Assignment 3: Data Exploration

Abhay V Rao, Section #4

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.**

```
#
getwd()
```

```
## [1] "C:/Users/av241/Documents/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: Neonicotinoids are systemic chemicals absorbed into plants, and can be present in pollen and nectar. These chemicals are less acutely toxic to mammals and other vertebrates, but are

highly toxic to pollinating insects.The long-lasting presence of neonicotinoids in plants raise concerns on their long-term impacts on pollinators.

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: The process of decomposition releases the carbon in litter and woody debris; hence studying these can give researchers insight into carbon emissions and net primary productivity.

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: *Samples are collected in randomly selected tower plot locations.* The size and number of tower plots surveyed depend on the stature of vegetation, and other characteristics. *Ground traps and elevated traps are installed, which are surveyed in the former case, once a year; in the latter, at more frequent intervals which are determined by the nature of the woodland, among other factors.

Obtain basic summaries of your data (Neonics)

5. What are the dimensions of the dataset?

```
dim(Neonics) # there are 4,623 objects of 30 variables.
```

```
## [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: The most studied effects are population (1803), followed by mortality (1493). This is likely because the researchers are most interested on the impact of neonics on insect populations, and their consequences on insect mortality.

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.Group)
```

```
##                                          Insects/Spiders
##                                                     3569
##                       Insects/Spiders; Standard Test Species
##                                                       27
```

```
## Insects/Spiders; Standard Test Species; U.S. Invasive Species
##                                                           667
##                        Insects/Spiders; U.S. Invasive Species
##                                                           360
```

```
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 top six species include five bee species, indicating that the researchers prioritize

gaining insight into the effects on 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"
```

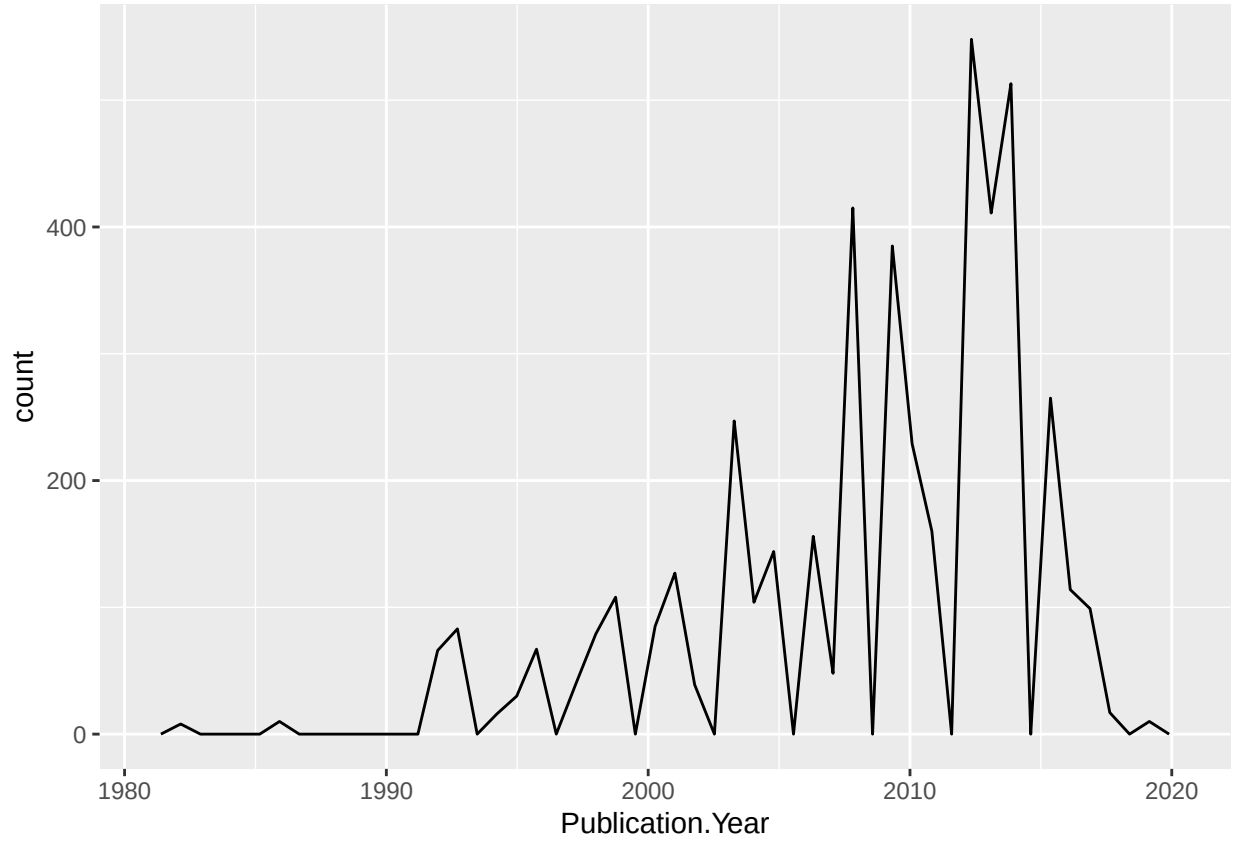
Answer: The concentration class is a factor.It appears that it is not considered to be numeric because some of the data includes characters, and symbols like "/". The characters are presumably included to express "no result" or a related message.

Explore your data graphically (Neonics)

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

