Assignment 3: Data Exploration

Mara Michel

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.

```
#install.packages("lubridate")
#install.packages("tidyverse")
#install.packages("here")
#install.packages("ggplot2")
library(here)
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: According to the NIH, neonicotinoids may have a negative impact on insects that are benefical and necessary to agriculture and our food systems, such as bees.

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:According to the U.S. Forest Service, woody debris provides habitats for organisms, provides nutrients back to the soil, and stores carbon. It's important to understand the role that forest debris play in carbon accounting.

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 and woody debris is sampled via both elevated and ground traps. 2.Ground traps are sampled once per year but the frequency of sampling for elevated traps depends on their location and weather conditions. 3.In 2020, the number of elevated traps were reduced but the number of ground traps remained the same.

Obtain basic summaries of your data (Neonics)

5. What are the dimensions of the dataset?

```
ncol(Neonics)
```

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

```
nrow(Neonics)
```

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

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 population (1803) and mortality (1493).When examining the effects of the insecticides, it would be important to study how they impact overall insect populations and how effective they are at killing insects, both intentionally and unintentionally.

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. . .]

```
sort(summary(Neonics$Species.Common.Name))
```

```
##                         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: Besides the "other" category, the six most commonly studied insects are the honey bee (667), parasitic wasp (285), buff tailed bumblebee (183), carniolan honey bee (152), bumble bee (140), and italian honeybee (113). These six species are all pollinators and our food systems rely heavily upon them to pollinate crops.

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?

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

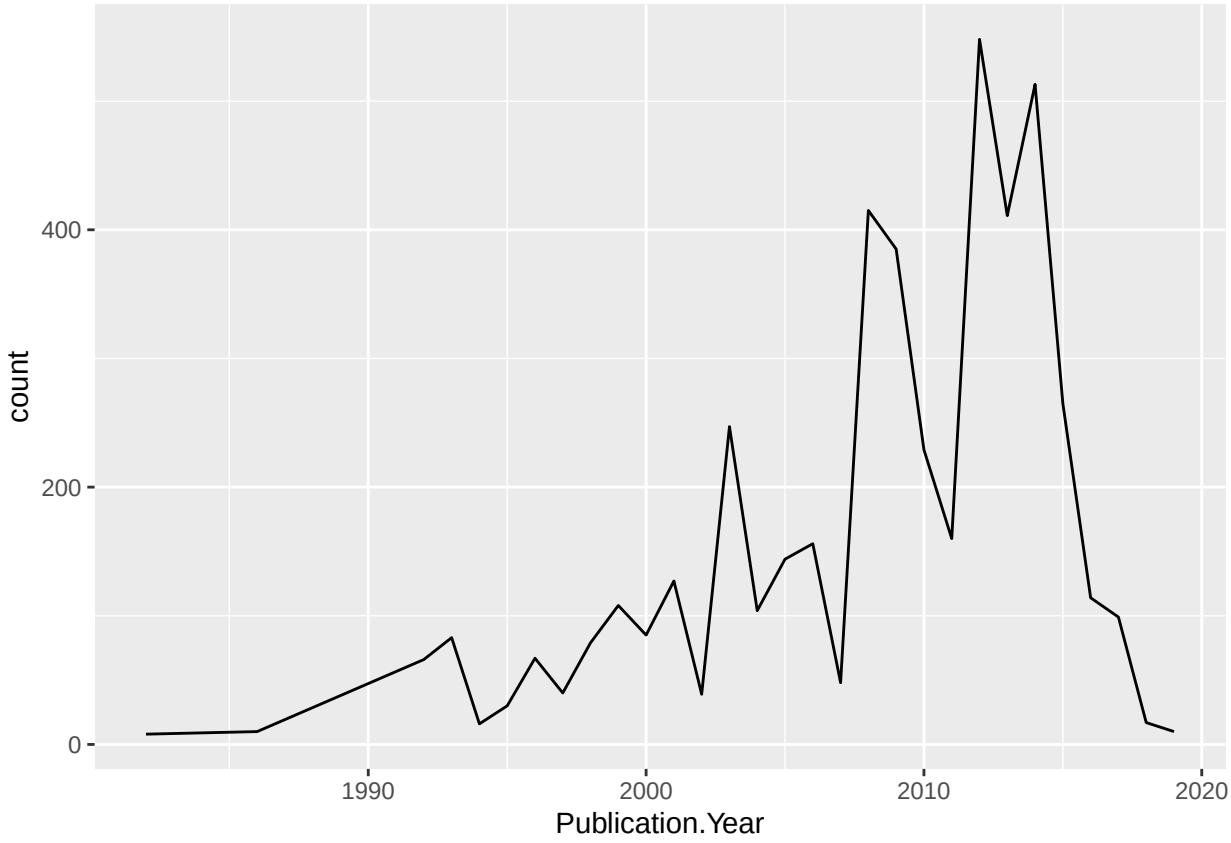
```
## [1] "factor"
```

Answer: Conc.1..Author is a factor because when we imported the data, we asked R to read strings as factors. Conc.1..Author was originally imported as a character value because there are characters in it (ex."NR/" appears multiple times).

Explore your data graphically (Neonics)

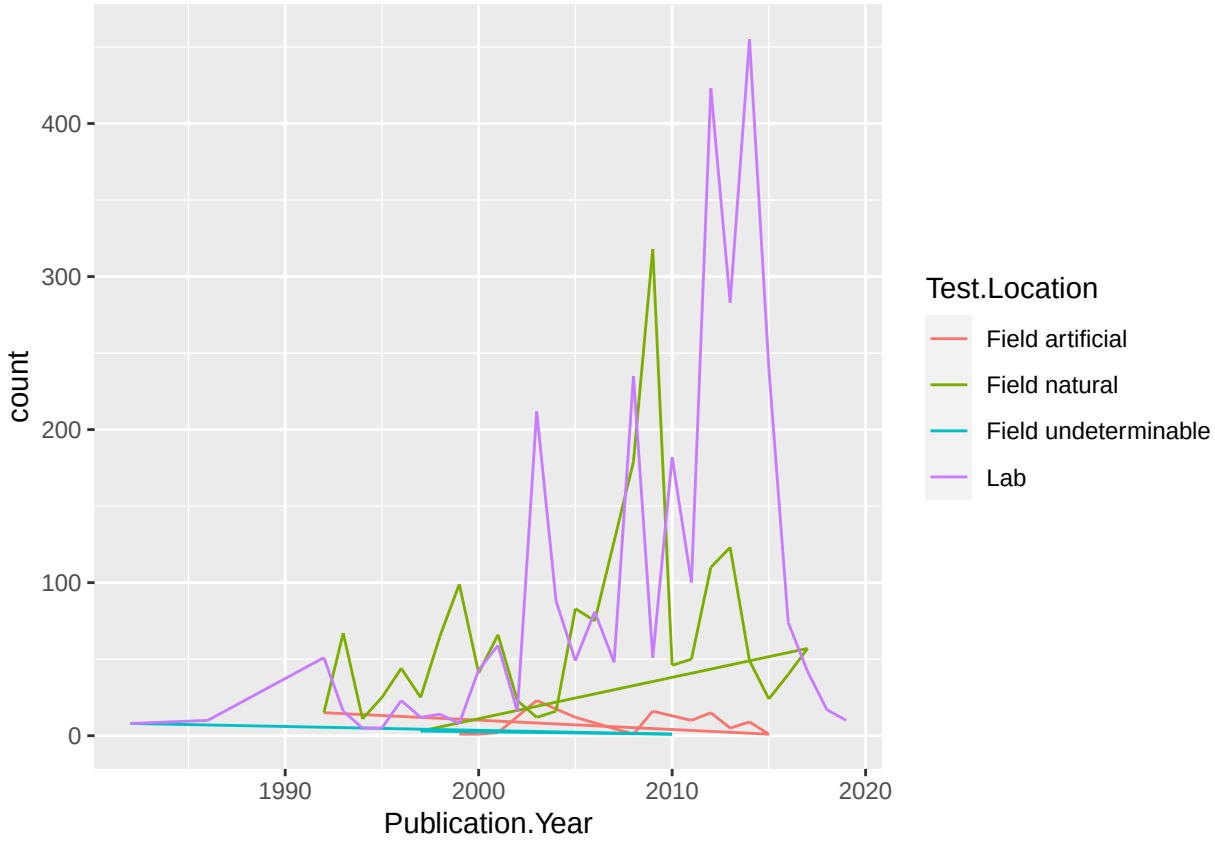
9. Using `geom_freqpoly`, generate a plot of the number of studies conducted by publication year.

```
library(ggplot2)
ggplot(data=Neonics) +
  geom_freqpoly(aes(Publication.Year),stat="count",group='Publication.Year')
```



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

```
ggplot(data=Neonics) +
  geom_freqpoly(aes(Publication.Year,color=Test.Location),stat="count",group='Publication.Year')
```



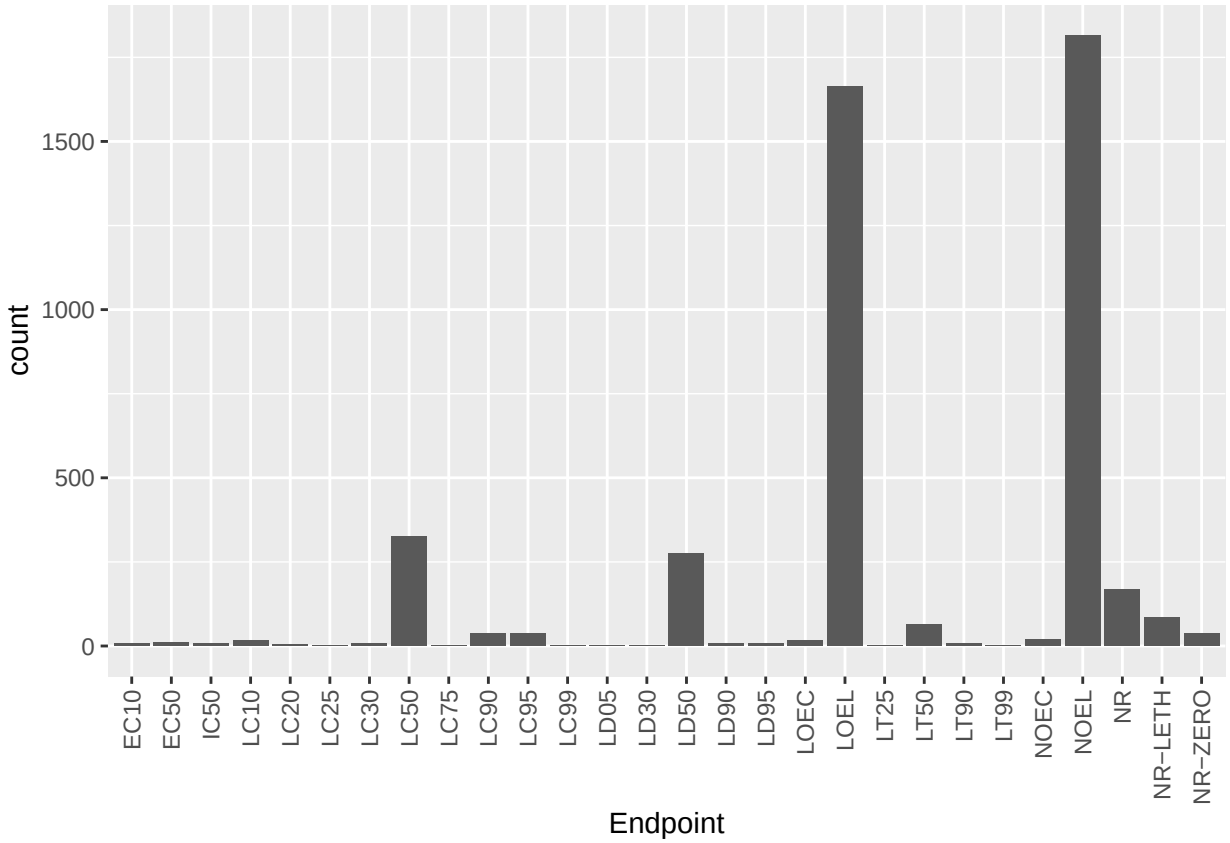
Interpret this graph. What are the most common test locations, and do they differ over time?

