Assignment 3: Data Exploration

Siyu Dong

Spring 2024

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.

```
#Check my working directory
getwd()
```

```
## [1] "/home/guest/EDA_Spring2024"
```

```
#Install packages and load them respectively
#install.packages("tidyverse")
library(tidyverse)
#install.packages("lubridate")
library(lubridate)
```

```r
#Upload the datasets and assign names for them with subcommand included
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)
#I can use the read.csv("./") then Tab to find the relative data path in the console pane
#but not in this pane
```

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: The reason maybe that there is potential adverse effects these insecticides can have on non-target organisms, particularly pollinators such as bees and other beneficial insects. Neonicotinoids are a class of systemic insecticides commonly used in agriculture to control pests. They are absorbed by plants and can be present in various plant tissues, including nectar and pollen.It is suggested that exposure to neonicotinoids can have detrimental effects on pollinators, leading to concerns about their impact on bee populations and overall ecosystem health. Bees and other pollinators play a crucial role in the pollination of many crops and wild plants, contributing to the biodiversity of ecosystems and supporting food production.

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 accumulation and decomposition of litter and woody debris play significant roles in shaping the structure and functioning of forest ecosystems. It contributes to our understanding of nutrient cycling, carbon sequestration, biodiversity, and the overall resilience of forested landscapes.

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.Functional Groups and Sorting: Mass data for each collection event are measured separately for different functional groups to an accuracy of 0.01 grams. The sorting process may identify functional groups even at weights below 0.01 grams, indicating their presence, although not at detectable masses. 2.Spatial Sampling Design: Litter and fine woody debris sampling occur at terrestrial NEON sites containing woody vegetation taller than 2m. Sampling is executed in tower plots, with the location of tower plots selected randomly within the 90% flux footprint of the primary and secondary airsheds. The number of plots sampled depends on the characteristics of the vegetation. 3. Temportal Sampling Design: Ground traps are sampled once per year, while elevated traps have varying sampling frequencies depending on the vegetation present at the site. For deciduous forest sites, frequent sampling (1x every 2 weeks) occurs during senescence, while evergreen sites have infrequent year-round sampling (1x every 1-2 months).

Obtain basic summaries of your data (Neonics)

5. What are the dimensions of the dataset?

```
dim(Neonics)
```

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

```
#The Neonics Dataset has 4623 rows
#The Neonics Dataset has 30 columns
```

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 top three effects found with using the 'summary' function is **Popuplation** (1803 times), **Mortality** (1493 times), and **Behavior** (360 times). The reasons account for their high frequency could be: 1. Study Focus: This investigation was designed with particular purpose of examing these three effects. 2. Data Availability: When conducting the research, these three parameters are more likely to be collected. 3. Ecological Significance: These three parameters do have their critical implications in ecological studies.

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...]

```
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
```

```
sort(summary(Neonics$Species.Common.Name))
```

```
##                         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
```

```
#Just as least-to-most order?
```

Answer: The six most commonly studied species are: Honey Bee, Parasitic Wasp, Buff Tailed Bumblebee, Carniolan Honey Bee, Bumble Bee, Italian Honeybee. They are crucial pollinators to our ecosystem and distribute vastly. Also, studying the effects of neonicotinoids on social behavior, communication, and colony dynamics provides insights into how these pesticides may impact entire colonies.

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.)
```

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

Answer: As outcome above, this value is a factor value. I think the main reason for that is when entering these concentrations data, there are different formats of them, like “0.6”, “0.1/”, which could not be distinguished when reading the dataset.

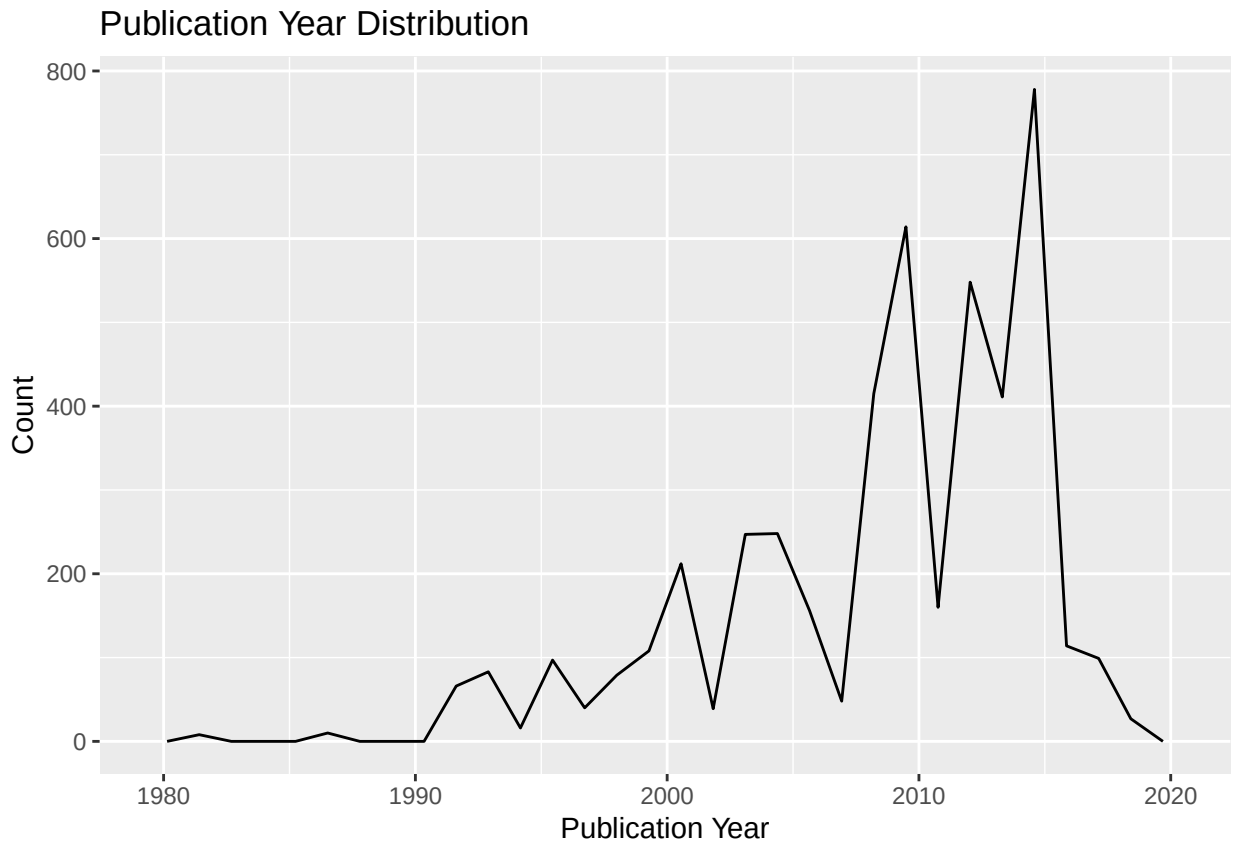
Explore your data graphically (Neonics)

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

```
library(ggplot2)

