Assignment 3: Data Exploration

Catherine Otero, Section #1

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 assignments tab 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 <Jan 31, 2022 at 7pm>.

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/Catherine/OneDrive - Duke University/GitHub_Spot/Environmental_Data_Analytics_2021/Envi
```

```
library(tidyverse)

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

Answer: Insects can be indicator species that show effects of neonicotinoids earlier than the effect on humans may be seen. Also, insects are often an indicator of ecosystem health, so seeing how insects react to neonicotinoids may help protect enviornmental health.

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 amount of litter and woody debris can show primary productivity of a system and habitat availability of organisms. It can also help plan for controlled burns and fire management.

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: *Litter is collected in PVC traps that are 0.5m squared and elevated above the groudn by about 80cm..* Litter has a butt end diameter of less than 2 cm and is less than 50 cm long. *The timing of target sampling depends on the vegetation type. Deciduous tree areas are sampled 1 time every 2 weeks during senescence. Evergreen areas are sampled year-round 1 time every 1-2 months.

Obtain basic summaries of your data (Neonics)

5. What are the dimensions of the dataset?

```
#4623 rows and 30 columns
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: Mortality (1493) and Population (1803) are the effects most studied. These effects are likely of interest to see the mortality effect of neonicotinoids on the total population.

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.

```r
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: Honey Bee (667), Parasitic Wasp (285), Buff Tailed Bumblebee (183), Carniolan Honey

Bee (152), Bumble Bee (140), Italian Honeybee (113)

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(Neonics$Conc.1..Author.)
```

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

Answer: Conc.1..Author is a factor because the numbers in the column are associated with values and are not values themselves.

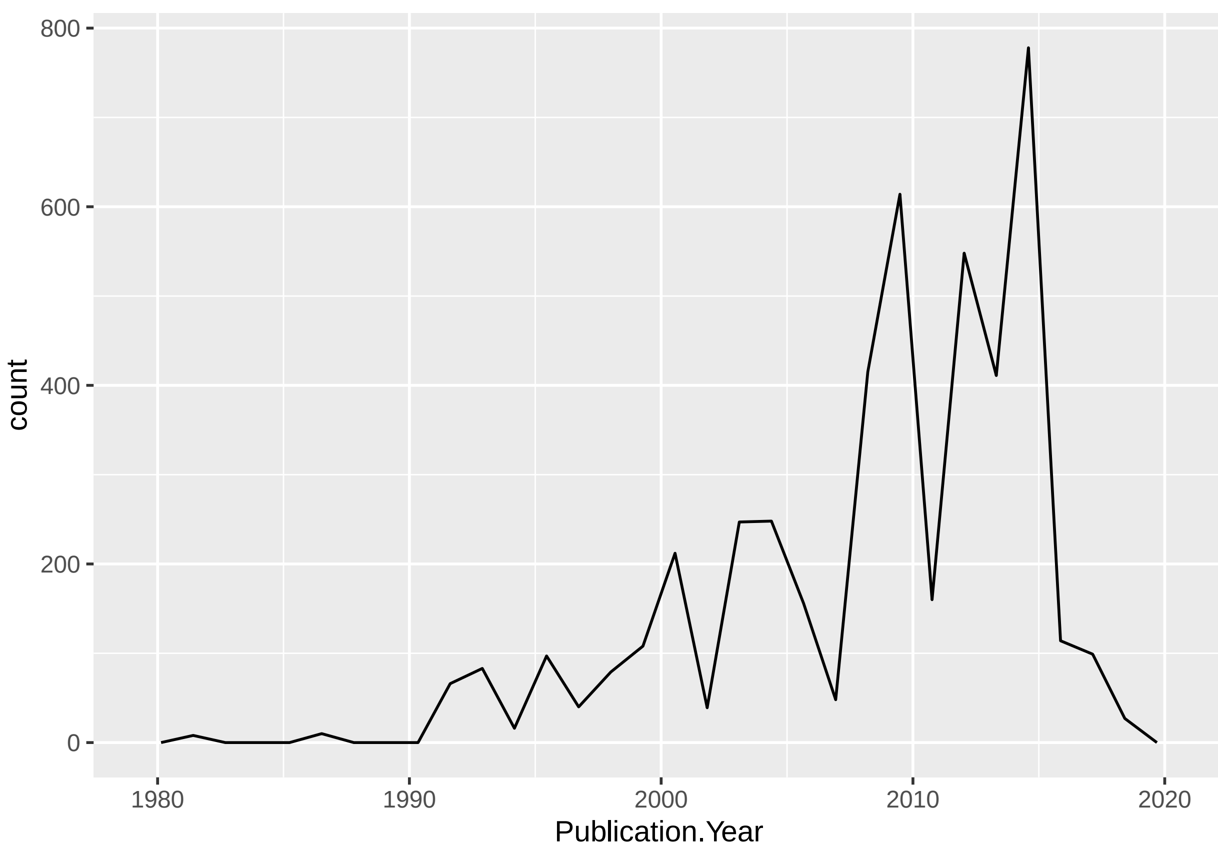
Explore your data graphically (Neonics)

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

```
class(Neonics$Publication.Year)
```

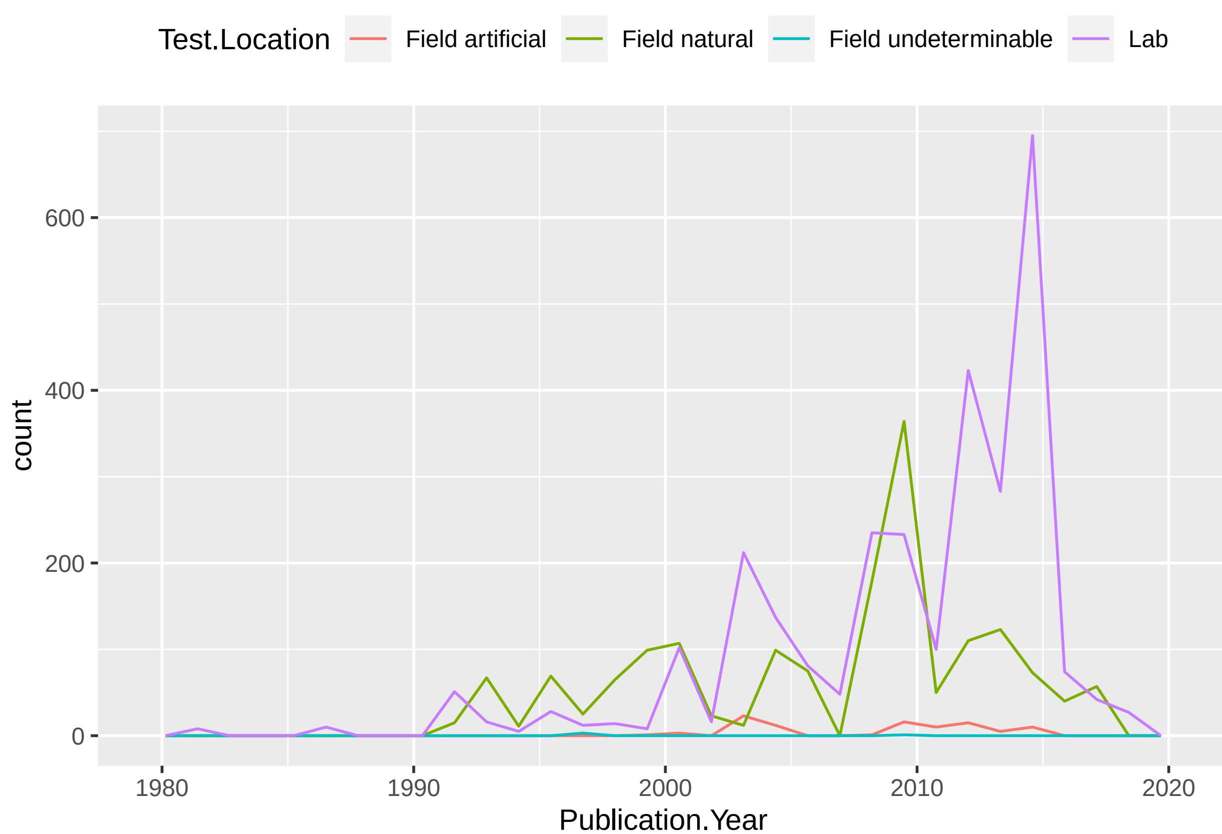
```
## [1] "integer"
```

```
ggplot(Neonics) +
  geom_freqpoly(aes(x = Publication.Year), bins = 30)
```



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 = 30) +
  theme(legend.position = "top")
```

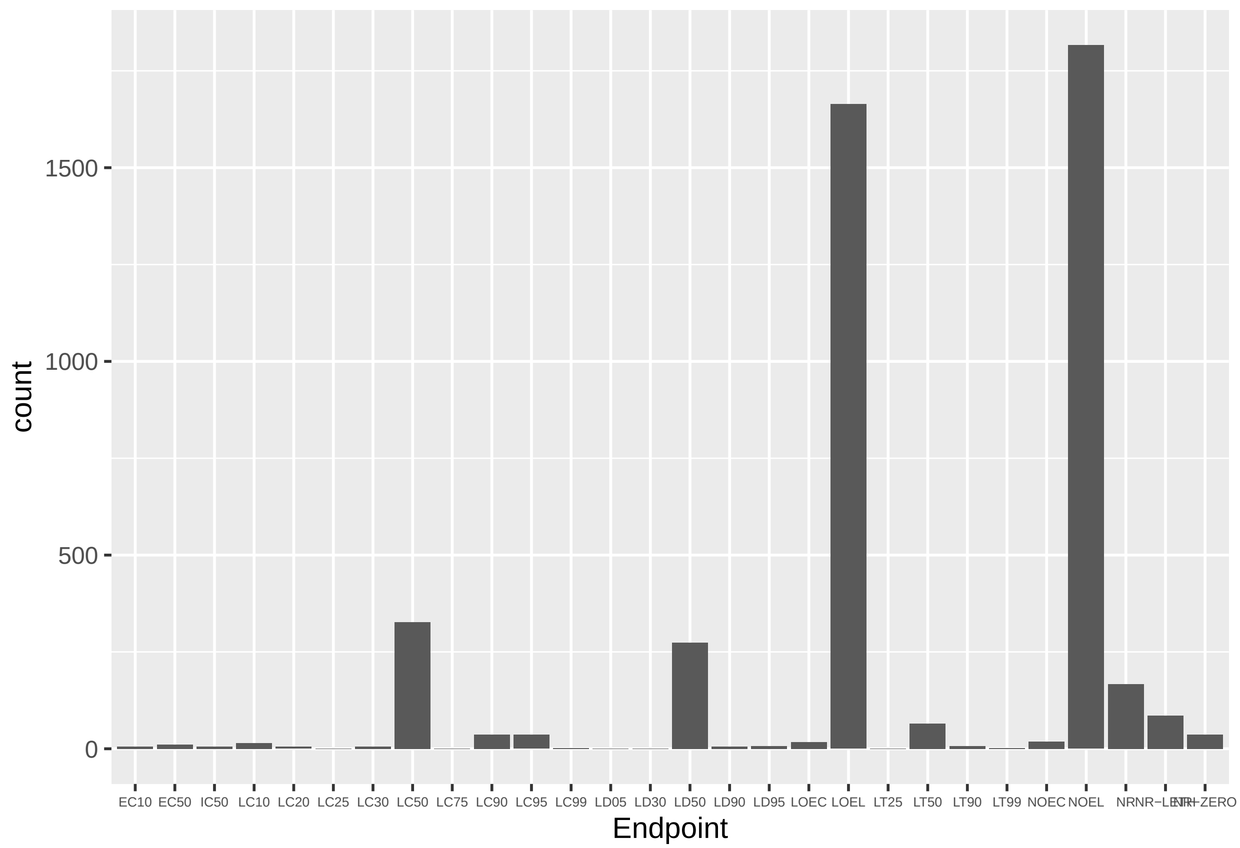


Interpret this graph. What are the most common test locations, and do they differ over time?

> Answer: The most common test locations are field natural and lab. There is a lot of variance overtime with the two most common test locations switching which is most popular. Field natural's last spike was in 2009 and lab's last spike was in 2014.

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(Neonics, aes(x = Endpoint)) +
  geom_bar() +
  theme(axis.text.x=element_text(size=5))
```



Answer: I coudn't figure out how to make my axis readable but the biggest bars are the most common endpoints, probably mortality-related.

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"
```

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
```

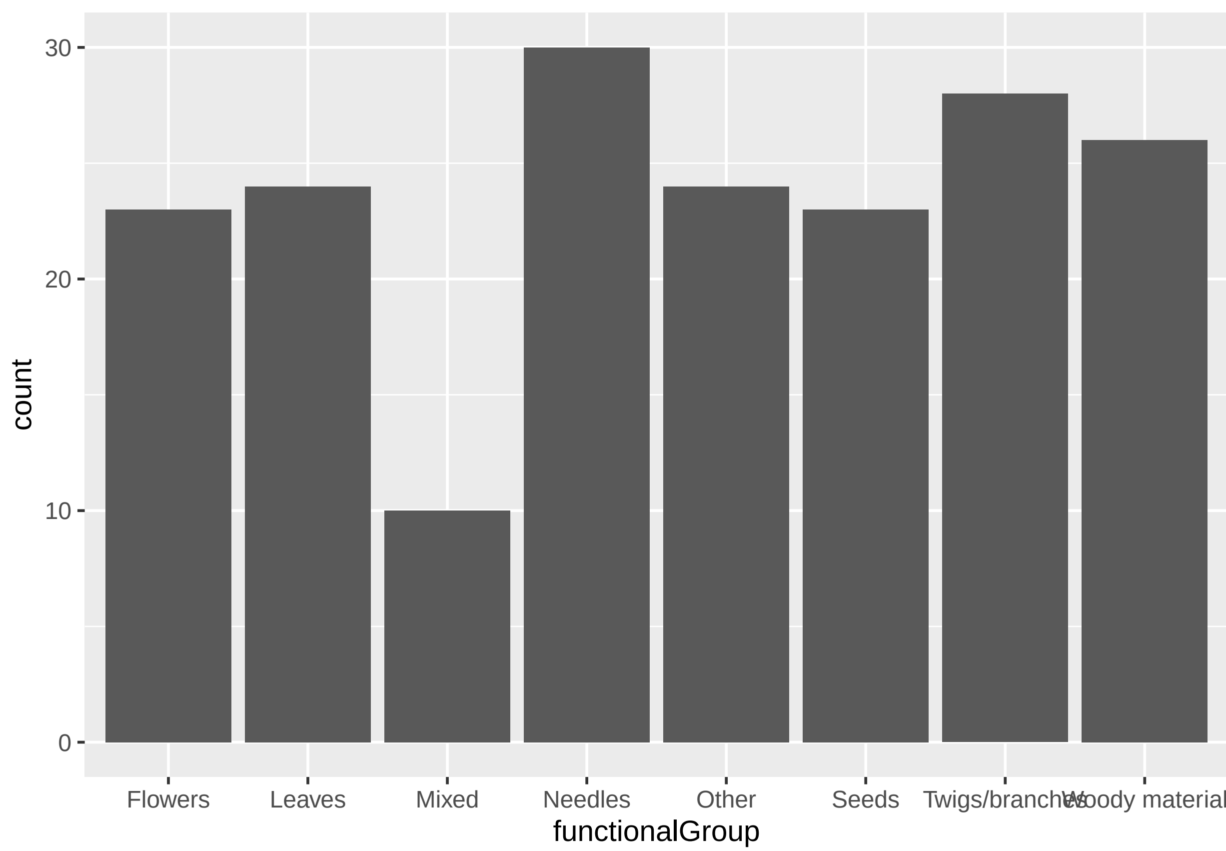
```
summary(Litter$plotID)
```

```
## 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: Unique provides a count of the unique values in the querried column along with the values while summary provides a count of the number of times a value appears in the coumn.

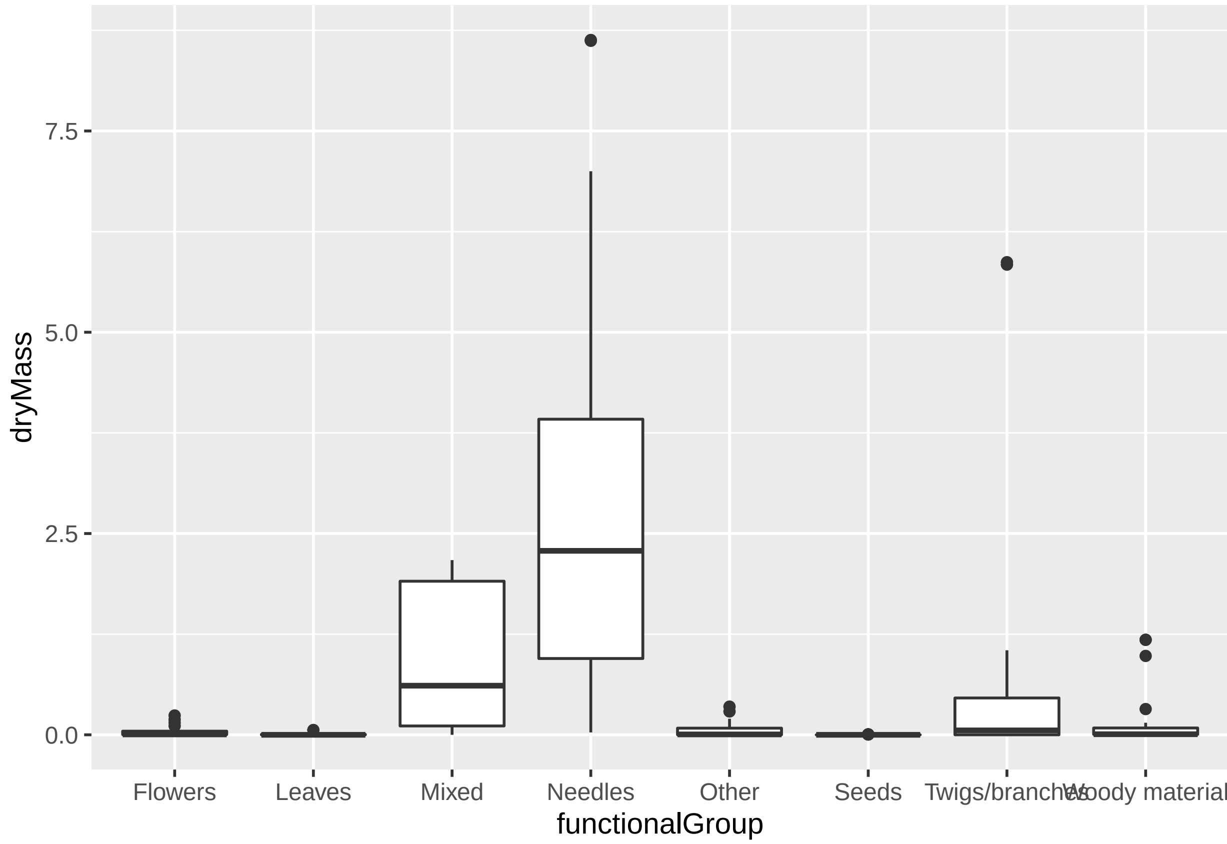
14. Create a bar graph of functionalGroup counts. This shows you what type of litter is collected at the Niwot Ridge sites. Notice that litter types are fairly equally distributed across the Niwot Ridge sites.

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

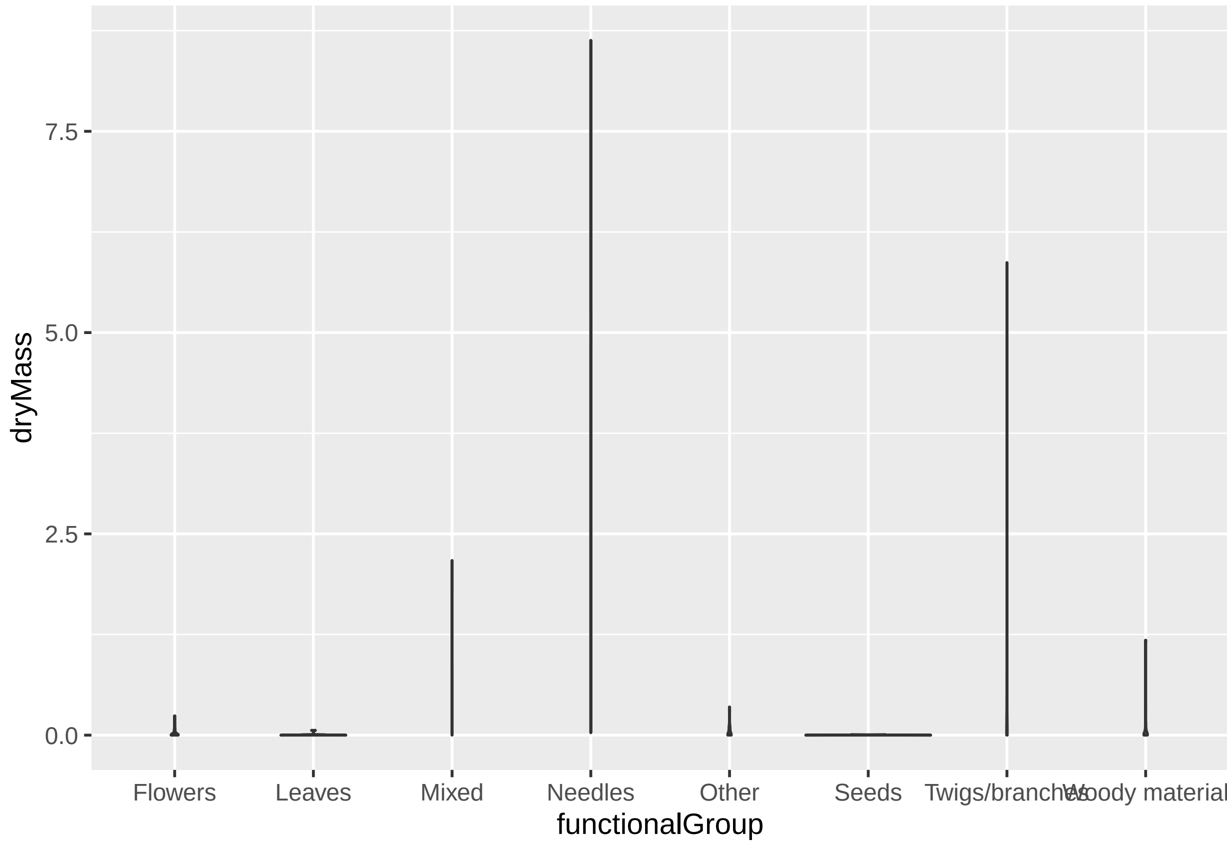


15. Using `geom_boxplot` and `geom_violin`, create a boxplot and a violin plot of dryMass by functionalGroup.

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




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



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

> Answer: A boxplot is more effective because it show distribution of dry mass by functional group while a violin plot shows a narrower range just within the "box" of the boxplot.

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

> Answer: Needles and Mixed have the highest biomass at these sites.