Assignment 5: Data Visualization

Yingchi Cheung

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 21th @ 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_PeterPau`
version) and the processed data file for the Niwot Ridge litter dataset (use the [`NEON_NIWO_Litter_mass_trap_Processed`
version).

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

```
# 1
getwd()
```

```
## [1] "/home/guest/ENV872/EDA-Fall2022"
```

```
library(tidyverse)
```

```
## -- Attaching packages --------------------------------------- tidyverse 1.3.2 --
## v ggplot2 3.3.6      v purrr   0.3.4
## v tibble  3.1.8      v dplyr   1.0.10
## v tidyr   1.2.0      v stringr 1.4.1
## v readr   2.1.2      v forcats 0.5.2
## -- 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(colormap)
NTLLTER <- read.csv("./Data/Processed/NTL-LTER_Lake_Chemistry_Nutrients_PeterPaul_Processed.csv",
    stringsAsFactors = TRUE)
NEON_NIWO <- read.csv("./Data/Processed/NEON_NIWO_Litter_mass_trap_Processed.csv",
    stringsAsFactors = TRUE)

# 2
NTLLTER$sampledate <- as.Date(NTLLTER$sampledate, format = "%Y-%m-%d")
class(NTLLTER$sampledate)
```

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

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

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

Define your theme

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

```
# 3
mytheme <- theme_classic(base_size = 12) + theme(axis.text = element_text(color = "black"),
    legend.position = "bottom")
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.

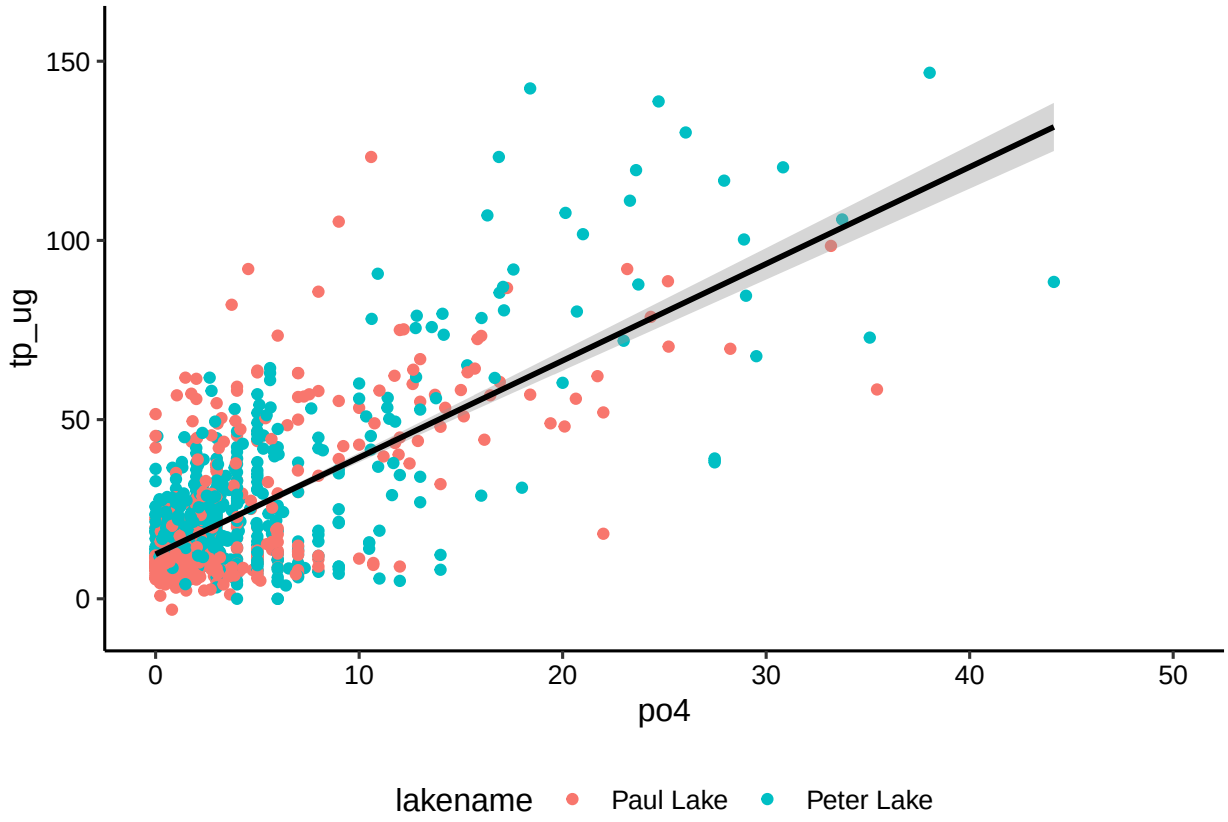
4. [NTL-LTER] Plot total phosphorus (`tp_ug`) by phosphate (`po4`), with separate aesthetics for Peter and Paul lakes. Add a line of best fit and color it black. Adjust your axes to hide extreme values (hint: change the limits using `xlim()` and/or `ylim()`).

