

Assignment 5: Data Visualization

Maggie O'Shea

OVERVIEW

This exercise accompanies the lessons in Environmental Data Analytics on Data Visualization

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., “Fay_A05_DataVisualization.Rmd”) prior to submission.

The completed exercise is due on Monday, February 14 at 7:00 pm.

Set up your session

1. Set up your session. Verify your working directory and load the tidyverse and cowplot packages. Upload the NTL-LTER processed data files for nutrients and chemistry/physics for Peter and Paul Lakes (use the tidy [NTL-LTER_Lake_Chemistry_Nutrients_PeterPaul_Processed.csv] version) and the processed data file for the Niwot Ridge litter dataset (use the [NEON_NIW0_Litter_mass_trap_Processed.csv] version).
2. Make sure R is reading dates as date format; if not change the format to date.

```
#1  
getwd()
```

```
## [1] "/Users/maggieoshea/Desktop/Spring 2022/Data Analytics/Environmental_Data_Analytics_2022/Assignm
```

```
#Installed ggplot separately because of problem with knitting
```

```
library(tidyr)  
library(cowplot)  
library(viridis)  
library(RColorBrewer)  
library(colormap)  
library(ggplot2)
```

```
peterpaulchemistry<-read.csv("../Data/Processed/NTL-LTER_Lake_Chemistry_Nutrients_PeterPaul_Processed.c  
stringsAsFactors = TRUE)
```

```
litter.masstrap<- read.csv("../Data/Processed/NEON_NIWO_Litter_mass_trap_Processed.csv",
  stringsAsFactors = TRUE)

#2
class(litter.masstrap$collectDate)

## [1] "factor"

class(peterpaulchemistry$sampledDate)

## [1] "factor"

peterpaulchemistry$sampledDate <- as.Date(peterpaulchemistry$sampledDate, format = "%Y-%m-%d")
litter.masstrap$collectDate <- as.Date(litter.masstrap$collectDate, format = "%Y-%m-%d")
```

Define your theme

3. Build a theme and set it as your default theme.

```
#3
A5theme <- theme_gray(base_size = 14) +
  theme(axis.text = element_text(color = "black"),
        legend.position = "right")

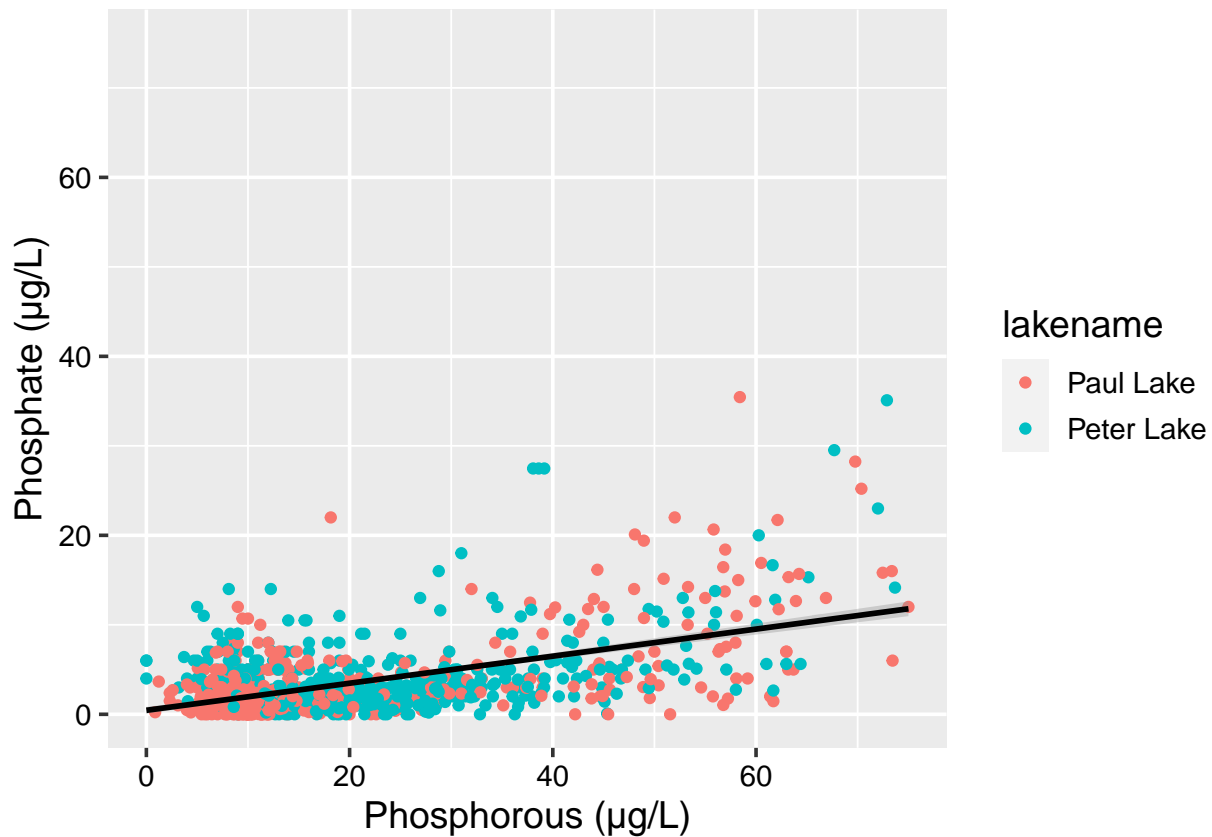
theme_set(A5theme)
```

