Assignment 3: Data Exploration

Jaimie Wargo

Fall 2023

OVERVIEW

This exercise accompanies the lessons in Environmental Data Analytics on Data Exploration.

Directions

1. Rename this file `<FirstLast>_A03_DataExploration.Rmd` (replacing `<FirstLast>` with your first and last name).
2. Change "Student Name" on line 3 (above) with your name.
3. Work through the steps, **creating code and output** that fulfill each instruction.
4. Assign a useful **name to each code chunk** and include ample **comments** with your code.
5. Be sure to **answer the questions** in this assignment document.
6. When you have completed the assignment, **Knit** the text and code into a single PDF file.
7. After Knitting, submit the completed exercise (PDF file) to the dropbox in Sakai.

TIP: If your code extends past the page when knit, tidy your code by manually inserting line breaks.

TIP: If your code fails to knit, check that no `install.packages()` or `View()` commands exist in your code.

Set up your R session

1. Check your working directory, load necessary packages (tidyverse, lubridate), 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 include the subcommand to read strings in as factors.

```
library(tidyverse)
library(lubridate)
library(here)
```

```
## Warning: package 'here' was built under R version 4.2.3
```

```
#Loading in useful libraries

#Loading in data and assigning to objects using the here() library, ensuring strings are factors.
Neonics <- read.csv(here('Data','Raw', 'ECOTOX_Neonicotinoids_Insects_raw.csv'),
```

```
                    stringsAsFactors = TRUE)
Litter <- read.csv(here('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 ecotoxicology of neonicotinoids on insects? Feel free to do a brief internet search if you feel you need more background information.

   Answer: Insects are a vital part of our ecosystem and have a variety of purposes, both in terms of human needs and for the environment in general. The impact of neonicotinoids on the health of insects could have drastic impacts on the environment, including human life, so it is important to understand and quantify these effects to regulate use of this product if necessary.

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 could relate to the frequency and intensity of wildfires. Additionally, litter can also be habitat for certain smaller creatures, like insects or even small birds and reptiles, so understanding the prevalence of it can potentially be used to estimate population viability.

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:

   1. Sampling is conducted at sites with woody plant greater than 2m tall
   2. One elevated trap and one ground trap is deployed per 400m2 area.
   3. Ground traps are sampled once per year, and elevated trap sampling varies based on vegetation type.

Obtain basic summaries of your data (Neonics)

5. What are the dimensions of the dataset?

```
dim(Neonics)
```

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

```
#dim() gives the number of rows and columns, in this case 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
```

```
#shows the number of observations for each factor
```

Answer: Population and Mortality are the most common effects studied. This makes sense because we would be interested in seeing the effect of the agrochemicals on the ability of these insects to live and maintain a steady population, as there would be negative ecosystem impacts if we significantly impacted this.

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.[TIP: The `sort()` command can sort the output of the summary command...]

```
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
```

```
#shows the number of observations for each factor
```

Answer: The most commonly studied species are the honey bee, parasitic wasp, buff tailed bumblebee, Carniolan honey bee, bumble bee, and Italian honey bee. These are all pollinator species. If pollinator species are negatively impacted by these chemicals, there will be direct impacts to our food production and entire agriculture system.

8. Concentrations are always a numeric value. What is the class of `Conc.1..Author.` column in the dataset, and why is it not numeric?

```
class(Neonics$Conc.1..Author.)
```

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

```
#shows the data type of the vector
```

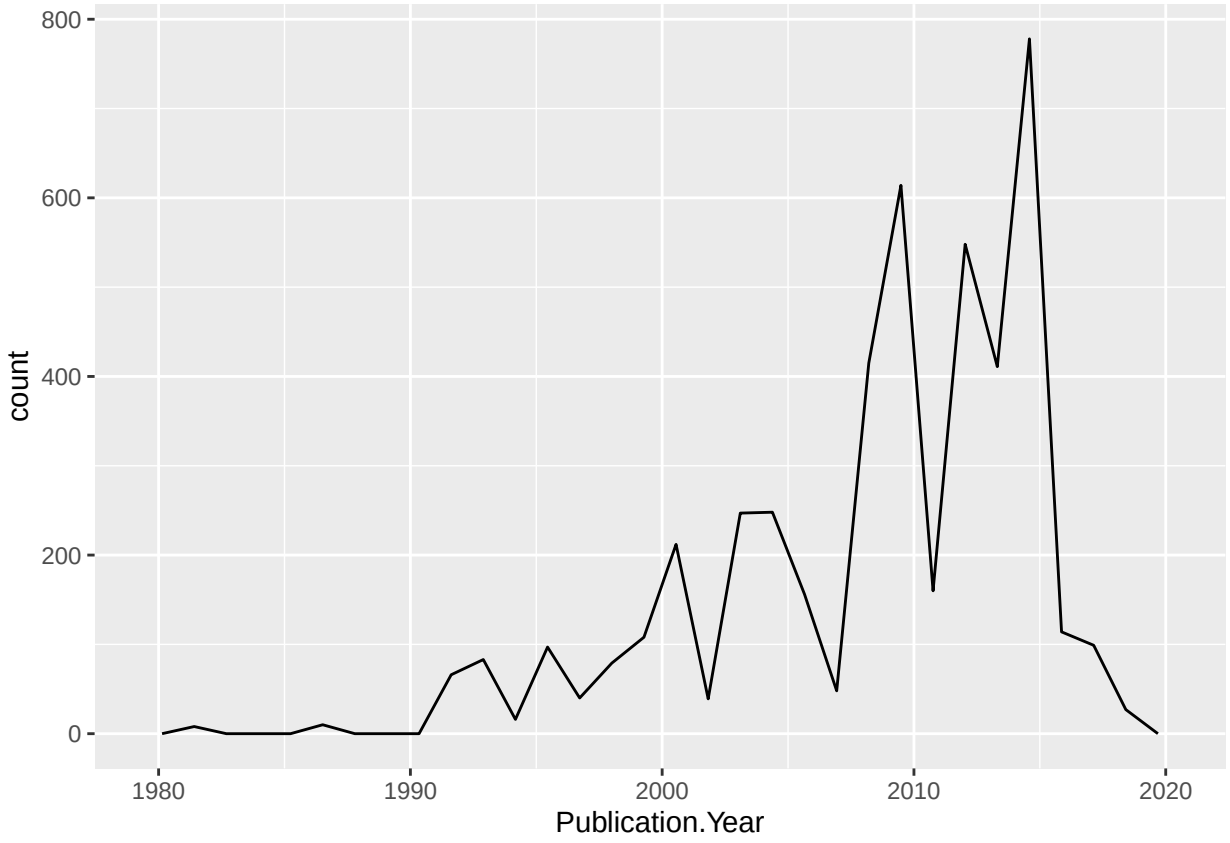
Answer: This vector is not numeric because there are values within it that contain character values, such as ~, /, and NR. Vectors can only have one data type, so it defaults to the most universal, which is character– more specifically, the class