Assignment 5: Data Visualization

Suad Muradov

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 getting wd, load tidyverse and cowplot

getwd()
```

```
## [1] "/Users/smuradov2021/Desktop/Desktop/MY DOCUMENTS/iMEP/Semester 4/ENV872/Environmental_Data_Analy
```

```
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   2.1.2     v forcats 0.5.1

## -- Conflicts ------------------------------------------ tidyverse_conflicts() --
## x dplyr::filter() masks stats::filter()
## x dplyr::lag()    masks stats::lag()
```

```
#install.packages("cowplot")
library(cowplot)
#uploading the processed files
```

```
PetPaul.processed<-read.csv("./Data/Processed/NTL-LTER_Lake_Chemistry_Nutrients_PeterPaul_Processed.csv
Litter<-read.csv("./Data/Processed/NEON_NIWO_Litter_mass_trap_Processed.csv",stringsAsFactors = TRUE)

#2 convert to date
PetPaul.processed$sampledate<-as.Date(PetPaul.processed$sampledate,"%Y-%m-%d")
Litter$collectDate<-as.Date(Litter$collectDate,"%Y-%m-%d")
```

Define your theme

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

```
#3 creating own theme
own.theme<-theme_classic(base_size = 18) +
  theme(axis.text = element_text(color = "black"),
        legend.position = "right", legend.title = element_text(size = 14))
```

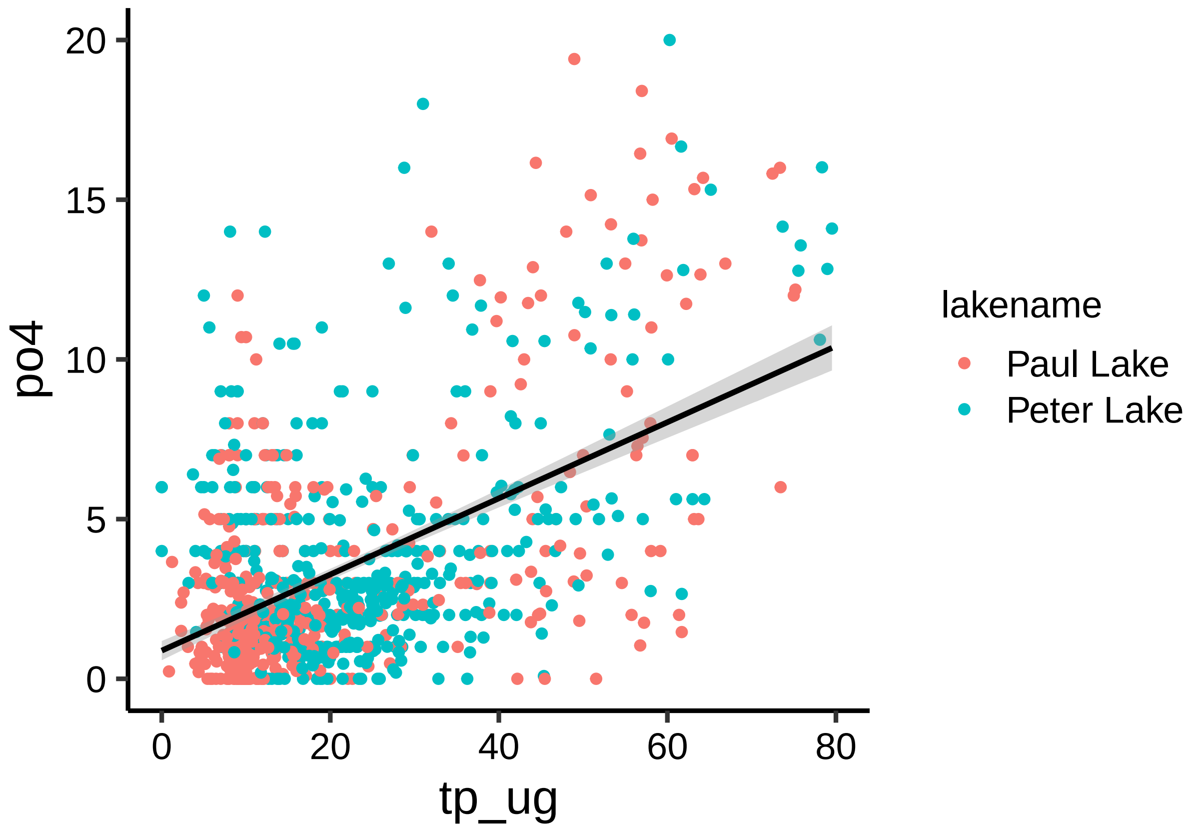
Create graphs

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()`).

```
#4 load ggplot2 and then plot PetPaul
library(ggplot2)
phosphate.vs.phosphorus<-
  ggplot(PetPaul.processed, aes(x = tp_ug, y = po4))+
  geom_point(aes(color=lakename)) +
  geom_smooth(method = lm,col='black')+
  xlim (0,80)+
  ylim(0,20)+
  own.theme
print(phosphate.vs.phosphorus)
```

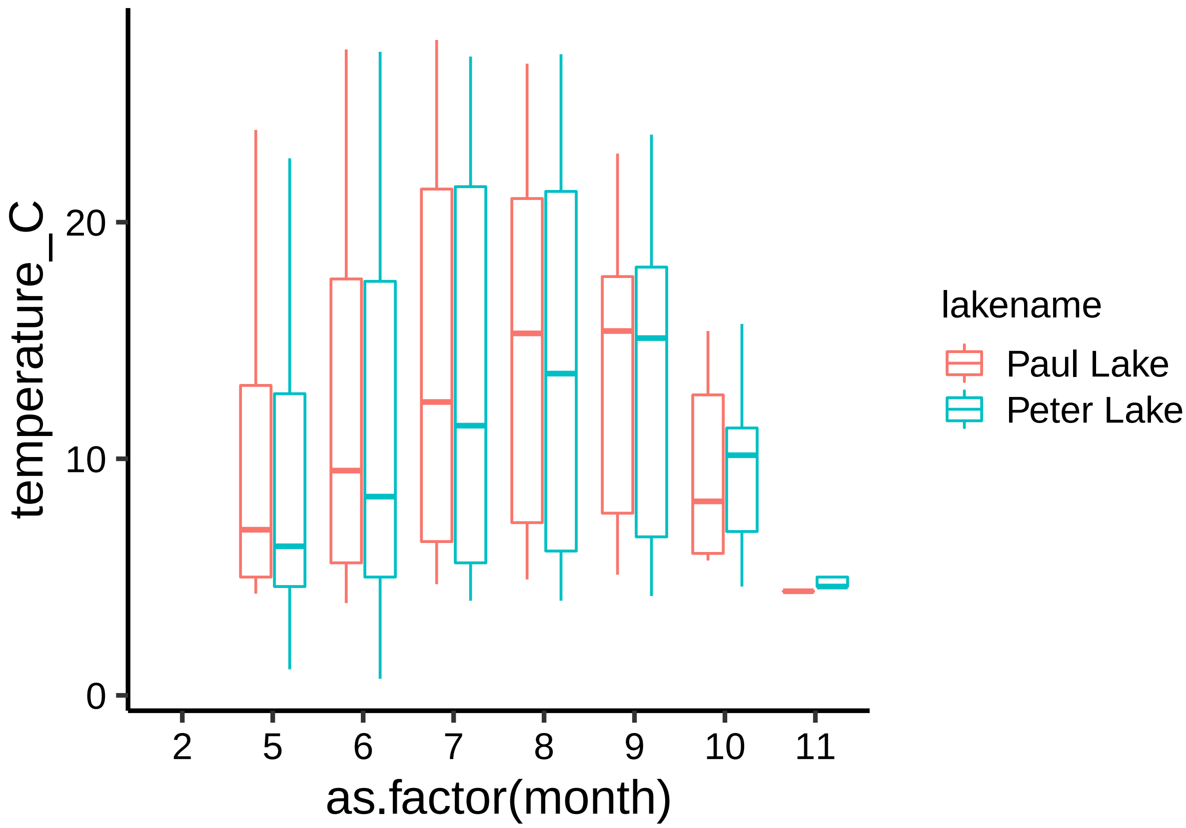
```
## `geom_smooth()` using formula 'y ~ x'

