Assignment 3: Data Exploration

Logan Dye

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. Be sure to **answer the questions** in this assignment document.
5. When you have completed the assignment, **Knit** the text and code into a single PDF file.
6. After Knitting, submit the completed exercise (PDF file) to the dropbox in Sakai.

The completed exercise is due on Sept 30th.

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 include the subcommand to read strings in as factors.

```
getwd()
```

```
## [1] "/home/guest/EDA-Fall2022/Assignments"
```

```
setwd("~/EDA-Fall2022")

#install.packages(tidyverse)
library(tidyverse)
#install.packages(lubridate)
library(lubridate)

Neonics <- read.csv("./Data/Raw/ECOTOX_Neonicotinoids_Insects_raw.csv", stringsAsFactors = TRUE)
Litter <- read.csv("./Data/Raw/NIWO_Litter/NEON_NIWO_Litter_massdata_2018-08_raw.csv", stringsAsFactors
```

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: We should be interested in the ecotoxigology of neonicitinoids because the neonics could have impacts on species outside of their intented use. For example, if a farmer wants to treat their crops for a specific pest and coats their field in neonics, they could unintentionally be killin other insect species that are important to the ecosystem. This is important knowledge to know for multiple reasons.

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: A lot if information can be gathered from looking at biomass overtime. From biomass you can infer potention fuel load on the forest floor that can increase instense wildfire risk. You can also tell if there is a shift in plant growth from year to year. If you collect biomass one year and it is very high, you can check what the weather conditions/distrubances were like a year ago that would increase the total biomass, or vica versa, you could see why there would be a decrease in biomass.

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. Ground traps are sampled once per year. Target sampling frequency for elevated traps varies by vegetaton present at the site, with frequent sampling (1x every 2weeks) in deciduous forest sites during senescence, and infrequent year-round sampling (1x every 1-2 months) at evergreen sites. Litter sampling of elevated traps may be discontinued for up to 6 months during the dormant season. 2. One litter trap pair (one elevated trap and one ground trap) is deployed for every 400 m2 plot area, resulting in 1-4 trap pairs per plot. Elevated litter traps only are used to collect litter and are each 0.5 square meters. Ground traps are each 3m by 0.5 m and collect both litter and find woody debree. 3. Litter is discribed as having a butt end <2cm and a total lenge <50 cm. Fine woody debree is discribed as having a butt end <2cm and a length >50cm.

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 most common effects studied are mortality and population. Since neonicitinoids are pesticides used to kill instects and impact their population numbers, it makes sense that these two effects would be the most commonly studied. Knowing what species will die and how poplations will responed is key information.

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.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 six most common are Honey Bee, Parasitic Wasp, Buff Tailed Bumblebee, Carniolan Honey Bee, Bumble Bee, and Italian Honeybee. All six of these species are insects that you would love to have in your farm/garden. All six species are terrific pollinators and the parasitic wasp is a natural insecticide that kills unwanted pests and keeps your plants healthy. These would be of upmost importance to study becasue, if the neonicitinoids you are putting out on your land kill all of your pollinators and natural insecticides then the net benefits of killing the upwanted pests will be negated by the net nagatives of killing off your pollinator species.

