Assignment 3: Data Exploration

Xia Meng Howey

Fall 2025

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 Canvas.

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. Load necessary packages (tidyverse, lubridate, here), check your current working directory 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.

```r
#load necessary packages
library(tidyverse)
library(lubridate)
library(here)
#check current working directory
getwd()
```

```
## [1] "/home/guest/R/EDE_Fall2025"
```

```r
#output is "/home/guest/R/EDE_Fall2025"
here()
```

```
## [1] "/home/guest/R/EDE_Fall2025"
```

```r
#import data
Neonics <- read.csv(
  file = here("./Data/Raw/ECOTOX_Neonicotinoids_Insects_raw.csv"),
  stringsAsFactors = TRUE
)
NeonicsRAW <- Neonics

Litter <- read.csv(
  file = here("./Data/Raw/NEON_NIWO_Litter_massdata_2018-08_raw.csv"),
  stringsAsFactors = TRUE
)
LitterRAW <- Litter
```

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: Since these neonicotinoids are insecticides that means that they are used to kill insects from damaging agricultural crops. To understand which neonicotinoids impact which insects and how, it would be useful to understand the ecotoxicology mechanisms involved.

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: Studying the litter and woody debris helps you understand the health and functioning of these ecosystems. They can give insights to processes like nutrient cycling, carbon storage, and habitat creation.

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. In spatial sampling design, sampling is executed at terrestrial sites that contain woody vegetation greater than 2m tall. 2. In temporal sampling design, ground traps are sampled once per year. 3. Material definitions: "litterfall" includes material dropped from the canopy, butt-end diameter < 2 cm and length < 50 cm; "fine woody debris" is similar diameter (< 2 cm) but length > 50 cm.

Obtain basic summaries of your data (Neonics)

5. What are the dimensions of the dataset?

```
dim(Neonics)
```

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

```
#dimensions are 4623 rows by 30 colummns
```

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? [Tip: The sort() command is useful for listing the values in order of magnitude...]

```
NEON_summary <- summary(as.factor(Neonics$Effect))
NEON_summary_sorted <- sort(NEON_summary)
NEON_summary_sorted
```

```
##       Hormone(s)        Histology       Physiology          Cell(s)
##                1                5                7                9
##     Biochemistry     Accumulation     Intoxication     Immunological
##               11               12               12               16
##       Morphology           Growth        Enzyme(s)         Genetics
##               22               38               62               82
##        Avoidance      Development     Reproduction Feeding behavior
##              102              136              197              255
##         Behavior        Mortality       Population
##              360             1493             1803
```

Answer:Most common effects are population, mortality, and behavior, population and mortality, the highest two, can show how effective the pesticides are at killing the pests, while looking at behavior can maybe lead to bugs being inactive to not damage crops or different behaviors like that

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: Explore the help on the summary() function, in particular the maxsum argument...]

```
NEON_commonname <- summary(as.factor(Neonics$Species.Common.Name))
NEON_commonname_sorted <- sort(NEON_commonname)
NEON_commonname_sorted
```

```
##                        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: From highest to lowest 1.Other 2.Honey Bee 3.Parasitic Wasp 4.Buff Tailed Bumblebee 5.Carniolan Honey Bee 6.Bumble Bee Most of these species are types of bees. Bees are pollinators, and thus an important part of the ecosystem.

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? [Tip: Viewing the dataframe may be helpful…]

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

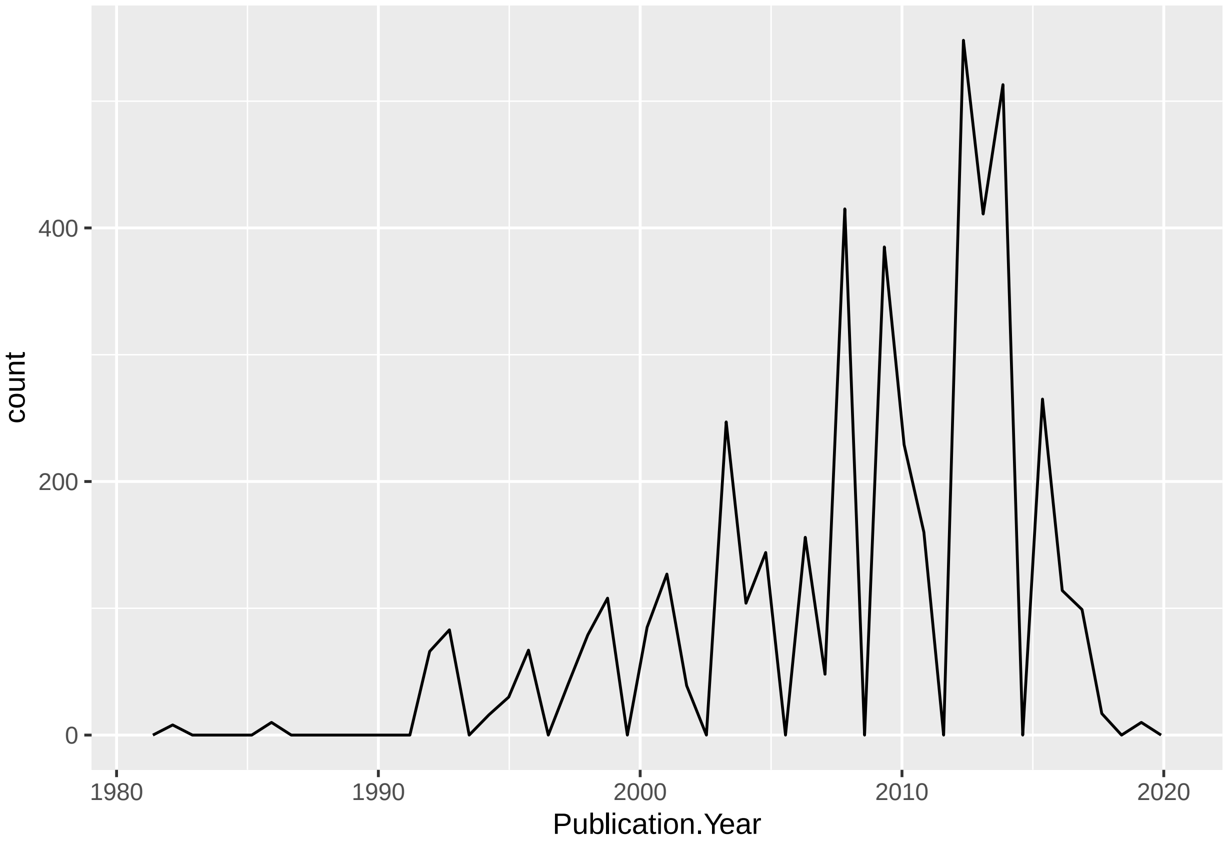
```
## [1] "factor"
```

Answer:This column is a factor, this is because while the column does include numbers, it also includes non number characters like letters, ~, /

Explore your data graphically (Neonics)

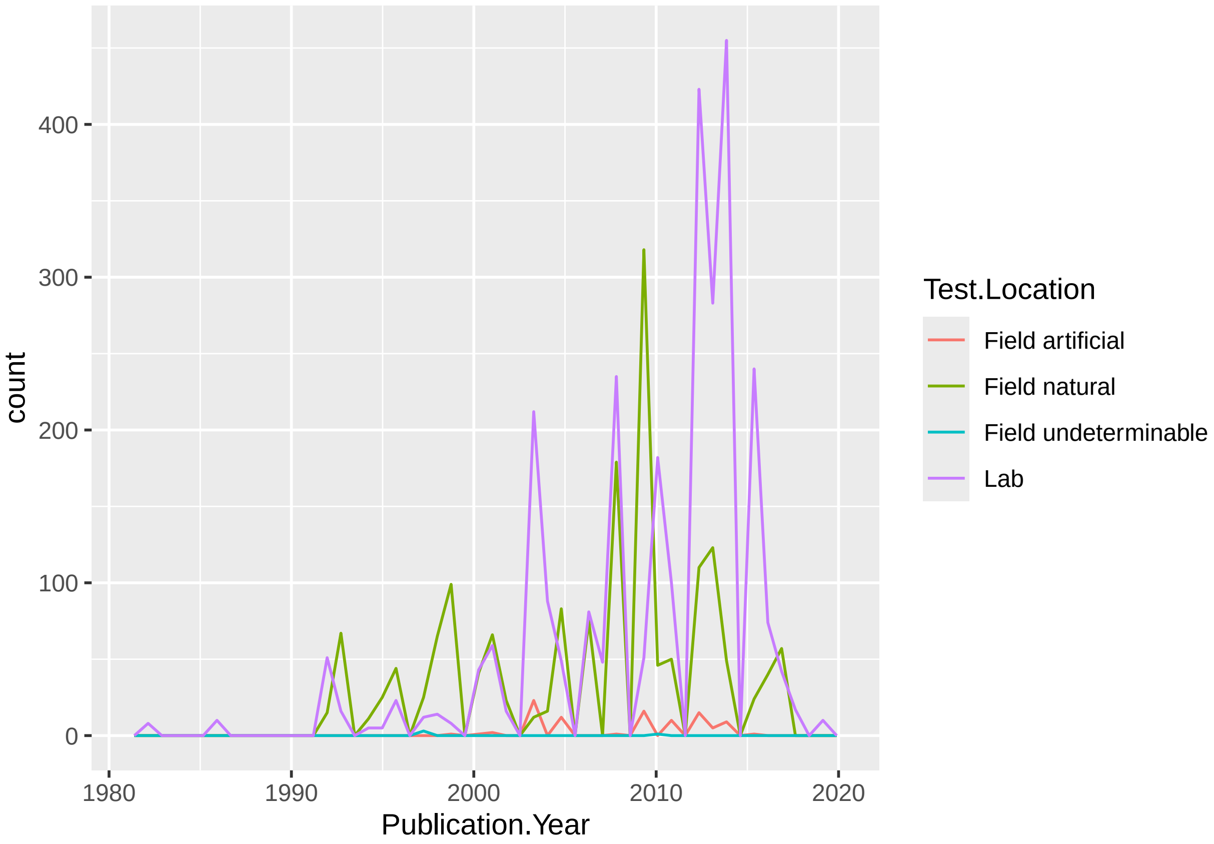
9. Using `geom_freqpoly`, generate a plot of the number of studies conducted by publication year.

