Assignment 5: Data Visualization

Cassidy White

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] "C:/Users/cassi/OneDrive - Duke University/Documents/School/Grad School/Spring 2022/Environmental
```

```
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)
library(ggplot2)
lake.chemistry<-read.csv("../Data/Processed/NTL-LTER_Lake_Chemistry_Nutrients_PeterPaul_Processed.csv",
Litter<-read.csv("../Data/Processed/NEON_NIWO_Litter_mass_trap_Processed.csv", stringsAsFactors = TRUE)
#2
```

```r
lake.chemistry$sampledate<-as.Date(lake.chemistry$sampledate)
lake.chemistry$month<-as.factor(lake.chemistry$month)
Litter$collectDate<-as.Date(Litter$collectDate)
```

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 = "right")
theme_set(mytheme)
```

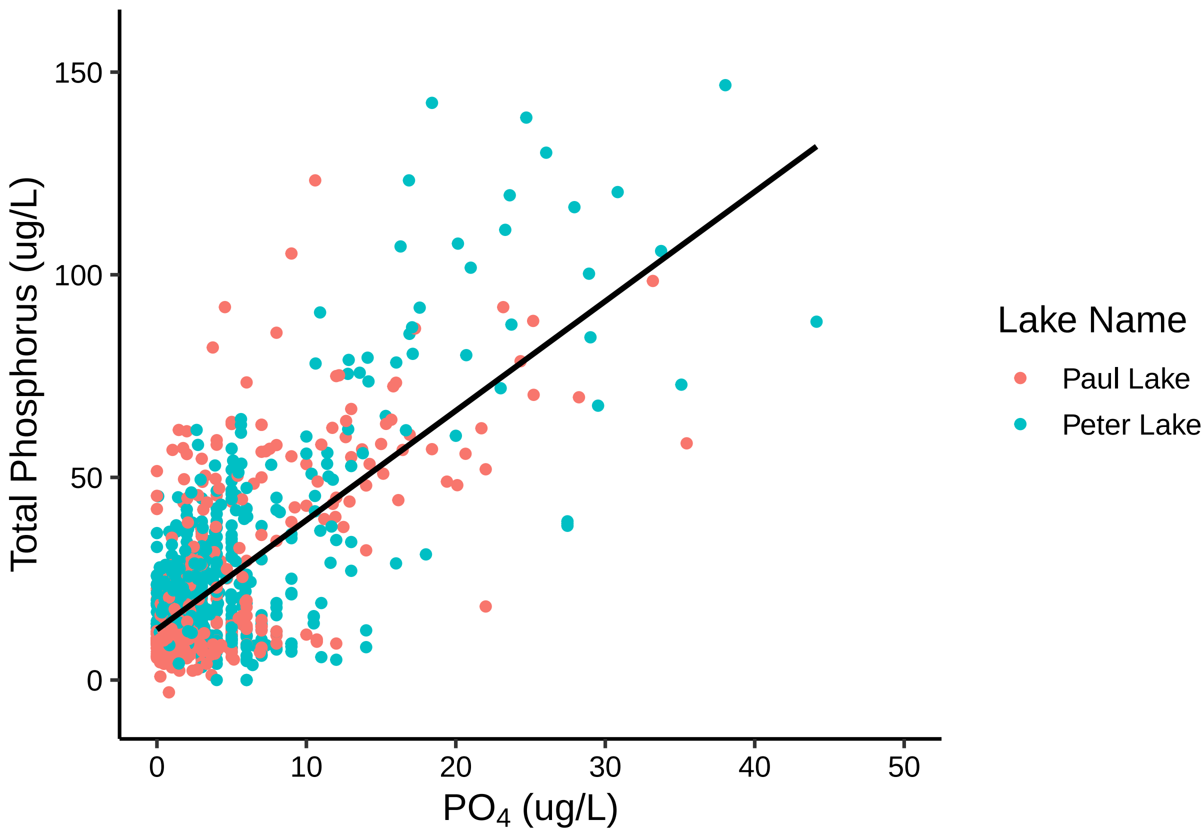
Create graphs

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()`).

```r
#4
ggplot(lake.chemistry, aes(x=po4, y=tp_ug)) +
  geom_point(aes(color = lakename))+
  xlab(expression("PO"[4]* " (ug/L)")) +
  ylab("Total Phosphorus (ug/L)") +
  labs(color = "Lake Name")+
  #facet_wrap(vars(lakename), nrow = 2) +
  xlim(0, 50)+
  geom_smooth(method = lm, se = FALSE, color = "black")
```

