Assignment 3: Data Exploration

Laura Exar

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.

```
getwd()
```

```
## [1] "/home/guest/R/EDA-Spring2023"
```

```
install.packages("tidyverse")  #installed tidyverse
install.packages("lubridate")  #installed lubridate

library(tidyverse)
library(lubridate)
```

```
Neonics <- read.csv("./Data/Raw/ECOTOX_Neonicotinoids_Insects_raw.csv", stringsAsFactors = TRUE)
# loaded and renamed the Neonics dataset
Litter <- read.csv("./Data/Raw/NEON_NIWO_Litter_massdata_2018-08_raw.csv", stringsAsFactors = TRUE)
# loaded and renamed the Litter dataset
```

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: Neonicitinoids are very toxic to pollinators and other insects and have been linked to a die off of insetc populations. Humans are particularly interested in studying these effects because can damage food webs and decrease biodiversity. Pollinators are also vital to agricultural production and plant growth, so declines in the populations are very concerning.

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 are ecologically important for insect populations because they provide habitat for many insect species that live in the forest. Because many species use litter and woody debris as habitat, selecting sample sites that have these features is a good choice for studying insect populations.

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. The litter and woody debris used are collected from elevated and ground traps. 2. Sampling of litter and woody debris is conducted at NEON sites that contain woody vegetation greater than 2 meters tall. 3. Depending on the vegetation, the trap placements within the plots were either targeted or random.

Obtain basic summaries of your data (Neonics)

5. What are the dimensions of the dataset?

The dataset has 4623 rows and 30 columns.

```
dim(Neonics)  #found the dimensions of the Neonics dataset
```

```
## [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)  #found the summary of the Effect column
```

```
##     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 effect is population (1803) and mortality (1493). These effects likely specifically of interest because they are related to the population collapse of the pollinators, which has significant impacts on agricultural production and plant growth–two things which researchers would be interested in studying.

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...]

```
species <- summary(Neonics$Species.Common.Name)
sort(species, 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
```

```
# found the summary of the most commonly studied species, sorted by decreasing
```

Answer: The six most common studies species, excluding the "other" category, are Honey Bee, Parasitic Wasp, Buff Tailed Bumblebee, Carniolan Honey Bee, Bumble Bee, Italian Honeybee. They are all types of bees/wasps, and they are all important pollinators. The fact that they are pollinators could be why they are being studied over other species, because pollinators are vital to both the ecosystem and agricultural production, and they are experiencing such extreme die-offs, particularly from neonicotinoids.

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"
```

```
# found the class of the concentration column
```

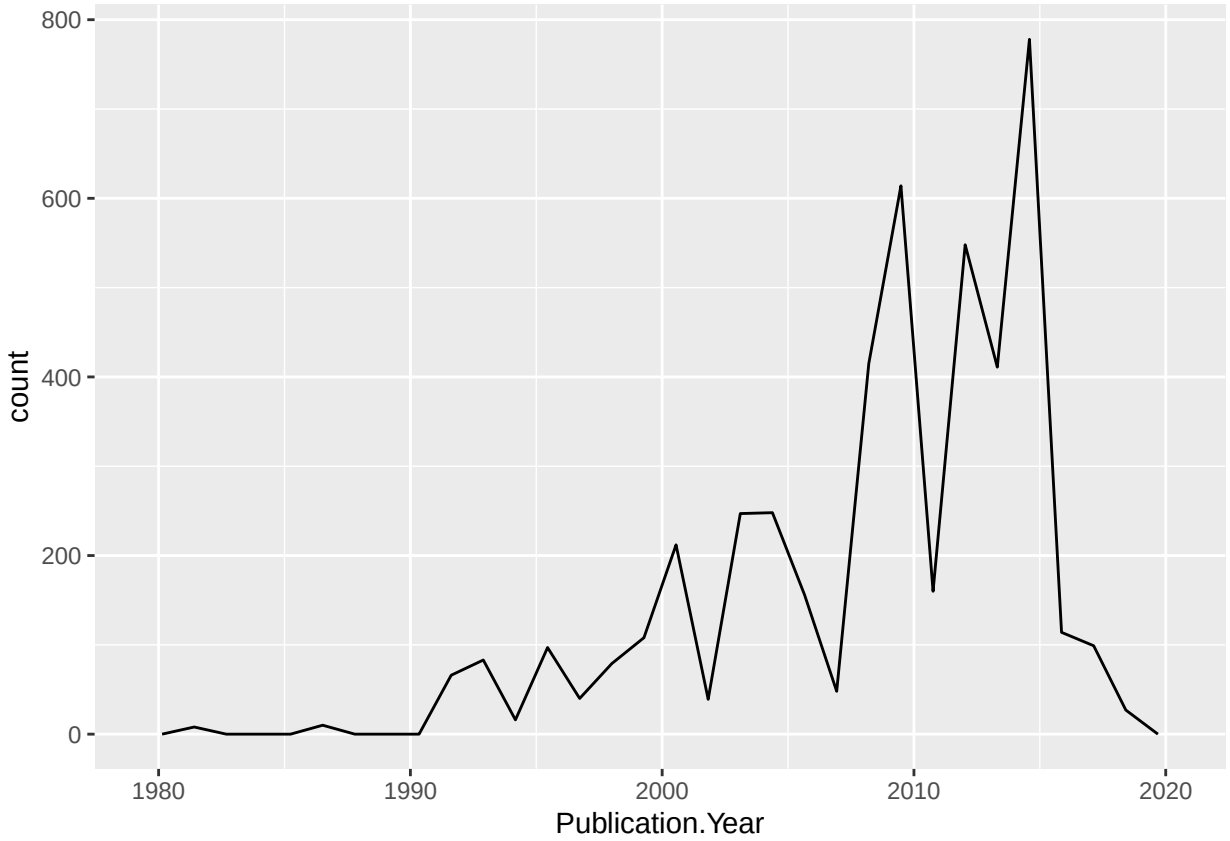
Answer: The 'Conc.1..Author' column is a factor. It is not numeric because some of the values have / in it, rather than the values just having numbers.

Explore your data graphically (Neonics)

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

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

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

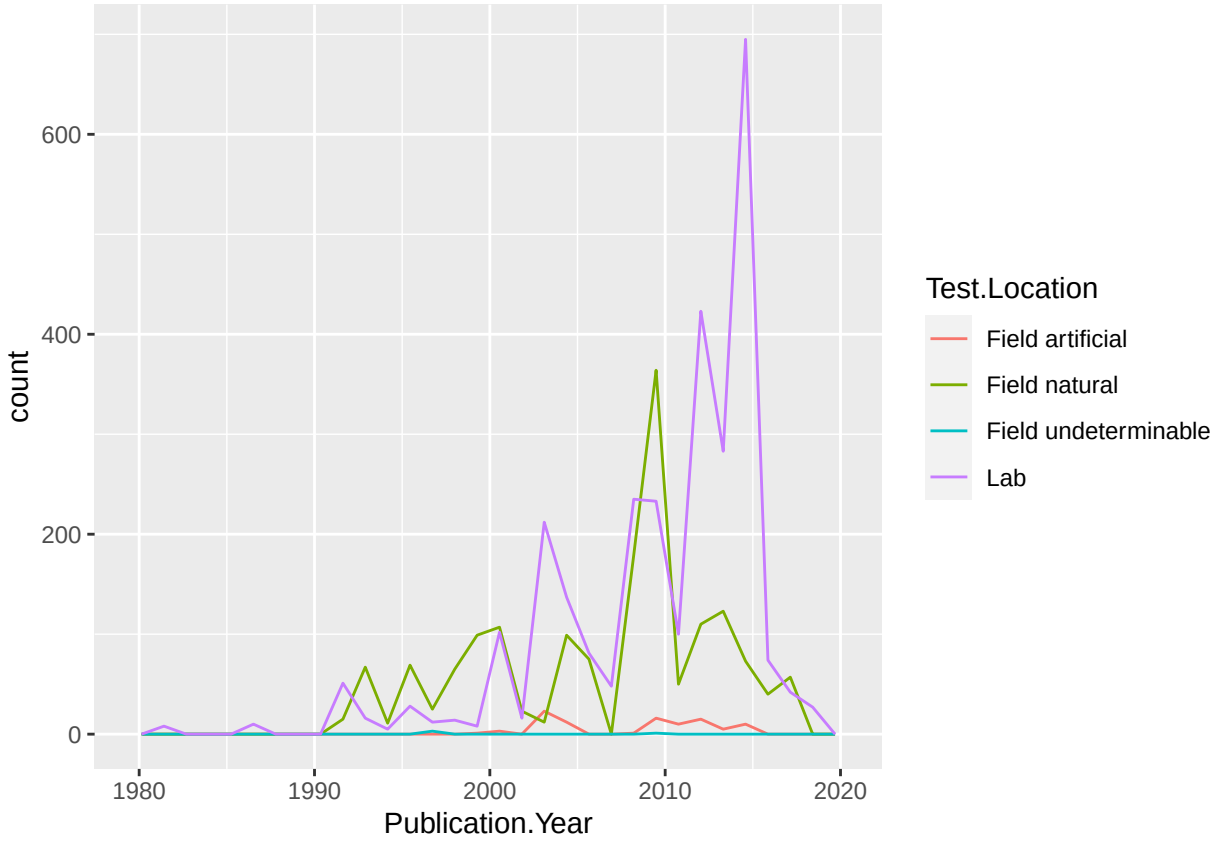



```
# created a plot of the publications produced each year
```

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`.
```




```
# created a plot of the publications produced each year, with different colors
# of test locations
```

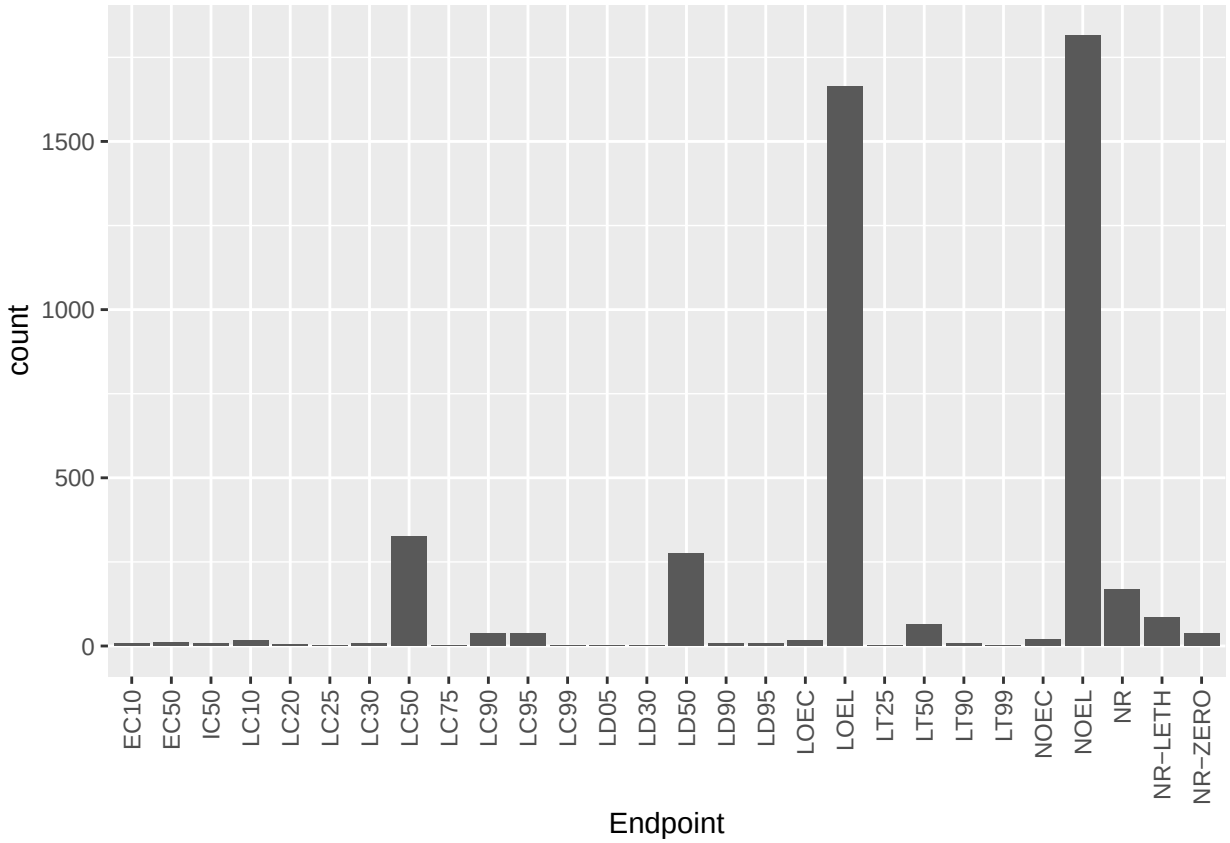
Interpret this graph. What are the most common test locations, and do they differ over time?

> Answer: Overall, the most common test location appears to be the lab, however, there is a spike in field natural test locations over lab right before 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, aes(x = Endpoint)) + geom_bar() + theme(axis.text.x = element_text(angle = 90,
    vjust = 0.5, hjust = 1))  #created a bar graph of endpoint counts
```



Answer: The two most common end points are NOEL and LOEL. NOEL, or "No-observable-effect-level," is defined as the "highest dose (concentration) producing effects not significantly different from responses of controls according to author's reported statistical test." LOEL, or the "Lowest-observable-effect-level," is defined as the "lowest dose (concentration) producing effects that were significantly different (as reported by authors) from responses of controls."

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.

The class of collectDate was not initally a date. The dates litter was sampled are the 2nd and the 30th of August 2018.

```
class(Litter$collectDate)
```

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

```
# determined the class of collectDate column
Litter$collectDate <- as.Date(Litter$collectDate)
# changed the class to a date
unique(Litter$collectDate)
```

```
## [1] "2018-08-02" "2018-08-30"
```

```
# determined which dates litter was sampled in August 2018
```

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`?

```
summary(Litter$plotID)  #summary of plots that were sampled
```

```
## 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
```

```
unique(Litter$plotID)  #determined the plots that were sampled
```

```
##  [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
```

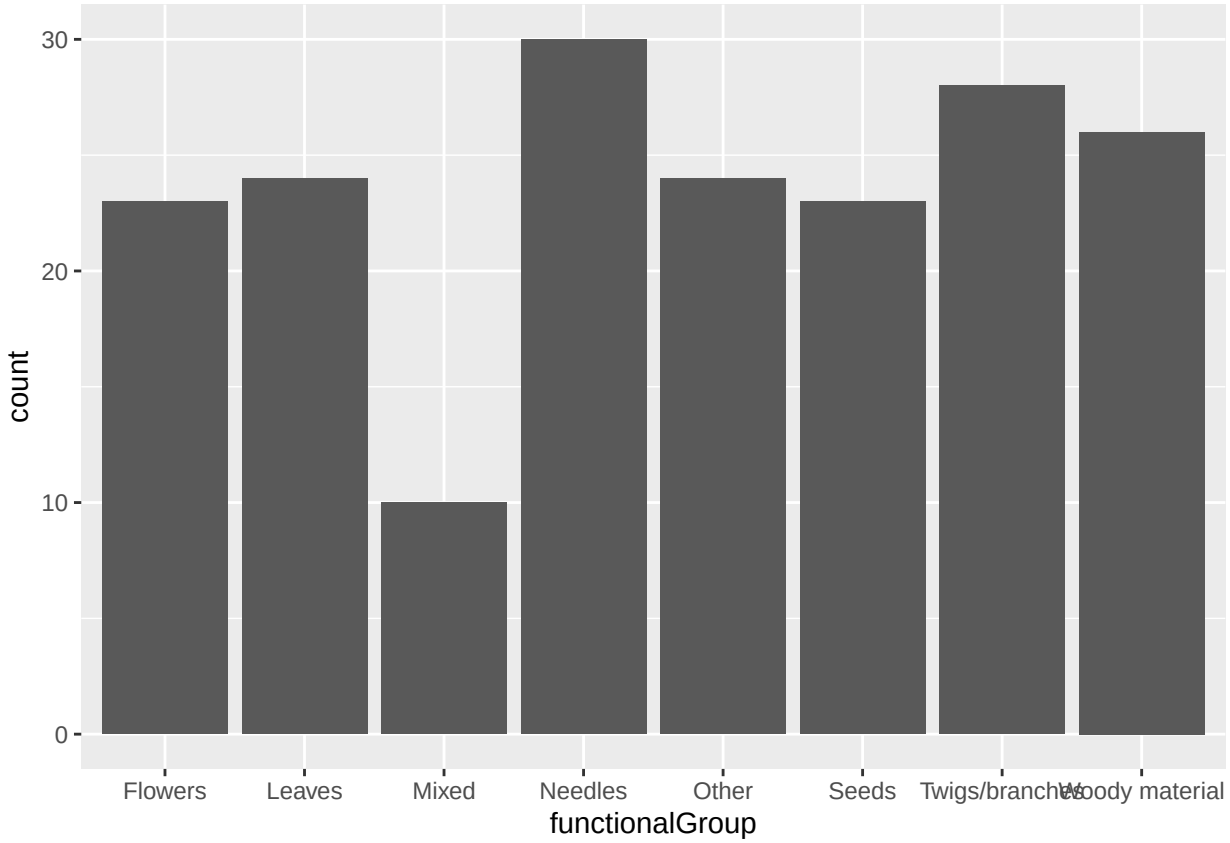
```
length(unique(Litter$plotID))  #found how many plots were sampled
```

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

Answer: 12 plots were sampled. Unique tells us how many of each plot there are. Summary tells us how many of each Plot ID exist in our dataset.

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.

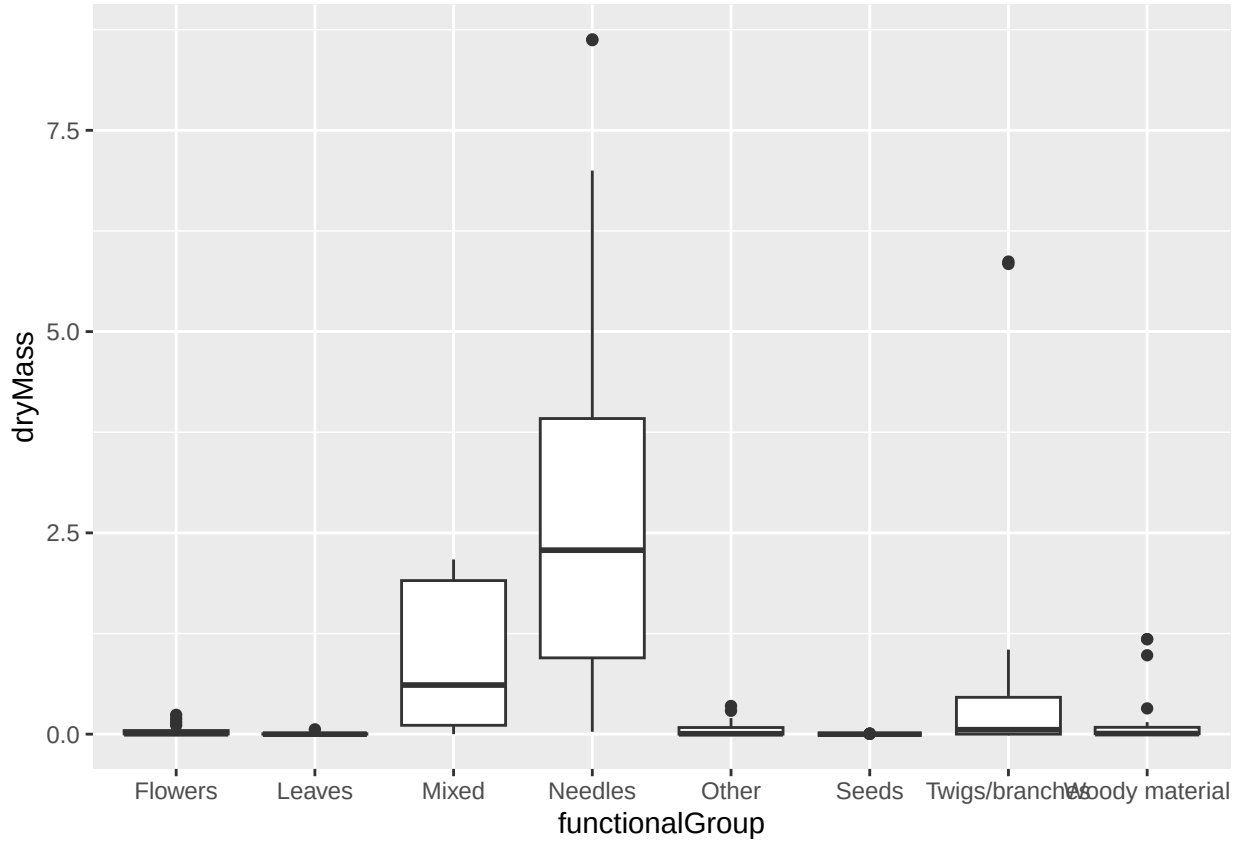
```
ggplot(Litter, aes(x = functionalGroup)) + geom_bar()
```




```
# created a bar graph of functionalGroups
```

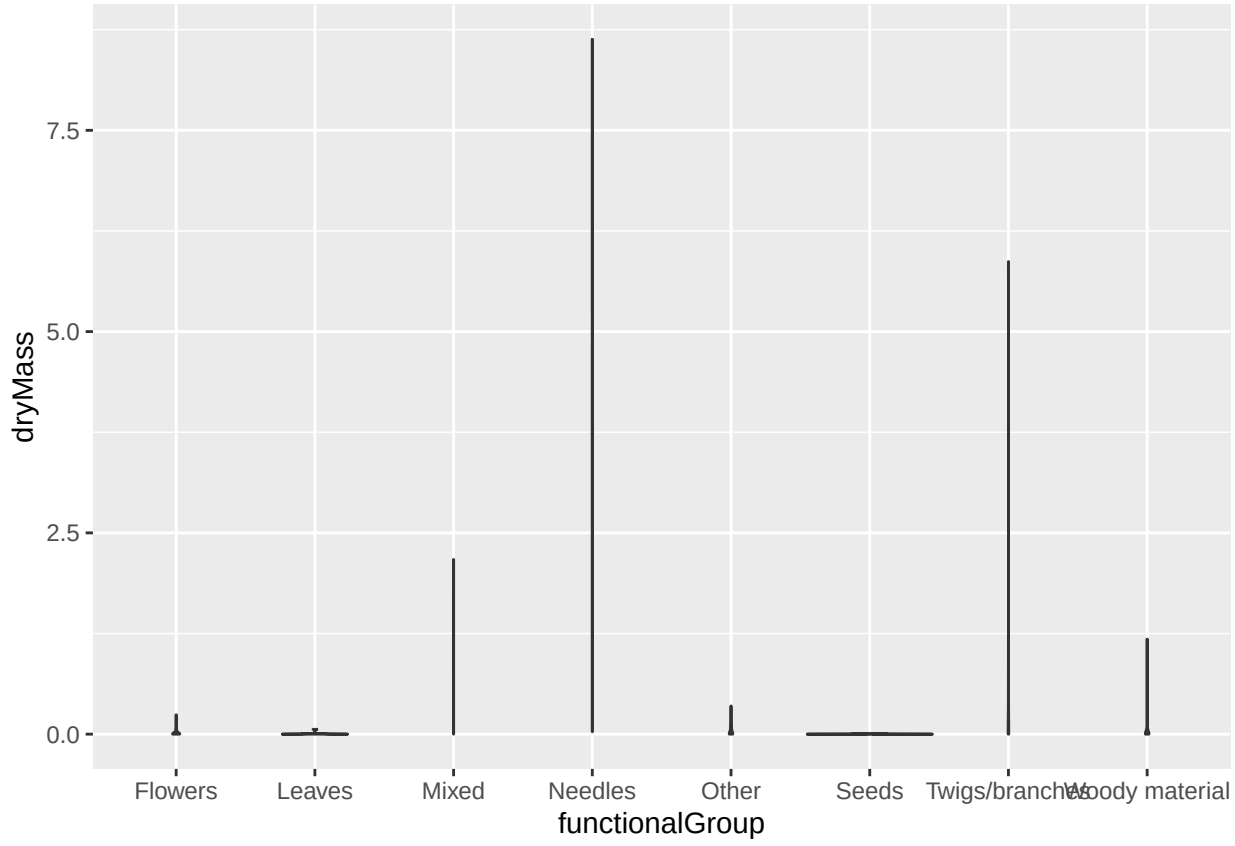
15. Using `geom_boxplot` and `geom_violin`, create a boxplot and a violin plot of dryMass by functionalGroup.

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




```
# created a boxplot of dryMass by functionalGroup

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




```
# created a violin plot of dryMass by functionalGroup
```

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

> Answer: The boxplot is a more effect visualization option than the violin plot because it shows the min and max values, as well as the distribution of the values by providing the first quartile, third quartile, and mean. The violin plot shows the density of the data at different values, which is not particularly helpful information for this questions.

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

> Answer: Needles have the highest biomass at the sites, followed by mixed litter.