```
p0 <- ggplot(Neonics) +
  geom_freqpoly(aes(x = Publication.Year), bins = 50
  )

print(p0)
```

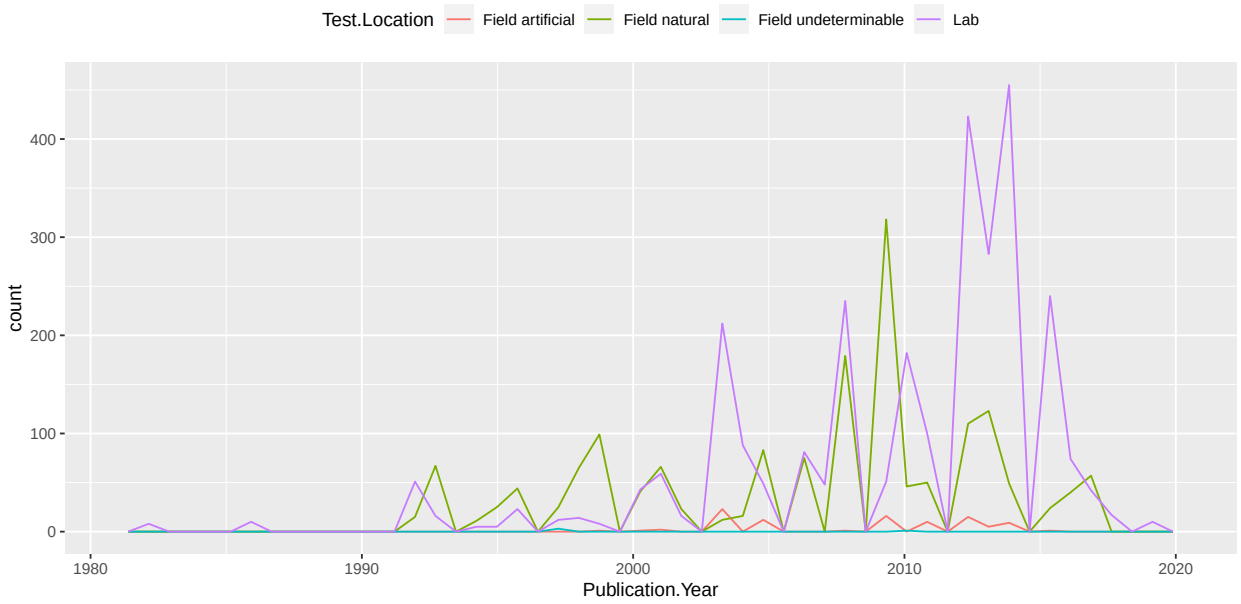


10. Reproduce the same graph but now add a color aesthetic so that different Test.Location are displayed as different colors.

```
#
p1 <- ggplot(Neonics) +
  geom_freqpoly(aes(x = Publication.Year, color = Test.Location), bins = 50) +
```

