# Assignment 5: Data Visualization

Karen Thornton

## 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] "/Users/karenthornton/Documents/School/Grad School/Year 1/Semester 2/EDA/Environmental_Data_Analy
```

```
library(tidyverse)
```

```
## -- Attaching packages --------------------------------------- tidyverse 1.3.1 --

## v ggplot2 3.3.5     v purrr   0.3.4
## v tibble  3.1.4     v dplyr   1.0.7
## v tidyr   1.1.3     v stringr 1.4.0
## v readr   2.0.1     v forcats 0.5.1

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

```
library(cowplot)
PPChemNutrients <- read.csv("/Users/karenthornton/Documents/School/Grad School/Year 1/Semester 2/EDA/Ent
Niwotlitter <- read.csv("/Users/karenthornton/Documents/School/Grad School/Year 1/Semester 2/EDA/Enviro
#2
class(Niwotlitter$collectDate)
```

```
## [1] "factor"
```

```
class(PPChemNutrients$sampledate)
```

```
## [1] "factor"
```

```
Niwotlitter$collectDate <-
  as.Date(Niwotlitter$collectDate, format="%Y-%m-%d")
PPChemNutrients$sampledate <-
  as.Date(PPChemNutrients$sampledate, format= "%Y-%m-%d")
```

## Define your theme

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

```
#3
mytheme <- theme_light(base_size = 16)+
  theme(legend.position = "top",
        legend.justification = "right",
        axis.text = element_text(color= "black"))

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
PhosPlot <- ggplot(PPChemNutrients,
                   aes(x= po4, y= tp_ug, color = lakename))+
  geom_point()+
  xlim(0,50)+
  ylim(0,150)+
  ylab("Total phosphorus") +
  xlab("Phosphate") +
  ggtitle("Total Phosphorus by Phosphate")+
  geom_smooth(method =lm, color = "black")+
  theme(legend.position = "right")+
```

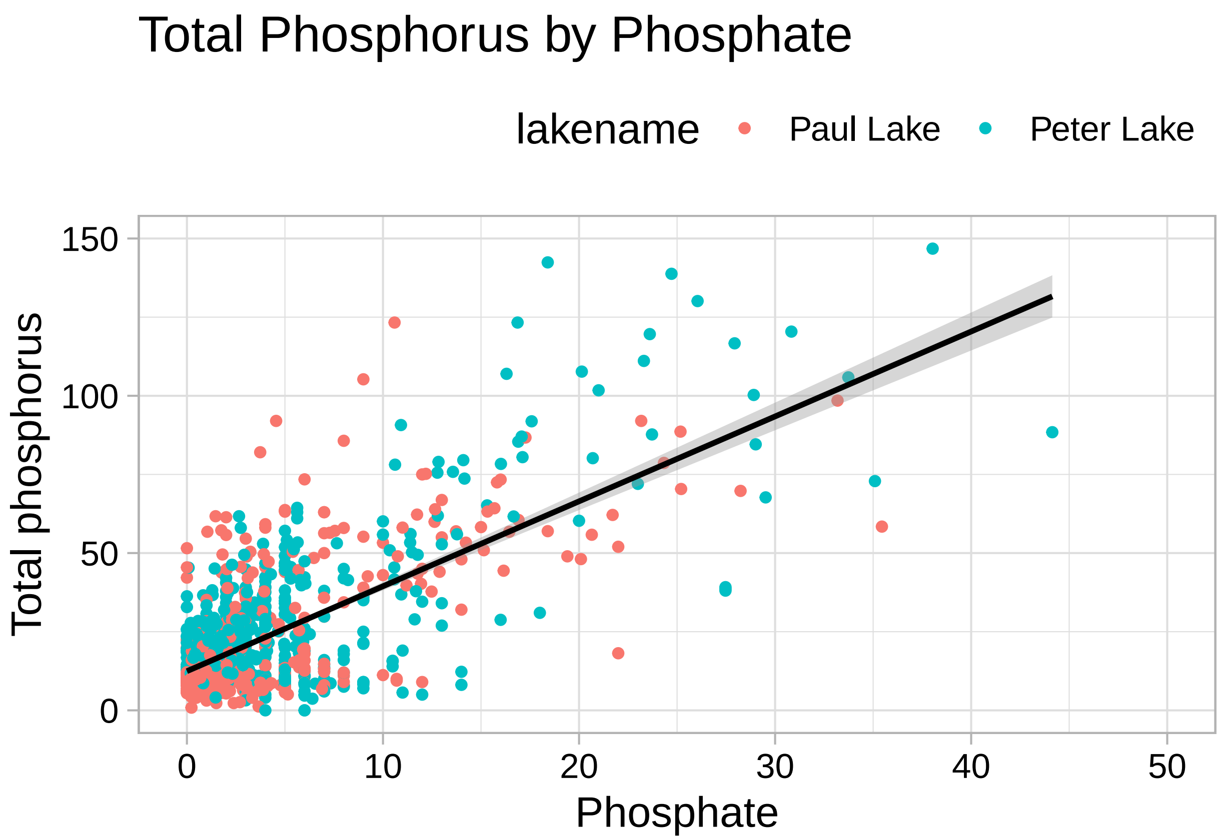
```
  mytheme

print(PhosPlot)
```

```
## `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).
```




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.

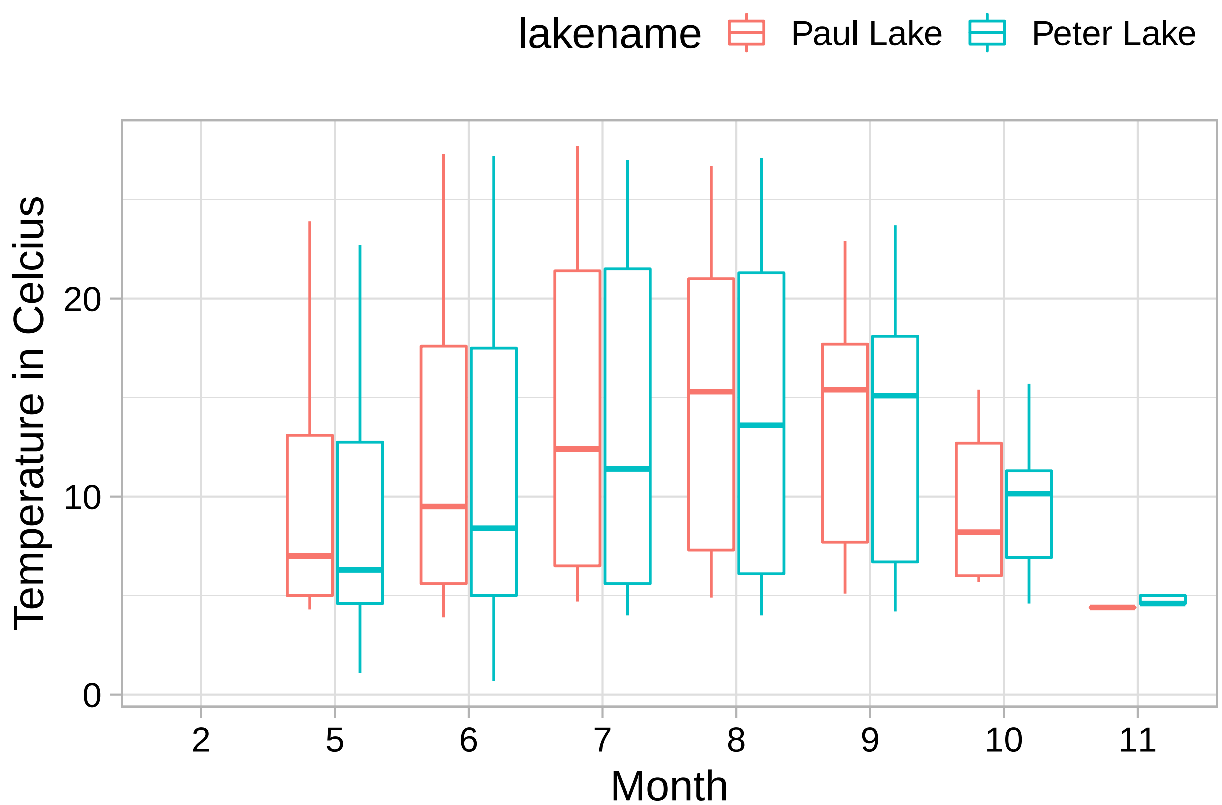
```
#5
PPChemNutrients$month <- as.factor(PPChemNutrients$month)
class(PPChemNutrients$month)
```

```
## [1] "factor"
```

```
PPChemTemp<-ggplot(PPChemNutrients,
                   aes(x= month, y= temperature_C,
                       color =lakename))+
```

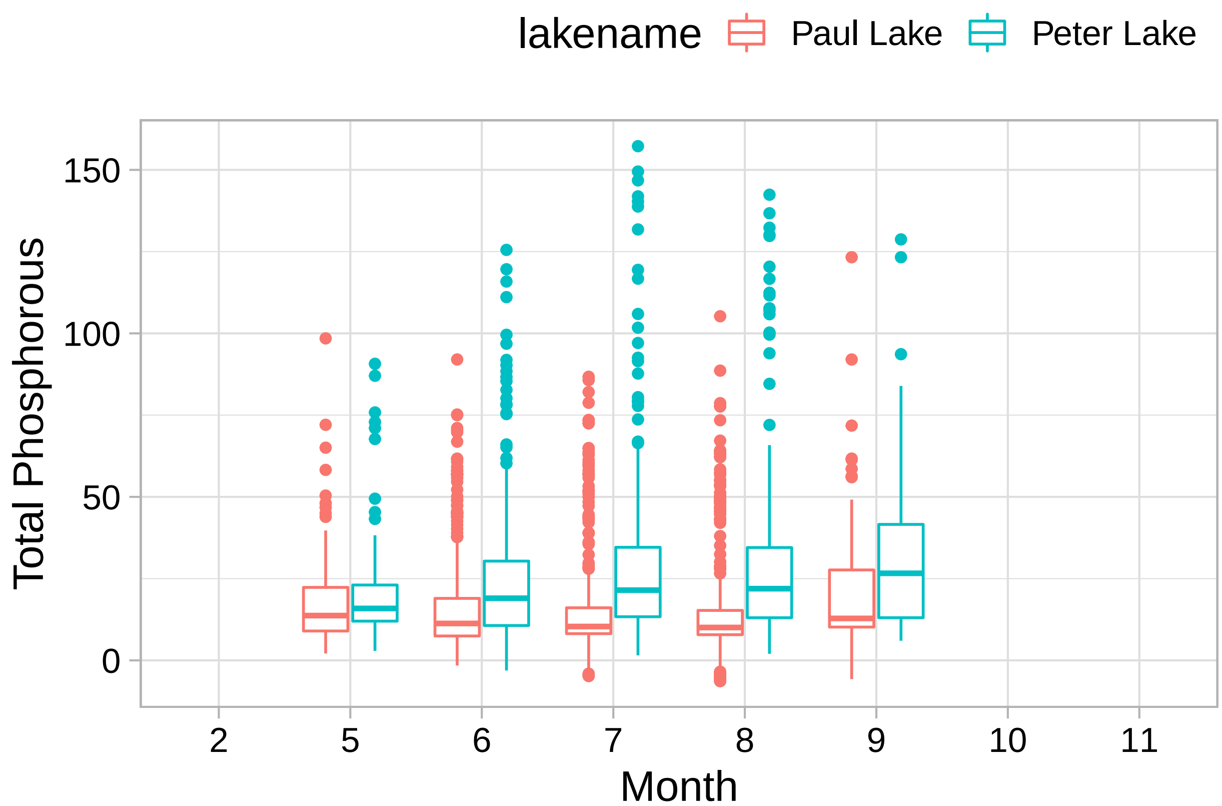
```
  geom_boxplot()+
  ylab("Temperature in Celcius")+
  xlab("Month")+
  theme(legend.position = "right")+
  mytheme
print(PPChemTemp)
```

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



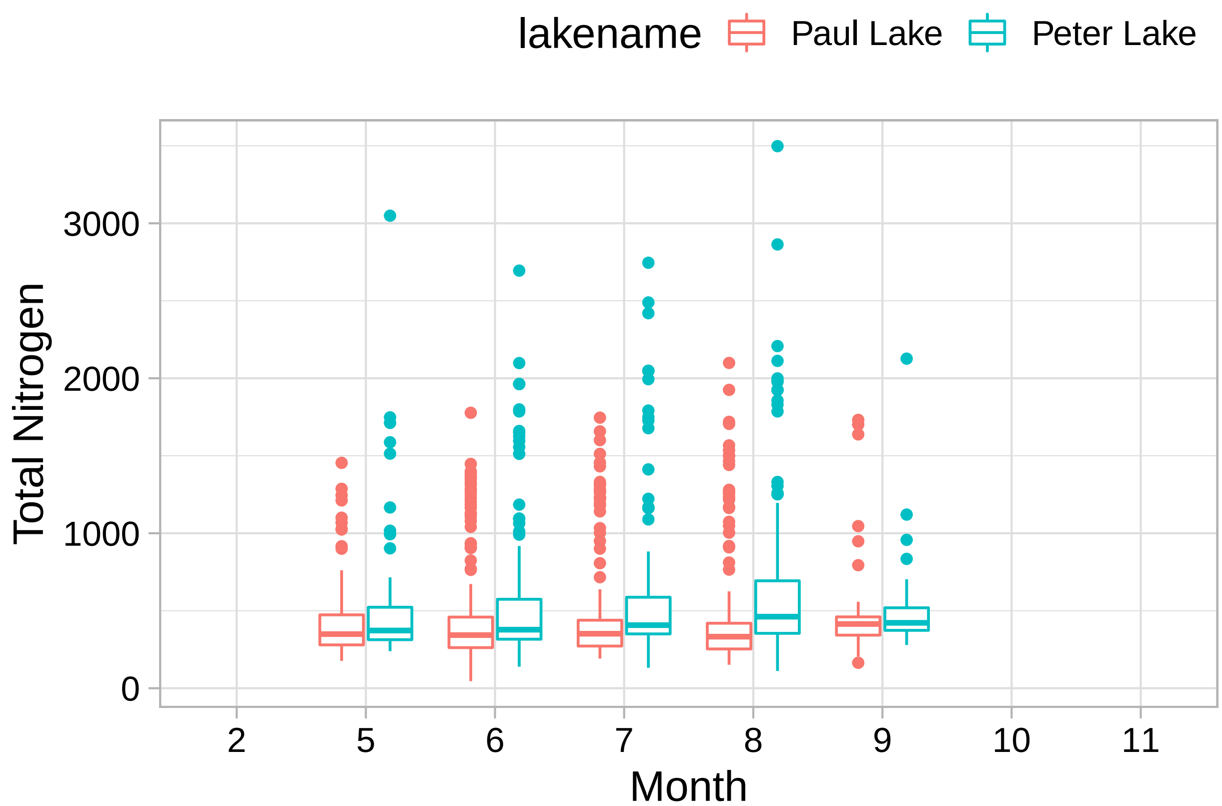

```
PPChemTP<-ggplot(PPChemNutrients,
                 aes(x= month, y= tp_ug,
                     color =lakename))+
  geom_boxplot()+
  ylab("Total Phosphorous")+
  xlab("Month")+
  theme(legend.position = "right")+
  mytheme
print(PPChemTP)
```

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




```
PPChemTN<-ggplot(PPChemNutrients,
                 aes(x= month, y= tn_ug,
                     color =lakename))+
  geom_boxplot()+
  ylab("Total Nitrogen")+
  xlab("Month")+
  theme(legend.position = "right")+
  mytheme
print(PPChemTN)
```

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




```
library(cowplot)
legendTN<- get_legend(PPChemTN+
                        theme(legend.position = "bottom"))
```

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

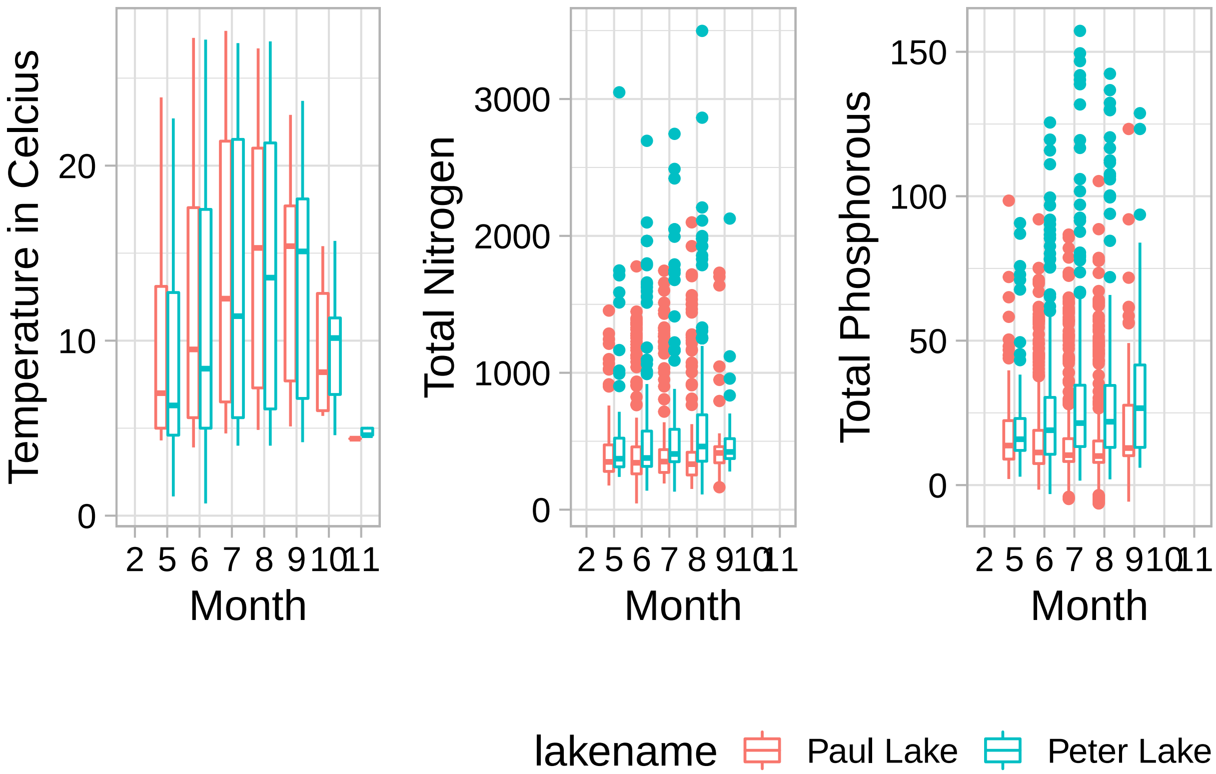
```
CombinedGrids<-
  plot_grid(PPChemTemp+
              theme(legend.position = "none"),
            PPChemTN+theme(legend.position = "none"),
            PPChemTP+theme(legend.position = "none"),
            align= "h", axis ="b",nrow=1)
```

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

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

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

```
FinalNumb5<- plot_grid(CombinedGrids, legendTN, ncol =1,
                       rel_heights = c(1,.3))
print(FinalNumb5)
```



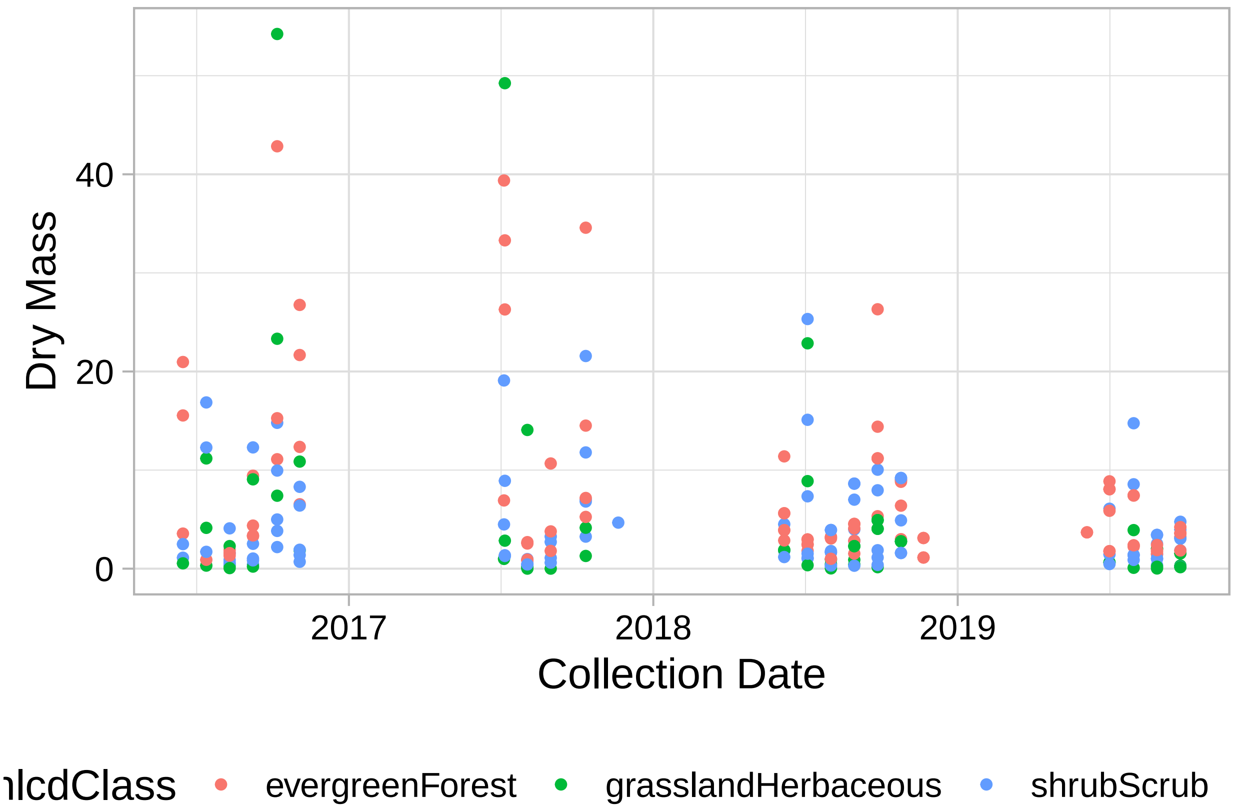

Question: What do you observe about the variables of interest over seasons and between lakes?

> Answer: The temperatures in the summer months (6,7,8) are higher than the other months. There are more points outside the upper quartile for Peter lake than Paul lake for total nitrogen and phosphorous. The numbers start to increase for nitrogen and phosphorous for the later months (7,8,9).

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.

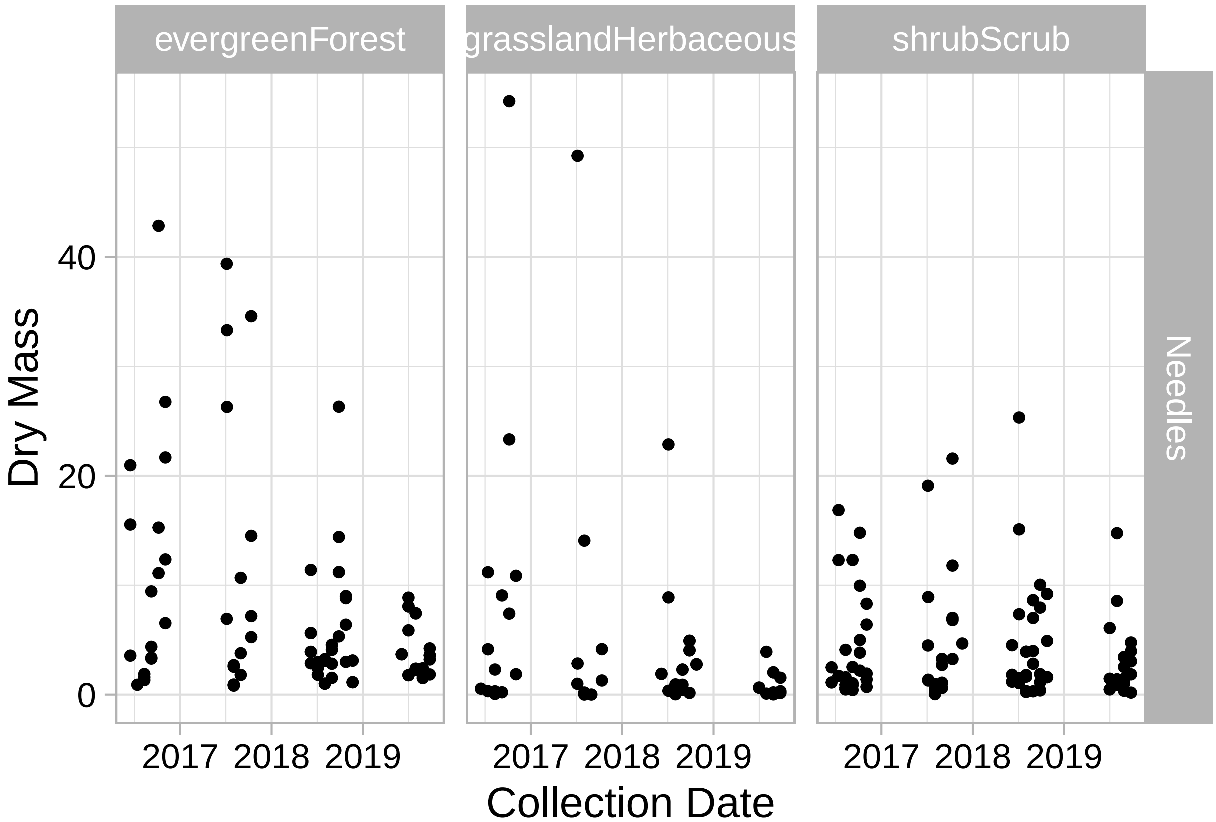
```
#6
LitterSub <- ggplot(subset(Niwotlitter,
                           functionalGroup == "Needles"),
                    aes(x=collectDate, y= dryMass, color = nlcdClass))+
  geom_point()+
  labs(x= "Collection Date", y= "Dry Mass")+
  theme(legend.position = "bottom")

print(LitterSub)
```




```
#7
LitterFac <- ggplot(subset(Niwotlitter,
                           functionalGroup == "Needles"),
                    aes(x=collectDate, y= dryMass))+
  geom_point()+
  facet_grid(functionalGroup ~ nlcdClass)+
  labs(x= "Collection Date", y= "Dry Mass")+
  theme(legend.position = "bottom")

print(LitterFac)
```




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

Answer:If you are trying to compare the three different classes then I think the plot with colored classes (plot 6) is better because it is easier to compare classes for each date. The faceted plot (plot 7) is good if you want to look at one class over the years, but not if you want to compare all the classes on each date.