Assignment 3: Data Exploration

Marisa Fajardo, 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.**

```
insects<- read.csv("../Data/Raw/ECOTOX_Neonicotinoids_Insects_raw.csv", stringsAsFactors = TRUE) #creat
litter<- read.csv("../Data/Raw/NEON_NIWO_Litter_massdata_2018-08_raw.csv", stringsAsFactors = TRUE) #cr
```

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 might want to know how effective these insecticides are for certain insect species, and we would want to know how lethal these toxins are for beneficial insect species such as pollinators. We would also want to know if insects are gaining resistants to insecticides, thus making them useless.

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: Wood litter and debris on the forest floor adds a great amount of nutrients to the soil and aids in keeping in moisture. Woody debris also provides habitat for small organisms. Studies on the amount of biomass on forest floors can be indicative of the nutrients in a forest ecosystem and biodiversity within 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: *Litter is collected in plots of varying sizes based on canopy cover* Litter trap placement in plots are be randomized in plots with >50% woody vegetation cover *in sites with <50% woody vegetation cover, litter trap placement will be targeted so that traps are placed underneath vegetation to catch litter

Obtain basic summaries of your data (Neonics)

5. What are the dimensions of the dataset?

```
dim(insects) #checking dimensions of insects
```

```
## [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(insects$Effect) #viewing summary of Effect column
```

```
##     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 population, followed by mortality, then behavior. Population would be of importance since we would want to identify species in an area. Mortality is also important to study, since it will show if insect death occurs in response to insecticides. Behavior is also important since it details responses of insects to the insecticides.

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(insects$Species.Common.Name) #viewing summary of species common name column
```

```
##                   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 6 most common insects are: Honey bee(667), parasitic wasp(285), buff tailed honeybee(183), Carniolan Honey Bee(152), Italian honeybee(113). These species are all pollinators. They may be of interest since they are essential to polinating plants and crops, therefore we would not want them to be affected by insecticides.

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(insects$Conc.1..Author.) #viewing class of conc 1 author column
```

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

Answer: The class is character. This is not numeric because the data is categorical.

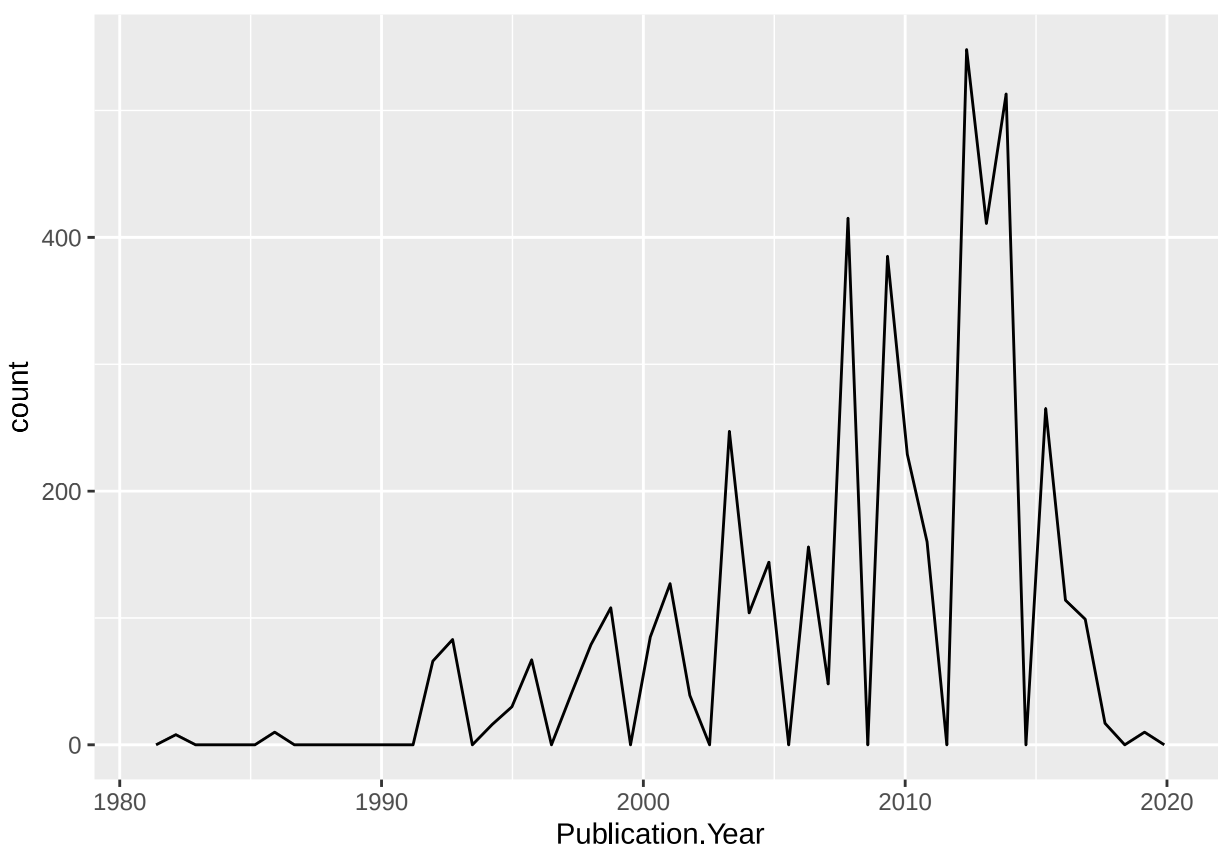
Explore your data graphically (Neonics)

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

```
library(ggplot2)
```

```
## Warning in register(): Can't find generic `scale_type` in package ggplot2 to
## register S3 method.
```

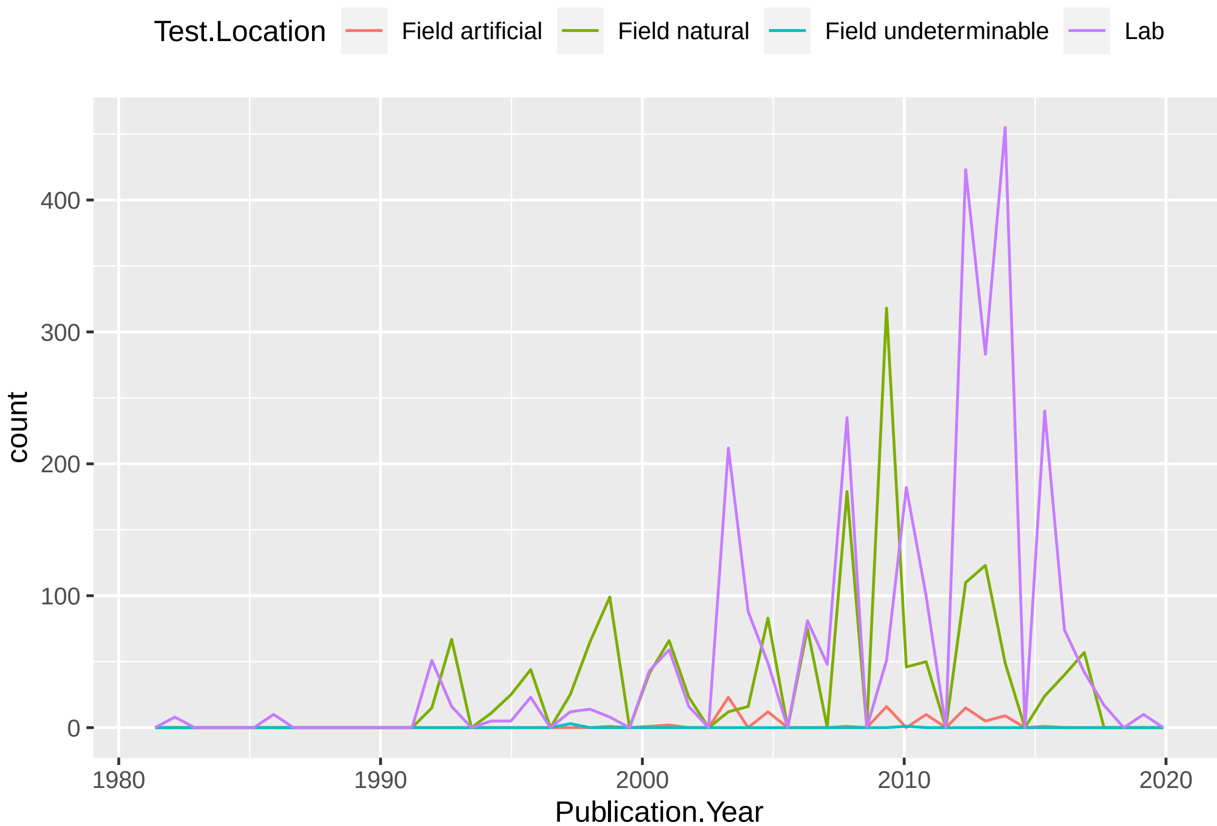
```
ggplot(insects) +
  geom_freqpoly(aes(x = Publication.Year), bins = 50) #creating simple line plot of number of studies p
```



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

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

  theme(legend.position = "top") #creating line plot that color codes test locations from the data
```

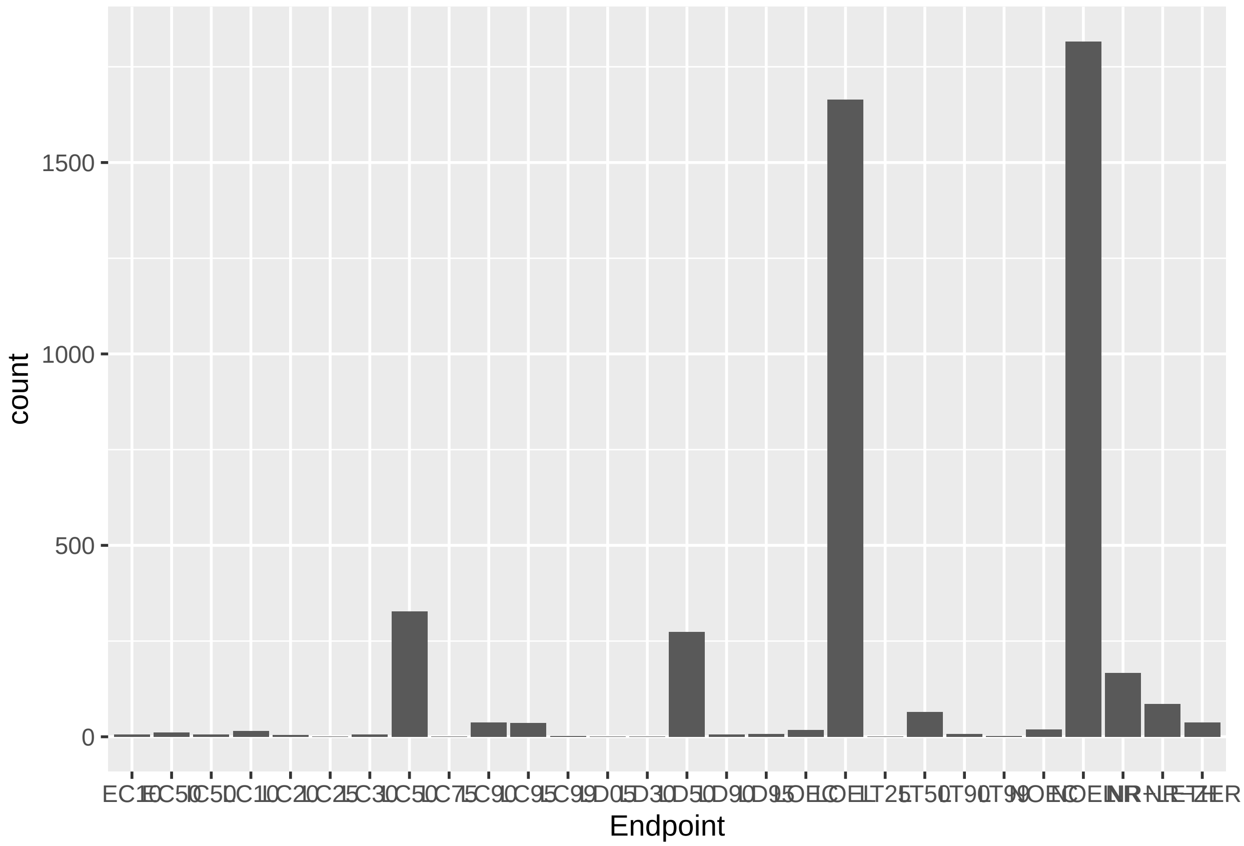


Interpret this graph. What are the most common test locations, and do they differ over time?

> Answer: The most common test locations are in the lab. These test locations frequencies ebb and flow, but overall increase over time, peaking at around 2014, then lowering again towards 2020.

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.

```r
ggplot(insects, aes(x = Endpoint)) +
  geom_bar() #creating bar graph to view counts of each endpoint type
```




```r
which.max(table(insects$Endpoint)) #checking the most common endpoint
```

```
## NOEL
##   25
```

> Answer: NOEL is the most common endpoint. It is a no observable effect level that refers to the highest concentration that would produce effects that are not significantly different from the responses of the controls. The second most common endpoint is LOEL(i believe that is what the graph says, but it is difficult to read since the endpoint names are on top of each other). LOEL is a lowest observable effect level that refers to the lowest concentration that that produces effects that are significantly different from the responses of 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.

```r
class(litter$collectDate) #viewing class of collection date
```

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

```r
View(litter$collectDate)
```

```r
litter$collectDate <- as.Date(litter$collectDate) #making this data recognizable as a date
class(litter$collectDate) #checking class of data
```

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

```r
unique(litter$collectDate)#checking collection dates
```

```
## [1] "2018-08-02" "2018-08-30"
```

collect Date is classified as a character, not a date, so I changed it to a 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`?

```r
unique(litter$plotID) #Looking at counts under plot id column
```

```
##  [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
```

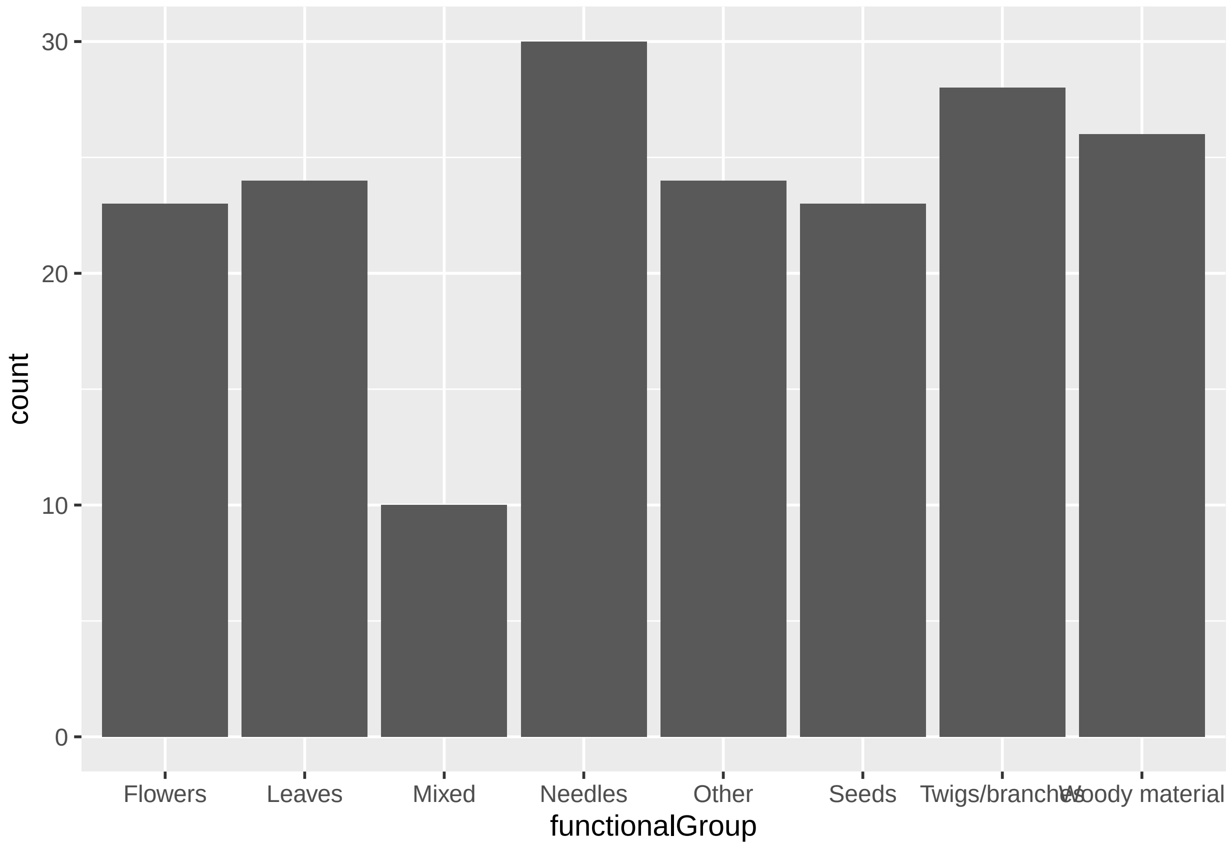
```r
summary(litter$plotID) #comparing to summary
```

```
## 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. The summary function gives an overview of the data frame, while the unique function allows you to identify specifics from the data frame.

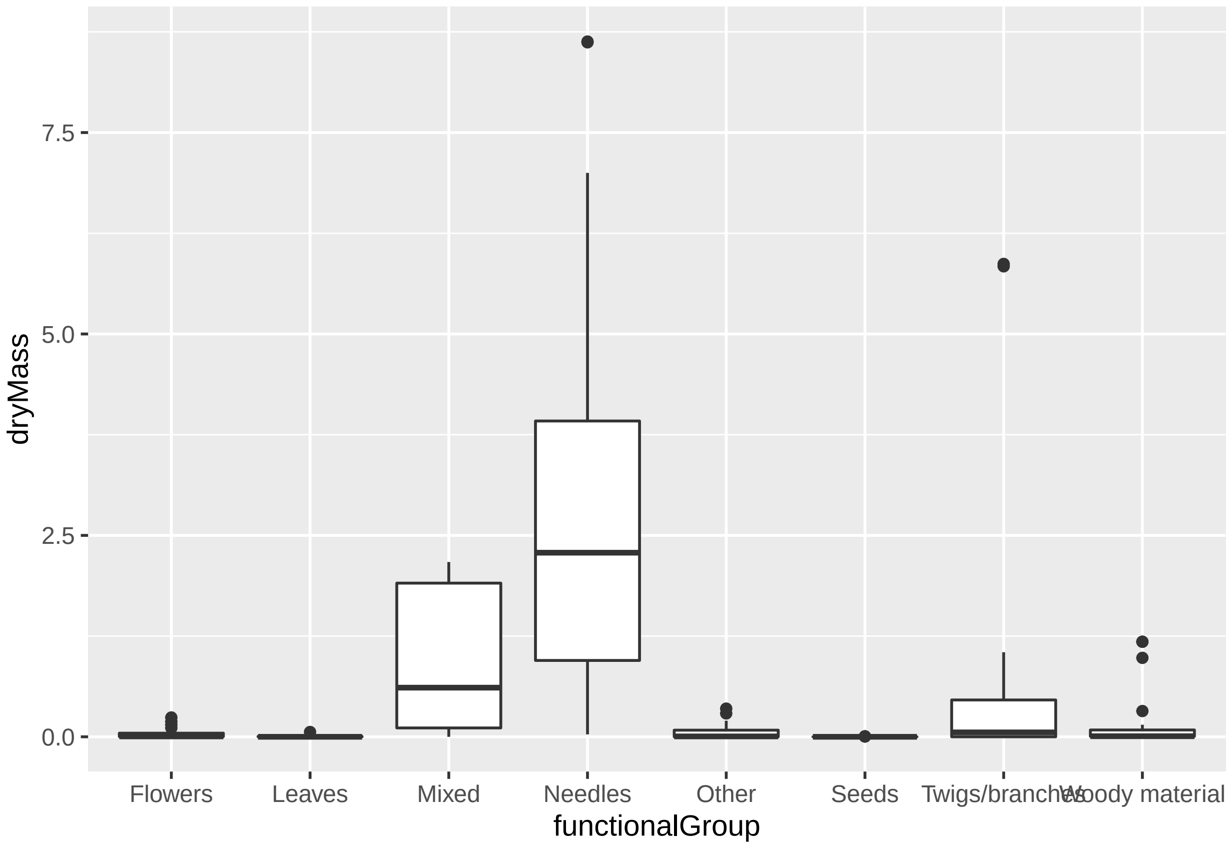
14. Create a bar graph of functionalGroup counts. This shows you what type of litter is collected at the Niwot Ridge sites. Notice that litter types are fairly equally distributed across the Niwot Ridge sites.

```r
ggplot(litter, aes(x = functionalGroup)) +
  geom_bar() #creating bar graph to view counts of each functional group type
```



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)) #creating box plot of drymass and functional grou
```

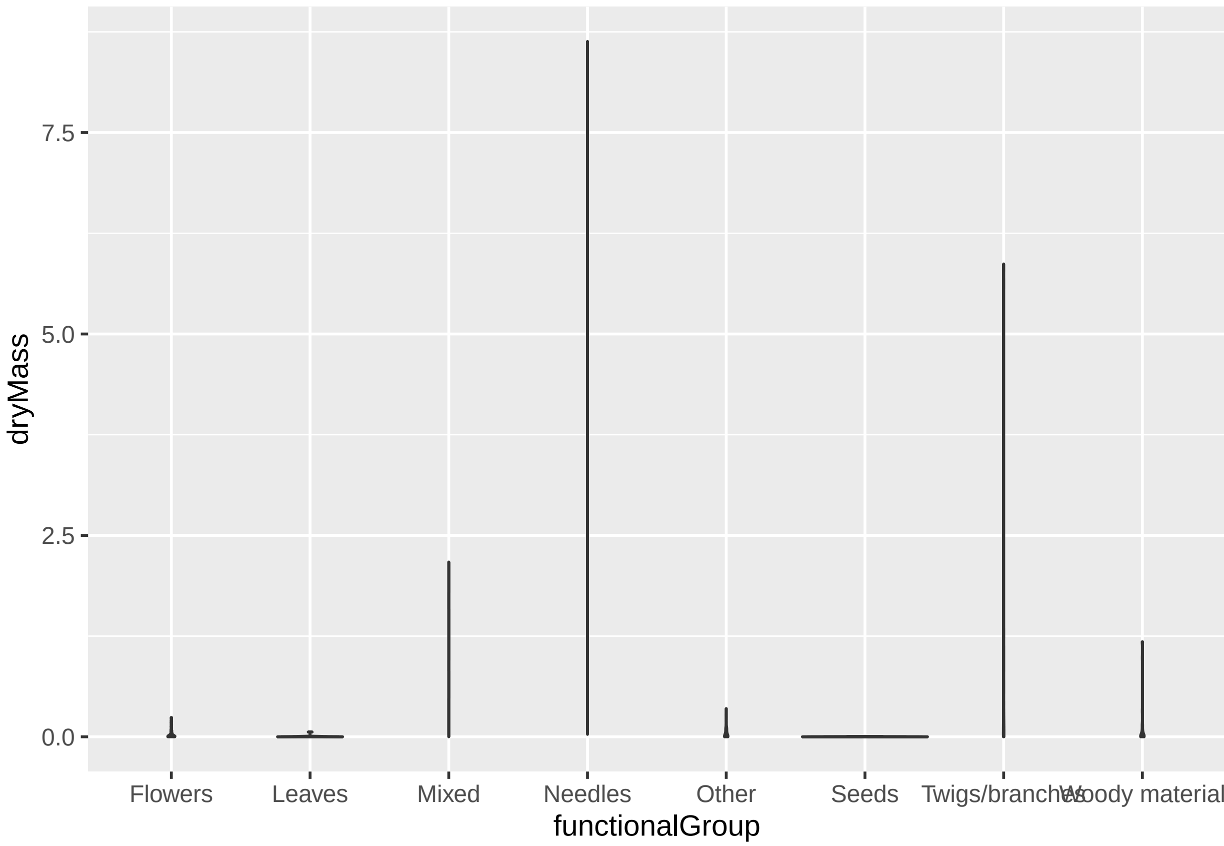



```
ggplot(litter) +
  geom_violin(aes(x = functionalGroup, y = dryMass),
              draw_quantiles = c(0.25, 0.5, 0.75)) #creating violin plot of drymass and functional grou
```

```
## 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: In the violin plot, the range of dryMass is too big, which causes the visualization of the functional group to become squished, making it difficult to read the data. The box plots are much easier to read and are a better data visualization in this case.

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

> Answer: Needles seem to have the highest biomass.