# Assignment 3_Data Exploration

Angelica Rodriguez

Fall 2025

## OVERVIEW

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

## Directions

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

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

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

---

## Set up your R session

1. Load necessary packages (tidyverse, lubridate, here), check your current working directory 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.

```
#Installing packages (tidyverse, lubridate, here)
#install.packages('tidyverse')
#install.packages('lubridate')
#install.packages('here')
#Get the link of the working directory in R
getwd()
```

```
## [1] "/home/guest/EDA_Spring2025"
```

```r
#get the library in package "Here"
#library(here)

#one way to access to the databases ECOTOX_Neonicotinoids_Insects_raw.csv
#and NEON_NIWO_Litter_massdata_2018-08_raw.csv vizualization from the console

#Read all libraries  (tidyverse); (lubridate); (here);(ggplot2)
library(tidyverse); library(lubridate); library(here); library(ggplot2)

#other way to access to the databases ECOTOX_Neonicotinoids_Insects_raw.csv
#and NEON_NIWO_Litter_massdata_2018-08_raw.csv

Neonics <- read.csv(here('Data/Raw/ECOTOX_Neonicotinoids_Insects_raw.csv'), stringsAsFactors = TRUE)
str(Neonics)#data visualization and look at them as factors
```

```
## 'data.frame':    4623 obs. of  30 variables:
##  $ CAS.Number                      : int  58842209 58842209 58842209 58842209 58842209 58842209 5884
##  $ Chemical.Name                   : Factor w/ 9 levels "(1E)-N-[(6-Chloro-3-pyridinyl)methyl]-N-ethy
##  $ Chemical.Grade                  : Factor w/ 9 levels "Analytical grade",..: 9 9 9 9 9 9 9 9 9 .
##  $ Chemical.Analysis.Method        : Factor w/ 5 levels "Measured","Not coded",..: 4 4 4 4 4 4 4 4 4
##  $ Chemical.Purity                 : Factor w/ 80 levels ">=98",">=99.0",..: 69 69 50 50 50 50 50 50
##  $ Species.Scientific.Name         : Factor w/ 398 levels "Acalolepta vastator",..: 69 69 248 248 24
##  $ Species.Common.Name             : Factor w/ 303 levels "Alfalfa Leafcutter Bee",..: 74 74 142 142
##  $ Species.Group                   : Factor w/ 4 levels "Insects/Spiders",..: 1 1 1 1 1 1 1 1 1 1 ..
##  $ Organism.Lifestage              : Factor w/ 20 levels "Adult","Cocoon",..: 1 1 19 19 19 1 19 1 1
##  $ Organism.Age                    : Factor w/ 39 levels "<=24","<=48",..: 39 39 39 39 39 36 39 36 3
##  $ Organism.Age.Units              : Factor w/ 11 levels "Day(s)","Days post-emergence",..: 9 9 4 4
##  $ Exposure.Type                   : Factor w/ 24 levels "Choice","Dermal",..: 23 23 11 11 11 11 11
##  $ Media.Type                      : Factor w/ 10 levels "Agar","Artificial soil",..: 7 7 3 3 3 3 3
##  $ Test.Location                   : Factor w/ 4 levels "Field artificial",..: 4 4 4 4 4 4 4 4 4 4 .
##  $ Number.of.Doses                 : Factor w/ 30 levels "' 4-5","' 4-7",..: 30 30 18 18 18 18 18 18
##  $ Conc.1.Type..Author.            : Factor w/ 3 levels "Active ingredient",..: 1 1 1 1 1 1 1 1 1 1
##  $ Conc.1..Author.                 : Factor w/ 1006 levels "<0.0004","<0.025",..: 639 510 813 622 44
##  $ Conc.1.Units..Author.           : Factor w/ 148 levels "%","% v/v","% w/v",..: 132 132 91 91 91 9
##  $ Effect                          : Factor w/ 19 levels "Accumulation",..: 16 16 16 16 16 16 16 16
##  $ Effect.Measurement              : Factor w/ 155 levels "Abundance","Accuracy of learned task, per
##  $ Endpoint                        : Factor w/ 28 levels "EC10","EC50",..: 15 15 8 8 8 8 8 8 8 8 ...
##  $ Response.Site                   : Factor w/ 19 levels "Abdomen","Brain",..: 14 14 14 14 14 14 14
##  $ Observed.Duration..Days.        : Factor w/ 361 levels "<.0002","<.0021",..: 145 145 145 145 145
##  $ Observed.Duration.Units..Days.  : Factor w/ 17 levels "Day(s)","Day(s) post-emergence",..: 1 1 1
##  $ Author                          : Factor w/ 433 levels "Abbott,V.A., J.L. Nadeau, H.A. Higo, and
##  $ Reference.Number                : int  107388 107388 103312 103312 103312 103312 103312 103312 10
##  $ Title                           : Factor w/ 458 levels "A Common Pesticide Decreases Foraging Suc
##  $ Source                          : Factor w/ 456 levels "Acta Hortic.1094:451-456",..: 295 295 296
##  $ Publication.Year                : int  1982 1982 1986 1986 1986 1986 1986 1986 1986 1986 ...
##  $ Summary.of.Additional.Parameters: Factor w/ 943 levels "Purity: \xca NC - NC | Organism Age: \xca
```

```r
Litter <- read.csv(here("Data/Raw/NEON_NIWO_Litter_massdata_2018-08_raw.csv"), stringsAsFactors = TRUE)
str(Litter)#data visualization and look at them as factors
```

```
## 'data.frame':    188 obs. of  19 variables:
##  $ uid                   : Factor w/ 188 levels "028eea3d-5c20-4afc-bb7e-a05bab305152",..: 84 96 85
```

```
## $ namedLocation          : Factor w/ 12 levels "NIWO_040.basePlot.ltr",..: 8 8 8 8 8 8 8 8 11 11 ..
## $ domainID               : Factor w/ 1 level "D13": 1 1 1 1 1 1 1 1 1 1 ...
## $ siteID                 : Factor w/ 1 level "NIWO": 1 1 1 1 1 1 1 1 1 1 ...
## $ plotID                 : Factor w/ 12 levels "NIWO_040","NIWO_041",..: 8 8 8 8 8 8 8 8 11 11 ...
## $ trapID                 : Factor w/ 12 levels "NIWO_040_205",..: 8 8 8 8 8 8 8 8 11 11 ...
## $ weighDate              : Factor w/ 2 levels "2018-08-06","2018-09-05": 1 1 1 1 1 1 1 1 1 1 ...
## $ setDate                : Factor w/ 2 levels "2018-07-05","2018-08-02": 1 1 1 1 1 1 1 1 1 1 ...
## $ collectDate            : Factor w/ 2 levels "2018-08-02","2018-08-30": 1 1 1 1 1 1 1 1 1 1 ...
## $ ovenStartDate          : Factor w/ 2 levels "2018-08-02T21:00Z",..: 1 1 1 1 1 1 1 1 1 1 ...
## $ ovenEndDate            : Factor w/ 2 levels "2018-08-06T18:02Z",..: 1 1 1 1 1 1 1 1 1 1 ...
## $ fieldSampleID          : Factor w/ 23 levels "NEON.LTR.NIWO040205.20180802",..: 14 14 14 14 14 14
## $ massSampleID           : Factor w/ 168 levels "NEON.LTR.NIWO040205.20180802.FLR",..: 102 101 103 9
## $ samplingProtocolVersion: Factor w/ 1 level "NEON.DOC.001710vE": 1 1 1 1 1 1 1 1 1 1 ...
## $ functionalGroup        : Factor w/ 8 levels "Flowers","Leaves",..: 7 6 8 1 8 4 5 2 1 8 ...
## $ dryMass                : num  0.4 0.005 0.04 0.005 0.07 1 0.2 0.005 0.19 1.18 ...
## $ qaDryMass              : Factor w/ 2 levels "N","Y": 1 1 2 1 1 1 1 1 1 2 ...
## $ remarks                : logi  NA NA NA NA NA NA ...
## $ measuredBy             : Factor w/ 2 levels "kstyers@battelleecology.org",..: 1 1 1 1 1 1 1 1 1
```

```
#Data view
# View(Litter)
# View(Neonics)
```

## Learn about your system

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

Answer: Neonicotinoids are a group of insecticides farmers use for pest control. It is known for having a low effect on mammals but is very toxic to pollinators, beneficial insects, and aquatic invertebrates. They are interested in ecotoxicology because its widespread use can have devastating effects on different ecosystems, destroying the population of pollinators and relevant insects that are vital for specific environments. (references https://www.xerces.org/)

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: According to the USDA, woody debris is an essential part of forest ecosystems because it has a relevant role in carbon budgets and the nutrient cycling of the soil in the forest." it is a source of energy for aquatic ecosystems, provides habitat for terrestrial and aquatic organisms, and contributes to structure and roughness, thereby influencing water flows and sediment."(USDA) It is a relevant source to understand how forest function and how they contribute to the global carbon cycle. (Retrieved from https://research.fs.usda.gov/treesearch/20001#:~:text=Woody%20debris%20is%20an%20important,influencing%20water%20flows%20and%20sediment)

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

Answer: 1.The sampling related with the type of litter collected. It is conducted using both Elevated and Ground Traps where the elevated are used to collect smaller litterfall, while the gorund traps are used for larger fine woody debris. 2. Sampling based on temporal variation by vegetation type. Ground traps are sampled once per year, and elevated traps are divided by Deciduous forests and Evergreen forests, where the first ones are sampled every 1-2 months year-round, and the second ones are sampled every two weeks during leaf senescence. 3.Stratification and ramdomizated spatial data: Sampling is conducted at terrestrial NEON sites containing woody vegetation greater than 2 m in height. Tower plot locations are randomly selected within the 90% flux footprint of the primary and secondary airsheds. At sites with >50% aboveground cover of woody vegetation, trap placement is random and uses a random grid. At sites with <50% cover, traps are placed beneath vegetation patches.

