Assignment 3: Data Exploration

Laurel Cohen, Section #1

OVERVIEW

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

Directions

1. Change "Student Name, Section #" on line 3 (above) with your name and section number.
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., "FirstLast_A03_DataExploration.Rmd") prior to submission.

The completed exercise is due on <>.

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 add the `stringsAsFactors = TRUE` parameter to the function when reading in the CSV files.**

```
setwd("/Users/Laurel/Documents/Information to Keep/Graduate School/Second Year/Second Semester/Environme
library(tidyverse)
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 ecotoxicologoy of neonicotinoids on insects? Feel free to do a brief internet search if you feel you need more background information.

   Answer: Neonicotinoids have been shown to be harmful to a wide array of insects, especially bees. Insects are a critical part of a healthy ecosystem, and bees especially are vital for pollinating crops. The ecotoxicology of neonicotinoids on insects is important to understand so farmers and other stakeholders (like eaters, or land managers, or neighbors, or state regulatory agencies who protect the environment or natural resources) can make informed decisions about how to raise crops economically as well as safely, for the environment and consumers.

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: The makeup of litter and woody debris that falls to the ground in forests has major implications for the nutrient content of the soil and the nutrient cycling and ecosystem function of the forest as a whole.

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 fine woody debris are sampled via elevated and ground traps, respectively. Sampling occurs only in tower plots. Ground traps are sampled once per year; target sampling frequency for elevated traps varies by vegetation present at the site, with frequent sampling (1x every 2weeks) in deciduous forest sites during senescence, and infrequent year-round sampling (1x every 1-2 months) at evergreen sites.

Obtain basic summaries of your data (Neonics)

5. What are the dimensions of the dataset?

```
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
##  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
##
##
##  Long-Term Effects of Imidacloprid on the Abundance of Surface- and Soil-Active Nontarget Fauna in Tu
##  Reduced Risk Insecticides to Control Scale Insects and Protect Natural Enemies in the Production and
##  Effects of Sublethal Doses of Acetamiprid and Thiamethoxam on the Behavior of the Honeybee (Apis mel
##  Exposure to Neonicotinoids Influences the Motor Function of Adult Worker Honeybees
##  Toxicity of Neonicotinoid Insecticides on Different Honey Bee Genotypes
##  Chronic Exposure of Imidacloprid and Clothianidin Reduce Queen Survival, Foraging, and Nectar Storin
##  (Other)
##                                           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 ingred
##  Purity: \xca NR - NR | Organism Age: \xca NR - NR Not reported | Conc 1 (Author): \xca Active ingred
##  Purity: \xca NR - NR | Organism Age: \xca NR - NR Not reported | Conc 1 (Author): \xca Active ingred
```

```
##  Purity: \xca NR - NR | Organism Age: \xca NR - NR Not reported | Conc 1 (Author): \xca Active ingre
##  Purity: \xca NR - NR | Organism Age: \xca NR - NR Not reported | Conc 1 (Author): \xca Active ingre
##  Purity: \xca NR - NR | Organism Age: \xca NR - NR Not reported | Conc 1 (Author): \xca Formulation 
##  (Other)
```

```
dim(Neonics)
```

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

```
#The dataset has 4623 rows and 30 columns.
```

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

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

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

Answer: Mortality and Population are the most commonly studied effects. These effects specifically may be of interest because they are upstream of all the other effects: if the creatures die, there is no other effect to study.

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.

```
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 in the dataset are Honey Bees, Parasitic Wasps, Buff Tailed Bumblebees, Carniolan Honey Bees, Bumble Bees, and Italian Honeybees. Most of these species are bees, and all are pollinators. These species might be of interest over other insects because if farmers use insecticides to kill nuisance pests that eat their crops, they might be sabotaging their efforts if the chemicals also kill the creatures that help pollinate the plants.

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

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

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

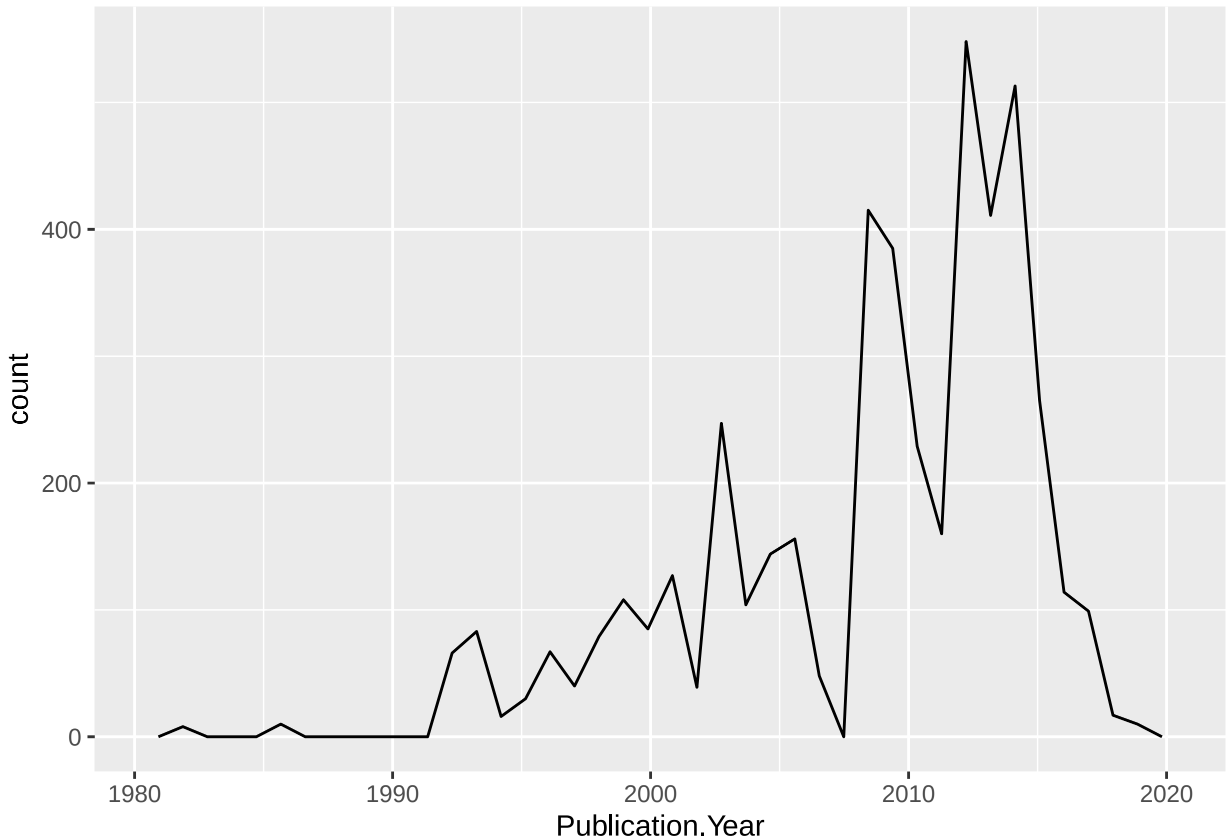
> Answer: Conc.1..Author is a factor variable. It is not numeric because it is a label that indicates the type of measurement each author took–whether a range or an exact number–and not the measurement value itself.

Explore your data graphically (Neonics)

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

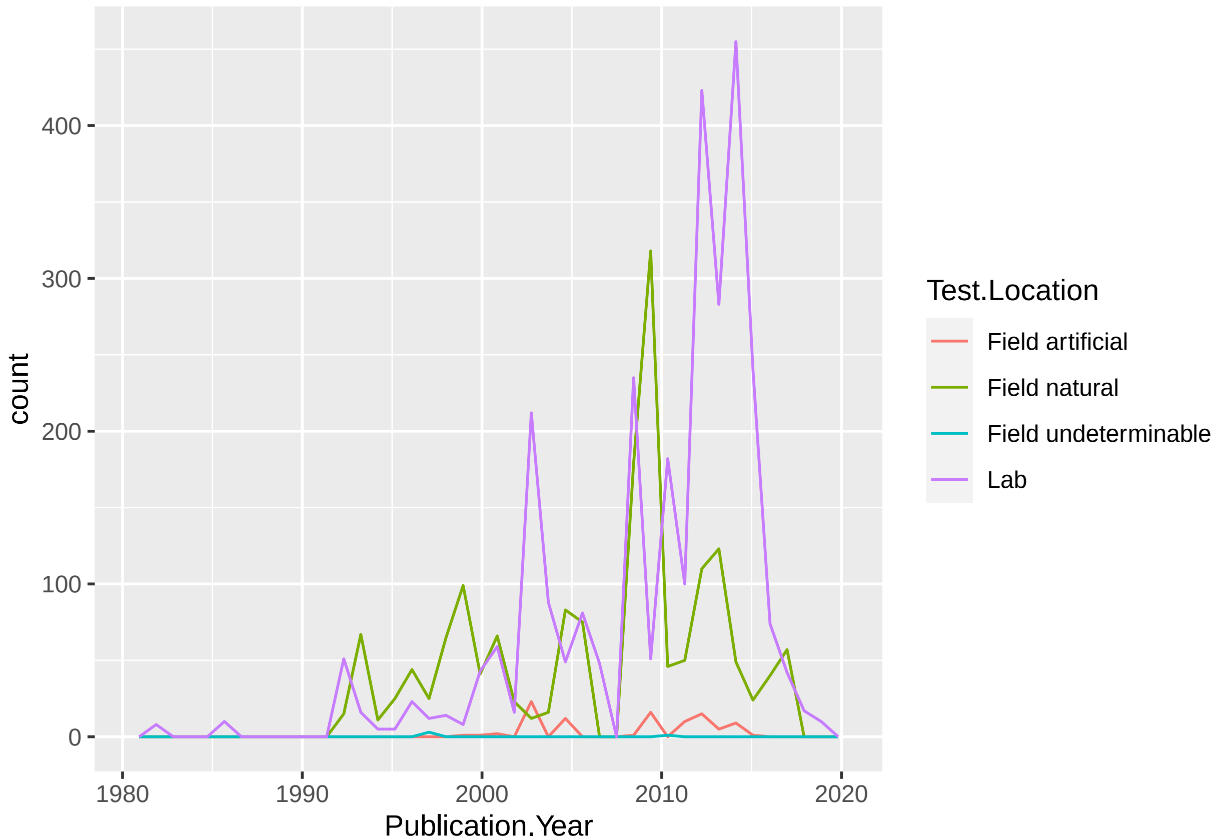
```r
library(dplyr)
library(ggplot2)

ggplot(Neonics) +
  geom_freqpoly(aes(x = Publication.Year), bins = 40)
```



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 = 40)
```

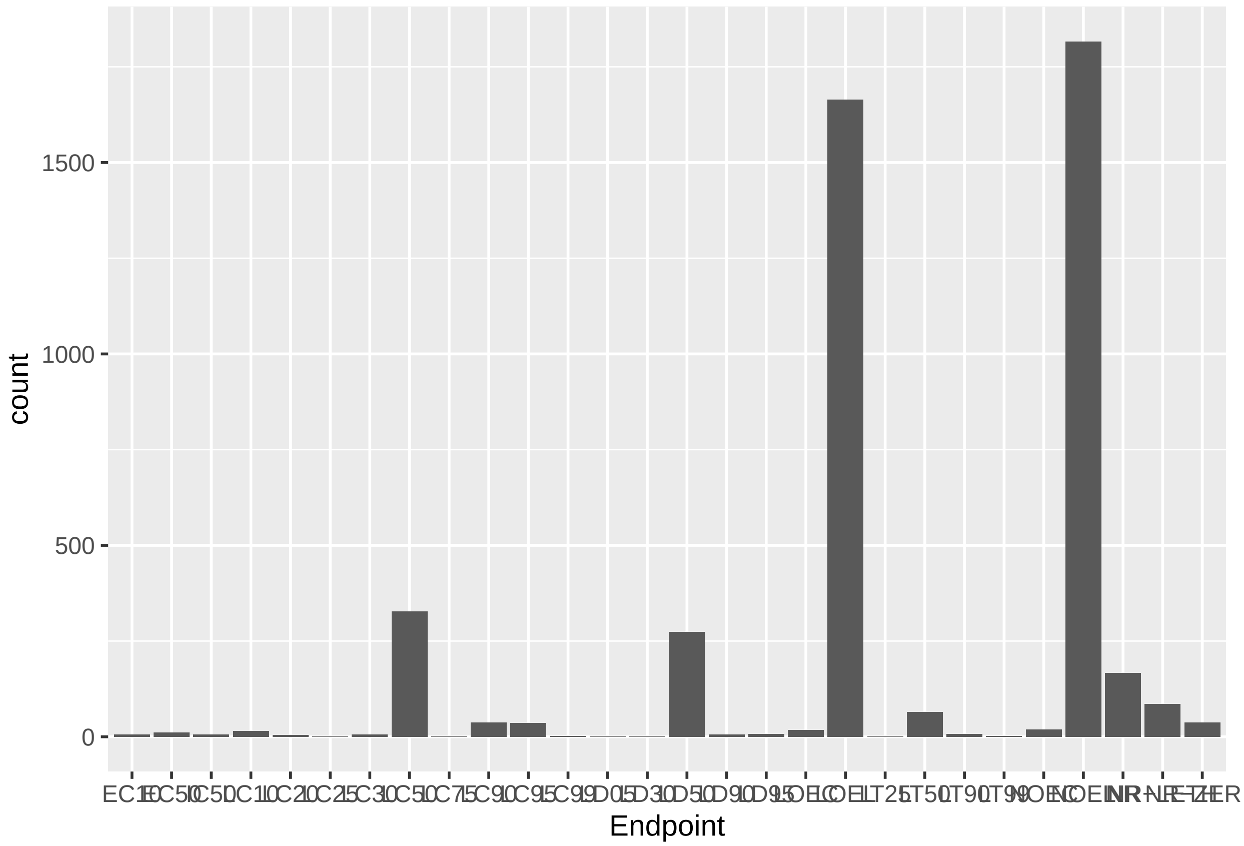


Interpret this graph. What are the most common test locations, and do they differ over time?

> Answer: The most common test location by overall number of studies is the lab, but in certain years, more natural field studies were published than lab studies. There are only a few artifial field lab studies, they were all conducted between 1995 and 2015, and the volume of those per year never even gets close to those of the other two categories.

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(Endpoint)) +
  geom_bar()
```



Answer: The two most common end points are NOEL and LOEL. NOEL is defined as, "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)." LOEL is defined as, "Lowest-observable-effect-level: lowest dose (concentration) producing effects that were significantly different (as reported by authors) from responses of controls (LOEAL/LOEC)."

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)
```

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

```
Litter$collectDate <- as.Date(Litter$collectDate, format = "%Y%m$d")
class(Litter$collectDate)
```

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

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`?

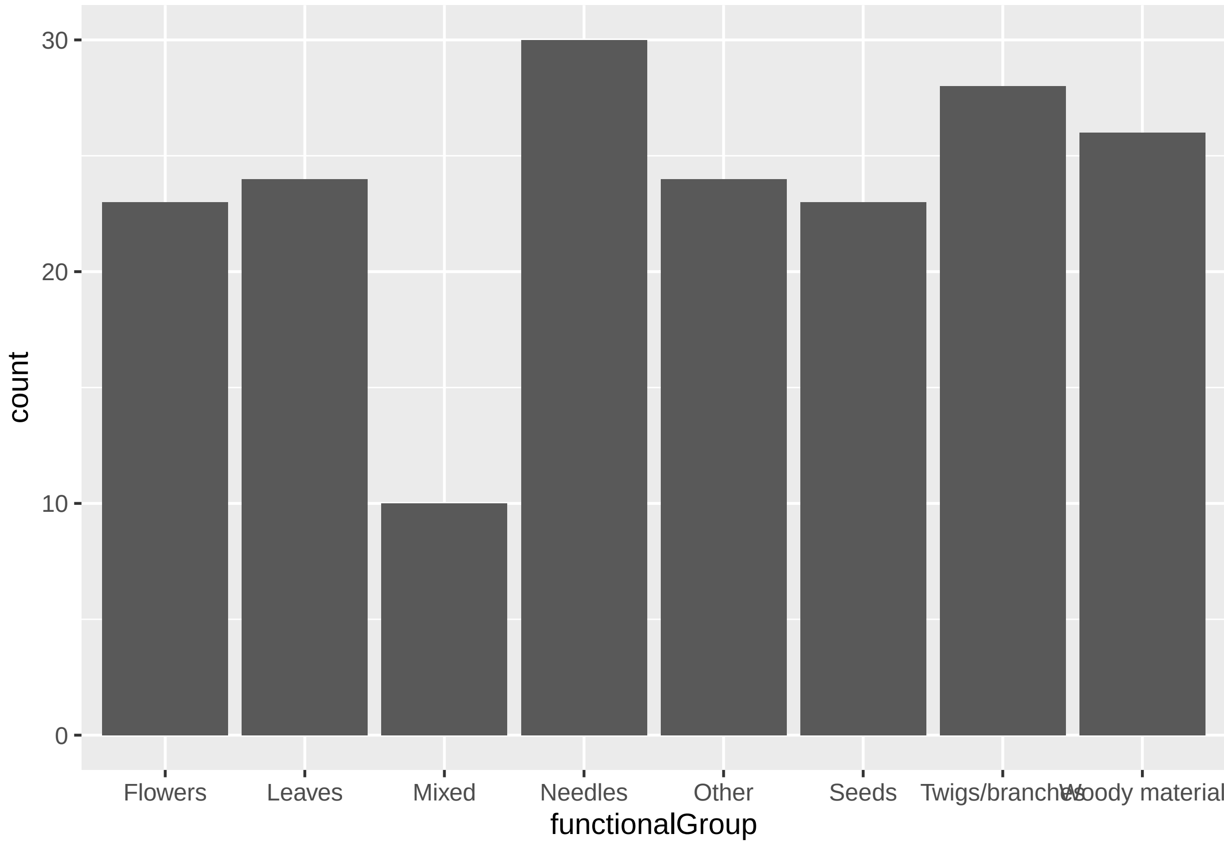
```
unique(Litter$plotID)
```

```
##  [1] NIWO_061 NIWO_064 NIWO_067 NIWO_040 NIWO_041 NIWO_063 NIWO_047 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 plots sampled at Niwot Ridge. The information obtained from 'unique' is different from that obtained from 'summary' because 'unique' eliminates duplicate data.

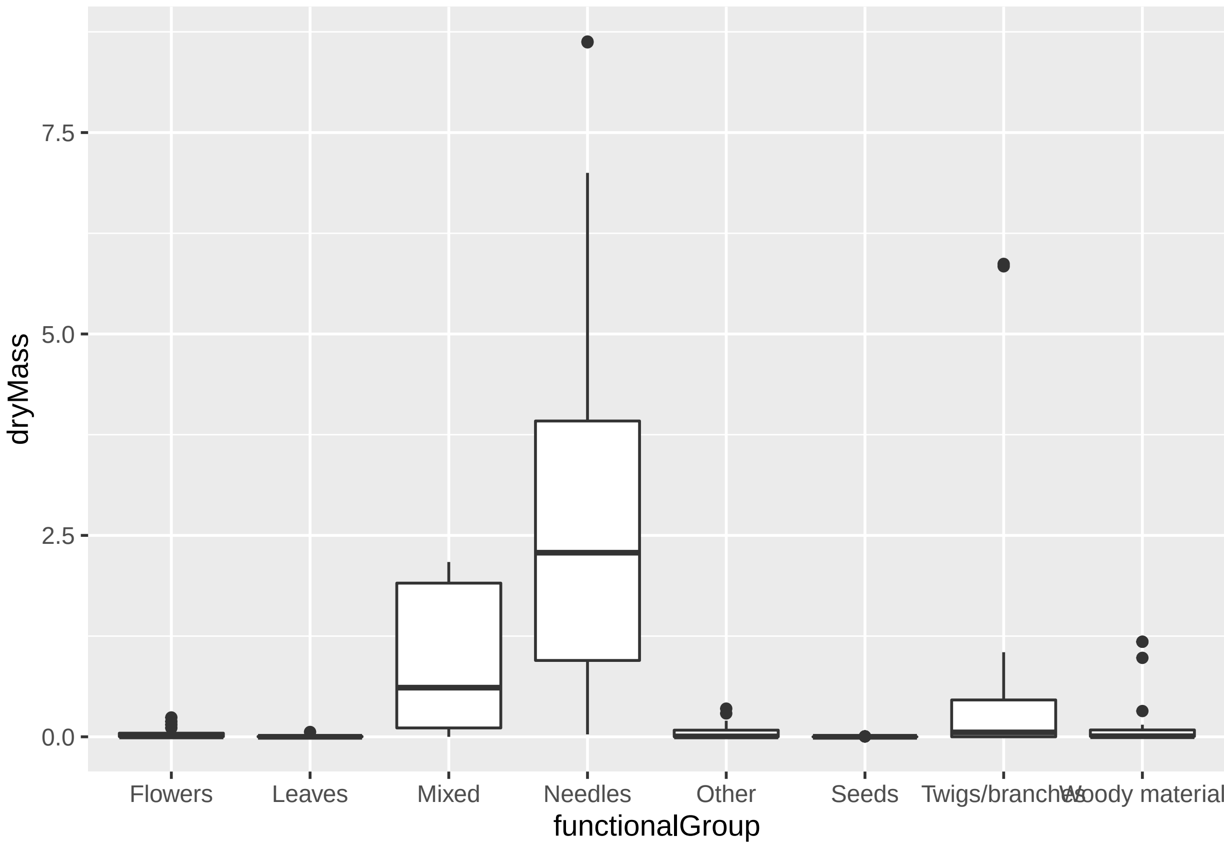
14. Create a bar graph of functionalGroup counts. This shows you what type of litter is collected at the Niwot Ridge sites. Notice that litter types are fairly equally distributed across the Niwot Ridge sites.

```
ggplot(Litter, aes(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 = functionalGroup, y = dryMass))
```

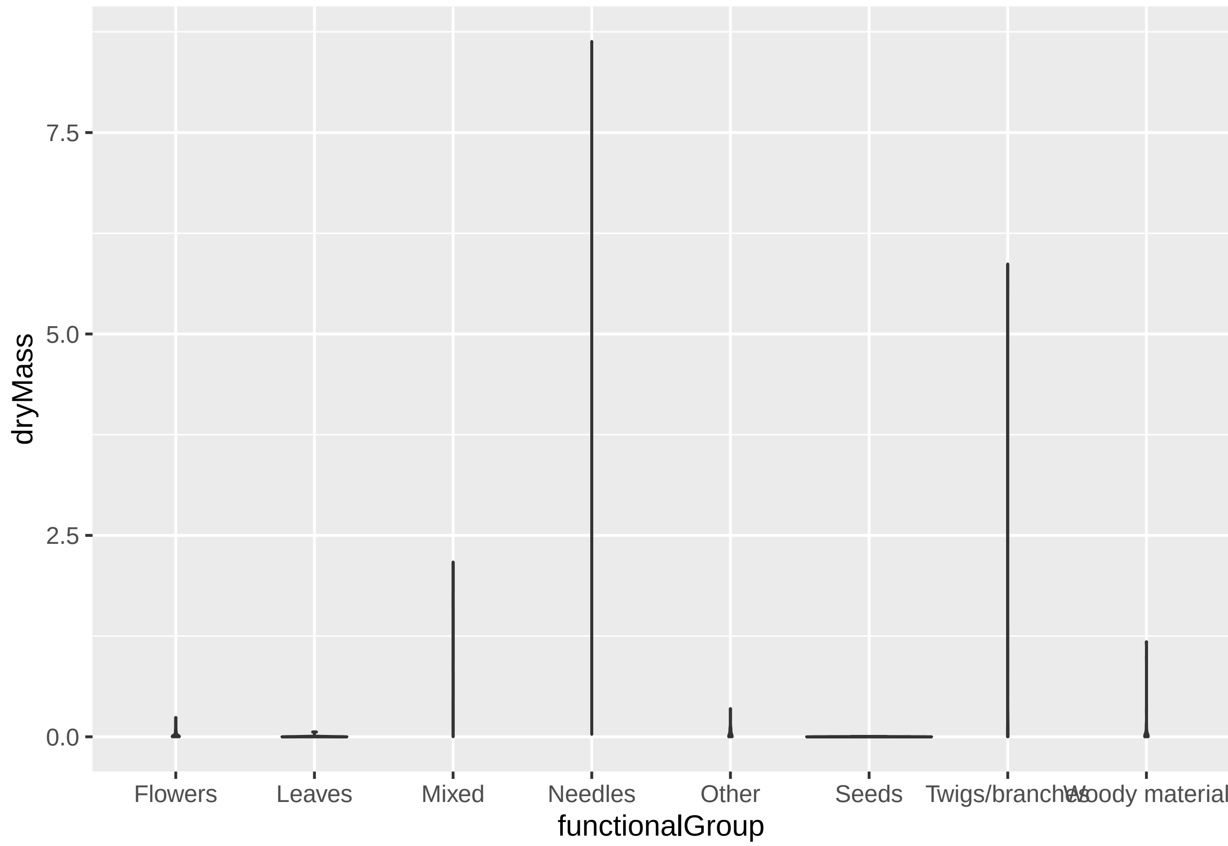



```r
ggplot(Litter) +
  geom_violin(aes(x = functionalGroup, y = dryMass),
              draw_quantiles = c(0.25, 0.5, 0.75))
```

```
## Warning in regularize.values(x, y, ties, missing(ties), na.rm = na.rm):
## collapsing to unique 'x' values

## Warning in regularize.values(x, y, ties, missing(ties), na.rm = na.rm):
## collapsing to unique 'x' values

## Warning in regularize.values(x, y, ties, missing(ties), na.rm = na.rm):
## collapsing to unique 'x' values
```



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

> Answer: The boxplot is a more effective visualization option than the violin plot here because the density distributions are very narrow, so using the violin plot returns essentially a bunch of straight lines that are hard to read and interpret.

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

> Answer: The litter type that tends to have the highest biomass at these sites is Needles.