Assignment 3: Data Exploration

Sam Saltman, Section #2

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() #check working directory
```

```
## [1] "/Users/samsaltman/Documents/R/Environmental_Data_Analytics_2022/Assignments"
```

```
library(tidyverse) # load necessary package

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: We may want to understand how this type of insecticide affects insects. Insects can harm crops; however, some insects can be good for agricultural purposes by pollinating crops and/or preying on harmful insects in the field. Data analysis on neonicotinoids can help us understand if an insecticide is successful at targeting harmful insects while not causing harms to insects that are beneficial to the agricultural process.

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: We may want to study the litter and woody debris in forests to determine the health of the forest. This kind of debris will decompose and provide nutrients to the soil. Litter will decompose faster than woody debris. Analyzing litter and woody debris can help researchers understand the replenishment rate of nutrients to the soil and the general lifecycle of a forest.

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 woody debris and litter samples were collected using both elevated and ground traps. * The research used a spatial sampling design to randomly select plot sites. The sites had to contain woody vegetation taller than 2 meters. * Also, the research had a temporal sampling design. Ground traps were measured once a year. Sampling of elevated traps occurred frequently throughout the year but varied depending on the vegetation in the sample.

Obtain basic summaries of your data (Neonics)

5. What are the dimensions of the dataset?

```
dim(Neonics) # dimensions
```

```
## [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?

```
Neonics_effects <- summary(Neonics$Effect)
Neonics_effects
```

```
##     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: Population (1803) & Mortality (1493) are the most common effects studied. Studying population may be of interest to see if the insecticide increases, decreases or has no effect on a population size over the course of the study. Studying mortality could help researchers understand the effectiveness of the insecticide as well as understand why population size may have changed.

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.

```
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: Top 6: Honey Bees, Parasitic Wasp, Buff Tailed Bumblebee,Carniolan Honey Bee, Bumble Bee, Italina Bee. Bees are the most studied because a healthy garden needs pollinators. Researchers want to make sure that these insects are not harmed by the insecticide. The Asian Lady Beetle is also frequently studied because it is beneficial to a farm by preying on harmful insects. The insecticide should not harm this species of insect either. On the other hand, the Japanese Beetle is studied frequently as they are a pest. The insecticide should be effective at reducing this insect population.

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.) #it's a factor
```

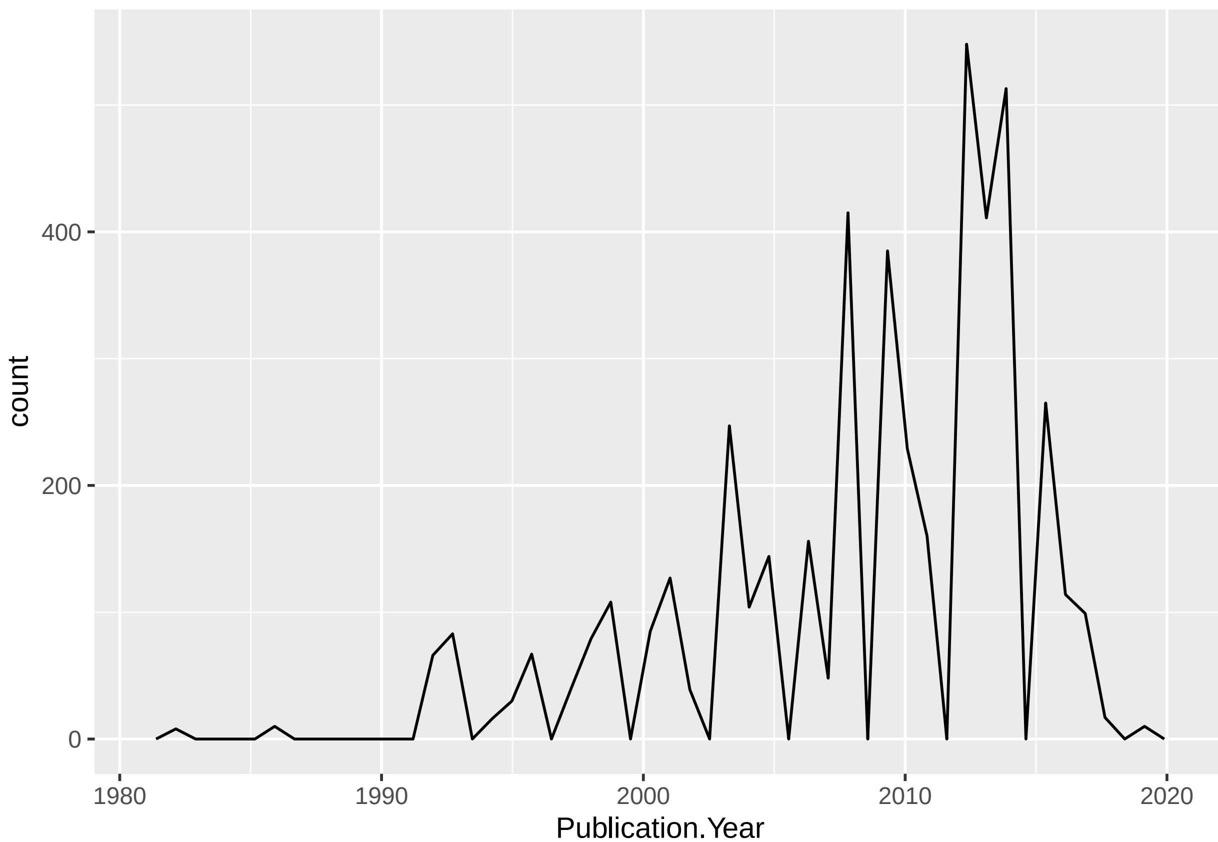
```
## [1] "factor"
```

Answer: It is not numeric because it contains both numbers and characters.

Explore your data graphically (Neonics)

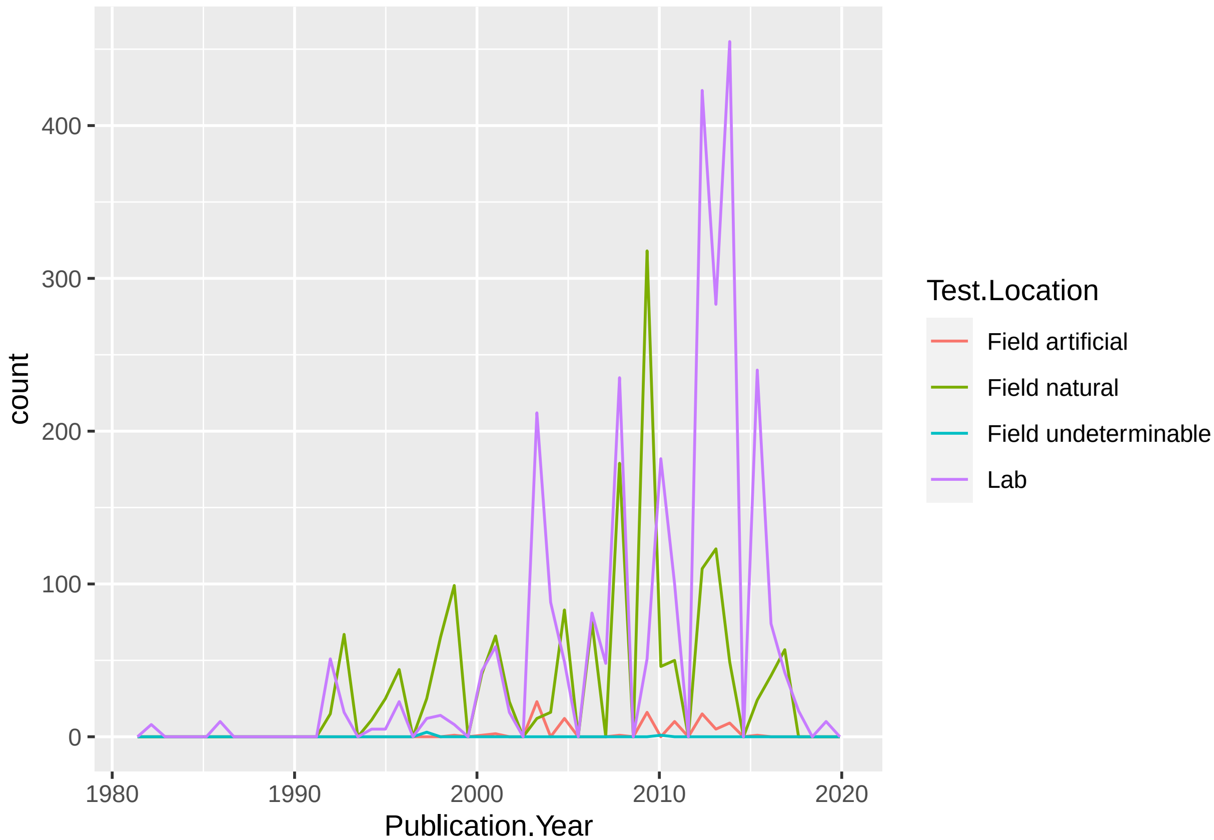
9. Using `geom_freqpoly`, generate a plot of the number of studies conducted by publication year.

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



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 = 50)
```

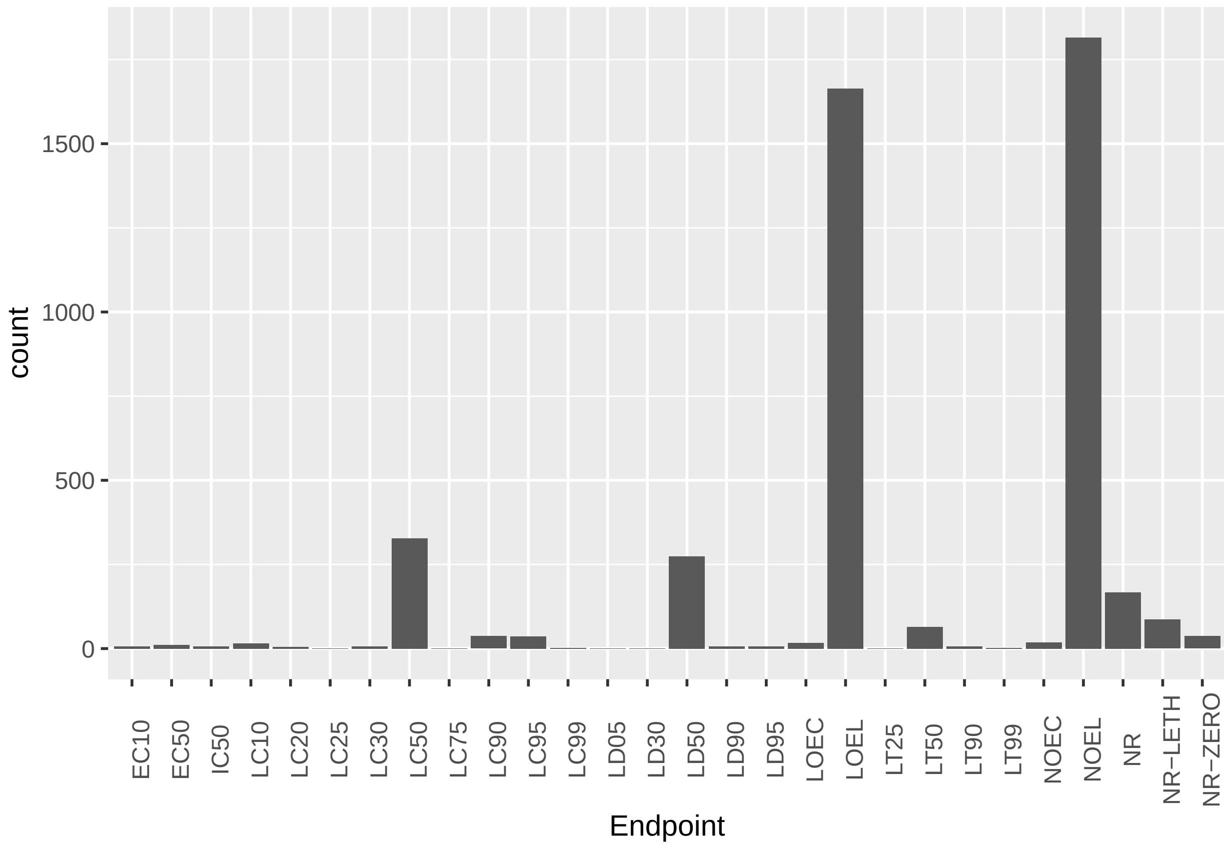


Interpret this graph. What are the most common test locations, and do they differ over time?

> Answer: Labs and field natural test locations are most commonly used. There is variability over time, but it appears that these locations have been used more often since the first recorded publish date in this dataset.

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)) +
  geom_bar(aes(x = Endpoint)) +
  theme(axis.text.x = element_text(angle = 90)) # used this code to make x axis labels visible
```



Answer: LOEL is the "lowest-observable-effect-level," and it is defined as the lowest dose that would produce a significantly different response from the controls. NOEL is "no-observable-effect-level," and it is defined as the highest dose that does not cause a significantly different response from the 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) #it's a factor to start. Can run this code again to see if conversion worked
```

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

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

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

```
unique(Litter$collectDate) #samples were collected on August 2nd and 30th in 2018
```

```
## [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`?

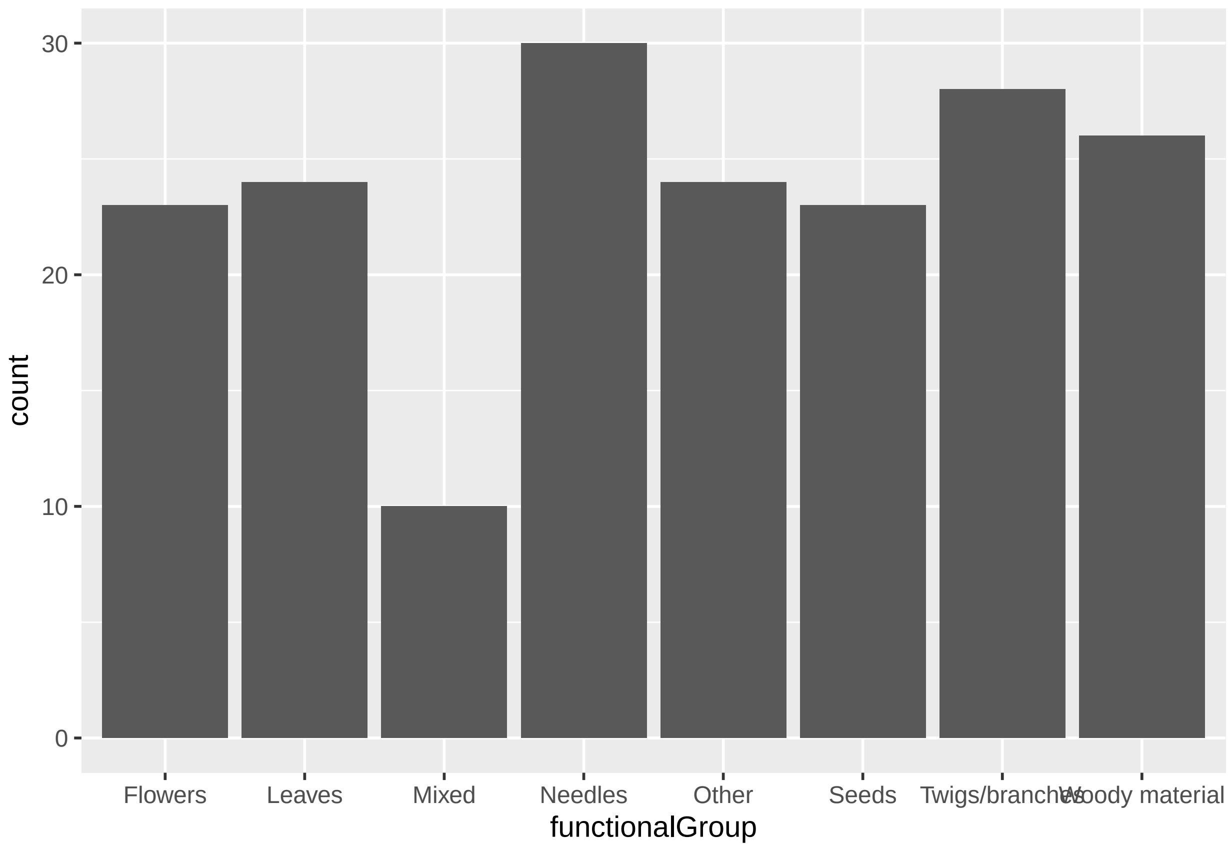
```
unique(Litter$plotID)
```

```
##  [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
```

Answer: The unique function lists every unique plot ID and gives the total of the different sites which is 12 in this case. We can infer the same info from the summary function; however, this function does not explicitly state how many unique plots there are. It's main function is to provide stats on each plot.

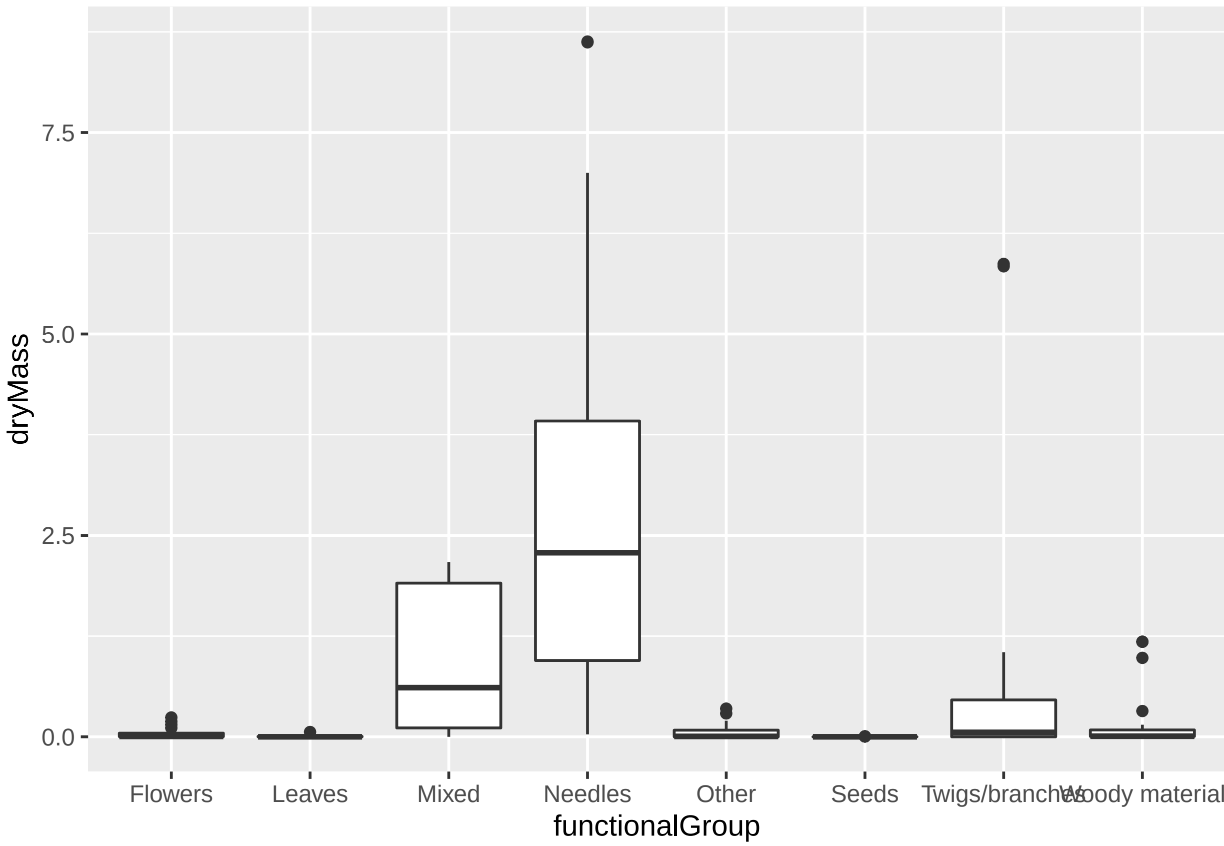
14. Create a bar graph of functionalGroup counts. This shows you what type of litter is collected at the Niwot Ridge sites. Notice that litter types are fairly equally distributed across the Niwot Ridge sites.

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

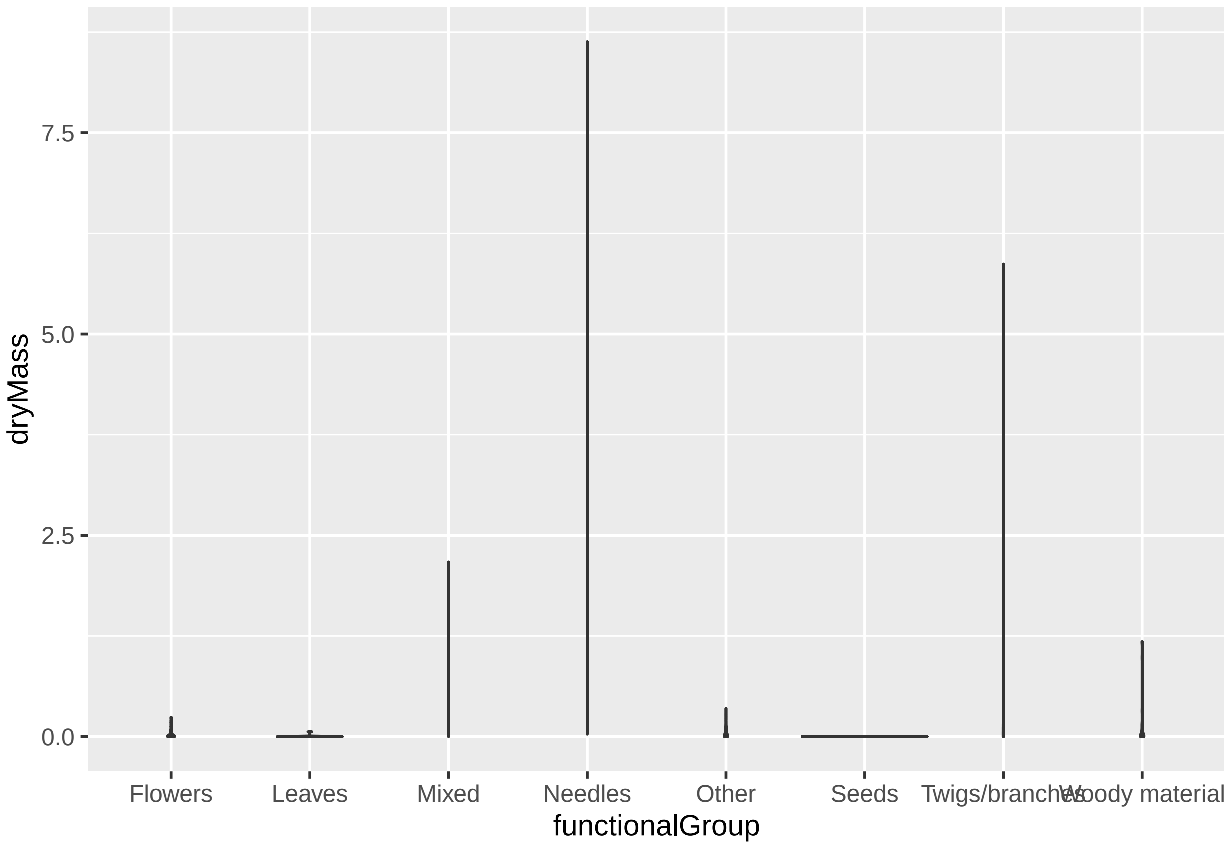


15. Using `geom_boxplot` and `geom_violin`, create a boxplot and a violin plot of dryMass by functionalGroup.

```
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: The boxplot helps us visualize the range of drymass between functionalgroups. We can clearly see the calculated averages between the samples. However, drymass accumulation in all functionalGroups varies between each sample. The violin is good for visually communicating if there is a certain range with many data points. This isn't the case in this chart, so the violin becomes less informative than the boxplot in this instance as a result.

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

> Answer: Needles