Assignment 5: Data Visualization

Student Name

OVERVIEW

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

Directions

1. Rename this file `<FirstLast>_A02_CodingBasics.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 to **answer the questions** in this assignment document.
5. When you have completed the assignment, **Knit** the text and code into a single PDF file.

The completed exercise is due on Friday, Oct 14th @ 5:00pm.

Set up your session

1. Set up your session. Verify your working directory and load the tidyverse, lubridate, & 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_PeterP`] version) and the processed data file for the Niwot Ridge litter dataset (use the [`NEON_NIWO_Litter_mass_trap_Processe`] version).

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

```
# 1
library(tidyverse)
library(lubridate)
library(cowplot)
ntl = read.csv(file = "E:/EDA-Fall2022/Data/Processed/NTL-LTER_Lake_Chemistry_Nutrients_PeterPaul_Proces
neon = read.csv(file = "E:/EDA-Fall2022/Data/Processed/NEON_NIWO_Litter_mass_trap_Processed.csv")

# 2
neon_1 = neon
neon_1$collectDate <- as.Date(neon_1$collectDate)
```

Define your theme

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

```r
# 3
kc_theme <- function() {
    theme_classic(base_size = 14) + theme(axis.text = element_text(color = "black"),
        legend.position = "top")
}
```

Create graphs

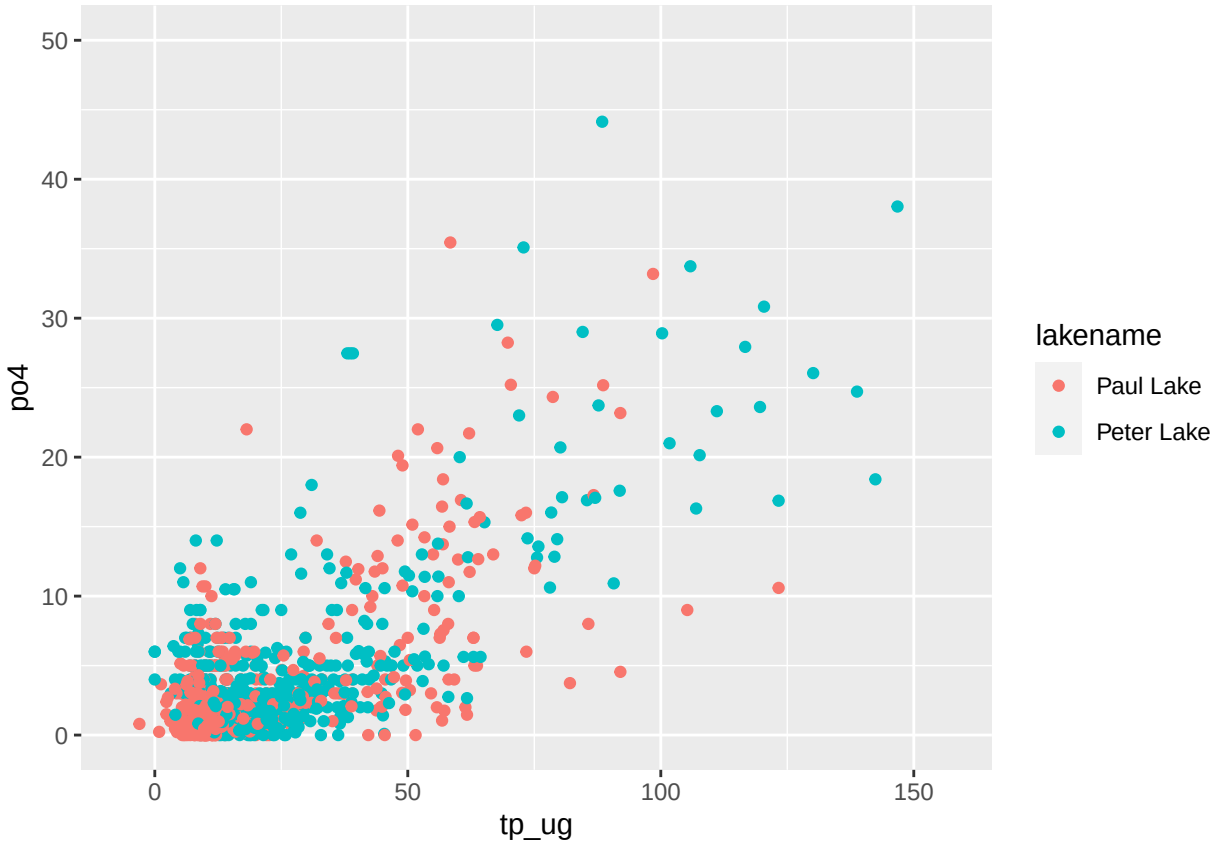
For numbers 4-7, create ggplot graphs and adjust aesthetics to follow best practices for data visualization. Ensure your theme, color palettes, axes, and additional aesthetics are edited accordingly.

