Assignment 3: Data Exploration

Charles :-)

Spring 2023

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.

```
tinytex::install_tinytex()
setwd("C:/Undergrad/Grad_School/General/Jasmine/R_stuff")
library(tidyverse)

#load datasests
Neonics <- read.csv("EDA-Spring2023/Data/Raw/ECOTOX_Neonicotinoids_Insects_raw.csv",stringsAsFactors = 

Litter <- read.csv("EDA-Spring2023/Data/Raw/NIWO_Litter/NEON_NIWO_Litter_massdata_2018-08_raw.csv",stri
```

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:

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:

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. 2. 3.

Obtain basic summaries of your data (Neonics)

5. What are the dimensions of the dataset?

```
dim(Neonics)
```

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

```
#dimensions are 4623x30
```

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?

```
#to see all effects
sort(summary(Neonics$Effect), decreasing=TRUE)
```

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

```
#population, mortality by a large margin. explain why below
```

Answer:

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...]

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

```
##                          Ant Family                         Apple Maggot
##                                   9                                    9
##              Glasshouse Potato Wasp                             Lacewing
##                                  10                                   10
##             Southern House Mosquito               Two Spotted Lady Beetle
##                                  10                                   10
##            Spotless Ladybird Beetle                   Braconid Parasitoid
##                                  11                                   12
##                          Common Thrip          Eastern Subterranean Termite
##                                  12                                   12
##                               Jassid                           Mite Order
##                                  12                                   12
##                           Pea Aphid                     Pond Wolf Spider
##                                  12                                   12
##              Armoured Scale Family                     Diamondback Moth
##                                  13                                   13
##                        Eulophid Wasp                    Monarch Butterfly
##                                  13                                   13
##                        Predatory Bug                Yellow Fever Mosquito
##                                  13                                   13
##                         Corn Earworm                    Green Peach Aphid
##                                  14                                   14
##                            House Fly                             Ox Beetle
##                                  14                                   14
##                   Red Scale Parasite                   Spined Soldier Bug
##                                  14                                   14
##              Western Flower Thrips Hemlock Woolly Adelgid Lady Beetle
##                                  15                                   16
##              Hemlock Wooly Adelgid                                 Mite
##                                  16                                   16
##                          Onion Thrip                Araneoid Spider Order
##                                  16                                   17
##                            Bee Order                      Egg Parasitoid
##                                  17                                   17
##                         Insect Class             Moth And Butterfly Order
##                                  17                                   17
##       Oystershell Scale Parasitoid            Black-spotted Lady Beetle
##                                  17                                   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
##               Codling Moth      Flatheaded Appletree Borer
##                          19                          20
##          Horned Oak Gall Wasp             Leaf Beetle Family
##                          20                          20
##             Potato Leafhopper      Tooth-necked Fungus Beetle
##                          20                          20
##                Argentine Ant                      Beetle
##                          21                          21
##                  Mason Bee                    Mosquito
##                          22                          22
##              Citrus Leafminer             Ladybird Beetle
##                          23                          23
##             Spider/Mite Class          Tobacco Flea Beetle
##                          24                          24
##                Chalcid Wasp          Convergent Lady Beetle
##                          25                          25
##                Stingless Bee          Ground Beetle Family
##                          25                          27
##           Rove Beetle Family                Tobacco Aphid
##                          27                          27
##                Scarab Beetle                Spring Tiphia
##                          29                          29
##                 Thrip Order          Ladybird Beetle Family
##                          29                          30
##                  Parasitoid                Braconid Wasp
##                          30                          33
##                Cotton Aphid                Predatory Mite
##                          33                          33
##          Sweetpotato Whitefly                 Aphid Family
##                          37                          38
##                Cabbage Looper          Buff-tailed Bumblebee
##                          38                          39
##                True Bug Order       Sevenspotted Lady Beetle
##                          45                          46
##                 Beetle Order    Snout Beetle Family, Weevil
##                          47                          47
##           Erythrina Gall Wasp              Parasitoid Wasp
##                          49                          51
##         Colorado Potato Beetle                Parastic Wasp
##                          57                          58
##           Asian Citrus Psyllid             Minute Pirate Bug
##                          60                          62
##              European Dark Bee                    Wireworm
##                          66                          69
##               Euonymus Scale             Asian Lady Beetle
##                          75                          76
##              Japanese Beetle              Italian Honeybee
##                          94                         113
##                  Bumble Bee           Carniolan Honey Bee
```

```
##                                140                                152
##              Buff Tailed Bumblebee                     Parasitic Wasp
##                                183                                285
##                          Honey Bee                            (Other)
##                                667                                670
```

```
#honey bee, parasitic wasp, buff tailed bumblebee, carniolan honey bee, bumble bee, italian honeybee
```

Answer:

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"
```

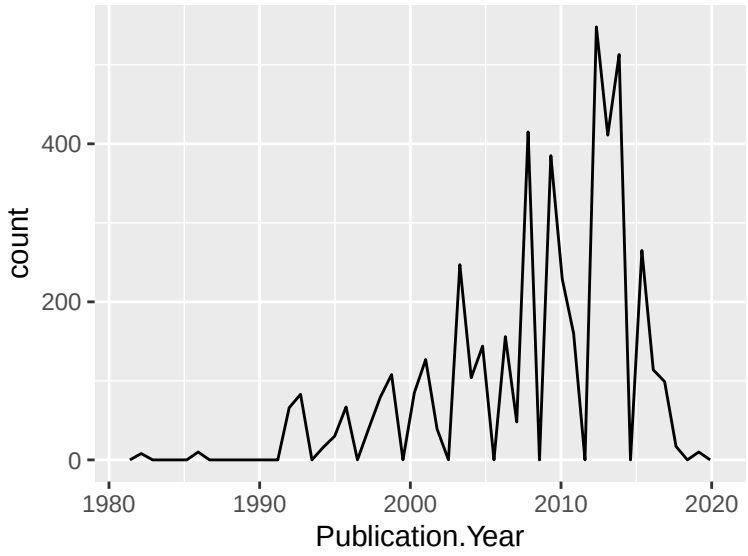
```
#its a factor class, factors are used to represent categorical data. Even though the concentrations are
```

Answer:

Explore your data graphically (Neonics)

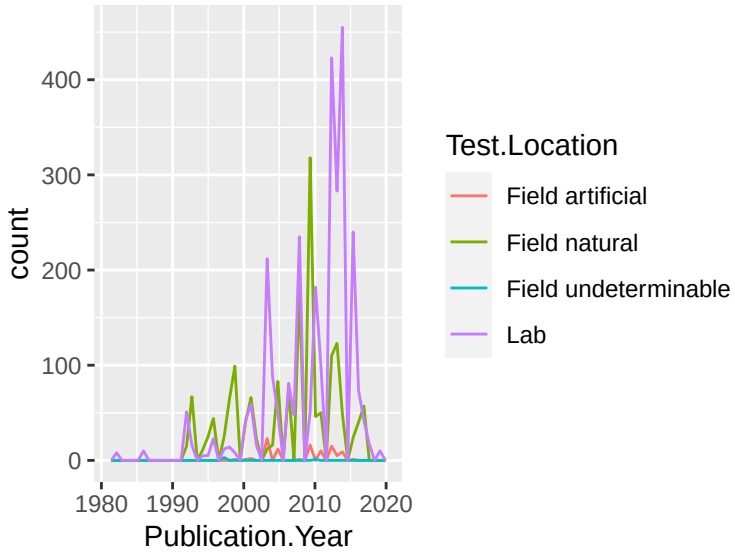
9. Using `geom_freqpoly`, generate a plot of the number of studies conducted by publication year.

```
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)
```



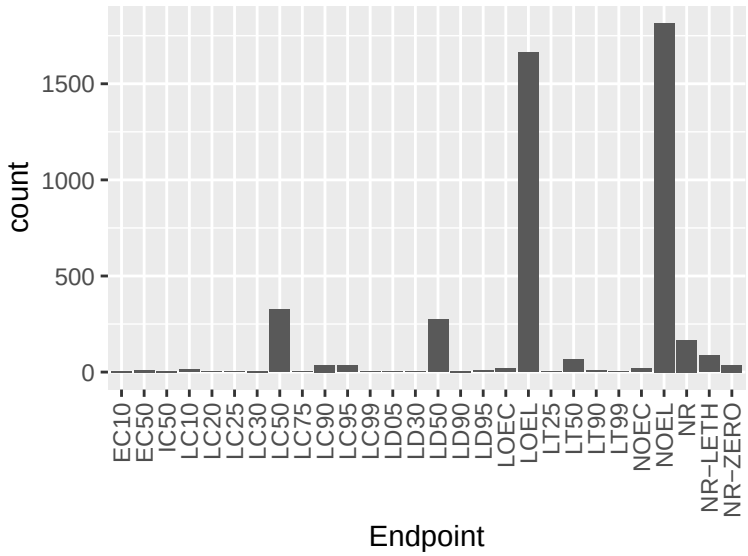
Interpret this graph. What are the most common test locations, and do they differ over time?

Answer: u do this

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, aes(x = Endpoint)) +
  geom_bar() +
  theme(axis.text.x = element_text(angle = 90, vjust = 0.5, hjust=1))
```



Answer: two most common are LOEC, NOEL - u look at the appendix i just coder man

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"
```

```
#its another factor class by default

#change to date
Litter$collectDate <- as.Date(Litter$collectDate, format = "%Y-%m-%d")

#confirm new class type
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`?

```
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
```

