Assignment 5: Data Visualization

Azura Liu

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  set up

getwd()
```

```
## [1] "C:/Users/Idae/Desktop/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  3.1.5     v dplyr   1.0.7
## v tidyr   1.1.4     v stringr 1.4.0
## v readr   2.0.2     v forcats 0.5.1

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

```
#install.packages("cowplot")
library(cowplot)

Lake.chem<-read.csv("../Data/Processed/NTL-LTER_Lake_Chemistry_Nutrients_PeterPaul_Processed.csv")
```

```
Lake.mass<-read.csv("../Data/Processed/NEON_NIWO_Litter_mass_trap_Processed.csv")

#2 as.Date

class(Lake.chem$sampledate)
```

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

```
Lake.chem$sampledate<-as.Date(Lake.chem$sampledate, format = "%Y-%m-%d")
class(Lake.chem$sampledate)
```

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

```
class(Lake.mass$collectDate)
```

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

```
Lake.mass$collectDate<-as.Date(Lake.mass$collectDate, format = "%Y-%m-%d")
class(Lake.mass$collectDate)
```

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

Define your theme

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

```
#3 set theme

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

theme_set(mytheme)
```

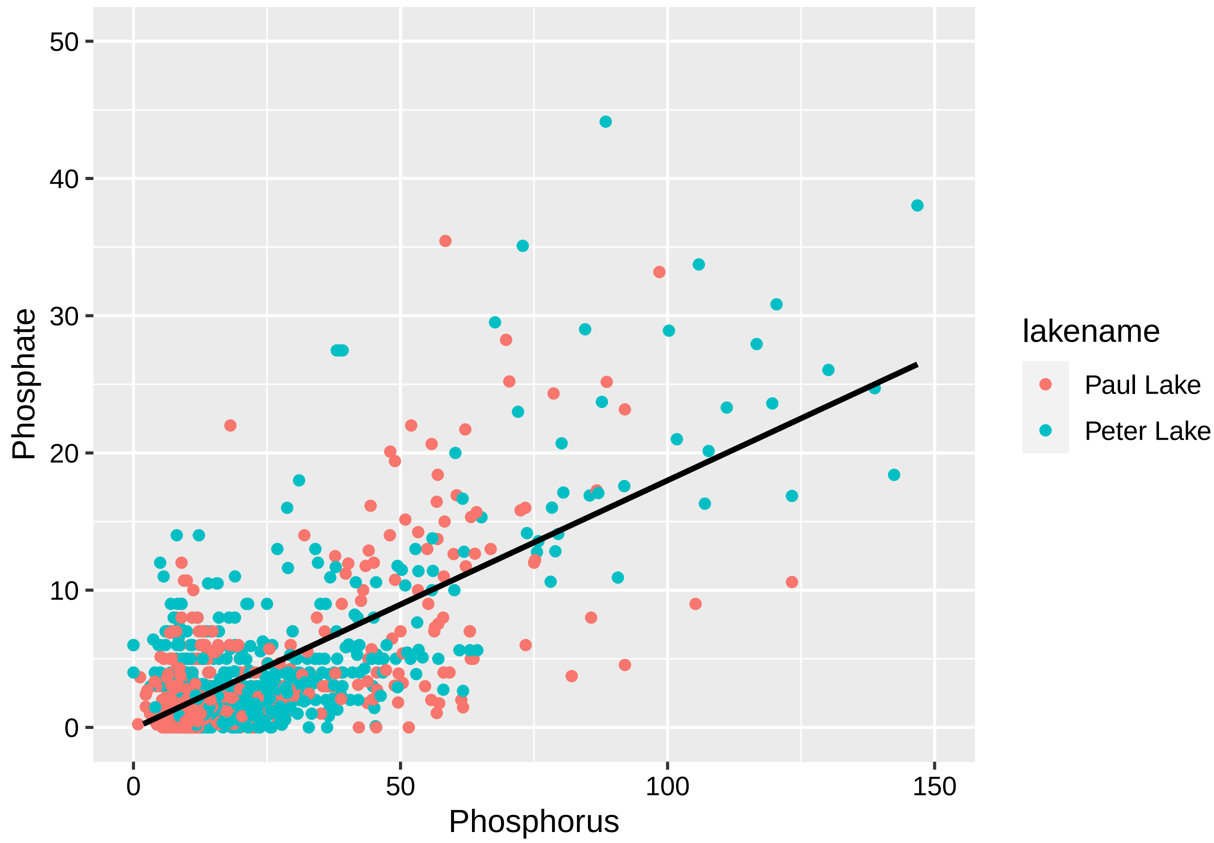
Create graphs

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
Pgraph <-
  ggplot(Lake.chem, aes(x = tp_ug, y = po4, color=lakename)) +
  geom_point() +
  xlim(0, 150) +
  ylim(0, 50) +
  geom_smooth(method = lm, se = FALSE, color = "black")+
  labs( x = "Phosphorus", y = "Phosphate")
print(Pgraph)
```

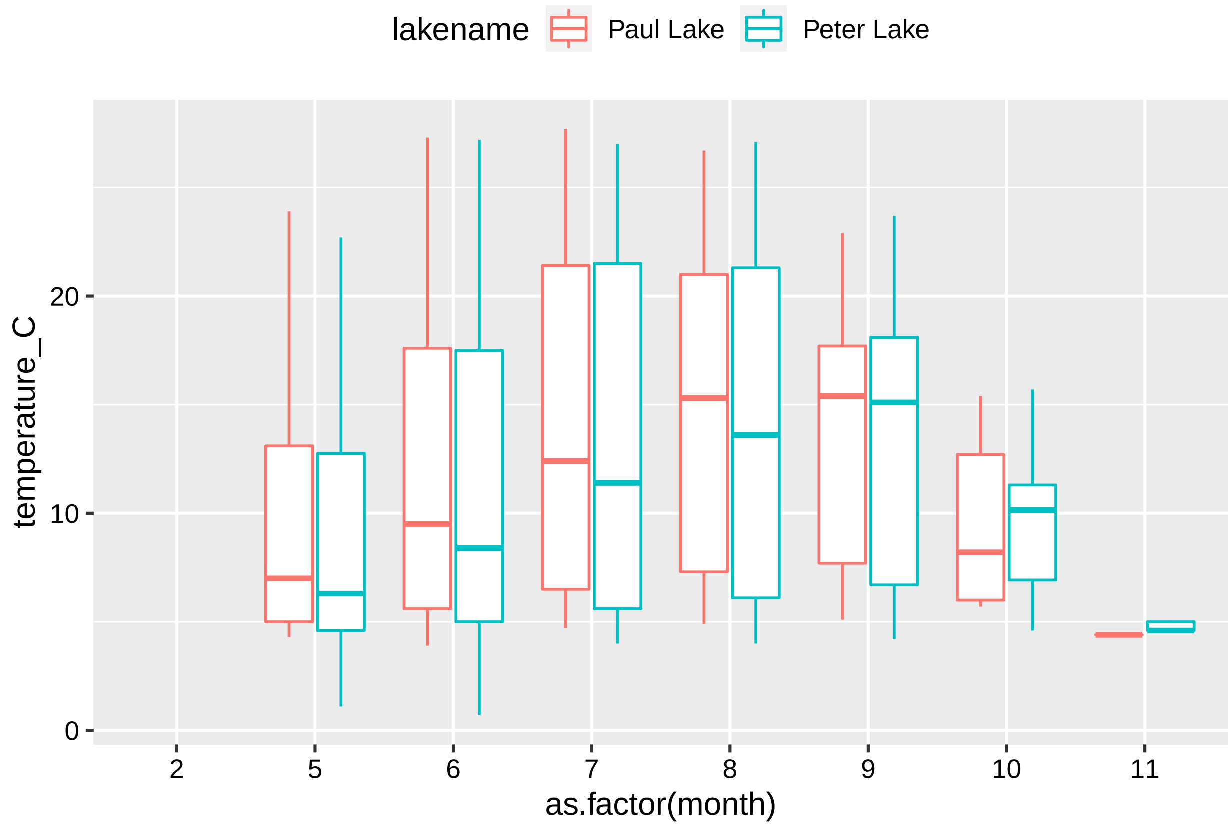
```
## `geom_smooth()` using formula 'y ~ x'

