Assignment 3: Data Exploration

Shivani Kuckreja, Section #2

OVERVIEW

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

Directions

1. Change "Student Name, Section #" on line 3 (above) with your name and section number.
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., "FirstLast_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. **Be sure to add the `stringsAsFactors = TRUE` parameter to the function when reading in the CSV files.**

```
library(tidyverse)
```

```
## Warning in system("timedatectl", intern = TRUE): running command 'timedatectl'
## had status 1
```

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

   Answer: Neonicotinoids are insectisides that are much more toxic to insects than they are to mammals and other organisms.

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 that falls to the forest ground can help predict the health (e.g., nutrient cycling and soil fertility) of an ecosystem.

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: The litter and woody debris is measured by mass. * Litter and fine woody debris is sampled at terrestrial NEON sites that contain woody vegetation more than 2m tall * Litter and fine woody debris is only sampled in tower plots * Tower plots are placed randomly within the 90% flux footprint of the primary and secondary airsheds

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 "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: Most common effects studied include mortality and population, which would be of interest because neonicotinoids are very toxic to 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? 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: 6 most commonly studied species: Honey Bee, Parasitic Wasp, Buff Tailed Bumblebee, Carniolan Honey Bee, Bumble Bee, and the Italian Honeybee. All six of these species are part of the Hymenoptera order of insects. Neonicotinoids are especially harmful to those of the Hymenoptera order of insects (specifically, to bees). Neonicotinoids are widely to blame for the recent decline in bee populations. Neonicotinoids are often absorbed by plants. When bees collect pollen from those plants, they often die from the toxicity of the neonicotinoids.

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.)
```

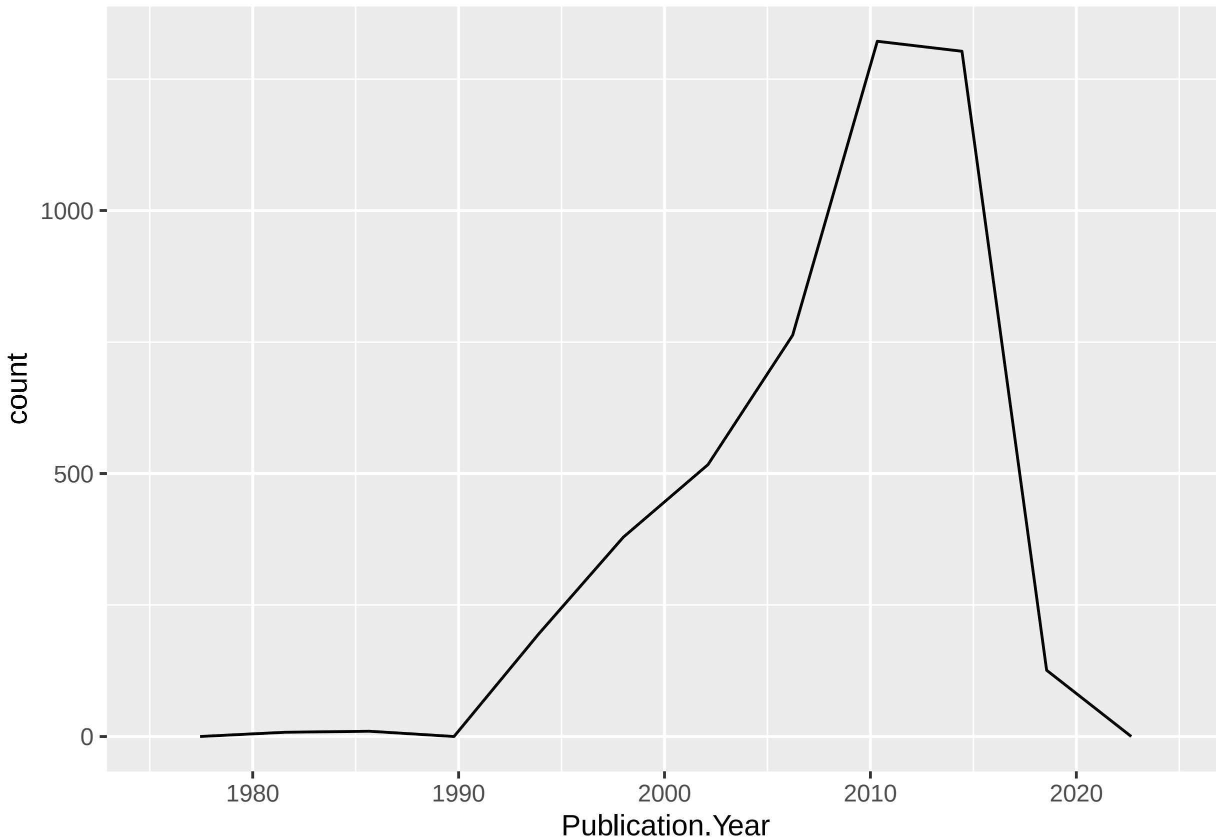
```
## [1] "factor"
```

Answer: While most in the class are numbers, many numbers are followed by slashes. In additionm, some rows do not have numbers–instead, they have “NR”s.

Explore your data graphically (Neonics)

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

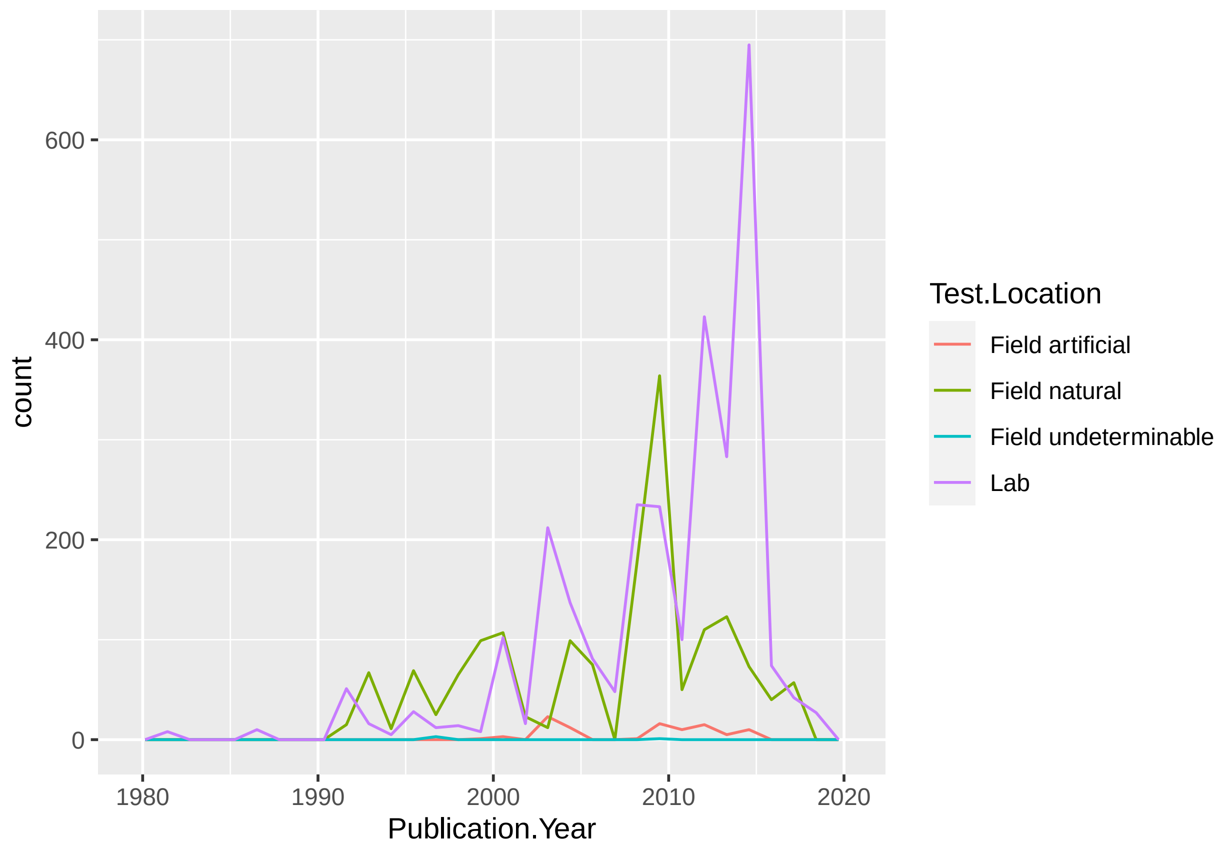
```r
ggplot(Neonics) +
  geom_freqpoly(aes(x = Publication.Year) , bins = 10)
```



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

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

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

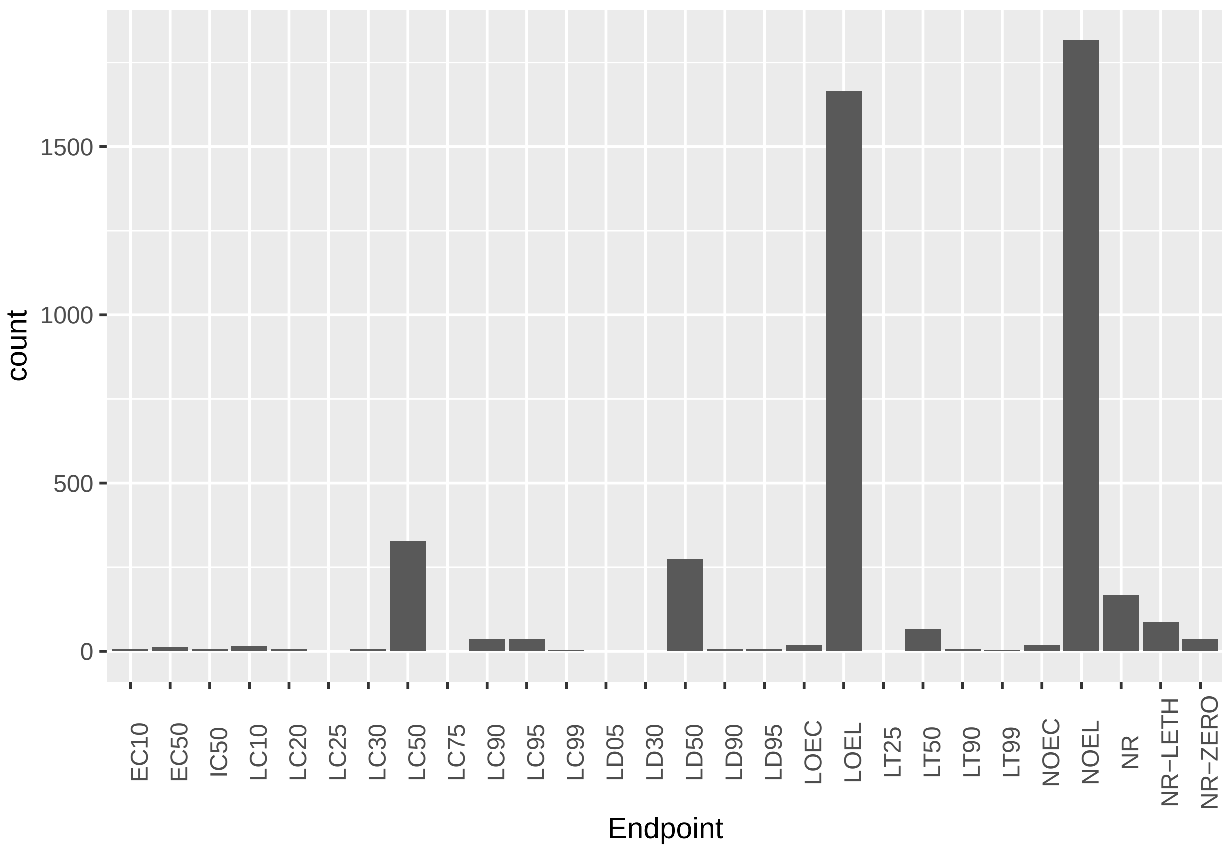


Interpret this graph. What are the most common test locations, and do they differ over time?

> Answer: Most common test locations are the lab and field natural. From about 1992 to 2000, field natural was more common than the lab. Since 2000, however, the lab has almost always been the more common test location compared to the field natural. Both test locations had little spikes over time and at least one big spike between 1990 and 2020.

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(x = Endpoint)) +
theme(axis.text.x = element_text(angle = 90)) #rotates x axis 90 degrees--learned in Office Hours!
```



Answer: The two most common endpoints are LOEL and NOEL. LOEL = Lowest-observable-effect-level: lowest dose (concentration) producing effects that were significantly different (as reported by authors) from responses of controls (LOEAL/LOEC) NOEL = No-observable-effect-level: highest dose (concentration) producing effects not significantly different from responses of controls according to author's reported statistical test (NOEAL/NOEC)

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) #It is not a date--it is a factor
```

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

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

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

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

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

Answer: The class of collectData was initially not a date–it was a factor. Litter was sampled on 2018-08-02 and 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`?

```
length(unique(Litter$plotID))
```

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

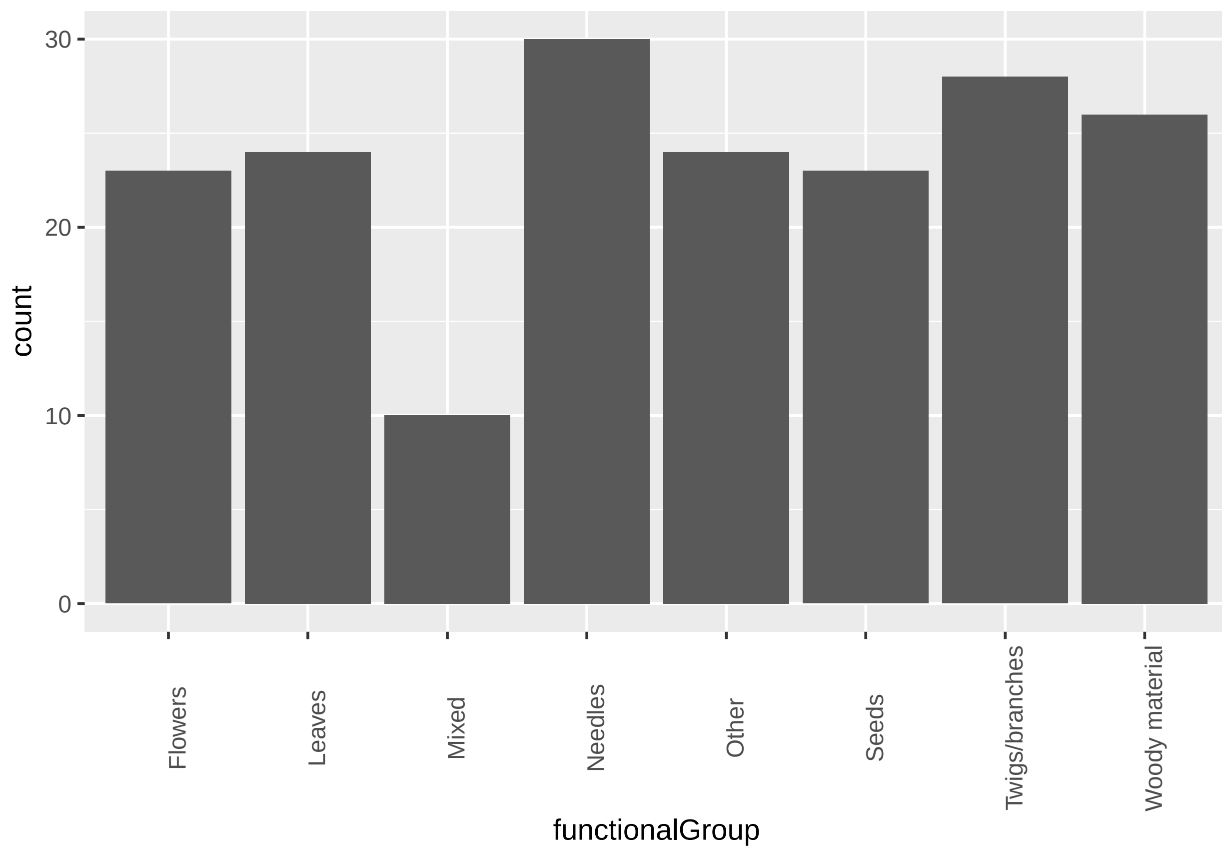
```
summary(Litter$plotID)
```

```
## 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: In Niwot Ridge, 12 different plots were sampled. "Unique" gives us the number of distinct plots sampled, whereas "summary" provides the frequency with which each plot was sampled, as well as tells us how many different plots there are (listed by name above).

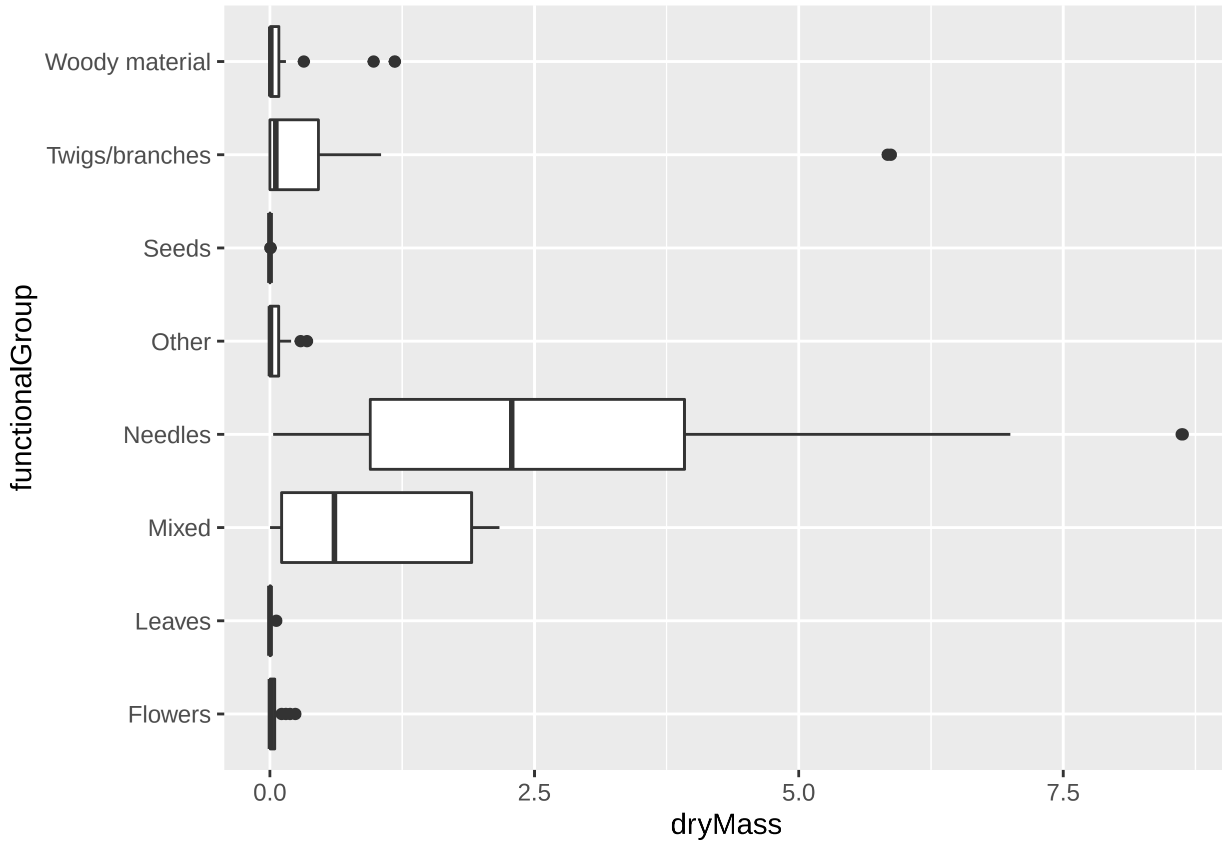
14. Create a bar graph of functionalGroup counts. This shows you what type of litter is collected at the Niwot Ridge sites. Notice that litter types are fairly equally distributed across the Niwot Ridge sites.

```
ggplot(Litter) +
  geom_bar(aes(x = functionalGroup)) +
theme(axis.text.x = element_text(angle = 90)) #rotates x axis 90 degrees--learned in Office
```

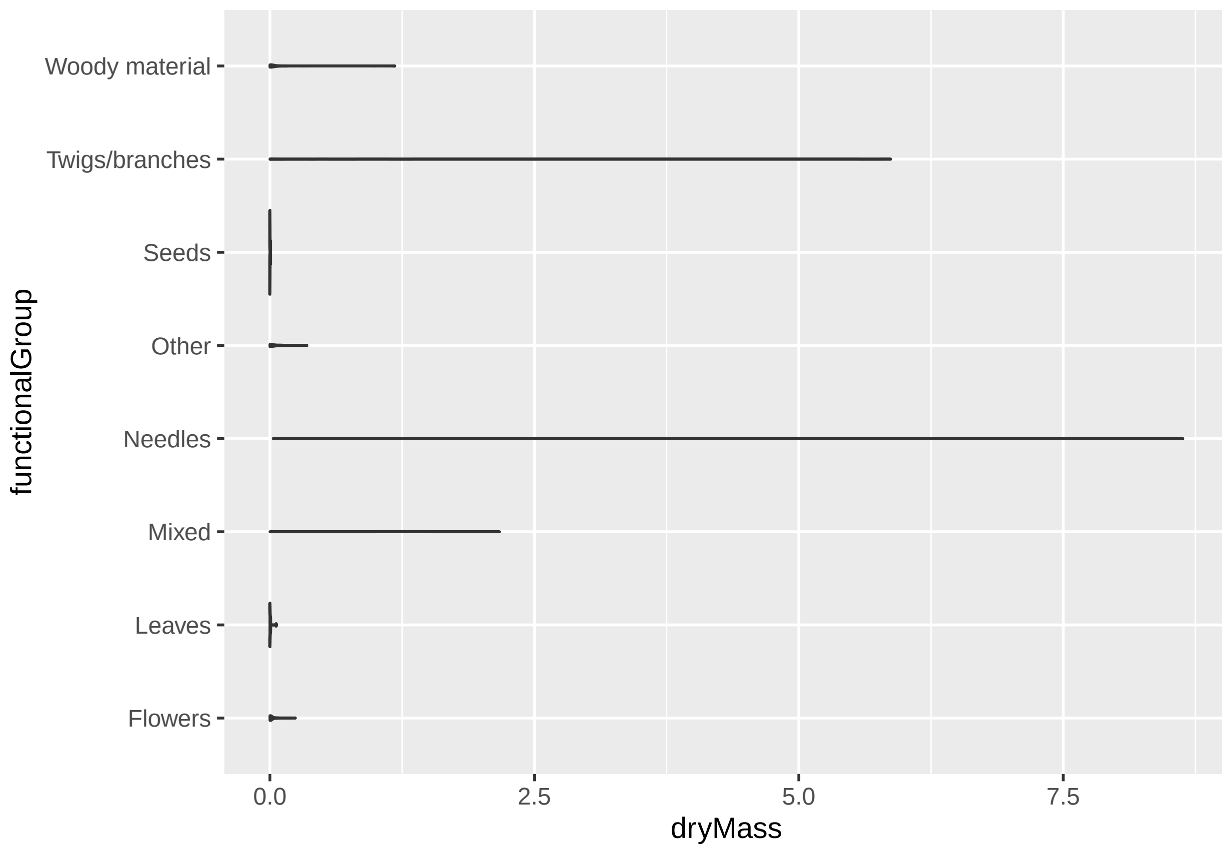


15. Using `geom_boxplot` and `geom_violin`, create a boxplot and a violin plot of dryMass by functionalGroup.

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




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



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

> Answer: Usually a violin plot is better than a boxplot because a violin plot shows the full distribution of data. In this case, however, the boxplot is more effective than the violin plot because the boxplot appears to show more of the distribution of data (including the mean, median, and outliers) whereas the violinlplot seems 1-dimensional.

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

> Answer: Needles and mixed