Assignment 5: Data Visualization

Student Name

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_NIWO_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] "Z:/EDA/Environmental_Data_Analytics_2022"
```

```
library(tidyverse)
```

```
## -- Attaching packages --------------------------------------- tidyverse 1.3.1 --

## v ggplot2 3.3.5     v purrr   0.3.4
## v tibble  3.1.6     v dplyr   1.0.7
## v tidyr   1.1.4     v stringr 1.4.0
## v readr   2.1.1     v forcats 0.5.1

## -- Conflicts ------------------------------------------ tidyverse_conflicts() --
## x dplyr::filter() masks stats::filter()
## x dplyr::lag()    masks stats::lag()
```

```
library(cowplot)

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

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

#2
NTL_LTER$sampledate<-
  as.Date(NTL_LTER$sampledate, format = "%Y-%m-%d")

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

class(Niwot_Ridge$collectDate)
```

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

```r
class(NTL_LTER$sampledate)#now they're dates!
```

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

Define your theme

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

```r
#3
mytheme <- theme_classic(base_size = 14)+
  theme(axis.text = element_text(color = "black"),
        legend.position = "top")
theme_set(mytheme)
```

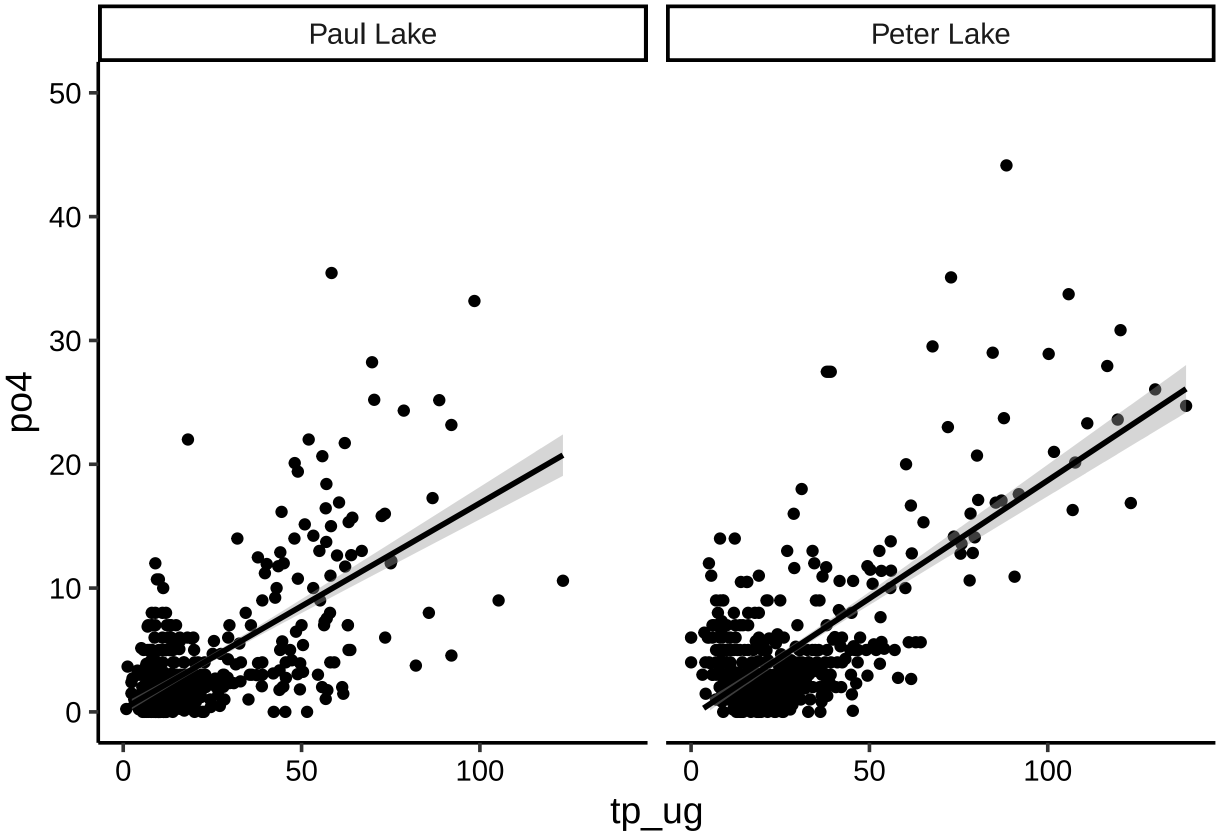
Create graphs

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()`).

```r
#4
tp_ug_vs_po4<-
  ggplot(NTL_LTER, aes(x = tp_ug, y = po4))+
  geom_point()+
  facet_wrap(vars(lakename))+
  xlim(0,140)+
  ylim(0,50)+
  geom_smooth(method = lm, color = "black")
print(tp_ug_vs_po4)
```

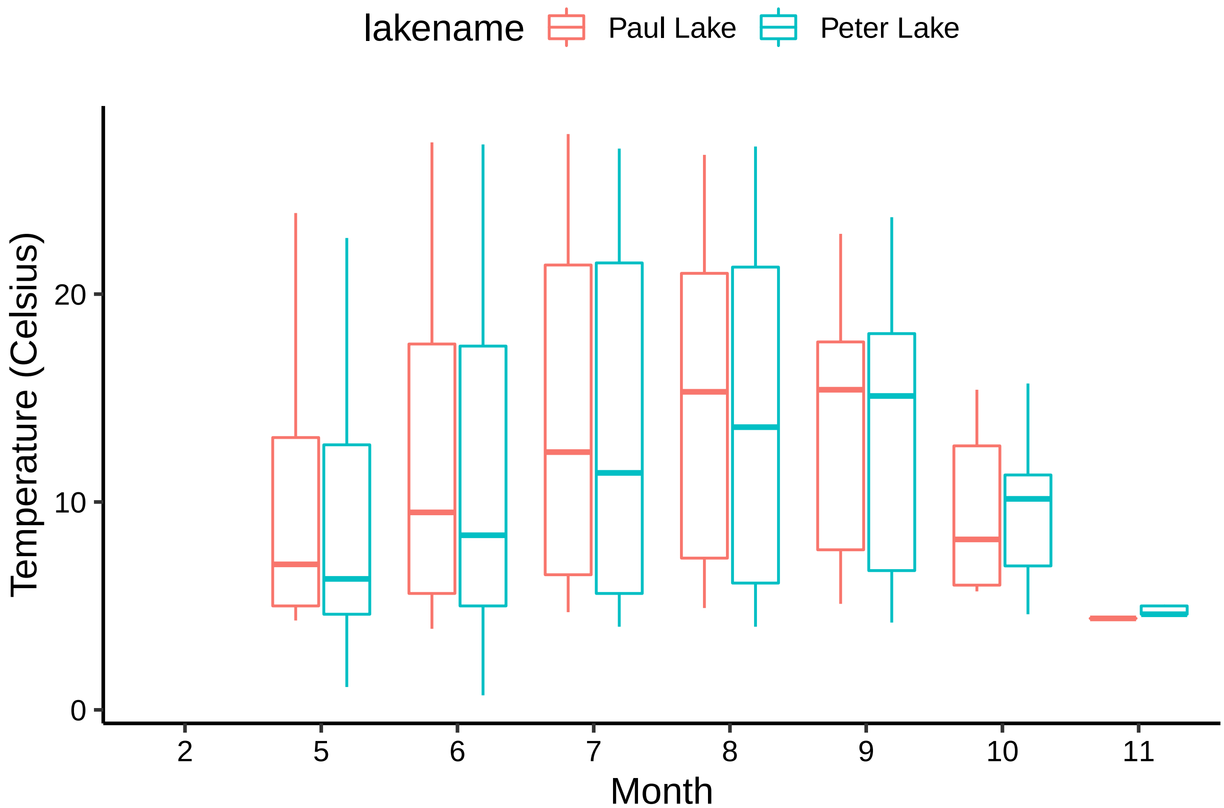
```
## `geom_smooth()` using formula 'y ~ x'

## Warning: Removed 21950 rows containing non-finite values (stat_smooth).

## Warning: Removed 21950 rows containing missing values (geom_point).

## Warning: Removed 2 rows containing missing values (geom_smooth).
```



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.

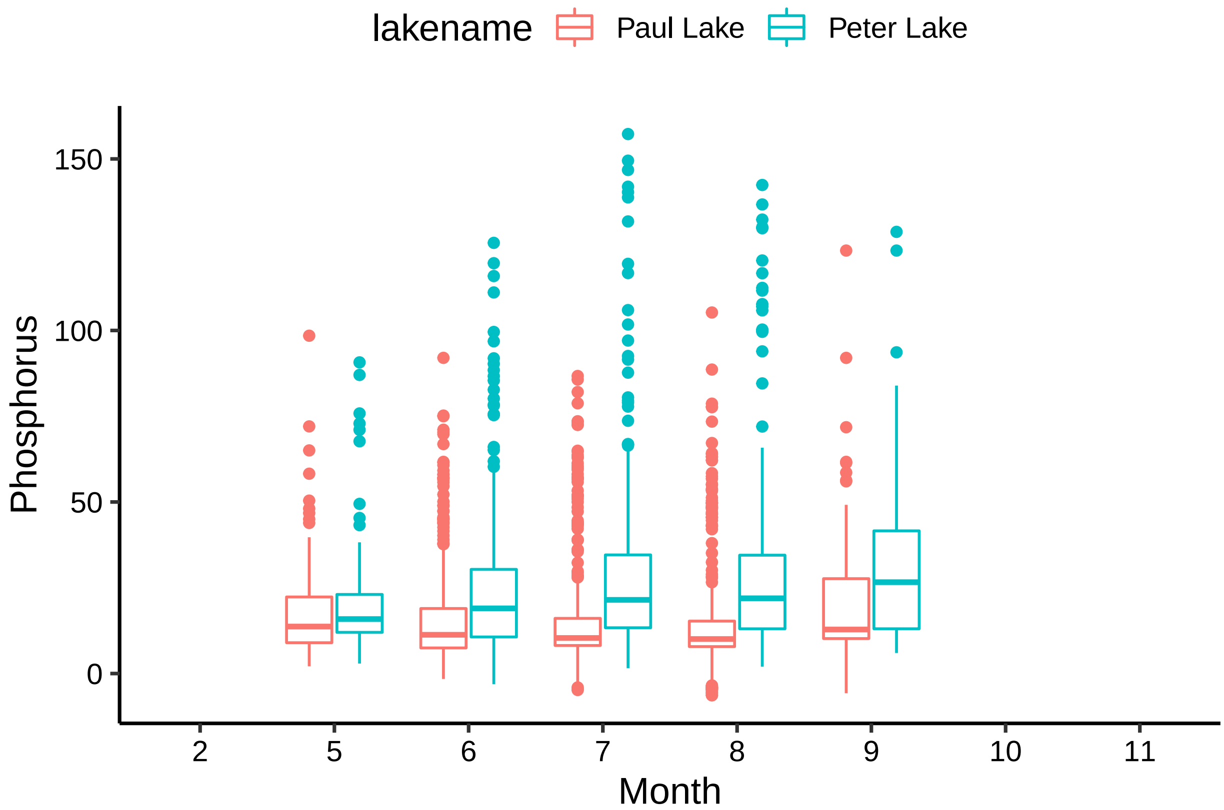
```
#5
NTL_LTER$month<-as.factor(NTL_LTER$month)
#forcing "month" to be a factor instead of a numeric
#variable so I can see multiple box plots

temp_boxplot<-
  ggplot(NTL_LTER, aes(x = month, y = temperature_C))+
  geom_boxplot(aes(color = lakename))+
  xlab("Month")+
  ylab("Temperature (Celsius)")
print(temp_boxplot)
```

```
## Warning: Removed 3566 rows containing non-finite values (stat_boxplot).
```



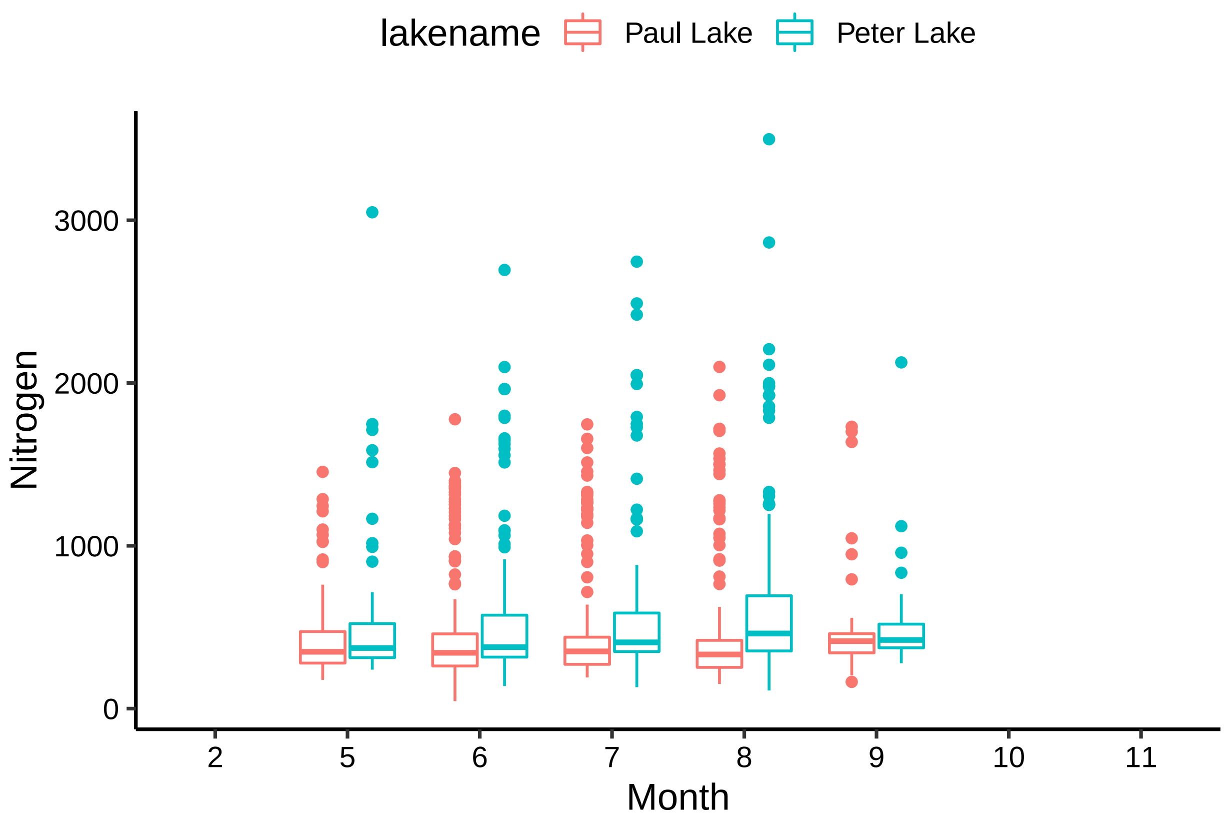

```
tp_boxplot<-
  ggplot(NTL_LTER, aes(x = month, y = tp_ug))+
  geom_boxplot(aes(color = lakename))+
  xlab("Month")+
  ylab("Phosphorus")
print(tp_boxplot)
```

```
## Warning: Removed 20729 rows containing non-finite values (stat_boxplot).
```




```r
tn_boxplot<-
  ggplot(NTL_LTER, aes(x = month, y = tn_ug))+
  geom_boxplot(aes(color = lakename))+
  xlab("Month")+
  ylab("Nitrogen")
print(tn_boxplot)
```

```
## Warning: Removed 21583 rows containing non-finite values (stat_boxplot).
```



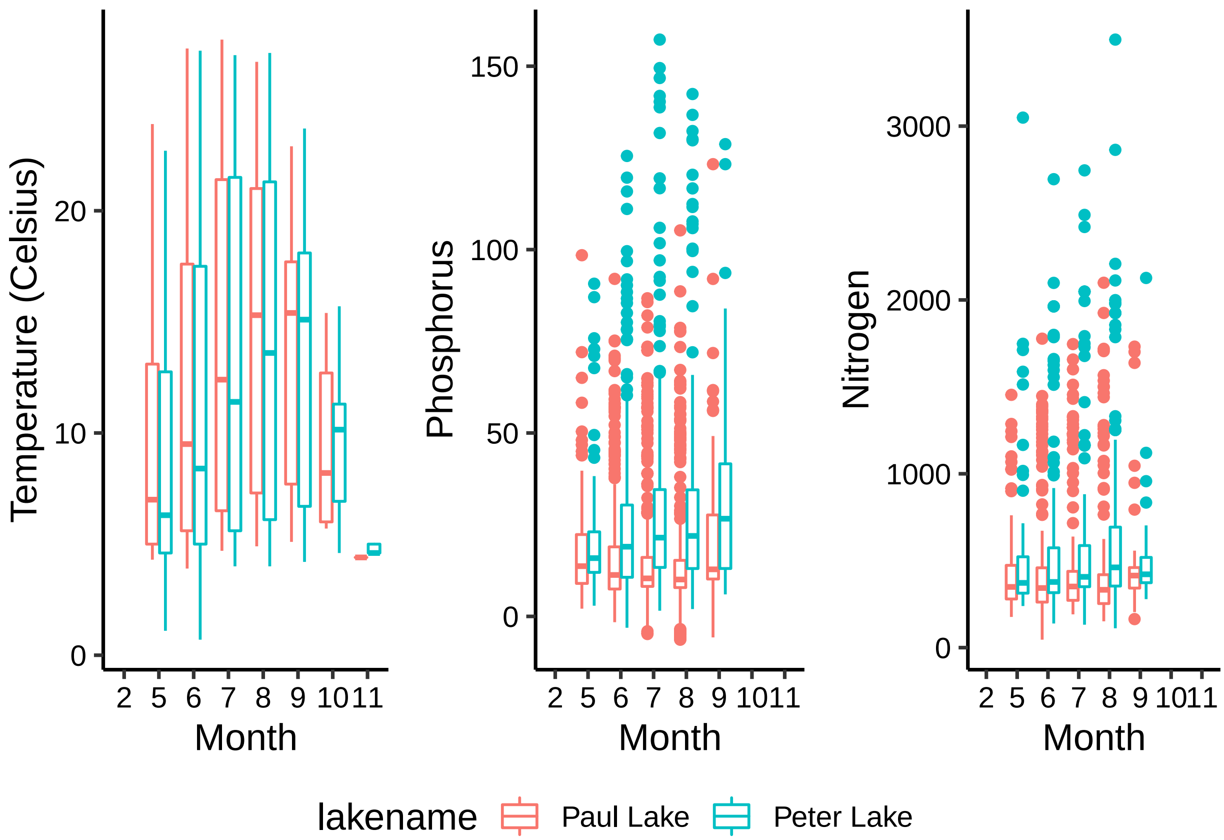

```
#cowplot:
three_plots<-plot_grid(temp_boxplot + theme(legend.position="none"),
                       tp_boxplot + theme(legend.position="none"),
                       tn_boxplot + theme(legend.position="none"),
                       nrow=1, align='h', rel_heights = c(1,1,1))
```

```
## Warning: Removed 3566 rows containing non-finite values (stat_boxplot).

## Warning: Removed 20729 rows containing non-finite values (stat_boxplot).

## Warning: Removed 21583 rows containing non-finite values (stat_boxplot).
```

```
#print(three_plots)

plot_legend<-get_legend(temp_boxplot + theme(legend.position = "bottom"))
```

```
## Warning: Removed 3566 rows containing non-finite values (stat_boxplot).
```

```
final_cowplot<-plot_grid(three_plots, plot_legend,ncol=1,rel_heights = c(1,.1))

print(final_cowplot)
```

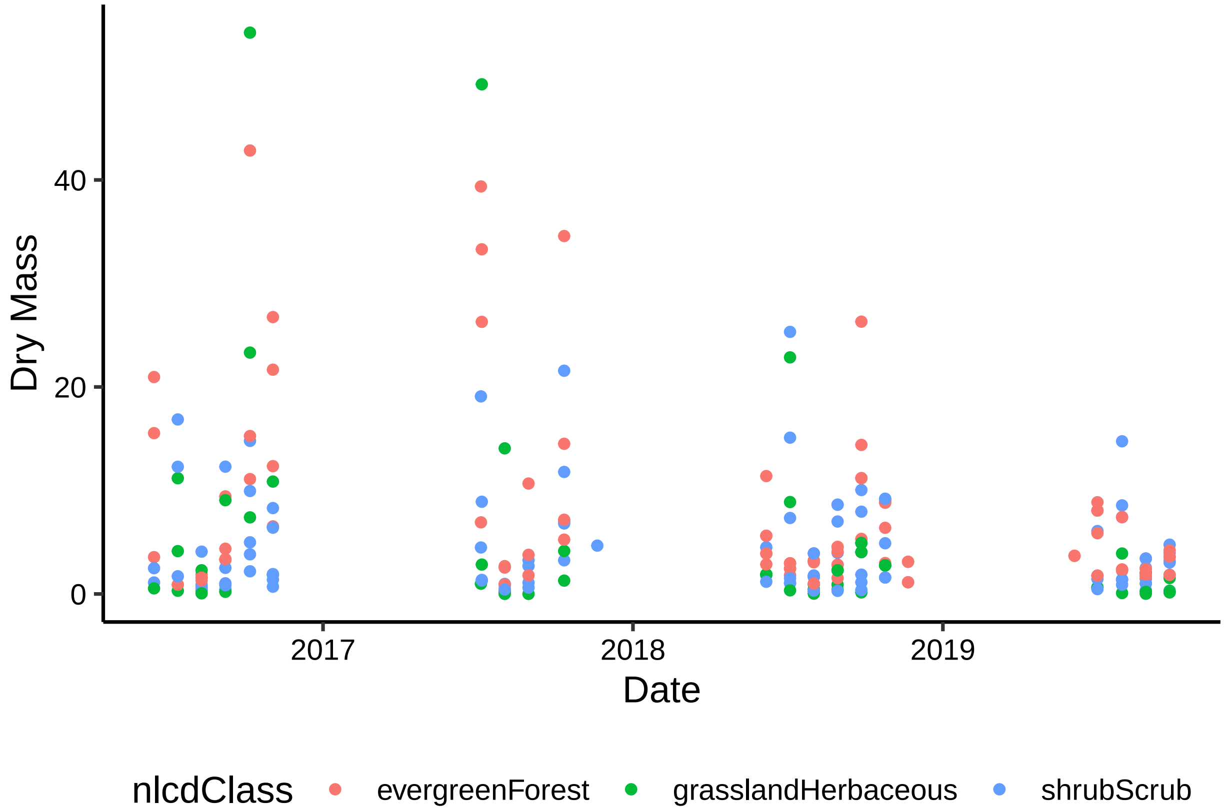


Question: What do you observe about the variables of interest over seasons and between lakes?

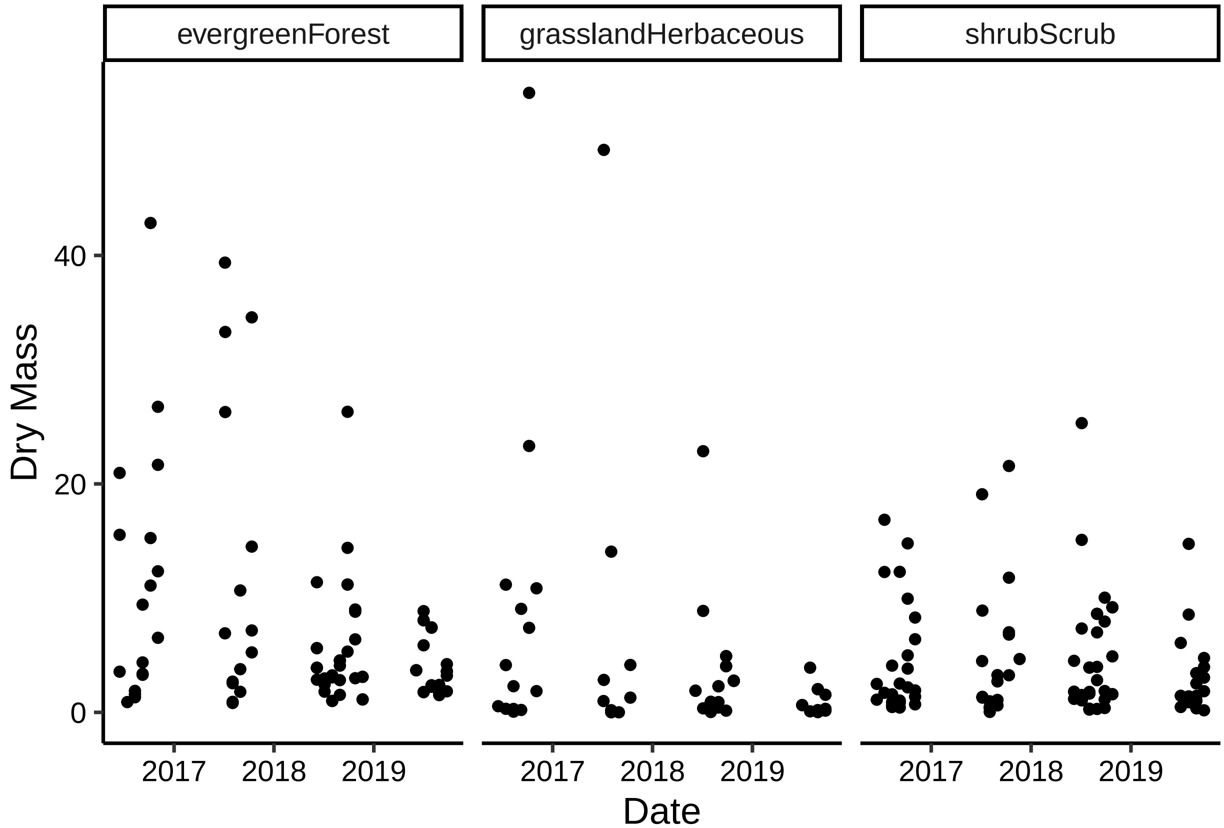
> Answer: I see that over the summer months, the temperature, nitrogen, and phosphorus levels all increased and seemed to peak aroun August. Particularly Peter Lake experienced higher levels of phosphorus and nitrogen in the summer months than Paul Lake.

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
#plot subset of the Litter dataset "needles"

Litter_Needles<-
  ggplot(filter(Niwot_Ridge, functionalGroup == "Needles"),
  aes(x = collectDate, y = dryMass, color = nlcdClass))+
  geom_point()+
  xlab("Date")+
  ylab("Dry Mass")+
  theme(legend.position = "bottom")
print(Litter_Needles)
```




```
#7
#separated by facets
Litter_Needles_Faceted<-
  ggplot(filter(Niwot_Ridge, functionalGroup == "Needles"),
  aes(x = collectDate, y = dryMass))+
  geom_point()+
  facet_wrap(vars(nlcdClass))+
  xlab("Date")+
  ylab("Dry Mass")
print(Litter_Needles_Faceted)
```



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

Answer: I think the faceted plot is more effective because it's easier to see. The color plot was difficult to see and make sense of based on location and date. This faceted plot makes it easy to see the dry mass and the different locations across years.