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/or `ylim()`).

```r
# 4
class(ntl)
```

```
## [1] "data.frame"
```

```r
p0 <- ggplot(ntl, aes(tp_ug, po4), warning = FALSE)
p0 + geom_point(aes(colour = lakename)) + ylim(0, 50)
```

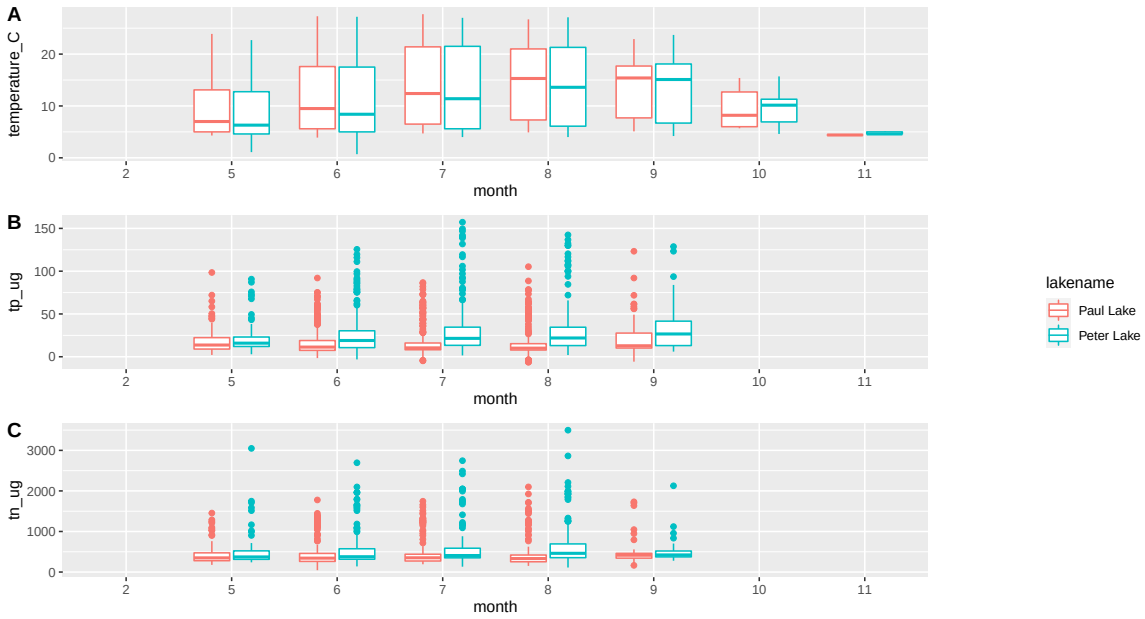


5. [NTL-LTER] Make three separate boxplots of (a) temperature, (b) TP, and

(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

# p <- ggplot(ntl, lakename = 'group') p + geom_boxplot(aes(month,
# temperature_C), color=factor(group))

ntl$month <- as.factor(ntl$month)

p1 <- ggplot(ntl, aes(month, temperature_C)) + geom_boxplot(aes(colour = lakename)) +
    theme(legend.position = "none")

p2 <- ggplot(ntl, aes(month, tp_ug)) + geom_boxplot(aes(colour = lakename)) + theme(legend.position = "

p3 <- ggplot(ntl, aes(month, tn_ug)) + geom_boxplot(aes(colour = lakename)) + theme(legend.position = "

p5 <- ggplot(ntl, aes(month, tn_ug)) + geom_boxplot(aes(colour = lakename))

p4 <- plot_grid(p1, p2, p3, labels = c("A", "B", "C"), nrow = 3, align = "hv")

legend <- get_legend(p5 + theme(legend.box.margin = margin(0, 0, 0, 2)))

p7 <- plot_grid(p4, legend, rel_widths = c(3, 1))

