Assignment 3: Data Exploration

Jack Alcorn

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.

```
getwd()
```

```
## [1] "Z:/ENV_872_Data_Analy/Environmental_Data_Analytics_2021/Assignments"
```

```
setwd("Z:/ENV_872_Data_Analy/Environmental_Data_Analytics_2021")
library(tidyverse)
```

```
## Warning: package 'tidyverse' was built under R version 4.0.3

## Warning: package 'ggplot2' was built under R version 4.0.3

## Warning: package 'tibble' was built under R version 4.0.3

## Warning: package 'tidyr' was built under R version 4.0.3

## Warning: package 'readr' was built under R version 4.0.3

## Warning: package 'purrr' was built under R version 4.0.3

## Warning: package 'dplyr' was built under R version 4.0.3

## Warning: package 'stringr' was built under R version 4.0.3

## Warning: package 'forcats' was built under R version 4.0.3
```

```
library(lubridate)
```

```
## Warning: package 'lubridate' was built under R version 4.0.3
```

```
Neonics <- read.csv("Z:/ENV_872_Data_Analy/Environmental_Data_Analytics_2021/Data/Raw/ECOTOX_Neonicotin
Litter <- read.csv('Z:/ENV_872_Data_Analy/Environmental_Data_Analytics_2021/Data/Raw/NEON_NIWO_Litter_ma
```

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: Neonics have been shown to have adverse effects on honey-bee populations and to also be harmful to birds by reducing overall insect populations.

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: Looking at forest litter can be useful in understanding the successional stage of a forest. It can also be used to analyze the ecosystem and animals who depend on litter and woody debris to live. Lastly, it could be used for wildfire management. Forest with an overabundance of litter and woody debris could need a prescribed burn.

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 was collected in sites where woody vegetation is greater than 2 meters tall. 40x40m and 20x20m plots with 10x10m and 1x1m nested subplots were created to sample the litter and woody debris. * deciduous forests were sampled during senescence and evergreen forests were sampled year round * In sites with with greater than 50% aerial cover of woody vegetaion greater than 2m in height, placement of litter traps is random * litter is defined as material that is dropped from the forest canopy and has a butt end diameter<2cm and a length <50 cm

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 "Effects" 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 and mortality are the the most common effects studied. These are important to monitor the number of insects/honey-bees in the ecosystem. Severe declines in population or severe increases in mortality of insects/honey-bees can have disastrous effects on animals and plants that depend on these insects and honey 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.

```
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: The Honey Bee, Parasitic Wasp, Buff Tailed Bumblebee, Carniolan Honey Bee, Bumble Bee, and Italian Honeybee are the six most studied species in the data set. All of these insects are pollinators and would be of importance to crops. Without these pollinators, crops would not be as productive.

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"
```

