Assignment 5: Data Visualization

Benjamin Culberson

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
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   1.1.2     v stringr 1.4.0
## v readr   1.4.0     v forcats 0.5.0

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

```
library(cowplot)
getwd()
```

```
## [1] "/Users/benculberson/Documents/Duke /Spring 2022/Environmental Data Analytics/Environmental_Data
```

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

#2
class(NTL_LTER.processed$sampledate)
```

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

```
class(Niwot_Ridge.processed$collectDate)
```

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

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

class(NTL_LTER.processed$sampledate)
```

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

```
class(Niwot_Ridge.processed$collectDate)
```

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

Define your theme

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

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

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.

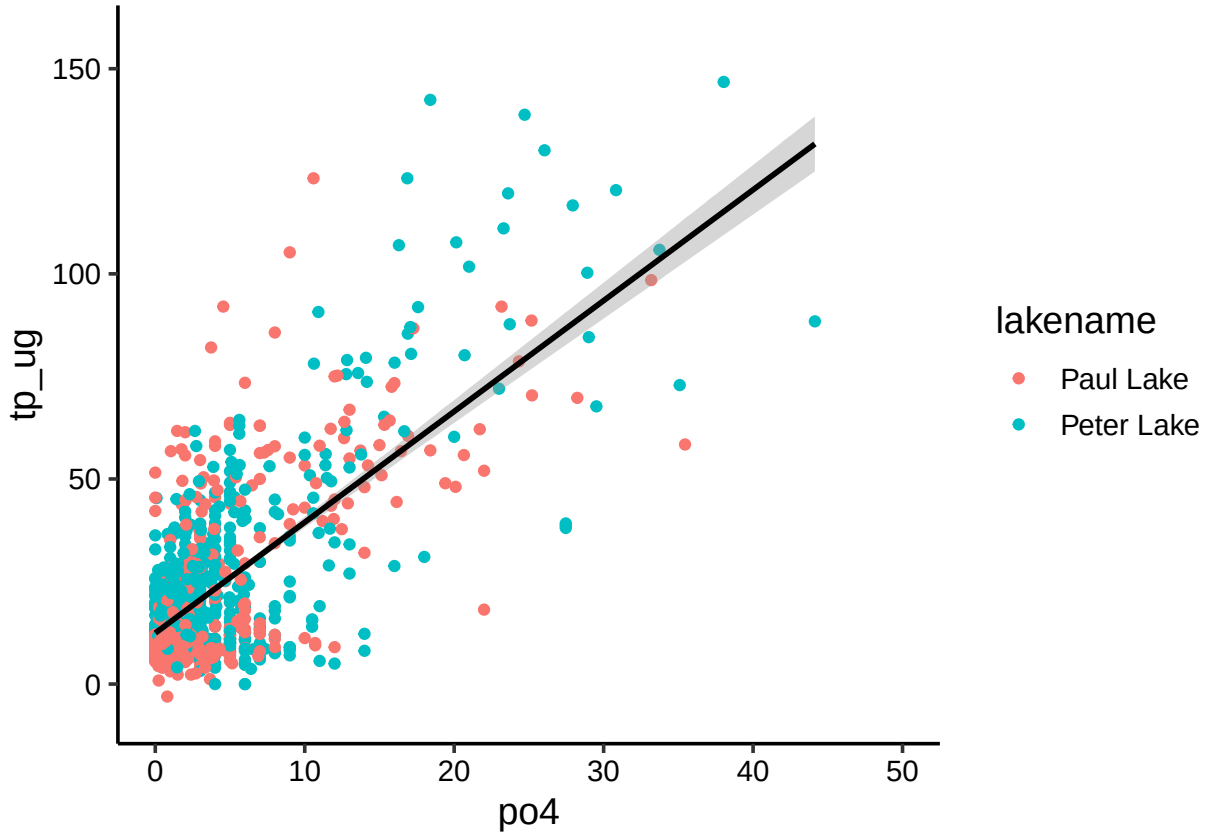
4. [NTL-LTER] Plot total phosphorus (`tp_ug`) by phosphate (`po4`), with separate aesthetics for Peter and Paul lakes. Add a line of best fit and color it black. Adjust your axes to hide extreme values (hint: change the limits using `xlim()` and `ylim()`).

```
#4
NTL_LTER_concentrations <-
  ggplot(NTL_LTER.processed, aes(x = po4, y = tp_ug, color = lakename)) +
  geom_point() +
  geom_smooth(method = lm, color = "black") +
  xlim(0,50)
print(NTL_LTER_concentrations)
```

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

## Warning: Removed 21947 rows containing non-finite values (stat_smooth).

## Warning: Removed 21947 rows containing missing values (geom_point).
```



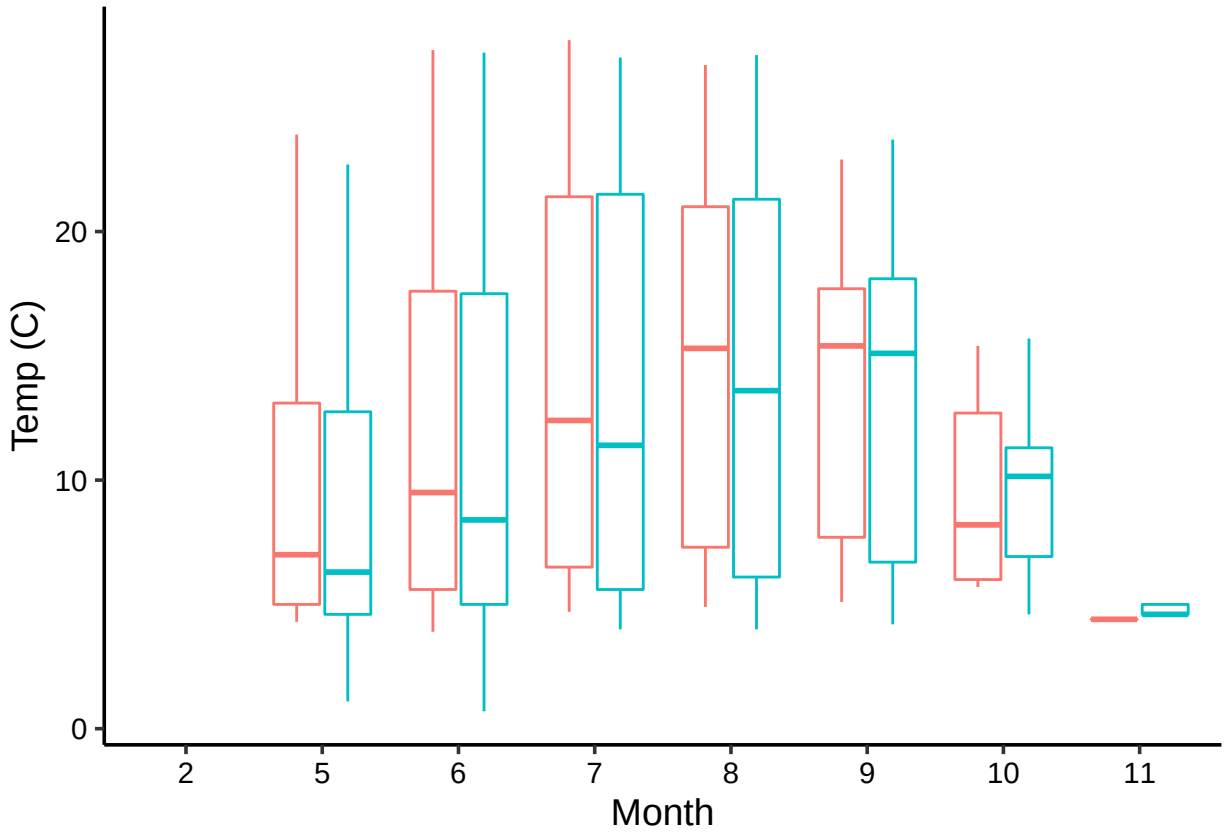
5. [NTL-LTER] Make three separate boxplots of (a) temperature, (b) TP, and (c) TN, with month as the x axis and lake as a color aesthetic. Then, create a cowplot that combines the three graphs. Make sure that only one legend is present and that graph axes are aligned.

```
#5
temperature_box <-ggplot(NTL_LTER.processed) +
  geom_boxplot(aes(x = as.factor(month), y = temperature_C, color = lakename)) +
```

```
  labs(y = "Temp (C)", x = "Month", color = "Lake") +
  theme(legend.position = "none")

print(temperature_box)
```

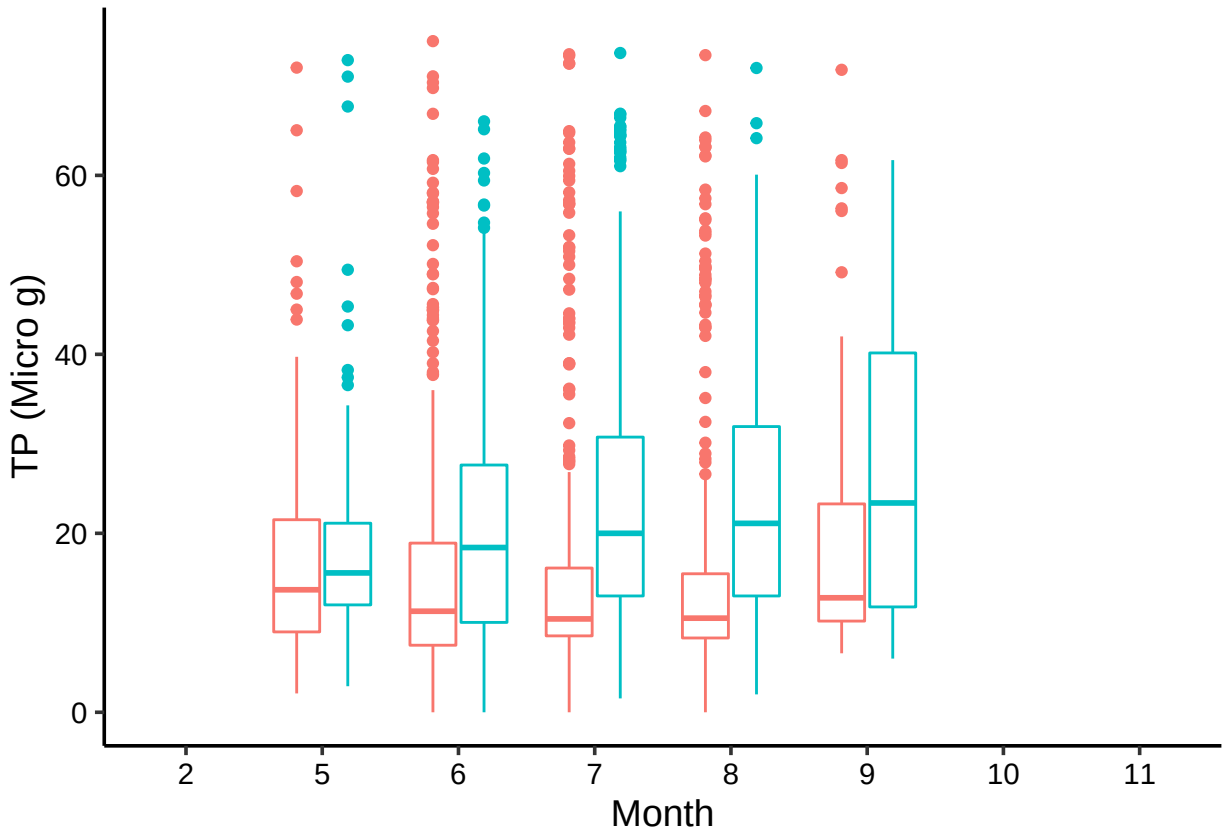
```
## Warning: Removed 3566 rows containing non-finite values (stat_boxplot).
```




```
TP_box <-ggplot(NTL_LTER.processed) +
  geom_boxplot(aes(x = as.factor(month), y = tp_ug, color = lakename)) +
  labs(y = "TP (Micro g)", x = "Month", color = "Lake") +
  ylim(0,75) +
  theme(legend.position = "none")

print(TP_box)
```

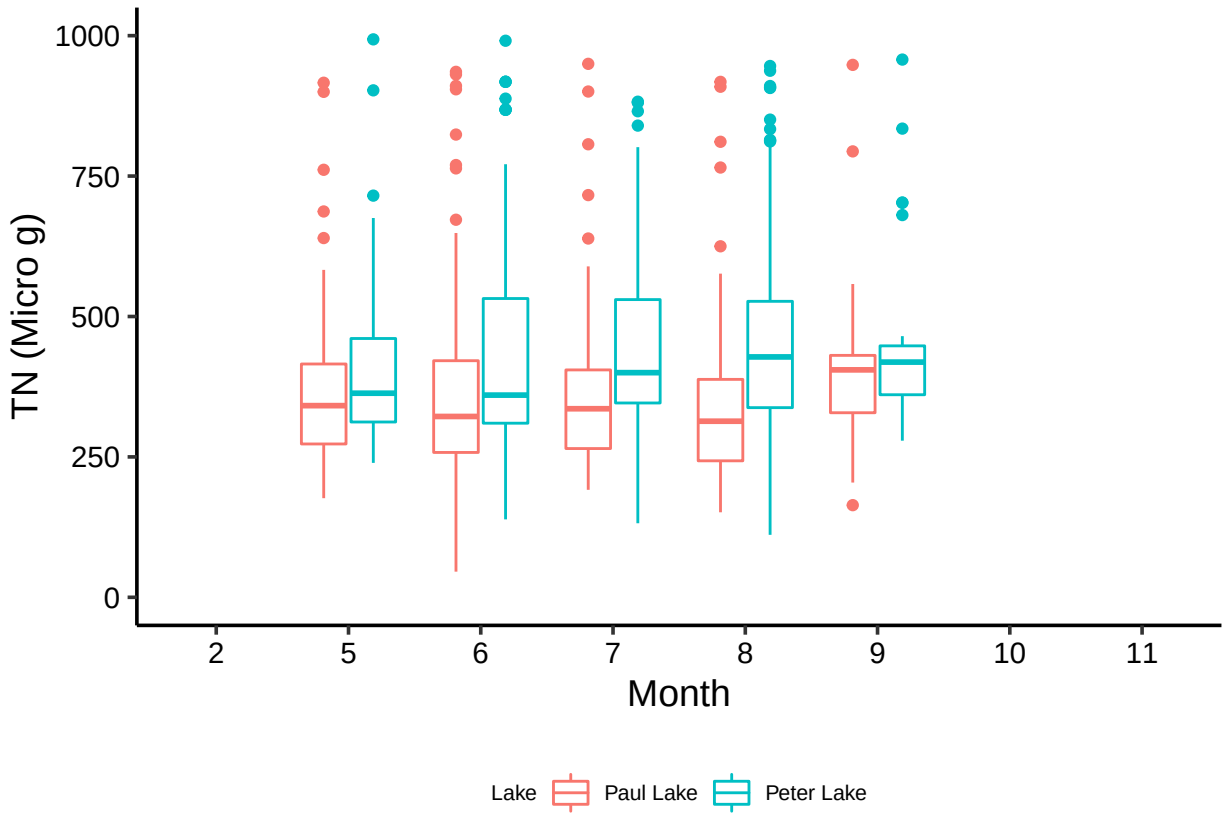
```
## Warning: Removed 20841 rows containing non-finite values (stat_boxplot).
```




```
TN_box <-ggplot(NTL_LTER.processed) +
  geom_boxplot(aes(x = as.factor(month), y = tn_ug, color = lakename)) +
  labs(y = "TN (Micro g)", x = "Month", color = "Lake") +
  ylim(0,1000) +
  theme(legend.position = "bottom",
        legend.text = element_text(size = 8), legend.title = element_text(size = 8))

print(TN_box)
```

```
## Warning: Removed 21730 rows containing non-finite values (stat_boxplot).
```




```
combined_cowplot <- plot_grid(temperature_box, TP_box, TN_box, nrow = 3,
                              align = 'hv', rel_heights = c(1, 1, 1.5)) +
  theme(legend.position = "bottom")
```

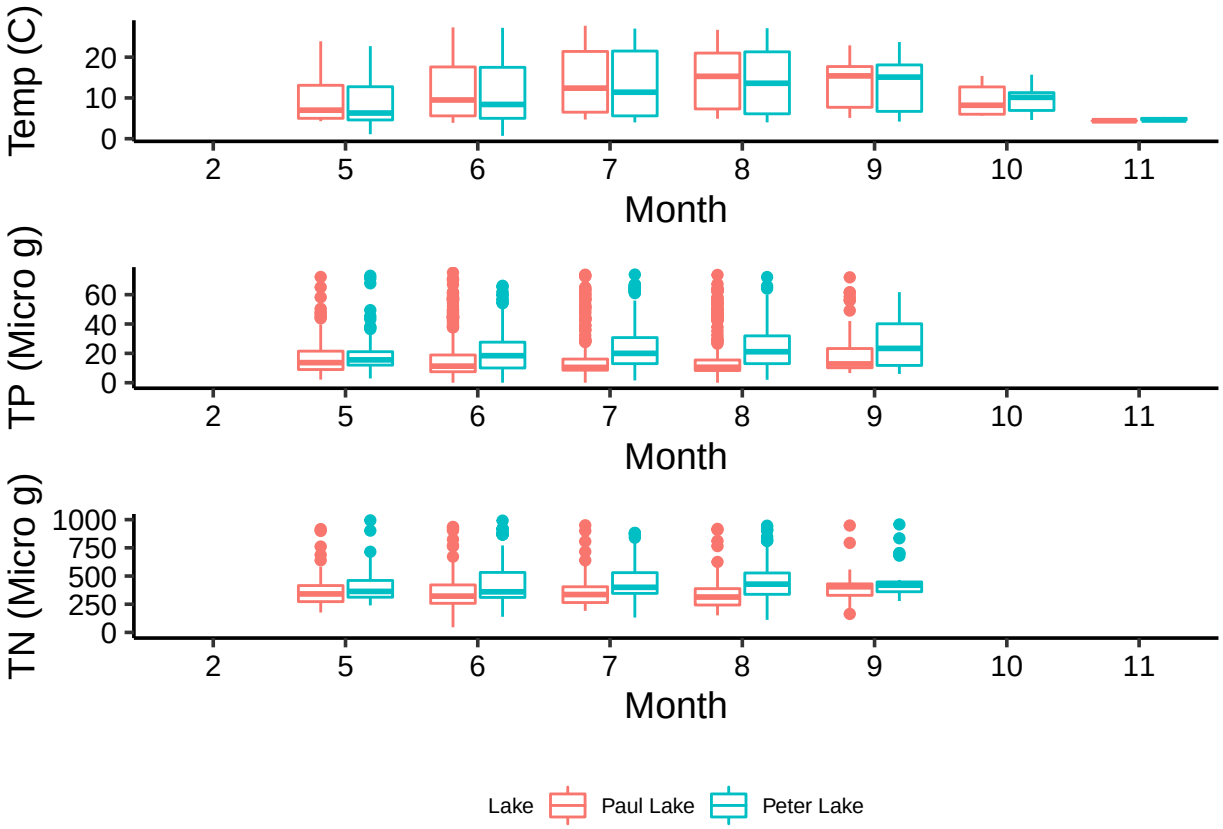
```
## Warning: Removed 3566 rows containing non-finite values (stat_boxplot).
```

```
## Warning: Removed 20841 rows containing non-finite values (stat_boxplot).
```

```
## Warning: Removed 21730 rows containing non-finite values (stat_boxplot).
```

```
## Warning: Graphs cannot be horizontally aligned unless the axis parameter is set.
## Placing graphs unaligned.
```

```
print(combined_cowplot)
```



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

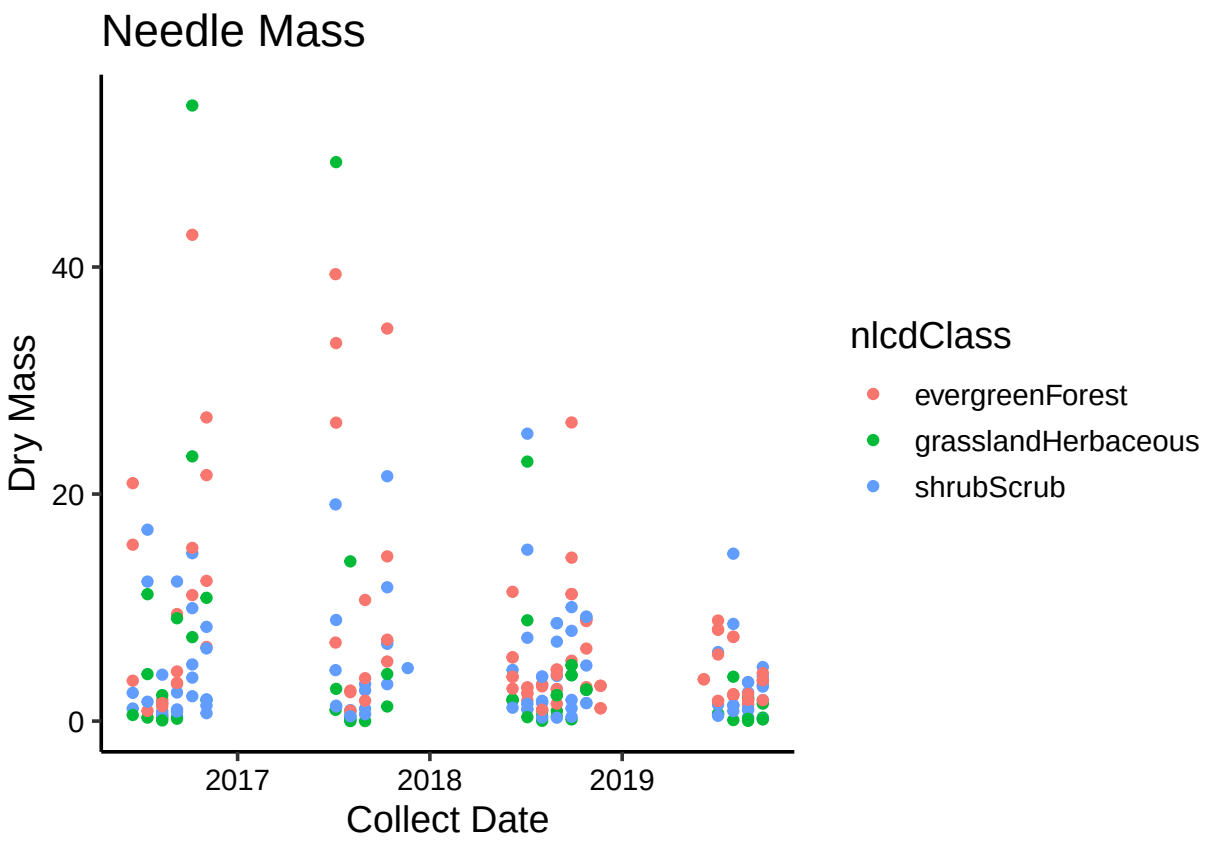
> Answer: It seems that the concentration of both TP and TN is greater in Peter Lake and that is true regardless of the month, although the disparity between the two lakes is greatest in the late summer (more so for TP than TN). Temperature is somewhat similar between the two lakes, although there is some variation between the two month to month. The temperature of both lakes also varies more month to month: hottest in the summer, coldest in the winter (although it is also sampled more frequently than the TP and TN concentrations and perhaps a better understanding of the temporal variation of TP and TN may be possible if they were sampled during the winter).

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.

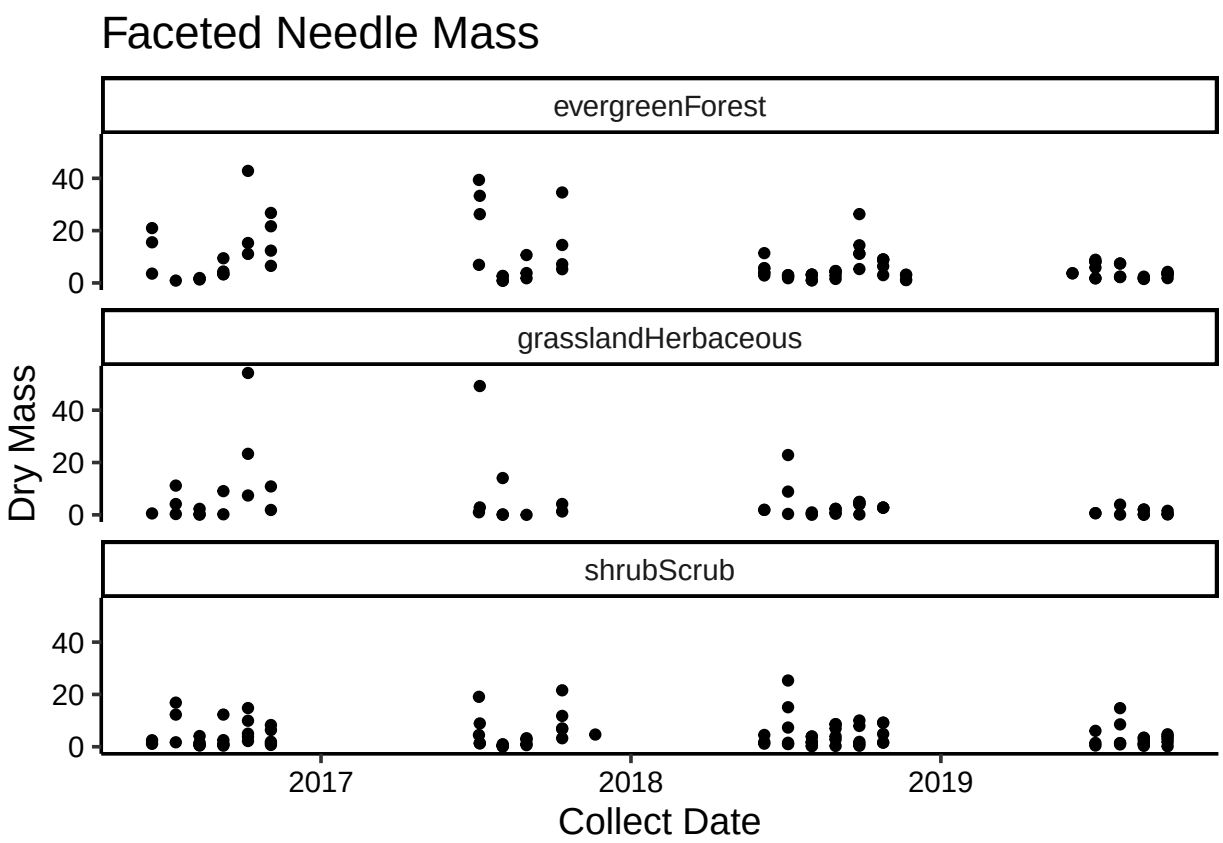
```
#6
Niwot_Ridge_Needles <-
  ggplot(subset(Niwot_Ridge.processed, functionalGroup == "Needles"),
         aes(x = collectDate, y = dryMass, color = nlcdClass)) +
  labs(title = "Needle Mass", x = "Collect Date", y= "Dry Mass" ) +
  geom_point()

print(Niwot_Ridge_Needles)
```




```
#7
Niwot_Ridge_Needles_facets <-
  ggplot(subset(Niwot_Ridge.processed, functionalGroup == "Needles"),
         aes(x = collectDate, y = dryMass)) +
  geom_point() +
  labs(title = "Faceted Needle Mass", x = "Collect Date", y= "Dry Mass" ) +
  facet_wrap(vars(nlcdClass), nrow = 3)

print(Niwot_Ridge_Needles_facets)
```



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

Answer: In this case, I think plot 7 is more effective. Plot 6 reminds me of a spaghetti plot where there is too much infomration on a single graph such that the viewer cannot follow. The coloration of plot 6 is not useful if one tries to look at each class seperately. Only in plot 7 is there a realistic means of viewing the data in terms of NLCD class and Collect Date. It is possible to view the data in plot 6 in terms of Collect Date, but colors don't split up the NLCD class in a clear way for the viewer.