Assignment 3: Data Exploration

Hanna Karnei

Fall 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)
library(ggplot2)
```

```
getwd()
```

```
## [1] "/Users/hannakarnei/Desktop/EDA/EDE_Fall2023"
```

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

   Answer: Many, if not all, insects are a crucial part of the ecosystem. We might want to know about the effects of neonicotinoids on insects to undestand how the insecticides affect invertebrates and to make inferences on their effects on the ecosystem at large.

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 and forest litter play an important role in the ecosystem, serving as a home to many terrestial organisms and participating in nutrient cycling. Studying it can help us better undrstand its role in stream and forest ecosystems.

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. Spatial Design: Sample locations are chosen randomly and grouped into plots. 2. Spatial Design: Traps inside plots are either placed randomly or intentionally, depending on the vegetation. 3. Temporal Design: The traps are sampled once per year; elevated traps are sampled once every 2 weeks or once every 1-2 months, based on the vegetation type.

Obtain basic summaries of your data (Neonics)

5. What are the dimensions of the dataset?

```
dim(Neonics) #4,623 observations (i.e. rows) of 30 variables (i.e. columns)
```

```
## [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?

```
sort(summary(Neonics$Effect))
```

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

Answer: The most common effects studied are the effects on population and mortality. These effects are important to study because they are a good proxy for the toxicity of Neonicotinoids for insects.

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. . .]

```
sort(summary(Neonics$Species.Common.Name))
```

```
##                         Ant Family                         Apple Maggot
##                                  9                                    9
##              Glasshouse Potato Wasp                            Lacewing
##                                 10                                   10
##             Southern House Mosquito              Two Spotted Lady Beetle
##                                 10                                   10
##            Spotless Ladybird Beetle                  Braconid Parasitoid
##                                 11                                   12
##                       Common Thrip         Eastern Subterranean Termite
##                                 12                                   12
##                             Jassid                           Mite Order
##                                 12                                   12
##                          Pea Aphid                    Pond Wolf Spider
##                                 12                                   12
##               Armoured Scale Family                    Diamondback Moth
##                                 13                                   13
##                       Eulophid Wasp                   Monarch Butterfly
##                                 13                                   13
##                       Predatory Bug                Yellow Fever Mosquito
##                                 13                                   13
##                        Corn Earworm                   Green Peach Aphid
##                                 14                                   14
##                           House Fly                            Ox Beetle
##                                 14                                   14
##                  Red Scale Parasite                   Spined Soldier Bug
##                                 14                                   14
##               Western Flower Thrips Hemlock Woolly Adelgid Lady Beetle
##                                 15                                   16
##               Hemlock Wooly Adelgid                                Mite
##                                 16                                   16
##                         Onion Thrip                Araneoid Spider Order
```

```
##                             16                                  17
##                     Bee Order                  Egg Parasitoid
##                             17                                  17
##                   Insect Class        Moth And Butterfly Order
##                             17                                  17
##    Oystershell Scale Parasitoid       Black-spotted Lady Beetle
##                             17                                  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
##                   Codling Moth       Flatheaded Appletree Borer
##                             19                                  20
##            Horned Oak Gall Wasp              Leaf Beetle Family
##                             20                                  20
##              Potato Leafhopper      Tooth-necked Fungus Beetle
##                             20                                  20
##                  Argentine Ant                          Beetle
##                             21                                  21
##                    Mason Bee                        Mosquito
##                             22                                  22
##               Citrus Leafminer                Ladybird Beetle
##                             23                                  23
##              Spider/Mite Class           Tobacco Flea Beetle
##                             24                                  24
##                   Chalcid Wasp          Convergent Lady Beetle
##                             25                                  25
##                   Stingless Bee           Ground Beetle Family
##                             25                                  27
##              Rove Beetle Family                  Tobacco Aphid
##                             27                                  27
##                   Scarab Beetle                  Spring Tiphia
##                             29                                  29
##                     Thrip Order          Ladybird Beetle Family
##                             29                                  30
##                    Parasitoid                  Braconid Wasp
##                             30                                  33
##                   Cotton Aphid                Predatory Mite
##                             33                                  33
##            Sweetpotato Whitefly                   Aphid Family
##                             37                                  38
##                  Cabbage Looper           Buff-tailed Bumblebee
##                             38                                  39
##                  True Bug Order         Sevenspotted Lady Beetle
##                             45                                  46
##                    Beetle Order       Snout Beetle Family, Weevil
##                             47                                  47
##             Erythrina Gall Wasp                Parasitoid Wasp
##                             49                                  51
##          Colorado Potato Beetle                  Parastic Wasp
```

```
##                             57                             58
##           Asian Citrus Psyllid            Minute Pirate Bug
##                             60                             62
##              European Dark Bee                      Wireworm
##                             66                             69
##                 Euonymus Scale            Asian Lady Beetle
##                             75                             76
##               Japanese Beetle             Italian Honeybee
##                             94                            113
##                    Bumble Bee          Carniolan Honey Bee
##                            140                            152
##          Buff Tailed Bumblebee               Parasitic Wasp
##                            183                            285
##                     Honey Bee                      (Other)
##                            667                            670
```

Answer: The most studied species (excluding "Other") are: 1. Honey Bee 2. Parasitic Wasp 3. Buff Tailed Bumblebee 4. Carniolan Honey Bee 5. Bumble Bee 6. Italian Honeybee All bees and wasps on the list belong to the order of Hymenoptera. It is so important to understand the effect of Neonicotinoids on bees and wasps because bees are pollinators in ecosystems and parastic wasps can contribute to pollination indirectly and contribute to pest control. Without them, our ecosystems would rapidly degrade and pests would proliferate.

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.)
```

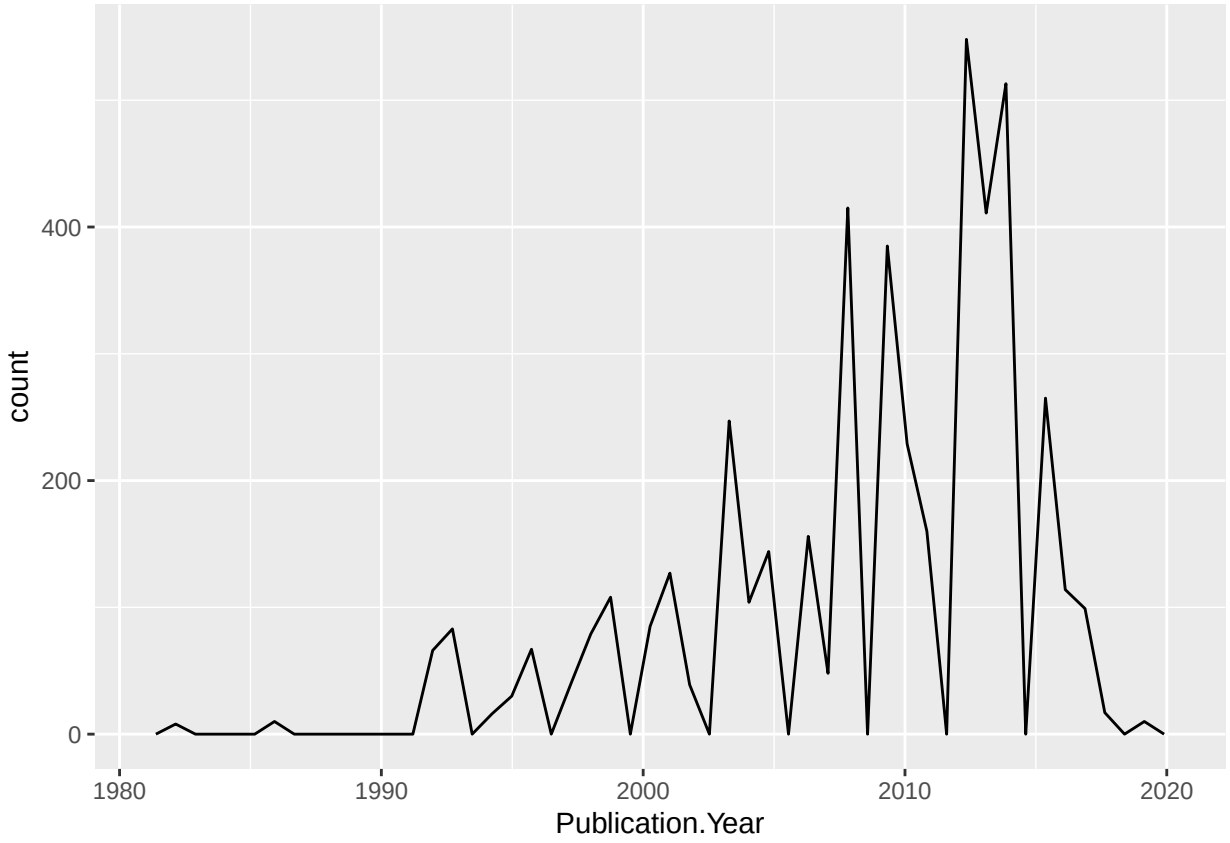
```
## [1] "factor"
```

Answer: Its class is "factor". The variable is not numberic because the records shows that this column contains special characters like, such as "/", in addition to numbers.

Explore your data graphically (Neonics)

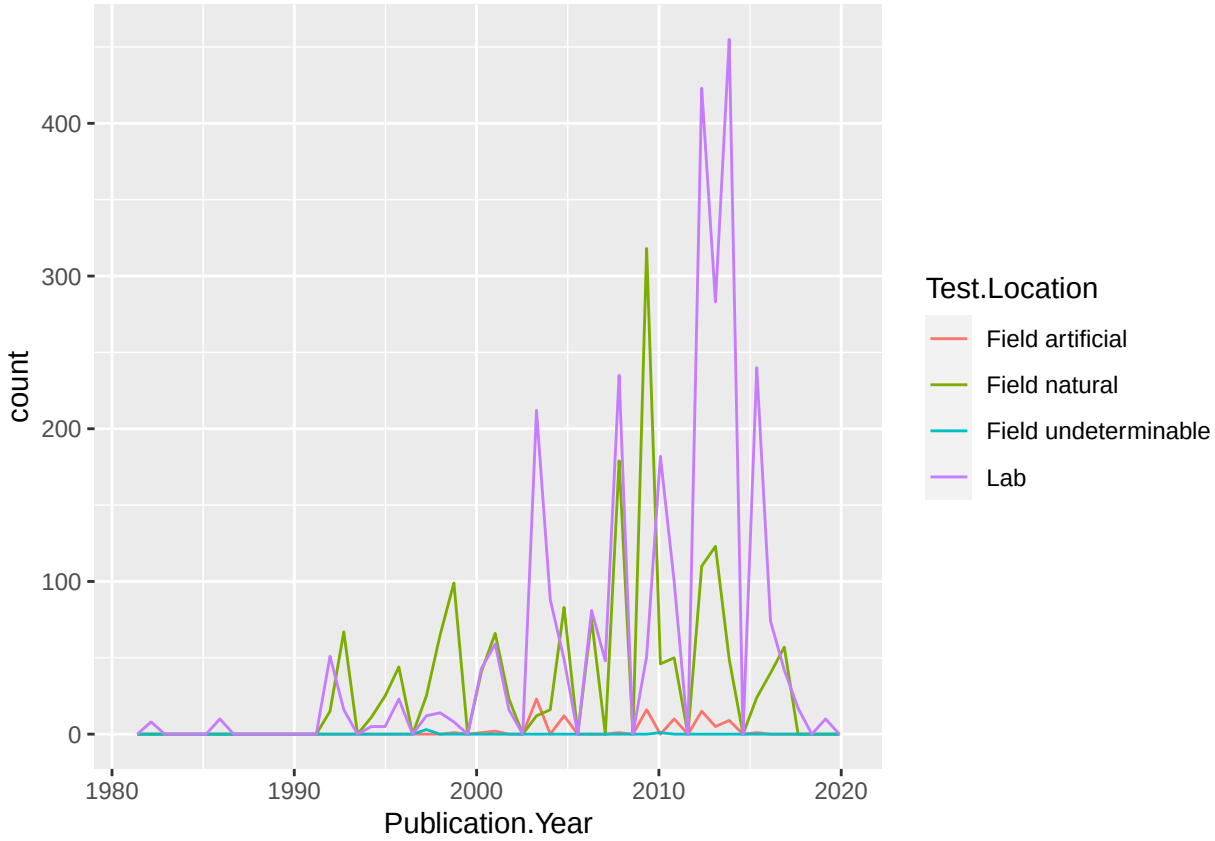
9. Using `geom_freqpoly`, generate a plot of the number of studies conducted by publication year.

```
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.

```
ggplot(Neonics) +
  geom_freqpoly(aes(x = Publication.Year, color=Test.Location), bins = 50)
```



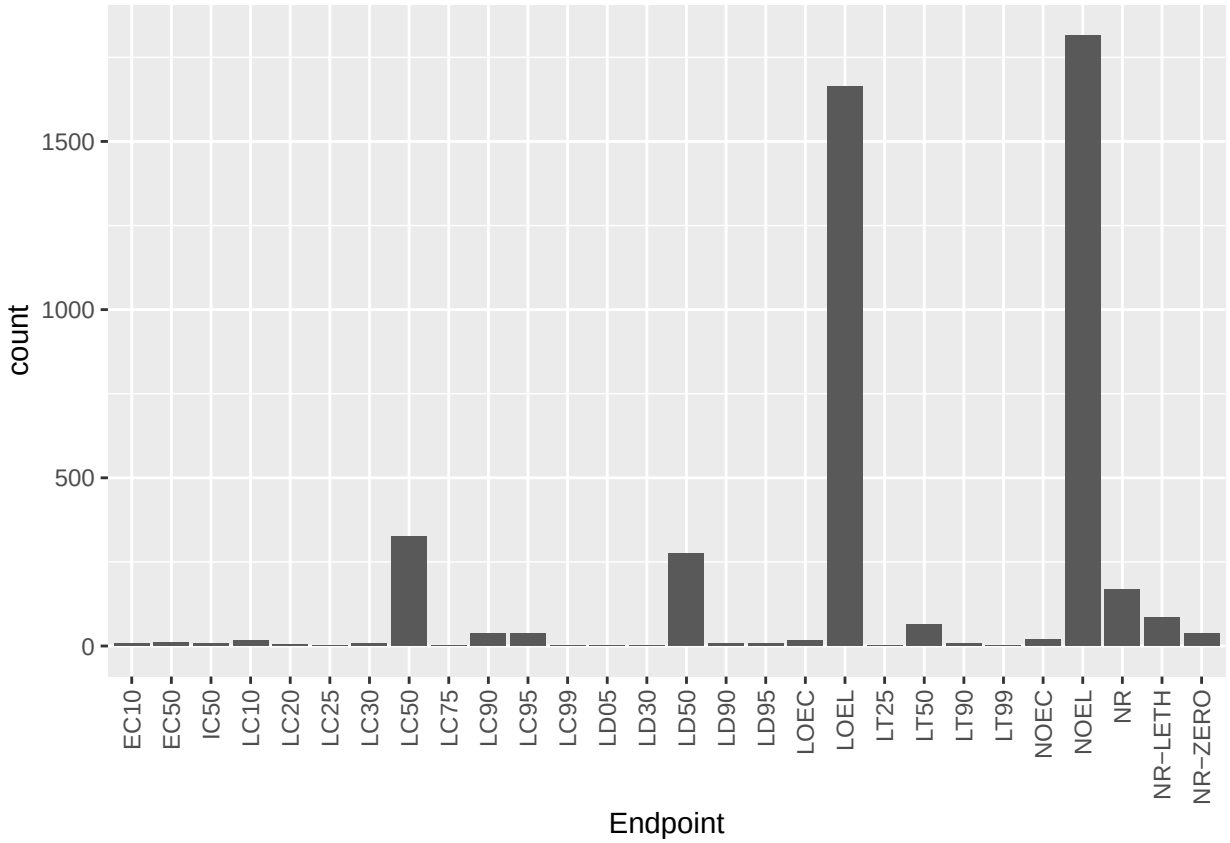
Interpret this graph. What are the most common test locations, and do they differ over time?

> Answer: The most common test locations are "lab" and "field natural". "Field natural" was more popular between early 1990 and 2000 and saw another spike in late 2000s. Between 1980 and 1990, as well as between mid-2000s and 2020 (save for the aforementioned spike for "field natural"), the lab was the most common test locaiton.

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



Answer: Two most common endpoints were NOEL and LOEL. NOEL was defined as "no-observable-effect-level", meaning that the highest concentration of the toxin did not produce effects that were different from the effects of the controls. LOEL was defined as "lowest-observable-effect-level" and was used to label the group for which the lowest concentration of the toxin produced effects that were significanlty different from the responses of the control group.

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_dates <- unique(Litter$collectDate)
unique_dates
```

```
## [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
```

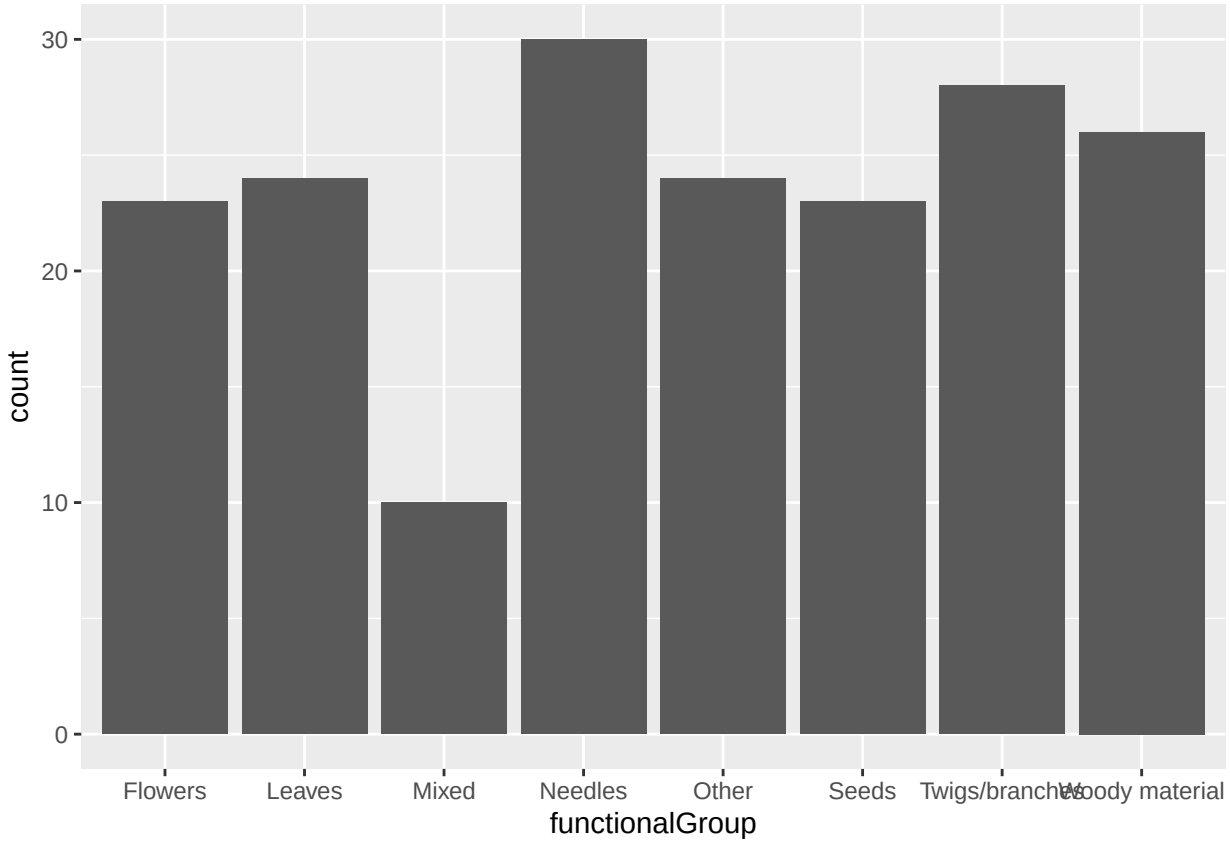
```
summary(Litter$plotID)
```

```
## 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: There were 12 unique plots examined at Niwot Ridge. The "Unique" function listed every unique plot name and a total number of plots ("12 Levels"), meanwhile the "summary" function provided the count of observations per each unique plot but did not provide the number of unique plots.

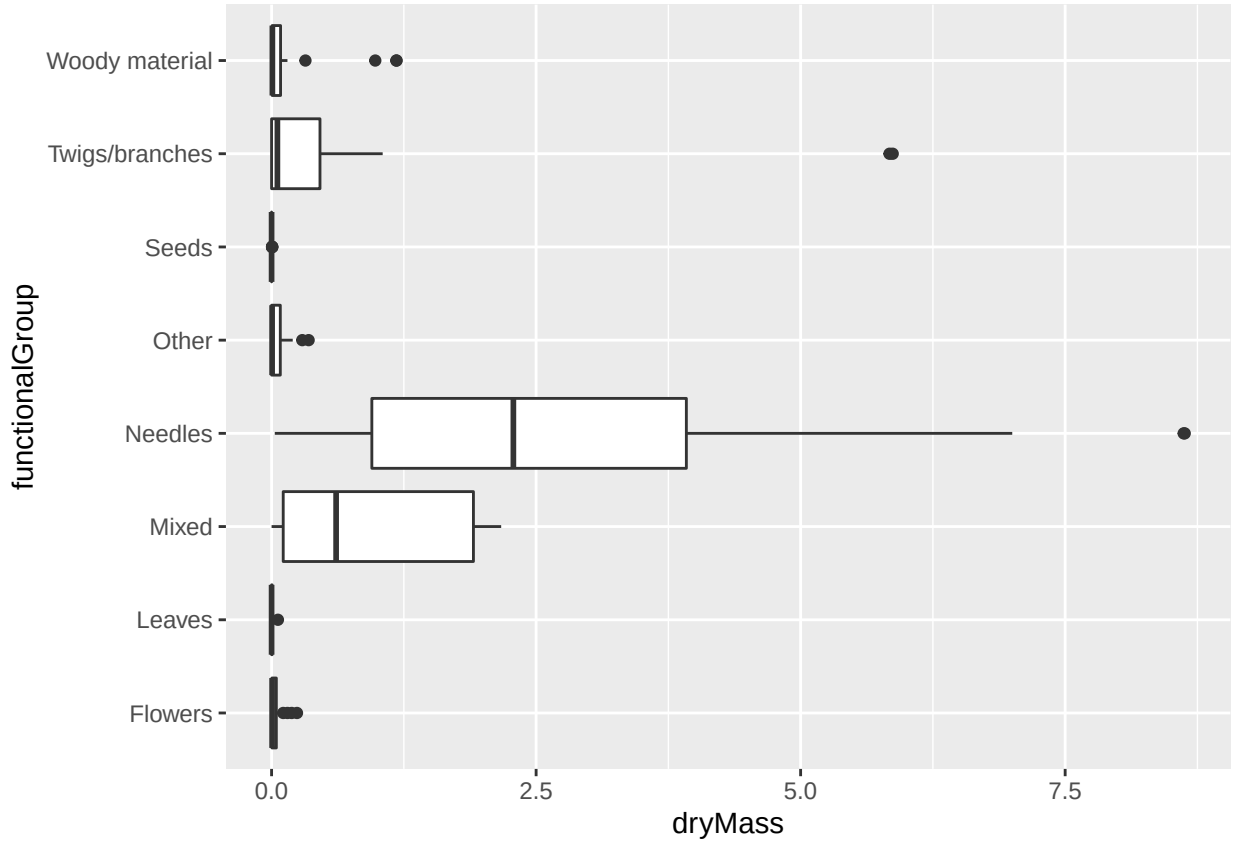
14. Create a bar graph of functionalGroup counts. This shows you what type of litter is collected at the Niwot Ridge sites. Notice that litter types are fairly equally distributed across the Niwot Ridge sites.

```
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 = dryMass, y = functionalGroup))
```

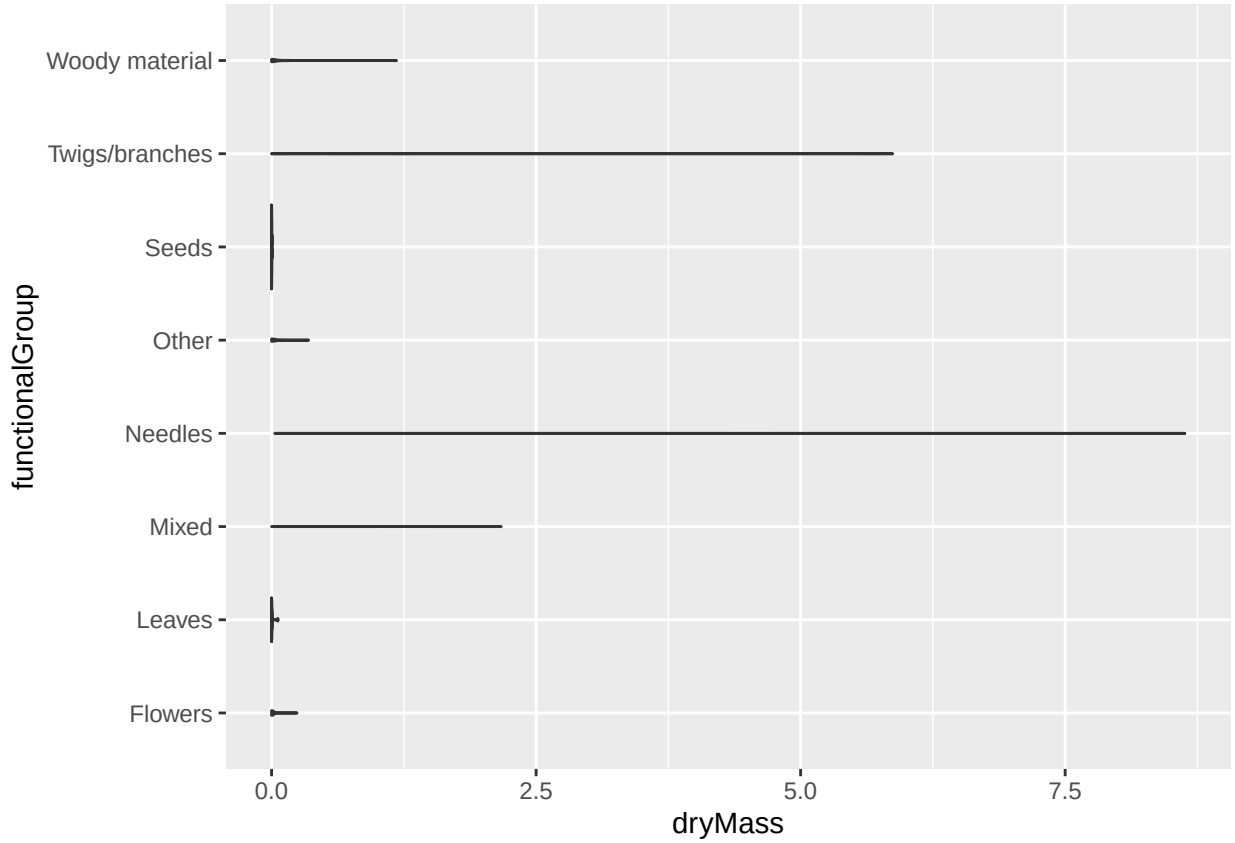



```
ggplot(Litter) +
  geom_violin(aes(x = dryMass, y = functionalGroup),
              draw_quantiles = c(0.25, 0.5, 0.75))
```

```
## 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 violin plot is used for displaying density distribution. In our case, the density distribution of dry mass is not variable, and so the violin plot only shows the range of data for each category, which is not very informative. In constrast, the boxplot allows us to see the distributoin of data clearly, including the interquartile range, outliers, and median.

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

> Answer: Needles, twigs/branches, and mixed dry mass tend to have the highest biomass.