Assignment 3: Data Exploration

Halina Malinowski, Section #3

OVERVIEW

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

Directions

1. Change "Student Name, Section #" on line 3 (above) with your name and section number.
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., "FirstLast_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. **Be sure to add the `stringsAsFactors = TRUE` parameter to the function when reading in the CSV files.**

```
getwd() #Check working directory
```

```
## [1] "C:/Users/Dell Laptop/Documents/GitHub/EDA/Assignments"
```

```
library(tidyverse) #load packages
#uploading 2 datasets
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 ecotoxicologoy of neonicotinoids on insects? Feel free to do a brief internet search if you feel you need more background information.

Answer: We might be interested in the ecotoxicology of neonicotinoids on insects because of their widespread use in agriculture and various effects on both targetted "pest" species as well as on non-target insect species. It would be advantageous to know how effective the noenicotinoids are at deterring pests and preserving crops. It is also critical to understand the widespread effects of neonicotinoids on other insects that are relied on for ecosystem services such as pollinating crops. There is evidence linked to the use of neonicotinoids and the decreases in essential pollinator populations such as bees. Other endangered insects, such as monarch butterflies, may also be negatively effected by neonicotinoids. Understanding the ecotoxicology of neonicotinoids on insects can demonstrate the positive and negative effects of these methods and the long term consequences of using such methods. Additionally, this can lead to the development of other pest control methods.

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: We may be interested in studying the litter and woody debris that falls in forests to gain a better understandning of the nutrients available to the forest, decomposition processes, and microbial activity. When woody debris and litter falls to the forest floor it creates a protective layer on the forest floor and begins to decompose. The material begins to go through decomposition from various microbes and bacteria. As the material breaks down nutrients from the litter leach back into the soil and become available for uptake by the forest. Therefore through the assessment the litter and decomposition on this litter we can understand nutrient and resource availability to the forest and can begin to understand components relating to forest productivity. We can also look at litter chemistry and stable isotopes to see what nutrients were available to the plants and were already absorbed while they were growing and the litter and wood were part of the parent tree.

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: *Litter and woody debris is sampled using a paired trap design: one ground trap and one elevated trap.* The placement of litter traps may be targeted or randomized depending on the vegetation within the plots. *Sampling times vary and ground traps are only sampled once per year, while sampling of elevated traps depends on the vegetation types present where deciduous forests are sample elevated traps 1x every 2 weeks and coniferous forest 1x every 1-2 months.

Obtain basic summaries of your data (Neonics)

5. What are the dimensions of the dataset?

```
#dimensions of dataset
(dim(Neonics))
```

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

```
(length(Neonics))
```

```
## [1] 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 five most common effects include Mortality, Population, Behavior, Feeding Behavior, and Reproduction. These effects might specifically be of interest to determine how insects respond to the pesticide. These effects show insect mortality most likely in response to the pesticide as well as population, behavior, and reproduction. This would demonstrate how the insects and their populations respond to the pestiscide either through mortality, change in behavior (avoidance or altering feeding behavior), and changes in population size.

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.

```
(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 common studied species are all hymenoptera and are specifically bees or wasps. All of these insects provide vital ecosystem services which are also important for the production of crops. The various bees and parasitic wasp can act as pollinators. Additionally, the parasitic wasp has the advantage that it can kill through its parasitism several species of insects often considered pests in agriculture. Parasitic wasps lay their eggs in the bodies of other insects and as they hatch the young eat the host. This way wasps may be helping to decrease the population of agricultural pests.

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.))
```

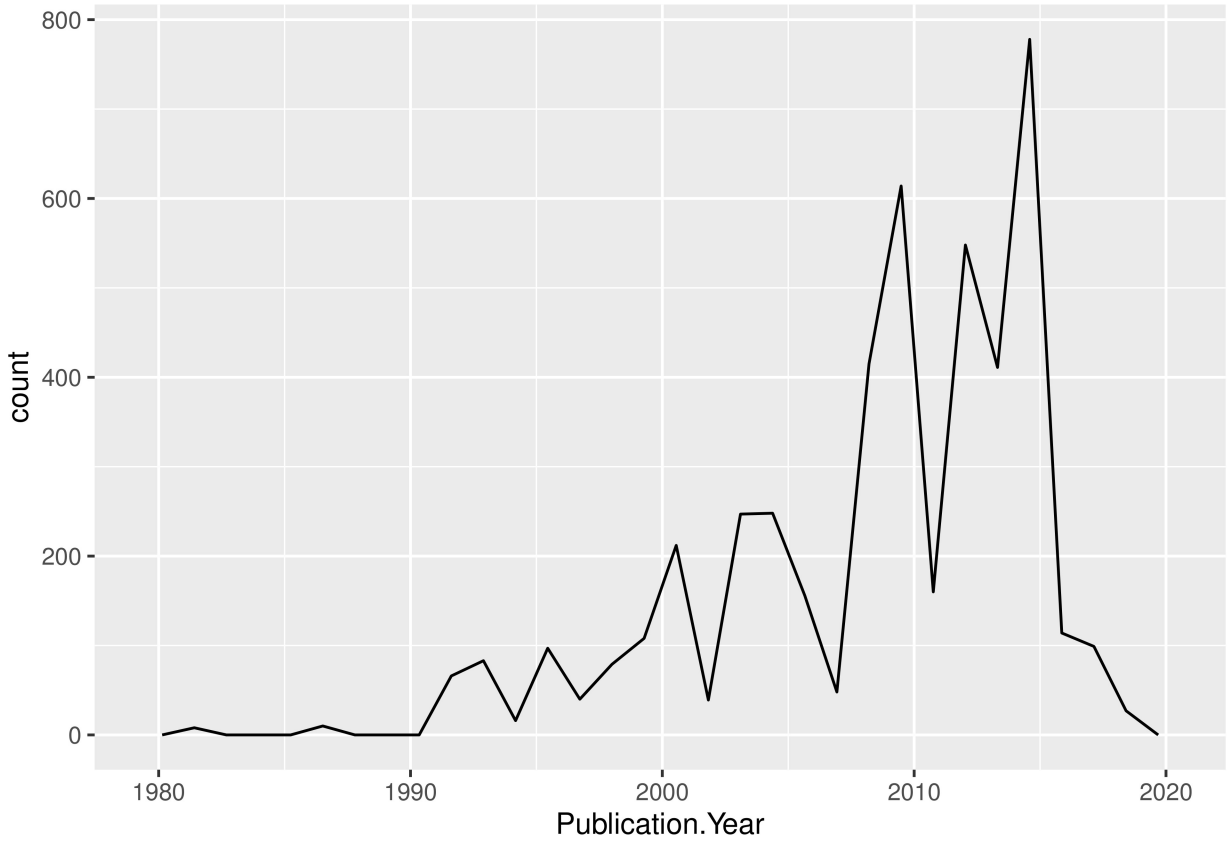
```
## [1] "factor"
```

Answer: The class of Conc.1..Author is a factor in this dataset instead of being numeric. This is because the values are not numeric rather they are characters so this has become categorical. Some examples of the input for Conc.1..Author. is "Active Ingredient", "Formulation", and "Not coded".

Explore your data graphically (Neonics)

