Assignment 3: Data Exploration

Logan Loadholtz

OVERVIEW

This exercise accompanies the lessons in Environmental Data Analytics on Data Exploration.

Directions

1. Change "Student Name" on line 3 (above) with your name.
2. Work through the steps, **creating code and output** that fulfill each instruction.
3. Be sure to **answer the questions** in this assignment document.
4. When you have completed the assignment, **Knit** the text and code into a single PDF file.
5. After Knitting, submit the completed exercise (PDF file) to the dropbox in Sakai. Add your last name into the file name (e.g., "Salk_A03_DataExploration.Rmd") prior to submission.

The completed exercise is due on <>.

Set up your R session

1. Check your working directory, load necessary packages (tidyverse), 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.

```
getwd()
```

```
## [1] "/Users/loganloadholtz/Documents/DATA/Environmental_Data_Analytics_2021/Assignments"
```

```
#Here I am checking the working directory by using getwd()
#install.packages("tidyverse")
library(tidyverse)

Neonics <- read.csv("~/Documents/DATA/Environmental_Data_Analytics_2021/Data/Raw/ECOTOX_Neonicotinoids_



#Neonics <- ECOTOX_Neonicotinoids_Insects_raw

Litter <- read.csv("~/Documents/DATA/Environmental_Data_Analytics_2021/Data/Raw/NEON_NIWO_Litter_massda

#Litter <- NEON_NIWO_Litter_massdata_2018.08_raw
```

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 ecotoxicologoy of neonicotinoids on insects? Feel free to do a brief internet search if you feel you need more background information.

Answer: Since neonicotinoids are class of insecticides, there is chance that they also impact specices that are not necessarily pests. For example, insects like bees are valuable for plants, even more so in agriculture, because they serve as pollinators. It is important to study the effects of neonicotinoids on many different insects to see if they are causing negative health impacts on insects that may be beneficial.

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: It is useful to study forest litter and woody debris because it helps recycle nutrients in the ecosystem and provides food and habitat for organisms.

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: Litter and fine woody debris are collected from elevated and ground traps, respectively. *Litter is defined as material with a butt end diameter <2cm and length <50 cm. Fine wood debris is material with butt end diameter <2cm and length >50cm.* evergreen sites are sampled 1x every 1-2 months, and deciduous forest sites are sampled frequently 1x every 2 weeks *Litter and fine woody debris data can be used to calculate aboveground net primary productivity and aboveground biomass at the plot site. Also, this data can provide information about vegetation carbon fluxes over time.

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 "Effects" column, determine the most common effects that are studied. Why might these effects specifically be of interest?

```
summary.factor(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 Effect is Population, with 1803 results, then Mortality with 1493 results.

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.

```
summary(as.factor(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 Wooly 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
```

```
# Here, I was able to sort the species from highest to lowest. The 6 most commonly studied species are
```

Answer: The six must studied species are honey bee (667), parasitic wasp (285), buff tailed bumblebee (183), carniolan honey bee (152), bumble bee (140), and italian honeybee (113)

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

```
class(Neonics$Conc.1..Author.)
```

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

Answer: This is character and not numeric because when it was imported, R read it as character instead of numeric.

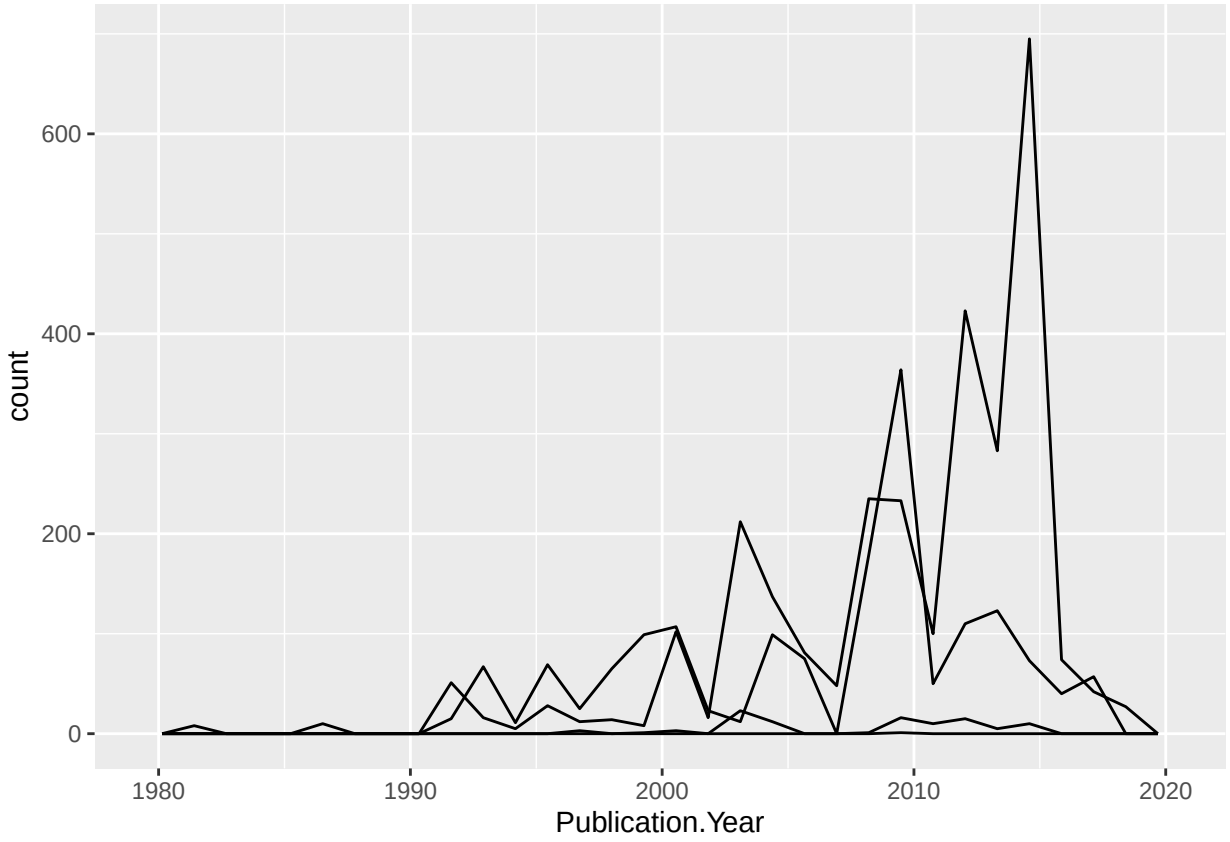
Explore your data graphically (Neonics)

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

```
ggplot(Neonics, aes(Publication.Year)) +geom_freqpoly(aes(group=Test.Location), bin=10)
```

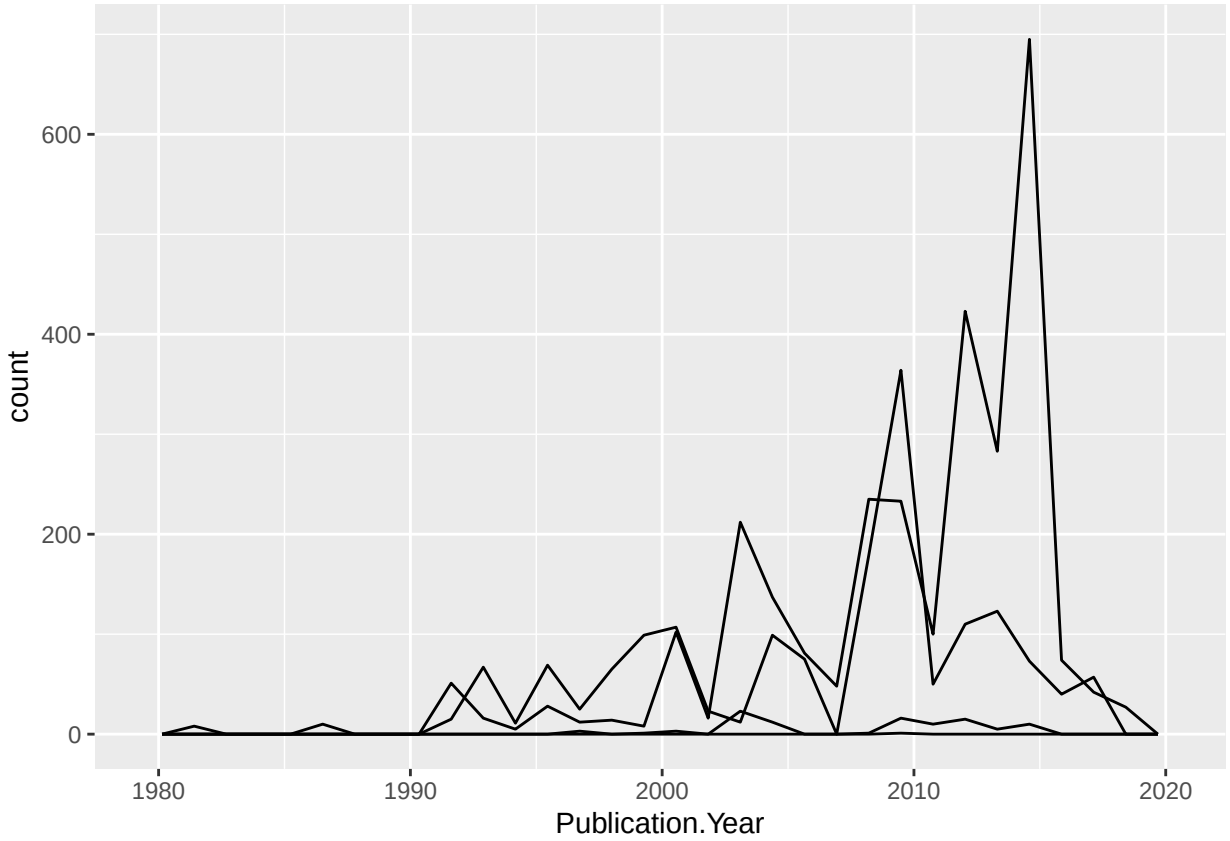
```
## Warning: Ignoring unknown parameters: bin
```

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




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

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

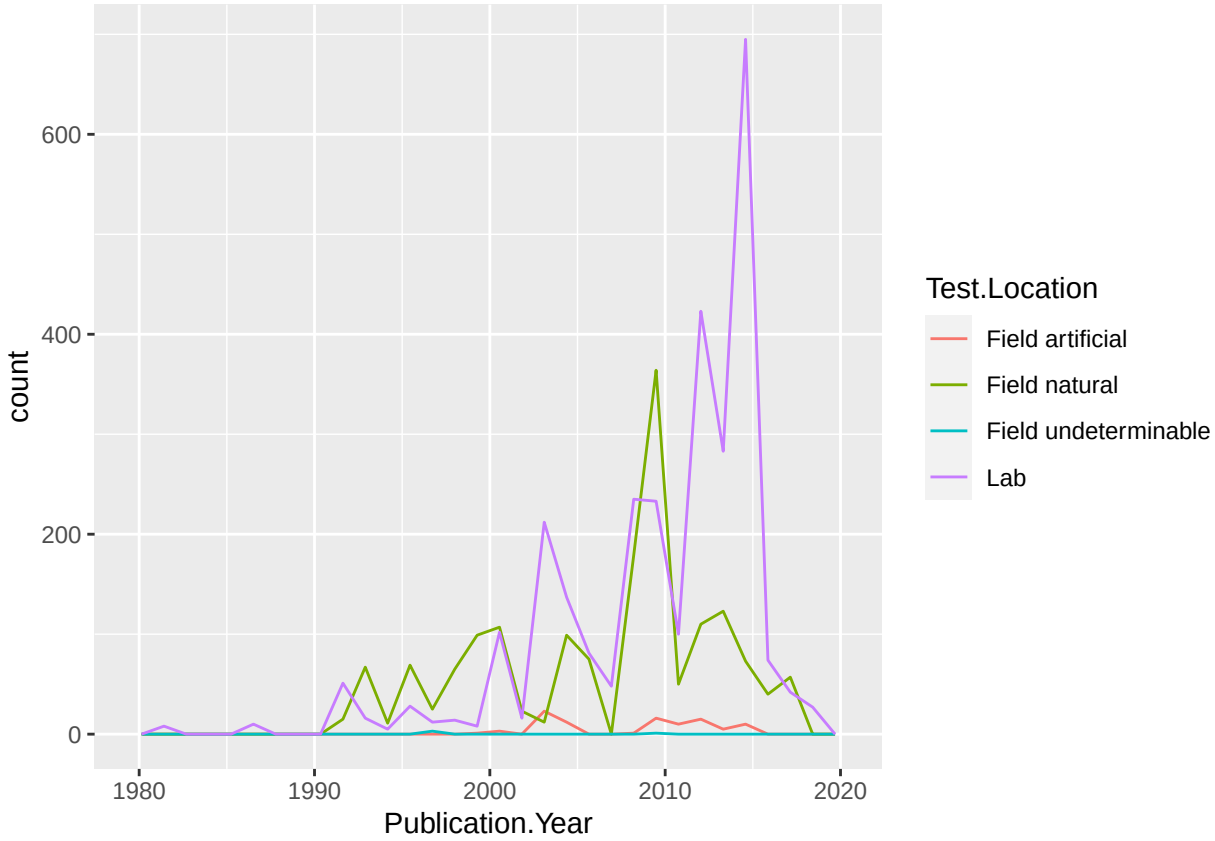


10. Reproduce the same graph but now add a color aesthetic so that different Test.Location are displayed as different colors.

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

```
## Warning: Ignoring unknown aesthetics: bin
```

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

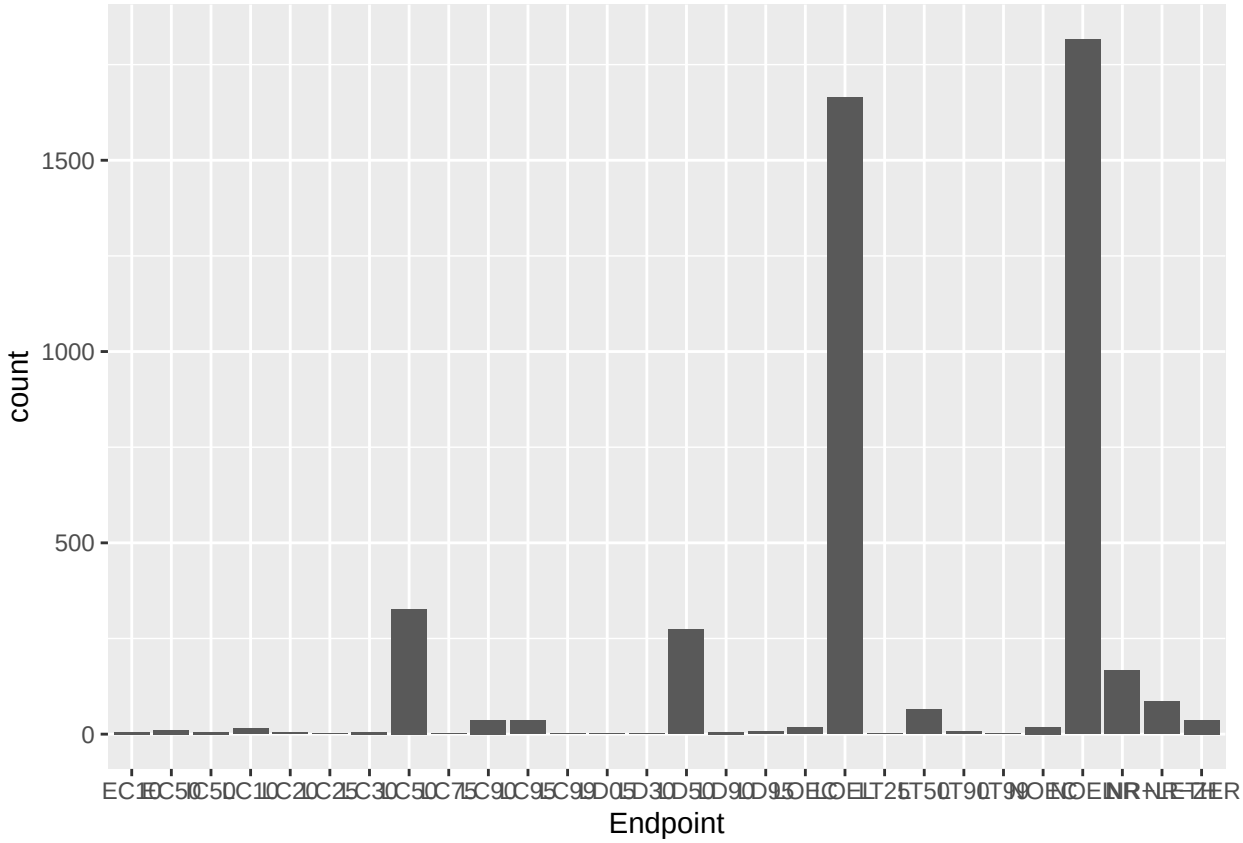


Interpret this graph. What are the most common test locations, and do they differ over time?

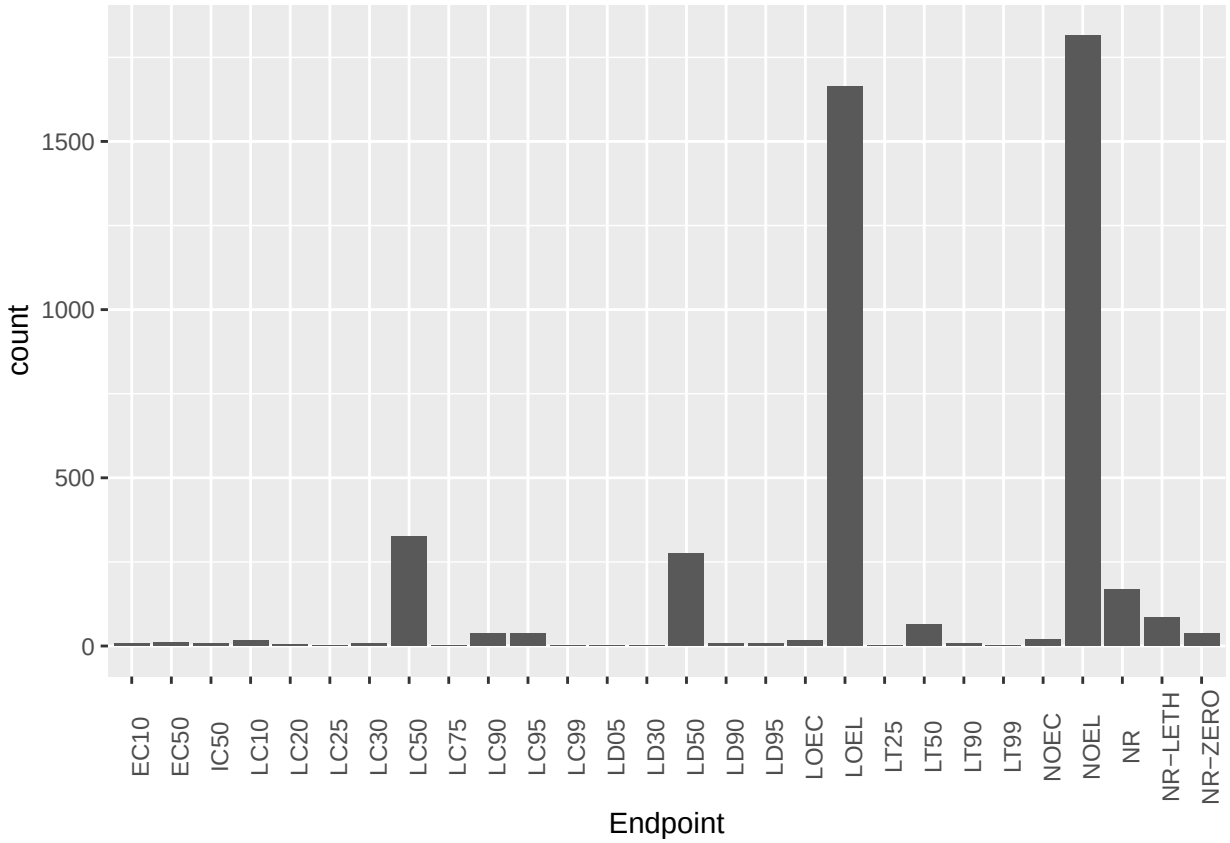
> Answer: The most common test location is in the lab. There are much more lab tests done as time goes on. In the 1980s, when the data begins, there are few to no lab tests. However, around 2015, there are almost 700 lab tests completed as shown by the purple on the graph.

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.

```
Endpoint_Count <- ggplot(Neonics) + geom_bar(aes(x=Endpoint))
Endpoint_Count
```




```
Endpoint_Count + theme(axis.text.x = element_text(angle=90))
```



Answer: The two most common endpoints are NOEL and LOEL. NOEL is the most common and it is defined as Terrestrial, No obserable effect level (NOEL). LOEL is also teresstrial and is lowest observable effect level

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.

```
class(Litter$collectDate)
```

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

```
#The class of collectDate is character. Therefore, we will need to change this to date format.

Litter$collectDate <- as.Date(Litter$collectDate, format= "%Y-%m-%d")
class(Litter$collectDate)
```

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

```
#Here, I used the as.Date function to change the date in character form to the date in Date form so tha
```

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`?

```
unique(Litter$plotID)
```

```
##  [1] NIWO_061 NIWO_064 NIWO_067 NIWO_040 NIWO_041 NIWO_063 NIWO_047 NIWO_051
##  [9] NIWO_058 NIWO_046 NIWO_062 NIWO_057
## 12 Levels: NIWO_040 NIWO_041 NIWO_046 NIWO_047 NIWO_051 NIWO_057 ... NIWO_067
```

```
summary(Litter)
```

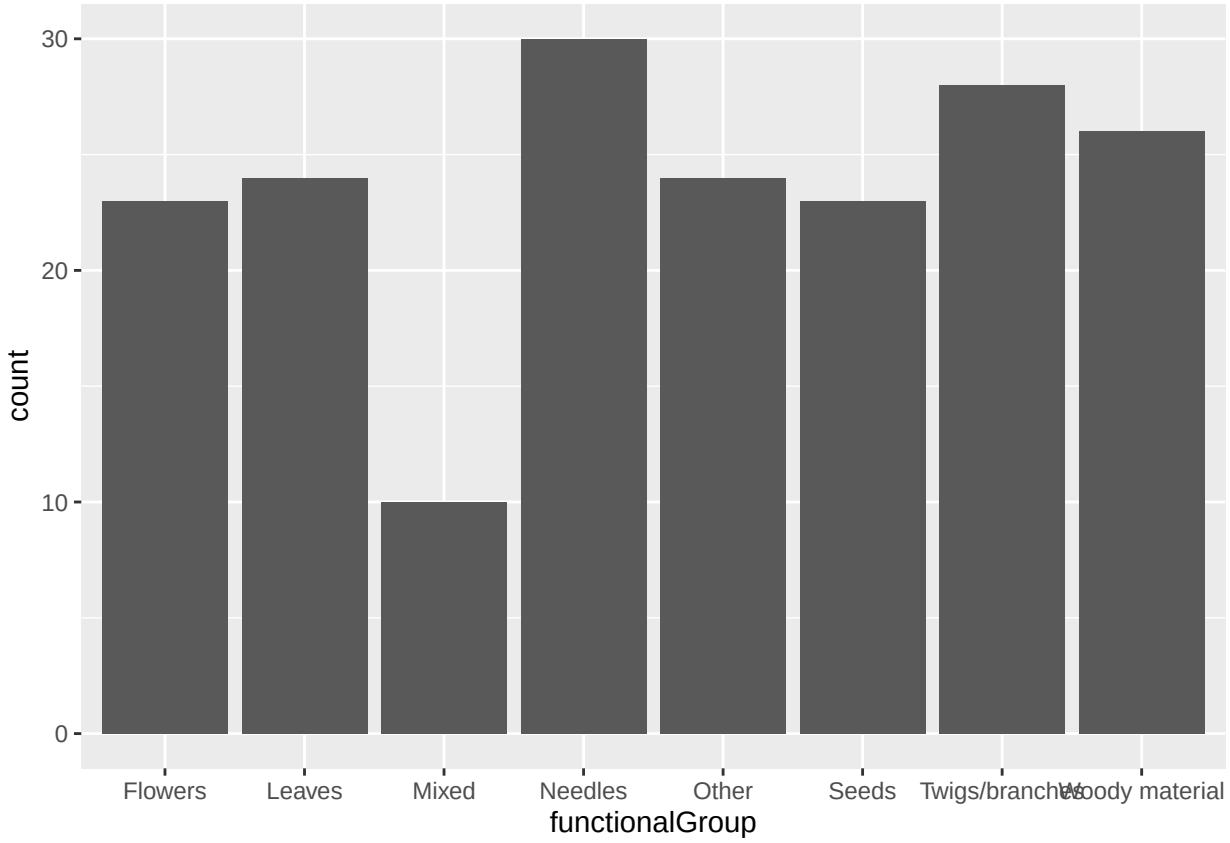
```
##                                     uid                     namedLocation
##  028eea3d-5c20-4afc-bb7e-a05bab305152:  1   NIWO_040.basePlot.ltr:20
##  06789d7b-b742-41d9-8556-79d23c193dc0:  1   NIWO_041.basePlot.ltr:19
##  07780a1e-8af9-4b8a-bb9b-be8add15a1e0:  1   NIWO_046.basePlot.ltr:18
##  0a6cae78-ea42-4e68-98c6-9d929068a38a:  1   NIWO_061.basePlot.ltr:17
##  0ae1c621-387e-42a9-bcf3-7ad1c9b97ab4:  1   NIWO_067.basePlot.ltr:17
##  0b274782-8e52-4f6a-bb17-36daa821f929:  1   NIWO_058.basePlot.ltr:16
##  (Other)                               :182   (Other)              :81
##  domainID   siteID         plotID            trapID        weighDate
##  D13:188   NIWO:188   NIWO_040:20   NIWO_040_205:20   2018-08-06:91
##                       NIWO_041:19   NIWO_041_059:19   2018-09-05:97
##                       NIWO_046:18   NIWO_046_155:18
##                       NIWO_061:17   NIWO_061_169:17
##                       NIWO_067:17   NIWO_067_017:17
##                       NIWO_058:16   NIWO_058_101:16
##                       (Other) :81   (Other)     :81
##        setDate    collectDate                  ovenStartDate
##  2018-07-05:91   Min.   :2018-08-02   2018-08-02T21:00Z:91
##  2018-08-02:97   1st Qu.:2018-08-02   2018-08-30T22:30Z:97
##                  Median :2018-08-30
##                  Mean   :2018-08-16
##                  3rd Qu.:2018-08-30
##                  Max.   :2018-08-30
##
##             ovenEndDate                       fieldSampleID
##  2018-08-06T18:02Z:91   NEON.LTR.NIWO041059.20180830: 11
##  2018-09-05T19:30Z:97   NEON.LTR.NIWO040205.20180802: 10
##                         NEON.LTR.NIWO040205.20180830: 10
##                         NEON.LTR.NIWO046155.20180802: 10
##                         NEON.LTR.NIWO058101.20180802:  9
##                         NEON.LTR.NIWO061169.20180802:  9
##                         (Other)                     :129
##                            massSampleID      samplingProtocolVersion
##  NEON.LTR.NIWO040205.20180802.MXT:  2   NEON.DOC.001710vE:188
##  NEON.LTR.NIWO040205.20180802.NDL:  2
##  NEON.LTR.NIWO040205.20180830.MXT:  2
##  NEON.LTR.NIWO040205.20180830.NDL:  2
##  NEON.LTR.NIWO041059.20180830.MXT:  2
##  NEON.LTR.NIWO041059.20180830.NDL:  2
##  (Other)                          :176
##        functionalGroup    dryMass       qaDryMass remarks
##  Needles       :30     Min.   :0.0000   N:168     Mode:logical
##  Twigs/branches:28     1st Qu.:0.0000   Y: 20     NA's:188
##  Woody material:26     Median :0.0050
##  Leaves        :24     Mean   :0.6115
##  Other         :24     3rd Qu.:0.3200
##  Flowers       :23     Max.   :8.6300
##  (Other)       :33
##                        measuredBy
##  kstyers@battelleecology.org:91
##  szrillo@battelleecology.org:97
##
```

```
##
##
##
##
```

Answer: There are 12 plots sampled at Niwot Ridge. Unique tells us how many unique plots were sampled. The summary function just gives us the number of total samples taken.

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.

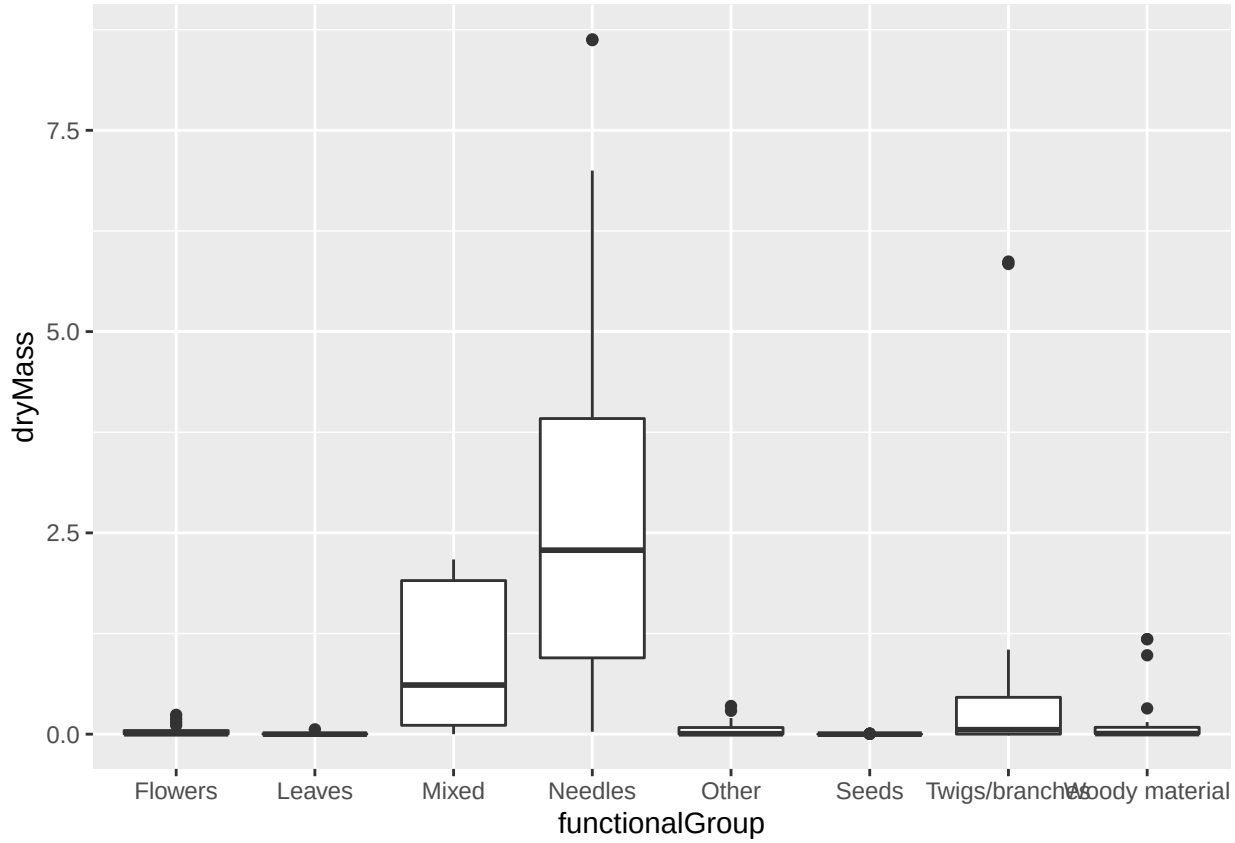
```
ggplot(Litter, aes(x=functionalGroup)) +geom_bar()
```




```
#counts of records in each functionalGroup category
```

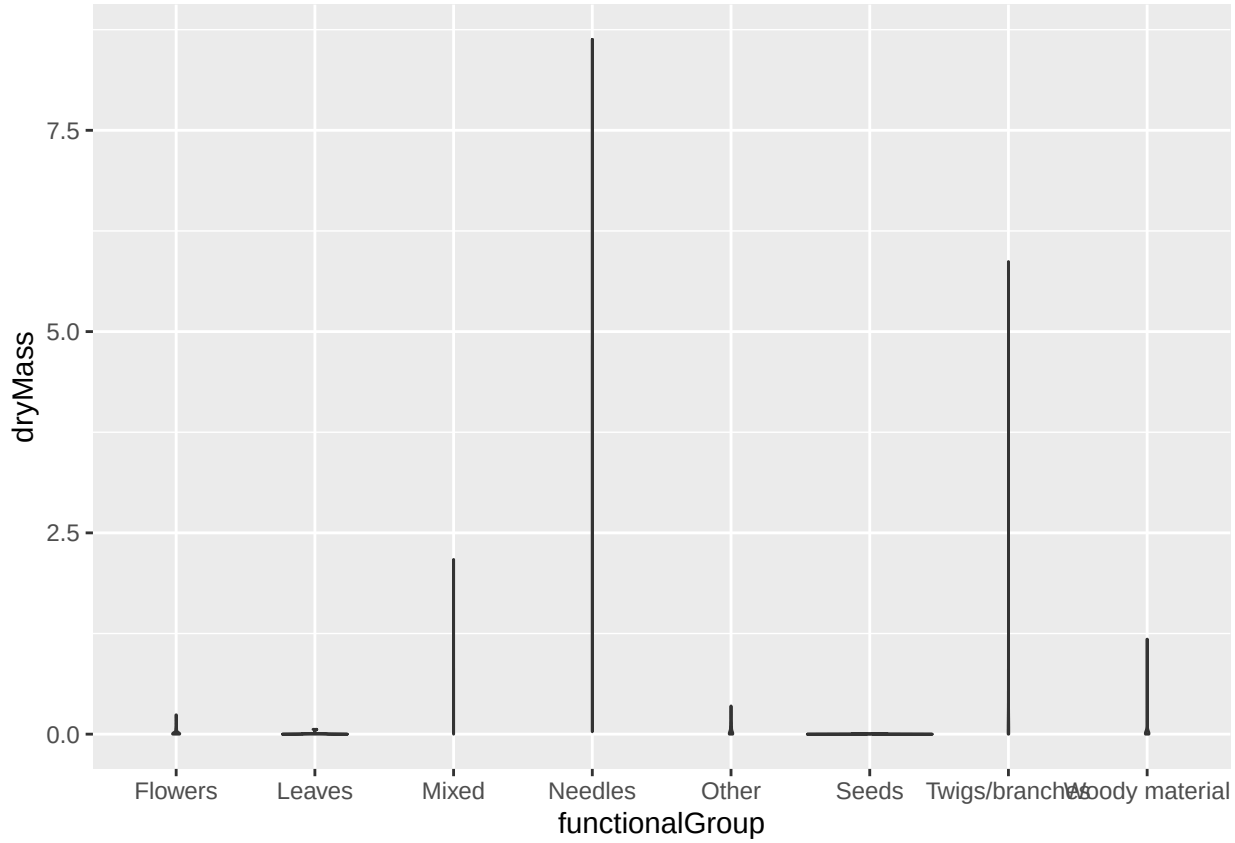
15. Using `geom_boxplot` and `geom_violin`, create a boxplot and a violin plot of dryMass by functionalGroup.

```
ggplot(Litter) + geom_boxplot(aes(x=functionalGroup, y=dryMass))
```




```
#Here I created a boxplot of the dryMass by functional group
```

```
ggplot(Litter) +geom_violin(aes(x=functionalGroup, y=dryMass))
```




```
#Here I created a violin plot of the dryMass by functionalGroup
```

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

> Answer: The boxplot is more effective at visualization than the violin for this example because we are able to see the distribution more clearly. We are able to see the summary statistics that boxplots show such as the median, and the quartiles, as well as any outliers. The violin plot for this case does not show the distribution of the values.

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

> Answer: Just by looking at the boxplot, needles have the highest biomass values at these sites. Needles has the highest median value out of all the litter types.