Assignment 5: Data Visualization

Jess Garcia

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_ChemistryNutrients_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. Confirm directory, load packages, & upload data

#Confirm workingdirectory
getwd()
```

```
## [1] "C:/Users/93jes/Documents/ENV872/Environmental_Data_Analytics_2021"
```

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

```
##
## Attaching package: 'lubridate'
## The following object is masked from 'package:cowplot':
##
##     stamp
## The following objects are masked from 'package:base':
##
##     date, intersect, setdiff, union
#Load data
ChemistryNutrients_PeterPaul <- read.csv("./Data/Processed/NTL-LTER_Lake_Chemistry_Nutrients_PeterPaul_
PeterPaul_Nutrients_Gathered <- read.csv("./Data/Processed/NTL-LTER_Lake_Nutrients_PeterPaulGathered_Pr
Niwot_Littertrap <- read.csv("./Data/Processed/NEON_NIWO_Litter_mass_trap_Processed.csv", stringsAsFact


#2. Confirm dates as date objects
class(ChemistryNutrients_PeterPaul$sampledate)
## [1] "factor"
class(PeterPaul_Nutrients_Gathered$sampledate)
## [1] "factor"
class(Niwot_Littertrap$collectDate)
## [1] "factor"
#Change from factor objects to date objects

ChemistryNutrients_PeterPaul$sampledate <- as.Date(ChemistryNutrients_PeterPaul$sampledate, format = "%
PeterPaul_Nutrients_Gathered$sampledate <- as.Date(PeterPaul_Nutrients_Gathered$sampledate, format = "%
Niwot_Littertrap$collectDate <- as.Date(Niwot_Littertrap$collectDate, format = "%Y-%m-%d")
```

Define your theme

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

```
#https://ggplot2.tidyverse.org/reference/theme.html

#Build default theme
JG_default_theme <- theme_bw(base_size = 12)+
  theme(axis.text = element_text(color = "black"),
        plot.title = element_text(color = "black", size = 16, face = "bold",
        hjust = 0.5))
```

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.

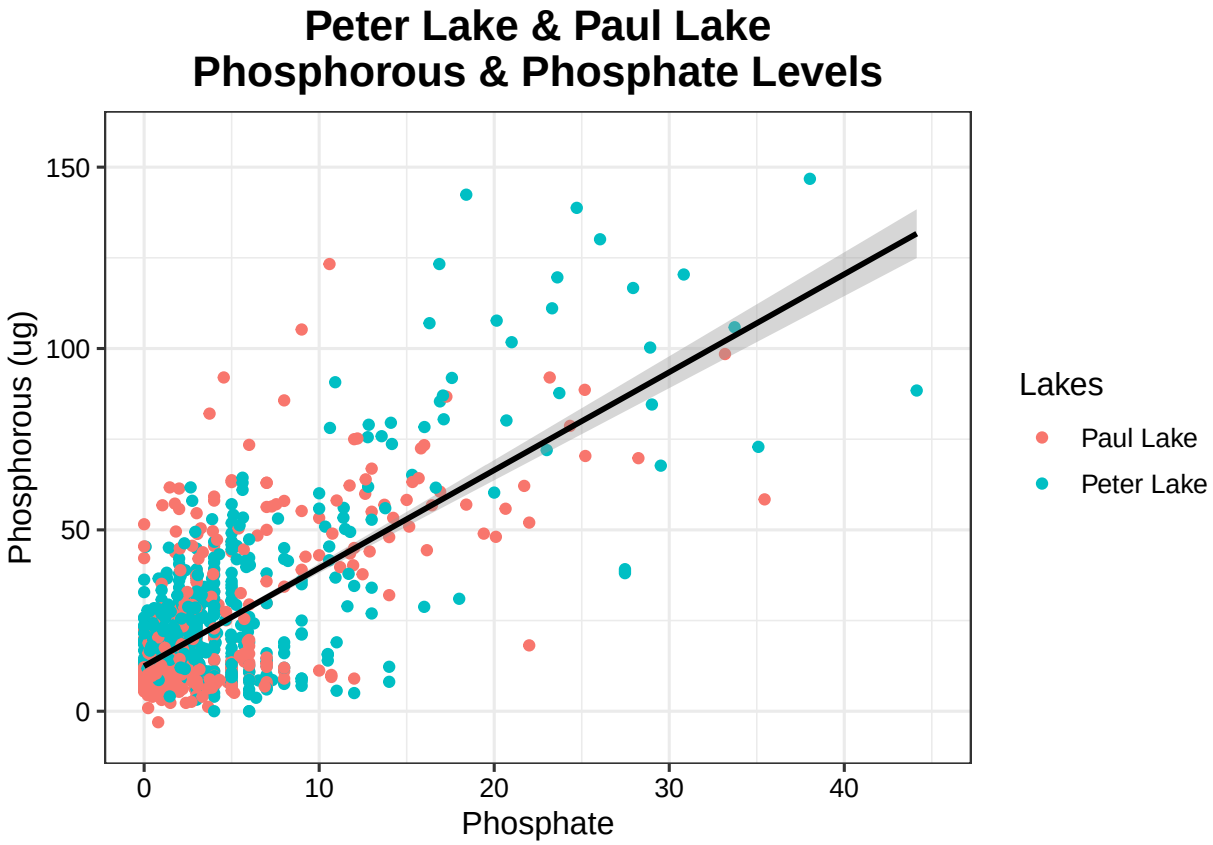
4. [NTL-LTER] Plot total phosphorus (tp_ug) by phosphate (po4), with separate aesthetics for Peter and Paul lakes. Add a line of best fit and color it black. Adjust your axes to hide extreme values.

```
#4. Create scatterplot for phosphorous and phosphate at Pete & Paul Lakes
PeterPaul.tp_ug.po4.plot <-
  ggplot(ChemistryNutrients_PeterPaul, aes(x=po4, y = tp_ug))+
  geom_point(aes(color=lakename))+
  JG_default_theme+
  scale_x_date(date_breaks = "2 months", date_labels = "%b %Y")+
  ggtitle("Peter Lake & Paul Lake\nPhosphorous & Phosphate Levels")+
  theme(plot.title = element_text(hjust = 0.5))+
  labs(x = "Phosphate", y= "Phosphorous (ug)", color = "Lakes")+
  xlim(0, 45)+
  geom_smooth(method = lm, color = "black")
```

```
## Scale for 'x' is already present. Adding another scale for 'x', which will
## replace the existing scale.
```

```
print(PeterPaul.tp_ug.po4.plot)
```

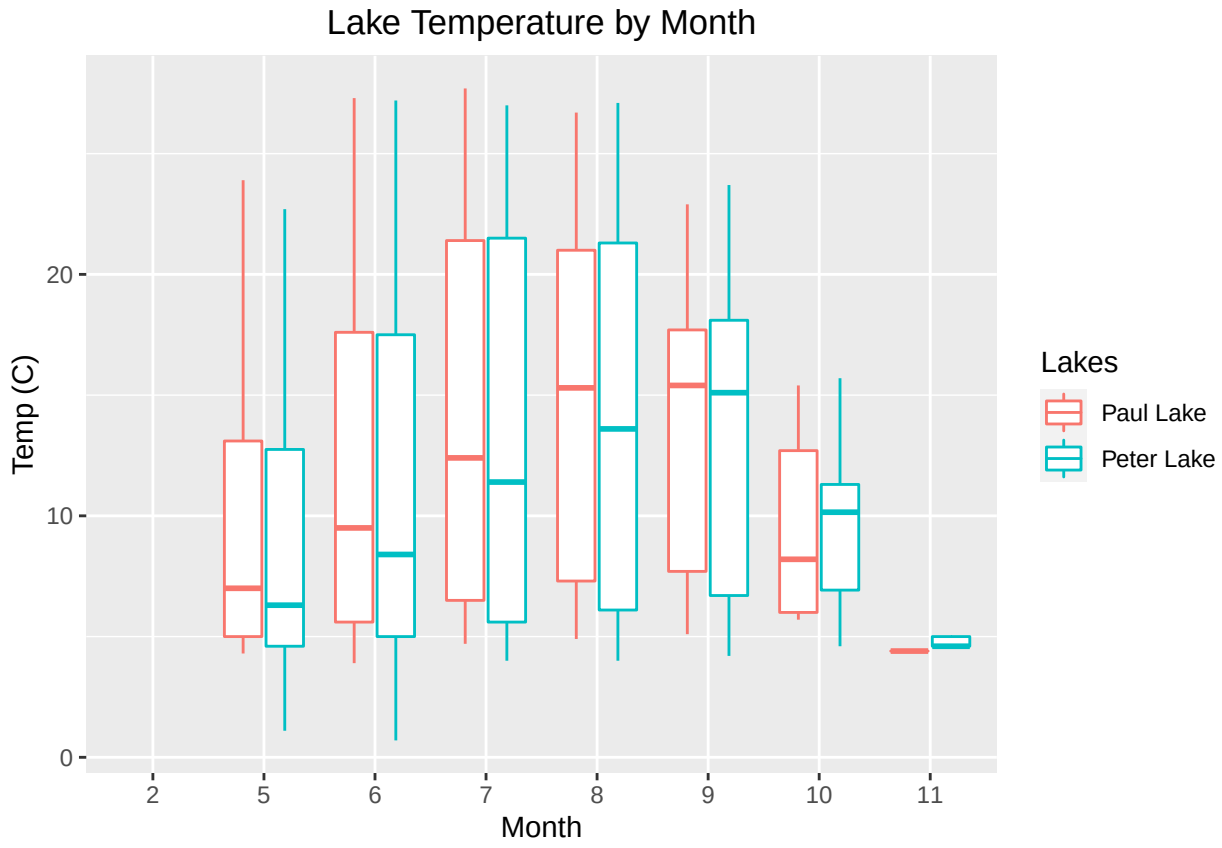
```
## `geom_smooth()` using formula 'y ~ x'
```



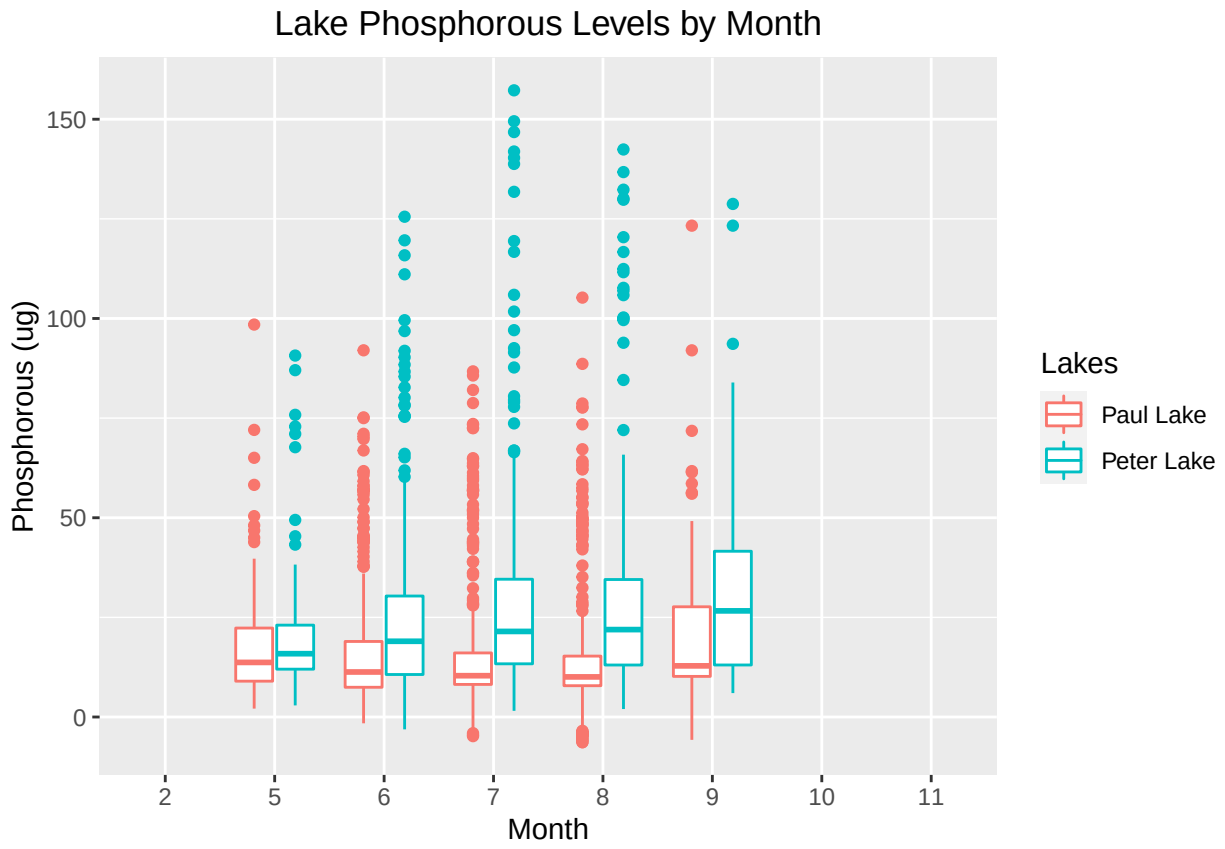
5. [NTL-LTER] Make three separate boxplots of (a) temperature, (b) TP, and (c) TN, with month as the x axis and lake as a color aesthetic. Then, create a cowplot that combines the three graphs. Make sure that only one legend is present and that graph axes are aligned.

```
#5. Create three separate boxplots
#First make month as factor and not integer
ChemistryNutrients_PeterPaul$month <- as.factor(ChemistryNutrients_PeterPaul$month)
```

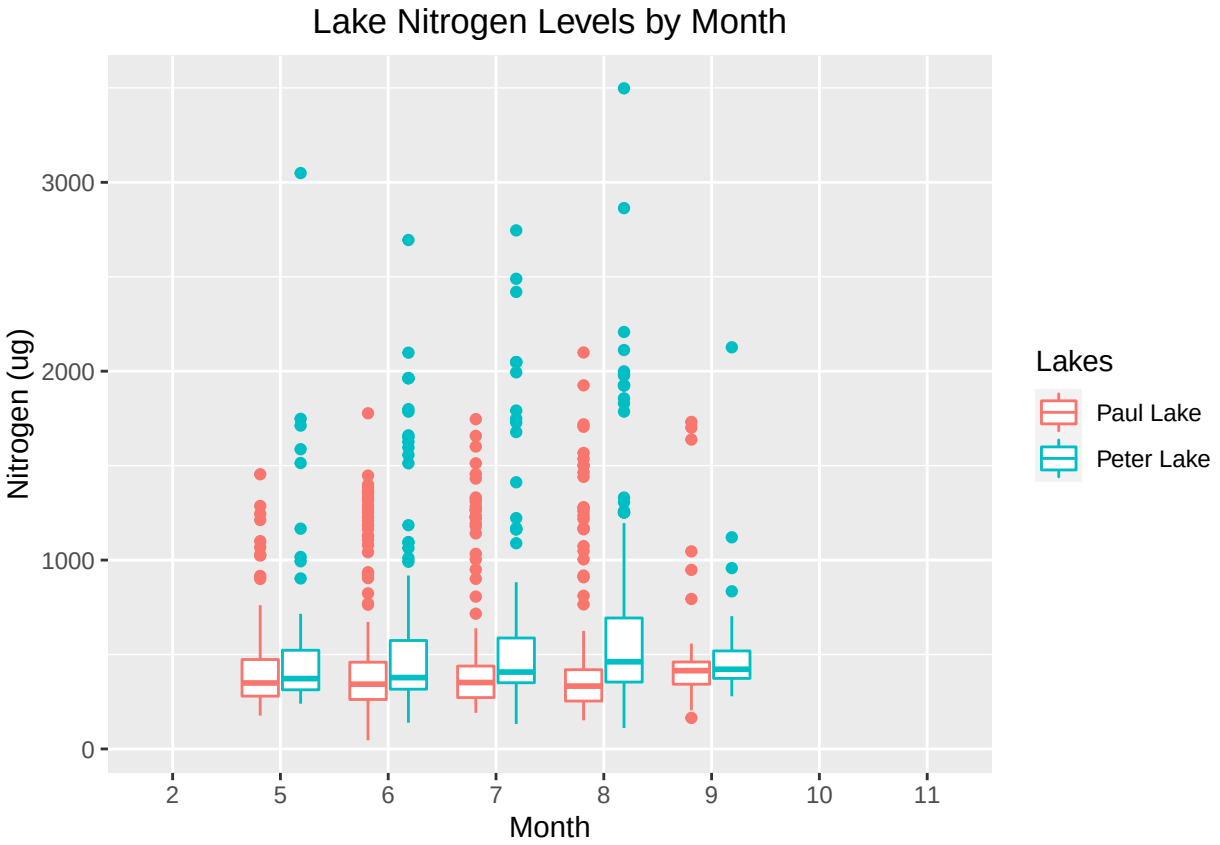
```
#Boxplot for temperature
PeterPaul_Temperature.boxplot <-
ggplot(ChemistryNutrients_PeterPaul, aes(x=month, y= temperature_C, color =lakename))+
  geom_boxplot()+
  ggtitle("Lake Temperature by Month")+
  theme(plot.title = element_text(hjust = 0.5))+
  labs(x= "Month", y = "Temp (C)", color = "Lakes")
print(PeterPaul_Temperature.boxplot)
```



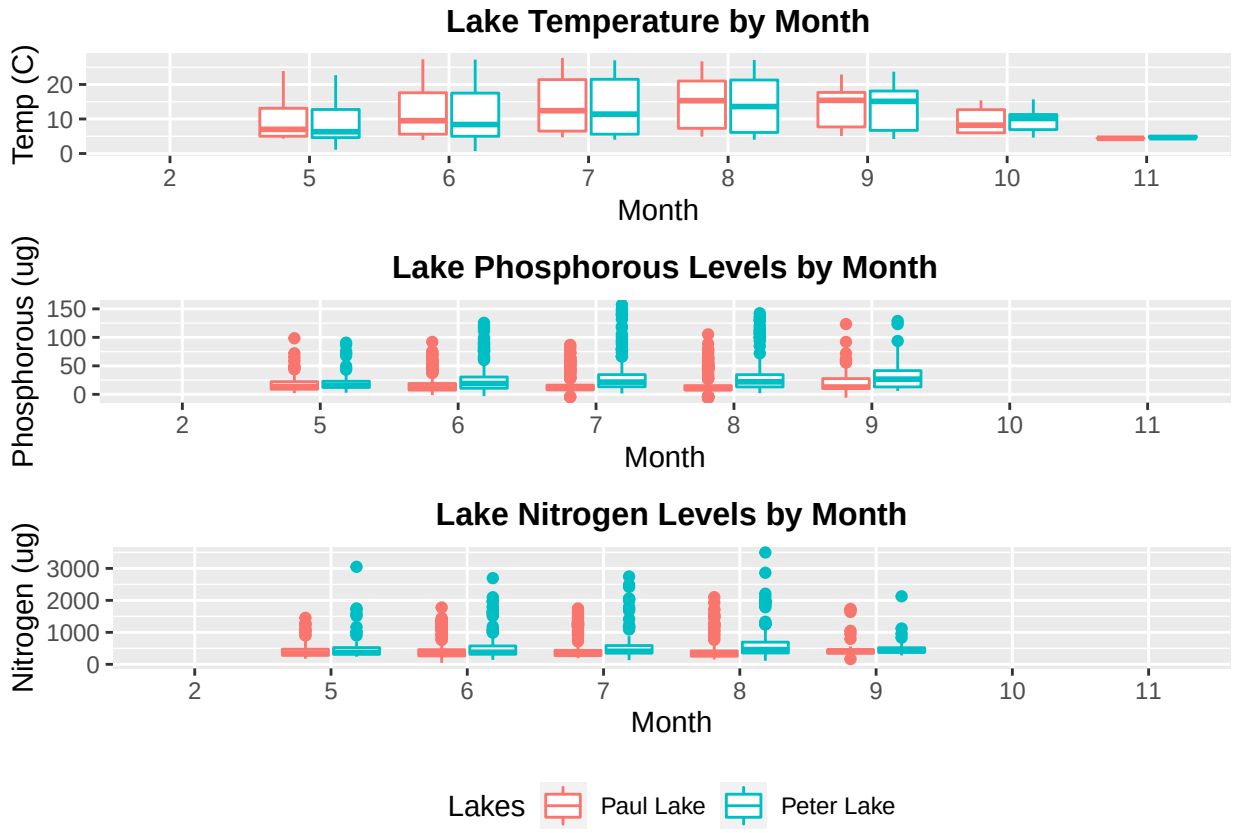

