Assignment 5: Data Visualization

Megan Lundequam

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
getwd()
```

```
## [1] "/Users/meganlundequam/Desktop/Spring 2022/Environmental Data Analytics/Git/Environmental_Data_An
```

```
library(tidyverse)
```

```
## -- 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)

PeterPaul.chem.nutrients <-
  read.csv("../Data/Processed/NTL-LTER_Lake_Chemistry_Nutrients_PeterPaul_Processed.csv")
Niwot.Ridge.Litter <-
```

```r
  read.csv("../Data/Processed/NEON_NIWO_Litter_mass_trap_Processed.csv")
#2
class(PeterPaul.chem.nutrients$sampledate)
```

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

```r
class(Niwot.Ridge.Litter$collectDate)
```

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

```r
PeterPaul.chem.nutrients$sampledate <- as.Date(
  PeterPaul.chem.nutrients$sampledate, format = "%Y-%m-%d")
Niwot.Ridge.Litter$collectDate <- as.Date(
  Niwot.Ridge.Litter$collectDate, format = "%Y-%m-%d")
```

Define your theme

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

```r
#3
A05theme <- theme_classic(base_size = 14) +
  theme(axis.text = element_text(color = "black"),
        legend.position = "right")

theme_set(A05theme)
```

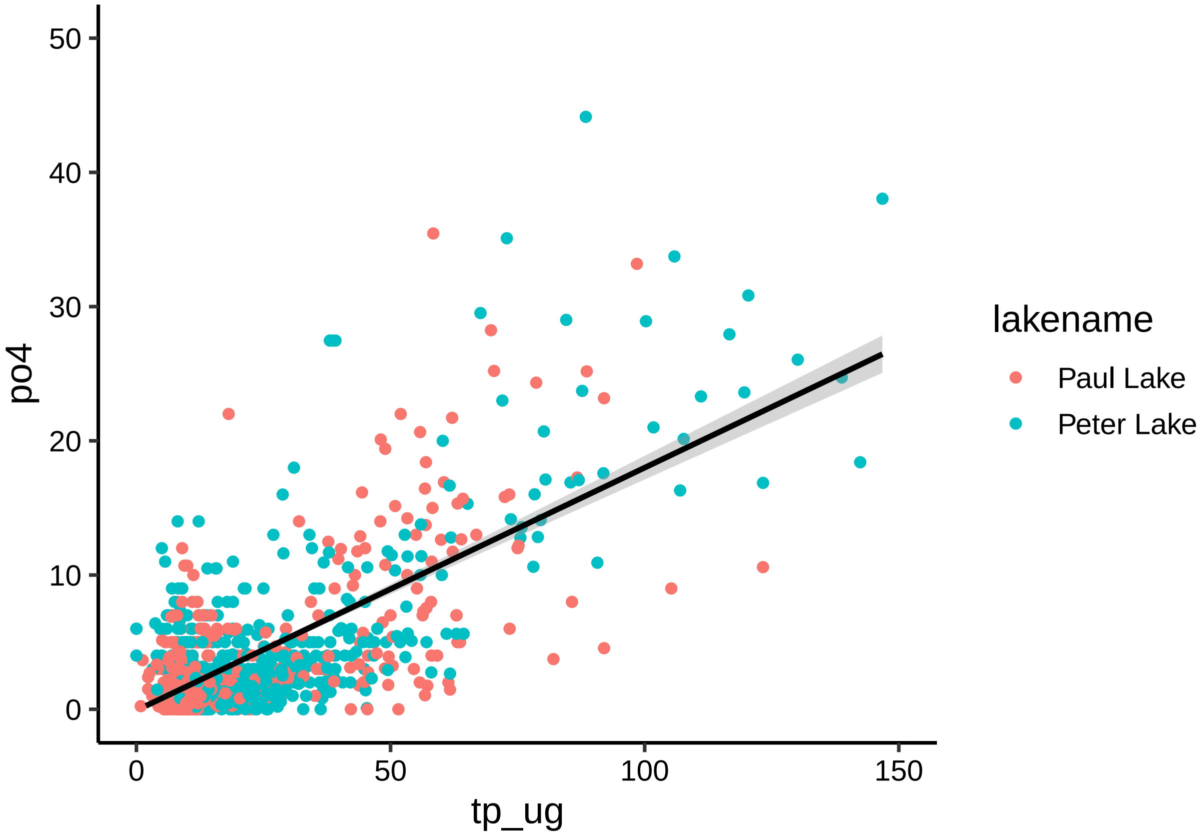
Create graphs

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
PhosphorusVPhosphate <-
  ggplot(PeterPaul.chem.nutrients, aes(x = tp_ug, y = po4, color = lakename)) +
  geom_point() +
  xlim(0, 150) +
  ylim(0, 50) +
  geom_smooth(method = lm, color = "black")
print(PhosphorusVPhosphate)
```

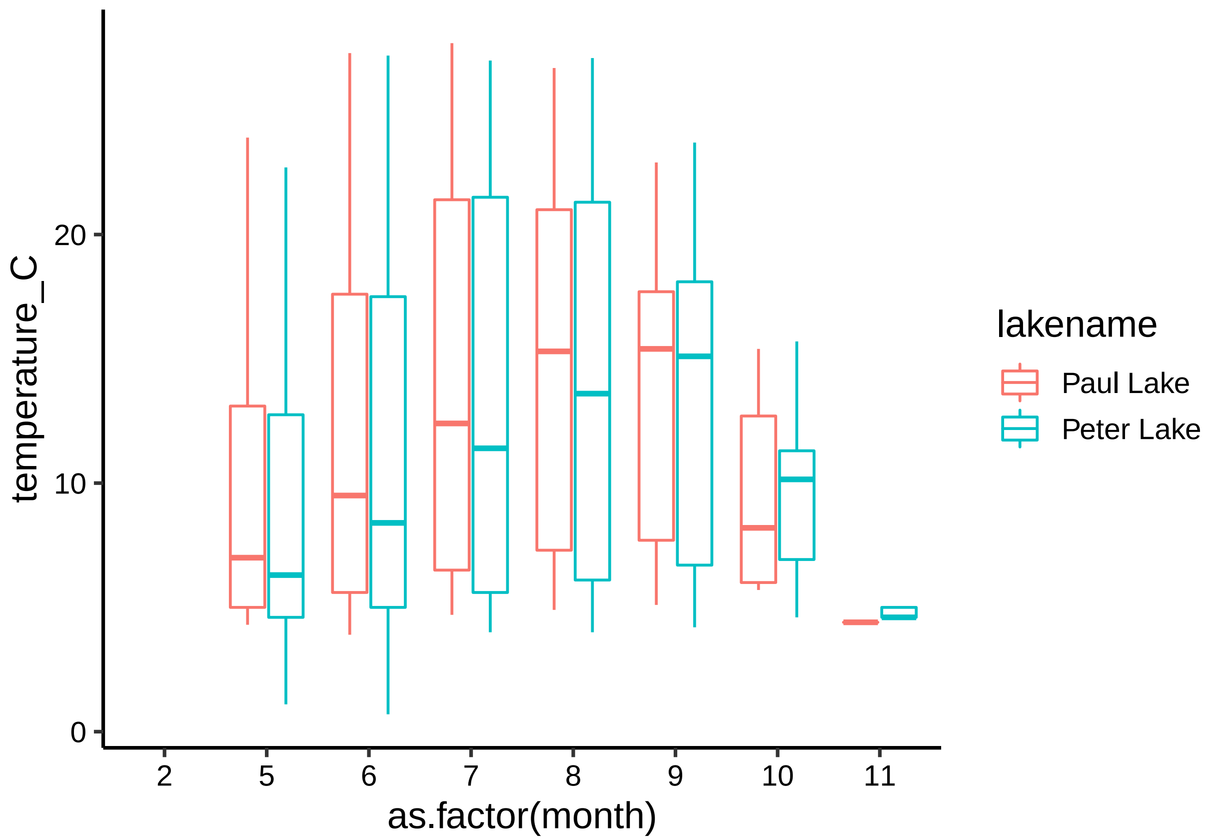
```
## `geom_smooth()` using formula 'y ~ x'

## 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.

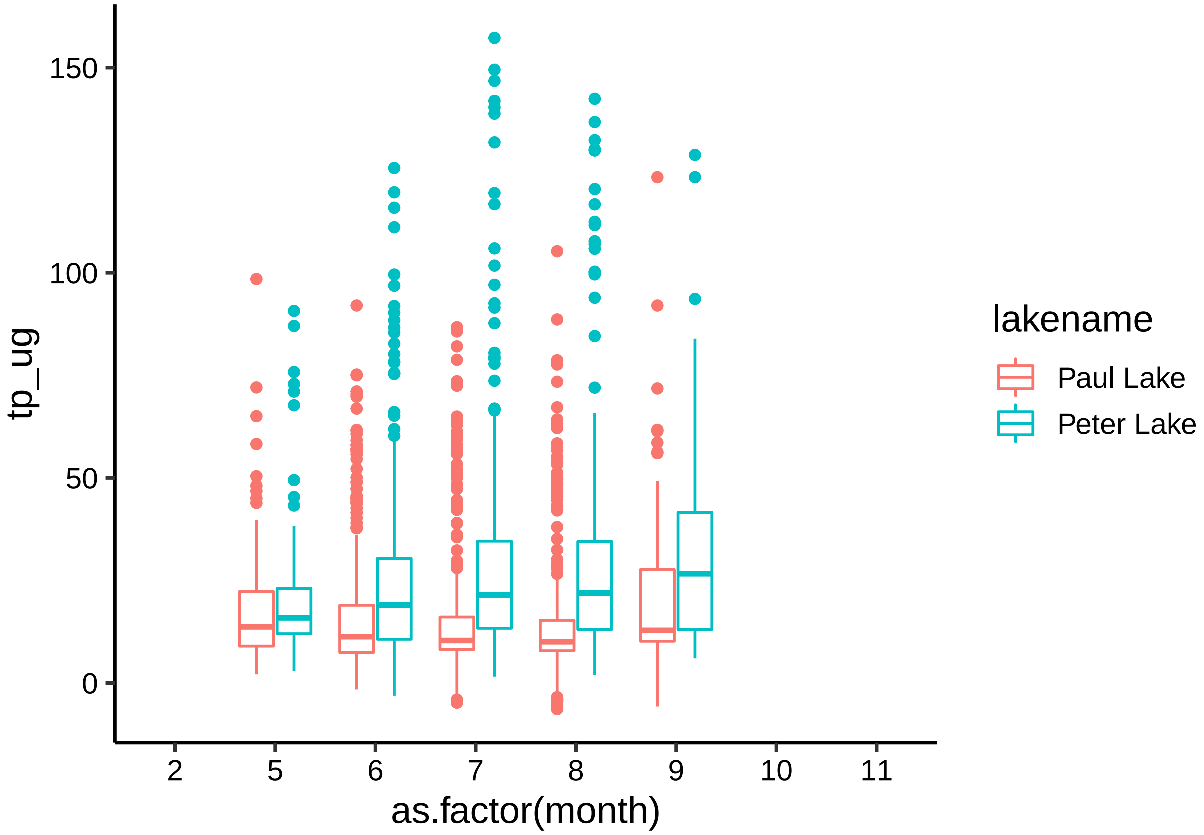
```
#5
Tempplot <-
  ggplot(PeterPaul.chem.nutrients, aes(x = as.factor(month), y = temperature_C)) +
  geom_boxplot(aes(color = lakename))
print(Tempplot)
```

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



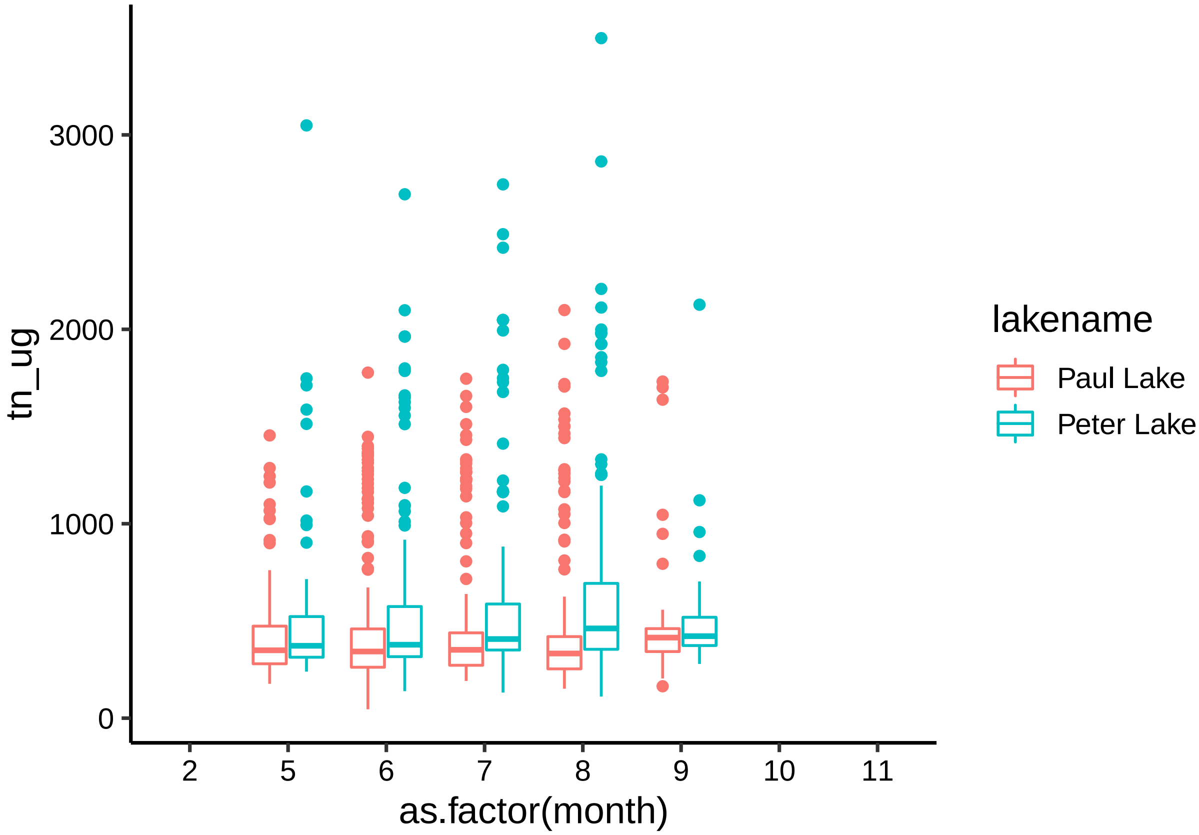

```
TPplot <-
  ggplot(PeterPaul.chem.nutrients, aes(x = as.factor(month), y = tp_ug)) +
  geom_boxplot(aes(color = lakename))
print(TPplot)
```

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




```
TNplot <-
  ggplot(PeterPaul.chem.nutrients, aes(x = as.factor(month), y = tn_ug)) +
  geom_boxplot(aes(color = lakename))
print(TNplot)
```

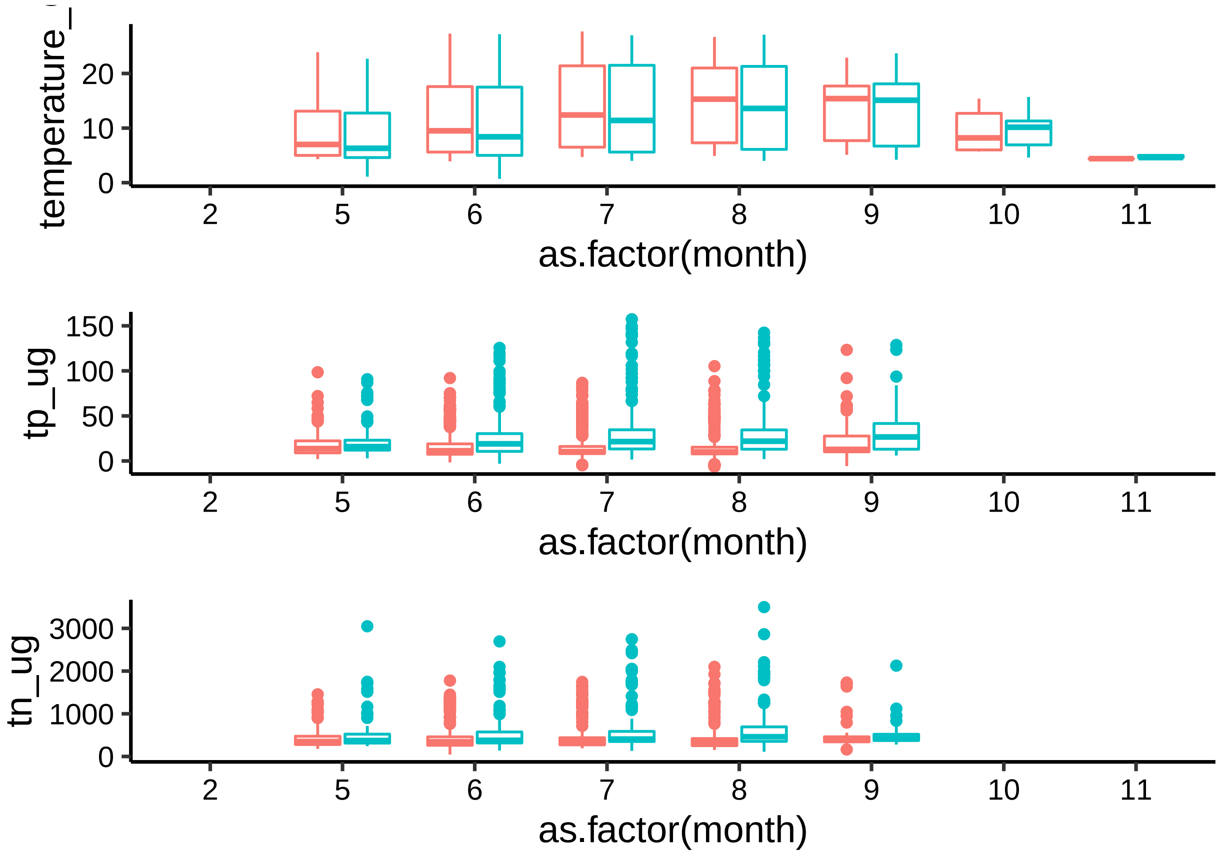
```
## Warning: Removed 21583 rows containing non-finite values (stat_boxplot).
```




