Assignment 3: Data Exploration

Chenjia Liu

Fall 2023

OVERVIEW

This exercise accompanies the lessons in Environmental Data Analytics on Data Exploration.

Directions

1. Rename this file `<FirstLast>_A03_DataExploration.Rmd` (replacing `<FirstLast>` with your first and last name).
2. Change "Student Name" on line 3 (above) with your name.
3. Work through the steps, **creating code and output** that fulfill each instruction.
4. Assign a useful **name to each code chunk** and include ample **comments** with your code.
5. Be sure to **answer the questions** in this assignment document.
6. When you have completed the assignment, **Knit** the text and code into a single PDF file.
7. After Knitting, submit the completed exercise (PDF file) to the dropbox in Sakai.

TIP: If your code extends past the page when knit, tidy your code by manually inserting line breaks.

TIP: If your code fails to knit, check that no `install.packages()` or `View()` commands exist in your code.

Set up your R session

1. Check your working directory, load necessary packages (tidyverse, lubridate), and upload two datasets: the ECOTOX neonicotinoid dataset (ECOTOX_Neonicotinoids_Insects_raw.csv) and the Niwot Ridge NEON dataset for litter and woody debris (NEON_NIWO_Litter_massdata_2018-08_raw.csv). Name these datasets "Neonics" and "Litter", respectively. Be sure to include the subcommand to read strings in as factors.

```
getwd()
```

```
## [1] "C:/Users/15638/Desktop/DUKE/FALL2023/ENV872/EDE_Fall2023"
```

```
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
```

```r
library(lubridate)

Neonics<- read.csv("./Data/Raw/ECOTOX_Neonicotinoids_Insects_raw.csv",stringsAsFactors = TRUE)
Litter<- read.csv("./Data/Raw/NEON_NIWO_Litter_massdata_2018-08_raw.csv",stringsAsFactors = TRUE)
```

Learn about your system

2. The neonicotinoid dataset was collected from the Environmental Protection Agency's ECOTOX Knowledgebase, a database for ecotoxicology research. Neonicotinoids are a class of insecticides used widely in agriculture. The dataset that has been pulled includes all studies published on insects. Why might we be interested in the ecotoxicology of neonicotinoids on insects? Feel free to do a brief internet search if you feel you need more background information.

   Answer:Neonicotinoids are active substances used in plant protection products to control harmful insects. Neonicotinoids transfuse into pollen grains, nectar, cell sap and the food synthesized in the plants treated by them. These chemicals are highly neurotoxic in their mode of action to insects and other arthropods and act on the nicotinic acetylcholine receptor

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:Litter and Woody debris are important parts of forest and stream ecosystems because it has a role in carbon budgets and nutrient cycling, is a source of energy for aquatic ecosystems, provides habitat for terrestrial and aquatic organisms, and contributes to structure and roughness, thereby influencing 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.Liter and fine woody debris sampling is executed at terrestrial NEON sites that contain woody vegetation >2m tall 2.One litter trap pair (one elevated trap and one ground trap) is deployed,either radomly or targated denpends on the aerial percent coverage, for every 400 m2 plot area, resulting in 1-4 trap pairs per plot. 3. Ground traps are sampled once per year. Target sampling frequency for elevated traps varies by vegetation present at the site.

Obtain basic summaries of your data (Neonics)

5. What are the dimensions of the dataset?

```r
dim(Neonics)
```

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

6. Using the `summary` function on the "Effect" column, determine the most common effects that are studied. Why might these effects specifically be of interest?

```r
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"
```

```r
summary(Neonics$Effect)
```

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

Answer:The most common effects are population and mortality. We are specifically intertested in these effects because we want to know how the presense of Neonics affect the population (reduce population), and the size of population is related to mortality, which could increase as the result of Neonics.

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...]

```r
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
```

Answer:The 6 most commonly studied species in the dataset from most to least are: 1. Honey Bee 2. Parasitic Wasp 3. Buff Tailed Bumblebee 4. Carniolan Honey Bee 5. Bumble Bee 6. Italian Honeybee. We are interested in bees becuase they are the major pollinator for the agriculture, and they are greatly affected by the use of systemic pesticides such as Neonics. Our food security and resources is strongly related to the presence of bees.

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"
```

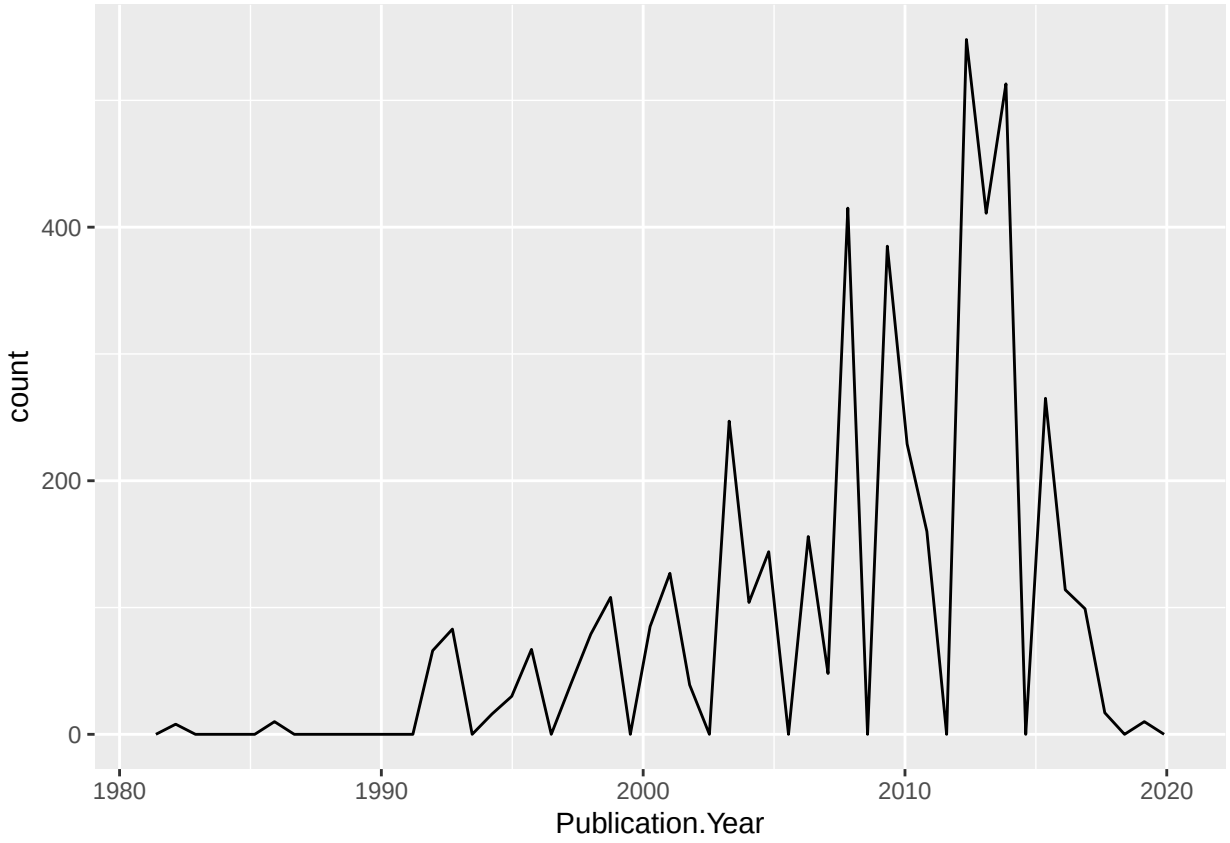
```
view(Neonics$Conc.1..Author.)
```

Answer:It is a factor. It is not numeric because there are not only number in this column, but there are other objects, such as 'NR'

Explore your data graphically (Neonics)

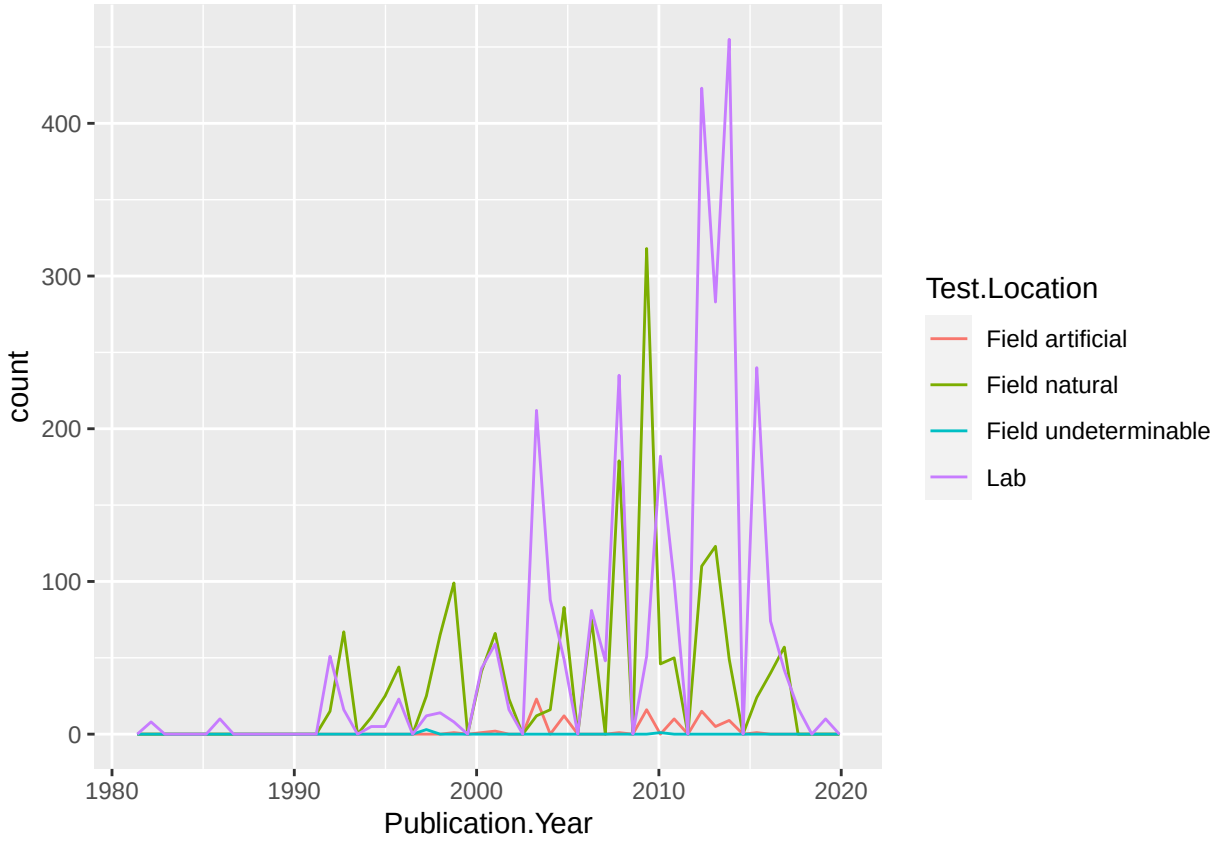
9. Using `geom_freqpoly`, generate a plot of the number of studies conducted by publication year.

```
ggplot(Neonics) +
  geom_freqpoly(aes(x = Publication.Year), bins = 50)
```



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

```
ggplot(Neonics) +
  geom_freqpoly(aes(x = Publication.Year,  color = Test.Location), bins = 50)
```



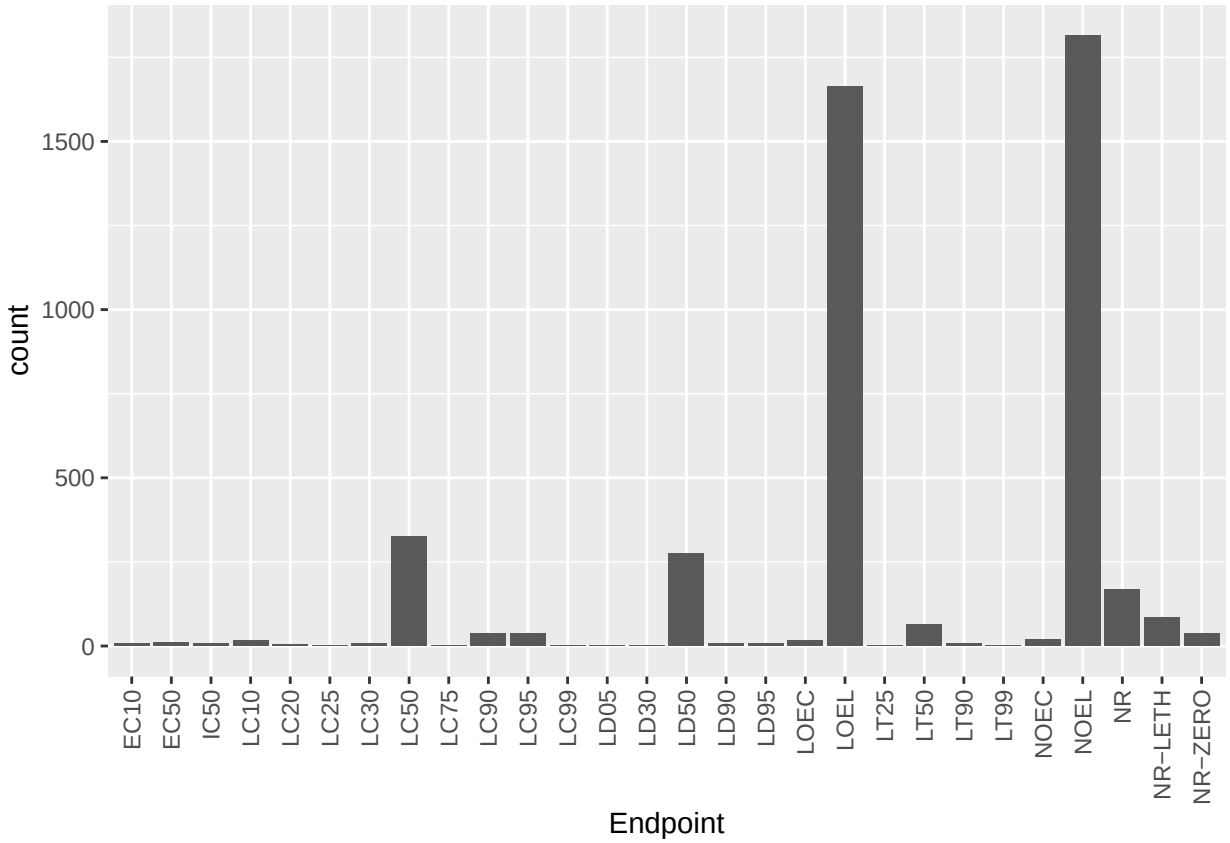
Interpret this graph. What are the most common test locations, and do they differ over time?

Answer:The most common location before 2000 is field natural. But after 2000, the lab became the most common location and the count of publication has also increase a lot. The publication reached peak between 2010 and 202.

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))
```



Answer:The most common endpoints are LOEL and NOEL. LOEL:Lowest-observable-effect-level: lowest dose (concentration) producing effects that were significantly different (as reported by authors) from responses of controls (LOEAL/LOEC) NOEL:No-observable-effect-level: highest dose (concentration) producing effects not significantly different from responses of controls according to author's reported statistical test (NOEAL/NOEC)

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$collectDate) #It's a factor
Litter$collectDate <- as.Date(Litter$collectDate, format = "%Y-%m-%d")
class(Litter$collectDate)
```

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

```
view(Litter)
unique(Litter$collectDate)
```

```
## [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$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
```

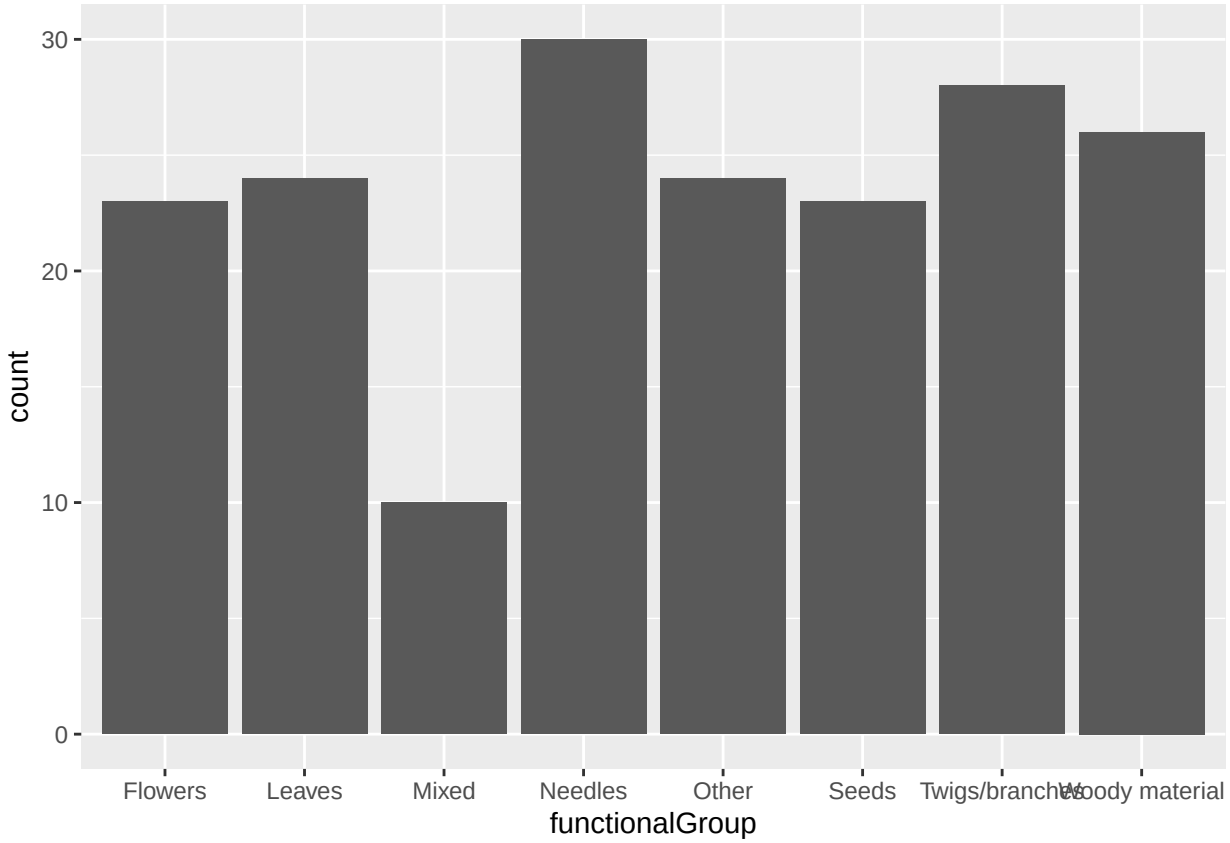
```
summary(Litter$namedLocation)
```

```
## 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:'Unique' is able to directly tell how many levels in the data, When using 'summary', we need to count number of levels but it also provide more information, ie. count of each location

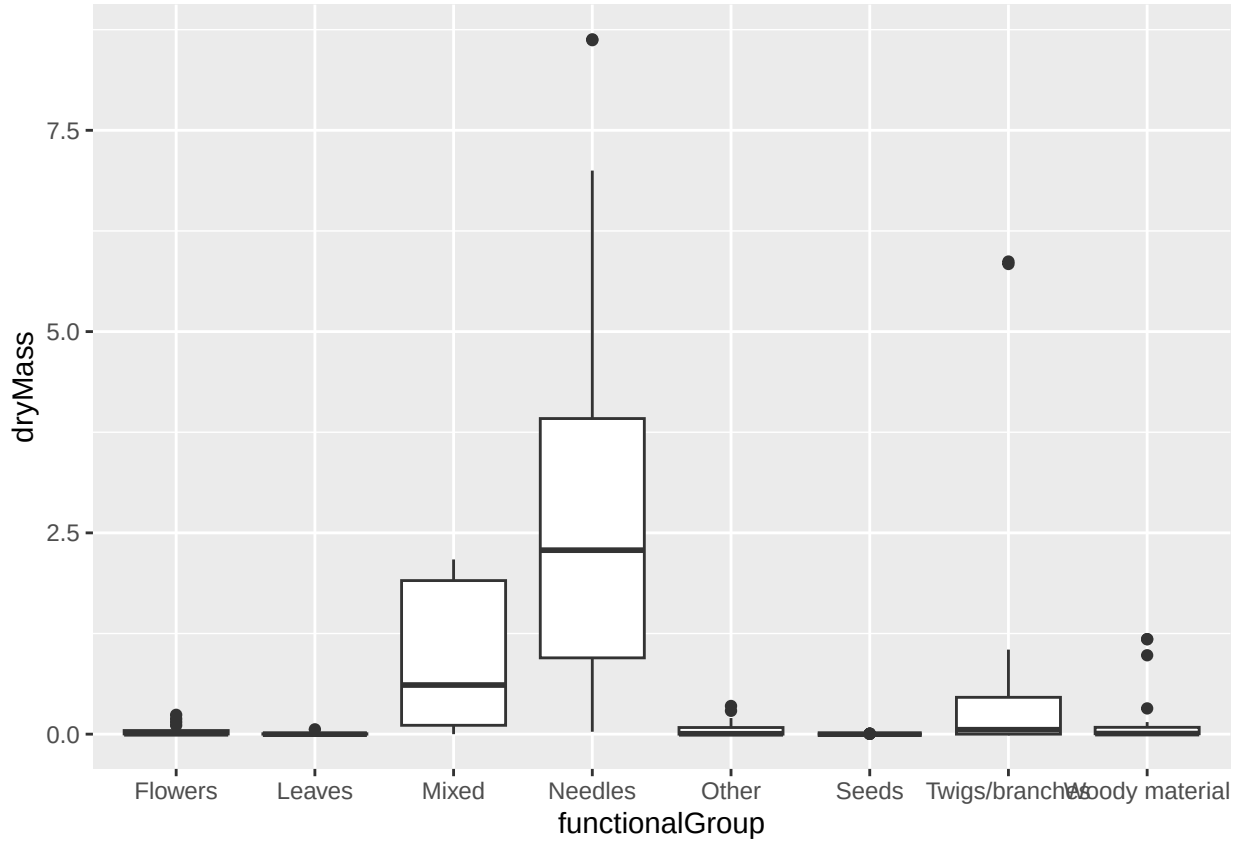
14. Create a bar graph of functionalGroup counts. This shows you what type of litter is collected at the Niwot Ridge sites. Notice that litter types are fairly equally distributed across the Niwot Ridge sites.

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

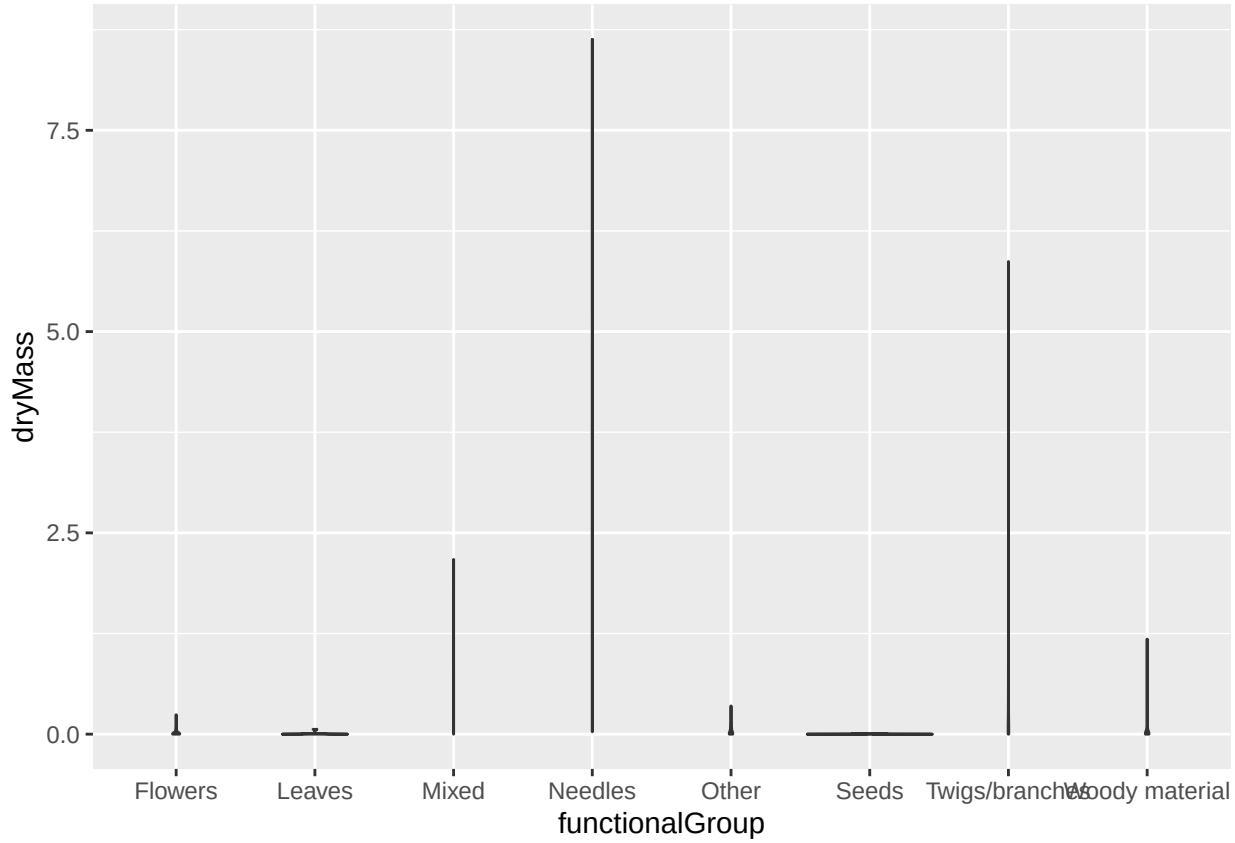


15. Using `geom_boxplot` and `geom_violin`, create a boxplot and a violin plot of dryMass by functionalGroup.

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




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



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

> Answer: In this case, only boxplot can show the IQR, median, and outliers, and the different length of the lines on both side of the box tell us the skewness of data. We can only see the range of our data from the violin plot in this case.

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

> Answer:Needles tend to have the highest biomass at these sites.