Assignment 5: Data Visualization

Nadia Swit

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
library(tidyverse)
```

```
## -- Attaching packages ---------------------------------------------------------------------------------- ti

## v ggplot2 3.3.2     v purrr   0.3.4
## v tibble  3.0.3     v dplyr   1.0.2
## v tidyr   1.1.2     v stringr 1.4.0
## v readr   1.3.1     v forcats 0.5.0

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

```
library(cowplot)

Chem.Nutrients <- read.csv("./Data/Processed/NTL-LTER_Lake_Chemistry_Nutrients_PeterPaul_Processed.csv")
Gathered.Nutrients <- read.csv("./Data/Processed/NTL-LTER_Lake_Nutrients_PeterPaulGathered_Processed.csv
Litter <- read.csv("./Data/Processed/NEON_NIWO_Litter_mass_trap_Processed.csv")


#2
Chem.Nutrients$sampledate <- as.Date(Chem.Nutrients$sampledate, format = "%Y-%m-%d")
Gathered.Nutrients$sampledate <- as.Date(Gathered.Nutrients$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.

```
mytheme <- theme_light(base_size = 12) +
  theme(axis.text = element_text(color = "black"),
        legend.position = "left",
        plot.title=element_text(hjust=0.5)) #alternative: legend.position + legend.justification

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.

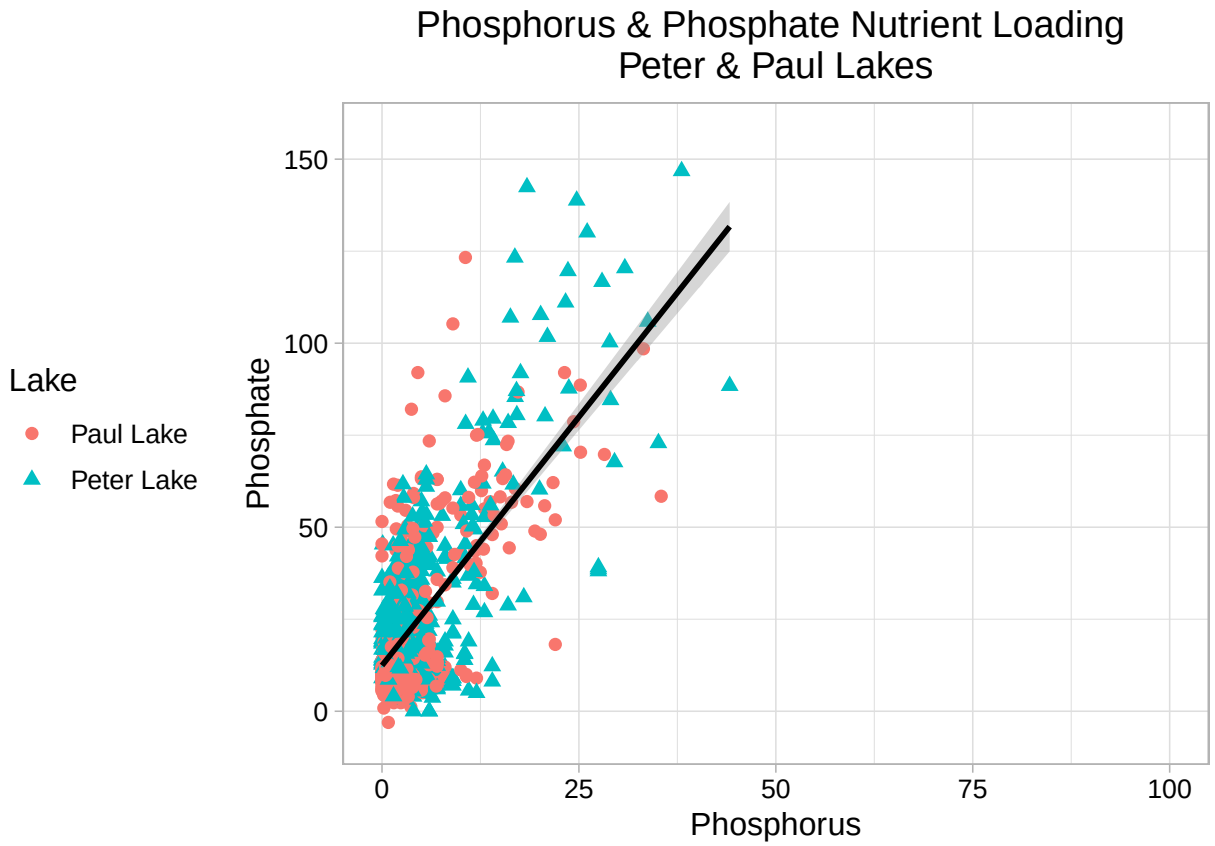
4. [NTL-LTER] Plot total phosphorus (`tp_ug`) by phosphate (`po4`), with separate aesthetics for Peter and Paul lakes. Add a line of best fit and color it black. Adjust your axes to hide extreme values.

```
Phosphorus.Phosphate <-
  ggplot(Chem.Nutrients, aes(x = po4, y = tp_ug)) +
  geom_point(aes(color=lakename, shape=lakename), size=2)+
  xlim(0,100) +
  labs(title="Phosphorus & Phosphate Nutrient Loading \nPeter & Paul Lakes", x="Phosphorus", y="Phospha
  geom_smooth(method=lm, color="black")
print(Phosphorus.Phosphate)
```

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

## Warning: Removed 21947 rows containing non-finite values (stat_smooth).

## Warning: Removed 21947 rows containing missing values (geom_point).
```

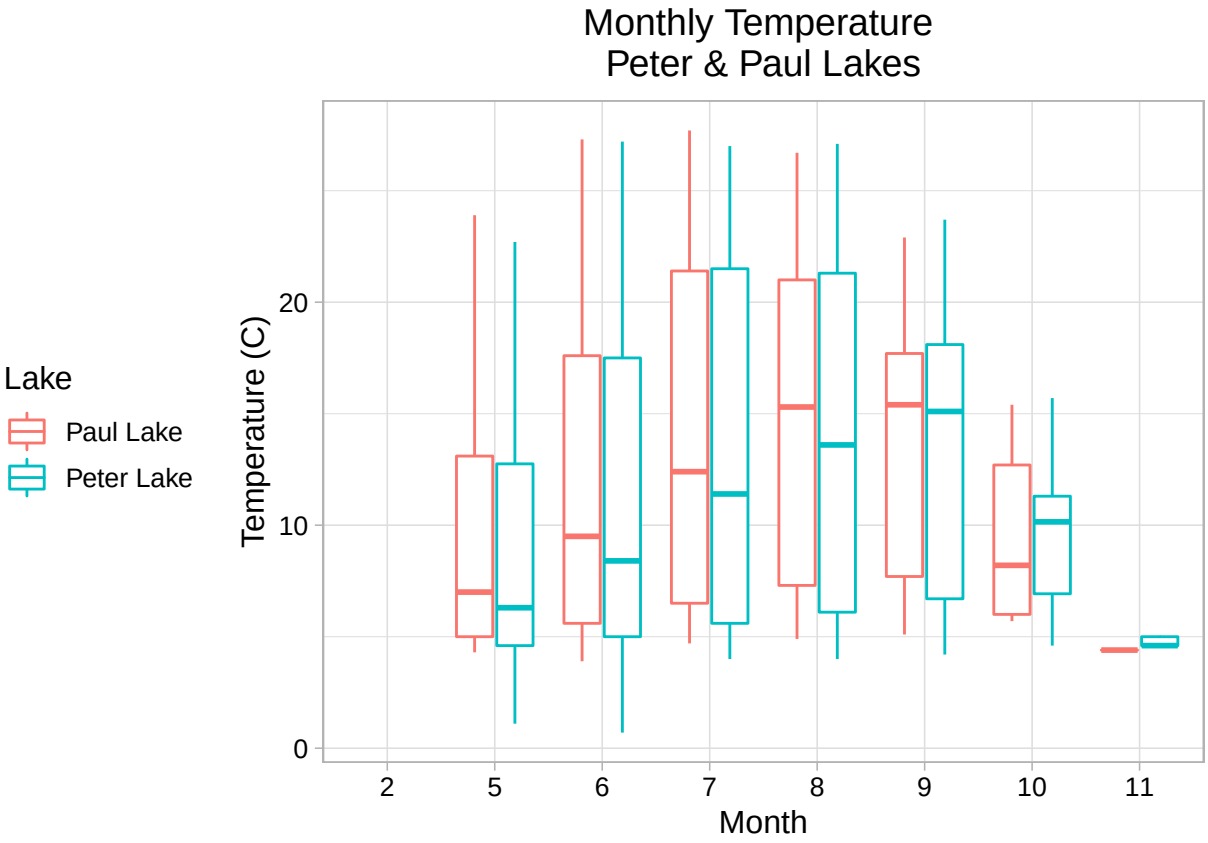



```
#able to change legend title because used color in aes of geom_point
```

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.

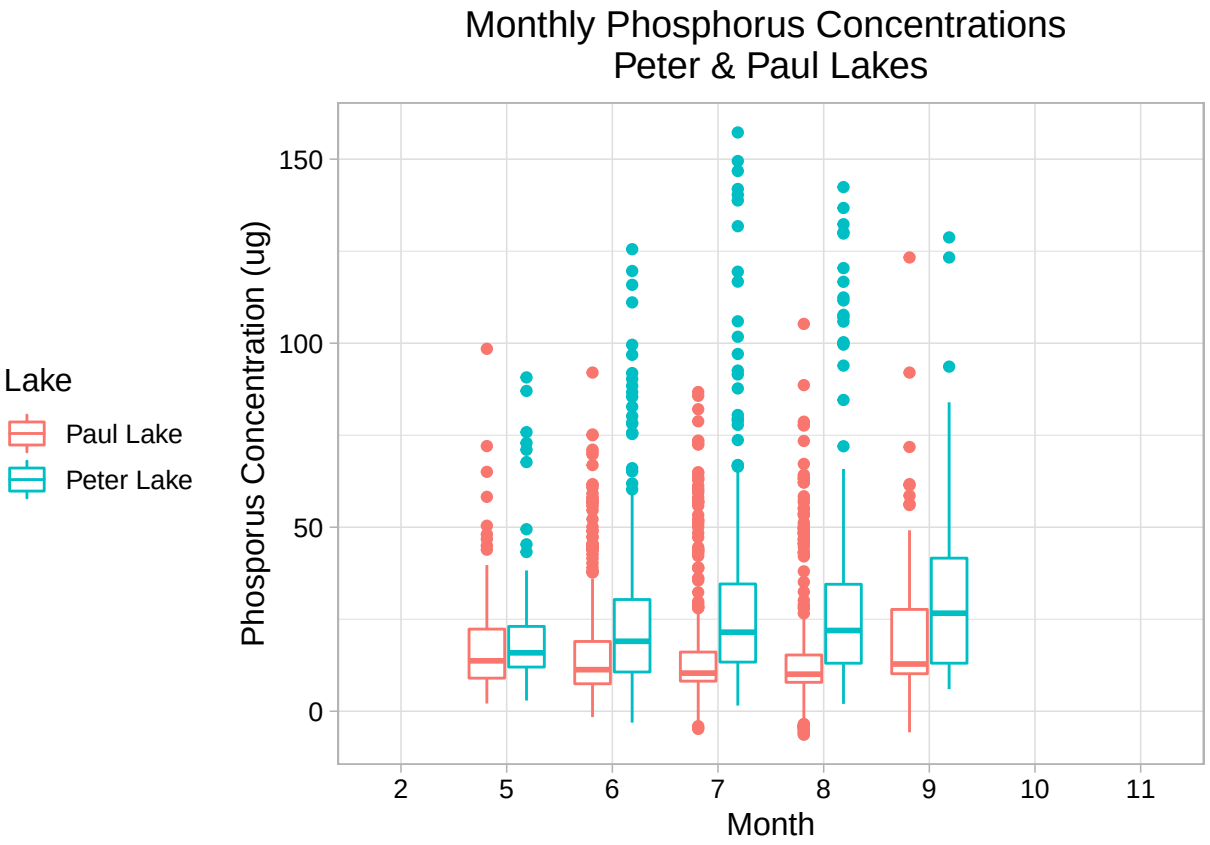
```
#temperature boxplot
Temp.Box <-
  ggplot(Chem.Nutrients, aes(x =as.factor(month), y = temperature_C)) +
  geom_boxplot(aes(color = lakename))+
  labs(title="Monthly Temperature \nPeter & Paul Lakes", x="Month", y="Temperature (C)", color="Lake")
print(Temp.Box)
```

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



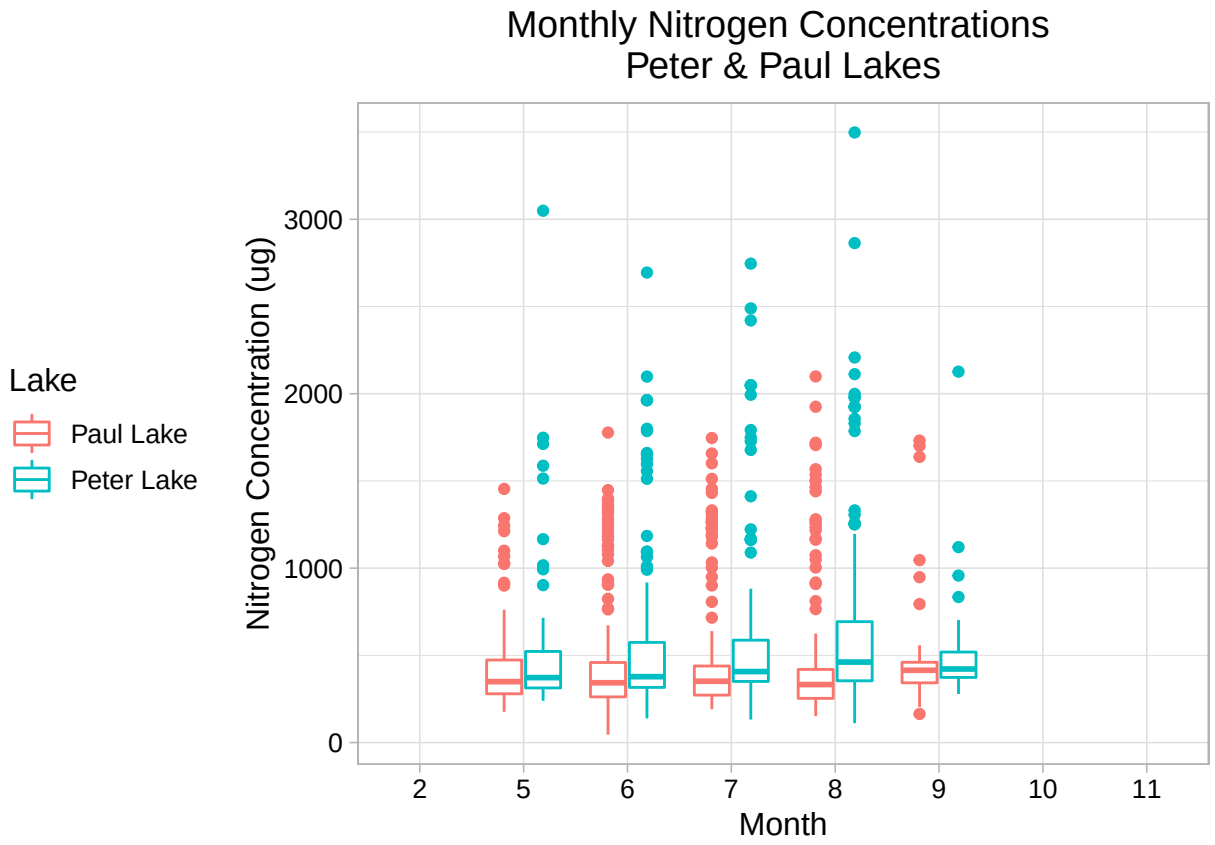

```
#total phosphorus boxplot
Phos.Box <-
  ggplot(Chem.Nutrients, aes(x = as.factor(month), y = tp_ug)) +
  geom_boxplot(aes(color = lakename))+
  labs(title="Monthly Phosphorus Concentrations \nPeter & Paul Lakes", x="Month", y="Phosporus Concentra
print(Phos.Box)
```

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




```
#total nitrogen boxplot
Nit.Box <-
  ggplot(Chem.Nutrients, aes(x = as.factor(month), y = tn_ug)) +
  geom_boxplot(aes(color = lakename))+
  labs(title="Monthly Nitrogen Concentrations \nPeter & Paul Lakes", x="Month", y="Nitrogen Concentrati
print(Nit.Box)
```

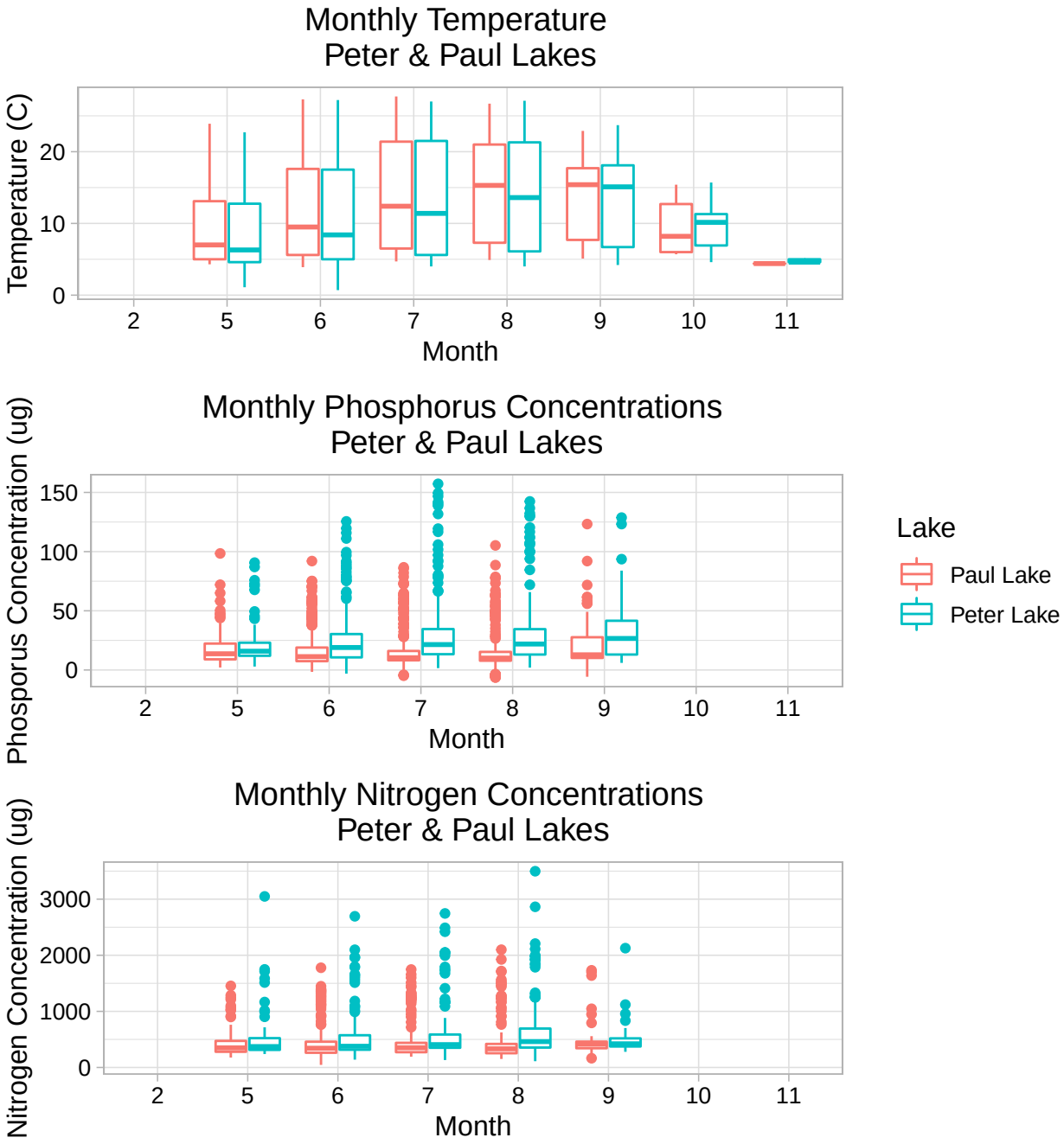
```
## Warning: Removed 21583 rows containing non-finite values (stat_boxplot).
```




```
#arrange all plots together
plot.boxes <- plot_grid(
  Temp.Box + theme(legend.position = "none"),
  Phos.Box + theme(legend.position = "none"),
  Nit.Box + theme(legend.position = "none"),
  nrow = 3
)

#define main legend for mega plot
legend <- get_legend(
  Temp.Box+theme(legend.box.margin = margin(0,0,0,12)))

