Assignment 3: Data Exploration

Grace Randall

Fall 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.

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

```
## [1] "/home/guest/module1/EDE_Fall2023"
```

```
#to install necessary packages if not already installed
#install.packages("tidyverse")
#install.packages("lubridate")

#load libraries
library("tidyverse")
library("lubridate")

#load data
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: This information could be used to assess the chemical's effectiveness as an insecticide. It could also be used to see what species of insect it may harm when released into the environment.

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:The litter that falls to the ground could be very improtant to measure for environmental modeling. Specifically litter plays a key role in nutrient cycling and carbon sequestration because it allows nutrients and carbon to be transfered from plant life to the soil.

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. ground traps are sampled once per year 2. litter is collected in elevated 0.5m2 traps and ground traps of 3m by 0.5m 3. one ground trap and one elevated trap was deployed for every 400m2 area

Obtain basic summaries of your data (Neonics)

5. What are the dimensions of the dataset?

```
dim(Neonics) #get dimensions of Neonics
```

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

   Answer: The neonics dataset has 30 columns and 4623 rows

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$Effect) # sumarize Neonics$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 that are studied are population and mortality. This might be specifically of interest because it is so relevant to environmental impact because if an insecticide causes changes in population it can cause a species to be threatened.

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. . .]

```
sort(summary(Neonics$Species.Common.Name)) #sumarize the species data and sort it by number of apperanc
```

```
##                         Ant Family                         Apple Maggot
##                                  9                                    9
##              Glasshouse Potato Wasp                            Lacewing
##                                 10                                   10
##             Southern House Mosquito              Two Spotted Lady Beetle
##                                 10                                   10
##            Spotless Ladybird Beetle                  Braconid Parasitoid
##                                 11                                   12
##                       Common Thrip         Eastern Subterranean Termite
##                                 12                                   12
##                             Jassid                           Mite Order
##                                 12                                   12
##                          Pea Aphid                    Pond Wolf Spider
##                                 12                                   12
##              Armoured Scale Family                    Diamondback Moth
##                                 13                                   13
##                       Eulophid Wasp                   Monarch Butterfly
##                                 13                                   13
##                       Predatory Bug                Yellow Fever Mosquito
##                                 13                                   13
##                        Corn Earworm                   Green Peach Aphid
##                                 14                                   14
##                           House Fly                           Ox Beetle
##                                 14                                   14
##                  Red Scale Parasite                  Spined Soldier Bug
##                                 14                                   14
##              Western Flower Thrips Hemlock Woolly Adelgid Lady Beetle
##                                 15                                   16
##              Hemlock Wooly Adelgid                                Mite
##                                 16                                   16
##                         Onion Thrip               Araneoid Spider Order
##                                 16                                   17
##                           Bee Order                      Egg Parasitoid
##                                 17                                   17
##                       Insect Class            Moth And Butterfly Order
##                                 17                                   17
##       Oystershell Scale Parasitoid           Black-spotted Lady Beetle
##                                 17                                   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
##                       Codling Moth          Flatheaded Appletree Borer
##                                 19                                   20
```

```
##              Horned Oak Gall Wasp                Leaf Beetle Family
##                                20                                20
##                 Potato Leafhopper         Tooth-necked Fungus Beetle
##                                20                                20
##                     Argentine Ant                            Beetle
##                                21                                21
##                         Mason Bee                           Mosquito
##                                22                                22
##                  Citrus Leafminer                  Ladybird Beetle
##                                23                                23
##                 Spider/Mite Class                Tobacco Flea Beetle
##                                24                                24
##                       Chalcid Wasp             Convergent Lady Beetle
##                                25                                25
##                     Stingless Bee               Ground Beetle Family
##                                25                                27
##                Rove Beetle Family                     Tobacco Aphid
##                                27                                27
##                     Scarab Beetle                     Spring Tiphia
##                                29                                29
##                      Thrip Order             Ladybird Beetle Family
##                                29                                30
##                       Parasitoid                     Braconid Wasp
##                                30                                33
##                     Cotton Aphid                   Predatory Mite
##                                33                                33
##                Sweetpotato Whitefly                      Aphid Family
##                                37                                38
##                     Cabbage Looper              Buff-tailed Bumblebee
##                                38                                39
##                     True Bug Order            Sevenspotted Lady Beetle
##                                45                                46
##                       Beetle Order         Snout Beetle Family, Weevil
##                                47                                47
##                Erythrina Gall Wasp                  Parasitoid Wasp
##                                49                                51
##              Colorado Potato Beetle                    Parastic Wasp
##                                57                                58
##                Asian Citrus Psyllid                Minute Pirate Bug
##                                60                                62
##                   European Dark Bee                         Wireworm
##                                66                                69
##                     Euonymus Scale                Asian Lady Beetle
##                                75                                76
##                   Japanese Beetle                  Italian Honeybee
##                                94                               113
##                        Bumble Bee               Carniolan Honey Bee
##                               140                               152
##              Buff Tailed Bumblebee                  Parasitic Wasp
##                               183                               285
##                         Honey Bee                           (Other)
##                               667                               670
```

Answer: the top 6 most commonly studdied species are the western honey bee (Apis mellifera),

the buff tailed bumblebee (Bombus terrestris), the parasitic wasp, the carniolan Honey Bee (Apis mellifera ssp. carnica), the common eastern bumble bee (Bombus impatiens), and the italian honey bee (Apis mellifera ssp. ligustica). All of these species are kinds of bees except for the parasitic wasp which is closely related. These species are likely most studied because bees are important polinators and bees have many populations that are under threat.

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.) #check class of `Conc.1..Author.`
```

