Assignment 3: Data Exploration

Shidi Dai

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. Be sure to **answer the questions** in this assignment document.
5. When you have completed the assignment, **Knit** the text and code into a single PDF file.
6. After Knitting, submit the completed exercise (PDF file) to the dropbox in Sakai.

The completed exercise is due on Sept 30th.

Set up your R session

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

```
getwd()
```

```
## [1] "/home/guest/R/EDA-Fall2022/Assignments"
```

```
#install.packages("tidyverse")
library(tidyverse)

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: It is important to improve and enhance the process of hazard assessments of ecotoxicology on ecology and human health. Thus, we are interested in learning more about the effects of the ecotoxicology of neonicotinoids on insects to evaluate its impact on the environment. Since neonicotinoid is widely used to kill insects, it might be a widespread environmental contaminants causing unexpected nontarget effects. Transmission through simple food chains might impact the entire food webs, causing unexpected effects in the environment. It might also break the biodiversity of the ecosystem. Thus, we are interested in the ecotoxicology of neonicotinoids on insects.

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 serve as the link between tree canopy and the soils beneath. They add nutrients accumulated from the biomass. They also influence forest productivity and tree growth. They play an important role in biogeochemicle cycling and tree growth, thus impacting the ecosystem of the forest. Litterfall and woody debris data may be used to estimate annual Aboveground Net Primary Productivity (ANPP) and aboveground biomass at plot, site, and continental scales. They also provide essential data for understanding vegetative carbon fluxes over time. Therefore, we are interested in studying litter and woody debris for environmental and ecological purposes.

4. How is litter and woody debris sampled as part of the NEON network? Read the NEON_Litterfall_UserGuide.pdf document to learn more. List three pieces of salient information about the sampling methods here:

   Answer: Litter and woody febris are collected from elevated and ground traps. All of those are from the spatial resolution of a single trap and the temporal resolution of a single collection event. Mass data for each collection event are measured separately to an accuracy of 0.01 grams. 1.Litter is collected in elevated 0.5m^2 PVC traps, and woody debris is collected in groud traps as longer material is not reliably collected by the elevated traps. 2.Using the spatial sampling design, litter and woody debris sampling occurs in tower plots that are selected randomly with forested tower airsheds or with low-statured vegetation over the tower airsheds. Trap placement within plots are targeted or randomized depending on the vegetation. 3. Under temporal sampling design, there is a time frequency for sampling. Ground traps are sampled once per year and elevated traps varies by vegetation present at the site.

Obtain basic summaries of your data (Neonics)

5. What are the dimensions of the dataset?

```
dim(Neonics) #4623 observations of 30 variables
```

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

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

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

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

```r
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: Most common effects: Population 1803, Mortality 1493, Behavior 360. Population is measurements and endpoints relating to a group of organisms or plants of the same species occupying the same area at a given time. Mortality is measurements and endpoints where the cause of death is by direct action of the chemical. These effects are mostly be of interest because they represent the most of how neonicotinoids are having an impact in the ecosystem, other species, and environment.

