Stawberries

Keyu Feng

2024-10-14

Strawberries: Data

This is a project abo ut acquiring strawberry data from the USDA-NASS system and then cleaning, organizing, and exploring the data in preparation for data analysis. To get started, I acquired the data from the USDA NASS system and downloaded them in a csv.

Data cleaning and organization references

"An introduction to data cleaning with R" by Edwin de Jonge and Mark van der Loo

"Problems, Methods, and Challenges in Comprehensive Data Cleansing" by Heiko Müller and Johann-Christoph Freytag

Questions about Strawberries

- 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?
- Do Strawberry farmers make money?
- How do the strawberries I buy get to my market?

Strawberry data source and parameters

The data set for this assignment has been selected from:

[USDA_NASS_strawb_2024SEP25.

The data have been stored on NASS here: USDA_NASS_strawb_2024SEP25 .

For the assignment, I stored the csv I downloaded on the MA615 Blackboard as strawberries25_v3.csv.

The data was originally collected at the county, state, and national levels, but the degree of missingness at the state level was too high, so I dropped the county-level data.

There are 5,359 rows and 21 column in the initial data set. The only complete year is 2022, although there is data for years 2018 through 2024.

To work with the data, define a function to remove columns with only single value in all its rows.

To work with this data, split the Census data from the Survey data.

Census data cleaning and organizing

we're examining census data because it's different from survey data

we're isolating organic data

Note that straw_cen has only one year: 2022

Current stats Census date has been isolated and split between Organic and Conventional strawberries

imputed values

We'll start with missing value interpolation for census data:

We will first interpolate the Bearing "Grown" values. Then we can interpolate Bearing and Non-bearing from those values as we discussed in lecture.

Now if GROWN is NA, there are three possibilities:

1 Bearing and Non-bearing are both NA
2 Neither Bearing nor Non-bearing are NA
3 One of Bearing nor Non-bearing are NA

The solution for each of these cases is different:

1 Interpolate from the total for area grown(a check of the dataframe shows you that the total is never NA). We'll try and calculate the average ratio of the missing parts across the non-missing data and split the difference from the total across those
2 Sum Bearing and non-bearing. Easy.

3 Calculate the average proportion of grown that bearing or non bearing(as the case may be) is for the state in question. Then we use that to figure out the value.

Applying our functions now to strawberry census:

Now that we have imputed the values, we still have an adjustment to make. Because we are using proportional averages, the values might no longer add up to our totals. So we need to write a function that will scale our imputed values so that the totals add up.

The happy news is because we are doing this by index, we still know which are the values we have imputed and therefore which values we need to scale.

Imputing acres bearing and non-bearing

We can go one better with imputation where we impute case 1 for the "GROWN" categories together with other case 1s for each. A slight modification to the function above plus some other changes achieves this.

CENSUS TABLES

Survey data cleaning and organizing

Shift data into alignment function

Examine Domain

now look at totals

there are two markets for Strawberries – Fresh Marketing and Processing

make a table for each

from the Survey Totals

we have reports for

Markets: Fresh and Processing Operations: Growing and Production

Florida - California - 2018 -2023

California and Florida chemicals

California and Florida fertilizers

```
library(ggplot2)
library(dplyr)
library(tidyverse)
library(PubChemR)
chemical<- read.csv("new.csv", header = TRUE)
```

#Question 1: Does different chemicals has different effects to the strawberry growth?

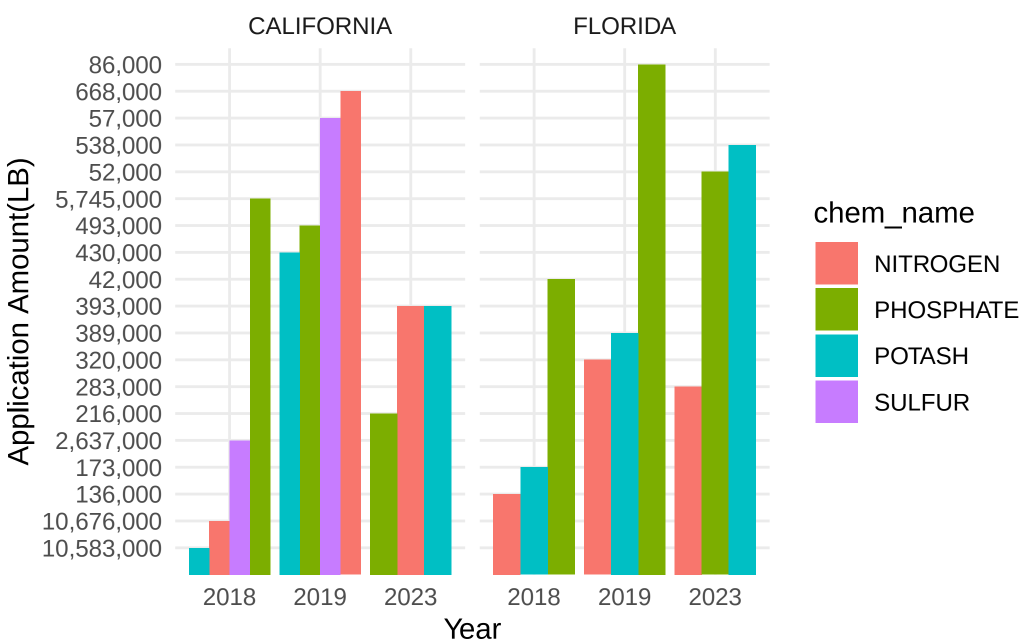
In the first step, I plan to use cleaned data named "new" from USDA-NASS data cleaning-ver7.qmd file. Since only Florida and California used chemicals, it is easy to start with total chemical used.

```
#Total Chemical
data_chem<- chemical %>%
  filter(Year %in% c("2018", "2019", "2023") & State %in% c("CALIFORNIA", "FLORIDA")& chem_na
filter(Value != "(D)") %>% filter(Value != "(NA)")
data_chem
```

```
   Year      State          mk2 measure other chem_name      Value
1  2023 CALIFORNIA APPLICATIONS      LB  <NA>  NITROGEN    393,000
2  2023 CALIFORNIA APPLICATIONS      LB  <NA> PHOSPHATE    216,000
3  2023 CALIFORNIA APPLICATIONS      LB  <NA>    POTASH    393,000
4  2023    FLORIDA APPLICATIONS      LB  <NA>  NITROGEN    283,000
5  2023    FLORIDA APPLICATIONS      LB  <NA> PHOSPHATE     52,000
6  2023    FLORIDA APPLICATIONS      LB  <NA>    POTASH    538,000
7  2019 CALIFORNIA APPLICATIONS      LB  <NA>  NITROGEN    668,000
8  2019 CALIFORNIA APPLICATIONS      LB  <NA> PHOSPHATE    493,000
9  2019 CALIFORNIA APPLICATIONS      LB  <NA>    POTASH    430,000
10 2019 CALIFORNIA APPLICATIONS      LB  <NA>    SULFUR     57,000
11 2019    FLORIDA APPLICATIONS      LB  <NA>  NITROGEN    320,000
12 2019    FLORIDA APPLICATIONS      LB  <NA> PHOSPHATE     86,000
13 2019    FLORIDA APPLICATIONS      LB  <NA>    POTASH    389,000
14 2018 CALIFORNIA APPLICATIONS      LB  <NA>  NITROGEN 10,676,000
15 2018 CALIFORNIA APPLICATIONS      LB  <NA> PHOSPHATE  5,745,000
```

```
16 2018 CALIFORNIA APPLICATIONS      LB  <NA>    POTASH 10,583,000
17 2018 CALIFORNIA APPLICATIONS      LB  <NA>    SULFUR  2,637,000
18 2018    FLORIDA APPLICATIONS      LB  <NA>  NITROGEN    136,000
19 2018    FLORIDA APPLICATIONS      LB  <NA> PHOSPHATE     42,000
20 2018    FLORIDA APPLICATIONS      LB  <NA>    POTASH    173,000
```

```r
# Check the range of values in Value
ggplot(data_chem, aes(x = factor(Year), y = Value, fill = chem_name))+
  geom_bar(stat = "identity", position = "dodge")+
  facet_wrap(~ State)+
  labs(
       x = "Year",
       y = "Application Amount(LB)")+
  theme_minimal()
```

