Assignment 3: Data Exploration

Yixin Fang

Spring 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.

```
library(tidyverse)
library(lubridate)
getwd()
```

```
## [1] "D:/DKU/2023_Spring/ENV872/EDA-Spring2023"
```

```
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) # Import datasets
```

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: Studies have shown that neonicotinoids may have unexpected effects on non-target species. For instance, neonicotinoids may transmit from plants to pollinators such as bees, which have significant value for food production. Neonicotinoids can also be toxic for other beneficial 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: Litter and woody debris is an important part of forest and stream ecosystems. It participates in the nutrient and carbon cycles and can serve as habitats for different species. It can also influence water flows and sediment transport.

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. Spatially, sampling sites are at terrestrial NEON sites with woody vegetation higher than 2 meters. 20m * 20m and 40m * 40m plots are created for sampling with 1 to 4 trap pairs for each plot. The plots should be away from roads, buildings, and other non-NEON infrastructures. 2. The deciduous forest sites are sampled every 2 weeks while evergreen sites are sampled every 1 to 2 months. 3. Mass data are collected for different functional groups: leaves, needles, twigs/branches, woody material, seeds, flowers and other non-woody reproductive structures, other, and mixed.

Obtain basic summaries of your data (Neonics)

5. What are the dimensions of the dataset?

```
dim(Neonics)
```

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

```
dim(Litter)
```

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

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) %>%
  sort(decreasing = TRUE) # Sort results by decreasing
```

```
##       Population        Mortality         Behavior Feeding behavior
##             1803             1493              360              255
##     Reproduction      Development        Avoidance         Genetics
##              197              136              102               82
##        Enzyme(s)           Growth       Morphology    Immunological
##               62               38               22               16
##     Accumulation     Intoxication     Biochemistry          Cell(s)
##               12               12               11                9
##       Physiology        Histology       Hormone(s)
##                7                5                1
```

Answer: Population is the most studied effect with a count of 1803. Mortality follows with a count of 1493. Because population and mortality are important indicators of population status. From the population and mortality data we can better understand population health and the insects' response to neonicotinoids.

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...]

```
summary(Neonics$Species.Common.Name) %>%
  sort(decreasing = TRUE)
```

```
##                          (Other)                       Honey Bee
##                              670                             667
##                   Parasitic Wasp            Buff Tailed Bumblebee
##                              285                             183
##               Carniolan Honey Bee                      Bumble Bee
##                              152                             140
##                  Italian Honeybee                  Japanese Beetle
##                              113                              94
##                Asian Lady Beetle                   Euonymus Scale
##                               76                              75
##                         Wireworm               European Dark Bee
##                               69                              66
##                Minute Pirate Bug              Asian Citrus Psyllid
##                               62                              60
##                    Parastic Wasp           Colorado Potato Beetle
##                               58                              57
##                  Parasitoid Wasp              Erythrina Gall Wasp
##                               51                              49
##                     Beetle Order       Snout Beetle Family, Weevil
##                               47                              47
##           Sevenspotted Lady Beetle                  True Bug Order
##                               46                              45
##              Buff-tailed Bumblebee                    Aphid Family
##                               39                              38
##                    Cabbage Looper              Sweetpotato Whitefly
```

```
##                             38                             37
##                 Braconid Wasp                  Cotton Aphid
##                             33                             33
##                Predatory Mite          Ladybird Beetle Family
##                             33                             30
##                   Parasitoid                 Scarab Beetle
##                             30                             29
##                Spring Tiphia                   Thrip Order
##                             29                             29
##           Ground Beetle Family              Rove Beetle Family
##                             27                             27
##                Tobacco Aphid                  Chalcid Wasp
##                             27                             25
##         Convergent Lady Beetle                 Stingless Bee
##                             25                             25
##              Spider/Mite Class            Tobacco Flea Beetle
##                             24                             24
##              Citrus Leafminer                Ladybird Beetle
##                             23                             23
##                     Mason Bee                       Mosquito
##                             22                             22
##                  Argentine Ant                        Beetle
##                             21                             21
##      Flatheaded Appletree Borer            Horned Oak Gall Wasp
##                             20                             20
##             Leaf Beetle Family             Potato Leafhopper
##                             20                             20
##      Tooth-necked Fungus Beetle                 Codling Moth
##                             20                             19
##       Black-spotted Lady Beetle                  Calico Scale
##                             18                             18
##            Fairyfly Parasitoid                   Lady Beetle
##                             18                             18
##         Minute Parasitic Wasps                     Mirid Bug
##                             18                             18
##               Mulberry Pyralid                      Silkworm
##                             18                             18
##                 Vedalia Beetle           Araneoid Spider Order
##                             18                             17
##                     Bee Order                 Egg Parasitoid
##                             17                             17
##                  Insect Class        Moth And Butterfly Order
##                             17                             17
##    Oystershell Scale Parasitoid Hemlock Woolly Adelgid Lady Beetle
##                             17                             16
##           Hemlock Wooly Adelgid                           Mite
##                             16                             16
##                   Onion Thrip           Western Flower Thrips
##                             16                             15
##                  Corn Earworm              Green Peach Aphid
##                             14                             14
##                     House Fly                      Ox Beetle
##                             14                             14
##             Red Scale Parasite              Spined Soldier Bug
```

```
##                                14                                14
##             Armoured Scale Family                  Diamondback Moth
##                                13                                13
##                      Eulophid Wasp                 Monarch Butterfly
##                                13                                13
##                      Predatory Bug             Yellow Fever Mosquito
##                                13                                13
##                Braconid Parasitoid                      Common Thrip
##                                12                                12
##       Eastern Subterranean Termite                             Jassid
##                                12                                12
##                         Mite Order                       Pea Aphid
##                                12                                12
##                   Pond Wolf Spider           Spotless Ladybird Beetle
##                                12                                11
##             Glasshouse Potato Wasp                          Lacewing
##                                10                                10
##            Southern House Mosquito            Two Spotted Lady Beetle
##                                10                                10
##                          Ant Family                      Apple Maggot
##                                 9                                 9
```

Answer: The six most studied species are Honey Bee, Parasitic Wasp, Buff Tailed Bumblebee, Carniolan Honey Bee, Bumble Bee, Italian Honeybee. They are all bees/wasps. Because bees are important pollinators and essential for food and honey production, people might be more interested to study them over other species.

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?

```
class(Neonics$Conc.1..Author.)
```

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

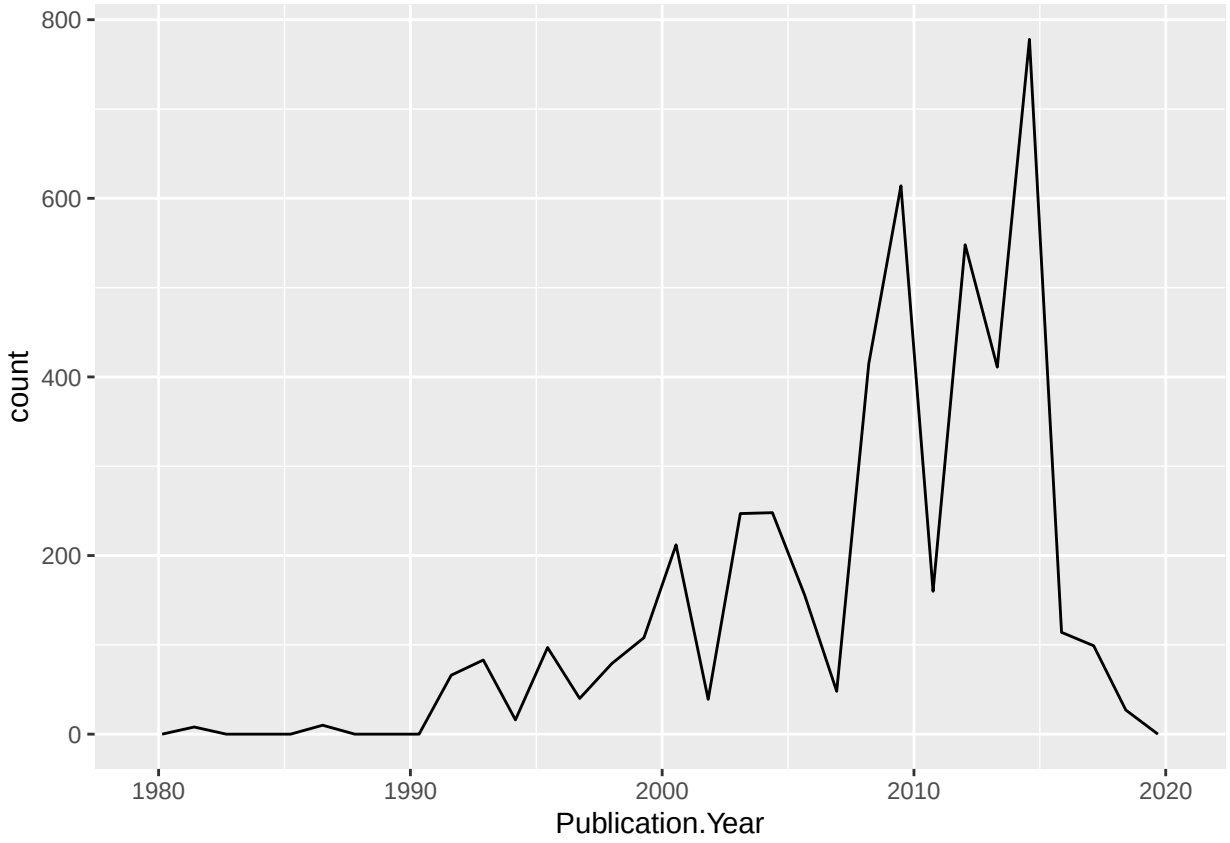
Answer: The class is factor. When we uploaded the dataset, we asked R to read strings as factor, and the "Conc.1..Author" column contains strings such as NR.

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))
```

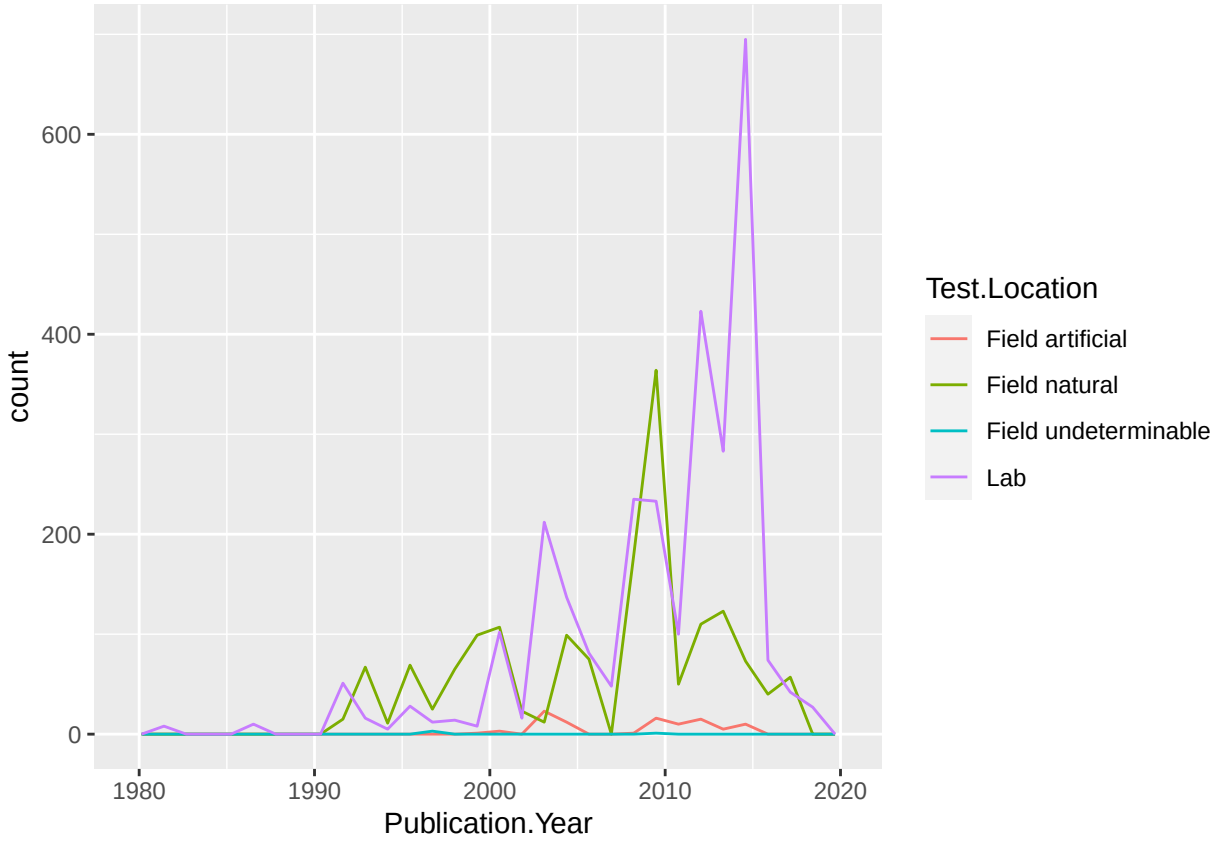
```
## `stat_bin()` using `bins = 30`. Pick better value with `binwidth`.
```



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

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

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



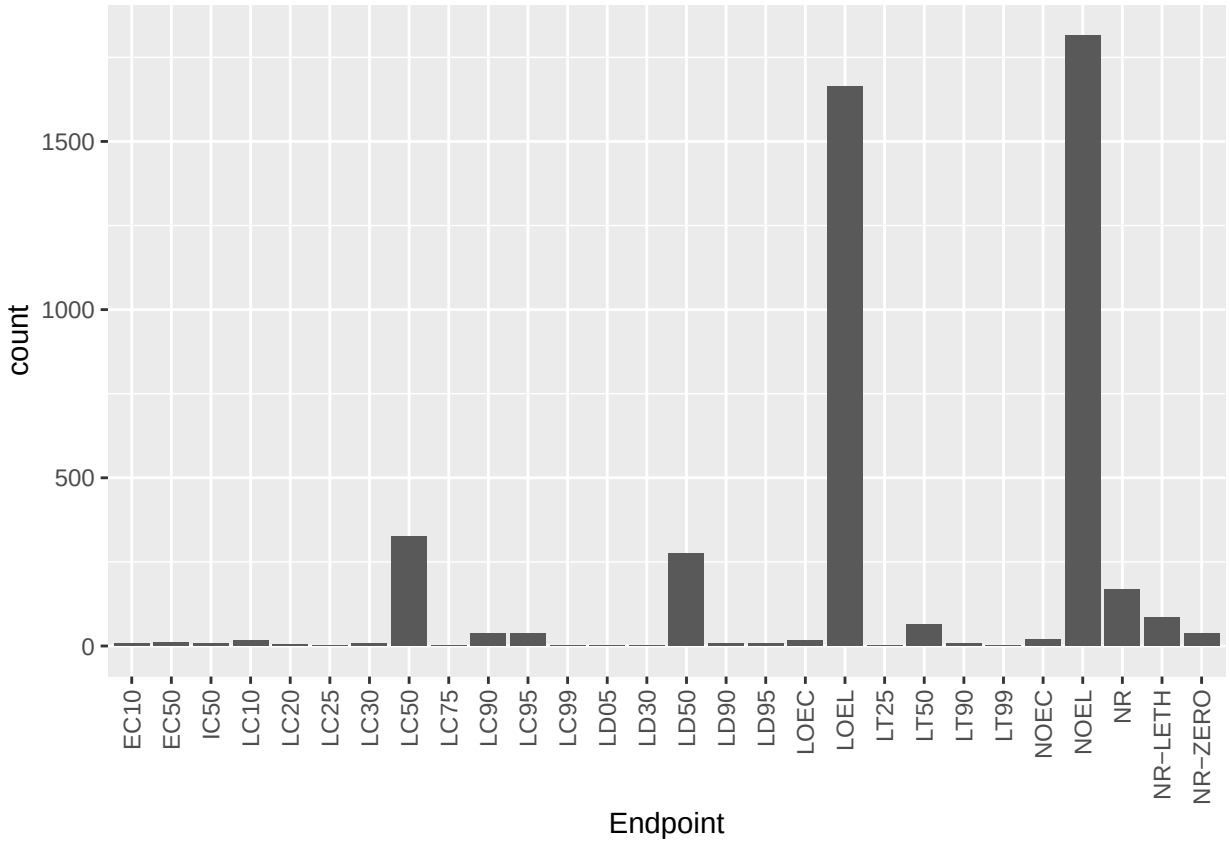
Interpret this graph. What are the most common test locations, and do they differ over time?

> Answer: Field artifical and Field natural are the two most common test locations. From 1990 to 2000, Field natural sites are mostly more than the artificial sites. Then, the artificals sites began to increase and exceeded the natural sites around 2010, while natural sites sharply decreased 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.

[**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...]

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



Answer: LOEL and NOEL are the two most common end points. LOEL stands for Lowest-observable-effect-level, which means the lowest concentration producing effects that were significantly different from responses of controls. NOEL stands for No-observable-effect-level, which means the highest concentration producing effects not significantly different from responses of controls according to author's reported statistical test.

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$collectDate <- ymd(Litter$collectDate)
class(Litter$collectDate)
```

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

```
unique(Litter$collectDate) # Litter was sampled on August 2nd and 30th.
```

```
## [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`?

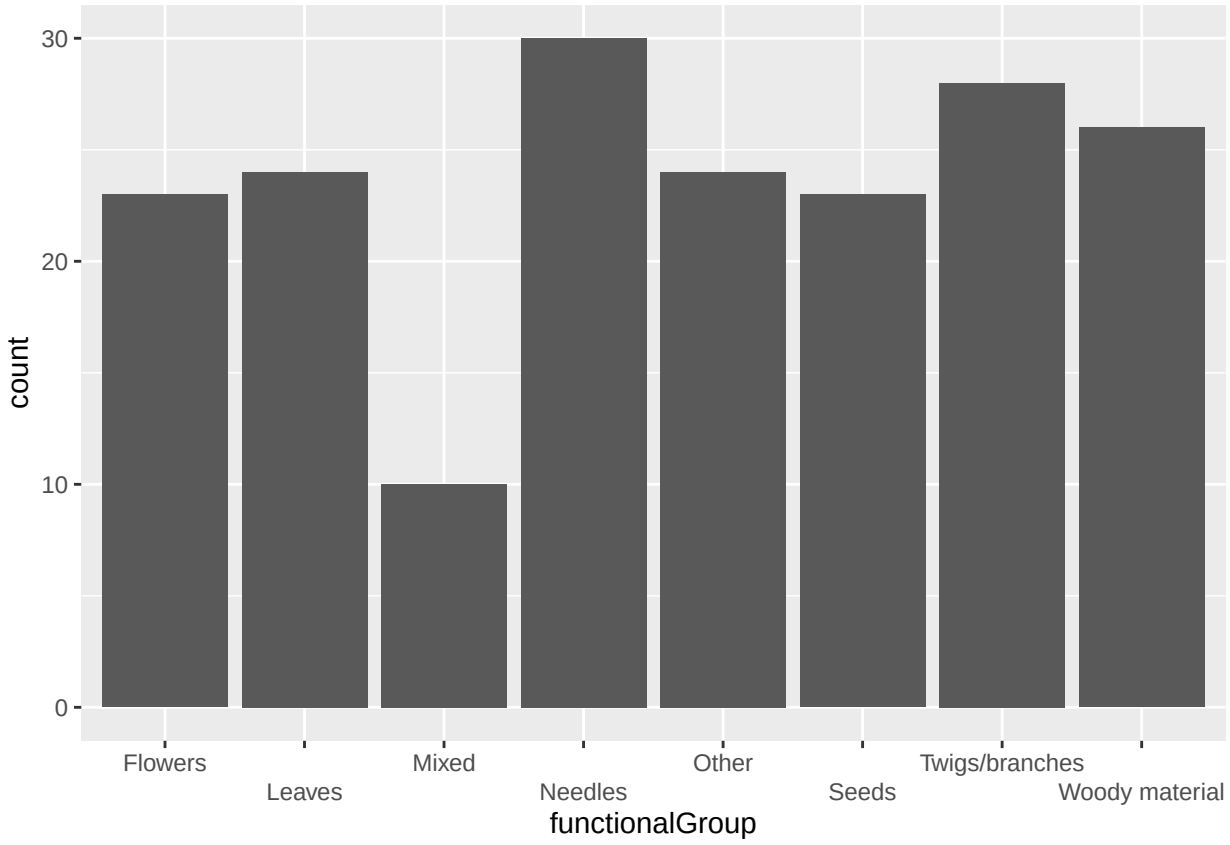
```
unique(Litter$plotID)
```

```
##  [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
```

Answer: 12 plots were sampled at Niwot Ridge. The "unique" function gives the name and number of the unique values in the column, while the "summary" function will summarize the data under each unique value separately.

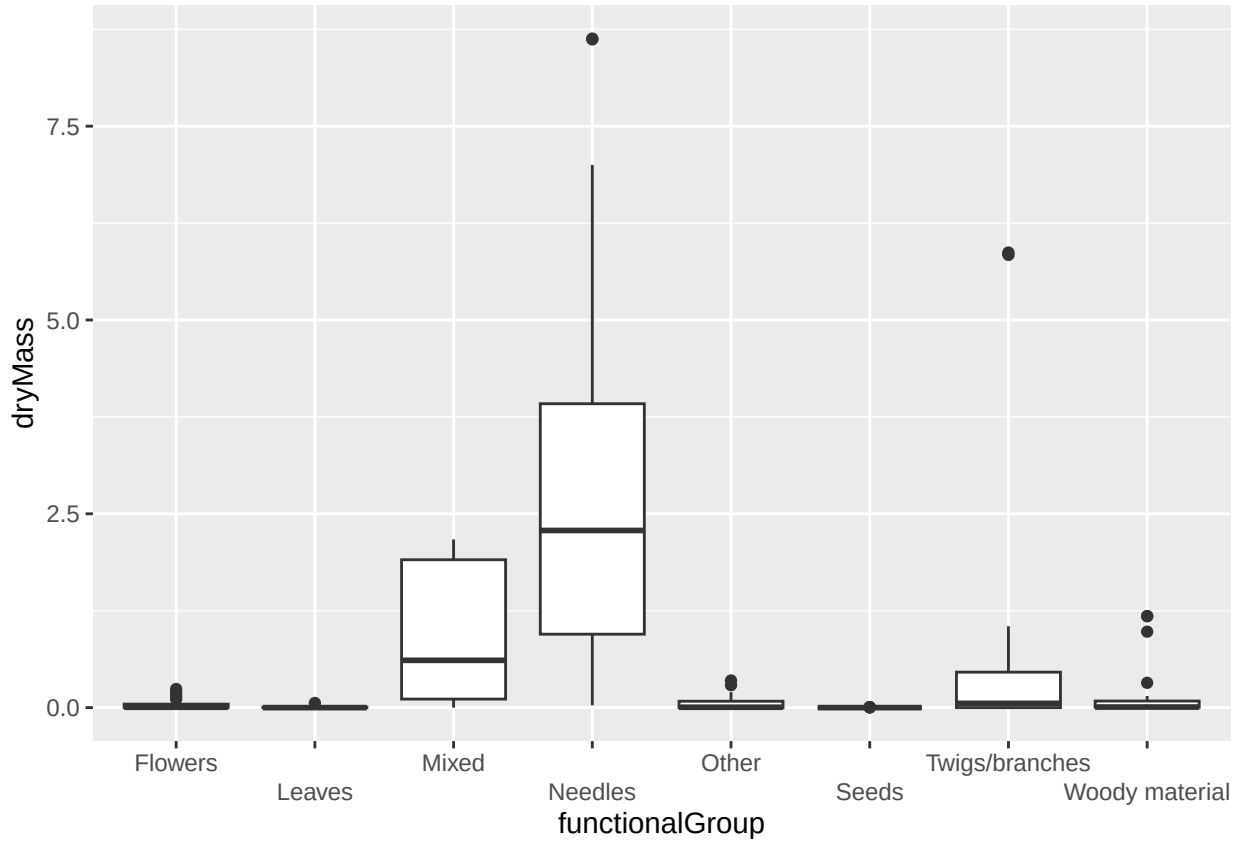
14. Create a bar graph of functionalGroup counts. This shows you what type of litter is collected at the Niwot Ridge sites. Notice that litter types are fairly equally distributed across the Niwot Ridge sites.

```
ggplot(Litter) +
  geom_bar(aes(x = functionalGroup)) +
  scale_x_discrete(guide = guide_axis(n.dodge = 2)) # Make overlapping x labels shift a step-down
```

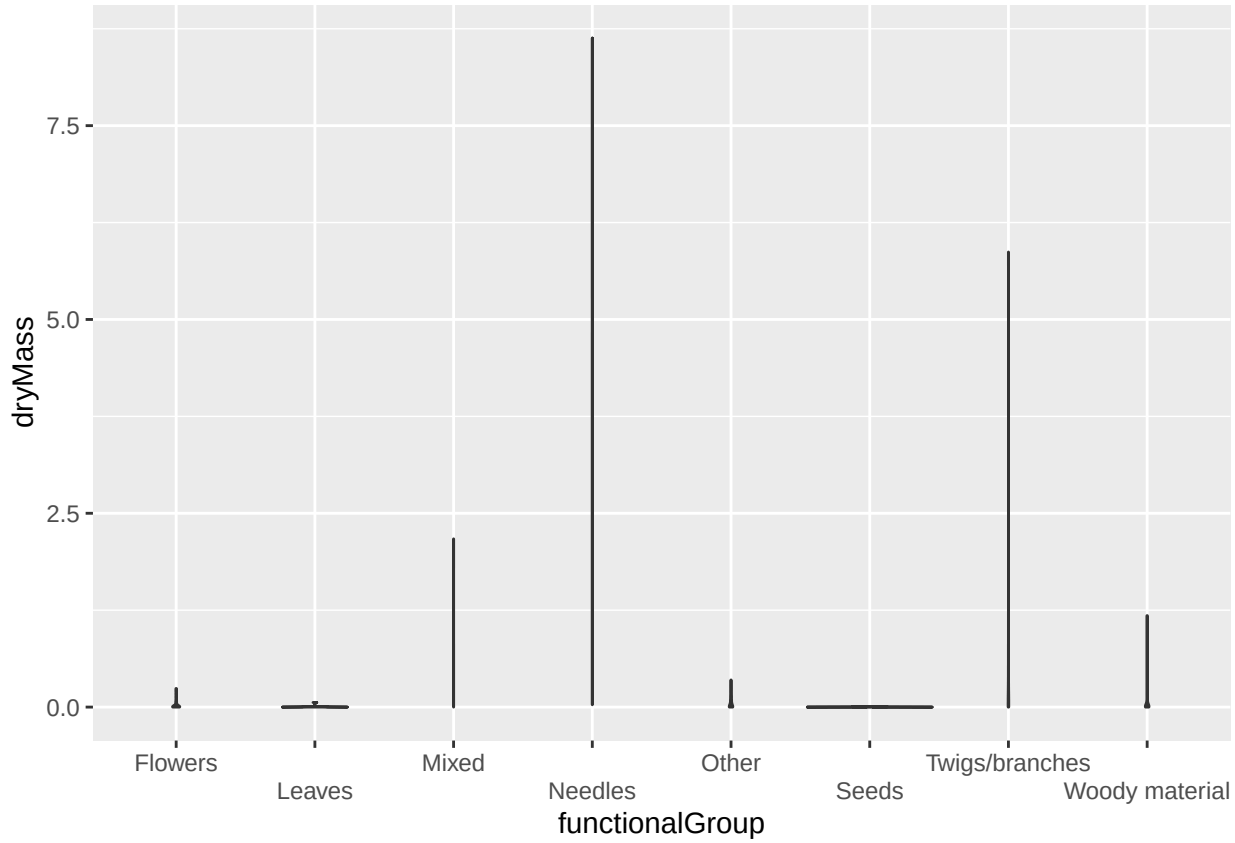


15. Using `geom_boxplot` and `geom_violin`, create a boxplot and a violin plot of dryMass by functionalGroup.

```
ggplot(Litter) +
  geom_boxplot(aes(x = functionalGroup, y = dryMass)) +
  scale_x_discrete(guide = guide_axis(n.dodge = 2)) # Make overlapping x labels shift a step-down
```




```
ggplot(Litter) +
  geom_violin(aes(x = functionalGroup, y = dryMass)) +
  scale_x_discrete(guide = guide_axis(n.dodge = 2)) # Make overlapping x labels shift a step-down
```



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

> Answer: Because the dry mass values are small and very similar within each functional group, so the violin plot couldn't show a clear distribution pattern.

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

> Answer: Needles tend to have the highest biomass at these sites.