## 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.

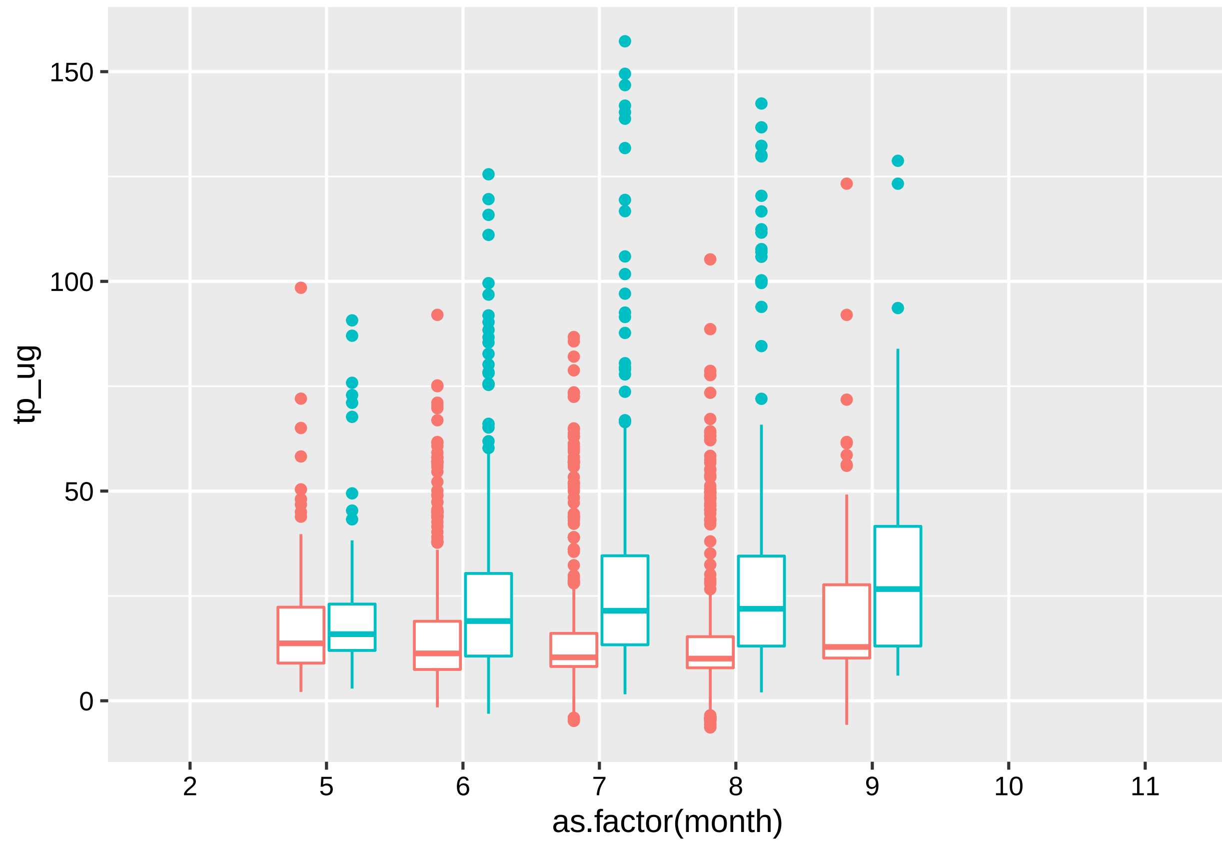
```
#5
Temp.box <-
  ggplot(Lake.chem, aes(x = as.factor(month), y = temperature_C, color = lakename)) +
  geom_boxplot()+
  theme(legend.position = "top")
print(Temp.box)
```

