EDAMA615

Hongnan Su

2024-12-08

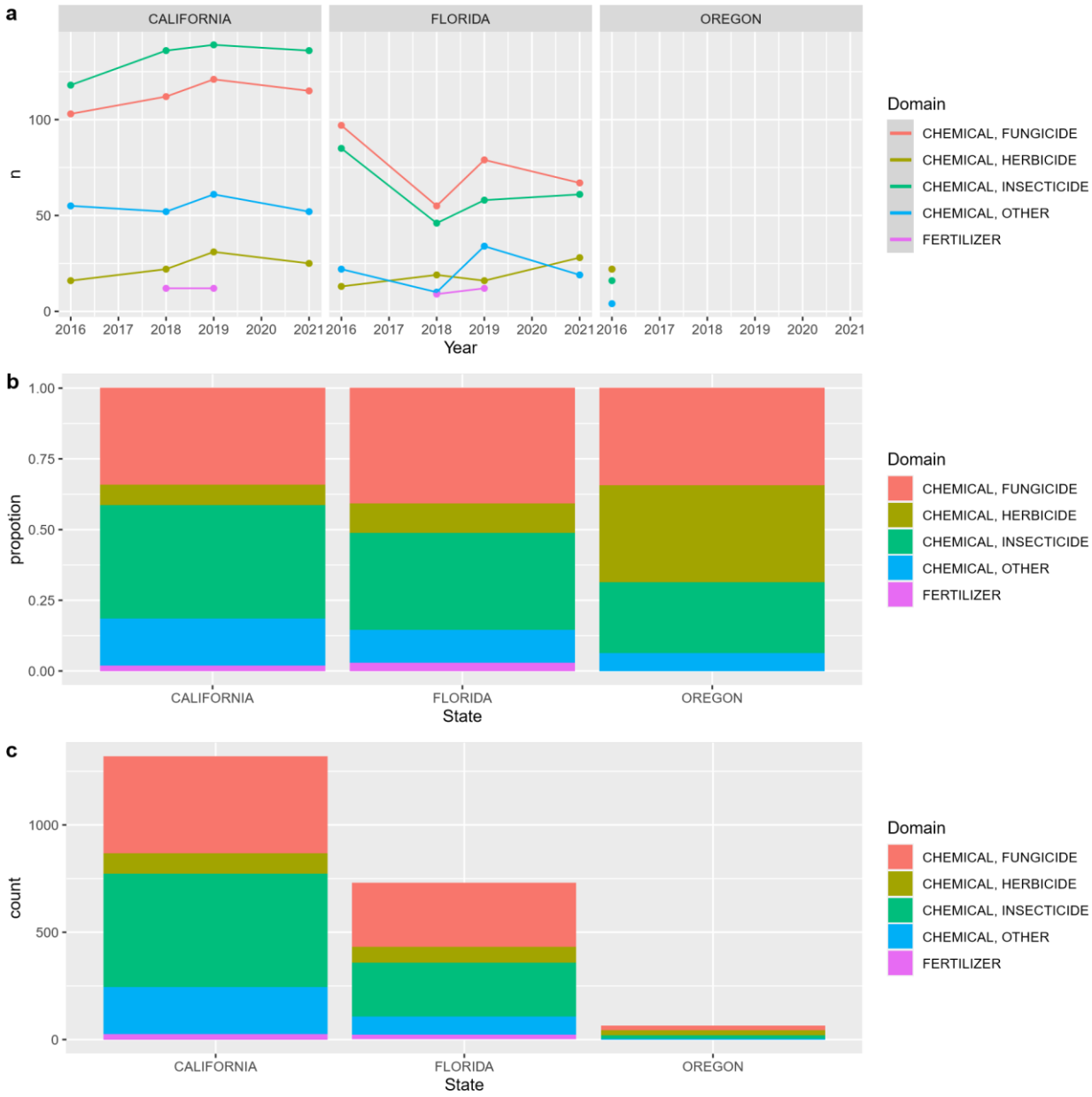
Chemicals

```
#subset the chemical rows
chemical <- filter(strawb, Domain != 'ORGANIC STATUS' & Domain != "TOTAL") head(chemical)
```

```
## # A tibble: 6 x 13
##            Program Year Period State                  State~1 Straw~2 type items units Domain Domai~3
##       <chr>          <dbl> <chr> <chr>                <dbl> <chr>             <chr> <chr> <chr> <chr> <chr>
  ## 1 SURVEY        2016 YEAR           CALIFOR~            6 STRAWB~ " BE~ " ME~ <NA> CHEMI~ CHEMIC~ ## 2
  SURVEY  2016 YEAR           CALIFOR~            6 STRAWB~ " BE~ " ME~ <NA> CHEMI~ CHEMIC~ ## 3 SURVEY
              2016 YEAR           CALIFOR~            6 STRAWB~ " BE~ " ME~ <NA> CHEMI~ CHEMIC~ ## 4 SURVEY
              2016 YEAR           CALIFOR~            6 STRAWB~ " BE~ " ME~ <NA> CHEMI~ CHEMIC~ ## 5 SURVEY
                      2016 YEAR           CALIFOR~            6 STRAWB~ " BE~ " ME~ <NA> CHEMI~ CHEMIC~
## 6 SURVEY        2016 YEAR        CALIFOR~                  6 STRAWB~ " BE~ " ME~ <NA> CHEMI~ CHEMIC~
## # ... with 2 more variables: Value <chr>, 'CV (%)' <chr>, and abbreviated ## #  variable names 1:
'State ANSI', 2: Strawberries, 3: 'Domain Category'
```

```
defaultW <- getOption("warn")

options(warn = -1) z<-
chemical %>% count(Year, Domain,
State)

dot_plot = ggplot(data = z,mapping = aes(x = Year, y = n, color=Domain)) + geom_point() +
   geom_line()+
   geom_smooth(method = 'loess',formula=y~x)+ facet_grid(~State)
bar1 = ggplot(chemical) + geom_bar(aes(x = State, fill = Domain), position = "fill")+labs(y='propotion')
bar2 = ggplot(chemical) + geom_bar(aes(x = State, fill = Domain))
plot_grid(dot_plot, bar1, bar2, labels = "auto", ncol = 1)
```

```
## geom_path: Each group consists of only one observation. Do you need to adjust ## the group
aesthetic?
```



Three states use chemicals in strawberry cultivation, California, Florida and Oregon. We combine three dimensions, year, states and chemical types into one plot. When we compare horizontally, we can see the number of chemicals used between different states. When comparing vertically, we can see the number of different chemicals used in the same state. In our plot, we can also find a preference for the type of chemical among states.

California is the state that uses the most chemicals, with the highest use of fungicides and insecticides. Based on the data we have, annual chemical use in it remained stable. Florida also likes to use these two chemicals as well, but it reduced their use of them. Oregon uses the least amount of chemicals.

Poisons

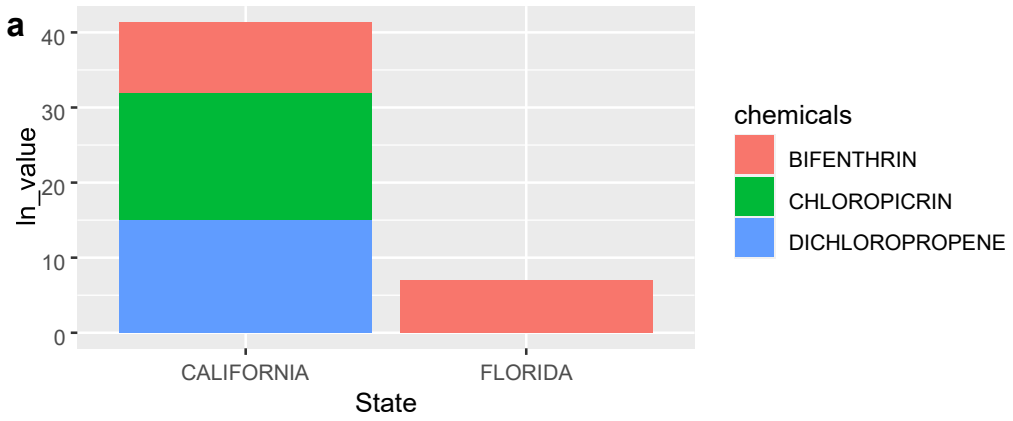
For poisons, we analyze them through states and types.

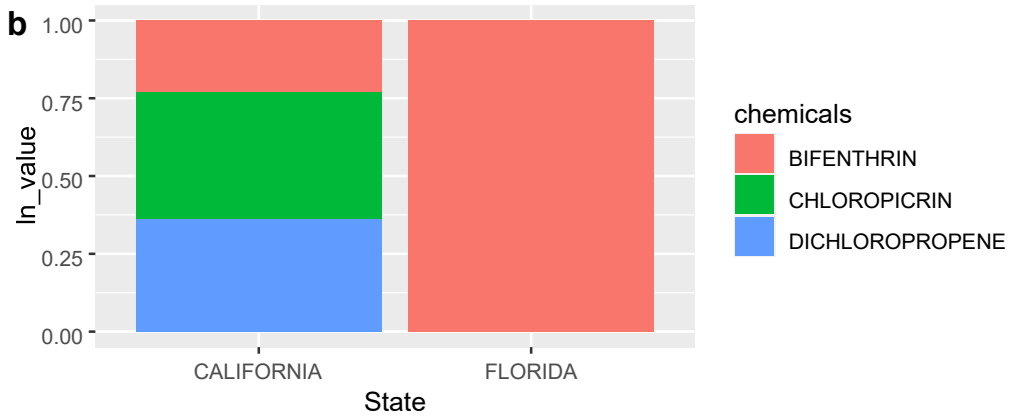
```
che<-strawb[grep("CARBENDAZIM|BIFENTHRIN|METHYL BROMIDE|CHLOROPICRIN|DICHLOROPROPENE"
                 , strawb$`Domain Category`), ]
head(che)
```

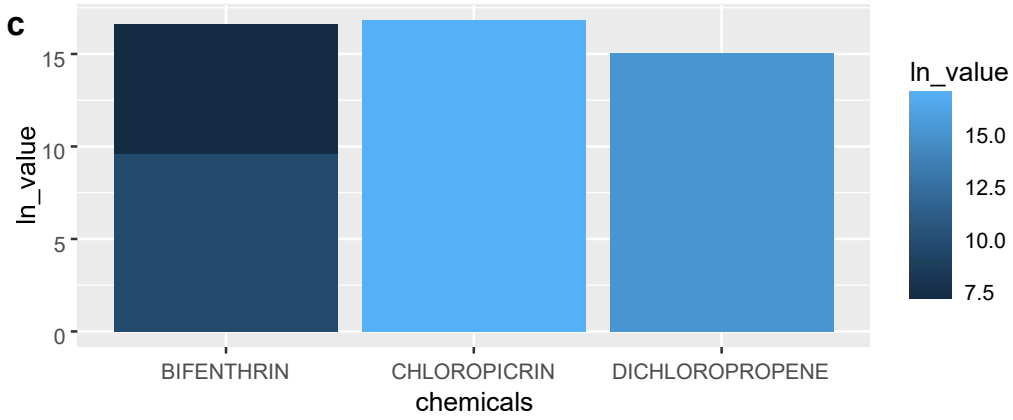
```
## # A tibble: 6 x 13
##   Program Year Period State  State~1 Straw~2 type  items units Domain Domai~3
##   <chr>   <dbl> <chr>  <chr>  <dbl> <chr>  <chr> <chr> <chr> <chr> <chr>
## 1 SURVEY   2016 YEAR   CALIFOR~      6 STRAWB~ " BE~ " ME~ <NA>  CHEMI~ CHEMIC~
## 2 SURVEY   2016 YEAR   CALIFOR~      6 STRAWB~ " BE~ " ME~ <NA>  CHEMI~ CHEMIC~
## 3 SURVEY   2016 YEAR   CALIFOR~      6 STRAWB~ " BE~ " ME~ <NA>  CHEMI~ CHEMIC~
## 4 SURVEY   2016 YEAR   CALIFOR~      6 STRAWB~ " BE~ " ME~ <NA>  CHEMI~ CHEMIC~
## 5 SURVEY   2016 YEAR   CALIFOR~      6 STRAWB~ " BE~ " ME~ " AV~ CHEMI~ CHEMIC~
## 6 SURVEY   2016 YEAR   CALIFOR~      6 STRAWB~ " BE~ " ME~ " AV~ CHEMI~ CHEMIC~
## # ... with 2 more variables: Value <chr>, 'CV (%)' <chr>, and abbreviated
## #   variable names 1: 'State ANSI', 2: Strawberries, 3: 'Domain Category'
```

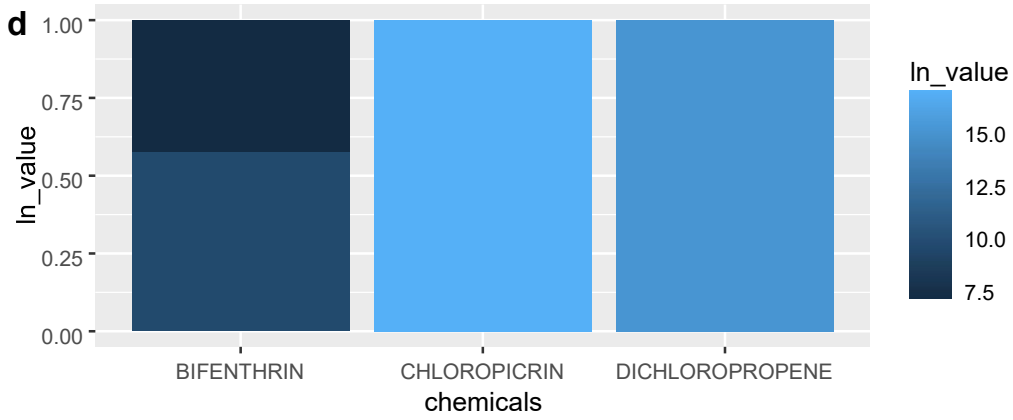
```
# group the value by states and chemicals and sum them
new<-aggregate(as.numeric(che$Value)~ che$State + che$`Domain Category`, data=che, FUN=
#rename the columns
names(new)[c(1,2,3)] =c("State","chemical","ln_value") #since the difference
of values are huge, take log new$ln_value = log(new$ln_value)
new$chemicals=c('BIFENTHRIN','BIFENTHRIN','CHLOROPICRIN','DICHLOROPROPENE')

#plot the them in stacks
bar3 = ggplot(new, aes(fill=chemicals, y= ln_value, x=State)) + geom_bar(position="stack", stat="identity")

bar4 = ggplot(new, aes(fill=chemicals, y= ln_value, x=State)) + geom_bar(position="fill", stat="identity")

bar5 = ggplot(new, aes(fill=ln_value, y=ln_value , x=chemicals)) + geom_bar(position="stack", stat="identity")

bar6 = ggplot(new, aes(fill=ln_value, y=ln_value , x=chemicals)) +
geom_bar(position="fill", stat="identity") plot_grid(bar3, bar4, bar5, bar6, labels =
"auto", ncol = 1)
```

sum)


a
40
30
20
10
0
ln_value
CALIFORNIA
FLORIDA
State
chemicals
BIFENTHRIN
CHLOROPICRIN
DICHLOROPROPENE


b
1.00
0.75
0.50
0.25
0.00
ln_value
CALIFORNIA
FLORIDA
State
chemicals
BIFENTHRIN
CHLOROPICRIN
DICHLOROPROPENE


c
15
10
5
0
ln_value
BIFENTHRIN
CHLOROPICRIN
DICHLOROPROPENE
chemicals
ln_value
15.0
12.5
10.0
7.5


d
1.00
0.75
0.50
0.25
0.00
ln_value
BIFENTHRIN
CHLOROPICRIN
DICHLOROPROPENE
chemicals
ln_value
15.0
12.5
10.0
7.5

This plot shows the poisons used in different states and their quantities. California used more poisons than Florida and it used the three poisons in similar proportions. Florida only used the Bifenthrin.

```
#This chunk is only for export the upcoming word cloud(delete '#' if needed)

#library(webshot)
#webshot::install_phantomjs()

#library("htmlwidgets")
```

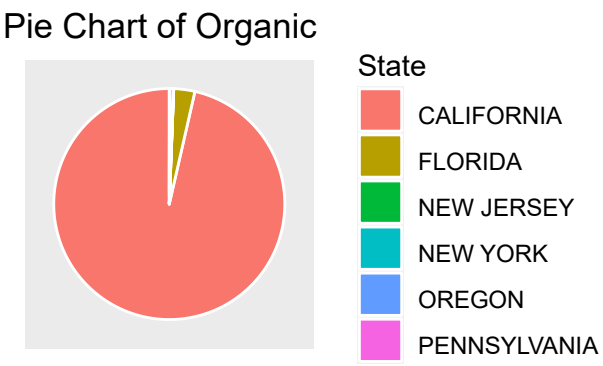
```
library(wordcloud2)

#make a table contains the chemical names with their frequency df1<-
as.data.frame(table(strawb_chem$chem_name))

#plot the wordcloud
wc = wordcloud2(df1, size = 0.4) wc
```

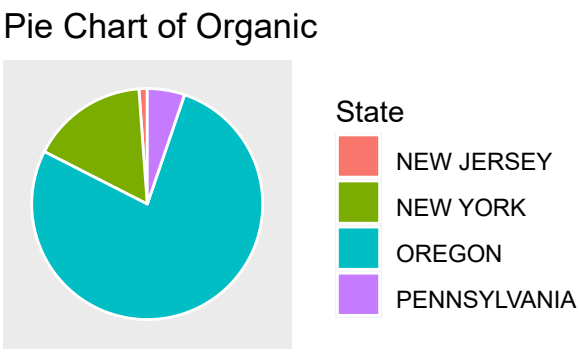
```
#export html and png file if needed (delete '#' if needed)
#saveWidget(wc,"tmp.html",selfcontained = F)
#webshot("tmp.html","fig_1.png", delay =5, vwidth = 1000, vheight=1000)
```

This word cloud shows the frequency of use of different poisons. **Organic**

```
library(tidyverse)
strawb_organicsub <- strawb_organic %>% filter(State != "CALIFORNIA" & State != "FLORIDA")
# group the value by states and chemicals and sum them
new<-aggregate(as.numeric(strawb_organic$Value)~ strawb_organic$State, data=strawb_organic, new2<-                FUN=sum)
aggregate(as.numeric(strawb_organicsub$Value)~ strawb_organicsub$State, data=
#rename the columns
names(new)[c(1,2)] =c("State","Value") names(new2)[c(1,2)]
=c("State","Value")
                                                                                                    strawb_organicsub, FU
```

```
# Barplot

organic3 <- ggplot(new, aes(x="",y=Value, fill=State)) +
  geom_bar(stat="identity", width=1,color="white") + coord_polar("y",
  start=0) + theme(axis.text = element_blank(), axis.ticks =
  element_blank(), panel.grid = element_blank())+
  labs(fill = "State", x =
         NULL, y = NULL,
         title = "Pie Chart of Organic")
organic4 <- ggplot(new2, aes(x="",y=Value, fill=State)) +
  geom_bar(stat="identity", width=1,color="white") + coord_polar("y",
  start=0) + theme(axis.text = element_blank(), axis.ticks =
  element_blank(), panel.grid = element_blank())+
  labs(fill = "State", x =
         NULL, y = NULL,
         title = "Pie Chart of Organic") plot_grid(organic3,organic4)
```





According to the plots, the sales value of organic strawberries in California is the highest and in New Jersey is the lowest one.