Assignment 3: Data Exploration

Jackie Van Der Hout, Section #1

OVERVIEW

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

Directions

1. Change "Student Name, Section #" on line 3 (above) with your name and section number.
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., "FirstLast_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. **Be sure to add the `stringsAsFactors = TRUE` parameter to the function when reading in the CSV files.**

```
getwd()
```

```
## [1] "C:/Users/Jackie/Box/Classes Spring 2022/Environmental Data Analytics/Environmental_Data_Analytic
```

```
library(tidyverse)
setwd("~/../Box/Classes Spring 2022/Environmental Data Analytics/")
Neonics <- read.csv("/Users/Jackie/Box/Classes Spring 2022/Environmental Data Analytics/Environmental_Da
#stringAsFactors changes characters to factor for analysis
Litter <- read.csv("/Users/Jackie/Box/Classes Spring 2022/Environmental Data Analytics/Environmental_Dat
```

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 effects of insecticides on insects would provide insights for applicators for what time of year and insect life stages the insecticide should be applied in order to be most effective. Additionally, it may be useful to know if a certain insecticide has effects on unintended beneficial insects, such as pollinators and benthic macroinvertibrates and other non-target species.

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: Long term ecological monitoring of litter and woody debri provides several applications. Forest managers and scientists may be interested in the biogeochemistry of forest litter to understand the factors that influence nutrient decomposition and cycling. Additionally, comparisons of nutrient cycling in different ecosystems and climates and the effects of climate change on decomposition rates and nutrient levels may be of interest to scientists. Forest litter decomposition not only stays within the forest ecosystem in which it falls but also travels through fluvial systems to downstream sites and has an impact on other both aquatic and terrestrial ecosystem biogeochemistry.

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: The National Ecological Observation Network (NEON) collects data on woody debri and litter both elevated and group traps. * Sampling is conducted at sites in the NEON network that have vegetation with a minumum heigh of 2m tall. *Airshed vegetation density is used to determine the sampling plot size for litter collection, with more plots used in less saturated plots.* Litter trap placement is randomized or targetted within a grid depending on the target vegetation type, while also making sure a minumum distance between sampling locations is maintained to rule out covariance between samples.

Obtain basic summaries of your data (Neonics)

5. What are the dimensions of the dataset? A: 4623 x 30

```
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?

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

```
##       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 effects studied are Population, Mortality, Behavior, Feeding behavior, Reproduction, Avoidance and Development. The effects of this insecticide on insects in these categories could be of interest to researches because they provide insight into the physiological effects of the insecticide on the basic life cycle functions of insects.