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



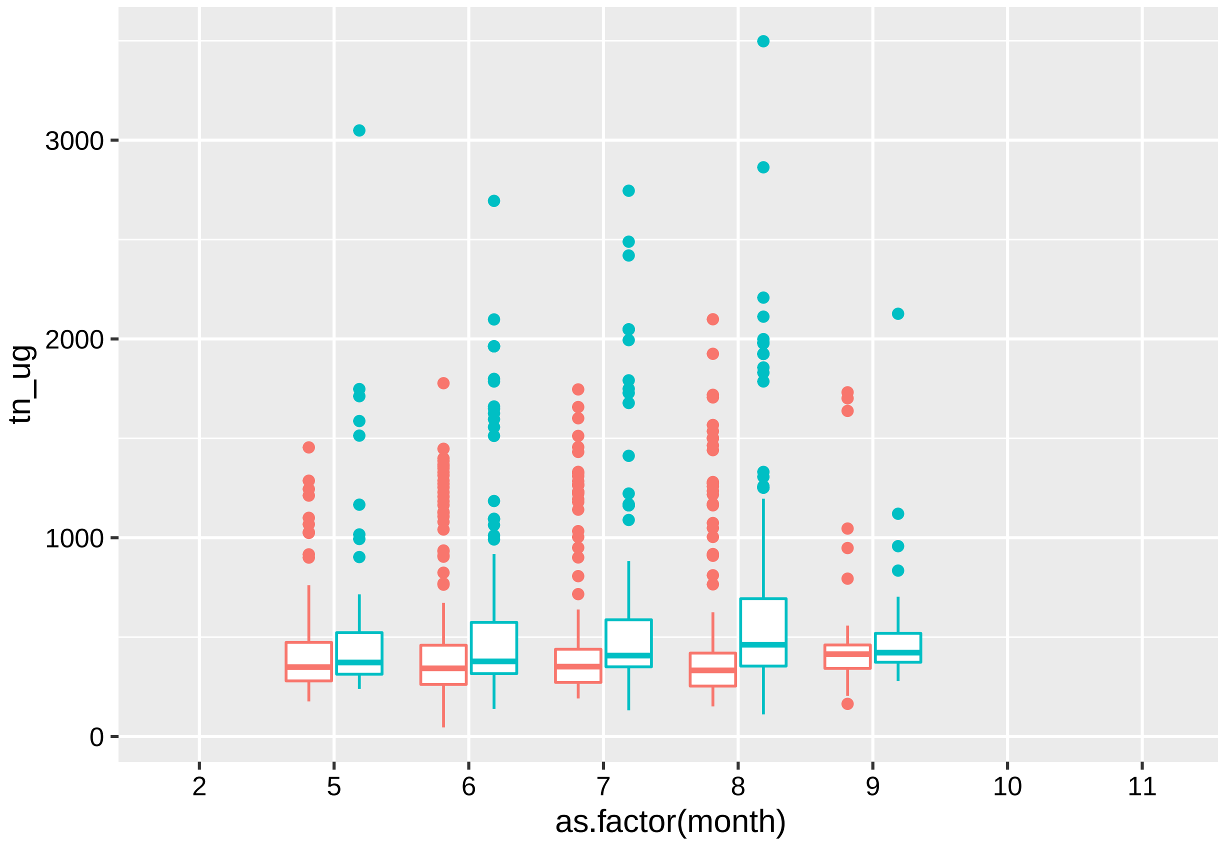

```
TP.box <-
  ggplot(Lake.chem, aes(x = as.factor(month), y = tp_ug, color = lakename)) +
  geom_boxplot() +
  theme(legend.position = "none")
print(TP.box)
```

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



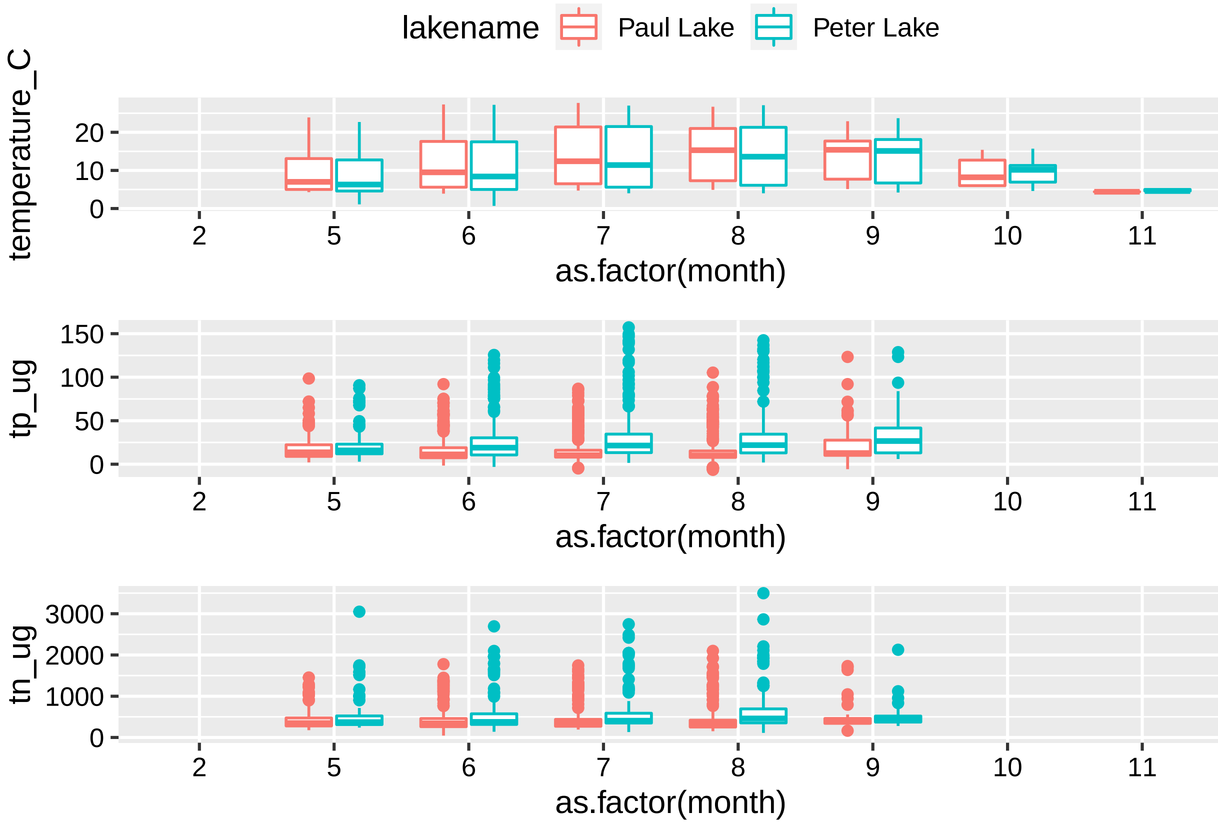

```
TN.box <-
  ggplot(Lake.chem, aes(x = as.factor(month), y = tn_ug, color = lakename)) +
  geom_boxplot() +
  theme(legend.position = "none")
print(TN.box)
```

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




```
plot_grid(Temp.box, TP.box, TN.box,nrow = 3, align = 'hv', rel_heights = c(1.25,1,1))
```

```
## 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).
## Warning: Graphs cannot be horizontally aligned unless the axis parameter is set.
## Placing graphs unaligned.
```

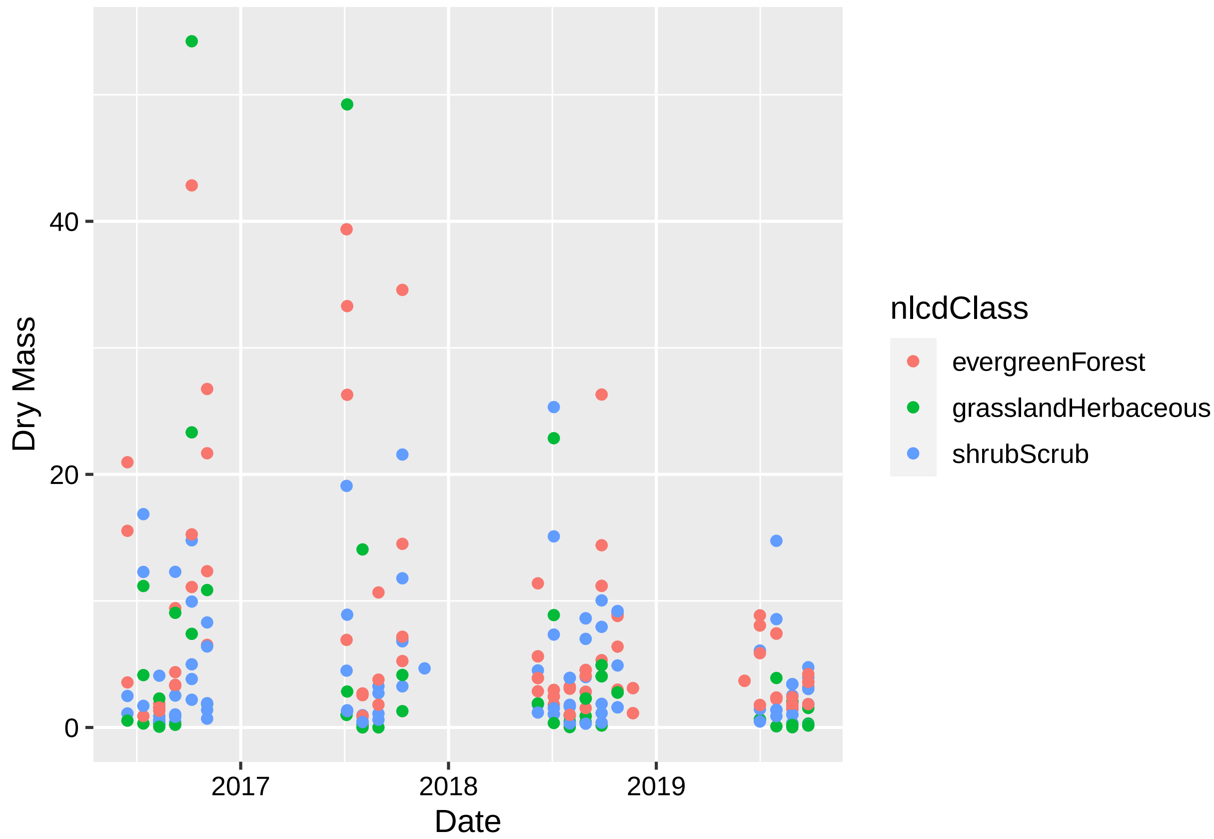


Question: What do you observe about the variables of interest over seasons and between lakes?

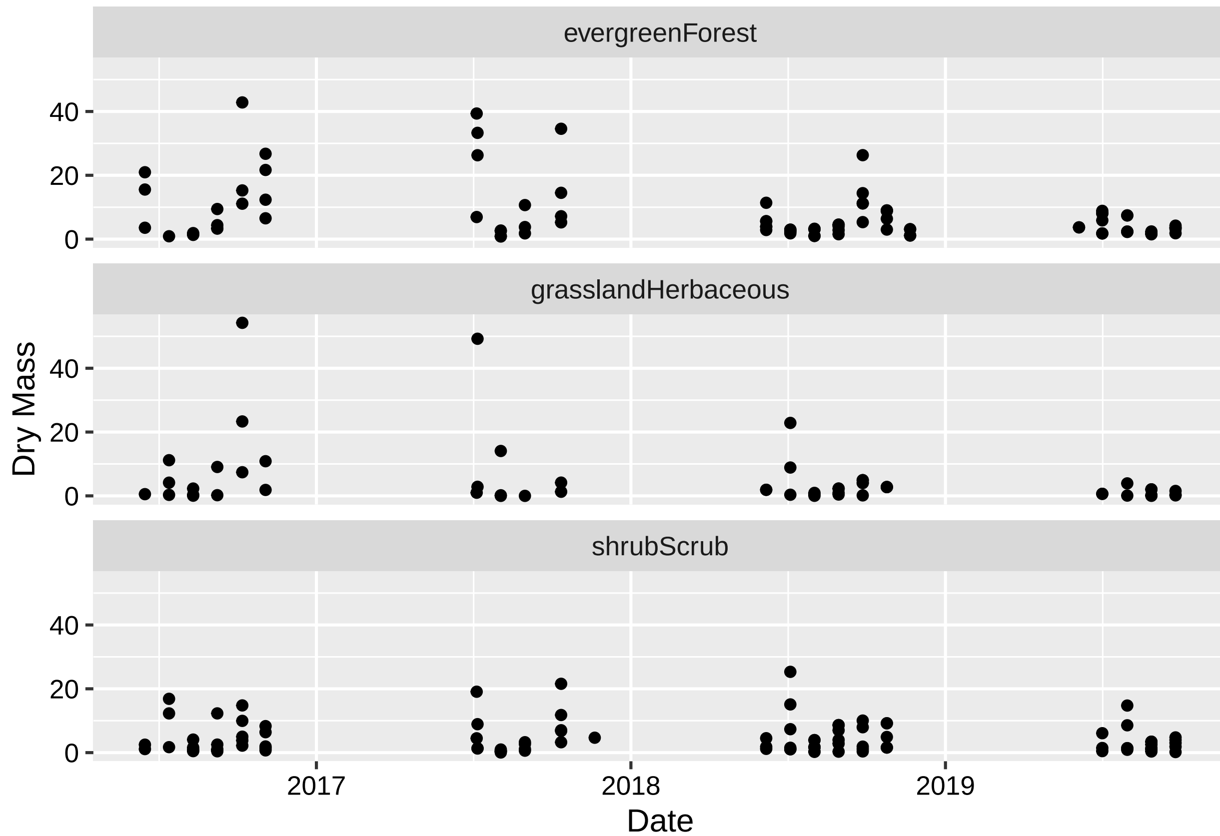
> Answer: Overall, as the temperature peaks during the summer months, so do the selected nutrients.

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
Litter.needle<-
  ggplot(filter(Lake.mass,functionalGroup == "Needles"),
         aes(x= collectDate, y= dryMass, color = nlcdClass))+
  geom_point()+
  labs( x = "Date", y = "Dry Mass")

print(Litter.needle)
```




```
#7
Litter.needle.faceted <-
 ggplot(filter(Lake.mass,functionalGroup == "Needles"), aes(x= collectDate, y= dryMass))+
  geom_point() +
  facet_wrap(vars(nlcdClass), nrow = 3)+
  labs( x = "Date", y = "Dry Mass")

print(Litter.needle.faceted)
```



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

> Answer: The faceted plot is much more effective for visualization because it providesyear to year, and both within and between group comparison. The colored dots as in #6 does not really help visualizing between groups, we just see an overall decrease in the dry mass for all litter types. For example, we can hardly tell from #6 that shrub scurb remained stable over the years, but #7 clearly shows that.