Assignment 3: Data Exploration

Analise Lindborg

OVERVIEW

This exercise accompanies the lessons in Environmental Data Analytics on Data Exploration.

Directions

1. Change "Student Name" on line 3 (above) with your name.
2. Work through the steps, **creating code and output** that fulfill each instruction.
3. Be sure to **answer the questions** in this assignment document.
4. When you have completed the assignment, **Knit** the text and code into a single PDF file.
5. After Knitting, submit the completed exercise (PDF file) to the dropbox in Sakai. Add your last name into the file name (e.g., "Salk_A03_DataExploration.Rmd") prior to submission.

The completed exercise is due on <>.

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.

```
knitr::opts_knit$set(root.dir = '/Users/analiselindborg/Desktop/Desktop - Analise's MacBook Pro/Data Ana
```

```
getwd()
```

```
## [1] "/Users/analiselindborg/Desktop/Desktop - Analise's MacBook Pro/Data Analytics/Environmental_Data
```

```
library(tidyverse)

Neonics <- read.csv("./Data/Raw/ECOTOX_Neonicotinoids_Insects_raw.csv")

Litter <- read.csv("./Data/Raw/NEON_NIWO_Litter_massdata_2018-08_raw.csv")
```

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 ecotoxicologoy of neonicotinoids on insects? Feel free to do a brief internet search if you feel you need more background information.

   Answer: We are interested in ecotoxicology of neonics because it is important to evaluate toxicity for all insect groups. Neonics are highly effective class of pesticides but they are not well targeted. They often are toxic to non-target groups of insects (honey bees being the most well-known example).

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 can be indicators of habitat suitability for different species (i.e. species like mice that need more cover may thrive in forest systems with more woody debris). It can also be an indicator of live:dead biomass, where increases in hte proportion of dead biomass can have implications for carbon sequestration.

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: *Samples are collected via ground and elevated traps and separated into plant functional groups.* Sample sites are selected randomly as 40 20x20m plots within a 90% flux footprint of the two airsheds *Ground traps are sampled once per year and elevated traps are sampled at intervals dependent on vegetation type

Obtain basic summaries of your data (Neonics)

5. What are the dimensions of the dataset?

```
#Neonics
dim(Neonics)
```

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

6. Using the `summary` function on the "Effects" column, determine the most common effects that are studied. Why might these effects specifically be of interest?

```
#Make factor first, otherwise summary() returns only the length and vector type
Neonics$Effect <- as.factor(Neonics$Effect)

#summary
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 that are studied are mortality and population. Population is often studied because it is a general parameter of how well a species is doing (assumption that large or stable populations are doing okay, declining populations may be impacted). Mortality is often studied because it is a difinitive measure of toxicity. If you expose certain species to a chemical and they die at a significant level, you can conclude that the chemical is likely toxic to that species.

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.

```
Neonics$Species.Common.Name<- as.factor(Neonics$Species.Common.Name)

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: Honey Bee, Parasitic Wasp, Buff Tailed Bumblebee, Carniolan Honey Bee, Bumble Bee, Italian Honeybee. Most of these species are in the bee family. They may be of interest over other species because bees have shown particular susceptability to neonics. Bee populations, particularly

honey bee and bumble bee populations, have been in rapid decline over the past few years and this is largely atrributed to the use of neonic pesticides.

