615 Assignment Strawberries 2

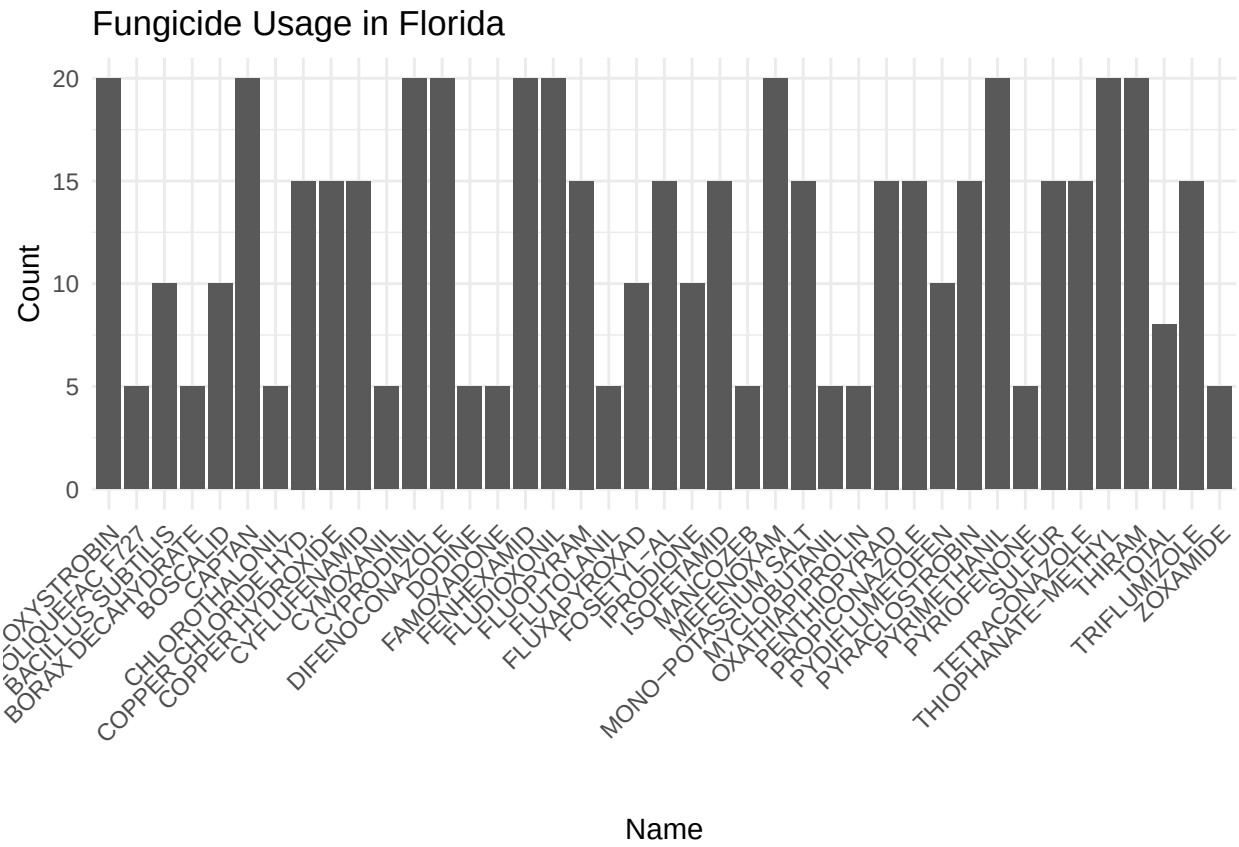
Haoran Cui

2024-10-21

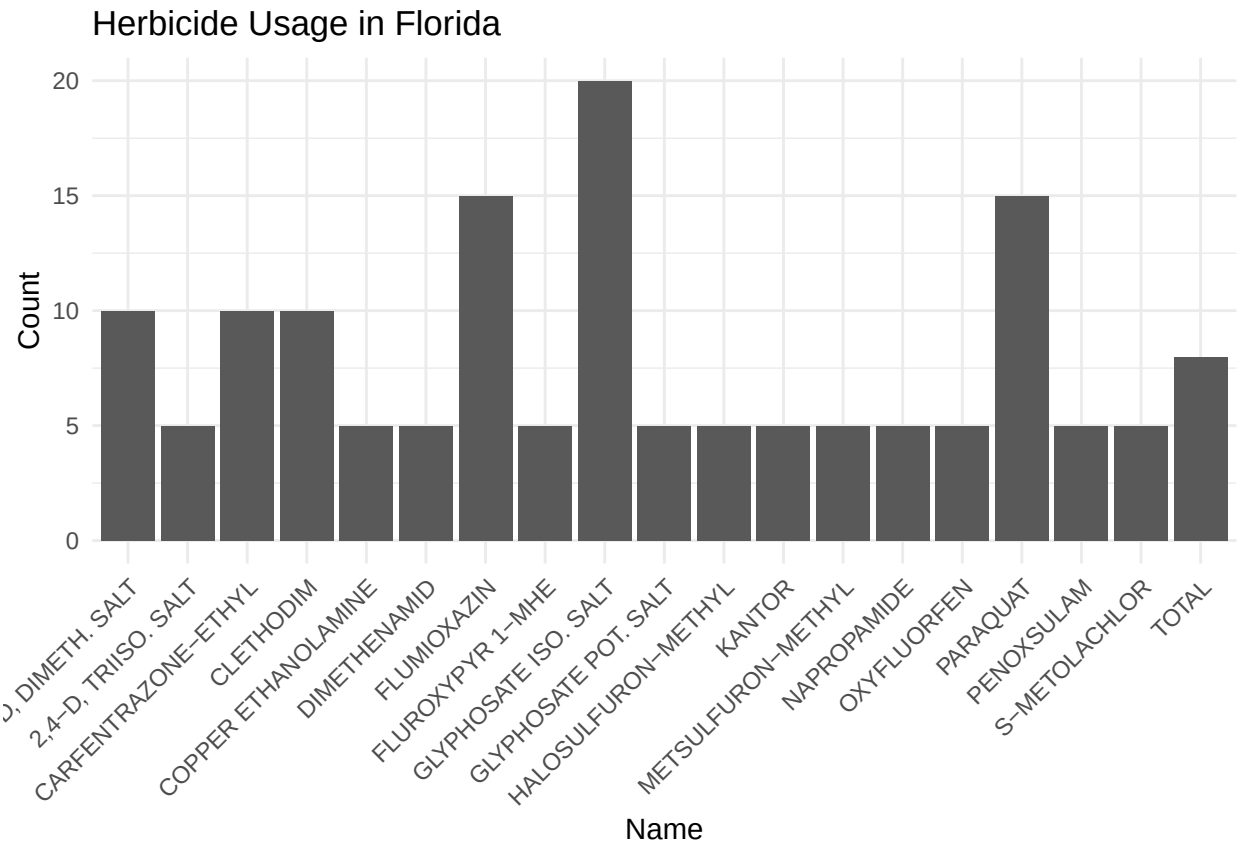
cleaned_strawberries.csv

```
strawberries <- read.csv("cleaned_strawberries.csv")

# Function to filter by category, state, and group by Name
filter_and_group <- function(data, category) {
  filtered_data <- subset(data, Category == category & State == "FLORIDA")
  grouped_data <- split(filtered_data, filtered_data$Name) # Group by Name
  return(grouped_data)
}

# Apply the function to each category
fungicide_florida_grouped <- filter_and_group(strawberries, "FUNGICIDE")
herbicide_florida_grouped <- filter_and_group(strawberries, "HERBICIDE")
insecticide_florida_grouped <- filter_and_group(strawberries, "INSECTICIDE")
other_florida_grouped <- filter_and_group(strawberries, "OTHER")
```

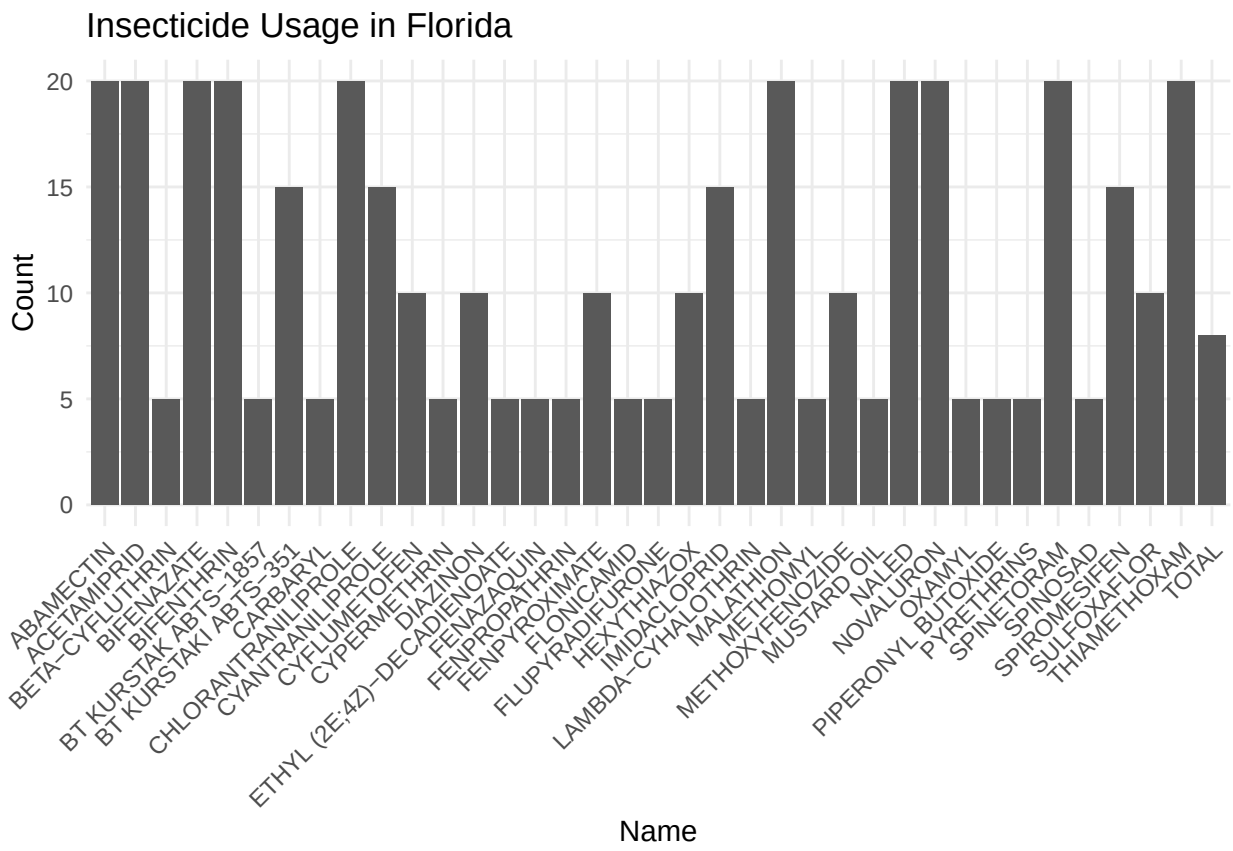
```
library(ggplot2)

# Function to create bar plots for each group
visualize_grouped_data <- function(grouped_data, title) {
  # Combine the data for easier plotting
  combined_data <- do.call(rbind, grouped_data)

  # Create a bar plot
  ggplot(combined_data, aes(x = Name)) +
    geom_bar() +
    labs(title = title, x = "Name", y = "Count") +
    theme_minimal() +
    theme(axis.text.x = element_text(angle = 45, hjust = 1))
}

# Visualize each category
visualize_grouped_data(fungicide_florida_grouped, "Fungicide Usage in Florida")
```



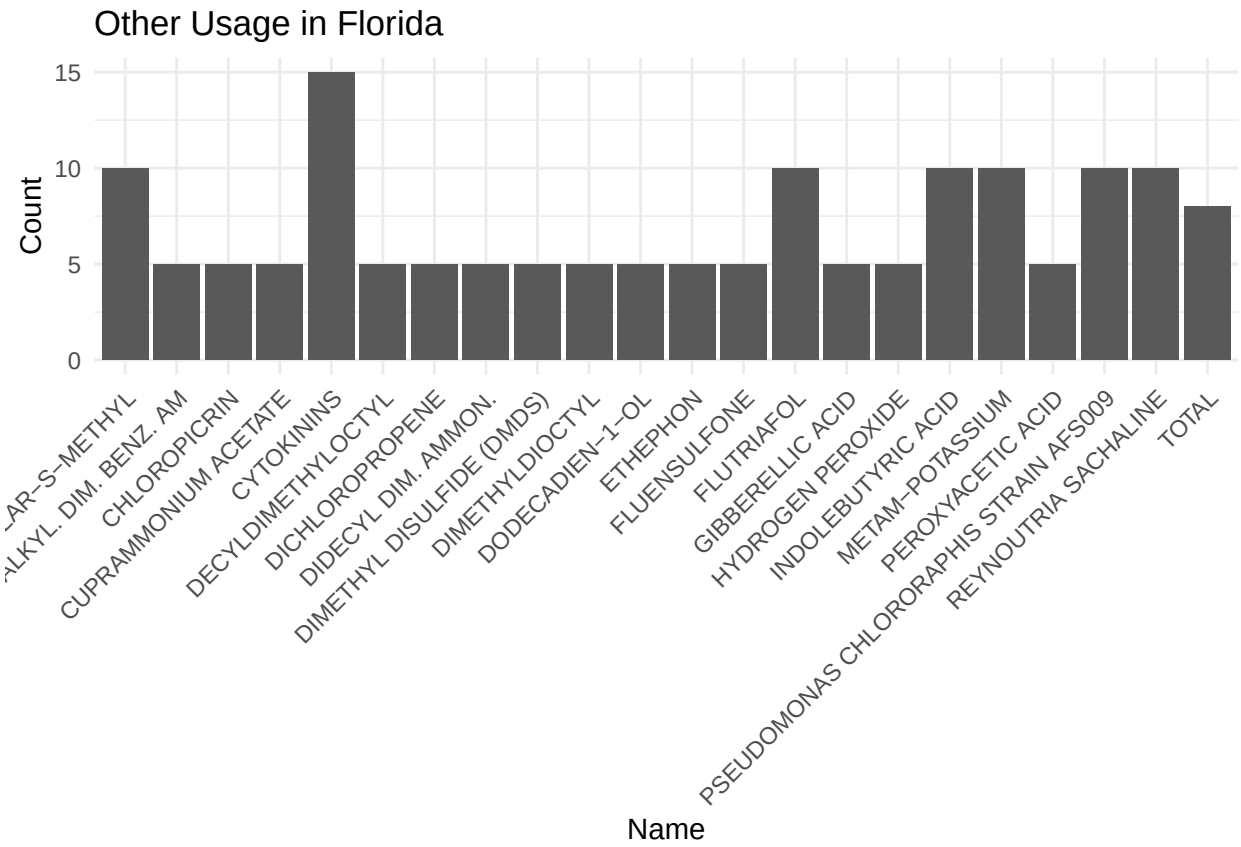

```
visualize_grouped_data(herbicide_florida_grouped, "Herbicide Usage in Florida")
```




```
visualize_grouped_data(insecticide_florida_grouped, "Insecticide Usage in Florida")
```




```
visualize_grouped_data(other_florida_grouped, "Other Usage in Florida")
```




```r
# Function to filter by category and state, then find the most and least frequent Name
find_most_least_frequent <- function(data, category) {
  # Filter data by category and state (Florida)
  filtered_data <- subset(data, Category == category & State == "FLORIDA")

  # Count occurrences of each Name
  name_counts <- table(filtered_data$Name)

  # Find the most frequent Name
  most_frequent <- names(name_counts[name_counts == max(name_counts)])

  # Find the least frequent Name
  least_frequent <- names(name_counts[name_counts == min(name_counts)])

  return(list("most_frequent" = most_frequent, "least_frequent" = least_frequent))
}

# Apply the function to each category
fungicide_florida_freq <- find_most_least_frequent(strawberries, "FUNGICIDE")
herbicide_florida_freq <- find_most_least_frequent(strawberries, "HERBICIDE")
insecticide_florida_freq <- find_most_least_frequent(strawberries, "INSECTICIDE")
other_florida_freq <- find_most_least_frequent(strawberries, "OTHER")

# Print the results for each category
print("Fungicide:")
```

```
## [1] "Fungicide:"
```

```
print(fungicide_florida_freq)
```

```
## $most_frequent
##  [1] "AZOXYSTROBIN"       "CAPTAN"             "CYPRODINIL"
##  [4] "DIFENOCONAZOLE"     "FENHEXAMID"         "FLUDIOXONIL"
##  [7] "MEFENOXAM"          "PYRIMETHANIL"       "THIOPHANATE-METHYL"
## [10] "THIRAM"
##
## $least_frequent
##  [1] "BACILLUS AMYLOLIQUEFAC F727" "BORAX DECAHYDRATE"
##  [3] "CHLOROTHALONIL"              "CYMOXANIL"
##  [5] "DODINE"                      "FAMOXADONE"
##  [7] "FLUTOLANIL"                  "MANCOZEB"
##  [9] "MYCLOBUTANIL"                "OXATHIAPIPROLIN"
## [11] "PYRIOFENONE"                 "ZOXAMIDE"
```

```
print("Herbicide:")
```

```
## [1] "Herbicide:"
```

```
print(herbicide_florida_freq)
```

```
## $most_frequent
## [1] "GLYPHOSATE ISO. SALT"
##
## $least_frequent
##  [1] "2,4-D, TRIISO. SALT"  "COPPER ETHANOLAMINE"  "DIMETHENAMID"
##  [4] "FLUROXYPYR 1-MHE"     "GLYPHOSATE POT. SALT" "HALOSULFURON-METHYL"
##  [7] "KANTOR"               "METSULFURON-METHYL"   "NAPROPAMIDE"
## [10] "OXYFLUORFEN"          "PENOXSULAM"           "S-METOLACHLOR"
```

```
print("Insecticide:")
```

```
## [1] "Insecticide:"
```

```
print(insecticide_florida_freq)
```

```
## $most_frequent
##  [1] "ABAMECTIN"           "ACETAMIPRID"         "BIFENAZATE"
##  [4] "BIFENTHRIN"          "CHLORANTRANILIPROLE" "MALATHION"
##  [7] "NALED"               "NOVALURON"           "SPINETORAM"
## [10] "THIAMETHOXAM"
##
## $least_frequent
##  [1] "BETA-CYFLUTHRIN"            "BT KURSTAK ABTS-1857"
##  [3] "CARBARYL"                   "CYPERMETHRIN"
##  [5] "ETHYL (2E;4Z)-DECADIENOATE" "FENAZAQUIN"
##  [7] "FENPROPATHRIN"              "FLONICAMID"
```

```
##  [9] "FLUPYRADIFURONE"            "LAMBDA-CYHALOTHRIN"
## [11] "METHOMYL"                   "MUSTARD OIL"
## [13] "OXAMYL"                     "PIPERONYL BUTOXIDE"
## [15] "PYRETHRINS"                 "SPINOSAD"
```

```
print("Other:")
```

```
## [1] "Other:"
```

```
print(other_florida_freq)
```

```
## $most_frequent
## [1] "CYTOKININS"
##
## $least_frequent
##  [1] "ALKYL. DIM. BENZ. AM"      "CHLOROPICRIN"
##  [3] "CUPRAMMONIUM ACETATE"      "DECYLDIMETHYLOCTYL"
##  [5] "DICHLOROPROPENE"           "DIDECYL DIM. AMMON."
##  [7] "DIMETHYL DISULFIDE (DMDS)" "DIMETHYLDIOCTYL"
##  [9] "DODECADIEN-1-OL"           "ETHEPHON"
## [11] "FLUENSULFONE"              "GIBBERELLIC ACID"
## [13] "HYDROGEN PEROXIDE"         "PEROXYACETIC ACID"
```

```
library(tidyverse)
```

```
## -- Attaching core tidyverse packages ------------------------ tidyverse 2.0.0 --
## v dplyr     1.1.4     v readr     2.1.5
## v forcats   1.0.0     v stringr   1.5.1
## v lubridate 1.9.3     v tibble    3.2.1
## v purrr     1.0.2     v tidyr     1.3.1
## -- Conflicts ---------------------------------------- tidyverse_conflicts() --
## x dplyr::filter() masks stats::filter()
## x dplyr::lag()    masks stats::lag()
## i Use the conflicted package (<http://conflicted.r-lib.org/>) to force all conflicts to become errors
```

```
library(PubChemR)

# Function to retrieve the GHS hazard statements with error handling
GHS_searcher <- function(result_json_object) {
  # Check if 'result', 'Hierarchies', and 'Hierarchy' exist and are not null
  if (!is.null(result_json_object[["result"]]) &&
      !is.null(result_json_object[["result"]][["Hierarchies"]]) &&
      !is.null(result_json_object[["result"]][["Hierarchies"]][["Hierarchy"]])) {

    hierarchy_list <- result_json_object[["result"]][["Hierarchies"]][["Hierarchy"]]

    # Loop through the hierarchy list and check for the GHS Classification
    for (i in seq_along(hierarchy_list)) {
      if (!is.null(hierarchy_list[[i]][["SourceName"]]) &&
          hierarchy_list[[i]][["SourceName"]] == "GHS Classification (UNECE)") {
        return(i)  # Return the index where GHS Classification is found
```

```
      }
    }
  }

  # If no GHS classification is found, return NA
  return(NA)
}

# Function to retrieve hazard details from the hierarchy with error handling
hazards_retriever <- function(index, result_json_object) {
  if (!is.na(index)) {
    hierarchy <- result_json_object[["result"]][["Hierarchies"]][["Hierarchy"]][[index]]
    if (!is.null(hierarchy[["Node"]])) {
      i <- 1
      output_list <- rep(NA, length(hierarchy[["Node"]]))

      while (i <= length(hierarchy[["Node"]]) &&
             !is.null(hierarchy[["Node"]][[i]][["Information"]][["Name"]]) &&
             str_detect(hierarchy[["Node"]][[i]][["Information"]][["Name"]], "H")) {
        output_list[i] <- hierarchy[["Node"]][[i]][["Information"]][["Name"]]
        i <- i + 1
      }

      return(output_list[!is.na(output_list)])  # Return non-NA hazard statements
    }
  }

  return(paste("No hazard information found"))
}

# Function to fetch and print hazard statements for a chemical
fetch_hazard_statements <- function(chemical_name) {
  result <- get_pug_rest(identifier = chemical_name, namespace = "name", domain = "compound", operation
  index <- GHS_searcher(result)
  if (!is.na(index)) {
    hazards <- hazards_retriever(index, result)
    return(hazards)
  } else {
    return(paste("No GHS classification found for", chemical_name))
  }
}

# Function to filter by category and state, then find the most and least frequent Name
find_most_least_frequent <- function(data, category) {
  # Filter data by category and state (Florida)
  filtered_data <- subset(data, Category == category & State == "FLORIDA")

  # Count occurrences of each Name
  name_counts <- table(filtered_data$Name)

  # Find the most frequent Name
  most_frequent <- names(name_counts[name_counts == max(name_counts)])
```

```r
  # Find the least frequent Name
  least_frequent <- names(name_counts[name_counts == min(name_counts)])

  return(list("most_frequent" = most_frequent, "least_frequent" = least_frequent))
}

# Assuming 'strawberries' data has already been loaded
# Retrieve the most and least frequent chemicals for each group
fungicide_florida_freq <- find_most_least_frequent(strawberries, "FUNGICIDE")
herbicide_florida_freq <- find_most_least_frequent(strawberries, "HERBICIDE")
insecticide_florida_freq <- find_most_least_frequent(strawberries, "INSECTICIDE")
other_florida_freq <- find_most_least_frequent(strawberries, "OTHER")

categories <- list(
  "Fungicide" = fungicide_florida_freq,
  "Herbicide" = herbicide_florida_freq,
  "Insecticide" = insecticide_florida_freq,
  "Other" = other_florida_freq
)

# Loop through each category to get hazard statements for the most and least frequent chemicals
for (category in names(categories)) {
  cat(paste("\nCategory:", category, "\n"))

  # Most frequent chemical
  most_frequent <- categories[[category]]$most_frequent
  cat(paste("Most frequent chemical:", most_frequent, "\n"))
  most_hazards <- fetch_hazard_statements(most_frequent)
  print(most_hazards)

  # Least frequent chemical
  least_frequent <- categories[[category]]$least_frequent
  cat(paste("Least frequent chemical:", least_frequent, "\n"))
  least_hazards <- fetch_hazard_statements(least_frequent)
  print(least_hazards)
}
```

```
##
## Category: Fungicide
## Most frequent chemical: AZOXYSTROBIN
##  Most frequent chemical: CAPTAN
##  Most frequent chemical: CYPRODINIL
##  Most frequent chemical: DIFENOCONAZOLE
##  Most frequent chemical: FENHEXAMID
##  Most frequent chemical: FLUDIOXONIL
##  Most frequent chemical: MEFENOXAM
##  Most frequent chemical: PYRIMETHANIL
##  Most frequent chemical: THIOPHANATE-METHYL
##  Most frequent chemical: THIRAM

## Request failed [404]. Retrying in 1.4 seconds...
## Request failed [404]. Retrying in 4.2 seconds...

##  [1] "No GHS classification found for AZOXYSTROBIN"
```

```
##  [2] "No GHS classification found for CAPTAN"
##  [3] "No GHS classification found for CYPRODINIL"
##  [4] "No GHS classification found for DIFENOCONAZOLE"
##  [5] "No GHS classification found for FENHEXAMID"
##  [6] "No GHS classification found for FLUDIOXONIL"
##  [7] "No GHS classification found for MEFENOXAM"
##  [8] "No GHS classification found for PYRIMETHANIL"
##  [9] "No GHS classification found for THIOPHANATE-METHYL"
## [10] "No GHS classification found for THIRAM"
## Least frequent chemical: BACILLUS AMYLOLIQUEFAC F727
##  Least frequent chemical: BORAX DECAHYDRATE
##  Least frequent chemical: CHLOROTHALONIL
##  Least frequent chemical: CYMOXANIL
##  Least frequent chemical: DODINE
##  Least frequent chemical: FAMOXADONE
##  Least frequent chemical: FLUTOLANIL
##  Least frequent chemical: MANCOZEB
##  Least frequent chemical: MYCLOBUTANIL
##  Least frequent chemical: OXATHIAPIPROLIN
##  Least frequent chemical: PYRIOFENONE
##  Least frequent chemical: ZOXAMIDE

## Request failed [404]. Retrying in 3.8 seconds...
## Request failed [404]. Retrying in 4.2 seconds...

##  [1] "No GHS classification found for BACILLUS AMYLOLIQUEFAC F727"
##  [2] "No GHS classification found for BORAX DECAHYDRATE"
##  [3] "No GHS classification found for CHLOROTHALONIL"
##  [4] "No GHS classification found for CYMOXANIL"
##  [5] "No GHS classification found for DODINE"
##  [6] "No GHS classification found for FAMOXADONE"
##  [7] "No GHS classification found for FLUTOLANIL"
##  [8] "No GHS classification found for MANCOZEB"
##  [9] "No GHS classification found for MYCLOBUTANIL"
## [10] "No GHS classification found for OXATHIAPIPROLIN"
## [11] "No GHS classification found for PYRIOFENONE"
## [12] "No GHS classification found for ZOXAMIDE"
##
## Category: Herbicide
## Most frequent chemical: GLYPHOSATE ISO. SALT

## Request failed [404]. Retrying in 3.9 seconds...
## Request failed [404]. Retrying in 3 seconds...

## [1] "No GHS classification found for GLYPHOSATE ISO. SALT"
## Least frequent chemical: 2,4-D, TRIISO. SALT
##  Least frequent chemical: COPPER ETHANOLAMINE
##  Least frequent chemical: DIMETHENAMID
##  Least frequent chemical: FLUROXYPYR 1-MHE
##  Least frequent chemical: GLYPHOSATE POT. SALT
##  Least frequent chemical: HALOSULFURON-METHYL
##  Least frequent chemical: KANTOR
##  Least frequent chemical: METSULFURON-METHYL
```

```
##  Least frequent chemical: NAPROPAMIDE
##  Least frequent chemical: OXYFLUORFEN
##  Least frequent chemical: PENOXSULAM
##  Least frequent chemical: S-METOLACHLOR

## Request failed [404]. Retrying in 2.3 seconds...
## Request failed [404]. Retrying in 7.1 seconds...

##  [1] "No GHS classification found for 2,4-D, TRIISO. SALT"
##  [2] "No GHS classification found for COPPER ETHANOLAMINE"
##  [3] "No GHS classification found for DIMETHENAMID"
##  [4] "No GHS classification found for FLUROXYPYR 1-MHE"
##  [5] "No GHS classification found for GLYPHOSATE POT. SALT"
##  [6] "No GHS classification found for HALOSULFURON-METHYL"
##  [7] "No GHS classification found for KANTOR"
##  [8] "No GHS classification found for METSULFURON-METHYL"
##  [9] "No GHS classification found for NAPROPAMIDE"
## [10] "No GHS classification found for OXYFLUORFEN"
## [11] "No GHS classification found for PENOXSULAM"
## [12] "No GHS classification found for S-METOLACHLOR"
##
## Category: Insecticide
## Most frequent chemical: ABAMECTIN
##  Most frequent chemical: ACETAMIPRID
##  Most frequent chemical: BIFENAZATE
##  Most frequent chemical: BIFENTHRIN
##  Most frequent chemical: CHLORANTRANILIPROLE
##  Most frequent chemical: MALATHION
##  Most frequent chemical: NALED
##  Most frequent chemical: NOVALURON
##  Most frequent chemical: SPINETORAM
##  Most frequent chemical: THIAMETHOXAM

## Request failed [404]. Retrying in 1.1 seconds...
## Request failed [404]. Retrying in 5.1 seconds...

##  [1] "No GHS classification found for ABAMECTIN"
##  [2] "No GHS classification found for ACETAMIPRID"
##  [3] "No GHS classification found for BIFENAZATE"
##  [4] "No GHS classification found for BIFENTHRIN"
##  [5] "No GHS classification found for CHLORANTRANILIPROLE"
##  [6] "No GHS classification found for MALATHION"
##  [7] "No GHS classification found for NALED"
##  [8] "No GHS classification found for NOVALURON"
##  [9] "No GHS classification found for SPINETORAM"
## [10] "No GHS classification found for THIAMETHOXAM"
## Least frequent chemical: BETA-CYFLUTHRIN
##  Least frequent chemical: BT KURSTAK ABTS-1857
##  Least frequent chemical: CARBARYL
##  Least frequent chemical: CYPERMETHRIN
##  Least frequent chemical: ETHYL (2E;4Z)-DECADIENOATE
##  Least frequent chemical: FENAZAQUIN
##  Least frequent chemical: FENPROPATHRIN
```

```
##  Least frequent chemical: FLONICAMID
##  Least frequent chemical: FLUPYRADIFURONE
##  Least frequent chemical: LAMBDA-CYHALOTHRIN
##  Least frequent chemical: METHOMYL
##  Least frequent chemical: MUSTARD OIL
##  Least frequent chemical: OXAMYL
##  Least frequent chemical: PIPERONYL BUTOXIDE
##  Least frequent chemical: PYRETHRINS
##  Least frequent chemical: SPINOSAD

## Request failed [404]. Retrying in 1.8 seconds...
## Request failed [404]. Retrying in 7.2 seconds...

##  [1] "No GHS classification found for BETA-CYFLUTHRIN"
##  [2] "No GHS classification found for BT KURSTAK ABTS-1857"
##  [3] "No GHS classification found for CARBARYL"
##  [4] "No GHS classification found for CYPERMETHRIN"
##  [5] "No GHS classification found for ETHYL (2E;4Z)-DECADIENOATE"
##  [6] "No GHS classification found for FENAZAQUIN"
##  [7] "No GHS classification found for FENPROPATHRIN"
##  [8] "No GHS classification found for FLONICAMID"
##  [9] "No GHS classification found for FLUPYRADIFURONE"
## [10] "No GHS classification found for LAMBDA-CYHALOTHRIN"
## [11] "No GHS classification found for METHOMYL"
## [12] "No GHS classification found for MUSTARD OIL"
## [13] "No GHS classification found for OXAMYL"
## [14] "No GHS classification found for PIPERONYL BUTOXIDE"
## [15] "No GHS classification found for PYRETHRINS"
## [16] "No GHS classification found for SPINOSAD"
##
## Category: Other
## Most frequent chemical: CYTOKININS

## Request failed [404]. Retrying in 1.7 seconds...
## Request failed [404]. Retrying in 2.8 seconds...

## [1] "No GHS classification found for CYTOKININS"
## Least frequent chemical: ALKYL. DIM. BENZ. AM
##  Least frequent chemical: CHLOROPICRIN
##  Least frequent chemical: CUPRAMMONIUM ACETATE
##  Least frequent chemical: DECYLDIMETHYLOCTYL
##  Least frequent chemical: DICHLOROPROPENE
##  Least frequent chemical: DIDECYL DIM. AMMON.
##  Least frequent chemical: DIMETHYL DISULFIDE (DMDS)
##  Least frequent chemical: DIMETHYLDIOCTYL
##  Least frequent chemical: DODECADIEN-1-OL
##  Least frequent chemical: ETHEPHON
##  Least frequent chemical: FLUENSULFONE
##  Least frequent chemical: GIBBERELLIC ACID
##  Least frequent chemical: HYDROGEN PEROXIDE
##  Least frequent chemical: PEROXYACETIC ACID

## Request failed [404]. Retrying in 2.2 seconds...
## Request failed [404]. Retrying in 1 seconds...
```

```
##  [1] "No GHS classification found for ALKYL. DIM. BENZ. AM"
##  [2] "No GHS classification found for CHLOROPICRIN"
##  [3] "No GHS classification found for CUPRAMMONIUM ACETATE"
##  [4] "No GHS classification found for DECYLDIMETHYLOCTYL"
##  [5] "No GHS classification found for DICHLOROPROPENE"
##  [6] "No GHS classification found for DIDECYL DIM. AMMON."
##  [7] "No GHS classification found for DIMETHYL DISULFIDE (DMDS)"
##  [8] "No GHS classification found for DIMETHYLDIOCTYL"
##  [9] "No GHS classification found for DODECADIEN-1-OL"
## [10] "No GHS classification found for ETHEPHON"
## [11] "No GHS classification found for FLUENSULFONE"
## [12] "No GHS classification found for GIBBERELLIC ACID"
## [13] "No GHS classification found for HYDROGEN PEROXIDE"
## [14] "No GHS classification found for PEROXYACETIC ACID"
```