Assignment 5: Data Visualization

Leyi Ruan

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 Tuesday, February 23 at 11:59 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 (both the tidy [`NTL-LTER_Lake_Chemistry_Nutrients_PeterPaul_Processed.csv`] and the gathered [`NTL-LTER_Lake_Nutrients_PeterPaulGathered_Processed.csv`] versions) and the processed data file for the Niwot Ridge litter dataset.

```
getwd()
```

```
## [1] "/Users/ruanleyi/Documents/Year2/872 Environmental Data Analytics/Environmental_Data_Analytics_2
```

```
library(tidyverse)
```

```
## -- Attaching packages --------------------------------------- tidyverse 1.3.0 --
```

```
## v ggplot2 3.3.3     v purrr   0.3.4
## v tibble  3.0.5     v dplyr   1.0.3
## v tidyr   1.1.1     v stringr 1.4.0
## v readr   1.3.1     v forcats 0.5.0
```

```
## -- 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", stringsAsFacto
PeterPaul.chem.nutrients.gathered <-
  read.csv("../Data/Processed/NTL-LTER_Lake_Nutrients_PeterPaulGathered_Processed.csv", stringsAsFactors
Niwot_Ridge.sum <-
  read.csv("../Data/Processed/NEON_NIWO_Litter_mass_trap_Processed.csv", stringsAsFactors = TRUE)
```

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

```
#1
PeterPaul.chem.nutrients$sampledate <- as.Date(PeterPaul.chem.nutrients$sampledate, format = "%Y-%m-%d")

PeterPaul.chem.nutrients.gathered$sampledate <- as.Date(PeterPaul.chem.nutrients.gathered$sampledate, fo

Niwot_Ridge.sum$collectDate<-as.Date(Niwot_Ridge.sum$collectDate, format = "%Y-%m-%d")
```

Define your theme

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

```
mytheme <- theme_classic(base_size = 13) +
  theme(axis.text = element_text(color = "black"),
        legend.position = "top")
theme_set(mytheme)
```

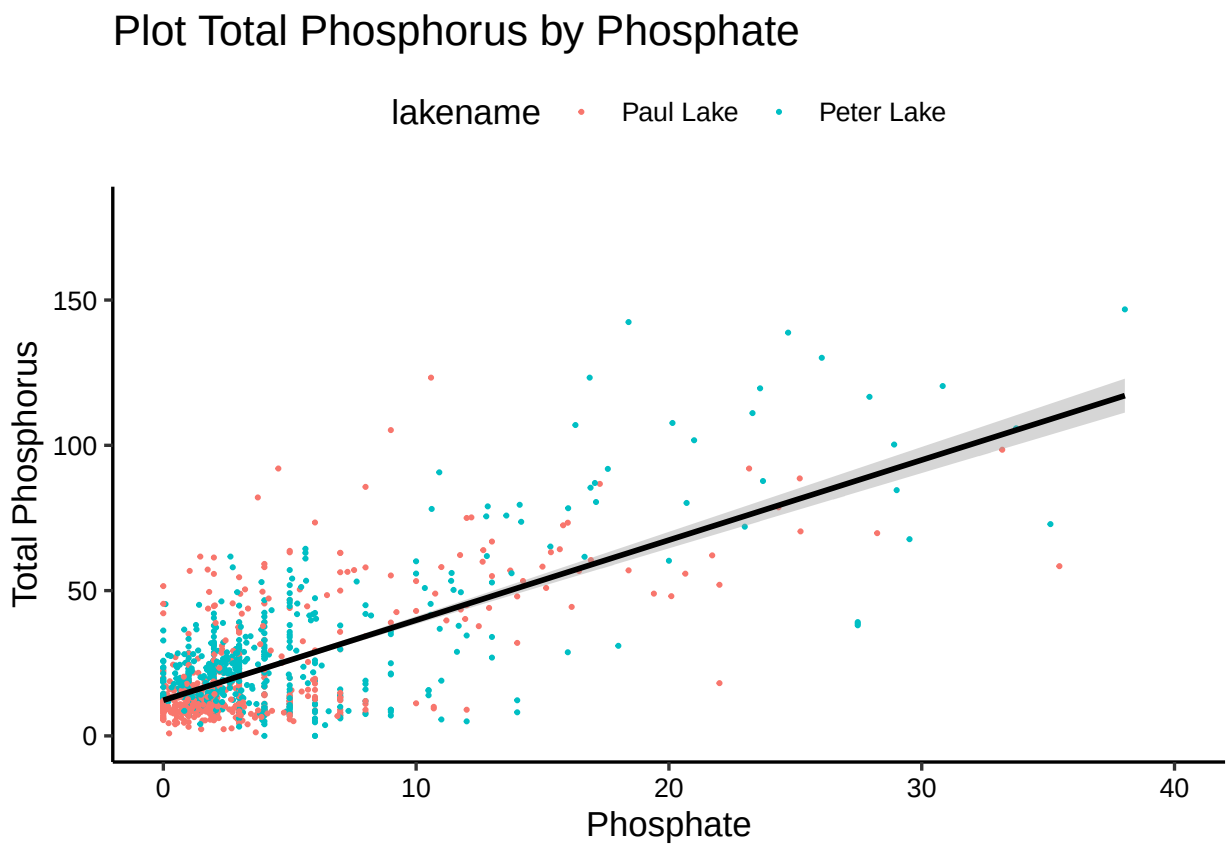
Create graphs

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.

```
plot4 <- ggplot(PeterPaul.chem.nutrients, aes(x = po4, y = tp_ug, color=lakename)) +
  geom_point(size=0.4) +
  geom_smooth(method = lm, color="black") +
  labs(y="Total Phosphorus", x="Phosphate", title="Plot Total Phosphorus by Phosphate") +
  xlim(0, 40) +
  ylim(0, 180)
print(plot4)
```

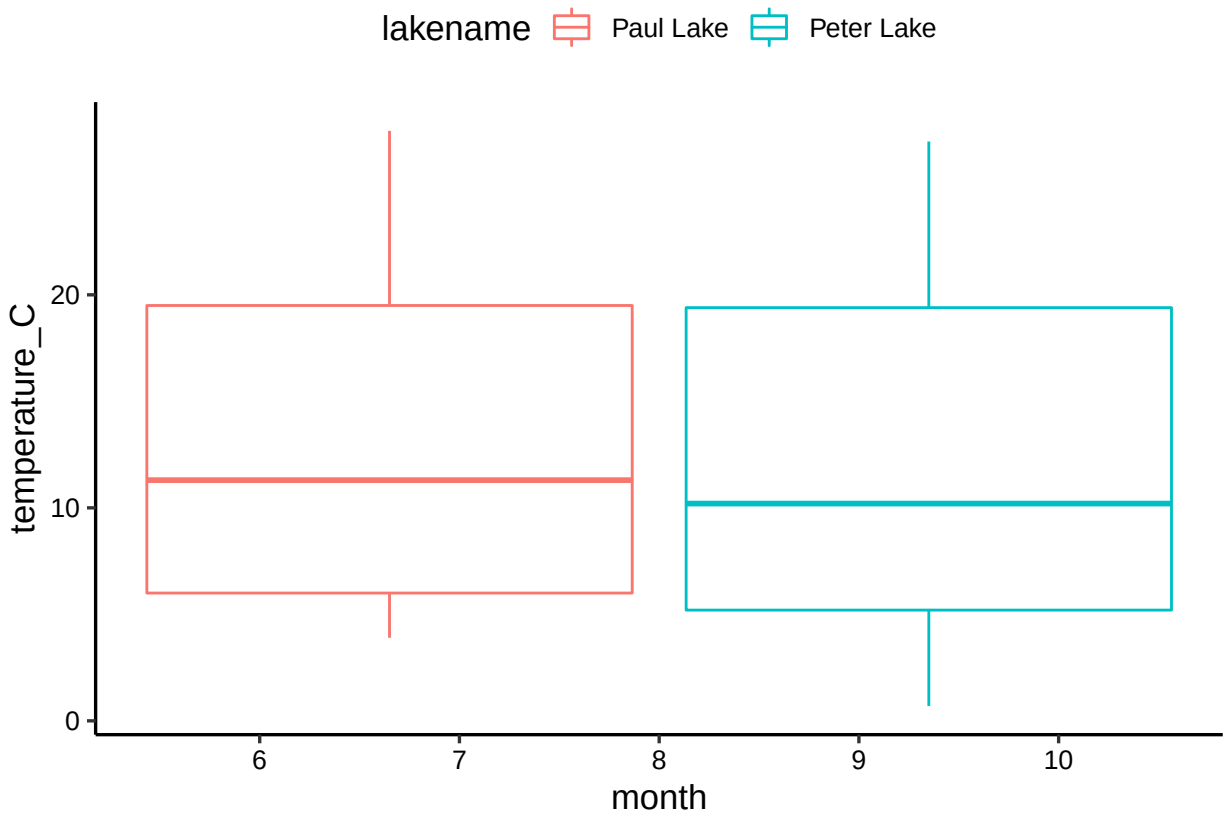
```
## `geom_smooth()` using formula 'y ~ x'

## Warning: Removed 21949 rows containing non-finite values (stat_smooth).

## Warning: Removed 21949 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.

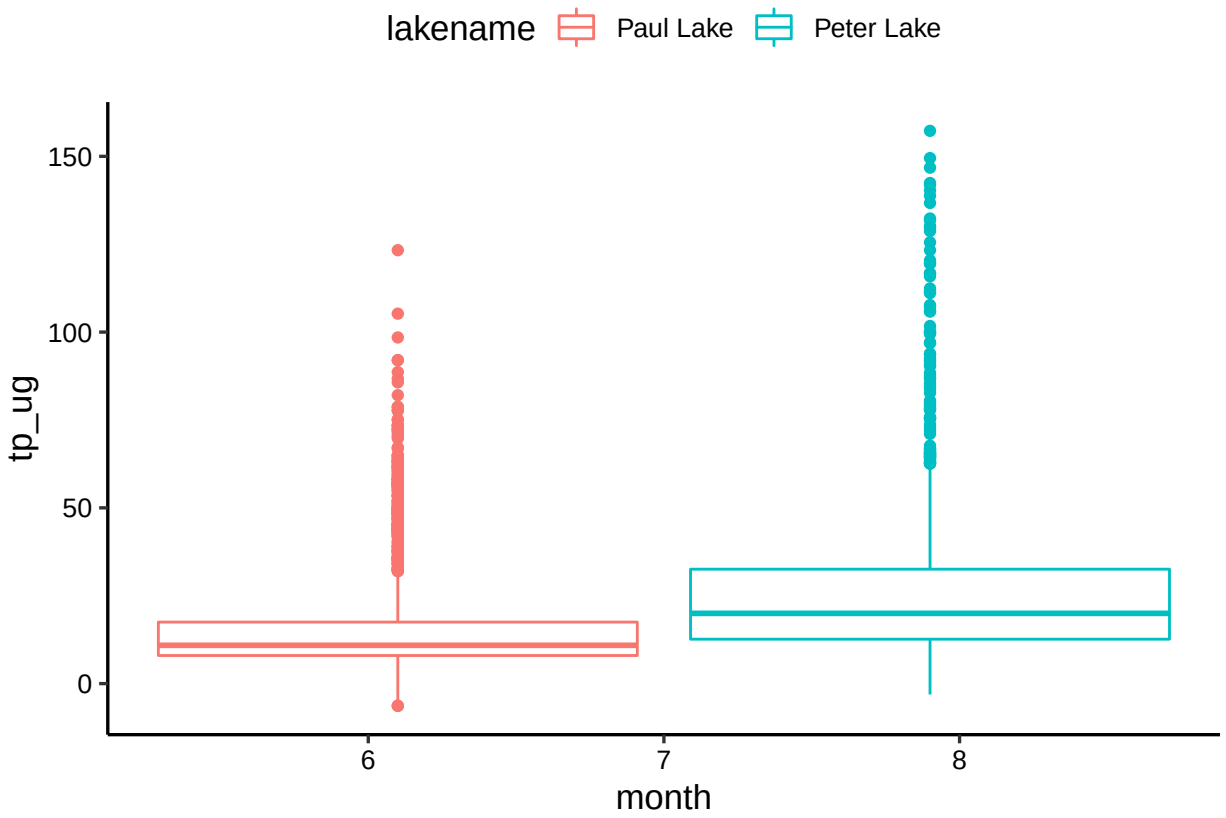
```
# Box and whiskers plot
plot5.1 <-
  ggplot(PeterPaul.chem.nutrients, aes(x = month, y = temperature_C, color=lakename)) +
  geom_boxplot()
print(plot5.1)
```

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



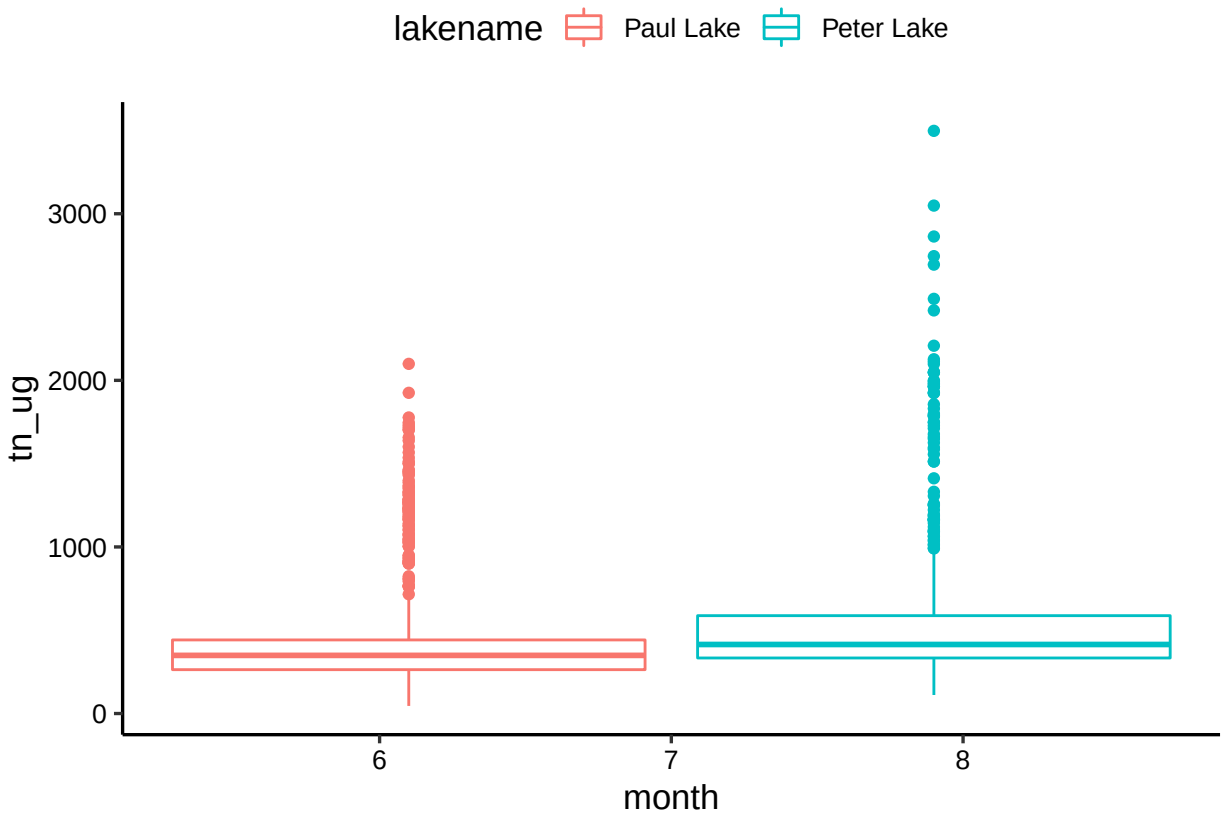

```
plot5.2 <-
  ggplot(PeterPaul.chem.nutrients, aes(x = month, y = tp_ug, color=lakename)) +
  geom_boxplot()
print(plot5.2)
```

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



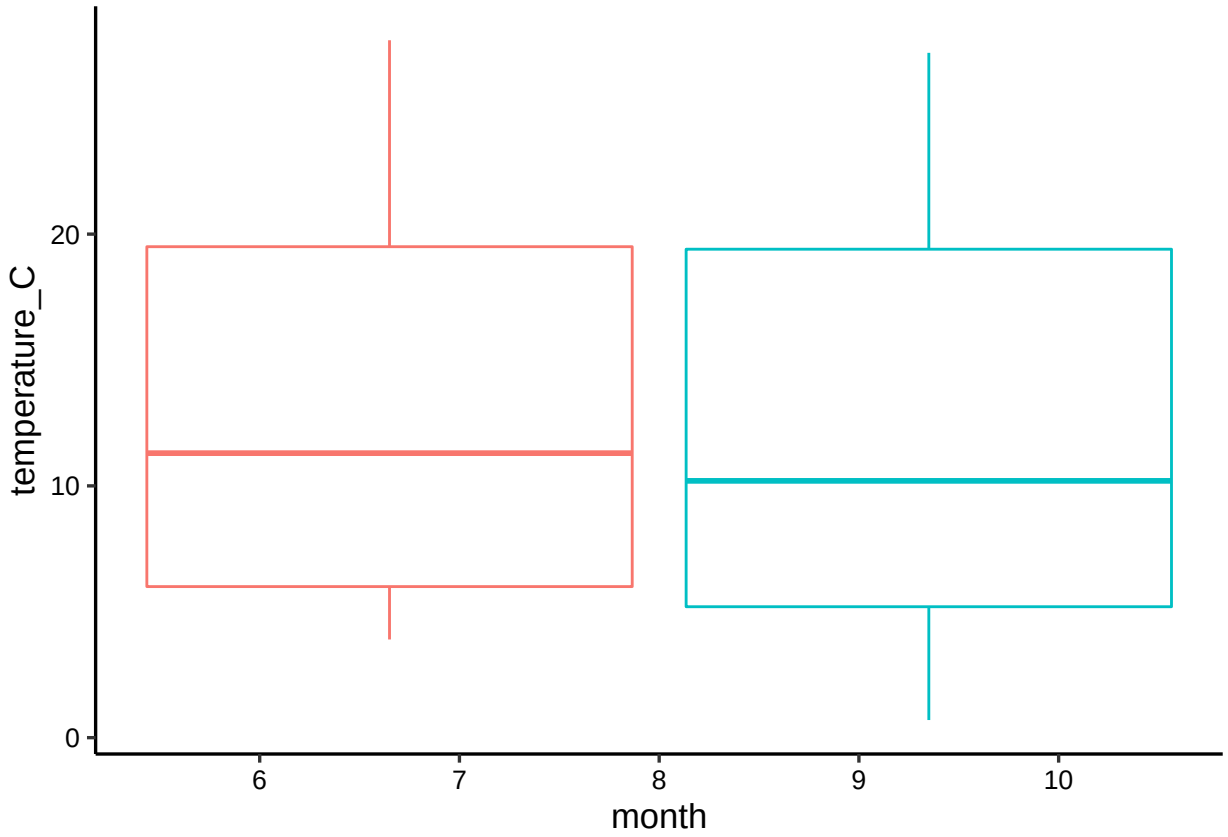

```
plot5.3 <-
  ggplot(PeterPaul.chem.nutrients, aes(x = month, y = tn_ug, color=lakename)) +
  geom_boxplot()
print(plot5.3)
```

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



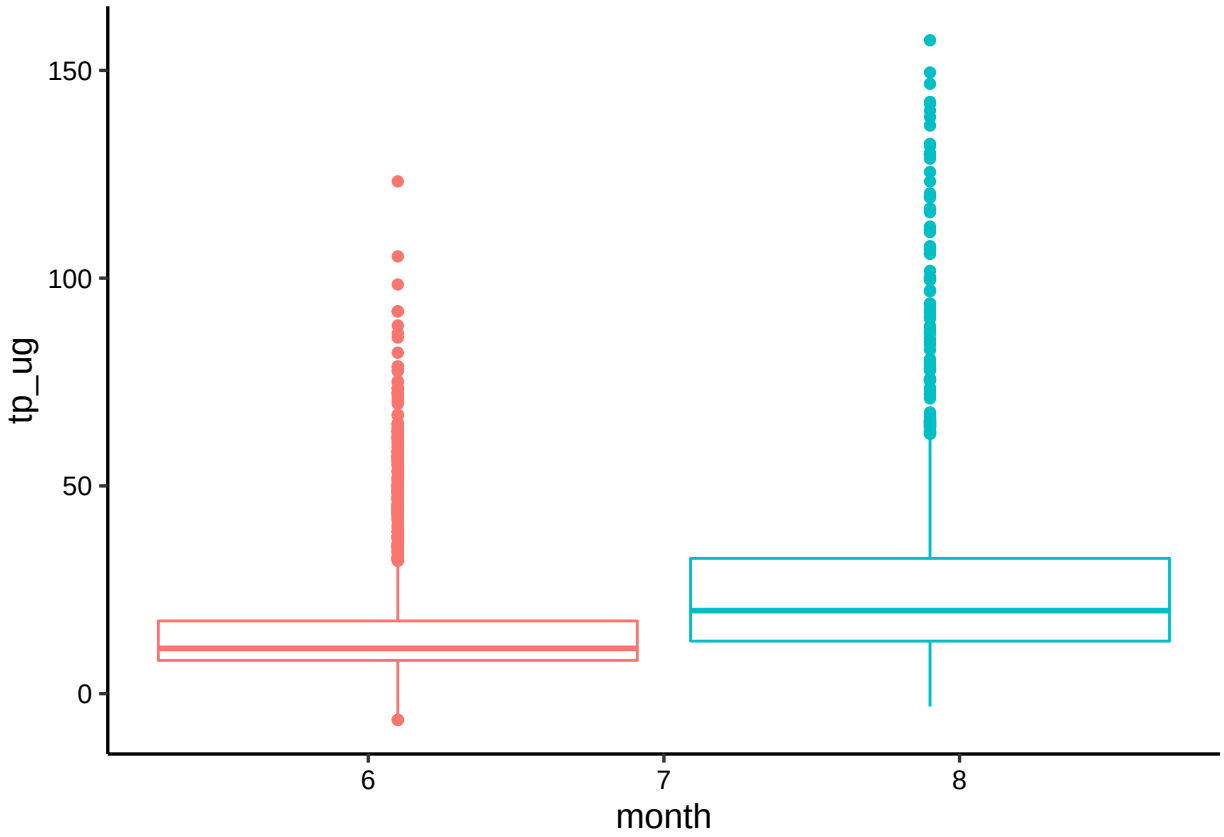

```
# Make two plots legend disappear
plot5.1.2 <-
  ggplot(PeterPaul.chem.nutrients, aes(x = month, y = temperature_C, color=lakename)) +
  geom_boxplot()+
  theme(legend.position="none")
print(plot5.1.2)
```

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




```
plot5.2.2 <-
  ggplot(PeterPaul.chem.nutrients, aes(x = month, y = tp_ug, color=lakename)) +
  geom_boxplot() +
  theme(legend.position="none")
print(plot5.2.2)
```

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

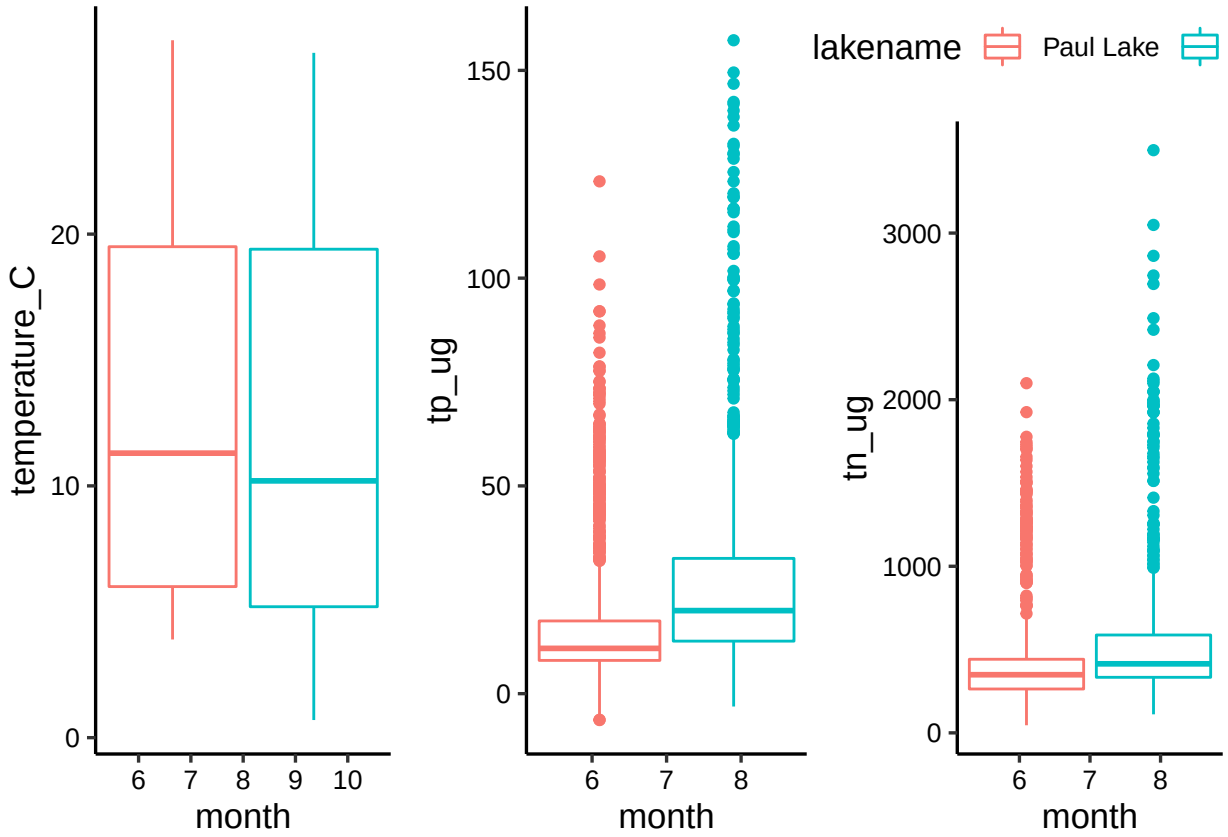



```
library(cowplot)
plot_grid(plot5.1.2, plot5.2.2, plot5.3, nrow = 1, align = 'h', axis = 'lr')
```

```
## 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).
```

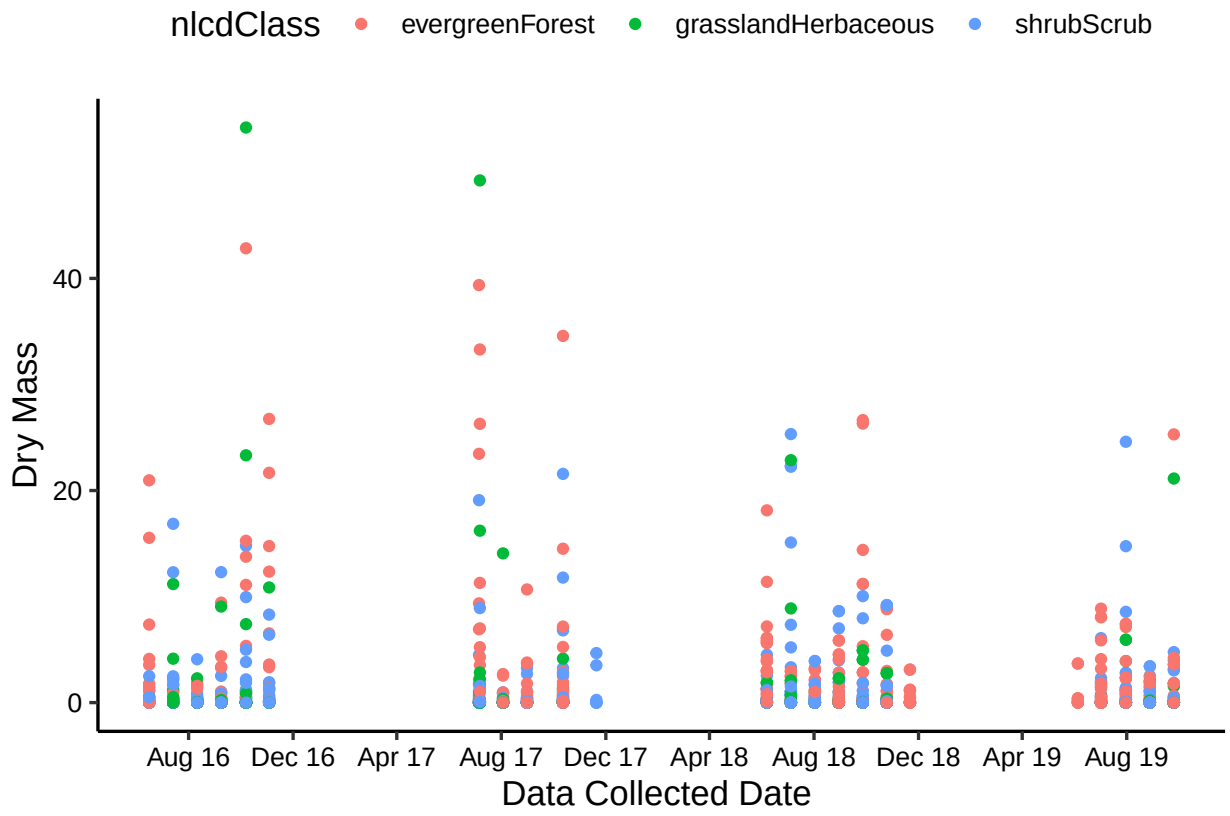


Question: What do you observe about the variables of interest over seasons and between lakes?

Answer: from the means and distribution of obversations, Peter lake seems to have a relatively lower temperature, higher phosphorus, and higher phosphate than Paul Lake over season.

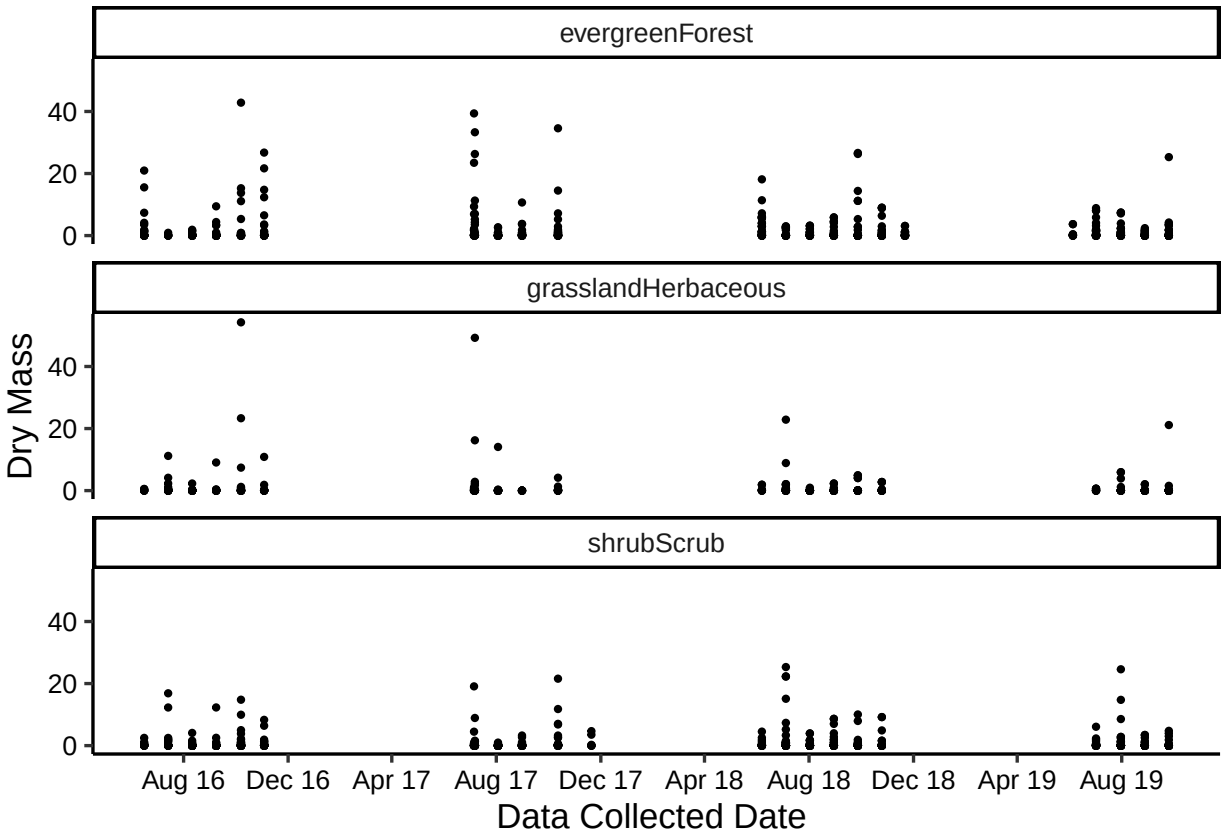
6. [Niwot Ridge] Plot a subset of the litter dataset by displaying only the "Needles" functional group. Plot the dry mass of needle litter by date and separate by NLCD class with a color aesthetic. (no need to adjust the name of each land use)

```
plot6<-
  ggplot(subset(Niwot_Ridge.sum, functionalGroup="Needles"), aes(x=collectDate, y=dryMass, color=nlcdCla
  geom_point()+
  scale_x_date(date_breaks = "4 months", date_labels = "%b %y")+
  labs(y="Dry Mass", x="Data Collected Date ") +
  mytheme
print(plot6)
```



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

```
plot7<-
  ggplot(subset(Niwot_Ridge.sum, functionalGroup="Needles"), aes(x=collectDate, y=dryMass))+
  geom_point(size=0.8)+
  facet_wrap(vars(nlcdClass), nrow=3)+
  scale_x_date(date_breaks = "4 months", date_labels = "%b %y")+
  labs(y="Dry Mass", x="Data Collected Date ")
print(plot7)
```



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

Answer: I think plot 7 is more effective since it compares the amount of dry mass among three land use groups, thus it display more information than plot 6.