Assignment 3: Data Exploration

Clara Fast, Section #3

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.**

```
#Check working directory
getwd()
```

```
## [1] "/Users/clara/Desktop/DATAN/Environmental_Data_Analytics_2022"
```

```
#Install necessary packages
require("tidyverse")

#Upload necessary datasets
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 are interested in the ecotoxicology of neonicotinoids on insects to evaluate the ramifications of using this to the environment. The insectiside may do a lot more harm than good to surrounding biodiversity. For example, the insectiside may be absorbed by plants and become present in pollen and nectar which consequently poisons feeding bees and other pollen vectors. It is also necessary to evaluate whether these insectisides are directly or indirectly harmful to humans.

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: There are multiple reasons why we may be interested in studying litter and woody debris on forest grounds. Firstly, it is one way to get more information on how and at what rate nutrients are cycled in an ecosystem. Secondly, as woody debris and forest litter increases fuel load in many cases, it is important to keep track of their abundance and other relevant patterns. Thirdly, they may serve as indicators of the health of the forest. Additionally this data may simply reveal what species are present and abundant in the forest canopy.

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 fine woody debris sampled using tower plots, which are selected randomly within a 90% flux footprint of primary and secondary airsheds. * In sites with forested tower airsheds, litter sampling takes place in 20 40mx40m tower plots. * In sites with low-statured vegetation over the tower airsheds, litter sampling takes place in 4 40mx40m and 26 20mx20m tower plots.

Obtain basic summaries of your data (Neonics)

5. What are the dimensions of the dataset?

```
print(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 studied are population (1803), mortality (1493), and behavior (360). These are of interest to determine whether the insectiside has altered population sizes of insects, in other words, to evaluate whether the insectiside harms insects, and/or alters their feeding patterns.

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 species are: Honey bee; Parasitic wasp; Buff tailed bumblebee; Carniolan honey bee; Bumblee bee; and the Italian honeybee. These are all pollinator species/vectors of pollen. They may be of interest because they determine plant yield and consequently directly impact human food abundance.

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

```r
print(class(Neonics$Conc.1..Author.))
```

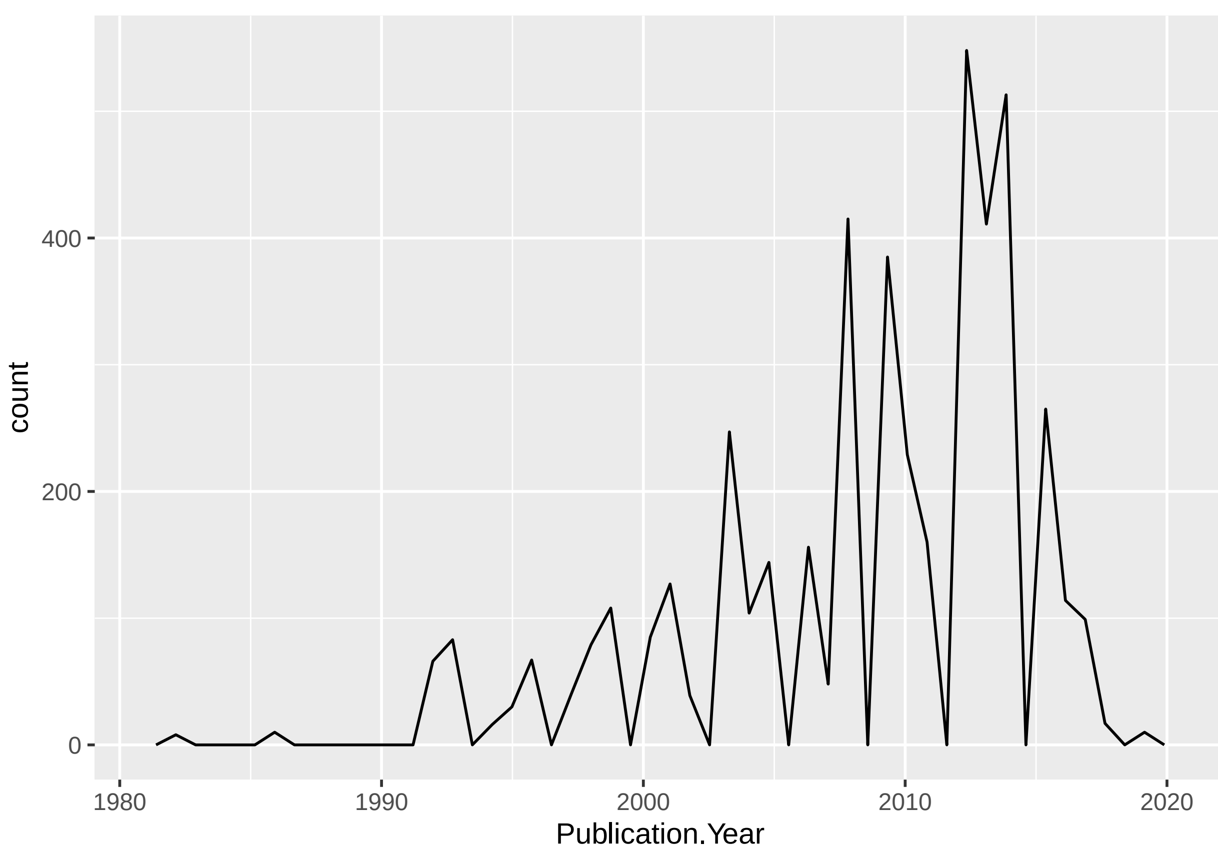
```
## [1] "factor"
```

Answer: The class of Conc.1..Author is factor. It is not numeric because there are characters present in the data.

Explore your data graphically (Neonics)

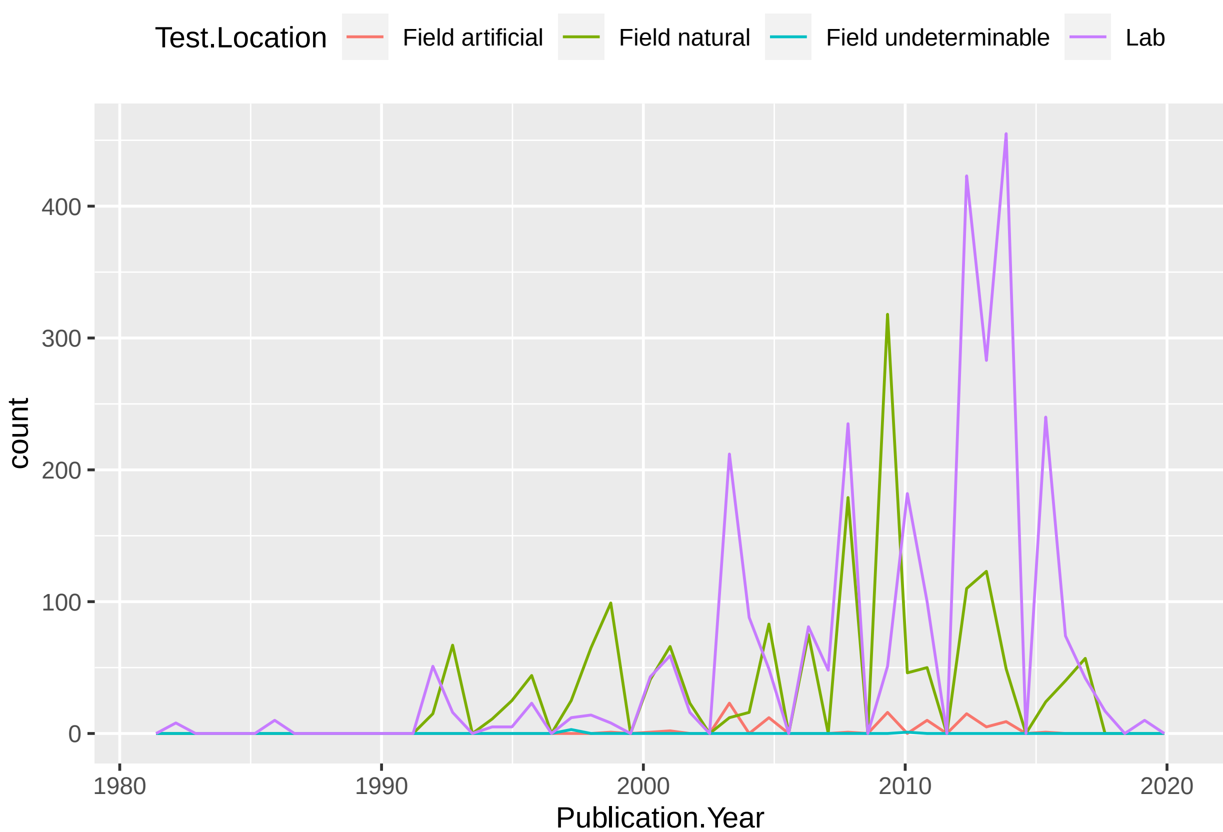
9. Using `geom_freqpoly`, generate a plot of the number of studies conducted by publication year.

```r
#Generate plot of 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.

```
#Generate previous plot and add test location variable
ggplot(Neonics) +
  geom_freqpoly(aes(x = Publication.Year, color = Test.Location), bins = 50) +
  theme(legend.position = "top")
```

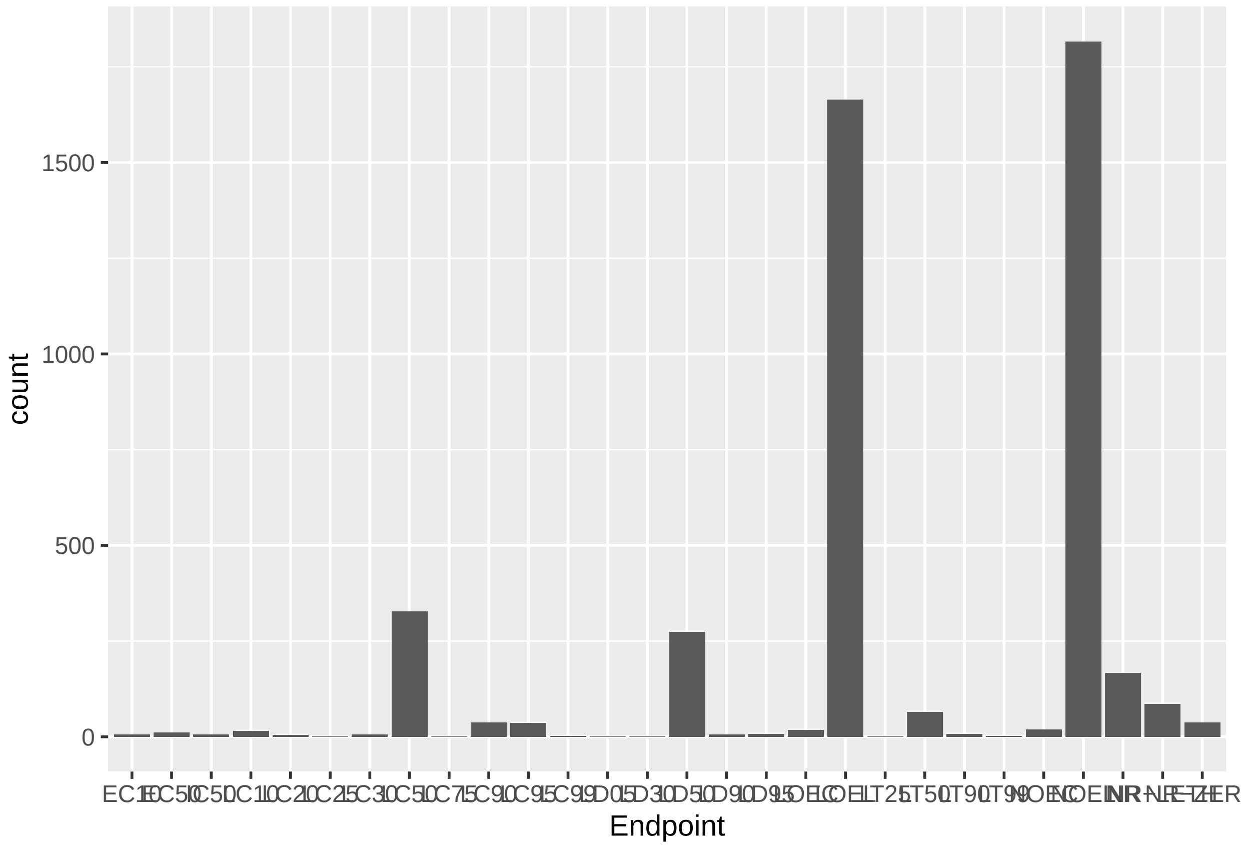


Interpret this graph. What are the most common test locations, and do they differ over time?

> Answer: The most common test locations are the lab and the natural field. Previously the natural field and the lab test locations alternated in popularity, but more recently, the lab appears to be the most common location to conduct the tests.

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.

