Assignment 5: Data Visualization

Pierre Mishra

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., "Salk_A05_DataVisualization.Rmd") prior to submission.

The completed exercise is due on Tuesday, February 11 at 1:00 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 (tidy and gathered) and the processed data file for the Niwot Ridge litter dataset.

2. Make sure R is reading dates as date format; if not change the format to date.

```
#1
getwd()
```

```
## [1] "C:/Users/Peaceful Pierre/Documents/Academics/Spring 2020/Environmental Data Analytics/Environmer
```

```
#setwd("Academics/Spring 2020/Environmental Data Analytics/Environmental_Data_Analytics_2020/Assignment
library("tidyverse")
library("cowplot")
peterpaul.chem.nutrients <-
  read.csv("../Data/Processed/NTL-LTER_Lake_Chemistry_Nutrients_PeterPaul_Processed.csv")
peterpaul.chem.nutrients.gathered <-
  read.csv("../Data/Processed/NTL-LTER_Lake_Nutrients_PeterPaulGathered_Processed.csv")
litter <-
  read.csv("../Data/Processed/NEON_NIWO_Litter_mass_trap_Processed.csv")

#2
peterpaul.chem.nutrients$sampledate <-
  as.Date(peterpaul.chem.nutrients$sampledate, format = "%Y-%m-%d")
```

```
peterpaul.chem.nutrients.gathered$sampledate <-
  as.Date(peterpaul.chem.nutrients.gathered$sampledate, format = "%Y-%m-%d")
litter$collectDate <- as.Date(litter$collectDate,format = "%Y-%m-%d")
```

Define your theme

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

```
peaceful.theme <- theme_classic(base_size = 14) +
  theme(axis.text = element_text(color = "black"),
        legend.position = "right")

theme_set(peaceful.theme)
```

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.

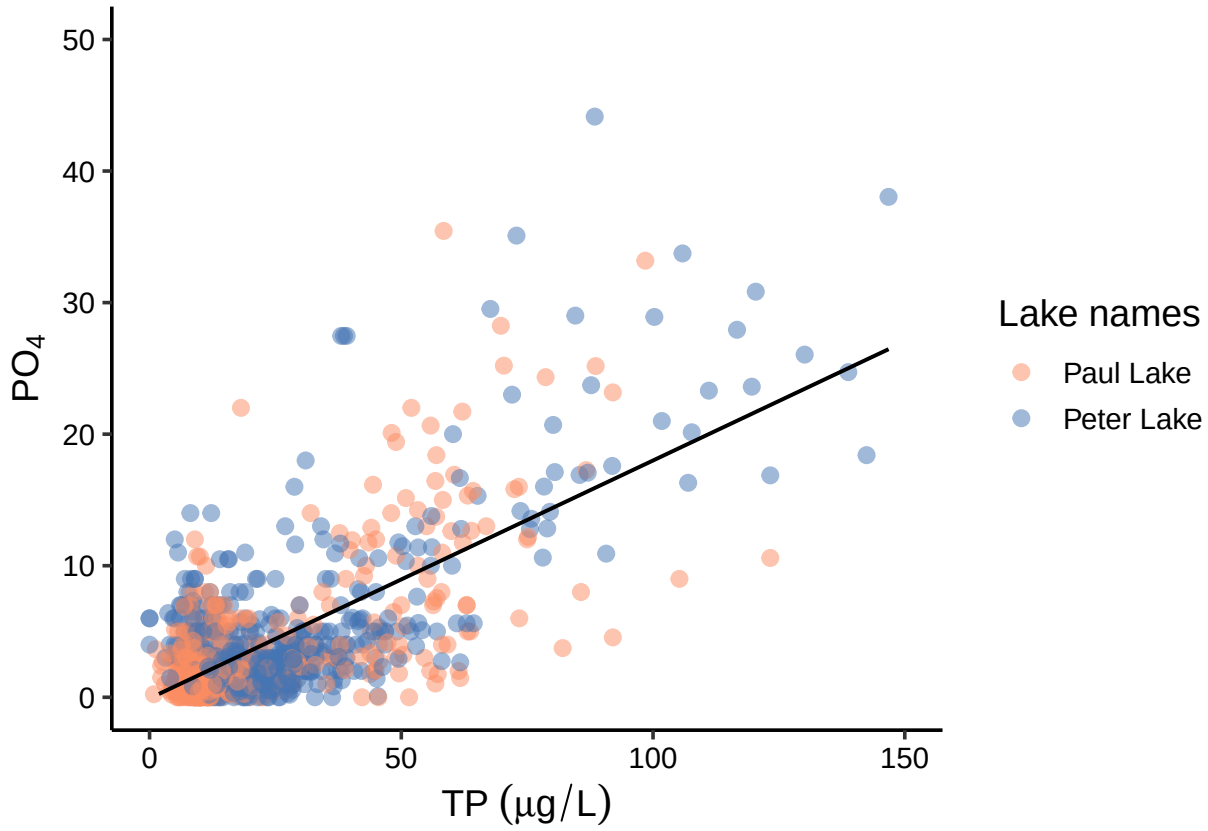
4. [NTL-LTER] Plot total phosphorus by phosphate, with separate aesthetics for Peter and Paul lakes. Add a line of best fit and color it black. Adjust your axes to hide extreme values.

```
ggplot(peterpaul.chem.nutrients, aes(x = tp_ug, y = po4, color = lakename)) +
  geom_point(alpha = 0.5, size = 2.5) +
  geom_smooth(aes(x = tp_ug, y = po4), method = lm,
              se = FALSE, inherit.aes = FALSE,
              col = "black", size = 0.7) +
  ylim (c(0,50)) +
  xlim (c(0,150)) +
  labs(x = expression(TP ~ (mu*g / L)),
       y = expression(PO[4]),
       color = "Lake names") +
  scale_color_manual(values=c("#fc8d62", "#4575b4"))
```

```
## Warning: Removed 21948 rows containing non-finite values (stat_smooth).
```

