# Assignment 5: Data Visualization

Rebecca Murphy

## 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 Set up session

getwd() # get working directory
```

```
## [1] "Z:/ENV872/Environmental_Data_Analytics_2022/Assignments"
```

```
library(tidyverse) #load packages
```

```
## -- 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()
```

```
library(cowplot)

#upload data
Lake.chemistry.processed <-
```

```r
read.csv("../Data/Processed/NTL-LTER_Lake_Chemistry_Nutrients_PeterPaul_Processed.csv",
                              stringsAsFactors = TRUE)


Litter.processed <-
read.csv("../Data/Processed/NEON_NIWO_Litter_mass_trap_Processed.csv",
                              stringsAsFactors = TRUE)

#2 Check that dates are in date format

class(Lake.chemistry.processed$sampledate)
```

```
## [1] "factor"
```

```r
class(Litter.processed$collectDate)
```

```
## [1] "factor"
```

```r
#Both are listed as factors

Lake.chemistry.processed$sampledate <- as.Date(Lake.chemistry.processed$sampledate,
                                                format = "%Y-%m-%d")

Litter.processed$collectDate <- as.Date(Litter.processed$collectDate,
                                         format = "%Y-%m-%d")

#After checking again, both are now dates
```

## Define your theme

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

```r
#3 Defining my theme

#Building a theme
new.theme <- theme_gray(base_size = 14) +
  theme(axis.text = element_text(color = "black"),
        legend.position = "bottom")

#Making it my default theme
theme_set(new.theme)
```

## Create graphs

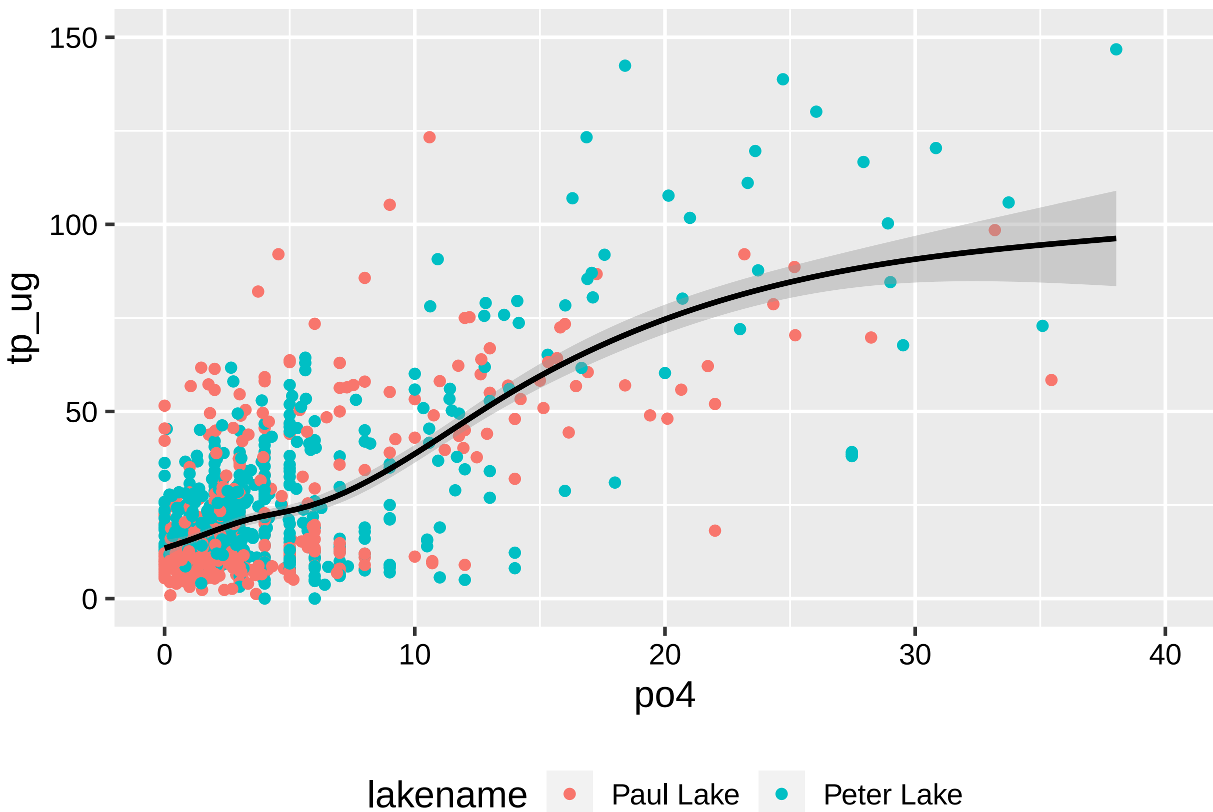
For numbers 4-7, create ggplot graphs and adjust aesthetics to follow best practices for data visualization. Ensure your theme, color palettes, axes, and additional aesthetics are edited accordingly.

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()`).