```
# 4
PvsPO4 <- ggplot(NTLLTER, aes(x = po4, y = tp_ug)) + geom_point(aes(color = lakename)) +
    xlim(0, 50) + geom_smooth(method = lm, color = "black")
print(PvsPO4)
```

```
## `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.

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

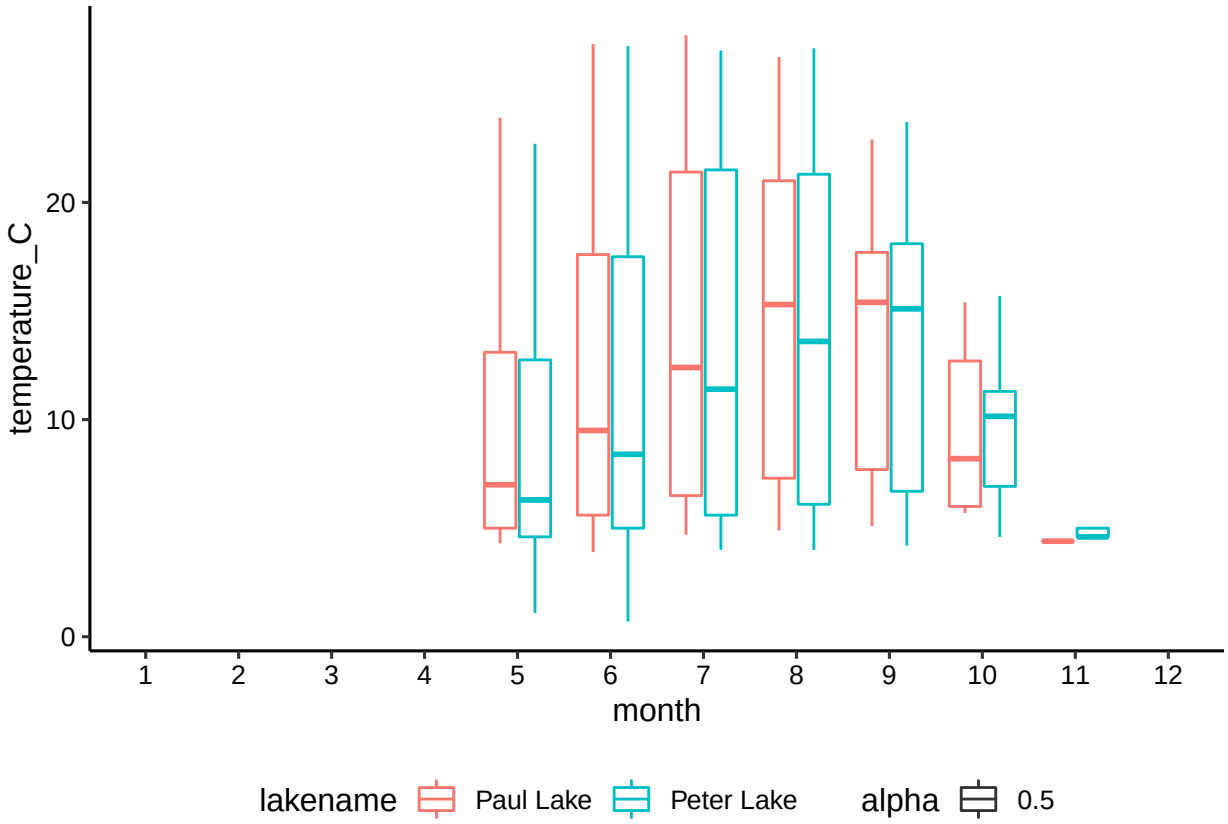
```
# 5
NTLLTER$month <- factor(NTLLTER$month, levels = c(1:12))
library(cowplot)
```

```
##
## Attaching package: 'cowplot'
```

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

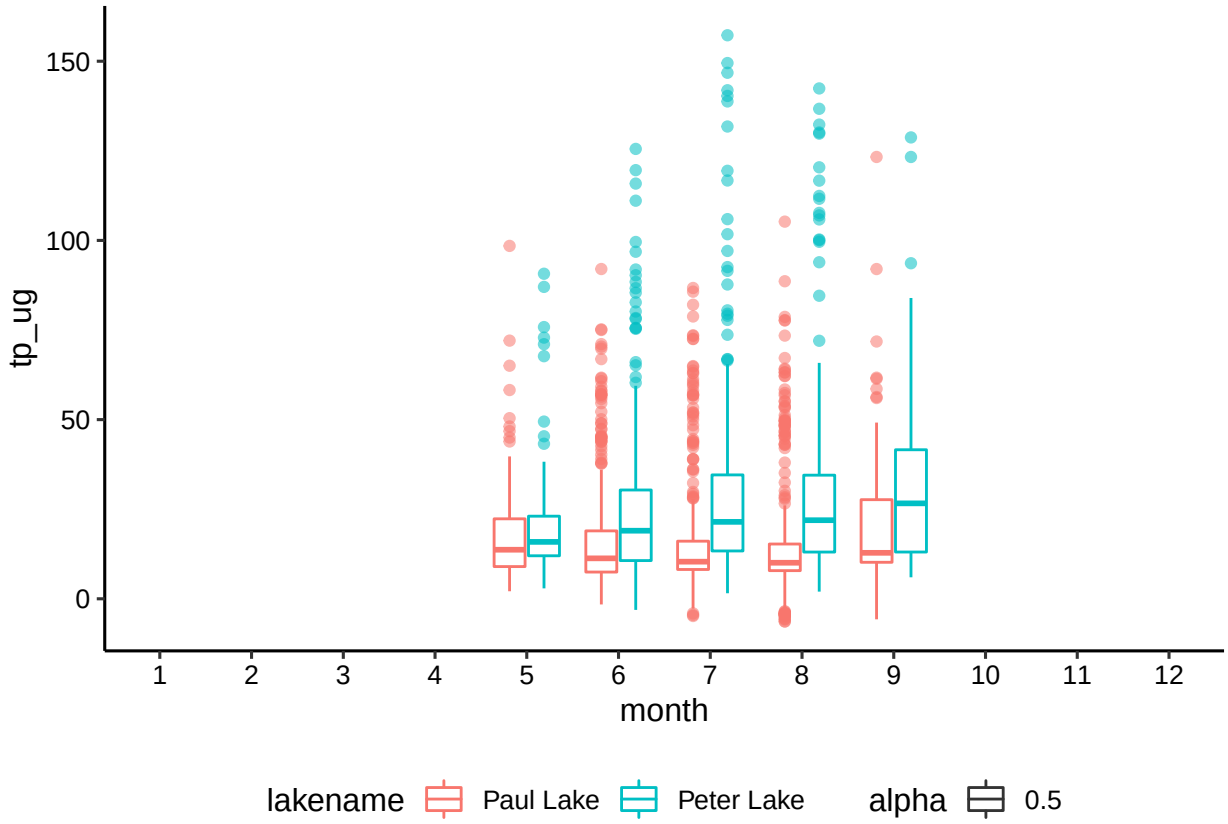
```
Temperature <- ggplot(NTLLTER, aes(x = month, y = temperature_C)) + geom_boxplot(aes(color = lakename,
    alpha = 0.5)) + scale_x_discrete(drop = FALSE)
print(Temperature)
```

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



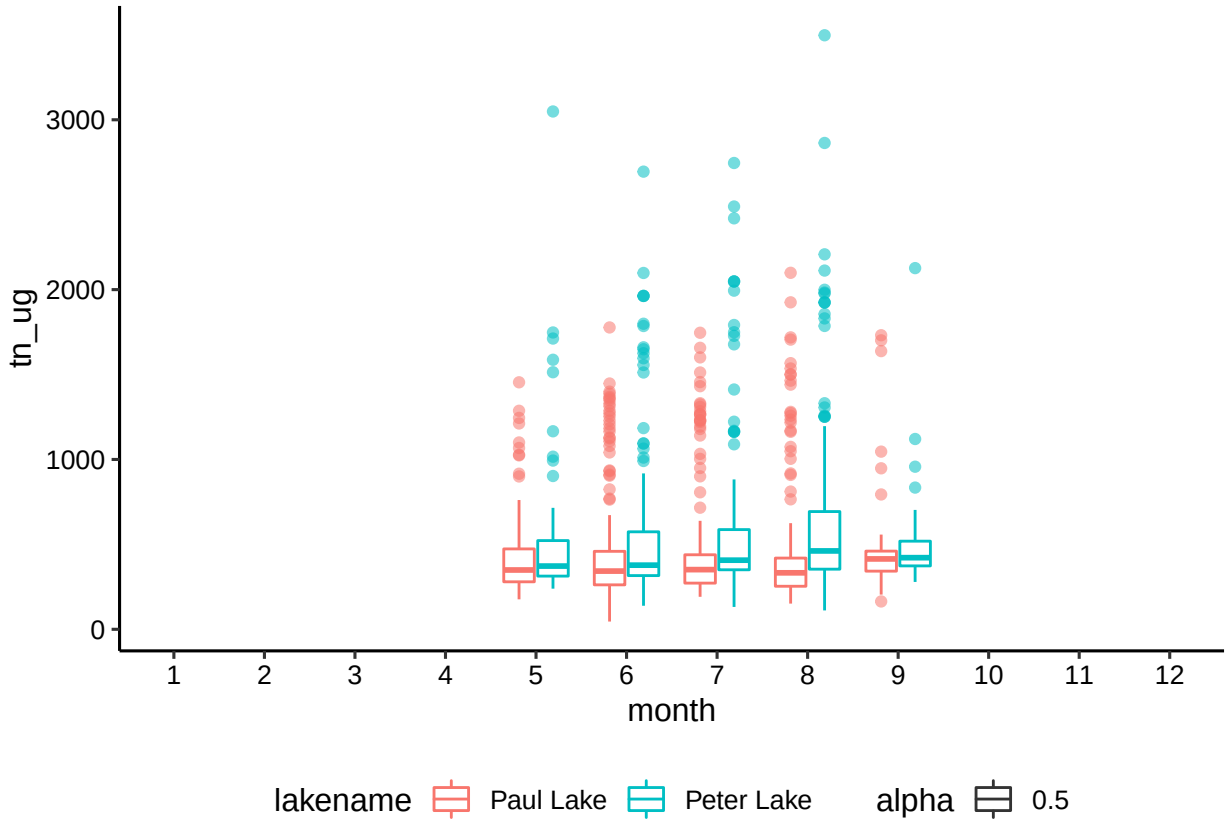

```
TP <- ggplot(NTLLTER, aes(x = month, y = tp_ug)) + geom_boxplot(aes(color = lakename,
    alpha = 0.5)) + scale_x_discrete(drop = FALSE)
print(TP)
```

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




```
TN <- ggplot(NTLLTER, aes(x = month, y = tn_ug)) + geom_boxplot(aes(color = lakename,
    alpha = 0.5)) + scale_x_discrete(drop = FALSE)
print(TN)
```

