Assignment 3: Data Exploration

Kathlyn MacDonald, Section #2

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.**

```
#Checking the working directory
getwd()
```

```
## [1] "C:/Users/kmac9/OneDrive/Documents/Duke/Year 1/Spring 2022/ENV Data/Environmental_Data_Analytics_
```

```
#loading packages
library(tidyverse)
#loading the data
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
```

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 are toxic to certain bugs, and keeping track of which insecticides are placed and where is useful for keeping track of species degredation. These pesticides also impact other organismal life, and should be carefully tracked.

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: Tracking the litter/debris that fall to the ground is important for tracking soil quality and ecosystem function. Ecosystem functions that are can be tracked include decomposition, carbon storage and nutrient cycling.These values can help estimate net priary productivity at the plot, site, and continental scale.

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 has a diameter of <2cm and length of <50cm, whereas fine wood debris has a diameter of <2cm and a lenght >50cm* Used elevated 0.5^2m mesh baskets to trap debris *Sampling is done in locations that contain woody vegetation >2m tall

Obtain basic summaries of your data (Neonics)

5. What are the dimensions of the dataset?

```
dim(Neonics)
```

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

```
#4623 rows, 30 columns
```

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?

```
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 and mortality (by far). These are most interesting and thus most studied, because they offer insight on how the neonics effect the species in quesiton. Considering it is a chemical created to reduce insect population, it is understandable that the chemical would effect population levels and mortality rates.

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 most studied species are: honey bee, parasitic wasp, buff tailed bumblebee, carniolan honey bee, bumble bee, italian honey bee. Bees are of great importance to the farming community, because they are essential pollinators and the success of the farm is related to the bees ability to pollinate their plants. If pollinators populations go down too far, it would be detrimental to our farming community and thus food system. Parasitic wasps are also an important study species because they are harmful to the bee population and can cause colony collapse.

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"
```

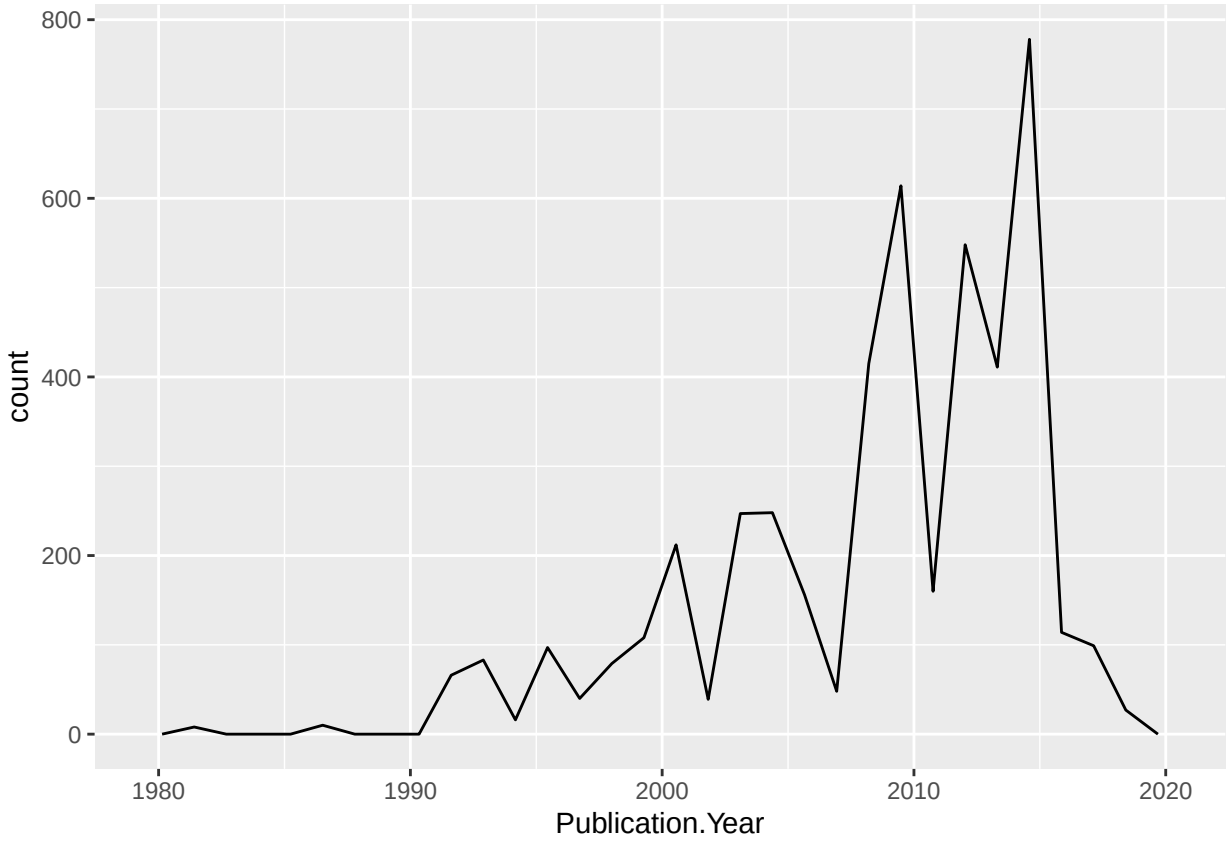
Answer: These concentrations are listed as a "factor" data set, so that they can be grouped into categories depending on the concentration. There are many concentrations at the same value, so these can be grouped together.

Explore your data graphically (Neonics)

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

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

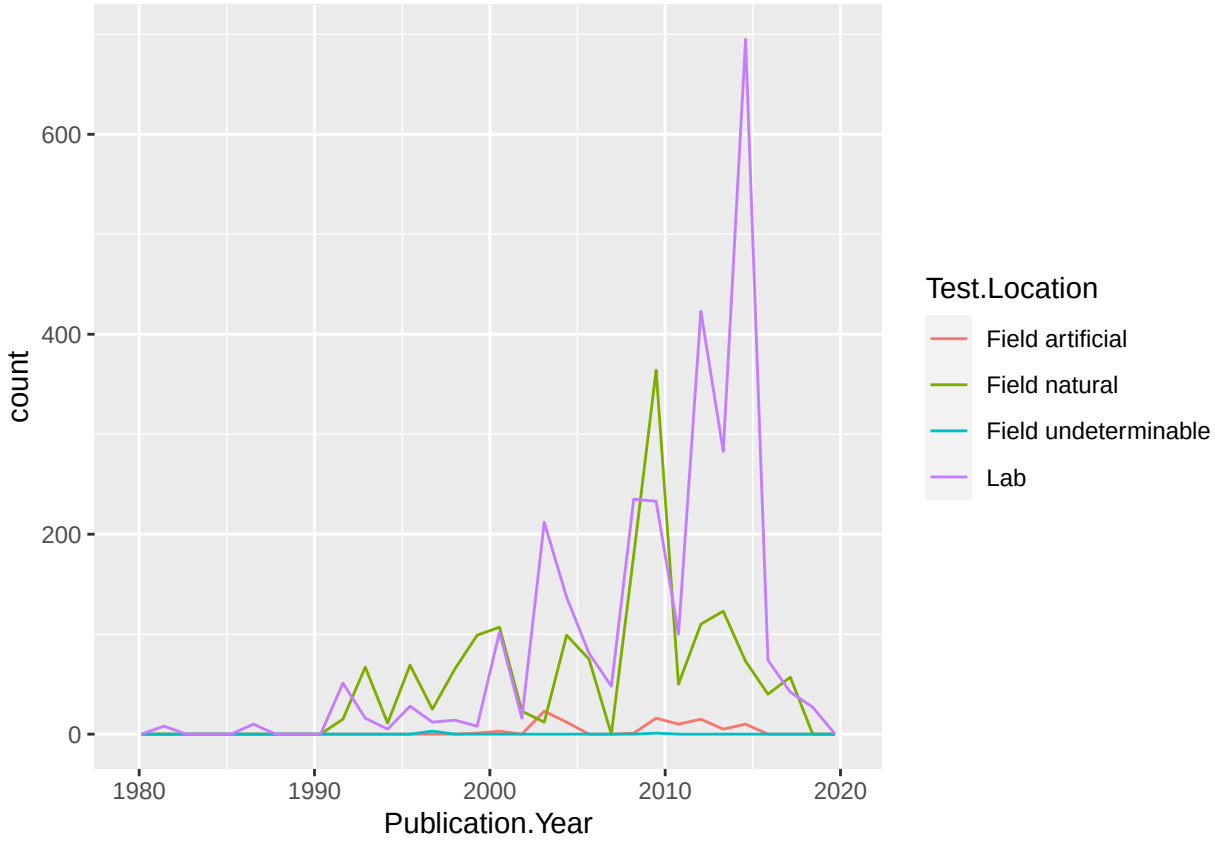
```
## `stat_bin()` using `bins = 30`. Pick better value with `binwidth`.
```



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

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

```
## `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?

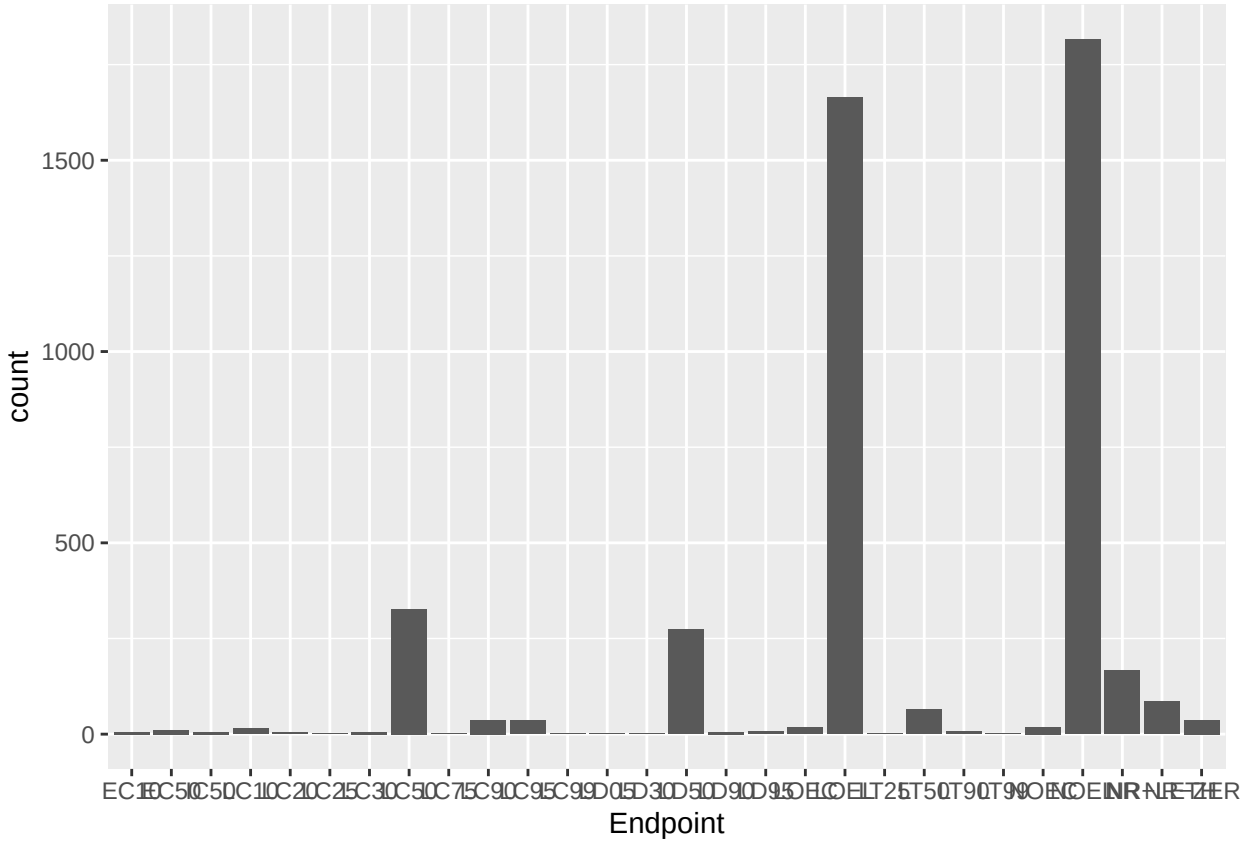
> Answer: The lab and out in the natural field are the most common locations. Testing in the field peaks earlier (in 2010) compared to in the lab which peaks in ~2015. Both grow over time, but eventually peak and drop in use.

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.

```
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
```

```
ggplot(Neonics, aes(x = Endpoint)) +
  geom_bar()
```



Answer: The two most common endpoints are LOEL and NOEL with greater than 1500 counts. LOEL is defined in the appendix as the lowest observable effect level (the lowest concentration that produces significant response). NOEL is defined as no observed effect level (the highest cocentration that does not produce significant response).

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.

```
#Classifying collectDate as a date not a factor
class(Litter$collectDate)
```

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

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

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

```
#What dates was litter sampled in August?
unique(Litter$collectDate)
```

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

Answer: August 2nd and 30th

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
## 12 Levels: NIWO_040.basePlot.ltr ... NIWO_067.basePlot.ltr
```

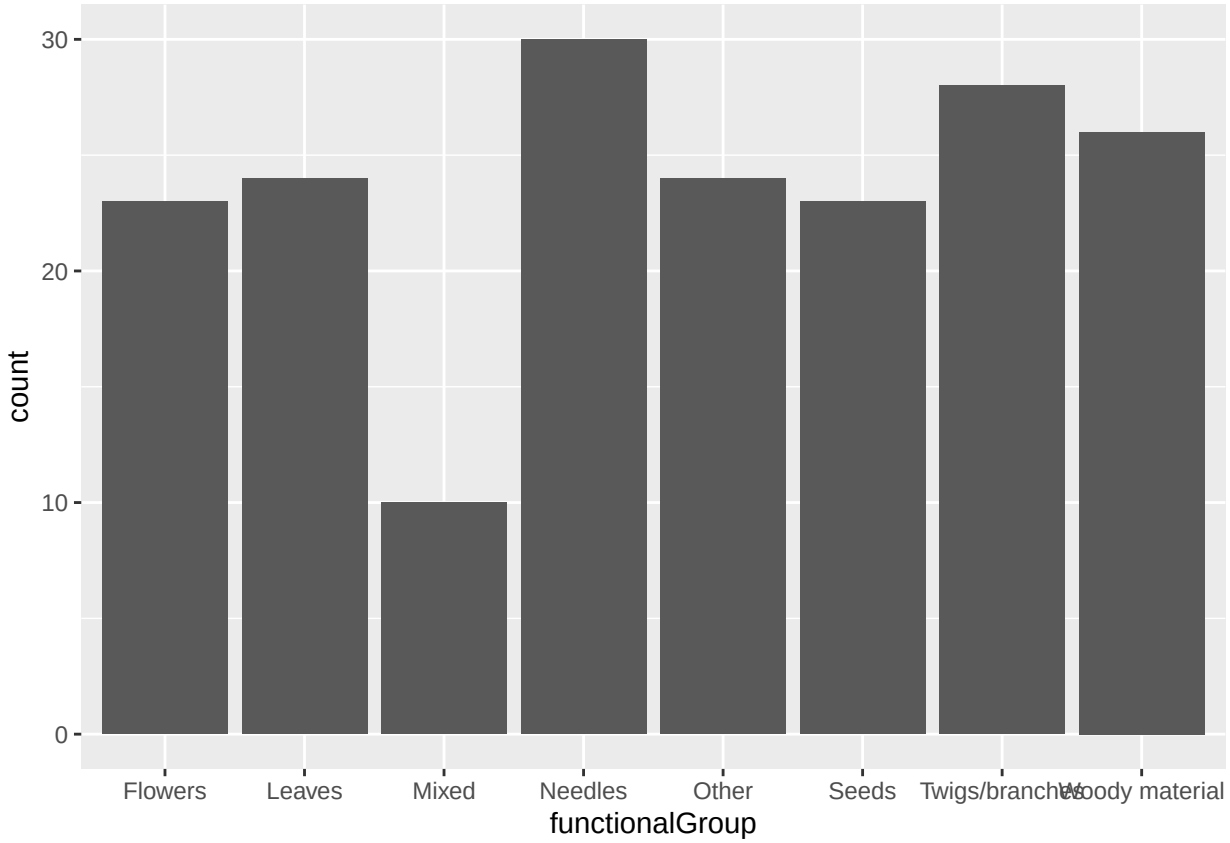
```
summary(Litter$namedLocation)
```

```
## 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: There are 12 different plots that were sampled at Niwot Ridge. The unique function removes all the duplicates from the selected data frame. Using this I am able to find all the different or 'unique' variables in the frame. The summary function, allows you to see the # of cases within each variable

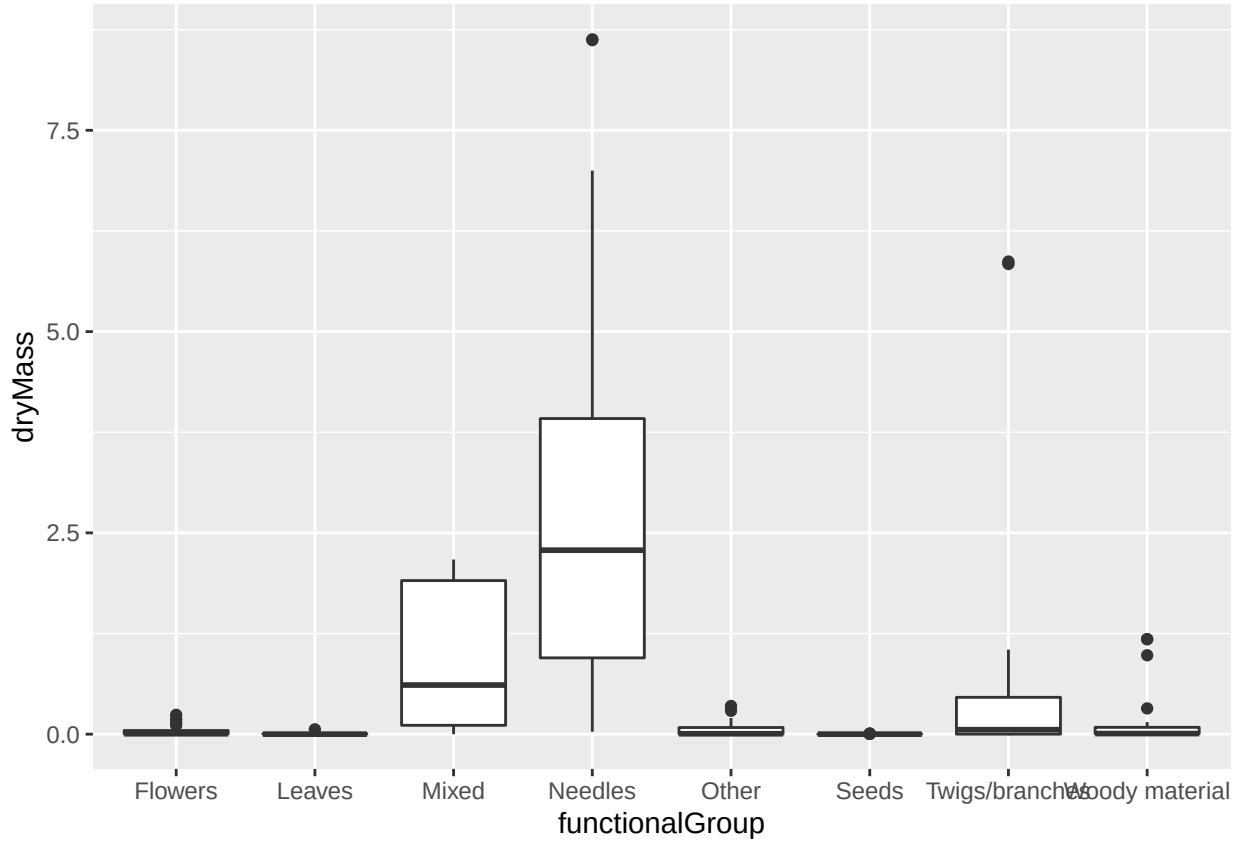
14. Create a bar graph of functionalGroup counts. This shows you what type of litter is collected at the Niwot Ridge sites. Notice that litter types are fairly equally distributed across the Niwot Ridge sites.

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

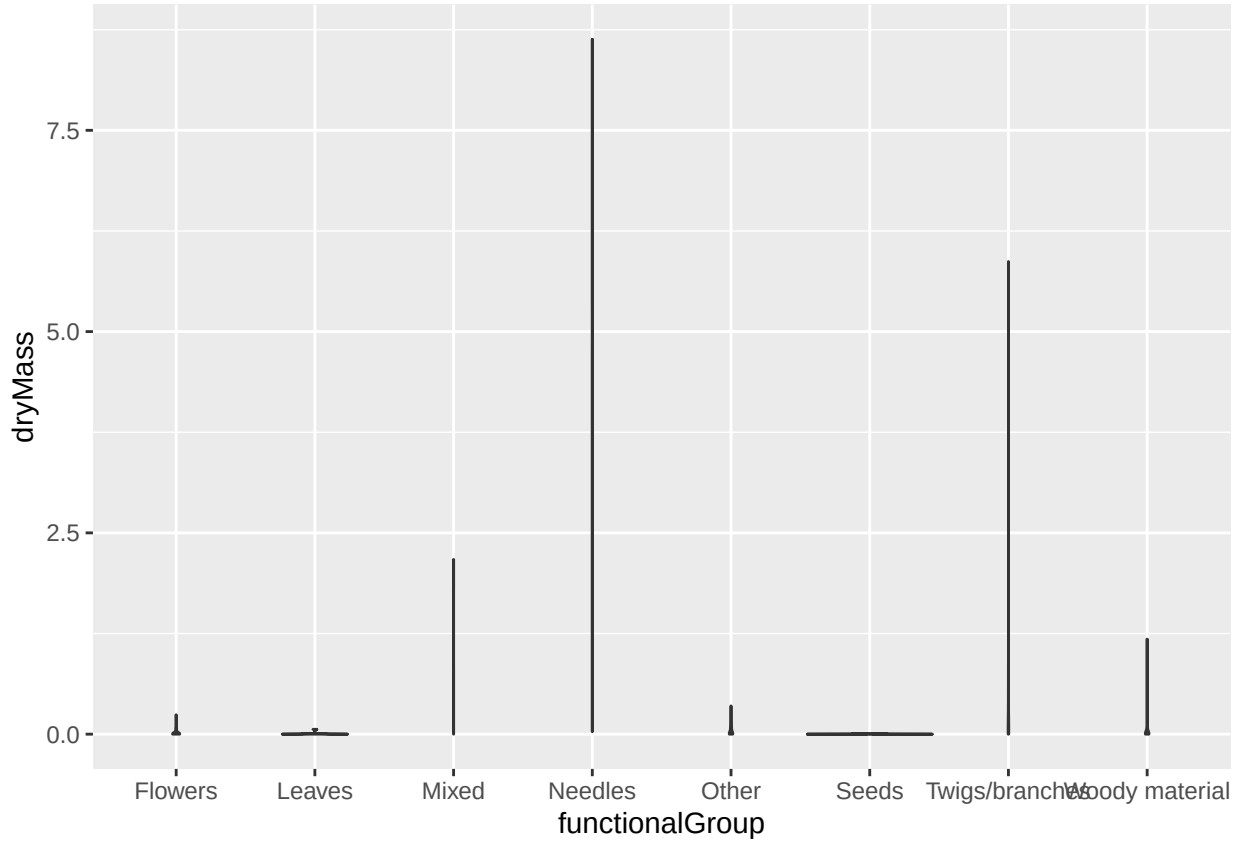


15. Using `geom_boxplot` and `geom_violin`, create a boxplot and a violin plot of dryMass by functionalGroup.

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




```
#violin plot
ggplot(Litter, aes(y=dryMass, x=functionalGroup)) +
  geom_violin()
```



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

> Answer: Violin plots allow you to see the density of each functional group at a certain mass. This is not very effective in measuring this data, because the mass is numerical and not categorical, meaning that the mass of each litter item is different for most every case. Because we looking at the plot on a small scale, the violin plot turns out to be a straight line. With the expception leaves and seeds, which have a very small mass, and are counted as zero.The boxplot allows you to see an average distribution. Which is more useful in this case.

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

> Answer: Needles and then mixed litter