Assignment 5: Data Visualization

Jessica Citrola

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

getwd()
```

```
## [1] "/Users/jessicacitrola/Documents/ENV872/Environmental_Data_Analytics_2022/Assignments"
```

```
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   1.1.4     v stringr 1.4.0
## v readr   2.0.2     v forcats 0.5.1
```

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

```
library(cowplot)
library(lubridate)
```

```
##
## Attaching package: 'lubridate'

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

## The following objects are masked from 'package:base':
##
##     date, intersect, setdiff, union
```

```
Lake.Chemistry.Nutrient <- read.csv("../Data/Processed/NTL-LTER_Lake_Chemistry_Nutrients_PeterPaul_Proc

Niwot.Ridge.Litter <- read.csv("../Data/Processed/NEON_NIWO_Litter_mass_trap_Processed.csv", stringsAsFa

#2

class(Lake.Chemistry.Nutrient$sampledate)
```

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

```
Lake.Chemistry.Nutrient$sampledate <- ymd(Lake.Chemistry.Nutrient$sampledate)
class(Lake.Chemistry.Nutrient$sampledate)
```

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

```
class(Niwot.Ridge.Litter$collectDate)
```

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

```
Niwot.Ridge.Litter$collectDate <- ymd(Niwot.Ridge.Litter$collectDate)
class(Niwot.Ridge.Litter$collectDate)
```

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

Define your theme

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

```r
#3

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

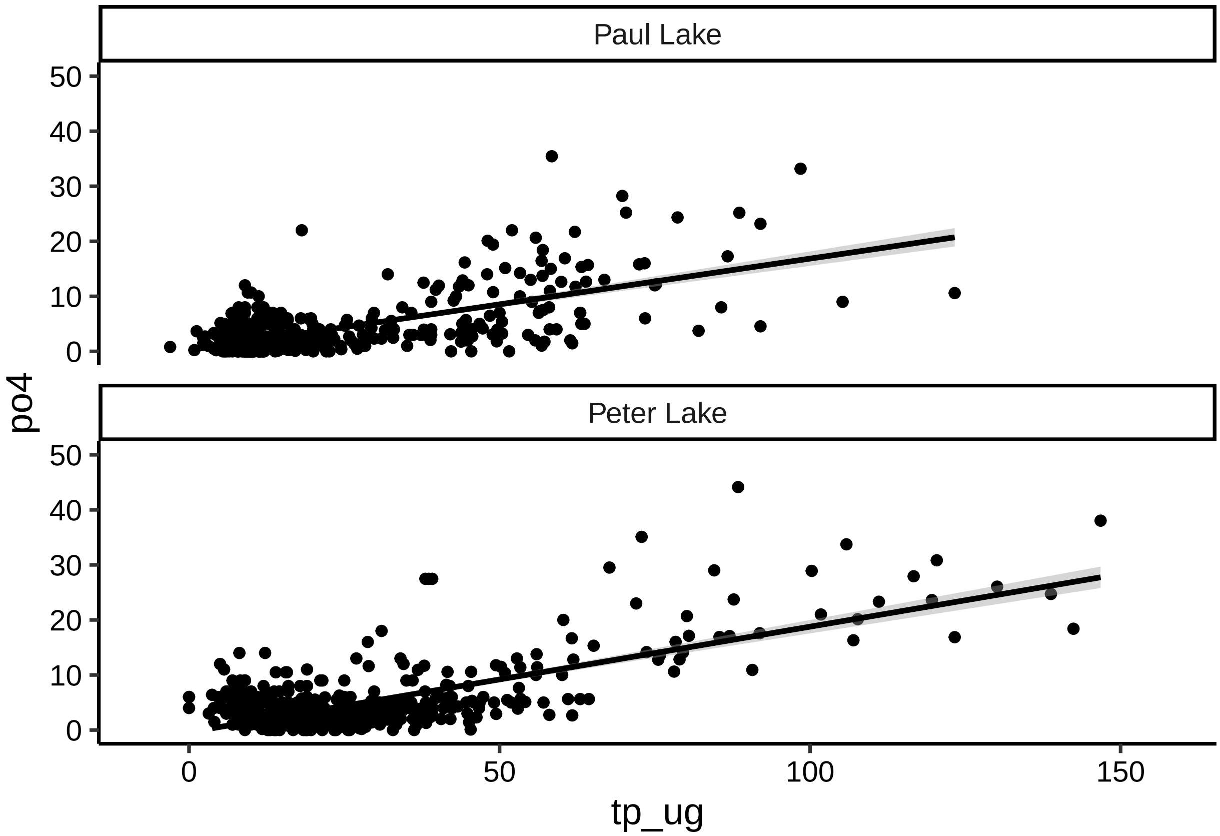
Create graphs

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()`).

```r
#4

Lake.Nutrient.Plot <-   ggplot(Lake.Chemistry.Nutrient, aes(x = tp_ug, y = po4)) +
  mytheme +
  geom_point() +
  ylim(0, 50) +
  geom_smooth(method = lm, col = "black") +
  facet_wrap(vars(lakename), nrow = 3)
print(Lake.Nutrient.Plot)
```

```
## `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).

## 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.

```
#5
class(Lake.Chemistry.Nutrient$month)
```

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

```
Lake.Chemistry.Nutrient$month <- as.factor(Lake.Chemistry.Nutrient$month)
class(Lake.Chemistry.Nutrient$month)
```

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

```
Lake.Boxplot.T <-
  ggplot(Lake.Chemistry.Nutrient, aes(x = month, y = temperature_C, color = lakename)) +
  theme(legend.position = "none") +
  geom_boxplot()

Lake.Boxplot.TP <-
  ggplot(Lake.Chemistry.Nutrient, aes(x = month, y = tp_ug, color = lakename)) +
  theme(legend.position = "none") +
  geom_boxplot()

Lake.Boxplot.TN <-
```

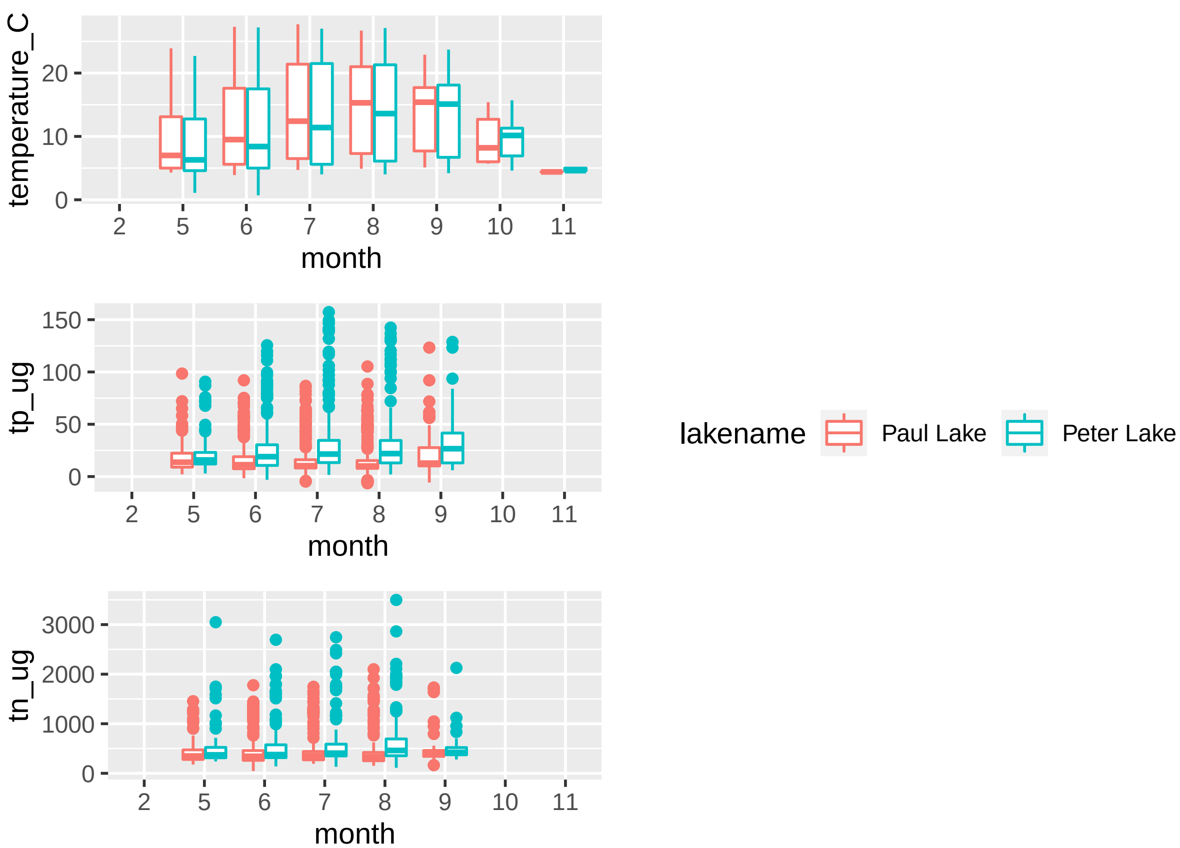
```
  ggplot(Lake.Chemistry.Nutrient, aes(x = month, y = tn_ug, color = lakename)) +
  theme(legend.position = "none") +
  geom_boxplot()

plot.grid <- plot_grid(Lake.Boxplot.T, Lake.Boxplot.TP, Lake.Boxplot.TN,
                       nrow = 3, align = 'h', rel_heights = c(3, 3, 3))
```

```
## 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).
```

```
legend <- get_legend(
  Lake.Boxplot.T +
    theme(legend.position = "bottom")
)
```

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

```
plot_grid(plot.grid, legend)
```

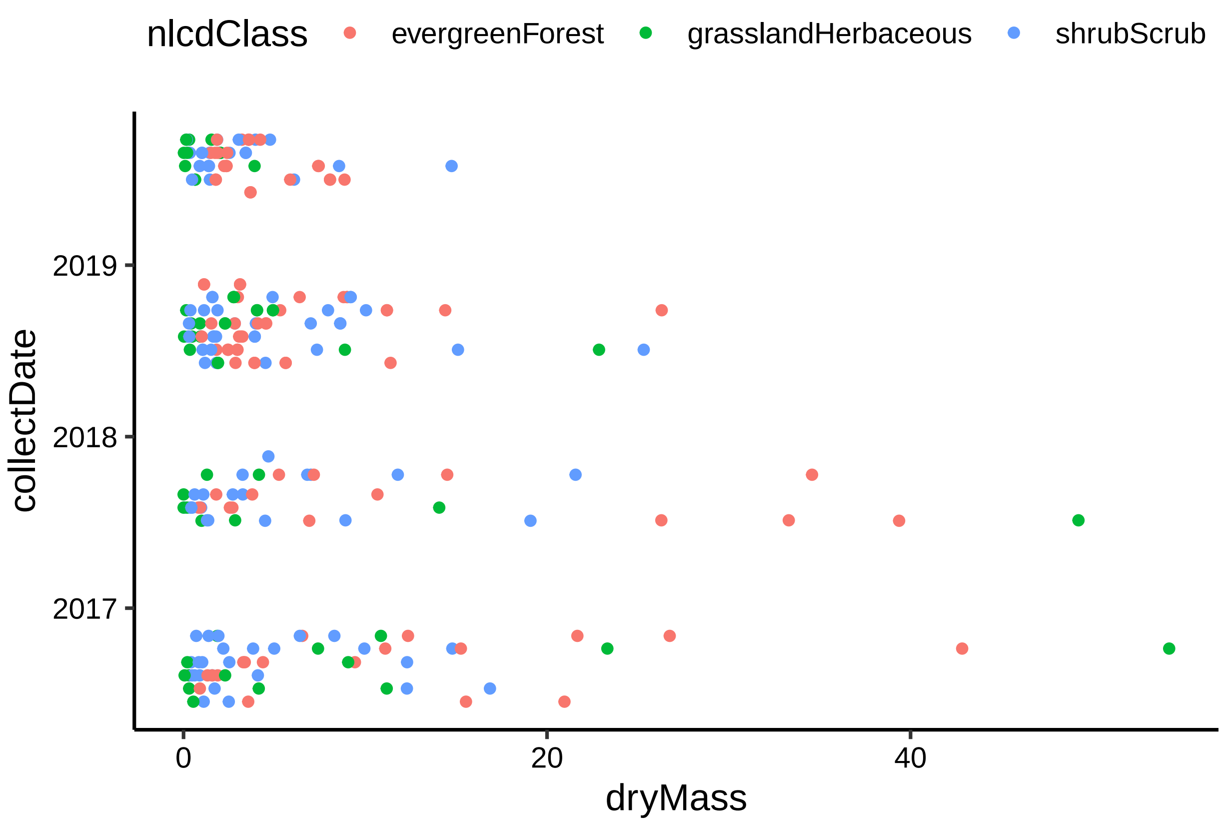


Question: What do you observe about the variables of interest over seasons and between lakes?

Answer: The temperature readings are relatively similar between both lakes, with higher average and peak temperatures in the later summer and the lowest temperatures in November. TP average for Peter lake is consistently higher than Paul lake throughout the months, with many high outliers observed in the warmer months. TP averges at Paul lake do not change significantly over the months, but outliers are still observed in each month. The TN averages also do not change significantly over the seasons. However, both lakes have many outlier readings throughout the warmer months with Peter lake showing the highest TN readings during the summer.

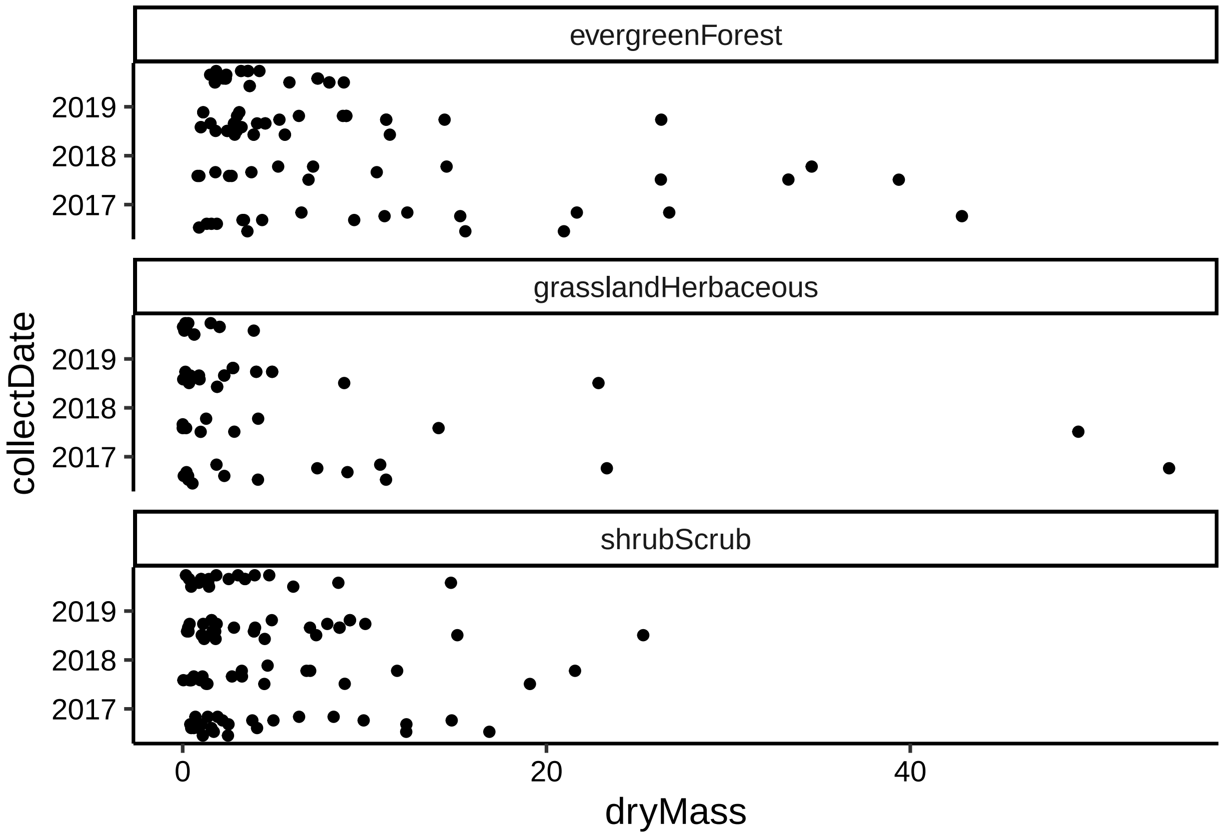
6. [Niwot Ridge] Plot a subset of the litter dataset by displaying only the "Needles" functional group. Plot the dry mass of needle litter by date and separate by NLCD class with a color aesthetic. (no need to adjust the name of each land use)
7. [Niwot Ridge] Now, plot the same plot but with NLCD classes separated into three facets rather than separated by color.

```
#6

Niwot.Plot <-
  ggplot(subset(Niwot.Ridge.Litter, functionalGroup == "Needles"),
  aes(y = collectDate, x = dryMass, color = nlcdClass)) +
  mytheme +
  geom_point()
print(Niwot.Plot)
```




```
#7

Niwot.Plot.Facet <-
```

```r
  ggplot(subset(Niwot.Ridge.Litter, functionalGroup == "Needles"),
  aes(y = collectDate, x = dryMass)) +
  geom_point() +
  mytheme +
  facet_wrap(vars(nlcdClass), nrow = 3)
print(Niwot.Plot.Facet)
```



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

> Answer: I believe that #7 is more effective because in #6, the data is more clumped together and harder to read when all of the nlcd classes are shown at the same time. In #7, I am able to get a better picture on the dry mass for each year and each type of nlcd class. The data in #7 appears more organized and easier to read.