Assignment 5: Data Visualization

Kathlyn MacDonald

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/kmac9/OneDrive/Documents/Duke/Year 1/Spring 2022/ENV Data/Environmental_Data_Analytics
```

```
setwd("c:/Users/kmac9/OneDrive/Documents/Duke/Year 1/Spring 2022/ENV Data/Environmental_Data_Analytics_

library(tidyverse)
```

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

## v ggplot2 3.3.5     v purrr   0.3.4
## v tibble  3.1.6     v dplyr   1.0.7
## v tidyr   1.1.4     v stringr 1.4.0
## v readr   2.1.1     v forcats 0.5.1
```

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

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

  Peter.Paul.chem.nutrients <-read.csv("./Data/Processed/NTL-LTER_Lake_Chemistry_Nutrients_PeterPaul_Pr


NEON.NIWO.litter <- read.csv("./Data/Processed/NEON_NIWO_Litter_mass_trap_Processed.csv", stringsAsFact

#2
Peter.Paul.chem.nutrients$sampledate <-as.Date(Peter.Paul.chem.nutrients$sampledate, format = "%Y-%m-%d


NEON.NIWO.litter$collectDate <- as.Date(NEON.NIWO.litter$collectDate, format = "%Y-%m-%d")
```

Define your theme

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

```
#3

mytheme <- theme_classic(base_size = 12)+ theme(axis.text = element_text(color = "black"),
```

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()`).

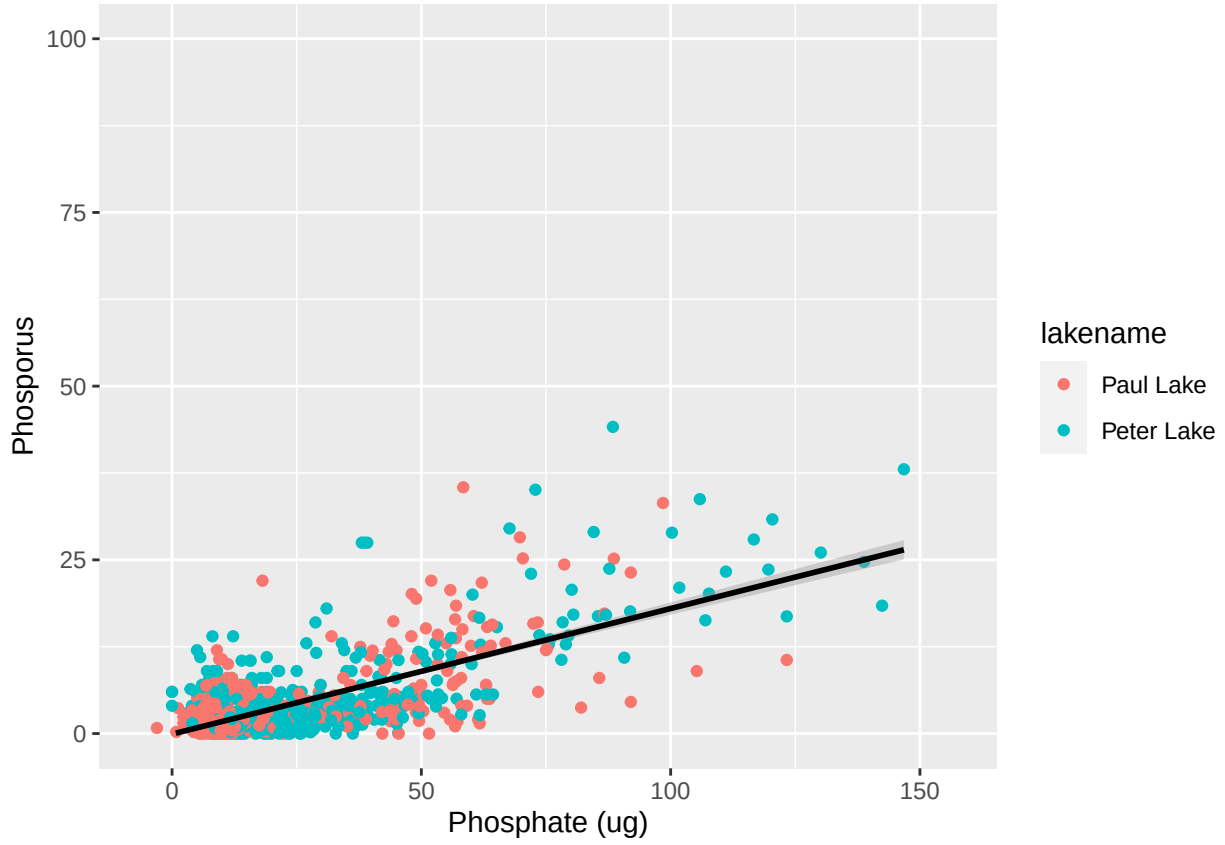
```
#4

NutrientPlot <-
  ggplot(Peter.Paul.chem.nutrients, aes(x = tp_ug, y = po4, color = lakename)) +
  geom_point() +
  geom_smooth(method= lm, color = "black") +
   ylim(0,100)+
  ylab("Phosporus")+
  xlab("Phosphate (ug)")
print(NutrientPlot)
```

```
## `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).
```

```
## Warning: Removed 2 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.

```
#5

Peter.Paul.chem.nutrients$month <- as.factor(Peter.Paul.chem.nutrients$month)

TemperaturePlot <- ggplot(Peter.Paul.chem.nutrients) +
  geom_boxplot(aes(x = month, y = temperature_C, color = lakename)) + xlab(NULL) +
  ylab("Temperature (C)") +
  geom_text(x = 1, y = 25,  label = "Temperature")

#print(TemperaturePlot)

Totalphosphorus <-ggplot(Peter.Paul.chem.nutrients) +
  geom_boxplot(aes(x= month, y =tp_ug, color = lakename)) +xlab(NULL)+
  geom_text(x = 1, y = 150,  label = "Total Phosphorous")+
  ylab("ug")

#print(Totalphosphorus)
```

```
Totalnitrogen <-ggplot(Peter.Paul.chem.nutrients) +
  geom_boxplot(aes(x= month, y = tn_ug, color = lakename)) +
  geom_text(x = 1, y = 3000,  label = "Total Nitrogen") +
  xlab("Month") +
  ylab("ug")
#print(Totalnitrogen)

NTL.legend <- get_legend(TemperaturePlot)
```

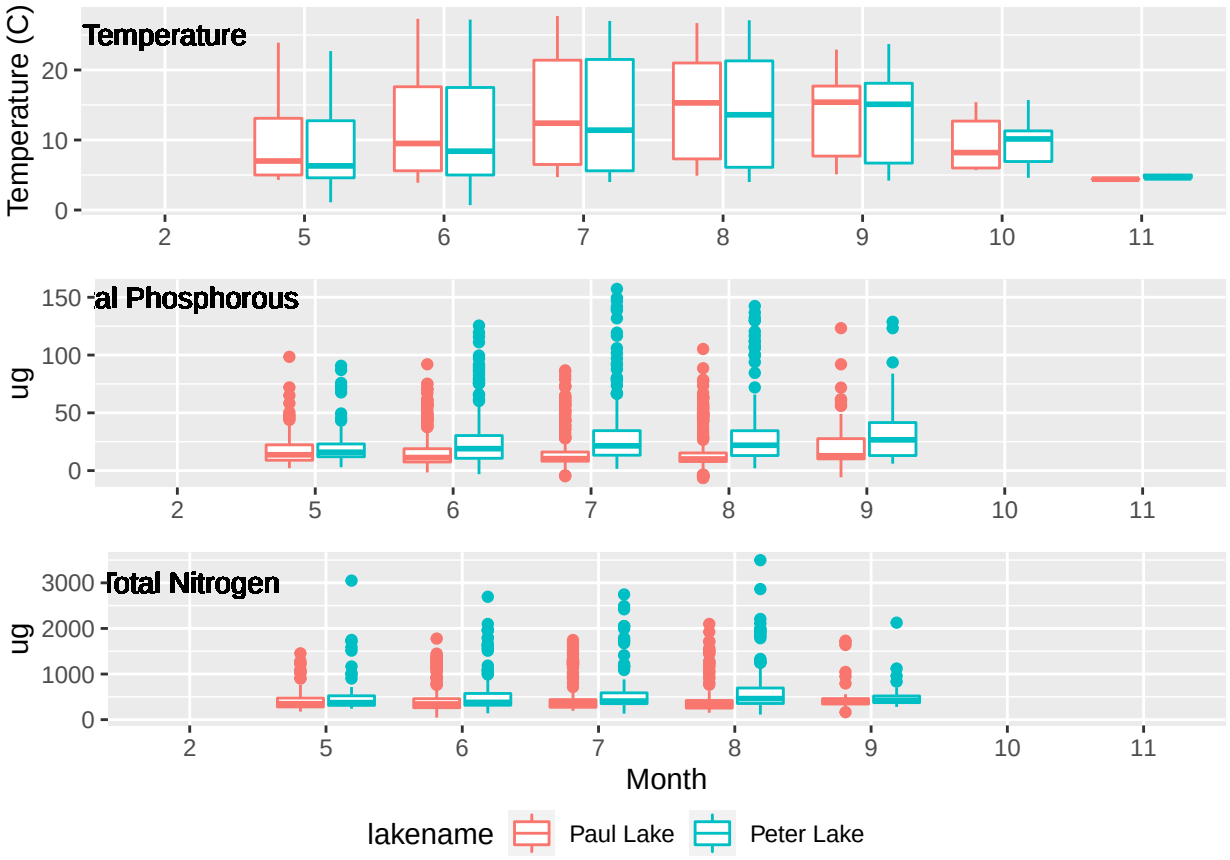
```
## Warning: Removed 3566 rows containing non-finite values (stat_boxplot).
```

```
Legend <- get_legend(
  TemperaturePlot +
    guides(color = guide_legend(nrow = 1)) +
    theme(legend.position ="bottom")
)
```

