Assignment 5: Data Visualization

Aye Nyein Thu

Fall 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.

```
#1a Loading packages
library(tidyverse)
library(lubridate)
library(here)
library(cowplot)

#1b Loading other packages for plots
library(ggridges)
library(viridis)
library(RColorBrewer)
library(colormap)
library(ggthemes)

#1c Verifying home directory
here()
```

```
## [1] "/home/guest/EDE_Fall2024"
```

```
#1d Assigning a string variable to the data location folder
processed_data = './Data/Processed_KEY/'

#1e Reading in data
NTL_LTER <- read.csv(
  here(processed_data,
       "NTL-LTER_Lake_Chemistry_Nutrients_PeterPaul_Processed.csv"),
  stringsAsFactors = TRUE
)

Niwot_Ridge <- read.csv(
  here(processed_data,
     "NEON_NIWO_Litter_mass_trap_Processed.csv"),
  stringsAsFactors = TRUE
)

#2 Changing date formats
class(NTL_LTER$sampledate)
```

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

```
NTL_LTER$sampledate <- ymd(NTL_LTER$sampledate)
class(NTL_LTER$sampledate)
```

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

```
class(Niwot_Ridge$collectDate)
```

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

```
Niwot_Ridge$collectDate <- ymd(Niwot_Ridge$collectDate)
class(Niwot_Ridge$collectDate)
```

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

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

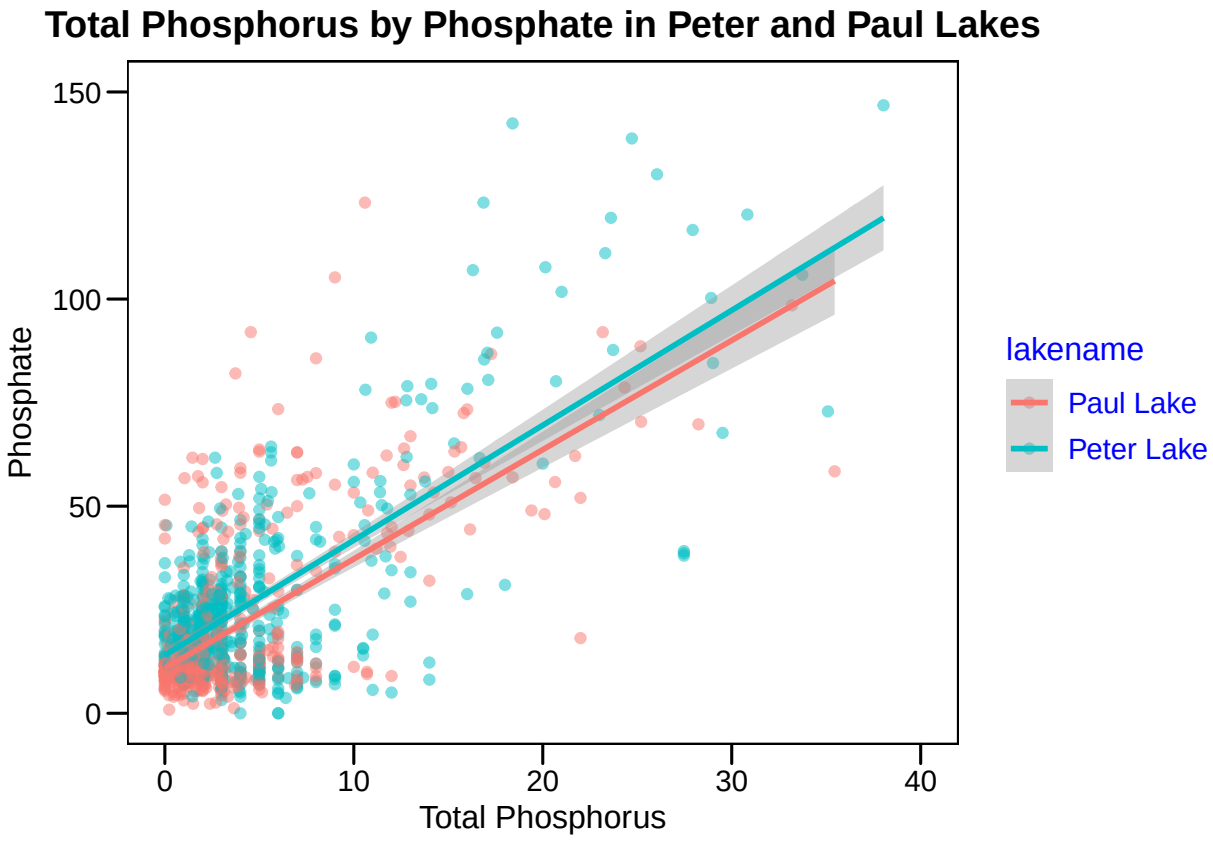
```r
#3 Defining theme
mytheme <- theme_base() +
  theme(
  plot.background = element_blank(),
  plot.title = element_text(size=14, face="bold", hjust = 0.5),
  axis.title = element_text(size=12),
  axis.text = element_text(size=11),
  legend.title = element_text(size=12, color="blue"),
  legend.text = element_text(size=11, color="blue"),
  panel.grid = element_blank(),
  legend.position = "right"
)
```

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()`).

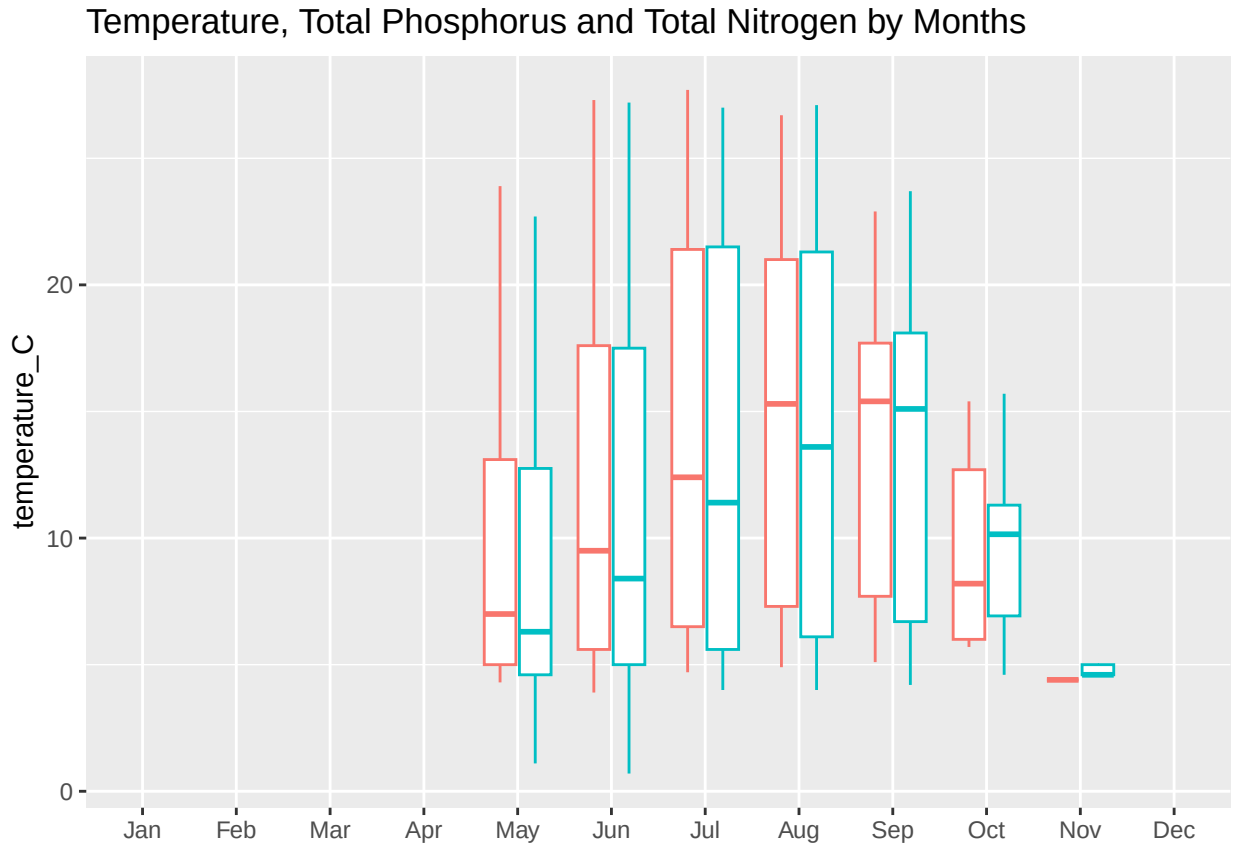
```r
#4 Concentration of Total Phosphorus by Phosphate in Peter and Paul Lakes
Graph4 <- NTL_LTER %>%
  ggplot(aes(x = po4, y = tp_ug,color = lakename)) +
  geom_point(alpha = 0.5) +
  geom_smooth(method = lm) +
  xlim(0,40) +
  ylim(0,150) +
  xlab(expression("Total Phosphorus")) +
  ylab(expression("Phosphate")) +
  ggtitle("Total Phosphorus by Phosphate in Peter and Paul Lakes") +
  mytheme

print(Graph4)
```



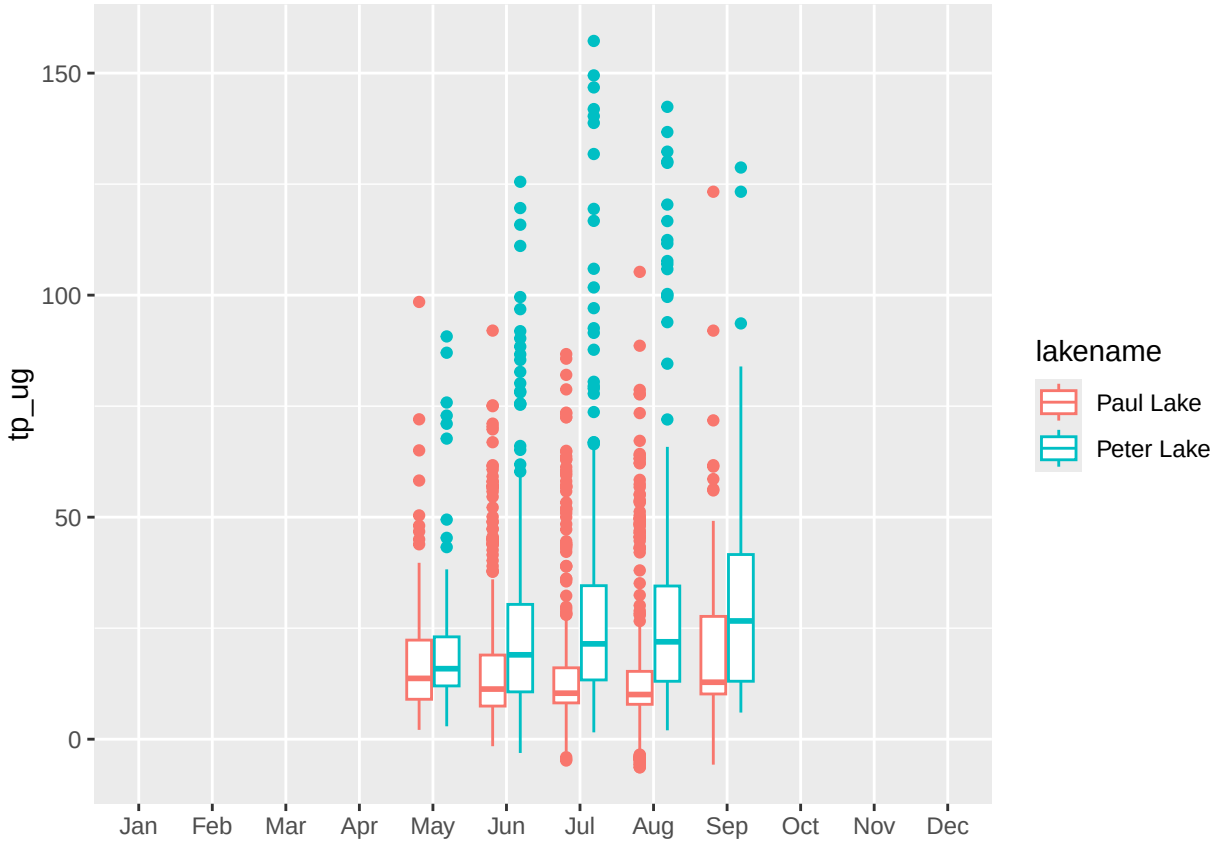
5. [NTL-LTER] Make three separate boxplots of (a) temperature, (b) TP, and (c) TN, with month as the x axis and lake as a color aesthetic. Then, create a cowplot that combines the three graphs. Make sure that only one legend is present and that graph axes are aligned.

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`.

```
#5a Boxplot of Temperature
Graph5a <- NTL_LTER %>%
  ggplot(aes(x = factor(month, levels = 1:12, labels = month.abb),
             y = temperature_C,
             color = lakename)) +
  scale_x_discrete(drop = F) +
  geom_boxplot() +
  theme(axis.title.x = element_blank(), legend.position = "none") +
  ggtitle("Temperature, Total Phosphorus and Total Nitrogen by Months")

print(Graph5a)
```

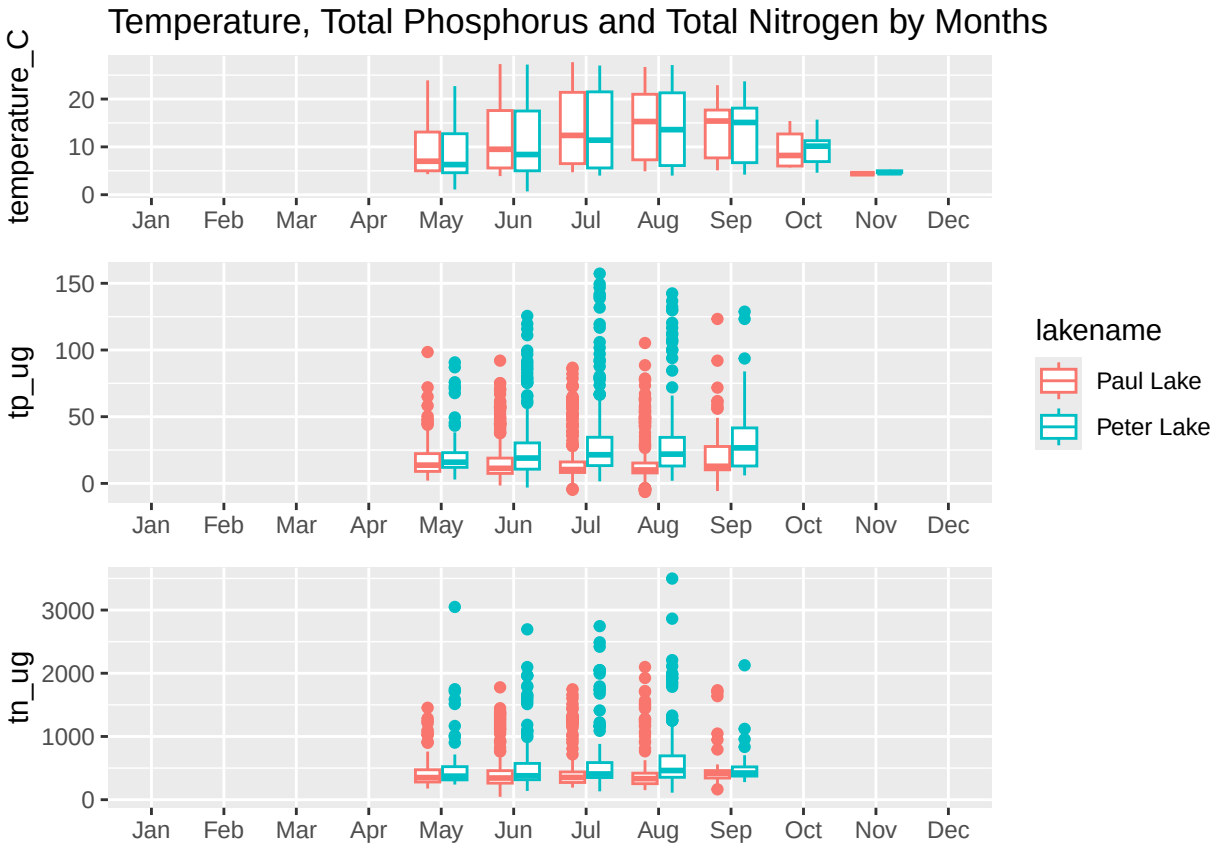



```r
#5b Boxplot of Total Phosphorus
Graph5b <- NTL_LTER %>%
  ggplot(aes(x = factor(month, levels = 1:12, labels = month.abb),
             y = tp_ug,
             color = lakename)) +
  scale_x_discrete(drop = F) +
  geom_boxplot() +
  theme(axis.title.x = element_blank())

print(Graph5b)
```




```
#5c Boxplot of Total Nitrogen
Graph5c <- NTL_LTER %>%
  ggplot(aes(x = factor(month, levels = 1:12, labels = month.abb),
             y = tn_ug,
             color = lakename)) +
  scale_x_discrete(drop = F) +
  geom_boxplot() +
  theme(axis.title.x = element_blank(), legend.position = "none")

print(Graph5c)
```


```
#5d Cowplot
Graph5d <- plot_grid(
  Graph5a, Graph5b, Graph5c,
  nrow=3, ncol = 1,
  align = "v",
  rel_heights = c(1.25,1.5,1.5)
)
print(Graph5d)
```



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

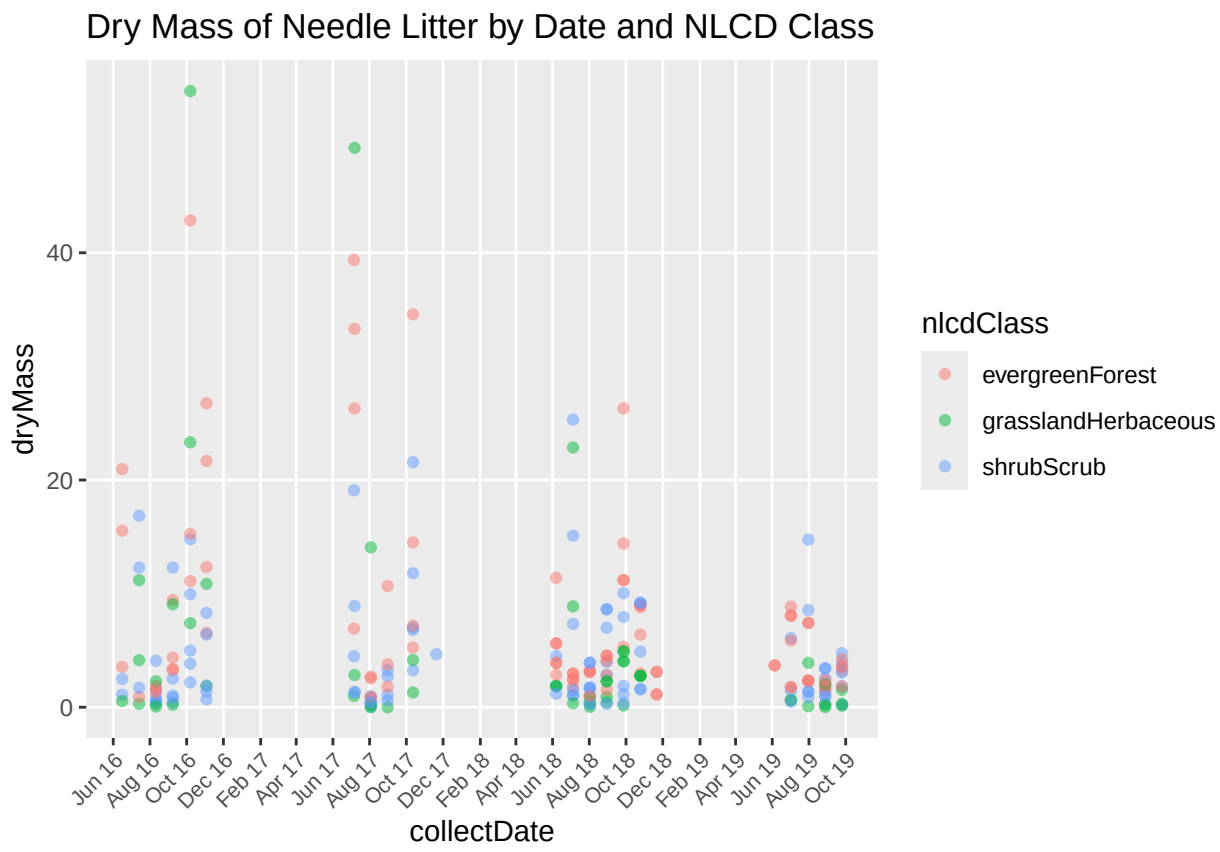
> Answer: The temperatures are usually higher during the summer and the average temperature of Paul Lake is higher Peter Lake except in October and November. Both Total Phosphorus and Total Nitrogen concentration in the lakes have large data spread and the spread is higher in Peter Lake. On the average term, Total Nitrogen contents in both lakes do not have significant differences. On the data spread term, total nitrogen contents reach their peak in August in both lakes. Compared to Total Nitrogen, the variation of average Total Phosphorus is higher.

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)

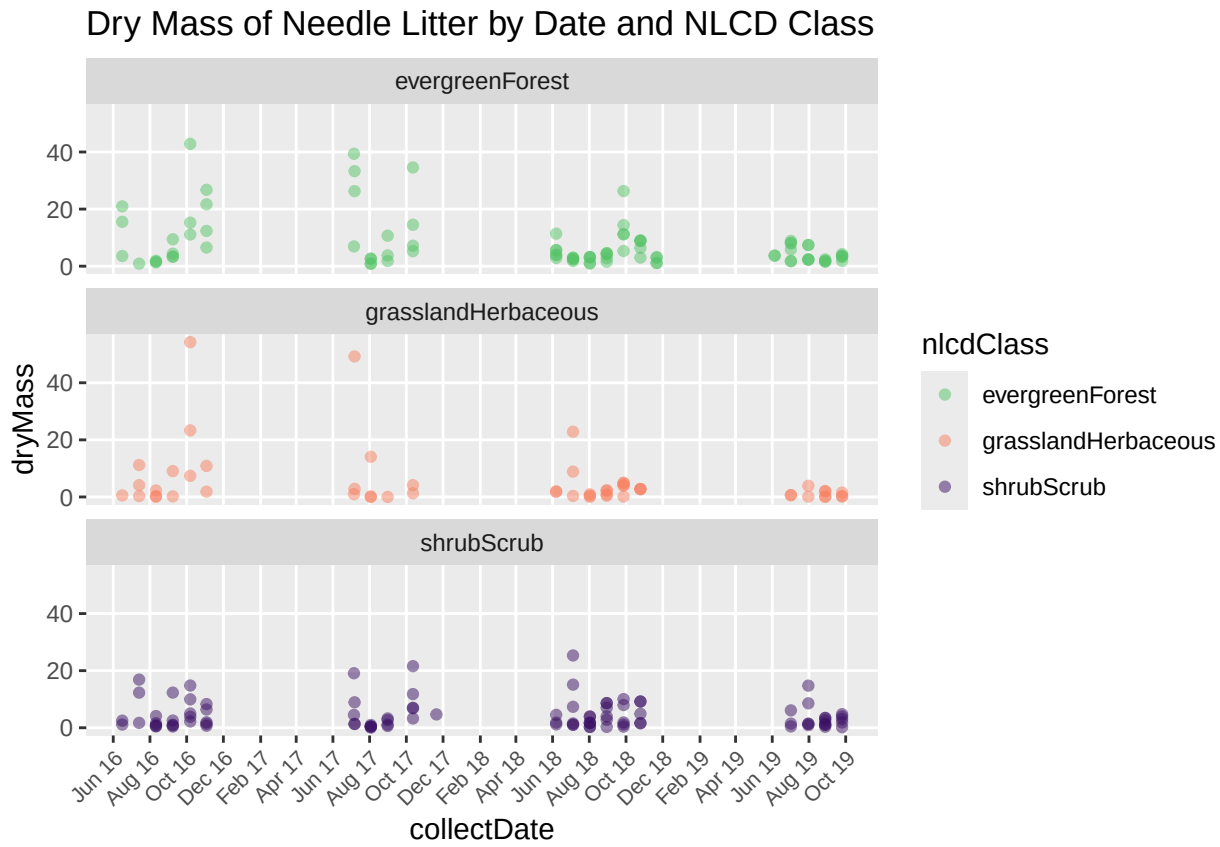
7. [Niwot Ridge] Now, plot the same plot but with NLCD classes separated into three facets rather than separated by color.

```
#6 Dry Mass of Needle Litter by Date and NLCD Class
Graph6 <- Niwot_Ridge %>%
  filter(functionalGroup == "Needles") %>%
  ggplot(aes(x = collectDate, y = dryMass, color = nlcdClass)) +
  geom_point(alpha=0.5) +
  scale_x_date(limits = as.Date(c("2016-06-16", "2019-09-25")),
               date_breaks = "2 months",
               date_labels = "%b %y") +
  ggtitle("Dry Mass of Needle Litter by Date and NLCD Class") +
  theme(axis.text.x = element_text(size=8, angle = 45,  hjust = 1)) +
```

```
  theme(panel.grid.minor = element_blank())

print(Graph6)
```




```
#7 Dry Mass of Needle Litter by Date and NLCD Class with Facets
Graph7 <- Niwot_Ridge %>%
  filter(functionalGroup == "Needles") %>%
  ggplot(aes(x = collectDate, y = dryMass, color=nlcdClass)) +
  geom_point(alpha=0.5) +
  facet_wrap(
    facets = vars(nlcdClass),
    nrow = 3, ncol = 1
  ) +
  scale_color_manual(values = c("#55c467ff","#f88061ff","#3b1165ff")) +
  scale_x_date(limits = as.Date(c("2016-06-16", "2019-09-25")),
               date_breaks = "2 months",
               date_labels = "%b %y") +
  ggtitle("Dry Mass of Needle Litter by Date and NLCD Class") +
  theme(axis.text.x = element_text(size=8, angle = 45,  hjust = 1)) +
  theme(panel.grid.minor = element_blank())

print(Graph7)
```



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

Answer: The plot 7 is more effective than the plot 6 as we can do the comparison better with plot 7. Although plot 6 different the dryMass by ncldClass colors, it is not visually clear to determine which class has the highest dryMass. In graph 7, we can easily tell which one is the highest.