Assignment 3: Data Exploration

Xianhang Xie

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.

```
# load necessary packages
library(tidyverse)
```

```
## Warning:   'tidyverse' R 4.1.3

## Warning:   'ggplot2' R 4.1.3

## Warning:   'tibble' R 4.1.3

## Warning:   'tidyr' R 4.1.3
```

```
## Warning:  'readr' R 4.1.3

## Warning:  'purrr' R 4.1.3

## Warning:  'dplyr' R 4.1.3

## Warning:  'stringr' R 4.1.3

## Warning:  'forcats' R 4.1.3
```

```
library(lubridate)
```

```
## Warning:  'lubridate' R 4.1.3

## Warning:  'timechange' R 4.1.3
```

```
# Check your working directory
getwd()
```

```
## [1] "C:/Users/11764/Desktop/EDA-Spring2023/Assignments"
```

```
# upload two datasets
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 a fairly new class of insecticides that are widely used in agriculture because they are systemic. This means that the plant can take up the pesticide and be protected from pests for the rest of its life. But because the pesticide affects the whole plant, it can also be found in the nectar and pollen of treated plants, which can be harmful to pollinators and other good bugs. Insects are very important for keeping ecosystems in balance, and they are also important for agriculture because they pollinate plants. The use of neonicotinoids and other insecticides can cause insect populations to drop, which can have big and long-term effects on the environment. Ecotoxicology of neonicotinoids on insects is a very important area of research that helps us learn more about how these insecticides affect non-target species and how to make their effects on the environment as small as possible.

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: Reasons why we might want to study forest litter and woody debris: (1) Nutrient cycling: In forests, litter and woody debris help the nutrient cycle by breaking down and releasing nutrients into the soil, which plants can then use. (2) Soil formation: Soil is made when dead leaves and pieces of wood pile up. This helps keep the soil's structure and stability. (3) Biodiversity: Litter and woody debris are homes for many different kinds of animals, such as insects, fungi, and small mammals. (4) Carbon sequestration: Forests are important carbon sinks, and litter and woody debris play a role in the carbon cycle by storing carbon as organic matter. (5) Forest management: Knowing what role litter and woody debris play in forest ecosystems can help guide forest management practices like harvesting and replanting and make sure they are sustainable.

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. Tower plot locations are chosen at random within the 90% flux footprint of the primary and secondary airsheds (and other areas close to the airsheds, if needed to give enough space between plots). 2. In places with forested tower airsheds, 20 40m *x* 40m plots are planned for lier sampling. In places where there is low vegetation over the tower airsheds, lier sampling is planned to take place in 4 40m * 40m tower plots and 26 20m *x* 20m plots. This is so that soil sampling can take place at the same time. 3. Depending on the plant, traps can be placed in plots in a planned or random way. In places where more than half of the trees are taller than 2 meters, the placement of lizard traps is random and uses the same list of grid cell locations that is used to collect herbaceous clippings and bryophytes (Figure 1) (AD[12], AD[13]).

Obtain basic summaries of your data (Neonics)

5. What are the dimensions of the dataset?

```
dim(Neonics)
```

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

The dimension of Neonics dataset is 4623 rows and 30 columns.

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)
```

```
##     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 is Population (1803), Mortality (1493) and Behavior (360). Because they are the most important topics for a given species.

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 
##                             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: Honey Bee (667); Parasitic Wasp (285); Buff Tailed Bumblebee (183); Carniolan Honey Bee (152); Bumble Bee (140); Italian Honeybee (113). They are all 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"
```

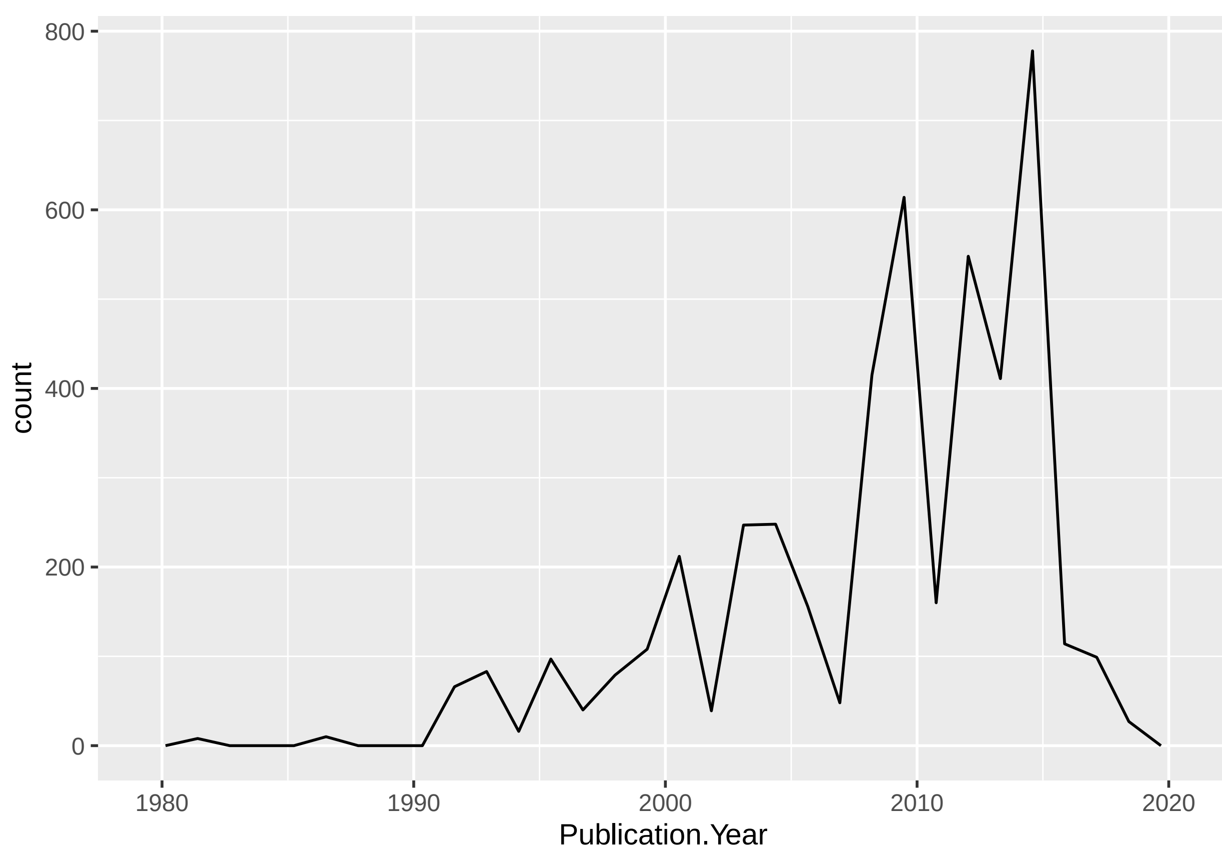
Answer: It is categorical (factor). It is not numeric because some of the concentration is less than some value.

Explore your data graphically (Neonics)

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

```
Neonics |> ggplot(aes(Publication.Year)) + geom_freqpoly()
```

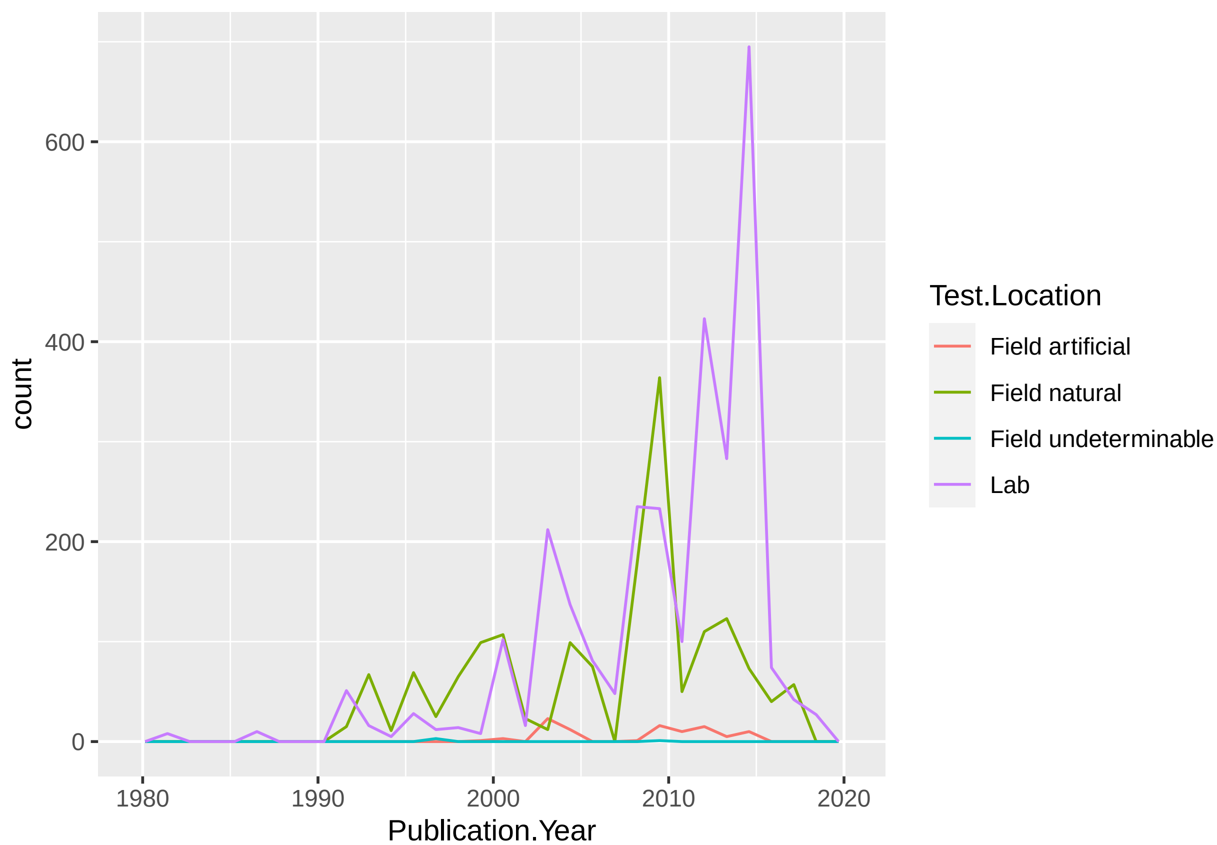
```
## `stat_bin()` using `bins = 30`. Pick better value with `binwidth`.
```



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

```
Neonics |> ggplot(aes(Publication.Year, color = Test.Location)) +
  geom_freqpoly()
```

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



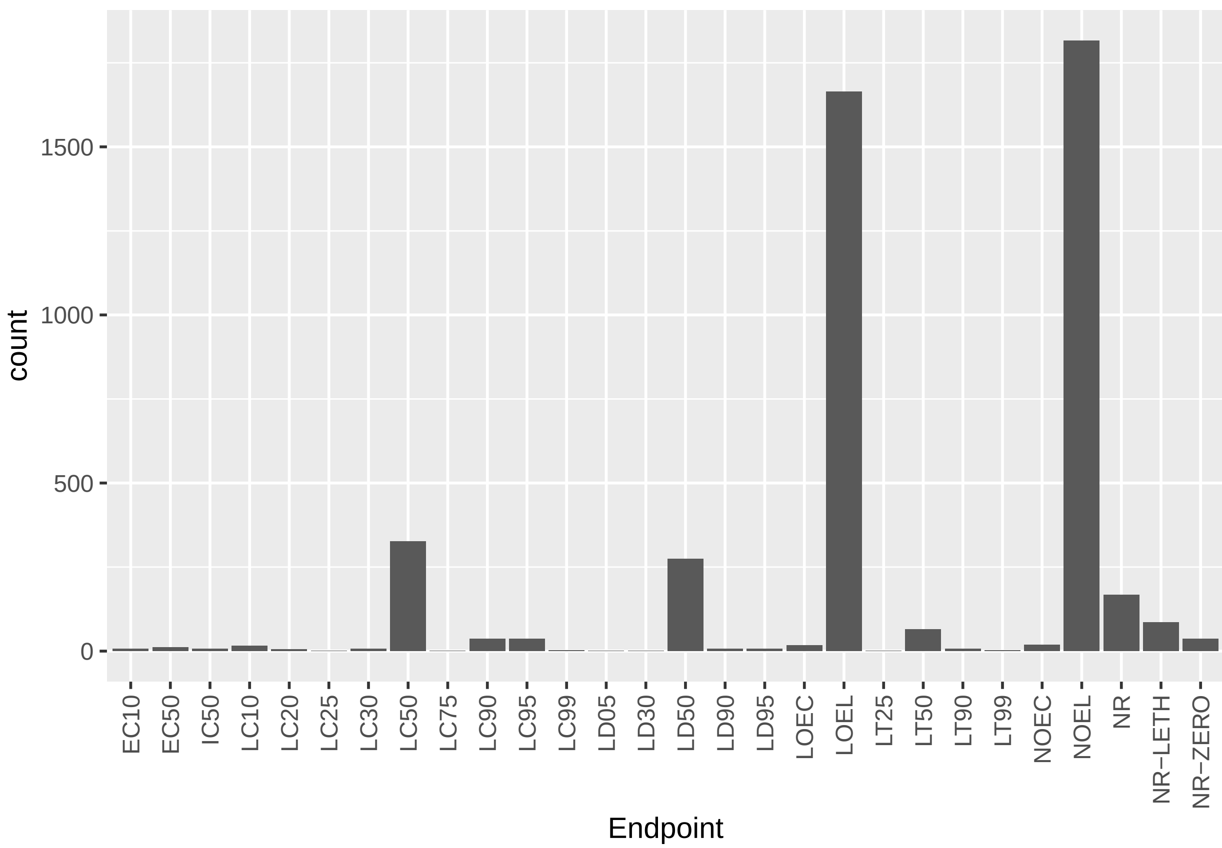
Interpret this graph. What are the most common test locations, and do they differ over time?

> Answer: In 1990-2000, it was field natural. After 2010, it is Lab.

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…]

```
Neonics |> ggplot(aes(Endpoint)) + geom_bar() +
 theme(axis.text.x = element_text(angle = 90, vjust = 0.5, hjust=1))
```



Answer: NOEL, LOEL

NOEL: No-observable-effect-level: highest dose (concentration) producing effects not significantly different from responses of controls according to author's reported statistical test

LOEL: Lowest-observable-effect-level: lowest dose (concentration) producing effects that were significantly different (as reported by authors) from responses of controls (LOEAL/LOEC)

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$collectDate = date(Litter$collectDate)
```

```
## Warning: tz(): Don't know how to compute timezone for object of class factor;
## returning "UTC".
```

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

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

It is not a date. It is a factor. “2018-08-02” “2018-08-30” was sampled.

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
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
```

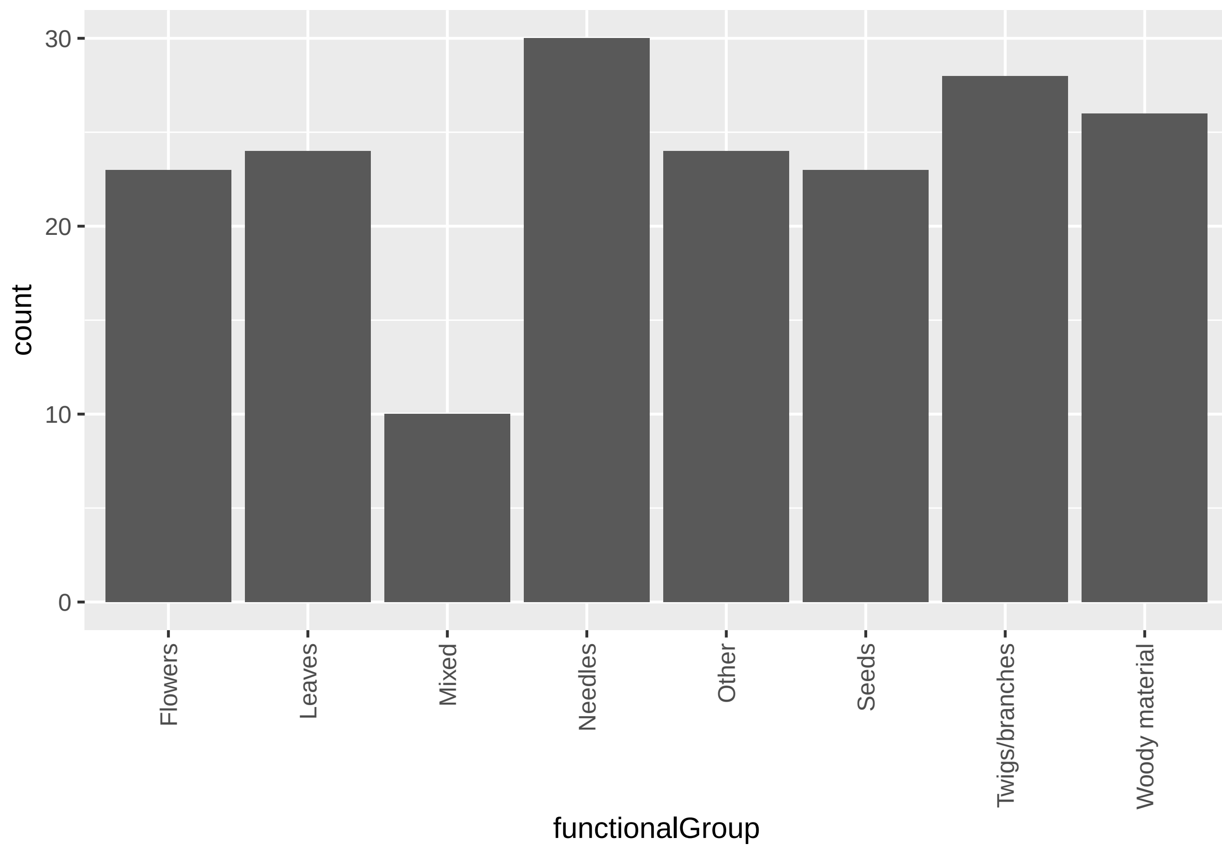
```r
summary(Litter$namedLocation)
```

```
## 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: 12 plots were sampled. The unique only tells the different unique things, and the summary tells you how many are in one category.

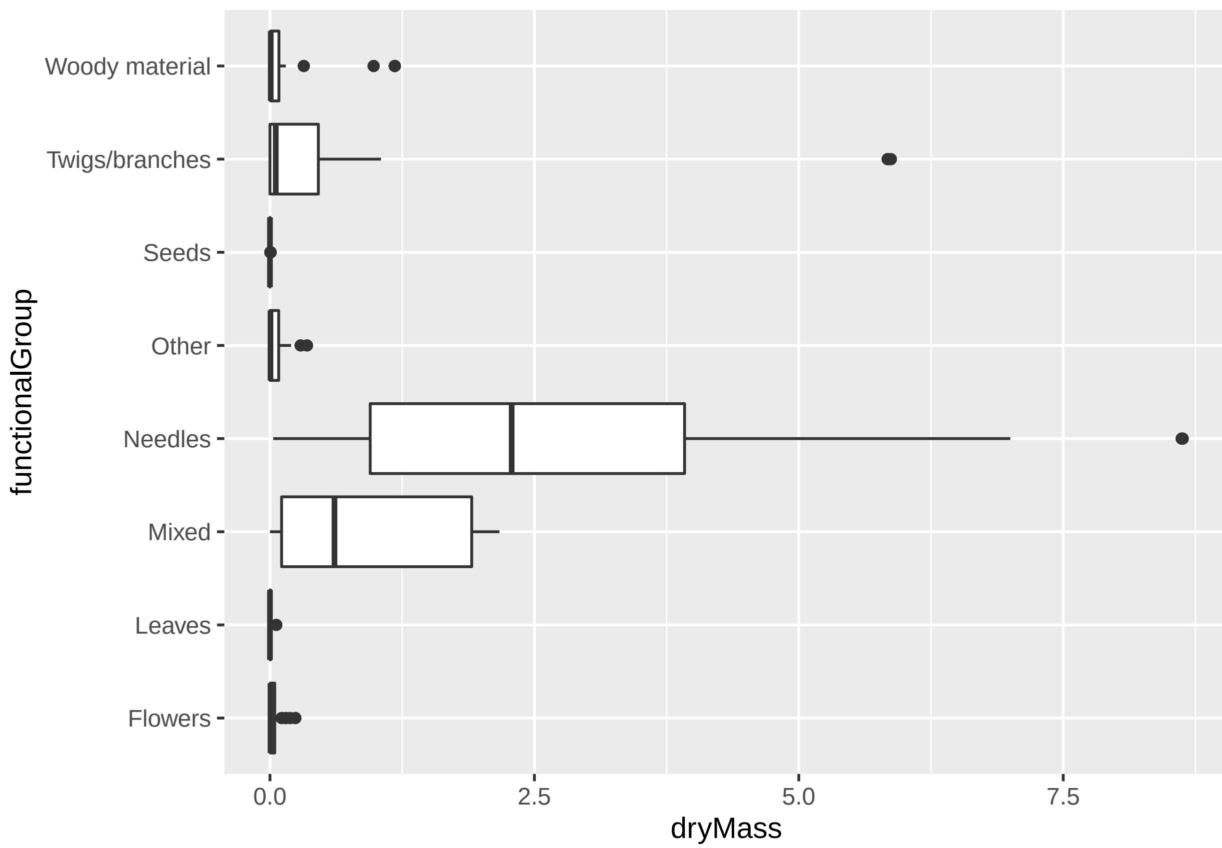
14. Create a bar graph of functionalGroup counts. This shows you what type of litter is collected at the Niwot Ridge sites. Notice that litter types are fairly equally distributed across the Niwot Ridge sites.

```r
Litter |> ggplot(aes(functionalGroup)) +
  geom_bar() +
 theme(axis.text.x = element_text(angle = 90, vjust = 0.5, hjust=1))
```

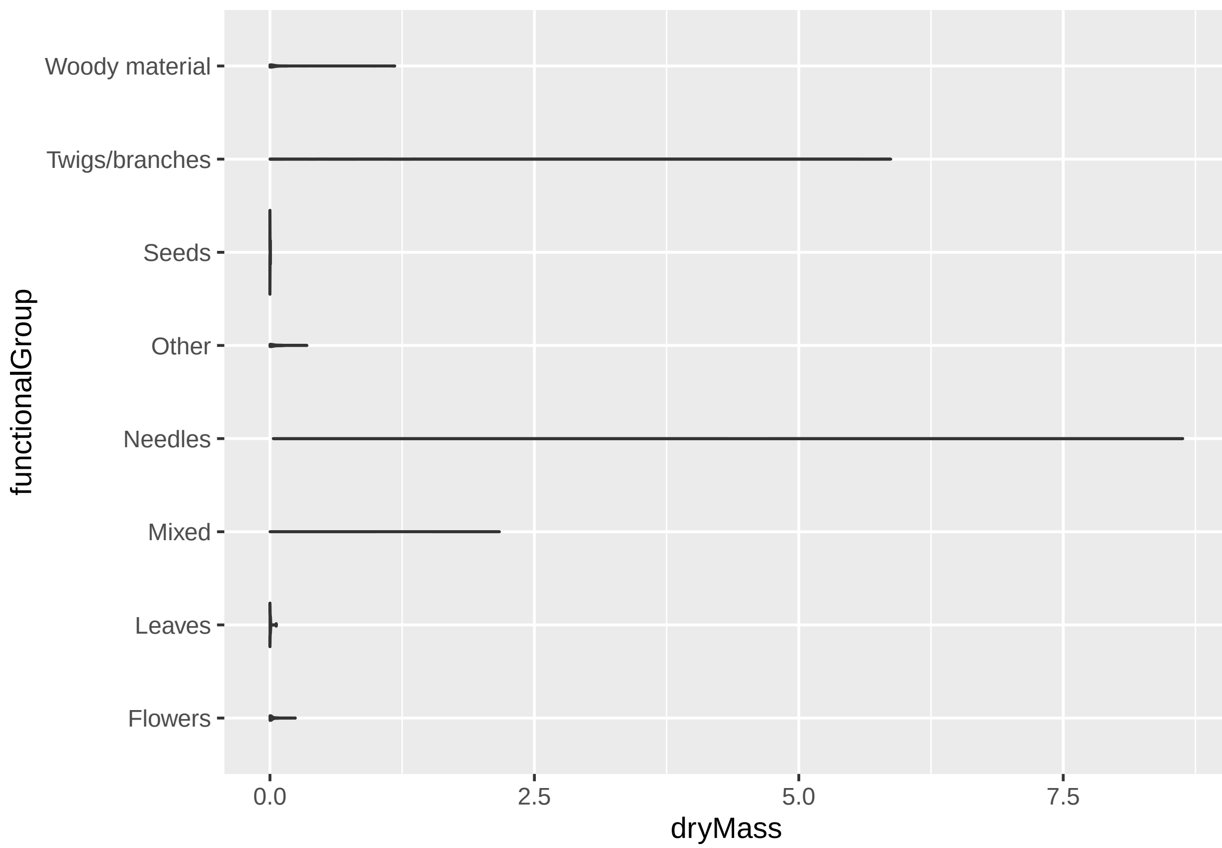


15. Using `geom_boxplot` and `geom_violin`, create a boxplot and a violin plot of dryMass by functionalGroup.

```
Litter |> ggplot(aes(dryMass, functionalGroup)) +
  geom_boxplot()
```




```
Litter |> ggplot(aes(dryMass, functionalGroup)) +
  geom_violin()
```



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

> Answer: Because the violin plot we cannot find enough information, mostly we can only see the outliers.

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

> Answer: Needles.