Assignment 5: Data Visualization

Amanda Booth

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.

```
knitr::opts_chunk$set(echo = TRUE)
#1

#Check working directory
getwd()
```

```
## [1] "/Users/amandabooth/Documents/GitHub/Environmental_Data_Analytics_2022"
```

```
#Load packages
library(tidyverse)
library(cowplot)

#Upload files
NTL <- read.csv("./Data/Processed/NTL-LTER_Lake_Chemistry_Nutrients_PeterPaul_Processed.csv")
NEON <- read.csv("./Data/Processed/NEON_NIWO_Litter_mass_trap_Processed.csv")
```

```
#2

#Check date class
class(NTL$sampledate)
```

```
## [1] "character"
```

```
class(NEON$collectDate)
```

```
## [1] "character"
```

```
#Change date class

NTL$sampledate <- as.Date(NTL$sampledate, format = "%Y-%m-%d")
NEON$collectDate <- as.Date(NEON$collectDate, format = "%Y-%m-%d")
```

Define your theme

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

```
#3

#Build theme
mytheme <- theme_linedraw(base_size = 10) +
  theme(axis.text = element_text(color = "grey"),
        legend.position = "bottom")

#Set as default theme
theme_set(mytheme)
```

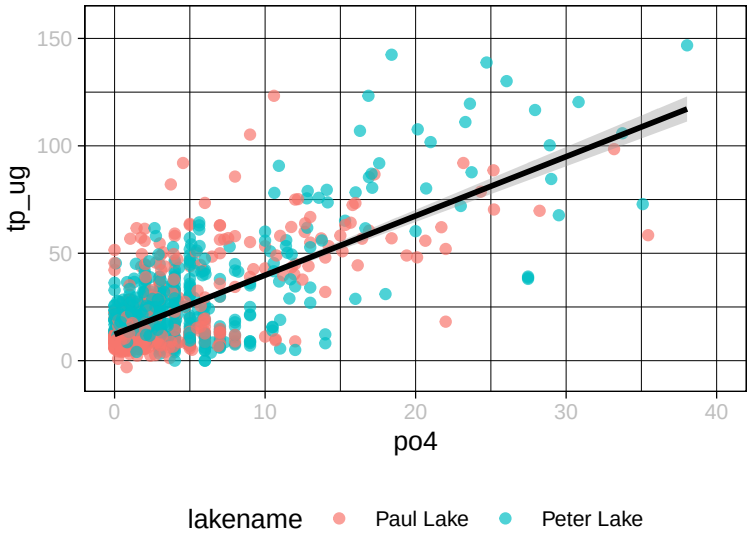
Create graphs

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()`).

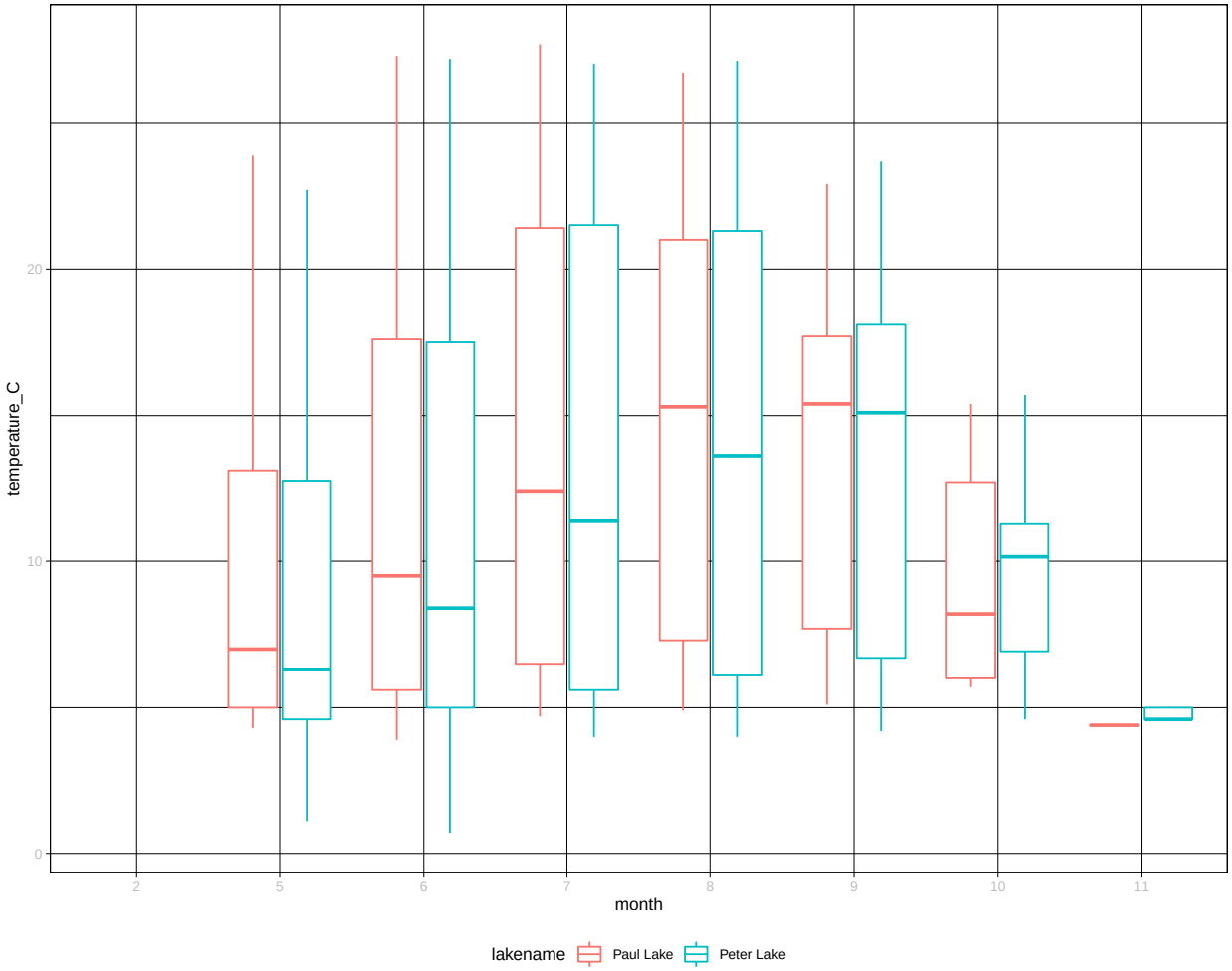
```
#4

Plot1 <- ggplot(NTL, aes(x = po4, y = tp_ug)) +
  geom_point(aes(color = lakename), alpha = 0.7) +
  xlim(0, 40) +
   geom_smooth(method='lm', color = "black")
print(Plot1)
```

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

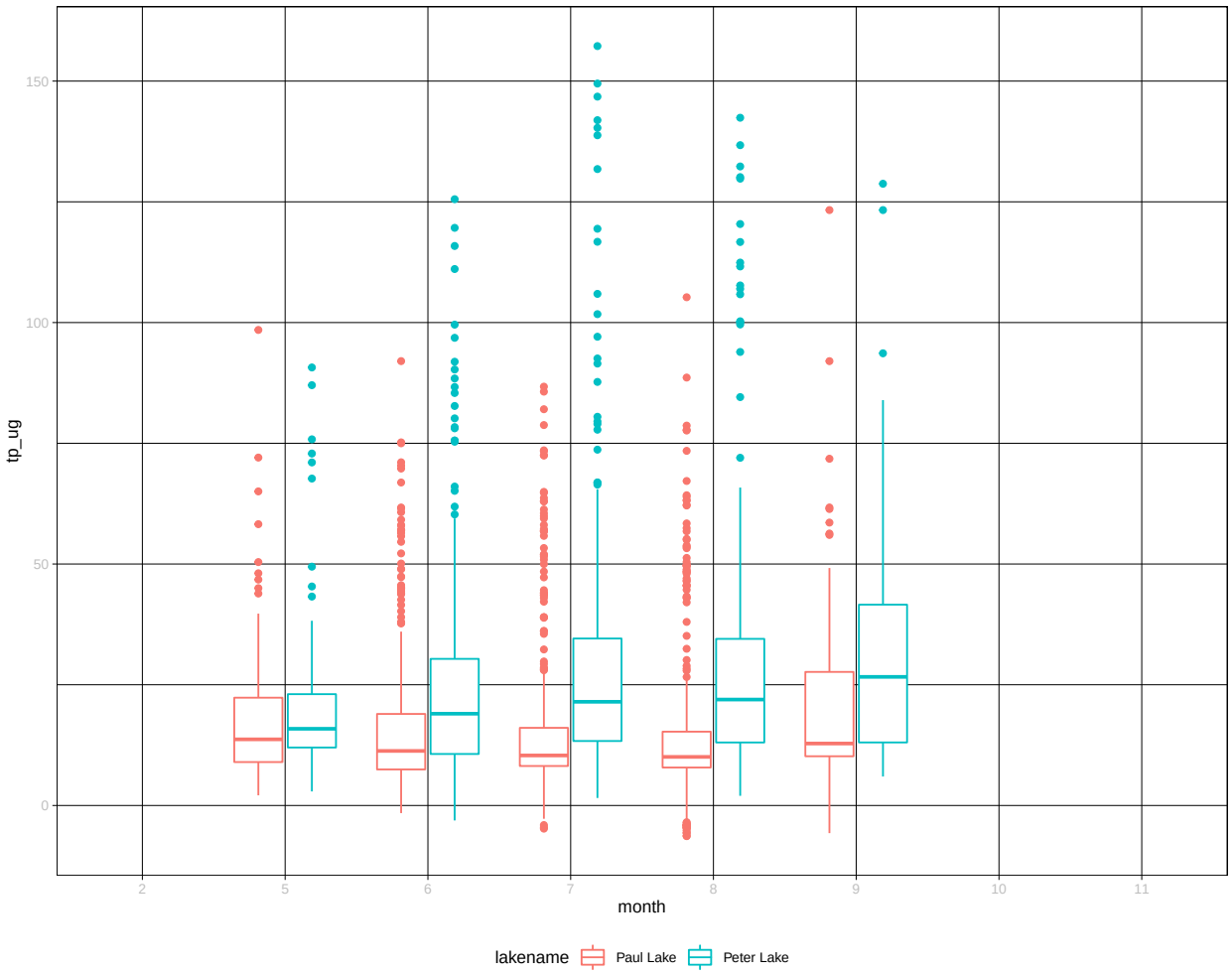


5. [NTL-LTER] Make three separate boxplots of (a) temperature, (b) TP, and (c) TN, with month as the x axis and lake as a color aesthetic. Then, create a cowplot that combines the three graphs. Make sure that only one legend is present and that graph axes are aligned.

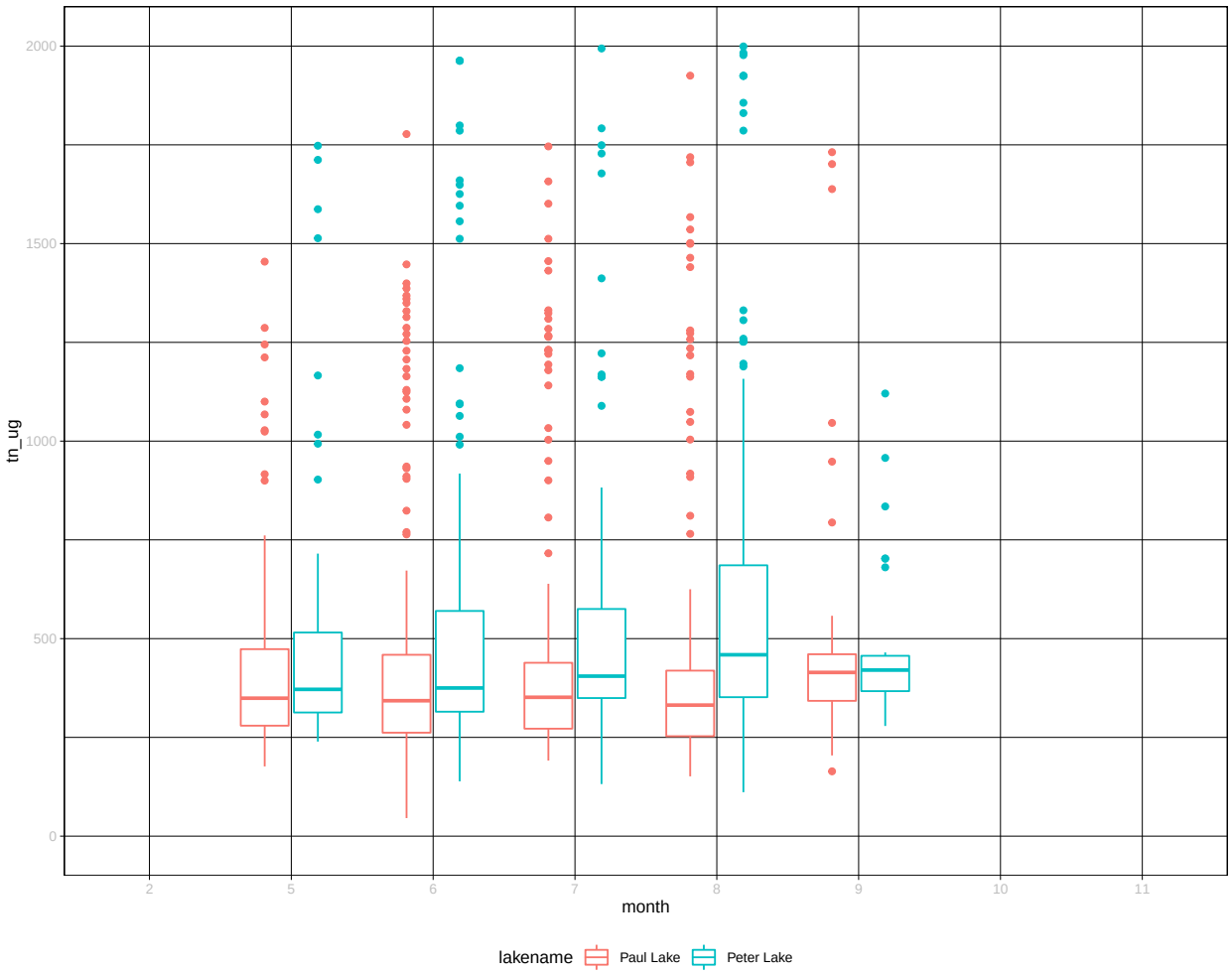
```
#5


#Boxplots

NTL$month <- as.factor(NTL$month)

Boxplot1 <- ggplot(NTL) +
  geom_boxplot(aes(x = month, y = temperature_C, color = lakename))
print(Boxplot1)
```



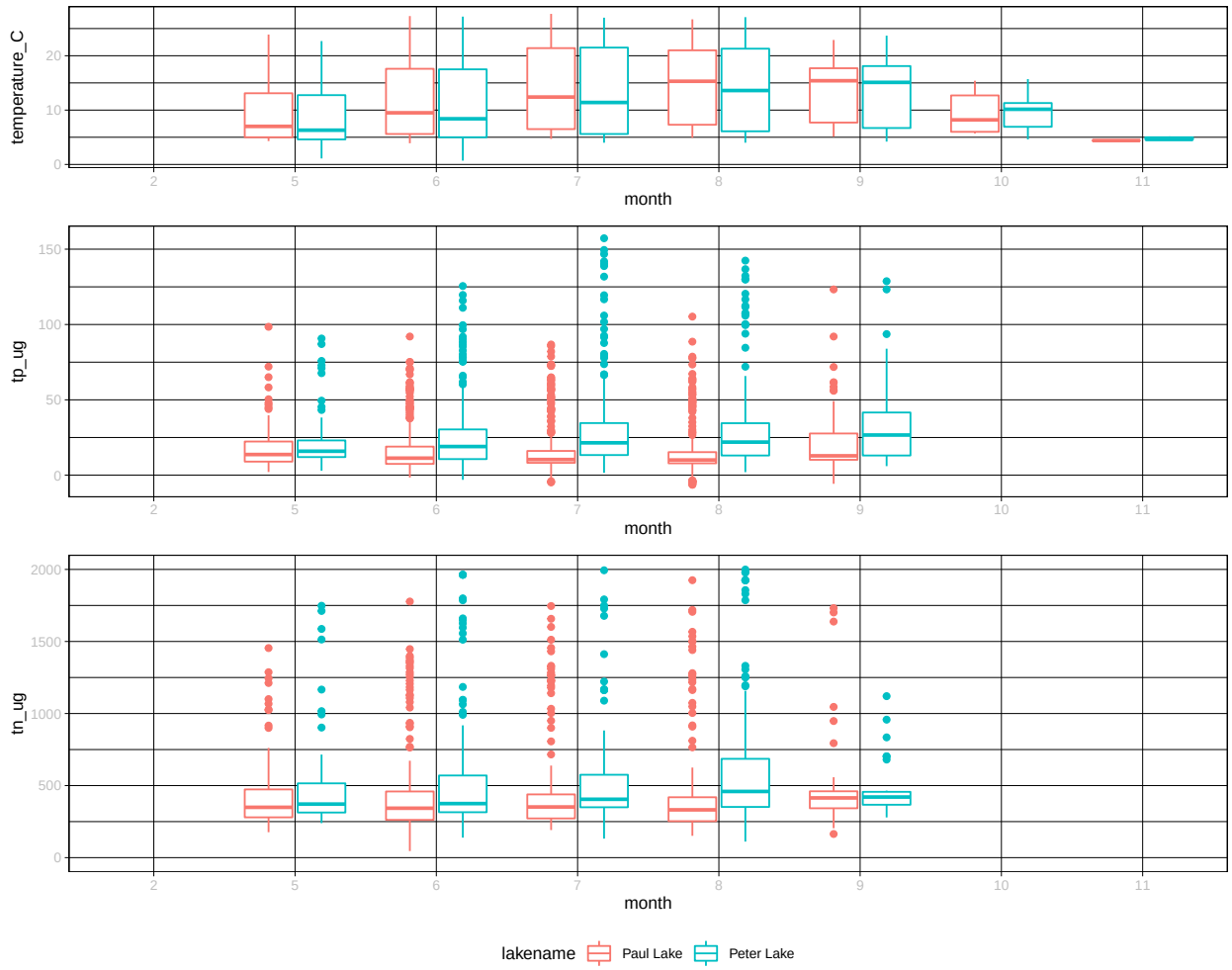

```
Boxplot2 <- ggplot(NTL) +
  geom_boxplot(aes(x = month, y = tp_ug, color = lakename))
print(Boxplot2)
```




```
Boxplot3 <- ggplot(NTL) +
  geom_boxplot(aes(x = month, y = tn_ug, color = lakename)) +
  ylim(0, 2000)
print(Boxplot3)
```




```
#Boxplots combined

Combined_boxplots <- plot_grid(Boxplot1 + theme(legend.position="none"), Boxplot2 + theme(legend.positi
print(Combined_boxplots)
```



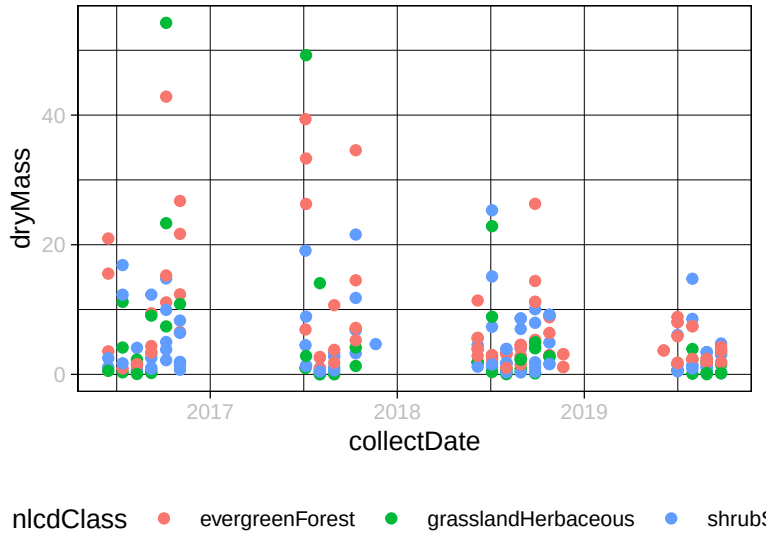
Question: What do you observe about the variables of interest over seasons and between lakes?

> Answer: The lakes tend to be fairly similar temperatures in any given month. The temperature increases for both lakes in the spring, summer, and early fall months. The tp_ug and tn_ug in the lakes are fairly close in May, but during the summer months there is an increase in Peter Lake. In September we can observe Paul Lake has increased tp_ug and tn_ug, lessening the gap between the two lakes once again.

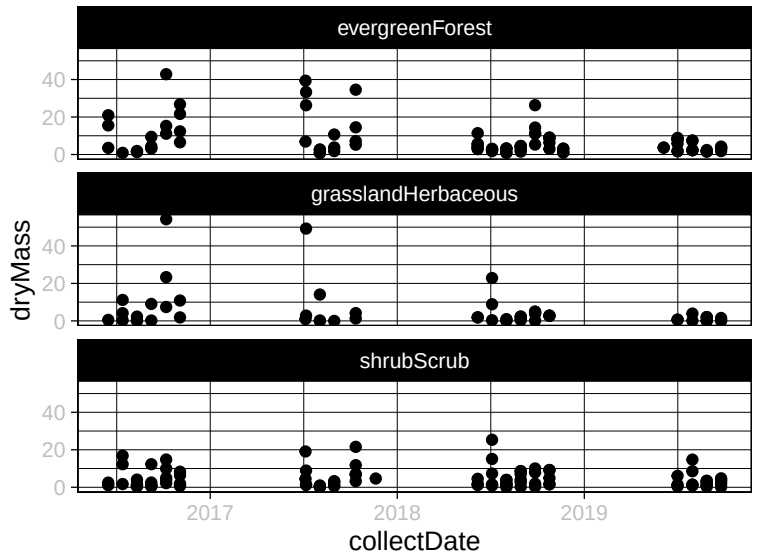
6. [Niwot Ridge] Plot a subset of the litter dataset by displaying only the "Needles" functional group. Plot the dry mass of needle litter by date and separate by NLCD class with a color aesthetic. (no need to adjust the name of each land use)

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

```
#6

NEON_plot1 <- ggplot(filter(NEON, functionalGroup == "Needles")) +
  geom_point(aes(x = collectDate, y = dryMass, color = nlcdClass))
print(NEON_plot1)
```




```
#7

NEON_plot2 <- ggplot(filter(NEON, functionalGroup == "Needles")) +
  geom_point(aes(x = collectDate, y = dryMass)) +
  facet_wrap(vars(nlcdClass), nrow = 3)
print(NEON_plot2)
```



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

Answer:Plot 7 is more effective, because isolating the NLCD classes makes it easier to see each of their dry mass trends over time. Plot 6, while distinguishing them by color, really only does a good job of showing general dry mass trends regardless of NLCD class.