Assignment 5: Data Visualization

Leah Li

Fall 2024

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. Read in 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 in the Processed_KEY folder) and the processed data file for the Niwot Ridge litter dataset (use the `NEON_NIWO_Litter_mass_trap_Processed.csv` version, again from the Processed_KEY folder).

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

```
#1
#Load packages
library(tidyverse)
library(lubridate)
library(here)
library(cowplot)
library(ggplot2)

#Check working directory
getwd()
```

```
## [1] "/home/guest/EDE_Fall2024"
```

```r
here()
```

```
## [1] "/home/guest/EDE_Fall2024"
```

```r
#2
#Load Data
Nutrients_PeterPaul.data <- read.csv(
  file = here("./Data/Processed_KEY/NTL-LTER_Lake_Chemistry_Nutrients_PeterPaul_Processed.csv"),
  stringsAsFactors = TRUE)

Litter_Mass_Trap.data <- read.csv(
  file = here("./Data/Processed_KEY/NEON_NIWO_Litter_mass_trap_Processed.csv"),
  stringsAsFactors = TRUE)
```

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

```r
#3
mytheme <- theme(
  plot.background = element_rect(fill = "lightblue", color = "black", size = 1),
  plot.title = element_text(face = "bold", size = 16, color = "darkblue", hjust = 0.5),
  axis.title = element_text(face = "italic", size = 14, color = "darkgreen"),
  axis.text = element_text(size = 12, color = "black"),
  axis.ticks = element_line(color = "red"),
  legend.title = element_text(face = "bold", color = "darkred"),
  legend.text = element_text(size = 12)
)
```

Create graphs

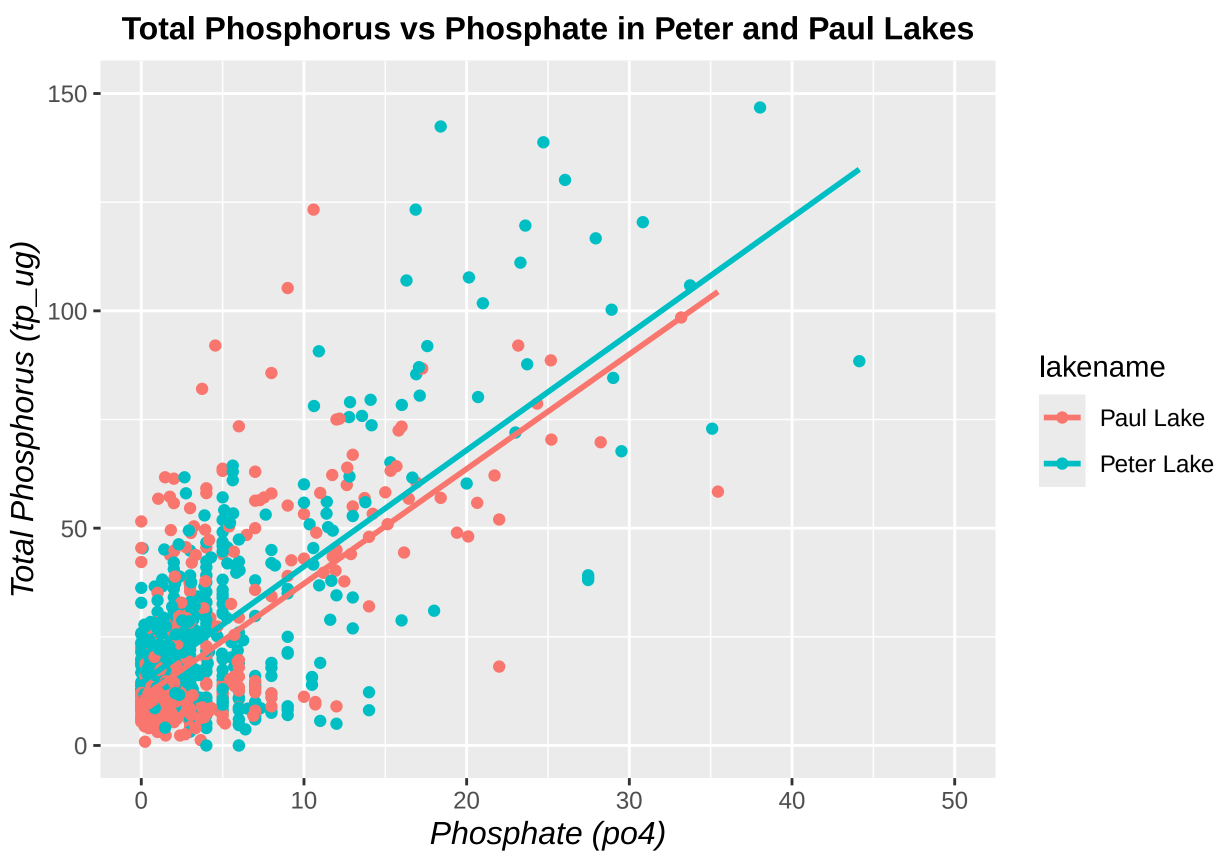
For numbers 4-7, create ggplot graphs and adjust aesthetics to follow best practices for data visualization. Ensure your theme, color palettes, axes, and additional aesthetics are edited accordingly.

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

```r
#4

ggplot(Nutrients_PeterPaul.data,
  aes(x = po4, y = tp_ug, color = lakename)) +
  geom_point() +  # Scatter plot
  geom_smooth(method = "lm", se = FALSE) +  # Add line of best fit
```

```r
  xlim(0, 50) +  # Adjust axis limits to avoid extreme values
  ylim(0, 150) +
  labs(title = "Total Phosphorus vs Phosphate in Peter and Paul Lakes",
       x = "Phosphate (po4)",
       y = "Total Phosphorus (tp_ug)") +
  theme(
    plot.title = element_text(face = "bold", size = 12, hjust = 0.5),
    axis.title = element_text(size = 12, face = "italic")
  )
```

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

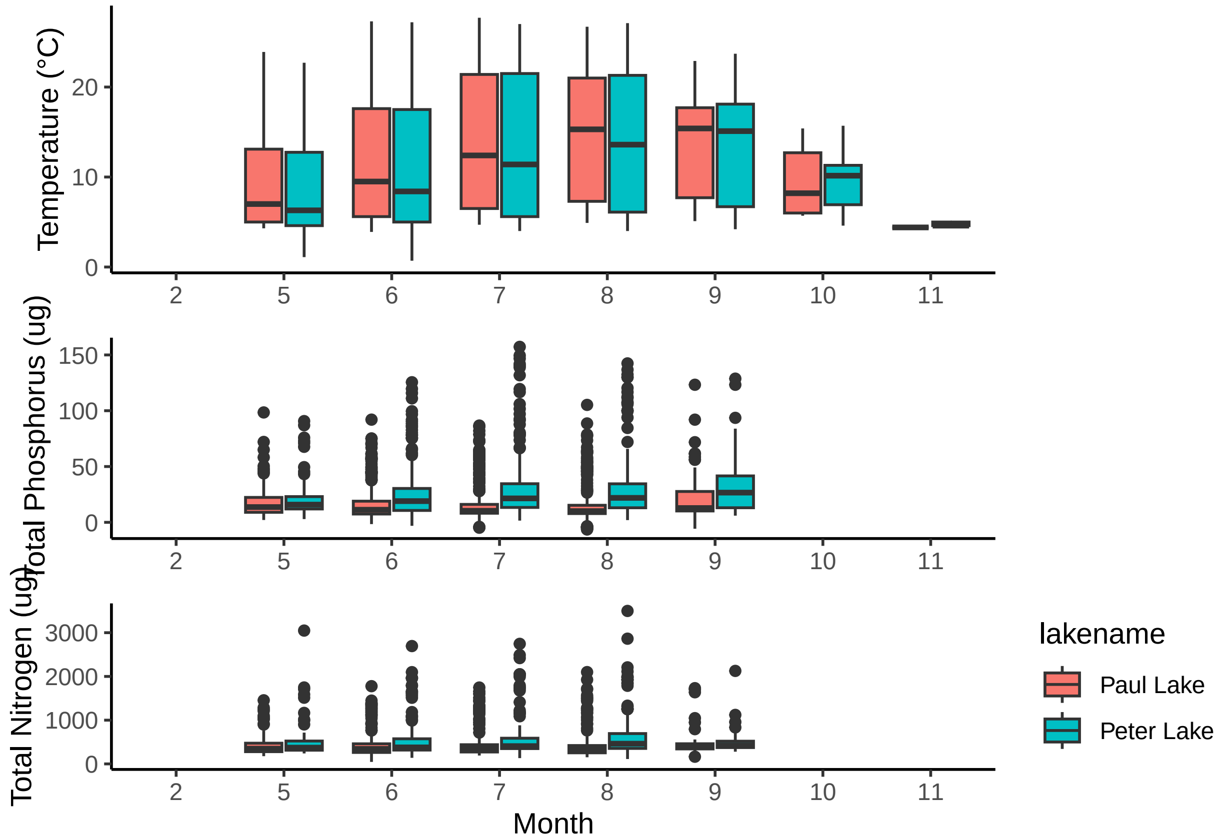


5. [NTL-LTER] Make three separate boxplots of (a) temperature, (b) TP, and (c) TN, with month as the x axis and lake as a color aesthetic. Then, create a cowplot that combines the three graphs. Make sure that only one legend is present and that graph axes are aligned.

Tips: * Recall the discussion on factors in the lab section as it may be helpful here. * Setting an axis title in your theme to `element_blank()` removes the axis title (useful when multiple, aligned plots use the same axis values) * Setting a legend's position to "none" will remove the legend from a plot. * Individual plots can have different sizes when combined using `cowplot`.

```r
#5

# Convert month to a factor for plotting
```

```r
Nutrients_PeterPaul.data$month <- as.factor(Nutrients_PeterPaul.data$month)

# Create the three separate plots

#Temperature boxplot
temp_plot <-
  ggplot(Nutrients_PeterPaul.data, aes(x = month, y = temperature_C, fill = lakename)) +