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

```
sort(summary(Neonics$Conc.1..Author.)) #take a look at what options there are for the input of this fie
```

```
##    0.02/   0.025/     0.29    37.5/       4/        5   0.047/    0.25/
##       11       11       11       11       11       11       12       12
##    0.28/    1.28/    1.81/      112      150     2.5/       25      60/
##       12       12       12       12       12       12       12       12
##      75/    0.17/      125       14       16       17   0.0001  0.0004/
##       12       13       13       13       13       13       14       14
##   0.084/     0.15      0.6    12.5/   144.0/     350/    40.0/      48/
##       14       14       14       14       14       14       14       14
##       56      84/    0.053     0.24     0.28     125/        9 0.00355/
##       14       14       15       15       15       15       15       16
##      0.1      0.4     150/      300      80/    0.18/     0.3/     1000
##       16       16       16       16       16       17       17       17
##       40    0.005     0.4/     0.05      1.5    2.60/    20.0/        6
##       17       18       18       20       20       20       20       20
##    6.80/    62.5/    0.336     1.5/    0.01/    1000/       3/     100/
##       20       20       21       21       22       22       22       23
##        3    0.56/     0.2/    0.027      2.4      12/    25.0/   0.048/
##       23       24       25       26       26       27       28       30
##    0.15/       1/       48      0.5    0.125     500/       50    0.45/
##       30       30       30       32       33       33       36       40
##     1.0/    2.27/     0.1/    0.05/     0.45     0.03     0.5/      50/
##       40       40       42       43       43       44       45       51
##      100   0.053/       10       2/    0.40/     1023        1       NR
##       56       59       62       63       69       80       82       94
##      NR/      10/    0.37/  (Other)
##      108      127      208     1817
```

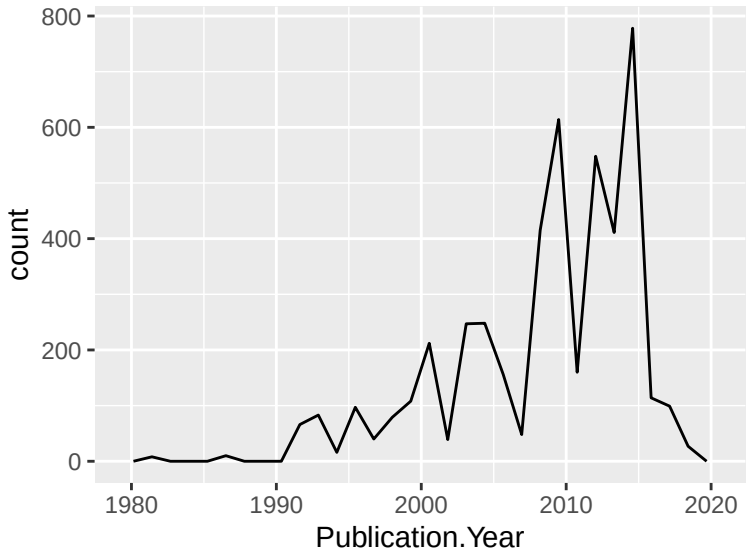
Answer: The class of `Conc.1..Author.` is factor. This is due to there being a few entries that are not numeric including the NR and also the / that is at the end of the numbered entries that prevents it from being read as a number. It is instead read as a string which is taken in as a factor.

