

Assignment 5: Data Visualization

Emily Wood

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_PeterPaul version) and the processed data file for the Niwot Ridge litter dataset (use the [NEON_NIWO_Litter_mass_trap_Processed version).
2. Make sure R is reading dates as date format; if not change the format to date.

```
# 1
getwd()

## [1] "/home/guest/EDA_2022/EDA-Fall2022"

library(tidyverse)

## -- Attaching packages ----- tidyverse 1.3.2 --
## v ggplot2 3.3.6      v purrr   0.3.4
## v tibble  3.1.8      v dplyr  1.0.10
## v tidyr   1.2.0      v stringr 1.4.1
## v readr   2.1.2      v forcats 0.5.2
## -- Conflicts ----- tidyverse_conflicts() --
## x dplyr::filter() masks stats::filter()
## x dplyr::lag()    masks stats::lag()

library(lubridate)

##
## Attaching package: 'lubridate'
##
## The following objects are masked from 'package:base':
##
```

```
##      date, intersect, setdiff, union
library(cowplot)

##
## Attaching package: 'cowplot'
##
## The following object is masked from 'package:lubridate':
##
##      stamp
PeterPaulchemnutrients <- read.csv("./Data/Processed/NTL-LTER_Lake_Chemistry_Nutrients_PeterPaul_Processed.csv",
  stringsAsFactors = T)

PeterPaulchemPhys <- read.csv("./Data/Processed/NTL-LTER_Lake_ChemistryPhysics_PeterPaul_Processed.csv",
  stringsAsFactors = T)

NIWOTRidgeLitter <- read.csv("./Data/Processed/NEON_NIWO_Litter_mass_trap_Processed.csv",
  stringsAsFactors = T)

# 2

class(PeterPaulchemnutrients$sampleddate)

## [1] "factor"
class(NIWOTRidgeLitter$collectDate)

## [1] "factor"
PeterPaulchemnutrients$sampleddate <- as.Date(PeterPaulchemnutrients$sampleddate, format = "%Y-%m-%d")
NIWOTRidgeLitter$collectDate <- as.Date(NIWOTRidgeLitter$collectDate, format = "%Y-%m-%d")
```

Define your theme

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

```
# 3

mynewtheme <- theme_grey(base_size = 12) + theme(axis.text = element_text(color = "Dark Green"),
  legend.position = "top")

theme_set(mynewtheme)
```

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.

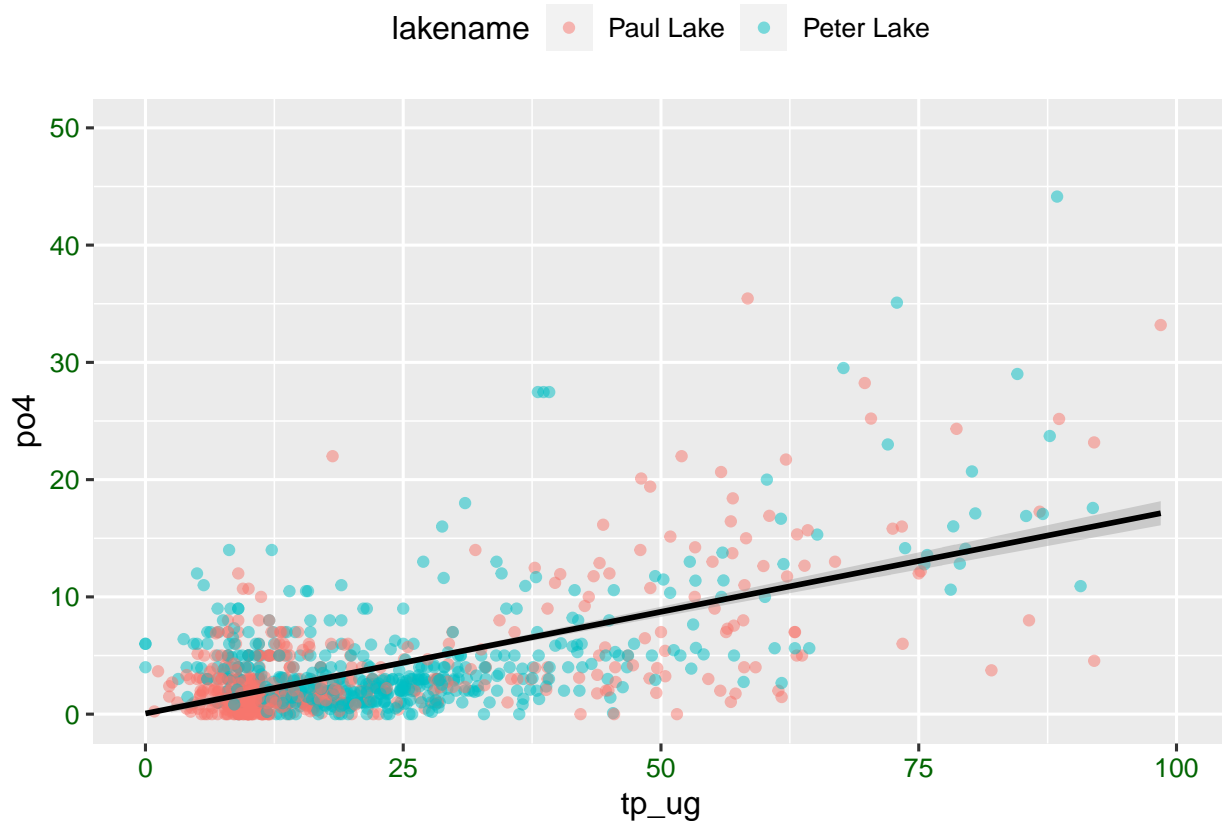
4. [NTL-LTER] Plot total phosphorus (tp_ug) by phosphate (po4), with separate aesthetics for Peter and Paul lakes. Add a line of best fit and color it black. Adjust your axes to hide extreme values (hint: change the limits using `xlim()` and/or `ylim()`).

```
# 4

Graph1 <- ggplot(PeterPaulchemnutrients, aes(x = tp_ug, y = po4)) + geom_point(aes(color = lakename),
  alpha = 0.5) + geom_smooth(method = lm, color = "black") + xlim(0, 100) + ylim(0,
```

```
50)
print(Graph1)

## `geom_smooth()` using formula 'y ~ x'
## Warning: Removed 21964 rows containing non-finite values (stat_smooth).
## Warning: Removed 21964 rows containing missing values (geom_point).
```



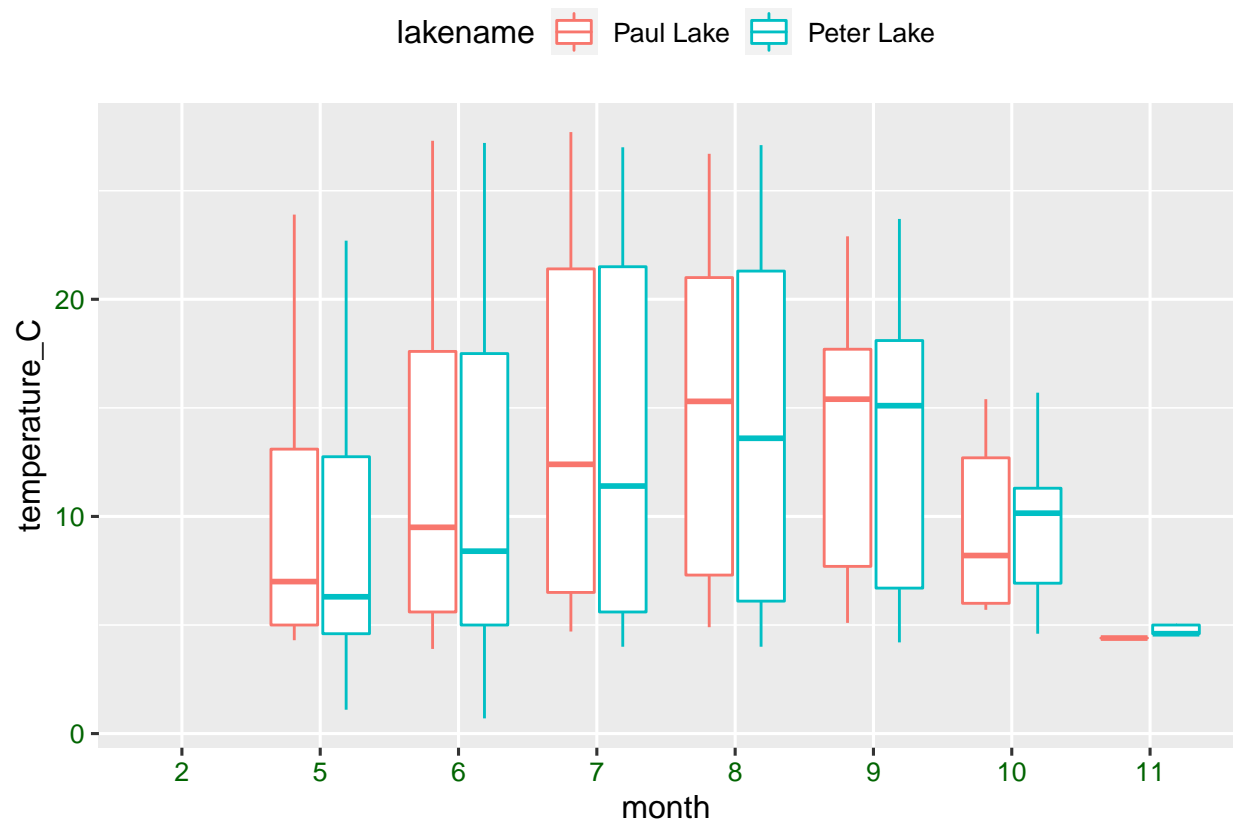
5. [NTL-LTER] Make three separate boxplots of (a) temperature, (b) TP, and (c) TN, with month as the x axis and lake as a color aesthetic. Then, create a cowplot that combines the three graphs. Make sure that only one legend is present and that graph axes are aligned.
- Tip: R has a built-in variable called `month.abb` that returns a list of months; see <https://r-lang.com/month-abb-in-r-with-example>

```
# 5

PeterPaulchemnutrients$month <- factor(PeterPaulchemnutrients$month, levels = c(1:12))

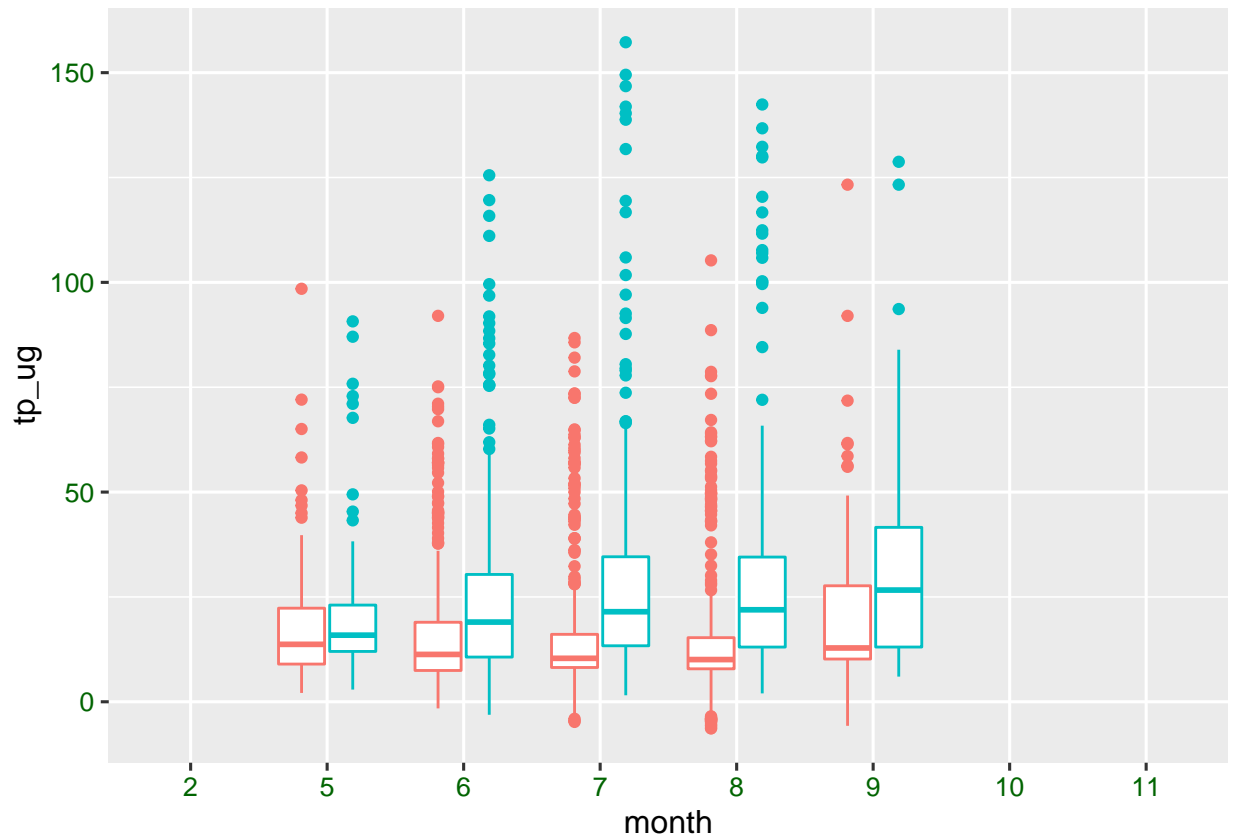
Boxplot_Temp <- ggplot(PeterPaulchemnutrients, aes(x = month, y = temperature_C,
  color = lakename)) + geom_boxplot() + theme(legend.position = "top")
plot(Boxplot_Temp)

## Warning: Removed 3566 rows containing non-finite values (stat_boxplot).
```



