

Assignment 5: Data Visualization

Karina Leung

Spring 2023

OVERVIEW

This exercise accompanies the lessons in Environmental Data Analytics on Data Visualization

Directions

1. Rename this file `<FirstLast>_A05_DataVisualization.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 your code is tidy; use line breaks to ensure your code fits in the knitted output.
5. Be sure to **answer the questions** in this assignment document.
6. When you have completed the assignment, **Knit** the text and code into a single PDF file.

Set up your session

1. Set up your session. Load the tidyverse, lubridate, here & cowplot packages, and verify your home directory. 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 1.3.2 --
## v ggplot2 3.4.1      v purrr   1.0.1
## v tibble  3.1.8      v dplyr   1.1.0
## v tidyr   1.3.0      v stringr 1.5.0
## v readr   2.1.4      v forcats 1.0.0
## -- 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(formatR)  
library(here)
```

```
## here() starts at C:/Users/ktlro/OneDrive/Documents/EDA-Spring2023
```

```
library(cowplot)
```

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

```
library(viridis)
```

```
## Loading required package: viridisLite
```

```
knitr::opts_chunk$set(tidy.opts=list(width.cutoff=80), tidy=TRUE)
```

```
here()
```

```
## [1] "C:/Users/ktlro/OneDrive/Documents/EDA-Spring2023"
```

```
PeterPaul.chem.nutrients <- read.csv(here("Data/Processed/NTL-LTER_Lake_Chemistry_Nutrients_PeterPaul_P"))
```

```
NEON.litter.mass <- read.csv(here("Data/Processed/NEON_NIWO_Litter_mass_trap_Processed.csv"), stringsAsFactors=FALSE)
```

```
#2
```

```
PeterPaul.chem.nutrients$sampleddate <- ymd(PeterPaul.chem.nutrients$sampleddate)  
NEON.litter.mass$collectDate <- ymd(NEON.litter.mass$collectDate)
```

Define your theme

3. Build a theme and set it as your default theme. Customize the look of at least two of the following:

- Plot background
- Plot title

- Axis labels
- Axis ticks/gridlines
- Legend

3

```
mytheme <- theme_classic(base_size = 12) + theme(axis.text = element_text(color = "black",
  size = 12), plot.title.position = "plot", plot.title = element_text(size = 16,
  face = "bold"), axis.ticks = element_line(color = "black", linewidth = 0.5),
  legend.background = element_rect(color = "grey", fill = "white"), legend.title = element_text(color
  legend.position = "right")
```

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

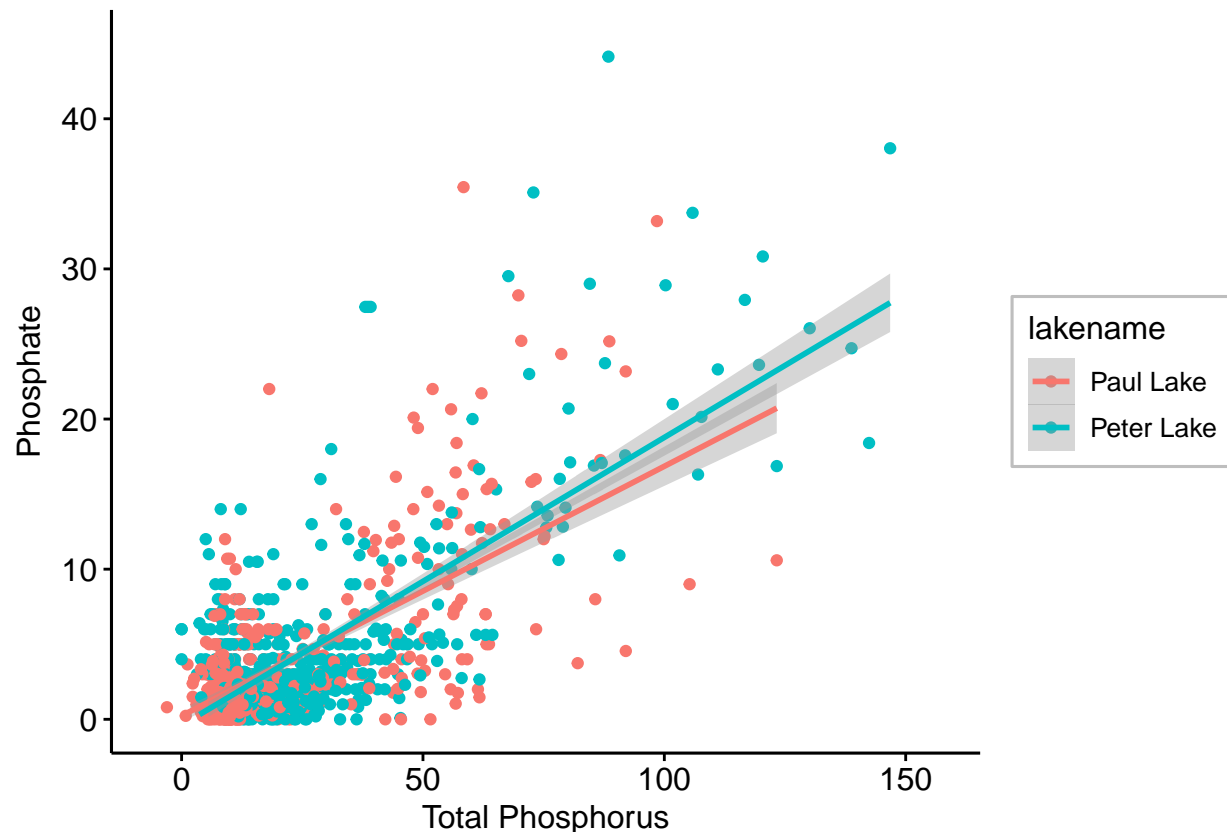
```
TPvsPo <- ggplot(PeterPaul.chem.nutrients, aes(x = tp_ug, y = po4, color = lakename)) +
  geom_point() + mytheme + ylab(expression(y = "Phosphate")) + xlab(expression(x = "Total Phosphorus"))
  geom_smooth(method = lm, formula = y ~ x) + ylim(0, 45)

print(TPvsPo)
```

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

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

```
## Warning: Removed 4 rows containing missing values ('geom_smooth()').
```



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 Checking month class
```

```
class(PeterPaul.chem.nutrients$month)
```

```
## [1] "integer"
```

```
unique(PeterPaul.chem.nutrients$month)
```

```
## [1] 5 6 7 8 9 10 11 2
```

```
# Changing month to show all 12 and abbreviate
```

```
factor(PeterPaul.chem.nutrients$month, levels = 1:12, labels = month.abb)
```

```
##      [1] May May May May May May May May May May May May May May May May May May May
##     [19] May May May May May May May May May May May May May May May May May May May
##     [37] May May May Jun Jun Jun Jun Jun Jun Jun Jun Jun Jun Jun Jun Jun Jun Jun Jun
##     [55] Jun Jun Jun Jun Jun Jun Jun Jun Jun Jun Jun Jun Jun Jun Jun Jun Jun Jun Jun Jun
```

[illegible]

[illegible]

[illegible]

[illegible]

[illegible]

[illegible]

[illegible]

[illegible]

[illegible]

[illegible]

[illegible]

[illegible]

[illegible]

[illegible]

[illegible]

[illegible]

[illegible]

[illegible]

[illegible]

[illegible]

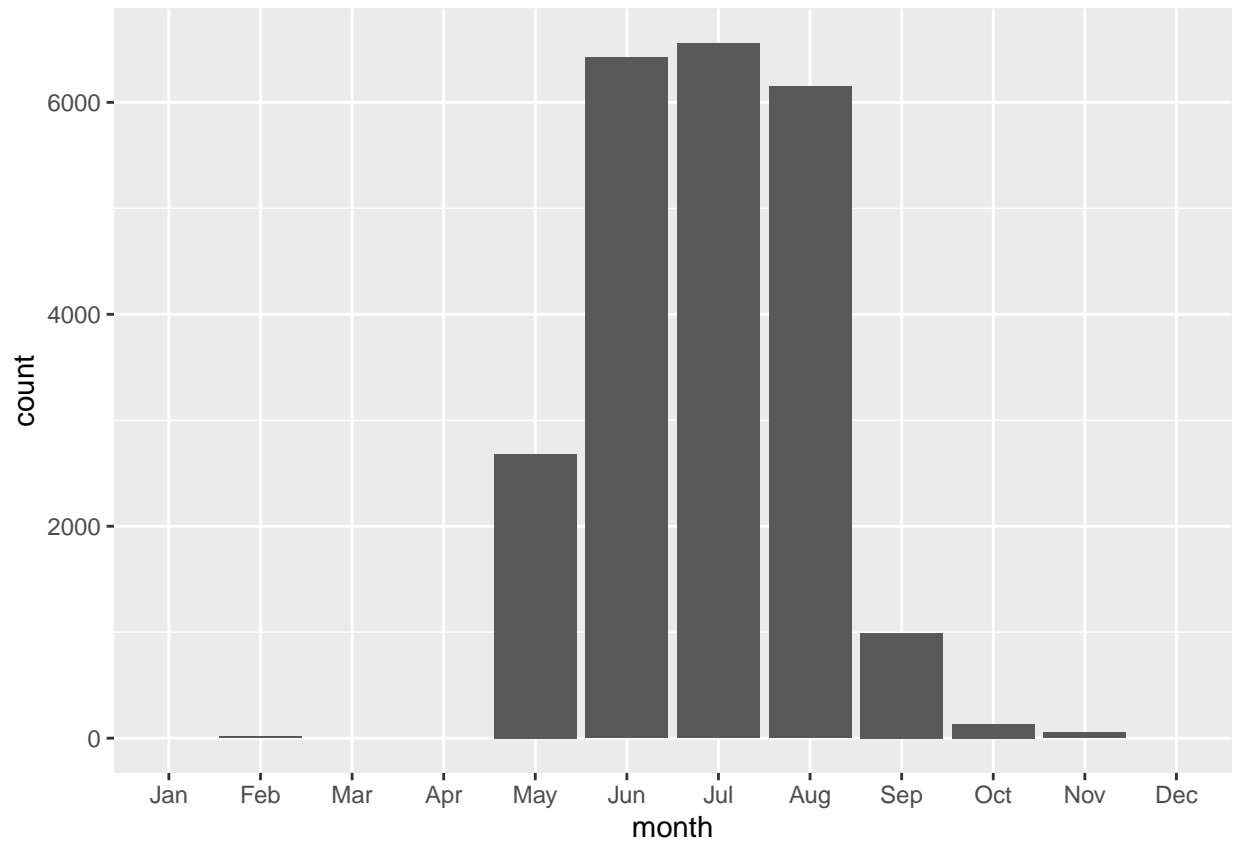
[illegible]

[illegible]

##	[21457]	Jul	Jul	Jul	Jul	Jul	Jul	Jul	Jul	Jul	Jul	Jul	Jul	Jul	Jul	Jul	Jul	Jul
##	[21475]	Jul	Jul	Jul	Jul	Jul	Jul	Jul	Jul	Jul	Jul	Aug	Aug	Aug	Aug	Aug	Aug	Aug
##	[21493]	Aug	Aug	Aug	Aug	Aug	Aug	Aug	Aug	Aug	Aug	Aug	Aug	Aug	Aug	Aug	Aug	Aug
##	[21511]	Aug	Aug	Aug	Aug	Aug	Aug	Aug	Aug	Aug	Aug	Aug	Aug	Aug	Aug	Aug	Aug	Aug
##	[21529]	Aug	Aug	Aug	Aug	Aug	Aug	Aug	Aug	Aug	Aug	Aug	Aug	Aug	Aug	Aug	Aug	Aug
##	[21547]	Aug	Aug	Aug	Aug	Aug	Aug	Aug	Aug	Aug	Aug	Aug	Aug	Aug	Aug	Aug	Aug	Aug
##	[21565]	Aug	Aug	Aug	Aug	Aug	Aug	Aug	Aug	Aug	Aug	Aug	Aug	Aug	Aug	Aug	Aug	Aug
##	[21583]	Aug	Aug	Aug	Aug	Aug	Aug	Aug	Aug	Aug	Aug	Aug	Aug	Aug	Aug	Aug	Aug	Aug
##	[21601]	Aug	Aug	Aug	Aug	Aug	Aug	Aug	Aug	Aug	Aug	Aug	Aug	Aug	May	May	May	May
##	[21619]	May	May	May	May	May	Jun	Jun	Jun	Jun	Jun	Jun	Jun	Jun	Jun	Jun	Jun	Jun
##	[21637]	Jun	Jun	Jun	Jun	Jun	Jun	Jun	Jun	Jun	Jun	Jun	Jun	Jun	Jun	Jun	Jun	Jun
##	[21655]	Jun	Jun	Jun	Jul	Jul	Jul	Jul	Jul	Jul	Jul	Jul	Jul	Jul	Jul	Jul	Jul	Jul
##	[21673]	Jul	Jul	Jul	Jul	Jul	Jul	Jul	Jul	Jul	Jul	Jul	Aug	Aug	Aug	Aug	Aug	Aug
##	[21691]	Aug	Aug	Aug	Aug	Aug	Aug	Aug	Aug	Aug	Aug	Aug	Aug	Aug	Aug	Aug	Aug	Aug
##	[21709]	Aug	Sep	Sep	Sep	Sep	Sep	Sep	Sep	Sep	May	May	May	May	May	May	May	May
##	[21727]	May	May	May	May	May	May	May	May	May	May	May	May	May	May	May	Jun	Jun
##	[21745]	Jun	Jun	Jun	Jun	Jun	Jun	Jun	Jun	Jun	Jun	Jun	Jun	Jun	Jun	Jun	Jun	Jun
##	[21763]	Jun	Jun	Jun	Jun	Jun	Jun	Jun	Jun	Jun	Jun	Jun	Jun	Jun	Jun	Jun	Jun	Jun
##	[21781]	Jun	Jun	Jul	Jul	Jul	Jul	Jul	Jul	Jul	Jul	Jul	Jul	Jul	Jul	Jul	Jul	Jul
##	[21799]	Jul	Jul	Jul	Jul	Jul	Jul	Jul	Jul	Jul	Jul	Jul	Jul	Jul	Jul	Jul	Jul	Jul
##	[21817]	Jul	Jul	Jul	Jul	Jul	Jul	Jul	Jul	Jul	Jul	Jul	Jul	Jul	Jul	Jul	Jul	Jul
##	[21835]	Jul	Jul	Jul	Jul	Jul	Jul	Aug	Aug	Aug	Aug	Aug	Aug	Aug	Aug	Aug	Aug	Aug
##	[21853]	Aug	Aug	Aug	Aug	Aug	Aug	Aug	Aug	Aug	Aug	Aug	Aug	Aug	Aug	Aug	Aug	Aug
##	[21871]	Aug	Aug	Aug	Aug	Aug	Aug	Aug	Aug	Aug	Aug	Aug	Aug	Aug	Sep	Sep	Sep	Sep
##	[21889]	Sep	Sep	Sep	Sep	Sep	Sep	Sep	Sep	Sep	Sep	Sep	Sep	Sep	Sep	May	May	May
##	[21907]	May	May	May	May	May	May	May	May	May	May	May	May	Jun	Jun	Jun	Jun	Jun
##	[21925]	Jun	Jun	Jun	Jun	Jun	Jun	Jun	Jun	Jun	Jun	Jun	Jun	Jun	Jun	Jun	Jun	Jun
##	[21943]	Jun	Jun	Jun	Jun	Jun	Jun	Jun	Jun	Jun	Jun	Jun	Jun	Jun	Jun	Jun	Jul	Jul
##	[21961]	Jul	Jul	Jul	Jul	Jul	Jul	Jul	Jul	Jul	Jul	Jul	Jul	Jul	Jul	Jul	Jul	Jul
##	[21979]	Jul	Jul	Jul	Jul	Jul	Jul	Jul	Jul	Aug	Aug	Aug	Aug	Aug	Aug	Aug	Aug	Aug
##	[21997]	Aug	Aug	Aug	Aug	Aug	Aug	Aug	Aug	Aug	Aug	Aug	Aug	Aug	Aug	Aug	Aug	Aug
##	[22015]	Aug	Aug	Aug	Aug	Aug	Aug	Aug	Aug	Aug	Aug	Aug	Aug	Aug	Aug	Sep	Sep	Sep
##	[22033]	Sep	Sep	Sep	Sep	Sep	May	May	May	May	May	May	May	May	May	May	May	May
##	[22051]	May	May	May	May	Jun	Jun	Jun	Jun									

```
## [22429] Jun Jun Jun Jun Jun Jun Jun Jun Jun Jun Jun Jun Jun Jun Jun Jun Jun Jun Jun Jun
## [22447] Jun Jun Jun Jun Jun Jun Jun Jun Jun Jun Jun Jun Jun Jul Jul Jul Jul Jul Jul Jul
## [22465] Jul Jul Jul Jul Jul Jul Jul Jul Jul Jul Jul Jul Jul Jul Jul Jul Jul Jul Jul Jul
## [22483] Jul Jul Jul Jul Jul Jul Jul Jul Jul Jul Jul Jul Jul Jul Jul Jul Jul Aug Aug Aug Aug
## [22501] Aug Aug Aug Aug Aug Aug Aug Aug Aug Aug Aug Aug Aug Aug Aug Aug Aug Aug Aug Aug Aug
## [22519] Aug Aug Aug Aug Aug Aug Aug Aug Aug Aug Sep Sep Sep Sep Sep Sep Sep Sep May May May
## [22537] May May May Jun Jun Jun Jun Jun Jun Jun Jun Jun Jun Jun Jun Jun Jun Jun Jun Jun Jun
## [22555] Jun Jun Jun Jun Jun Jun Jun Jun Jun Jun Jun Jun Jun Jun Jun Jun Jun Jun Jun Jun Jul Jul
## [22573] Jul Jul Jul Jul Jul Jul Jul Jul Jul Jul Jul Jul Jul Jul Jul Jul Jul Jul Jul Jul Jul
## [22591] Jul Jul Jul Jul Jul Jul Jul Jul Jul Jul Jul Jul Jul Jul Jul Aug Aug Aug Aug Aug Aug Aug
## [22609] Aug Aug Aug Aug Aug Aug Aug Aug Aug Aug Aug Aug Aug Aug Aug Aug Aug Aug Aug Aug Aug
## [22627] Aug Aug Aug Aug Aug Aug Aug Aug Aug Aug Aug Aug Aug Aug Aug Aug Aug Aug May May May
## [22645] May May May May Jun Jun Jun Jun Jun Jun Jun Jun Jun Jun Jun Jul Jul Jul Jul Jul
## [22663] Jul Jul Aug Aug Aug Aug Aug Aug Aug Aug Aug Aug May May May May May May May May May
## [22681] Jun Jun Jun Jun Jun Jun Jun Jun Jun Jun Jun Jun Jun Jun Jun Jun Jun Jun Jun Jun Jun
## [22699] Jun Jun Jun Jun Jun Jun Jun Jun Jun Jun Jun Jun Jun Jun Jun Jun Jun Jun Jun Jun Jun
## [22717] Jun Jun Jun Jun Jun Jun Jun Jun Jun Jun Jun Jun Jun Jun Jun Jun Jun Jun Jun Jun Jun
## [22735] Jul Jul Jul Jul Jul Jul Jul Jul Jul Jul Jul Jul Jul Jul Jul Jul Jul Jul Jul Jul Jul
## [22753] Jul Jul Jul Jul Jul Jul Jul Jul Jul Jul Jul Jul Jul Jul Jul Jul Jul Jul Jul Jul Jul
## [22771] Jul Jul Jul Jul Jul Jul Jul Jul Jul Jul Jul Jul Jul Jul Jul Jul Jul Jul Aug Aug Aug Aug
## [22789] Aug Aug Aug Aug Aug Aug Aug Aug Aug Aug Aug Aug Aug Aug Aug Aug Aug Aug Aug Aug Aug
## [22807] Aug Aug Aug Aug Aug Aug Aug Aug Aug Aug Aug Aug Aug Aug Aug Aug Aug Aug Aug Aug Aug
## [22825] Aug Aug Aug Aug Aug Aug Aug Aug Aug Aug Aug Aug Aug Aug Aug Aug Aug Aug Sep Sep May
## [22843] May May May May May May May May May May Jun Jun Jun Jun Jun Jun Jun Jun Jun Jun
## [22861] Jun Jun Jun Jun Jun Jun Jun Jun Jun Jun Jun Jun Jun Jun Jun Jun Jun Jun Jun Jun Jun
## [22879] Jun Jun Jun Jun Jun Jun Jun Jun Jun Jun Jun Jun Jun Jun Jun Jun Jun Jun Jun Jun Jun
## [22897] Jun Jun Jun Jun Jun Jun Jun Jun Jun Jun Jun Jun Jul Jul Jul Jul Jul Jul Jul Jul Jul
## [22915] Jul Jul Jul Jul Jul Jul Jul Jul Jul Jul Jul Jul Jul Jul Jul Jul Jul Jul Jul Jul Jul
## [22933] Jul Jul Jul Jul Jul Jul Jul Jul Jul Jul Jul Jul Jul Jul Jul Jul Jul Jul Jul Jul Jul
## [22951] Jul Jul Jul Jul Jul Jul Jul Jul Jul Jul Jul Aug Aug Aug Aug Aug Aug Aug Aug Aug Aug
## [22969] Aug Aug Aug Aug Aug Aug Aug Aug Aug Aug Aug Aug Aug Aug Aug Aug Aug Aug Aug Aug Aug
## [22987] Aug Aug Aug Aug Aug Aug Aug Aug Aug Aug Aug Aug Aug Aug Aug Aug Aug Aug Aug Aug Aug
## [23005] Aug Aug Aug Aug
## Levels: Jan Feb Mar Apr May Jun Jul Aug Sep Oct Nov Dec
```

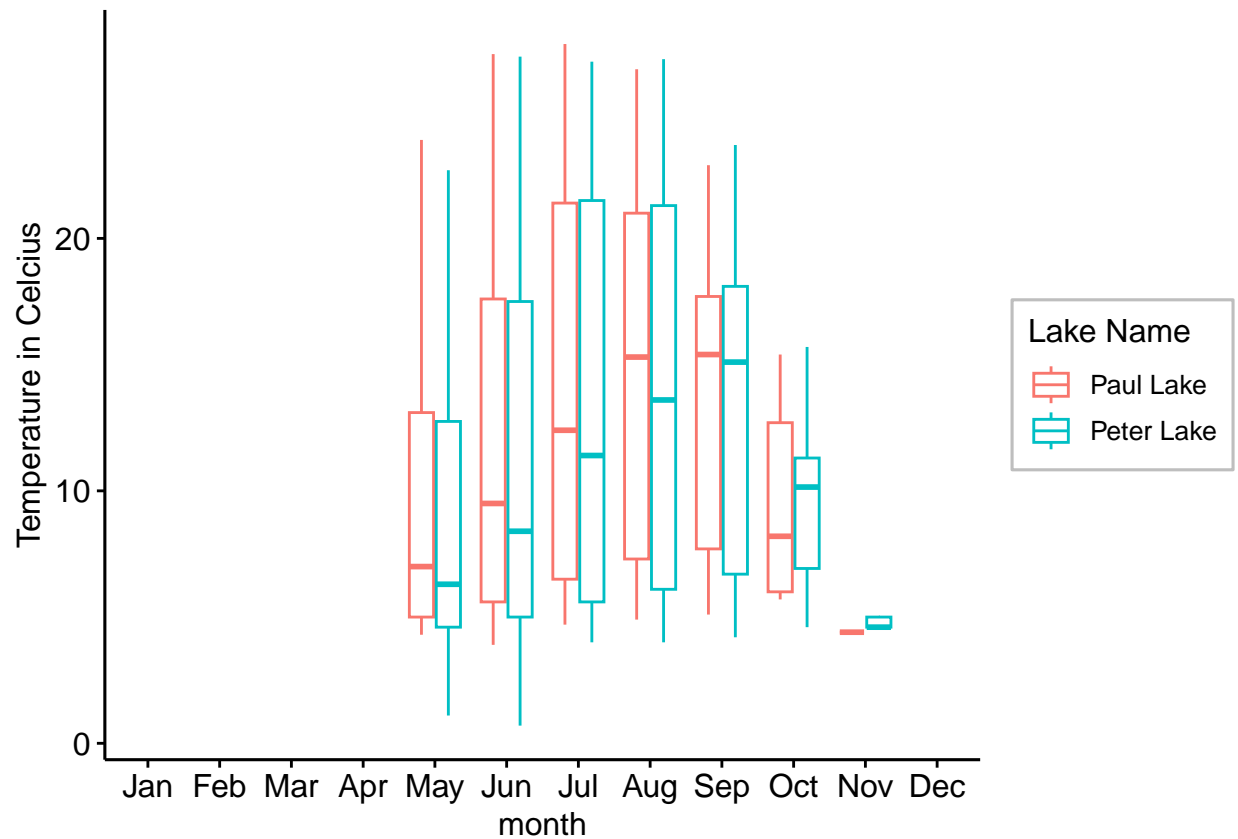
```
# Making sure month shows up the way I want in a graph
ggplot(PeterPaul.chem.nutrients, aes(x = factor(month, levels = 1:12, labels = month.abb))) +
  geom_bar() + scale_x_discrete(name = "month", drop = FALSE)
```



```
# Plotting each of the three box plots
```

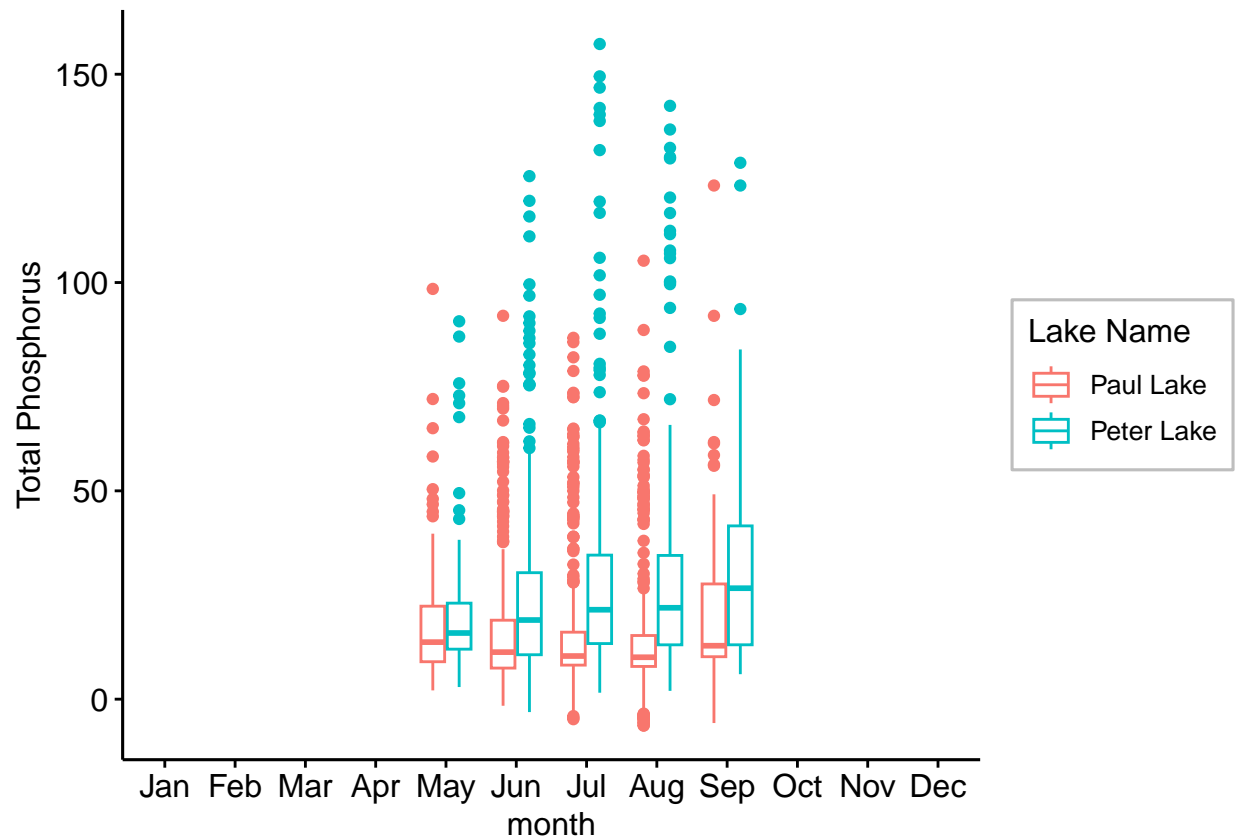
```
PeterPaul.temp <- ggplot(PeterPaul.chem.nutrients, aes(x = month, y = temperature_C)) +  
  geom_boxplot(aes(x = factor(month, levels = 1:12, labels = month.abb), color = lakename)) +  
  scale_x_discrete(name = "month", drop = FALSE) + labs(x = "Month", y = "Temperature in Celcius",  
    color = "Lake Name") + mytheme  
print(PeterPaul.temp)
```

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



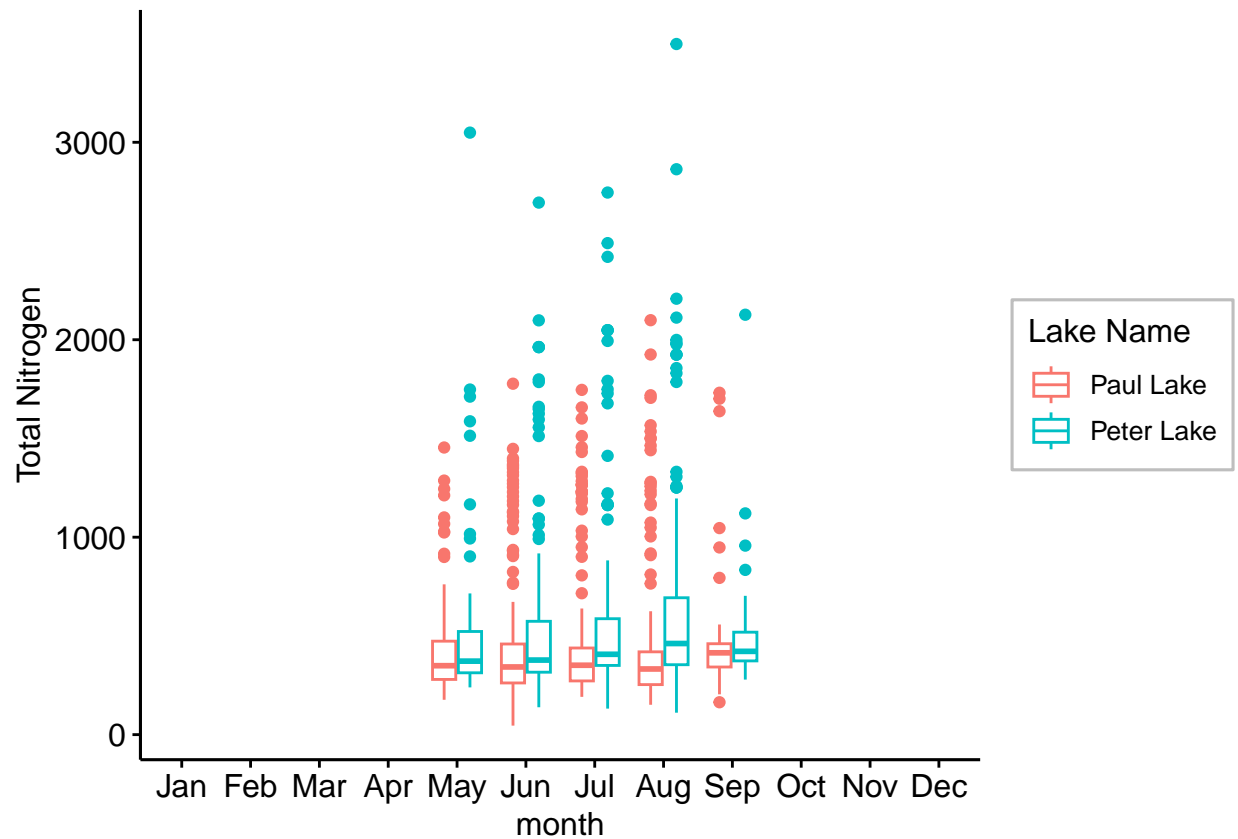
```
PeterPaul.TP <- ggplot(PeterPaul.chem.nutrients, aes(x = month, y = tp_ug)) + geom_boxplot(aes(x = factor(
  levels = 1:12, labels = month.abb), color = lakename)) + scale_x_discrete(name = "month",
  drop = FALSE) + labs(x = "Month", y = "Total Phosphorus", color = "Lake Name") +
  mytheme
print(PeterPaul.TP)
```

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



```
PeterPaul.TN <- ggplot(PeterPaul.chem.nutrients, aes(x = month, y = tn_ug)) + geom_boxplot(aes(x = factor(
  levels = 1:12, labels = month.abb), color = lakename)) + scale_x_discrete(name = "month",
  drop = FALSE) + labs(x = "Month", y = "Total Nitrogen", color = "Lake Name") +
  mytheme
print(PeterPaul.TN)
```

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

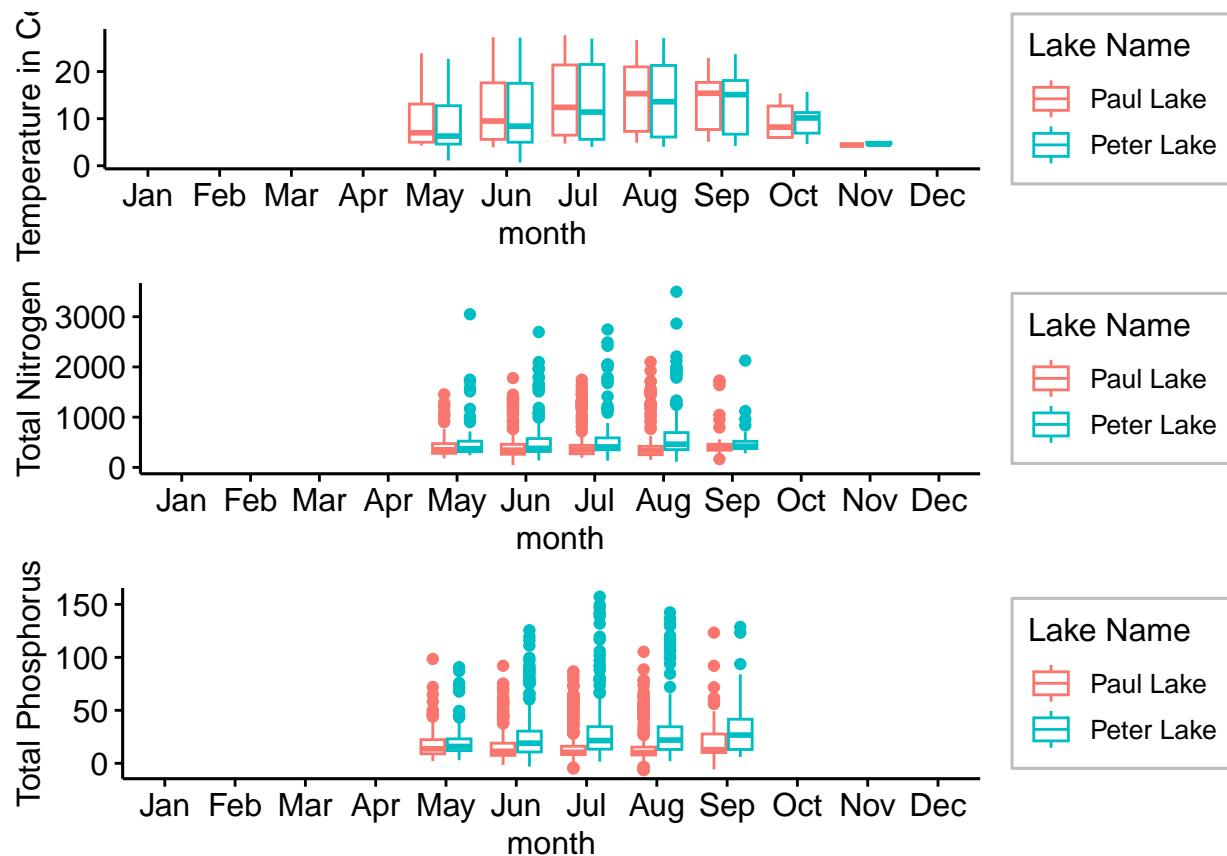


```
# Putting all three boxplots together
plot_grid(PeterPaul.temp, PeterPaul.TN, PeterPaul.TP, nrow = 3, align = "h", rel_heights = c(1.25,
  1.5, 1.5))
```

```
## 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: There appears to be more phosphorus and nitrogen during months with higher median temperatures, particularly July, August, and September.

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 Checking collectDate class and turning into factor
class(NEON.litter.mass$collectDate)
```

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

```
unique(NEON.litter.mass$collectDate)
```

```
## [1] "2016-06-16" "2016-07-14" "2016-08-11" "2016-09-08" "2016-10-07"
## [6] "2016-11-03" "2017-07-06" "2017-07-07" "2017-08-03" "2017-08-31"
## [11] "2017-10-12" "2017-11-20" "2018-06-07" "2018-07-05" "2018-08-02"
## [16] "2018-08-30" "2018-09-27" "2018-10-25" "2018-11-21" "2019-06-05"
## [21] "2019-07-02" "2019-07-31" "2019-08-28" "2019-09-25"
```

```
as.factor(NEON.litter.mass$collectDate)
```

[illegible]

[illegible]

[illegible]

[illegible]

[illegible]

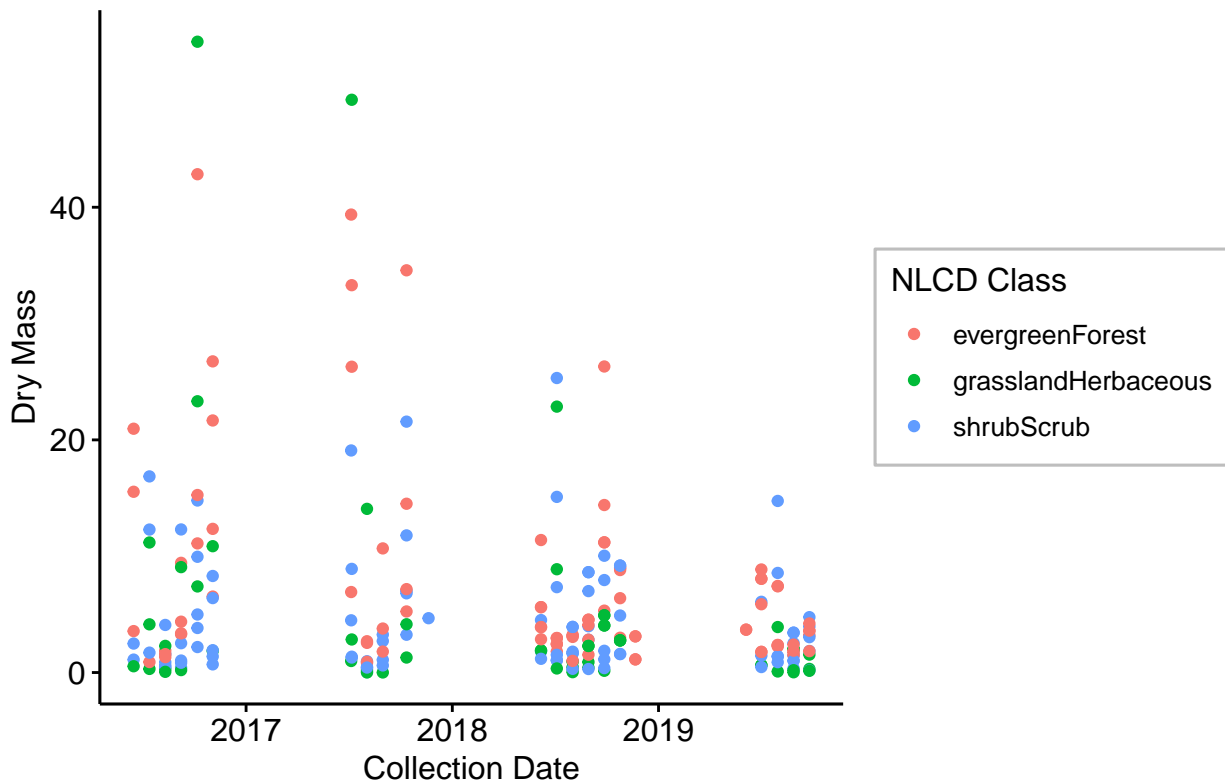
```
## [1603] 2019-09-25 2019-09-25 2019-09-25 2019-09-25 2019-09-25 2019-09-25 2019-09-25
## [1609] 2019-09-25 2019-09-25 2019-09-25 2019-09-25 2019-09-25 2019-09-25 2019-09-25
## [1615] 2019-09-25 2019-09-25 2019-09-25 2019-09-25 2019-09-25 2019-09-25 2019-09-25
## [1621] 2019-09-25 2019-09-25 2019-09-25 2019-09-25 2019-09-25 2019-09-25 2019-09-25
## [1627] 2019-09-25 2019-09-25 2019-09-25 2019-09-25 2019-09-25 2019-09-25 2019-09-25
## [1633] 2019-09-25 2019-09-25 2019-09-25 2019-09-25 2019-09-25 2019-09-25 2019-09-25
## [1639] 2019-09-25 2019-09-25 2019-09-25 2019-09-25 2019-09-25 2019-09-25 2019-09-25
## [1645] 2019-09-25 2019-09-25 2019-09-25 2019-09-25 2019-09-25 2019-09-25 2019-09-25
## [1651] 2019-09-25 2019-09-25 2019-09-25 2019-09-25 2019-09-25 2019-09-25 2019-09-25
## [1657] 2019-09-25 2019-09-25 2019-09-25 2019-09-25 2019-09-25 2019-09-25 2019-09-25
## [1663] 2019-09-25 2019-09-25 2019-09-25 2019-09-25 2019-09-25 2019-09-25 2019-09-25
## [1669] 2019-09-25 2019-09-25 2019-09-25 2019-09-25 2019-09-25 2019-09-25 2019-09-25
## [1675] 2019-09-25 2019-09-25 2019-09-25 2019-09-25 2019-09-25 2019-09-25 2019-09-25
## [1681] 2019-09-25 2019-09-25 2019-09-25 2019-09-25 2019-09-25 2019-09-25 2019-09-25
## [1687] 2019-09-25 2019-09-25 2019-09-25 2019-09-25 2019-09-25 2019-09-25 2019-09-25
## 24 Levels: 2016-06-16 2016-07-14 2016-08-11 2016-09-08 ... 2019-09-25
```

```
# Generating subset plot
```

```
Litter.subset <- ggplot(subset(NEON.litter.mass, functionalGroup == "Needles"), aes(x = collectDate,
  y = dryMass, color = nlcdClass)) + geom_point() + labs(title = "Needle Dry Mass Across Dates and CL
  y = "Dry Mass", x = "Collection Date", color = "NLCD Class") + mytheme
```

```
print(Litter.subset)
```

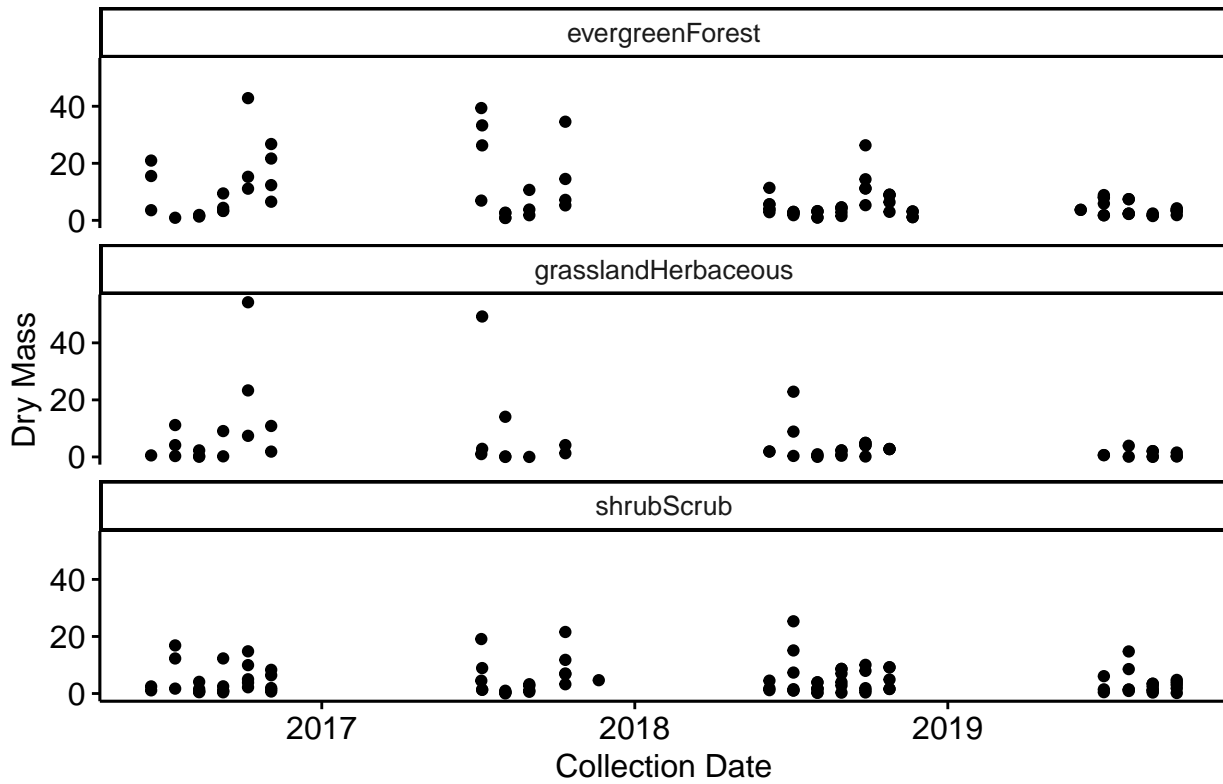
Needle Dry Mass Across Dates and Class



```
# 7
Litter.subset.facet <- ggplot(subset(NEON.litter.mass, functionalGroup == "Needles"),
  aes(x = collectDate, y = dryMass)) + geom_point() + labs(title = "Needle Dry Mass Across Dates and (",
  y = "Dry Mass", x = "Collection Date") + mytheme + facet_wrap(vars(nlcdClass),
  nrow = 3)

print(Litter.subset.facet)
```

Needle Dry Mass Across Dates and Class



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

Answer: I think displaying the data by facet makes it easier to visually see how needle dry mass varied by class across the collection dates. When the data is separated by only color, all the points are on one graph which makes it difficult to discern the individual differences between each class.