Assignment 3: Data Exploration

Alicia Zhao

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 Tuesday, January 28 at 1:00 pm.

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] "/Users/mac/Desktop/Data Analytics/Environmental_Data_Analytics_2020/Assignments"
```

```
library(tidyverse)
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: We may be interested in studying the ecotoxicology of neonicotinoids on insects to see if they are in fact effective, given the harm that they may incur on the environment and secondarily, humans.

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 may be interested in studying litter and and woody debris to understand how nutrient concentrations are changing in the forest ecosystems.

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 debirs are collected from elevated and ground traps, respectively. One litter trap pair is deployed for every 400 m2 plot area, resulting in 1-4 trap pairs per plot. Ground traps are sampled once per year while 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 infrequent year-round sampling (1x every 1-2 months) at evergreen sites. * Masses reported following processes are reported at the spatial resolution of a single trap (trapID) and the temporal resolution of a single collection event (daysofTrapping). * Weights <.01g are reported are reported and may indicate presence of a given funcitonal group, identified in the sorting process, but not present at the detectable masses.

Obtain basic summaries of your data (Neonics)

5. What are the dimensions of the dataset?

```r
dim(Neonics)
```

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

Answer: Dimensions are 4623 rows (observations) and 30 columns (variables).

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

```r
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 commonly studied effects are population, mortality and behavior. Population and mortality give us an idea of the number of insects that are able to persist after an insecticide is used. Additionally, behavioral changes in insects could also provide evidence on how effective an insecticide is.

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
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 commonly studied species are the honey bee, the parasitic wasp and the buff tailed bumblebee. These species are all beneficial to agriculture, either through their ability to pollinate or their control agricultural pests.

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.)
```

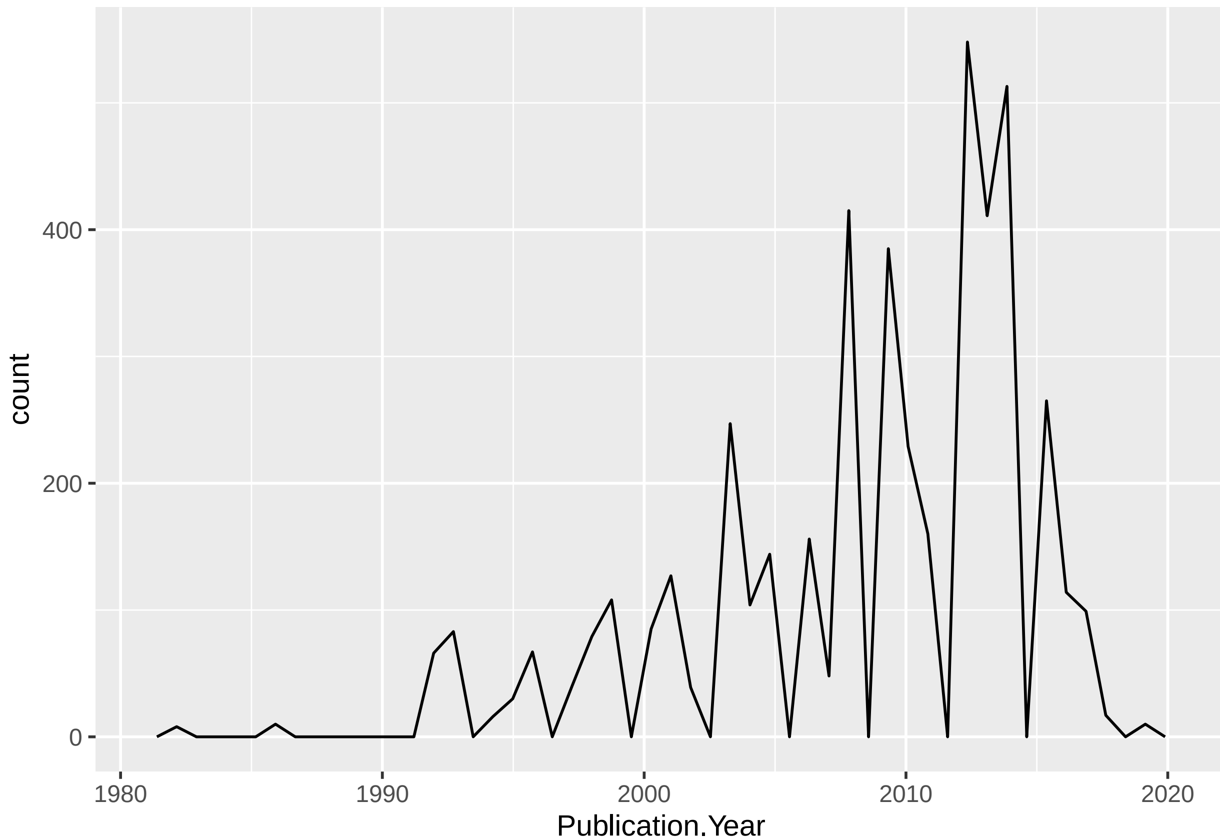
```
## [1] "factor"
```

Answer: The class is 'factor' because the data consists of numbers such as "144.0/", which are not read as numeric.

Explore your data graphically (Neonics)

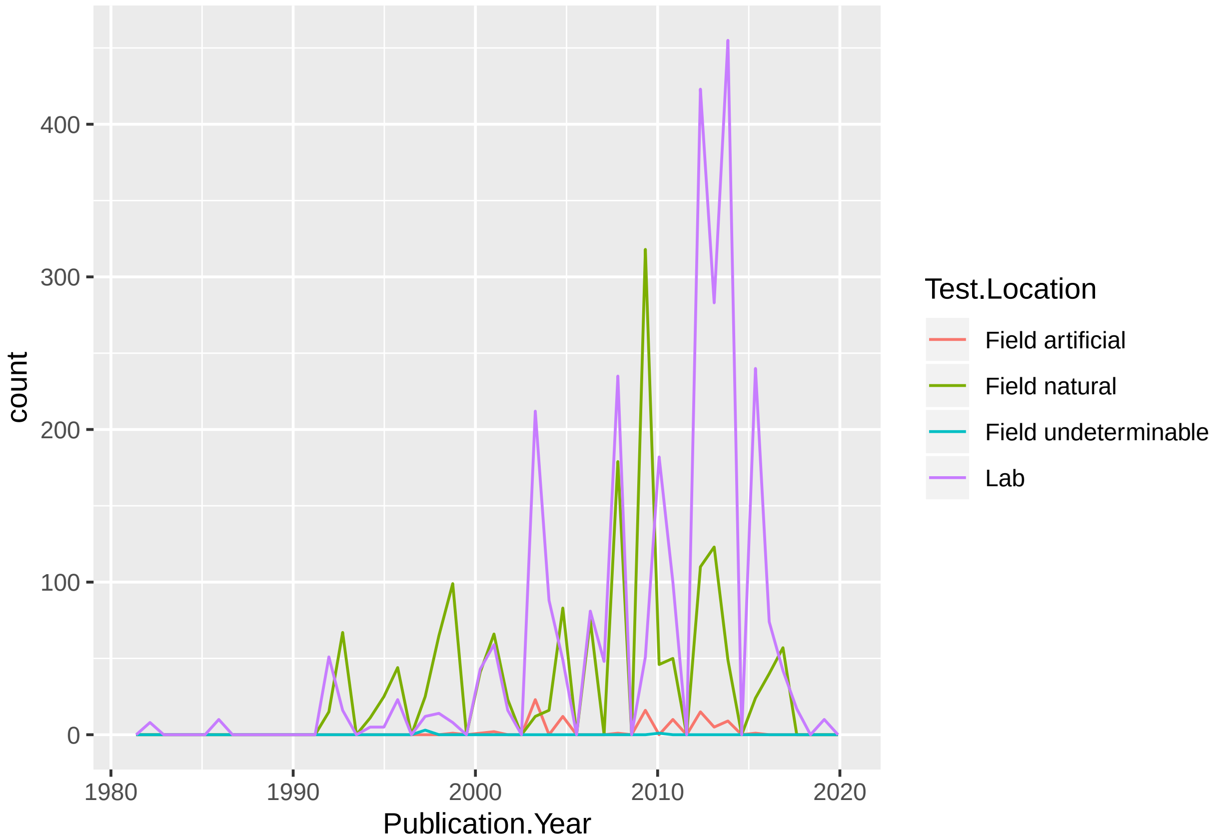
9. Using `geom_freqpoly`, generate a plot of the number of studies conducted by publication year.

```
ggplot(Neonics) +
  geom_freqpoly(aes(x = Publication.Year), bins = 50)
```



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 = 50)
```

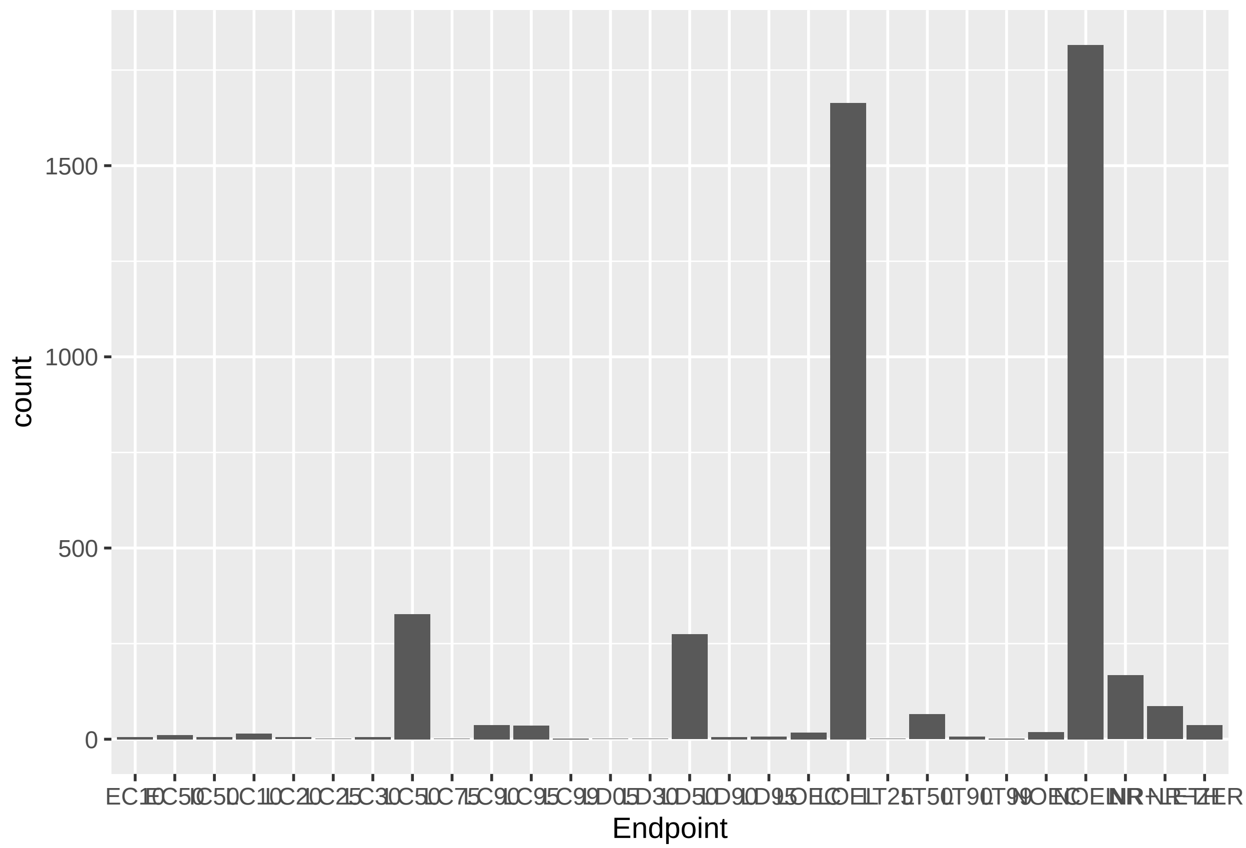


Interpret this graph. What are the most common test locations, and do they differ over time?

> Answer: According to the graph, the most common test locations are lab and field natural. The number of studies for these test locations varies over time. The number of studies for the lab setting peaked between approximately 2012 and 2015, and the number of studies for the field natural study peaked around 2009.

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()
```



Answer: The two most common end points are NOEL, followed by LOEL. NOEL(No-observable-effect-level) is defined as the highest dose (concentration) producing effects not significantly different from responses of controls according to author's reported statistical test. LOEL (Lowest-observable-effect-level) is dfined as the lowest dose (concentration) producing effects that were 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"
```

```
# Class of collectDate is a factor, not a date

Litter$collectDate <- as.Date(Litter$collectDate, format = "%Y-%m-%d")
# Change factor to date

class(Litter$collectDate)
```

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

```
# Confirm that collectDate is now classified as a date

unique(Litter$collectDate)
```

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

```r
# Use unique function to determine which dates litter was sampled in August 2018
```

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`?

```r
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
```

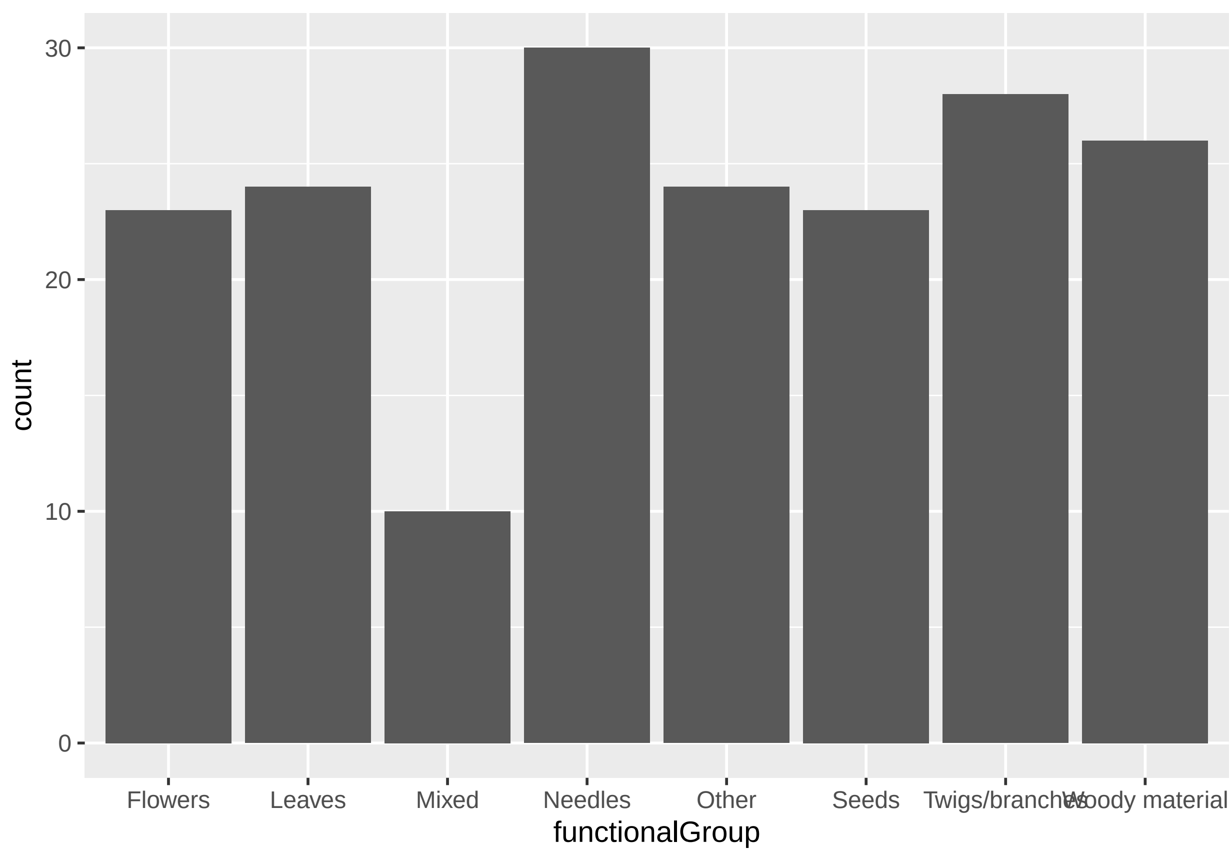
```r
summary(Litter$plotID)
```

```
## 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: The unique function tells you how many categories there are, while the summary function simply lists all of categories and their corresponding counts, without indicating how many categories there are in total.

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.

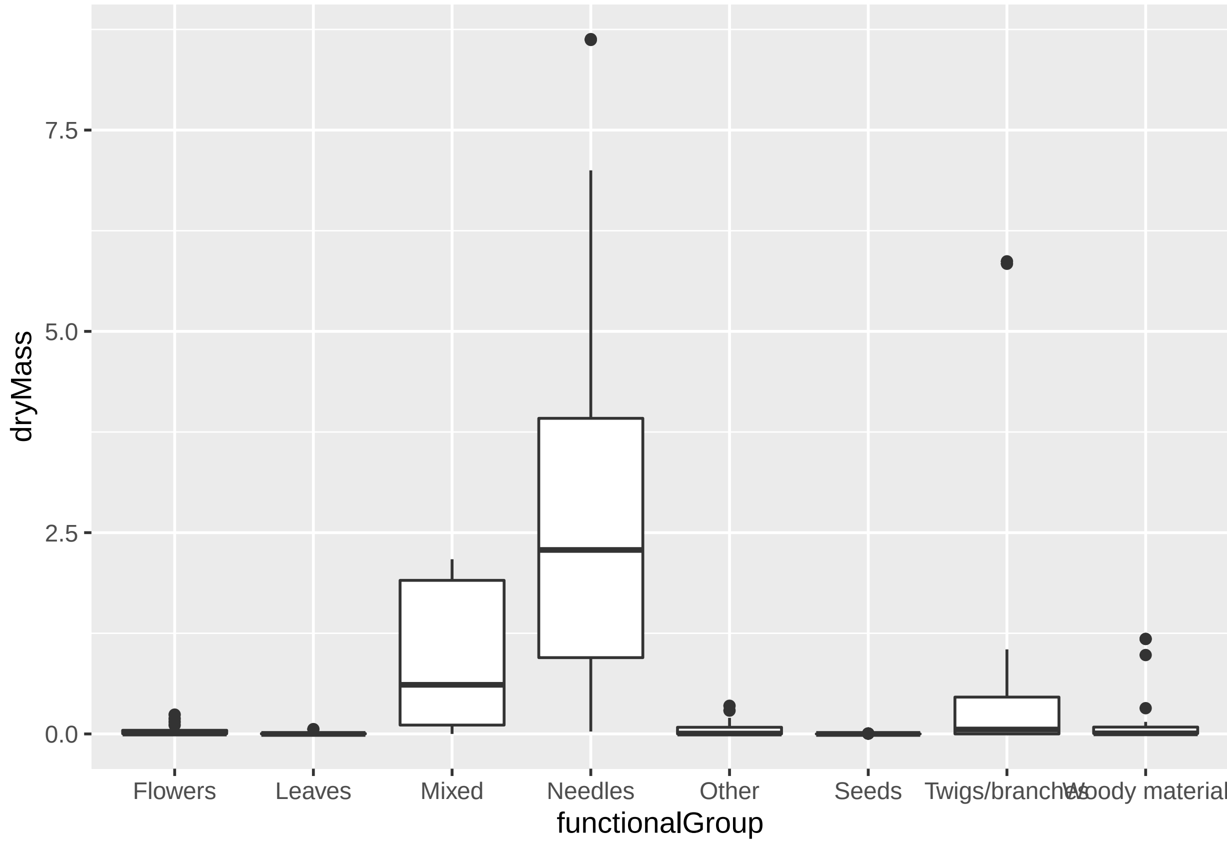
```r
ggplot(Litter, aes(x = functionalGroup)) +
  geom_bar()
```



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

Group.

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

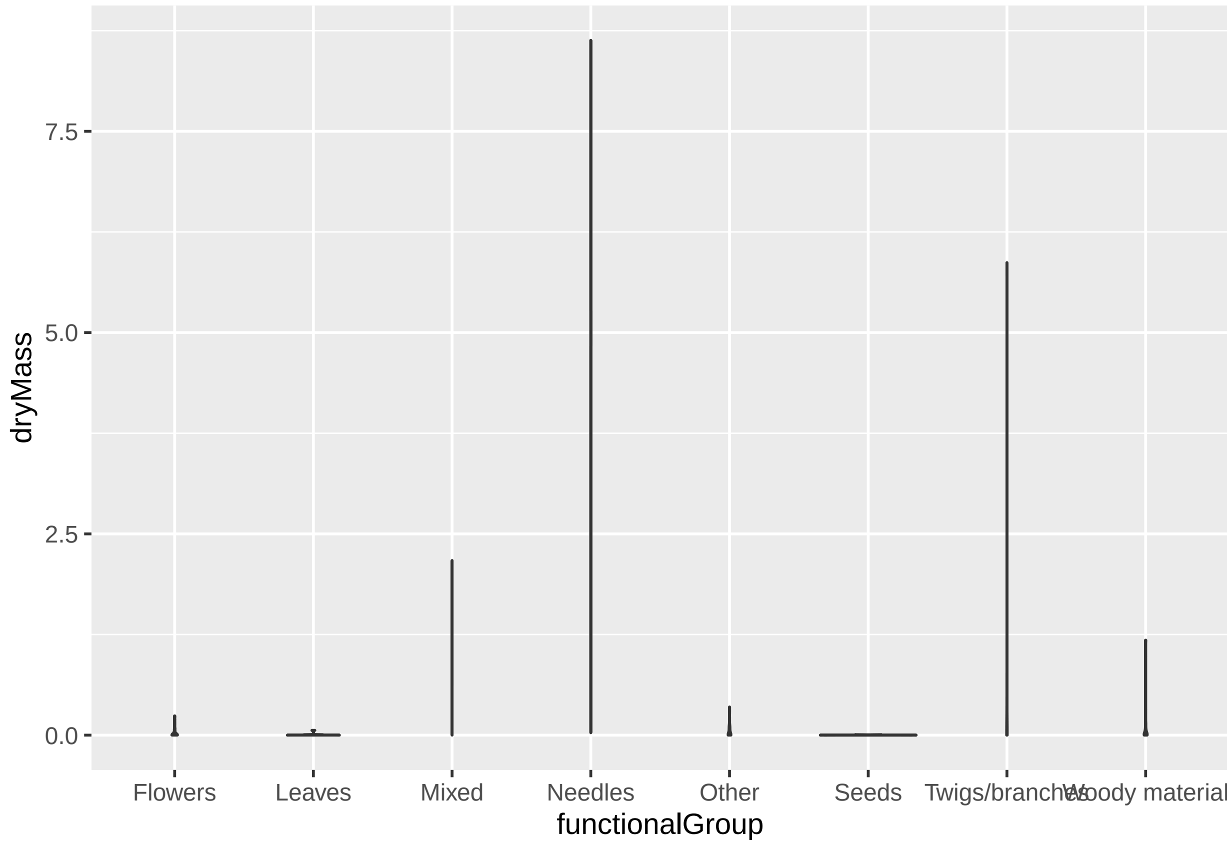



```r
ggplot(Litter) +
  geom_violin(aes( x = functionalGroup, y = dryMass),
  draw_quantiles = c(0.25, 0.5,0.75),
      scale="count")
```

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

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

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



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

> Answer: In this instance, the distribution of drymass varies significantly between the function groups. The boxplot allows us to more clearly see the median, outliers and IQR for some of the functional groups, whereas we cannot easily distinguish those points in the violin plot.

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

> Answer: Needles and Mixed litter tend to have the highest biomass at these sites.