Assignment 5: Data Visualization

Abby Liu

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 Import libraries & Read datasets
library(tidyverse);library(lubridate);library(here);library(cowplot)
here()
```

```
## [1] "/home/guest/EDA-Spring2023"
```

```
PeterPaul.chem.nutrients <- read.csv(
  here("Data/Processed/NTL-LTER_Lake_Chemistry_Nutrients_PeterPaul_Processed.csv"),
  stringsAsFactors = TRUE)

NIWO.litter <- read.csv(
  here("Data/Processed/NEON_NIWO_Litter_mass_trap_Processed.csv"),
  stringsAsFactors = TRUE)
```

```r
#2 Change to date format
PeterPaul.chem.nutrients$sampledate <- ymd(PeterPaul.chem.nutrients$sampledate)
NIWO.litter$collectDate <- ymd(NIWO.litter$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

```r
#3 Build a theme
my_theme <-  theme(
    text = element_text(color = 'mistyrose4', family=''),
    plot.background = element_rect(fill = 'mistyrose', colour = "mistyrose4", linewidth = 1),
    panel.background = element_rect(fill = 'ivory'),
    plot.title = element_text(size = rel(1.5), color = 'mistyrose4'),
    axis.text = element_text(color = "mistyrose4", size = rel(0.8)),
    axis.title = element_text(color = 'mistyrose4', size = rel(1)),
    legend.text = element_text(color='mistyrose4'),
    legend.background = element_rect(color='mistyrose4', fill = 'ivory'),
    legend.title = element_text(color='mistyrose4')
  )
theme_set(my_theme)
```

