Assignment 3: Data Exploration

Griffin Bird

Spring 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.

```
setwd("C:/Users/17038/Documents/EDA Spring 2023/")
getwd()
```

```
## [1] "C:/Users/17038/Documents/EDA Spring 2023"
```

```
library(tidyverse)
library(lubridate)

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 ecotoxicology of neonicotinoids on insects? Feel free to do a brief internet search if you feel you need more background information.

   Answer: Decreasing populations of pollinators is a pressing problem, and it's been established that insecticeds contibute to excess pollinator death. Makes sense that we would want to look the one of the most widely used classes of insecticides and their toxological impact on insects.

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: Forest litter influences the temperature, duration, and veracity of wildfires, so studying forest litter would give us a better idea of how a wildfire will act in a given area. Seeing how wildfires have plagued western states recently, I can imagine folks in CO are interested in what will fuel the next fire.

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. The locations of tower plots, where woody debris sampling takes place, is randomly selected. 2. Trap placement within plots is either targeted or randomized, depending on vegetation. In sites with more than 50% aerial cover of woody vegetation greater than 2m in height, trap placement is randomized. In sites with less than 50% cover of woody vegetation or heterogenously distributed, patchy vegetation, trap placement is targeted.
   3. Ground traps are sampled once per year, and elevated traps are sampled at different frequencies depending on what vegetation is present.

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: The two most commonly studied effects are Population and Mortality. These effects would be of particular interest because pesticide is known to harm insect populations, particularly pollinator populations. I imagine scientists are trying to figure out exactly how pollinators are affected by this class of pesticides.

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) %>%
  sort(decreasing = TRUE)
```

```
##                            (Other)                          Honey Bee
##                                670                                667
##                     Parasitic Wasp              Buff Tailed Bumblebee
##                                285                                183
##               Carniolan Honey Bee                          Bumble Bee
##                                152                                140
##                   Italian Honeybee                    Japanese Beetle
##                                113                                 94
##                  Asian Lady Beetle                     Euonymus Scale
##                                 76                                 75
##                           Wireworm                 European Dark Bee
##                                 69                                 66
##                  Minute Pirate Bug                Asian Citrus Psyllid
##                                 62                                 60
##                      Parastic Wasp             Colorado Potato Beetle
##                                 58                                 57
##                    Parasitoid Wasp                  Erythrina Gall Wasp
##                                 51                                 49
##                       Beetle Order        Snout Beetle Family, Weevil
##                                 47                                 47
##           Sevenspotted Lady Beetle                     True Bug Order
##                                 46                                 45
##               Buff-tailed Bumblebee                      Aphid Family
##                                 39                                 38
##                     Cabbage Looper                Sweetpotato Whitefly
##                                 38                                 37
##                      Braconid Wasp                        Cotton Aphid
##                                 33                                 33
##                     Predatory Mite              Ladybird Beetle Family
```

```
##                            33                             30
##                    Parasitoid                   Scarab Beetle
##                            30                             29
##                 Spring Tiphia                    Thrip Order
##                            29                             29
##          Ground Beetle Family             Rove Beetle Family
##                            27                             27
##                 Tobacco Aphid                   Chalcid Wasp
##                            27                             25
##        Convergent Lady Beetle                  Stingless Bee
##                            25                             25
##             Spider/Mite Class            Tobacco Flea Beetle
##                            24                             24
##              Citrus Leafminer                Ladybird Beetle
##                            23                             23
##                     Mason Bee                       Mosquito
##                            22                             22
##                 Argentine Ant                         Beetle
##                            21                             21
##    Flatheaded Appletree Borer           Horned Oak Gall Wasp
##                            20                             20
##            Leaf Beetle Family              Potato Leafhopper
##                            20                             20
##    Tooth-necked Fungus Beetle                   Codling Moth
##                            20                             19
##     Black-spotted Lady Beetle                   Calico Scale
##                            18                             18
##           Fairyfly Parasitoid                    Lady Beetle
##                            18                             18
##        Minute Parasitic Wasps                      Mirid Bug
##                            18                             18
##              Mulberry Pyralid                       Silkworm
##                            18                             18
##                Vedalia Beetle           Araneoid Spider Order
##                            18                             17
##                     Bee Order                 Egg Parasitoid
##                            17                             17
##                  Insect Class      Moth And Butterfly Order
##                            17                             17
##  Oystershell Scale Parasitoid Hemlock Woolly Adelgid Lady Beetle
##                            17                             16
##          Hemlock Wooly Adelgid                           Mite
##                            16                             16
##                   Onion Thrip          Western Flower Thrips
##                            16                             15
##                   Corn Earworm              Green Peach Aphid
##                            14                             14
##                     House Fly                      Ox Beetle
##                            14                             14
##             Red Scale Parasite             Spined Soldier Bug
##                            14                             14
##          Armoured Scale Family                Diamondback Moth
##                            13                             13
##                 Eulophid Wasp               Monarch Butterfly
```

```
##                                 13                               13
##                      Predatory Bug           Yellow Fever Mosquito
##                                 13                               13
##                 Braconid Parasitoid                    Common Thrip
##                                 12                               12
##        Eastern Subterranean Termite                          Jassid
##                                 12                               12
##                          Mite Order                       Pea Aphid
##                                 12                               12
##                    Pond Wolf Spider        Spotless Ladybird Beetle
##                                 12                               11
##                 Glasshouse Potato Wasp                     Lacewing
##                                 10                               10
##               Southern House Mosquito        Two Spotted Lady Beetle
##                                 10                               10
##                          Ant Family                    Apple Maggot
##                                  9                                9
```

