Assignment 3: Data Exploration

Dori Rathmell

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. Be sure to **answer the questions** in this assignment document.
5. When you have completed the assignment, **Knit** the text and code into a single PDF file.
6. After Knitting, submit the completed exercise (PDF file) to the dropbox in Sakai.

The completed exercise is due on Sept 30th.

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 include the subcommand to read strings in as factors.

```
#install.packages('formatR')
knitr::opts_chunk$set(tidy.opts=list(width.cutoff=80), tidy=TRUE) #wraps text for knitted file




getwd() #shows working directory
```

```
## [1] "/home/guest/R/EDA Fall/Assignments"
```

```
library(tidyverse) #load tidyverse package
library(ggplot2)
#open Neonics file
Neonics <- read.csv("../Data/Raw/ECOTOX_Neonicotinoids_Insects_raw.csv", stringsAsFactors = TRUE)
#open Litter file
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: Insecticides used for agricultural production can severely impact the surrounding insect population, even insects for which the pesticide is not targeted.

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 release nutrients into the soil and are critical for maintaining soil health, and leaf litter/woody debris is an important habitat for insect species

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: 1. litter and woody debris is sampled at areas with woody vegetation that is greater than 2m 2. One litter trap is placed every 400 square plot area, with 1 - 4 trap pairs per plot 3. Depending on trap location, the trap placement can either be targeted or randomized

Obtain basic summaries of your data (Neonics)

5. What are the dimensions of the dataset?

```
# view(Neonics)
dim(Neonics)  #displays dimensions of Neonics files, there are 4623 rows of 30 columns
```

```
## [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?

```
summary(Neonics$Effect)  #summarizes data in effect column of Neonics file
```

```
##     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: Population, Mortality, and behavior are the three most common effects that are studied. This could be because they are broader topics, the other topics studied fall under the umbrella of these three topics, ie. reproduction, avoidance, and food behavior all fall under the 'behavior' category. Additionally all three of the most commonly studied factors are highly impacted by the presence of insecticides since both the population size, health and behavior of insects are all impacted with pesticides.

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)  #summarizes data of Species.Common.Name from the Neonics file
```

```
##                          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 honey bee, the parasitic wasp, the buff tailed honeybee, the Carniolan Honey Bee, the Bumble Bee and the Italian Honeybee are the 6 most commonly studied insects. This is likely because bees as a whole are incredibly endangered, with certain bee species being more endangered than others and parasitic wasps are a direct predator of bees and their presence negatively impacts surrounding bee populations. Additionally, parasitic wasps and bees are excellent pollinators,

they are highly vulnerable to 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.)  #displays the class status of Conc.1..Author
```

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

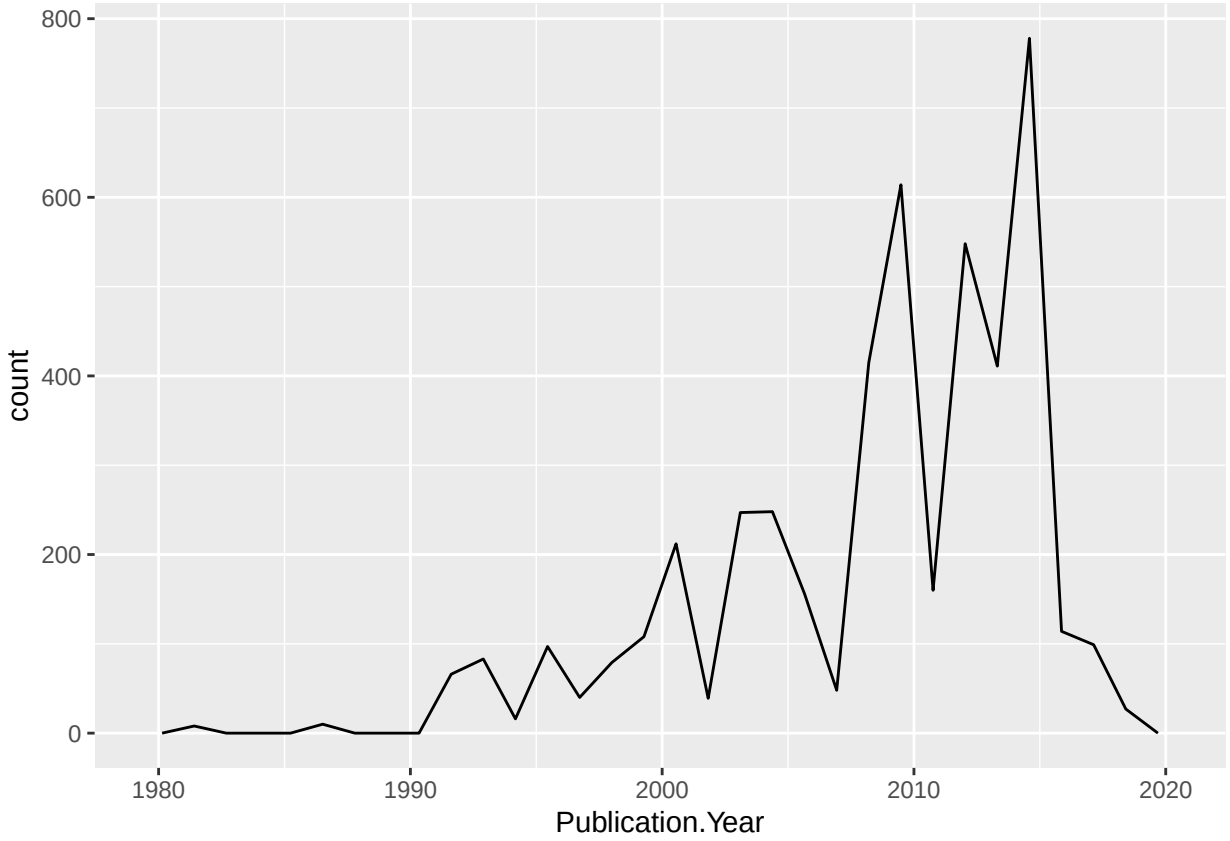
Answer:The class is factors, because some data is listed as "NR" (not recorded) strings and earlier in the assignment we ran the strings as factors function, making the data a factor class.

Explore your data graphically (Neonics)

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

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

```
## `stat_bin()` using `bins = 30`. Pick better value with `binwidth`.
```

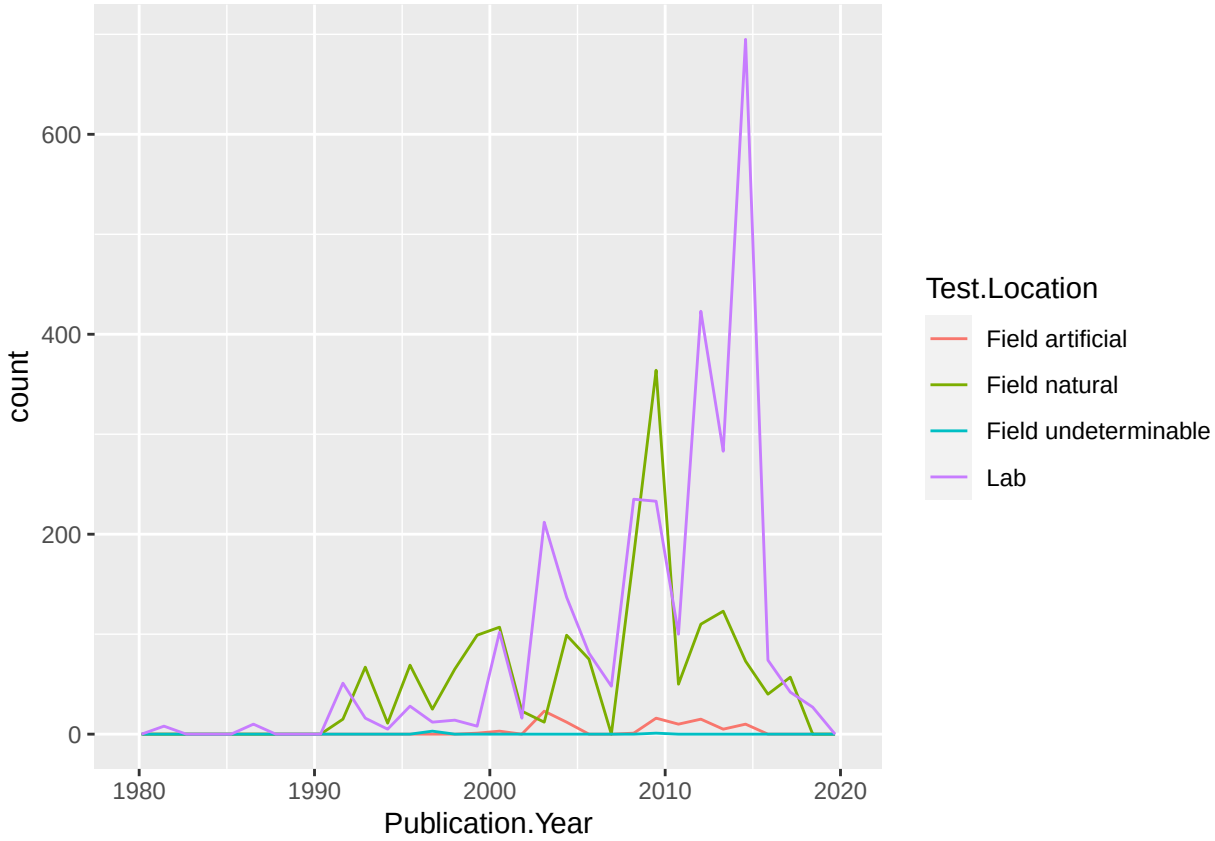



```
# creates a line graph publication year from the Neonics file
```

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))
```

```
## `stat_bin()` using `bins = 30`. Pick better value with `binwidth`.
```



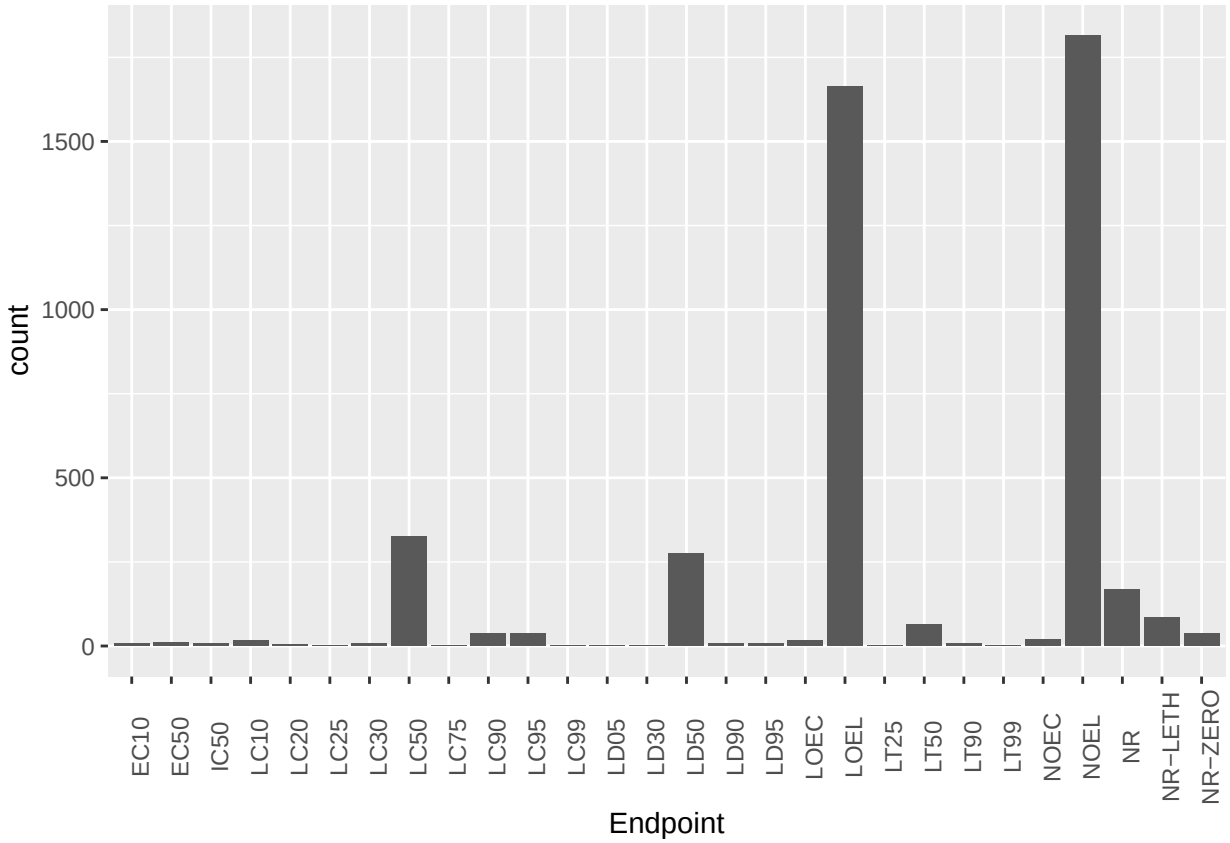

```
# creates the same graph, but differentiates publication year by test location
```

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

> Answer: The most common test location is the lab, but prior to 2010 the natural field was another common test location.

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() + theme(axis.text.x = element_text(angle = 90))
```




```
# creates a bar chart of the endpoints for the studies, also turned endpoint
# labels 90 degrees to read better
```

Answer: The two most common endpoints are LOEL (the lowest observable effect level), meaning the endpoint with the lowest dose that produced results different from experimental controls and NOEL (no observable effect level) the highest dose that does not produce effects different from the experimental 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)  #verifies which class the collectDate data is in the Litter file
```

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

```
collectDate_2 <- as.Date(Litter$collectDate)  #changes the class from factor to date
class(collectDate_2)  #verifies the class is now date
```

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

```
unique(collectDate_2)  #displays the two dates in august where collection happened
```

```
## [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)  #displays the different plots sampled at Niwot Ridge
```

```
##  [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
```

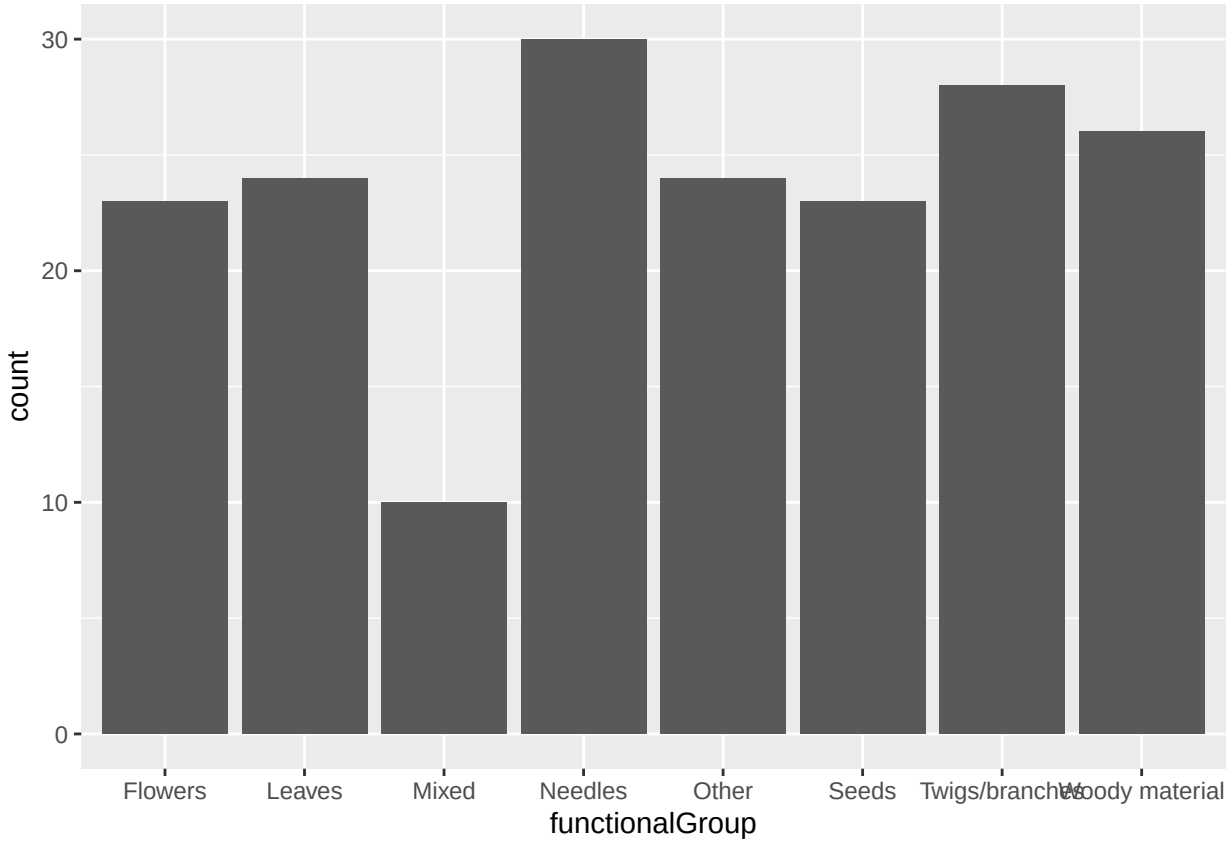
```
length(unique(Litter$plotID))  #shows the number of different plots
```

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

Answer: a summary gives you information and data on each plot, unique tells you the number of plots.

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.

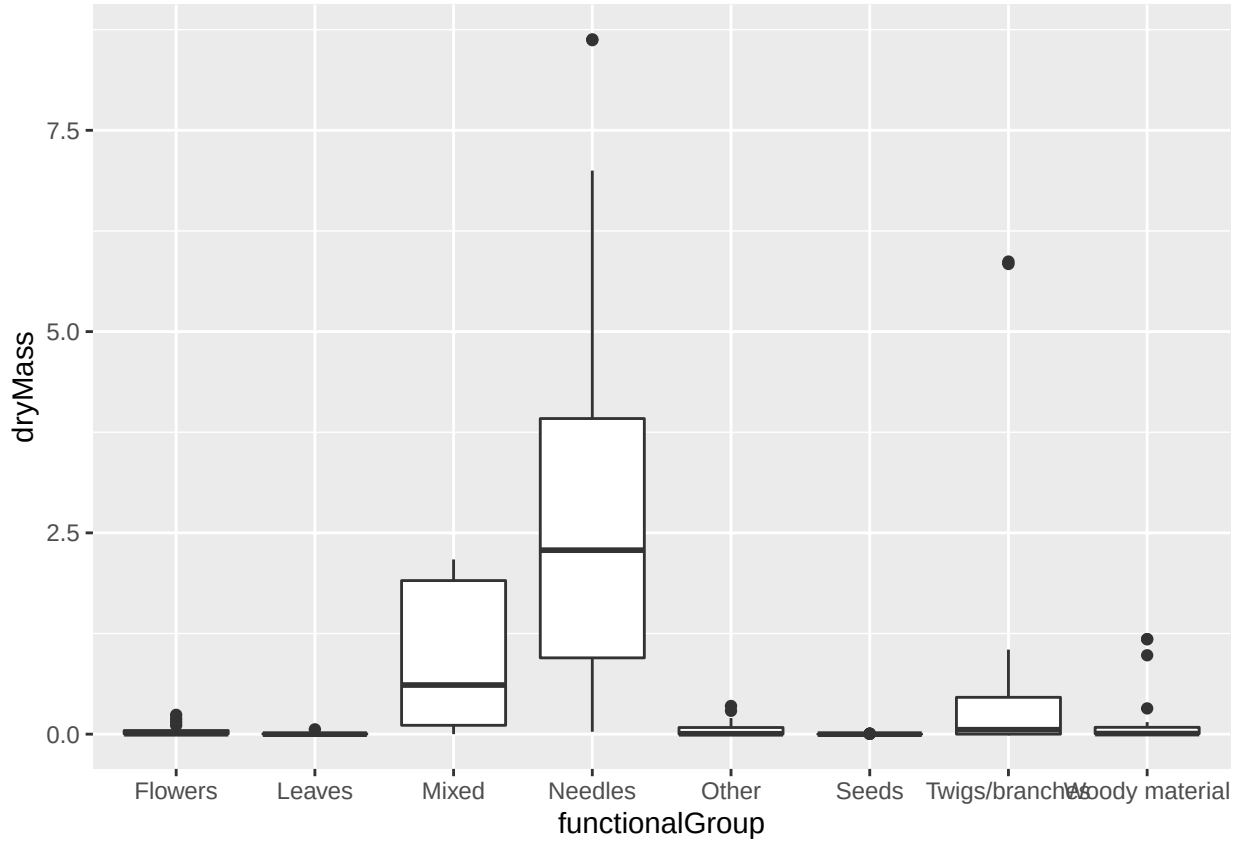
```
ggplot(Litter, aes(x = functionalGroup)) + geom_bar()
```



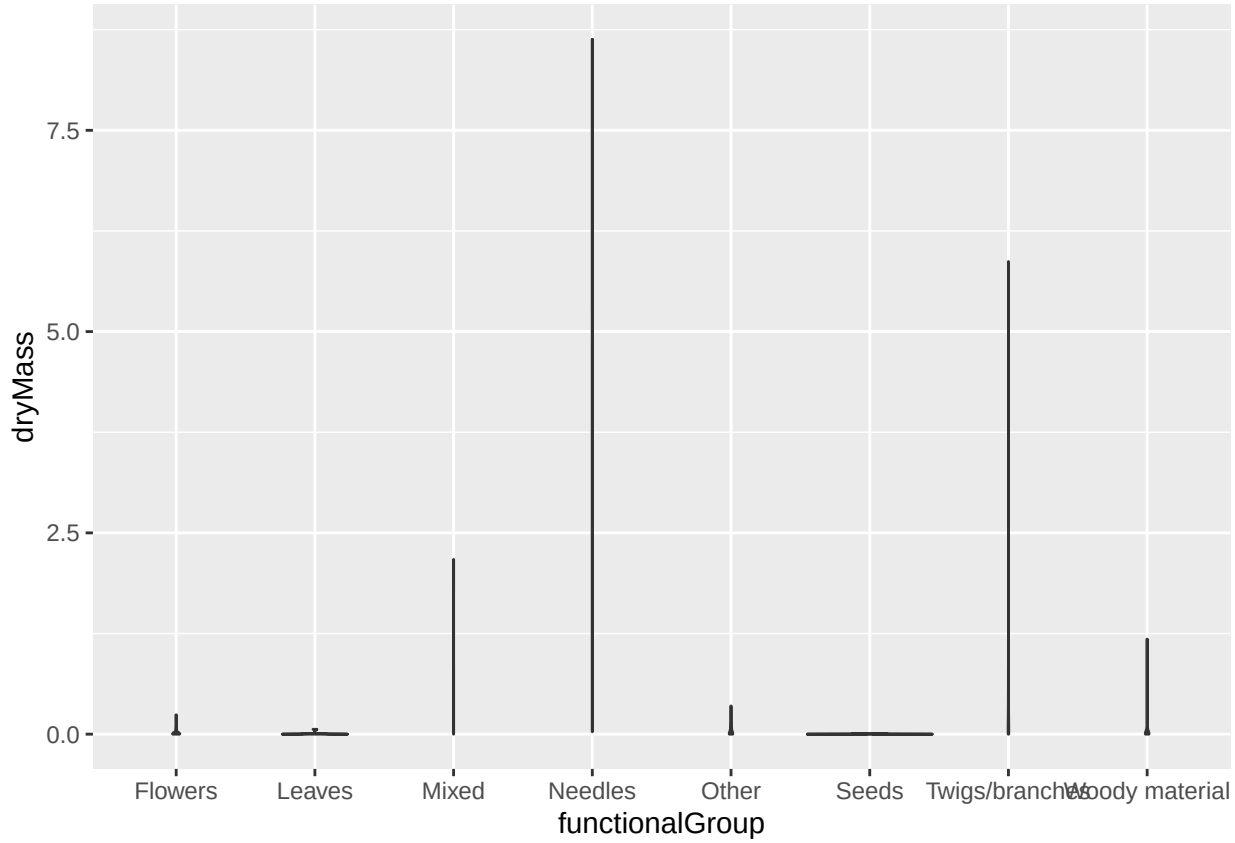

```
# creates a bar chart of the functional groups of leaf litter and woody debris.
```

15. Using `geom_boxplot` and `geom_violin`, create a boxplot and a violin plot of dryMass by functionalGroup.

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




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




```
# creates a box plot and violin plot of the functional group by the drymass.
```

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

> Answer: Because dry mass is studied in continuous variables that cannot be accurately plotted on a violin plot

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

> Answer: Needles have the highest biomass, followed by Mixed, and twigs/branches