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.

```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), 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 (excluding others 670) are honey bee 667, parasitic wasp 285, buff tailed bumblebee 183, carniolan honey bee 152, bumble bee 140, and Italian honeybee 113. They are mostly bees. Bees are of interest over other insects because they have the ability to spread neonicotinoids/or the effect of ecotoxicology while spreading pollens. They are also essential insects in reproductive of the ecosystem.

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

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

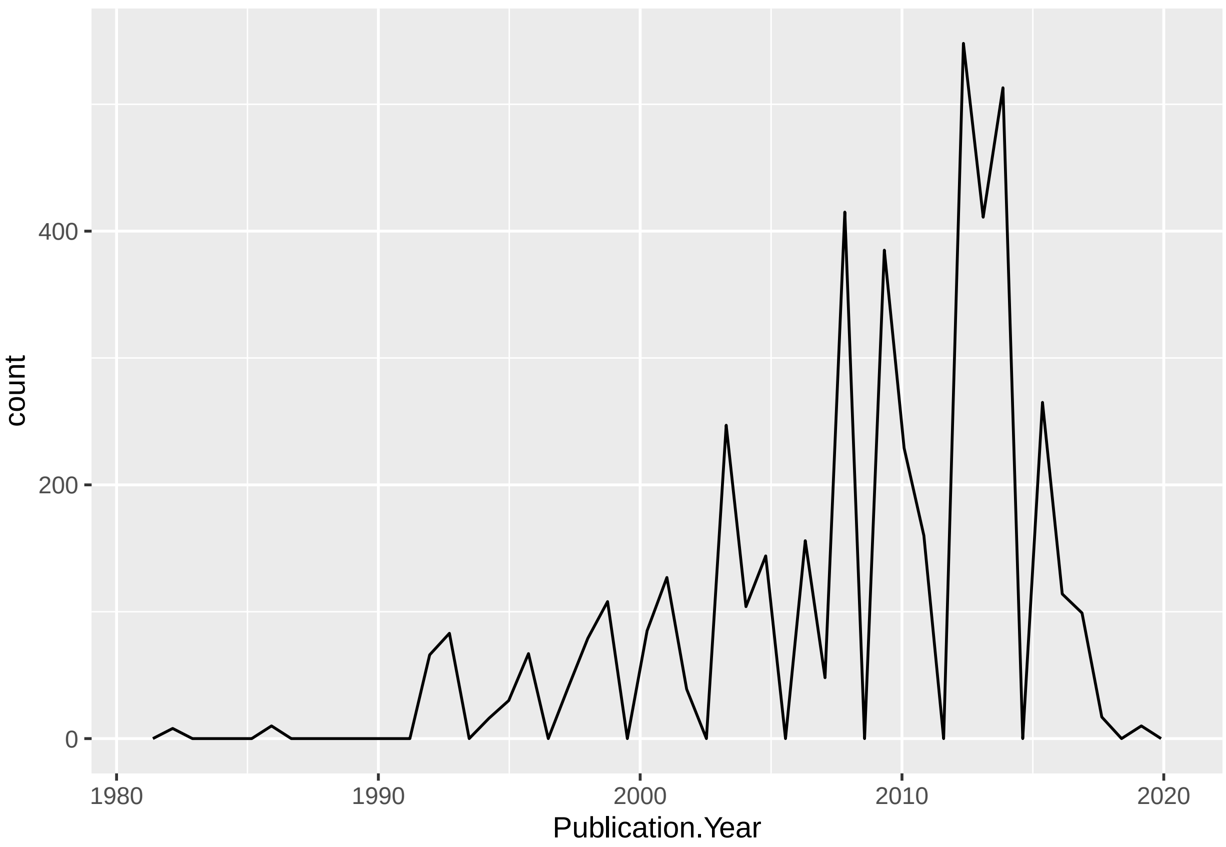
```
## [1] "factor"
```

Answer: It is a factor. Factors are used to represent categorical data. They are stored as integers and have labels associated with these unique integers. Factors can only contain a pre-defined set values, known as levels. In this case, there are 1006 levels, representing different categories.

Explore your data graphically (Neonics)

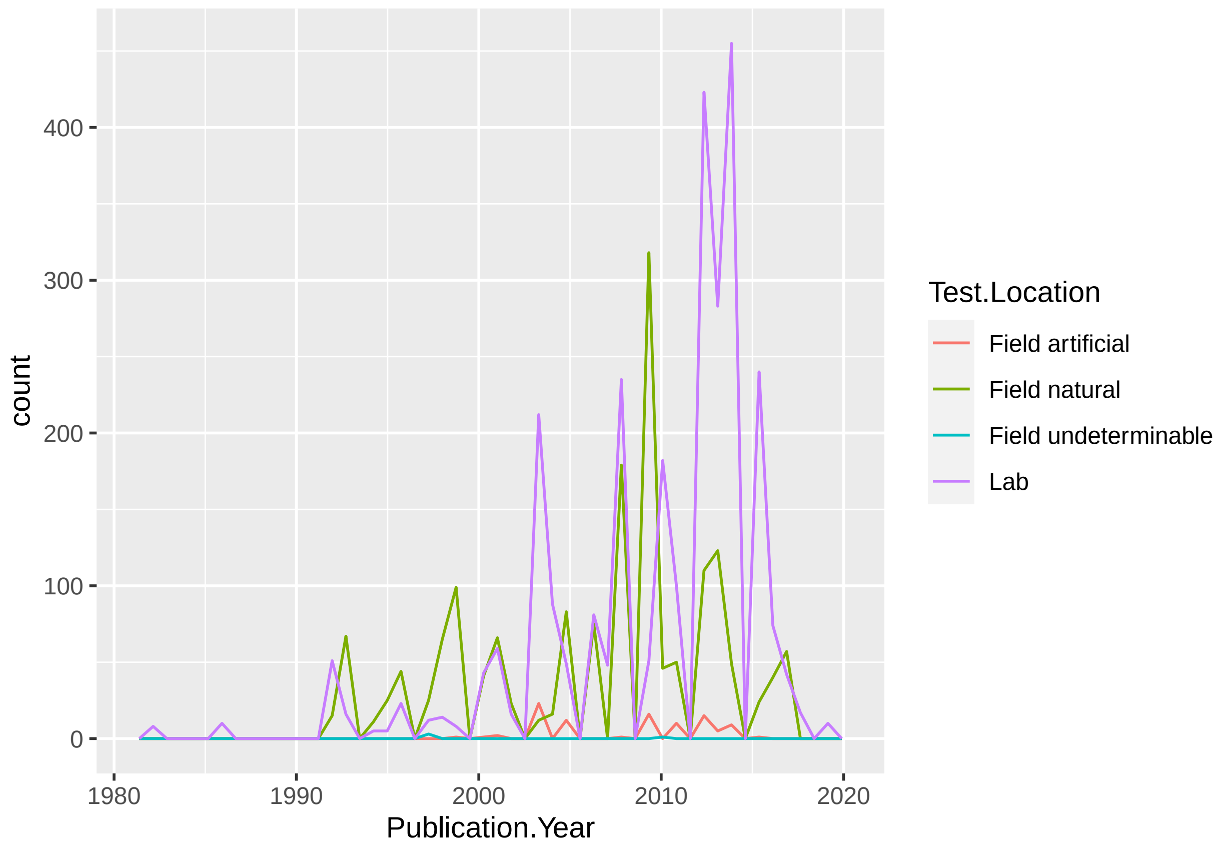
9. Using `geom_freqpoly`, generate a plot of the number of studies conducted by publication year.

```
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)
```

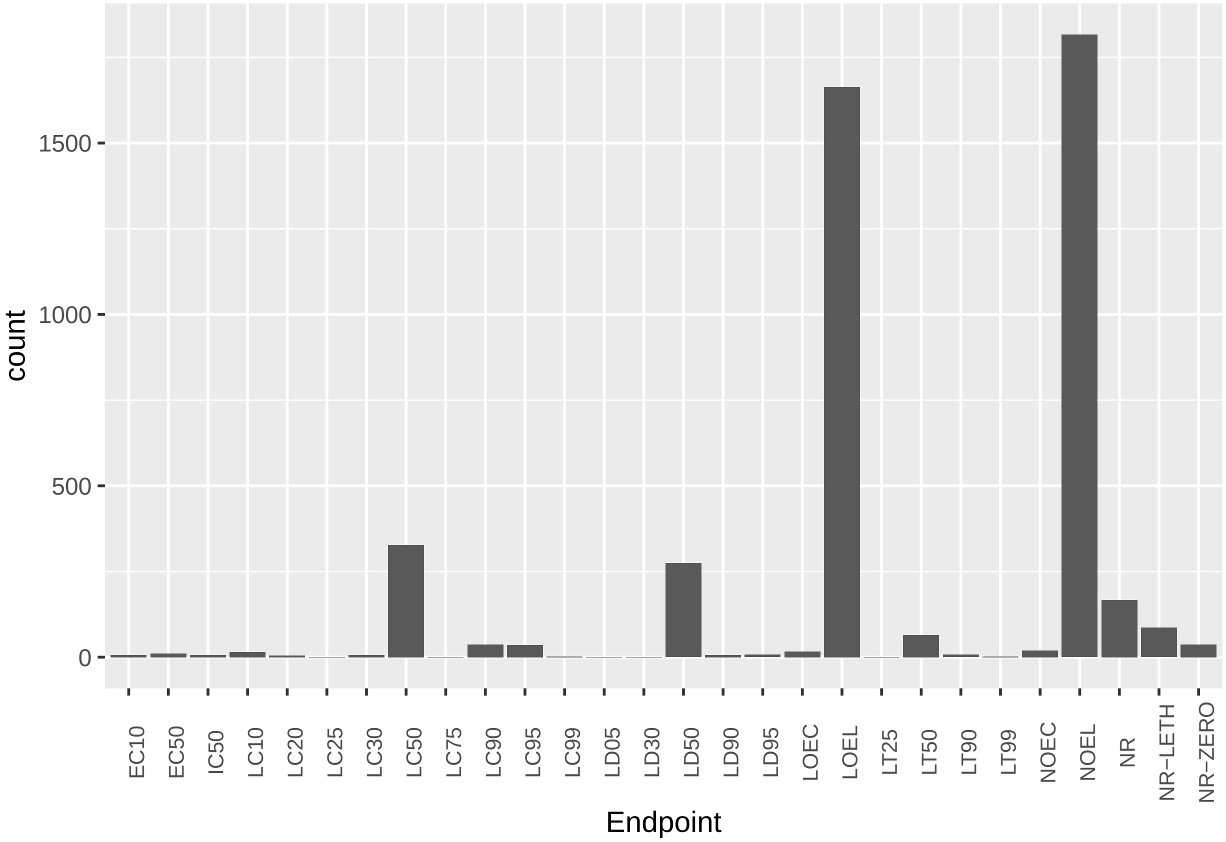


Interpret this graph. What are the most common test locations, and do they differ over time?

> Answer: The most common test location is lab. Field natural were the most common location for publication studies in the 1990s, early 2000s, and before 2010. Lab was most common before 1990, increased to become the most common location starting around 2003, and reached its maximum between 2010 and 2015. Relative much fewer studies were done in field artificial, which mostly occured between 2003 and 2015.

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.

```
ggplot(Neonics, aes(x = Endpoint)) +
  geom_bar()+
  theme(axis.text.x = element_text(angle = 90, size = 8))
```



Answer: The two most common endpoints are NOEL and LOEL. NOEL is No-observable-effect-level, defined as highest dose (concentration) producing effects not significantly different from responses of controls according to author's reported statistical test. LOEL is Lowest-observable-effect-level, defined as lowest dose (concentration) producing effects that were significantly different as reported by authors from responses of 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) #factor not date
```

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

```
Litter$collectDate <- as.Date(Litter$collectDate, format = "%Y-%m-%d")
unique(Litter$collectDate) #Litter was sampled on Aug. 2 and Aug. 30 in 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$plotID) #12 plots were sampled at Niwot Ridge.
```

```
##  [1] NIWO_061 NIWO_064 NIWO_067 NIWO_040 NIWO_041 NIWO_063 NIWO_047 NIWO_051
##  [9] NIWO_058 NIWO_046 NIWO_062 NIWO_057
## 12 Levels: NIWO_040 NIWO_041 NIWO_046 NIWO_047 NIWO_051 NIWO_057 ... NIWO_067
```

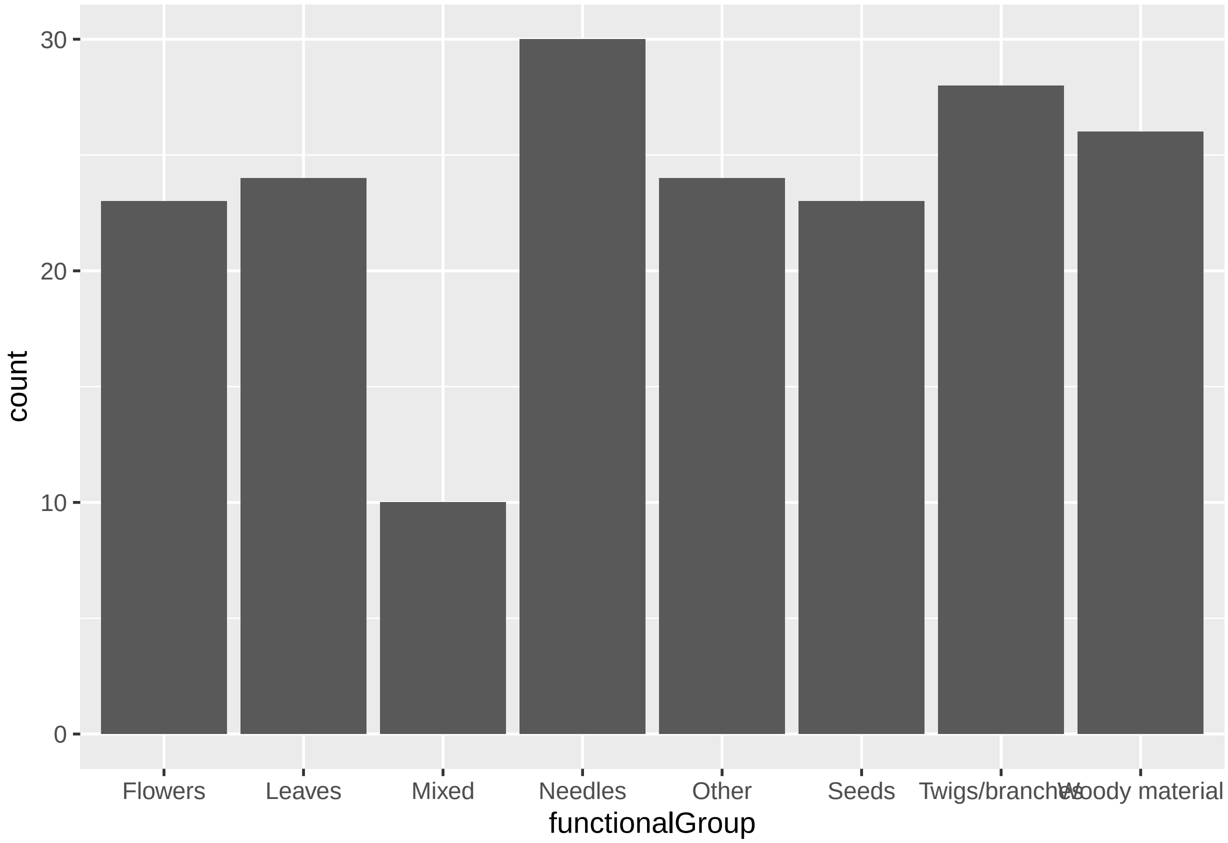
```
summary(Litter$plotID)
```

```
## NIWO_040 NIWO_041 NIWO_046 NIWO_047 NIWO_051 NIWO_057 NIWO_058 NIWO_061
##       20       19       18       15       14        8       16       17
## NIWO_062 NIWO_063 NIWO_064 NIWO_067
##       14       14       16       17
```

Answer: 12 plots were sampled at Niwot Ridge. Summary tells us how many samples we took in each of the 12 plots. Unique eliminates duplicate values (plots) and shows us a list of different plots we took samples from.

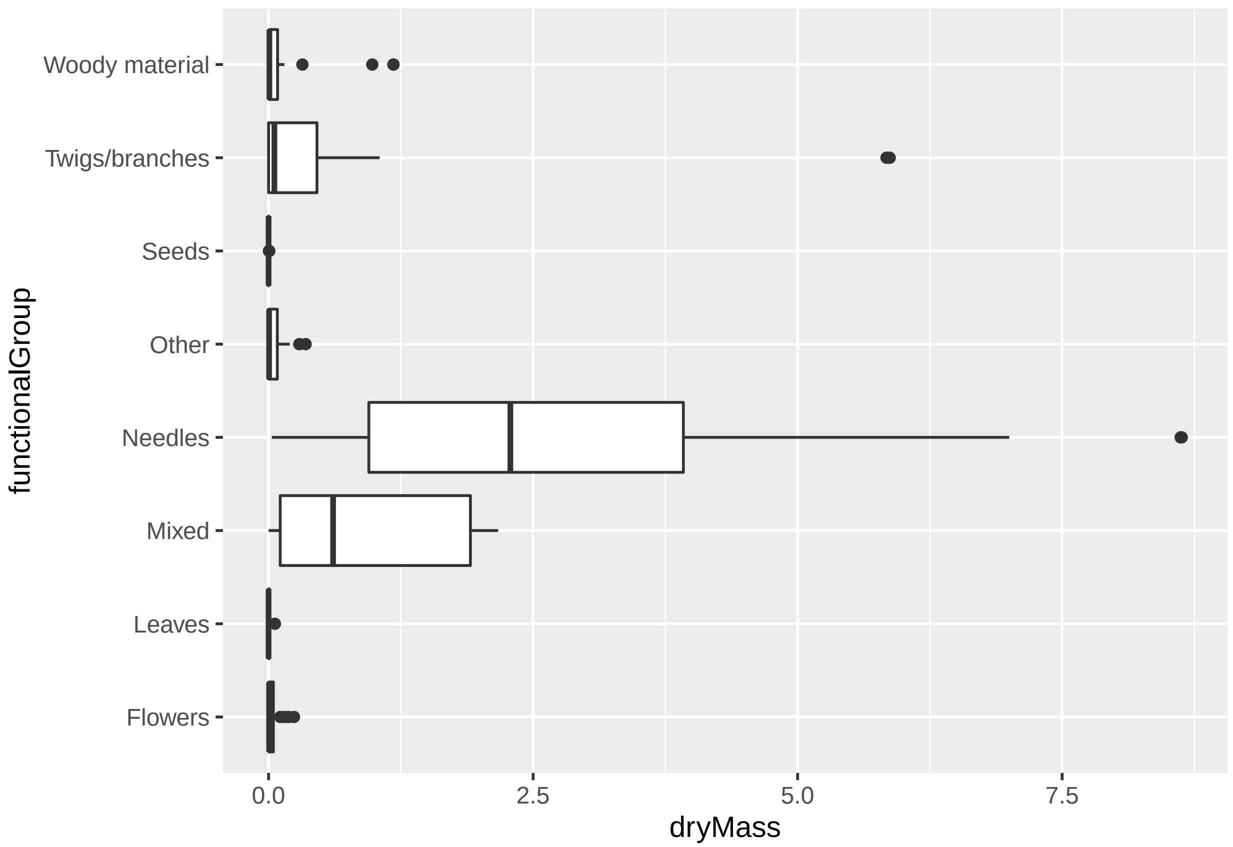
14. Create a bar graph of functionalGroup counts. This shows you what type of litter is collected at the Niwot Ridge sites. Notice that litter types are fairly equally distributed across the Niwot Ridge sites.

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

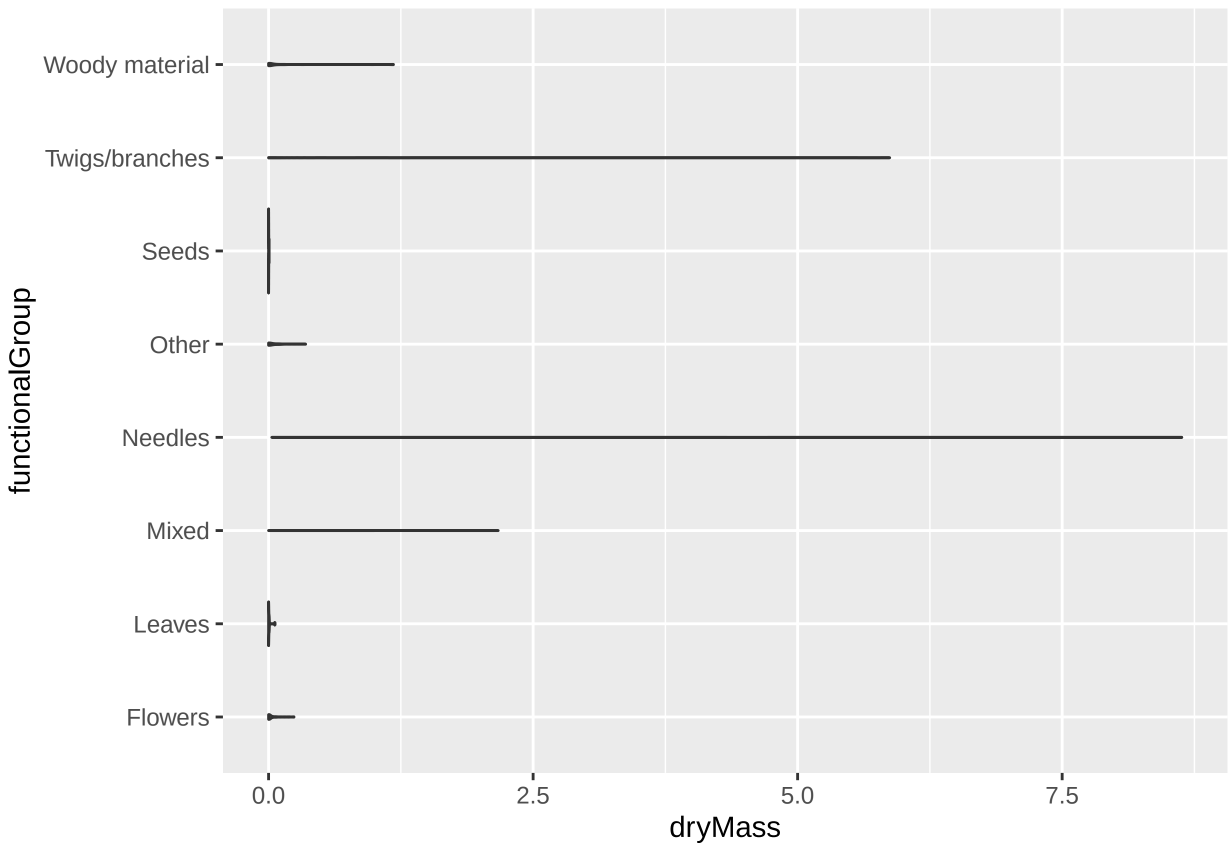


15. Using `geom_boxplot` and `geom_violin`, create a boxplot and a violin plot of dryMass by functionalGroup.

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



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



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

> Answer: A violin plot shows a kernel density estimation on the distribution. However, in this case, the data for groups are either together giving very high density at certain points, or having outliers very far from the main group resulting in density "lines" showing in the violin plot. On the other hand, boxplot gives us clearly the mean, 1st and 3rd quartile, and outlier information while plotting the distribution. Therefore, the violin plot is not an effective visualization option.

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

> Answer: Needles tend to have the highest biomass (highest dry mass and highest average dry mass) at these sites.