

# Assignment 5: Data Visualization

Elsie Liu

## OVERVIEW

This exercise accompanies the lessons in Environmental Data Analytics on Data Visualization

### Set up your session

1. Set up your session. Verify your working directory and load the tidyverse and cowplot packages.
2. Make sure R is reading dates as date format; if not change the format to date.

```
#1
getwd()

## [1] "/Users/Lsy/Box Sync/Duke/spring2022/872-EDA/GitRepository/Environmental_Data_Analytics_2022"
library(tidyverse)
library(cowplot)
chemnutr <- read.csv("./Data/Processed/NTL-LTER_Lake_Chemistry_Nutrients_PeterPaul_Processed.csv")
litter <- read.csv("./Data/Processed/NEON_NIWO_Litter_mass_trap_Processed.csv")
#2
class(chemnutr$sampleddate)

## [1] "character"
chemnutr$sampleddate <- as.Date(chemnutr$sampleddate)
class(litter$collectDate)

## [1] "character"
litter$collectDate <- as.Date(litter$collectDate)
```

### Define your theme

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

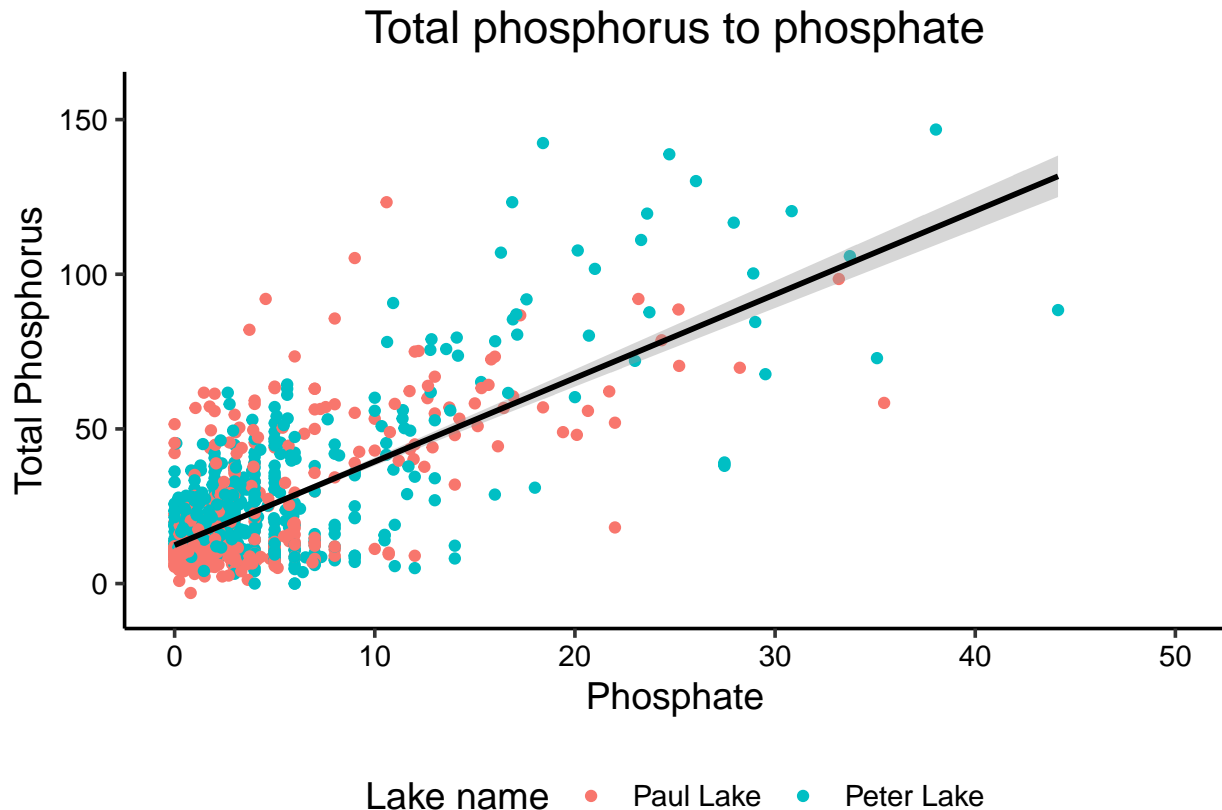
```
#3
mytheme <- theme_classic(base_size = 14) +
  theme(axis.text = element_text(color = "black"),
        legend.position = "bottom", plot.title = element_text(hjust = 0.5))
theme_set(mytheme)
```

### 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 = chemnutr, aes(x = po4, y = tp_ug))+
  geom_point(aes(color = lakename))+
  xlim(0,50)+
  xlab("Phosphate")+
  ylab("Total Phosphorus")+
  labs(title="Total phosphorus to phosphate", colour = "Lake name")+
  geom_smooth(method = lm, colour="black")
```



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
#boxplot
ChemNutrplot1 <- ggplot(data = chemnutr, aes(x = month, y=temperature_C, color=lakename))+
  geom_boxplot()+
  labs(title="Boxplot of Temperature", colour = "Lake name")+
  theme(plot.title = element_text(size=8))

ChemNutrplot2 <- ggplot(data = chemnutr, aes(x = month, y=tp_ug, color=lakename))+
  geom_boxplot()+
  labs(title="Boxplot of Total Phosphorus", colour = "Lake name")+
  theme(plot.title = element_text(size=8))

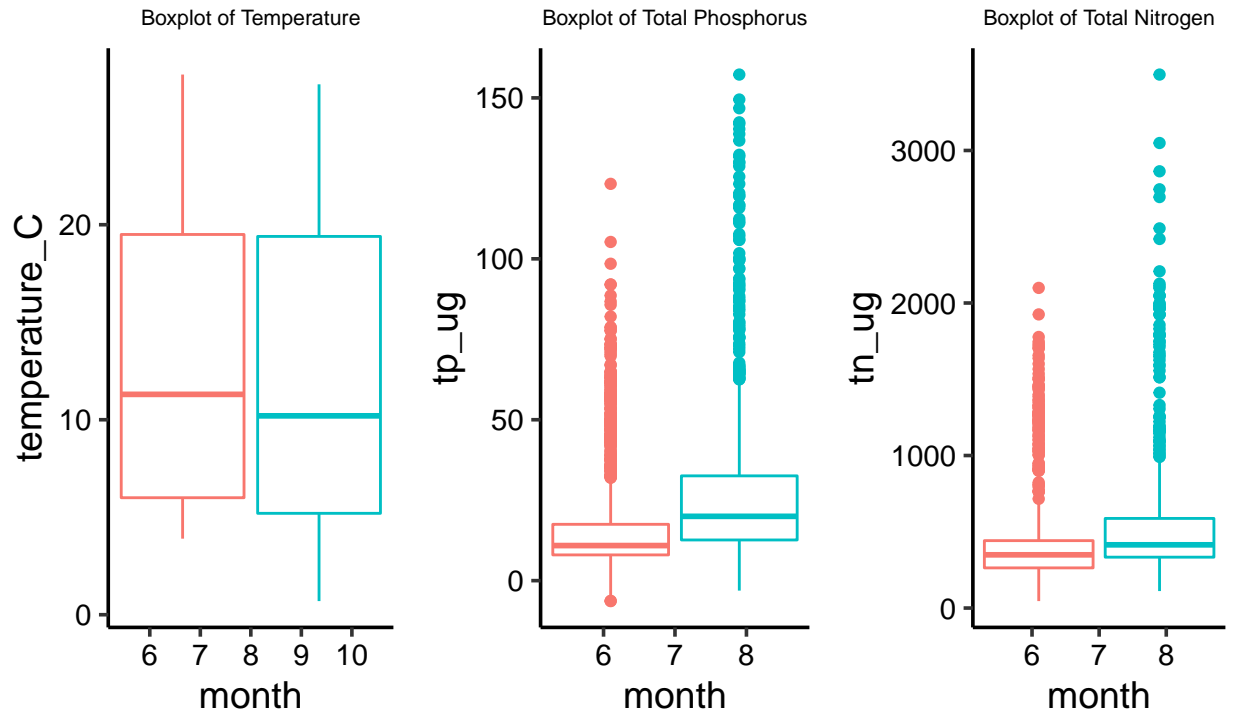
ChemNutrplot3 <- ggplot(data = chemnutr, aes(x = month, y=tn_ug, color=lakename))+
  geom_boxplot()+
  labs(title="Boxplot of Total Nitrogen", colour = "Lake name")+
  theme(plot.title = element_text(size=8))
```

```

theme(plot.title = element_text(size=8))

#cowplot
library(cowplot)
plot_grid(ChemNutrplot1,ChemNutrplot2,ChemNutrplot3, nrow = 1, ncol=3, align = 'h', rel_heights = c(1.2,

```



lake name Paul Lake    Lake name Paul Lake    Lake name Paul Lake

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

Answer: Temperature has more compact structure while the total phosphorous and total nitrogen has lots of outliers. Peter Lake also shows higher variance than Paul Lake.

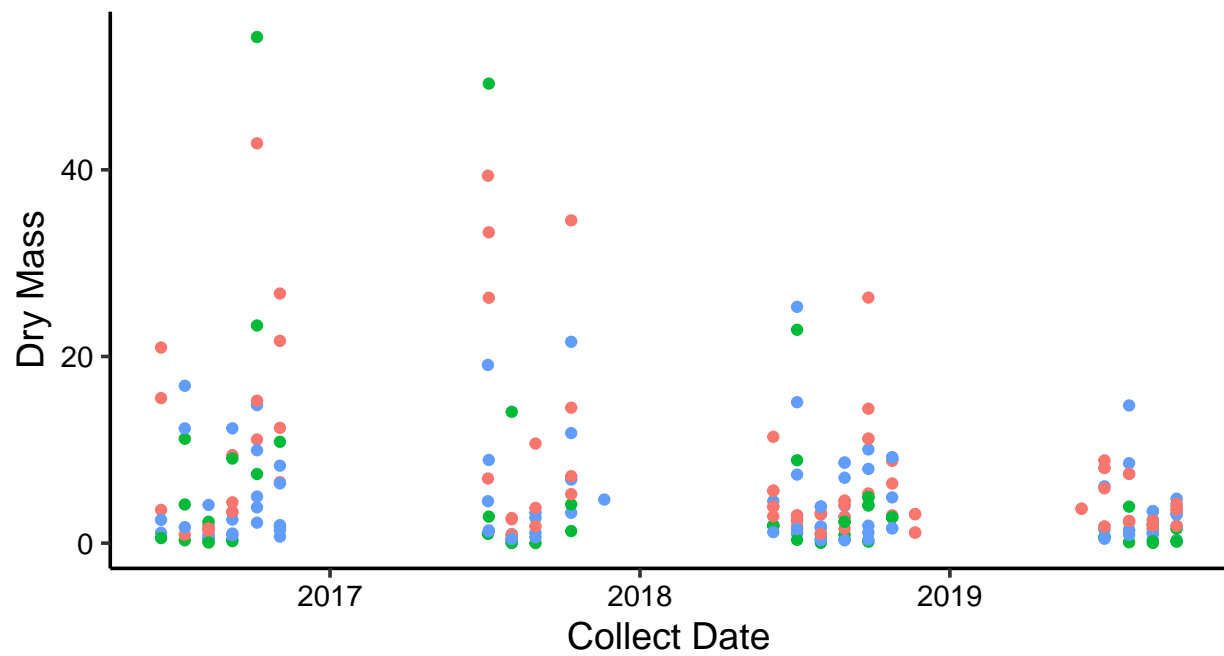
6. [Niwot Ridge] Plot a subset of the litter dataset by displaying only the “Needles” functional group. Plot the dry mass of needle litter by date and separate by NLCD class with a color aesthetic. (no need to adjust the name of each land use)
7. [Niwot Ridge] Now, plot the same plot but with NLCD classes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

```

#6
ggplot(data=subset(litter, functionalGroup == "Needles"), aes(x = collectDate,y = dryMass, color=nlcdCl
geom_point()+
xlab("Collect Date")+
ylab("Dry Mass")+
labs(title="Dry mass of needle litter by date", colour = "NLCD Class")

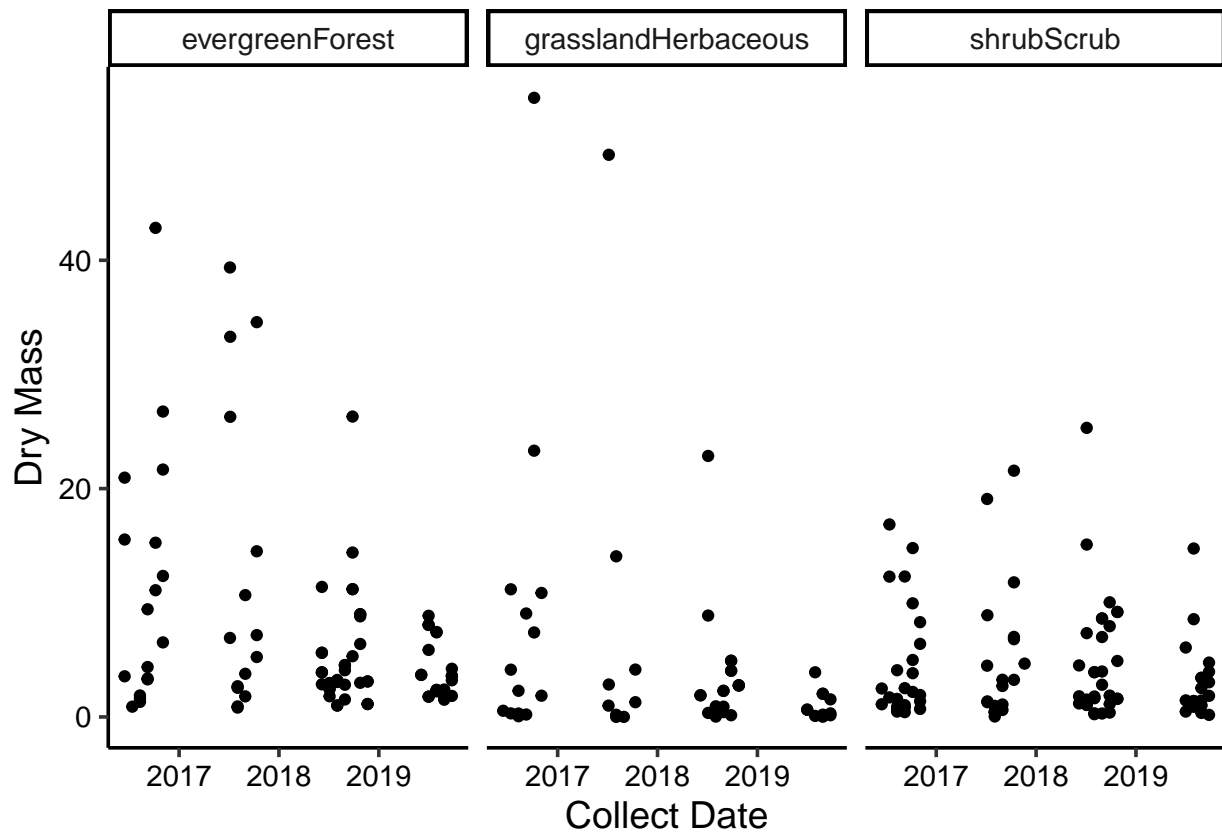
```

## Dry mass of needle litter by date



NLCD Class    ● evergreenForest    ● grasslandHerbaceous    ● shrubScrub

```
#7
ggplot(data=subset(litter, functionalGroup == "Needles"), aes(x = collectDate, y = dryMass))+
  geom_point()+
  facet_wrap(vars(nlcdClass))+
  xlab("Collect Date")+
  ylab("Dry Mass")
```



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

**Answer:** The plot from 7 is more effective as it is easier for viewers to compare through time within nld groups and see the unique pattern of each specific group.