Assignment 3: Data Exploration

Natasha Jacob, Section #3

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 <>.

```
library(knitr)
opts_chunk$set(tidy.opts=list(width.cutoff=60),tidy=TRUE, echo = T)
```

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.**

```
# Checking working directory
getwd()
```

```
## [1] "/Users/admin/Desktop/ENV872_EDA/Environmental_Data_Analytics_2022/Assignments"
```

```
# Loading tidyverse

library(tidyverse)
```

```
## -- Attaching packages --------------------------------------- tidyverse 1.3.1 --
```

```
## v ggplot2 3.3.5     v purrr   0.3.4
## v tibble  3.1.6     v dplyr   1.0.7
## v tidyr   1.1.4     v stringr 1.4.0
## v readr   2.1.1     v forcats 0.5.1
```

```
## -- Conflicts ------------------------------------------ tidyverse_conflicts() --
## x dplyr::filter() masks stats::filter()
## x dplyr::lag()    masks stats::lag()
```

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

   Answer: Neonicotinoids are a new class of insecticides chemically related to nicotine. Like nicotine, the neonicotinoids act on certain kinds of receptors in the nerve synapse. They are much more toxic to invertebrates, like insects, than they are to other animals. Neonicotinoids are water soluble which enables easy uptake by plants. New research done on Neonicotinoids points to potential toxicity to bees and other beneficial insects through low level contamination of nectar and pollen with neonicotinoid insecticides used in agriculture. Although these low level exposures do not normally kill bees directly, they may impact some bees' ability to foraging for nectar, learn and remember where flowers are located, and possibly impair their ability to find their way home to the nest or hive. Despite the controlled studies completed to date, the actual impact of neonicotinoid insecticides on honey bees in the field are difficult to measure.

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: Trees that fall and decay in the forest add nutrients to the forest soil and retain moisture in the forest. Fallen wood greater than 7 cm diameter is referred to as coarse woody debris. The time coarse woody litter takes to decompose is dependent upon moisture and temperature.Whereas, Fine woody material dries quickly and therefore decays slowly. Fine woody debris may act as a tinder that promotes the start and spreading of forest fires. Coarse woody debris adds long-lasting unique habitat structure and resources to both terrestrial and aquatic habitats, by being an important source of energy and nutrients for microorganisms and detritivores, trapping sediments, offering protection against harsh environmental conditions, and serving as refugium from predation to both consumers and inhabitants of decomposing wood. Coarse woody debris increases the diversity of biological communities in aquatic and terrestrial habitats.

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 sampling is executed at terrestrial NEON sites that contain woody vegetation >2m tall. * Along with most of NEON's plant productivity measurements, litter and fine woody debris sampling occurs only in tower plots. Locations of tower plots are selected randomly within the 90% flux footprint of the primary and secondary airsheds * Ground traps are sampled once per year. Target sampling frequency for elevated traps varies by vegetation present at the site, with frequent sampling (1x every 2weeks) in deciduous forest sites during senescence, and in- frequent year-round sampling (1x every 1-2 months) at evergreen sites.

Obtain basic summaries of your data (Neonics)

5. What are the dimensions of the dataset?

```
# Using the dim function to view the dimension of the
# dataset
dim(Neonics)  #The dataset has 4623 rows and 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)
```

```
##     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. These effects are commonly studied to monitor the effect of the Neonicotinoids on the 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: The six most commonly studied species in the dataset are Honey Bee, Parasitic Wasp, Buff Tailed Bumblebee, Carniolan Honey Bee, Bumble Bee and Italian Honeybee. These insects which belong to the same order 'Hymenoptera' are at greater risk since new research done on Neonicotinoids points to potential toxicity to bees and other beneficial insects through low level contamination of nectar and pollen with neonicotinoid insecticides used in agriculture.

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

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

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

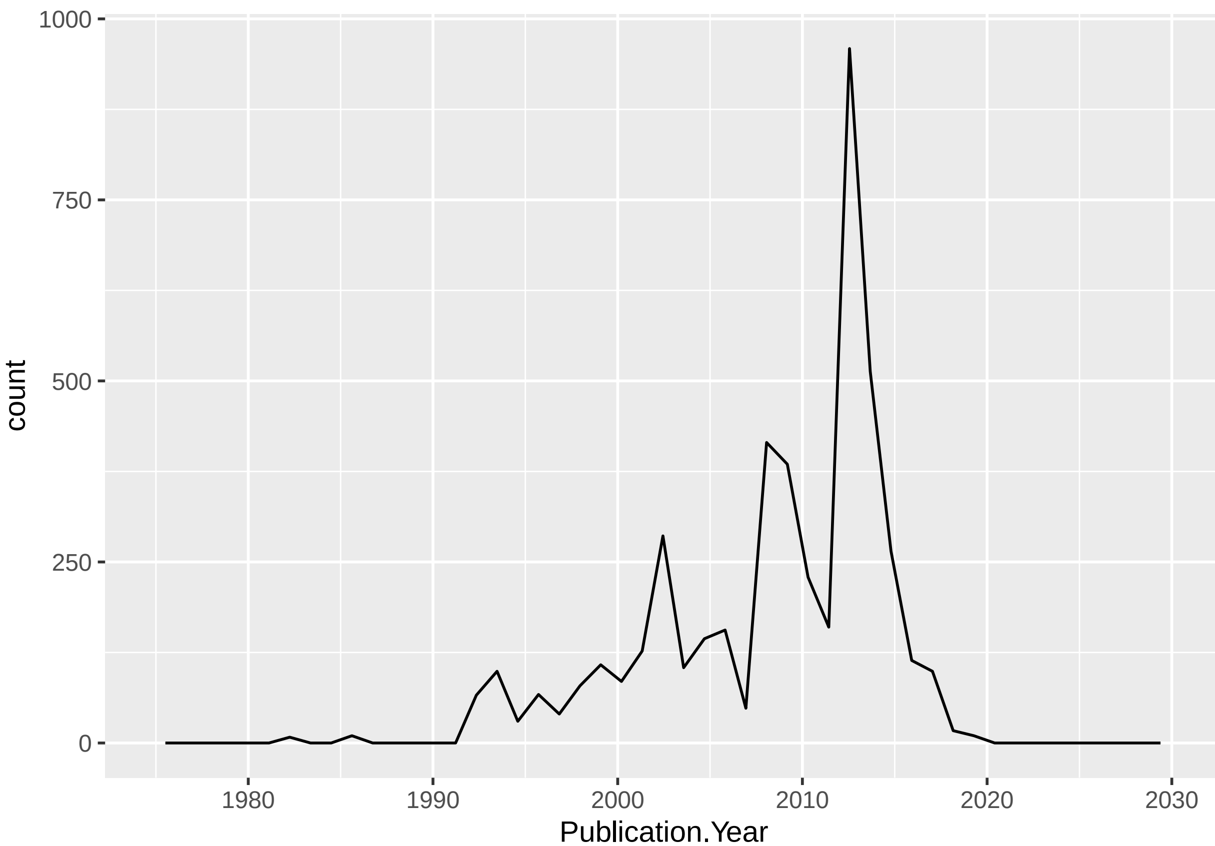
Answer: The dataset is not identified as numeric since there are other characters such as "NR", "~" and "/" present in the dataset which are not numeric values.

Explore your data graphically (Neonics)

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

```
ggplot(Neonics) + geom_freqpoly(aes(x = Publication.Year), bins = 50) +
    scale_x_continuous(limits = c(1975, 2030))
```

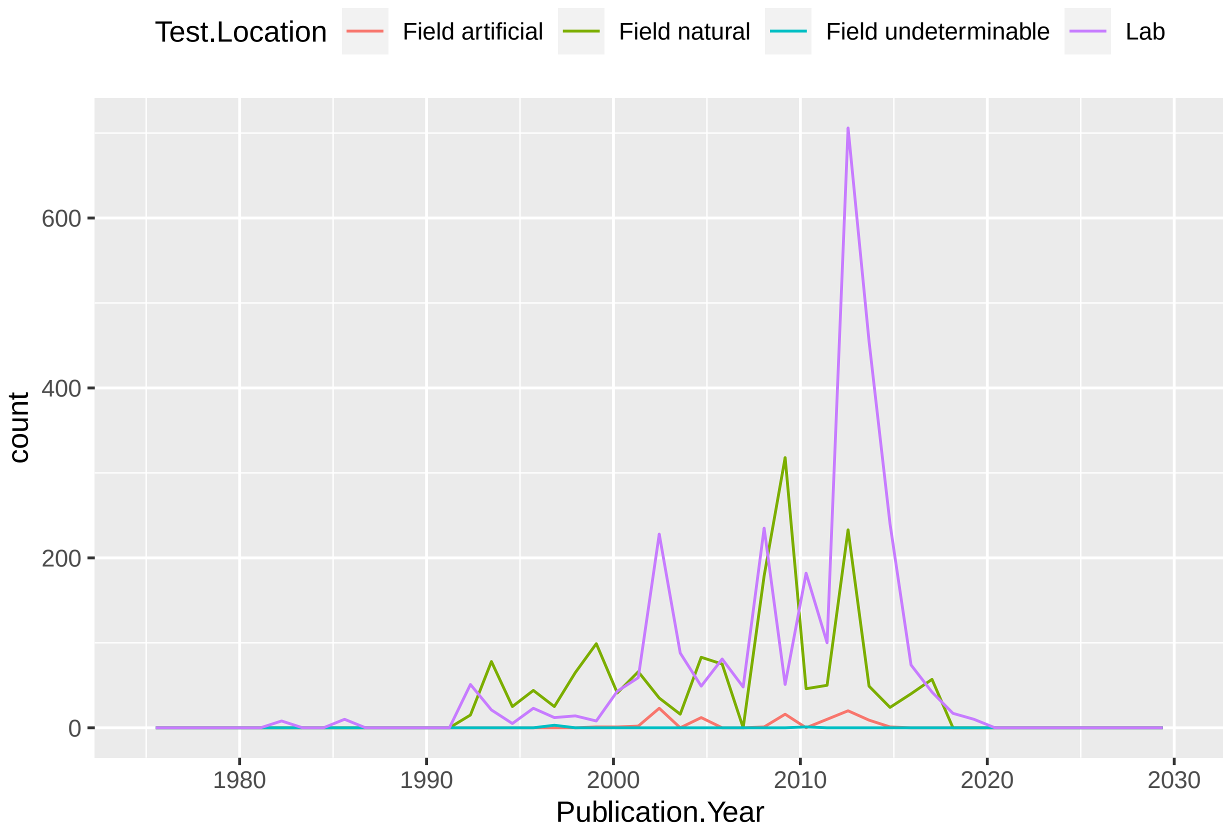
```
## Warning: Removed 3 row(s) containing missing values (geom_path).
```



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),
    bins = 50) + scale_x_continuous(limits = c(1975, 2030)) +
    theme(legend.position = "top")
```

```
## Warning: Removed 12 row(s) containing missing values (geom_path).
```

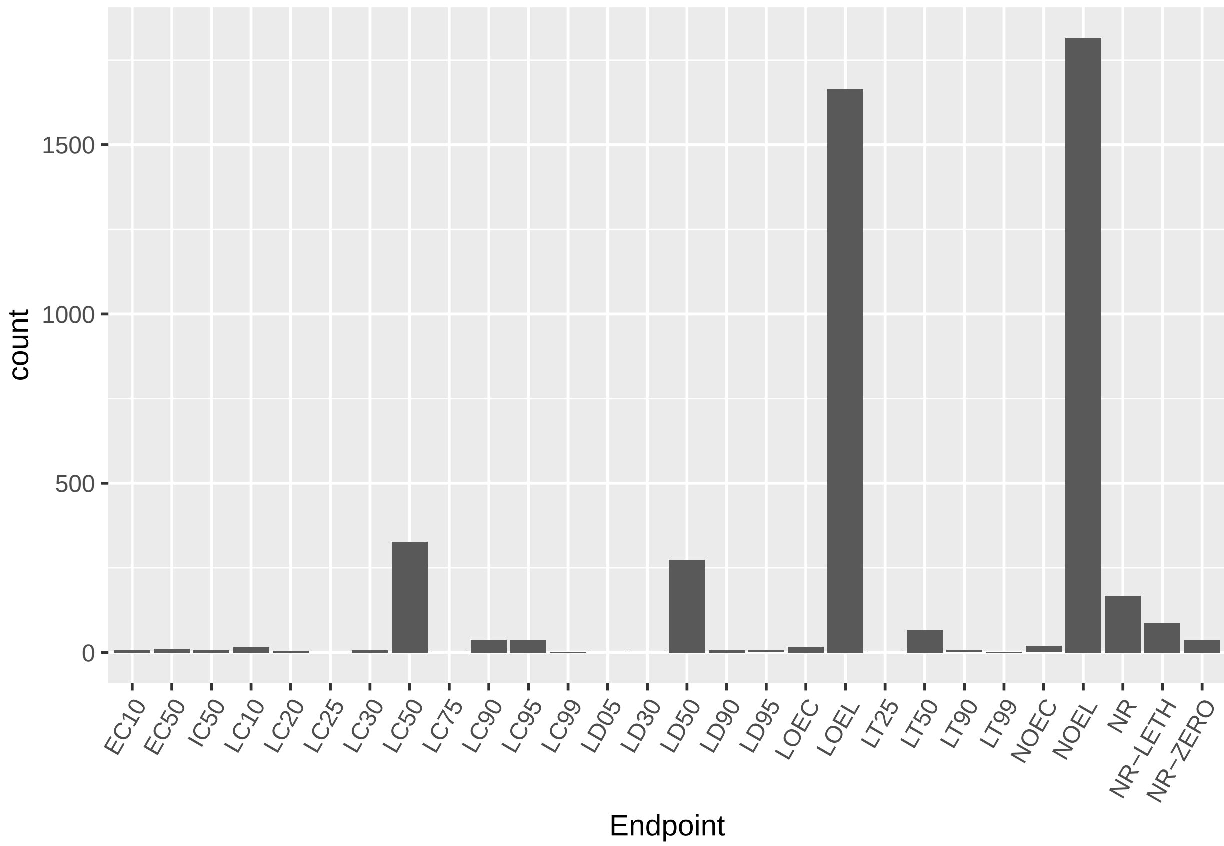


Interpret this graph. What are the most common test locations, and do they differ over time?

> Answer: The most common test locations are Field Natural and Lab. In the earlier years (1990's), Field Natural seems to have been a common test location. Around early 2000's during the boon of technology, labs were used as common test locations and we can observe a sharp peak at around 2013.

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_bargraph <- ggplot(Neonics, aes(x = Endpoint)) + geom_bar()
Endpoints_bargraph + theme(axis.text.x = element_text(angle = 60,
    hjust = 1))
```




```
unique(Neonics$Endpoint)
```

```
##  [1] LD50    LC50    IC50    LOEL    NR      NR-ZERO LC95    LC90    NOEL
## [10] LOEC    EC50    NOEC    NR-LETH LD90    LC10    LT50    LT90    LT25
## [19] LC75    LD30    EC10    LC25    LC20    LD05    LC99    LT99    LC30
## [28] LD95
## 28 Levels: EC10 EC50 IC50 LC10 LC20 LC25 LC30 LC50 LC75 LC90 LC95 LC99 ... NR-ZERO
```

```
summary(Neonics$Endpoint)
```

```
##    EC10    EC50    IC50    LC10    LC20    LC25    LC30    LC50    LC75    LC90
##       6      11       6      15       5       1       6     327       1      37
##    LC95    LC99    LD05    LD30    LD50    LD90    LD95    LOEC    LOEL    LT25
##      36       2       1       1     274       6       7      17    1664       1
##    LT50    LT90    LT99    NOEC    NOEL      NR NR-LETH NR-ZERO
##      65       7       2      19    1816     167      86      37
```

Answer: The two most common endpoints are LOEL and NOEL. LOEL - Lowest observable effect level, NOEL - No observable effect level.

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.

```r
# Determining the format
class(Litter$collectDate)
```

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

```r
# Changing format from factor to date
DateFormat <- as.Date(Litter$collectDate, format = "%Y-%m-%d")
DateFormat
```

```
##   [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"
```

```r
# Determining the new format
class(DateFormat)
```

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

```r
# Using unique function to determine which dates litter was
# sampled in August 2018
unique(Litter$collectDate)
```

```
## [1] 2018-08-02 2018-08-30
## Levels: 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`?

```r
unique(Litter$plotID)
```

```
##  [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
```

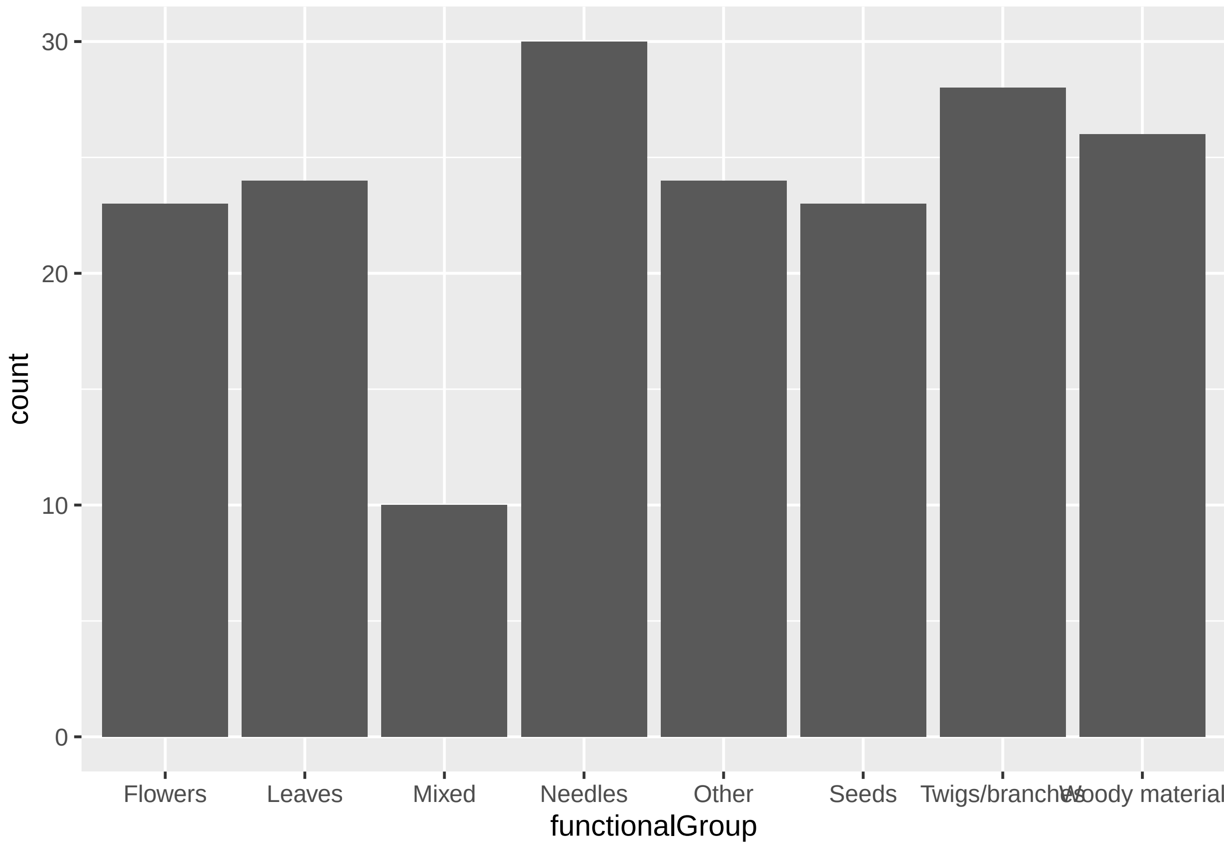
