Assignment 3: Data Exploration

Sokna Kry

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.

```
getwd()
```

```
## [1] "/Users/sokna/Documents/EDA-Spring2023"
```

```
library(tidyverse)
```

```
## -- Attaching packages --------------------------------------- tidyverse 1.3.2 --
## v ggplot2 3.4.0     v purrr   1.0.1
## v tibble  3.1.8     v dplyr   1.1.0
```

```
## v tidyr   1.3.0     v stringr 1.5.0
## v readr   2.1.3     v forcats 1.0.0
## -- Conflicts ------------------------------------------ tidyverse_conflicts() --
## x dplyr::filter() masks stats::filter()
## x dplyr::lag()    masks stats::lag()
```

```
library(lubridate)
```

```
##
## Attaching package: 'lubridate'
##
## The following objects are masked from 'package:base':
##
##     date, intersect, setdiff, union
```

```
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:Neonicotinoids are used in agriculture to protect crops from various insects. However, they might affect the evironment and non target insects such as bees, worms and birds. Insects (for example bees or worms) play an important role in maintain the ballance of ecosystem (for example in pollinating or decomposing). The ecotoxicology of neonicotinoids can help us to understand the impacts of neonicotinoids on the insects, risk on human health and ecological, and provide useful information for decision making on using neonicotinoids in agriculture sector.

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 play an important role in carbon budgets and nutrient cycling, and provide habitat and food for a variety of organisms, including insects, fungi, and mammals. The study of litter and woody debris that falls to the ground in forests is important as it can provide the information for land and resource managment. The information from the study will help in decision making on sustainable managment of forest and ecosystem to balance the ecological function.

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.**litter** is defined as material that is dropped from the forest canopy and has a but end diameter <2cm and a length <50 cm; this material is collected in elevated 0.5m2 PVC traps. **Fine wood debris** is defined as material that is dropped from the forest canopy and has a but end diameter <2cm and a length >50 cm; this material is collected in ground traps as longer material is not reliably collected by the elevated traps. The ground traps are 3m X 0.5 m rectangular areas. 2.**Spatial Sampling** in sites with forested tower airsheds, the litter sampling is targeted to tak place in 20 40mx40m plots. In sites with low-statured vegetation over the tower airsheds, litter sampling is targeted to take place in 4 40mx40m tower plots plus 26 20mx20m plots. 3. **Temporal Sampling** Ground traps are sampled once per year. Target sampling frequency for elevated traps varies by vegetarion present at the site, with frequent sampling (1x every 2 weeks) in deciduos forest sites during senescence, and in frequent year-round sampling (1 x every 1-2 months) at evergreen sites.

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 are Population (1803) and Mortality (1493). Thes effects specifically are of interest because it provide insight into potential harm on insects from using neonicotinoids. For example, neonicotinoids cost harm on population and mortality of insects (bees), resulting in less bees for pollinating , causing inbalance in ecosystem. The results from the effects study can be used as a back up for decision making in establishing policy for pest control or how to use neonicotinoids in more sustainable way.

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…]

```r
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 6 most commly studied species in the datset are Honey Bees (667), Parasitc Wasp (285), Buff Tailed Bumblebee (183), Carnilan Honey Bee (152), Bumble Bee (140), and Italian Honeybee (113). One important thing that these species have in common is a role in pollination. They are the pollinators in ecosystem. They are in our interest to study over other insects because they play an important roles in our food chain and maintain the ballance of ecosystem. With the study, we can understand the effects of their decline in population and address that problem.

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.)
```

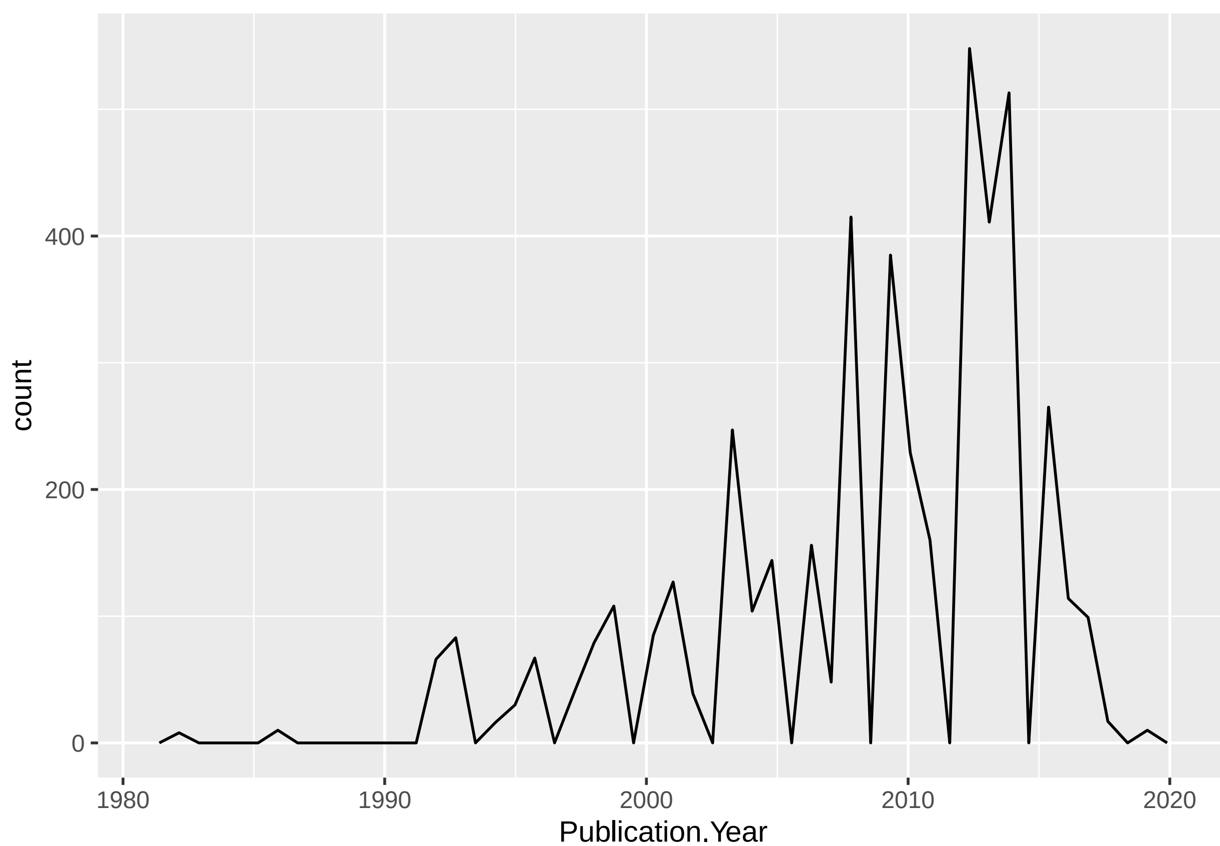
```
## [1] "factor"
```

Answer:The class of Conc.1..Author. is factor. IT is not numberic because some rows were recorded with NR (charactors). Therefore, the combination of charactors and numberic are considered as categorical data instead of numerical data.

Explore your data graphically (Neonics)

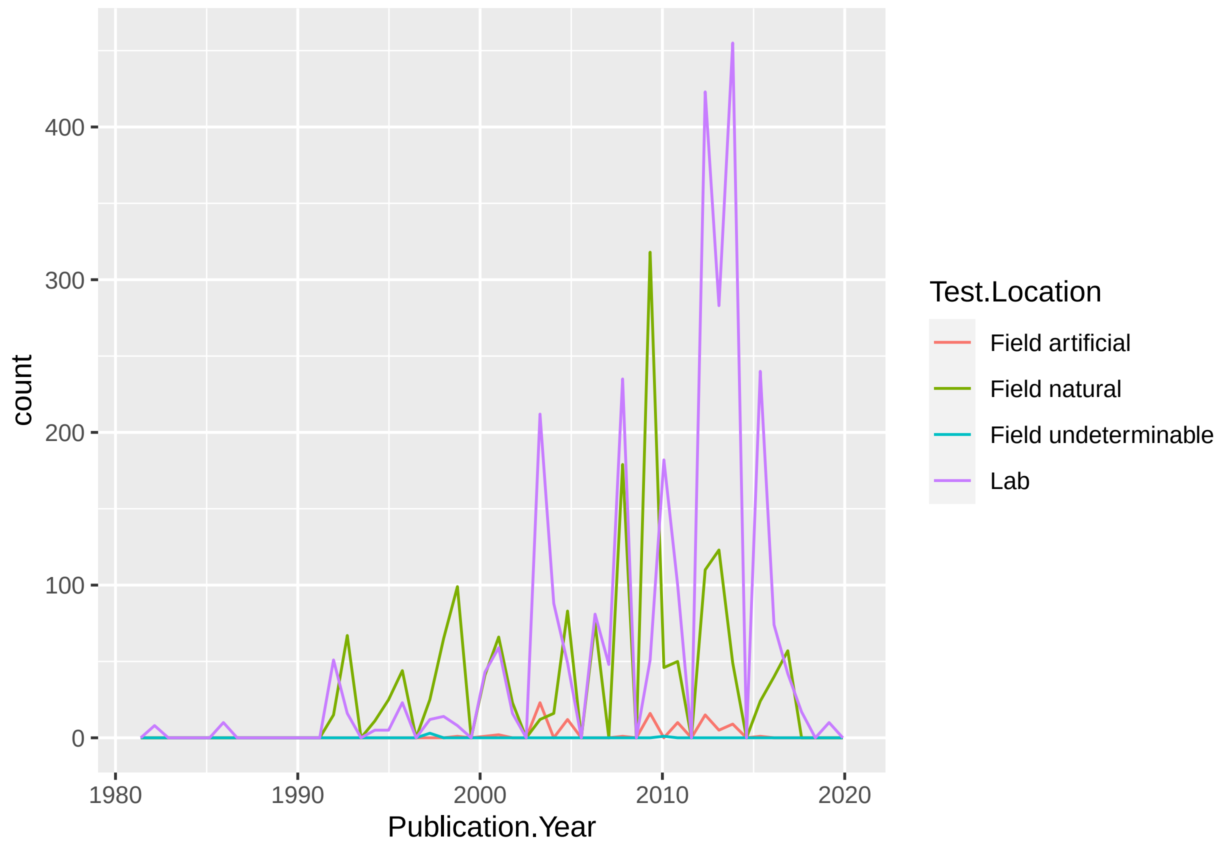
9. Using `geom_freqpoly`, generate a plot of the number of studies conducted by publication year.

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



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)
```



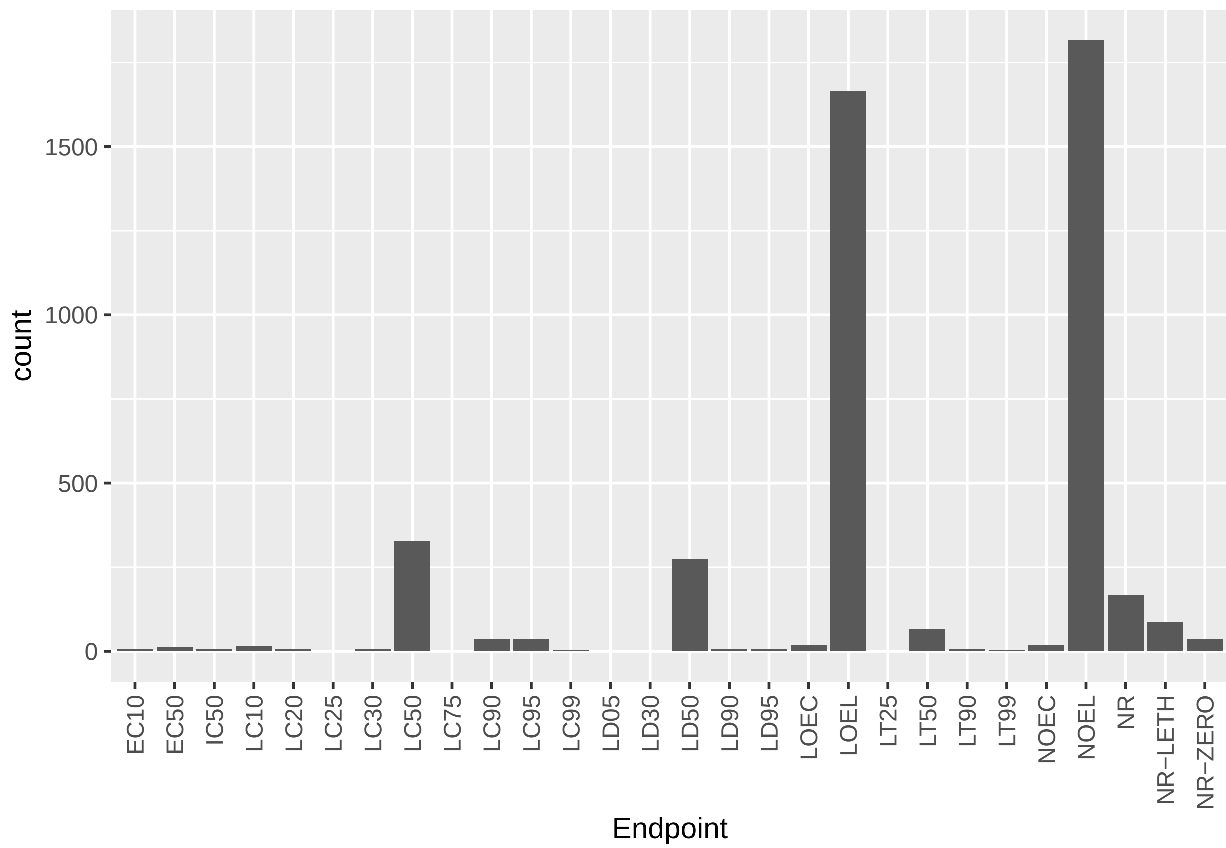
Interpret this graph. What are the most common test locations, and do they differ over time?

> Answer: The most common test locations are Lab and Field natural. They differ over time. In 2009, test at field natural was at the maximum of over 300. In 2014, test in the Lab was at the it highest points of around 450.

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




```
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 Endponts they defined are NOEL (1816) and LOEL (1664). LOEL (Lowest-observable-effect-leve) is defined as lowest dose (concentration) producing effects that were signigicantly different from responses of controls. NOEL (No-observable-effect-level) is defined as the highest dose (concentration) producting 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$collectDate <- as.Date(Litter$collectDate, format = "%Y-%m-%d")

class(Litter$collectDate)
```

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

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_plots<- unique(Litter$siteID[Litter$siteID=="NIWO"])

summary(unique_plots)
```

```
## NIWO
##    1
```

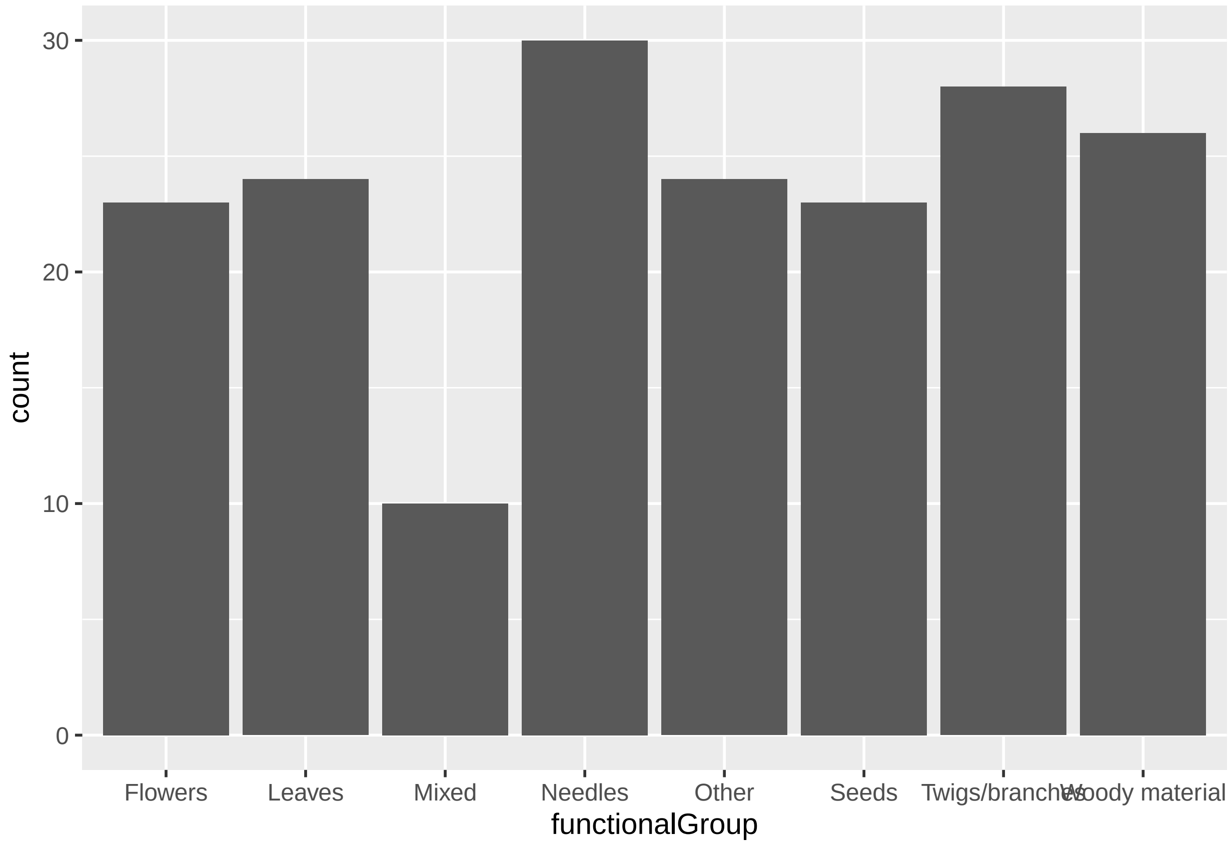
```
summary(Litter$siteID)
```

```
## NIWO
##  188
```

> Answer: The summary shows that the number of NIWO plots is 188. But the Unique plot is only 1. This is different because in summary, all unique values in a column are counted (counted all different values and the same values), while in unique, it counts only the different values.

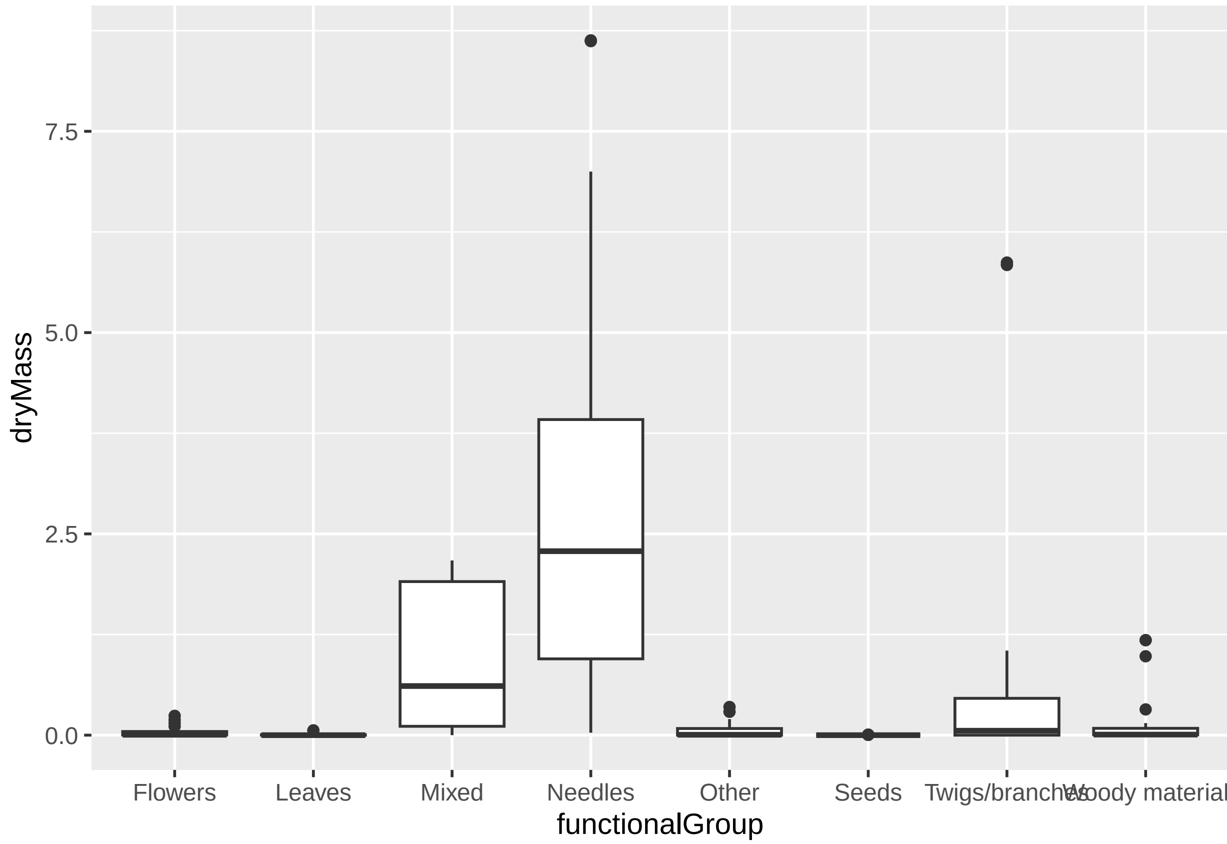
14. Create a bar graph of functionalGroup counts. This shows you what type of litter is collected at the Niwot Ridge sites. Notice that litter types are fairly equally distributed across the Niwot Ridge sites.

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

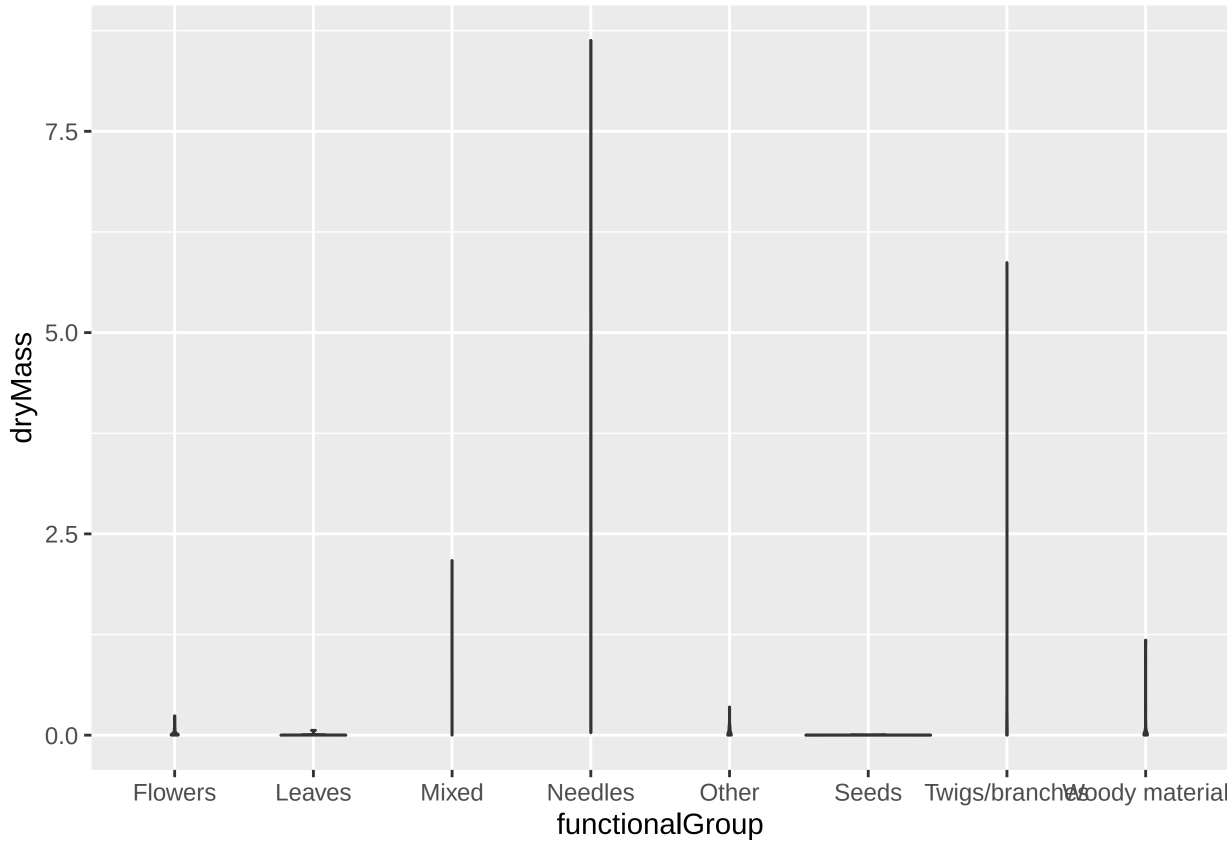


15. Using `geom_boxplot` and `geom_violin`, create a boxplot and a violin plot of dryMass by functionalGroup.

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




```
ggplot(Litter) +
  geom_violin(aes(x = functionalGroup, y = dryMass),
              draw_quantiles = c(0.25, 0.5, 0.75))
```



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

> Answer:Boxplot is more effective visualization option than the violin plot in this case because it is easier to compare the median and quartiles across the different categories of Functional-Group. The boxplot also provides a clear distribution of data, which is easier to spot outliers and skewness. While the violin plot shows the disbution of each category, but it is rather hard to interpret.

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

> Answer: By looking at the boxplox, Needles tend to have the highest biomass at these sites, then follow by Mixed.