  geom_boxplot() +
  labs(x = "Month", y = "Temperature (°C)") +
  theme_classic() +
  theme(axis.title.x = element_blank(),  # Remove x-axis title for alignment
        legend.position = "none")  # Remove legend for this plot

#TP boxplot
tp_plot <- ggplot(Nutrients_PeterPaul.data, aes(x = month, y = tp_ug, fill = lakename)) +
  geom_boxplot() +
  labs(x = "Month", y = "Total Phosphorus (ug)") +
  theme_classic() +
  theme(axis.title.x = element_blank(),  # Remove x-axis title for alignment
        legend.position = "none")  # Remove legend for this plot

#TN boxplot
tn_plot <- ggplot(Nutrients_PeterPaul.data, aes(x = month, y = tn_ug, fill = lakename)) +
  geom_boxplot() +
  labs(x = "Month", y = "Total Nitrogen (ug)") +
  theme_classic()

# Combine the three plots into one using cowplot
combined_plot <- plot_grid(
  temp_plot, tp_plot, tn_plot,
  ncol = 1, align = "v", axis = "lr",
  rel_heights = c(1.25, 1, 1)
  # Make the temperature plot larger (1.25x the height of the others)
)

final_plot <- plot_grid(combined_plot, ncol = 1, rel_heights = c(1, 0.1))

# Display the final combined plot
print(final_plot)
```



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

Answer:

The temperature in both lakes shows a clear seasonal trend, increasing from February to August, and then gradually decreasing towards November. This pattern suggests typical seasonal warming during the summer and cooling in the fall. However, Peter Lake seems to exhibit slightly higher temperatures compared to Paul Lake, especially during the peak summer months (June to August), though the overall range of temperature is similar in both lakes.