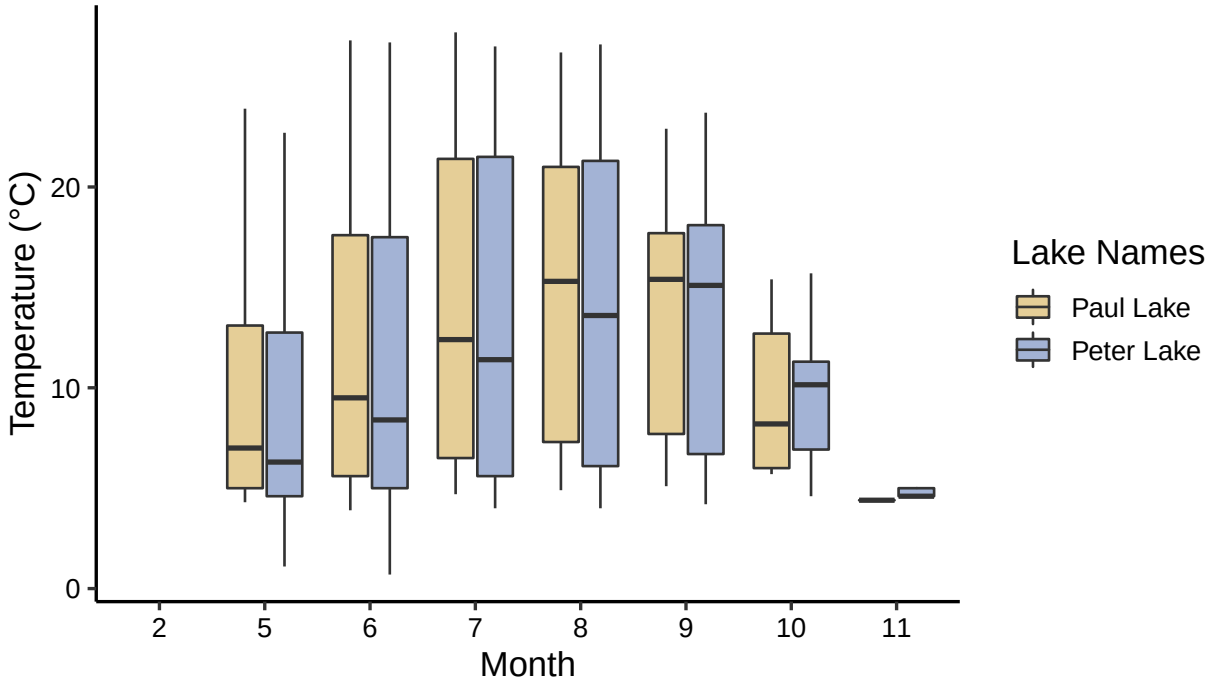
```
## Warning: Removed 21948 rows containing missing values (geom_point).
```

```
## Warning: Removed 1 rows containing missing values (geom_smooth).
```

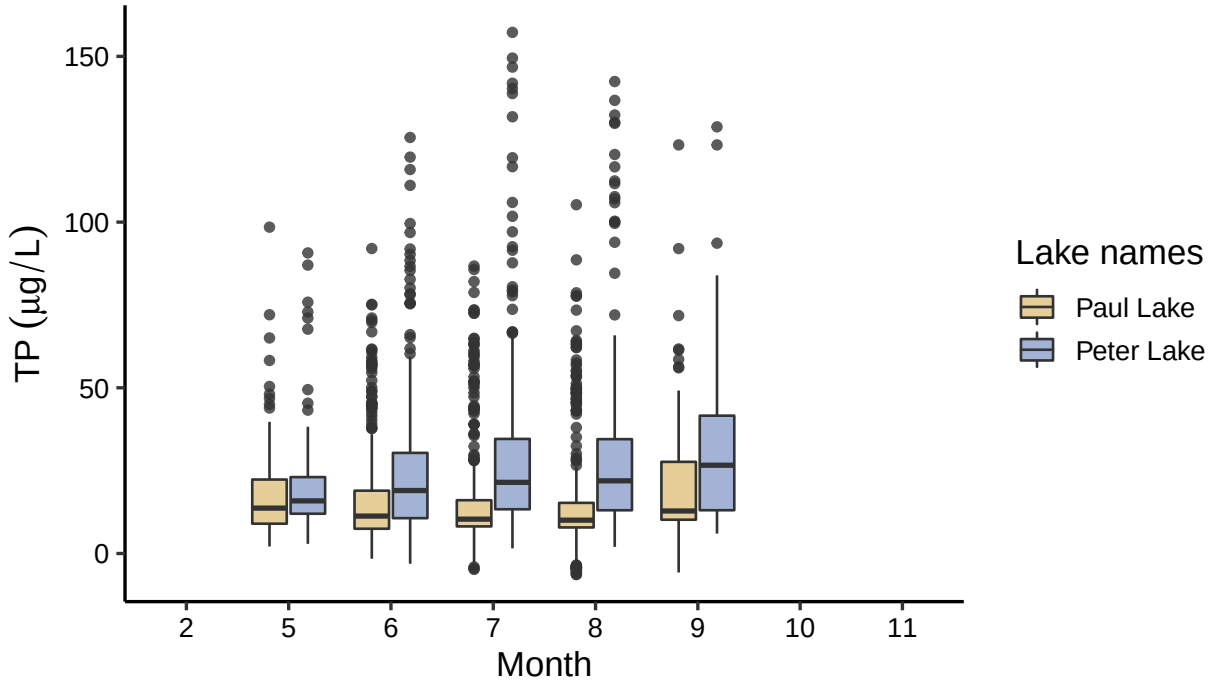


5. [NTL-LTER] Make three separate boxplots of (a) temperature, (b) TP, and (c) TN, with month as the x axis and lake as a color aesthetic. Then, create a cowplot that combines the three graphs. Make sure that only one legend is present and that graph axes are aligned.

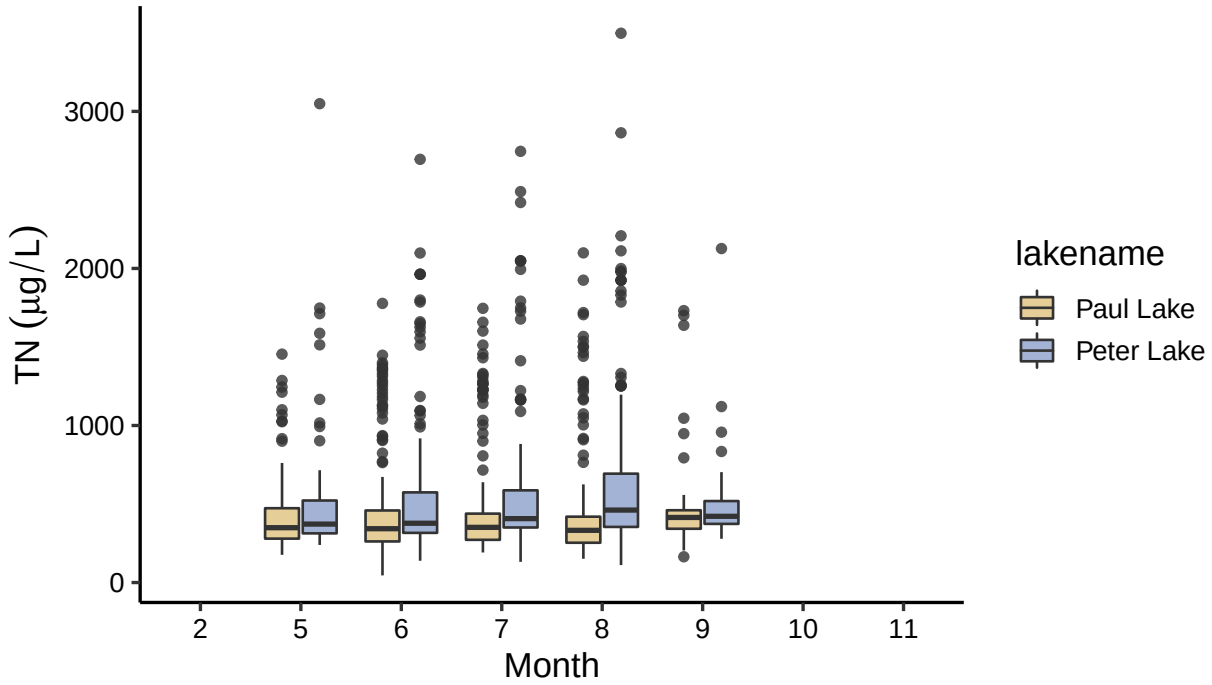
```
# (a)
temp <- ggplot(peterpaul.chem.nutrients, aes(x = as.factor(month), y = temperature_C)) +
         geom_boxplot(aes(fill = lakename), alpha = 0.8) +
         scale_fill_manual(values = c("#dfc27d", "#8da0cb")) +
         labs(x = "Month",
              y = "Temperature (°C)",
              fill = "Lake Names")
print(temp)
```