```
#Boxplot for TP
PeterPaul_TP.boxplot <-
ggplot(ChemistryNutrients_PeterPaul, aes(x=month, y= tp_ug, color =lakename))+
  geom_boxplot()+
ggtitle("Lake Phosphorous Levels by Month")+
  theme(plot.title = element_text(hjust = 0.5))+
  labs(x= "Month", y = "Phosphorous (ug)", color = "Lakes")
print(PeterPaul_TP.boxplot)
```




```
#Boxplot for TN
PeterPaul_TN.boxplot <-
ggplot(ChemistryNutrients_PeterPaul, aes(x=month, y= tn_ug, color =lakename))+
  geom_boxplot()+
  ggtitle("Lake Nitrogen Levels by Month")+
  theme(plot.title = element_text(hjust = 0.5))+
  labs(x= "Month", y = "Nitrogen (ug)", color = "Lakes")
print(PeterPaul_TN.boxplot)
```




```
#Combine the threee boxplots using cowplot
PeterPaul_combined_boxplot <-
plot_grid(PeterPaul_Temperature.boxplot +theme(legend.position="none")+ theme(plot.title = element_text
PeterPaul_TP.boxplot  +theme(legend.position="none")+ theme(plot.title = element_text(hjust = 0.5, size
PeterPaul_TN.boxplot +theme(legend.position ="bottom")+ theme(plot.title = element_text(hjust = 0.5, si
nrow=3, align = 'h', rel_heights = c(1,1,1.5))

print(PeterPaul_combined_boxplot)
```



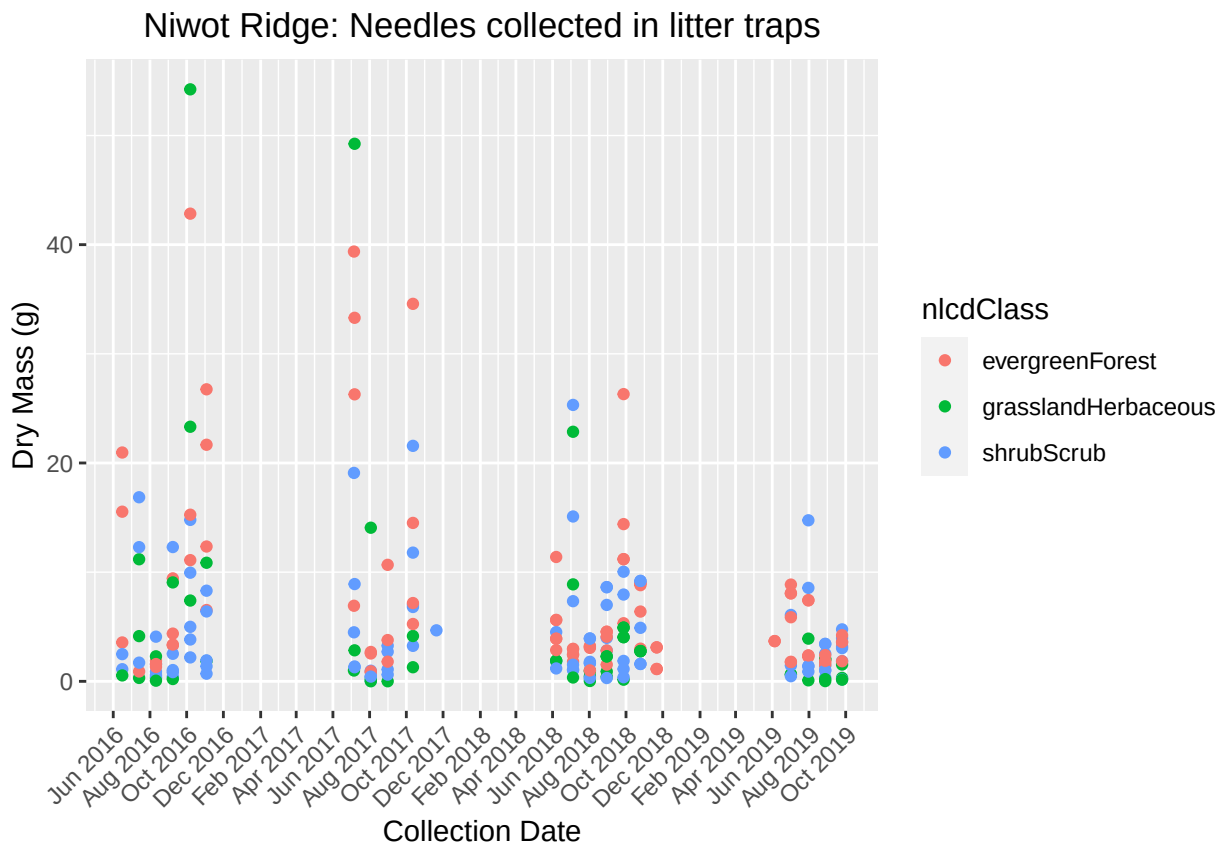
Question: What do you observe about the variables of interest over seasons and between lakes?

> Answer: In the warmer months there are higher concentrations of phosphorous and nitrogen in these lakes. Also, Peter Lake tends to have higher levels of nitrogen and phosphorous regardless of the month compared to Paul Lake; i.e. the