Assignment 5: Data Visualization

Kayla Emerson

Fall 2024

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. Read in 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 in the Processed_KEY folder) and the processed data file for the Niwot Ridge litter dataset (use the `NEON_NIWO_Litter_mass_trap_Processed.csv` version, again from the Processed_KEY folder).

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

```
#1
#load packages
#install.packages("thematic")
library(tidyverse)
```

```
## -- Attaching core tidyverse packages ------------------------ tidyverse 2.0.0 --
## v dplyr     1.1.4     v readr     2.1.5
## v forcats   1.0.0     v stringr   1.5.1
## v ggplot2   3.5.1     v tibble    3.2.1
## v lubridate 1.9.3     v tidyr     1.3.1
## v purrr     1.0.2
## -- Conflicts ------------------------------------------ tidyverse_conflicts() --
## x dplyr::filter() masks stats::filter()
## x dplyr::lag()    masks stats::lag()
## i Use the conflicted package (<http://conflicted.r-lib.org/>) to force all conflicts to become errors
```

```
library(here)
```

```
## here() starts at /home/guest/EDE_Fall2024
```

```
library(lubridate)
library(cowplot)
```

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

```
library(thematic)
library(ggthemes)
```

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

```
#verify home directory
here()
```

```
## [1] "/home/guest/EDE_Fall2024"
```

```
getwd()
```

```
## [1] "/home/guest/EDE_Fall2024"
```

```
#read in NTL-LTER processed data files
#make processed data object
processed_data = "./Data/Processed_KEY"

#Read in data
PeterPaul.chem.nutrients <- read.csv(
  here(processed_data,"NTL-LTER_Lake_Chemistry_Nutrients_PeterPaul_Processed.csv"),
  stringsAsFactors = TRUE)

Neon_Niwo_Litter <- read.csv(
  here(processed_data,"NEON_NIWO_Litter_mass_trap_Processed.csv"),
  stringsAsFactors = TRUE)
#2
#check date format
class(PeterPaul.chem.nutrients$sampledate)
```

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

```
class(Neon_Niwo_Litter$collectDate)
```

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

```
#change to date
PeterPaul.chem.nutrients$sampledate <- ymd(PeterPaul.chem.nutrients$sampledate)
class(PeterPaul.chem.nutrients$sampledate)
```

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

```
Neon_Niwo_Litter$collectDate <- ymd(Neon_Niwo_Litter$collectDate)
class(Neon_Niwo_Litter$collectDate)
```

```
## [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

```
#3
#build theme!
#customize line color and width, legend title and location, and plot background
kme_theme <- theme_base() +
  theme(
    line = element_line(
      color='black',
      linewidth =0.5
    ),
    plot.background =   element_rect(
      fill = 'white'
    ),
    legend.background = element_rect(
      color='grey',
      fill = 'aliceblue'
    ),
    legend.title = element_text(
      color= 'darkblue',
      size = 12
    ),
    legend.position = 'right',
    plot.title = element_text(
      size = 12
    )

  )
theme_set(kme_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 line(s) of best fit using the `lm` method. Adjust your axes to hide extreme values (hint: change the limits using `xlim()` and/or `ylim()`).

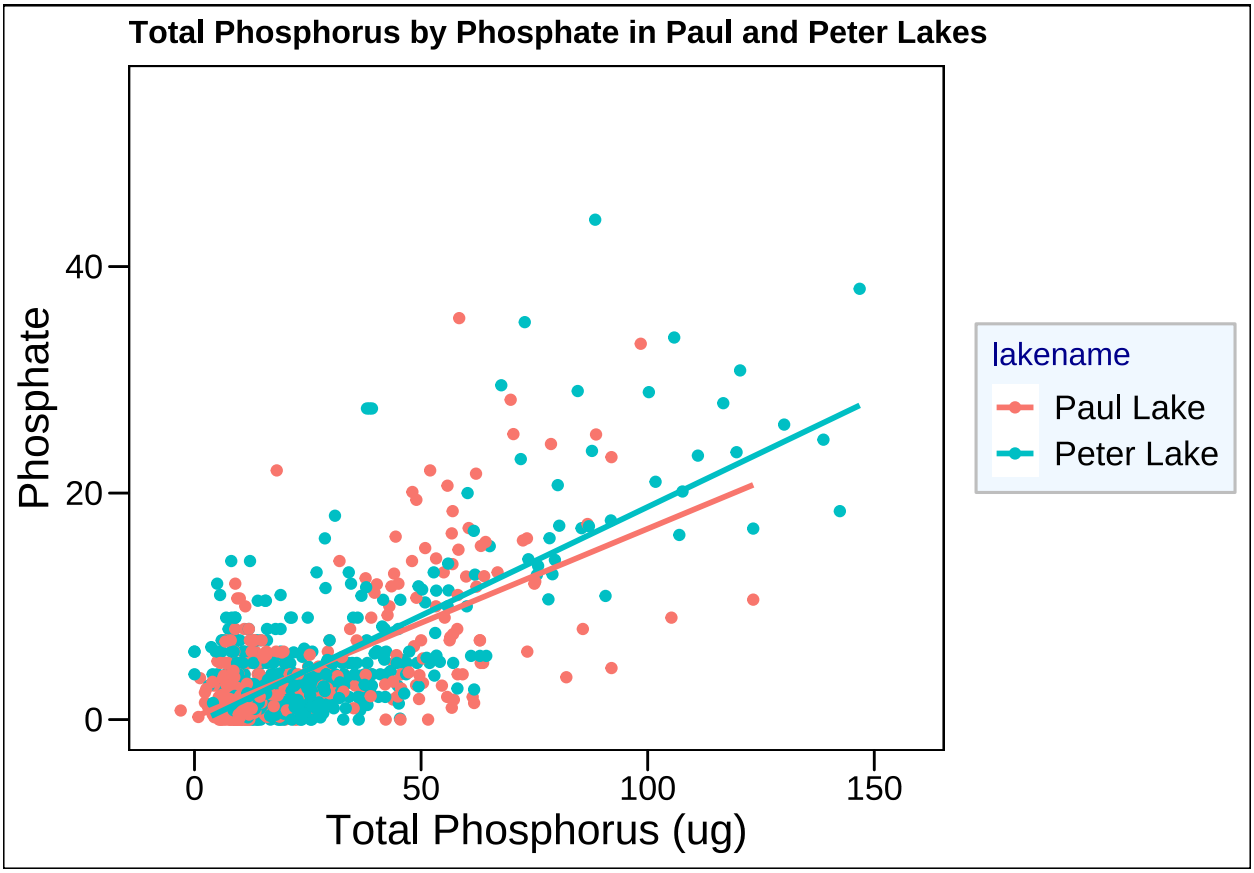
```
#4
# plot total phosphorus by phosphate, color by lake name, include lines of best
#fit, remove extreme values

plot_four <- PeterPaul.chem.nutrients %>%
  ggplot(aes(
      x = tp_ug,
      y = po4,
      color = lakename
      ),
      alpha = 0.25) +
  geom_point() +
  ylim(0,55)  +
  geom_smooth(
    method = lm,
    se = FALSE) +
  labs(
    title = "Total Phosphorus by Phosphate in Paul and Peter Lakes",
    y = "Phosphate",
    x = "Total Phosphorus (ug)"
  )

plot_four
```

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

```
## Warning: Removed 21947 rows containing non-finite outside the scale range
## (`stat_smooth()`).
```

```
## Warning: Removed 21947 rows containing missing values or values outside the scale range
## (`geom_point()`).
```

```
## Warning: Removed 4 rows containing missing values or values outside the scale range
## (`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.

Tips: * Recall the discussion on factors in the lab section as it may be helpful here. * Setting an axis title in your theme to `element_blank()` removes the axis title (useful when multiple, aligned plots use the same axis values) * Setting a legend's position to "none" will remove the legend from a plot. * Individual plots can have different sizes when combined using `cowplot`.