#add legend
plot_grid(plot.boxes, legend, rel_widths = c(2,0.5))
```



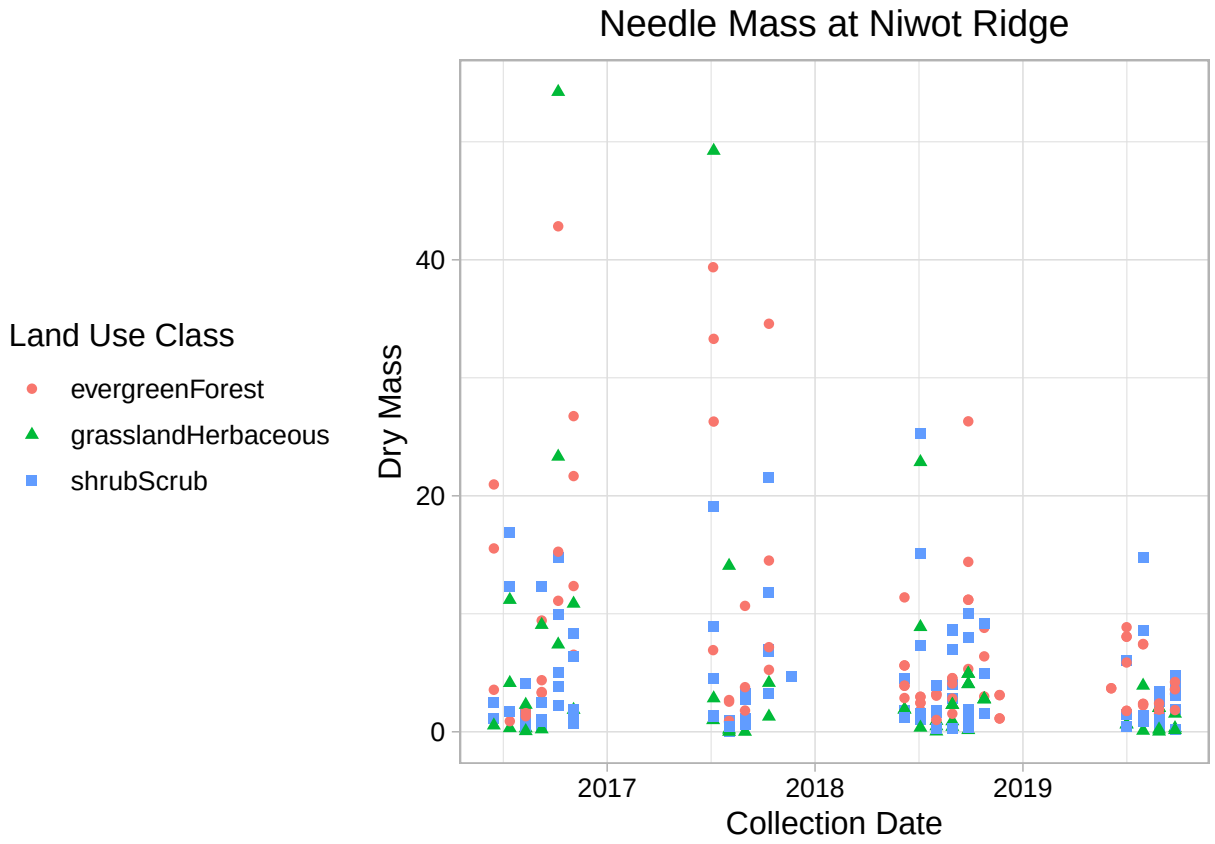
Question: What do you observe about the variables of interest over seasons and between lakes?

Answer: Temperature range between Peter and Paul Lake is similar for each of ther erpoted months. Median temperature between the two lakes is roughly similar for each of the reported months, however, there is a larger difference in October. Paul Lake has a lower median temperature. However, there is a greater difference in monthly phosphorus concentrations, and there are a significant number of outliers within each month between the lakes. Paul lake has some lower phosphorus concentration values that are outliers and Peter lake does not. These lower outliers occur in late summer, in July and August. Likewise, the majority of outliers also occur during late summmer. Similarly, the nitrogen concentrations between the two lakes have outliers in each month. However, Peter lake has significantly higher outliers. Higher maximum values

and outliers also occur more often in late summer.

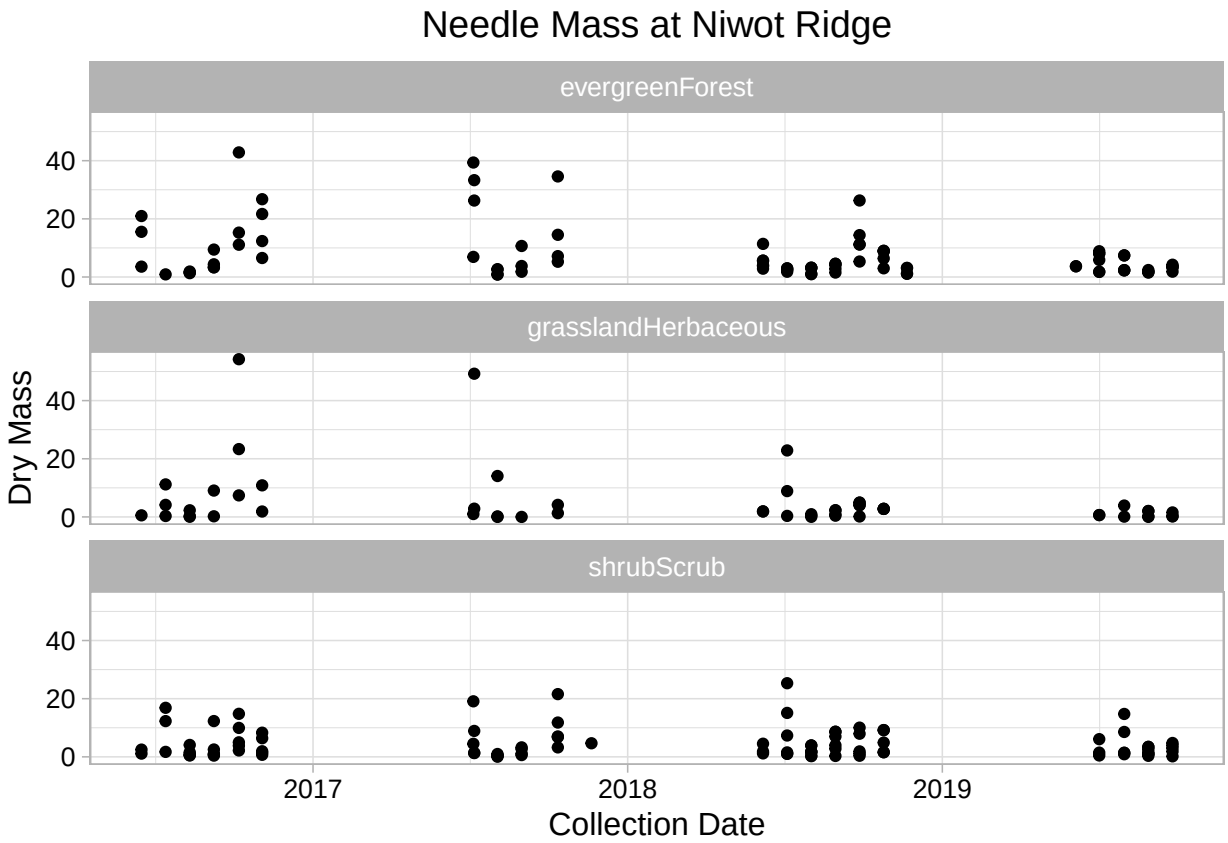
6. [Niwot Ridge] Plot a subset of the litter dataset by displaying only the "Needles" functional group. Plot the dry mass of needle litter by date and separate by NLCD class with a color aesthetic. (no need to adjust the name of each land use)

```
Needles <- ggplot(filter(Litter, functionalGroup=="Needles"), aes(x=collectDate, y=dryMass))+
  geom_point(aes(color=nlcdClass, shape=nlcdClass))+
  labs(title="Needle Mass at Niwot Ridge", x="Collection Date", y="Dry Mass", color="Land Use Class", sh
print(Needles)
```



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

```
Needles.2 <- ggplot(filter(Litter, functionalGroup=="Needles"), aes(x=collectDate, y=dryMass))+
  geom_point()+
  facet_wrap(vars(nlcdClass), nrow=3)+
  labs(title="Needle Mass at Niwot Ridge", x="Collection Date", y="Dry Mass")
print(Needles.2)
```



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

Answer:The faceted data (plot 7) is able to represent the different land use classes more effectively. However, plot 6, allows you to visualize the difference in dry mass of needles by viewing them all together. Although, it is more difficuly to differentiate (plot 6).