Assignment 5: Data Visualization

Ana Bishop

Spring 2023

OVERVIEW

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

Directions

1. Rename this file `<FirstLast>_A05_DataVisualization.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 your code is tidy; use line breaks to ensure your code fits in the knitted output.
5. Be sure to **answer the questions** in this assignment document.
6. When you have completed the assignment, **Knit** the text and code into a single PDF file.

Set up your session

1. Set up your session. Load the tidyverse, lubridate, here & cowplot packages, and verify your home directory. 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 load libraries
library(tidyverse)
```

```
## -- Attaching packages --------------------------------------- tidyverse 1.3.2 --
## v ggplot2 3.4.0      v purrr   0.3.5
## v tibble  3.1.8      v dplyr   1.0.10
## v tidyr   1.2.1      v stringr 1.5.0
## v readr   2.1.3      v forcats 0.5.2
## -- Conflicts ------------------------------------------ tidyverse_conflicts() --
## x dplyr::filter() masks stats::filter()
## x dplyr::lag()    masks stats::lag()
```

```
library(lubridate)
```

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

```
library(here)
```

```
## here() starts at /Users/anabishop/Documents/Data Analytics/ENV872
```

```
library(cowplot)
```

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

```
library(ggthemes)
```

```
##
## Attaching package: 'ggthemes'
##
## The following object is masked from 'package:cowplot':
##
##     theme_map
```

```
# check wd
getwd()  #good
```

```
## [1] "/Users/anabishop/Documents/Data Analytics/ENV872"
```

```
here()  #good
```

```
## [1] "/Users/anabishop/Documents/Data Analytics/ENV872"
```

```
# load data: NTL-LTER processed
# nutrient data, chem/physics data, and
# Niwot Ridge litter data
pp_chem_nutrients <- read.csv(here("Data/Processed_KEY/NTL-LTER_Lake_Chemistry_Nutrients_PeterPaul_Proc
    stringsAsFactors = TRUE)

pp_chem_physics <- read.csv(here("Data/Processed_KEY/NTL-LTER_Lake_ChemistryPhysics_PeterPaul_Processed
    stringsAsFactors = TRUE)
```

```r
niwotridge <- read.csv(here("Data/Processed_KEY/NEON_NIWO_Litter_mass_trap_Processed.csv"),
    stringsAsFactors = TRUE)

# 2
class(pp_chem_nutrients$sampledate)  #factor
```

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

```r
class(pp_chem_physics$sampledate)  #factor
```

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

```r
class(niwotridge$collectDate)  #factor
```

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

```r
# changing to date format
pp_chem_nutrients$sampledate <- ymd(pp_chem_nutrients$sampledate)
pp_chem_physics$sampledate <- ymd(pp_chem_physics$sampledate)
niwotridge$collectDate <- ymd(niwotridge$collectDate)

class(pp_chem_nutrients$sampledate)  #date
```

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

```r
class(pp_chem_physics$sampledate)  #date
```

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

```r
class(niwotridge$collectDate)  #date
```

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

Define your theme

3. Build a theme and set it as your default theme. Customize the look of at least two of the following:

- Plot background
- Plot title
- Axis labels
- Axis ticks/gridlines
- Legend

```r
# 3
my_theme <- theme_base() + theme(rect = element_rect(color = "black",
    fill = "white"), plot.title = element_text(size = 15,
    hjust = 0.5), axis.text = element_text(color = "black",
    size = 10), panel.grid.major = element_line(color = "lightgray"),
```

```r
    plot.background = element_rect(color = "black",
        fill = "lightblue"), legend.position = "right",
    )

