

Assignment 3: Data Exploration

Sammy DiLoreto

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.

```
#Check working directory
getwd() #NOT WORKING NOW?????
```

```
## [1] "/Users/sammydiloreto/Library/CloudStorage/Box-Box/ENV872-EDA/EDA-Spring2023/Assignments"
```

```
#Load necessary packages
library(tidyverse)
library(lubridate)
#Upload two data sets
Neonics <- read.csv("../Data/Raw/ECOTOX_Neonicotinoids_Insects_raw.csv", stringsAsFactors = T)
Litter <- read.csv("../Data/Raw/NEON_NIWO_Litter_massdata_2018-08_raw.csv", stringsAsFactors = T)
```

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 are used to get rid of insects that pose harm to crops, although some insecticides impact other insect and organism that are not targeted, called off-target effects. Understanding the ecotoxicology of neonicotinoids is important to understand how these compounds target insects and cause toxicity and which insects may be more impacted. This knowledge informs regulation in terms of which plants and insects neonicotinoids can be applied to in order to limit off-target effects. It is known that neonicotinoids affect bees which are crucial to the growth of many plants so its important to understand how the affected and how use of these pesticides can be limited to reduce impacts to bees.

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:

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. 2. 3.

Obtain basic summaries of your data (Neonics)

5. What are the dimensions of the dataset?

```
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?

```
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	

Answer: The most common effects that are studied are mortality and population. These effects are of interest because...

- 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...]

```
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 Woolly 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

```
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: Honey Bee, Parasitic Wasp, Buff Tailed Bumblebee, Carniolan Honey Bee, Bumble Bee, and Italian Honeybee are the six most commonly studied species in this dataset, aside from “other.” These are the most commonly studies because neonicotinoids are toxic to bees. Pollen can contain nenonics which is how bees are in direct contact with these pesticides. more???

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.) #ALL COMING UP AS FACTOR???
```

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

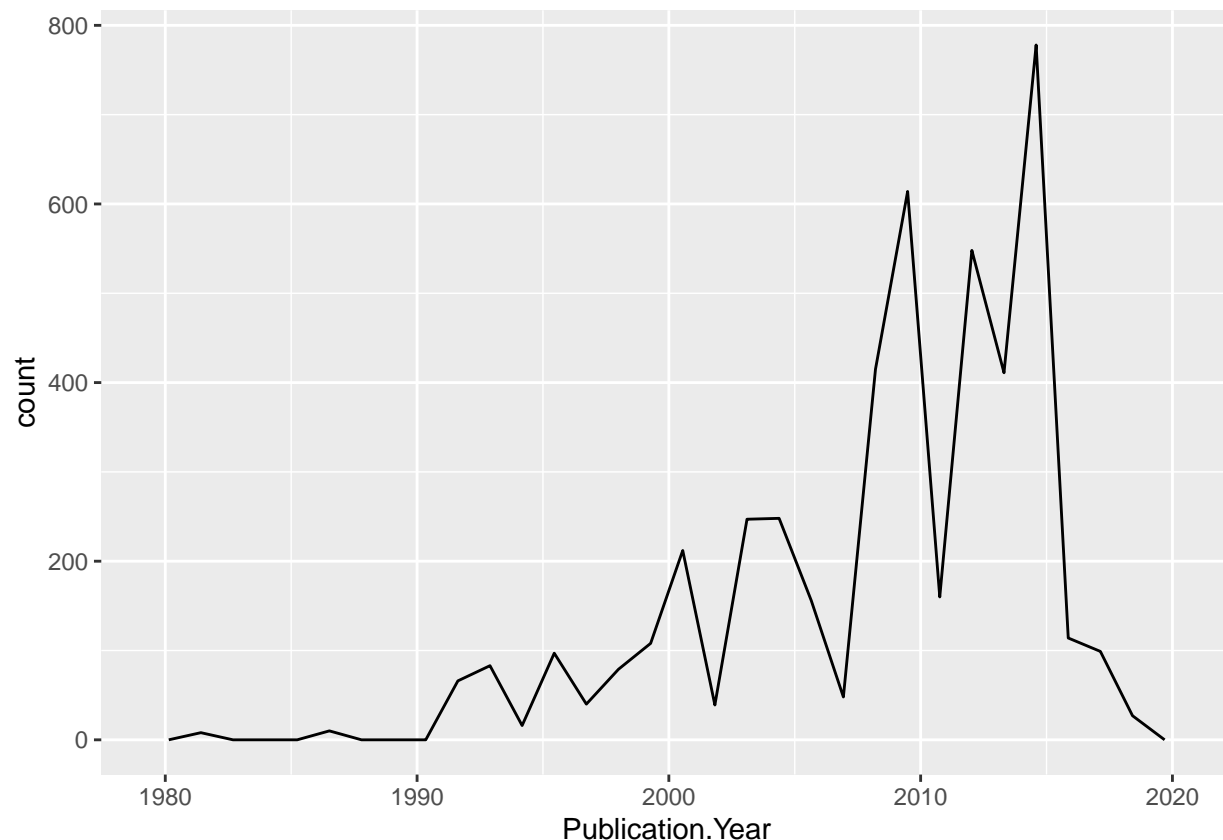
Answer:

Explore your data graphically (Neonics)

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

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

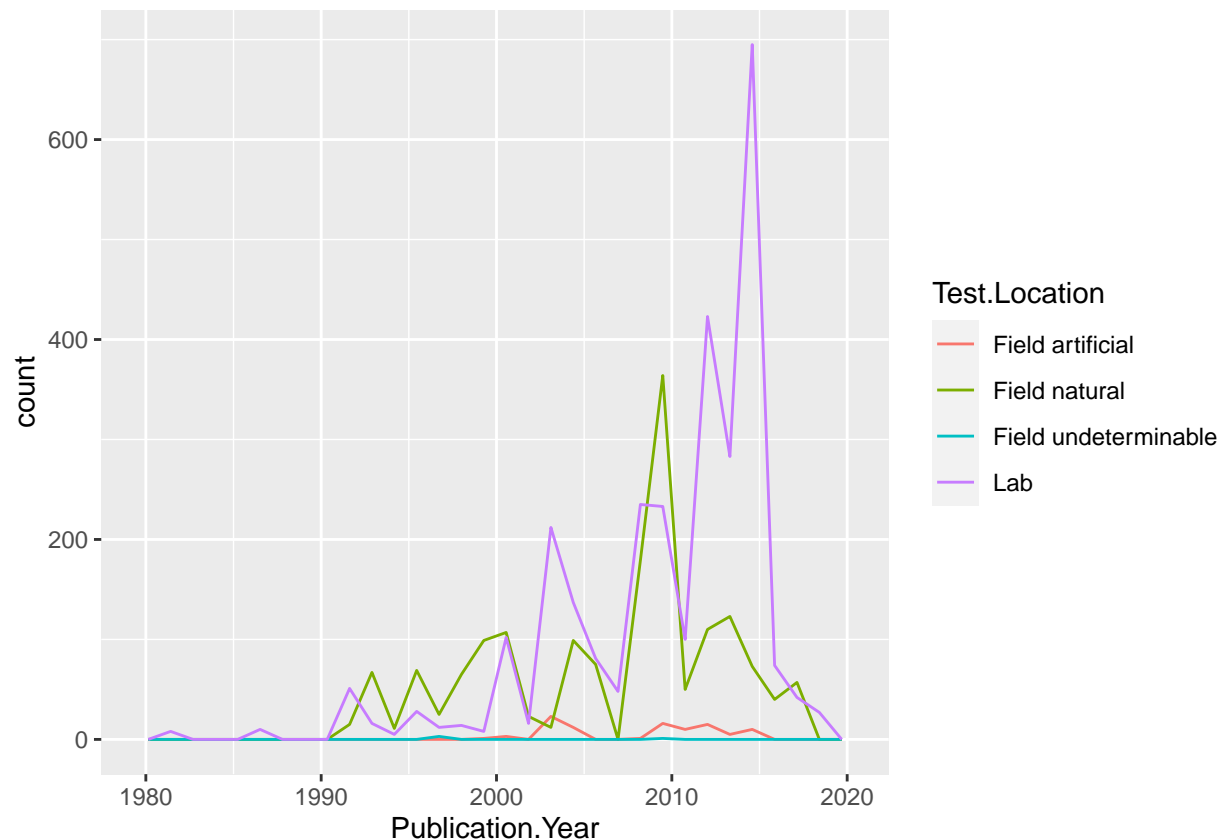
```
## 'stat_bin()' using 'bins = 30'. Pick better value with 'binwidth'.
```



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

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

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



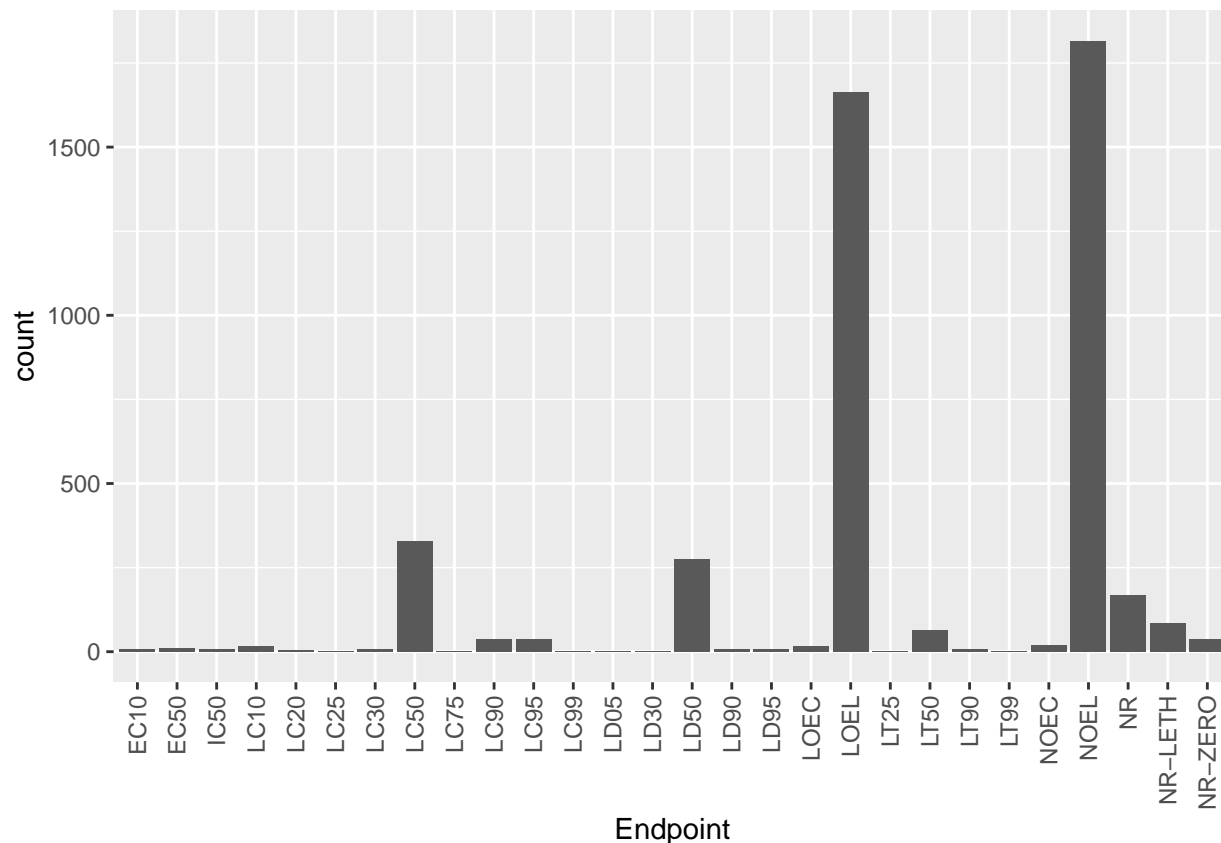
Interpret this graph. What are the most common test locations, and do they differ over time?

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.

[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...]

```
ggplot(Neonics, aes(x = Endpoint)) +  
  geom_bar()+  
  theme(axis.text.x = element_text(angle = 90, vjust = 0.5, hjust=1))
```

Answer: LOEL and NOEL. LOEL is the lowest observed effect level at which there is a statistically significant result. NOEL is the highest level at which there is no observed effect level.

Explore your data (Litter)

- 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.

```
#Determine the class of collectDate
class(Litter$collectDate)
```

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

```
#Change collectDate to date format
Litter$collectDate <- ymd(Litter$collectDate)
#Determine which dates litter was sampled in August 2018
#unique(Litter$collectDate) ????
```

- 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 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: