Assignment 3: Data Exploration

Tani Valdez Rivas

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.

```
# Setting working directory
getwd()
```

```
## [1] "/home/guest/R/EDE_Fall2023"
```

```
# Installing packages
library(tidyverse)
library(lubridate)

# Uploading the data sets
```

```
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: Neonicotinoids were first registered for use in the early 1990s. It was not until the mid-2000s did the use of neonicotinoid use exploded in popularity in the agricultural setting. Although the application of insecticide has shifted to seed coating, neoincotinoids have a strong effect on non-target organisms. Pollinators and aquatic insects have been exposed to neoincotinoids. It has reduced their longevity, behavior, etc.

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 wood debris is important for nutrient cycling and moisture in forests. Addittionally, it provides habitat for terrestrial organisms and plays an important role in carbon sequestration. It shapes the structure and roughness of the forest ground, influencing 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. Sampling sites must contain woody vegetation that >2 meters tall. 2. Depending on the vegetation, trap placement within plots are either random or not random. 3. Ground traps are sampled once a year.

Obtain basic summaries of your data (Neonics)

5. What are the dimensions of the dataset?

```
# Finding the dimensions using dim() function
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?

```r
# Finding the dimensions on Effect
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: Mortality, population, and behavior seem to be the most common effects that are studied. Neonicotinoids have been shown to be toxic to target and non-target organisms. By understanding the phsyioloigcal and neural effects on these species, it might lead scientists and interested parties to adjust the amount of insecticide used during application or switch to a new insecticide.

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...]

```r
# Using the summary function to extract Species Common Name
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 are Honey bees, Parasitic Wasps, Buff Tailed Bumblebees, Cariolan Honey Bees, Bumble Bees, Italian Honeybees. Bees are in an important pollinator in agricultural ecosystems. However, they are very susceptible to illness or death insecticides and are exposed to them frequently.

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?

```
# Using the class() function to determine Conc.1..Author class
class(Neonics$Conc.1..Author.)
```

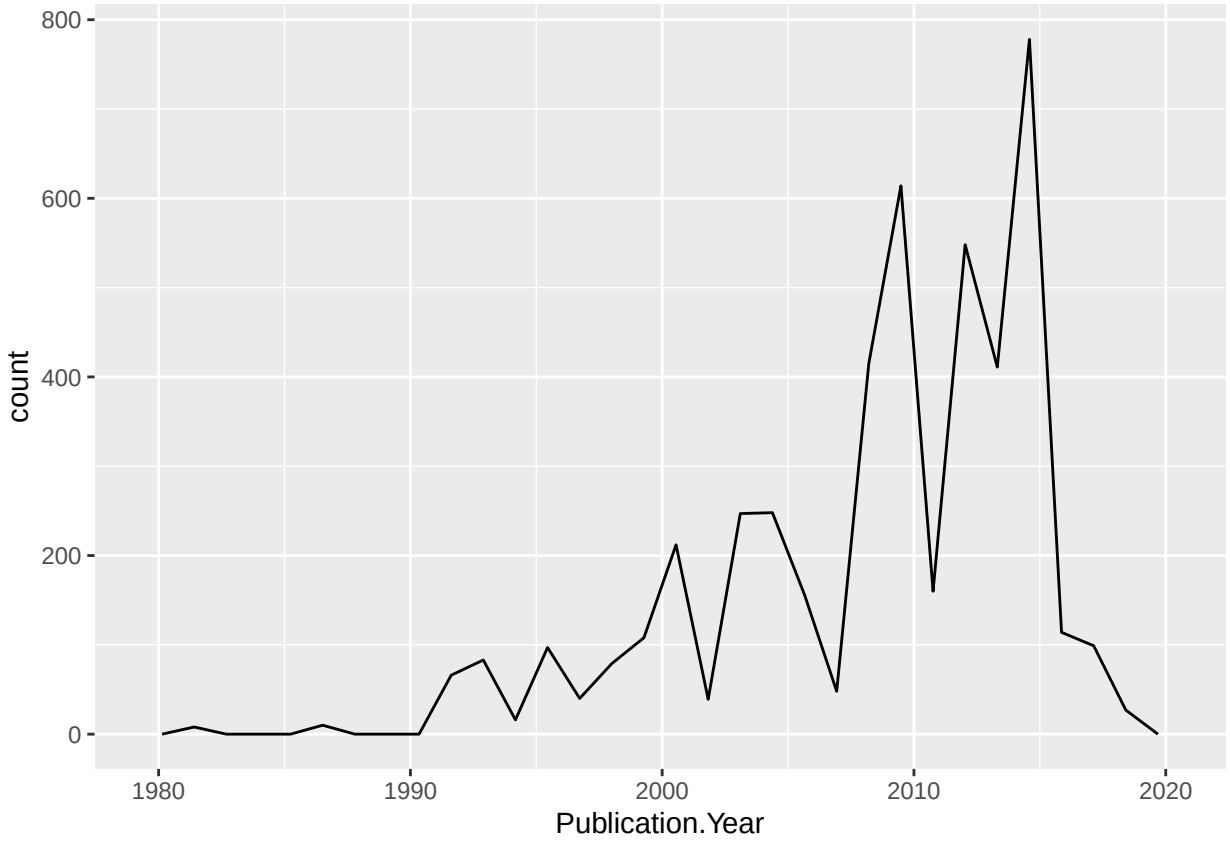
```
## [1] "factor"
```

Answer: Conc.1..Author are factors and not numeric. Conc.1..Author is a string that is repres-nting that value, so that means that Conc.1..Author has multiple categories.

Explore your data graphically (Neonics)

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

```
# Creating ggplot first to then use geom_freqpoly
ggplot(Neonics) +
  geom_freqpoly(aes(x = Publication.Year), bins = 30)
```



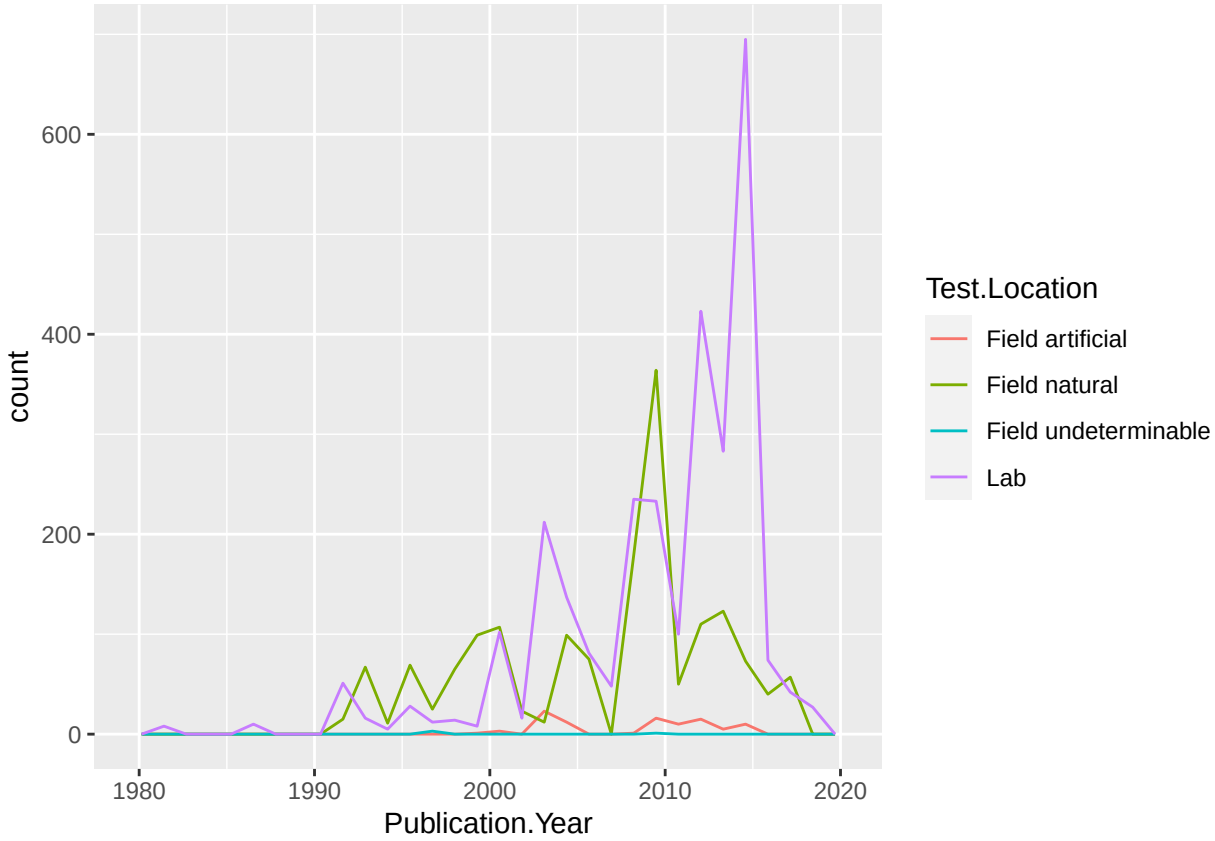
10. Reproduce the same graph but now add a color aesthetic so that different Test.Location are displayed as different colors.

```
# Installing ggplot2 as my plots were not visible
install.packages("ggplot2")
```

```
## Installing package into '/home/guest/R/x86_64-pc-linux-gnu-library/4.3'
## (as 'lib' is unspecified)
```

```
library(ggplot2)

# Creating a ggplot and differentiating test locations by color
ggplot(Neonics) +
  geom_freqpoly(aes(x = Publication.Year, color = Test.Location), bins = 30)
```



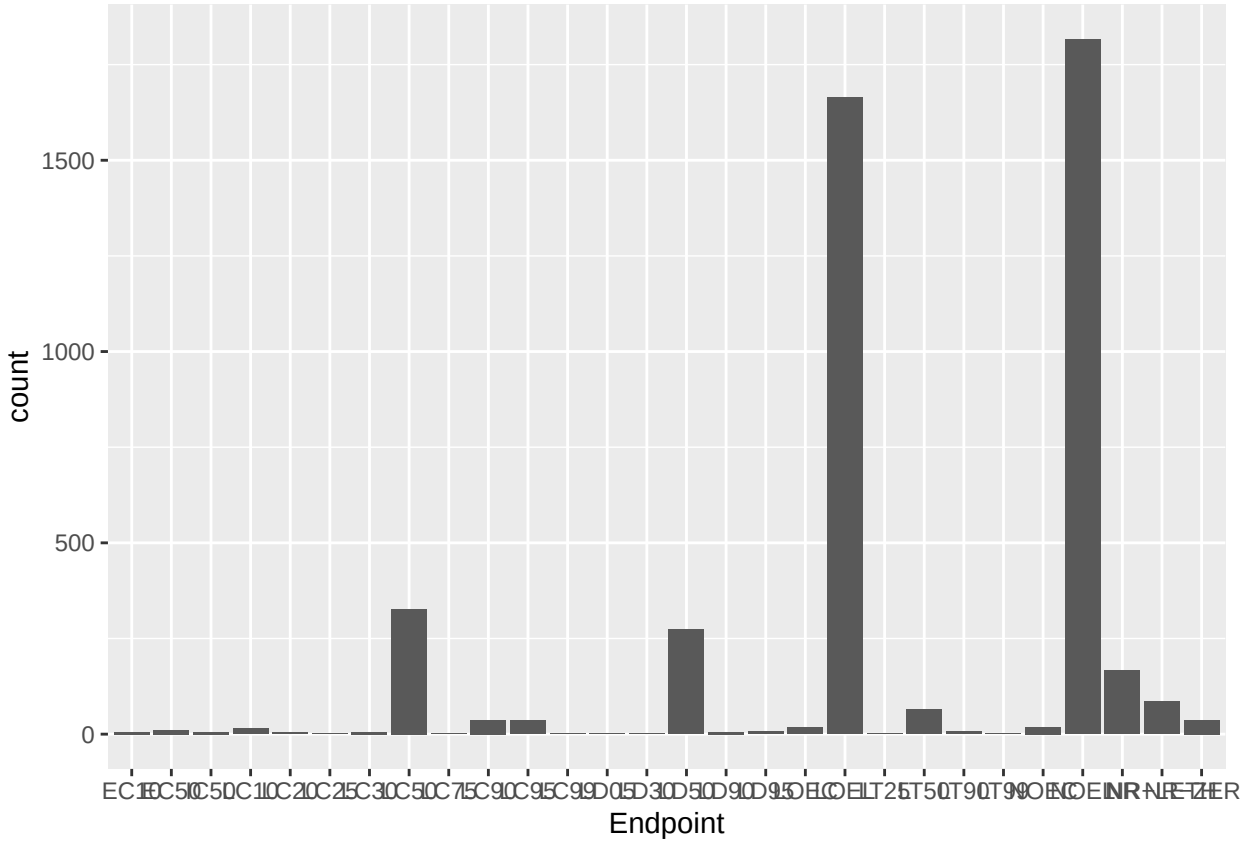
Interpret this graph. What are the most common test locations, and do they differ over time?

> Answer: The two most common locations are the labs and natural fields. Initially (between 1980 - 1990), the most common test sites were in undetermine fields and labs. But as the use of insecticide increased, it was easier for scientist to go actual locations and conduct surveys.

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...]

```
# Using ggplot and then geom-bar to create a bar graph of endpoint counts
ggplot(Neonics, aes(x = Endpoint)) +
  geom_bar()
```




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

```
## List of 1
##  $ axis.text.x:List of 11
##   ..$ family       : NULL
##   ..$ face         : NULL
##   ..$ colour       : NULL
##   ..$ size         : NULL
##   ..$ hjust        : num 1
##   ..$ vjust        : num 0.5
##   ..$ angle        : num 90
##   ..$ lineheight   : NULL
##   ..$ margin       : NULL
##   ..$ debug        : NULL
##   ..$ inherit.blank: logi FALSE
##   ..- attr(*, "class")= chr [1:2] "element_text" "element"
##  - attr(*, "class")= chr [1:2] "theme" "gg"
##  - attr(*, "complete")= logi FALSE
##  - attr(*, "validate")= logi TRUE
```

Answer: NOEL and LOEL are the two most common points. NOEL is defined as no-observable-effect-level. This means that the highest concentration of insecticide and its effect is not different from the responses of the control. LOEL is defined as Lowest-observable-effect-level. This indicates that the effects from a low concentration were different from the control.

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.

```
# Determining class of Litter$collect
class(Litter$collectDate) # it is a factor
```

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

```
# Changing collectDate from factor to date
Litter_CollectData <- ymd(Litter$collectDate)
Litter_CollectData
```

```
##   [1] "2018-08-02" "2018-08-02" "2018-08-02" "2018-08-02" "2018-08-02"
##   [6] "2018-08-02" "2018-08-02" "2018-08-02" "2018-08-02" "2018-08-02"
##  [11] "2018-08-02" "2018-08-02" "2018-08-02" "2018-08-02" "2018-08-02"
##  [16] "2018-08-02" "2018-08-02" "2018-08-02" "2018-08-02" "2018-08-02"
##  [21] "2018-08-02" "2018-08-02" "2018-08-02" "2018-08-02" "2018-08-02"
##  [26] "2018-08-02" "2018-08-02" "2018-08-02" "2018-08-02" "2018-08-02"
##  [31] "2018-08-02" "2018-08-02" "2018-08-02" "2018-08-02" "2018-08-02"
##  [36] "2018-08-02" "2018-08-02" "2018-08-02" "2018-08-02" "2018-08-02"
##  [41] "2018-08-02" "2018-08-02" "2018-08-02" "2018-08-02" "2018-08-02"
##  [46] "2018-08-02" "2018-08-02" "2018-08-02" "2018-08-02" "2018-08-02"
##  [51] "2018-08-02" "2018-08-02" "2018-08-02" "2018-08-02" "2018-08-02"
##  [56] "2018-08-02" "2018-08-02" "2018-08-02" "2018-08-02" "2018-08-02"
##  [61] "2018-08-02" "2018-08-02" "2018-08-02" "2018-08-02" "2018-08-02"
##  [66] "2018-08-02" "2018-08-02" "2018-08-02" "2018-08-02" "2018-08-02"
##  [71] "2018-08-02" "2018-08-02" "2018-08-02" "2018-08-02" "2018-08-02"
##  [76] "2018-08-02" "2018-08-02" "2018-08-02" "2018-08-02" "2018-08-02"
##  [81] "2018-08-02" "2018-08-02" "2018-08-02" "2018-08-02" "2018-08-02"
##  [86] "2018-08-02" "2018-08-02" "2018-08-02" "2018-08-02" "2018-08-02"
##  [91] "2018-08-02" "2018-08-30" "2018-08-30" "2018-08-30" "2018-08-30"
##  [96] "2018-08-30" "2018-08-30" "2018-08-30" "2018-08-30" "2018-08-30"
## [101] "2018-08-30" "2018-08-30" "2018-08-30" "2018-08-30" "2018-08-30"
## [106] "2018-08-30" "2018-08-30" "2018-08-30" "2018-08-30" "2018-08-30"
## [111] "2018-08-30" "2018-08-30" "2018-08-30" "2018-08-30" "2018-08-30"
## [116] "2018-08-30" "2018-08-30" "2018-08-30" "2018-08-30" "2018-08-30"
## [121] "2018-08-30" "2018-08-30" "2018-08-30" "2018-08-30" "2018-08-30"
## [126] "2018-08-30" "2018-08-30" "2018-08-30" "2018-08-30" "2018-08-30"
## [131] "2018-08-30" "2018-08-30" "2018-08-30" "2018-08-30" "2018-08-30"
## [136] "2018-08-30" "2018-08-30" "2018-08-30" "2018-08-30" "2018-08-30"
## [141] "2018-08-30" "2018-08-30" "2018-08-30" "2018-08-30" "2018-08-30"
## [146] "2018-08-30" "2018-08-30" "2018-08-30" "2018-08-30" "2018-08-30"
## [151] "2018-08-30" "2018-08-30" "2018-08-30" "2018-08-30" "2018-08-30"
## [156] "2018-08-30" "2018-08-30" "2018-08-30" "2018-08-30" "2018-08-30"
## [161] "2018-08-30" "2018-08-30" "2018-08-30" "2018-08-30" "2018-08-30"
## [166] "2018-08-30" "2018-08-30" "2018-08-30" "2018-08-30" "2018-08-30"
## [171] "2018-08-30" "2018-08-30" "2018-08-30" "2018-08-30" "2018-08-30"
## [176] "2018-08-30" "2018-08-30" "2018-08-30" "2018-08-30" "2018-08-30"
## [181] "2018-08-30" "2018-08-30" "2018-08-30" "2018-08-30" "2018-08-30"
## [186] "2018-08-30" "2018-08-30" "2018-08-30"
```

```r
# Determining class of Litter_collectDate
class(Litter_CollectData)
```

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

```r
# Finding dates litter were sampled in August 2018
unique(Litter_CollectData)
```

```
## [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`?

```r
# Using unique () to determine the plots sampled at Niwot Ridge
unique(Litter$namedLocation)
```

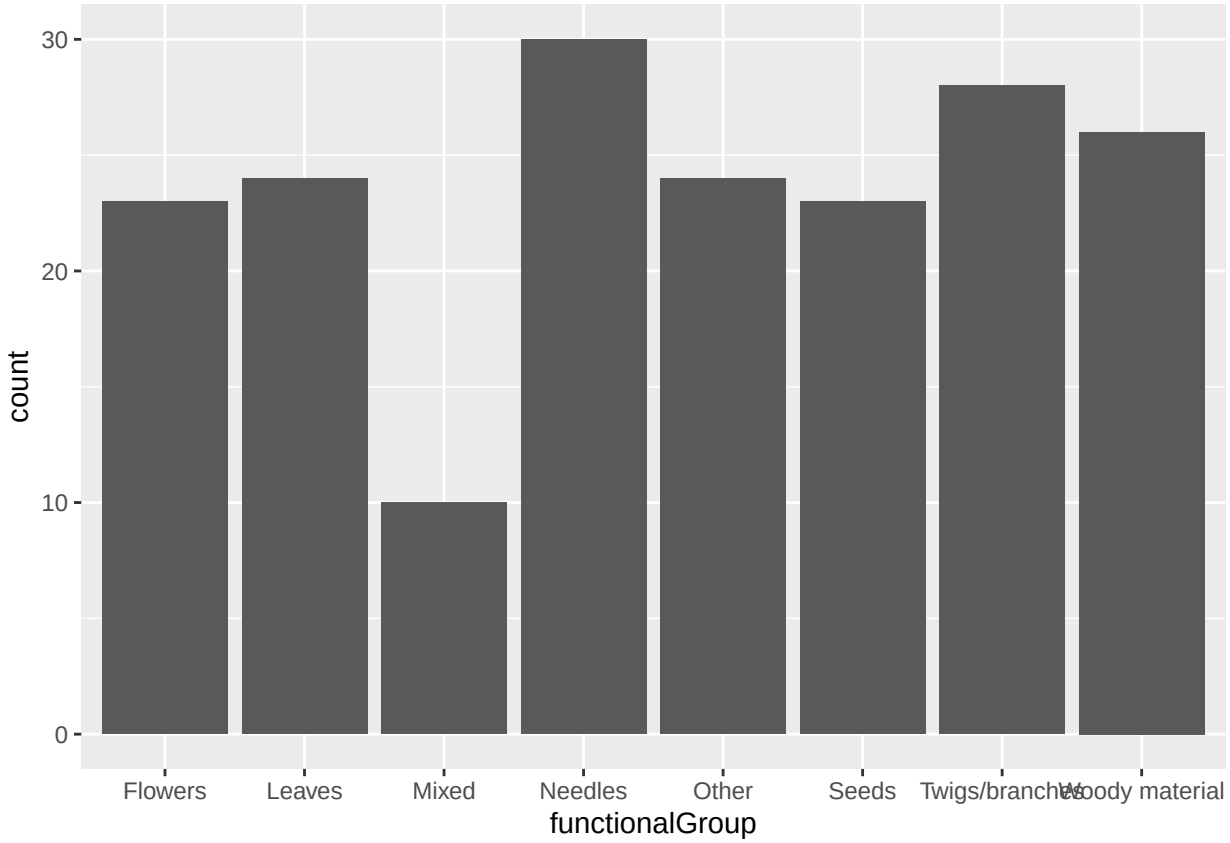
```
##  [1] NIWO_061.basePlot.ltr NIWO_064.basePlot.ltr NIWO_067.basePlot.ltr
##  [4] NIWO_040.basePlot.ltr NIWO_041.basePlot.ltr NIWO_063.basePlot.ltr
##  [7] NIWO_047.basePlot.ltr NIWO_051.basePlot.ltr NIWO_058.basePlot.ltr
## [10] NIWO_046.basePlot.ltr NIWO_062.basePlot.ltr NIWO_057.basePlot.ltr
## 12 Levels: NIWO_040.basePlot.ltr ... NIWO_067.basePlot.ltr
```

```r
# From this function, it has been determined that 12 unique plots were sampled
```

Answer: The unqiue () function is able to return a vector, data frame, etc but removes duplicate features, returning unqiue object. The summary () function provides information of the raw date. It provides the returns the minimum, maximum, mean, median, etc. of the vector. The summary () function counts the number of times a value appears in a certain category.

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.

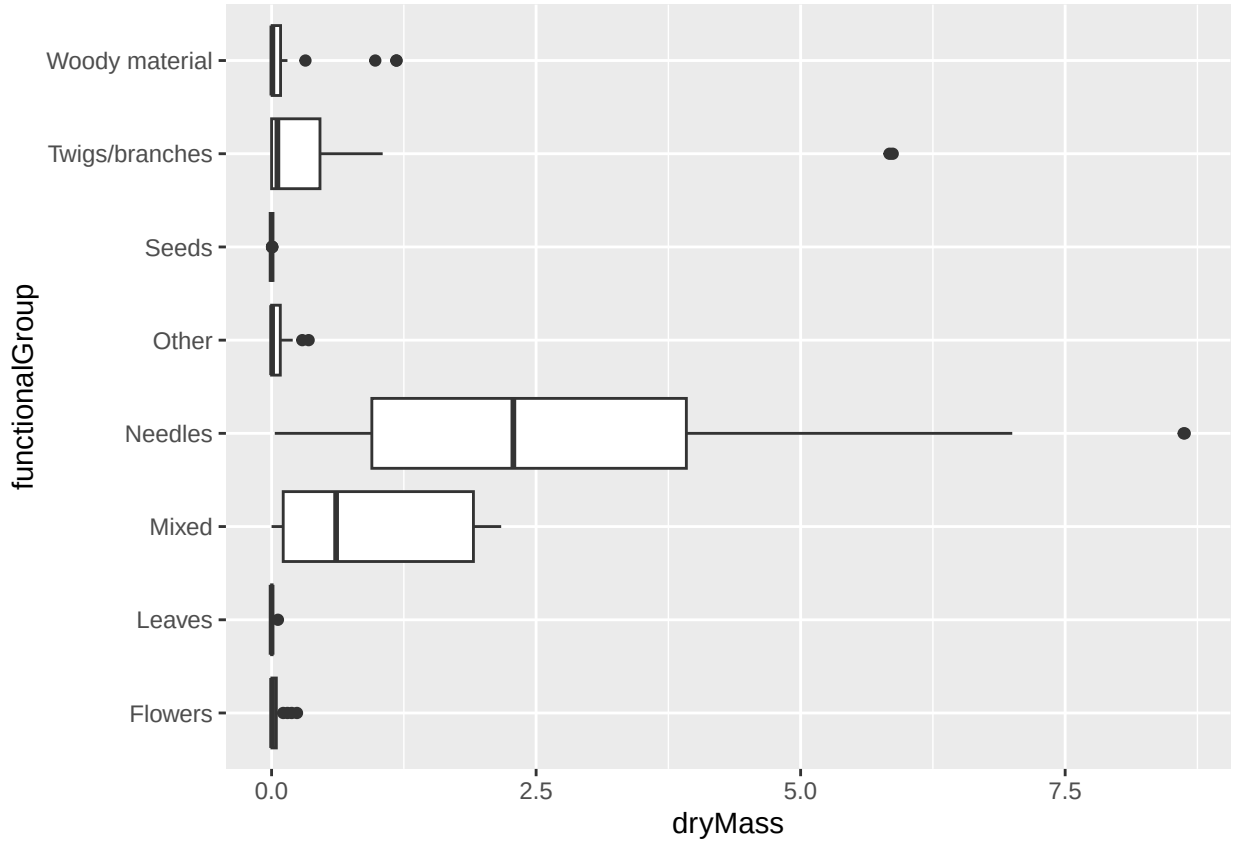
```r
# # Using ggplot and then geom-bar to create a bar graph of functional Group count
ggplot(Litter, aes(x = functionalGroup)) +
  geom_bar()
```



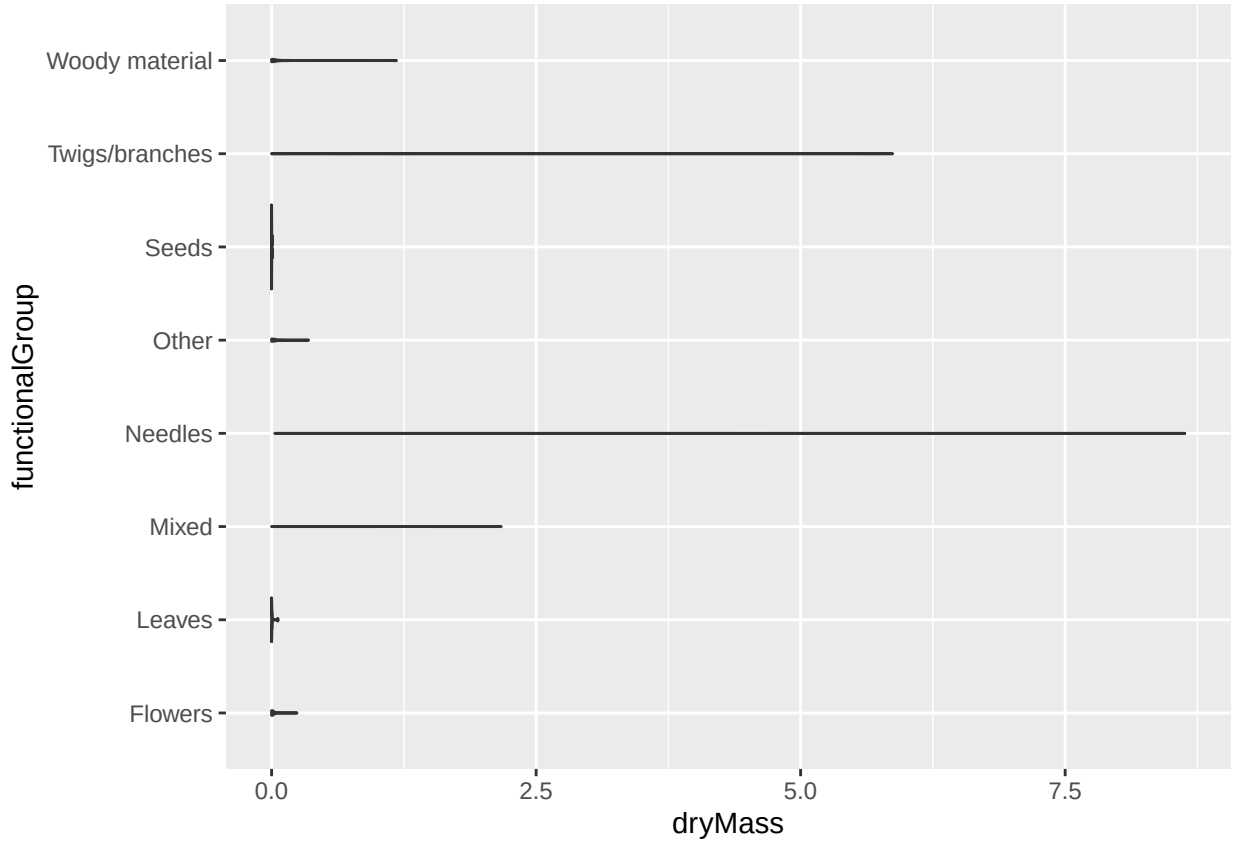
15. Using `geom_boxplot` and `geom_violin`, create a boxplot and a violin plot of dryMass by functionalGroup.

```
# Using the geomboxplot and geom_violing functions to create plots

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




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



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

> Answer: The boxplot provides a better visual compared to the violin plot as it demonstrates the minimum, median, maximum. Additionally, it shows the interquartile range so it better demonstrates how close the dry mass weight was in distribution. The boxplot also showed outliers in the data. The violin plot just demonstrates the range of the dry mass.

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

> Answer: Needles and twigs/branches showed to have the highest biomass.