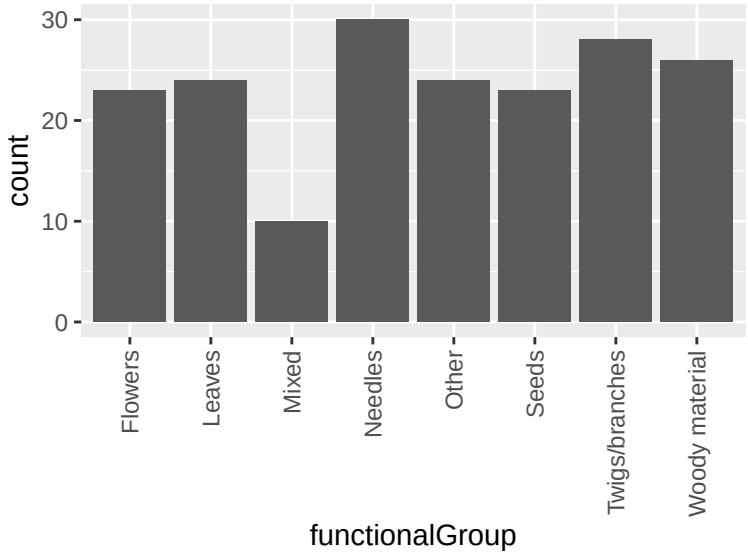
```
unique(Litter$plotID)
```

```
##  [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
```

Answer: unique doesn't include duplicates, and printed output is more compact. I think the answer they're looking for is 12 - but unique didn't really help me because the data is all factor classes

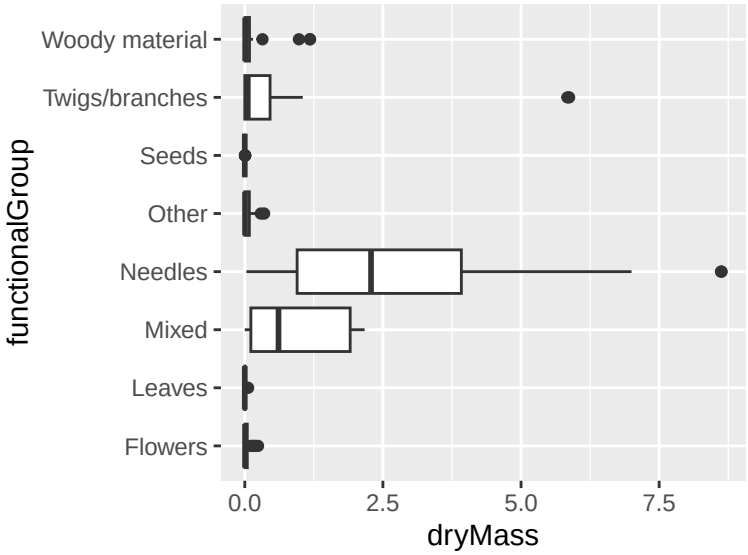
14. Create a bar graph of functionalGroup counts. This shows you what type of litter is collected at the Niwot Ridge sites. Notice that litter types are fairly equally distributed across the Niwot Ridge sites.

```
ggplot(Litter, aes(x = functionalGroup)) +
  geom_bar() +
  theme(axis.text.x = element_text(angle = 90, vjust = 0.5, hjust=1))
```

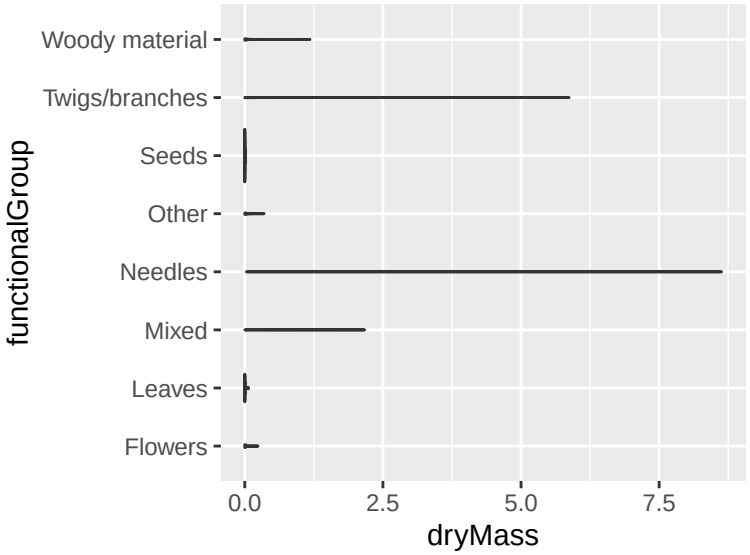


15. Using `geom_boxplot` and `geom_violin`, create a boxplot and a violin plot of dryMass by functionalGroup.

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




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



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

Answer: boxy

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

Answer: idk u read the plot