## Obtain basic summaries of your data (Neonics)

5. What are the dimensions of the dataset?

```r
#To see al characteristics of the data, type of data,
#number of observations, and number of variables.
str(Neonics)
```

```
## 'data.frame':    4623 obs. of  30 variables:
##  $ CAS.Number                      : int  58842209 58842209 58842209 58842209 58842209 58842209 5884
##  $ Chemical.Name                   : Factor w/ 9 levels "(1E)-N-[(6-Chloro-3-pyridinyl)methyl]-N-ethy
##  $ Chemical.Grade                  : Factor w/ 9 levels "Analytical grade",..: 9 9 9 9 9 9 9 9 9 9 .
##  $ Chemical.Analysis.Method        : Factor w/ 5 levels "Measured","Not coded",..: 4 4 4 4 4 4 4 4 4
##  $ Chemical.Purity                 : Factor w/ 80 levels ">=98",">=99.0",..: 69 69 50 50 50 50 50 50
##  $ Species.Scientific.Name         : Factor w/ 398 levels "Acalolepta vastator",..: 69 69 248 248 24
##  $ Species.Common.Name             : Factor w/ 303 levels "Alfalfa Leafcutter Bee",..: 74 74 142 142
##  $ Species.Group                   : Factor w/ 4 levels "Insects/Spiders",..: 1 1 1 1 1 1 1 1 1 1 ..
##  $ Organism.Lifestage              : Factor w/ 20 levels "Adult","Cocoon",..: 1 1 19 19 19 1 19 1 1
##  $ Organism.Age                    : Factor w/ 39 levels "<=24","<=48",..: 39 39 39 39 39 36 39 36 3
##  $ Organism.Age.Units              : Factor w/ 11 levels "Day(s)","Days post-emergence",..: 9 9 4 4
##  $ Exposure.Type                   : Factor w/ 24 levels "Choice","Dermal",..: 23 23 11 11 11 11 11
##  $ Media.Type                      : Factor w/ 10 levels "Agar","Artificial soil",..: 7 7 3 3 3 3 3
##  $ Test.Location                   : Factor w/ 4 levels "Field artificial",..: 4 4 4 4 4 4 4 4 4 4 .
##  $ Number.of.Doses                 : Factor w/ 30 levels "' 4-5"," ' 4-7",..: 30 30 18 18 18 18 18 18
##  $ Conc.1.Type..Author.            : Factor w/ 3 levels "Active ingredient",..: 1 1 1 1 1 1 1 1 1 1
##  $ Conc.1..Author.                 : Factor w/ 1006 levels "<0.0004","<0.025",..: 639 510 813 622 44
##  $ Conc.1.Units..Author.           : Factor w/ 148 levels "%","% v/v","% w/v",..: 132 132 91 91 91 9
##  $ Effect                          : Factor w/ 19 levels "Accumulation",..: 16 16 16 16 16 16 16 16
##  $ Effect.Measurement              : Factor w/ 155 levels "Abundance","Accuracy of learned task, per
##  $ Endpoint                        : Factor w/ 28 levels "EC10","EC50",..: 15 15 8 8 8 8 8 8 8 8 ...
##  $ Response.Site                   : Factor w/ 19 levels "Abdomen","Brain",..: 14 14 14 14 14 14 14
##  $ Observed.Duration..Days.        : Factor w/ 361 levels "<.0002","<.0021",..: 145 145 145 145 145
##  $ Observed.Duration.Units..Days.  : Factor w/ 17 levels "Day(s)","Day(s) post-emergence",..: 1 1 1
##  $ Author                          : Factor w/ 433 levels "Abbott,V.A., J.L. Nadeau, H.A. Higo, and
##  $ Reference.Number                : int  107388 107388 103312 103312 103312 103312 103312 103312 10
##  $ Title                           : Factor w/ 458 levels "A Common Pesticide Decreases Foraging Suc
##  $ Source                          : Factor w/ 456 levels "Acta Hortic.1094:451-456",..: 295 295 296
##  $ Publication.Year                : int  1982 1982 1986 1986 1986 1986 1986 1986 1986 1986 ...
##  $ Summary.of.Additional.Parameters: Factor w/ 943 levels "Purity: \xca NC - NC | Organism Age: \xca
```

```r
#To see the number of observations and number of variables
dim(Neonics)
```

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

```r
#4623 obs. of  30 variables
```

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? [Tip: The `sort()` command is useful for listing the values in order of magnitude...]

```r
#data available in the Effect column
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 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 Formulation I
##  (Other)
```

```r
summary(sort(Neonics$Effect, decreasing = TRUE)) #To review the
```

```
##     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
#To review and organize the data available in the Effect column
sort(table(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
```

```r
summary(sort(Neonics$Species.Common.Name, decreasing = TRUE))
```

```
##                            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
```

```r
#Example nesting a sort function inside the summary to get it to sort the bugs
#in order of observations by their common name.
```

Answer: According to the data this column is to follow the diferent reaction that insects had to certain types of chemicals. The most common effects studied Population,Mortality, Behavior, Feeding behavior and Reproduction

7. Using the `summary` function, determine the six most commonly studied species in the dataset (common name). What do these species have in common, and why might they be of interest over other insects? Feel free to do a brief internet search for more information if needed.[TIP: Explore the help on the `summary()` function, in particular the `maxsum` argument…]

```r
summary(Neonics$Species.Common.Name, maxsum = 6)
```

```
##            Honey Bee        Parasitic Wasp Buff Tailed Bumblebee
##                  667                   285                   183
##   Carniolan Honey Bee            Bumble Bee               (Other)
##                  152                   140                  3196
```

```r
# calling a summary with the added function that it take the top ten
# names as rated by the number of observations and rank them.
```

Answer: 1) Honey Bee 2)Parasitic Wasp 3)Buff 4)Tailed Bumblebee 5)Carniolan 6) Bumble Bee

8. Concentrations are always a numeric value. What is the class of `Conc.1..Author.` column in the dataset, and why is it not numeric? [Tip: Viewing the dataframe may be helpful…]

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

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

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

```
##    0.37/      10/      NR/       NR        1     1023    0.40/       2/
##      208      127      108       94       82       80       69       63
##       10   0.053/      100      50/     0.5/     0.03    0.05/     0.45
##       62       59       56       51       45       44       43       43
##     0.1/    0.45/     1.0/    2.27/       50    0.125     500/      0.5
##       42       40       40       40       36       33       33       32
##   0.048/    0.15/       1/       48    25.0/      12/    0.027      2.4
##       30       30       30       30       28       27       26       26
##     0.2/    0.56/     100/        3    0.01/    1000/       3/    0.336
##       25       24       23       23       22       22       22       21
##     1.5/     0.05      1.5    2.60/    20.0/        6    6.80/    62.5/
##       21       20       20       20       20       20       20       20
##    0.005     0.4/    0.18/     0.3/     1000       40 0.00355/      0.1
##       18       18       17       17       17       17       16       16
##      0.4     150/      300      80/    0.053     0.24     0.28     125/
##       16       16       16       16       15       15       15       15
##        9   0.0001  0.0004/   0.084/     0.15      0.6    12.5/   144.0/
##       15       14       14       14       14       14       14       14
##     350/    40.0/      48/       56      84/    0.17/      125       14
##       14       14       14       14       14       13       13       13
##       16       17   0.047/    0.25/    0.28/    1.28/    1.81/      112
##       13       13       12       12       12       12       12       12
##      150     2.5/       25      60/      75/    0.02/   0.025/     0.29
##       12       12       12       12       12       11       11       11
##    37.5/       4/        5  (Other)
##       11       11       11     1817
```

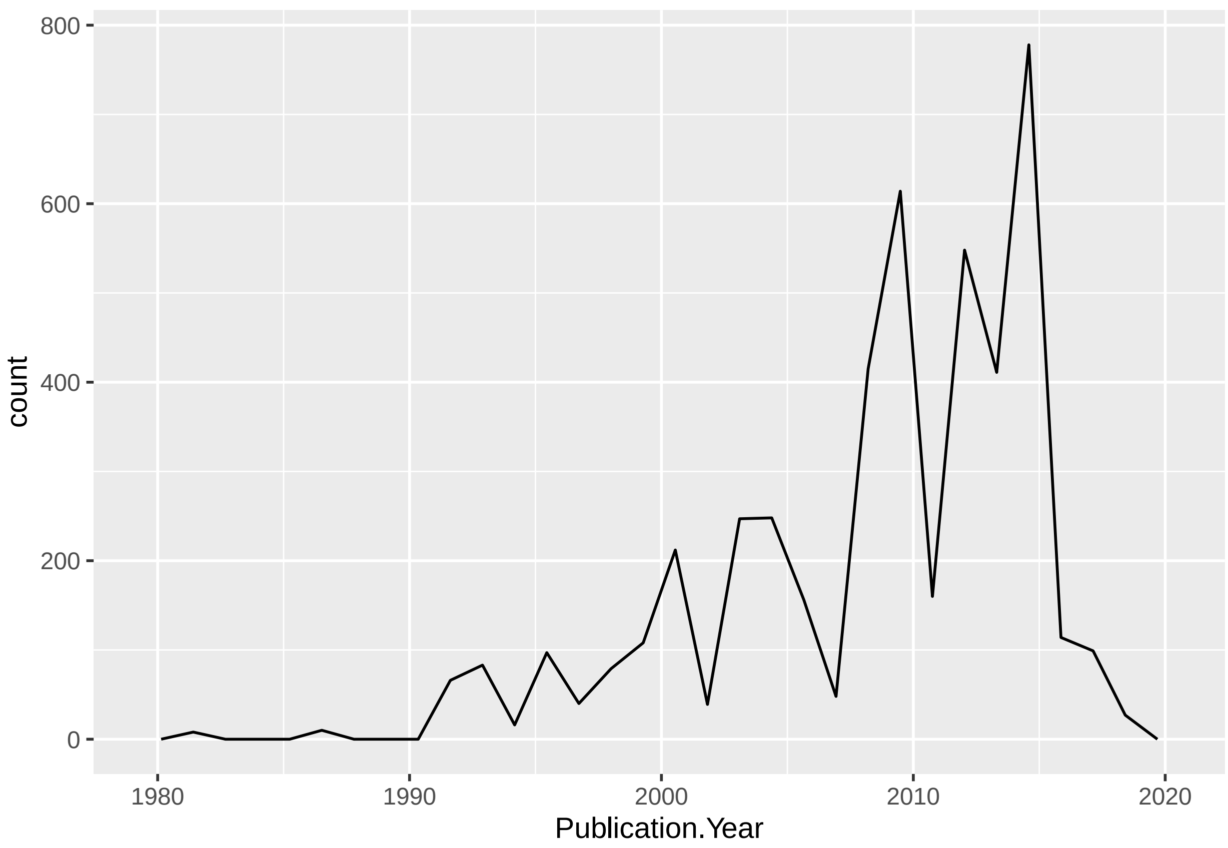
Answer: When summarizing, the “Conc.1..Author” column has different characters that R takes as factors. The presence of NR and / in different numbers makes the program understand that the data are factors and not numbers.

## Explore your data graphically (Neonics)

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

```r
ggplot(Neonics, aes(x = Publication.Year)) +
  geom_freqpoly() ##to generate a frequency polygon that has 'Publication.Year' on the X axis
```

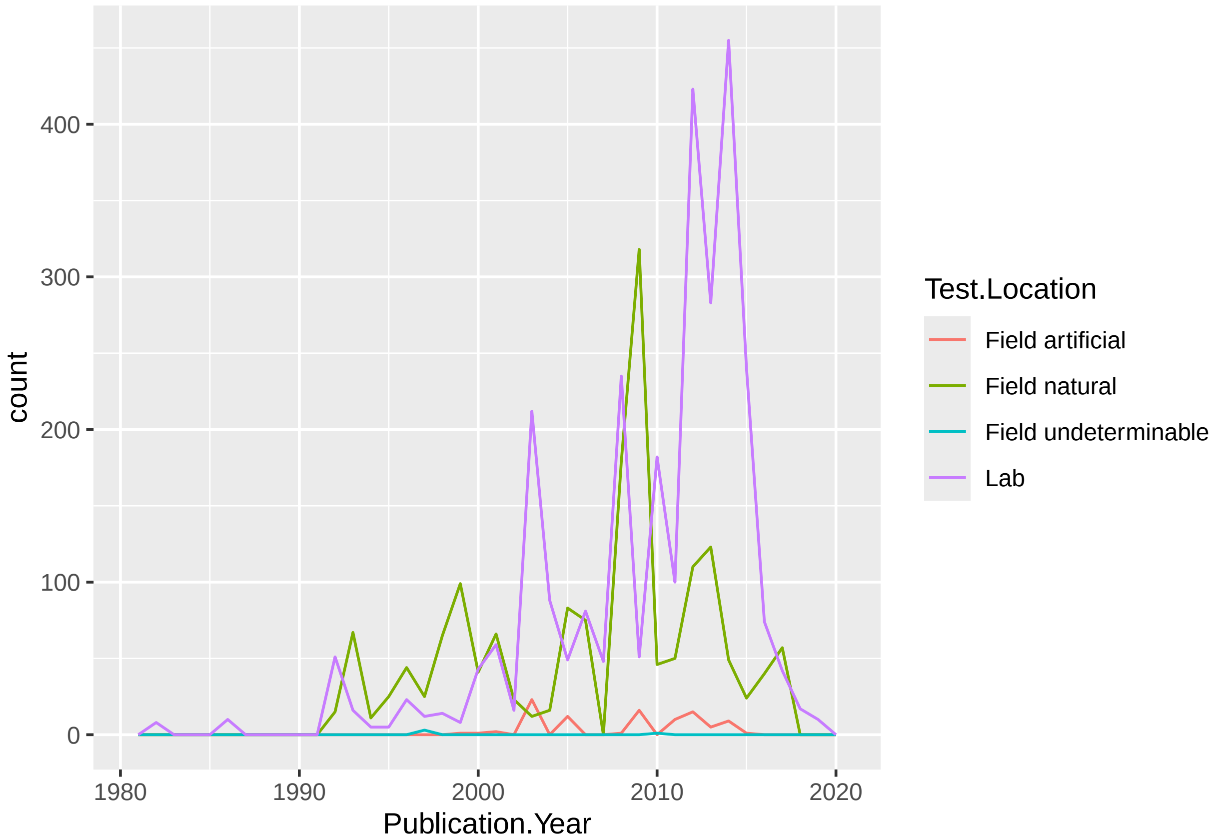
```
## `stat_bin()` using `bins = 30`. Pick better value with `binwidth`.
```




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

```
ggplot(Neonics, aes(x = Publication.Year, color = Test.Location)) +
  geom_freqpoly(binwidth = 1, size = 0.5)#to generate a frequency polygon that has 'Publication.Year' o
```

```
## Warning: Using ‘size‘ aesthetic for lines was deprecated in ggplot2 3.4.0.
## i Please use ‘linewidth‘ instead.
## This warning is displayed once every 8 hours.
## Call ‘lifecycle::last_lifecycle_warnings()‘ to see where this warning was
## generated.
```



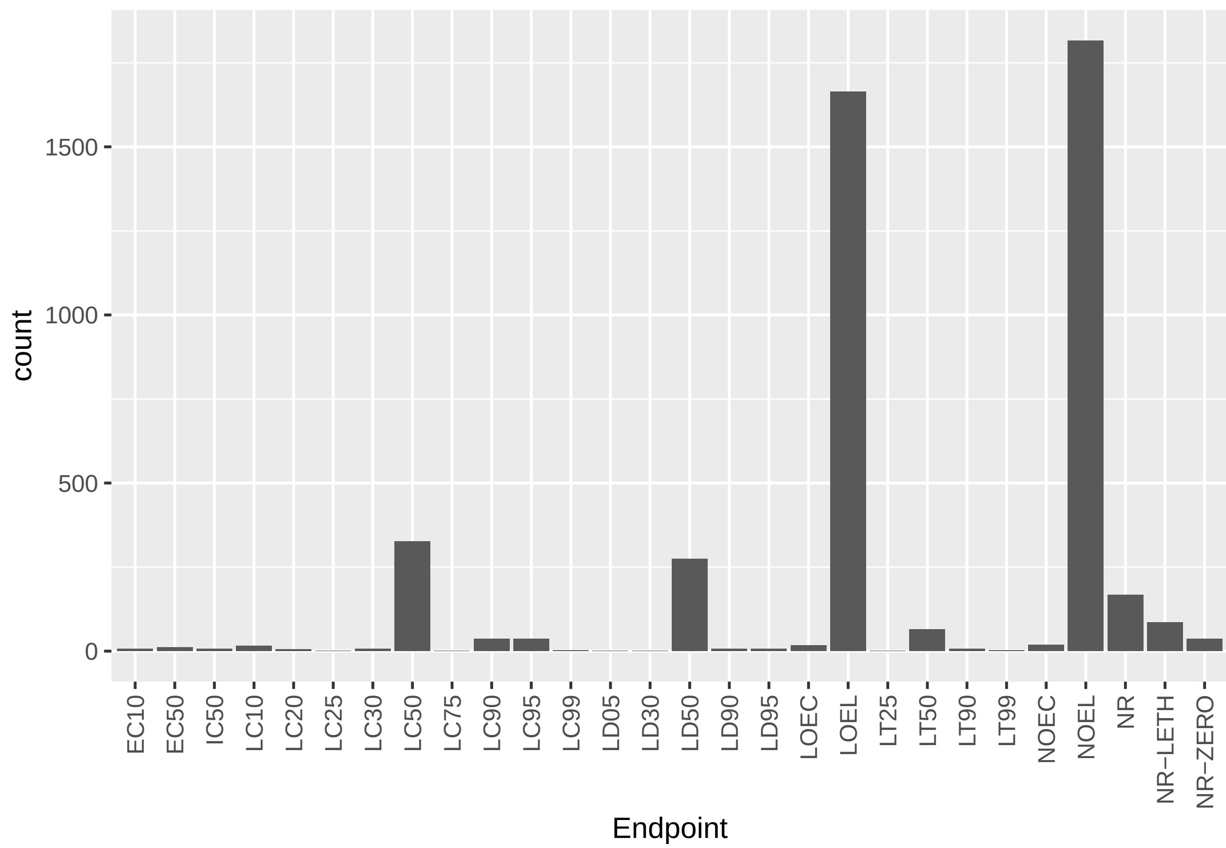

Interpret this graph. What are the most common test locations, and do they differ over time?

Answer: The graph illustrates that from 1990 to 2000, the most common test location was the natural field. However, starting in the early 2000s, this trend began to shift, with laboratories gaining prominence. This change became most pronounced during the decade from 2010 to 2020, when laboratory testing reached its peak.

11. Create a bar graph of Endpoint counts. What are the two most common end points, and how are they defined? Consult the ECOTOX_CodeAppendix for more information.

[**TIP**: Add `theme(axis.text.x = element_text(angle = 90, vjust = 0.5, hjust=1))` to the end of your plot command to rotate and align the X-axis labels…]

```
ggplot(Neonics, aes(x = Endpoint,)) +
  geom_bar()+ theme(axis.text.x = element_text(angle = 90, vjust = 0.5, hjust=1))
```




```
#ggplot to create a graphic of Endpoint.
#geom_bar() to ask R that the graphic must be a bar graph
#`theme(axis.text.x = element_text(angle = 90, vjust = 0.5, hjust=1))`
#to give characteristics to the graph
```

Answer: NOEL anf LOEL are the two most common end points. NOEL: No-observable-effect-level: highest dose (concentration) producing effects not significantly different from responses of controls according to author's reported statistical test (NOEAL/NOEC) LOEL: 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.

```
#Asking R to give me a summary of the different columns
str(Litter)
```

```
## 'data.frame':    188 obs. of  19 variables:
##  $ uid                    : Factor w/ 188 levels "028eea3d-5c20-4afc-bb7e-a05bab305152",..: 84 96 85
##  $ namedLocation          : Factor w/ 12 levels "NIWO_040.basePlot.ltr",..: 8 8 8 8 8 8 8 8 11 11 ..
##  $ domainID               : Factor w/ 1 level "D13": 1 1 1 1 1 1 1 1 1 1 ...
##  $ siteID                 : Factor w/ 1 level "NIWO": 1 1 1 1 1 1 1 1 1 1 ...
```

```
##  $ plotID                  : Factor w/ 12 levels "NIWO_040","NIWO_041",..: 8 8 8 8 8 8 8 8 11 11 ...
##  $ trapID                  : Factor w/ 12 levels "NIWO_040_205",..: 8 8 8 8 8 8 8 8 11 11 ...
##  $ weighDate               : Factor w/ 2 levels "2018-08-06","2018-09-05": 1 1 1 1 1 1 1 1 1 1 ...
##  $ setDate                 : Factor w/ 2 levels "2018-07-05","2018-08-02": 1 1 1 1 1 1 1 1 1 1 ...
##  $ collectDate             : Factor w/ 2 levels "2018-08-02","2018-08-30": 1 1 1 1 1 1 1 1 1 1 ...
##  $ ovenStartDate           : Factor w/ 2 levels "2018-08-02T21:00Z",..: 1 1 1 1 1 1 1 1 1 1 ...
##  $ ovenEndDate             : Factor w/ 2 levels "2018-08-06T18:02Z",..: 1 1 1 1 1 1 1 1 1 1 ...
##  $ fieldSampleID           : Factor w/ 23 levels "NEON.LTR.NIWO040205.20180802",..: 14 14 14 14 14 14
##  $ massSampleID            : Factor w/ 168 levels "NEON.LTR.NIWO040205.20180802.FLR",..: 102 101 103 9
##  $ samplingProtocolVersion: Factor w/ 1 level "NEON.DOC.001710vE": 1 1 1 1 1 1 1 1 1 1 ...
##  $ functionalGroup         : Factor w/ 8 levels "Flowers","Leaves",..: 7 6 8 1 8 4 5 2 1 8 ...
##  $ dryMass                 : num  0.4 0.005 0.04 0.005 0.07 1 0.2 0.005 0.19 1.18 ...
##  $ qaDryMass               : Factor w/ 2 levels "N","Y": 1 1 2 1 1 1 1 1 1 2 ...
##  $ remarks                 : logi  NA NA NA NA NA NA ...
##  $ measuredBy              : Factor w/ 2 levels "kstyers@battelleecology.org",..: 1 1 1 1 1 1 1 1 1 1
```

```r
# #Asking R to give me specyfic information about collectDate
class(Litter$collectDate)
```

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

```r
#Changing the format
Litter$collectDate <- as.Date(Litter$collectDate, format = "%Y-%m-%d")
class(Litter$collectDate)#Verifying the change
```

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

```r
# format(Litter$collectDate, "%Y-%m") == "2018-08" filter dates sampled in August 2018
unique_dates <- unique(Litter$collectDate[format(Litter$collectDate, "%Y-%m") == "2018-08"])


#unique_dates <- unique(Litter$collectDate) create a characteristic.
print(unique_dates)
```

```
## [1] "2018-08-02" "2018-08-30"
```

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

```r
unique(Litter$plotID) #to show different Unique values in plotID
```

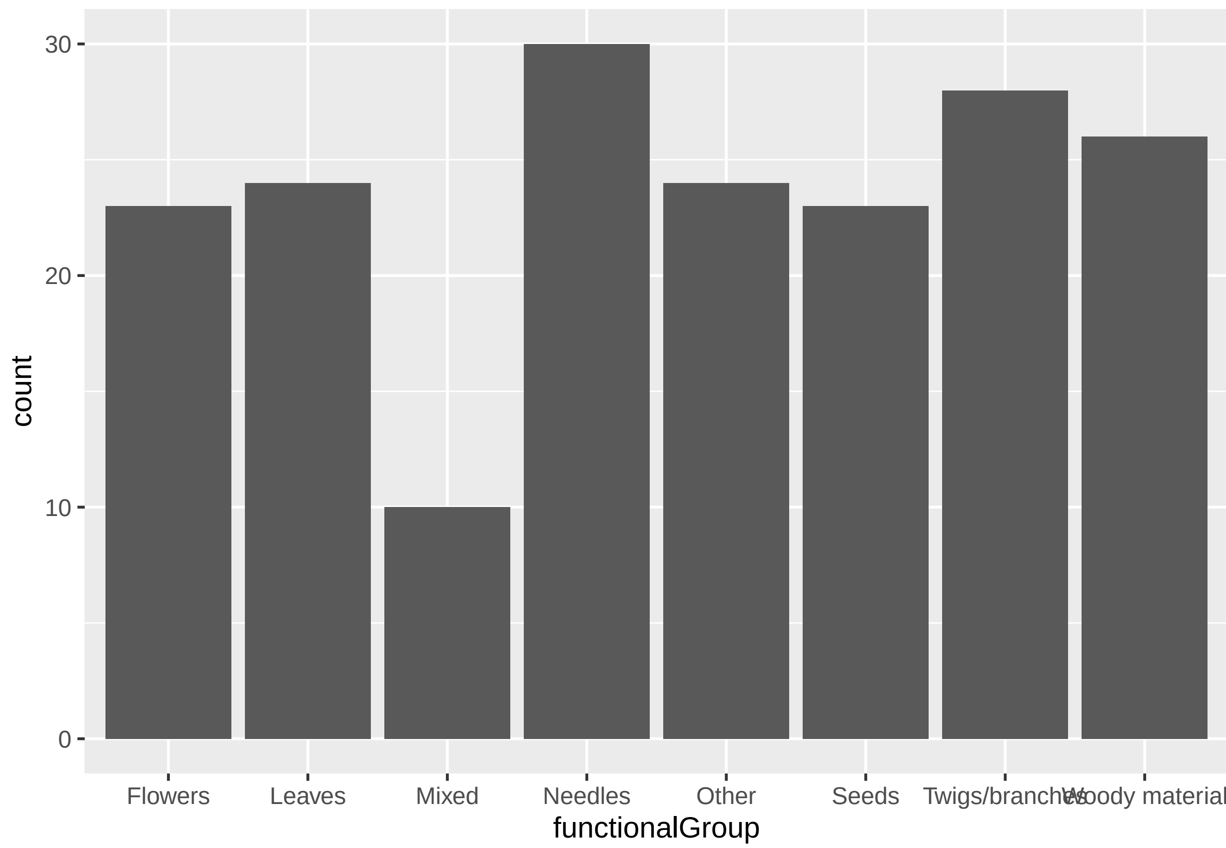
```
##  [1] NIWO_061 NIWO_064 NIWO_067 NIWO_040 NIWO_041 NIWO_063 NIWO_047 NIWO_051
##  [9] NIWO_058 NIWO_046 NIWO_062 NIWO_057
## 12 Levels: NIWO_040 NIWO_041 NIWO_046 NIWO_047 NIWO_051 NIWO_057 ... NIWO_067
```

Answer:

It returned the distinct (non-duplicate) values from plotID. It shows 12 unique plots. The difference betwen the `unique` function and `summary`is that `unique`brings just the non-duplicate values, and `summary`shows all the values and its characteristics.

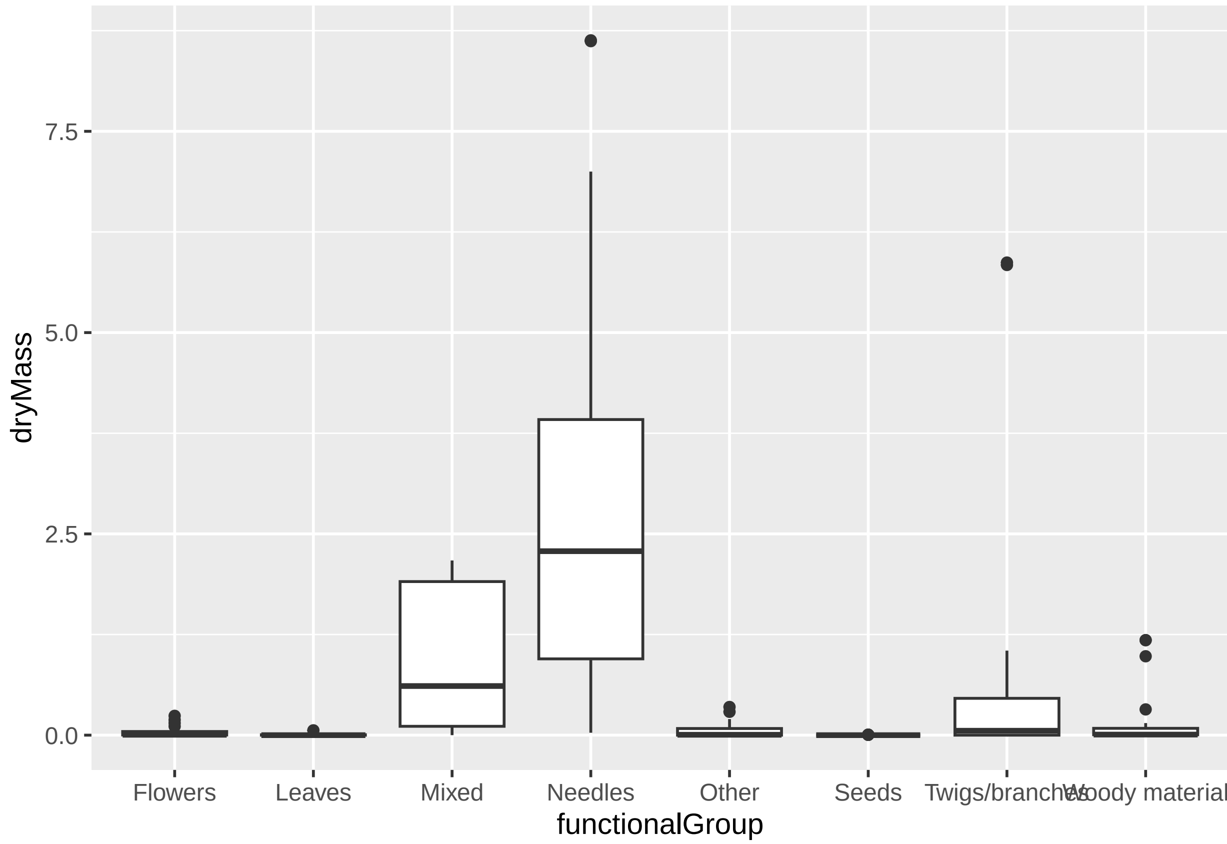
14. Create a bar graph of functionalGroup counts. This shows you what type of litter is collected at the Niwot Ridge sites. Notice that litter types are fairly equally distributed across the Niwot Ridge sites.

```
ggplot(Litter, aes(x = functionalGroup)) +  geom_bar() #Create a simple chart bar
```

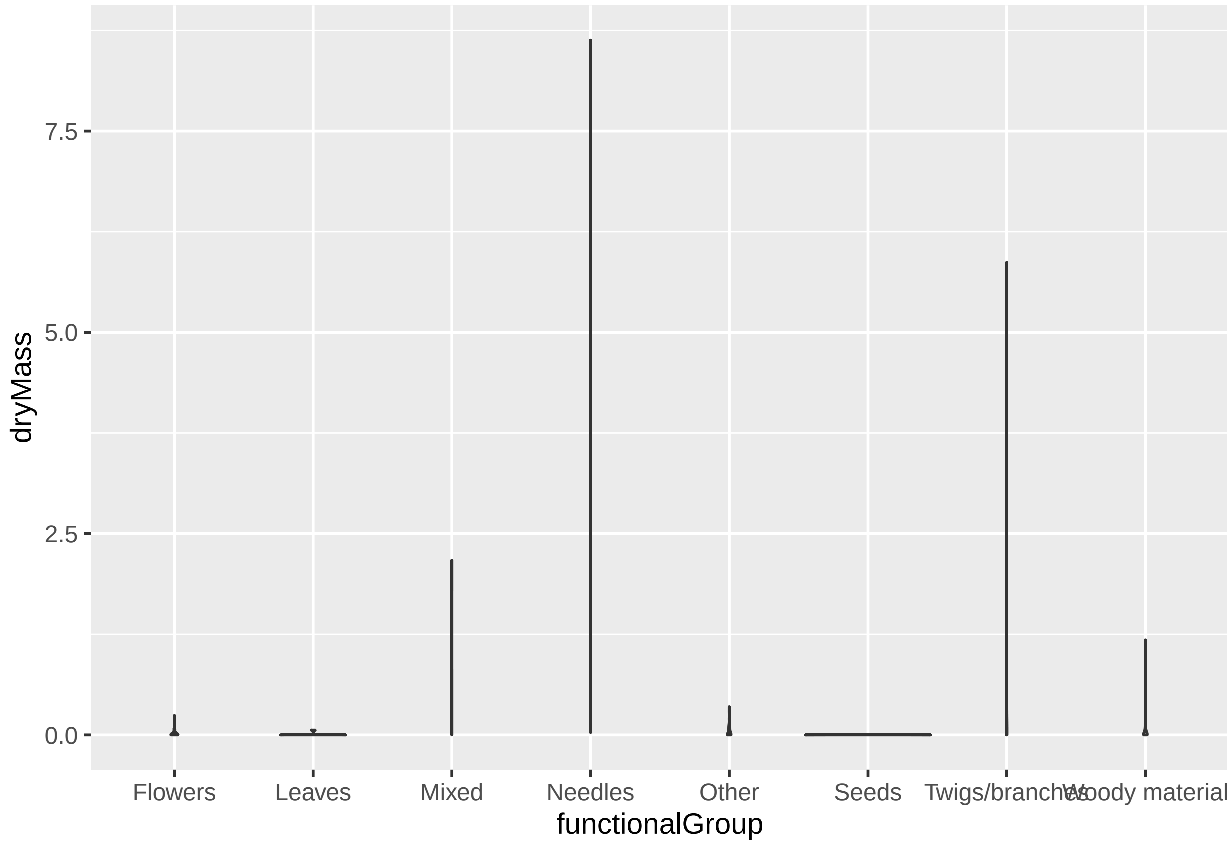



15. Using `geom_boxplot` and `geom_violin`, create a boxplot and a violin plot of dryMass by functionalGroup.

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




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




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

Answer: Because in this case box plot gives more descriptive information, it shows the distribution but at the same time it shows the quartiles and it is better to compare. The violin just gives information of distribution.

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

Answer: Needles, Mixed and twigs.