

Assignment 3: Data Exploration

Fei Wu

Spring 2024

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.

```
#install.packages("tidyverse")
library(tidyverse)

Neonics <- read.csv("./Data/Raw/Neonics.csv", stringsAsFactors=TRUE)
Litter <- read.csv("./Data/Raw/NIWO_Litter/Litter.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: Key reasons for interest include concerns about the impact on non-target organisms, particularly pollinators like bees, potential effects on biodiversity and ecosystems, water contamination risks, regulatory considerations for pesticide use, and potential indirect effects on human health through the food chain. Understanding the ecotoxicological effects informs regulatory decisions and aids in mitigating harm to both ecosystems and human well-being.

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: Fallen debris plays a key role in nutrient cycling, providing habitats for diverse organisms, and influencing carbon sequestration. Monitoring litter and woody debris at Niwot Ridge contributes valuable insights into long-term ecosystem health, helping assess the impact of climate, disturbance, and land use changes. Additionally, it aids in fire ecology research, as the debris can act as both fuel and a barrier. Overall, this research enhances our understanding of how forest ecosystems function, adapt, and respond to environmental changes, providing essential information for effective ecosystem management and conservation strategies.

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. Ground taps are sampled once a year 2. Litter and fine woody debris are collected from elevated and ground traps respectively. 3. Litter is defined as material that is dropped from the forest canopy and has a butt end diameter <2cm and a length <50cm

Obtain basic summaries of your data (Neonics)

5. What are the dimensions of the dataset?

```
Neonics.data.complete <- na.omit(Neonics)
dim(Neonics)
```

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

```
dim(Neonics.data.complete)
```

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

```
## Dimensions is 4623,30
```

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: Mortality and Population is the most common effectss.

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...]

```
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

#Honey Bee, Parasitic Wasp, Buff Tailed Bumblebee, Carniolan Honey Bee, Bumble Bee, 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?

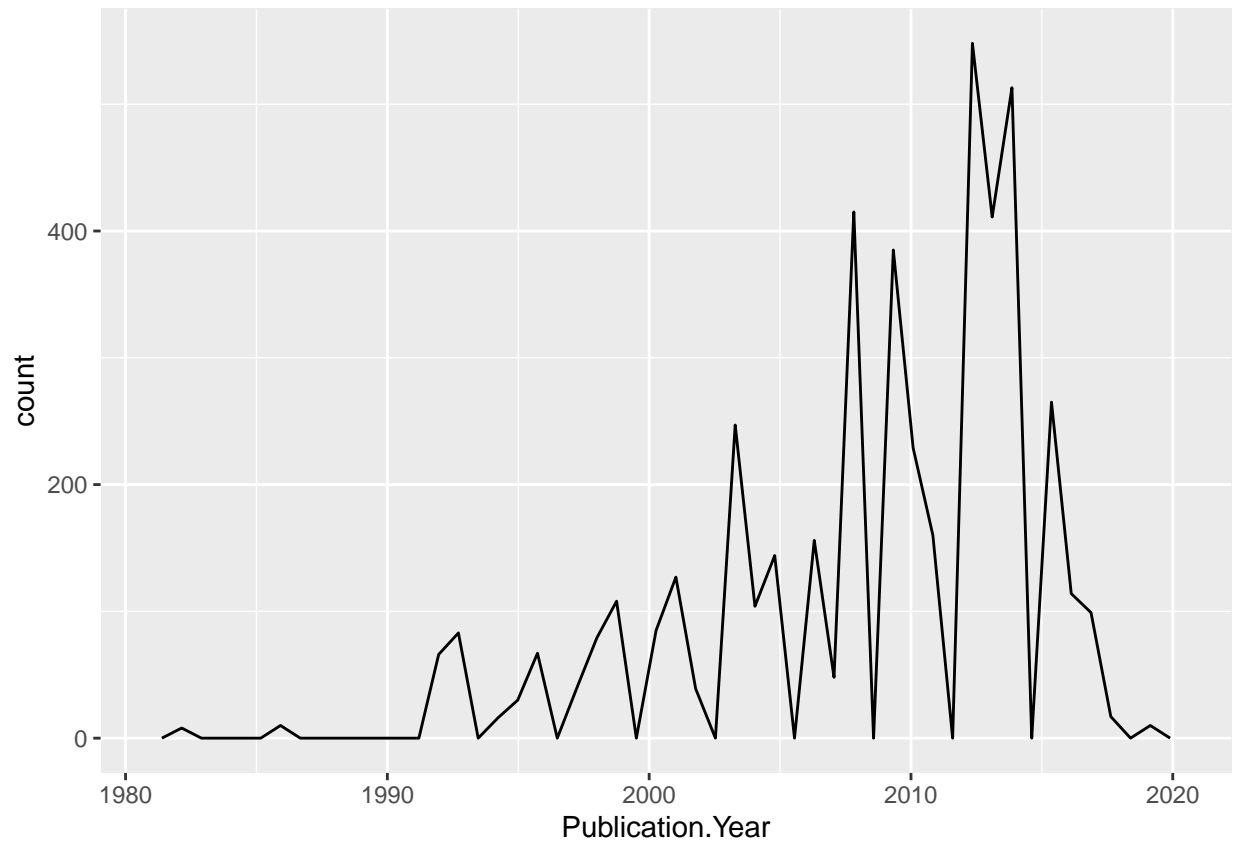
```
class(Neonics$Conc.1..Author)
```

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

Answer: It's factor because they are not just all numbers and we read them as strings earlier. ##
Explore your data graphically (Neonics)

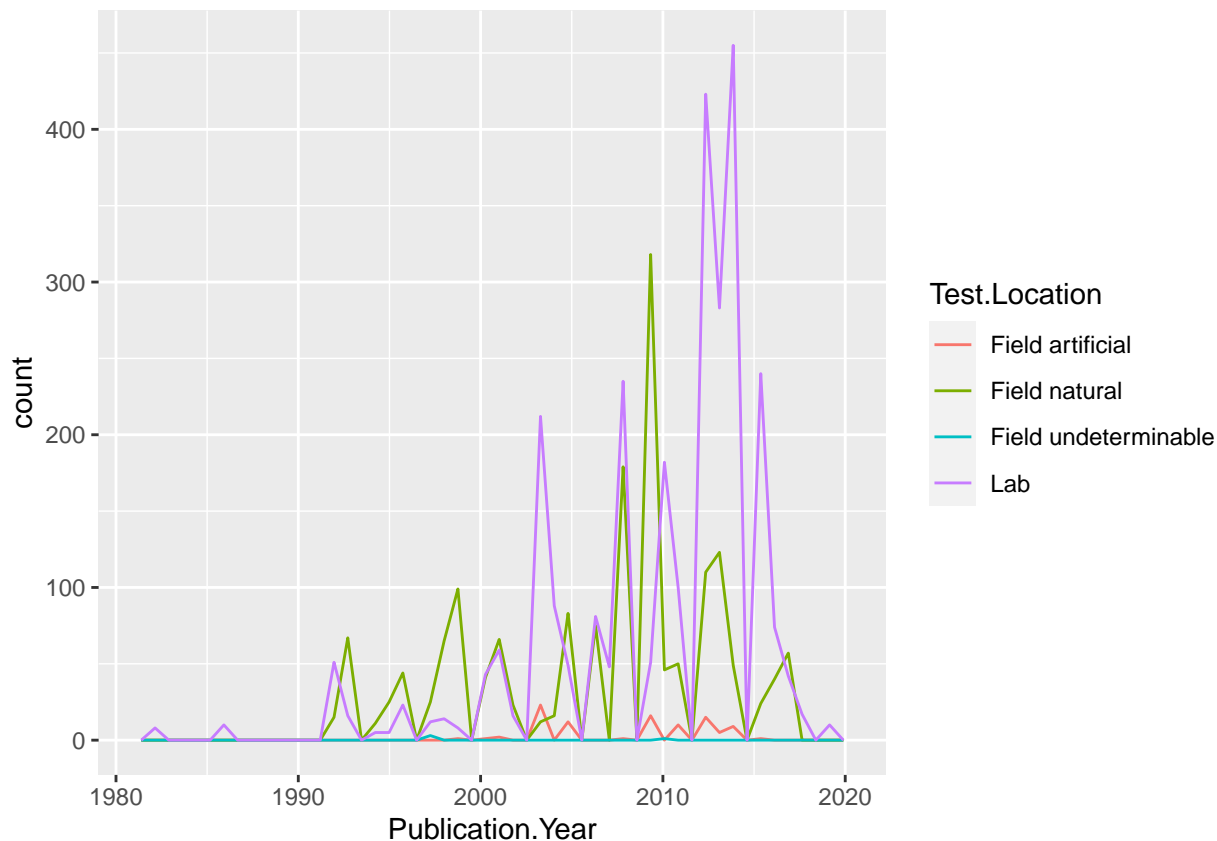
9. Using `geom_freqpoly`, generate a plot of the number of studies conducted by publication year.

```
ggplot(Neonics) +  
  geom_freqpoly(aes(x = Publication.Year), bins = 50)
```



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 = 50)
```



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

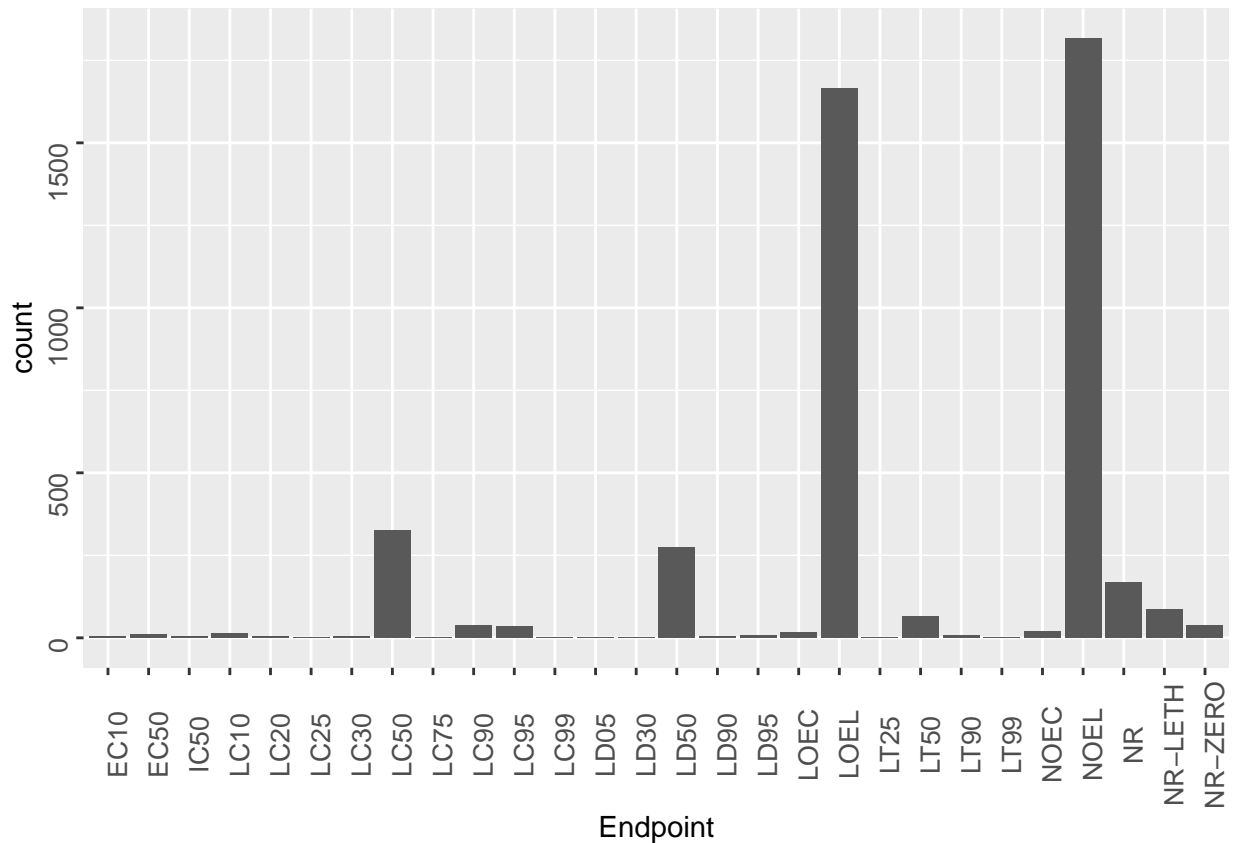
Answer: Lab. they do differ over time. Between 1990 to 2000, field natural is low but becomes high in 2010. Field artificial did not appear until late 2000 to 2010. Lab was always there and peaked between 2010 to 2020.

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_bar(aes(x = Endpoint), bins = 50) + theme(axis.text= element_text(size=10, angle=90))
```

```
## Warning in geom_bar(aes(x = Endpoint), bins = 50): Ignoring unknown parameters:
## 'bins'
```



Answer: LOEL and NOEL

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"
```

```
Litter$collectDate <-as.Date(Litter$collectDate)
class(Litter$collectDate)
```

```
## [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`?

```
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
```



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

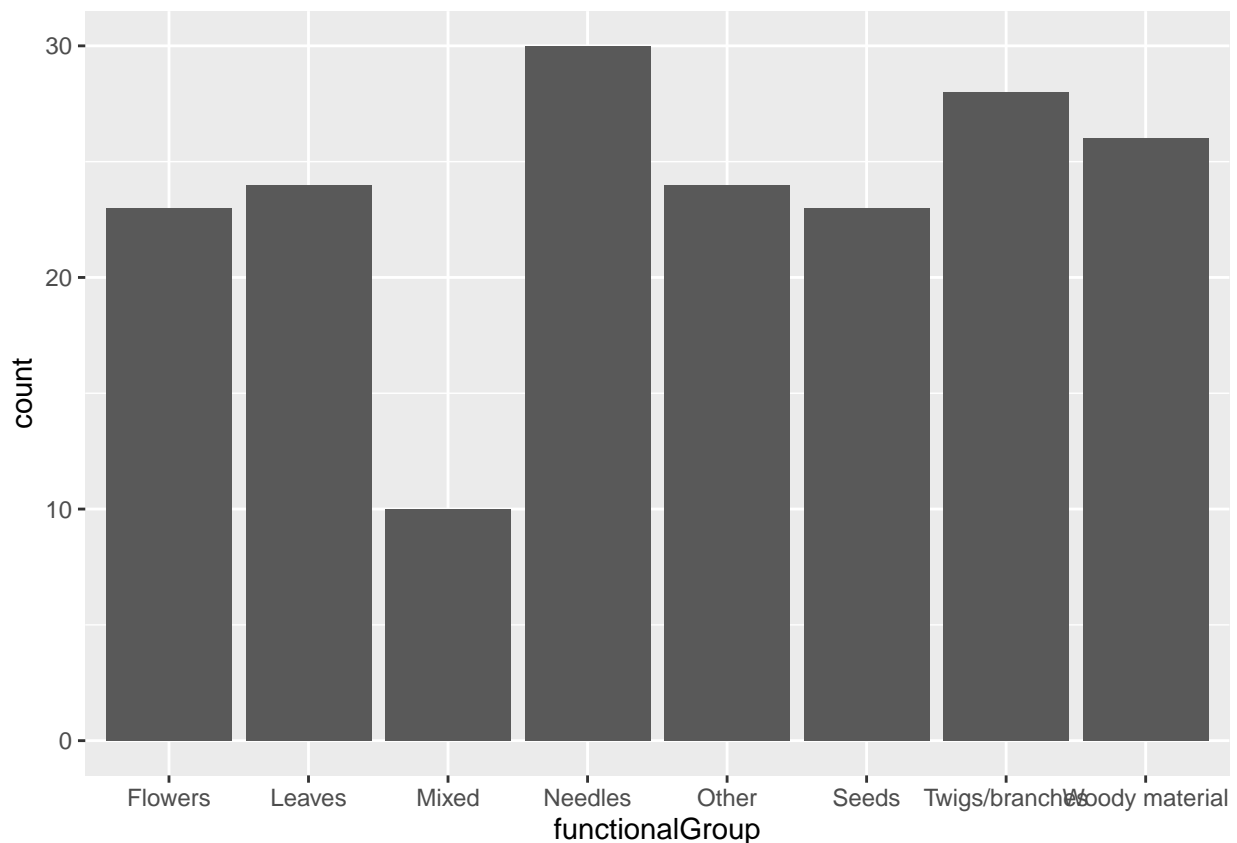
```
## NIWO_040 NIWO_041 NIWO_046 NIWO_047 NIWO_051 NIWO_057 NIWO_058 NIWO_061
##      20      19      18      15      14       8      16      17
## NIWO_062 NIWO_063 NIWO_064 NIWO_067
##      14      14      16      17
```

Answer: levels means listing out every unique value ; summary will tell us the unique values and the number of them appear

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), bins = 50)
```

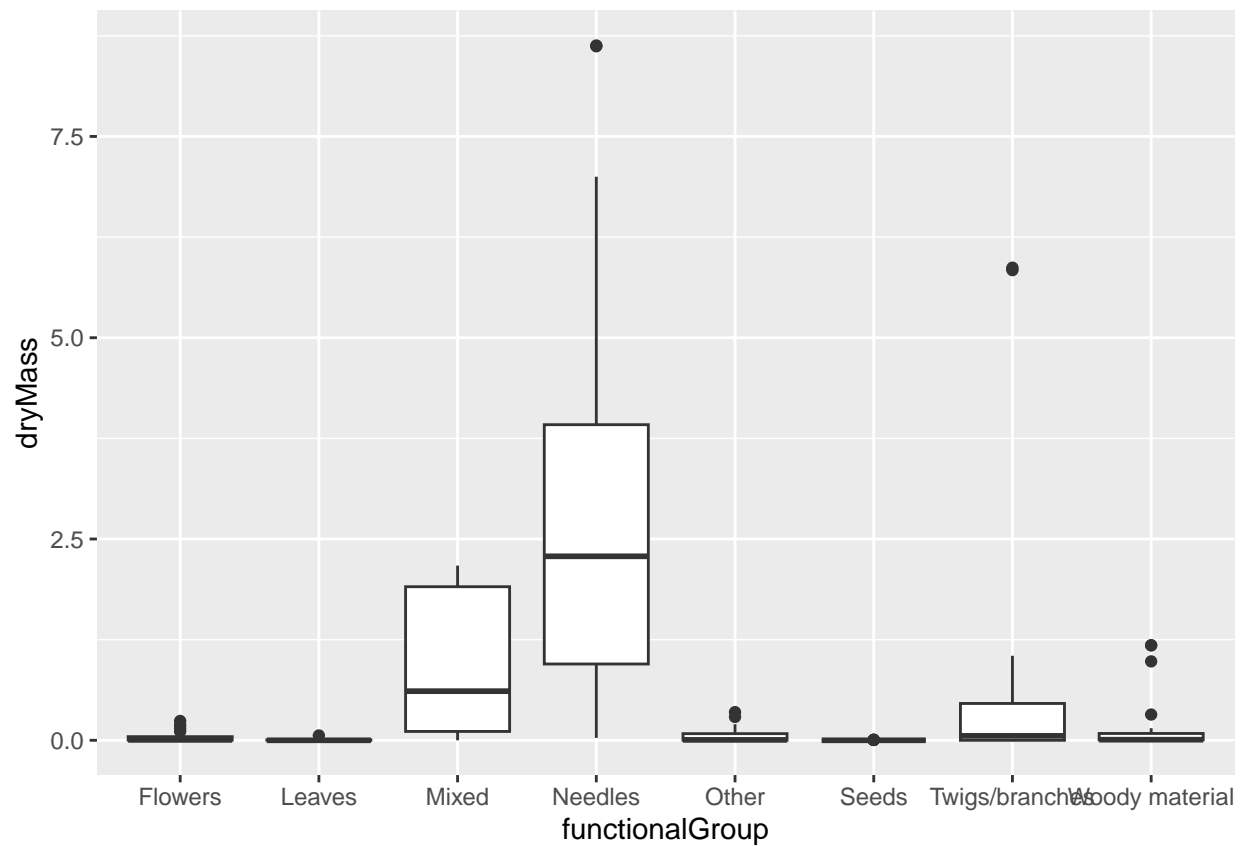
```
## Warning in geom_bar(aes(x = functionalGroup), bins = 50): Ignoring unknown
## parameters: 'bins'
```



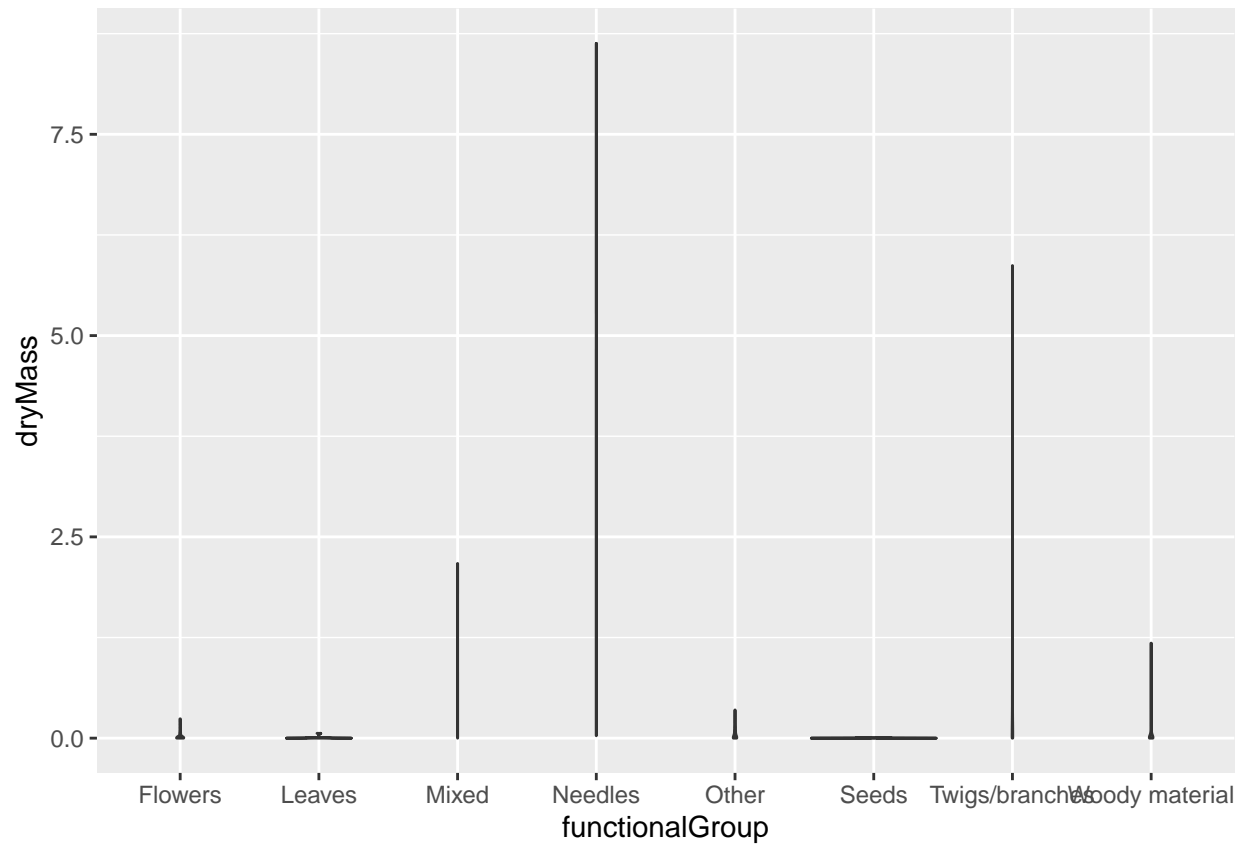
15. Using `geom_boxplot` and `geom_violin`, create a boxplot and a violin plot of `dryMass` by `functionalGroup`.

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

```
## Warning in geom_boxplot(aes(x = functionalGroup, y = dryMass), bins = 50):  
## Ignoring unknown parameters: 'bins'
```



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



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

#Answer: not enough values for violin.

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

Answer: #Needles