Assignment 3: Data Exploration

Jess Ozog, 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.**

```
# load tidyverse
library(tidyverse)

# working directory
setwd("C:/ENV872/Environmental_Data_Analytics_2022")

# read in the data and rename
Neonics <- read.csv("./Data/Raw/ECOTOX_Neonicotinoids_Insects_raw.csv",
                    stringsAsFactors = TRUE, header = TRUE)
Litter <- read.csv("./Data/Raw/NEON_NIWO_Litter_massdata_2018-08_raw.csv",
                   stringsAsFactors = TRUE, header = 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: Neonicotinoid insecticides may have different effects on different species, which may result in non-target insects being effected by the insecticide.

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: It is important to understand the litter and woody debris that is on the forest floor because both provide important habitats for terrestrial organisms, can influence sediment and water transport, and assist with nutrient cycling.

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 sites much have woody vegetation more than 2m tall. * Sampling can only be conducted in tower plots. * Site vegetation determines the frequency at which traps are sampled, with deciduous sites being sampled more frequently than evergreen sites.

Obtain basic summaries of your data (Neonics)

5. What are the dimensions of the dataset?

```
dim(Neonics)
```

```
## [1] 4623   30
```

```
# Dataset dimensions: 4623 rows and 30 columns
```

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 common effects studied were population (1803), mortality (1493), behavior (360), and feeding behavior (255). These effects are of interest because they all relate to the survival of the insects being studied and could be devastating to the population.

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.

```
sort(summary(Neonics$Species.Common.Name))
```

```
##                         Ant Family                         Apple Maggot
##                                  9                                    9
##              Glasshouse Potato Wasp                            Lacewing
##                                 10                                   10
##             Southern House Mosquito             Two Spotted Lady Beetle
##                                 10                                   10
##           Spotless Ladybird Beetle                 Braconid Parasitoid
##                                 11                                   12
##                        Common Thrip        Eastern Subterranean Termite
##                                 12                                   12
##                              Jassid                           Mite Order
##                                 12                                   12
##                           Pea Aphid                    Pond Wolf Spider
##                                 12                                   12
##              Armoured Scale Family                    Diamondback Moth
##                                 13                                   13
##                       Eulophid Wasp                   Monarch Butterfly
##                                 13                                   13
##                       Predatory Bug                Yellow Fever Mosquito
##                                 13                                   13
##                        Corn Earworm                   Green Peach Aphid
##                                 14                                   14
##                           House Fly                           Ox Beetle
##                                 14                                   14
##                  Red Scale Parasite                  Spined Soldier Bug
##                                 14                                   14
##              Western Flower Thrips Hemlock Woolly Adelgid Lady Beetle
##                                 15                                   16
##              Hemlock Wooly Adelgid                                Mite
##                                 16                                   16
##                         Onion Thrip               Araneoid Spider Order
##                                 16                                   17
##                            Bee Order                      Egg Parasitoid
##                                 17                                   17
##                        Insect Class            Moth And Butterfly Order
##                                 17                                   17
##       Oystershell Scale Parasitoid           Black-spotted Lady Beetle
##                                 17                                   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
##                       Codling Moth          Flatheaded Appletree Borer
##                                 19                                   20
```

```
##              Horned Oak Gall Wasp                Leaf Beetle Family
##                                20                                20
##                 Potato Leafhopper         Tooth-necked Fungus Beetle
##                                20                                20
##                     Argentine Ant                            Beetle
##                                21                                21
##                         Mason Bee                           Mosquito
##                                22                                22
##                  Citrus Leafminer                  Ladybird Beetle
##                                23                                23
##                 Spider/Mite Class                Tobacco Flea Beetle
##                                24                                24
##                       Chalcid Wasp             Convergent Lady Beetle
##                                25                                25
##                     Stingless Bee               Ground Beetle Family
##                                25                                27
##                Rove Beetle Family                     Tobacco Aphid
##                                27                                27
##                     Scarab Beetle                     Spring Tiphia
##                                29                                29
##                      Thrip Order             Ladybird Beetle Family
##                                29                                30
##                       Parasitoid                     Braconid Wasp
##                                30                                33
##                     Cotton Aphid                   Predatory Mite
##                                33                                33
##               Sweetpotato Whitefly                      Aphid Family
##                                37                                38
##                    Cabbage Looper              Buff-tailed Bumblebee
##                                38                                39
##                    True Bug Order            Sevenspotted Lady Beetle
##                                45                                46
##                      Beetle Order         Snout Beetle Family, Weevil
##                                47                                47
##                Erythrina Gall Wasp                  Parasitoid Wasp
##                                49                                51
##              Colorado Potato Beetle                    Parastic Wasp
##                                57                                58
##               Asian Citrus Psyllid                Minute Pirate Bug
##                                60                                62
##                  European Dark Bee                         Wireworm
##                                66                                69
##                     Euonymus Scale                Asian Lady Beetle
##                                75                                76
##                   Japanese Beetle                  Italian Honeybee
##                                94                               113
##                        Bumble Bee               Carniolan Honey Bee
##                               140                               152
##              Buff Tailed Bumblebee                  Parasitic Wasp
##                               183                               285
##                         Honey Bee                           (Other)
##                               667                               670
```

Answer: The most commonly studied species include: honey bee (n = 667), parasitic wasp (n =

285), buff tailed bumblebee (n = 183), carniolan honey bee (n = 152), bumble bee (n = 140), and italian honeybee (n = 113). All these species are pollinators and play a vital role in ecosystem function, as well as in agriculture.

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.)
```

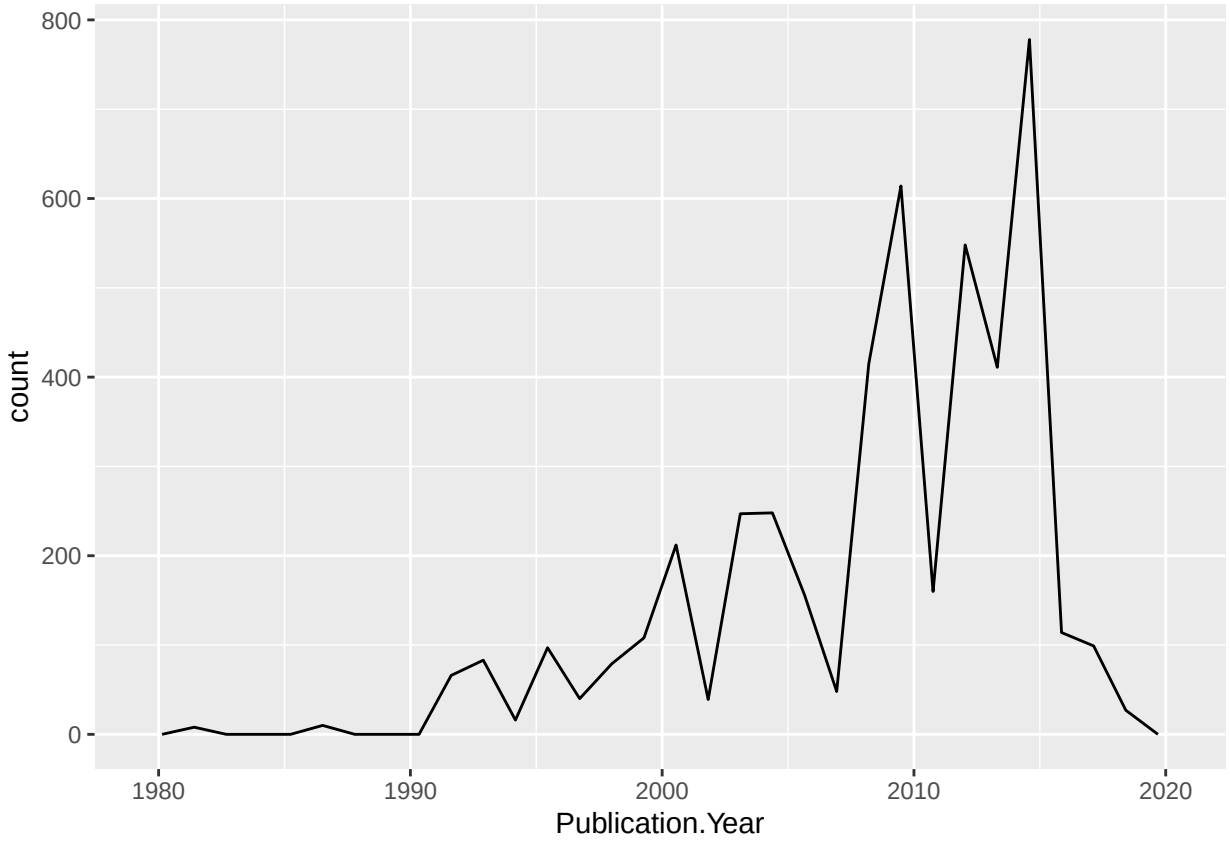
```
## [1] "factor"
```

Answer: Con.1..Author is a factor in this dataset. This is not numeric because there are other symbols in the data, such as backslashes (eg. 0.012/).

Explore your data graphically (Neonics)

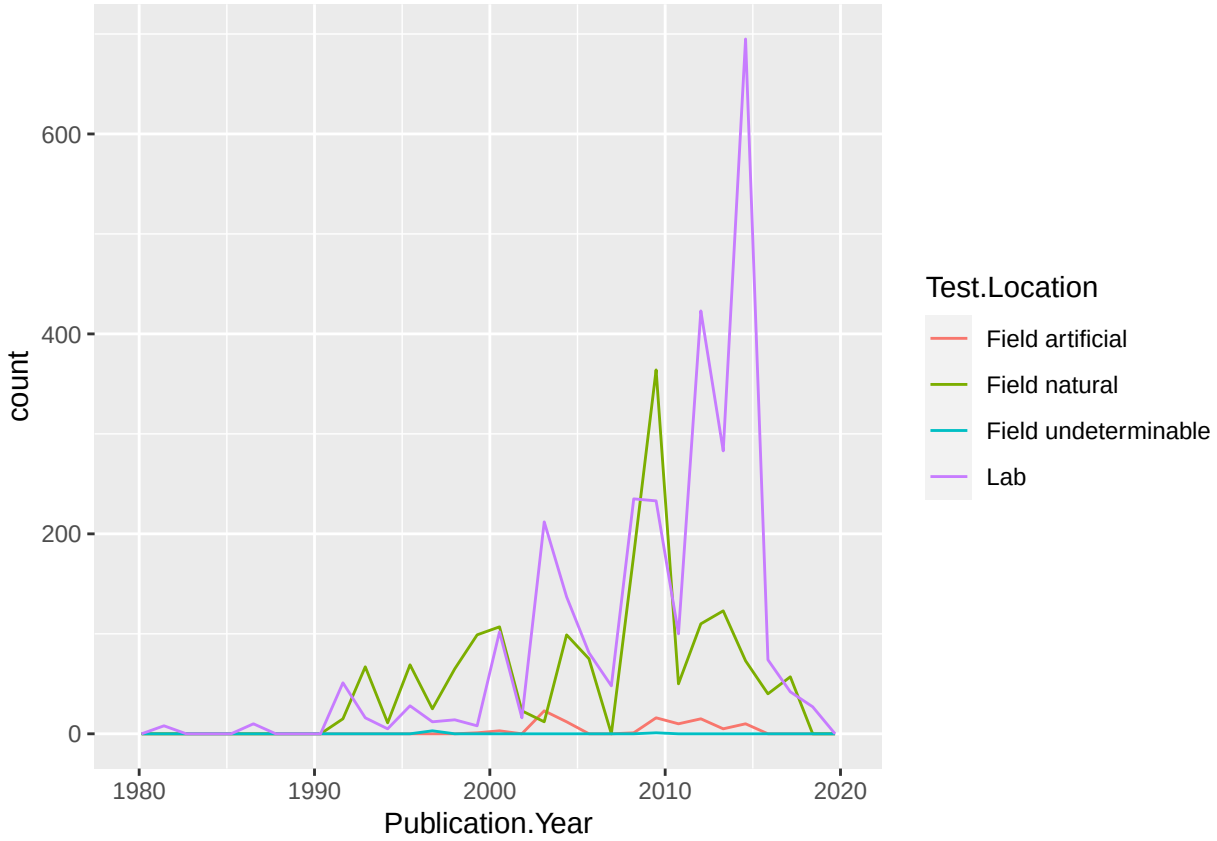
9. Using `geom_freqpoly`, generate a plot of the number of studies conducted by publication year.

```
ggplot(Neonics) +
  geom_freqpoly(aes(x=Publication.Year), bins = 30)
```



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 = 30)
```

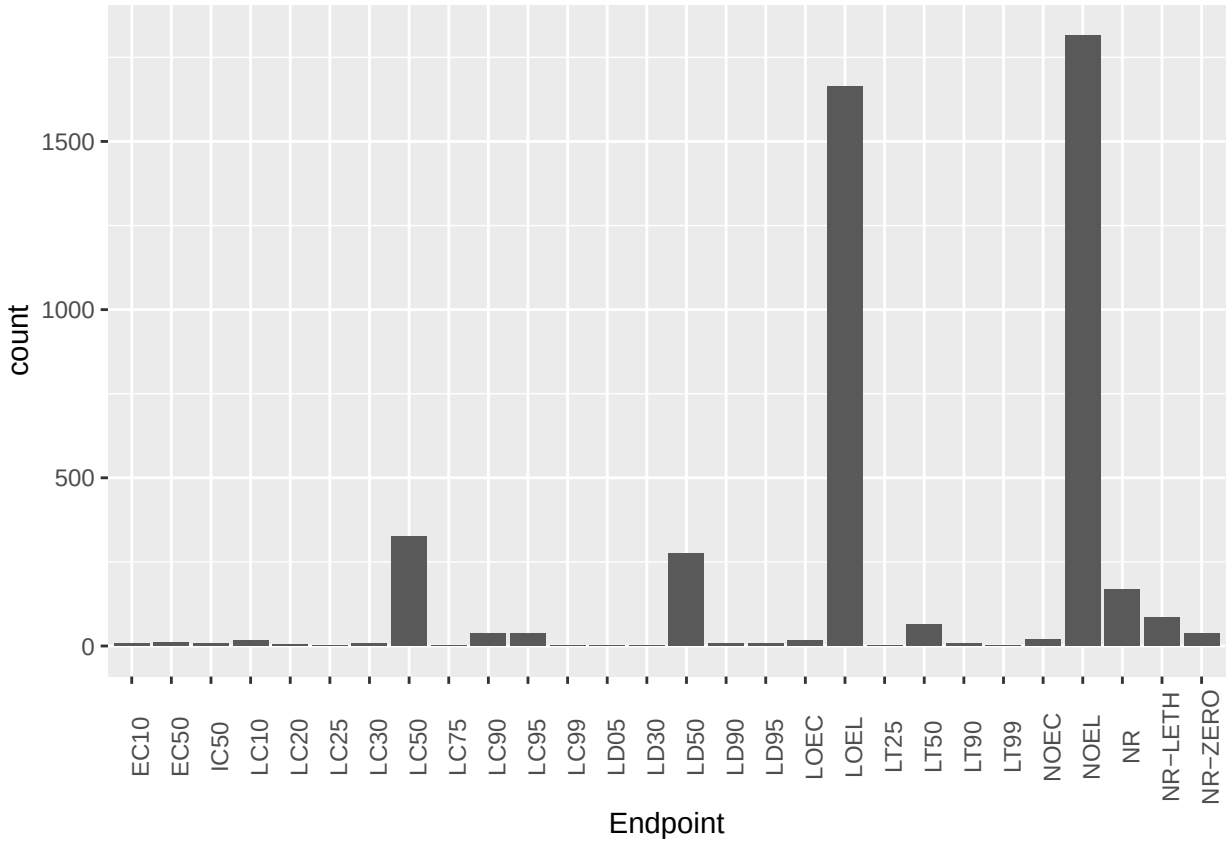


Interpret this graph. What are the most common test locations, and do they differ over time?

> Answer: The most common test locations are lab and field natural. Both locations started off on roughly the same path. Field natural peaked around 2010 and then began to decline, however, lab continued to increase and peaked around 2015. Both lab and natural field were low in late 2010's into 2020.

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)) +
  geom_bar() +
  theme(axis.text.x = element_text(angle = 90))
```



Answer: The two most common endpoints are NOEL and LOEL. NOEL is defined as "no observable effect level" and is used when the effects of the highest dose is not significantly different from the controls. LOEL is defined as "lowest observable effect level" and is used when the effects of the lowest dose is signficantly different from the 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"
```

```
# collectDate is a factor

# change to date format
Litter$collectDate <- as.Date(Litter$collectDate, format = "%Y-%m-%d")
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)
```

```
##  [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
```

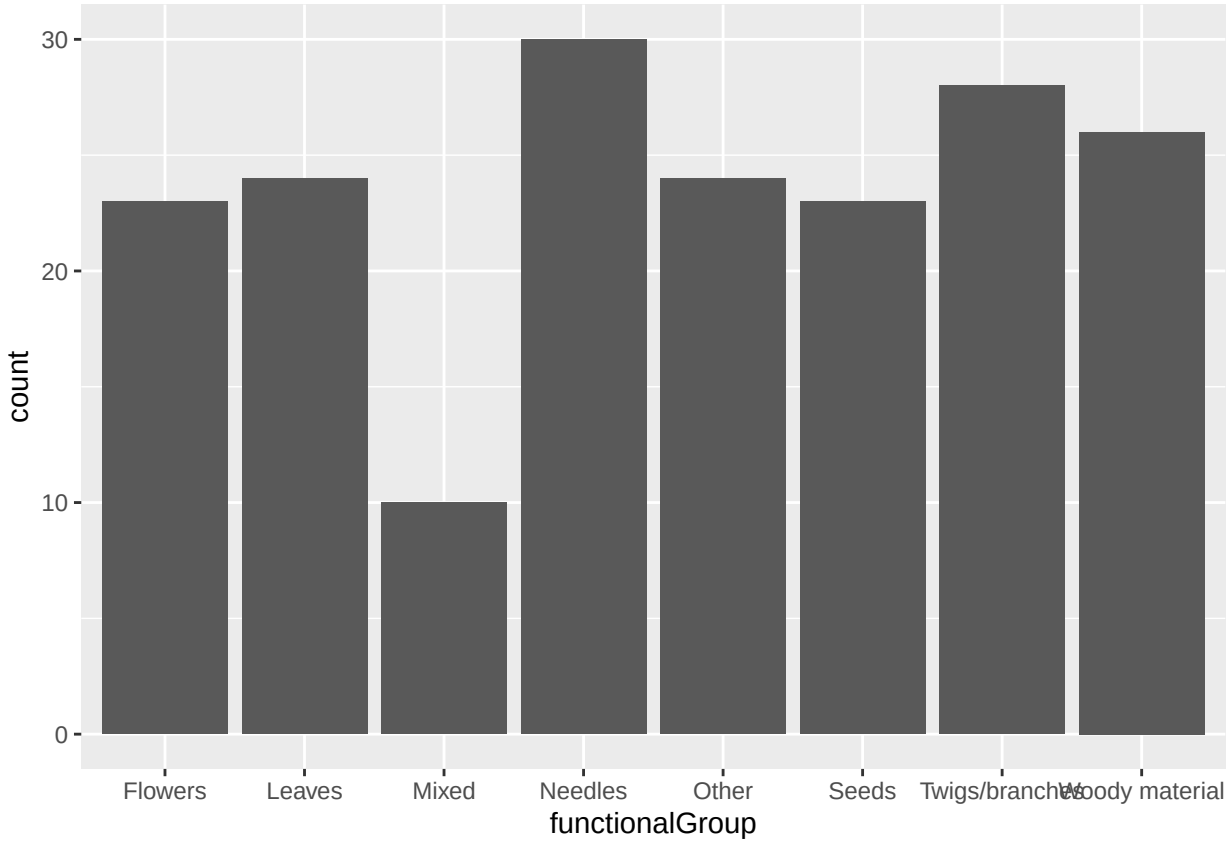
```
summary(Litter$plotID)
```

```
## 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. The unique function returns back each unique value that is present in the data, however, it does not indicate the amount that unique value appears in the data. The summary function returns back each unique value, as well as the total number of times that value appears in the data.

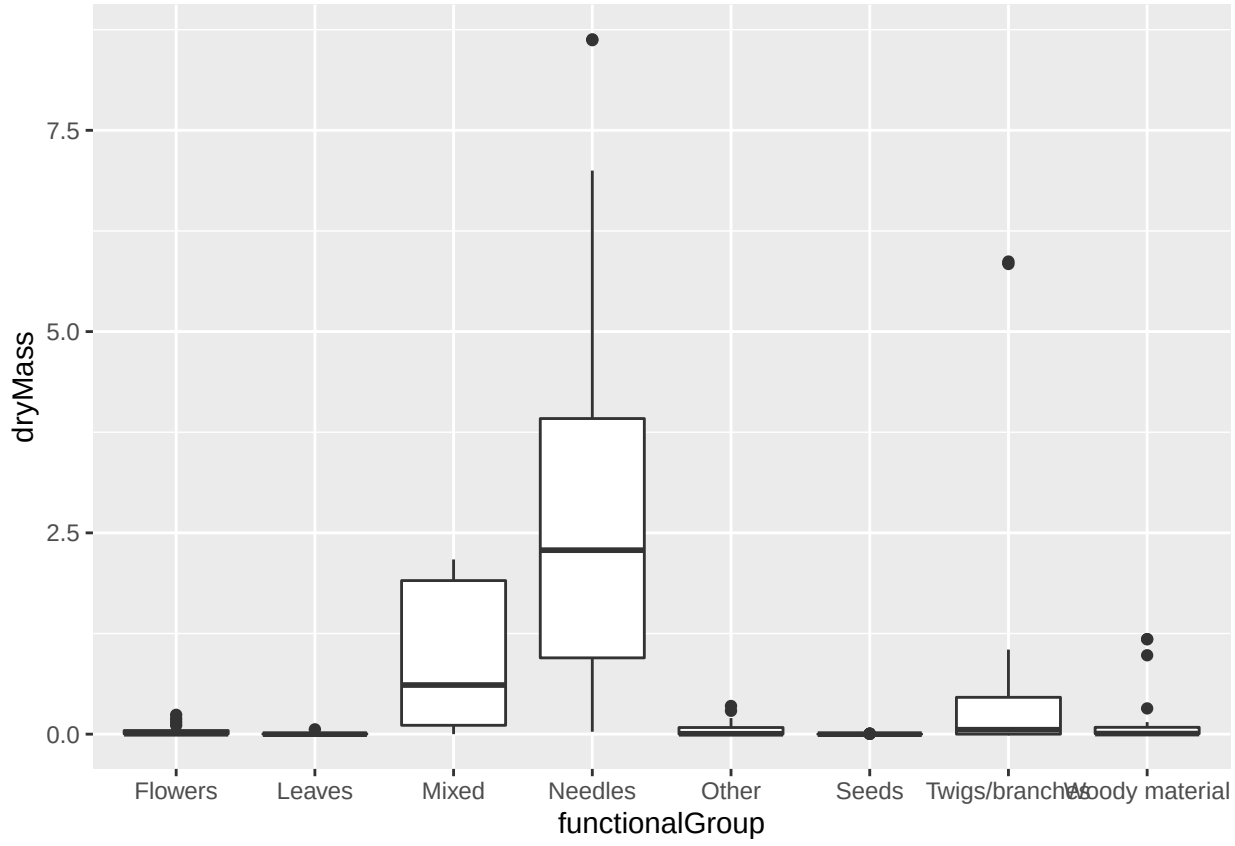
14. Create a bar graph of functionalGroup counts. This shows you what type of litter is collected at the Niwot Ridge sites. Notice that litter types are fairly equally distributed across the Niwot Ridge sites.

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

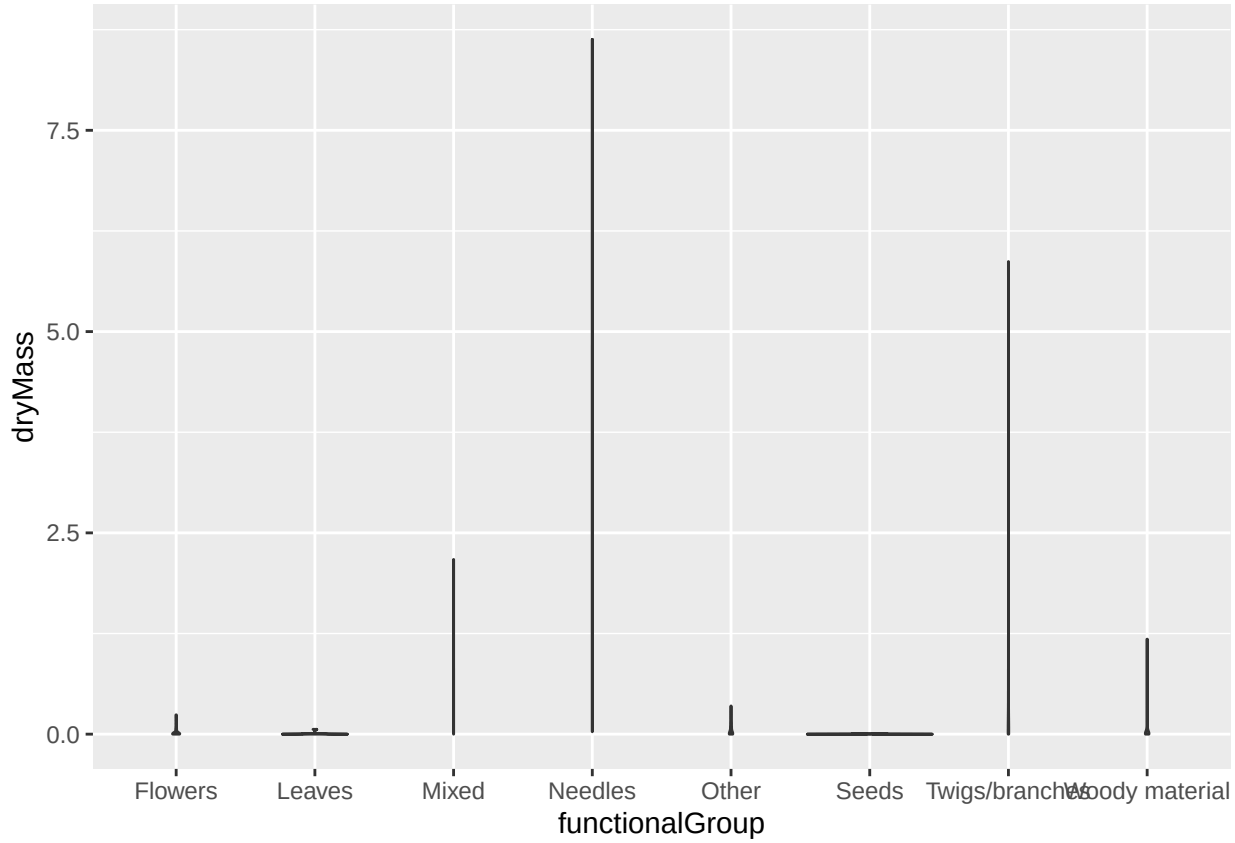


15. Using `geom_boxplot` and `geom_violin`, create a boxplot and a violin plot of dryMass by functionalGroup.

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




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



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

> Answer: A boxplot is more effective for visualization with this data because it allows us to see the IQR of the data, as well as the mean and if the data is skewed. Based on the data, the violin plot is not useful for showing the mean or the distribution of the data due to the plots showing up as lines.

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

> Answer: The litter types with the highest biomass are needles and mixed.