Assignment 3: Data Exploration

Lambert Ngenzi, Section 1

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 <>. ##comment code

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] "C:/Users/ln113/Documents/EDA_R/Environmental_Data_Analytics_2022/Assignments"
```

```
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: Most of insects are either pollinators (moving pollen) or/and carriers of diseases (harmful to the environment), it is important to undestand insectidides applied to crops or other because potential effects they may cause to our environment or about their impact on the health of insect/species populations.

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 know litter or woody debris arehomes to different insects that can be moved via different means (wind, water, species such as human . . .), It is important to know how spatially and temporally these insects traveled in a region or place. Let's say when these insects end up in rivers or any body of water, aquatic species fed up on these insects, which we or other species also end up consuming. Food chain concept.

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 sampled is done at terrestrial NEON sites containing woody vegetation >2m tall. They also provide essential data for a better understanding of vegetative carbon fluxes over time. *Spatial Sampling Deesign: Uses terrestrial NEON sites containing woody vegetation >2m tall* Sampling takes place in 20 40m x 40m plots. *Temporal Sampling Design:Grounds traps sampled once per year

Obtain basic summaries of your data (Neonics)

5. What are the dimensions of the dataset? 4623 observations and 30 variables

```
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?

```
Only_Effect <- summary(Neonics$Effect)
Only_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 effects studied are, population. It is important to know the effect of insecticde on insects population and being able to track that over time is very useful.

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.

```
Most_Common_name <- summary(Neonics$Species.Common.Name)
Most_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
```

```
Top6 <- head(sort(Most_Common_name, decreasing = TRUE), n=6)
Top6
```

```
##                (Other)             Honey Bee        Parasitic Wasp
##                    670                   667                   285
## Buff Tailed Bumblebee   Carniolan Honey Bee            Bumble Bee
##                    183                   152                   140
```

Answer: They are all pollinators and belong to the same order. Our crop yield and health depends heavily on these pollinators.

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: It is a factor, because the column contains a mixture of letters and numbers. So the software considers the entire column as a factor

Explore your data graphically (Neonics)

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

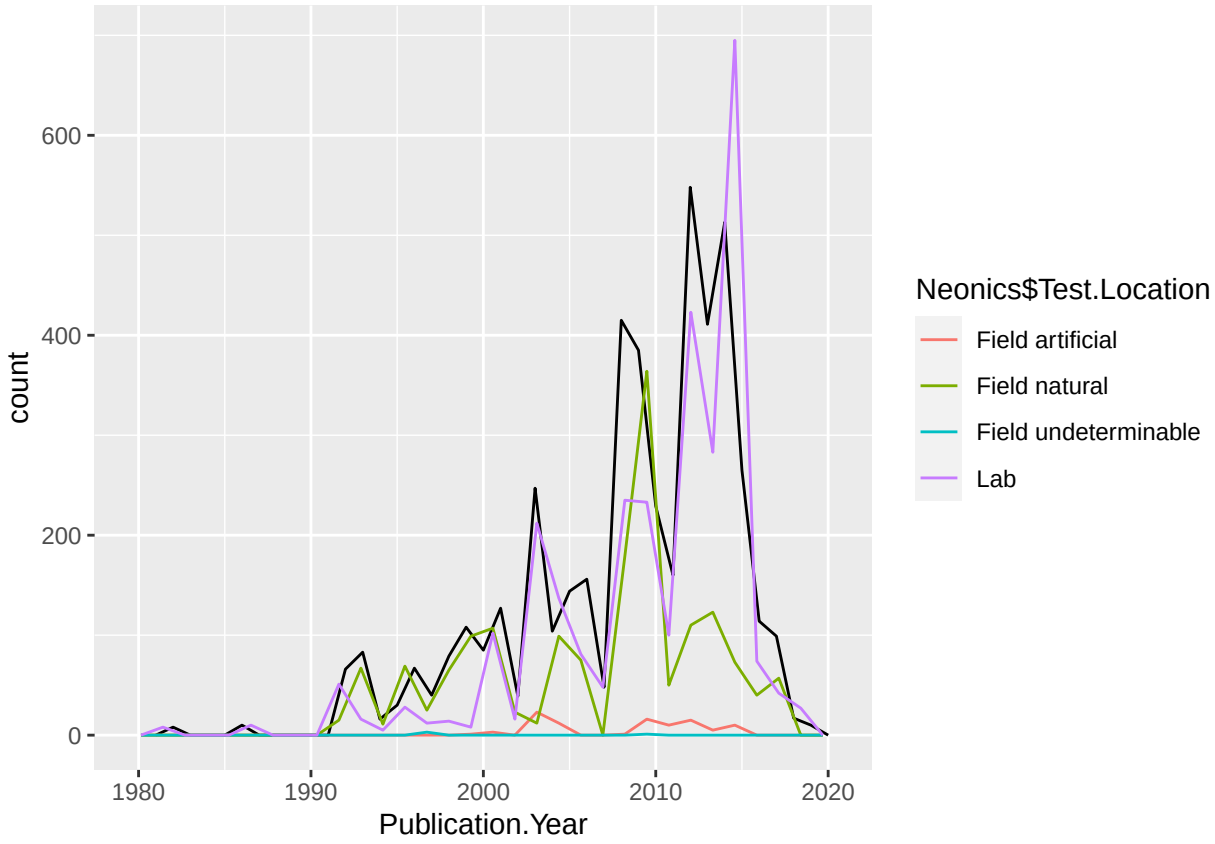
```
PubYear_Count <- ggplot(Neonics, aes(Publication.Year)) +
  geom_freqpoly(binwidth=1)
```

