Assignment 5: Data Visualization

Xuening Tang

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_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/xueningtang/Desktop/R lab/EDA-Fall2022/Assignments"
```

```
library(tidyverse)
```

```
## Warning: package 'tidyverse' was built under R version 4.1.2

## -- Attaching packages --------------------------------------- tidyverse 1.3.2 --v ggplot2 3.3.5
## v tibble  3.1.8      v dplyr   1.0.10
## v tidyr   1.2.1      v stringr 1.4.0
## v readr   2.1.2      v forcats 0.5.1

## Warning: package 'tibble' was built under R version 4.1.2
## Warning: package 'tidyr' was built under R version 4.1.2

## Warning: package 'readr' was built under R version 4.1.2

## Warning: package 'dplyr' was built under R version 4.1.2

## -- 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(cowplot)
```

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

```
PeterPaul.chem <-
  read.csv(
    "../Data/Processed/NTL-LTER_Lake_Chemistry_Nutrients_PeterPaul_Processed.csv",
    stringsAsFactors = TRUE)
Niwot.litter <-
  read.csv(
    "../Data/Processed/NEON_NIWO_Litter_mass_trap_Processed.csv",
    stringsAsFactors = TRUE)

#2
PeterPaul.chem$sampledate <- as.Date(
  PeterPaul.chem$sampledate, format = "%Y-%m-%d")
Niwot.litter$collectDate <- as.Date(
  Niwot.litter$collectDate,format = "%Y-%m-%d")
```

Define your theme

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

```
#3
my.theme <- theme_classic(base_size = 14) +
  theme(axis.text = element_text(color = "black"),
        legend.position = "top")
theme_set(my.theme)
```

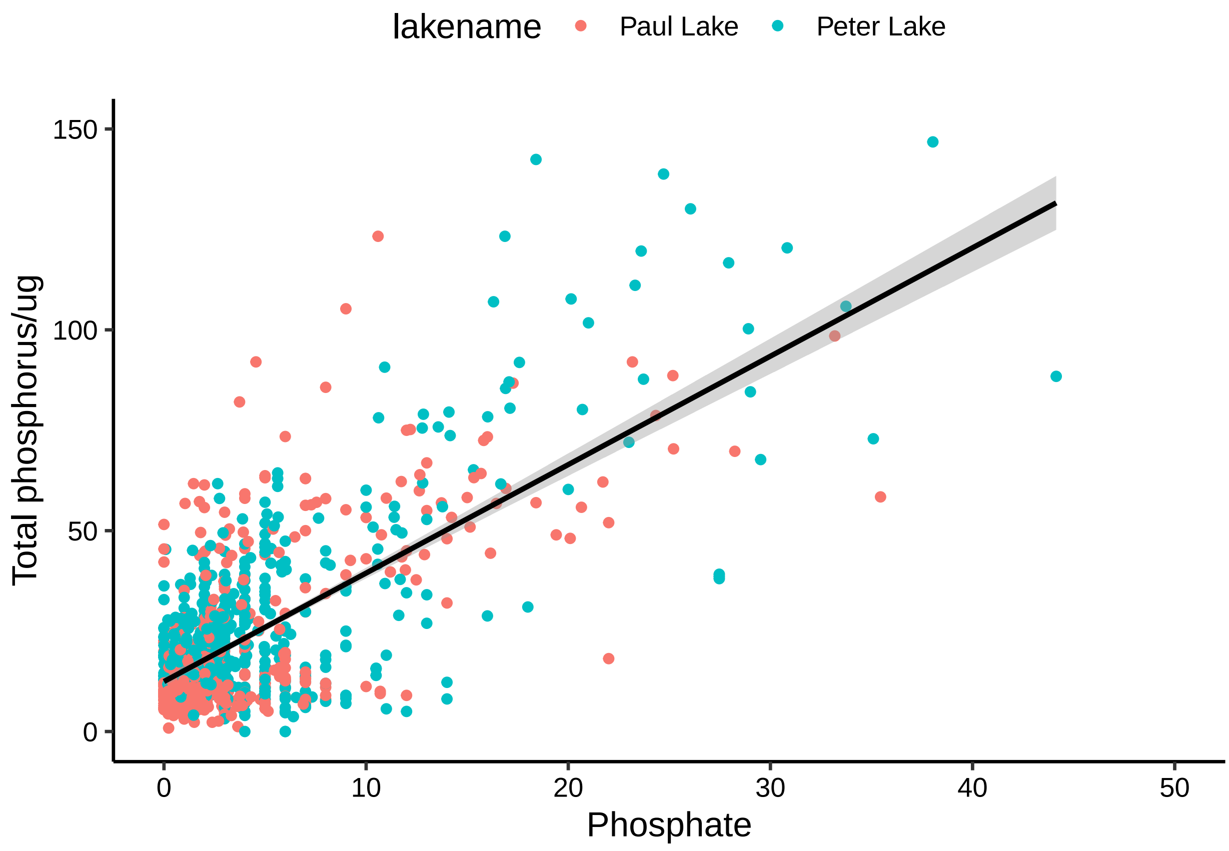
Create graphs

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()`).

```
#4
Plot1 <-
  ggplot(PeterPaul.chem, aes(y=tp_ug, x=po4))+
  geom_point(aes(color = lakename))+
  ylab("Total phosphorus/ug")+
  xlab("Phosphate")+
  ylim(0,150)+
  xlim(0,50)+
  geom_smooth(method = lm,color="black")
print(Plot1)
```

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

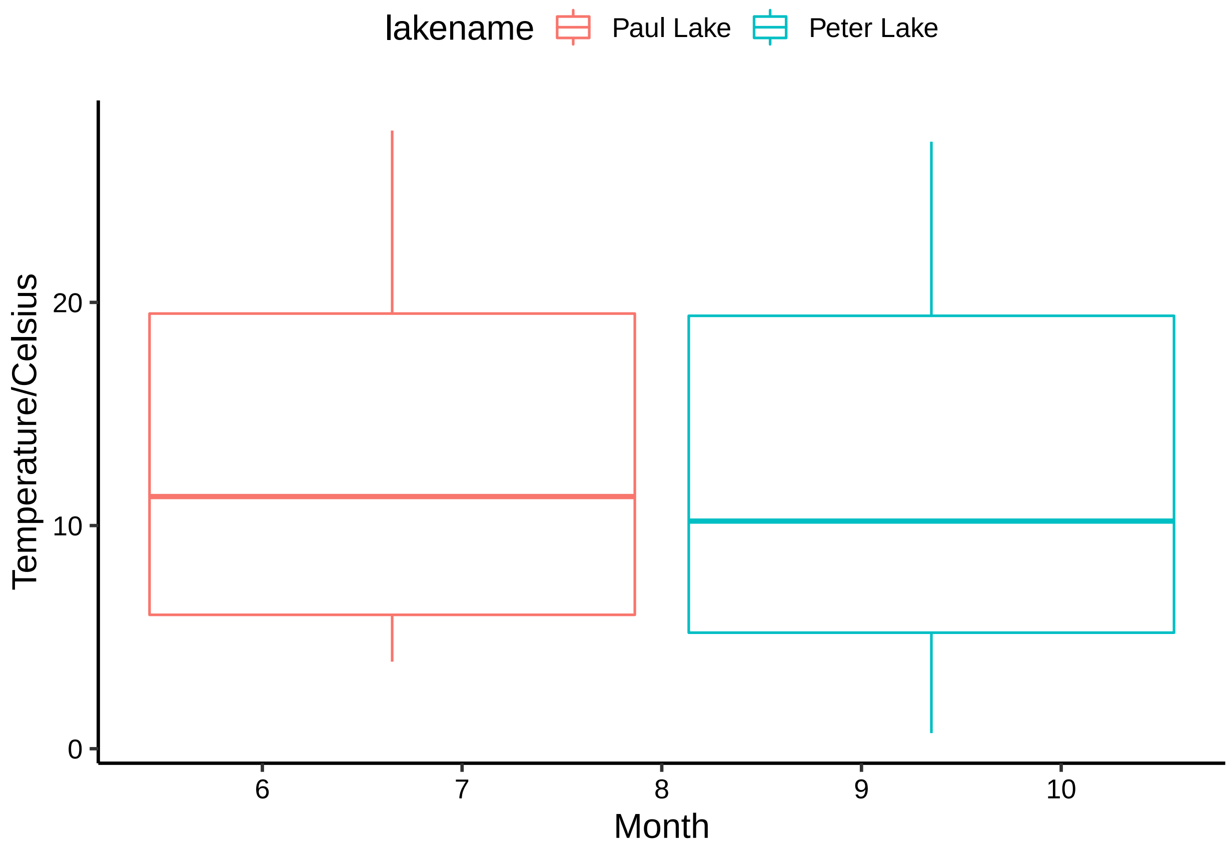


5. [NTL-LTER] Make three separate boxplots of (a) temperature, (b) TP, and

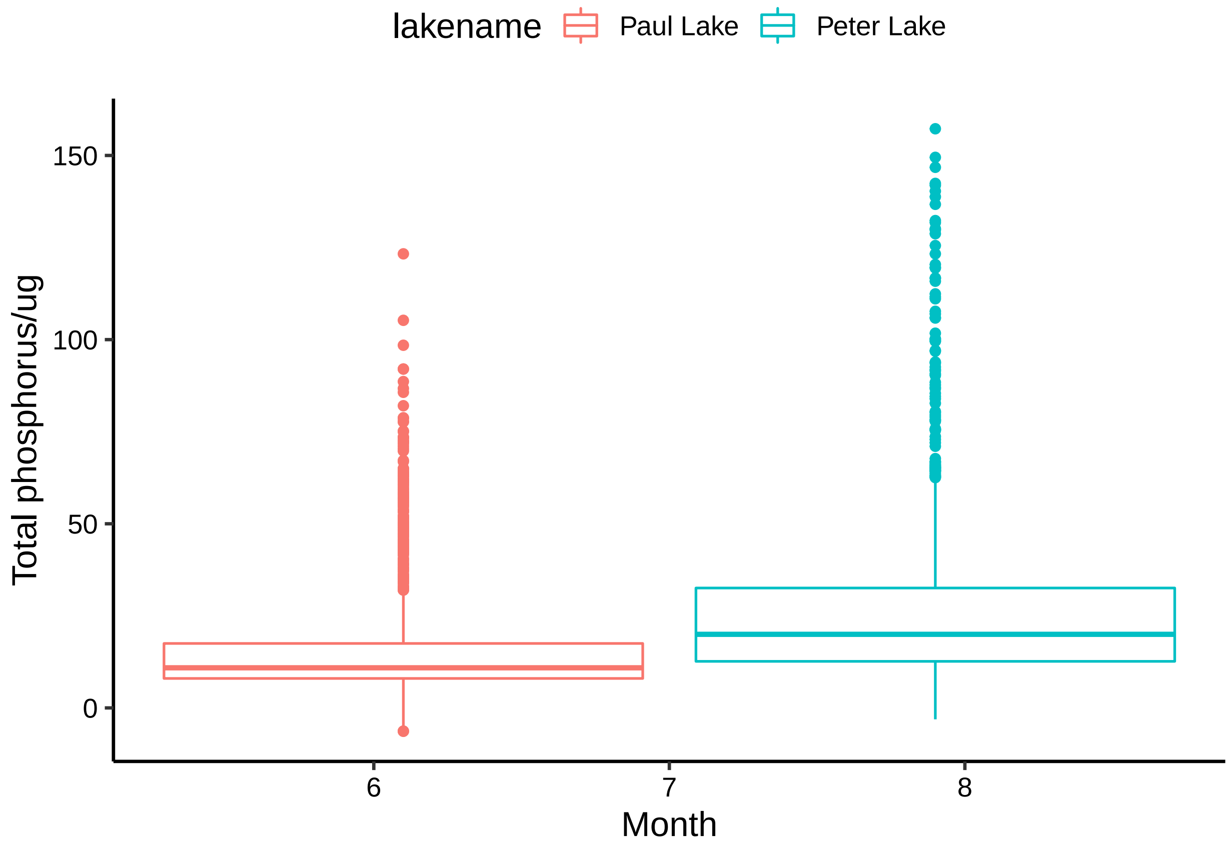
(c) TN, with month as the x axis and lake as a color aesthetic. Then, create a cowplot that combines the three graphs. Make sure that only one legend is present and that graph axes are aligned.

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

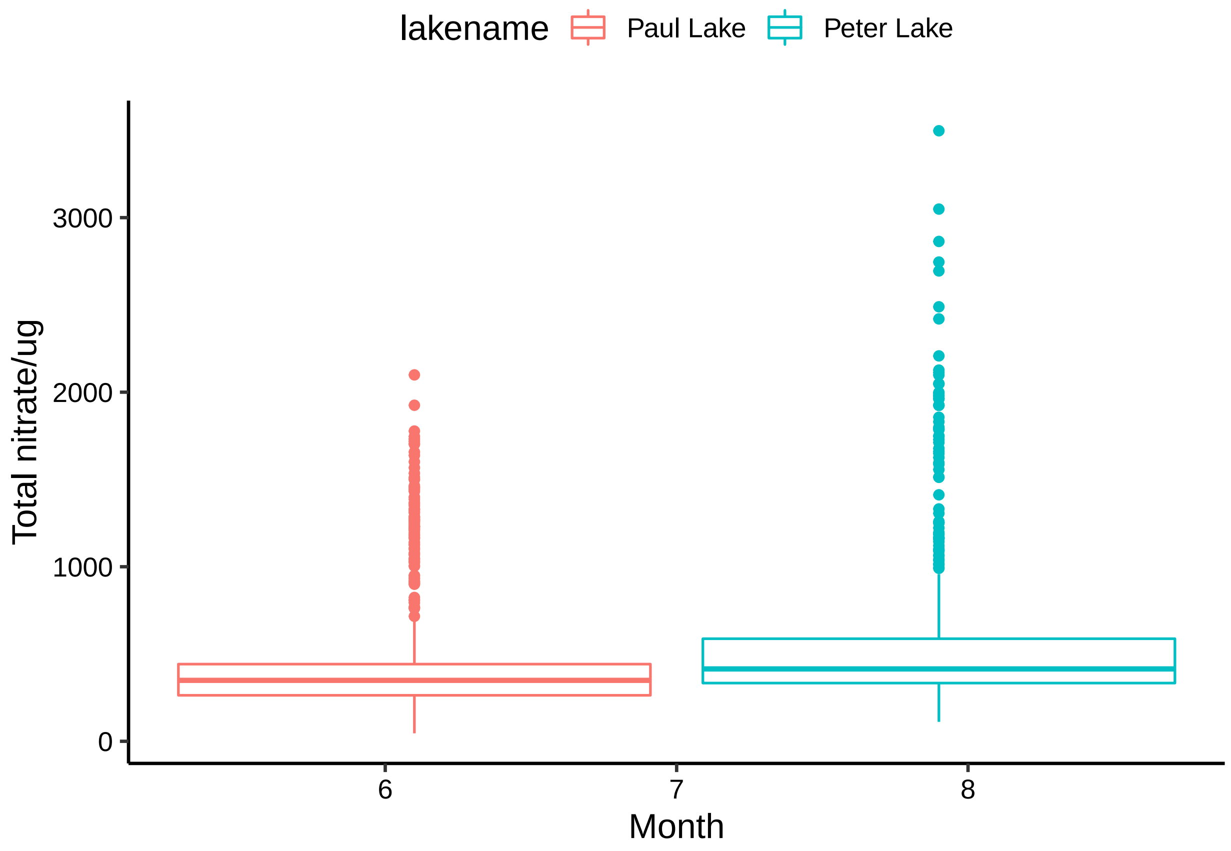
```
#5
Box1 <-
  ggplot(PeterPaul.chem, aes(x = month, y = temperature_C)) +
  geom_boxplot(aes(color = lakename))+
  ylab("Temperature/Celsius")+
  xlab("Month")
print(Box1)
```



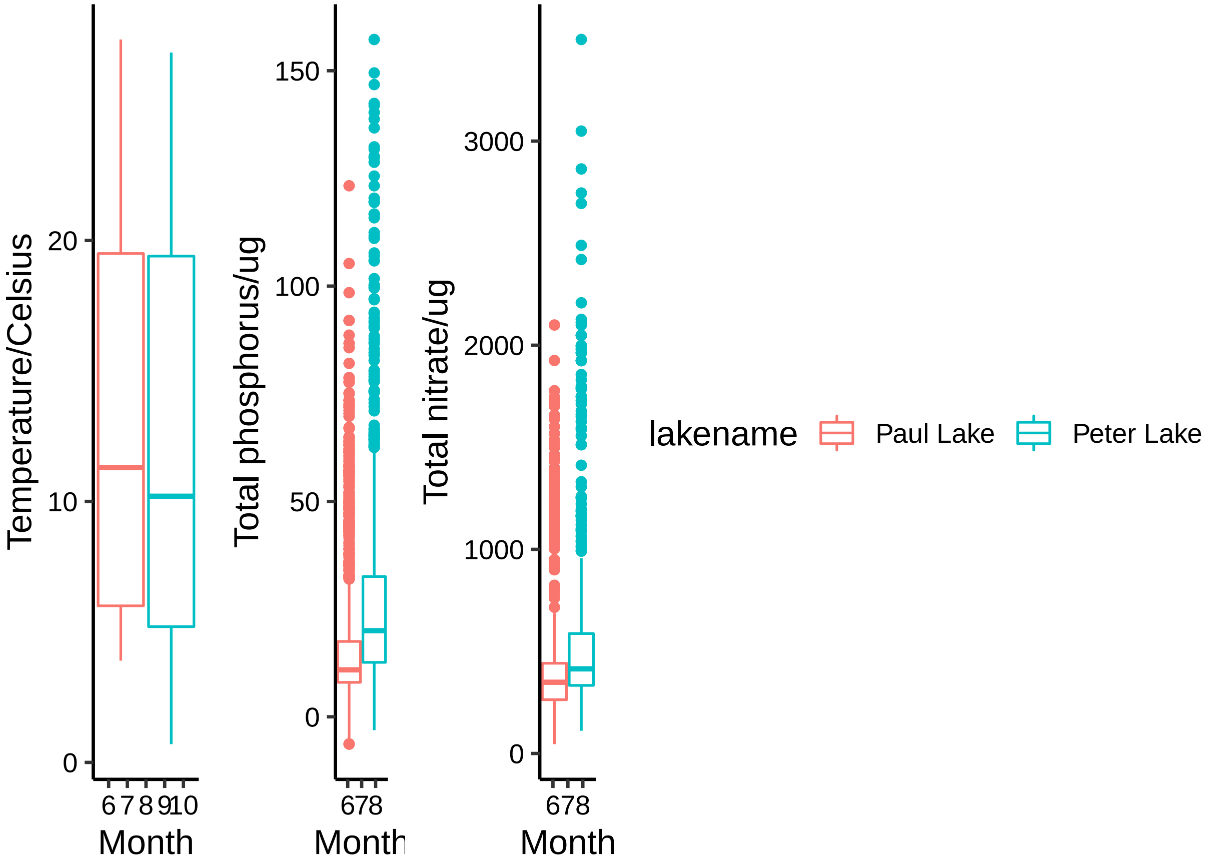
```
Box2 <-
  ggplot(PeterPaul.chem, aes(x = month, y = tp_ug)) +
  geom_boxplot(aes(color = lakename))+
  ylab("Total phosphorus/ug")+
  xlab("Month")
print(Box2)
```



```
Box3 <-
  ggplot(PeterPaul.chem, aes(x = month, y = tn_ug)) +
  geom_boxplot(aes(color = lakename))+
  ylab("Total nitrate/ug")+
  xlab("Month")
print(Box3)
```



```
Box1a <- Box1 + theme(legend.position='none')
Box2b <- Box2 + theme(legend.position='none')
Box3c <- Box3 + theme(legend.position='none')
Gridplot <- plot_grid(Box1a, Box2b, Box3c, nrow = 1, align = 'h', rel_widths = c(1.2,1,1.1))
legend <- get_legend(Box1)
plot_grid(Gridplot, legend)
```

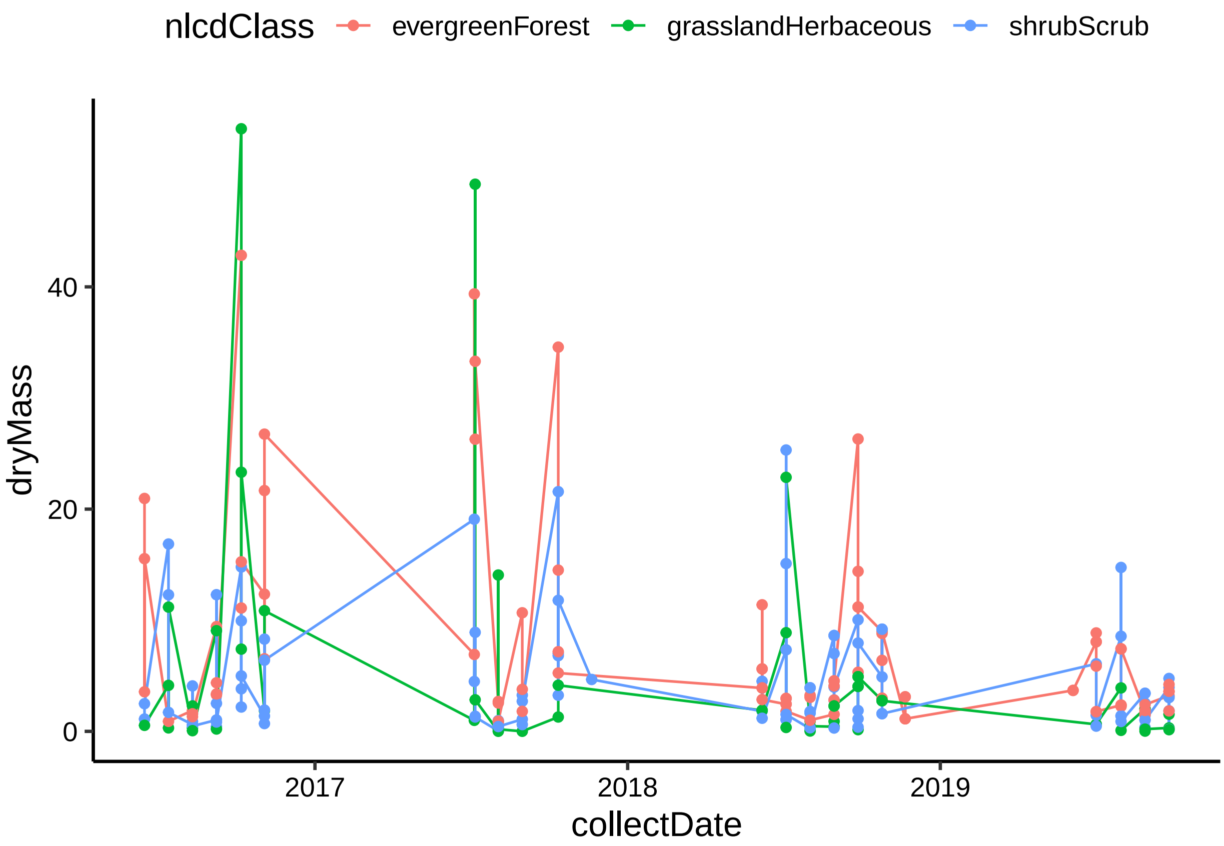


Question: What do you observe about the variables of interest over seasons and between lakes?

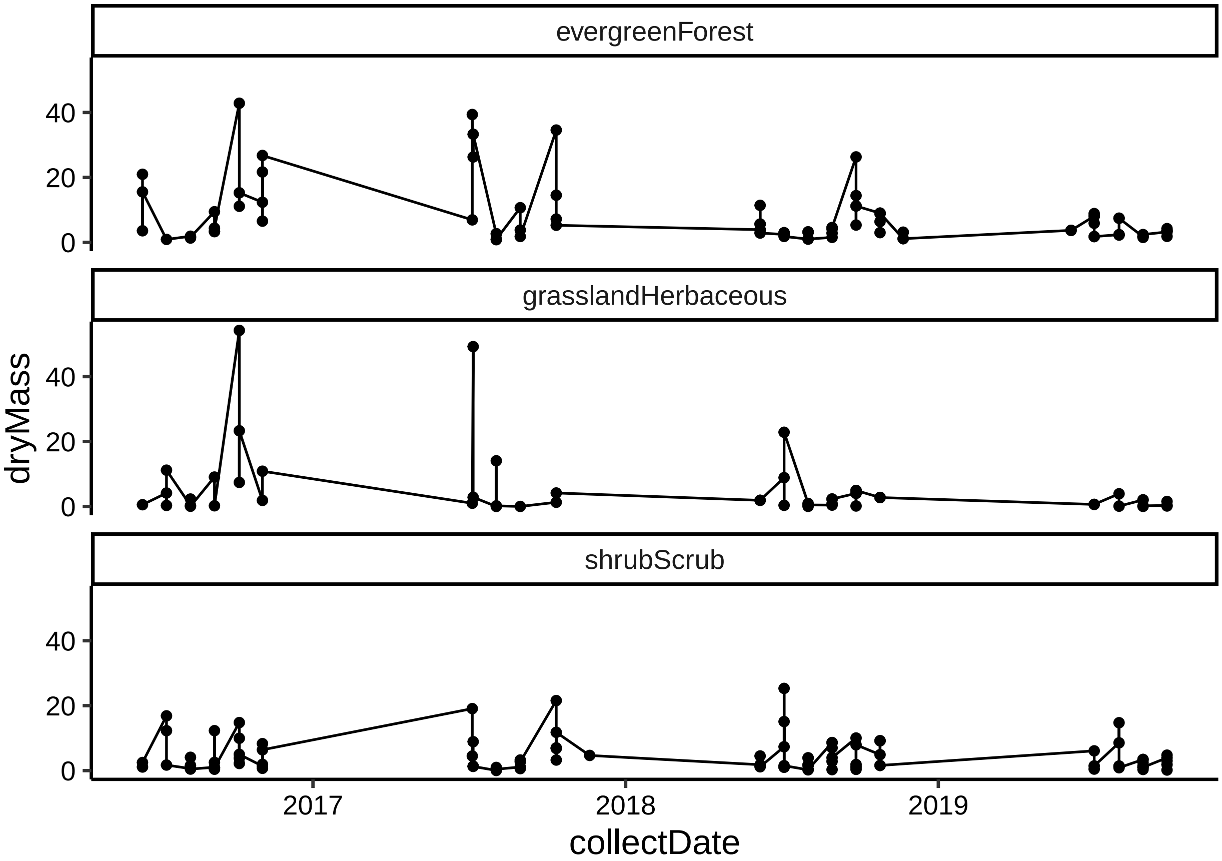
> Answer: In these three boxplots, I observed the seasonal change of temperature, phosphorus concentration and nitrate concentration, it is obvious to see the overall temperature in fall is much lower than that in summer as the the chemical levels going up. The overall chemicals in Peter lake is higher than Paul lake according to the mean value and data range. The average temperature in Peter lake is lower than that in Paul 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
Plot2 <-
  ggplot(subset(Niwot.litter, functionalGroup == "Needles"),
         aes(x = collectDate, y = dryMass)) +
  geom_line(aes(color = nlcdClass))+
  geom_point(aes(color = nlcdClass))
print(Plot2)
```



```
#7
Plot3 <-
  ggplot(subset(Niwot.litter, functionalGroup == "Needles"),
         aes(x = collectDate, y = dryMass)) +
  geom_line()+
  geom_point()+
  facet_wrap(vars(nlcdClass), nrow = 3)
print(Plot3)
```



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

> Answer: I think plot in 7 is more effective because we are more likely willing to observe the trend of dry mass change during the time in one class rather than comparison of dry mass among different classess. So the plot in 6 actully can not express data well and hard to distinguish the time change trend.