Assignment03

Zhiteng Ma

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. Be sure to **answer the questions** in this assignment document.
5. When you have completed the assignment, **Knit** the text and code into a single PDF file.
6. After Knitting, submit the completed exercise (PDF file) to the dropbox in Sakai.

The completed exercise is due on Sept 30th.

Set up your R session

1. Check your working directory, load necessary packages (tidyverse), 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] "C:/Users/Zhiteng Ma/Desktop"

Neonics<-read.csv('c:/Users/Zhiteng Ma/Desktop/EDA-Fall2022/Data/Raw/EC
OTOX_Neonicotinoids_Insects_raw.csv', stringsAsFactors =TRUE)
Litter<-read.csv('c:/Users/Zhiteng Ma/Desktop/EDA-Fall2022/Data/Raw/NEO
N_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 have been widely used since their introduction in the 1980s. As a typical third-generation neonicotinoid insecticide, it has been registered and used in more than 20 countries worldwide because of its high efficiency and broad spectrum. With the extensive use of neonicotinoid pesticides in recent years, the potential harm to non-target organisms and the environmental risks caused by a large amount of input in agricultural production has attracted people's attention. Therefore, elucidating the biological activity and ecotoxicology of neonicotinoid pesticides and revealing the fate of their environmental behavior is of great significance for the ecological risk assessment and scientific and rational use of neonicotinoid pesticides.

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 that falls to the ground in forests will affect the ecological environment of the forest. The destruction of forest ecosystems may lead to violent typhoons along the coast, flooding, sand and dust weather, accelerated soil erosion, land desertification, reduced freshwater resources necessary for human survival and global warming, accelerated species extinction, and severe damage to natural organisms The balance of the ecosystem such as material exchange and energy flow between the environment and the environment. At the same time, the survival and development of human beings are threatened.

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.Along with most of NEON's plant productvity measurements, sampling for this product occurs only in tower plots (AD[06]). Locatons of tower plots are selected randomly within the 90% flux footprint of the primary and secondary airsheds . In sites with forested tower airsheds, the liter sampling is targeted to take place in 20 40m x 40m plots. 2.In sites with low-statured vegetaton over the tower airsheds, liter sampling is targeted to take place in 4 40m x 40m tower plots plus 26 20m x 20m plots. One liter trap pair is deployed for every 400 m2 plot area, resultng in 1-4 trap pairs per plot. In some cases, available space, plot spacing requirements, and/or the tower airshed size restricts the number of plots that can be sampled for liter below 20 (forested) or 30 (low-stature). 3. Specifically, plot edges must be separated by a distance 150% of one edge of the plot; plot centers must be greater than 50m from large paved roads and plot edges must be 10m from two-track dirt roads; plot centers must be 50m from buildings and other non-NEON infrastructure; streams larger than 1m must not intersect plots.

Obtain basic summaries of your data (Neonics)

5. What are the dimensions of the dataset?

```
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?

```
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:Population are studied most.

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.

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

##                            Honey Bee                            Parasitic Was
p
##                                  667                                       28
5
##                 Buff Tailed Bumblebee                     Carniolan Honey Be
e
##                                  183                                       15
2
##                            Bumble Bee                          Italian Honeybe
e
##                                  140                                       11
3
```

```
##                    Japanese Beetle                   Asian Lady Beetl
e
##                                 94                                  7
6
##                     Euonymus Scale                           Wirewor
m
##                                 75                                  6
9
##                  European Dark Bee                   Minute Pirate Bu
g
##                                 66                                  6
2
##                Asian Citrus Psyllid                       Parastic Was
p
##                                 60                                  5
8
##             Colorado Potato Beetle                    Parasitoid Was
p
##                                 57                                  5
1
##                 Erythrina Gall Wasp                        Beetle Orde
r
##                                 49                                  4
7
##        Snout Beetle Family, Weevil           Sevenspotted Lady Beetl
e
##                                 47                                  4
6
##                     True Bug Order                Buff-tailed Bumblebe
e
##                                 45                                  3
9
##                       Aphid Family                      Cabbage Loope
r
##                                 38                                  3
8
##                Sweetpotato Whitefly                      Braconid Was
p
##                                 37                                  3
3
##                       Cotton Aphid                      Predatory Mit
e
##                                 33                                  3
3
##             Ladybird Beetle Family                          Parasitoi
d
##                                 30                                  3
0
##                       Scarab Beetle                       Spring Tiphi
a
```