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.)
```

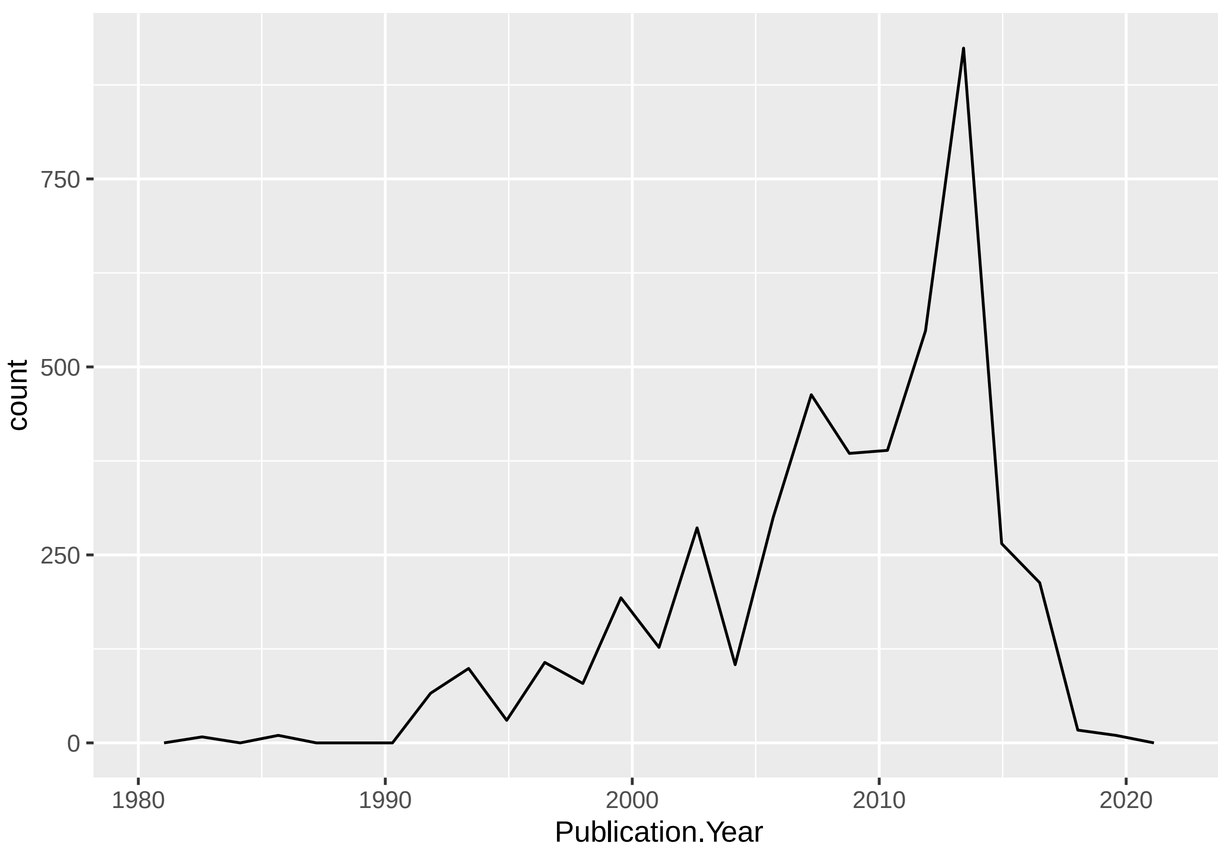
```
## [1] "factor"
```

Answer: The class is "factor." I belive that it is not listed as numeric because there are numbers data entries under this column that are not numberic. For example, many entries are "NR/" or are just "NR."

Explore your data graphically (Neonics)

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

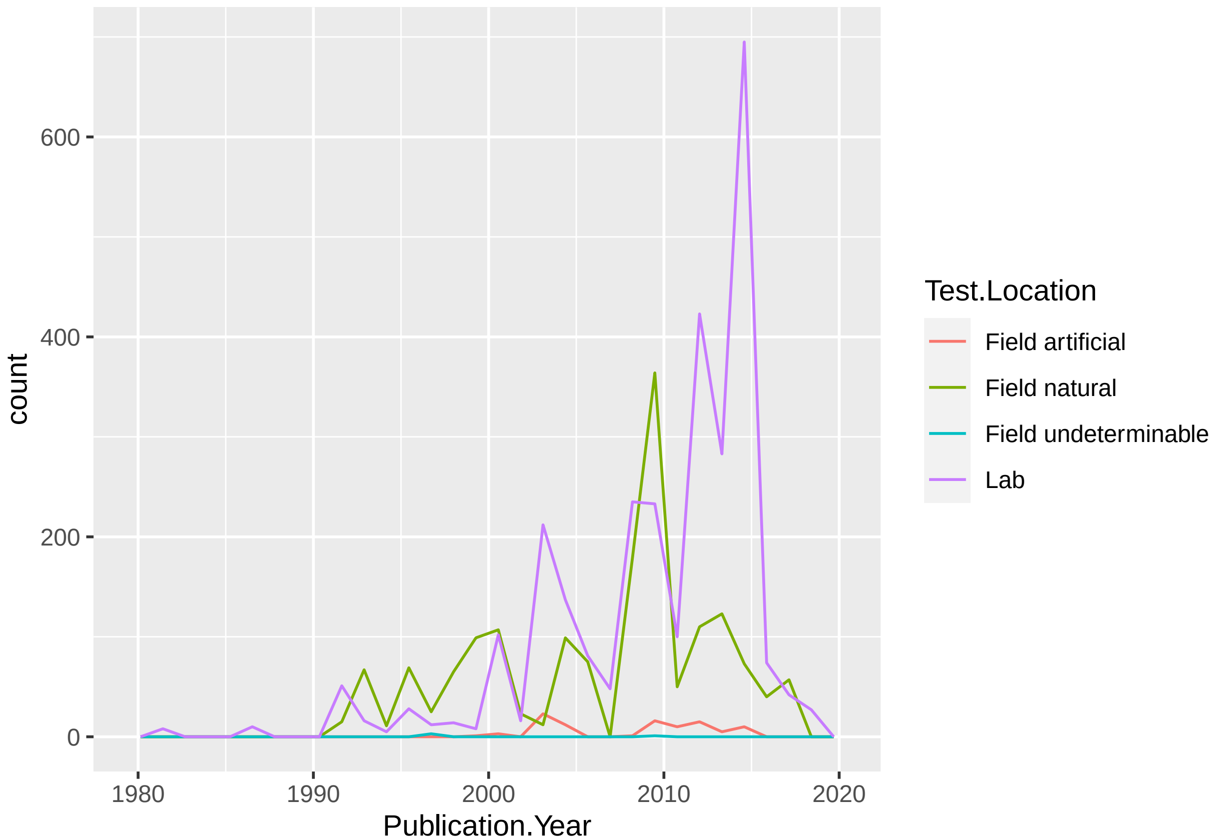
```
ggplot(Neonics) + geom_freqpoly(aes(x = Publication.Year), bins = 25)
```



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