```
  theme(legend.position = "top")

print(p1)
```

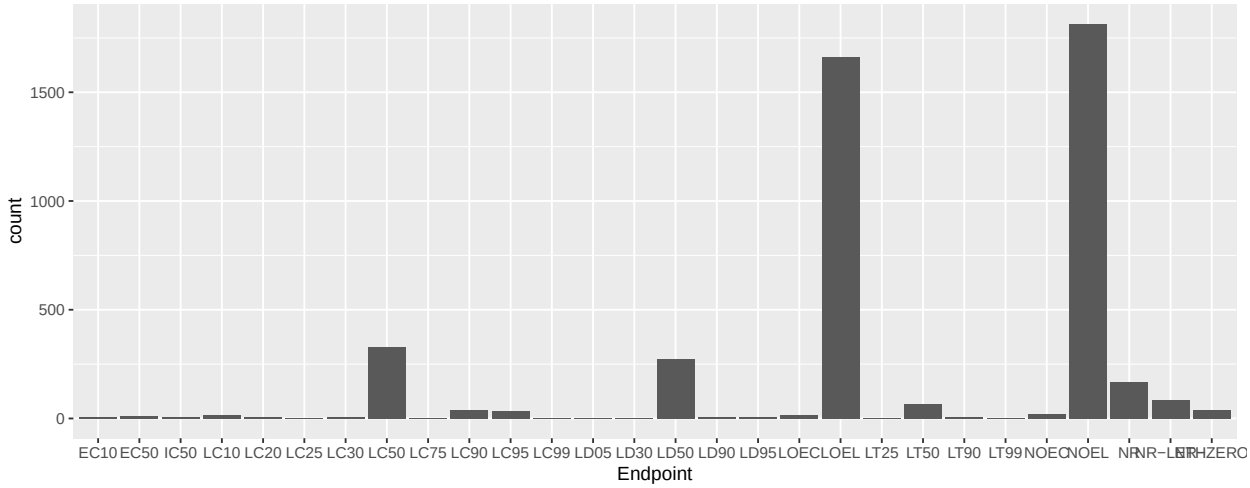


Interpret this graph. What are the most common test locations, and do they differ over time?

> Answer: 'Laboratory' and 'field natural' are the most common test locations, and they do vary over time; although the peaks - and therefore counts - of conditions associated with lab tests, tend to be higher, especially after the year 2000.

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.

```
#
p3 <- ggplot(Neonics)+
  geom_bar(aes(x= Endpoint))
print(p3)
```



```
#
summary(Neonics$Endpoint)
```

```
##    EC10    EC50    IC50    LC10    LC20    LC25    LC30    LC50    LC75    LC90
##       6      11       6      15       5       1       6     327       1      37
##    LC95    LC99    LD05    LD30    LD50    LD90    LD95    LOEC    LOEL    LT25
##      36       2       1       1     274       6       7      17    1664       1
##    LT50    LT90    LT99    NOEC    NOEL      NR NR-LETH NR-ZERO
##      65       7       2      19    1816     167      86      37
```

Answer: The two most common endpoints are NOEL and LOEL: No observable effect level and lowest observable effect level.

NOEL is an endpoint signifying that the highest dose (concentration) produces effects not significantly different from responses of control.

LOEL is an endpoint signifying that the lowest dose (concentration) produces effects that *were* significantly different from responses of 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)
```

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

```
#
Litter$collectDate <- as.Date(Litter$collectDate, format = "%y/%m/%d")
```

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, incomparables = FALSE)
```