```
##                         29                          2
9
##                Thrip Order        Ground Beetle Famil
y
##                         29                          2
7
##          Rove Beetle Family               Tobacco Aphi
d
##                         27                          2
7
##                Chalcid Wasp     Convergent Lady Beetl
e
##                         25                          2
5
##              Stingless Bee            Spider/Mite Clas
s
##                         25                          2
4
##        Tobacco Flea Beetle             Citrus Leafmine
r
##                         24                          2
3
##             Ladybird Beetle                   Mason Be
e
##                         23                          2
2
##                    Mosquito               Argentine An
t
##                         22                          2
1
##                      Beetle Flatheaded Appletree Bore
r
##                         21                          2
0
##       Horned Oak Gall Wasp           Leaf Beetle Famil
y
##                         20                          2
0
##           Potato Leafhopper Tooth-necked Fungus Beetl
e
##                         20                          2
0
##                Codling Moth   Black-spotted Lady Beetl
e
##                         19                          1
8
##                Calico Scale           Fairyfly Parasitoi
d
##                         18                          1
8
```

```
##                              Lady Beetle             Minute Parasitic Wasp
s
##                                       18                                 1
8
##                                Mirid Bug                   Mulberry Pyrali
d
##                                       18                                 1
8
##                                 Silkworm                     Vedalia Beetl
e
##                                       18                                 1
8
##                     Araneoid Spider Order                         Bee Orde
r
##                                       17                                 1
7
##                           Egg Parasitoid                      Insect Clas
s
##                                       17                                 1
7
##                Moth And Butterfly Order       Oystershell Scale Parasitoi
d
##                                       17                                 1
7
## Hemlock Woolly Adelgid Lady Beetle              Hemlock Wooly Adelgi
d
##                                       16                                 1
6
##                                     Mite                       Onion Thri
p
##                                       16                                 1
6
##                   Western Flower Thrips                      Corn Earwor
m
##                                       15                                 1
4
##                        Green Peach Aphid                         House Fl
y
##                                       14                                 1
4
##                                Ox Beetle                 Red Scale Parasit
e
##                                       14                                 1
4
##                      Spined Soldier Bug               Armoured Scale Famil
y
##                                       14                                 1
3
##                         Diamondback Moth                      Eulophid Was
p
```

```
##                           13                                  1
3
##            Monarch Butterfly                       Predatory Bu
g
##                           13                                  1
3
##        Yellow Fever Mosquito                 Braconid Parasitoi
d
##                           13                                  1
2
##                 Common Thrip      Eastern Subterranean Termit
e
##                           12                                  1
2
##                       Jassid                          Mite Orde
r
##                           12                                  1
2
##                    Pea Aphid                     Pond Wolf Spide
r
##                           12                                  1
2
##      Spotless Ladybird Beetle             Glasshouse Potato Was
p
##                           11                                  1
0
##                     Lacewing             Southern House Mosquit
o
##                           10                                  1
0
##       Two Spotted Lady Beetle                          Ant Famil
y
##                           10
9
##                 Apple Maggot                               (Other)

##                            9                                 67
0
```

Answer: 1.Six most commonly studied species in the dataset is Honey Bee, Parasitic Wasp, Buff Tailed Bumblebee, Carniolan Honey Bee, Bumble Bee, Italian Honeybee. Honeybees are the most famous species of Hymenoptera. They can pollen up to 1,000 flowers per day. Honey-producing bees collect pollen from about 300 flowers a day. Bees carry pollen to other plants, which promotes plant reproduction. 80% of all flowering plants are pollinated by insects.

8. Concentrations are always a numeric value. What is the class of Conc.1..Author. in the dataset, and why is it not numeric?

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

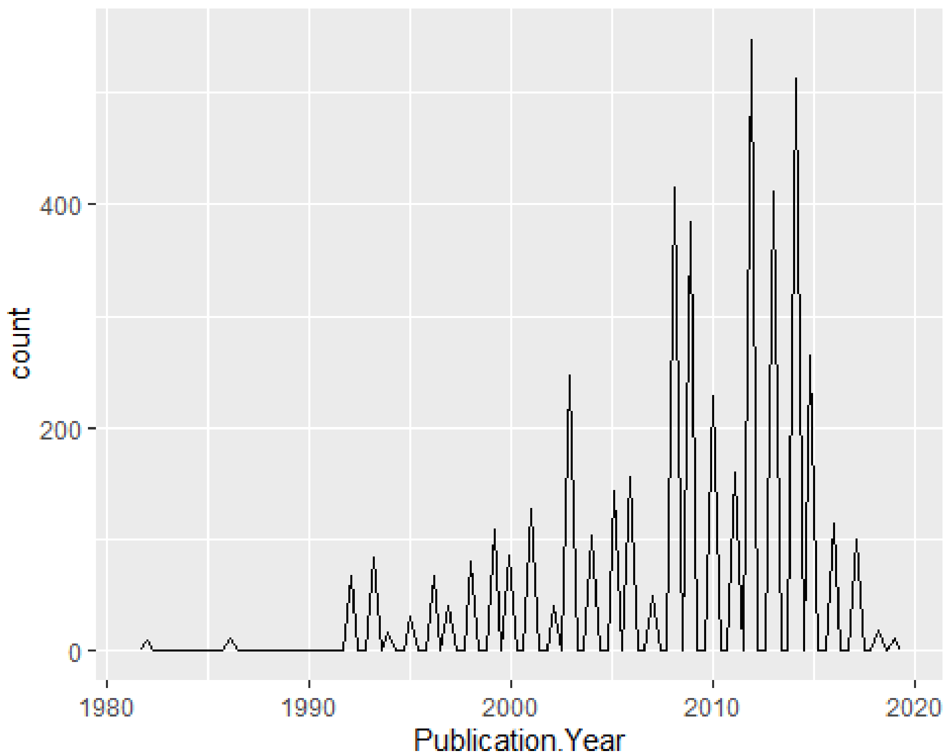
```
## [1] "factor"
```

Answer: It is nor numeric, it is factor. There are N/A in the Conc.1..Author.

Explore your data graphically (Neonics)

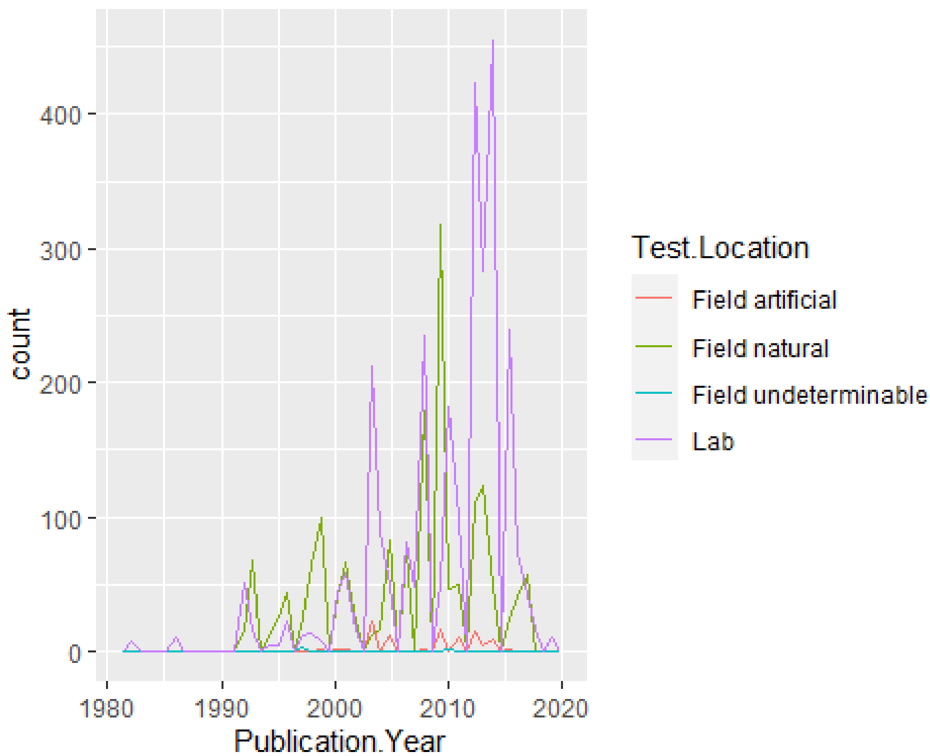
9. Using `geom_freqpoly`, generate a plot of the number of studies conducted by publication year.

```
library(ggplot2)
  ggplot(Neonics) +
  geom_freqpoly(aes(x = Publication.Year), bins = 100)
```



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

```
library(ggplot2)
  ggplot(Neonics) +
  geom_freqpoly(aes(x = Publication.Year,color = Test.Location), bins =
 50)
```

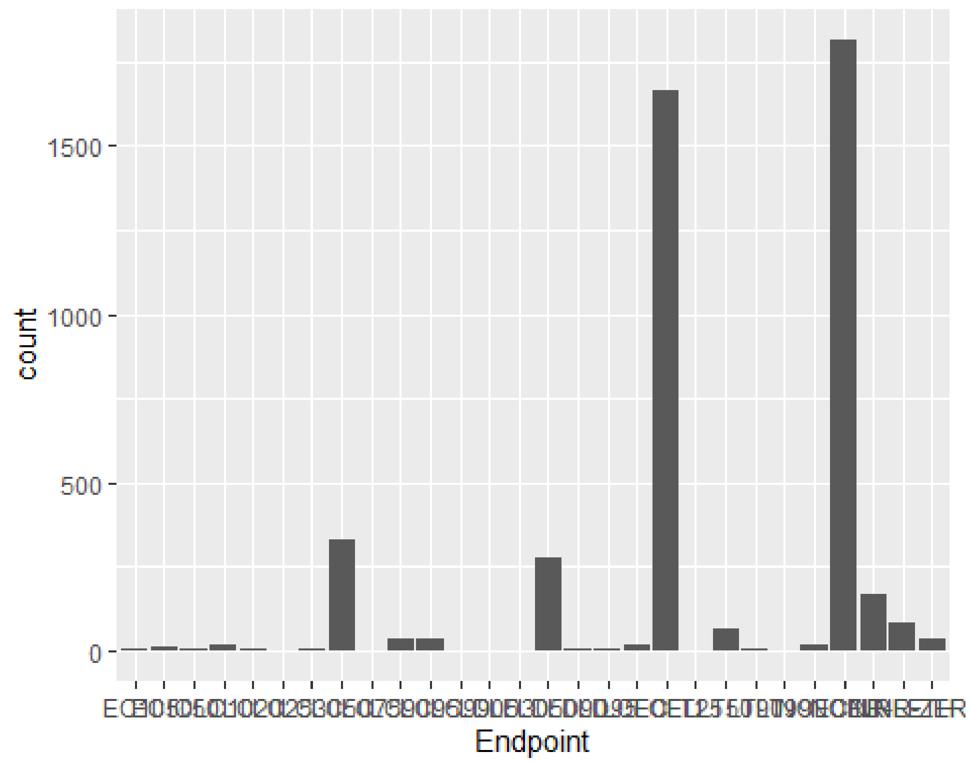


Interpret this graph. What are the most common test locations, and do they differ over time?

Answer:They increased over time before 2014 and then decreased after 2014

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.

```
library(ggplot2)
  ggplot(Neonics, aes(x = Endpoint)) +
  geom_bar()
```



Answer:What are the two most common end points is LOEL and NOEL.

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$collectDate = as.Date(Litter$collectDate)
class(Litter$collectDate)

## [1] "Date"

unique(Litter$collectDate)

## [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`?

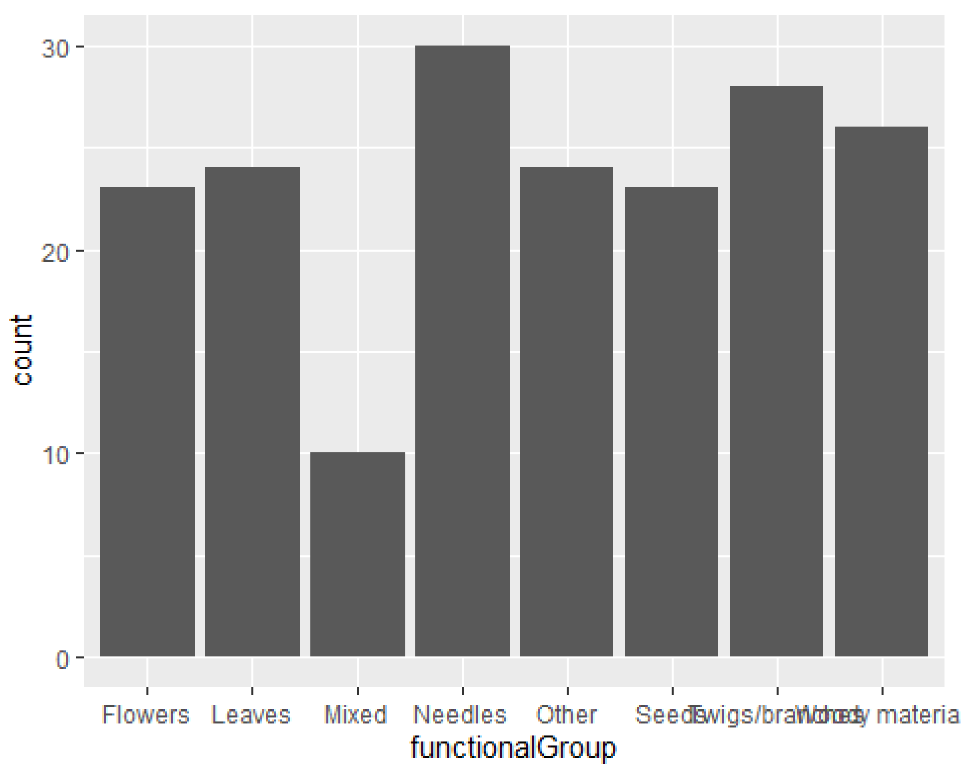
```
unique(Litter$plotID)

##  [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: Unique can display the type of information and the length of the information, but summary is only statistical information.

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.

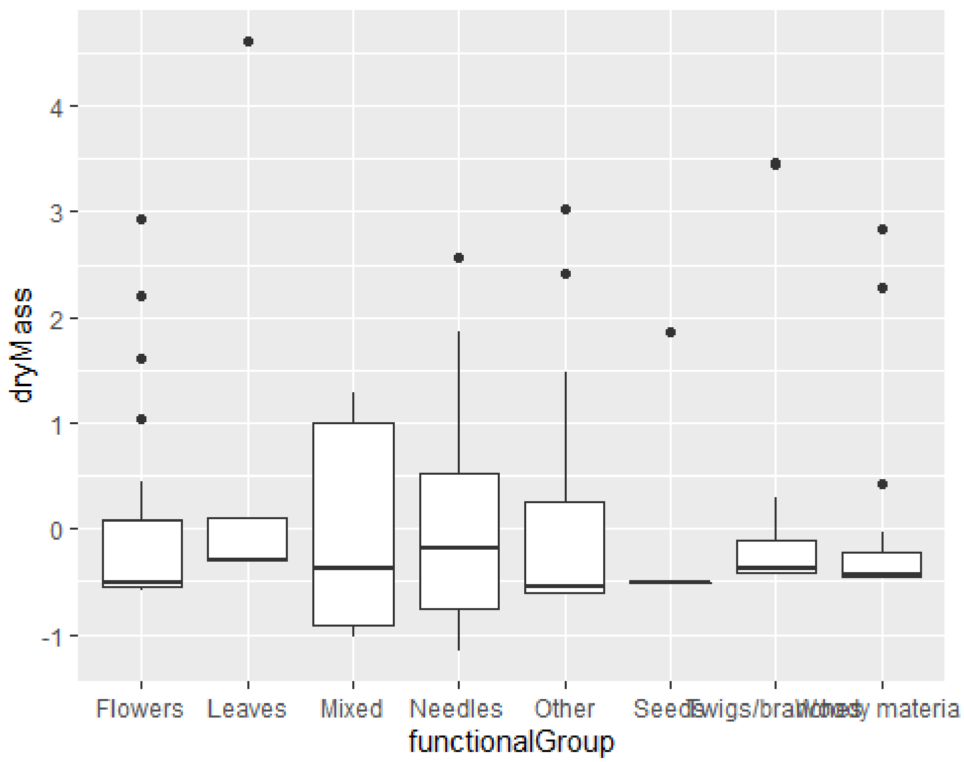
```
library(ggplot2)
ggplot(Litter, aes(x = functionalGroup)) +
  geom_bar()
```



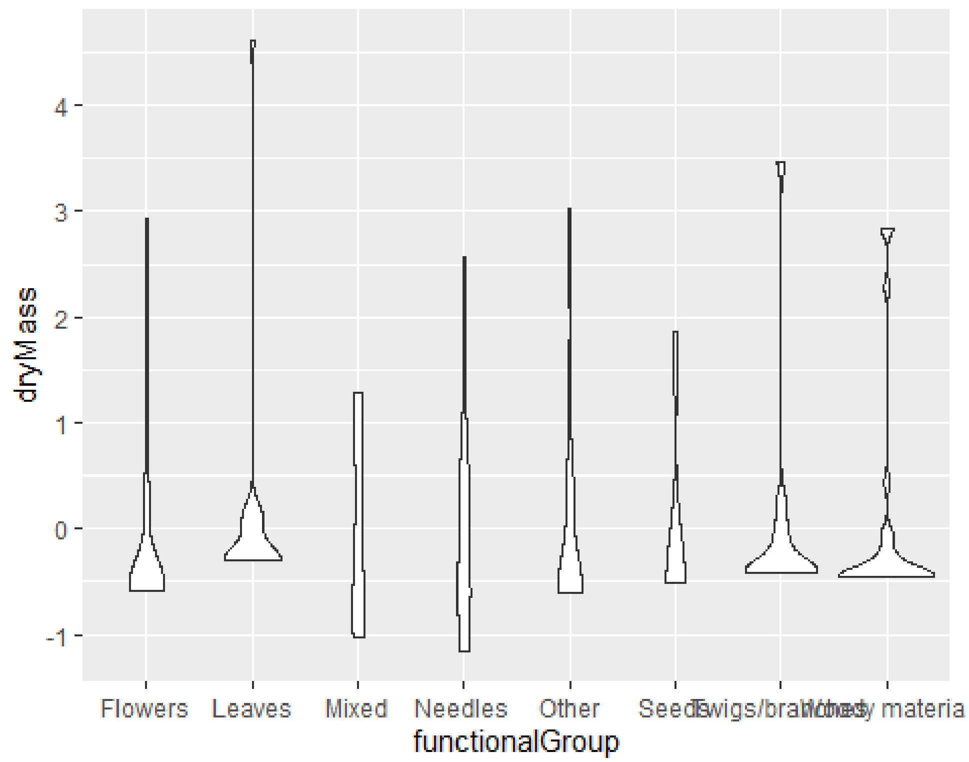
15. Using `geom_boxplot` and `geom_violin`, create a boxplot and a violin plot of dryMass by functionalGroup.

```
library(ggplot2)
for (func in unique(Litter$functionalGroup)){
  idx = Litter$functionalGroup == func
```

```
  data = Litter$dryMass[idx]
  m = mean(data)
  v = sd(data)
  nor_data = (data-m)/v
  Litter$dryMass[idx] = nor_data
}
m = mean(Litter)

## Warning in mean.default(Litter): 参数不是数值也不是逻辑值：回覆 NA

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




```
library(ggplot2)
ggplot(Litter)+
  geom_violin(aes(x =  functionalGroup, y = dryMass))
```



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

> Answer: In this case, boxplots provide summary statistics (boxes and expanded lines) and direct data visualization (outliers). Box plots use commonly used statistics and can provide key information about the location and dispersion of data, especially when comparing different parent data. It's more intuitive than a violin plot.

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

> Answer:Mixed