# Assignment 3: Data Exploration

Robert Hill

Spring 2023

## OVERVIEW

This exercise accompanies the lessons in Environmental Data Analytics on Data Exploration.

## Directions

1. Rename this file `<FirstLast>_A03_DataExploration.Rmd` (replacing `<FirstLast>` with your first and last name).
2. Change "Student Name" on line 3 (above) with your name.
3. Work through the steps, **creating code and output** that fulfill each instruction.
4. Assign a useful **name to each code chunk** and include ample **comments** with your code.
5. Be sure to **answer the questions** in this assignment document.
6. When you have completed the assignment, **Knit** the text and code into a single PDF file.
7. After Knitting, submit the completed exercise (PDF file) to the dropbox in Sakai.

**TIP**: If your code extends past the page when knit, tidy your code by manually inserting line breaks.

**TIP**: If your code fails to knit, check that no `install.packages()` or `View()` commands exist in your code.

---

## Set up your R session

1. Check your working directory, load necessary packages (tidyverse, lubridate), and upload two datasets: the ECOTOX neonicotinoid dataset (ECOTOX_Neonicotinoids_Insects_raw.csv) and the Niwot Ridge NEON dataset for litter and woody debris (NEON_NIWO_Litter_massdata_2018-08_raw.csv). Name these datasets "Neonics" and "Litter", respectively. Be sure to include the subcommand to read strings in as factors.

```
getwd() #finds the working directory so I can  point to the file within the wd
```

```
## [1] "C:/Users/Robert Hill/Documents/EDA-Spring2023-RH"
```

```
library(tidyverse)
Neonics <- read.csv("./Data/Raw/ECOTOX_Neonicotinoids_Insects_raw.csv",stringsAsFactors = T)
Litter <- read.csv("./Data/Raw/NEON_NIWO_Litter_massdata_2018-08_raw.csv",stringsAsFactors = T)
#added stringasfactors so the strings are read as vectors rather than strings
```

## Learn about your system

2. The neonicotinoid dataset was collected from the Environmental Protection Agency's ECOTOX Knowledgebase, a database for ecotoxicology research. Neonicotinoids are a class of insecticides used widely in agriculture. The dataset that has been pulled includes all studies published on insects. Why might we be interested in the ecotoxicology of neonicotinoids on insects? Feel free to do a brief internet search if you feel you need more background information.

   Answer: neonicotinoids are ach receptor antagonist and leads to paralysis (and inevitably death) of insect. It isn't toxic to mammals, however it is very persistent in the environment. As such, we care about its ecotoxicology on insects so that we may make recommendations to minimize the amount used to reduce the amount remaining persistently in the environment.

3. The Niwot Ridge litter and woody debris dataset was collected from the National Ecological Observatory Network, which collectively includes 81 aquatic and terrestrial sites across 20 ecoclimatic domains. 32 of these sites sample forest litter and woody debris, and we will focus on the Niwot Ridge long-term ecological research (LTER) station in Colorado. Why might we be interested in studying litter and woody debris that falls to the ground in forests? Feel free to do a brief internet search if you feel you need more background information.

   Answer: the buildup of litter and woody debris provides an organic and nutrient rich environment for the soil ecosystem. Knowing this information is important in understanding soil nutrient cycling and ecosystem dynamics.

4. How is litter and woody debris sampled as part of the NEON network? Read the NEON_Litterfall_UserGuide.pdf document to learn more. List three pieces of salient information about the sampling methods here:

   Answer: 1. materials are collected in elevated 0.5m^2 PVC litter traps places 80cm above ground OR in 3x0.5 m rectangular ground traps (typically in pairs) 2. Total sampling locations are 20 plors with forested tower airsheds and 30 plots with low-statured vegatation over the tower airshed (total of 50) 3. One random or targeted sample pair is taken for every 400 m^2 plot area with miscelaneous restrictions on sampling location within a plot area biweekly for deciduous forests and every 1-2 months for evergreen sites

## Obtain basic summaries of your data (Neonics)

5. What are the dimensions of the dataset?

```
dim(Neonics) #provides number of row columns in data
```

```
## [1] 4623   30
```

6. Using the `summary` function on the "Effect" column, determine the most common effects that are studied. Why might these effects specifically be of interest?

```
summary(Neonics$Effect)
```

```
##     Accumulation        Avoidance         Behavior     Biochemistry
##               12              102              360               11
##          Cell(s)      Development        Enzyme(s) Feeding behavior
```

```
##               9            136             62            255
##        Genetics         Growth      Histology     Hormone(s)
##              82             38              5              1
##   Immunological   Intoxication     Morphology      Mortality
##              16             12             22           1493
##      Physiology     Population   Reproduction
##               7           1803            197
```

Answer: The most common effects are mortality and population. These are of interests because it is linked efficacy of the insecticide for the purpose for which it is used.

7. Using the `summary` function, determine the six most commonly studied species in the dataset (common name). What do these species have in common, and why might they be of interest over other insects? Feel free to do a brief internet search for more information if needed.[TIP: The `sort()` command can sort the output of the summary command…]

```
summary(Neonics$Species.Common.Name)
```

```
##                          Honey Bee                     Parasitic Wasp
##                                667                                285
##              Buff Tailed Bumblebee                Carniolan Honey Bee
##                                183                                152
##                         Bumble Bee                  Italian Honeybee
##                                140                                113
##                    Japanese Beetle                  Asian Lady Beetle
##                                 94                                 76
##                     Euonymus Scale                           Wireworm
##                                 75                                 69
##                  European Dark Bee                  Minute Pirate Bug
##                                 66                                 62
##                Asian Citrus Psyllid                     Parastic Wasp
##                                 60                                 58
##             Colorado Potato Beetle                    Parasitoid Wasp
##                                 57                                 51
##                Erythrina Gall Wasp                       Beetle Order
##                                 49                                 47
##        Snout Beetle Family, Weevil            Sevenspotted Lady Beetle
##                                 47                                 46
##                     True Bug Order              Buff-tailed Bumblebee
##                                 45                                 39
##                       Aphid Family                     Cabbage Looper
##                                 38                                 38
##                Sweetpotato Whitefly                      Braconid Wasp
##                                 37                                 33
##                       Cotton Aphid                     Predatory Mite
##                                 33                                 33
##             Ladybird Beetle Family                         Parasitoid
##                                 30                                 30
##                      Scarab Beetle                      Spring Tiphia
##                                 29                                 29
##                        Thrip Order                Ground Beetle Family
##                                 29                                 27
##                 Rove Beetle Family                      Tobacco Aphid
```

```
##                                   27                              27
##                         Chalcid Wasp         Convergent Lady Beetle
##                                   25                              25
##                        Stingless Bee              Spider/Mite Class
##                                   25                              24
##                  Tobacco Flea Beetle               Citrus Leafminer
##                                   24                              23
##                      Ladybird Beetle                      Mason Bee
##                                   23                              22
##                             Mosquito                  Argentine Ant
##                                   22                              21
##                               Beetle    Flatheaded Appletree Borer
##                                   21                              20
##                   Horned Oak Gall Wasp             Leaf Beetle Family
##                                   20                              20
##                    Potato Leafhopper    Tooth-necked Fungus Beetle
##                                   20                              20
##                          Codling Moth     Black-spotted Lady Beetle
##                                   19                              18
##                         Calico Scale            Fairyfly Parasitoid
##                                   18                              18
##                          Lady Beetle        Minute Parasitic Wasps
##                                   18                              18
##                            Mirid Bug               Mulberry Pyralid
##                                   18                              18
##                             Silkworm                 Vedalia Beetle
##                                   18                              18
##                Araneoid Spider Order                      Bee Order
##                                   17                              17
##                       Egg Parasitoid                   Insect Class
##                                   17                              17
##             Moth And Butterfly Order  Oystershell Scale Parasitoid
##                                   17                              17
## Hemlock Woolly Adelgid Lady Beetle        Hemlock Wooly Adelgid
##                                   16                              16
##                                 Mite                    Onion Thrip
##                                   16                              16
##                Western Flower Thrips                   Corn Earworm
##                                   15                              14
##                    Green Peach Aphid                      House Fly
##                                   14                              14
##                            Ox Beetle             Red Scale Parasite
##                                   14                              14
##                   Spined Soldier Bug         Armoured Scale Family
##                                   14                              13
##                     Diamondback Moth                  Eulophid Wasp
##                                   13                              13
##                    Monarch Butterfly                  Predatory Bug
##                                   13                              13
##                Yellow Fever Mosquito            Braconid Parasitoid
##                                   13                              12
##                         Common Thrip  Eastern Subterranean Termite
##                                   12                              12
##                               Jassid                     Mite Order
```

```
##                     12                          12
##                Pea Aphid             Pond Wolf Spider
##                     12                          12
##   Spotless Ladybird Beetle     Glasshouse Potato Wasp
##                     11                          10
##                 Lacewing      Southern House Mosquito
##                     10                          10
##    Two Spotted Lady Beetle                 Ant Family
##                     10                           9
##             Apple Maggot                     (Other)
##                      9                         670
```

Answer: This is interesting (and intuitive) because neonicotinoids are responsible for colony callapse disorder. Most of the top species are in fact bees.

8. Concentrations are always a numeric value. What is the class of `Conc.1..Author.` column in the dataset, and why is it not numeric?

```
class(Neonics$Conc.1..Author.) #this determines the data class of this column
```

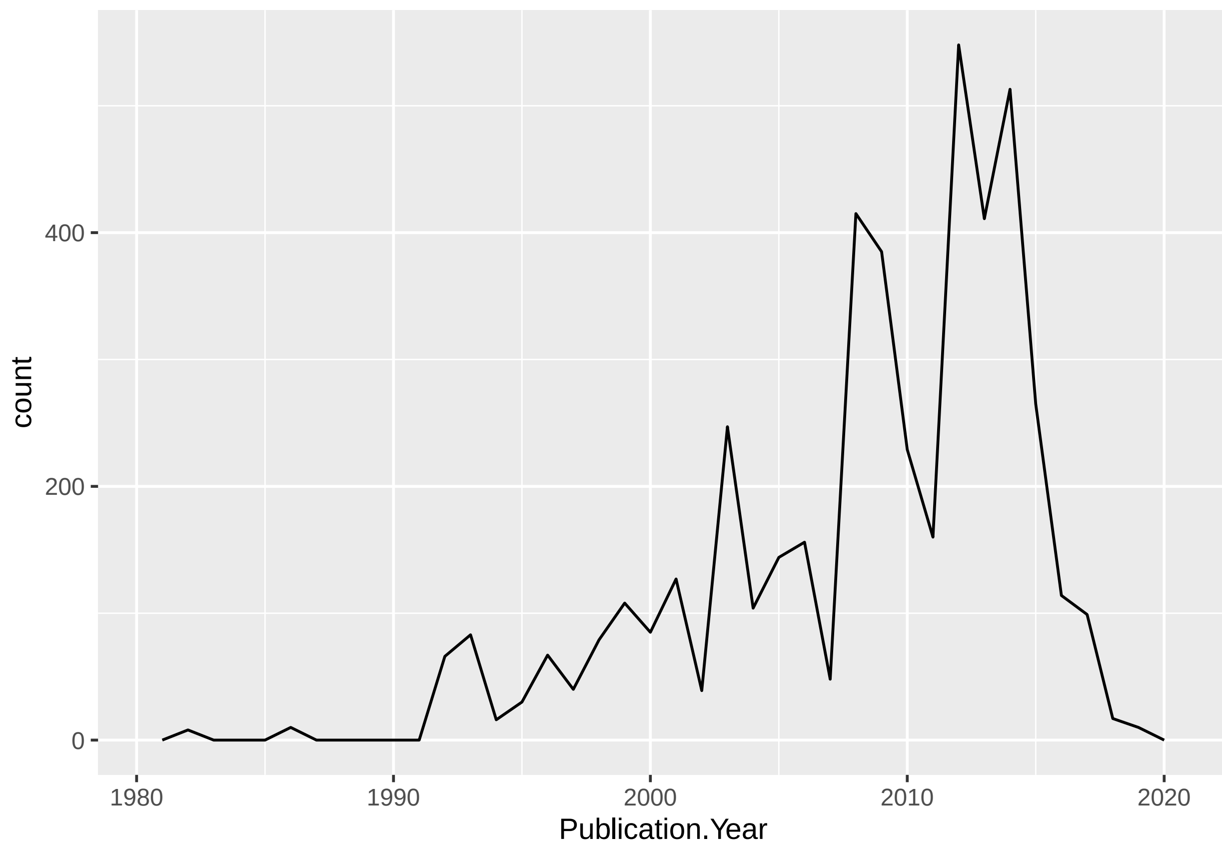
```
## [1] "factor"
```

Answer: the class is not numeric because there are non-numeric symbols in the cells; such as <, /, and ~.

## Explore your data graphically (Neonics)

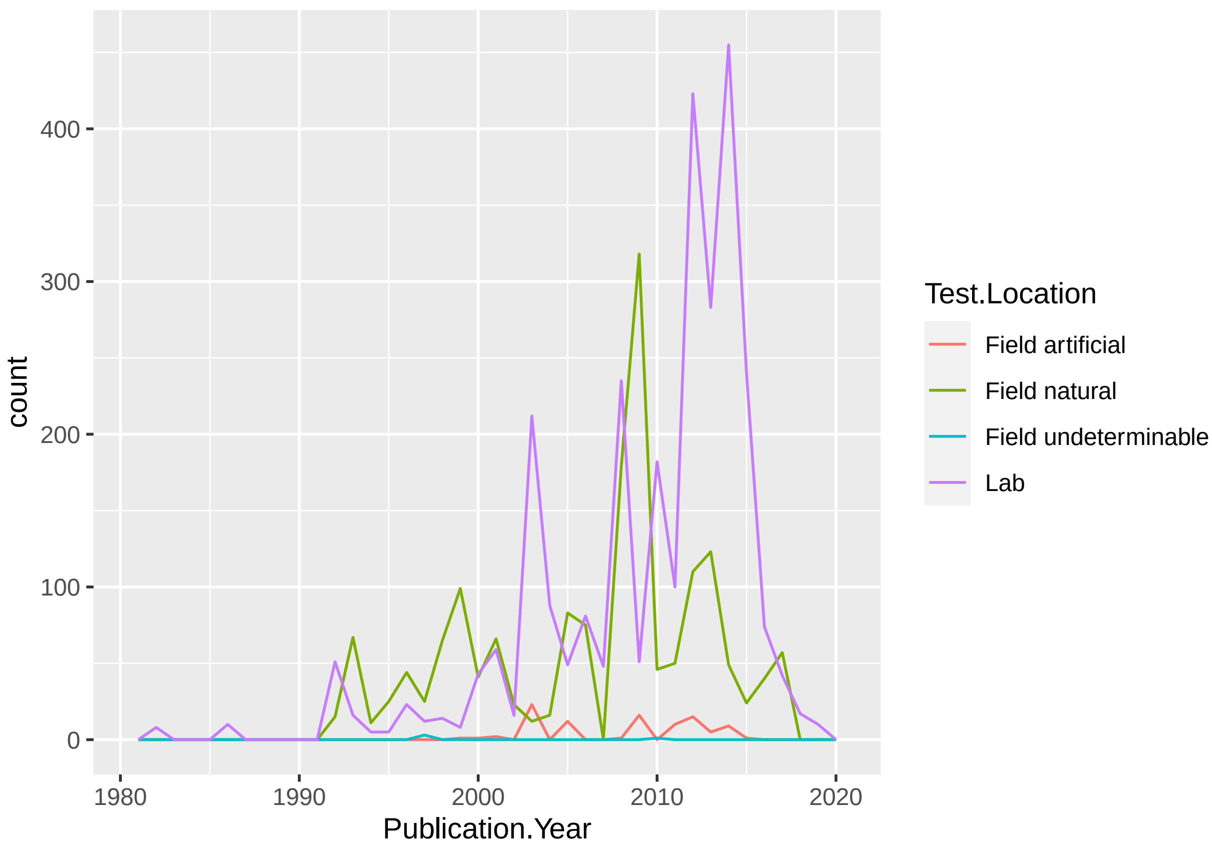
9. Using `geom_freqpoly`, generate a plot of the number of studies conducted by publication year.

```
#create a frequency line graph of studies over time
ggplot(Neonics) +
  geom_freqpoly(aes(x = Publication.Year), bins=38)
```




10. Reproduce the same graph but now add a color aesthetic so that different Test.Location are displayed as different colors.

```
#create a frequency line graph of studies over time, separated based on testing environment
ggplot(Neonics) +
  geom_freqpoly(aes(x = Publication.Year, color = Test.Location), bins=38)
```



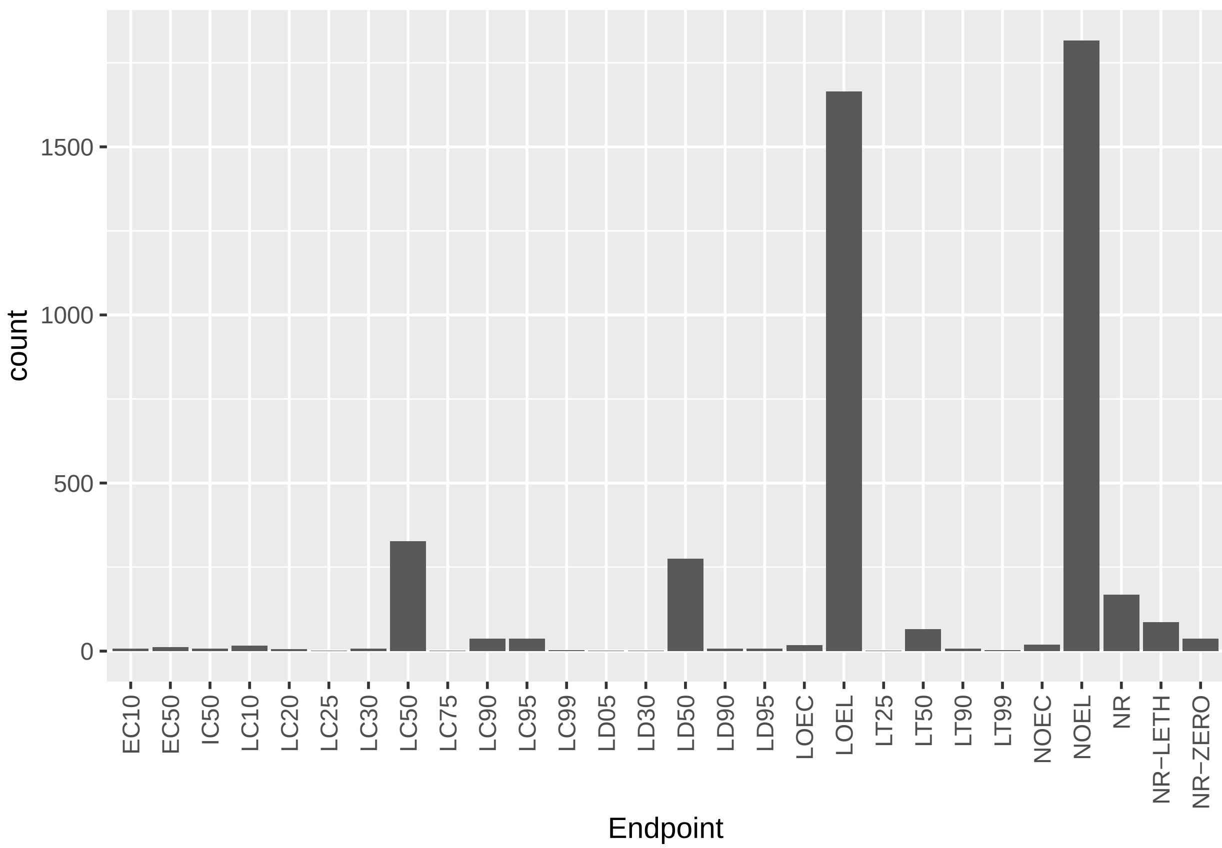

Interpret this graph. What are the most common test locations, and do they differ over time?

> Answer: This graph portrays the frequency of studies conducted on neonicotinoids categorized by where the study occured (in different types of field settings or in the lab). In the 90s, the most prominent study location was in the natural field environment; however as we move into the 21st century, studies became for frequent in laboratory settings. The frequency of natural field studies didn't change too drastically, but a large addition of lab studies began to emerge into the 21st century.

11. Create a bar graph of Endpoint counts. What are the two most common end points, and how are they defined? Consult the ECOTOX_CodeAppendix for more information.

[**TIP**: Add `theme(axis.text.x = element_text(angle = 90, vjust = 0.5, hjust=1))` to the end of your plot command to rotate and align the X-axis labels…]

```
#create a bar frequency plot of common measurable endpoints with default aesthetics [aes()]
#and adjusted the text with theme() so that is is legible and not overlapping
ggplot(Neonics) +
  geom_bar(aes(x = Endpoint)) +
  theme(axis.text.x = element_text(angle = 90, vjust = 0.5, hjust=1))
```




> Answer: the most common endpoints are the LOEL (lowest observable effect level) which is the lowest concentration for which there was a statistically significant response and NOEL (no observable effect level) which is the highest concentration for which there is no statistically significant response.

## Explore your data (Litter)

12. Determine the class of collectDate. Is it a date? If not, change to a date and confirm the new class of the variable. Using the `unique` function, determine which dates litter was sampled in August 2018.

```
class(Litter$collectDate) #this determines the class of the collectDate column
```

```
## [1] "factor"
```

```
library(lubridate)
```

```
##
## Attaching package: 'lubridate'

## The following objects are masked from 'package:base':
##
##     date, intersect, setdiff, union
```

```
CollectDate <-ymd(Litter$collectDate) #identify collectDate as a Date class
class(CollectDate) #confirms CollectDate is a date class
```

```
## [1] "Date"
```

```
unique(CollectDate) #this returns a list of unique values during which litter was sampled
```

```
## [1] "2018-08-02" "2018-08-30"
```

13. Using the `unique` function, determine how many plots were sampled at Niwot Ridge. How is the information obtained from `unique` different from that obtained from `summary`?

```
unique(Litter$plotID) #number of unique plot IDs (aka # of plots sampled)
```

```
##  [1] NIWO_061 NIWO_064 NIWO_067 NIWO_040 NIWO_041 NIWO_063 NIWO_047 NIWO_051
##  [9] NIWO_058 NIWO_046 NIWO_062 NIWO_057
## 12 Levels: NIWO_040 NIWO_041 NIWO_046 NIWO_047 NIWO_051 NIWO_057 ... NIWO_067
```

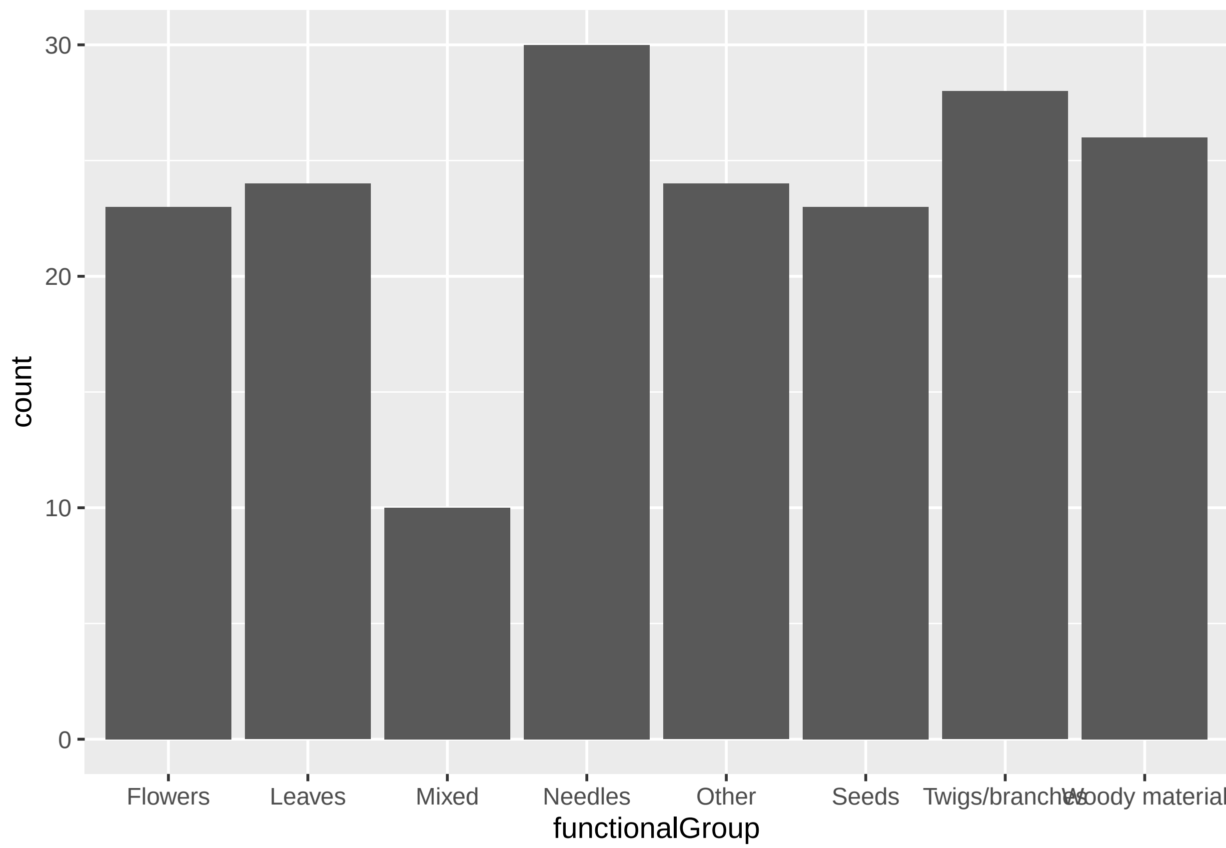
```
summary(Litter$plotID) #ran summary function to compare output
```

```
## NIWO_040 NIWO_041 NIWO_046 NIWO_047 NIWO_051 NIWO_057 NIWO_058 NIWO_061
##       20       19       18       15       14        8       16       17
## NIWO_062 NIWO_063 NIWO_064 NIWO_067
##       14       14       16       17
```

> Answer: 'unique' provides information regarding the number of unique values (12) while 'summary' only lists the unique values, requiring us to count how many unique values ourselves.

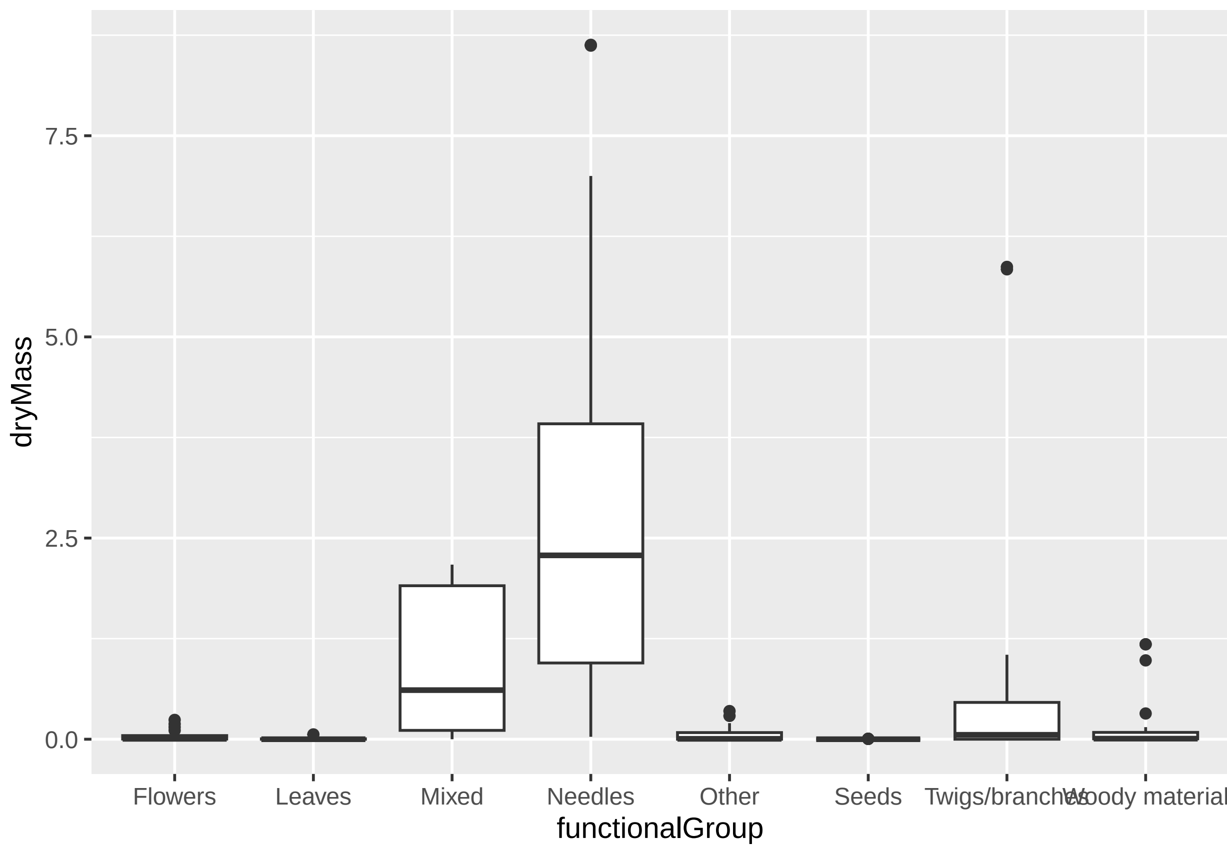
14. Create a bar graph of functionalGroup counts. This shows you what type of litter is collected at the Niwot Ridge sites. Notice that litter types are fairly equally distributed across the Niwot Ridge sites.

```
ggplot(Litter) +
  geom_bar(aes(functionalGroup))
```

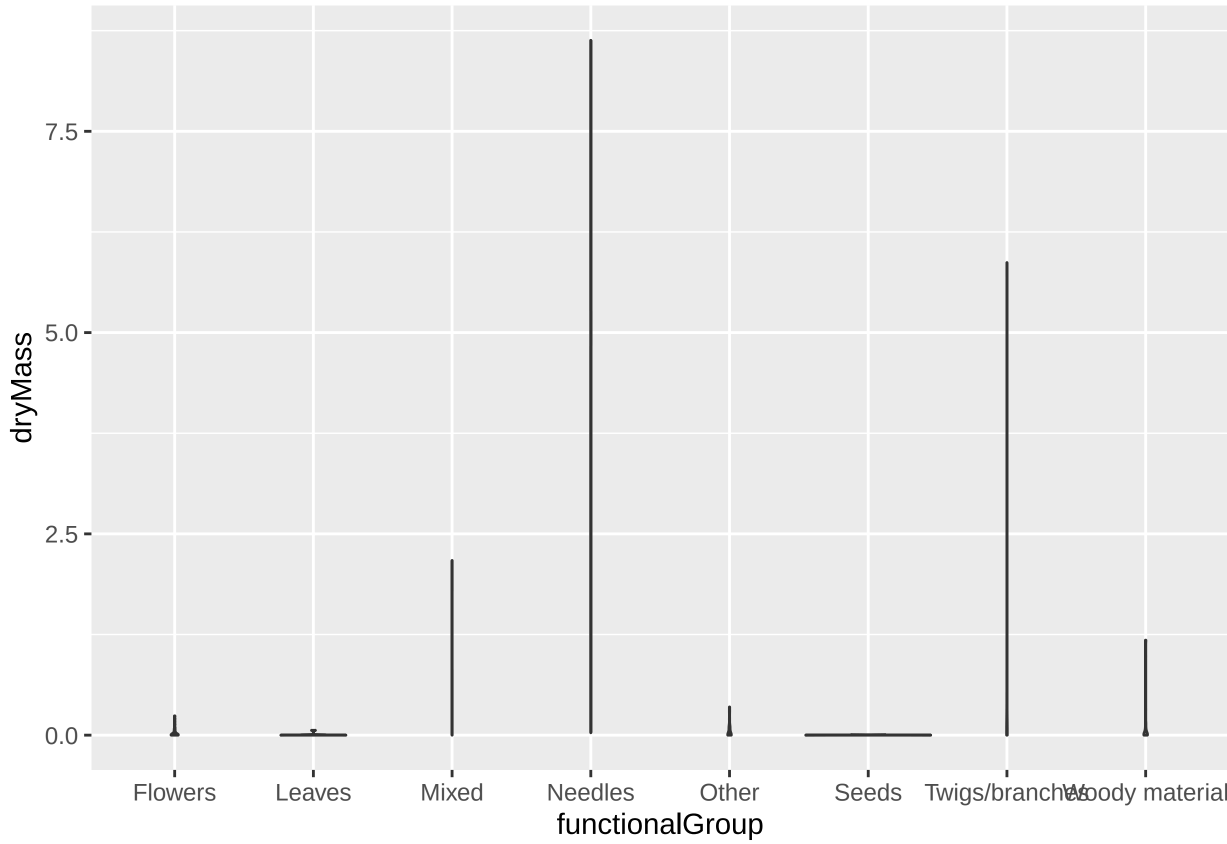



15. Using `geom_boxplot` and `geom_violin`, create a boxplot and a violin plot of dryMass by functionalGroup.

```
ggplot(Litter) +
  geom_boxplot(aes(functionalGroup,dryMass))
```




```
ggplot(Litter) +
  geom_violin(aes(functionalGroup,dryMass))
```




Why is the boxplot a more effective visualization option than the violin plot in this case?

> Answer: The boxplot provides more visual statistical information like as median and outliers. Although a violin plot can show this information and more, there is not high enough density of discrete values for the plot to be effective.

What type(s) of litter tend to have the highest biomass at these sites?

> Answer: Needles tend to have the highest biomass at these sites.