Assignment 3: Data Exploration

Kaichun Yang

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.

```
# Load packages

library(dplyr)
library(ggplot2)
library(tidyverse)

# Load dataset
Neonics <- read.csv(file = 'E:/EDA-Fall2022/Data/Raw/ECOTOX_Neonicotinoids_Insects_raw.csv')
Litter <- read.csv(file = 'E:/EDA-Fall2022/Data/Raw/NEON_NIWO_Litter_massdata_2018-08_raw.csv')
class(Neonics)
```

```
## [1] "data.frame"
```

```
class(Neonics$Effect)
```

```
## [1] "character"
```

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: Since 1990s, neonicotinoids have become the most widely used insecticides in the world. Their toxicity is less than active ingredients and traditional classes of insecticides. Yet recent research shows neonicotinoids have become widespread environmental contaminants. Study the exotoxicology of neonicotinoids on insects can help people understand the impact of neonicotinoids on environment, agriculture and human health.

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: Study the forest debri can help people understand the component of forest ecosystems and the ecological functions, including nutrient cycling, energy flow and habitat for wildlife.

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.Litter and fine woody debris are collected from the elevated traps or ground traps. 2.All data are collected at the spatial resolution and the temporal resolution of a single event. 3.Litter is defined the material with butt end diameter <2cm and length <50cm, fine wood debris is defined as materials with but end diameter <2cm and length >50cm.

Obtain basic summaries of your data (Neonics)

5. What are the dimensions of the dataset?

```
Neonics_dimension <- dim_desc(Neonics)
Litter_dimension <- dim_desc(Litter)
print (Neonics_dimension)
```

```
## [1] "[4,623 x 30]"
```

```
print (Litter_dimension)
```

```
## [1] "[188 x 19]"
```

```
#The dimension of dataset 'Litter' and ' Neonics' is [188 x 19] and [4623 x 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?

```
Neonics$Effect <- as.factor(Neonics$Effect)
class(Neonics$Effect)
```

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

```
N_S <- summary(Neonics$Effect)
sort(N_S)
```

```
##       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 effect that is studied is 'Population' and all other variables are sorted as above.

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.

```
Neonics$Species.Common.Name <- as.factor(Neonics$Species.Common.Name)
class(Neonics$Species.Common.Name)
```

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

```
N_S2 <- summary(Neonics$Species.Common.Name)
sort(N_S2)
```

```
##                          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 six most commonly studied species are Honey Bee (667), Parasitic Wasp (285), Buff Tailed Bumblebee (183), Carniolan Honey Bee (152), Bumble Bee (140), Italian Honeybee (113). The bee species are most common studied, which may because they are widely distributed in the whole ecosystem and have more contact with plants to pollination.

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

```
Neonics$Conc.1..Author. <- as.numeric(Neonics$Conc.1..Author.)
```

```
## Warning: Ç¿ÖÆ¸Ä±ä¹ý³ÌÖÐ²úÉúÁËNA
```

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

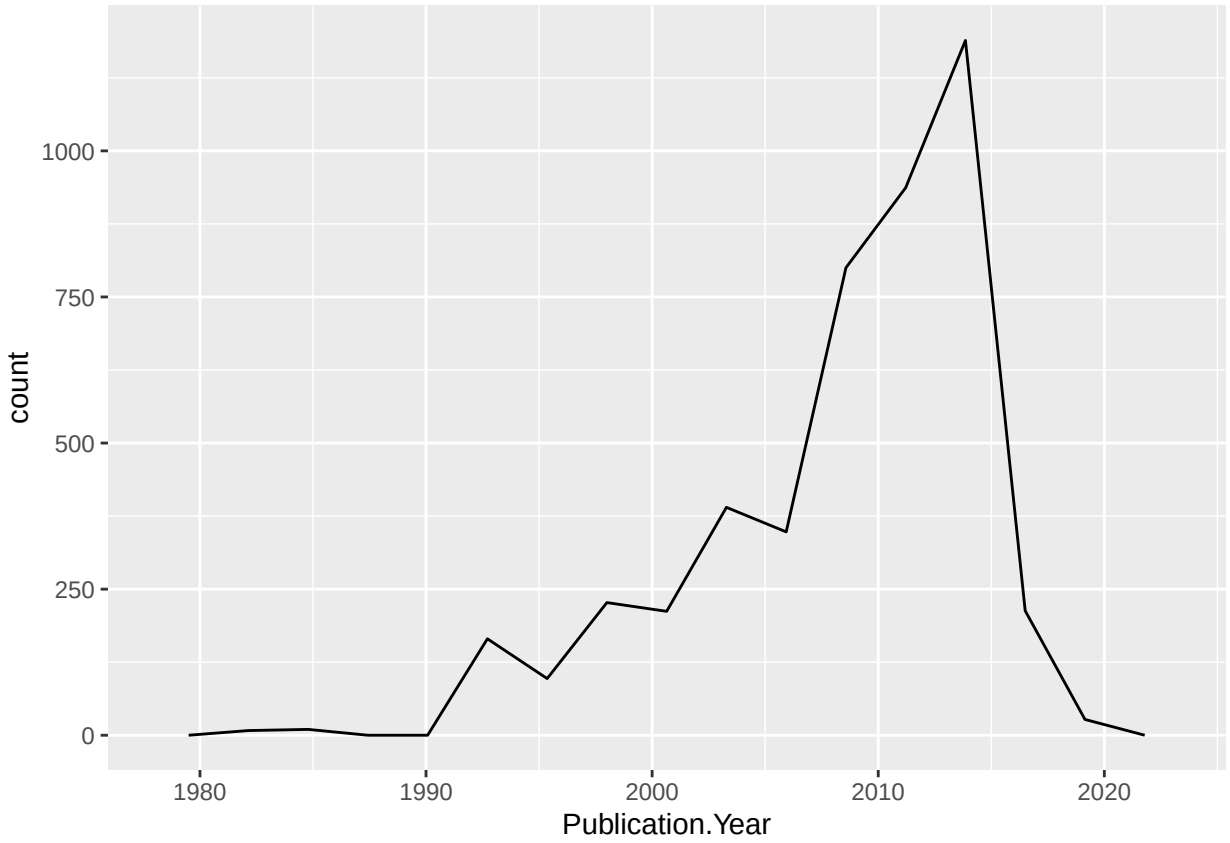
```
## [1] "numeric"
```

Answer: The class of Conc.1..Author is character since some of the data have sign but it can be converted. Here it shows it as numeric since I converted it.

Explore your data graphically (Neonics)

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

```
p0 <- ggplot(data = Neonics, mapping = aes(x=Publication.Year))
p0 + geom_freqpoly(bins = 15)
```



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