ggplot(Neonics) +
  geom_freqpoly(aes(x = Publication.Year)) +
  labs(title = "Publication Year Distribution",
       x = "Publication Year",
       y = "Count")
```

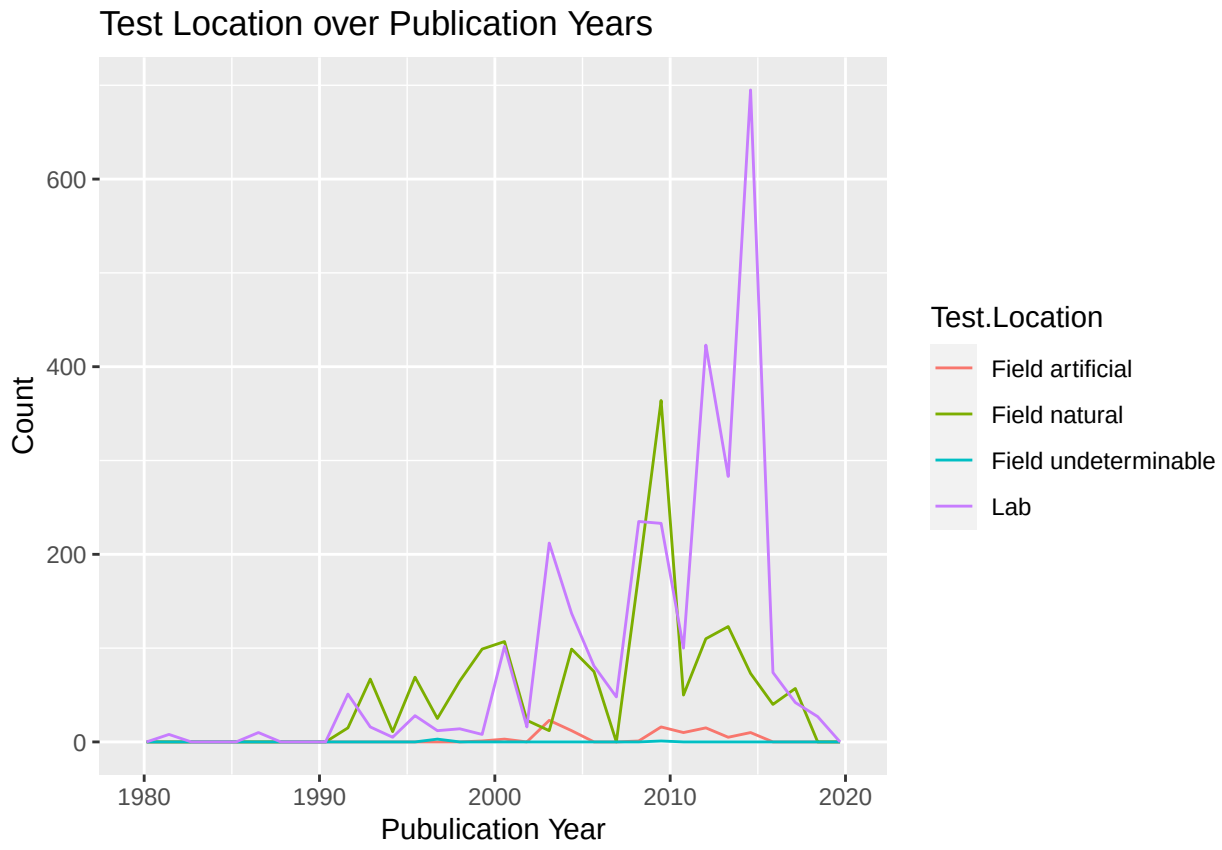
```
## `stat_bin()` using `bins = 30`. Pick better value with `binwidth`.
```



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)) +
  labs(title = "Test Location over Publication Years",
       x = "Pubulication Year",
       y = "Count")
```

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



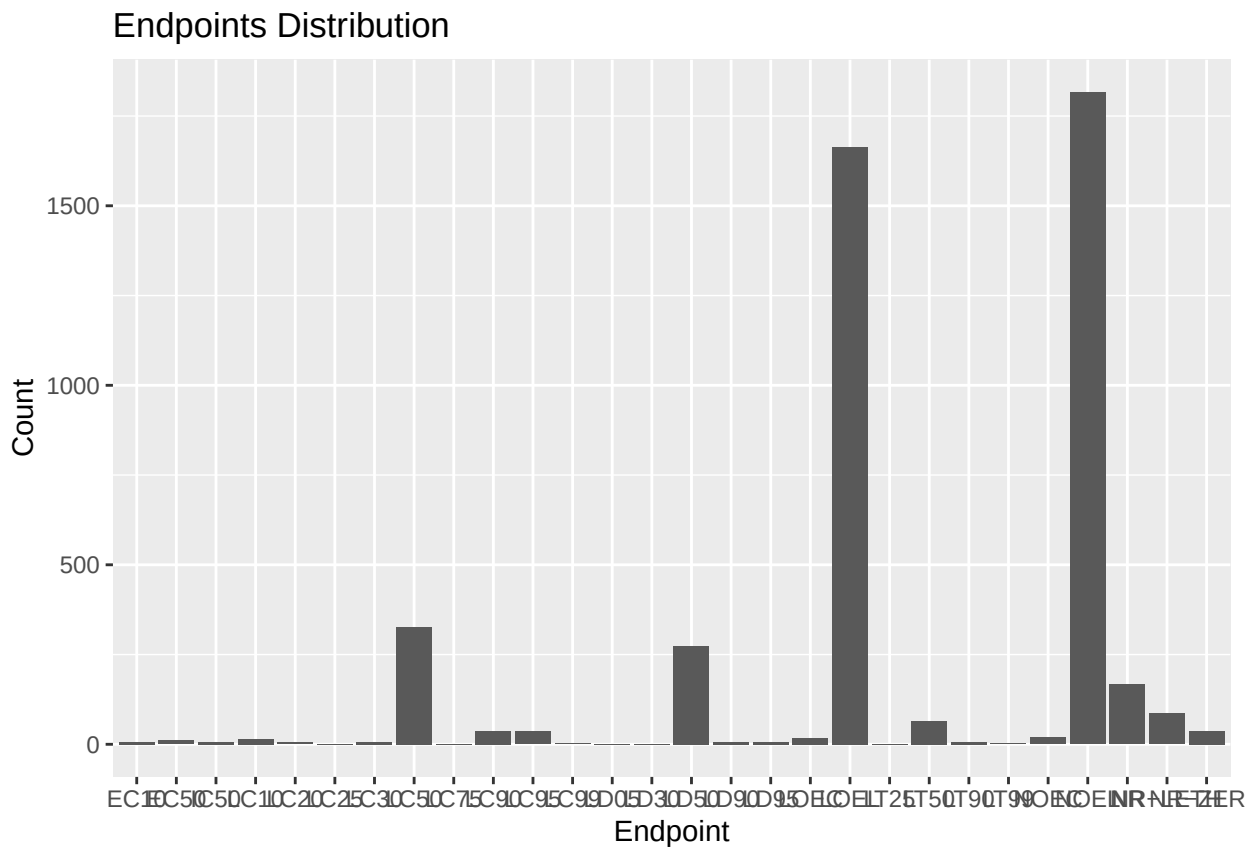
Interpret this graph. What are the most common test locations, and do they differ over time?

> Answer: Test locationa are selected through different time periods. From 1990 to 2000, Field natural is the top choice; from 2010 to 2020, Lab is the top choice. From 2000 to 2010, both Field natural and Lad have certain amount but changed greatly.

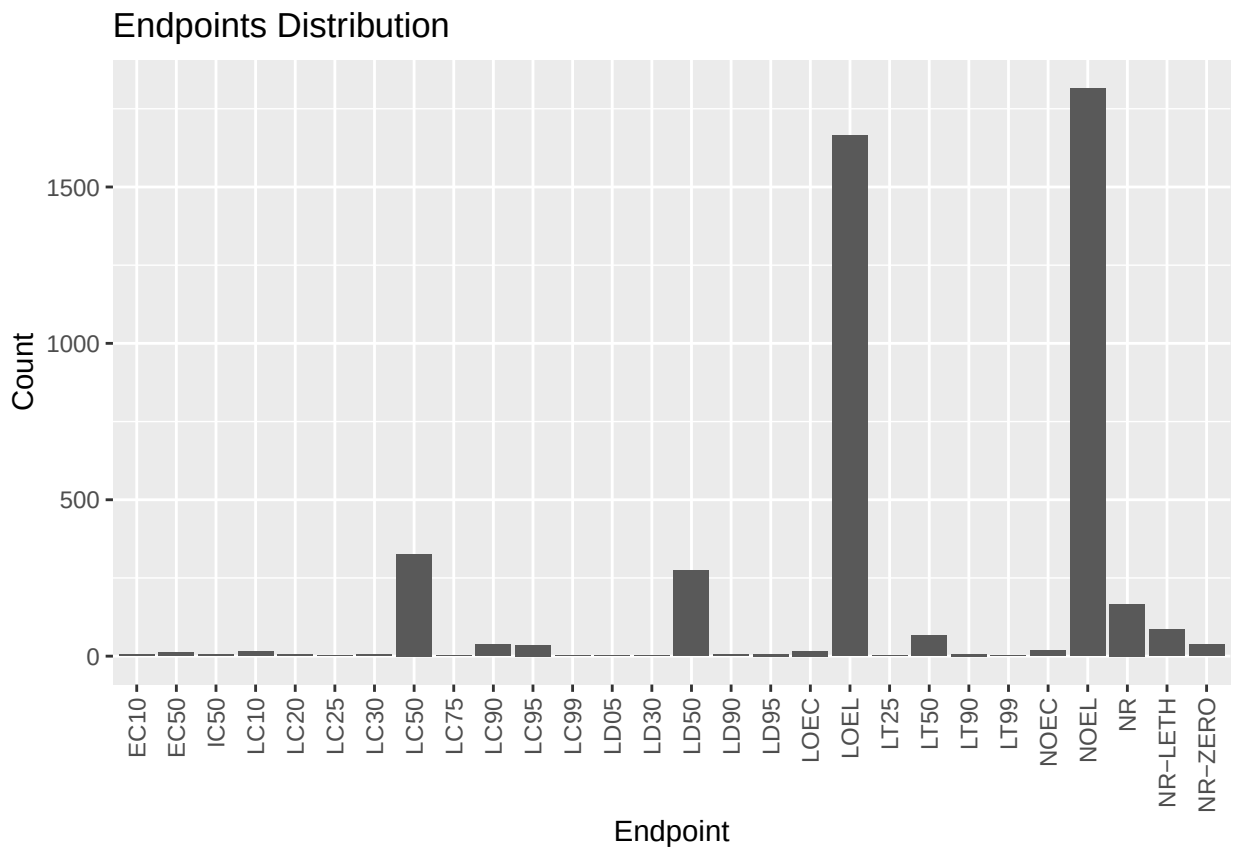
11. Create a bar graph of Endpoint counts. What are the two most common end points, and how are they defined? Consult the ECOTOX_CodeAppendix for more information.

[**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...]

```
#try1
ggplot(Neonics) +
  geom_bar(aes(x = Endpoint)) +
  labs(title = "Endpoints Distribution",
       x = "Endpoint",
       y = "Count")
```




```
#try2
ggplot(Neonics) +
  geom_bar(aes(x = Endpoint))+
  labs(title = "Endpoints Distribution",
       x = "Endpoint",
       y = "Count") +
  theme(axis.text.x = element_text(angle = 90, vjust = 0.5, hjust=1))
```




```
#Question1: Do we have code for making the exact count number showing on the top of each bar;
#Question2: Is it possible to rotate the x axis and y axis?
```

Answer: NOEL and LOEL are the two most common end points. The endpoint of NOEL is used for terrestial database, and is defined as the abbreviation for No-Observable-Effect-Level. In the context of ECOTOX_CodeAppendix, it refers to the highest dose or concentration of a substance at which no observable effects are detected in an experimental study, and these effects are not significantly different from the responses of control groups, as determined by the author's reported statistical test. The endpoint of LOEL is used for terrestial database, and is defined as the abbreviation for Lowest-Observable-Effect-Level. In the context of ECOTOX_CodeAppendix, it refers to the lowest dose or concentration of a substance at which observable effects are detected in an experimental study.

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)
```

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

```r
#It is factor, not a date

Litter$collectDate <- as.Date(Litter$collectDate)
#Since the default setting for R to converse to the Date format is %Y-%m-%d or %Y/%m/%d,
#here I just simply use this function.

class(Litter$collectDate)
```

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

```r
#Now the collectDate vector is confirmed as Date

unique(Litter$collectDate)
```

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

```r
#Aug 2 and 30 were sampled in August 2018
```

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$namedLocation)
```

```
##  [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
```

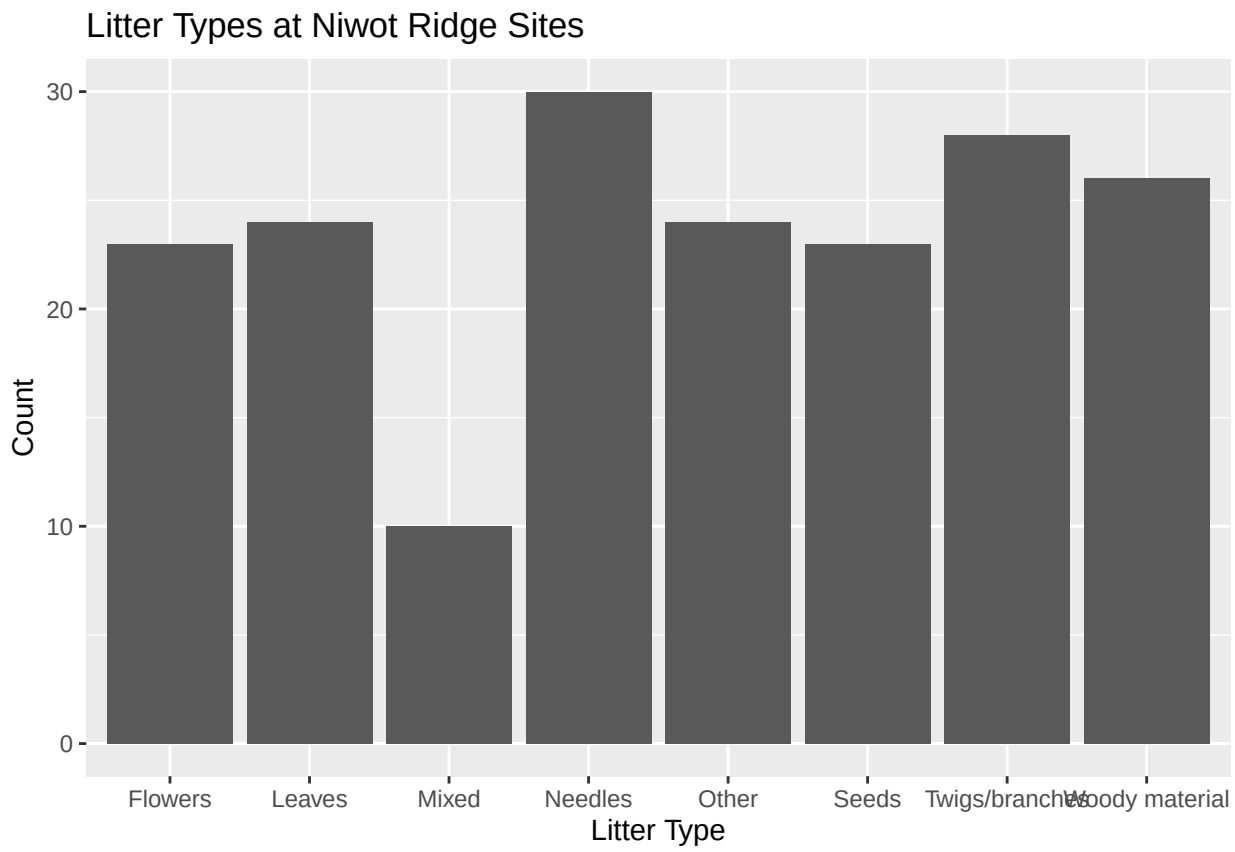
```r
summary(Litter$namedLocation)
```

```
## 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
```

Answer: In total, there are 12 plots sampled at Niwot Ridge. For the **unique** function, it only extract the sampled plots appeared in this **namedLocation** vector, while the **summary** function list all of them with the order of the most to the least with respective frequency. However, the **unique** function does mention that there are 12 levels in this vector but I could not totally understand it.

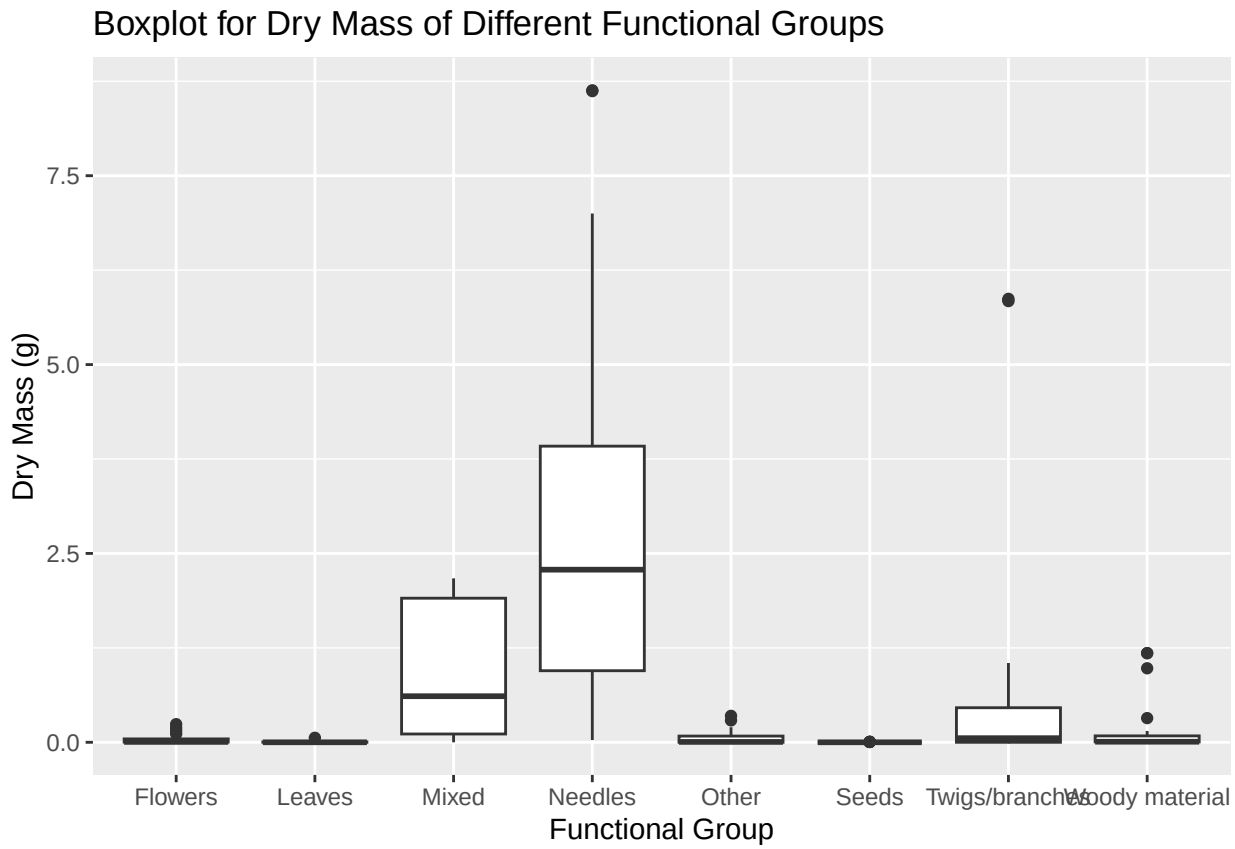
14. Create a bar graph of functionalGroup counts. This shows you what type of litter is collected at the Niwot Ridge sites. Notice that litter types are fairly equally distributed across the Niwot Ridge sites.

```
ggplot(Litter)+
  geom_bar(aes(x = functionalGroup))+
  labs(title = "Litter Types at Niwot Ridge Sites",
       x = "Litter Type",
       y = "Count")
```

