Assignment 3: Data Exploration

Britney Pepper, Section #1

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 <1/31/22>.

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.**

```
library(knitr)
opts_chunk$set(tidy.opts=list(width.cutoff=60),tidy=TRUE, echo = T)

#set working directory
getwd()
```

```
## [1] "/Users/britneypepper/Desktop/ENVIRON 872 and L/GitHubRepositories/Environmental_Data_Analytics_1
```

```
setwd("/Users/britneypepper/Desktop/ENVIRON 872 and L/GitHubRepositories/Environmental_Data_Analytics_20

#install necessary packages
#install.packages("dplyr")
#install.packages("ggplot2")
#install.packages("lubridate")

library(lubridate)
library(dplyr)
library(ggplot2)
```

```
library(tidyverse)
library(lubridate)

#load the data
Neonics <- read.csv("/Users/britneypepper/Desktop/ENVIRON 872 and L/GitHubRepositories/Environmental_Da
                    stringsAsFactors = TRUE)

Litter <- read.csv("/Users/britneypepper/Desktop/ENVIRON 872 and L/GitHubRepositories/Environmental_Data
                   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 might be interested in this because it helps control or harm insect populations and in turn allow for the production of more crops. Also, if we learn more about which biological and chemical agents are most effective in insects, then we can more easily target species or reduce damage to species and in turn the benefits/harms to agriculture. This in turn could make everything more efficient as there is more pollination of plants, or it could be worse if harmful insect populations are not monitored. It could also help farmers to determine how much insecticide they should or if they should switch insecticides.

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 woddy debris that fall to the ground in forests can have significant impacts on the health of ecosystem. One reason for studying would be to monitor or control forest fires. If there is lots of leaf litter and woody debris, this can easily catch fire and spread far and for long periods of time. However, this can also be beneficial because it allows for some plant die-off to allow for new plants to take their place. It also allows for nutrients to re-enter the system through decomposition.

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 occurs at terrestrial NEON sites that contain woody vegetation >2m tall* litter sampling is targeted to take place in 20, 40m x 40m plots *plot edges must be separated by a distance 150% of one edge of the plot

Obtain basic summaries of your data (Neonics)

5. What are the dimensions of the dataset?

```
dim(Neonics)
```

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

```
#returned 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?

```
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: These effects might specifically be of interest because they look at the attributes that the insets portray. It would help show whether or not the insecticide is effective on the population.

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
```

```
#most commonly studied species: Honey Bee (667), Parasitic Wasp (285),
#Buff Tailed Bumblebee (183), Carniolan Honey Bee (152),
#Bumble Bee (140), and Italian Honeybee (113).
```

Answer: The most commonly studied species are Honey Bee (667), Parasitic Wasp (285), Buff Tailed Bumblebee (183), Carniolan Honey Bee (152), Bumble Bee (140), and Italian Honeybee (113). All of these species have stingers and have hives. These species might be of interest over others because of the endangerment that they face and the potential threats we can face if we loose them. Bees are one of the main sources of pollination of plants. Honey bees in particular produce most of the honey that we consume. They faced disease in recent years that has lowered their populations and made honey scarce while decreasing the yield of agricultural crops. Therefore, it becomes imporatant to study bees to help our ensure that we do not loose important species. Parasitic wasps are a little different, but they have been known to harm bee populations, so it is likely included in many studies because it is a direct threat to bees.

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

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

```
##    0.37/      10/      NR/      NR        1     1023    0.40/       2/
##      208      127      108       94       82       80       69       63
##       10   0.053/      100      50/     0.5/     0.03    0.05/     0.45
##       62       59       56       51       45       44       43       43
```

```
##     0.1/    0.45/     1.0/    2.27/       50    0.125     500/      0.5
##       42       40       40       40       36       33       33       32
##   0.048/    0.15/       1/       48    25.0/      12/    0.027      2.4
##       30       30       30       30       28       27       26       26
##     0.2/    0.56/     100/        3    0.01/    1000/       3/    0.336
##       25       24       23       23       22       22       22       21
##     1.5/    0.05      1.5    2.60/    20.0/        6    6.80/    62.5/
##       21       20       20       20       20       20       20       20
##    0.005     0.4/    0.18/     0.3/     1000       40 0.00355/      0.1
##       18       18       17       17       17       17       16       16
##      0.4     150/      300      80/    0.053     0.24     0.28     125/
##       16       16       16       16       15       15       15       15
##        9   0.0001  0.0004/   0.084/     0.15      0.6    12.5/   144.0/
##       15       14       14       14       14       14       14       14
##     350/    40.0/      48/       56       84/    0.17/      125       14
##       14       14       14       14       14       13       13       13
##       16       17   0.047/    0.25/    0.28/    1.28/    1.81/      112
##       13       13       12       12       12       12       12       12
##      150     2.5/       25      60/      75/    0.02/   0.025/     0.29
##       12       12       12       12       12       11       11       11
##    37.5/       4/        5  (Other)
##       11       11       11     1817
```

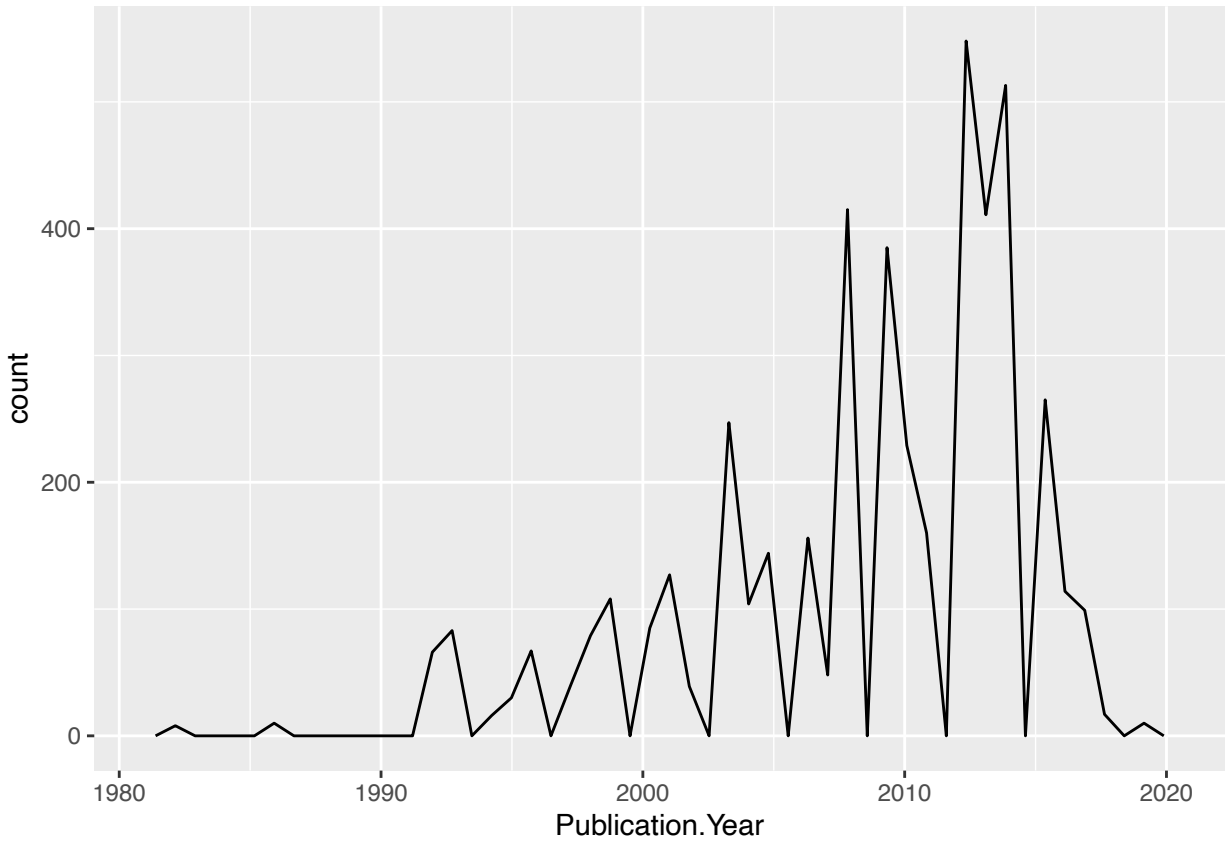
```
#producing a new data frame to see the data
conc_1_author <- data.frame(Neonics$Conc.1..Author.)
#shows multiple types of variables like fractions, whole numbers, decimals, etc.

#testing to see what values are given to the column of data
conc_1_author_test <- Neonics$Conc.1..Author.
#shows that it is considered a "Factor" that has 1006 levels.
```

