Introduction

The goal of this project was to clean, organize, and explore the USDA strawberry dataset in and prepare it for analysis. This phase focused on separating organic and non-organic data and further cleaning chemical data used in strawberry cultivation by splitting it into three relevant columns: chemical use, chemical name, and chemical code.

Intital Data Analysis

I started off by loading the dataset and checking the structure

```
# Load necessary libraries
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(ggplot2)
library(tidyr)

# Loading the dataset
strawberry_data <- read.csv('strawberries25_v3.csv')

# Displaying the structure of the dataset
str(strawberry_data)
```

```
## 'data.frame':    12669 obs. of  21 variables:
##  $ Program         : chr  "CENSUS" "CENSUS" "CENSUS" "CENSUS" ...
##  $ Year            : int  2022 2022 2022 2022 2022 2022 2022 2022 2022 2022 ...
##  $ Period          : chr  "YEAR" "YEAR" "YEAR" "YEAR" ...
##  $ Week.Ending     : logi  NA NA NA NA NA NA ...
##  $ Geo.Level       : chr  "COUNTY" "COUNTY" "COUNTY" "COUNTY" ...
##  $ State           : chr  "ALABAMA" "ALABAMA" "ALABAMA" "ALABAMA" ...
##  $ State.ANSI      : int  1 1 1 1 1 1 1 1 1 1 ...
##  $ Ag.District     : chr  "BLACK BELT" "BLACK BELT" "BLACK BELT" "BLACK BELT" ...
##  $ Ag.District.Code: int  40 40 40 40 40 40 40 40 40 40 ...
##  $ County          : chr  "BULLOCK" "BULLOCK" "BULLOCK" "BULLOCK" ...
##  $ County.ANSI     : int  11 11 11 11 11 11 101 101 101 101 ...
##  $ Zip.Code        : logi  NA NA NA NA NA NA ...
##  $ Region          : logi  NA NA NA NA NA NA ...
##  $ watershed_code  : int  0 0 0 0 0 0 0 0 0 0 ...
##  $ Watershed       : logi  NA NA NA NA NA NA ...
##  $ Commodity       : chr  "STRAWBERRIES" "STRAWBERRIES" "STRAWBERRIES" "STRAWBERRIES" ...
```

```
##  $ Data.Item       : chr  "STRAWBERRIES - ACRES BEARING" "STRAWBERRIES - ACRES GROWN" "STRAWBERRIES -
##  $ Domain          : chr  "TOTAL" "TOTAL" "TOTAL" "TOTAL" ...
##  $ Domain.Category : chr  "NOT SPECIFIED" "NOT SPECIFIED" "NOT SPECIFIED" "NOT SPECIFIED" ...
##  $ Value           : chr  " (D)" "3" " (D)" "1" ...
##  $ CV....          : chr  "(D)" "15.7" "(D)" "(L)" ...
```

```r
colnames(strawberry_data)
```

```
##  [1] "Program"          "Year"             "Period"           "Week.Ending"
##  [5] "Geo.Level"        "State"            "State.ANSI"       "Ag.District"
##  [9] "Ag.District.Code" "County"           "County.ANSI"      "Zip.Code"
## [13] "Region"           "watershed_code"   "Watershed"        "Commodity"
## [17] "Data.Item"        "Domain"           "Domain.Category"  "Value"
## [21] "CV...."
```

We viewed all the columns and understand the data based on the above. Based on this, our columns of interest are- Program: Type of data collection program (e.g., Census, Survey). Domain and Domain Category: Fields that describe whether the data is organic or non-organic, and include information about chemicals.

```r
# To Get an overview of the unique values in the "Domain" column
unique_domains <- unique(strawberry_data$Domain)
print(unique_domains)
```

```
## [1] "TOTAL"                 "AREA GROWN"            "ORGANIC STATUS"
## [4] "CHEMICAL, FUNGICIDE"   "CHEMICAL, INSECTICIDE" "CHEMICAL, OTHER"
## [7] "CHEMICAL, HERBICIDE"   "FERTILIZER"
```

```r
# Get unique values from the "Domain.Category" column
unique_domain_categories <- unique(strawberry_data$Domain.Category)

print(unique_domain_categories)
```

```
##   [1] "NOT SPECIFIED"
##   [2] "AREA GROWN: (0.1 TO 0.9 ACRES)"
##   [3] "AREA GROWN: (1.0 TO 4.9 ACRES)"
##   [4] "AREA GROWN: (100 OR MORE ACRES)"
##   [5] "AREA GROWN: (15.0 TO 24.9 ACRES)"
##   [6] "AREA GROWN: (25.0 TO 49.9 ACRES)"
##   [7] "AREA GROWN: (5.0 TO 14.9 ACRES)"
##   [8] "AREA GROWN: (50.0 TO 99.9 ACRES)"
##   [9] "ORGANIC STATUS: (NOP USDA CERTIFIED)"
##  [10] "CHEMICAL, FUNGICIDE: (OXATHIAPIPROLIN = 128111)"
##  [11] "CHEMICAL, INSECTICIDE: (CYCLANILIPROLE = 26202)"
##  [12] "CHEMICAL, INSECTICIDE: (PERMETHRIN = 109701)"
##  [13] "CHEMICAL, OTHER: (ISARIA FUMOSOROSEA STRAIN FE 9901 = 115003)"
##  [14] "CHEMICAL, FUNGICIDE: (AZOXYSTROBIN = 128810)"
##  [15] "CHEMICAL, FUNGICIDE: (BACILLUS AMYLOLIQUEFACIENS STRAIN D747 = 16482)"
##  [16] "CHEMICAL, FUNGICIDE: (BACILLUS SUBTILIS = 6479)"
##  [17] "CHEMICAL, FUNGICIDE: (BLAD = 30006)"
##  [18] "CHEMICAL, FUNGICIDE: (BORAX DECAHYDRATE = 11102)"
##  [19] "CHEMICAL, FUNGICIDE: (BOSCALID = 128008)"
```

```
##  [20] "CHEMICAL, FUNGICIDE: (BT SUBSP KURSTAKI EVB-113-19 = 6544)"
##  [21] "CHEMICAL, FUNGICIDE: (CAPTAN = 81301)"
##  [22] "CHEMICAL, FUNGICIDE: (CYFLUFENAMID = 555550)"
##  [23] "CHEMICAL, FUNGICIDE: (CYPRODINIL = 288202)"
##  [24] "CHEMICAL, FUNGICIDE: (DIFENOCONAZOLE = 128847)"
##  [25] "CHEMICAL, FUNGICIDE: (FENHEXAMID = 90209)"
##  [26] "CHEMICAL, FUNGICIDE: (FLUDIOXONIL = 71503)"
##  [27] "CHEMICAL, FUNGICIDE: (FLUOPYRAM = 80302)"
##  [28] "CHEMICAL, FUNGICIDE: (FLUXAPYROXAD = 138009)"
##  [29] "CHEMICAL, FUNGICIDE: (FOSETYL-AL = 123301)"
##  [30] "CHEMICAL, FUNGICIDE: (ISOFETAMID = 270000)"
##  [31] "CHEMICAL, FUNGICIDE: (MEFENOXAM = 113502)"
##  [32] "CHEMICAL, FUNGICIDE: (MONO-POTASSIUM SALT = 76416)"
##  [33] "CHEMICAL, FUNGICIDE: (MYCLOBUTANIL = 128857)"
##  [34] "CHEMICAL, FUNGICIDE: (PENTHIOPYRAD = 90112)"
##  [35] "CHEMICAL, FUNGICIDE: (POLYOXIN D ZINC SALT = 230000)"
##  [36] "CHEMICAL, FUNGICIDE: (PROPICONAZOLE = 122101)"
##  [37] "CHEMICAL, FUNGICIDE: (PYDIFLUMETOFEN = 90110)"
##  [38] "CHEMICAL, FUNGICIDE: (PYRACLOSTROBIN = 99100)"
##  [39] "CHEMICAL, FUNGICIDE: (PYRIMETHANIL = 288201)"
##  [40] "CHEMICAL, FUNGICIDE: (QUINOLINE = 55459)"
##  [41] "CHEMICAL, FUNGICIDE: (SULFUR = 77501)"
##  [42] "CHEMICAL, FUNGICIDE: (TETRACONAZOLE = 120603)"
##  [43] "CHEMICAL, FUNGICIDE: (THIOPHANATE-METHYL = 102001)"
##  [44] "CHEMICAL, FUNGICIDE: (THIRAM = 79801)"
##  [45] "CHEMICAL, FUNGICIDE: (TOTAL)"
##  [46] "CHEMICAL, FUNGICIDE: (TRIFLOXYSTROBIN = 129112)"
##  [47] "CHEMICAL, FUNGICIDE: (TRIFLUMIZOLE = 128879)"
##  [48] "CHEMICAL, HERBICIDE: (CARFENTRAZONE-ETHYL = 128712)"
##  [49] "CHEMICAL, HERBICIDE: (FLUMIOXAZIN = 129034)"
##  [50] "CHEMICAL, HERBICIDE: (OXYFLUORFEN = 111601)"
##  [51] "CHEMICAL, HERBICIDE: (PENDIMETHALIN = 108501)"
##  [52] "CHEMICAL, HERBICIDE: (TOTAL)"
##  [53] "CHEMICAL, INSECTICIDE: (ABAMECTIN = 122804)"
##  [54] "CHEMICAL, INSECTICIDE: (ACEQUINOCYL = 6329)"
##  [55] "CHEMICAL, INSECTICIDE: (ACETAMIPRID = 99050)"
##  [56] "CHEMICAL, INSECTICIDE: (AZADIRACHTIN = 121701)"
##  [57] "CHEMICAL, INSECTICIDE: (BEAUVERIA BASSIANA = 128924)"
##  [58] "CHEMICAL, INSECTICIDE: (BIFENAZATE = 586)"
##  [59] "CHEMICAL, INSECTICIDE: (BIFENTHRIN = 128825)"
##  [60] "CHEMICAL, INSECTICIDE: (BT KURSTAK ABTS-1857 = 6523)"
##  [61] "CHEMICAL, INSECTICIDE: (BT KURSTAKI ABTS-351 = 6522)"
##  [62] "CHEMICAL, INSECTICIDE: (BT KURSTAKI SA-11 = 6519)"
##  [63] "CHEMICAL, INSECTICIDE: (CANOLA OIL = 11332)"
##  [64] "CHEMICAL, INSECTICIDE: (CHLORANTRANILIPROLE = 90100)"
##  [65] "CHEMICAL, INSECTICIDE: (CHROMOBAC SUBTSUGAE PRAA4-1 CELLS AND SPENT MEDIA = 16329)"
##  [66] "CHEMICAL, INSECTICIDE: (CYANTRANILIPROLE = 90098)"
##  [67] "CHEMICAL, INSECTICIDE: (CYFLUMETOFEN = 138831)"
##  [68] "CHEMICAL, INSECTICIDE: (ETOXAZOLE = 107091)"
##  [69] "CHEMICAL, INSECTICIDE: (FENBUTATIN-OXIDE = 104601)"
##  [70] "CHEMICAL, INSECTICIDE: (FENPROPATHRIN = 127901)"
##  [71] "CHEMICAL, INSECTICIDE: (FENPYROXIMATE = 129131)"
##  [72] "CHEMICAL, INSECTICIDE: (FLONICAMID = 128016)"
##  [73] "CHEMICAL, INSECTICIDE: (FLUPYRADIFURONE = 122304)"
```

```
##  [74] "CHEMICAL, INSECTICIDE: (HEXYTHIAZOX = 128849)"
##  [75] "CHEMICAL, INSECTICIDE: (IMIDACLOPRID = 129099)"
##  [76] "CHEMICAL, INSECTICIDE: (LAMBDA-CYHALOTHRIN = 128897)"
##  [77] "CHEMICAL, INSECTICIDE: (MALATHION = 57701)"
##  [78] "CHEMICAL, INSECTICIDE: (METHOXYFENOZIDE = 121027)"
##  [79] "CHEMICAL, INSECTICIDE: (NALED = 34401)"
##  [80] "CHEMICAL, INSECTICIDE: (NEEM OIL = 25006)"
##  [81] "CHEMICAL, INSECTICIDE: (NEEM OIL, CLAR. HYD. = 25007)"
##  [82] "CHEMICAL, INSECTICIDE: (NOVALURON = 124002)"
##  [83] "CHEMICAL, INSECTICIDE: (PIPERONYL BUTOXIDE = 67501)"
##  [84] "CHEMICAL, INSECTICIDE: (PYRETHRINS = 69001)"
##  [85] "CHEMICAL, INSECTICIDE: (PYRIDABEN = 129105)"
##  [86] "CHEMICAL, INSECTICIDE: (SPINETORAM = 110007)"
##  [87] "CHEMICAL, INSECTICIDE: (SPINOSAD = 110003)"
##  [88] "CHEMICAL, INSECTICIDE: (THIAMETHOXAM = 60109)"
##  [89] "CHEMICAL, INSECTICIDE: (TOTAL)"
##  [90] "CHEMICAL, OTHER: (ACIBENZOLAR-S-METHYL = 61402)"
##  [91] "CHEMICAL, OTHER: (CAPSICUM OLEORESIN EXTRACT = 70704)"
##  [92] "CHEMICAL, OTHER: (CHLOROPICRIN = 81501)"
##  [93] "CHEMICAL, OTHER: (DICHLOROPROPENE = 29001)"
##  [94] "CHEMICAL, OTHER: (FLUTRIAFOL = 128940)"
##  [95] "CHEMICAL, OTHER: (GARLIC OIL = 128827)"
##  [96] "CHEMICAL, OTHER: (HYDROGEN PEROXIDE = 595)"
##  [97] "CHEMICAL, OTHER: (IRON PHOSPHATE = 34903)"
##  [98] "CHEMICAL, OTHER: (METALDEHYDE = 53001)"
##  [99] "CHEMICAL, OTHER: (METAM-POTASSIUM = 39002)"
## [100] "CHEMICAL, OTHER: (METAM-SODIUM = 39003)"
## [101] "CHEMICAL, OTHER: (PEROXYACETIC ACID = 63201)"
## [102] "CHEMICAL, OTHER: (PSEUDOMONAS CHLORORAPHIS STRAIN AFS009 = 6800)"
## [103] "CHEMICAL, OTHER: (REYNOUTRIA SACHALINE = 55809)"
## [104] "CHEMICAL, OTHER: (TOTAL)"
## [105] "FERTILIZER: (NITROGEN)"
## [106] "FERTILIZER: (PHOSPHATE)"
## [107] "FERTILIZER: (POTASH)"
## [108] "FERTILIZER: (SULFUR)"
## [109] "CHEMICAL, FUNGICIDE: (PYRIOFENONE = 28828)"
## [110] "CHEMICAL, FUNGICIDE: (ZOXAMIDE = 101702)"
## [111] "CHEMICAL, HERBICIDE: (METSULFURON-METHYL = 122010)"
## [112] "CHEMICAL, HERBICIDE: (PENOXSULAM = 119031)"
## [113] "CHEMICAL, HERBICIDE: (S-METOLACHLOR = 108800)"
## [114] "CHEMICAL, INSECTICIDE: (BETA-CYFLUTHRIN = 118831)"
## [115] "CHEMICAL, INSECTICIDE: (ETHYL (2E;4Z)-DECADIENOATE = 144022)"
## [116] "CHEMICAL, INSECTICIDE: (OXAMYL = 103801)"
## [117] "CHEMICAL, OTHER: (CUPRAMMONIUM ACETATE = 36011)"
## [118] "CHEMICAL, OTHER: (DODECADIEN-1-OL = 129028)"
## [119] "CHEMICAL, OTHER: (FLUENSULFONE = 50410)"
## [120] "CHEMICAL, OTHER: (GIBBERELLIC ACID = 43801)"
## [121] "CHEMICAL, FUNGICIDE: (BACILLUS AMYLOLIQUEFAC F727 = 16489)"
## [122] "CHEMICAL, FUNGICIDE: (CHLOROTHALONIL = 81901)"
## [123] "CHEMICAL, FUNGICIDE: (COPPER CHLORIDE HYD. = 23501)"
## [124] "CHEMICAL, FUNGICIDE: (COPPER HYDROXIDE = 23401)"
## [125] "CHEMICAL, FUNGICIDE: (CYMOXANIL = 129106)"
## [126] "CHEMICAL, FUNGICIDE: (FAMOXADONE = 113202)"
## [127] "CHEMICAL, FUNGICIDE: (IPRODIONE = 109801)"
```

```
## [128] "CHEMICAL, FUNGICIDE: (MANCOZEB = 14504)"
## [129] "CHEMICAL, HERBICIDE: (2,4-D, DIMETH. SALT = 30019)"
## [130] "CHEMICAL, HERBICIDE: (CLETHODIM = 121011)"
## [131] "CHEMICAL, HERBICIDE: (GLYPHOSATE ISO. SALT = 103601)"
## [132] "CHEMICAL, HERBICIDE: (PARAQUAT = 61601)"
## [133] "CHEMICAL, INSECTICIDE: (DIAZINON = 57801)"
## [134] "CHEMICAL, INSECTICIDE: (METHOMYL = 90301)"
## [135] "CHEMICAL, INSECTICIDE: (SULFOXAFLOR = 5210)"
## [136] "CHEMICAL, OTHER: (CYTOKININS = 116801)"
## [137] "CHEMICAL, OTHER: (INDOLEBUTYRIC ACID = 46701)"
## [138] "CHEMICAL, FUNGICIDE: (BACILLUS AMYLOLIQUEFACIENS MBI 600 = 129082)"
## [139] "CHEMICAL, FUNGICIDE: (BACILLUS PUMILUS = 6485)"
## [140] "CHEMICAL, FUNGICIDE: (COPPER OCTANOATE = 23306)"
## [141] "CHEMICAL, FUNGICIDE: (POTASSIUM BICARBON. = 73508)"
## [142] "CHEMICAL, FUNGICIDE: (STREPTOMYCES LYDICUS = 6327)"
## [143] "CHEMICAL, HERBICIDE: (GLYPHOSATE POT. SALT = 103613)"
## [144] "CHEMICAL, HERBICIDE: (NAPROPAMIDE = 103001)"
## [145] "CHEMICAL, INSECTICIDE: (BT KURSTAKI EG7841 = 6453)"
## [146] "CHEMICAL, INSECTICIDE: (BT SUB AIZAWAI GC-91 = 6426)"
## [147] "CHEMICAL, INSECTICIDE: (BUPROFEZIN = 275100)"
## [148] "CHEMICAL, INSECTICIDE: (BURKHOLDERIA A396 CELLS & MEDIA = 6534)"
## [149] "CHEMICAL, INSECTICIDE: (HELICOVERPA ZEA NPV = 107300)"
## [150] "CHEMICAL, INSECTICIDE: (PETROLEUM DISTILLATE = 63503)"
## [151] "CHEMICAL, INSECTICIDE: (POTASSIUM SALTS = 79021)"
## [152] "CHEMICAL, INSECTICIDE: (PYRIPROXYFEN = 129032)"
## [153] "CHEMICAL, INSECTICIDE: (SPIROMESIFEN = 24875)"
## [154] "CHEMICAL, OTHER: (CAPRIC ACID = 128955)"
## [155] "CHEMICAL, OTHER: (CAPRYLIC ACID = 128919)"
## [156] "CHEMICAL, OTHER: (MINERAL OIL = 63502)"
## [157] "CHEMICAL, OTHER: (PAECILOMYCES FUMOSOR = 115002)"
## [158] "CHEMICAL, OTHER: (POTASSIUM SILICATE = 72606)"
## [159] "CHEMICAL, HERBICIDE: (COPPER ETHANOLAMINE = 24409)"
## [160] "CHEMICAL, HERBICIDE: (DIMETHENAMID = 129051)"
## [161] "CHEMICAL, HERBICIDE: (FLUROXYPYR 1-MHE = 128968)"
## [162] "CHEMICAL, HERBICIDE: (HALOSULFURON-METHYL = 128721)"
## [163] "CHEMICAL, HERBICIDE: (KANTOR = 129108)"
## [164] "CHEMICAL, INSECTICIDE: (CARBARYL = 56801)"
## [165] "CHEMICAL, INSECTICIDE: (FENAZAQUIN = 44501)"
## [166] "CHEMICAL, OTHER: (ETHEPHON = 99801)"
## [167] "CHEMICAL, FUNGICIDE: (BACILLUS SUBT. GB03 = 129068)"
## [168] "CHEMICAL, FUNGICIDE: (TRICHODERMA HARZ. = 119202)"
## [169] "CHEMICAL, HERBICIDE: (GLUFOSINATE-AMMONIUM = 128850)"
## [170] "CHEMICAL, HERBICIDE: (SULFENTRAZONE = 129081)"
## [171] "CHEMICAL, INSECTICIDE: (CHLORPYRIFOS = 59101)"
## [172] "CHEMICAL, INSECTICIDE: (SOYBEAN OIL = 31605)"
## [173] "CHEMICAL, INSECTICIDE: (ZETA-CYPERMETHRIN = 129064)"
## [174] "CHEMICAL, OTHER: (AUREOBASIDIUM PULLULANS DSM 14940 = 46010)"
## [175] "CHEMICAL, OTHER: (AUREOBASIDIUM PULLULANS DSM 14941 = 36010)"
## [176] "CHEMICAL, OTHER: (BT KURSTAKI SA-12 = 6518)"
## [177] "CHEMICAL, OTHER: (GLIOCLADIUM VIRENS = 129000)"
## [178] "CHEMICAL, OTHER: (TRICHODERMA VIRENS STRAIN G-41 = 176604)"
## [179] "CHEMICAL, FUNGICIDE: (DODINE = 44301)"
## [180] "CHEMICAL, FUNGICIDE: (FLUTOLANIL = 128975)"
## [181] "CHEMICAL, HERBICIDE: (2,4-D, TRIISO. SALT = 30035)"
```

```
## [182] "CHEMICAL, INSECTICIDE: (CYPERMETHRIN = 109702)"
## [183] "CHEMICAL, OTHER: (ALKYL. DIM. BENZ. AM = 69105)"
## [184] "CHEMICAL, OTHER: (DECYLDIMETHYLOCTYL = 69165)"
## [185] "CHEMICAL, OTHER: (DIDECYL DIM. AMMON. = 69166)"
## [186] "CHEMICAL, OTHER: (DIMETHYLDIOCTYL = 69149)"
## [187] "CHEMICAL, INSECTICIDE: (CYFLUMETOFEN = 138831)"
## [188] "CHEMICAL, INSECTICIDE: (EMAMECTIN BENZOATE = 122806)"
## [189] "CHEMICAL, INSECTICIDE: (SPIROTETRAMAT = 392201)"
## [190] "CHEMICAL, INSECTICIDE: (MUSTARD OIL = 4901)"
## [191] "CHEMICAL, OTHER: (DIMETHYL DISULFIDE (DMDS) = 29088)"
```

We filtered the data into two distinct categories: organic and non-organic. This was done based on the Domain and Domain Category columns, which indicate whether the data is associated with organic cultivation.

Splitting organic and non-organic data

```
organic_data <- strawberry_data %>%
  filter(grepl("ORGANIC", Domain) | grepl("ORGANIC", Domain.Category))

print(dim(organic_data))
```

```
## [1] 732  21
```

```
head(organic_data)
```

```
##   Program Year Period Week.Ending Geo.Level    State State.ANSI Ag.District
## 1  CENSUS 2021   YEAR          NA  NATIONAL US TOTAL         NA
## 2  CENSUS 2021   YEAR          NA  NATIONAL US TOTAL         NA
## 3  CENSUS 2021   YEAR          NA  NATIONAL US TOTAL         NA
## 4  CENSUS 2021   YEAR          NA  NATIONAL US TOTAL         NA
## 5  CENSUS 2021   YEAR          NA  NATIONAL US TOTAL         NA
## 6  CENSUS 2021   YEAR          NA  NATIONAL US TOTAL         NA
##   Ag.District.Code County County.ANSI Zip.Code Region watershed_code Watershed
## 1               NA                 NA       NA     NA              0        NA
## 2               NA                 NA       NA     NA              0        NA
## 3               NA                 NA       NA     NA              0        NA
## 4               NA                 NA       NA     NA              0        NA
## 5               NA                 NA       NA     NA              0        NA
## 6               NA                 NA       NA     NA              0        NA
##      Commodity                                            Data.Item
## 1 STRAWBERRIES                STRAWBERRIES, ORGANIC - ACRES HARVESTED
## 2 STRAWBERRIES STRAWBERRIES, ORGANIC - OPERATIONS WITH AREA HARVESTED
## 3 STRAWBERRIES          STRAWBERRIES, ORGANIC - OPERATIONS WITH SALES
## 4 STRAWBERRIES    STRAWBERRIES, ORGANIC - PRODUCTION, MEASURED IN CWT
## 5 STRAWBERRIES           STRAWBERRIES, ORGANIC - SALES, MEASURED IN $
## 6 STRAWBERRIES         STRAWBERRIES, ORGANIC - SALES, MEASURED IN CWT
##           Domain                      Domain.Category       Value CV....
## 1 ORGANIC STATUS ORGANIC STATUS: (NOP USDA CERTIFIED)       5,301   43.5
## 2 ORGANIC STATUS ORGANIC STATUS: (NOP USDA CERTIFIED)         546    6.3
## 3 ORGANIC STATUS ORGANIC STATUS: (NOP USDA CERTIFIED)         546    6.3
## 4 ORGANIC STATUS ORGANIC STATUS: (NOP USDA CERTIFIED)   1,495,299   51.2
```

```
## 5 ORGANIC STATUS ORGANIC STATUS: (NOP USDA CERTIFIED) 335,964,420   45.7
## 6 ORGANIC STATUS ORGANIC STATUS: (NOP USDA CERTIFIED)   1,494,673   51.3
```

```
non_organic_data <- strawberry_data %>%
  filter(!grepl("ORGANIC", Domain) & !grepl("ORGANIC", Domain.Category))

print(dim(non_organic_data))
```

```
## [1] 11937    21
```

```
head(non_organic_data)
```

```
##   Program Year Period Week.Ending Geo.Level   State State.ANSI Ag.District
## 1  CENSUS 2022   YEAR          NA    COUNTY ALABAMA          1  BLACK BELT
## 2  CENSUS 2022   YEAR          NA    COUNTY ALABAMA          1  BLACK BELT
## 3  CENSUS 2022   YEAR          NA    COUNTY ALABAMA          1  BLACK BELT
## 4  CENSUS 2022   YEAR          NA    COUNTY ALABAMA          1  BLACK BELT
## 5  CENSUS 2022   YEAR          NA    COUNTY ALABAMA          1  BLACK BELT
## 6  CENSUS 2022   YEAR          NA    COUNTY ALABAMA          1  BLACK BELT
##   Ag.District.Code  County County.ANSI Zip.Code Region watershed_code Watershed
## 1               40 BULLOCK          11       NA     NA              0        NA
## 2               40 BULLOCK          11       NA     NA              0        NA
## 3               40 BULLOCK          11       NA     NA              0        NA
## 4               40 BULLOCK          11       NA     NA              0        NA
## 5               40 BULLOCK          11       NA     NA              0        NA
## 6               40 BULLOCK          11       NA     NA              0        NA
##      Commodity                                    Data.Item Domain
## 1 STRAWBERRIES                   STRAWBERRIES - ACRES BEARING  TOTAL
## 2 STRAWBERRIES                     STRAWBERRIES - ACRES GROWN  TOTAL
## 3 STRAWBERRIES                 STRAWBERRIES - ACRES NON-BEARING  TOTAL
## 4 STRAWBERRIES     STRAWBERRIES - OPERATIONS WITH AREA BEARING  TOTAL
## 5 STRAWBERRIES       STRAWBERRIES - OPERATIONS WITH AREA GROWN  TOTAL
## 6 STRAWBERRIES STRAWBERRIES - OPERATIONS WITH AREA NON-BEARING  TOTAL
##   Domain.Category Value CV....
## 1   NOT SPECIFIED   (D)    (D)
## 2   NOT SPECIFIED     3   15.7
## 3   NOT SPECIFIED   (D)    (D)
## 4   NOT SPECIFIED     1    (L)
## 5   NOT SPECIFIED     6   52.7
## 6   NOT SPECIFIED     5   47.6
```

Organic Data: Contains records where "ORGANIC" appears in the Domain or Domain Category. Non-Organic Data: Contains all other records that do not have "ORGANIC" in these fields

Splitting the Chemical Data

```
# Filtering out the chemical-related data from the non-organic dataset
chemical_data <- non_organic_data %>%
  filter(grepl("CHEMICAL", Domain) | grepl("CHEMICAL", Domain.Category))
```

```r
# Splitting the "Domain Category" column into three new columns: Use, Chemical Name, and Code
chemical_data_clean <- chemical_data %>%
  separate(Domain.Category, into = c("Use", "Chemical_Name", "Chemical_Code"),
           sep = ": | = ", remove = FALSE, extra = "merge", fill = "right")

# Cleaning the "Chemical_Code" to retain only numeric values
chemical_data_clean <- chemical_data_clean %>%
  mutate(Chemical_Code = as.numeric(gsub("[^0-9]", "", Chemical_Code)))

# Viewing the cleaned chemical data
head(chemical_data_clean)
```

```
##   Program Year Period Week.Ending Geo.Level      State State.ANSI Ag.District
## 1  SURVEY 2023   YEAR          NA     STATE CALIFORNIA          6
## 2  SURVEY 2023   YEAR          NA     STATE CALIFORNIA          6
## 3  SURVEY 2023   YEAR          NA     STATE CALIFORNIA          6
## 4  SURVEY 2023   YEAR          NA     STATE CALIFORNIA          6
## 5  SURVEY 2023   YEAR          NA     STATE CALIFORNIA          6
## 6  SURVEY 2023   YEAR          NA     STATE CALIFORNIA          6
##   Ag.District.Code County County.ANSI Zip.Code Region watershed_code Watershed
## 1               NA                 NA       NA     NA              0        NA
## 2               NA                 NA       NA     NA              0        NA
## 3               NA                 NA       NA     NA              0        NA
## 4               NA                 NA       NA     NA              0        NA
## 5               NA                 NA       NA     NA              0        NA
## 6               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
## 2
## 3
## 4
## 5
## 6
```

```
# Saving the cleaned chemical data
write.csv(chemical_data_clean, '~/Desktop/cleaned_chemical_data.csv')
```

```
chemical_data <- non_organic_data %>%
  filter(grepl("CHEMICAL", Domain) | grepl("CHEMICAL", Domain.Category))

chemical_data_clean <- chemical_data %>%
  separate(Domain.Category, into = c("Use", "Chemical_Name", "Chemical_Code"),
           sep = ": | = ", remove = FALSE, extra = "merge", fill = "right") %>%
  mutate(Chemical_Code = as.numeric(gsub("[^0-9]", "", Chemical_Code)),
         Use = gsub("CHEMICAL, ", "", Use))

head(chemical_data_clean)
```

```
##   Program Year Period Week.Ending Geo.Level      State State.ANSI Ag.District
## 1  SURVEY 2023   YEAR          NA     STATE CALIFORNIA          6
## 2  SURVEY 2023   YEAR          NA     STATE CALIFORNIA          6
## 3  SURVEY 2023   YEAR          NA     STATE CALIFORNIA          6
## 4  SURVEY 2023   YEAR          NA     STATE CALIFORNIA          6
## 5  SURVEY 2023   YEAR          NA     STATE CALIFORNIA          6
## 6  SURVEY 2023   YEAR          NA     STATE CALIFORNIA          6
##   Ag.District.Code County County.ANSI Zip.Code Region watershed_code Watershed
## 1               NA                 NA       NA     NA              0        NA
## 2               NA                 NA       NA     NA              0        NA
## 3               NA                 NA       NA     NA              0        NA
## 4               NA                 NA       NA     NA              0        NA
## 5               NA                 NA       NA     NA              0        NA
## 6               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         Use
## 1               CHEMICAL, FUNGICIDE: (OXATHIAPIPROLIN = 128111)   FUNGICIDE
## 2               CHEMICAL, INSECTICIDE: (CYCLANILIPROLE = 26202) INSECTICIDE
## 3                  CHEMICAL, INSECTICIDE: (PERMETHRIN = 109701) INSECTICIDE
## 4 CHEMICAL, OTHER: (ISARIA FUMOSOROSEA STRAIN FE 9901 = 115003)       OTHER
## 5               CHEMICAL, FUNGICIDE: (OXATHIAPIPROLIN = 128111)   FUNGICIDE
## 6               CHEMICAL, INSECTICIDE: (CYCLANILIPROLE = 26202) INSECTICIDE
##                       Chemical_Name Chemical_Code Value CV....
## 1                   (OXATHIAPIPROLIN        128111   (D)
## 2                    (CYCLANILIPROLE         26202   (D)
## 3                       (PERMETHRIN        109701   (D)
## 4 (ISARIA FUMOSOROSEA STRAIN FE 9901        115003  (NA)
## 5                   (OXATHIAPIPROLIN        128111   (D)
## 6                    (CYCLANILIPROLE         26202   (D)
```

```
write.csv(chemical_data_clean, 'cleaned_chemical_data.csv')
```

We can now Specify the type of chemical, such as fungicide, insecticide, herbicide; The name of the chemical compound applied (e.g., BACILLUS SUBTILIS); A unique numeric code associated with the chemical.

Viewing the Final Column names-

```
colnames(strawberry_data)
```

```
##  [1] "Program"          "Year"             "Period"           "Week.Ending"
##  [5] "Geo.Level"        "State"            "State.ANSI"       "Ag.District"
##  [9] "Ag.District.Code" "County"           "County.ANSI"      "Zip.Code"
## [13] "Region"           "watershed_code"   "Watershed"        "Commodity"
## [17] "Data.Item"        "Domain"           "Domain.Category"  "Value"
## [21] "CV...."
```

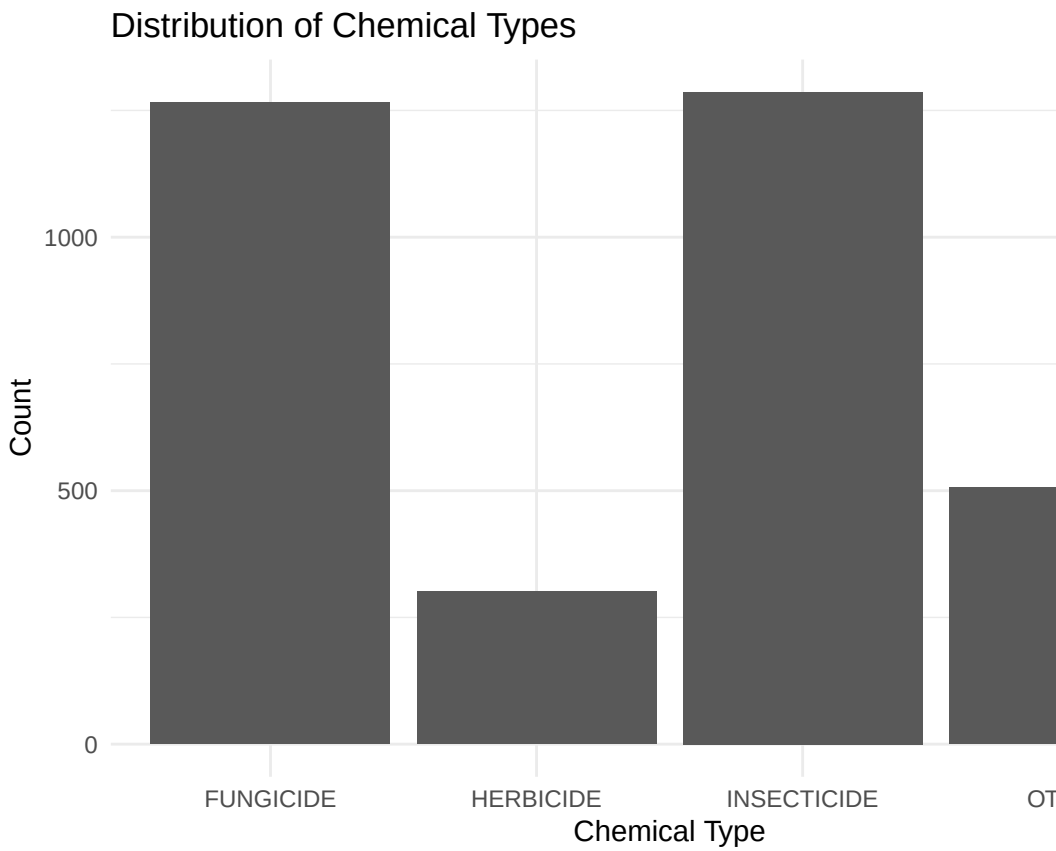
```
colnames(chemical_data_clean)
```

```
##  [1] "Program"          "Year"             "Period"           "Week.Ending"
##  [5] "Geo.Level"        "State"            "State.ANSI"       "Ag.District"
##  [9] "Ag.District.Code" "County"           "County.ANSI"      "Zip.Code"
## [13] "Region"           "watershed_code"   "Watershed"        "Commodity"
## [17] "Data.Item"        "Domain"           "Domain.Category"  "Use"
## [21] "Chemical_Name"    "Chemical_Code"    "Value"            "CV...."
```

Data Visualisations

```
chemical_summary <- chemical_data_clean %>%
  group_by(Use) %>%
  summarise(Count = n())

ggplot(chemical_summary, aes(x = Use, y = Count)) +
  geom_bar(stat = "identity") +
  labs(title = "Distribution of Chemical Types", x = "Chemical Type", y = "Count") +
  theme_minimal()
```

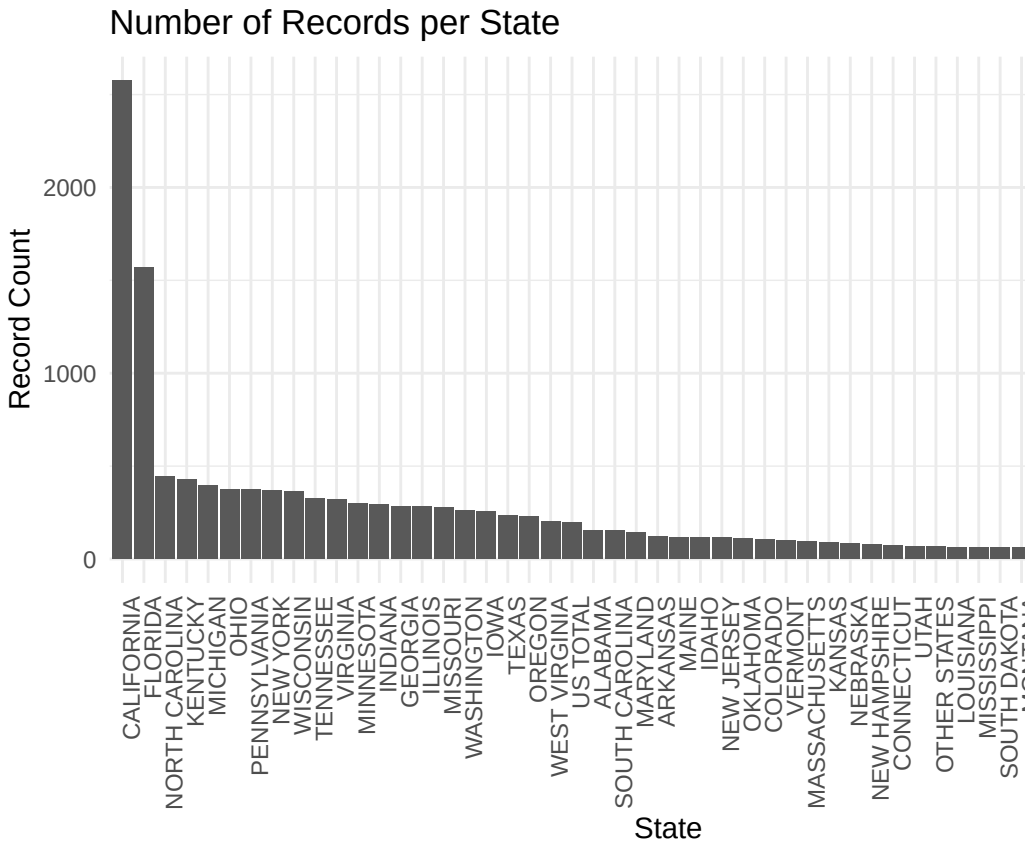


Plot of the Chemical Types-
Fungicides and Insecticides dominate the chemical applications in the dataset, with both categories having over 1,000 records each. Herbicides are applied far less frequently in comparison to fungicides and insecticides. The "Other" category shows a moderate level of use, which might include chemicals that don't fit neatly into fungicide, insecticide, or herbicide categories.

```
# Count records per state
state_summary <- strawberry_data %>%
  group_by(State) %>%
  summarise(Record_Count = n())

# Plot
```

```
ggplot(state_summary, aes(x = reorder(State, -Record_Count), y = Record_Count)) +
  geom_bar(stat = "identity") +
  labs(title = "Number of Records per State", x = "State", y = "Record Count") +
  theme_minimal() +
  theme(axis.text.x = element_text(angle = 90, hjust = 1))
```



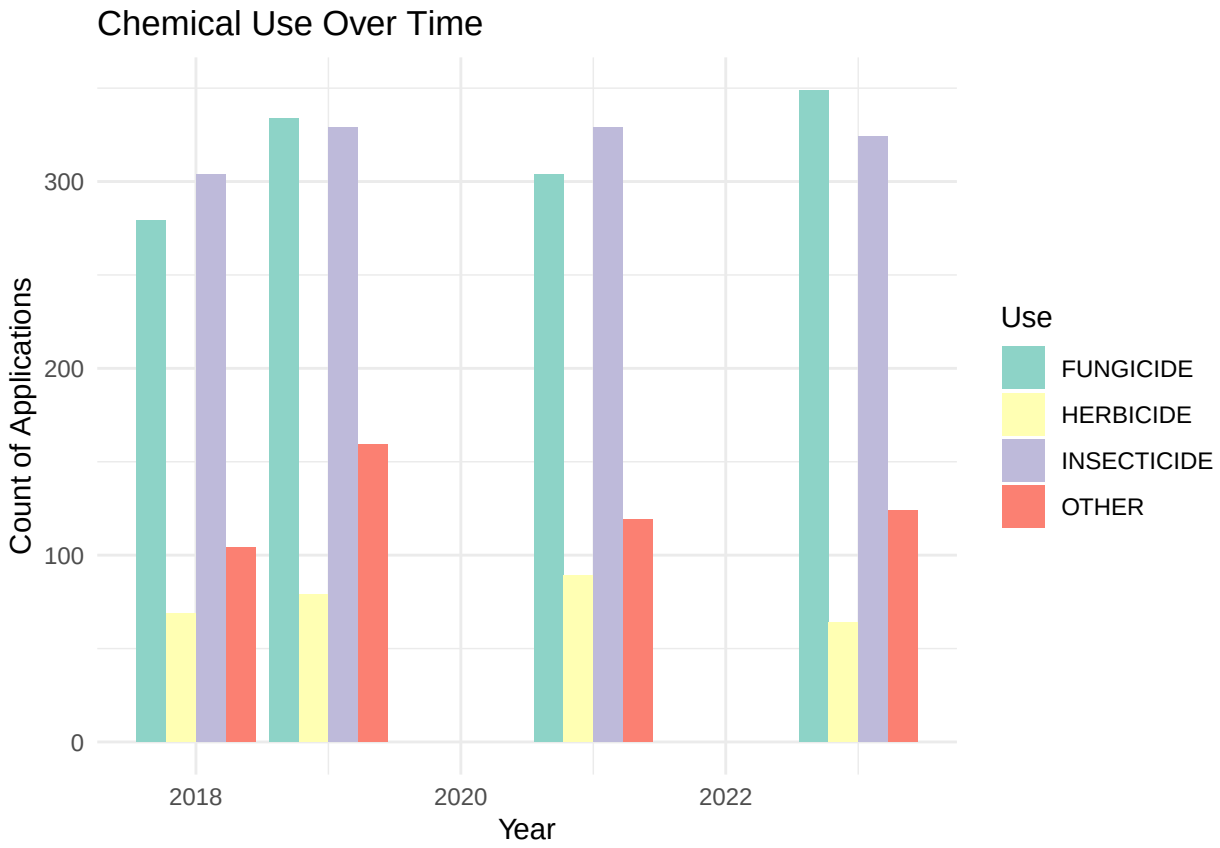
Number of Records Per State-
It seems likes California has the highest records for Strawberry Cultivation followe by Florida, North Carolina and others.

```
# Summarising chemical data by year and chemical type
chemical_by_year <- chemical_data_clean %>%
  group_by(Year, Use) %>%
  summarise(Count = n())
```

Chemical Use over time-

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

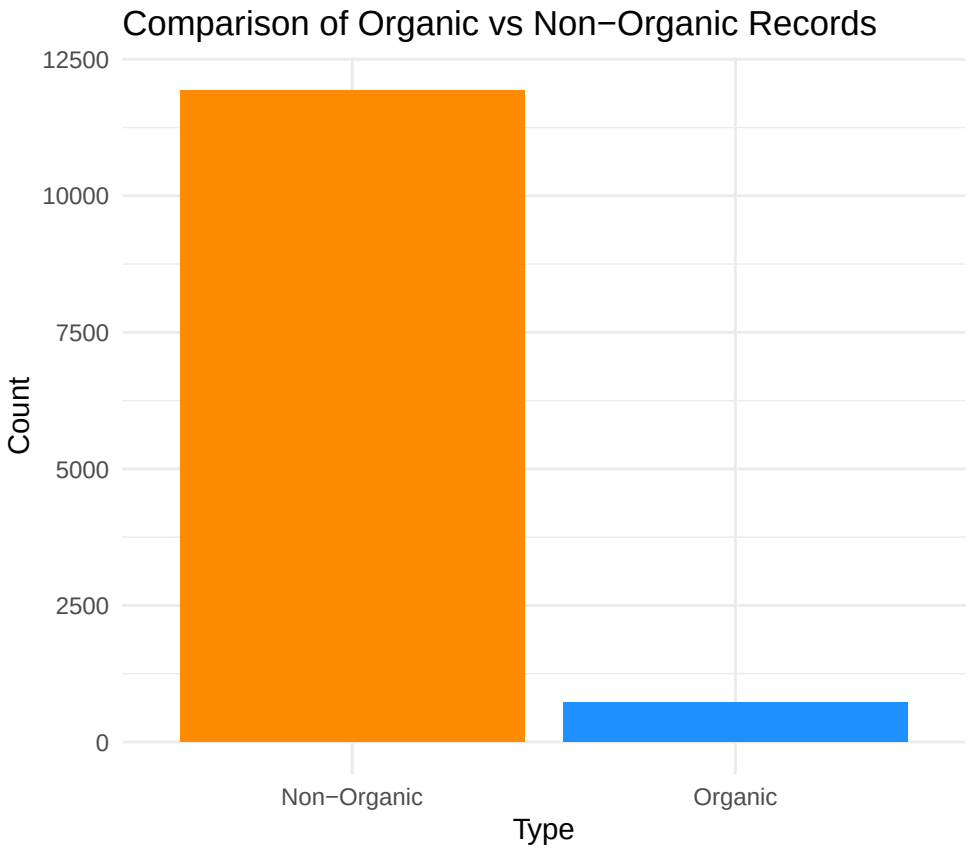
```
# Plot
ggplot(chemical_by_year, aes(x = Year, y = Count, fill = Use)) +
  geom_bar(stat = "identity", position = "dodge") +
  labs(title = "Chemical Use Over Time", x = "Year", y = "Count of Applications") +
  theme_minimal() +
  scale_fill_brewer(palette = "Set3")
```



The x-axis represents the years, while the y-axis indicates the count of chemical applications. The colors represent different chemical types: fungicide (green), herbicide (yellow), insecticide (purple), and other chemicals (red). Both fungicide and insecticide use has remained pretty stable between 2018 and 2022, with around 300 applications per year for each chemical type. Fungicide appears to be a little more used than insecticide. Herbicide use remains consistently low compared to fungicides and insecticides. Other chemicals see a notable increase in usage over time, particularly in 2020 and 2022. This category might contain newer or more specialized chemicals being adopted more frequently.

```
# Comparing organic and non-organic data
organic_vs_non_organic <- data.frame(
  Type = c("Organic", "Non-Organic"),
  Count = c(nrow(organic_data), nrow(non_organic_data))
)

# Plot
ggplot(organic_vs_non_organic, aes(x = Type, y = Count, fill = Type)) +
  geom_bar(stat = "identity") +
  labs(title = "Comparison of Organic vs Non-Organic Records", x = "Type", y = "Count") +
  theme_minimal() +
  scale_fill_manual(values = c("darkorange", "dodgerblue"))
```



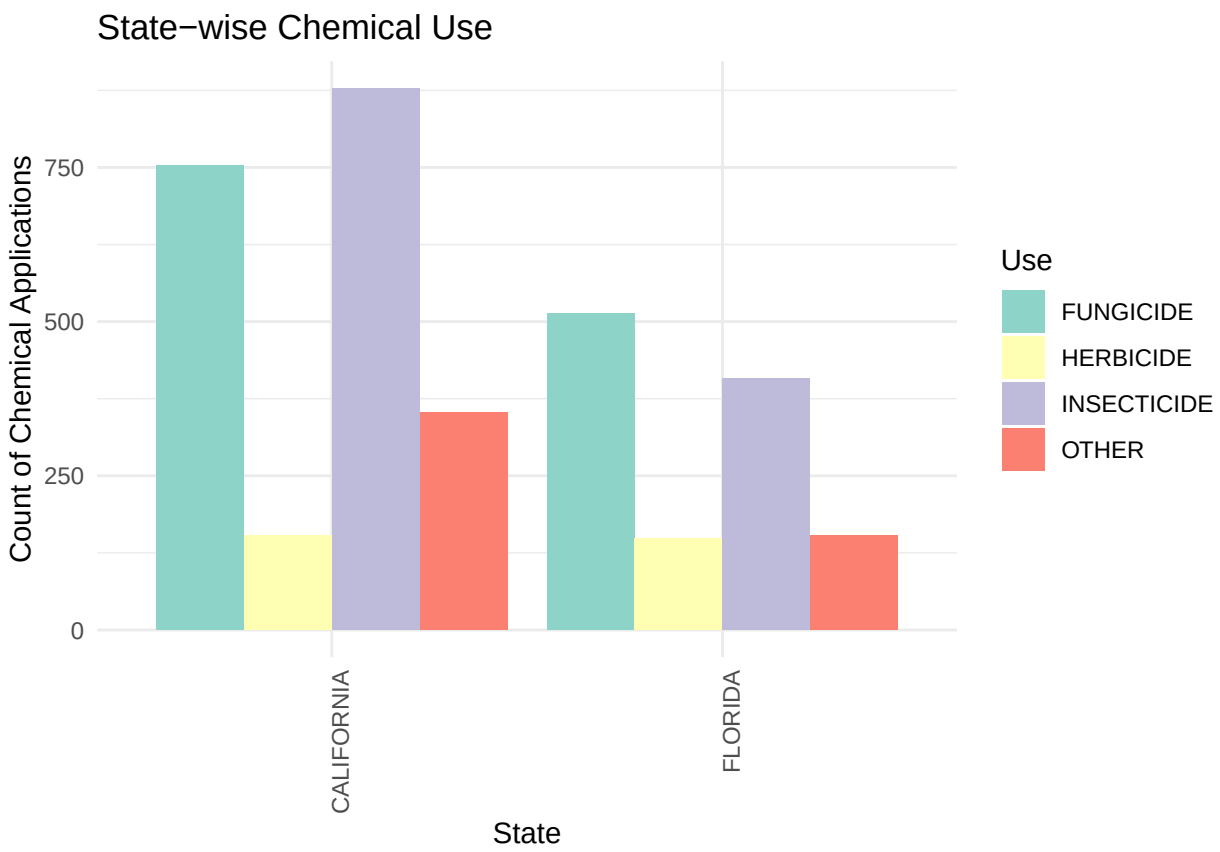
Organic vs Non-Organic Records-
The non-organic category dominates the dataset, with approximately 12,000 records.

```
# Grouping by state and chemical type
chemical_by_state <- chemical_data_clean %>%
  group_by(State, Use) %>%
  summarise(Count = n())
```

State-wise Chemical Use-

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

```
# Plot
ggplot(chemical_by_state, aes(x = reorder(State, -Count), y = Count, fill = Use)) +
  geom_bar(stat = "identity", position = "dodge") +
  labs(title = "State-wise Chemical Use", x = "State", y = "Count of Chemical Applications") +
  theme_minimal() +
  theme(axis.text.x = element_text(angle = 90, hjust = 1)) +
  scale_fill_brewer(palette = "Set3")
```



California has the highest number of chemical applications across all chemical types, particularly for insecticides and fungicides. This aligns with California's the fact that it was at the top in strawberry production, as observed in earlier plots. Fungicide is the dominant chemical type used in California, closely followed by insecticides. Florida shows a similar pattern, but the total count of chemical applications is notably lower than in California.

Conclusion-

The USDA strawberry dataset analysis provided several insights into chemical use, regional farming practices, and data distribution between organic and non-organic farming methods. The data has also been Split in a way where it can be used in the future.