

Assignment 3: Data Exploration

Julia Kagiliery

Fall 2024

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 Canvas.

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. Load necessary packages (tidyverse, lubridate, here), check your current working directory 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 sub-command to read strings in as factors.

```
library(tidyverse)
library(lubridate)
library(here)

getwd()
```

```
## [1] "/Users/juliakagiliery/Library/Mobile Documents/com~apple~CloudDocs/GitHub Links/EDAClas2025"
```

```
Neonics <- read.csv(here("Data/Raw/ECOTOX_Neonicotinoids_Insects_raw.csv"), stringsAsFactors = TRUE)
Litter <- read.csv(here("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: I might be interested in the ecotoxicology of these pesticides on insects because I want to understand the impact of using pesticides on agricultural fields for surrounding areas. The pesticides may become runoff from the farms and then start impacting non-target insects (like bees or lady bugs which provide benefits to the environment like pollination and pest control). Understanding this ecotoxicology of different pesticides might help me select a pesticide that minimizes the adverse effects for non-target species.

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 that fall to the ground in forests can become fuel for forest fires. Having an excess amount of this debris on the ground might be correlated to the severity of a possible fire and hence understanding the litter and debris might inform forest management practices (such as controlled burns) to protect against forest fire.

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 is defined as material that drops from the forest canopy with a butt end diameter less than 2 cm and a length less than 50 cm which is collected in an elevated (at 80 cm off the ground) 0.5m² PVC trap. Fine wood debris is defined material that is dropped from the forest canopy but exceeds the 2 cm diameter and the 50 cm length which is collected in ground traps which have an area of 3 m x 0.5 m. 2. The sampling is executed at sites that contain vegetation that are taller than 2 meters. 3. Ground traps are sampled once per year, frequency of elevated trap sampling depends on the vegetation present at the site which ranges from once every other week to once every one or two months.

Obtain basic summaries of your data (Neonics)

5. What are the dimensions of the dataset?

```
print(dim(Neonics))
```

```
## [1] 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? [Tip: The `sort()` command is useful for listing the values in order of magnitude...]

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: Explore the help on the `summary()` function, in particular the `maxsum` argument...]

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? [Tip: Viewing the dataframe may be helpful...]

Answer:

Explore your data graphically (Neonics)

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

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: