Assignment 3: Data Exploration

Yikai Jing, Section #4

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/me/Environmental_Data_Analytics_2022/Assignments"
```

```
library(tidyverse)

Neonics <- read.csv("~/Environmental_Data_Analytics_2022/Data/Raw/ECOTOX_Neonicotinoids_Insects_raw.csv
Litter <- read.csv("~/Environmental_Data_Analytics_2022/Data/Raw/NEON_NIWO_Litter_massdata_2018-08_raw.
```

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 are now the most widely used insecticides in the world. They are also less toxic to vertebrates than common older insecticides due to their increased selectivity to insect acetylcholine receptors in the brain. These benefits have led to their widespread use in agriculture and residential areas; however, they have been under scrutiny due to their persistence in the soil, ability to leach into the environment, high water solubility, and potential negative health implications for non-target organisms such as pollinators. The contradictory findings for the

effects of neonicotinoids on insects has caused them to be a very controversial topic for policy decisions.

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: Trees that fall and decay in the forest add nutrients to the forest soil and retain moisture in the forest. The forest floor, roots, and fine woody litter become increasingly important contributors to soil organic matter as the intensity of forest management increases and the contribution of large woody litter decreases.

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: *One litter trap pair (one elevated trap and one ground trap) is deployed for every 400 m2 plot area, resultng in 1-4 trap pairs per plot. Trap placement within plots may be either targeted or randomized, depending on the vegetaton.* Litter and fine woody debris sampling is executed at terrestrial NEON sites that contain woody vegetation >2m tall. Locations of tower plots are selected randomly within the 90% flux footprint of the primary andsecondary airsheds (and additional areas in close proximity to the airshed, as necessary to accommodate sufficientspacing between plots). *Ground traps are sampled once per year. Target sampling frequency for elevated traps varies by vegetation present at the site, with frequent sampling (1x every 2weeks) in deciduous forest sites during senescence, and infrequent year-round sampling (1x every 1-2 months) at evergreen sites.

Obtain basic summaries of your data (Neonics)

5. What are the dimensions of the dataset?

```
dim(Neonics)
```

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

```
str(Neonics)
```

```
## 'data.frame':    4623 obs. of  30 variables:
##  $ CAS.Number                     : int  58842209 58842209 58842209 58842209 58842209 58842209 5884
##  $ Chemical.Name                  : chr  "Tetrahydro-2-(nitromethylene)-2H-1,3-thiazine" "Tetrahydr
##  $ Chemical.Grade                 : chr  "Technical grade, technical product, technical formulation
##  $ Chemical.Analysis.Method       : chr  "Unmeasured" "Unmeasured" "Unmeasured" "Unmeasured" ...
##  $ Chemical.Purity                : chr  "99" "99" "95" "95" ...
##  $ Species.Scientific.Name        : chr  "Araecerus fasciculatus" "Araecerus fasciculatus" "Musca d
##  $ Species.Common.Name            : chr  "Coffee Bean Weevil" "Coffee Bean Weevil" "House Fly" "Hou
##  $ Species.Group                  : chr  "Insects/Spiders" "Insects/Spiders" "Insects/Spiders" "Ins
##  $ Organism.Lifestage             : chr  "Adult" "Adult" "Young" "Young" ...
##  $ Organism.Age                   : chr  "NR" "NR" "NR" "NR" ...
##  $ Organism.Age.Units             : chr  "Not reported" "Not reported" "Hour(s)" "Hour(s)" ...
##  $ Exposure.Type                  : chr  "Topical, general" "Topical, general" "Food" "Food" ...
##  $ Media.Type                     : chr  "No substrate" "No substrate" "Filter paper" "Filter paper
##  $ Test.Location                  : chr  "Lab" "Lab" "Lab" "Lab" ...
##  $ Number.of.Doses                : chr  "NR" "NR" "11" "11" ...
##  $ Conc.1.Type..Author.           : chr  "Active ingredient" "Active ingredient" "Active ingredient
##  $ Conc.1..Author.                : chr  "27.2" "19.7" "47" "25" ...
##  $ Conc.1.Units..Author.          : chr  "ug/g bdwt" "ug/g bdwt" "mg/L" "mg/L" ...
```

```
##  $ Effect                          : chr  "Mortality" "Mortality" "Mortality" "Mortality" ...
##  $ Effect.Measurement              : chr  "Mortality" "Mortality" "Mortality" "Mortality" ...
##  $ Endpoint                        : chr  "LD50" "LD50" "LC50" "LC50" ...
##  $ Response.Site                   : chr  "Not reported" "Not reported" "Not reported" "Not reported
##  $ Observed.Duration..Days.        : chr  "1" "1" "1" "1" ...
##  $ Observed.Duration.Units..Days.  : chr  "Day(s)" "Day(s)" "Day(s)" "Day(s)" ...
##  $ Author                          : chr  "Childers,C.C., and H.N. Nigg" "Childers,C.C., and H.N. Nig
##  $ Reference.Number                : int  107388 107388 103312 103312 103312 103312 103312 103312 103
##  $ Title                           : chr  "Contact Toxicity of Insecticides to Adults of the Coffee
##  $ Source                          : chr  "J. Econ. Entomol.75(3): 556-559" "J. Econ. Entomol.75(3):
##  $ Publication.Year                : int  1982 1982 1986 1986 1986 1986 1986 1986 1986 1986 ...
##  $ Summary.of.Additional.Parameters: chr  "Purity: \xca NR - NR | Organism Age: \xca NR - NR Not rep
```

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(as.factor(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
```

```
sort(table(Neonics$Effect),decreasing=TRUE)[1:2]
```

```
##
## Population  Mortality
##       1803       1493
```

Answer: Population and mortality are the most studied effect. This may tell us how will the tested amount of Neonicotinoids affect the insects' mortality so we can make assumptions regarding the safeness and useage of the chemicals.

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(as.factor(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
```

```
sort(table(Neonics$Species.Common.Name),decreasing=TRUE)[1:6]
```

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

Answer: Honey Bee, Parasitic Wasp, Buff Tailed Bumblebee, Carniolan Honey Bee, Bumble Bee, Italian Honeybee These six are all bees or they are species very close to the bee. And I think they are all considered to be beneficial insects. they might be of interest over other insects because bees are the major pollinator of the plants so they may have more contact with plants and vegetation. And this fact that beneficial insects are killed by insecticides illustrates the contradiction or disadvantage of such insecticides.

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] "character"
```

Answer:It's read as "factor" in this file. It is not numeric because in this dataset concentration

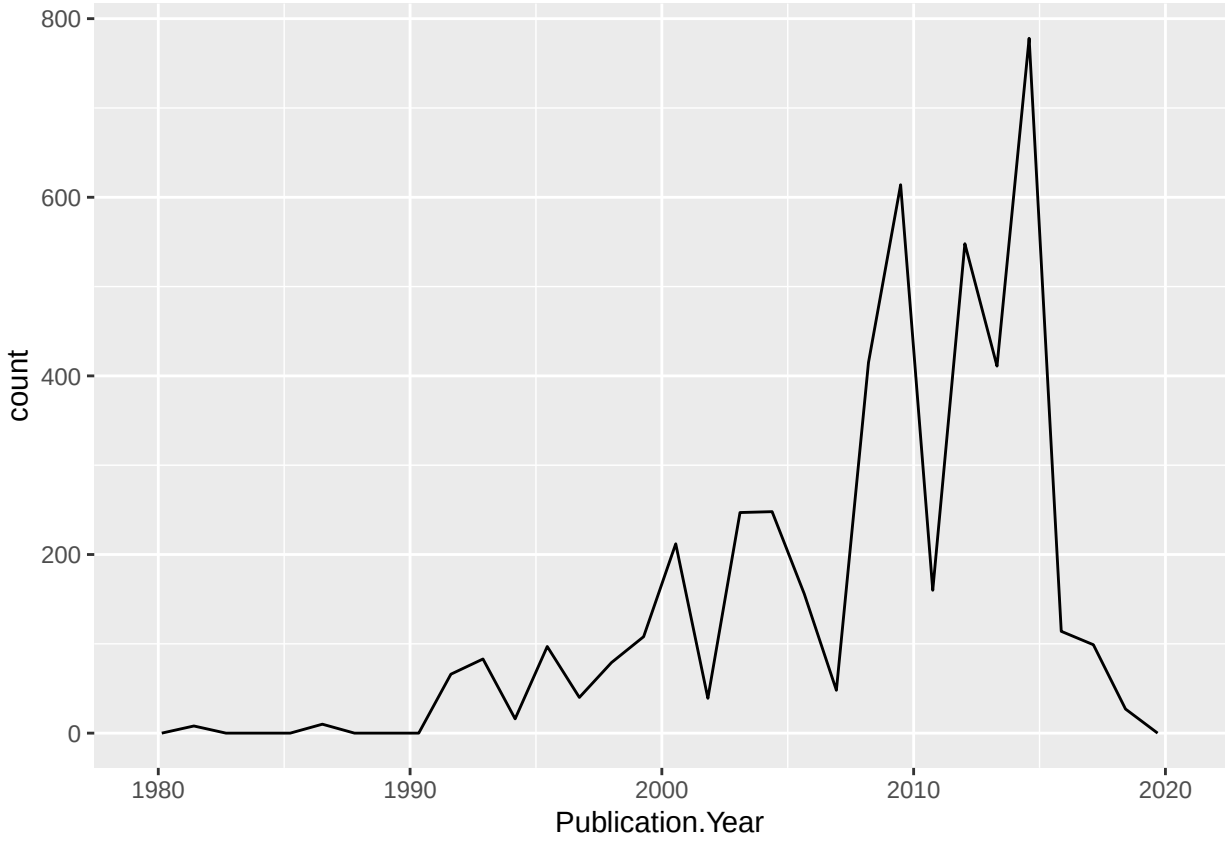
has different units and different expressions. Two rows adjacent to each other may not necessarily suitable for mathemetically computation.

Explore your data graphically (Neonics)

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

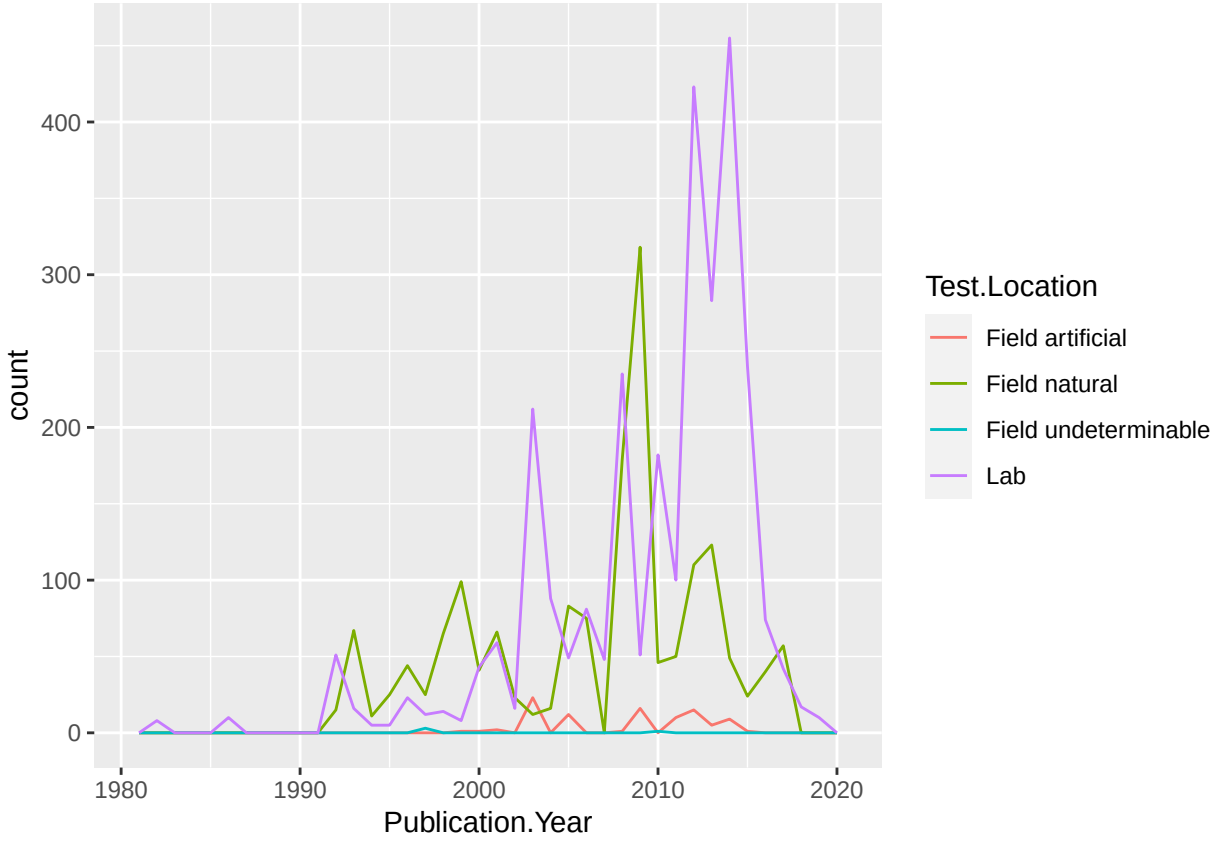
```
ggplot(Neonics)+
  geom_freqpoly(aes(x = Publication.Year))
```

```
## `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)+
  geom_freqpoly(aes(x = Publication.Year, color = Test.Location), binwidth = 1)
```

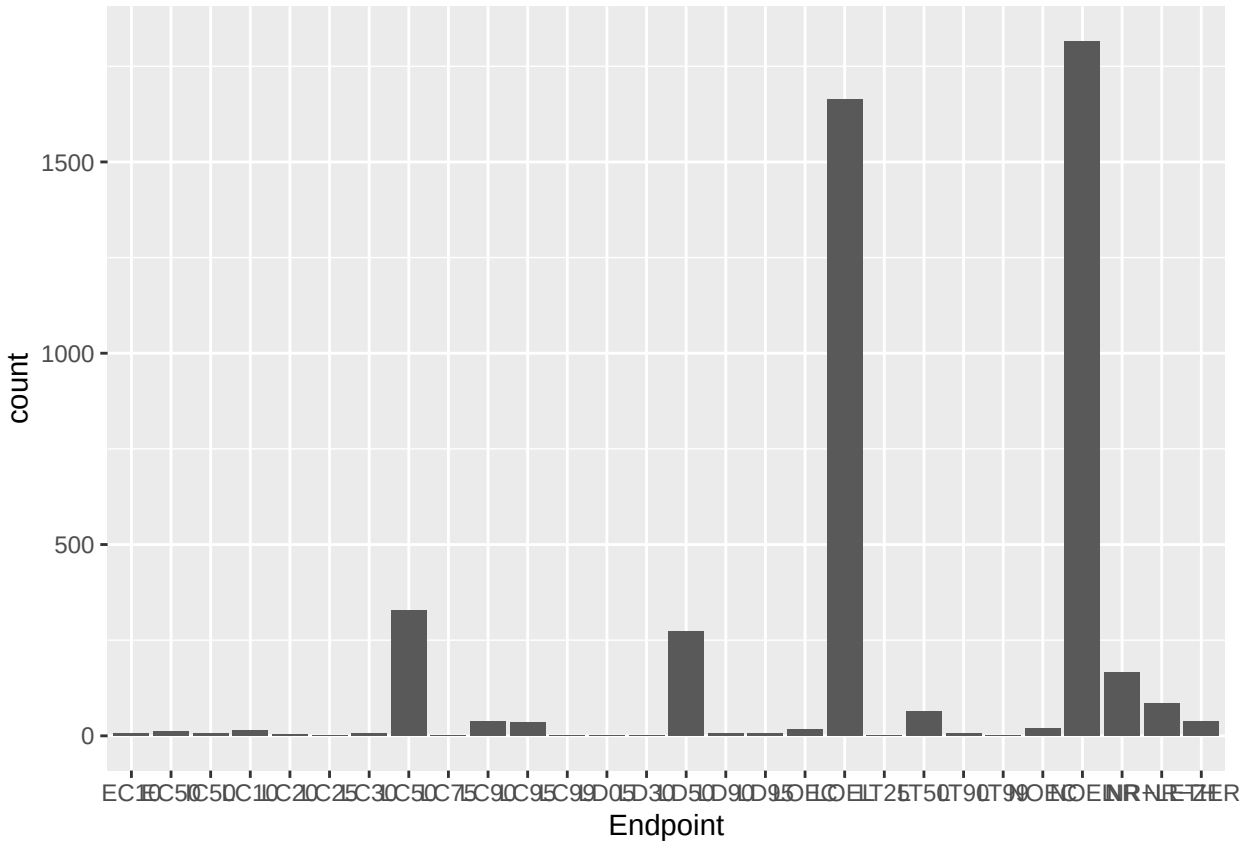


Interpret this graph. What are the most common test locations, and do they differ over time?

Answer: According to this graph, lab would be the most common test location, and it was between 2003 to 2008 and after 2010. Field natural is also common during the period of 1993 to 2002, 2008 to 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)+
  geom_bar(aes(x = Endpoint))
```



Answer: LOEL and NOEL are the two most common end points. NOEL means No-observableeffect-level for terrestrial database, while LOEL means Lowest-observable-effect-level for terrestrial database.

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] "character"
```

```
str(Litter$collectDate)
```

```
##  chr [1:188] "2018-08-02" "2018-08-02" "2018-08-02" "2018-08-02" ...
```

```
Litter$collectDate <- as.Date(Litter$collectDate, format = "%Y-%m-%d")
str(Litter$collectDate)
```

```
##  Date[1:188], format: "2018-08-02" "2018-08-02" "2018-08-02" "2018-08-02" "2018-08-02" ...
```

```
class(Litter$collectDate)
```

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

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

```
## [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$siteID)
```

```
## [1] "NIWO"
```

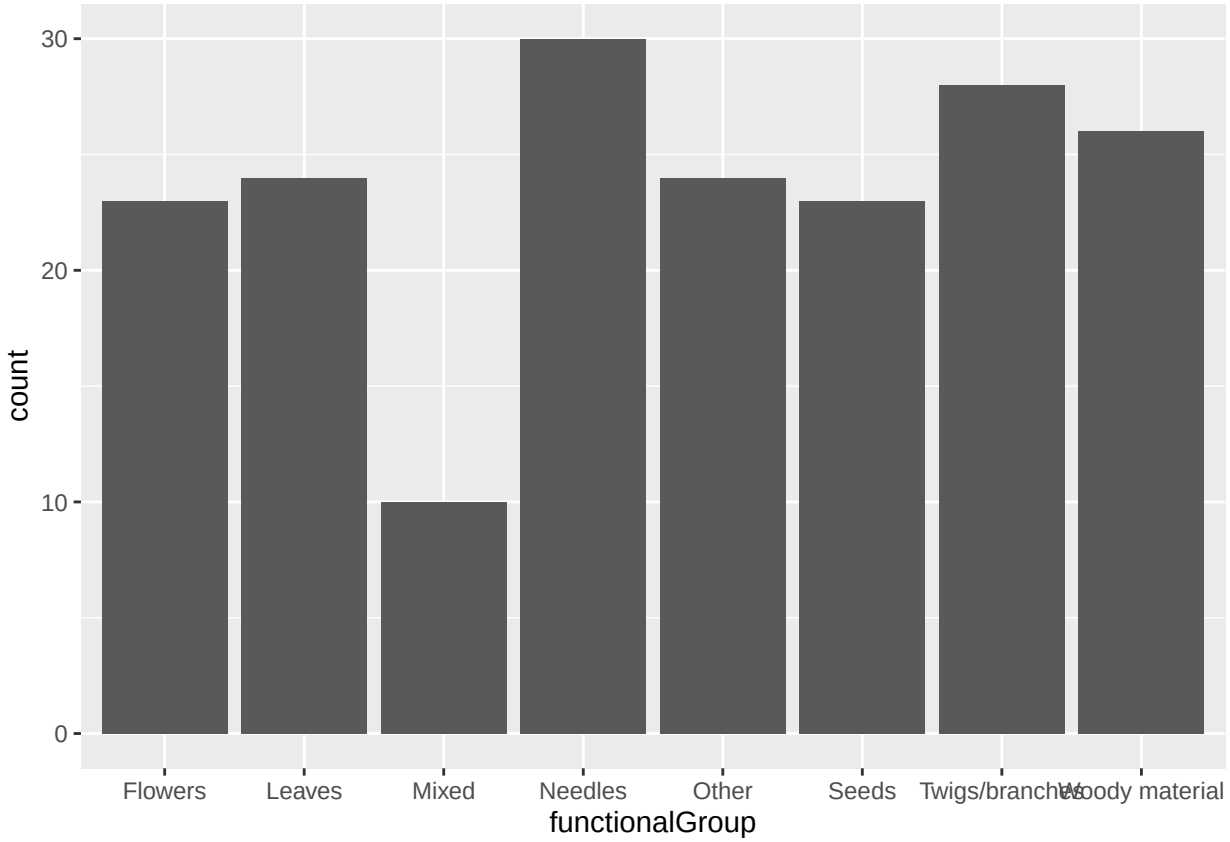
```
summary(Litter$siteID)
```

```
##    Length     Class      Mode
##       188 character character
```

Answer: There are 188 plots, all of which were sampled at Niwot Ridge. 'unique' function only provides a list of vectors that the df included without duplication, while 'summary' function tells the vector's length, class, and mode.

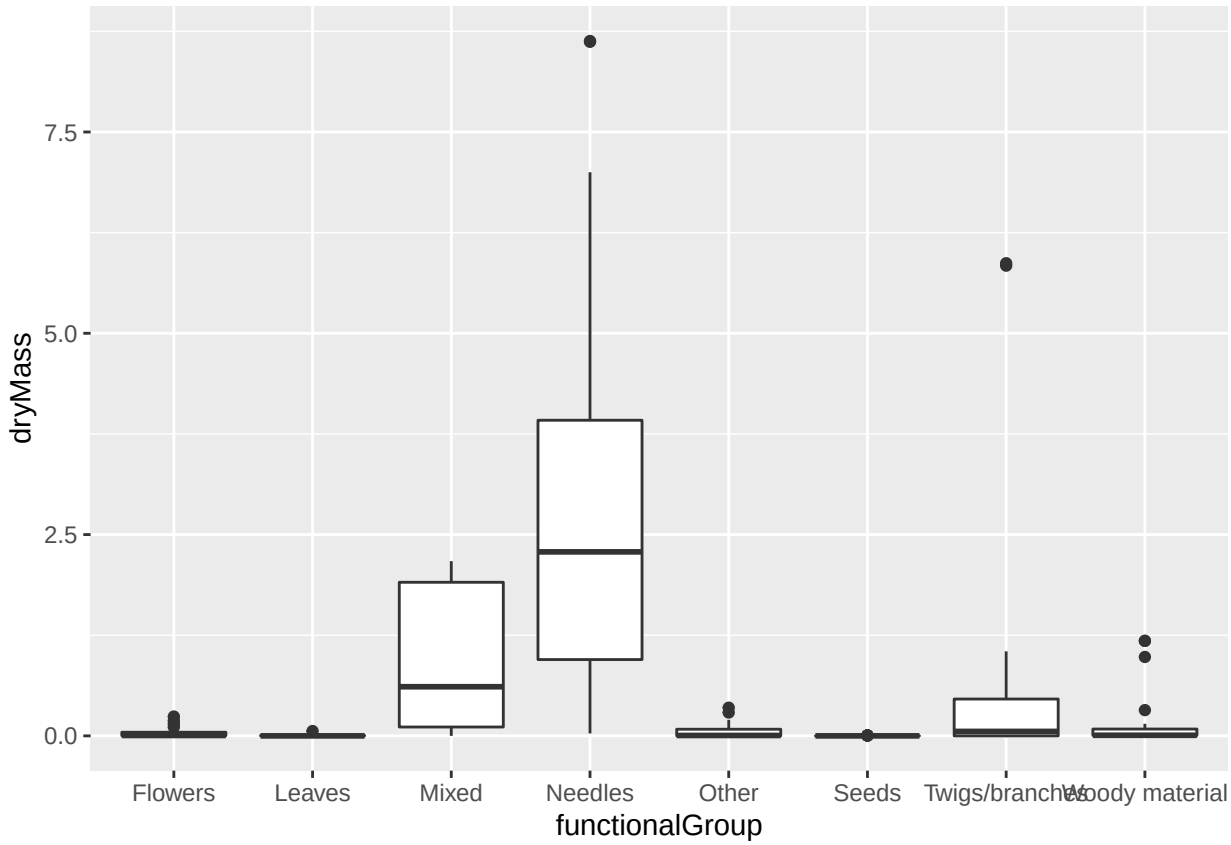
14. Create a bar graph of functionalGroup counts. This shows you what type of litter is collected at the Niwot Ridge sites. Notice that litter types are fairly equally distributed across the Niwot Ridge sites.

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



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))
```

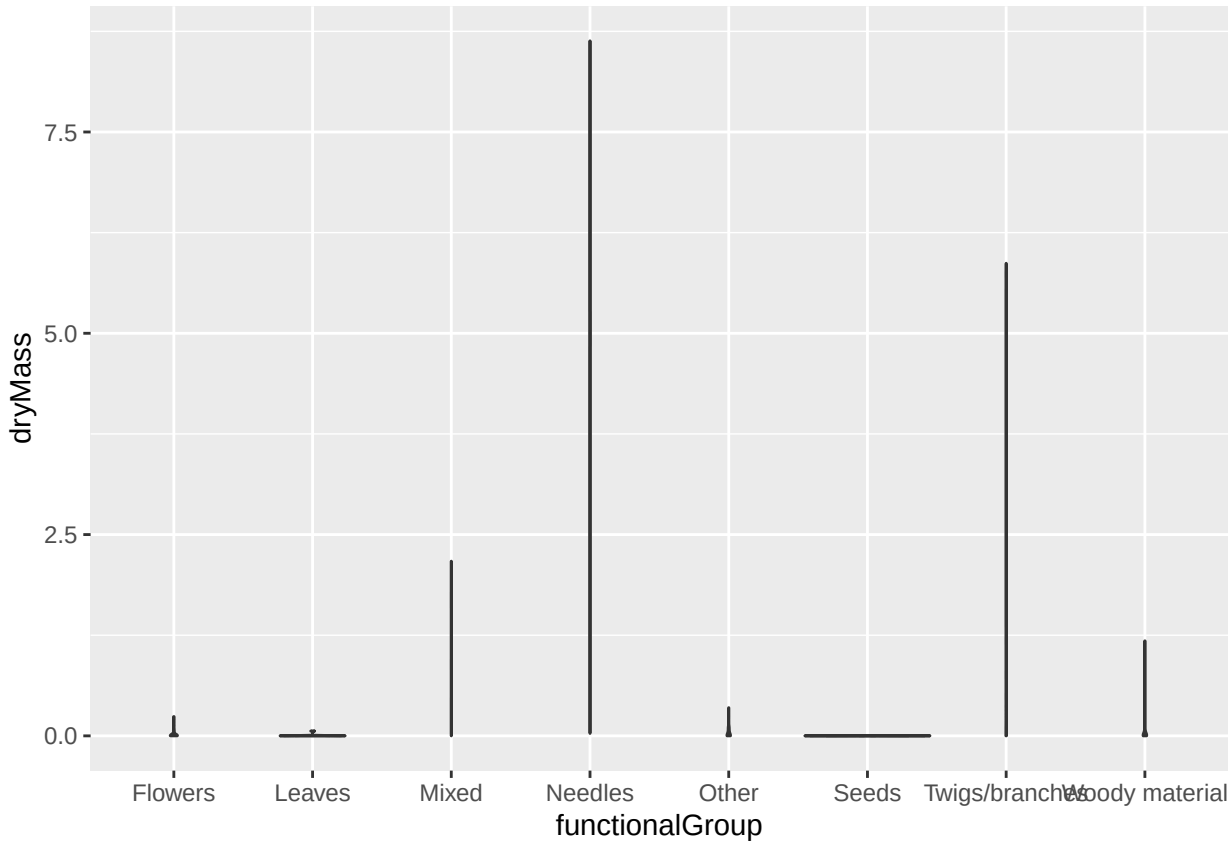



```
ggplot(Litter)+
  geom_violin(aes(x = functionalGroup, y = dryMass),
              draw_quantiles = c(0.25, 0.5, 0.75))
```

```
## 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: boxplot has a steady shape of "box" that performs better regardless of the data. Violin plot is dependent on the distribution of data to draw its shape so that a very small unit of data, such as dryMass in our case, may lead to ineffective presentation.

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

> Answer: Needles