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

```
plot_grid(TemperaturePlot + theme(legend.position = "none"), Totalphosphorus + theme(legend.position =
```

```
## 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 vertically aligned unless the axis parameter is set.
## Placing graphs unaligned.

## Warning: Graphs cannot be horizontally aligned unless the axis parameter is set.
## Placing graphs unaligned.
```



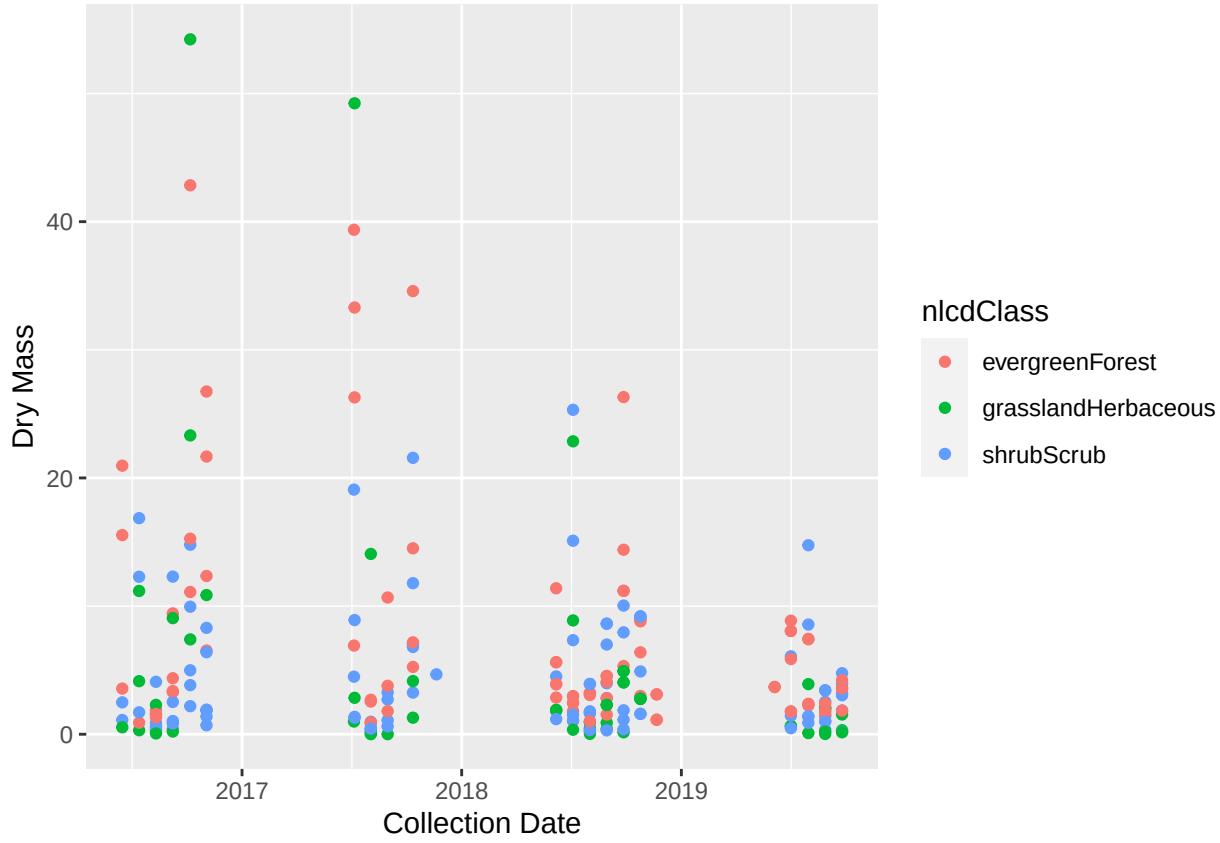
Question: What do you observe about the variables of interest over seasons and between lakes?

