Assignment 3: Data Exploration

Jinglin Zhang

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.

```
setwd("~/EDA-Spring2023")

library(tidyverse)
library(lubridate)

Neonics <- read.csv("./Data/Raw/ECOTOX_Neonicotinoids_Insects_raw.csv",stringsAsFactors = TRUE)
Litter <- read.csv("./Data/Raw/NEON_NIWO_Litter_massdata_2018-08_raw.csv",stringsAsFactors = TRUE)
```

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: because neonicotinoid is a kind of pesticides widly allpied in agriculture, and abundant researches revealed that this toxin can permanently binding to the nerve cells of insects, overstimulating and destroying them. They can remain in the soil, food, insect and human bodies through the food chain, which bring a potential risk of physical health. checking the the remained neonicotinoid in insect can indicate the pollution level of neonicotinoid and be helpful to assess the potentical risk.

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: because the forest litter and woody debris are organic matters from forest organism, they can indicate the forest situation, and they can be easilly be collected and analyzed.

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.litter trap includes 4 40mx40m tower plots and 26 20mx20m plots. 2.a plot includes several clip cells, trap will be placed either targeted or randomized, a 1m buffer will applied around the edge of the plot and subplots. 3. ground traps are sampled once a year but target sampling time is varied by vegitation present at the site.

Obtain basic summaries of your data (Neonics)

5. What are the dimensions of the dataset?

```
dim(Litter)
```

```
## [1] 188  19
```

```
dim(Neonics)
```

```
## [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 effect is population.the effect on population, specifically, the abundance, can discribe the status quo of population in a marco perspective. the effect of pesticides can be analyzed by comparing the population of treatment groups and control groups.

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...]

```
outputSpecies <- summary(Neonics$Species.Common.Name)
sort(outputSpecies)
```

```
##                          Ant Family                         Apple Maggot
##                                   9                                    9
##              Glasshouse Potato Wasp                             Lacewing
##                                  10                                   10
##             Southern House Mosquito             Two Spotted Lady Beetle
##                                  10                                   10
##            Spotless Ladybird Beetle                 Braconid Parasitoid
##                                  11                                   12
##                        Common Thrip        Eastern Subterranean Termite
##                                  12                                   12
##                               Jassid                           Mite Order
##                                  12                                   12
##                           Pea Aphid                   Pond Wolf Spider
##                                  12                                   12
##               Armoured Scale Family                   Diamondback Moth
##                                  13                                   13
##                        Eulophid Wasp                  Monarch Butterfly
##                                  13                                   13
##                        Predatory Bug               Yellow Fever Mosquito
##                                  13                                   13
##                         Corn Earworm                  Green Peach Aphid
##                                  14                                   14
##                            House Fly                            Ox Beetle
##                                  14                                   14
##                  Red Scale Parasite                  Spined Soldier Bug
##                                  14                                   14
##              Western Flower Thrips Hemlock Woolly Adelgid Lady Beetle
```

```
##                              15                                16
##              Hemlock Wooly Adelgid                              Mite
##                              16                                16
##                       Onion Thrip              Araneoid Spider Order
##                              16                                17
##                         Bee Order                     Egg Parasitoid
##                              17                                17
##                      Insect Class           Moth And Butterfly Order
##                              17                                17
##       Oystershell Scale Parasitoid          Black-spotted Lady Beetle
##                              17                                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
##                      Codling Moth          Flatheaded Appletree Borer
##                              19                                20
##                Horned Oak Gall Wasp                 Leaf Beetle Family
##                              20                                20
##                  Potato Leafhopper          Tooth-necked Fungus Beetle
##                              20                                20
##                     Argentine Ant                            Beetle
##                              21                                21
##                        Mason Bee                          Mosquito
##                              22                                22
##                   Citrus Leafminer                   Ladybird Beetle
##                              23                                23
##                  Spider/Mite Class                 Tobacco Flea Beetle
##                              24                                24
##                      Chalcid Wasp             Convergent Lady Beetle
##                              25                                25
##                     Stingless Bee               Ground Beetle Family
##                              25                                27
##                Rove Beetle Family                      Tobacco Aphid
##                              27                                27
##                     Scarab Beetle                     Spring Tiphia
##                              29                                29
##                       Thrip Order             Ladybird Beetle Family
##                              29                                30
##                        Parasitoid                     Braconid Wasp
##                              30                                33
##                      Cotton Aphid                     Predatory Mite
##                              33                                33
##                Sweetpotato Whitefly                       Aphid Family
##                              37                                38
##                     Cabbage Looper              Buff-tailed Bumblebee
##                              38                                39
##                     True Bug Order            Sevenspotted Lady Beetle
##                              45                                46
##                       Beetle Order          Snout Beetle Family, Weevil
```

```
##                                47                                47
##               Erythrina Gall Wasp                  Parasitoid Wasp
##                                49                                51
##            Colorado Potato Beetle                    Parastic Wasp
##                                57                                58
##              Asian Citrus Psyllid                 Minute Pirate Bug
##                                60                                62
##                 European Dark Bee                          Wireworm
##                                66                                69
##                    Euonymus Scale                 Asian Lady Beetle
##                                75                                76
##                  Japanese Beetle                  Italian Honeybee
##                                94                               113
##                       Bumble Bee                Carniolan Honey Bee
##                               140                               152
##             Buff Tailed Bumblebee                     Parasitic Wasp
##                               183                               285
##                        Honey Bee                           (Other)
##                               667                               670
```

Answer: they all belong to apidae family, it might because many researches pointed out neonics is a major problem for the death of massive bee, and the livelihood of bees have strong connection with human food security, thus bees are research hotspots in neonicotinoids.

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.)
```

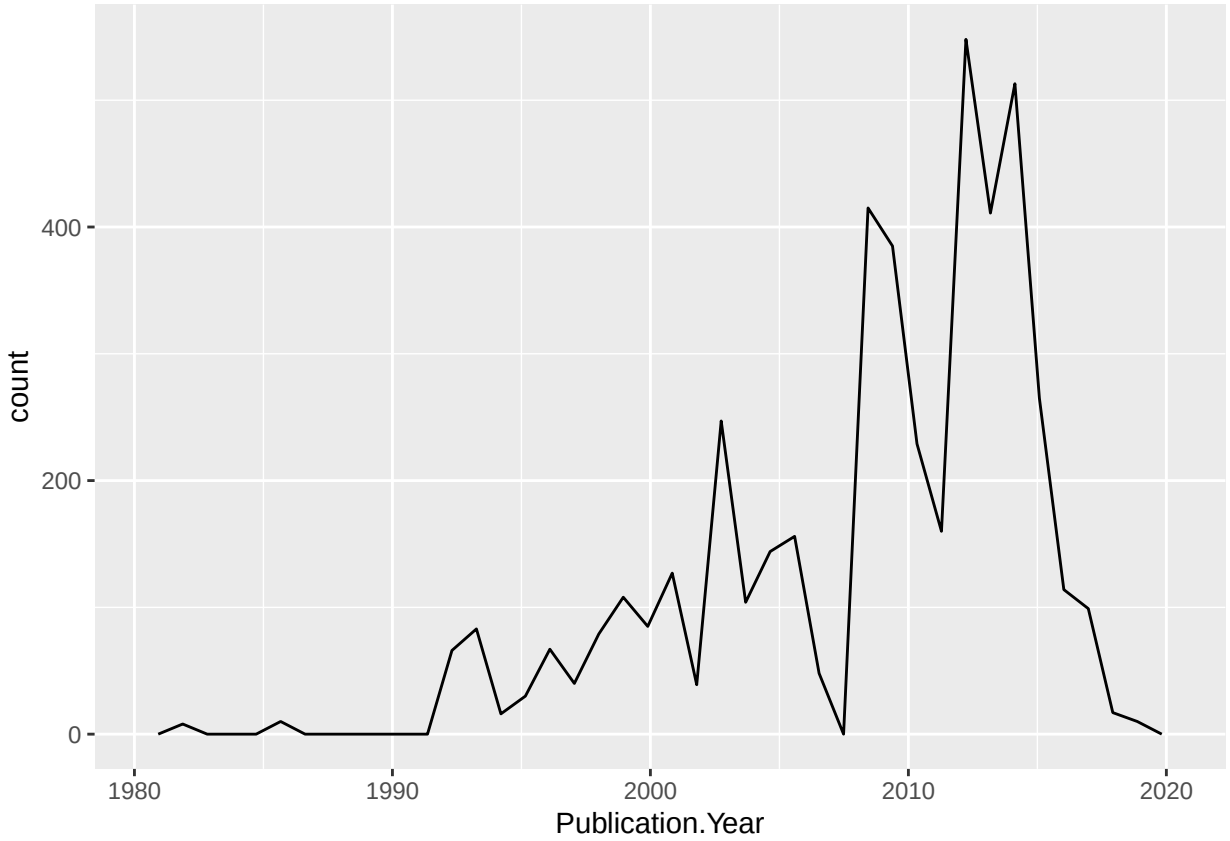
```
## [1] "factor"
```

Answer:because factor values are used for categorical data.

Explore your data graphically (Neonics)

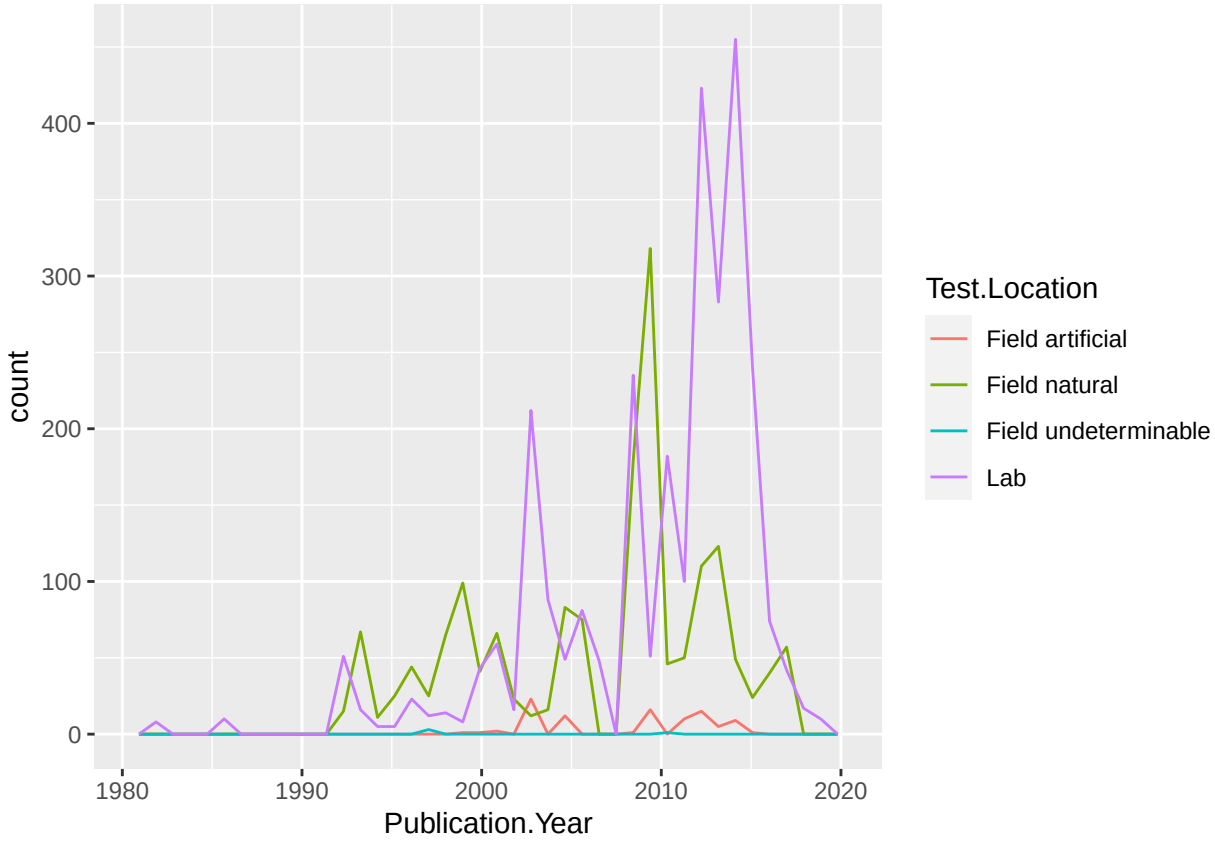
9. Using `geom_freqpoly`, generate a plot of the number of studies conducted by publication year.

```
ggplot(Neonics) +
  geom_freqpoly(aes(x = Publication.Year), bins = 40)
```



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

```
ggplot(Neonics) +
  geom_freqpoly(aes(x = Publication.Year, color = Test.Location), bins = 40)
```



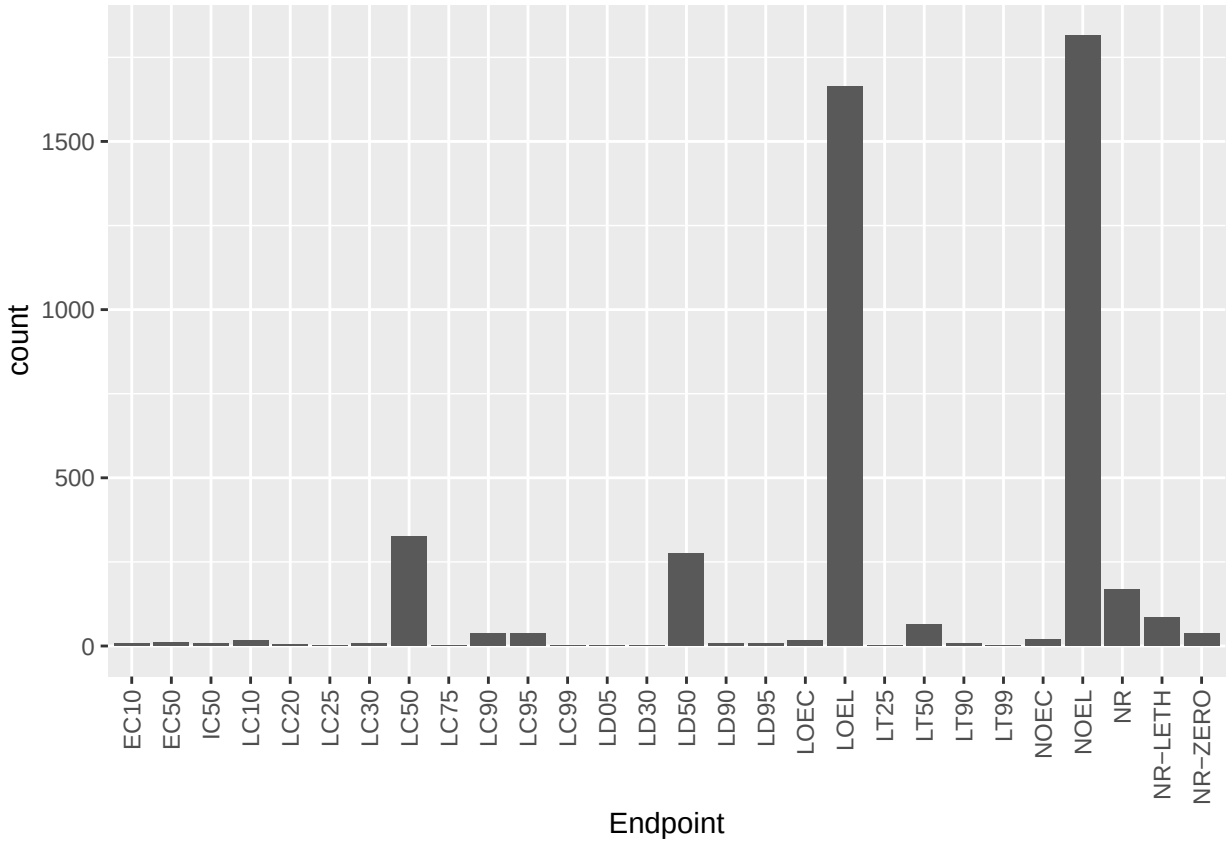
Interpret this graph. What are the most common test locations, and do they differ over time?

Answer:the most comment test location after 2010 is lab, before 2010, the most common location is natural field, but occasionally is lab.

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…]

```
ggplot(Neonics) +
  geom_bar(aes(x = Endpoint)) +
    theme(axis.text.x = element_text(angle = 90, vjust = 0.5, hjust=1))
```



Answer:two most comment endpoints are LOEC and NOEL, LOEC means "Lowest observable effect concentration". NOEL means "highest dose (concentration) producing effects not significantly different from responses of controls".

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)
```

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

```
Litter$collectDate <- as.Date(Litter$collectDate, format = "%Y-%m-%d")
class(Litter$collectDate)
```

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

```
unique(Litter$collectDate)
```

```
## [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)
```

```
##  [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
```

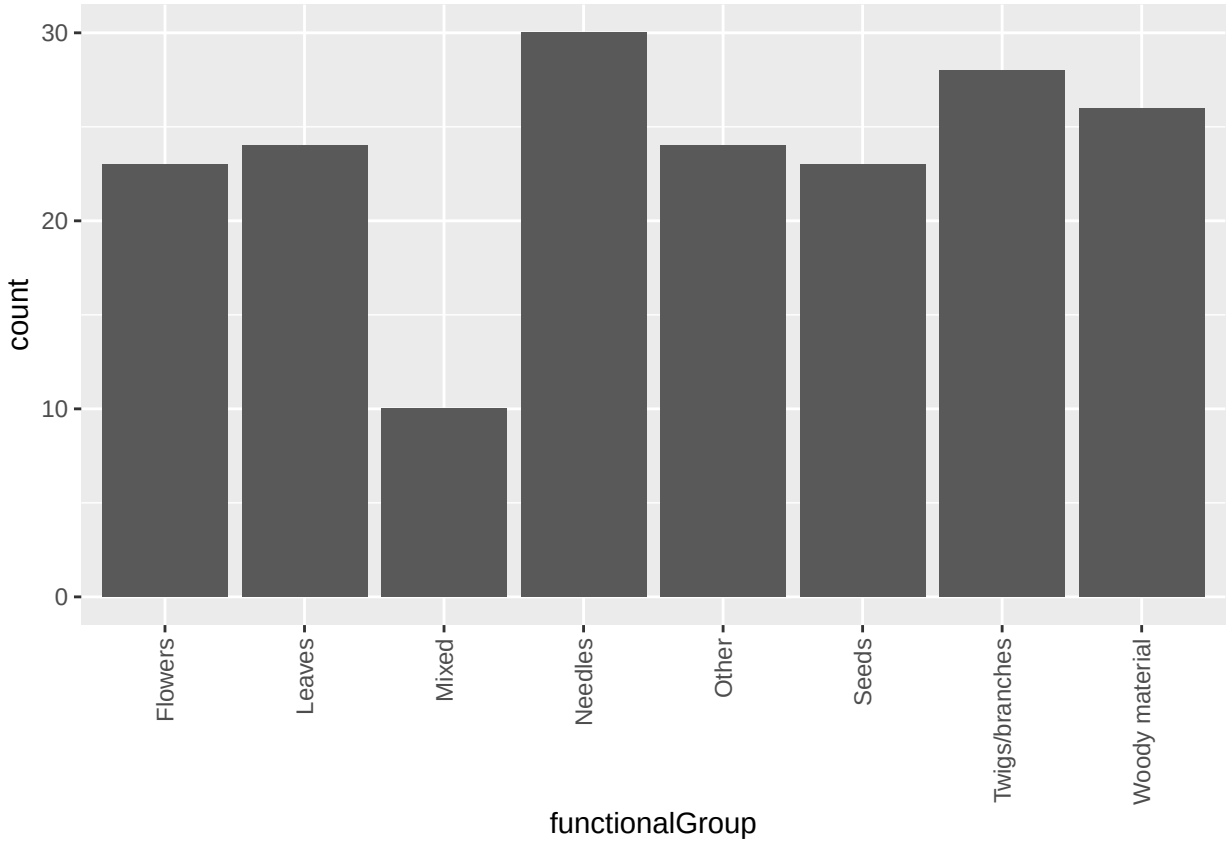
```
summary(Litter$plotID)
```

```
## 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: both unique and summary function can give information that which plots are sampled, but unique function will not provide the count of eath plot.

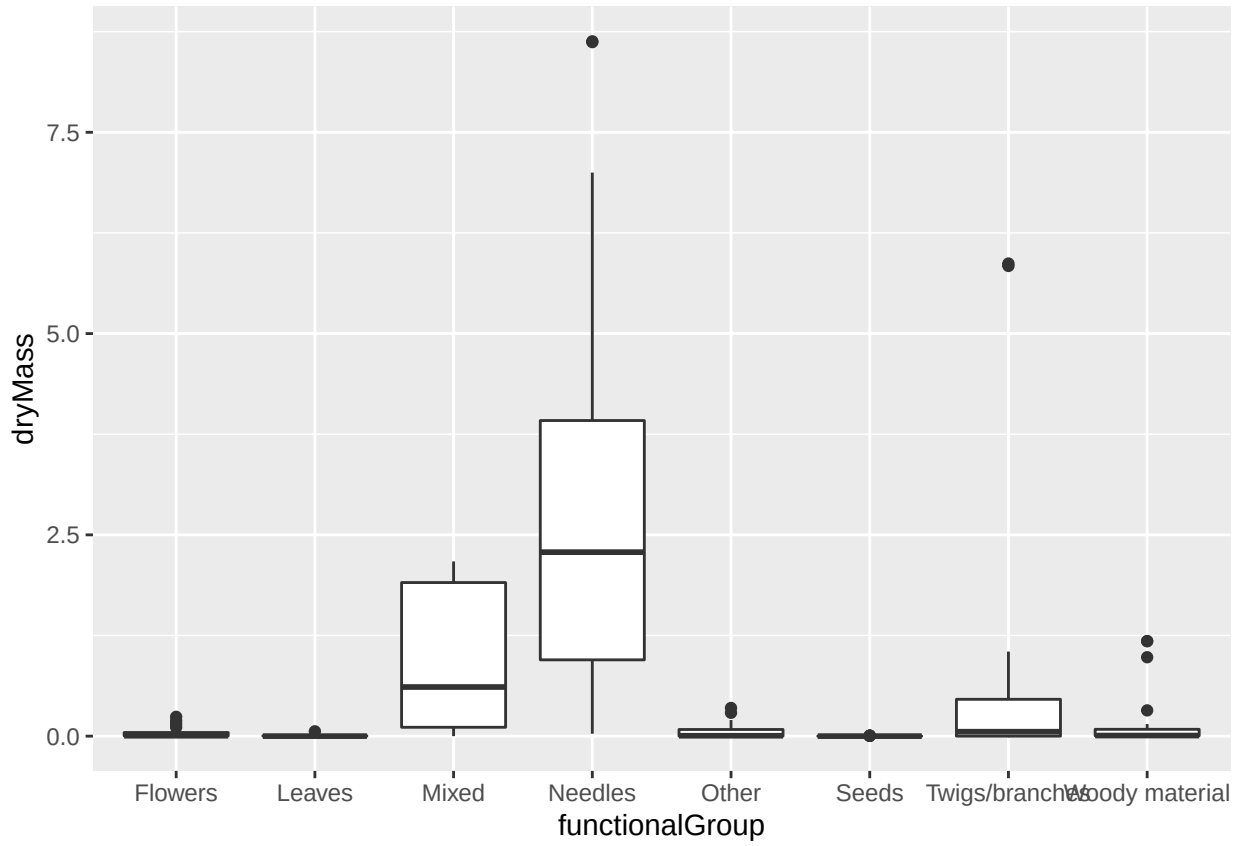
14. Create a bar graph of functionalGroup counts. This shows you what type of litter is collected at the Niwot Ridge sites. Notice that litter types are fairly equally distributed across the Niwot Ridge sites.

```
ggplot(Litter) +
  geom_bar(aes(x = functionalGroup)) +
  theme(axis.text.x = element_text(angle = 90, vjust = 0.5, hjust=1))
```

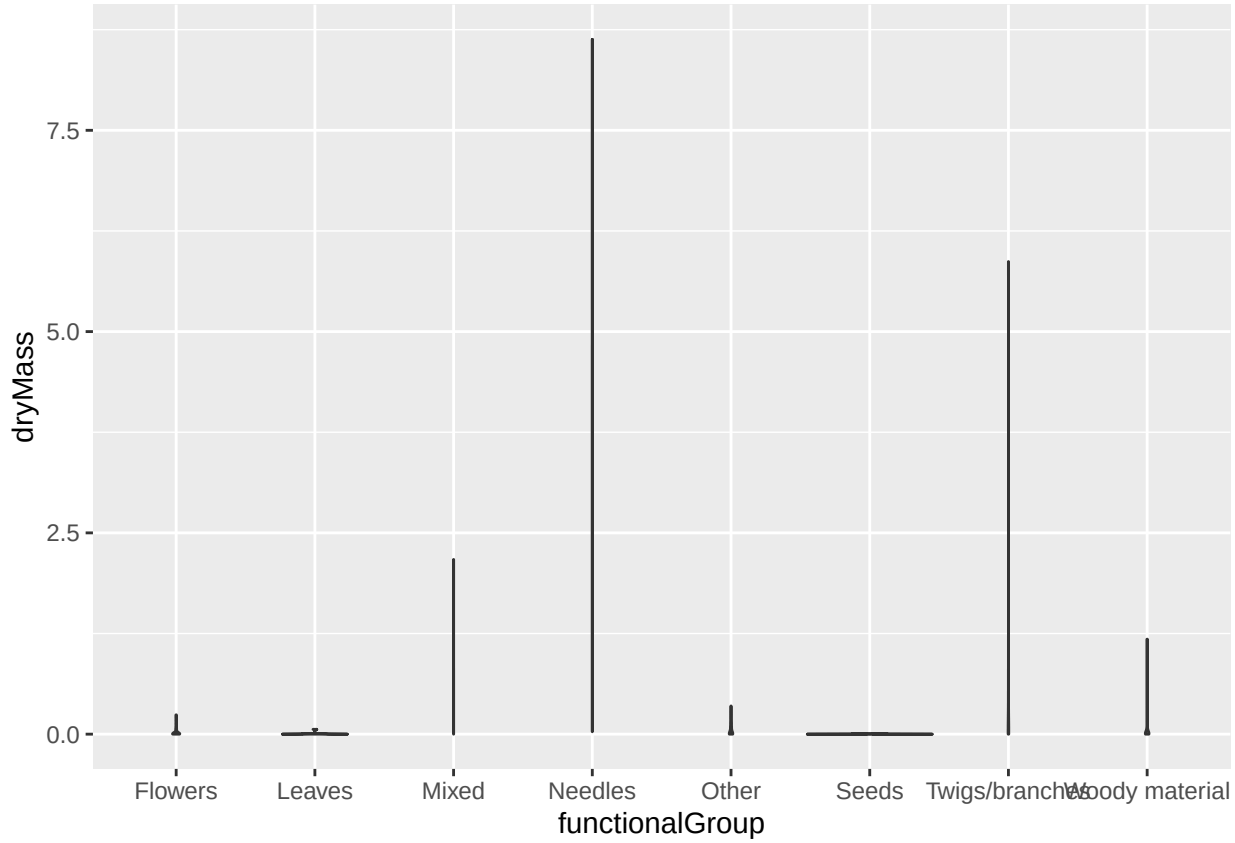


15. Using `geom_boxplot` and `geom_violin`, create a boxplot and a violin plot of dryMass by functionalGroup.

```
ggplot(Litter) +
  geom_boxplot(aes(x = functionalGroup, y = dryMass))
```




```
ggplot(Litter) +
  geom_violin(aes(x = functionalGroup, y = dryMass))
```



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

> Answer: because in this dataset, the most drymass value distribute concentrated, the advantages of violin plot (show the probability density of the data) cannot be shown in this dataset. in this dataset, boxplot can show the variation, mean better than violin plot.

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

> Answer:needles