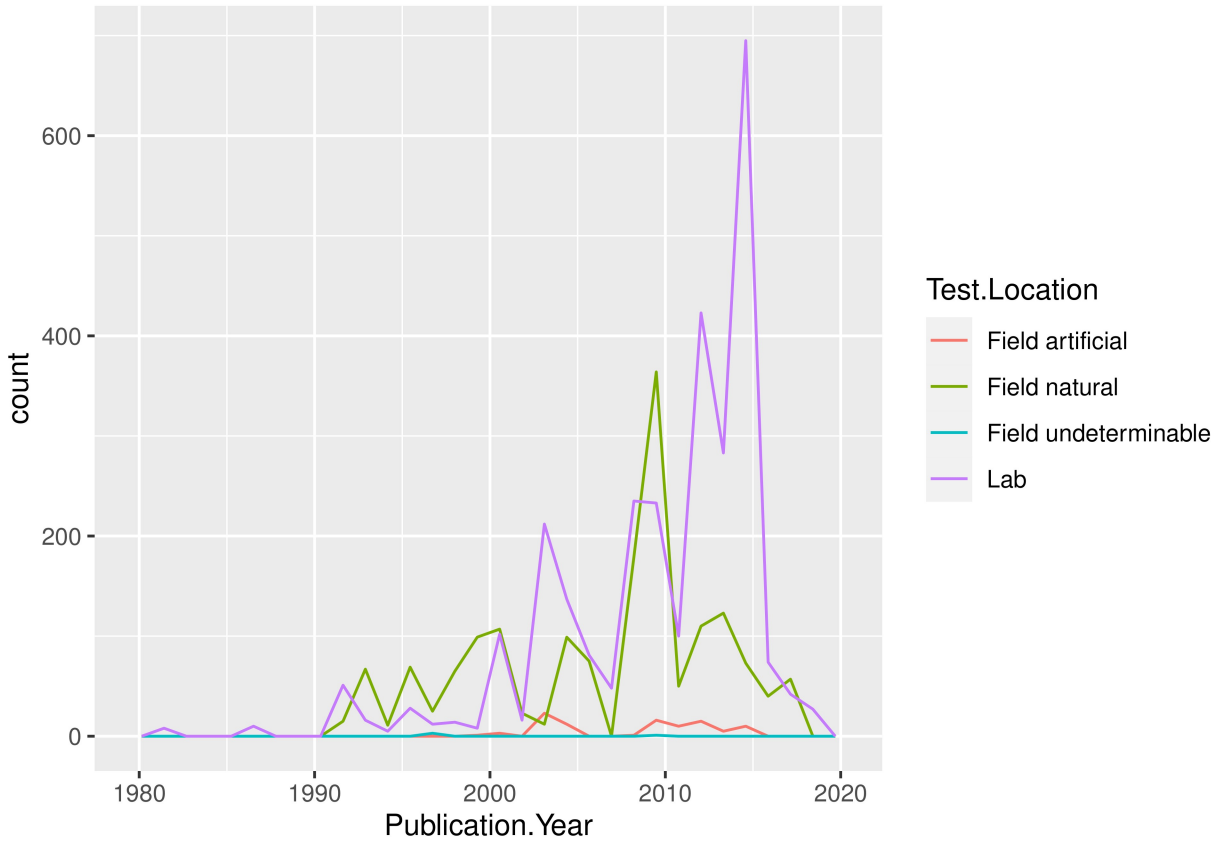
9. Using `geom_freqpoly`, generate a plot of the number of studies conducted by publication year.

```
Studies_per_year <- ggplot(Neonics)+
  geom_freqpoly(aes(x = Publication.Year), bins = 30)
Studies_per_year
```



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

```
Studies_per_year_color <- ggplot(Neonics)+
  geom_freqpoly(aes(x = Publication.Year, color = Test.Location), bins = 30)
Studies_per_year_color
```

