Assignment 5: Data Visualization

Austin Guimond

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_PeterPa` 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
getwd()
```

```
## [1] "/home/guest/R/EDA-Fall2022"
```

```
library(tidyverse)
```

```
## -- Attaching packages --------------------------------------- tidyverse 1.3.2 --
## v ggplot2 3.3.6      v purrr   0.3.4
## v tibble  3.1.8      v dplyr   1.0.10
## v tidyr   1.2.0      v stringr 1.4.1
## v readr   2.1.2      v forcats 0.5.2
## -- Conflicts ------------------------------------------ tidyverse_conflicts() --
## x dplyr::filter() masks stats::filter()
## x dplyr::lag()    masks stats::lag()
```

```r
library(lubridate)
```

```
##
## Attaching package: 'lubridate'
##
## The following objects are masked from 'package:base':
##
##     date, intersect, setdiff, union
```

```r
library(cowplot)
```

```
##
## Attaching package: 'cowplot'
##
## The following object is masked from 'package:lubridate':
##
##     stamp
```

```r
# Read lake chemistry csv
Lake.Chemistry <- read.csv("./Data/Processed/NTL-LTER_Lake_Chemistry_Nutrients_PeterPaul_Processed.csv")
# Read litter table
Litter.Mass <- read.csv("./Data/Processed/NEON_NIWO_Litter_mass_trap_Processed.csv")
# 2 change format for all dates
class(Lake.Chemistry$sampledate)
```

```
## [1] "character"
```

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

```
## [1] "character"
```

```r
Lake.Chemistry$sampledate <- as.Date(Lake.Chemistry$sampledate, format = "%Y-%m-%d")
Litter.Mass$collectDate <- as.Date(Litter.Mass$collectDate, format = "%Y-%m-%d")
class((Lake.Chemistry$sampledate))
```

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

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

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

Define your theme

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

```r
# 3 change and set plot theme
mytheme <- theme_grey(base_size = 8) + theme(axis.text = element_text(color = "black"),
    legend.position = "bottom")
theme_set(mytheme)
```

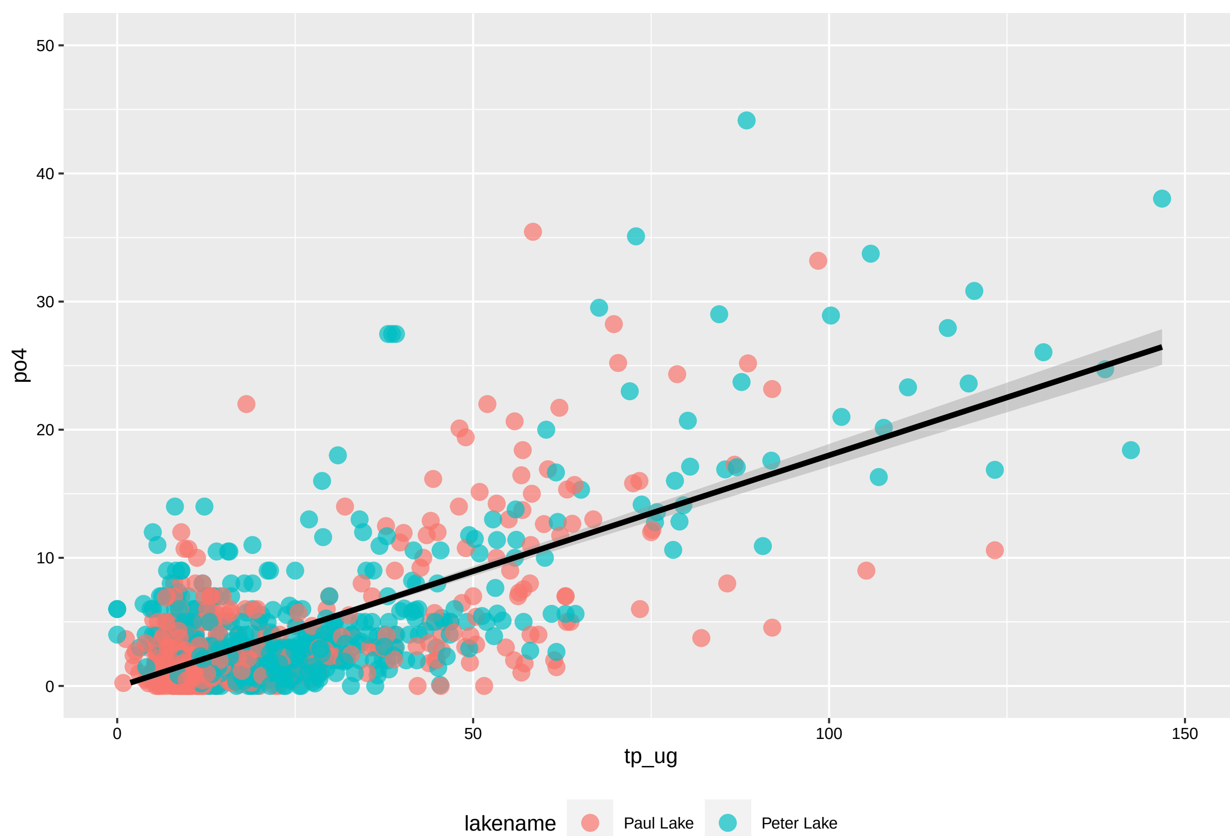
Create graphs

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
Phosphorous.Plot <- ggplot(Lake.Chemistry, aes(x = tp_ug, y = po4, color = lakename)) +
    geom_point(alpha = 0.7, size = 2.5) + xlim(0, 150) + ylim(0, 50) + geom_smooth(method = lm,
    col = "black")
print(Phosphorous.Plot)
```

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

## Warning: Removed 21948 rows containing non-finite values (stat_smooth).

## Warning: Removed 21948 rows containing missing values (geom_point).

## Warning: Removed 1 rows containing missing values (geom_smooth).
```



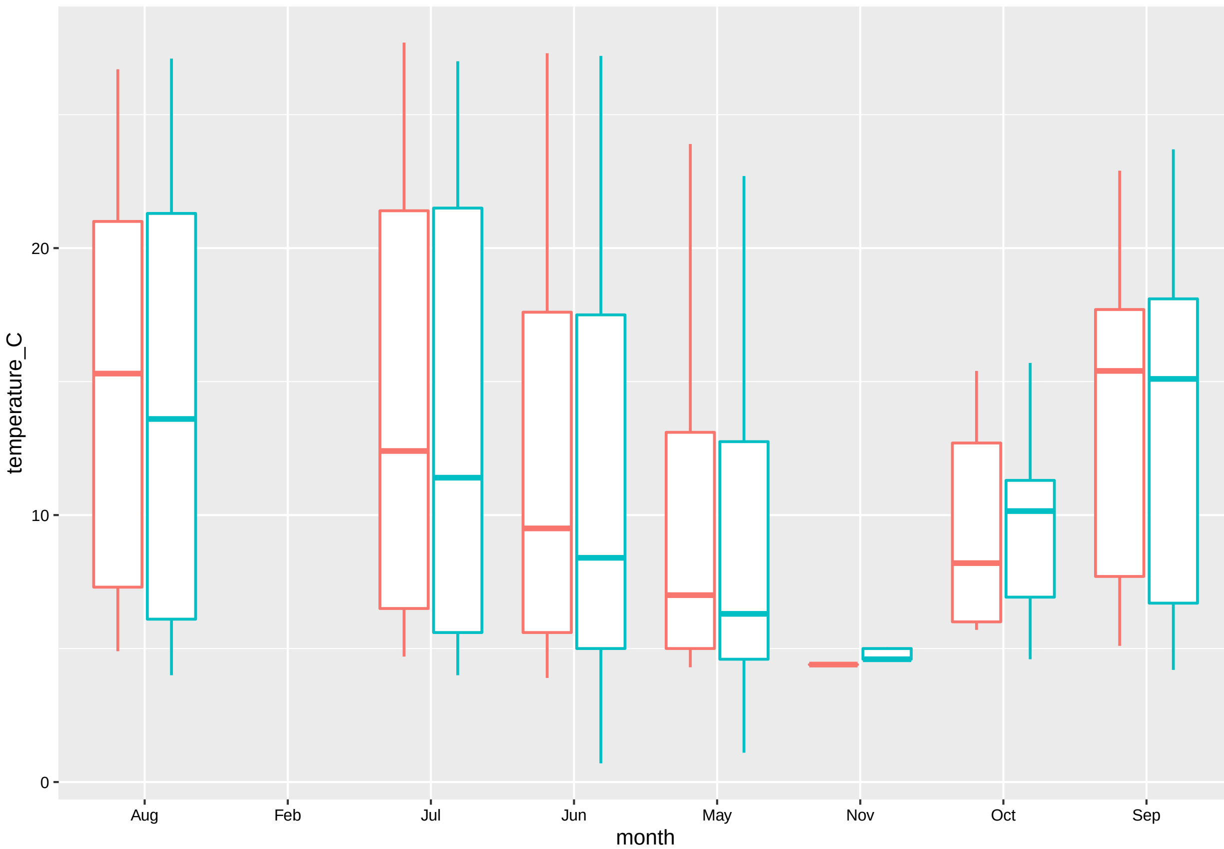
5. [NTL-LTER] Make three separate boxplots of (a) temperature, (b) TP, and

(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.

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

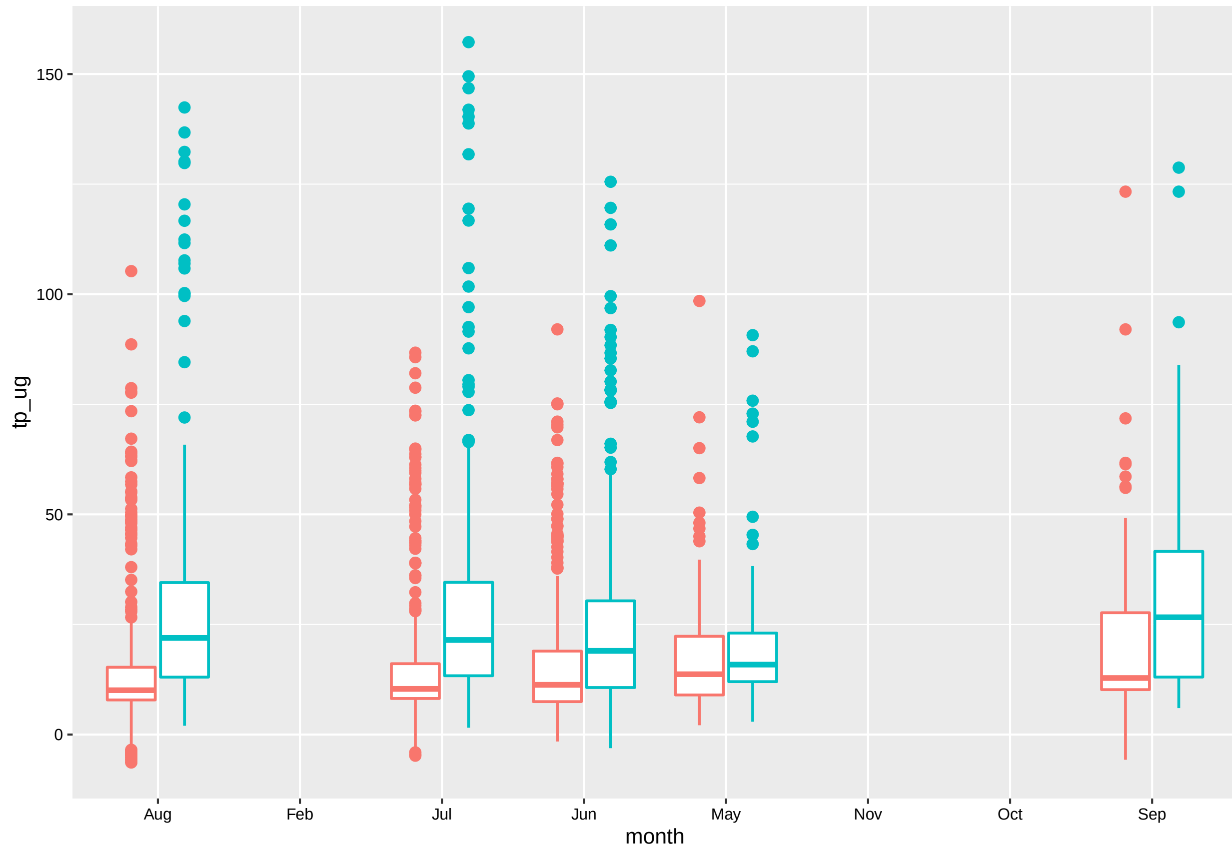
```
# 5 Change months from numbers to abbreviations
Lake.Chemistry.Month <- mutate(Lake.Chemistry, month = month.abb[month])