```
#Pick out different locataions and plot as their publication year
class(Neonics$Test.Location)
```

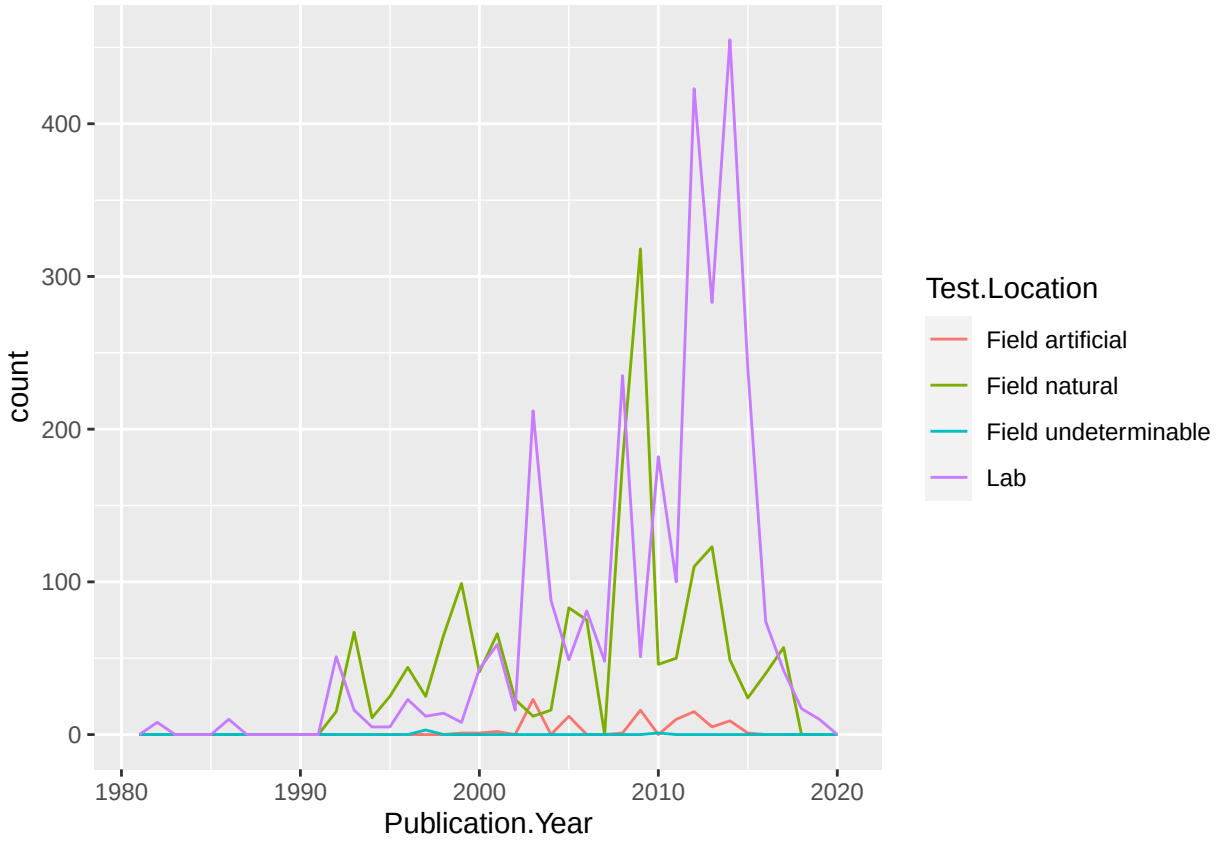
```
## [1] "character"
```

```
Neonics$Test.Location <- as.factor(Neonics$Test.Location)
summary(Neonics$Test.Location)
```

```
##     Field artificial        Field natural Field undeterminable
##                   96                 1663                    4
##                  Lab
##                 2860
```

```
library(grid)

ggplot(Neonics, aes(Publication.Year, colour = Test.Location)) +
  geom_freqpoly(binwidth = 1)
```



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

> Answer: Lab (2860), has a peak during 2010 -2020, increases before ~2015 and decreases sharply until 2020 Field natural (1663), roughly, increaes before ~2010 and decreases from ~2012 until 2020 Field artificial (96), has no obvious trend, three peaks at ~2003, ~2008 and ~2012 Field undeterminable (4), 3 cases in 1997 and 1 case in 2010

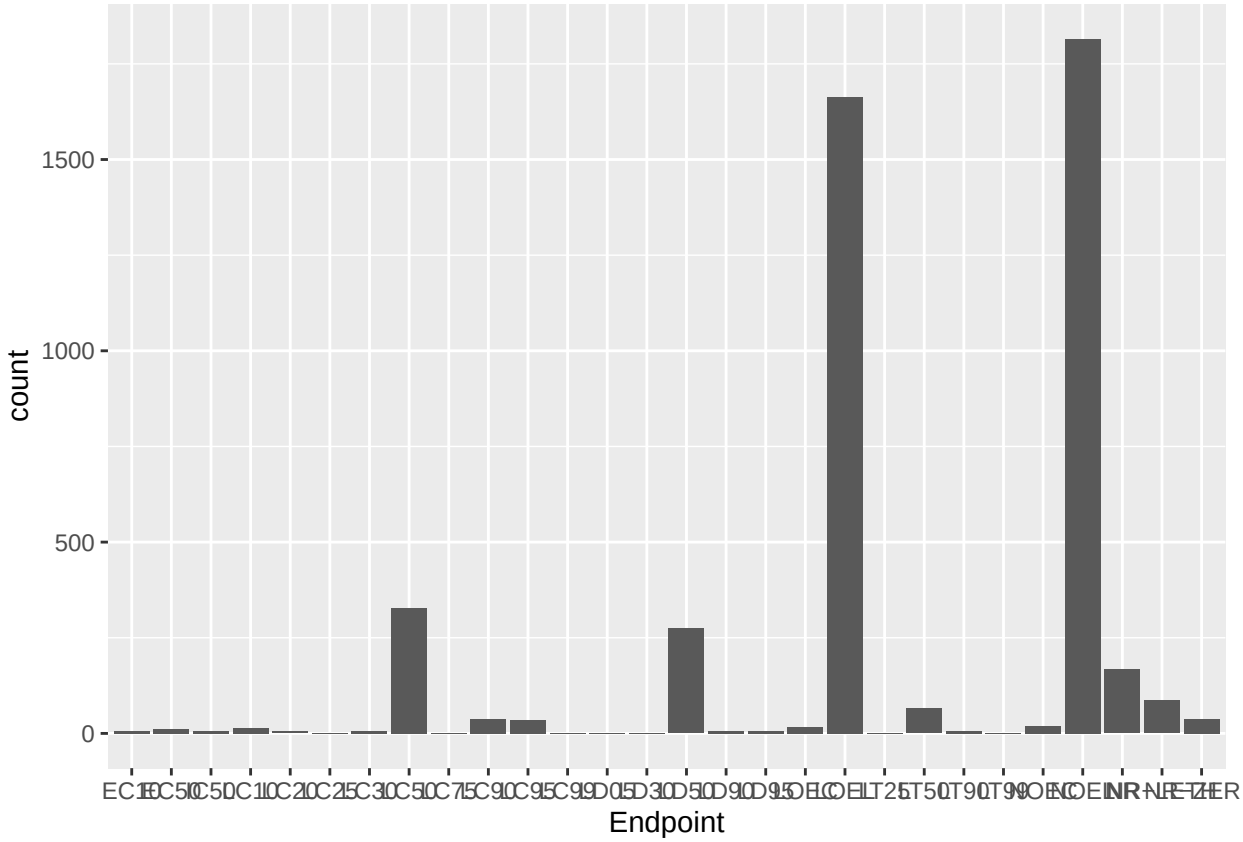
11. Create a bar graph of Endpoint counts. What are the two most common end points, and how are they defined? Consult the ECOTOX_CodeAppendix for more information.

```
Neonics$Endpoint <- as.factor(Neonics$Endpoint)
summary(Neonics$Endpoint)
```

```
##     EC10     EC50     IC50     LC10     LC20     LC25     LC30     LC50     LC75     LC90
##        6       11        6       15        5        1        6      327        1       37
##     LC95     LC99     LD05     LD30     LD50     LD90     LD95     LOEC     LOEL     LT25
##       36        2        1        1      274        6        7       17     1664        1
##     LT50     LT90     LT99     NOEC     NOEL       NR  NR-LETH  NR-ZERO
##       65        7        2       19     1816      167       86       37
```

```
p1 <- ggplot(data = Neonics, mapping = aes(x=Endpoint))
p1 + geom_bar(bins = 10)
```

```
## Warning: Ignoring unknown parameters: bins
```



Answer: The two most common end points are: Noel (1816) and LOEL (1664)

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.

```
# Format data to Date
class(Litter$collectDate)
```

```
## [1] "character"
```

```
Litter$collectDate <- as.Date(Litter$collectDate)

# Determine the dates
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$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"
```

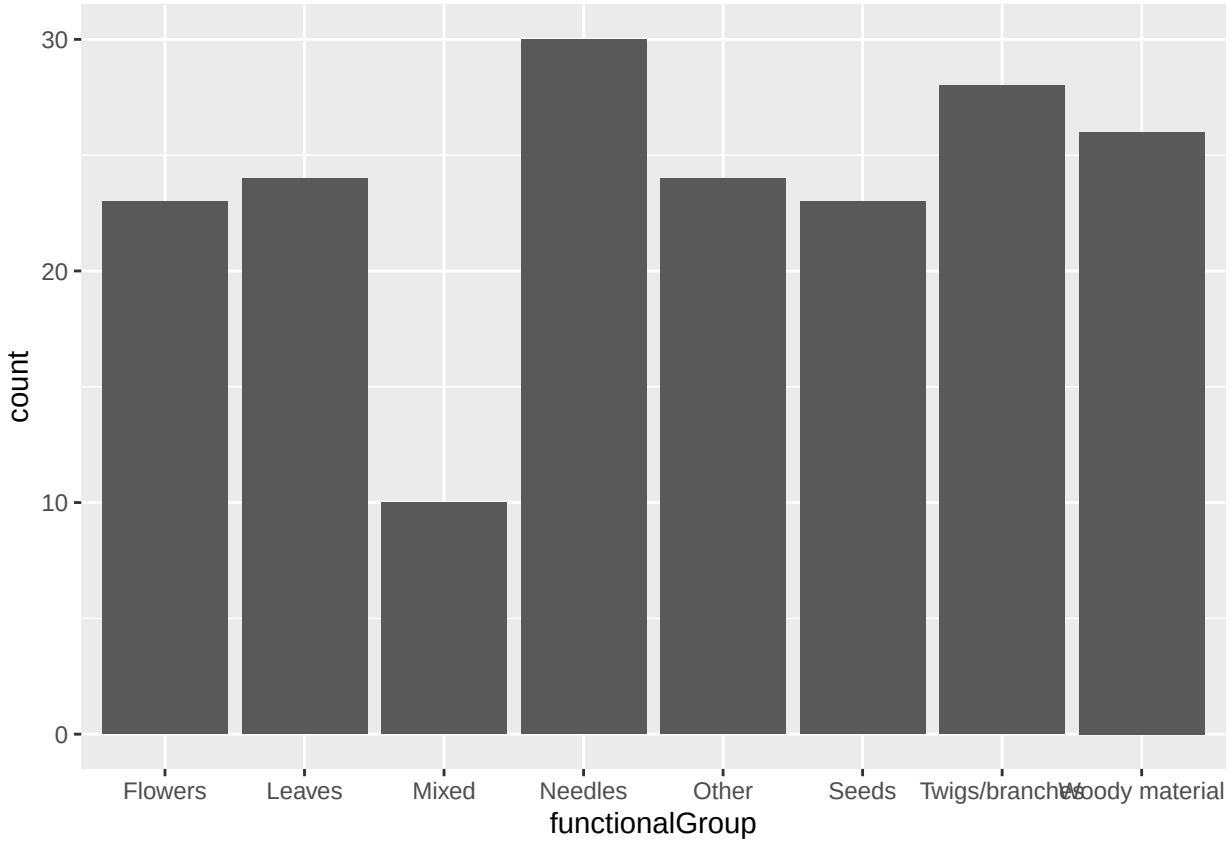
```
Litter$namedLocation <- as.factor(Litter$namedLocation)
summary(Litter$namedLocation)
```

```
## NIWO_040.basePlot.ltr NIWO_041.basePlot.ltr NIWO_046.basePlot.ltr
##                    20                    19                    18
## NIWO_047.basePlot.ltr NIWO_051.basePlot.ltr NIWO_057.basePlot.ltr
##                    15                    14                     8
## NIWO_058.basePlot.ltr NIWO_061.basePlot.ltr NIWO_062.basePlot.ltr
##                    16                    17                    14
## NIWO_063.basePlot.ltr NIWO_064.basePlot.ltr NIWO_067.basePlot.ltr
##                    14                    16                    17
```

Answer: 1. 12 plots were sampled at Miwot Ridge, 40, 41, 46, 47, 51, 57, 58, 61, 62, 63, 64, 67
2. Summary also print the number of each plot (when we set Litter$namedLocation as factor)

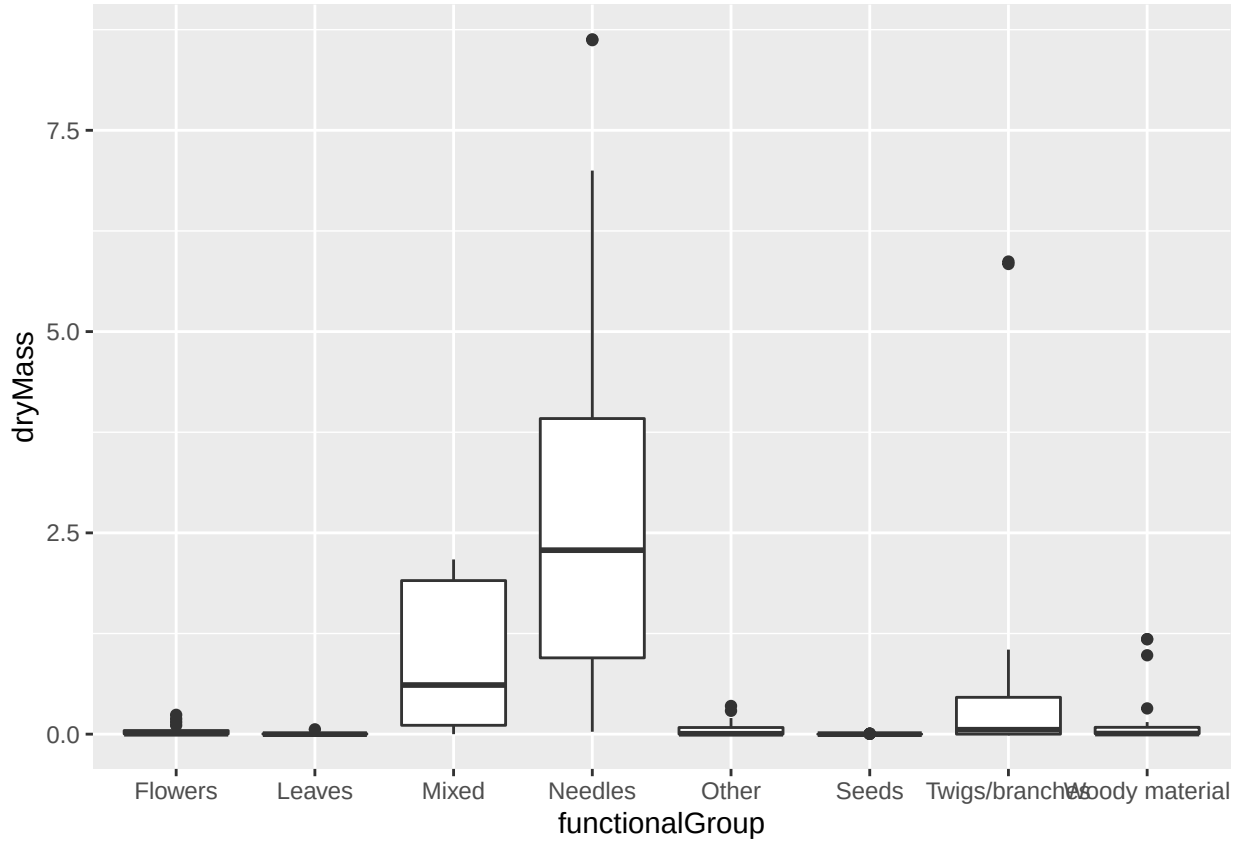
14. Create a bar graph of functionalGroup counts. This shows you what type of litter is collected at the Niwot Ridge sites. Notice that litter types are fairly equally distributed across the Niwot Ridge sites.

```
p2 <- ggplot(data = Litter, aes(
    x = functionalGroup
))
p2 + geom_bar()
```

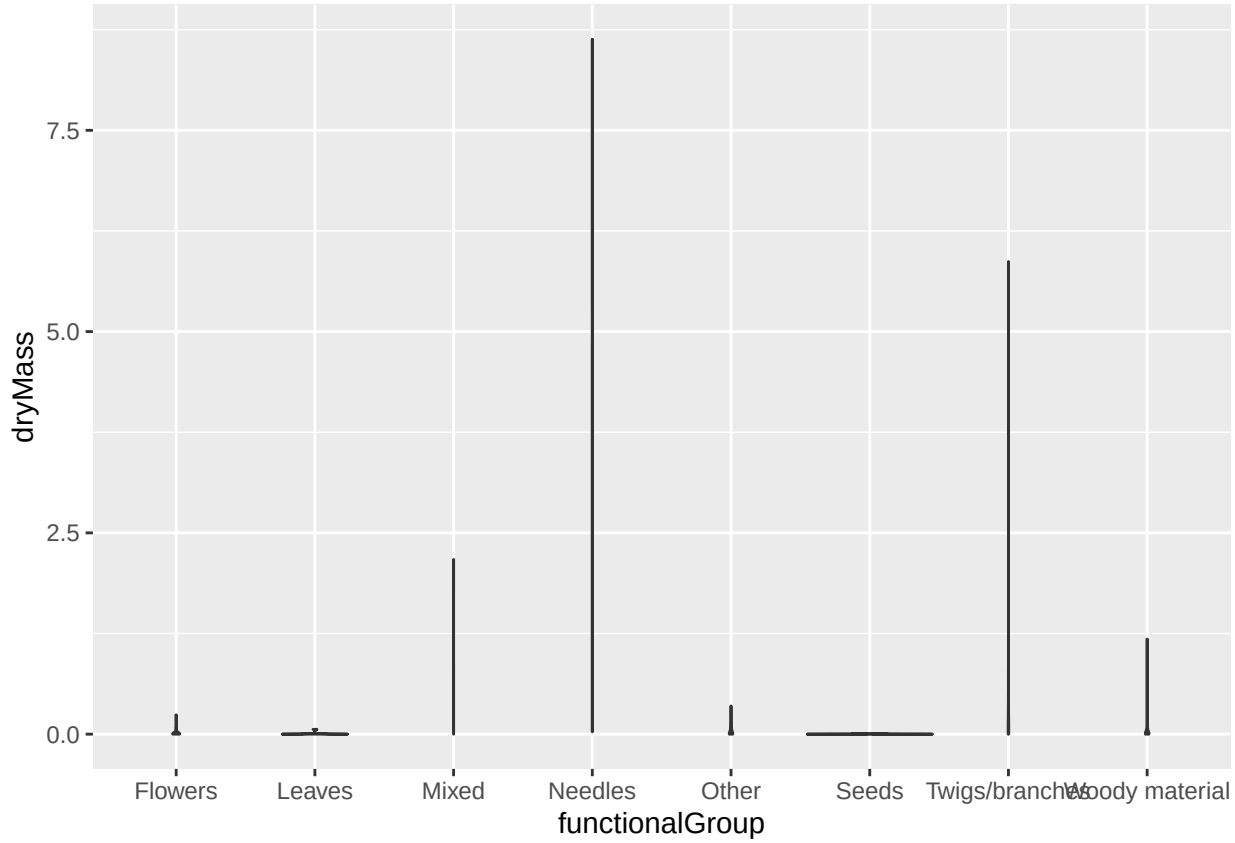


15. Using `geom_boxplot` and `geom_violin`, create a boxplot and a violin plot of dryMass by functionalGroup.

```
p3 <- ggplot(data = Litter, aes(
    x = functionalGroup,
    y = dryMass
))
p3 + geom_boxplot()
```




```
p4 <- ggplot(data = Litter, aes(
    x = functionalGroup,
    y = dryMass
))
p4 + geom_violin()
```



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

> Answer: The amount of data in this case is not large enough to support violin plot. So the violin plot would be very 'narrow' and the side peaks are not obvious since the data point amount at each value are similar. But boxplot calculate and present the 25th and 75th percentiles even if the data is sparse.

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

> Answer: Needles is the type which tends to have highest biomass at these sites.