Assignment 3: Data Exploration

Shana Shapiro, 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 <January 30, 2022>.

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.**

```
library(ggplot2)
```

```
## Warning: package 'ggplot2' was built under R version 4.0.5
```

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

```
getwd()
```

```
## [1] "Z:/EnvironmentalDataAnalytics/Environmental_Data_Analytics_2022/Assignments"
```

```
ecotox <- read.csv("../Data/Raw/ECOTOX_Neonicotinoids_Insects_raw.csv",stringsAsFactors = TRUE)
neon <- 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 might be interested in the ecotoxicology of neonicotinoids of insects because they can negatively impact non-target insects. Trace amounts of neonicotinoids can linger in plant pollen and kill bee populations. Decimation of the already-struggling species and other crucial pollinators can have cascading negative impacts on the global food system and ecosystem functions.

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: Trees are important to forest function throughout their life cycle, including after dropping branches or trunks that make up woody debris. The material can provide habitat and contributes to carbon and nitrogen cycling. The long-term component of the NEON research is also important as they can better assess the impacts of woody debris and litter over time.

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: *Sampling differed temporally between deciduous and evergreen sites. Deciduous sites were sampled frequently during senescence and evergreen sites were samples infrequently throughout the year.* Locations of the tower plots were selected randomly within the 90% flux footprint of the primary airshed. *In sites where the majority of aerial cover is woody vegetation, placement of litter traps is random and uses a list of grid cell locations being used for herbaceous clip harvest and bryophyte sampling.

Obtain basic summaries of your data (Neonics)

5. What are the dimensions of the dataset?

```
dim(ecotox)
```

```
## [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(ecotox$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 effect on Population and Mortality are of interest because negative impacts of neonicotinoids will specifically impact those aspects. Population and Mortality measurements can provide quantitative evidence of negative impacts of the pesticide.

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(ecotox$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: Honey Bee (667), Parasitic Wasp (285), Buff Tailed Bumblebee (183), Carniolan Honey Bee (152), Bumble Bee (140), Italian Honeybee (113). Each of the top six species are insect species that are important for pollination or are species that may be indirectly affected by neonicotinoid pesticides. These species may be have higher interest over other insects because pollinators are crucial to the human food system.

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(ecotox$Conc.1..Author.)
```

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

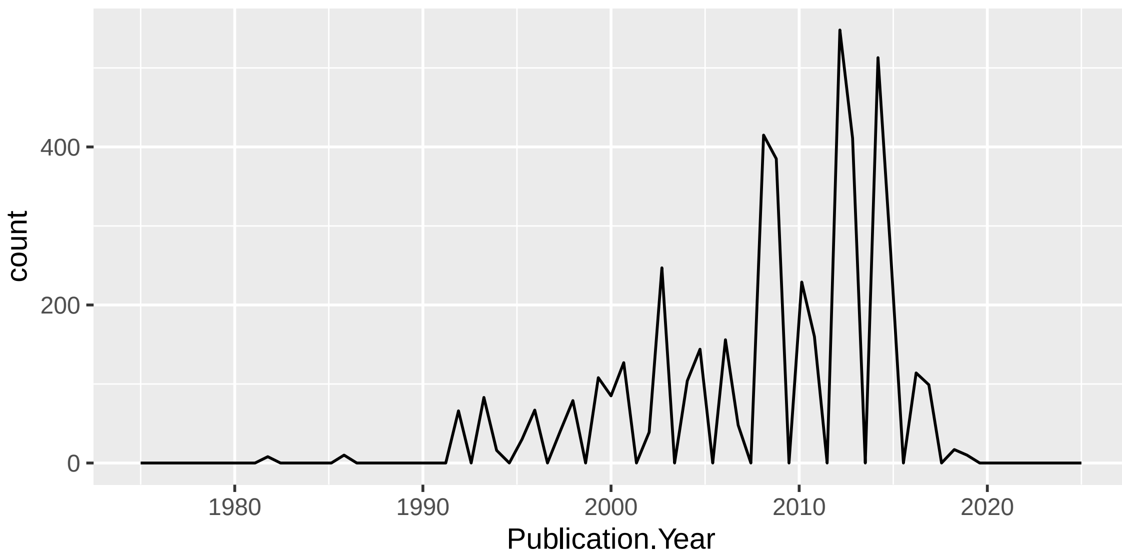
> Answer: While the values are mostly numeric, some are NR (Not Reported) or there are backslashes on some values. Since there are characters that are not only numeric, the class of the entire column is not considered numeric and is considered factor.

Explore your data graphically (Neonics)

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

```
ggplot(ecotox) +
  geom_freqpoly(aes(x = Publication.Year), bins = 75) +
  scale_x_continuous(limits = c(1975,2025)) +
  theme(legend.position = "top")
```

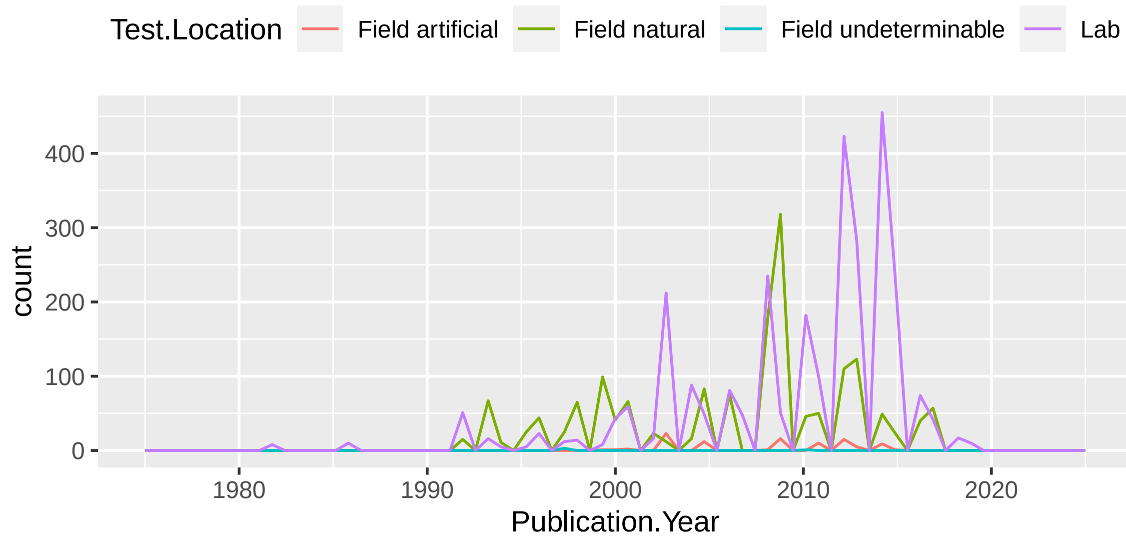
```
## Warning: Removed 2 row(s) containing missing values (geom_path).
```



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

```
ggplot(ecotox) +
  geom_freqpoly(aes(x = Publication.Year, color = Test.Location), bins = 75) +
  scale_x_continuous(limits = c(1975,2025)) +
  theme(legend.position = "top")
```