```r
ggplot(Neonics, aes(Publication.Year, after_stat(count), colour = Test.Location)) +
    geom_freqpoly()
```

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

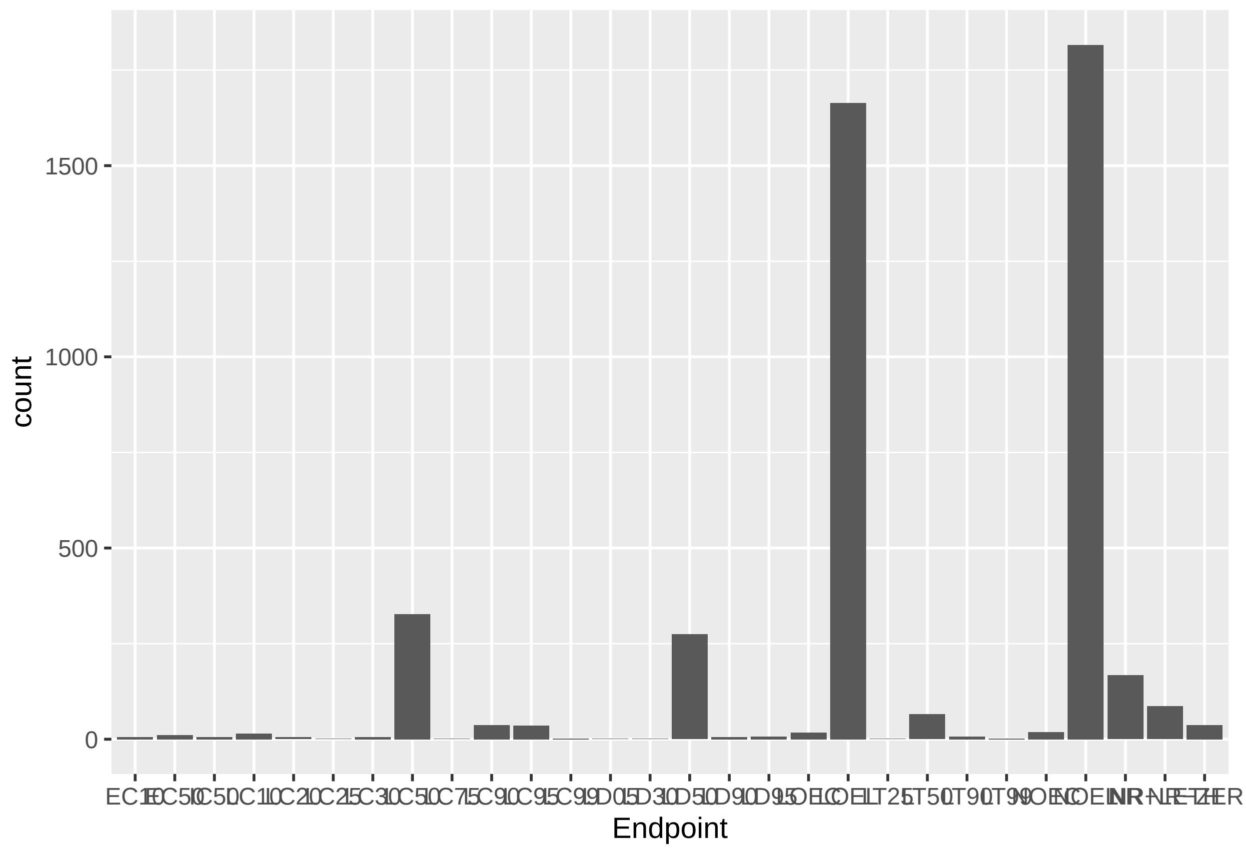


Interpret this graph. What are the most common test locations, and do they differ over time?

> Answer: Lab and Field Natural are for and away the most common test locations. Over time, it appears as though testing in the lab has overtaken testing in field natural. Particularly after the year 2000. I am assuming that this is becasue better lab methods are being developed that can replicate the natural field setting accuratly while also controlling for extraneous variables.

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.