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.

```r
NeonicsSpecies <- summary(Neonics$Species.Common.Name)
sort(NeonicsSpecies, decreasing = TRUE)
```

```
##                          (Other)                        Honey Bee
##                              670                              667
##                   Parasitic Wasp            Buff Tailed Bumblebee
##                              285                              183
##              Carniolan Honey Bee                       Bumble Bee
##                              152                              140
##                 Italian Honeybee                  Japanese Beetle
##                              113                               94
##                Asian Lady Beetle                   Euonymus Scale
##                               76                               75
##                         Wireworm                European Dark Bee
##                               69                               66
##                Minute Pirate Bug             Asian Citrus Psyllid
##                               62                               60
##                    Parastic Wasp           Colorado Potato Beetle
##                               58                               57
##                  Parasitoid Wasp              Erythrina Gall Wasp
##                               51                               49
##                     Beetle Order      Snout Beetle Family, Weevil
##                               47                               47
##         Sevenspotted Lady Beetle                   True Bug Order
##                               46                               45
##            Buff-tailed Bumblebee                     Aphid Family
##                               39                               38
##                   Cabbage Looper             Sweetpotato Whitefly
##                               38                               37
##                    Braconid Wasp                     Cotton Aphid
##                               33                               33
##                   Predatory Mite           Ladybird Beetle Family
##                               33                               30
##                       Parasitoid                    Scarab Beetle
##                               30                               29
##                    Spring Tiphia                      Thrip Order
##                               29                               29
##               Ground Beetle Family                 Rove Beetle Family
##                                 27                                 27
##                      Tobacco Aphid                       Chalcid Wasp
##                                 27                                 25
##             Convergent Lady Beetle                      Stingless Bee
##                                 25                                 25
##                  Spider/Mite Class                Tobacco Flea Beetle
##                                 24                                 24
##                   Citrus Leafminer                    Ladybird Beetle
##                                 23                                 23
##                          Mason Bee                           Mosquito
##                                 22                                 22
##                      Argentine Ant                             Beetle
##                                 21                                 21
##         Flatheaded Appletree Borer               Horned Oak Gall Wasp
##                                 20                                 20
##                 Leaf Beetle Family                  Potato Leafhopper
##                                 20                                 20
##         Tooth-necked Fungus Beetle                       Codling Moth
##                                 20                                 19
##          Black-spotted Lady Beetle                       Calico Scale
##                                 18                                 18
##                Fairyfly Parasitoid                        Lady Beetle
##                                 18                                 18
##             Minute Parasitic Wasps                          Mirid Bug
##                                 18                                 18
##                   Mulberry Pyralid                           Silkworm
##                                 18                                 18
##                     Vedalia Beetle              Araneoid Spider Order
##                                 18                                 17
##                          Bee Order                     Egg Parasitoid
##                                 17                                 17
##                       Insect Class           Moth And Butterfly Order
##                                 17                                 17
##       Oystershell Scale Parasitoid Hemlock Woolly Adelgid Lady Beetle
##                                 17                                 16
##              Hemlock Wooly Adelgid                               Mite
##                                 16                                 16
##                        Onion Thrip              Western Flower Thrips
##                                 16                                 15
##                       Corn Earworm                  Green Peach Aphid
##                                 14                                 14
##                          House Fly                          Ox Beetle
##                                 14                                 14
##                 Red Scale Parasite                 Spined Soldier Bug
##                                 14                                 14
##              Armoured Scale Family                   Diamondback Moth
##                                 13                                 13
##                      Eulophid Wasp                  Monarch Butterfly
##                                 13                                 13
##                      Predatory Bug              Yellow Fever Mosquito
##                                 13                                 13
##                Braconid Parasitoid                       Common Thrip
##                                 12                                 12
##       Eastern Subterranean Termite                             Jassid 
##                                 12                                 12 
##                         Mite Order                          Pea Aphid 
##                                 12                                 12 
##                   Pond Wolf Spider           Spotless Ladybird Beetle 
##                                 12                                 11 
##             Glasshouse Potato Wasp                           Lacewing 
##                                 10                                 10 
##            Southern House Mosquito             Two Spotted Lady Beetle 
##                                 10                                 10 
##                         Ant Family                       Apple Maggot 
##                                  9                                  9
```

```
#you can also do this, which just gives top four:
head(summary(Neonics$Species.Common.Name))
```

```
##            Honey Bee        Parasitic Wasp Buff Tailed Bumblebee 
##                  667                  285                  183 
##   Carniolan Honey Bee           Bumble Bee      Italian Honeybee 
##                  152                  140                  113
```

> Answer: Other than the "other" category, the top six species are Honey Bee, Parasitic Wasp Buff, Tailed Bumblebee, Carniolan Honey Bee, the Bumblee Bee and the Italian Honeybee. With the exception of the wasp, these are all pollinator species beneficial to crop plants and native plants. The bees and wasps may be being studied frequently to ensure that insecticides applied do not adversely affect these populations.