```r
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.

```r
Pub_test_fig <- ggplot(Neonics) +
  geom_freqpoly(aes(x = Publication.Year, color = Test.Location), bins = 50 )
Pub_test_fig
```



```r
# Plotting the frequency polygon
```

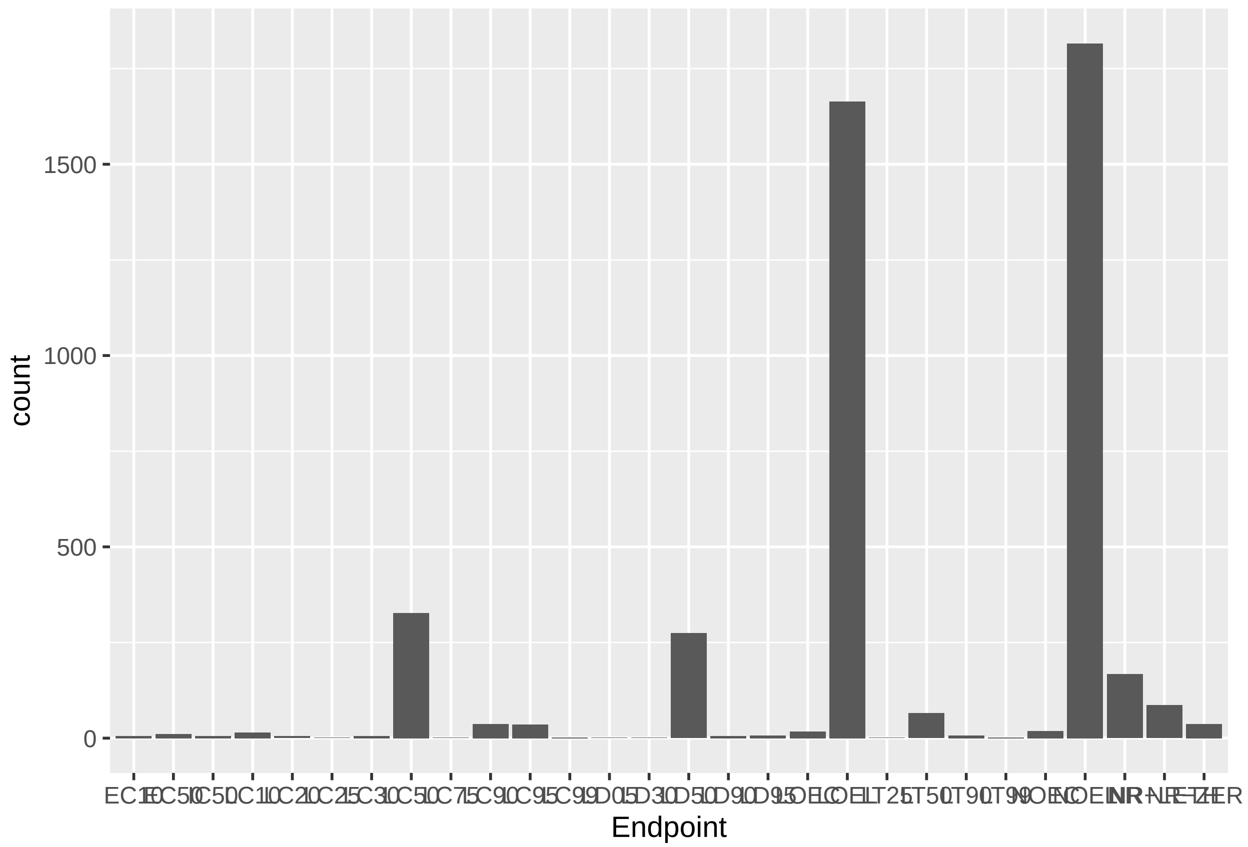
Interpret this graph. What are the most common test locations, and do they differ over time?

> Answer:The most common test location is the lab, but they do differ over time with sometimes the field natural being the greatest, like just before 2010.

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…]

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



Answer:NOEL and LOEL are the two most common end points. LOEL is the Lowest Observed Effect Level - the lowest concentration where there is a statistically adverse effect. NOEL is the no observed effect level - the highest concentration where there is no statistically observed effect.

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"
```

```
#answer is factor, not date
Litter$collectDate <- as.Date(Litter$collectDate, format = "%Y-%m-%d")
class(Litter$collectDate)
```

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

13. Using the `unique` function, list the different plotIDs sampled at Niwot Ridge. How is the information obtained from `unique` different from that obtained from `summary`?

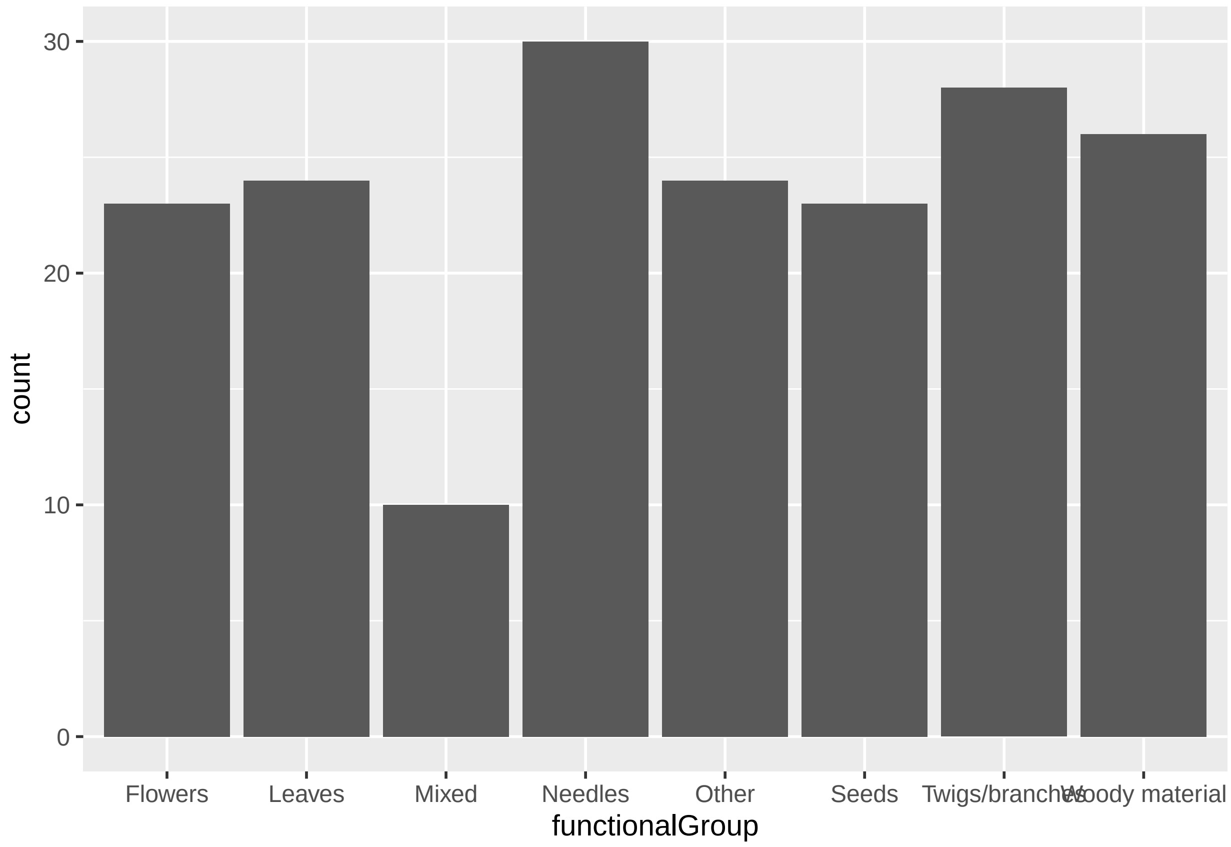
```
unique(Litter$plotID)
```

```
##  [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
```

Answer:12 levels - NIWO_061 NIWO_064 NIWO_067 NIWO_040 NIWO_041 NIWO_063 NIWO_047 NIWO_051 NIWO_058 NIWO_046 NIWO_062 NIWO_057 summary will tell you the count of each level, while unique just shows how many different types there are.

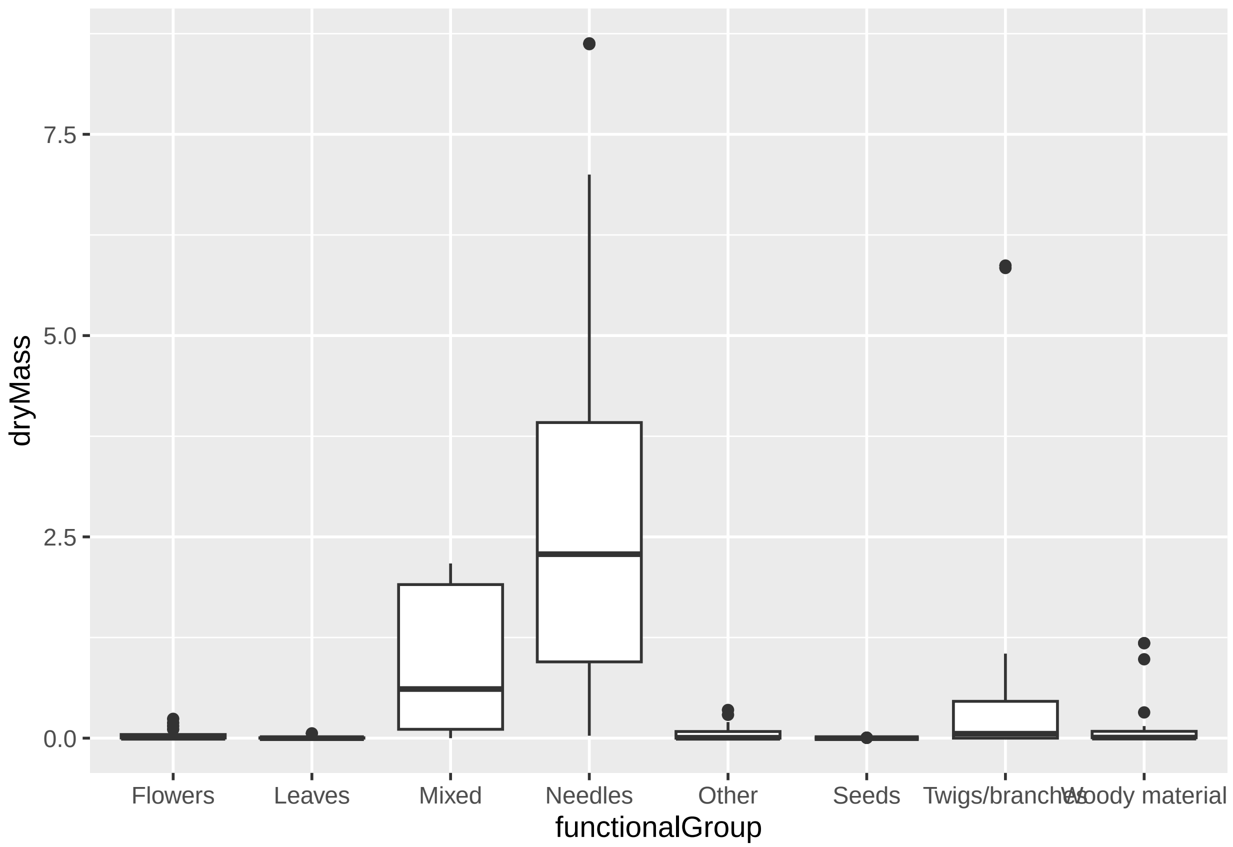
14. Create a bar graph of functionalGroup counts. This shows you what type of litter is collected at the Niwot Ridge sites. Notice that litter types are fairly equally distributed across the Niwot Ridge sites.

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

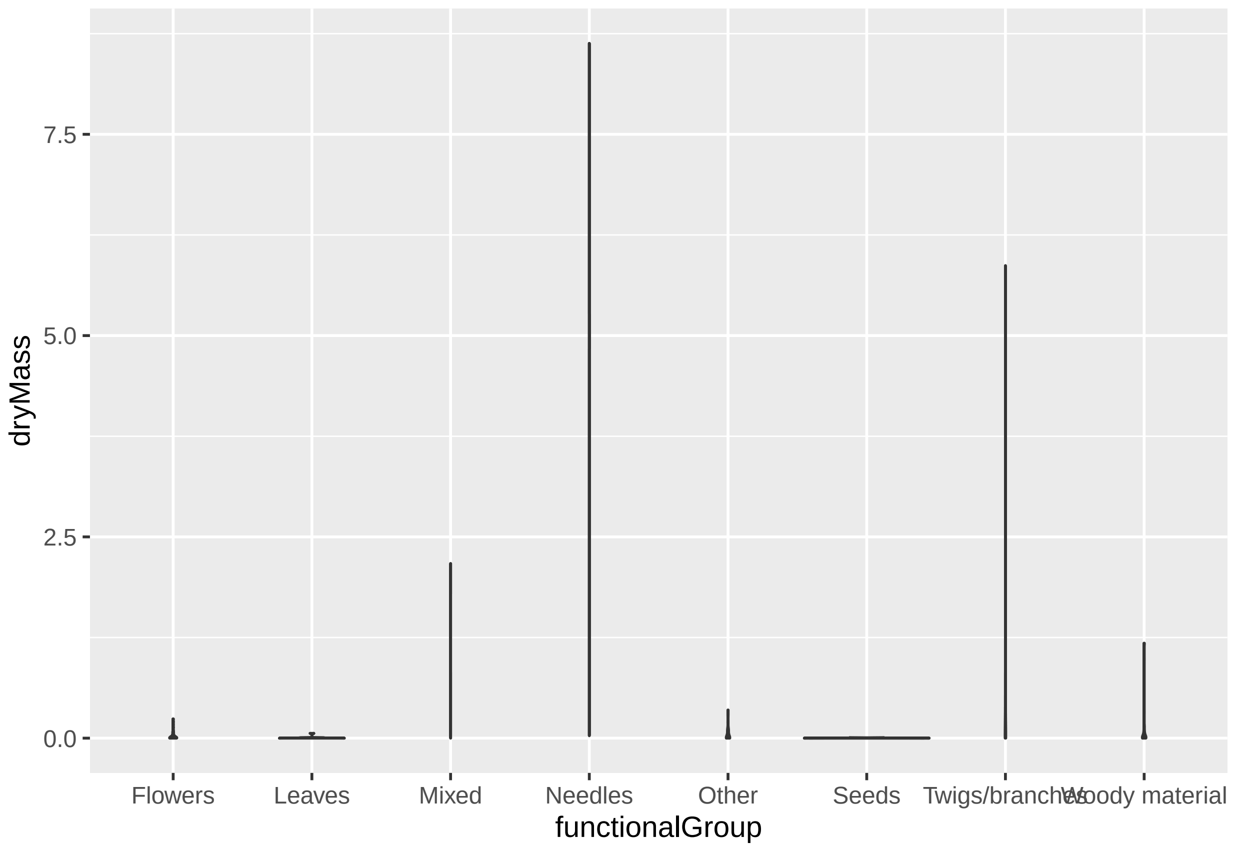


15. Using `geom_boxplot` and `geom_violin`, create a boxplot and a violin plot of dryMass by functionalGroup.

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



```
ggplot(Litter) +
  geom_violin(aes(x = functionalGroup, y = dryMass),
              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:In this case, the violin plots are not really showing violin shape or wide density(where the data is more concentrated), so the visual is lost and you are left with thin vertical lines which do not portray your data to viewers as nicely as the box plot does.

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

> Answer: Needles has the highest, then mixed, then twigs/branches.