Assignment 3: Data Exploration

Amanda Braun

OVERVIEW

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

Directions

Change "Student Name" on line 3 (above) with your name.
Work through the steps, **creating code and output** that fulfill each instruction.
Be sure to **answer the questions** in this assignment document.
When you have completed the assignment, **Knit** the text and code into a single PDF file.
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

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/amandabraun/Documents/Classes Spring 2020/Data
Analytics/Environmental_Data_Analytics_2020"
library(tidyverse)

Neonics <- read.csv("./Data/Raw/ECOTOX_Neonicotinoids_Insects_raw.csv")

view(Neonics)

Litter<-read.csv("./Data/Raw/NEON_NIWO_Litter_massdata_2018-08_raw.csv")
#had to change object name from file name to not include a dash -

view(Litter)
```

Learn about your system

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 ecotoxicologoy of neonicotinoids on insects? Feel free to do a brief internet search if you feel you need more background information.

Answer: We could use the data to determine the efficacy of using neonictinoids in eradicating different classes of insects. We could look at the amount of neonictinoids used, and the amount of insects that existed after the application. We could look at the amount of time between application of neonictinoids and decline in insect population.

The Niwot Ridge litter and woody debris dataset was collected from the National Ecological Observatory Network, which collectively includes 81 aquatic and terrestrial sites across 20 ecoclimatic domains. 32 of these sites sample forest litter and woody debris, and we will focus on the Niwot Ridge long-term ecological research (LTER) station in Colorado. Why might we be interested in studying litter and woody debris that falls to the ground in forests? Feel free to do a brief internet search if you feel you need more background information.

Answer: The amount and condition of woody debris and litter is emerging as a major factor that influences forest ecosystem processes such as carbon cycling, fire behavior, and tree regeneration

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: *Trap placement within plots may be either targeted or randomized, depending on the vegetaion. In sites with > 50% aerial cover of woody vegetaion >2m in height, placement of litter traps is random. In sites*

with < 50% cover of woody vegetation that is heterogeneously distributed, trap placement is targeted beneath vegetation. The finest resolution at which spatial data are reported is a single trap. *Ground traps are sampled once per year. Elevated traps are sampled more frequently. Deciduous forest sites are sampled once every two weeks, and evergren sites are sampled once every one to two months.

Obtain basic summaries of your data (Neonics)

What are the dimensions of the dataset?

```
dim(Neonics)
```

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

5.There are 30 columns of variables and 4623 rows of observed data in the Neonictinoids dataset.

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

```
summary(Neonics)
```

```
##    CAS.Number
##  Min.   : 58842209
##  1st Qu.:138261413
##  Median :138261413
##  Mean   :147651982
##  3rd Qu.:153719234
##  Max.   :210880925
##
##                                                    Chemical.Name
##  (2E)-1-[(6-Chloro-3-pyridinyl)methyl]-N-nitro-2-imidazolidinimine                :2658
##  3-[(2-Chloro-5-thiazolyl)methyl]tetrahydro-5-methyl-N-nitro-4H-1,3,5-oxadiazin-4-imine: 686

##  [C(E)]-N-[(2-Chloro-5-thiazolyl)methyl]-N'-methyl-N''-nitroguanidine             : 452
##  (1E)-N-[(6-Chloro-3-pyridinyl)methyl]-N'-cyano-N-methylethanimidamide            : 420
##  N''-Methyl-N-nitro-N'-[(tetrahydro-3-furanyl)methyl]guanidine                    : 218
##  [N(Z)]-N-[3-[(6-Chloro-3-pyridinyl)methyl]-2-thiazolidinylidene]cyanamide        : 128
##  (Other)                                                                          :  61
##                                    Chemical.Grade
##  Not reported                                 :3989
##  Technical grade, technical product, technical formulation: 422
##  Pestanal grade                               :  93
##  Not coded                                    :  53
##  Commercial grade                             :  27
##  Analytical grade                             :  15
##  (Other)                                      :  24
##                                 Chemical.Analysis.Method
##  Measured                                     : 230
##  Not coded                                    :  51
##  Not reported                                 :   5
##  Unmeasured                                   :4321
##  Unmeasured values (some measured values reported in article):  16
##
##
##  Chemical.Purity              Species.Scientific.Name
##  NR     :2502    Apis mellifera              : 667
##  25     : 244    Bombus terrestris           : 183
##  50     : 200    Apis mellifera ssp. carnica : 152
##  20     : 189    Bombus impatiens            : 140
##  70     : 112    Apis mellifera ssp. ligustica: 113
##  75     :  89    Popillia japonica           :  94
##  (Other):1287    (Other)                     :3274
##           Species.Common.Name
##  Honey Bee           : 667
##  Parasitic Wasp      : 285
##  Buff Tailed Bumblebee: 183
##  Carniolan Honey Bee : 152
##  Bumble Bee          : 140
##  Italian Honeybee    : 113
##  (Other)             :3083
##                                    Species.Group
##  Insects/Spiders                              :3569
##  Insects/Spiders; Standard Test Species       :  27
##  Insects/Spiders; Standard Test Species; U.S. Invasive Species: 667
##  Insects/Spiders; U.S. Invasive Species       : 360
##
##
##
##    Organism.Lifestage  Organism.Age            Organism.Age.Units
##  Not reported:2271     NR     :3851   Not reported       :3515
##  Adult       :1222     2      : 111   Day(s)             : 327
##  Larva       : 437     3      : 105   Instar             : 255
##  Multiple    : 285     <24    :  81   Hour(s)            : 241
```

```
##  Multiple  : 285    <24   :  81   Hour(s)          : 241
##  Egg       : 128    4     :  81   Hours post-emergence:  99
##  Pupa      :  69    1     :  59   Year(s)          :  64
##  (Other)   : 211    (Other): 335   (Other)         : 122
##                  Exposure.Type        Media.Type
##  Environmental, unspecified:1599   No substrate:2934
##  Food                       :1124   Not reported: 663
##  Spray                      : 393   Natural soil: 393
##  Topical, general           : 254   Litter      : 264
##  Ground granular            : 249   Filter paper: 230
##  Hand spray                 : 210   Not coded   :  51
##  (Other)                    : 794   (Other)     :  88
##             Test.Location  Number.of.Doses       Conc.1.Type..Author.
##  Field artificial   :  96   2     :2441    Active ingredient:3161
##  Field natural      :1663   3     : 499    Formulation      :1420
##  Field undeterminable:   4   5     : 314    Not coded        :  42
##  Lab                :2860   6     : 230
##                             4     : 221
##                             NR    : 217
##                             (Other): 701
##  Conc.1..Author. Conc.1.Units..Author.             Effect
##  0.37/  : 208    AI kg/ha  : 575       Population      :1803
##  10/    : 127    AI mg/L   : 298       Mortality       :1493
##  NR/    : 108    AI lb/acre: 277       Behavior        : 360
##  NR     :  94    AI g/ha   : 241       Feeding behavior: 255
##  1      :  82    ng/org    : 231       Reproduction    : 197
##  1023   :  80    ppm       : 180       Development     : 136
##  (Other):3924    (Other)   :2821       (Other)         : 379
##             Effect.Measurement    Endpoint                 Response.Site
##  Abundance            :1699     NOEL   :1816   Not reported          :4349
##  Mortality            :1294     LOEL   :1664   Midgut or midgut gland:  63
##  Survival             : 133     LC50   : 327   Not coded             :  51
##  Progeny counts/numbers: 120     LD50   : 274   Whole organism        :  41
##  Food consumption     : 103     NR     : 167   Hypopharyngeal gland  :  27
##  Emergence            :  98     NR-LETH:  86   Head                  :  23
##  (Other)              :1176     (Other): 289   (Other)               :  69
##  Observed.Duration..Days.       Observed.Duration.Units..Days.
##  1      : 713            Day(s)             :4394
##  2      : 383            Emergence          :  70
##  NR     : 355            Growing season     :  48
##  7      : 207            Day(s) post-hatch  :  20
##  3      : 183            Day(s) post-emergence:  17
##  0.0417 : 133            Tiller stage       :  15
##  (Other):2649            (Other)            :  59
##                                                  Author
##  Peck,D.C.                                             : 208
##  Frank,S.D.                                            : 100
##  El Hassani,A.K., M. Dacher, V. Gary, M. Lambin, M. Gauthier, and C. Armengaud:  96
##  Williamson,S.M., S.J. Willis, and G.A. Wright         :  93
##  Laurino,D., A. Manino, A. Patetta, and M. Porporato   :  88
##  Scholer,J., and V. Krischik                           :  82
##  (Other)                                               :3956
##  Reference.Number
##  Min.   :   344
##  1st Qu.:108459
##  Median :165559
##  Mean   :142189
##  3rd Qu.:168998
##  Max.   :180410
##
##                                                                                                  Title
##  Long-Term Effects of Imidacloprid on the Abundance of Surface- and Soil-Active Nontarget
Fauna in Turf                                 : 200
##  Reduced Risk Insecticides to Control Scale Insects and Protect Natural Enemies in the
Production and Maintenance of Urban Landscape Plants: 100
##  Effects of Sublethal Doses of Acetamiprid and Thiamethoxam on the Behavior of the Honeybee
(Apis mellifera)                              :  96
##  Exposure to Neonicotinoids Influences the Motor Function of Adult Worker Honeybees
                                              :  93
##  Toxicity of Neonicotinoid Insecticides on Different Honey Bee Genotypes
                                              :  88
##  Chronic Exposure of Imidacloprid and Clothianidin Reduce Queen Survival, Foraging, and
Nectar Storing in Colonies of Bombus impatiens     :  82
##  (Other)                                                                                         :3964
##                                        Source     Publication.Year
##  Agric. For. Entomol.11(4): 405-419          : 200   Min.   :1982
```

