Assignment 3: Data Exploration

Kamil Orozco

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.

```
# importing the data sets
NeonicsData <- read.csv("./Data/Raw/ECOTOX_Neonicotinoids_Insects_raw.csv", stringsAsFactors = TRUE)
#2nd data set
LitterData <- 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: Insecticides widely used by agriculutre in general are known to show detrimental effects in humans and in important agriculutral species (pollinating species, eg: bees). Ecotoxicoligists are interested in improving insecticides as our population numbers continue to climb; as does the need for food. The reason these insecticides work is because they target a certain neurologic or developmental function of the insect which is made possible by the insect not having the enzyme(s) to break that particular insecticide down. With that said, while insects have very different body types to humans, if the active chemical in the insecticide isn't broken down by the insect, soemthing would have to eventually which can lead to a forever chemical if nothing breaks it down.

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: Increased litter and foliage/woody debris in both terrestrial and aquatic environments, decreases the amount of sunlight penetration that occurs underneath the water/soil surface. This can have many adverse effects on trophic systems that require producer organisms or vegetatin as its primary tier food source.

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. Dry weight of litter and woody debris are measured 2. Stable isotopes are measured 3. Placement of litter traps (20 m) in random catchment areas and harvesting of vegetation surrounding these areas.

Obtain basic summaries of your data (Neonics)

5. What are the dimensions of the dataset?

```
#calling the dimension of the dataset
dim(NeonicsData)
```

```
## [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?

```
#calling the summary for the Effect column
summary(NeonicsData$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
```

Answer: The three most common effects are: Population and Mortality. Population and mortality numbers being affected is the overall goal of insecticides. The thrid most common, Behavior, is interesting to note because that doesn't guarantee that the number of behaviors changed were killed but it doesnt signify a chemical imbalance within the insect that could distrupt other things like reproduction or development.

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. . .]

```
#using summary to determine 6 most common studied species
SixSpecies <- summary(NeonicsData$Species.Common.Name)
#sorting of the assigned summary
sort(SixSpecies)
```

```
##                         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
```

Answer: Six most commonly studied species are: Honey Bees (667), Parasitic wasps (285), Buff tailed bumble bee (183), Carniolan Honey Bee (152), Bumble Bee (140), Italian honey bee (113). A commonality of all six species is that they are all pollinators. I think honey bees are at the top of the list because they are not only important for ecosystem purposes but they also carry a large economic/societal purpose for human consumption of honey. I couldn't wrap my head around why parasitic wasps would be useful but with a quick google search I found that these wasps prey on crop pests and implant their eggs into these pest species so that more wasps contribute to killing them. They are a very important garden/farm maintenance species.

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?

```
#calling of the class type for Concentration 1 Author from neonics data
class(NeonicsData$Conc.1..Author.)
```

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

Answer: The concentration column is a factor class; factors are stored as integers even though they look and behave as characters. Because factors can only contain pre-defined set values (levels), then by default, R always sorts levels in alphabetical order.

Explore your data graphically (Neonics)

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

```
#creating a frequency graph
ggplot(NeonicsData) +
  geom_freqpoly(aes(x = Publication.Year), bins = 10) + #inserting the outer boundaries of the years pr
  scale_x_continuous(limits = c(1980, 2023))
```

```
## Error in ggplot(NeonicsData): could not find function "ggplot"
```

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

```
#reproducing same graph and adding more aesthetic elements
ggplot(NeonicsData.complete) +
  geom_freqpoly(aes(x = Publication.Year, color = Test.Location), bins = 10) +
  scale_x_continuous(limits = c(1980, 2023)) +
  theme(legend.position = "top") #since I will have multiple locations, I need a key
```

```
## Error in ggplot(NeonicsData.complete): could not find function "ggplot"
```

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

> Answer: The most common test location is within a laboratory. Second most common is out in the field of a natural environment. I honestly would have thought that the second most common would have been in "Field Artificial" since that is in between being in the lab and in the natural environment. There was a steep decline in both testing locations prior to the Covid-19 pandemic which is interesting.

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...]

```
#creating a bar graph of Endpoints so that counts will show on y axis
ggplot(NeonicsData.complete, aes(x = Endpoint)) +
  geom_bar() +
  theme(axis.text.x = element_text(angle = 90, vjust = 0.5, hjust =1))
```

```
## Error in ggplot(NeonicsData.complete, aes(x = Endpoint)): could not find function "ggplot"
```

> Answer: The two most common endpoints are LOEL and NOEL; both describing concentration measurment, lowest observed effect level and no observed effect level. LOEL: The lowest dose where there is adverse effects in the exposed organisms compared to the controls. NOEL: The highest dose that does not cause any adverse effect.

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.

```
#clearing out any NA values from raw data
LitterData.complete <- na.omit(LitterData)
#determining class
class(LitterData$collectDate)
```

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

```
#setting factor class as date class
LitterData.complete$collectDate <- as.Date(LitterData.complete$collectDate, format= "%m/%d/%y")
#checking for date class confirmation
class(LitterData.complete$collectDate)
```

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

```
#using unique function to determine how many August 2018 litter collection dates there were
unique(LitterData$collectDate,"08","2018"=TRUE)
```

```
## [1] 2018-08-02 2018-08-30
## Levels: 2018-08-02 2018-08-30
```

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`?

```
#using the unique function to count how many sampled plots were at Niwot Ridge
unique(LitterData.complete$namedLocation, "Niwot Ridge"=TRUE)
```

```
## factor(0)
## 12 Levels: NIWO_040.basePlot.ltr ... NIWO_067.basePlot.ltr
```

Answer: The summary function reduces a data frame to a one vector/value summary of its values. Whereas the unique function identifies, eliminates/deletes duplicate values or rows in a vector, data frame, or matrix.

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.

```
#creating a bar graph of functional groups in litter data
ggplot(LitterData, aes(x = functionalGroup)) +
  geom_bar()
```

```
## Error in ggplot(LitterData, aes(x = functionalGroup)): could not find function "ggplot"
```

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

```
# boxplot of dry mass and functional group
ggplot(LitterData) + #using "complete" data was not useful this go around
  geom_boxplot(aes(x = dryMass, y = functionalGroup))
```

```
## Error in ggplot(LitterData): could not find function "ggplot"
```

```
# violin plot of dry mass and functional group
ggplot(LitterData) +
  geom_violin(aes(x = dryMass, y = functionalGroup),
              draw_quantiles = c(0.25, 0.50, 0.75))
```

```
## Error in ggplot(LitterData): could not find function "ggplot"
```

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

> Answer: Since the data is not multimodal, there isn't really a use for a violon plot that is good at showing distribution over time. In other words, there aren't multiple peaks, therefore a boxplot does a great job at displaying this particular dataset.

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

> Answer: Needles, twigs/branches, and mixed.