## Warning: Removed 21995 rows containing non-finite values (stat_smooth).

## Warning: Removed 21995 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.

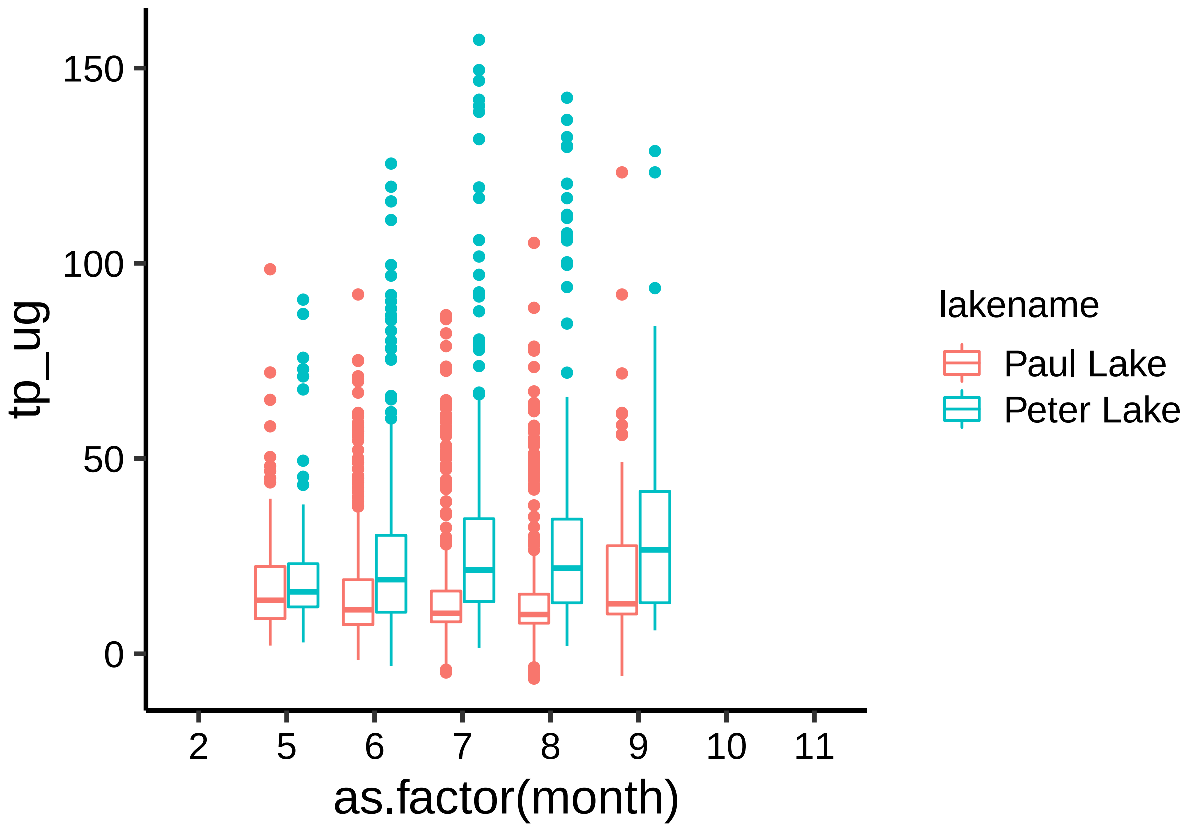
```
#5
boxTemp<-
  ggplot(PetPaul.processed,aes(x=as.factor(month),y=temperature_C))+
  geom_boxplot(aes(color=lakename))+
  own.theme
print(boxTemp)
```

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



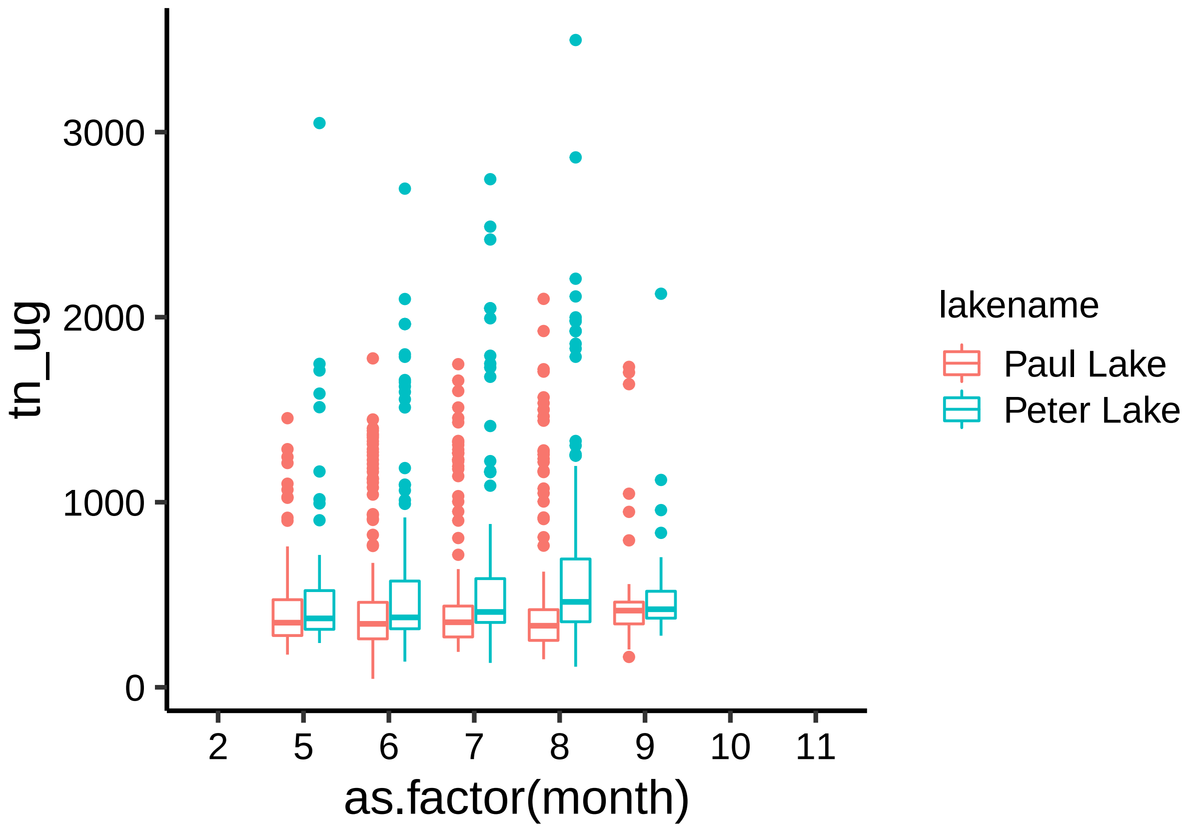

```
boxTP<-
  ggplot(PetPaul.processed,aes(x=as.factor(month),y=tp_ug))+
  geom_boxplot(aes(color=lakename))+
  own.theme
print(boxTP)
```