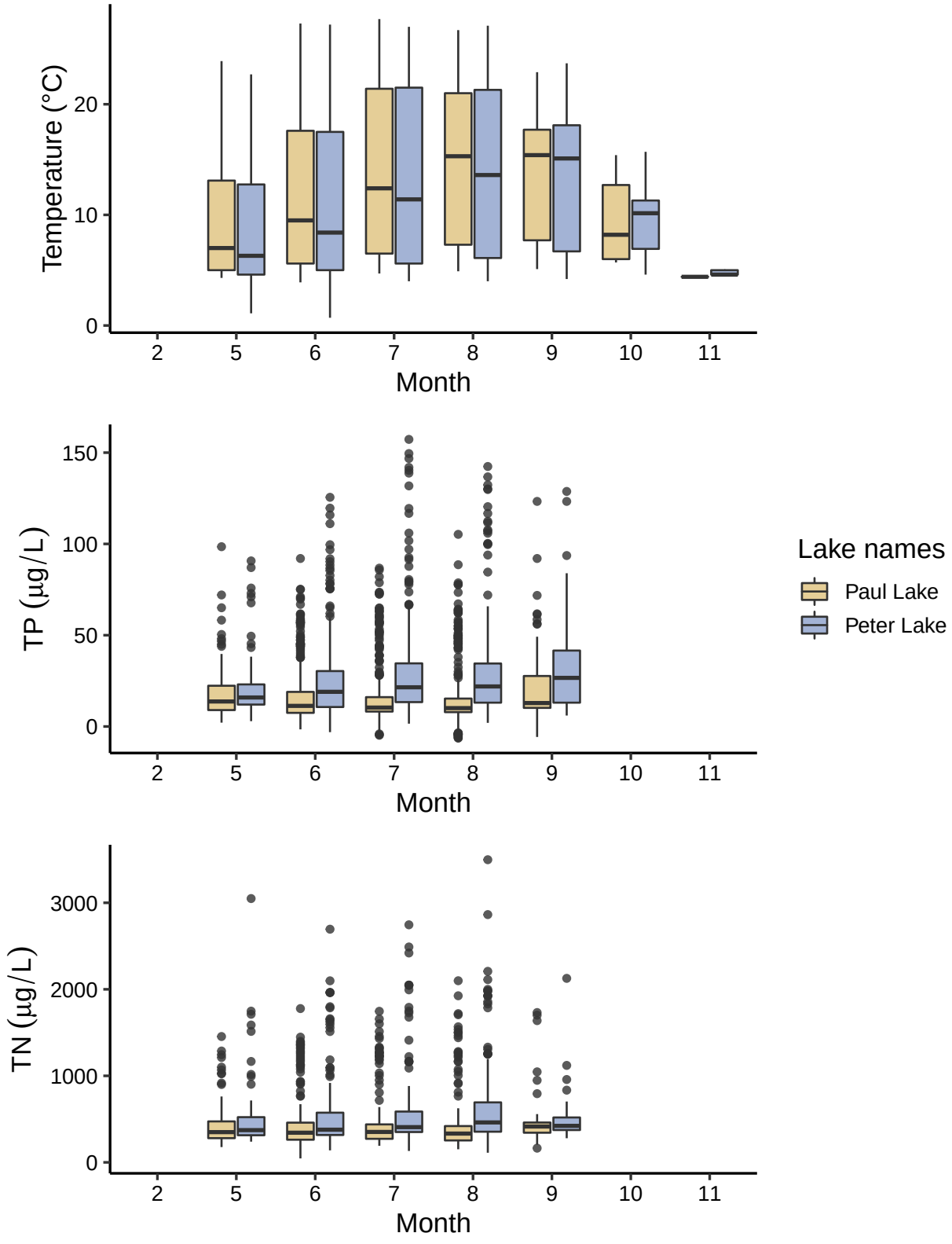

```
#(b)
phosphorus <- ggplot(peterpaul.chem.nutrients, aes(x = as.factor(month), y = tp_ug)) +
  geom_boxplot(aes(fill = lakename), alpha = 0.8) +
  scale_fill_manual(values = c("#dfc27d", "#8da0cb")) +
  labs(x = "Month",
       y = expression(TP ~ (mu*g / L)),
       fill = "Lake names")
print(phosphorus)
```




```
#(c)
nitrogen <- ggplot(peterpaul.chem.nutrients, aes(x = as.factor(month), y = tn_ug)) +
  geom_boxplot(aes(fill = lakename), alpha = 0.8) +
  scale_fill_manual(values = c("#dfc27d", "#8da0cb")) +
  labs(x = "Month",
       y = expression(TN ~ (mu*g / L)))
print(nitrogen)
```



```r
#(d)
plot_grid(temp + theme(legend.position="none"),
          phosphorus, nitrogen + theme(legend.position="none"),
          nrow = 3, axis = 'lr', align = 'v')
```



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

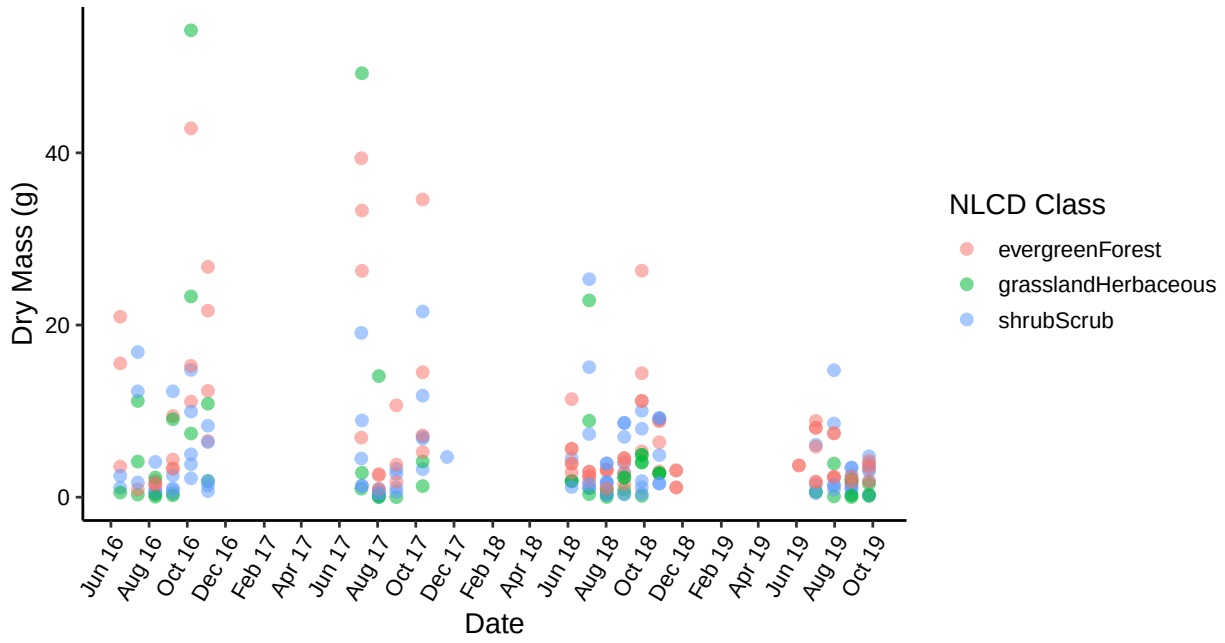
Answer: Temperatures of both the lakes increase during the summer time as expected. Till

August, the median temperature for Paul Lake is higher than that of Peter Lake. However, after September, the median temperature of Paul Lake falls faster than that of Peter Lake. After May, the median total phosphorus in Peter Lake is higher than that in Paul lake and the difference is increasing due to a slight increase in phosphorus in Peter Lake. In Fall, Paul Lake shows the highest increase in its median total phosphorus levels compared to its earlier phosphorus levels which were slighlty decreasing from Spring to Summer. Overall, PeterLake shows higher variation in its total phosphorus levels throughout the year compared to that of Paul Lake and is increasing from Spring to Fall compared to its earlier values. Total nitrogen also seems to be higher in Peter Lake where its 75 percentile values are consistently higher than those of Paul Lake. However, in September, the interquatile range of total nitrogen in Peter Lake decreased compared to its previous values.

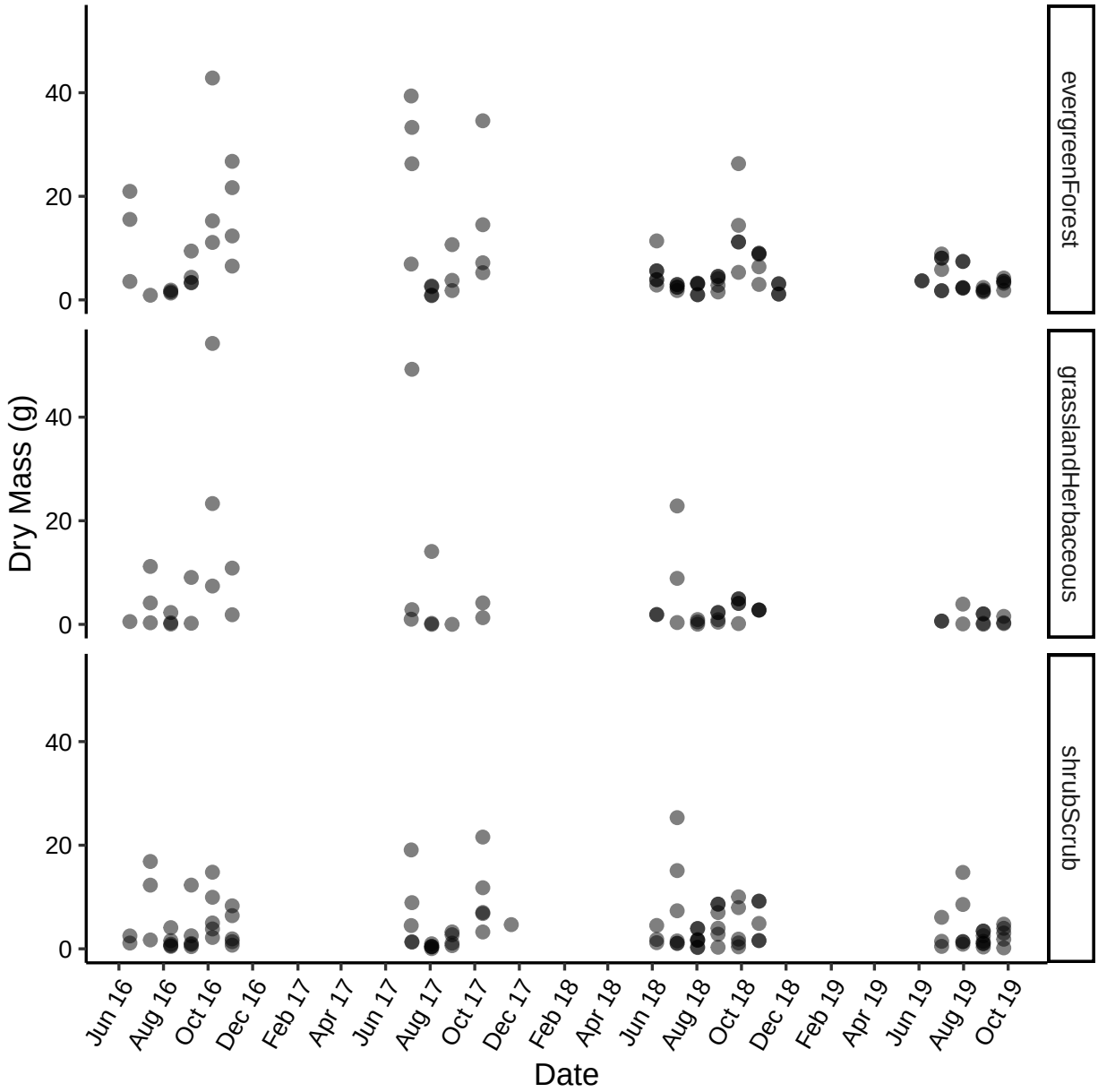
6. [Niwot Ridge] Plot a subset of the litter dataset by displaying only the "Needles" functional group. Plot the dry mass of needle litter by date and separate by NLCD class with a color aesthetic. (no need to adjust the name of each land use)

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, functionalGroup == "Needles"),
       aes(x = collectDate, y = dryMass, color = nlcdClass)) +
  geom_point(size = 2.5, alpha = 0.55) +
  scale_x_date(limits = as.Date(c("2016-06-16", "2019-09-25")),
    date_breaks = "2 months", date_labels = "%b %y") +
  theme(axis.text.x = element_text(angle = 60,  hjust = 1)) +
  labs(x = "Date",
       y = "Dry Mass (g)",
       fill = "Lake names",
       color = "NLCD Class")
```



```r
#7.
ggplot(subset(litter, functionalGroup == "Needles"),
       aes(x = collectDate, y = dryMass)) +
  geom_point(size = 2.5, alpha = 0.5) +
  scale_x_date(limits = as.Date(c("2016-06-16", "2019-09-25")),
    date_breaks = "2 months", date_labels = "%b %y") +
  theme(axis.text.x = element_text(angle = 60,  hjust = 1)) +
  labs(x = "Date",
       y = "Dry Mass (g)",
       fill = "Lake names",
       color = "NLCD Class") +
  facet_grid(nlcdClass ~ .)
```



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

Answer: I think the plot in 7 is more effective than the one in 6 because one can clearly distinguish among the dry mass of the all the NLCD classes. There is no overlap of values. In plot 6, although dry mass of each NLCD class had different colors to distinguish, a lot of them overlap each other at low values of dry mass even after setting transparency in the points, thereby, creating confusion for the reader.