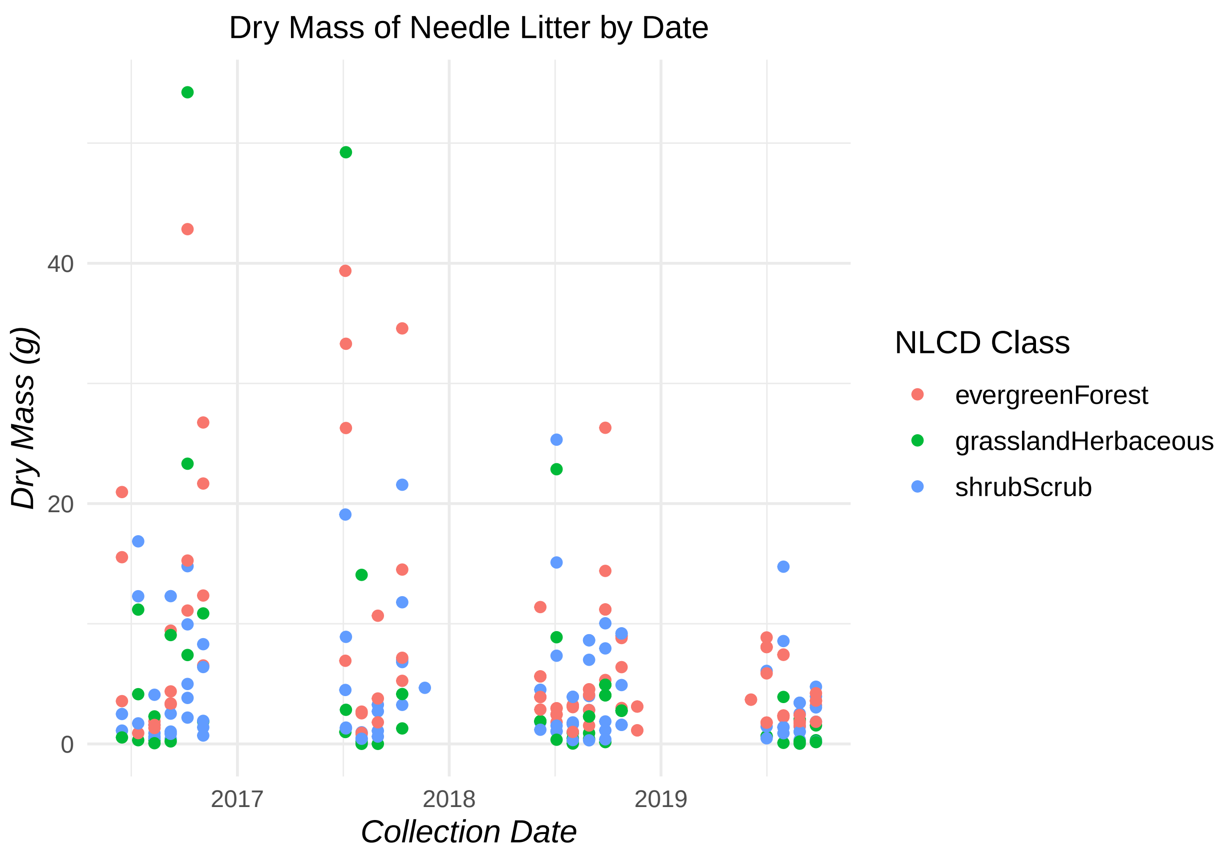
Total phosphorus levels show some variability throughout the year, with more pronounced outliers in the summer months (June to September), especially in Peter Lake. Both lakes exhibit relatively stable median phosphorus levels, but Peter Lake tends to have a wider range and higher concentration of outliers compared to Paul Lake, indicating potential nutrient spikes in the summer, possibly due to increased biological activity or external inputs.

Total nitrogen levels remain fairly consistent across the months, with small fluctuations and some outliers. Both lakes have similar patterns, though Peter Lake consistently shows a slightly higher spread in nitrogen levels, particularly during the summer months. Despite this, there is no major seasonal trend observed in nitrogen levels, with concentrations remaining relatively stable throughout the year.

