Assignment 3: Data Exploration

Azura Liu, Section #4

OVERVIEW

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

Directions

1. Change "Student Name, Section #" on line 3 (above) with your name and section number.
2. Work through the steps, **creating code and output** that fulfill each instruction.
3. Be sure to **answer the questions** in this assignment document.
4. When you have completed the assignment, **Knit** the text and code into a single PDF file.
5. After Knitting, submit the completed exercise (PDF file) to the dropbox in Sakai. Add your last name into the file name (e.g., "FirstLast_A03_DataExploration.Rmd") prior to submission.

The completed exercise is due on <>.

Set up your R session

1. Check your working directory, load necessary packages (tidyverse), 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 add the `stringsAsFactors = TRUE` parameter to the function when reading in the CSV files.**

```
getwd()
```

```
## [1] "C:/Users/Idae/Desktop/ENV872/Environmental_Data_Analytics_2022/Assignments"
```

```
#install.packages("tidyverse") #just voiding the code so it does not keep installing it
library("tidyverse")
ecotox<-read.csv("../Data/Raw/ECOTOX_Neonicotinoids_Insects_raw.csv")
litter<-read.csv("../Data/Raw/NEON_NIWO_Litter_massdata_2018-08_raw.csv")
```

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 ecotoxicologoy of neonicotinoids on insects? Feel free to do a brief internet search if you feel you need more background information.

   Answer: This study may help the regulation and development of insecticides for commercial use, especially when neonicotinoids exist as many plants' natural defense mechanism. The regulation agency can determine whether the insecticide is harmful to non-pest species (e.g. pollinators); companies can look for more effective or safer formulas for their products.

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 litter and woody debris tell us a lot about the forest. It tells you the forest type, ecosystem health, wildlife status, the rate of decomposition, etc. Using this information we can manage our forests more effectively. For example, a bare forest floor can mean the presence of invasive or heavily populated earthworms.

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: The data came from: *spatial sampling (litter traps);* temporal sampling (repeated data collection); *individual reports.

Obtain basic summaries of your data (Neonics)

5. What are the dimensions of the dataset?

```
dim(ecotox)
```

```
## [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?

```
sort(summary(as.factor(ecotox$Effect)), decreasing = TRUE)
```

```
##       Population        Mortality         Behavior Feeding behavior
##             1803             1493              360              255
##     Reproduction      Development        Avoidance         Genetics
##              197              136              102               82
##        Enzyme(s)           Growth       Morphology    Immunological
##               62               38               22               16
##     Accumulation     Intoxication     Biochemistry          Cell(s)
##               12               12               11                9
##       Physiology        Histology       Hormone(s)
##                7                5                1
```

Answer: The most common effects are "Population" and "Mortality". They might be of interest because they can be the most straightforward representation of the effect of insecticides/neonics.

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.

```
sort(summary(as.factor(ecotox$Species.Common.Name)), decreasing = TRUE)
```

```
##                            (Other)                         Honey Bee
##                                670                               667
##                       Parasitic Wasp               Buff Tailed Bumblebee
##                                285                               183
##                 Carniolan Honey Bee                        Bumble Bee
##                                152                               140
##                    Italian Honeybee                     Japanese Beetle
##                                113                                94
##                  Asian Lady Beetle                     Euonymus Scale
##                                 76                                75
```

```
##                      Wireworm                European Dark Bee
##                            69                               66
##              Minute Pirate Bug             Asian Citrus Psyllid
##                            62                               60
##                 Parastic Wasp          Colorado Potato Beetle
##                            58                               57
##               Parasitoid Wasp             Erythrina Gall Wasp
##                            51                               49
##                  Beetle Order      Snout Beetle Family, Weevil
##                            47                               47
##         Sevenspotted Lady Beetle                 True Bug Order
##                            46                               45
##            Buff-tailed Bumblebee                   Aphid Family
##                            39                               38
##                 Cabbage Looper           Sweetpotato Whitefly
##                            38                               37
##                  Braconid Wasp                  Cotton Aphid
##                            33                               33
##                 Predatory Mite          Ladybird Beetle Family
##                            33                               30
##                     Parasitoid                  Scarab Beetle
##                            30                               29
##                  Spring Tiphia                   Thrip Order
##                            29                               29
##             Ground Beetle Family             Rove Beetle Family
##                            27                               27
##                  Tobacco Aphid                  Chalcid Wasp
##                            27                               25
##           Convergent Lady Beetle                 Stingless Bee
##                            25                               25
##                Spider/Mite Class             Tobacco Flea Beetle
##                            24                               24
##                Citrus Leafminer                 Ladybird Beetle
##                            23                               23
##                      Mason Bee                        Mosquito
##                            22                               22
##                  Argentine Ant                          Beetle
##                            21                               21
##       Flatheaded Appletree Borer           Horned Oak Gall Wasp
##                            20                               20
##              Leaf Beetle Family              Potato Leafhopper
##                            20                               20
##       Tooth-necked Fungus Beetle                  Codling Moth
##                            20                               19
##        Black-spotted Lady Beetle                  Calico Scale
##                            18                               18
##              Fairyfly Parasitoid                   Lady Beetle
##                            18                               18
##           Minute Parasitic Wasps                     Mirid Bug
##                            18                               18
##                Mulberry Pyralid                      Silkworm
##                            18                               18
##                  Vedalia Beetle           Araneoid Spider Order
##                            18                               17
```

```
##                          Bee Order                      Egg Parasitoid
##                                 17                                  17
##                       Insect Class            Moth And Butterfly Order
##                                 17                                  17
##       Oystershell Scale Parasitoid Hemlock Woolly Adelgid Lady Beetle
##                                 17                                  16
##              Hemlock Wooly Adelgid                               Mite
##                                 16                                  16
##                        Onion Thrip              Western Flower Thrips
##                                 16                                  15
##                       Corn Earworm                  Green Peach Aphid
##                                 14                                  14
##                           House Fly                           Ox Beetle
##                                 14                                  14
##                 Red Scale Parasite                 Spined Soldier Bug
##                                 14                                  14
##              Armoured Scale Family                   Diamondback Moth
##                                 13                                  13
##                      Eulophid Wasp                  Monarch Butterfly
##                                 13                                  13
##                      Predatory Bug              Yellow Fever Mosquito
##                                 13                                  13
##                Braconid Parasitoid                         Common Thrip
##                                 12                                  12
##       Eastern Subterranean Termite                               Jassid
##                                 12                                  12
##                         Mite Order                          Pea Aphid
##                                 12                                  12
##                   Pond Wolf Spider           Spotless Ladybird Beetle
##                                 12                                  11
##             Glasshouse Potato Wasp                           Lacewing
##                                 10                                  10
##            Southern House Mosquito            Two Spotted Lady Beetle
##                                 10                                  10
##                          Ant Family                       Apple Maggot
##                                  9                                   9
```

Answer: Honey bee and Parasitic Wasp are the most common species.

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

```
class(ecotox$Conc.1..Author.)
```

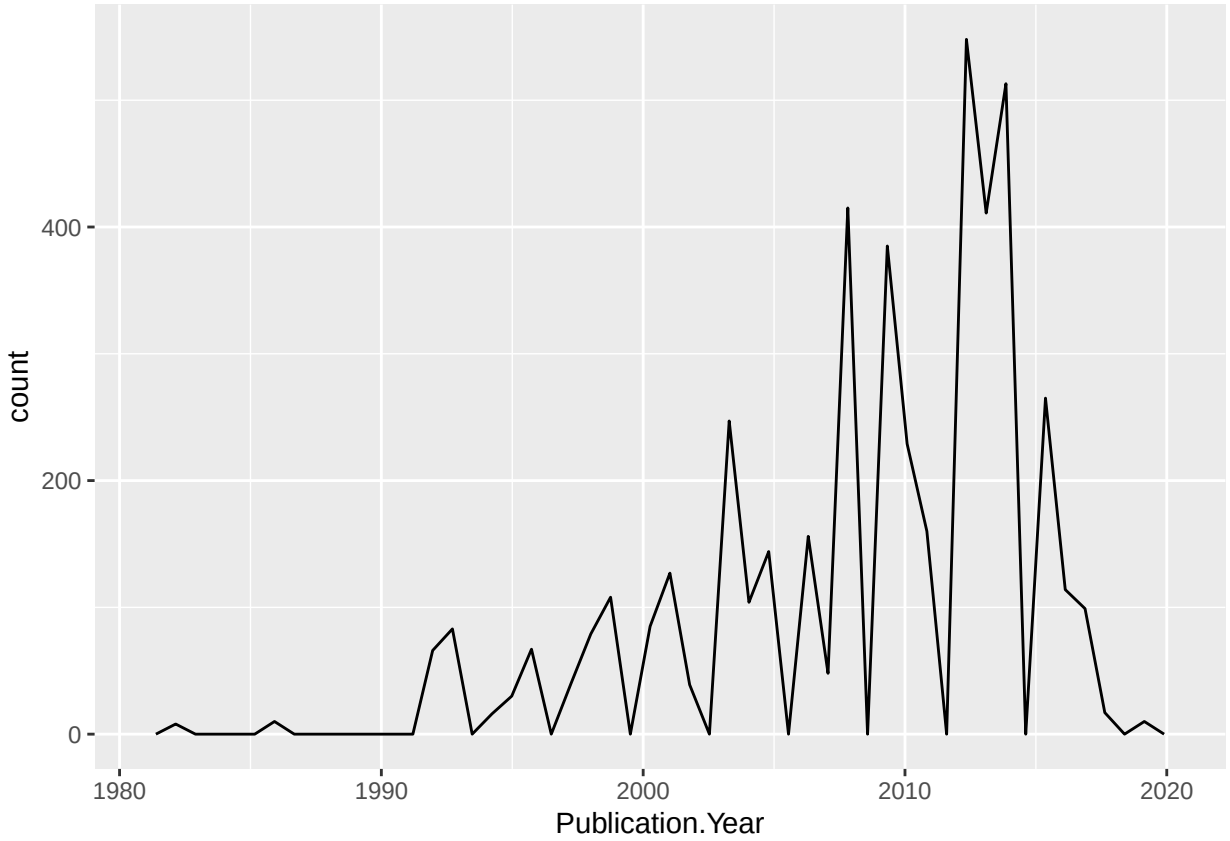
```
## [1] "character"
```

Answer: Invalid or missing values are presented in the column, making it non-numeric.

Explore your data graphically (Neonics)

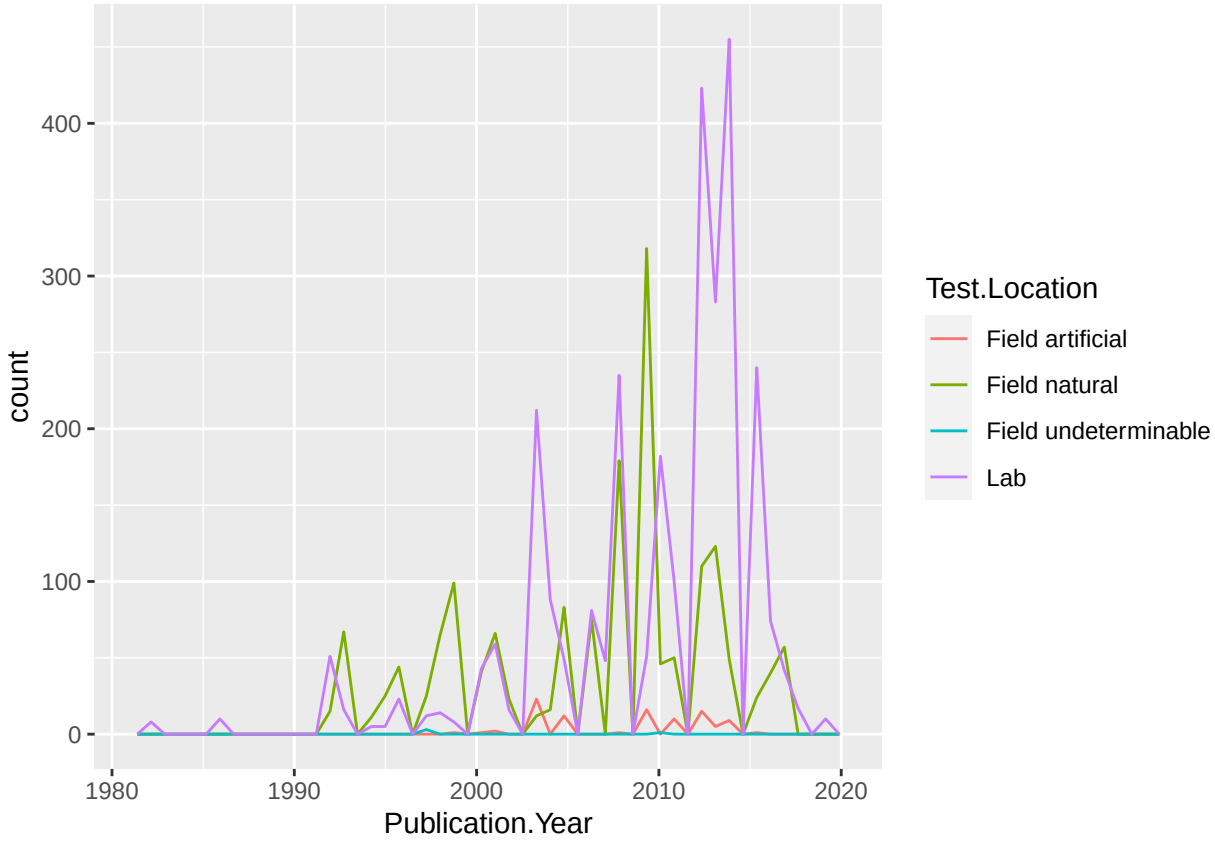
9. Using `geom_freqpoly`, generate a plot of the number of studies conducted by publication year.

```
ggplot(ecotox) +
  geom_freqpoly(aes(x = Publication.Year), bins = 50)
```



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

```
ggplot(ecotox) +
  geom_freqpoly(aes(x = Publication.Year, color= Test.Location ), bins = 50)
```

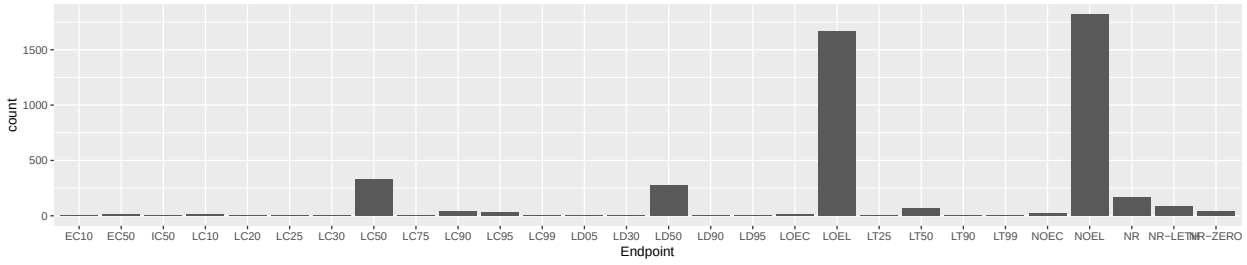


Interpret this graph. What are the most common test locations, and do they differ over time?

Answer: The most common test location (other than undetermined ones) shifted from filed natural to field artifitial and to 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.

```
ggplot(ecotox, aes(x = Endpoint) )+
  geom_bar()
```



Answer: The most common end points are LD50 and NOEL. LD50 refers to the lethal dose to 50% of test animals; NOEL means the highest dose (concentration) producing effects not significantly different from responses of controls according to author's reported statistical test.

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] "character"
```

```
#not a date
litter$collectDate<- as.Date(litter$collectDate, format = "%y-%m-%d")
class(litter$collectDate)
```

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

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`?

```
sort(summary(as.factor(litter$plotID)), decreasing = TRUE)
```

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

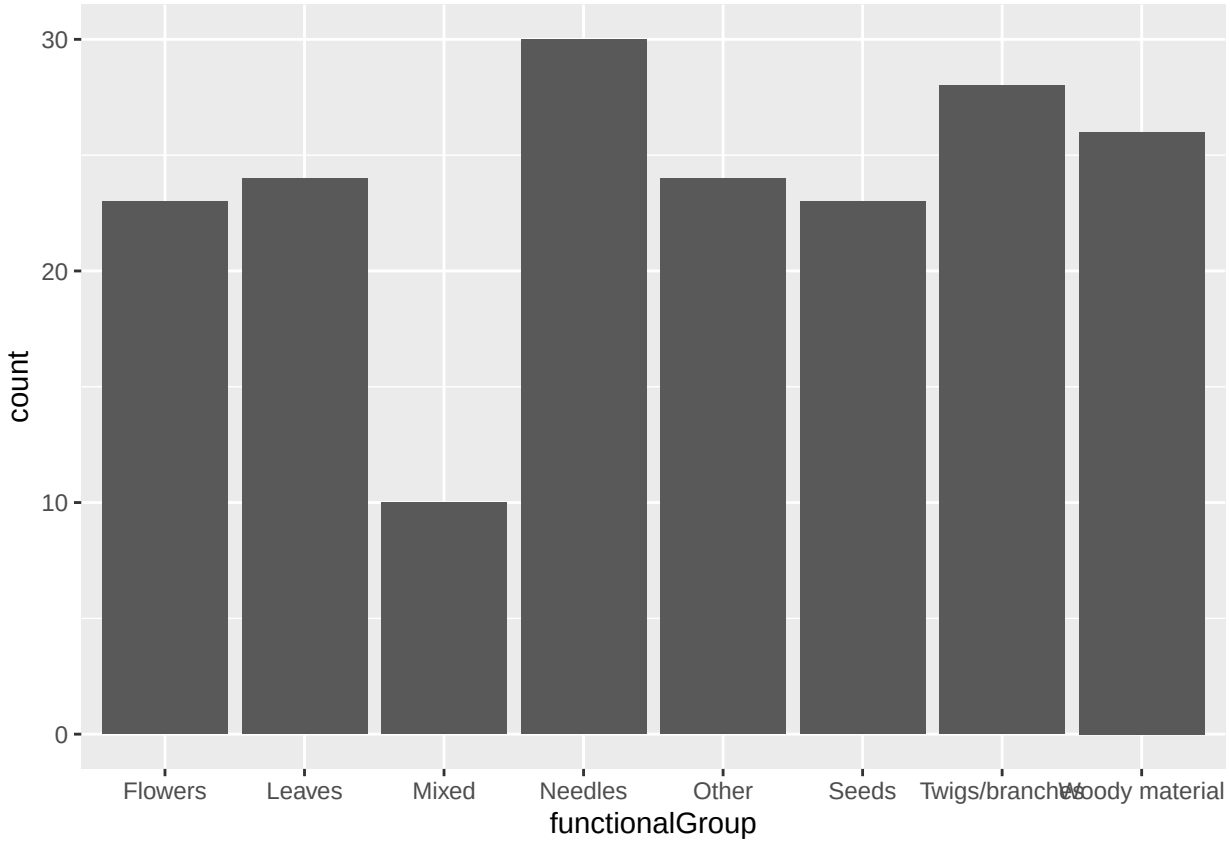
```
sort(unique(litter$plotID))
```

```
##  [1] "NIWO_040" "NIWO_041" "NIWO_046" "NIWO_047" "NIWO_051" "NIWO_057"
##  [7] "NIWO_058" "NIWO_061" "NIWO_062" "NIWO_063" "NIWO_064" "NIWO_067"
```

Answer: The "unique" function only shows the unique values, not including the counts for each.

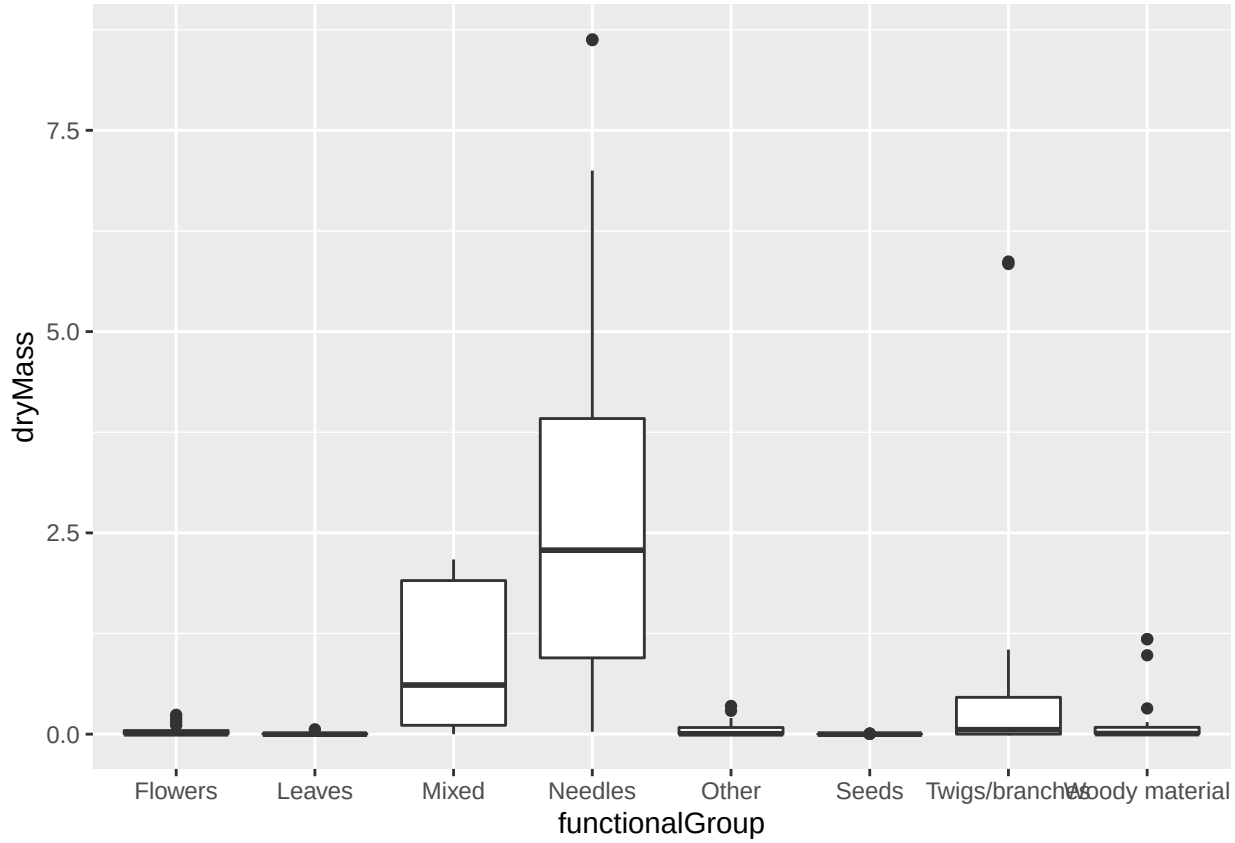
14. Create a bar graph of functionalGroup counts. This shows you what type of litter is collected at the Niwot Ridge sites. Notice that litter types are fairly equally distributed across the Niwot Ridge sites.

```
ggplot(litter, aes(x =functionalGroup ) )+
  geom_bar()
```



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))
```

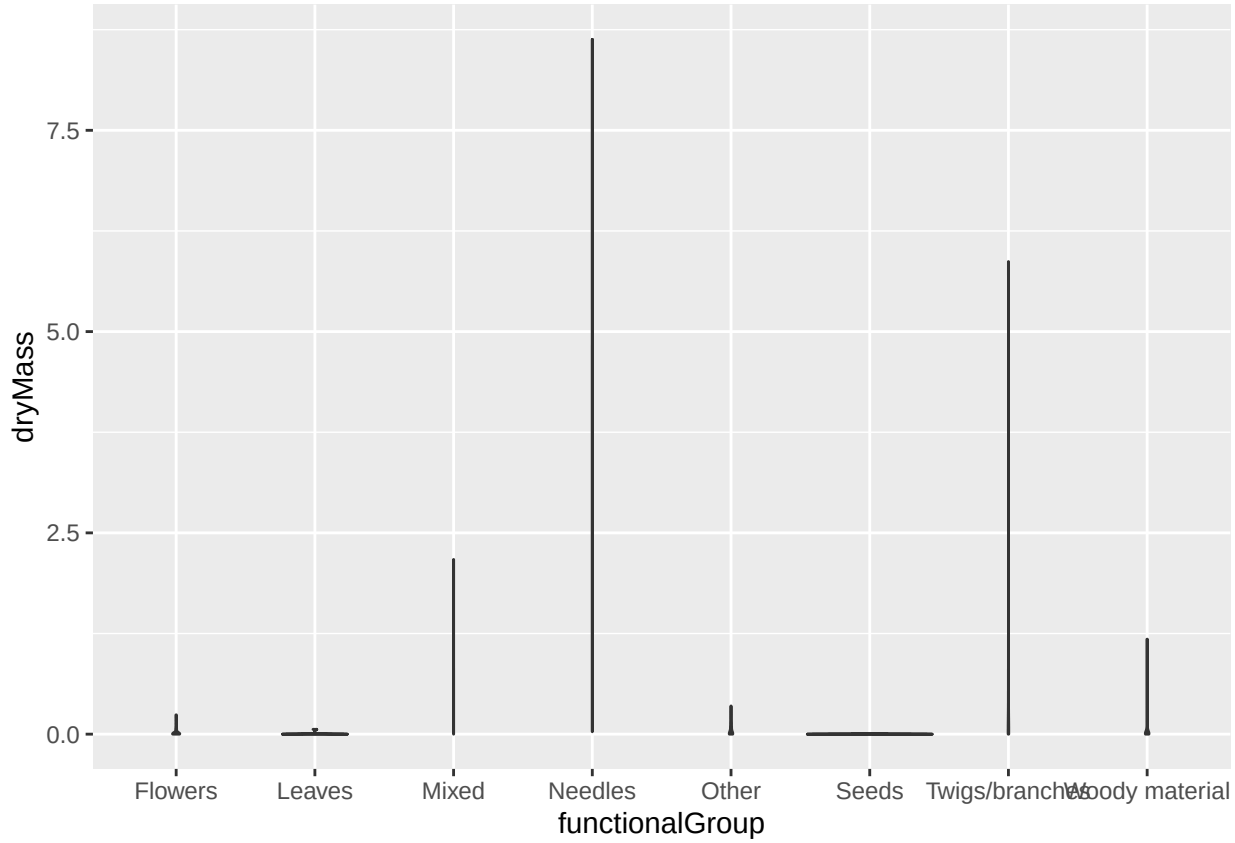



```
ggplot(litter) +
  geom_violin(aes(x = functionalGroup, y = dryMass),
              draw_quantiles = c(0.25, 0.5, 0.75))
```

```
## Warning in regularize.values(x, y, ties, missing(ties), na.rm = na.rm):
## collapsing to unique 'x' values

## Warning in regularize.values(x, y, ties, missing(ties), na.rm = na.rm):
## collapsing to unique 'x' values

## Warning in regularize.values(x, y, ties, missing(ties), na.rm = na.rm):
## collapsing to unique 'x' values
```



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

> Answer: The boxplot does not only show the counts but also gives us an idea what the distributions look like. In this case, the counts of the functional groups barely tell us anything. But the bpxplot can visualize the dry mass distribution that actually tell us which litter types are more out there.

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

> Answer: Needles have the highest biomass.