# 615 strawberry report

Chi Zhang

2020/10/18

# 1.Data cleaning and orgnization:

## Acquire and read the data

These data were collected from the USDA database selector: https://quickstats.nass.usda.gov (https://quickstats.nass.usda.gov)

The data were stored online (https://quickstats.nass.usda.gov/results/D416E96E-3D5C-324C-9334-1D38DF88FFF1) and then downloaded as a CSV file.

First of all, we read the data, and make a list of columns with one unique value and list its names.

Now we are going to clean the redundant and meaningless data of the original data set. Removing the 1 unique columns then state name and ANSI are redundant.

| Year | Period | State | Commodity | Data Item | Domain | Domain Category | Value |
|---|---|---|---|---|---|---|---|
| 2019 | MARKETING YEAR | CALIFORNIA | BLUEBERRIES | BLUEBERRIES, TAME - PRICE RECEIVED, MEASURED IN $ / LB | TOTAL | NOT SPECIFIED | 2.85 |
| 2019 | MARKETING YEAR | CALIFORNIA | BLUEBERRIES | BLUEBERRIES, TAME, FRESH MARKET - PRICE RECEIVED, MEASURED IN $ / LB | TOTAL | NOT SPECIFIED | 3.56 |
| 2019 | MARKETING YEAR | CALIFORNIA | BLUEBERRIES | BLUEBERRIES, TAME, PROCESSING - PRICE RECEIVED, MEASURED IN $ / LB | TOTAL | NOT SPECIFIED | 0.29 |
| 2019 | MARKETING YEAR | CALIFORNIA | RASPBERRIES | RASPBERRIES - PRICE RECEIVED, MEASURED IN $ / LB | TOTAL | NOT SPECIFIED | 2.69 |
| 2019 | MARKETING YEAR | CALIFORNIA | RASPBERRIES | RASPBERRIES, FRESH MARKET - PRICE RECEIVED, MEASURED IN $ / LB | TOTAL | NOT SPECIFIED | D. |
| 2019 | MARKETING YEAR | CALIFORNIA | RASPBERRIES | RASPBERRIES, PROCESSING - PRICE RECEIVED, MEASURED IN $ / LB | TOTAL | NOT SPECIFIED | D. |

1)Cleaning the data of strawpberry with time period “YEAR”.

2)Separate strawberry’s domain and domain category.

3)Eliminating the redundancy.

4)Clean it up.

Here is the cleaned strawberry data.

| Year | State | type | Measures | Materials | Chemical | Value |
|---|---|---|---|---|---|---|
| 2019 | CALIFORNIA | BEARING - APPLICATIONS | MEASURED IN LB | (AZOXYSTROBIN = 128810) | FUNGICIDE | 5,500 |
| 2019 | CALIFORNIA | BEARING - APPLICATIONS | MEASURED IN LB | (BACILLUS AMYLOLIQUEFACIENS MBI 600 = 129082) | FUNGICIDE | (NA) |
| 2019 | CALIFORNIA | BEARING - APPLICATIONS | MEASURED IN LB | (BACILLUS AMYLOLIQUEFACIENS STRAIN D747 = 16482) | FUNGICIDE | (NA) |
| 2019 | CALIFORNIA | BEARING - APPLICATIONS | MEASURED IN LB | (BACILLUS PUMILUS = 6485) | FUNGICIDE | (NA) |
| 2019 | CALIFORNIA | BEARING - APPLICATIONS | MEASURED IN LB | (BACILLUS SUBT. GB03 = 129068) | FUNGICIDE | (NA) |
| 2019 | CALIFORNIA | BEARING - APPLICATIONS | MEASURED IN LB | (BACILLUS SUBTILIS = 6479) | FUNGICIDE | (NA) |

# 2.Exploratory data analysis:

## 2.1Background

We obtained data on three berry species from USDA's website from 2015 to 2019. They are the data about the origin, field and value of raspberry, strawberry and blueberry. However, there are many columns of meaningless data as well as many blanks and duplications in the database, so I first cleaned the data, and then sorted out the useful data of strawberry for data analysis. I would like to use the variables and values to find out the relationship between them, so as to provide some help for strawberry production.