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



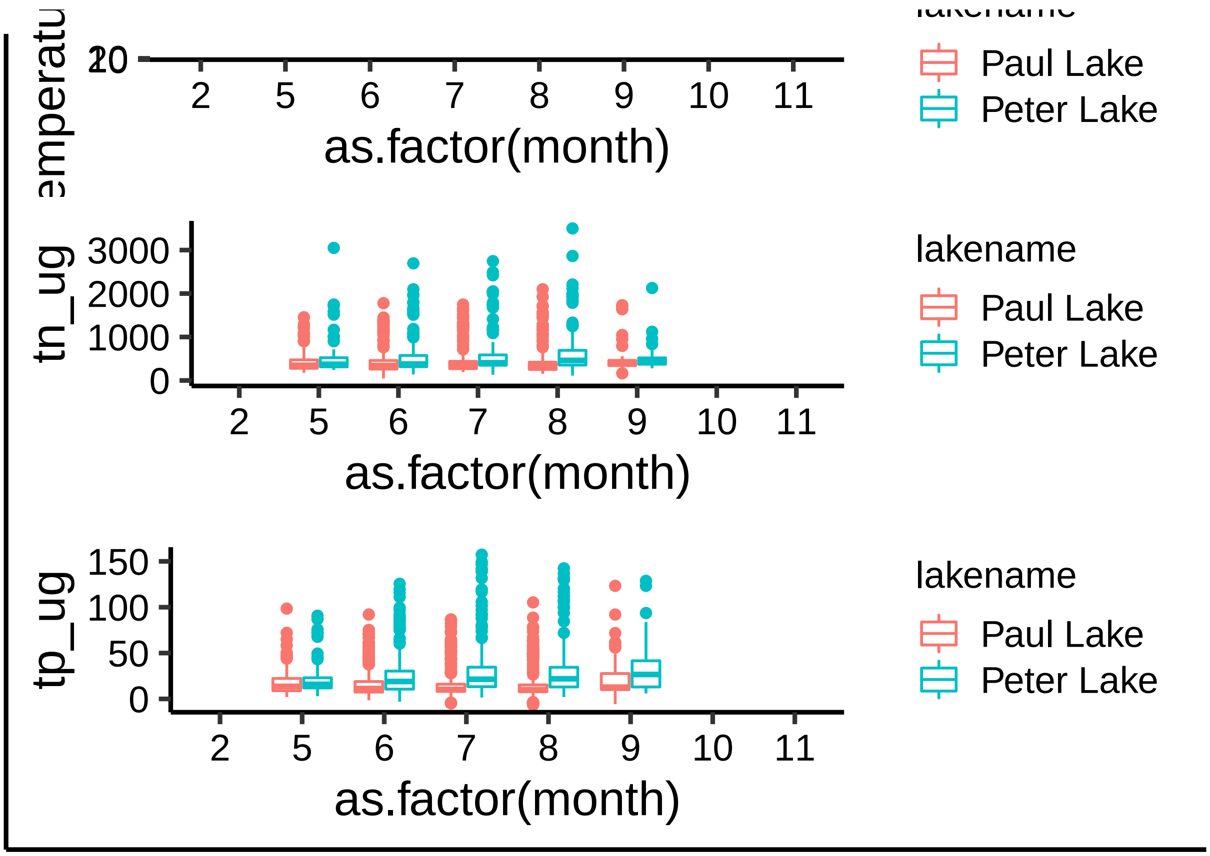

```
boxTN<-
  ggplot(PetPaul.processed,aes(x=as.factor(month),y=tn_ug))+
  geom_boxplot(aes(color=lakename))+
  own.theme
print(boxTN)
```

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




```
plot_grid(boxTemp,boxTN,boxTP, nrow = 3, align = 'h', rel_heights = c(0.75, 1.5, 1.5))+
  own.theme
```

```
## Warning: Removed 3566 rows containing non-finite values (stat_boxplot).
## Warning: Removed 21583 rows containing non-finite values (stat_boxplot).
## Warning: Removed 20729 rows containing non-finite values (stat_boxplot).
```

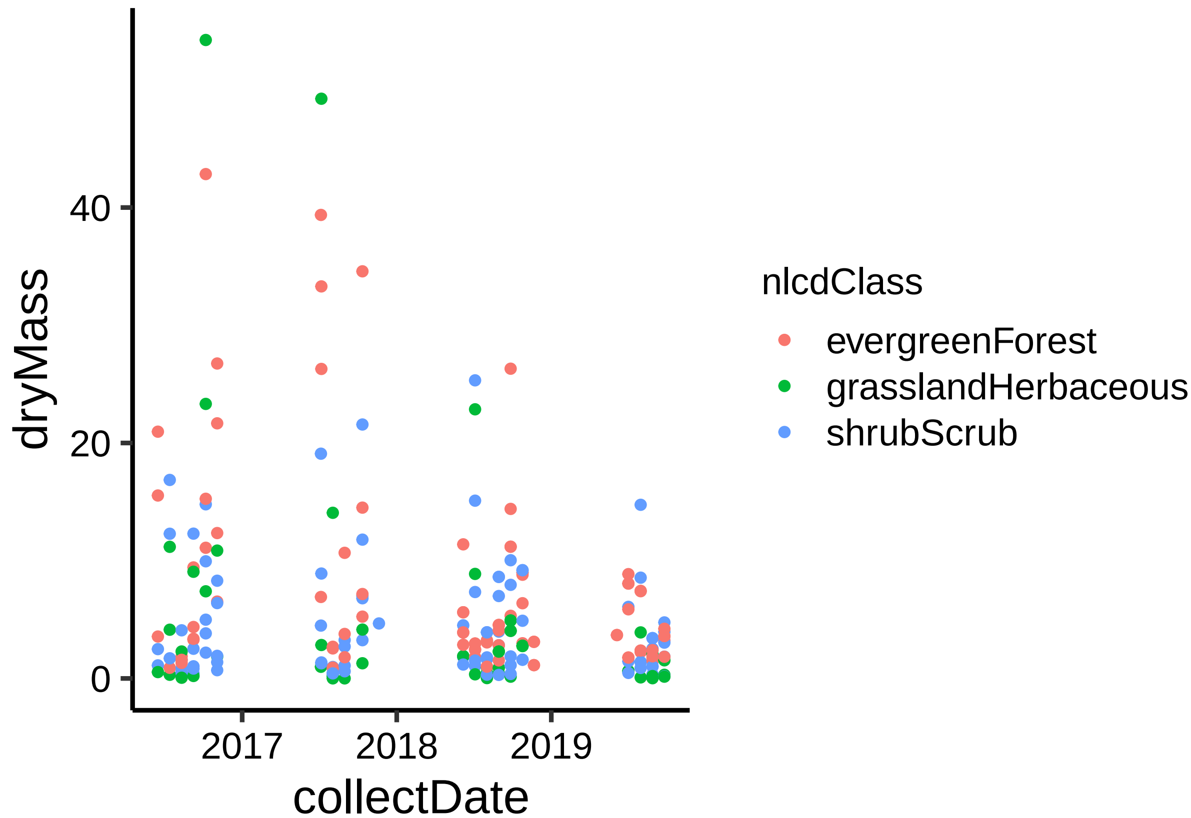


Question: What do you observe about the variables of interest over seasons and between lakes?

