17/17 - excellent!

# Assignment 3: Data Exploration

Rachel Gonsenhauser

## 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. 2/2

```
getwd()
```

```
## [1] "/Users/rachelgonsenhauser/Documents/Environmental_Data_Analytics_2020/Assignments/Assignment 3"
```

```
library(tidyverse)
Neonics <- read.csv("~/Documents/Environmental_Data_Analytics_2020/Data/Raw/ECOTOX_Neonicotinoids_Insec
# View(Neonics)
Litter <- read.csv("~/Documents/Environmental_Data_Analytics_2020/Data/Raw/NEON_NIWO_Litter_massdata_20
# View(Litter)
```

## 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: The use of neonicotinoids has been associated with adverse ecological impacts, especially honey bee population collapse and bird population declines due to the reduciton in insect populations. As such, understanding the ecotoxicology of neonicotinoids on insects is crucial in understanding how these biotic populations respond to certain dosages of the insecticide and, consequently, how best to manage these populations. 1/1

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 can provide many benefits, including nutrient cycling, providing habitat for forest species, and promoting soil stabilization and improved infiltration. However, large quantitites of litter and woody debris on the ground of forests can provide fuel for wildfire events, leading to prolonged, and more severe forest fires. 1/1

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 debris is sampled at terrestrial NEON sites that house woody vegetation >2m tall, with sampling occuring in tower plots whose locations were selected randomly within the 90% flux footprint of primary and secondary airsheds.* Trap placement within plots were targeted or randomized depending on vegetation (e.g. in sites with more than 50% aerial cover of woody vegetation >2m tall, placement of traps was raondomized). *Ground traps are sampled once per year whereas sampling frequency for elevated traps varies by site vegetation, with deciduous forest sites being sampled more frequently than evergreen sites. 1/1

## Obtain basic summaries of your data (Neonics)

5. What are the dimensions of the dataset? 0.5/0.5

```r
dim(Neonics)
```

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

   Answer: The dimensions of the dataset are 4,623 rows by 30 columns.

6. Using the `summary` function, determine the most common effects that are studied. Why might these effects specifically be of interest? 1/1

```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 common effects studied include population, mortality, behavior, feeding behavior, reproduction, and development. These effects were measured by abudance, mortality, survival, progency count, food consumption, and emergence, respectively. These effects might be of interest because testing these effects can reveal information about insect populations' responses to different doses of and means of exposure to neonicotinoids. Studying these effects in this context also allows cross-species comparisons of how the effects of different exposure manifest in different 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? 1/1

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 six most commonly studied species in the dataset are the honey bee, parasitic wasp, buff tailed bumblebee, carniolan honey bee, bumblee bee, and italian honeybee. These species are all bees and wasps and may be of particular interest over other insect species because neonicotinoids can kill bees and parasitic wasps as they can contaminate pollen and nectar that these insects feed on. As such, it makes sense that these species would be of particular interest in neonicotinoud ecotoxicity studies over other insect species such as beetles and spiders who do not feed on pollen and nectar.

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.)
```

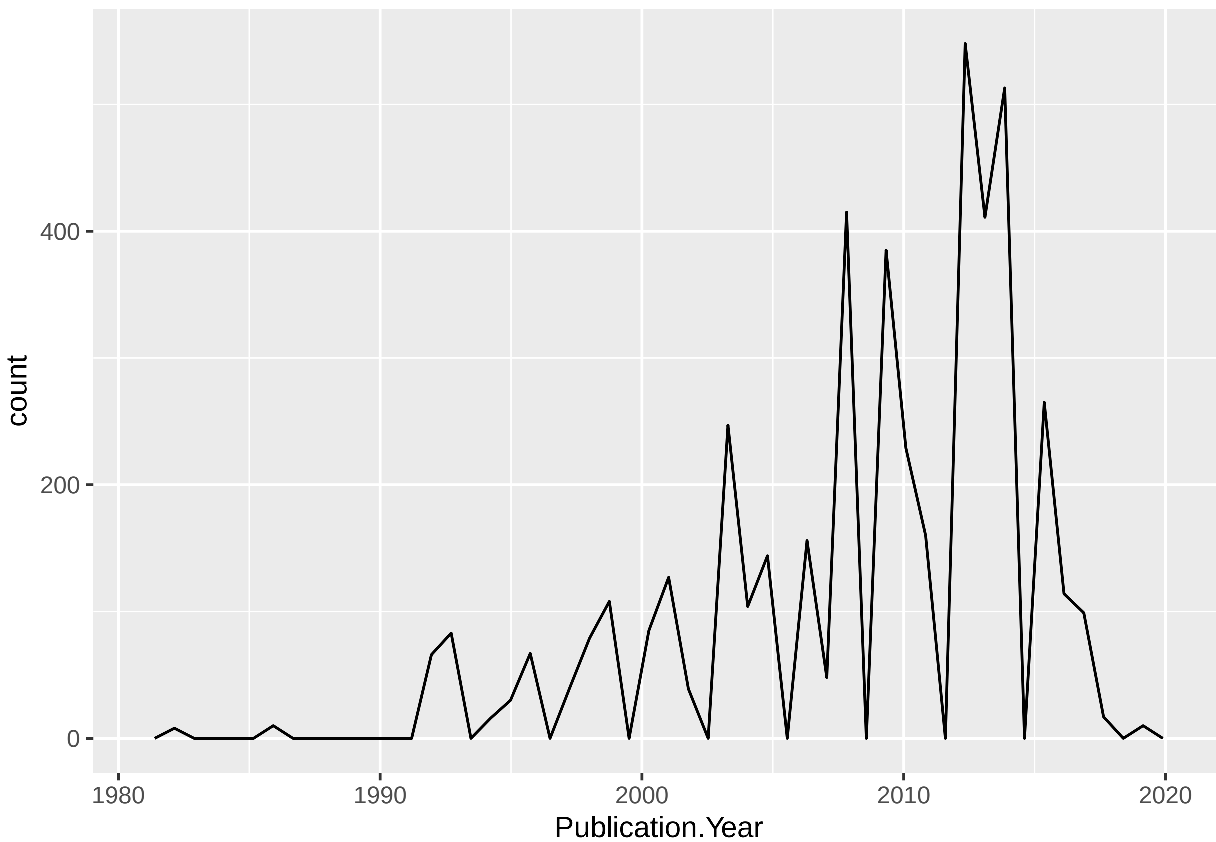
```
## [1] "factor"
```

Answer: The class of Conc.1..Author in the dataset is "factor". The class of these values is not numeric because the data in this column contains non-numeric information (e.g. slashes and letters). As such, this column cannot be made a numeric class of data. 1/1

## Explore your data graphically (Neonics)

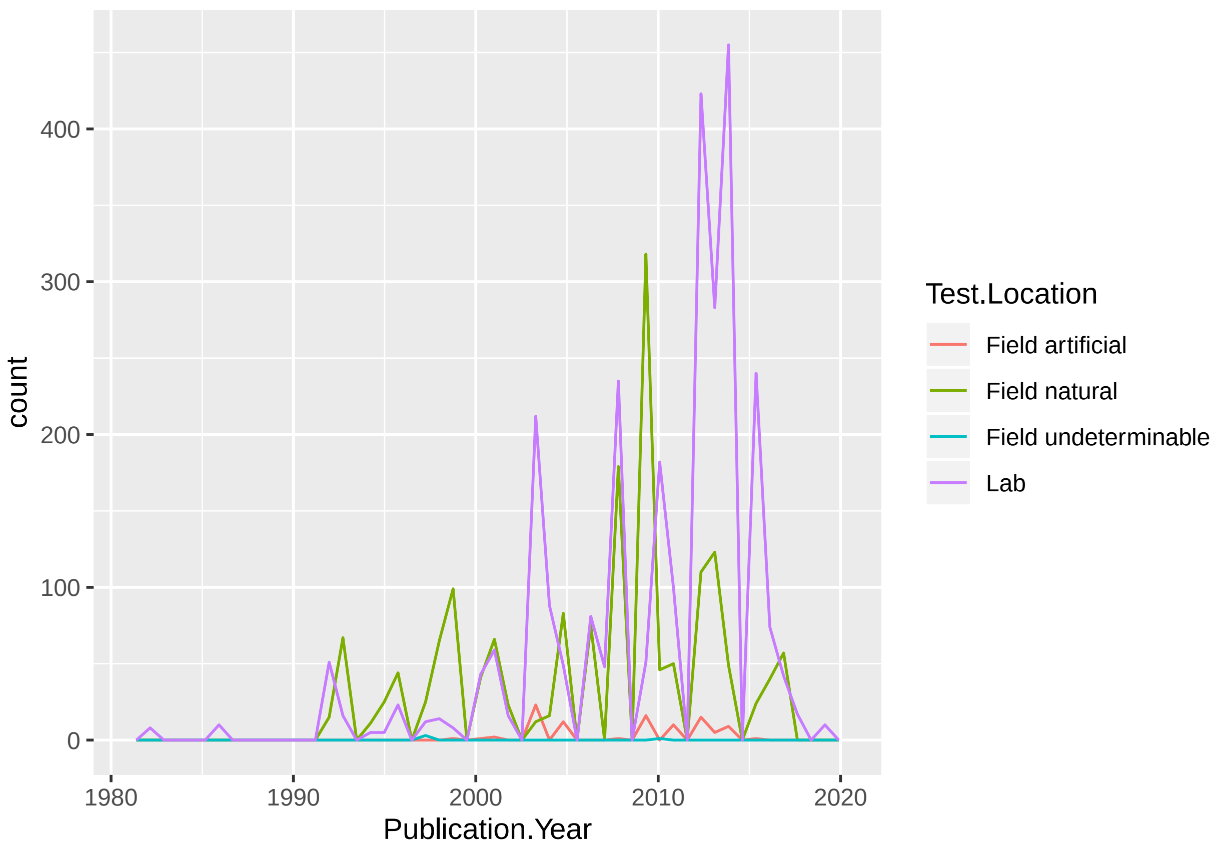
9. Using `geom_freqpoly`, generate a plot of the number of studies conducted by publication year. 1/1

```
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)
```

