Assignment 3: Data Exploration

Megan Lundequam, 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()
```

```
## [1] "/Users/meganlundequam/Desktop/Spring 2022/Environmental Data Analytics/Git/Environmental_Data_An
```

```
library(tidyverse)
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: My initial reaction before conducting an internet search is that information about the ecotoxicology of neonicotinoids on insects could provide insight into how these insecticides bioaccumulate in organisms that ingest them. If studies show that these insecticides are turning up in larger quantities in insects that eat those insects that have come into contact with the pesticide, this could suggest that these toxins travel through food chains and could end up harming untargetted species, including humans. After an internet search, some of the most common discussions were around the effects of neonicotinoids on bees. If bees are being negatively impacted by these chemicals, this could have adverse impacts on all species that require pollination by bees and therefore disrupt an entire system as opposed to just the target insects.

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: Forest litter and woody debris can be an indicator of overall forest health. Litter and debris can provide insights into ecosystem functions, wildlife composition through potential wildlife habitat, biodiversity of dead-wood dependent organisms, tree regeneration, wildfire risk, nutrient cycles, and carbon stocks and cycling. And over time, this data can reveal how the above factors have changed in relation to different events and paint a picture of forest changes over time. The NEON userguide specifically states annual Aboveground Net Primary Productivity and aboveground biomass as information that can be derived from litterfall and find woody debris data. Additionally, they can provide essential data for understanding vegetative carbon fluxes over time.

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: * Sampling occurs only in tower plots whos locations are selected randomly within the 90% flux footprint of the primary and secondary airsheds. * 20 40m x 40m plots are placed to collect litter samples in forested tower airsheds and in sites with low-saturated vegetation over the tower airsheds, litter sampling is to take place in 4 40m x 40m plots and 26 20m x 20m plots. * Plots must be sepatated by a distance 150% of one edge of the plot but trap placement within the plot can be targeted or randomized, depending on the vegetation. * Ground traps are sampled once per year while elevated trap sampling varies.

Obtain basic summaries of your data (Neonics)

5. What are the dimensions of the dataset?

```
dim(Neonics)
```

```
## [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?

```
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 Popultion and Mortality, with Behavior and Feeding behavior falling as far 3rd and 4ths. Population and Mortality are arguably the most important effects to understand because population could tell you how persistent the chemical is and mortality could reveal whether or not the chemical is lethal to the insects that come into contact with it.

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 six most commonly studied species in the dataset are Honey Bees, Parasitic Wasps, Buff Tailed Bumblebees, Carniolan Honey Bees, Bumble Bees, and Italian Honeybees. Something that all of these insects have in common (apart from parasitic wasps) is the fact that they are pollinators. This quality makes them of interest over other insects because pollinators are vital to agricultural production and using a pesticide that could be lethal to these species would be very contradictory to the purpose of using a pesticide on agricultural fields in the first place.

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.)
```

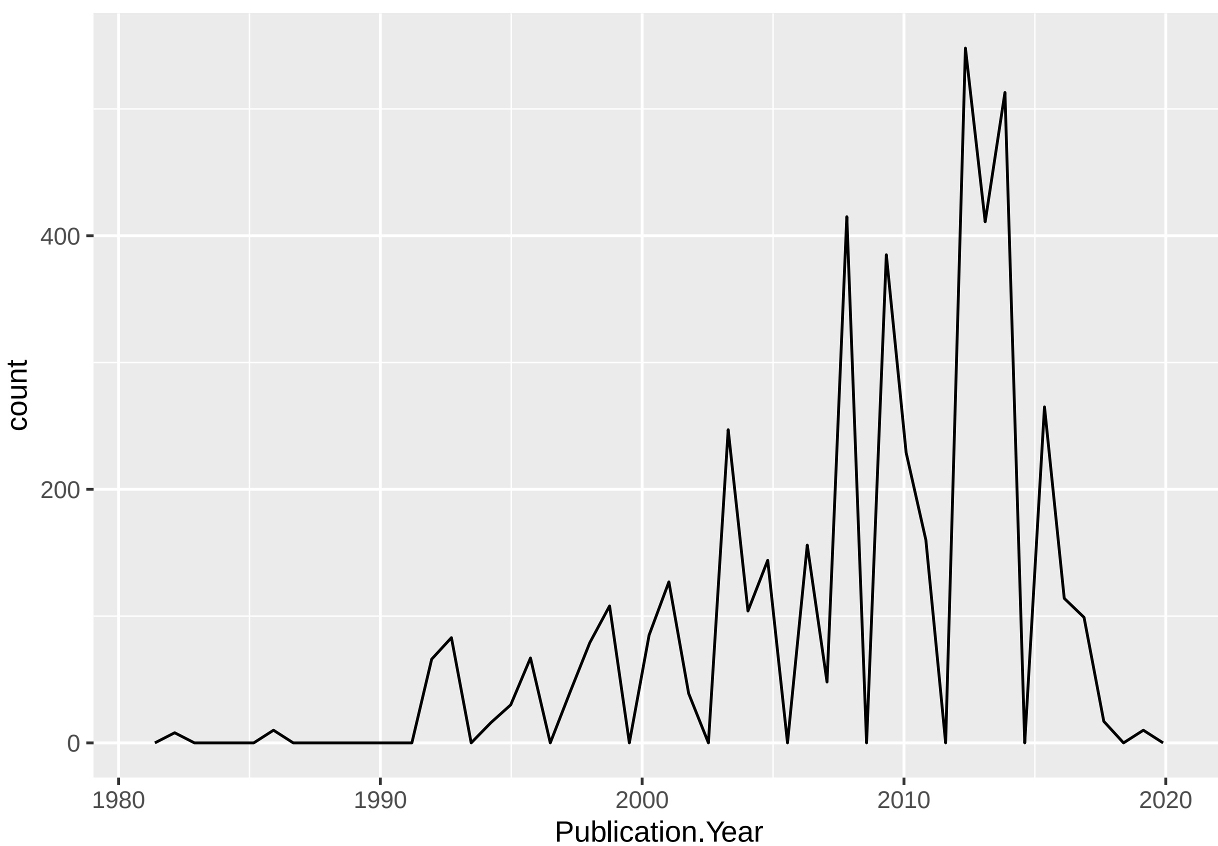
```
## [1] "factor"
```

Answer: The class of Conc.1..Author. is “factor”. This could be because the values for this variable sometimes have a “/” after the number, rendering them as categories as opposed to numeric values when R reads them.

Explore your data graphically (Neonics)

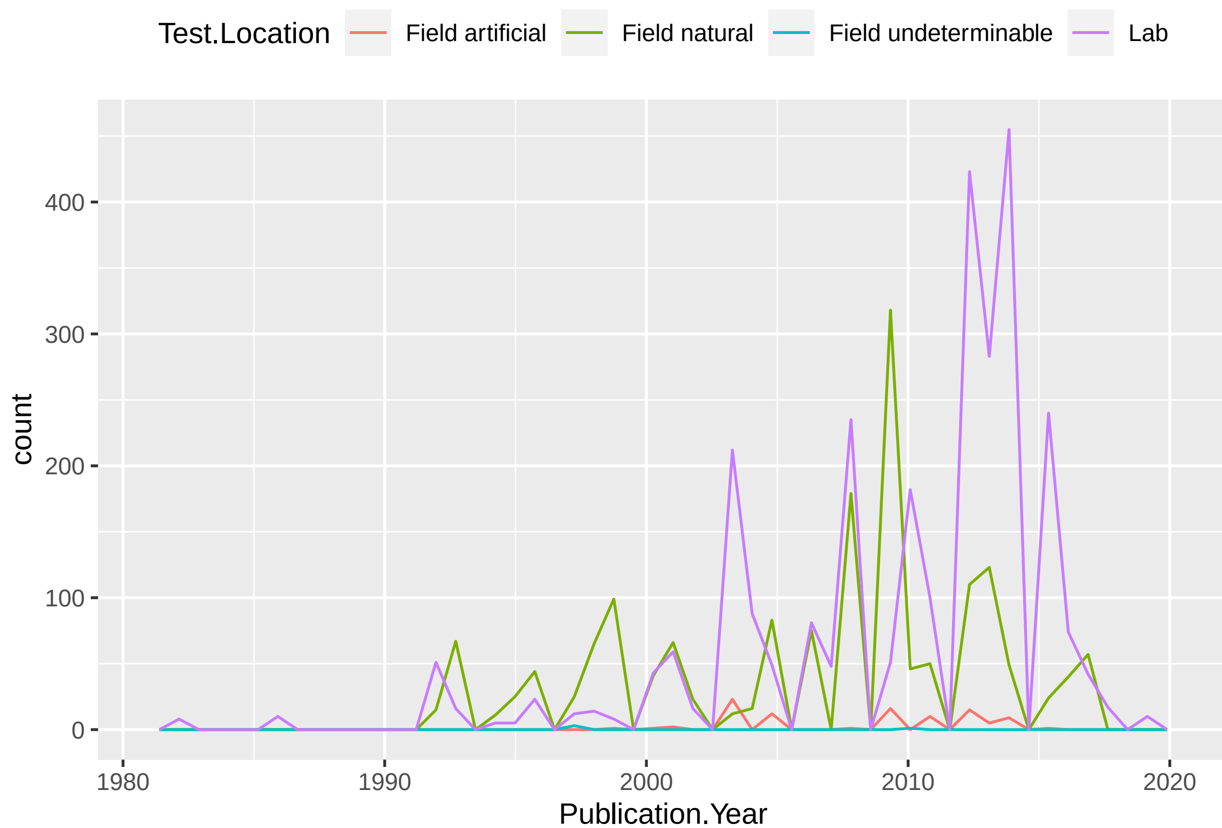
9. Using `geom_freqpoly`, generate a plot of the number of studies conducted by publication year.

```
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
ggplot(Neonics) +
  geom_freqpoly(aes(x = Publication.Year, color = Test.Location), bins = 50) +
  theme(legend.position = "top")
```

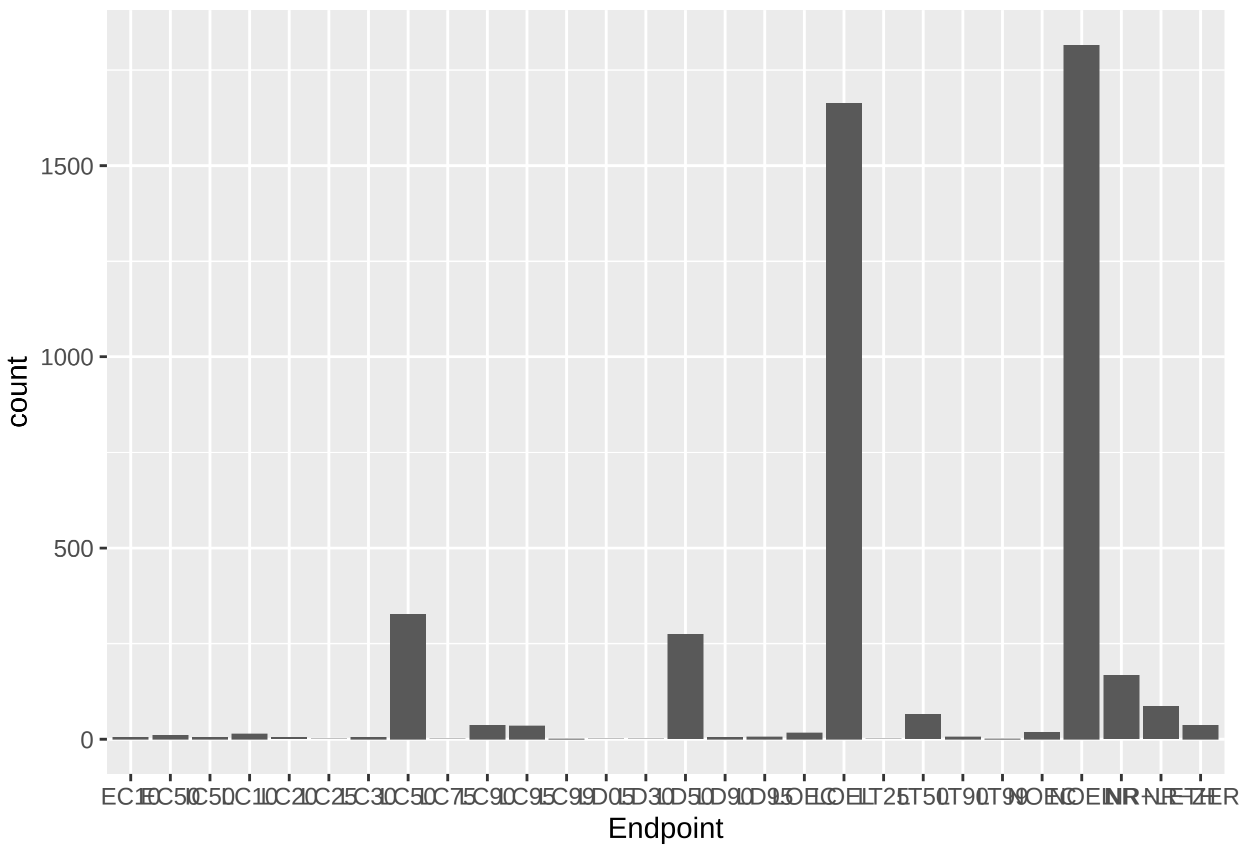


Interpret this graph. What are the most common test locations, and do they differ over time?

> Answer: Yes! This graph shows that lab tests and natural tests in the field are the most common test locations. This has changed with time, however, as around 2011, the number of lab tests dramatically increased and far surpassed the number of field tests and has consistently stayed the most common location despite being outweighed by field tests just a couple of years prior.

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.

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



Answer: LOEL and NOEL are the two most common endpoints according to this dataset. LOEL is the Lowest-observable-effect-level, i.e. the lowest dose (concentration) producing effects that were significantly different (as reported by authors) from responses of controls (LOEAL/LOEC). NOEL endpoint represents the No-observable-effect-level, i.e. the highest dose (concentration) producing effects not significantly different from responses of controls according to author's reported statistical test (NOEAL/NOEC).

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"
```

```
class(Litter)
```

```
## [1] "data.frame"
```

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

```
## [1] NA
```

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`?

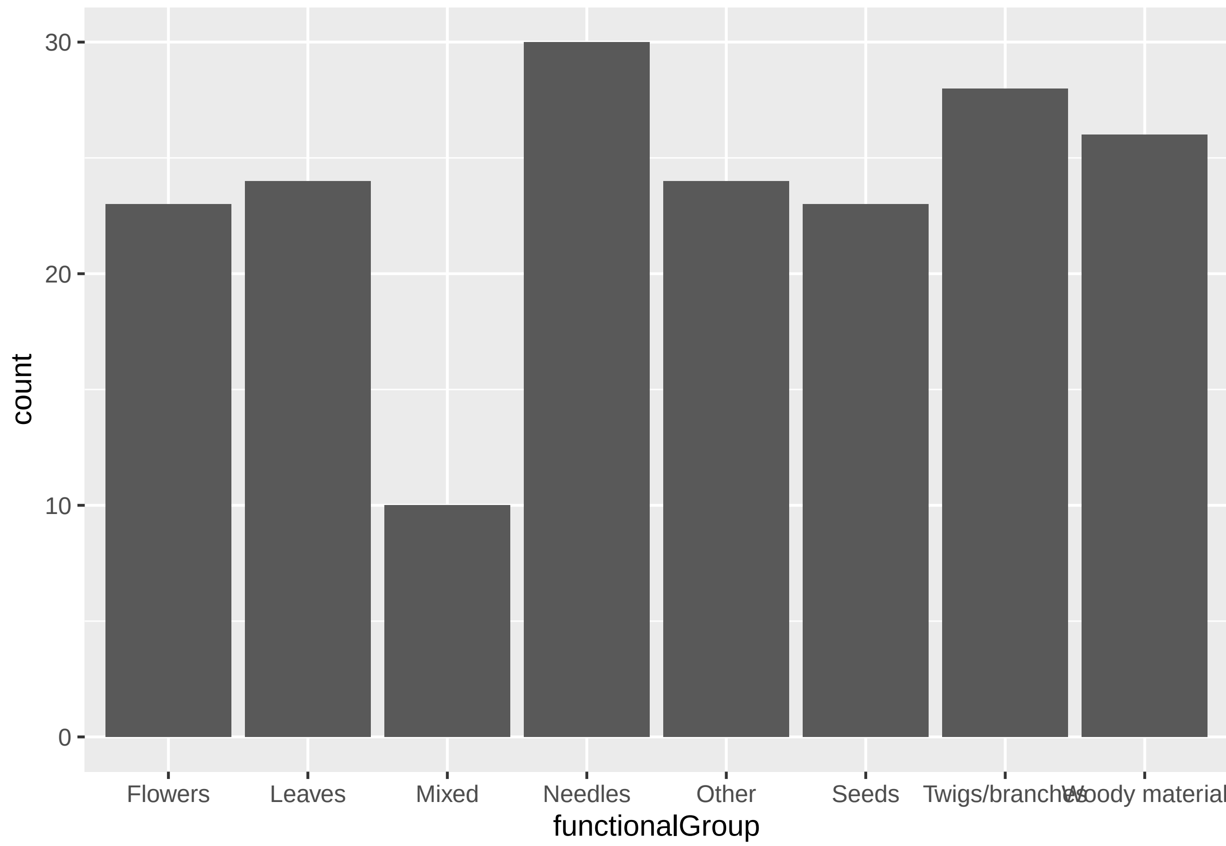
```
unique(Litter$plotID)
```

```
##  [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
```

Answer: 12 plots were sampled. This information is different than what the function summary tells us because summary reports the number of data entries for the plot variable whereas unique finds those values that are unique and does not report duplicates. This allows us to see how many different plots there are rather than how many data entries there are.

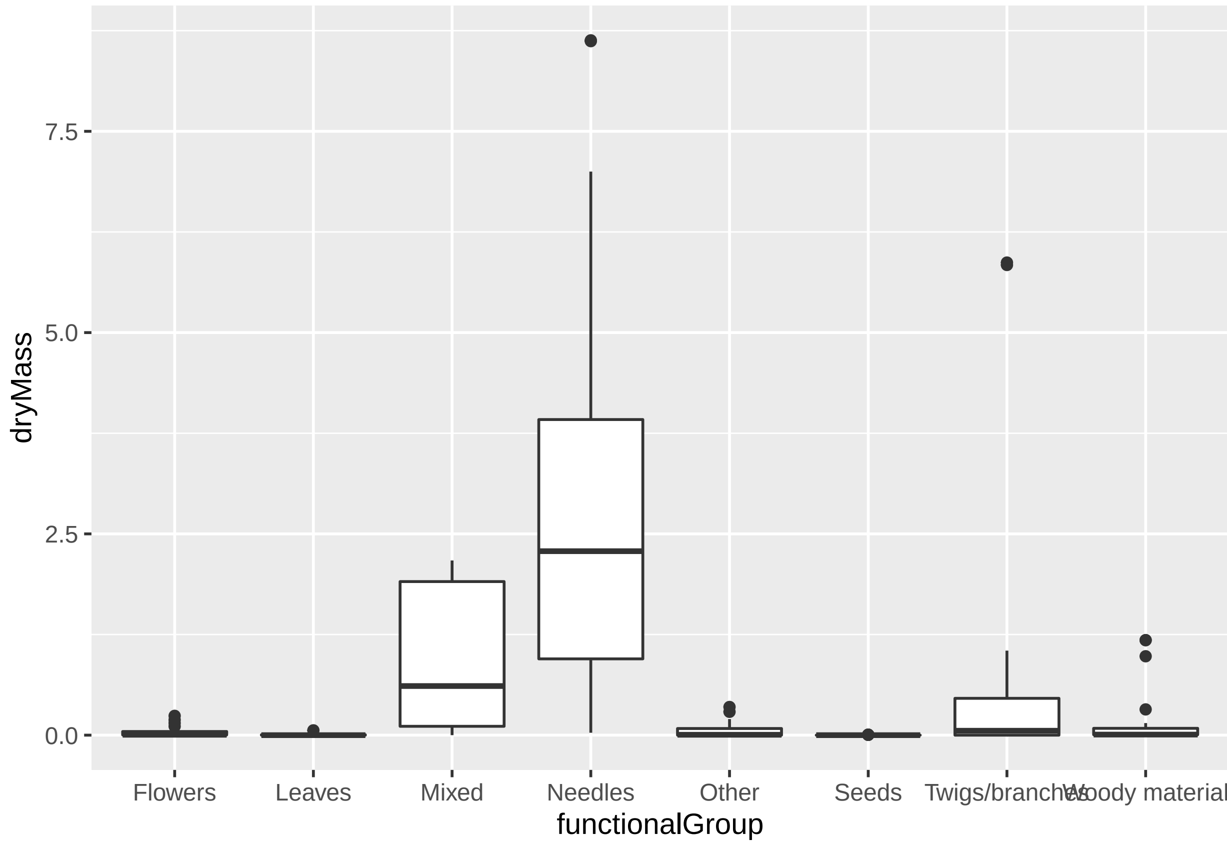
14. Create a bar graph of functionalGroup counts. This shows you what type of litter is collected at the Niwot Ridge sites. Notice that litter types are fairly equally distributed across the Niwot Ridge sites.

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

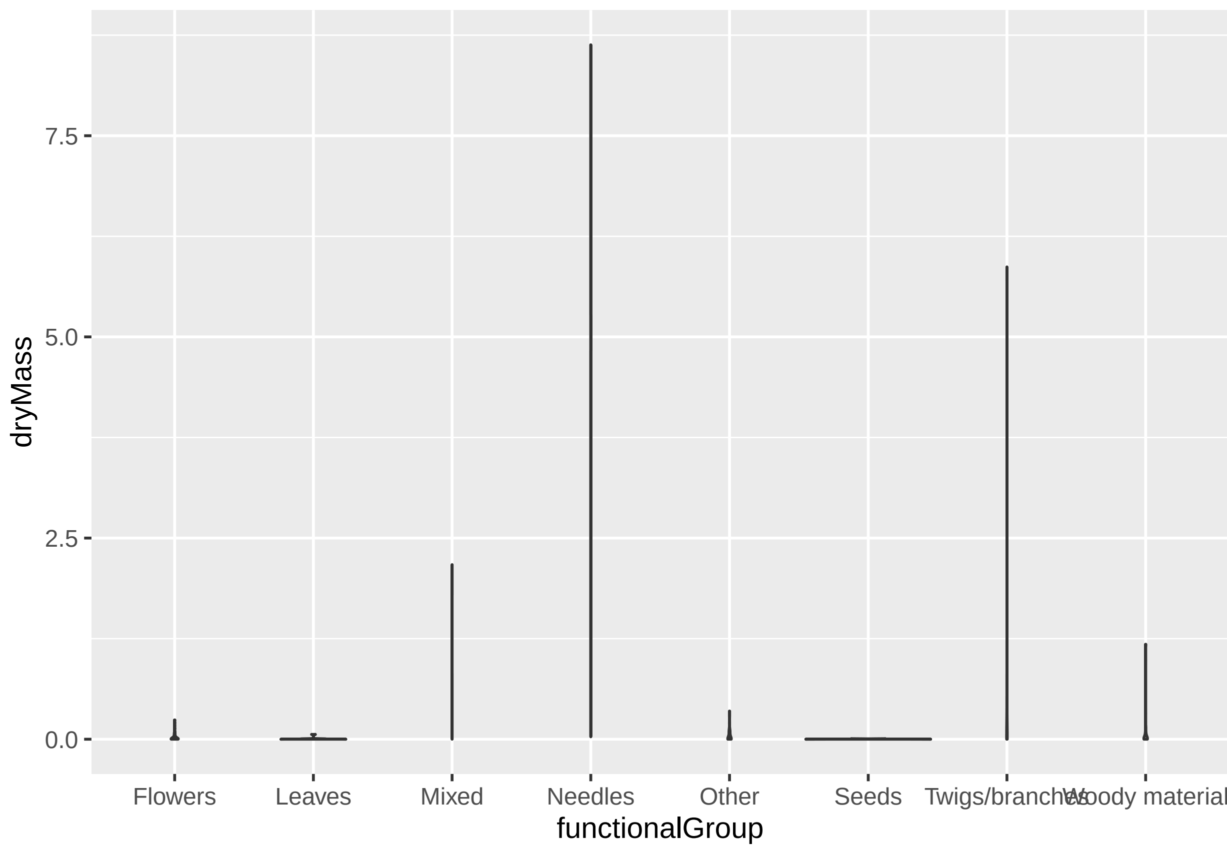


15. Using `geom_boxplot` and `geom_violin`, create a boxplot and a violin plot of dryMass by functionalGroup.

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




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



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 a violin plot shows the range of values and distribution within that range with the width representing how many data points are distributed within that range, but in this case, the data points within mass ranges are so few and concentrated in low numbers that no visual depiction can really be seen.

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

> Answer: Needles and mixed!