```
##  Environ. Entomol.41(2): 377-386             : 100   1st Qu.:2005
##  Arch. Environ. Contam. Toxicol.54(4): 653-661:  96   Median :2010
##  Ecotoxicology23:1409-1418                   :  93   Mean   :2008
##  Bull. Insectol.66(1): 119-126               :  88   3rd Qu.:2013
##  PLoS One9(3): 14 p.                         :  82   Max.   :2019
##  (Other)                                     :3964
##  Summary.of.Additional.Parameters

##  Purity: \xca NR - NR | Organism Age: \xca NR - NR Not reported | Conc 1 (Author): \xca
Active ingredient NR/ - NR/ AI kg/ha | Duration (Days): \xca NR - NR NR | Conc 2 (Author): \xca
NR (NR - NR) NR | Conc 3 (Author): \xca NR (NR - NR) NR  : 389
##  Purity: \xca NR - NR | Organism Age: \xca NR - NR Not reported | Conc 1 (Author): \xca
Active ingredient NR - NR AI lb/acre | Duration (Days): \xca NR - NR NR | Conc 2 (Author): \xca
NR (NR - NR) NR | Conc 3 (Author): \xca NR (NR - NR) NR  : 138
##  Purity: \xca NR - NR | Organism Age: \xca NR - NR Not reported | Conc 1 (Author): \xca
Active ingredient NR - NR AI kg/ha | Duration (Days): \xca NR - NR NR | Conc 2 (Author): \xca
NR (NR - NR) NR | Conc 3 (Author): \xca NR (NR - NR) NR    : 136
##  Purity: \xca NR - NR | Organism Age: \xca NR - NR Not reported | Conc 1 (Author): \xca
Active ingredient NR/ - NR/ AI lb/acre | Duration (Days): \xca NR - NR NR | Conc 2 (Author):
\xca NR (NR - NR) NR | Conc 3 (Author): \xca NR (NR - NR) NR: 124
##  Purity: \xca NR - NR | Organism Age: \xca NR - NR Not reported | Conc 1 (Author): \xca
Active ingredient NR - NR AI ng/org | Duration (Days): \xca NR - NR NR | Conc 2 (Author): \xca
NR (NR - NR) NR | Conc 3 (Author): \xca NR (NR - NR) NR   :  94
##  Purity: \xca NR - NR | Organism Age: \xca NR - NR Not reported | Conc 1 (Author): \xca
Formulation NR - NR ml/ha | Duration (Days): \xca NR - NR NR | Conc 2 (Author): \xca NR (NR -
NR) NR | Conc 3 (Author): \xca NR (NR - NR) NR            :  80
##  (Other)
                                                                :3662
```

Answer: The effects that are studied are abundance, mortality, survival, food consumption, progeny count and emergence. These effects might be of interest because we can determine the number of insects that Neonictinoids kill, the percentage of an insect population that is killed or effected by the Neonictinoid and how Neonictinoids affect reproduction.

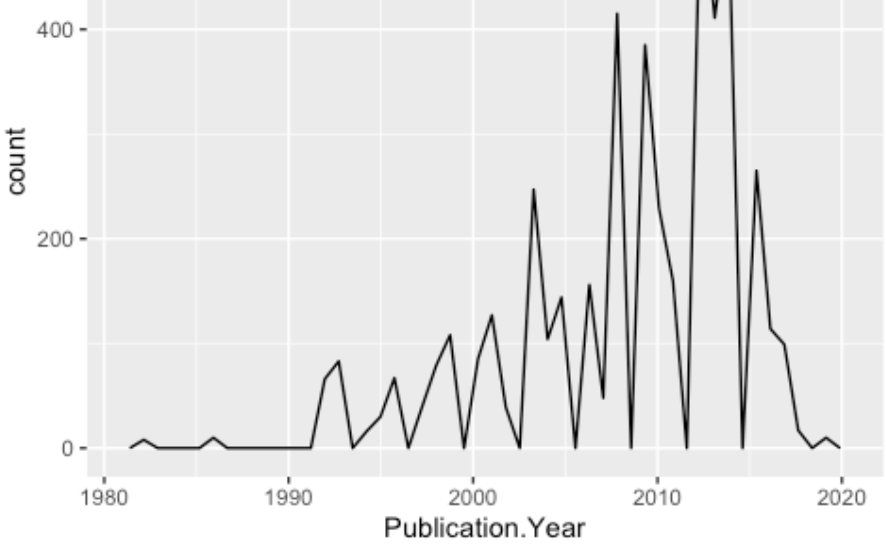
Using the summary function, determine the six most commonly studied species in the dataset (common name). What do these species have in common, and why might they be of interest over other insects? Feel free to do a brief internet search for more information if needed.

```
summary(Neonics)
##    CAS.Number
##  Min.   : 58842209
##  1st Qu.:138261413
##  Median :138261413
##  Mean   :147651982
##  3rd Qu.:153719234
##  Max.   :210880925
##
##                                                    Chemical.Name
##  (2E)-1-[(6-Chloro-3-pyridinyl)methyl]-N-nitro-2-imidazolidinimine                :2658
##  3-[(2-Chloro-5-thiazolyl)methyl]tetrahydro-5-methyl-N-nitro-4H-1,3,5-oxadiazin-4-imine: 686

##  [C(E)]-N-[(2-Chloro-5-thiazolyl)methyl]-N'-methyl-N''-nitroguanidine             : 452
##  (1E)-N-[(6-Chloro-3-pyridinyl)methyl]-N'-cyano-N-methylethanimidamide            : 420
##  N''-Methyl-N-nitro-N'-[(tetrahydro-3-furanyl)methyl]guanidine                    : 218
##  [N(Z)]-N-[3-[(6-Chloro-3-pyridinyl)methyl]-2-thiazolidinylidene]cyanamide        : 128
##  (Other)                                                                          :  61
##                                    Chemical.Grade
##  Not reported                                :3989
##  Technical grade, technical product, technical formulation: 422
##  Pestanal grade                              :  93
##  Not coded                                   :  53
##  Commercial grade                            :  27
##  Analytical grade                            :  15
##  (Other)                                     :  24
##                                 Chemical.Analysis.Method
##  Measured                                    : 230
##  Not coded                                   :  51
##  Not reported                                :   5
##  Unmeasured                                  :4321
##  Unmeasured values (some measured values reported in article):  16
##
##
##  Chemical.Purity                Species.Scientific.Name
##  NR     :2502    Apis mellifera              : 667
##  25     : 244    Bombus terrestris           : 183
##  50     : 200    Apis mellifera ssp. carnica : 152
##  20     : 189    Bombus impatiens            : 140
```

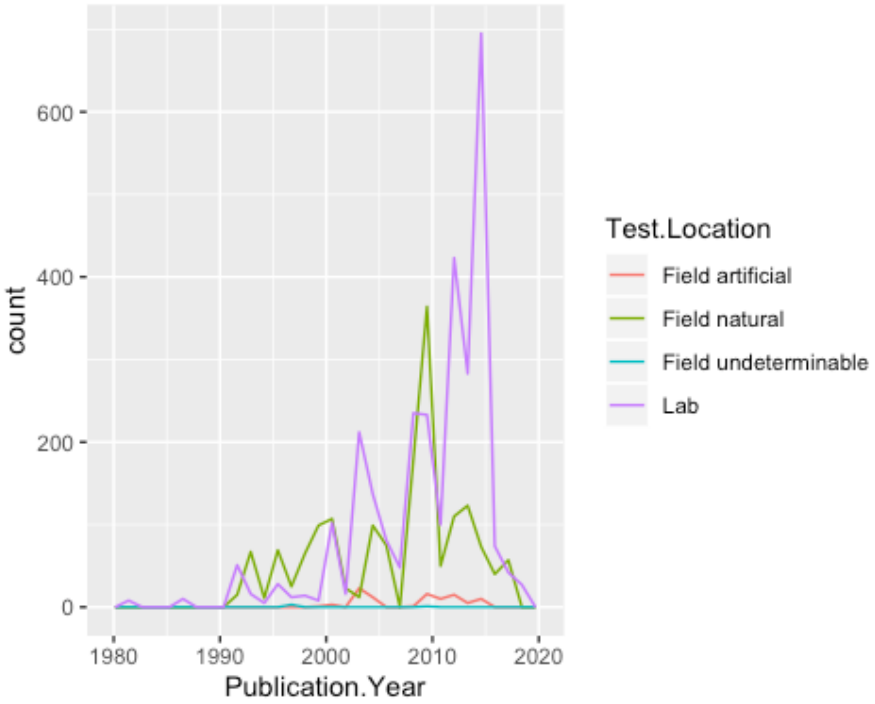
```
## 20    : 189   Bombus impatiens          : 140
## 70    : 112   Apis mellifera ssp. ligustica: 113
## 75    :  89   Popillia japonica          :  94
## (Other):1287   (Other)                   :3274
##           Species.Common.Name
## Honey Bee           : 667
## Parasitic Wasp      : 285
## Buff Tailed Bumblebee: 183
## Carniolan Honey Bee  : 152
## Bumble Bee          : 140
## Italian Honeybee    : 113
## (Other)             :3083
##                                          Species.Group
## Insects/Spiders                                  :3569
## Insects/Spiders; Standard Test Species                :  27
## Insects/Spiders; Standard Test Species; U.S. Invasive Species: 667
## Insects/Spiders; U.S. Invasive Species                : 360
##
##
##
##    Organism.Lifestage  Organism.Age          Organism.Age.Units
## Not reported:2271     NR     :3851   Not reported       :3515
## Adult       :1222     2      : 111   Day(s)             : 327
## Larva       : 437     3      : 105   Instar             : 255
## Multiple    : 285     <24    :  81   Hour(s)            : 241
## Egg         : 128     4      :  81   Hours post-emergence:  