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() #plot time series of number of publication
```

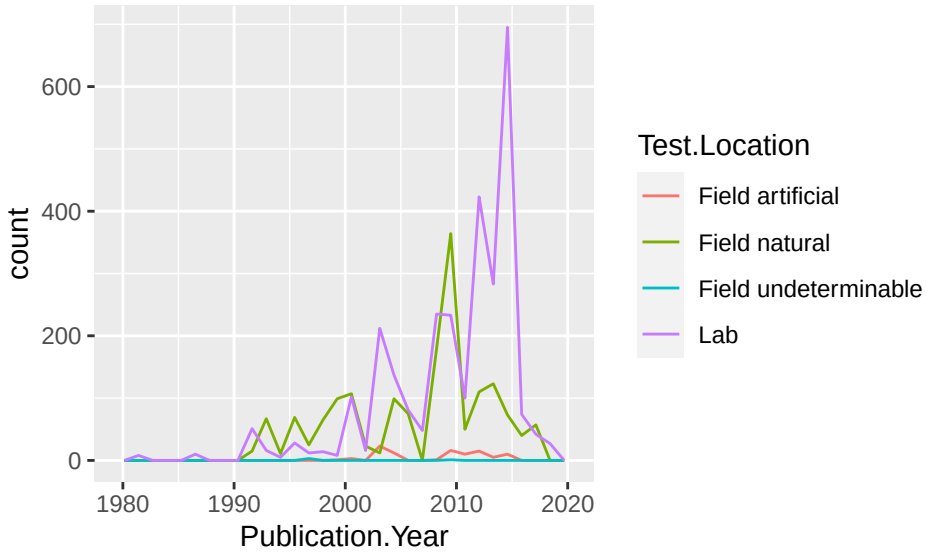
```
## `stat_bin()` using `bins = 30`. Pick better value with `binwidth`.
```



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

```
ggplot(Neonics)+
  geom_freqpoly(aes(x = Publication.Year, color = Test.Location)) #plot time series of number of public
```

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



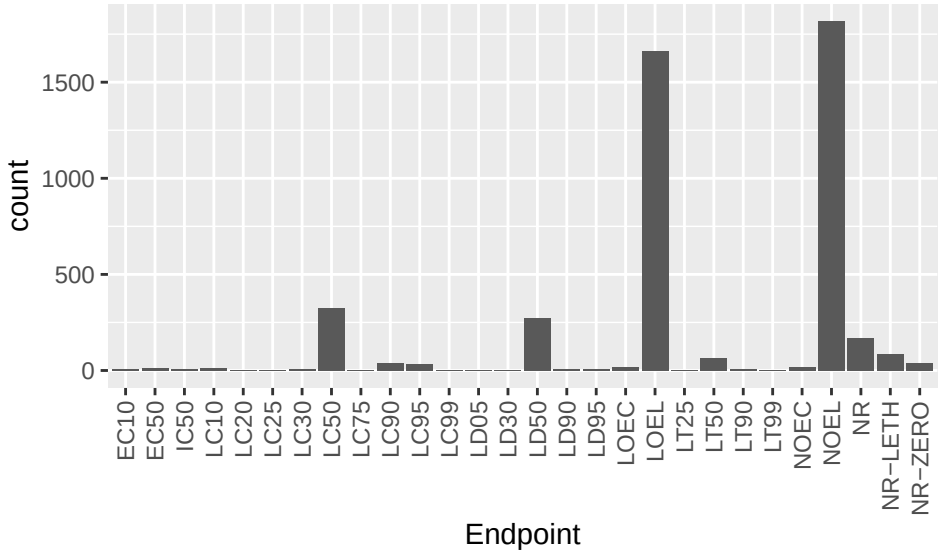
Interpret this graph. What are the most common test locations, and do they differ over time?

> Answer: The most common test locations are natural fields and labs. Field tests reached their peak popularity shortly before 2010. Lab tests Reached their peak popularity in 2015.

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...]

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



Answer: the two most common endpoints are LOEC (Lowest observable effect concentration) and NOCE (No observable effect concentration)

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) #check class of collect date
```

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

```
Litter$collectDate <- as.Date(Litter$collectDate,format = "%Y-%m-%d") #update class to be date instead
Litter$collectDate # printo out new collect date data
```

```
##   [1] "2018-08-02" "2018-08-02" "2018-08-02" "2018-08-02" "2018-08-02"
##   [6] "2018-08-02" "2018-08-02" "2018-08-02" "2018-08-02" "2018-08-02"
##  [11] "2018-08-02" "2018-08-02" "2018-08-02" "2018-08-02" "2018-08-02"
##  [16] "2018-08-02" "2018-08-02" "2018-08-02" "2018-08-02" "2018-08-02"
##  [21] "2018-08-02" "2018-08-02" "2018-08-02" "2018-08-02" "2018-08-02"
##  [26] "2018-08-02" "2018-08-02" "2018-08-02" "2018-08-02" "2018-08-02"
##  [31] "2018-08-02" "2018-08-02" "2018-08-02" "2018-08-02" "2018-08-02"
##  [36] "2018-08-02" "2018-08-02" "2018-08-02" "2018-08-02" "2018-08-02"
##  [41] "2018-08-02" "2018-08-02" "2018-08-02" "2018-08-02" "2018-08-02"
##  [46] "2018-08-02" "2018-08-02" "2018-08-02" "2018-08-02" "2018-08-02"
##  [51] "2018-08-02" "2018-08-02" "2018-08-02" "2018-08-02" "2018-08-02"
##  [56] "2018-08-02" "2018-08-02" "2018-08-02" "2018-08-02" "2018-08-02"
##  [61] "2018-08-02" "2018-08-02" "2018-08-02" "2018-08-02" "2018-08-02"
##  [66] "2018-08-02" "2018-08-02" "2018-08-02" "2018-08-02" "2018-08-02"
##  [71] "2018-08-02" "2018-08-02" "2018-08-02" "2018-08-02" "2018-08-02"
##  [76] "2018-08-02" "2018-08-02" "2018-08-02" "2018-08-02" "2018-08-02"
##  [81] "2018-08-02" "2018-08-02" "2018-08-02" "2018-08-02" "2018-08-02"
##  [86] "2018-08-02" "2018-08-02" "2018-08-02" "2018-08-02" "2018-08-02"
##  [91] "2018-08-02" "2018-08-30" "2018-08-30" "2018-08-30" "2018-08-30"
```

```
##  [96] "2018-08-30" "2018-08-30" "2018-08-30" "2018-08-30" "2018-08-30"
## [101] "2018-08-30" "2018-08-30" "2018-08-30" "2018-08-30" "2018-08-30"
## [106] "2018-08-30" "2018-08-30" "2018-08-30" "2018-08-30" "2018-08-30"
## [111] "2018-08-30" "2018-08-30" "2018-08-30" "2018-08-30" "2018-08-30"
## [116] "2018-08-30" "2018-08-30" "2018-08-30" "2018-08-30" "2018-08-30"
## [121] "2018-08-30" "2018-08-30" "2018-08-30" "2018-08-30" "2018-08-30"
## [126] "2018-08-30" "2018-08-30" "2018-08-30" "2018-08-30" "2018-08-30"
## [131] "2018-08-30" "2018-08-30" "2018-08-30" "2018-08-30" "2018-08-30"
## [136] "2018-08-30" "2018-08-30" "2018-08-30" "2018-08-30" "2018-08-30"
## [141] "2018-08-30" "2018-08-30" "2018-08-30" "2018-08-30" "2018-08-30"
## [146] "2018-08-30" "2018-08-30" "2018-08-30" "2018-08-30" "2018-08-30"
## [151] "2018-08-30" "2018-08-30" "2018-08-30" "2018-08-30" "2018-08-30"
## [156] "2018-08-30" "2018-08-30" "2018-08-30" "2018-08-30" "2018-08-30"
## [161] "2018-08-30" "2018-08-30" "2018-08-30" "2018-08-30" "2018-08-30"
## [166] "2018-08-30" "2018-08-30" "2018-08-30" "2018-08-30" "2018-08-30"
## [171] "2018-08-30" "2018-08-30" "2018-08-30" "2018-08-30" "2018-08-30"
## [176] "2018-08-30" "2018-08-30" "2018-08-30" "2018-08-30" "2018-08-30"
## [181] "2018-08-30" "2018-08-30" "2018-08-30" "2018-08-30" "2018-08-30"
## [186] "2018-08-30" "2018-08-30" "2018-08-30"
```

```
class(Litter$collectDate) # check that class has been updated
```

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

```
unique(Litter$collectDate) # find unique collection dates
```

```
## [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`?

