Assignment 3: Data Exploration

Eva May

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 <>.

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/evamay/Desktop/ENV872/Environmental.Data.Analytics.2021/Assignments"
```

```
setwd("~/Desktop/ENV872/Environmental.Data.Analytics.2021")
library(tidyverse)
Neonics <- read.csv("./Data/Raw/ECOTOX_Neonicotinoids_Insects_raw.csv")
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: Insects - be they pests or pollinators - interact with, consume, and land upon various plant species. Therefore, any pesticides and insecticides used on crops are bound to have an impact on the insects interacting with those crops. While this is the point of using insecticides - either killing or fending off harmful pests - there can sometimes be unintended consequences when these insecticides impact other, non-target insects. Understanding their ecotoxicological impacts will allow researchers to better understand the harm that neonicotinoid use could bring upon beneficial insects like butterflies, bees, and insects that consume typical agricultural pests. Ecotoxicology research could, for example, point to acceptable levels at which neonicotinoids can be used for their desired effect without harming other insect populations or the broader ecosystems that may rely on some of these insects. There is actually a current push from some NGOs to stop the use of neonicotinoids specifically because of their unintended negative impacts on broader ecosystems and important insect species.

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: Woody debris and litter that fall to the forest floor can be an important component of the overall forest ecosystem, and the structure of this debris can help structure the forest floor environment. Coarser debris that does not decompose as quickly can provide a habitat for forest floor organisms. More importantly, though, is the decomposition process of what lies on the forest floor. As debris and litter decompose, the nutrients they hold - which likely vary depending on the specific type of debris - will be upcycled back into the environment via pathways such as through tree roots in the soil. Additionally, the quantity of woody debris and litter can help shape which decomposers and detritivores are able to live in the forest, thus having biodiversity implications.

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: Researchers set up elevated and on-the-ground traps to catch falling and fallen debris and litter, then measured the mass of each collection trap. They used a random sampling design for tower plots where litter and debris were collected. Traps were set up randomly within plots that had >50% vegetation cover, whereas plots with patchy vegetation had traps that were targeted toward areas with heavier vegetation densities. Traps were sampled annually on the ground but were sampled more frequently if they were elevated. The frequency of elevated trap sampling depended on the kind of forest each trap was in. It is clear that protocols for traps were put in place to ensure that each trap was placed in a qualified area of each plot. *

Obtain basic summaries of your data (Neonics)

5. What are the dimensions of the dataset?

```
dim(Neonics)
```

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

```
#4623 rows, 30 columns
```

   Answer: Neonics has 4623 rows and 30 columns.

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

```
Neonics$Effect <- as.factor(Neonics$Effect)
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: Population and mortality are the most commonly studied effects on insects. If we consider the implications of negative impacts on non-target insect species, population-wide effects would be more devastating than effects on individuals alone. Mortality is an important effect because it is - on the scale of individuals - the most devastating negative effect. If an entire local bee population is impacted by an insecticide, or if the insecticide causes death in bees, this would have grim results for the success of crops that rely on pollination. Conversely, if an insecticide has ecotoxicological effects on an entire population of boll weevils, or if it causes high mortality rates in boll weevils, this would suggest that it is beneficial to use on potato crops, for example.

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.

```
Neonics$Species.Common.Name <- as.factor(Neonics$Species.Common.Name)
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: The 6 most commonly studied insects - in descending order - are honeybee, parasitc wasp, buff tailed bumblebee, Carniolan honey bee, bumble bee, and Italian honeybee. These are all pollinators. As alluded to earlier, this makes sense, as environmentally sound insecticides should do as little damage to important pollinators as possible. An insecticide's purpose is to protect agricultural crops, but these crops often cannot grow without pollinators, so ensuring the health of pollinators is part of ensuring agricultural success. Other insects on this list are mostly pests. While it is important that insecticides prevent these pests from ruining crops, most ecotoxicology studies done would focus on non-target effects of the insecticides, as it is expected that the ecotoxicological effects on pests like beetles and worms are already vetted due to the nature of the product.

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] "character"
```

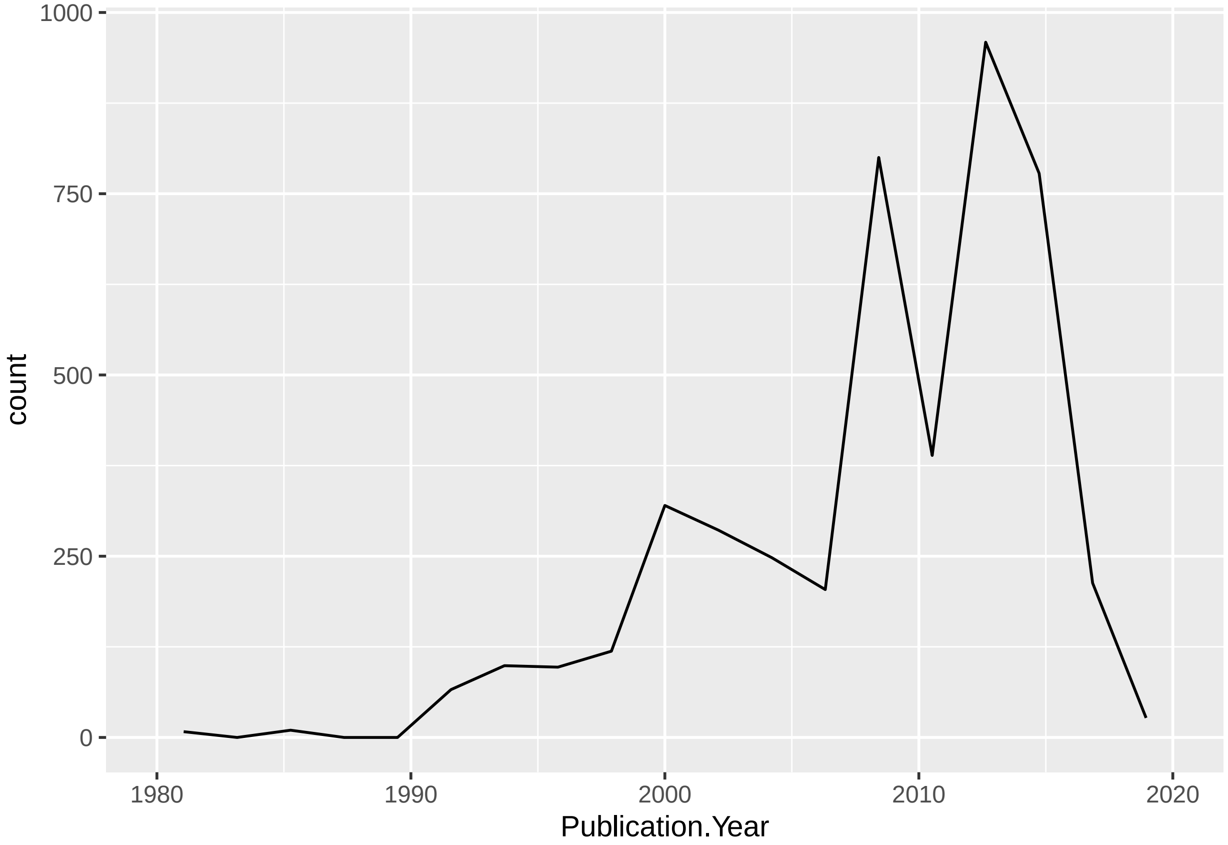
Answer: This column is composed of character values. This is because not all values in this column are simple numerical ones. Some include symbols such as ~ and /, while others are NR (not reported aka no data). Because this column has to contain only 1 kind of data (unlike a list, which can contain multiple data types), all of the values are read as character strings.

Explore your data graphically (Neonics)

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

