

# Stawberries: exploratory data analysis

Chang Shi

2023 Oct 23

## Assignment

Using our class discussions and this document as a starting point, produce an EDA report. The report should describe the data itself so that readers understand the data sources used in the report and how you cleaned and organized the data for analysis.

The sections below suggest how the report might be organized. The report should be succinct, communicating the information that you believe will be helpful to someone doing a fuller analysis of the data or using the data for model building. Implementation details should be included in commentary that is included in code.

Sections of the document as it was originally presented in class have been commented so that you can see them in the code.

## Data acquisition and assessment

- Data sources
- Assumptions and motivations

## Data cleaning and organization

Outline the approach taken to clean and organize the data.

## References

### Material about strawberries

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

### Technical 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 issues 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 insights.

## Initial questions

- Initial questions about strawberries, the data, and about the work you are undertaking. Write these before you begin working.

## The data

Describe the source and original condition of the data: organization, problems with the data that needed to be addressed and so on. Cite data sources.

The data set for this assignment has been selected from: USDA\_NASS The data have been stored on NASS here: USDA\_NASS\_strawb\_2023SEP19

Make relevant observations in the document and in your code about data. Add commentary to the code so that another analysts could use or extend your code.

Discuss missing data, including how you handled it. Be careful to point out where NA's are being produced during processing and are not data missing in the original data.

Where it is relevant, include information of how you have organized the data for analysis. It might, for example, be helpful to know that there is both agricultural census data and survey data. It might be helpful to discuss data that appears to be redundant between these two sources.

Make sure you include details in your discussion and in your code about other data and information you used in your work. Cite sources and provide detail that would allow another analyst to reproduce your work.

```
## Rows: 4,314
## Columns: 10
## $ Program      <chr> "CENSUS", "CENSUS", "CENSUS", "CENSUS", "CENSUS", "C-
## $ Year          <dbl> 2021, 2021, 2021, 2021, 2021, 2021, 2021, 2021, 2021, 2021~
## $ Period        <chr> "YEAR", "YEAR", "YEAR", "YEAR", "YEAR", "YEAR", "YEA~
## $ State         <chr> "ALASKA", "ALASKA", "ALASKA", "ALASKA", "ALASKA", "A~
## $ 'State ANSI'  <chr> "02", "02", "02", "02", "02", "02", "02", "06", "06"~
## $ 'Data Item'   <chr> "STRAWBERRIES, ORGANIC - OPERATIONS WITH SALES", "ST~
## $ Domain        <chr> "ORGANIC STATUS", "ORGANIC STATUS", "ORGANIC STATUS"~
## $ 'Domain Category' <chr> "ORGANIC STATUS: (NOP USDA CERTIFIED)", "ORGANIC STA~
## $ Value         <chr> "2", "(D)", "(D)", "(D)", "2", "(D)", "(D)", "142", ~
## $ 'CV (%)'      <chr> "(H)", "(D)", "(D)", "(D)", "(H)", "(D)", "(D)", "19~

## [1] "Every row has value in the State column."
```

```
plot1_data <- strawberry %>%
  select(Year, State, `Data Item`, Value) %>%
  filter((Year == 2021) & (`Data Item` == "STRAWBERRIES, ORGANIC - OPERATIONS WITH SALES"))

# Convert Value column to numeric after handling non-numeric characters
plot1_data$Value <- as.numeric(gsub("[^0-9]", "", plot1_data$Value))

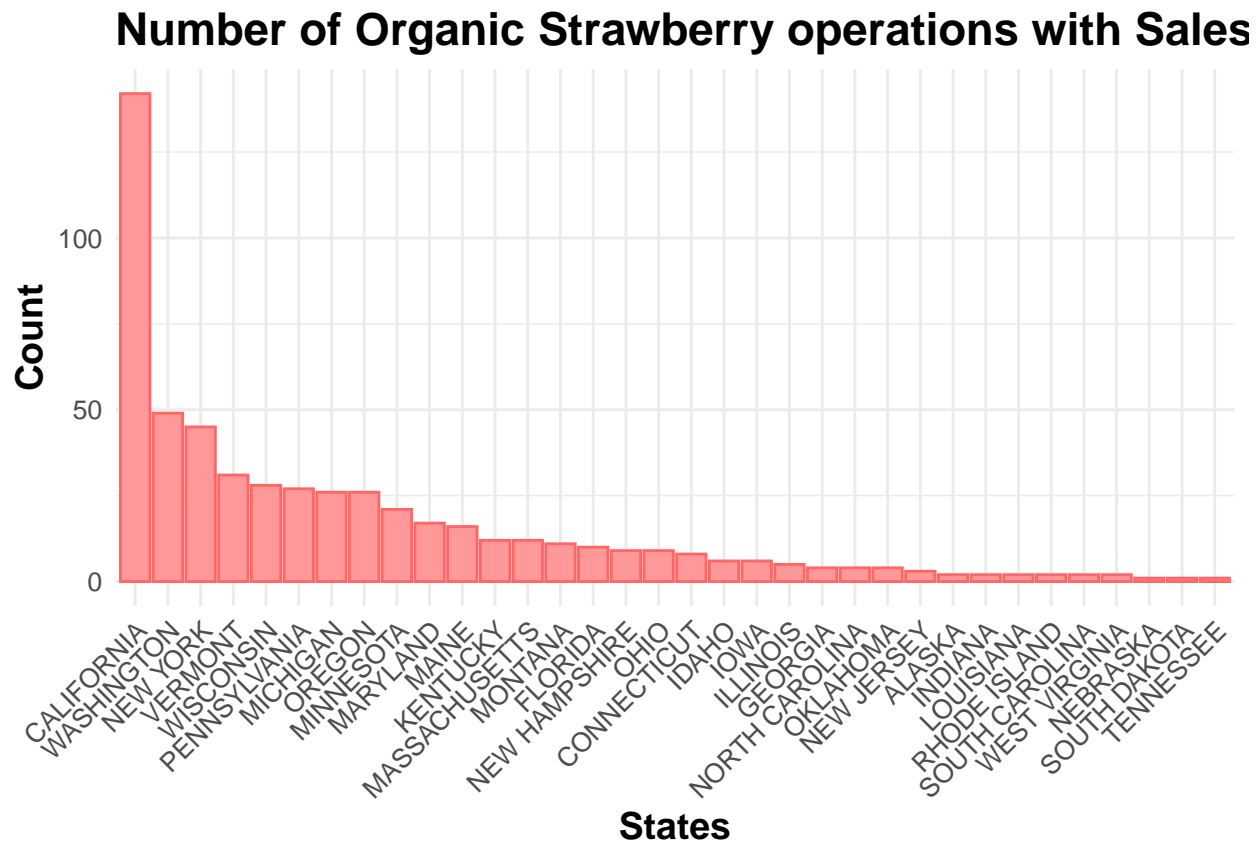
# Arrange data in descending order of Value
plot1_data <- plot1_data %>%
  arrange(desc(Value))

# Create and optimize the plot
ggplot(plot1_data, aes(x=reorder(State, -Value), y=Value)) +
  geom_bar(stat = "identity", fill = "#FF9999", color = "#FF6666") +
```

```

theme_minimal() +
theme(axis.text.x=element_text(angle=45, hjust=1, size=10),
      axis.text.y=element_text(size=10),
      title=element_text(size=14, face="bold")) +
labs(x = "States", y = "Count",
     title = "Number of Organic Strawberry operations with Sales in 2021")

```



```

# Adjusting the code based on the correct column names
plot2_data <- strawberry %>%
  select(Year, State, `Data Item`, Domain, Value) %>%
  filter((Year == 2021) &
         (Domain == "ORGANIC STATUS") &
         (`Data Item` == "STRAWBERRIES, ORGANIC - SALES, MEASURED IN $") &
         (Value != "(D)"))

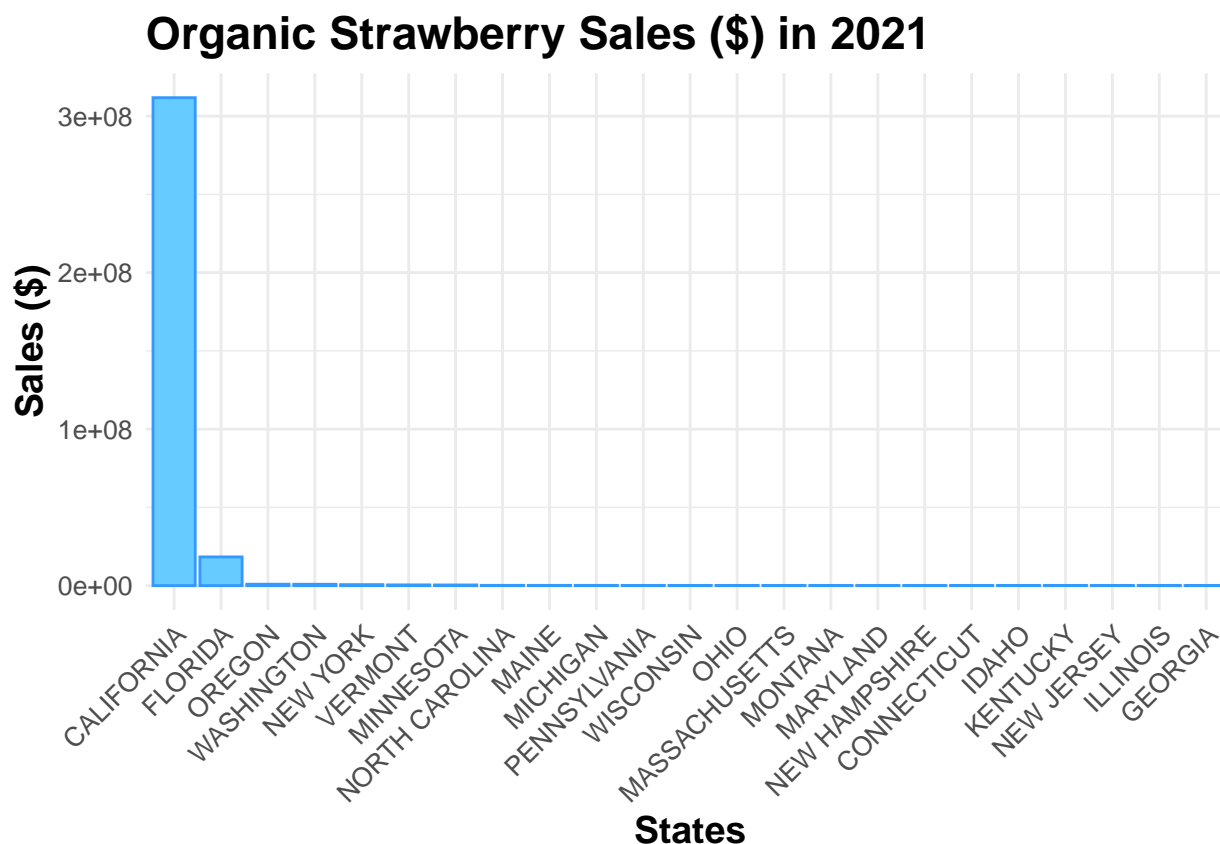
# Convert Value column to numeric after handling non-numeric characters
plot2_data$Value <- as.numeric(gsub("[^0-9]", "", plot2_data$Value))

# Arrange data in descending order of Value
plot2_data <- plot2_data %>%
  arrange(desc(Value))

# Create and optimize the plot
ggplot(plot2_data, aes(x=reorder(State, -Value), y=Value)) +
  geom_bar(stat = "identity", fill = "#66CCFF", color = "#3399FF") +

