EDA Document for MA615 Midterm Project

Andrew Sisitzky, Daniel Xu, Guangze Yu, Yuyang Li

November 2nd, 2021

Topic we study:

How the human-harm pesticide usage differ as time passes in the major strawberries producing areas in the US.

Data wrangle and clean:

We began by get the data from USDA about strawberries harvest and treatments across the US in year 2016, 2018 and 2019 and a dataset about the pesticide types with their toxicity towards human(Carcinogen, Hormone.Disruptor, Neurotoxins and Developmental.or.Reproductive.Toxins) and bee.

For data cleaning: Firstly, we merge those two dataset by the pesticide type and subset all the strawberries treated not as organic.

Secondly, we design a variable called: toxicity level for human. To get the value for this variable, we first assigned the toxicity level for Carcinogen from 1 to 4, Hormone.Disruptor range from 1 and 2, Neurotoxins range from 1 and 2 and Developmental.or.Reproductive.Toxins range from 1 and 2, and add them to a new toxcity level range from 4 to 10, higher the toxcity level is, the more toxic the pesticide type is.

Thirdly, we filtered the merged dataset to only include chemicals that were listed in our toxicity dataset. Therefore we only were left with the observations for which we had relevant data regarding their toxicity to humans.

Clean for EDA: In order to do our EDA, we subset the data by each year and with a certain measurement. And we add a column store the frequency of chemical type for "Total Pesticide usage: 2016 vs. 2018 vs. 2019" and "Comparison of Pesdicide usage: 2016 vs. 2018 vs. 2019" these two plots. We also create three datasets storing the frequency of toxicity level to humans of each year to do the plot: "Pesticide Toxicitylevelhuman Frequency 2016 vs. 2018 vs. 2019".

The Original dataframe:

```
##   Program Year Period Geo.Level      State State.ANSI    Commodity Strawberries
## 1  CENSUS 2019   YEAR     STATE CALIFORNIA          6 STRAWBERRIES STRAWBERRIES
## 2  CENSUS 2019   YEAR     STATE CALIFORNIA          6 STRAWBERRIES STRAWBERRIES
## 3  CENSUS 2019   YEAR     STATE CALIFORNIA          6 STRAWBERRIES STRAWBERRIES
## 4  CENSUS 2019   YEAR     STATE CALIFORNIA          6 STRAWBERRIES STRAWBERRIES
## 5  CENSUS 2019   YEAR     STATE CALIFORNIA          6 STRAWBERRIES STRAWBERRIES
## 6  CENSUS 2019   YEAR     STATE CALIFORNIA          6 STRAWBERRIES STRAWBERRIES
##                             items                         discription
## 1  ORGANIC - OPERATIONS WITH SALES                                <NA>
## 2                 ORGANIC - SALES                       MEASURED IN $
## 3                 ORGANIC - SALES                     MEASURED IN CWT
## 4                         ORGANIC  FRESH MARKET - OPERATIONS WITH SALES
## 5                         ORGANIC                 FRESH MARKET - SALES
## 6                         ORGANIC                 FRESH MARKET - SALES
##             units         Domain Chemical                      Domain.Category
## 1            <NA> ORGANIC STATUS     <NA> ORGANIC STATUS: (NOP USDA CERTIFIED)
```

```
## 2            <NA> ORGANIC STATUS     <NA> ORGANIC STATUS: (NOP USDA CERTIFIED)
## 3            <NA> ORGANIC STATUS     <NA> ORGANIC STATUS: (NOP USDA CERTIFIED)
## 4            <NA> ORGANIC STATUS     <NA> ORGANIC STATUS: (NOP USDA CERTIFIED)
## 5    MEASURED IN $ ORGANIC STATUS     <NA> ORGANIC STATUS: (NOP USDA CERTIFIED)
## 6  MEASURED IN CWT ORGANIC STATUS     <NA> ORGANIC STATUS: (NOP USDA CERTIFIED)
##         Value CV....
## 1         174      8
## 2 300,277,717   33.1
## 3   1,384,016   30.4
## 4         170      8
## 5 275,716,713   35.5
## 6   1,177,214   33.7
```

After the merge and add toxicitylevel for human and bee:

```
##   Year      State   Chemical                        Chemicaltype Value
## 1 2019 CALIFORNIA  FUNGICIDE                        AZOXYSTROBIN 5,500
## 2 2019 CALIFORNIA  FUNGICIDE     BACILLUSAMYLOLIQUEFACIENSMBI600  (NA)
## 3 2019 CALIFORNIA  FUNGICIDE BACILLUSAMYLOLIQUEFACIENSSTRAIND747  (NA)
## 4 2019 CALIFORNIA  FUNGICIDE                     BACILLUSPUMILUS  (NA)
## 5 2019 CALIFORNIA  FUNGICIDE                   BACILLUSSUBT.GB03  (NA)
## 6 2019 CALIFORNIA  FUNGICIDE                    BACILLUSSUBTILIS  (NA)
##         measurement Carcinogen Hormone.Disruptor Neurotoxins
## 1  MEASURED IN LB      unknown           unknown     unknown
## 2  MEASURED IN LB      unknown           unknown     unknown
## 3  MEASURED IN LB      unknown           unknown     unknown
## 4  MEASURED IN LB      unknown           unknown     unknown
## 5  MEASURED IN LB      unknown           unknown     unknown
## 6  MEASURED IN LB      unknown           unknown     unknown
##   Developmental.or.Reproductive.Toxins toxicitylevelhuman Bee.Toxins
## 1                              unknown                  4    unknown
## 2                              unknown                  4    unknown
## 3                              unknown                  4    unknown
## 4                              unknown                  4    unknown
## 5                              unknown                  4    unknown
## 6                              unknown                  4    unknown
##   toxicitylevelbee
## 1                1
## 2                1
## 3                1
## 4                1
## 5                1
## 6                1
```

The dataset ready for EDA:

```
##   Year      State   Chemical Chemicaltype Value       measurement Carcinogen
## 1 2019 CALIFORNIA  FUNGICIDE AZOXYSTROBIN    NA  MEASURED IN LB      unknown
## 2 2019 CALIFORNIA  FUNGICIDE     BOSCALID    NA  MEASURED IN LB     possible
## 3 2019 CALIFORNIA  FUNGICIDE       CAPTAN    NA  MEASURED IN LB        known
## 4 2019 CALIFORNIA  FUNGICIDE   CYPRODINIL    NA  MEASURED IN LB      unknown
## 5 2019 CALIFORNIA  FUNGICIDE   FENHEXAMID    NA  MEASURED IN LB      unknown
## 6 2019 CALIFORNIA  FUNGICIDE  FLUDIOXONIL    NA  MEASURED IN LB      unknown
##   Hormone.Disruptor Neurotoxins Developmental.or.Reproductive.Toxins
## 1           unknown     unknown                              unknown
## 2           unknown     unknown                              unknown
```

```
## 3           unknown     unknown                              unknown
## 4           unknown     unknown                              unknown
## 5           unknown     unknown                              unknown
## 6           unknown     unknown                              unknown
##   toxicitylevelhuman Bee.Toxins toxicitylevelbee
## 1                  4    unknown                1
## 2                  5    unknown                1
## 3                  7    unknown                1
## 4                  4    unknown                1
## 5                  4    unknown                1
## 6                  4     slight                2
```

The dataset ready for maps:

```
##     State mean_toxicity      lat      long group order
## 1 ALABAMA            NA 30.38968 -87.46201     1     1
## 2 ALABAMA            NA 30.37249 -87.48493     1     2
## 3 ALABAMA            NA 30.37249 -87.52503     1     3
## 4 ALABAMA            NA 30.33239 -87.53076     1     4
## 5 ALABAMA            NA 30.32665 -87.57087     1     5
## 6 ALABAMA            NA 30.32665 -87.58806     1     6
```

The subset for Total Pesticide usage

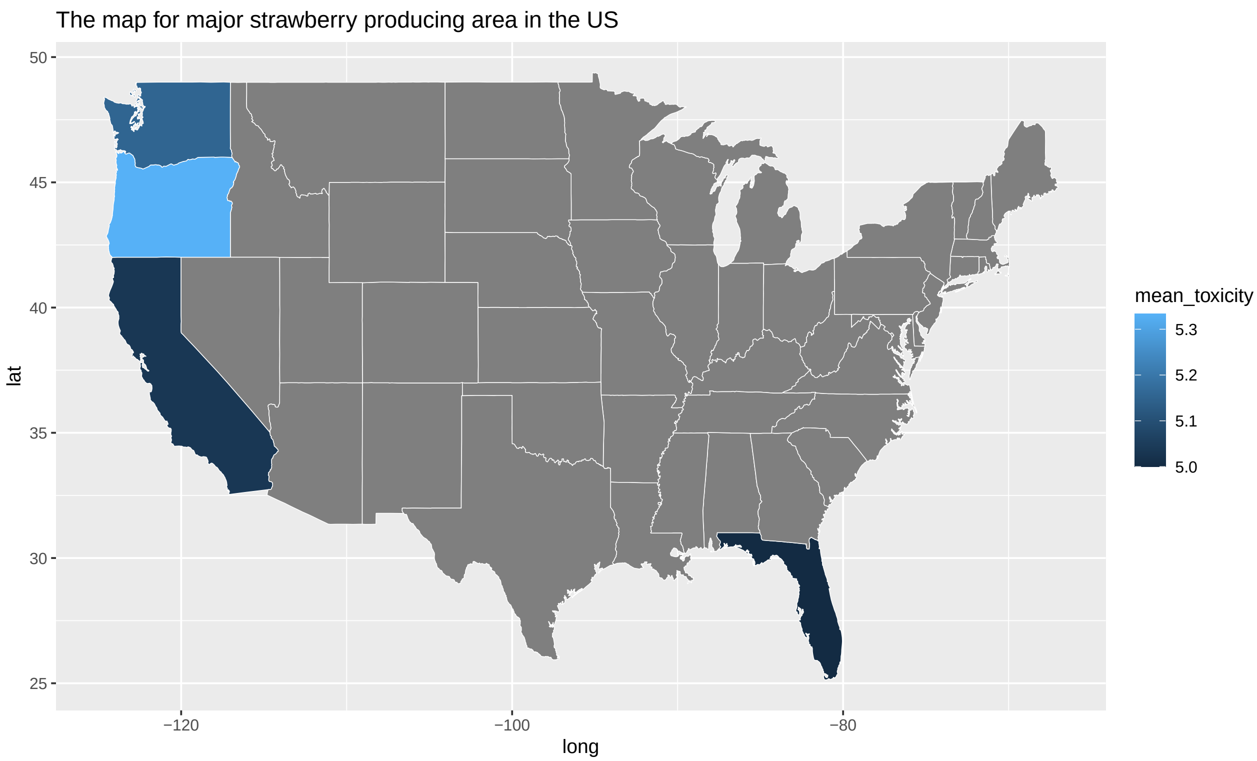
```
##   Chemicaltype Freq year
## 1  ACETAMIPRID    3 2016
## 2 AZOXYSTROBIN    2 2016
## 3   BIFENAZATE    2 2016
## 4   BIFENTHRIN    4 2016
## 5     BOSCALID    4 2016
## 6       CAPTAN    4 2016
```

The dataset for Pesticide Toxicitylevelhuman Frequency 2016:

```
##   Toxicityhumenlevel Freq FreqPerc
## 2                  5    4     0.13
## 3                  6   10     0.32
## 4                  7   13     0.42
## 5                  8    3     0.10
## 6                 10    1     0.03
```

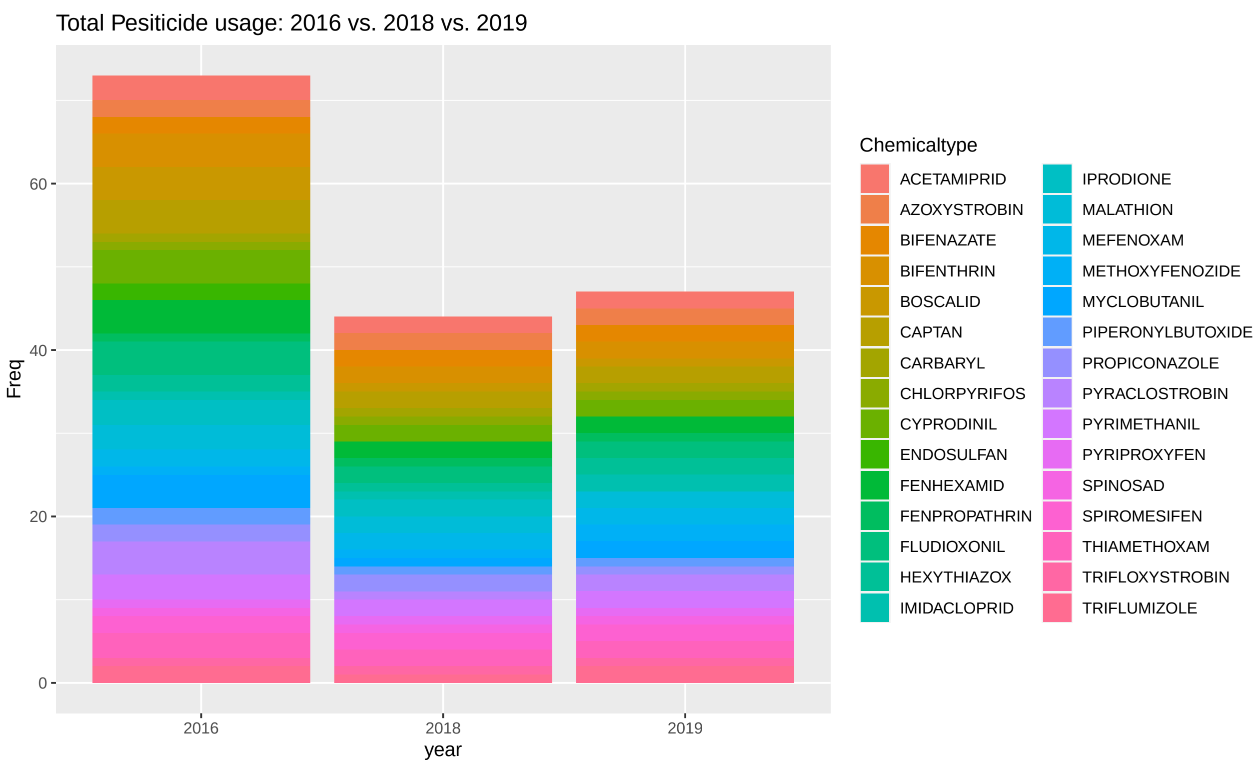
The map for major strawberry producing area in the US(including the mean toxicity level for human)

In our initial analysis of the data, we created a map to visualize the mean toxicity of all observations for each state. We took data across all years, calculated the means, and plotted the results on a map of the United States. We found that California and Florida on average had lower human toxicity levels in strawberry treatments than Washington and Oregon. It appears that Oregon has the highest mean toxicity levels out of the four states from which we have data.



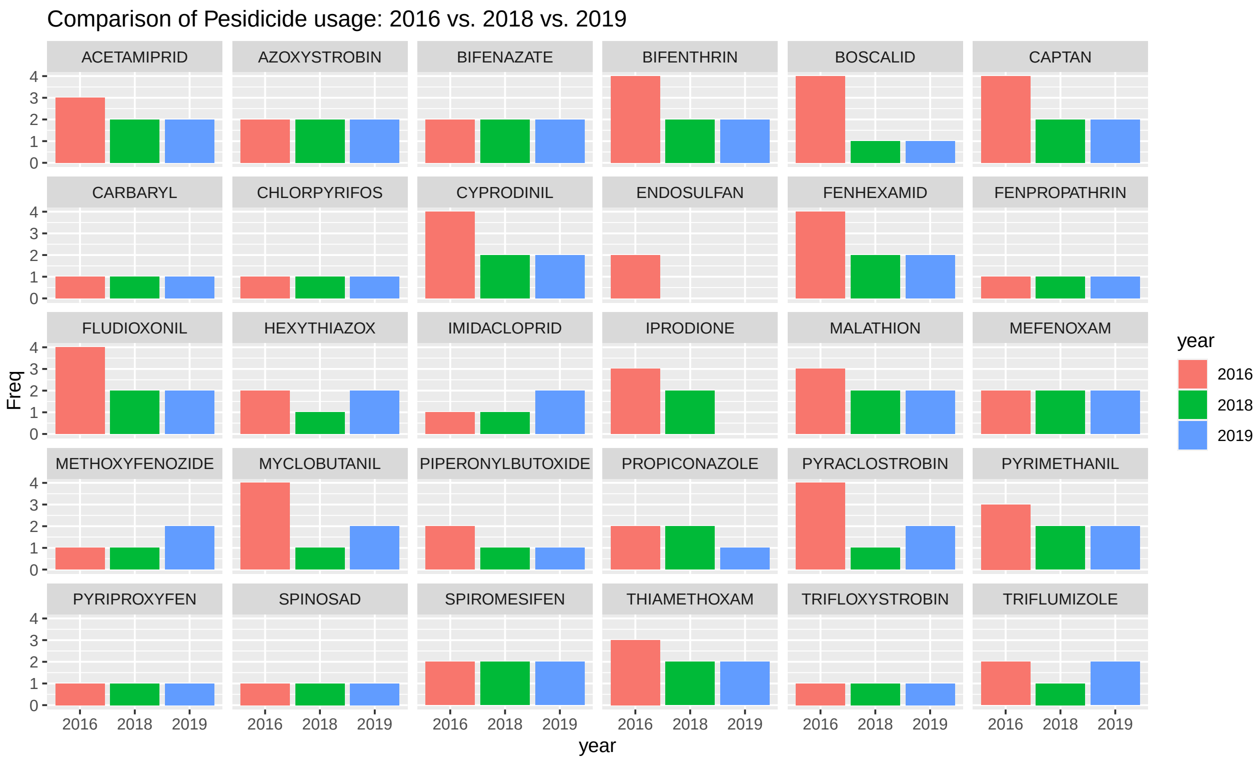
Total Pesticide usage: 2016 vs. 2018 vs. 2019

After our initial analysis of our data, we plot the total pesticide usage comparison for four states from 2016 to 2019 except 2017. From the bar chart, we can see that the pesticide usage is getting lower as time passes.



Comparison of Pesdicide usage: 2016 vs. 2018 vs. 2019

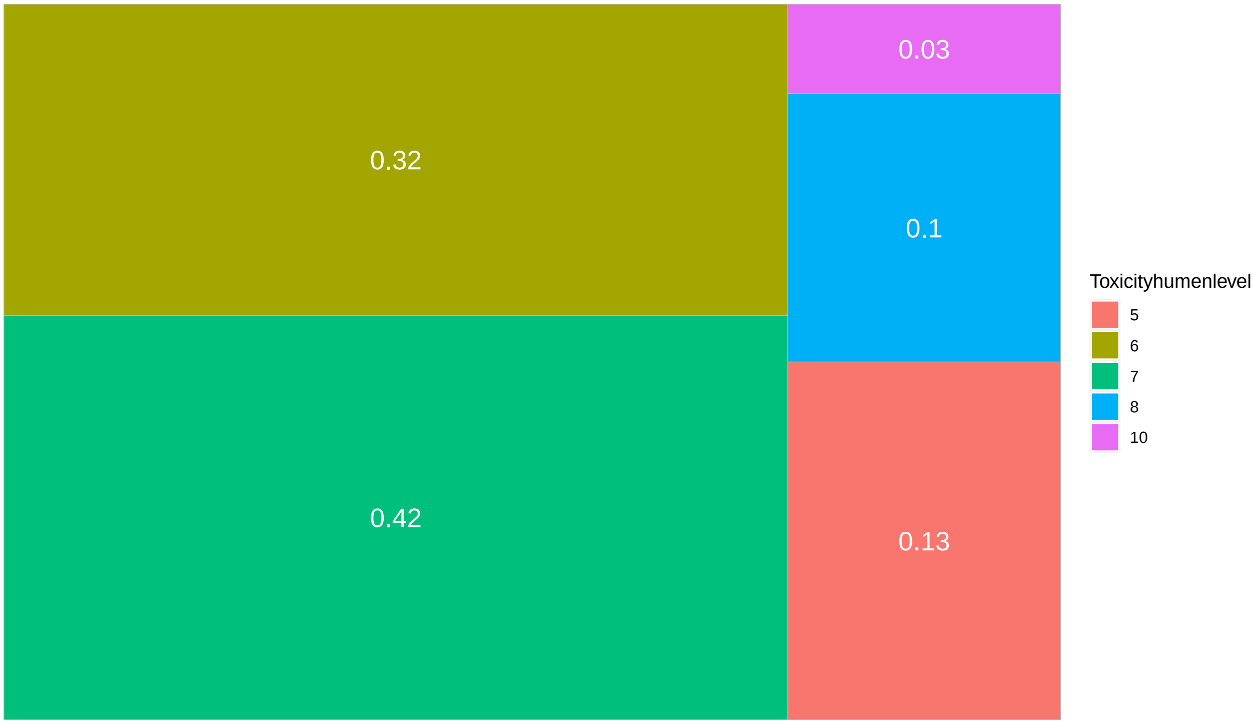
After we looked at the total usage of pesticides in our previous plot, we decided that we wanted to look deeper into the usage of pesticide over the years of data that we have access to. For this plot, we looked at each chemical type and plotted its frequency of usage across 3 years.



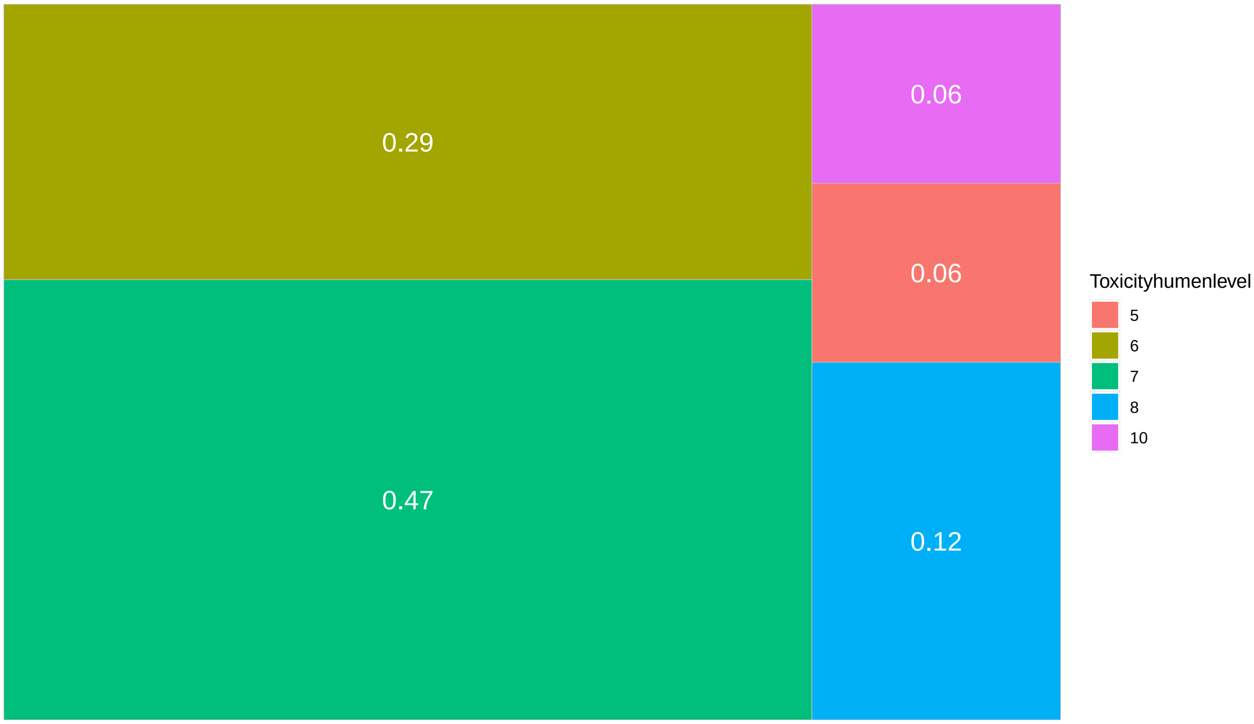
Pesticide Toxicitylevelhuman Frequency: 2016 vs. 2018 vs. 2019

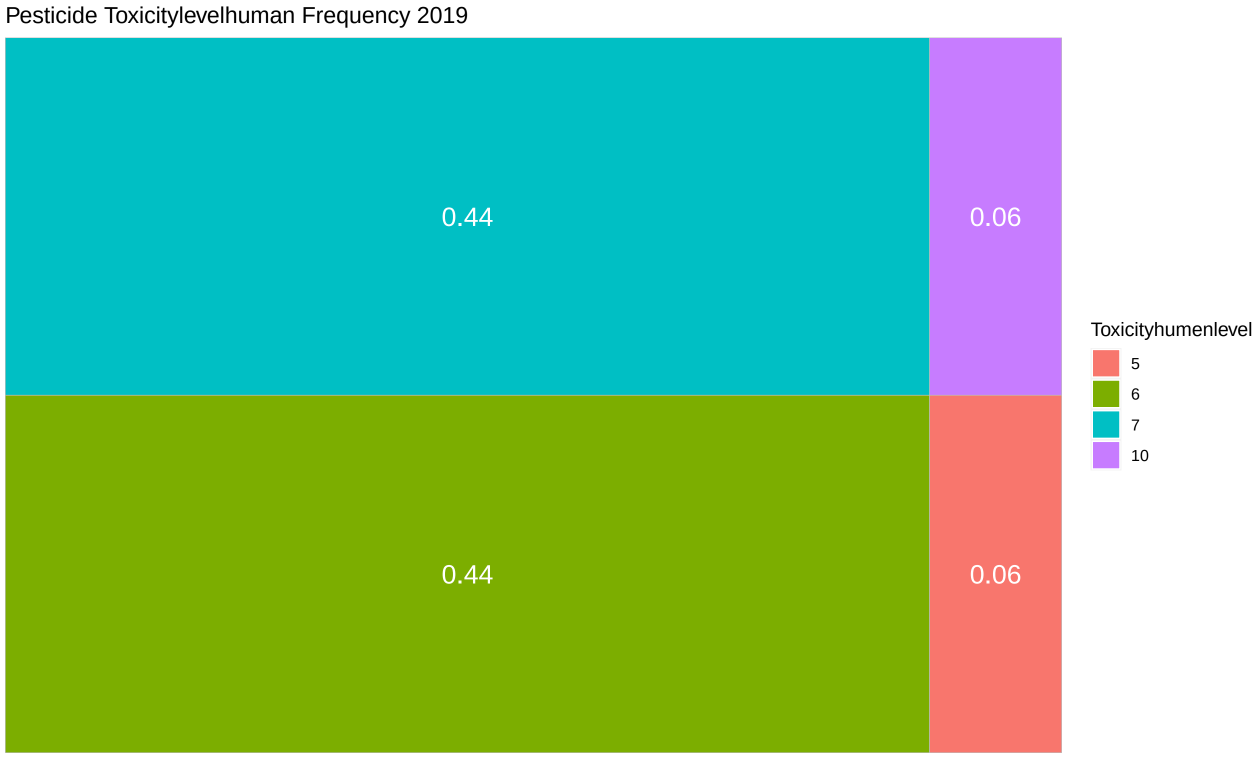
Now as we know the specific usage of each chemical type through 3 years. We want to dive deeper into the distribution of toxicity level that related to human for each year.

Pesticide Toxicitylevelhuman Frequency 2016



Pesticide Toxicitylevelhuman Frequency 2018





After we view all the three plots about the distribution, we conclude that althought the total amount of chemical be used across the 3 years decreased, the propotion of toxicitylevel remains relatively the same. We can say that even though the amount of chemicals they are treating the strawberries with is decreasing, those that they are using are still known to be toxic to humans.