Assignment 5: Data Visualization

Zhenghao Lin

Fall 2023

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
#Read libraries
library(tidyverse)
```

```
## -- 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
```

```r
library(lubridate)
library(here)
```

```
## here() starts at /Users/lzh/Desktop/EDE_Fall2023
```

```r
library(cowplot)
```

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

```r
library(ggridges)
library(viridis)
```

```
## Loading required package: viridisLite
```

```r
library(RColorBrewer)
library(colormap)
library(ggthemes)
```

```
##
## Attaching package: 'ggthemes'
##
## The following object is masked from 'package:cowplot':
##
##      theme_map
```

```r
#Read datasets
PP_chemnutr <- read.csv(
  "/Users/lzh/Desktop/EDE_Fall2023/Data/Processed/Processed_KEY/NTL-LTER_Lake_Chemistry_Nutrients_Peterl
  )
NIWO <- read.csv(
  "/Users/lzh/Desktop/EDE_Fall2023/Data/Processed/Processed_KEY/NEON_NIWO_Litter_mass_trap_Processed.csv
)
#2
#change data types of date columns into date format
PP_chemnutr$sampledate <- ymd(PP_chemnutr$sampledate)
NIWO$collectDate <- ymd(NIWO$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
#Build a them
my_theme <- theme_base() +
  theme(
    line = element_line(
      color='red',
      linewidth =0.5
    ),
    legend.background = element_rect(
      color='black',
      fill = 'grey'
    ),
    legend.title = element_text(
      color='white'
    )

  )

theme_set(my_theme)
```
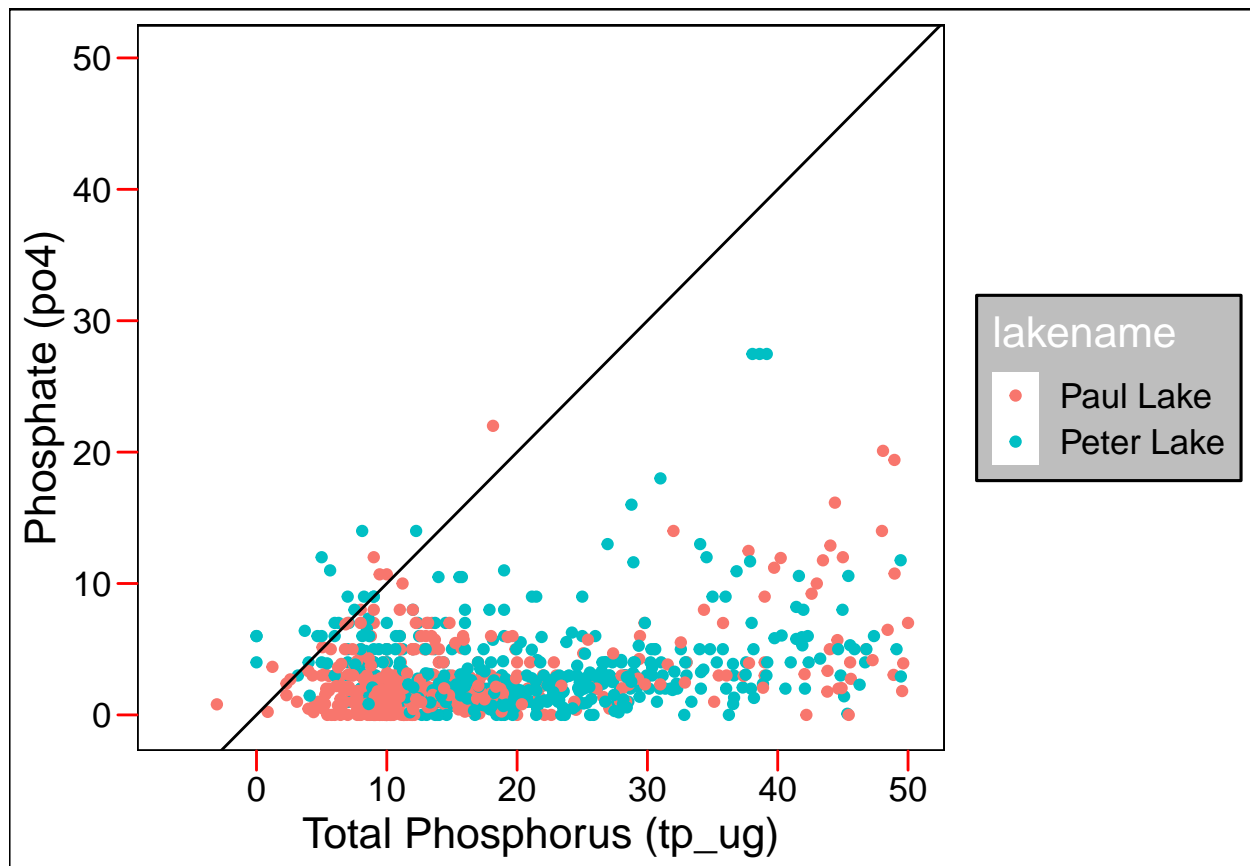
Create graphs

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 a line of best fit and color it black. Adjust your axes to hide extreme values (hint: change the limits using `xlim()` and/or `ylim()`).

```
#4

tpug_po4 <- PP_chemnutr %>%
  ggplot(aes(tp_ug, po4, color = lakename)) +
  geom_point() +
  ylim(NA,50) + xlim(NA,50) +
  geom_abline() +
  labs(x = "Total Phosphorus (tp_ug)", y = "Phosphate (po4)") +
  my_theme
tpug_po4
```

```
## Warning: Removed 22061 rows containing missing values ('geom_point()').
```

3

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.

Tip: * Recall the discussion on factors in the previous section as it may be helpful here. * R has a built-in variable called `month.abb` that returns a list of months;see https://r-lang.com/month-abb-in-r-with-example

```
#5
# Create separate boxplots for temperature, TP, and TN

boxplot_temp <- PP_chemnutr %>%
  ggplot(aes(x = month, y = temperature_C, fill = lakename)) +
  geom_boxplot() +
  xlim(6,12) +
  labs(title = "Temperature_C") +
  theme(legend.position = "none") +
  theme(plot.title = element_text(size = 10))

boxplot_tp <- PP_chemnutr %>%
  ggplot(aes(x = month, y = tp_ug, fill = lakename)) +
  geom_boxplot() +
  xlim(6,12) +
  labs(title = "Total Phosphorus (TP)") +
  theme(legend.position = "bottom") +
  theme(legend.title = element_text(size = 10))+
```

```r
  theme(plot.title = element_text(size = 10))

boxplot_tn <- PP_chemnutr %>%
  ggplot(aes(x = month, y = tn_ug, fill = lakename)) +
  geom_boxplot() +
  xlim(6,12) +
  labs(title = "Total Nitrogen (TN)") +
  theme(legend.position = "none") +
  theme(plot.title = element_text(size = 12))

# Combine the individual plots into a single cowplot
combined_plot <- plot_grid(boxplot_temp, boxplot_tp, boxplot_tn, ncol = 3, align = "v")
```

Warning: Removed 2698 rows containing missing values (`stat_boxplot()`).

Warning: Removed 3184 rows containing non-finite values (`stat_boxplot()`).

## Warning: Removed 2698 rows containing missing values (`stat_boxplot()`).

Warning: Removed 18243 rows containing non-finite values (`stat_boxplot()`).

Warning: Removed 2698 rows containing missing values (`stat_boxplot()`).

Warning: Removed 19058 rows containing non-finite values (`stat_boxplot()`).

```r
#Picture is too large to be displayed in the window, zoom in to view the full graph.
combined_plot
```
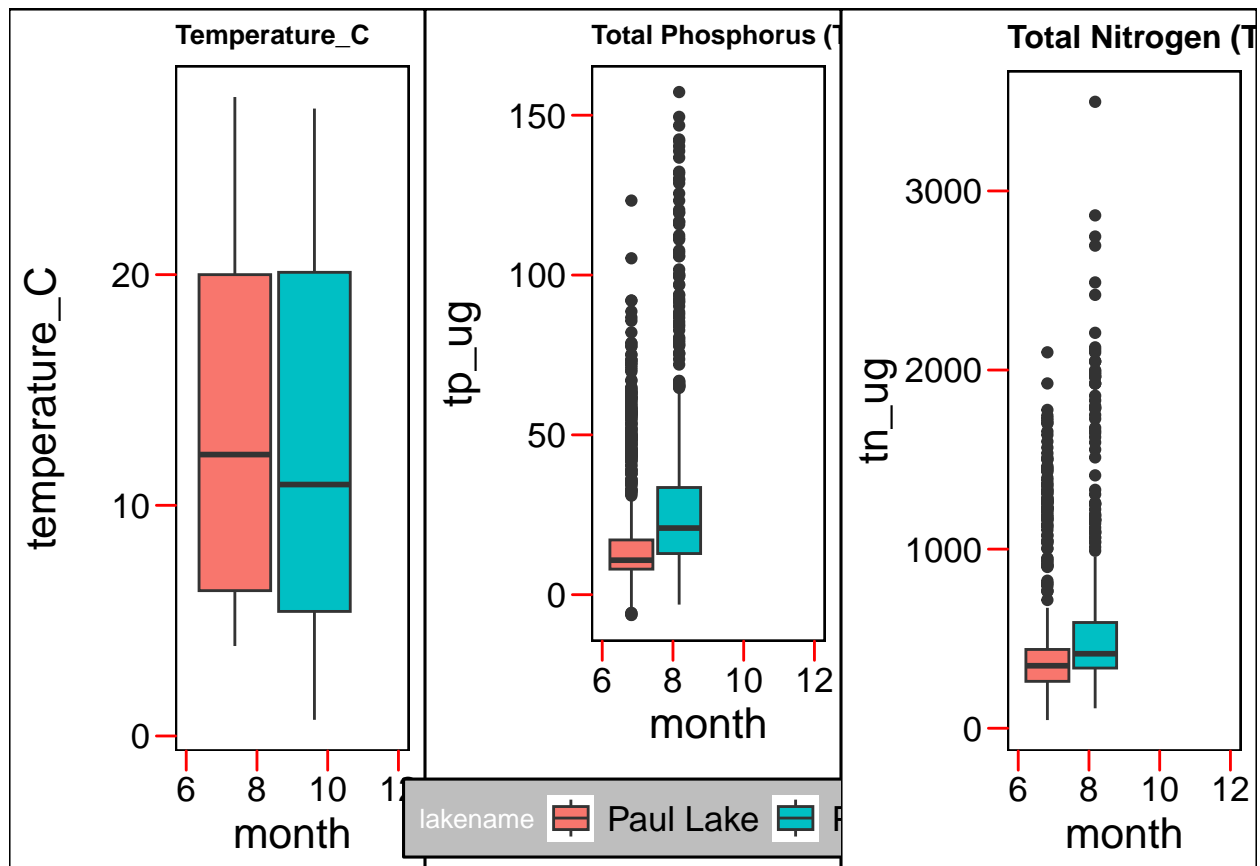
Question: What do you observe about the variables of interest over seasons and between lakes?

Answer: Season: Most observations are taken from June to September, which is during the summer. Lakes: It's noticable that Peter lake, while the temperature is similar to Paul lake, do have higher total Phosphorus and nitrogen 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

# Filter the dataset to include only the "Needles" functional group
needles_subset <- NIWO %>%
  filter(functionalGroup == "Needles")

# Plot the dry mass of needle litter by date and separate by NLCD class with a color aesthetic
Needles1 <- ggplot(needles_subset, aes(x = collectDate, y = dryMass, color = nlcdClass)) +
  geom_point() +
  labs(title = "Dry Mass of Needle Litter by Date",
       x = "Date",
       y = "Dry Mass") +
  scale_color_discrete(name = "NLCD Class") +
```

```
  theme(legend.position = "right")
```
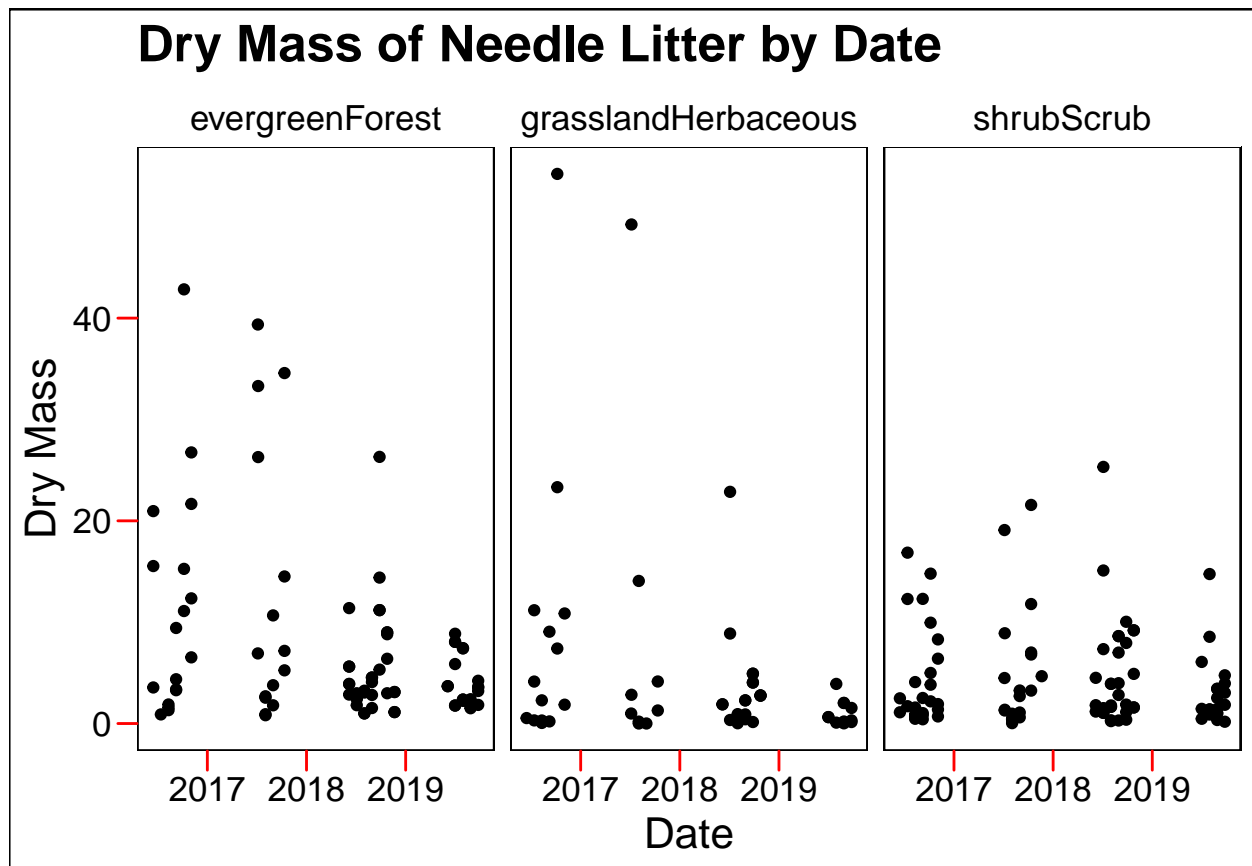
Needles1
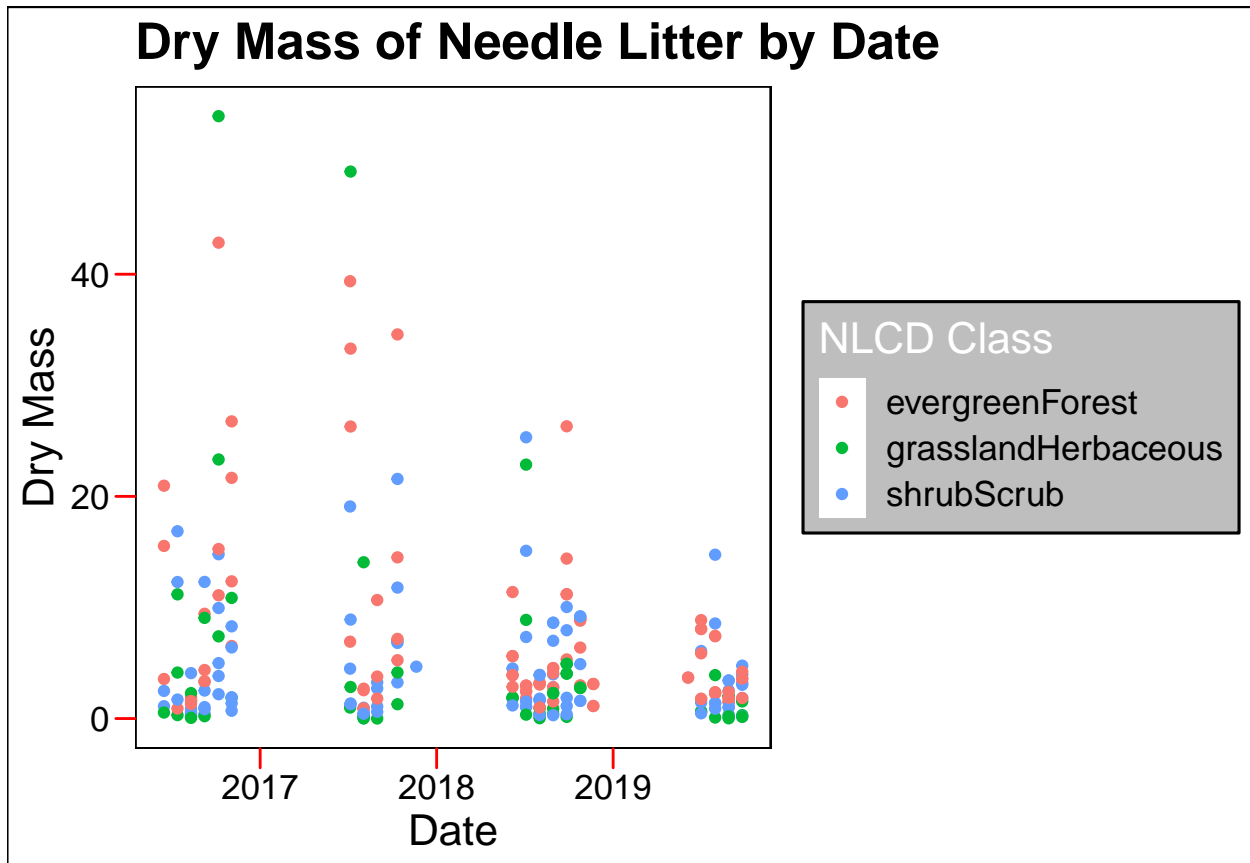

```
#7

# Plot the same data with NLCD classes separated into three facets
Needles2 <- ggplot(needles_subset, aes(x = collectDate, y = dryMass)) +
  geom_point() +
  facet_wrap(~ nlcdClass, nrow = 1) +  # Separate by NLCD class in facets
  labs(title = "Dry Mass of Needle Litter by Date",
       x = "Date",
       y = "Dry Mass")

# Display the plot
Needles2
```
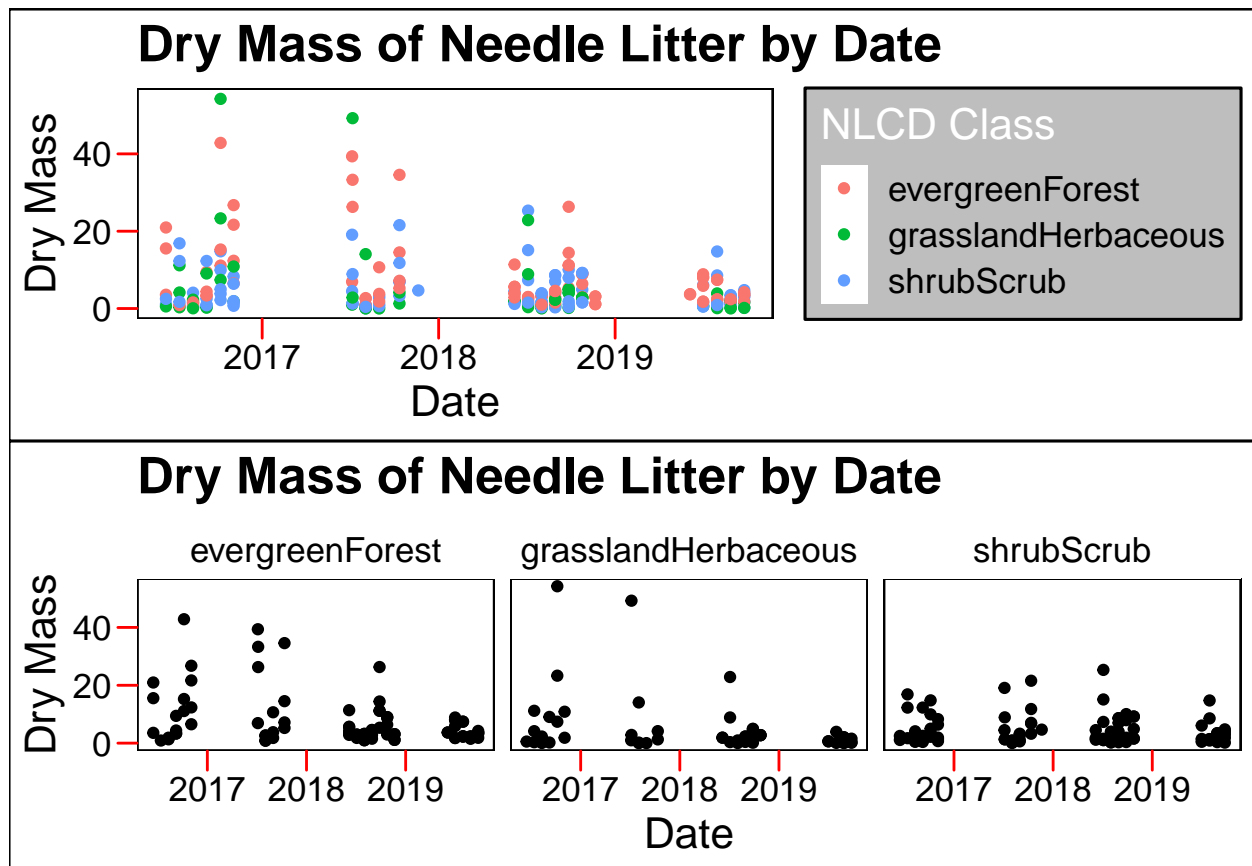
Dry Mass of Needle Litter by Date

```
combined_Needles <- plot_grid(Needles1, Needles2,  nrow = 2, align = "h")
```

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

```
combined_Needles
```

Dry Mass of Needle Litter by Date

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

Answer: I think that with this amount of dataset, if we want to compare the dry mass of different NLCD class at different year, the facet graph should be more effective than plots with color mapping. From the facet, not only can I see the difference among each NLCD class of the dry mass of Needles at each year, but I can also compare the difference at each year of every class.