```

```
theme_minimal() +
theme(axis.text.x=element_text(angle=45, hjust=1, size=10),
      axis.text.y=element_text(size=10),
      title=element_text(size=14, face="bold")) +
labs(x = "States", y = "Sales ($)",
     title = "Organic Strawberry Sales ($) in 2021")
```



## EDA

Once the data has been cleaned and organized, you must conduct your own EDA. Be sure to include a discussion of your analysis of the chemical information, including citations for data and other information you have used. Visualizations should play a key role in your analysis. Plots should be labeled and captioned.

```
# glimpse(strwb_survey)

## find strwb_survey columns that should be split into columns
##

## this section will produce tables listing
## the variables in the columns of strwb_survey
## remove the table you won't use

# distinct(strwb_survey[,3]) |> kable()
# c4 <- distinct(strwb_survey[,4])
```

```
#
# c6 <- distinct(strwb_survey[,6])
# c7 <- distinct(strwb_survey[,7])
# c8 <- distinct(strwb_survey[,8])

# c3 |> kable()
# c4 |> kable()
#
# c6 |> kable()
# c7 |> kable()
# c8 |> kable()

# View the structure of strwb_survey dataframe
glimpse(strwb_survey)

## Rows: 3,450
## Columns: 10
## $ Program      <chr> "SURVEY", "SURVEY", "SURVEY", "SURVEY", "SURVEY", "S~
## $ Year          <dbl> 2022, 2022, 2022, 2022, 2022, 2022, 2022, 2022, 2022~
## $ Period        <chr> "MARKETING YEAR", "MARKETING YEAR", "MARKETING YEAR"~
## $ State         <chr> "CALIFORNIA", "CALIFORNIA", "CALIFORNIA", "FLORIDA",~
## $ 'State ANSI'  <chr> "06", "06", "06", "12", "12", "12", NA, NA, NA, "06"~
## $ 'Data Item'   <chr> "STRAWBERRIES - PRICE RECEIVED, MEASURED IN $ / CWT"~
## $ Domain        <chr> "TOTAL", "TOTAL", "TOTAL", "TOTAL", "TOTAL", "TOTAL"~
## $ 'Domain Category' <chr> "NOT SPECIFIED", "NOT SPECIFIED", "NOT SPECIFIED", "~
## $ Value         <chr> "108", "(D)", "(D)", "169", "(D)", "(D)", "0", "135"~
## $ 'CV (%)'      <chr> NA, NA, NA, NA, NA, NA, NA, NA, NA, NA, NA, ~
```

```
# Extract distinct values from various columns of strwb_survey
c3 <- distinct(strwb_survey[,3])
c4 <- distinct(strwb_survey[,4])
c6 <- distinct(strwb_survey[,6])
c7 <- distinct(strwb_survey[,7])
c8 <- distinct(strwb_survey[,8])
c5 <- distinct(strwb_survey[,5])

# Display the distinct values using kable for a better presentation
kable(c3, caption = "Distinct values from column 3")
```

Table 1: Distinct values from column 3

Period
MARKETING YEAR
YEAR
YEAR - AUG FORECAST

```
kable(c4, caption = "Distinct values from column 4")
```

Table 2: Distinct values from column 4

State
CALIFORNIA
FLORIDA
OTHER STATES
NEW YORK
NORTH CAROLINA
OREGON
WASHINGTON
MICHIGAN
OHIO
PENNSYLVANIA
WISCONSIN

```
kable(c5, caption = "Distinct values from column 5")
```

Table 3: Distinct values from column 5

State ANSI
06
12
NA
36
37
41
53
26
39
42
55

```
kable(c6, caption = "Distinct values from column 6")
```

Table 4: Distinct values from column 6

Data Item
STRAWBERRIES - PRICE RECEIVED, MEASURED IN \$ / CWT
STRAWBERRIES, FRESH MARKET - PRICE RECEIVED, MEASURED IN \$ / CWT
STRAWBERRIES, PROCESSING - PRICE RECEIVED, MEASURED IN \$ / CWT
STRAWBERRIES - PRODUCTION, MEASURED IN \$
STRAWBERRIES - PRODUCTION, MEASURED IN CWT
STRAWBERRIES, FRESH MARKET - PRODUCTION, MEASURED IN \$
STRAWBERRIES, FRESH MARKET, UTILIZED - PRODUCTION, MEASURED IN CWT
STRAWBERRIES, NOT SOLD - PRODUCTION, MEASURED IN CWT
STRAWBERRIES, PROCESSING - PRODUCTION, MEASURED IN \$
STRAWBERRIES, PROCESSING, UTILIZED - PRODUCTION, MEASURED IN CWT
STRAWBERRIES, UTILIZED - PRODUCTION, MEASURED IN CWT
STRAWBERRIES - PRODUCTION, MEASURED IN TONS

Data Item
STRAWBERRIES, BEARING - APPLICATIONS, MEASURED IN LB
STRAWBERRIES, BEARING - APPLICATIONS, MEASURED IN LB / ACRE / APPLICATION, AVG
STRAWBERRIES, BEARING - APPLICATIONS, MEASURED IN LB / ACRE / YEAR, AVG
STRAWBERRIES, BEARING - APPLICATIONS, MEASURED IN NUMBER, AVG
STRAWBERRIES, UTILIZED - PRODUCTION, MEASURED IN TONS
STRAWBERRIES, PROCESSING - PRICE RECEIVED, MEASURED IN \$ / TON
STRAWBERRIES, PROCESSING, UTILIZED - PRODUCTION, MEASURED IN TONS
STRAWBERRIES, PROCESSING - PRODUCTION, MEASURED IN CWT

```
kable(c7, caption = "Distinct values from column 7")
```

Table 5: Distinct values from column 7

Domain
TOTAL
CHEMICAL, FUNGICIDE
CHEMICAL, HERBICIDE
CHEMICAL, INSECTICIDE
CHEMICAL, OTHER
FERTILIZER

```
kable(c8, caption = "Distinct values from column 8")
```

Table 6: Distinct values from column 8

Domain Category
NOT SPECIFIED
CHEMICAL, FUNGICIDE: (AZOXYSTROBIN = 128810)
CHEMICAL, FUNGICIDE: (BACILLUS AMYLOLIQUEFAC F727 = 16489)
CHEMICAL, FUNGICIDE: (BACILLUS AMYLOLIQUEFACIENS MBI 600 = 129082)
CHEMICAL, FUNGICIDE: (BACILLUS AMYLOLIQUEFACIENS STRAIN D747 = 16482)
CHEMICAL, FUNGICIDE: (BACILLUS PUMILUS = 6485)
CHEMICAL, FUNGICIDE: (BACILLUS SUBTILIS = 6479)
CHEMICAL, FUNGICIDE: (BLAD = 30006)
CHEMICAL, FUNGICIDE: (BORAX DECAHYDRATE = 11102)
CHEMICAL, FUNGICIDE: (BOSCALID = 128008)
CHEMICAL, FUNGICIDE: (BT SUBSP KURSTAKI EVB-113-19 = 6544)
CHEMICAL, FUNGICIDE: (CAPTAN = 81301)
CHEMICAL, FUNGICIDE: (COPPER OCTANOATE = 23306)
CHEMICAL, FUNGICIDE: (CYFLUFENAMID = 555550)
CHEMICAL, FUNGICIDE: (CYPRODINIL = 288202)
CHEMICAL, FUNGICIDE: (DIFENOCONAZOLE = 128847)
CHEMICAL, FUNGICIDE: (FENHEXAMID = 90209)
CHEMICAL, FUNGICIDE: (FLUDIOXONIL = 71503)
CHEMICAL, FUNGICIDE: (FLUOPYRAM = 80302)
CHEMICAL, FUNGICIDE: (FLUXAPYROXAD = 138009)
CHEMICAL, FUNGICIDE: (FOSETYL-AL = 123301)
CHEMICAL, FUNGICIDE: (ISOSETAMID = 270000)

