Assignment 5: Data Visualization

Mandy Hooks

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., "Fay_A05_DataVisualization.Rmd") prior to submission.

The completed exercise is due on Tuesday, February 23 at 11:59 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 (both the tidy [`NTL-LTER_Lake_Chemistry_Nutrients_PeterPaul_Processed.csv`] and the gathered [`NTL-LTER_Lake_Nutrients_PeterPaulGathered_Processed.csv`] versions) 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
getwd()
```

```
## [1] "/Users/mandyhooks/Environmental_Data_Analytics_2021/Assignments"
```

```
library(tidyverse)
```

```
## -- Attaching packages --------------------------------------- tidyverse 1.3.0 --
```

```
## v ggplot2 3.3.3     v purrr   0.3.4
## v tibble  3.0.6     v dplyr   1.0.4
## v tidyr   1.1.2     v stringr 1.4.0
## v readr   1.4.0     v forcats 0.5.1
```

```
## -- Conflicts ------------------------------------------ tidyverse_conflicts() --
## x dplyr::filter() masks stats::filter()
## x dplyr::lag()    masks stats::lag()
```

```
library(cowplot)
library(ggplot2)

LakeChem<-read.csv("/Users/mandyhooks/Environmental_Data_Analytics_2021//Data/Processed/NTL-LTER_Lake_C
GathNut<-read.csv("/Users/mandyhooks/Environmental_Data_Analytics_2021//Data/Processed/NTL-LTER_Lake_Nu
```

```
NiwotRidge<-read.csv("/Users/mandyhooks/Environmental_Data_Analytics_2021//Data/Processed/NEON_NIWO_Lit

#2

LakeChem$sampledate<-as.Date(LakeChem$sampledate, format = "%Y-%m-%d")
GathNut$sampledate<-as.Date(GathNut$sampledate, format = "%Y-%m-%d")
NiwotRidge$collectDate<-as.Date(NiwotRidge$collectDate, format = "%Y-%m-%d")
```

Define your theme

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

```
A5theme <- theme_classic(base_size = 20) +
  theme(axis.text = element_text(color = "black"),
        legend.position = "top")
```

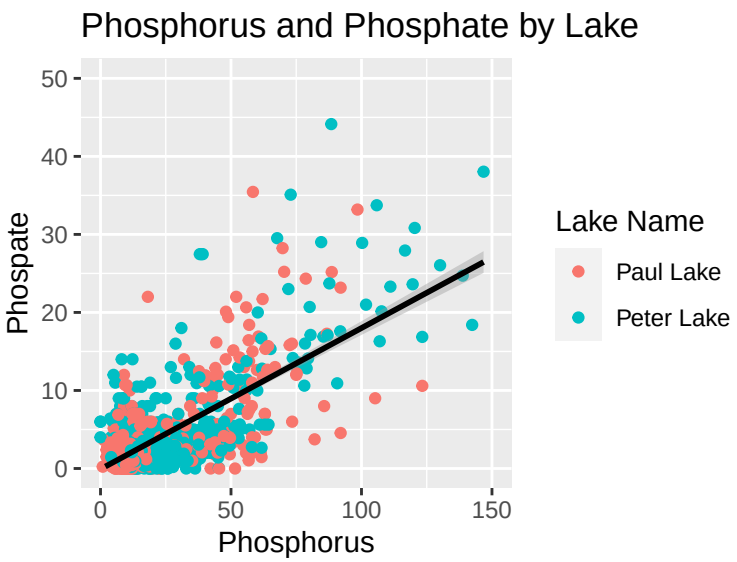
Create graphs

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.

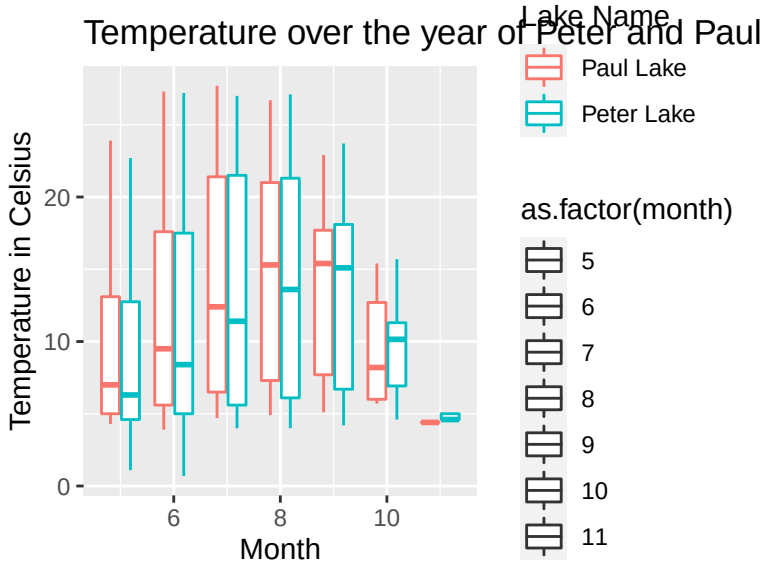
```
phosplot<- ggplot(LakeChem, aes(x=tp_ug, y=po4, color=lakename))+
  geom_point(alpha=1)+
  geom_smooth(method=lm, color="black")+
  ylim(0,50)+
  xlim(0,150)+
  labs(x="Phosphorus", y="Phospate", title="Phosphorus and Phosphate by Lake", color="Lake Name")
print(phosplot)
```

