# Assignment 3: Data Exploration

Ayoung Kim

Fall 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.

```
#Check working directory
getwd()
```

```
## [1] "/home/guest/EDE_Fall2023"
```

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

library(tidyverse)
library(lubridate)


#Name datasets
Neonics <-read.csv("./Data/Raw/ECOTOX_Neonicotinoids_Insects_raw.csv",stringsAsFactors = TRUE)
Litter <-read.csv("./Data/Raw/NIWO_Litter/NEON_NIWO_Litter_massdata_2018-08_raw.csv",stringsAsFactors =
```

```
#I intalled two pacakges, tidyverse and lubridate, and imported two datasets from the raw data folder.A
```

## 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: Because using neonicotinoids is having a negative impacts on ecological system, especially on pollinators and water(aquatic) ecosystem. #From Xerces Society

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: Because studying litter and woody debris can be helpful to figure out the impacts of neonicotinoids on ecological system.

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. Litterfall and fine woody debris sampling 2.Dry weight of litterfall and fine woody debris collected from litter traps by plant functional type 3. Associated metadata from input data

## Obtain basic summaries of your data (Neonics)

5. What are the dimensions of the dataset?

```
#Dimension of Neonics (Used dim to get dimensions. 4623 Rows, 30 Columns)
dim(Neonics)
```

```
## [1] 4623   30
```

```
#Dimension of Litter(Used dim function to get dimension. 188 Rows, 19 Columns)
dim(Litter)
```

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

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?

```
#Most common effects studied
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 is Population with 1803. Population would be of interest because studying population can help understand how many speicies and populations are affected by the neonicide overall.

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. . . ]

```
#The six most commonly studied species
summary(Neonics$Species.Common.Name)
```

```
##                          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 for 7: The most commonly studied species are Honey Bee, Parasitic Wasp, Buff Tailed Bumblebee, Carniolan Honey Bee, Bumble Bee, and Italian Honey Bee.

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?

```
#The class of 'Con.1..Author' column
class(Neonics$Conc.1..Author.)
```

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

Answer: The class of 'Con.1..Author' is factor. That is because I imported the datasets into a dataframe.

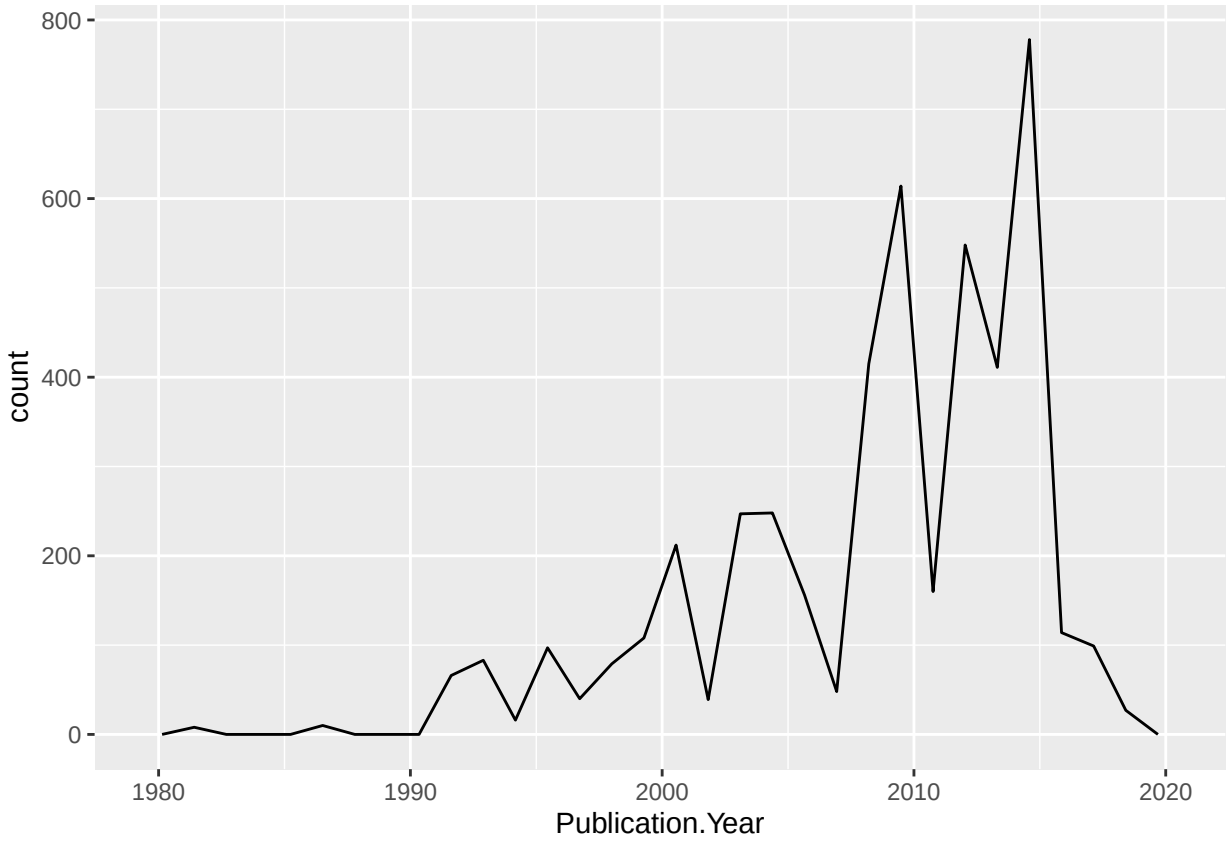
## Explore your data graphically (Neonics)

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

```
#Frequency Line Graph with Publication Year (I used ggplot+geom_freqpoly to get frequency line graph of

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

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

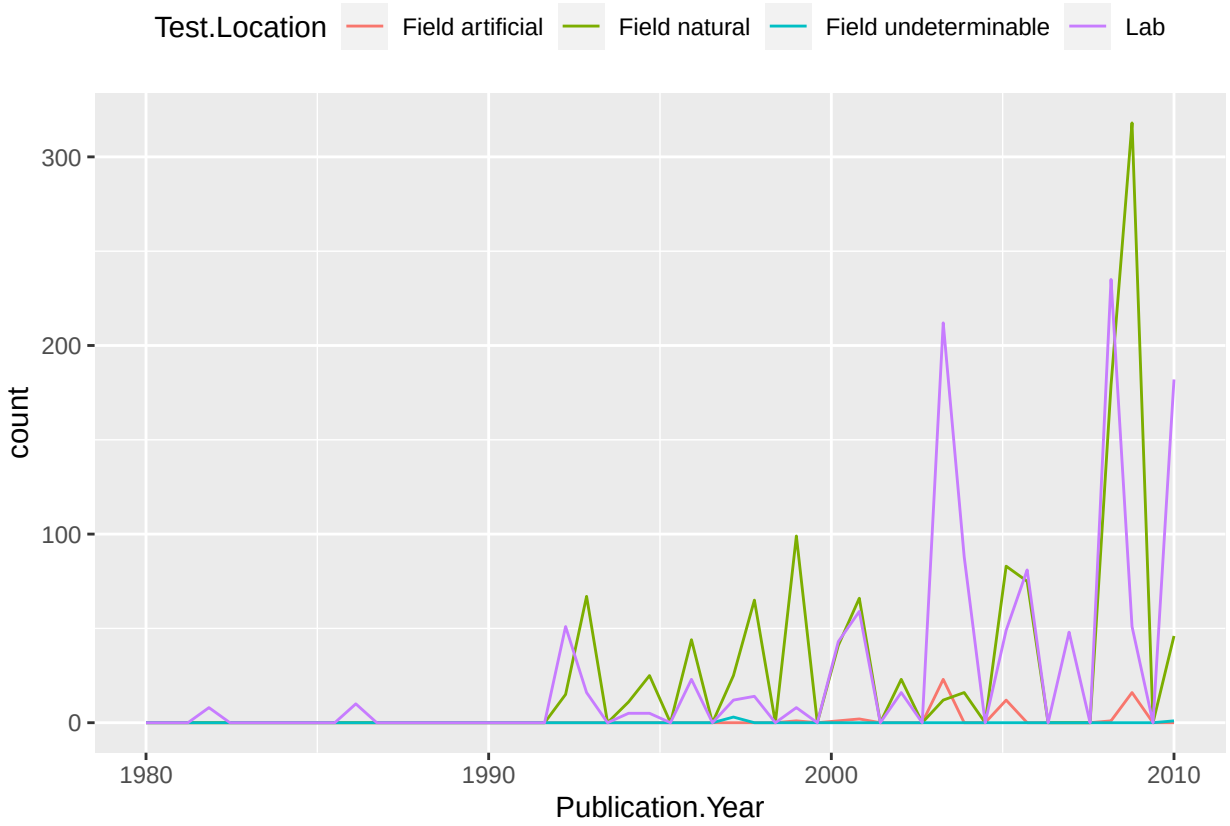


10. Reproduce the same graph but now add a color aesthetic so that different Test.Location are displayed as different colors.

```
#Frequency Line Graph with a color aesthetic
ggplot(Neonics)+
  geom_freqpoly(aes(x=Publication.Year, color=Test.Location), bins=50)+
  scale_x_continuous(limits=c(1980,2010))+
  theme(legend.position = "top")
```

```
## Warning: Removed 2137 rows containing non-finite values (`stat_bin()`).
```

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



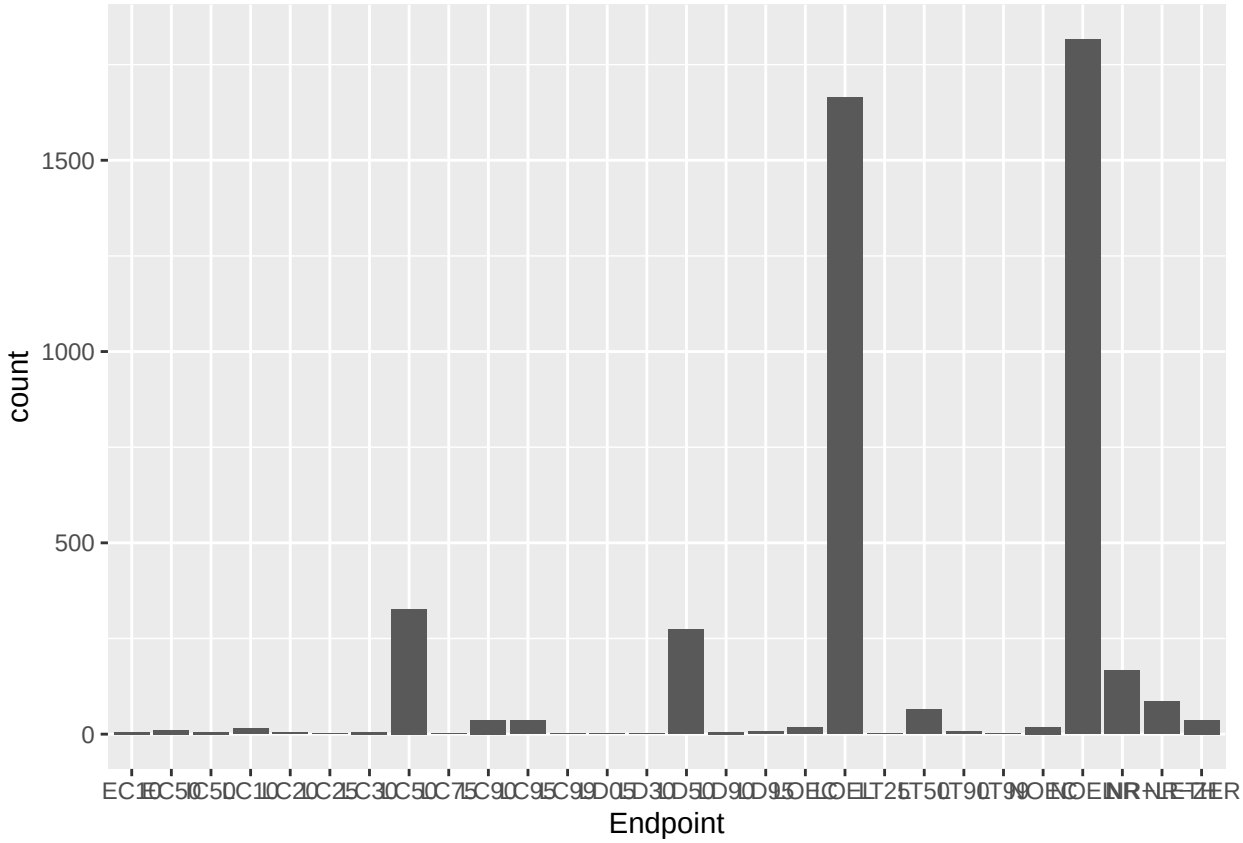

Interpret this graph. What are the most common test locations, and do they differ over time?

> Answer: The most common test locations are lab and Field natural tests.The lab tests increased after 2000 and reached the peak in 2007-2009 (approximately). The Field natural tests were popular steadily from the early 1990 and reached the peak in the same period as the lab test.

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...]

```
#Bar graph of Endpoint counts
ggplot(Neonics,aes(x=Endpoint))+
  geom_bar()
```




```
theme(axis.text.x = element_text(angle = 90, vjust = 0.5, hjust=1))
```

```
## List of 1
##  $ axis.text.x:List of 11
##   ..$ family       : NULL
##   ..$ face         : NULL
##   ..$ colour       : NULL
##   ..$ size         : NULL
##   ..$ hjust        : num 1
##   ..$ vjust        : num 0.5
##   ..$ angle        : num 90
##   ..$ lineheight   : NULL
##   ..$ margin       : NULL
##   ..$ debug        : NULL
##   ..$ inherit.blank: logi FALSE
##   ..- attr(*, "class")= chr [1:2] "element_text" "element"
##  - attr(*, "class")= chr [1:2] "theme" "gg"
##  - attr(*, "complete")= logi FALSE
##  - attr(*, "validate")= logi TRUE
```

Answer:The two most common end points are NOEL and LOEL. NOEL is No-oberservable-effect-level and LOEL is Lowest-observable-effect-level.

## 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.

```
#Determine the class of collectDate (Class: Factor)
class(Litter$collectDate)
```

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

```
#Change to a date (Class of collectDate: "date")
Litter$collectDate<-as.Date(Litter$collectDate,format="%Y-%M-%d")
class(Litter$collectDate)
```

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

```
#Which dates litter was sampled in Aug 2018 ("2018-09-02" and "2018-09-30")
unique(Litter$collectDate)
```

```
## [1] "2018-09-02" "2018-09-30"
```

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`?

```
#How many plots were sampled using Unique and Summary functions
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
```

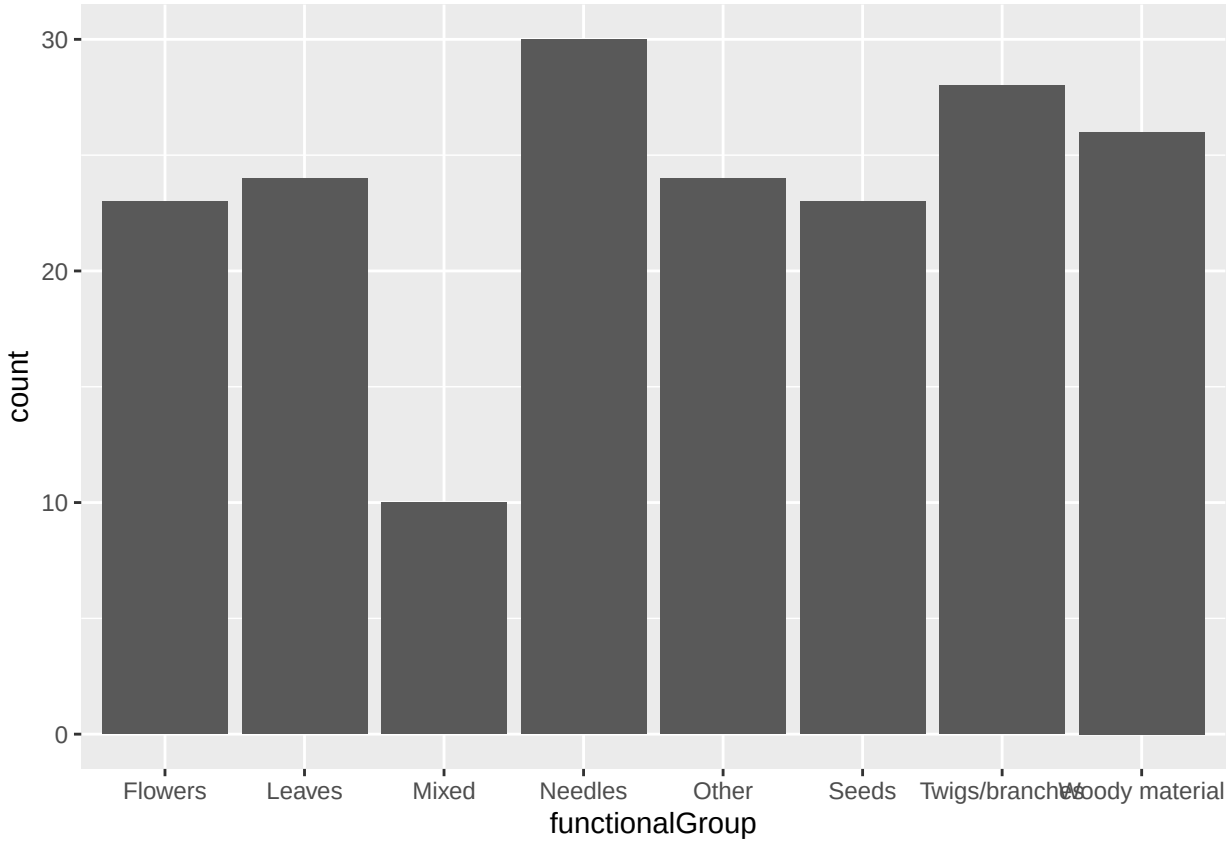
```
summary(Litter$plotID)
```

```
## 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: 12 plots were sampled at Niwot Ridge. Though both unique and summary showed 12 plots, but using summary, I could get how many values are falling into each plot as well.

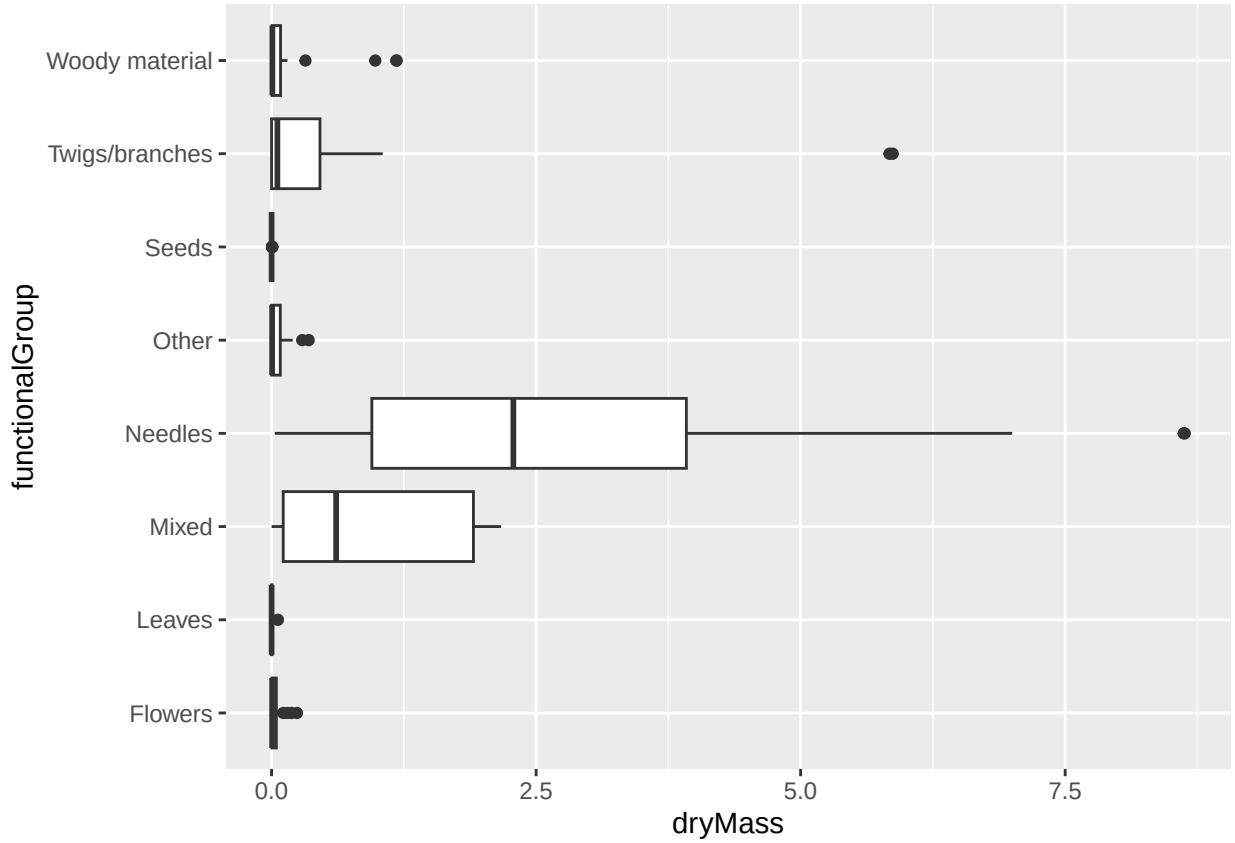
14. Create a bar graph of functionalGroup counts. This shows you what type of litter is collected at the Niwot Ridge sites. Notice that litter types are fairly equally distributed across the Niwot Ridge sites.

```
#Bar graph of functionalGroup
ggplot(Litter,aes(x=functionalGroup))+
  geom_bar()
```

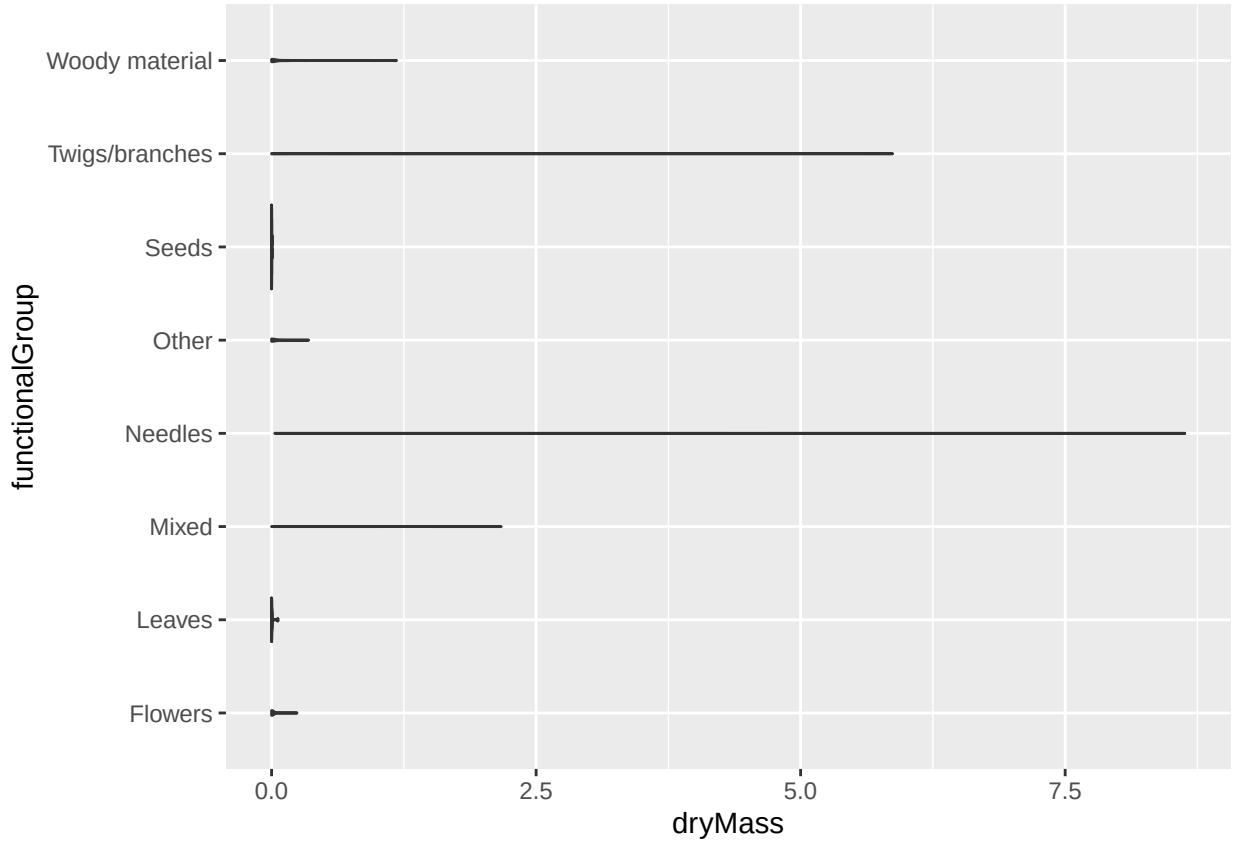



15. Using `geom_boxplot` and `geom_violin`, create a boxplot and a violin plot of dryMass by functionalGroup.

```
#Boxplot of dryMass by functionalGroup (drymass to x axis, functionalGroup to y axis)
ggplot(Litter)+
  geom_boxplot(aes(x=dryMass, y=functionalGroup))
```




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




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

Answer: Because I can easily find outliers and easy to interpret.

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

Answer: Needles