Assignment 5: Data Visualization

Blair Johnson

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:/ENV872/Environmental_Data_Analytics_2022/Assignments"
```

```
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)

#loading data
Peter.Paul.Lake<-read.csv("../Data/Processed/NTL-LTER_Lake_Chemistry_Nutrients_PeterPaul_Processed.csv"
NIWO.Litter<-read.csv("../Data/Processed/NEON_NIWO_Litter_mass_trap_Processed.csv", stringsAsFactors = T
```

```
#2
class(Peter.Paul.Lake$sampledate)
```

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

```
Peter.Paul.Lake$sampledate <-as.Date(Peter.Paul.Lake$sampledate)


class(NIWO.Litter$collectDate)
```

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

```
NIWO.Litter$collectDate <-as.Date(NIWO.Litter$collectDate)

#also converting month to a factor for the boxplots
Peter.Paul.Lake$month <-as.factor(Peter.Paul.Lake$month)
```

Define your theme

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

```
#3
mytheme <- theme_gray(base_size=12) +
  theme(axis.text = element_text(color="black"),
        legend.position = "right")
theme_set(mytheme)
```

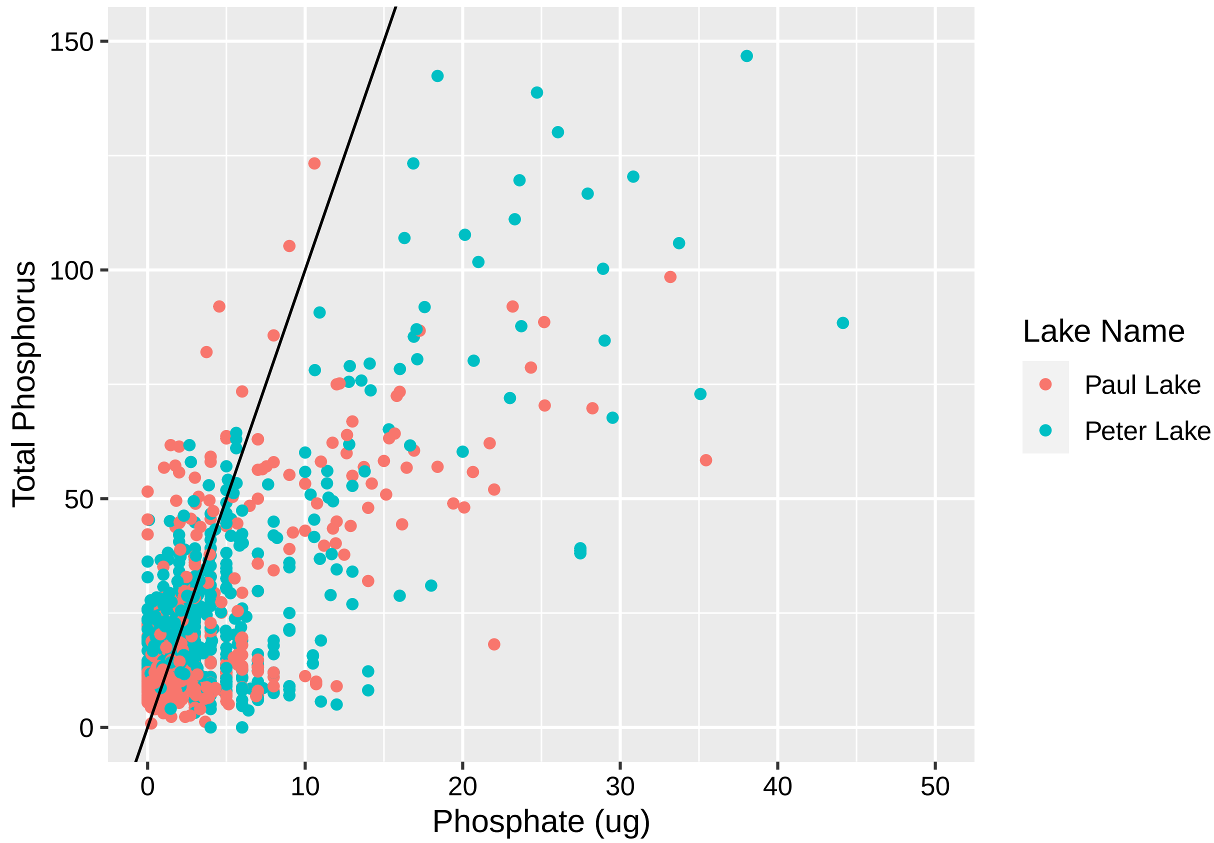
Create graphs

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()`).

```
#4
Phosphorus.by.phosphate <-
  ggplot(Peter.Paul.Lake) +
  geom_point(aes(x=po4, y=tp_ug, color=lakename)) +
  geom_abline(aes(slope=10, intercept=0)) +
  xlim(0,50) +
  ylim(0,150) +labs(color="Lake Name", x="Phosphate (ug)", y="Total Phosphorus")
print(Phosphorus.by.phosphate)
```

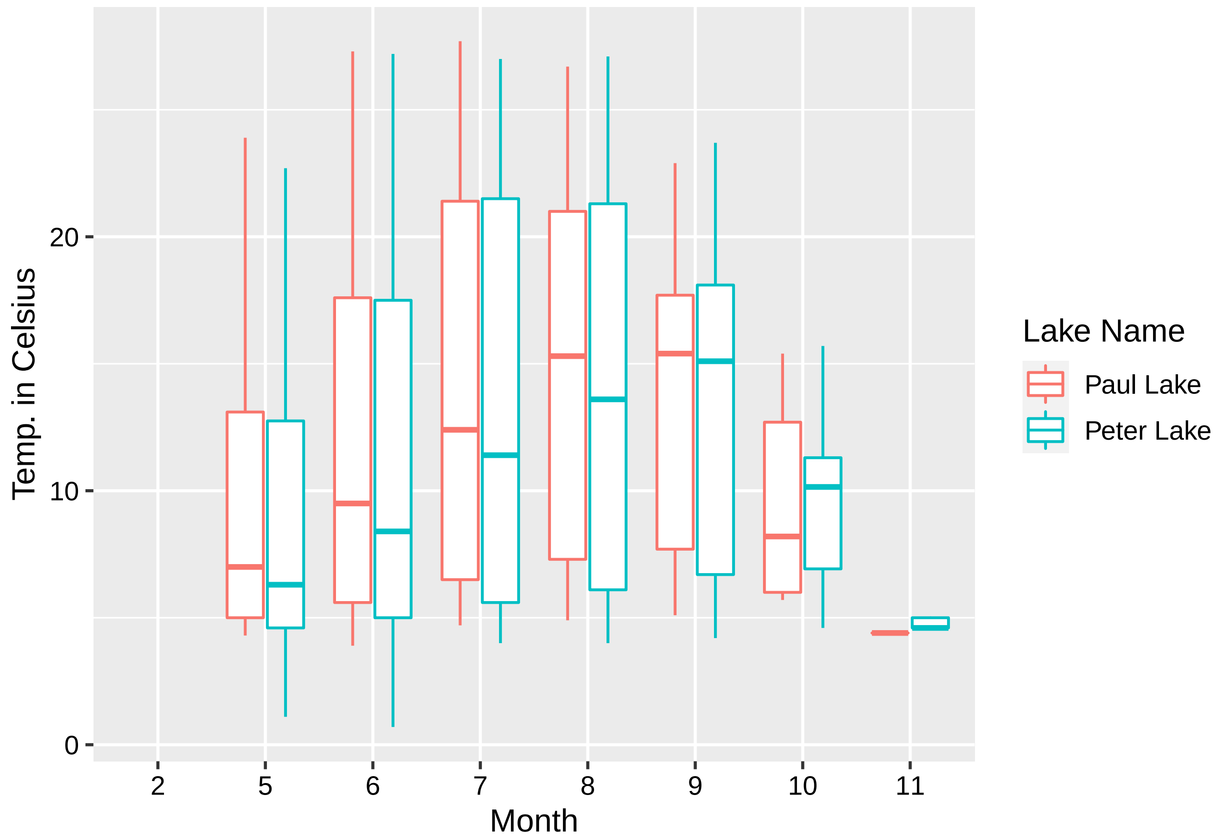
```
## Warning: Removed 21948 rows containing missing values (geom_point).
```



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.

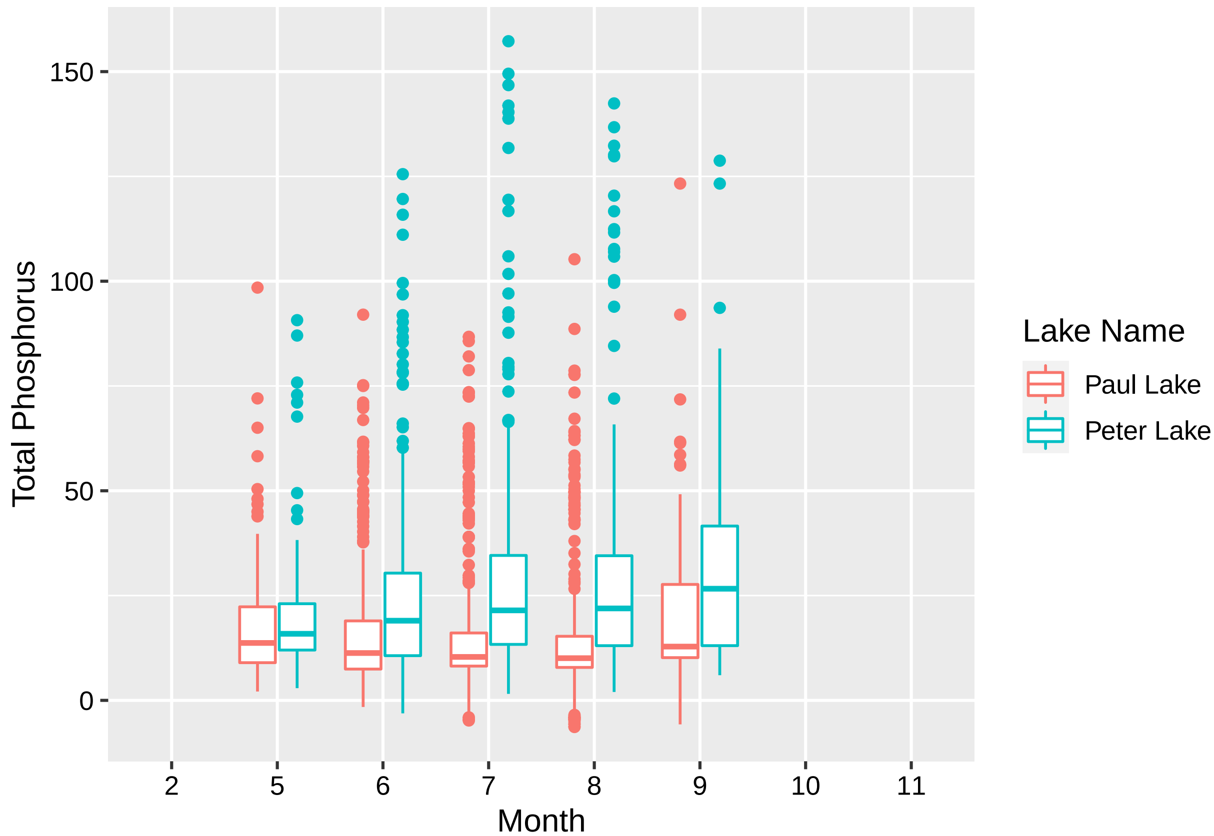
```
#5
plot.temp <- ggplot(Peter.Paul.Lake, aes(x=month, y=temperature_C)) +
  geom_boxplot(aes(color=lakename)) +
  ylab("Temp. in Celsius") +
  xlab("Month") +
labs(color = "Lake Name")
print(plot.temp)
```

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



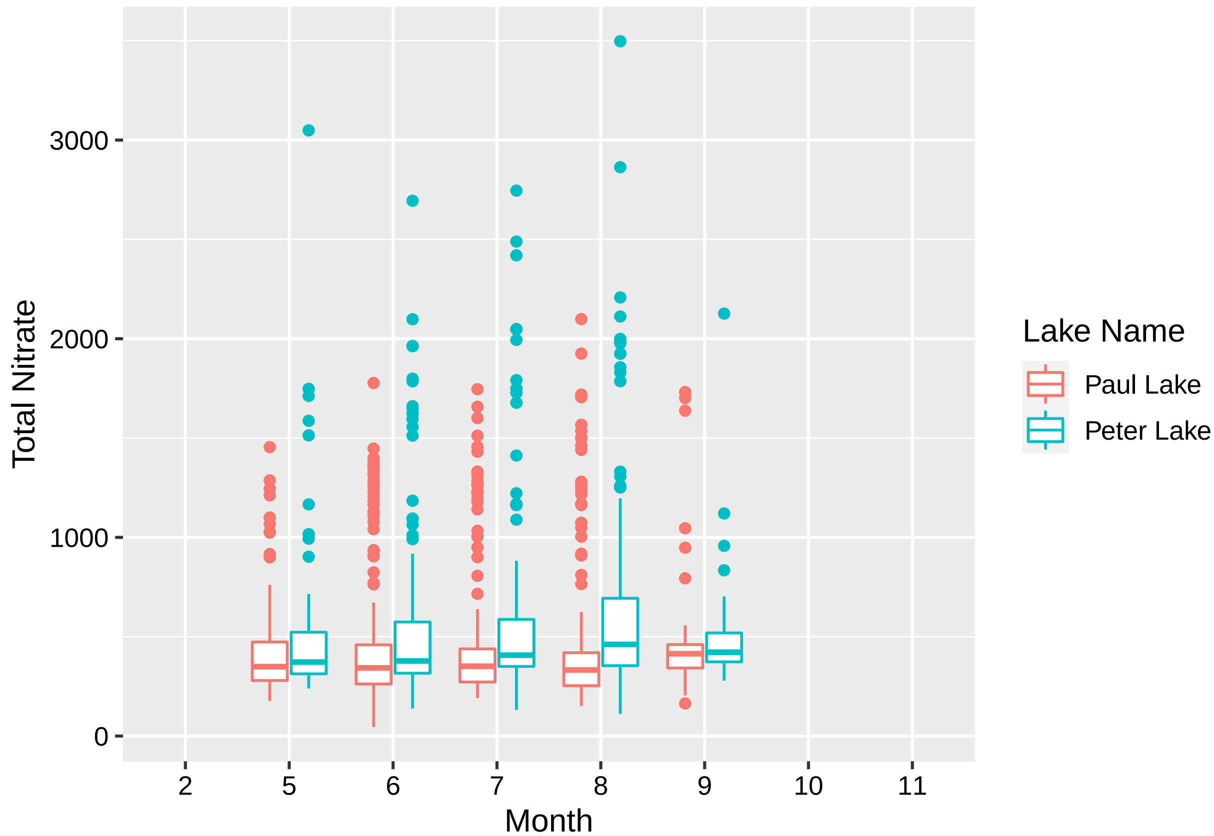

```
plot.TP <- ggplot(Peter.Paul.Lake, aes(x=month, y=tp_ug)) +
  geom_boxplot(aes(color=lakename)) +
  ylab("Total Phosphorus") +
  xlab("Month") +
  labs(color = "Lake Name")
print(plot.TP)
```

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




```
plot.TN <- ggplot(Peter.Paul.Lake, aes(x=month, y=tn_ug, color=lakename)) +geom_boxplot()  +
  xlab("Month") +
  ylab("Total Nitrate") +
  labs(color = "Lake Name")
print(plot.TN)
```

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



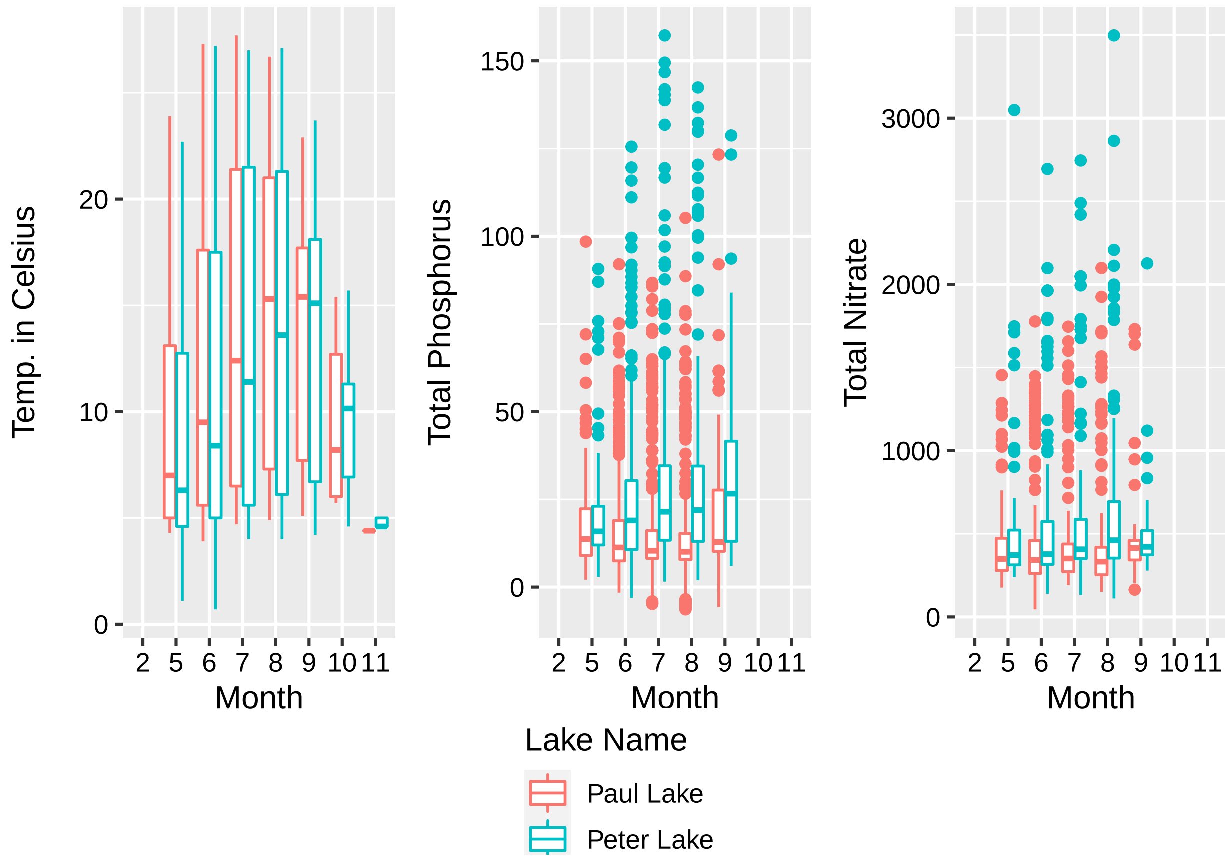

```
#Plotting each variable without the legend
plot.temp.axis <- plot.temp + theme(legend.position = "none")
plot.TP.axis <- plot.TP + theme(legend.position = "none")
plot.TN.axis <- plot.TN +theme(legend.position = "none")

#plot grid without legend
plot.no.legend <- plot_grid(plot.temp.axis, plot.TP.axis, plot.TN.axis, ncol=3, align ='v', axis ='tb',
```

```
## 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).
```

```
#Pulling legend from one of the plots
legend <-get_legend(plot.temp)
```

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

```
#Combining plots and 1 legend
plot.legend <- plot_grid(plot.no.legend, legend, nrow=2, rel_heights= c(3,0.5))
plot.legend
```

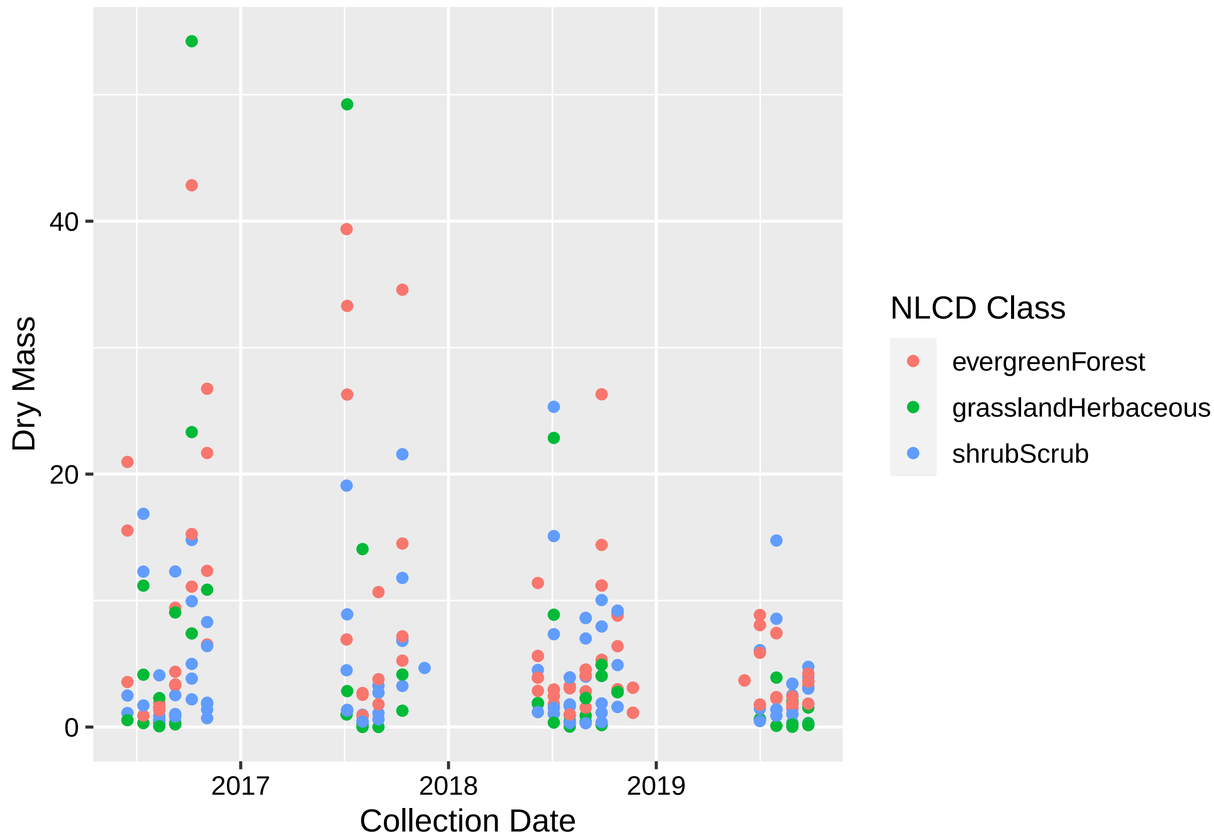


Question: What do you observe about the variables of interest over seasons and between lakes?

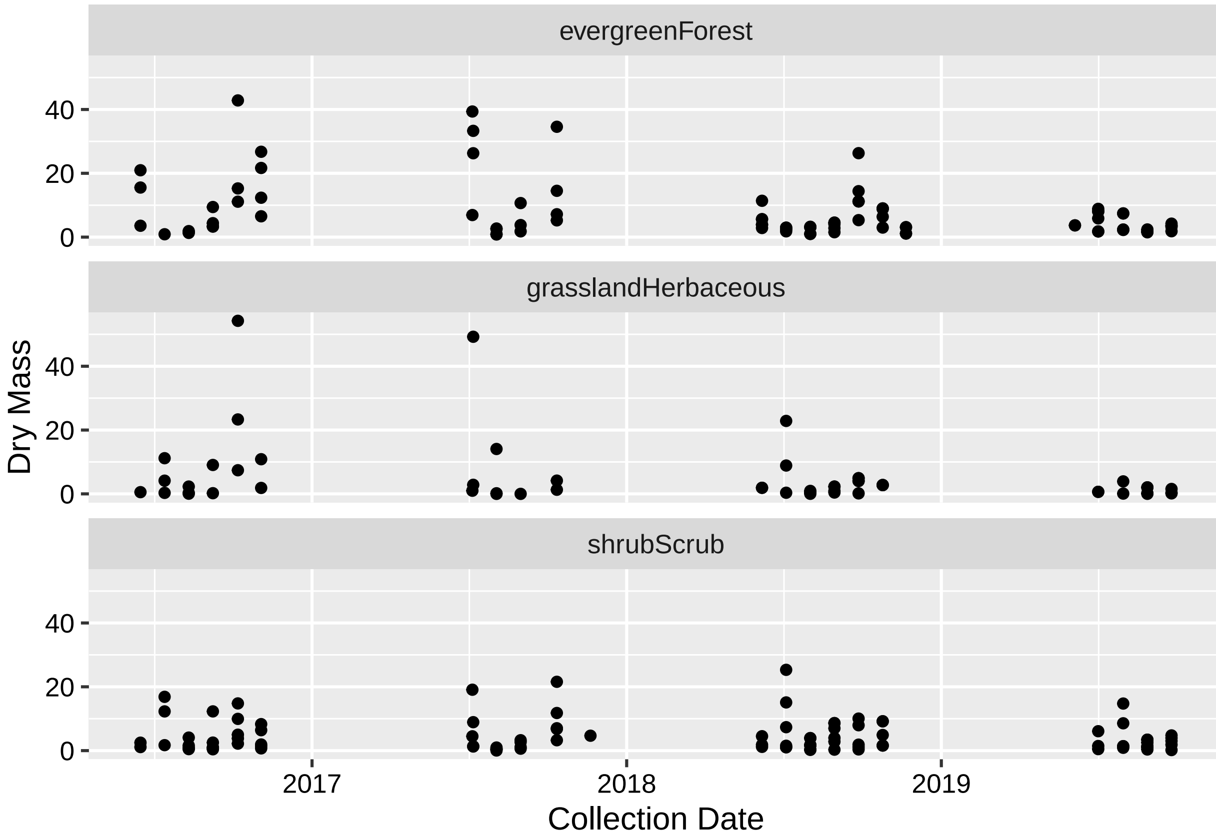
> Answer: From these observations, I observe that Paul Lake is slightly warmer than Peter Lake for the majority of the year (With the exception of November). The lakes also have high concentrations of total phosphorus in the summer months with Peter Lake containing higher concentrations of phosphorus than Paul Lake. In addition, the concentration of total nitrate is higher in the lakes during warmer summer months (ex. July and August) and Peter Lake's nitrates concentrations tend to run higher than those in 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

NIWOT.mass.colors <-
  ggplot(subset(NIWO.Litter, functionalGroup =="Needles"),
         aes(x = collectDate, y = dryMass)) +
  geom_point(aes(color= nlcdClass)) +
  labs(x = "Collection Date", y = "Dry Mass", color = "NLCD Class")
print(NIWOT.mass.colors)
```




```
#7
NIWOT.mass.facets <-
  ggplot(subset(NIWO.Litter, functionalGroup =="Needles"),
         aes(x=collectDate, y=dryMass)) +
            geom_point() +
  facet_wrap(vars(nlcdClass), nrow=3) +
  labs(x = "Collection Date", y = "Dry Mass")
print(NIWOT.mass.facets)
```



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

Answer: I think that plot 7 is more effective at displaying the differences in dry mass across NLCD classes. By separating the dry masses of needles based off of the NLCD classes we are better able to visualize the uniformity and outliers in dry mass values. From looking at plot 7, I can deduce that evergreen forests and herbaceous grasslands tend to have higher dry masses than shrubs. When the values are on the same plot such as in plot 6, the scatterplots are more likely to overlap so we can miss out on certain trends.