Assignment 3: Data Exploration

Candela Cerpa

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 data sets: the ECOTOX neonicotinoid data set (ECOTOX_Neonicotinoids_Insects_raw.csv) and the Niwot Ridge NEON data set for litter and woody debris (NEON_NIWO_Litter_massdata_2018-08_raw.csv). Name these data sets "Neonics" and "Litter", respectively. Be sure to include the subcommand to read strings in as factors.

```r
getwd() #checking working directory
```

```
## [1] "Z:/EDE_Fall2023"
```

```r
library(tidyverse) #loading necessary package
```

```
## -- Attaching core tidyverse packages ------------------------ tidyverse 2.0.0 --
## v dplyr     1.1.3     v readr     2.1.4
## v forcats   1.0.0     v stringr   1.5.0
## v ggplot2   3.4.3     v tibble    3.2.1
## v lubridate 1.9.2     v tidyr     1.3.0
## v purrr     1.0.2
## -- Conflicts ------------------------------------------ tidyverse_conflicts() --
## x dplyr::filter() masks stats::filter()
## x dplyr::lag()    masks stats::lag()
## i Use the conflicted package (<http://conflicted.r-lib.org/>) to force all conflicts to become errors
```

```r
library(lubridate) #loading necessary package
Neonics <- read.csv("./Data/Raw/ECOTOX_Neonicotinoids_Insects_raw.csv") #import
#data with a relative path, naming it
Litter <- read.csv("./Data/Raw/NEON_NIWO_Litter_massdata_2018-08_raw.csv") #import
#data with a relative path, naming it
```

Learn about your system

2. The neonicotinoid data set 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 data set 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 impact non-target insects, such as bees, and aquatic invertebrate communities. Bees are very important pollinators, and the toxicity they have been experiencing for decades because of insecticides (like Neonicotinoids also others) has led to a significant decrease in bee population, which can be disastrous for ecosystems. Aquatic ecosystems are also impacted by neonicotinoids and it's difficult to remove pollutants from waterways.

3. The Niwot Ridge litter and woody debris data set 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 are important in forest and stream ecosystems because they are a part of carbon budgets and nutrient cycling, provide habitats for terrestrial and aquatic organisms, and influence water flows and sediment transport.

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. Sampling sites range from a size of 20m x 20 m (low–statured vegetation over the tower airsheds) to 40 m x 40 m (forested tower airsheds). 2. Little trap placement within plots may be targeted or randomnized, depending on the vegetation. If the site has 50% aerial cover of woody vegetation >2m in height, litter traps are randomly placed. Otherwise, there's a described method on how to place the traps. 3. Group traps are sampled once annually, while sampling for elevated traps varies by the site's vegetation; decidious forest sites should have frequent sampling (once every 2 weeks) during senescence, while evergreen sites need infrequent year-round sampling (once every 1-2 months).

Obtain basic summaries of your data (Neonics)

5. What are the dimensions of the data set?

```
dim(Neonics) # 188 rows, 19 columns
```

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

```
dim(Litter) # 4623 rows, 30 columns
```

```
## [1] 188  19
```

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?

```
Neonics$Effect <- as.factor(Neonics$Effect) #convert the character values in the
#column into factors. This allows us to quantify character values, particularly
#when they have a fixed and known set of possible values
summary(Neonics$Effect) #summarizes the column to find the most common values
```

```
##     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
```

```
sort(summary(Neonics$Effect), decreasing = TRUE)
```

```
##       Population        Mortality         Behavior Feeding behavior
##             1803             1493              360              255
##     Reproduction      Development        Avoidance         Genetics
##              197              136              102               82
##        Enzyme(s)           Growth       Morphology    Immunological
##               62               38               22               16
##     Accumulation     Intoxication     Biochemistry          Cell(s)
##               12               12               11                9
##       Physiology        Histology       Hormone(s)
##                7                5                1
```

Answer: The two most common studied effects are Population (1803 examples) and Mortality (1493). Mortality is likely studied to measure when the cause of death is by direct action of the chemical. Population measures the group of organisms of the same species in the same area at a given time. This would allow for a measure of biodiversity and, depending on how it is measured, may allow for a calculation of mortality rate.