is factor since we changed string to factor when loading in the data frame.

Explore your data graphically (Neonics)

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

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

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

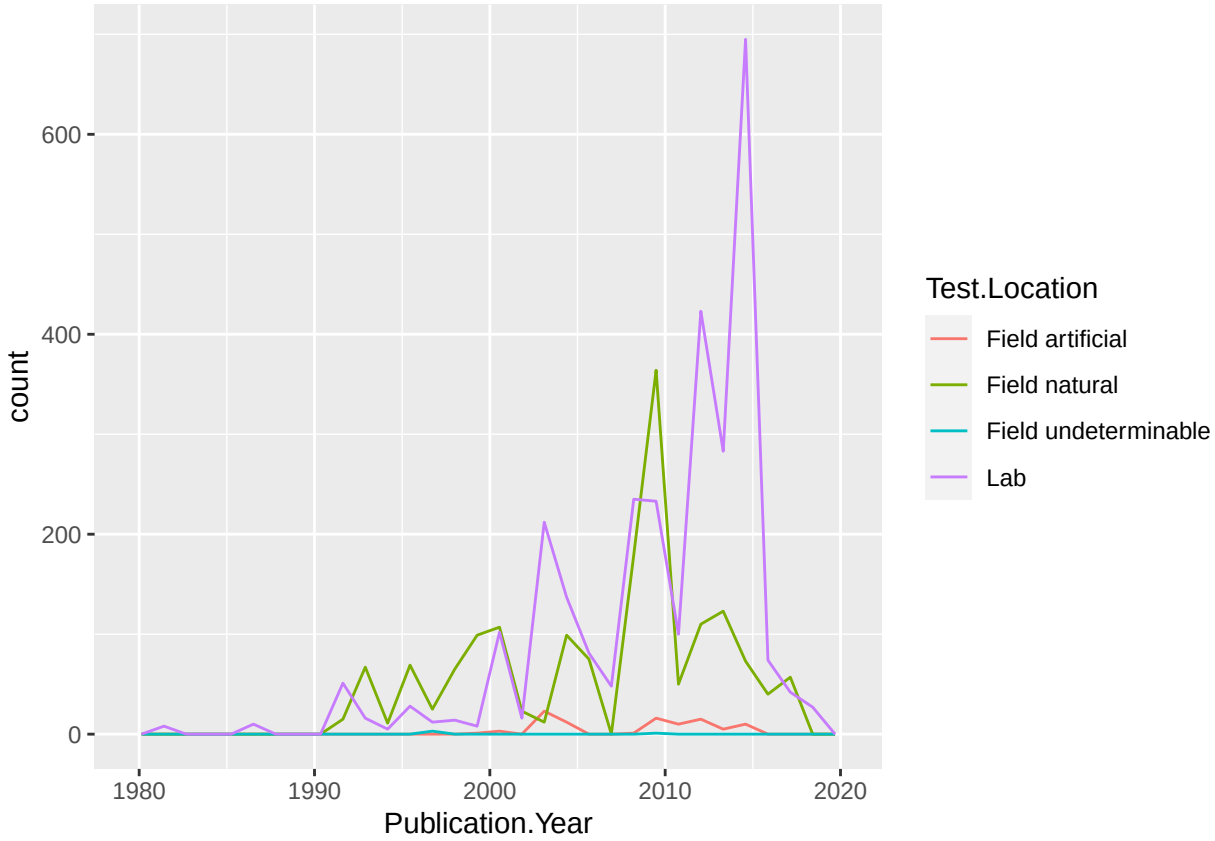



```
#I used the ggplot function and specified the data as the Neonics df and that I wanted the x value to b
```

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

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

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




```
#Using color= in the aes() argument does the requested task
```

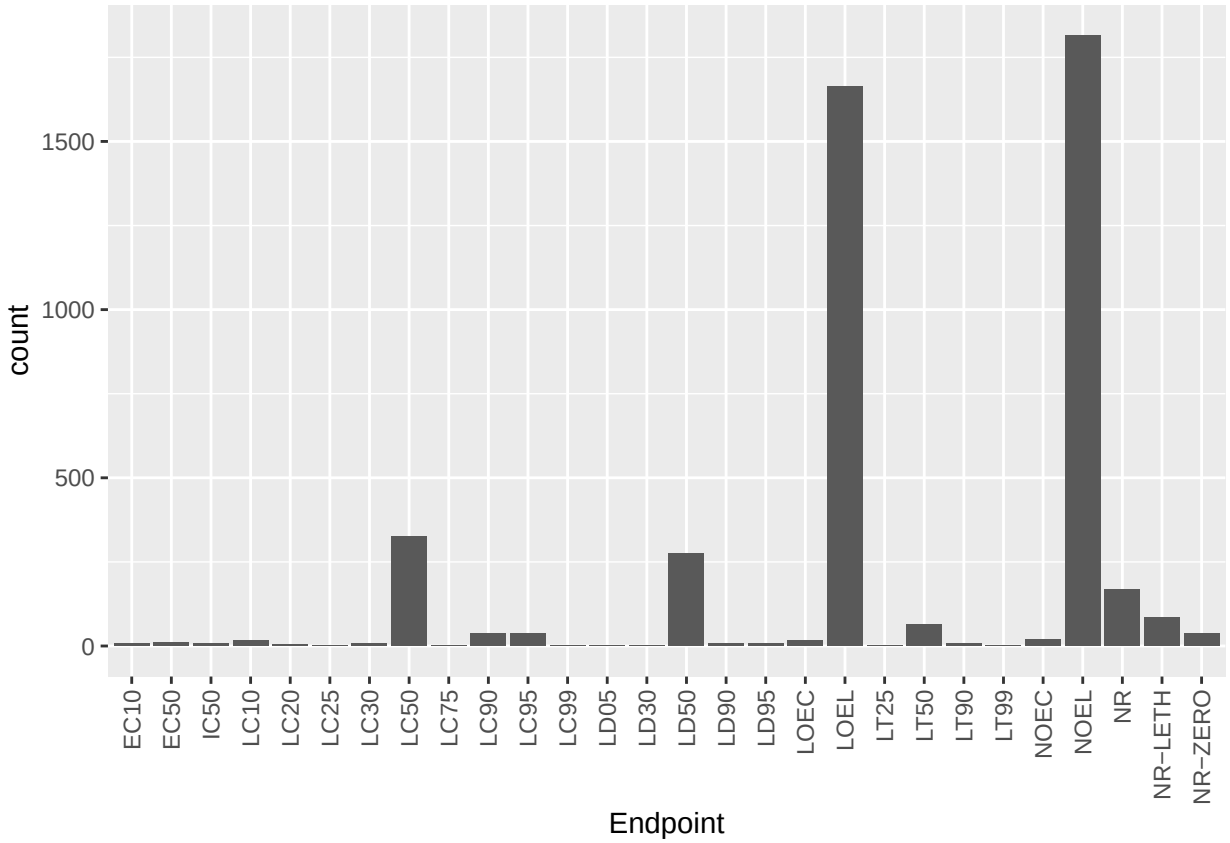
Interpret this graph. What are the most common test locations, and do they differ over time?

> Answer: The most common test locations are a natural field and in the lab. Lab tests have become the most prevalent in the last decade, but natural fields were more used in the late 2000s and between 1990 and 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.

[**TIP**: Add `theme(axis.text.x = element_text(angle = 90, vjust = 0.5, hjust=1))` to the end of your plot command to rotate and align the X-axis labels...]

```
ggplot(data=Neonics, aes(x=Endpoint))+
  geom_bar()+
  theme(axis.text.x = element_text(angle = 90, vjust = 0.5, hjust=1))
```




```
#geom_bar adds the barplot to the graph, displaying the counts of the Endpoint factor
```

Answer: The most common endpoints are LOEL and NOEL. LOEL is the lowest-observable-effect-level, which is the lowest dose of the chemical that produces an effect on the organism. NOEL is no-observable-effect-level, which is the highest dose able to be administered without observing a significant difference from the control.

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) #class is factor
```

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

```
Litter$collectDate <- ymd(Litter$collectDate)
#reassigning the vector to the appropriate date after it is transformed with ymd() from the lubridate p
class(Litter$collectDate) #class is now date
```

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

```r
unique(Litter$collectDate) #gives unique values listed in the vector
```

```
## [1] "2018-08-02" "2018-08-30"
```

```r
#Litter was collected on August 2 and August 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`?

```r
unique(Litter$plotID)
```

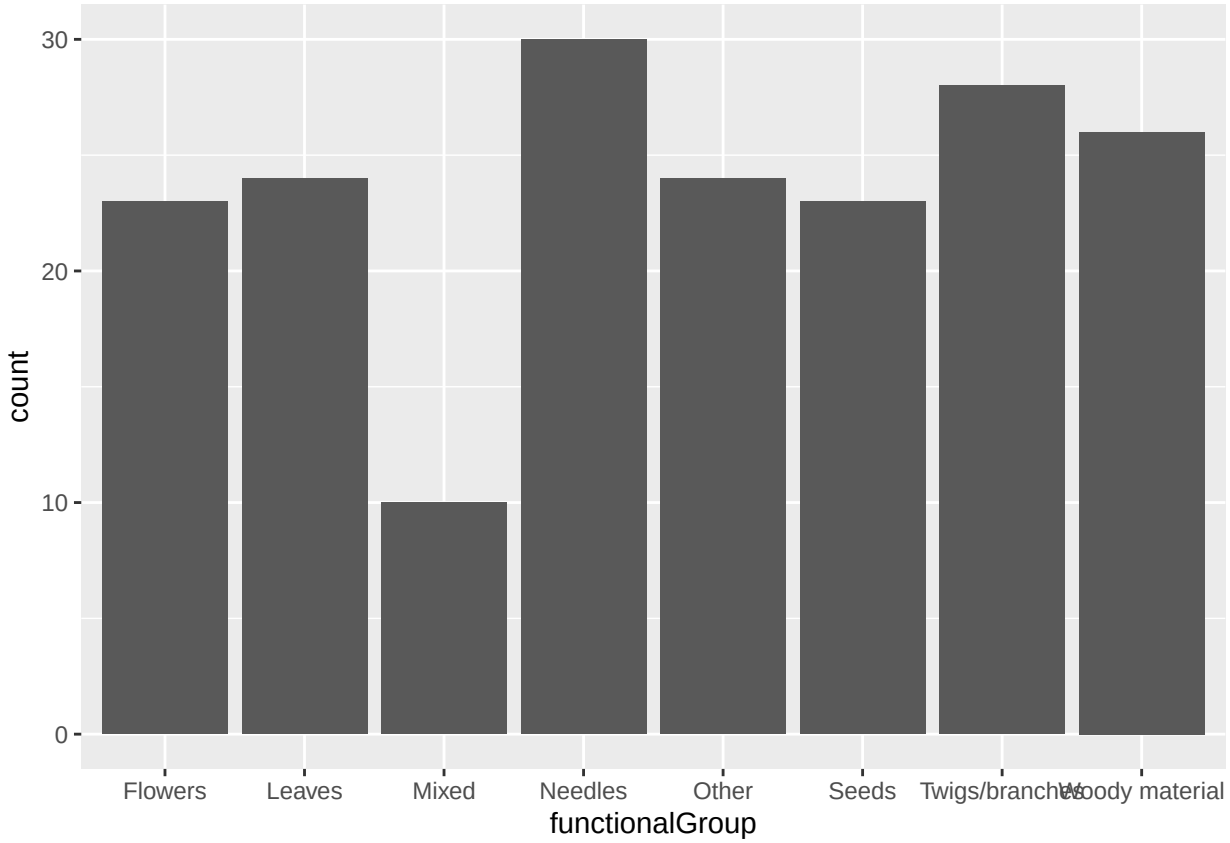
```
##  [1] NIWO_061 NIWO_064 NIWO_067 NIWO_040 NIWO_041 NIWO_063 NIWO_047 NIWO_051
##  [9] NIWO_058 NIWO_046 NIWO_062 NIWO_057
## 12 Levels: NIWO_040 NIWO_041 NIWO_046 NIWO_047 NIWO_051 NIWO_057 ... NIWO_067
```

```r
#shows the unique values in the vector
```

Answer: 12 plots were sampled. Summary shows the number of observations from each plotID, but unique shows just the different plot IDs. Unique also presents the levels associated with this vector.

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.

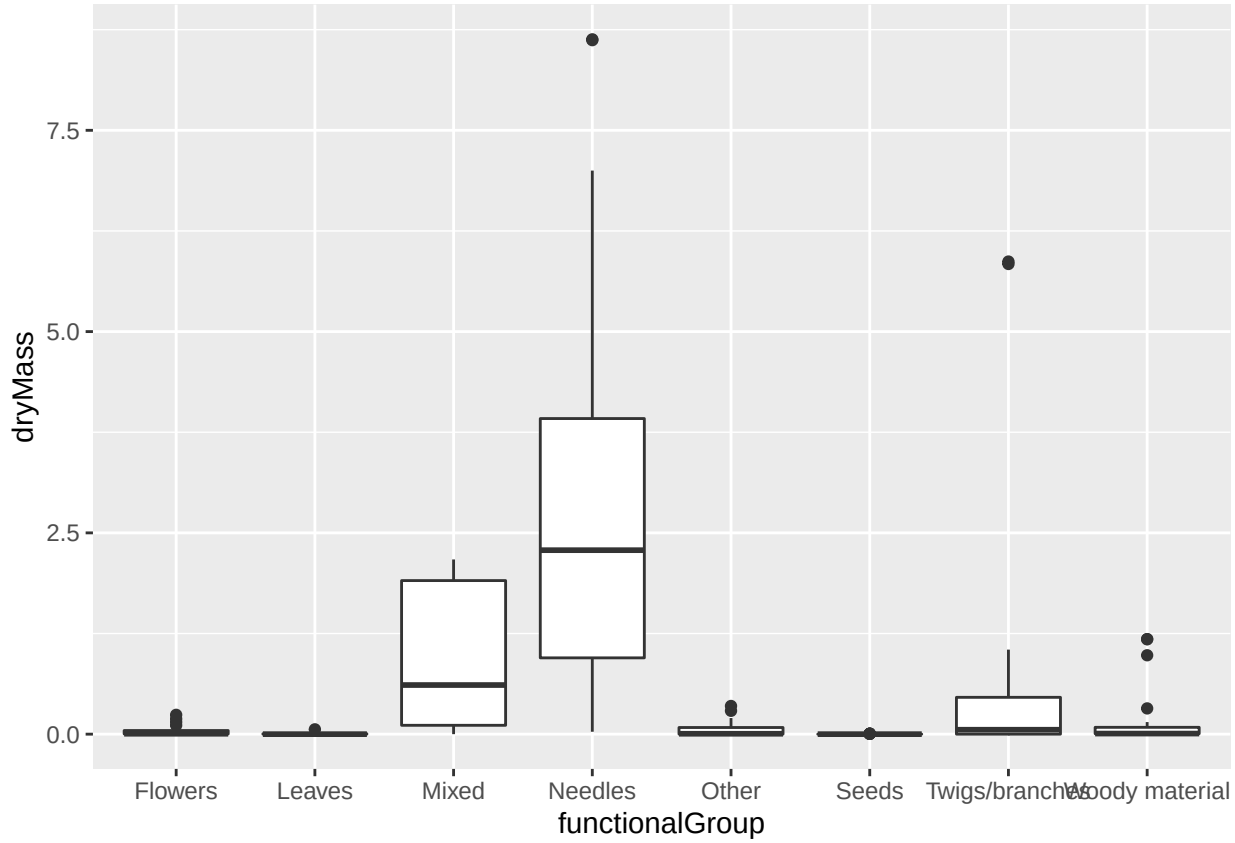
```r
ggplot(Litter, aes(x=functionalGroup))+
  geom_bar()
```




```
#creates barplot as done previously
```

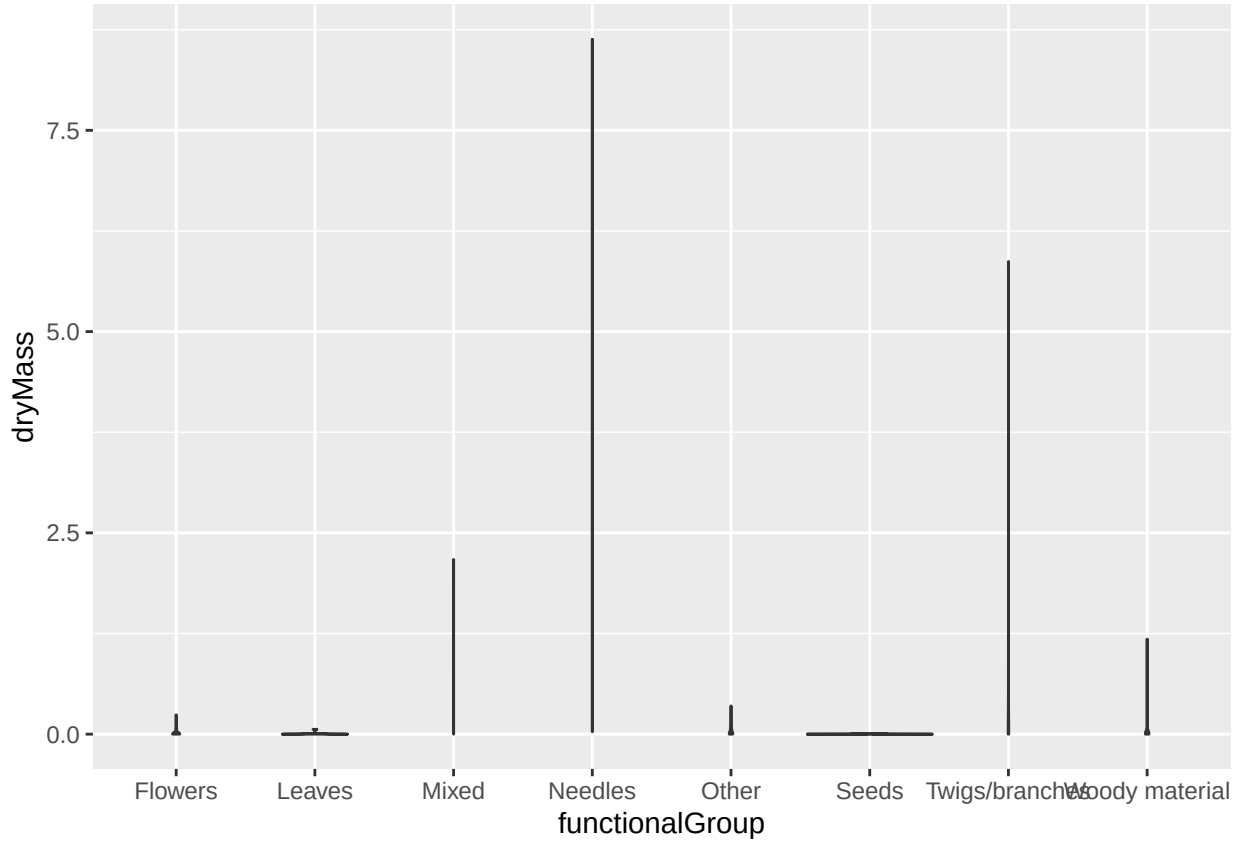
15. Using `geom_boxplot` and `geom_violin`, create a boxplot and a violin plot of dryMass by functionalGroup.

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




```
#geom_boxplot creates a boxplot with the associated x and y variables as listed in the aes argument
```

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




```
#geom_violin creates a violin plot with the associated x and y variables as listed in the aes argument,
```

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

> Answer: In this case, boxplot is better because the scale is skewed due to the outliers of needles and twigs, so it is difficult to observe the distribution of values recorded in the other functional groups. Boxplots showing the IQR and min and max are less affected, but there are still issues with that plot.

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

> Answer: Needles, mixed, and twigs/branches have the highest observed biomass at these sites.