---

Domain Category

---

CHEMICAL, FUNGICIDE: (MEFENOXAM = 113502)  
CHEMICAL, FUNGICIDE: (MONO-POTASSIUM SALT = 76416)  
CHEMICAL, FUNGICIDE: (MYCLOBUTANIL = 128857)  
CHEMICAL, FUNGICIDE: (PENTHIOPYRAD = 90112)  
CHEMICAL, FUNGICIDE: (POLYOXIN D ZINC SALT = 230000)  
CHEMICAL, FUNGICIDE: (POTASSIUM BICARBON. = 73508)  
CHEMICAL, FUNGICIDE: (PROPICONAZOLE = 122101)  
CHEMICAL, FUNGICIDE: (PYRACLOSTROBIN = 99100)  
CHEMICAL, FUNGICIDE: (PYRIMETHANIL = 288201)  
CHEMICAL, FUNGICIDE: (QUINOLINE = 55459)  
CHEMICAL, FUNGICIDE: (STREPTOMYCES LYDICUS = 6327)  
CHEMICAL, FUNGICIDE: (SULFUR = 77501)  
CHEMICAL, FUNGICIDE: (TETRACONAZOLE = 120603)  
CHEMICAL, FUNGICIDE: (THIOPHANATE-METHYL = 102001)  
CHEMICAL, FUNGICIDE: (THIRAM = 79801)  
CHEMICAL, FUNGICIDE: (TOTAL)  
CHEMICAL, FUNGICIDE: (TRIFLOXYSTROBIN = 129112)  
CHEMICAL, FUNGICIDE: (TRIFLUMIZOLE = 128879)  
CHEMICAL, HERBICIDE: (CARFENTRAZONE-ETHYL = 128712)  
CHEMICAL, HERBICIDE: (FLUMIOXAZIN = 129034)  
CHEMICAL, HERBICIDE: (GLYPHOSATE ISO. SALT = 103601)  
CHEMICAL, HERBICIDE: (GLYPHOSATE POT. SALT = 103613)  
CHEMICAL, HERBICIDE: (NAPROPAMIDE = 103001)  
CHEMICAL, HERBICIDE: (OXYFLUORFEN = 111601)  
CHEMICAL, HERBICIDE: (PARAQUAT = 61601)  
CHEMICAL, HERBICIDE: (PENDIMETHALIN = 108501)  
CHEMICAL, HERBICIDE: (TOTAL)  
CHEMICAL, INSECTICIDE: (ABAMECTIN = 122804)  
CHEMICAL, INSECTICIDE: (ACEQUINOCYL = 6329)  
CHEMICAL, INSECTICIDE: (ACETAMIPRID = 99050)  
CHEMICAL, INSECTICIDE: (AZADIRACHTIN = 121701)  
CHEMICAL, INSECTICIDE: (BEAUVERIA BASSIANA = 128924)  
CHEMICAL, INSECTICIDE: (BIFENAZATE = 586)  
CHEMICAL, INSECTICIDE: (BIFENTHRIN = 128825)  
CHEMICAL, INSECTICIDE: (BT KURSTAK ABTS-1857 = 6523)  
CHEMICAL, INSECTICIDE: (BT KURSTAKI ABTS-351 = 6522)  
CHEMICAL, INSECTICIDE: (BT KURSTAKI EG7841 = 6453)  
CHEMICAL, INSECTICIDE: (BT KURSTAKI SA-11 = 6519)  
CHEMICAL, INSECTICIDE: (BT SUB AIZAWAI GC-91 = 6426)  
CHEMICAL, INSECTICIDE: (BUPROFEZIN = 275100)  
CHEMICAL, INSECTICIDE: (BURKHOLDERIA A396 CELLS & MEDIA = 6534)  
CHEMICAL, INSECTICIDE: (CANOLA OIL = 11332)  
CHEMICAL, INSECTICIDE: (CHLORANTRANILIPROLE = 90100)  
CHEMICAL, INSECTICIDE: (CHROMOBAC SUBTSUGAE PRAA4-1 CELLS AND SPENT MEDIA = 16329)  
CHEMICAL, INSECTICIDE: (CYANTRANILIPROLE = 90098)  
CHEMICAL, INSECTICIDE: (CYFLUMETOFEN = 138831)  
CHEMICAL, INSECTICIDE: (DIAZINON = 57801)  
CHEMICAL, INSECTICIDE: (ETOXAZOLE = 107091)  
CHEMICAL, INSECTICIDE: (FENBUTATIN-OXIDE = 104601)  
CHEMICAL, INSECTICIDE: (FENPROPATHRIN = 127901)  
CHEMICAL, INSECTICIDE: (FENPYROXIMATE = 129131)



---

Domain Category

---

CHEMICAL, INSECTICIDE: (FLONICAMID = 128016)  
CHEMICAL, INSECTICIDE: (FLUPYRADIFURONE = 122304)  
CHEMICAL, INSECTICIDE: (HELICOVERPA ZEA NPV = 107300)  
CHEMICAL, INSECTICIDE: (HEXYTHIAZOX = 128849)  
CHEMICAL, INSECTICIDE: (IMIDACLOPRID = 129099)  
CHEMICAL, INSECTICIDE: (MALATHION = 57701)  
CHEMICAL, INSECTICIDE: (METHOXYFENOZIDE = 121027)  
CHEMICAL, INSECTICIDE: (NALED = 34401)  
CHEMICAL, INSECTICIDE: (NEEM OIL = 25006)  
CHEMICAL, INSECTICIDE: (NEEM OIL, CLAR. HYD. = 25007)  
CHEMICAL, INSECTICIDE: (NOVALURON = 124002)  
CHEMICAL, INSECTICIDE: (PETROLEUM DISTILLATE = 63503)  
CHEMICAL, INSECTICIDE: (PIPERONYL BUTOXIDE = 67501)  
CHEMICAL, INSECTICIDE: (POTASSIUM SALTS = 79021)  
CHEMICAL, INSECTICIDE: (PYRETHRINS = 69001)  
CHEMICAL, INSECTICIDE: (PYRIDABEN = 129105)  
CHEMICAL, INSECTICIDE: (PYRIPROXYFEN = 129032)  
CHEMICAL, INSECTICIDE: (SPINETORAM = 110007)  
CHEMICAL, INSECTICIDE: (SPINOSAD = 110003)  
CHEMICAL, INSECTICIDE: (SPIROMESIFEN = 24875)  
CHEMICAL, INSECTICIDE: (THIAMETHOXAM = 60109)  
CHEMICAL, INSECTICIDE: (TOTAL)  
CHEMICAL, OTHER: (ACIBENZOLAR-S-METHYL = 61402)  
CHEMICAL, OTHER: (CAPRIC ACID = 128955)  
CHEMICAL, OTHER: (CAPRYLIC ACID = 128919)  
CHEMICAL, OTHER: (CAPSICUM OLEORESIN EXTRACT = 70704)  
CHEMICAL, OTHER: (CHLOROPICRIN = 81501)  
CHEMICAL, OTHER: (DICHLOROPROPENE = 29001)  
CHEMICAL, OTHER: (FLUTRIAFOL = 128940)  
CHEMICAL, OTHER: (GARLIC OIL = 128827)  
CHEMICAL, OTHER: (HYDROGEN PEROXIDE = 595)  
CHEMICAL, OTHER: (IRON PHOSPHATE = 34903)  
CHEMICAL, OTHER: (METAM-POTASSIUM = 39002)  
CHEMICAL, OTHER: (MINERAL OIL = 63502)  
CHEMICAL, OTHER: (PAECILOMYCES FUMOSOR = 115002)  
CHEMICAL, OTHER: (PEROXYACETIC ACID = 63201)  
CHEMICAL, OTHER: (POTASSIUM SILICATE = 72606)  
CHEMICAL, OTHER: (PSEUDOMONAS CHLORORAPHIS STRAIN AFS009 = 6800)  
CHEMICAL, OTHER: (REYNOUtria SACHALINE = 55809)  
CHEMICAL, OTHER: (TOTAL)  
CHEMICAL, FUNGICIDE: (PYDIFLUMETOFEN = 90110)  
CHEMICAL, HERBICIDE: (CLETHODIM = 121011)  
CHEMICAL, HERBICIDE: (COPPER ETHANOLAMINE = 24409)  
CHEMICAL, HERBICIDE: (DIMETHENAMID = 129051)  
CHEMICAL, HERBICIDE: (FLUROXYPYR 1-MHE = 128968)  
CHEMICAL, HERBICIDE: (HALOSULFURON-METHYL = 128721)  
CHEMICAL, HERBICIDE: (KANTOR = 129108)  
CHEMICAL, INSECTICIDE: (CARBARYL = 56801)  
CHEMICAL, INSECTICIDE: (FENAZAQUIN = 44501)  
CHEMICAL, INSECTICIDE: (SULFOXAFLOr = 5210)  
CHEMICAL, OTHER: (CYTOKININS = 116801)  
CHEMICAL, OTHER: (ETHEPHON = 99801)

---

Domain Category

---

CHEMICAL, OTHER: (INDOLEBUTYRIC ACID = 46701)  
CHEMICAL, FUNGICIDE: (BACILLUS SUBT. GB03 = 129068)  
CHEMICAL, FUNGICIDE: (COPPER HYDROXIDE = 23401)  
CHEMICAL, FUNGICIDE: (TRICHODERMA HARZ. = 119202)  
CHEMICAL, HERBICIDE: (GLUFOSINATE-AMMONIUM = 128850)  
CHEMICAL, HERBICIDE: (SULFENTRAZONE = 129081)  
CHEMICAL, INSECTICIDE: (CHLORPYRIFOS = 59101)  
CHEMICAL, INSECTICIDE: (SOYBEAN OIL = 31605)  
CHEMICAL, INSECTICIDE: (ZETA-CYPERMETHRIN = 129064)  
CHEMICAL, OTHER: (AUREOBASIDIUM PULLULANS DSM 14940 = 46010)  
CHEMICAL, OTHER: (AUREOBASIDIUM PULLULANS DSM 14941 = 36010)  
CHEMICAL, OTHER: (BT KURSTAKI SA-12 = 6518)  
CHEMICAL, OTHER: (GLIOCLADIUM VIRENS = 129000)  
CHEMICAL, OTHER: (METALDEHYDE = 53001)  
CHEMICAL, OTHER: (METAM-SODIUM = 39003)  
CHEMICAL, OTHER: (TRICHODERMA VIRENS STRAIN G-41 = 176604)  
FERTILIZER: (NITROGEN)  
FERTILIZER: (PHOSPHATE)  
FERTILIZER: (POTASH)  
FERTILIZER: (SULFUR)  
CHEMICAL, FUNGICIDE: (COPPER CHLORIDE HYD. = 23501)  
CHEMICAL, FUNGICIDE: (DODINE = 44301)  
CHEMICAL, FUNGICIDE: (FLUTOLANIL = 128975)  
CHEMICAL, HERBICIDE: (2,4-D, DIMETH. SALT = 30019)  
CHEMICAL, HERBICIDE: (2,4-D, TRIISO. SALT = 30035)  
CHEMICAL, INSECTICIDE: (CYPERMETHRIN = 109702)  
CHEMICAL, OTHER: (ALKYL. DIM. BENZ. AM = 69105)  
CHEMICAL, OTHER: (DECYLDIMETHYLOCTYL = 69165)  
CHEMICAL, OTHER: (DIDECYL DIM. AMMON. = 69166)  
CHEMICAL, OTHER: (DIMETHYLDIOCTYL = 69149)  
CHEMICAL, FUNGICIDE: (IPRODIONE = 109801)  
CHEMICAL, INSECTICIDE: (CYFLUMETOFEN = 138831)  
CHEMICAL, INSECTICIDE: (EMAMECTIN BENZOATE = 122806)  
CHEMICAL, INSECTICIDE: (LAMBDA-CYHALOTHRIN = 128897)  
CHEMICAL, INSECTICIDE: (SPIROTETRAMAT = 392201)  
CHEMICAL, INSECTICIDE: (MUSTARD OIL = 4901)  
CHEMICAL, OTHER: (DIMETHYL DISULFIDE (DMDS) = 29088)  
CHEMICAL, FUNGICIDE: (COPPER OXIDE = 25601)  
CHEMICAL, HERBICIDE: (AMMONIUM PELARGONATE = 31802)  
CHEMICAL, INSECTICIDE: (FLUBENDIAMIDE = 27602)  
CHEMICAL, OTHER: (METHYL BROMIDE = 53201)  
CHEMICAL, FUNGICIDE: (CHLOROTHALONIL = 81901)  
CHEMICAL, FUNGICIDE: (CYAZOFAMID = 85651)  
CHEMICAL, FUNGICIDE: (MANCOZEB = 14504)  
CHEMICAL, INSECTICIDE: (ENDOSULFAN = 79401)  
CHEMICAL, OTHER: (CAPSAICIN = 70701)  
CHEMICAL, OTHER: (HARPIN A B PROTEIN = 6506)  
CHEMICAL, HERBICIDE: (CLOPYRALID MONO SALT = 117401)  
CHEMICAL, HERBICIDE: (SIMAZINE = 80807)  
CHEMICAL, HERBICIDE: (TERBACIL = 12701)  
CHEMICAL, INSECTICIDE: (FERRIC SODIUM EDTA = 139114)  
CHEMICAL, HERBICIDE: (CLOMAZONE = 125401)

---

```
## [1] "MARKETING YEAR"      "YEAR"      "YEAR - AUG FORECAST"

##      MARKETING YEAR      YEAR  YEAR - AUG FORECAST
##      "marketing year"    "year" "year - Aug Forecast"

## # A tibble: 6 x 1
##   Domain
##   <chr>
## 1 TOTAL
## 2 CHEMICAL, FUNGICIDE
## 3 CHEMICAL, HERBICIDE
## 4 CHEMICAL, INSECTICIDE
## 5 CHEMICAL, OTHER
## 6 FERTILIZER

## # A tibble: 1 x 1
##   temp1a
##   <chr>
## 1 STRAWBERRIES

## [1] " BEARING - APPLICATIONS"

## character(0)

## [1] "TOTAL"      "CHEMICAL"    "FERTILIZER"

## [1] NA      " FUNGICIDE"  " HERBICIDE"  " INSECTICIDE" " OTHER"
```

