Assignment 5: Data Visualization

Langston Alexander

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
library(cowplot)
```

```
## Warning in register(): Can't find generic `scale_type` in package ggplot2 to
## register S3 method.
```

```
library(tidyverse)
library(wesanderson)
peter_paulLakes <-read.csv("C:/Users/lwa8/Documents/R/ENV872/Environmental_Data_Analytics_2022/Data/Proc
                           stringsAsFactors = TRUE)
niwot_litter <- read.csv("C:/Users/lwa8/Documents/R/ENV872/Environmental_Data_Analytics_2022/Data/Proces
                         stringsAsFactors = TRUE)

#2

str(peter_paulLakes)
```

```
## 'data.frame':    23008 obs. of  15 variables:
##  $ lakename       : Factor w/ 2 levels "Paul Lake","Peter Lake": 1 1 1 1 1 1 1 1 1 1 ...
##  $ year4          : int  1984 1984 1984 1984 1984 1984 1984 1984 1984 1984 ...
##  $ daynum         : int  148 148 148 148 148 148 148 148 148 148 ...
##  $ month          : int  5 5 5 5 5 5 5 5 5 5 ...
```

```
##  $ sampledate     : Factor w/ 1103 levels "1984-05-27","1984-05-28",..: 1 1 1 1 1 1 1 1 1 1 ...
##  $ depth          : num  0 0.25 0.5 0.75 1 1.5 2 3 4 5 ...
##  $ temperature_C  : num  14.5 NA NA NA 14.5 NA 14.2 11 7 6.1 ...
##  $ dissolvedOxygen: num  9.5 NA NA NA 8.8 NA 8.6 11.5 11.9 2.5 ...
##  $ irradianceWater: num  1750 1550 1150 975 870 610 420 220 100 34 ...
##  $ irradianceDeck : num  1620 1620 1620 1620 1620 1620 1620 1620 1620 1620 ...
##  $ tn_ug          : num  NA NA NA NA NA NA NA NA NA NA ...
##  $ tp_ug          : num  NA NA NA NA NA NA NA NA NA NA ...
##  $ nh34           : num  NA NA NA NA NA NA NA NA NA NA ...
##  $ no23           : num  NA NA NA NA NA NA NA NA NA NA ...
##  $ po4            : num  NA NA NA NA NA NA NA NA NA NA ...
```

```
str(niwot_litter)
```

```
## 'data.frame':    1692 obs. of  13 variables:
##  $ plotID          : Factor w/ 12 levels "NIWO_040","NIWO_041",..: 9 8 9 11 7 7 4 4 4 4 ...
##  $ trapID          : Factor w/ 15 levels "NIWO_040_139",..: 11 10 11 13 9 9 5 5 5 5 ...
##  $ collectDate     : Factor w/ 24 levels "2016-06-16","2016-07-14",..: 1 1 1 1 1 1 1 1 1 1 ...
##  $ functionalGroup : Factor w/ 8 levels "Flowers","Leaves",..: 6 5 8 6 4 2 2 6 7 8 ...
##  $ dryMass         : num  0 0.27 0.12 0 1.11 0 0 0 0.07 0.02 ...
##  $ qaDryMass       : Factor w/ 2 levels "N","Y": 1 1 1 1 2 1 1 1 1 1 ...
##  $ subplotID       : int  31 41 31 32 32 32 40 40 40 40 ...
##  $ decimalLatitude : num  40.1 40 40.1 40 40 ...
##  $ decimalLongitude: num  -106 -106 -106 -106 -106 ...
##  $ elevation       : num  3477 3413 3477 3373 3446 ...
##  $ nlcdClass       : Factor w/ 3 levels "evergreenForest",..: 3 1 3 1 3 3 2 2 2 2 ...
##  $ plotType        : Factor w/ 1 level "tower": 1 1 1 1 1 1 1 1 1 1 ...
##  $ geodeticDatum   : Factor w/ 1 level "WGS84": 1 1 1 1 1 1 1 1 1 1 ...
```

```
peter_paulLakes$sampledate <- as.Date(peter_paulLakes$sampledate,
                                      format = "%Y-%m-%d")
niwot_litter$collectDate <- as.Date(niwot_litter$collectDate,
                                    format = "%Y-%m-%d")
```

Define your theme

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

