Assignment 3: Data Exploration

Student Name

Fall 2023

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 Sakai.

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. Check your working directory, load necessary packages (tidyverse, lubridate), 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.

```
getwd()
```

```
## [1] "/home/guest/EDE_Fall2023"
```

```
#install.packages("tidyverse")
#install.packages("lubridate")
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 ecotoxicology of neonicotinoids on insects? Feel free to do a brief internet search if you feel you need more background information.

   Answer:Neonictonoids are a class of pesticidies that kill insects by inhibiting their nervous system fuction. Their initial use was to target insects and pests that impact crop production and quality, but the mechanism they target is found widely across insects making it a non target pesticide. From an ecotoxicity perpsective, folks are interested in learning about the wide ranging impact of this pesticide and its effect on pollinator species such as bees.

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:Litter and woody debris play an essential role in the ecosystem by

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
dim(Neonics)
```

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

```r
# 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?

```r
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: Based on the summary Mortality, Population, Behavior, Feeding Behavior, Development and Reproduction are commonly studied endpoints. This makes sense because if a pesticide causes death, mortality and population can be used to track the status of an insect species in the environment. If a pesticide does not cause, it could impact other functions such as reproduction or development which has larger species impact. Lastly, studying Behavior and feeding behavior gives insight into the insects' role in the environment, helping us predict what ecosystems may collapse without their presence.

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: The `sort()` command can sort the output of the summary command...]

```r
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
sort(summary(Neonics$Species.Common.Name))
```

```
##                      Ant Family                    Apple Maggot
##                               9                               9
##          Glasshouse Potato Wasp                        Lacewing
##                              10                              10
##          Southern House Mosquito         Two Spotted Lady Beetle
##                              10                              10
##         Spotless Ladybird Beetle              Braconid Parasitoid
##                              11                              12
##                    Common Thrip     Eastern Subterranean Termite
##                              12                              12
##                          Jassid                      Mite Order
##                              12                              12
##                       Pea Aphid                Pond Wolf Spider
##                              12                              12
##           Armoured Scale Family                 Diamondback Moth
##                              13                              13
##                   Eulophid Wasp                Monarch Butterfly
##                              13                              13
##                   Predatory Bug           Yellow Fever Mosquito
##                              13                              13
##                    Corn Earworm                Green Peach Aphid
##                              14                              14
##                       House Fly                       Ox Beetle
##                              14                              14
##              Red Scale Parasite               Spined Soldier Bug
##                              14                              14
##          Western Flower Thrips Hemlock Woolly Adelgid Lady Beetle
##                              15                              16
##           Hemlock Wooly Adelgid                            Mite
##                              16                              16
##                     Onion Thrip            Araneoid Spider Order
##                              16                              17
##                       Bee Order                  Egg Parasitoid
##                              17                              17
##                    Insect Class         Moth And Butterfly Order
##                              17                              17
##     Oystershell Scale Parasitoid         Black-spotted Lady Beetle
##                              17                              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
##                    Codling Moth        Flatheaded Appletree Borer
##                              19                              20
##              Horned Oak Gall Wasp              Leaf Beetle Family
##                              20                              20
##               Potato Leafhopper       Tooth-necked Fungus Beetle
```

```
##                               20                                 20
##                     Argentine Ant                             Beetle
##                               21                                 21
##                        Mason Bee                           Mosquito
##                               22                                 22
##                  Citrus Leafminer                    Ladybird Beetle
##                               23                                 23
##                Spider/Mite Class                Tobacco Flea Beetle
##                               24                                 24
##                     Chalcid Wasp            Convergent Lady Beetle
##                               25                                 25
##                    Stingless Bee               Ground Beetle Family
##                               25                                 27
##               Rove Beetle Family                      Tobacco Aphid
##                               27                                 27
##                    Scarab Beetle                      Spring Tiphia
##                               29                                 29
##                      Thrip Order            Ladybird Beetle Family
##                               29                                 30
##                       Parasitoid                      Braconid Wasp
##                               30                                 33
##                     Cotton Aphid                     Predatory Mite
##                               33                                 33
##            Sweetpotato Whitefly                       Aphid Family
##                               37                                 38
##                   Cabbage Looper              Buff-tailed Bumblebee
##                               38                                 39
##                   True Bug Order           Sevenspotted Lady Beetle
##                               45                                 46
##                     Beetle Order      Snout Beetle Family, Weevil
##                               47                                 47
##               Erythrina Gall Wasp                    Parasitoid Wasp
##                               49                                 51
##           Colorado Potato Beetle                      Parastic Wasp
##                               57                                 58
##              Asian Citrus Psyllid                  Minute Pirate Bug
##                               60                                 62
##                European Dark Bee                           Wireworm
##                               66                                 69
##                   Euonymus Scale                  Asian Lady Beetle
##                               75                                 76
##                  Japanese Beetle                  Italian Honeybee
##                               94                                113
##                       Bumble Bee                Carniolan Honey Bee
##                              140                                152
##             Buff Tailed Bumblebee                    Parasitic Wasp
##                              183                                285
##                        Honey Bee                            (Other)
##                              667                                670
```

