Assignment 3: Data Exploration

Amrita Sood

OVERVIEW

This exercise accompanies the lessons in Environmental Data Analytics on Data Exploration.

Directions

1. Change "Student Name" on line 3 (above) with your name.
2. Work through the steps, **creating code and output** that fulfill each instruction.
3. Be sure to **answer the questions** in this assignment document.
4. When you have completed the assignment, **Knit** the text and code into a single PDF file.
5. After Knitting, submit the completed exercise (PDF file) to the dropbox in Sakai. Add your last name into the file name (e.g., "Salk_A03_DataExploration.Rmd") prior to submission.

The completed exercise is due on <>.

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.

```
#checking working directory
getwd()

## [1] "/Users/amritasood/Desktop/CLASSES MEM/Spring 2021/EDA -
ENV872/Environmental_Data_Analytics_2021/Assignments"

#set absolute working directory
setwd("/Users/amritasood/Desktop/CLASSES MEM/Spring 2021/EDA -
ENV872/Environmental_Data_Analytics_2021")

#install packages
library(tidyverse)

#upload datasets
Neonics <- read.csv("../Data/Raw/ECOTOX_Neonicotinoids_Insects_raw.csv")
Litter <- read.csv("../Data/Raw/NEON_NIWO_Litter_massdata_2018-08_raw.csv")
```

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 ecotoxicologoy of neonicotinoids on insects? Feel free to do a brief internet search if you feel you need more background information.

   Answer: The impact of this class of insecticides on pollinating insects such as honey bees and native bees is a cause for concern. Because they are systemic chemicals absorbed into the plant, neonicotinoids can be present in pollen and nectar, making them toxic to pollinators that feed on them. Neonicotinoids can kill beneficial insects such as honey bees, hoverflies, and parasitic wasps. Thus, we are interested in the ecotoxicology of neonicotinoids on insects such as bees.

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: We are interested in studying the litter and woody debris that falls to the ground in forests because it may provide useful insight about the nutrient cycle; help us understand the impact of different pesticides or insecticides and there leftover remanants. It is highly useful to gain a better understanding of nutrient availability and habitat information for different microbes and organisms.

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: Litter and fine woody debris sampling is conducted at terrestrial NEON sites that contain woody vegeta□on >2m tall. *Spatial Design : One litter trap pair (one elevated trap and one ground trap) is deployed for every 400 m2 plot area, resultng in 1-4 trap pairs per plot. Trap placement within plots may be either targeted or randomized, depending on the vegetation. Locations of tower plots are selected randomly within the 90% flux footprint of the primary and secondary airsheds (and additional areas in close proximity to the airshed, as necessary to accommodate sufficient spacing between plots). So the sampling spatial design was both randomized and targeted.* Temporal Sampling Design: Ground traps are sampled once per year. Target sampling frequency for elevated traps varies by vegetation present at the site, with frequent sampling 1x every 2 weeks) in deciduous forest sites during senescence, and year-round sampling (1x every 1-2 months) at evergreen sites. Sampling occured at different frequencies for different plants/trees. *Location of sampling was selected randomly within the 90% flux footprint of the primary and secondary airsheds. The available space,

plot spacing requirements, and/or the tower airshed size restricts the number of plots that can be sampled for litter.

Obtain basic summaries of your data (Neonics)

5. What are the dimensions of the dataset?

```
view(Neonics)
view(Litter)
#dimensions of datasets

dim(Neonics)

## [1] 4623   30

dim(Litter)

## [1] 188  19
```

Length of neonics dataset is 30 and width is 4623. Length of Litter dataset is 19 and width is 188.

6. Using the summary function on the "Effects" column, determine the most common effects that are studied. Why might these effects specifically be of interest?

```
#summary of the effects column
common_effects<-as.factor(Neonics$Effect)
summary(common_effects)

##     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: We are studying the ecotoxicological effect of neonictonoids on insects and these effects such as immnology or genetics or intoxication may be useful in deepening our understanding about how the effects of this insecticide on different insects such as bees.

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.

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

```
##                          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: These species all belong to same clas/family. They are all flower visiting insects. Neonicotinoid has a a toxicological effect on bees specifically and maybe some of these other insects listed above.

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

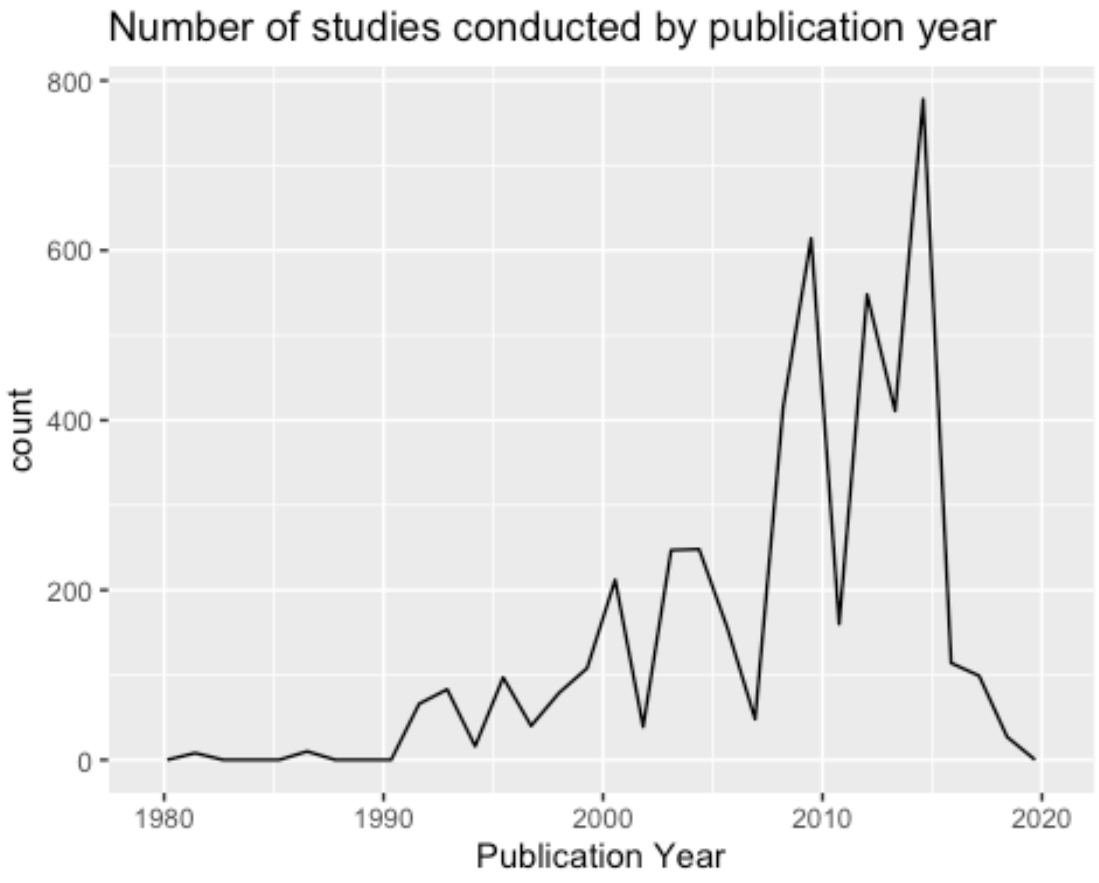
```
class(Neonics$Conc.1..Author.)

## [1] "character"
```

Answer: Concnentrations are always numeric however, in this dataset Conc.1.Author has other non numeric charcaters mentioneds such as "~" which makes R read it as a character instead of numeric.

Explore your data graphically (Neonics)

