Assignment 3: Data Exploration

Cristiana Falvo

OVERVIEW

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

Directions

1. Change "Student Name" on line 3 (above) with your name.
2. Work through the steps, **creating code and output** that fulfill each instruction.
3. Be sure to **answer the questions** in this assignment document.
4. When you have completed the assignment, **Knit** the text and code into a single PDF file.
5. After Knitting, submit the completed exercise (PDF file) to the dropbox in Sakai. Add your last name into the file name (e.g., "Salk_A03_DataExploration.Rmd") prior to submission.

The completed exercise is due on Tuesday, January 28 at 1:00 pm.

Set up your R session

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.

```
getwd()
```

```
## [1] "/Users/cristiana/Documents/Duke/DataAnalytics/Environmental_Data_Analytics_2020/Assignments"
```

```
library(tidyverse)

Neonics <- read.csv("~/Documents/Duke/DataAnalytics/Environmental_Data_Analytics_2020/Data/Raw/ECOTOX_N
Litter <- read.csv("../Data/Raw/NEON_NIWO_Litter_massdata_2018-08_raw.csv")
```

Learn about your system

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

   Answer: It's important to understand the effects that insecticides (especially ones that are widely used) have on insects so we can determine the proper levels that should be allowed to be released into the environment.

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: Understanding the composition of the material that falls on the forest floor can help us understand the chemicals that will later be mixed into the earth and cycled through the atmosphere.

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: * ground traps are sampled once per year * deciduous trees are sampled more often than evergreen trees * locations of tower plots are selected randomly

Obtain basic summaries of your data (Neonics)

5. What are the dimensions of the dataset?

```r
dim(Neonics)
```

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

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

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

Answer: Chloro-3-pyridinyl)methyl]-N-nitro-2-imidazolidinimine , perhaps this is the most common chemical

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.

```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)
```

Answer: Honey Bee, Parasitic Wasp, Buff Tailed Bumblebee, Carniolan Honey Bee, Bumble Bee, Italian Honeybee; these are probably important species for agriculture

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

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

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

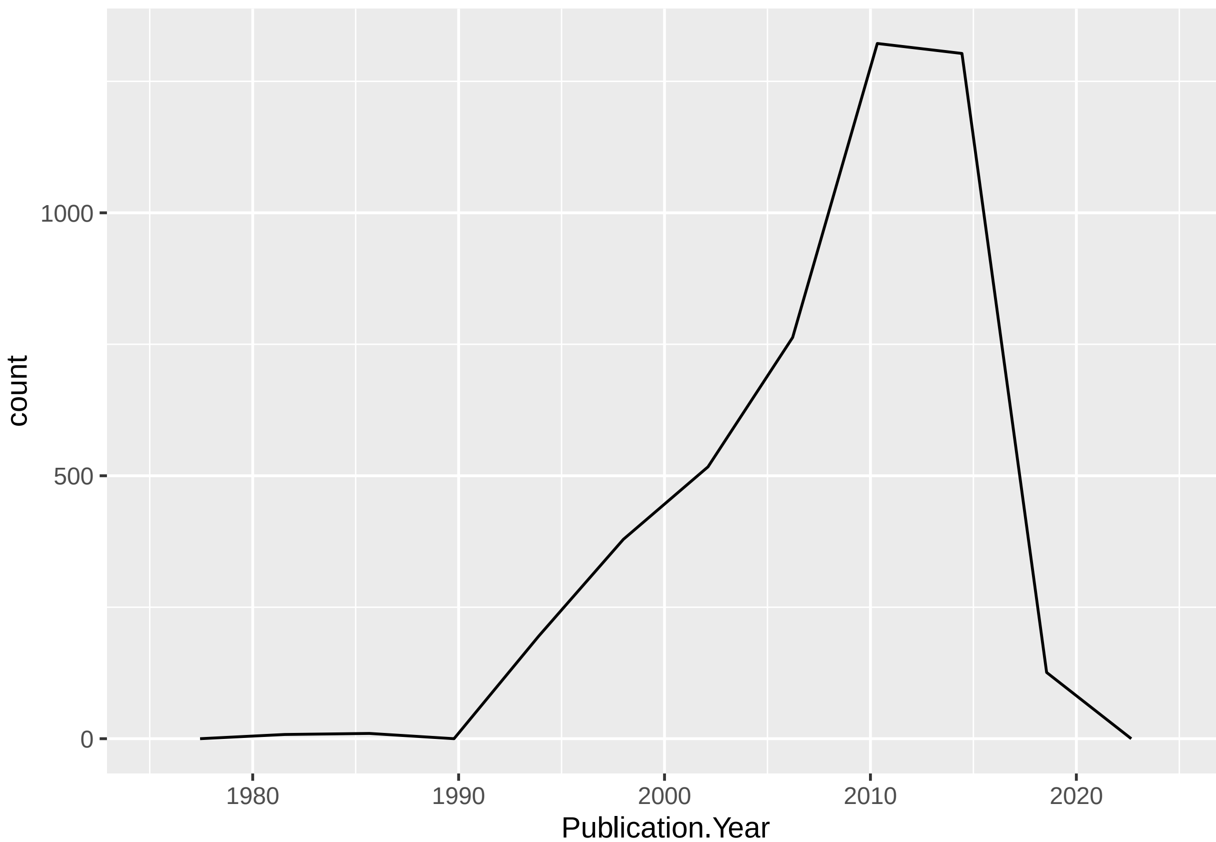
Answer: the class is factor because it has a list of possible values rather than a range of numbers

Explore your data graphically (Neonics)

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

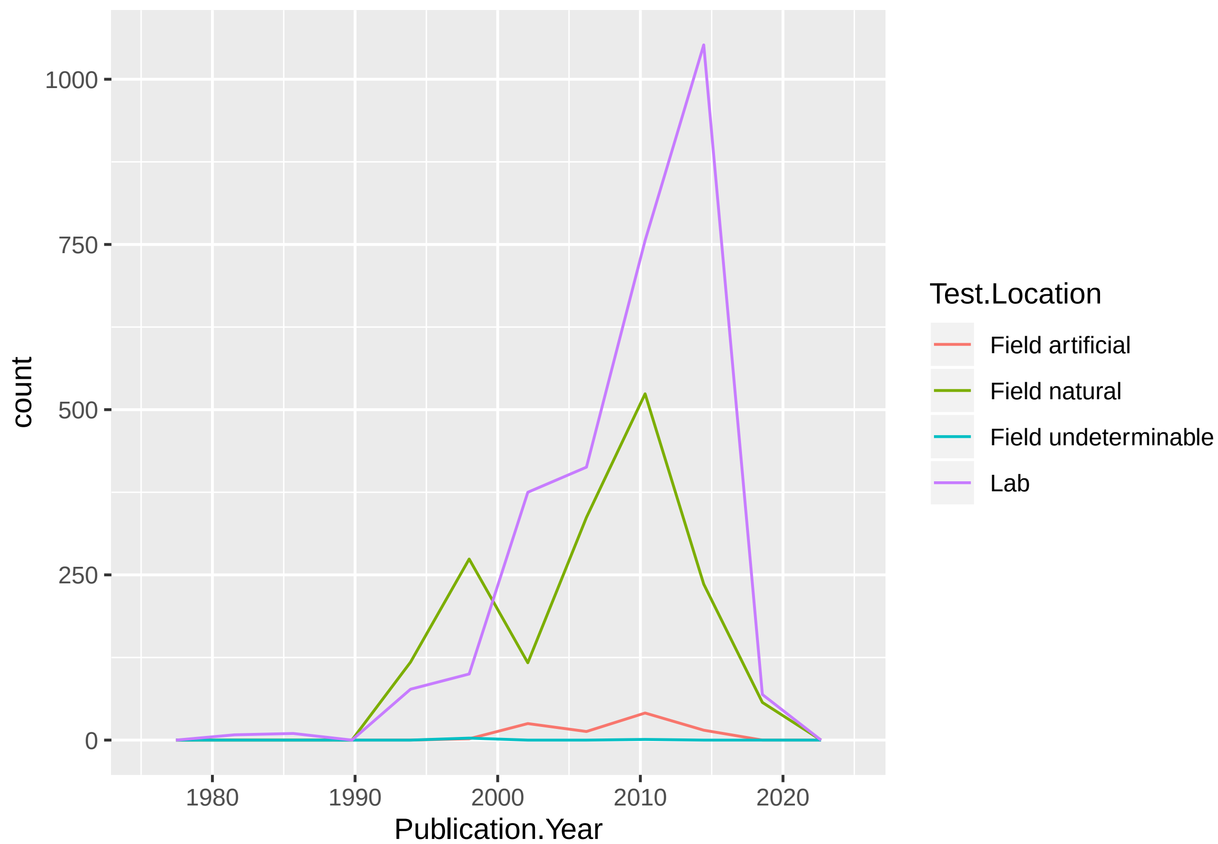
```r
view(Neonics)

ggplot(Neonics) +
  geom_freqpoly(aes(x = Publication.Year), bins = 10)
```



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

```r
ggplot(Neonics) +
  geom_freqpoly(aes(x = Publication.Year, color = Test.Location), bins = 10)
```



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

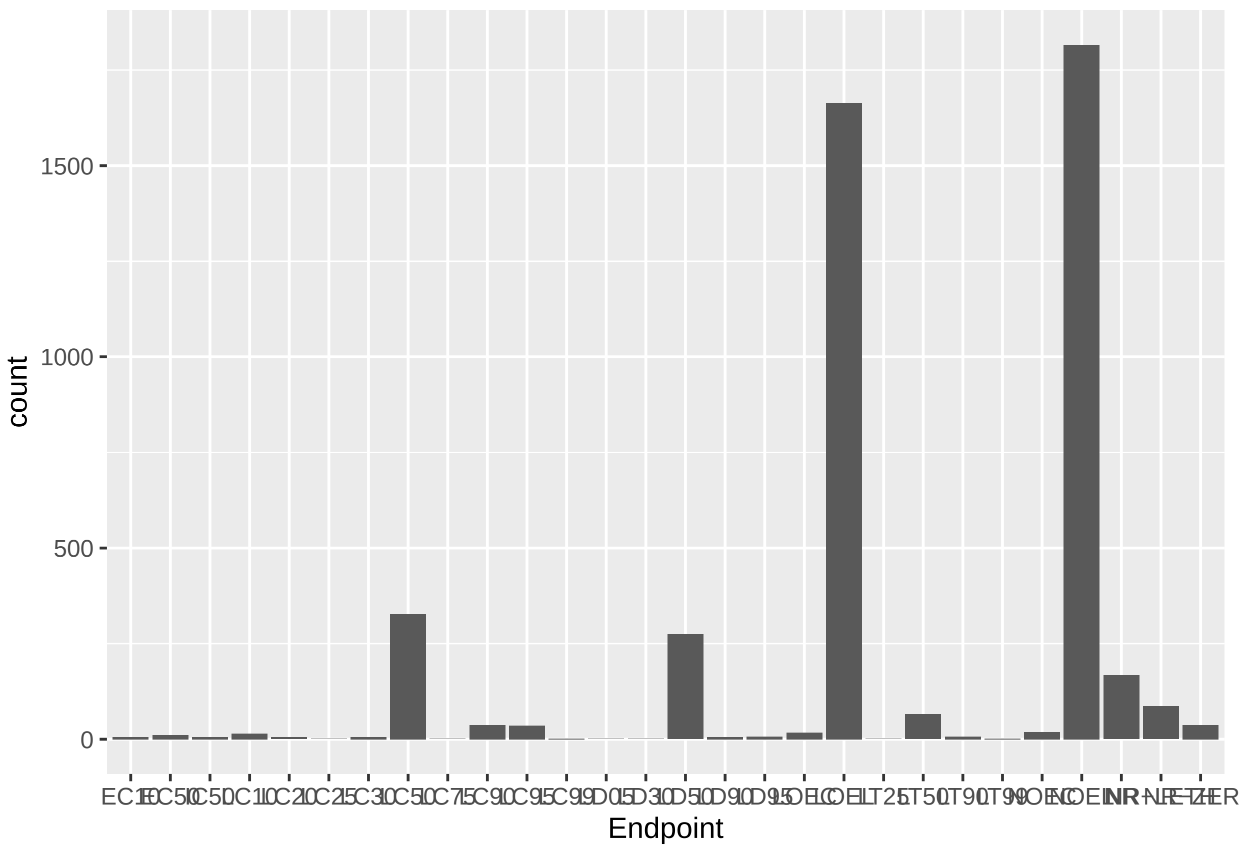
```
summary(Neonics$Test.Location)
```

```
##     Field artificial        Field natural Field undeterminable
##                   96                 1663                    4
##                  Lab
##                 2860
```

Answer: the lab is the most common test location, and this data does change over time. For example between 1990 and 1995 the field was the most common test location.

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

```
ggplot(Neonics, aes(x = Endpoint), bins = 75) +
  geom_bar()
```



Answer: The two most common end points are LOEL (lowest observable effect level, from the terrestrial dataset) and NOEL (no observable effect level, also from the terrestrial dataset)

Explore your data (Litter)

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

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

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

```
head(Litter)
```

```
##                                    uid         namedLocation domainID siteID
## 1 7f065fec-bcb2-4af9-b742-8e520fab7f6e NIWO_061.basePlot.ltr      D13   NIWO
## 2 88df210b-1445-4c3f-b19e-5dabd9305c6e NIWO_061.basePlot.ltr      D13   NIWO
## 3 7f3c549c-1dfa-43bf-a485-c7c2bcb31fd6 NIWO_061.basePlot.ltr      D13   NIWO
## 4 97806ab5-42d2-49c0-8463-db48cd5eab12 NIWO_061.basePlot.ltr      D13   NIWO
## 5 9d7c89f5-85f8-47b6-b415-1ae208580e6f NIWO_061.basePlot.ltr      D13   NIWO
## 6 6ca7a3e8-4d9e-4062-91a0-845f23b5b925 NIWO_061.basePlot.ltr      D13   NIWO
##     plotID       trapID  weighDate    setDate collectDate     ovenStartDate
## 1 NIWO_061 NIWO_061_169 2018-08-06 2018-07-05  2018-08-02 2018-08-02T21:00Z
## 2 NIWO_061 NIWO_061_169 2018-08-06 2018-07-05  2018-08-02 2018-08-02T21:00Z
## 3 NIWO_061 NIWO_061_169 2018-08-06 2018-07-05  2018-08-02 2018-08-02T21:00Z
## 4 NIWO_061 NIWO_061_169 2018-08-06 2018-07-05  2018-08-02 2018-08-02T21:00Z
## 5 NIWO_061 NIWO_061_169 2018-08-06 2018-07-05  2018-08-02 2018-08-02T21:00Z
## 6 NIWO_061 NIWO_061_169 2018-08-06 2018-07-05  2018-08-02 2018-08-02T21:00Z
```

```
##          ovenEndDate                fieldSampleID
## 1 2018-08-06T18:02Z NEON.LTR.NIWO061169.20180802
## 2 2018-08-06T18:02Z NEON.LTR.NIWO061169.20180802
## 3 2018-08-06T18:02Z NEON.LTR.NIWO061169.20180802
## 4 2018-08-06T18:02Z NEON.LTR.NIWO061169.20180802
## 5 2018-08-06T18:02Z NEON.LTR.NIWO061169.20180802
## 6 2018-08-06T18:02Z NEON.LTR.NIWO061169.20180802
##                       massSampleID samplingProtocolVersion functionalGroup
## 1 NEON.LTR.NIWO061169.20180802.TWI       NEON.DOC.001710vE  Twigs/branches
## 2 NEON.LTR.NIWO061169.20180802.SDS       NEON.DOC.001710vE           Seeds
## 3 NEON.LTR.NIWO061169.20180802.WDY       NEON.DOC.001710vE  Woody material
## 4 NEON.LTR.NIWO061169.20180802.FLR       NEON.DOC.001710vE         Flowers
## 5 NEON.LTR.NIWO061169.20180802.WDY       NEON.DOC.001710vE  Woody material
## 6 NEON.LTR.NIWO061169.20180802.NDL       NEON.DOC.001710vE         Needles
##   dryMass qaDryMass remarks                   measuredBy
## 1   0.400         N      NA kstyers@battelleecology.org
## 2   0.005         N      NA kstyers@battelleecology.org
## 3   0.040         Y      NA kstyers@battelleecology.org
## 4   0.005         N      NA kstyers@battelleecology.org
## 5   0.070         N      NA kstyers@battelleecology.org
## 6   1.000         N      NA kstyers@battelleecology.org
```

```r
Litter$collectDate <- as.Date(Litter$collectDate, format = "%y-%m-%d" )
class(Litter$collectDate) # this field is now recognized by R as a date format
```

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

```r
# hm..now all collectDate values are 'NA' ..
  ## maybe I overwrote the values? maybe I needed to dictate a format change..

unique(Litter$setDate)
```

```
## [1] 2018-07-05 2018-08-02
## Levels: 2018-07-05 2018-08-02
```

```r
View(Litter)
```

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$namedLocation)
```

```
##  [1] NIWO_061.basePlot.ltr NIWO_064.basePlot.ltr NIWO_067.basePlot.ltr
##  [4] NIWO_040.basePlot.ltr NIWO_041.basePlot.ltr NIWO_063.basePlot.ltr
##  [7] NIWO_047.basePlot.ltr NIWO_051.basePlot.ltr NIWO_058.basePlot.ltr
## [10] NIWO_046.basePlot.ltr NIWO_062.basePlot.ltr NIWO_057.basePlot.ltr
## 12 Levels: NIWO_040.basePlot.ltr ... NIWO_067.basePlot.ltr
```

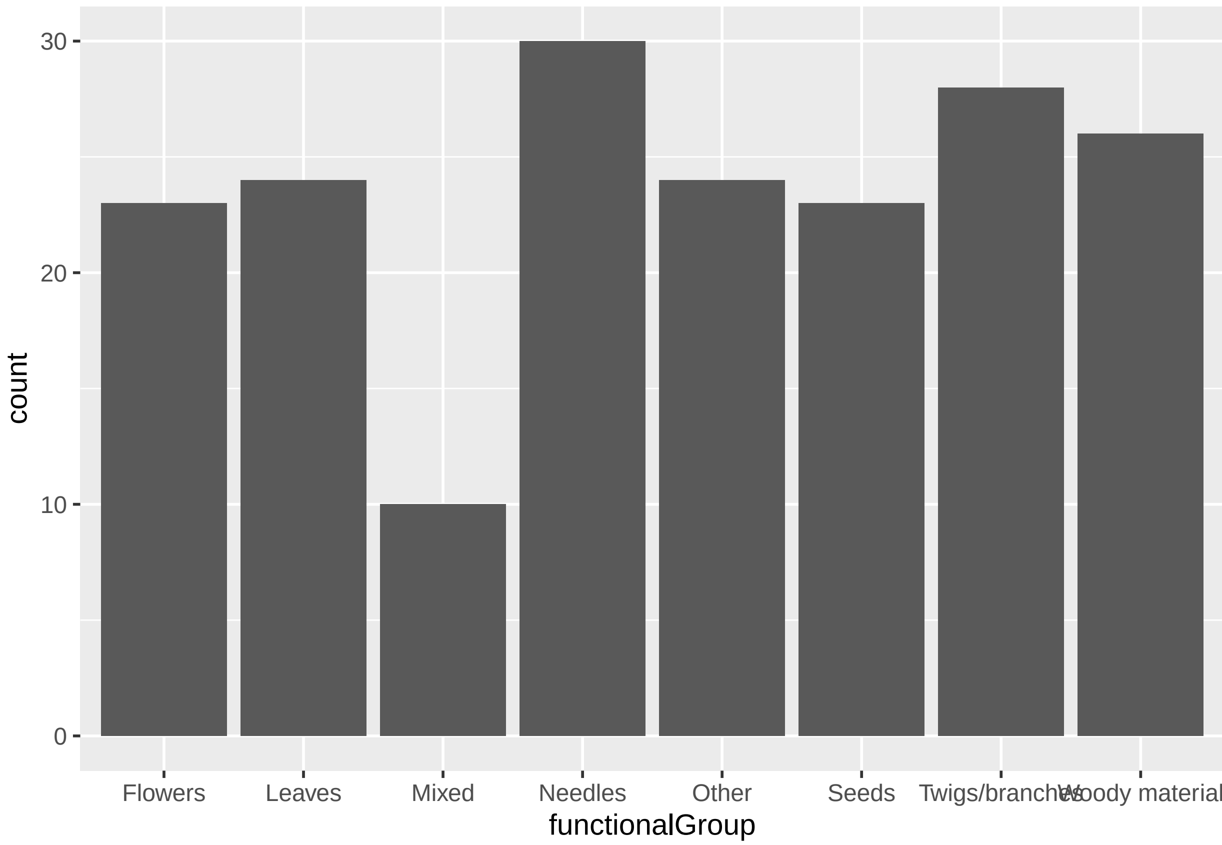
```r
unique(Litter$siteID)
```

```
## [1] NIWO
## Levels: NIWO
```

> Answer: By running unique setDate I was able to see that August 2nd was the only day sampled in August 2018. Niwot was the only site surveyed. Unique gives unique values while summary gives all values.

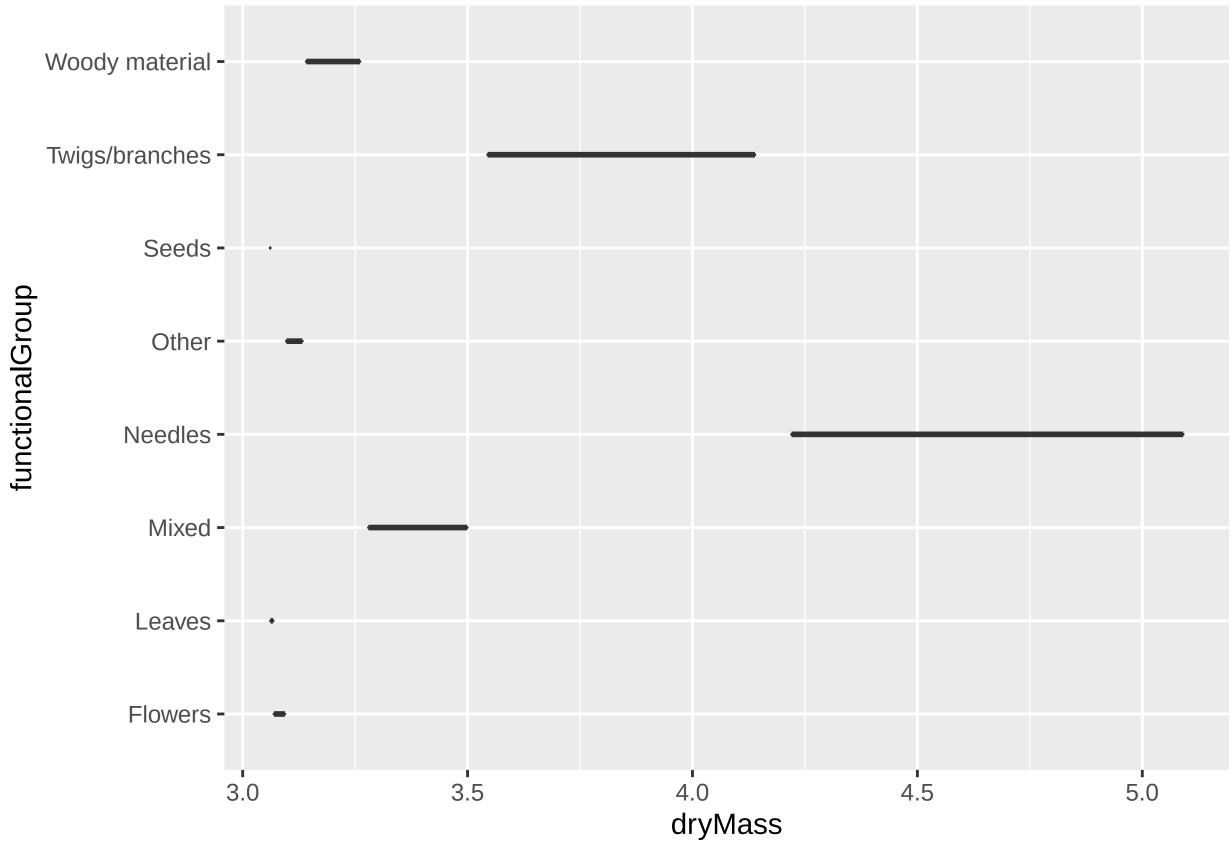
14. Create a bar graph of functionalGroup counts. This shows you what type of litter is collected at the Niwot Ridge sites. Notice that litter types are fairly equally distributed across the Niwot Ridge sites.

```r
ggplot(Litter, aes(x = functionalGroup), bins = 75) +
  geom_bar()
```



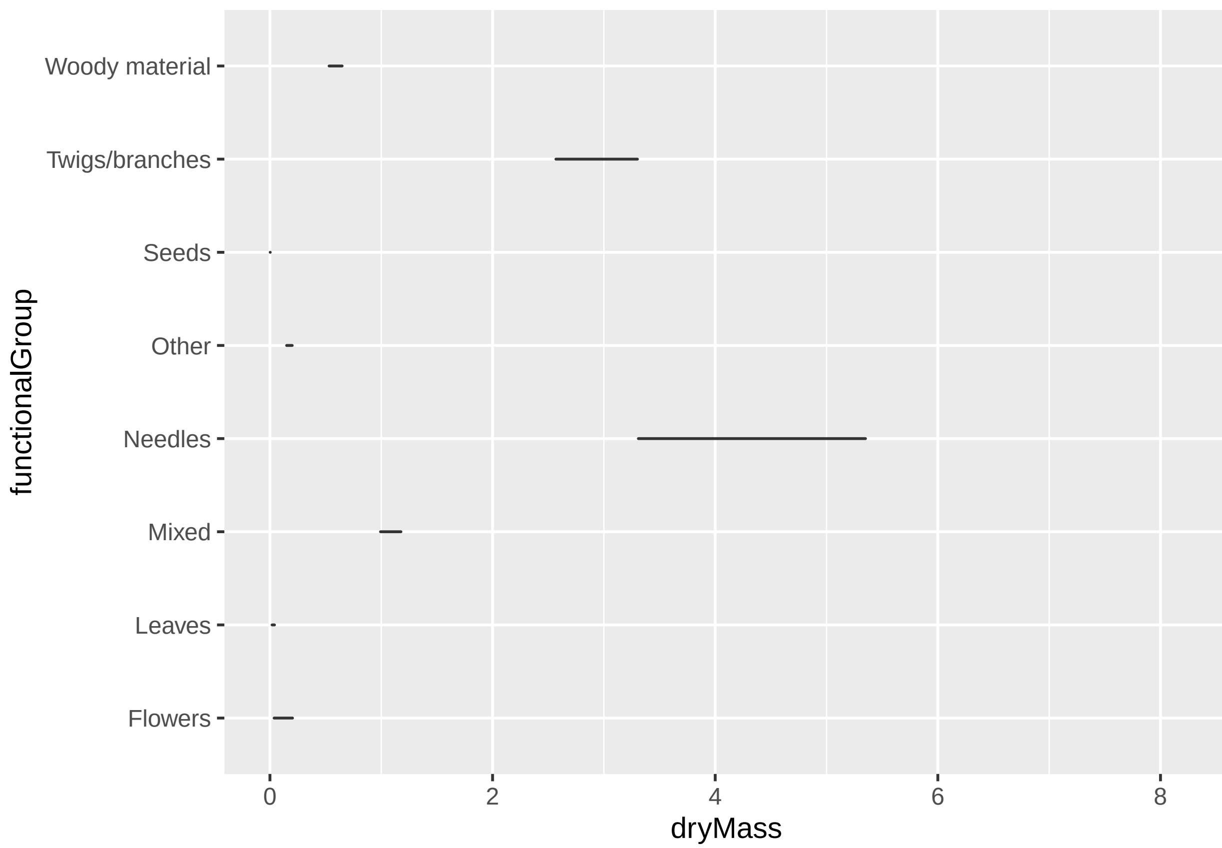
15. Using `geom_boxplot` and `geom_violin`, create a boxplot and a violin plot of dryMass by functionalGroup.

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




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

```
## Warning: position_dodge requires non-overlapping x intervals
```



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

> Answer: Box plot was more effective, violin doesn't seem to work here.

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

> Answer: Needles tend to have the highest biomass.