plot(p7)
```



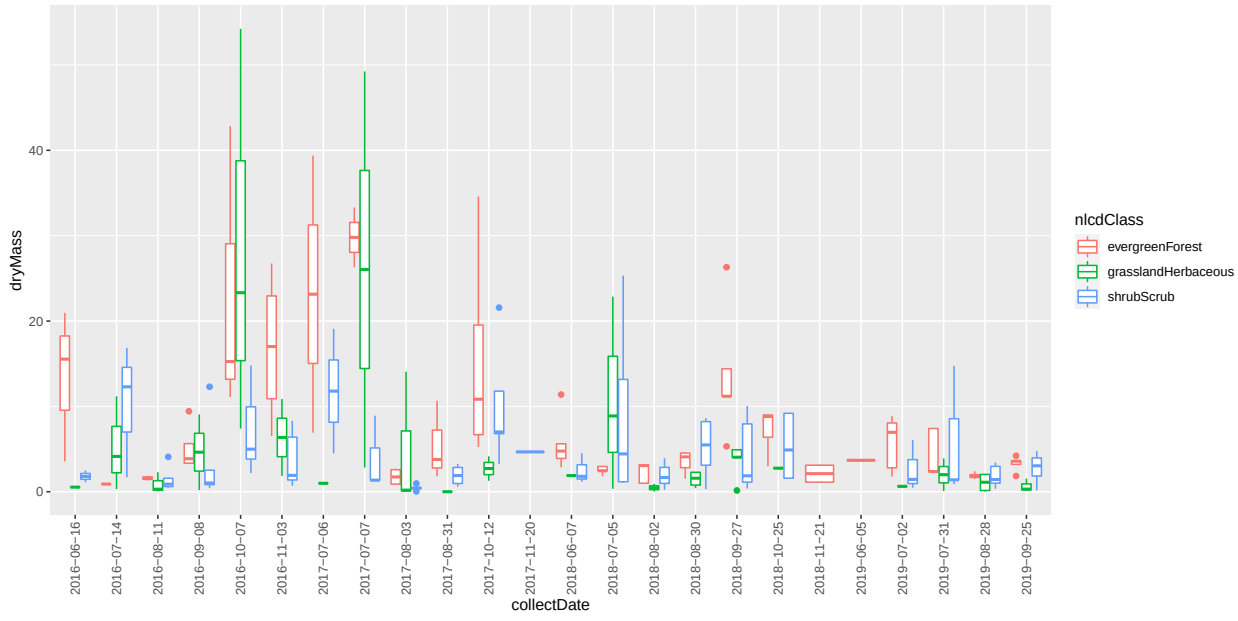
Question: What do you observe about the variables of interest over seasons and between lakes?

Answer: Temperature doesn't change too much, Peter Lake always has higher tp_ug and tn_ng than Paul Lake

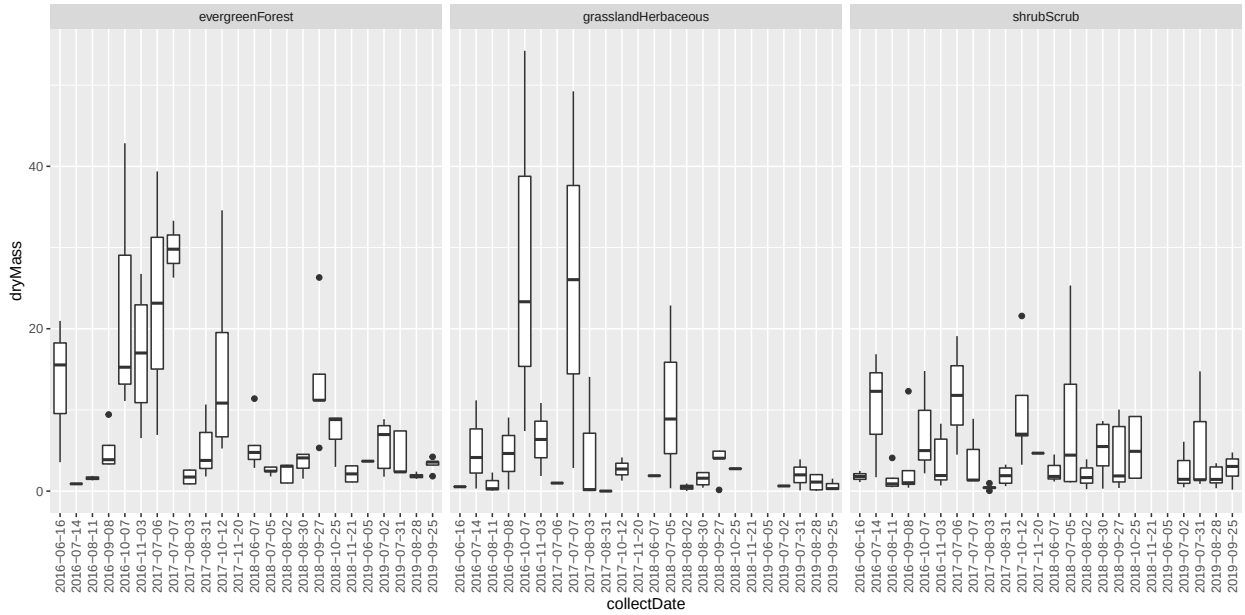
6. [Niwot Ridge] Plot a subset of the litter dataset by displaying only the "Needles" functional group. Plot the dry mass of needle litter by date and separate by NLCD class with a color aesthetic. (no need to adjust the name of each land use)

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

```
# 6
neon_ndls <- filter(neon_1, functionalGroup == "Needles")

neon_ndls$collectDate <- as.factor(neon_ndls$collectDate)

p6 <- ggplot(neon_ndls, aes(collectDate, dryMass)) + geom_boxplot(aes(colour = nlcdClass)) +
    theme(axis.text.x = element_text(angle = 90, vjust = 0.5, hjust = 1))

plot(p6)
```




```
# 7
FacetPlot1 = ggplot(neon_ndls, aes(collectDate, dryMass)) + geom_boxplot() + facet_grid(~nlcdClass) +
    theme(axis.text.x = element_text(angle = 90, vjust = 0.5, hjust = 1))

plot(FacetPlot1)
```



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

Answer: 7 is more effective to see the trend of each dataset and 6 is more effectice to compare the valuse in each date.