Assignment 3: Data Exploration

Hannah Smith

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 Tuesday, January 28 at 1:00 pm.

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/hannahsmith/Documents/ENV872 Data Analytics/Environmental_Data_Analytics_2020"
```

```
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: Neonicotinoids were originally thought to be toxic to only a few insects. However, neonicotinoids affect the central nervous system of honeybees and other animals that influence their ability to forage and fly without being lethal. These insects that are potentially affected by neonicotinoids are essential to stability of the environment.

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 woody debris act as habitats for various animals and nutriets to the ecosystem as they decompose. By looking at the types of woody debris, the habitats within the ecosystem can be more well understood.

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: * Litter is defined as debris that falls from the forest canopy with a butt diameter of less than 2cm and a length of less than 50cm, while fine woody debris has a butt diameter of less than 2cm and a length of greater than 50 cm. * Ground traps are only sampled once per year while elevated traps are sampled once per two weeks for deciduous forests and once per month for evergreen forests. * Trap placements depend on vegetation. If the an area has less than 50% cover of woody vegetation, trap placement is more targeted than if the forest has greater than 50% vegetation cover.

Obtain basic summaries of your data (Neonics)

5. What are the dimensions of the dataset?

```
dim(Neonics)
```

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

```
# 4,623 rows, 30 columns
```

6. Using the `summary` function, 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
```

```
#Population, Mortality, Behavior, Feeding Behavior, and Development are the most common effects that ar
```

Answer: Population is the most common effect that is studied because if the population of one type of insect changes, the entire ecosystem can be affected. Mortality follows closely behind the effect on population because it examines the death rate of an individual insect rather than the effect on entire popultion. Taken in conjuction, the mortality rate could be an indicator for why a population might be decreasing. However, the general behavior of an insect can affect feeding patterns which in turn could cause decreased development or reproductivity of the individual, thus affecting the entire population. In short, mortality, behavior, feeding behavior, and development of individual insects can greatly impact the population as a whole. Each individual effect must be studied to determine the exact cause of the change in a population. Death (mortality) is generally the first assumption for decrease of a population, but the other effects should be considered after the initial study of mortality.

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: The Honey Bee, Parasitic Wasp, Buff Tailed Bumblebee, Carniolan Honey Bee, Bumble Bee, and Italian Honeybee are the six most commonly studied insects relating to neonicotinoids.

Bees are essential pollinators for plants important to humans, such as fruit and vegetables. Without them, we would lose a crucial step in the development of our plants. The parastitic wasp use other crop-killing insects as hosts, thereby protecting agricultural crops.

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: It is a factor because there is a limited number of values.

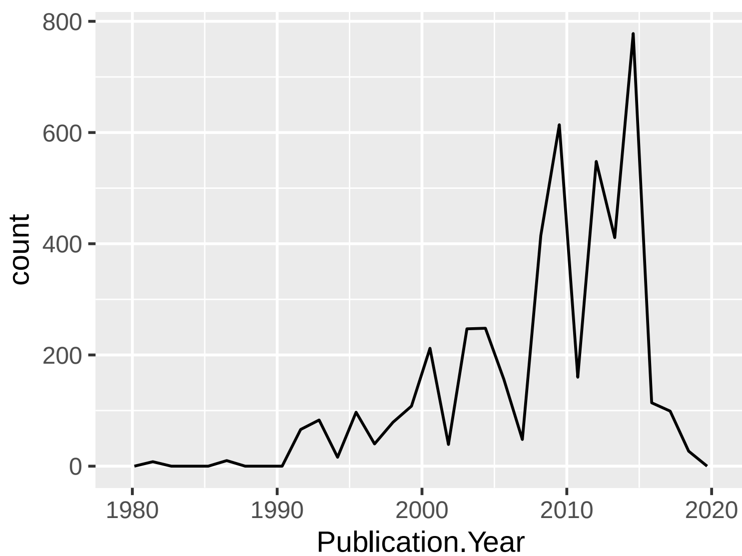
Explore your data graphically (Neonics)

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

```
publicationyearplot <- ggplot(Neonics) +
  geom_freqpoly(aes(x = Publication.Year))

print(publicationyearplot)
```

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

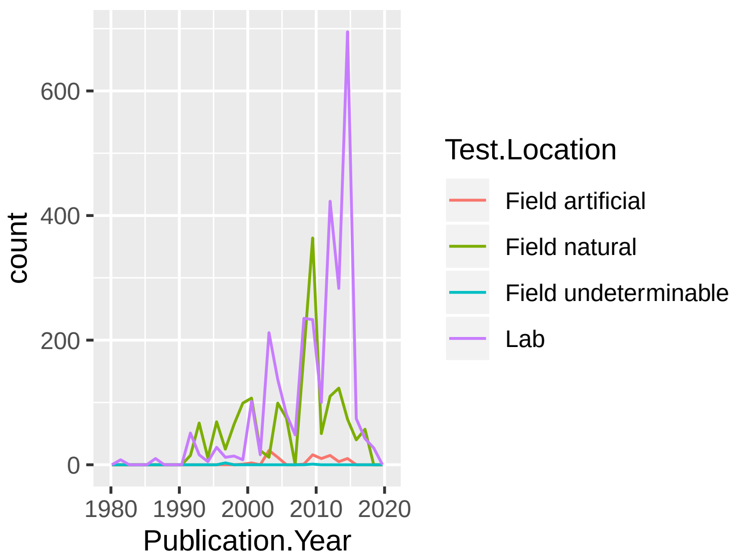


10. Reproduce the same graph but now add a color aesthetic so that different Test.Location are displayed as different colors.

```
location <- ggplot(Neonics) +
  geom_freqpoly(aes(x = Publication.Year, color = Test.Location))

print(location)
```

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



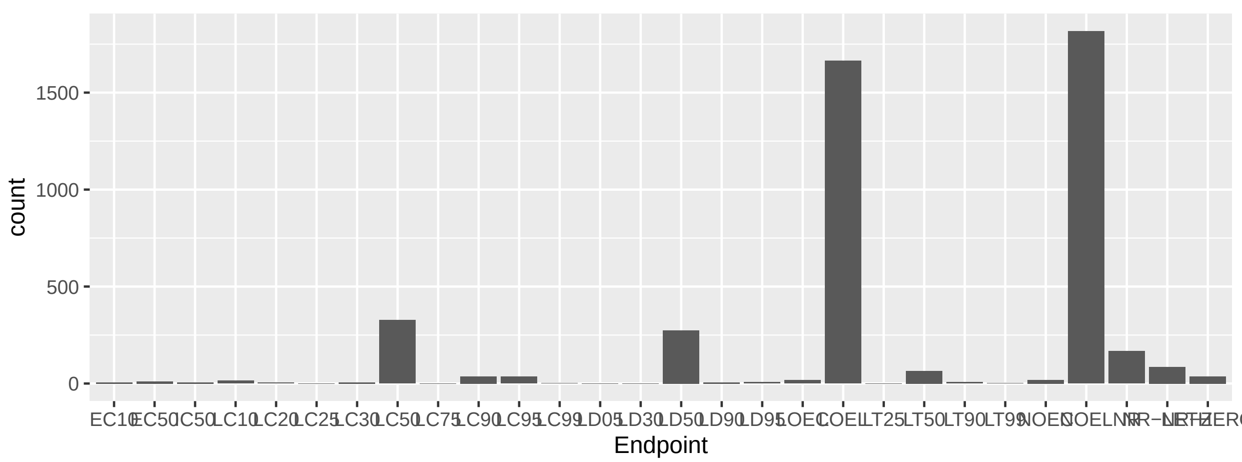
Interpret this graph. What are the most common test locations, and do they differ over time?

Answer:

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.

```
endpoint <- ggplot(Neonics) +
  geom_bar(aes(x = Endpoint))

print(endpoint)
```



Answer: The two most common end points are LOEC and NOEC, meaning lowest observable effect concentration and no observable effect concentration, respectively. LOEC means the lowest tested concentration that is significantly different from the control, and NOEC is the tested concentration right below the LOEC that is not significantly different from the control.

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.

```r
class(Litter$collectDate)
```

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

```r
#no, it is a factor
Litter$collectDate <- as.Date(Litter$collectDate, format = "%m/%d/%y")
class(Litter$collectDate)
```

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

```r
#now it's a date!
unique(Litter$collectDate)
```

```
## [1] NA
```

```r
#litter was sampled on August 2 and August 30 in 2018
```

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`?

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

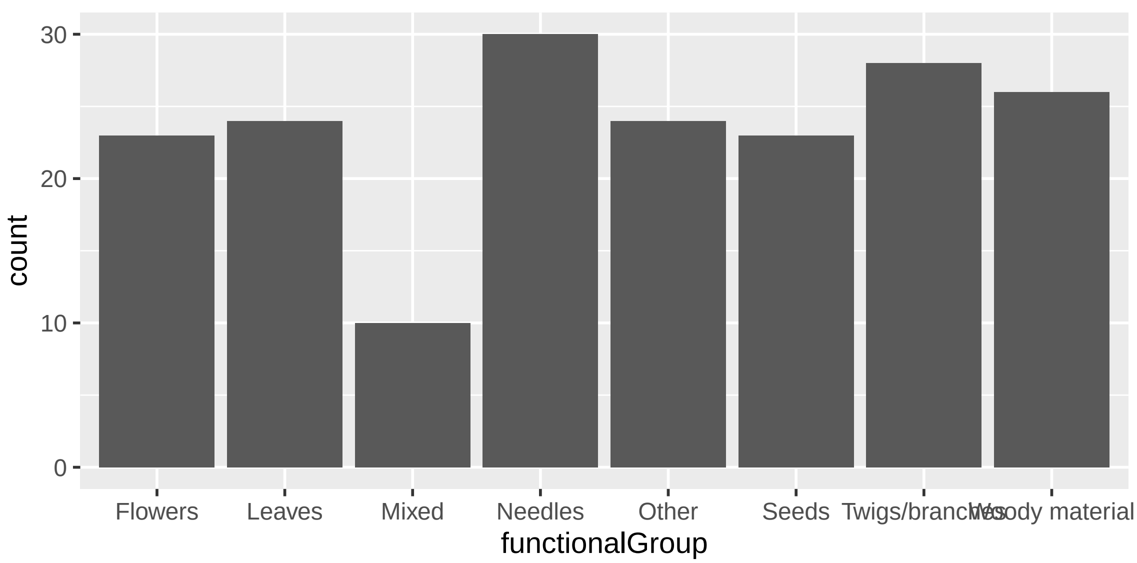
```
##  [1] NIWO_061 NIWO_064 NIWO_067 NIWO_040 NIWO_041 NIWO_063 NIWO_047 NIWO_051
##  [9] NIWO_058 NIWO_046 NIWO_062 NIWO_057
## 12 Levels: NIWO_040 NIWO_041 NIWO_046 NIWO_047 NIWO_051 NIWO_057 ... NIWO_067
```

> Answer: The unique function omits duplicate elements whereas summary does not, so you get a more precise count of how many plots were actually sampled.

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.

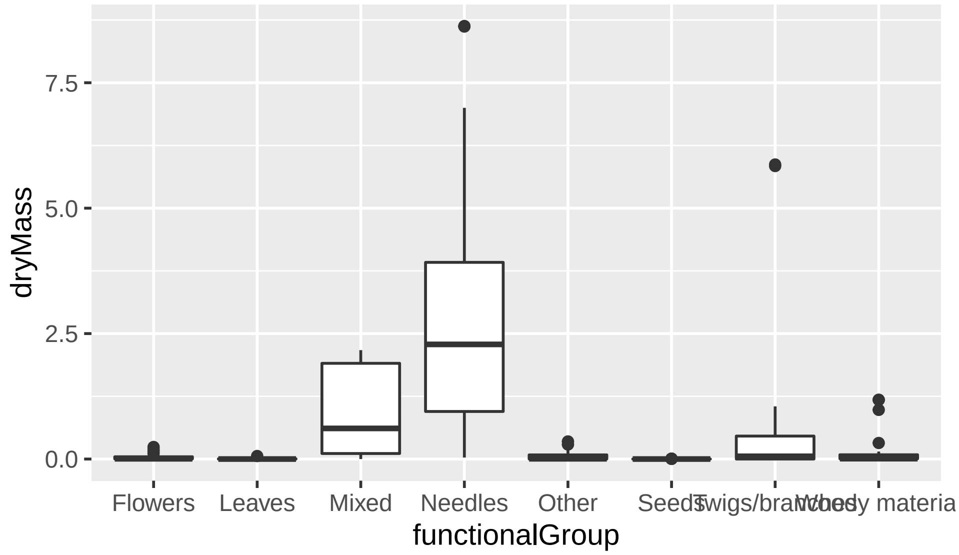
```r
littertype <- ggplot(Litter) +
  geom_bar(aes(x = functionalGroup))

print(littertype)
```

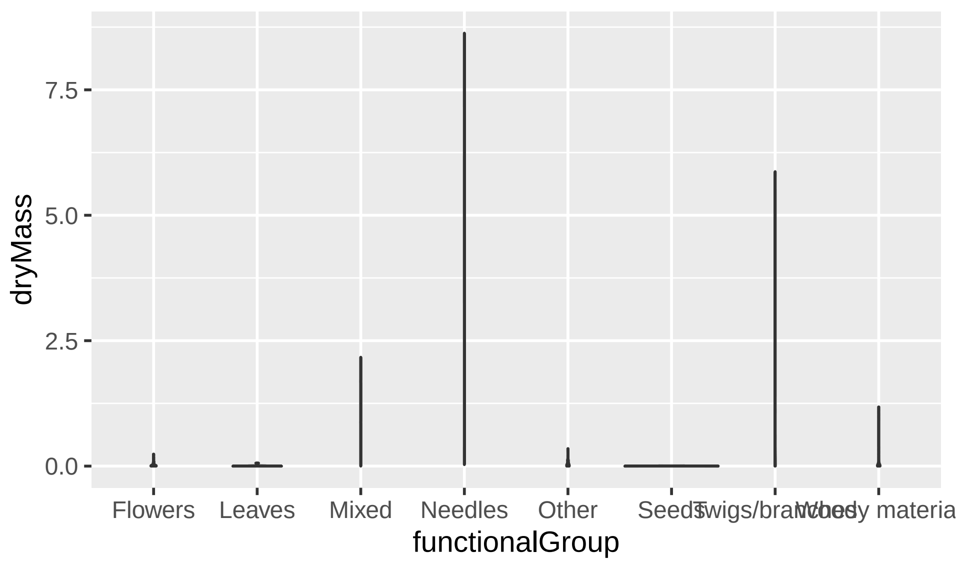


15. Using `geom_boxplot` and `geom_violin`, create a boxplot and a violin plot of dryMass by functionalGroup.

```
drymassboxplot <- ggplot(Litter) +
  geom_boxplot(aes(x = functionalGroup, y = dryMass))
print(drymassboxplot)
```




```
drymassviolin <- ggplot(Litter) +
  geom_violin(aes(x = functionalGroup, y = dryMass))
print(drymassviolin)
```



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

> Answer: The boxplot shows the distribution of the data better than the violin plot. Specifically, the boxplot shows the median, the interquartile range, and outliers of each type of functional group, whereas the violin plot is not clear on that for this dataset.

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

> Answer: The needles and mixed have the highest biomass at these stie, but an outlier of the twigs/branches is also high in biomass.