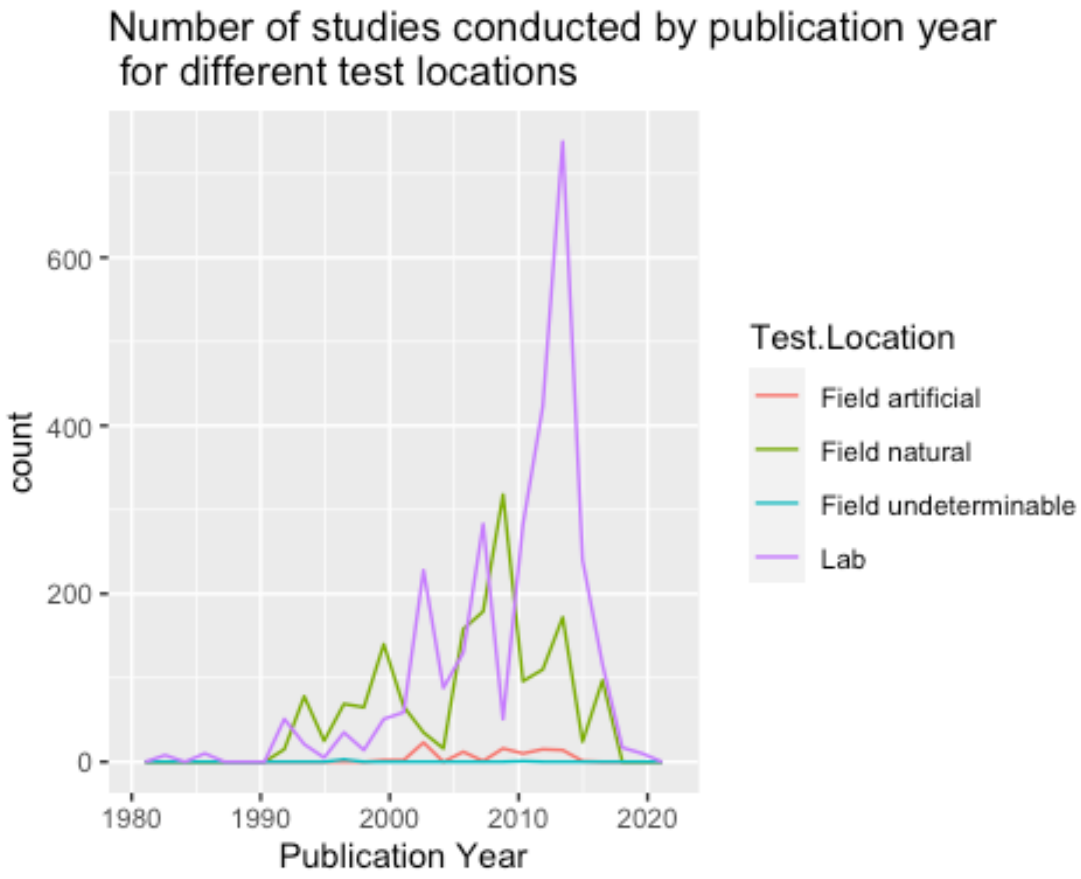
9. Using `geom_freqpoly`, generate a plot of the number of studies conducted by publication year.

```
ggplot(Neonics) +
  geom_freqpoly(aes(x = Publication.Year), bins = 30) + ggtitle("Number of
studies conducted by publication year") + labs(x = "Publication 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), bins = 25)+
ggtitle("Number of studies conducted by publication year \n for different
test locations") + labs(x = "Publication Year")
```

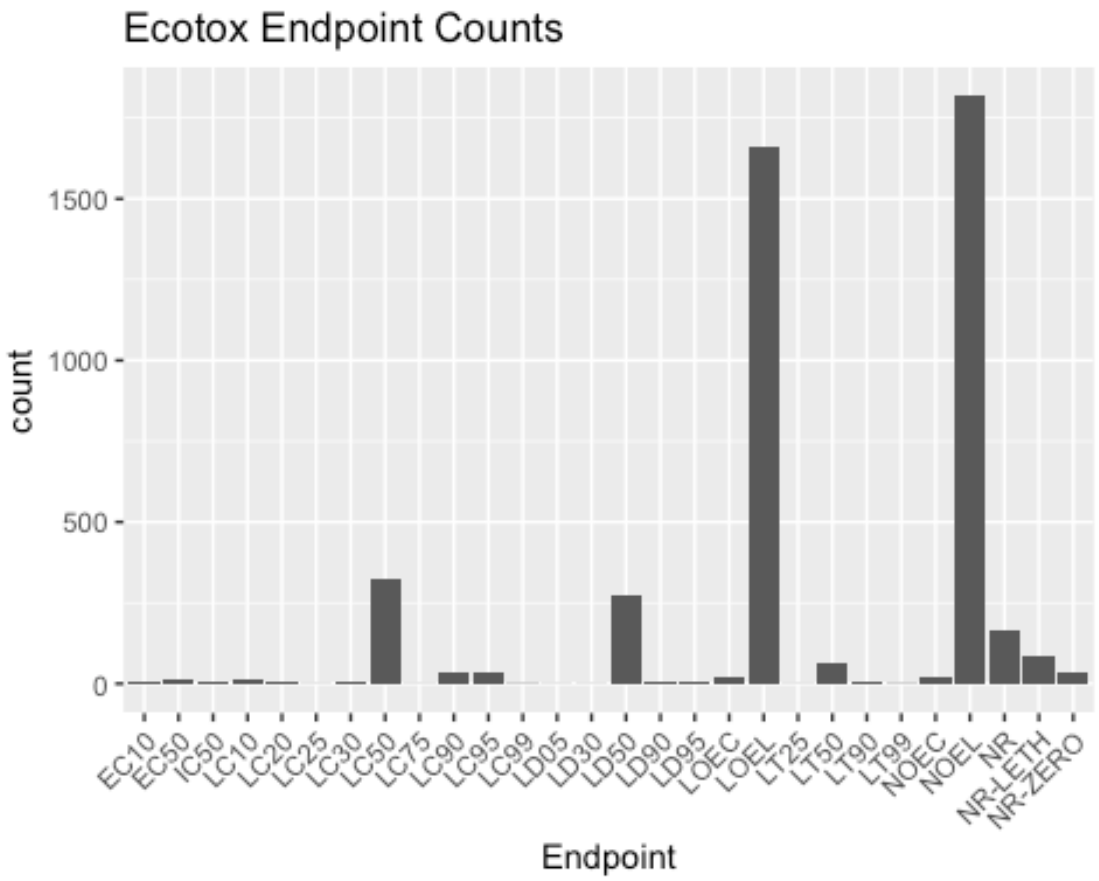


Interpret this graph. What are the most common test locations, and do they differ over time?

Answer: Lab is the most common test location. The second most common is Field Natural.

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.

```
ggplot(Neonics, aes(x = Endpoint)) +
  geom_bar()+ theme(axis.text.x = element_text(angle = 45, hjust =1) )+
ggtitle("Ecotox Endpoint Counts")
```



Answer: NOEL (No-observable-effect-level) and LOEL (Lowest-observable-effect-level)are the 2 most common endpoints. LOEL is defined as Lowest-observable-effect-level: lowest dose (concentration) producing effects that were significantly different (as reported by authors) from responses of controls. NOEL is defined as No-observable-effect-level: highest dose (concentration) producing effects not significantly different from responses of controls according to author's reported statistical test.

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] "character"

Litter$collectDate <- as.Date(Litter$collectDate, format = "%Y-%m-%d")
class(Litter$collectDate)

## [1] "Date"

August2018 <- unique(Litter$collectDate)
August2018
```

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

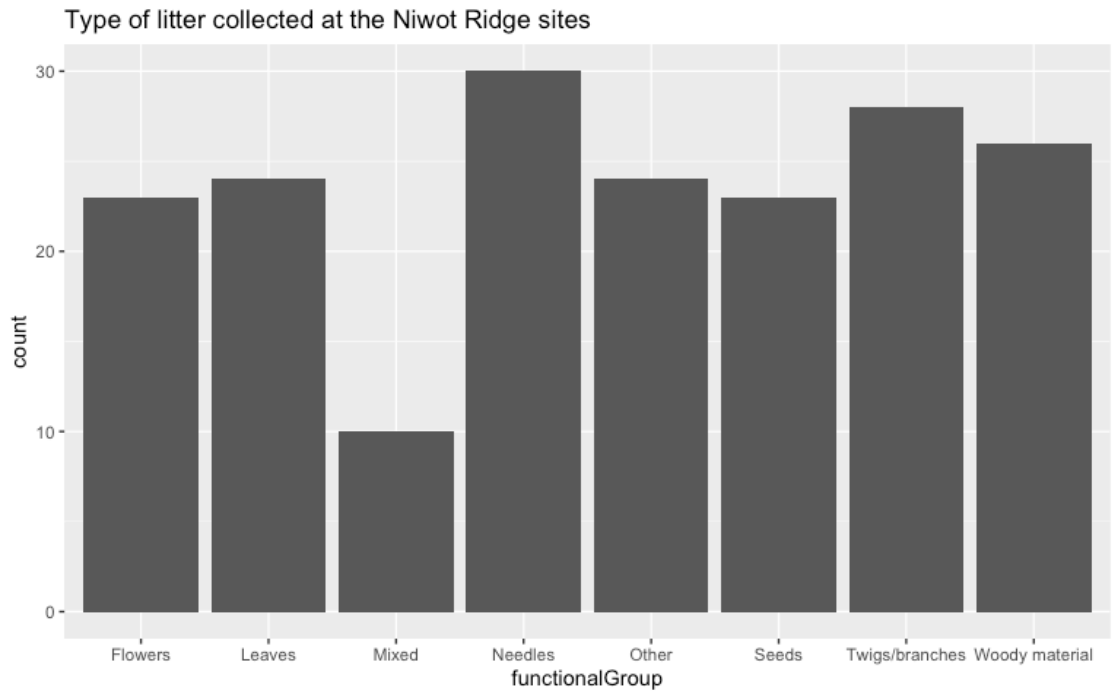
Collect date was a character so I converted it into date. 13. Using the `unique` function, determine how many plots were sampled at Niwot Ridge. How is the information obtained from `unique` different from that obtained from `summary`?

```
plots_sampled <- unique(Litter$plotID)
plots_sampled

##  [1] "NIWO_061" "NIWO_064" "NIWO_067" "NIWO_040" "NIWO_041" "NIWO_063"
##  [7] "NIWO_047" "NIWO_051" "NIWO_058" "NIWO_046" "NIWO_062" "NIWO_057"

plots <- as.factor(Litter$plotID)
summary(plots)

## 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: 12 plots were sampled. According to summary there are also 12 plots. The information obtained is same.

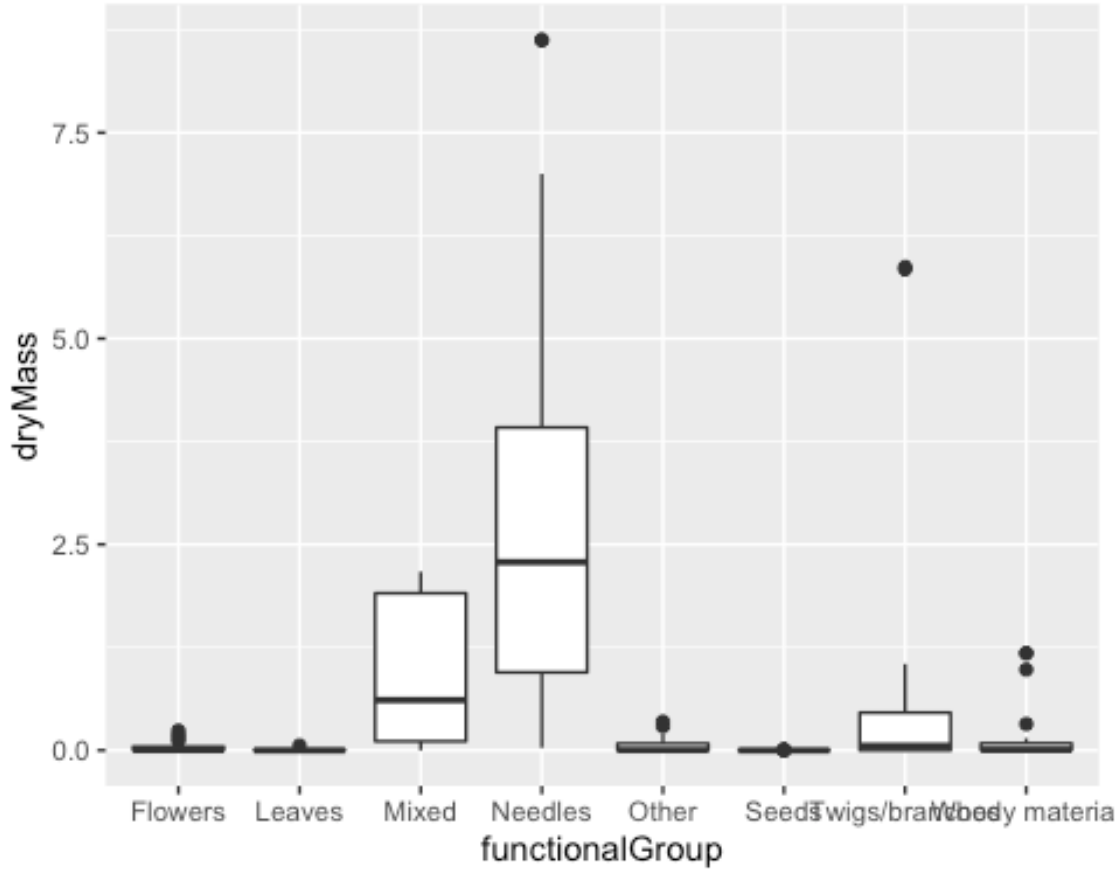
14. Create a bar graph of functionalGroup counts. This shows you what type of litter is collected at the Niwot Ridge sites. Notice that litter types are fairly equally distributed across the Niwot Ridge sites.

```
ggplot(Litter, aes(x = functionalGroup)) +
  geom_bar()+ ggtitle("Type of litter collected at the Niwot Ridge sites")
```

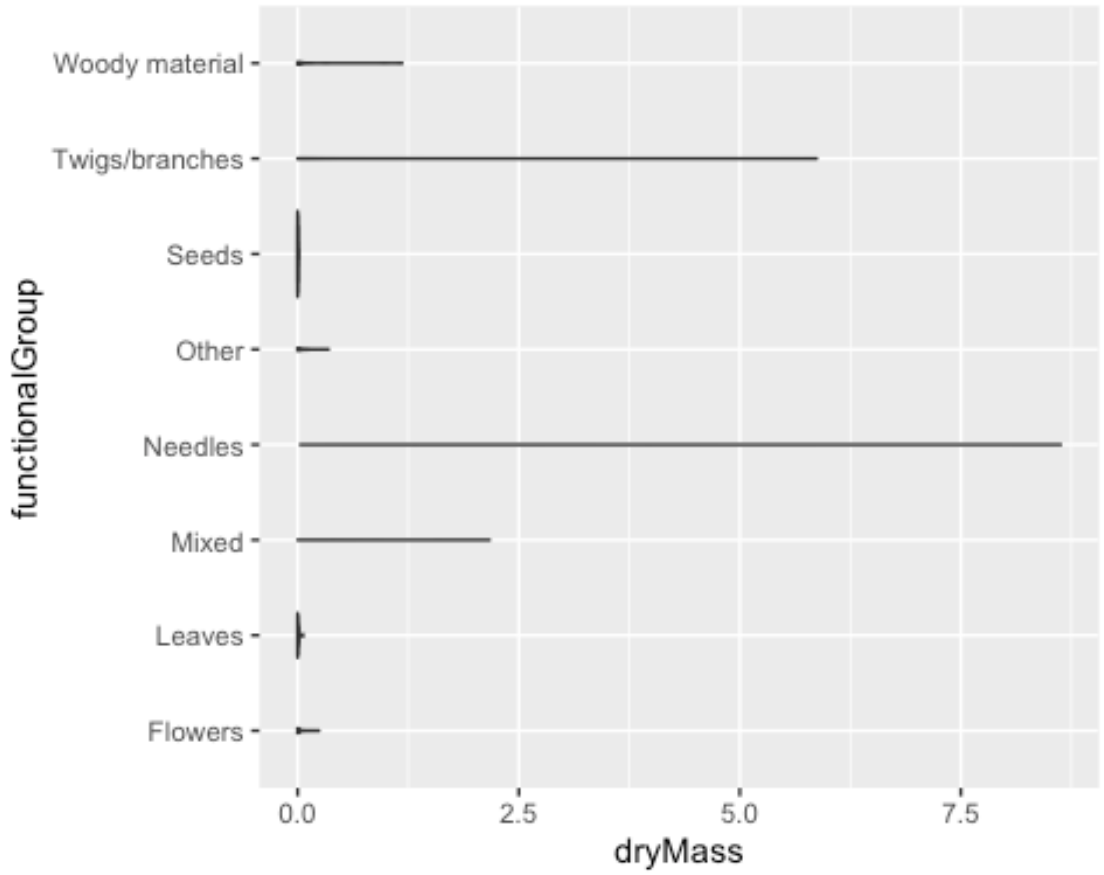


15. Using `geom_boxplot` and `geom_violin`, create a boxplot and a violin plot of dryMass by functionalGroup.

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




```
ggplot(Litter) +
  geom_violin(aes(x = dryMass, y = functionalGroup),
              draw_quantiles = c(0.25, 0.5, 0.75))

## Warning in regularize.values(x, y, ties, missing(ties), na.rm = na.rm):
## collapsing to unique 'x' values

## Warning in regularize.values(x, y, ties, missing(ties), na.rm = na.rm):
## collapsing to unique 'x' values

## Warning in regularize.values(x, y, ties, missing(ties), na.rm = na.rm):
## collapsing to unique 'x' values
```



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

> Answer: Violin plots show probability density of a rotated kernel on each side. In this case boxplot is more effective because the data distribution is more clearly visible and easy to read. The violin plot doesnt provide any useful information. A violin plot is more useful when the data is multimodal but in our case we can visualize the data better with just a boxplot.

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

> Answer: Needles