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

```
PubYear_Count + geom_freqpoly(aes(x = Publication.Year, color= Neonics$Test.Location))
```

```
## Warning: Use of `Neonics$Test.Location` is discouraged. Use `Test.Location`
## instead.
```

```
## `stat_bin()` using `bins = 30`. Pick better value with `binwidth`.
```



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

Answer:We see a trend of increasing publication since the 1980 but with a seasonal occurence

over year. Before 2014, there have been a sudden decrease of publication. The most common test location are in the lab

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.

```
Endpoint_count <- ggplot(Neonics, aes(Endpoint)) +
  geom_bar()
```

Answer: The two most common end points are: 1. NOEL: No-observable-effect-level, the highest dose or concentration not significant 2. LOEL:Lowest-observable-effect-level, lowest concentration significant

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"
```

```
#  will change the format from factor to Date
 as.Date(Litter$collectDate, format = "%Y-%m-%d")
```

```
##   [1] "2018-08-02" "2018-08-02" "2018-08-02" "2018-08-02" "2018-08-02"
##   [6] "2018-08-02" "2018-08-02" "2018-08-02" "2018-08-02" "2018-08-02"
##  [11] "2018-08-02" "2018-08-02" "2018-08-02" "2018-08-02" "2018-08-02"
##  [16] "2018-08-02" "2018-08-02" "2018-08-02" "2018-08-02" "2018-08-02"
##  [21] "2018-08-02" "2018-08-02" "2018-08-02" "2018-08-02" "2018-08-02"
##  [26] "2018-08-02" "2018-08-02" "2018-08-02" "2018-08-02" "2018-08-02"
##  [31] "2018-08-02" "2018-08-02" "2018-08-02" "2018-08-02" "2018-08-02"
##  [36] "2018-08-02" "2018-08-02" "2018-08-02" "2018-08-02" "2018-08-02"
##  [41] "2018-08-02" "2018-08-02" "2018-08-02" "2018-08-02" "2018-08-02"
##  [46] "2018-08-02" "2018-08-02" "2018-08-02" "2018-08-02" "2018-08-02"
##  [51] "2018-08-02" "2018-08-02" "2018-08-02" "2018-08-02" "2018-08-02"
##  [56] "2018-08-02" "2018-08-02" "2018-08-02" "2018-08-02" "2018-08-02"
##  [61] "2018-08-02" "2018-08-02" "2018-08-02" "2018-08-02" "2018-08-02"
##  [66] "2018-08-02" "2018-08-02" "2018-08-02" "2018-08-02" "2018-08-02"
##  [71] "2018-08-02" "2018-08-02" "2018-08-02" "2018-08-02" "2018-08-02"
##  [76] "2018-08-02" "2018-08-02" "2018-08-02" "2018-08-02" "2018-08-02"
##  [81] "2018-08-02" "2018-08-02" "2018-08-02" "2018-08-02" "2018-08-02"
##  [86] "2018-08-02" "2018-08-02" "2018-08-02" "2018-08-02" "2018-08-02"
##  [91] "2018-08-02" "2018-08-30" "2018-08-30" "2018-08-30" "2018-08-30"
##  [96] "2018-08-30" "2018-08-30" "2018-08-30" "2018-08-30" "2018-08-30"
## [101] "2018-08-30" "2018-08-30" "2018-08-30" "2018-08-30" "2018-08-30"
## [106] "2018-08-30" "2018-08-30" "2018-08-30" "2018-08-30" "2018-08-30"
## [111] "2018-08-30" "2018-08-30" "2018-08-30" "2018-08-30" "2018-08-30"
## [116] "2018-08-30" "2018-08-30" "2018-08-30" "2018-08-30" "2018-08-30"
## [121] "2018-08-30" "2018-08-30" "2018-08-30" "2018-08-30" "2018-08-30"
## [126] "2018-08-30" "2018-08-30" "2018-08-30" "2018-08-30" "2018-08-30"
## [131] "2018-08-30" "2018-08-30" "2018-08-30" "2018-08-30" "2018-08-30"
## [136] "2018-08-30" "2018-08-30" "2018-08-30" "2018-08-30" "2018-08-30"
## [141] "2018-08-30" "2018-08-30" "2018-08-30" "2018-08-30" "2018-08-30"
## [146] "2018-08-30" "2018-08-30" "2018-08-30" "2018-08-30" "2018-08-30"
## [151] "2018-08-30" "2018-08-30" "2018-08-30" "2018-08-30" "2018-08-30"
## [156] "2018-08-30" "2018-08-30" "2018-08-30" "2018-08-30" "2018-08-30"
```

```
## [161] "2018-08-30" "2018-08-30" "2018-08-30" "2018-08-30" "2018-08-30"
## [166] "2018-08-30" "2018-08-30" "2018-08-30" "2018-08-30" "2018-08-30"
## [171] "2018-08-30" "2018-08-30" "2018-08-30" "2018-08-30" "2018-08-30"
## [176] "2018-08-30" "2018-08-30" "2018-08-30" "2018-08-30" "2018-08-30"
## [181] "2018-08-30" "2018-08-30" "2018-08-30" "2018-08-30" "2018-08-30"
## [186] "2018-08-30" "2018-08-30" "2018-08-30"
```

```
##Determine dates litter was sampled
Litter_Sampled_August <- unique(Litter$collectDate, fromLast = 2018-08)
Litter_Sampled_August
```

```
## [1] 2018-08-02 2018-08-30
## Levels: 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`?

```
Ploting_Niwot <- unique(Litter$siteID)
Ploting_Niwot
```

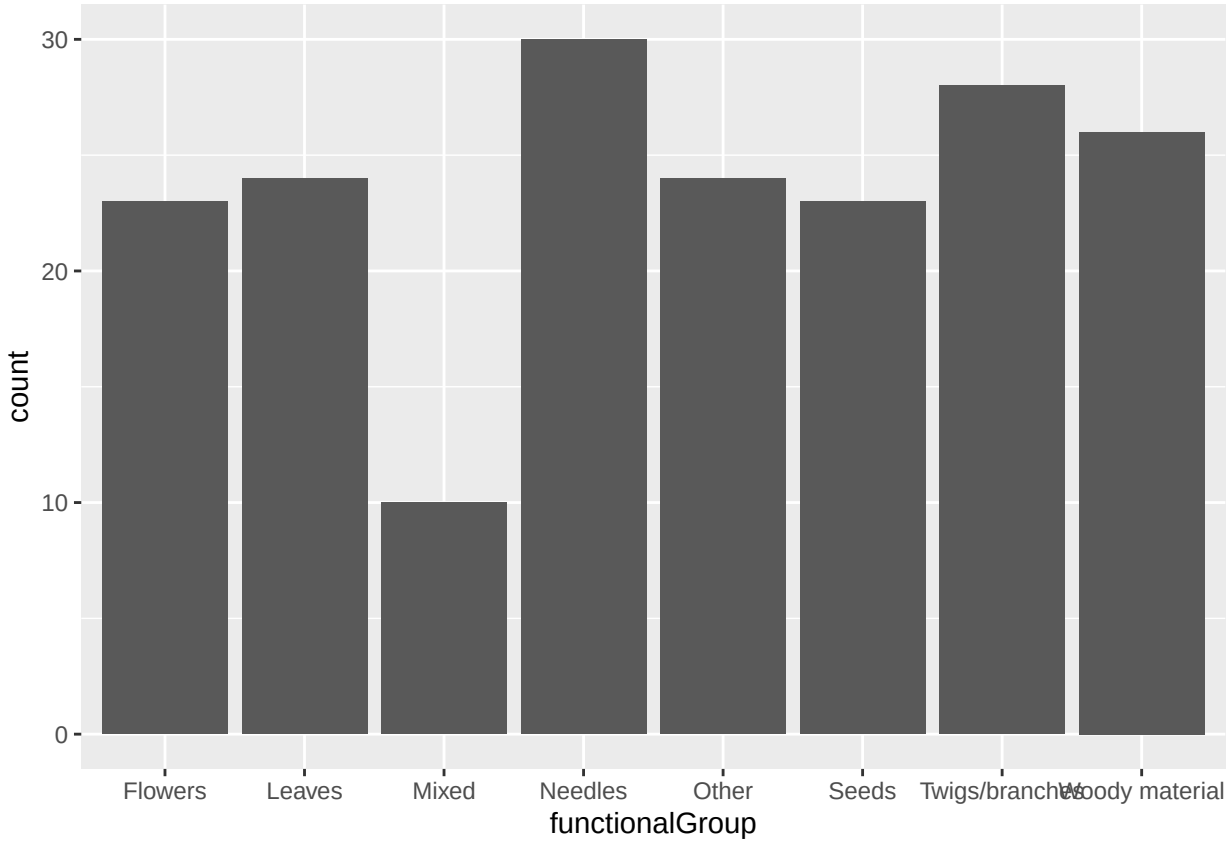
```
## [1] NIWO
## Levels: NIWO
```

```
## Create a name function for site ID
```

Answer: Unique function only give you the unique value (eliminated duplicated vlues such as here, Niwo came back) While summary function accounts all the data and doesn't do well with factors

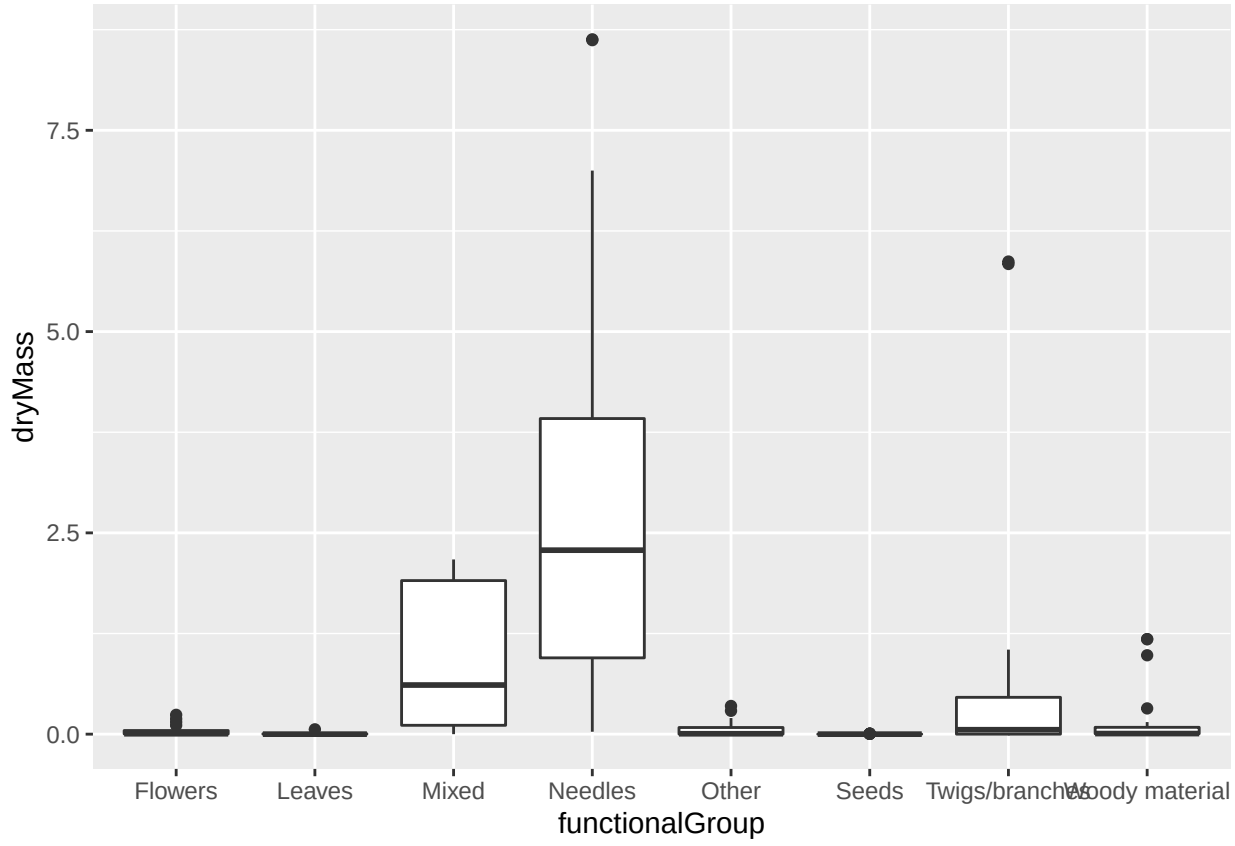
14. Create a bar graph of functionalGroup counts. This shows you what type of litter is collected at the Niwot Ridge sites. Notice that litter types are fairly equally distributed across the Niwot Ridge sites.

```
FunctionalGroup_plot <- ggplot(Litter, aes(functionalGroup)) +
                                geom_bar()
FunctionalGroup_plot
```

