Assignment 3: Data Exploration

Yuechen Huang

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.

```
library(tidyverse)
```

```
## -- 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
```

```
library(lubridate)

Neonics <- read.csv('D:/ENV872_DataExploration/ENV872_DataExploration_Fall2023/Data/Raw/ECOTOX_Neonicot
Litter <- read.csv('D:/ENV872_DataExploration/ENV872_DataExploration_Fall2023/Data/Raw/NEON_NIWO_Litter
```

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 specificity of these neonicotinoids is very important. If these neonicorinoids have a high specificity, other insects (that have important ecological impacts) can still be alive. Human and other mammals will also be safe. The knowing the ecotoxicology helps us to understand the specificity and protect our food and water resources for the current and future generations.

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: Woody debris has important roles in carbon recycle and can provide habitat to terrestrial and aquatic creatures. Litter debris, however, can be a source of plastic pollution. These litter debris can affect soil quality, negatively influence human and other animals' health. Therefore, studying litter and woody debris is important to forest ecology and health of different creatures.

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. Litter and fine woody debris sampling is executed at terrestrial NEON sites that contain woody vegetation >2m tall. 2. 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 2 weeks) in deciduous forest sites during senescence, and infrequent year-round sampling (1x every 1-2 months) at evergreen sites. 3. In sites with forested tower airsheds, the litter sampling is targeted to take place in 20 40m x 40m plots.

Obtain basic summaries of your data (Neonics)

5. What are the dimensions of the dataset?

```
colnames(Neonics)
```

```
##  [1] "CAS.Number"                         "Chemical.Name"
##  [3] "Chemical.Grade"                     "Chemical.Analysis.Method"
##  [5] "Chemical.Purity"                    "Species.Scientific.Name"
##  [7] "Species.Common.Name"                "Species.Group"
##  [9] "Organism.Lifestage"               "Organism.Age"
## [11] "Organism.Age.Units"               "Exposure.Type"
## [13] "Media.Type"                       "Test.Location"
## [15] "Number.of.Doses"                  "Conc.1.Type..Author."
## [17] "Conc.1..Author."                  "Conc.1.Units..Author."
## [19] "Effect"                           "Effect.Measurement"
## [21] "Endpoint"                         "Response.Site"
## [23] "Observed.Duration..Days."         "Observed.Duration.Units..Days."
## [25] "Author"                           "Reference.Number"
## [27] "Title"                            "Source"
## [29] "Publication.Year"                 "Summary.of.Additional.Parameters"
```

```
# column names refer to the dimensions of dataset
```

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: Population seems to be of interest as 1803 studies are population studies and mortality is the second (1493). The reason why population study is the most popular catagory may be that population study is the basic study for ecotoxicity. Most of the studies need to be done in population level to confirm the toxicity level before further research.

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), decreasing = TRUE) # sort the output of species common names
```

```
##                            (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: Honey bees are the most popular research subject because honey bees are ecologically important making them easy to transport pollutants. They are resilient to environmental stress. References: Cunningham MM, Tran L, McKee CG, et al. Honey bees as biomonitors of environmental contaminants, pathogens, and climate change. Ecol Ind. 2022;134:108457. https://www.sciencedirect.com/science/article/pii/S1470160X21011225. doi: 10.1016/j.ecolind.2021.108457.

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.)
```

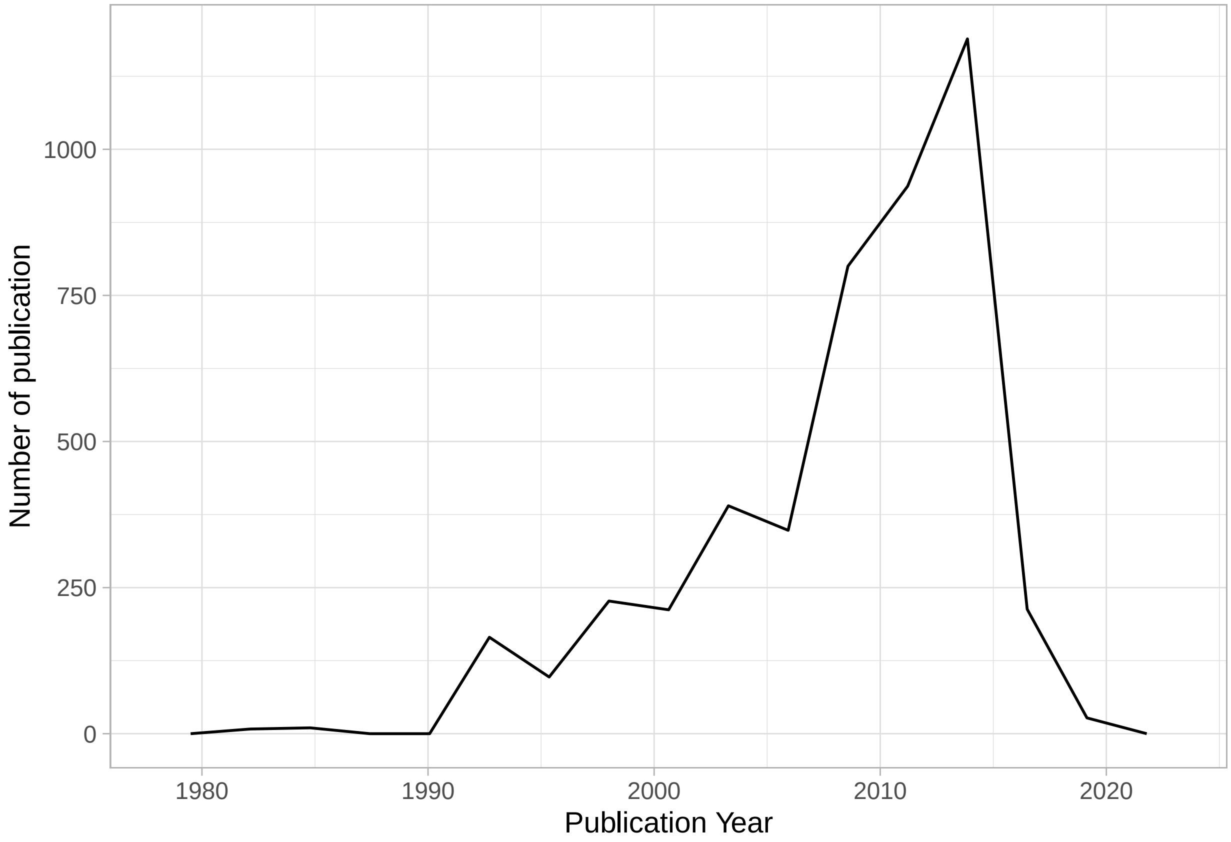
```
## [1] "factor"
```

Answer: The class of `Conc.1..Author.` column is factor because some of the values are 'NR' and some values contains '/'.

Explore your data graphically (Neonics)

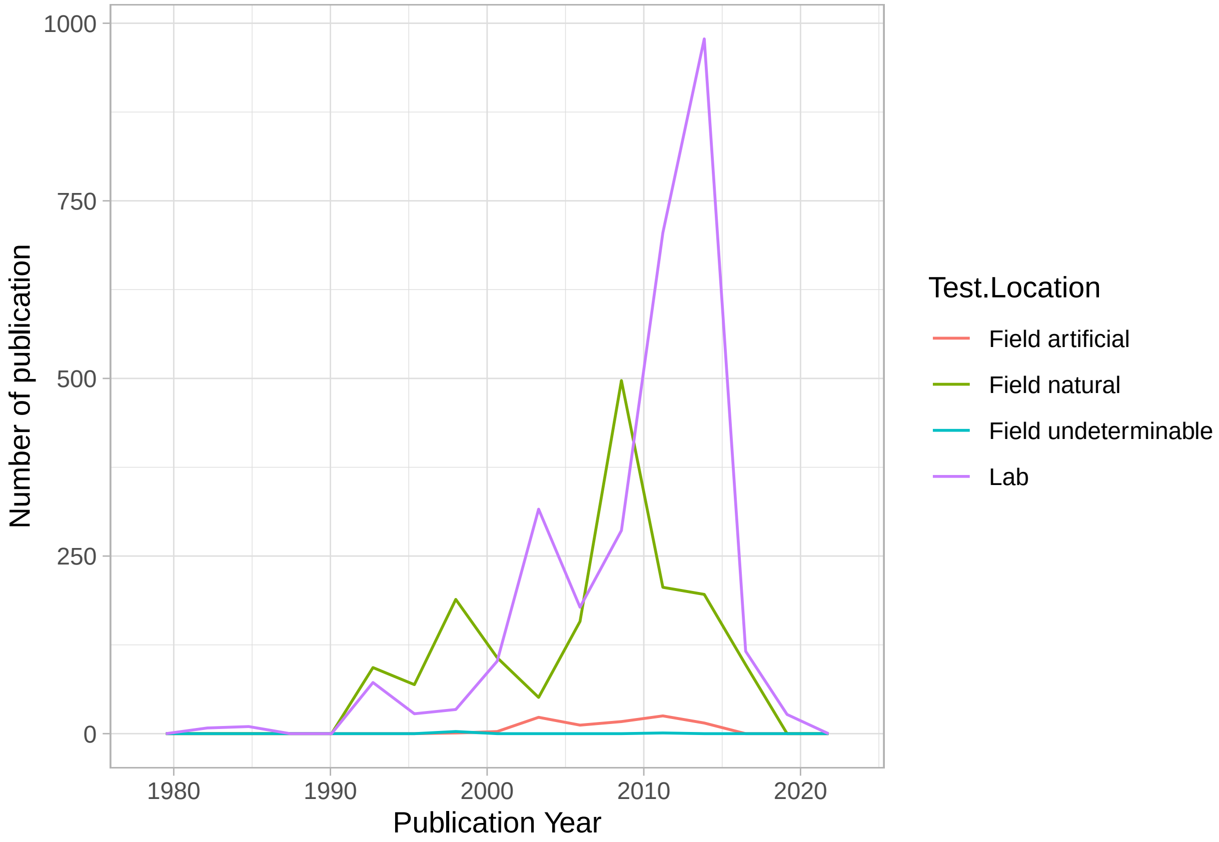
9. Using `geom_freqpoly`, generate a plot of the number of studies conducted by publication year.

```
plot1_NeoFreq <- ggplot(Neonics) +
  geom_freqpoly(aes(x = Publication.Year), bins = 15) + theme_light() + labs(x = 'Publication Year', y =
plot1_NeoFreq
```



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

```
plot2_NeoFreqColor <- ggplot(Neonics) +
  geom_freqpoly(aes(x = Publication.Year, color = Test.Location), bins = 15) + theme_light() + labs(x =
plot2_NeoFreqColor
```



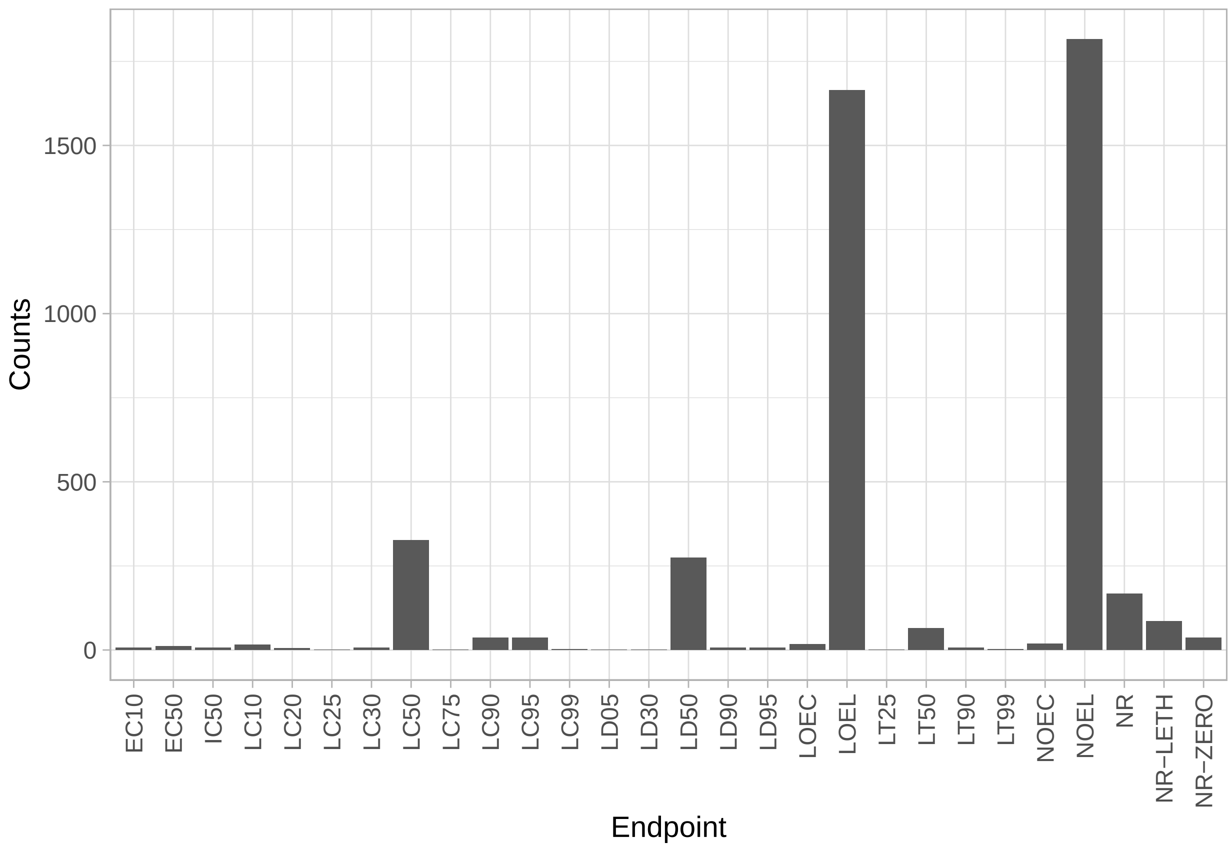
Interpret this graph. What are the most common test locations, and do they differ over time?

> Answer: The most common test locations are lab and field natural. Lab experiements have a trend of increasing from 1980 to around 2014 and decreasing afterwards. Similarly, the number of natual field experiments increases and decreases. The peak of the publications of natural field experiement is around 2009.

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…]

```
plot3_NeoBar <- ggplot(Neonics) +
  geom_bar(aes(x = Endpoint)) + theme_light() + theme(axis.text.x = element_text(angle = 90, vjust = 0.
plot3_NeoBar
```



Answer: NOEL and LOEL are the two most common end points. NOEL (No-observable-effect-level) and LOEL (Lowest-observable-effect-level) are common Endpoints for terrestrial database.

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) # the class of collectDate is factor
```

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

```
Litter$collectDate <- as.Date(Litter$collectDate, Format = "%Y-%m-%d") # change factor into date
class(Litter$collectDate) # check the class again
```

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

```
unique(Litter$collectDate) # to see which dates litter was sampled
```

```
## [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$fieldSampleID) # to see how many unique values are in the column fieldSampleID
```

```
##  [1] NEON.LTR.NIWO061169.20180802 NEON.LTR.NIWO064103.20180802
##  [3] NEON.LTR.NIWO067017.20180802 NEON.LTR.NIWO040205.20180802
##  [5] NEON.LTR.NIWO041059.20180802 NEON.LTR.NIWO063062.20180802
##  [7] NEON.LTR.NIWO047197.20180802 NEON.LTR.NIWO051045.20180802
##  [9] NEON.LTR.NIWO058101.20180802 NEON.LTR.NIWO046155.20180802
## [11] NEON.LTR.NIWO062050.20180802 NEON.LTR.NIWO040205.20180830
## [13] NEON.LTR.NIWO041059.20180830 NEON.LTR.NIWO047197.20180830
## [15] NEON.LTR.NIWO051045.20180830 NEON.LTR.NIWO058101.20180830
## [17] NEON.LTR.NIWO063062.20180830 NEON.LTR.NIWO046155.20180830
## [19] NEON.LTR.NIWO062050.20180830 NEON.LTR.NIWO061169.20180830
## [21] NEON.LTR.NIWO064103.20180830 NEON.LTR.NIWO057081.20180830
## [23] NEON.LTR.NIWO067017.20180830
## 23 Levels: NEON.LTR.NIWO040205.20180802 ... NEON.LTR.NIWO067017.20180830
```

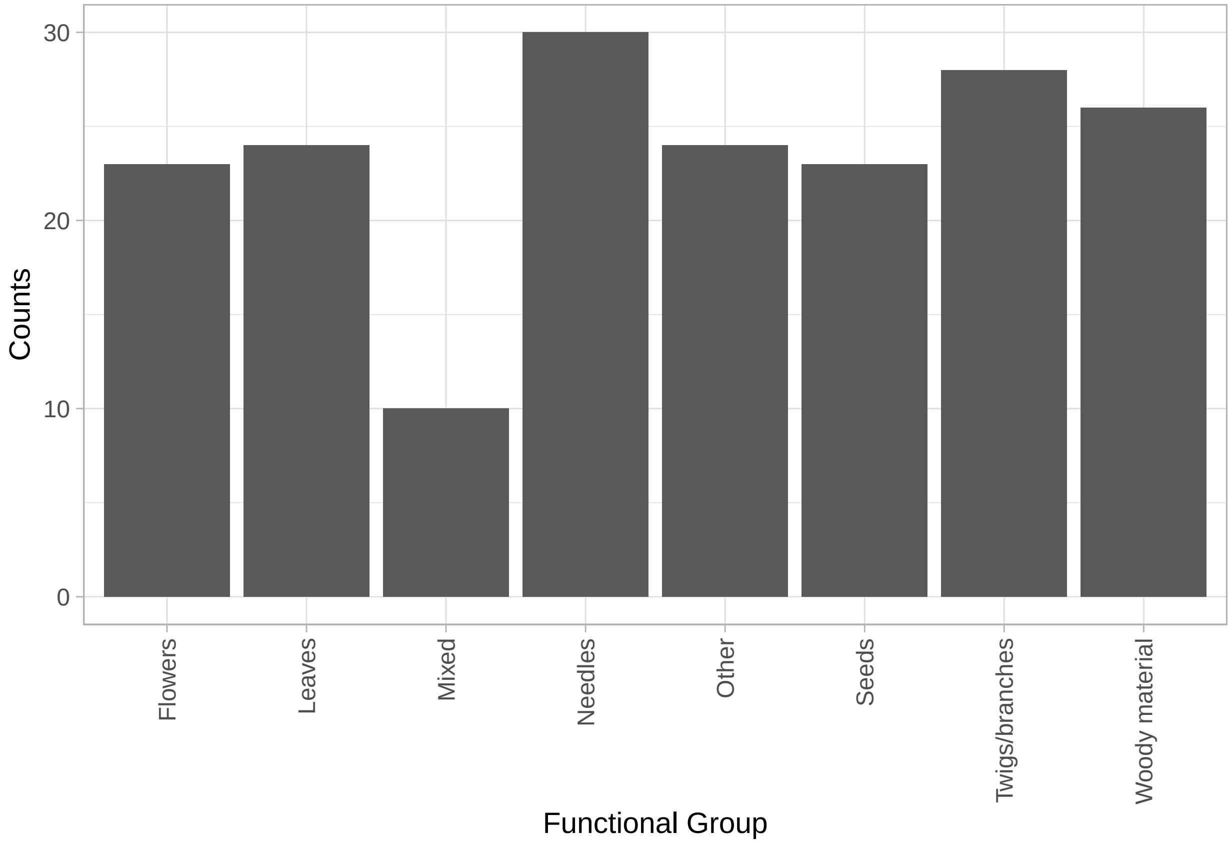
```
summary(Litter$fieldSampleID)
```

```
## NEON.LTR.NIWO040205.20180802 NEON.LTR.NIWO040205.20180830
##                           10                           10
## NEON.LTR.NIWO041059.20180802 NEON.LTR.NIWO041059.20180830
##                            8                           11
## NEON.LTR.NIWO046155.20180802 NEON.LTR.NIWO046155.20180830
##                           10                            8
## NEON.LTR.NIWO047197.20180802 NEON.LTR.NIWO047197.20180830
##                            8                            7
## NEON.LTR.NIWO051045.20180802 NEON.LTR.NIWO051045.20180830
##                            7                            7
## NEON.LTR.NIWO057081.20180830 NEON.LTR.NIWO058101.20180802
##                            8                            9
## NEON.LTR.NIWO058101.20180830 NEON.LTR.NIWO061169.20180802
##                            7                            9
## NEON.LTR.NIWO061169.20180830 NEON.LTR.NIWO062050.20180802
##                            8                            7
## NEON.LTR.NIWO062050.20180830 NEON.LTR.NIWO063062.20180802
##                            7                            7
## NEON.LTR.NIWO063062.20180830 NEON.LTR.NIWO064103.20180802
##                            7                            8
## NEON.LTR.NIWO064103.20180830 NEON.LTR.NIWO067017.20180802
##                            8                            8
## NEON.LTR.NIWO067017.20180830
##                            9
```

Answer: Total of 23 plots were sampled at Niwot Ridge. Summary() returns the sample sites but also how many results are in the same site while unique() only returns the name of each unique sample site.

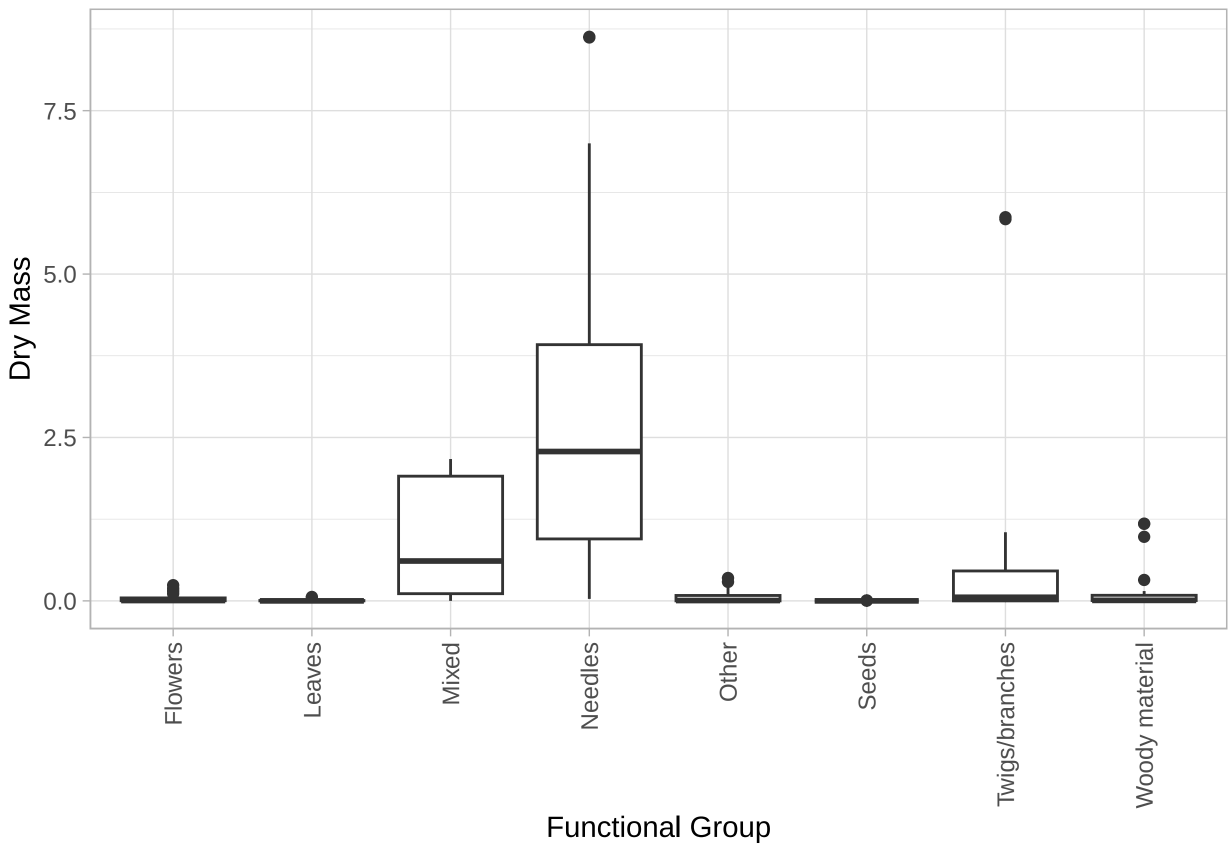
14. Create a bar graph of functionalGroup counts. This shows you what type of litter is collected at the Niwot Ridge sites. Notice that litter types are fairly equally distributed across the Niwot Ridge sites.

```
plot4_litterBar <- ggplot(Litter) +
  geom_bar(aes(x = functionalGroup)) + theme_light() + theme(axis.text.x = element_text(angle = 90, vju
plot4_litterBar
```

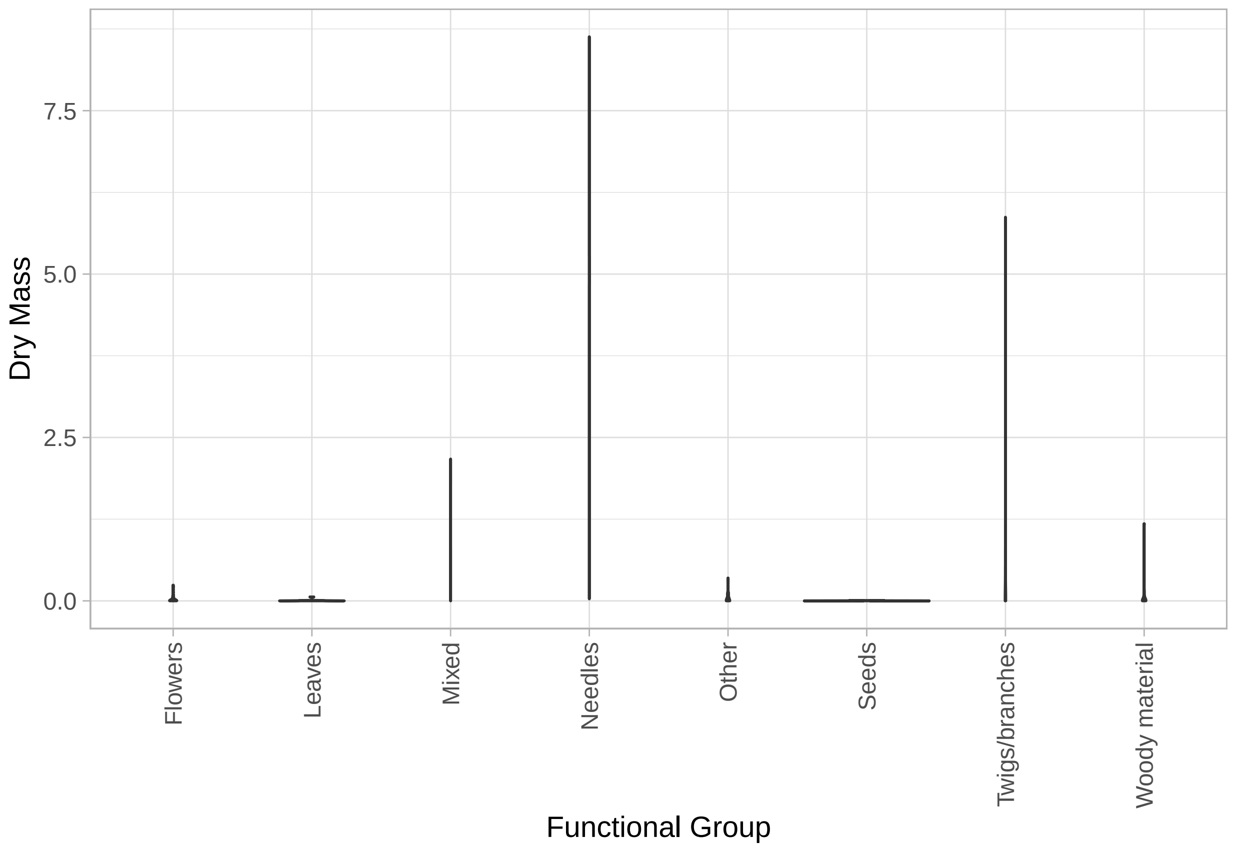


15. Using `geom_boxplot` and `geom_violin`, create a boxplot and a violin plot of dryMass by functionalGroup.

```
# box plot
plot5_litterBox <- ggplot(Litter) +
  geom_boxplot(aes(x = functionalGroup, y = dryMass)) + theme_light() + theme(axis.text.x = element_tex
plot5_litterBox
```



```
# violin plot
plot6_litterVio <- ggplot(Litter) +
  geom_violin(aes(x = functionalGroup, y = dryMass)) + theme_light() + theme(axis.text.x = element_text
plot6_litterVio
```



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

> Answer: Violin plots may not work well when the sample size is small or there are multiple peaks rather than having a unimodal distribution. Therefore, in our case, the boxplot can be a more effection option which can also clearly show outliers.

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

> Answer: Needles