Assignment 3: Data Exploration

Atalie Fischer, 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 <January 31, 2021 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] "/Users/ataliefischer/Desktop/EDA/Environmental_Data_Analytics_2022/Assignments"
```

```
setwd("/Users/ataliefischer/Desktop/EDA/Environmental_Data_Analytics_2022/")

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: There are many reasons why we would use data on the ecotoxicology of neonicotinoids on insects. One reason this data are useful is to determine the effectiveness of the insecticide following application. The Neonics data show the insecticide dose that caused an effect in various species, as well as information about how the organism was exposed. The data also include the source of where the data were found, which is useful in determining the credibility of data.

Another reason why the ecotoxicology of neonicotinoids of insects is to determine whether the insecticide is species specific and only affecting the target species, or if there may be effects on non-target species.

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: Studying litter and woody debris that falls to the ground in forests is useful for understanding the net annual primary productivity of the forest, which is a metric for forest health. It can also inform nutrient availability, which can inform the types of organisms that inhabit the soil. These metrics are useful in studying the health and resilience of the ecosystem.

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 and woody debris are sampled using elevated and ground traps, respectively. * Sampling only occurs at plots with vegetation greater than 2m tall. * Ground traps are sampled once per year. Elevated traps are sampled every two weeks in deciduous forests and every month in evergreen forests. * Sampling sites are distributed using plot-subplot-nested subplot layouts.

Obtain basic summaries of your data (Neonics)

5. What are the dimensions of the dataset?

```
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 effects that are studied are mortality and population effects. Mortality is of interest because it shows the most extreme toxicological endpoint. Population effects are of interest in studying the long-term toxicological effects within the ecosystem.

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: The six most commonly studied speices in the Neonics dataset are Honey Bee (667), Parasitic Wasp (285), Buff Tailed Bumblebee (183), Carniolan Honey Bee (152), Bumble Bee (140), and Italian Honeybee (113). These are all bees, which are important for ecosystem functioning and are also sensive species to toxins.

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.)
```

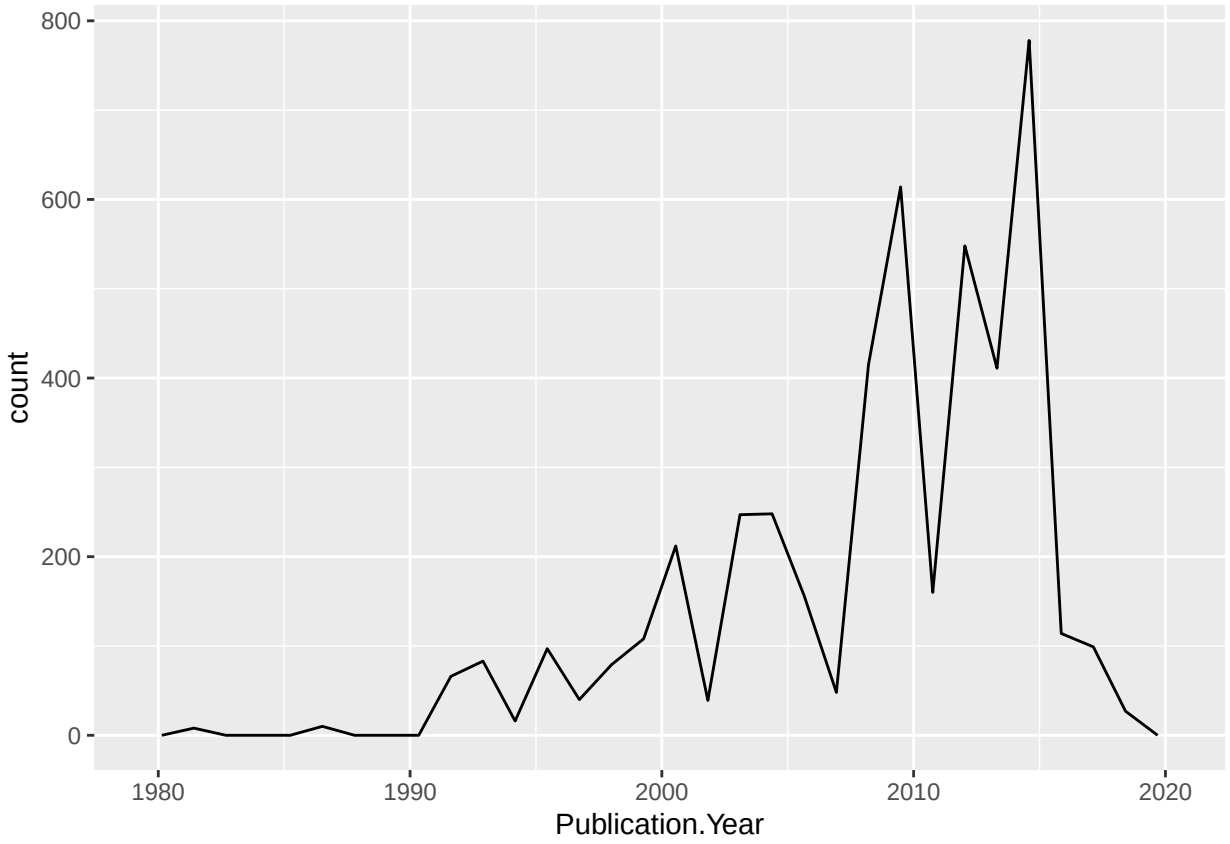
```
## [1] "factor"
```

Answer: The class of the concentration column in the Neonics dataset is "factor" because the row entries include both numeric and character values. For example, some entires were not reported and are logged as "NR" instead of a numeric concentration.

Explore your data graphically (Neonics)

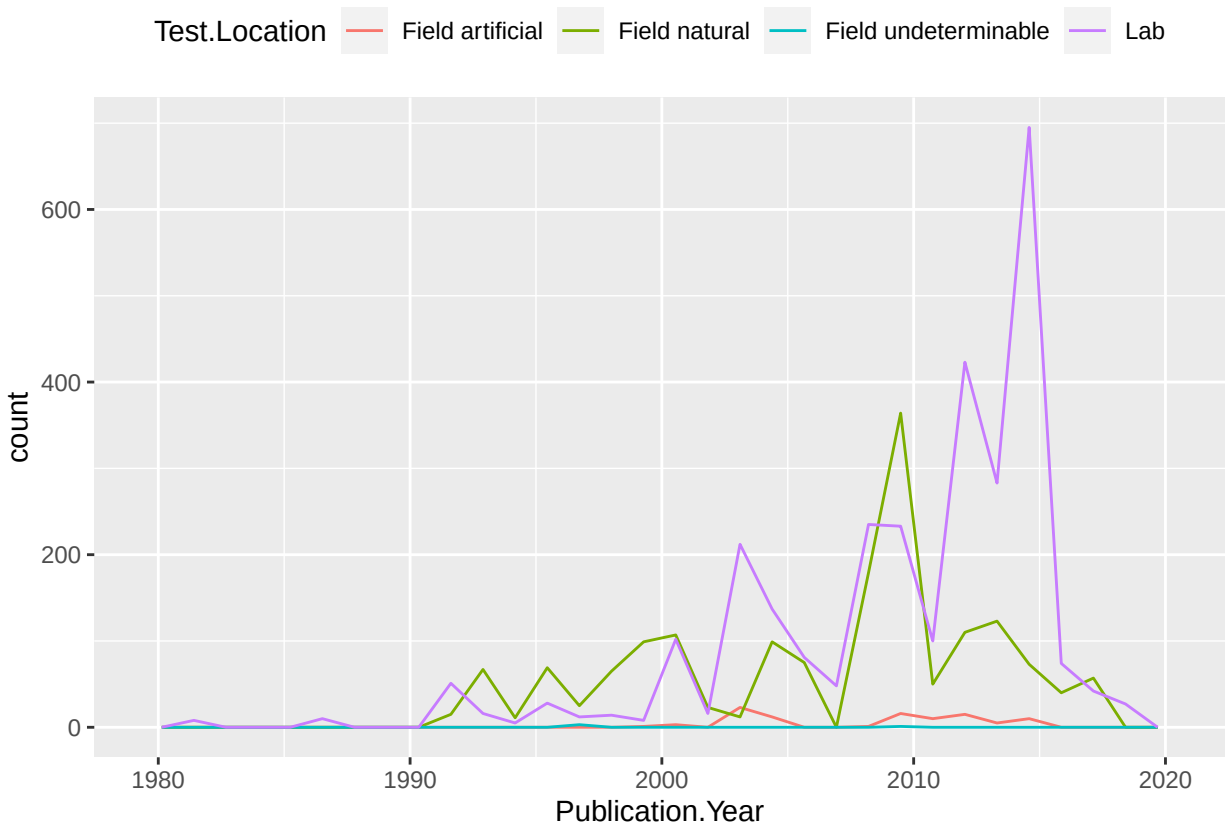
9. Using `geom_freqpoly`, generate a plot of the number of studies conducted by publication year.