```
#3

mytheme <- theme_light(base_size = 10)+
  theme(axis.text = element_text(color = "Black"),
        plot.title =element_text(size=12, face="bold", hjust = 0.5),
        )

theme_set(mytheme)
```

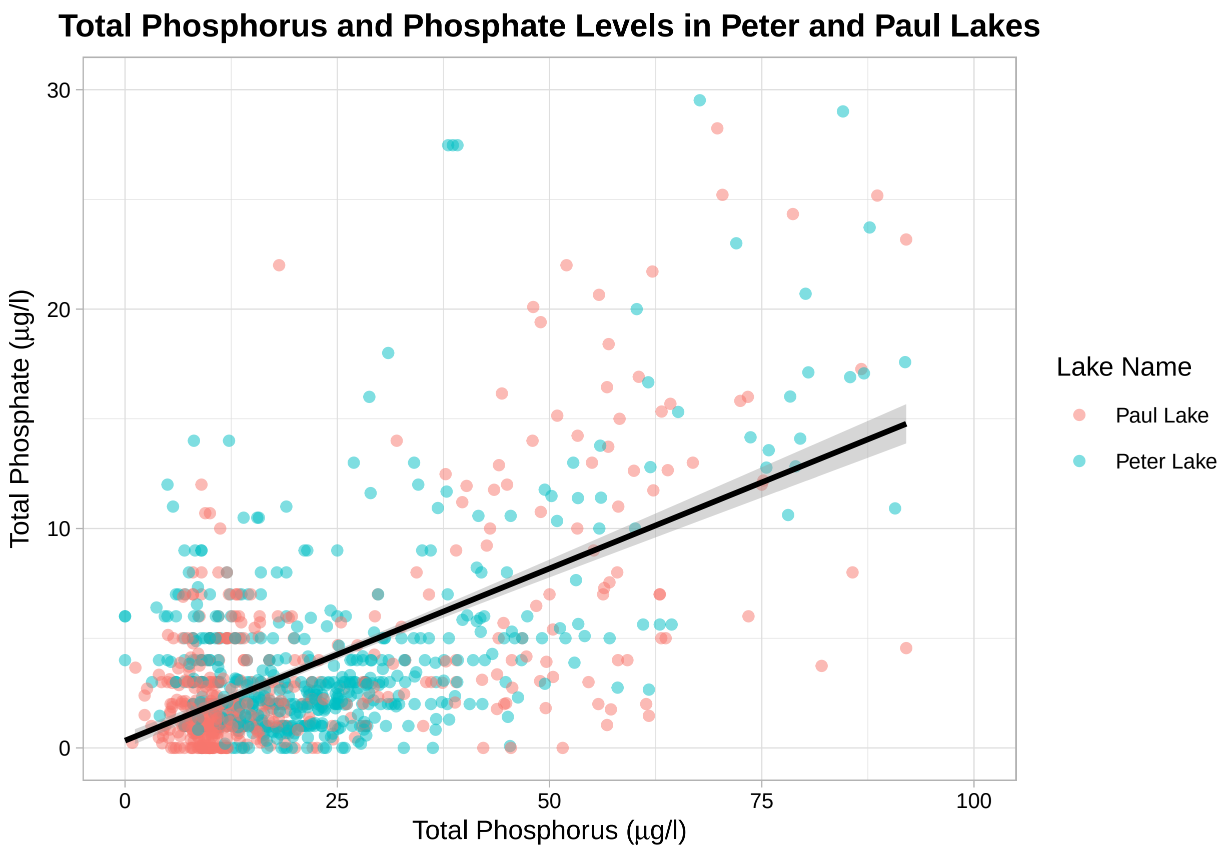
Create graphs

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

ggplot(peter_paulLakes,
       aes(x = tp_ug, y = po4, color = lakename))+
  geom_point(alpha = 0.5)+
  geom_smooth(method = "lm", color = "Black")+
  ylim(0, 30)+
  xlim(0,100)+
  labs(title = "Total Phosphorus and Phosphate Levels in Peter and Paul Lakes",
       x = expression(paste("Total Phosphorus (", mu, "g/l)")),
       y = expression(paste("Total Phosphate (", mu, "g/l)")),
       color = "Lake Name")
```

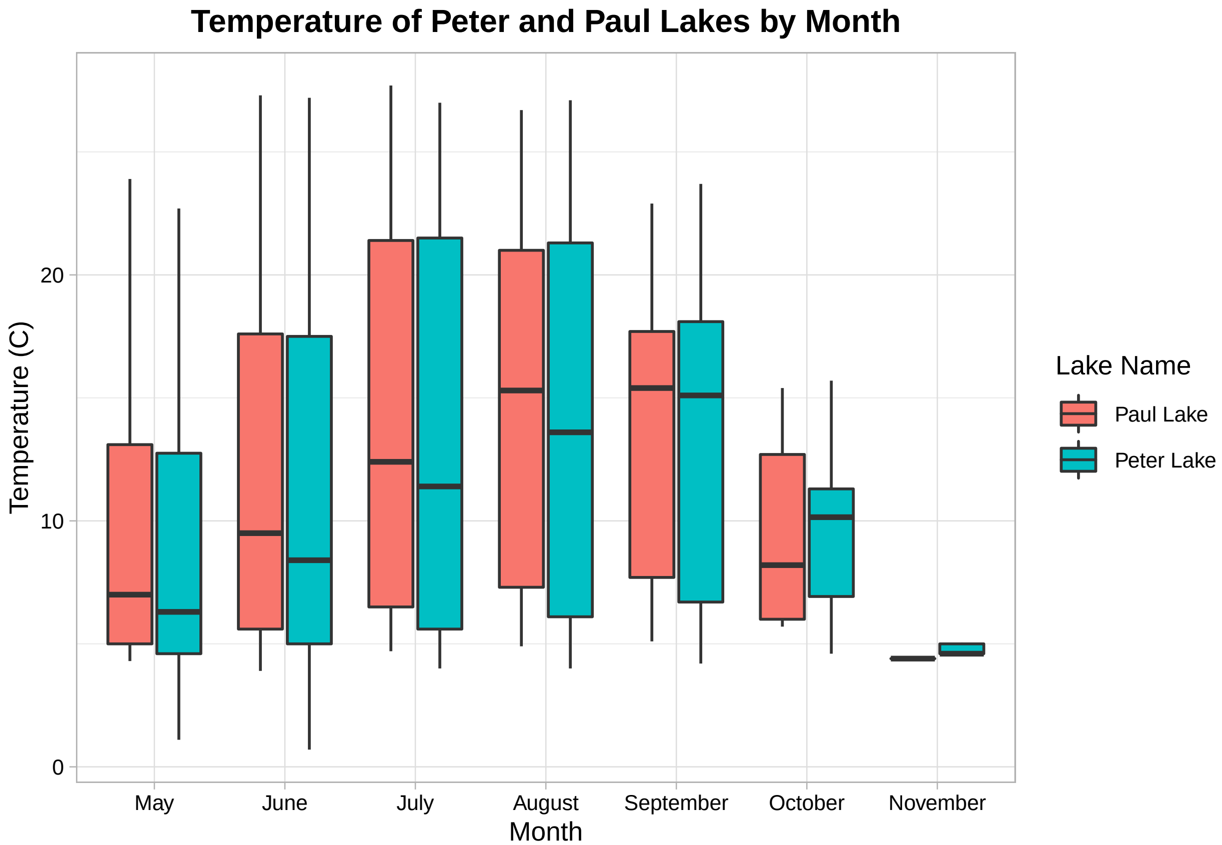
```
## `geom_smooth()` using formula 'y ~ x'
```



