Assignment 3: Data Exploration

Camila Zarate Ospina

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.

```
getwd()
```

```
## [1] "/Users/camilazarate/OneDrive - Duke University/2 Second semester/Data analytics/Environmental_Da
```

```
setwd("/Users/camilazarate/OneDrive - Duke University/2 Second semester/Data analytics/Environmental_Dat
```

```
library(tidyverse)
```

```
## -- Attaching packages ---------------------------------------------------------------------- tidyvers
## v ggplot2 3.3.2     v purrr   0.3.4
## v tibble  3.0.3     v dplyr   1.0.2
## v tidyr   1.1.2     v stringr 1.4.0
## v readr   1.3.1     v forcats 0.5.0
## -- Conflicts ------------------------------------------------------------------------- tidyverse_con
## x dplyr::filter() masks stats::filter()
## x dplyr::lag()    masks stats::lag()
```

```
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")
```

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: Neonicotinoids insecticide is linked to adverse effects in nature, including negative impacts on bee's colonies and other polinators. Some bird populations have decrease as a consequence of the use of Neonicotinoid, which erradicated all the insects they eat. As a consequence, this insecticide has been restricted in european countries and a few states in the US. Ecotoxicology is the science that studies the effects of toxic chemicals - like Neonicotinoids - on biological organisms at the organisms, community, ecosystem and biosphere level. One might be interested in the ecotoxicologoy of neonicotinoids on insects because in the long run, negative effects on insects can translate into reduced polinators, directly affecting our food systems.

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 plays an important role in carbon budgets and nutrient cycles and provides habitat for terrestrial and aquatic organisms. Likewise, litter plays a role by transferring nutrients from the aboveground biomass to the soil and in general improving soil quality. Together, they can give us great information regarding the health of the forest and potential disruptions of nutrient cycles.

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: * Random sampling. Litter and woody debris are sampled at terrestrial NEON sites. The sampling process occurs in tower plots whose location is randomly selected. Depending on the site of the forest (low-statured vegetation, forested, low-statue) tower plots are placed at specific distance from each other. * Depending on the vegetation (% aerial cover of woody vegetation and height), trap placement within plots may be either targeted or randomized. * Sampling frequency also depends on the vegetation: Ground traps - once per year, elevated traps in deciduous forest - once every two weeks, elevated traps at evergreen sites - once every 1-2 months.

Obtain basic summaries of your data (Neonics)

5. What are the dimensions of the dataset?

```
dim(Neonics) # row column
```

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

```
dim(Litter)
```

```
## [1] 188  19
```

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?

```
# Option 1: Create and object and change the Effect column from string
# to factor so the function summary works.
# Effect <- as.factor(Neonics$Effect)
# Effect
```

```
# summary(Effect)

# Option 2: Add the clause stringsAsFactors = TRUE when reading the csv and use 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: Effects are defined as a response that directly results from the action of a chemical stressor. Mortality and population are the two most common effects. This is of interest because high mortality rates and lower population of insects might affect the food chain that depends on the insects, ultimately affecting other insects, mammals and birds.

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: Honey Bee (667), Parasitic Wasp (285), Buff Tailed Bumblebee (183), Carniolan Honey Bee (152), Bumble Bee (140), Italian Honeybee (113). They all have in common that they're all bees, which might be of special interest since bees are one of the most important polinators, thereby playing an esential role in food systems.

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.)
```

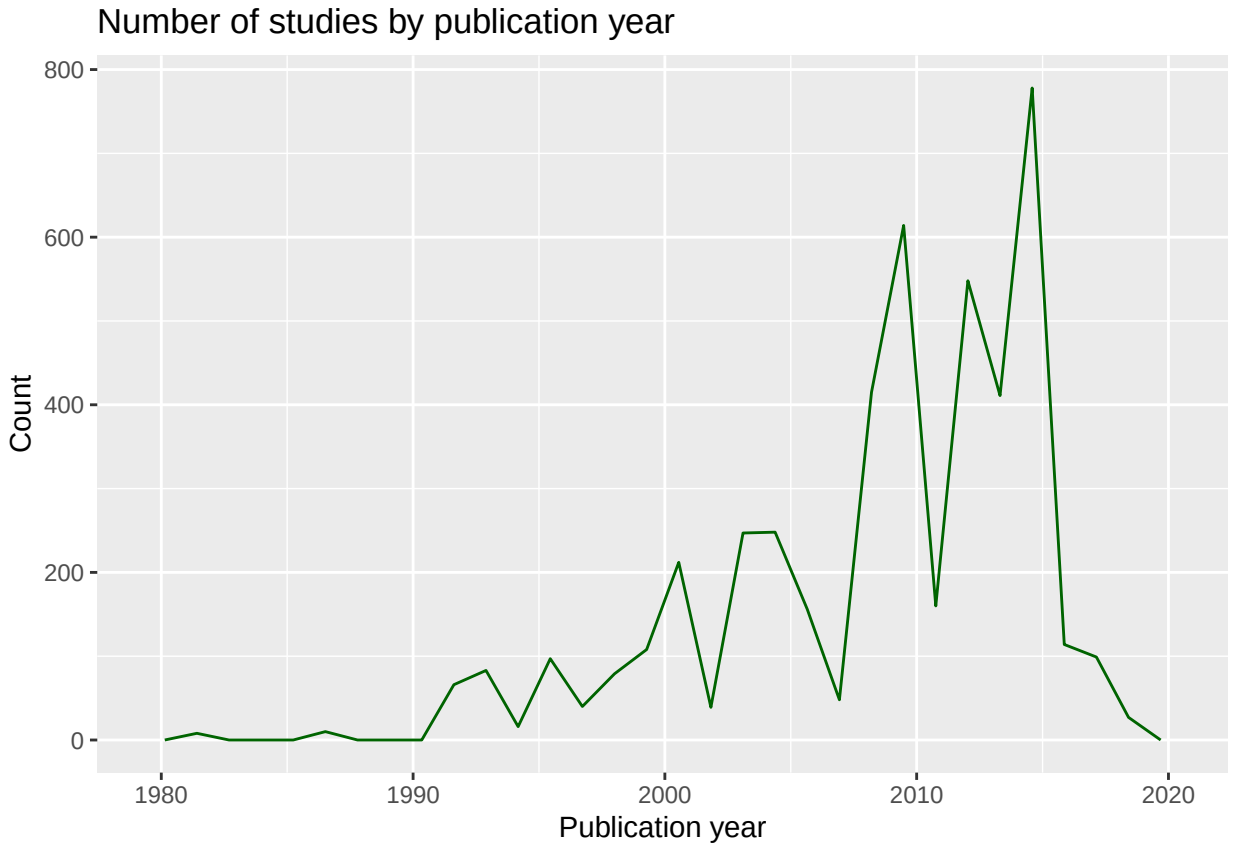
```
## [1] "factor"
```

Answer: Some rows have symbols like / or ~.

Explore your data graphically (Neonics)

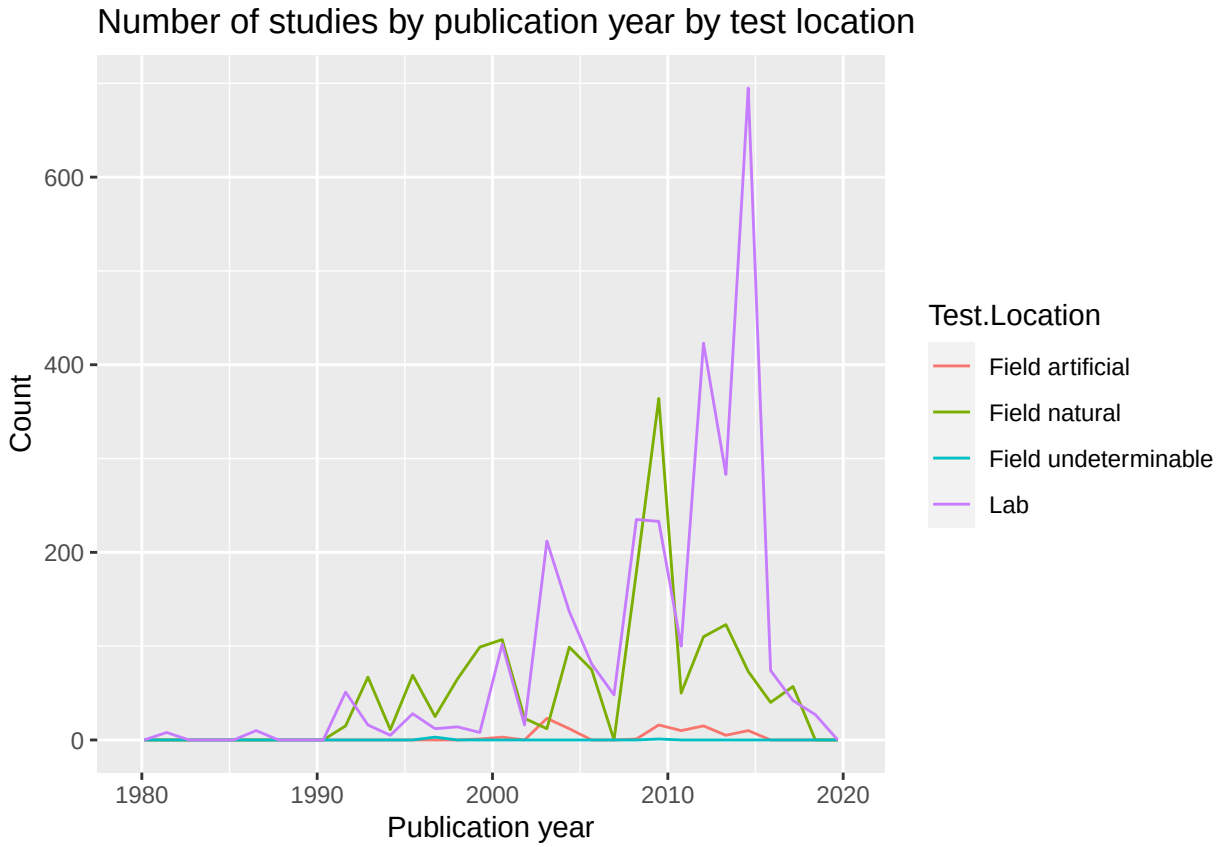
9. Using `geom_freqpoly`, generate a plot of the number of studies conducted by publication year.

```
ggplot(Neonics) +
  geom_freqpoly(aes(x = Publication.Year), bins = 30, color = "darkgreen") +
  labs(title = "Number of studies by publication year", x = "Publication year", y = "Count")
```



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

```
ggplot(Neonics) +
  geom_freqpoly(aes(x = Publication.Year, color = Test.Location), bins = 30) +
  labs(title = "Number of studies by publication year by test location",
       x = "Publication year", y = "Count")
```

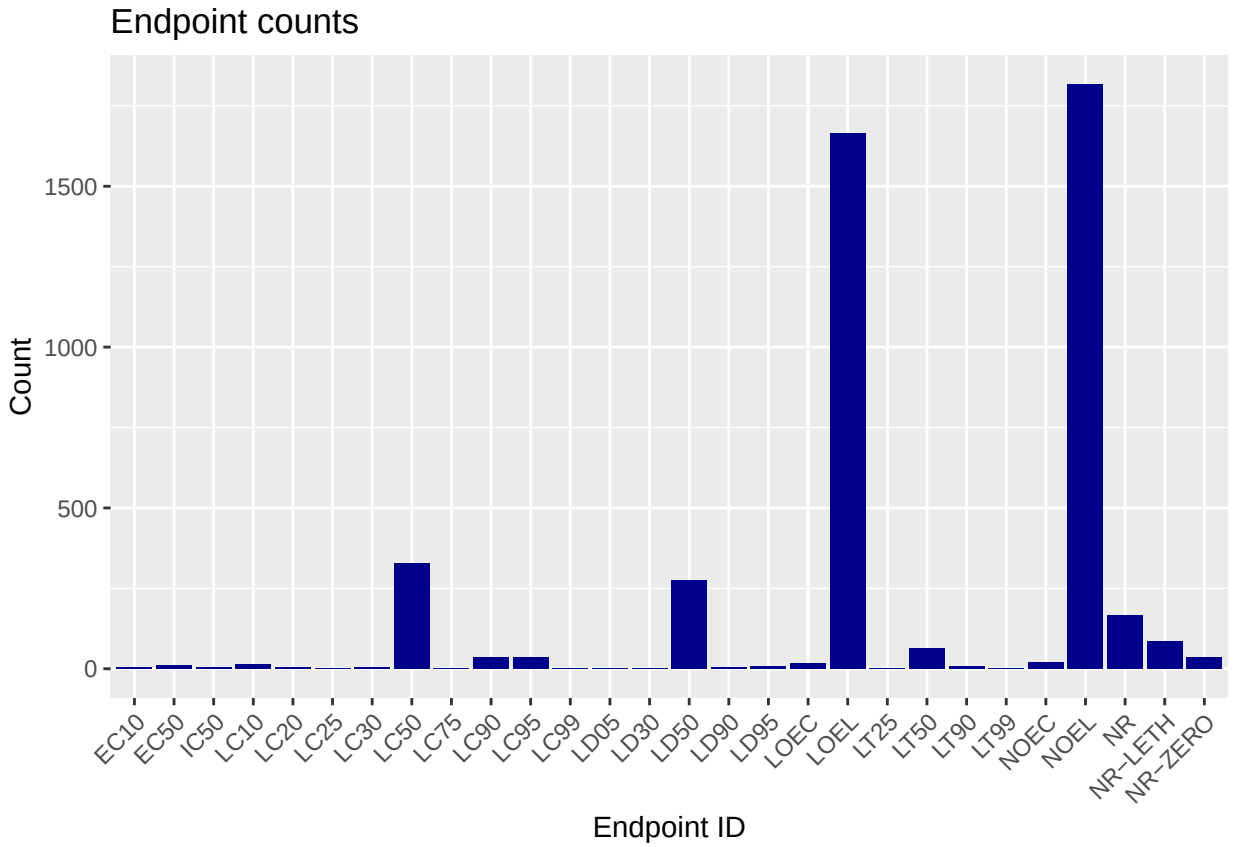


Interpret this graph. What are the most common test locations, and do they differ over time?

Answer: The most common test locations are Lab and Field natural. Overall, the location Field natural has had a steady number of studies conducted around 100 studies, except for the year 2010 where it had a peack of around 370 studies. On the contrary, Lab has seen a progressive increase of number of studies over the years, reaching a peak of 650 studies in 2015.

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.

```
ggplot(Neonics, aes(x = Endpoint)) +
  geom_bar(fill = "darkblue") +
  theme(axis.text.x = element_text(angle = 45, hjust = 1 )) +
  labs(title= "Endpoint counts", x = "Endpoint ID", y = "Count")
```



Answer: The most common two endpoits are NOEL (No-observable-effect-level) and LOEL (Lowest-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] "character"
```

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

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

```
August2018Samples <- unique(Litter$collectDate)
August2018Samples
```

```
## [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`?

```
PlotsSampled <- unique(Litter$plotID)
PlotsSampled
```

```
##  [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"
```

```
length(PlotsSampled)
```

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

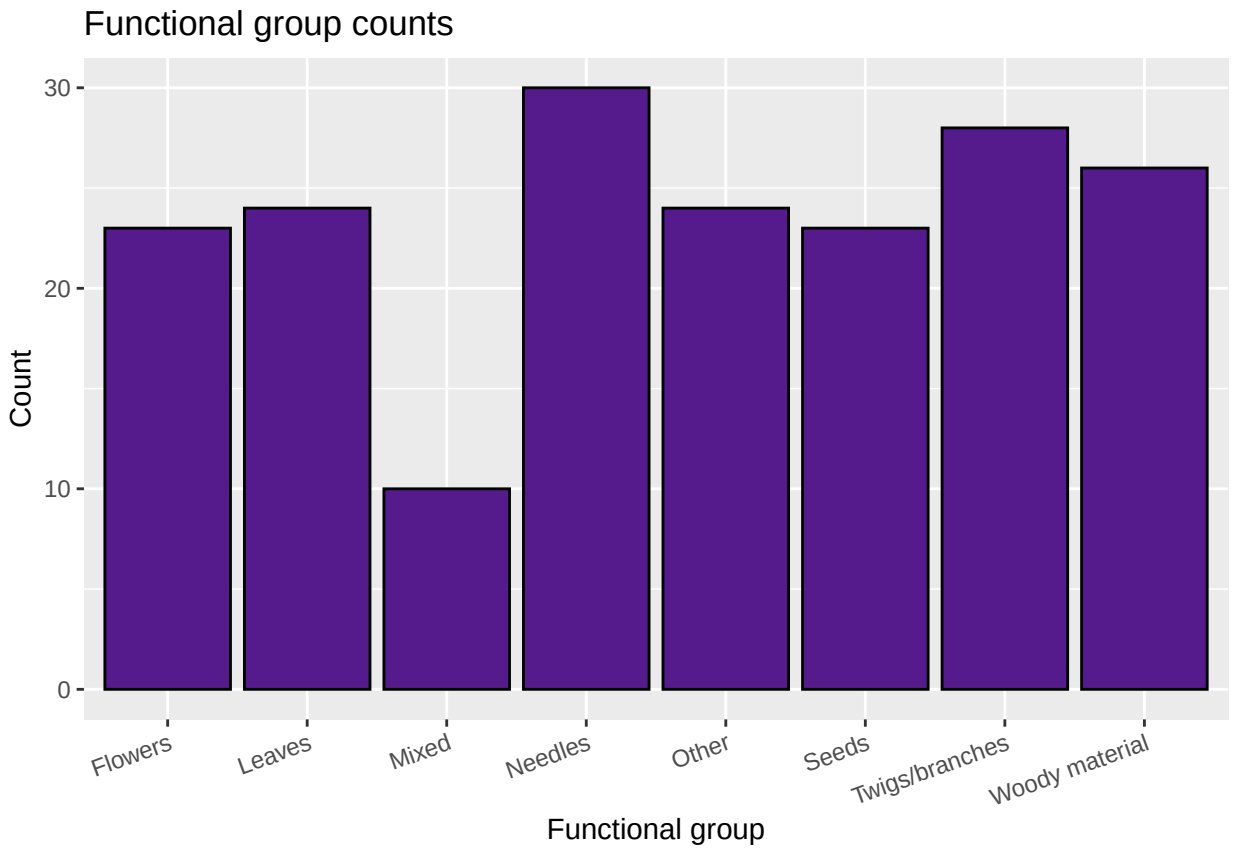
```
Plots_Sampled <- as.factor(Litter$plotID)
summary(PlotsSampled)
```

```
##    Length     Class      Mode
##        12 character character
```

Answer: The information obtained from Unique are all of the different objects/names/categories in a column. If the objects are repeated, unique won't show the repetitions. Summary gives information on object's characteristics like lenght and class.

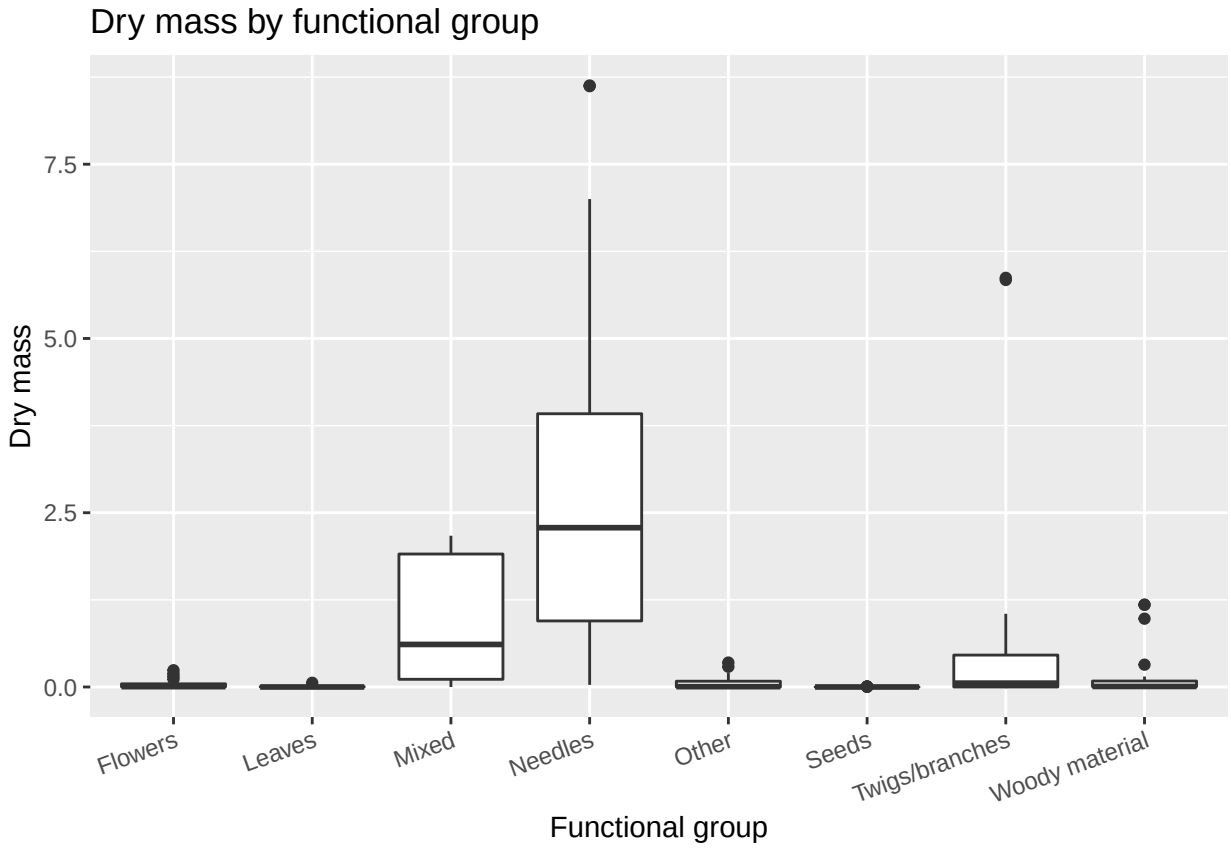
14. Create a bar graph of functionalGroup counts. This shows you what type of litter is collected at the Niwot Ridge sites. Notice that litter types are fairly equally distributed across the Niwot Ridge sites.

```
ggplot(Litter, aes(x = functionalGroup)) +
  geom_bar(fill = "purple4", color = "black") +
  theme(axis.text.x = element_text(angle = 20, hjust = 1 )) +
  labs(title="Functional group counts", x="Functional group", y="Count")
```



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)) +
  theme(axis.text.x = element_text(angle = 20, hjust = 1 )) +
  labs(title="Dry mass by functional group", x="Functional group", y="Dry mass")
```

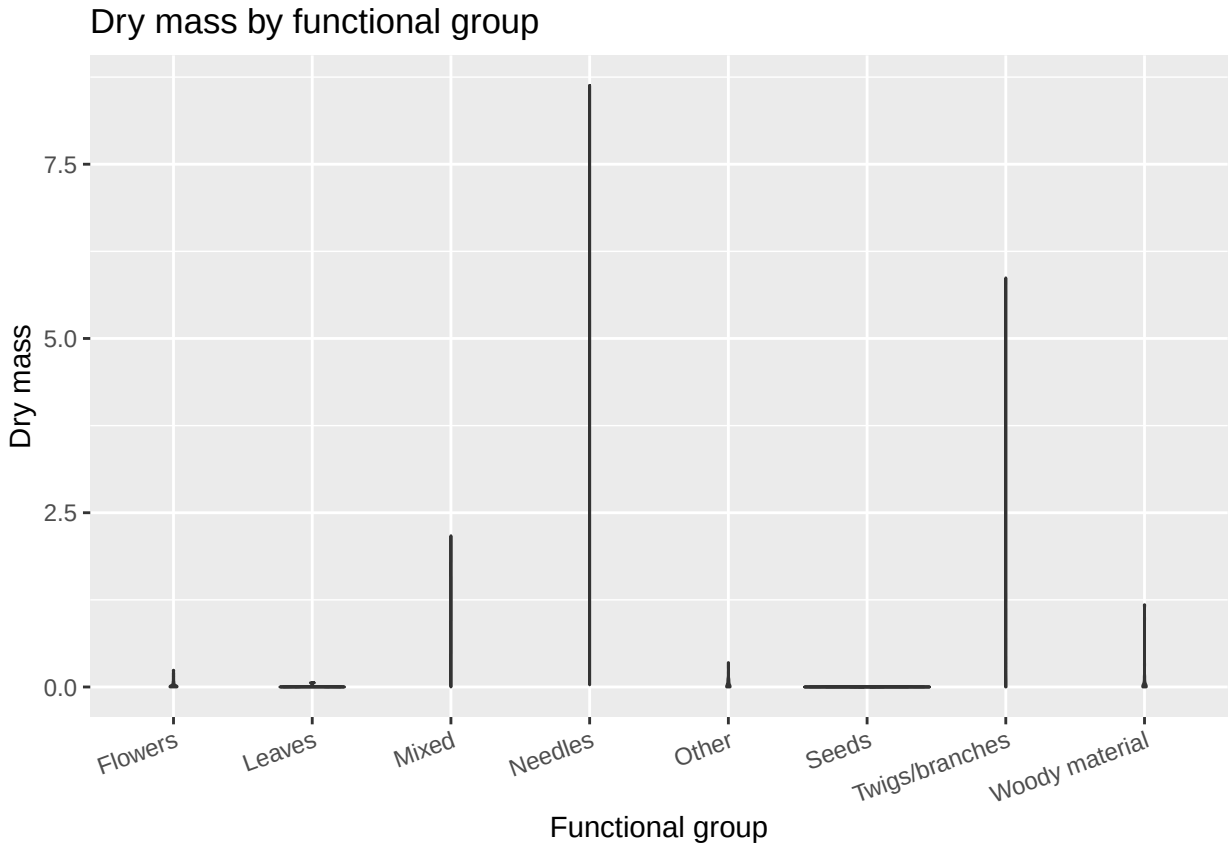



```
ggplot(Litter) +
  geom_violin(aes(x = functionalGroup, y = dryMass),
              draw_quantiles = c(0.25, 0.5, 0.75)) +
   theme(axis.text.x = element_text(angle = 20, hjust = 1 )) +
   labs(title="Dry mass by functional group", x="Functional group", y="Dry mass")
```

```
## 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 shows summary statistics of the data, and the violin plot shows the distribution of the data. In this case, the violin plot doesn't show a good visualization because the data is extremely skewed.

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

> Answer: Needles, followed by twings/branches.