Table 7: 1-value columns dropped

col_name	col_val
Program	SURVEY
Period	YEAR
temp1a	STRAWBERRIES
temp1b	NA
temp2	BEARING - APPLICATIONS
temp22	CHEMICAL
temp42	CHEMICAL
CV (%)	NA

```
library(dplyr)
# Assuming you've already loaded tidyr, dplyr, and ggplot2

# Filter the dataset for a specific metric (e.g., sales, production)
# Here, I am assuming the metric is in the 'Value' column and selecting data for the year 2021
time_series_data <- strawberry %>%
  filter(Year == 2021 & !is.na(Value) & Value != "(D)") %>%
  select(Year, Value)

# Convert the 'Value' column to numeric
time_series_data$Value <- as.numeric(gsub(",", "", time_series_data$Value))
```

```
## Warning: NAs introduced by coercion
```

```
# Plot the PACF  
# Filter the dataset for a specific metric and remove NAs  
time_series_data <- strawberry %>%  
  filter(Year == 2021 & !is.na(Value) & Value != "(D)") %>%  
  select(Year, Value) %>%  
  filter(!is.na(Value))  
  
time_series_data$Value[time_series_data$Value == "(D)"] <- 0  
  
# Convert the 'Value' column to numeric  
time_series_data$Value <- as.numeric(gsub(",", "", time_series_data$Value))
```

```
## Warning: NAs introduced by coercion
```

```
# Ensure there are no NAs  
time_series_data <- na.omit(time_series_data)  
  
# Plot the PACF  
pacf(time_series_data$Value, main="Partial Autocorrelation for Selected Metric")
```

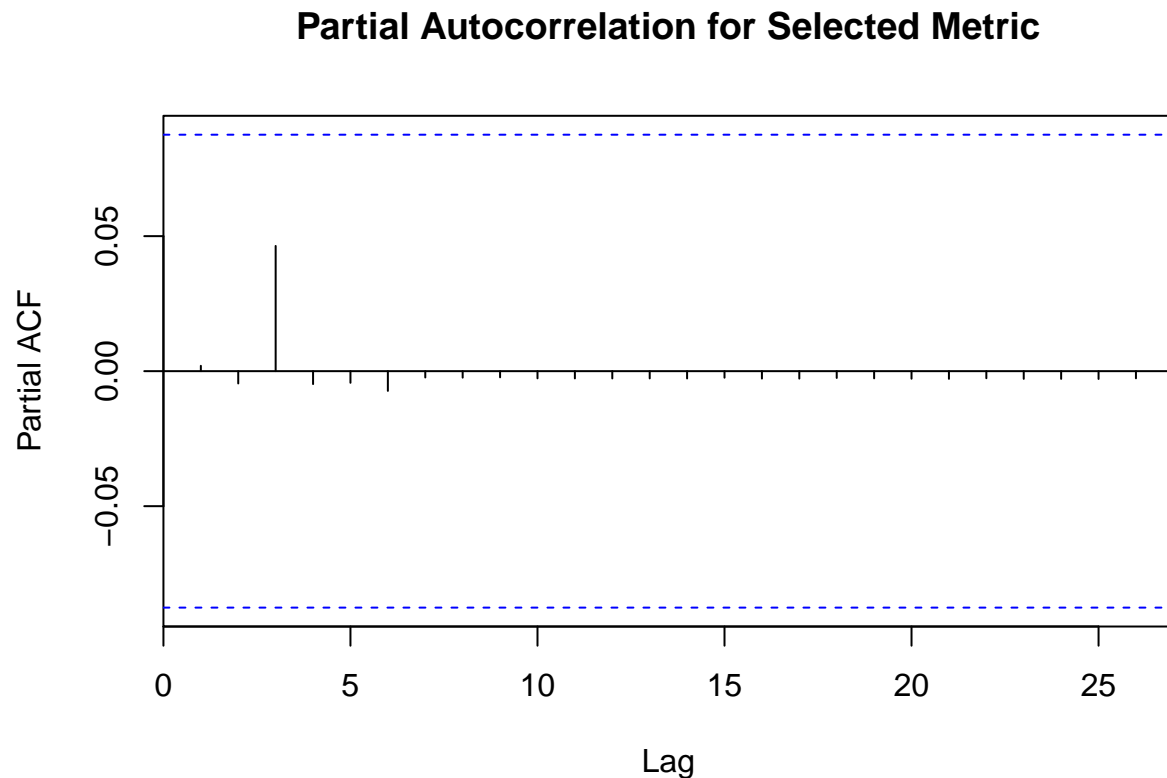


Table 8: dropping 1-value cols - mkt

col_name	col_val
Program	SURVEY
temp1a	STRAWBERRIES
temp23	NA
temp43	NA
temp44	NA
temp45	NA
CV (%)	NA

**These references have been left in the document to help while you are writing. Cite those you use and drop the rest from the final document.**

NASS help

Quick Stats Glossary

Quick Stats Column Definitions

stats by subject

for EPA number lookup epa numbers

Active Pesticide Product Registration Informational Listing

pc number input pesticide chemical search

toxic chemical dashboard

ACToR – Aggregated Computational Toxicology Resource

comptox dashboard

pubChem

The EPA PC (Pesticide Chemical) Code is a unique chemical code number assigned by the EPA to a particular pesticide active ingredient, inert ingredient or mixture of active ingredients.

## Investigating toxic pesticides

start here with chem PC code

step 2 to get label (with warnings) for products using the chemical

International Chemical safety cards

Pesticide Product and Label System

Search by Chemical

CompTox Chemicals Dashboard

Active Pesticide Product Registration Informational Listing

OSHA chemical database

Pesticide Ingredients

NPIC Product Research Online (NPRO)

Databases for Chemical Information

Pesticide Active Ingredients  
TSCA Chemical Substance Inventory  
glyphosate