Assignment 5: Data Visualization

Student Name

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.

18 points total

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.

1 point 1/2 for packages, 1/2 for datasets

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

1 point 1/2 point missing for incorrect or if all three datasets are not changed

```
#1
library(tidyverse)
library(cowplot)
Litter <- read.csv("../Data/Processed/NEON_NIWO_Litter_mass_trap_Processed.csv")
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")


#2
Litter$collectDate <- as.Date(Litter$collectDate, "%Y-%m-%d")
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")
```

Define your theme

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

1 point

```
#3
theme_set(theme_classic())
```

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.

5 points, 1 point each:

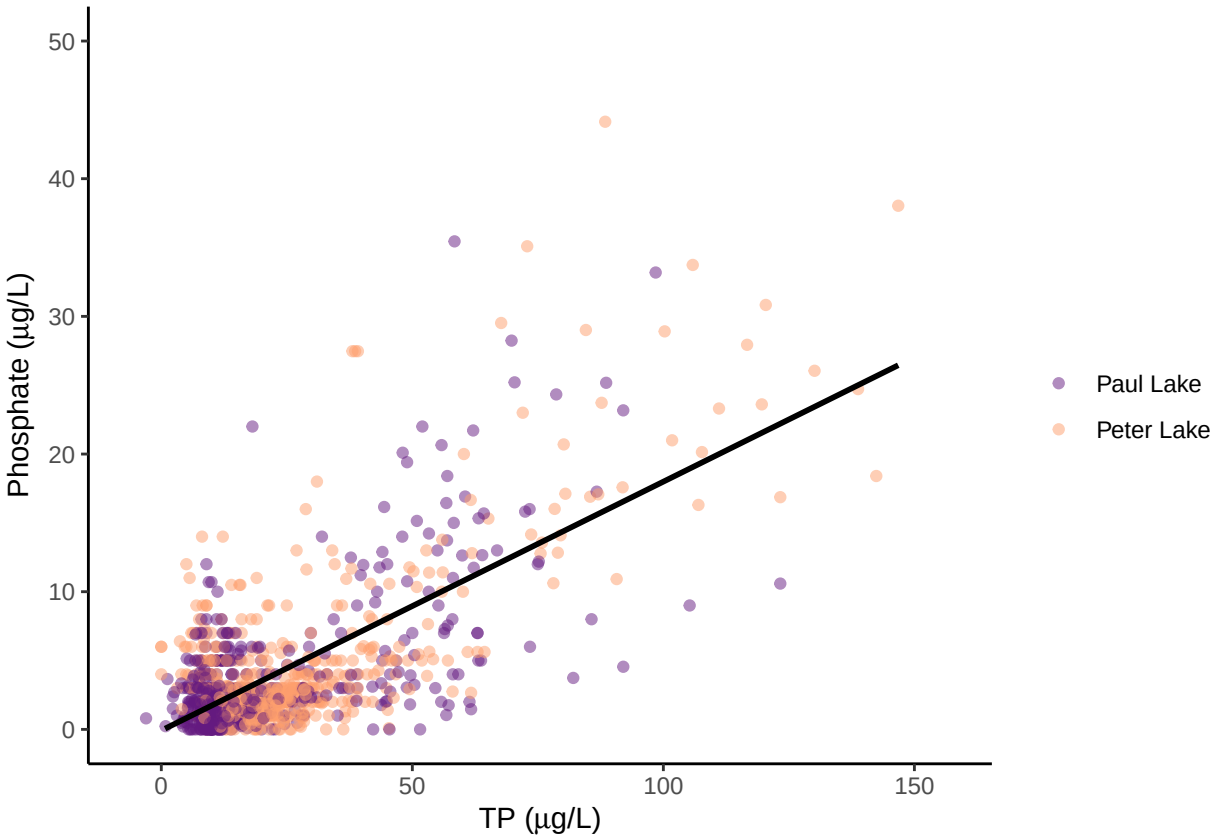
- correct dataset and variables chosen (TP, PO4, lakename)
- color or other aesthetic scale for lake edited
- line of best fit: geom_smooth with black color
- x axis, y axis, and color aesthetics have new labels. Symbols and super/subscripts formatted correctly, as needed.
- scale_y_continuous reset near ~50

```
#4
PhosphorusPlot <-
ggplot(PeterPaul.chem.nutrients, aes(x = tp_ug, y = po4, color = lakename)) +
  geom_point(alpha = 0.5) +
  geom_smooth(method = "lm", se = FALSE, color = "black") +
  scale_y_continuous(limits = c(0, 50)) +
  scale_color_viridis_d(option = "magma", begin = 0.3, end = 0.8) +
  labs(x = expression(paste("TP (", mu, "g/L)")), y = expression(paste(Phosphate," (", mu, "g/L)")),
       color = "")
print(PhosphorusPlot)
```

```
## 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 surface (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.

6 points, 1 point each:

- dataset subsetted for surface depths
- correct dataset and variables displayed (month, lake, temp, TP, TN)
- color aesthetic chosen and consistent
- x axis, y axis, and color aesthetics have new labels. Symbols and super/subscripts formatted correctly, as needed.
- only one legend appears
- graph axes aligned with align function in cowplot

```
#5
Tempbymonth <-
ggplot(subset(PeterPaul.chem.nutrients, depth < 0.25)) +
  geom_boxplot(aes(x = as.factor(month), y = temperature_C, color = lakename)) +
  theme(legend.position = "top") +
  scale_color_viridis_d(option = "magma", begin = 0.3, end = 0.8) +
  labs(x = "Month", y = expression("Temperature ("*~degree*C*")"), color = "")

TPbymonth <-
ggplot(subset(PeterPaul.chem.nutrients, depth < 0.25)) +
  geom_boxplot(aes(x = as.factor(month), y = tp_ug, color = lakename)) +
  theme(legend.position = "none") +
  scale_color_viridis_d(option = "magma", begin = 0.3, end = 0.8) +
  labs(x = "Month", y = expression(paste("TP (", mu, "g/L)")))
```

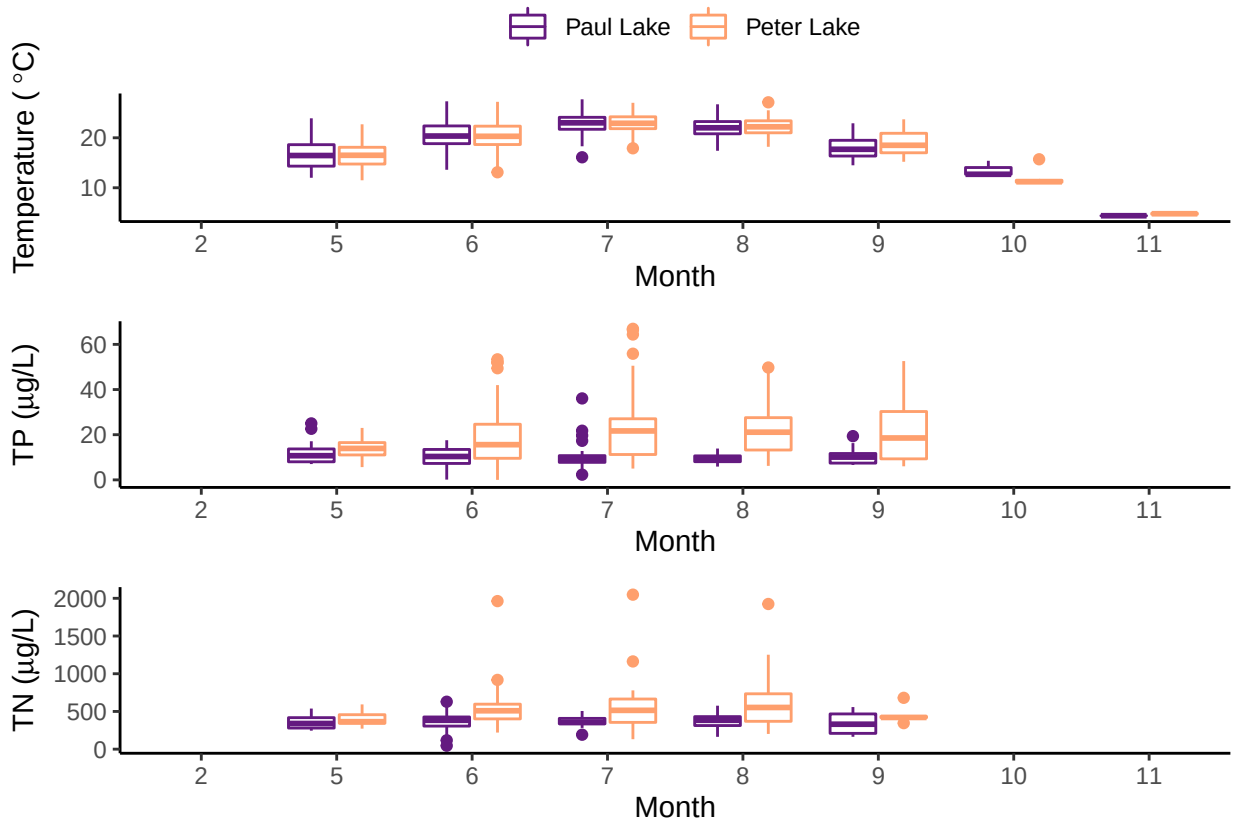
```
TNbymonth <-
ggplot(subset(PeterPaul.chem.nutrients, depth < 0.25)) +
  geom_boxplot(aes(x = as.factor(month), y = tn_ug, color = lakename)) +
  theme(legend.position = "none") +
  scale_color_viridis_d(option = "magma", begin = 0.3, end = 0.8) +
  labs(x = "Month", y = expression(paste("TN (", mu, "g/L)")))

plot_grid(Tempbymonth, TPbymonth, TNbymonth, nrow = 3,
          align = "v", rel_heights = c(1.25, 1, 1))
```

```
## Warning: Removed 17 rows containing non-finite values (stat_boxplot).
```

```
## Warning: Removed 806 rows containing non-finite values (stat_boxplot).
```

```
## Warning: Removed 872 rows containing non-finite values (stat_boxplot).
```



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

3 points, 1 point each:

> Answer: Temperature higher in summer months, little differences between lakes. TP fairly consistent across months, Peter Lake higher than Paul. TN fairly sonsistent across months, Peter Lake higher than Paul.

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)

4 points, 1 point each:

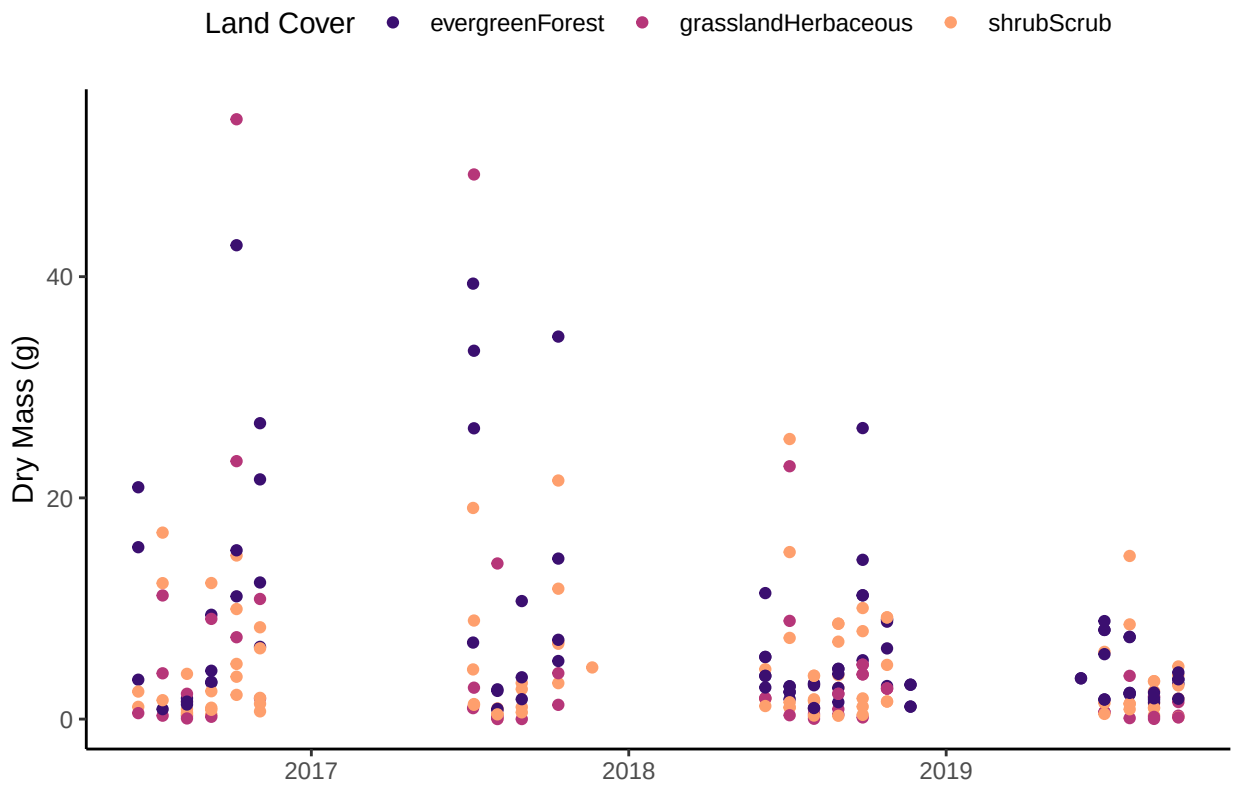
- correct dataset and variables (date, mass, land cover class)
- subsetted with just needles
- color aesthetic adjusted
- axes and color aesthetics labels adjusted accordingly

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

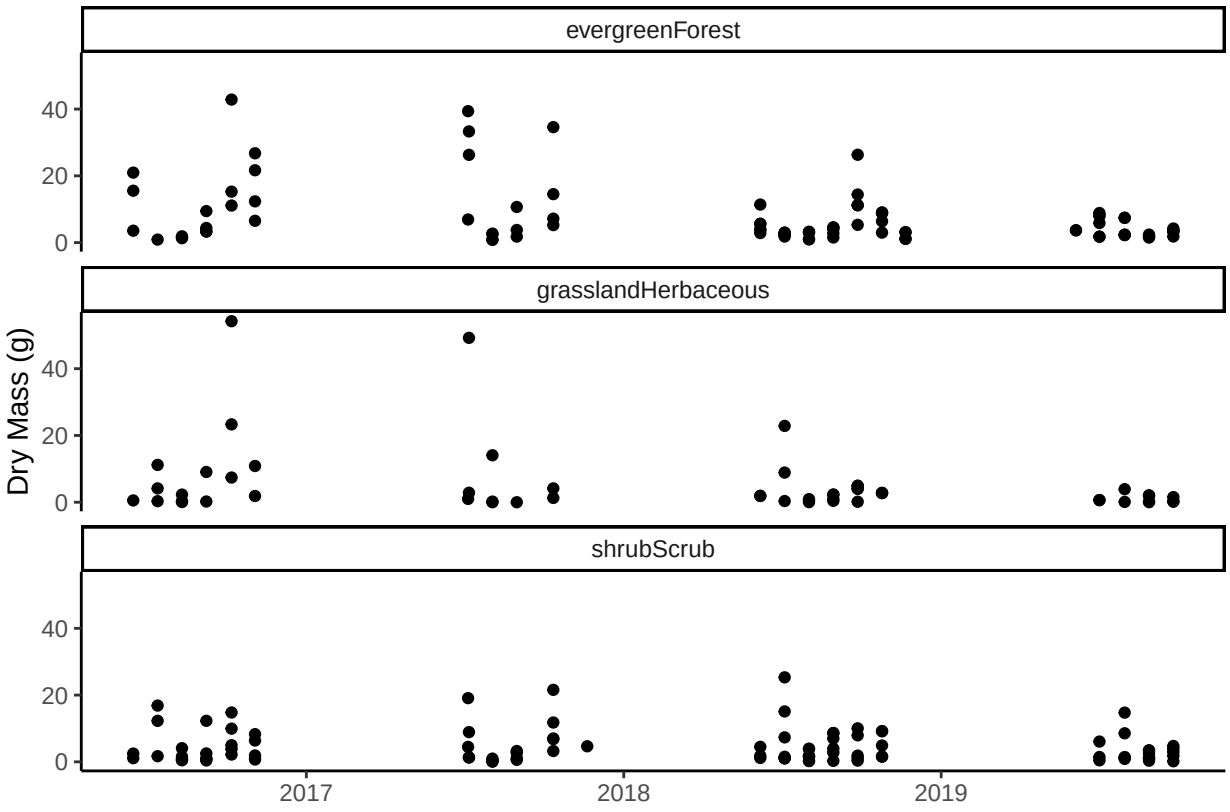
1 point

- formatted same as previous graph, but facet_grid or facet_wrap is substituted for a color scale

```
Needleplot1 <-
ggplot(subset(Litter, functionalGroup == "Needles")) +
  geom_point(aes(x = collectDate, y = dryMass, color = nlcdClass)) +
  scale_color_viridis_d(option = "magma", begin = 0.2, end = 0.8) +
  theme(legend.position = "top") +
  labs(x = "", y = "Dry Mass (g)", color = "Land Cover")
print(Needleplot1)
```




```
Needleplot2 <-
ggplot(subset(Litter, functionalGroup == "Needles")) +
  geom_point(aes(x = collectDate, y = dryMass)) +
  facet_wrap(vars(nlcdClass), nrow = 3) +
  labs(x = "", y = "Dry Mass (g)")
 print(Needleplot2)
```



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

1 point for thoughtful answer

Answer: