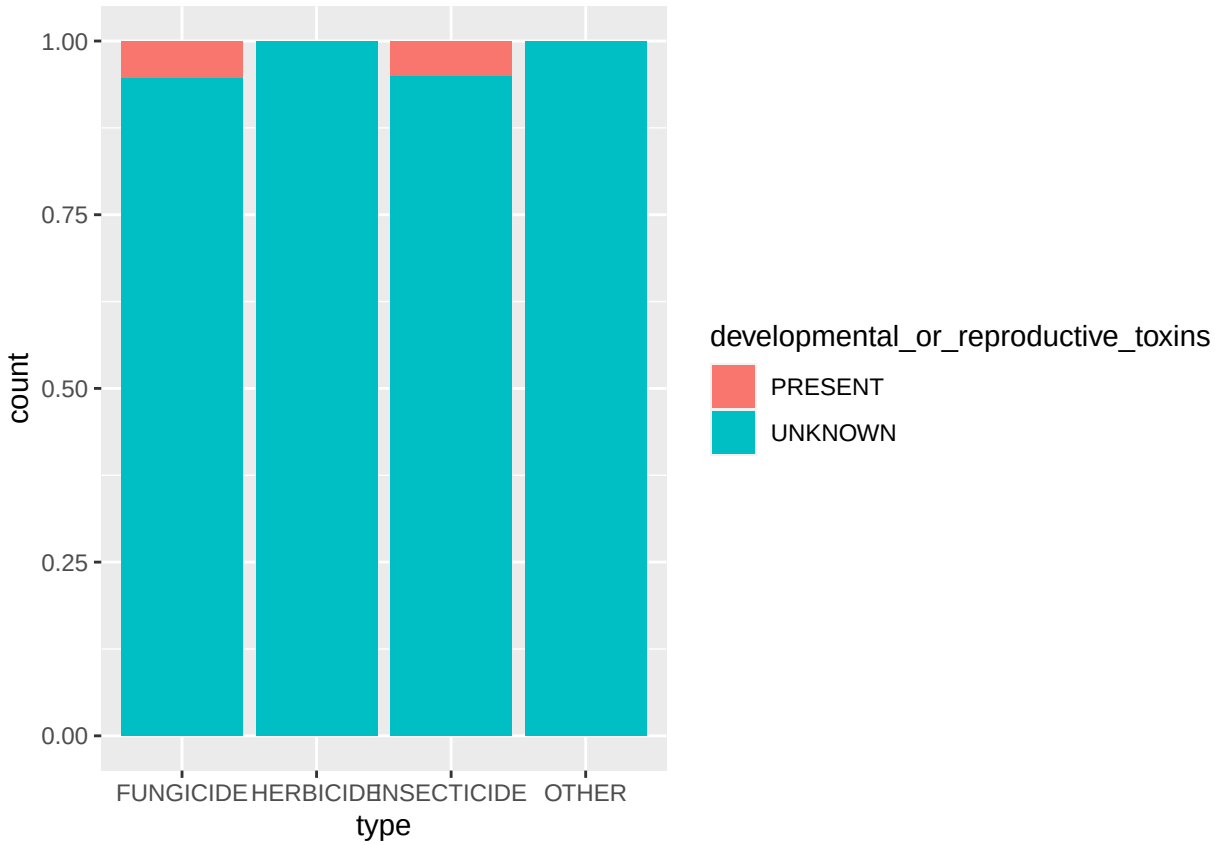


Figure.toxicity of Carcinogen, the majority area is the missing data, so we can only focus on the fungicides and the insecticides. On these two types of chemicals, farmers are more likely to use possible toxic.

Figure.toxicity of Hormone Disruption, Neurotoxins and Developmental Or Reproductive Toxins, the lake blue area is unknown data, we can only have some information about the type of fungicides and insecticides. If we take more data, we might know the frequency of that the farmers used which toxicity level of the chemicals.

The Toxicity to Bees

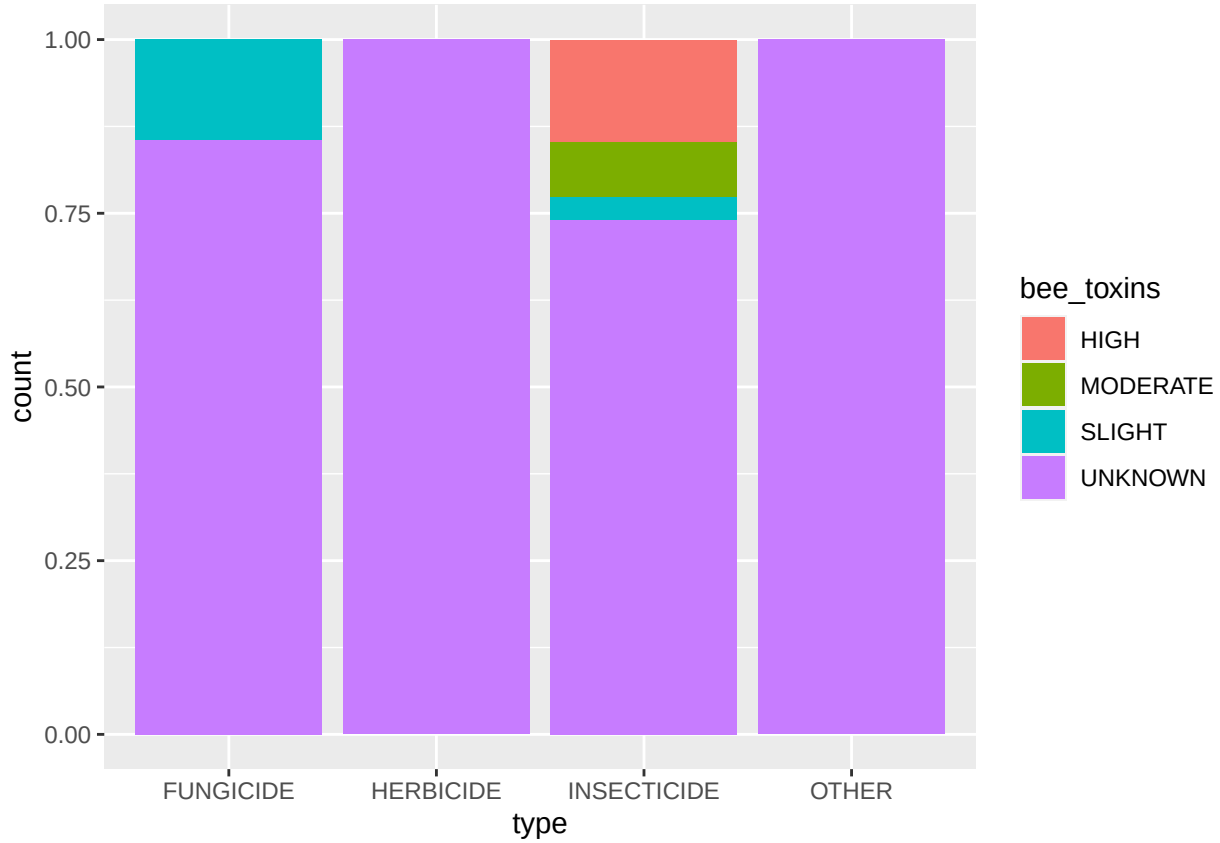
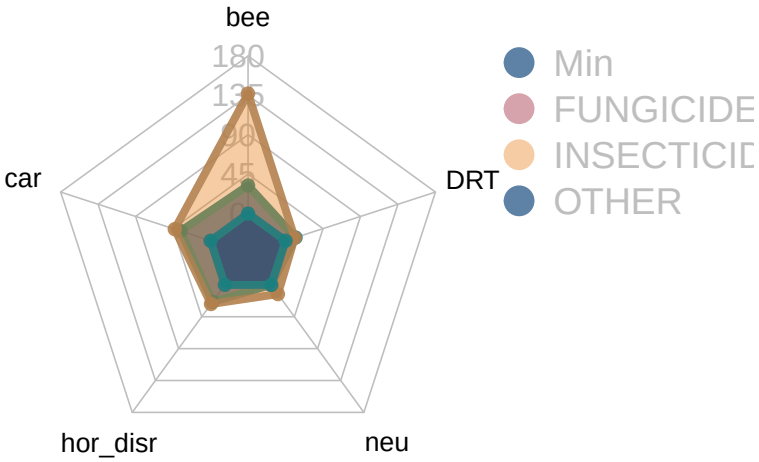


Figure.toxicity to bees, we can easily observe that when the farmers use the fungicide, they would like to use the slightly toxic chemicals, and when the farmers use the insecticides, they prefer to use highly toxic chemicals.

The Whole Toxicity Level



The radar plot below whose area represents the total toxicity level. As the plot shows, insecticides have the largest area. We may think that when planting strawberries, farmers tend to use more toxic insecticides.

Shiny App

Here is the link for our shiny dashboard: link

Result

The order of most chemicals used is California, Florida, Oregon, Washington. We also see that more toxic chemicals are preferred when considering specific type of chemicals or state.

Limitations and Improvements

There is a lot of missing data. And we are not sure if there is a bias when collecting the data. Therefore we could not make too much inference about chemicals in strawberry based on the data we have. We would like the get more data for future exploration if possible. According to our exploratory analysis process, we find that farmers may not always choose harmless or less harmful pesticides. Thus, based on the existing data, we can only say that the choice of type of chemical and its toxicity level might be varied by states. What's more, we may focus on the chemical usage in California further as it is one of the largest strawberries produce in United States.

References

- Leaflet for R
- Shiny Dashboard
- Lecture notes from MA 615
- Data comes from the US Department of Agriculture National Agricultural Statistics Service