```
##  [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)
```

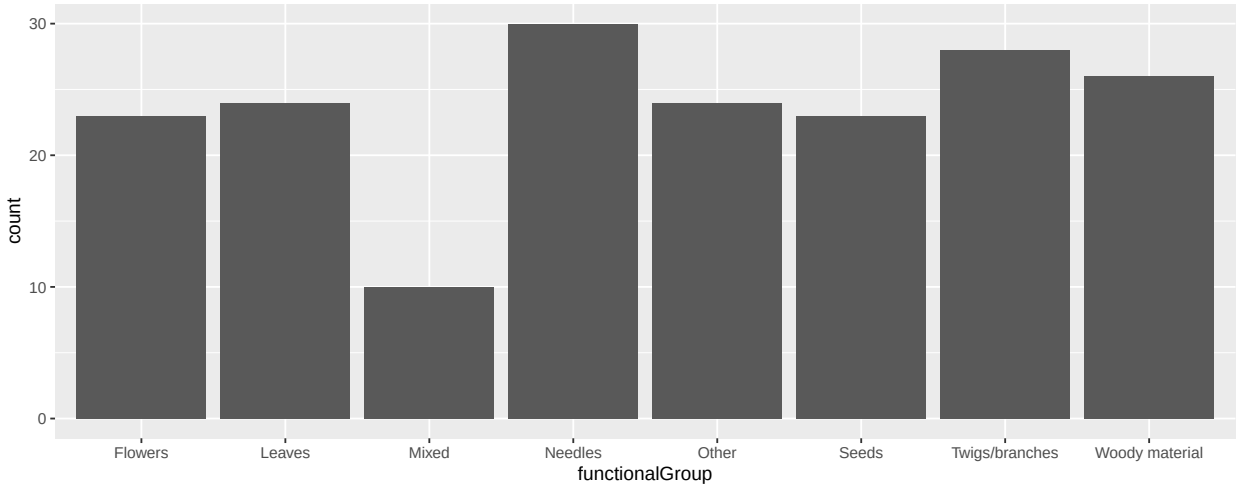
```
## NIWO_040 NIWO_041 NIWO_046 NIWO_047 NIWO_051 NIWO_057 NIWO_058 NIWO_061
##       20       19       18       15       14        8       16       17
## NIWO_062 NIWO_063 NIWO_064 NIWO_067
##       14       14       16       17
```

Answer: The unique function tells us that there are 12 levels or types of plots listed. The summary function organizes the data by a given plot - and is useful in determining *how many* data points exist for a given plot.

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.

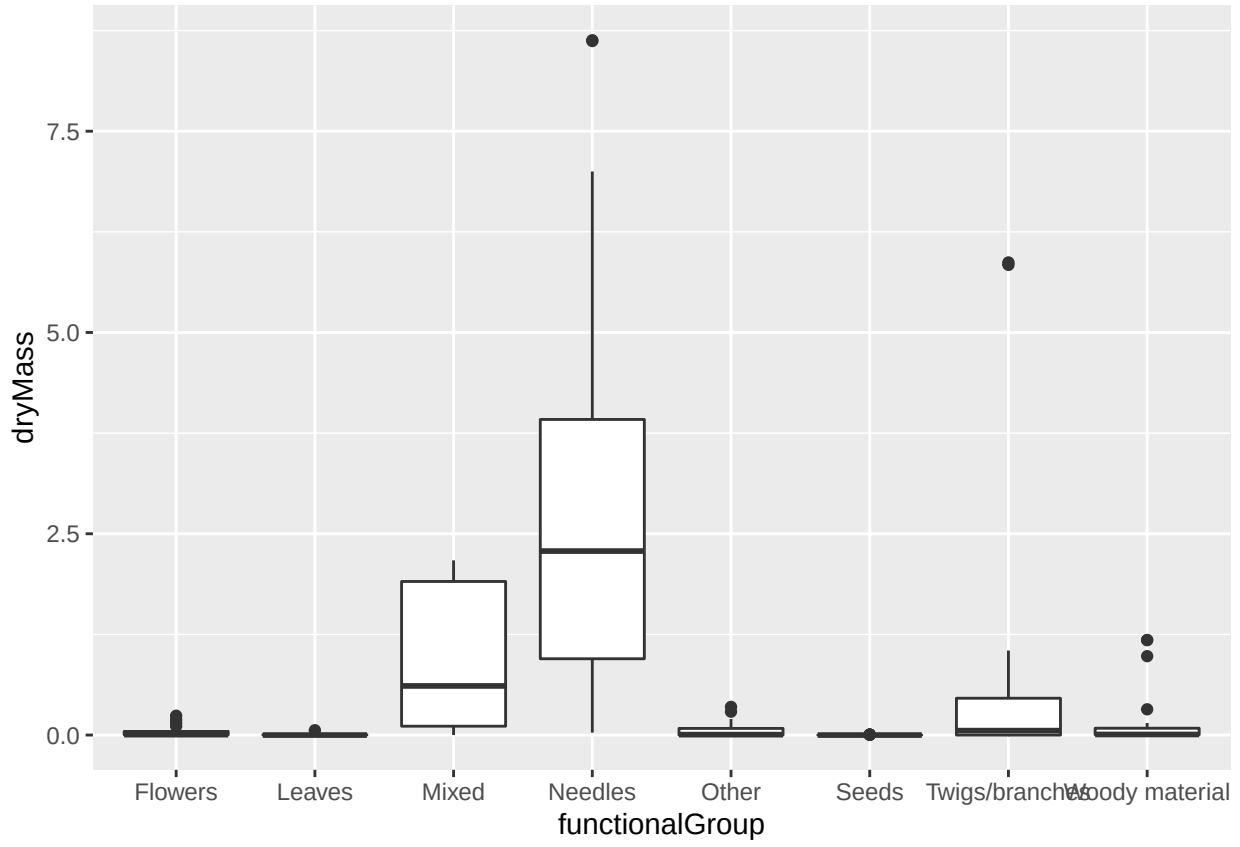
```
#
p4 <- ggplot(Litter) +
```