Answer: I believe that it is not numeric because the class Conc.1..Author has multiple types of variables in each row. There were fractions, whole numbers, decimals, values of "NR/", etc. I think that because there are many different levels that R cannot process the values as numeric. I belive that the class is considered a Factor.

Explore your data graphically (Neonics)

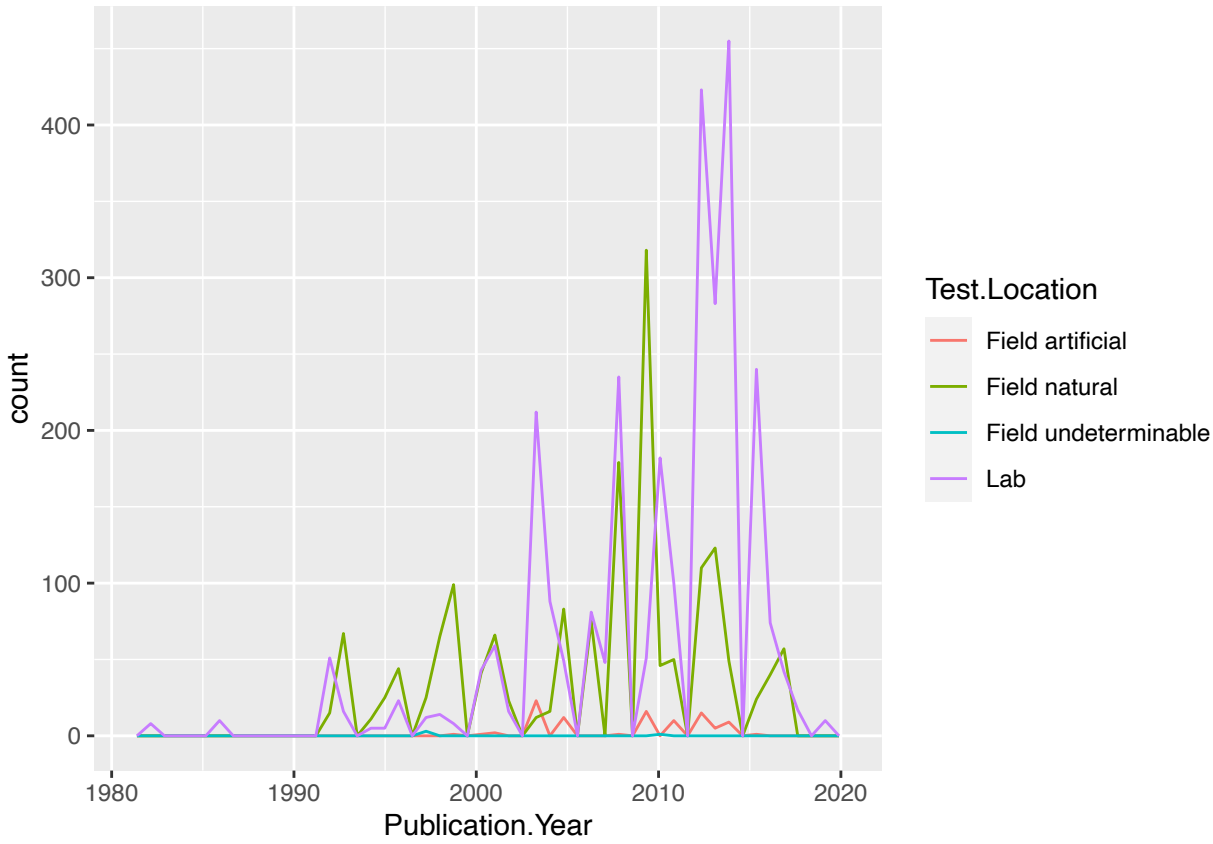
9. Using `geom_freqpoly`, generate a plot of the number of studies conducted by publication year.

```
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)
```

