Assignment 5: Data Visualization

Yosia Theo Napitupulu

OVERVIEW

This exercise accompanies the lessons in Environmental Data Analytics on Data Visualization

Directions

1. Rename this file `<FirstLast>_A02_CodingBasics.Rmd` (replacing `<FirstLast>` with your first and last name).
2. Change "Student Name" on line 3 (above) with your name.
3. Work through the steps, **creating code and output** that fulfill each instruction.
4. Be sure to **answer the questions** in this assignment document.
5. When you have completed the assignment, **Knit** the text and code into a single PDF file.

The completed exercise is due on Friday, Oct 14th @ 5:00pm.

Set up your session

1. Set up your session. Verify your working directory and load the tidyverse, lubridate, & 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_PeterP`
version) and the processed data file for the Niwot Ridge litter dataset (use the [`NEON_NIWO_Litter_mass_trap_Processe`
version).

2. Make sure R is reading dates as date format; if not change the format to date.

```
# 1 Set up, locating working directory
getwd()
```

```
## [1] "E:/ENV872/EDA-Fall2022"
```

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

NTL.chem.nutrient.data.PeterPaul <- read.csv("./Data/Processed/NTL-LTER_Lake_Chemistry_Nutrients_PeterP
    stringsAsFactors = TRUE)

NIWO.litter.data <- read.csv("./Data/Processed/NEON_NIWO_Litter_mass_trap_Processed.csv",
    stringsAsFactors = TRUE)
```

```
# 2 Change format date
NTL.chem.nutrient.data.PeterPaul$sampledate <- as.Date(NTL.chem.nutrient.data.PeterPaul$sampledate,
    format = "%Y-%m-%d")

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

class(NTL.chem.nutrient.data.PeterPaul$sampledate)
```

```
## [1] "Date"
```

```
class(NIWO.litter.data$collectDate)
```

```
## [1] "Date"
```

Define your theme

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

```
# 3 Set theme
mytheme <- theme_classic(base_size = 14) + theme(axis.text = element_text(color = "black"),
    legend.position = "top")
theme_set(mytheme)
```

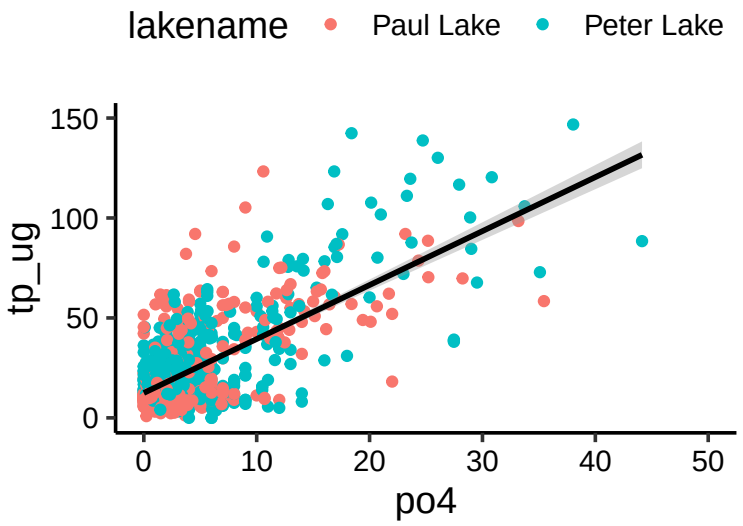
Create graphs

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 (hint: change the limits using `xlim()` and/or `ylim()`).

```
# 4 total phosphorus (`tp_ug`) by phosphate (`po4)
nutrient.plot <- ggplot(NTL.chem.nutrient.data.PeterPaul, aes(x = po4, y = tp_ug,
    color = lakename)) + geom_point() + geom_smooth(method = lm, color = "black") +
    ylim(0, 150) + xlim(0, 50)
print(nutrient.plot)
```

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



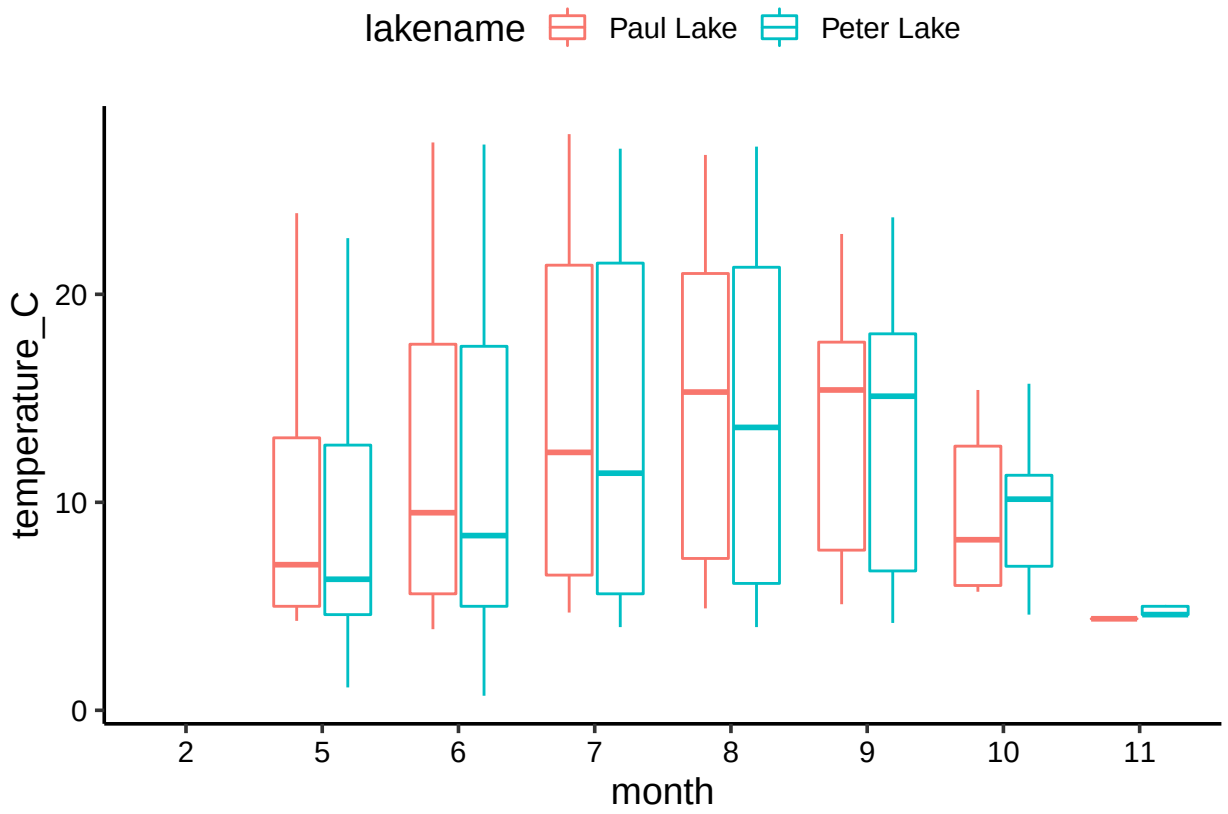
5. [NTL-LTER] Make three separate boxplots of (a) temperature, (b) TP, and

(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.

Tip: R has a build in variable called `month.abb` that returns a list of months; see https://r-lang.com/month-abb-in-r-with-example

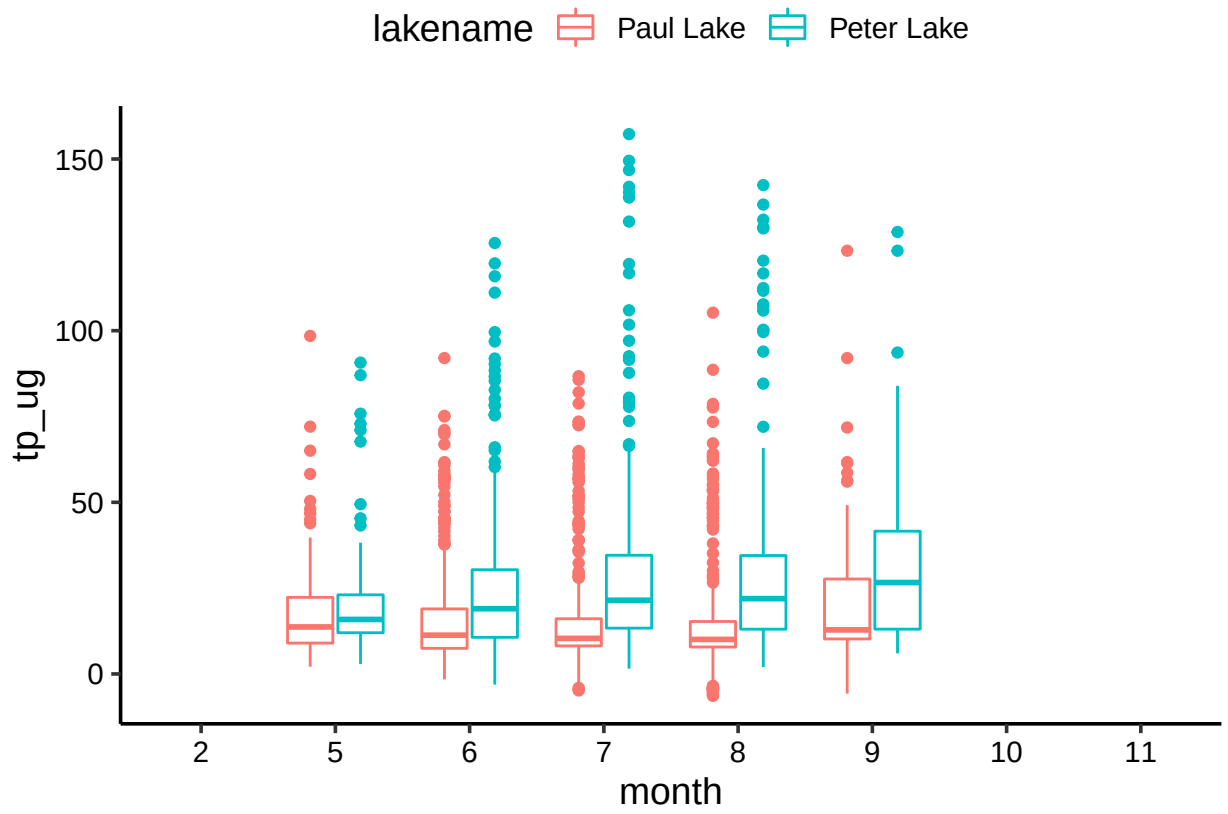
```
# 5 Plots for temperature, TP and TN change month from int to factor
NTL.chem.nutrient.data.PeterPaul$month <- as.factor(NTL.chem.nutrient.data.PeterPaul$month)


# boxplot for temperature
temp.plot <- ggplot(NTL.chem.nutrient.data.PeterPaul, aes(x = month, y = temperature_C)) +
    geom_boxplot(aes(color = lakename))
print(temp.plot)
```

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



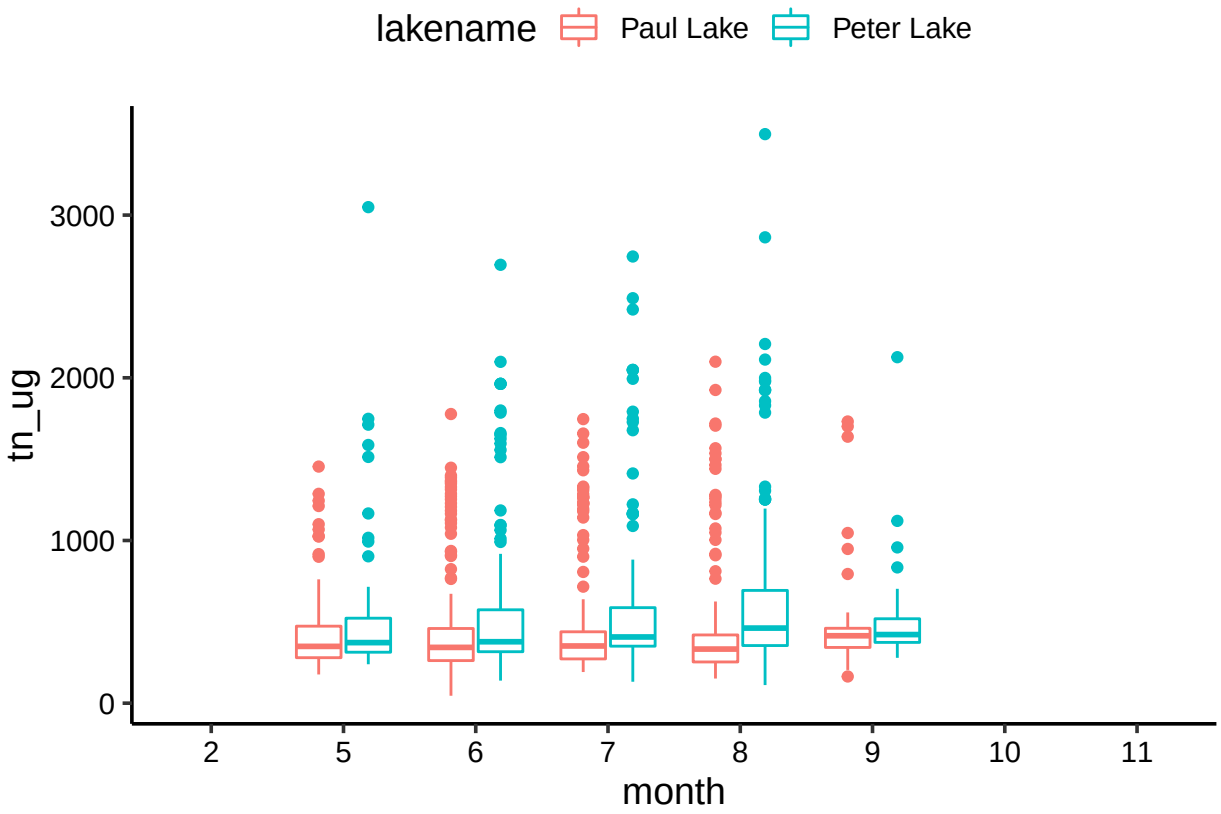

```
# boxplot for TP
TP.plot <- ggplot(NTL.chem.nutrient.data.PeterPaul, aes(x = month, y = tp_ug)) +
    geom_boxplot(aes(color = lakename))
print(TP.plot)
```

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




```
# boxplot for TN
TN.plot <- ggplot(NTL.chem.nutrient.data.PeterPaul, aes(x = month, y = tn_ug)) +
    geom_boxplot(aes(color = lakename))
print(TN.plot)
```

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



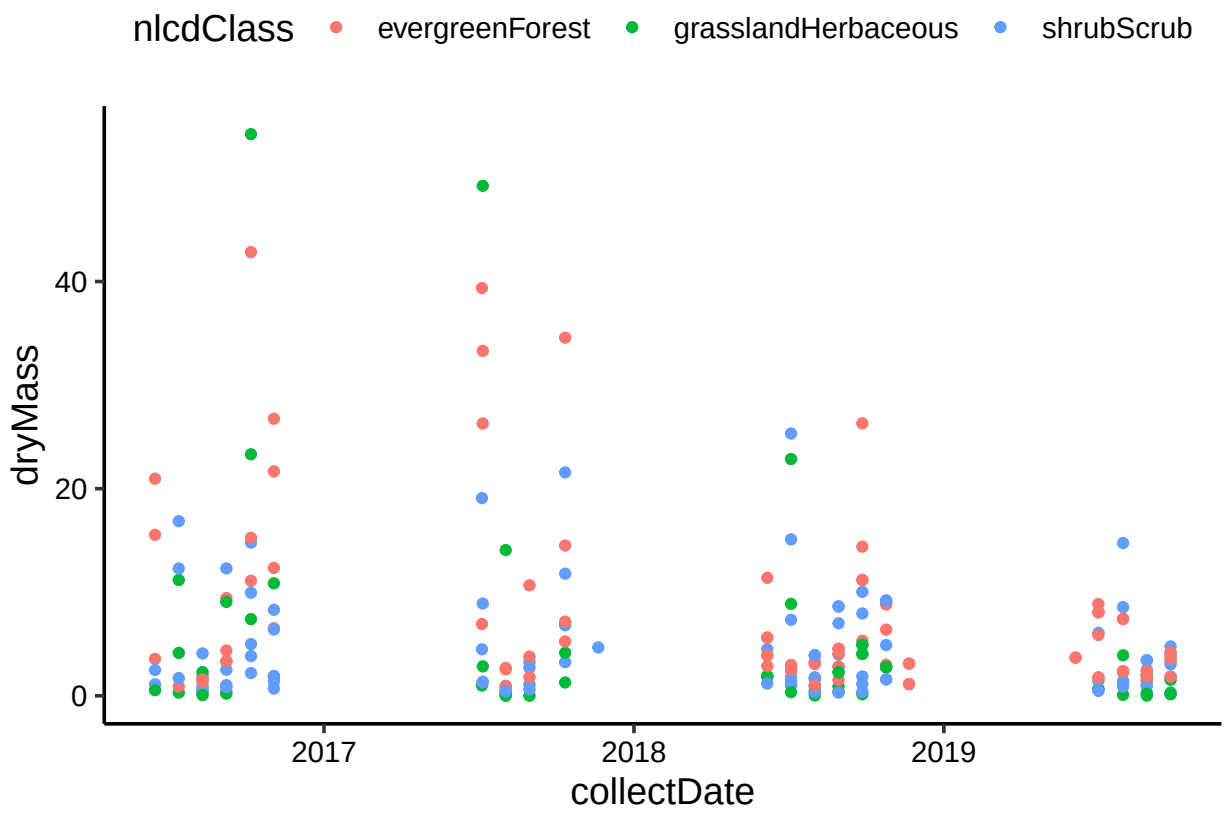
Question: What do you observe about the variables of interest over seasons and between lakes?

> Answer: The both lakes only have a contrast difference of median temperatures value on August and october, while the other months seems quite similar with Peter Lake has a higher median value only on October. while regarding the tp_ug and tn_ug, Peter Lake has a higher median value of tp_ug and tn_ug through the year of observation. Another interesting information is about the number of outlier which both lakes has pretty high data outlier in tp_ug and tn_ug. However, Peter Lake still becoming the lake who has more outlier in both tp_ug and tn_ug.

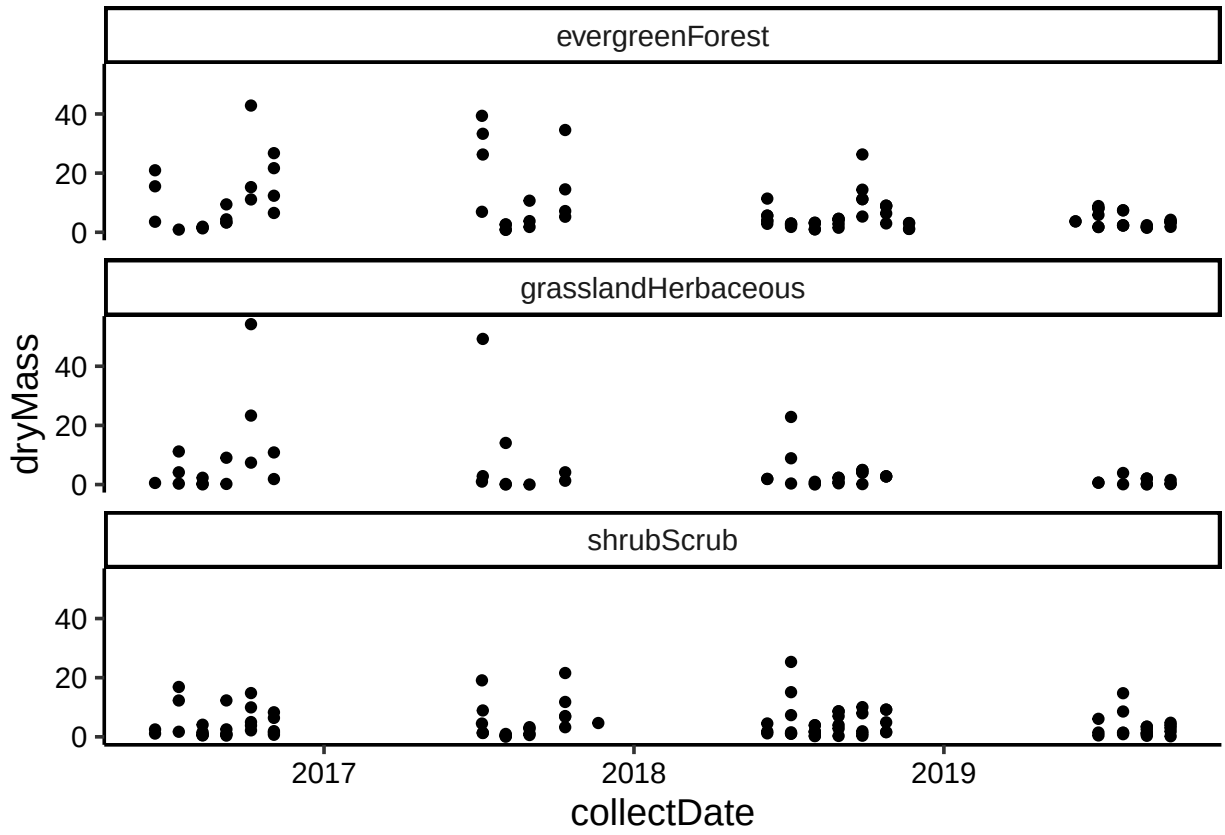
6. [Niwot Ridge] Plot a subset of the litter dataset by displaying only the "Needles" functional group. Plot the dry mass of needle litter by date and separate by NLCD class with a color aesthetic. (no need to adjust the name of each land use)

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

```
# 6 Plotting the dry mass of needle litter
ggplot(subset(NIWO.litter.data, functionalGroup == "Needles"), aes(x = collectDate,
    y = dryMass)) + geom_point(aes(color = nlcdClass))
```




```
# 7 Plotting the dry mass of needle litter with 3 facets of nlcd class
ggplot(subset(NIWO.litter.data, functionalGroup == "Needles"), aes(x = collectDate,
    y = dryMass)) + 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 think the plot 7 would be more effective because it has a more obvious picture and visualization style. The plot 7 has segmenting the variables in each category, then it becomes more easy to measure and compare between three of each class in each year, rather than the coloured plot 6.