IreneChang_A03_DataExploration.Rmd

Irene Chang

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/EDA/EDA-Spring2023"
```

```
library(tidyverse)
library (lubridate)
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: You might want to look at how effective neonicotinoids are on insects at different stages of their life. Did the use of a neonicotinoid lead to the mortality of a specific class on insects? What was the toxcicity of the chemical used? A person may also be interest in whether or not they have negative externalities on the environment, ie. soil quality, but this was not studied in this dataset.

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: You would be interested in studying litter and woody debris that falls on the ground to better understand the chemical makeup (such as the carbon and nitrogen concentrations) of the debris over time and how changes may affect decomposition and plant productivity. Increased nitrogen in debris can potentially be associated with soil carbon sequesteration.

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. Clear definition of litter and woody debris. For example, litter is defined as "dropped from the forest canopy and has a butt end diameter <2cm and a length of <50cm". 2. Each sample is sorted by functional group. 3. Then each sample is dried and weighed.

Obtain basic summaries of your data (Neonics)

5. What are the dimensions of the dataset?

```
dim(Litter) #188 rows and 19 columns
```

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

```
dim(Neonics) #4623 rows and 30 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?

```
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:The most common effects that are studied are mortality (1493 observations), population (1803), and behavior (360). These may be of interest to understand what effect the insecticide had on the mortality, population, and behavior of insects. Changes in these effects are informative to the effectiveness of the 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…]

```
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 6 most commonly studied species are: Other (670), Honey Bee (667), Parasitic Wasp (285), Buff Tailed Bumblebee (183), Carniolan Honey Bee (152), Bumble Bee (140), and Italian Honeybee (113) if not including other. These species all come from the same Hymenoptera family. They may be of interest over other insects to see if the effects of insecticides differentiate among different bee species. It may also be that they would like to specifically target this family with the insecticide, so they are testing the effects it has within the hymenoptera family.

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"
```

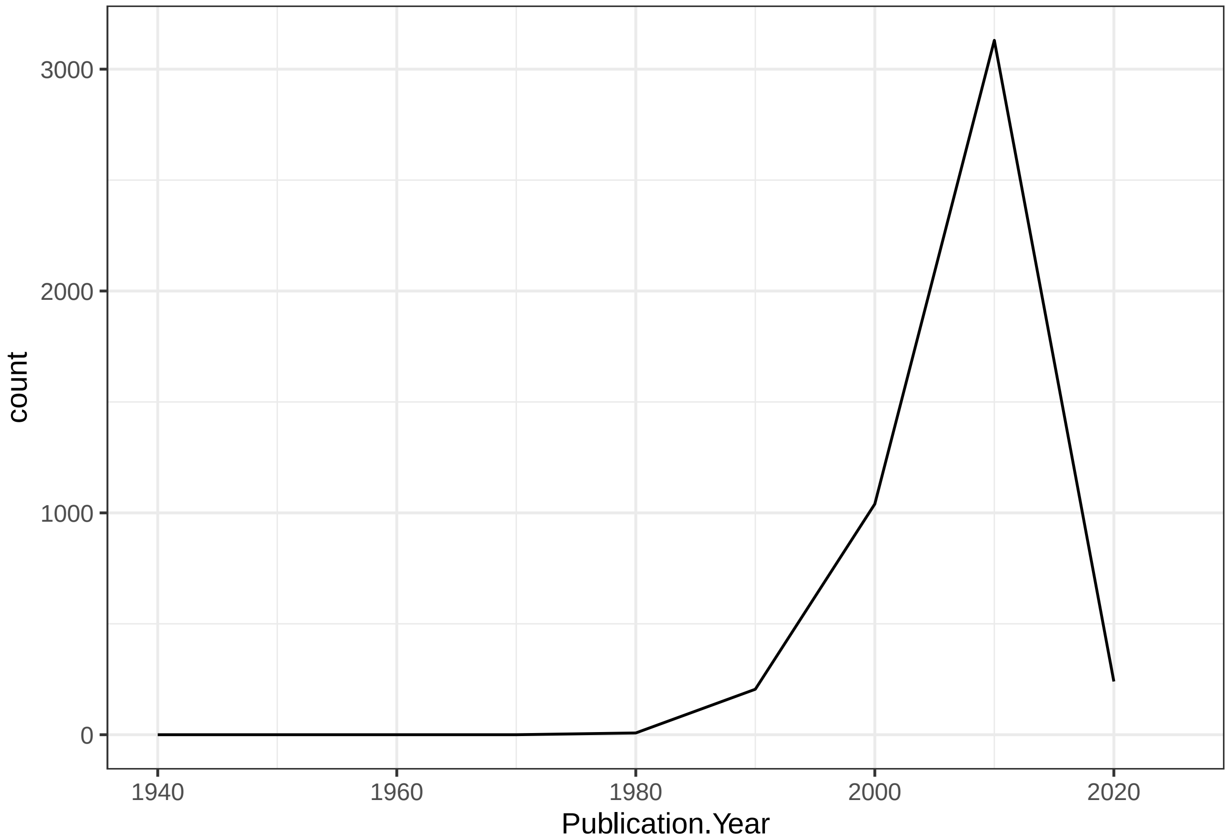
Answer:The class is a factor. These can be used to represent categorical data and are stored as integers, which makes them easy for statistical analysis and plotting.

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(binwidth=10)+
  xlim(c(1940,2025))+
  theme_bw()
```

```
## Warning: Removed 2 rows containing missing values (‘geom_path()‘).
```

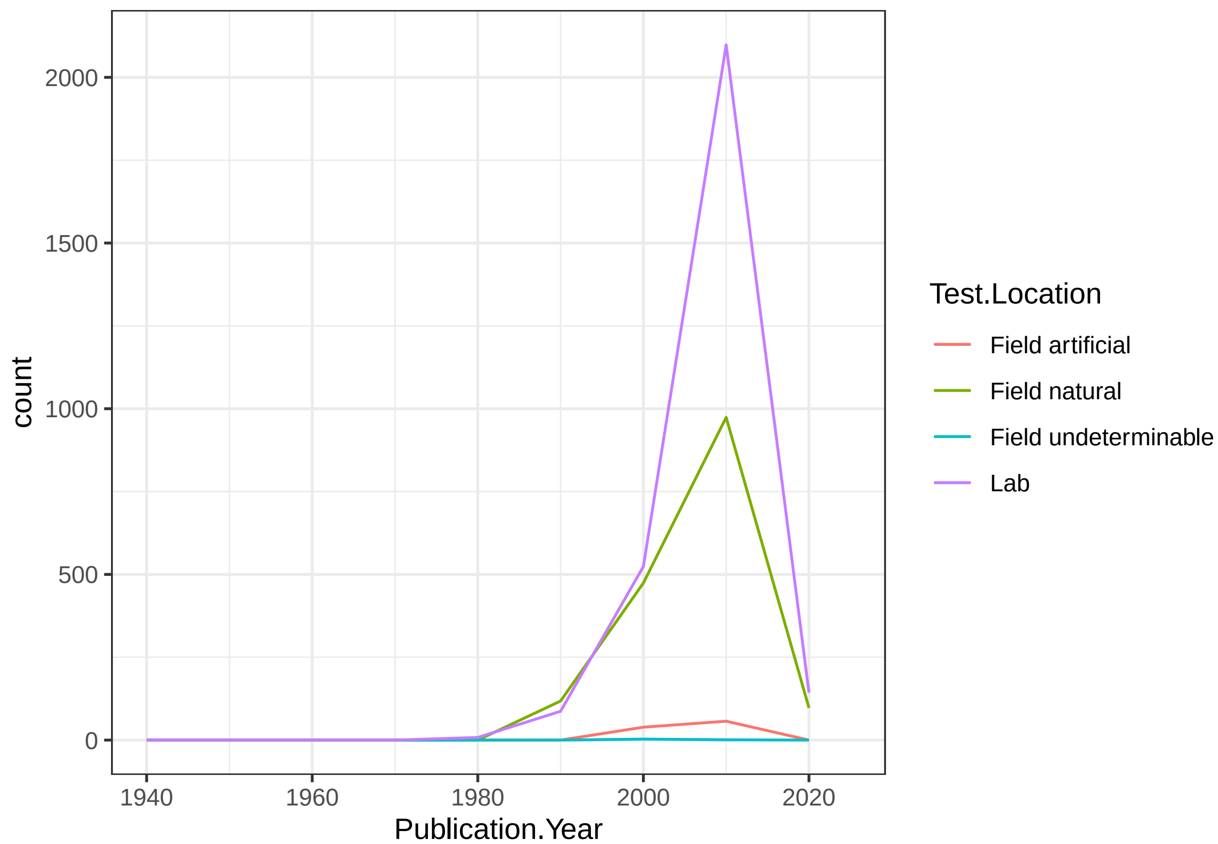



```
#if you have a smaller binwidth, the buckets of data that you are putting it in is smaller
#set the limits by data and making it into a vector
#theme_bw() makes it so that the color scheme of the chart is black and white
```

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

```
ggplot(Neonics, aes(x=Publication.Year, color=Test.Location)) +
  geom_freqpoly(binwidth=10)+
  xlim(c(1940,2025))+
  theme_bw()
```

```
## Warning: Removed 8 rows containing missing values (‘geom_path()‘).
```



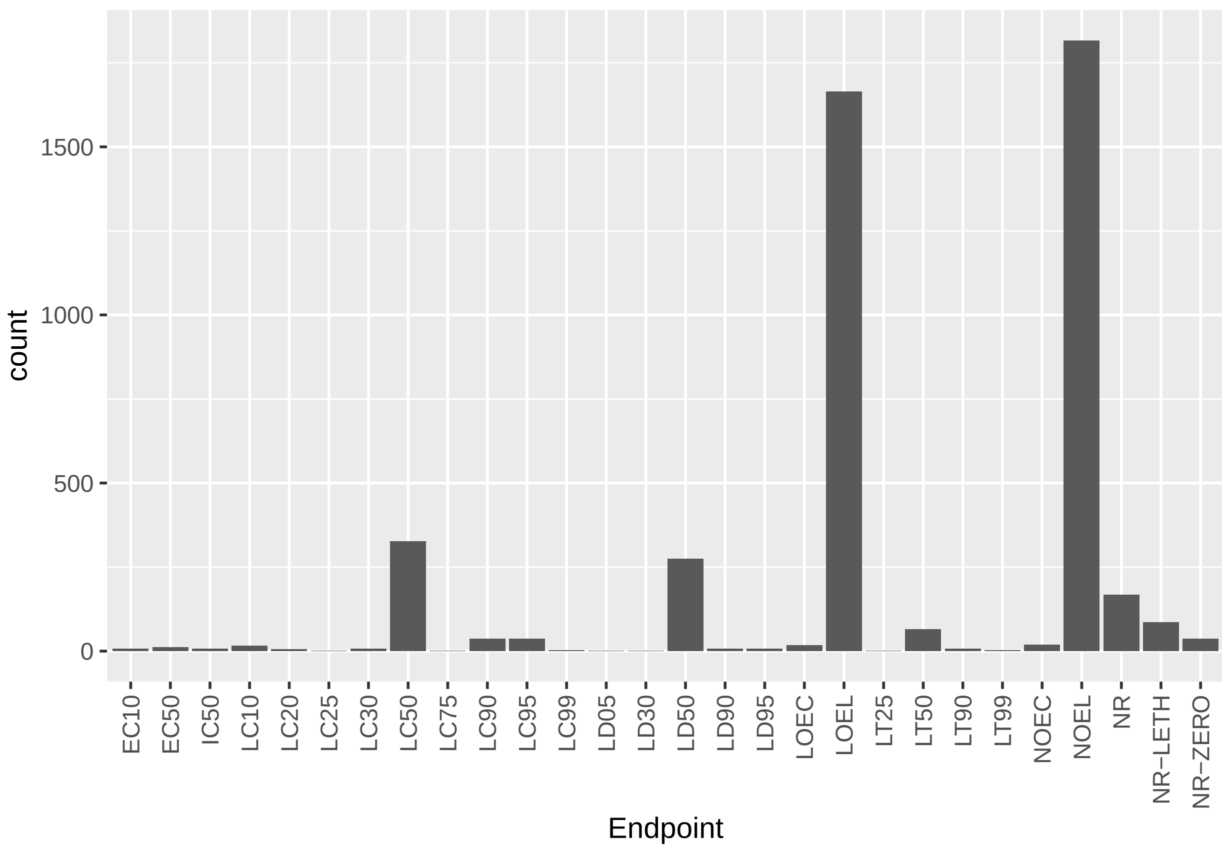
Interpret this graph. What are the most common test locations, and do they differ over time?

> Answer: The most common is lab, though there was a brief time between 1980 to 1995 where Field natural was more common. The number of observations goes up drastically around 1995, and the lab test location becomes extremely common and field natural rises, but not to the extent that lab does.

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: The most common end points are NOEL and LOEL. For LOEL, the database use was terrestrial and is the lowest-observable effect level: on the lowest dose producing effects that were significantly different from the response of controls. For NOEL, the database was terrestrial and is the was the no observable effect level: the highest dose producing effects not significantly different from responses 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.

```
class(Litter$collectDate) #this is a factor.
```

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

```
CD.Collection.Date <- ymd(Litter$collectDate)
class(CD.Collection.Date)
```

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

```
unique(Litter$collectDate)
```

```
## [1] 2018-08-02 2018-08-30
## Levels: 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`?

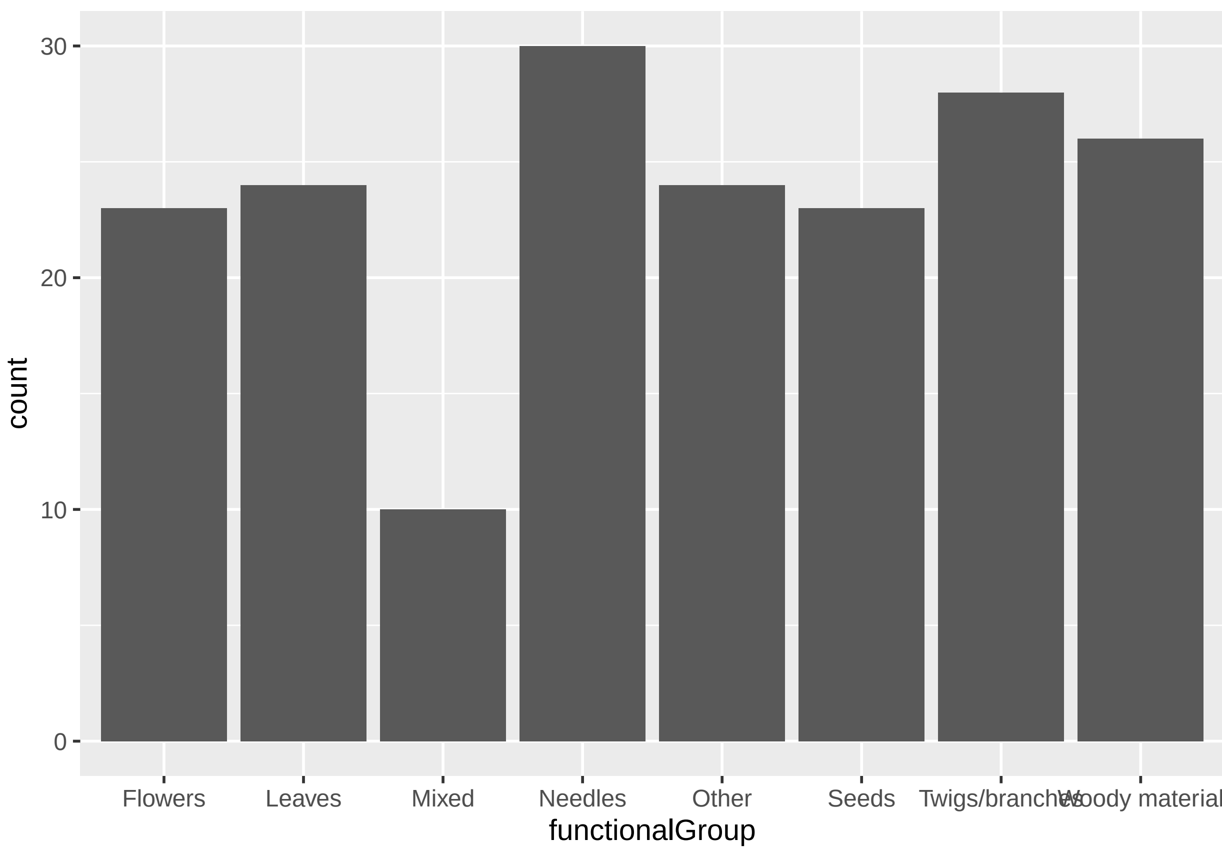
```
unique(Litter$plotID)
```

```
##  [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: The unique function will get rid of duplicates in the data.

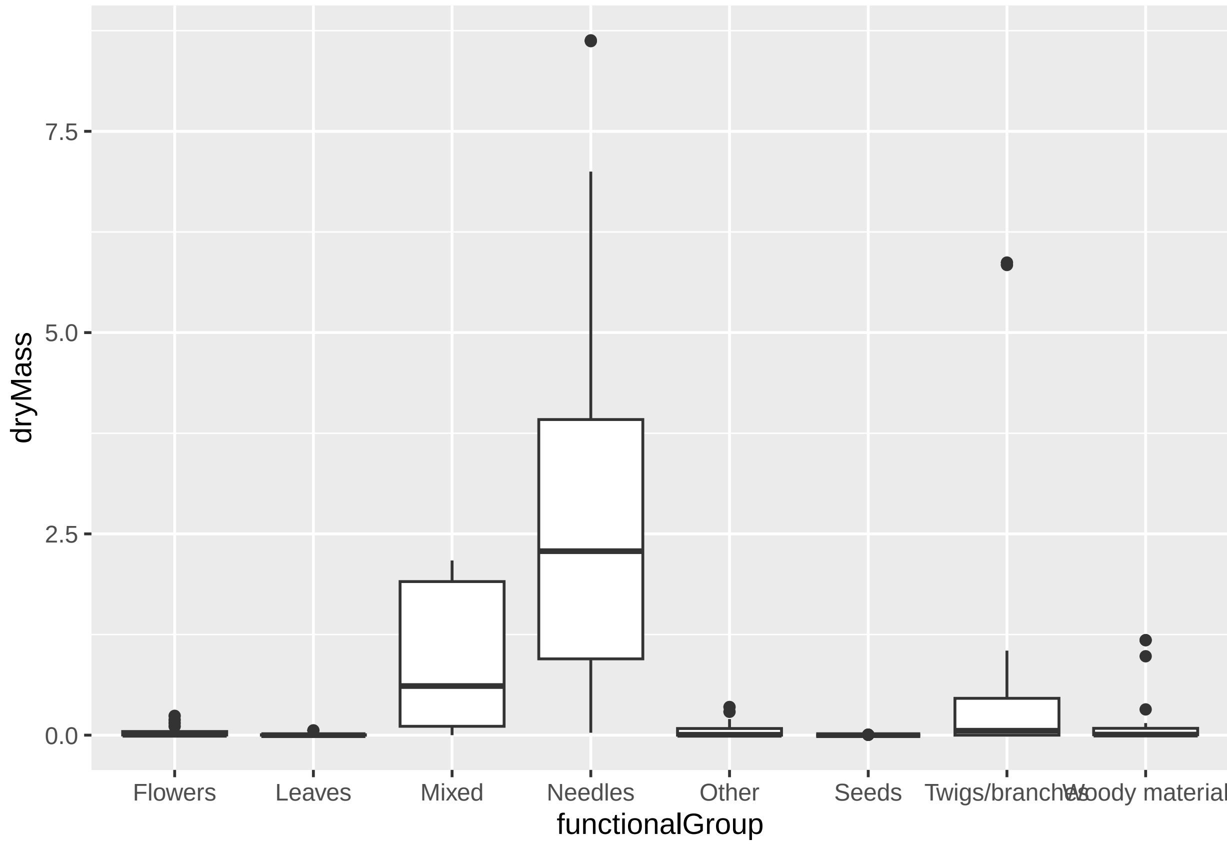
14. Create a bar graph of functionalGroup counts. This shows you what type of litter is collected at the Niwot Ridge sites. Notice that litter types are fairly equally distributed across the Niwot Ridge sites.

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

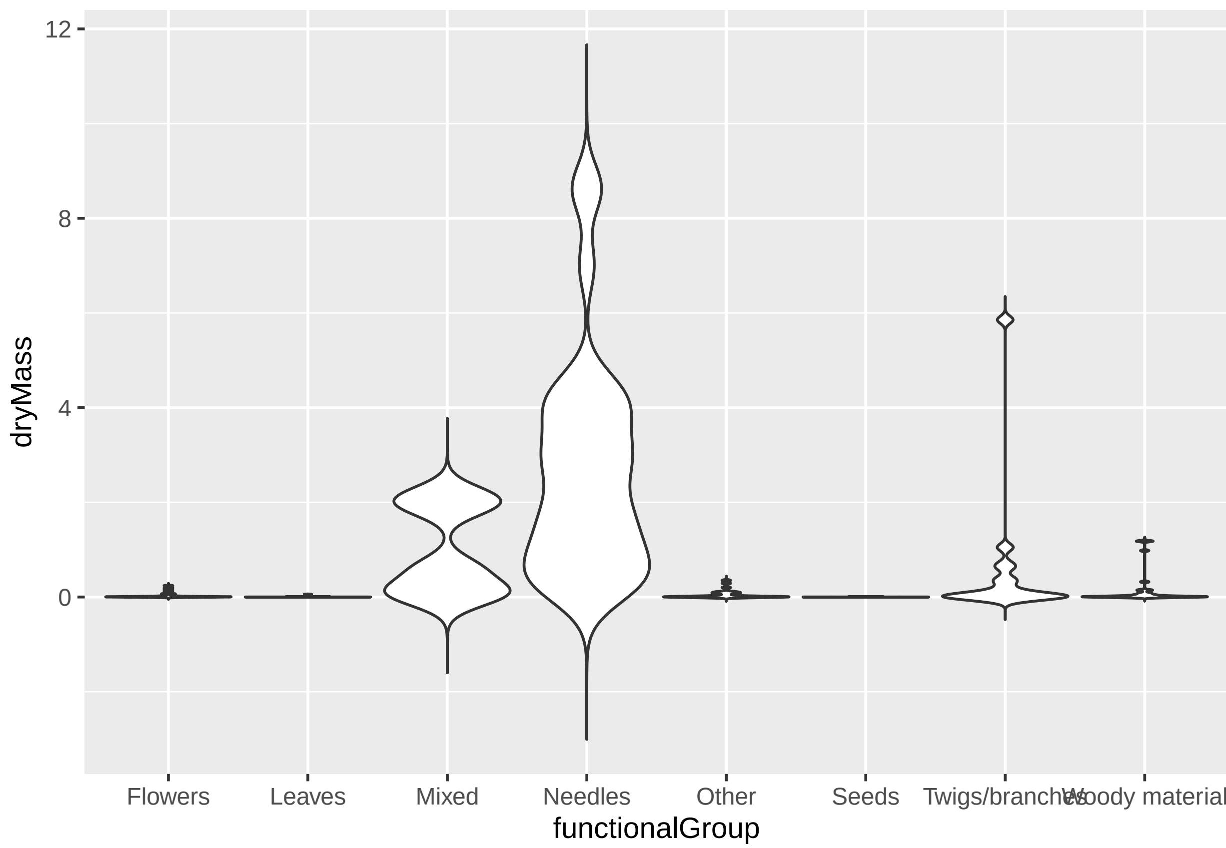


15. Using `geom_boxplot` and `geom_violin`, create a boxplot and a violin plot of dryMass by functionalGroup.

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




```r
ggplot(Litter, aes(x=functionalGroup, y=dryMass)) +
geom_violin(scale = "width", trim = FALSE, adjust = 0.5)
```



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

> Answer: The boxplot is more effective because it gives a more clear picture of the summary statistics. In general for the violin plot, you need enough data for it to be effective and unless setting parameters for the violin plot, there is little data you can obtain from first glance.

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

> Answer: Needles

#check output