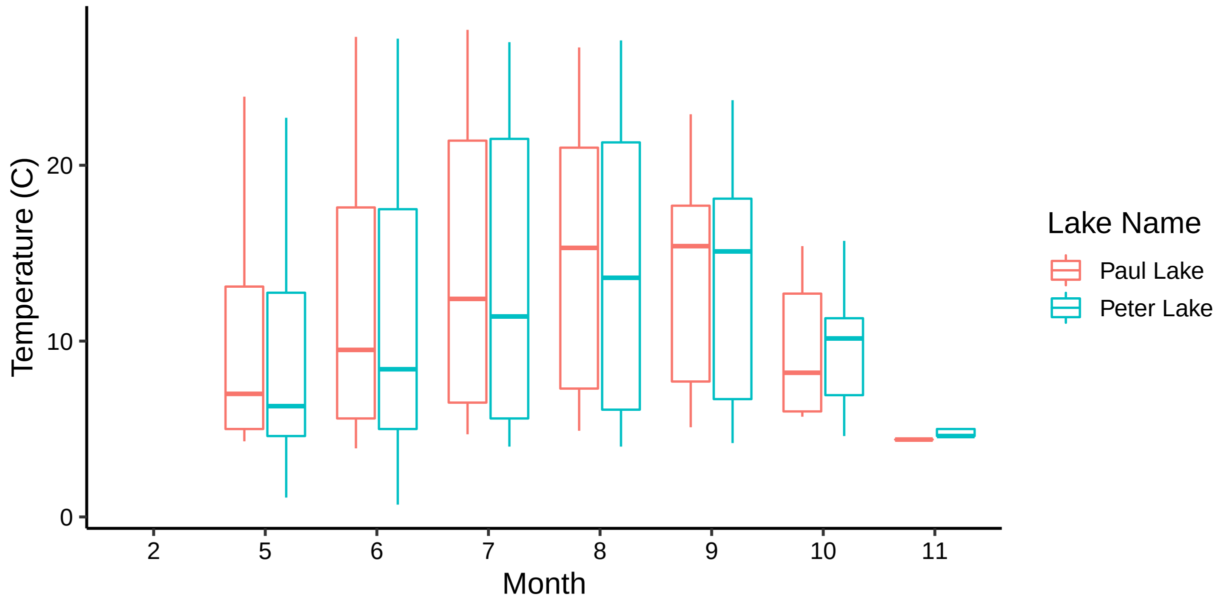
```
## `geom_smooth()` using formula 'y ~ x'

## Warning: Removed 21947 rows containing non-finite values (stat_smooth).

## Warning: Removed 21947 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.

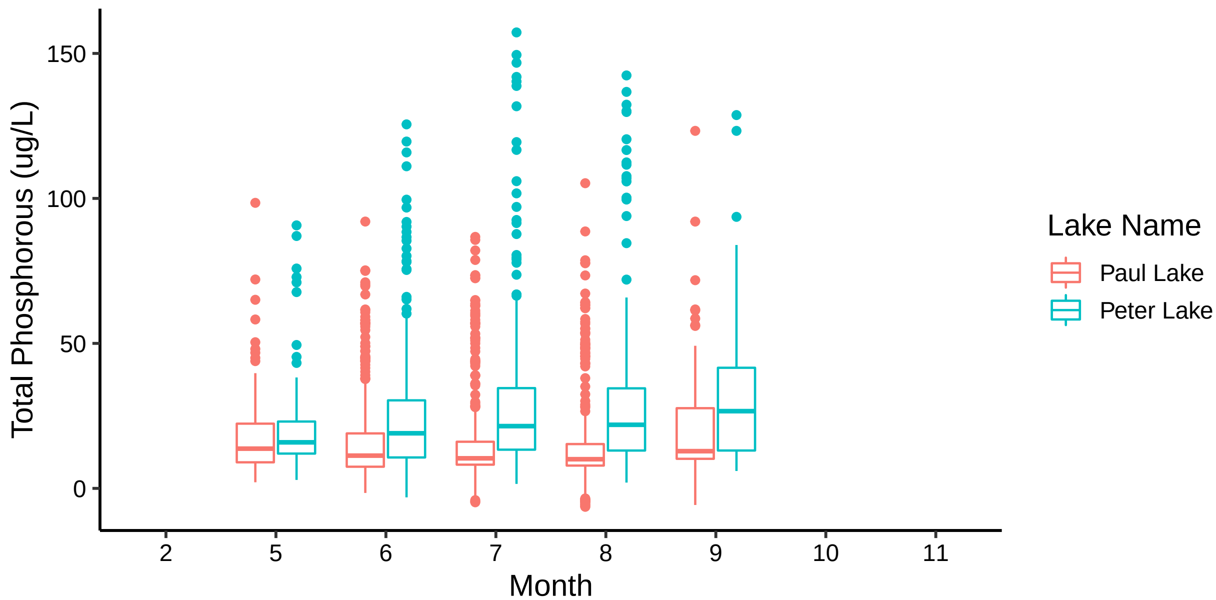
```
#5
