Stawberry3

Ziheng Li

2024-10-31

```
library(magrittr)
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(stringr)
library(readr)
library(ggplot2)
library(tidyverse)
```

```
## -- Attaching core tidyverse packages ------------------------ tidyverse 2.0.0 --
## v forcats   1.0.0     v tibble    3.2.1
## v lubridate 1.9.3     v tidyr     1.3.1
## v purrr     1.0.2
```

```
## -- Conflicts ------------------------------------------ tidyverse_conflicts() --
## x tidyr::extract()   masks magrittr::extract()
## x dplyr::filter()    masks stats::filter()
## x dplyr::lag()       masks stats::lag()
## x purrr::set_names() masks magrittr::set_names()
## i Use the conflicted package (<http://conflicted.r-lib.org/>) to force all conflicts to become errors
```

```
library(PubChemR)
```

```
strawberry_chem <- read_csv("survey_d_chem.csv",show_col_types = FALSE)
```

```
## New names:
## * `` -> `...1`
```

```r
chemical_type <- strawberry_chem$type
head(chemical_type)
```

```
## [1] "FUNGICIDE"   "INSECTICIDE" "INSECTICIDE" "OTHER"       "FUNGICIDE"
## [6] "INSECTICIDE"
```

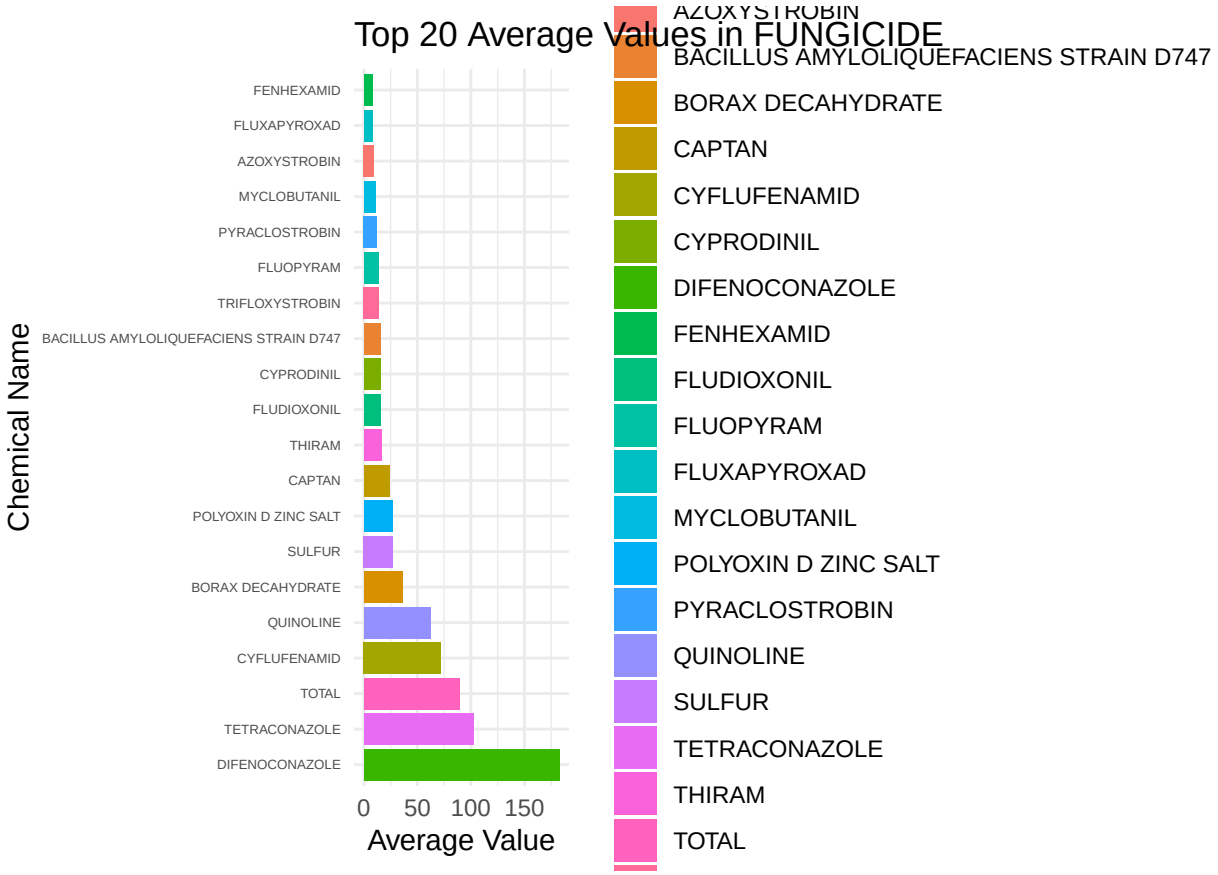
```r
california_data <- filter(strawberry_chem, State == "CALIFORNIA")
unique_types <- unique(california_data$type)
print(unique_types)
```

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

```r
strawberry_chem$Value <- as.numeric(as.character(strawberry_chem$Value))
```

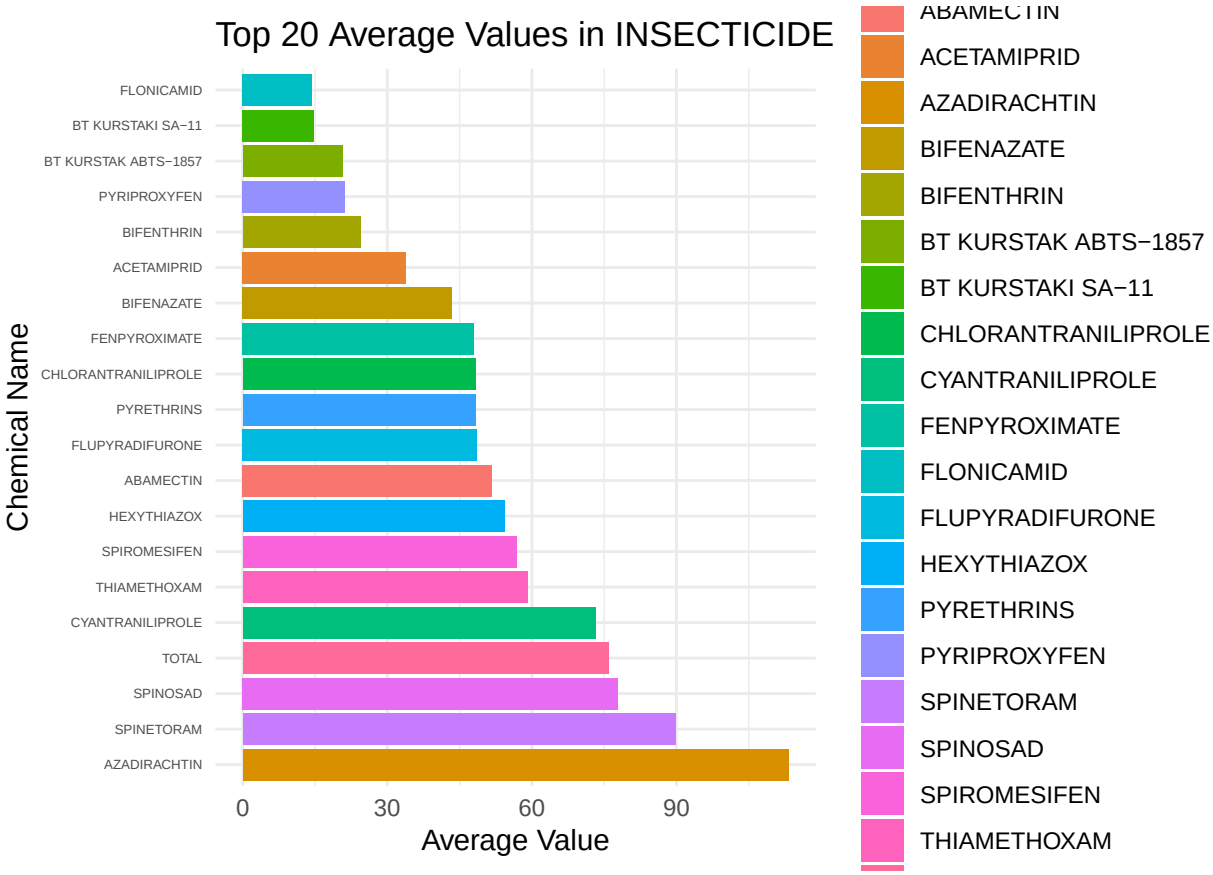
```
## Warning: NAs introduced by coercion
```

```r
# Define a function to filter data by type, compute mean values, and plot top 20 chemicals
plot_top_20_chemicals <- function(type_name) {
  # Filter data by type and calculate mean values
  type_data <- strawberry_chem %>%
    filter(type == type_name) %>%
    group_by(chem_name) %>%
    summarise(mean_value = mean(Value, na.rm = TRUE), .groups = 'drop') %>%
    arrange(desc(mean_value)) %>%
    slice_head(n = 20)

  # Plotting the top 20 chemicals
  plot <- ggplot(type_data, aes(x = reorder(chem_name, -mean_value), y = mean_value, fill = chem_name))
    geom_col() +
    coord_flip() +  # Flips the axes for better label readability
    labs(x = "Chemical Name", y = "Average Value", title = paste("Top 20 Average Values in", type_name))
    theme_minimal() +
    theme(axis.text.y = element_text(size = 5))  # Adjust text size for readability

  print(plot)
}

# Apply the function to each type
plot_top_20_chemicals("FUNGICIDE")
```



Fungicides: The chart shows a range of chemicals for which the mean values for phenyl ether metronidazole and tetraconazole are significantly higher, indicating higher utilization, efficacy or cost.

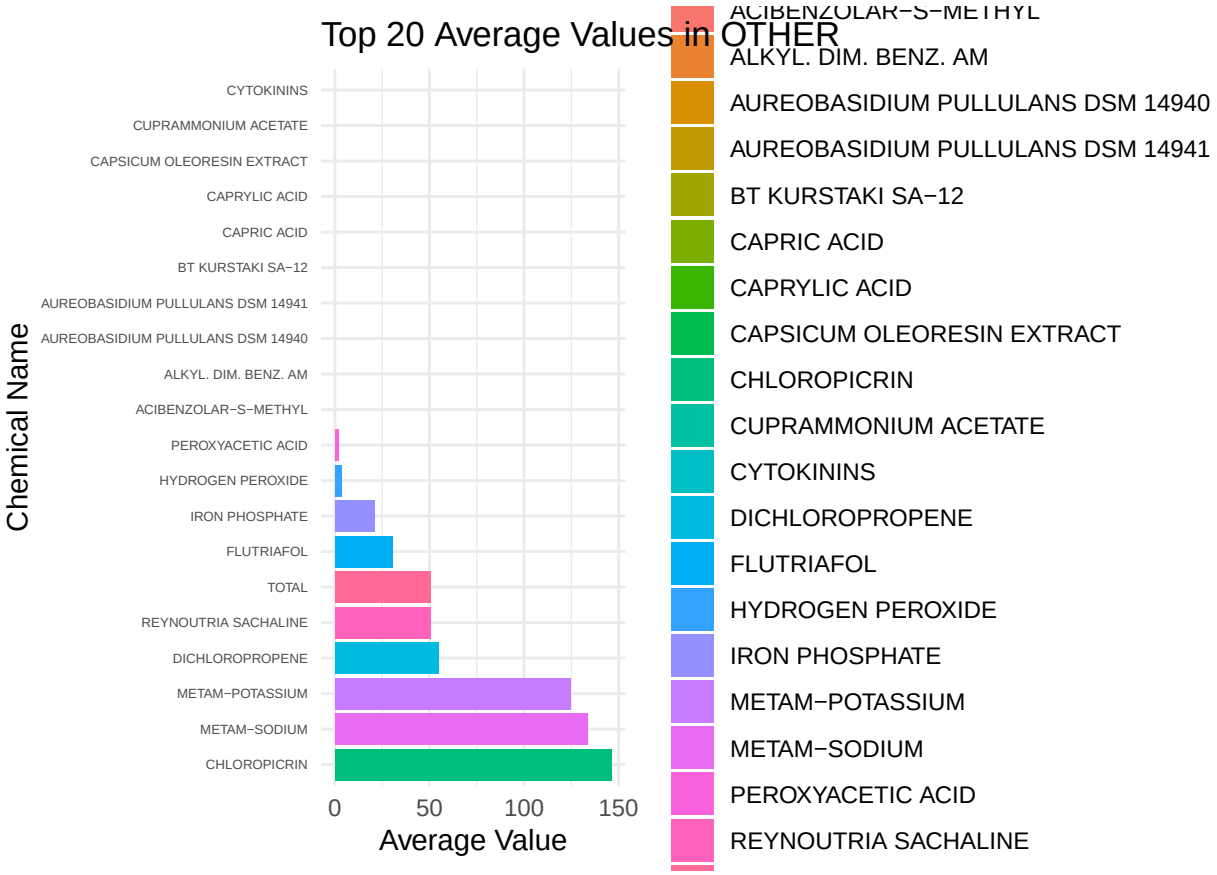
```
plot_top_20_chemicals("INSECTICIDE")
```



Insecticides: Azadirachtin and Spinetoram lead in average values, suggesting their major roles in insect control. Azadirachtin, derived from neem trees, also suggests a preference for bio-based insecticides.The wide range of chemicals indicates a variety of approaches to pest management, reflecting different application requirements and target species.

```
plot_top_20_chemicals("OTHER")
```

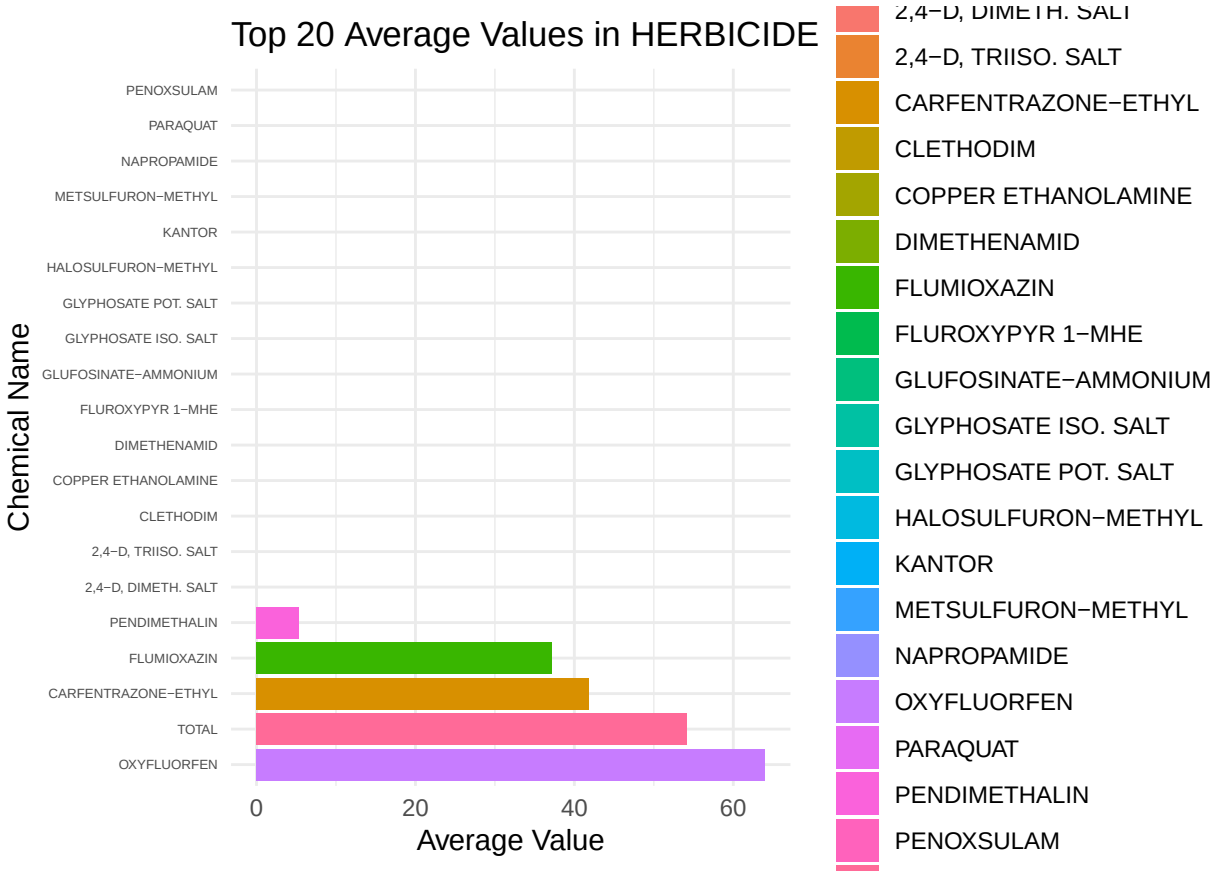
```
## Warning: Removed 10 rows containing missing values or values outside the scale range
## (`geom_col()`).
```



Other: The high average values of these chemicals could reflect their crucial role in preparing fields for planting by controlling soil-borne pests and diseases.

```
plot_top_20_chemicals("HERBICIDE")
```

```
## Warning: Removed 15 rows containing missing values or values outside the scale range
## (`geom_col()`).
```



Herbicides: The dominance of Oxyfluorfen and Paraquat in average values points to their widespread use, possibly due to their effectiveness in controlling a broad range of weeds.The presence of several different active ingredients suggests a strategic approach to herbicide resistance management.

```
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 Classification (UN
      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(hierarchy[["Node
    output_list[i]<-hierarchy[["Node"]][[i]][["Information"]][["Name"]]
    i<-i+1
  }
  return(output_list[!is.na(output_list)])
}
```

```r
chemical_vec<-c("azoxystrobin","flumioxazin")

result_f<-get_pug_rest(identifier = "flumioxazin", namespace = "name", domain = "compound",operation="cl

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

```
##  [1] "H360: May damage fertility or the unborn child [Danger Reproductive toxicity]"
##  [2] "H300: Health Hazards"
##  [3] "Hazard Statement Codes"
##  [4] "H361d: Suspected of damaging the unborn child [Warning Reproductive toxicity]"
##  [5] "H400: Very toxic to aquatic life [Warning Hazardous to the aquatic environment, acute hazard]"
##  [6] "H400: Environmental Hazards"
##  [7] "H410: Very toxic to aquatic life with long lasting effects [Warning Hazardous to the aquatic en
##  [8] "Hazardous to the aquatic environment, acute hazard"
##  [9] "Environmental Hazards"
## [10] "Hazard Classes"
## [11] "Hazardous to the aquatic environment, long-term hazard"
```

```r
result_d<-get_pug_rest(identifier = "hexythiazox", namespace = "name", domain = "compound",operation="cl

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

```
##  [1] "H320: Causes eye irritation [Warning Serious eye damage/eye irritation]"
##  [2] "H300: Health Hazards"
##  [3] "Hazard Statement Codes"
##  [4] "H400: Very toxic to aquatic life [Warning Hazardous to the aquatic environment, acute hazard]"
##  [5] "H400: Environmental Hazards"
##  [6] "H410: Very toxic to aquatic life with long lasting effects [Warning Hazardous to the aquatic en
##  [7] "Hazardous to the aquatic environment, acute hazard"
##  [8] "Environmental Hazards"
##  [9] "Hazard Classes"
## [10] "Hazardous to the aquatic environment, long-term hazard"
```