Answer:The 6 most commonly studied species are the Honey Bee, Parisitic Wasp, Buff Tailed Bumble Bee, Carniolan Honey Bee and Italian Honeybee.

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?

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

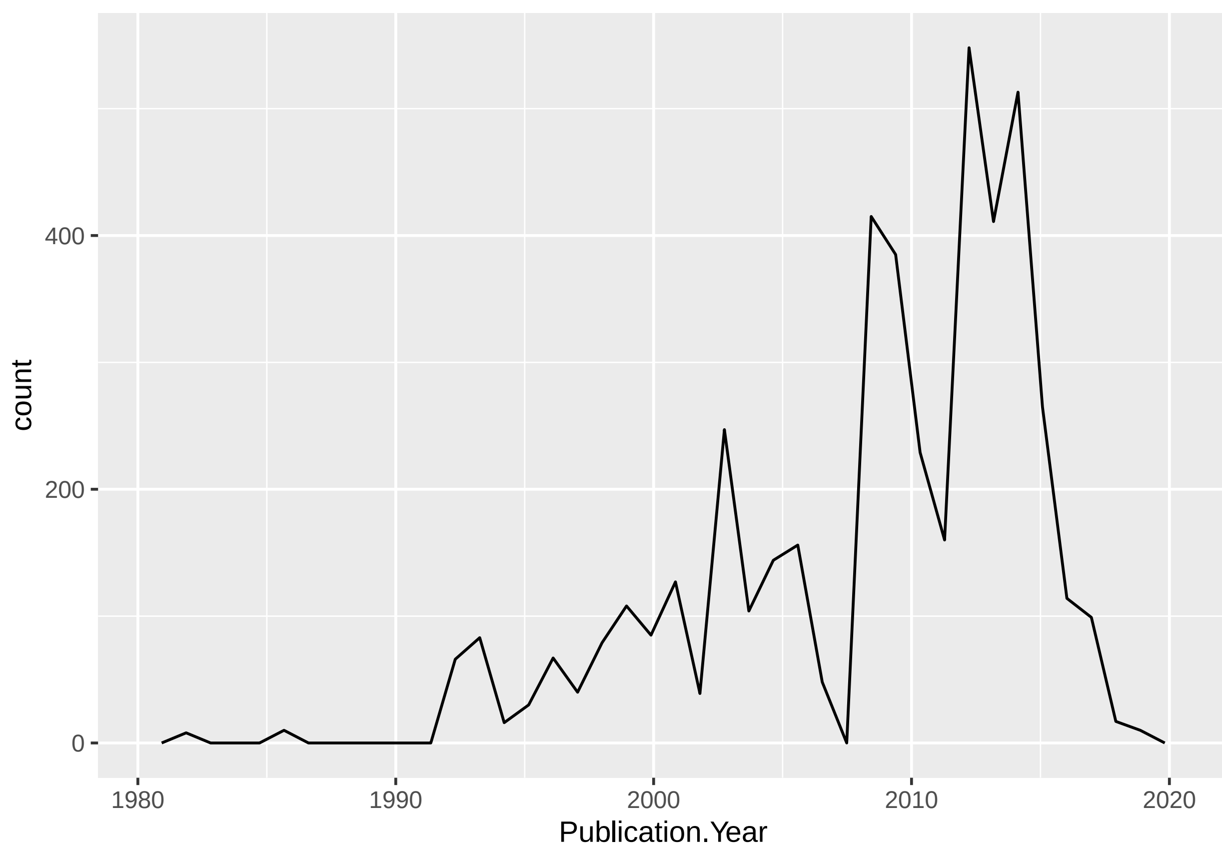
```
## [1] "factor"
```

Answer:

Explore your data graphically (Neonics)

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

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

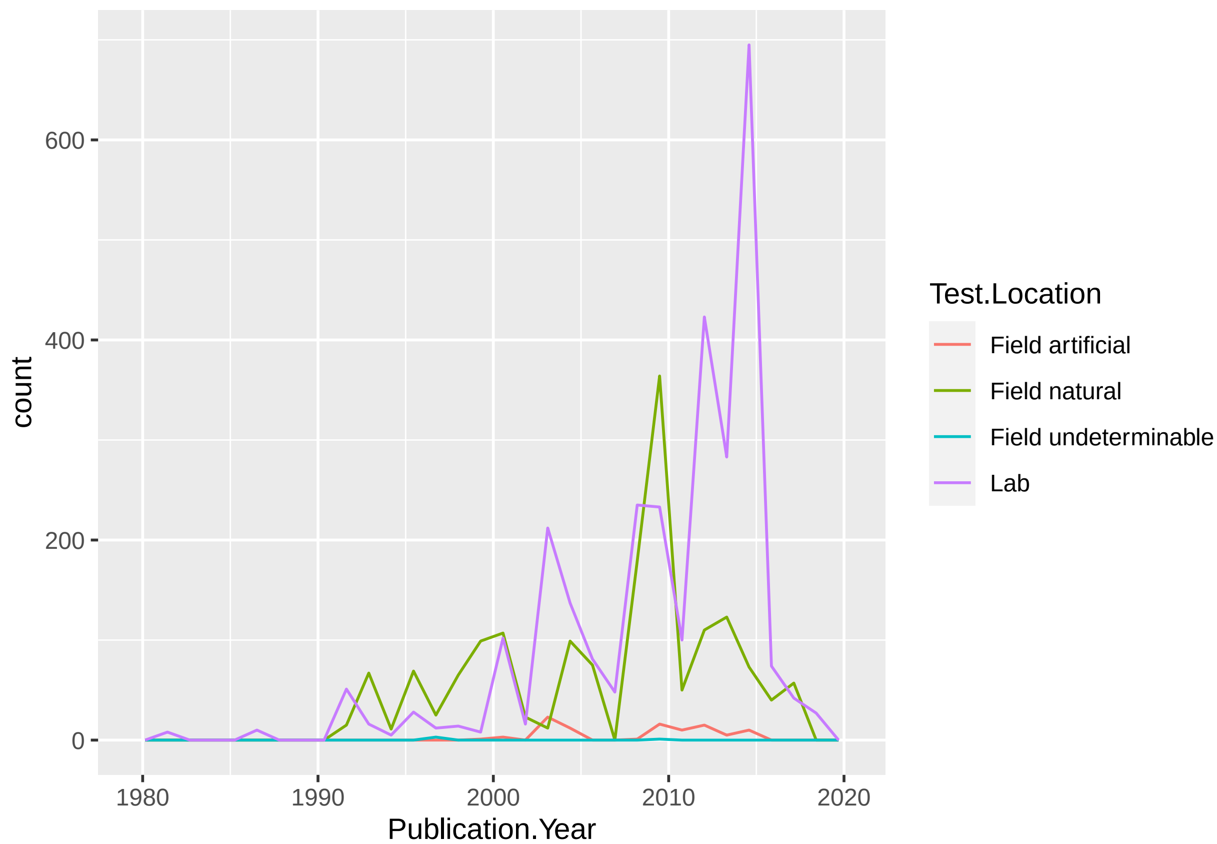


10. Reproduce the same graph but now add a color aesthetic so that different Test.Location are displayed as different colors.

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

```
## Warning in geom_freqpoly(aes(x = Publication.Year, color = Test.Location, :
## Ignoring unknown aesthetics: bins
```

```
## `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?

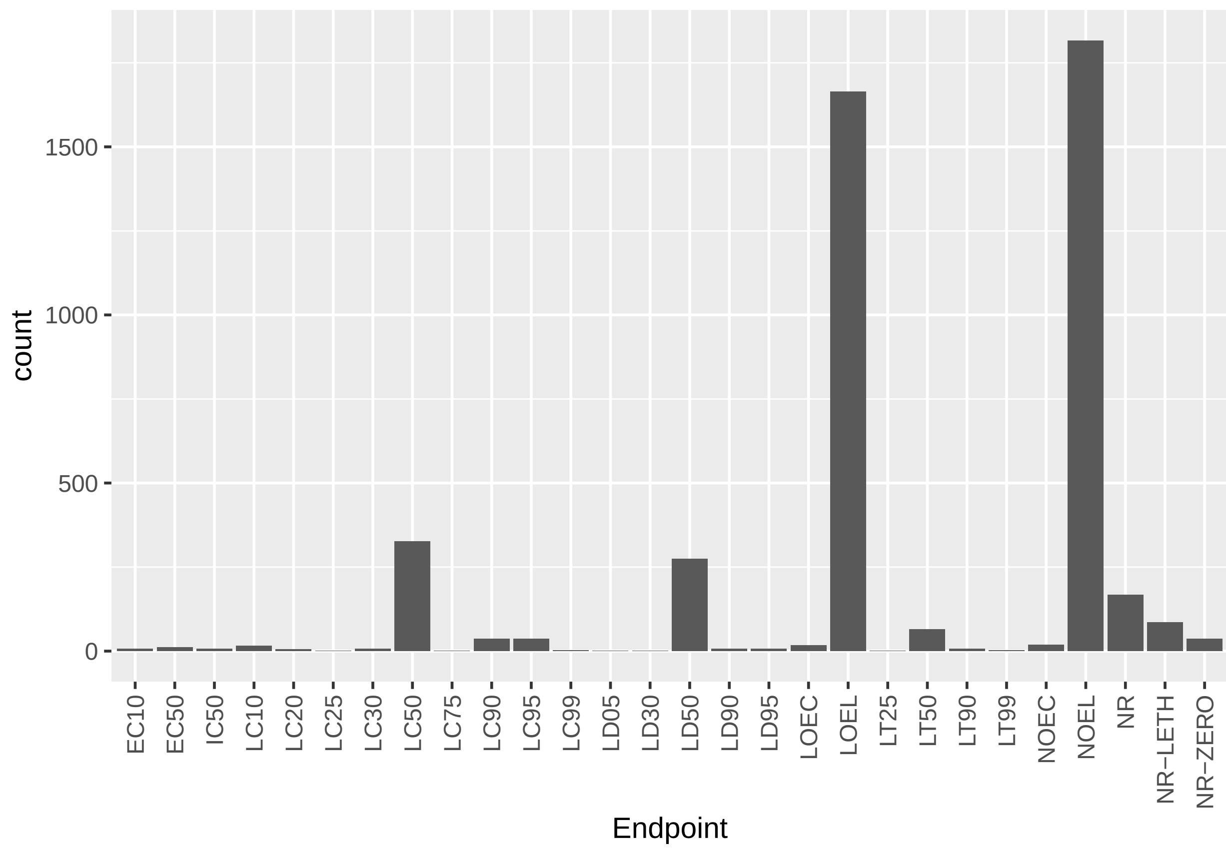
> Answer:

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) + geom_histogram(aes(x=Endpoint),stat="count") + theme(axis.text.x = element_text(angle
```

```
## Warning in geom_histogram(aes(x = Endpoint), stat = "count"): Ignoring unknown
## parameters: `binwidth`, `bins`, and `pad`
```



Answer:

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.

```r
class(Litter$collectDate)
```

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

```r
library(lubridate)
date_new <- ymd(Litter$collectDate)
date_new
```

```
##   [1] "2018-08-02" "2018-08-02" "2018-08-02" "2018-08-02" "2018-08-02"
##   [6] "2018-08-02" "2018-08-02" "2018-08-02" "2018-08-02" "2018-08-02"
##  [11] "2018-08-02" "2018-08-02" "2018-08-02" "2018-08-02" "2018-08-02"
##  [16] "2018-08-02" "2018-08-02" "2018-08-02" "2018-08-02" "2018-08-02"
##  [21] "2018-08-02" "2018-08-02" "2018-08-02" "2018-08-02" "2018-08-02"
##  [26] "2018-08-02" "2018-08-02" "2018-08-02" "2018-08-02" "2018-08-02"
##  [31] "2018-08-02" "2018-08-02" "2018-08-02" "2018-08-02" "2018-08-02"
##  [36] "2018-08-02" "2018-08-02" "2018-08-02" "2018-08-02" "2018-08-02"
```

```
##  [41] "2018-08-02" "2018-08-02" "2018-08-02" "2018-08-02" "2018-08-02"
##  [46] "2018-08-02" "2018-08-02" "2018-08-02" "2018-08-02" "2018-08-02"
##  [51] "2018-08-02" "2018-08-02" "2018-08-02" "2018-08-02" "2018-08-02"
##  [56] "2018-08-02" "2018-08-02" "2018-08-02" "2018-08-02" "2018-08-02"
##  [61] "2018-08-02" "2018-08-02" "2018-08-02" "2018-08-02" "2018-08-02"
##  [66] "2018-08-02" "2018-08-02" "2018-08-02" "2018-08-02" "2018-08-02"
##  [71] "2018-08-02" "2018-08-02" "2018-08-02" "2018-08-02" "2018-08-02"
##  [76] "2018-08-02" "2018-08-02" "2018-08-02" "2018-08-02" "2018-08-02"
##  [81] "2018-08-02" "2018-08-02" "2018-08-02" "2018-08-02" "2018-08-02"
##  [86] "2018-08-02" "2018-08-02" "2018-08-02" "2018-08-02" "2018-08-02"
##  [91] "2018-08-02" "2018-08-30" "2018-08-30" "2018-08-30" "2018-08-30"
##  [96] "2018-08-30" "2018-08-30" "2018-08-30" "2018-08-30" "2018-08-30"
## [101] "2018-08-30" "2018-08-30" "2018-08-30" "2018-08-30" "2018-08-30"
## [106] "2018-08-30" "2018-08-30" "2018-08-30" "2018-08-30" "2018-08-30"
## [111] "2018-08-30" "2018-08-30" "2018-08-30" "2018-08-30" "2018-08-30"
## [116] "2018-08-30" "2018-08-30" "2018-08-30" "2018-08-30" "2018-08-30"
## [121] "2018-08-30" "2018-08-30" "2018-08-30" "2018-08-30" "2018-08-30"
## [126] "2018-08-30" "2018-08-30" "2018-08-30" "2018-08-30" "2018-08-30"
## [131] "2018-08-30" "2018-08-30" "2018-08-30" "2018-08-30" "2018-08-30"
## [136] "2018-08-30" "2018-08-30" "2018-08-30" "2018-08-30" "2018-08-30"
## [141] "2018-08-30" "2018-08-30" "2018-08-30" "2018-08-30" "2018-08-30"
## [146] "2018-08-30" "2018-08-30" "2018-08-30" "2018-08-30" "2018-08-30"
## [151] "2018-08-30" "2018-08-30" "2018-08-30" "2018-08-30" "2018-08-30"
## [156] "2018-08-30" "2018-08-30" "2018-08-30" "2018-08-30" "2018-08-30"
## [161] "2018-08-30" "2018-08-30" "2018-08-30" "2018-08-30" "2018-08-30"
## [166] "2018-08-30" "2018-08-30" "2018-08-30" "2018-08-30" "2018-08-30"
## [171] "2018-08-30" "2018-08-30" "2018-08-30" "2018-08-30" "2018-08-30"
## [176] "2018-08-30" "2018-08-30" "2018-08-30" "2018-08-30" "2018-08-30"
## [181] "2018-08-30" "2018-08-30" "2018-08-30" "2018-08-30" "2018-08-30"
## [186] "2018-08-30" "2018-08-30" "2018-08-30"
```

```r
class(date_new)
```

```
## [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`?

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

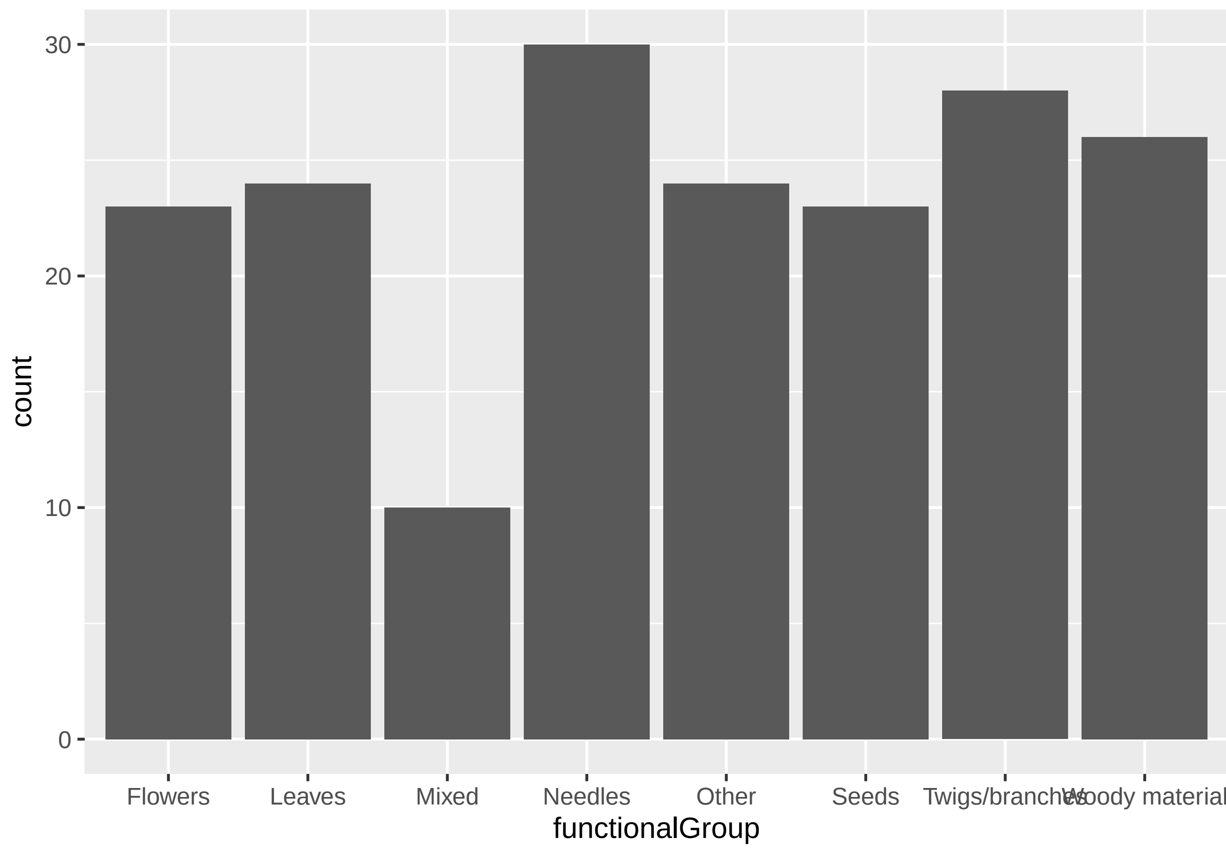
```
##  [1] NIWO_061 NIWO_064 NIWO_067 NIWO_040 NIWO_041 NIWO_063 NIWO_047 NIWO_051
##  [9] NIWO_058 NIWO_046 NIWO_062 NIWO_057
## 12 Levels: NIWO_040 NIWO_041 NIWO_046 NIWO_047 NIWO_051 NIWO_057 ... NIWO_067
```

Answer:

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.

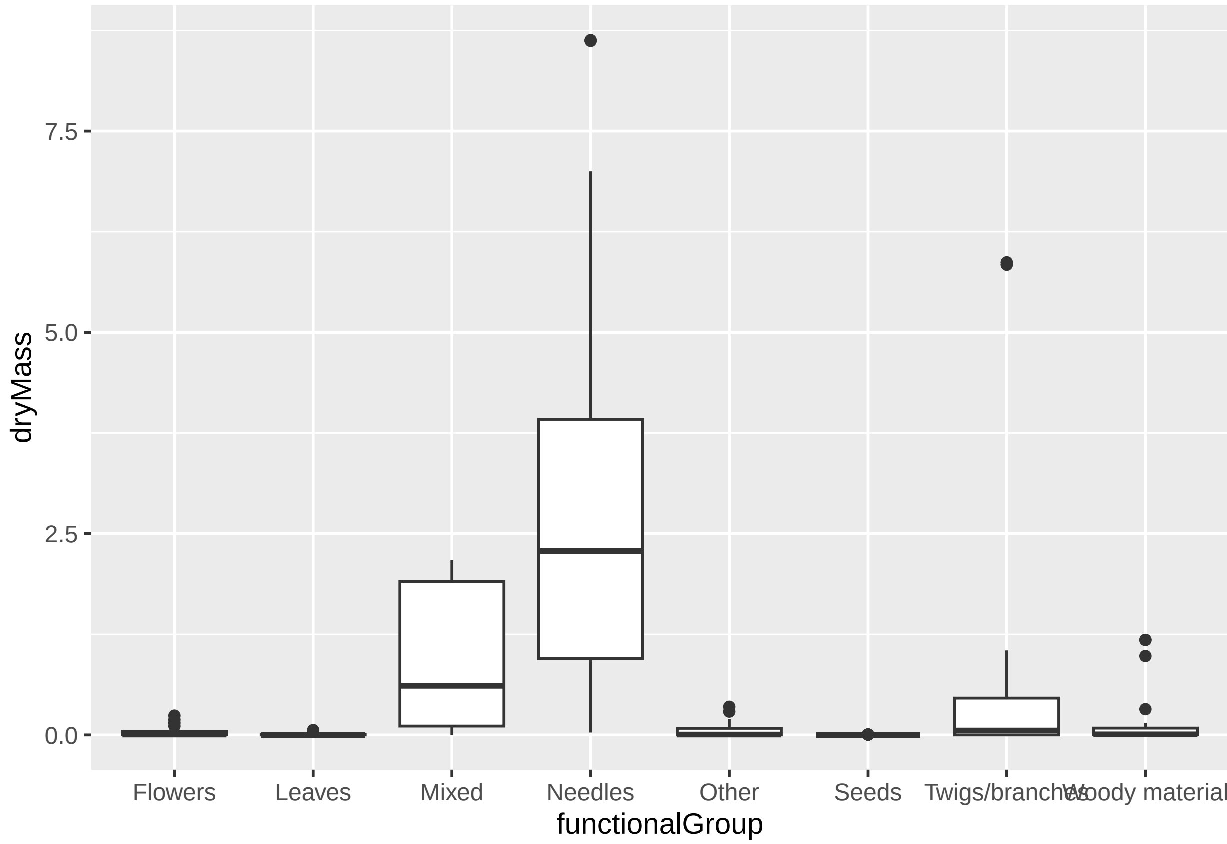
```r
ggplot(Litter) + geom_histogram(aes(x=functionalGroup),stat="count")
```

```
## Warning in geom_histogram(aes(x = functionalGroup), stat = "count"): Ignoring
## unknown parameters: `binwidth`, `bins`, and `pad`
```

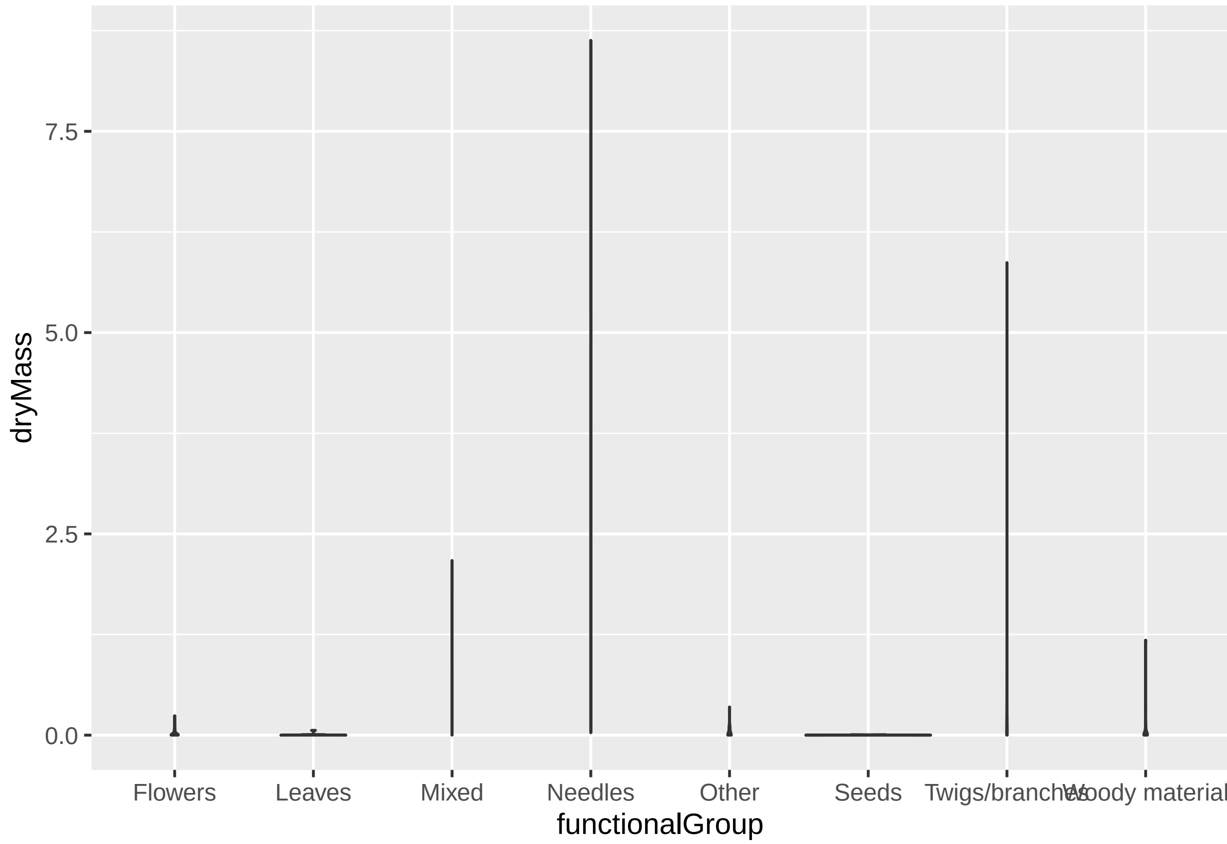


15. Using `geom_boxplot` and `geom_violin`, create a boxplot and a violin plot of dryMass by functionalGroup.

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




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



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: