Assignment 5: Data Visualization

Amrita Sood

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.

```
getwd()
```

```
## [1] "/Users/amritasood/Desktop/CLASSES MEM/Spring 2021/EDA - ENV872/Environmental_Data_Analytics_202
```

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

```
library(tidyverse)
```

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

```
## v ggplot2 3.3.3     v purrr   0.3.4
## v tibble  3.0.5     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(ggridges)
library(viridis)
```

```
## Loading required package: viridisLite
```

```
library(RColorBrewer)
library(colormap)
```

```
#1
PeterPaul.chem.nutrients <-
  read.csv("./Data/Processed/NTL-LTER_Lake_Chemistry_Nutrients_PeterPaul_Processed.csv", stringsAsFacto
PeterPaul.chem.nutrients.gathered <-
  read.csv("./Data/Processed/NTL-LTER_Lake_Nutrients_PeterPaulGathered_Processed.csv", stringsAsFactors
NiwotRidge.litter <-
  read.csv("./Data/Processed/NEON_NIWO_Litter_mass_trap_Processed.csv", stringsAsFactors = TRUE)


#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, fo
NiwotRidge.litter$collectDate <- as.Date(NiwotRidge.litter$collectDate, format = "%Y-%m-%d")
```

Define your theme

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

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

theme_set(mytheme)
```

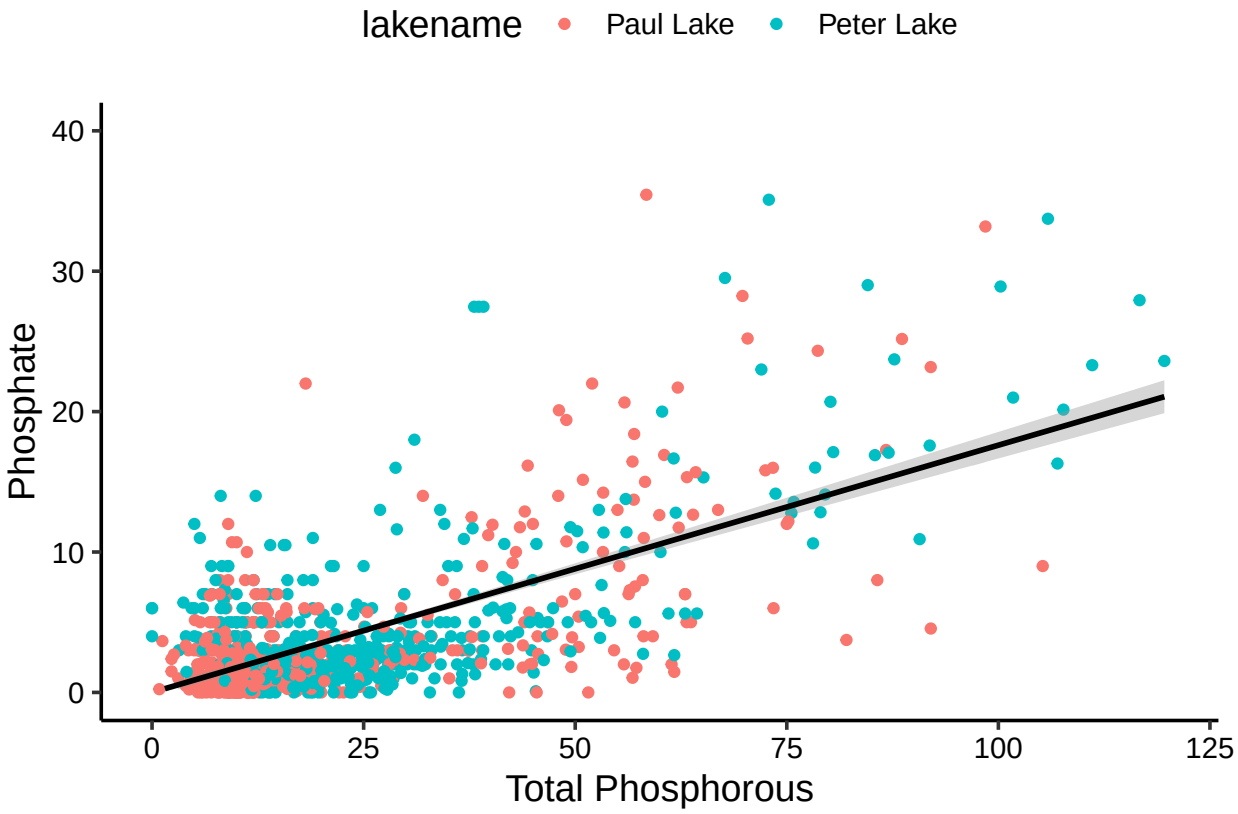
Create graphs

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.

```
Plot.1 <- ggplot(PeterPaul.chem.nutrients, aes(x = tp_ug, y = po4)) + geom_point(aes(color=lakename))+la
  ylim(0, 40)+
  geom_smooth(method=lm, color="black")
print(Plot.1)
```

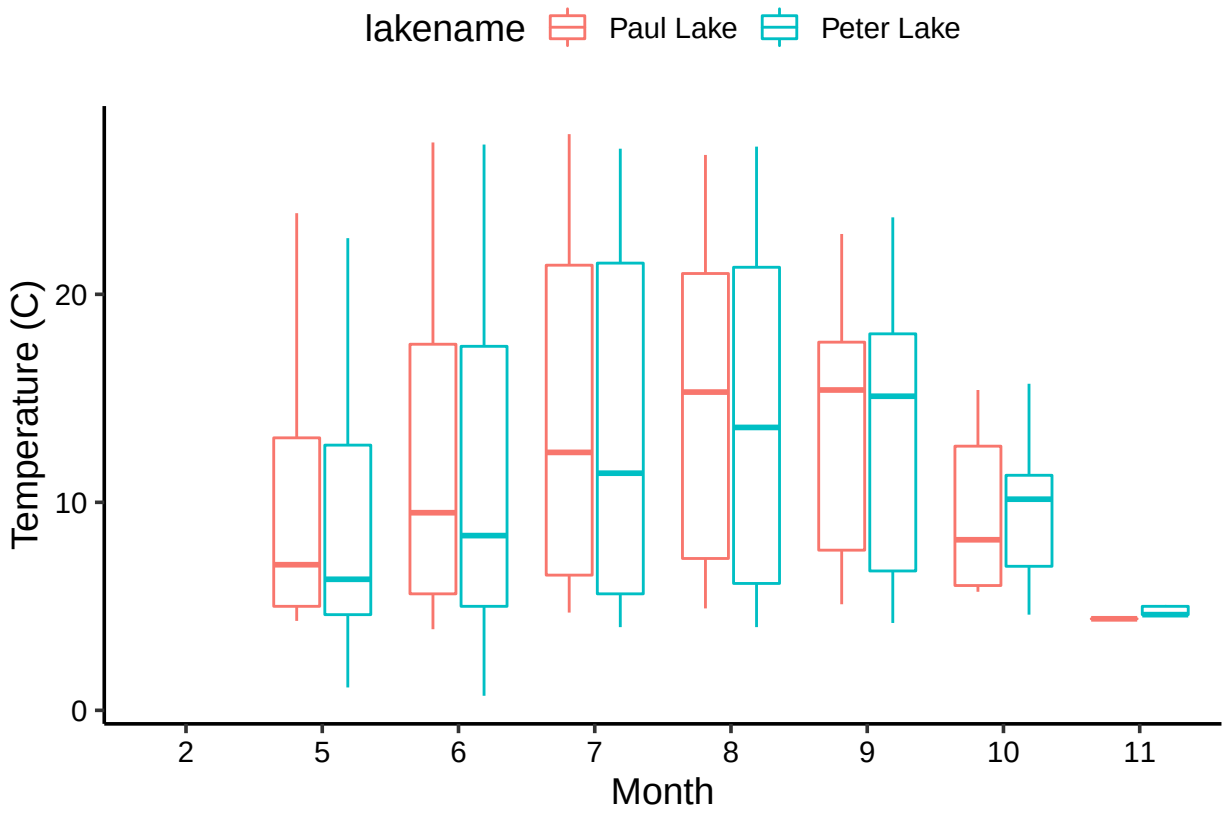
```
## `geom_smooth()` using formula 'y ~ x'

## Warning: Removed 21956 rows containing non-finite values (stat_smooth).

## Warning: Removed 21956 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.

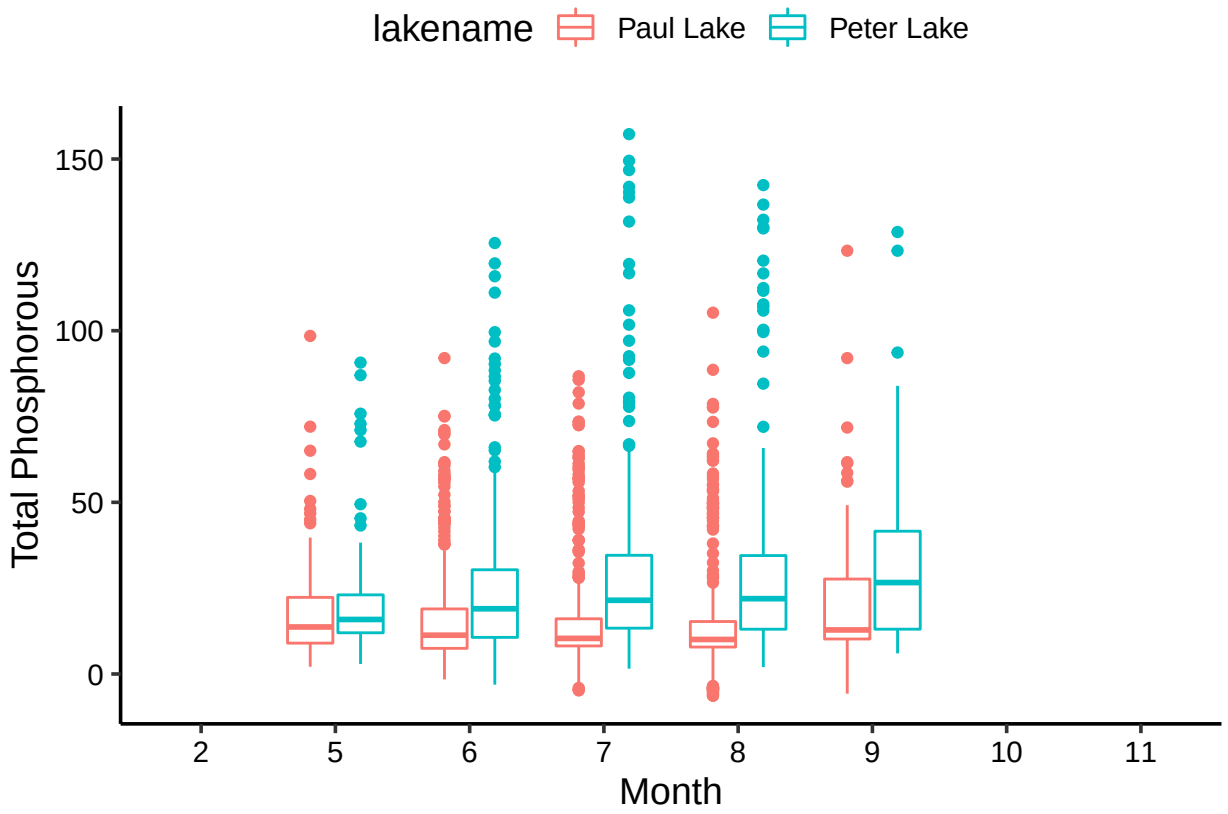
```
PeterPaul.chem.nutrients$month <- as.factor(PeterPaul.chem.nutrients$month)

#A. Temperature
Plot.2 <-
  ggplot(PeterPaul.chem.nutrients, aes(x = month, y = temperature_C, color=lakename)) +
  geom_boxplot() + labs(fill = "Lake Name", x="Month", y="Temperature (C)")

print(Plot.2)
```