```
#5
#change month to a factor and by unique name
class(PeterPaul.chem.nutrients$month)
```

```
## [1] "integer"
```

```
unique(PeterPaul.chem.nutrients$month)
```

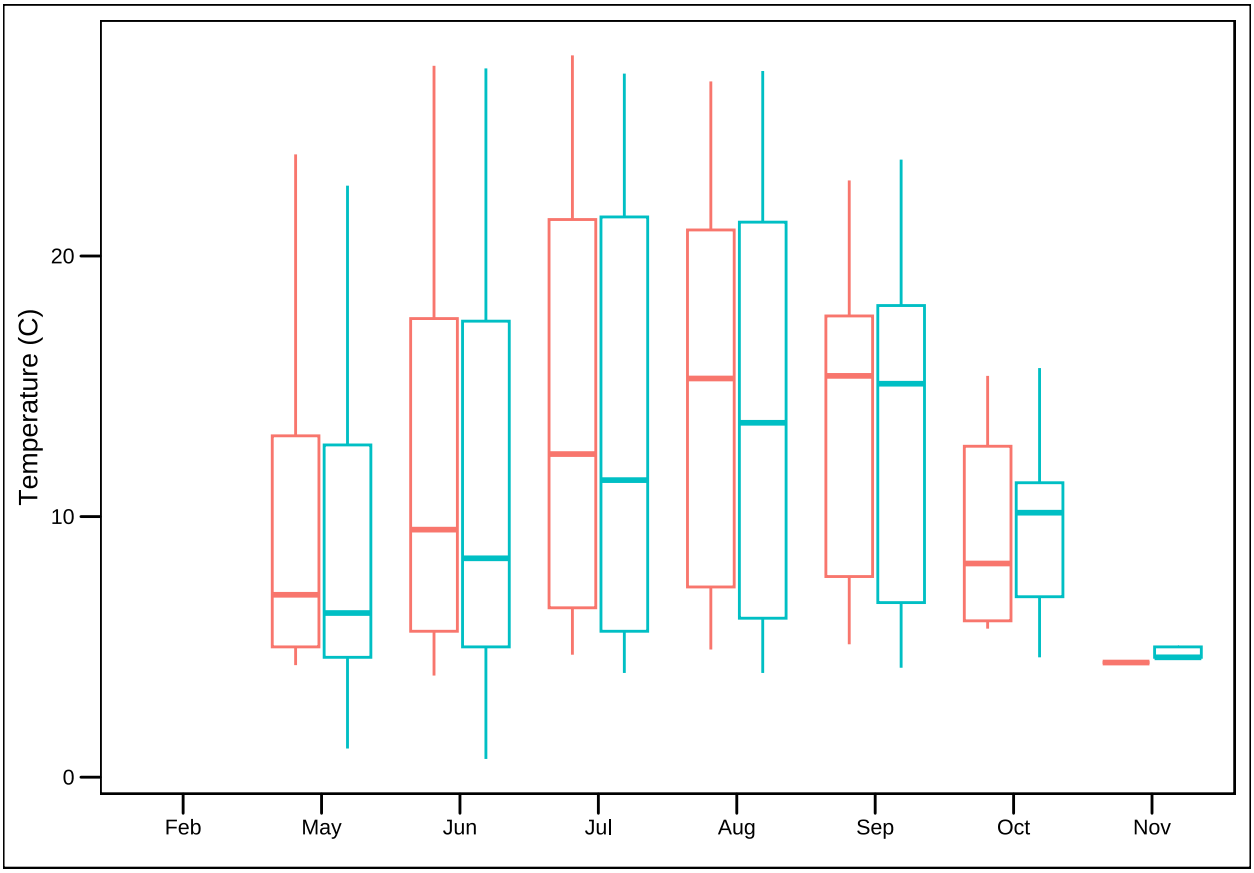
```
## [1]  5  6  7  8  9 10 11  2
```

```
PeterPaul.chem.nutrients$month <- factor(PeterPaul.chem.nutrients$month,
       levels = 1:12,
       labels = month.abb
       )
class(PeterPaul.chem.nutrients$month)
```

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

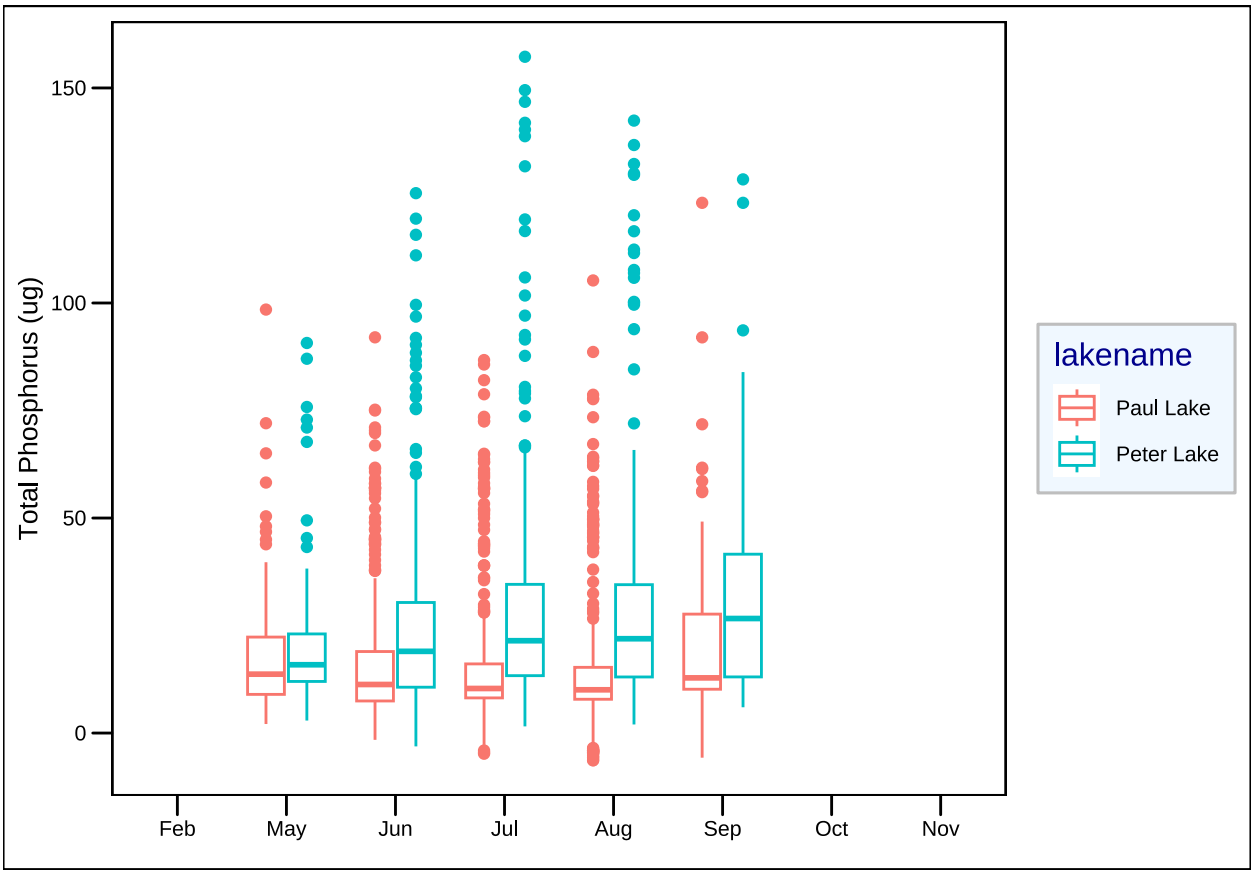
```
#plot boxplot of temperature by month
plot_five_a <- PeterPaul.chem.nutrients %>%
  ggplot(aes(
      x = month,
      y = temperature_C,
      color = lakename
      ),
      alpha = 0.25) +
  geom_boxplot() +
  theme(legend.position = "none",
        text = element_text(size = 10)) +
  labs(
    x = element_blank(),
    y = "Temperature (C)"
  )

plot_five_a
```

```
## Warning: Removed 3566 rows containing non-finite outside the scale range
## (`stat_boxplot()`).
```



```
#plot boxplot of TP by month
plot_five_b <- PeterPaul.chem.nutrients %>%
  ggplot(aes(
      x = month,
      y = tp_ug,
      color = lakename
      ),
      alpha = 0.25) +
  geom_boxplot() +
  theme(legend.position = "right",
        text = element_text(size = 10)) +
  labs(
    x = element_blank(),
    y = "Total Phosphorus (ug)"
  )

plot_five_b
```

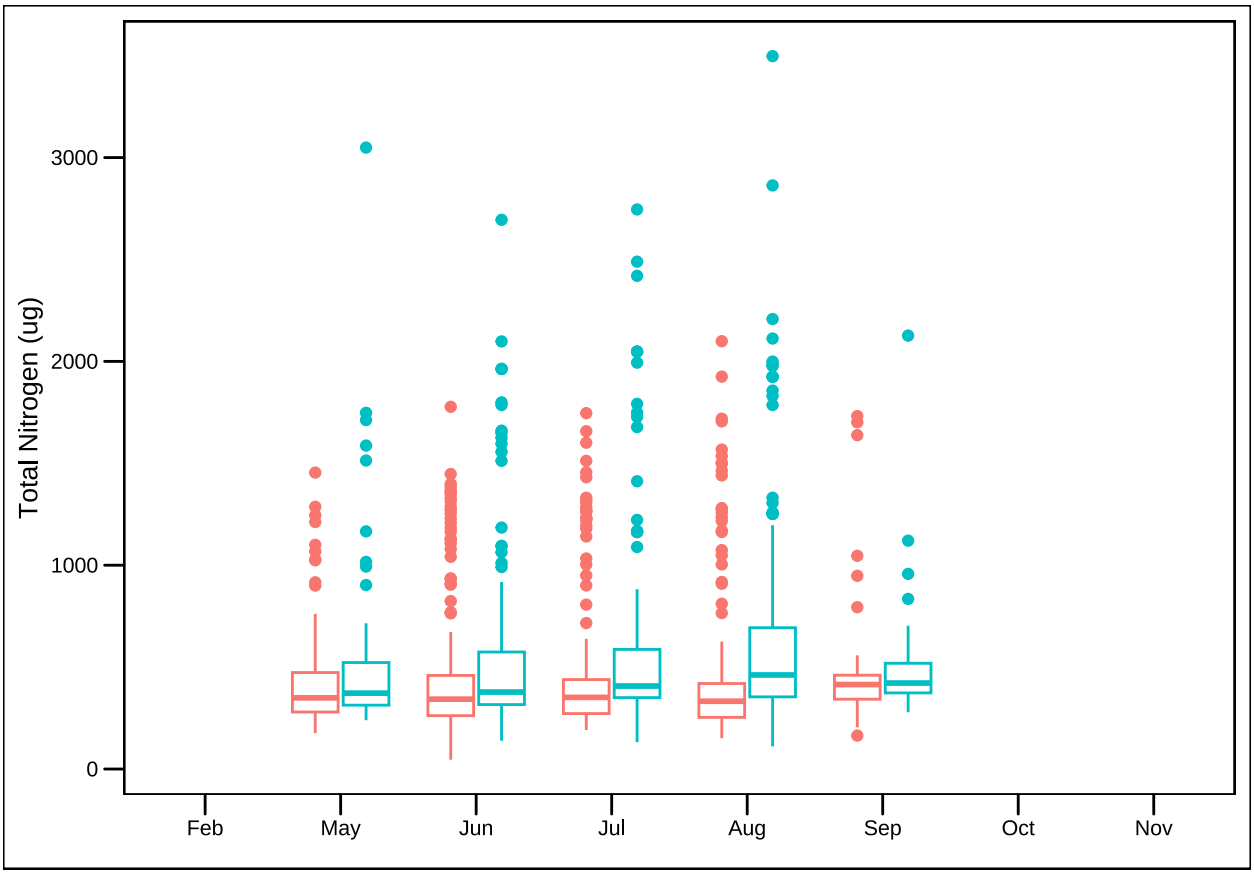
```
## Warning: Removed 20729 rows containing non-finite outside the scale range
## (‘stat_boxplot()‘).
```




```
#plot boxplot of TN by month
plot_five_c <- PeterPaul.chem.nutrients %>%
  ggplot(aes(
      x = month,
```

