Assignment 5: Data Visualization

Kaitlyn Elliott Section 4

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] "C:/Users/15408/Documents/Environmental_Data_Analytics_2022"
```

```
library(tidyverse)
```

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

```
## v ggplot2 3.3.3     v purrr   0.3.4
## v tibble  3.1.6     v dplyr   1.0.7
## v tidyr   1.1.3     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)
ntl_lter_peter_paul_chemistry<-read.csv("./Data/Processed/NTL-LTER_Lake_Chemistry_Nutrients_PeterPaul_P

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

```
#2
ntl_lter_peter_paul_chemistry$sampledate<-as.Date(ntl_lter_peter_paul_chemistry$sampledate)
niwot_litter_prodessed$collectDate<-as.Date(niwot_litter_prodessed$collectDate)
```

Define your theme

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

```
#3
theme_groovy<-theme_classic(base_size = 12) +
  theme(axis.text = element_text(color = "orange"),
        legend.position = "right")
theme_set(theme_groovy)
```

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.

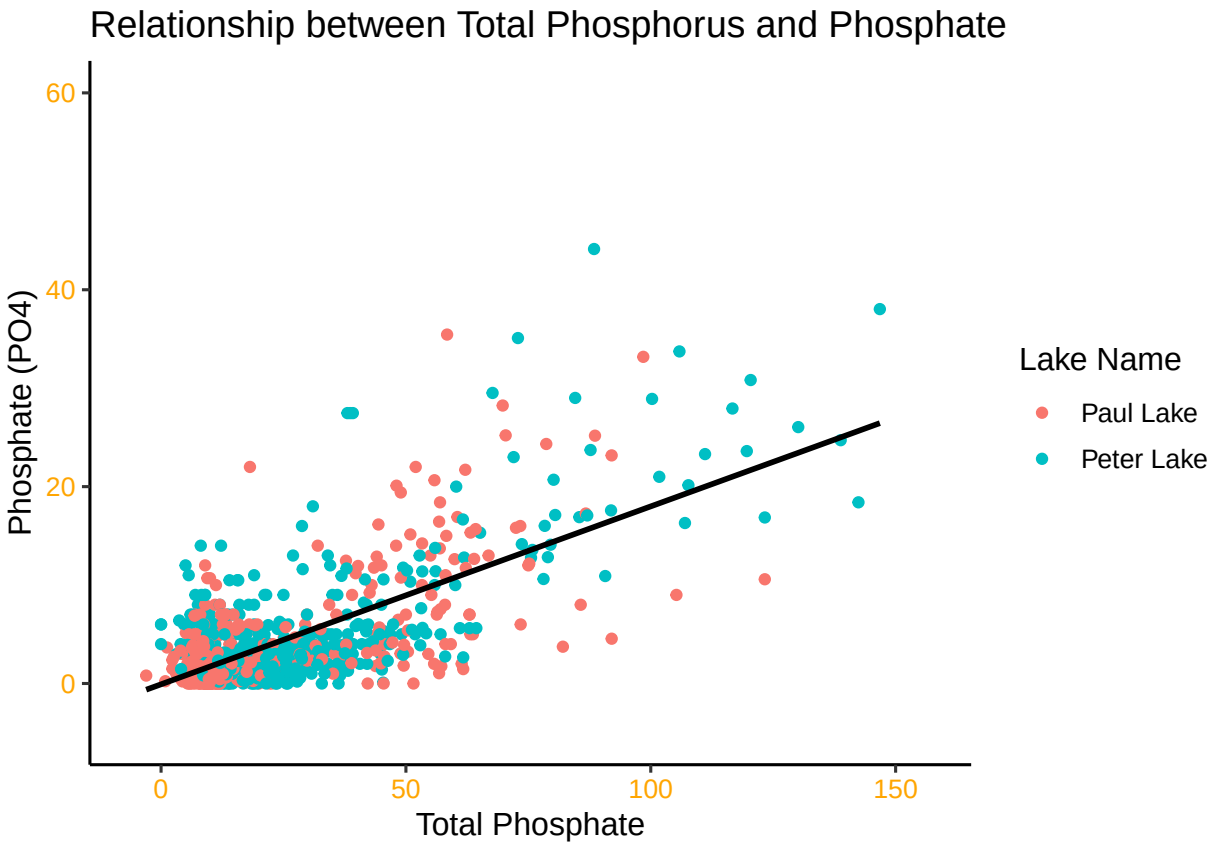
4. [NTL-LTER] Plot total phosphorus (`tp_ug`) by phosphate (`po4`), with separate aesthetics for Peter and Paul lakes. Add a line of best fit and color it black. Adjust your axes to hide extreme values (hint: change the limits using `xlim()` and `ylim()`).

```
#4
ggplot(ntl_lter_peter_paul_chemistry, aes(x=tp_ug, y=po4, color=lakename))+
  geom_point()+
  ylim(-5,60)+geom_smooth(method="lm", se=FALSE, color="black", aes(x=tp_ug, y=po4))+
  labs(title="Relationship between Total Phosphorus and Phosphate", color="Lake Name")+
  ylab("Phosphate (PO4)")+xlab("Total 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).
```

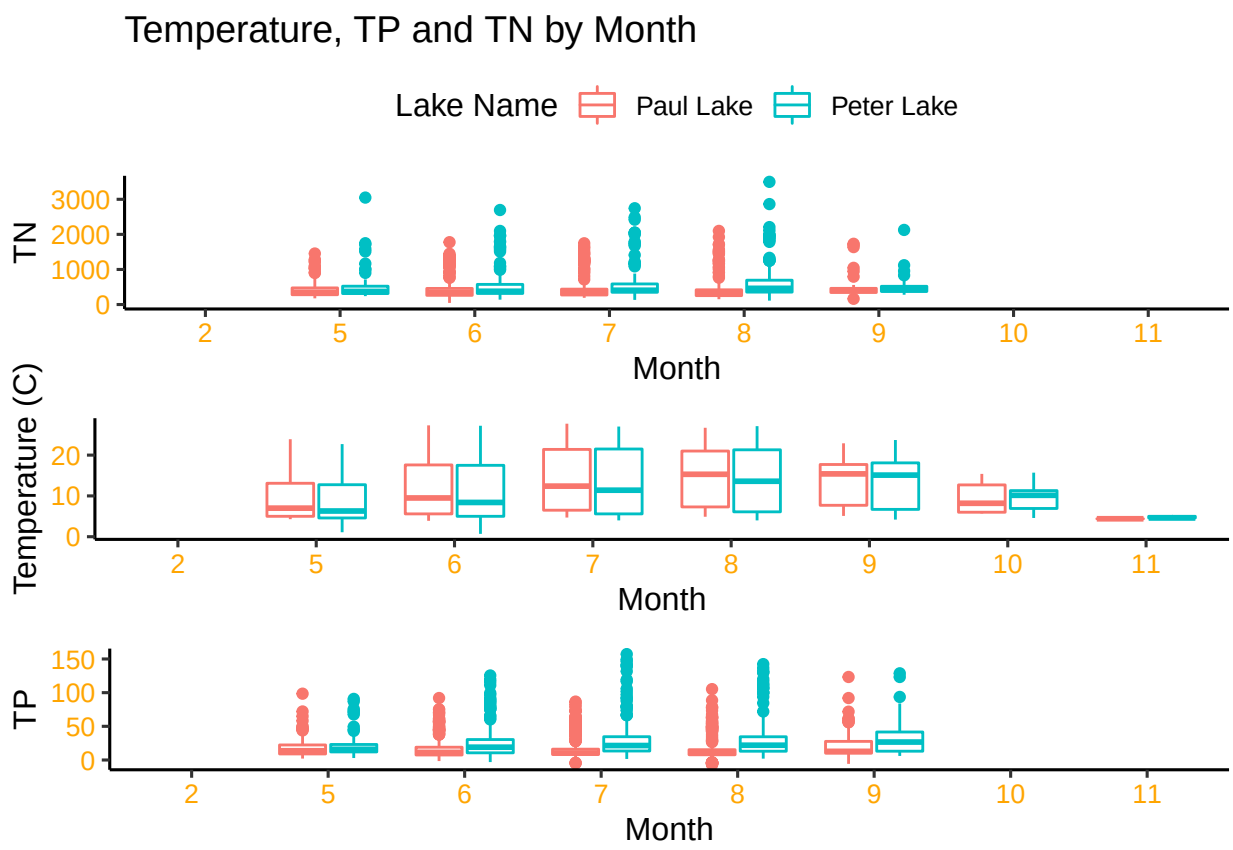


5. [NTL-LTER] Make three separate boxplots of (a) temperature, (b) TP, and (c) TN, with month as the x axis and lake as a color aesthetic. Then, create a cowplot that combines the three graphs. Make sure that only one legend is present and that graph axes are aligned.

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

temp_plot<-ggplot(ntl_lter_peter_paul_chemistry,aes(x=month,y=temperature_C, color=lakename))+
  geom_boxplot()+labs( color="Lake Name")+ylab("Temperature (C)")+xlab("Month")+
  theme(legend.position = "none")

tp_plot<-ggplot(ntl_lter_peter_paul_chemistry,aes(x=month,y=tp_ug,color=lakename))+
  geom_boxplot()+ylab("TP")+xlab("Month")+theme(legend.position = "none")

tn_plot<-ggplot(ntl_lter_peter_paul_chemistry,aes(x=month,y=tn_ug,color=lakename))+
  geom_boxplot()+ylab("TN")+xlab("Month")+
  labs(title = "Temperature, TP and TN by Month", color="Lake Name")+
  theme(legend.position = "top")

plot_grid(tn_plot,temp_plot,tp_plot, nrow=3, align = "h", rel_heights = c(1.75,1,1))
```

```
## Warning: Removed 21583 rows containing non-finite values (stat_boxplot).
## Warning: Removed 3566 rows containing non-finite values (stat_boxplot).
## Warning: Removed 20729 rows containing non-finite values (stat_boxplot).
## Warning: Graphs cannot be horizontally aligned unless the axis parameter is set.
```

```
## Placing graphs unaligned.
```



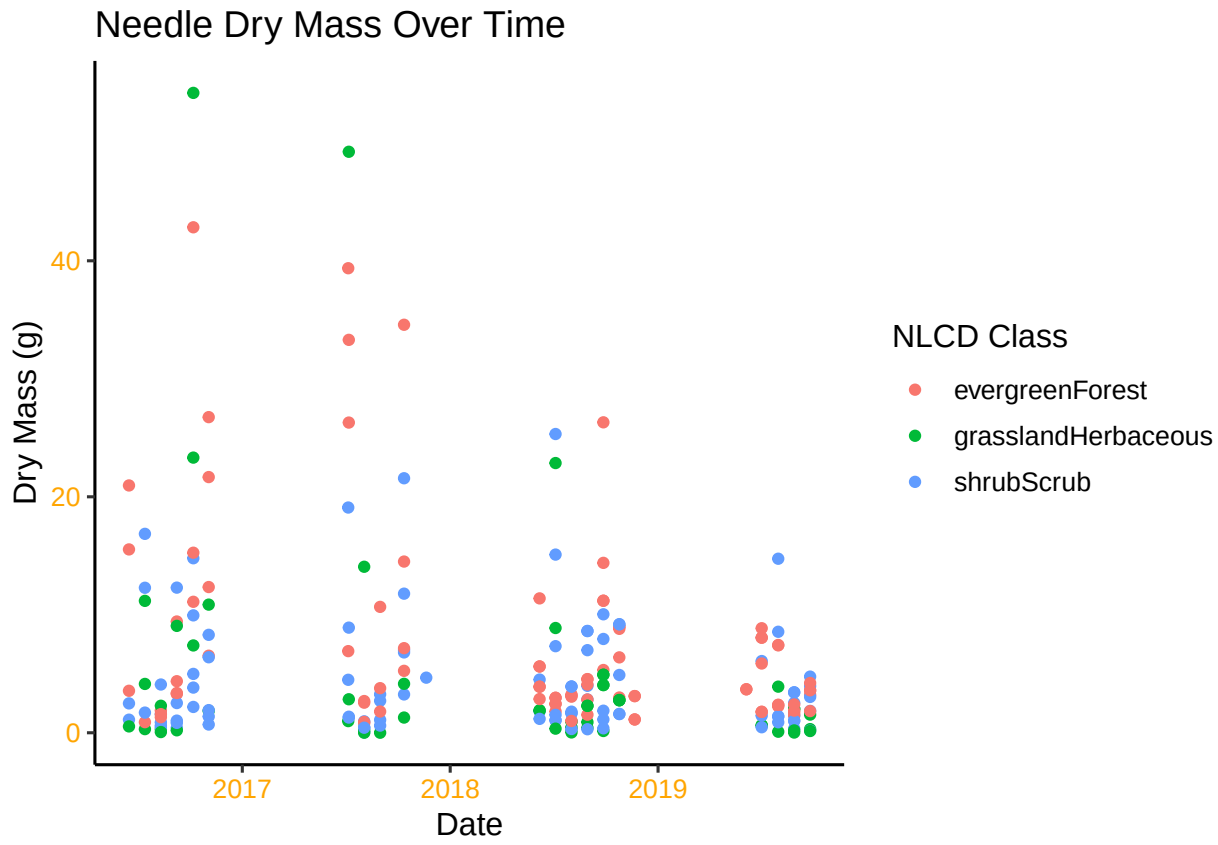
Question: What do you observe about the variables of interest over seasons and between lakes?

