Assignment 5: Data Visualization

Xiangtian 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(cowplot)
library(ggplot2)
PeterPaul.chem.nutrients <-
  read.csv("../Data/Processed/NTL-LTER_Lake_Chemistry_Nutrients_PeterPaul_Processed.csv")
PeterPaul.gather.nutrients <-
  read.csv("../Data/Processed/NTL-LTER_Lake_Nutrients_PeterPaulGathered_Processed.csv")
Litter.mass.trap <-
  read.csv("../Data/Processed/NEON_NIWO_Litter_mass_trap_Processed.csv")
#2
class(PeterPaul.chem.nutrients$sampledate)
```

```
## [1] "factor"
```

```
class(PeterPaul.gather.nutrients$sampledate)
```

```
## [1] "factor"
```

```
class(Litter.mass.trap$collectDate)
```

```
## [1] "factor"
```

```
Litter.mass.trap$collectDate <-
  as.Date(Litter.mass.trap$collectDate, format ="%Y-%m-%d")
PeterPaul.gather.nutrients$sampledate <-
  as.Date(PeterPaul.gather.nutrients$sampledate, format = "%Y-%m-%d" )
PeterPaul.chem.nutrients$sampledate <-
  as.Date(PeterPaul.chem.nutrients$sampledate, format = "%Y-%m-%d")
```

Define your theme

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

```
mytheme <- theme_classic(base_size = 12) +
  theme(axis.text = element_text(color = "black"),
        legend.position = "top")
theme_set(mytheme)
```

Create graphs

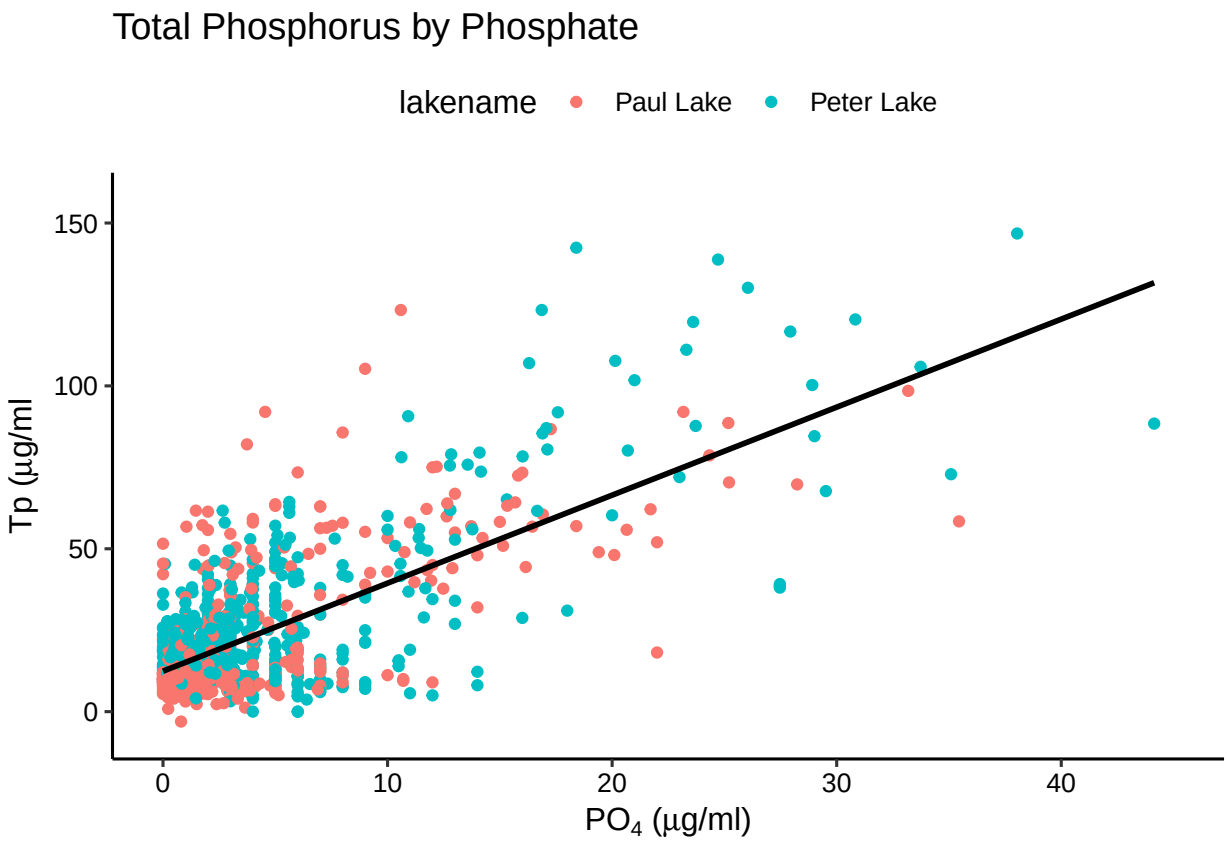
For numbers 4-7, create ggplot graphs and adjust aesthetics to follow best practices for data visualization. Ensure your theme, color palettes, axes, and additional aesthetics are edited accordingly.

4. [NTL-LTER] Plot total phosphorus 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.

