Assignment 3: Data Exploration

Andreana Chou

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(tidyverse)
library(tidyverse)          #installed and loaded tidyverse

#install.packages(lubridate)
library(lubridate)          #installed and loaded 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: Neonicotinoids can impact the biodiversity and abundance of insects. Insects are pollinators and crucial to agriculture. Insects are also a source of food for key species in the food chain.

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: Woody debris and litter in forests are critical to the ecosystem's carbon and nutrient cycle. It can provide food for organisms, and affect the absorption of water and buildup of sediments in forests.

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 can be grouped as: leaves, needles, twigs/branches, woody material, seeds, flowers, lichen/mosses, unsorted material 2. Trap placement within plots may be targeted or randomized depending on vegetation 3. Ground traps are sampled once a year, while elevated traps are sampled at varying frequencies depending on vegetation

Obtain basic summaries of your data (Neonics)

5. What are the dimensions of the dataset?

```
Neonics_dimensions <- dim(Neonics)
print(Neonics_dimensions)
```

```
## [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?

```
Effect_summary <- summary(Neonics$Effect)
print(Effect_summary)
```

```
##     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 effects are of interest to scientists because a chemical may have varying impacts on insects. Some chemicals may alter behavior, or stunt development, or increase mortality, or accumulate in organisms. The summary function counts the number of different observed effects for each categorical group.

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...]

```
common_species <- summary(Neonics$Species.Common.Name)
print(sort(common_species))
```

```
##                          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: The six most commonly studied species are: honey bee, parasitic wasp, Buff Tailed bumblebee, Carniolan honey bee, bumble bee, Italian honeybee. These insects all belong to the order Hymenoptera and are pollinators. Pollinators are likely studied more frequently because of their importance to the agricultural system.

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?

```
Conc_1_Author_class <- class(Neonics$Conc.1..Author.)
print(Conc_1_Author_class)
```

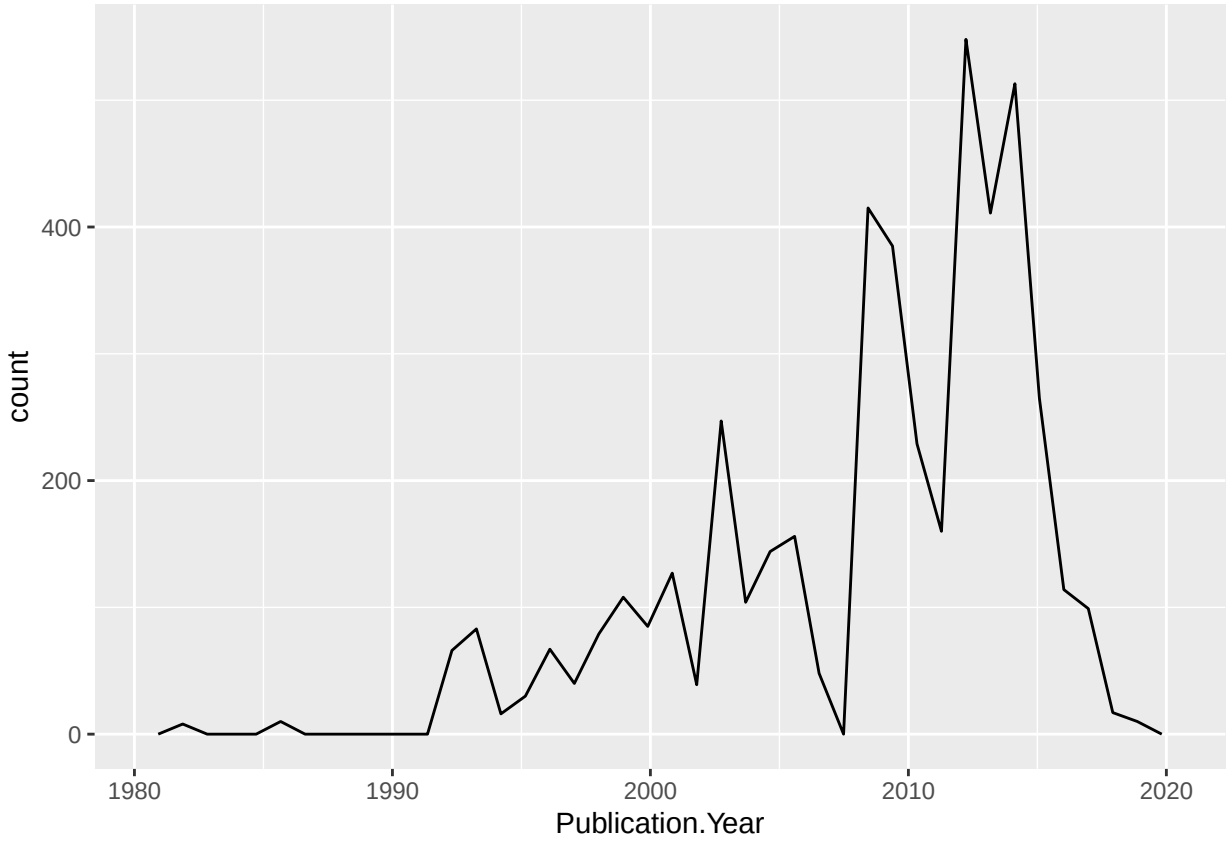
```
## [1] "factor"
```

Answer: The class of 'Conc.1...Author' is factor. It is not numeric because there are several values in Conc.I...Author that have other symbols such as "~" and "/".

Explore your data graphically (Neonics)

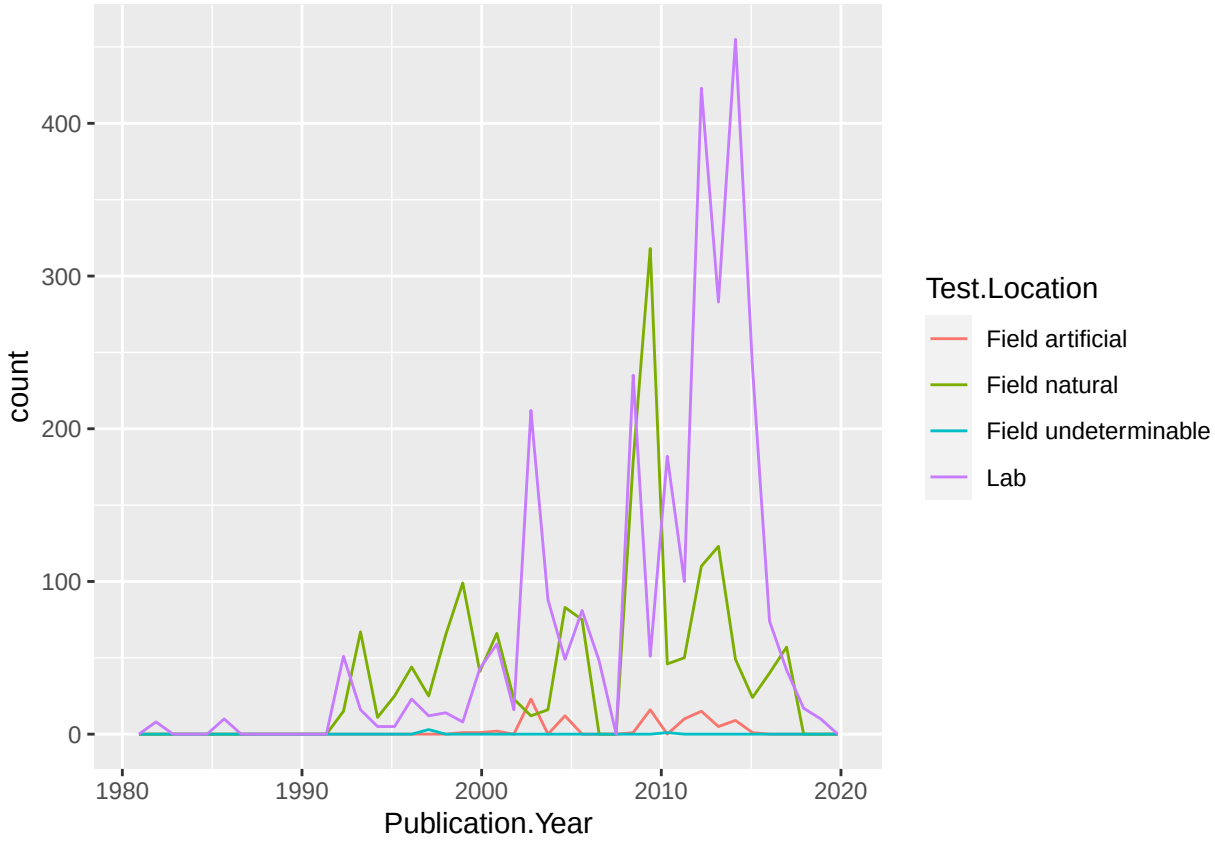
9. Using `geom_freqpoly`, generate a plot of the number of studies conducted by publication year.

```
studies_year <- ggplot(Neonics) + geom_freqpoly(aes(x=Publication.Year), bins=40)

#set bins equal to 40 due to range of ~40 years

studies_year
```



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

```
studies_year_location <- ggplot(Neonics) + geom_freqpoly(aes(x=Publication.Year, color=Test.Location), l

studies_year_location
```



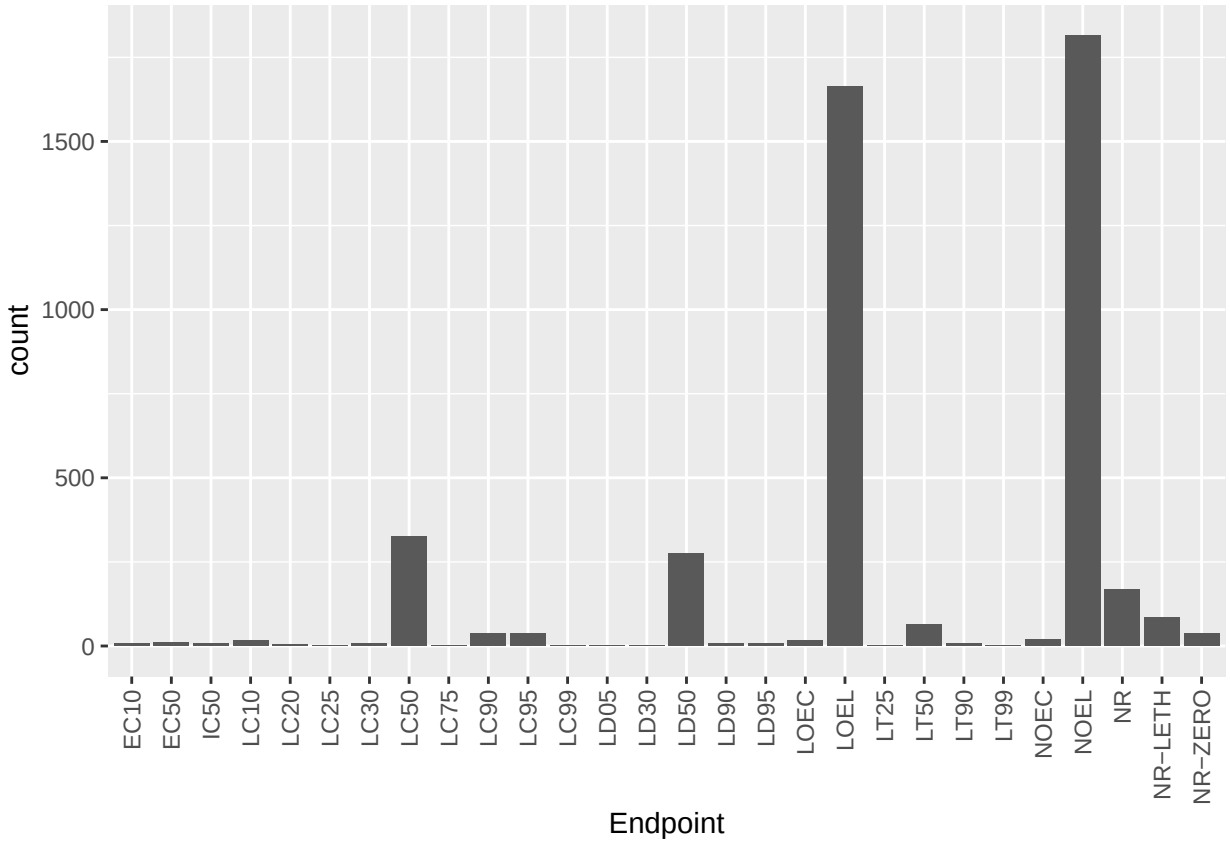
Interpret this graph. What are the most common test locations, and do they differ over time?

> Answer: The most common location is field natural, which occurred around 2010. Overall, field natural remained the dominant location throughout the data collection period.

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...]

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



Answer: The two most common endpoints are NOEL and LOEL. NOEL stands for no-observable-effects-level at the highest dose/concentration level. LOEL stands for lowest-observable-effect-level at the lowest dose/concentration 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.

```
class(Litter$collectDate)     #collectDate is factor
```

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

```
Litter$collectDate <- ymd(Litter$collectDate)
#used year-month-date function from lubridate package

class(Litter$collectDate)     #collectDate is date
```

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

```
August_dates <- unique(Litter$collectDate)
#unique function for collectDate column
August_dates
```

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

Answer: Litter was sampled during August 2 and August 30 in 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`?

```
unique_plots <- unique(Litter$plotID)      #unique function for plotID column
print(unique_plots)       #print list of unique plots
```

```
##  [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
```

```
length(unique_plots)      #length() to count number of characters
```

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

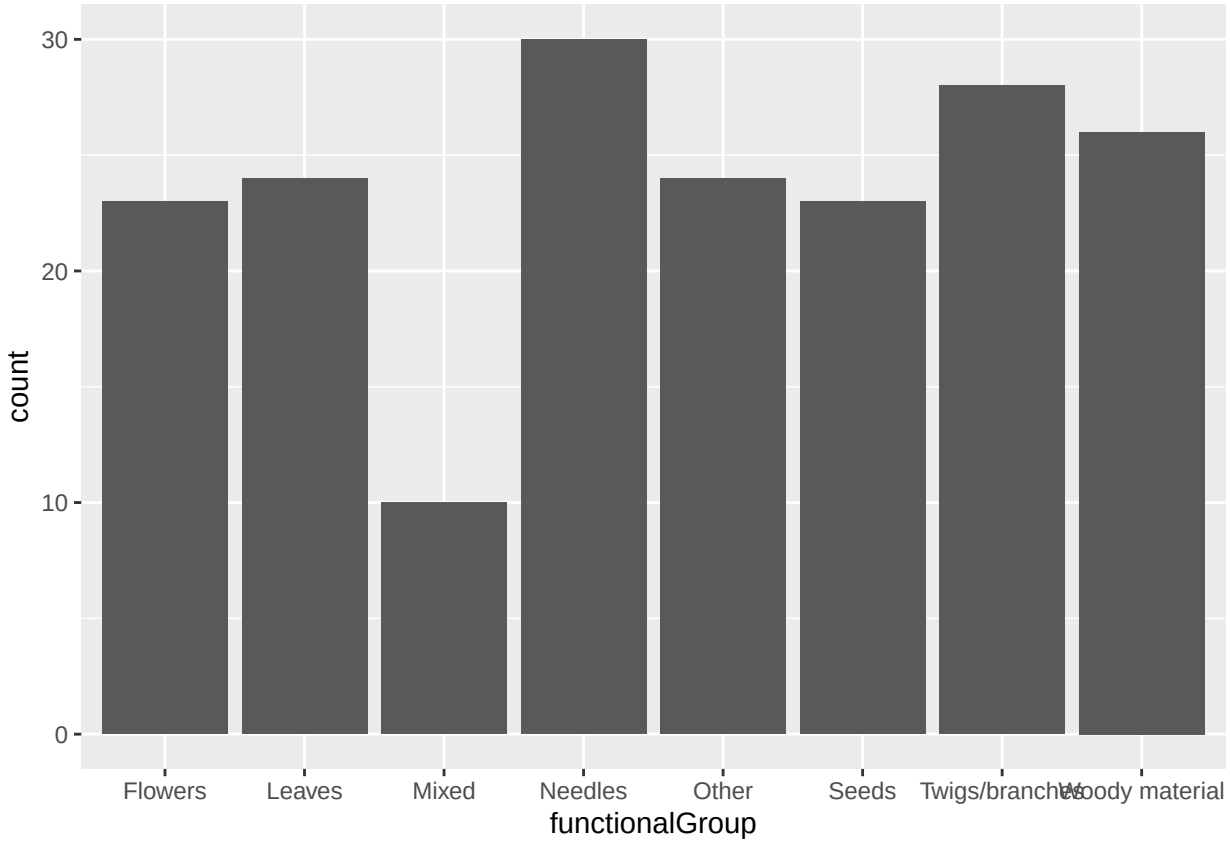
```
plot_sum <- summary(Litter$plotID)
print(plot_sum)
```

```
## 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 unique plots were sampled at Niwot Ridge. The unique function gives us the total number of unique plots in the Litter dataset. The summary function gives us all the unique plots, but tallies the frequency of samples for each unique plot.

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.

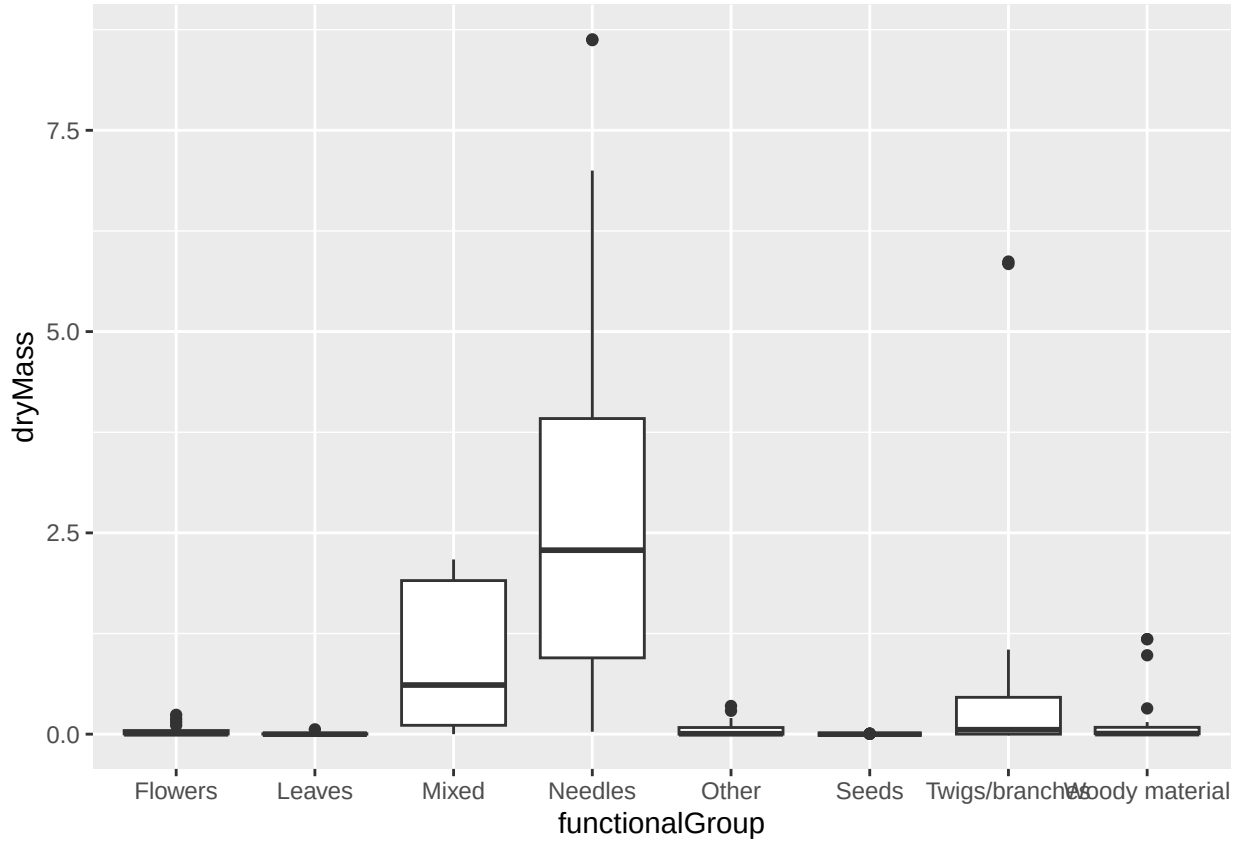
```
litter_functional_group <- ggplot(Litter, aes(x=functionalGroup))+geom_bar()
litter_functional_group
```



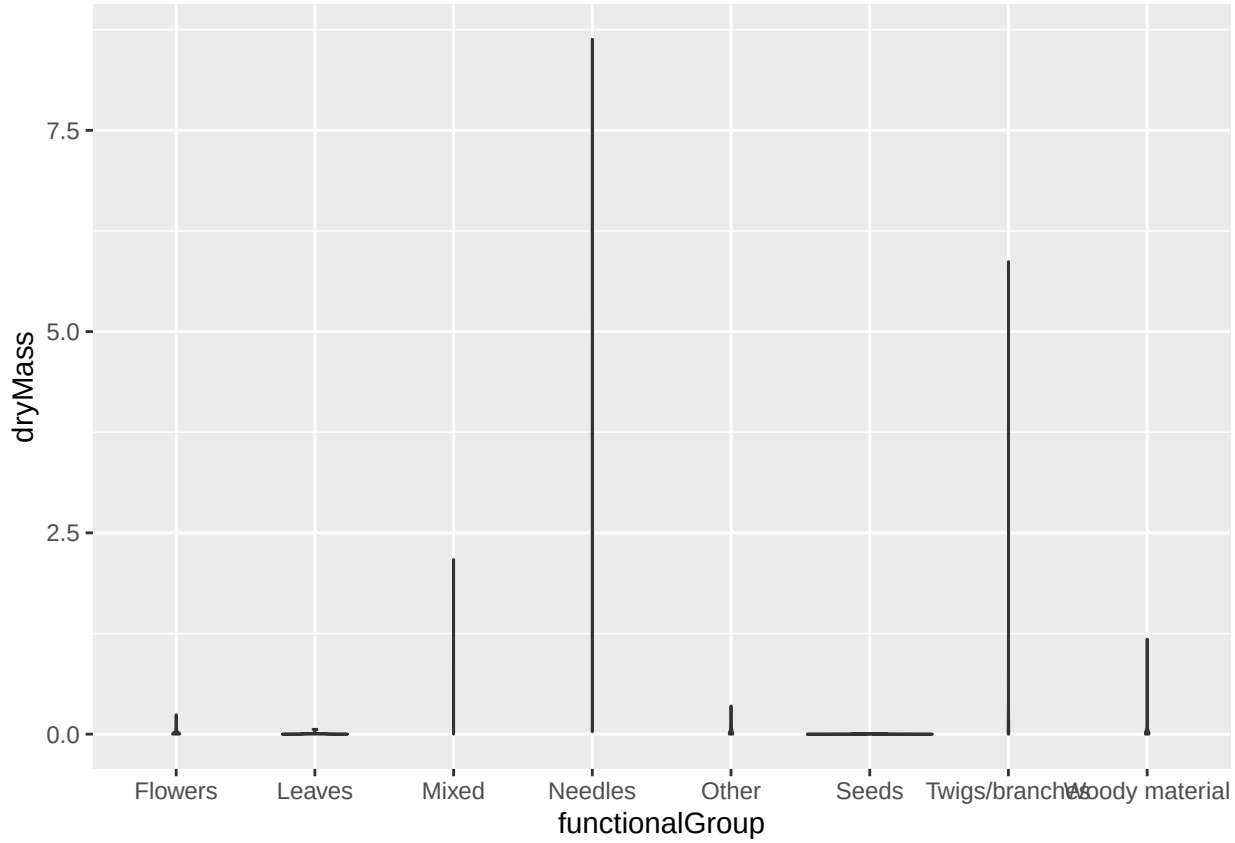

```
#set categories as functionalGroup categories
```

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

```
dryMass_box <- ggplot(Litter) + geom_boxplot(aes(x=functionalGroup, y=dryMass))
dryMass_box     #set functionalGroup as x-axis, measure of dryMass as y-axis
```




```
dryMass_violin <- ggplot(Litter) + geom_violin(aes(x=functionalGroup, y=dryMass))
dryMass_violin  #set functionalGroup as x-axis, measure of dryMass as y-axis
```



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

> Answer: The violin plot compressed all the dryMass data, making it difficult to interpret the data.The boxplot also more effectively demonstrated the outliers and median of each functional group's dry mass measurements.

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

> Answer: Needles (median is the highest out of all other litter groups)