Assignment 3: Data Exploration

Megan McClaugherty

OVERVIEW

This exercise accompanies the lessons in Environmental Data Analytics on Data Exploration.

Directions

1. Rename this file `<FirstLast>_A03_DataExploration.Rmd` (replacing `<FirstLast>` with your first and last name).
2. Change "Student Name" on line 3 (above) with your name.
3. Work through the steps, **creating code and output** that fulfill each instruction.
4. Be sure to **answer the questions** in this assignment document.
5. When you have completed the assignment, **Knit** the text and code into a single PDF file.
6. After Knitting, submit the completed exercise (PDF file) to the dropbox in Sakai.

The completed exercise is due on Sept 30th.

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 include the subcommand to read strings in as factors.

```
getwd()
```

```
## [1] "/home/guest/EDA-Fall2022/Assignments"
```

```
# library(tidyverse)
library(ggplot2)
Neonics <- read.csv("/home/guest/EDA-Fall2022/Data/Raw/ECOTOX_Neonicotinoids_Insects_raw.csv",
    stringsAsFactors = TRUE)
Litter <- read.csv("/home/guest/EDA-Fall2022/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 Knowledge

> Answer: Insects perform a wide variety of ecosystem services that directly benefit humans, particular

3.  The Niwot Ridge litter and woody debris dataset was collected from the National Ecological Observat

> Answer: Leaf litter and woody debris in forests provide nutrients for soils as they decompose, help r
```

4. How is litter and woody debris sampled as part of the NEON network? Read the NEON_Litterfall_UserGu

> Answer:
1.Samples are taken from sites with woody vegetation greater than 2m tall
2. The number and size of plots depends on the stature of the vegetation around the tower airshed. One
3. Ground traps are checked annually. Elevated traps in deciduous sites are checked once every two wee

Obtain basic summaries of your data (Neonics)

5. What are the dimensions of the dataset?

```r
dim(Neonics)  #the dim function finds the dimensions of the dataframe
```

```
## [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?

```r
summary(Neonics$Effect)  #Using the summary function to find how how of each type of effect was observe
```

```
##     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: These effects are of interest because it is useful to see how the insects responsed to exposure to the chemicals and to see how many experienced each type of effect.

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)  #using the summary function to find the most commonly studied spe
```

```
##                          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 used species were Honey Bees, Parasitic Wasps, Buff Tailed BumbleBee, Carniolan Honey Bees, Bumble Bees, and Italian Honey Bees. Many of these species are pollinators that provide valuable agricultural services for humans. If we use insecticides to target crop pests, we would want to understand the impacts on these pollinator species. Parasitic wasps target species that may be considered agricultural pests, so similarly we would want to understand how insecticides affect these species.

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.)  #Determining the class of the Conc.1..Author column.
```

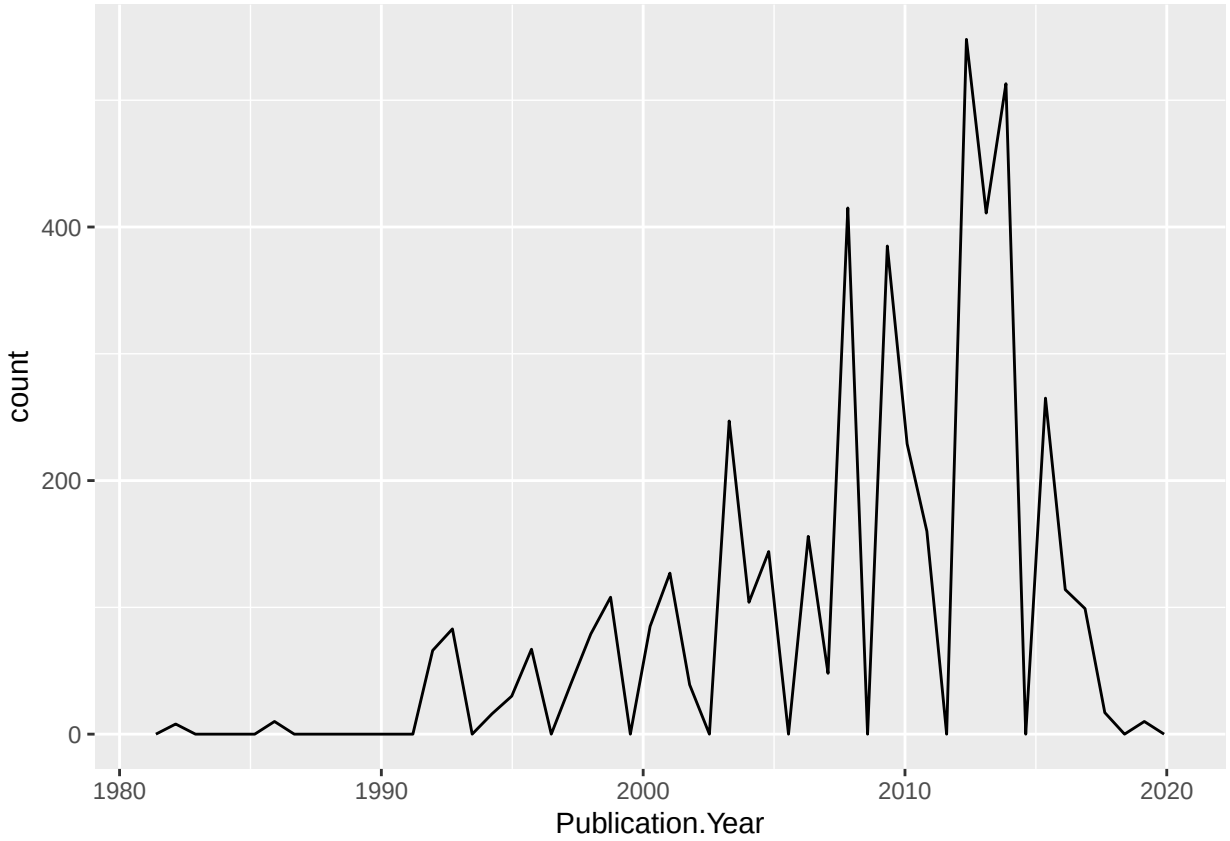
```
## [1] "factor"
```

Answer: Conc.1..Author is a factor. They represent different categories of concentrations because these concentrations depend on several other factors, like the chemical type and concentration type, so they aren't universal concentrations.

Explore your data graphically (Neonics)

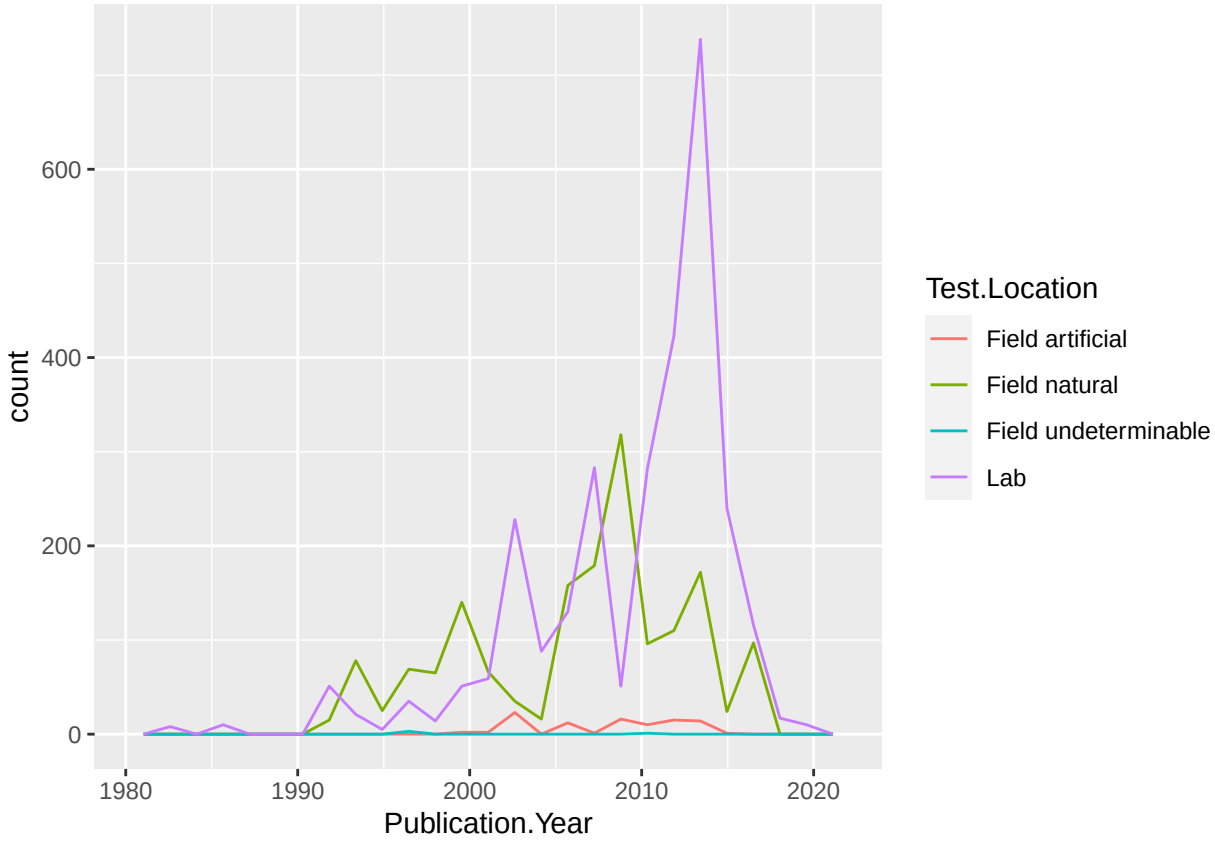
9. Using `geom_freqpoly`, generate a plot of the number of studies conducted by publication year.

```
ggplot(Neonics) + geom_freqpoly(aes(x = Publication.Year), bins = 50)  #Using a frequency line graph to
```



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 = 25)  #Determining the number of publications per year by the location of each test/experimen
```

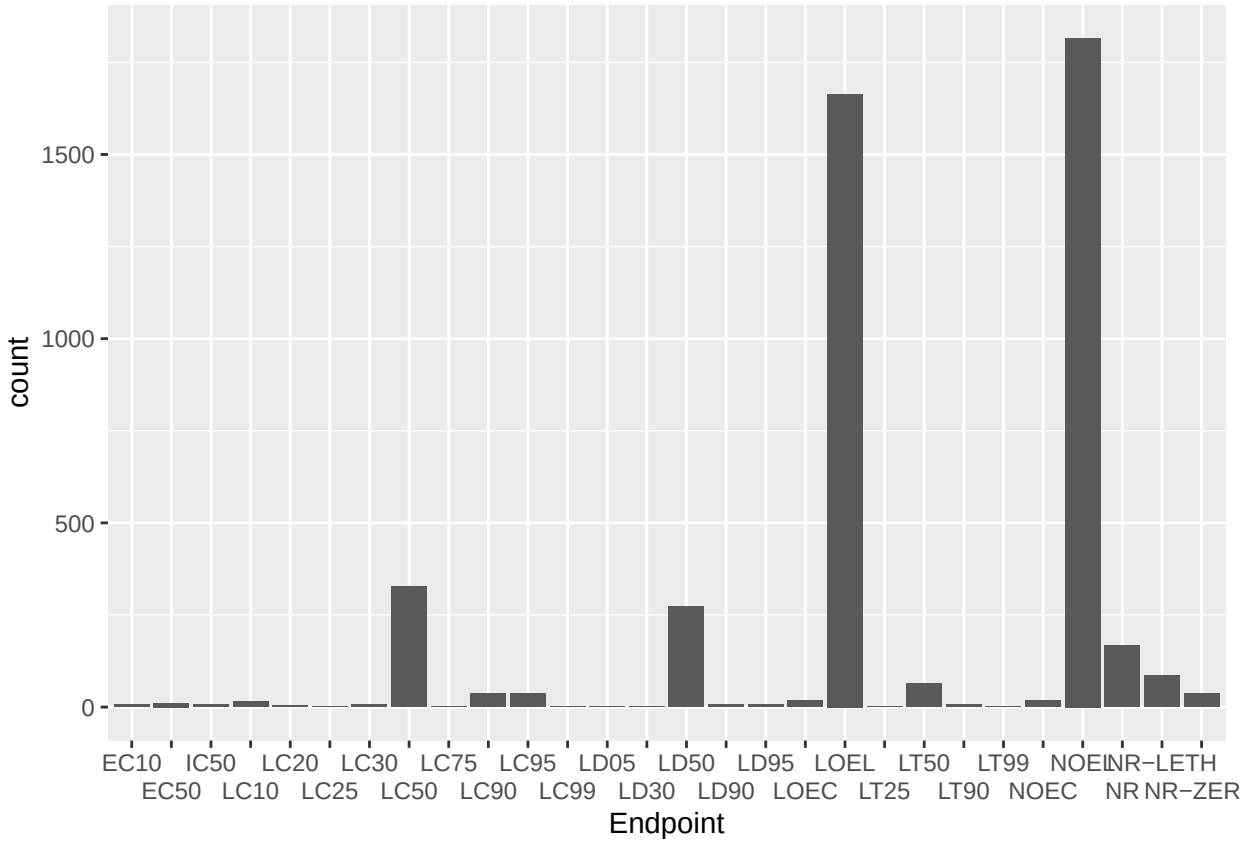


Interpret this graph. What are the most common test locations, and do they differ over time?