```r
ggplot(Neonics) +
  geom_freqpoly(aes(x=Publication.Year), bins = 20) +
  scale_x_continuous(limits = c(1980, 2020))
```

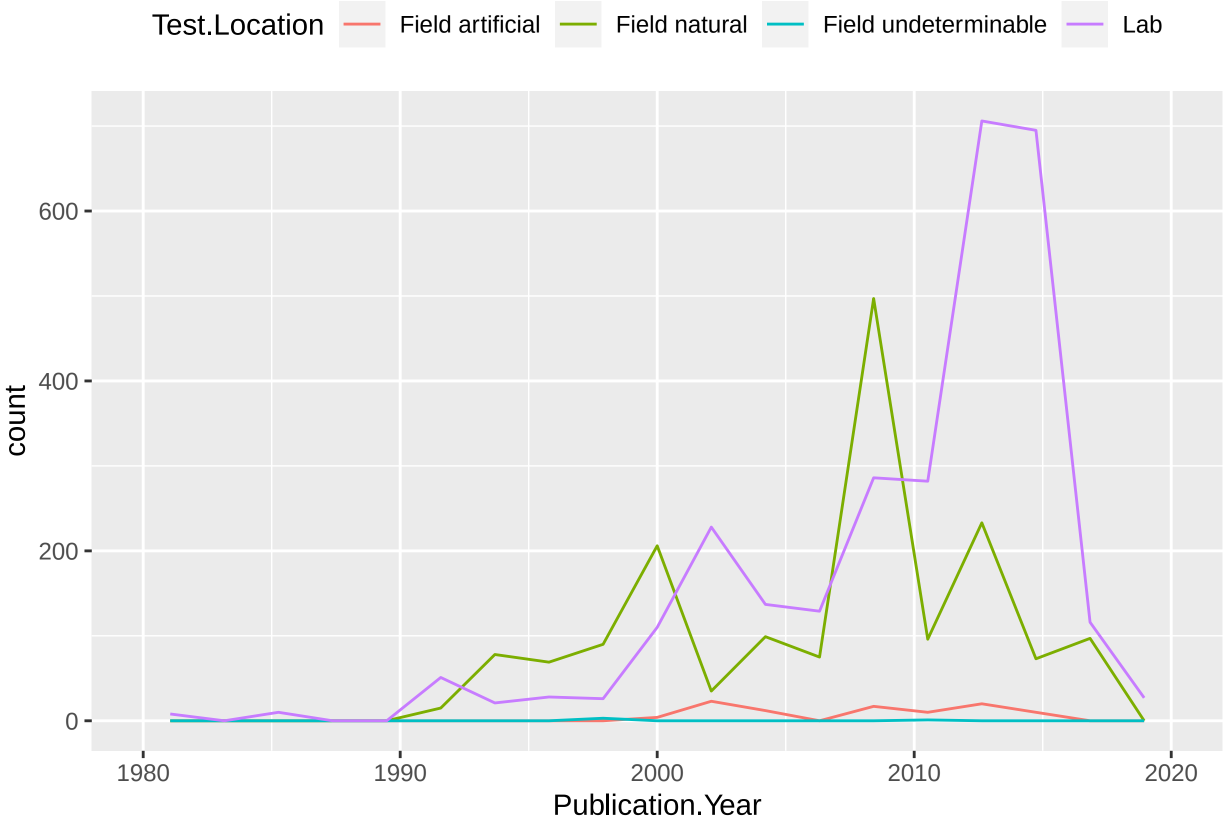
```
## Warning: Removed 2 row(s) containing missing values (geom_path).
```



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 = 20) +
  scale_x_continuous(limits = c(1980, 2020)) +
  theme(legend.position = "top")
```

```
## Warning: Removed 8 row(s) containing missing values (geom_path).
```