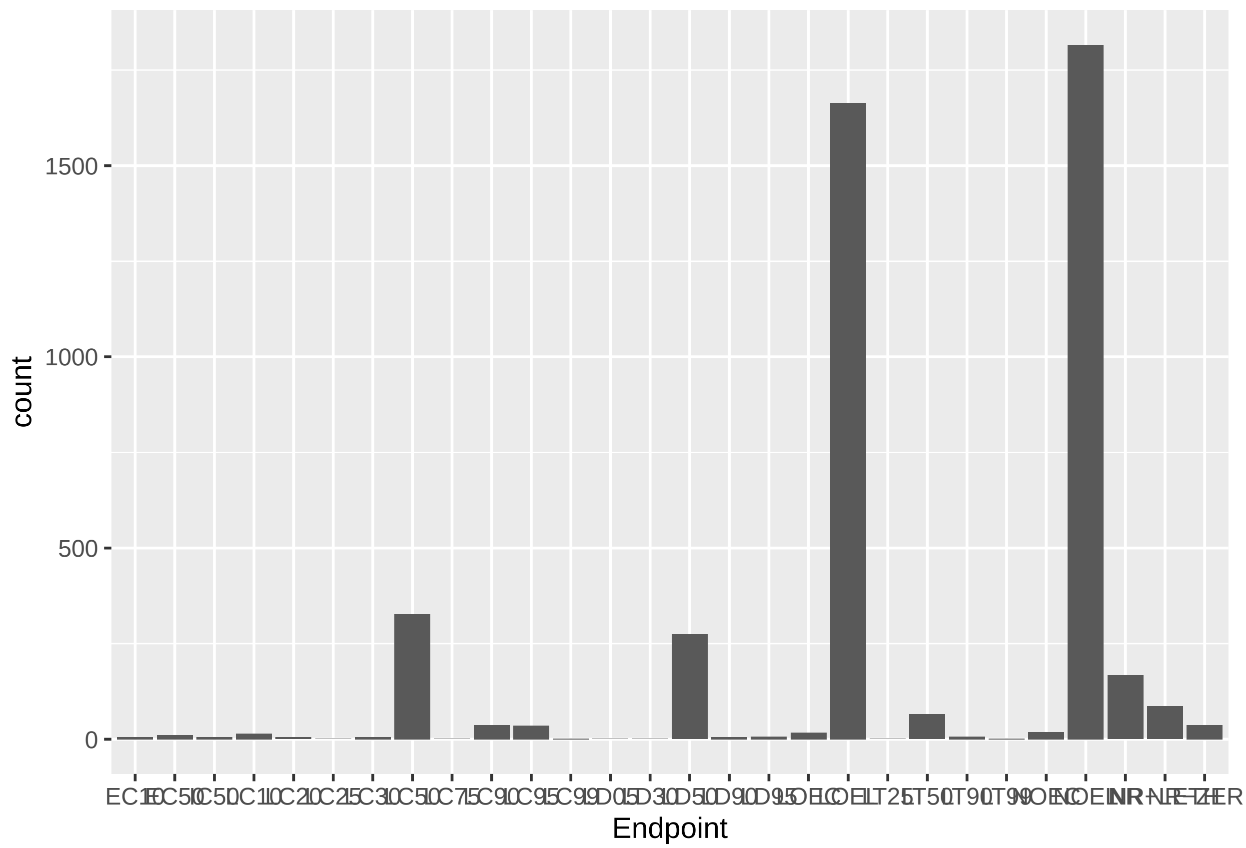



Interpret this graph. What are the most common test locations, and do they differ over time?

> Answer: The most common test locations are "field natural" and "lab". The frequency of testing in the "field natural" location peaks around 2008, with less frequent testing before and after this period. By contrast, testing in "lab" locations increases dramatically with time, peaking around 2012 and 2014, tapering slightly after this, but still remaining quite high until around 2017. 1/1

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 endpoints are LOEL and NOEL. As per the ECOTOX_CodeAppendix, LOEL is defined as the "lowest-observable-effect-level", or the lowest dose producing effects that were significantly different from responses of controls. NOEL is defined as "no-observable-effect-level", or the highest dose producing effects not significantly different from responses of controls according to author's reported statistical test. 1/1

## 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. 1.5/1.5

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

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

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

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

```
collection_date <- Litter$collectDate
collection_date
```

```
##   [1] "2018-08-02" "2018-08-02" "2018-08-02" "2018-08-02" "2018-08-02"
##   [6] "2018-08-02" "2018-08-02" "2018-08-02" "2018-08-02" "2018-08-02"
##  [11] "2018-08-02" "2018-08-02" "2018-08-02" "2018-08-02" "2018-08-02"
##  [16] "2018-08-02" "2018-08-02" "2018-08-02" "2018-08-02" "2018-08-02"
```

```
##  [21] "2018-08-02" "2018-08-02" "2018-08-02" "2018-08-02" "2018-08-02"
##  [26] "2018-08-02" "2018-08-02" "2018-08-02" "2018-08-02" "2018-08-02"
##  [31] "2018-08-02" "2018-08-02" "2018-08-02" "2018-08-02" "2018-08-02"
##  [36] "2018-08-02" "2018-08-02" "2018-08-02" "2018-08-02" "2018-08-02"
##  [41] "2018-08-02" "2018-08-02" "2018-08-02" "2018-08-02" "2018-08-02"
##  [46] "2018-08-02" "2018-08-02" "2018-08-02" "2018-08-02" "2018-08-02"
##  [51] "2018-08-02" "2018-08-02" "2018-08-02" "2018-08-02" "2018-08-02"
##  [56] "2018-08-02" "2018-08-02" "2018-08-02" "2018-08-02" "2018-08-02"
##  [61] "2018-08-02" "2018-08-02" "2018-08-02" "2018-08-02" "2018-08-02"
##  [66] "2018-08-02" "2018-08-02" "2018-08-02" "2018-08-02" "2018-08-02"
##  [71] "2018-08-02" "2018-08-02" "2018-08-02" "2018-08-02" "2018-08-02"
##  [76] "2018-08-02" "2018-08-02" "2018-08-02" "2018-08-02" "2018-08-02"
##  [81] "2018-08-02" "2018-08-02" "2018-08-02" "2018-08-02" "2018-08-02"
##  [86] "2018-08-02" "2018-08-02" "2018-08-02" "2018-08-02" "2018-08-02"
##  [91] "2018-08-02" "2018-08-30" "2018-08-30" "2018-08-30" "2018-08-30"
##  [96] "2018-08-30" "2018-08-30" "2018-08-30" "2018-08-30" "2018-08-30"
## [101] "2018-08-30" "2018-08-30" "2018-08-30" "2018-08-30" "2018-08-30"
## [106] "2018-08-30" "2018-08-30" "2018-08-30" "2018-08-30" "2018-08-30"
## [111] "2018-08-30" "2018-08-30" "2018-08-30" "2018-08-30" "2018-08-30"
## [116] "2018-08-30" "2018-08-30" "2018-08-30" "2018-08-30" "2018-08-30"
## [121] "2018-08-30" "2018-08-30" "2018-08-30" "2018-08-30" "2018-08-30"
## [126] "2018-08-30" "2018-08-30" "2018-08-30" "2018-08-30" "2018-08-30"
## [131] "2018-08-30" "2018-08-30" "2018-08-30" "2018-08-30" "2018-08-30"
## [136] "2018-08-30" "2018-08-30" "2018-08-30" "2018-08-30" "2018-08-30"
## [141] "2018-08-30" "2018-08-30" "2018-08-30" "2018-08-30" "2018-08-30"
## [146] "2018-08-30" "2018-08-30" "2018-08-30" "2018-08-30" "2018-08-30"
## [151] "2018-08-30" "2018-08-30" "2018-08-30" "2018-08-30" "2018-08-30"
## [156] "2018-08-30" "2018-08-30" "2018-08-30" "2018-08-30" "2018-08-30"
## [161] "2018-08-30" "2018-08-30" "2018-08-30" "2018-08-30" "2018-08-30"
## [166] "2018-08-30" "2018-08-30" "2018-08-30" "2018-08-30" "2018-08-30"
## [171] "2018-08-30" "2018-08-30" "2018-08-30" "2018-08-30" "2018-08-30"
## [176] "2018-08-30" "2018-08-30" "2018-08-30" "2018-08-30" "2018-08-30"
## [181] "2018-08-30" "2018-08-30" "2018-08-30" "2018-08-30" "2018-08-30"
## [186] "2018-08-30" "2018-08-30" "2018-08-30"
```

```r
length(collection_date)
```

```
## [1] 188
```

```r
unique_collection_date <- unique(collection_date)
unique_collection_date
```

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

```r
length(unique_collection_date)
```

```
## [1] 2
```

Answer: Litter was sampled on August 2, 2018 and August 30, 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
plots <-Litter$plotID
length(plots)
```

```
## [1] 188
```

```
unique_plots <- unique(plots)
unique_plots
```

```
##  [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
```

```
length(unique_plots)
```

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

```
summary(unique_plots)
```

```
## 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
```

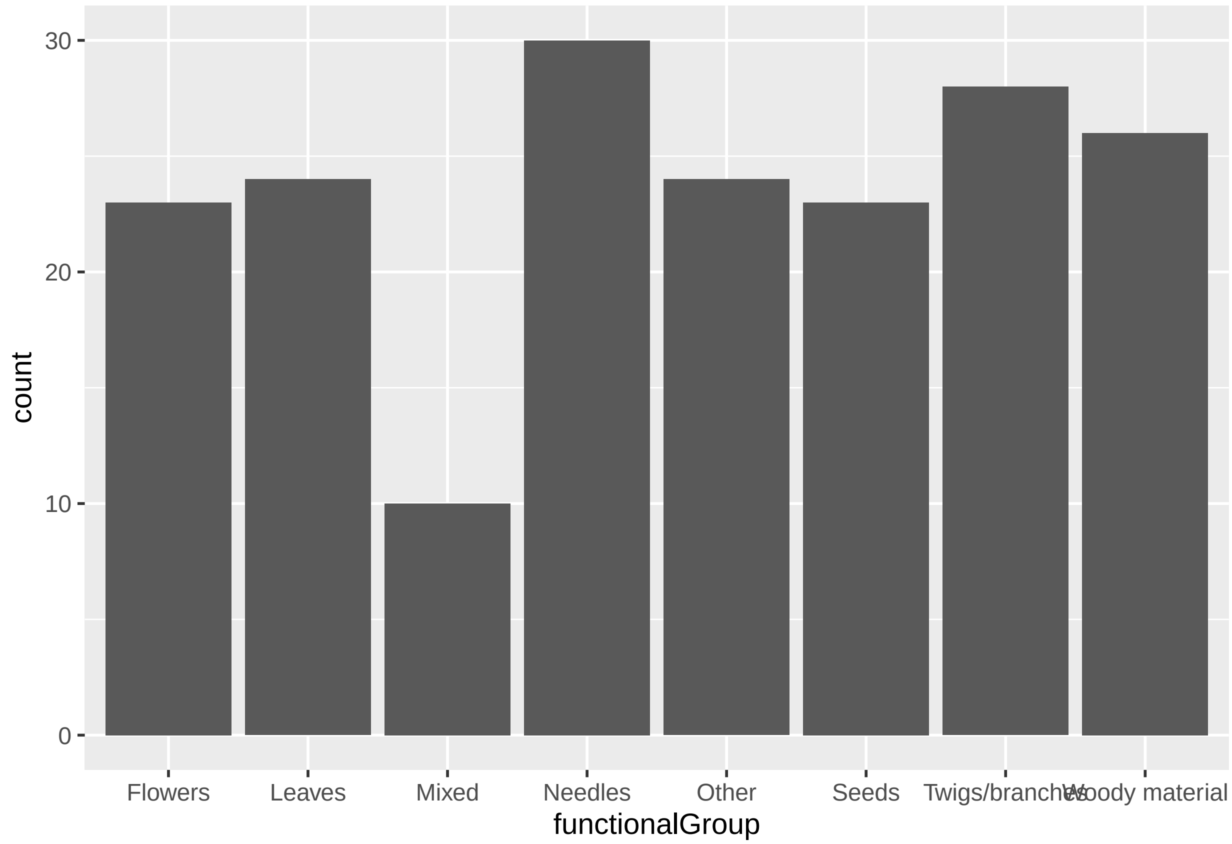
```
summary(plots)
```

```
## 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 reveals that 12 plots were sampled at Niwot Ridge. The 'unique' function presents a list of site numbers, explaining that they exist at 12 levels. Conversely, the 'summary' function provides a list of site numbers accompanied by counts of each plot name. 1/1

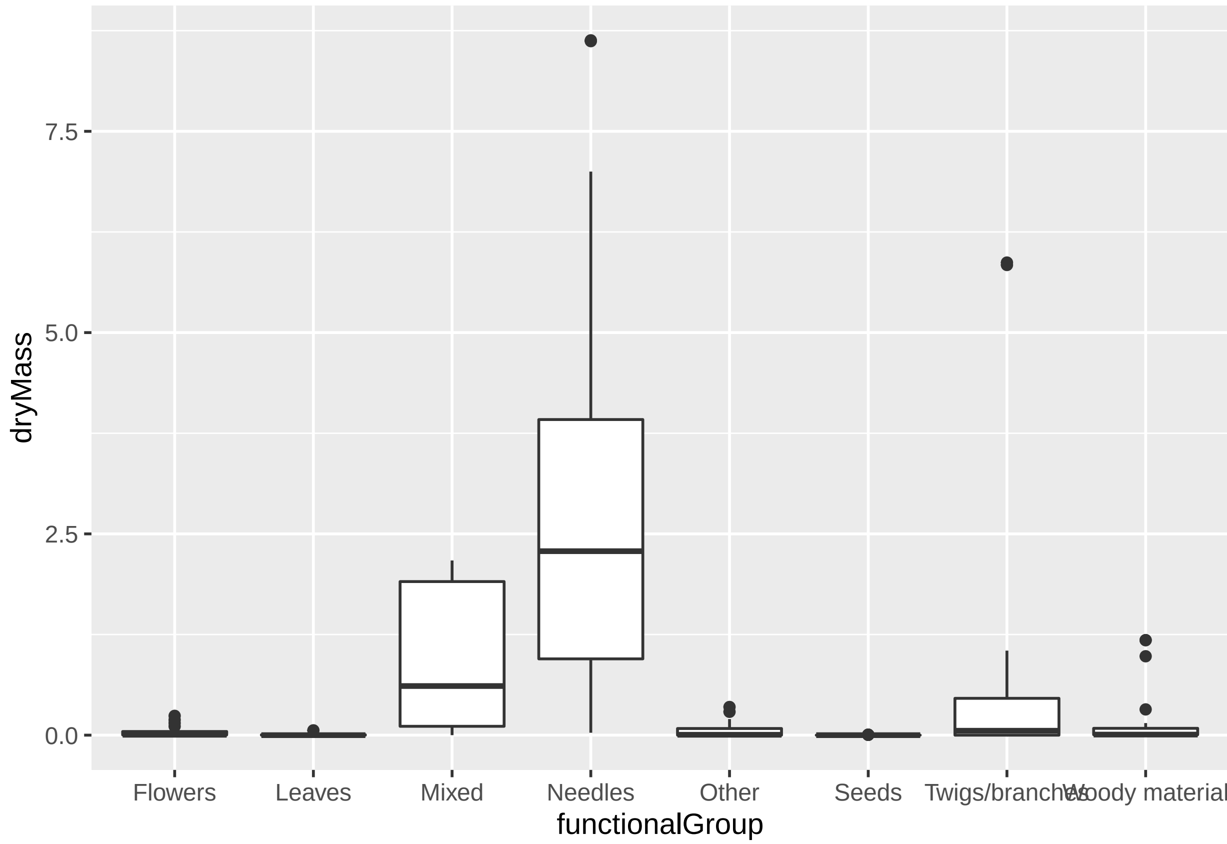
14. Create a bar graph of functionalGroup counts. This shows you what type of litter is collected at the Niwot Ridge sites. Notice that litter types are fairly equally distributed across the Niwot Ridge sites. 1/1

```
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.

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

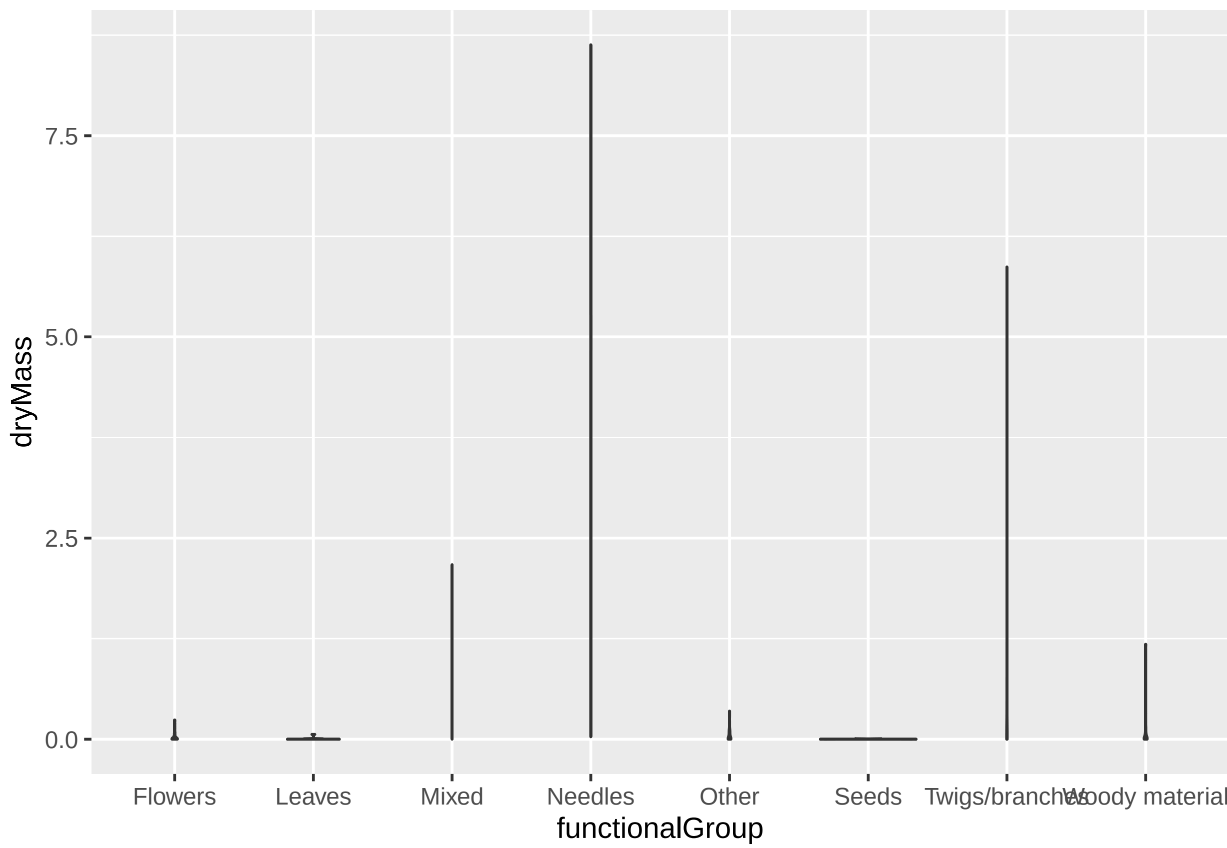



```
ggplot(Litter) +
  geom_violin(aes(x = functionalGroup, y = dryMass),
              draw_quantiles = c(0.25, 0.5, 0.75),
                   scale = "count")
```

```
## 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: The boxplot is able to show the distribution of dryMass for each functional group, wherein the median and spread of dryMass in each case can be compared side by side. By contrast, the violin plot is only able to show the spread of dryMass values for each group in this case; since the data is skewed it's very hard to see the data in the violin plot format.

2/2

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

> Answer: The mixed and needles functional groups have the highest biomass at these sites. Additionally, the twigs/branches group has an outlier with a particuarly high biomass value.