Assignment 5: Data Visualization

Isaac Benaka

OVERVIEW

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

Directions

1. Rename this file `<FirstLast>_A02_CodingBasics.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 to **answer the questions** in this assignment document.
5. When you have completed the assignment, **Knit** the text and code into a single PDF file.

The completed exercise is due on Friday, Oct 14th @ 5:00pm.

Set up your session

1. Set up your session. Verify your working directory and load the tidyverse, lubridate, & cowplot packages. Upload the NTL-LTER processed data files for nutrients and chemistry/physics for Peter and Paul Lakes (use the tidy [NTL-LTER_Lake_Chemistry_Nutrients_PeterPa version) and the processed data file for the Niwot Ridge litter dataset (use the [NEON_NIWO_Litter_mass_trap_Processe version).

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

```
# 1
getwd()
```

```
## [1] "/Users/isaacbenaka/Desktop/Fall 2022/872 - Data Analytics/EDA-Fall2022"
```

```
library(tidyverse)
```

```
## -- Attaching packages --------------------------------------- tidyverse 1.3.2 --
## v ggplot2 3.3.6      v purrr   0.3.4
## v tibble  3.1.8      v dplyr   1.0.10
## v tidyr   1.2.1      v stringr 1.4.1
## v readr   2.1.2      v forcats 0.5.2
## -- Conflicts ------------------------------------------ tidyverse_conflicts() --
## x dplyr::filter() masks stats::filter()
## x dplyr::lag()    masks stats::lag()
```

```
library(lubridate)
```

```
##
## Attaching package: 'lubridate'
##
## The following objects are masked from 'package:base':
##
##     date, intersect, setdiff, union
```

```
library(cowplot)
```

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

```
PeterPaul.chem.nutrients <- read.csv("./Data/Processed/NTL-LTER_Lake_Chemistry_Nutrients_PeterPaul_Proce
    stringsAsFactors = T)
Litter <- read.csv("./Data/Processed/NEON_NIWO_Litter_mass_trap_Processed.csv")
# 2
class(Litter$collectDate)
```

```
## [1] "character"
```

```
Litter$collectDate <- as.Date(Litter$collectDate,
    format = "%Y-%m-%d")
class(PeterPaul.chem.nutrients$sampledate)
```

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

```
PeterPaul.chem.nutrients$sampledate <- as.Date(PeterPaul.chem.nutrients$sampledate,
    format = "%Y-%m-%d")
class(PeterPaul.chem.nutrients$month)
```

```
## [1] "integer"
```

```
PeterPaul.chem.nutrients$month <- factor(PeterPaul.chem.nutrients$month,
    levels = c(1:12))
```

Define your theme

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

```
# 3
iketheme <- theme_bw(base_size = 8) + theme(axis.text = element_text(color = "black"),
    legend.position = "top")
theme_set(iketheme)
```

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
UGvsPO4plot <- ggplot(PeterPaul.chem.nutrients,
    (aes(x = tp_ug, y = po4, color = lakename))) +
    ylim(0, 50) + geom_point(alpha = 0.5) +
    geom_smooth(method = "lm", se = FALSE,
        color = "black") + xlab("Total phosphorus") +
    ylab("Phosphate")
print(UGvsPO4plot)
```

```
## '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).

## Warning: Removed 2 rows containing missing values (geom_smooth).
```


3
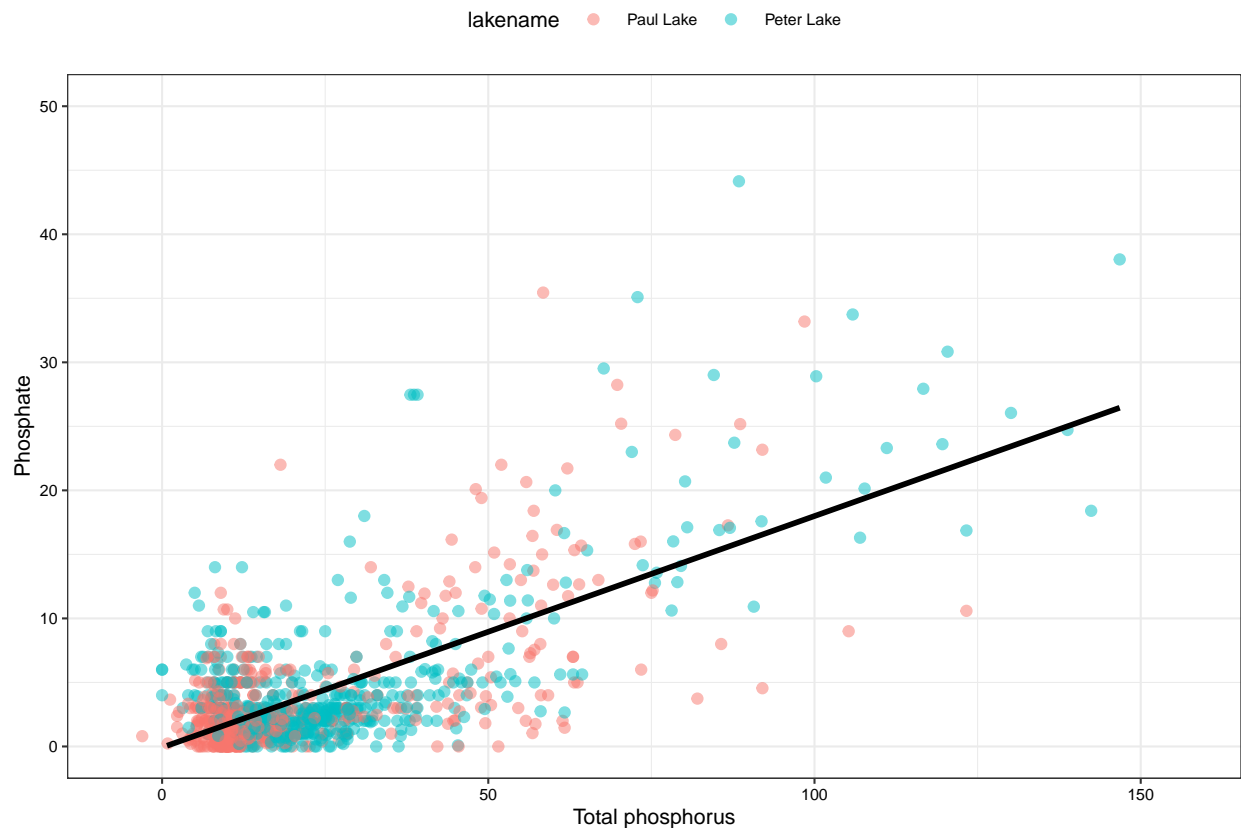
5. [NTL-LTER] Make three separate boxplots of (a) temperature, (b) TP, and

(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: R has a build in variable called `month.abb` that returns a list of months; see https://r-lang.com/month-abb-in-r-with-example
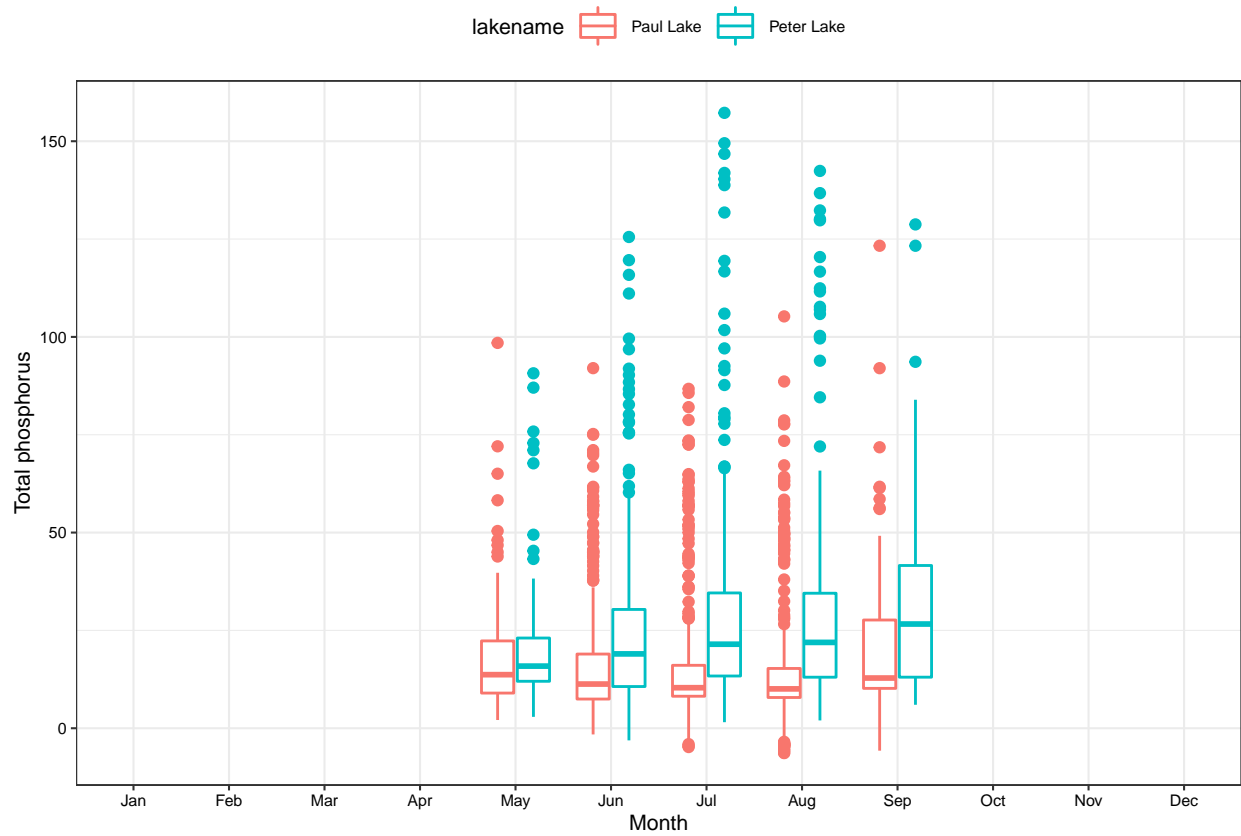
```
# 5
Tplot <- ggplot(PeterPaul.chem.nutrients,
    aes(x = month, y = temperature_C)) +
    geom_boxplot(aes(color = lakename)) +
    ylab("Temperature (C)") + xlab("Month") +
    scale_x_discrete(labels = month.abb,
        drop = FALSE)
print(Tplot)
```

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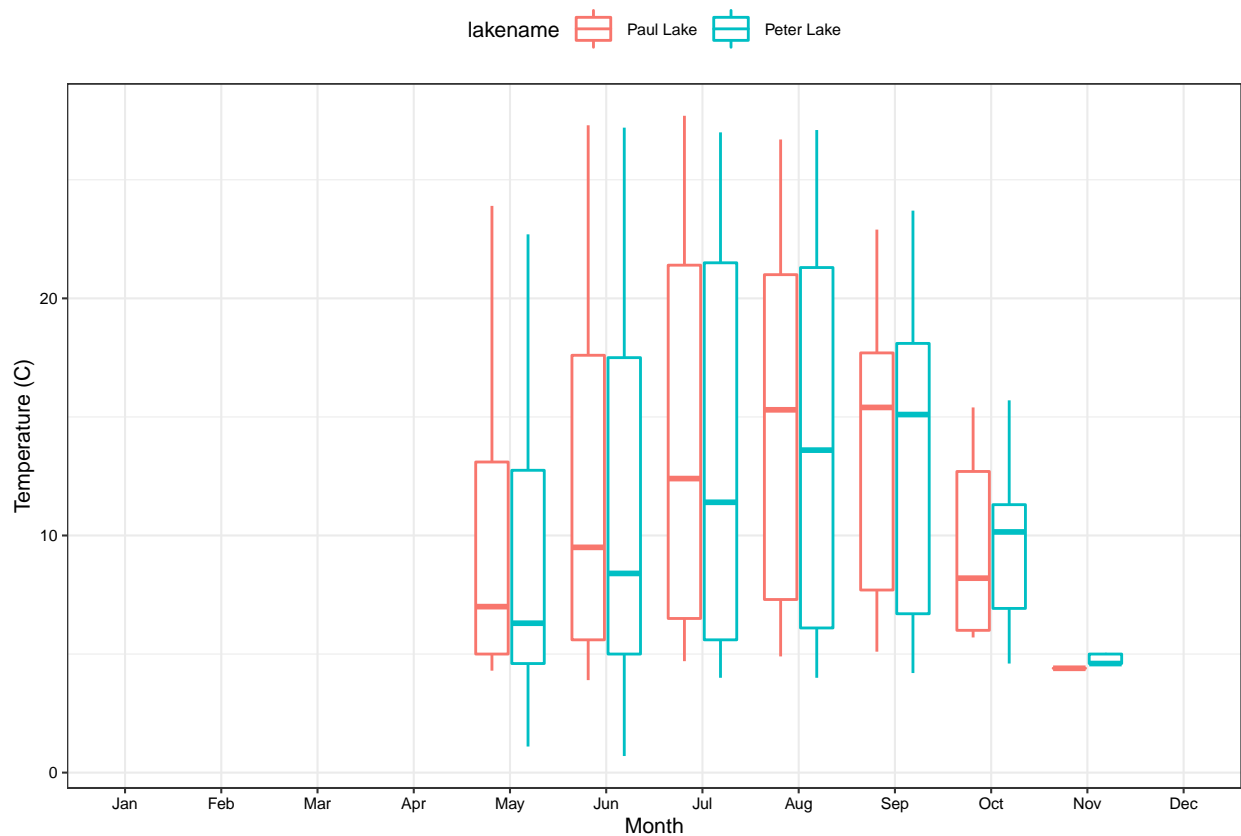