7. Using the `summary` function, determine the six most commonly studied species in the data set (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...]

```
Neonics$Species.Common.Name <- as.factor(Neonics$Species.Common.Name) #convert
#the character values in the column into factors. This allows us to quantify
#character values, particularly when they have a fixed and known set of possible
#values
summary(Neonics$Species.Common.Name) #summarizes the column to find the most
```

```
##                          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
```

```r
#common values
sort(summary(Neonics$Species.Common.Name), decreasing = TRUE)
```

```
##                            (Other)                          Honey Bee
##                                670                                667
##                     Parasitic Wasp              Buff Tailed Bumblebee
##                                285                                183
##                Carniolan Honey Bee                         Bumble Bee
##                                152                                140
##                   Italian Honeybee                    Japanese Beetle
##                                113                                 94
##                  Asian Lady Beetle                     Euonymus Scale
##                                 76                                 75
##                           Wireworm                  European Dark Bee
##                                 69                                 66
##                  Minute Pirate Bug               Asian Citrus Psyllid
##                                 62                                 60
##                      Parastic Wasp             Colorado Potato Beetle
##                                 58                                 57
##                    Parasitoid Wasp                Erythrina Gall Wasp
##                                 51                                 49
##                       Beetle Order        Snout Beetle Family, Weevil
##                                 47                                 47
##           Sevenspotted Lady Beetle                     True Bug Order
##                                 46                                 45
##              Buff-tailed Bumblebee                       Aphid Family
##                                 39                                 38
##                     Cabbage Looper               Sweetpotato Whitefly
##                                 38                                 37
##                      Braconid Wasp                       Cotton Aphid
##                                 33                                 33
##                     Predatory Mite             Ladybird Beetle Family
##                                 33                                 30
##                         Parasitoid                      Scarab Beetle
##                                 30                                 29
##                      Spring Tiphia                        Thrip Order
##                                 29                                 29
##               Ground Beetle Family                 Rove Beetle Family
##                                 27                                 27
##                      Tobacco Aphid                       Chalcid Wasp
##                                 27                                 25
##             Convergent Lady Beetle                      Stingless Bee
##                                 25                                 25
##                  Spider/Mite Class                Tobacco Flea Beetle
##                                 24                                 24
##                   Citrus Leafminer                    Ladybird Beetle
##                                 23                                 23
##                          Mason Bee                           Mosquito
##                                 22                                 22
##                      Argentine Ant                             Beetle
##                                 21                                 21
##         Flatheaded Appletree Borer               Horned Oak Gall Wasp
##                                 20                                 20
##                 Leaf Beetle Family                  Potato Leafhopper
##                                 20                                 20
##         Tooth-necked Fungus Beetle                       Codling Moth
##                                 20                                 19
##          Black-spotted Lady Beetle                       Calico Scale
##                                 18                                 18
##                Fairyfly Parasitoid                        Lady Beetle
##                                 18                                 18
##             Minute Parasitic Wasps                          Mirid Bug
##                                 18                                 18
##                   Mulberry Pyralid                           Silkworm
##                                 18                                 18
##                     Vedalia Beetle              Araneoid Spider Order
##                                 18                                 17
##                          Bee Order                     Egg Parasitoid
##                                 17                                 17
##                       Insect Class           Moth And Butterfly Order
##                                 17                                 17
##       Oystershell Scale Parasitoid Hemlock Woolly Adelgid Lady Beetle
##                                 17                                 16
##              Hemlock Wooly Adelgid                               Mite
##                                 16                                 16
##                        Onion Thrip              Western Flower Thrips
##                                 16                                 15
##                       Corn Earworm                  Green Peach Aphid
##                                 14                                 14
##                          House Fly                          Ox Beetle
##                                 14                                 14
##                 Red Scale Parasite                 Spined Soldier Bug
##                                 14                                 14
##              Armoured Scale Family                   Diamondback Moth
##                                 13                                 13
##                      Eulophid Wasp                  Monarch Butterfly
##                                 13                                 13
##                      Predatory Bug              Yellow Fever Mosquito
##                                 13                                 13
##                Braconid Parasitoid                       Common Thrip
##                                 12                                 12
##       Eastern Subterranean Termite                             Jassid
##                                 12                                 12
##                         Mite Order                          Pea Aphid
##                                 12                                 12
##                   Pond Wolf Spider           Spotless Ladybird Beetle
##                                 12                                 11
##             Glasshouse Potato Wasp                           Lacewing
##                                 10                                 10
##            Southern House Mosquito            Two Spotted Lady Beetle
##                                 10                                 10
##                         Ant Family                       Apple Maggot
##                                  9                                  9
```

Answer: The six most commonly studied species are the Honey Bee (667), Parasitic Wasp (285), Buff Tailed Bumblebee (183), Carniolan Honey Bee (152), Bumble Bee (140), and Italian Honey Bee (113). All of these species are pollinators are not target insects of the insecticide. They are likely studied most because of the aforementioned broad decrease of pollinator (specially bees) populations because of insecticide use.

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

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

```
## [1] "character"
```

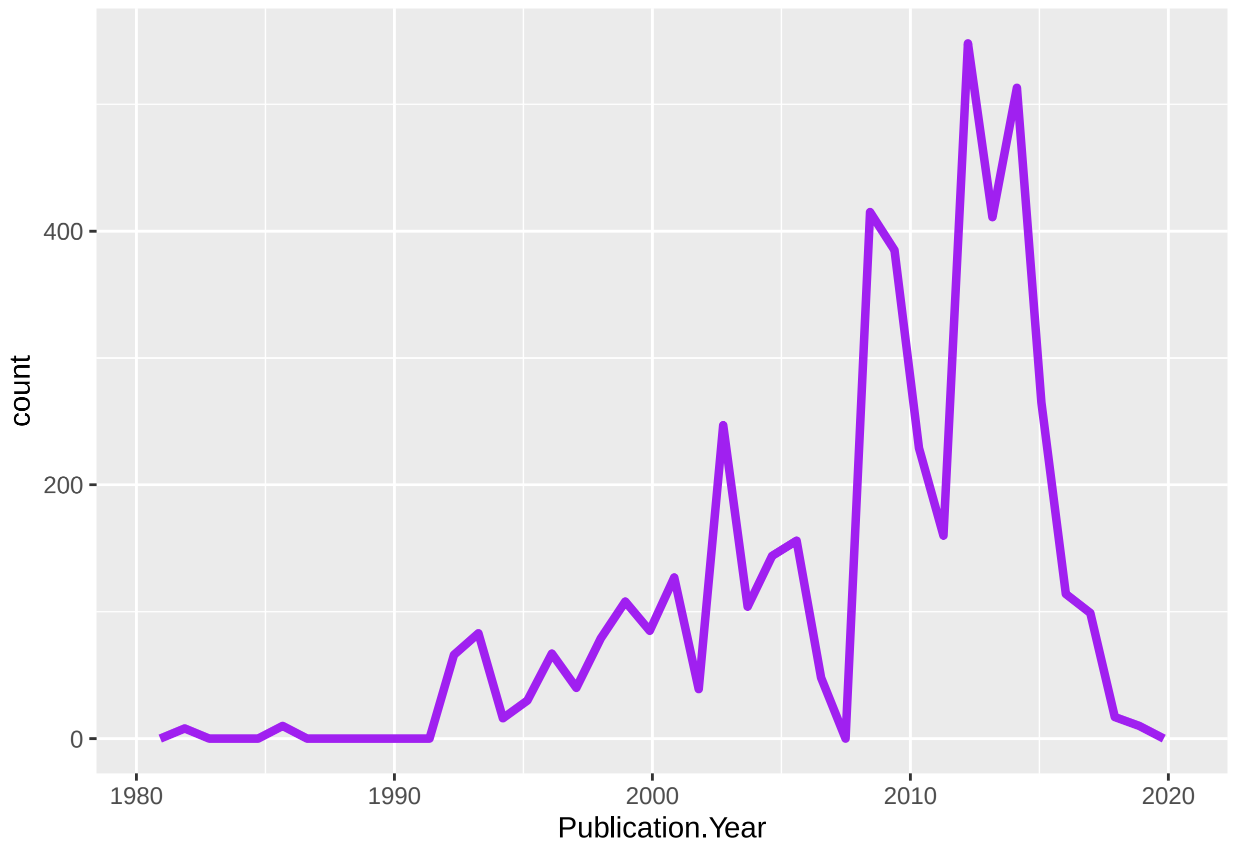
Answer: The values in the Conc.1..Author. column are characters. The numeric concentrations were reported as characters because not all elements are reported on the same unit and they are not measuring the same thing. Conc.1.Type..Author. outlines what the row is measuring (such as active ingredient), while Conc.1.Units..Author. determines what unit is to be used. This also allows for non-numeric characters, such as ~ to denote approximation and /. By listing the values as characters, the dataset authors disencourage summarizations of a column that is not standardized.

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 = 40, color = "purple", size = 1.5)
```

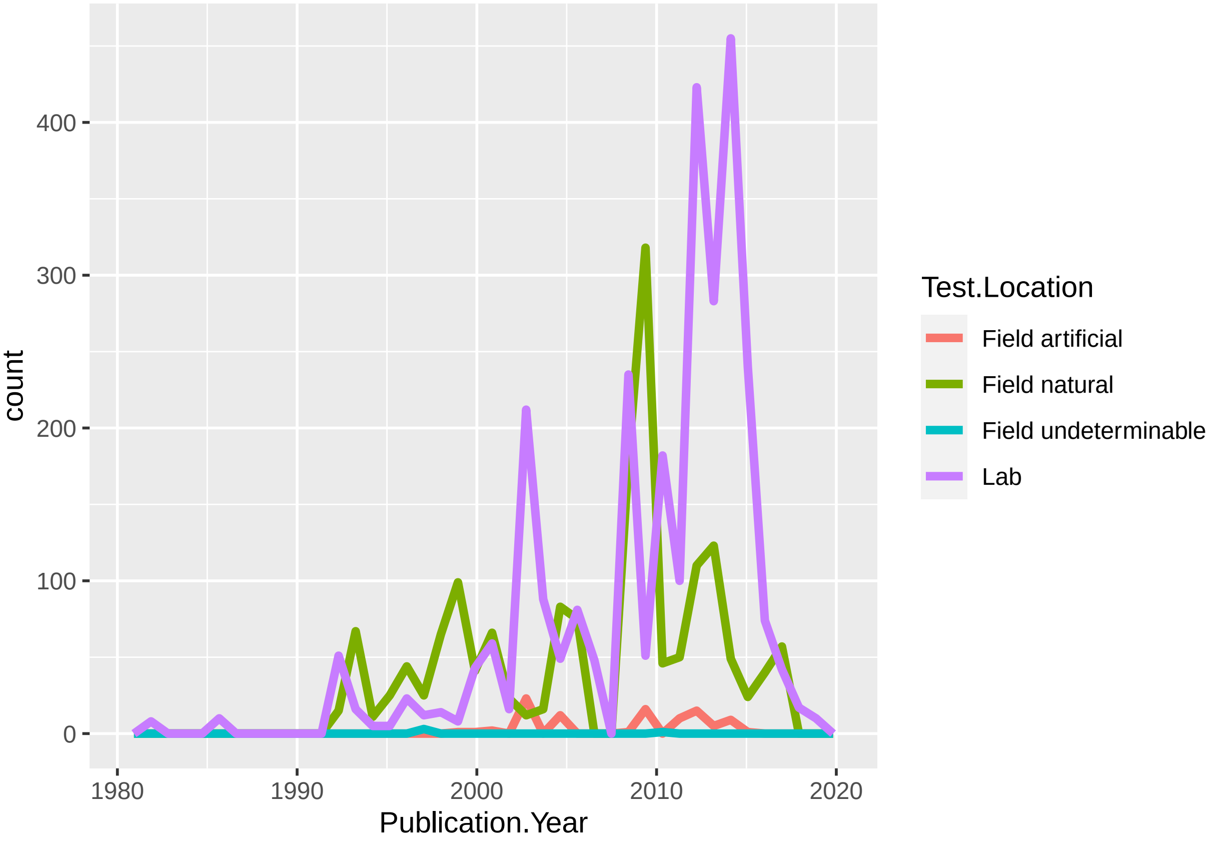
```
## Warning: Using ‘size‘ aesthetic for lines was deprecated in ggplot2 3.4.0.
## i Please use ‘linewidth‘ instead.
## This warning is displayed once every 8 hours.
## Call ‘lifecycle::last_lifecycle_warnings()‘ to see where this warning was
## generated.
```



```
#I graphed the specific column, increased the number of bins to see subtler
#trends over the years (40 is the difference between 2020 and 1980), and changed
#the line color and size to make it more visible.
```

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 = 40, size = 1.5)
```



```
#I kept the same basic graph but added the color determinant to be the elements
#in TestLocation to create 4 separate lines
```

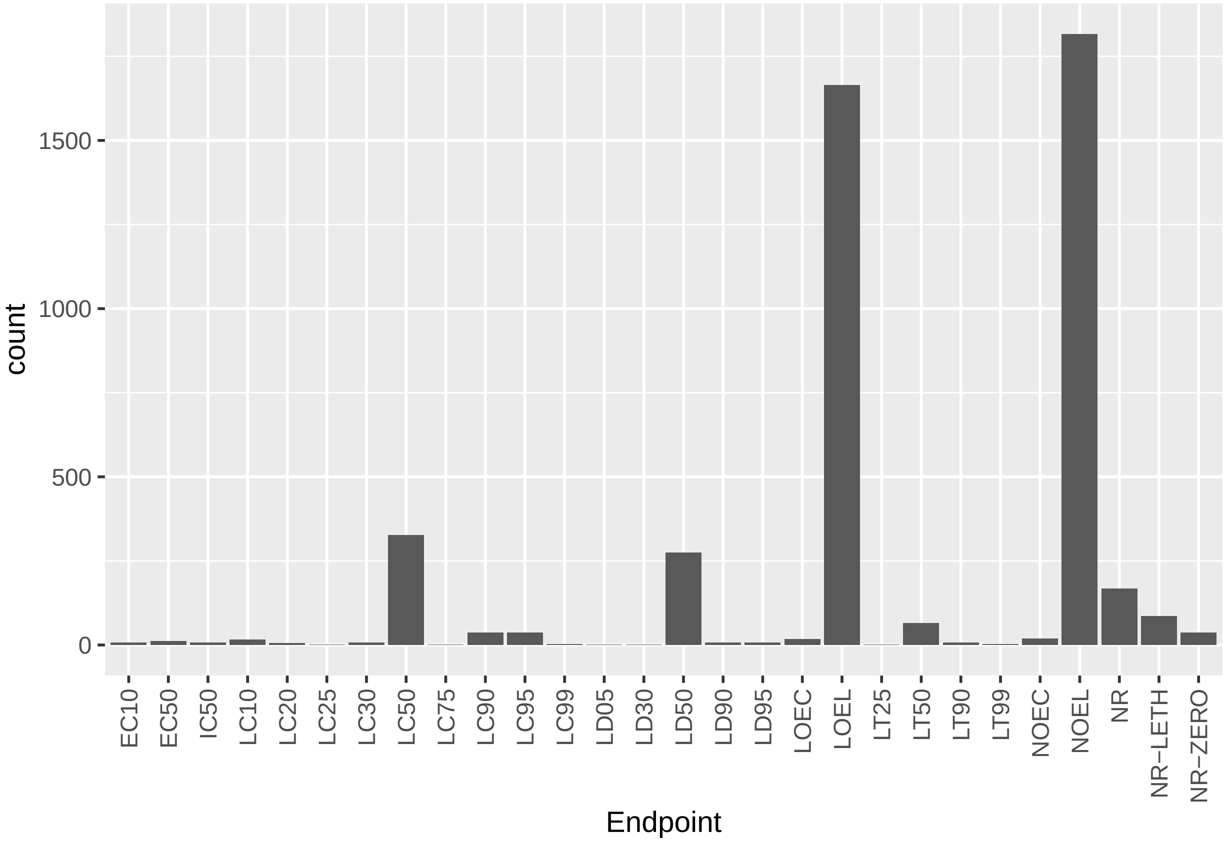
Interpret this graph. What are the most common test locations, and do they differ over time?

> Answer: By having the elements in Test.Location be the determining factor for the coloring, it creates a line graph for each related element; in this case, a graph of the studies published by year, by test location. The four types of location are Field Artificial, Field Natural, Field Undeterminable, and Lab. Lab is by far the most frequent, though there are times (such as in the 90s and right before 2010) when Field Natural (the second most popular type, is the most used test location for the studies published that year. Field artificial was used sometimes, but not every year and not many published studies use it. Field Undeterminable is an unused Test Location type in published studies.

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)) #rotates
```



```
#and aligns the x-axis labels
```

Answer: The two most common end points are NOEL and LOEL. NOEL is No-Observable-Effect-Level, whereused to note the highest concentration producing effects that aren't significantly different from control responses. LOEL is Lowest-Observable-Effect-Level, used to note the lowest concentration producing effects that are significantly different from controls.

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 the class of the column; it is currently character
```

```
## [1] "character"
```

```
Litter$collectDate <- as.Date(Litter$collectDate) #Transform column to Date format
class(Litter$collectDate) #confirm the new class of the Collect Date column
```

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

```
unique(Litter$collectDate) #determine when litter was sampled in Aug 2018
```

```
## [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`?

```
unique(Litter$siteID)
```

```
## [1] "NIWO"
```

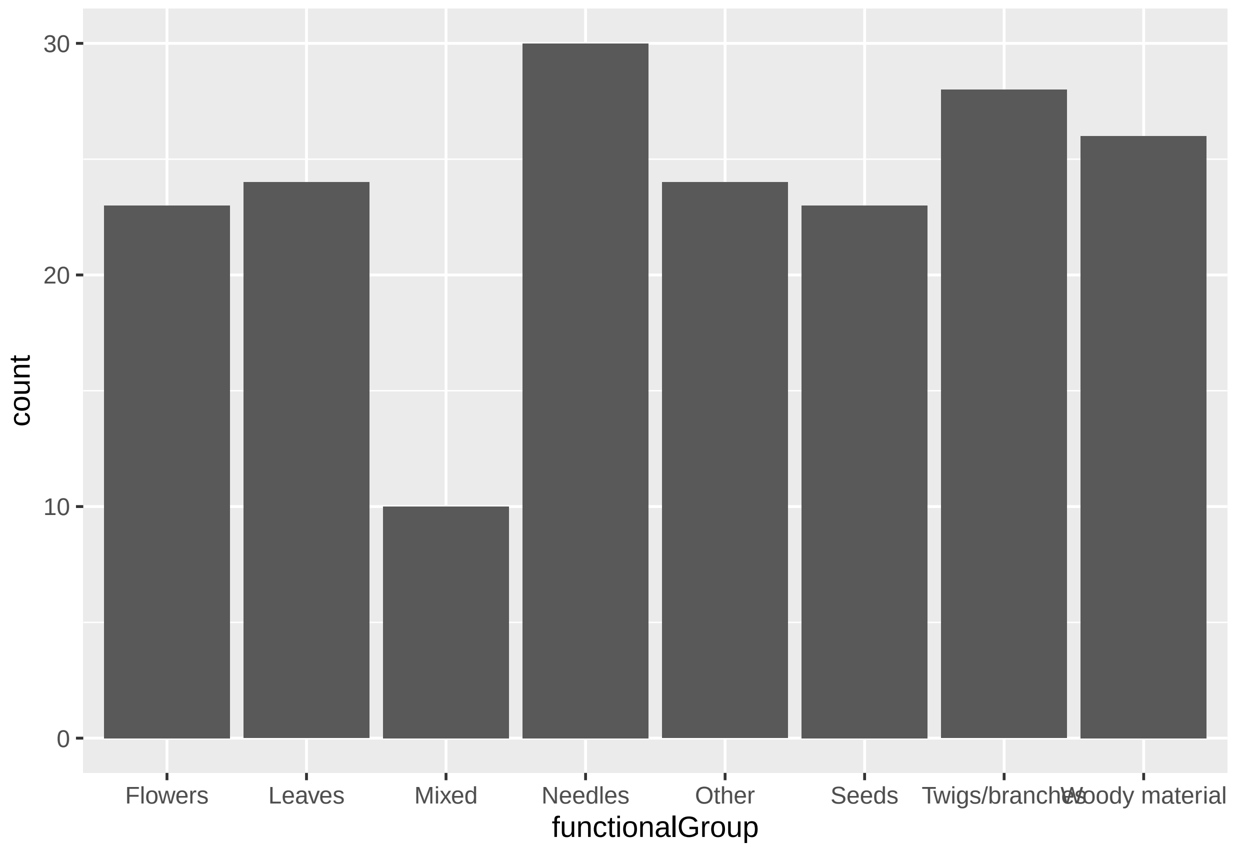
```
Litter$siteID <- as.factor(Litter$siteID) #change it to factor as before to make
#it possible to count them in summary
summary(Litter$siteID)
```

```
## NIWO
##  188
```

> Answer: All 188 plots were sampled at Niwot Ridge. The information is the same, just shown differently. Unique tells you the amount of unique values there are, so if there's only one value throughout the column and there are no NA/empty cells, you can assume its the length of your dataset. Summary, on the other hand, tells you the value in the column. If you turn it into an element, it can tell you the number of occurences of every value listed, but since there's only one here, it lists the length of the dataset.

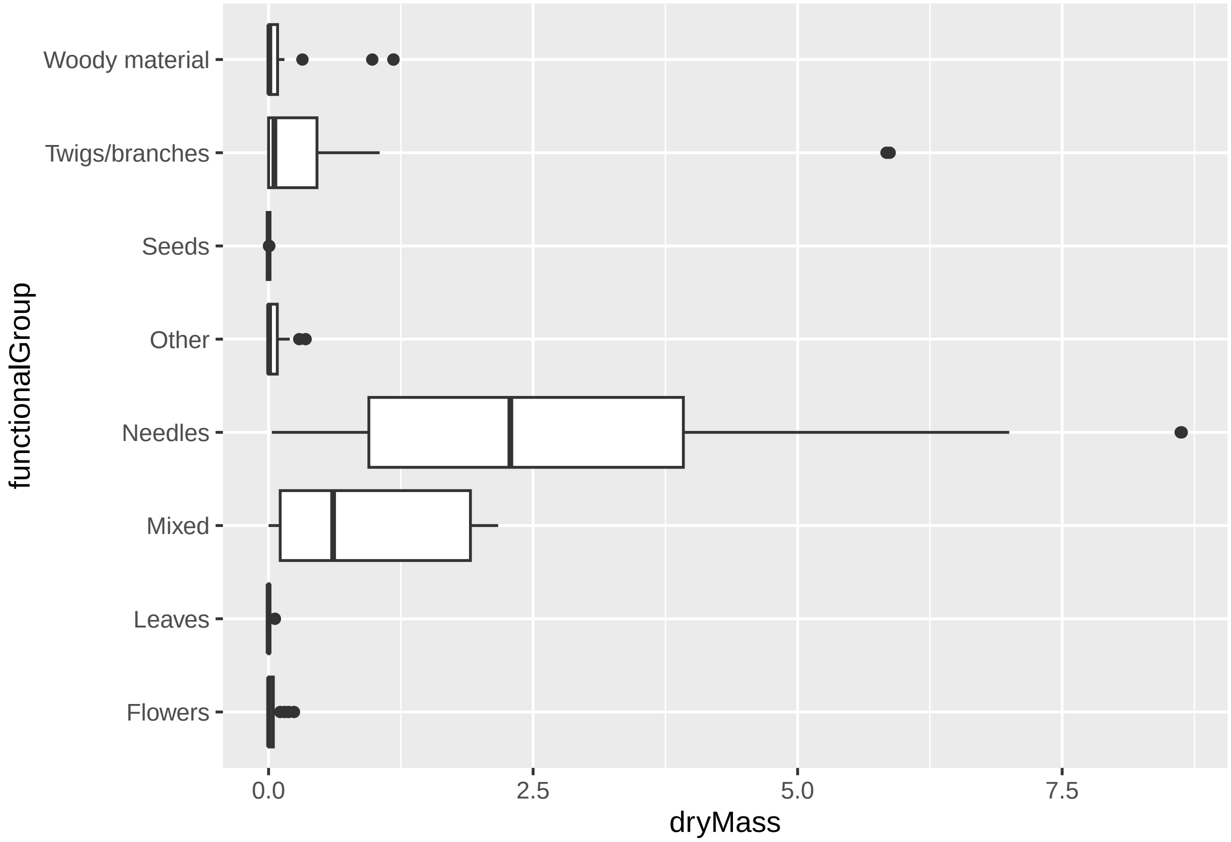
14. Create a bar graph of functionalGroup counts. This shows you what type of litter is collected at the Niwot Ridge sites. Notice that litter types are fairly equally distributed across the Niwot Ridge sites.

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

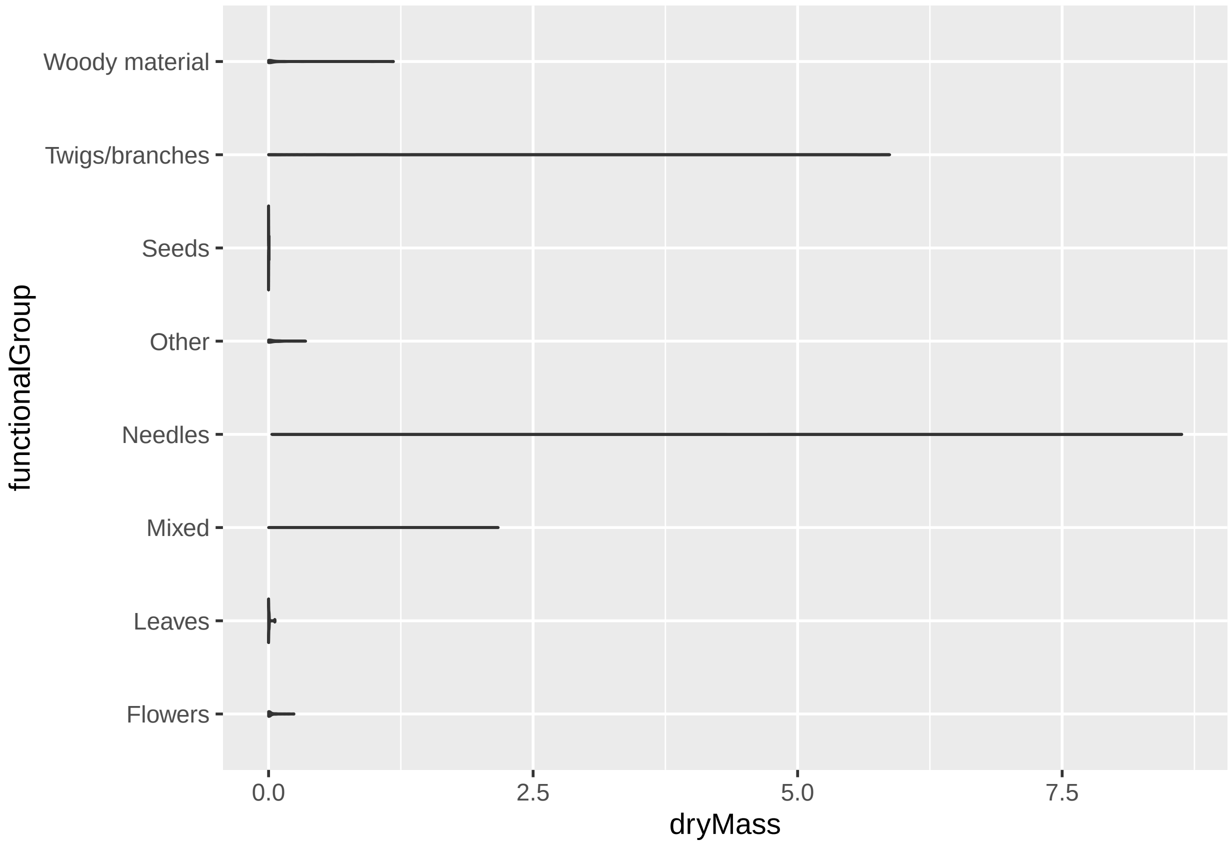


15. Using `geom_boxplot` and `geom_violin`, create a box plot and a violin plot of dryMass by functionalGroup.

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



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



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

> Answer: The violin plot is eant to show the distribution and probability density at different values, but this dataset is too disperse to easily visualize any trend with this plot. The boxplot, on the other hand, always hs a visible box to visualize the median, quartiles, and outliers.

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

> Answer: Needles.