# plot temperature
Temperature.Box <- ggplot(Lake.Chemistry.Month, aes(x = month, y = temperature_C)) +
    geom_boxplot(aes(color = lakename)) + theme(legend.position = "none")
print(Temperature.Box)
```

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



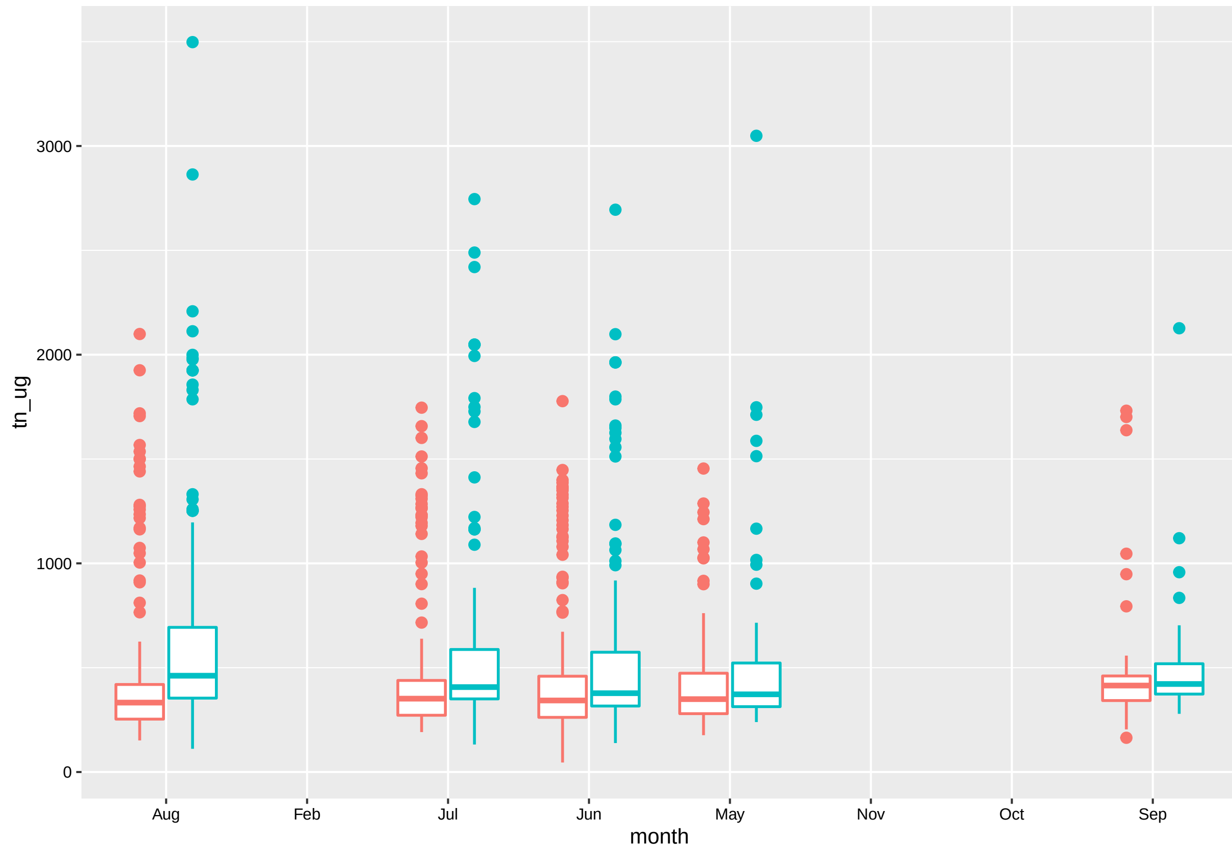

```
# plot TP
TP.Box <- ggplot(Lake.Chemistry.Month, aes(x = month, y = tp_ug)) + geom_boxplot(aes(color = lakename))
    theme(legend.position = "none")
print(TP.Box)
```

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




```r
# plot TN
TN.Box <- ggplot(Lake.Chemistry.Month, aes(x = month, y = tn_ug)) + geom_boxplot(aes(color = lakename))
    theme(legend.position = "none")
print(TN.Box)
```

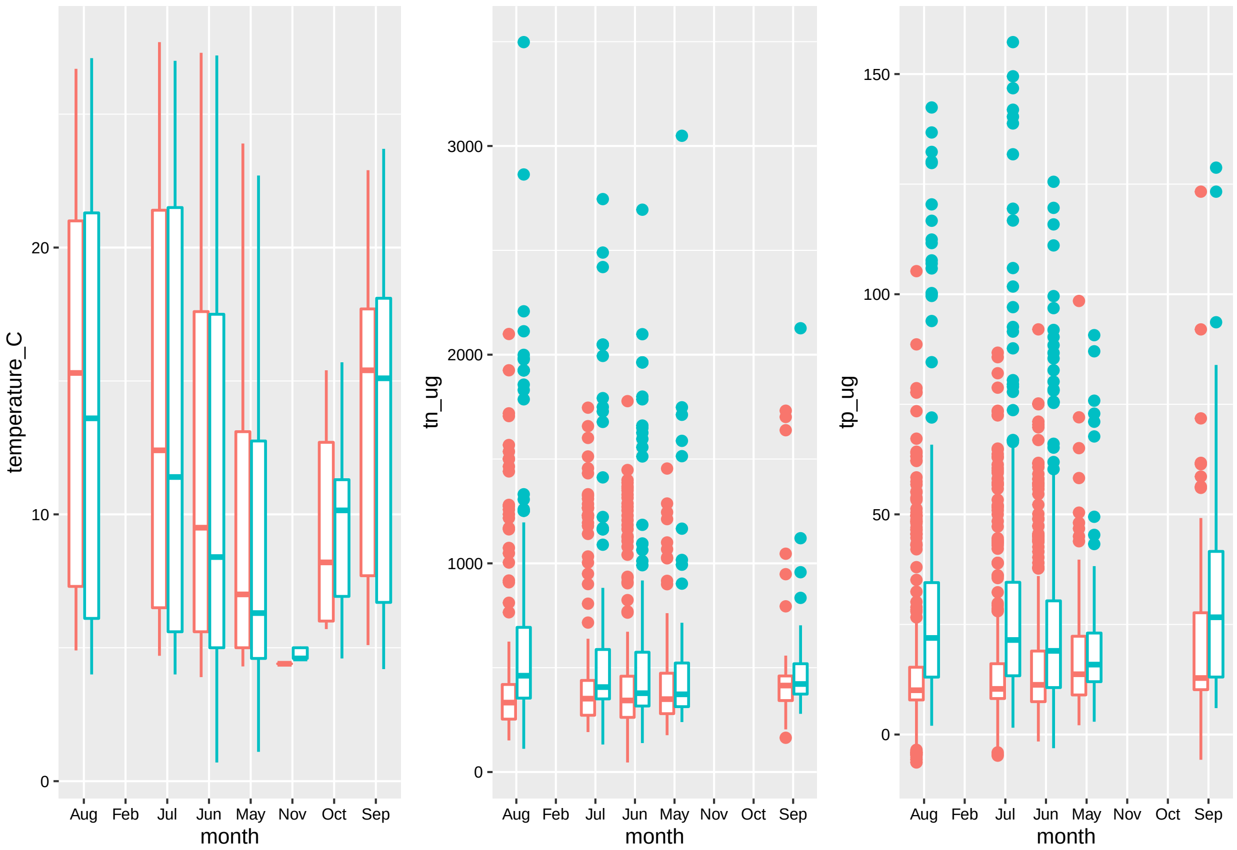
```
## Warning: Removed 21583 rows containing non-finite values (stat_boxplot).
```




```r
# plot all
Combined.Box <- plot_grid(Temperature.Box, TN.Box, TP.Box, ncol = 3, nrow = 1)
```

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

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

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

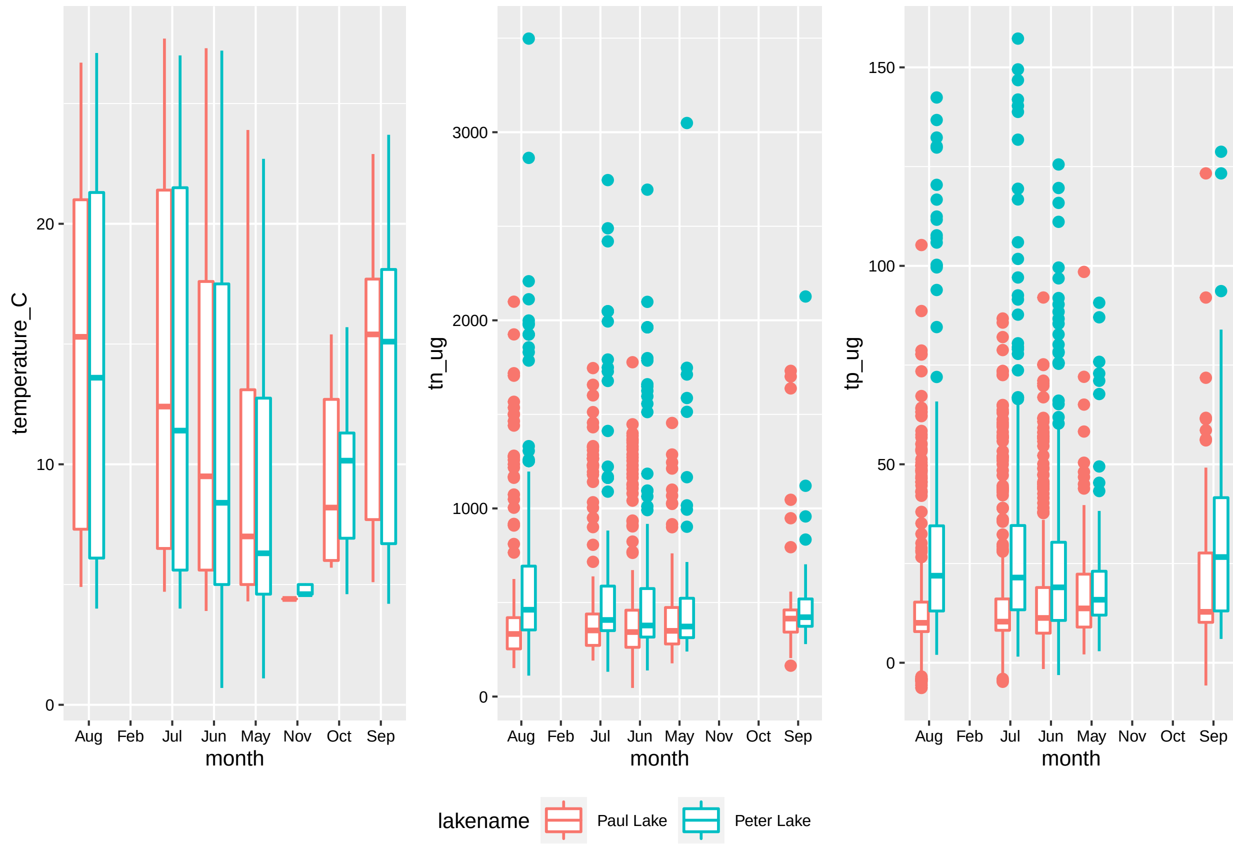
```r
print(Combined.Box)
```




```
# generate legend
Box.Legend <- ggplot(Lake.Chemistry, aes(x = month, y = temperature_C)) + geom_boxplot(aes(color = laken

# plot all with single legend
Combined.Legend <- get_legend(Box.Legend)
```

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

```
One.Legend.Box <- plot_grid(Combined.Box, Combined.Legend, ncol = 1, rel_heights = c(1,
    0.1))
print(One.Legend.Box)
```

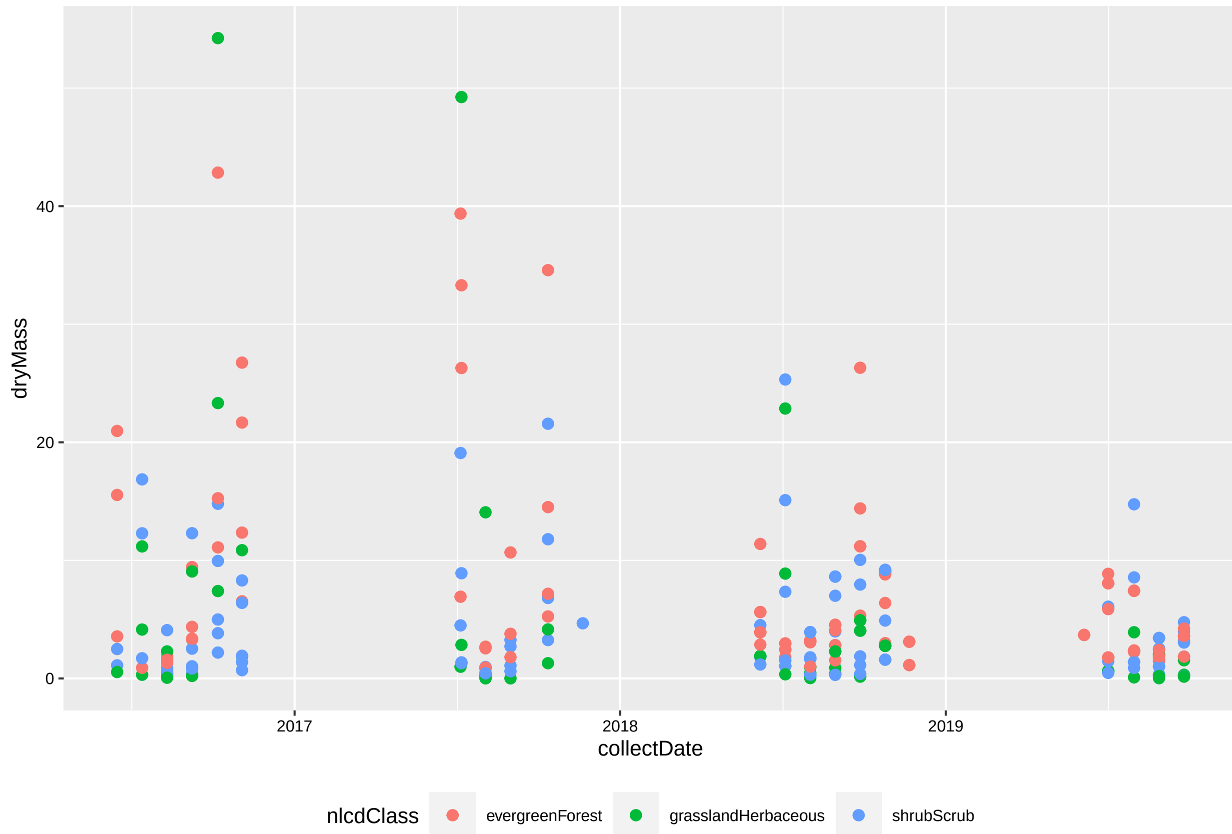


Question: What do you observe about the variables of interest over seasons and between lakes?

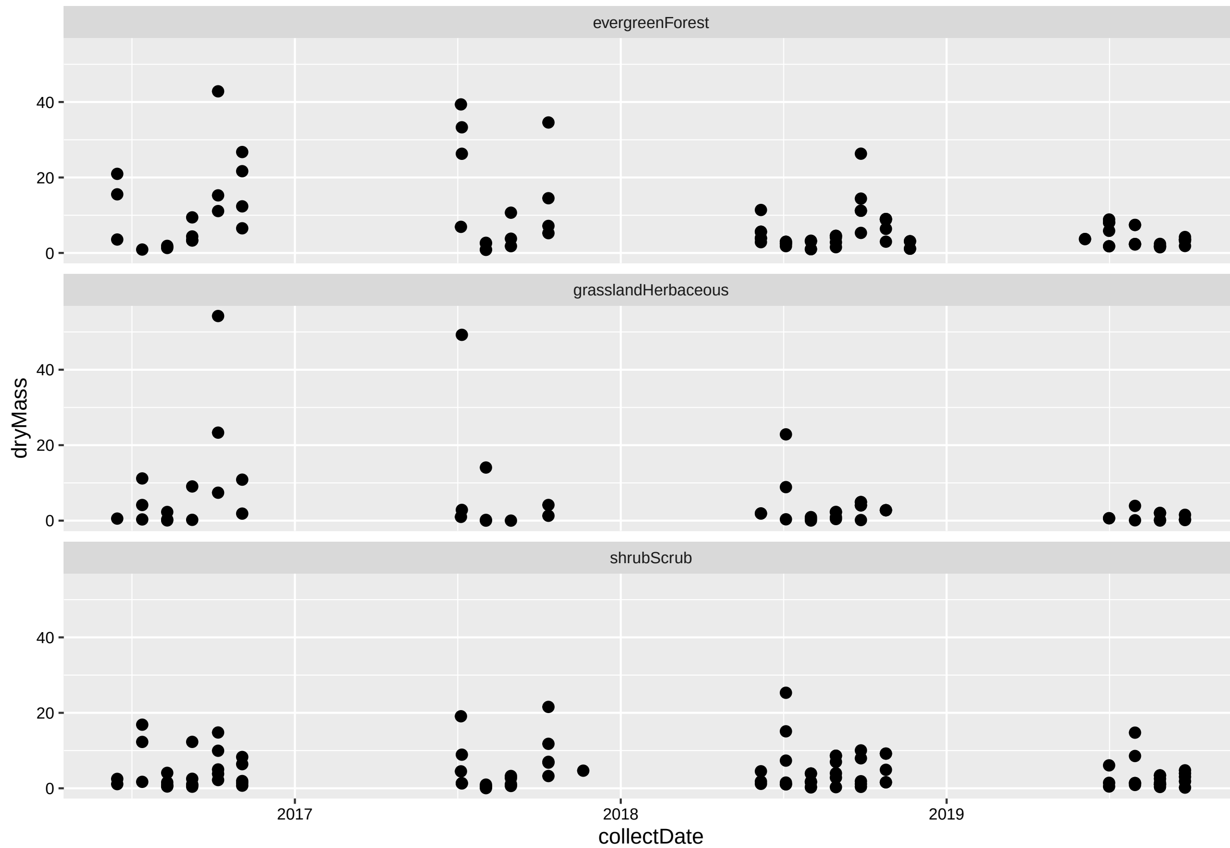
> Answer: One of the trends that can be viewed between the lakes is that on average, Paul Lake was warmer than Peter Lake. Peter lake consistently had higher TN and TP as well as more high outlier values. Seasonally, both lakes had low temperatures in May and June that steadily increased until September. TN and TP values were conistent over most parts of the year for both lakes. However, the variablility and presence of oulier values increased later in the summer, especially in August.

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 plot needle drymass by year
Litter.Needles.Plot <- ggplot(subset(Litter.Mass, functionalGroup == "Needles"),
    aes(x = collectDate, y = dryMass, color = nlcdClass)) + geom_point()
print(Litter.Needles.Plot)
```




```
# 7 plot needle drymass sperated by class
Litter.Needles.Facet.Plot <- ggplot(subset(Litter.Mass, functionalGroup == "Needles"),
    aes(x = collectDate, y = dryMass, )) + geom_point() + facet_wrap(vars(nlcdClass),
    nrow = 3)
print(Litter.Needles.Facet.Plot)
```



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

Answer: Plot #7 is more informative becasue it seperates out the data so more conclusions can be drawn. By examining figure 7, I can see that drymass for needles was highest before 2019, while 2019 was an espcially low year. By faceting the graphs I can also confirm that the trend in declined needle mass after 2018 was the same in each of the study ecosytems. By looking at plot 6, I can make the same broad conlsuion that needlmass has declined over time, but it is harder to visually interprit that the trend was consistent across all ecosytem types.