Assignment 3: Data Exploration

Jiahuan Li

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.

```r
# install.packages("tidyverse")
# install.packages("lubridate")

# check directory
getwd()
```

```
## [1] "d:/Users/Lijh/Desktop/872 R & data analytics/ENV872"
```

```r
# upload datasets
Neonics <- read.csv("Data/Raw/ECOTOX_Neonicotinoids_Insects_raw.csv", stringsAsFactors = TRUE)

Litter <- read.csv("Data/Raw/NEON_NIWO_Litter_massdata_2018-08_raw.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: That is because if we want to widely apply certain insecticides, we should be cautious about their environmental impacts. Some insecticides will possibly cause severe but unexpected harm which even overweights their benefit. And ecotoxicology is an important perspective to evaluate the potential risks of neonicotinoid insecticides. Actually, there is literature pointing out that neonicotinoid Insecticides could be a major threat to pollinating insects such as bees. They may eliminate not only pests but also pollinators through similar biochemical processes.

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: That is because litterfall in terrestrial ecosystems represents an important pathway for nutrient return to the soil. The data products from litterfall and fine woody debris sampling can provide mass data for plant functional groups. Furthermore, data on litterfall and fine woody debris can be used to calculate annual Aboveground Net Primary Productivity (ANPP) and aboveground biomass at plot, site, and continental levels. They can also provide critical information for understanding the evolution of vegetative carbon fluxes.

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. Litter sampling is planned to take place in 20 40m x 40m plots at sites with forested tower airsheds. And it is designed for 4 40m x 40m tower plots and 26 20m x 20m plots in sites with low-statured vegetation over the tower airsheds. For every 400 mˆ2 plot area, one litter trap pair (one elevated trap and one ground trap) is deployed, resulting in 1-4 trap pairs per plot.
   2. Depending on the vegetation, trap placement within plots can be either targeted or randomized. Litter trap placement is random in sites with >50% aerial cover of woody vegetation >2m in height.
   3. Ground traps are sampled once per year. The frequency of target sampling for elevated traps varies depending on the vegetation present at the site, with frequent sampling (1x every 2 weeks) in deciduous forest sites during senescence and infrequent year-round sampling (1x every 1-2 months) at evergreen sites.

Obtain basic summaries of your data (Neonics)

5. What are the dimensions of the dataset?

```
dimensions = dim(Neonics)

# Dataset dimensions: rows = 4623 and columns = 30
print(paste("The dimension of data frame is:", dimensions[1], dimensions[2]))
```

```
## [1] "The dimension of data frame 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?

   Answer: Population and mortality are the most common effects with the frequency of 1803 and 1493, respectively. That may be because they are closely related to the potential environmental harm of insecticides. Scientists may focus more on the changes of population abundance and mortality rate of certain insect groups, which could be straightforward and important indicators for the environmental impacts caused by the insecticides.

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), decreasing = TRUE)
```

```
##                            (Other)                          Honey Bee
##                                670                                667
##                     Parasitic Wasp              Buff Tailed Bumblebee
##                                285                                183
##                Carniolan Honey Bee                         Bumble Bee
##                                152                                140
##                   Italian Honeybee                    Japanese Beetle
##                                113                                 94
##                  Asian Lady Beetle                     Euonymus Scale
##                                 76                                 75
##                           Wireworm                  European Dark Bee
##                                 69                                 66
##                  Minute Pirate Bug               Asian Citrus Psyllid
##                                 62                                 60
##                      Parastic Wasp             Colorado Potato Beetle
##                                 58                                 57
##                    Parasitoid Wasp                Erythrina Gall Wasp
##                                 51                                 49
##                       Beetle Order        Snout Beetle Family, Weevil
##                                 47                                 47
##           Sevenspotted Lady Beetle                     True Bug Order
##                                 46                                 45
##              Buff-tailed Bumblebee                       Aphid Family
##                                 39                                 38
##                     Cabbage Looper               Sweetpotato Whitefly
##                                 38                                 37
##                      Braconid Wasp                       Cotton Aphid
##                                 33                                 33
##                     Predatory Mite             Ladybird Beetle Family
##                                 33                                 30
##                         Parasitoid                      Scarab Beetle
##                                 30                                 29
##                      Spring Tiphia                        Thrip Order
##                                 29                                 29
##               Ground Beetle Family                 Rove Beetle Family
##                                 27                                 27
##                      Tobacco Aphid                       Chalcid Wasp
##                                 27                                 25
##             Convergent Lady Beetle                      Stingless Bee
##                                 25                                 25
##                  Spider/Mite Class                Tobacco Flea Beetle
##                                 24                                 24
##                   Citrus Leafminer                    Ladybird Beetle
##                                 23                                 23
##                          Mason Bee                           Mosquito
##                                 22                                 22
##                      Argentine Ant                             Beetle
##                                 21                                 21
##         Flatheaded Appletree Borer               Horned Oak Gall Wasp
##                                 20                                 20
##                 Leaf Beetle Family                  Potato Leafhopper
##                                 20                                 20
##         Tooth-necked Fungus Beetle                       Codling Moth
##                                 20                                 19
##          Black-spotted Lady Beetle                       Calico Scale
##                                 18                                 18
##                Fairyfly Parasitoid                        Lady Beetle
##                                 18                                 18
##             Minute Parasitic Wasps                          Mirid Bug
##                                 18                                 18
##                   Mulberry Pyralid                           Silkworm
##                                 18                                 18
##                     Vedalia Beetle              Araneoid Spider Order
##                                 18                                 17
##                          Bee Order                     Egg Parasitoid
##                                 17                                 17
##                       Insect Class           Moth And Butterfly Order
##                                 17                                 17
##       Oystershell Scale Parasitoid Hemlock Woolly Adelgid Lady Beetle
##                                 17                                 16
##              Hemlock Wooly Adelgid                               Mite
##                                 16                                 16
##                        Onion Thrip              Western Flower Thrips
##                                 16                                 15
##                       Corn Earworm                  Green Peach Aphid
##                                 14                                 14
##                          House Fly                          Ox Beetle
##                                 14                                 14
##                 Red Scale Parasite                 Spined Soldier Bug
##                             14                            14
##           Armoured Scale Family             Diamondback Moth
##                             13                            13
##                   Eulophid Wasp            Monarch Butterfly
##                             13                            13
##                    Predatory Bug        Yellow Fever Mosquito
##                             13                            13
##              Braconid Parasitoid                 Common Thrip
##                             12                            12
##       Eastern Subterranean Termite                    Jassid
##                             12                            12
##                      Mite Order                    Pea Aphid
##                             12                            12
##                 Pond Wolf Spider    Spotless Ladybird Beetle
##                             12                            11
##            Glasshouse Potato Wasp                    Lacewing
##                             10                            10
##           Southern House Mosquito     Two Spotted Lady Beetle
##                             10                            10
##                      Ant Family                 Apple Maggot
##                              9                             9
```

Answer: The six most commonly studied species are Honey Bee, Parasitic Wasp, Buff Tailed Bumblebee, Carniolan Honey Bee, Bumble Bee, and Italian Honeybee. These species have high genetic similarity and all belong to Apidae according to biotaxonomy. They are of special interest because they are regarded as the best pollinators of crops around the world. Thus, the declining population of these important pollinators day by day is a major threat to the agriculture. And according to the literature, the use of neonicotinoid pesticide is considered to be the major reason for the decline of bees in the world.

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"
```

