Assignment 3: Data Exploration

Nadia Swit

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() #changed global options to current working directory so that it referred to main folder instea

library(tidyverse)
library(lubridate)
Neonics <- read.csv("./Data/Raw/ECOTOX_Neonicotinoids_Insects_raw.csv")
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: Neonicotinoids can inadvertently harm beneficial insects, such as bees and other pollinators. Furthermore, other non-target organisms, including humans, can be experiencing environmental risks associated with the pesticides. As many are water-soluble, chemical compounds can be absorbed and processed by plants and crops and can be dispersed throughout the environment. Repeated first-hand exposure without proper PPE can cause negative side-effects to humans, in addition to impacting fetal development and causing birth defects.

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: Forest litter and woody debris can serve as indicators to the health of the forest. The amount of organic matter can impact the quality of habitat for organisms within the forest, as well as forage material. Composition of the litter and debris can also relay changing forest dynamics, such as if other tree species are dominating forests and altering the environment. Furthermore, the amount of debris, especially leaf litter, can also serve as a proxy for changes in climate and signify if seasons are changing at unsual rates. Woody debris, especially large woody debris, is important for sediment retention and stream channel morphology.

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 was collected from individual sampling bouts and collected from elevated and ground traps. * Mass data was recorded to the closest 0.01 g and sorted based on functional group. * Location of tower plots are selected randomly based on its relative location to primary and secondary airsheds, and depending on vegatitve cover and height, plot size would vary.

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 "Effects" 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
```

```
#included as.factor because effect is a character
```

Answer: This information will indicate what effects the neonicotinoids have on the targeted species group. The most common effects including mortality and population growth.

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
```

Answer: The 6 most common species are honey bees, parasitic wasps, buff tailed bumblebees, Carniolan honey bee, bumble bee, and Italian honey bee. The majority of these species are important pollinators, therefore it would be determinental if they suffered effects from the pesticide. However, parasitc wasps would be a crucial targeted species to reduce pollinator mortality.

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.)
```

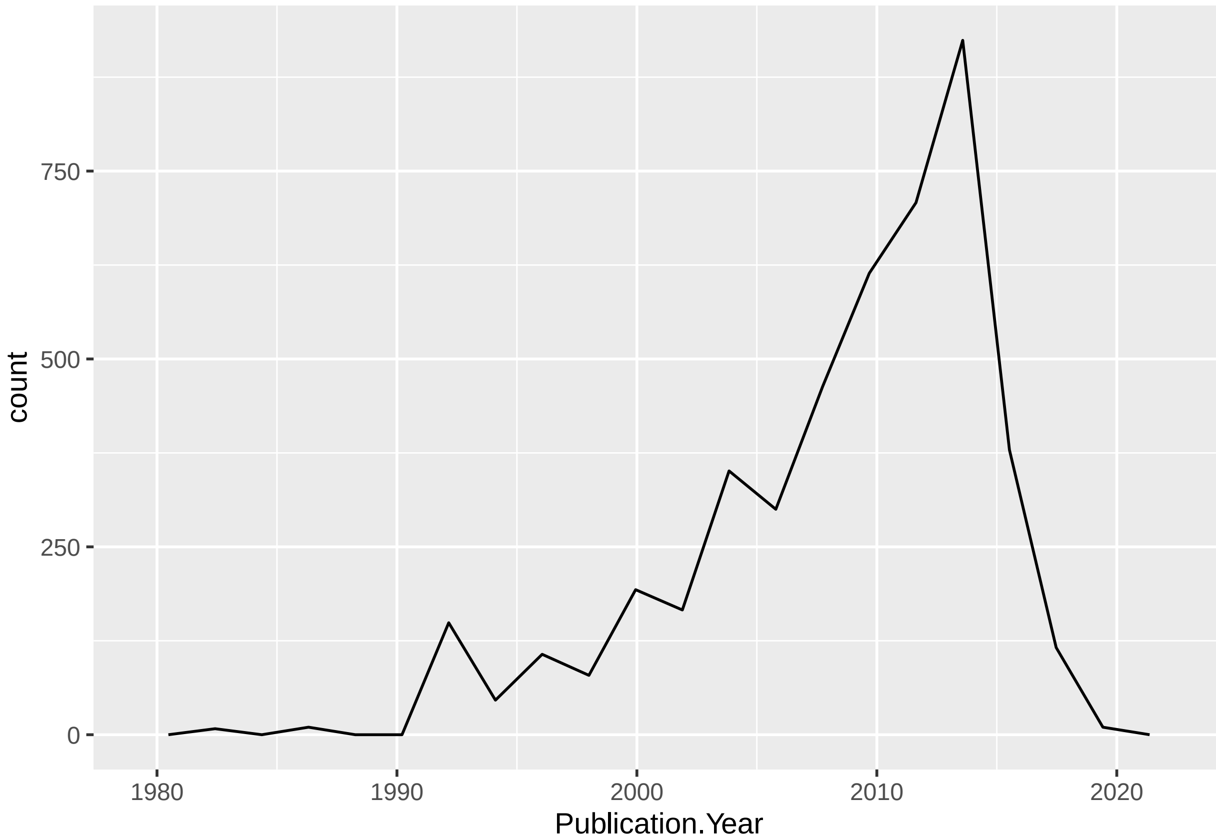
```
## [1] "character"
```

Answer: There are non-numerical values within the class of Conc.1…Author. Since these are not all numbers it makes the value character

Explore your data graphically (Neonics)

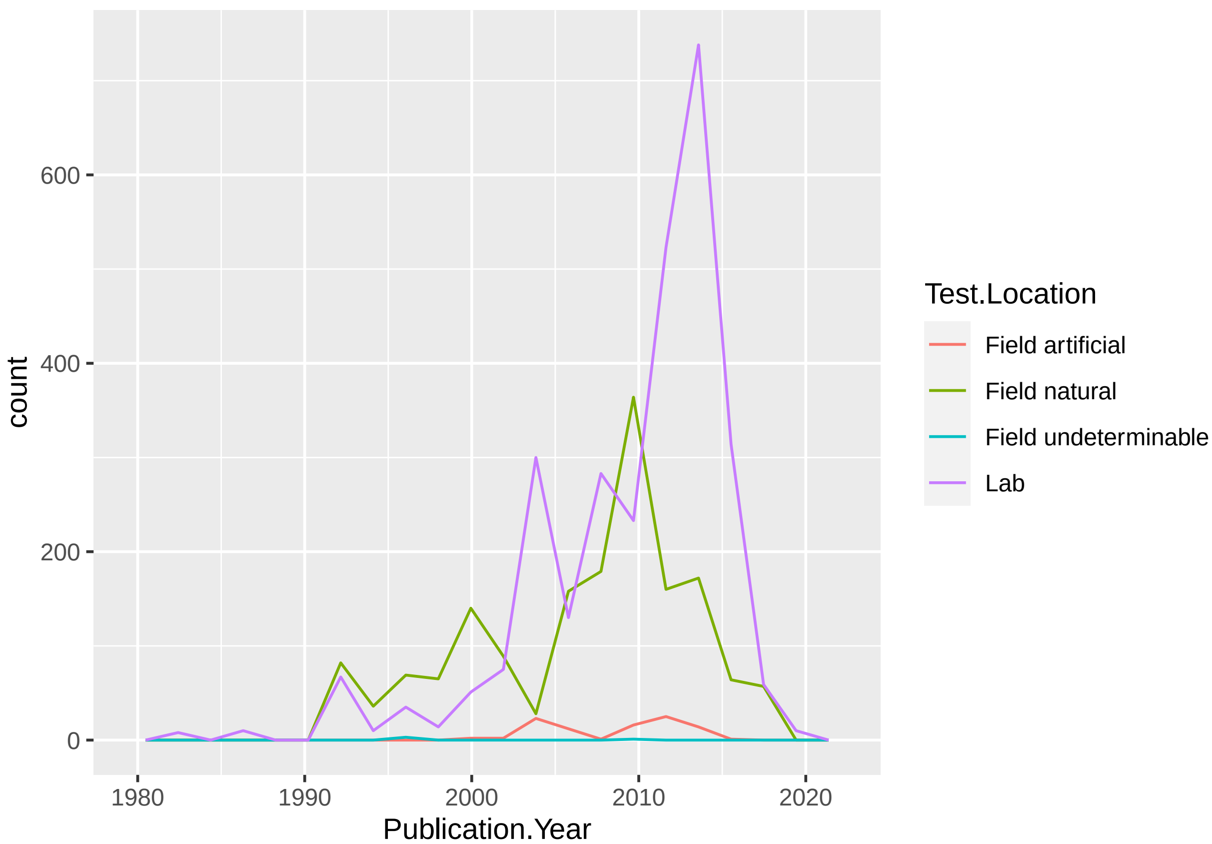
9. Using `geom_freqpoly`, generate a plot of the number of studies conducted by publication year.

