Assignment 3: Data Exploration

Srishti Mutha

Spring 2023

OVERVIEW

This exercise accompanies the lessons in Environmental Data Analytics on Data Exploration.

Directions

1. Rename this file `<FirstLast>_A03_DataExploration.Rmd` (replacing `<FirstLast>` with your first and last name).
2. Change "Student Name" on line 3 (above) with your name.
3. Work through the steps, **creating code and output** that fulfill each instruction.
4. Assign a useful **name to each code chunk** and include ample **comments** with your code.
5. Be sure to **answer the questions** in this assignment document.
6. When you have completed the assignment, **Knit** the text and code into a single PDF file.
7. After Knitting, submit the completed exercise (PDF file) to the dropbox in Sakai.

TIP: If your code extends past the page when knit, tidy your code by manually inserting line breaks.

TIP: If your code fails to knit, check that no `install.packages()` or `View()` commands exist in your code.

Set up your R session

1. Check your working directory, load necessary packages (tidyverse, lubridate), 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 include the subcommand to read strings in as factors.

```
getwd()
```

```
## [1] "/Users/cherry/Desktop/EDA/EDA-Spring2023"
```

```
#install.packages("tidyverse")
#install.packages("lubridate")

library(tidyverse)
library(lubridate)
```

```r
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 ecotoxicology of neonicotinoids on insects? Feel free to do a brief internet search if you feel you need more background information.

   Answer:neonicotinoids are some of the most deadly pesticides ever created. The problem is that they kill indiscriminately, exterminating not only "pest" insects but also countless butterflies, bees, and other wildlife. When it is applied directly to the soil around a plant's roots, or as a coating on a plant seed, which the plant then literally soaks up as it grows. That makes the plant itself—including its nectar, pollen, leaves, stems, and fruit—toxic. Thus it is intersting to study about it and researching more about its effects.

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: The litter and woody debris that falls to the ground are intersting to study as they affect the health of the soil. They fall and decay in the forest add nutrients to the forest soil and retain moisture in the forest. Moreover, fine woody material dries quickly and therefore decays slowly. Fine woody debris may act as a tinder that promotes the start and spreading of forest fires. Sampling the extent of fine woody debris indicates fire risk and gives clues about the rate of forest decay.

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 sampling is executed at terrestrial NEON sites that contain woody vegetation on >2m tall. Along with most of NEON's plant productivity measurements, sampling for this product occurs only in tower plots (AD[06]). Locations of tower plots are selected randomly within the 90% flux footprint of the primary and secondary airsheds 1. 2. 3.

Obtain basic summaries of your data (Neonics)

5. What are the dimensions of the dataset?

```r
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?

```
colnames(Neonics)
```

```
##  [1] "CAS.Number"                "Chemical.Name"
##  [3] "Chemical.Grade"            "Chemical.Analysis.Method"
##  [5] "Chemical.Purity"           "Species.Scientific.Name"
##  [7] "Species.Common.Name"       "Species.Group"
##  [9] "Organism.Lifestage"        "Organism.Age"
## [11] "Organism.Age.Units"        "Exposure.Type"
## [13] "Media.Type"                "Test.Location"
## [15] "Number.of.Doses"           "Conc.1.Type..Author."
## [17] "Conc.1..Author."           "Conc.1.Units..Author."
## [19] "Effect"                    "Effect.Measurement"
## [21] "Endpoint"                  "Response.Site"
## [23] "Observed.Duration..Days."  "Observed.Duration.Units..Days."
## [25] "Author"                    "Reference.Number"
## [27] "Title"                     "Source"
## [29] "Publication.Year"          "Summary.of.Additional.Parameters"
```

```
x <- sort(Neonics$Effect, descending=TRUE)
summary(x)
```

```
##     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 that are studied are population and mortality

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.[TIP: The `sort()` command can sort the output of the summary command...]

```
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
```

3

```
##                               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 Honey Bee, Parasitic Wasp, Buff Tailed Bumblebee, Carniolan Honey Bee, Bumble Bee and Italian Honeybee. What they have in common is that most cooonly studies species are bees as neonicotinoid insecticides appear the most toxic to bees.

8. Concentrations are always a numeric value. What is the class of `Conc.1..Author.` column in the dataset, and why is it not numeric?

```
class(Neonics$Conc.1..Author.)
```
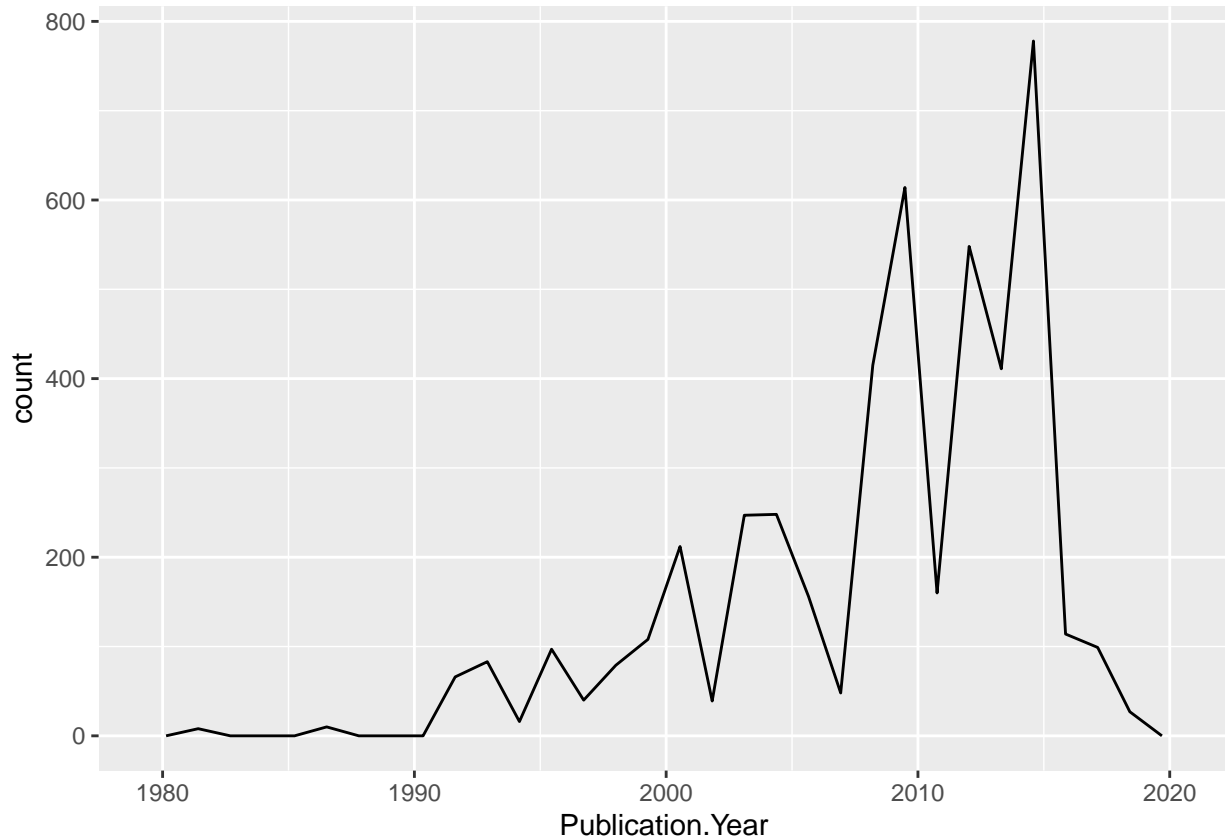
```
## [1] "factor"
```

Answer:The class of the column is Factor as we had added a suncommand to add the strings as factor.

Explore your data graphically (Neonics)

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

```
ggplot(data = Neonics, mapping = aes(x=Publication.Year)) +
  geom_freqpoly()
```
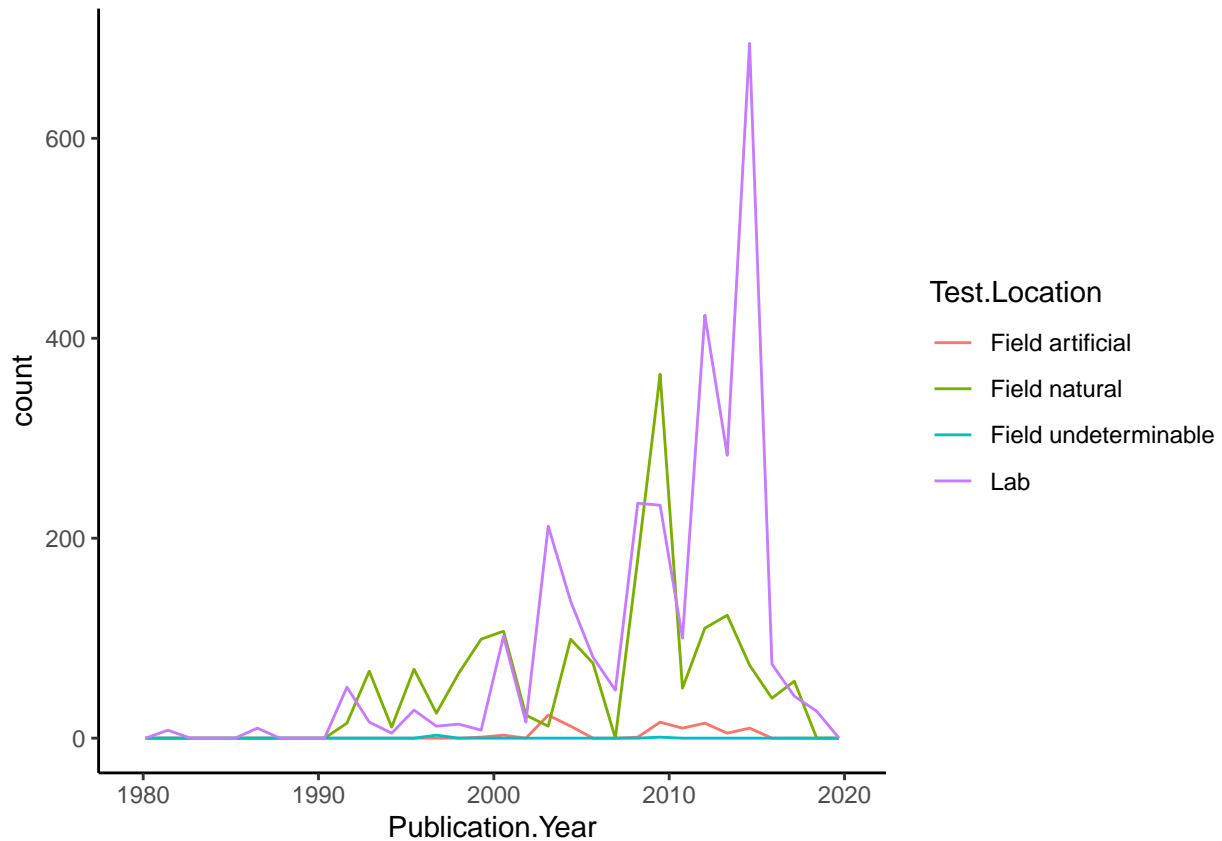
`stat_bin()` using `bins = 30`. Pick better value with `binwidth`.

10. Reproduce the same graph but now add a color aesthetic so that different Test.Location are displayed as different colors.

```
ggplot(data = Neonics, mapping = aes(x=Publication.Year, colour=Test.Location)) +
  geom_freqpoly() +
  theme_classic()
```

`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: Lab is the most common test location, yes they differ over time. In 2010, feild natural's count was more than Lab which changed later.

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.

[**TIP**: Add `theme(axis.text.x = element_text(angle = 90, vjust = 0.5, hjust=1))` to the end of your plot command to rotate and align the X-axis labels...]

```
ggplot(data = Neonics, mapping = aes(x=Endpoint)) +
  geom_bar() +
  theme_light() +
  theme(axis.text.x = element_text(angle = 90, vjust = 0.5, hjust=1))
```
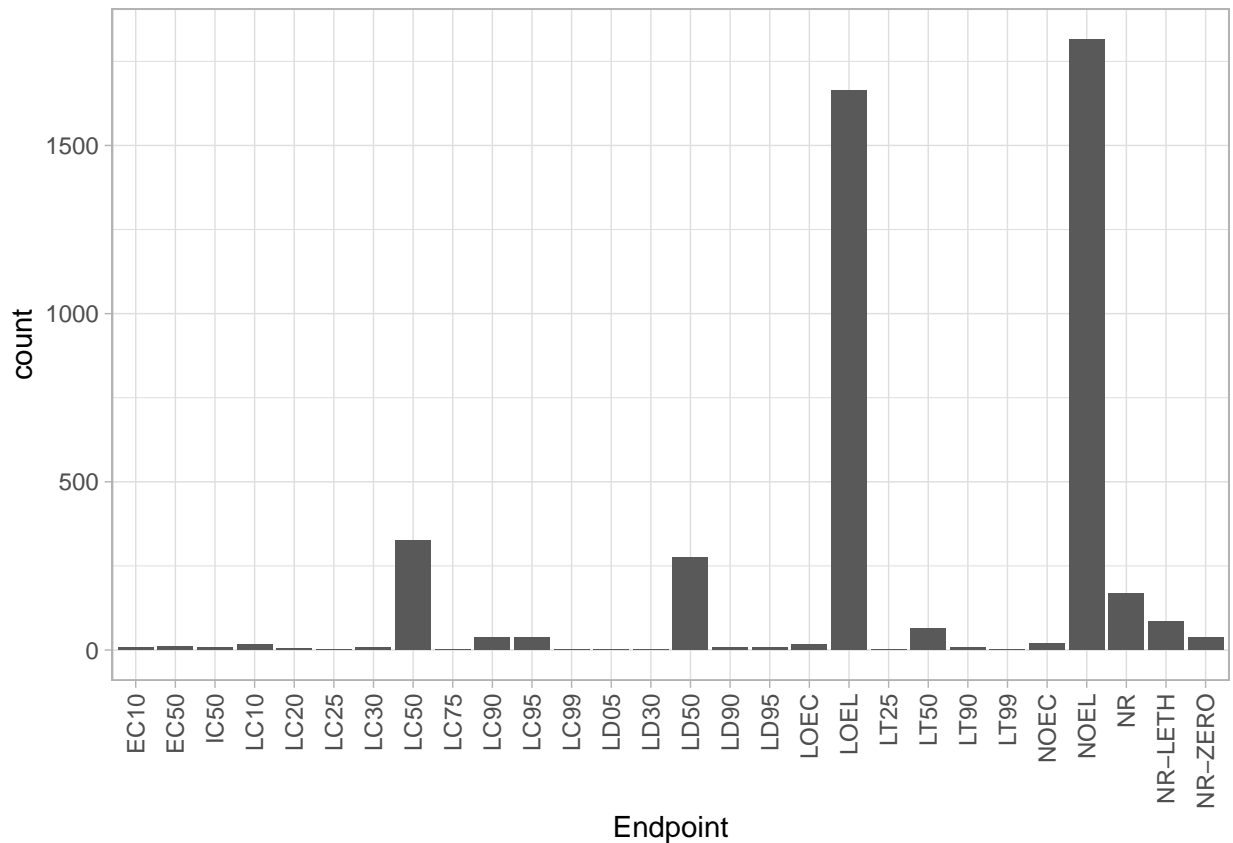
Answer: The most common end points are LOEL and NOEL

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$collectDate <- ymd(Litter$collectDate)
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$namedLocation)
```

```
##  [1] NIWO_061.basePlot.ltr NIWO_064.basePlot.ltr NIWO_067.basePlot.ltr
##  [4] NIWO_040.basePlot.ltr NIWO_041.basePlot.ltr NIWO_063.basePlot.ltr
##  [7] NIWO_047.basePlot.ltr NIWO_051.basePlot.ltr NIWO_058.basePlot.ltr
## [10] NIWO_046.basePlot.ltr NIWO_062.basePlot.ltr NIWO_057.basePlot.ltr
## 12 Levels: NIWO_040.basePlot.ltr ... NIWO_067.basePlot.ltr
```

```
summary(Litter$namedLocation)
```

```
## NIWO_040.basePlot.ltr NIWO_041.basePlot.ltr NIWO_046.basePlot.ltr
##                    20                    19                    18
## NIWO_047.basePlot.ltr NIWO_051.basePlot.ltr NIWO_057.basePlot.ltr
##                    15                    14                     8
## NIWO_058.basePlot.ltr NIWO_061.basePlot.ltr NIWO_062.basePlot.ltr
##                    16                    17                    14
## NIWO_063.basePlot.ltr NIWO_064.basePlot.ltr NIWO_067.basePlot.ltr
##                    14                    16                    17
```
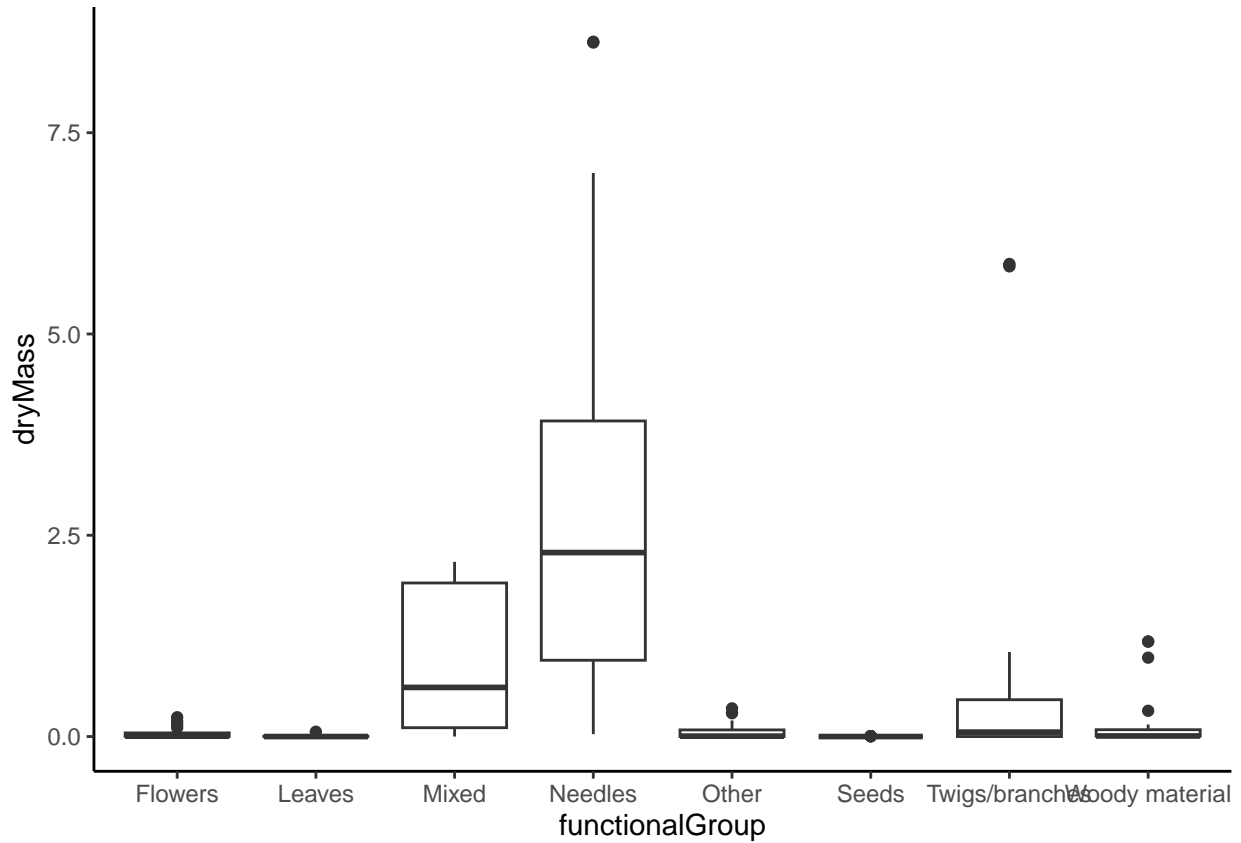
Answer:...

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(data = Litter, mapping = aes(y = functionalGroup)) +
  geom_bar() +
  theme_classic()
```
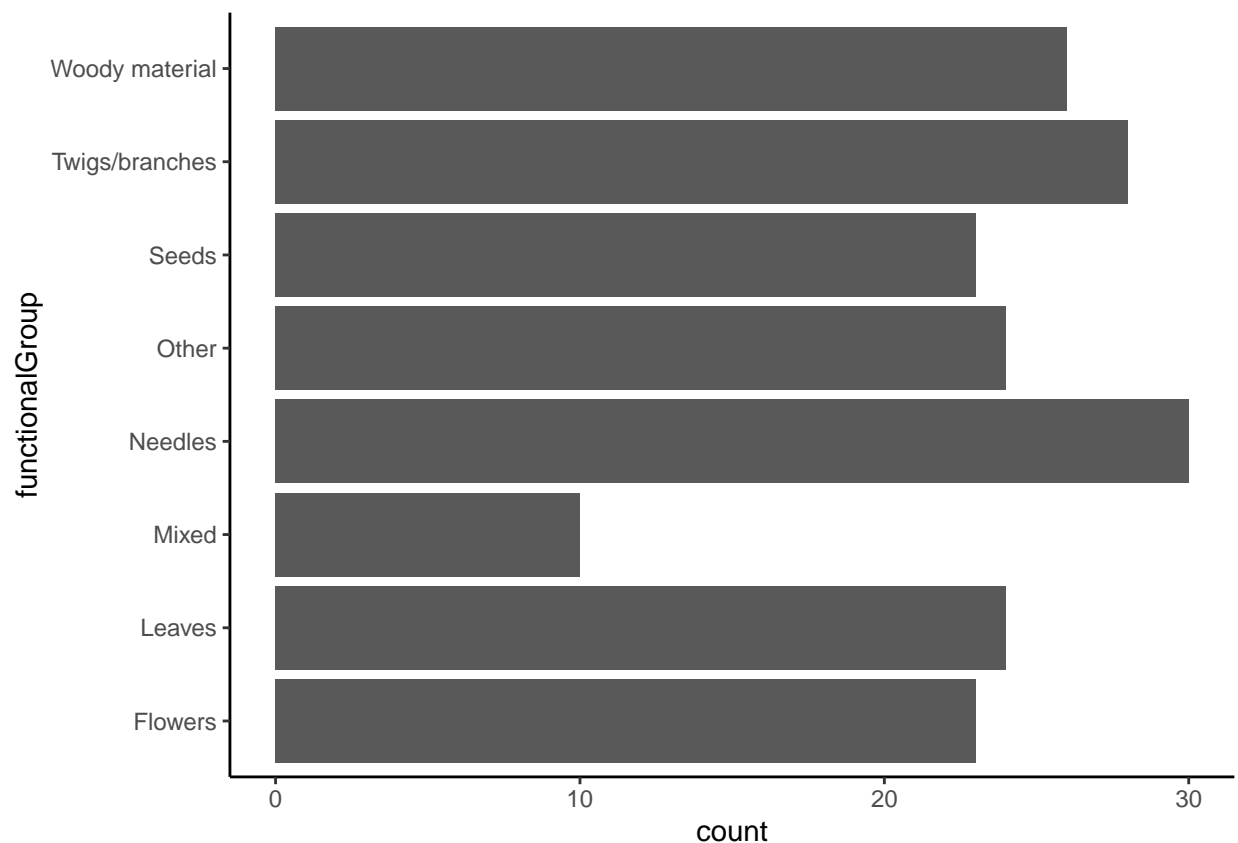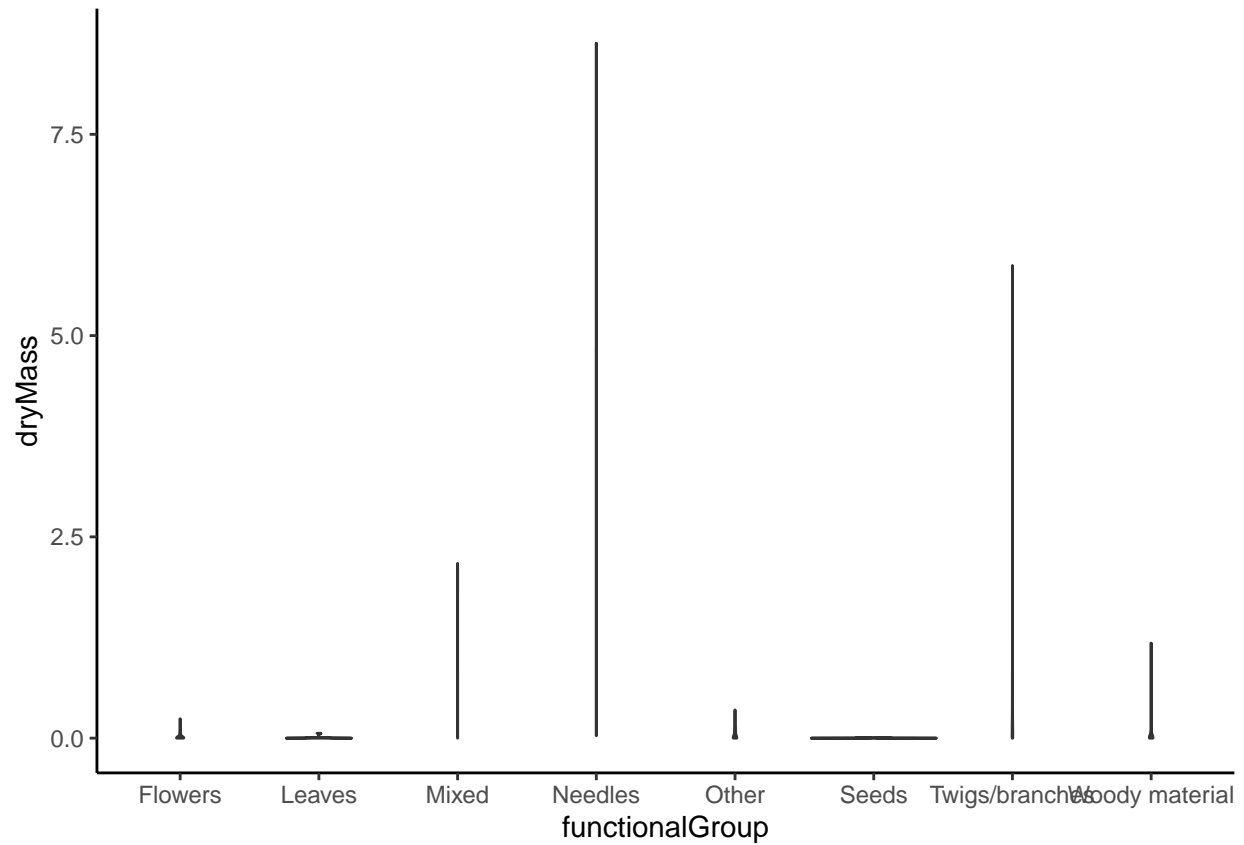
15. Using `geom_boxplot` and `geom_violin`, create a boxplot and a violin plot of dryMass by functional-Group.

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



```
ggplot(data = Litter, mapping = aes(y = dryMass, x = functionalGroup)) +
  geom_violin() +
  theme_classic()
```

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

Answer: Boxplot is more effective visualization as violin plot is very confusing

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

Answer:Needles