Assignment 3: Data Exploration

Kelsie Roberton, 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.**

```
getwd()
```

```
## [1] "/Users/kelsieroberton/RStudio/Environmental_Data_Analytics_2022"
```

```
#install package
#install.packages("tidyverse")
library(tidyverse)

#import data sets

#Absolute file path (not recommended)
#read.csv(ECOTOX_Neonicotinoids_Insects_raw.csv)

#Relative file path (friendly for users regardless of machine!)
Neonics <-read.csv ("./Data/Raw/ECOTOX_Neonicotinoids_Insects_raw.csv", stringsAsFactors= TRUE)
Litter <-read.csv("./Data/Raw/NEON_NIWO_Litter_massdata_2018-08_raw.csv", stringsAsFactors = TRUE)
```

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: Neonicotinoids is type of pesticide used in agriculture to protect crops from insects. They can also be used for other purposes, such as killing insects in homes, controlling fleas on pets, and protecting trees from invasive insects (e.g. Emerald Ash Borer). Although neonicotinoids are considered low toxicity to mammals and humans in comparison to insecticides, ihoney bees exposed to neonicotinoids can experience severe effects: reduce taste sensitivity, slower mobility, and ultimately—death. It is one of the largest factors to be found causing the decline of the honey bee population. This is extremely detrimental since bees play a huge part in the support of tree, flower, and food plants. Without bees, there is a high likelihood that all of humankind would not be able to live (75% of the world's plants are flowered by bees),.

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: Wooden debris is the byproduct of decomposed logs on the forest floor. It provide nutrients to plants, habitat for wildlife, and food for insects and microorganisms. Firstly, wooden debris can provide a significant amount of organic matter to the soil, which is crucial for tree growth. The decaying matter has the capability to provide nitrogen, potassium, and phosphate into the soil to be reused by other plants. The health of the soil is directly correlated to the health of the trees. Secondly, wooden debris and organic litter material help protect and trap carbon in the soil; increasing the amount of stored carbon (SARE.org).

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: *Sampling for litter and fine woody debris occurs only in tower plots. The locations of tower plots are selected randomly within the 90% plux footprint of the primary and secondary airsheds.* Plot edges must be separated by a distance of 150% of one edge of the plot * Trap placement within plots may be either targete or randomized, depending on the vegetation. In sites with 50% aerial cover of woody vegetation >2m in height, placement of litter traps is random and utilizes the randomized list of grid cell locations. In sites with <50% cover of woody vegetation, sites with patchy vegetation, trap placement is targeted such that only areas beneath qualifying vegetation are considered for trap placement.

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 "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: A couple of the most common effects studied are population and mortality. These are both important variables to consider to understand if the invertabrate population size is affected by the chemical stressor, Neonicotinoids. In turn, it may be useful to see if the presence of the insecticide resulted in death.

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: Five out of the six most common species are bees. The other species is classified as a 'wasp.' However, both bees and wasps are members of the Hymenoptera order of insects. Wasps and honey bees do vary in their physical bodies, so it is quite interesting to see how neonicotinoids may affect a member of the same insect family based on its physical composition.

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"
```

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

```
##    0.37/      10/      NR/       NR        1     1023    0.40/       2/
##      208      127      108       94       82       80       69       63
##       10   0.053/      100      50/     0.5/     0.03    0.05/     0.45
##       62       59       56       51       45       44       43       43
##     0.1/    0.45/     1.0/    2.27/       50    0.125     500/      0.5
##       42       40       40       40       36       33       33       32
##   0.048/    0.15/       1/       48    25.0/      12/    0.027      2.4
##       30       30       30       30       28       27       26       26
##     0.2/    0.56/     100/        3    0.01/    1000/       3/    0.336
##       25       24       23       23       22       22       22       21
##     1.5/     0.05      1.5    2.60/    20.0/        6    6.80/    62.5/
##       21       20       20       20       20       20       20       20
##    0.005     0.4/    0.18/     0.3/     1000       40 0.00355/      0.1
##       18       18       17       17       17       17       16       16
##      0.4     150/      300      80/    0.053     0.24     0.28     125/
##       16       16       16       16       15       15       15       15
##        9   0.0001  0.0004/   0.084/     0.15      0.6    12.5/   144.0/
##       15       14       14       14       14       14       14       14
##     350/    40.0/      48/       56      84/    0.17/      125       14
##       14       14       14       14       14       13       13       13
##       16       17   0.047/    0.25/    0.28/    1.28/    1.81/      112
##       13       13       12       12       12       12       12       12
##      150     2.5/       25      60/      75/    0.02/   0.025/     0.29
##       12       12       12       12       12       11       11       11
##    37.5/       4/        5  (Other)
##       11       11       11     1817
```

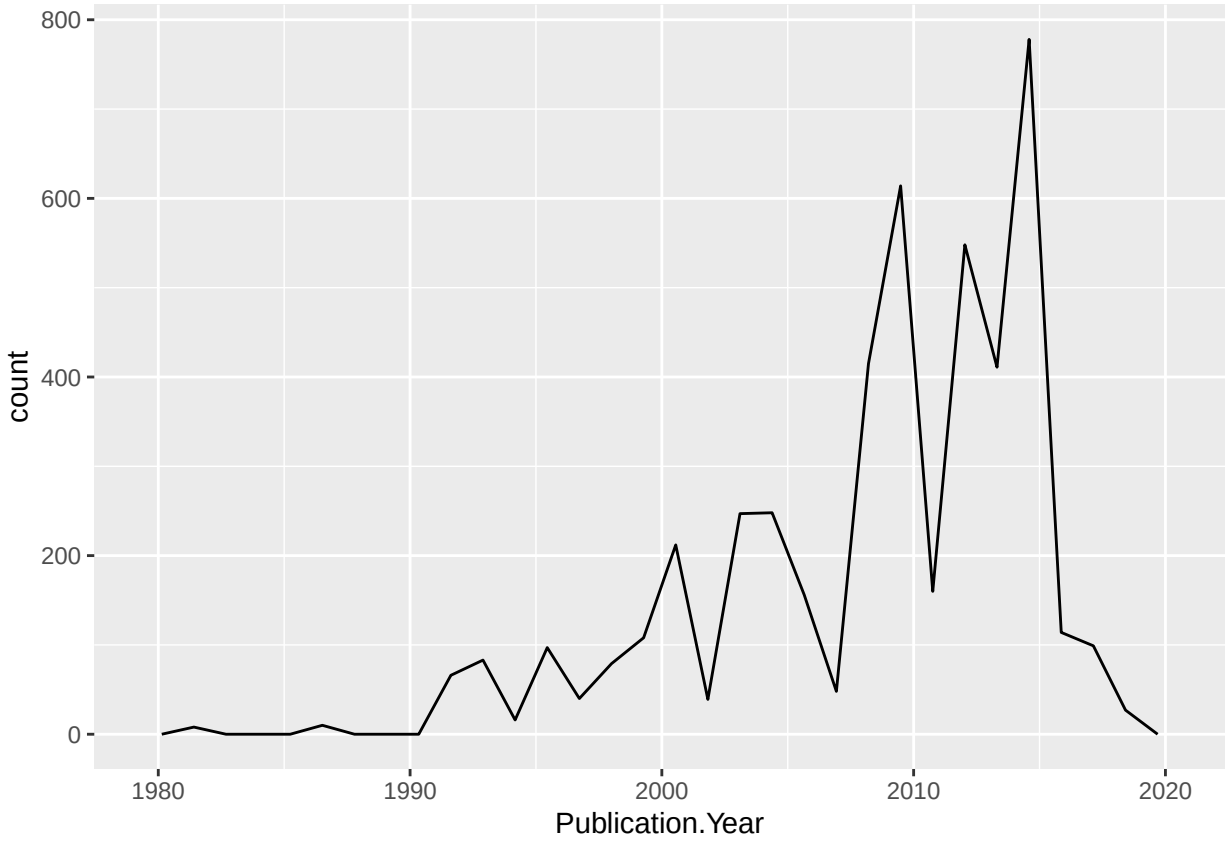
Answer:There are different types of data. There is numerical and categorical data. Numerical data is measured data. You are actually measuring the amount of something; actual values have meaning and can do math with them (you can count them). Numerical data also has continous values (measured/discrete values) On the other hand, categorical data can be a number, but it is nominal data. The value of the number doesn't mean anything, it is just assigned to that item. E.g. zipcodes

Explore your data graphically (Neonics)

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