median concentration values are higher in Peter Lake than Paul Lake.

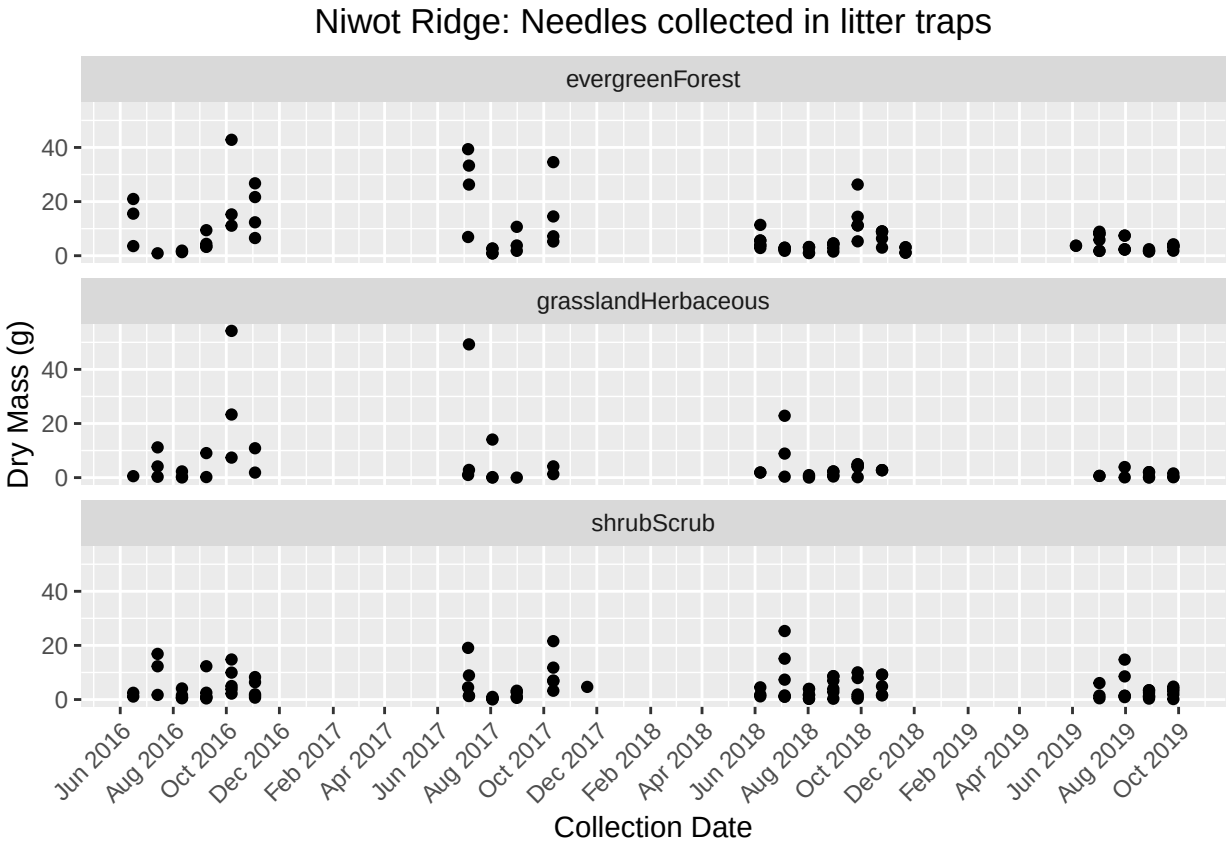
6. [Niwot Ridge] Plot a subset of the litter dataset by displaying only the "Needles" functional group. Plot the dry mass of needle litter by date and separate by NLCD class with a color aesthetic. (no need to adjust the name of each land use)

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

