Strawberry_Final_Report

Haochen Li

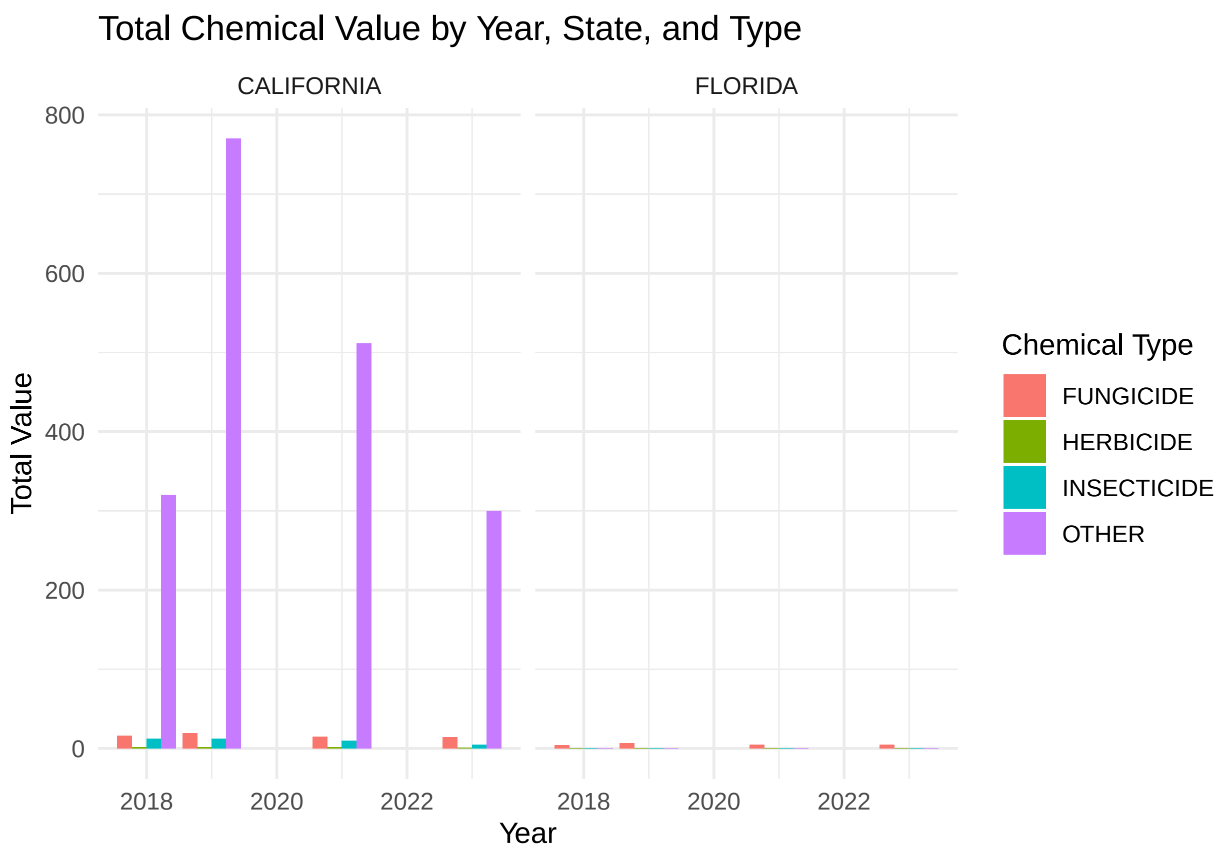
2024-11-18

Strawberries are one of the most widely cultivated and consumed fruits in the United States, cherished for their nutritional value and taste. However, their cultivation often involves significant chemical inputs, including fertilizers and pesticides, to maintain high yields and manage pests. This study aims to investigate the toxicity and chemical usage in strawberry production across different states, with a particular focus on California, the largest strawberry-producing state in the country. By analyzing chemical usage data, this research seeks to identify patterns, regional differences, and potential implications for environmental and human health. Insights from this analysis can inform sustainable agricultural practices and policy decisions for reducing the environmental footprint of strawberry cultivation while maintaining productivity.

```
grouped_data <- subset_data %>%
  group_by(Year, State, type) %>%
  summarise(Total_Value = sum(Value, na.rm = TRUE))
```

```
## `summarise()` has grouped output by 'Year', 'State'. You can override using the
## `.groups` argument.
```

```
# Create a ggplot to visualize the total Value by Year, State, and Type
ggplot(grouped_data, aes(x = Year, y = Total_Value, fill = type)) +
  geom_bar(stat = "identity", position = "dodge") +
  facet_wrap(~ State) +
  labs(title = "Total Chemical Value by Year, State, and Type",
       x = "Year",
       y = "Total Value",
       fill = "Chemical Type") +
  theme_minimal()
```



The graph shows total chemical usage by year and type in California and Florida.

California dominates with significantly higher chemical usage, particularly in the "Other" category, which peaks in 2020 and 2022. Fungicides, herbicides, and insecticides contribute minimally.

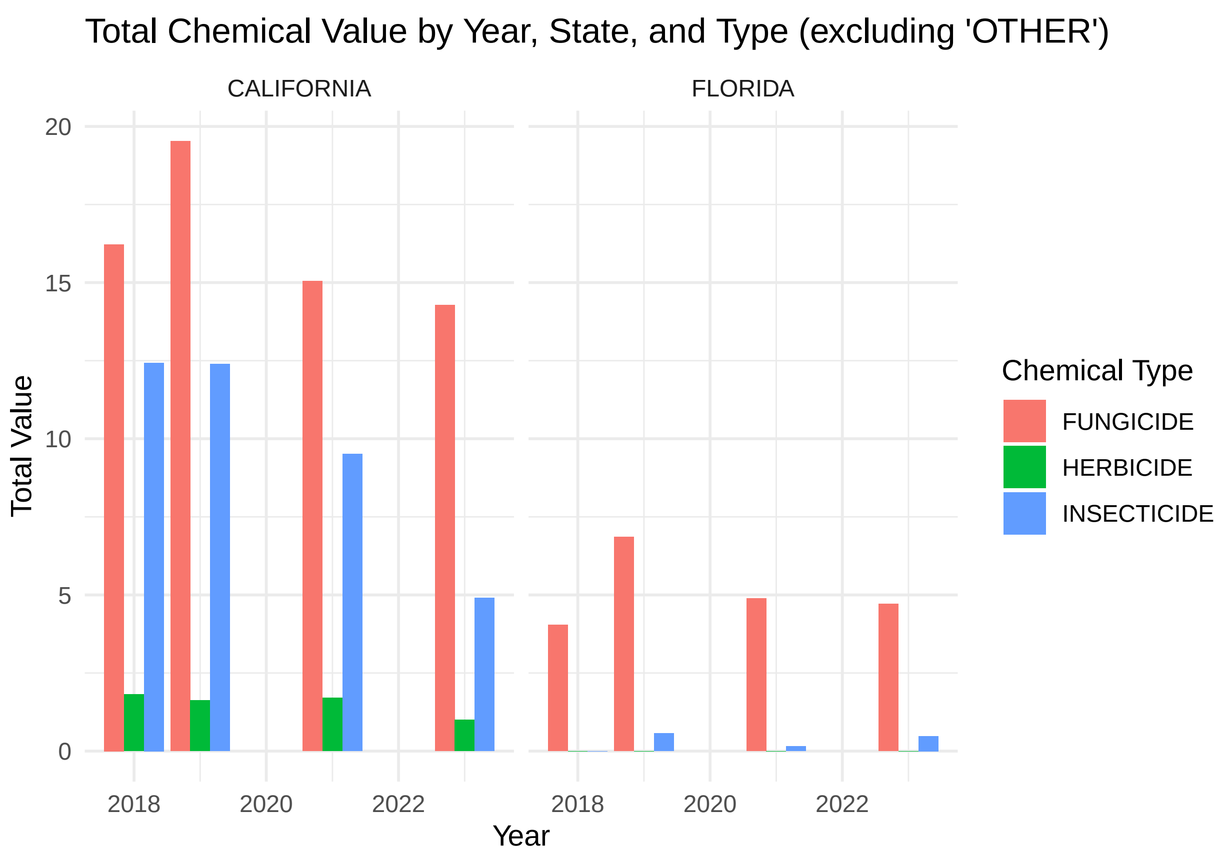
Florida's usage is much lower overall, with "Other" chemicals still being the largest category, though at much smaller values.

```r
subset_data$Value <- as.numeric(subset_data$Value)

filtered_other <- subset_data %>%
  filter(type != "OTHER")

# Group the filtered data by Year, State, and Type, and summarize the Value column
filtered_other <- filtered_other %>%
  group_by(Year, State, type) %>%
  summarise(Total_Value = sum(Value, na.rm = TRUE))
```

The high prevalence of "Other" chemicals in California suggests a need for further investigation into their specific components and impact.

```
## 'summarise()' has grouped output by 'Year', 'State'. You can override using the
## '.groups' argument.
```

```r
# Create a ggplot to visualize the total Value by Year, State, and Type (excluding "OTHER")
ggplot(filtered_other, aes(x = Year, y = Total_Value, fill = type)) +
  geom_bar(stat = "identity", position = "dodge") +
  facet_wrap(~ State) +
  labs(title = "Total Chemical Value by Year, State, and Type (excluding 'OTHER')",
       x = "Year",
       y = "Total Value",
       fill = "Chemical Type") +
  theme_minimal()
```



California Dominates:

Fungicides (red) are the most widely used chemical type across all years in California.

Insecticides (blue) are the second most prominent type, with noticeable usage across 2018, 2020, and 2022.

Herbicides (green) have minimal usage compared to fungicides and insecticides. Florida's Chemical Usage:

Fungicides dominate in Florida but on a much smaller scale compared to California.

Insecticides and herbicides are rarely used.

California shows consistent fungicide usage, peaking in 2018 and 2022. Florida exhibits lower, consistent fungicide use with no significant trends.

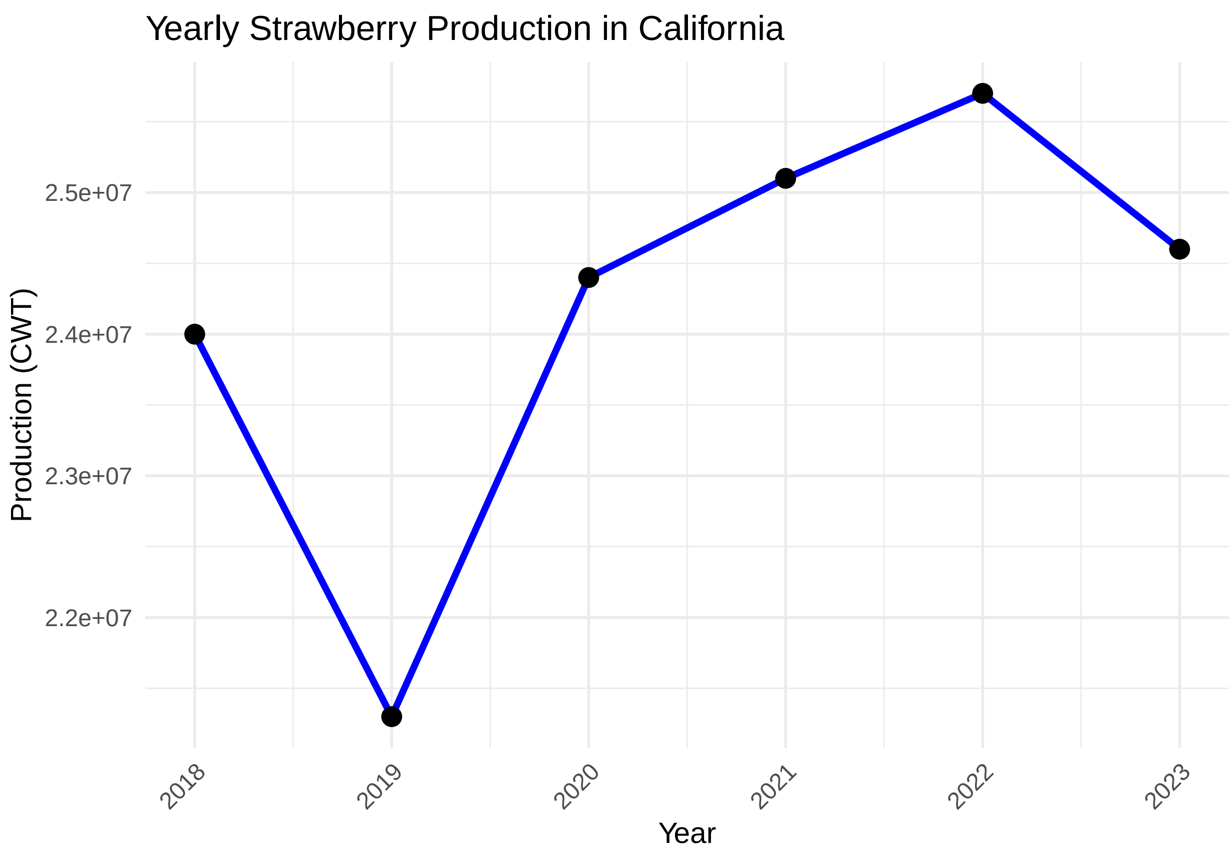
Fungicides are critical in both states, especially in California, where they form the bulk of chemical usage.

Minimal herbicide usage indicates reliance on other methods for weed control.

```r
california_production_data <- survey_d_total %>%
  filter(State == "CALIFORNIA"& product_price == "PRODUCTION" & measure == "CWT"& Period == "YEAR")

ggplot(california_production_data, aes(x = Year, y = as.numeric(gsub(",", "", Value)))) +
  geom_line(group = 1, color = "blue", size = 1.2) + # Line plot
  geom_point(size = 3) +             # Points on the plot
  labs(title = "Yearly Strawberry Production in California",
       x = "Year",
       y = "Production (CWT)") +
  theme_minimal() +
  theme(axis.text.x = element_text(angle = 45, hjust = 1))
```

The graph highlights California's larger agricultural chemical footprint compared to Florida.

```
## Warning: Using ‘size‘ aesthetic for lines was deprecated in ggplot2 3.4.0.
## i Please use ‘linewidth‘ instead.
## This warning is displayed once every 8 hours.
## Call ‘lifecycle::last_lifecycle_warnings()‘ to see where this warning was
## generated.
```

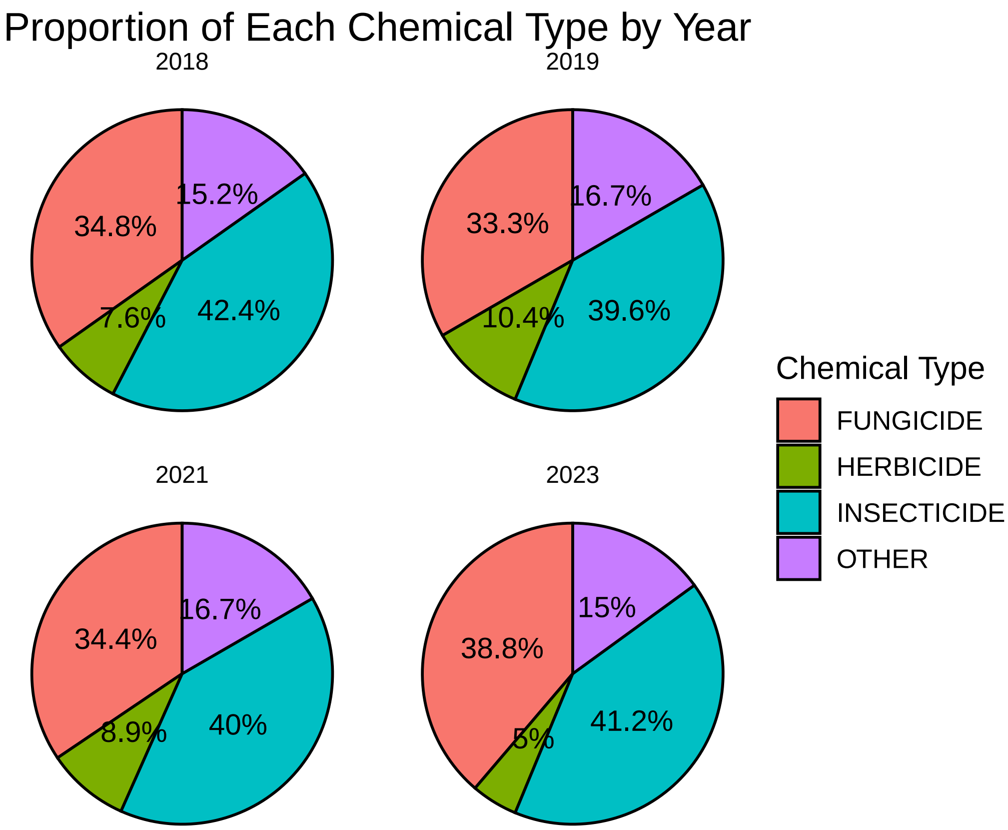


The graph depicts the yearly strawberry production in California from 2018 to 2023, measured in CWT (hundredweight). Production starts at approximately 24 million CWT in 2018, drops sharply in 2019 to around 22 million CWT, then steadily increases from 2020 to 2022, peaking at about 25.5 million CWT. In 2023, production slightly declines, falling back to around 24.5 million CWT.

```
yearly_proportion <- subset_data %>%
  filter(State == "CALIFORNIA") %>%
  group_by(Year, type) %>%
  summarise(count = n(), .groups = "drop") %>%
  group_by(Year) %>%
  mutate(proportion = (count / sum(count)) * 100)

ggplot(yearly_proportion, aes(x = "", y = proportion, fill = type)) +
  geom_bar(stat = "identity", width = 1, color = "black") +
  coord_polar("y", start = 0) +
  labs(title = "Proportion of Each Chemical Type by Year", fill = "Chemical Type") +
  geom_text(aes(label = paste0(round(proportion, 1), "%")),
            position = position_stack(vjust = 0.5)) +
  theme_void() +
  theme(legend.title = element_text(size = 12),
        legend.text = element_text(size = 10),
```

```
      plot.title = element_text(hjust = 0.5, size = 15)) +
  facet_wrap(~ Year)
```

The significant dip in 2019 could indicate challenges such as adverse weather conditions, pest outbreaks, or market disruptions. The consistent growth from 2020 to 2022 reflects recovery, possibly due to improved agricultural practices, favorable climate, or increased demand. However, the slight decline in 2023 suggests potential new pressures on production, such as resource limitations or environmental factors.



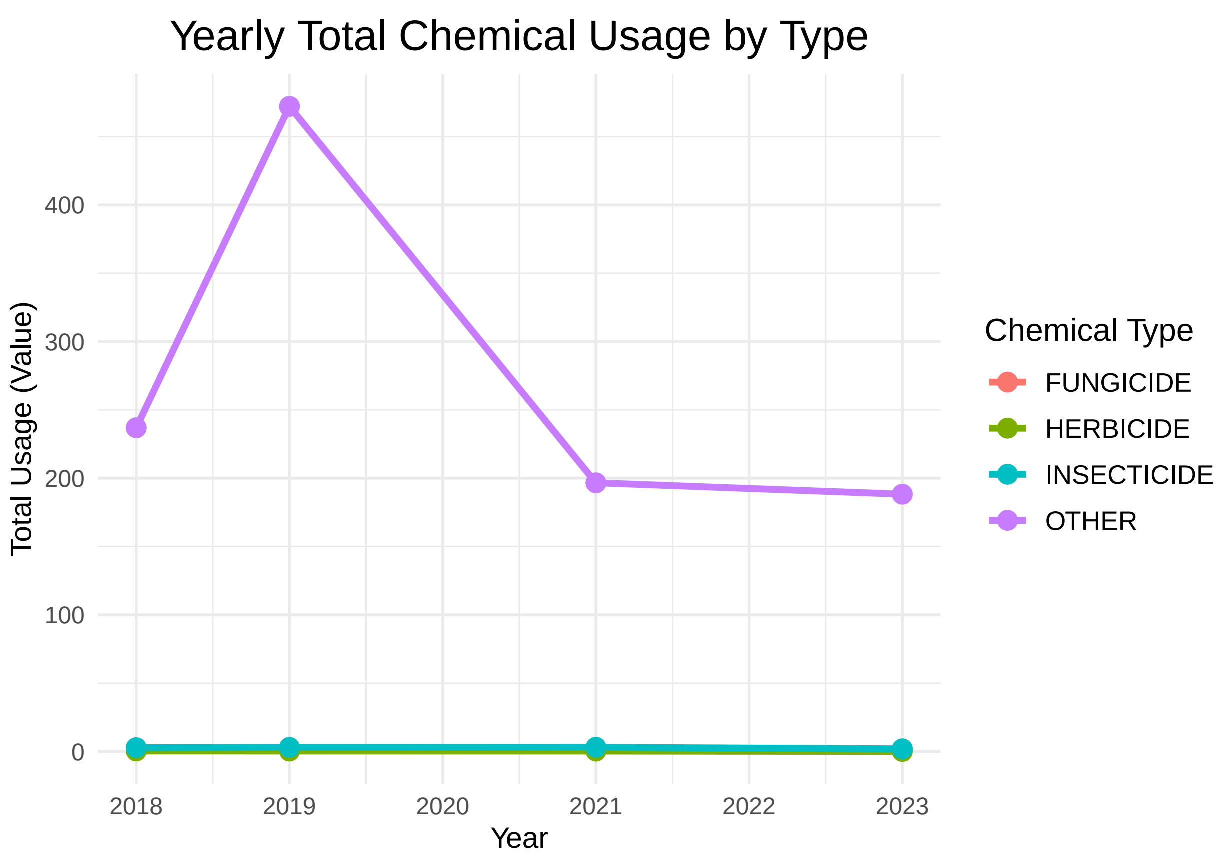

```
fungicides_data <- subset(subset_data, type == "FUNGICIDE")

# Summing up the 'Value' column for each fungicide
fungicide_usage <- aggregate(Value ~ chem_name, data = fungicides_data, sum)

# Sorting by usage in descending order
most_used_fungicides <- fungicide_usage[order(-fungicide_usage$Value), ]

# Selecting the top 3 most-used fungicides
top_fungicides <- head(most_used_fungicides, 3)
print(top_fungicides)
```

The pie chart shows that insecticides account for the largest share of chemical usage at 40.8%, followed by fungicides at 35.2%. Herbicides (8.1%) and "Other" chemicals (15.9%) have com-

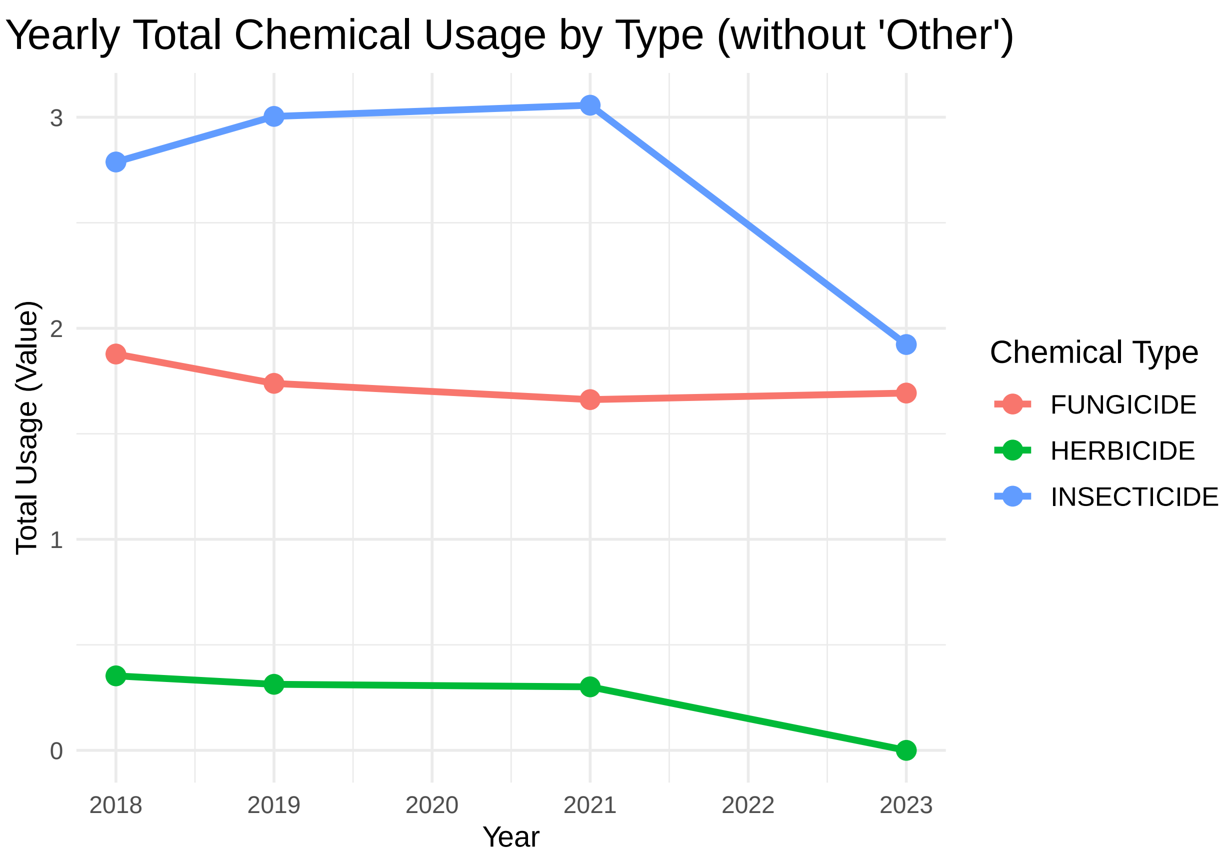
paratively smaller proportions, highlighting a focus on pest and disease control in chemical applications.

```
##    chem_name  Value
## 38    SULFUR 17.771
## 41    THIRAM 16.846
## 5     CAPTAN 14.965
```

```r
# Step 1: Summarize total usage by Year and Type
type_yearly_usage <- dangerous_chemicals %>%
  group_by(Year, type) %>%
  summarize(Total_Usage = sum(Value, na.rm = TRUE), .groups = "drop")

# Step 2: Plot yearly total usage for each type
ggplot(type_yearly_usage, aes(x = Year, y = Total_Usage, color = type, group = type)) +
  geom_line(size = 1.2) +
  geom_point(size = 3) +
  labs(title = "Yearly Total Chemical Usage by Type",
       x = "Year",
       y = "Total Usage (Value)",
       color = "Chemical Type") +
  theme_minimal() +
  theme(plot.title = element_text(hjust = 0.5, size = 16),
        legend.title = element_text(size = 12),
        legend.text = element_text(size = 10))
```



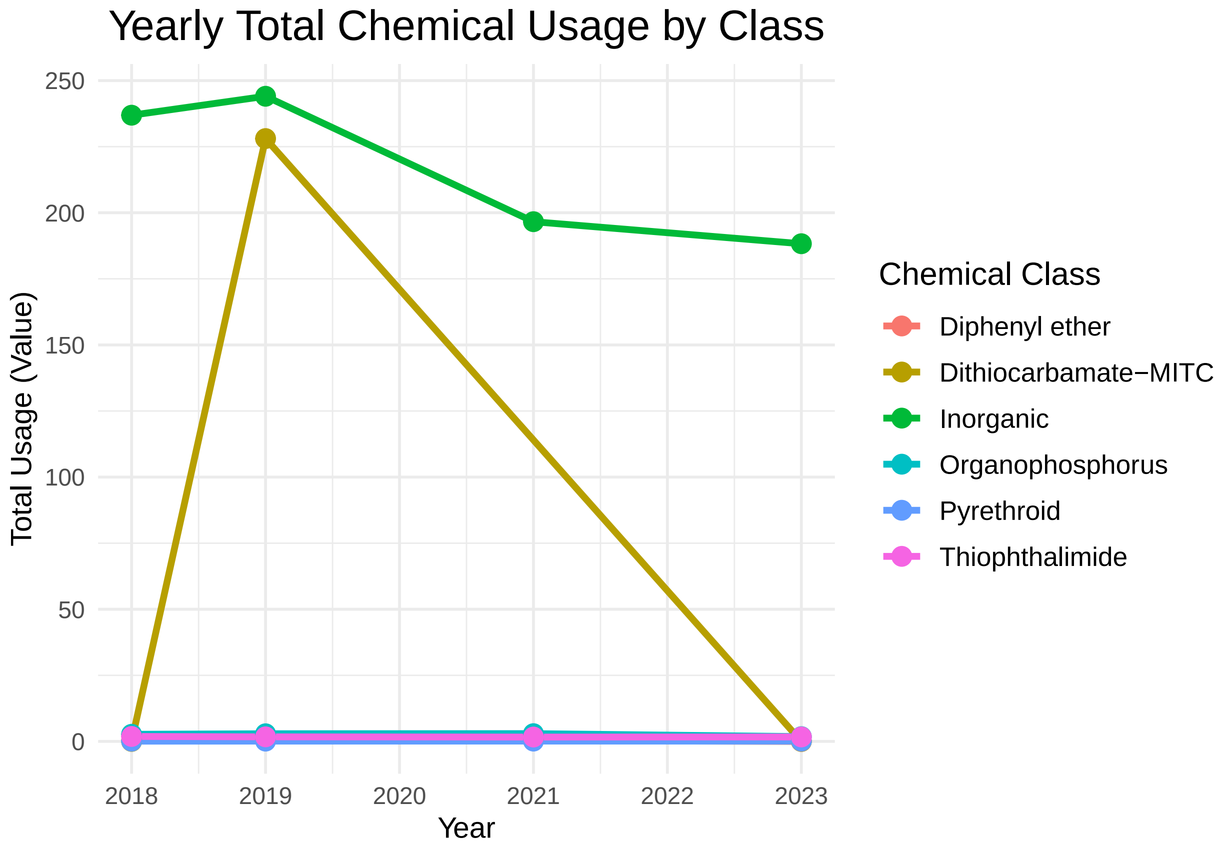
```r
type_yearly_usage1 <- dangerous_chemicals %>%
  filter(type != "OTHER") %>%  # Exclude "Other" type
  group_by(Year, type) %>%
  summarize(Total_Usage = sum(Value, na.rm = TRUE), .groups = "drop")

ggplot(type_yearly_usage1, aes(x = Year, y = Total_Usage, color = type, group = type)) +
  geom_line(size = 1.2) +
  geom_point(size = 3) +
  labs(title = "Yearly Total Chemical Usage by Type (without 'Other')",
       x = "Year",
       y = "Total Usage (Value)",
       color = "Chemical Type") +
  theme_minimal() +
  theme(plot.title = element_text(hjust = 0.5, size = 16),
        legend.title = element_text(size = 12),
        legend.text = element_text(size = 10))
```




```r
# Step 1: Summarize total usage by Year and Chemical Class
class_yearly_usage <- dangerous_chemicals %>%
  group_by(Year, Chemical.Class) %>%
  summarize(Total_Usage = sum(Value, na.rm = TRUE), .groups = "drop")
```

```r
# Step 2: Filter out "Other" or irrelevant classes if needed
class_yearly_usage <- class_yearly_usage %>%
  filter(Chemical.Class != "Other")  # Adjust the condition as needed

# Step 3: Plot yearly total usage for each chemical class
ggplot(class_yearly_usage, aes(x = Year, y = Total_Usage, color = Chemical.Class, group = Chemical.Class
  geom_line(size = 1.2) +
  geom_point(size = 3) +
  labs(title = "Yearly Total Chemical Usage by Class",
       x = "Year",
       y = "Total Usage (Value)",
       color = "Chemical Class") +
  theme_minimal() +
  theme(plot.title = element_text(hjust = 0.5, size = 16),
        legend.title = element_text(size = 12),
        legend.text = element_text(size = 10))
```

The graph illustrates the yearly total usage of fungicides, herbicides, and insecticides from 2018 to 2023, excluding the "Other" category. Insecticides show the highest usage but exhibit a noticeable decline after 2021, possibly due to regulatory changes or alternative practices. Fungicide usage remains relatively stable, indicating its consistent importance in agricultural practices. Herbicides have the lowest usage and a slight declining trend, reflecting limited and targeted application. This suggests a shift in chemical reliance over the years, with insecticides seeing reduced dependence.



The trend shows that Inorganic and Organophosphorus dominate in total usage butare declining over time, suggesting a shift in chemical use patterns.

The drop in Dithiocarbamate-MITC usage highlights a significant change, possibly due to stricter safety concerns or a switch to alternatives.

Organophosphorus: Commonly used as insecticides, organophosphorus chemicals are known for their acute toxicity to humans (causing neurological and respiratory effects) and potential for groundwater contamination and bioaccumulation; while usage is declining, monitoring remains essential due to its toxic effects.

Inorganic: Often used as fungicides and general-purpose agricultural chemicals, inorganic compounds can persist in the environment, causing heavy metal pollution and soil quality degradation; their high and stable usage highlights the need for continued environmental impact monitoring.

Dithiocarbamate-MITC: These fungicides and soil fumigants are linked to respiratory and skin irritation in humans and can produce toxic degradation products; their sharp decline in usage after 2019, likely due to regulatory changes, reduces their current priority.

Pyrethroid: Widely used as insecticides targeting specific pests, pyrethroids are less toxic to humans but highly toxic to aquatic organisms due to their persistence in water bodies; despite low but consistent usage, they require attention in regions near water.

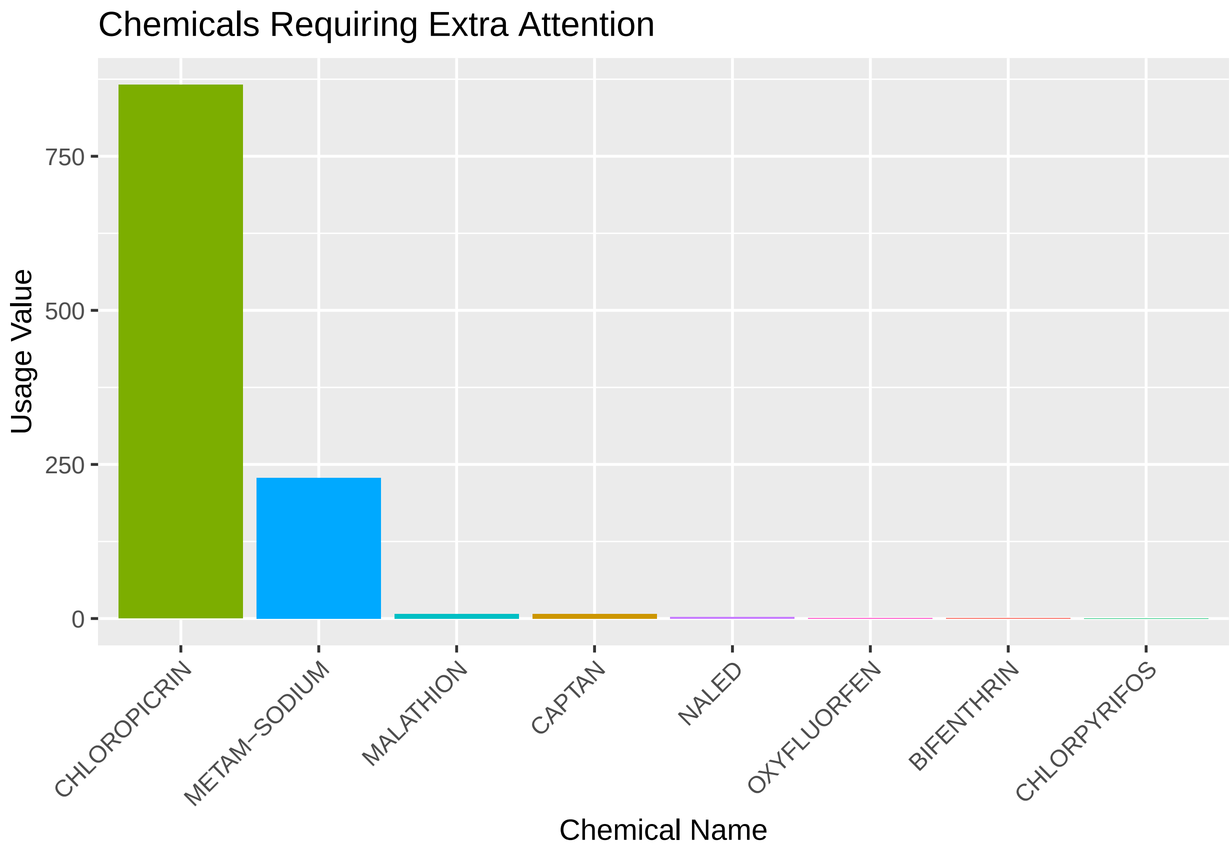
Diphenyl Ether: Used primarily as herbicides for weed control, diphenyl ethers pose risks of phytotoxicity to non-target plants and biodiversity; with low and stable usage, they are a lower priority for concern.

Thiophthalimide: Fungicides targeting specific plant pathogens, thiophthalimides exhibit low toxicity and environmental persistence; their very low usage suggests minimal immediate attention is needed.

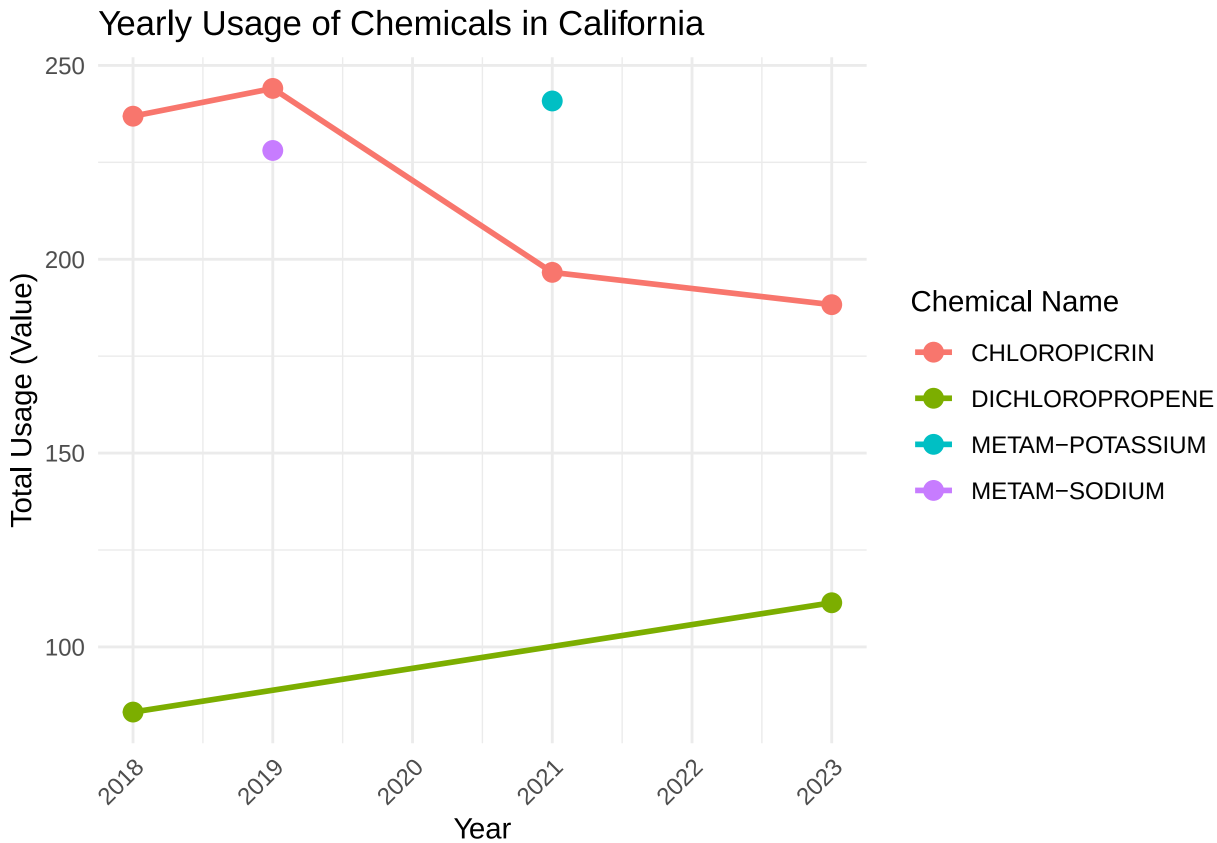
```r
# Create a bar chart of the most used dangerous chemicals
ggplot(dangerous_chemicals, aes(x = reorder(chem_name, -Value), y = Value, fill = chem_name)) +
  geom_bar(stat = "identity") +
  labs(title = "Chemicals Requiring Extra Attention", x = "Chemical Name", y = "Usage Value") +
  theme(axis.text.x = element_text(angle = 45, hjust = 1)) +
  theme(legend.position = "none")
```

These chemicals require our attention due to their significant usage and associated risks. Chloropicrin and Metam-Sodium dominate in usage, with Chloropicrin showing the highest value, indicating its widespread application and potential environmental and health impacts. Organophosphorus, inorganic chemicals, and pyrethroids also need close monitoring because of their acute toxicity to humans and ecosystems, persistence, and environmental risks. Summing up these classes, their critical usage trends

and concerns emphasize the need for regulatory oversight and alternative solutions.




```r
california_data <- subset(subset_data, State == "CALIFORNIA")

top_two_chemicals <- california_data %>%
  group_by(Year, chem_name) %>%
  summarize(Total_Usage = sum(Value, na.rm = TRUE), .groups = "drop") %>%
  group_by(Year) %>%
  slice_max(order_by = Total_Usage, n = 2)



ggplot(top_two_chemicals, aes(x = Year, y = Total_Usage, color = chem_name, group = chem_name)) +
  geom_line(size = 1) +
  geom_point(size = 3) +
  labs(title = "Yearly Usage of Chemicals in California",
       x = "Year",
       y = "Total Usage (Value)",
       color = "Chemical Name") +
  theme_minimal() +
  theme(axis.text.x = element_text(angle = 45, hjust = 1))
```



The graph shows the yearly usage of four chemicals in California between 2018 and 2023. CHLOROPICRIN was the most used chemical at the start, but its usage steadily declined over time. DICHLOROPROPENE usage, on the other hand, gradually increased each year. METAM-POTASSIUM saw a significant jump in usage in 2023, while METAM-SODIUM was only used in 2019 and 2020. The trends suggest changing preferences or regulations in chemical use over the years.

The chemicals in the graph, like CHLOROPICRIN and DICHLOROPROPENE, are widely used in farming to control pests and diseases, helping to increase crop production. However, they can be harmful to us and the environment. For example, CHLOROPICRIN can irritate our lungs and contribute to air pollution, while DICHLOROPROPENE can contaminate groundwater and may pose cancer risks. It's important to ask if using more of these chemicals really leads to higher production or if there's a point where it causes more harm than good. By looking at the trends, we can start to understand whether these chemicals are being used responsibly and how they might affect not just farming, but also our health and the world around us.