Create graphs

For numbers 4-7, create ggplot graphs and adjust aesthetics to follow best practices for data visualization. Ensure your theme, color palettes, axes, and additional aesthetics are edited accordingly.

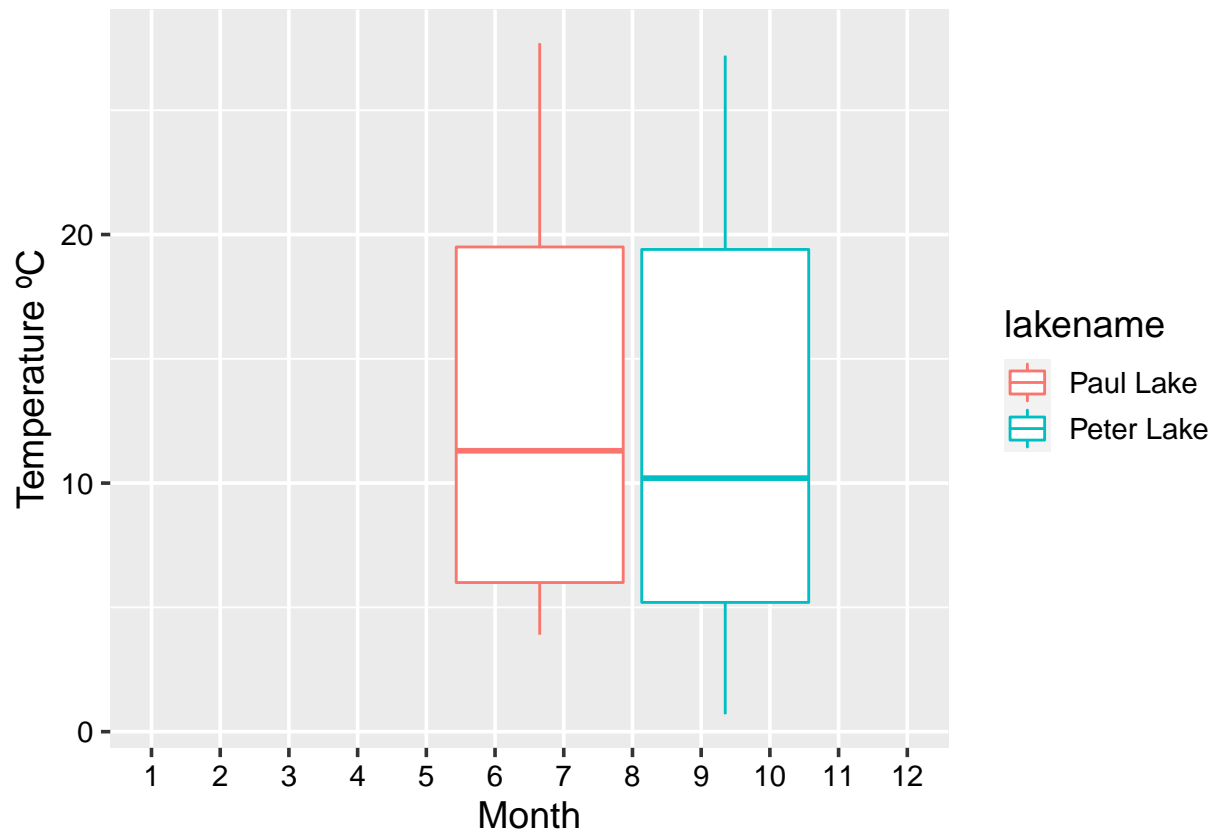
4. [NTL-LTER] Plot total phosphorus (tp_ug) by phosphate (po4), with separate aesthetics for Peter and Paul lakes. Add a line of best fit and color it black. Adjust your axes to hide extreme values (hint: change the limits using `xlim()` and `ylim()`).

```
#4
ggplot(peterpaulchemistry, aes(x = tp_ug, y = po4)) +
  geom_point(aes(color=lakename))+
  labs(y= "Phosphate (µg/L)", x = "Phosphorous (µg/L)")+
  geom_smooth(method = lm, color="black")+
  xlim(0, 75) +
  ylim(0, 75)
```

