Assignment 3: Data Exploration

Garrett Pearce, Section #4

OVERVIEW

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

Directions

1. Change "Student Name, Section #" on line 3 (above) with your name and section number.
2. Work through the steps, **creating code and output** that fulfill each instruction.
3. Be sure to **answer the questions** in this assignment document.
4. When you have completed the assignment, **Knit** the text and code into a single PDF file.
5. After Knitting, submit the completed exercise (PDF file) to the dropbox in Sakai. Add your last name into the file name (e.g., "FirstLast_A03_DataExploration.Rmd") prior to submission.

The completed exercise is due on <>.

Set up your R session

1. Check your working directory, load necessary packages (tidyverse), and upload two datasets: the ECOTOX neonicotinoid dataset (ECOTOX_Neonicotinoids_Insects_raw.csv) and the Niwot Ridge NEON dataset for litter and woody debris (NEON_NIWO_Litter_massdata_2018-08_raw.csv). Name these datasets "Neonics" and "Litter", respectively. **Be sure to add the `stringsAsFactors = TRUE` parameter to the function when reading in the CSV files.**

```
getwd()
```

```
## [1] "C:/Users/dgp20/Documents/ENV 872/Environmental_Data_Analytics_2022/Assignments"
```

```
library(tidyverse)

Neonics <- read.csv('../Data/Raw/ECOTOX_Neonicotinoids_Insects_raw.csv', stringsAsFactors = TRUE)
Litter <- read.csv('../Data/Raw/NEON_NIWO_Litter_massdata_2018-08_raw.csv', stringsAsFactors = TRUE)
```

Learn about your system

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

   Answer: We may be interested in how neonicotoids affect pest species of insects versus how they affect insect species that may be less harmful, or even beneficial, to agriculture. Ecotoxicology reports could also be utilized to improve harmful insect reduction.

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: Leaf litter and woody debris can be indicative of soil quality and therefore forest health. Studying these aspects can paint a picture of forest health over a long period of time.

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: *Sampling takes place at random NEON sites that have woody vegetation >2m tall.* "Trap pairs", made up of one ground-level trap and one elevated trap, are used in every 400 m^2 plot area. *Ground traps are sampled once a year, while elevated trap sampling varies based on vegetation levels, either 1x per week or 1x per 1-2 months.

Obtain basic summaries of your data (Neonics)

5. What are the dimensions of the dataset?

```
dim(Neonics)
```

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

Answer: The dataset is 4,623 rows by 30 columns, or 4,623 observations of 30 data points.

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

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

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

Answer: The three most common effects studied are Reproduction, Mortality, and Population. This makes sense, as population size is a good indicator of species health, and mortality and birth rates are both key determinants of population size and health. These are all good effects to study when determining how neonics affect insect populations.