```r
#4 Plotting total tp_ug x po4

Plot1 <- ggplot(Lake.chemistry.processed, aes(x = po4, y = tp_ug, color = lakename)) +
  geom_point() +
  geom_smooth(color = "black") +
```

```
  xlim(0, 40) +
  ylim(0, 150)
print(Plot1)
```

```
## `geom_smooth()` using method = 'gam' and formula 'y ~ s(x, bs = "cs")'

## Warning: Removed 21949 rows containing non-finite values (stat_smooth).

## Warning: Removed 21949 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 Boxplots for temp, TP, and TN

#changing months to categories first...

lake.month <- as.factor(Lake.chemistry.processed$month)


#Create individual boxplots
Plot2 <- ggplot(Lake.chemistry.processed, aes(x = lake.month,
                                              y = tp_ug, color = lakename)) +
geom_boxplot() +
theme(legend.position = "none")

Plot3 <- ggplot(Lake.chemistry.processed, aes(x = lake.month,
```

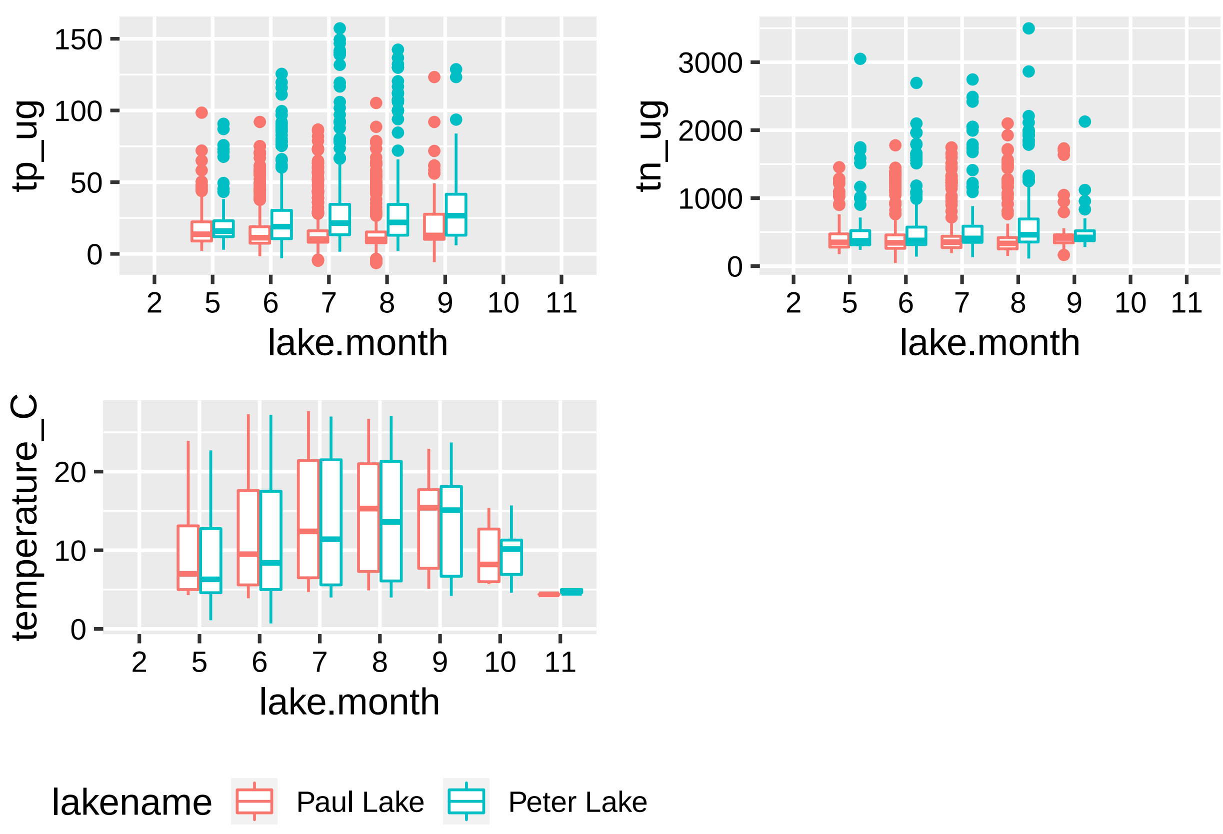
```
                                                    y = tn_ug, color = lakename)) +
geom_boxplot() +
theme(legend.position = "none")

Plot4 <-
  ggplot(Lake.chemistry.processed,
                aes(x = lake.month, y = temperature_C, color = lakename)) +
  geom_boxplot()

 #have to make separate legend as template
legend <- get_legend(Plot2)
```

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

```
#Create cowplot to combine all three graphs
plot.all <- plot_grid(Plot2, Plot3, Plot4,
                      nrow = 2,
                      legend,
                      rel_heights = c(1,1.25))
```

```
## Warning: Removed 20729 rows containing non-finite values (stat_boxplot).
## Warning: Removed 21583 rows containing non-finite values (stat_boxplot).
## Warning: Removed 3566 rows containing non-finite values (stat_boxplot).
```

```
print(plot.all)
```

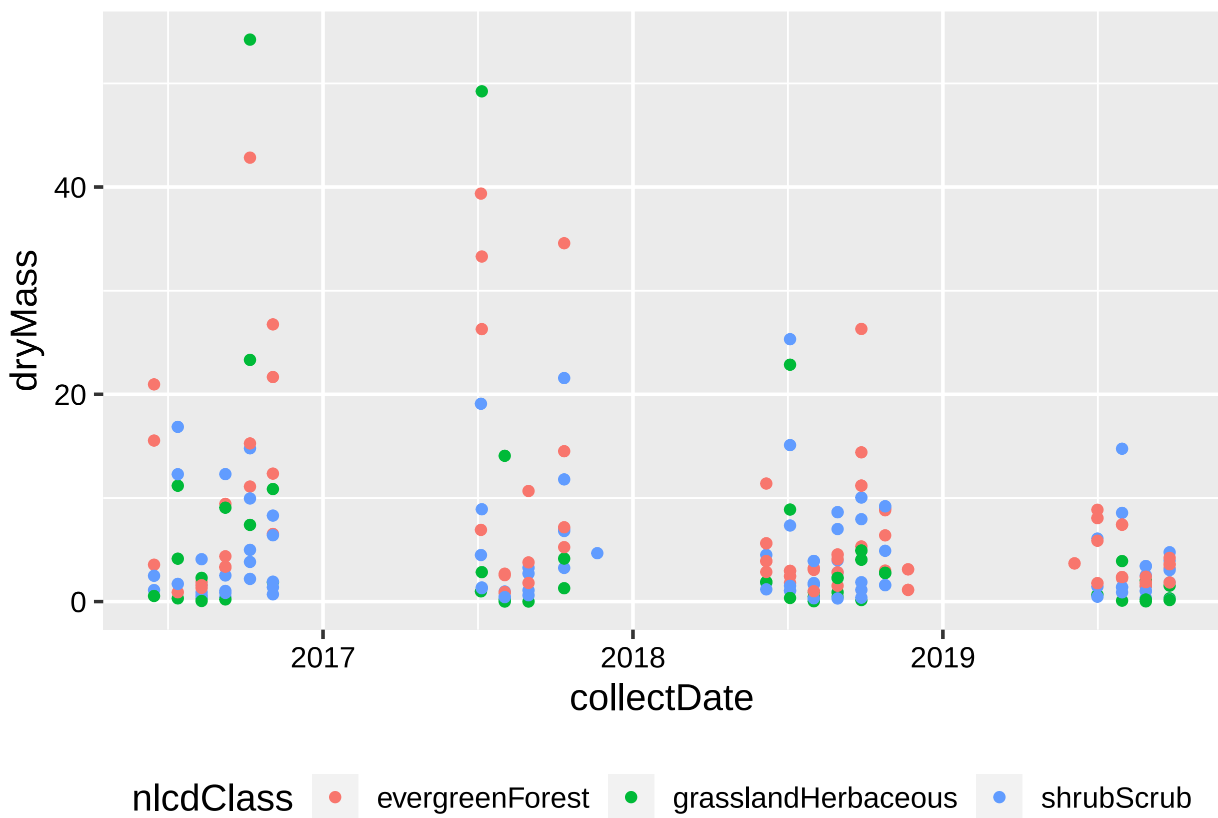


Question: What do you observe about the variables of interest over seasons and between lakes?

> Answer: On average, it appears that Peter Lake has higher nutrient values (phosphorus and nitrogen), particularly in late summer/early fall months. Both lakes have similar temperature values across months (Paul Lake is slightly higher on average), with peak average temperatures occuring in August and September.

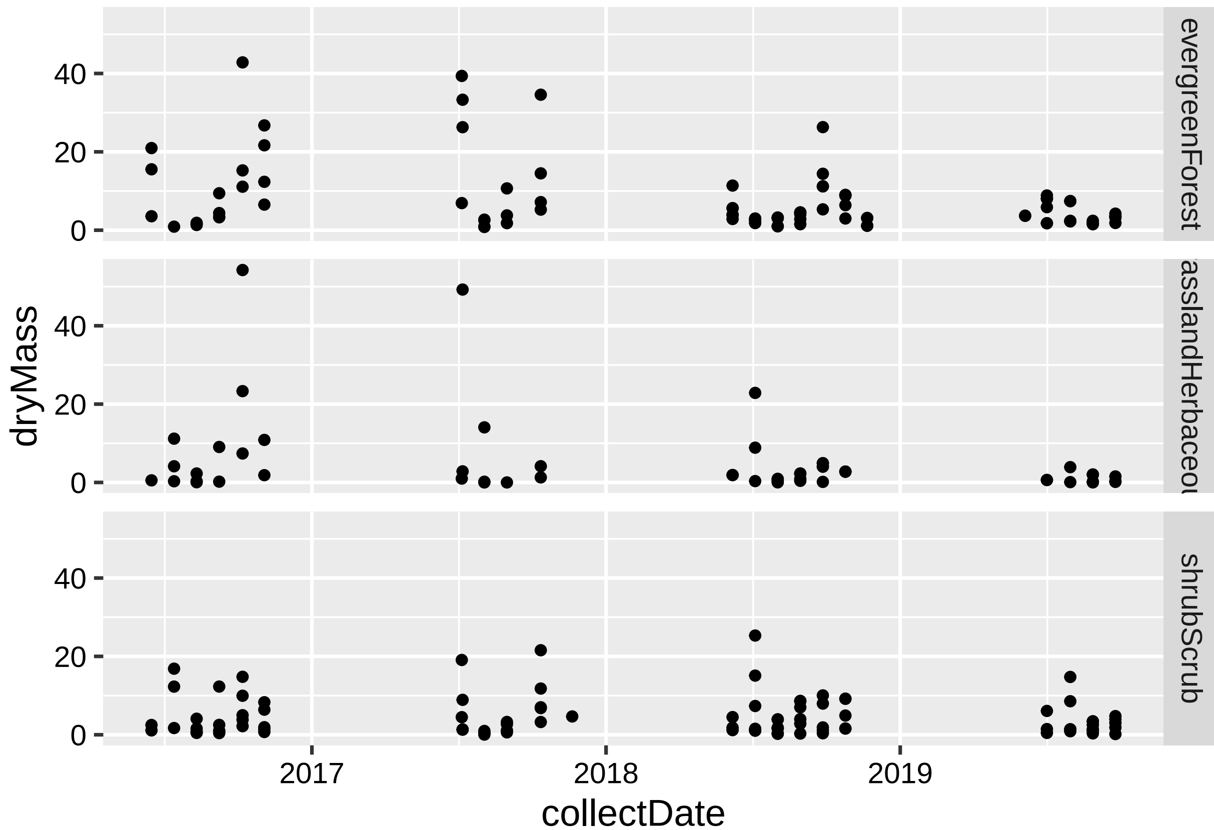
6. [Niwot Ridge] Plot a subset of the litter dataset by displaying only the "Needles" functional group. Plot the dry mass of needle litter by date and separate by NLCD class with a color aesthetic. (no need to adjust the name of each land use)
7. [Niwot Ridge] Now, plot the same plot but with NLCD classes separated into three facets rather than separated by color.

```
#6 Plotting dry mass of needle litter by date

Plot5 <- ggplot(subset(Litter.processed, functionalGroup == "Needles"),
                aes(x = collectDate, y = dryMass, color = nlcdClass)) +
  geom_point()

print(Plot5)
```




```
#7

Plot6 <- ggplot(subset(Litter.processed, functionalGroup == "Needles"),
                aes(x = collectDate, y = dryMass)) +
  geom_point() +
  facet_grid(vars(nlcdClass))
```

```
print(Plot6)
```




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

> Answer: I found Plot6 (separated into 3 facets) easier to digest and discern than coloration of the classes. Specifically, when looking at both counts and spread of needle dry mass over time, it is easier to visualize changes per nlcd class when separated.