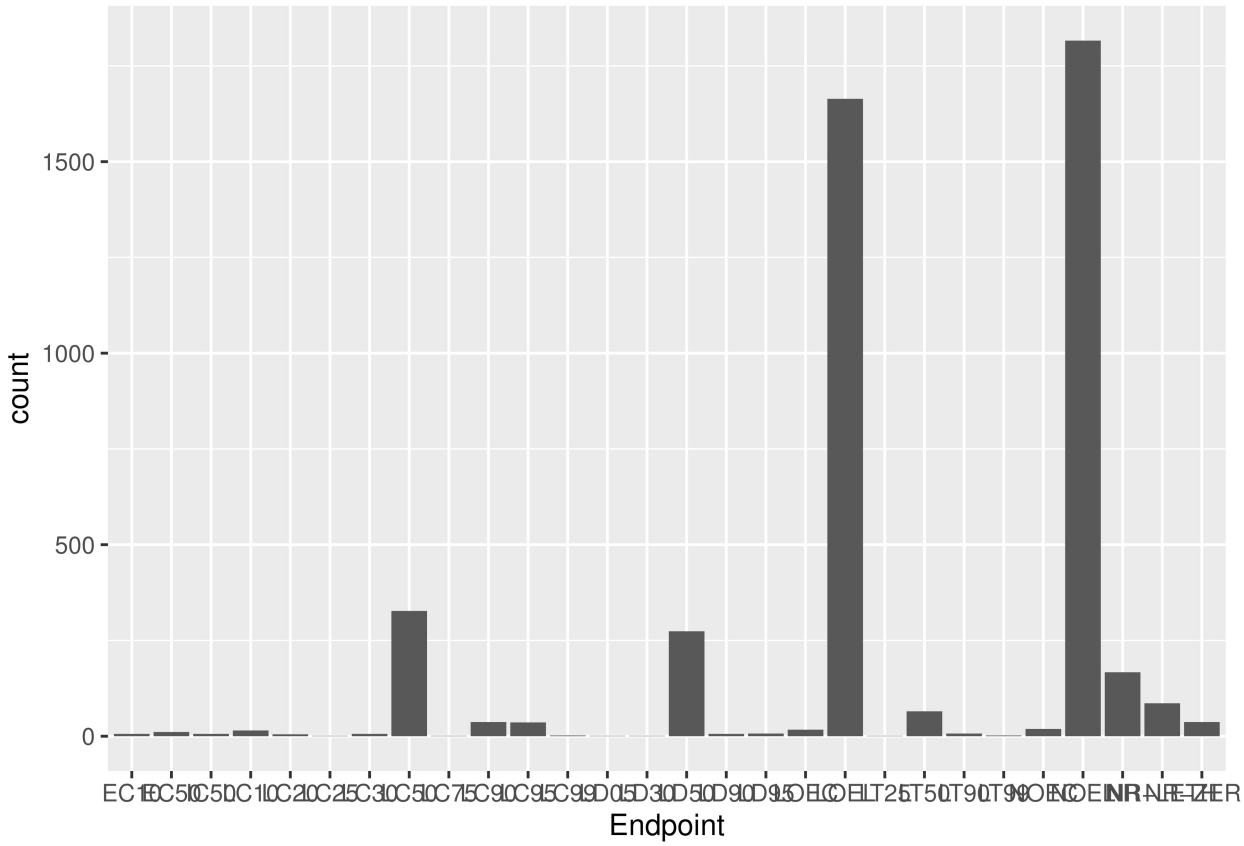


Interpret this graph. What are the most common test locations, and do they differ over time?

> Answer: The most common locations are lab and field natural. These locations do change a bit in frequency over time. Around 2010 the natural field is the most common while roughly from 2010 on the lab is the most common test location. This may be because tests done in the lab can control for other external factors to ensure that the results can be attributed to the pesticide. Additionally, laboratory work may have improved in the last decade allowing for laboratory tests to become more accurate and potentially cheaper.

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_bars <- ggplot(Neonics, aes( x = Endpoint))+
  geom_bar()
Endpoint_bars
```



Answer: The two most common end point are LOEL and NOEL. LOEL is defined as the "Lowest observable effect level" while NOEL is defined as "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.

```
(class(Litter$collectDate))
```

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

```
Litter$collectDate <- as.Date(Litter$collectDate, format = "%Y-%m-%d")
(class(Litter$collectDate))
```

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

```
(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 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
```

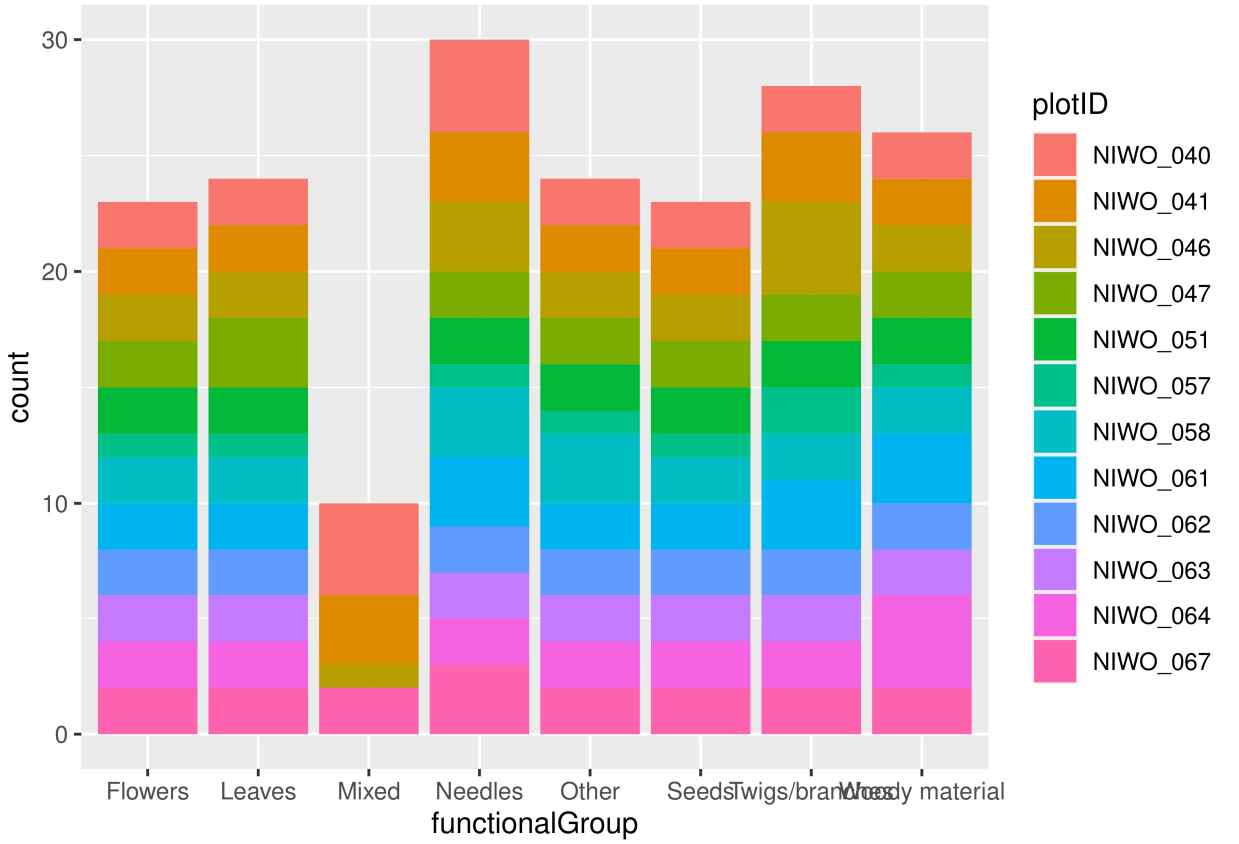
```
(summary(Litter$plotID))
```

```
## 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: Twelve plots were sampled at Niwot Ridge. The information between unique and summary is different that summary gives you the total of how many samples were taken at each plot as well.

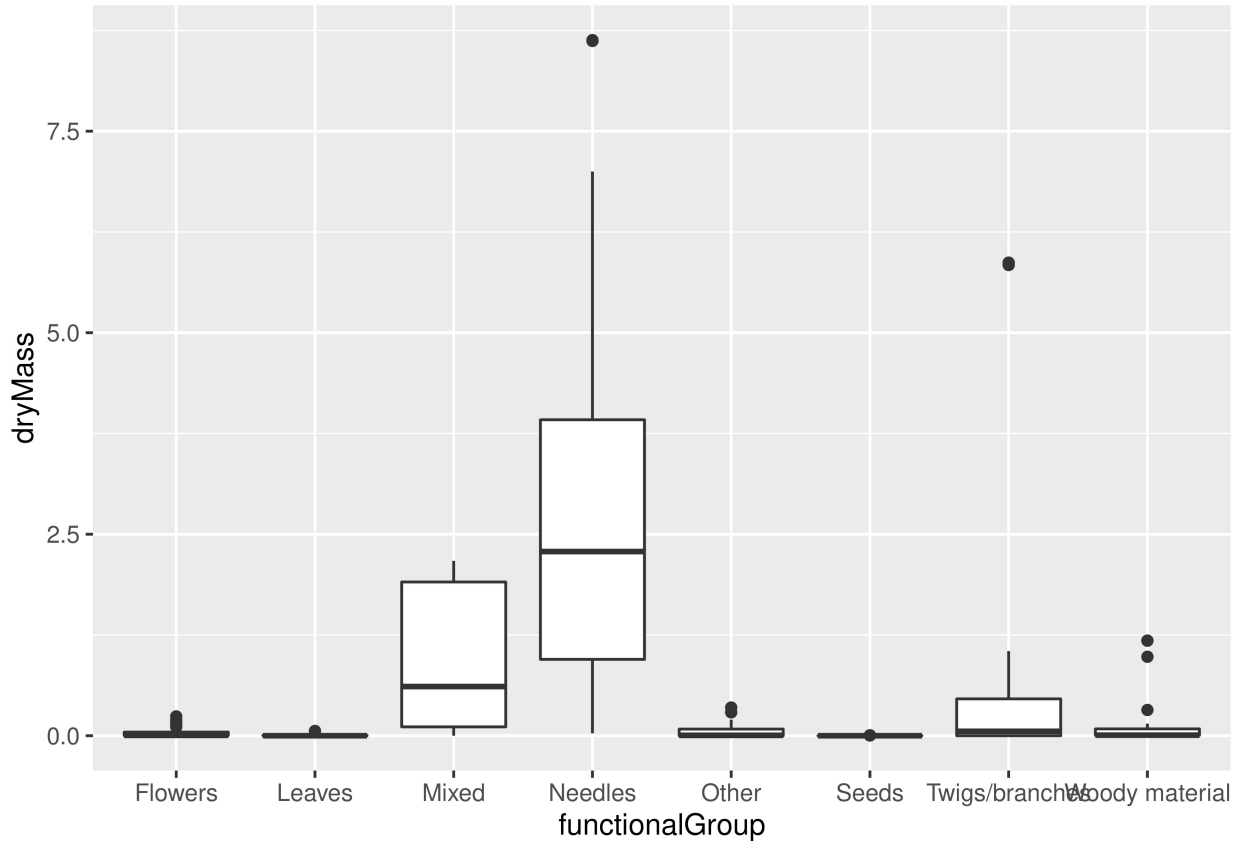
14. Create a bar graph of functionalGroup counts. This shows you what type of litter is collected at the Niwot Ridge sites. Notice that litter types are fairly equally distributed across the Niwot Ridge sites.

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



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

```
#Boxplot
Boxplot <- ggplot(Litter) +
  geom_boxplot(aes(x = functionalGroup, y = dryMass))
Boxplot
```

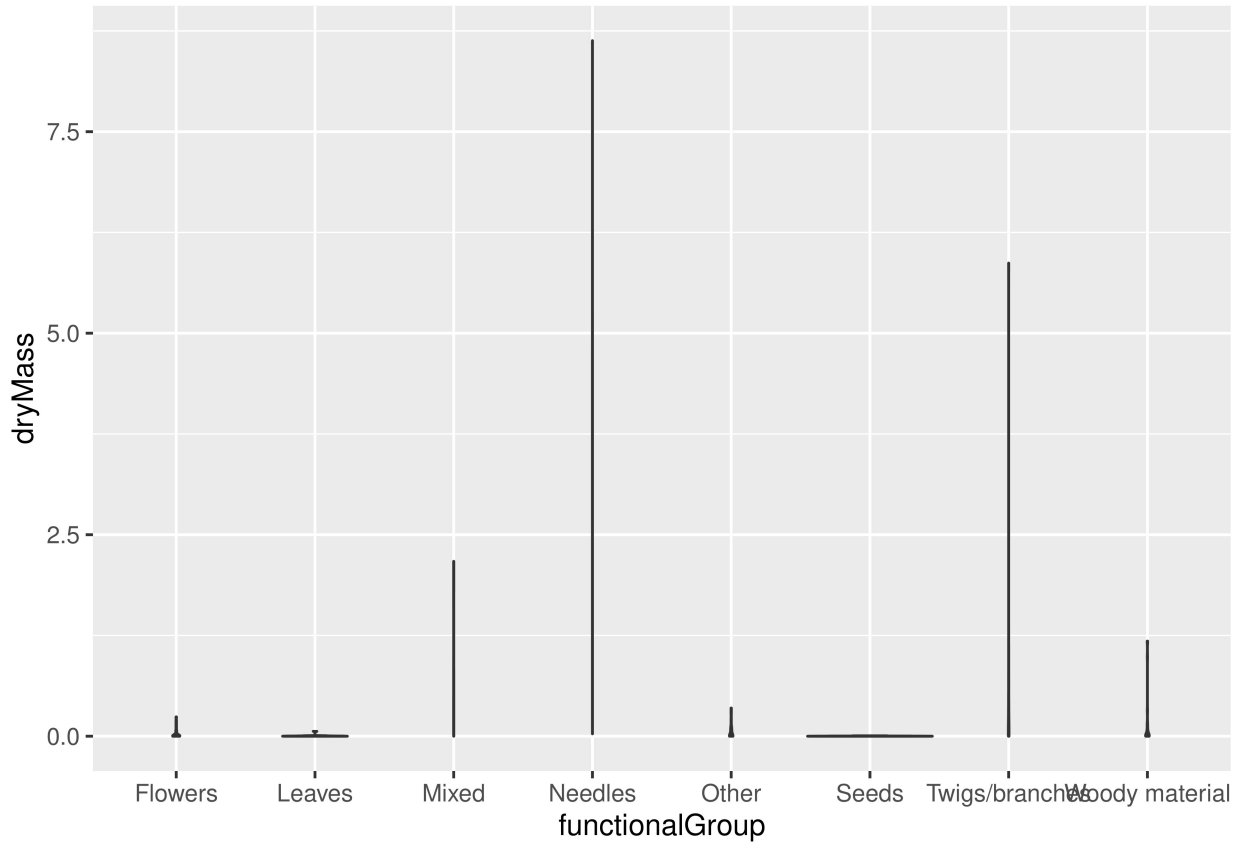



```
#Violin Plot
Violin <- ggplot(Litter) +
  geom_violin(aes(x = functionalGroup, y = dryMass),
              draw_quantiles = c(0.25, 0.5, 0.75))
Violin
```

```
## Warning in regularize.values(x, y, ties, missing(ties), na.rm = na.rm):
## collapsing to unique 'x' values

## Warning in regularize.values(x, y, ties, missing(ties), na.rm = na.rm):
## collapsing to unique 'x' values

## Warning in regularize.values(x, y, ties, missing(ties), na.rm = na.rm):
## collapsing to unique 'x' values
```



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

> Answer: The boxplot is a more effective visualization option than the violin plot because of the draw_quantiles. The draw quantiles do not fit the scale of the data and so the plot looks greatly distorted. Violin draw quantiles cannot exceed 1. Therefore, the boxplot displays the information much more clearly and to scale.

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

> Answer: Needles has the highest biomass at these sites followed by mixed and twigs/branches.