```r
summary(Litter$plotID)
```

```
## 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: Twelve plots were sampled at Niwot Ridge. While the unique function provies the plot ID and the total number of plots that were sampled at Niwot Ridge, the summary function provides information on the number of each individual plots among the twelve plots.

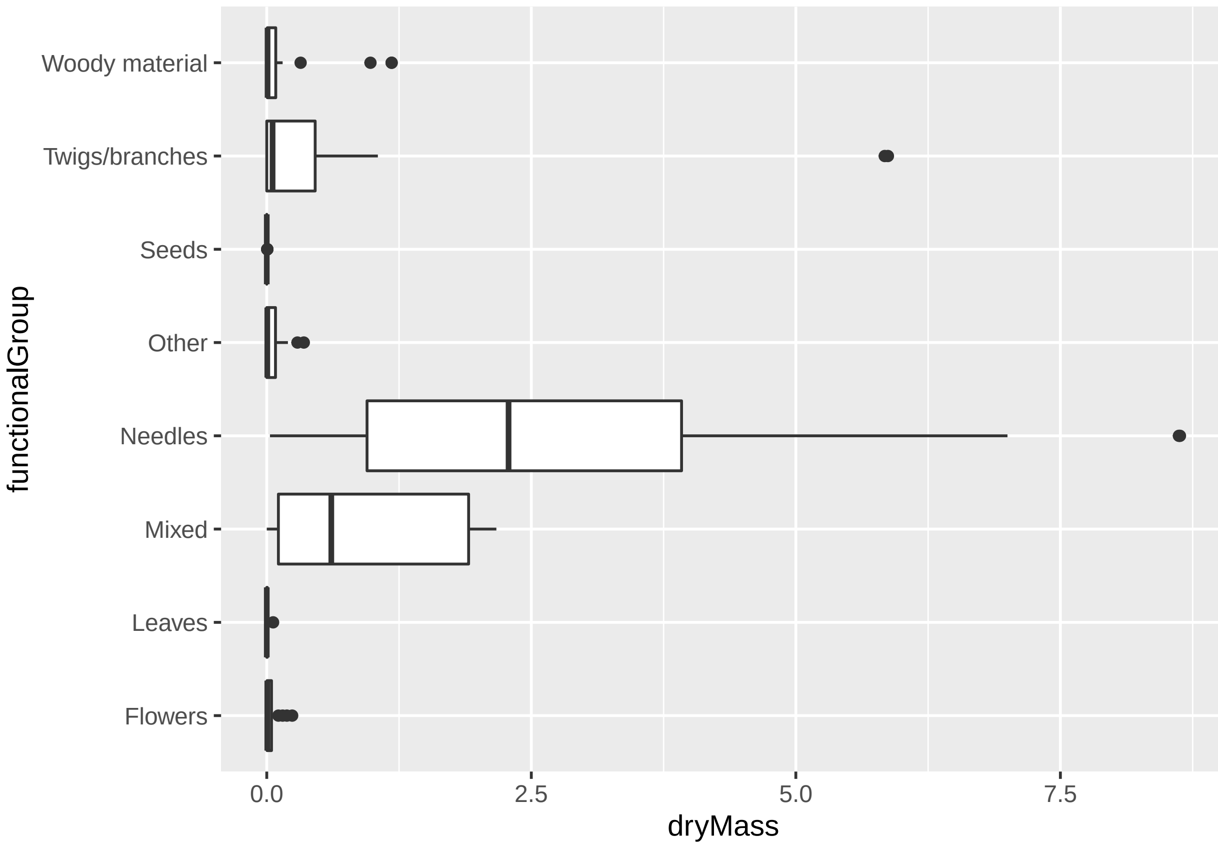
14. Create a bar graph of functionalGroup counts. This shows you what type of litter is collected at the Niwot Ridge sites. Notice that litter types are fairly equally distributed across the Niwot Ridge sites.

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



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,
    group = cut_width(functionalGroup, 1)))
```

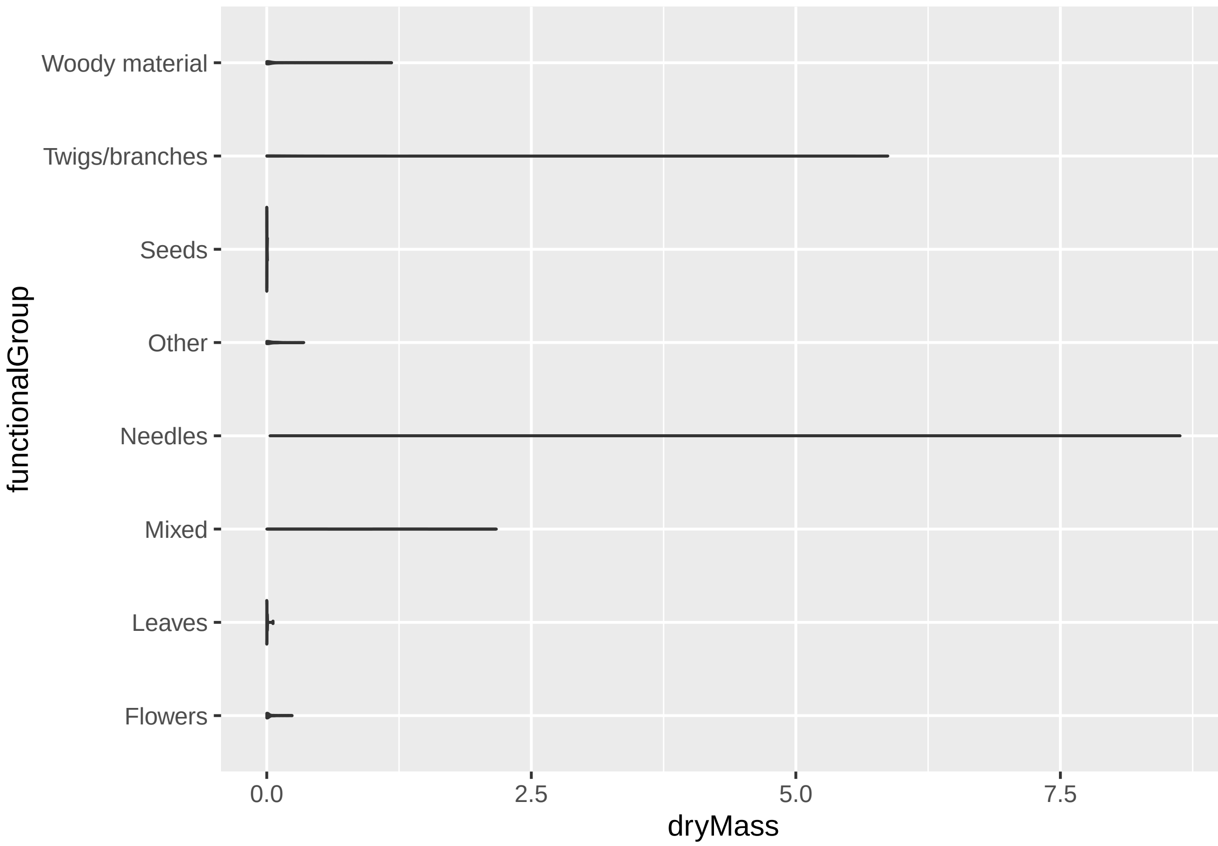



```
ggplot(Litter) + geom_violin(aes(x = dryMass, y = functionalGroup),
    draw_quantiles = c(0.25, 0.5, 0.75))
```

```
## Warning in regularize.values(x, y, ties, missing(ties), na.rm = na.rm):
## collapsing to unique 'x' values

## Warning in regularize.values(x, y, ties, missing(ties), na.rm = na.rm):
## collapsing to unique 'x' values

## Warning in regularize.values(x, y, ties, missing(ties), na.rm = na.rm):
## collapsing to unique 'x' values
```



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

> Answer: The boxplot is a more effective visualization option than the violin plot in this case since we can clearly identify the Interquartile range and the median of our data along with the outliers. Whereas, in the violin plot which displays density distributions, we are not able to view the violins since the data is not distributed evenly.

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

> Answer: Needles seems to have the highest biomass, followed by mixed.