Answer: The six most commonly studied species, by common name, are: Honey Bee, Parasitic Wasp, Buff Tailed Bumblebe, Carniolan Honey Bee, Bumble Bee, Italian Honeybee. As it would happen, all of these species are pollinators! These are probably the most frequently studied for that exact reason, their populations are declining and we want to know what role insecticides play in that.

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.)
```

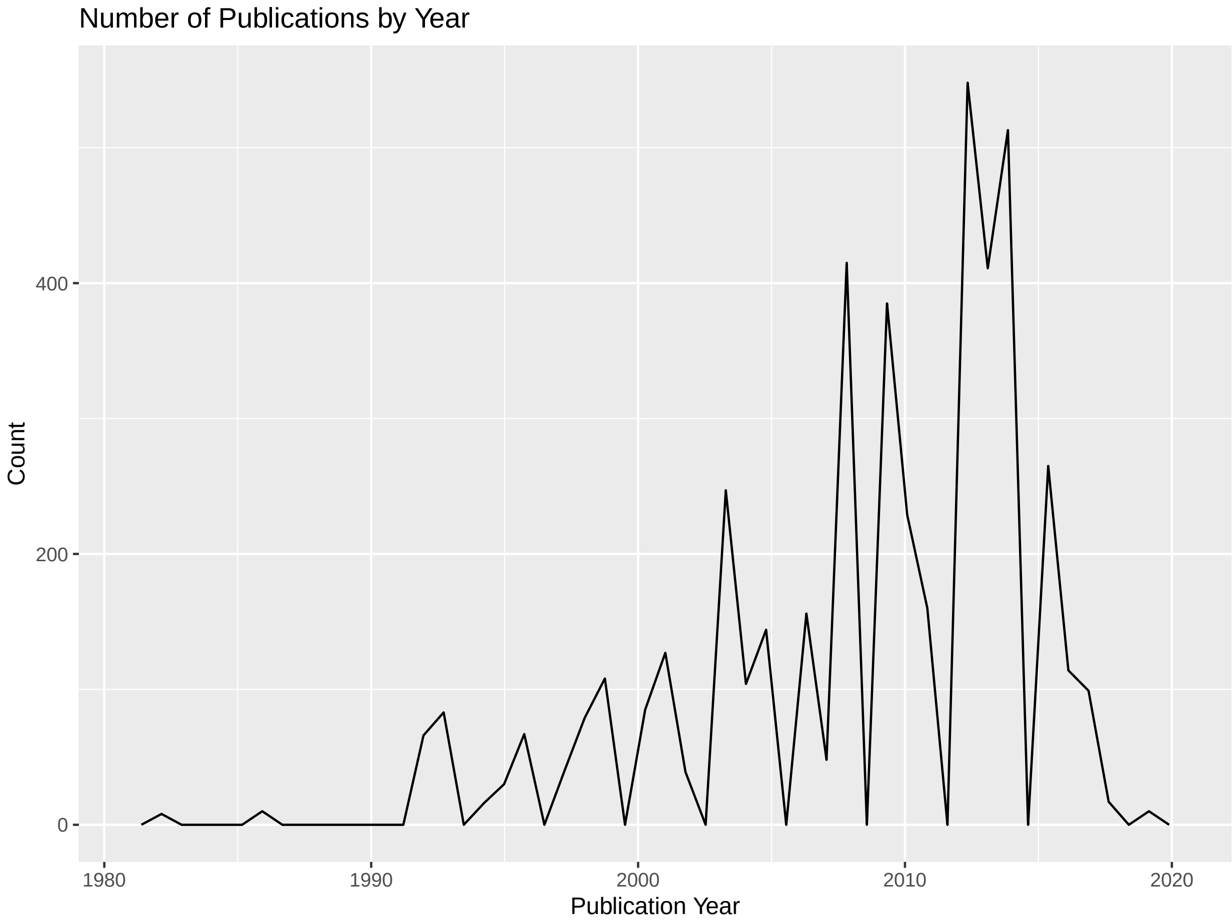
```
## [1] "factor"
```

Answer: `Conc.1..Author.` is factor data. R imported this column as factors because the column contains some non-numeric values, a lot of "/"s. R is kicking it all into factors because of those non-numeric characters.

Explore your data graphically (Neonics)

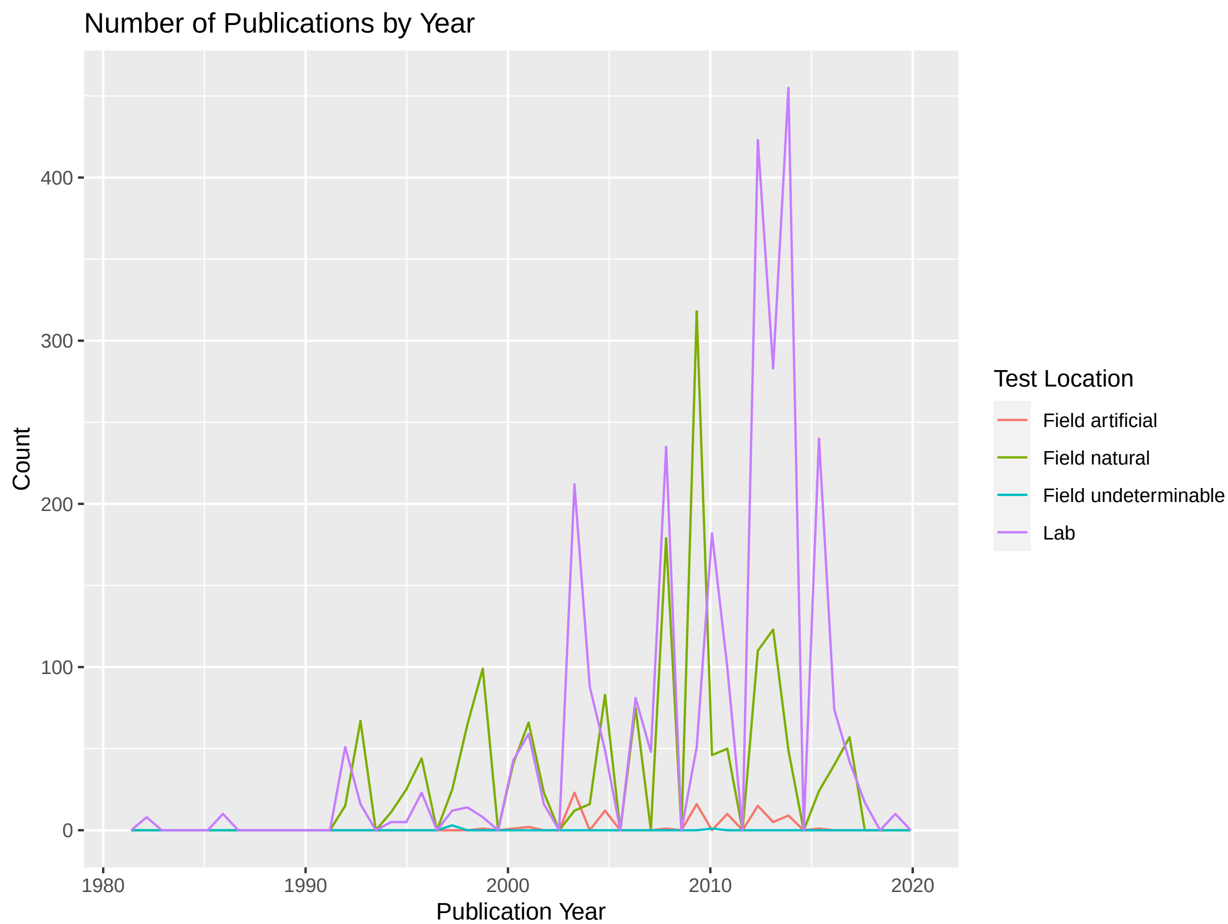
9. Using `geom_freqpoly`, generate a plot of the number of studies conducted by publication year.