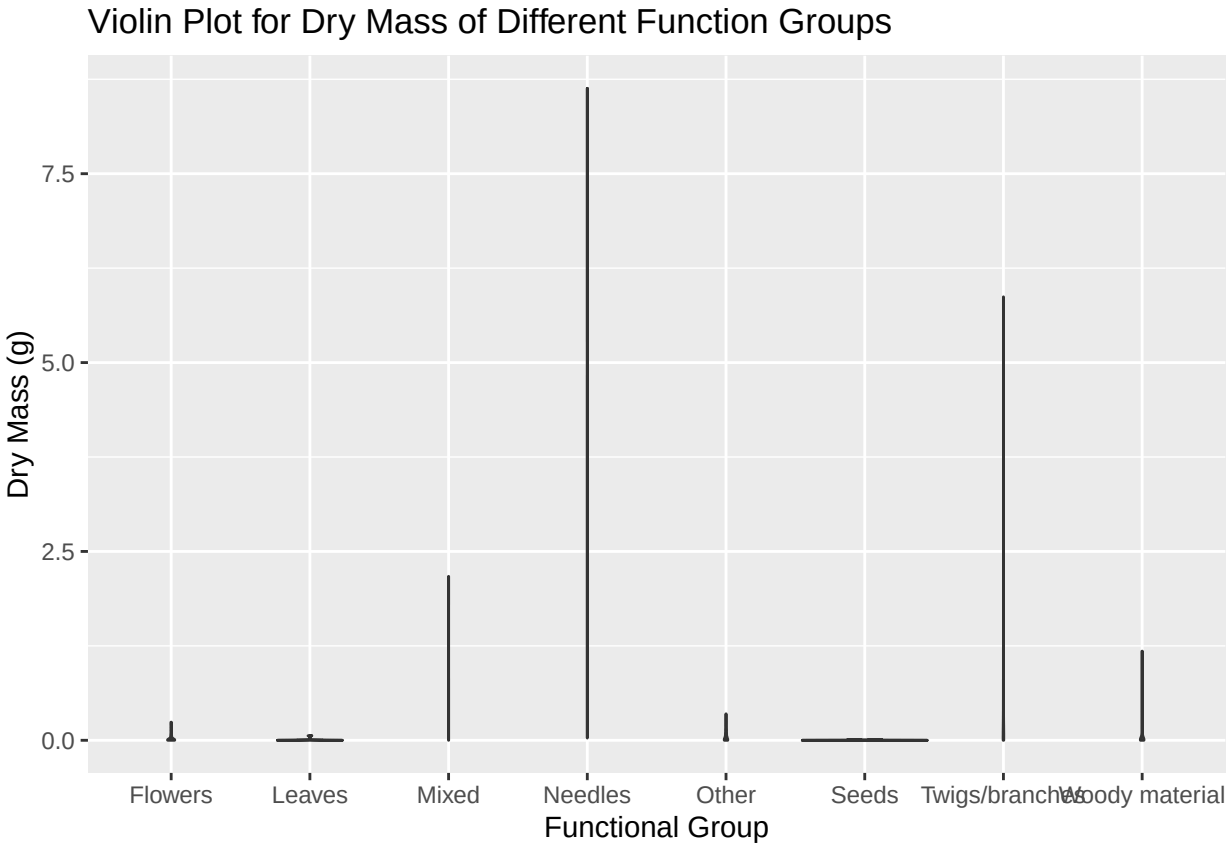


15. Using `geom_boxplot` and `geom_violin`, create a boxplot and a violin plot of dryMass by functionalGroup.

```
ggplot(Litter)+
  geom_boxplot(aes(x=functionalGroup, y=dryMass)) +
  labs(title = "Boxplot for Dry Mass of Different Functional Groups",
       x = "Functional Group",
       y = "Dry Mass (g)")
```




```
ggplot(Litter)+
  geom_violin(aes(x=functionalGroup, y=dryMass)) +
  labs(title = "Violin Plot for Dry Mass of Different Function Groups",
       x = "Functional Group",
       y = "Dry Mass (g)")
```



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

> Answer: Boxplot is simpler and more straightforward than violin plot in providing a clear representation of the median and quartiles in a concise manner, and in highlighting outliers through individual points beyond the whiskers.

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

> Answer: Needless. The interquartile range represented by the box are rather big and it has the highest median. Meanwhile, the whiskers extend greatly among all functional groups. Additionally, there's even a outlier point which has the highest weight.