Assignment 5: Data Visualization

Curtis Cha

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 1.3.1 --
```

```
## v ggplot2 3.3.5     v purrr   0.3.4
## v tibble  3.1.6     v dplyr   1.0.7
## v tidyr   1.1.3     v stringr 1.4.0
## v readr   2.0.1     v forcats 0.5.1
```

```
## -- 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
```

```
setwd("../Data")
PeterPaul <- read.csv(
  "Processed/NTL-LTER_Lake_Chemistry_Nutrients_PeterPaul_Processed.csv",
  header = T, stringsAsFactors = T)
NIWO <- read.csv(
  "Processed/NEON_NIWO_Litter_mass_trap_Processed.csv",
  header = T, stringsAsFactors = T)

#2
NIWO$collectDate <- ymd(NIWO$collectDate)
PeterPaul$sampledate <- ymd(PeterPaul$sampledate)
class(NIWO$collectDate)
```

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

```
class(PeterPaul$sampledate)
```

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

Define your theme

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

```
#3

mytheme <- theme_classic(base_size = 13) +
  theme(axis.text = element_text(color = "blue"),
        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 `ylim()`).

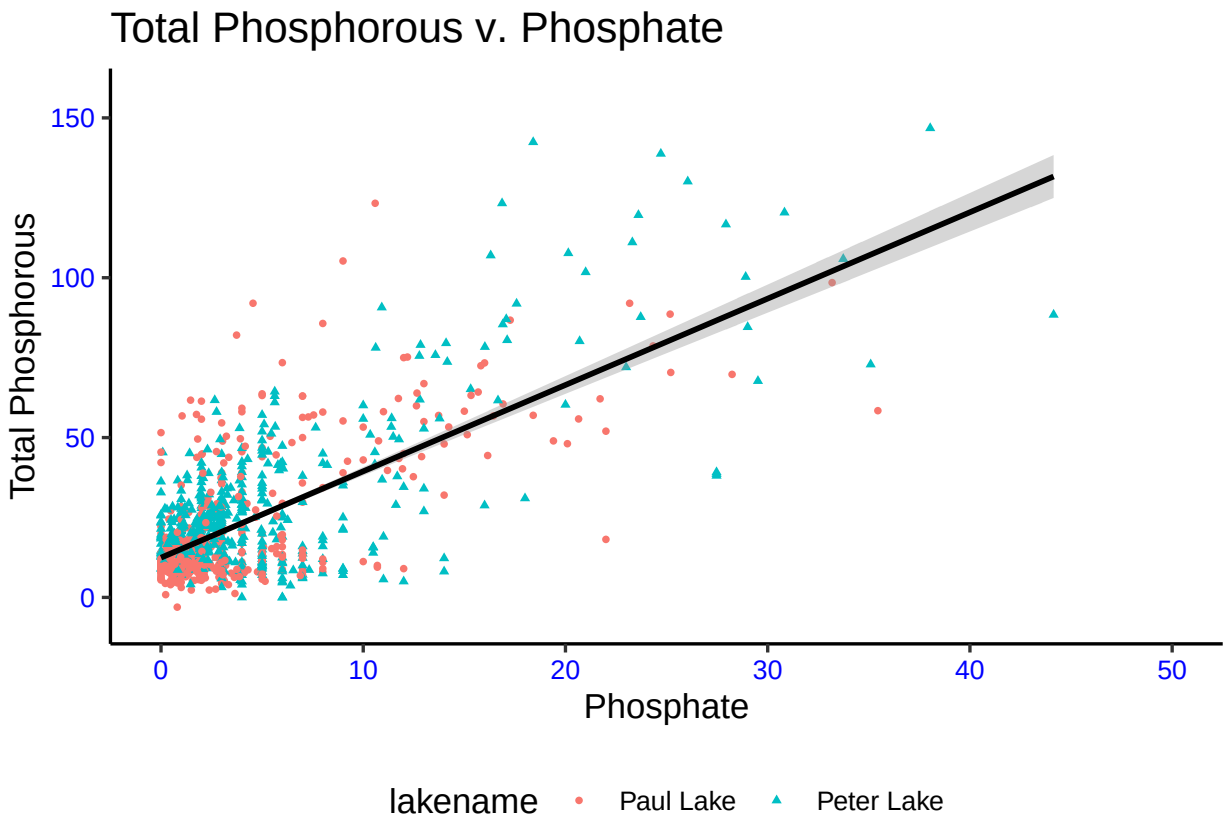
```
#4

ggplot(data = PeterPaul, aes(x = po4, y = tp_ug)) +
  geom_point(size = 1, aes(color = lakename, shape = lakename)) +
  geom_smooth(method = 'lm', color = 'black') +
  xlim(c(0,50)) +
  labs(y = "Total Phosphorous", x = "Phosphate",
       title = "Total Phosphorous v. Phosphate")
```

```
## `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
T_C <- ggplot(data = PeterPaul,
              aes(y = temperature_C, x = month(sampledate), color = lakename)) +
  geom_point() +
  labs(y = "Temperature(C)", x = "Month", color = "Lake")

P_ug <- ggplot(data = PeterPaul,
               aes(y = tp_ug, x = month(sampledate), color = lakename)) +
  geom_point() +
  labs(y = "Total Phosphorous(ug)", x = "Month", color = "Lake")

N_ug <- ggplot(data = PeterPaul,
               aes(y = tn_ug, x = month(sampledate), color = lakename)) +
  geom_point() +
  labs(y = "Total Nitrogen(ug)", x = "Month", color = "Lake")

library(cowplot)
```

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

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

```
##     stamp
```

```
grid <- plot_grid(T_C + theme(legend.position = 'none'),
                  P_ug + theme(legend.position = 'none'),
                  N_ug + theme(legend.position = 'none'),
                  nrow = 1, align = 'h', rel_heights = c(1.25, 1))
```

```
## Warning: Removed 3566 rows containing missing values (geom_point).
```

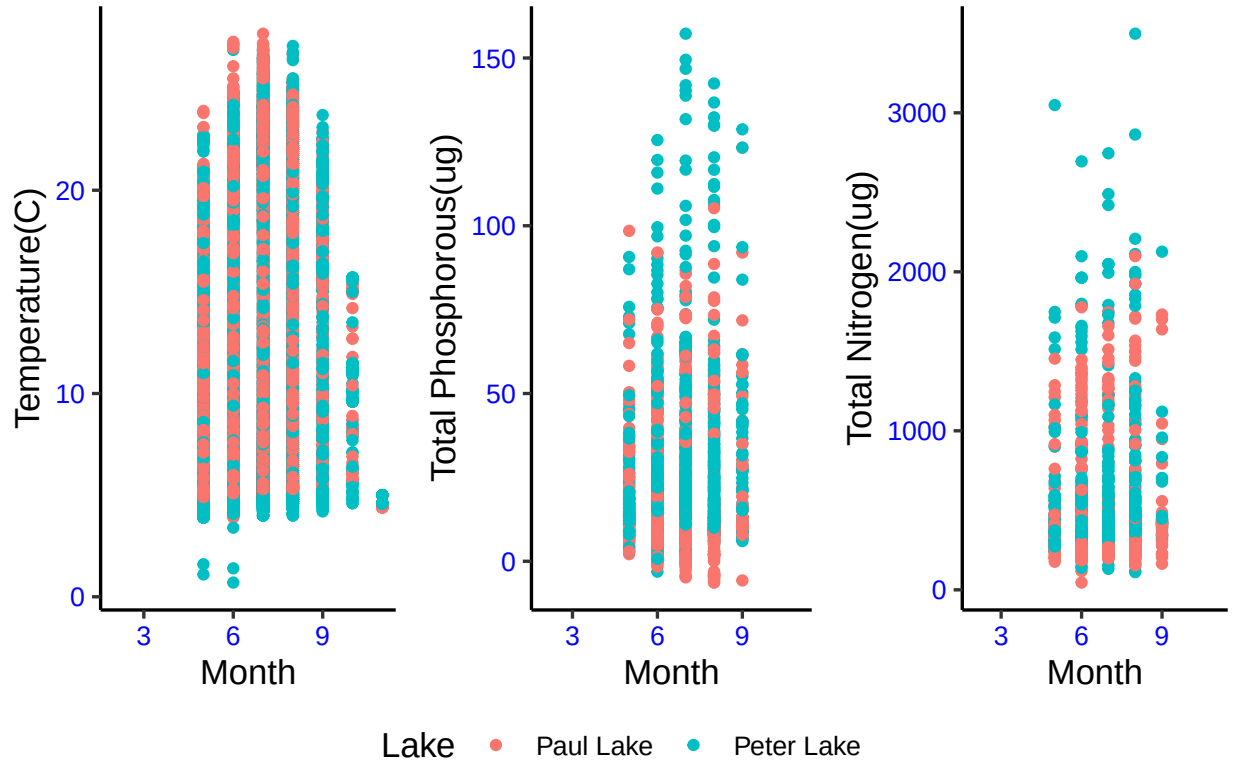
```
## Warning: Removed 20729 rows containing missing values (geom_point).
```

```
## Warning: Removed 21583 rows containing missing values (geom_point).
```

```
title <- ggdraw() +
  draw_label("Variables by Month",
    fontface = 'bold', x = 0, hjust = 0)

plot_grid(title, grid, get_legend(T_C), nrow =3, rel_heights = c(.1, 1, .1))
```

```
## Warning: Removed 3566 rows containing missing values (geom_point).
```



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

> Answer: From the combined plots of Temperature, Phosphorous and Nitrogen, I observe a peak in all three variables during the months between month 6 and 8. Since these months tend to be the summer months, the rise in temperature is associated with the season. Of the two lakes, Peter Lake has higher concentrations of Nitrogen and Phosphorous. However, there is no difference observed in temperature betweeen the two lakes throughout the months.

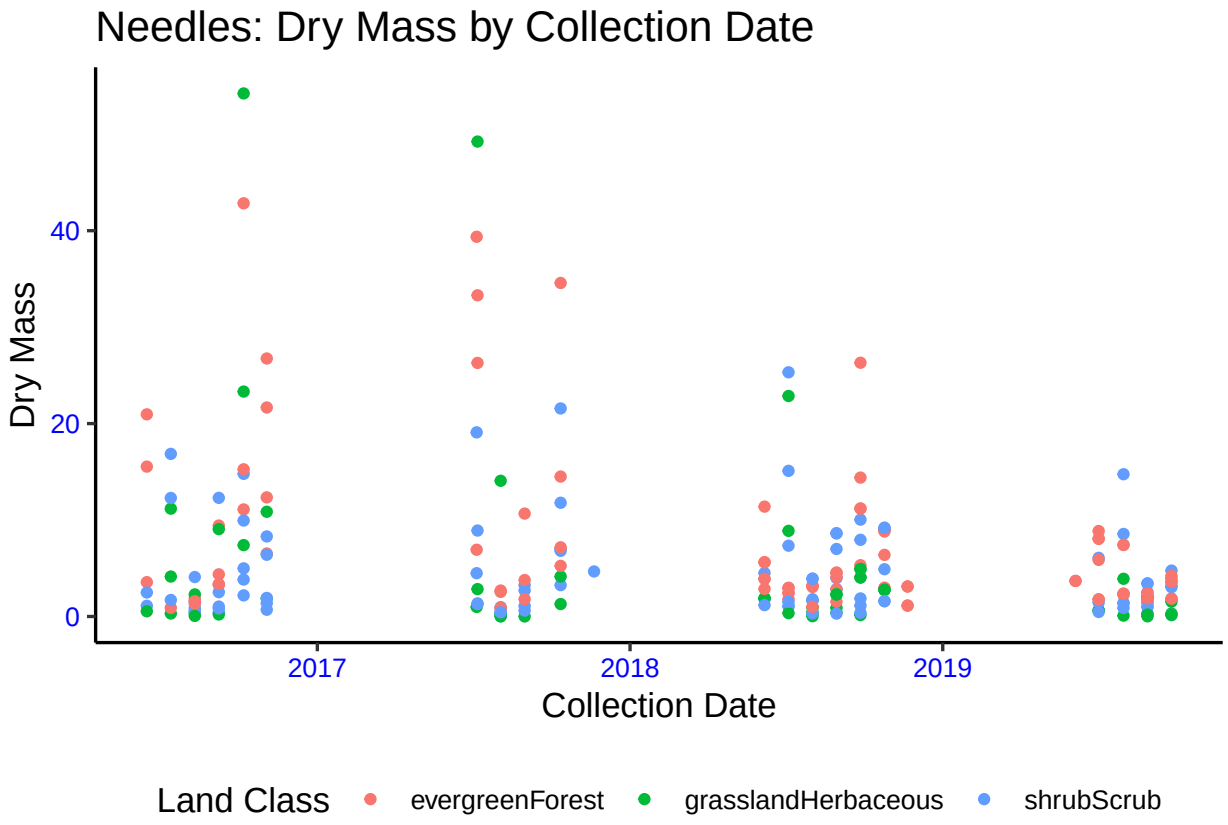
6. [Niwot Ridge] Plot a subset of the litter dataset by displaying only the "Needles" functional group. Plot

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.