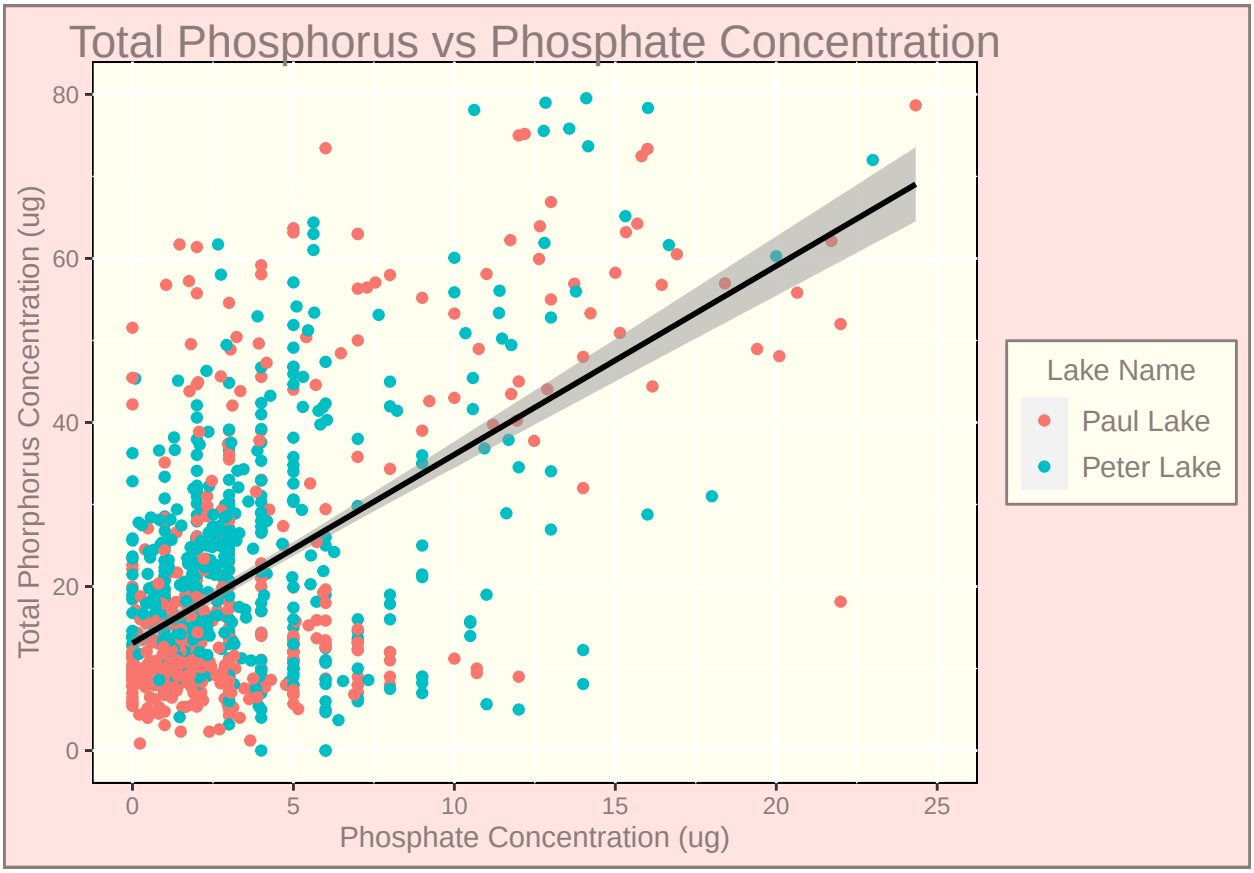
Create graphs

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()`).

```r
#4
p_vs_po4_plot <- PeterPaul.chem.nutrients %>%
  ggplot(
    mapping = aes(
      x=po4,
      y=tp_ug,
      color=lakename)
    )+
  geom_point(size=1.5)+
  labs(x = "Phosphate Concentration (ug)",
       y= "Total Phorphorus Concentration (ug)",
       title = "Total Phosphorus vs Phosphate Concentration",
       color = "Lake Name")+
```

```
  #facet_wrap(vars(lakename), nrow = 1)+
  geom_smooth(
    method = "lm",
    alpha = 0.5,
    mapping = aes(
      x=po4,
      y=tp_ug),
    color='black')+
  xlim(0,25)+
  ylim(0,80)+
  my_theme
p_vs_po4_plot
```



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
#Temperature plot
temp_plot <- PeterPaul.chem.nutrients %>%
  ggplot(
```

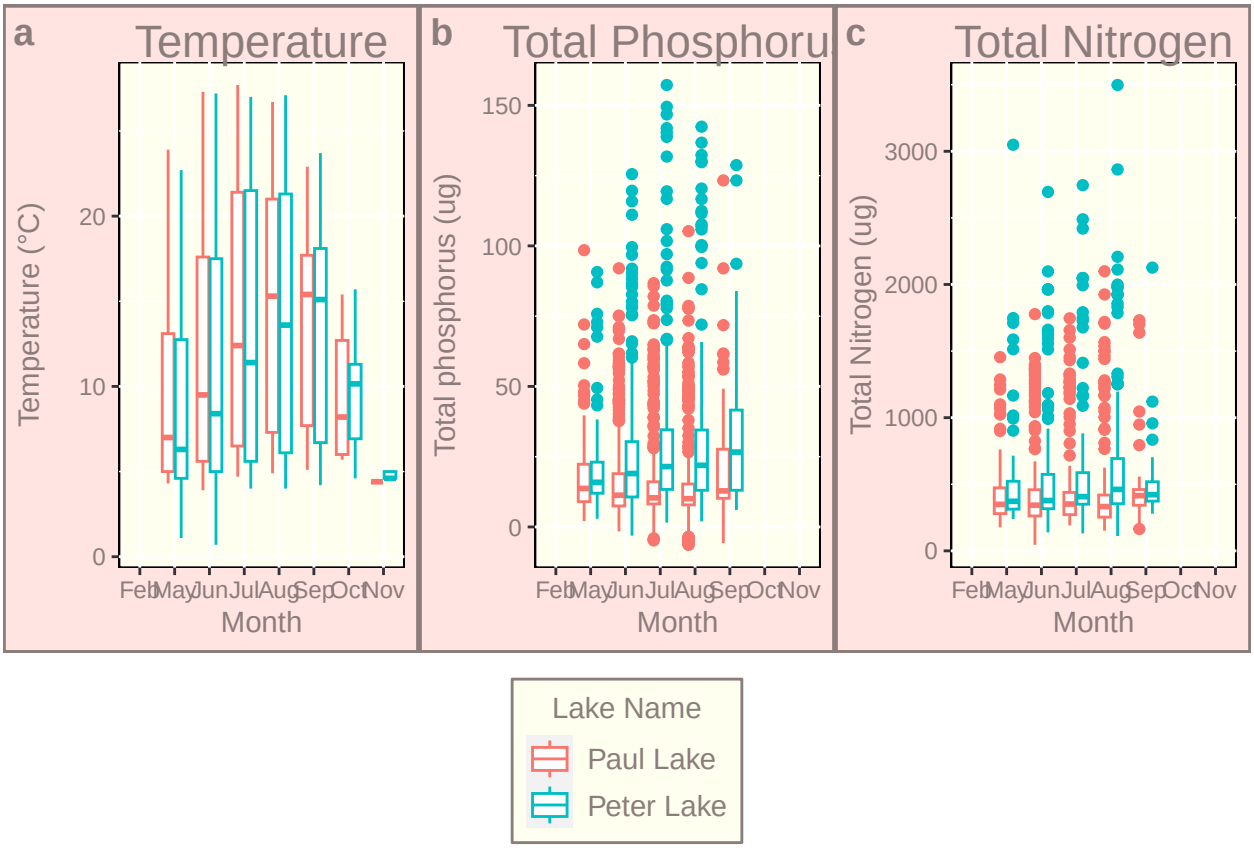
```
    mapping = aes(
      x=factor(month, levels = 1:12, labels = month.abb),
      y=temperature_C,
      color=lakename)
    )+
  geom_boxplot()+
  labs(x = "Month",
       y= "Temperature (°C)",
       title="Temperature",
       color = "Lake Name")+
  #scale_x_discrete(drop=FALSE) +
  my_theme

#TP plot
tp_plot <- PeterPaul.chem.nutrients %>%
  ggplot(
    mapping = aes(
      x=factor(month, levels = 1:12, labels = month.abb),
      y=tp_ug,
      color=lakename)
    )+
  geom_boxplot()+
  labs(x = "Month",
       y= "Total phosphorus (ug)",
       title="Total Phosphorus",
       color = "Lake Name")+
  #scale_x_discrete(drop=FALSE) +
  my_theme

#TN plot
tn_plot <- PeterPaul.chem.nutrients %>%
  ggplot(
    mapping = aes(
      x=factor(month, levels = 1:12, labels = month.abb),
      y=tn_ug,
      color=lakename)
    )+
  geom_boxplot()+
  labs(x = "Month",
       y= "Total Nitrogen (ug)",
       title="Total Nitrogen",
       color = "Lake Name")+
  #scale_x_discrete(drop=FALSE) +
  my_theme

#Combine plots
legend = get_legend(temp_plot)
combined_plot <- plot_grid(
  temp_plot + theme(legend.position="none"),
  tp_plot + theme(legend.position="none"),
  tn_plot + theme(legend.position="none"),
  labels = 'auto',
  align = 'vh',
```

```
  nrow = 1
)
final_plot <- plot_grid(combined_plot, legend, nrow=2, rel_heights = c(3,1))
final_plot
```



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

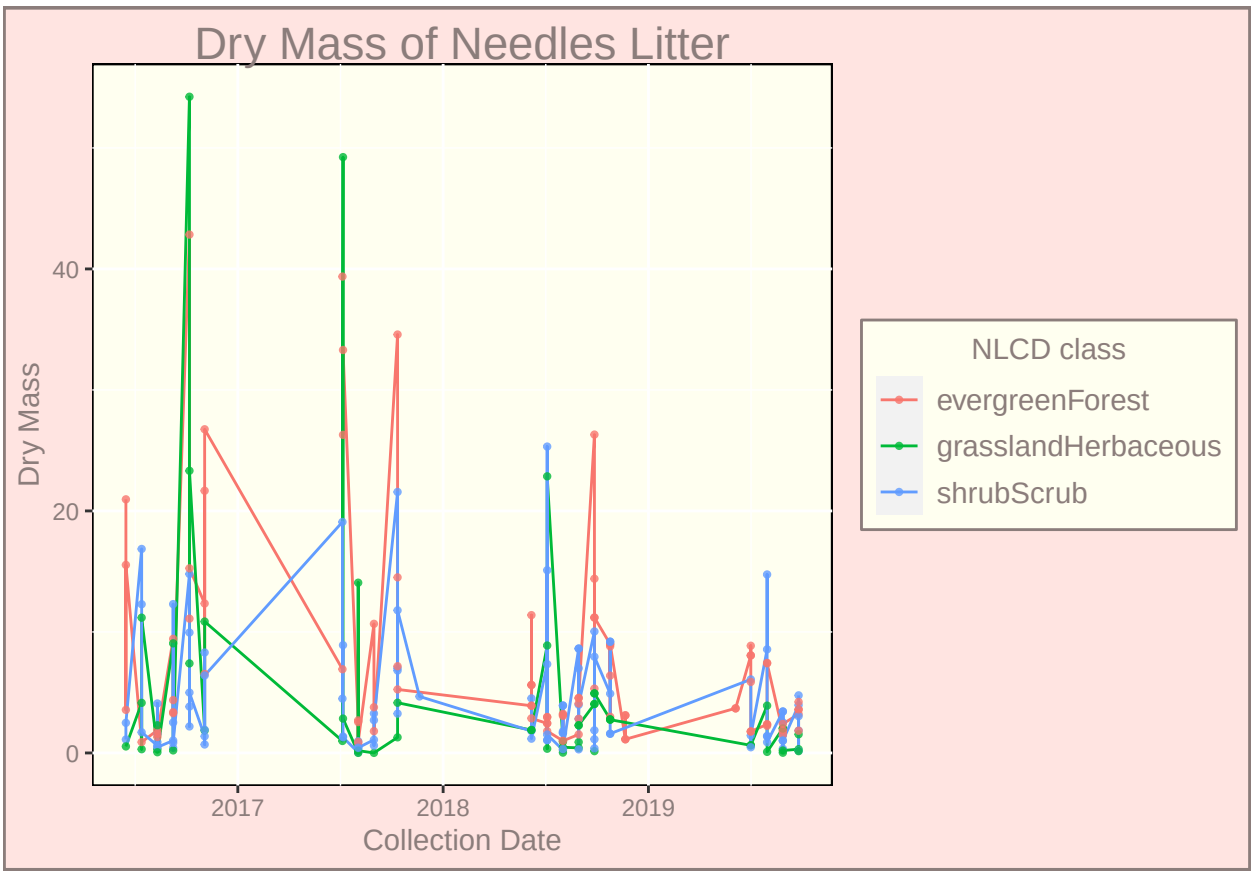
> Answer: For temperature, it goes up and then down in both lakes. For total phosphorus, the concentation in Paul Lake remain realitively steady while that in Peter Lake increases over season. For total nitrogen, the seasonal changes are relatively small and follow similar pattern of total phosphorus concentration.

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.

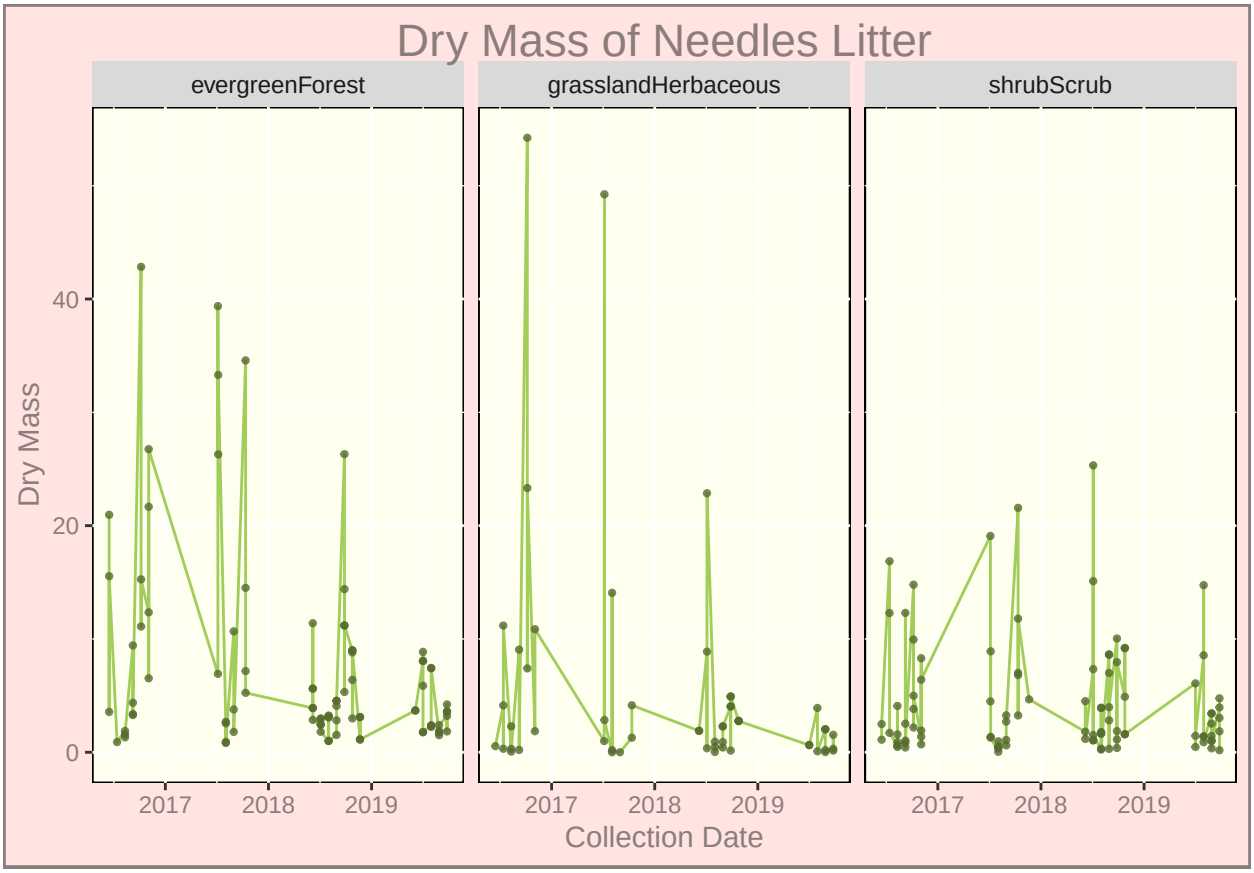
```
#6 Seprate by color
needle_plot <- NIWO.litter %>%
  filter(functionalGroup == 'Needles') %>%
  ggplot(
    mapping = aes(
      x=collectDate,
```

```
      y=dryMass,
      color=nlcdClass)
    )+
  geom_line()+
  geom_point(size=0.8, alpha=0.8)+
  labs(
    x='Collection Date',
    y='Dry Mass',
    title='Dry Mass of Needles Litter',
    color='NLCD class'
  )+
  my_theme
needle_plot
```




```
#7 Three facets
needle_plot_facet <- NIWO.litter %>%
  filter(functionalGroup == 'Needles') %>%
  ggplot(
    mapping = aes(
      x=collectDate,
      y=dryMass)
    )+
  geom_line(color='darkolivegreen3')+
  geom_point(color='darkolivegreen', size=0.8, alpha=0.8)+
  labs(
```

```
    x='Collection Date',
    y='Dry Mass',
    title='Dry Mass of Needles Litter'
  )+
  facet_wrap(vars(nlcdClass), nrow = 1)+
  my_theme
needle_plot_facet
```



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

Answer: I think plot 7 is more effective because there are a lot of variations in the dry mass so plot 6 can get pretty messy. In comparison, plot 7 is easier to read.