Assignment 5: Data Visualization

Jared Wang

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., "Salk_A05_DataVisualization.Rmd") prior to submission.

The completed exercise is due on Tuesday, February 11 at 1:00 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 (tidy and gathered) and the processed data file for the Niwot Ridge litter dataset.

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

```
#1
library(tidyverse)
library(lubridate)
library(ggplot2)
library(ggthemes)
library(cowplot)

getwd()

df.LTER <-
  read.csv("../Data/Processed/NTL-LTER_Lake_Chemistry_Nutrients_PeterPaul_Processed.csv")
df.LITTER <-
  read.csv("../Data/Processed/NEON_NIWO_Litter_mass_trap_Processed.csv")

#2
class(df.LITTER$collectDate)
df.LITTER <- mutate(df.LITTER, collectDate = as.Date(collectDate, format = "%Y-%m-%d"))
class(df.LTER$sampledate)
df.LTER <- mutate(df.LTER, sampledate = as.Date(sampledate, format = "%Y-%m-%d"))
```

Define your theme

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

```
theme.hc01 <- theme_hc() +
  theme(axis.title = element_text(family = "serif", size = (10)),
        axis.text = element_text(family = "serif", size = (8), color = "black"),
        legend.title = element_text(size = 10, family = "serif"),
```

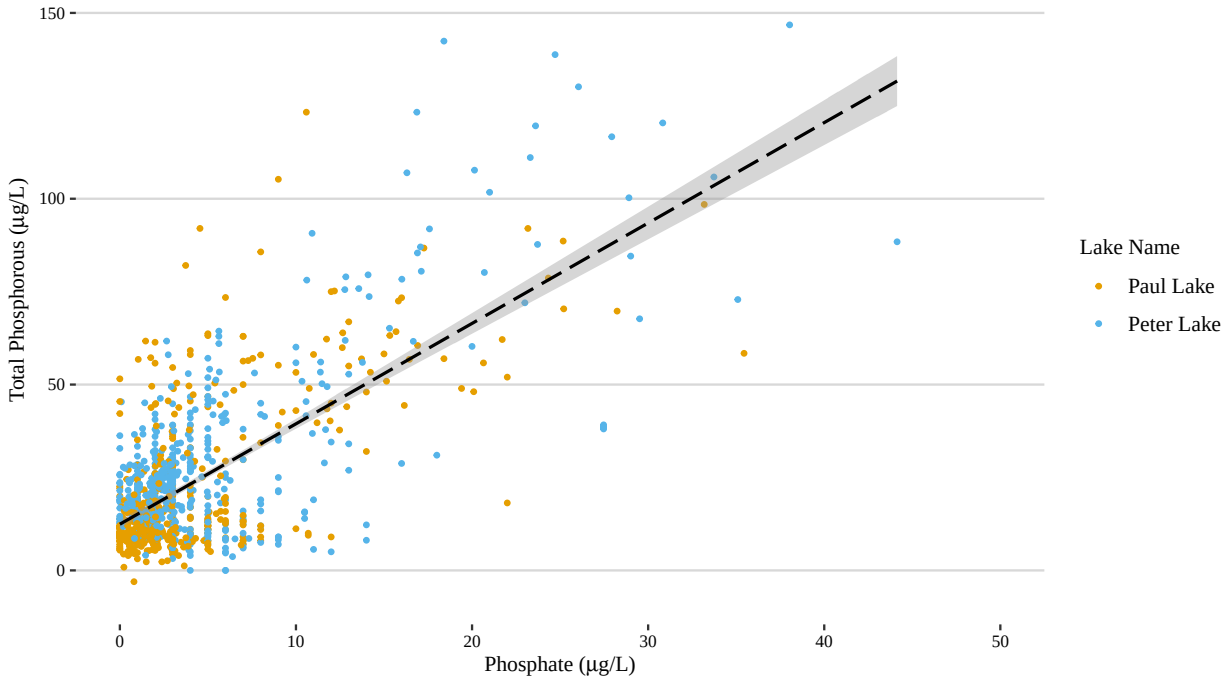
```
       legend.text = element_text(size = 10, family = "serif"),
       legend.key = element_rect(color = NA, fill = NA),
       legend.background = element_rect(color = NA, fill = NA),
       legend.position = "right")

theme.hc01.nolegend <- theme_hc() +
  theme(axis.title = element_text(family = "serif", size = (10)),
       axis.text = element_text(family = "serif", size = (8), color = "black"),
       legend.title = element_text(size = 10, family = "serif"),
       legend.text = element_text(size = 10, family = "serif"),
       legend.key = element_rect(color = NA, fill = NA),
       legend.background = element_rect(color = NA, fill = NA),
       legend.position = "none")

theme.hc01.legendin <- theme_hc() +
  theme(axis.title = element_text(family = "serif", size = (10)),
       axis.text = element_text(family = "serif", size = (8), color = "black"),
       legend.title = element_text(size = 10, family = "serif"),
       legend.text = element_text(size = 10, family = "serif"),
       legend.key = element_rect(color = NA, fill = NA),
       legend.background = element_rect(color = NA, fill = NA),
       legend.position = c(0.8, 0.8))
```

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 by phosphate, 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.

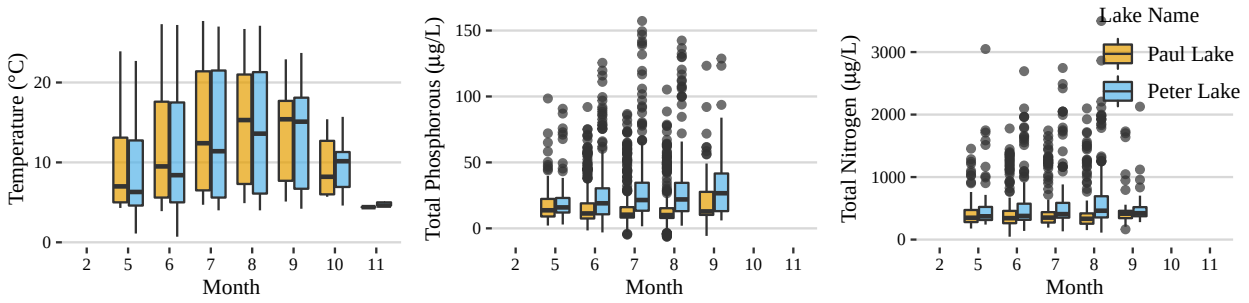
```
scat.po4 <- ggplot(df.LTER) +
  geom_point(aes(x = po4, y = tp_ug, color = lakename), size = 0.8, alpha = 1) +
  geom_smooth(aes(x = po4, y = tp_ug, color = lakename),
              method = lm,
              lty = 5, lwd = 0.7, color = "black") +
  scale_x_continuous(limits = c(0, 50)) +
  labs(x = expression(paste("Phosphate (", mu, "g/L)")),
       y = expression(paste("Total Phosphorous (", mu, "g/L)")),
       color = "Lake Name") +
  scale_color_manual(values = c("#E69F00", "#56B4E9")) +
  theme.hc01
plot(scat.po4)
```



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.

```
#convert month to factor
df.LTER.monthfact <- df.LTER %>%
  mutate(month = as.factor(month))

#temp
box.temp <- ggplot(df.LTER.monthfact) +
  geom_boxplot(aes(x = month, y = temperature_C, fill = lakename),
               alpha = 0.7) +
  labs(x = "Month", y = "Temperature (°C)", color = "Lake Name") +
  #scale_color_manual(values = c("cadetblue4", "cornsilk3")) +
  scale_fill_manual(values = c("#E69F00", "#56B4E9")) +
  theme.hc01.nolegend
#plot(box.temp)

#TP
box.tp <- ggplot(df.LTER.monthfact) +
  geom_boxplot(aes(x = month, y = tp_ug, fill = lakename),
               alpha = 0.7) +
  labs(x = "Month",
       y = expression(paste("Total Phosphorous (", mu, "g/L)")),
       fill = "Lake Name") +
  scale_fill_manual(values = c("#E69F00", "#56B4E9")) +
  theme.hc01.nolegend
#plot(box.tp)

#TN
```

```
box.tn <- ggplot(df.LTER.monthfact) +
  geom_boxplot(aes(x = month, y = tn_ug, fill = lakename),
               alpha = 0.7) +
  labs(x = "Month",
       y = expression(paste("Total Nitrogen (", mu, "g/L)")),
       fill = "Lake Name") +
  scale_fill_manual(values = c("#E69F00", "#56B4E9")) +
  theme.hc01.legendin
#plot(box.tn)

plot_grid(box.temp, box.tp, box.tn, nrow = 1)
```



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

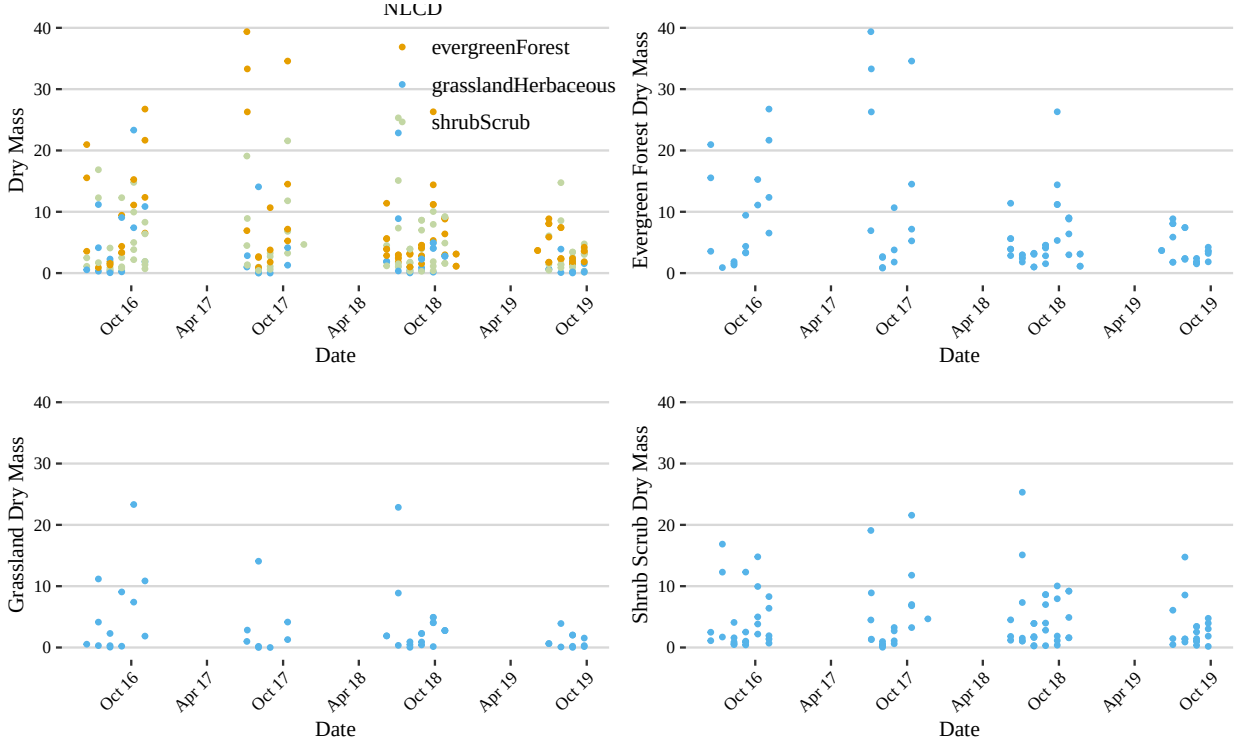
> Answer: Generally, there is a higher variation in temperature, total phosphorous, and total nutrient in summer than in spring and fall. Peter Lake generally has slightly higher nutrient and phosphorous concentration than Paul Lake.

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.

```
#question 6
df.LITTER.NE <- df.LITTER %>%
  filter(functionalGroup == "Needles") %>%
  mutate(collectDate = as.Date(collectDate, format = "%Y-%m-%d"))

scat.needle <- ggplot(df.LITTER.NE) +
  geom_point(aes(x = collectDate, y = dryMass, color = nlcdClass),
             size = 0.8, alpha = 1) +
  labs(x = "Date", y = "Dry Mass", color = "NLCD") +
  scale_y_continuous(limits = c(0, 40)) +
  scale_x_date(date_breaks = "6 months", date_labels = "%b %y") +
  scale_color_manual(values = c("#E69F00", "#56B4E9", "#C3D7A4")) +
  theme.hc01.legendin +
  theme(axis.text.x = element_text(angle = 45,  hjust = 1))
#plot(scat.needle)

#question 7
df.LITTER.NE.wide <- df.LITTER.NE %>%
  spread(key = nlcdClass, value = dryMass)
```

```r
scat.needle.evgr <- ggplot(df.LITTER.NE.wide) +
  geom_point(aes(x = collectDate, y = evergreenForest),
             color = "#56B4E9", size = 0.8, alpha = 1) +
  labs(x = "Date", y = "Evergreen Forest Dry Mass") +
  scale_y_continuous(limits = c(0, 40)) +
  scale_x_date(date_breaks = "6 months", date_labels = "%b %y") +
  theme.hc01 +
  theme(axis.text.x = element_text(angle = 45,  hjust = 1))
#plot(scat.needle.evgr)

scat.needle.gras <- ggplot(df.LITTER.NE.wide) +
  geom_point(aes(x = collectDate, y = grasslandHerbaceous),
             color = "#56B4E9", size = 0.8, alpha = 1) +
  labs(x = "Date", y = "Grassland Dry Mass") +
  scale_y_continuous(limits = c(0, 40)) +
  scale_x_date(date_breaks = "6 months", date_labels = "%b %y") +
  theme.hc01 +
  theme(axis.text.x = element_text(angle = 45,  hjust = 1))
#plot(scat.needle.gras)

scat.needle.shru <- ggplot(df.LITTER.NE.wide) +
  geom_point(aes(x = collectDate, y = shrubScrub),
             color = "#56B4E9", size = 0.8, alpha = 1) +
  labs(x = "Date", y = "Shrub Scrub Dry Mass") +
  scale_y_continuous(limits = c(0, 40)) +
  scale_x_date(date_breaks = "6 months", date_labels = "%b %y") +
  theme.hc01 +
  theme(axis.text.x = element_text(angle = 45,  hjust = 1))
#plot(scat.needle.shru)

plot_grid(scat.needle, scat.needle.evgr,
          scat.needle.gras, scat.needle.shru, nrow = 2)
```



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

Answer: Plots 7 are more effective in terms of showing the dry mass, because looking at one type of land use at a time reduces distraction. If the purpose is to compare dry mass by land use type, using a boxplot should be more effective than the scatter plot (plot 6).