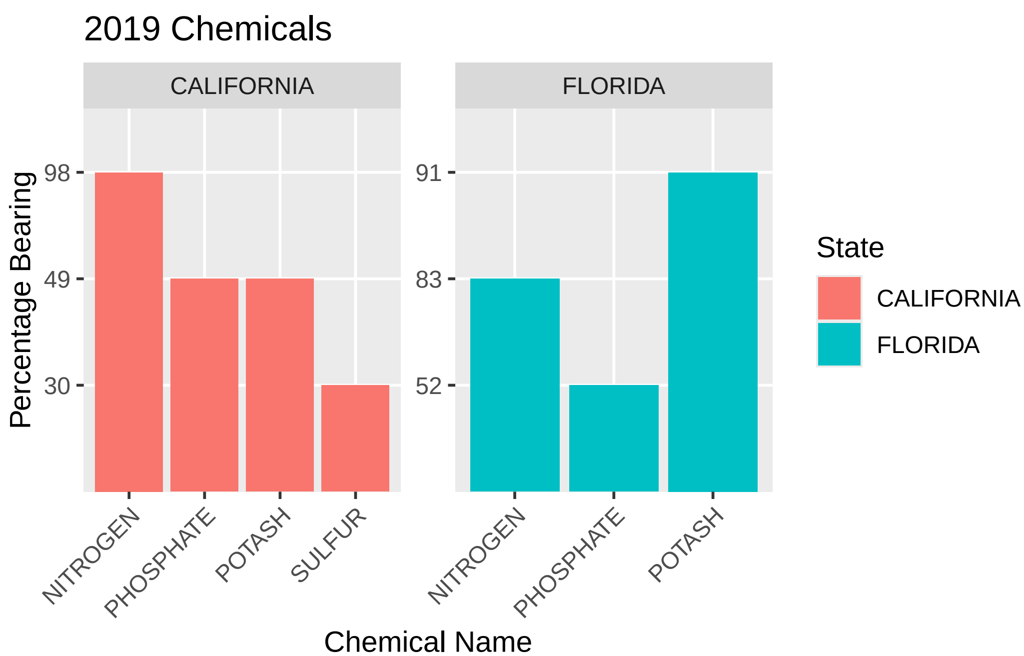


From the histogram, we can find that In both states, Sulfur is consistently the least used chemical across all years, while Potash and Phosphate dominate in terms of usage in various years.