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

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

```
##                     Honey Bee                Parasitic Wasp
##                           667                           285
##          Buff Tailed Bumblebee           Carniolan Honey Bee
##                           183                           152
##                     Bumble Bee              Italian Honeybee
##                           140                           113
##                Japanese Beetle             Asian Lady Beetle
##                            94                            76
```

```
##                 Euonymus Scale                        Wireworm
##                             75                              69
##              European Dark Bee               Minute Pirate Bug
##                             66                              62
##            Asian Citrus Psyllid                   Parastic Wasp
##                             60                              58
##         Colorado Potato Beetle                 Parasitoid Wasp
##                             57                              51
##            Erythrina Gall Wasp                    Beetle Order
##                             49                              47
##   Snout Beetle Family, Weevil        Sevenspotted Lady Beetle
##                             47                              46
##                 True Bug Order          Buff-tailed Bumblebee
##                             45                              39
##                   Aphid Family                  Cabbage Looper
##                             38                              38
##            Sweetpotato Whitefly                   Braconid Wasp
##                             37                              33
##                   Cotton Aphid                  Predatory Mite
##                             33                              33
##         Ladybird Beetle Family                      Parasitoid
##                             30                              30
##                  Scarab Beetle                   Spring Tiphia
##                             29                              29
##                    Thrip Order            Ground Beetle Family
##                             29                              27
##             Rove Beetle Family                   Tobacco Aphid
##                             27                              27
##                   Chalcid Wasp         Convergent Lady Beetle
##                             25                              25
##                  Stingless Bee               Spider/Mite Class
##                             25                              24
##            Tobacco Flea Beetle                Citrus Leafminer
##                             24                              23
##                Ladybird Beetle                       Mason Bee
##                             23                              22
##                       Mosquito                   Argentine Ant
##                             22                              21
##                         Beetle    Flatheaded Appletree Borer
##                             21                              20
##            Horned Oak Gall Wasp              Leaf Beetle Family
##                             20                              20
##              Potato Leafhopper    Tooth-necked Fungus Beetle
##                             20                              20
##                   Codling Moth       Black-spotted Lady Beetle
##                             19                              18
##                   Calico Scale             Fairyfly Parasitoid
##                             18                              18
##                    Lady Beetle          Minute Parasitic Wasps
##                             18                              18
##                      Mirid Bug                Mulberry Pyralid
##                             18                              18
##                       Silkworm                  Vedalia Beetle
##                             18                              18
```

```
##             Araneoid Spider Order                          Bee Order
##                                17                                 17
##                   Egg Parasitoid                       Insect Class
##                                17                                 17
##          Moth And Butterfly Order       Oystershell Scale Parasitoid
##                                17                                 17
## Hemlock Woolly Adelgid Lady Beetle             Hemlock Wooly Adelgid
##                                16                                 16
##                              Mite                        Onion Thrip
##                                16                                 16
##             Western Flower Thrips                        Corn Earworm
##                                15                                 14
##                 Green Peach Aphid                          House Fly
##                                14                                 14
##                         Ox Beetle                 Red Scale Parasite
##                                14                                 14
##                Spined Soldier Bug              Armoured Scale Family
##                                14                                 13
##                  Diamondback Moth                       Eulophid Wasp
##                                13                                 13
##                 Monarch Butterfly                       Predatory Bug
##                                13                                 13
##             Yellow Fever Mosquito                 Braconid Parasitoid
##                                13                                 12
##                      Common Thrip        Eastern Subterranean Termite
##                                12                                 12
##                            Jassid                          Mite Order
##                                12                                 12
##                         Pea Aphid                    Pond Wolf Spider
##                                12                                 12
##          Spotless Ladybird Beetle             Glasshouse Potato Wasp
##                                11                                 10
##                          Lacewing             Southern House Mosquito
##                                10                                 10
##            Two Spotted Lady Beetle                         Ant Family
##                                10                                  9
##                      Apple Maggot                            (Other)
##                                 9                                670
```

Answer: Not counting the "Other: category, the six most common species in order are Honey Bee, Parasitic Wasp, Buff Tailed Bumblebee, Carniolan Honeybee, Bumblebee, and the Italian Honeybee.The multiple bee species all make sense here- these are important pollinators in their ecosystems, and therefore are important to study- researchers may want to ensure that neonics use does not correlate with population decline in these bee species. the one outlier here is the parasitic wasp. This may be studied so heavily because, like the bee species, these wasps have a massive effect on their ecosystems, since they kill other insects to lay their eggs. Since these wasps may have a large adverse effect on ecosystems and agriculture, researchers may instead want to ensure that neonics are sufficiently decreasing parasitic wasp populations.

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

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

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

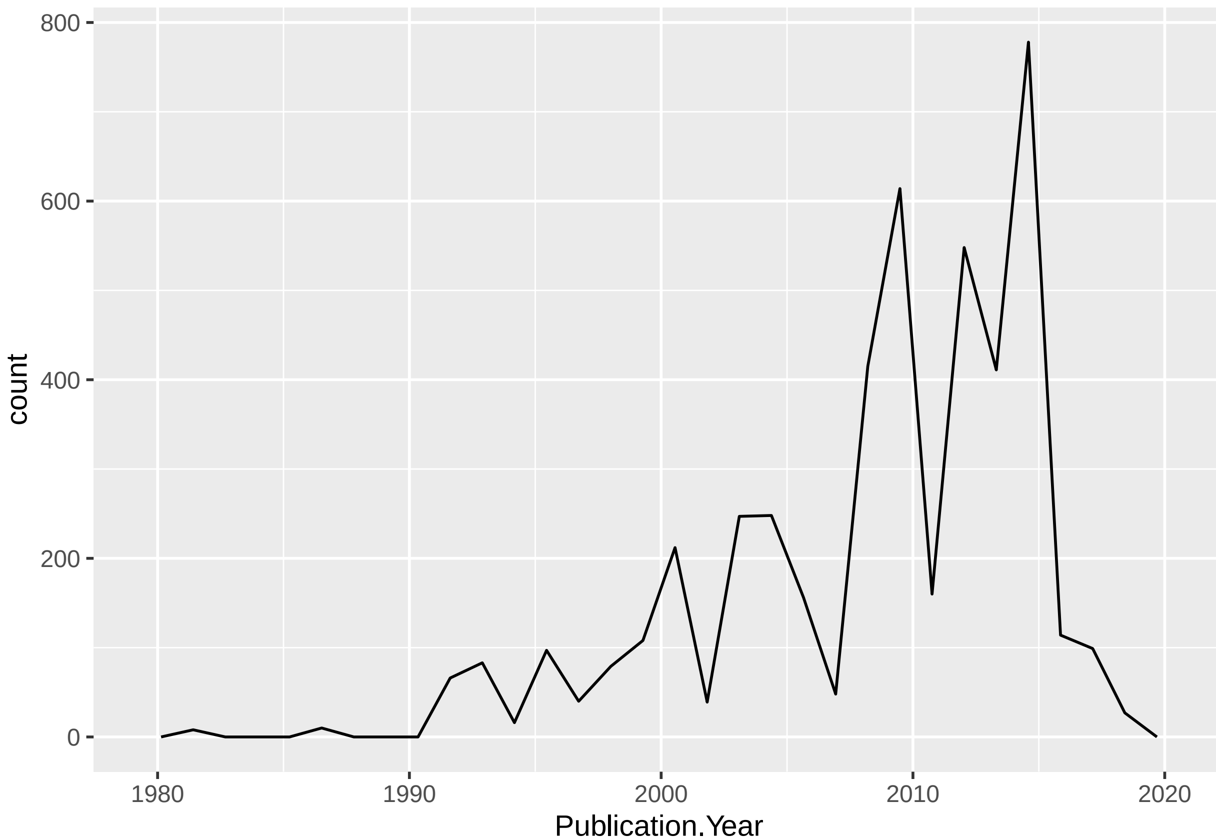
Answer: Conc.1..Author. is labelled as class "factor" here.This may be because the contents of the column in question are all categorical and non-numerical.

Explore your data graphically (Neonics)

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

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

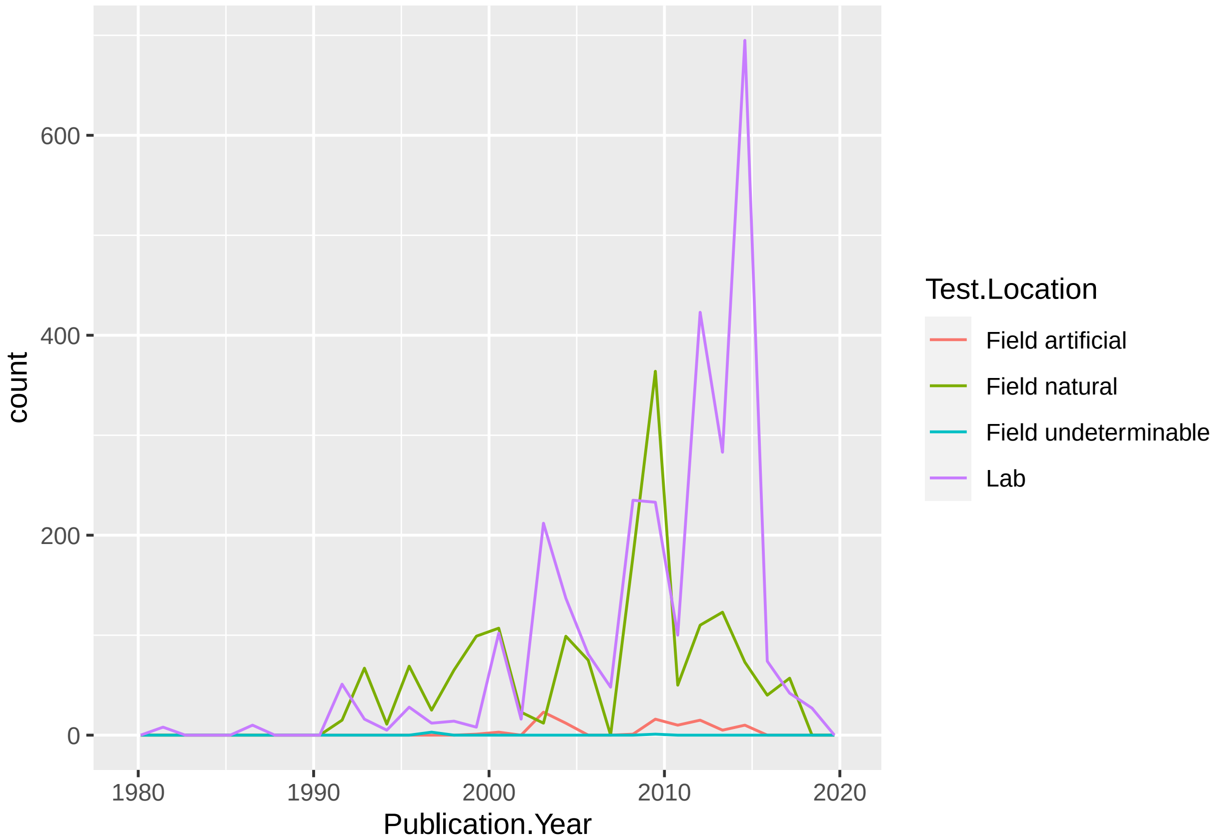
```
## `stat_bin()` using `bins = 30`. Pick better value with `binwidth`.
```



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()
```

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

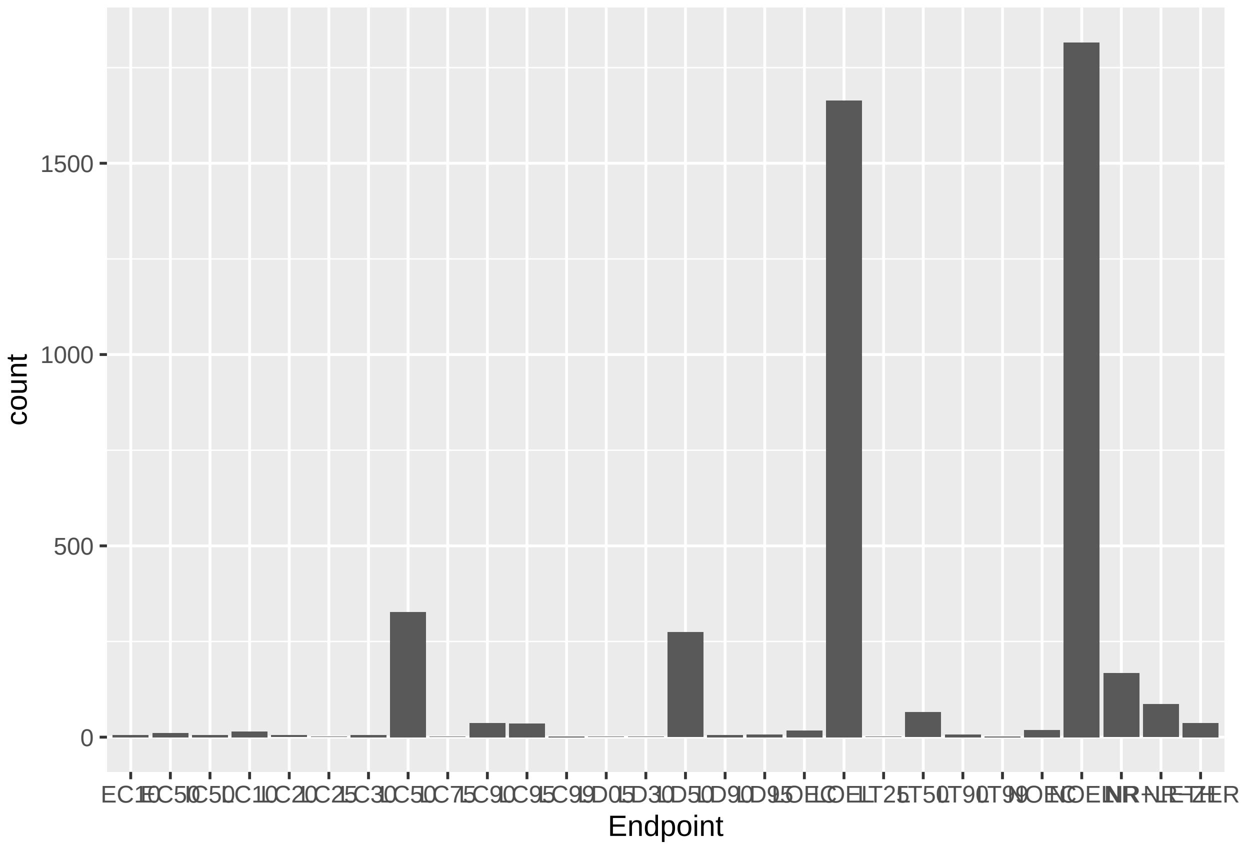


Interpret this graph. What are the most common test locations, and do they differ over time?

> Answer: Lab tests seem to be the most common generally. This seems to vary a bit over time- for example, in the mid-to-late 1990s and the late 2000s, natural field tests were used more often.Largely, though, lab locations dominate.

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), binwidth = 50, cex = 0.5) +
  geom_bar()
```



Answer: The most common endpoints are NOEL(no observable effect level) and LOEL (lowest observable effect level), suggesting that neonics had, on average, little to no effects on observed populations.

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"
```

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

class(realdate)
```

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

```
unique(realdate)
```

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

Litter was collected on 2018-08-02 and 2018-08-30

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

```
unique(Litter$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
```

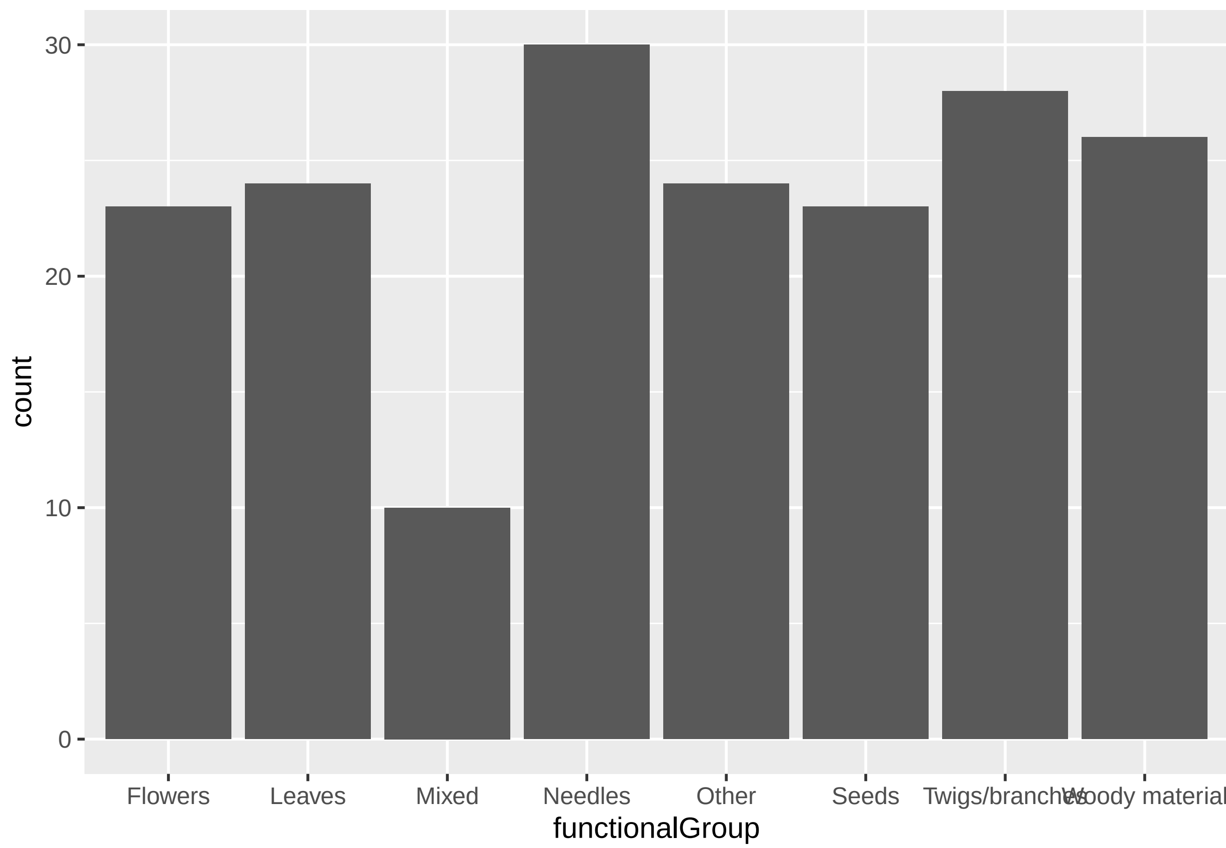
```
summary(Litter$namedLocation)
```

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

Answer: It appears that 12 plots were sampled at Niwot Ridge. This lines up with the output from the summary function.

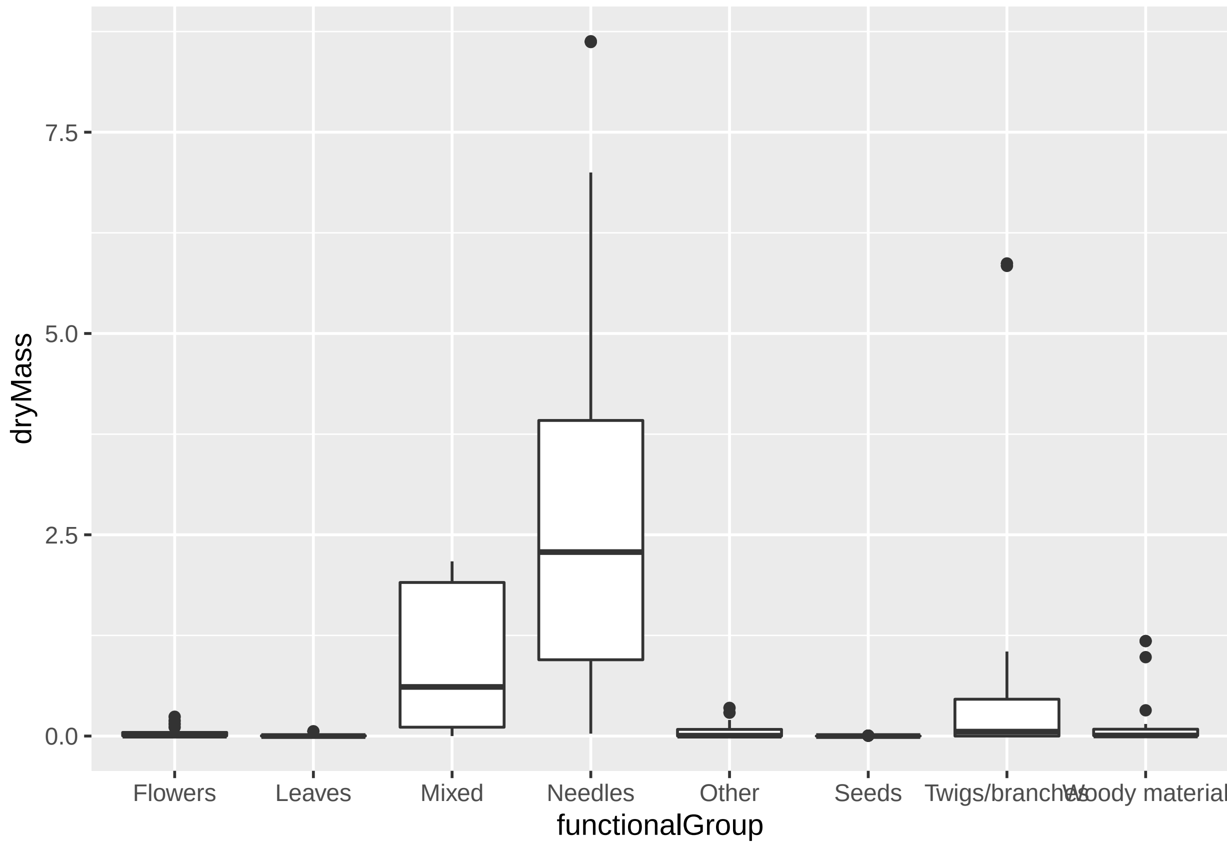
14. Create a bar graph of functionalGroup counts. This shows you what type of litter is collected at the Niwot Ridge sites. Notice that litter types are fairly equally distributed across the Niwot Ridge sites.

```
ggplot(Litter, aes(x = functionalGroup), binwidth = 50) +
  geom_bar()
```

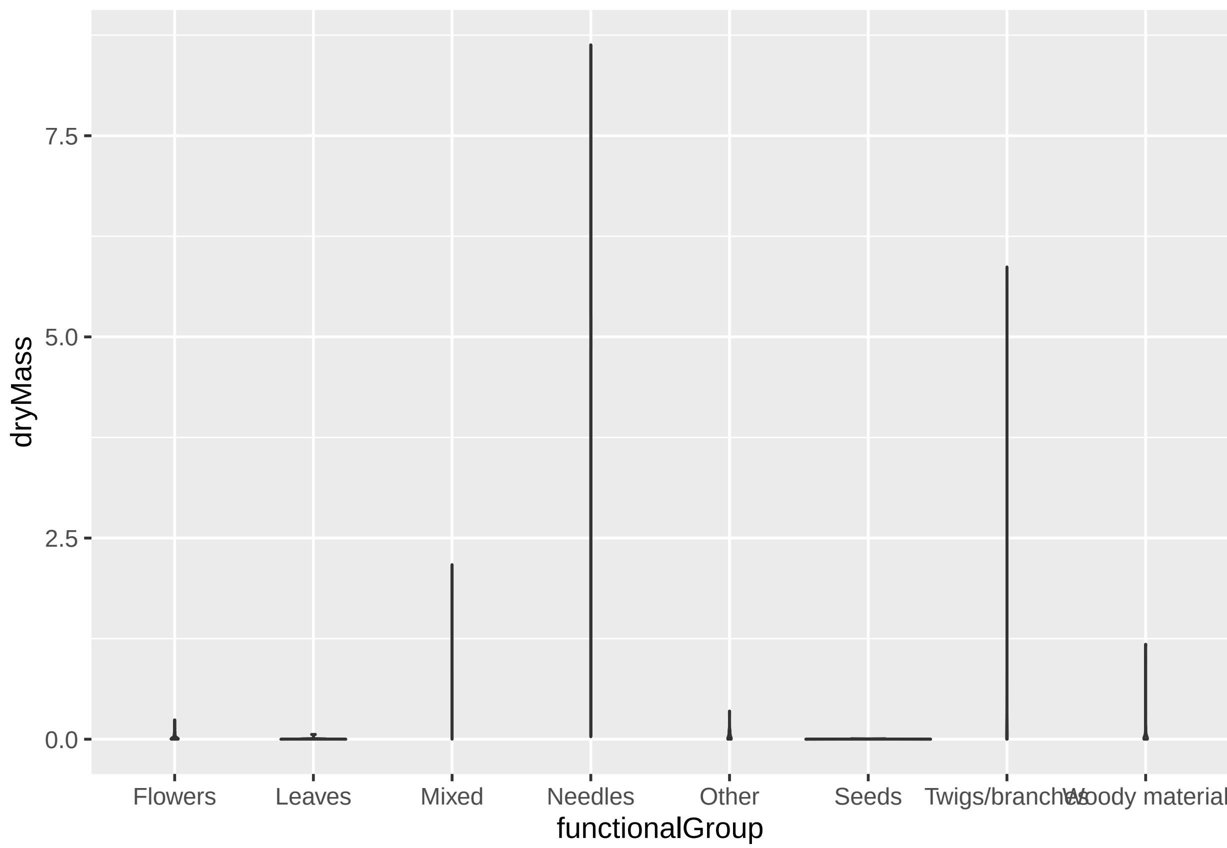


15. Using `geom_boxplot` and `geom_violin`, create a boxplot and a violin plot of dryMass by functionalGroup.

```
ggplot(Litter, aes(x = functionalGroup, y=dryMass), binwidth = 50) +
  geom_boxplot()
```




```
ggplot(Litter, aes(x = functionalGroup, y=dryMass), binwidth = 50) +
  geom_violin()
```



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

> Answer: Having the mean/mode and interquartile ranges is helpful for this data visualization, while the violin plot does not contain this more in-depth information.The violin plot's probability density visualization is not as helpful here.

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

> Answer: Needles (and "Mixed").