# set as default
theme_set(my_theme)
```

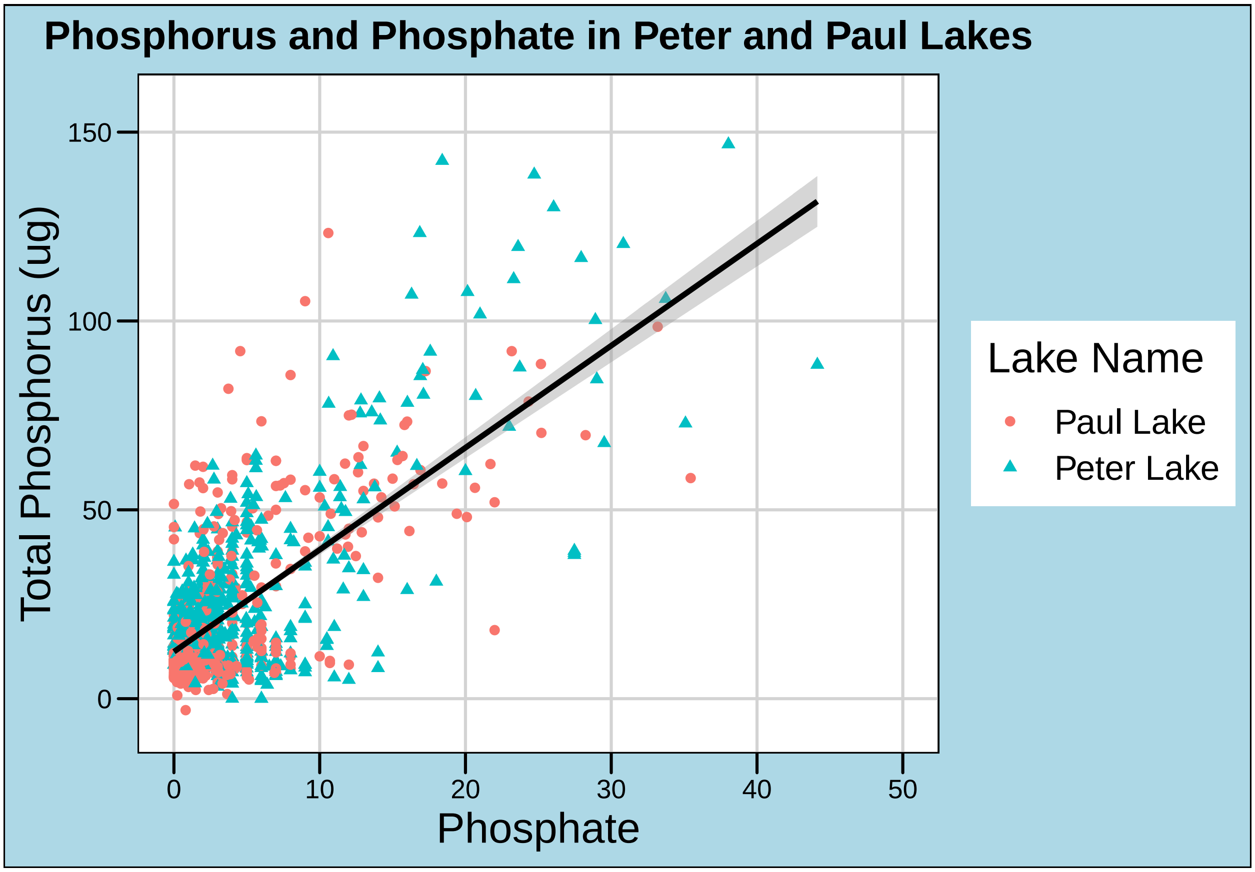
Create graphs

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/or `ylim()`).

```r
# 4
phosphorus_phosphate <- ggplot(pp_chem_nutrients,
    aes(x = po4, y = tp_ug)) + geom_point(aes(color = lakename,
    shape = lakename)) + geom_smooth(method = lm,
    color = "black") + xlim(0, 50) + labs(x = "Phosphate",
    y = "Total Phosphorus (ug)", color = "Lake Name",
    shape = "Lake Name", title = "Phosphorus and Phosphate in Peter and Paul Lakes")

print(phosphorus_phosphate)
```

```
## `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()`).
```

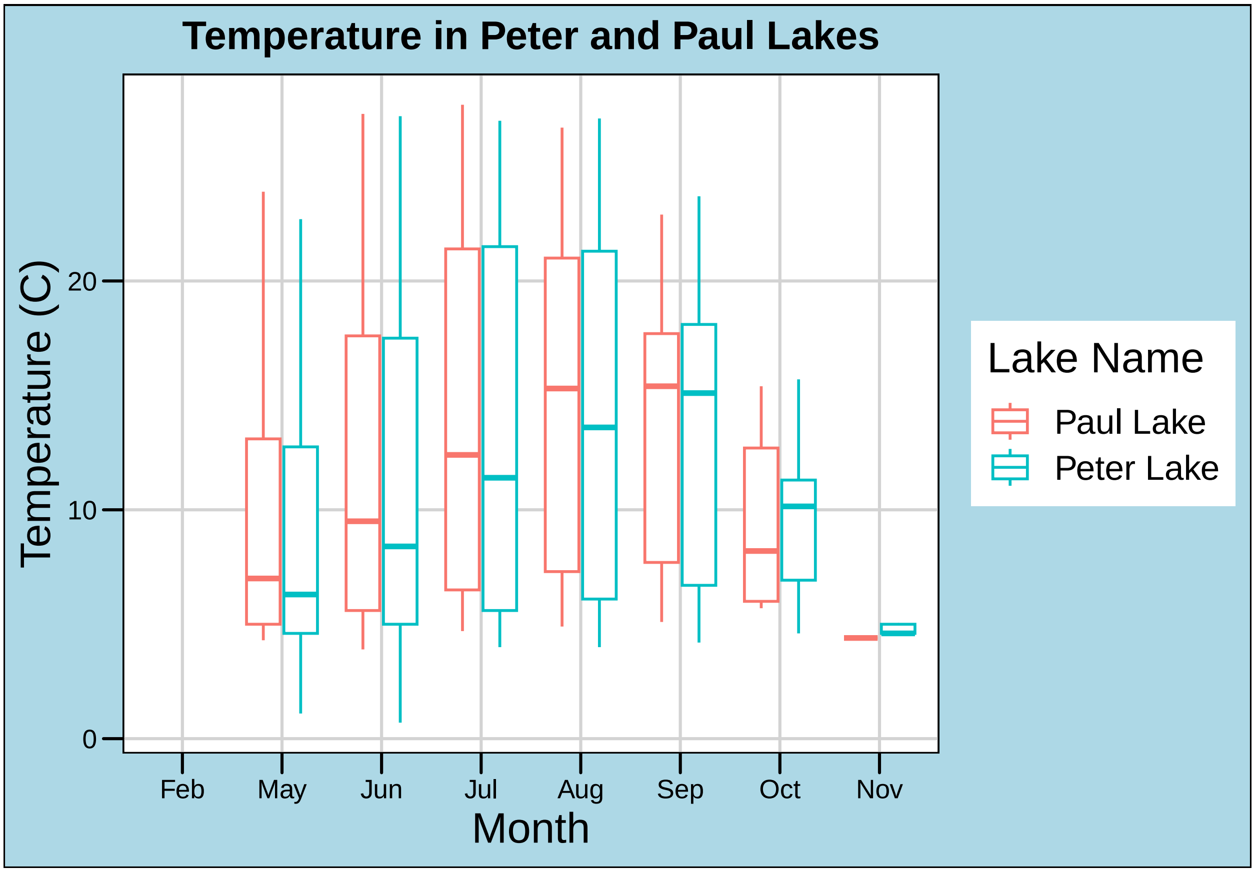


5. [NTL-LTER] Make three separate boxplots of (a) temperature, (b) TP, and (c) TN, with month as the x axis and lake as a color aesthetic. Then, create a cowplot that combines the three graphs. Make sure that only one legend is present and that graph axes are aligned.

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

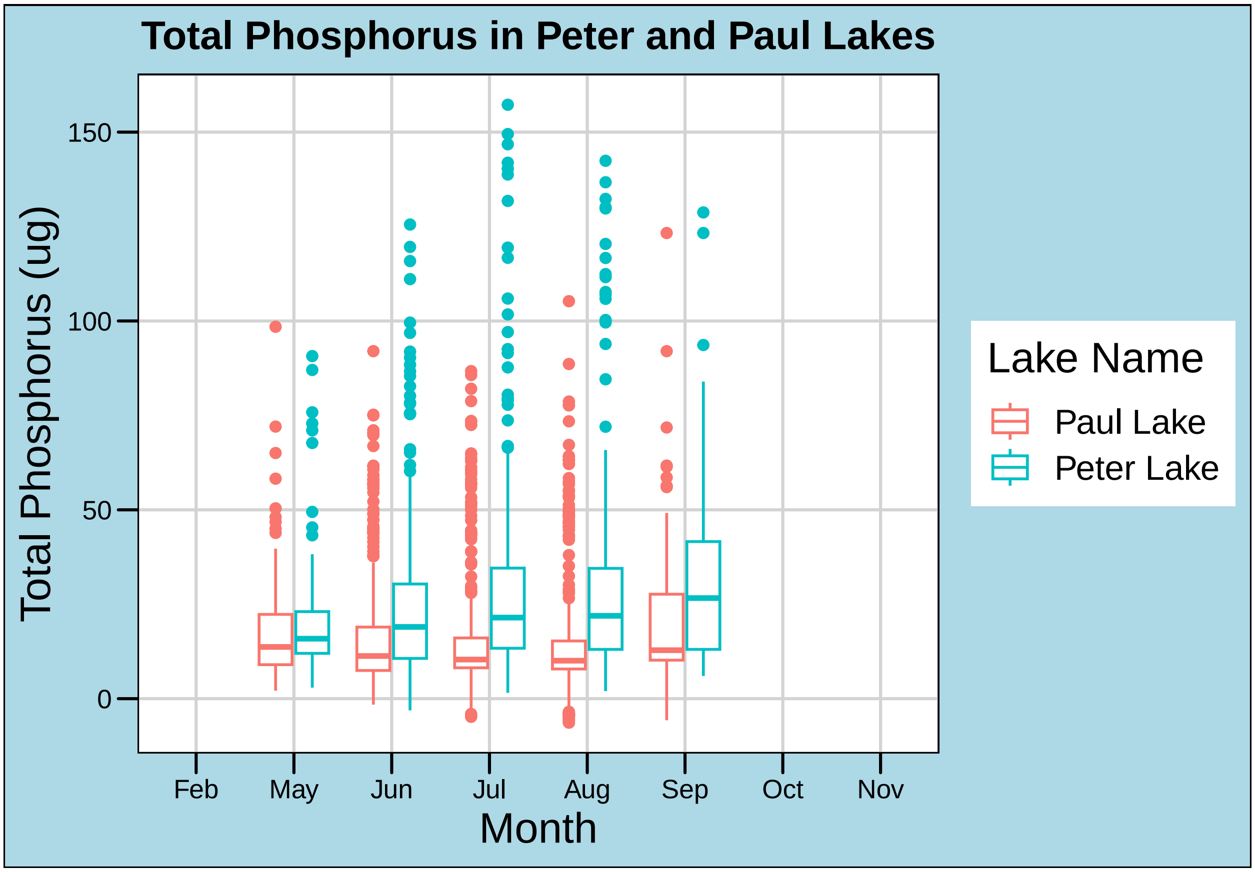
```
# 5
pp_chem_nutrients$month <- as.factor(pp_chem_nutrients$month)  #change month column to a factor

# temp boxplot
temp <- ggplot(pp_chem_nutrients, aes(x = factor(month,
    level = 1:12, labels = month.abb), y = temperature_C)) +
    geom_boxplot(aes(color = lakename)) +
    labs(x = "Month", y = "Temperature (C)",
        color = "Lake Name", title = "Temperature in Peter and Paul Lakes")

print(temp)
```

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



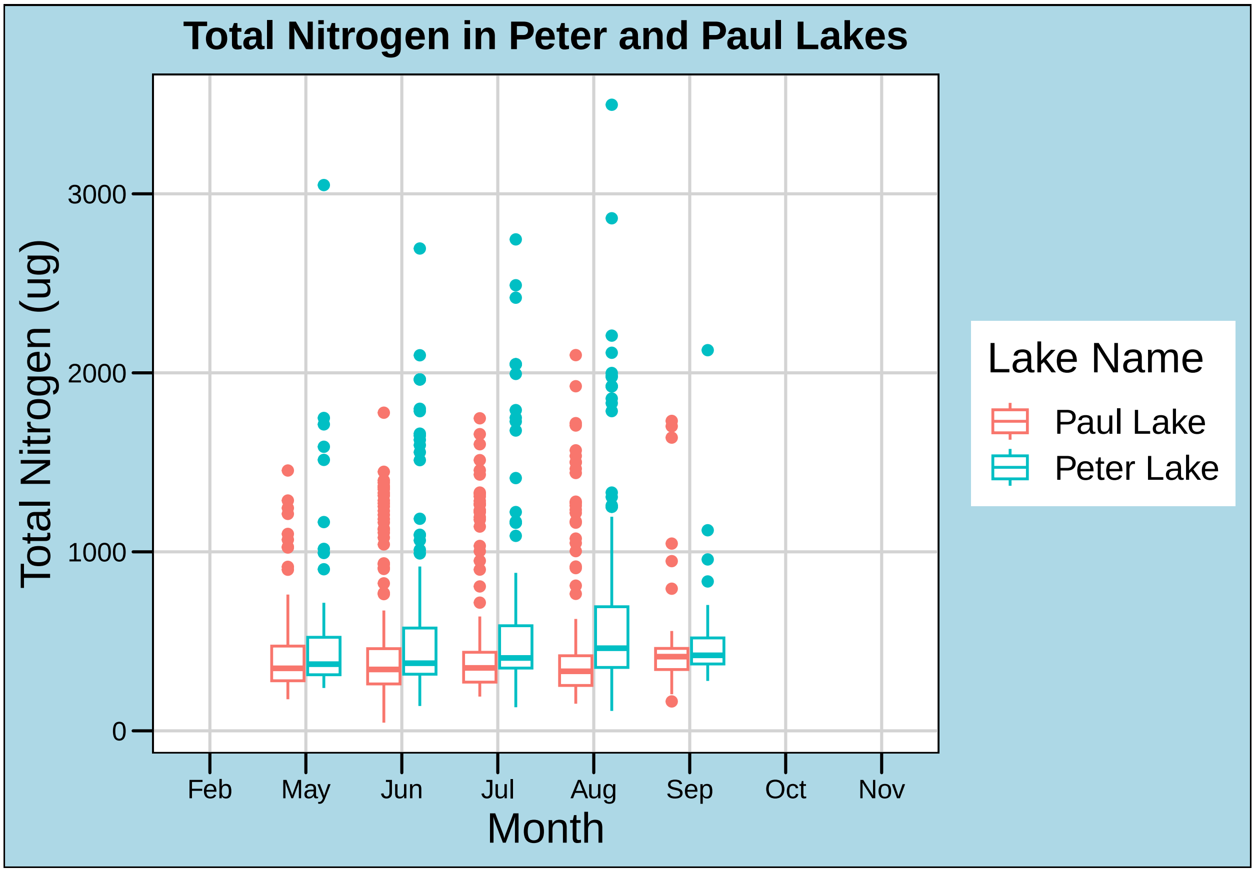

```r
# phosphorus boxplot
tp <- ggplot(pp_chem_nutrients, aes(x = factor(month,
    level = 1:12, labels = month.abb), y = tp_ug)) +
    geom_boxplot(aes(color = lakename)) +
    labs(x = "Month", y = "Total Phosphorus (ug)",
        color = "Lake Name", title = "Total Phosphorus in Peter and Paul Lakes")

print(tp)
```

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




```
# nitrogen boxplot
tn <- ggplot(pp_chem_nutrients, aes(x = factor(month,
    level = 1:12, labels = month.abb), y = tn_ug)) +
    geom_boxplot(aes(color = lakename)) +
    labs(x = "Month", y = "Total Nitrogen (ug)",
        color = "Lake Name", title = "Total Nitrogen in Peter and Paul Lakes")

print(tn)
```

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




```
# cowplot
lake_characteristics <- plot_grid(temp +
    theme(legend.position = "top"), tp +
    theme(legend.position = "none"), tn +
    theme(legend.position = "none"), nrow = 3,
    align = "h", rel_heights = c(3, 2.5,
        2.5))
```

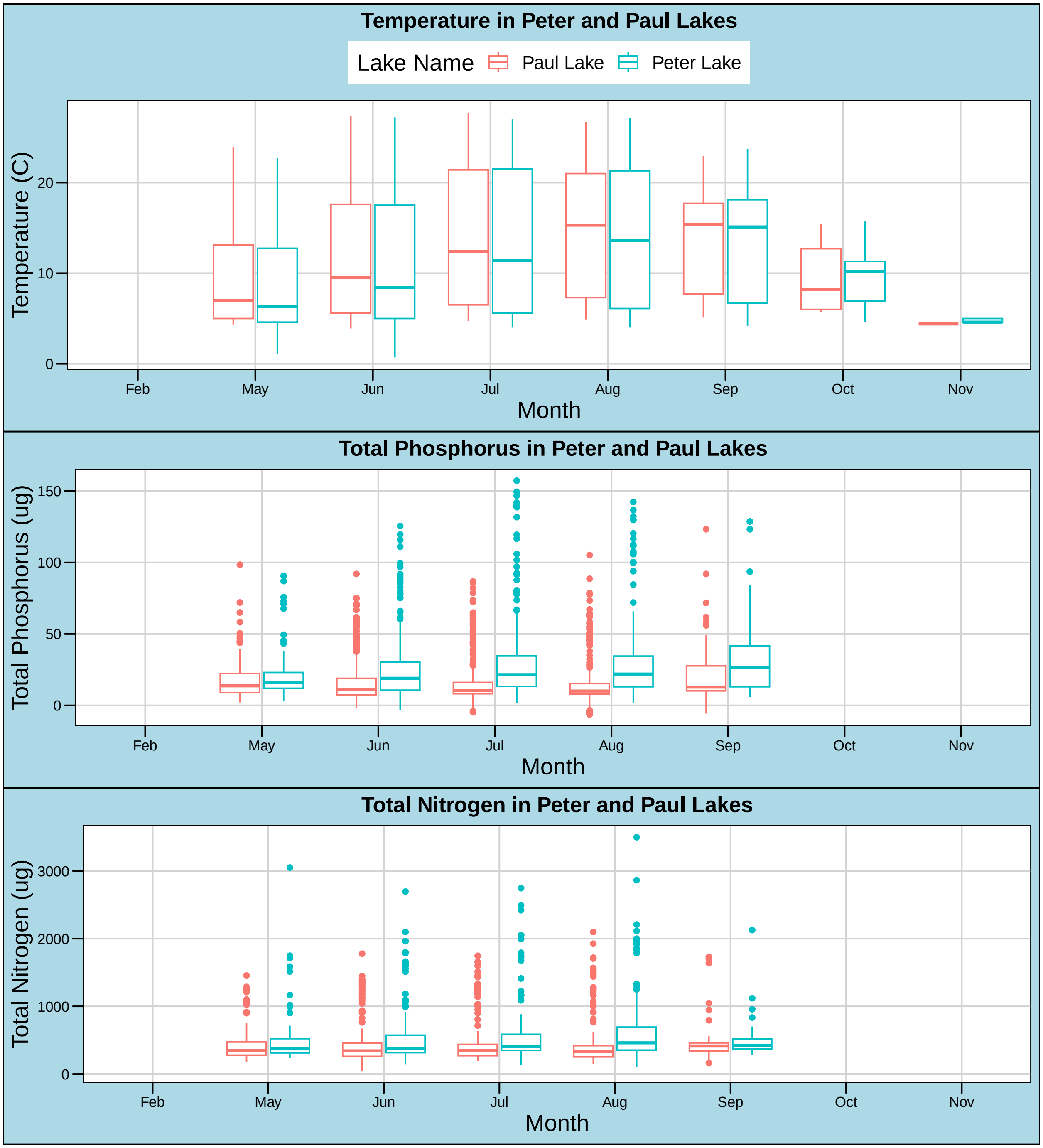
```
## Warning: Removed 3566 rows containing non-finite values (`stat_boxplot()`).
```

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

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

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

```
print(lake_characteristics)
```



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

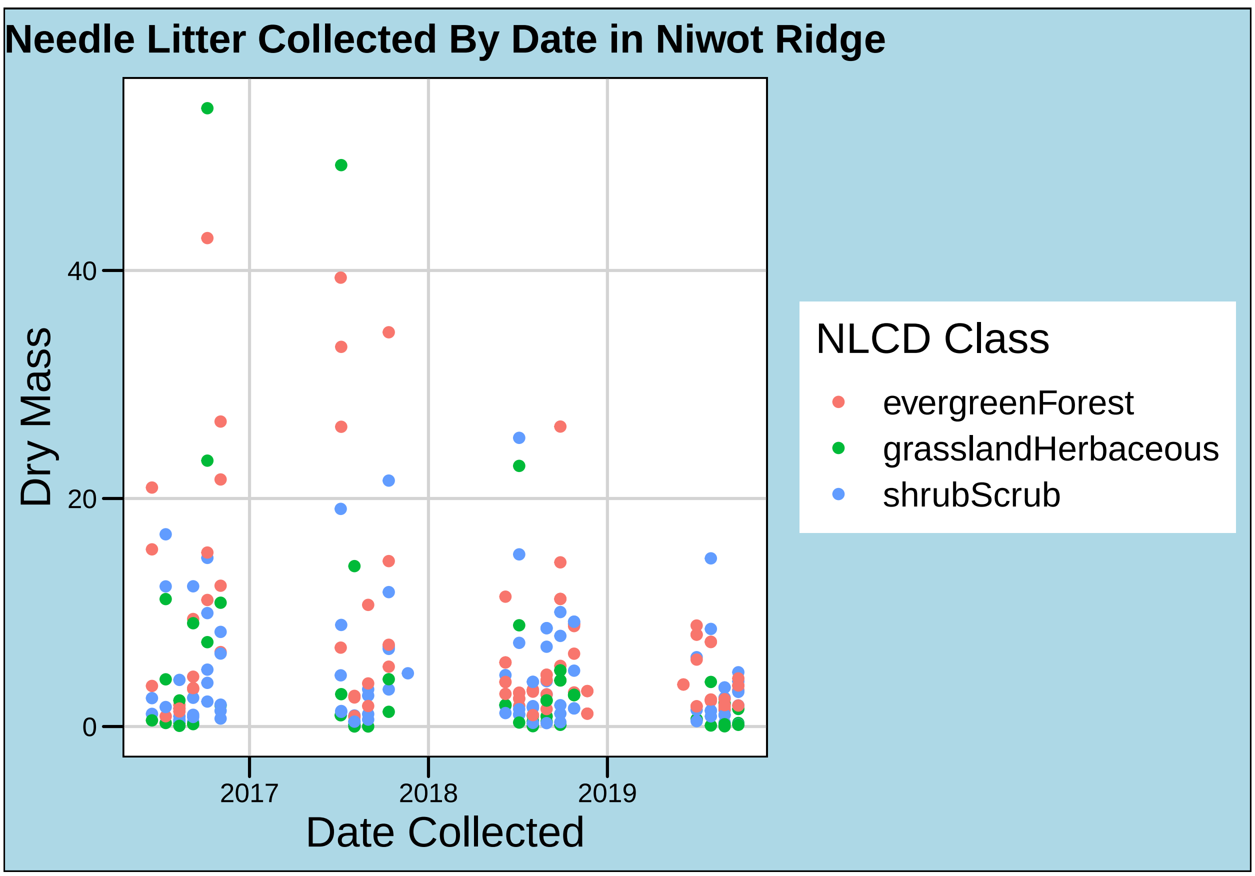
> Answer: Paul and Peter Lake had similar temperatures to each other throughout the entire study period. However, Paul Lake had a higher median temperature than Peter Lake for nearly all of the months. Unsurprisingly, the water temperature increased during the summer months, and was lower in the spring and fall. Total phosphorus had an inverse relationship between the two lakes over the study period; Peter Lake's amount slowly increased throughout the summer into the fall, while Paul Lake slowly declined throughout the summer, with a slight increase in the fall. However, the values in both lakes stayed relatively close to each other throughout the study period. The biggest discrepancy happened in July, when Peter Lake had much higher values

than Paul Lake. Finally, Peter Lake consistently had higher nitrogen amounts than Paul Lake, although the values remained fairly close to each other overall. Again, Peter Lake experienced a small but consistent increase throughout the summer, with the exception of September, while Peter Lake remained relatively stable throughout the study period.

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)

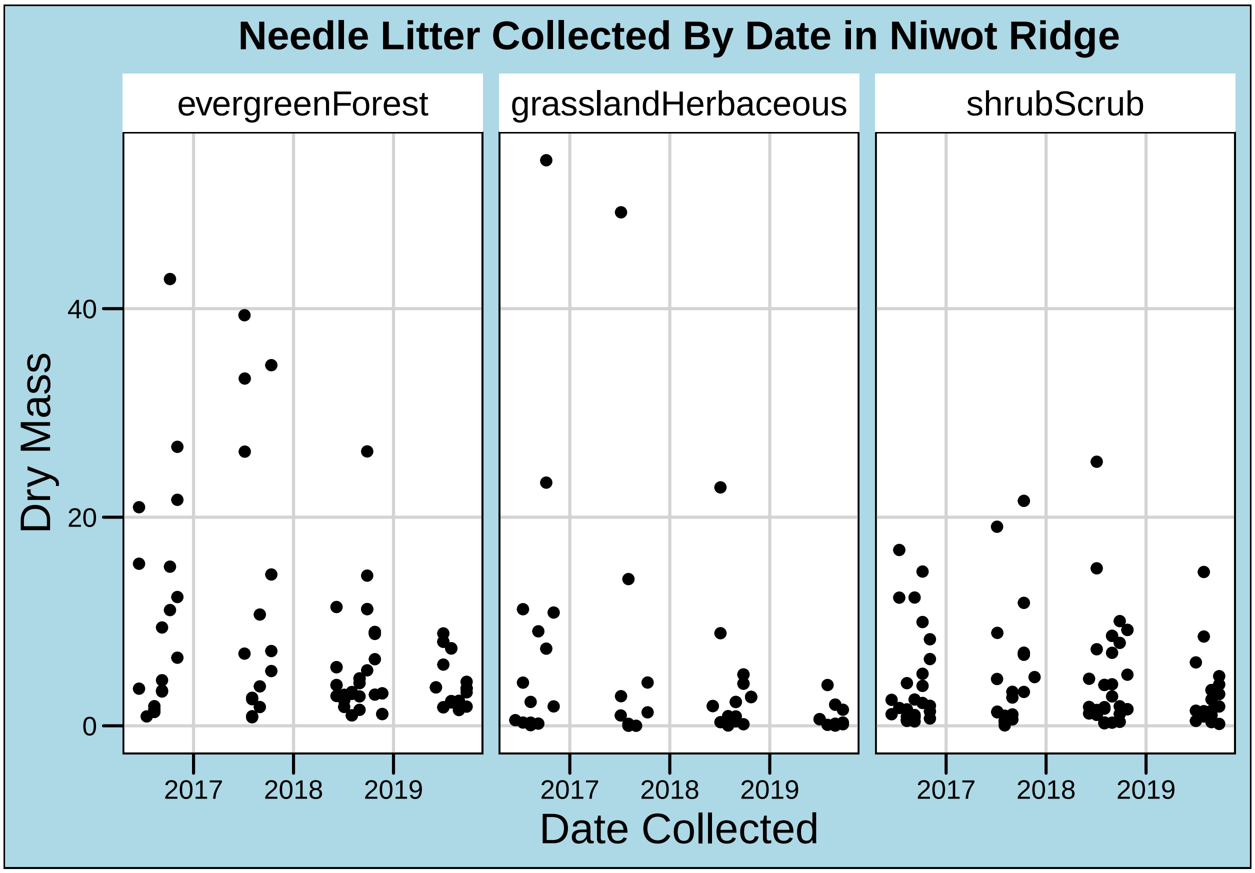
