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.

# install.packages("tidyverse")  
# install.packages("lubridate")  
  
# check directory  
getwd()

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

# 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

1. 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.

1. 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.

1. 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)

1. 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"

1. 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:

1. 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…]

Answer:

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

Answer:

## Explore your data graphically (Neonics)

1. Using geom\_freqpoly, generate a plot of the number of studies conducted by publication year.
2. 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:

1. 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)

1. 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.
2. 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:

1. 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.
2. 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: