Assignment 5: Data Visualization

Kendra Sultzer

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 Tuesday, February 23 at 11:59 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 (both the tidy [`NTL-LTER_Lake_Chemistry_Nutrients_PeterPaul_Processed.csv`] and the gathered [`NTL-LTER_Lake_Nutrients_PeterPaulGathered_Processed.csv`] versions) 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/kendr/Documents/Spring2021Classes/ENV872/Environmental_Data_Analytics_2021"
```

```
library(tidyverse)
```

```
## -- Attaching packages --------------------------------------- tidyverse 1.3.0 --

## v ggplot2 3.3.3     v purrr   0.3.4
## v tibble  3.0.6     v dplyr   1.0.3
## v tidyr   1.1.2     v stringr 1.4.0
## v readr   1.4.0     v forcats 0.5.1

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

```
library(cowplot)
LakeChem.Nutrients.PP <- read.csv("./Data/Processed/NTL-LTER_Lake_Chemistry_Nutrients_PeterPaul_Processe
Lake.Nutrients.PP.Gathered <- read.csv("./Data/Processed/NTL-LTER_Lake_Nutrients_PeterPaulGathered_Proce
Niwot.Litter <- read.csv("./Data/Processed/NEON_NIWO_Litter_mass_trap_Processed.csv")

#2
class(Lake.Nutrients.PP.Gathered$sampledate) #was character, not date
```

```
## [1] "character"
```

```
Lake.Nutrients.PP.Gathered$sampledate <- as.Date(Lake.Nutrients.PP.Gathered$sampledate,
                                                   format = "%Y-%m-%d")
class(Lake.Nutrients.PP.Gathered$sampledate) #now reading as date
```

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

```
class(LakeChem.Nutrients.PP$sampledate) #was character, not date
```

```
## [1] "character"
```

```
LakeChem.Nutrients.PP$sampledate <- as.Date(LakeChem.Nutrients.PP$sampledate,
                                              format = "%Y-%m-%d")
class(LakeChem.Nutrients.PP$sampledate) #now reading as date
```

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

```
class(Niwot.Litter$collectDate) #was character, not date
```

```
## [1] "character"
```

```
Niwot.Litter$collectDate <- as.Date(Niwot.Litter$collectDate, format="%Y-%m-%d")
class(Niwot.Litter$collectDate) #now defined as date
```

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

Define your theme

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

```
mytheme <- theme_classic(base_size = 12) +
  theme(axis.text = element_text(color = "black"),
        legend.position = "right",
        legend.box.background = element_rect(color="black"), legend.background = element_blank(),
        plot.title=element_text(hjust = 0.5))
#makes font size bigger, puts legend on right, puts box around legend, centers title in middle

theme_set(mytheme) #setting for all subsequent plots
```

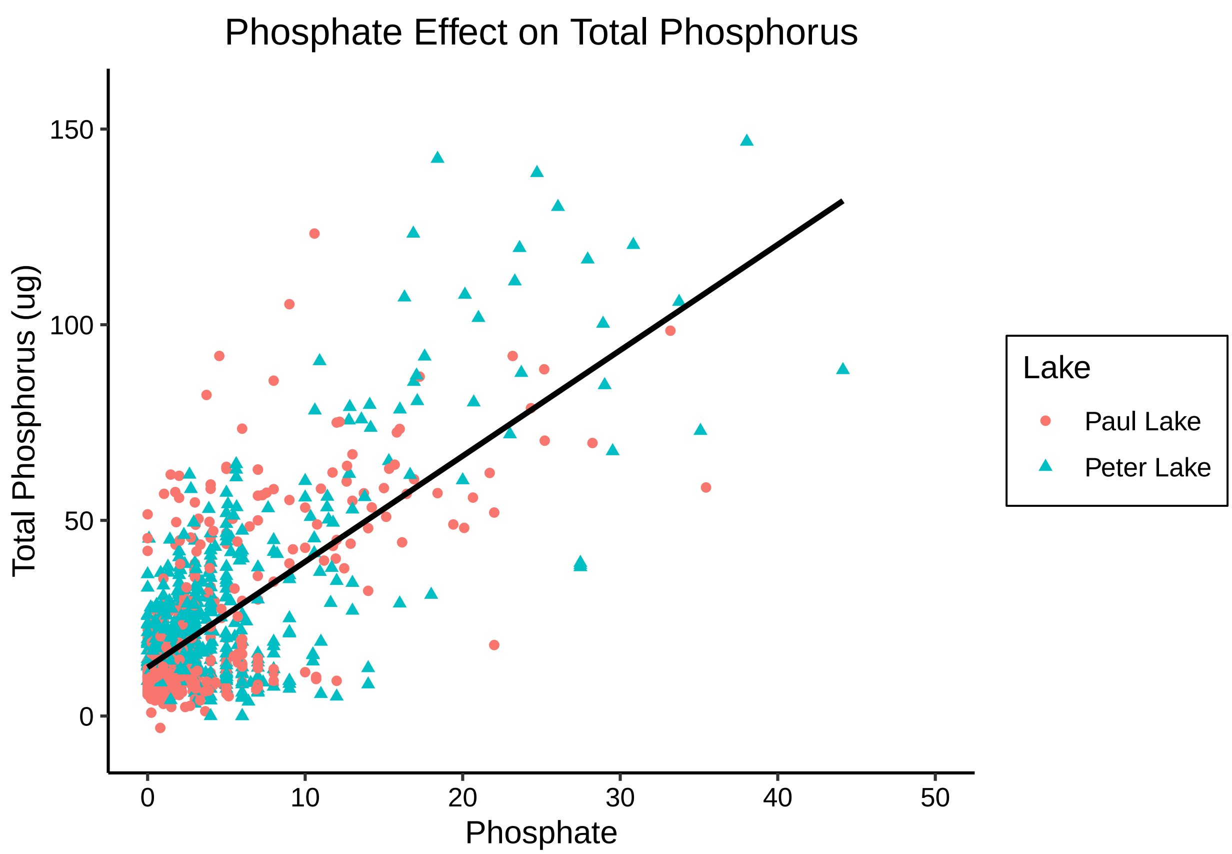
Create graphs

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.

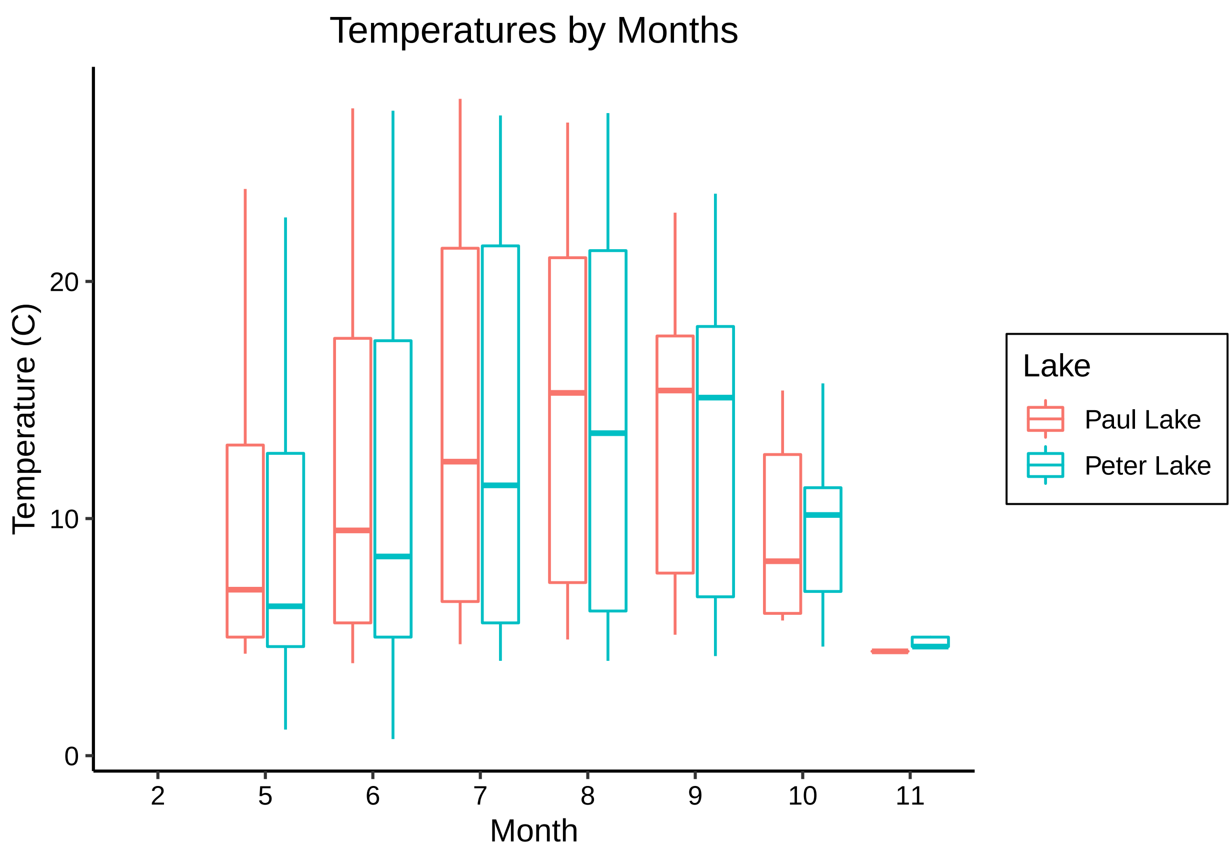
```
Plot1 <- ggplot(LakeChem.Nutrients.PP, aes(x=po4, y=tp_ug))+
  geom_point(aes(color=lakename, shape=lakename))+
  xlim(0,50)+
  labs(title="Phosphate Effect on Total Phosphorus",x="Phosphate", y="Total Phosphorus (ug)",
       color="Lake", shape="Lake")+
  geom_smooth(method=lm, se=FALSE, color="black")
print(Plot1)
```

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

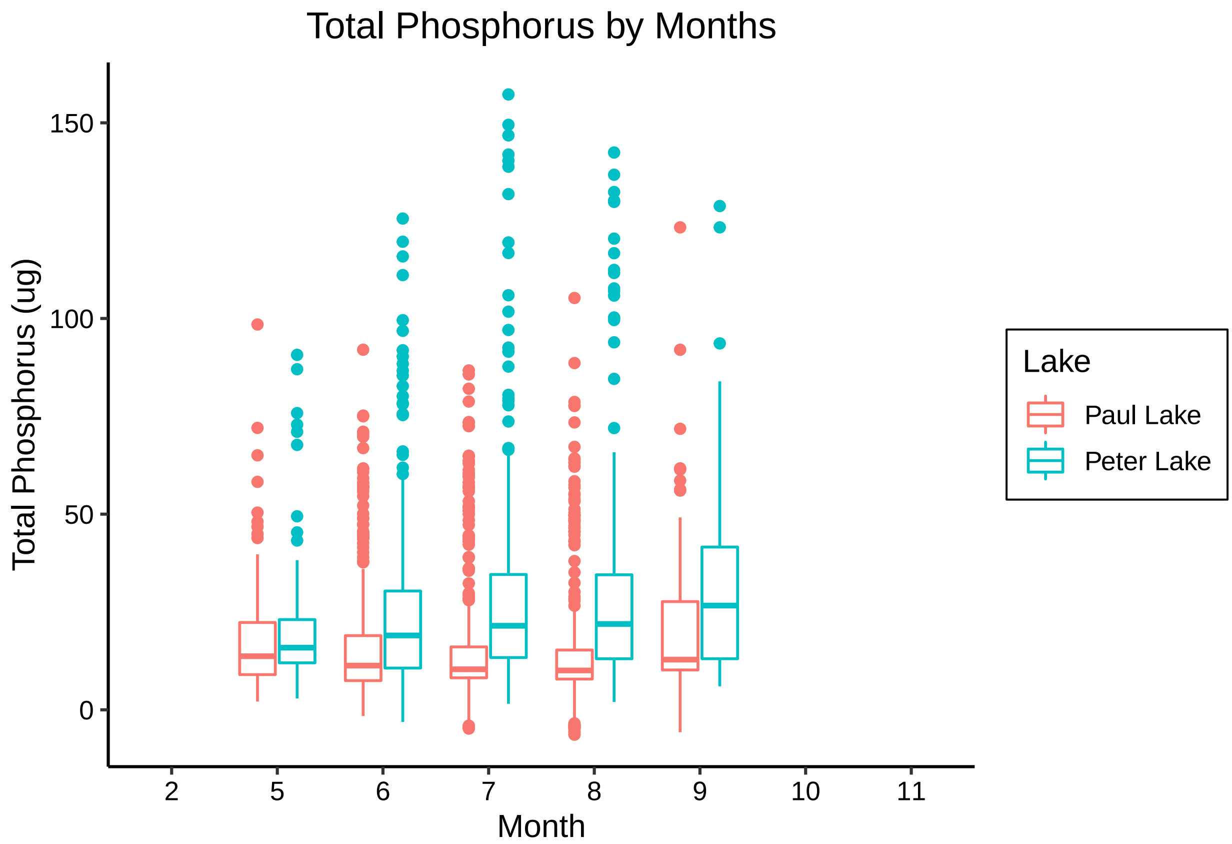


5. [NTL-LTER] Make three separate boxplots of (a) temperature, (b) TP, and (c) TN, with month as the x axis and lake as a color aesthetic. Then, create a cowplot that combines the three graphs. Make sure that only one