```
## `geom_smooth()` using formula 'y ~ x'
```

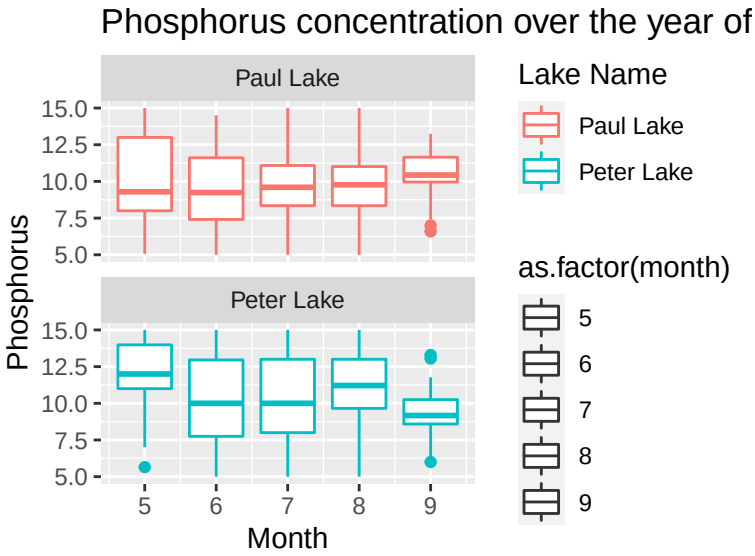


5. [NTL-LTER] Make three separate boxplots of (a) temperature, (b) TP, and (c) TN, with month as the x axis and lake as a color aesthetic. Then, create a cowplot that combines the three graphs. Make sure that only one legend is present and that graph axes are aligned.

```
#Temperature Boxplot
boxtemp<-ggplot(LakeChem, aes(x=month, y=temperature_C, shape=as.factor(month), color=lakename))+
  geom_boxplot()+
  labs(x="Month", y="Temperature in Celsius", color="Lake Name", title="Temperature over the year of Pe
print(boxtemp)
```

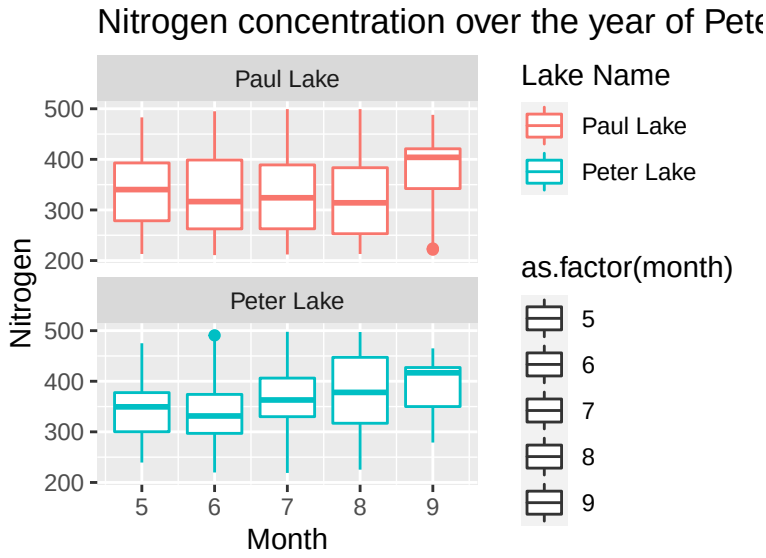



```
#TP Boxplot
boxtp<- ggplot(LakeChem, aes(x=month, y=tp_ug, shape= as.factor(month), color=lakename))+
  geom_boxplot()+
  ylim(5,15)+
  facet_wrap(vars(lakename), nrow = 2)+
  labs(x="Month", y="Phosphorus", color="Lake Name", title="Phosphorus concentration over the year of P
print(boxtp)
```




```
#TN Boxplot
boxtn<- ggplot(LakeChem, aes(x=month, y=tn_ug, color= lakename, shape =as.factor(month)))+
  geom_boxplot()+
  ylim(210,500)+
```

```
  facet_wrap(vars(lakename), nrow = 2)+
  labs(x="Month", y="Nitrogen", color="Lake Name", title="Nitrogen concentration over the year of Peter
print(boxtn)
```



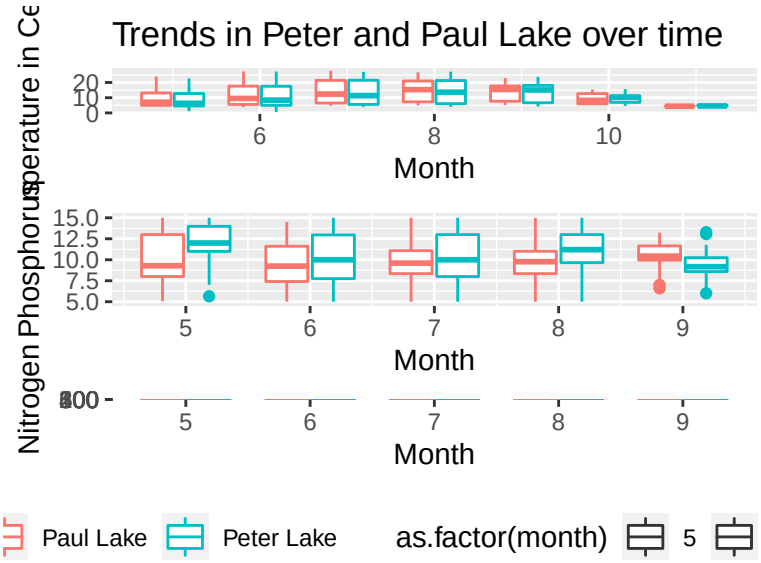

```
#Cowplot

p1<-ggplot(LakeChem, aes(x=month, y=temperature_C, color=lakename, shape= as.factor(month)))+
  geom_boxplot()+ #I don't think you need aes here again, can you remove and try?
  #facet_wrap(vars(lakename), nrow = 2)+
  labs(x="Month", y="Temperature in Celsius", color="Lake Name", title = "Trends in Peter and Paul Lake

p2<-ggplot(LakeChem, aes(x=month, y=tp_ug, color=lakename, shape= as.factor(month)))+
  geom_boxplot()+ #I don't think you need aes here again, can you remove and try?
  ylim(5,15)+
  #facet_wrap(vars(lakename), nrow = 2)+
  labs(x="Month", y="Phosphorus", color="Lake Name")

p3<-ggplot(LakeChem, aes(x=month, y=tn_ug, color=lakename, shape=as.factor(month)))+
  geom_boxplot()+ #I don't think you need aes here again, can you remove and try?
  ylim(120,600)+
  #facet_wrap(vars(lakename), nrow = 2)+
  labs(x="Month", y="Nitrogen", color="Lake Name")


prow <- plot_grid(
  p1 + theme(legend.position="none"),
  p2 + theme(legend.position="none"),
  p3 + theme(legend.position="bottom"),
  align="hv",
  ncol=1,
  rel_widths=c(1,1,1)
)
prow
```

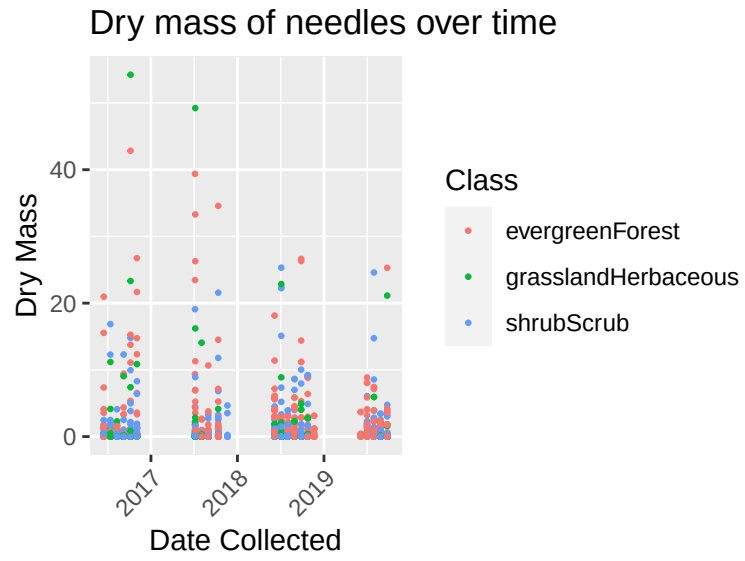


Question: What do you observe about the variables of interest over seasons and between lakes?

Answer: Overall temperartures and concentrations of the chemicals tended to be lower in Paul Lake, while Peter Lake was overall warmer and had higher concentrations of the chemicals.

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)

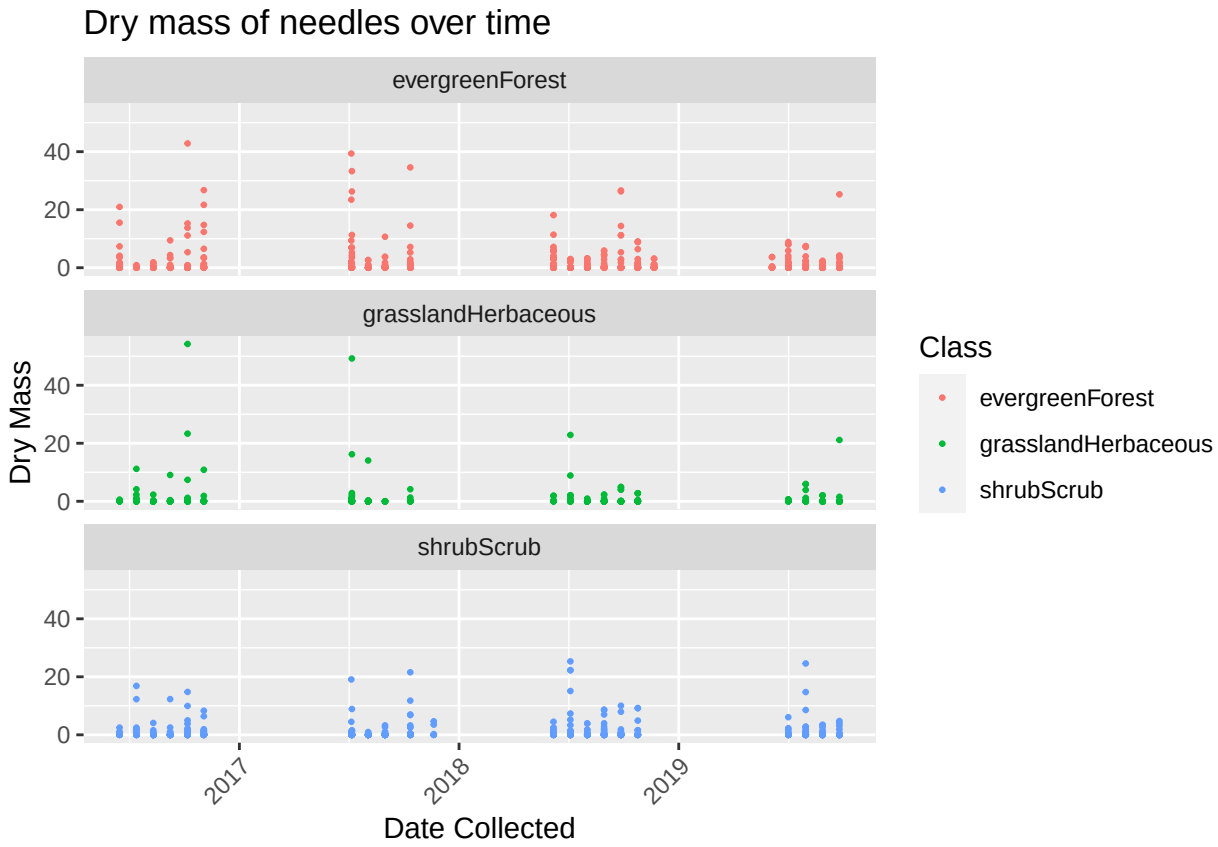
```
ggplot(subset(NiwotRidge, functionalGroup=Needles), aes(x=collectDate, y=dryMass, color=nlcdClass))+
  geom_point(size=.5)+
  labs(title="Dry mass of needles over time", x="Date Collected", y="Dry Mass", color="Class")+
  theme(axis.text.x = element_text(angle=45, hjust=1))
```



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

```
ggplot(subset(NiwotRidge, functionalGroup=Needles), aes(x=collectDate, y=dryMass, color=nlcdClass))+
  geom_point(size=.5)+
  labs(title="Dry mass of needles over time", x="Date Collected", y="Dry Mass", color="Class")+
```

```
theme(axis.text.x = element_text(angle=45, hjust=1))+
facet_wrap("nlcdClass",nrow=3)
```



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

Answer: I think 7 is because it is easier to compare the different classes, and the plots are less crowded and overwhelming than in 6.