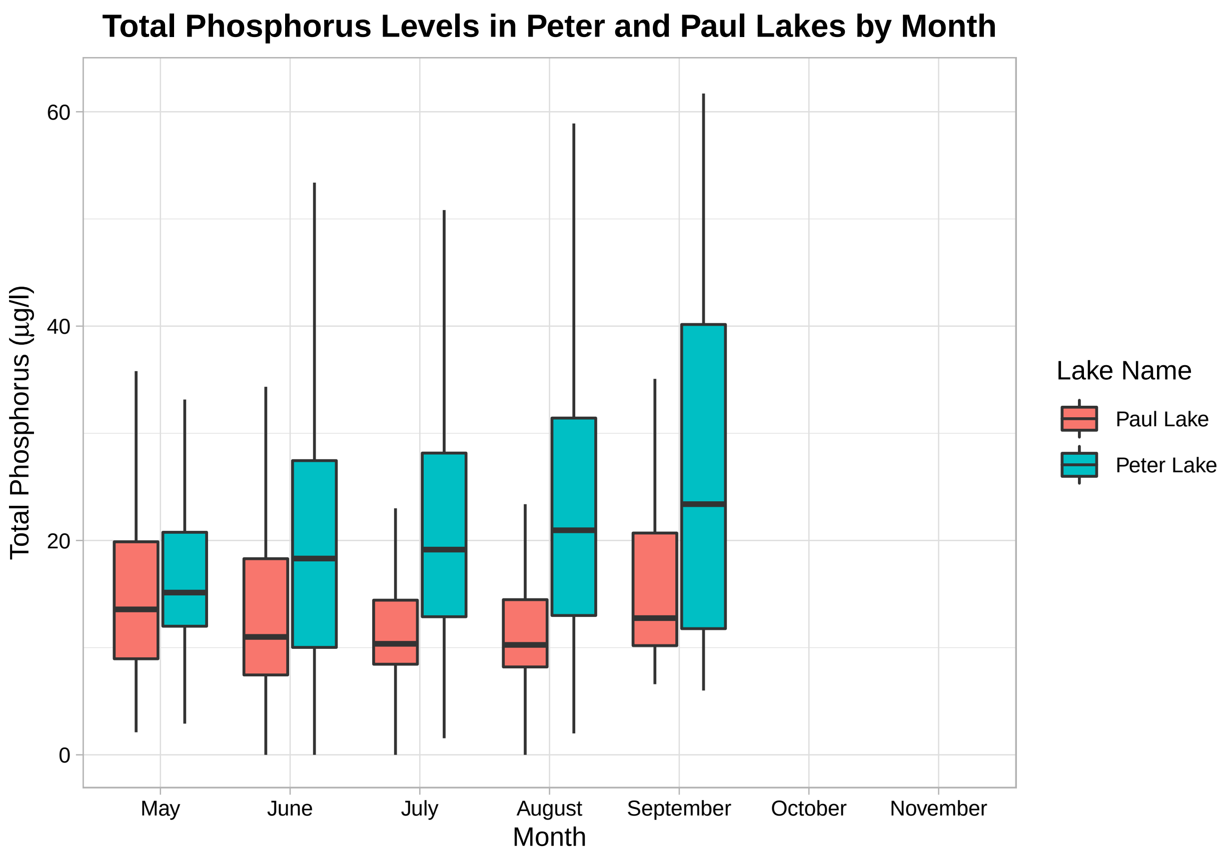
5. [NTL-LTER] Make three separate boxplots of (a) temperature, (b) TP, and (c) TN, with month as the x axis and lake as a color aesthetic. Then, create a cowplot that combines the three graphs. Make sure that only one legend is present and that graph axes are aligned.

```r
#5
boxplot_base <- ggplot(peter_paulLakes)+
  xlab("Month")+
   scale_x_discrete(limits = c("5", "6", "7", "8", "9", "10", "11"),
                    labels = c("May", "June", "July", "August", "September", "October", "November"))

#Separate Box Plots
boxplot_base+
```

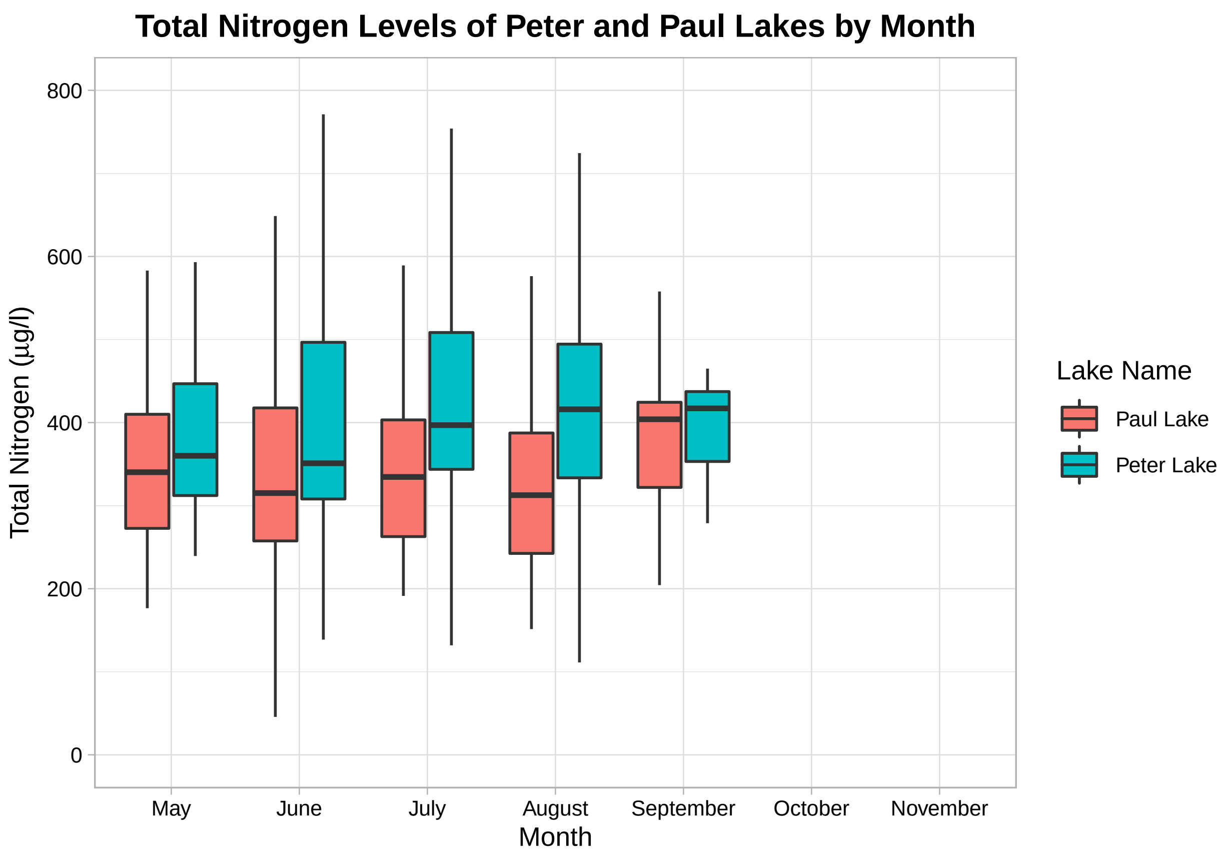
```r
    geom_boxplot(aes(x = as.factor(month),
                     y = temperature_C,
                     fill = lakename))+
  labs(title = "Temperature of Peter and Paul Lakes by Month",
       y = "Temperature (C)",
       fill = "Lake Name")
```




```r
boxplot_base+
    geom_boxplot(aes(x = as.factor(month),
                     y = tp_ug, fill = lakename),
                 outlier.shape = NA)+
  labs(title = "Total Phosphorus Levels in Peter and Paul Lakes by Month",
       y = expression(paste("Total Phosphorus (", mu, "g/l)")),
       fill = "Lake Name")+
   ylim(0,62)
```




```
boxplot_base+
    geom_boxplot(aes(x = as.factor(month),
                     y = tn_ug, fill = lakename),
                 outlier.shape = NA)+
  labs(title = "Total Nitrogen Levels of Peter and Paul Lakes by Month",
       y= expression(paste("Total Nitrogen (", mu, "g/l)")),
        fill = "Lake Name")+
  ylim(0,800)
```

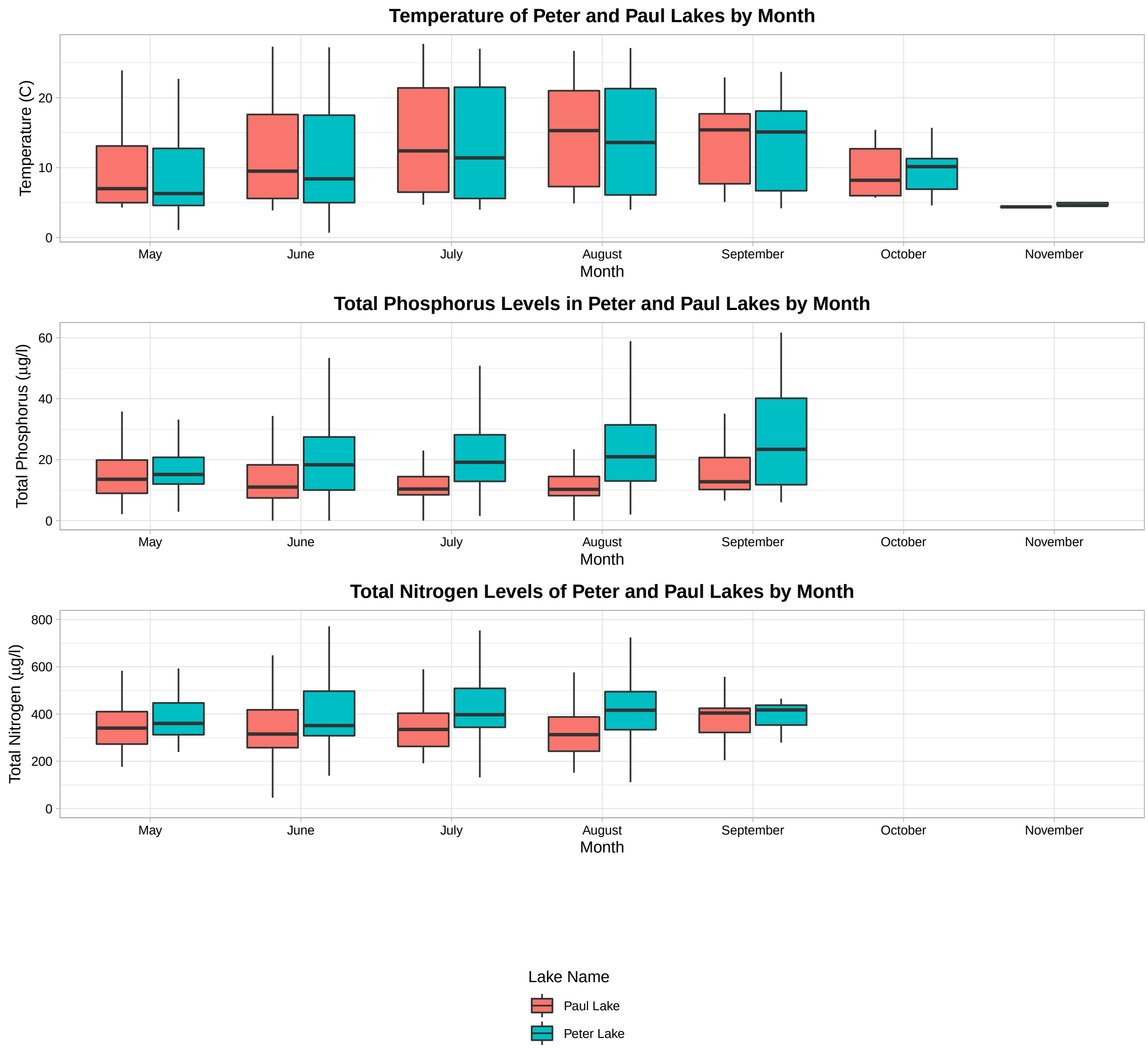



```
#Box Plots without legends for the combined plot
temp <- boxplot_base+
    geom_boxplot(aes(x = as.factor(month),
                     y = temperature_C,
                     fill = lakename))+
  labs(title = "Temperature of Peter and Paul Lakes by Month",
       y = "Temperature (C)",
       fill = "Lake Name")+
  theme(legend.position = "none")



phos <- boxplot_base+
    geom_boxplot(aes(x = as.factor(month),
                     y = tp_ug, fill = lakename),
                    outlier.shape = NA)+
  labs(title = "Total Phosphorus Levels in Peter and Paul Lakes by Month",
       y = expression(paste("Total Phosphorus (", mu, "g/l)")),
        fill = "Lake Name")+
   ylim(0,62)+
  theme(legend.position = "none")


nitro <- boxplot_base+
    geom_boxplot(aes(x = as.factor(month),
                     y = tn_ug, fill = lakename),
```

```r
                 outlier.shape = NA)+
  labs(title = "Total Nitrogen Levels of Peter and Paul Lakes by Month",
       y= expression(paste("Total Nitrogen (", mu, "g/l)")),
        fill = "Lake Name")+
  ylim(0,800)+
  theme(legend.position = "none")

legend <- get_legend(
  temp +
    theme(legend.position = "right"))
```

```r
plot_grid(temp, phos, nitro, legend,
          ncol = 1, axis = c("l"), align = "v")
```

```
## Warning: Removed 16 rows containing missing values (stat_boxplot).

## Warning: Removed 3550 rows containing non-finite values (stat_boxplot).

## Warning: Removed 16 rows containing missing values (stat_boxplot).

## Warning: Removed 20877 rows containing non-finite values (stat_boxplot).

## Warning: Removed 16 rows containing missing values (stat_boxplot).

## Warning: Removed 21757 rows containing non-finite values (stat_boxplot).
```

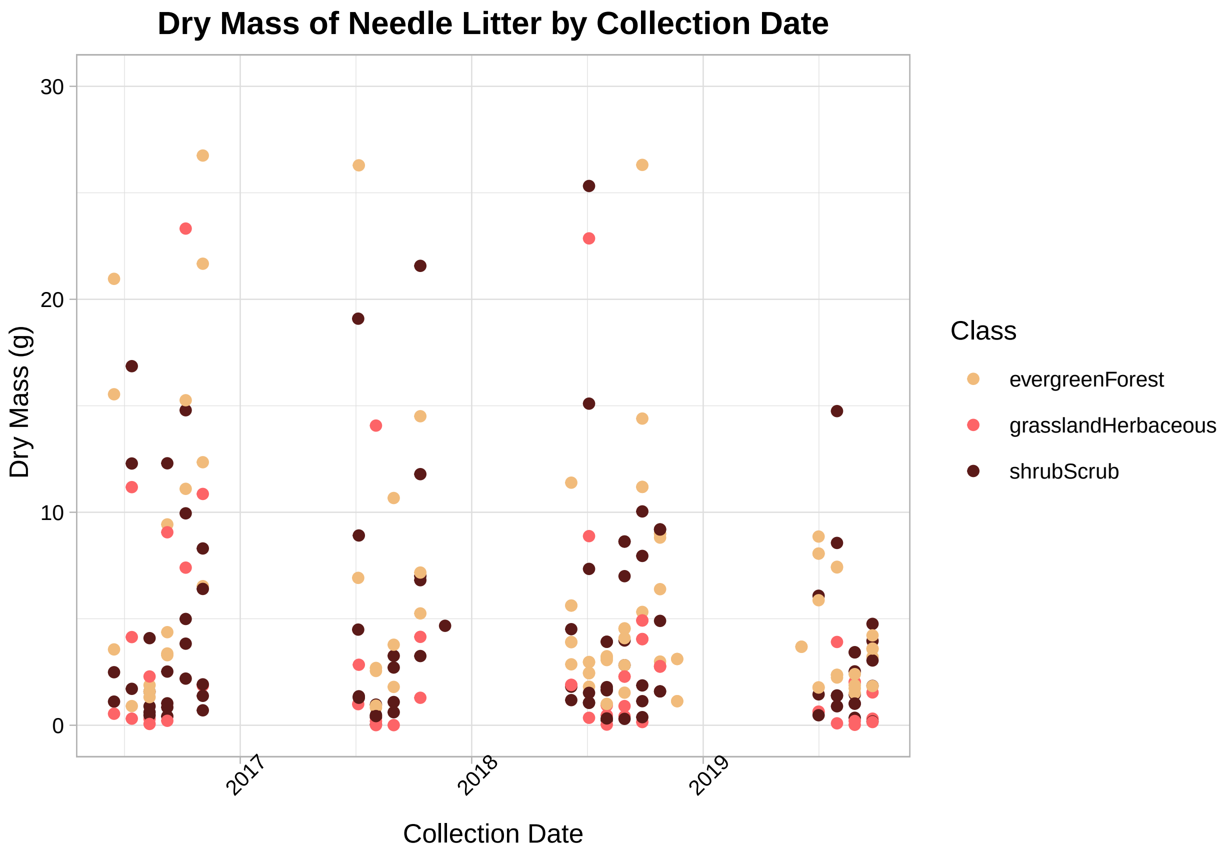


Question: What do you observe about the variables of interest over seasons and between lakes?

> Answer: The temperature follows a similar pattern across both lakes. They both warm up during the summer and cool off through the fall. For phosphorus levels there is a divergence between the two lakes from May through September. They have nearly the same median level of phosporus in May, bbut Peter Lake increases slightly each month until September and Paul Lake slightly decreases. For nitrogen, Peter Lake continually increases from June to September, while Paul Lake hovers between about 325 and 350 micrograms per liter from May to August, then jumps up to over 400 in September.

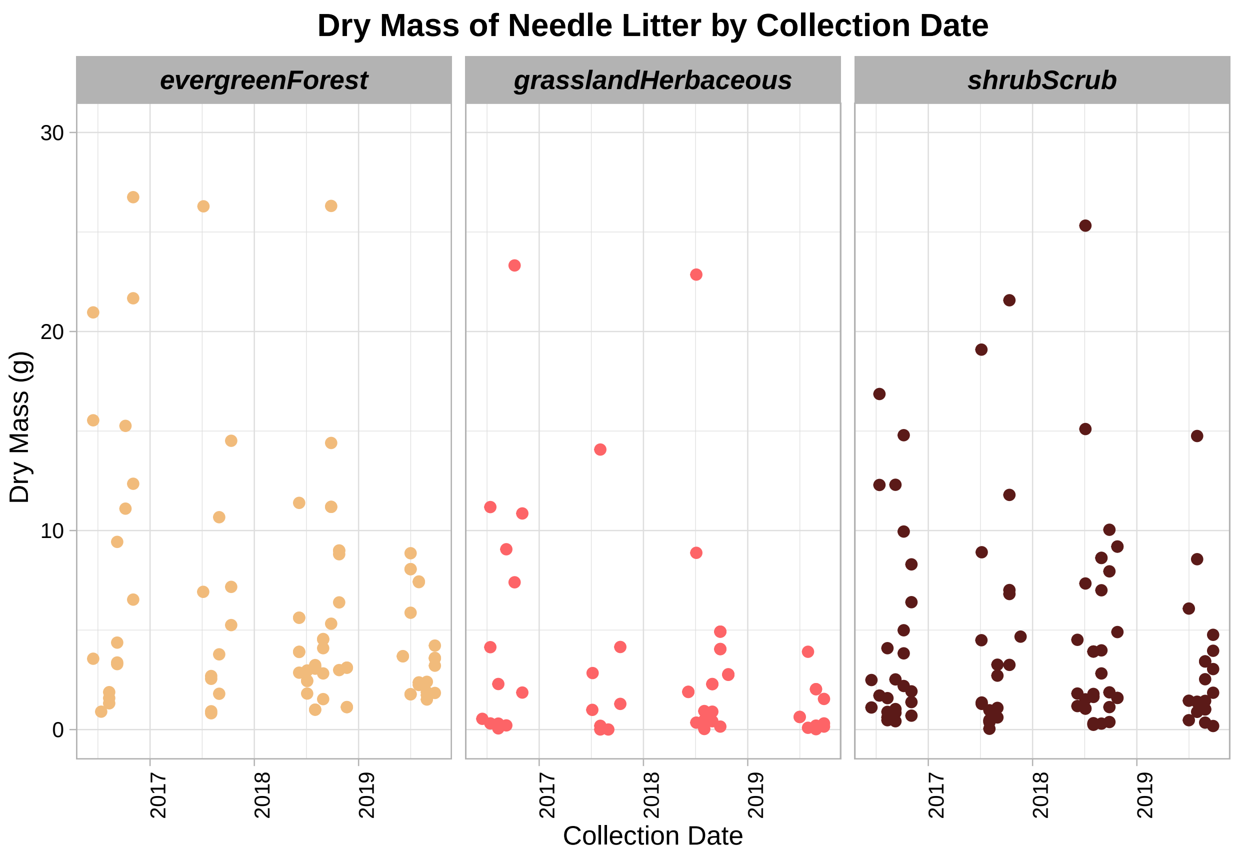
6. [Niwot Ridge] Plot a subset of the litter dataset by displaying only the "Needles" functional group. Plot the dry mass of needle litter by date and separate by NLCD class with a color aesthetic. (no need to adjust the name of each land use)
7. [Niwot Ridge] Now, plot the same plot but with NLCD classes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

```r
#6
ggplot(subset(niwot_litter, functionalGroup == "Needles"),
       aes(x = collectDate, y = dryMass, color = nlcdClass))+
  geom_point()+
  labs(title = "Dry Mass of Needle Litter by Collection Date",
       y = " Dry Mass (g)",
       x = "Collection Date",
       color = "Class")+
  ylim(0,30)+
  scale_color_manual(values = wes_palette("GrandBudapest1", 3))+
  theme(axis.text.x = element_text(angle = 45))
```




```r
#7

ggplot(subset(niwot_litter, functionalGroup == "Needles"),
       aes(x = collectDate, y = dryMass, color = nlcdClass))+
  geom_point()+
  labs(title = "Dry Mass of Needle Litter by Collection Date",
       y = " Dry Mass (g)",
       x = "Collection Date",
       color = "Class")+
  ylim(0,30)+
  scale_color_manual(labels = c("Evergreen Forest", "Herbaceous Grassland", "Scrub Shrub"),
                     values = wes_palette("GrandBudapest1", 3))+
  theme(axis.text.x = element_text(angle = 90),
        strip.text.x = element_text(
```

```
        size = 10, color = "black", face = "bold.italic"),
        legend.position = "none")+
  facet_wrap(vars(nlcdClass))
```



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

> Answer: To me, the faceted plots visualize the data more effectively. There is simply too many data points to parse out any patterns visually. By separating the points into 3 plots we can both see the patterns internal to each NLCD class, but also compare the patterns across classes more easily.