```
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)
```

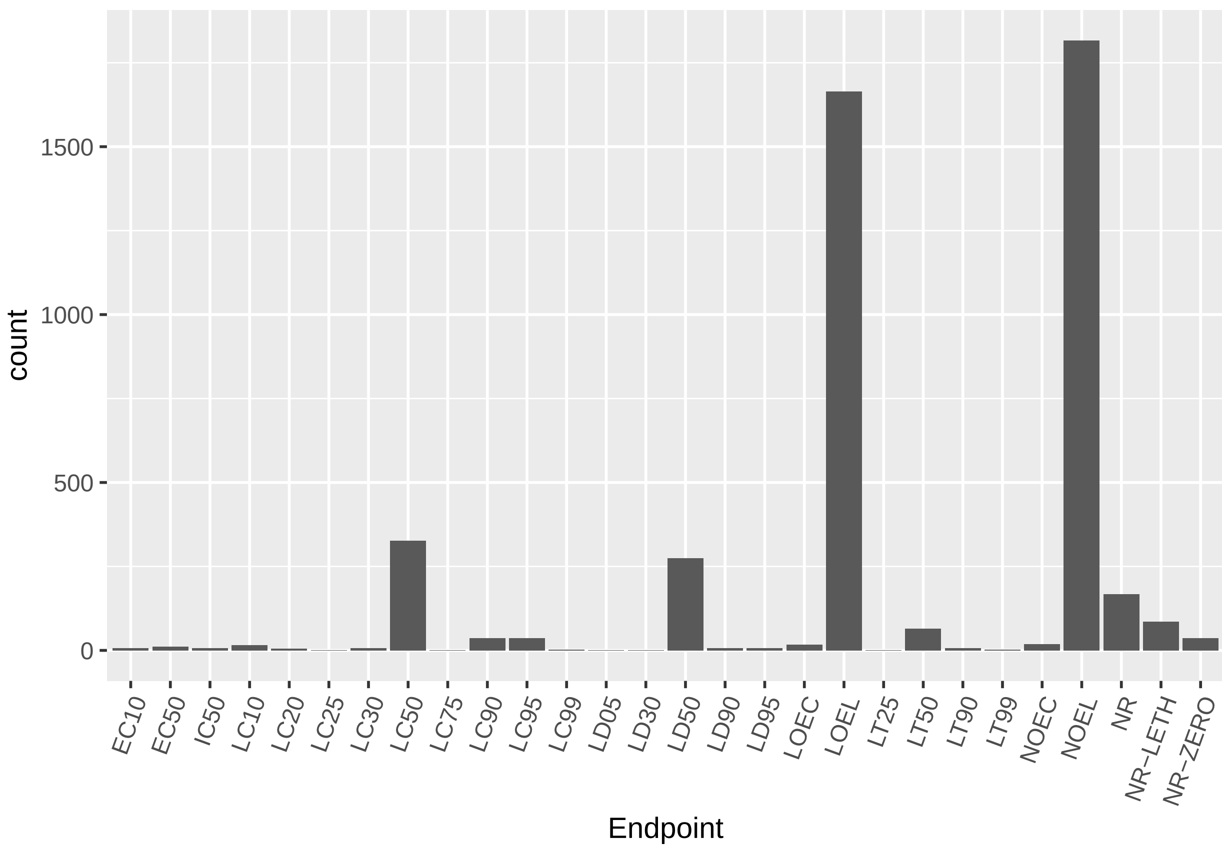


Interpret this graph. What are the most common test locations, and do they differ over time?

> Answer: The most common test locations are in the lab and in the natural field settings. Lab settings increased after 2000, and sharply increased in 2010. Natural field settings increased at a more gradual rate between 1990 and 2000 with some drops, but experienced a significant drop around 2005. Subsequently they increased in 2010 but then decreased again.

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 (Endpoint)) +
  theme(axis.text.x = element_text(angle=70, hjust=1))
```



Answer: The two most common Endpoints are LOEL and NOEL. LOEL, lowest-observable-effect-level, is a terrestiral endpoint. It reports the lowest dose (concentration) that produces effects that were significantly different from control responses. NOEL, also a terrestrial endoint, is no observed effects residue. This reports the highest residue concentration that produces effects, which is not significantly different from control responses.

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) #collectDate originally a character
```

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

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

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$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"
```

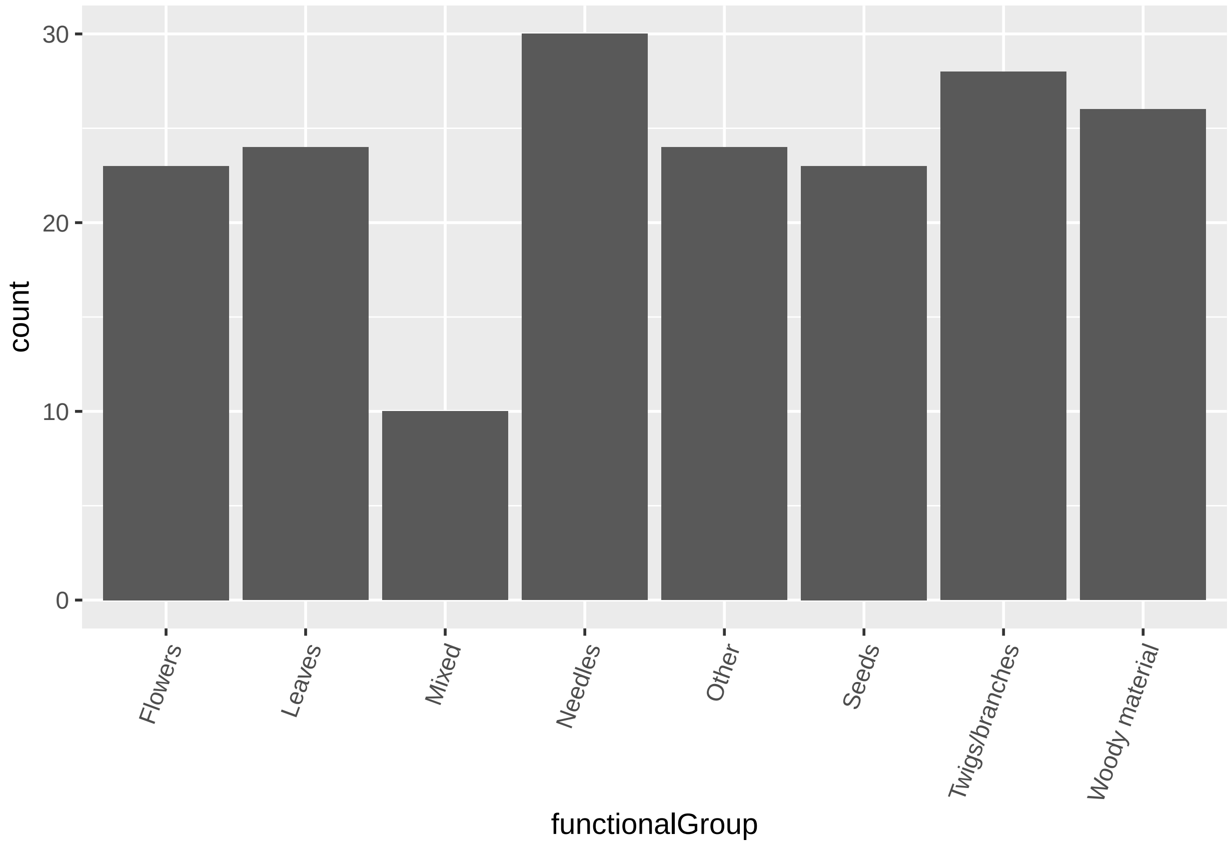
```
summary(Litter$plotID)
```

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

> Answer: The unique function presents the unique values within the field. For plotID, there are 12 different types of plots, where samples correspond to one of those types. This differes from the summary function, where it reorts the total number of observations within the field. This reports that there are 188 observations.

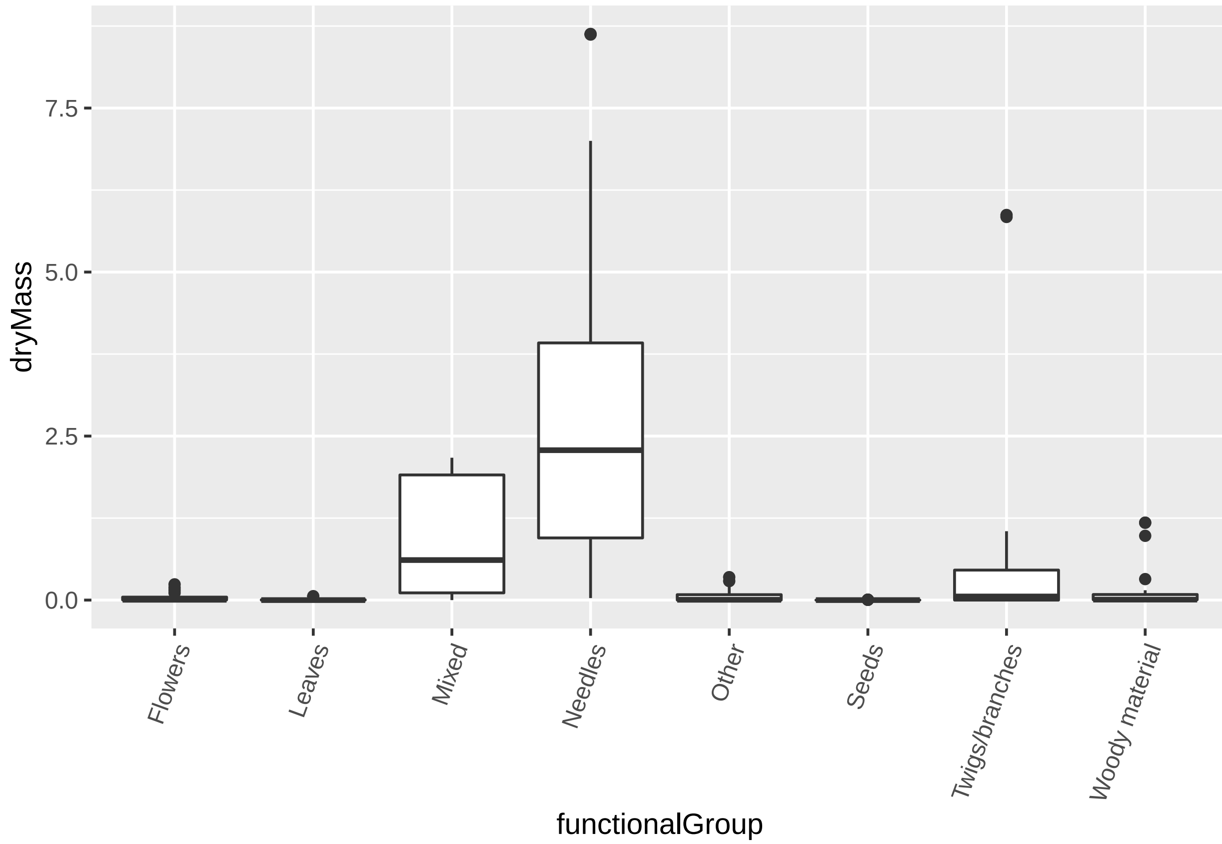
14. Create a bar graph of functionalGroup counts. This shows you what type of litter is collected at the Niwot Ridge sites. Notice that litter types are fairly equally distributed across the Niwot Ridge sites.

```
ggplot(Litter) +
  geom_bar(aes (functionalGroup)) +
  theme(axis.text.x = element_text(angle=70, hjust=1))
```



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=70, hjust=1))
```

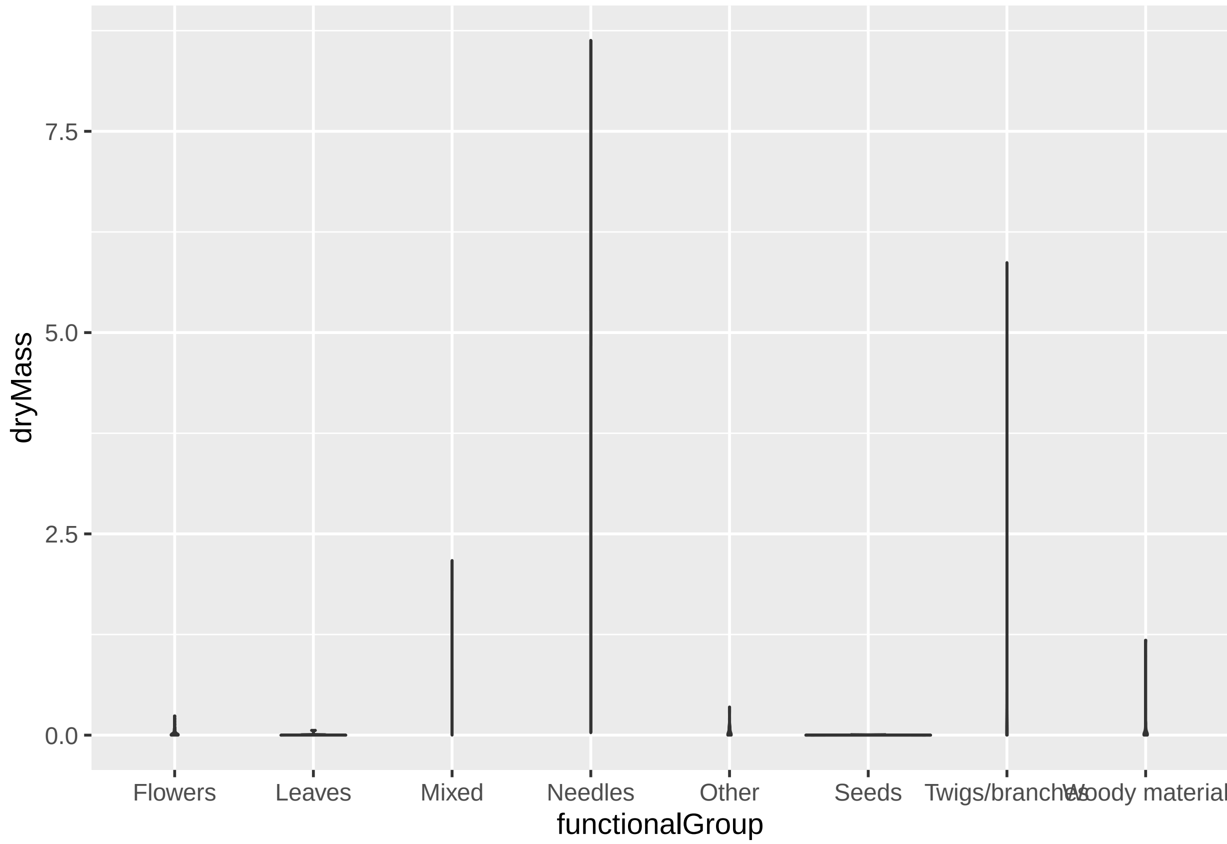



```
ggplot(Litter) +
  geom_violin(aes(x = functionalGroup, y = dryMass),
              draw_quantiles = c(0.25, 0.5, 0.75))
```

```
## 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: The boxplot is better at displaying the spread of teh data for each functional group.Even though the range for some of the groups was very small, you are also able to visualize outliers. However, the violin plot is not an adequate visualization. Since the values in each functional group are similar and the distribution density is small, the plot is not capable of displaying the probability distrubution within the plot since the values are too similar.

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

> Answer: Mixed litter and needle litter have the highest biomass at the sample sites.