```
      y = tn_ug,
      color = lakename
      ),
      alpha = 0.25) +
  geom_boxplot() +
  theme(legend.position = "none",
        text = element_text(size = 10)) +
  labs(
    x = element_blank(),
    y = "Total Nitrogen (ug)"
  )

plot_five_c
```

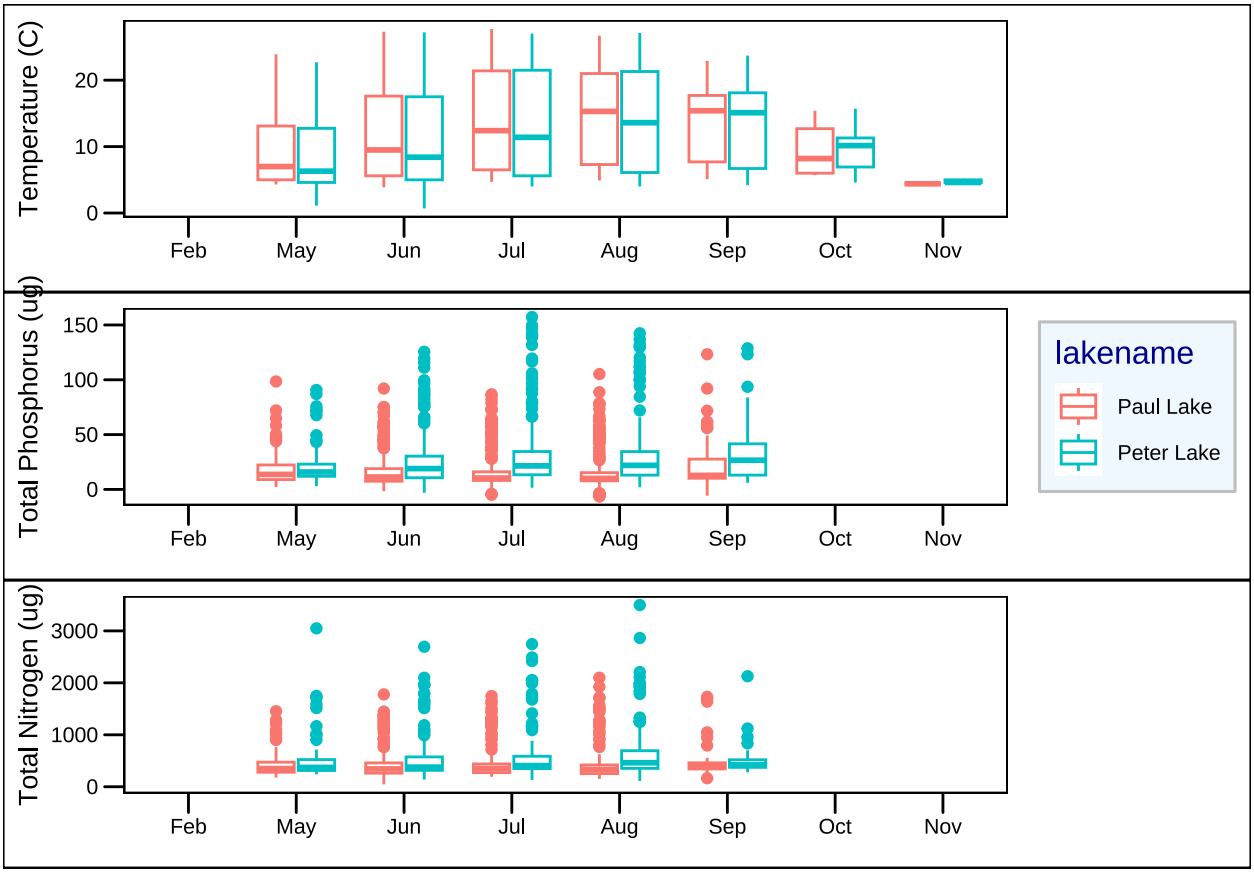
```
## Warning: Removed 21583 rows containing non-finite outside the scale range
## (‘stat_boxplot()‘).
```




```
#display all three plots together using cowplot
plot_grid(plot_five_a, plot_five_b, plot_five_c, nrow = 3, align = "v")
```

```
## Warning: Removed 3566 rows containing non-finite outside the scale range
## (‘stat_boxplot()‘).

