Assignment 3: Data Exploration

Reino Hyyppa, Section #02

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.**

```
# check wd
getwd()
```

```
## [1] "/Users/reinohyyppa/Desktop/Duke MEM/Spring 22 /ENV872/Environmental_Data_Analytics_2022/Assignme
```

```
# load packages
library(tidyverse)

# initialize data
Neonics <- read.csv("../Data/Raw/ECOTOX_Neonicotinoids_Insects_raw.csv", stringsAsFactors = TRUE)

# initialize data
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: From my understanding, harmful chemicals used in neonicotinoids are dangerous to insects, including pollinators which are important in sustaining a healthy ecosystem.

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 floor can be beneficial to health of the forest and the wildlife that live in them. In addition, wood debris is a fuel source for widlfires, and monitoring litter and woody debris can help with wildfire management.

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 are only sampled in tower plots. * Ground traps are sampled once per year. * Trap placement within plots can be either targeted or randomized, depending on the vegetation.

Obtain basic summaries of your data (Neonics)

5. What are the dimensions of the dataset?

```
# check dimension of Neonics dataset
dim(Neonics)
```

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

   Answer: Rows: 4623; Columns: 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: The effects of neonics are important to study in order to determine the predomoninant effects that neonics have on insects health and survival.

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: The six most commonly studies insects are Honey bee, parasitic wasp, buff tailed bumblebee, Carniolan Honey Bee, Bumble Bee, Italian Honeybee.All six of these species are pollinators. One reason why pollinators might be of greater interest than other insects, is the role that pollinators play in overall ecosystem health.

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

```
Neonics$Conc.1..Author.[1:6]
```

```
## [1] 27.2 19.7 47   25   13   268
## 1006 Levels: <0.0004 <0.025 <0.088 <0.5 <1.5 <10/ <2.5/ <4.00 <5.00 ... NR/
```

```
class(Neonics$Conc.1..Author.)
```

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

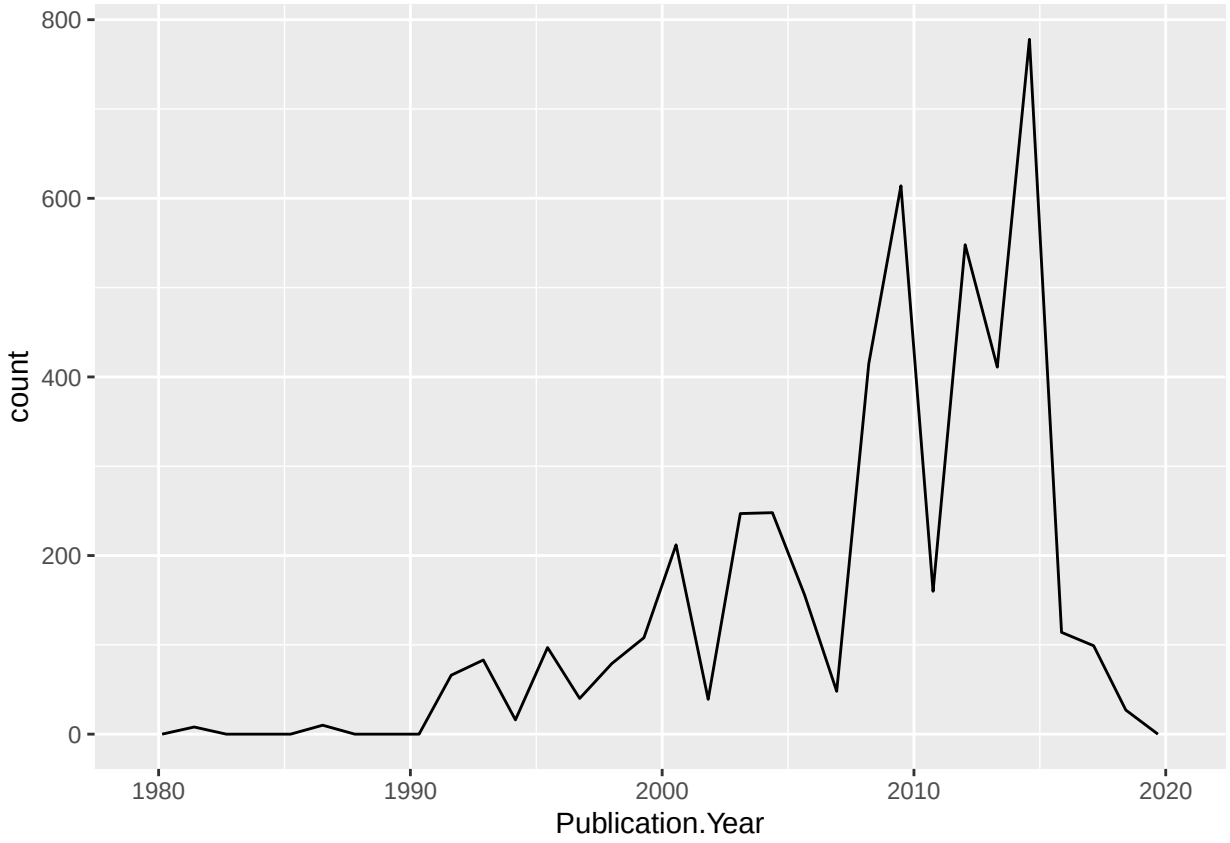
Answer: Conc.1..Author. is classified as a factor. This variable is stored as a factor because it seems like there's a "/" stored in a lot of the concentration values. Because of this, R is intepretating concentrations as a factor instead of a numeric value.

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()
```

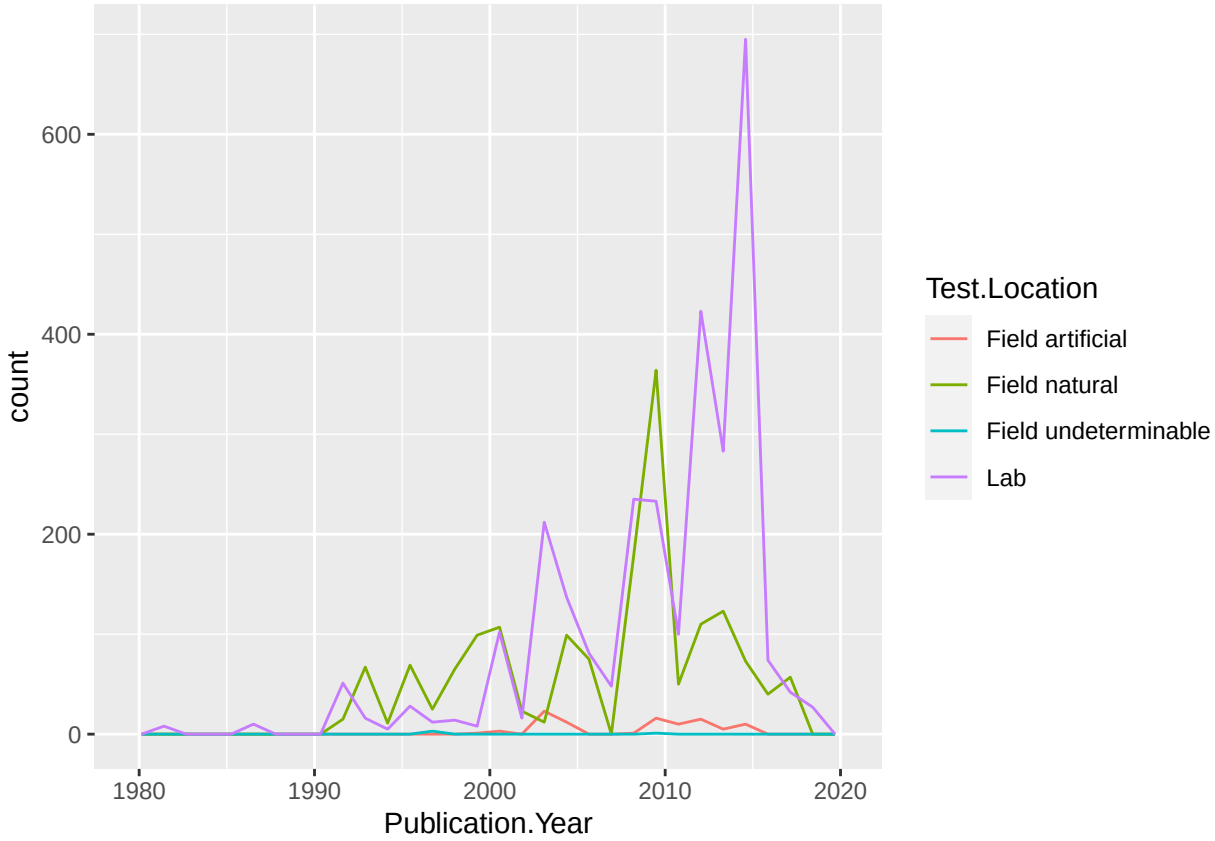
```
## `stat_bin()` using `bins = 30`. Pick better value with `binwidth`.
```



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

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

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

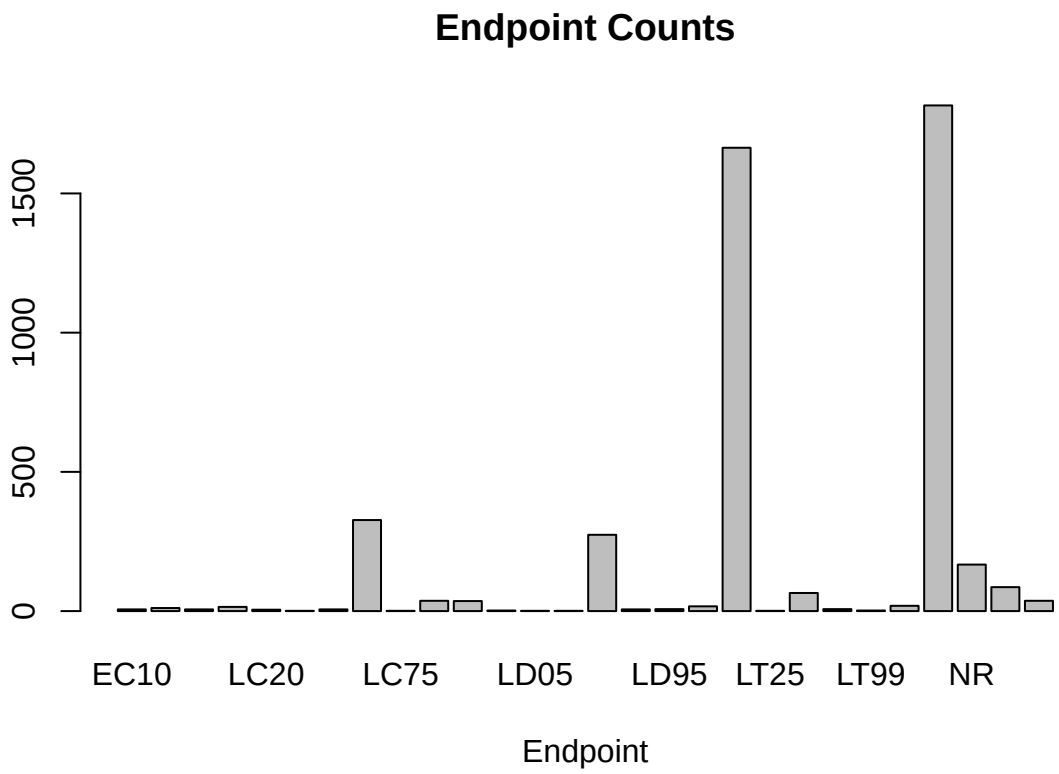


Interpret this graph. What are the most common test locations, and do they differ over time?

Answer:

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.

```
endpoints <- table(Neonics$Endpoint)

barplot(endpoints, main = "Endpoint Counts",
        xlab = "Endpoint")
```



Answer: Based on the barchart plot, the two most common endpoints are NOEl and LOEL. They are defined as no-observable-effect level and lowest-observable-effect-level, respectively.

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.

```
# check class
class(Litter$collectDate)
```

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

```
# set date
Litter$collectDate <- as.Date(Litter$collectDate)

# use unique function to determine
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`?

```
df_plotID <- unique(Litter$plotID)
length(df_plotID)
```

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

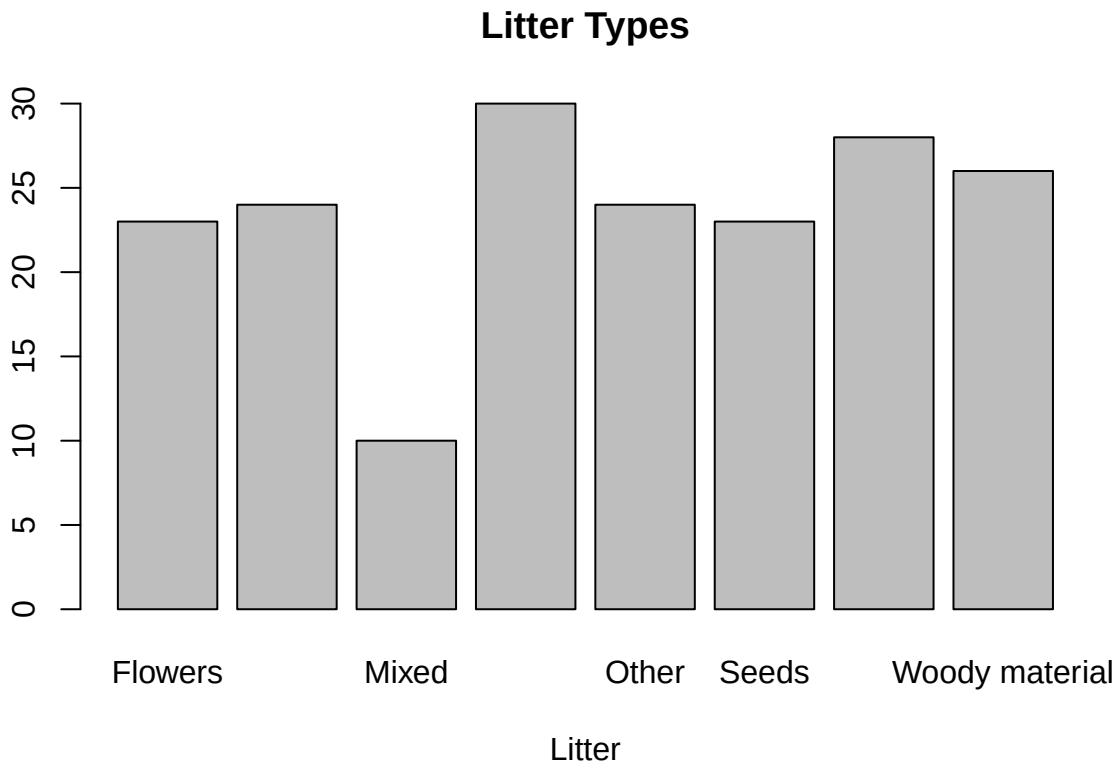
```
summary(df_plotID)
```

```
## NIWO_040 NIWO_041 NIWO_046 NIWO_047 NIWO_051 NIWO_057 NIWO_058 NIWO_061
##        1        1        1        1        1        1        1        1
## NIWO_062 NIWO_063 NIWO_064 NIWO_067
##        1        1        1        1
```

Answer: The unique function produces individual results for each plot samples at Niwot Ridge, whereas, the summary function summarizes all of the plots into a single count.

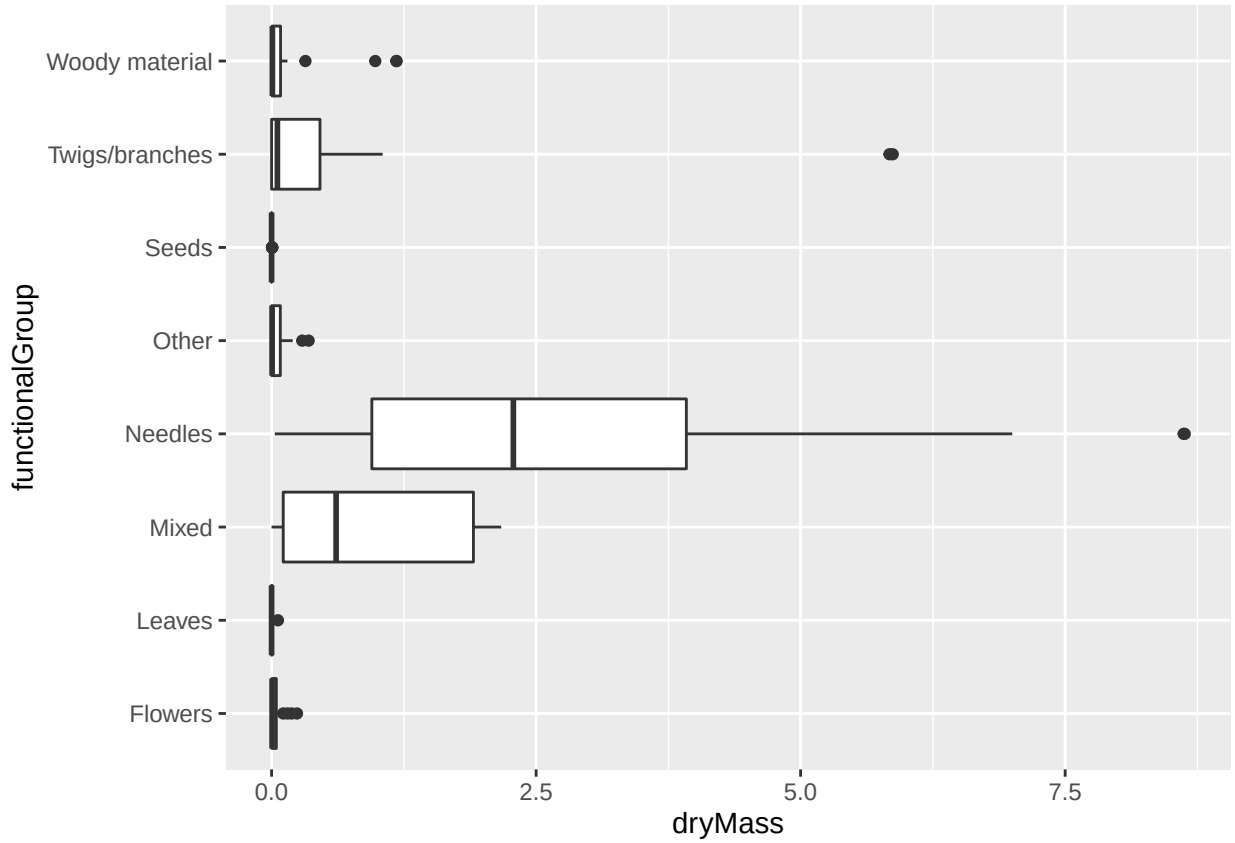
14. Create a bar graph of functionalGroup counts. This shows you what type of litter is collected at the Niwot Ridge sites. Notice that litter types are fairly equally distributed across the Niwot Ridge sites.

```
LitterType <- table(Litter$functionalGroup)

barplot(LitterType, main = "Litter Types",
        xlab = "Litter")
```

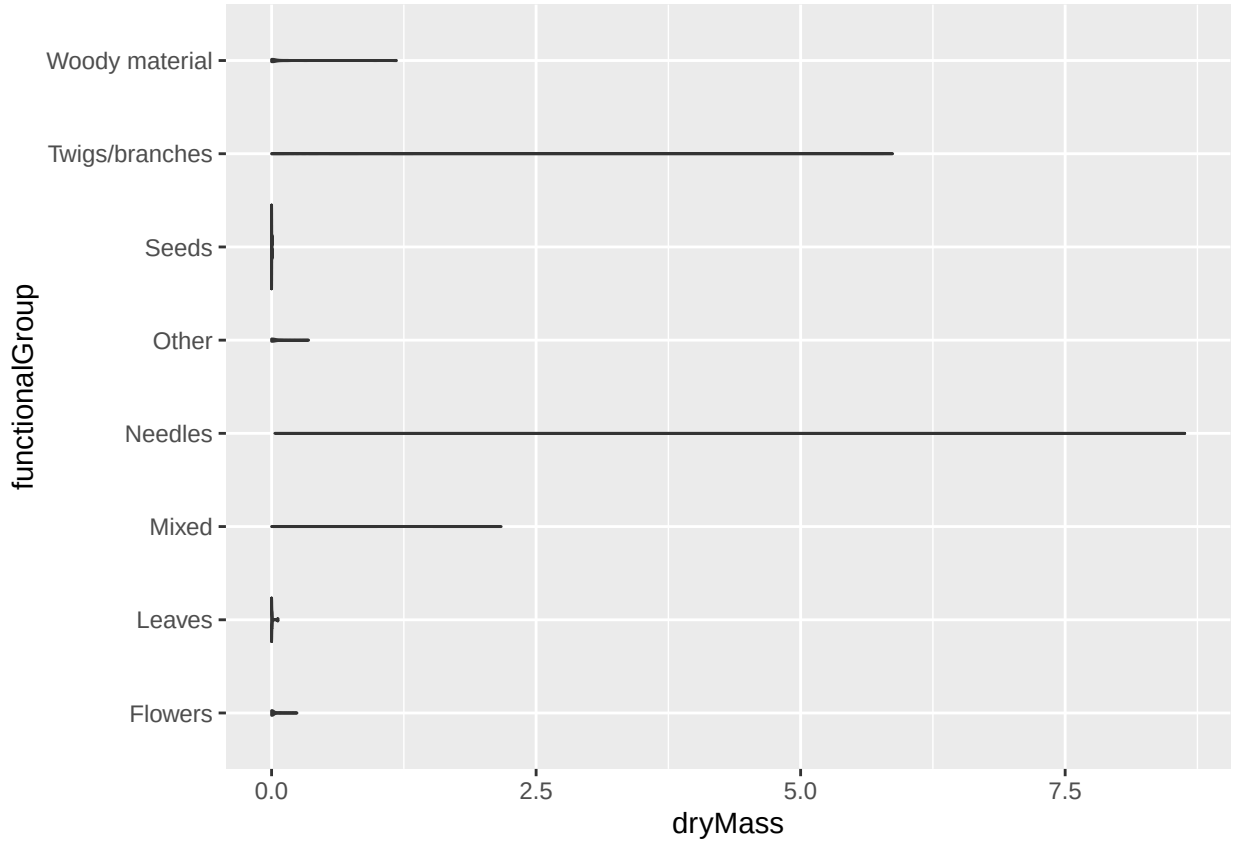


15. Using `geom_boxplot` and `geom_violin`, create a boxplot and a violin plot of dryMass by functionalGroup.

```
x1 <- Litter$dryMass
x2 <- Litter$functionalGroup

# create a boxplot
ggplot(Litter, aes(x=dryMass, y=functionalGroup)) +
  geom_boxplot()
```




```
# create a violin plot
ggplot(Litter, aes(x=dryMass, y=functionalGroup)) +
  geom_violin()
```



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

> Answer: A boxplot is a more effective visualization tool for several reasons. First, off a boxplot shows the full distribution of the data and, therefore, outliers are clearly identifiable. Additionally, a boxplot displays summary statistics including median and the interquartile range which can be useful to analyze your data.

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

> Answer: Needles and twigs/branches tend to have the highest biomas as these sights.