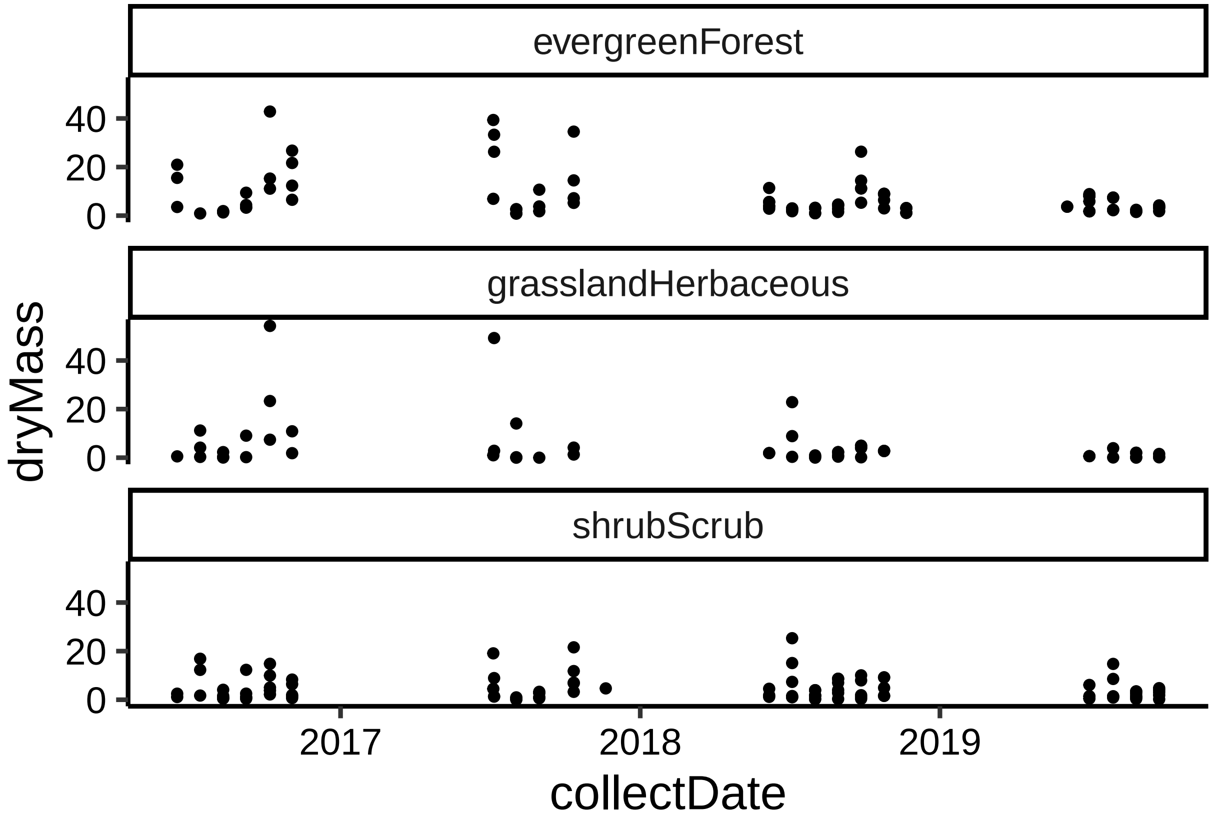
> Answer: We can see by looking at the tempretaure plot that for the first 9 months Paul Lake had higher or equal temperature than Peter Lake while during the 10th month Peter Lake got way more hotter than Paul Lake. Coming to the total nitrogen, it was dominant in Peter Lake across the timeline while the total phosphate was somewhat equal in both lakes during 5th month, the rest of the time Peter Lake being dominant in this component as well.

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 Plot nlcdClass values of dry mass of Needles litter over the dates
Needles.by.nlcd.color<-
  ggplot(subset(Litter, functionalGroup=="Needles"),aes(x=collectDate,y=dryMass))+
  geom_point(aes(color=nlcdClass))+
  own.theme
print(Needles.by.nlcd.color)
```




```
#7 Plot nlcdClass values of dry mass of Needles litter over the dates with facets
Needles.by.nlcd.facets<-
ggplot(subset(Litter, functionalGroup=="Needles"),aes(x=collectDate,y=dryMass))+
  geom_point()+
  facet_wrap(vars(nlcdClass),nrow=3)+
  own.theme
print(Needles.by.nlcd.facets)
```



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

> Answer: I think this question is a bit tricky: from one side using facets is relatively more handy since the observations of different NLCD classes do not bump into each other and really provide the picture of dry mass of Needles per each class; however, coloring helps better with comparing the data since the balck dots that are placed in separate spaces can be confusing in terms of knowing how the observations of each class performed over the given time parameter.