#Plot “PCT OF AREA BEARING”

```r
data_chem1<- chemical %>%
  filter(Year == 2019 & State %in% c("CALIFORNIA", "FLORIDA")& chem_name %in% c("NITROGEN",
```

```
filter(Value != "(D)") %>% filter(Value != "(NA)")
#Plot the growth rate under different chemicals.
ggplot(data_chem1, aes(x = chem_name, y = Value, fill = State))+
  geom_bar(stat = "identity", position = "dodge")+
  facet_wrap(~ State, scales = "free")+
  labs(title = "2019 Chemicals",
       x = "Chemical Name",
       y = "Percentage Bearing",
       fill = "State")+
  theme(axis.text.x = element_text(angle= 45, hjust =1))
```




```
  theme_minimal()
```

```
List of 136
 $ line                            :List of 6
  ..$ colour       : chr "black"
  ..$ linewidth    : num 0.5
  ..$ linetype     : num 1
  ..$ lineend      : chr "butt"
  ..$ arrow        : logi FALSE
  ..$ inherit.blank: logi TRUE
  ..- attr(*, "class")= chr [1:2] "element_line" "element"
```

```
$ rect                           :List of 5
 ..$ fill         : chr "white"
 ..$ colour       : chr "black"
 ..$ linewidth    : num 0.5
 ..$ linetype     : num 1
 ..$ inherit.blank: logi TRUE
 ..- attr(*, "class")= chr [1:2] "element_rect" "element"
$ text                           :List of 11
 ..$ family       : chr ""
 ..$ face         : chr "plain"
 ..$ colour       : chr "black"
 ..$ size         : num 11
 ..$ hjust        : num 0.5
 ..$ vjust        : num 0.5
 ..$ angle        : num 0
 ..$ lineheight   : num 0.9
 ..$ margin       : 'margin' num [1:4] 0points 0points 0points 0points
 .. ..- attr(*, "unit")= int 8
 ..$ debug        : logi FALSE
 ..$ inherit.blank: logi TRUE
 ..- attr(*, "class")= chr [1:2] "element_text" "element"
$ title                          : NULL
$ aspect.ratio                   : NULL
$ axis.title                     : NULL
$ axis.title.x                   :List of 11
 ..$ family       : NULL
 ..$ face         : NULL
 ..$ colour       : NULL
 ..$ size         : NULL
 ..$ hjust        : NULL
 ..$ vjust        : num 1
 ..$ angle        : NULL
 ..$ lineheight   : NULL
 ..$ margin       : 'margin' num [1:4] 2.75points 0points 0points 0points
 .. ..- attr(*, "unit")= int 8
 ..$ debug        : NULL
 ..$ inherit.blank: logi TRUE
 ..- attr(*, "class")= chr [1:2] "element_text" "element"
$ axis.title.x.top               :List of 11
 ..$ family       : NULL
 ..$ face         : NULL
 ..$ colour       : NULL
 ..$ size         : NULL
```

```
  ..$ hjust        : NULL
  ..$ vjust        : num 0
  ..$ angle        : NULL
  ..$ lineheight   : NULL
  ..$ margin       : 'margin' num [1:4] 0points 0points 2.75points 0points
  .. ..- attr(*, "unit")= int 8
  ..$ debug        : NULL
  ..$ inherit.blank: logi TRUE
  ..- attr(*, "class")= chr [1:2] "element_text" "element"
 $ axis.title.x.bottom             : NULL
 $ axis.title.y                    :List of 11
  ..$ family       : NULL
  ..$ face         : NULL
  ..$ colour       : NULL
  ..$ size         : NULL
  ..$ hjust        : NULL
  ..$ vjust        : num 1
  ..$ angle        : num 90
  ..$ lineheight   : NULL
  ..$ margin       : 'margin' num [1:4] 0points 2.75points 0points 0points
  .. ..- attr(*, "unit")= int 8
  ..$ debug        : NULL
  ..$ inherit.blank: logi TRUE
  ..- attr(*, "class")= chr [1:2] "element_text" "element"
 $ axis.title.y.left               : NULL
 $ axis.title.y.right              :List of 11
  ..$ family       : NULL
  ..$ face         : NULL
  ..$ colour       : NULL
  ..$ size         : NULL
  ..$ hjust        : NULL
  ..$ vjust        : num 1
  ..$ angle        : num -90
  ..$ lineheight   : NULL
  ..$ margin       : 'margin' num [1:4] 0points 0points 0points 2.75points
  .. ..- attr(*, "unit")= int 8
  ..$ debug        : NULL
  ..$ inherit.blank: logi TRUE
  ..- attr(*, "class")= chr [1:2] "element_text" "element"
 $ axis.text                       :List of 11
  ..$ family       : NULL
  ..$ face         : NULL
  ..$ colour       : chr "grey30"
```

```
 ..$ size         : 'rel' num 0.8
 ..$ hjust        : NULL
 ..$ vjust        : NULL
 ..$ angle        : NULL
 ..$ lineheight   : NULL
 ..$ margin       : NULL
 ..$ debug        : NULL
 ..$ inherit.blank: logi TRUE
 ..- attr(*, "class")= chr [1:2] "element_text" "element"
$ axis.text.x                     :List of 11
 ..$ family       : NULL
 ..$ face         : NULL
 ..$ colour       : NULL
 ..$ size         : NULL
 ..$ hjust        : NULL
 ..$ vjust        : num 1
 ..$ angle        : NULL
 ..$ lineheight   : NULL
 ..$ margin       : 'margin' num [1:4] 2.2points 0points 0points 0points
 .. ..- attr(*, "unit")= int 8
 ..$ debug        : NULL
 ..$ inherit.blank: logi TRUE
 ..- attr(*, "class")= chr [1:2] "element_text" "element"
$ axis.text.x.top                 :List of 11
 ..$ family       : NULL
 ..$ face         : NULL
 ..$ colour       : NULL
 ..$ size         : NULL
 ..$ hjust        : NULL
 ..$ vjust        : num 0
 ..$ angle        : NULL
 ..$ lineheight   : NULL
 ..$ margin       : 'margin' num [1:4] 0points 0points 2.2points 0points
 .. ..- attr(*, "unit")= int 8
 ..$ debug        : NULL
 ..$ inherit.blank: logi TRUE
 ..- attr(*, "class")= chr [1:2] "element_text" "element"
$ axis.text.x.bottom              : NULL
$ axis.text.y                     :List of 11
 ..$ family       : NULL
 ..$ face         : NULL
 ..$ colour       : NULL
 ..$ size         : NULL
```

```
  ..$ hjust        : num 1
  ..$ vjust        : NULL
  ..$ angle        : NULL
  ..$ lineheight   : NULL
  ..$ margin       : 'margin' num [1:4] 0points 2.2points 0points 0points
  .. ..- attr(*, "unit")= int 8
  ..$ debug        : NULL
  ..$ inherit.blank: logi TRUE
  ..- attr(*, "class")= chr [1:2] "element_text" "element"
 $ axis.text.y.left                : NULL
 $ axis.text.y.right               :List of 11
  ..$ family       : NULL
  ..$ face         : NULL
  ..$ colour       : NULL
  ..$ size         : NULL
  ..$ hjust        : num 0
  ..$ vjust        : NULL
  ..$ angle        : NULL
  ..$ lineheight   : NULL
  ..$ margin       : 'margin' num [1:4] 0points 0points 0points 2.2points
  .. ..- attr(*, "unit")= int 8
  ..$ debug        : NULL
  ..$ inherit.blank: logi TRUE
  ..- attr(*, "class")= chr [1:2] "element_text" "element"
 $ axis.text.theta                 : NULL
 $ axis.text.r                     :List of 11
  ..$ family       : NULL
  ..$ face         : NULL
  ..$ colour       : NULL
  ..$ size         : NULL
  ..$ hjust        : num 0.5
  ..$ vjust        : NULL
  ..$ angle        : NULL
  ..$ lineheight   : NULL
  ..$ margin       : 'margin' num [1:4] 0points 2.2points 0points 2.2points
  .. ..- attr(*, "unit")= int 8
  ..$ debug        : NULL
  ..$ inherit.blank: logi TRUE
  ..- attr(*, "class")= chr [1:2] "element_text" "element"
 $ axis.ticks                      : list()
  ..- attr(*, "class")= chr [1:2] "element_blank" "element"
 $ axis.ticks.x                    : NULL
 $ axis.ticks.x.top                : NULL
```

```
$ axis.ticks.x.bottom            : NULL
$ axis.ticks.y                   : NULL
$ axis.ticks.y.left              : NULL
$ axis.ticks.y.right             : NULL
$ axis.ticks.theta               : NULL
$ axis.ticks.r                   : NULL
$ axis.minor.ticks.x.top         : NULL
$ axis.minor.ticks.x.bottom      : NULL
$ axis.minor.ticks.y.left        : NULL
$ axis.minor.ticks.y.right       : NULL
$ axis.minor.ticks.theta         : NULL
$ axis.minor.ticks.r             : NULL
$ axis.ticks.length              : 'simpleUnit' num 2.75points
 ..- attr(*, "unit")= int 8
$ axis.ticks.length.x            : NULL
$ axis.ticks.length.x.top        : NULL
$ axis.ticks.length.x.bottom     : NULL
$ axis.ticks.length.y            : NULL
$ axis.ticks.length.y.left       : NULL
$ axis.ticks.length.y.right      : NULL
$ axis.ticks.length.theta        : NULL
$ axis.ticks.length.r            : NULL
$ axis.minor.ticks.length        : 'rel' num 0.75
$ axis.minor.ticks.length.x      : NULL
$ axis.minor.ticks.length.x.top  : NULL
$ axis.minor.ticks.length.x.bottom: NULL
$ axis.minor.ticks.length.y      : NULL
$ axis.minor.ticks.length.y.left : NULL
$ axis.minor.ticks.length.y.right : NULL
$ axis.minor.ticks.length.theta  : NULL
$ axis.minor.ticks.length.r      : NULL
$ axis.line                      : list()
 ..- attr(*, "class")= chr [1:2] "element_blank" "element"
$ axis.line.x                    : NULL
$ axis.line.x.top                : NULL
$ axis.line.x.bottom             : NULL
$ axis.line.y                    : NULL
$ axis.line.y.left               : NULL
$ axis.line.y.right              : NULL
$ axis.line.theta                : NULL
$ axis.line.r                    : NULL
$ legend.background              : list()
 ..- attr(*, "class")= chr [1:2] "element_blank" "element"
```

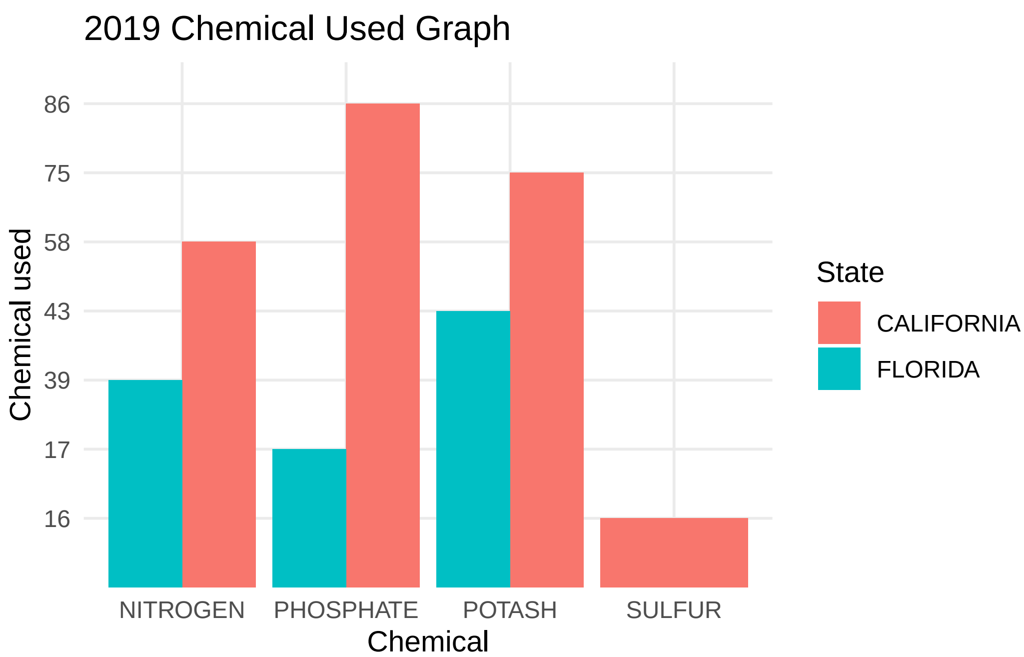
```
$ legend.margin               : 'margin' num [1:4] 5.5points 5.5points 5.5points 5.5poi
 ..- attr(*, "unit")= int 8
$ legend.spacing              : 'simpleUnit' num 11points
 ..- attr(*, "unit")= int 8
$ legend.spacing.x            : NULL
$ legend.spacing.y            : NULL
$ legend.key                  : list()
 ..- attr(*, "class")= chr [1:2] "element_blank" "element"
$ legend.key.size             : 'simpleUnit' num 1.2lines
 ..- attr(*, "unit")= int 3
$ legend.key.height           : NULL
$ legend.key.width            : NULL
$ legend.key.spacing          : 'simpleUnit' num 5.5points
 ..- attr(*, "unit")= int 8
$ legend.key.spacing.x        : NULL
$ legend.key.spacing.y        : NULL
$ legend.frame                : NULL
$ legend.ticks                : NULL
$ legend.ticks.length         : 'rel' num 0.2
$ legend.axis.line            : NULL
$ legend.text                 :List of 11
 ..$ family       : NULL
 ..$ face         : NULL
 ..$ colour       : NULL
 ..$ size         : 'rel' num 0.8
 ..$ hjust        : NULL
 ..$ vjust        : NULL
 ..$ angle        : NULL
 ..$ lineheight   : NULL
 ..$ margin       : NULL
 ..$ debug        : NULL
 ..$ inherit.blank: logi TRUE
 ..- attr(*, "class")= chr [1:2] "element_text" "element"
$ legend.text.position        : NULL
$ legend.title                :List of 11
 ..$ family       : NULL
 ..$ face         : NULL
 ..$ colour       : NULL
 ..$ size         : NULL
 ..$ hjust        : num 0
 ..$ vjust        : NULL
 ..$ angle        : NULL
 ..$ lineheight   : NULL
```

```
  ..$ margin       : NULL
  ..$ debug        : NULL
  ..$ inherit.blank: logi TRUE
  ..- attr(*, "class")= chr [1:2] "element_text" "element"
 $ legend.title.position           : NULL
 $ legend.position                 : chr "right"
 $ legend.position.inside          : NULL
 $ legend.direction                : NULL
 $ legend.byrow                    : NULL
 $ legend.justification            : chr "center"
 $ legend.justification.top        : NULL
 $ legend.justification.bottom     : NULL
 $ legend.justification.left       : NULL
 $ legend.justification.right      : NULL
 $ legend.justification.inside     : NULL
 $ legend.location                 : NULL
 $ legend.box                      : NULL
 $ legend.box.just                 : NULL
 $ legend.box.margin               : 'margin' num [1:4] 0cm 0cm 0cm 0cm
  ..- attr(*, "unit")= int 1
 $ legend.box.background           : list()
  ..- attr(*, "class")= chr [1:2] "element_blank" "element"
 $ legend.box.spacing              : 'simpleUnit' num 11points
  ..- attr(*, "unit")= int 8
  [list output truncated]
 - attr(*, "class")= chr [1:2] "theme" "gg"
 - attr(*, "complete")= logi TRUE
 - attr(*, "validate")= logi TRUE
```

#Plot “LB / ACRE / YEAR”

```
data_chem2<- chemical %>%
  filter(Year == 2019 & State %in% c("CALIFORNIA", "FLORIDA")& chem_name %in% c("NITROGEN", "
filter(Value != "(D)") %>% filter(Value != "(NA)")

#Plot the use of chemical graph
ggplot(data_chem2, aes(x = chem_name, y = Value, fill = State))+
  geom_bar(stat = "identity", position = "dodge")+
  labs(title = "2019 Chemical Used Graph",
       x = "Chemical",
       y = "Chemical used")+
  theme_minimal()
```



Compare the two plots, we can easily find that California uses more chemicals than Florida in 2019. Among them, California uses Nitrogen most while Florida uses Potash most.

Different chemicals have different effects to strawberry bearing. While some may restrict the growth, others can help increase it. For example, enough Nitrogen has positive effects to the area bearing, but too much Potash makes the area bearing smaller.

#Question 2: How does the Fungicide affect the strawberry growth?

Chemicals are used in strawberry Sterilization. I wrote a new csv file named "new1" for my EDA. It contains all information about chemical types and names.

Take AZOXYSTROBIN as an example, lets check if it is hazard to humans.

```
chem<- read.csv("new1.csv")

chem_data<- chem %>%
  filter(type== "FUNGICIDE")%>% filter(Value != "(D)") %>% filter(Value != "(NA)")
head(chem_data)
```

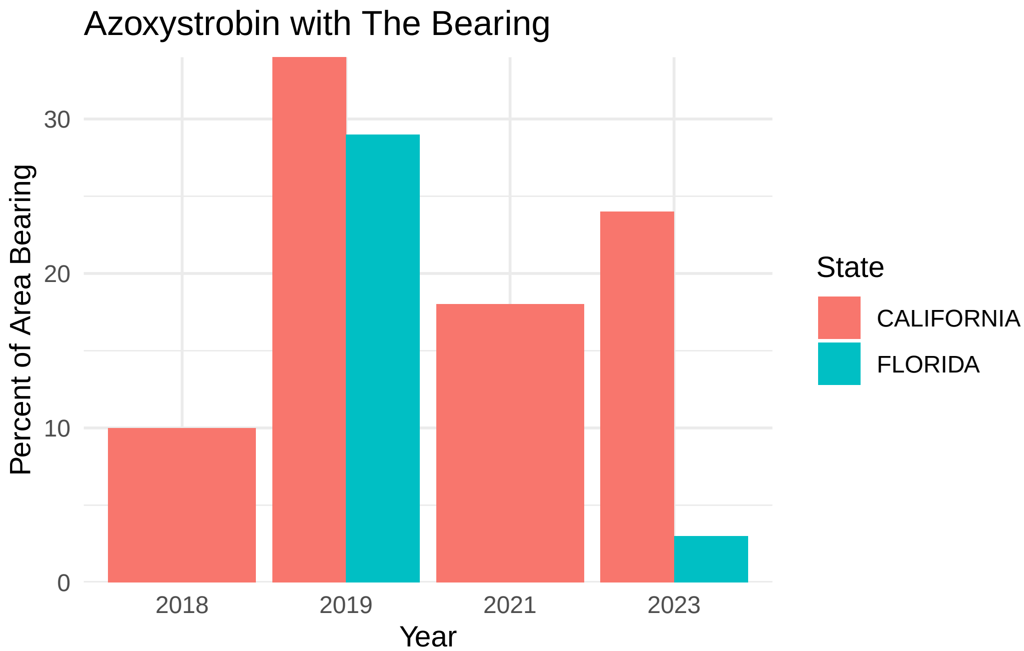
```
  Year      State     mk1          mk2 measure other      type
1 2023 CALIFORNIA BEARING APPLICATIONS      LB  <NA> FUNGICIDE
2 2023 CALIFORNIA BEARING APPLICATIONS      LB  <NA> FUNGICIDE
3 2023 CALIFORNIA BEARING APPLICATIONS      LB  <NA> FUNGICIDE
4 2023 CALIFORNIA BEARING APPLICATIONS      LB  <NA> FUNGICIDE
```

```
5 2023 CALIFORNIA BEARING APPLICATIONS      LB  <NA> FUNGICIDE
6 2023 CALIFORNIA BEARING APPLICATIONS      LB  <NA> FUNGICIDE
          chem_name chem_index   Value
1       AZOXYSTROBIN     128810   3,300
2 BORAX DECAHYDRATE      11102   2,800
3          BOSCALID     128008   6,600
4            CAPTAN      81301 603,100
5        CYPRODINIL     288202  30,300
6        FENHEXAMID      90209   8,600
```

```r
#Compare the use of Azoxystrobin
chem_data1<- chem_data %>% filter(State %in% c("FLORIDA", "CALIFORNIA") & mk2=="TREATED" & cl
chem_data1
```

```
  Year      State     mk1     mk2             measure other      type
1 2023 CALIFORNIA BEARING TREATED PCT OF AREA BEARING   AVG FUNGICIDE
2 2023    FLORIDA BEARING TREATED PCT OF AREA BEARING   AVG FUNGICIDE
3 2021 CALIFORNIA BEARING TREATED PCT OF AREA BEARING   AVG FUNGICIDE
4 2019 CALIFORNIA BEARING TREATED PCT OF AREA BEARING   AVG FUNGICIDE
5 2019    FLORIDA BEARING TREATED PCT OF AREA BEARING   AVG FUNGICIDE
6 2018 CALIFORNIA BEARING TREATED PCT OF AREA BEARING   AVG FUNGICIDE
     chem_name chem_index Value
1 AZOXYSTROBIN     128810    24
2 AZOXYSTROBIN     128810     3
3 AZOXYSTROBIN     128810    18
4 AZOXYSTROBIN     128810    34
5 AZOXYSTROBIN     128810    29
6 AZOXYSTROBIN     128810    10
```

```r
chem_data1$Year <- as.factor(chem_data1$Year)
chem_data1$Value <- as.numeric(chem_data1$Value)
ggplot(chem_data1, aes(x= Year, y= Value, fill= State))+
  geom_bar(stat = "identity", position = "dodge")+
  scale_y_continuous(limits = c(0, max(chem_data1$Value, na.rm = TRUE)), expand = c(0, 0))+
  labs(title = "Azoxystrobin with The Bearing",
       x = "Year",
       y = "Percent of Area Bearing")+
  theme_minimal()
```



From the graph, we can find that Florida didn't use Azoxystrobin in 2018, and 2021.

```
chem_data<- chem %>%
  filter(type== "FUNGICIDE")%>% filter(Value != "(D)") %>% filter(Value != "(NA)")
head(chem_data)
```

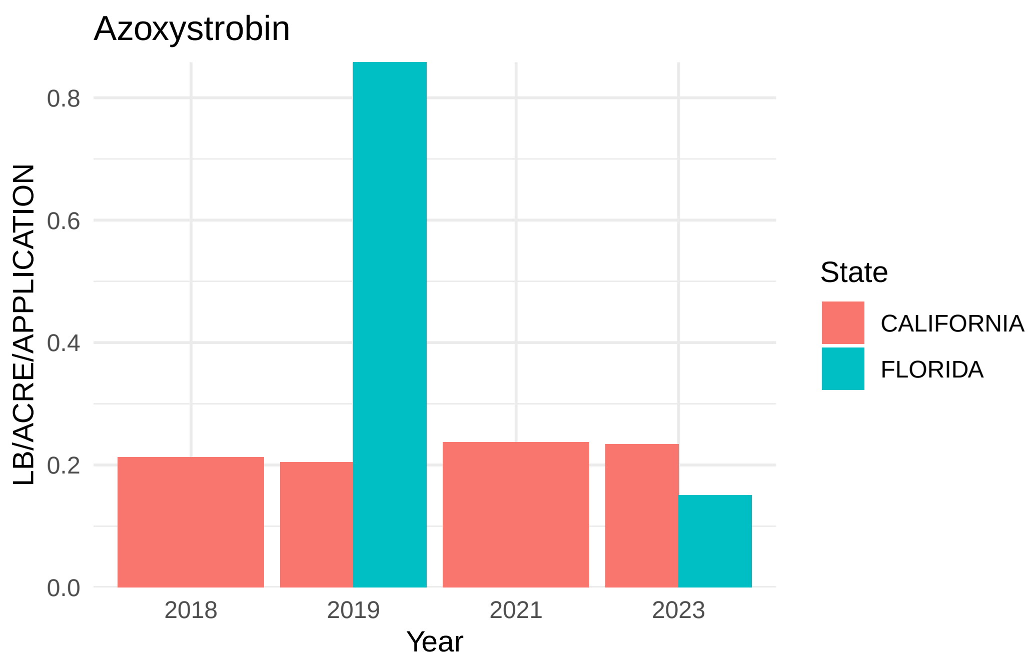
```
  Year      State     mk1          mk2 measure other      type
1 2023 CALIFORNIA BEARING APPLICATIONS      LB  <NA> FUNGICIDE
2 2023 CALIFORNIA BEARING APPLICATIONS      LB  <NA> FUNGICIDE
3 2023 CALIFORNIA BEARING APPLICATIONS      LB  <NA> FUNGICIDE
4 2023 CALIFORNIA BEARING APPLICATIONS      LB  <NA> FUNGICIDE
5 2023 CALIFORNIA BEARING APPLICATIONS      LB  <NA> FUNGICIDE
6 2023 CALIFORNIA BEARING APPLICATIONS      LB  <NA> FUNGICIDE
          chem_name chem_index   Value
1      AZOXYSTROBIN     128810   3,300
2 BORAX DECAHYDRATE      11102   2,800
3          BOSCALID     128008   6,600
4            CAPTAN      81301 603,100
5        CYPRODINIL     288202  30,300
6        FENHEXAMID      90209   8,600
```

```r
chem_data1<- chem_data %>% filter(State %in% c("FLORIDA", "CALIFORNIA") & measure=="LB / ACRE
chem_data1
```

```
  Year      State     mk1          mk2                  measure other      type
1 2023 CALIFORNIA BEARING APPLICATIONS LB / ACRE / APPLICATION   AVG FUNGICIDE
2 2023    FLORIDA BEARING APPLICATIONS LB / ACRE / APPLICATION   AVG FUNGICIDE
3 2021 CALIFORNIA BEARING APPLICATIONS LB / ACRE / APPLICATION   AVG FUNGICIDE
4 2019 CALIFORNIA BEARING APPLICATIONS LB / ACRE / APPLICATION   AVG FUNGICIDE
5 2019    FLORIDA BEARING APPLICATIONS LB / ACRE / APPLICATION   AVG FUNGICIDE
6 2018 CALIFORNIA BEARING APPLICATIONS LB / ACRE / APPLICATION   AVG FUNGICIDE
     chem_name chem_index Value
1 AZOXYSTROBIN     128810 0.234
2 AZOXYSTROBIN     128810 0.151
3 AZOXYSTROBIN     128810 0.237
4 AZOXYSTROBIN     128810 0.205
5 AZOXYSTROBIN     128810 0.858
6 AZOXYSTROBIN     128810 0.213
```

```r
chem_data1$Year <- as.factor(chem_data1$Year)
chem_data1$Value <- as.numeric(chem_data1$Value)

ggplot(chem_data1, aes(x= Year, y= Value, fill= State))+
  geom_bar(stat = "identity", position = "dodge")+
  scale_y_continuous(limits = c(0, max(chem_data1$Value, na.rm = TRUE)), expand = c(0, 0))+
  labs(title = "Azoxystrobin",
       x = "Year",
       y = "LB/ACRE/APPLICATION")+
  theme_minimal()
```



From the research, the best rate for Azoxystrobin is about 0.65-0.975 lb/Acre/Application. Using the histogram, we can find the highest rates are below the limit, so we can conclude that the use of it is normative.

However, it is hard to predict the relationship between the use of Azoxystrobin and the bearing rate of strawberries. In our histograms, we can find that the use of Azoxystrobin in California are similar in 2018 and 2019, but the bearing area under it have a large difference between 2018 and 2019. However, in Florida, although there is 0.8 lb/Acre/Application, the percent of area bearing is still higher. Therefore, there might be other factors that affect the growth of strawberries.

#Question 3: How does a hazard pesticide affect the growth of strawberries?

Next goal is trying to explore the use of deadly carcinogens. Malathion is an organophosphate insecticide which acts as an acetylcholinesterase inhibitor. We start by choosing the true columns.

```
chem_data2<- chem %>% filter(type=="INSECTICIDE") %>% filter(State %in% c("FLORIDA", "CALIFOR
chem_data2
```

```
   Year      State    mk1          mk2                   measure other
1  2023 CALIFORNIA BEARING APPLICATIONS                       LB  <NA>
2  2023 CALIFORNIA BEARING APPLICATIONS LB / ACRE / APPLICATION   AVG
3  2023 CALIFORNIA BEARING APPLICATIONS        LB / ACRE / YEAR   AVG
```

```
4  2023 CALIFORNIA BEARING APPLICATIONS                  NUMBER   AVG
5  2023 CALIFORNIA BEARING      TREATED     PCT OF AREA BEARING   AVG
6  2023    FLORIDA BEARING APPLICATIONS                      LB  <NA>
7  2023    FLORIDA BEARING APPLICATIONS LB / ACRE / APPLICATION   AVG
8  2023    FLORIDA BEARING APPLICATIONS        LB / ACRE / YEAR   AVG
9  2023    FLORIDA BEARING APPLICATIONS                  NUMBER   AVG
10 2023    FLORIDA BEARING      TREATED     PCT OF AREA BEARING   AVG
11 2021 CALIFORNIA BEARING APPLICATIONS                      LB  <NA>
12 2021 CALIFORNIA BEARING APPLICATIONS LB / ACRE / APPLICATION   AVG
13 2021 CALIFORNIA BEARING APPLICATIONS        LB / ACRE / YEAR   AVG
14 2021 CALIFORNIA BEARING APPLICATIONS                  NUMBER   AVG
15 2021 CALIFORNIA BEARING      TREATED     PCT OF AREA BEARING   AVG
16 2021    FLORIDA BEARING APPLICATIONS                      LB  <NA>
17 2021    FLORIDA BEARING APPLICATIONS LB / ACRE / APPLICATION   AVG
18 2021    FLORIDA BEARING APPLICATIONS        LB / ACRE / YEAR   AVG
19 2021    FLORIDA BEARING APPLICATIONS                  NUMBER   AVG
20 2021    FLORIDA BEARING      TREATED     PCT OF AREA BEARING   AVG
21 2019 CALIFORNIA BEARING APPLICATIONS                      LB  <NA>
22 2019 CALIFORNIA BEARING APPLICATIONS LB / ACRE / APPLICATION   AVG
23 2019 CALIFORNIA BEARING APPLICATIONS        LB / ACRE / YEAR   AVG
24 2019 CALIFORNIA BEARING APPLICATIONS                  NUMBER   AVG
25 2019 CALIFORNIA BEARING      TREATED     PCT OF AREA BEARING   AVG
26 2019    FLORIDA BEARING APPLICATIONS                      LB  <NA>
27 2019    FLORIDA BEARING APPLICATIONS LB / ACRE / APPLICATION   AVG
28 2019    FLORIDA BEARING APPLICATIONS        LB / ACRE / YEAR   AVG
29 2019    FLORIDA BEARING APPLICATIONS                  NUMBER   AVG
30 2019    FLORIDA BEARING      TREATED     PCT OF AREA BEARING   AVG
31 2018 CALIFORNIA BEARING APPLICATIONS                      LB  <NA>
32 2018 CALIFORNIA BEARING APPLICATIONS LB / ACRE / APPLICATION   AVG
33 2018 CALIFORNIA BEARING APPLICATIONS        LB / ACRE / YEAR   AVG
34 2018 CALIFORNIA BEARING APPLICATIONS                  NUMBER   AVG
35 2018 CALIFORNIA BEARING      TREATED     PCT OF AREA BEARING   AVG
36 2018    FLORIDA BEARING APPLICATIONS                      LB  <NA>
37 2018    FLORIDA BEARING APPLICATIONS LB / ACRE / APPLICATION   AVG
38 2018    FLORIDA BEARING APPLICATIONS        LB / ACRE / YEAR   AVG
39 2018    FLORIDA BEARING APPLICATIONS                  NUMBER   AVG
40 2018    FLORIDA BEARING      TREATED     PCT OF AREA BEARING   AVG
          type chem_name chem_index  Value
1  INSECTICIDE MALATHION      57701 19,400
2  INSECTICIDE MALATHION      57701  1.807
3  INSECTICIDE MALATHION      57701  2.398
4  INSECTICIDE MALATHION      57701    1.3
5  INSECTICIDE MALATHION      57701     19
```

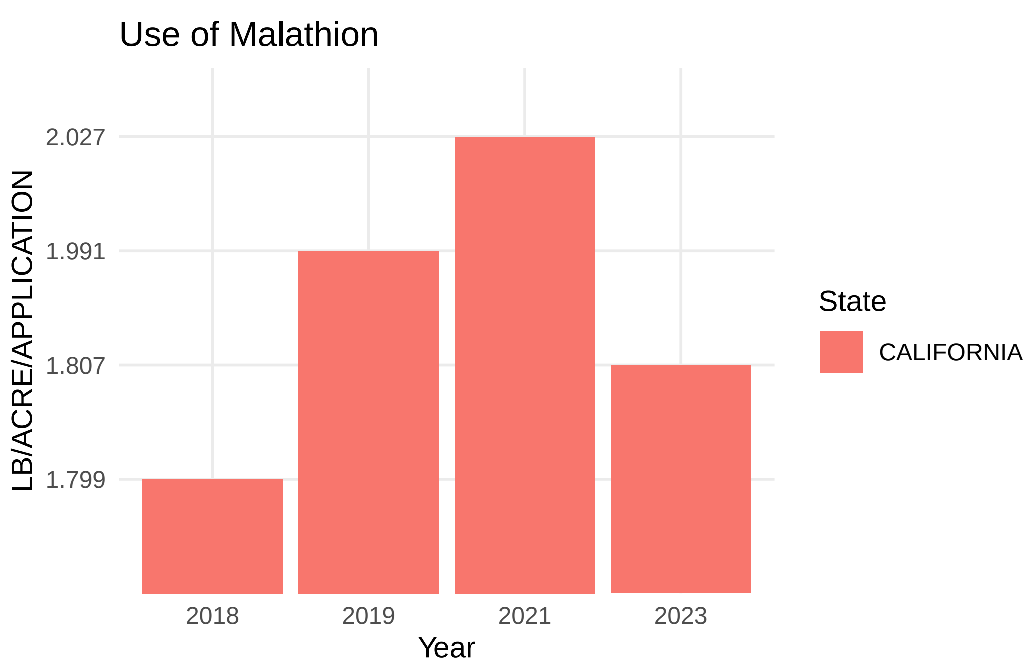
```
6  INSECTICIDE MALATHION      57701    (D)
7  INSECTICIDE MALATHION      57701    (D)
8  INSECTICIDE MALATHION      57701    (D)
9  INSECTICIDE MALATHION      57701    (D)
10 INSECTICIDE MALATHION      57701    (D)
11 INSECTICIDE MALATHION      57701 29,100
12 INSECTICIDE MALATHION      57701  2.027
13 INSECTICIDE MALATHION      57701  2.443
14 INSECTICIDE MALATHION      57701    1.2
15 INSECTICIDE MALATHION      57701     30
16 INSECTICIDE MALATHION      57701    (D)
17 INSECTICIDE MALATHION      57701    (D)
18 INSECTICIDE MALATHION      57701    (D)
19 INSECTICIDE MALATHION      57701    (D)
20 INSECTICIDE MALATHION      57701    (D)
21 INSECTICIDE MALATHION      57701 56,700
22 INSECTICIDE MALATHION      57701  1.991
23 INSECTICIDE MALATHION      57701  5.571
24 INSECTICIDE MALATHION      57701    2.8
25 INSECTICIDE MALATHION      57701     28
26 INSECTICIDE MALATHION      57701    (D)
27 INSECTICIDE MALATHION      57701    (D)
28 INSECTICIDE MALATHION      57701    (D)
29 INSECTICIDE MALATHION      57701    (D)
30 INSECTICIDE MALATHION      57701    (D)
31 INSECTICIDE MALATHION      57701  8,000
32 INSECTICIDE MALATHION      57701  1.799
33 INSECTICIDE MALATHION      57701  3.062
34 INSECTICIDE MALATHION      57701    1.7
35 INSECTICIDE MALATHION      57701      7
36 INSECTICIDE MALATHION      57701    (D)
37 INSECTICIDE MALATHION      57701    (D)
38 INSECTICIDE MALATHION      57701    (D)
39 INSECTICIDE MALATHION      57701    (D)
40 INSECTICIDE MALATHION      57701    (D)
```

```r
chem_data2<- chem_data2 %>% filter(measure== "LB / ACRE / APPLICATION") %>% filter(Value != '
chem_data2
```

```
  Year      State     mk1          mk2                   measure other
1 2023 CALIFORNIA BEARING APPLICATIONS LB / ACRE / APPLICATION   AVG
2 2021 CALIFORNIA BEARING APPLICATIONS LB / ACRE / APPLICATION   AVG
```

```
3 2019 CALIFORNIA BEARING APPLICATIONS LB / ACRE / APPLICATION   AVG
4 2018 CALIFORNIA BEARING APPLICATIONS LB / ACRE / APPLICATION   AVG
        type chem_name chem_index Value
1 INSECTICIDE MALATHION      57701 1.807
2 INSECTICIDE MALATHION      57701 2.027
3 INSECTICIDE MALATHION      57701 1.991
4 INSECTICIDE MALATHION      57701 1.799
```

```
#
chem_data2$Year <- as.factor(chem_data2$Year)
ggplot(chem_data2, aes(x= Year, y= Value, fill= State))+
  geom_bar(stat = "identity", position = "dodge")+
  labs(title = "Use of Malathion",
       x = "Year",
       y = "LB/ACRE/APPLICATION")+
  theme_minimal()
```




```
#Input hazard information

GHS_searcher<-function(result_json_object){
  result<-result_json_object
  for (i in 1:length(result[["result"]][["Hierarchies"]][["Hierarchy"]])){
```

```
    if(result[["result"]][["Hierarchies"]][["Hierarchy"]][[i]][["SourceName"]]=="GHS Classifi
      return(i)
    }

  }
}

hazards_retriever<-function(index,result_json_object){
  result<-result_json_object
  hierarchy<-result[["result"]][["Hierarchies"]][["Hierarchy"]][[index]]
  i<-1
  output_list<-rep(NA,length(hierarchy[["Node"]]))
  while(str_detect(hierarchy[["Node"]][[i]][["Information"]][["Name"]],"H") & i<length(hiera
    output_list[i]<-hierarchy[["Node"]][[i]][["Information"]][["Name"]]
    i<-i+1
  }
  return(output_list[!is.na(output_list)])
}
chemical_vec<-c("azoxystrobin","malathion")
#
result_f<-get_pug_rest(identifier = "azoxystrobin", namespace = "name", domain = "compound",o

hazards_retriever(GHS_searcher(result_f),result_f)
```

```
[1] "H331: Toxic if inhaled [Danger Acute toxicity, inhalation]"
[2] "H300: Health Hazards"
[3] "Hazard Statement Codes"
[4] "H370: Causes damage to organs [Danger Specific target organ toxicity, single exposure]"
[5] "H400: Very toxic to aquatic life [Warning Hazardous to the aquatic environment, acute ha
[6] "H400: Environmental Hazards"
[7] "H410: Very toxic to aquatic life with long lasting effects [Warning Hazardous to the aqu
```

```
result_d<-get_pug_rest(identifier = "malathion", namespace = "name", domain = "compound",oper

hazards_retriever(GHS_searcher(result_d),result_d)
```

```
 [1] "H302: Harmful if swallowed [Warning Acute toxicity, oral]"
 [2] "H300: Health Hazards"
 [3] "Hazard Statement Codes"
 [4] "H317: May cause an allergic skin reaction [Warning Sensitization, Skin]"
 [5] "H320: Causes eye irritation [Warning Serious eye damage/eye irritation]"
```

```
 [6] "H331: Toxic if inhaled [Danger Acute toxicity, inhalation]"
 [7] "H341: Suspected of causing genetic defects [Warning Germ cell mutagenicity]"
 [8] "H350: May cause cancer [Danger Carcinogenicity]"
 [9] "H370: Causes damage to organs [Danger Specific target organ toxicity, single exposure]"
[10] "H372: Causes damage to organs through prolonged or repeated exposure [Danger Specific t
[11] "H373: May causes damage to organs through prolonged or repeated exposure [Warning Speci
[12] "H400: Very toxic to aquatic life [Warning Hazardous to the aquatic environment, acute h
[13] "H400: Environmental Hazards"
[14] "H410: Very toxic to aquatic life with long lasting effects [Warning Hazardous to the aq
```

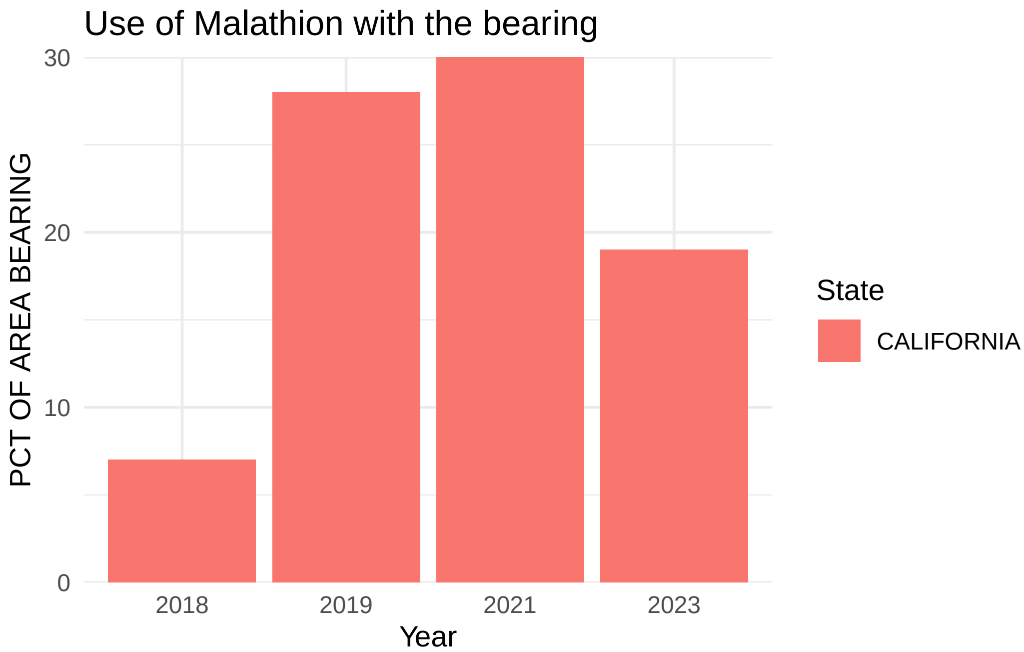
According to the records, Malathion can damage organs and genetic defects as well as being environment hazards, the best rate for using is 1.25 - 2.5 lb/ Acre/ Application. In California, the highest rate is 2.027 lb/ Acre/ Application while the lowest rate is 1.799 lb/Acre/Application, both of them are under the limit. We can then explore the bearing area to find the relationship between them.

```r
chem_data3<- chem %>% filter(type=="INSECTICIDE") %>% filter(State %in% c("FLORIDA", "CALIFOR
chem_data3<- chem_data3 %>% filter(measure== "PCT OF AREA BEARING") %>% filter(Value != "(D)"
chem_data3
```

```
  Year      State     mk1     mk2             measure other        type
1 2023 CALIFORNIA BEARING TREATED PCT OF AREA BEARING   AVG INSECTICIDE
2 2021 CALIFORNIA BEARING TREATED PCT OF AREA BEARING   AVG INSECTICIDE
3 2019 CALIFORNIA BEARING TREATED PCT OF AREA BEARING   AVG INSECTICIDE
4 2018 CALIFORNIA BEARING TREATED PCT OF AREA BEARING   AVG INSECTICIDE
  chem_name chem_index Value
1 MALATHION      57701    19
2 MALATHION      57701    30
3 MALATHION      57701    28
4 MALATHION      57701     7
```

```r
chem_data3$Year <- as.factor(chem_data3$Year)
chem_data3$Value <- as.numeric(chem_data3$Value)

ggplot(chem_data3, aes(x= Year, y= Value, fill= State))+
  geom_bar(stat = "identity", position = "dodge")+
  scale_y_continuous(limits = c(0, max(chem_data3$Value, na.rm = TRUE)), expand = c(0, 0))+
  labs(title = "Use of Malathion with the bearing",
       x = "Year",
       y = "PCT OF AREA BEARING")+
  theme_minimal()
```



We can find that the use of Malathion truly has some good effect to the bearing rate, even though it is dangerous to humans. For instance, in 2018, the use of Malathion was little, leading to the low bearing rate, this is because Malathion is a great insecticide, in that year, little use of Malathion could not restrict the pests.

Chemicals used in strawberry cultivaion

Six deadly carcinogens from WHO list

captafol

ethylene dibromide also

glyphosate See also 1
2
3
4

malathion 1 2

diazinon 1 2 3

dichlorophenyltrichloroethane (DDT) 1 2 [3](https://www.epa.gov/ingredients-used-pesticide-products/ddt-brief-history-and-status#:~:text=DDT%20(dichloro%2Ddiphenyl%2Dtrichloroethane,both%20mi

For contrast

Azadirachtin 1 2 3

Sources of agricultural chemical information

for EPA number lookup epa numbers

Active Pesticide Product Registration Informational Listing

CAS for Methyl Bromide

pesticide chemical search

toxic chemical 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

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