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



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

```
ggplot(Neonics, aes(x = Publication.Year, color = Test.Location)) +
  geom_freqpoly() +
  theme(legend.position = "top")
```

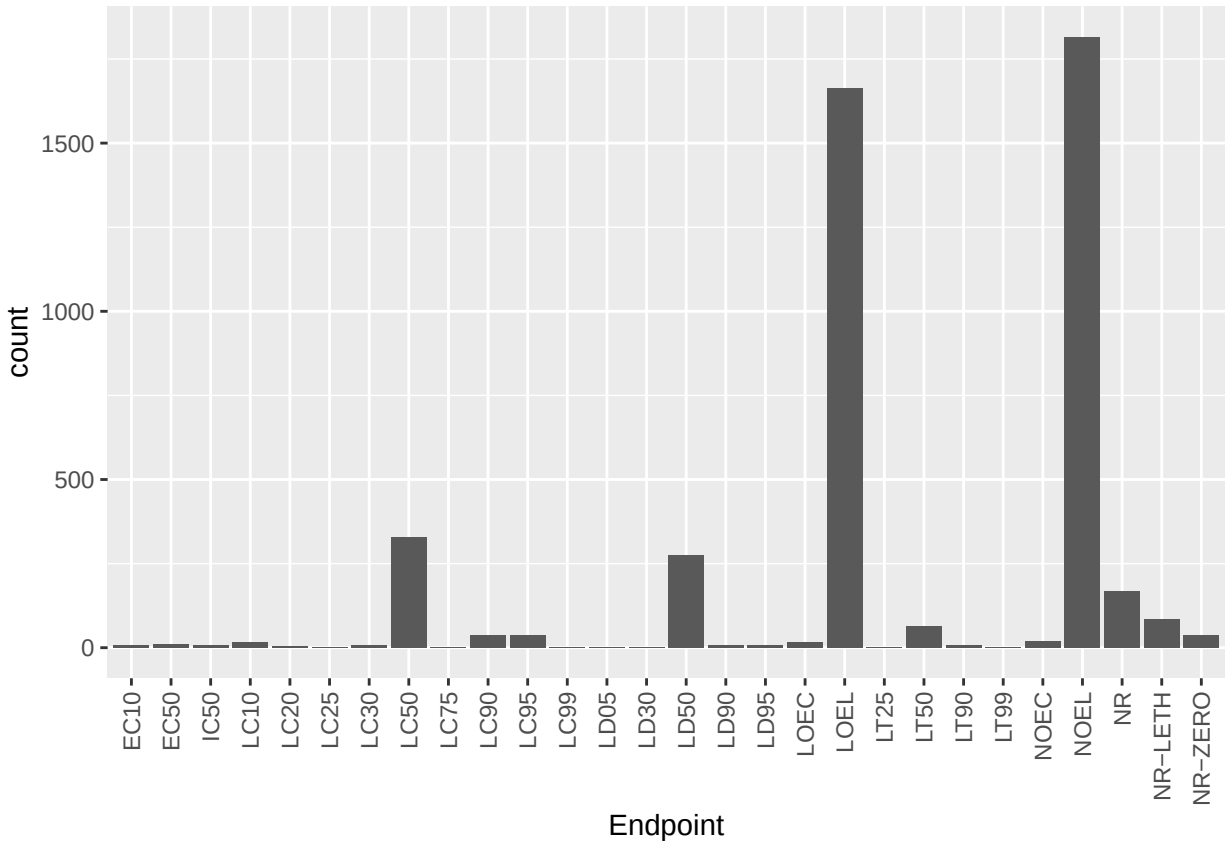


Interpret this graph. What are the most common test locations, and do they differ over time?

Answer: Laboratory and natural field studies are the most common test locations. Laboratory studies are increasing over time, while natural field studies have been declining since around 2010.

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(angle = 90, vjust = 0.5, hjust=1))
```



Answer: The two most common end points are LOEL (lowest observed effect level) and NOEL (no observed effect level). LOELs are the lowest dose concentration at which an effect is observed. NOELs are the highest dose concentration at which no effect is observed.

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"
```

```
collectDate_Aug18 <- unique(Litter$collectDate)
print(collectDate_Aug18)
```

```
## [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`?

```
NiwotRidge_plots <- unique(Litter$plotID)
print(NiwotRidge_plots)
```

```
##  [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
```

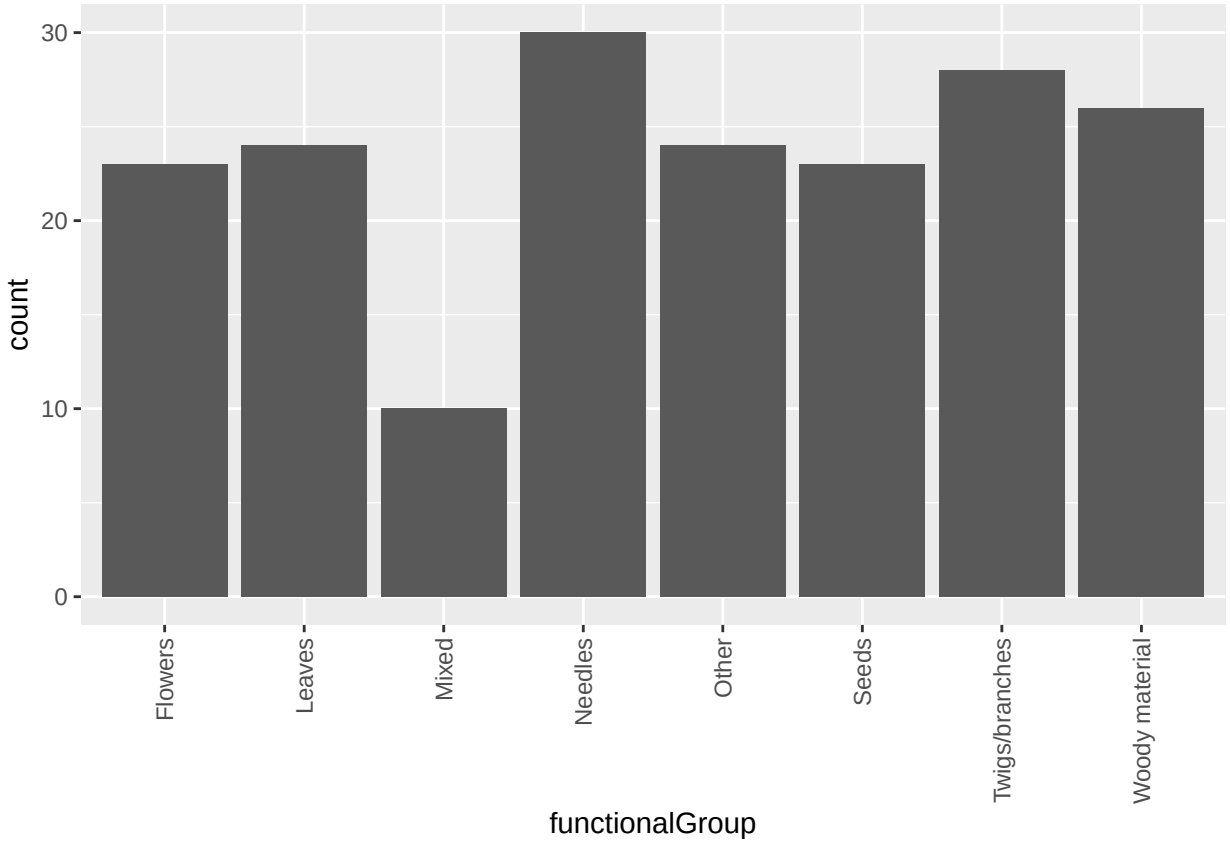
```
summary(Litter$plotID)
```

```
## 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: The "unique" function gives the number of unique entries in that column by displaying the number of levels within the vector. There are 12 plots sampled at Niwot Ridge. The "summary" function gives the number of samples taken at each plot.

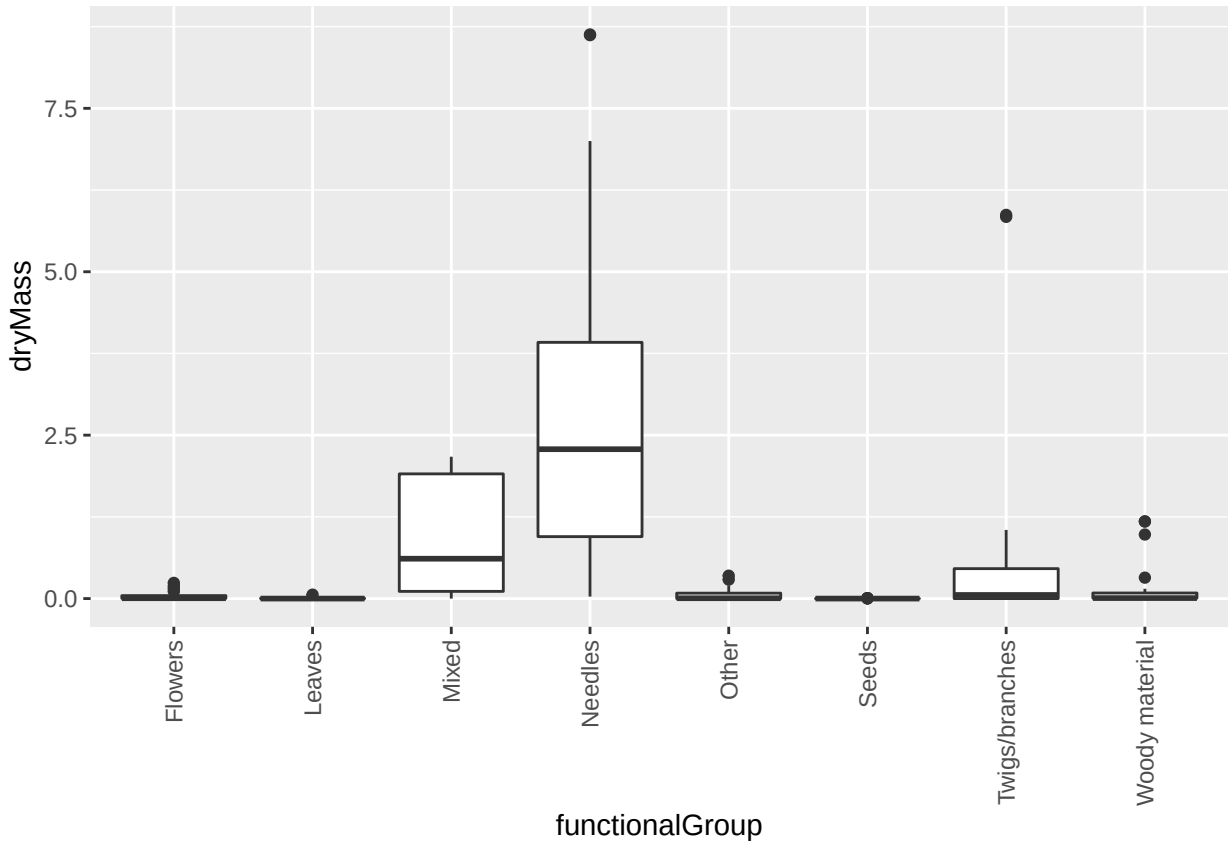
14. Create a bar graph of functionalGroup counts. This shows you what type of litter is collected at the Niwot Ridge sites. Notice that litter types are fairly equally distributed across the Niwot Ridge sites.

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

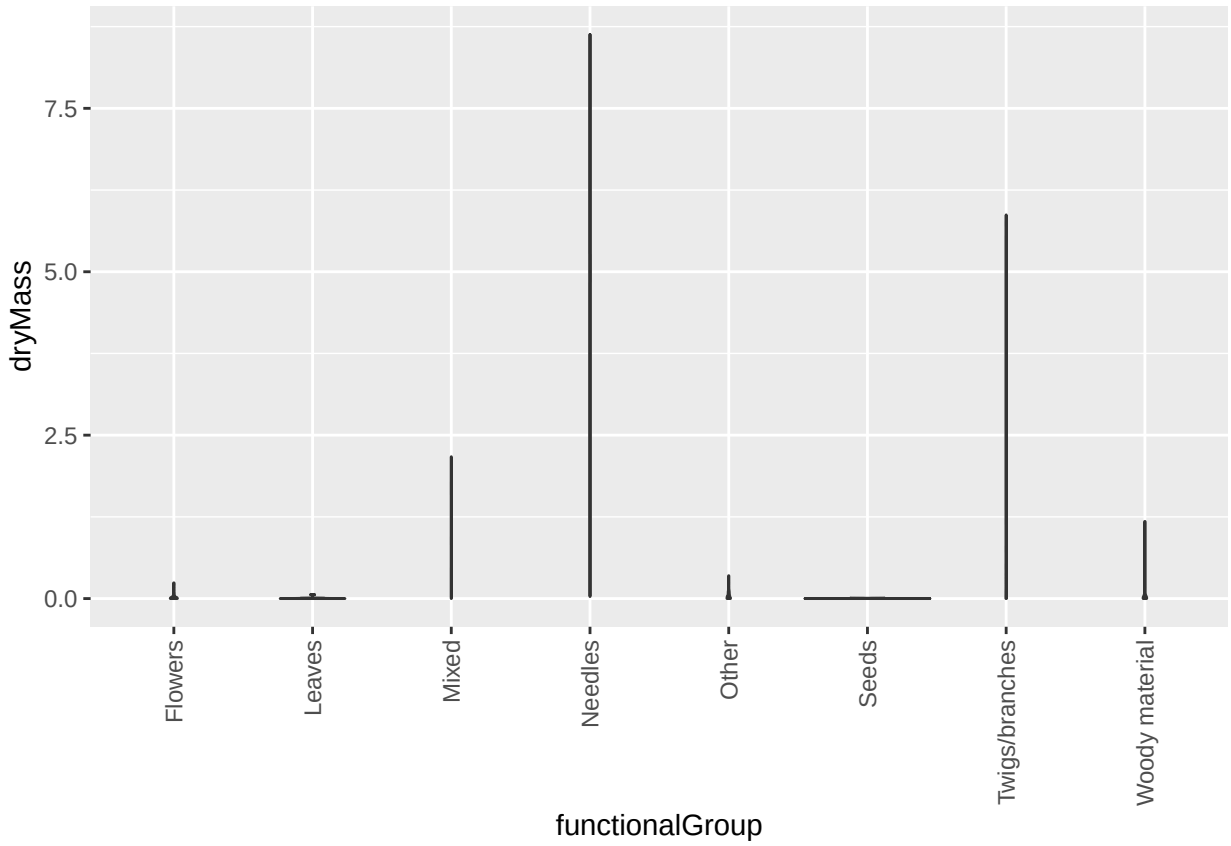



```
ggplot(Litter) +
  geom_violin(aes(x=functionalGroup, y=dryMass),
              draw_quantiles = c(0.25, 0.5, 0.75)) +
  theme(axis.text.x = element_text(angle = 90, vjust = 0.5, hjust=1))
```

```
## 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: A violin plot includes the frequency of each observation. However, in this case, there are few observations for each value, so each functional group appears as a straight line with no clear distinction between the quartiles. Thus, the boxplot is a more effective visualization option because the quartiles and outliers are clearly distinguished.

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.