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




```
month.abb
```

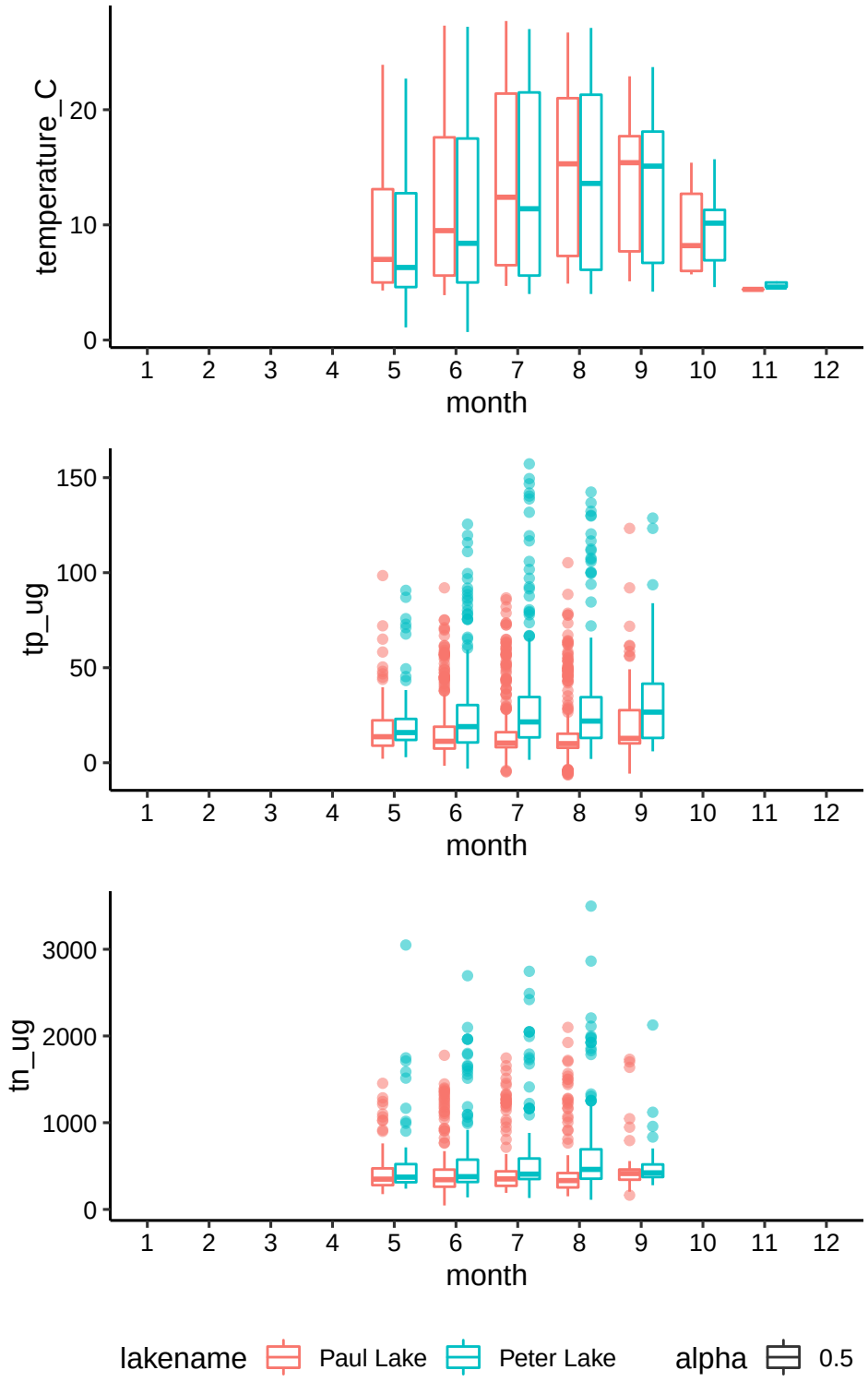
```
##  [1] "Jan" "Feb" "Mar" "Apr" "May" "Jun" "Jul" "Aug" "Sep" "Oct" "Nov" "Dec"
```

```
plot_grid(Temperature + theme(legend.position = "none"), TP + theme(legend.position = "none"),
    TN, ncol = 1, align = "v", axis = "l", rel_heights = c(1, 1, 1.2))
```

```
## 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?

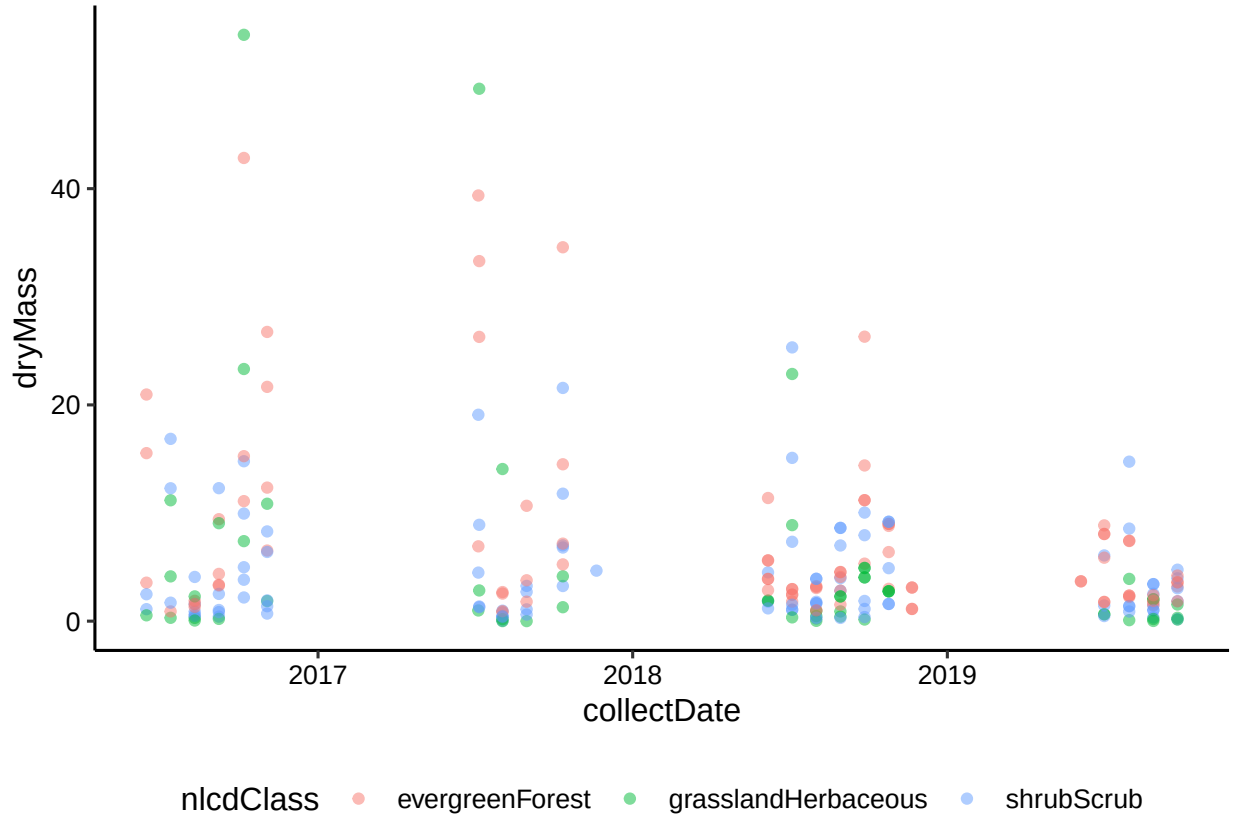
Answer: Peter Lake has a higher temperature than Paul Lake in November. But, the temperatures of both lakes in the rest of the months are not significantly different from others. Peter Lake's TP amount is likely to be higher than Paul Lake's in July and August. Overall, Peter Lack has higher outliers than Paul Lake, but not in March. In August, there is likely to be a higher amount of

TN in Peter Lake than in Paul Lake. Peter Lack has higher outliers than Paul Lake throughout all year.

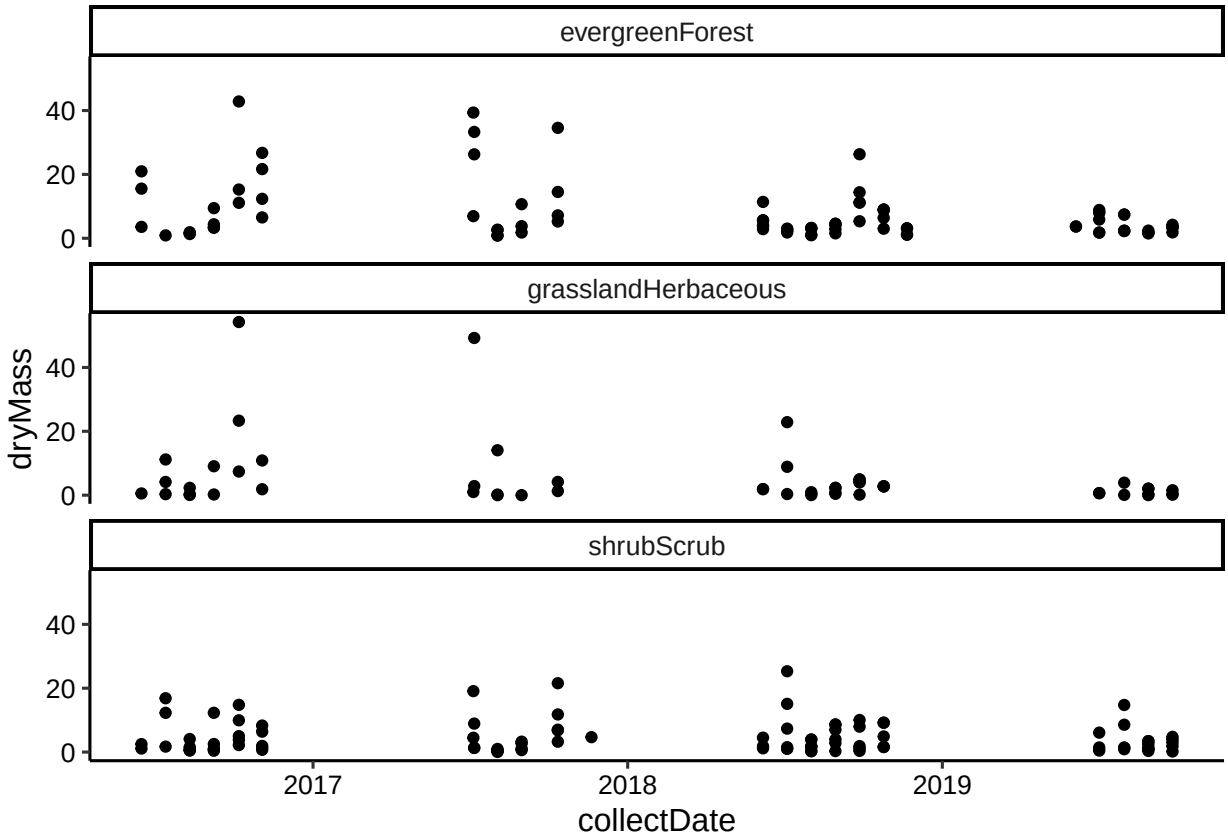
6. [Niwot Ridge] Plot a subset of the litter dataset by displaying only the "Needles" functional group. Plot the dry mass of needle litter by date and separate by NLCD class with a color aesthetic. (no need to adjust the name of each land use)

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

```
# 6
Needles <- filter(NEON_NIWO, functionalGroup == "Needles")
litter_color <- ggplot(Needles, aes(x = collectDate, y = dryMass)) + geom_point(aes(color = nlcdClass),
    alpha = 0.5)
print(litter_color)
```




```
# 7
litter_facets <- ggplot(Needles, aes(x = collectDate, y = dryMass)) + geom_point() +
    facet_wrap(vars(nlcdClass), nrow = 3)
print(litter_facets)
```



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

Answer: The plot created in question 7 is more effective. The separation into three facets looks clearer than the color separation in question 6. In question 7, We can easily compare them by year and NLCD classes because they are not crowded together. But in question 6, even though they are categorized by year, it is still hard to tell these NLCD classes apart, let alone compare them.