```
#check our date column
ggplot(Neonics, aes(x = Publication.Year)) +
  geom_freqpoly()
```

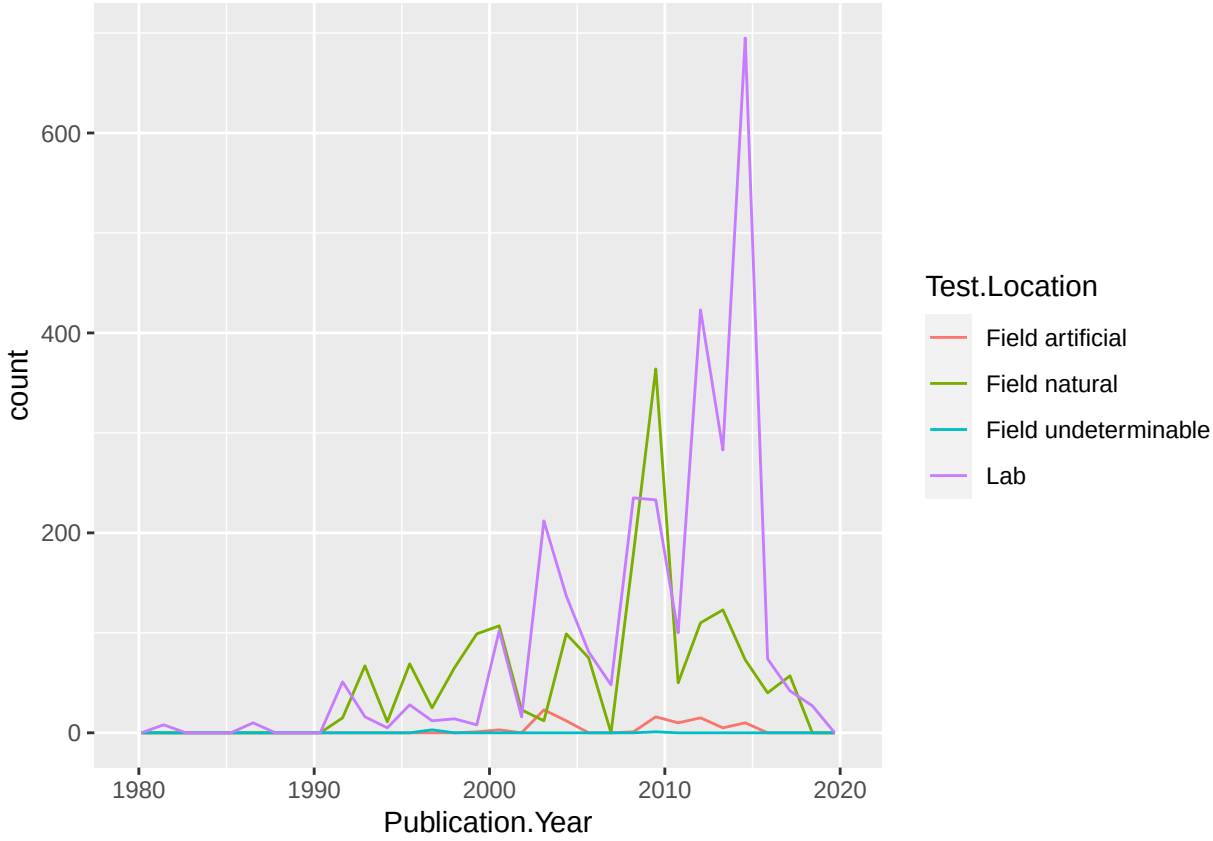
```
## `stat_bin()` using `bins = 30`. Pick better value with `binwidth`.
```



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`.
```

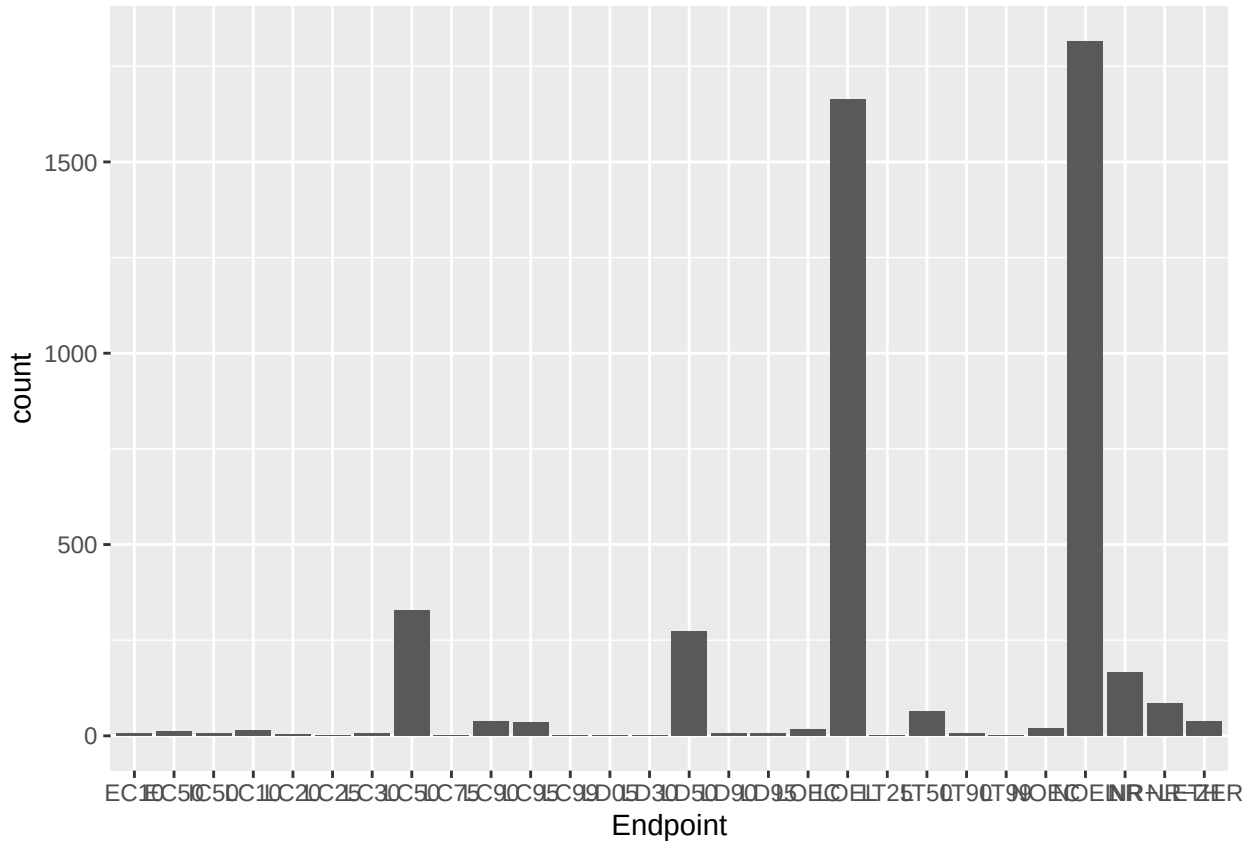


Interpret this graph. What are the most common test locations, and do they differ over time?

> Answer: The most common test locations are the lab, by a significant amount, followed by the field natural. There is a large increase in lab test locations between 2010 and 2015, and drastically drop right after 2015. The spike for field natural occurs before the lab spike in 2009, and then drastically drops in 2011. It looks like all test locations begin to drop after the 2015 publication year.

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()
```




```
summary(Neonics$Endpoint) #readability purposes
```

```
##     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
```

Answer: The two most common endpoints look to be LOEL and NOEL. LOEL database usage is terrestial that is defined as the lowest observable effect level. It is the lowest dose (concentration) producing effects that were significantly different (as reported by authors) from responses of controls. NOEL is also another endpoints that has a database usage for terrestial. It shows "no-observable-effect-level: highest dose (concentration) producing effects not significantly different from responses of controls according to author's reported statistical test," according to the ECOTOX_CodeAppendix.

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 = "%m/%d/%y")
```

```
unique(Litter$collectDate)
```

```
## [1] NA
```

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`?

```
df_uni <- unique(Litter$plotID)
length(df_uni)
```

```
## [1] 12
```

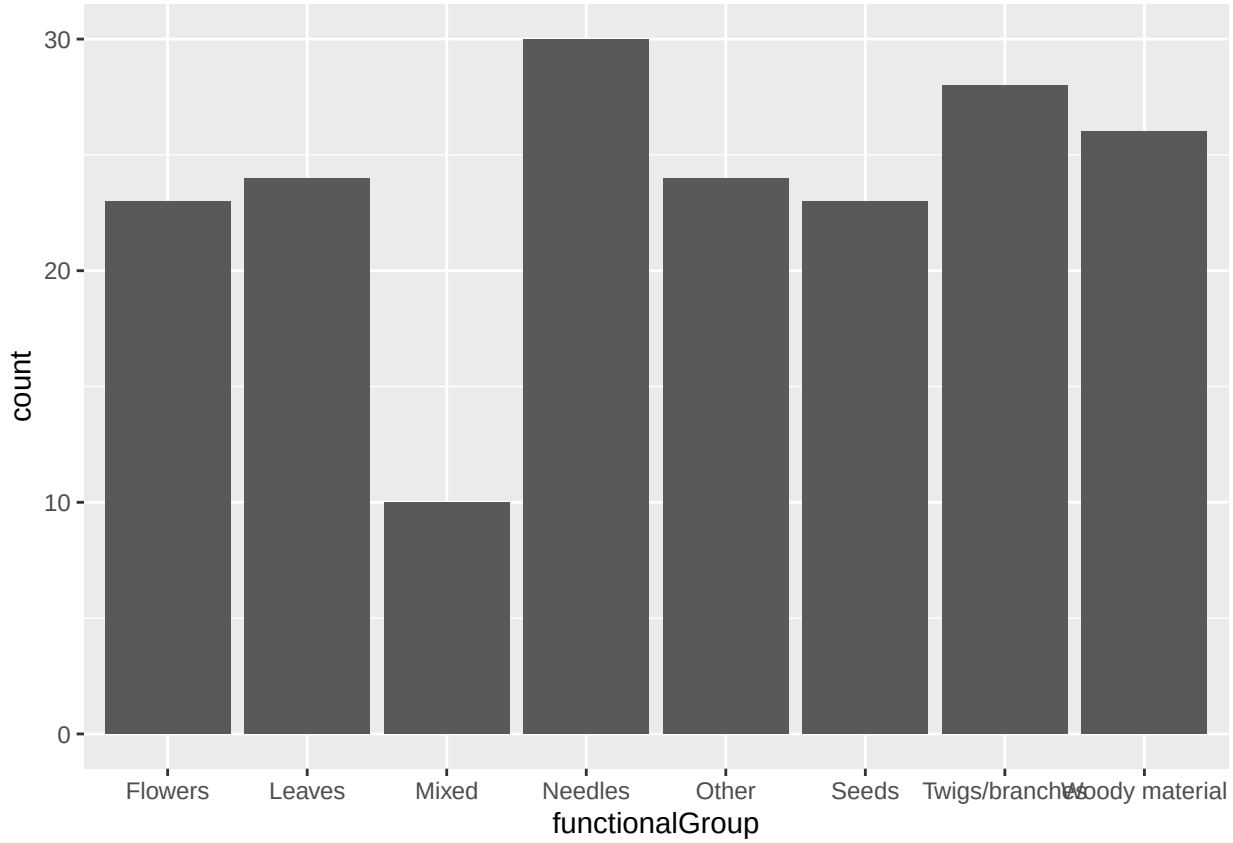
```
summary(df_uni)
```

```
## NIWO_040 NIWO_041 NIWO_046 NIWO_047 NIWO_051 NIWO_057 NIWO_058 NIWO_061
##        1        1        1        1        1        1        1        1
## NIWO_062 NIWO_063 NIWO_064 NIWO_067
##        1        1        1        1
```

Answer: The unique() function in R is used to eliminate or delete the duplicate values or the rows present in the vector, data frame, or matrix. In the 'summary' function, it will produce all summaries of the results.

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.

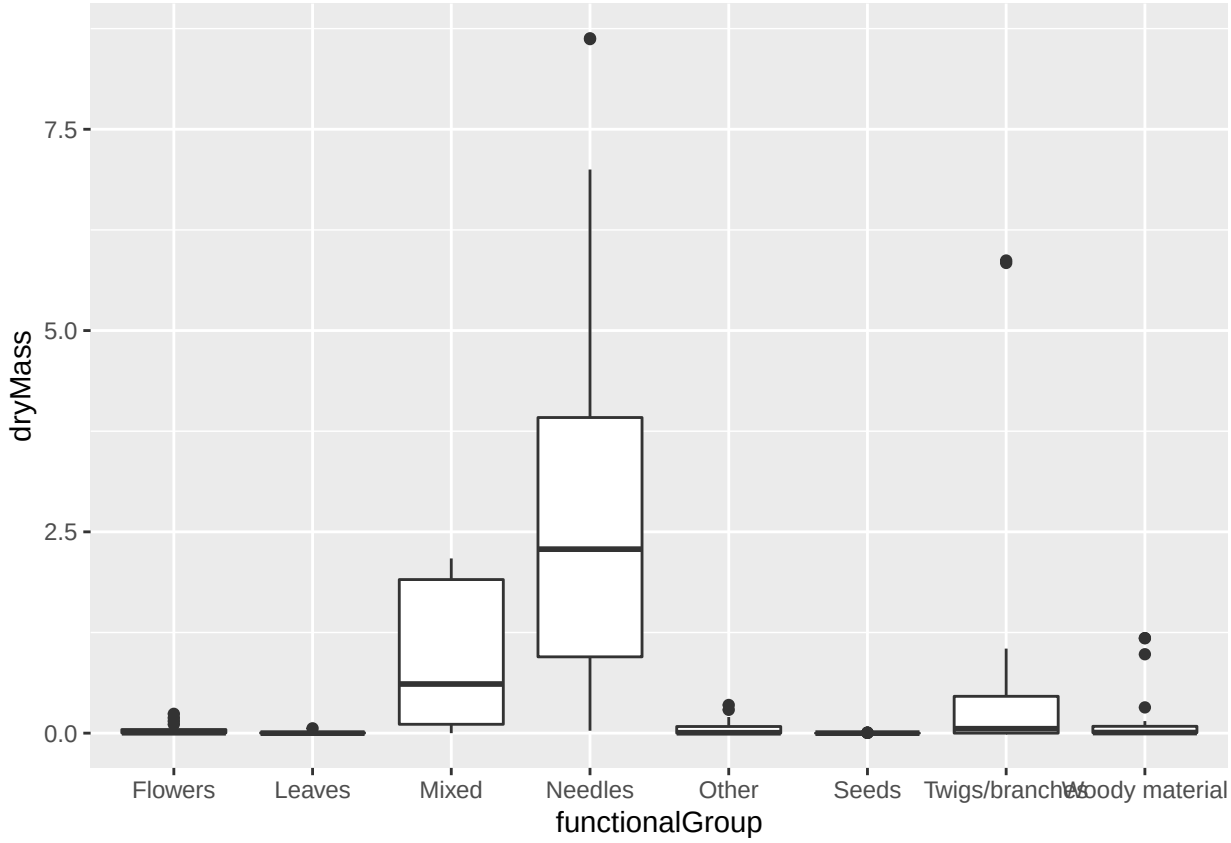
```
ggplot(Litter, aes(x= functionalGroup)) +
  geom_bar()
```



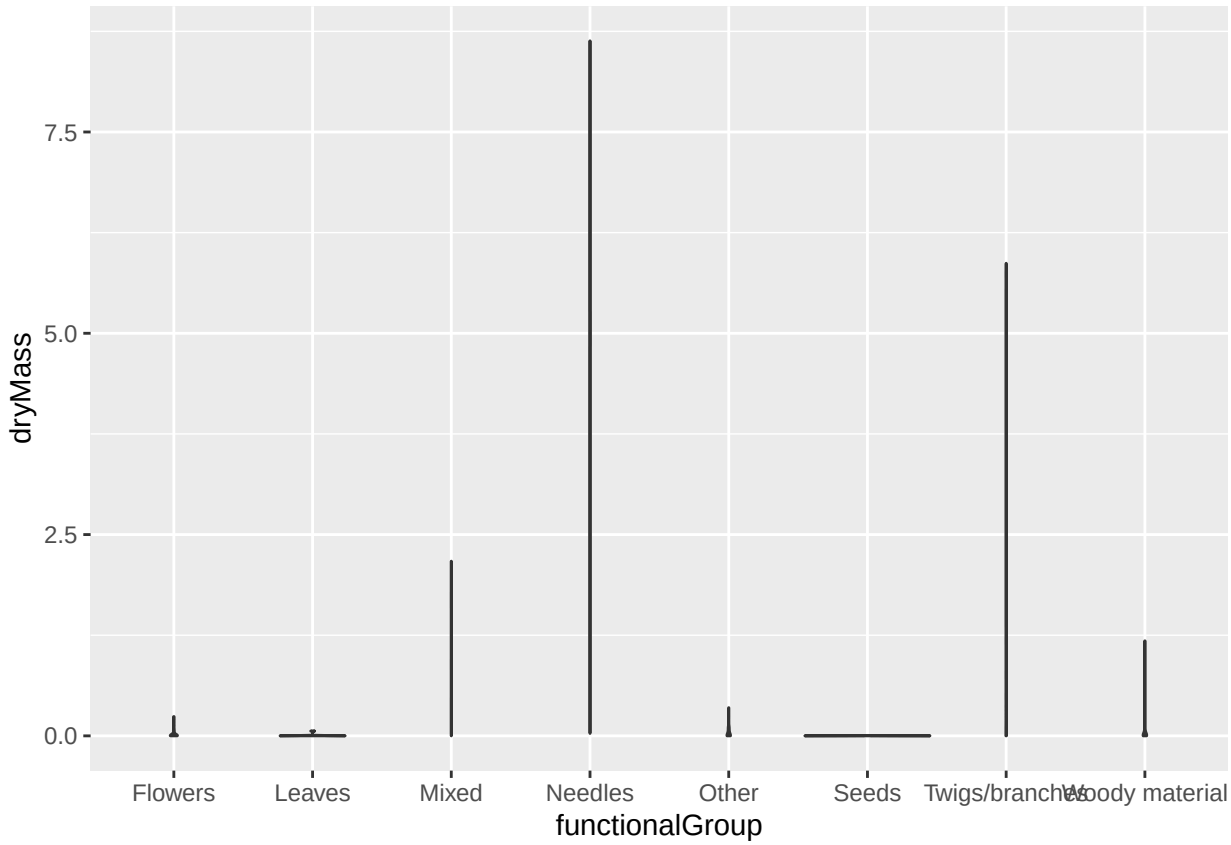
15. Using `geom_boxplot` and `geom_violin`, create a boxplot and a violin plot of dryMass by functional-

Group.

```
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 bloxplot is a good way to show a graphical image of the concentration of data. They can show how far the extreme values are from most of the data. A boxplot is constructed from the five values: min, first quartile, median, third quartile, and maximum value. These value are compared to see how close other data values are to them. Boxplot is mainly used for summary statistics only whereas the violin plot also takes into consideration the density of each variable.

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

> Answer: The type of litter (s) that tends to have the highest biomass at these sites are needles and twigs/branches.