99
## Pupa        :  69     1      :  59   Year(s)            :  64
## (Other)     : 211     (Other): 335   (Other)            : 122
##                 Exposure.Type        Media.Type
## Environmental, unspecified:1599   No substrate:2934
## Food                    :1124   Not reported: 663
## Spray                   : 393   Natural soil: 393
## Topical, general        : 254   Litter      : 264
## Ground granular         : 249   Filter paper: 230
## Hand spray              : 210   Not coded   :  51
## (Other)                 : 794   (Other)     :  88
##            Test.Location  Number.of.Doses       Conc.1.Type..Author.
## Field artificial    :  96   2      :2441    Active ingredient:3161
## Field natural       :1663   3      : 499    Formulation      :1420
## Field undeterminable:   4   5      : 314    Not coded        :  42
## Lab                 :2860   6      : 230
##                             4      : 221
##                             NR     : 217
##                             (Other): 701
## Conc.1..Author. Conc.1.Units..Author.            Effect
## 0.37/  : 208    AI kg/ha  : 575       Population      :1803
## 10/    : 127    AI mg/L   : 298       Mortality       :1493
## NR/    : 108    AI lb/acre: 277       Behavior        : 360
## NR     :  94    AI g/ha   : 241       Feeding behavior: 255
## 1      :  82    ng/org    : 231       Reproduction    : 197
## 1023   :  80    ppm       : 180       Development     : 136
## (Other):3924    (Other)   :2821       (Other)         : 379
##            Effect.Measurement    Endpoint                Response.Site
## Abundance            :1699     NOEL   :1816   Not reported         :4349
## Mortality            :1294     LOEL   :1664   Midgut or midgut gland:  63
## Survival             : 133     LC50   : 327   Not coded            :  51
## Progeny counts/numbers: 120     LD50   : 274   Whole organism       :  41
## Food consumption      : 103     NR     : 167   Hypopharyngeal gland  :  27
## Emergence            :  98     NR-LETH:  86   Head                 :  23
## (Other)              :1176     (Other): 289   (Other)              :  69
## Observed.Duration..Days.       Observed.Duration.Units..Days.
## 1      : 713            Day(s)              :4394
## 2      : 383            Emergence           :  70
## NR     : 355            Growing season      :  48
## 7      : 207            Day(s) post-hatch   :  20
## 3      : 183            Day(s) post-emergence:  17
## 0.0417 : 133            Tiller stage        :  15
## (Other):2649            (Other)             :  59
##                                                        Author
## Peck,D.C.                                                  : 208
## Frank,S.D.                                                 : 100
## El Hassani,A.K., M. Dacher, V. Gary, M. Lambin, M. Gauthier, and C. Armengaud:  96
## Williamson,S.M., S.J. Willis, and G.A. Wright                    :  93
## Laurino,D., A. Manino, A. Patetta, and M. Porporato                :  88
## Scholer,J., and V. Krischik                                    :  82
## (Other)                                                    :3956
```

```
##  Reference.Number
##  Min.   :   344
##  1st Qu.:108459
##  Median :165559
##  Mean   :142189
##  3rd Qu.:168998
##  Max.   :180410
##
##                                                
```



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, colour = Test.Location))
## `stat_bin()` using `bins = 30`. Pick better value with `binwidth`.
```

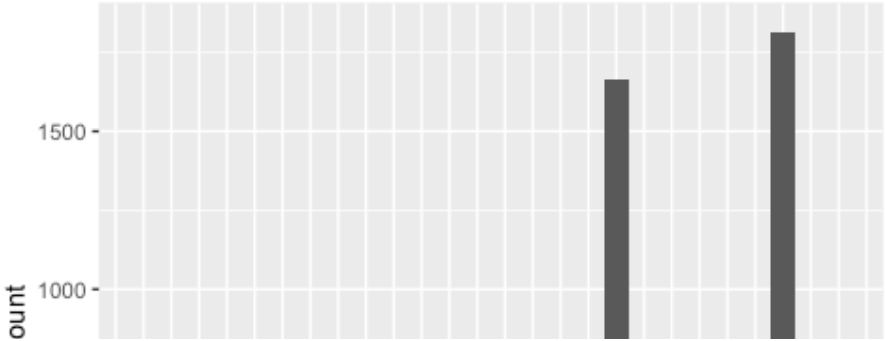


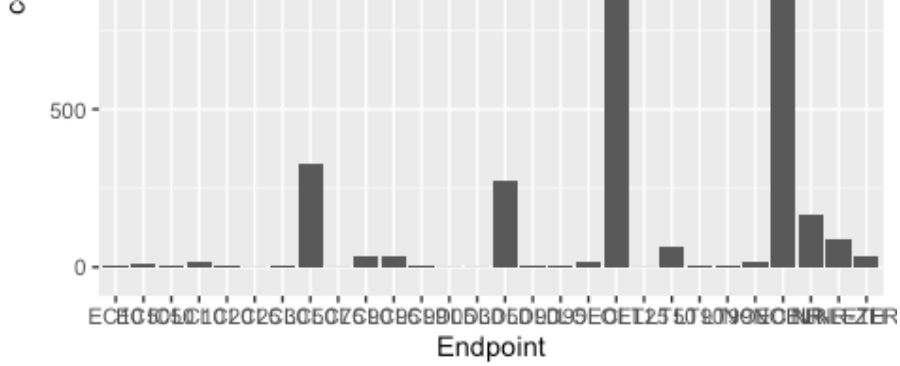
Interpret this graph. What are the most common test locations, and do they differ over time?

Answer: Field natural and lab are the most common test locations. Between 1990 and 2000 field natural and lab had about the same count. Between 2010 and 2020, lab tests became much more prevalent (6 to 10 times more prevalent) than field natural test locations.

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()
```





Answer: The two more common end points are LOEL and NOEL.

Explore your data (Litter)

Determine the class of collectDate. Is it a date? If not, change to a date and confirm the new class of the variable. Using the unique function, determine which dates litter was sampled in August 2018.

```
class(Litter$collectDate)
## [1] "factor"
Litter$collectDate <- format(Litter$collectDate, format = "%y-%m-%d")
Litter$collectDate <- as.Date(Litter$collectDate, format = "%y%m%d")
```

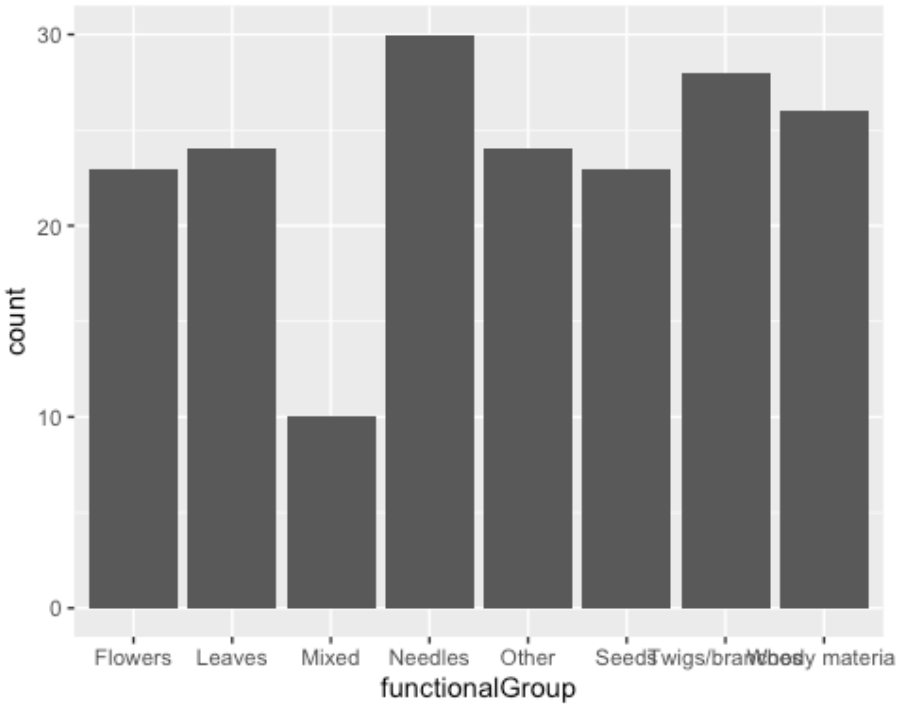
Using the unique function, determine how many plots were sampled at Niwot Ridge. How is the information obtained from unique different from that obtained from summary?