> Answer: Peter lake tends to have more nitrogen than Paul lake throughout the year. In both lakes you can see a seasonal temperature change with July and August being the warmest months and Paul lake is slightly warmer on average. I don't notice much of a seasonal differene in TN and TP, except for it slightly increases from May until September.

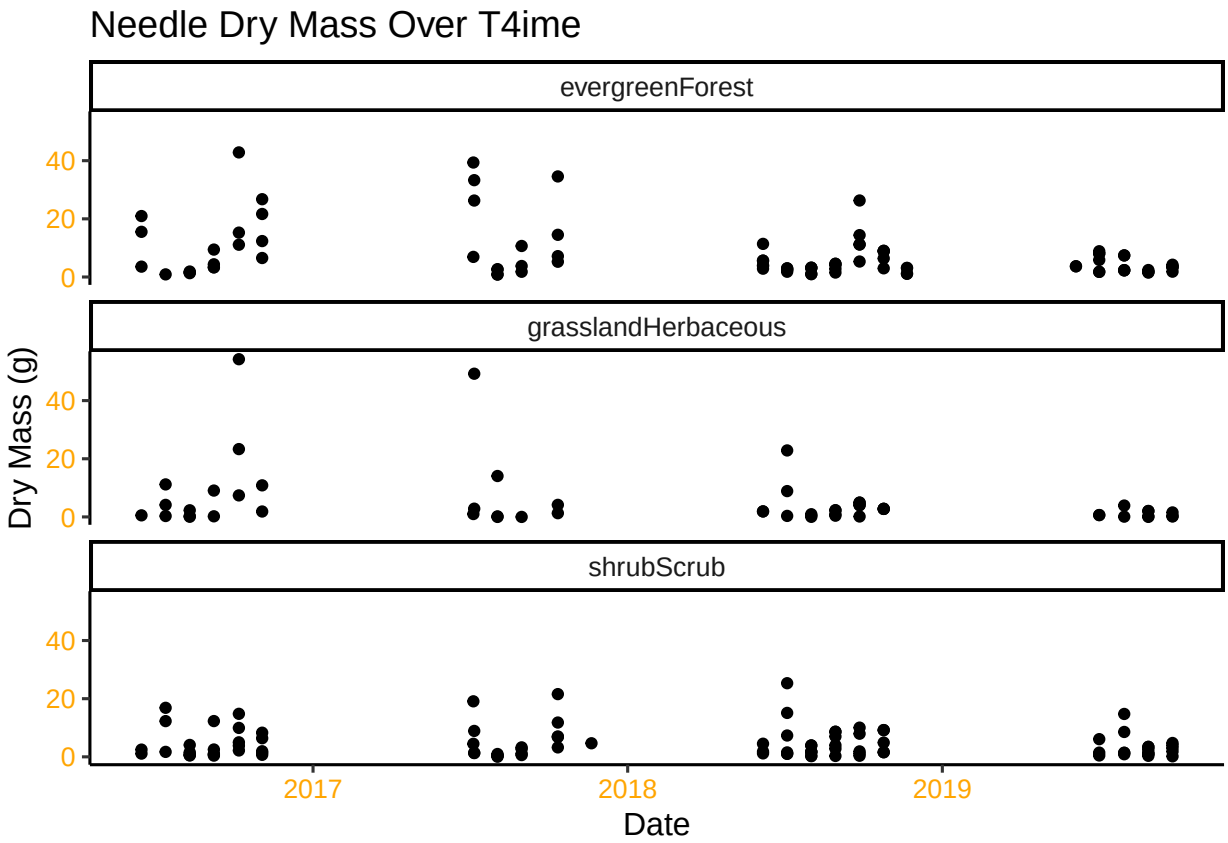
6. [Niwot Ridge] Plot a subset of the litter dataset by displaying only the "Needles" functional group. Plot the dry mass of needle litter by date and separate by NLCD class with a color aesthetic. (no need to adjust the name of each land use)

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

```
#6
ggplot(subset(niwot_litter_prodessed, functionalGroup== "Needles"),
  aes(x=collectDate, y=dryMass, color=nlcdClass))+geom_point()+
  labs(title="Needle Dry Mass Over Time", color="NLCD Class")+
  ylab("Dry Mass (g)")+xlab("Date")
```




```
#7
ggplot(subset(niwot_litter_prodessed, functionalGroup== "Needles"),
       aes(x=collectDate, y=dryMass))+geom_point()+
  labs(title="Needle Dry Mass Over T4ime")+ylab("Dry Mass (g)")+xlab("Date")+
  facet_wrap(vars(nlcdClass), nrow=3)
```



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

Answer: I think 7 is more effective because the data is less crowded together and its easier to see the differences between the categories when they are not overlapping eachother.