415_Midterm_Project

Haoyi Liu

Before Starting the Project

Data Set

The csv file named "strawb_mar6.csv", which contains strawberry cultivation data in different states and measurements.

Main Focus

1. Pick three chemical treatments used for conventional strawberries in both states and contrast their use. Try to find chemicals with divergent use patterns between the states. Produce, tables, plots, and descriptions of the chemicals, how they are used, and how their use differs between California and Florida.
2. Compare the production and sales of organic and conventional strawberries and strawberries sold for processing. Show differences in price and volume between California and Florida. How do price, cost, and volume relationships change over the years?

Data Cleaning

Install packages

```
# install required packages
#install.packages("knitr")
#install.packages("kableExtra")
#install.packages("tidyverse")
#install.packages("stringr")
#install.packages("ggplot2")
```

```
library(knitr)
library(kableExtra)
library(tidyverse)
```

```
-- Attaching core tidyverse packages ------------------------ tidyverse 2.0.0 --
v dplyr     1.1.4     v readr     2.1.5
v forcats   1.0.0     v stringr   1.5.1
v ggplot2   3.5.1     v tibble    3.2.1
v lubridate 1.9.4     v tidyr     1.3.1
v purrr     1.0.2
-- Conflicts ------------------------------------------ tidyverse_conflicts() --
x dplyr::filter()     masks stats::filter()
x dplyr::group_rows() masks kableExtra::group_rows()
x dplyr::lag()        masks stats::lag()
i Use the conflicted package (<http://conflicted.r-lib.org/>) to force all conflicts to becor
```

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

Read File (From `USDA-NASS starwberries.qmd`)

```
# Label for the code chunk
#| label: read data - glimpse

# Read the CSV file "strawb_mar6.csv" into a data frame called 'strawberry'
# - 'col_names = TRUE' tells read_csv to treat the first row as column names
# - 'show_col_types = FALSE' suppresses printing of column type information
strawberry <- read_csv("strawb_mar6.csv",
                       col_names = TRUE,
                       show_col_types = FALSE)

# Source the file "my_functions.R" to load custom functions defined in that script
source("my_functions.R")
```

```
# remove columns from a data frame that contain only a single unique value
strawb <- strawberry |> drop_one_value_col()
```

Exploring Data

```r
# assume data is a tibble
# n_show is the number of rows to show

show_unique <- function(data, nrows=10 ){
  # make a tibble items to hold the data to show
  # browser()
  a <- nrows * dim(data)[2]  # number of cells in items
  items <- rep(" ", a) # items will coerce everything to char
  dim(items) <- c(nrows ,dim(data)[2]) # shape items
  items <- as_tibble(items)
  colnames(items) <- colnames(data)
  # browser()
  for(i in 1:dim(data)[2]){

    col_items <- unique(data[,i])
    # row_ex is the number of rows needed
    # to make the column length conformable with items
    row_ex <- nrows - dim(col_items)[1]
    if(row_ex >= 0){
      ex_rows <- tibble(rep(" ",row_ex))
      colnames(ex_rows) <- colnames(col_items)
      col_add <- rbind2(col_items, ex_rows)

    } else if(row_ex < 0){
      col_add <- col_items[1:10,]

    }

    items[,i] <- col_add

  }

  return(items)
}

#test <- show_unique(strawb, 10)
```

```r
#|label: split strawb into census and survey pieces

strw_census <- strawb |> filter(Program == "CENSUS")

strw_survey <- strawb |> filter(Program == "SURVEY")
```

```
nrow(strawb) == (nrow(strw_census) + nrow(strw_survey))
```

```
[1] TRUE
```

```
# Remove columns that contain only a single unique value
s_census <- strw_census |> drop_one_value_col(prt_val = TRUE)
```

```
[1] "Looking for single value columns in data frame:  strw_census"
[1] "Columns dropped:"
    Program      Period Week Ending
   "CENSUS"      "YEAR"          NA
```

```
s_survey <- strw_survey |> drop_one_value_col(prt_val = TRUE)
```

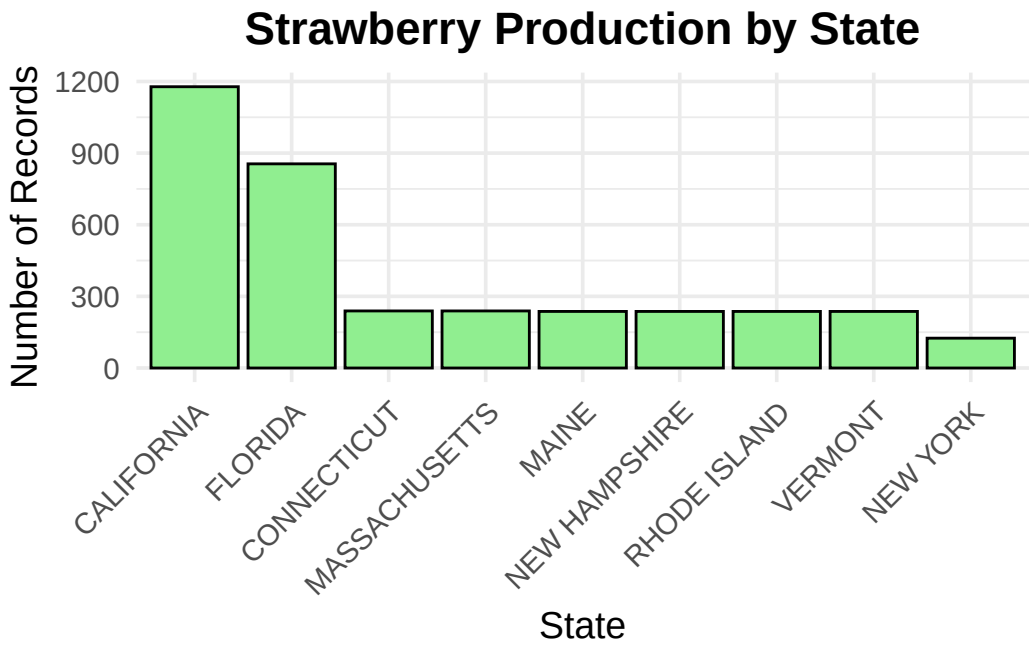
```
[1] "Looking for single value columns in data frame:  strw_survey"
[1] "Columns dropped:"
       Program      Commodity         CV (%)
      "SURVEY" "STRAWBERRIES"             NA
```

```
# Preview up to 10 unique values per column in each data set
unique_cen <- s_census |> show_unique(nrows = 10)
unique_sur <- s_survey |> show_unique(nrows = 10)

# Drop redundant or unnecessary Data
strw_census <- s_census |> select(-`State ANSI`)

# Remove 'State ANSI', 'Week Ending', and 'Period' for Data Cleaning
strw_survey <- s_survey |> select(-`State ANSI`, -`Week Ending`, -Period)

# Removing intermediate and temporary data set
rm(s_census, s_survey, strawberry, items)
```

Strawberry Growth Location

```
# Starberries' grown place :)
state_all <- strawb |> distinct(State)
state_all1 <- strawb |> group_by(State) |> count()
```

```r
# Improved bar plot
ggplot(state_all1, aes(x = reorder(State, -n), y = n)) +
  geom_bar(stat = "identity", fill = "lightgreen", color = "black") +
  labs(
    title = "Strawberry Production by State",
    x = "State",
    y = "Number of Records"
  ) +
  theme_minimal(base_size = 14) +
  theme(
    axis.text.x = element_text(angle = 45, hjust = 1),
    plot.title = element_text(face = "bold", hjust = 0.5)
  )
```




```r
# Show unique values in the 'Domain' column from both datasets
unique(strw_census$Domain)
```

```
[1] "NET GAIN"             "TOTAL"                "NET LOSS"
[4] "AREA OPERATED"        "ECONOMIC CLASS"       "FARM SALES"
[7] "NAICS CLASSIFICATION" "ORGANIC STATUS"
```

```
unique(strw_survey$Domain)
```

```
[1] "TOTAL"                 "CHEMICAL, FUNGICIDE"   "CHEMICAL, INSECTICIDE"
[4] "CHEMICAL, OTHER"       "CHEMICAL, HERBICIDE"   "FERTILIZER"
```

Part 1: Chemical Analysis

Gather Chemical Information Data

```
#Select all the rows that contains chemical use information
chemical <- strw_survey[str_detect(strw_survey$`Domain Category`, regex("chemical", ignore_ca

#Taking only CA and FL state data of chemical usage
chemical_CA <- chemical[chemical$State == "CALIFORNIA", ]
chemical_FL <- chemical[chemical$State == "FLORIDA", ]
print(chemical_CA)
```

```
# A tibble: 1,011 x 9
    Year State      Fruit   Category Item  Metric Domain `Domain Category` Value
   <dbl> <chr>      <chr>   <chr>    <chr> <chr>  <chr>  <chr>              <chr>
 1  2023 CALIFORNIA STRAWB~ " MEASU~  <NA> <NA>   CHEMI~ CHEMICAL, FUNGIC~ (D)
 2  2023 CALIFORNIA STRAWB~ " MEASU~  <NA> <NA>   CHEMI~ CHEMICAL, INSECT~ (D)
 3  2023 CALIFORNIA STRAWB~ " MEASU~  <NA> <NA>   CHEMI~ CHEMICAL, INSECT~ (D)
 4  2023 CALIFORNIA STRAWB~ " MEASU~  <NA> <NA>   CHEMI~ CHEMICAL, OTHER:~ (NA)
 5  2023 CALIFORNIA STRAWB~ " MEASU~ " AV~ <NA>   CHEMI~ CHEMICAL, FUNGIC~ (D)
 6  2023 CALIFORNIA STRAWB~ " MEASU~ " AV~ <NA>   CHEMI~ CHEMICAL, INSECT~ (D)
 7  2023 CALIFORNIA STRAWB~ " MEASU~ " AV~ <NA>   CHEMI~ CHEMICAL, INSECT~ (D)
 8  2023 CALIFORNIA STRAWB~ " MEASU~ " AV~ <NA>   CHEMI~ CHEMICAL, OTHER:~ (NA)
 9  2023 CALIFORNIA STRAWB~ " MEASU~ " AV~ <NA>   CHEMI~ CHEMICAL, FUNGIC~ (D)
10  2023 CALIFORNIA STRAWB~ " MEASU~ " AV~ <NA>   CHEMI~ CHEMICAL, INSECT~ (D)
# i 1,001 more rows
```

```
print(chemical_FL)
```

```
# A tibble: 691 x 9
    Year State   Fruit      Category Item  Metric Domain `Domain Category` Value
   <dbl> <chr>   <chr>      <chr>    <chr> <chr>  <chr>  <chr>              <chr>
 1  2023 FLORIDA STRAWBERR~ " MEASU~ <NA>  <NA>   CHEMI~ CHEMICAL, FUNGIC~ (D)
```

```
 2  2023 FLORIDA STRAWBERR~ " MEASU~ <NA>  <NA>   CHEMI~ CHEMICAL, FUNGIC~ (D)
 3  2023 FLORIDA STRAWBERR~ " MEASU~ <NA>  <NA>   CHEMI~ CHEMICAL, FUNGIC~ (D)
 4  2023 FLORIDA STRAWBERR~ " MEASU~ <NA>  <NA>   CHEMI~ CHEMICAL, HERBIC~ (D)
 5  2023 FLORIDA STRAWBERR~ " MEASU~ <NA>  <NA>   CHEMI~ CHEMICAL, HERBIC~ (D)
 6  2023 FLORIDA STRAWBERR~ " MEASU~ <NA>  <NA>   CHEMI~ CHEMICAL, HERBIC~ (D)
 7  2023 FLORIDA STRAWBERR~ " MEASU~ <NA>  <NA>   CHEMI~ CHEMICAL, INSECT~ (D)
 8  2023 FLORIDA STRAWBERR~ " MEASU~ <NA>  <NA>   CHEMI~ CHEMICAL, INSECT~ (D)
 9  2023 FLORIDA STRAWBERR~ " MEASU~ <NA>  <NA>   CHEMI~ CHEMICAL, INSECT~ (D)
10  2023 FLORIDA STRAWBERR~ " MEASU~ <NA>  <NA>   CHEMI~ CHEMICAL, OTHER:~ (D)
# i 681 more rows
```

```
rm(chemical)
```

Chemical 1: Captan

```
# First we check the use of Captan
# Use library of stringr functions to select the captan information
# Filter rows from chemical_CA
captan_CA <- chemical_CA[str_detect(chemical_CA$`Domain Category`, regex("captan", ignore_cas

# Filter rows from chemical_FL
captan_FL <- chemical_FL[str_detect(chemical_FL$`Domain Category`, regex("captan", ignore_cas

captan_combined <- rbind(captan_CA, captan_FL)

rm(captan_CA)
rm(captan_FL)

print(captan_combined)
```

```
# A tibble: 20 x 9
    Year State      Fruit   Category Item  Metric Domain `Domain Category` Value
   <dbl> <chr>      <chr>   <chr>    <chr> <chr>  <chr>  <chr>              <chr>
 1  2023 CALIFORNIA STRAWB~ " BEARI~ " ME~  <NA>  CHEMI~ CHEMICAL, FUNGIC~ 603,~
 2  2023 CALIFORNIA STRAWB~ " BEARI~ " ME~ " AVG" CHEMI~ CHEMICAL, FUNGIC~ 1.693
 3  2023 CALIFORNIA STRAWB~ " BEARI~ " ME~ " AVG" CHEMI~ CHEMICAL, FUNGIC~ 15.9~
 4  2023 CALIFORNIA STRAWB~ " BEARI~ " ME~ " AVG" CHEMI~ CHEMICAL, FUNGIC~ 9.4
 5  2023 CALIFORNIA STRAWB~ " BEARI~ " ME~ " AVG" CHEMI~ CHEMICAL, FUNGIC~ 88
 6  2021 CALIFORNIA STRAWB~ " BEARI~ " ME~  <NA>  CHEMI~ CHEMICAL, FUNGIC~ 253,~
 7  2021 CALIFORNIA STRAWB~ " BEARI~ " ME~ " AVG" CHEMI~ CHEMICAL, FUNGIC~ 1.662
```

```
 8  2021 CALIFORNIA STRAWB~ " BEARI~ " ME~ " AVG" CHEMI~ CHEMICAL, FUNGIC~ 14.2~
 9  2021 CALIFORNIA STRAWB~ " BEARI~ " ME~ " AVG" CHEMI~ CHEMICAL, FUNGIC~ 8.6
10  2021 CALIFORNIA STRAWB~ " BEARI~ " ME~ " AVG" CHEMI~ CHEMICAL, FUNGIC~ 45
11  2023 FLORIDA    STRAWB~ " BEARI~ " ME~  <NA>  CHEMI~ CHEMICAL, FUNGIC~ 144,~
12  2023 FLORIDA    STRAWB~ " BEARI~ " ME~ " AVG" CHEMI~ CHEMICAL, FUNGIC~ 2.012
13  2023 FLORIDA    STRAWB~ " BEARI~ " ME~ " AVG" CHEMI~ CHEMICAL, FUNGIC~ 10.5~
14  2023 FLORIDA    STRAWB~ " BEARI~ " ME~ " AVG" CHEMI~ CHEMICAL, FUNGIC~ 5.2
15  2023 FLORIDA    STRAWB~ " BEARI~ " ME~ " AVG" CHEMI~ CHEMICAL, FUNGIC~ 96
16  2021 FLORIDA    STRAWB~ " BEARI~ " ME~  <NA>  CHEMI~ CHEMICAL, FUNGIC~ 135,~
17  2021 FLORIDA    STRAWB~ " BEARI~ " ME~ " AVG" CHEMI~ CHEMICAL, FUNGIC~ 2.025
18  2021 FLORIDA    STRAWB~ " BEARI~ " ME~ " AVG" CHEMI~ CHEMICAL, FUNGIC~ 21.3~
19  2021 FLORIDA    STRAWB~ " BEARI~ " ME~ " AVG" CHEMI~ CHEMICAL, FUNGIC~ 10.5
20  2021 FLORIDA    STRAWB~ " BEARI~ " ME~ " AVG" CHEMI~ CHEMICAL, FUNGIC~ 61
```

```
# Now we get the table of captan use data in both California and Florida
```

```
# now we start to reorganize the data in the order of category at first, Item for the second
```

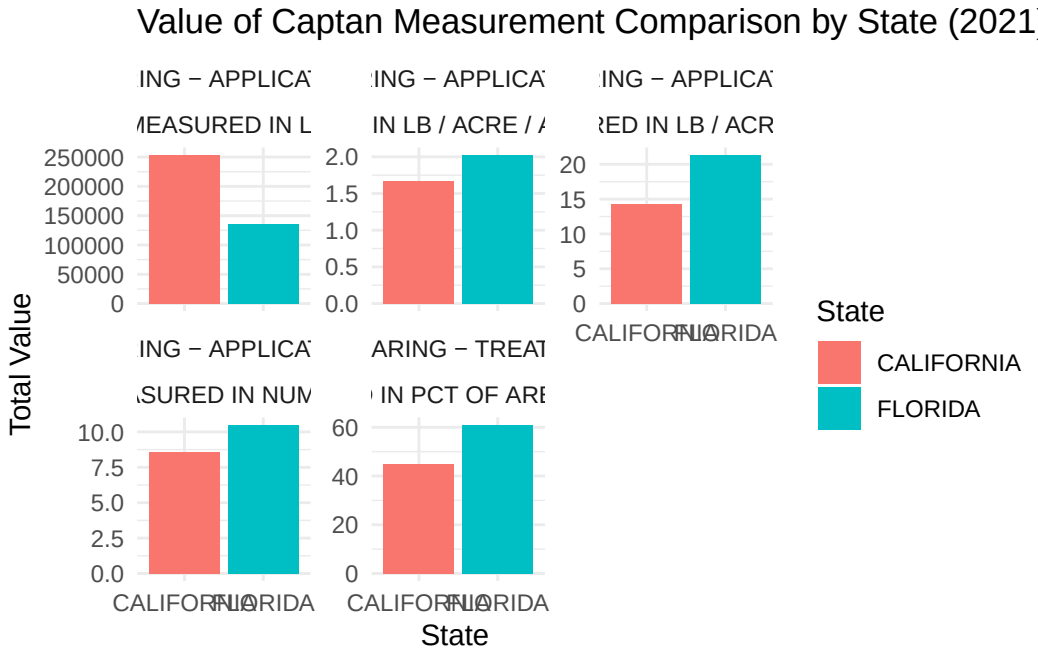
```
captan <- captan_combined |> group_by(Category, Item, State, Year) |> count(Value)

print(captan)
```

```
# A tibble: 20 x 6
# Groups:   Category, Item, State, Year [20]
   Category                  Item                        State  Year Value     n
   <chr>                     <chr>                       <chr> <dbl> <chr> <int>
 1 " BEARING - APPLICATIONS" " MEASURED IN LB"           CALI~  2021 253,~     1
 2 " BEARING - APPLICATIONS" " MEASURED IN LB"           CALI~  2023 603,~     1
 3 " BEARING - APPLICATIONS" " MEASURED IN LB"           FLOR~  2021 135,~     1
 4 " BEARING - APPLICATIONS" " MEASURED IN LB"           FLOR~  2023 144,~     1
 5 " BEARING - APPLICATIONS" " MEASURED IN LB / ACRE / ~ CALI~  2021 1.662     1
 6 " BEARING - APPLICATIONS" " MEASURED IN LB / ACRE / ~ CALI~  2023 1.693     1
 7 " BEARING - APPLICATIONS" " MEASURED IN LB / ACRE / ~ FLOR~  2021 2.025     1
 8 " BEARING - APPLICATIONS" " MEASURED IN LB / ACRE / ~ FLOR~  2023 2.012     1
 9 " BEARING - APPLICATIONS" " MEASURED IN LB / ACRE / ~ CALI~  2021 14.2~     1
10 " BEARING - APPLICATIONS" " MEASURED IN LB / ACRE / ~ CALI~  2023 15.9~     1
11 " BEARING - APPLICATIONS" " MEASURED IN LB / ACRE / ~ FLOR~  2021 21.3~     1
12 " BEARING - APPLICATIONS" " MEASURED IN LB / ACRE / ~ FLOR~  2023 10.5~     1
13 " BEARING - APPLICATIONS" " MEASURED IN NUMBER"       CALI~  2021 8.6       1
14 " BEARING - APPLICATIONS" " MEASURED IN NUMBER"       CALI~  2023 9.4       1
15 " BEARING - APPLICATIONS" " MEASURED IN NUMBER"       FLOR~  2021 10.5      1
```

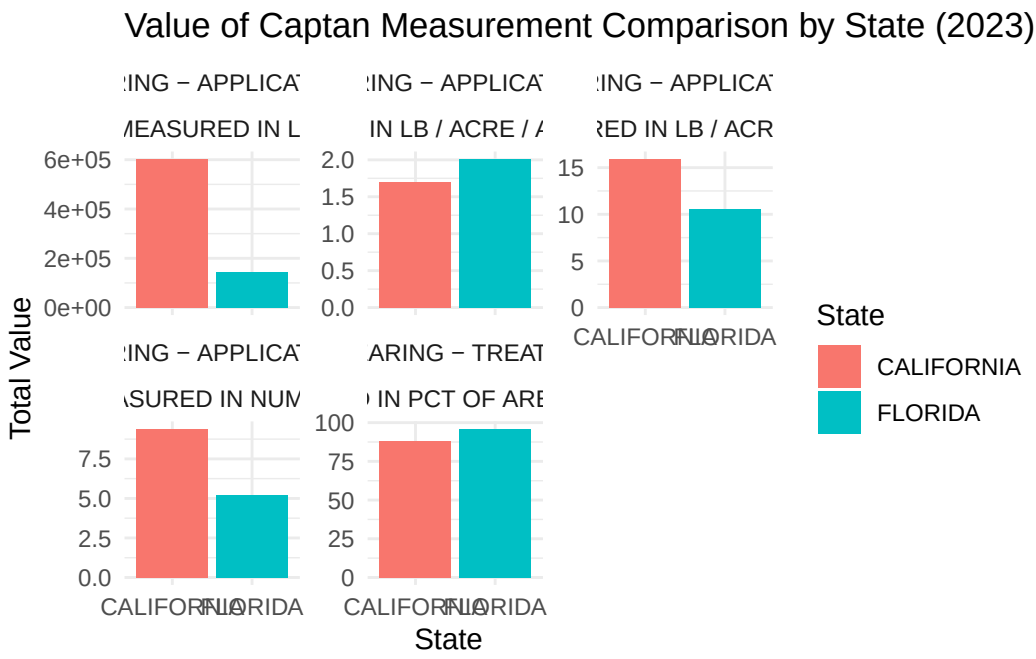
```
16 " BEARING - APPLICATIONS" " MEASURED IN NUMBER"       FLOR~  2023 5.2       1
17 " BEARING - TREATED"      " MEASURED IN PCT OF AREA ~ CALI~  2021 45        1
18 " BEARING - TREATED"      " MEASURED IN PCT OF AREA ~ CALI~  2023 88        1
19 " BEARING - TREATED"      " MEASURED IN PCT OF AREA ~ FLOR~  2021 61        1
20 " BEARING - TREATED"      " MEASURED IN PCT OF AREA ~ FLOR~  2023 96        1
```

```
# This table  directly shows the transition of captan use from year 2021 to 2023 in each stat
```

```
#Then we can make a bar chart to compare the captan use in California and Florida

#Generate Data from 2021 and 2023
captan_2021 <- captan[str_detect(captan$Year, regex("2021", ignore_case = TRUE)), ]
captan_2023 <- captan[str_detect(captan$Year, regex("2023", ignore_case = TRUE)), ]

#Ensure the value under Value column is numeric, so in graph it can provide the correct relat
captan_2021$Value <- as.numeric(gsub(",", "", captan_2021$Value))
captan_2023$Value <- as.numeric(gsub(",", "", captan_2023$Value))

#Then we can make the graph (code perfection by chatgpt, for a better visual effect)
ggplot(captan_2021, aes(x = State, y = Value, fill = State)) +
  geom_bar(stat = "identity", position = "dodge") +
  facet_wrap(~ Category + Item, scales = "free_y") +
  labs(title = "Value of Captan Measurement Comparison by State (2021)",
       x = "State",
       y = "Total Value") +
  theme_minimal()
```




```
ggplot(captan_2023, aes(x = State, y = Value, fill = State)) +
  geom_bar(stat = "identity", position = "dodge") +
  facet_wrap(~ Category + Item, scales = "free_y") +
  labs(title = "Value of Captan Measurement Comparison by State (2023)",
       x = "State",
       y = "Total Value") +
  theme_minimal()
```




```
#Remove intermediate Data set
rm(captan_2021,captan_2023)
```

From the graph of 2021, California uses the fungicide that contains captan (a certified carcinogenic chemical) more than Florida in general. However, when we look up into other measurement, the average detection of captan use in California is less than Florida. This is because California Strawberry Production Firm is larger than Florida in size.

In 2023, There are some changes in the data. The measure in lb/acre/year and measure in number of California firm is higher than Florida firm, meaning that California uses captan fungicide more often per year, and the average number that detected captan is higher than Florida.

Chemical 2: Thiram

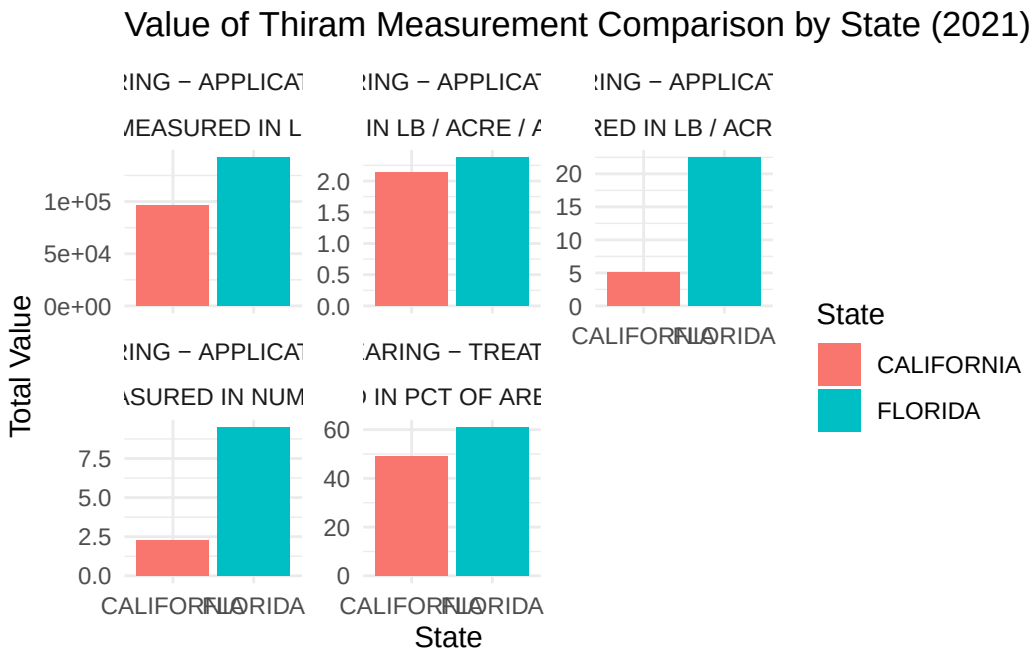
```
#Follows the same step as captan
#Filter rows form chemical_CA
thiram_CA <- chemical_CA[str_detect(chemical_CA$`Domain Category`, regex("thiram", ignore_cas

# Filter rows from chemical_FL
thiram_FL <- chemical_FL[str_detect(chemical_FL$`Domain Category`, regex("thiram", ignore_cas
```

```r
#Get the thiram information data and gather them into one data set
thiram_combined <- rbind(thiram_CA, thiram_FL)

#remove intermediate data set
rm(thiram_CA)
rm(thiram_FL)

#Table of Comparison of different year by same state
thiram <- thiram_combined |> group_by(Category, Item, State, Year) |> count(Value)

print(thiram)
```

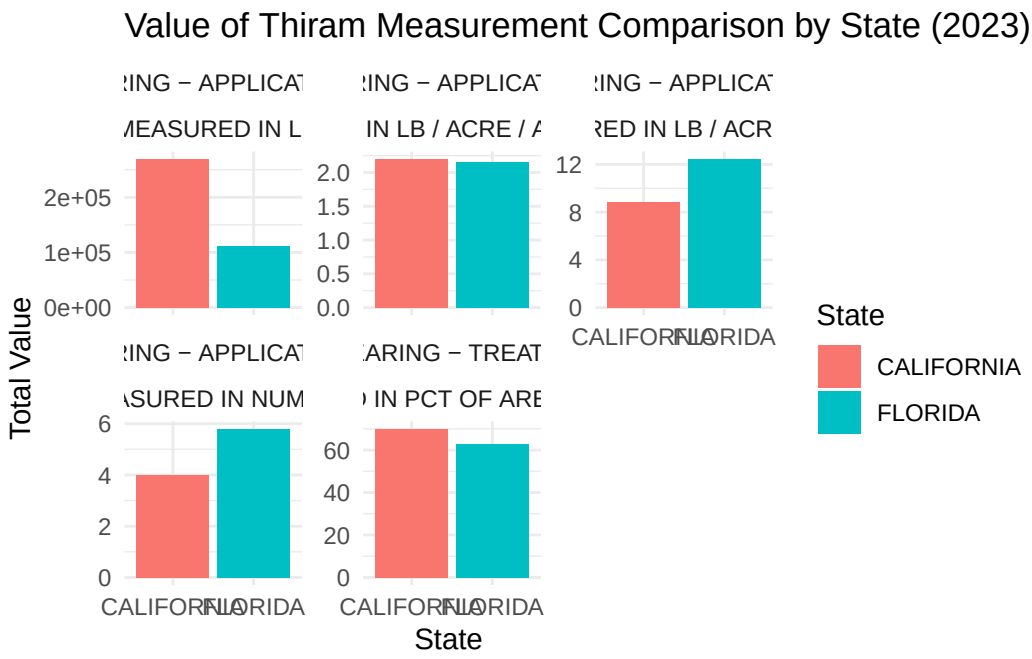
```
# A tibble: 20 x 6
# Groups:   Category, Item, State, Year [20]
   Category                  Item                          State  Year Value     n
   <chr>                     <chr>                         <chr> <dbl> <chr> <int>
 1 " BEARING - APPLICATIONS" " MEASURED IN LB"             CALI~  2021 96,3~     1
 2 " BEARING - APPLICATIONS" " MEASURED IN LB"             CALI~  2023 269,~     1
 3 " BEARING - APPLICATIONS" " MEASURED IN LB"             FLOR~  2021 142,~     1
 4 " BEARING - APPLICATIONS" " MEASURED IN LB"             FLOR~  2023 112,~     1
 5 " BEARING - APPLICATIONS" " MEASURED IN LB / ACRE / ~ CALI~  2021 2.144     1
 6 " BEARING - APPLICATIONS" " MEASURED IN LB / ACRE / ~ CALI~  2023 2.201     1
 7 " BEARING - APPLICATIONS" " MEASURED IN LB / ACRE / ~ FLOR~  2021 2.38      1
 8 " BEARING - APPLICATIONS" " MEASURED IN LB / ACRE / ~ FLOR~  2023 2.156     1
 9 " BEARING - APPLICATIONS" " MEASURED IN LB / ACRE / ~ CALI~  2021 5.029     1
10 " BEARING - APPLICATIONS" " MEASURED IN LB / ACRE / ~ CALI~  2023 8.873     1
11 " BEARING - APPLICATIONS" " MEASURED IN LB / ACRE / ~ FLOR~  2021 22.5~     1
12 " BEARING - APPLICATIONS" " MEASURED IN LB / ACRE / ~ FLOR~  2023 12.4~     1
13 " BEARING - APPLICATIONS" " MEASURED IN NUMBER"        CALI~  2021 2.3       1
14 " BEARING - APPLICATIONS" " MEASURED IN NUMBER"        CALI~  2023 4         1
15 " BEARING - APPLICATIONS" " MEASURED IN NUMBER"        FLOR~  2021 9.5       1
16 " BEARING - APPLICATIONS" " MEASURED IN NUMBER"        FLOR~  2023 5.8       1
17 " BEARING - TREATED"      " MEASURED IN PCT OF AREA ~ CALI~  2021 49        1
18 " BEARING - TREATED"      " MEASURED IN PCT OF AREA ~ CALI~  2023 70        1
19 " BEARING - TREATED"      " MEASURED IN PCT OF AREA ~ FLOR~  2021 61        1
20 " BEARING - TREATED"      " MEASURED IN PCT OF AREA ~ FLOR~  2023 63        1
```

```r
#Rearrange the data for bar chart
thiram_2021 <- thiram[str_detect(thiram$Year, regex("2021", ignore_case = TRUE)), ]
thiram_2023 <- thiram[str_detect(thiram$Year, regex("2023", ignore_case = TRUE)), ]
```

```r
#Ensure the value under Value column is numeric, so in graph it can provide the right relatio
thiram_2021$Value <- as.numeric(gsub(",", "", thiram_2021$Value))
thiram_2023$Value <- as.numeric(gsub(",", "", thiram_2023$Value))
#Same year, same domain category and item, but different state record comparison (code perfe

ggplot(thiram_2021, aes(x = State, y = Value, fill = State)) +
  geom_bar(stat = "identity", position = "dodge") +
  facet_wrap(~ Category + Item, scales = "free_y") +
  labs(title = "Value of Thiram Measurement Comparison by State (2021)",
       x = "State",
       y = "Total Value") +
  theme_minimal()
```




```r
ggplot(thiram_2023, aes(x = State, y = Value, fill = State)) +
  geom_bar(stat = "identity", position = "dodge") +
  facet_wrap(~ Category + Item, scales = "free_y") +
  labs(title = "Value of Thiram Measurement Comparison by State (2023)",
       x = "State",
       y = "Total Value") +
  theme_minimal()
```




```
# Remove intermediate data set
rm(thiram_2021,thiram_2023)
```

Chemical 3: Propiconazole

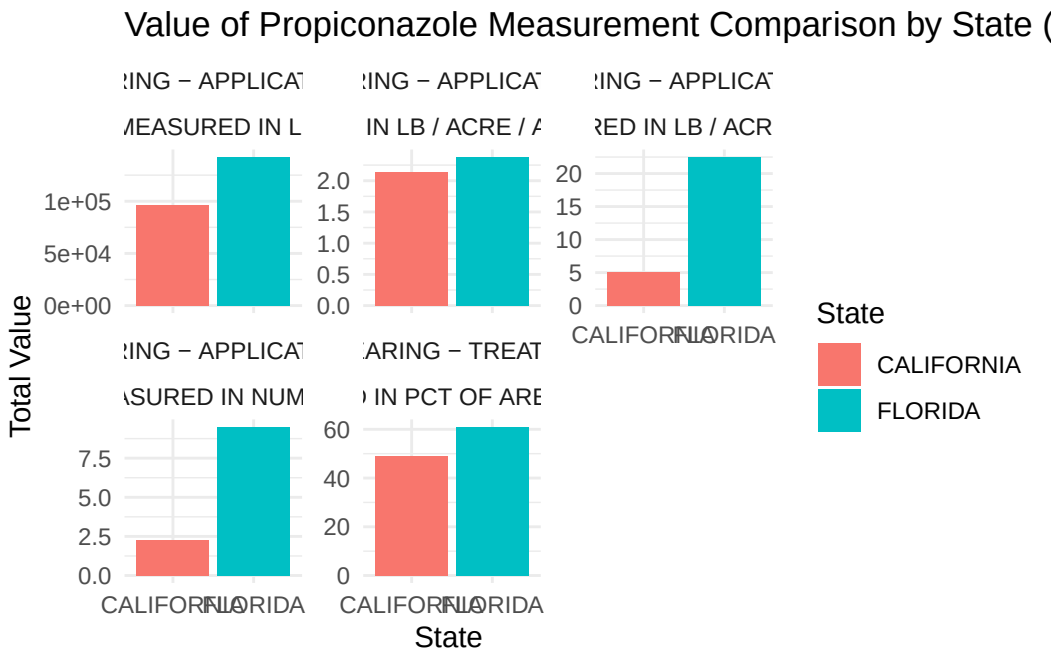
```
#Follows the same step as captan
#Filter rows form chemical_CA
ppz_CA <- chemical_CA[str_detect(chemical_CA$`Domain Category`, regex("thiram", ignore_case =

# Filter rows from chemical_FL
ppz_FL <- chemical_FL[str_detect(chemical_FL$`Domain Category`, regex("thiram", ignore_case =

#Get the thiram information data and gather them into one data set
ppz_combined <- rbind(ppz_CA, ppz_FL)

#remove intermediate data set
rm(ppz_CA)
rm(ppz_FL)

#Table of Comparison of different year by same state
ppz1 <- ppz_combined |> group_by(Category, Item, State, Year) |> count(Value)
```

```r
print(ppz1)
```

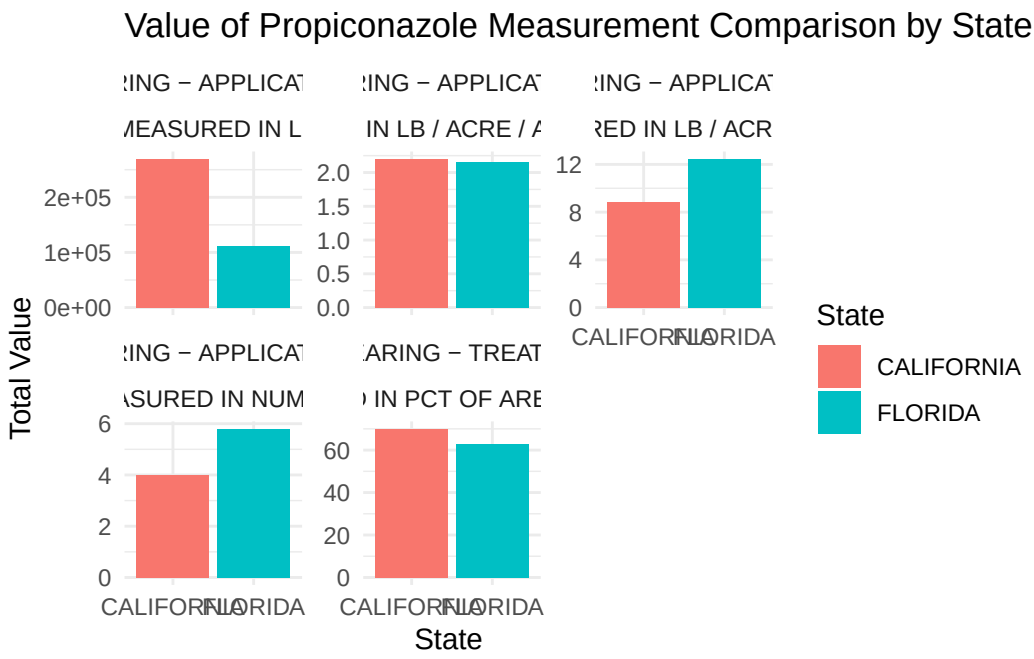
```
# A tibble: 20 x 6
# Groups:   Category, Item, State, Year [20]
   Category                  Item                          State  Year Value     n
   <chr>                     <chr>                         <chr> <dbl> <chr> <int>
 1 " BEARING - APPLICATIONS" " MEASURED IN LB"             CALI~  2021 96,3~     1
 2 " BEARING - APPLICATIONS" " MEASURED IN LB"             CALI~  2023 269,~     1
 3 " BEARING - APPLICATIONS" " MEASURED IN LB"             FLOR~  2021 142,~     1
 4 " BEARING - APPLICATIONS" " MEASURED IN LB"             FLOR~  2023 112,~     1
 5 " BEARING - APPLICATIONS" " MEASURED IN LB / ACRE / ~ CALI~  2021 2.144     1
 6 " BEARING - APPLICATIONS" " MEASURED IN LB / ACRE / ~ CALI~  2023 2.201     1
 7 " BEARING - APPLICATIONS" " MEASURED IN LB / ACRE / ~ FLOR~  2021 2.38      1
 8 " BEARING - APPLICATIONS" " MEASURED IN LB / ACRE / ~ FLOR~  2023 2.156     1
 9 " BEARING - APPLICATIONS" " MEASURED IN LB / ACRE / ~ CALI~  2021 5.029     1
10 " BEARING - APPLICATIONS" " MEASURED IN LB / ACRE / ~ CALI~  2023 8.873     1
11 " BEARING - APPLICATIONS" " MEASURED IN LB / ACRE / ~ FLOR~  2021 22.5~     1
12 " BEARING - APPLICATIONS" " MEASURED IN LB / ACRE / ~ FLOR~  2023 12.4~     1
13 " BEARING - APPLICATIONS" " MEASURED IN NUMBER"        CALI~  2021 2.3       1
14 " BEARING - APPLICATIONS" " MEASURED IN NUMBER"        CALI~  2023 4         1
15 " BEARING - APPLICATIONS" " MEASURED IN NUMBER"        FLOR~  2021 9.5       1
16 " BEARING - APPLICATIONS" " MEASURED IN NUMBER"        FLOR~  2023 5.8       1
17 " BEARING - TREATED"      " MEASURED IN PCT OF AREA ~ CALI~  2021 49        1
18 " BEARING - TREATED"      " MEASURED IN PCT OF AREA ~ CALI~  2023 70        1
19 " BEARING - TREATED"      " MEASURED IN PCT OF AREA ~ FLOR~  2021 61        1
20 " BEARING - TREATED"      " MEASURED IN PCT OF AREA ~ FLOR~  2023 63        1
```

```r
#Rearrange the data for bar chart
ppz_2021 <- ppz1[str_detect(ppz1$Year, regex("2021", ignore_case = TRUE)), ]
ppz_2023 <- ppz1[str_detect(ppz1$Year, regex("2023", ignore_case = TRUE)), ]

#Ensure the value under Value column is numeric, so in graph it can provide the right relatio
ppz_2021$Value <- as.numeric(gsub(",", "", ppz_2021$Value))
ppz_2023$Value <- as.numeric(gsub(",", "", ppz_2023$Value))
#Same year, same domain category and item, but different state record comparison (code perfe

ggplot(ppz_2021, aes(x = State, y = Value, fill = State)) +
  geom_bar(stat = "identity", position = "dodge") +
  facet_wrap(~ Category + Item, scales = "free_y") +
  labs(title = "Value of Propiconazole Measurement Comparison by State (2021)",
       x = "State",
```

```
       y = "Total Value") +
  theme_minimal()
```




```
ggplot(ppz_2023, aes(x = State, y = Value, fill = State)) +
  geom_bar(stat = "identity", position = "dodge") +
  facet_wrap(~ Category + Item, scales = "free_y") +
  labs(title = "Value of Propiconazole Measurement Comparison by State (2023)",
       x = "State",
       y = "Total Value") +
  theme_minimal()
```




```
# Remove intermediate data set
rm(ppz_2021,ppz_2023)
```