```
TPplot <- ggplot(PeterPaul.chem.nutrients,
    (aes(x = month, y = tp_ug))) + geom_boxplot(aes(color = lakename)) +
    ylab("Total phosphorus") + xlab("Month") +
    scale_x_discrete(labels = month.abb,
        drop = FALSE)
print(TPplot)
```

```
## Warning: Removed 20729 rows containing non-finite values (stat_boxplot).
```



```
TNplot <- ggplot(PeterPaul.chem.nutrients,
    (aes(x = month, y = tn_ug))) + geom_boxplot(aes(color = lakename)) +
    ylab("Phosphate") + xlab("Month") + scale_x_discrete(labels = month.abb,
    drop = FALSE)
print(TNplot)
```

```
## Warning: Removed 21583 rows containing non-finite values (stat_boxplot).
```
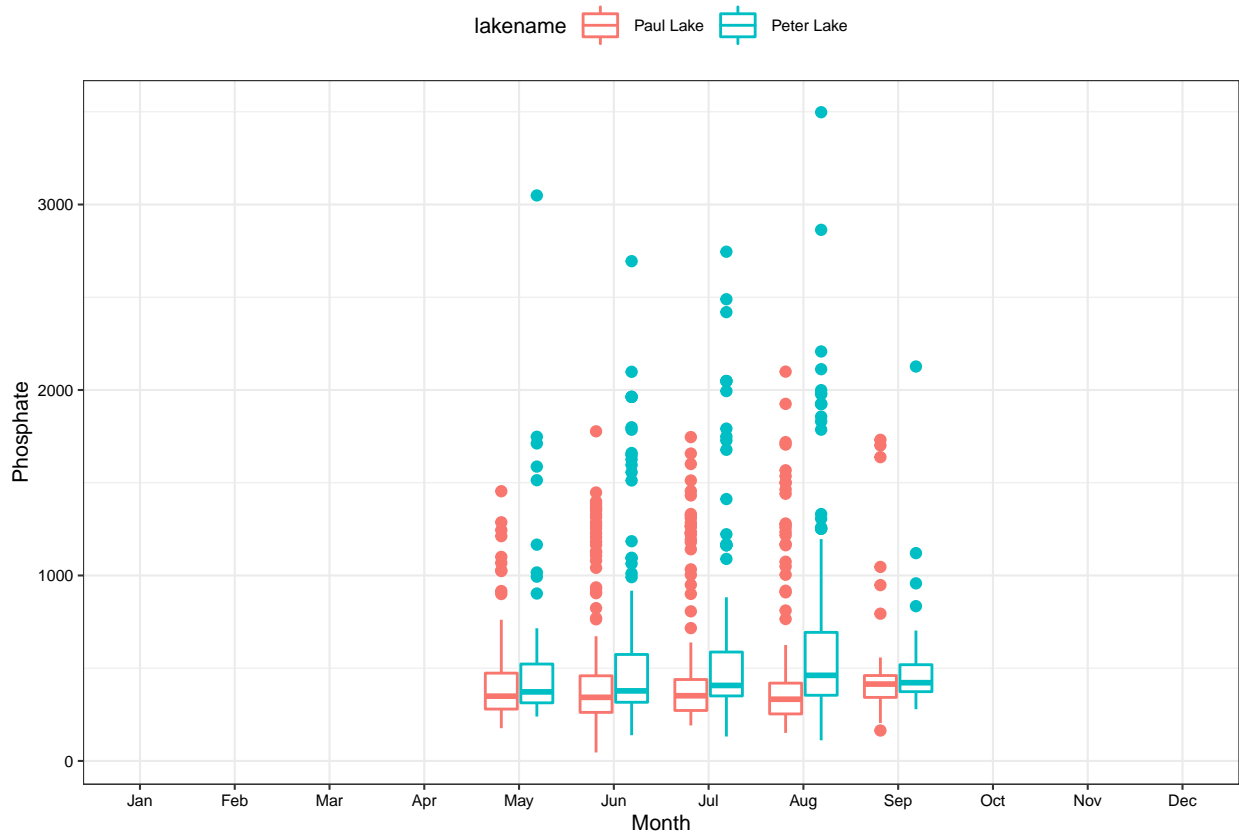
```r
combinedplot <- plot_grid(Tplot + theme(legend.position = "none"),
    TPplot + theme(legend.position = "none"),
    TNplot + theme(legend.position = "bottom"),
    rel_heights = c(2, 2, 4), ncol = 1, align = "v",
    axis = "b")
```

```
## 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).
```

```r
print(combinedplot)
```
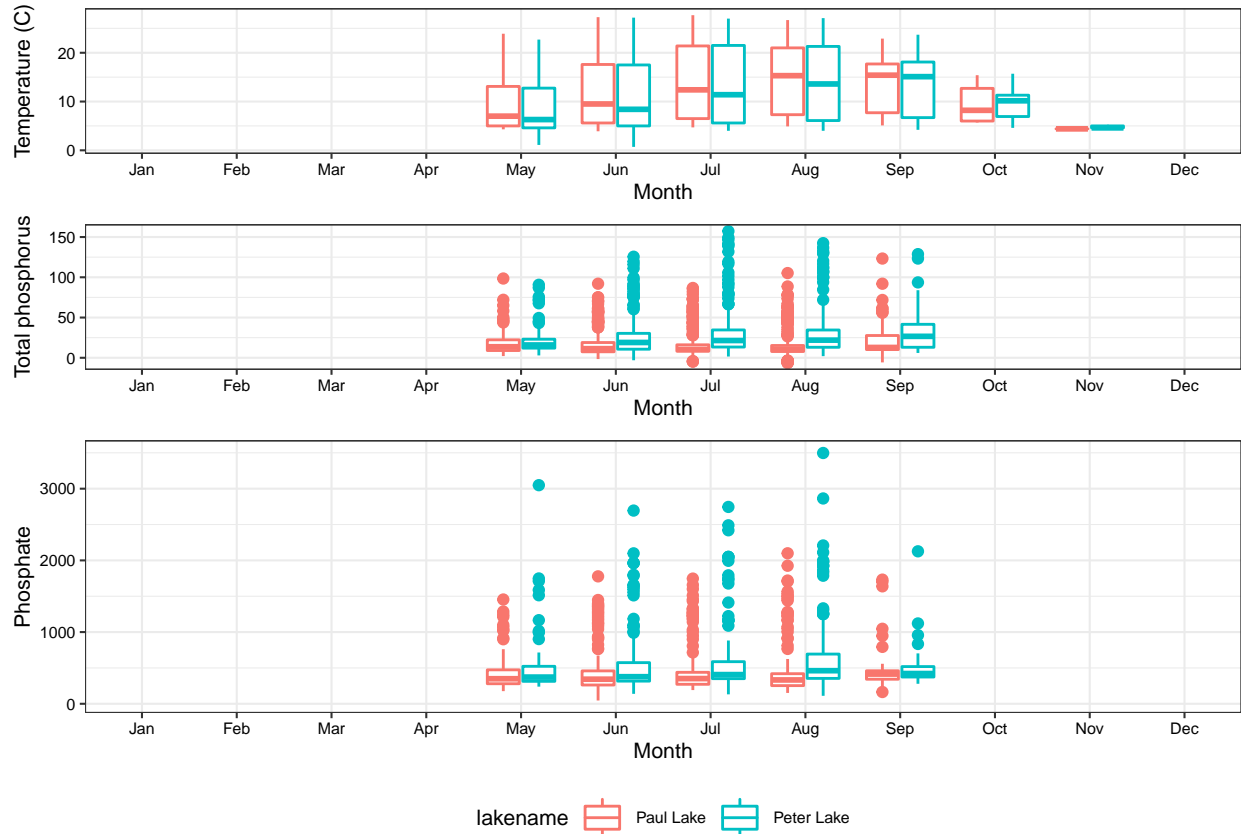
Question: What do you observe about the variables of interest over seasons and between lakes?

Answer: Temperature for both lakes rises during the summer months and begins to decrease going into the autumnal months. Overall, phosphorus measurements were fairly consistent for both lakes over time, but did rise in September. Peter Lake has a higher phosphorus concentration compared to Paul Lake. Similarly, nitrogen levels were also consistent over time, and Peter Lake had slightly higher nitrogen, but the means were fairly close. Peter Lake had more outlying measurements for both nitrogen and phosphorus compared to 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
Litter.needles <- Litter %>% #Filter out all litter groups except for "Needles"
  filter(Litter$functionalGroup == "Needles")

DryMassplot <- ggplot(Litter.needles)+
  geom_point(aes(x = collectDate, y = dryMass, color=nlcdClass))+
  ylab("Needle dry mass (g)")+
  xlab("Date of collection")+
  scale_x_date(limits = as.Date(c("2016-01-01", "2019-12-31")),
    date_breaks = "3 months", date_labels = "%b %Y")
print(DryMassplot)
```
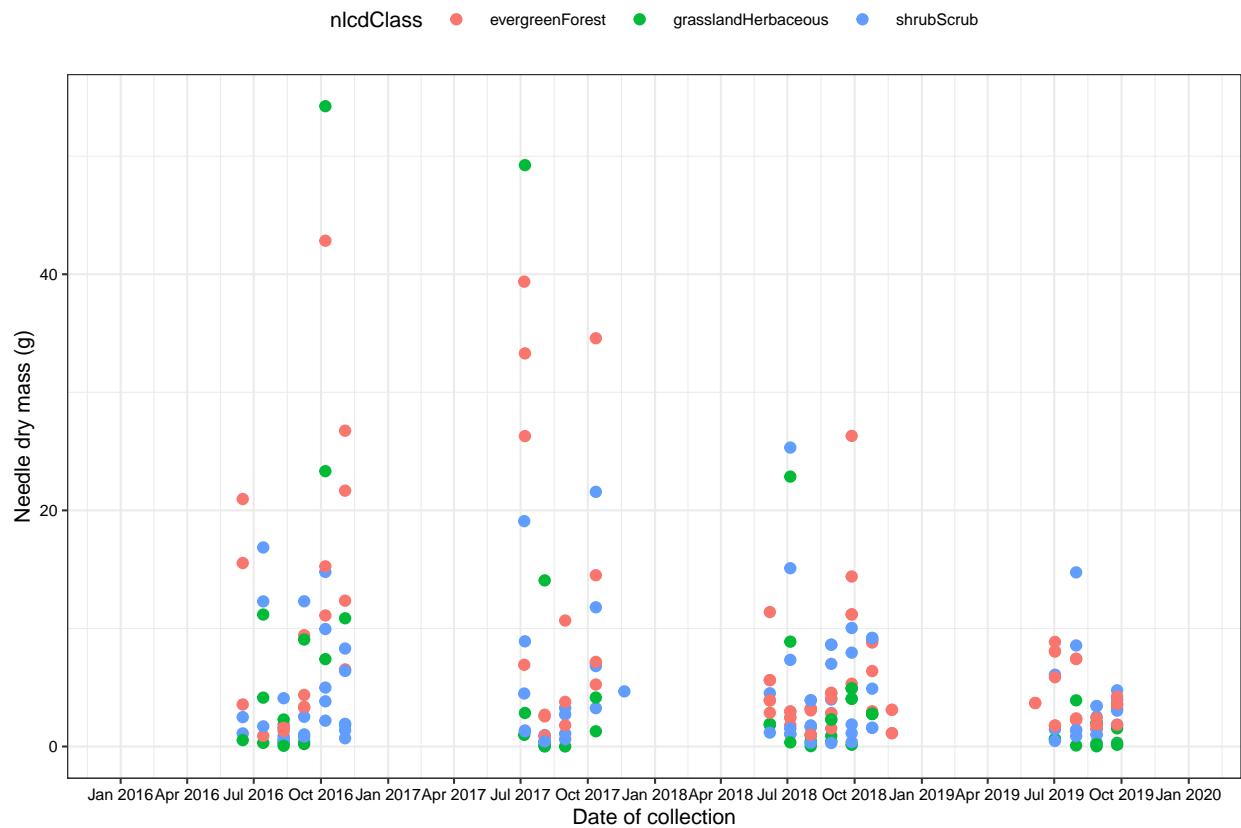
```
#7
DryMassfacet <- ggplot(Litter.needles)+
  geom_point(aes(x = collectDate, y = dryMass))+
  facet_wrap(vars(nlcdClass), nrow = 3)+
  ylab("Needle dry mass (g)")+