## Warning: Removed 20729 rows containing non-finite outside the scale range
## (‘stat_boxplot()‘).
```

```
## Warning: Removed 21583 rows containing non-finite outside the scale range
## (`stat_boxplot()`).
```



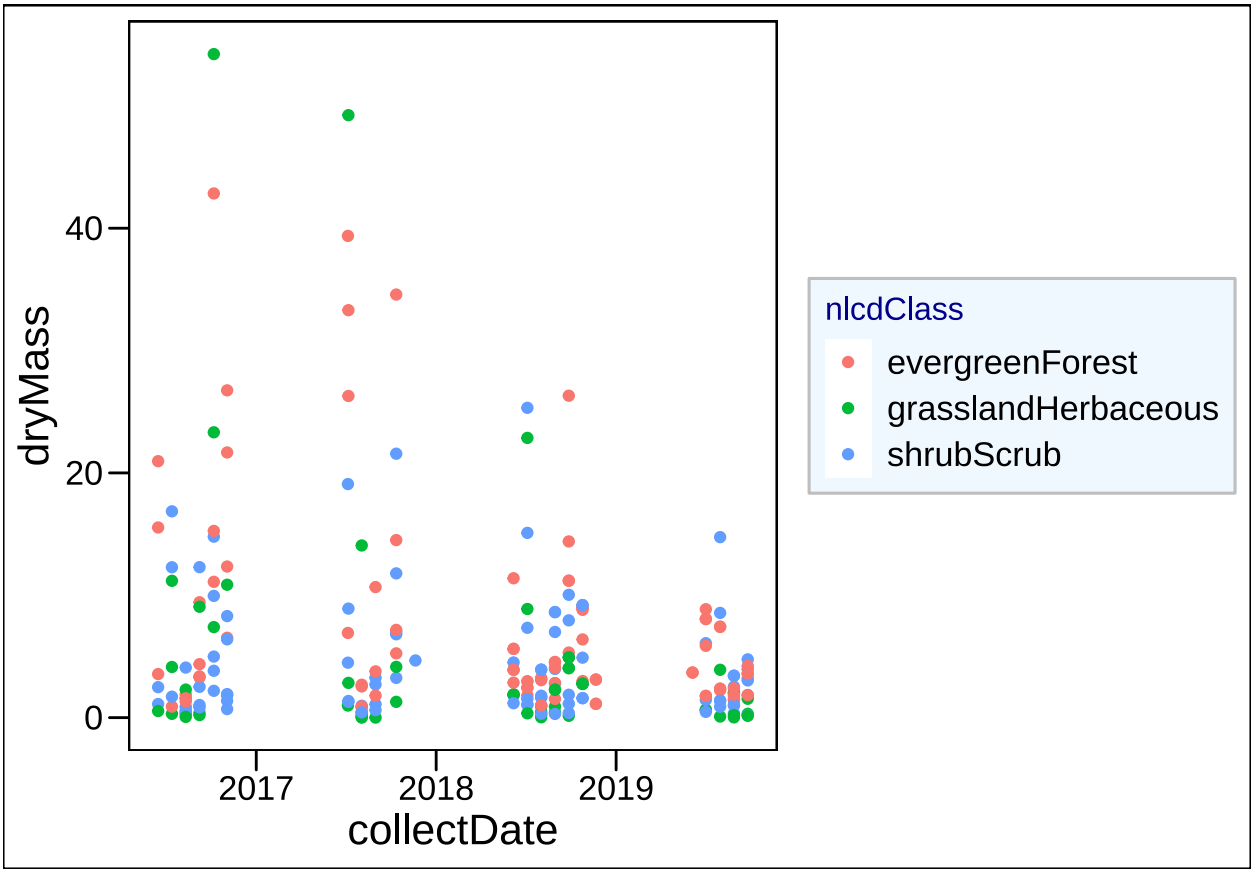
Question: What do you observe about the variables of interest over seasons and between lakes?

> Answer: The variables of interest are higher during the warmer months. Temperature appears to be similar between the two lakes but total phosphorus and total nitrogen are higher in Peter Lake.

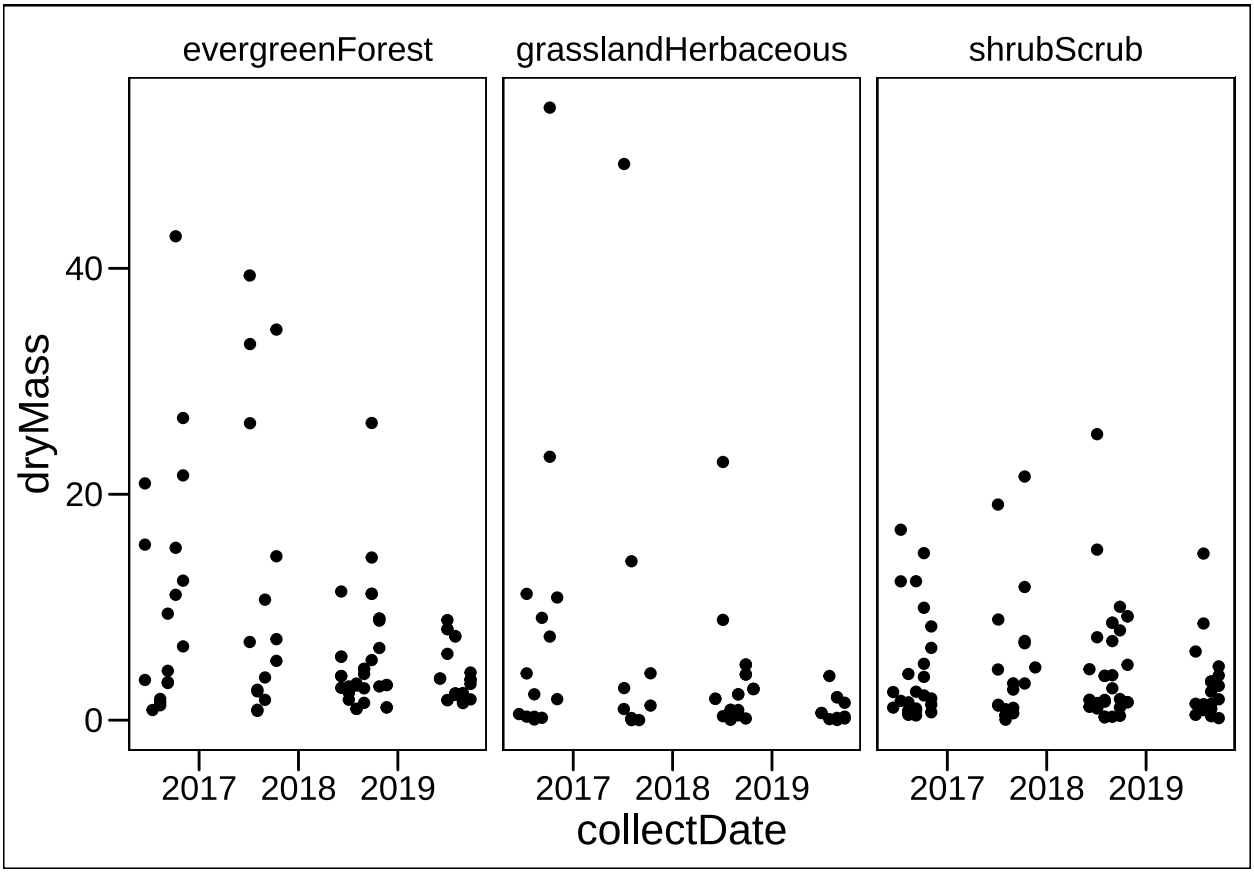
6. [Niwot Ridge] Plot a subset of the litter dataset by displaying only the "Needles" functional group. Plot the dry mass of needle litter by date and separate by NLCD class with a color aesthetic. (no need to adjust the name of each land use)

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

```
#6
#subset litter dataset, plot dry mass of needle litter by data and different land classes by color
plot_six <- Neon_Niwo_Litter %>%
   filter(functionalGroup == "Needles") %>%
  ggplot(aes(
      x=collectDate,
      y=dryMass,
      color=nlcdClass)) +
  geom_point()

plot_six
```




```
#7
#repeat plot but use three facets for land classes instead of separating by color
plot_seven <- Neon_Niwo_Litter %>%
   filter(functionalGroup == "Needles") %>%
  ggplot(aes(
      x=collectDate,
      y=dryMass)) +
  geom_point() +
  facet_wrap(facets=vars(nlcdClass), nrow = 1, ncol =3)

plot_seven
```



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

Answer: Plot 7 is more effetive at showing the difference between land class and dry mass by collection date because you can view all of the data side by side, with the same scale, instead of having to decipher between the colors. From plot 7 it is clear that evergreen forest and grassland herbaceous land classes have greater dry masses than shrub scrub.