```
  geom_bar(aes(x= functionalGroup))
print(p4)
```

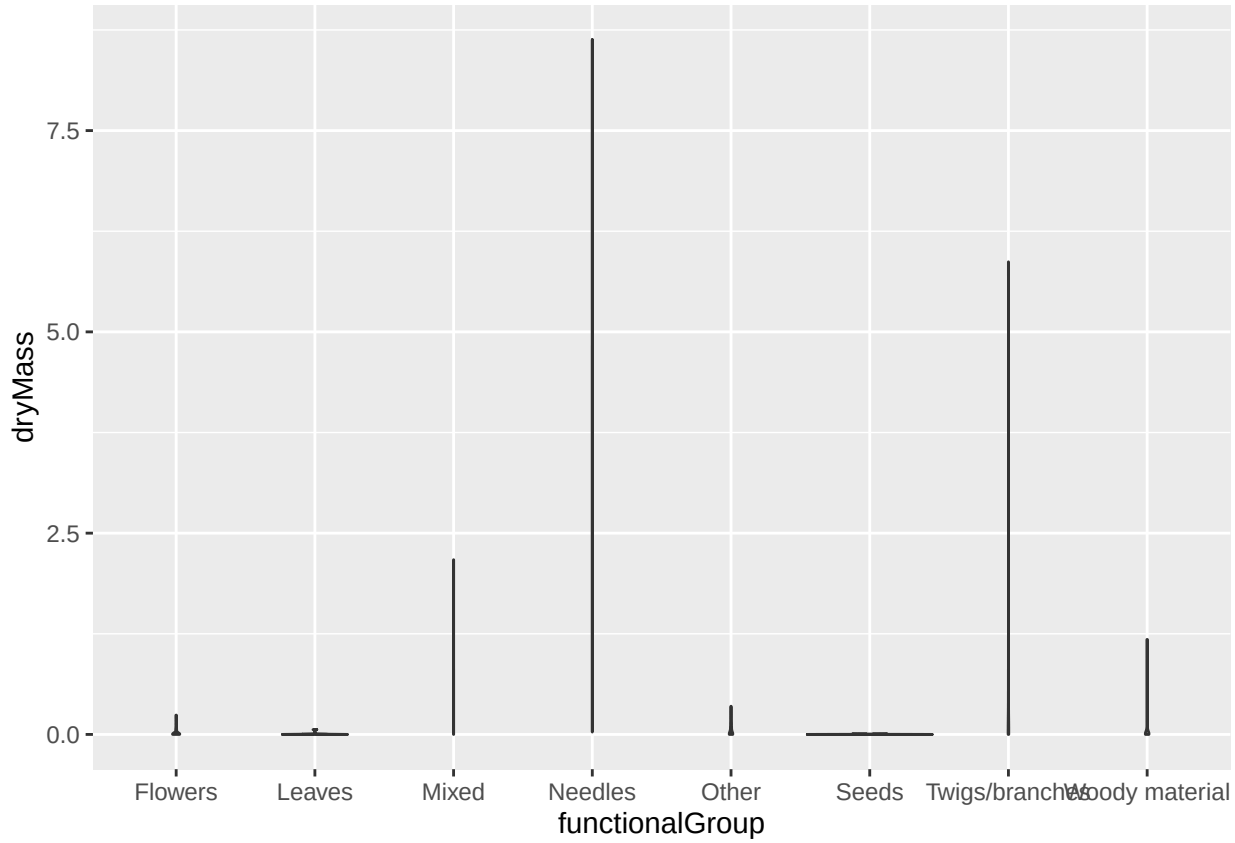


15. Using `geom_boxplot` and `geom_violin`, create a boxplot and a violin plot of dryMass by functionalGroup.

```
p5 <- ggplot(Litter)+
  geom_boxplot(aes(x=functionalGroup, y=dryMass))
print(p5)
```



```
p6 <- ggplot(Litter)+
  geom_violin(aes(x=functionalGroup, y=dryMass))
print(p6)
```



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

> Answer: The boxplot shares more insight that the violin plot, because the data is sparsely distributed within categories such as mixed, needles, twigs and woody material.

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

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