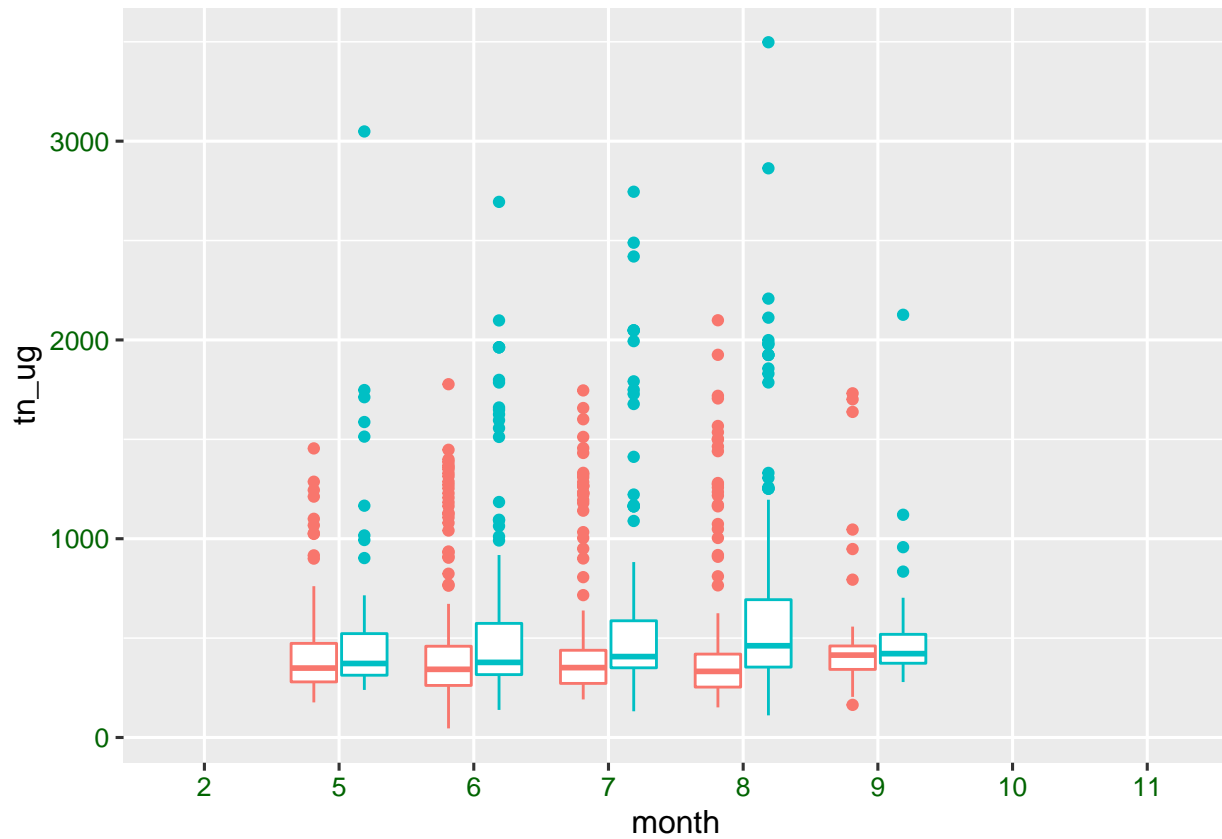
```
Boxplot_TP <- ggplot(PeterPaulchemnutrients, aes(x = month, y = tp_ug, color = lakename)) +
  geom_boxplot() + theme(legend.position = "none")
plot(Boxplot_TP)
```

```
## Warning: Removed 20729 rows containing non-finite values (stat_boxplot).
```



```
Boxplot_TN <- ggplot(PeterPaulchemnutrients, aes(x = month, y = tn_ug, color = lakename)) +
  geom_boxplot() + theme(legend.position = "none")
plot(Boxplot_TN)
```

```
## Warning: Removed 21583 rows containing non-finite values (stat_boxplot).
```



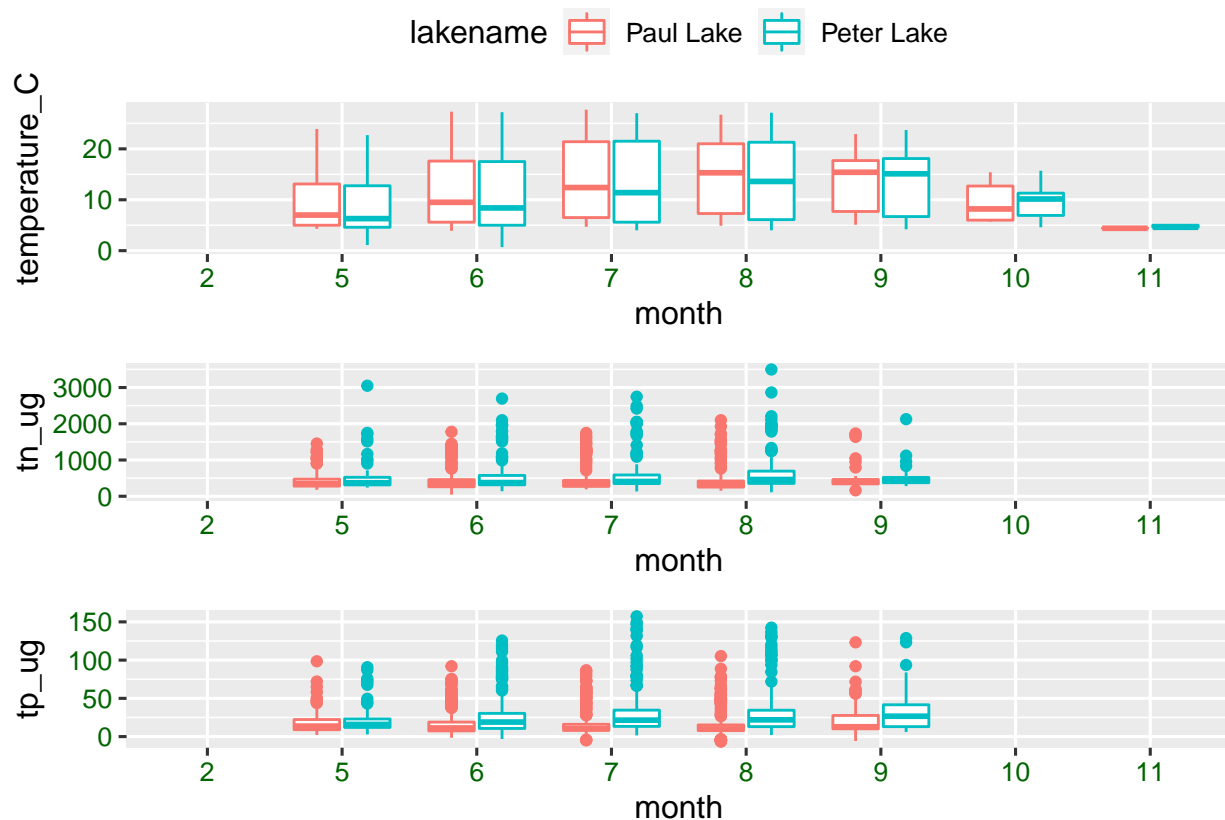
```
Boxplot_All <- plot_grid(Boxplot_Temp, Boxplot_TN, Boxplot_TP, nrow = 3, axis = "b",
  align = "v", rel_heights = c(1.5, 1, 1))
```

```
## Warning: Removed 3566 rows containing non-finite values (stat_boxplot).
```

```
## Warning: Removed 21583 rows containing non-finite values (stat_boxplot).
```

```
## Warning: Removed 20729 rows containing non-finite values (stat_boxplot).
```

```
plot(Boxplot_All)
```



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

Answer: I notice that the temperature stays about the same between each lake for each month. It also rises in the summer months as expected. I notice that Peter lake has more outliers for each month where TN was measured compared to Paul lake. Peter also has higher TN overall across all months than Paul Lake. TP also is higher in Peter lake and has more high outliers 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

```
Graph2 <- ggplot(subset(NIWOTRidgeLitter, functionalGroup == "Needles"), aes(x = collectDate,
  y = dryMass)) + geom_point(aes(color = nlcdClass))
plot(Graph2)
```



7

```
Graph3 <- ggplot(subset(NIWOTRidgeLitter, functionalGroup == "Needles"), aes(x = collectDate,
  y = dryMass)) + geom_point() + facet_grid("nlcdClass")
plot(Graph3)
```




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

Answer: I believe the plot from question six is the most effective because the colors on the same plot allow you to make an easier comparison between the nlcd classes. That being said, it's difficult to tell how many of each type is in each class because the colored point overlap. If that is the intended purpose, one might prefer the plots from question seven.