  xlab("Date of collection")+
  scale_x_date(limits = as.Date(c("2016-01-01", "2019-12-31")),
    date_breaks = "3 months", date_labels = "%b %Y")
print(DryMassfacet)
```
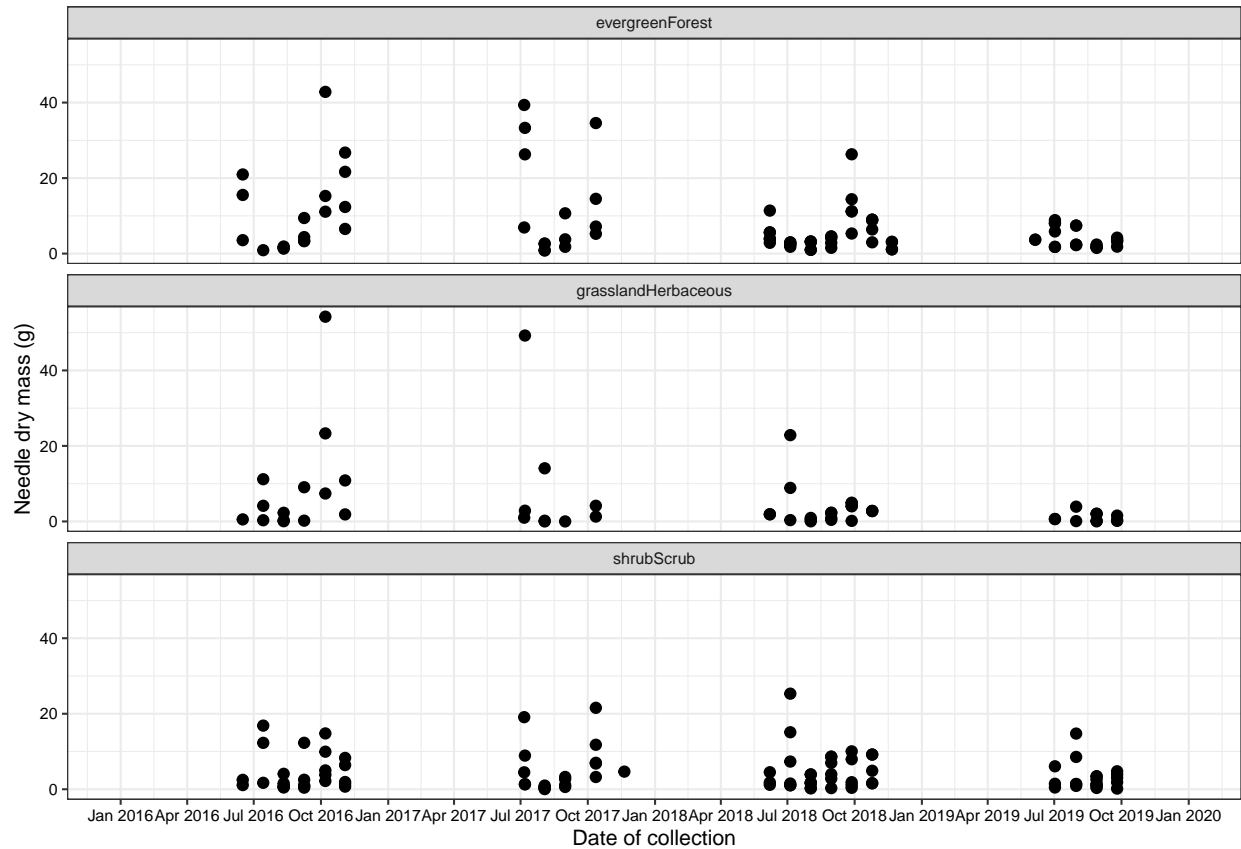
Question: Which of these plots (6 vs. 7) do you think is more effective, and why?

Answer: The plot that is separated into facets is a more effective presentation of the data because it is easier to discern trends of the separate litter groups. When overlayed on eachother with different colors separating them, the litter groups are harder to identify from one another because the functional group dots can overlap. With the facet wrapped plots, it is also easier to see the overall trend of which NLCD class has more dry mass.