legend is present and that graph axes are aligned.

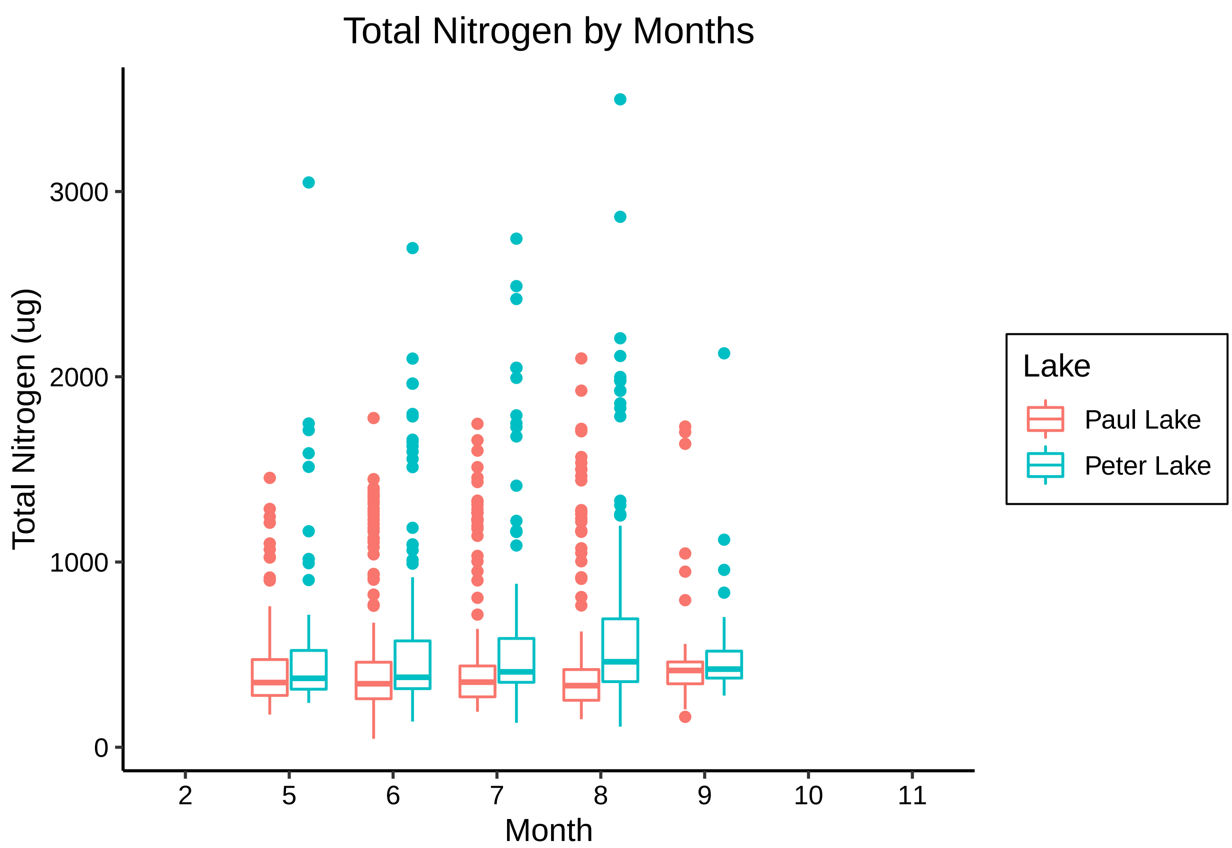
```r
Plot2 <- ggplot(LakeChem.Nutrients.PP)+
  geom_boxplot(aes(x=as.factor(month), y=temperature_C, color=lakename))+
  labs(title="Temperatures by Months", x="Month", y="Temperature (C)",color="Lake")
print(Plot2)
```



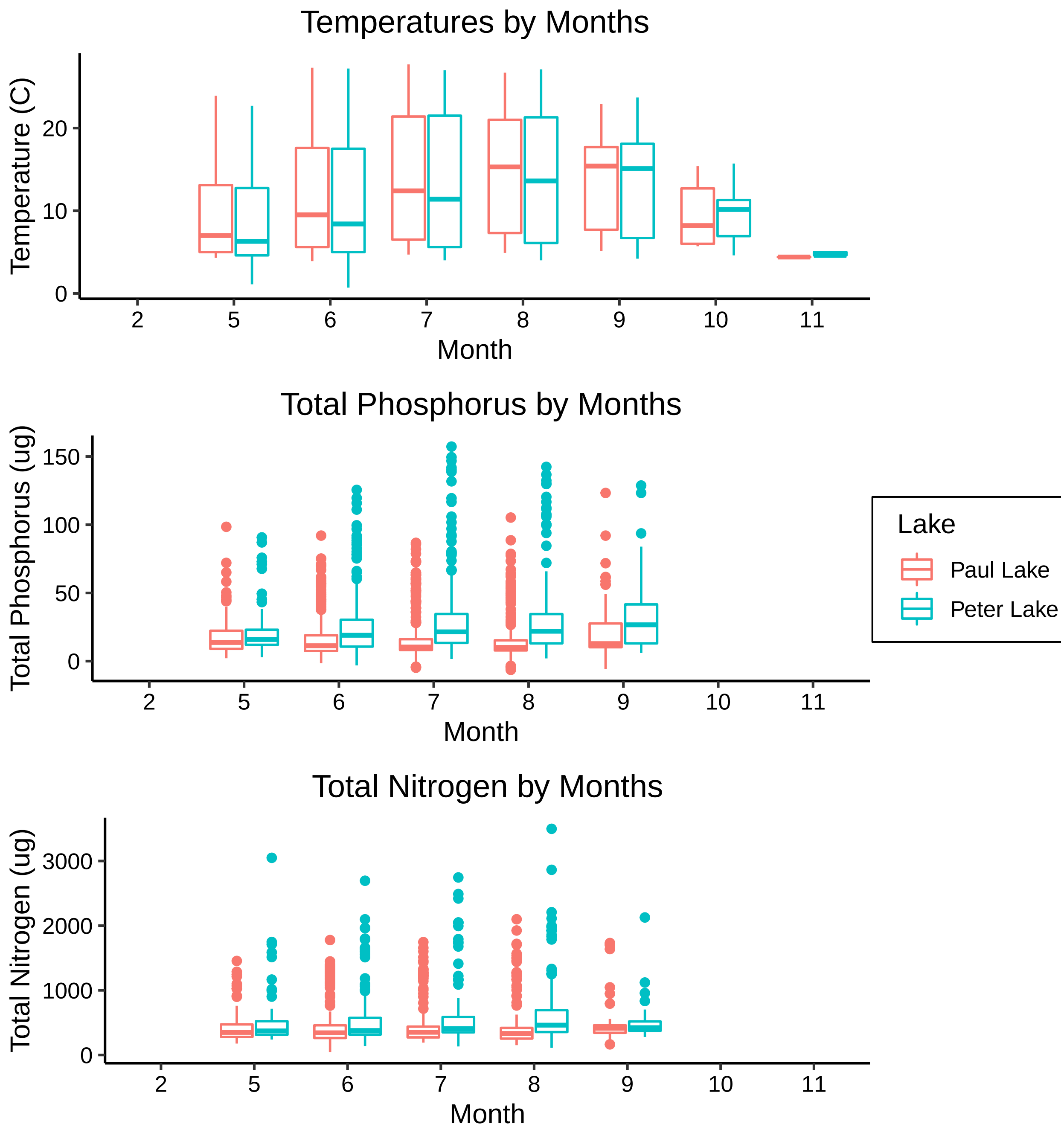

```r
Plot3 <- ggplot(LakeChem.Nutrients.PP)+
  geom_boxplot(aes(x=as.factor(month), y=tp_ug, color=lakename))+
  labs(title="Total Phosphorus by Months", x="Month", y="Total Phosphorus (ug)",color="Lake")
print(Plot3)
```




```
Plot4 <- ggplot(LakeChem.Nutrients.PP)+
  geom_boxplot(aes(x=as.factor(month), y=tn_ug, color=lakename))+
  labs(title="Total Nitrogen by Months", x="Month", y="Total Nitrogen (ug)",color="Lake")
print(Plot4)
```




```
Cowplot1 <- plot_grid(
  Plot2+theme(legend.position = "none"),
  Plot3+theme(legend.position = "none"),
  Plot4+theme(legend.position = "none"),
  nrow=3
)
legend <- get_legend(Plot3+theme(legend.box.margin = margin(0,0,0,5)))
#creating legend component to add back into boxplot
Cowplot2 <- plot_grid(Cowplot1, legend, rel_widths = c(2,0.4))
print(Cowplot2)
```

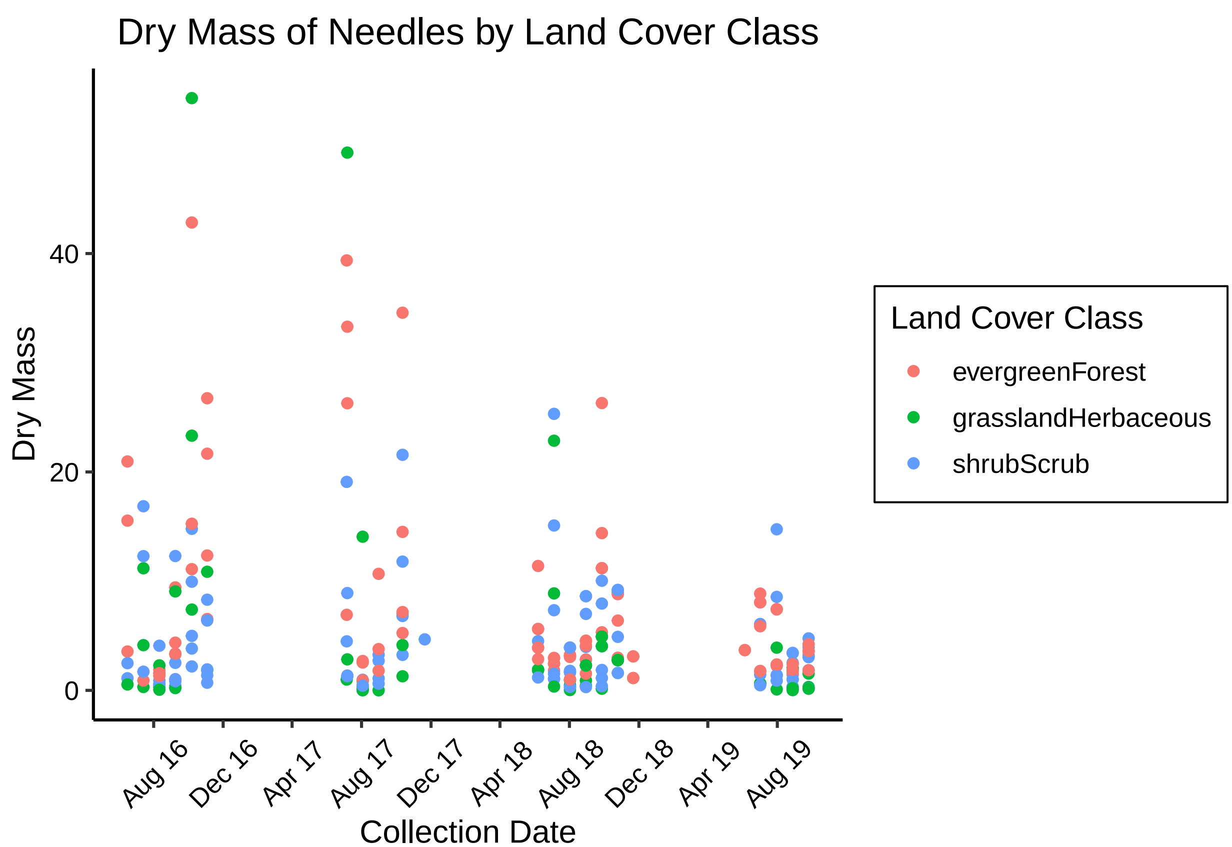


Question: What do you observe about the variables of interest over seasons and between lakes?