```
class(Litter$plotID)
## [1] "factor"
unique(Litter$plotID)
##  [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
```

Answer: There were 12 different plots sampled at Niwot Ridge. The unique funciton pulls up just the number of different type of factors of the variable. Summary pulls up all of the different factors for all the different variables.

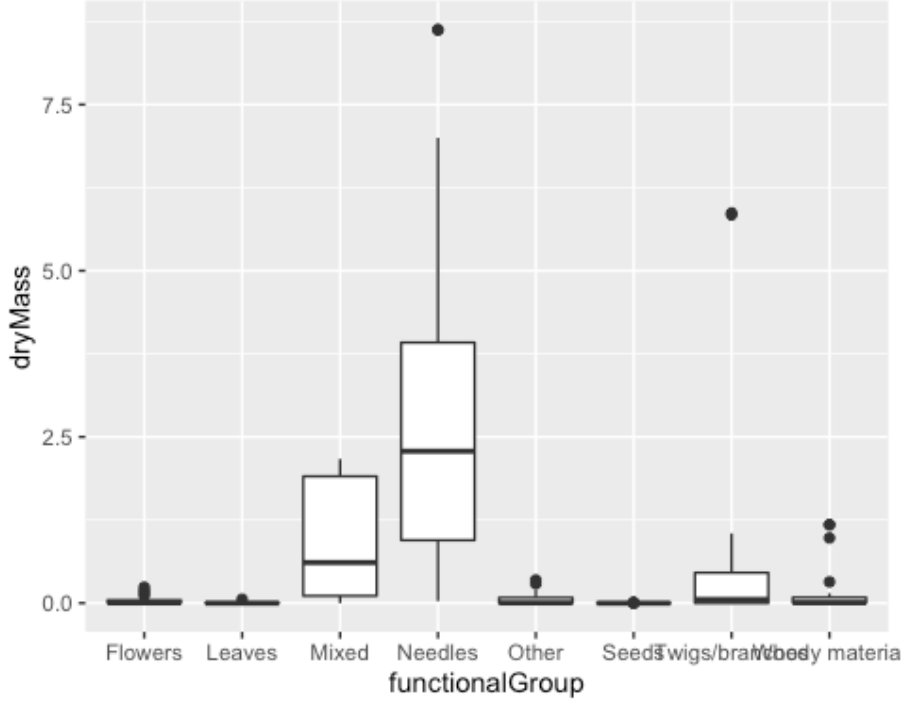
Create a bar graph of functionalGroup counts. This shows you what type of litter is collected at the Niwot Ridge sites. Notice that litter types are fairly equally distributed across the Niwot Ridge sites.

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

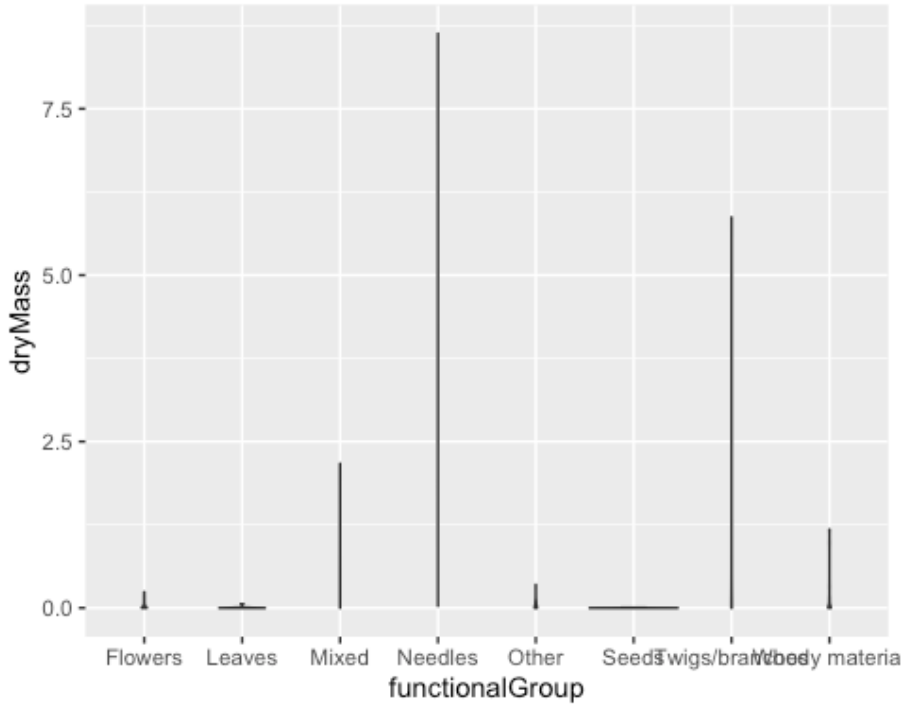


Using geom_boxplot and geom_violin, create a boxplot and a violin plot of dryMass by functionalGroup.

```
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: The boxplot is more effective. The violin plot is not showing the median, interquartile spread, or outliers like the boxplot.

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

Answer: Needles and Twigs/Branches