Strawberry

2024-10-28

Introduction

This project aims to perform EDA on the Chemicals Dataset. I have previously already performed data cleaning and have created a new cleaned chemical data file. The aim is to see if certain chemicals are used more frequently in certain states, and there might be regional trends in the use of fungicides, insecticides, and other chemicals.

Loading necessary libraries and dataset

```r
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
```

```r
library(ggplot2)


# Load the cleaned chemical data
cleaned_data <- read.csv("cleaned_chemical_data.csv")

# Inspect the first few rows of the data
head(cleaned_data)
```

```
##   X Program Year Period Week.Ending Geo.Level      State State.ANSI Ag.District
## 1 1  SURVEY 2023   YEAR          NA     STATE CALIFORNIA          6          NA
## 2 2  SURVEY 2023   YEAR          NA     STATE CALIFORNIA          6          NA
## 3 3  SURVEY 2023   YEAR          NA     STATE CALIFORNIA          6          NA
## 4 4  SURVEY 2023   YEAR          NA     STATE CALIFORNIA          6          NA
## 5 5  SURVEY 2023   YEAR          NA     STATE CALIFORNIA          6          NA
## 6 6  SURVEY 2023   YEAR          NA     STATE CALIFORNIA          6          NA
##   Ag.District.Code County County.ANSI Zip.Code Region watershed_code Watershed
```

```
## 1              NA     NA          NA       NA     NA             0        NA
## 2              NA     NA          NA       NA     NA             0        NA
## 3              NA     NA          NA       NA     NA             0        NA
## 4              NA     NA          NA       NA     NA             0        NA
## 5              NA     NA          NA       NA     NA             0        NA
## 6              NA     NA          NA       NA     NA             0        NA
##     Commodity
## 1 STRAWBERRIES
## 2 STRAWBERRIES
## 3 STRAWBERRIES
## 4 STRAWBERRIES
## 5 STRAWBERRIES
## 6 STRAWBERRIES
##                                                                Data.Item
## 1                          STRAWBERRIES - APPLICATIONS, MEASURED IN LB
## 2                          STRAWBERRIES - APPLICATIONS, MEASURED IN LB
## 3                          STRAWBERRIES - APPLICATIONS, MEASURED IN LB
## 4                          STRAWBERRIES - APPLICATIONS, MEASURED IN LB
## 5 STRAWBERRIES - APPLICATIONS, MEASURED IN LB / ACRE / APPLICATION, AVG
## 6 STRAWBERRIES - APPLICATIONS, MEASURED IN LB / ACRE / APPLICATION, AVG
##                 Domain
## 1   CHEMICAL, FUNGICIDE
## 2 CHEMICAL, INSECTICIDE
## 3 CHEMICAL, INSECTICIDE
## 4       CHEMICAL, OTHER
## 5   CHEMICAL, FUNGICIDE
## 6 CHEMICAL, INSECTICIDE
##                                                  Domain.Category
## 1              CHEMICAL, FUNGICIDE: (OXATHIAPIPROLIN = 128111)
## 2              CHEMICAL, INSECTICIDE: (CYCLANILIPROLE = 26202)
## 3                 CHEMICAL, INSECTICIDE: (PERMETHRIN = 109701)
## 4 CHEMICAL, OTHER: (ISARIA FUMOSOROSEA STRAIN FE 9901 = 115003)
## 5              CHEMICAL, FUNGICIDE: (OXATHIAPIPROLIN = 128111)
## 6              CHEMICAL, INSECTICIDE: (CYCLANILIPROLE = 26202)
##                    Use                     Chemical_Name Chemical_Code Value
## 1   CHEMICAL, FUNGICIDE                  (OXATHIAPIPROLIN        128111   (D)
## 2 CHEMICAL, INSECTICIDE                   (CYCLANILIPROLE         26202   (D)
## 3 CHEMICAL, INSECTICIDE                       (PERMETHRIN        109701   (D)
## 4       CHEMICAL, OTHER (ISARIA FUMOSOROSEA STRAIN FE 9901        115003  (NA)
## 5   CHEMICAL, FUNGICIDE                  (OXATHIAPIPROLIN        128111   (D)
## 6 CHEMICAL, INSECTICIDE                   (CYCLANILIPROLE         26202   (D)
##   CV....
## 1     NA
## 2     NA
## 3     NA
## 4     NA
## 5     NA
## 6     NA
```

Exploratory Data Analysis

We shall group the data by state and chemical use (e.g., fungicide, insecticide, etc.) to explore how different chemicals are distributed across regions. This helps us see patterns and differences in chemical usage between states. By grouping the data this way, we can better understand how each state contributes to overall chemical usage in strawberry farming, and whether certain chemicals are more prevalent in particular regions.

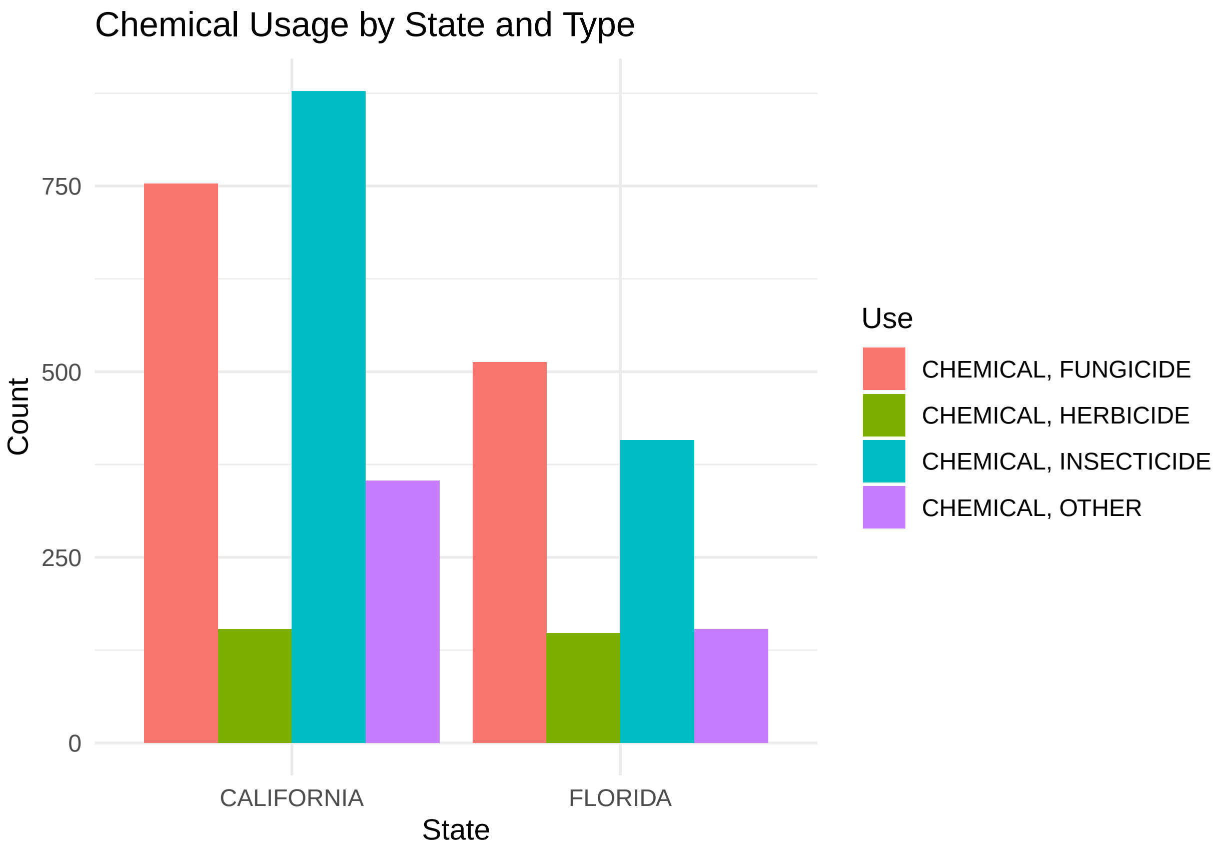
```r
# Group by state and chemical type, and calculate counts
chemical_usage_by_state <- cleaned_data %>%
  group_by(State, Use) %>%
  summarize(Count = n(), .groups = 'drop')

# Inspect the first few rows of the result
head(chemical_usage_by_state)
```

```
## # A tibble: 6 x 3
##   State      Use                   Count
##   <chr>      <chr>                 <int>
## 1 CALIFORNIA CHEMICAL, FUNGICIDE     753
## 2 CALIFORNIA CHEMICAL, HERBICIDE     153
## 3 CALIFORNIA CHEMICAL, INSECTICIDE   878
## 4 CALIFORNIA CHEMICAL, OTHER         353
## 5 FLORIDA    CHEMICAL, FUNGICIDE     513
## 6 FLORIDA    CHEMICAL, HERBICIDE     148
```

From the grouped data, we observed that California is a major user of all chemical types, particularly insecticides (878 applications) and fungicides (753 applications). Florida also has significant chemical usage, especially with fungicides (513 applications) and herbicides (148 applications). This indicates that California and Florida may have more intensive chemical use compared to other regions, which could be due to their larger strawberry production scales.

Now, we generate a bar plot to visually compare chemical usage by state and type. This helps us easily identify which chemicals are most commonly used and in which regions.

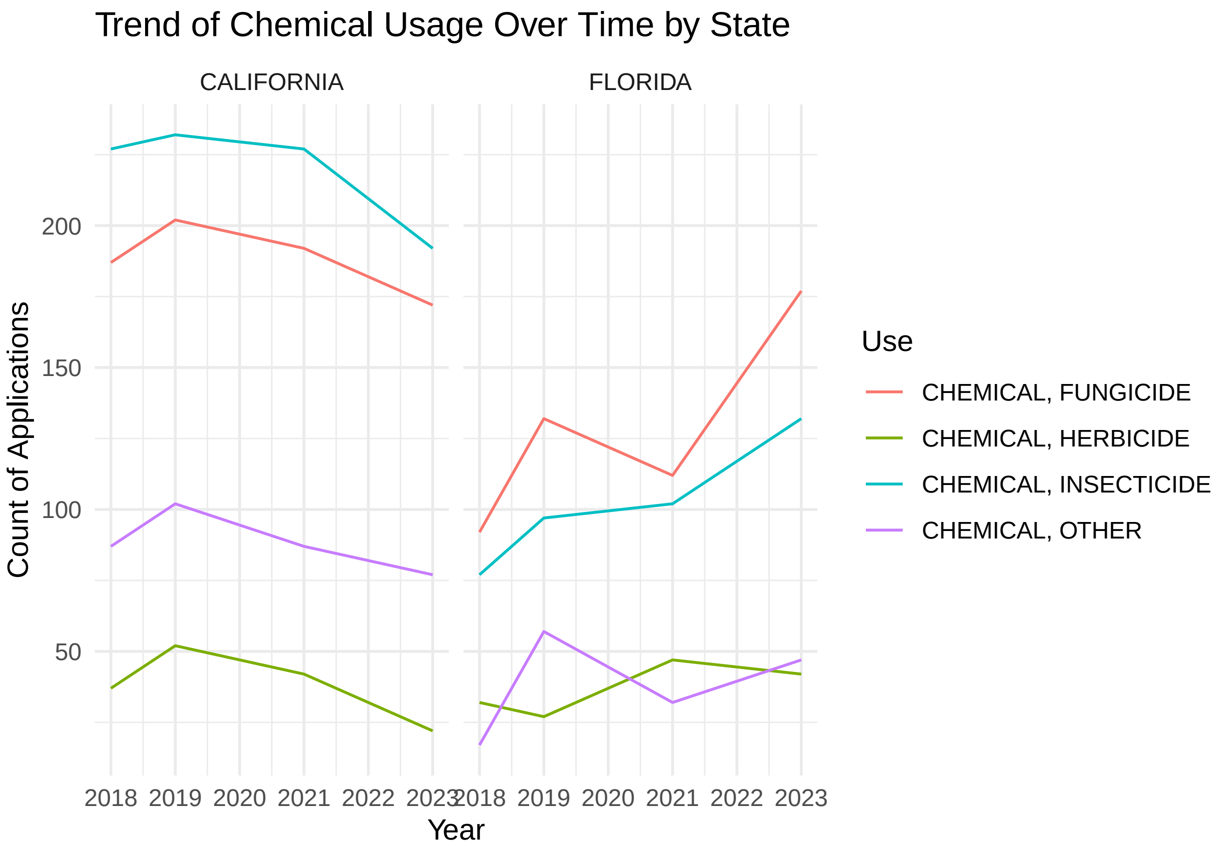
```r
# Visualizing the distribution of chemical usage by state and chemical type
ggplot(chemical_usage_by_state, aes(x = State, y = Count, fill = Use)) +
  geom_bar(stat = "identity", position = "dodge") +
  labs(title = "Chemical Usage by State and Type", x = "State", y = "Count") +
  theme_minimal()
```



This shows that insecticides and fungicides are the dominant chemicals used in both states, but with different proportions. California uses more insecticides, whereas Florida uses more fungicides. Herbicides are used much less in both states, which might reflect different pest management practices or crop growing conditions.

Next, we do the trend analysis over time for the chemical usage. We'll look at how the usage of different chemicals has changed across the years for each state.

```
# First, let's ensure that the 'Year' column is numeric
cleaned_data$Year <- as.numeric(cleaned_data$Year)

# Group by year, state, and chemical use type to see how usage changes over time
chemical_trend_over_time <- cleaned_data %>%
  group_by(Year, State, Use) %>%
  summarize(Count = n(), .groups = 'drop')

# Visualize the trend of chemical usage over time
ggplot(chemical_trend_over_time, aes(x = Year, y = Count, color = Use)) +
  geom_line() +
  facet_wrap(~State) +
  labs(title = "Trend of Chemical Usage Over Time by State", x = "Year", y = "Count of Applications") +
  theme_minimal()
```



Here, we group the data by Year, State, and Use to see the count of chemical applications for each year and category. We create a line plot to visualize how the use of different chemical types (fungicides, herbicides, etc.) has changed over time in each state.

The increasing use of fungicides and herbicides in both states in recent years could indicate a shift toward more aggressive pest management or a reaction to changing environmental or agricultural conditions. In California, the decline in insecticide usage suggests a potential reduction in pest issues, the use of alternative pest control methods, or environmental regulations limiting their application.

Next, we can analyze the most commonly used chemicals within each type (fungicide, insecticide, herbicide, etc.) to see which specific chemicals are the most prevalent

```
# Group by chemical name and type to find the most commonly used chemicals
chemical_breakdown <- cleaned_data %>%
  group_by(Chemical_Name, Use) %>%
  summarize(Count = n(), .groups = 'drop') %>%
  arrange(desc(Count))

# Display the top 10 most frequently used chemicals
head(chemical_breakdown, 10)
```

```
## # A tibble: 10 x 3
##    Chemical_Name         Use                    Count
##    <chr>                 <chr>                  <int>
##  1 (ABAMECTIN            CHEMICAL, INSECTICIDE     40
##  2 (ACETAMIPRID          CHEMICAL, INSECTICIDE     40
##  3 (AZOXYSTROBIN         CHEMICAL, FUNGICIDE       40
##  4 (BIFENAZATE           CHEMICAL, INSECTICIDE     40
```

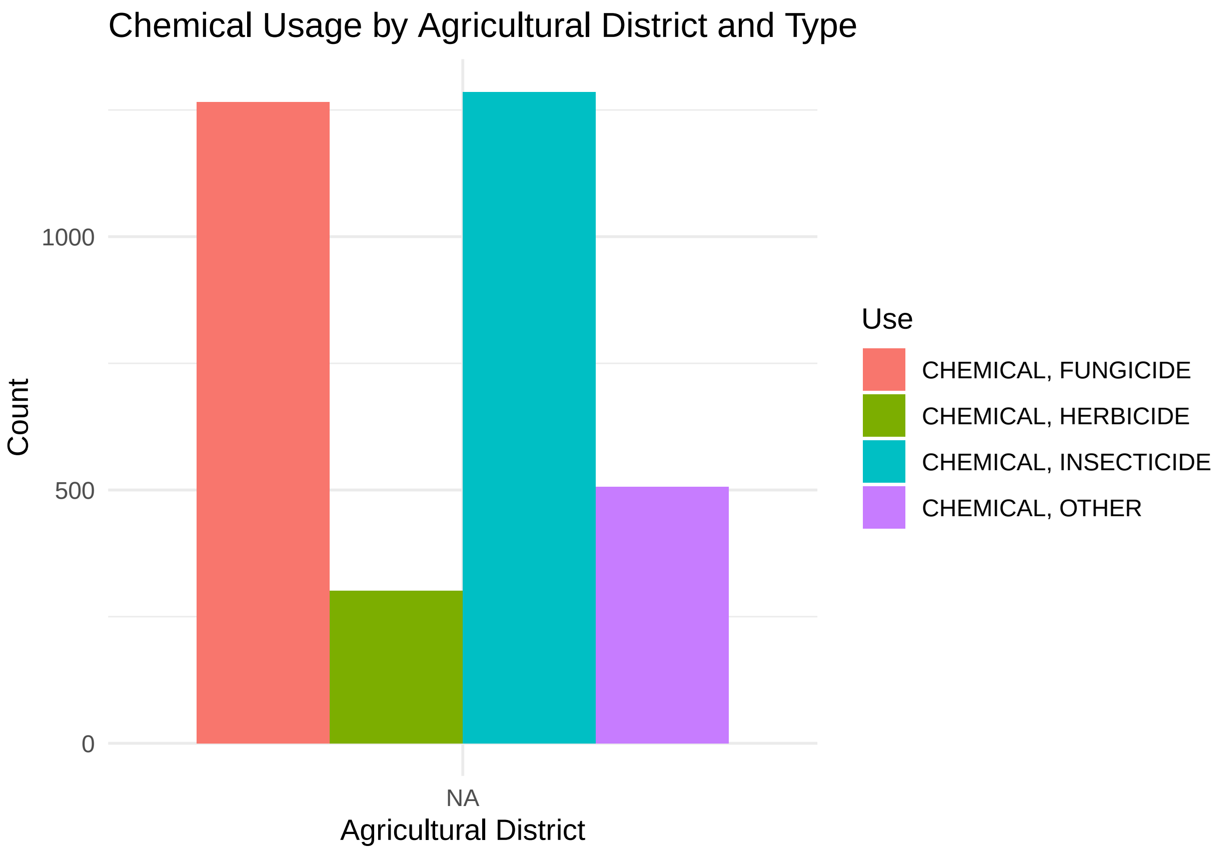
```
## 5 (BIFENTHRIN          CHEMICAL, INSECTICIDE    40
## 6 (CAPTAN              CHEMICAL, FUNGICIDE      40
## 7 (CHLORANTRANILIPROLE CHEMICAL, INSECTICIDE    40
## 8 (CYPRODINIL          CHEMICAL, FUNGICIDE      40
## 9 (DIFENOCONAZOLE      CHEMICAL, FUNGICIDE      40
## 10 (FENHEXAMID         CHEMICAL, FUNGICIDE      40
```

The top 10 most frequently used chemicals consist of: Insecticides such as Abamectin, Acetamiprid, Azoxystrobin, Bifenazate, and Bifenthrin. Fungicides such as Captan, Cyprodinil, and Fenhexamid.

Interestingly, all of these chemicals have the same count of 40 applications. This suggests that there is a consistent level of usage for these chemicals, indicating that they are equally important in managing strawberry farming challenges like pests and fungi.

Next, we can analyze how the distribution of chemical types (fungicides, insecticides, etc.) varies across different regions, like agricultural districts or broader geographic areas.

```r
# Group by agricultural district (if available) and chemical type to see the distribution
chemical_by_region <- cleaned_data %>%
  group_by(Ag.District, Use) %>%
  summarize(Count = n(), .groups = 'drop')

# Visualize the distribution of chemical types by agricultural district
ggplot(chemical_by_region, aes(x = Ag.District, y = Count, fill = Use)) +
  geom_bar(stat = "identity", position = "dodge") +
  labs(title = "Chemical Usage by Agricultural District and Type", x = "Agricultural District", y = "Co
  theme_minimal()
```

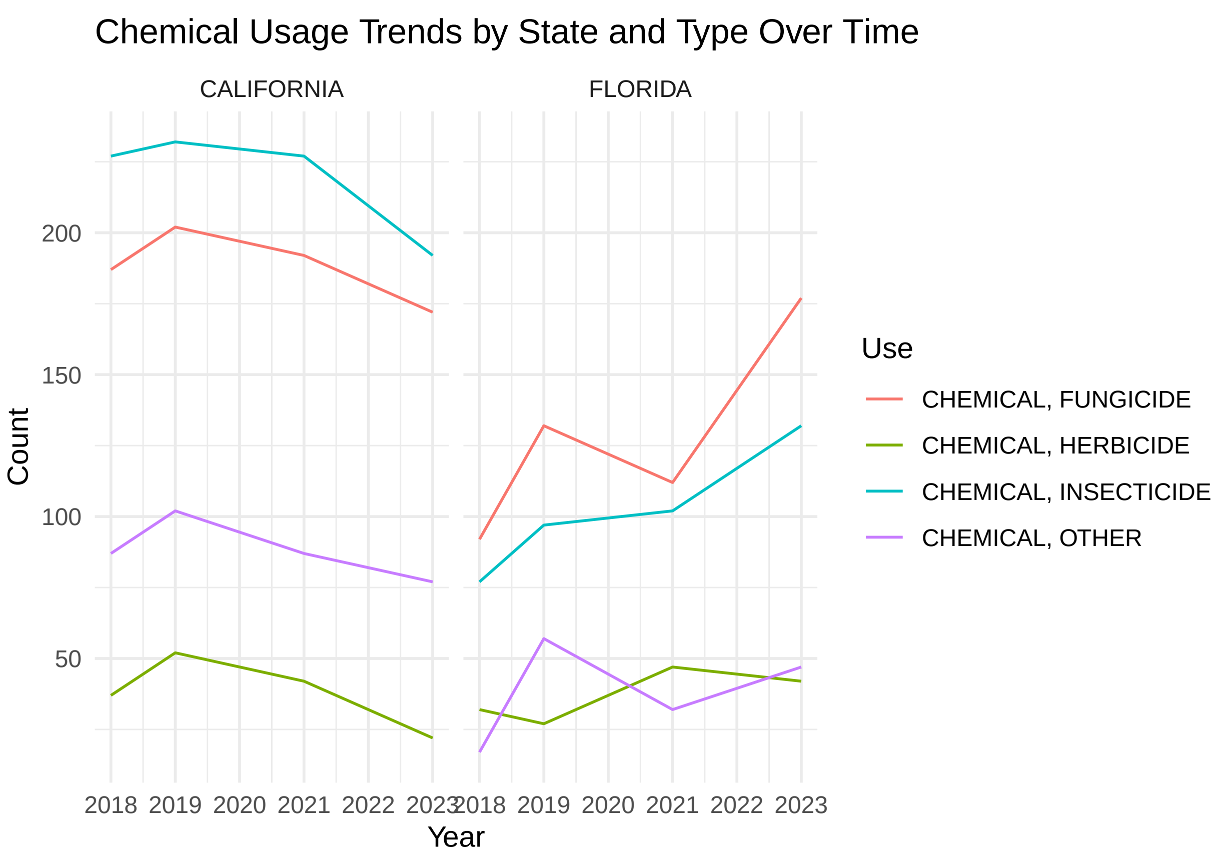


from the bar chart, it seems that the “Agricultural District” field has some missing or undefined values, as indicated by the "NA" label. However, the chart still provides useful insights into the types of chemicals being used.

The chart indicates a heavy use of insecticides and fungicides, with insecticides being particularly dominant in the regions for which data is available. Fungicide usage is significant as well, but herbicides have a much lower representation compared to the other chemical types. The presence of "NA" in the agricultural district could mean missing or undefined data in the dataset for certain districts.

Next, we plot a line chart to visualize how chemical usage has evolved over time in California and Florida, broken down by chemical type (fungicide, insecticide, etc.). This analysis can helps us explore the long-term trends in chemical applications, allowing us to see whether the use of certain chemicals is increasing or decreasing in key strawberry-producing states.

```r
# Group by state, year, and chemical type to analyze trends over time in each state
chemical_trend_by_state_time <- cleaned_data %>%
  group_by(State, Year, Use) %>%
  summarize(Count = n(), .groups = 'drop')

# Visualize chemical usage over time by state and type
ggplot(chemical_trend_by_state_time, aes(x = Year, y = Count, color = Use)) +
  geom_line() +
  facet_wrap(~State) +
  labs(title = "Chemical Usage Trends by State and Type Over Time", x = "Year", y = "Count") +
  theme_minimal()
```



The contrasting trends between California and Florida suggest regional differences in pest management strategies, regulatory environments, or environmental factors affecting strawberry farming. The spike in chemical use in Florida from 2020 onward may be due to external factors liks climate change(might be worth exploring)

#Conclusion

Throughout our exploratory data analysis (EDA), we investigated the patterns of chemical usage in strawberry farming across different states, over time, and across chemical types.

California used more insecticides, while Florida used more fungicides.T his likely reflects different pest and disease pressures in these regions. Chemical use in California declined, while Florida showed an increase starting in 2020. Maybe regulatory factors in California may have reduced use, while Florida's increase could be due to rising pest/disease issues. Common chemicals like Abamectin (insecticide) and Captan (fungicide) were heavily used in both states. These chemicals are likely the most effective for addressing common strawberry farming challenges. Insecticides and fungicides dominated in all districts, with herbicides used less frequently. This reflects the primary pest and disease concerns in strawberry cultivation.