Assignment 3: Data Exploration

Emma Brentjens

OVERVIEW

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

Directions

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

The completed exercise is due on Sept 30th.

Set up your R session

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

```
#1.
##checking working directory
getwd()
```

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

```
##setting working directory to raw data folder (???????)
#setwd("/home/guest/R/EDA-Fall2022/Data/Raw")

##loading tidyverse package
library(tidyverse)
library(ggplot2)

##uploading and naming datasets
Neonics <- read.csv("Copyof_ECOTOX_Neonicotinoids_Insects_raw.csv", stringsAsFactors = T)
Litter <- read.csv("Copyof_NEON_NIWO_Litter_massdata_2018-08_raw.csv", stringsAsFactors = T)
#####this seems to only work when running in console when I tried to set wd to raw data folder
```

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: This research can provide information on the efficacy of neonicotinoids to protect crops from insect pests. It is important to understand which doses are effective as using too little of the insecticide would not produce the desired effect but using too much may be harmful to nontarget species, like bees (Texas A&M AgriLife Extension, n.d.). Citation: Texas A&M AgriLife Extension. (n.d.). What is a neonicotinoid? Insects in the City. Retrieved September 26, 2022, from https://citybugs.tamu.edu/factsheets/ipm/what-is-a-neonicotinoid/

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 amount of leaf litter and debris on the groun can influence soil moisture and nutrient cycling. As leaf litter decomposes, it is broken down into compounds that plants can absorb as they grow (Giweta, 2020). Leaf litter types can also provide information about plant composition in ecosystems (Giweta, 2020). Citation: Giweta, M. (2020). Role of litter production and its decomposition, and factors affecting the processes in a tropical forest ecosystem: A review. Journal of Ecology and Environment, 44(1), 11. https://doi.org/10.1186/s41610-020-0151-2

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. 2. 3.

Obtain basic summaries of your data (Neonics)

5. What are the dimensions of the dataset?

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

```r
##finding number of rows and columns
nrow(Neonics)
```

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

```r
ncol(Neonics)
```

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

```r
##finding the class of the dataset
class(Neonics)
```

```
## [1] "data.frame"
```

```r
#5.
##The "Neonics" dataframe has 30 columns and 4623 rows/observations
```

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?

```r
#6.
##obtaining summary statistics of effects studied
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: Most common effects studied: 1. Population (1803) 2. Mortality (1493) 3. Behavior (360) The researchers are interested in these effects because they are measures of how the insecticides impact insects. Population abundance and mortality demonstrate how insecticides impact insect viability (yes??) while insect behavior could include important life history activities like feeding and mating.

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.

```
#7.
##obtaining summary statistics of species studied
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: Most common species studied: 1. Honey Bee (667) 2. Parasitic Wasp (285) 3. Buff

Tailed Bumblebee (183) [interest over other insects??????????????]

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

```
#8.
##determining class of "Conc.1..Author" column
class(Neonics$Conc.1..Author.)
```

