

MA615 strawberry EDA

Hongnan Su

10/19/2024

```
knitr::opts_chunk$set(echo = TRUE)
library(dplyr)
```

```
##
## Attaching package: 'dplyr'

## The following objects are masked from 'package:stats':
##
##   filter, lag

## The following objects are masked from 'package:base':
##
##   intersect, setdiff, setequal, union
```

```
library(stringr)
library(ggplot2)
library(patchwork)
```

Deal with data

```
strawberry<-read.csv("strawberry_cleaned.csv")
```

```
#First I check data of California
california<- strawberry[strawberry$State == "CALIFORNIA", ]
#"Domain" column means the types used in strawberries planting,I want to find all kind of chemical thing
california_1<-california%>%filter(str_detect(Domain, "CHEMICAL"))

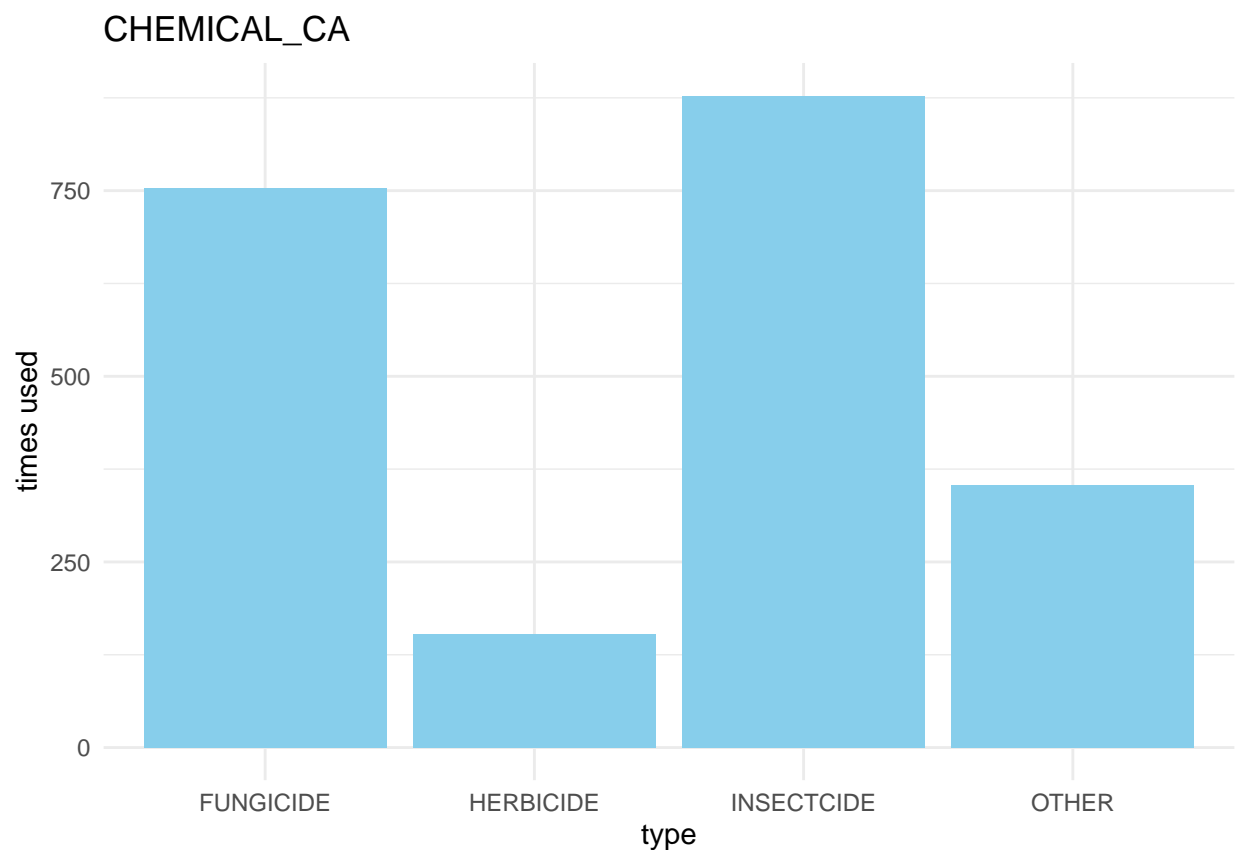
count_CA<-california_1%>%count(Domain)
count_CA
```

```
##           Domain    n
## 1  CHEMICAL, FUNGICIDE 753
## 2  CHEMICAL, HERBICIDE 153
## 3 CHEMICAL, INSECTICIDE 878
## 4    CHEMICAL, OTHER 353
```

We can find that fungicide has been used 753 times, herbicide 153times, insecticide 878 times, and other substances 353 times.

```
#Use ggplot to draw a bar chart to show this.
chemical_CA<- data.frame(category = c("FUNGICIDE", "HERBICIDE", "INSECTICIDE", "OTHER"),value = c(753,153,878,353))

plot_CA<-ggplot(chemical_CA, aes(x = category, y = value)) +
  geom_bar(stat = "identity", fill = "skyblue") +
  labs(title = "CHEMICAL_CA", x = "type", y = "times used") +
  theme_minimal()
plot_CA
```



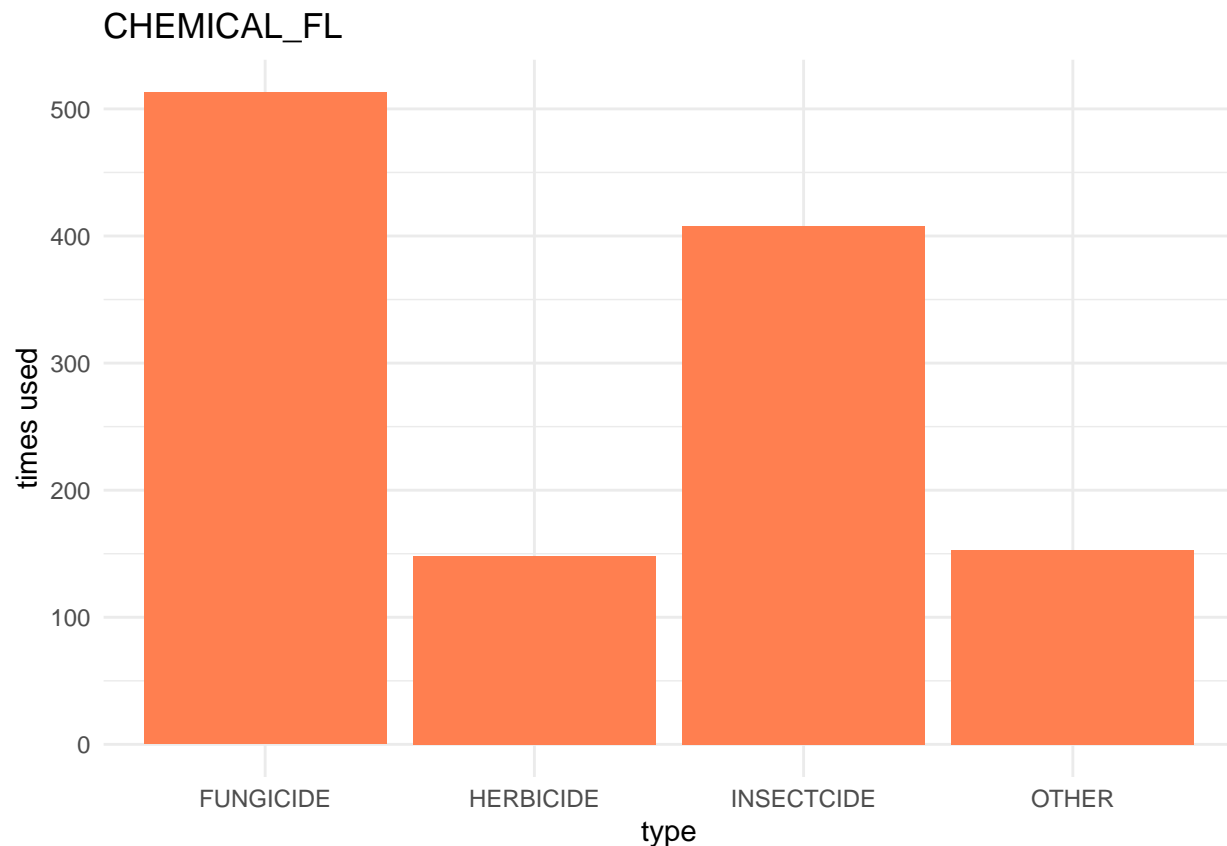
```
#Now I want to know same data in Florida
florida<- strawberry[strawberry$State == "FLORIDA", ]

#Same,find chemical things in Florida
florida_1<-florida%>%filter(str_detect(Domain, "CHEMICAL"))
count_FL<-florida_1%>%count(Domain)
count_FL
```

```
##           Domain    n
## 1  CHEMICAL, FUNGICIDE 513
## 2  CHEMICAL, HERBICIDE 148
## 3 CHEMICAL, INSECTICIDE 408
## 4      CHEMICAL, OTHER 153
```

```
chemical_FL<- data.frame(category = c("FUNGICIDE", "HERBICIDE","INSECTIDE", "OTHER"),value = c(513, 140, 410, 155))

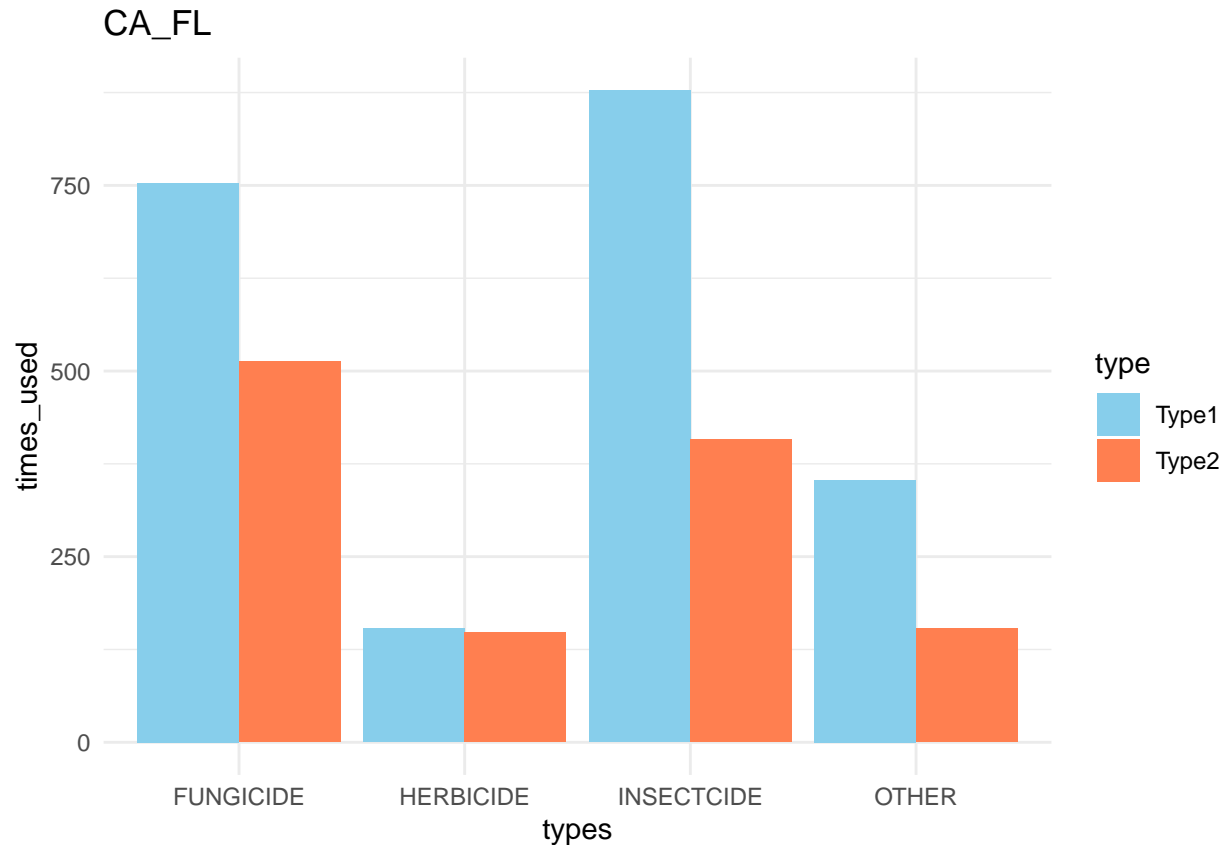
plot_FL<-ggplot(chemical_FL, aes(x = category, y = value)) +
  geom_bar(stat = "identity", fill = "coral") +
  labs(title = "CHEMICAL_FL", x = "type", y = "times used") +
  theme_minimal()
plot_FL
```



#If I want to check the difference of two states use of these 4 chemical things,I can combine this two

```
combine_data <- data.frame(category=rep(c("FUNGICIDE","HERBICIDE","INSECTIDE", "OTHER"), 2),value = c(513, 140, 410, 155, 513, 140, 410, 155))

combine_plot<-ggplot(combine_data, aes(x = category, y = value, fill = type)) +
  geom_bar(stat = "identity", position = "dodge") +
  scale_fill_manual(values = c("Type1" = "skyblue", "Type2" = "coral")) +
  labs(title = "CA_FL", x = "types", y = "times_used") +
  theme_minimal()
combine_plot
```



As we can see, CA's total usage is larger than FL, and the maximum difference is in insecticide.

EDA part

Q1: The chart shows California uses a lot more insecticide than Florida, why?

Answer

- **Crop Variety and Scale:** California cultivates a wide variety of crops, including fruits, nuts, and vegetables. These crops are often more sensitive to pests, leading to a greater need for insecticide to protect them.
- **Climatic Conditions:** The climatic conditions in California may be more conducive to the survival and reproduction of certain pests, necessitating more frequent use of insecticides by farmers.
- **Agricultural Scale:** Agriculture in California is generally larger in scale and more commercialized, which often results in higher frequency and quantity of chemical use.
- **Pest Management Strategies:** California may employ more concentrated pest management strategies, including preventive use of insecticides.
- **Regulations and Oversight:** Differences in regulations and oversight regarding agricultural management and pesticide use between the two states may also influence the amount of insecticides used.

Q2: In California, insecticide is used more than fungicide, but in Florida, fungicide is used more than insecticide. Why?

Answer

- **Crop Types:** California grows a large number of fruits, nuts, and vegetables, which are often more susceptible to insect damage, necessitating a higher use of insecticides. In contrast, Florida's crops, such as citrus fruits, may be more prone to fungal diseases.
- **Climatic Conditions:** California's climate is conducive to the survival and reproduction of certain pests, leading farmers to use insecticides more frequently. Conversely, Florida's humid climate favors the growth of fungi, increasing the use of fungicides.