Assignment 3: Data Exploration

Nikki Shintaku

OVERVIEW

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

Directions

1. Change "Student Name" on line 3 (above) with your name.
2. Work through the steps, **creating code and output** that fulfill each instruction.
3. Be sure to **answer the questions** in this assignment document.
4. When you have completed the assignment, **Knit** the text and code into a single PDF file.
5. After Knitting, submit the completed exercise (PDF file) to the dropbox in Sakai. Add your last name into the file name (e.g., "Salk_A03_DataExploration.Rmd") prior to submission.

The completed exercise is due on Tuesday, January 28 at 1:00 pm.

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.

```
getwd()
```

```
## [1] "/Users/nikkishintaku/Desktop/Environmental872/Environmental_Data_Analytics_2020"
```

```
library(knitr)
opts_knit$set(root.dir = "/Users/nikkishintaku/Desktop/Environmental872/Environmental_Data_Analytics_20
```

```
library(tidyverse)

#reading in Neonics and Litter datasets
Neonics <- read.csv("./Data/Raw/ECOTOX_Neonicotinoids_Insects_raw.csv")
Litter <- read.csv("./Data/Raw/NEON_NIWO_Litter_massdata_2018-08_raw.csv")
```

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: If companies are using neonicotinoids as insecticides on agriculture, then we need to know the ecotoxicology of them in order to assess if the neonicotinoids are going to be harmful to us when we eat the crops they are put on. We also would need to make sure that it is targeting the correct insect and not harming other animals.

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: Studying litter and woody debris that falls to the ground is important data that can be used to estimate annual aboveground net primary productivity and aboveground biomass in the region. It can also help to understand vegetative carbon fluxes over time in the particular ecoclimate domain. Litter and woody debris hold an important role in carbon budgets and nutrient cycling.

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: *Sampling locations were selected randomly and sampling occurred in tower plots. In sites with forested tower airsheds, there was 20 40mX40m plots. In sites with low-statured vegetation, there was 4 40mx40m plots and 26 20mx20m plots.*
   There was one elevated and one ground trap was deployed for every 400m^2 plot area. Trap placements within plots were targeted or randomized, depending on vegetation *Ground traps were sampled once per year. Elevated traps sampling frequency varied by vegetation present at the site with 1x every 2 weeks in deciduous forest and 1x every 1-2 months in evergreen sites.

Obtain basic summaries of your data (Neonics)

5. What are the dimensions of the dataset?

```
#dimensions of Neonics
dim(Neonics)
```

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

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

```
#summary of only the Effects column of the Neonics dataset
summary(Neonics$Effect) #effect group
```

```
##     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
summary(Neonics$Effect.Measurement) #effect and measurement
```

```
##                                            Abundance
##                                                 1699
##                                            Mortality
##                                                 1294
##                                             Survival
##                                                  133
##                               Progeny counts/numbers
##                                                  120
##                                     Food consumption
##                                                  103
##                                            Emergence
##                                                   98
##                       Search/explore/forage behavior
##                                                   96
##                            Feeding behavior, general
##                                                   92
##                                   Chemical avoidance
##                                                   65
##                                               Weight
##                                                   48
##            Distance moved, change in direct movement
##                                                   38
##                                     Feeding behavior
##                                                   36
##                                      Flying behavior
##                                                   30
##                Accuracy of learned task, performance
##                                                   28
##                                            Sex ratio
##                                                   27
##                                            Fecundity
##                                                   26
##                                   Stimulus avoidance
##                                                   26
##                                    Righting response
##                                                   24
##                                             Lifespan
##                                                   23
##                                        Acquired task
##                                                   22
##                                                Hatch
##                                                   21
##                                   Predatory behavior
##                                                   21
##                                 Acetylcholinesterase
##                                                   20
##                                                 Walk
##                                                   19
##                                   Freezing behavior
##                                                  18
##                      Reproductive success (general)
##                                                  17
##        Slowed, Retarded, Delayed or Non-development
##                                                  17
##                                            Grooming
##                                                  16
##                                            Diameter
##                                                  14
##                                             Residue
##                                                  12
##                                   Activity, general
##                                                  11
##                                      Food avoidance
##                                                  11
##                                             Control
##                                                   9
##                      Developmental changes, general
##                                                   9
##                          Intrinsic rate of increase
##                                                   9
##                                    Pollen collected
##                                                   9
##                                                Size
##                                                   9
##                                            Esterase
##                                                   8
##                               Intoxication, general
##                                                   8
##                         Mortality/survival, general
##                                                   8
##      Population change (change in N/change in time)
##                                                   8
##                                         Smell/Sniff
##                                                   8
##                                             Biomass
##                                                   7
##                                       Catalase mRNA
##                                                   7
##                                     Generation time
##                                                   7
##                                            Infected
##                                                   7
##                                         Orientation
##                                                   7
##                            Population doubling time
##                                                   7
##                              Population growth rate
##                                                   7
##                                        Sealed brood
##                                                   7
##                                   Vitellogenin mRNA
##                                                   7
##                                        Ali esterase 
##                                                   6 
## Apoptosis, programmed cell death, DNA fragmentation 
##                                                   6 
##                                    Carboxylesterase 
##                                                   6 
##                                            Hemocyte 
##                                                   6 
##                                           Knockdown 
##                                                   6 
##                                           Viability 
##                                                   6 
##                                          Extinction 
##                                                   5 
##                               Net Reproductive Rate 
##                                                   5 
##                                  Polyphenol oxidase 
##                                                   5 
##                                    Prey penetration 
##                                                   5 
##                                            Pupation 
##                                                   5 
##                               Reproducing organisms 
##                                                   5 
##   Amount or percent animals infested with parasites 
##                                                   4 
##              Continual reinforcement task performed 
##                                                   4 
##                                     Defensin 1 mRNA 
##                                                   4 
##                                 Diversity, Evenness 
##                                                   4 
##              Encapsulation or Melanization Response 
##                                                   4 
##                          General biochemical effect 
##                                                   4 
##                           Glutathione S-transferase 
##                                                   4 
##                       Histological changes, general 
##                                                   4 
##                                     Life expectancy 
##                                                   4 
##                         Thioredoxin peroxidase mRNA 
##                                                   4 
##                      Vanin-like protein 1-like mRNA 
##                                                   4 
##                                   Bees wax produced 
##                                                   3 
##                         Behavioral changes, general 
##                                                   3 
##                                            Catalase 
##                                                   3 
##                                       Cell turnover 
##                                                   3 
##                                           Cytochrome P-450
##                                                          3
##                                               Feeding time
##                                                          3
##                                                     Length
##                                                          3
##                                            Protein, total
##                                                          3
##                                                Respiration
##                                                          3
##                                Response time to a stimulus
##                                                          3
##                                                      Stage
##                                                          3
##                                     Time to first progeny
##                                                          3
##                                             Trehalase mRNA
##                                                          3
##                                       Alkaline phosphatase
##                                                          2
##                    Carboxylesterase clade I, member 1 mRNA
##                                                          2
##                                            Centractin mRNA
##                                                          2
##                                           Chitinase 5 mRNA
##                                                          2
##                                  Colony maintenance (bees)
##                                                          2
##                                                  COX2 mRNA
##                                                          2
##                                     Endoplasmin-like mRNA
##                                                          2
##                                          Gamete production
##                                                          2
##                                Glucose dehydrogenase 2 mRNA
##                                                          2
##                               Glucosinolate sulphatase mRNA
##                                                          2
##                          Glutathione peroxidase-like 1 mRNA
##                                                          2
##                          Glutathione peroxidase-like 2 mRNA
##                                                          2
##                                                    (Other)
##                                                         77
```

Answer: The most common effect is population and mortality with the most common effect and measurement being abundance and mortality. Using the summary function is an easy way to quickly see how many cases of each effect were found in all the studies about the effects of neonicotinoids on insects. This gives a clear idea of more specific effects the neonicotinoids have on insects and can be used to steer further research and understanding of how it can help or hurt agriculture growth.

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
```

Answer: The six most commonly studied species are Honey Bee, Parasitic Wasp, Buff Trailed Bumblbee, Carniolan Honey Bee, Bumble Bee, and Italian Honeybee. All of these species, except the parasitic wasp, is a pollinator. Pollinators are important for plants to reproduce and grow so studying the effects of neonicotinoids on these pollinators is of interest to make sure the insecticide aren't harming their presence.

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"
```

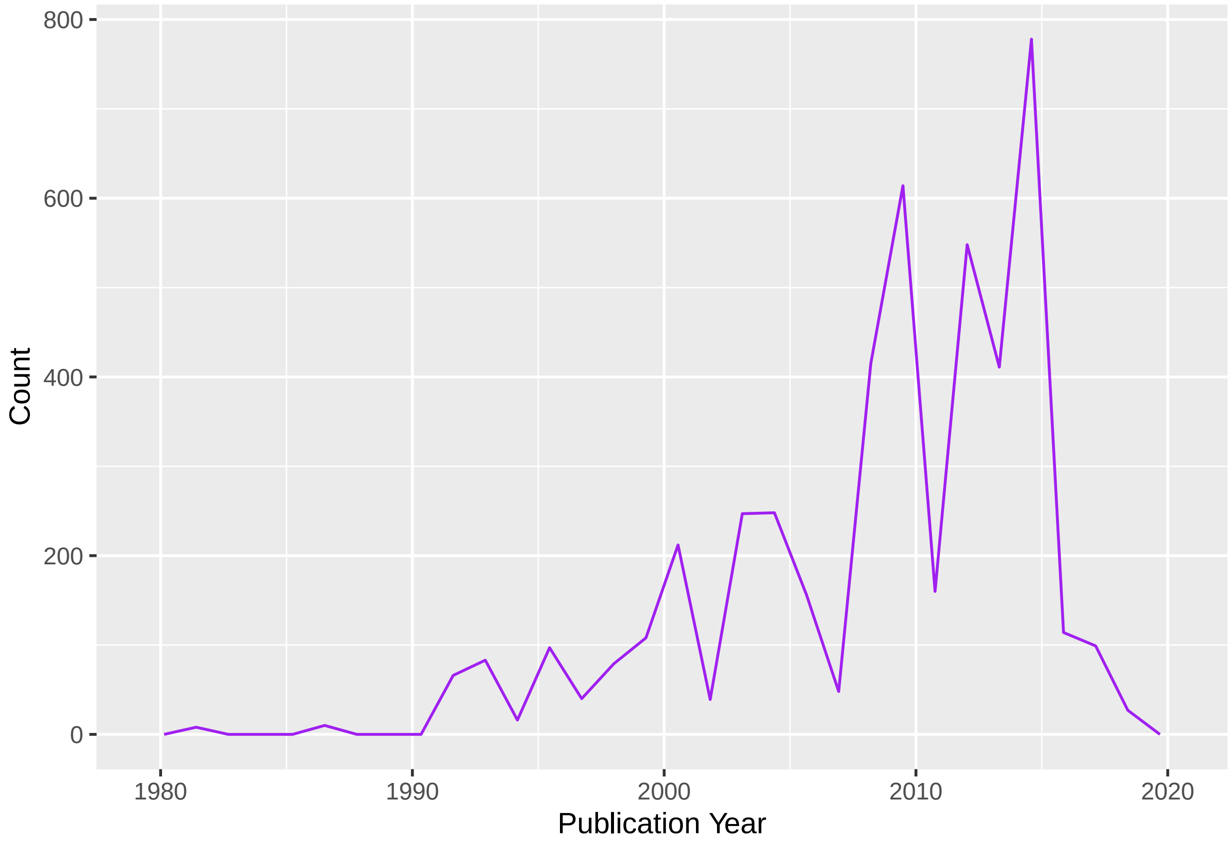
Answer: The class of Conc.1..Author is a factor because each of the numbers have a / placed after it, thus, it can not be classed as numeric.

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), color="Purple")+
  labs(x="Publication Year", y="Count")
```

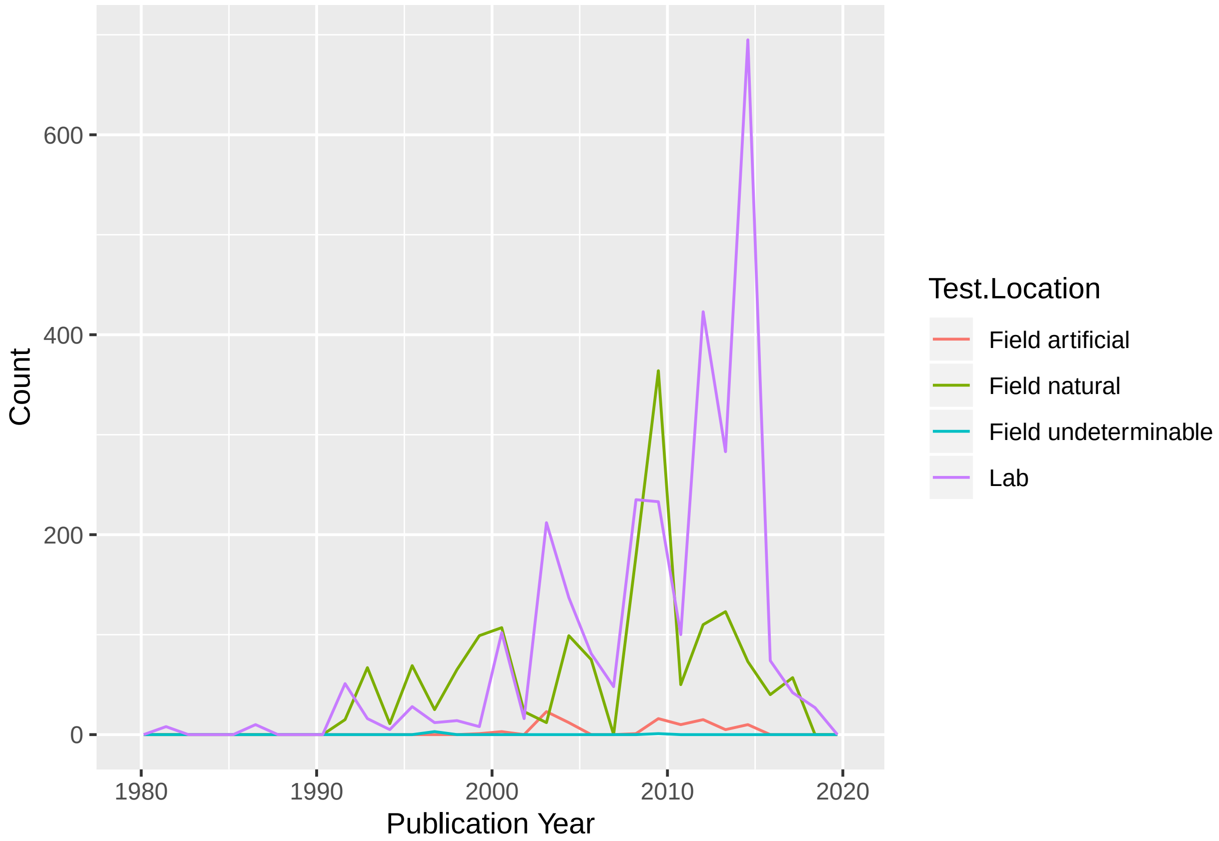
```
## `stat_bin()` using `bins = 30`. Pick better value with `binwidth`.
```



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(x="Publication Year", y="Count")
```

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

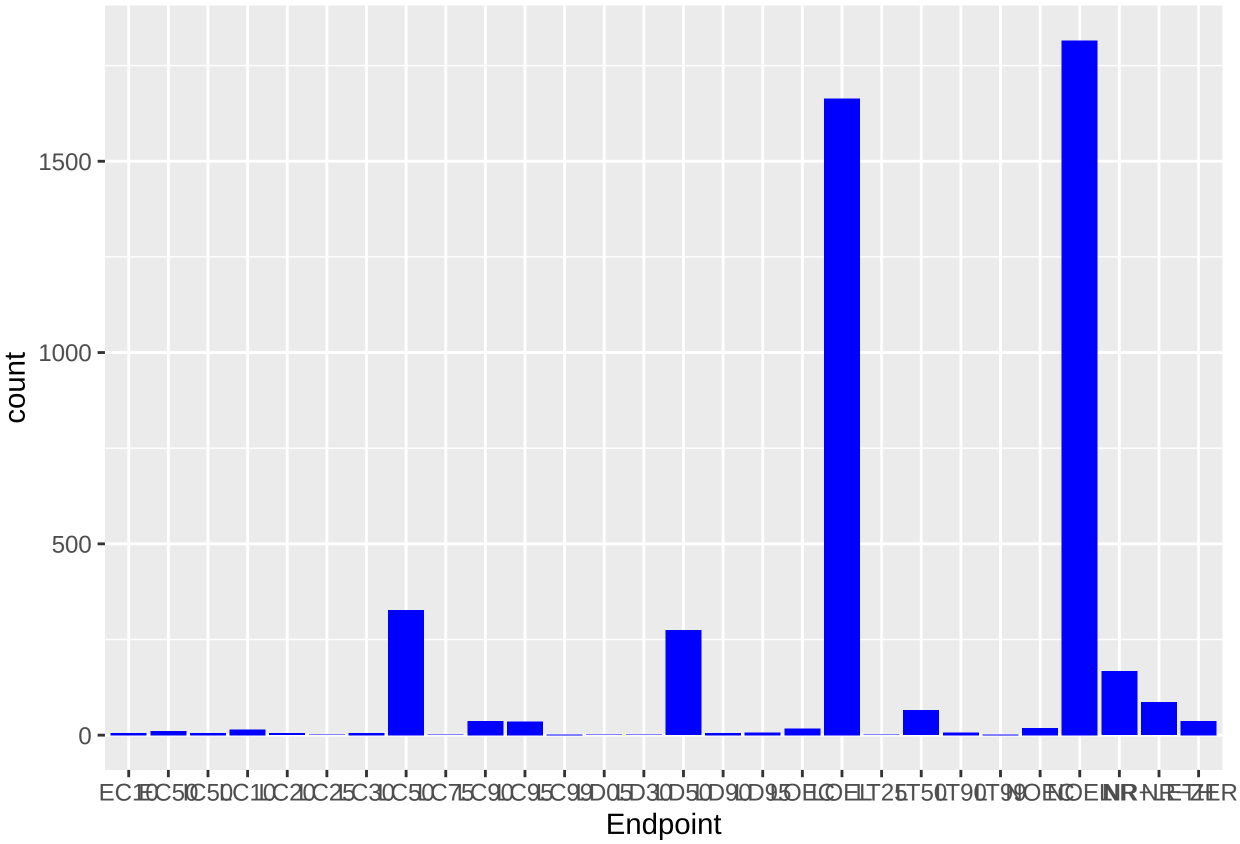


Interpret this graph. What are the most common test locations, and do they differ over time?

> Answer: The most common test location was a lab, and the second most common was in the field naturally. Between the 1990s and 2000s, field natural was used more than the lab, but as time continued into the 2010s, lab as the test location increased 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.

```
ggplot(Neonics)+
  geom_bar(aes(x = Endpoint), fill = "blue1")
```



Answer: The two most common end points are NOEL and LOEL. NOEL stands for no-observable-effect-level. This means that the higest dose producing effects are not siginificantly different from the reponses of controls. LOEL stands for lowest-observable-effect level. This means the lowest dose producing effects that were significantly different from reponse 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) #class is factor
```

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

```
#change from factor to date format
Litter$collectDate <- as.Date(Litter$collectDate, format = "%Y-%m-%d")
class(Litter$collectDate) #class is a date
```

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

```
#Use Unique function to see what dates were sampled in August 2018
august_2018 <- unique(format(Litter$collectDate, "2018-08-%d"))
august_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`?

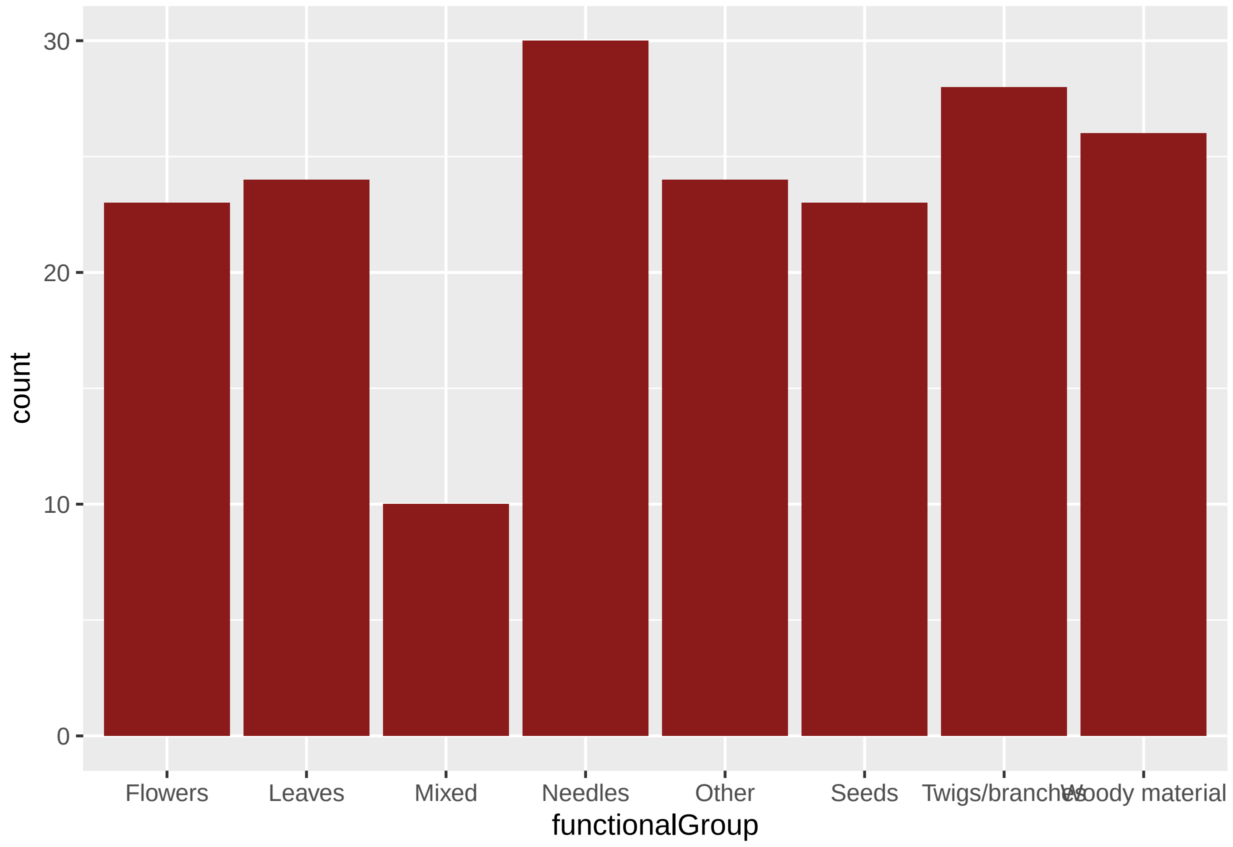
```
unique(Litter$plotID)
```

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

Answer: There were 12 plots that were sampled at Niwot Ridge. Information from Unique is different from Summary because Unique returns the names of all the different entries in the column that you want. Whereas Summary returns the number of times each of those entries comes up in the column.

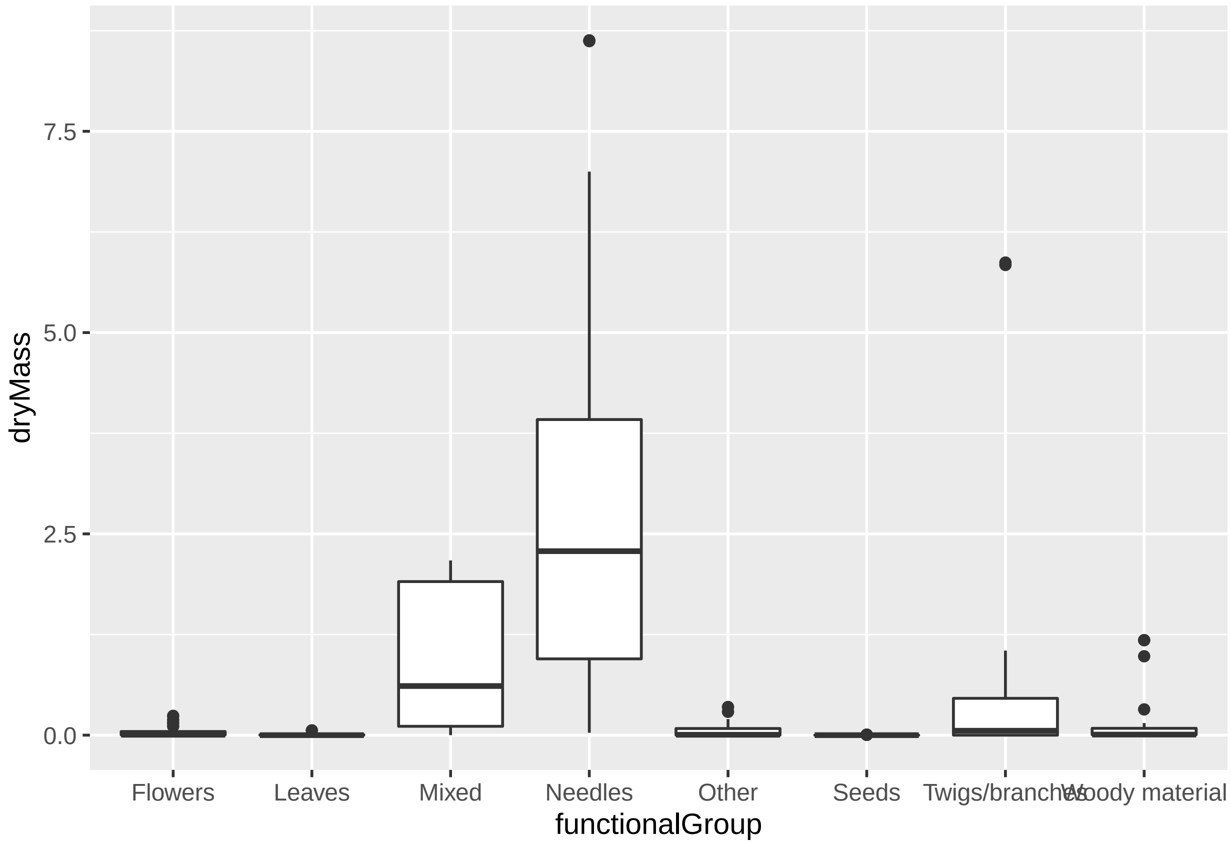
14. Create a bar graph of functionalGroup counts. This shows you what type of litter is collected at the Niwot Ridge sites. Notice that litter types are fairly equally distributed across the Niwot Ridge sites.

```
ggplot(Litter)+
  geom_bar(aes(x = functionalGroup), fill = "firebrick4")
```

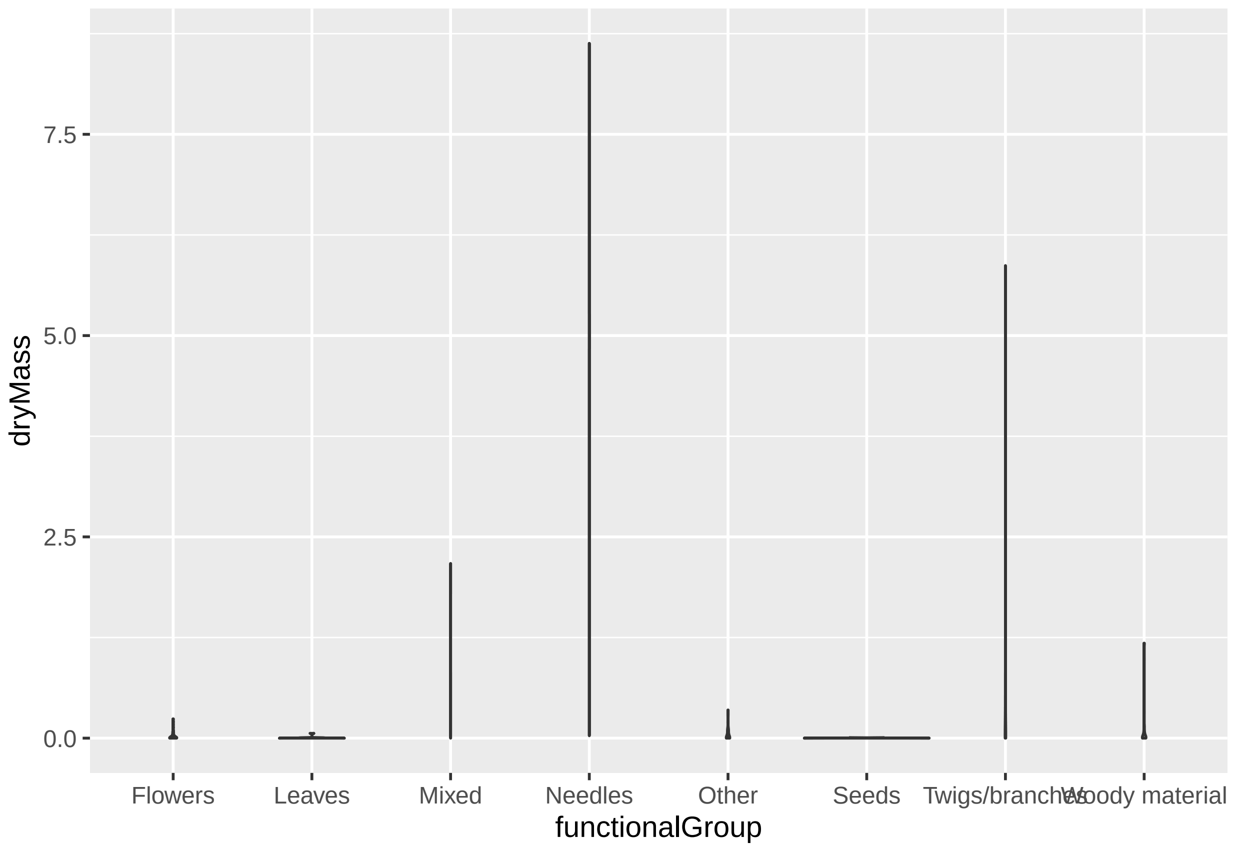


15. Using `geom_boxplot` and `geom_violin`, create a boxplot and a violin plot of dryMass by functionalGroup.

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



```
#violin
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: The violin plot does not show the distribution of the dry mass of each functional group as well as the boxplot does. The boxplot shows outliers and the skew of the data more so than the violin plot.

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

> Answer: Needles have a high biomass at these sites followed by litter that is mixed.