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

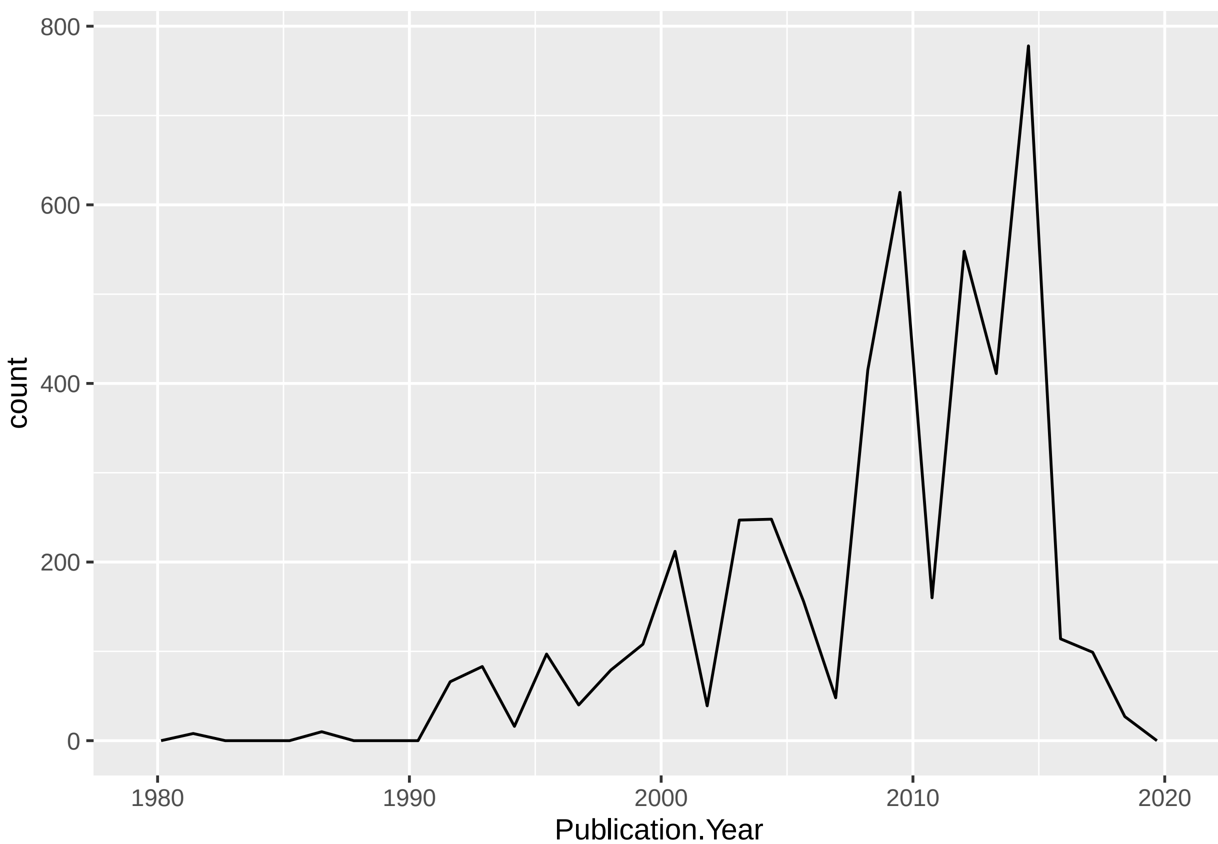
Answer: The class of "Conc.1..Author" is "factor." [why not numeric? not sure what column means???]

Explore your data graphically (Neonics)

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

```
#9
##creating plot of studies by publication year
studies_by_year <- ggplot(data=Neonics, aes(x=Publication.Year))+
  geom_freqpoly()
studies_by_year
```

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

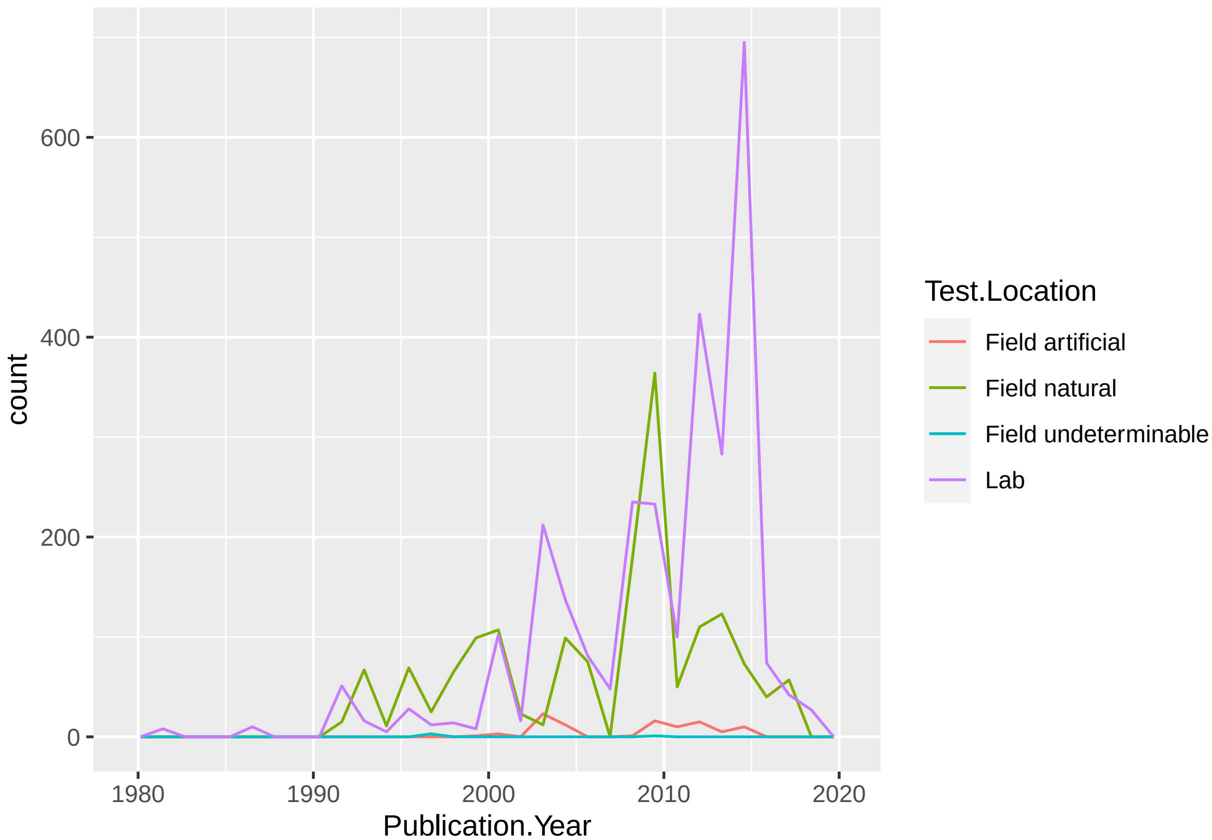



```
##change bin width???????
```

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

```
#10
##creating plot of studies by publication year and test location
studies_by_year2 <- ggplot(data=Neonics, aes(x=Publication.Year, color=Test.Location))+
  geom_freqpoly()

studies_by_year2
```

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

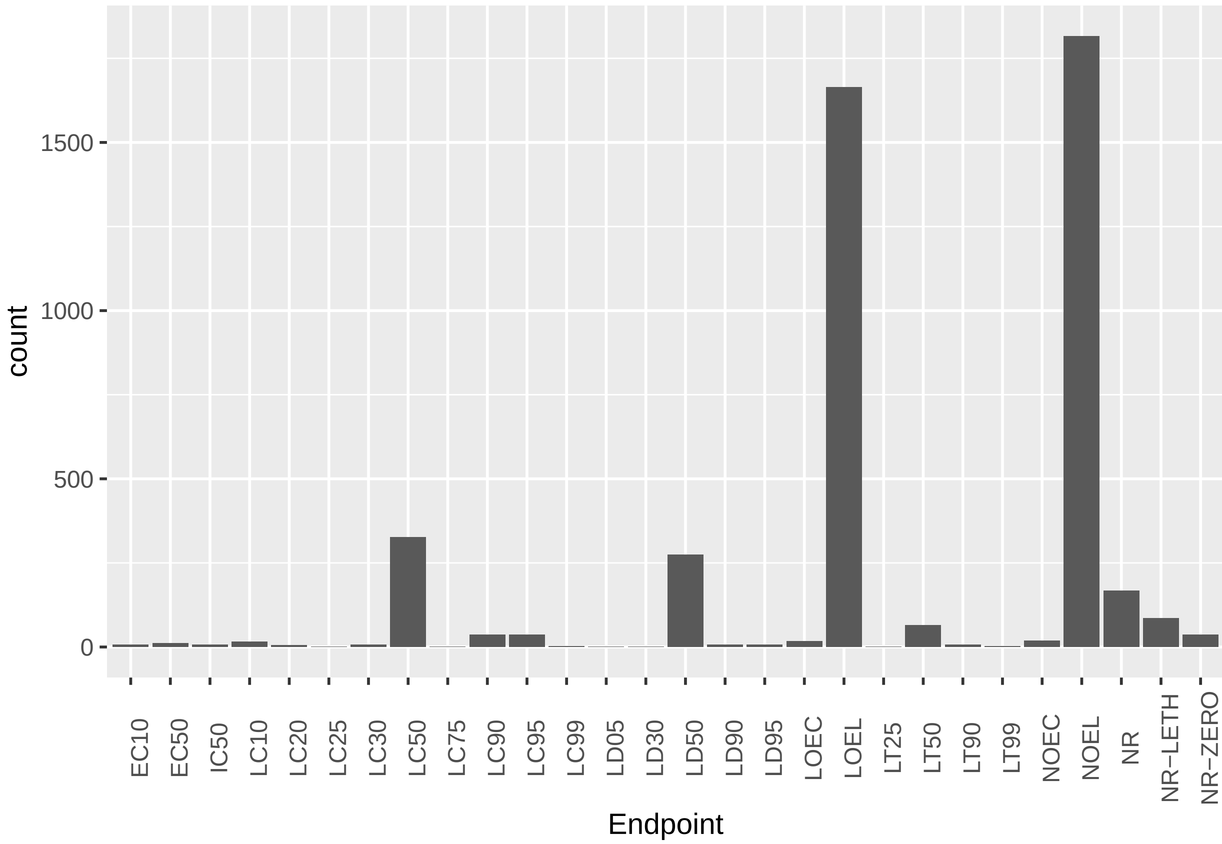


Interpret this graph. What are the most common test locations, and do they differ over time?

> Answer: The most common test locations are labs and natural fields. The number of lab studies generally seems to have increased over time (before 2020) while the number of natural field studies peaked at around 2010 and has declined since then (perhaps due to increasing popularity of lab tests).

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.

```
#11
##creating bar graph for endpoint counts
endpoint_counts_graph <- ggplot(data=Neonics, aes(x=Endpoint))+
  geom_bar()+
  theme(axis.text.x=element_text(angle=90))

##output for endpoint counts graph
endpoint_counts_graph
```



Answer: The two most common endpoints are NOEL (no-observable-effect-level) and LOEL (lowest-observable-effect-level). The NOEL is defined as the greatest concentration of chemical that does not cause an effect significantly different than the control. The LOEL refers to the lowest chemical concentration that causes an effect that varies significantly from the control.

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.

```
#12
##determining the class of the litter collect date
class(Litter$collectDate) ##class = "factor"
```

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

```
Litter$collectDate_date <- as.Date(Litter$collectDate) ##get NAs when specifying format

class(Litter$collectDate_date)
```

```
## [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`?

```
#13.
##determining number of plots sampled at Niwot Ridge
length(unique(Litter$plotID)) ###is this okay??????????
```

```
## [1] 12
```

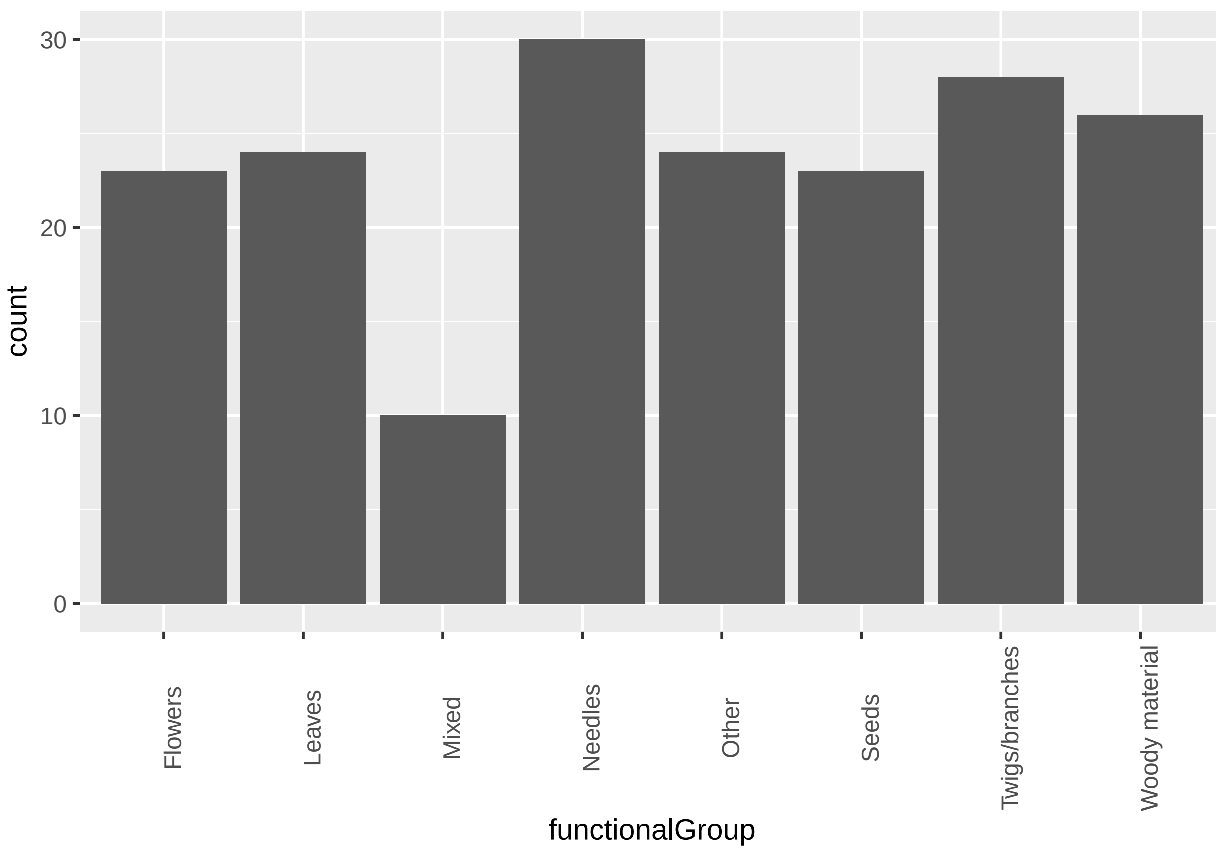
```r
##running summary command on plots
summary(Litter$plotID)
```

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

> Answer: There were 12 plots sampled at Niwot Ridge. A summary of Litter$plotID gives you the number of observations at each plot while the unique function provides the number of plots studied.

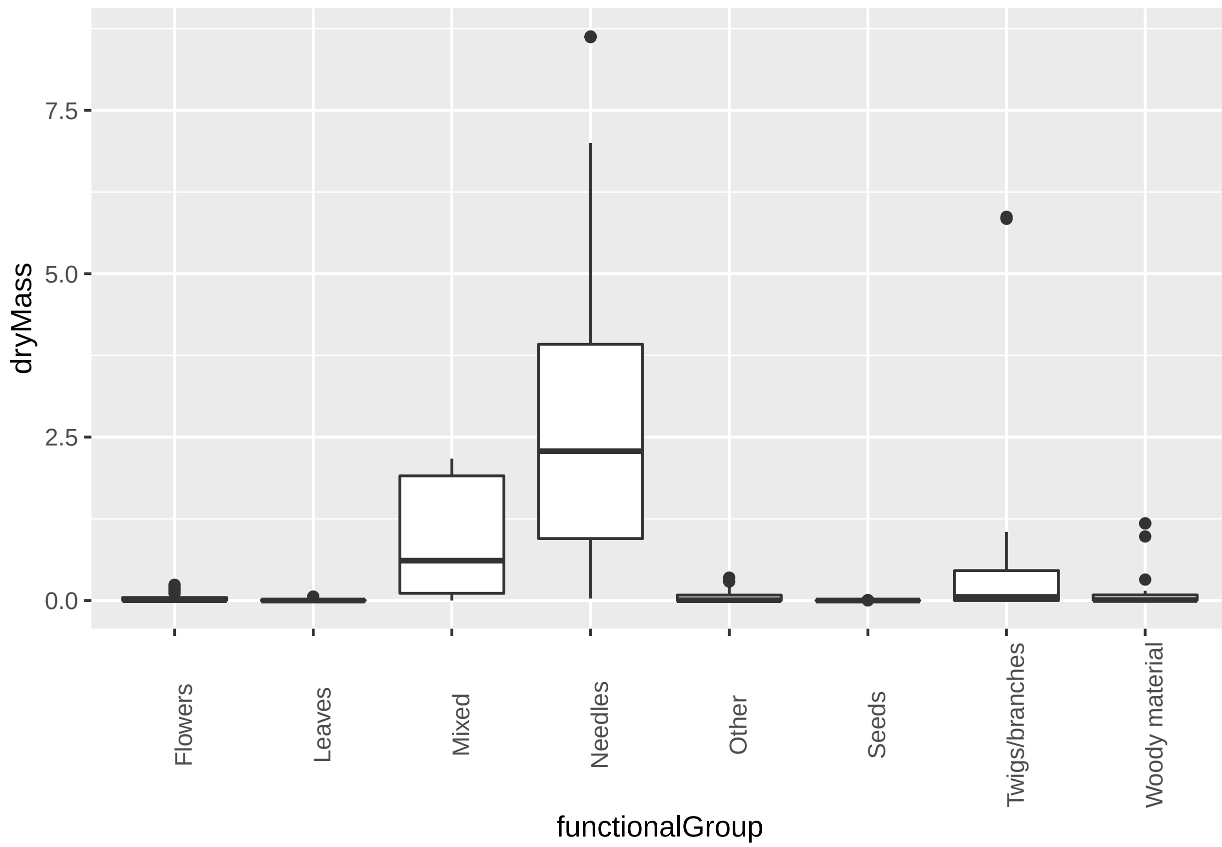
14. Create a bar graph of functionalGroup counts. This shows you what type of litter is collected at the Niwot Ridge sites. Notice that litter types are fairly equally distributed across the Niwot Ridge sites.

```r
#14
##creating functionalGroup counts bar graph
functionalGroup_bar_graph <- ggplot(data=Litter, aes(x=functionalGroup))+
  geom_bar()+
  theme(axis.text.x=element_text(angle=90))

##output for functionalGroup counts bar graph
functionalGroup_bar_graph
```



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

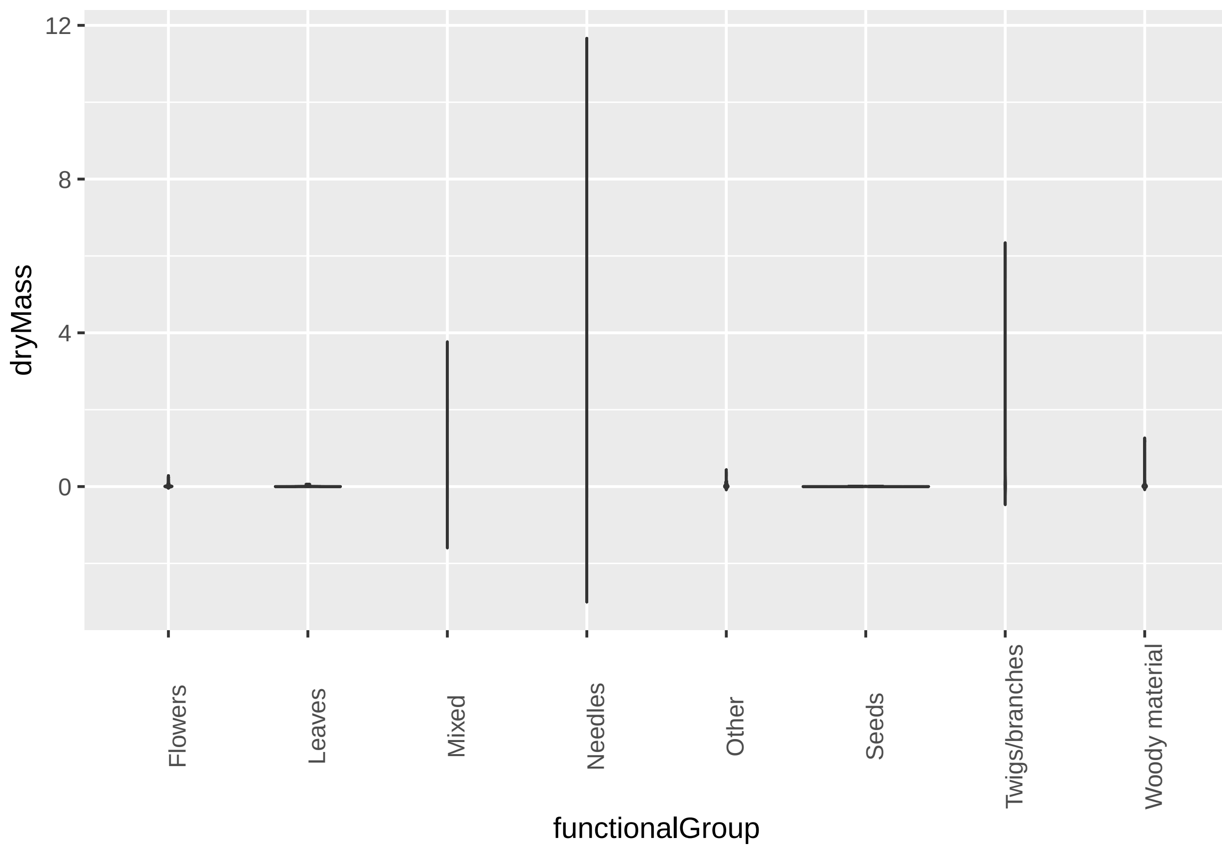
```r
#15
##functionalGroup boxplot
functionalGroup_boxplot <- ggplot(data=Litter, aes(x=functionalGroup, y=dryMass))+
  geom_boxplot()+
  theme(axis.text.x=element_text(angle=90))

##boxplot output
functionalGroup_boxplot
```




```r
##functionalGroup violin plot
functionalGroup_violin <- ggplot(data=Litter, aes(x=functionalGroup, y=dryMass))+
  geom_violin(pt.size = 10, trim=F)+
  theme(axis.text.x=element_text(angle=90))
```

```
## Warning: Ignoring unknown parameters: pt.size
```

```r
##violin plot output
functionalGroup_violin ##violin plot just coming out as vertical lines???????????????
```



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

Answer:

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

Answer: The type of litter with the highest biomass is needles, followed by mixed litter and twigs and branches.