> Answer: The Lab and Field Natural are the most common test locations. This changes slightly over time. From the early 1990s into 2000, Field Natural was slightly more common but from that time forward, Lab test locations were relatively as common until about 2010 when Lab tests far exceeded all other test locations.

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)) + scale_x_discrete(guide = guide_axis(n.dodge = 2))   #usi
```



Answer: The two most common endpoints are LOEL (Lowest Observable Effects Level), or the lowest concentration producing effects that were significantly different from the controls, and the NOEL (No Observable Effect Level), the highest concentration that produced effects that were not significantly different than those observed in 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)  #looking at how RStudio is interpreting the collectDate column.
```

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

```
library(lubridate)
```

```
##
## Attaching package: 'lubridate'
```

```
## The following objects are masked from 'package:base':
##
##     date, intersect, setdiff, union
```

```
Litter$collectDate <- as.Date(Litter$collectDate, format = "%Y-%m-%d")  #using Lubridate to correct the
class(Litter$collectDate)
```

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

```
unique(Litter$collectDate)
```

```
## [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)  #using the unique function to see how many plots were sampled and comparing it t
```

```
##  [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
```

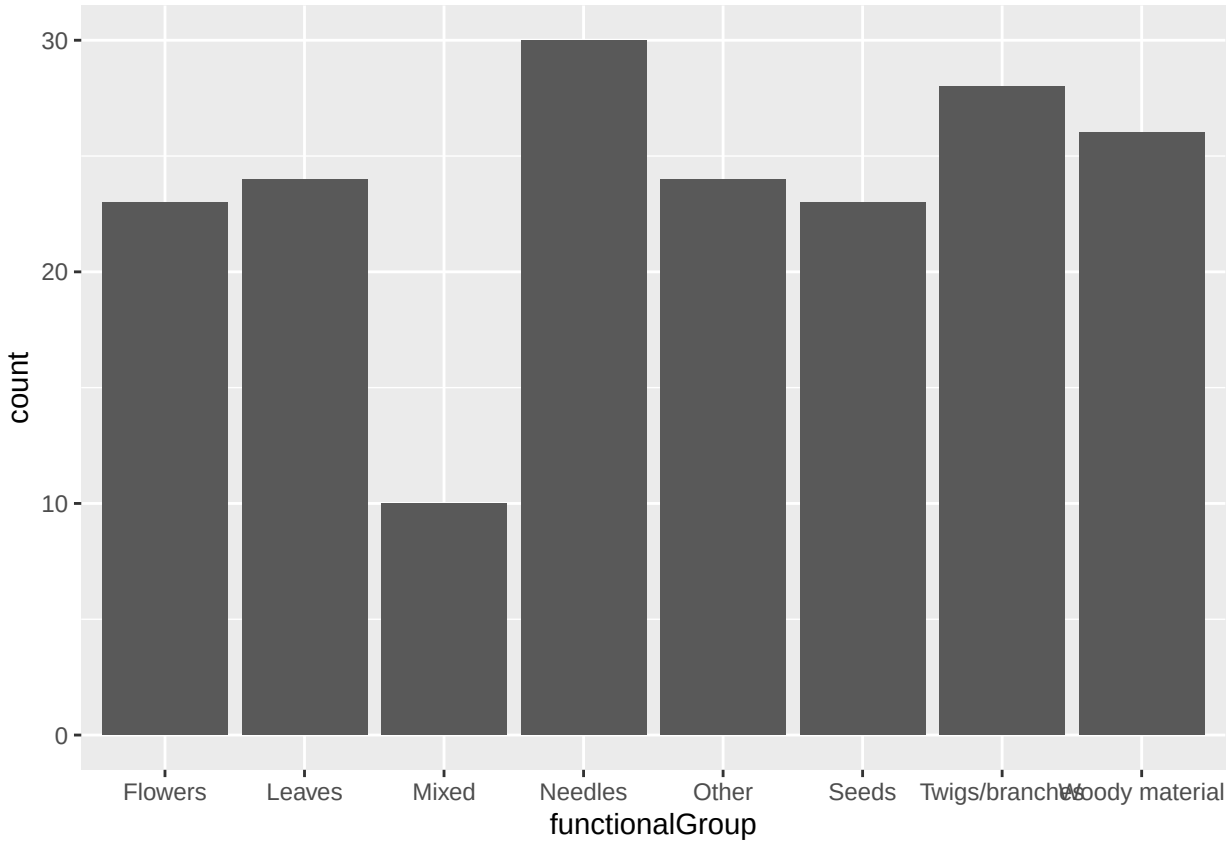
```
summary(Litter$plotID)
```

```
## 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: 12 plots were sampled at Niwot Ridge. Unique only returns a list of values that don't have duplicates based on your selection criteria. Summary will give a count of how many entries there are for each of your selections.

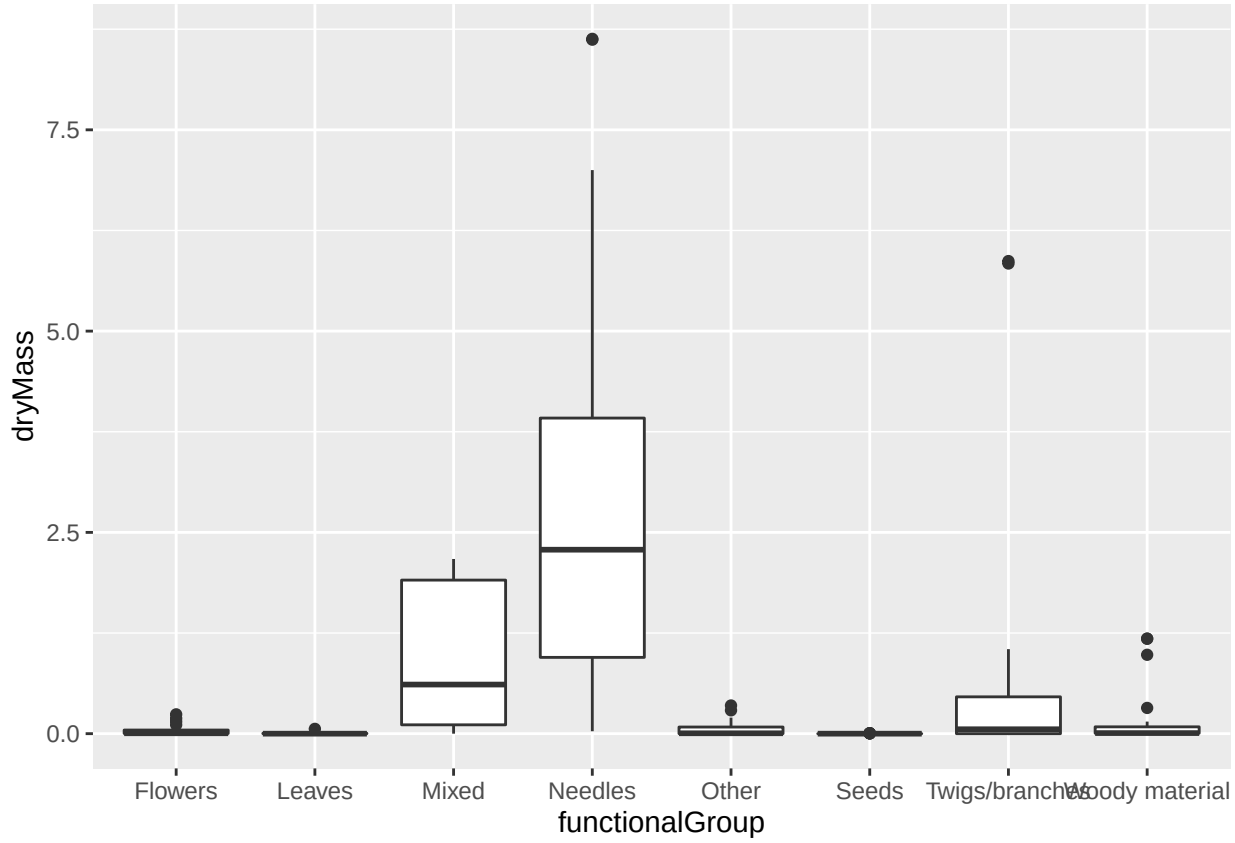
14. Create a bar graph of functionalGroup counts. This shows you what type of litter is collected at the Niwot Ridge sites. Notice that litter types are fairly equally distributed across the Niwot Ridge sites.

```
ggplot(Litter) + geom_bar(aes(x = functionalGroup))  #creating a bar graph showing what types of litter
```

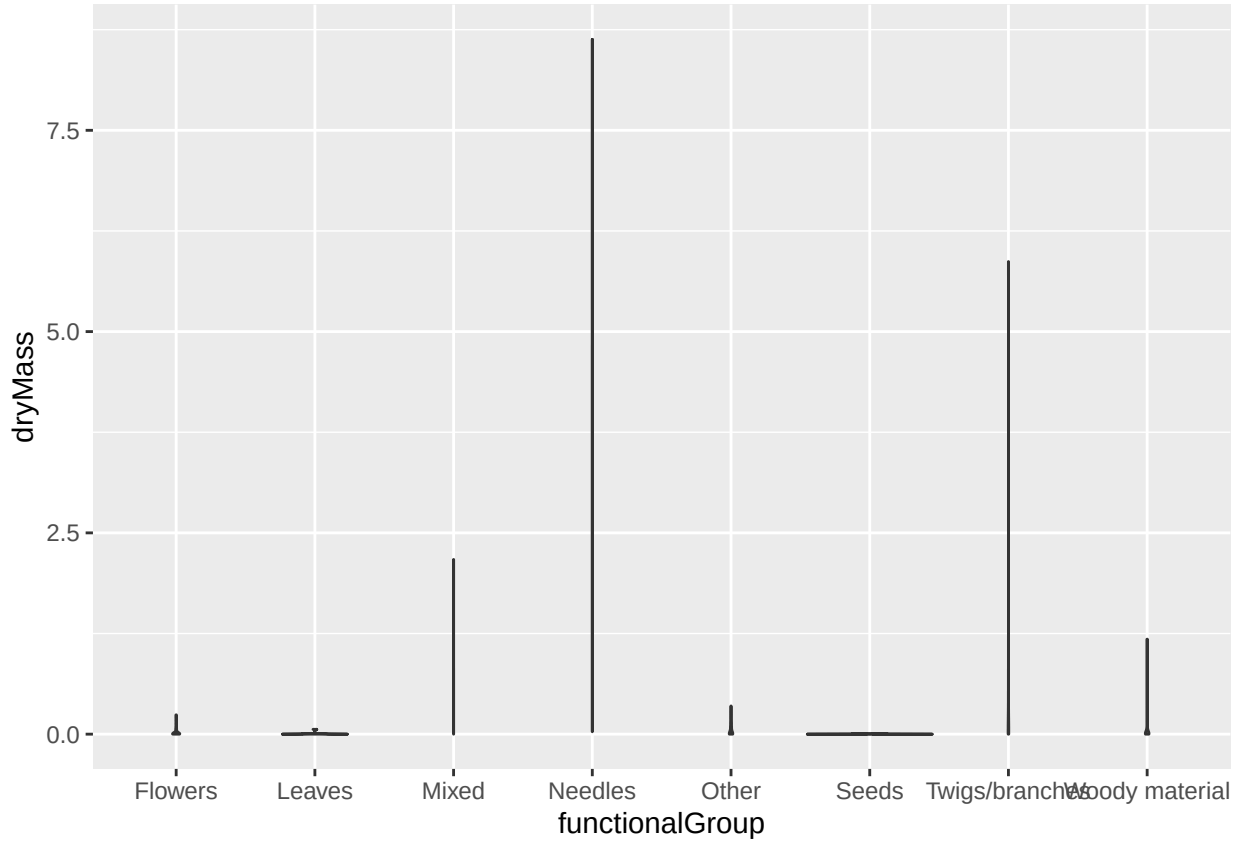


15. Using `geom_boxplot` and `geom_violin`, create a boxplot and a violin plot of dryMass by functionalGroup.

```
ggplot(Litter) + geom_boxplot(aes(x = functionalGroup, y = dryMass))  #using a bar graph to visualize t
```




```
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: In this case the boxplot is a more effective visualization option because without further manipulation, the violin plot is too compressed to reveal anything about the data.

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

> Answer: Needles, mixed, and twigs/branches have the highest biomass at these sites.