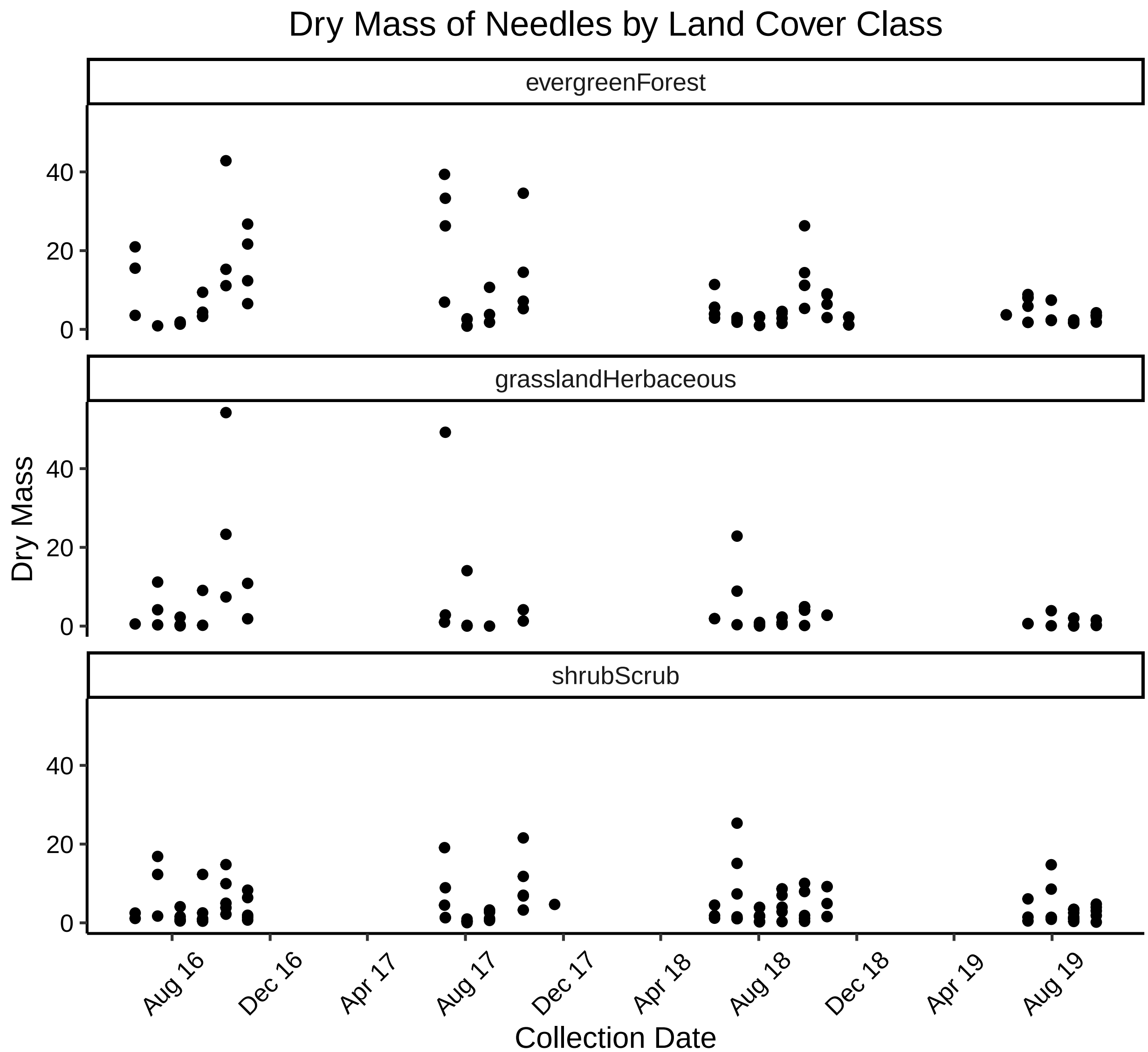
> Answer: The amounts of nitrogen don't vary drastically over the seasons, but Peter Lake appears to always be higher in total nitrogen than Paul Lake (although not very much in most months). There are also a large amount of outliers in total nitrogen and phosphorus throughout the months (maybe from nutrient loading in the water). Peter Lake also has higher total phosphorus amounts than Paul Lake, and the differences between total phosphorus amounts appear to increase over the year. In contrast to these other trends, it appears that Peter lake usually has lower temperatures than Paul Lake by month, and the temperatures in general fluctuate by season as one would expect.

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
Needles.Plot <- ggplot(subset(Niwot.Litter, functionalGroup=="Needles"))+
  geom_point(aes(x=collectDate, y=dryMass, color=nlcdClass))+
  labs(title="Dry Mass of Needles by Land Cover Class", x="Collection Date", y = "Dry Mass",
       color="Land Cover Class")+
  scale_x_date(date_breaks = "4 months", date_labels = "%b %y")+
  theme(axis.text.x = element_text(angle=45,vjust=0.5))
print(Needles.Plot)
```




```
#7
Needles.Plot2 <- ggplot(subset(Niwot.Litter, functionalGroup == "Needles"))+
  geom_point(aes(x=collectDate, y=dryMass))+
  labs(title="Dry Mass of Needles by Land Cover Class", x="Collection Date", y = "Dry Mass")+
  facet_wrap(vars(nlcdClass), nrow = 3) +
 scale_x_date(date_breaks = "4 months", date_labels = "%b %y")+
  theme(axis.text.x = element_text(angle=45,vjust=0.5))
print(Needles.Plot2)
```



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

> Answer: Plot #6 is a bit messy because it's hard to distinguish trends within the land cover classes- the colored points just all overlap. Plot #7 is nicer for seeing the trends within landcover classe over time since the plots are separated by landcover.