> Answer: Peter Lake tends to have higher levels of phosphorus than Paul Lake, throughout the warmer months that data was collected. Nitrogen levels remain fairly steady and similar for both lakes.

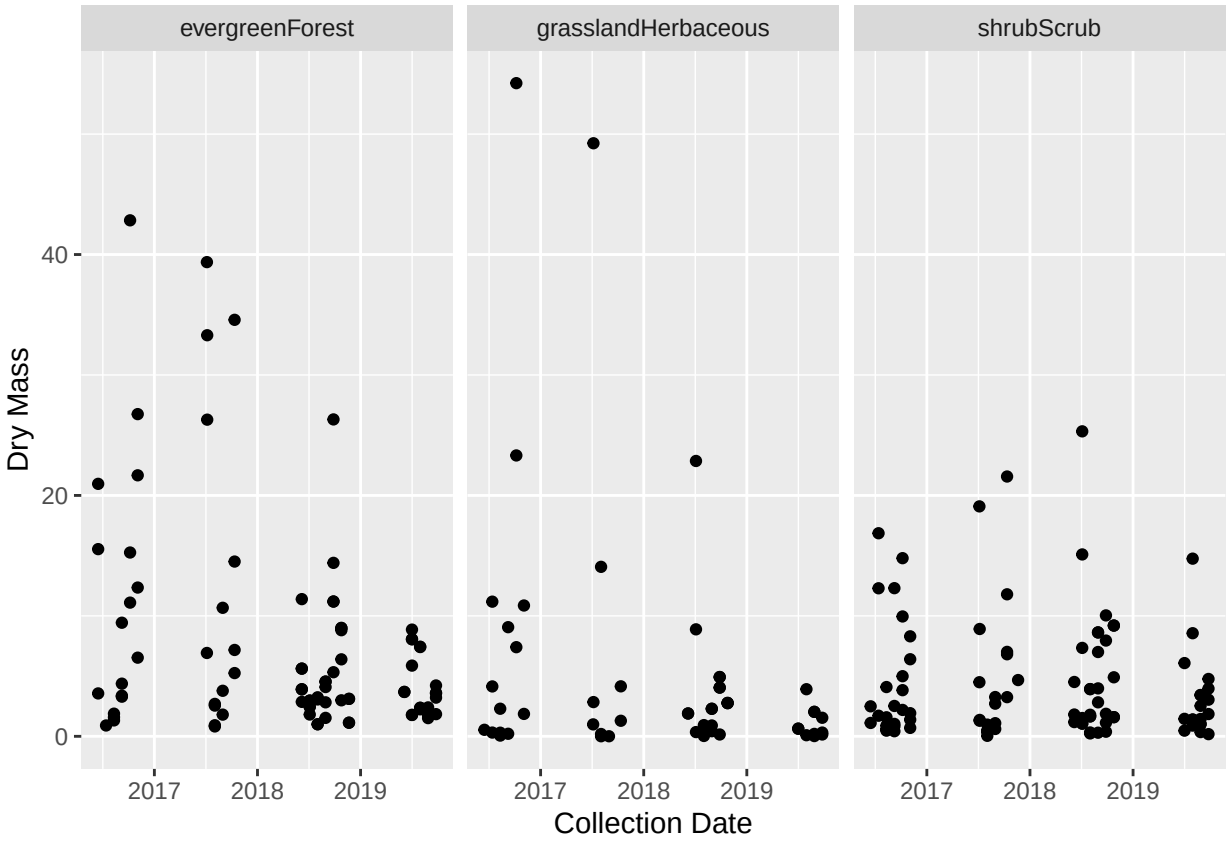
6. [Niwot Ridge] Plot a subset of the litter dataset by displaying only the "Needles" functional group. Plot the dry mass of needle litter by date and separate by NLCD class with a color aesthetic. (no need to adjust the name of each land use)

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

```
#6

LitterNeedlesPlot <-
  ggplot(subset(NEON.NIWO.litter, functionalGroup == "Needles"), aes(x=collectDate, y = dryMass, color =
  geom_point()+
  ylab("Dry Mass") +
  xlab("Collection Date")
print(LitterNeedlesPlot)
```




```
#7
LitterNeedlesPlotfaceted <-
  ggplot(subset(NEON.NIWO.litter, functionalGroup == "Needles"), aes(x=collectDate, y = dryMass))+
  geom_point()+
  ylab("Dry Mass") +
  xlab("Collection Date") +
  facet_wrap(~nlcdClass)
print(LitterNeedlesPlotfaceted)
```



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

Answer: The faceted plot is much easier to read, it is hard to distinguish between all the colors when they are mixed together. But seperated out, I can start to see trends and then compare between the classes. Also, you get a lot of overlap in plot 6, which makes it hard to read what is really going on.