Assignment 5: Data Visualization

Joshua Frear

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
#getwd()
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)

Lakedata_tidy <- read.csv("./Data/Processed/NTL-LTER_Lake_Chemistry_Nutrients_PeterPaul_Processed.csv",
Lakedata_gathered <- read.csv("./Data/Processed/NTL-LTER_Lake_Nutrients_PeterPaulGathered_Processed.csv
litter <- read.csv("./Data/Processed/NEON_NIWO_Litter_mass_trap_Processed.csv", stringsAsFactors = TRUE

#2
#str(litter)
```

```
#str(Lakedata_gathered)
#str(Lakedata_tidy)
#dates are characters, so using as.Date() to fix
litter$collectDate <- as.Date(litter$collectDate)
Lakedata_gathered$sampledate <- as.Date(Lakedata_gathered$sampledate)
Lakedata_tidy$sampledate <- as.Date(Lakedata_tidy$sampledate)
```

Define your theme

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

```
mytheme <- theme_light(base_size = 12) +
  theme(legend.position = "top")

theme_set(mytheme)
```

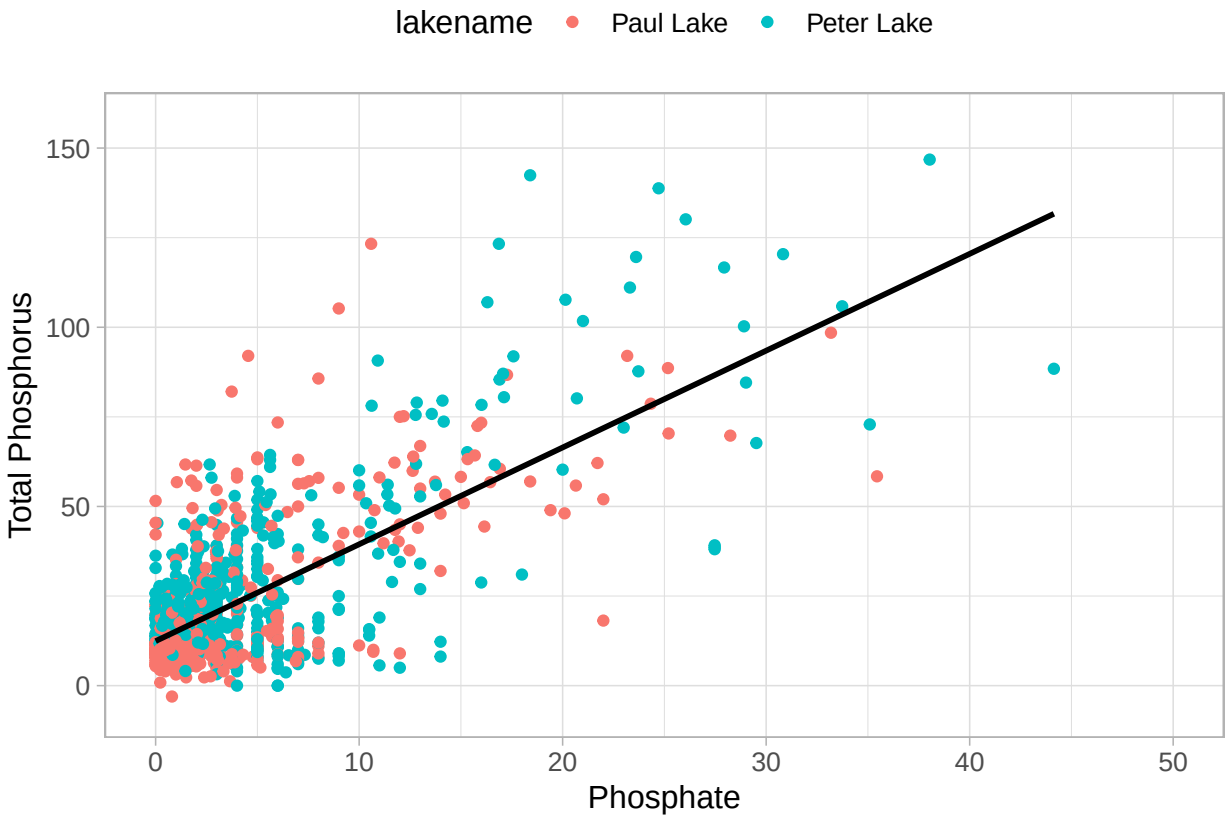
Create graphs

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.

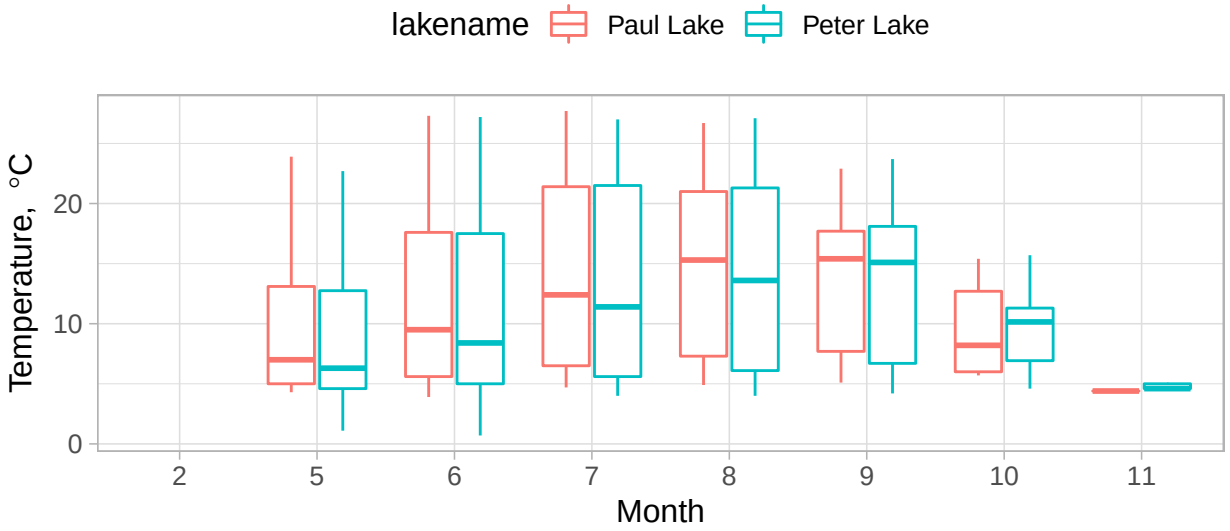
```
#4
plot1 <- ggplot(Lakedata_tidy, aes(x = po4, y = tp_ug, color = lakename)) + geom_point() + xlim(0,50) +
print(plot1)
```

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

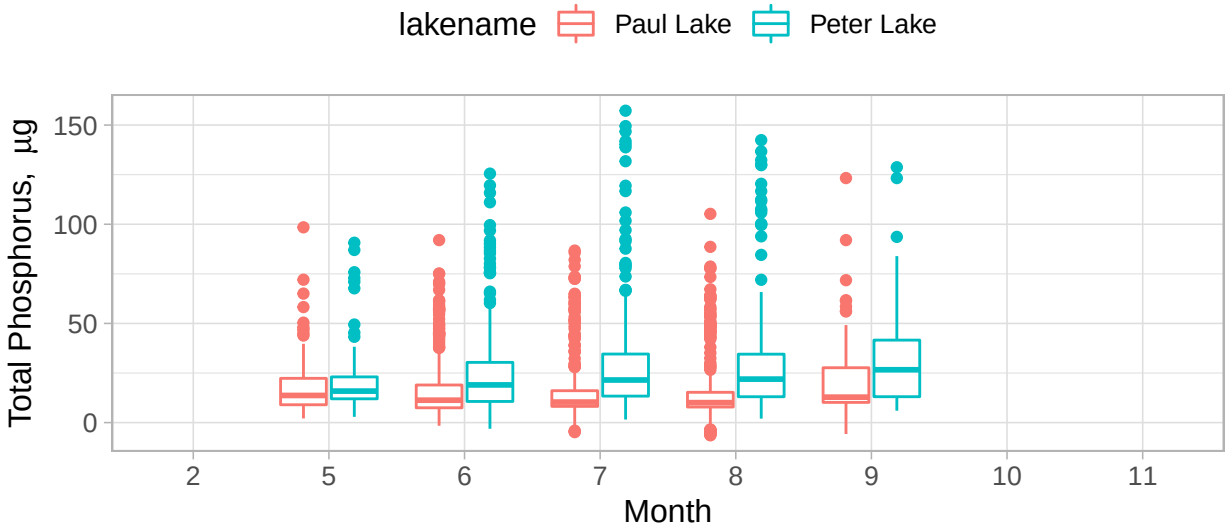


5. [NTL-LTER] Make three separate boxplots of (a) temperature, (b) TP, and (c) TN, with month as the x axis and lake as a color aesthetic. Then, create a cowplot that combines the three graphs. Make sure that only one legend is present and that graph axes are aligned.

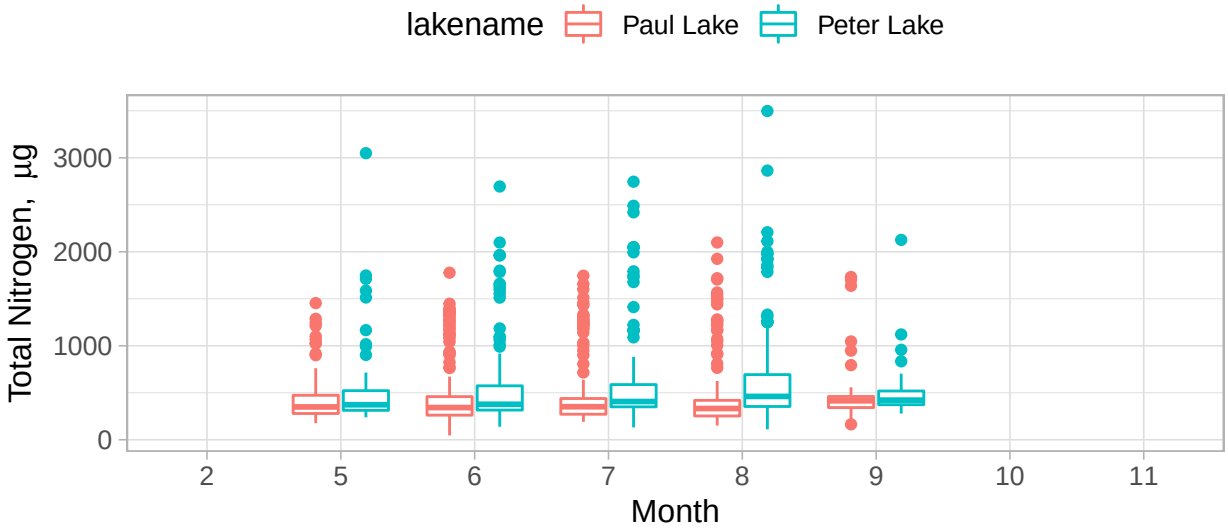
```
#5
library(cowplot)
library(ggplot2)

boxplot1 <- ggplot(Lakedata_tidy, aes(x = as.factor(month), y = temperature_C, color = lakename)) + geo
print(boxplot1)
```



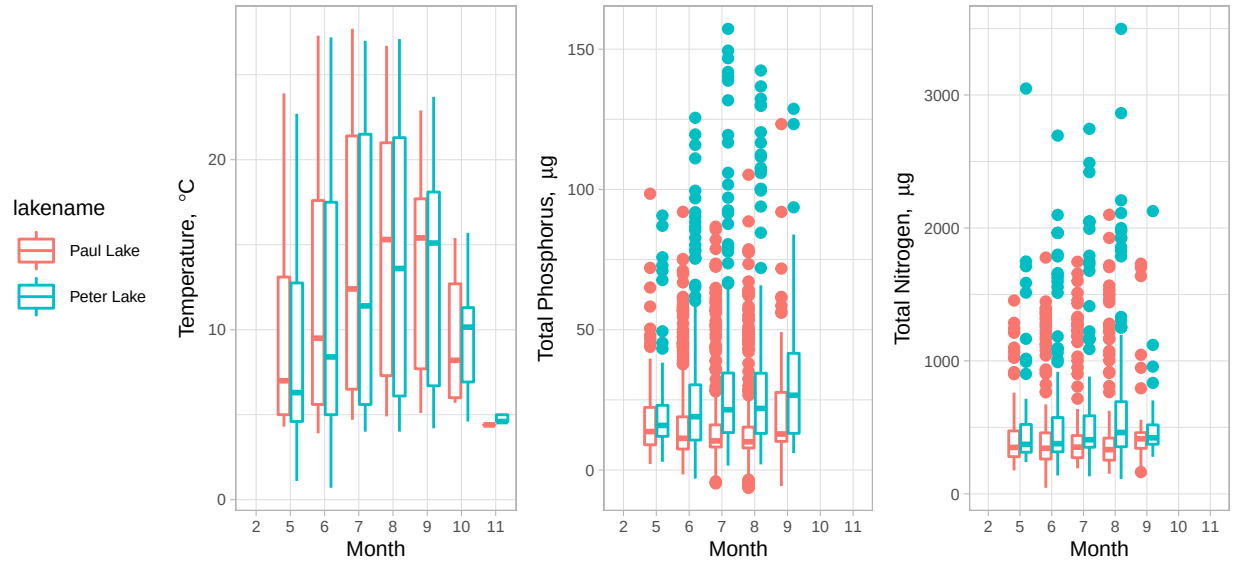