```
#6. Niwot Ridge - Needles plot w/NLCD classes separated by color
Niwot_NeedlesMass_color.plot <-
  ggplot(subset(Niwot_Littertrap, functionalGroup == "Needles"),
  aes(x=collectDate, y=dryMass, color = nlcdClass))+
  geom_point()+
  scale_x_date(date_breaks = "2 months", date_labels = "%b %Y")+
  theme(axis.text.x = element_text(angle=45, hjust=1.0))+
  ggtitle("Niwot Ridge: Needles collected in litter traps")+
  theme(plot.title = element_text(hjust = 0.5))+
  labs(x = "Collection Date", y= "Dry Mass (g)")

print(Niwot_NeedlesMass_color.plot)
```




```
#7. Niwot Ridge - Needles plot w/NLCD classes separated by facets

Niwot_NeedlesMass_facets.plot <-
  ggplot(subset(Niwot_Littertrap, functionalGroup == "Needles"),
  aes(x=collectDate, y=dryMass))+
  geom_point()+
  facet_wrap(vars(nlcdClass), nrow =3)+
  scale_x_date(date_breaks = "2 months", date_labels = "%b %Y")+
    theme(axis.text.x = element_text(angle=45, hjust=1.0))+
  ggtitle("Niwot Ridge: Needles collected in litter traps")+
    theme(plot.title = element_text(hjust = 0.5))+
  labs(x = "Collection Date", y= "Dry Mass (g)")

print(Niwot_NeedlesMass_facets.plot)
```



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

Answer: I think plot 7 is probably more effective than plot 6 only because I am able to tell more clearly how the type of forest (nlcdClass) relates to the dry mass collected. Both 6 and 7 make it obvious which months do not have litter collected so they are even there. Plot 6 is more useful if I'm interested in generally seeing what the spread of dry mass collected is generally, but I think in this case the site does matter and so plot 7 is more helpful.