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



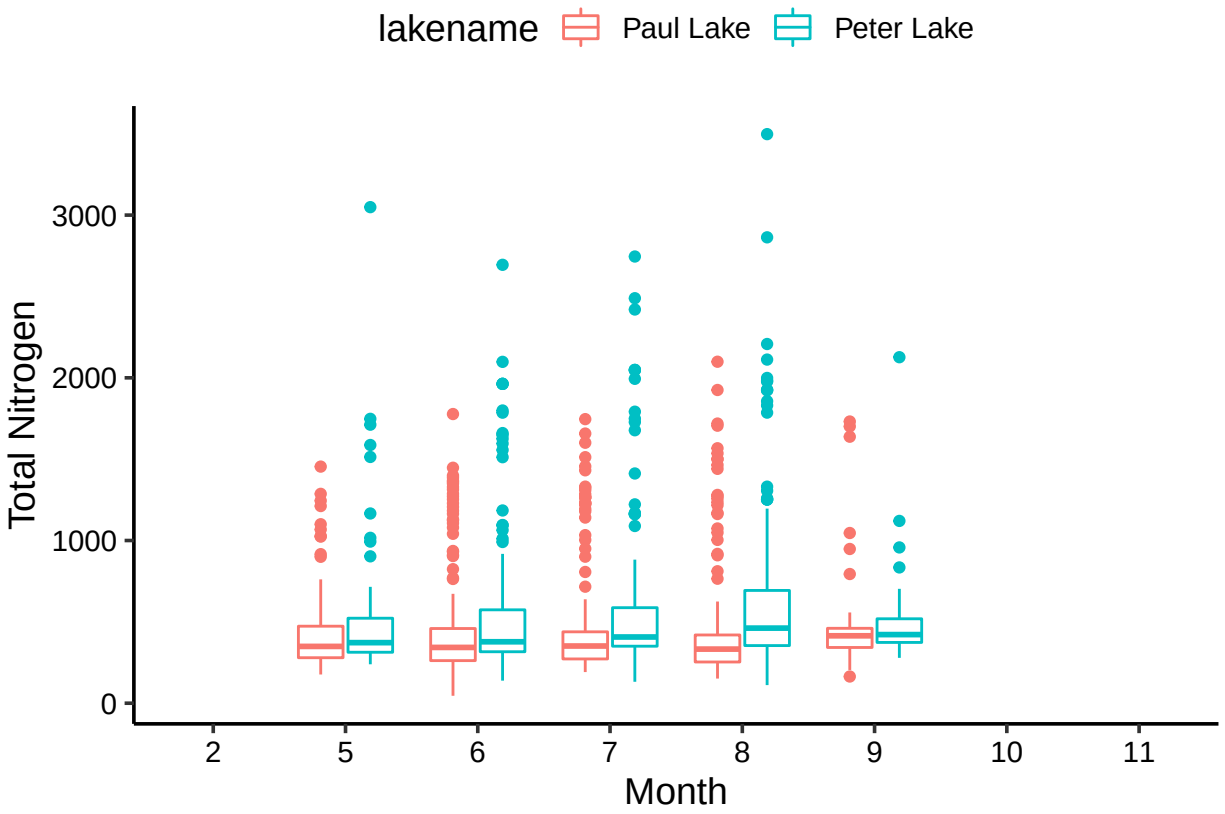

```
#B. TP
Plot.3 <-
  ggplot(PeterPaul.chem.nutrients, aes(x = month, y = tp_ug, color=lakename)) +
  geom_boxplot() + labs(x="Month", y="Total Phosphorous")
print(Plot.3)
```

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




```
#C. TN
Plot.4 <-
  ggplot(PeterPaul.chem.nutrients, aes(x = month, y = tn_ug, color=lakename)) +
  geom_boxplot() + labs(x="Month", y="Total Nitrogen")
print(Plot.4)
```

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



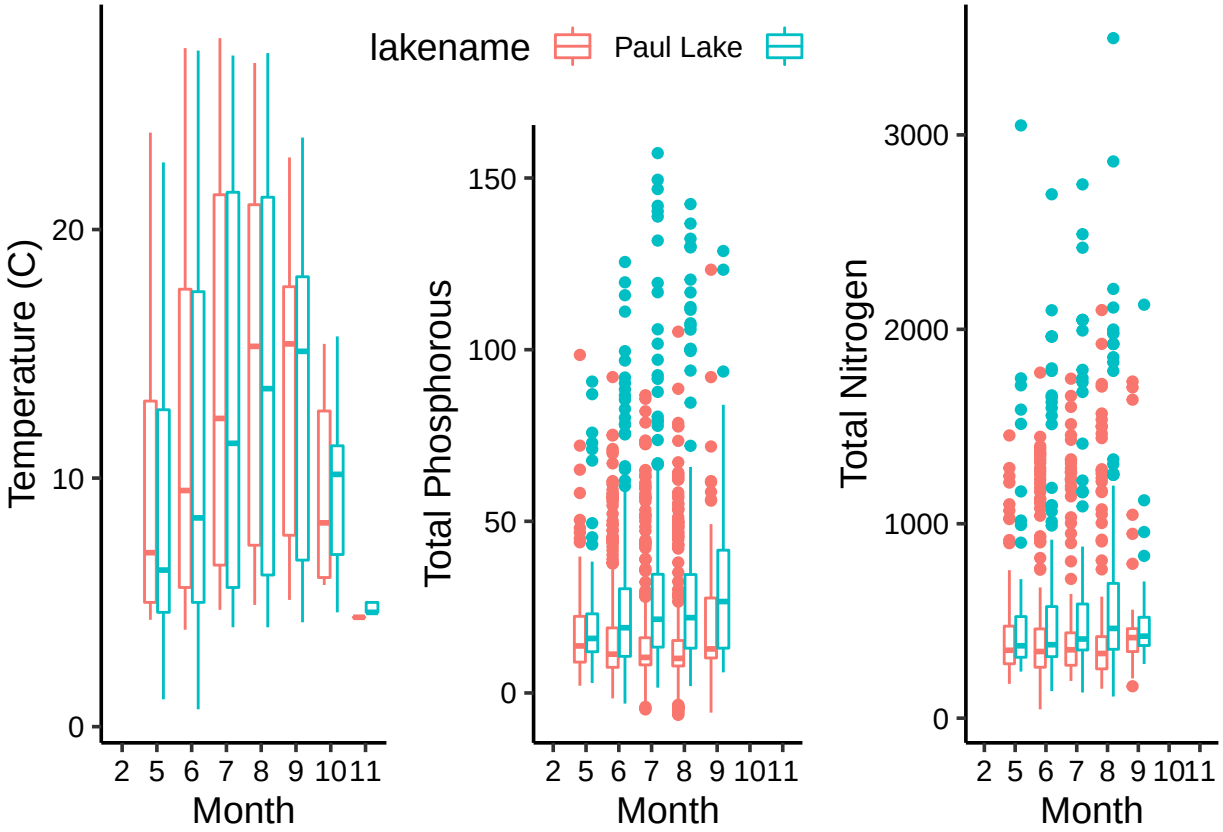

```
#Cowplot to compare the graphs
library(cowplot)
 plot_grid(Plot.2 + theme( legend.position ="none"), Plot.3, Plot.4+ theme(legend.position ="none"), nr
```

```
## 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.
```

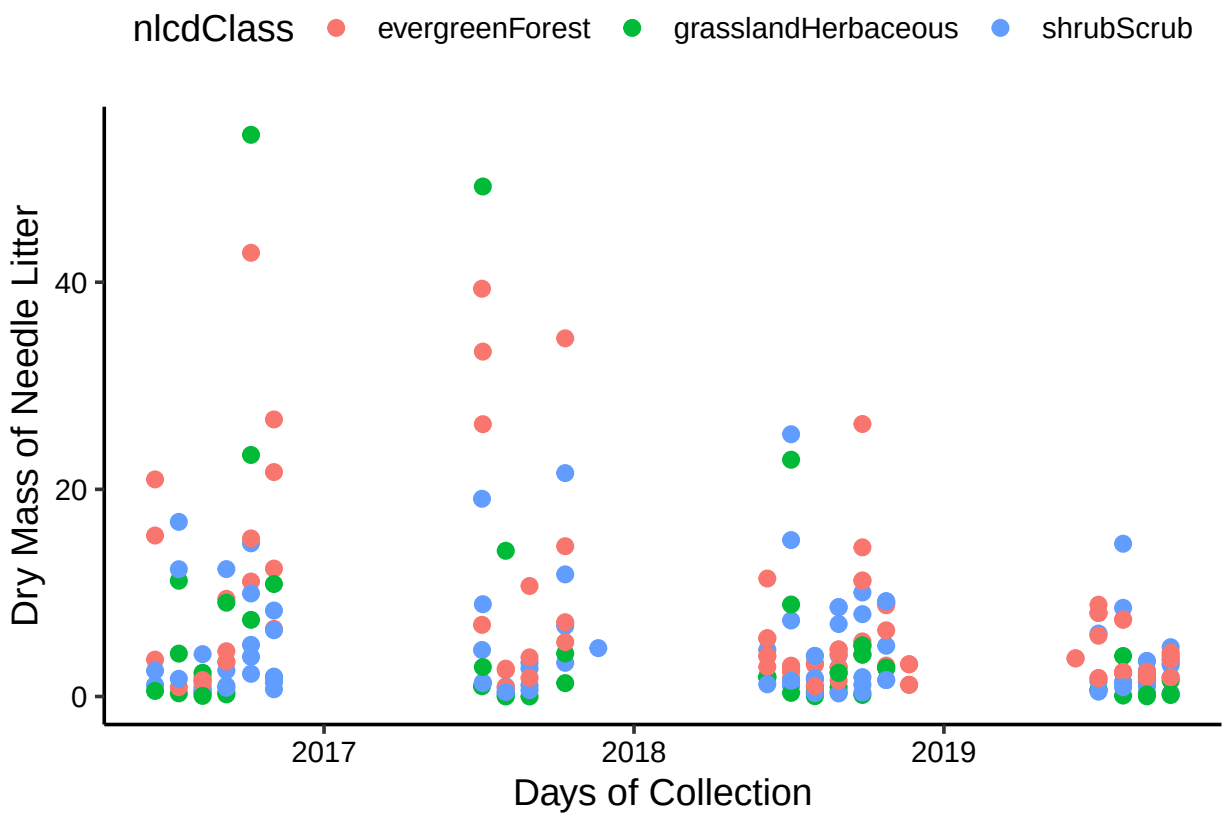


Question: What do you observe about the variables of interest over seasons and between lakes?

Answer: The concentration of nitrigen and phosphorous levels are consistentl higher for Peter lake. As expected, the temperatures are higher in summer months comapred to the winter months. The total phosphorous and total nitrogen levels are higher in summer months (7th and 8th month) compared to other colder months for both Peter and Paul lake

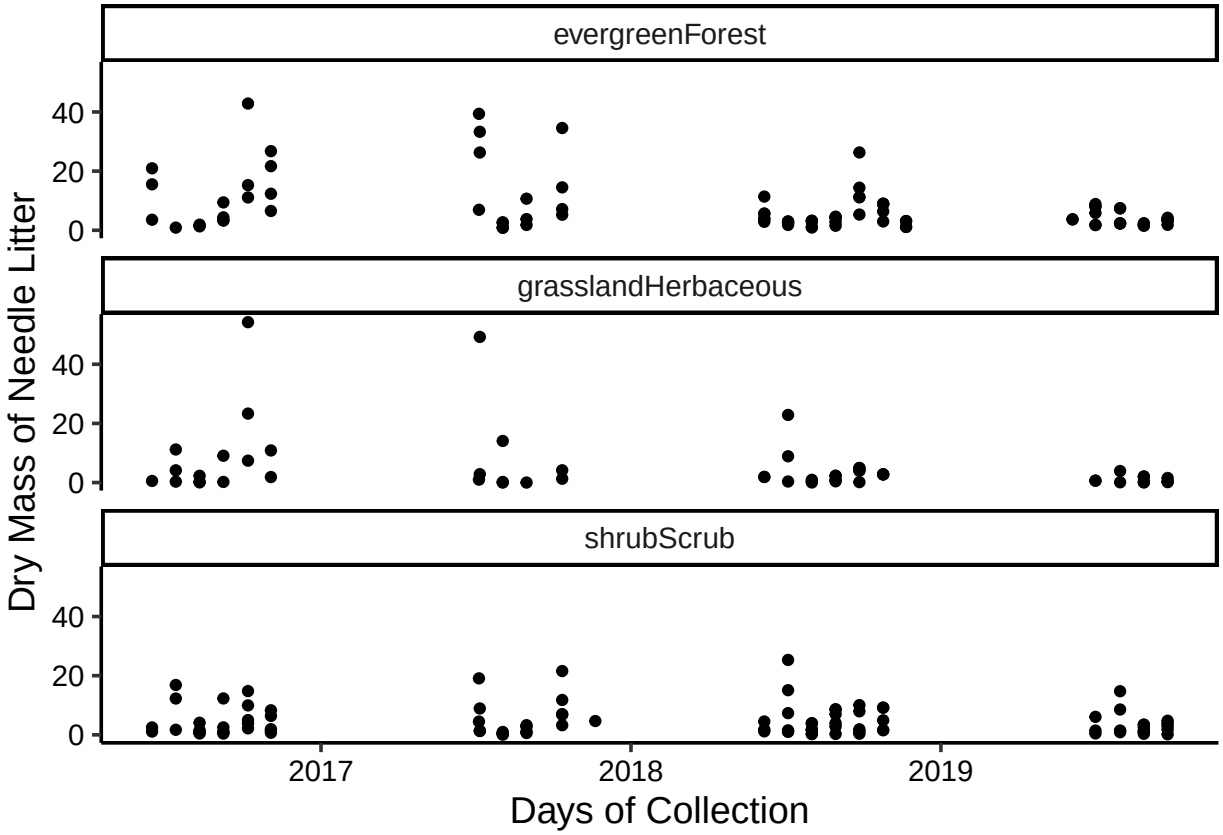
6. [Niwot Ridge] Plot a subset of the litter dataset by displaying only the "Needles" functional group. Plot the dry mass of needle litter by date and separate by NLCD class with a color aesthetic. (no need to adjust the name of each land use)

```
Plot.5 <-
  ggplot(subset(NiwotRidge.litter, functionalGroup == "Needles"),
         aes(y = dryMass, x = collectDate)) + geom_point(aes(color=nlcdClass), size =2.5)+ labs(y = "Dry
print(Plot.5)
```



7. [Niwot Ridge] Now, plot the same plot but with NLCD classes separated into three facets rather than separated by color.

```
Plot.6 <-
  ggplot(subset(NiwotRidge.litter, functionalGroup == "Needles"),
         aes(y = dryMass, x = collectDate)) + geom_point()+ labs(y = "Dry Mass of Needle Litter", x = "
print(Plot.6)
```



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

Answer:Plot 7 is more efficient because it shows us the data for each 3 -"Evergeen Forest", "Grass Land Herbaceous", "Shrub Scrub" separately making it easier for us to interpret the data.