Temp.box<-ggplot(lake.chemistry, aes(x=month, y = temperature_C))+
  geom_boxplot(aes(color = lakename))+
  labs(x="Month", y="Temperature (C)")+
  labs(color = "Lake Name")
print(Temp.box)
```

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



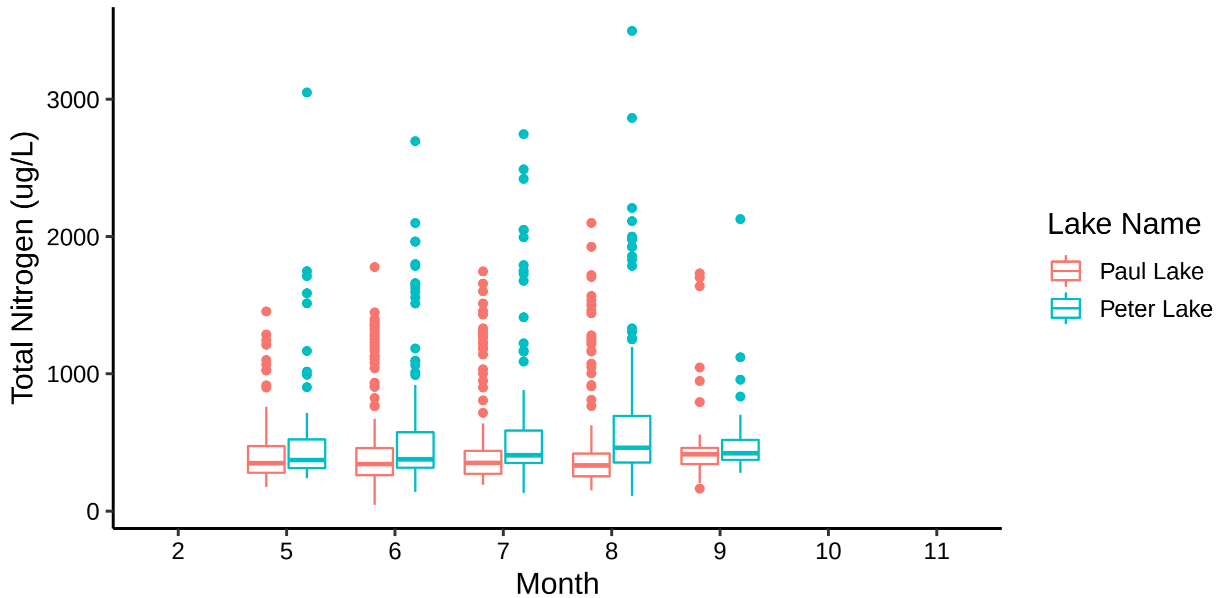