```r
ggplot(Neonics, aes(x = Endpoint)) + geom_bar()
```



Answer: The two most common endpints are LOEL and NOEL. LOEL is the "Lowest-observable-effect-level: lowest dose (concentration) producing effects that were significantly different (as reported by authors) from responses of controls" and NOEL is "No-observable-effect-level: highest dose (concentration) producing effects not significantly different from responses of controls according to author's reported statistical test." This is important information to have when you are attemping to set dosing standards.

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")

unique(Litter$collectDate, 2018 - 8)
```

```
## [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
```

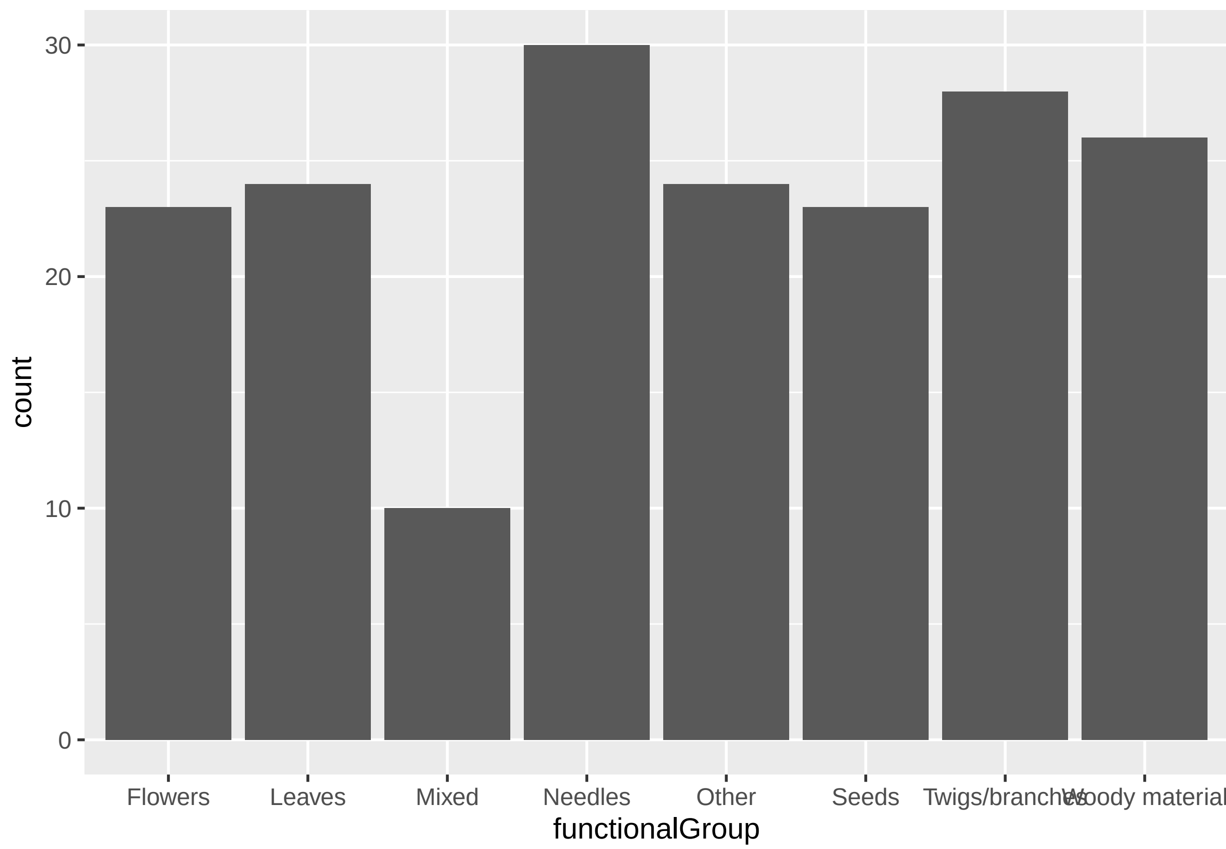
```
summary(Litter$plotID)
```

```
## 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:There are 12 different plots that were sampled at Niwot Ridge. The information you get from "unique" is different from "summary," because all it gives is the individual unique values and nothing else. For this particular question, all it lists are the names of each unique plot. When I run "summary" on the same information, the output gives the unique plots, but also the number of times each plot was sampled.

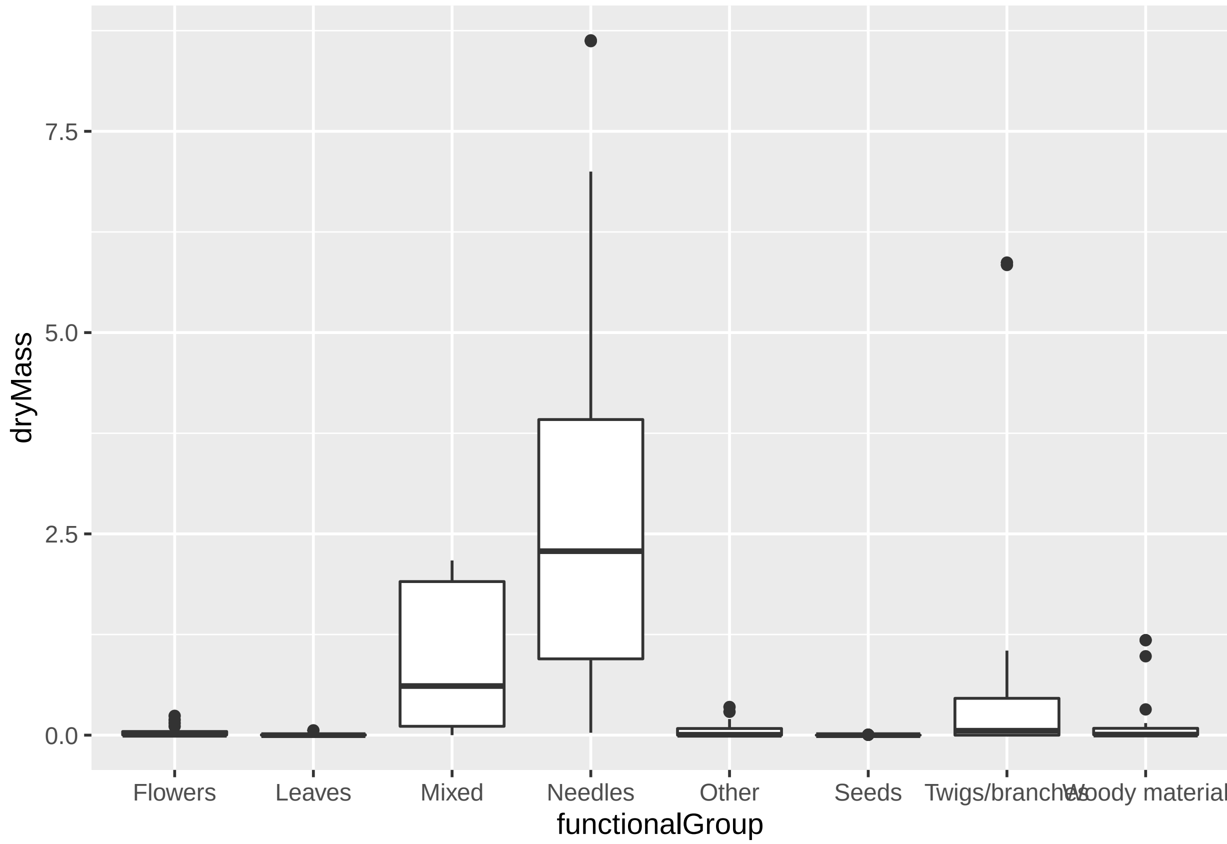
14. Create a bar graph of functionalGroup counts. This shows you what type of litter is collected at the Niwot Ridge sites. Notice that litter types are fairly equally distributed across the Niwot Ridge sites.

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

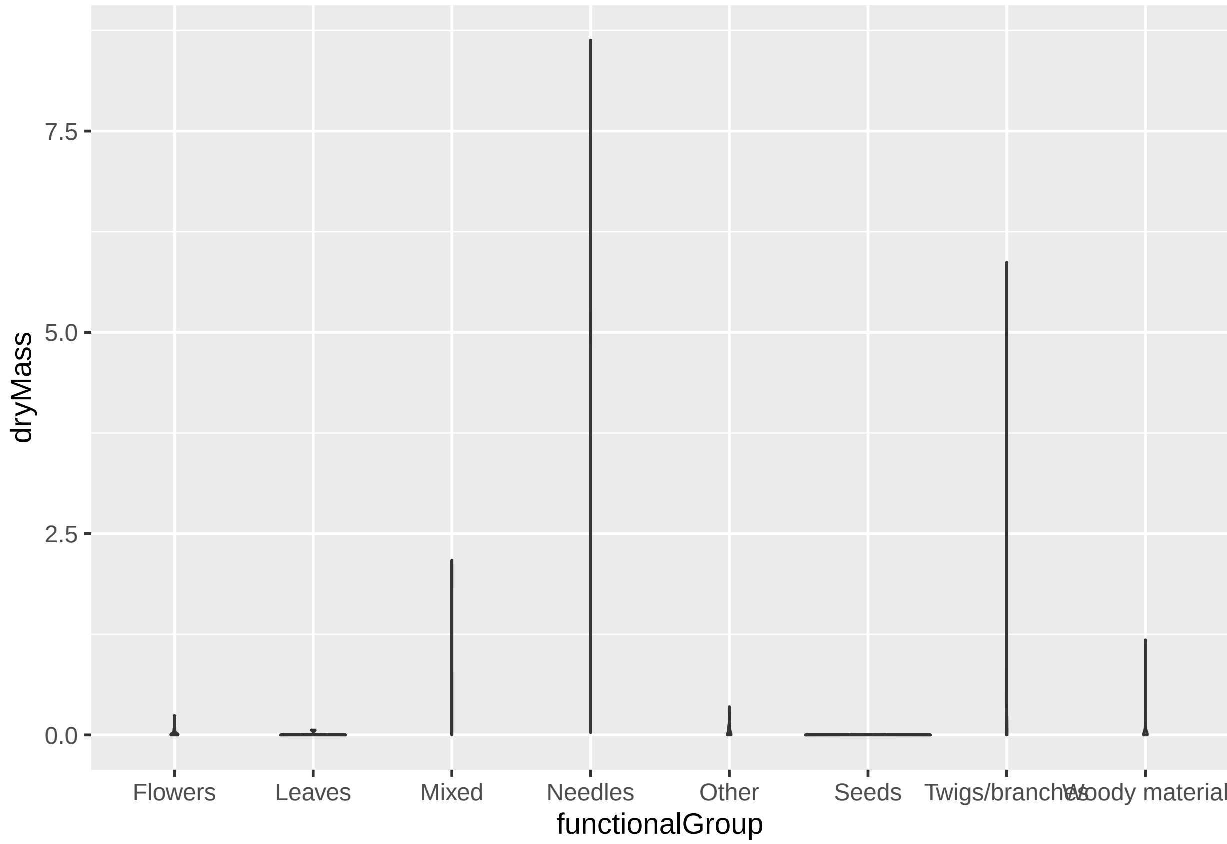


15. Using `geom_boxplot` and `geom_violin`, create a boxplot and a violin plot of dryMass by functionalGroup.

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




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



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

> Answer:There are a lot of "0" value points in this data set. For a boxplot, the 0's do not impact what the median and IQR's are so the boxplots are less influanced. However, when a violin plot is caclulating the densities, the high number of 0's greatly skew the data becasue the denses part of the dats is goig to be around "0."

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

> Answer: "Needles" has the highest biomass by far. Behind "needles" is "mixed" and "twigs/branches."