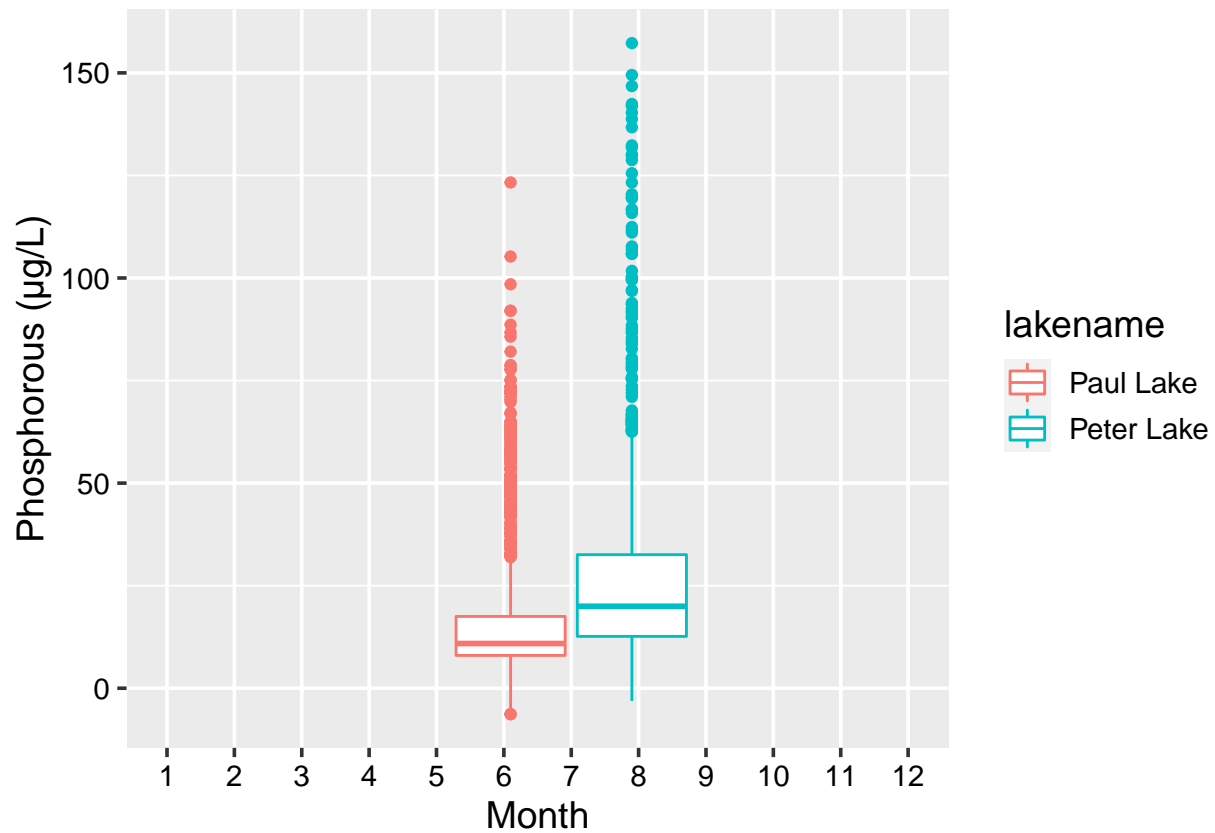


5. [NTL-LTER] Make three separate boxplots of (a) temperature, (b) TP, and (c) TN, with month as the x axis and lake as a color aesthetic. Then, create a cowplot that combines the three graphs. Make sure that only one legend is present and that graph axes are aligned.