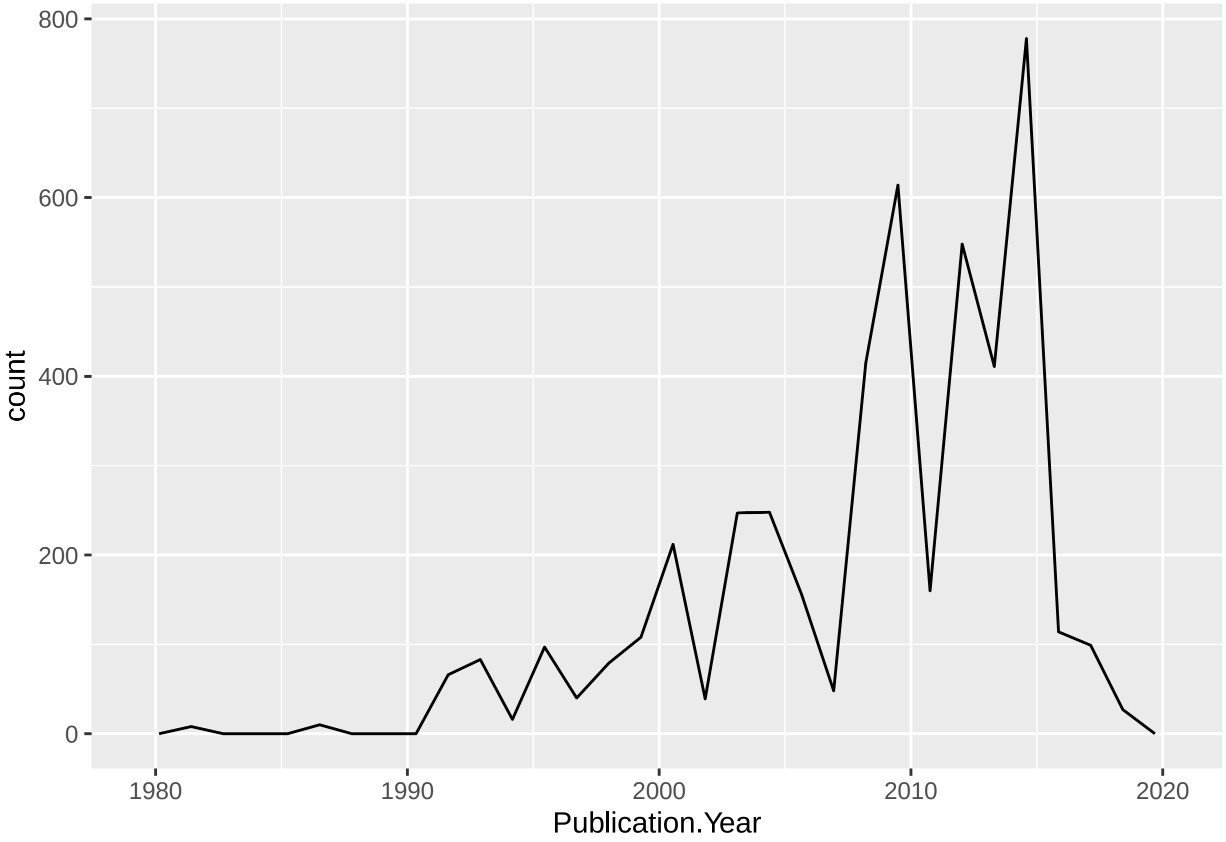
> Answer: It is a character. It could be becuase the concentrations are in different units.

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))
```

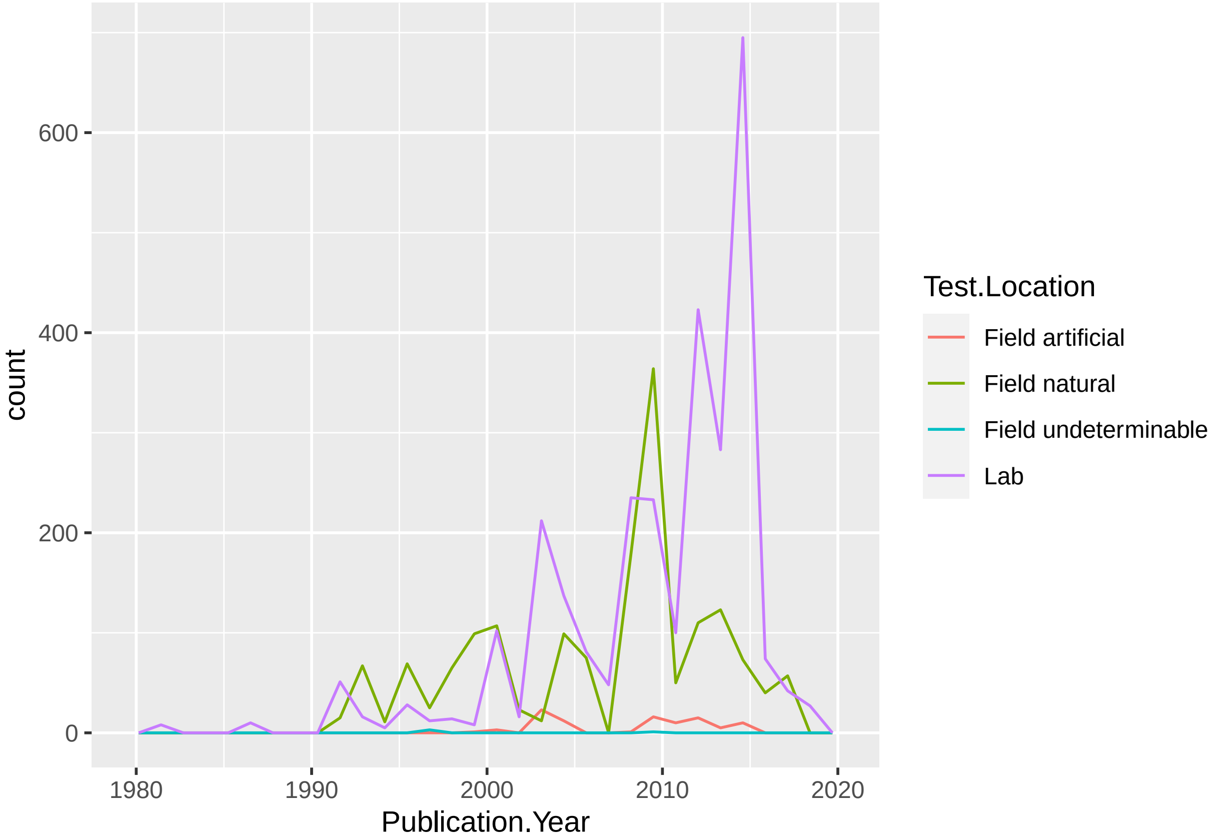
```
## `stat_bin()` using `bins = 30`. Pick better value with `binwidth`.
```



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

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

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

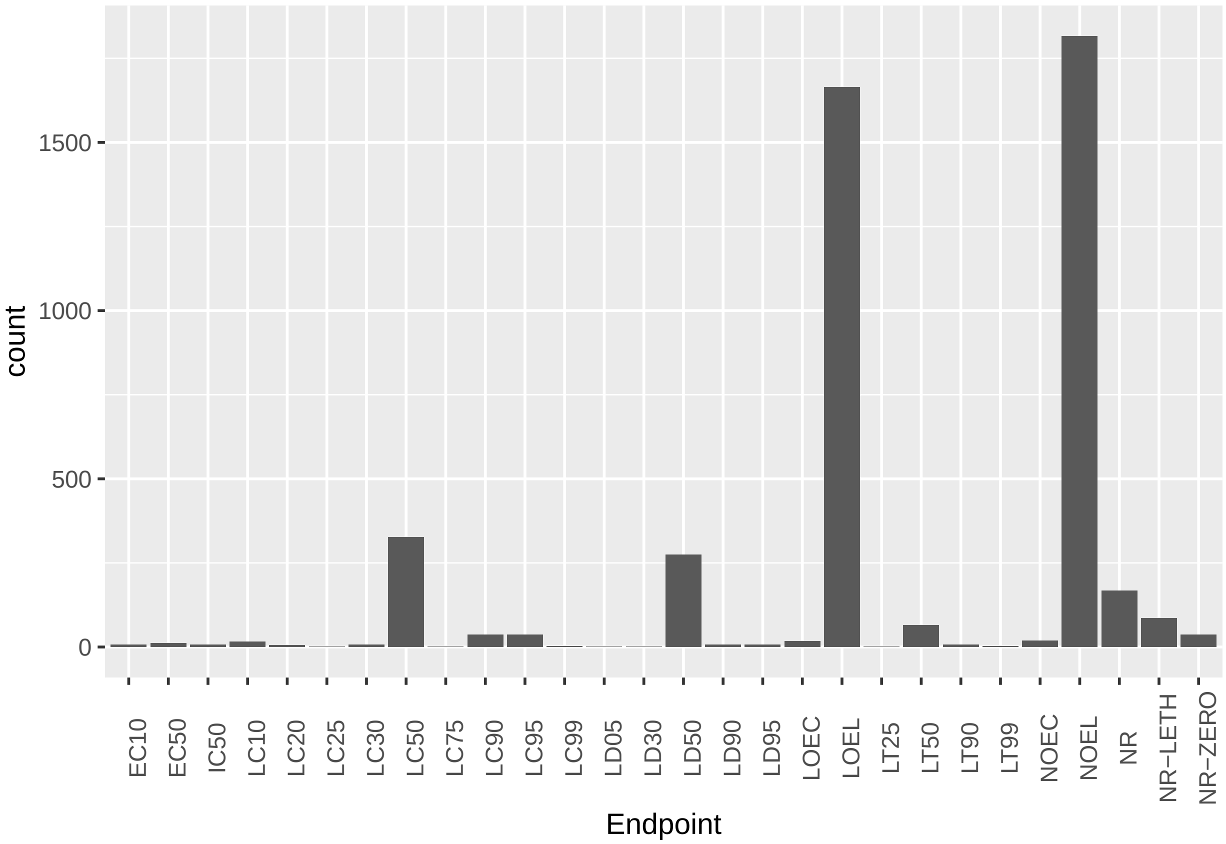


Interpret this graph. What are the most common test locations, and do they differ over time?

> Answer: The most common test locations are the lab and the natural field. The natural field has been pretty constant with a sharp spike around 2010. Lab locations have increase from 1990 to 2015 and decreased from 2015 to 2020.

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 = 90))
```