7. [Niwot Ridge] Now, plot the same plot but with NLCD classes separated into three facets rather than separated by color.

```
# 6
needle_litter <- niwotridge %>%
    filter(functionalGroup == "Needles") %>%
    ggplot(aes(x = collectDate, y = dryMass,
        color = nlcdClass)) + geom_point() +
    labs(x = "Date Collected", y = "Dry Mass",
        title = "Needle Litter Collected By Date in Niwot Ridge",
        color = "NLCD Class")

print(needle_litter)
```




```
# 7
needle_litter_faceted <- niwotridge %>%
```

```
    filter(functionalGroup == "Needles") %>%
    ggplot(aes(x = collectDate, y = dryMass)) +
    geom_point() + facet_wrap(vars(nlcdClass)) +
    labs(x = "Date Collected", y = "Dry Mass",
        title = "Needle Litter Collected By Date in Niwot Ridge")

print(needle_litter_faceted)
```



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

> Answer: Plot 7 is much more effective than plot 6 because it separates out each NLCD class in a way that can be visualized in its entirety. In plot 6, all of the data points are stacked on top of each other in a way that makes it nearly impossible to visualize the specific trends within each NLCD class. However, in plot 7, each class can be viewed within its own frame and compared across frames in a much simpler way. The data can be understood more effectively and accurately in this format.