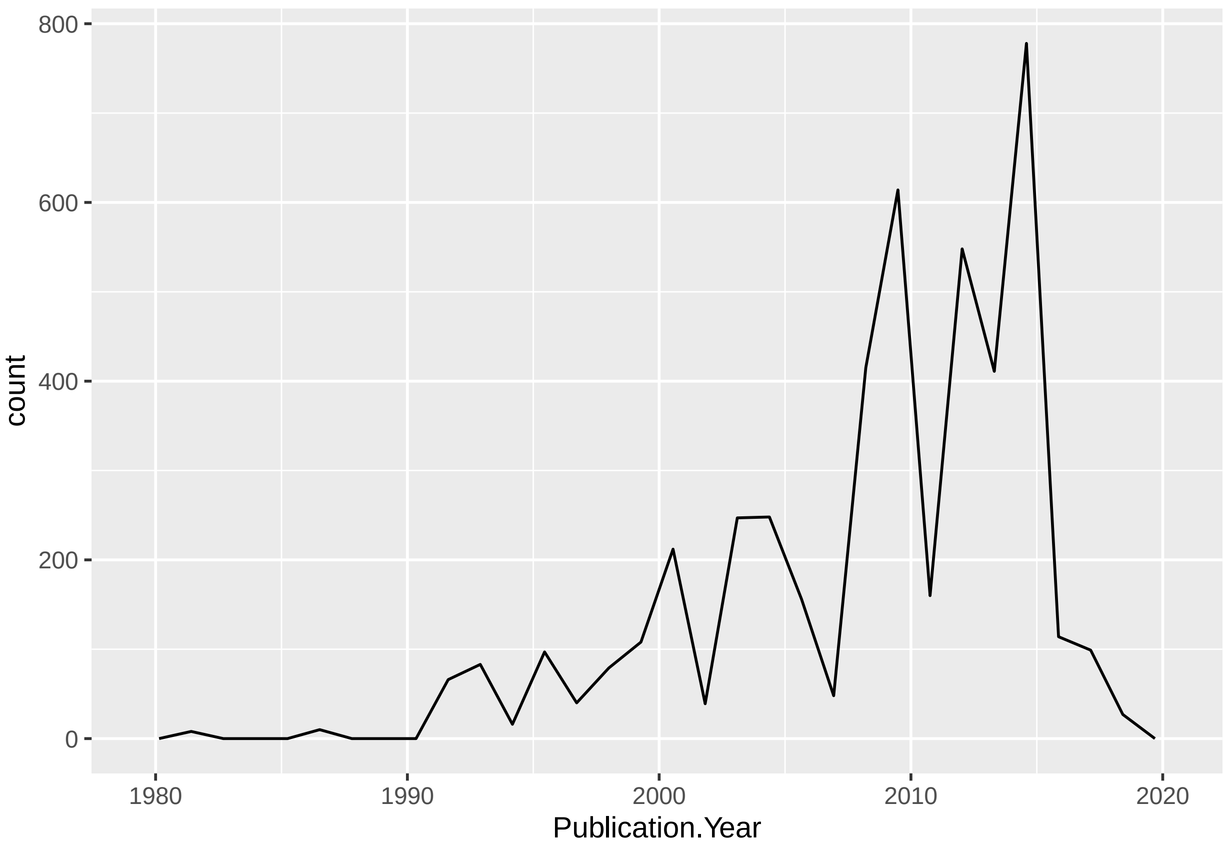
Answer: the class of this column is factor. It is not numeric because this column is not purely composed of numeric values. Some values in the column are not numbers and they are viewed as strings.

Explore your data graphically (Neonics)

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

```
library(ggplot2)
ggplot(Neonics, aes(Publication.Year)) + geom_freqpoly()
```

```
## '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.

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.

[**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…]

> Answer:

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.
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`?

> Answer:

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

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

> Answer:

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

> Answer: