

Assignment 5: Data Visualization

Fei Wu

Spring 2024

OVERVIEW

This exercise accompanies the lessons in Environmental Data Analytics on Data Visualization

Directions

1. Rename this file `<FirstLast>_A05_DataVisualization.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 your code is tidy; use line breaks to ensure your code fits in the knitted output.
 5. Be sure to **answer the questions** in this assignment document.
 6. When you have completed the assignment, **Knit** the text and code into a single PDF file.
-

Set up your session

1. Set up your session. Load the tidyverse, lubridate, here & cowplot packages, and verify your home directory. Read in 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_Processed.csv version in the Processed_KEY folder) and the processed data file for the Niwot Ridge litter dataset (use the NEON_NIWO_Litter_mass_trap_Processed.csv version, again from the Processed_KEY folder).
2. Make sure R is reading dates as date format; if not change the format to date.

#1

```
library(tidyverse);library(lubridate);library(here)
```

```
## -- Attaching core tidyverse packages ----- tidyverse 2.0.0 --
## v dplyr      1.1.3      v readr      2.1.4
## v forcats    1.0.0      v stringr    1.5.0
## v ggplot2     3.4.3      v tibble     3.2.1
## v lubridate  1.9.2      v tidyr      1.3.0
## v purrr      1.0.2
## -- Conflicts ----- tidyverse_conflicts() --
## x dplyr::filter() masks stats::filter()
## x dplyr::lag()     masks stats::lag()
## i Use the conflicted package (<http://conflicted.r-lib.org/>) to force all conflicts to become errors
## here() starts at /home/guest/EDA_Spring2024
```

```
library(cowplot)
```

```
##  
## Attaching package: 'cowplot'  
##  
## The following object is masked from 'package:lubridate':  
##  
## stamp
```

```
here()
```

```
## [1] "/home/guest/EDA_Spring2024"
```

```
Lake_Chemistry <- read.csv(  
  file=here("Data/Processed_KEY/NTL-LTER_Lake_Chemistry_Nutrients_PeterPaul_Processed.csv"),  
  stringsAsFactors = TRUE  
)  
Niwot_Ridge <- read.csv(  
  file=here("Data/Processed_KEY/NEON_NIWO_Litter_mass_trap_Processed.csv"),  
  stringsAsFactors = TRUE  
)  
  
#2  
Lake_Chemistry$sampldate <- ymd(Lake_Chemistry$sampldate)  
Niwot_Ridge$collectDate <- ymd(Niwot_Ridge$collectDate)
```

Define your theme

3. Build a theme and set it as your default theme. Customize the look of at least two of the following:

- Plot background
- Plot title
- Axis labels
- Axis ticks/gridlines
- Legend

```
#3  
my_theme <-  
  theme(  
    line = element_line(  
      color='red',  
      linewidth =2  
    ),  
    legend.background = element_rect(  
      color='grey',  
      fill = 'orange'  
    ),  
    legend.title = element_text(  
      color='blue'  
    )  
  )
```

```
)  
  
theme_set(my_theme)
```

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.

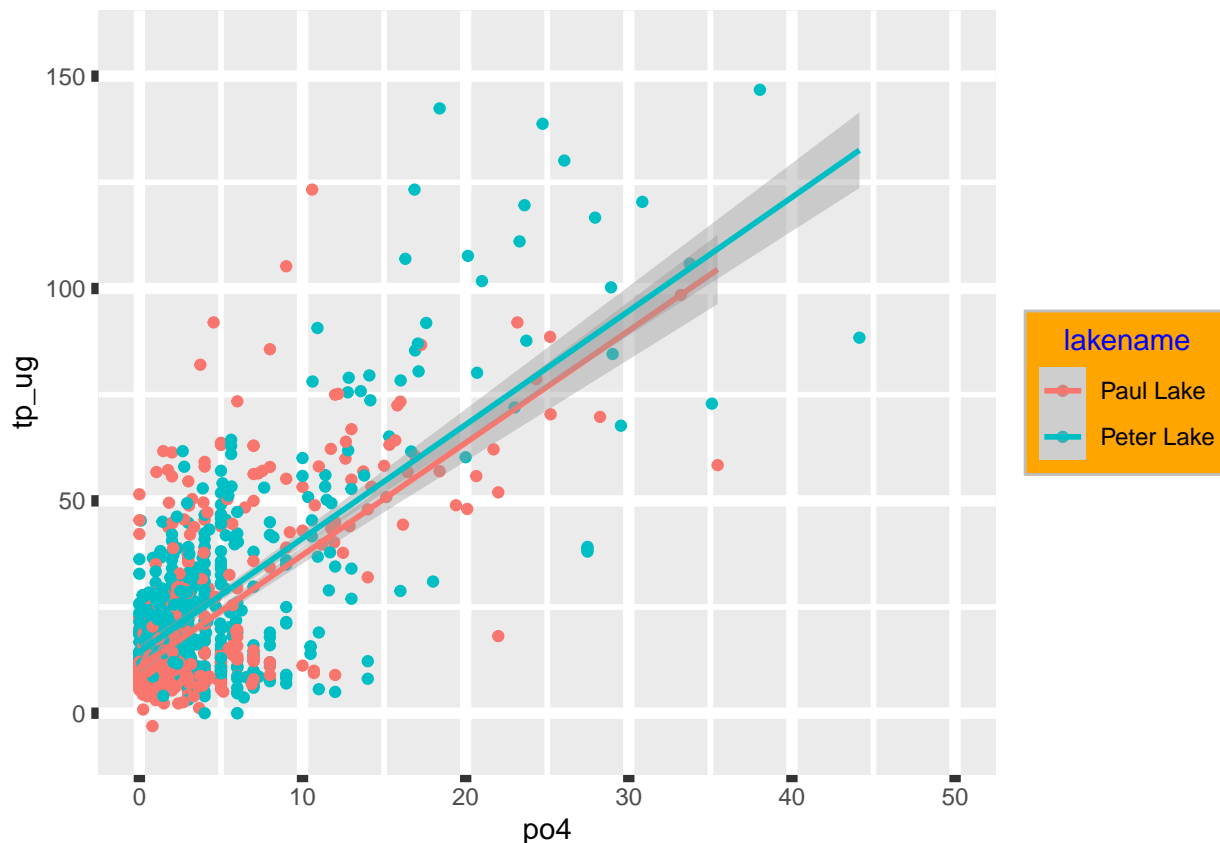
4. [NTL-LTER] Plot total phosphorus (**tp_ug**) by phosphate (**po4**), with separate aesthetics for Peter and Paul lakes. Add line(s) of best fit using the **lm** method. Adjust your axes to hide extreme values (hint: change the limits using **xlim()** and/or **ylim()**).

```
#4  
Lake_Chemistry %>%  
  ggplot(aes(x=po4, y=tp_ug, color=lakename))+  
  geom_point() +  
  geom_smooth(method='lm')+  
  xlim(0,50)
```

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

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

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



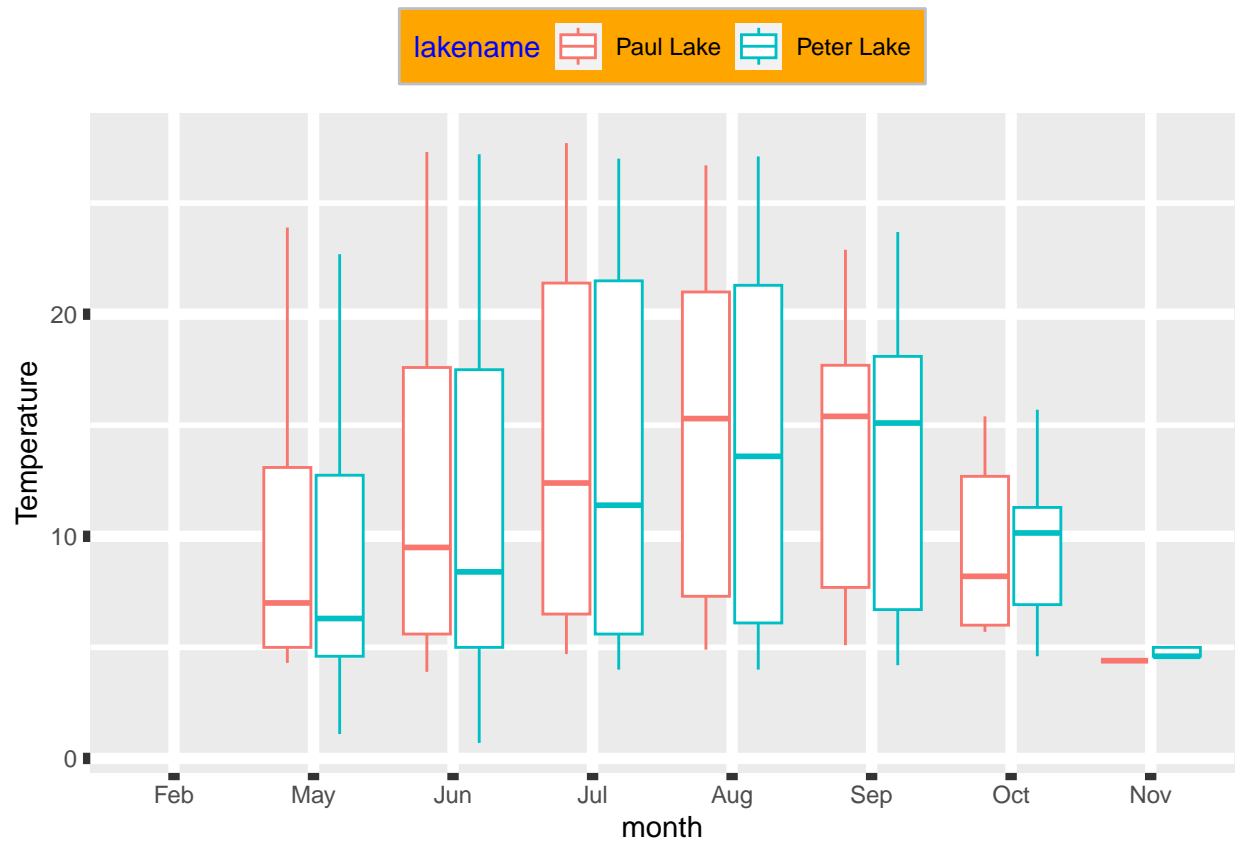
5. [NTL-LTER] Make three separate boxplots of (a) temperature, (b) TP, and (c) TN, with month as the x axis and lake as a color aesthetic. Then, create a cowplot that combines the three graphs. Make sure that only one legend is present and that graph axes are aligned.

Tips: * Recall the discussion on factors in the lab section as it may be helpful here. * Setting an axis title in your theme to `element_blank()` removes the axis title (useful when multiple, aligned plots use the same axis values) * Setting a legend's position to "none" will remove the legend from a plot. * Individual plots can have different sizes when combined using `cowplot`.

#5

```
p1<- Lake_Chemistry %>%
ggplot(aes(x=factor(month, levels=1:12, labels=month.abb),y=temperature_C, color=lakename))+
geom_boxplot()+
labs(x="month", y="Temperature")+theme(legend.position = "top")
p1
```

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



```
p2<- Lake_Chemistry %>%
ggplot(aes(x=factor(month, levels=1:12, labels=month.abb), y=tp_ug,color=lakename))+
geom_boxplot(show.legend=F)+
  labs(x="month",y="TP")

p3<- Lake_Chemistry %>%
ggplot(aes(x=factor(month, levels=1:12, labels=month.abb),y=tn_ug, color=lakename))+
geom_boxplot(show.legend=F) +
  labs(x="month", y="TN")

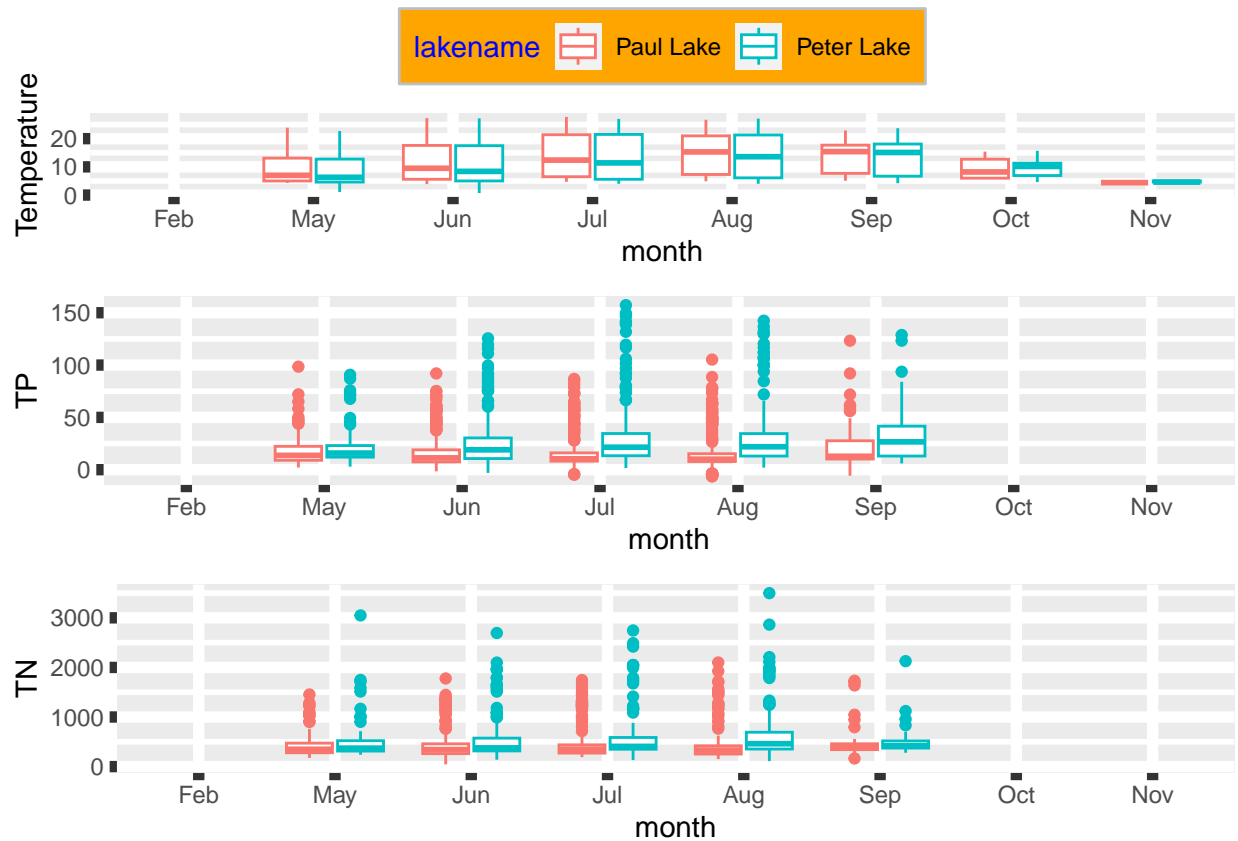
library(cowplot)
plot_grid(p1,p2,p3, nrow=3, align = "h")
```

```
## 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()').
```

```
## Warning: Graphs cannot be horizontally aligned unless the axis parameter is
## set. Placing graphs unaligned.
```



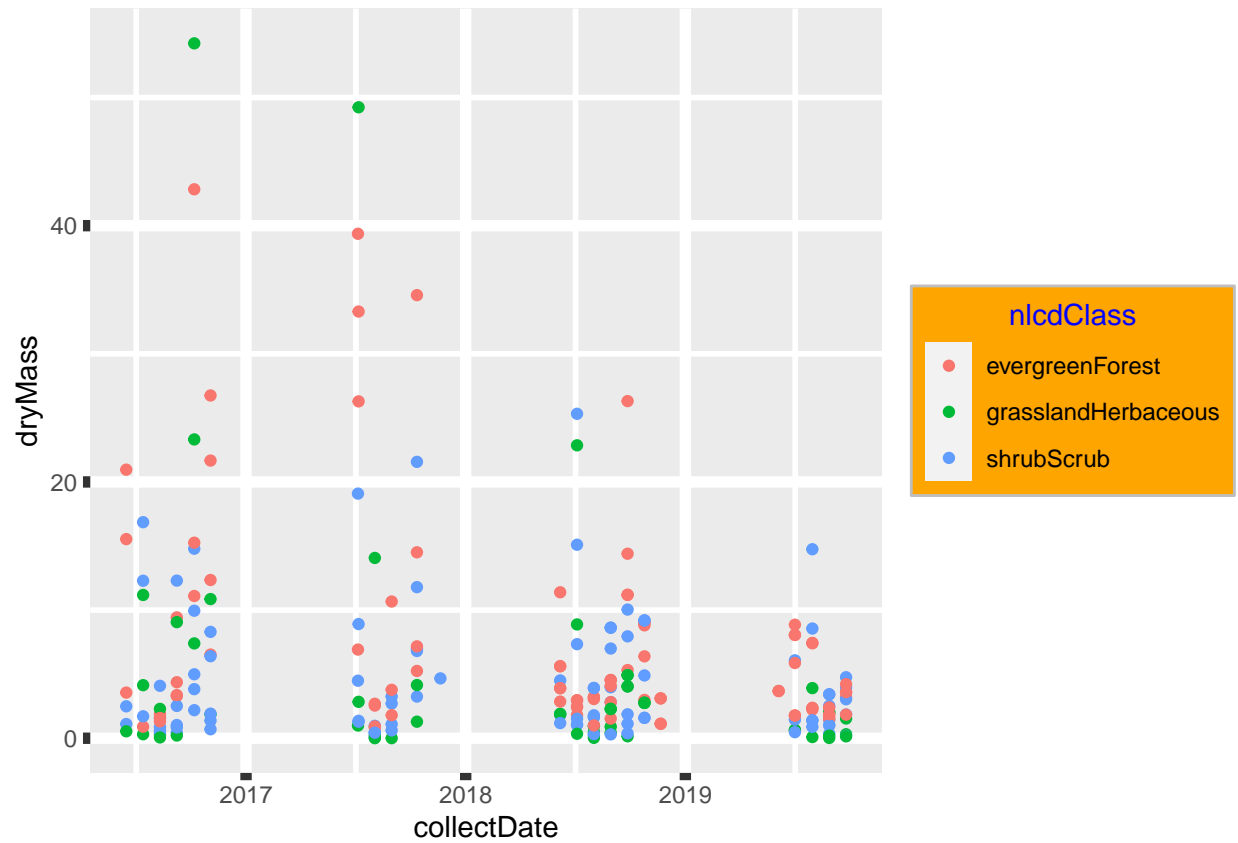
Question: What do you observe about the variables of interest over seasons and between lakes?

Answer: in Summer the tp is very high and the temperature is very high. Peter Lake is especially high compared to Paul lake. Same thing for TN that Peter Lake has higher TN in August and July than Paul lake.

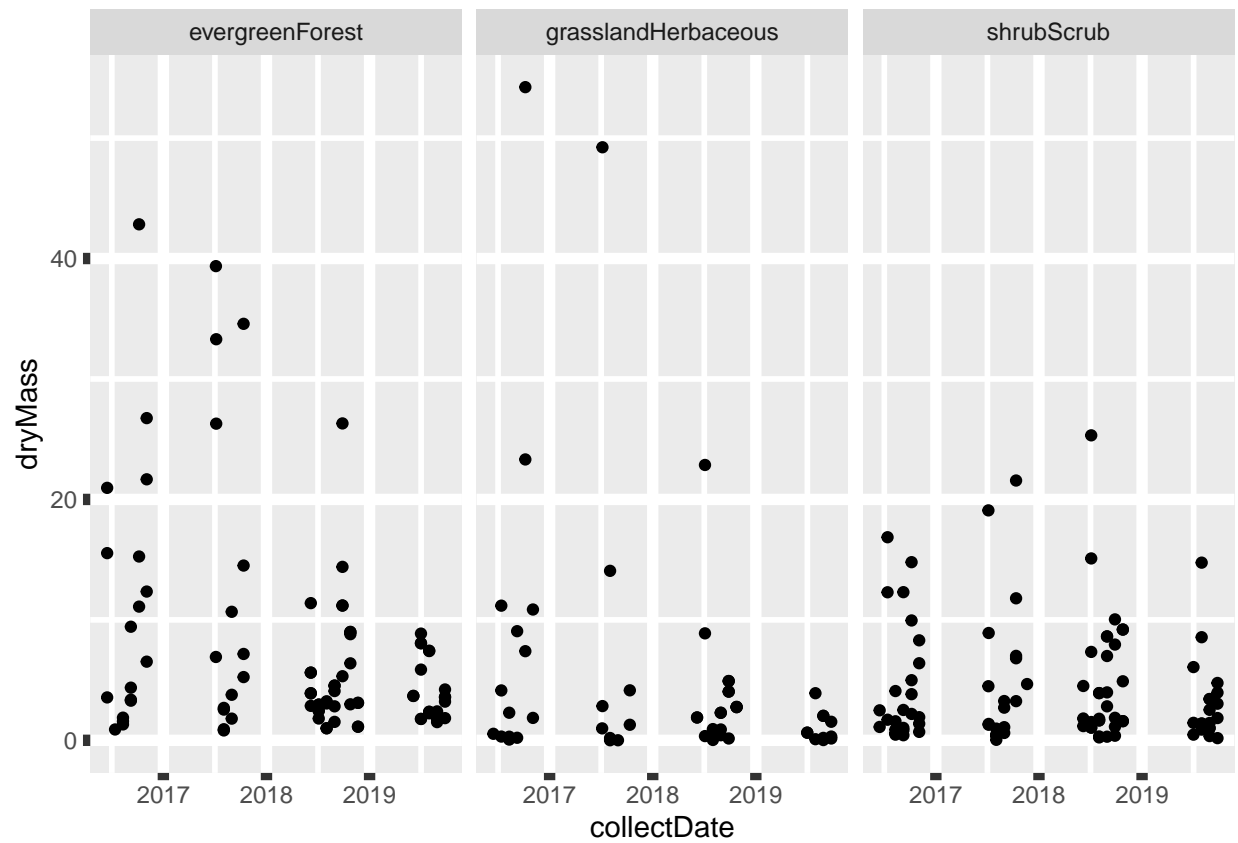
6. [Niwot Ridge] Plot a subset of the litter dataset by displaying only the “Needles” functional group. Plot the dry mass of needle litter by date and separate by NLCD class with a color aesthetic. (no need to adjust the name of each land use)
7. [Niwot Ridge] Now, plot the same plot but with NLCD classes separated into three facets rather than separated by color.

#6

```
Niwot_Ridge %>%
  filter(functionalGroup=="Needles") %>%
  ggplot(aes(x=collectDate, y=dryMass, color=nlcdClass)) +
  geom_point()
```



```
#7
Niwot_Ridge%>%
  filter(functionalGroup=="Needles") %>%
  ggplot(aes(x=collectDate,y=dryMass))+
  geom_point()+
  facet_wrap(vars(nlcdClass))
```



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

Answer: I think the first one with color is more effective because you can clearly compare with three classes. it's easier this way because the color differentiates the different drymass.