```
summary(Litter$namedLocation) #summarize locations
```

```
## 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
```

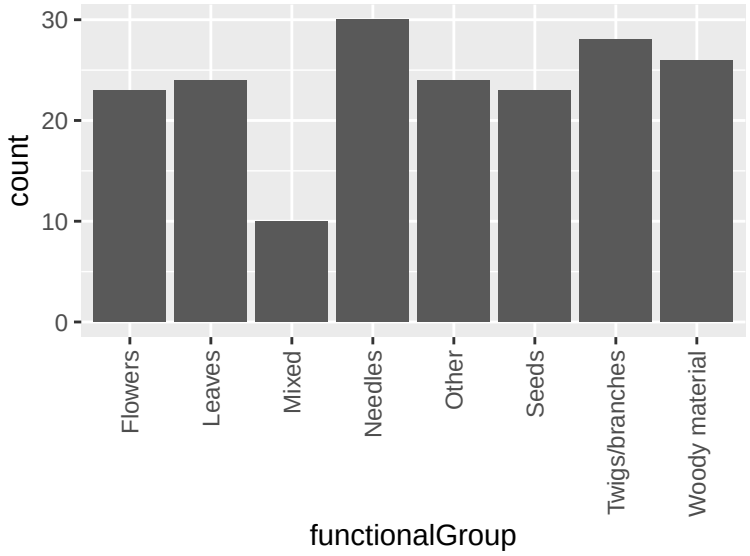
```
unique(Litter$namedLocation) # find unique locations
```

```
##  [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
```

Answer:12 plots were sampled at Niwot Ridge. Summary gives both the locations and how many records are at each location. Unique just gives a list of the different unique locations and a count of how many there are.

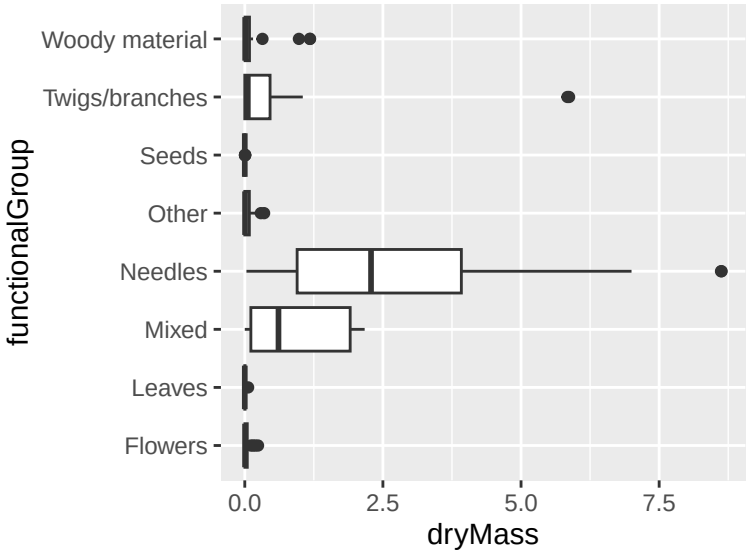
14. Create a bar graph of functionalGroup counts. This shows you what type of litter is collected at the Niwot Ridge sites. Notice that litter types are fairly equally distributed across the Niwot Ridge sites.

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

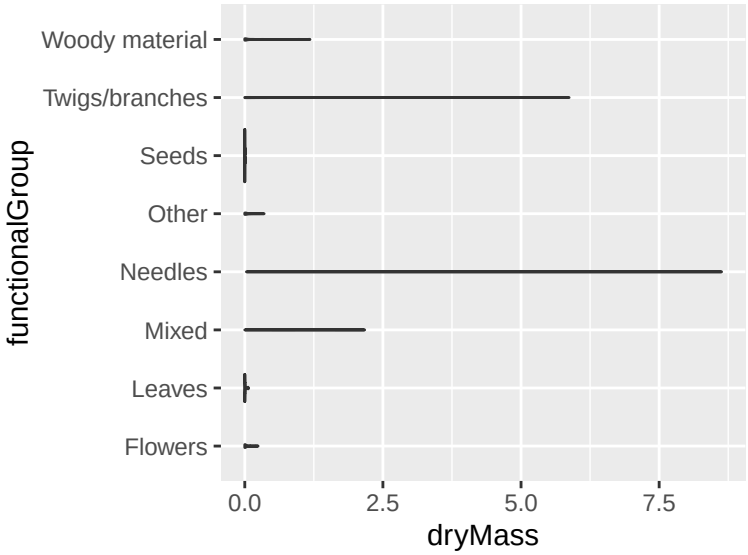


15. Using `geom_boxplot` and `geom_violin`, create a boxplot and a violin plot of dryMass by functionalGroup.

```
ggplot(Litter) +
  geom_boxplot(aes(x = dryMass, y = functionalGroup)) # plot boxplot of dry mass by functional group
```




```
ggplot(Litter) +
  geom_violin(aes(x = dryMass, y = functionalGroup)) # plot boxplot of dry mass by functional group
```




```
# I am not sure if it was supposed to have x and y axis but I think it looks best this way
```

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

> Answer: the box plot is more effective because the violin plots don't show the difference between where the majority of the measurments are and the outliers. the boxplots make it clear where the outliers are

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

> Answer: needles