```
ggplot(Neonics) +
 geom_freqpoly(aes(x = Publication.Year), bins = 50) + ggtitle("Number of Publications by Year") + xlab
```



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 = 50) + ggtitle("Number of Publica
```



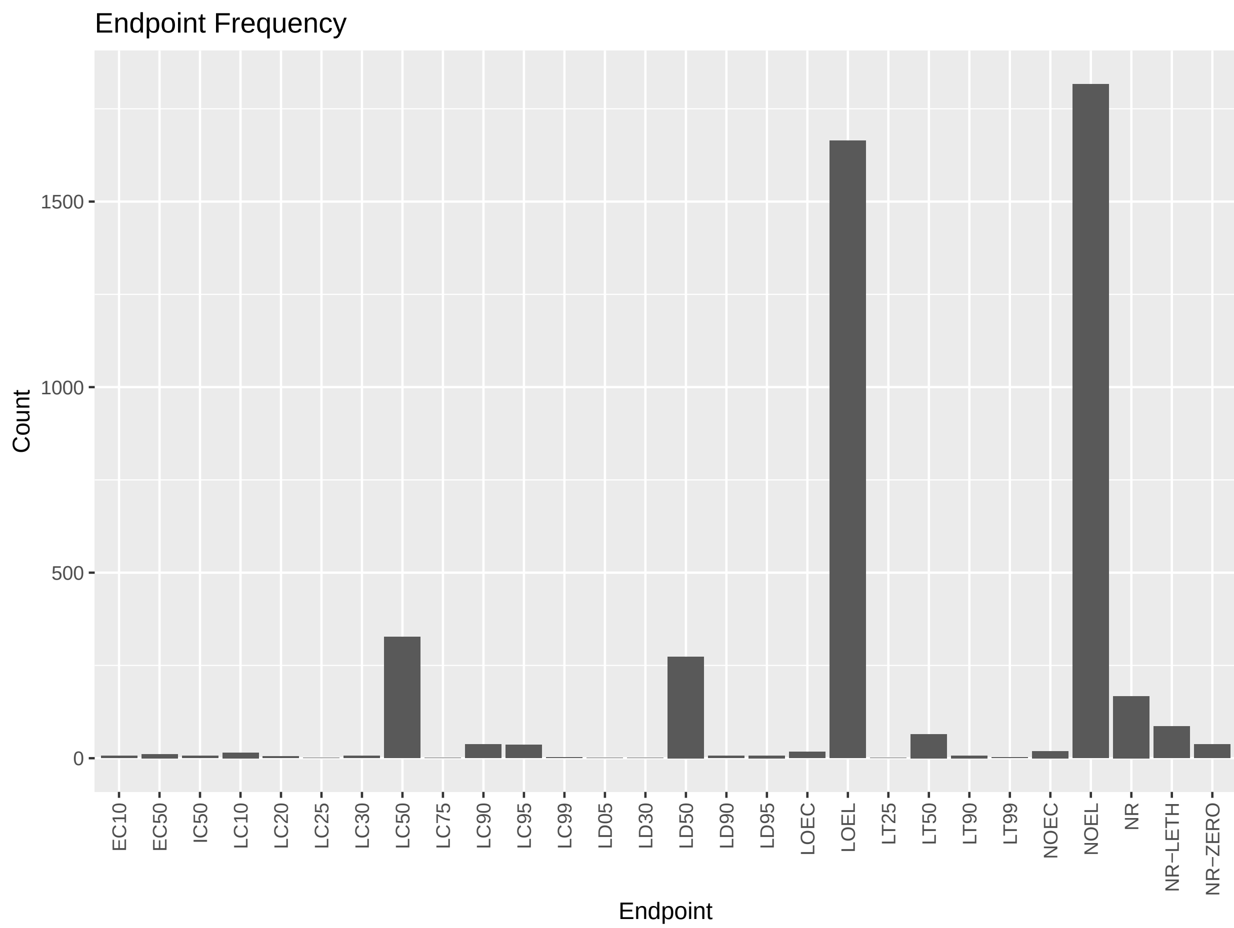
Interpret this graph. What are the most common test locations, and do they differ over time?

> Answer: The most common test locations are the lab and field natural. Field natural and lab experiments frequency are close through 2008 or so, but come 2015 lab is much more frequent than field natural.

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



Answer: The two most common endpoints are LOEL and NOEL. Loel is the lowest observable effect level at which a pesticide produced effects significantly different than the response to controls. Noel is the no-observable-effect-level which is the highest dose producing effects not 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$collectDate <- ymd(Litter$collectDate)
class(Litter$collectDate)
```

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

```r
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`?

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

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

```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
```

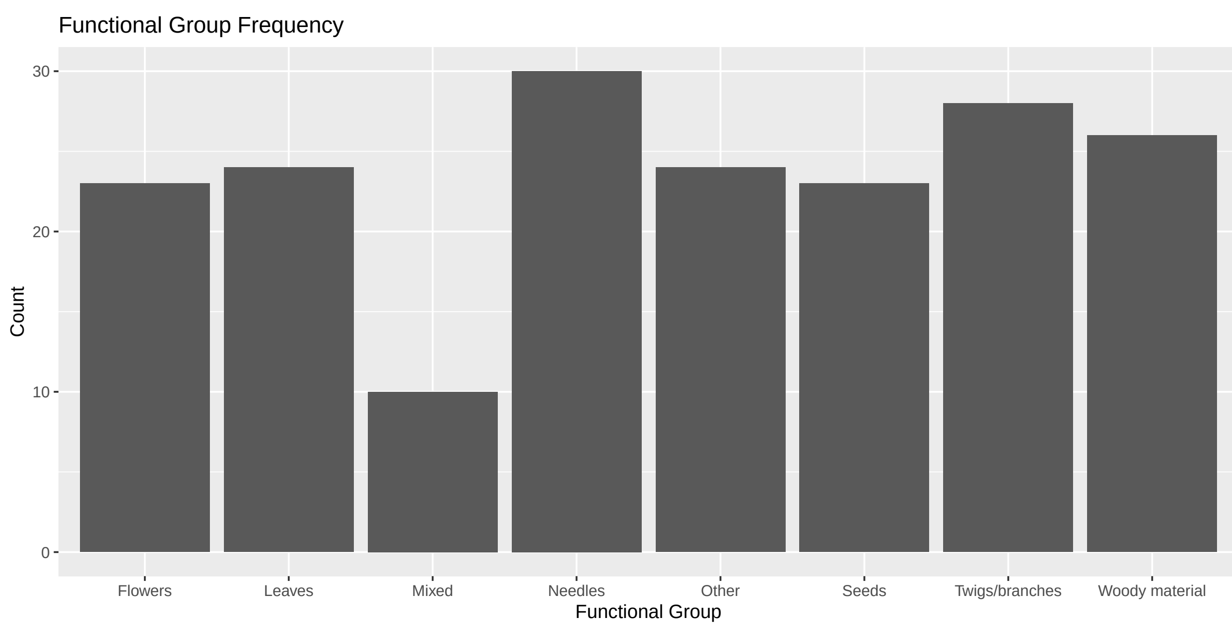
```r
summary(Litter$plotID)
```

```
## 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. "Unique" generates a string of all unique values in a vector, "Summary" will do that and show a frequency count for each unique value.

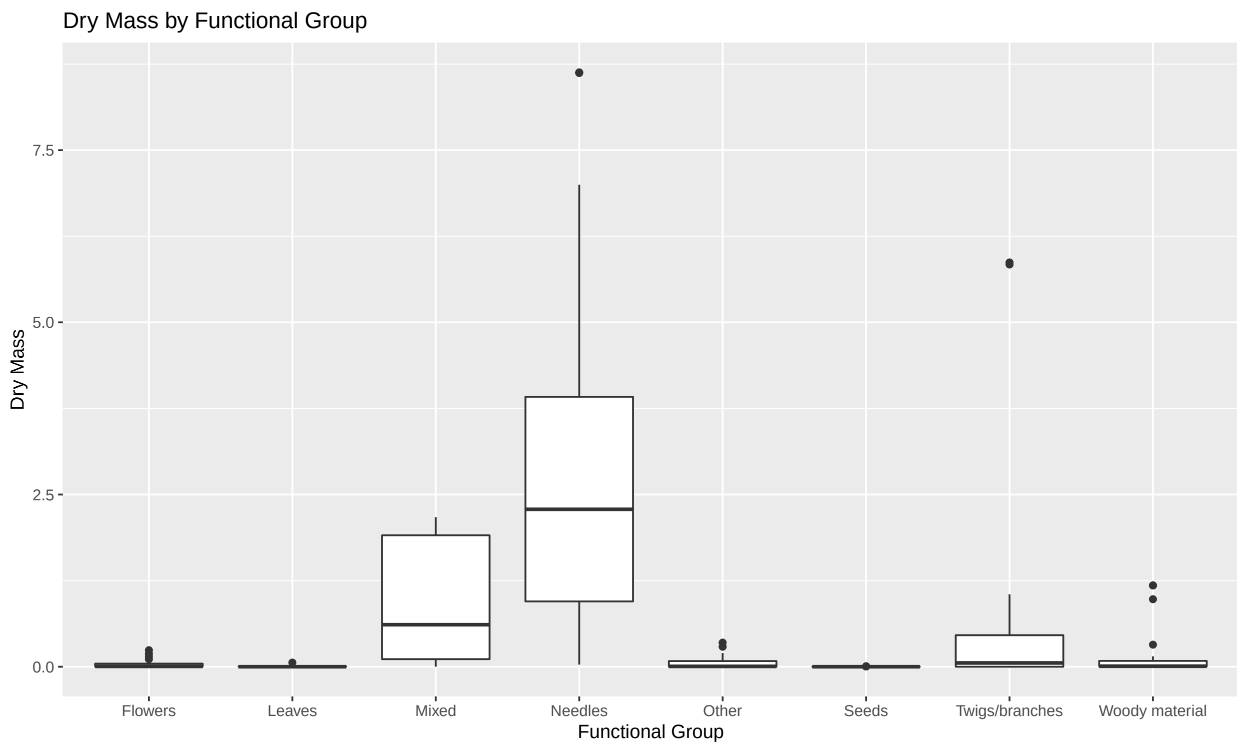
14. Create a bar graph of functional Group counts. This shows you what type of litter is collected at the Niwot Ridge sites. Notice that litter types are fairly equally distributed across the Niwot Ridge sites.

```r
ggplot(Litter) +
 geom_bar(aes(functionalGroup)) + xlab("Functional Group") + ylab("Count") + ggtitle("Functional Group F
```

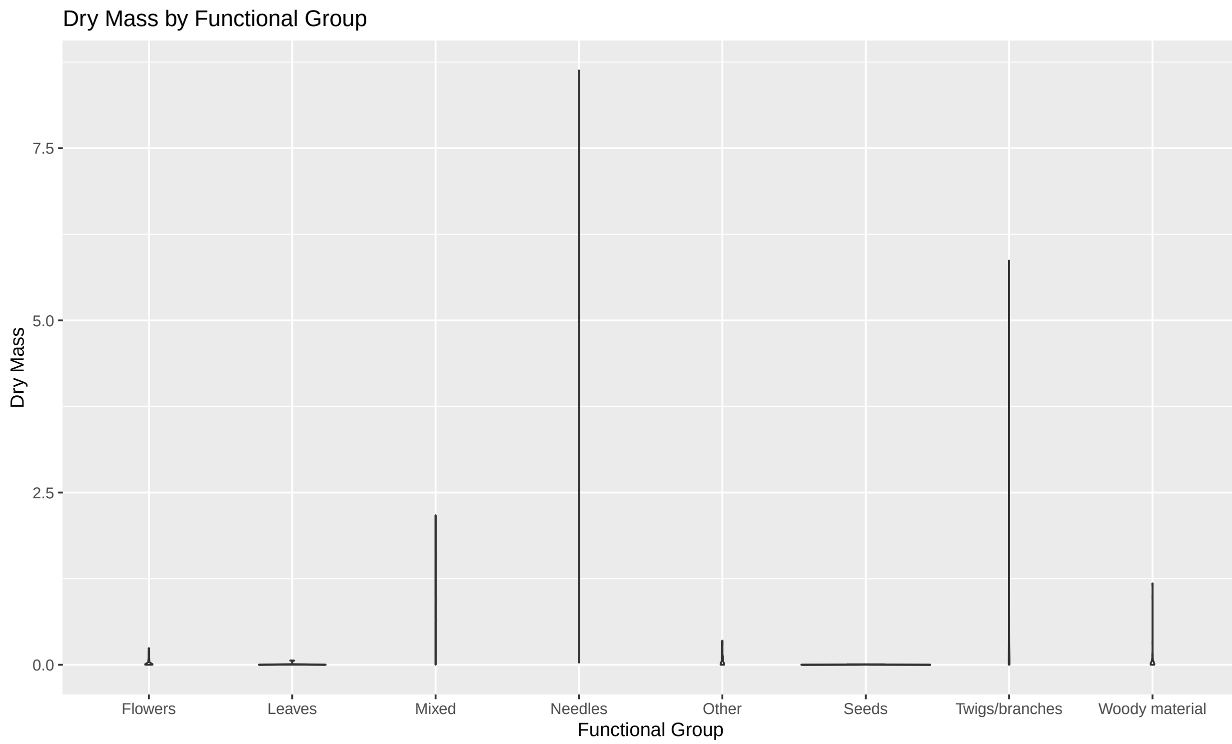


15. Using `geom_boxplot` and `geom_violin`, create a boxplot and a violin plot of dryMass by functionalGroup.

```
ggplot(Litter)+
  geom_boxplot(aes(functionalGroup, dryMass)) + xlab("Functional Group") + ylab("Dry Mass") + ggtitle("I
```




```
ggplot(Litter)+
  geom_violin(aes(functionalGroup, dryMass)) + xlab("Functional Group") + ylab("Dry Mass") + ggtitle("Dr
```



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

> Answer: So many funcitonal groups are being compared that it's difficult to see the definition of the "violin" for any one group, the box and whisker plot is much easier to read.

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

> Answer: Needles, twigs/branches, and mixed litter have the highest biomass.