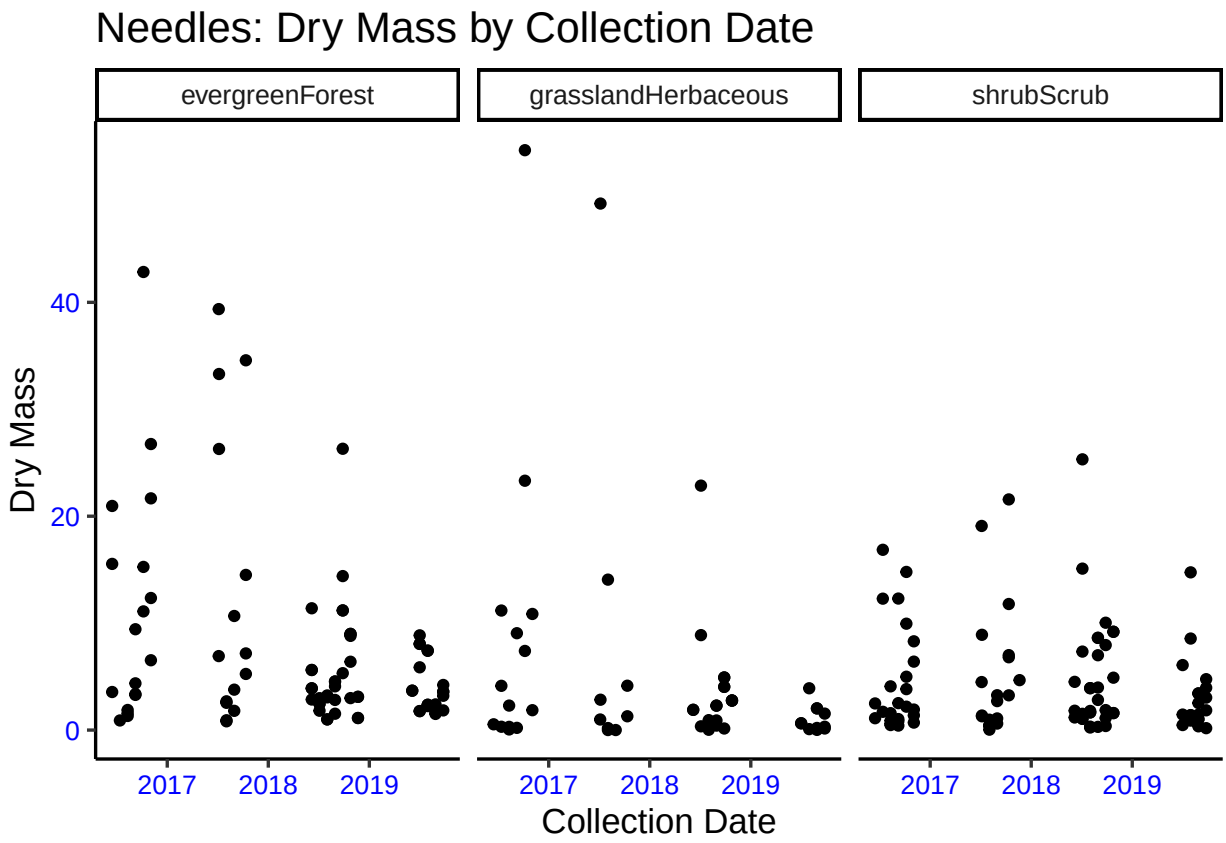
```
#6

ggplot(data = filter(NIWO, functionalGroup == "Needles"),
       aes(y = dryMass, x = collectDate)) +
  geom_point(aes(color = nlcdClass)) +
  labs(x = "Collection Date", y = "Dry Mass", color = "Land Class",
       title = "Needles: Dry Mass by Collection Date")
```




```
#7

ggplot(data = filter(NIWO, functionalGroup == "Needles"),
       aes(y = dryMass, x = collectDate)) +
  geom_point() +
  facet_wrap(vars(nlcdClass)) +
  labs(x = "Collection Date", y = "Dry Mass", color = "Land Class",
       title = "Needles: Dry Mass by Collection Date")
```



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

Answer: The plot for number 7 is useful because we can better compare the drymass between Land Classes than the plot in number 6. The plot in number 6, the dry mass points are overlaying each other and difficult to discern the differences between the three. This is due to similar collection dates for the three land classes. The plot in number 7 separates the land class' dry mass so they can be more easily compared.