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] "character"
```

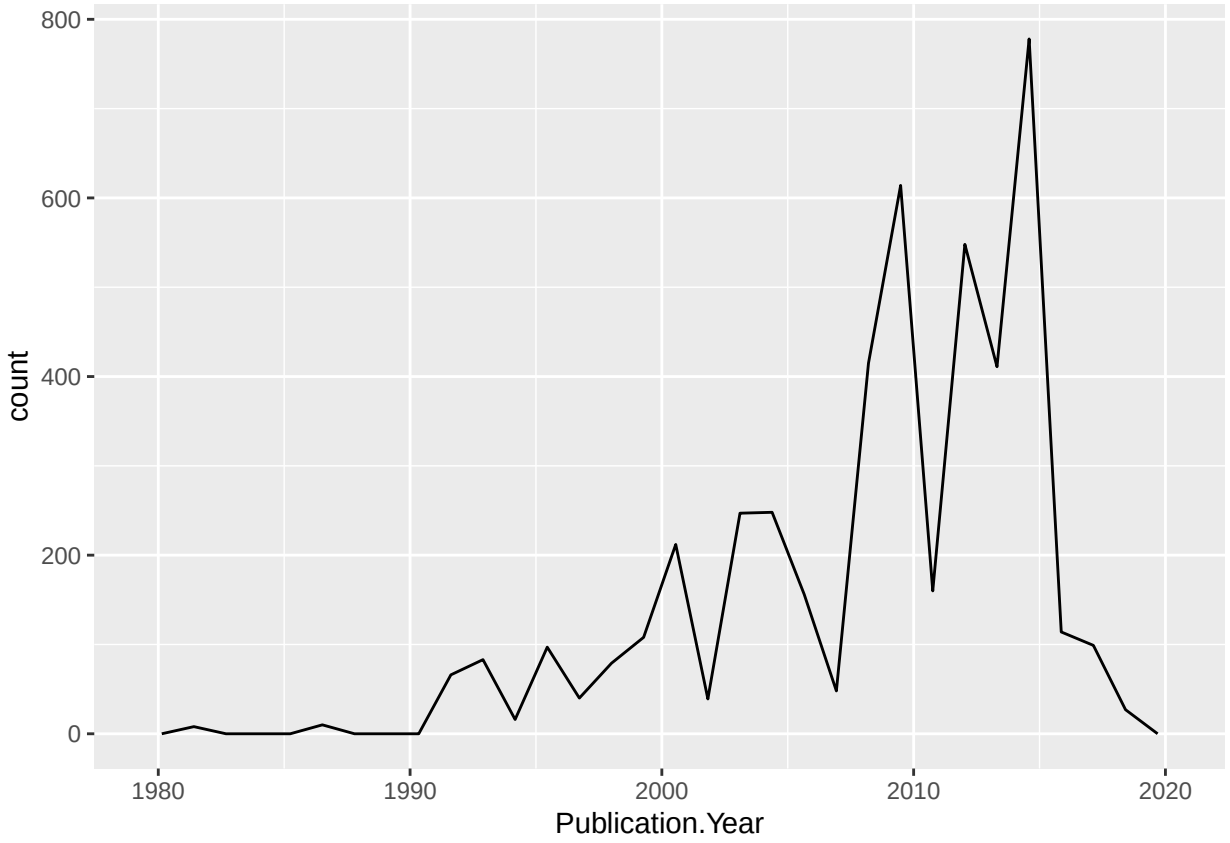
Answer: It is a character because the data column contains non-numeric characters (~, /, NR, etc.)

Explore your data graphically (Neonics)

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

```
pub.plot <- ggplot(Neonics, aes(Publication.Year)) +
  geom_freqpoly()
pub.plot
```

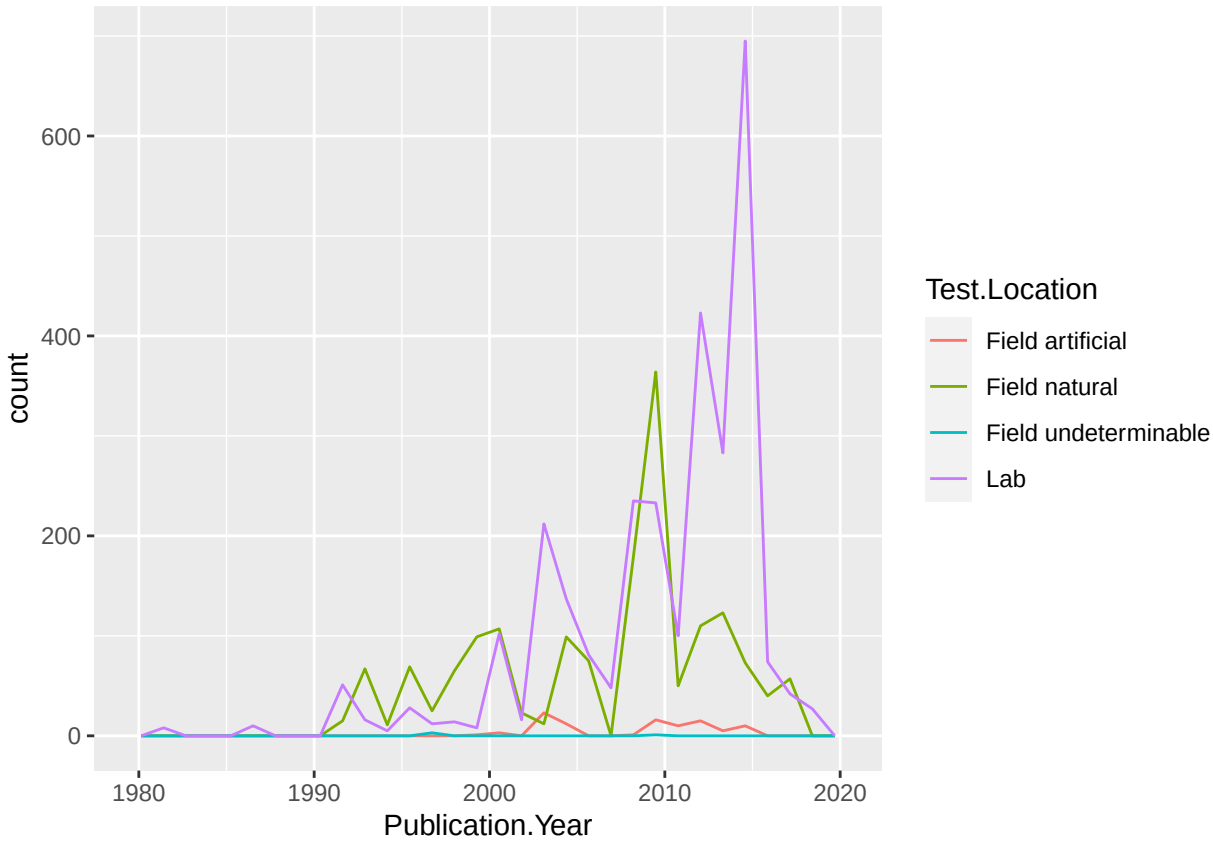
```
## `stat_bin()` using `bins = 30`. Pick better value with `binwidth`.
```



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

```
pub.plot.color <- ggplot(Neonics, aes(Publication.Year, color = Test.Location)) +
  geom_freqpoly()
pub.plot.color
```

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

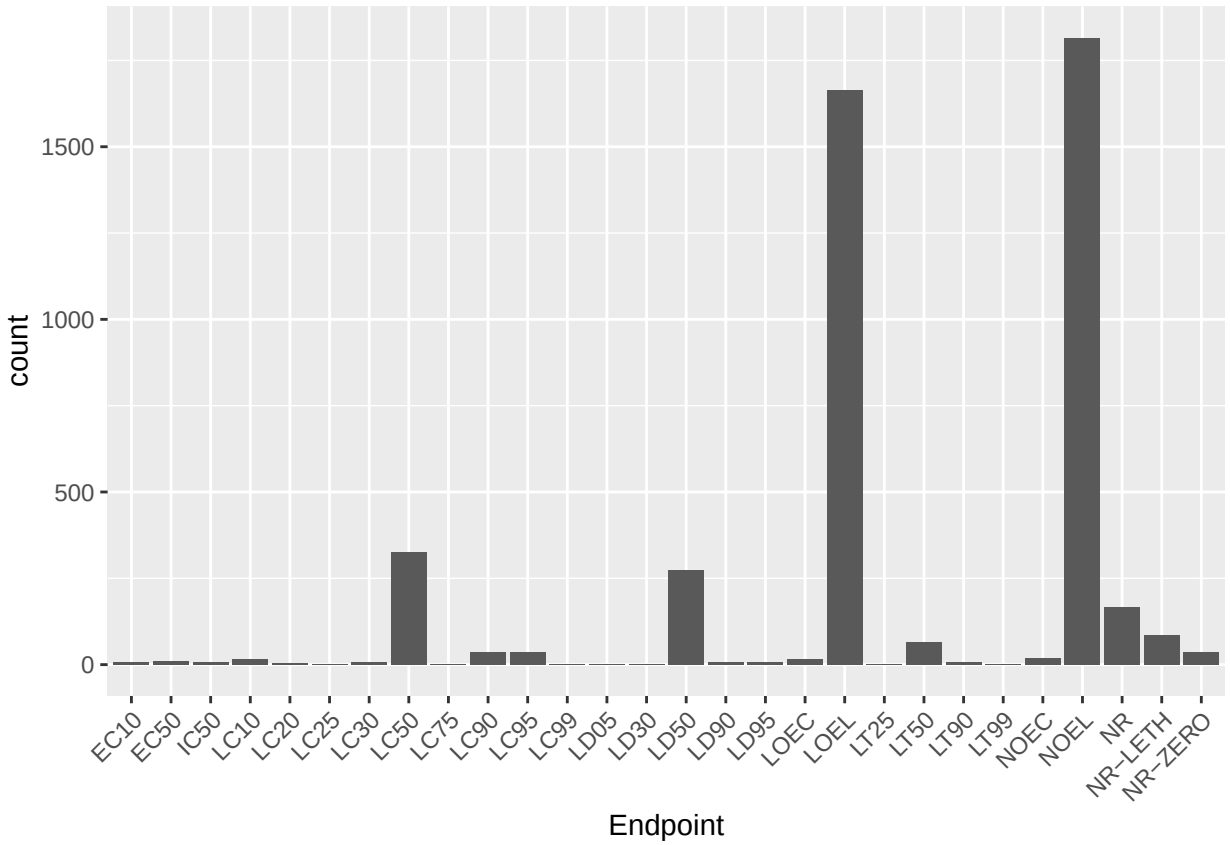


Interpret this graph. What are the most common test locations, and do they differ over time?

Answer: The most common test locations are field natural and lab. Both gradually increased over time, with lab studies taking over significantly around 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.

```
endpoint.plot <- ggplot(Neonics, aes(Endpoint)) +
  geom_bar() +
  #add theme to see labels, otherwise they overlap
  theme(axis.text.x = element_text(angle = 45, hjust = 1))
endpoint.plot
```



Answer: The two most common endpoints are LOEL and NOEL. These are the lowest observed effect concentration and the no observed effect concentration, respectively. The LOEC is the lowest concentration of a chemical where a statistically significant effect is observed on the study species (effect being the toxicity endpoint being studied (mortality, weight gain, etc.). The NOEC is the highest concentration for which there is no statistically significant adverse effect observed.

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) #character
```

```
## [1] "character"
```

```
#converting to date
Litter$collectDate <- as.Date(Litter$collectDate, format = "%Y-%m-%d")

#litter samples in August 2018
unique(Litter$collectDate)
```

```
## [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`?

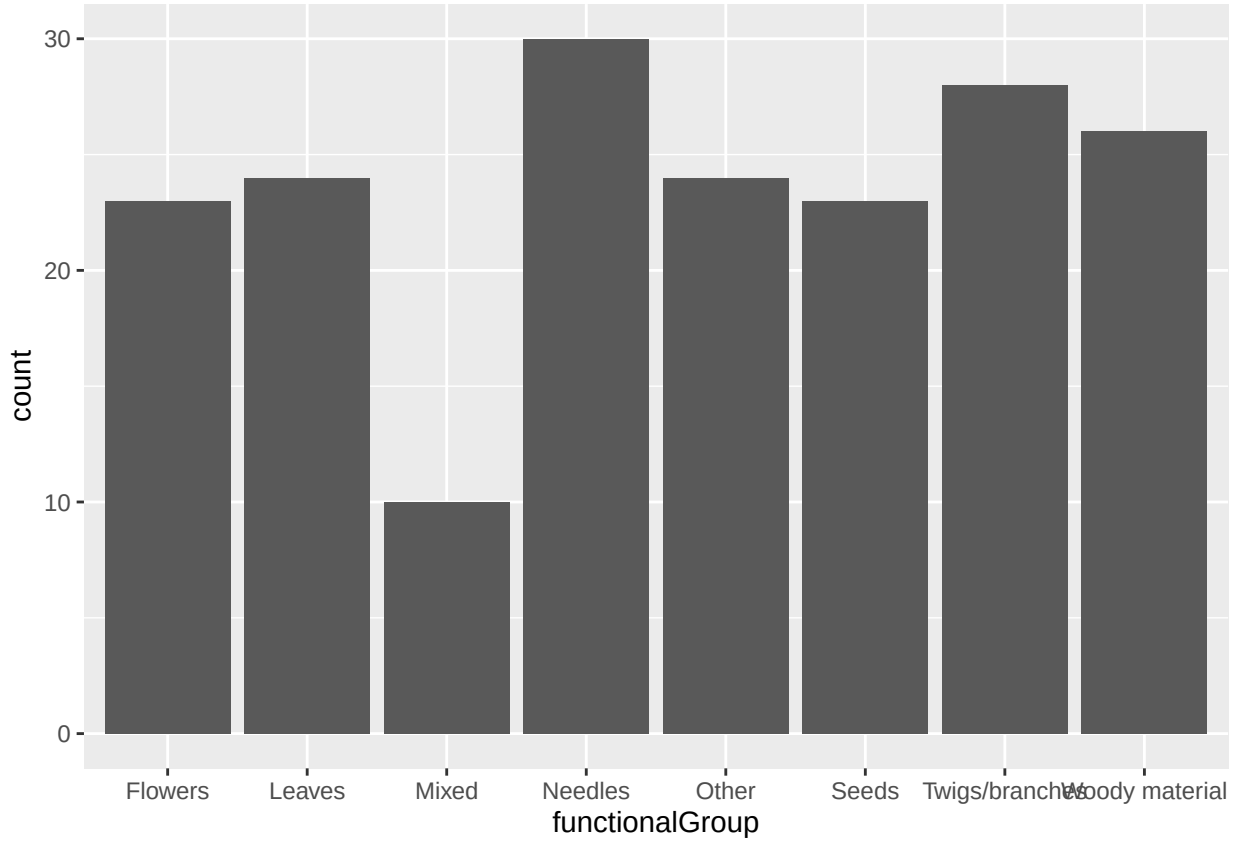
```
unique(Litter$plotID)
```

```
##  [1] "NIWO_061" "NIWO_064" "NIWO_067" "NIWO_040" "NIWO_041" "NIWO_063"
##  [7] "NIWO_047" "NIWO_051" "NIWO_058" "NIWO_046" "NIWO_062" "NIWO_057"
```

Answer: 12 plots were sampled. Unique is different because it returns a list of all unique values in a column, but does not tell you the frequency or count of those values.

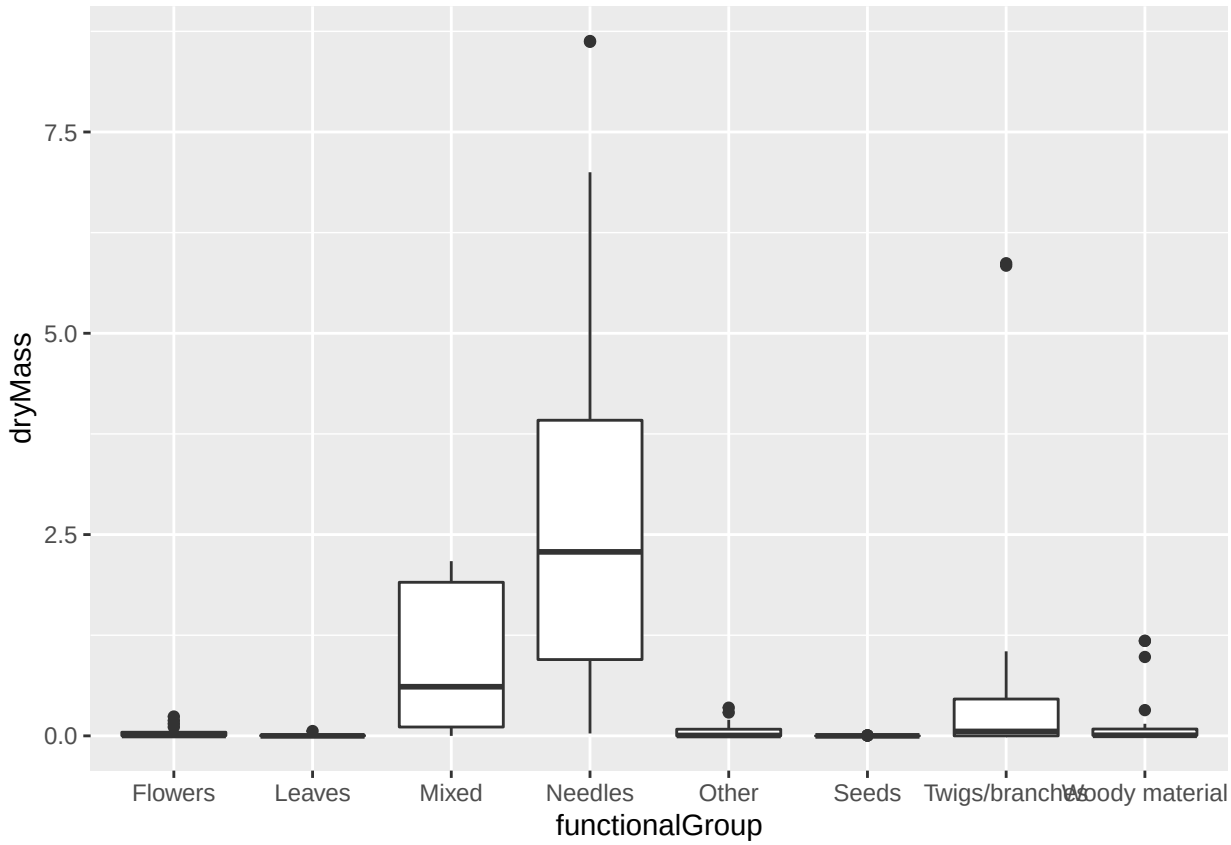
14. Create a bar graph of functionalGroup counts. This shows you what type of litter is collected at the Niwot Ridge sites. Notice that litter types are fairly equally distributed across the Niwot Ridge sites.

```
fun.group.plot <- ggplot(Litter, aes(functionalGroup)) +
  geom_bar()
fun.group.plot
```

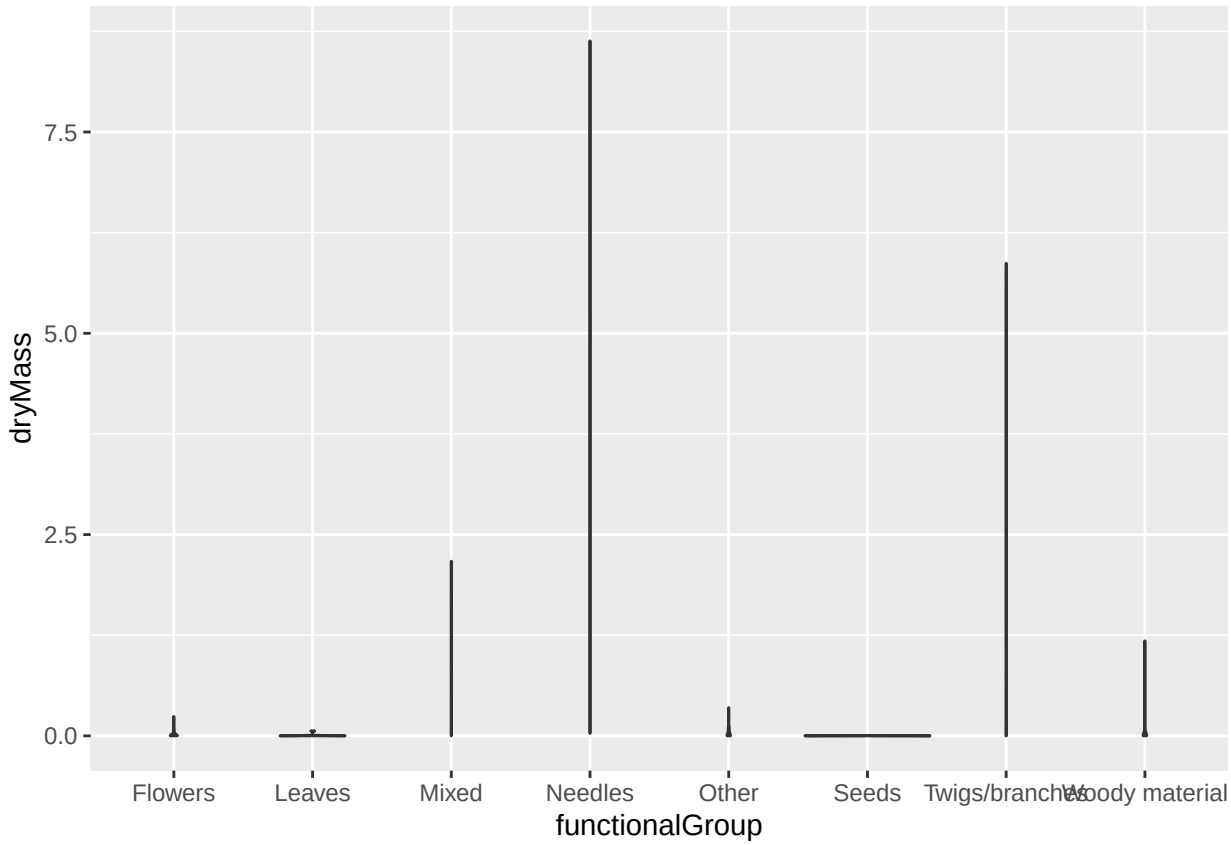


15. Using `geom_boxplot` and `geom_violin`, create a boxplot and a violin plot of dryMass by functionalGroup.

```
#boxplot
dry.mass.boxplot <- ggplot(Litter, aes(x = functionalGroup, y = dryMass)) +
  geom_boxplot()
dry.mass.boxplot
```




```
#violin plot
dry.mass.violin <- ggplot(Litter, aes(x = functionalGroup, y = dryMass)) +
  geom_violin()
dry.mass.violin
```



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

> Answer: The boxplot is a more effective visualization because it shows the summary statics and clear comparison of groups. Violin plots are useful for showing more complex parameters like distribution of the data, but in this case it is just lines.

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

> Answer: Needles have the highest biomass, while mixed litter is the second highest.