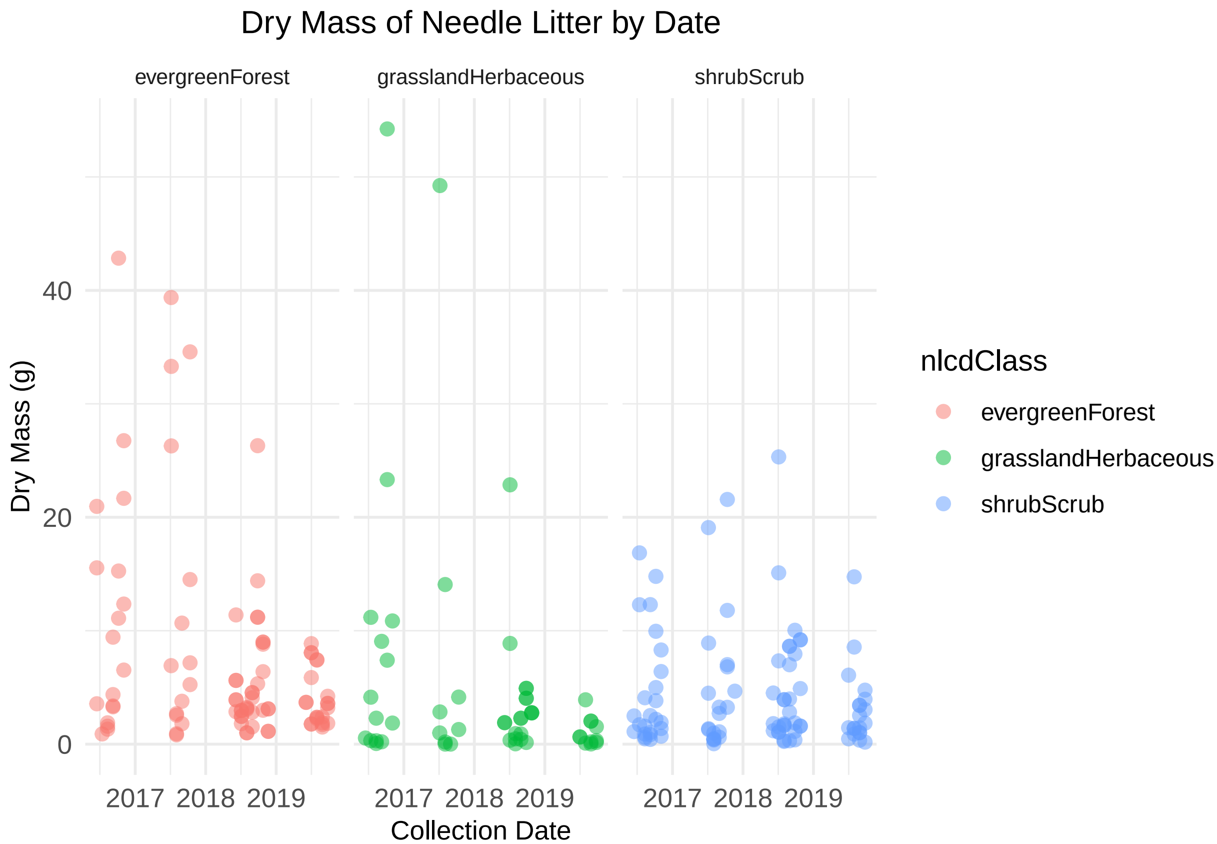
6. [Niwot Ridge] Plot a subset of the litter dataset by displaying only the “Needles” functional group. Plot the dry mass of needle litter by date and separate by NLCD class with a color aesthetic. (no need to adjust the name of each land use)

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

```
#6

Litter_Mass_sub.data <- subset(Litter_Mass_Trap.data, functionalGroup == "Needles")

# Convert the collectDate to Date format
Litter_Mass_sub.data$collectDate <- as.Date(Litter_Mass_sub.data$collectDate)

# Plot dry mass of needle litter by date
#Separated by NLCD class with a color aesthetic
plot6<-
ggplot(Litter_Mass_sub.data, aes(x = collectDate, y = dryMass, color = nlcdClass)) +
  geom_point() +
  labs(title = "Dry Mass of Needle Litter by Date",
       x = "Collection Date",
       y = "Dry Mass (g)",
       color = "NLCD Class") +  # Label for the legend
  theme_minimal() +
  theme(
    plot.title = element_text(size = 12, hjust = 0.5),
    axis.title = element_text(size = 12, face = "italic"),
    legend.title = element_text(size = 12),
    legend.text = element_text(size = 10),
  )
print(plot6)
```




```r
#7

plot7<-
  ggplot(Litter_Mass_sub.data, aes(x = collectDate, y = dryMass)) +
  geom_point(aes(color = nlcdClass), size = 2, alpha = 0.5) +
  labs(title = "Dry Mass of Needle Litter by Date",
       x = "Collection Date",
       y = "Dry Mass (g)") +
  # Create separate facets for each NLCD class
  facet_wrap(~ nlcdClass) +
  theme_minimal() +
  theme(
     plot.title = element_text(size = 12, hjust=0.5),
    axis.title = element_text(size = 10),
    axis.text = element_text(size = 10),
    strip.text = element_text(size = 8),
  )
print(plot7)
```



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

Answer:Plot 7 is more effective because it separates the NLCD classes into distinct facets, allowing for a clearer comparison between Evergreen Forest, Grassland Herbaceous, and Shrub Scrub. The use of separate panels prevents the data points from overlapping like in Plot 6, making it easier to identify trends within each class. By utilizing both color and faceting, Plot 7 reduces visual clutter and enhances the readability of the data, especially for viewers who might struggle to differentiate between colors.