```
Phosph <- ggplot (PeterPaul.chem.nutrients, aes(po4, tp_ug, color = lakename))+
              geom_point() +
              geom_smooth(method = "lm", color="black", se =F) +
              xlim(0,45) +
              xlab(expression("PO"[4]*" ("* mu*"g/ml)")) +
              ylab(expression("Tp ("*mu*"g/ml")) +
              labs(title = "Total Phosphorus by Phosphate")
print(Phosph)
```

```
## Warning: Removed 21947 rows containing non-finite values (stat_smooth).
```

```
## Warning: Removed 21947 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.

```
box.temperature <- ggplot(PeterPaul.chem.nutrients)+
                    geom_boxplot(aes(factor(month), temperature_C,
                                     color = lakename))
box.tp <- ggplot(PeterPaul.chem.nutrients,
                         aes(factor(month), tp_ug, color = lakename))+
                    geom_boxplot()
box.TN <- ggplot(PeterPaul.chem.nutrients,
                         aes(factor(month), tn_ug, color = lakename))+
                    geom_boxplot()
box.all <- plot_grid(box.temperature + theme(legend.position="none"),
          box.tp + theme(legend.position="none"),
          box.TN + theme(legend.position="none"),
          align = 'v', nrow = 3)
```

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

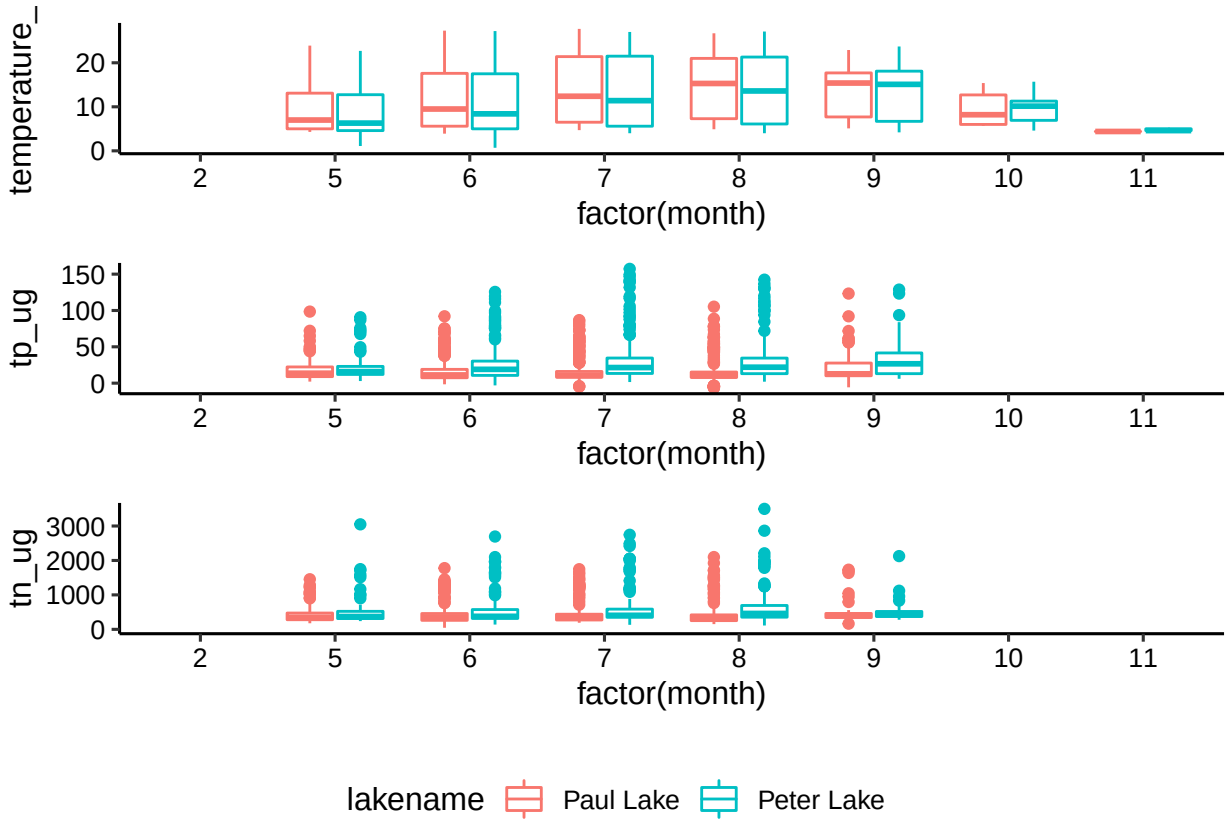
```
## Warning: Removed 20729 rows containing non-finite values (stat_boxplot).
```

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

```
legend_box <- get_legend(box.TN +
        theme(legend.position = "bottom"))
```

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

```
plot_grid(box.all, legend_box, ncol = 1, rel_heights = c(1, .2 ))
```



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

Answer:

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

Answer: