Assignment 5: Data Visualization

Kristine Swann

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

PeterPaul.chem.nutrients <-
  read.csv("./Data/Processed/NTL-LTER_Lake_Chemistry_Nutrients_PeterPaul_Processed.csv")
PeterPaul.chem.nutrients.gathered <-
  read.csv("./Data/Processed/NTL-LTER_Lake_Nutrients_PeterPaulGathered_Processed.csv")
litter <- read.csv("./Data/Processed/NEON_NIWO_Litter_mass_trap_Processed.csv")

#2
class(PeterPaul.chem.nutrients.gathered$sampledate)
```

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

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

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

```
PeterPaul.chem.nutrients$sampledate <- as.Date(
  PeterPaul.chem.nutrients$sampledate, format = "%Y-%m-%d")
PeterPaul.chem.nutrients.gathered$sampledate <- as.Date(
  PeterPaul.chem.nutrients.gathered$sampledate, format = "%Y-%m-%d")
class(PeterPaul.chem.nutrients.gathered$sampledate)
```

```
## [1] "Date"
```

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

```
## [1] "Date"
```

```
class(litter$collectDate)
```

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

```
litter$collectDate <- as.Date(litter$collectDate, format = "%Y-%m-%d")
```

Define your theme

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

```
blahtheme <- theme_classic(base_size = 14) +
  theme(axis.text = element_text(color = "salmon"), axis.ticks = element_line(colour = 'darkseagreen4',
        legend.position = "bottom")


theme_set(blahtheme)
```

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.

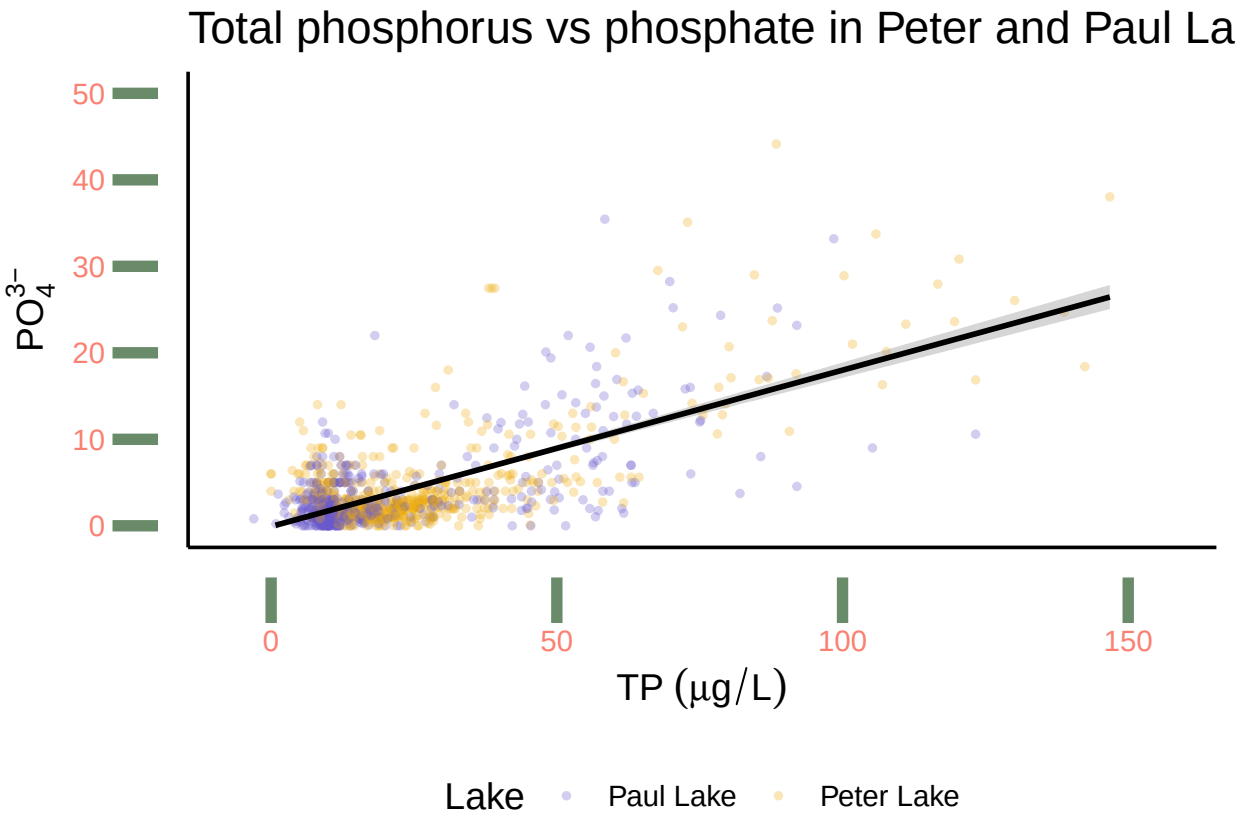
```
pxp <- ggplot(PeterPaul.chem.nutrients, aes(x = tp_ug, y = po4, color=lakename))+
  scale_color_manual(name = "Lake", values = c("Paul Lake" = "slateblue3", "Peter Lake" = "darkgoldenroc
  geom_point(alpha=0.3, size=1)+
  ylim(0,50) +
  geom_smooth(method = lm, color='black')+
 labs( x = expression(TP ~ (mu*g / L)), y = expression(PO[4]^{"3-"}), title= "Total phosphorus vs phospl

print(pxp)
```

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

```
## Warning: Removed 21947 rows containing missing values (geom_point).
```

```
## Warning: Removed 2 rows containing missing values (geom_smooth).
```



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.

```
library(cowplot)
```

```
##
## ************************************************************
## Note: As of version 1.0.0, cowplot does not change the
##   default ggplot2 theme anymore. To recover the previous
##   behavior, execute:
##   theme_set(theme_cowplot())
## ************************************************************
```

```
class(PeterPaul.chem.nutrients$month)
```

```
## [1] "integer"
```

```
PeterPaul.chem.nutrients$month <- factor(PeterPaul.chem.nutrients$month )
class(PeterPaul.chem.nutrients$month)
```

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

```
plot_wq <- function(ystuff) {
  ystuff <- enquo(ystuff)
  ggplot(PeterPaul.chem.nutrients, aes(x= month, !!ystuff, color=lakename)) +
    geom_boxplot() +
```

```
  labs(x = "Month", color = "Lake") +
  scale_y_continuous(expand = c(0, 0)) +
 scale_color_brewer(palette = "Dark2")+
    theme(plot.margin = margin(6, 0, 6, 0))
}

# make three plots
p1 <- plot_wq(temperature_C)+labs(y = "Temp (C)")
p2 <- plot_wq(tp_ug)+ labs(y = expression(TP ~ (mu*g / L)))
p3 <- plot_wq(tn_ug)+ labs(y = expression(TN ~ (mu*g / L)))

#make title
title <- ggplot()+
  labs(title = "Some water quality parameters by lake and month")
# make legend
legend <- get_legend(p1 +
    guides(color = guide_legend(nrow = 1)) +
    theme(legend.position = "bottom"))
```

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

```
# arrange the three plots in a single row
threescompany <- plot_grid(
  p1 + theme(legend.position="none"),
  p2 + theme(legend.position="none"),
  p3 + theme(legend.position="none"),
  align = 'vh',
  hjust = -1,
  nrow = 1
)
```

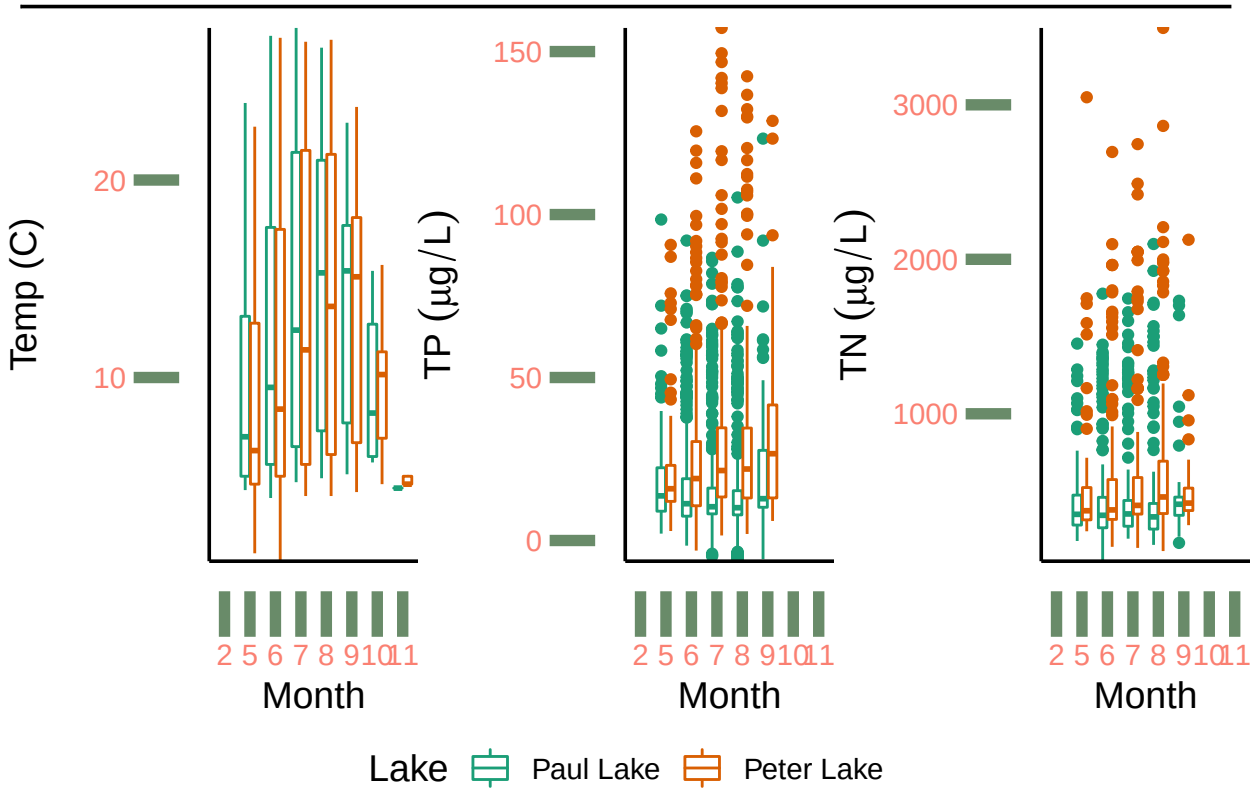
```
## Warning: Removed 3566 rows containing non-finite values (stat_boxplot).

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

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

```
# add the titl and legend
tada <- plot_grid(title, threescompany, legend, ncol = 1, rel_heights = c(.1, 1, .1))

print(tada)
```



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

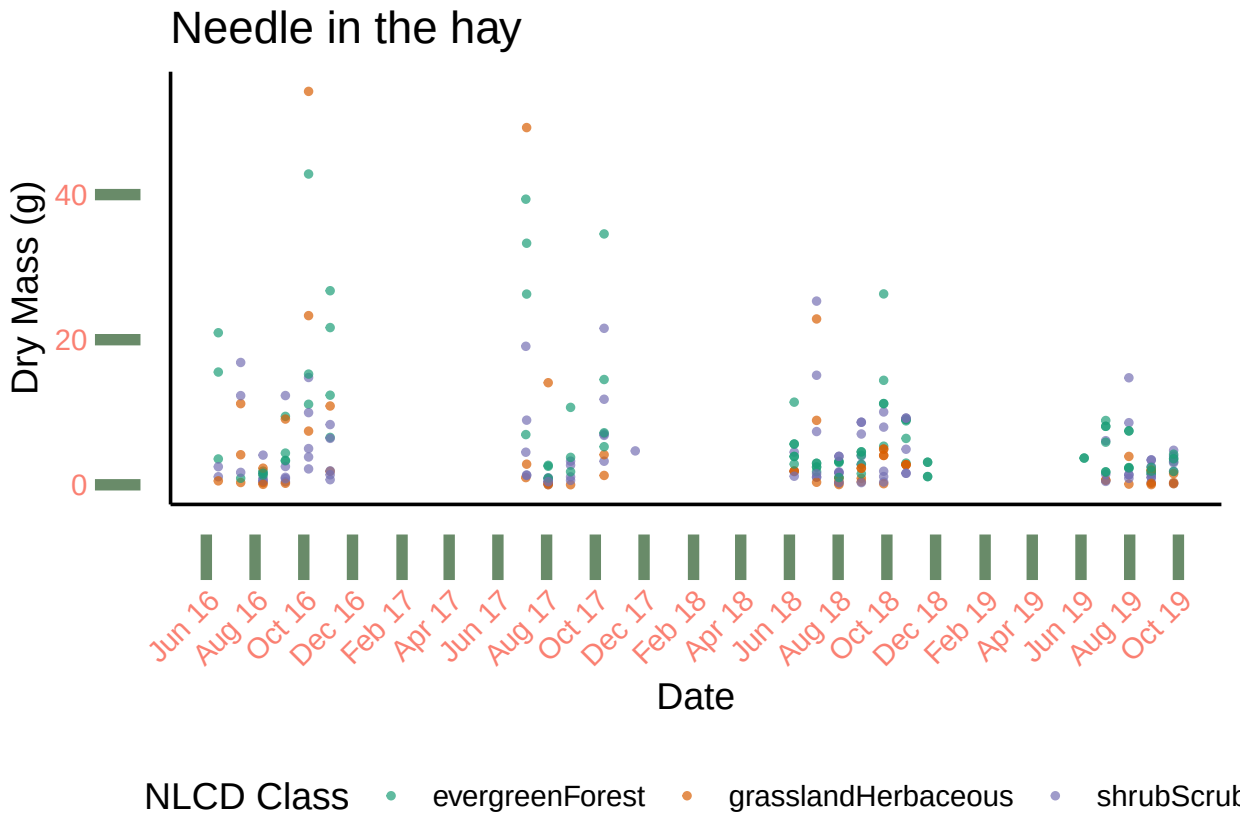
Answer: Temperature looks like it has no outliers and the two lakes are fairly similar in their distribution. TP and TN distributions are higher for Peter Lake in summer months, and both nutrient measures have a lot of right-tailed outliers.

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)

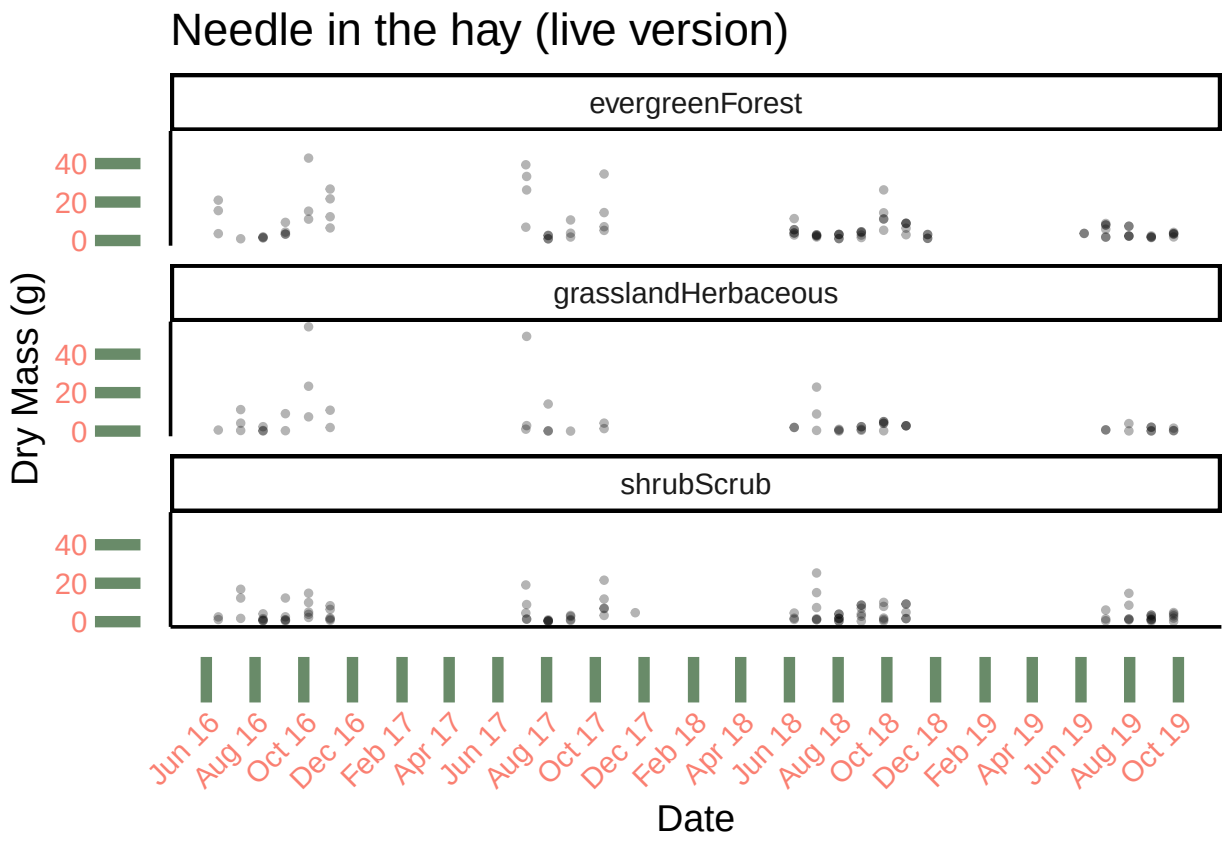
7. [Niwot Ridge] Now, plot the same plot but with NLCD classes separated into three facets rather than separated by color.

```
needle1 <- ggplot(filter(litter, functionalGroup=="Needles"), aes(x = collectDate, y = dryMass, color=nl
  scale_color_brewer(palette = "Dark2")+
  geom_point(alpha=0.7, size=1)+
 labs( x = "Date", y = "Dry Mass (g)", title= "Needle in the hay", color = "NLCD Class")+
  scale_x_date(date_breaks = "2 months", date_labels = "%b %y")+
  theme(axis.text.x = element_text(angle = 45, hjust = 1))

print(needle1)
```




```
needle2 <- ggplot(filter(litter, functionalGroup=="Needles"), aes(x = collectDate, y = dryMass))+
  geom_point(alpha=0.3, size=1)+
 labs( x = "Date", y = "Dry Mass (g)", title= "Needle in the hay (live version)")+
  scale_x_date(date_breaks = "2 months", date_labels = "%b %y")+
  theme(axis.text.x = element_text(angle = 45, hjust = 1))+
facet_wrap(vars(nlcdClass), nrow = 3)
print(needle2)
```



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

Answer: Plot 6 isn't great, but it's better than the faceted one. You can't easily compare the data in 7. Neither are great for assessing differences by NLCD because it's just the raw data scattered by date. It would be more functional to do boxplots or violin plots by NLCD by month or year. (BTW, title here is a Elliott Smith reference)