> Answer: The most common test locations have been Lab and Field Natural. The two have generally alternated as the most common types over time. However, Lab tests have dominated consistently since 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...]

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



Answer: The two most common end points are LOEL and NOEL. LOEL means "lowest observeable effect level" and NOEL means "No observable effect level."

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.

```
data.class(Litter$collectDate)
```

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

```
#collectDate is a factor
library("lubridate")
```

```
##
## Attaching package: 'lubridate'
```

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

```
Litter$collectDate<-ymd(Litter$collectDate)
Litter$collectDate
```

```
##   [1] "2018-08-02" "2018-08-02" "2018-08-02" "2018-08-02" "2018-08-02"
##   [6] "2018-08-02" "2018-08-02" "2018-08-02" "2018-08-02" "2018-08-02"
##  [11] "2018-08-02" "2018-08-02" "2018-08-02" "2018-08-02" "2018-08-02"
##  [16] "2018-08-02" "2018-08-02" "2018-08-02" "2018-08-02" "2018-08-02"
##  [21] "2018-08-02" "2018-08-02" "2018-08-02" "2018-08-02" "2018-08-02"
##  [26] "2018-08-02" "2018-08-02" "2018-08-02" "2018-08-02" "2018-08-02"
##  [31] "2018-08-02" "2018-08-02" "2018-08-02" "2018-08-02" "2018-08-02"
##  [36] "2018-08-02" "2018-08-02" "2018-08-02" "2018-08-02" "2018-08-02"
##  [41] "2018-08-02" "2018-08-02" "2018-08-02" "2018-08-02" "2018-08-02"
##  [46] "2018-08-02" "2018-08-02" "2018-08-02" "2018-08-02" "2018-08-02"
##  [51] "2018-08-02" "2018-08-02" "2018-08-02" "2018-08-02" "2018-08-02"
##  [56] "2018-08-02" "2018-08-02" "2018-08-02" "2018-08-02" "2018-08-02"
##  [61] "2018-08-02" "2018-08-02" "2018-08-02" "2018-08-02" "2018-08-02"
##  [66] "2018-08-02" "2018-08-02" "2018-08-02" "2018-08-02" "2018-08-02"
##  [71] "2018-08-02" "2018-08-02" "2018-08-02" "2018-08-02" "2018-08-02"
##  [76] "2018-08-02" "2018-08-02" "2018-08-02" "2018-08-02" "2018-08-02"
##  [81] "2018-08-02" "2018-08-02" "2018-08-02" "2018-08-02" "2018-08-02"
##  [86] "2018-08-02" "2018-08-02" "2018-08-02" "2018-08-02" "2018-08-02"
##  [91] "2018-08-02" "2018-08-30" "2018-08-30" "2018-08-30" "2018-08-30"
##  [96] "2018-08-30" "2018-08-30" "2018-08-30" "2018-08-30" "2018-08-30"
## [101] "2018-08-30" "2018-08-30" "2018-08-30" "2018-08-30" "2018-08-30"
## [106] "2018-08-30" "2018-08-30" "2018-08-30" "2018-08-30" "2018-08-30"
## [111] "2018-08-30" "2018-08-30" "2018-08-30" "2018-08-30" "2018-08-30"
## [116] "2018-08-30" "2018-08-30" "2018-08-30" "2018-08-30" "2018-08-30"
## [121] "2018-08-30" "2018-08-30" "2018-08-30" "2018-08-30" "2018-08-30"
## [126] "2018-08-30" "2018-08-30" "2018-08-30" "2018-08-30" "2018-08-30"
## [131] "2018-08-30" "2018-08-30" "2018-08-30" "2018-08-30" "2018-08-30"
## [136] "2018-08-30" "2018-08-30" "2018-08-30" "2018-08-30" "2018-08-30"
## [141] "2018-08-30" "2018-08-30" "2018-08-30" "2018-08-30" "2018-08-30"
## [146] "2018-08-30" "2018-08-30" "2018-08-30" "2018-08-30" "2018-08-30"
## [151] "2018-08-30" "2018-08-30" "2018-08-30" "2018-08-30" "2018-08-30"
## [156] "2018-08-30" "2018-08-30" "2018-08-30" "2018-08-30" "2018-08-30"
## [161] "2018-08-30" "2018-08-30" "2018-08-30" "2018-08-30" "2018-08-30"
## [166] "2018-08-30" "2018-08-30" "2018-08-30" "2018-08-30" "2018-08-30"
## [171] "2018-08-30" "2018-08-30" "2018-08-30" "2018-08-30" "2018-08-30"
## [176] "2018-08-30" "2018-08-30" "2018-08-30" "2018-08-30" "2018-08-30"
## [181] "2018-08-30" "2018-08-30" "2018-08-30" "2018-08-30" "2018-08-30"
## [186] "2018-08-30" "2018-08-30" "2018-08-30"
```

```
data.class(Litter$collectDate)
```

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

```
unique(Litter$collectDate)
```

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

```
#Litter was sampled on August 2,2018 and August 30,2018
```

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`?

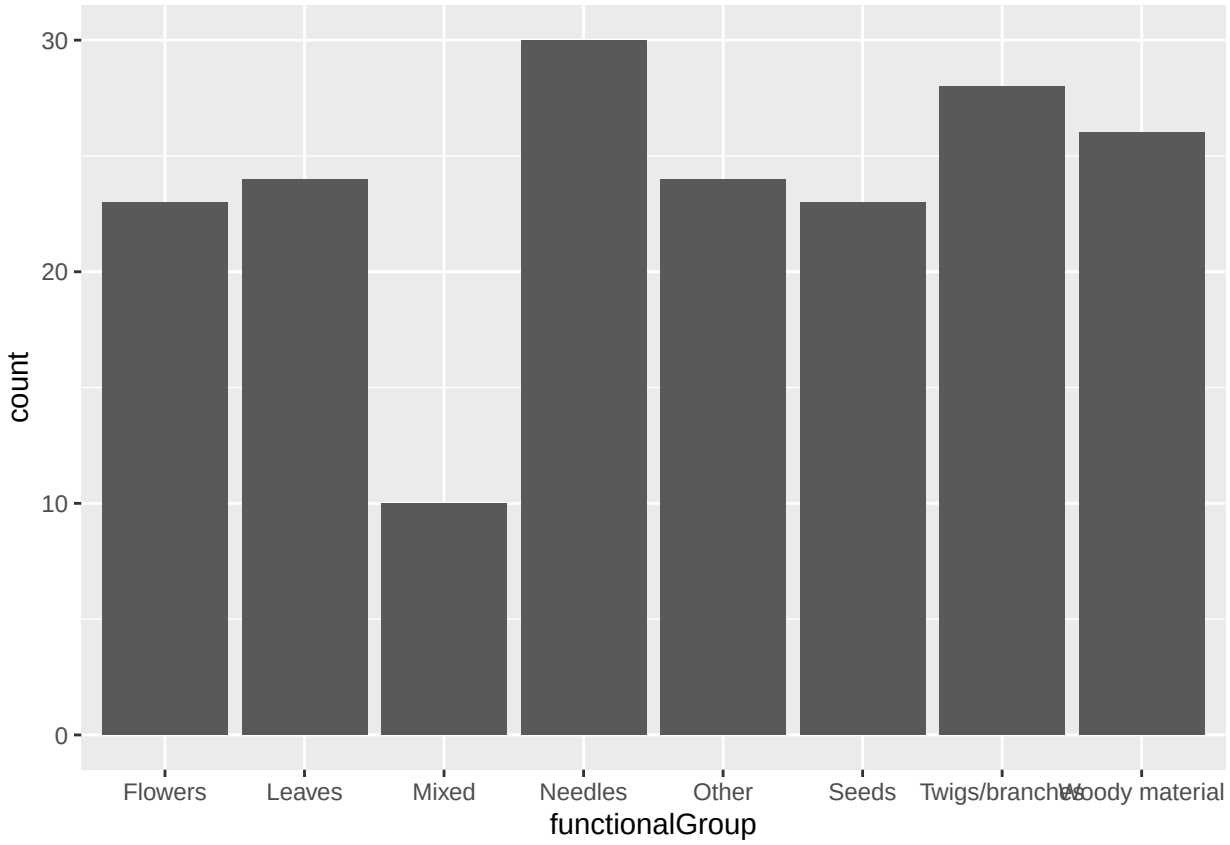
```
length(unique(Litter$plotID))
```

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

Answer:12 plots were sampled at Niwot Ridge. The 'unique' function determines how many unique values are in a list whereas the 'summary' function determines how many times each value repeats in a list.

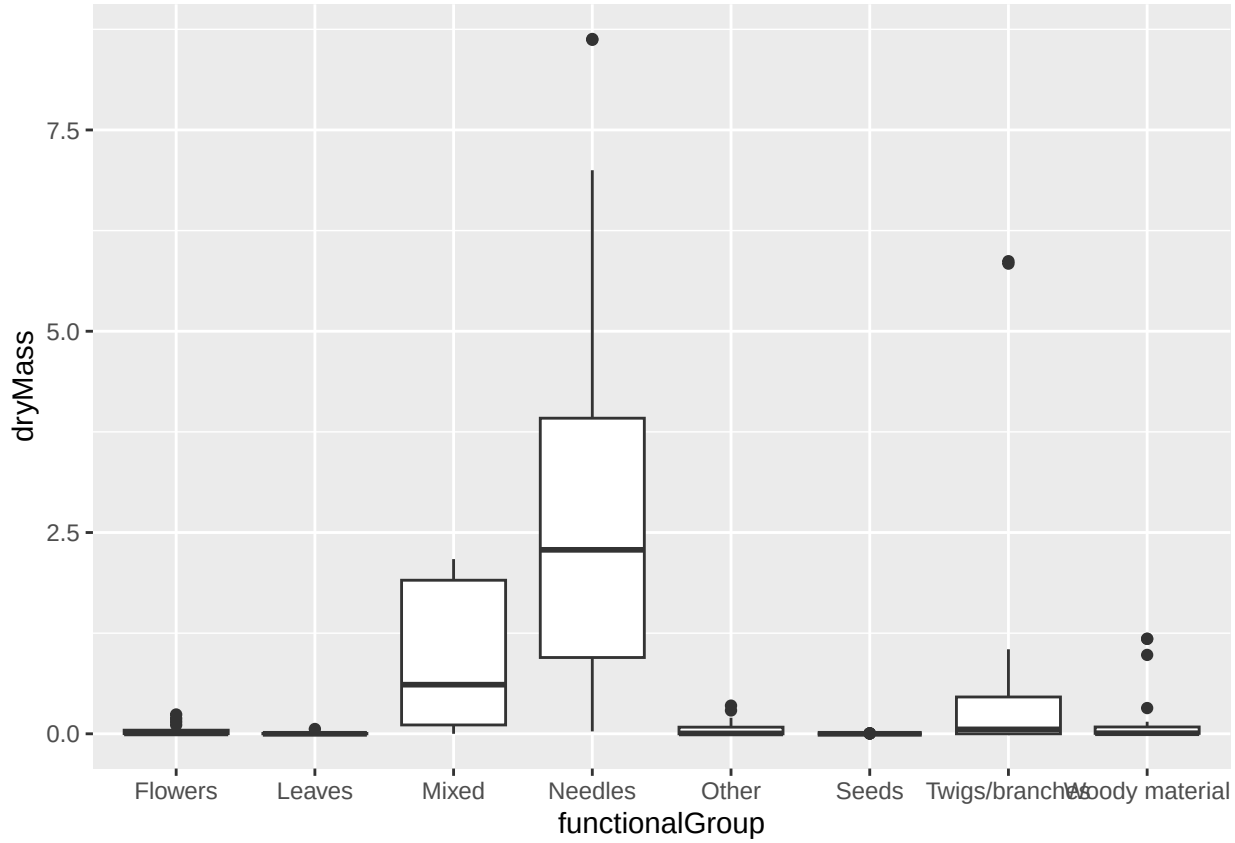
14. Create a bar graph of functionalGroup counts. This shows you what type of litter is collected at the Niwot Ridge sites. Notice that litter types are fairly equally distributed across the Niwot Ridge sites.

```
ggplot(data=Litter) +
  geom_bar(aes(functionalGroup),stat="count",group='functionalGroup')
```

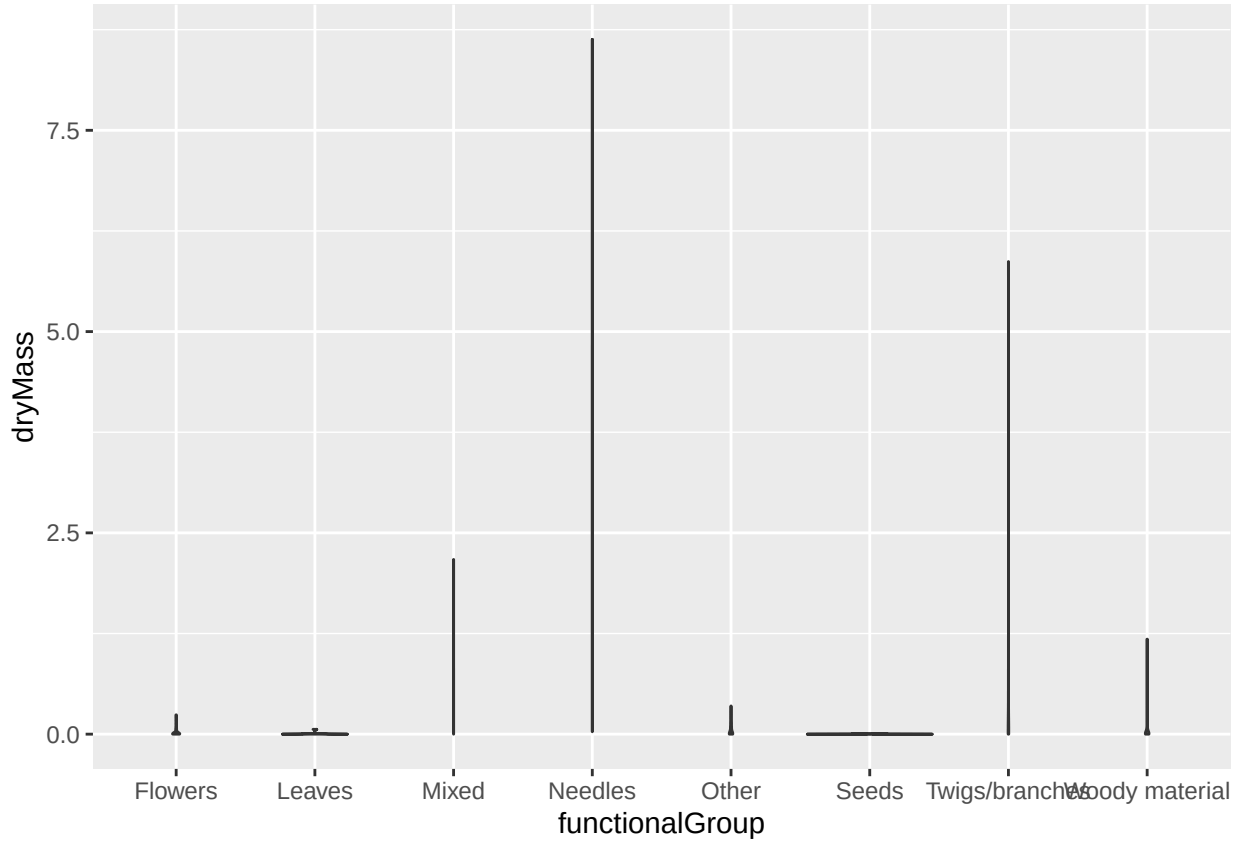


15. Using `geom_boxplot` and `geom_violin`, create a boxplot and a violin plot of dryMass by functionalGroup.

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




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



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

> Answer: A boxplot shows the interquartile range, helping us to observe the distribution of dry mass. The violin plot just shows the spread of values.

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

> Answer: Needles and "Mixed" litter have the highest biomass. Twigs/ branches has outliers on the higher range but overall its median dry mass is much lower than Needles and "Mixed".