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"
```

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

```
## [1] 27.2 19.7 47   25   13   268 
## 1006 Levels: ~10 ~30/ ~40/ ~41 <0.0004 <0.025 <0.088 <0.5 <1.5 <10/ ... NR/
```

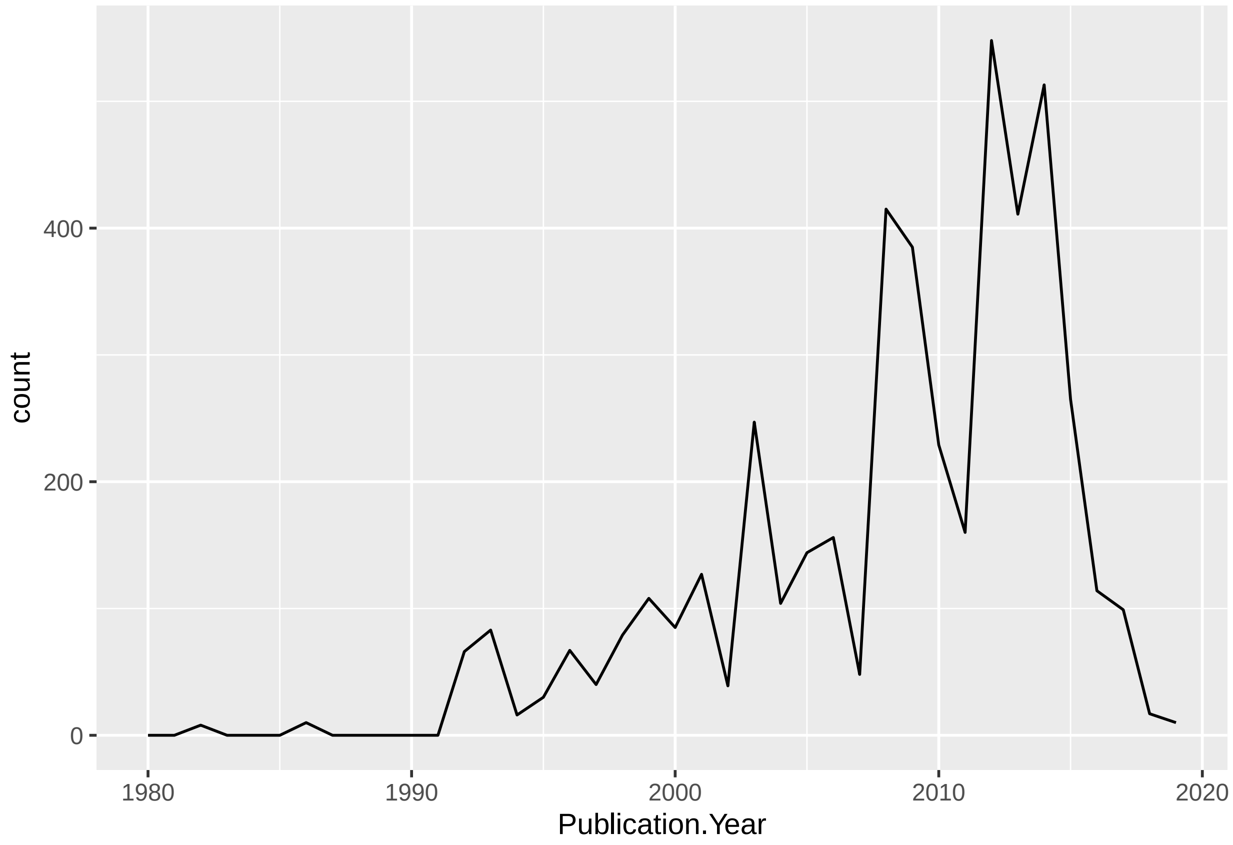
> Answer: The class of this variable is "factor". The reason it is a factor instead of a number is because the numbers have sympols in them such as > and ~ which when the dataset was imported using the stringsAsFactors = TRUE command labelled them as factors.

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 = 40)+ #chose 40 after trying many other values for bes
  scale_x_continuous(limits = c(1980, 2019))
```

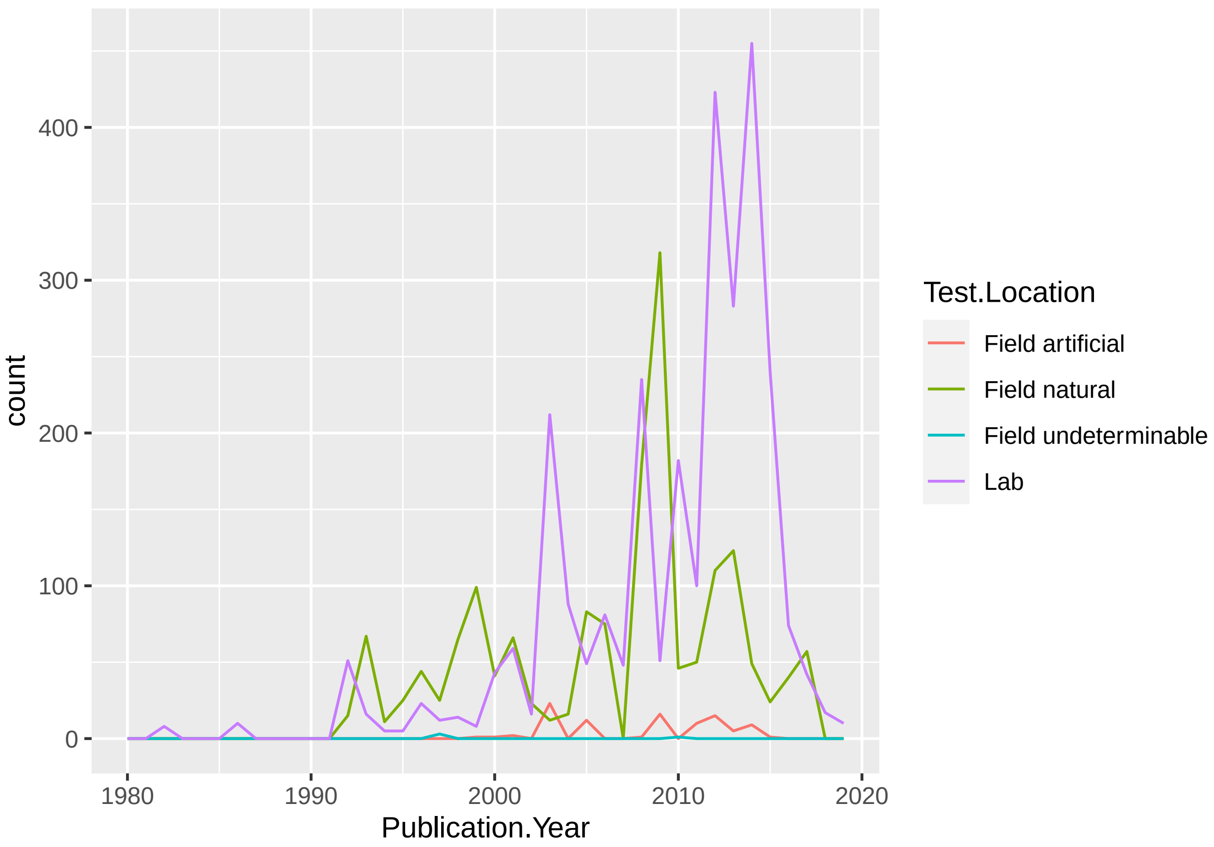
```
## Warning: Removed 2 row(s) containing missing values (geom_path).
```



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 = 40)+
  scale_x_continuous(limits = c(1980, 2019))
```

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

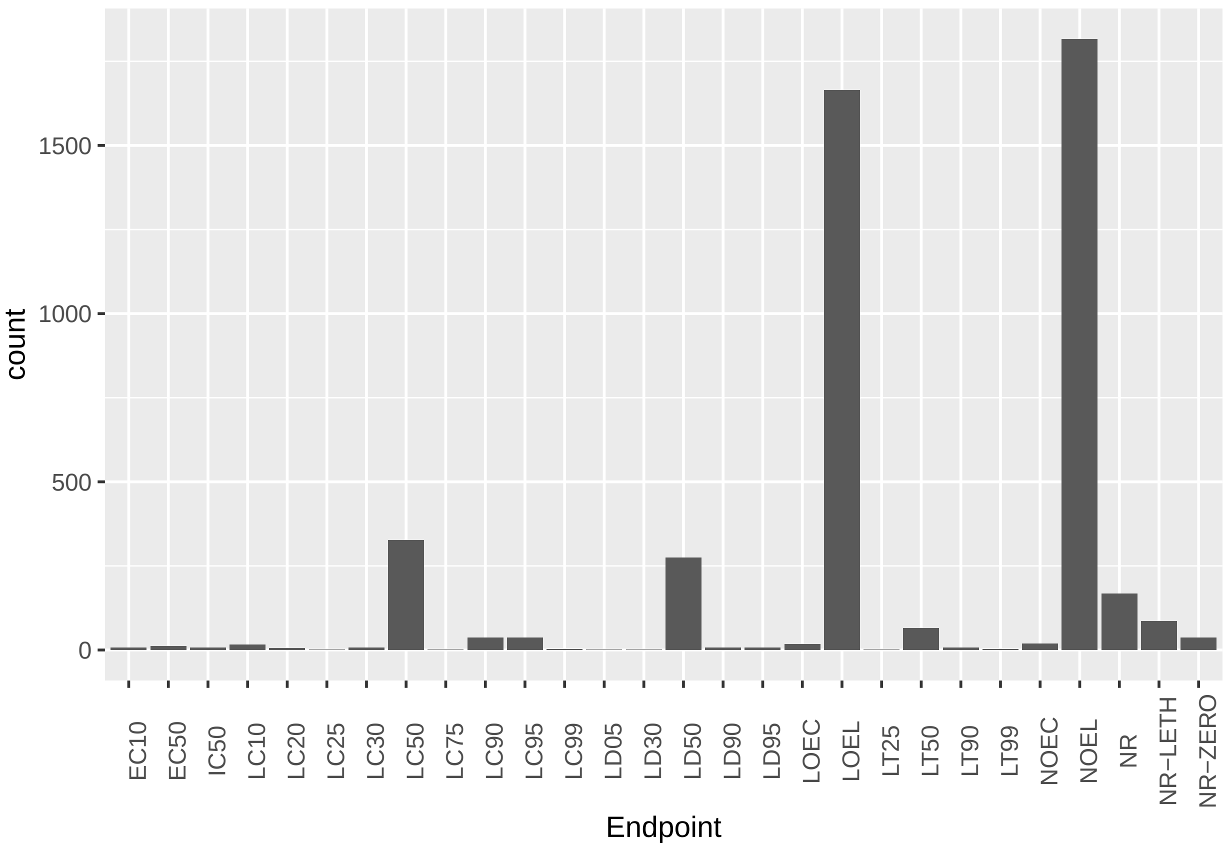


Interpret this graph. What are the most common test locations, and do they differ over time?

> Answer: From the 1980s until the early 2000s, the most common testing locations were "field natural". From the early 2000s until the present the lab studies have become the dominant study type and location.

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)+
  geom_bar(aes(Endpoint))+
  theme(axis.text.x = element_text(angle = 90)) #rotates text on X axis
```



Answer: The two most common endpoints by far are LOEL and NOEL. LOEL stands for "lowest observable effect level" which is the lowest concentration that produces effects that are significantly different than the controls, in a terrestrial environment. NOEL is also a terrestrial measurement and stands for "No Observable Effect Level" and is inversely the highest concentration which does not produce significantly different responses from those of the 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) #currently factor
```

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

```
library(lubridate)
```

```
##
## Attaching package: 'lubridate'
```

```
## The following objects are masked from 'package:base':
##
##     date, intersect, setdiff, union
```

```
Litter$collectDate <- ymd(Litter$collectDate) #overwritting litter column as a date
class(Litter$collectDate) #now R knows it is a date
```

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

```
unique(Litter$collectDate)#collected on two dates, "2018-08-02" and "2018-08-30"
```

```
## [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
```

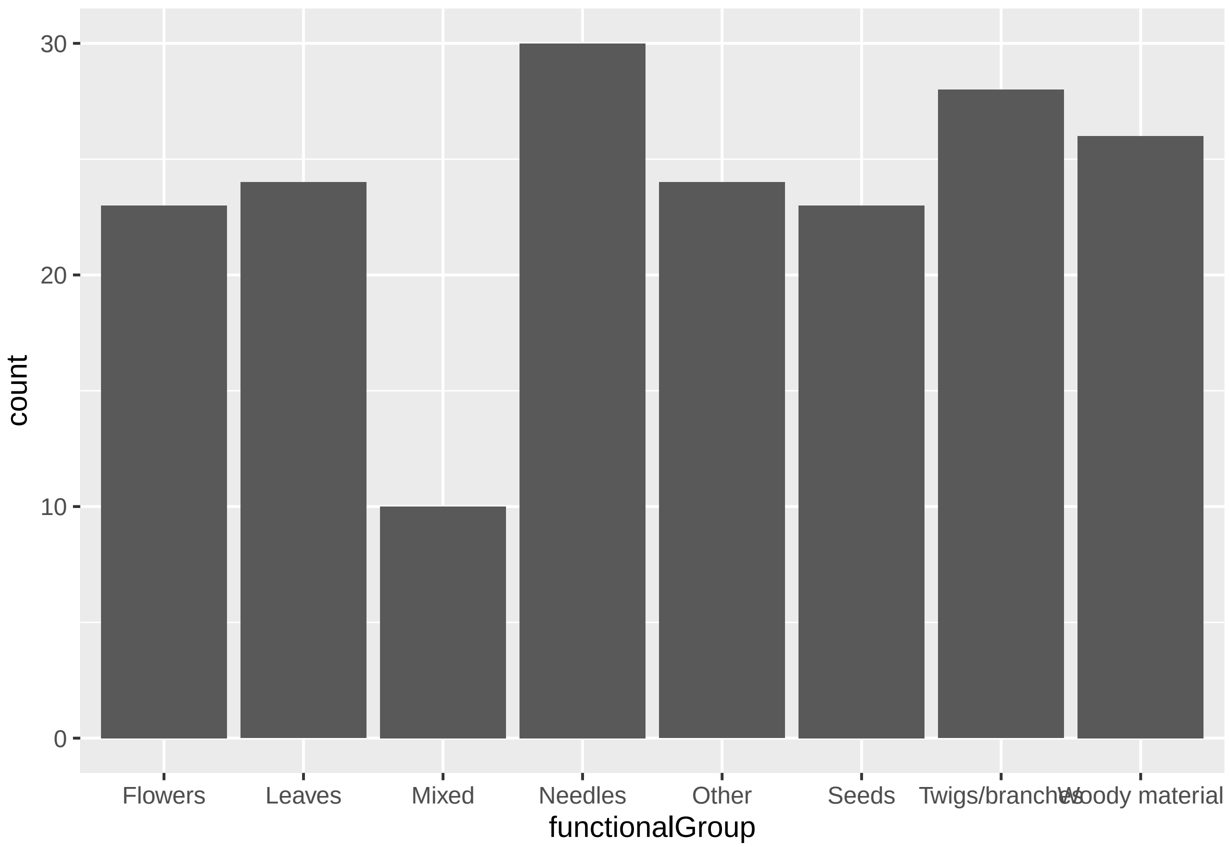
```
summary(Litter$plotID)
```

```
## 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 sample plots were used at Niwot Ridget. The unique function shows that there were 12 unique sites but not more information, whereas the summary function explains how many samples were taken at each of the 12 sites.

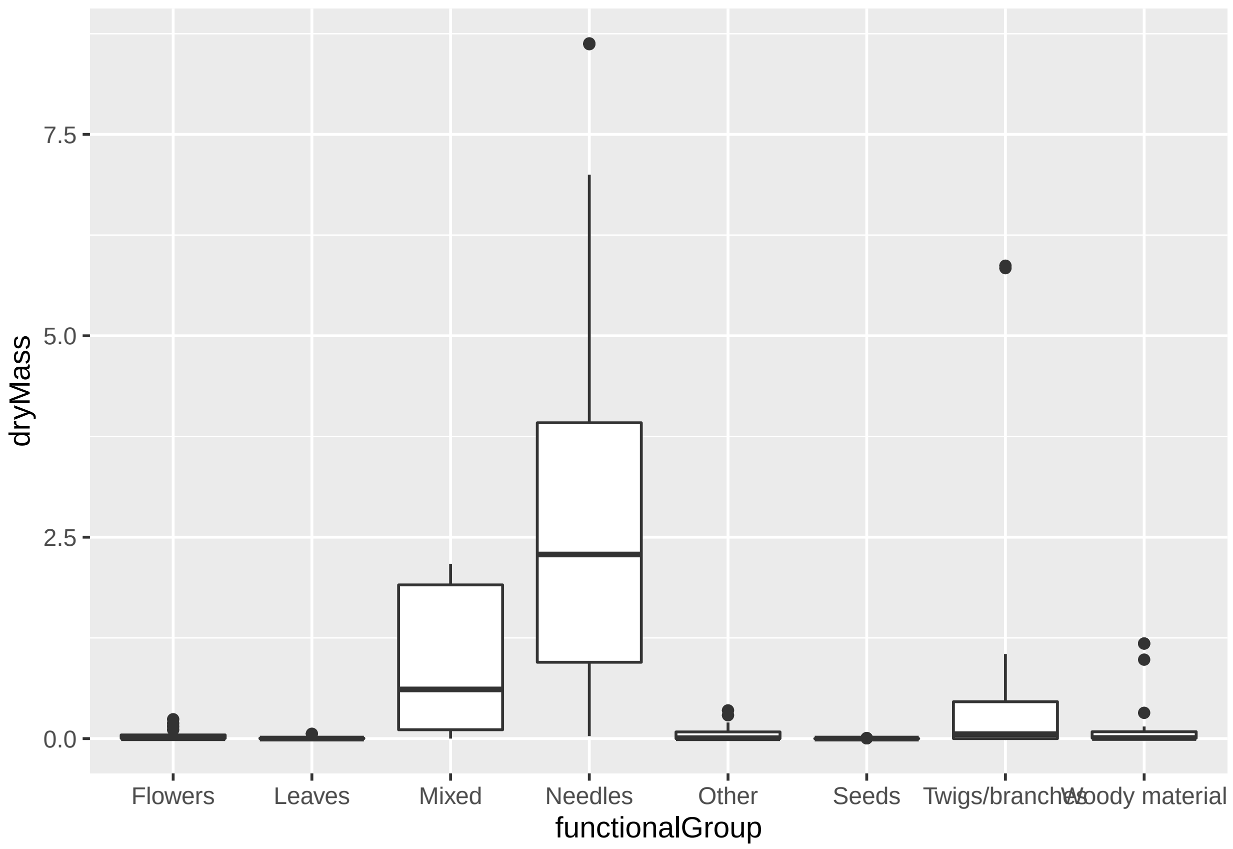
14. Create a bar graph of functionalGroup counts. This shows you what type of litter is collected at the Niwot Ridge sites. Notice that litter types are fairly equally distributed across the Niwot Ridge sites.

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

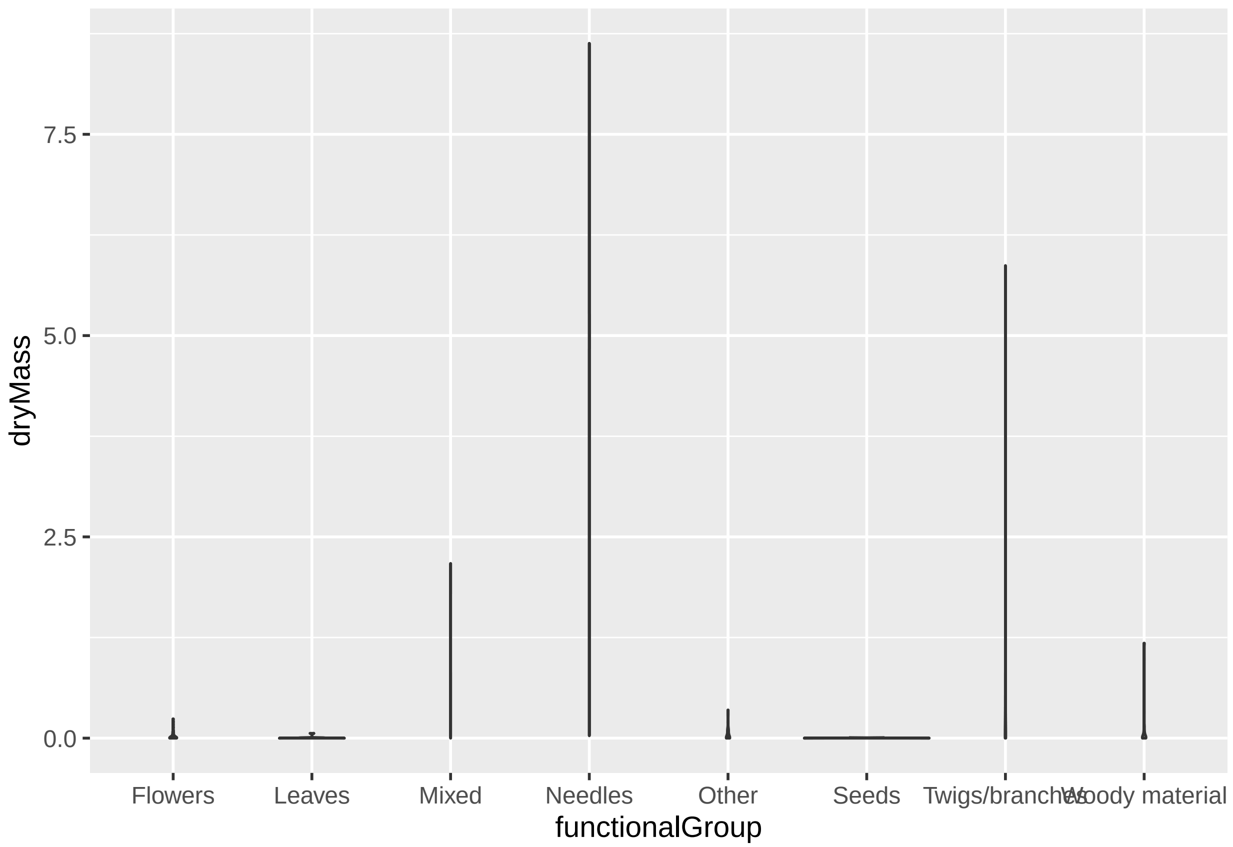


15. Using `geom_boxplot` and `geom_violin`, create a boxplot and a violin plot of dryMass by functionalGroup.

```
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: Box and whisker plots show distributions of the median values for continuous numerical variables, shown in the interqquatrile range. They show outliers from the IQR and the general pattern of the distribution and are a good way to quickly compare two variables, which is why they work well here. However, the violin plots are meant to show density distributions within the box component of the box and whisker plots, however here the distribution does not appear, showing only a line of variability in range.

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

> Answer: Needles tend to have the higest biomass.