```
TP.box<-ggplot(lake.chemistry, aes(x=month, y = tp_ug))+
  geom_boxplot(aes(color = lakename))+
  labs(x="Month", y="Total Phosphorous (ug/L)")+
  labs(color = "Lake Name")
print(TP.box)
```

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




```
TN.box<-ggplot(lake.chemistry, aes(x=month, y = tn_ug))+
  geom_boxplot(aes(color = lakename))+
  labs(x="Month", y="Total Nitrogen (ug/L)")+
  labs(color = "Lake Name")
print(TN.box)
```

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

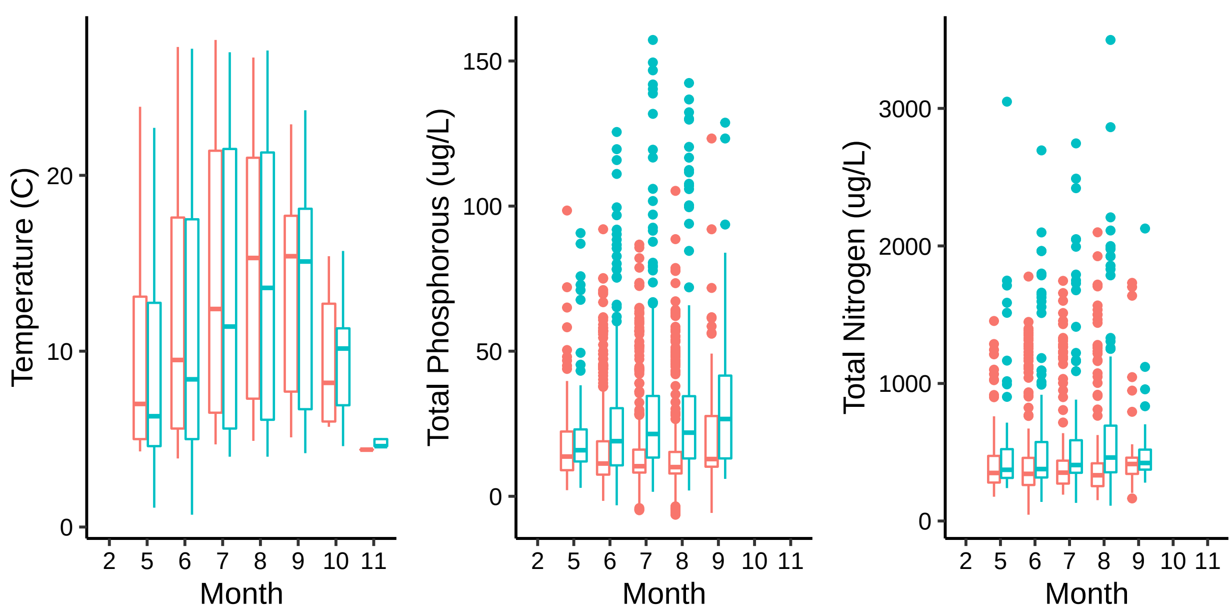



```r
Temp.box.nolegend<-Temp.box +theme(legend.position = "none")
TP.box.nolegend<-TP.box +theme(legend.position = "none")
TN.box.nolegend<-TN.box +theme(legend.position = "none")
```

```r
plot1<-plot_grid(Temp.box.nolegend, TP.box.nolegend, TN.box.nolegend, nrow =1, align = 'h', rel_heights
```

```
## 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).
```

```r
print(plot1)
```



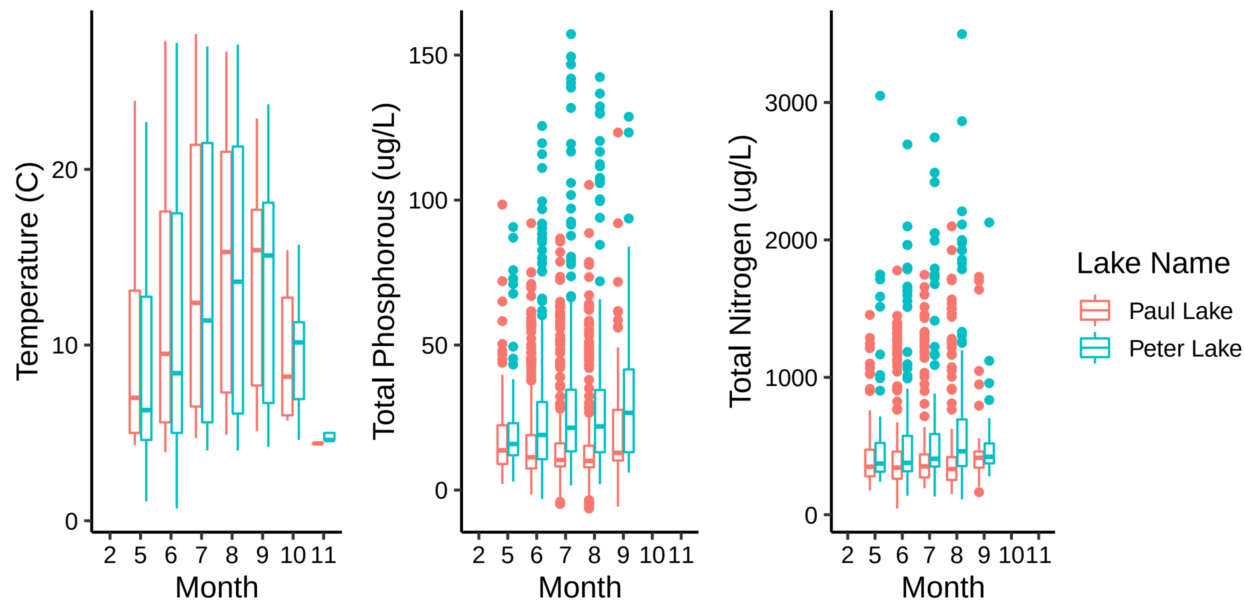
```
legend<-get_legend(Temp.box)
```

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

```
legend
```

```
## TableGrob (5 x 5) "guide-box": 2 grobs
##                                   z     cells                  name
## 99_3e8bde505edc50a9718d7cf1ebfefe6d 1 (3-3,3-3)                guides
##                                   0 (2-4,2-4) legend.box.background
##                                             grob
## 99_3e8bde505edc50a9718d7cf1ebfefe6d gtable[layout]
##                                   zeroGrob[NULL]
```

```
plot2<-plot_grid(plot1, legend, rel_widths = c(3,.5))
print(plot2)
```

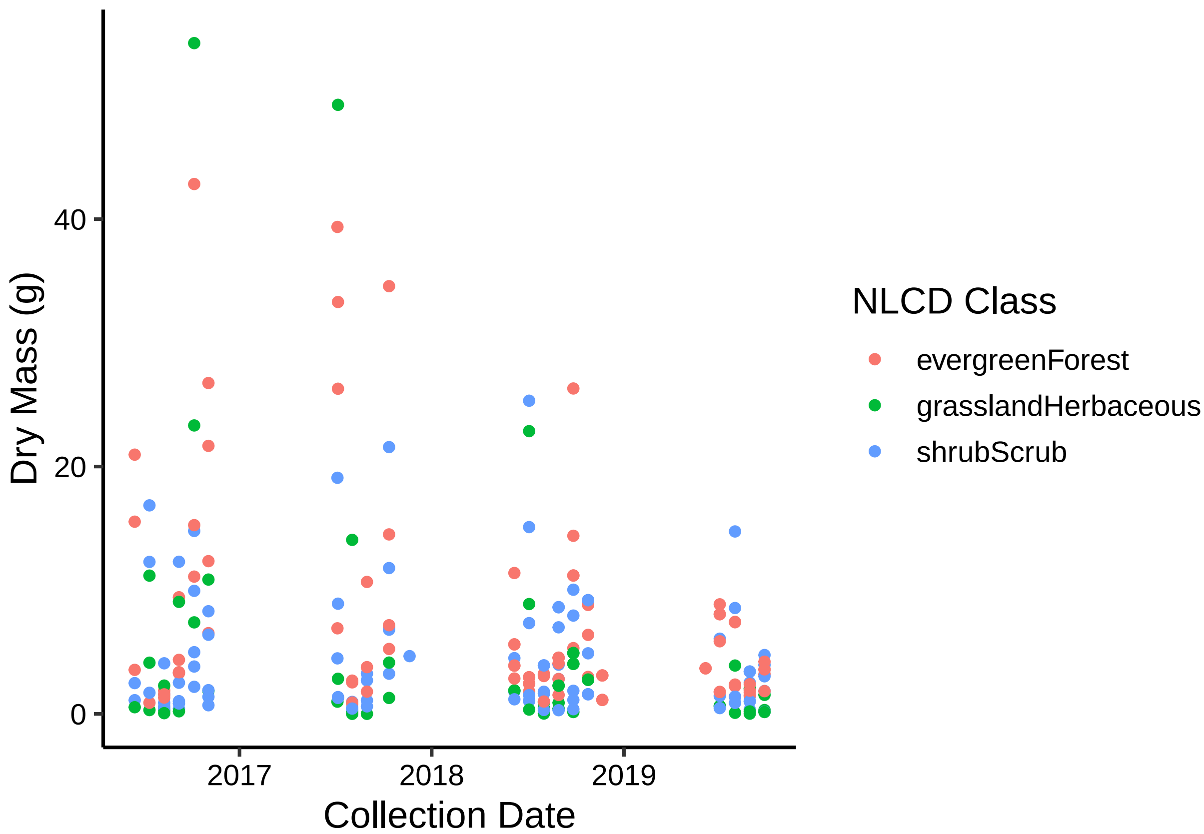


Question: What do you observe about the variables of interest over seasons and between lakes?

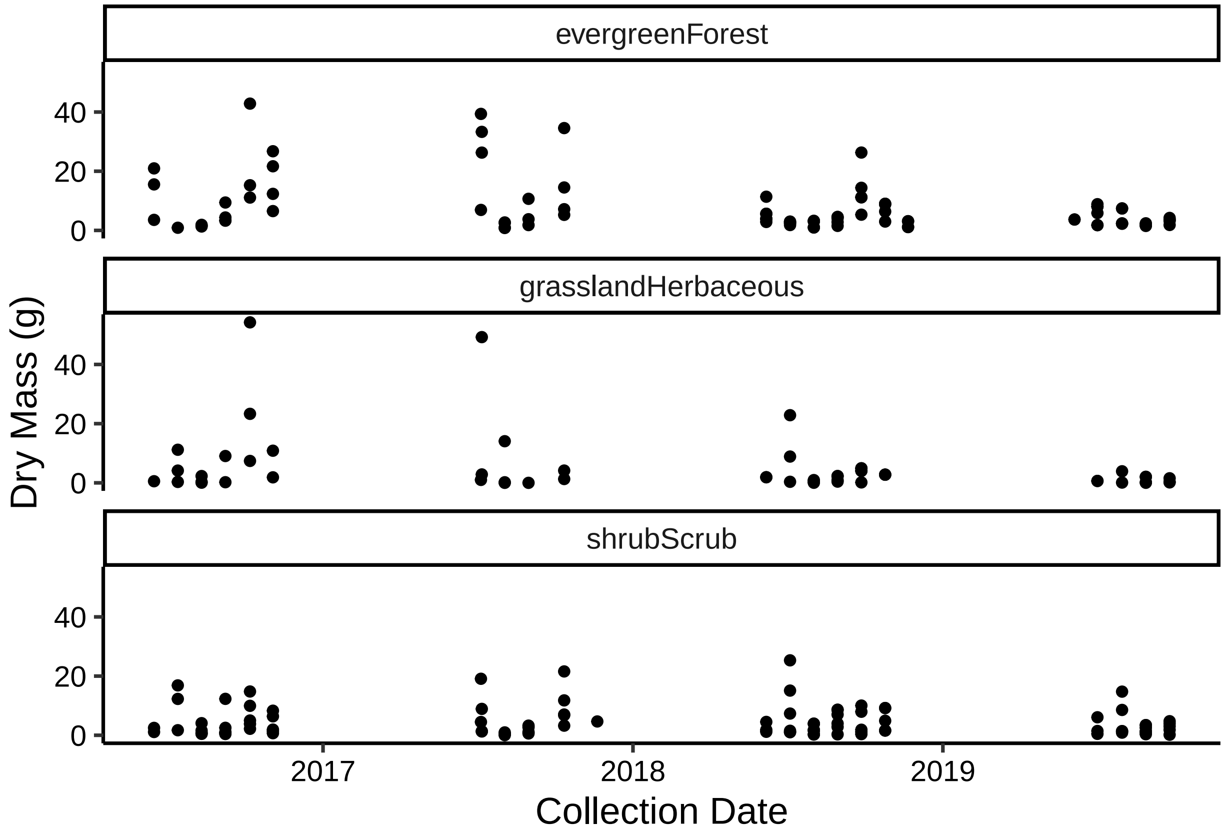
> Answer: Peter Lake has higher total phosphorous and total nitrogen compared to Paul Lake. For both lakes, temperature, total phosphorus, and total nitrogen are highest in the summer months.

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




```
#7
Litter.plot2<-ggplot(subset(Litter, functionalGroup == "Needles"), aes(x=collectDate, y=dryMass))+
  geom_point()+
  labs(x="Collection Date", y= "Dry Mass (g)", color = "NLCD Class")+
  facet_wrap(vars(nlcdClass), nrow = 3)
print(Litter.plot2)
```



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

Answer: I think the plot for Q6 is more effective because it combines everything in one graph. This is easier on the eye for comparison so you're not looking back and forth.