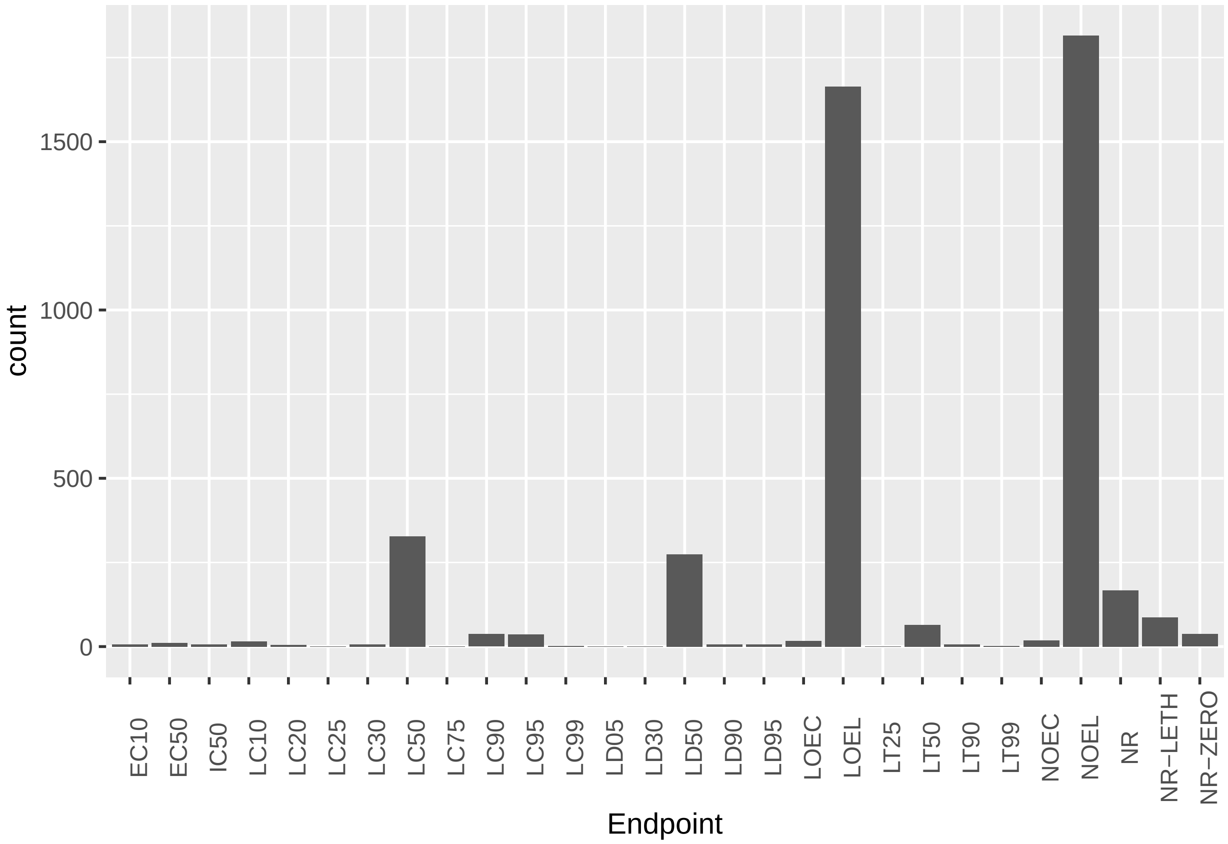


Interpret this graph. What are the most common test locations, and do they differ over time?

> Answer: The most common test locations are field natural and lab. While lab has most recently been the very dominant location, it was not consistently the most common location throughout the time period in this data. Field natural overtook lab in frequency in the years just before the beginning of the century and just before 2010. Field artificial is very uncommon, and field undeterminable is almost always 0.

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) +
  geom_bar(aes(x = Endpoint)) +
  theme(axis.text.x = element_text(angle=90))
```



Answer: The 2 most common endpoints are LOEL and NOEL. LOEL stands for Lowest Observable Effect Level - it is the lowest concentration [of the insecticide] that produces effects with a significant difference from control experiments. NOEL is No Observable Effect Level - it is the highest concentration that produces effects that are not significantly different from the control experiments. Identifying LOEL and NOEL as endpoints is a good way to make sure the concentration of neonicotinoids insects are exposed to will have a sufficient level for its use on pests and that this concentration will not have a significant effect on non-pests.

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] "character"
```

```
#character, not date format. current format is yyyy-mm-dd
Litter$collectDate <- as.Date(Litter$collectDate, format = "%Y-%m-%d")
class(Litter$collectDate)
```

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

```
unique(Litter$collectDate)
```

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

```
#litter was sampled on August 2nd and August 30th, 2018
```

Answer: Litter was sampled on the 2nd and 30th of August.

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`?

```
length(unique(Litter$plotID))
```

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

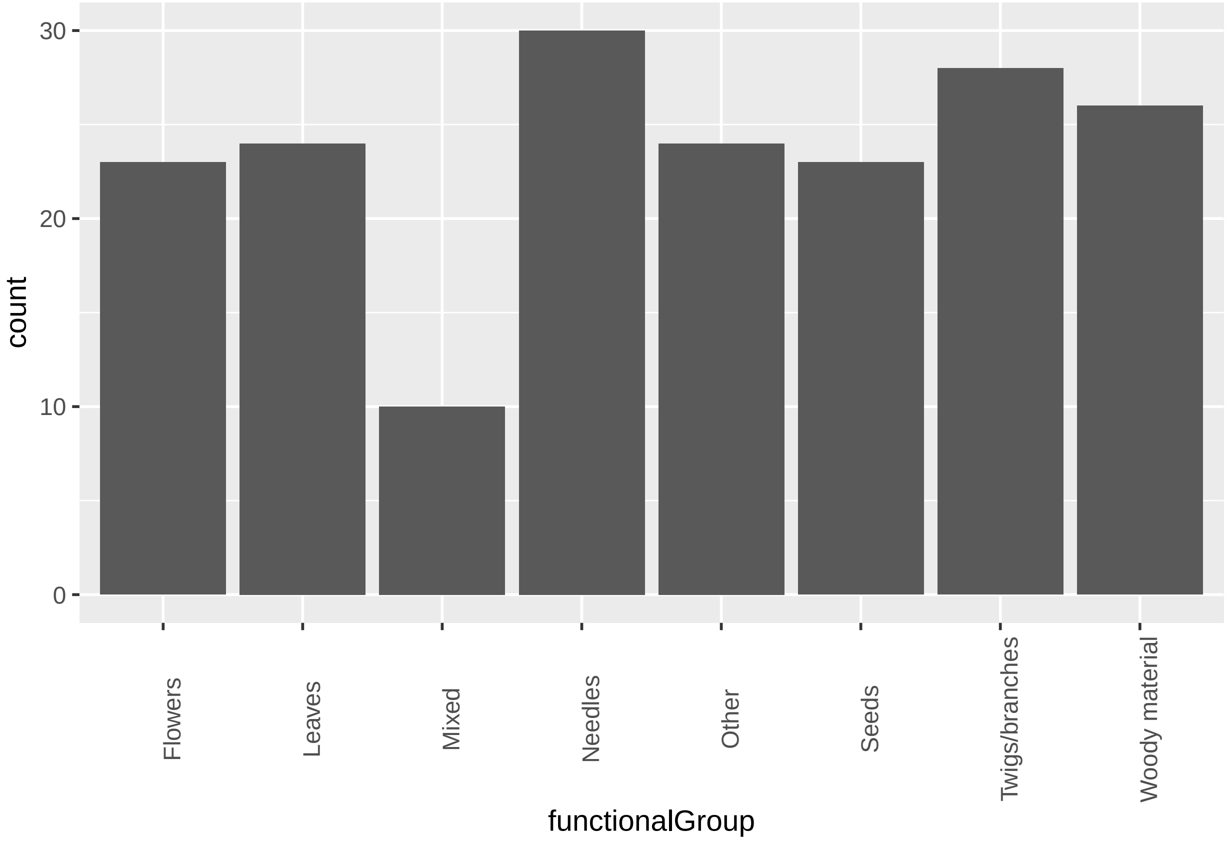
```
summary(Litter$plotID)
```

```
##    Length     Class      Mode 
##       188 character character
```

Answer: 12 plots were sampled at Niwot Ridge. Unique gives the unique objects in a column (or a data frame etc, but here we are specifying columns), whereas summary gives the summary statistics. In the case of a character column, summary will give you the number of objects in that column - NOT individual objects, but total objects, and not the objects themselves, but the number of them - as well as information about the type of data for those objects.

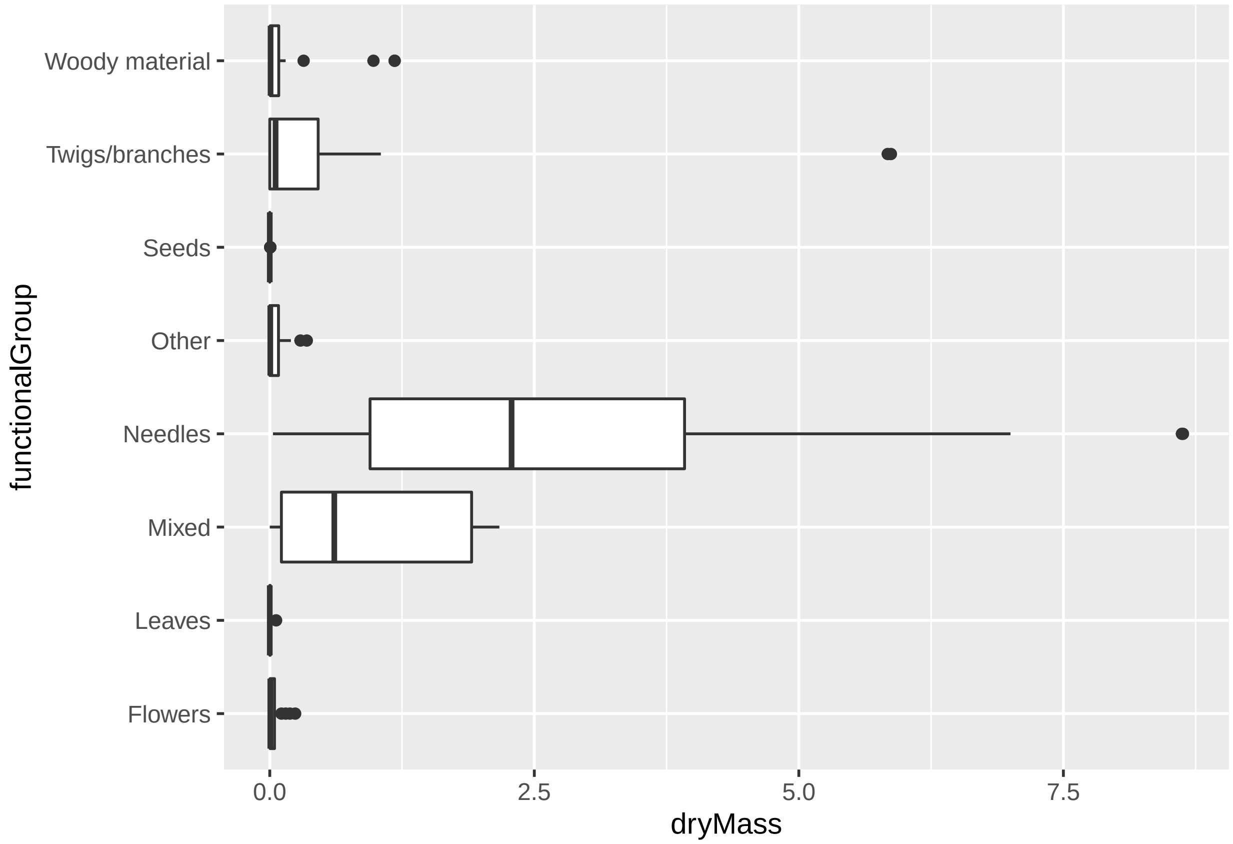
14. Create a bar graph of functionalGroup counts. This shows you what type of litter is collected at the Niwot Ridge sites. Notice that litter types are fairly equally distributed across the Niwot Ridge sites.

```
ggplot(Litter) +
  geom_bar(aes(x=functionalGroup)) +
  theme(axis.text.x = element_text(angle=90))
```

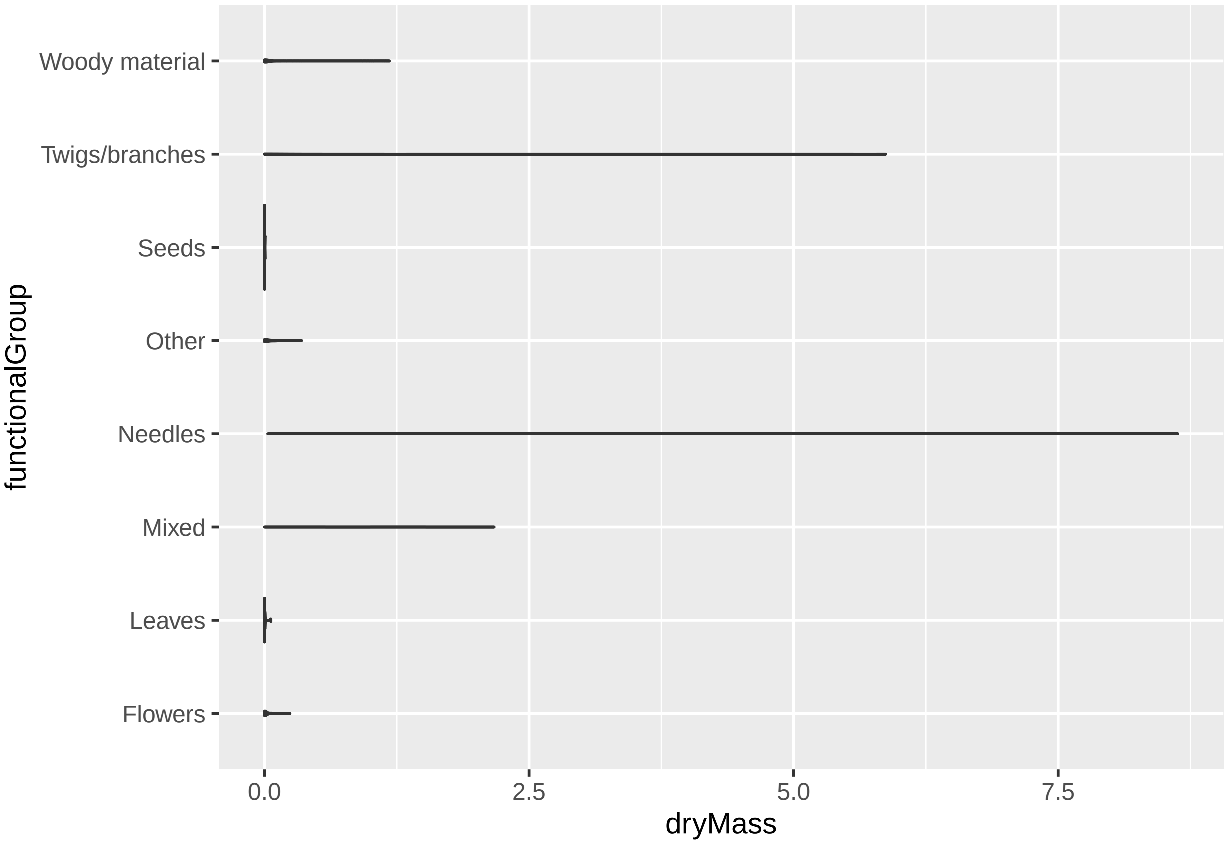


15. Using `geom_boxplot` and `geom_violin`, create a boxplot and a violin plot of dryMass by functionalGroup.

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



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



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

> Answer: Here, the violin plot looks more like flat, straight lines than it does normal, curvy violin plots. The curves in a normal violin plot come from differences in the distribution of values within each category (here, the frequency of biomass values for each litter type). In this dataset, there is not much variation amongst how values are distributed in several of the litter type categories. Additionally, outliers in categories like needles and twigs/branches elongate the violin plot, making it look like one long line and compressing our visualization of variation in the data closer to 0. In categories like seeds and leaves, we get vertical lines - there may be interesting visualization of distribution frequency in these plots, but the scale of our x-axis compresses this, as all of the data points in these categories are close to 0, but the graph is on a scale going up to 8. So, the nature of how the points are distributed, the x-axis scale, and outliers are all creating compressed-looking violin plots that are not as useful as boxplots for visualizing data distribution.

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

> Answer: Needles and mixed litter have the highest dry biomass at these sites. The biomass of the other litter materials tends to be quite small in the study areas.