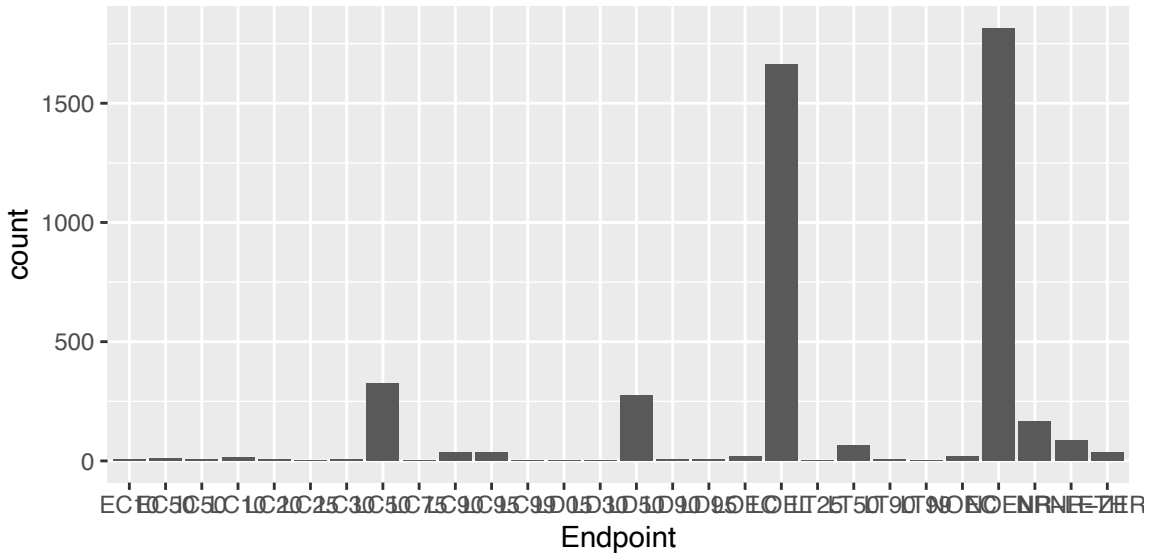


Interpret this graph. What are the most common test locations, and do they differ over time?

> Answer: The most common test locations are the Lab and Field Natural. These do differ over time; it appears that Field Natural is the most common from 1992-1999, 2008-2009, and a couple of scattered years. The rest of the time appears to favor the Lab.

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)) +
  geom_bar()
```




```
#two most common are LOEL and NOEL
```

Answer: The two most common are LOEL and NOEL. LOEL is defined as, "Lowest-observable-effect-level: lowest dose (concentration) producing effects that were significantly different (as reported by authors) from responses of controls (LOEAL/LOEC)." NOEL is defined as, "No-observable-effect-level: highest dose (concentration) producing effects not significantly different from responses of controls according to author's reported statistical test (NOEAL/NOEC)."

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.

```
#testing to see if collectDate is a date
test_date_collected <-ymd(Litter$collectedDate)
#conclusion: not a date


#Changing to a date
collectedDate <- Litter$collectDate
View(collectedDate)

date_Collected <- ymd(collectedDate)
View(date_Collected)
#Now in date format

#dates that Litter was sampled
unique(date_Collected)
```

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

```
#unique tells you how many unique IDs there are by listing them out
#--> can see that there are 12 and it lists them out

#summary
summary(Litter$plotID)
```

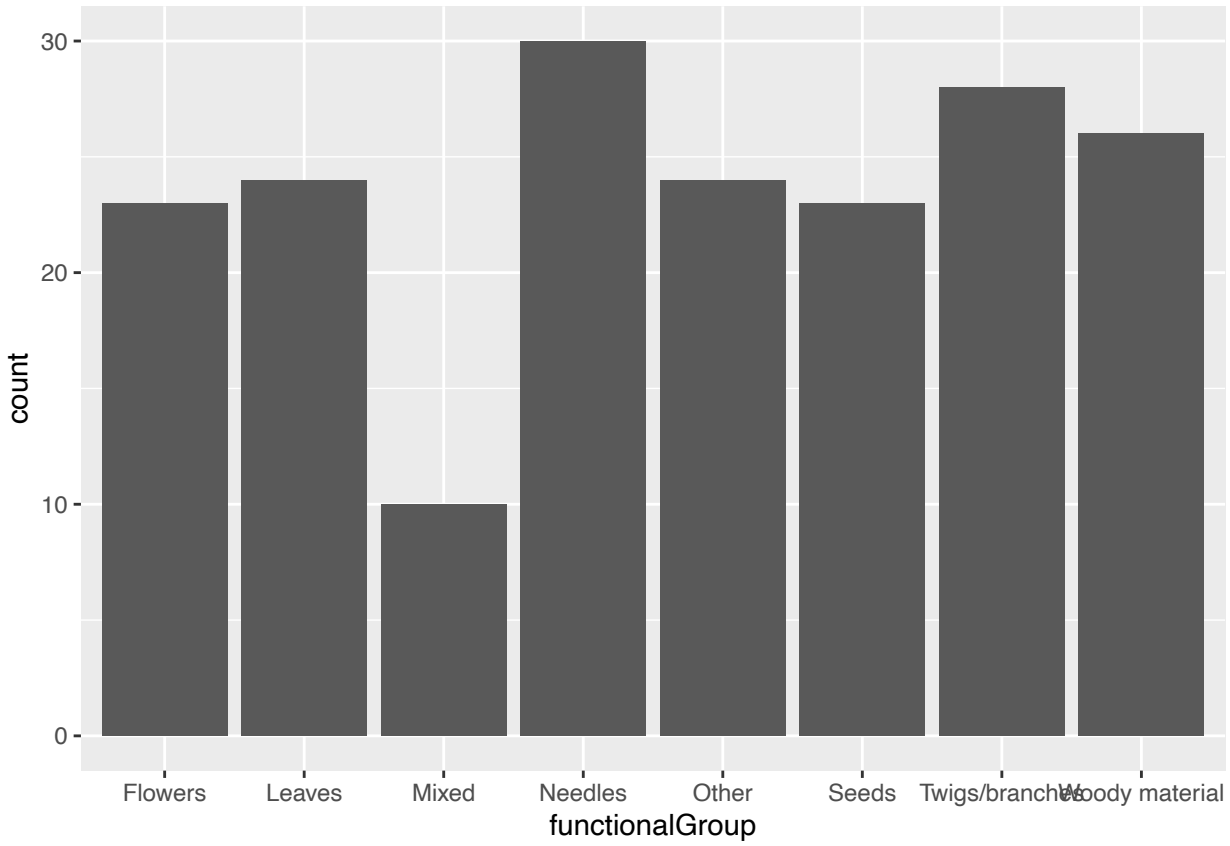
```
## NIWO_040 NIWO_041 NIWO_046 NIWO_047 NIWO_051 NIWO_057 NIWO_058 NIWO_061
##       20       19       18       15       14        8       16       17
## NIWO_062 NIWO_063 NIWO_064 NIWO_067
##       14       14       16       17
```

```
#summary tells you how many times each unique ID was sampled
#--> it still tells you how many plots were sampled, but you have to manually count that
```

Answer: When using the 'unique' function, you can see that there are '12 Levels', and it lists them out based on their ID. This is different from the 'summary' because that still gives you each ID, but it lists out how many times each plot was sampled. Summary still shows you how many plots were sampled, but it does not count them individually for you, so you would have to manually do that yourself.

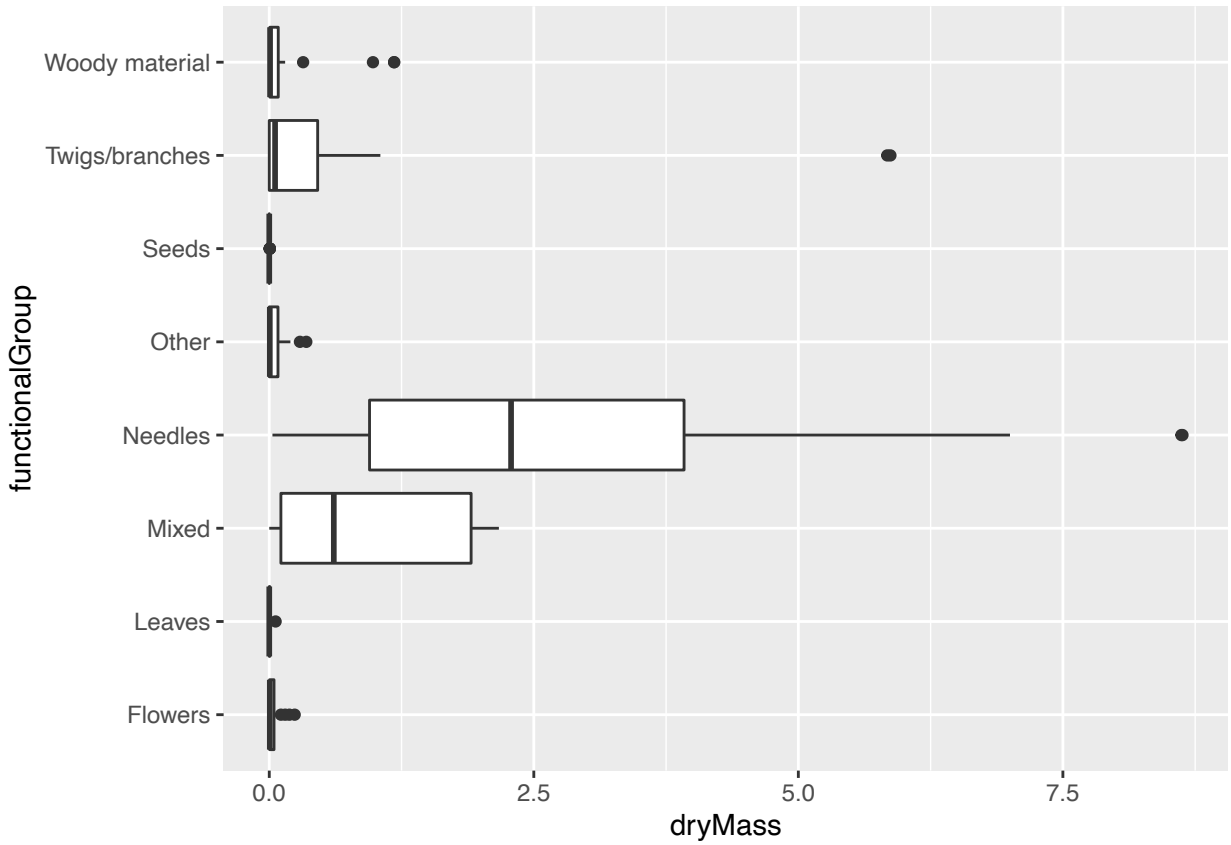
14. Create a bar graph of functionalGroup counts. This shows you what type of litter is collected at the Niwot Ridge sites. Notice that litter types are fairly equally distributed across the Niwot Ridge sites.

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



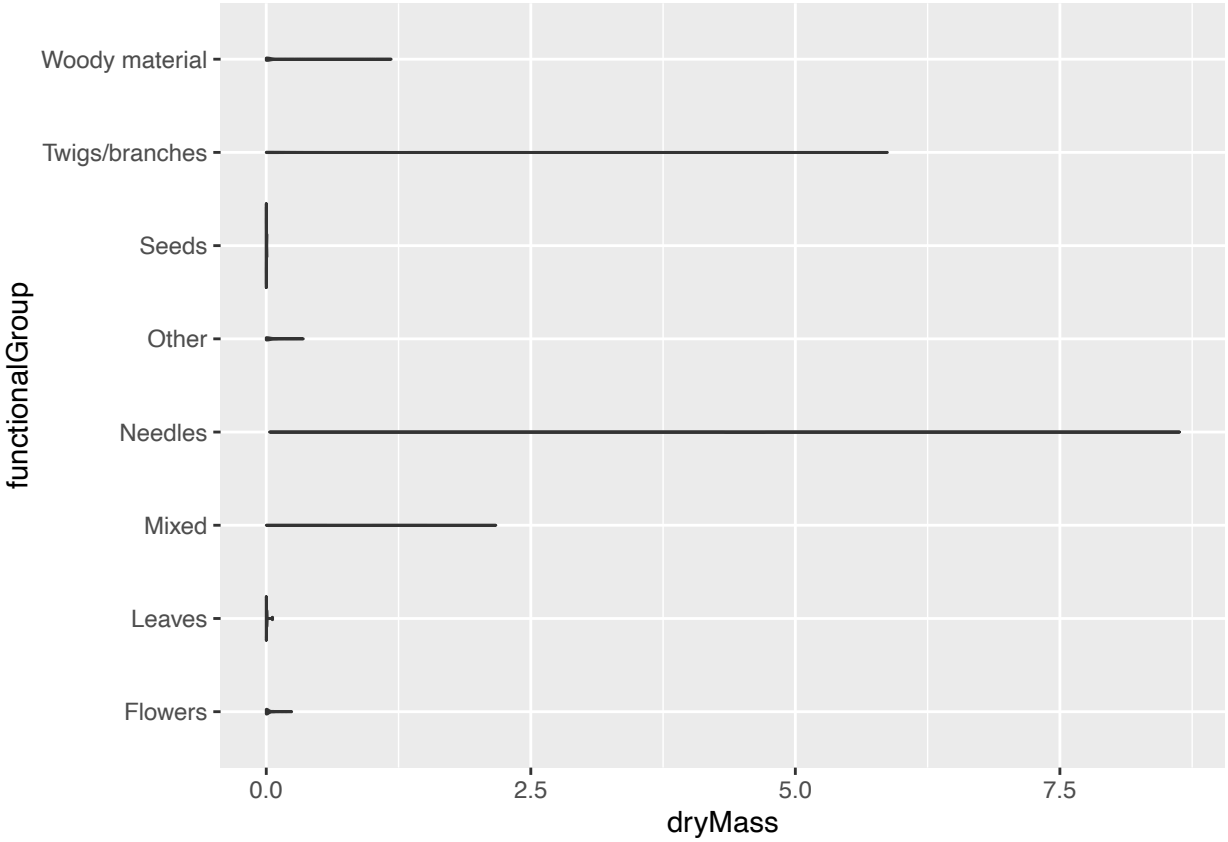
15. Using `geom_boxplot` and `geom_violin`, create a boxplot and a violin plot of dryMass by functionalGroup.

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




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

```
## Warning in regularize.values(x, y, ties, missing(ties), na.rm = na.rm):
## collapsing to unique 'x' values

## Warning in regularize.values(x, y, ties, missing(ties), na.rm = na.rm):
## collapsing to unique 'x' values

## Warning in regularize.values(x, y, ties, missing(ties), na.rm = na.rm):
## collapsing to unique 'x' values
```



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

> Answer: The boxplot gives you more information about the data. It shows you the range of values (including outliers), and gives you an idea of where the mean is. The violin plot just shows you the range.

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

> Answer: The type of litter that has the highest biomass at these sites is Needles. Needles is obviously the highest with a mean greater than the maximum values of majority of the other types of litter. The second highest biomass is Mixed litter, which is greater than the rest of the types of litter.