Assignment 3: Data Exploration

Kendra Sultzer

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() #was in assignments folder and wanted it to be broader folder
```

```
## [1] "C:/Users/kendr/Documents/Spring2021Classes/ENV872/Environmental_Data_Analytics_2021"
```

```
setwd("C:/Users/kendr/Documents/Spring2021Classes/ENV872/Environmental_Data_Analytics_2021")
```

```
library(tidyverse)
```

```
## -- Attaching packages --------------------------------------- tidyverse 1.3.0 --

## v ggplot2 3.3.3     v purrr   0.3.4
## v tibble  3.0.6     v dplyr   1.0.3
## v tidyr   1.1.2     v stringr 1.4.0
## v readr   1.4.0     v forcats 0.5.1

## -- Conflicts ------------------------------------------ tidyverse_conflicts() --
## x dplyr::filter() masks stats::filter()
## x dplyr::lag()    masks stats::lag()
```

```
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")
```

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: I did a project last year on neonicotinoid insecticides for my environmental politics class! Neonics work by overwhelming and killing the nerve cells of insects which is advantageous for eliminating crop pests, but may have serious implications in other regards. Neonics have been linked to a decrease in pollinator populations which affect the production of fruits and vegetables for human consumption. In addition, a decrease in pollinators will also be contributing to overall loss of biodiversity.

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: Litter and woody debris may have a large impact on smaller species living within forests. Herpetofauna, like salamanders, depend on litter and woody debris for shelter and places to reproduce, so changes in these things may greatly affect smaller animal populations at the microhabitat level. In a different way, it could be interesting to see how different management that may be occurring around the research stations may be affecting litter and woody debris. Lastly, it could be interesting to compare litter produced between sites to ascertain productivity of trees in those forests.

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 occurs at terrestrial NEON sites that have woody vegetation >2m tall
   litter is collected from elevated traps while fine woody debris is collected from ground traps
   *masses reported are from a single trap during a single collection event

Obtain basic summaries of your data (Neonics)

5. What are the dimensions of the dataset? Neonics has 4,623 rows and 30 columns

```
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: The two most common effects studied are population and mortality which makes sense because researchers are probably most interested in seeing how the chemicals are affecting a group of insects and whether or not they died specifically from the chemicals. These would be looking at whether the chemicals cause worst case scenarios for the population 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.

```
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 6 most commonly studied insects are the honey bee, parasitic wasp, buff tailed bumblebee, Carniolan honey bee, bumble bee, and Italian honeybee. This makes sense since these are all insects that are pollinating and drinking potentially infected nectar. Pollinators are of particular salience because without pollinators, having fruits and vegetables will be difficult!

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.)
```

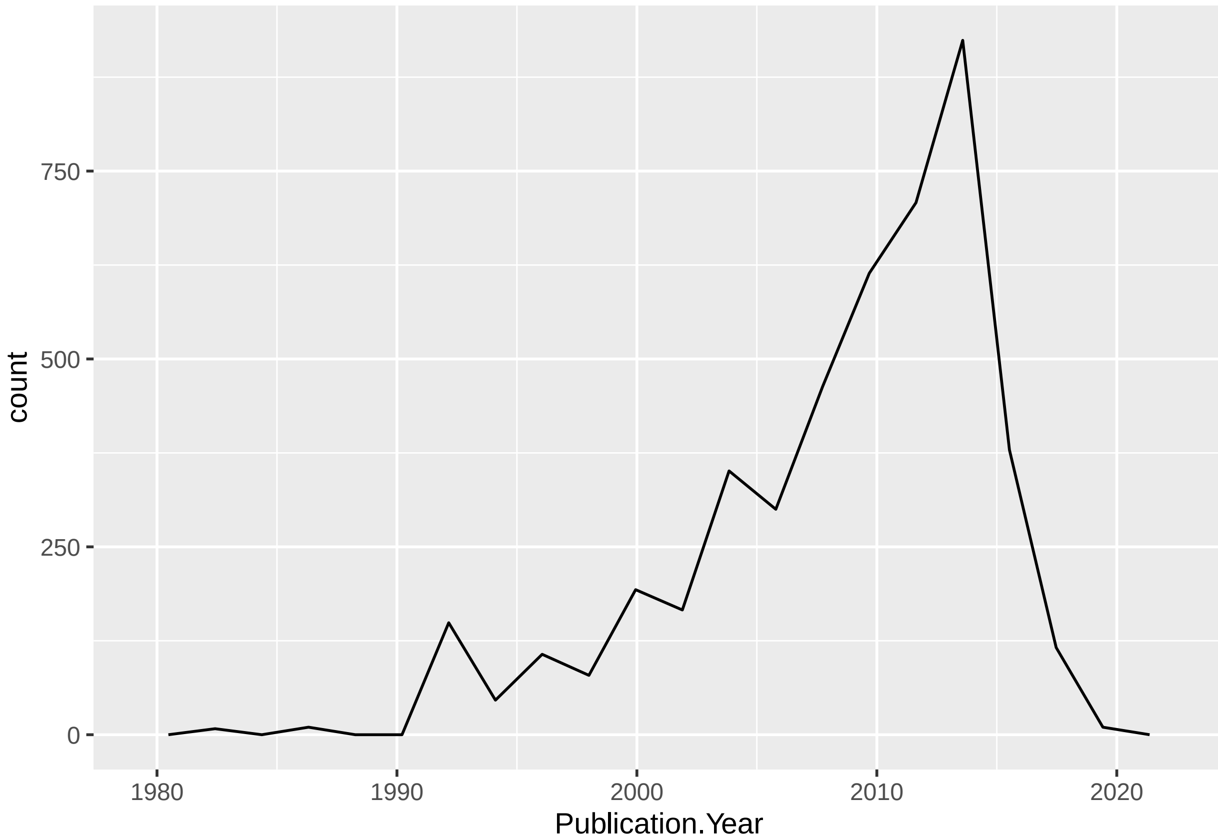
```
## [1] "factor"
```

Answer: The class is listed as a factor since I set strings to be factors when I read the data into R. It's not numeric because there are slashes and tildas along with the numbers in the column, so R thinks that it's a category as opposed to a number.

Explore your data graphically (Neonics)

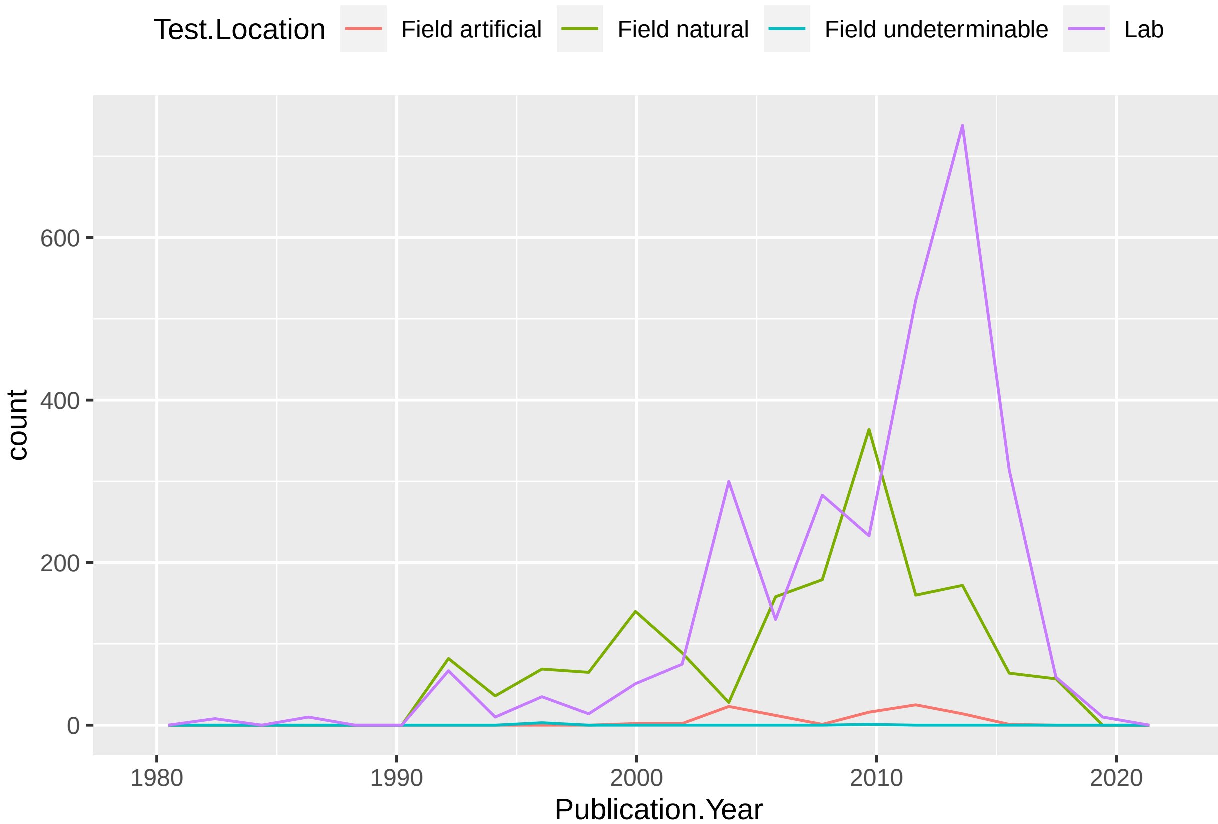
9. Using `geom_freqpoly`, generate a plot of the number of studies conducted by publication year.

```
ggplot(Neonics)+
geom_freqpoly(aes(x = Publication.Year),bins=20)
```



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=20)+
  theme(legend.position = "top")
```

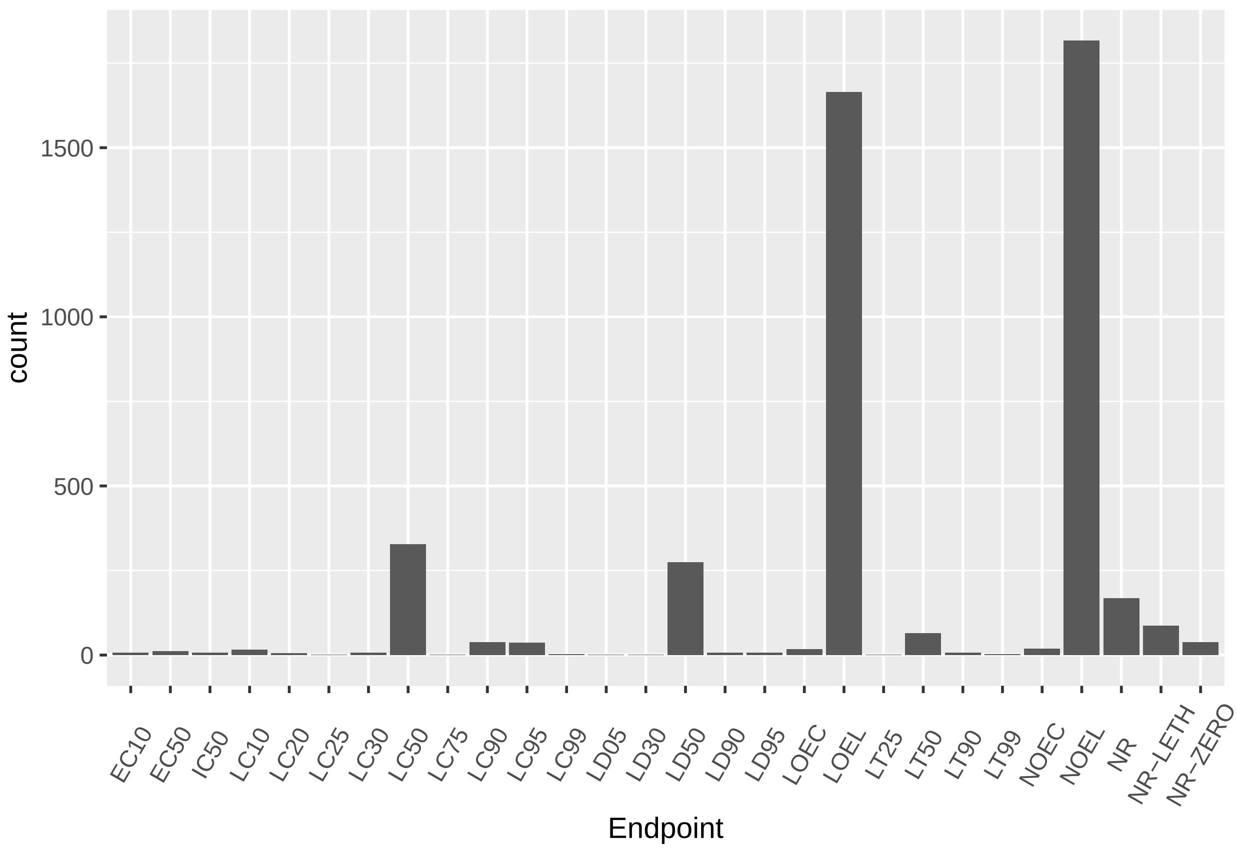


Interpret this graph. What are the most common test locations, and do they differ over time?

> Answer: The most common test locations are the lab (indoor setting) and then a natural field study. From 1990 to about 2002 or 2003, the natural field study was more common at which point laboratory study takes over and are mostly more common.

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), bins=20) +
  geom_bar() +
  theme(axis.text.x=element_text(angle=60, vjust=0.5))
```



Answer: The most common endpoint is NOEL, which means that there's no observable effect level. The highest concentration that produces effects is not significantly different from responses of controls. The second most common endpoint is LOEL, which means that there's the lowest observable effect level. The lowest concentration that produces effects were significantly different from 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) #listed as a character, not a date
```

```
## [1] "character"
```

```
Litter$collectDate <- as.Date(Litter$collectDate, format="%Y-%m-%d")
class(Litter$collectDate) #now listed as a date
```

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

```
unique(Litter$collectDate) #litter was sampled on August 2 and August 30, 2018
```

```
## [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`?

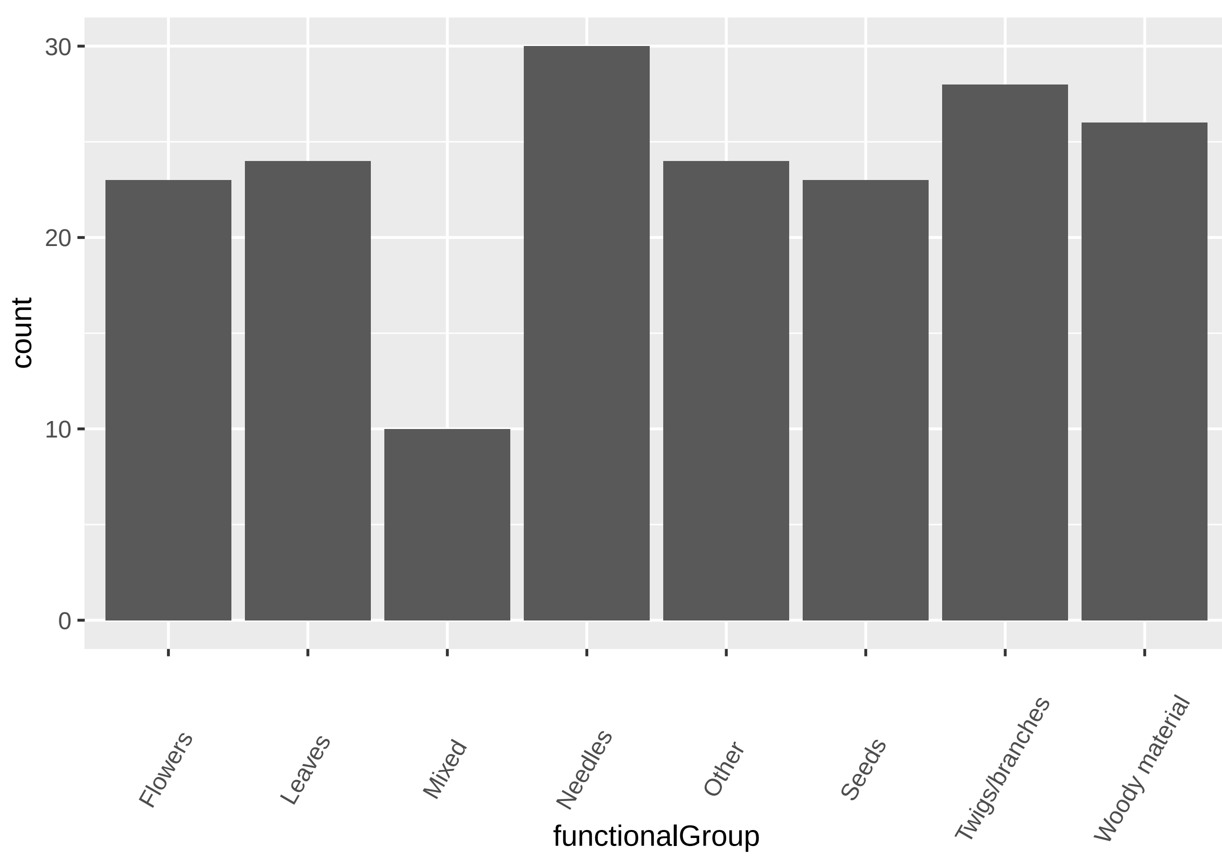
```
unique(Litter$plotID)
```

```
##  [1] "NIWO_061" "NIWO_064" "NIWO_067" "NIWO_040" "NIWO_041" "NIWO_063"
##  [7] "NIWO_047" "NIWO_051" "NIWO_058" "NIWO_046" "NIWO_062" "NIWO_057"
```

Answer: 12 plots were sampled at Niwot Ridge. The "unique" function just outputs a list of the categories within that variable whereas the "summary" function has a list of the categories within the variables along with a count of how many counts are within each category.

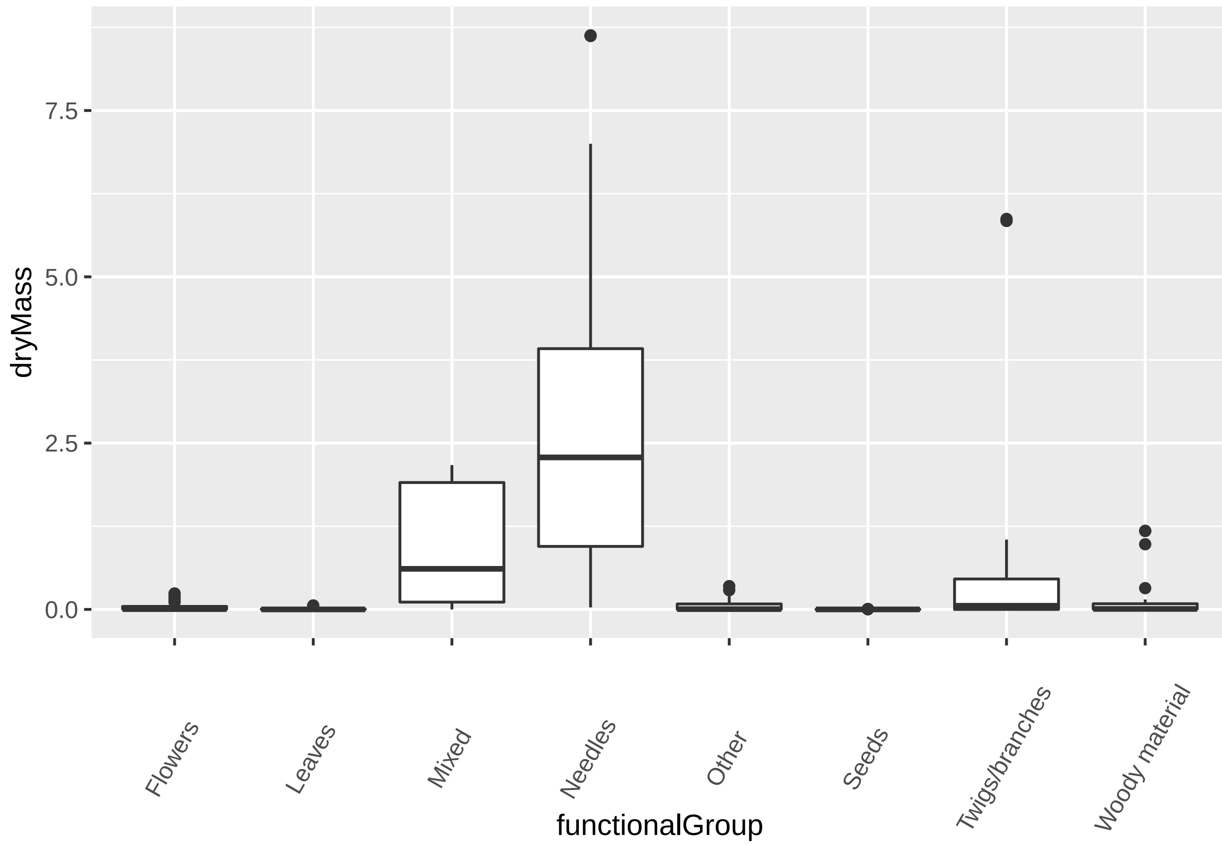
14. Create a bar graph of functionalGroup counts. This shows you what type of litter is collected at the Niwot Ridge sites. Notice that litter types are fairly equally distributed across the Niwot Ridge sites.

```
ggplot(Litter, aes(x=functionalGroup))+
  geom_bar()+
  theme(axis.text.x=element_text(angle=60, vjust=0.3))
```



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))+
  theme(axis.text.x=element_text(angle=60, vjust=0.3))
```

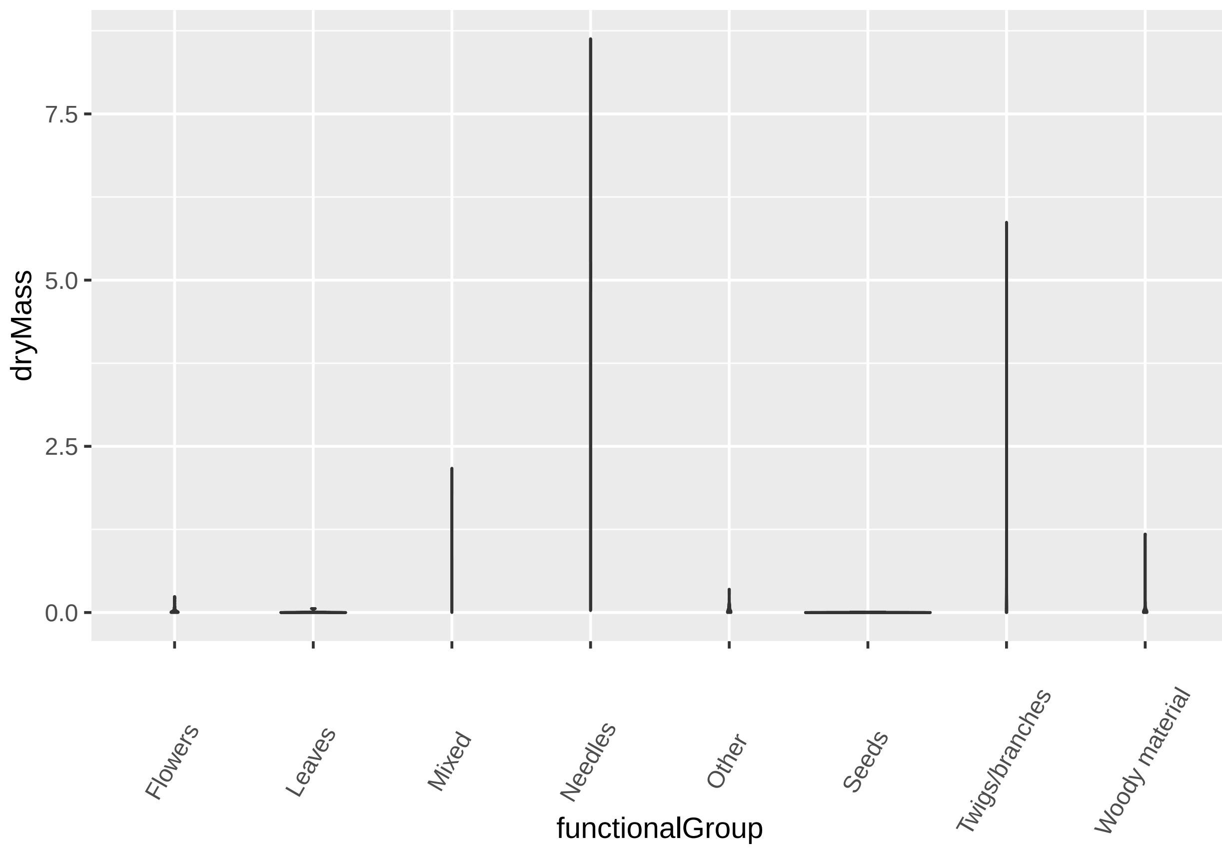



```
ggplot(Litter) +
  geom_violin(aes(x = functionalGroup, y=dryMass),
              draw_quantiles = c(0.25, 0.5, 0.75))+
  theme(axis.text.x=element_text(angle=60, vjust=0.3))
```

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

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

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



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

> Answer: In this case, most of the dry masses are very condensed within their functional group. This large amount of condensing offers some interpretation in the violin plot. However, the few functional groups with large amounts of variability in dry mass are very difficult to visualize in the violin plot and just get collapsed into straight lines.

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

> Answer: In general, the needle functional group has the highest biomass at these sites, with the mixed function group following up.