```
boxplot2 <- ggplot(Lakedata_tidy, aes(x = as.factor(month), y = tp_ug, color = lakename)) + geom_boxplo
print(boxplot2)
```




```
boxplot3 <- ggplot(Lakedata_tidy, aes(x = as.factor(month), y = tn_ug, color = lakename)) + geom_boxplo
print(boxplot3)
```




```
boxplot1a <- ggplot(Lakedata_tidy, aes(x = as.factor(month), y = temperature_C, color = lakename)) + geo

boxplot2a <- ggplot(Lakedata_tidy, aes(x = as.factor(month), y = tp_ug, color = lakename)) + geom_boxplo

boxplot3a <- ggplot(Lakedata_tidy, aes(x = as.factor(month), y = tn_ug, color = lakename)) + geom_boxplo

themesmall <- theme_light(base_size = 8) +
  theme(legend.position = "top")

theme_set(themesmall)

cowplot1 <- plot_grid(boxplot1a, boxplot2a, boxplot3a, nrow = 1, ncol = 3, align = 'h', rel_widths = c(1

print(cowplot1)
```

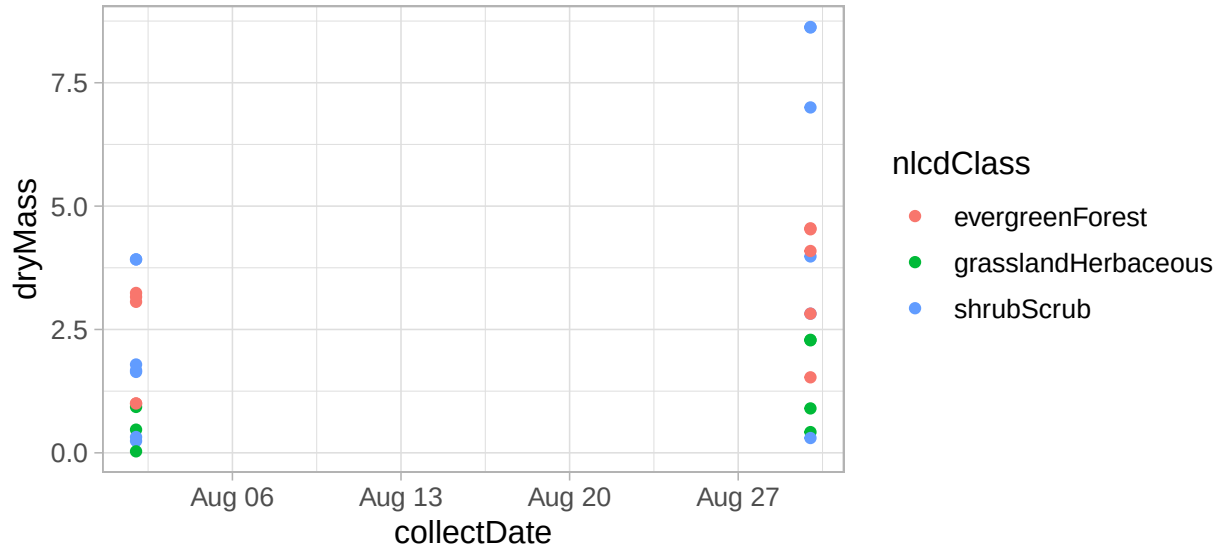


Question: What do you observe about the variables of interest over seasons and between lakes?

Answer: The variables increase during the year until July or August, at which point they decrease. They are generally higher in Peter Lake than Paul Lake for a given month

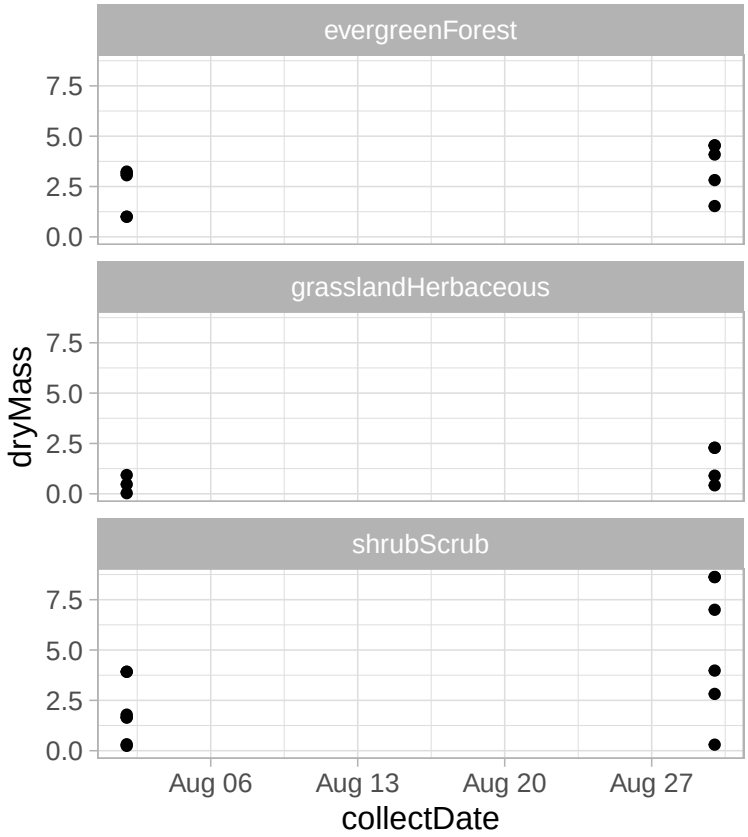
6. [Niwot Ridge] Plot a subset of the litter dataset by displaying only the "Needles" functional group. Plot the dry mass of needle litter by date and separate by NLCD class with a color aesthetic. (no need to adjust the name of each land use)

```
#6
theme_set(mytheme)

ggplot(litter[litter$functionalGroup == "Needles",],
       aes(y = dryMass, x = collectDate, color = nlcdClass)) + geom_point() + theme(legend.position = "
```



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

```
#7
ggplot(litter[litter$functionalGroup == "Needles",],
       aes(y = dryMass, x = collectDate)) + geom_point() + facet_wrap(vars(nlcdClass), nrow = 3)
```



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

Answer: I like 6 better - you get more information in a smaller space, and the general trend is easily visible. There are not enough data points to justify the arrangement in 7.