Answer: The two most common endpoints are LOEL and NOEL. LOEL is Lowest-observable-effect-level: lowest dose (concentration) producing effects that were significantly different and NOEL is No-observable-effect-level: highest dose (concentration) producing effects not significantly different from responses of controls.

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, format = "%Y-%m-%d")
class(Litter$collectDate)
```

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

```
?unique
```

```
## starting httpd help server ... done
```

```
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
```

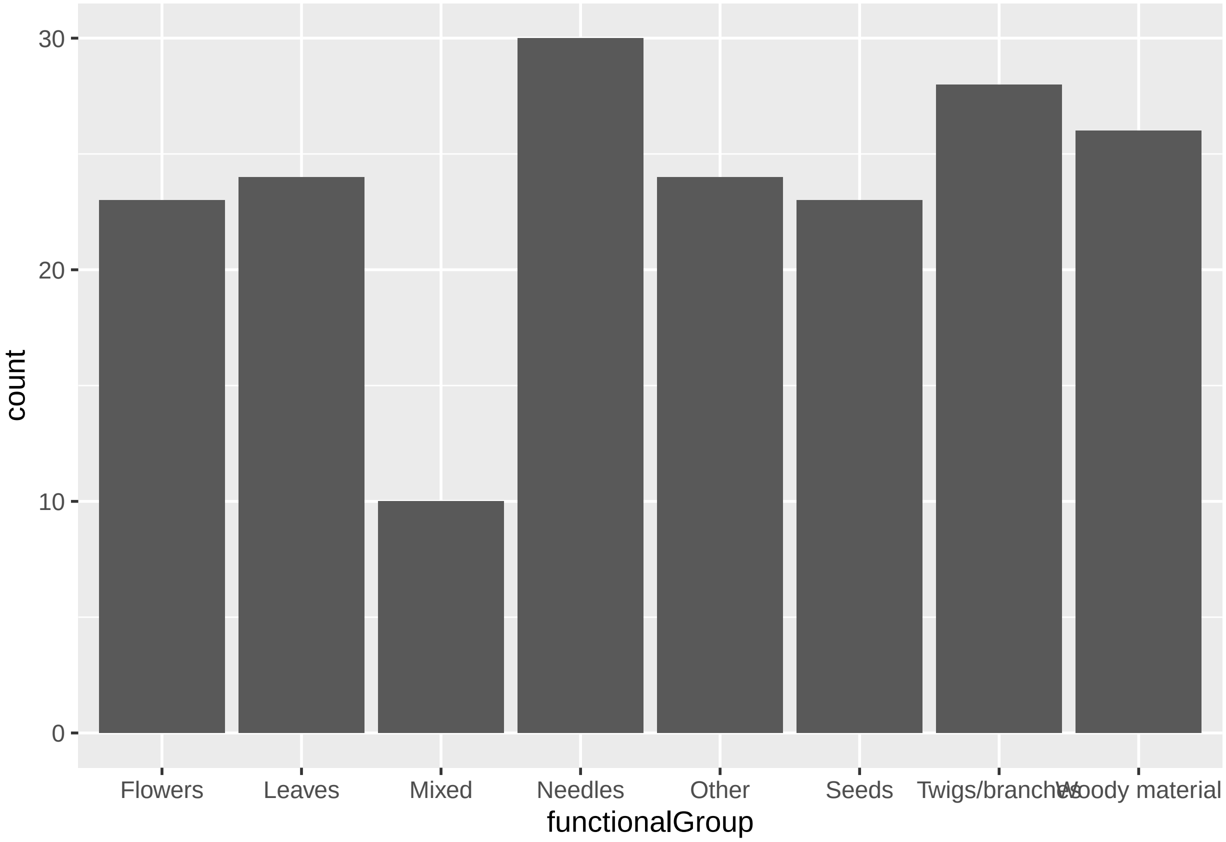
```
summary(Litter$plotID)
```

```
## 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: 12 different plots were sampled at NIWOT Ridge. The unique function tells you how many unique values are in a column. It eliminates duplicate values. The summary function tells us how many samples in each plot were taken.

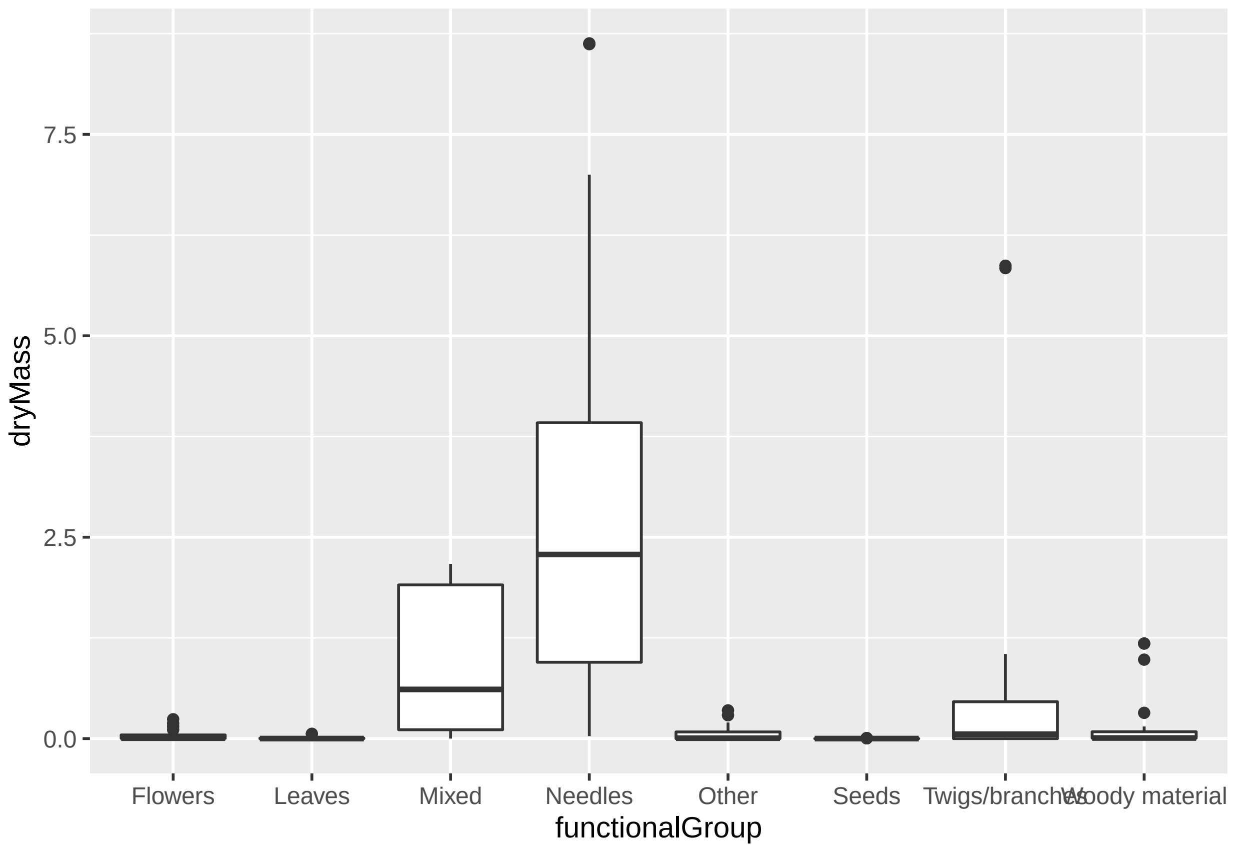
14. Create a bar graph of functionalGroup counts. This shows you what type of litter is collected at the Niwot Ridge sites. Notice that litter types are fairly equally distributed across the Niwot Ridge sites.

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

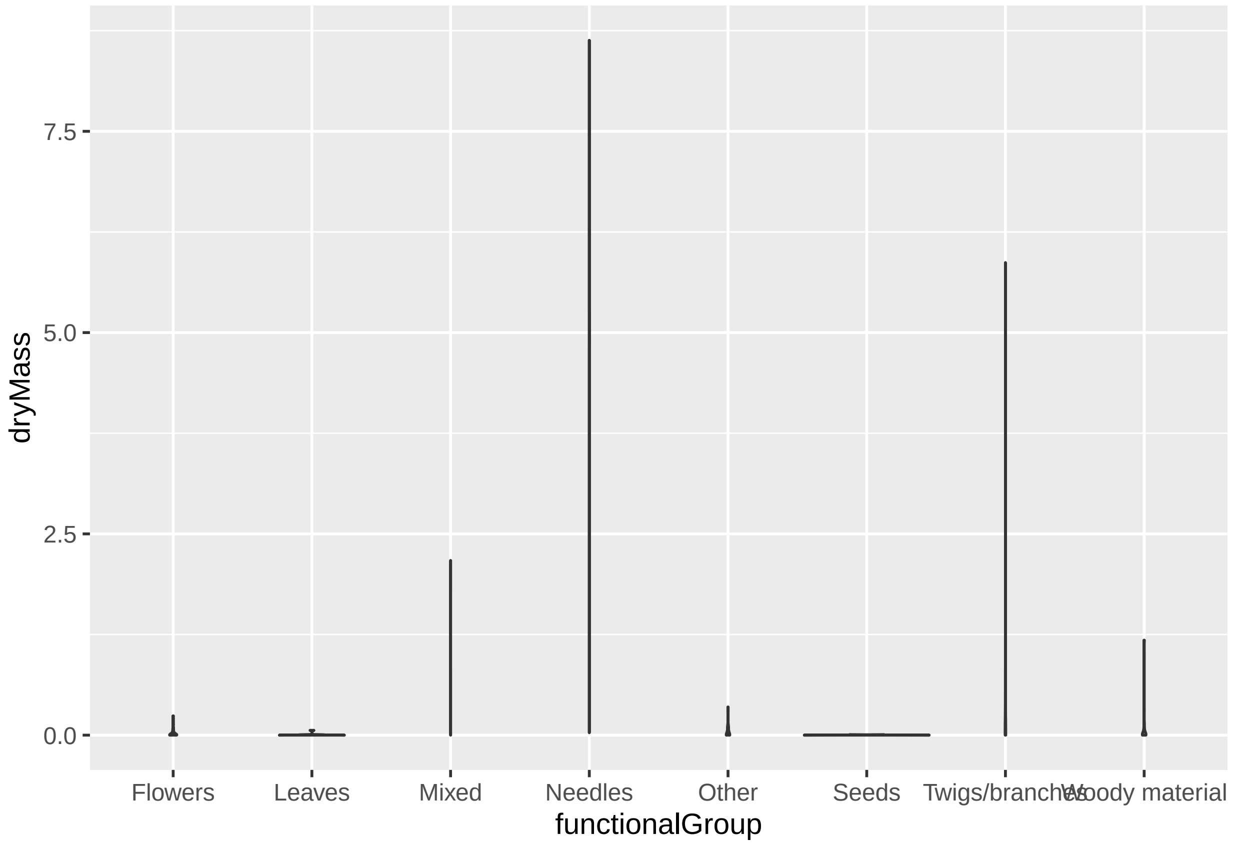


15. Using `geom_boxplot` and `geom_violin`, create a boxplot and a violin plot of dryMass by functionalGroup.

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



```
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: The box plot allows us to see the mean and interquartile ranges whereas the violin plot shows us the ammount of points in those ranges. The violin plot is not very useful in this situation becuase there is not an abundent amount of dry mass values for the functional group. Thus the box plot gives us a better visual becuase it shows the range of values.

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

> Answer: Needles and "mixed" tend to have the highest biomass at the sites.