Assignment 3: Data Exploration

Ben Egan, Section Thursday

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 <1.31>.

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.**

```
Neonics <- read.csv(file.choose(), stringsAsFactors = TRUE)
Litter <- read.csv(file.choose(), stringsAsFactors = TRUE)
library(ggplot2)
library(tidyverse)
```

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: Neonicotinoids are an effective inseticide and unfortunately that effect is fealt by Bee populations as well. Ecotoxicolgy data is important to see how the use of these insectides may be impacting Bee populations, water health and other envrionmental implications

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: Litter and woody debris are great sources of food for organisms/ bacteria/ mosses on the forest floor. This in turn creates healthy soil and healthy plants. Litter and woody debris can also drastically reduce soil erosion.

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: *The sampling of the Litter and Woody Debris happens in tower plots where debris falls into a marked off area that can be easily measured* Some of the traps are elevated off the ground and some are on the ground) *Depending on the the trap method and how wooded the are is, the sites can be sampled every 2 weeks to 6 Months

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: Some of the most common are Mortality and Popultaion. It's important to know this as the scientist are trying to measure the effect of the pesticides on eliminating Pests

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 common Are Honey Bee, Parasitic Wasp, Buff Tailed Bumblebee, Carniolan Honey Bee, and Bumble Bee. They are all Bees which are crucial to the pollination of the crops and the general health of the ecological environment.

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)
```

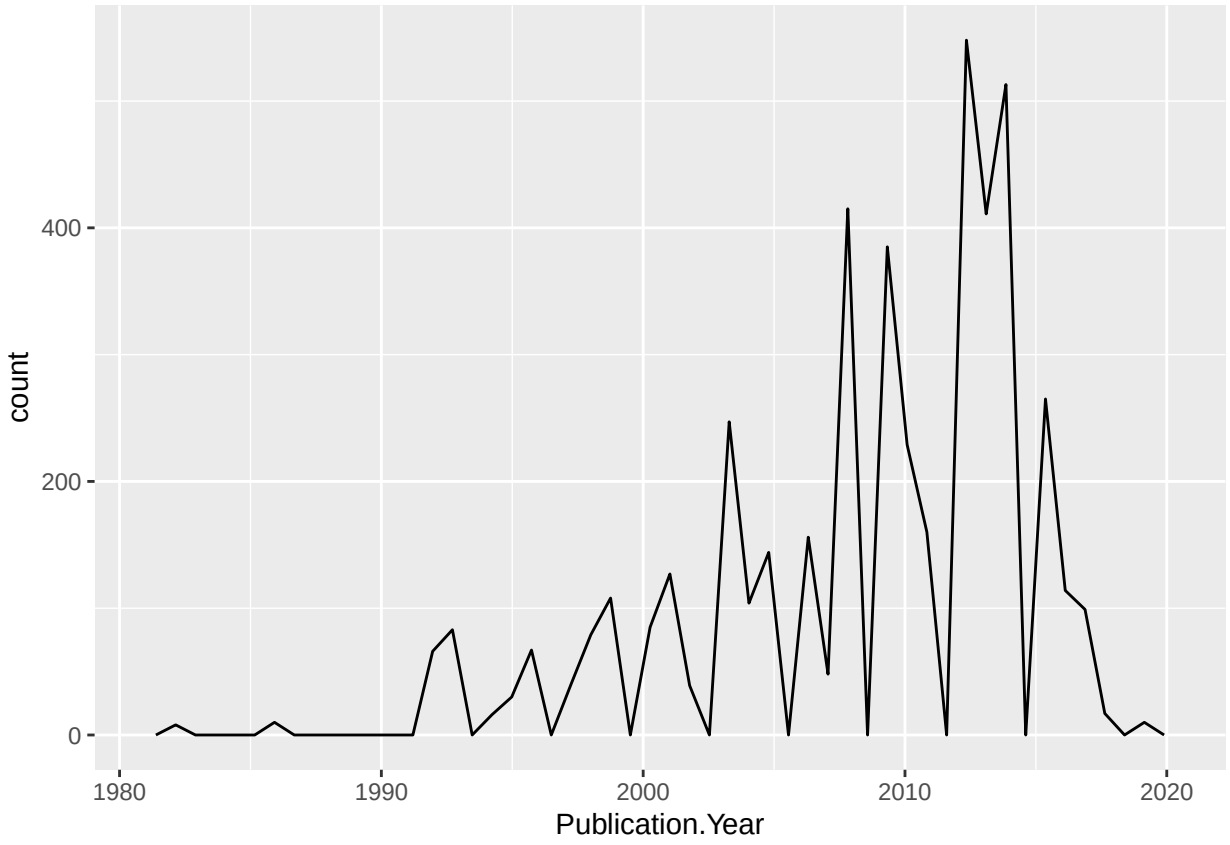
```
## [1] "factor"
```

Answer: The class is "factor" because some of the values have been entered as NR or with "/". If there is a non-numeric value than the whole row is recorded as Character

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)
```



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

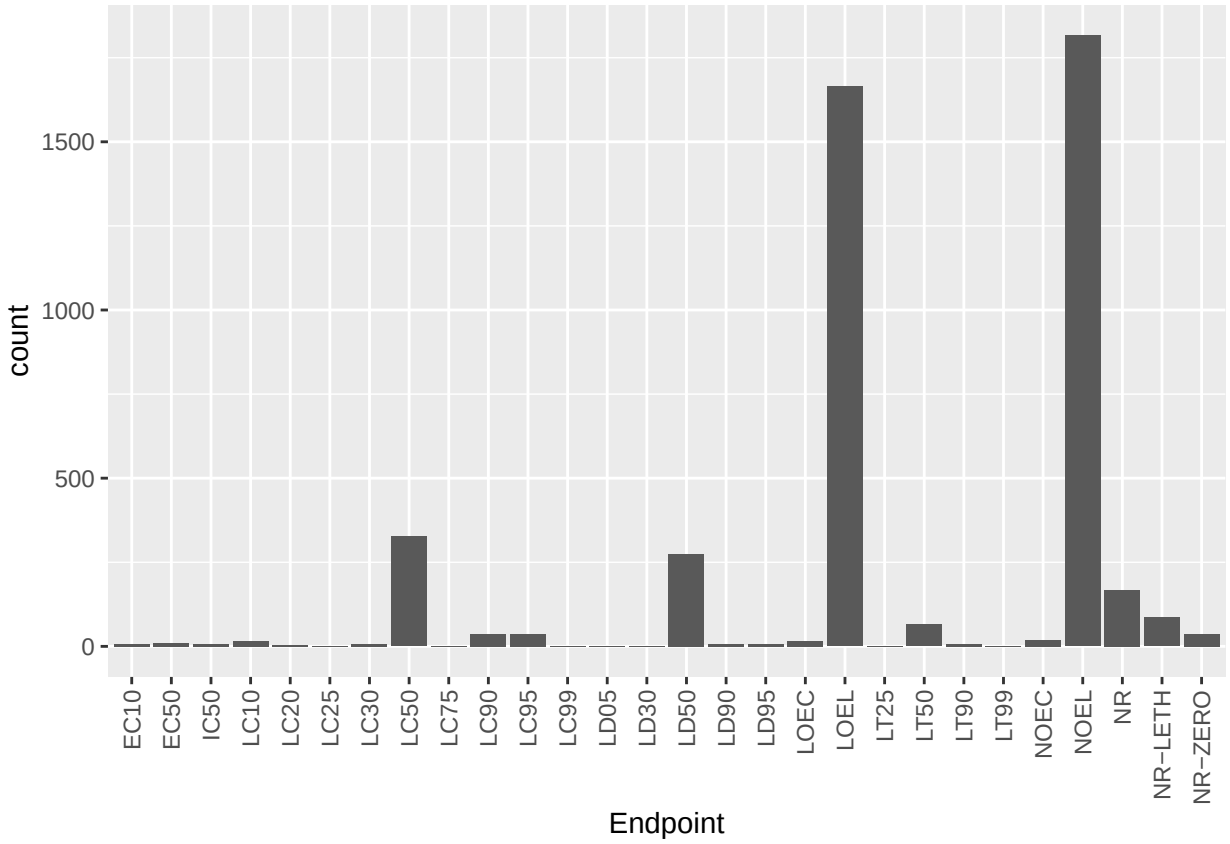
```
#ggplot(Neonics, aes(x = Publication.Year, color = Test.Location)) +
  #geom_freqpoly() + theme(legend.position = "top")
```

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

Answer: The most common are lab and Field Natural. And over time, Lab tests become much more popular while natural studies have come down a little since the mid 00's.

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), color = Endpoint) + geom_bar()+theme(axis.text.x = element_text(angl
```



Answer: The two Most common endpoints are LOEL And NOEL. a LOEL is the lowest observed effect level and the NOEL is the No Observed effect Level. For Toxicology studies these are teh most important distinguishers as they determine the to what the standards.

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")

Aug_Dates <- unique(Litter$collectDate)

print(Aug_Dates)
```

```
## [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`?

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

```
##  [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
```

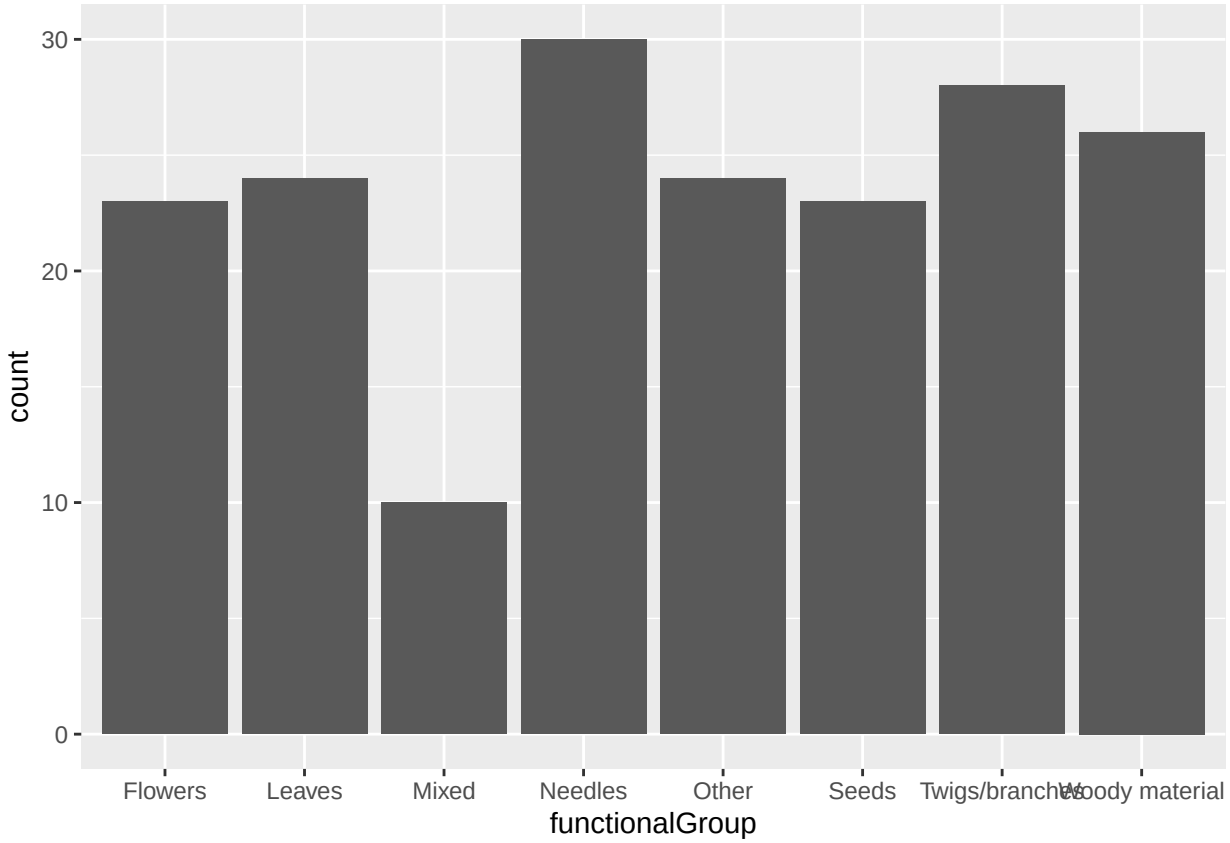
```
sum(summary(Niwot_Ridge))
```

```
## [1] 12
```

Answer: They are different as the Unique function lists all of the individual Plot names. My summary function has summed the unique values and leaves out the individual names.

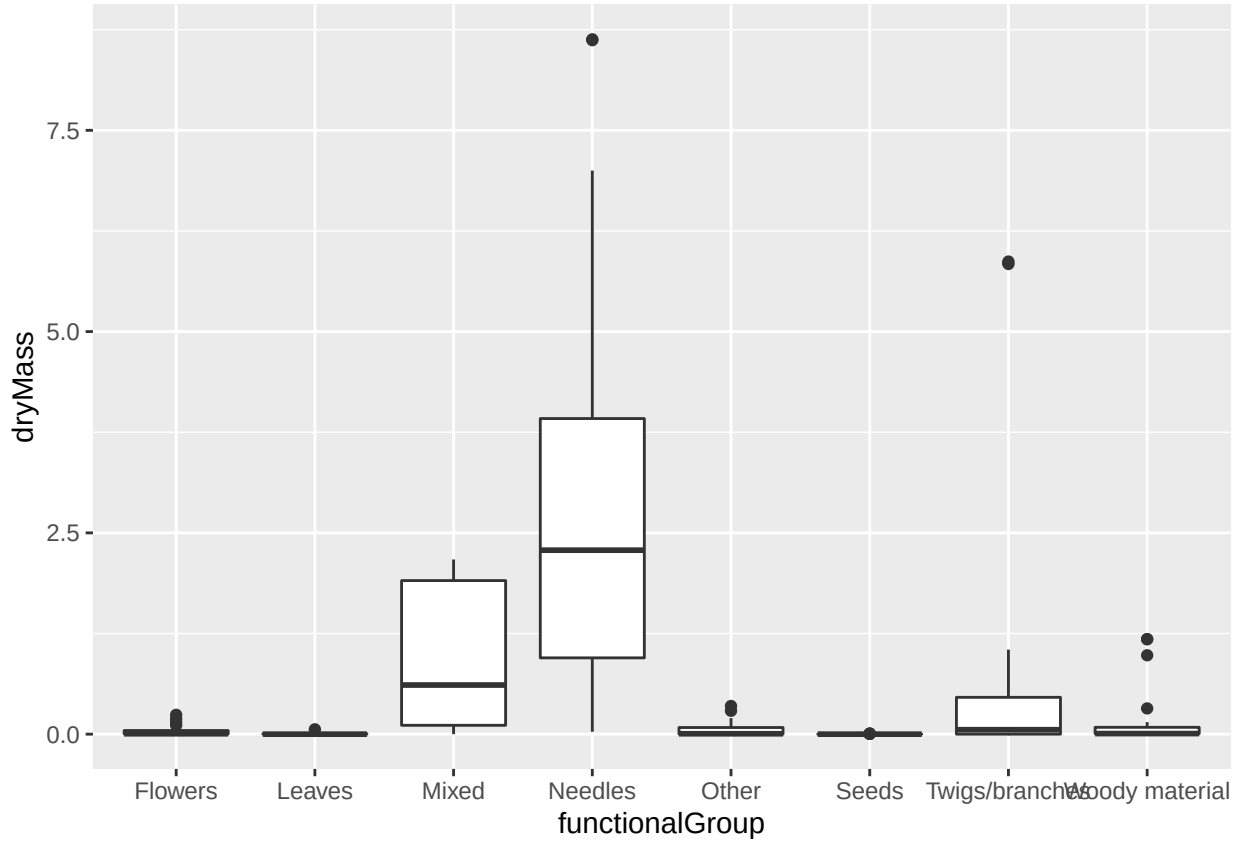
14. Create a bar graph of functionalGroup counts. This shows you what type of litter is collected at the Niwot Ridge sites. Notice that litter types are fairly equally distributed across the Niwot Ridge sites.

```
ggplot(Litter, aes(x = functionalGroup), color = functionalGroup) + geom_bar()
```

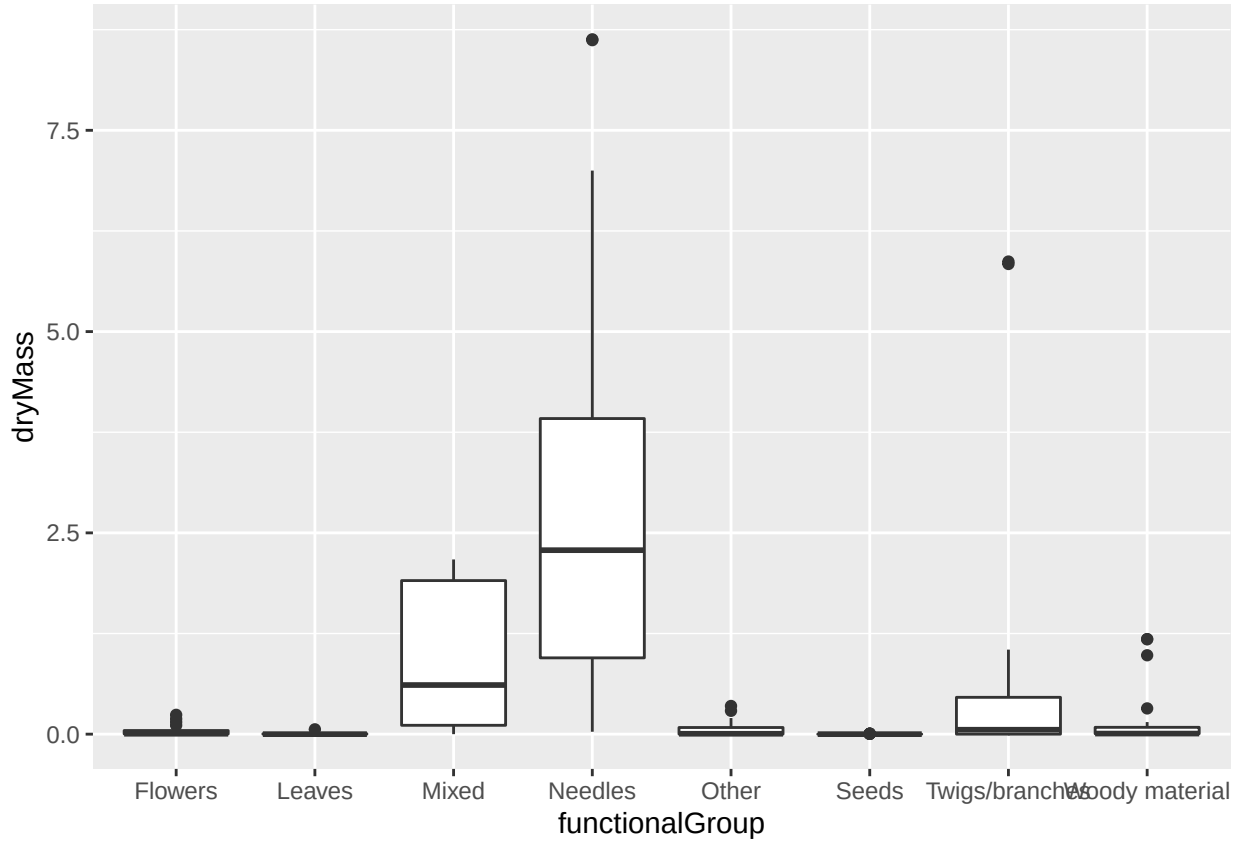


15. Using `geom_boxplot` and `geom_violin`, create a boxplot and a violin plot of dryMass by functionalGroup.

```
ggplot(Litter) + geom_boxplot(aes(x = functionalGroup, y = dryMass))
```




```
ggplot(Litter) + geom_boxplot(aes(x = functionalGroup, y = dryMass))
```

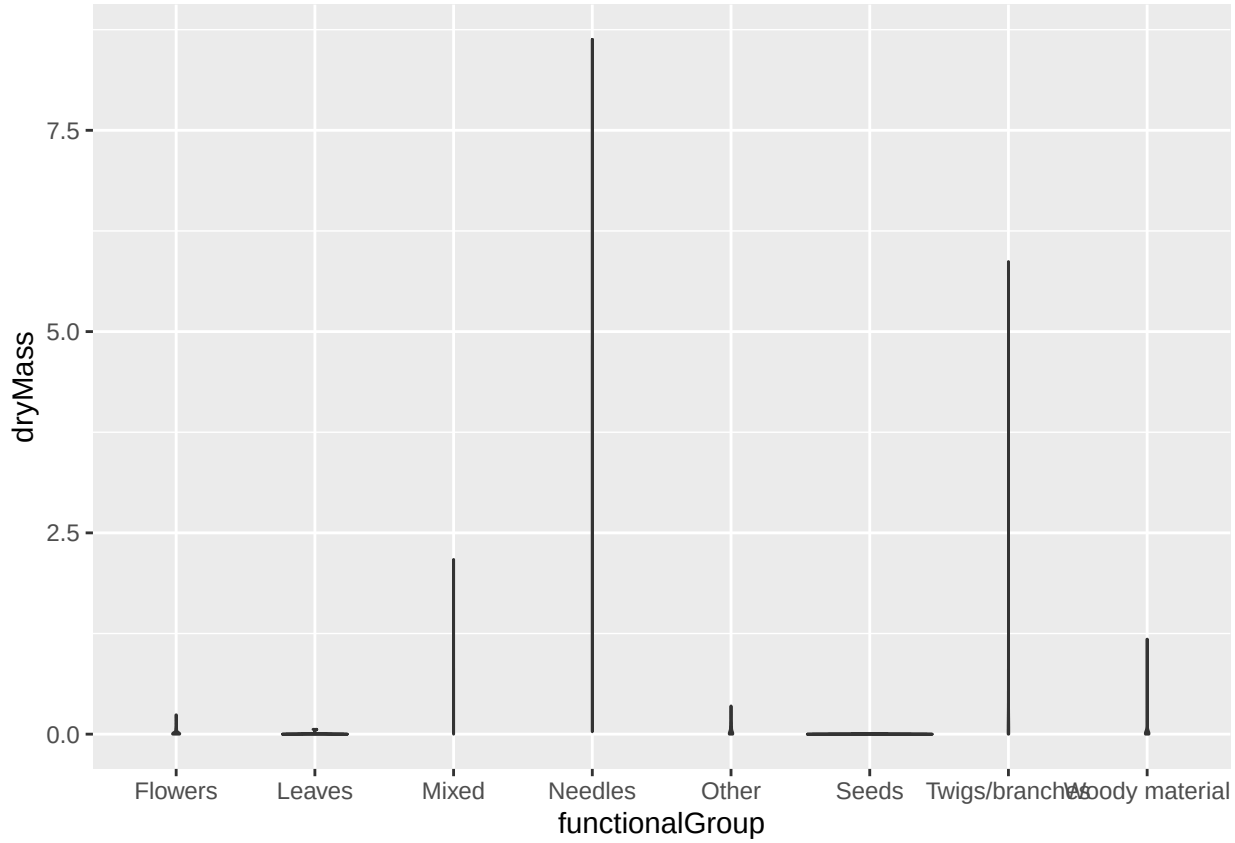



```
ggplot(Litter) +
  geom_violin(aes(x = functionalGroup, y = dryMass),
              draw_quantiles = c(0.25, 0.5, 0.75))
```

```
## 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 Box plot gives a lot more information that shows outliers and the quartiles. the Violin does not give the quartile information and distributes the full range evenly

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

> Answer: Needles have the highest biomass at these sites.