MA415 Midterm Project

PUBLISHED
March 6, 2025

EDA: Preparing Strawberry data for analysis

Due: March 21

As described in class, this document is a starter for the Midterm project.

Your assignment is to clean, organize, and explore the data. Turn in a report that describes how your work has set the stage for further analysis and model building.

The dataset contains strawberry farming data with details about conventional and organic cultivation. These data include information about chemicals used in strawberry farming, as well as sales, revenue and expense details.

While there is no “right answer” for this assignment, there are characteristics for the report that are essential. For example, sata visualization is critical. So is producing tabular presentations and document structure. Your target audience consists of analysts who may take the next steps with the data analysis and modeling.

Think of your report as a stage on which to showcase your ability to use R to work with data and produce professional reports. This is an opportunity to do some data storytelling.

Submit your report on or before March 21 using the Midterm portal on Blackboard.

Introduction: foundations

Before we begin to work with the strawberry data, let’s talk about how we will approach the work.

Data cleaning and organization

Cleaning and organizing data for analysis is an essential skill for data scientists. Serious data analyses must be presented with the data on which the results depend. The credibility of data analysis and modelling depends on the care taken in data preparation and organization.

References

In their handbook “An introduction to data cleaning with R” by Edwin de Jonge and Mark van der Loo, de Jonge and van der Loo go into detail about specific data cleaning isssues and how to handle them in R.

“Problems, Methods, and Challenges in Comprehensive Data Cleansing” by Heiko Müller and Johann-Christoph Freytag is a good companion to the de Jonge and van der Loo handbook, offering additional issues in their discussion.

Attitudes

Mechanistic descriptions of data cleaning methods are insufficient.

Data is the product (or by-product) of purposeful human activity

Much of the data used in analysis accessed on local databases or online which may create the impression that the data have been carefully curated. Beware. Data are produced by people for a purpose, with a point-of-view, and at a time and location that may affect the data. The provenance and lineage of the data are meta data you should include when reporting analysis. Data collection is purposeful human activity with all of the risks and weaknesses that are part of any purposeful human activity.

Data is language

Data has meaning. Data can be included in sentences related to the meaning of the data. Cleaning and organizing data should be informed by the meaning the data convey and how that meaning relates to the research you are doing do achieve this important result.

- Immerse yourself in the data. Put data into context.
- Visualize the data to find problems, confirm your understandings, and plan your data organization. People do a bad job of seeing meaningful patterns in data but a good job of seeing patterns of all kinds when data are rendered as plots. As you product and show visualizations, ask your self and those who view your presentations, “what do you see?” and “what do you wonder?”

Example: Strawberries

Public information

WHO says strawberries may not be so safe for you–2017March16

Pesticides + poison gases = cheap, year-round strawberries 2019March20

Multistate Outbreak of Hepatitis A Virus Infections Linked to Fresh Organic Strawberries-2022March5

Strawberry makes list of cancer-fighting foods-2023May31

What is the question?

- Where they are grown? By whom?
- Are they really loaded with carcinogenic poisons?
- Are they really good for your health? Bad for your health?
- Are organic strawberries carriers of deadly diseases?

- When I go to the market should I buy conventional or organic strawberries?

The data

The data set for this assignment has been selected from:

strawberries 2025march6

USDA NASS

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

Read the file

```
strawberry <- read_csv("strawb_mar6.csv",
                       col_names = TRUE,
                       show_col_types = FALSE)

source("my_functions.R")
```

Examine the data. How is it organized?

```
strawb <- strawberry |> drop_one_value_col()
```

```
# 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)
```

```
#|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
```

```
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
```

```
unique_sur <- s_survey |> show_unique(nrows = 10)

unique_cen <- s_census |> show_unique(nrows = 10)
```

```
strw_census <- s_census |> select(-`State ANSI`)

strw_survey <- s_survey |> select(-`State ANSI`, -`Week Ending`, -Period)

rm(s_census, s_survey, strawberry, strawb, items)
```

```
commod <- strw_census$Commodity |> unique()

#### split Data Item

strw_census <- strw_census |>
  separate_wider_delim(  cols = Commodity,
                         delim = ",",
                         names = c("INCOME",
                                   "NET CASH FARM",
                                   "STRAW"
                                               ),
                         too_many = "error",
                         names_sep = " ",
                         too_few = "align_start"
  )



inc <- strw_census$Fruit |> unique()

strw_census <- strw_census |>
  separate_wider_delim(  cols = Fruit,
                         delim = ",",
                         names = c("INCOME",
                                   "STRAWB"
                                               ),
                         too_many = "error",
                         too_few = "align_start"
  )
```

```
straw_cen_f <- strw_census |> filter(State == "FLORIDA")

straw_sur_f <- strw_survey |> filter(State == "FLORIDA")
straw_cen_c <- strw_census |> filter(State == "CALIFORNIA")
straw_sur_c <- strw_survey |> filter(State == "CALIFORNIA")

rm(strw_census, strw_survey, unique_cen, unique_sur)
```

Chemical Treatments Analysis

Begin with combining survey data

I want to use the survey data, specifically, in order to do an analysis that compares the usage of three chemicals across states and across years.

```
# A tibble: 6 × 9
   Year State      Fruit    Category Item  Metric Domain `Domain Category` Value
  <dbl> <chr>      <chr>    <chr>    <chr> <chr>  <chr>  <chr>             <chr>
1  2023 CALIFORNIA STRAWBE… " MEASU…  <NA> <NA>   TOTAL  NOT SPECIFIED     121
2  2023 CALIFORNIA STRAWBE… " FRESH… " ME… <NA>   TOTAL  NOT SPECIFIED     (D)
3  2023 CALIFORNIA STRAWBE… " PROCE… " ME… <NA>   TOTAL  NOT SPECIFIED     (D)
4  2023 CALIFORNIA STRAWBE…  <NA>     <NA> <NA>   TOTAL  NOT SPECIFIED     42,7…
5  2023 CALIFORNIA STRAWBE… " MEASU…  <NA> <NA>   CHEMI… CHEMICAL, FUNGIC… (D)
6  2023 CALIFORNIA STRAWBE… " MEASU…  <NA> <NA>   CHEMI… CHEMICAL, INSECT… (D)
```

I want to find what chemicals there are so I can narrow down what is suitable for analysis

```
# extract chemical names from the 'Domain Category' field
chemicals_df <- chemicals_df |>
  mutate(Chemical = str_extract(`Domain Category`, "\\((.*?)\\)")) |>
  mutate(Chemical = str_replace_all(Chemical, "[()]", "")) |>
  mutate(Chemical = str_trim(Chemical))

# unique chemicals to verify extraction
unique_chemicals <- unique(chemicals_df$Chemical)
print(unique_chemicals)
```

```
 [1] NA
 [2] "OXATHIAPIPROLIN = 128111"
 [3] "CYCLANILIPROLE = 26202"
 [4] "PERMETHRIN = 109701"
 [5] "ISARIA FUMOSOROSEA STRAIN FE 9901 = 115003"
 [6] "AZOXYSTROBIN = 128810"
 [7] "BACILLUS AMYLOLIQUEFACIENS STRAIN D747 = 16482"
 [8] "BACILLUS SUBTILIS = 6479"
 [9] "BLAD = 30006"
[10] "BORAX DECAHYDRATE = 11102"
[11] "BOSCALID = 128008"
[12] "BT SUBSP KURSTAKI EVB-113-19 = 6544"
[13] "CAPTAN = 81301"
[14] "CYFLUFENAMID = 555550"
[15] "CYPRODINIL = 288202"
[16] "DIFENOCONAZOLE = 128847"
[17] "FENHEXAMID = 90209"
[18] "FLUDIOXONIL = 71503"
[19] "FLUOPYRAM = 80302"
[20] "FLUXAPYROXAD = 138009"
[21] "FOSETYL-AL = 123301"
```

```
 [22] "ISOFETAMID = 270000"
 [23] "MEFENOXAM = 113502"
 [24] "MONO-POTASSIUM SALT = 76416"
 [25] "MYCLOBUTANIL = 128857"
 [26] "PENTHIOPYRAD = 90112"
 [27] "POLYOXIN D ZINC SALT = 230000"
 [28] "PROPICONAZOLE = 122101"
 [29] "PYDIFLUMETOFEN = 90110"
 [30] "PYRACLOSTROBIN = 99100"
 [31] "PYRIMETHANIL = 288201"
 [32] "QUINOLINE = 55459"
 [33] "SULFUR = 77501"
 [34] "TETRACONAZOLE = 120603"
 [35] "THIOPHANATE-METHYL = 102001"
 [36] "THIRAM = 79801"
 [37] "TOTAL"
 [38] "TRIFLOXYSTROBIN = 129112"
 [39] "TRIFLUMIZOLE = 128879"
 [40] "CARFENTRAZONE-ETHYL = 128712"
 [41] "FLUMIOXAZIN = 129034"
 [42] "OXYFLUORFEN = 111601"
 [43] "PENDIMETHALIN = 108501"
 [44] "ABAMECTIN = 122804"
 [45] "ACEQUINOCYL = 6329"
 [46] "ACETAMIPRID = 99050"
 [47] "AZADIRACHTIN = 121701"
 [48] "BEAUVERIA BASSIANA = 128924"
 [49] "BIFENAZATE = 586"
 [50] "BIFENTHRIN = 128825"
 [51] "BT KURSTAK ABTS-1857 = 6523"
 [52] "BT KURSTAKI ABTS-351 = 6522"
 [53] "BT KURSTAKI SA-11 = 6519"
 [54] "CANOLA OIL = 11332"
 [55] "CHLORANTRANILIPROLE = 90100"
 [56] "CHROMOBAC SUBTSUGAE PRAA4-1 CELLS AND SPENT MEDIA = 16329"
 [57] "CYANTRANILIPROLE = 90098"
 [58] "CYFLUMETOFEN = 138831"
 [59] "ETOXAZOLE = 107091"
 [60] "FENBUTATIN-OXIDE = 104601"
 [61] "FENPROPATHRIN = 127901"
 [62] "FENPYROXIMATE = 129131"
 [63] "FLONICAMID = 128016"
 [64] "FLUPYRADIFURONE = 122304"
 [65] "HEXYTHIAZOX = 128849"
 [66] "IMIDACLOPRID = 129099"
 [67] "LAMBDA-CYHALOTHRIN = 128897"
 [68] "MALATHION = 57701"
 [69] "METHOXYFENOZIDE = 121027"
 [70] "NALED = 34401"
 [71] "NEEM OIL = 25006"
 [72] "NEEM OIL, CLAR. HYD. = 25007"
```

```
 [73] "NOVALURON = 124002"
 [74] "PIPERONYL BUTOXIDE = 67501"
 [75] "PYRETHRINS = 69001"
 [76] "PYRIDABEN = 129105"
 [77] "SPINETORAM = 110007"
 [78] "SPINOSAD = 110003"
 [79] "THIAMETHOXAM = 60109"
 [80] "ACIBENZOLAR-S-METHYL = 61402"
 [81] "CAPSICUM OLEORESIN EXTRACT = 70704"
 [82] "CHLOROPICRIN = 81501"
 [83] "DICHLOROPROPENE = 29001"
 [84] "FLUTRIAFOL = 128940"
 [85] "GARLIC OIL = 128827"
 [86] "HYDROGEN PEROXIDE = 595"
 [87] "IRON PHOSPHATE = 34903"
 [88] "METALDEHYDE = 53001"
 [89] "METAM-POTASSIUM = 39002"
 [90] "METAM-SODIUM = 39003"
 [91] "PEROXYACETIC ACID = 63201"
 [92] "PSEUDOMONAS CHLORORAPHIS STRAIN AFS009 = 6800"
 [93] "REYNOUTRIA SACHALINE = 55809"
 [94] "NITROGEN"
 [95] "PHOSPHATE"
 [96] "POTASH"
 [97] "SULFUR"
 [98] "BACILLUS AMYLOLIQUEFAC F727 = 16489"
 [99] "BACILLUS AMYLOLIQUEFACIENS MBI 600 = 129082"
[100] "BACILLUS PUMILUS = 6485"
[101] "COPPER OCTANOATE = 23306"
[102] "POTASSIUM BICARBON. = 73508"
[103] "STREPTOMYCES LYDICUS = 6327"
[104] "GLYPHOSATE ISO. SALT = 103601"
[105] "GLYPHOSATE POT. SALT = 103613"
[106] "NAPROPAMIDE = 103001"
[107] "PARAQUAT = 61601"
[108] "BT KURSTAKI EG7841 = 6453"
[109] "BT SUB AIZAWAI GC-91 = 6426"
[110] "BUPROFEZIN = 275100"
[111] "BURKHOLDERIA A396 CELLS & MEDIA = 6534"
[112] "DIAZINON = 57801"
[113] "HELICOVERPA ZEA NPV = 107300"
[114] "PETROLEUM DISTILLATE = 63503"
[115] "POTASSIUM SALTS = 79021"
[116] "PYRIPROXYFEN = 129032"
[117] "SPIROMESIFEN = 24875"
[118] "CAPRIC ACID = 128955"
[119] "CAPRYLIC ACID = 128919"
[120] "MINERAL OIL = 63502"
[121] "PAECILOMYCES FUMOSOR = 115002"
[122] "POTASSIUM SILICATE = 72606"
[123] "PYRIOFENONE = 28828"
```

```
[124] "ZOXAMIDE = 101702"
[125] "METSULFURON-METHYL = 122010"
[126] "PENOXSULAM = 119031"
[127] "S-METOLACHLOR = 108800"
[128] "BETA-CYFLUTHRIN = 118831"
[129] "ETHYL 2E;4Z"
[130] "OXAMYL = 103801"
[131] "CUPRAMMONIUM ACETATE = 36011"
[132] "DODECADIEN-1-OL = 129028"
[133] "FLUENSULFONE = 50410"
[134] "GIBBERELLIC ACID = 43801"
[135] "CHLOROTHALONIL = 81901"
[136] "COPPER CHLORIDE HYD. = 23501"
[137] "COPPER HYDROXIDE = 23401"
[138] "CYMOXANIL = 129106"
[139] "FAMOXADONE = 113202"
[140] "IPRODIONE = 109801"
[141] "MANCOZEB = 14504"
[142] "2,4-D, DIMETH. SALT = 30019"
[143] "CLETHODIM = 121011"
[144] "METHOMYL = 90301"
[145] "SULFOXAFLOR = 5210"
[146] "CYTOKININS = 116801"
[147] "INDOLEBUTYRIC ACID = 46701"
[148] "COPPER ETHANOLAMINE = 24409"
[149] "DIMETHENAMID = 129051"
[150] "FLUROXYPYR 1-MHE = 128968"
[151] "HALOSULFURON-METHYL = 128721"
[152] "KANTOR = 129108"
[153] "CARBARYL = 56801"
[154] "FENAZAQUIN = 44501"
[155] "ETHEPHON = 99801"
```

Filtering for valid chemicals

```
# A tibble: 10 × 10
    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 more variable: Chemical <chr>
```

To compare usage, I need to remove values that have (D) or (NA) so that there are not any unsuitable entries being used. Since we are only comparing 3 different chemicals, I would like all 3 of them to have number values to get a meaningful insight.

```
# A tibble: 10 × 10
    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… 3,300
 2  2023 CALIFORNIA STRAWB… " BEARI… " ME… <NA>   CHEMI… CHEMICAL, FUNGIC… 2,800
 3  2023 CALIFORNIA STRAWB… " BEARI… " ME… <NA>   CHEMI… CHEMICAL, FUNGIC… 6,600
 4  2023 CALIFORNIA STRAWB… " BEARI… " ME… <NA>   CHEMI… CHEMICAL, FUNGIC… 603,…
 5  2023 CALIFORNIA STRAWB… " BEARI… " ME… <NA>   CHEMI… CHEMICAL, FUNGIC… 30,3…
 6  2023 CALIFORNIA STRAWB… " BEARI… " ME… <NA>   CHEMI… CHEMICAL, FUNGIC… 8,600
 7  2023 CALIFORNIA STRAWB… " BEARI… " ME… <NA>   CHEMI… CHEMICAL, FUNGIC… 22,4…
 8  2023 CALIFORNIA STRAWB… " BEARI… " ME… <NA>   CHEMI… CHEMICAL, FUNGIC… 14,6…
 9  2023 CALIFORNIA STRAWB… " BEARI… " ME… <NA>   CHEMI… CHEMICAL, FUNGIC… 7,100
10  2023 CALIFORNIA STRAWB… " BEARI… " ME… <NA>   CHEMI… CHEMICAL, FUNGIC… 7,200
# ℹ 1 more variable: Chemical <chr>
```

Here, I want to isolate the name of the chemicals being used from the 'Chemical' column in chemicals_df from the numbers so I can later filter through based just on chemical name.

```
# A tibble: 71 × 1
   Chemical
   <chr>
 1 AZOXYSTROBIN
 2 BORAX DECAHYDRATE
 3 BOSCALID
 4 CAPTAN
 5 CYPRODINIL
 6 FENHEXAMID
 7 FLUDIOXONIL
 8 FLUOPYRAM
 9 FLUXAPYROXAD
10 MEFENOXAM
# ℹ 61 more rows
```

Here, I am looking chemicals that will fit the criteria that I want for my analysis. I want chemicals that have records both in California and Florida so there is not any missing data when making visualizations.

```
[1] "ABAMECTIN"   "CAPTAN"      "CYPRODINIL"  "FLUDIOXONIL" "NOVALURON"
[6] "THIRAM"      "TOTAL"
```

Now that I have a list of viable options, I chose Abamectin, Fludioxonil, and Thiram to compare across the two states. This is a summary, mostly to

make sure that I have 12 records like I expect.

```
# no commas and converting to numeric
chemicals_df <- chemicals_df |>
  mutate(Value = as.numeric(gsub(",", "", Value)))
```

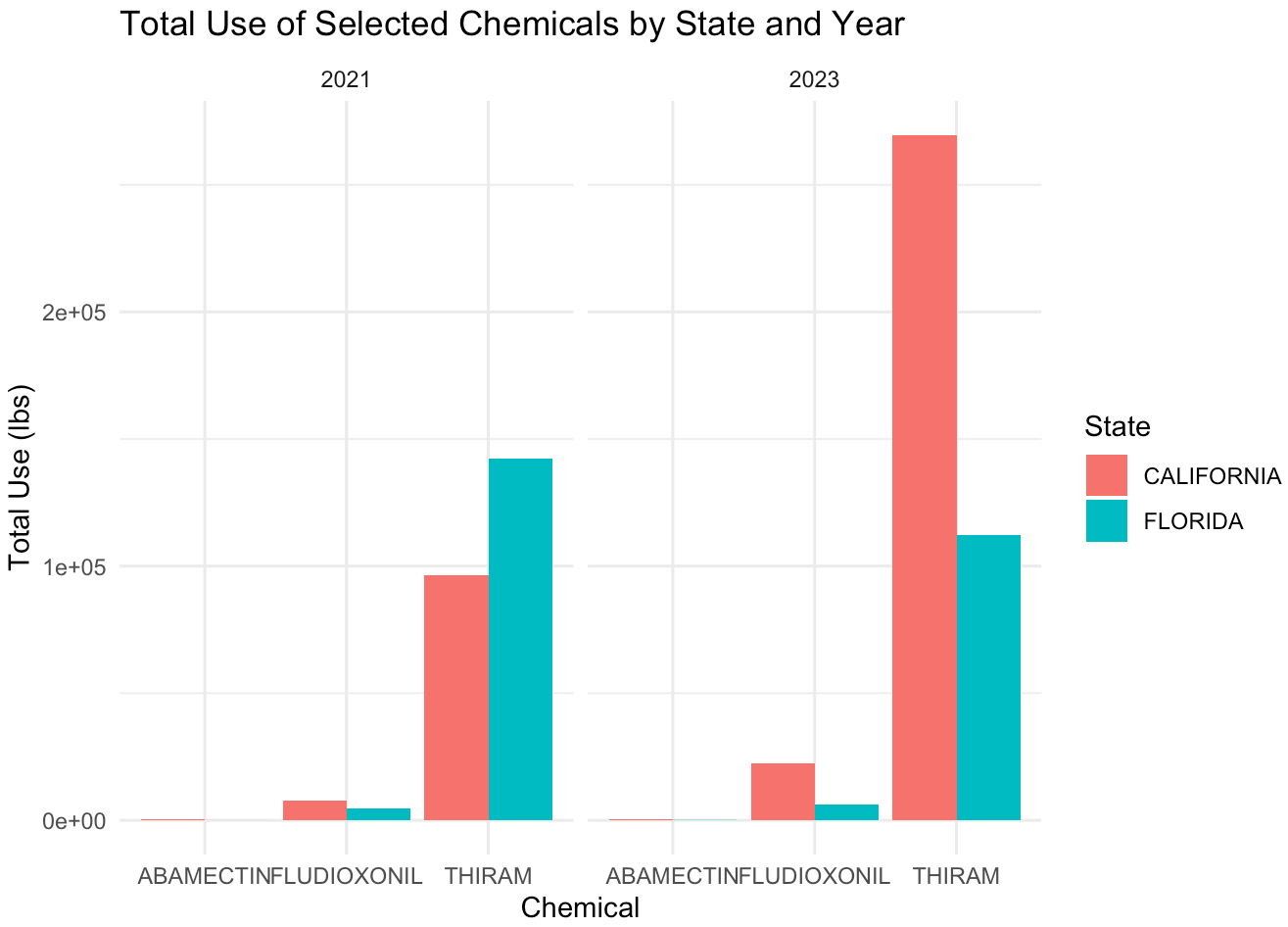
```
Warning: There was 1 warning in `mutate()`.
ℹ In argument: `Value = as.numeric(gsub(",", "", Value))`.
Caused by warning:
! NAs introduced by coercion
```

```
selected_chems <- c("ABAMECTIN", "FLUDIOXONIL", "THIRAM")

# summarize by year and state
chemical_summary <- chemicals_df |>
  filter(Chemical %in% selected_chems) |>
  group_by(Year, State, Chemical) |>
  summarize(Total_Use_Lbs = sum(Value, na.rm = TRUE), .groups = "drop")

print(chemical_summary)
```

```
# A tibble: 12 × 4
    Year State      Chemical    Total_Use_Lbs
   <dbl> <chr>      <chr>              <dbl>
 1  2021 CALIFORNIA ABAMECTIN          221.
 2  2021 CALIFORNIA FLUDIOXONIL       7551.
 3  2021 CALIFORNIA THIRAM           96358.
 4  2021 FLORIDA    ABAMECTIN           9.15
 5  2021 FLORIDA    FLUDIOXONIL       4570.
 6  2021 FLORIDA    THIRAM          142495.
 7  2023 CALIFORNIA ABAMECTIN          331.
 8  2023 CALIFORNIA FLUDIOXONIL      22491.
 9  2023 CALIFORNIA THIRAM          269585.
10  2023 FLORIDA    ABAMECTIN          115.
11  2023 FLORIDA    FLUDIOXONIL       6195.
12  2023 FLORIDA    THIRAM          112183.
```

Total Use of Selected Chemicals by State and Year

```
ggplot(chemical_summary, aes(x = Chemical, y = Total_Use_Lbs, fill = State)) +
  geom_bar(stat = "identity", position = "dodge") +
  facet_wrap(~ Year) +
  labs(title = "Total Use of Selected Chemicals by State and Year",
       x = "Chemical",
       y = "Total Use (lbs)") +
  theme_minimal()
```



Chemical Usage from 2021 to 2023 by State

Useage from 2021 to 2023 can be seen here. It seens like in both California and Florida, useage for Abamectin and Fludioxonil went up between from 2021 to 2023. However, Thiram useage went up for California went up while it went down for Florida. I hypothesize this is due to an alternative fungicide that does a better job that replaced Florida's use of Thiram.

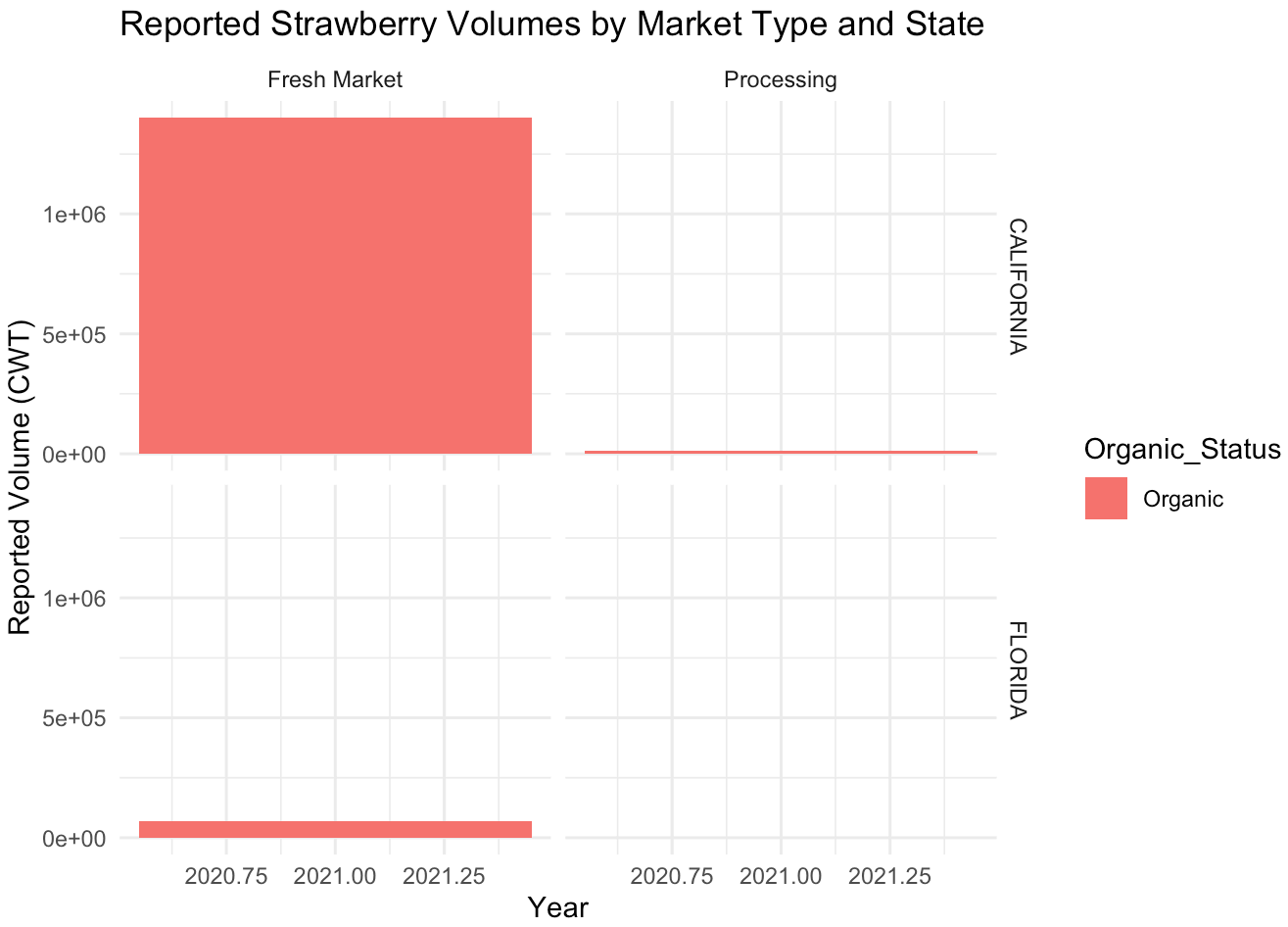
```
ggplot(chemical_summary, aes(x = Year, y = Total_Use_Lbs, color = State, group = State))
  geom_line(size = 1) +
  geom_point(size = 3) +
  geom_text(aes(label = round(Total_Use_Lbs, 1)), vjust = -0.5, size = 3) +
  facet_wrap(~ Chemical, scales = "free_y") +
  labs(title = "Chemical Usage from 2021 to 2023 by State",
       x = "Year",
       y = "Total Use (lbs)") +
  theme_minimal() +
  theme(axis.text.x = element_text(angle = 45, hjust = 1, size = 10),
        axis.text.y = element_text(size = 10),
        plot.title = element_text(size = 14, face = "bold"))
```

```
Warning: Using `size` aesthetic for lines was deprecated in ggplot2 3.4.0.
ℹ Please use `linewidth` instead.
```

Analysis on Production & Sales of Different Types of Strawberries

```
Attaching package: 'scales'

The following object is masked from 'package:purrr':

    discard

The following object is masked from 'package:readr':

    col_factor
```

```r
ggplot(volume_summary, aes(x = Year, y = Reported_Volume, fill = Organic_Status)) +
  geom_col(position = "dodge") +
  facet_grid(State ~ Market_Type) +
  labs(
    title = "Reported Strawberry Volumes by Market Type and State",
    y = "Reported Volume (CWT)",
    x = "Year"
  ) +
  theme_minimal()
```



Based on the USDA census data, strawberries for both California and Florida were mostly organic production. Florida shows more distribution across different categories including processing.