```
TempvTPvTN <- plot_grid(
  Tempplot + theme(legend.position="none"),
  TPplot + theme(legend.position="none"),
  TNplot + theme(legend.position="none"),
  nrow = 3, align = 'v', axis = "l")
```

```
## 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).
```

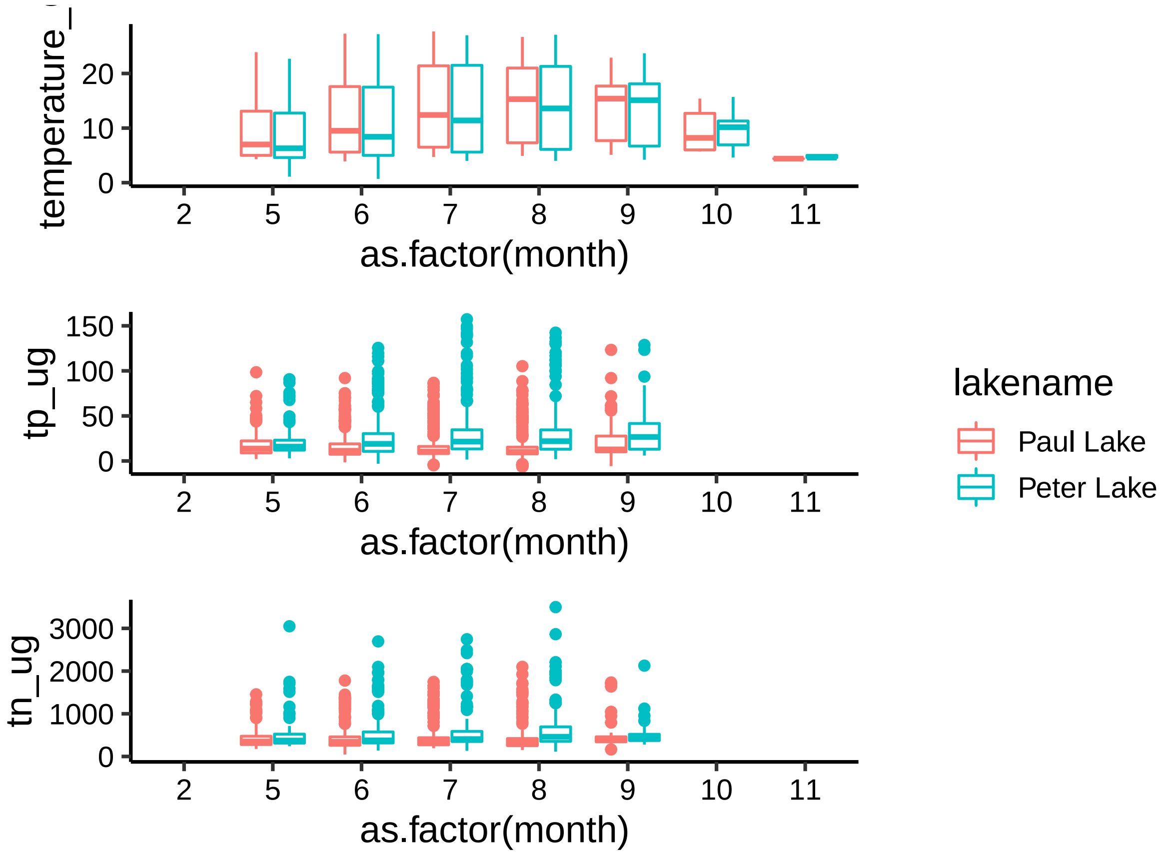
```
print(TempvTPvTN)
```




```
legend <- get_legend(
  # create some space to the left of the legend
  Tempplot + theme(legend.box.margin = margin(0, 1, 0, 1))
)
```

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

```
TempvTPvTN.legend <- plot_grid(TempvTPvTN, legend,
                               rel_widths = c(1, .4)
                               )
print(TempvTPvTN.legend)
```

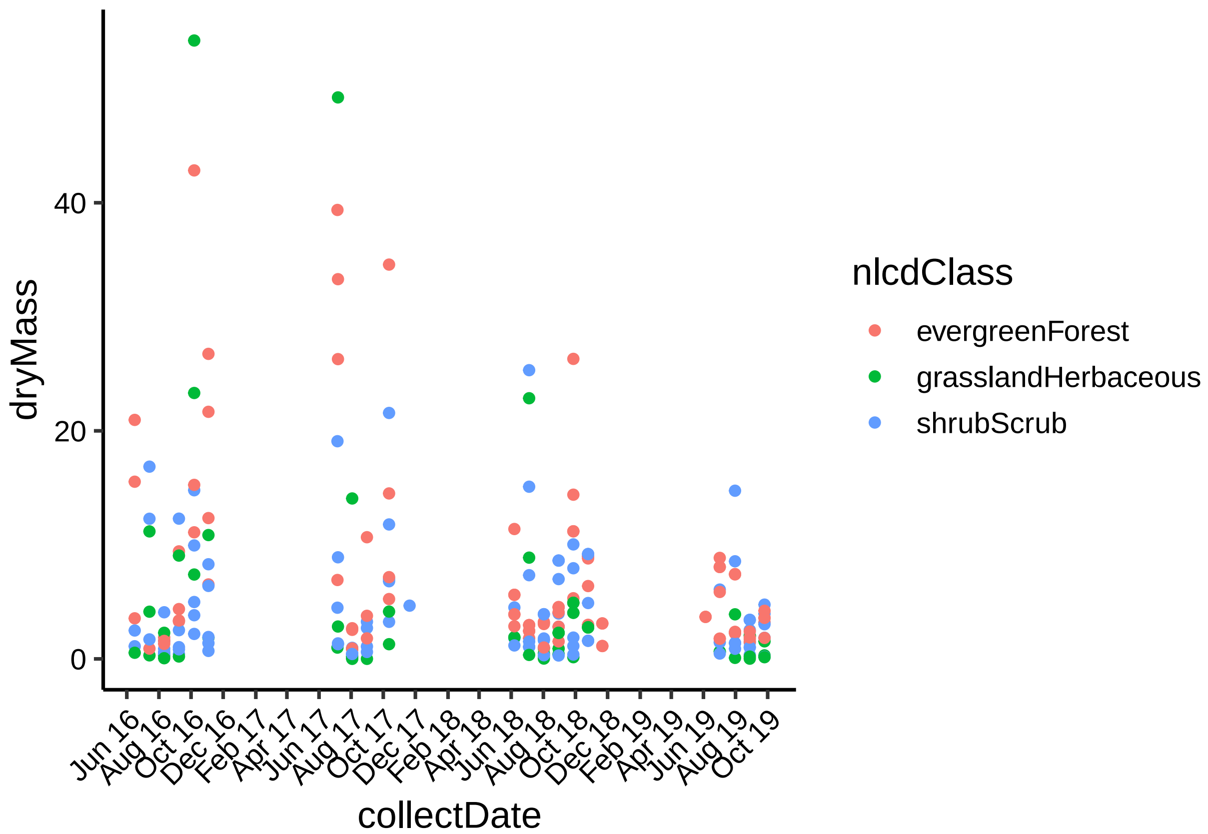


Question: What do you observe about the variables of interest over seasons and between lakes?

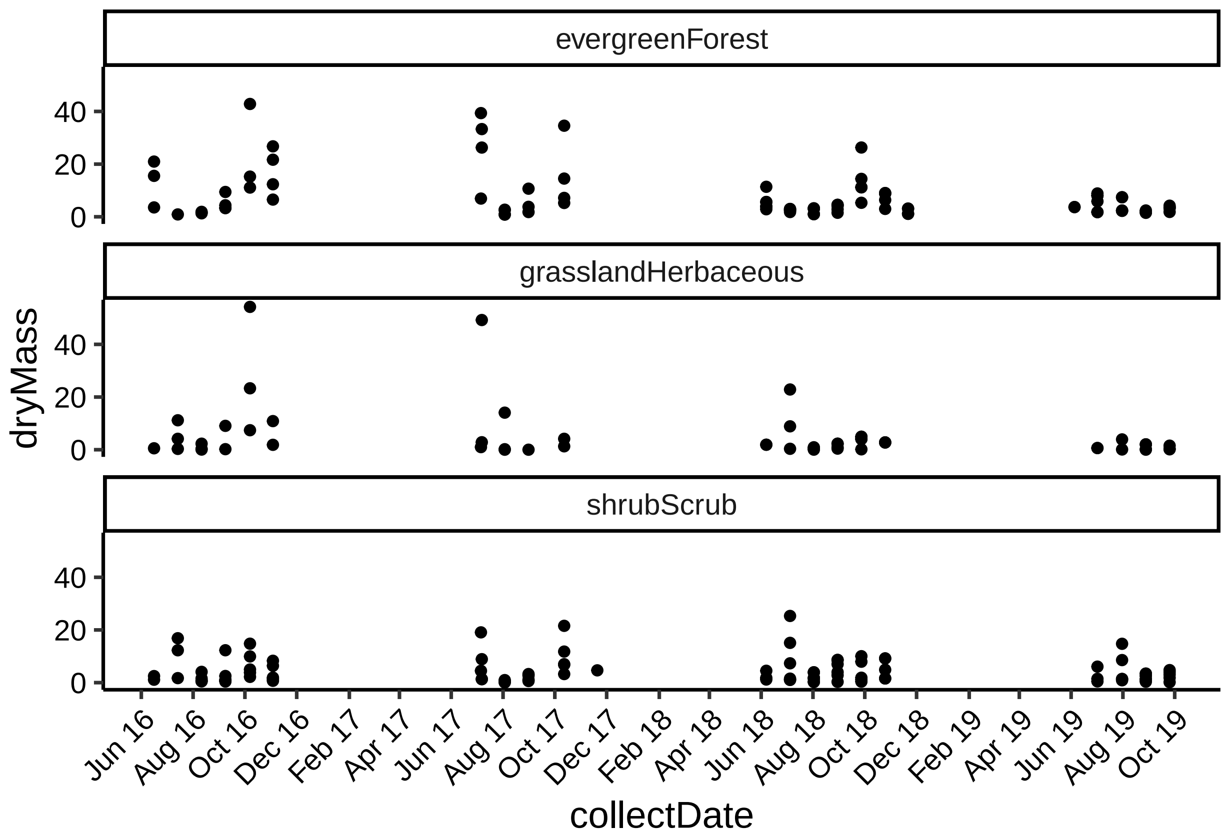
> Answer: Peter Lake seems to have slightly higher TP and TN values compared to Paul Lake, but only slightly. However, Peter Lake appears to have outliers for TP and TN values that are higher than those for Paul Lake. Both lakes display almost identical temperature fluctuations with the temperatures increasing as May progresses to September, with September showing the highest mean temperatures compared to the other months. Despite temperature fluctuations, the boxplots do not depict a significant increase in TP or TN concentrations congruent with temperature changes, but perhaps a trend would be more visible if we were to zoom in on the data and thereby eliminate some of the outliers.

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
Niwot.Litter.Needles <-
  ggplot(subset(Niwot.Ridge.Litter, functionalGroup == "Needles"),
         aes(x = collectDate, y = dryMass, color = nlcdClass)) +
  geom_point() +
  scale_x_date(date_breaks = "2 months", date_labels = "%b %y") +
  theme(axis.text.x = element_text(angle = 45,  hjust = 1))
print(Niwot.Litter.Needles)
```




```
#7
Niwot.Litter.Needles.Faceted <-
  ggplot(subset(Niwot.Ridge.Litter, functionalGroup == "Needles"),
         aes(x = collectDate, y = dryMass)) +
  geom_point() +
    scale_x_date(date_breaks = "2 months", date_labels = "%b %y") +
  theme(axis.text.x = element_text(angle = 45,  hjust = 1)) +
  facet_wrap(vars(nlcdClass), nrow = 3)
print(Niwot.Litter.Needles.Faceted)
```



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

> Answer: I think that the preferred plot depends on what you are trying to glean from the visualization. Plot 6 would be preferred if you are more focused on examining how all of the litter mass has changed over time with some insight into the spread of the different litter types compared to eachother, while Plot 7 (the faceted plot) is more effective for seeing values and spread for each individual litter type. Plot 6 makes it difficult to interpret any values for individual litter types whereas Plot 7, with the litter types separated out, is more effective for conveying individual litter data.