```
#Generate bar graph of Endpoint counts
ggplot(Neonics, aes(x = Endpoint)) +
  geom_bar()
```



Answer: The two most common end points are LOEL and NOEL. LOEL: Lowest-observable-effect-level; It is the lowest dose (concentration) producing effects that were significantly different from responses of controls. NOEL: No-observable-effect-level; It is the highest dose (concentration) producing effects not significantly different from responses of controls according to the 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.

```
#Determine class of collectDate
print(class(Litter$collectDate))
```

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

```
#Change class to date
Litter$collectDate<-as.Date(Litter$collectDate)
print(class(Litter$collectDate))
```

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

```r
#Determine which dates litter was sampled in August 2018
print(unique(Litter$collectDate, incomparables = FALSE))
```

```
## [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`?

```r
#Determine how many plots sampled at Niwot Ridge
print(unique(Litter$plotID, incomparables = FALSE))
```

```
##  [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
```

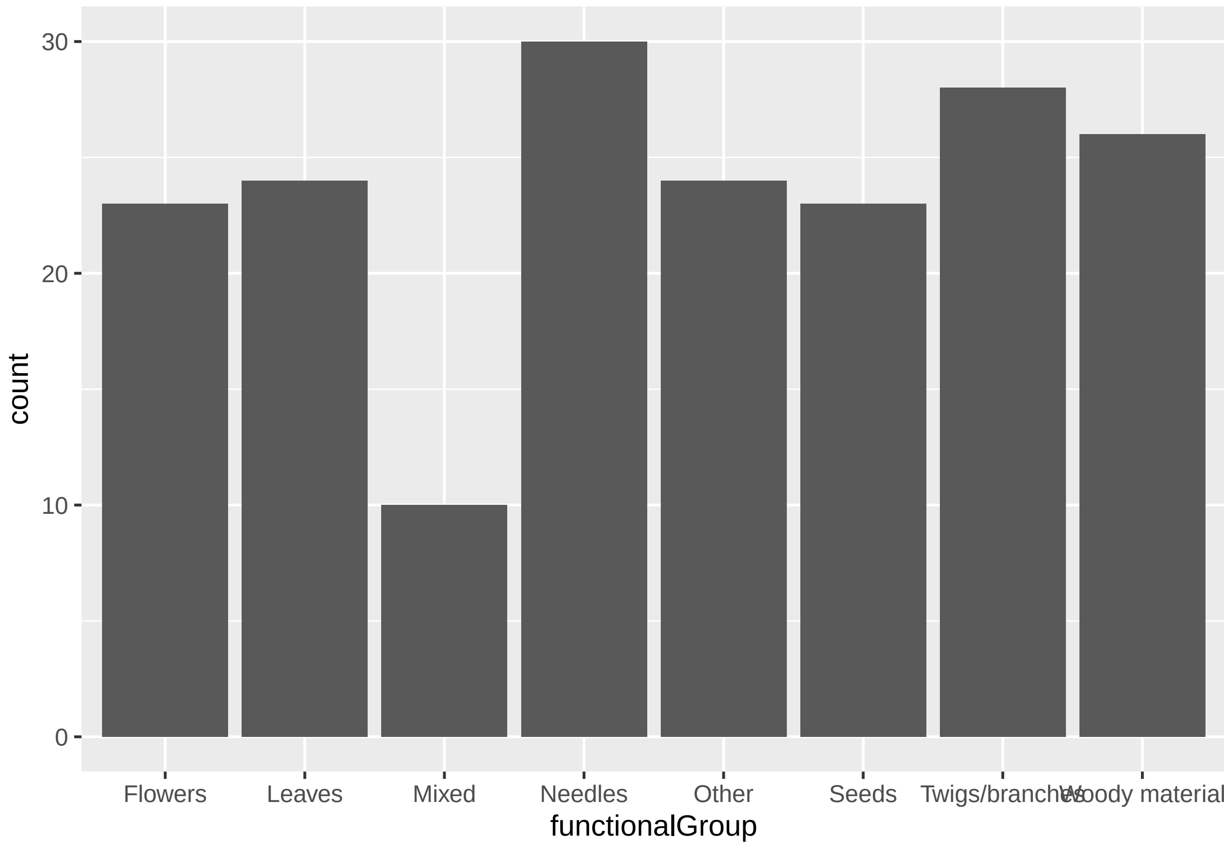
```r
summary(Litter$plotID)
```

```
## 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: 12 plots were sampled at Niwot Ridge. Summary gives an overview of the plot ID data whereas the unique function allows you to identify and return a specific vector, data frame, or array without including duplicates. The unique function explicitly identifies that number of plots, but

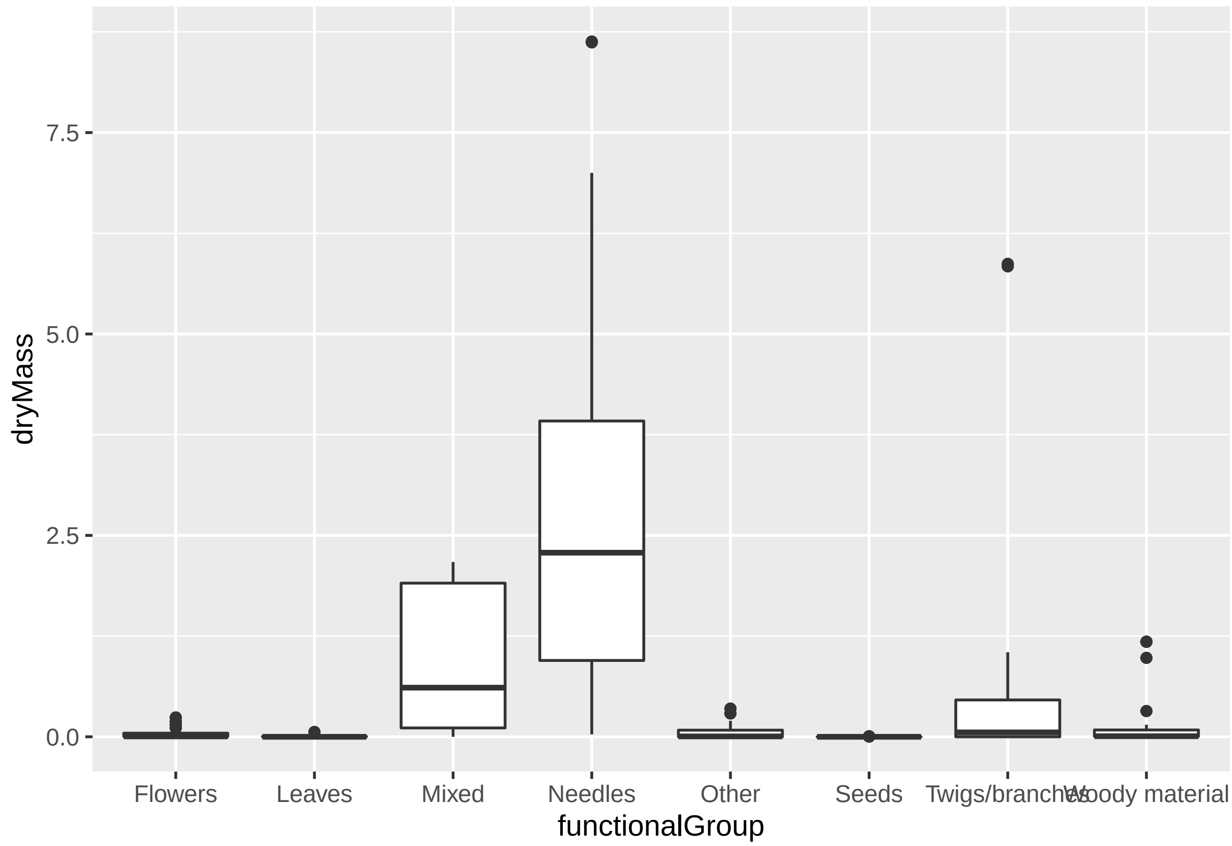
14. Create a bar graph of functionalGroup counts. This shows you what type of litter is collected at the Niwot Ridge sites. Notice that litter types are fairly equally distributed across the Niwot Ridge sites.

```r
#Generate bar graph of functionalGroup counts
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.

```
#Generate boxplot of dryMass by functionalGroup
ggplot(Litter) +
  geom_boxplot(aes(x = functionalGroup, y = dryMass))
```

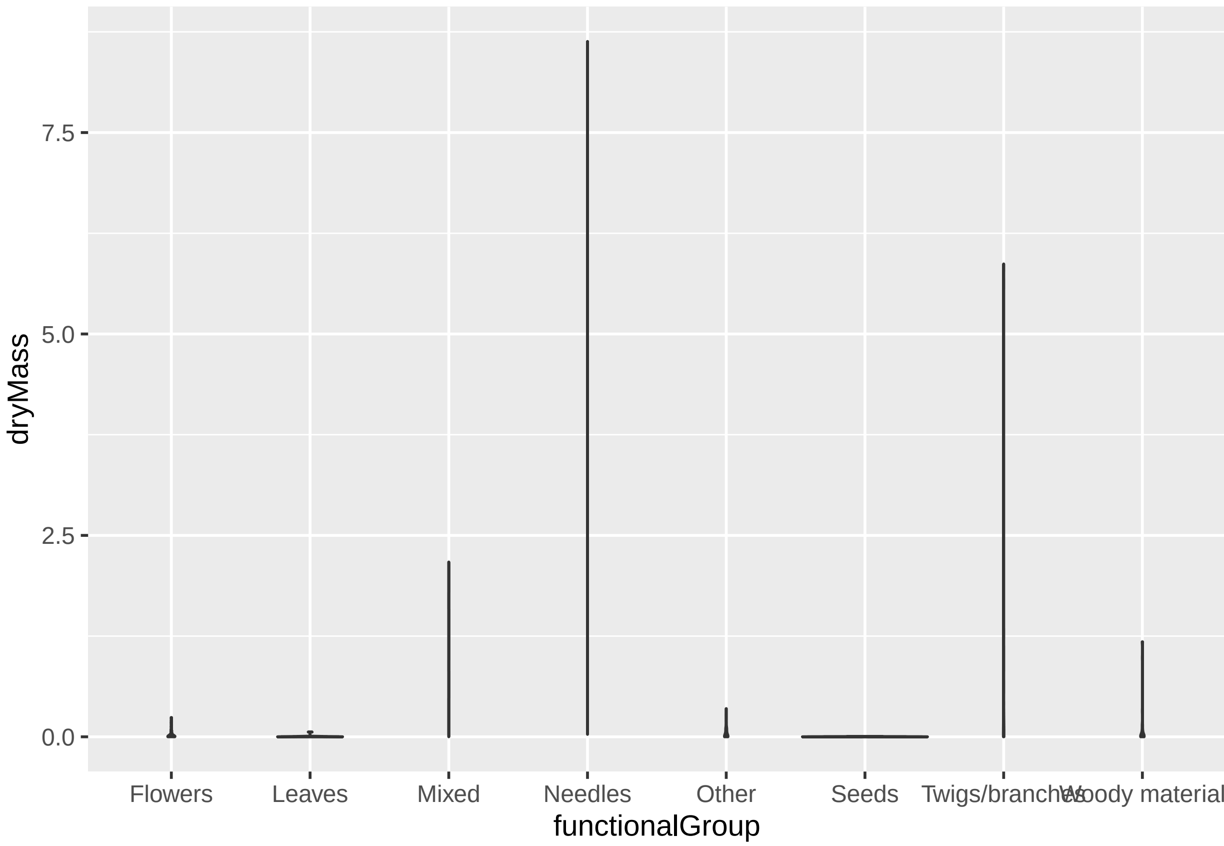



```
#Generate violin plot of dryMass by functionalGroup
ggplot(Litter) +
  geom_violin(aes(x = functionalGroup, y = dryMass),
              draw_quantiles = c(0.25, 0.5, 0.75))
```

```
## 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 violin plot is not an effective visualization option in this case because it does not depict the spread of the data.

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

> Answer: Needles tends to have the highest biomass at these sites, as evident from the boxplot.