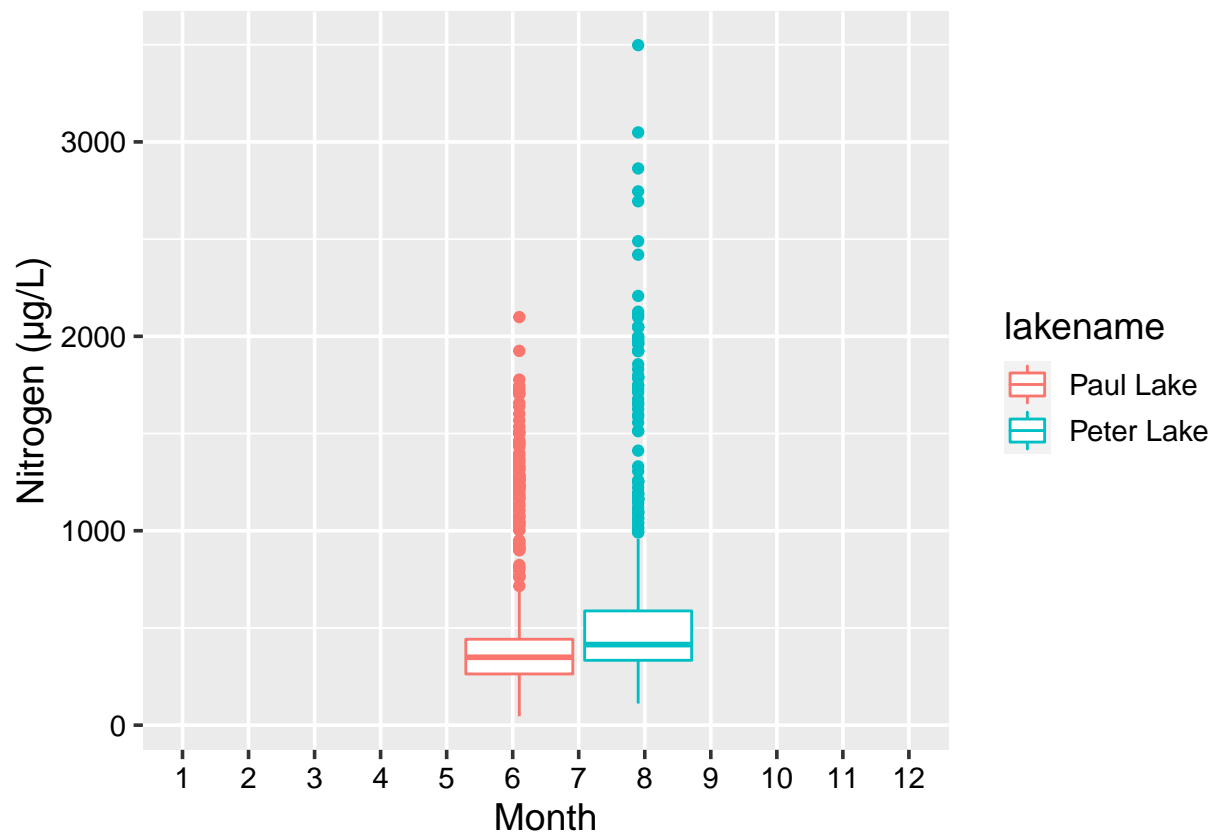
```
#5
#Individual box plots
temperature <-
  ggplot(peterpaulchemistry, aes(x = month, y = temperature_C, color=lakename)) +
  geom_boxplot()+
  labs(y= "Temperature °C", x = "Month")+
  scale_x_discrete(
    limits=c("1","2","3", "4", "5", "6", "7", "8", "9", "10", "11", "12"))
print(temperature)
```



```
TP <-
  ggplot(peterpaulchemistry, aes(x = month, y = tp_ug, color=lakename)) +
  geom_boxplot()+
  labs(y="Phosphorous (µg/L)", x="Month")+
  scale_x_discrete(
    limits=c("1","2","3", "4", "5", "6", "7", "8", "9", "10", "11", "12"))
print(TP)
```

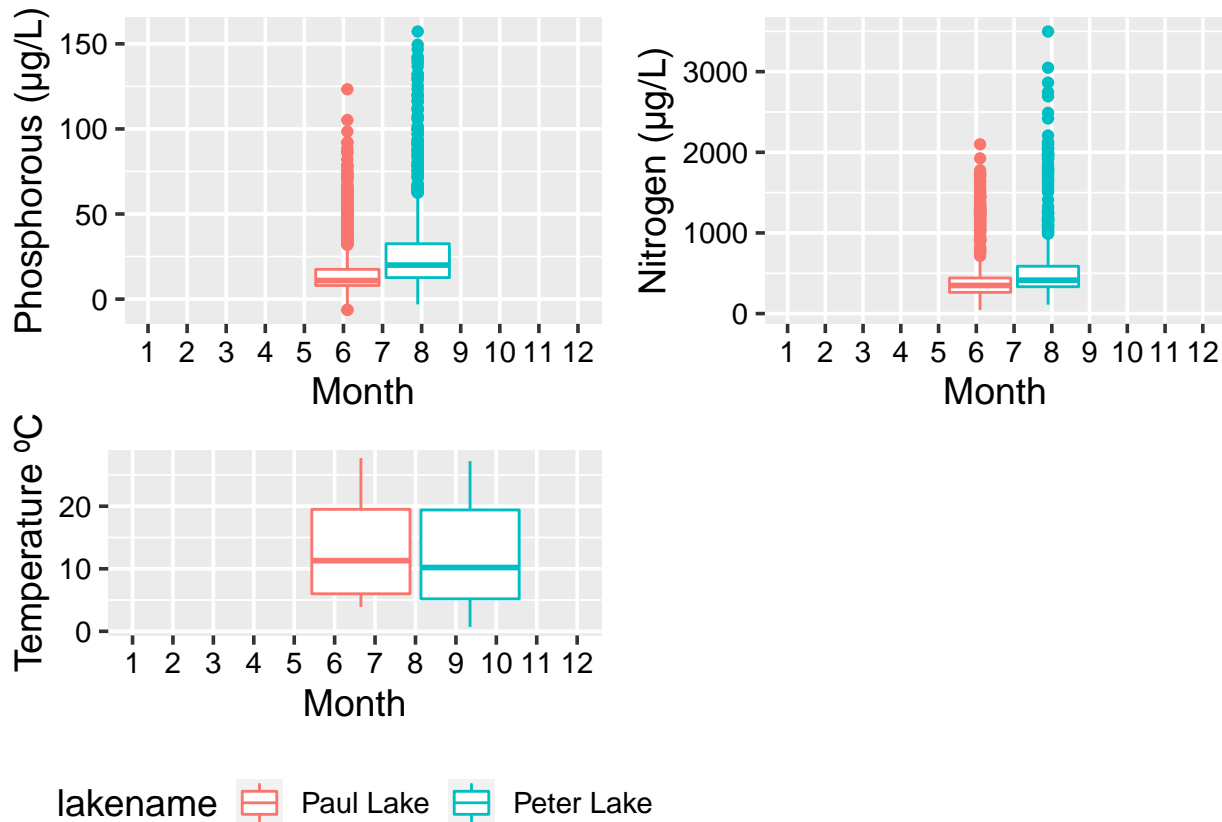


```
TN <-
  ggplot(peterpaulchemistry, aes(x = month, y = tn_ug, color=lakename)) +
  geom_boxplot()+
  labs(y="Nitrogen (µg/L)", x="Month")+
  scale_x_discrete(
    limits=c("1","2","3", "4", "5", "6", "7", "8", "9", "10", "11", "12"))
print(TN)
```



```
# Combination with one legend

plot_grid(TP +
  theme(legend.position="none"),
  TN+
  theme(legend.position="none"),
  temperature+theme(legend.position = "bottom"),
  nrow = 2, rel_heights= c(4, 4, 6), rel_widths = c(3, 3, 3))
```

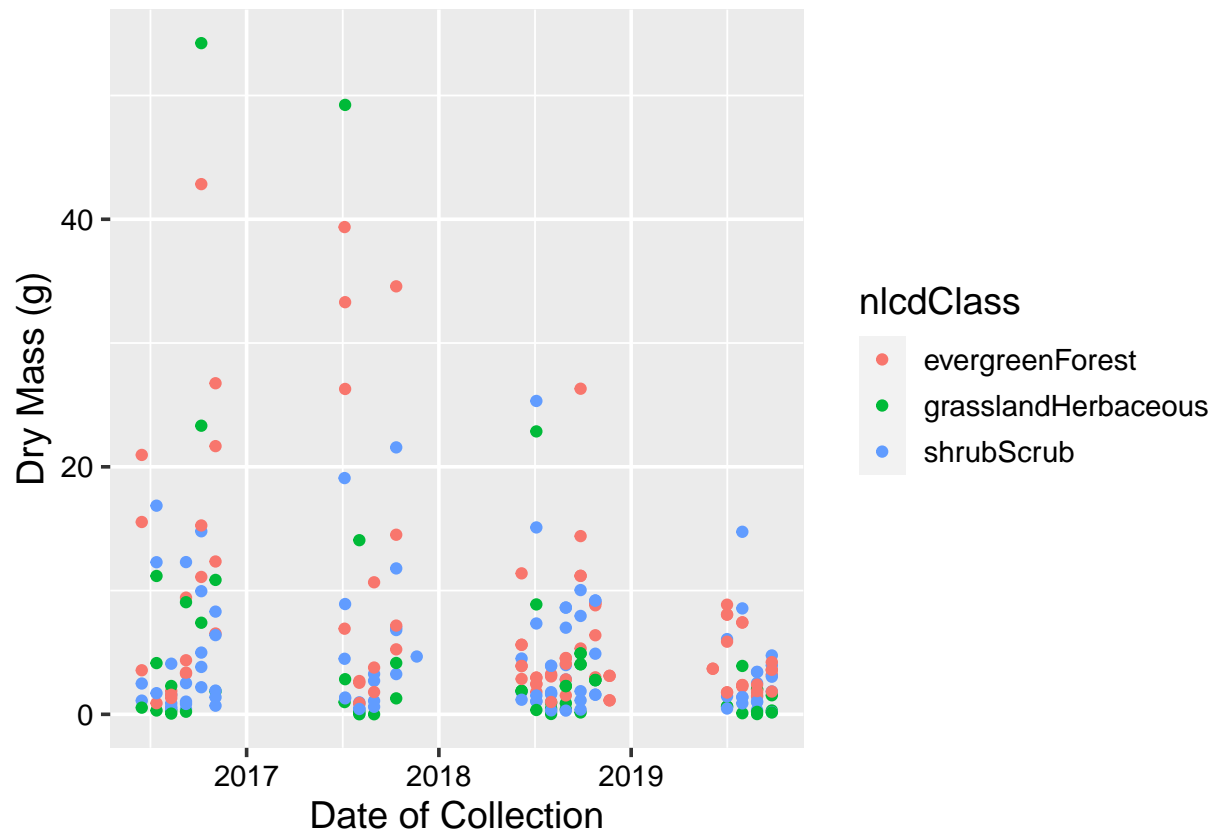


Question: What do you observe about the variables of interest over seasons and between lakes?

Answer: For each of the variables, the observations for Paul Lake are concentrated near month 6 (June) and for Peter Lake near August (8). This is more true for the nitrogen and phosphorous observations, with temperature observations ranging from mid-May to July for Paul Lake and from August to mid-October for Peter Lake. While temperature does not have outliers, both nitrogen and phosphorous have outliers in both Peter and Paul Lake. The outliers are concentrated above the median, with only one outlier below the median for phosphorous and none for nitrogen.

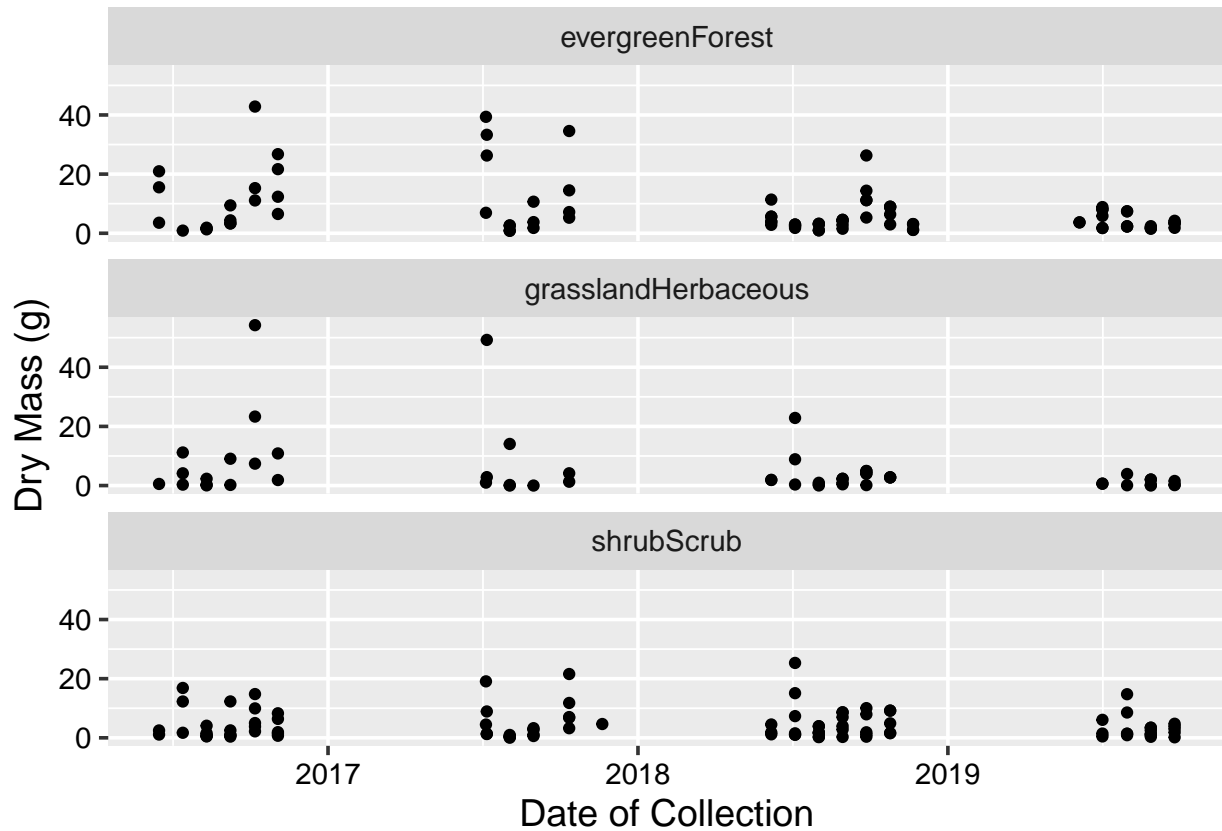
6. [Niwot Ridge] Plot a subset of the litter dataset by displaying only the “Needles” functional group. Plot the dry mass of needle litter by date and separate by NLCD class with a color aesthetic. (no need to adjust the name of each land use)
7. [Niwot Ridge] Now, plot the same plot but with NLCD classes separated into three facets rather than separated by color.

```
#6
ggplot(subset(litter.masstrap, functionalGroup == "Needles"),
  aes(x = collectDate, y = dryMass)) +
  labs(y = "Dry Mass (g)", x = "Date of Collection") +
  geom_point(aes(color = nlcdClass))
```



#7

```
ggplot(subset(litter.masstrap, functionalGroup == "Needles"),
  aes(x = collectDate, y = dryMass)) +
  labs(y= "Dry Mass (g)", x = "Date of Collection")+
  geom_point() +
  facet_wrap(vars(nlcdClass), nrow = 3)
```

Question: Which of these plots (6 vs. 7) do you think is more effective, and why?

Answer: The second plot (7) shows the distribution of observations for each NLCD class more clearly and is thus a more effective visualization of the data. In the first plot (question 6), it is difficult to discern the NLCD class, particularly of observations with low dry mass given the density of observations near zero. Plot 7 allows the viewer to clearly see the dry masses of each NLCD class while also allowing for comparison between the classes.