```
## Warning: Removed 8 row(s) containing missing values (geom_path).
```

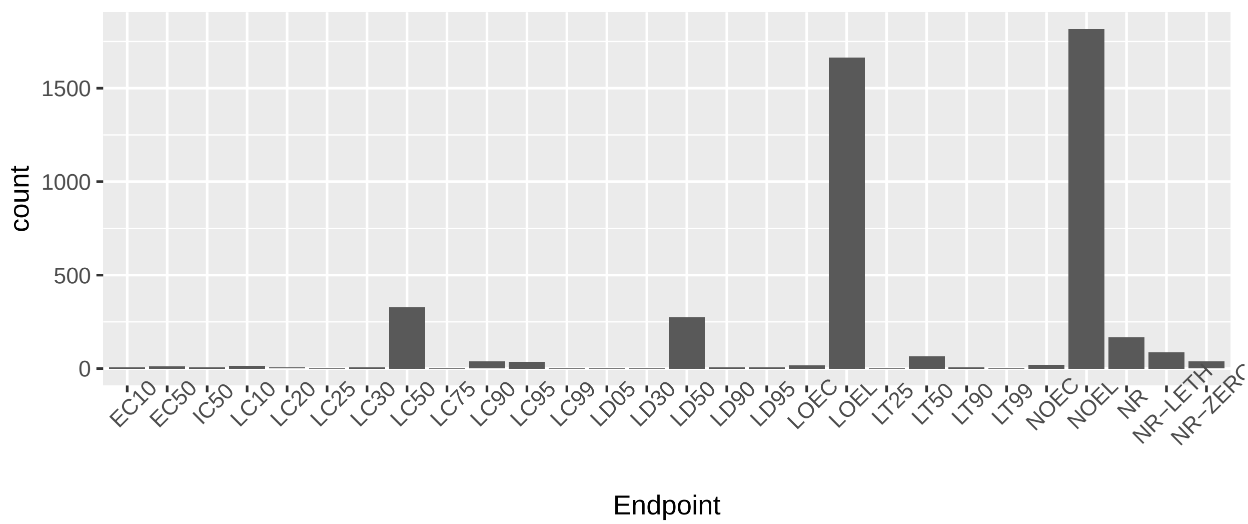


Interpret this graph. What are the most common test locations, and do they differ over time?

> Answer: The most common test locations appear to be in the field. The common test locations differs over time, however. Before the 2000's, test locations in the field used to be the most common location. Post-2000's, the lab became the most common test location, though there were more tests overall.

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(ecotox, aes(x = Endpoint)) +
  theme(axis.text.x = element_text(angle = 45)) +
  geom_bar()
```



> Answer: The two most common endpoints are the LOEC (lowest observable effect level) and NOEC (no observable effect level). The endpoints indicate that there was mortality.

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(neon$collectDate)
```

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

```
#class "factor". Must change to date
neon$dateDate <- as.Date(neon$collectDate, format = "%Y-%m-%d")
class(neon$dateDate)
```

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

```
unique(neon$dateDate)
```

```
## [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(neon$plotID)
```

```
##  [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
```

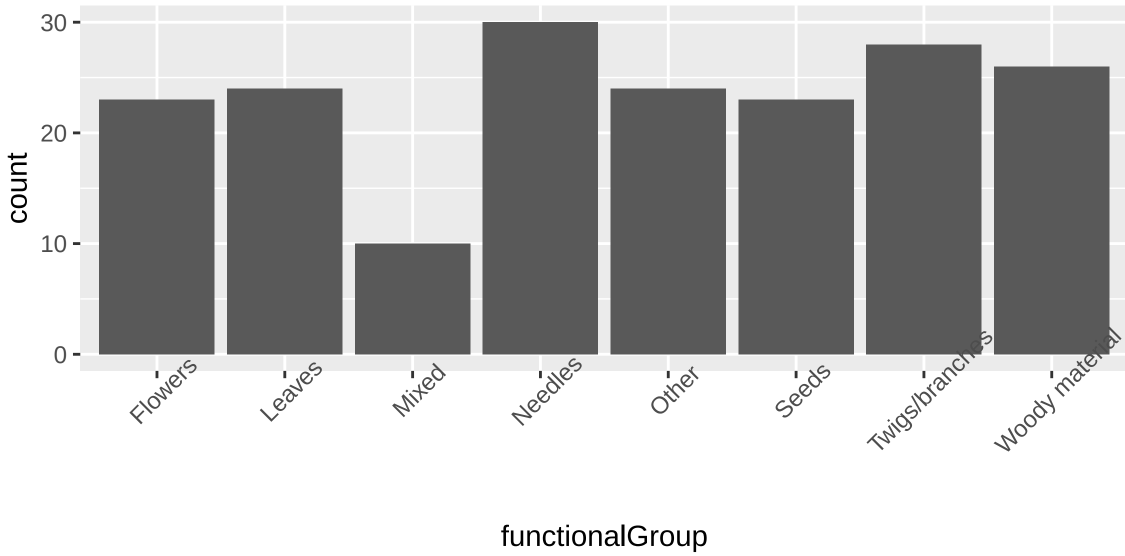
```
summary(neon$plotID)
```

```
## 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: The 'unique' function provides the plot IDs sampled at Niwot ridge. While summary also provides the plot IDs, it also provides the summary or number of instances of that plot ID.

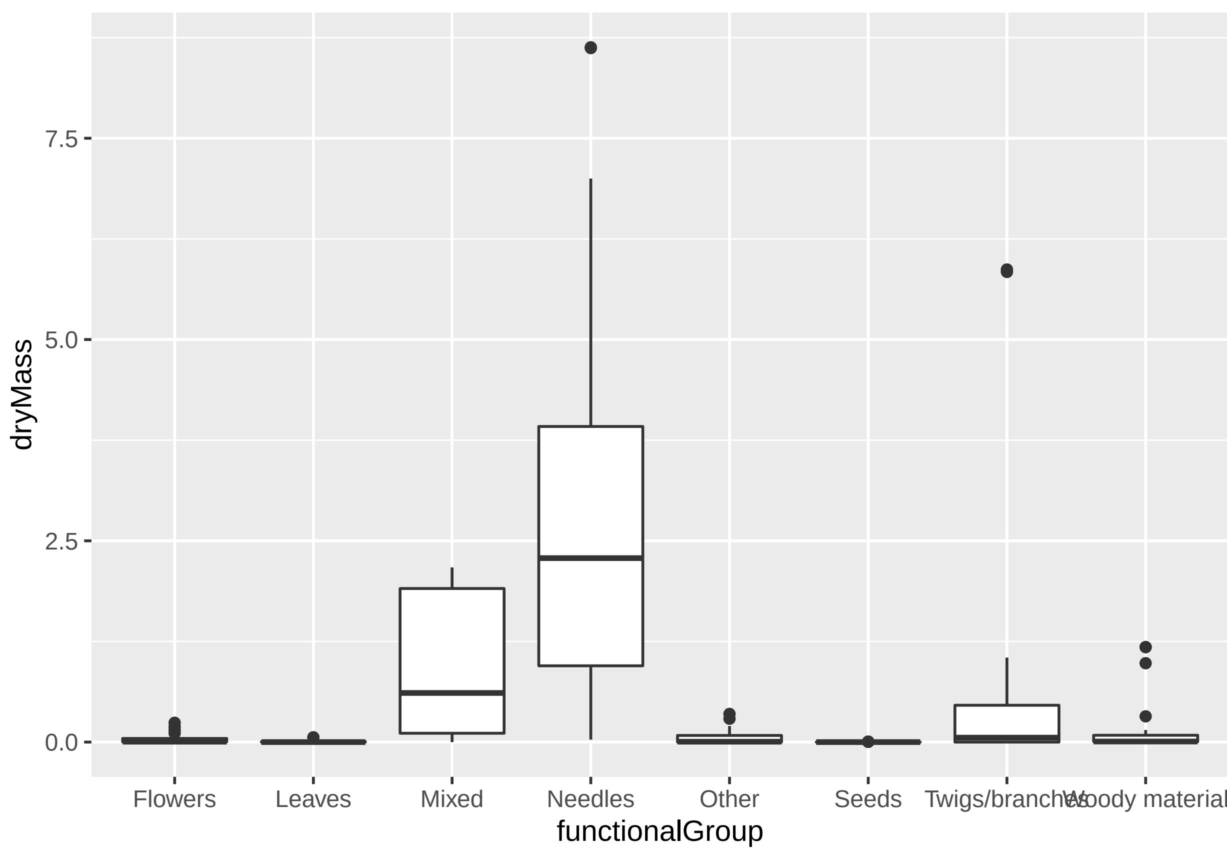
14. Create a bar graph of functionalGroup counts. This shows you what type of litter is collected at the Niwot Ridge sites. Notice that litter types are fairly equally distributed across the Niwot Ridge sites.

```
ggplot(neon, aes(x = functionalGroup)) +
  theme(axis.text.x = element_text(angle = 45)) +
  geom_bar()
```



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

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



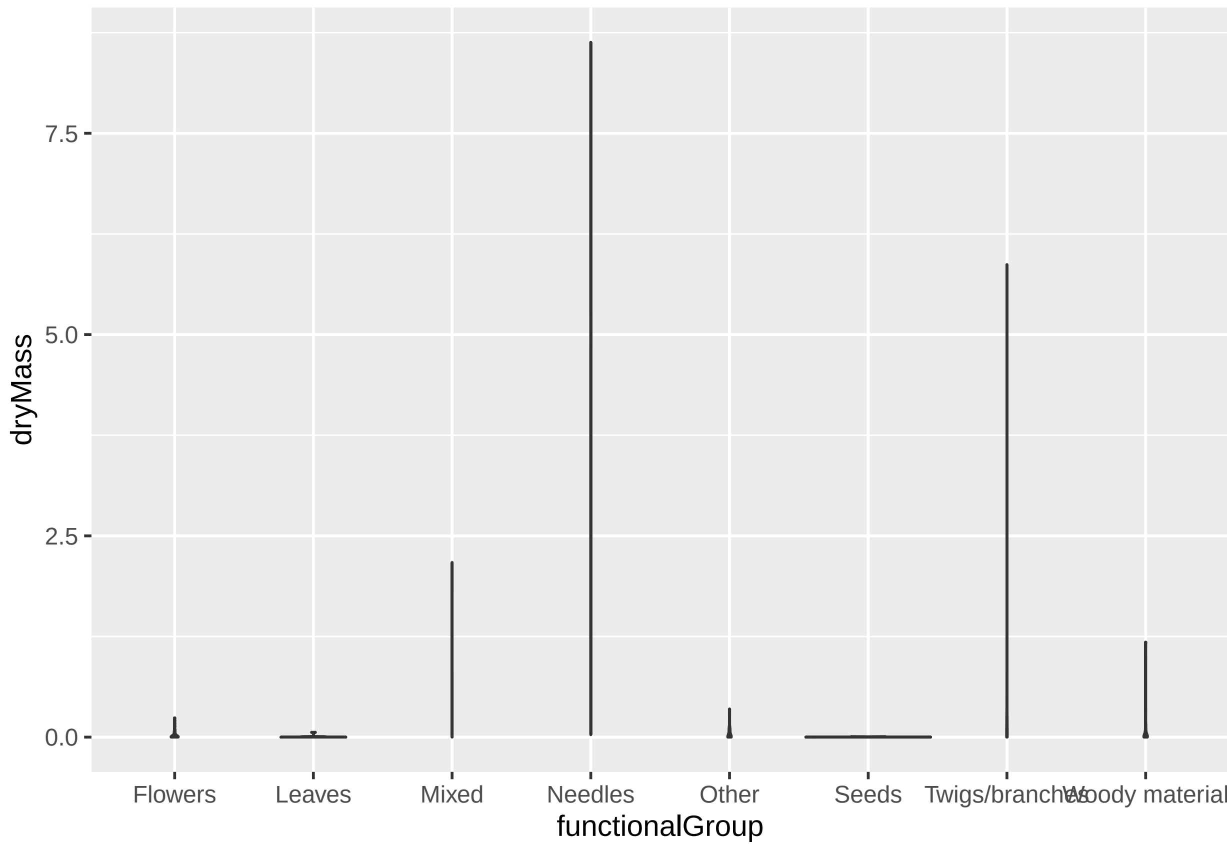

```
#ggplot(USGS.flow.data) +
  #geom_boxplot(aes(x = gage.height.mean, y = discharge.mean, group = cut_width(gage.height.mean, 1)))
```

```
#
ggplot(neon) +
  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 boxplot can show the outliers and range of the functional group. The violin plot is heavily stretched.

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

> Answer: Needles