## 2.2 Variables

year- the year of the list of data

State - the place of origin

type - the present state

Measures - what kind of information measured

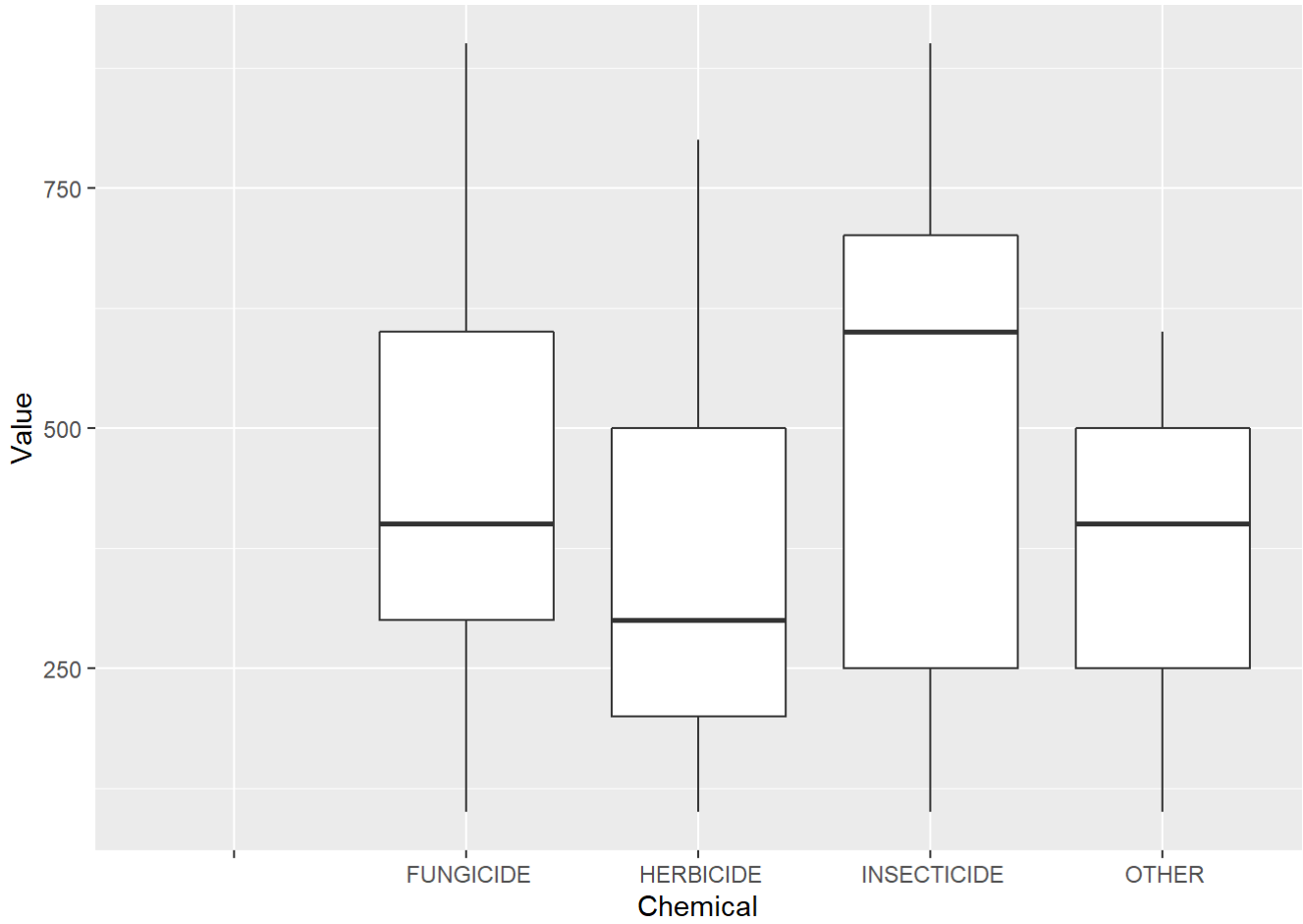
Materials - Substances in chemicals

Chemical - what type of chemicals used

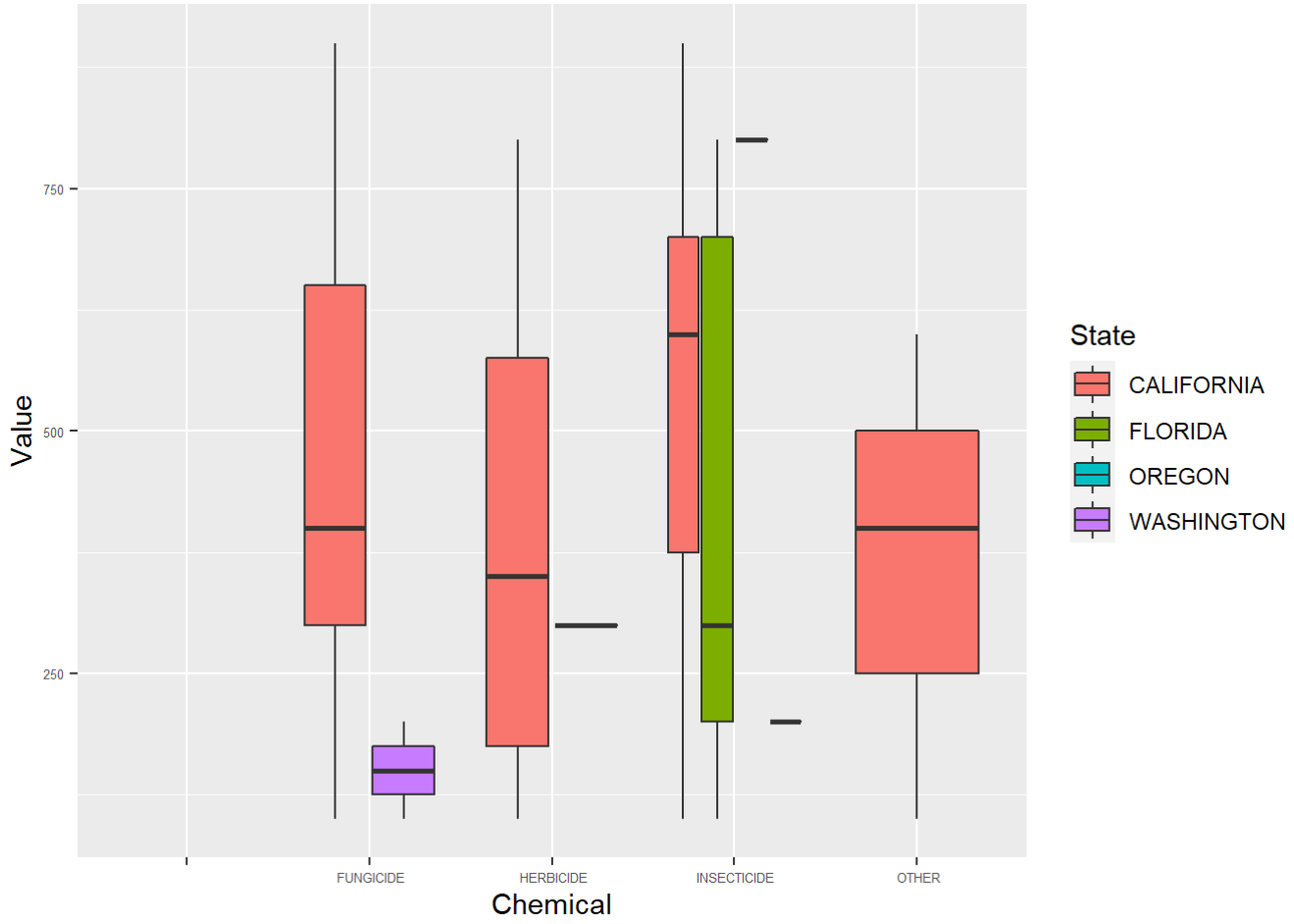
Value - value of chemicals

## 2.3 EDA

First, I would like to make graphs to compare the values of different chemicals used for strawberry. Let us see what is going on.




I filtered data with measured in lb to compare different chemicals with their values. We can see that Insecticide has the highest value, and followed by Fungicide, then, herbicide and other(Nitrogen, Sulfur, Potash) have lowest value.




Now I added the continent variable to see if I can find any regional clues. The results showed that almost all kind of chemicals were used in California, Insecticide was the most common one, and Herbicide was least used; surprisingly, in Florida, people only used Insecticide as chemical, and in Washington state, people only used Fungicide as chemical.

## 2.4 Summary and conclusion

After a simple data analysis, I found some interesting things. For strawberries, Insecticide is the most used chemical for growing strawberries. California is the state with the most diverse chemicals, and almost every chemical is used; on the contrary, some states, such as Florida and Washington, use only one specific chemical to grow strawberries. The above analysis is based on the data about strawberries I cleaned, which has uncertainty and can only be used as a reference.

# 3 References

David Morison, Exploratory data analysis into the relationship between different types of crime in London“https://towardsdatascience.com/exploratory-data-analysis-into-the-relationship-between-different-types-of-crime-in-london-20c328e193ff (https://towardsdatascience.com/exploratory-data-analysis-into-the-relationship-between-different-types-of-crime-in-london-20c328e193ff)”

USDA, National Agricultural Statistics Service “https://quickstats.nass.usda.gov/results/D416E96E-3D5C-324C-9334-1D38DF88FFF1 (https://quickstats.nass.usda.gov/results/D416E96E-3D5C-324C-9334-1D38DF88FFF1)”

R for Data Science, Garrett Grolemund, Hadley Wickham“https://r4ds.had.co.nz/ (https://r4ds.had.co.nz/)”

Almost All American Strawberries Are Grown With Toxic Chemicals, NICOLA TWILLEYCYNTHIA GRABERGASTROPOD SEPTEMBER 1, 2019“https://www.theatlantic.com/science/archive/2019/09/toxic-chemicals-used-grow-strawberries/596968/ (https://www.theatlantic.com/science/archive/2019/09/toxic-

chemicals-used-grow-strawberries/596968/)”