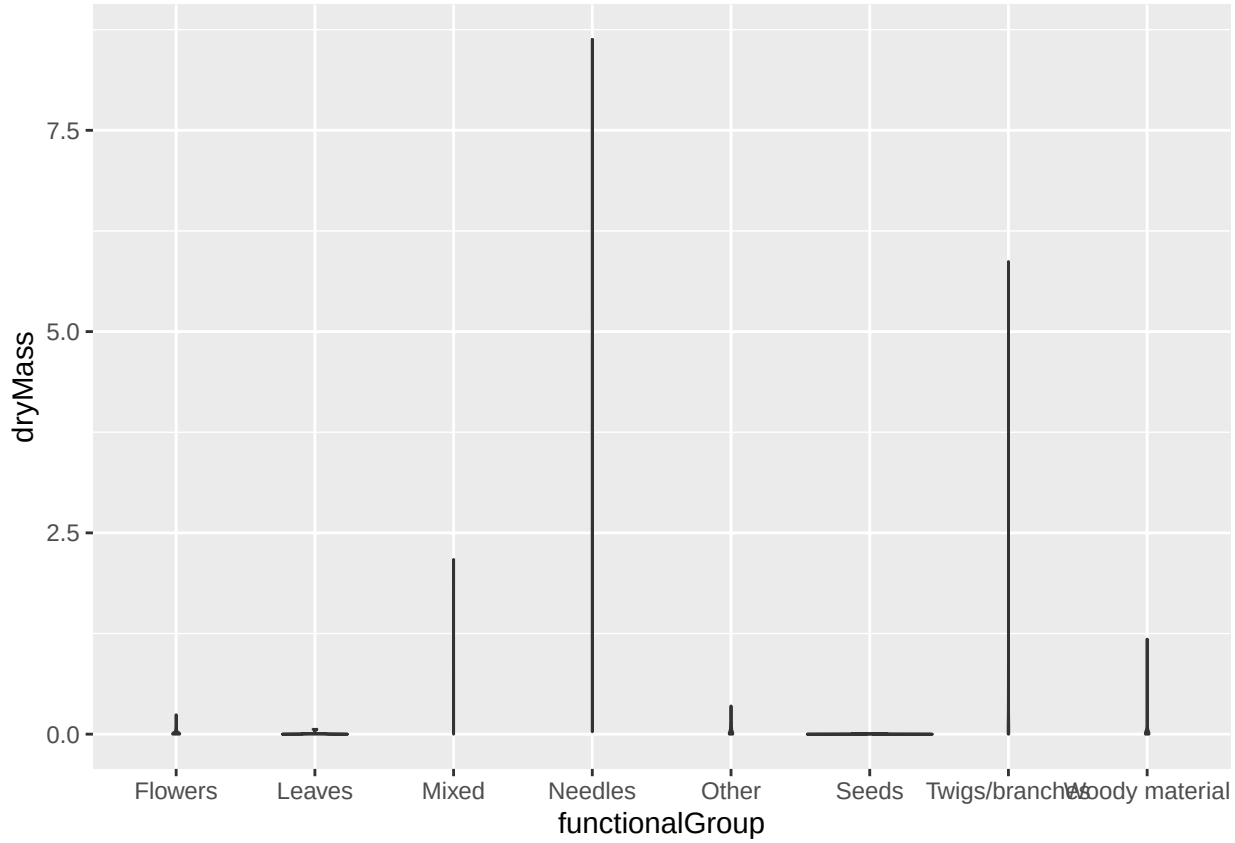


15. Using `geom_boxplot` and `geom_violin`, create a boxplot and a violin plot of dryMass by functionalGroup.

```
ggplot(Litter, aes(functionalGroup, dryMass)) +
  geom_boxplot()
```




```
ggplot(Litter, aes(functionalGroup, dryMass)) +
  geom_violin()
```



Why is the boxplot a more effective visualization option than the violin plot in this case?

> Answer: The boxplot gives you a better idea of the distribution of the data of each type while the violin plot is less informative about the data you are computing. As a visualization tool, it is limited

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

> Answer: Needles follow by Twigs/branches have highest biomass respectively