Assignment 5: Data Visualization

Nadia Barbo

Spring 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. 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_PeterPaul_Processed.csv` version) and the processed data file for the Niwot Ridge litter dataset (use the `NEON_NIWO_Litter_mass_trap_Processed.csv` version).

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

```r
#1
library(tidyverse); library(lubridate); library(here); library(cowplot) #loading
→   necessary packages
getwd() #checking my working directory
```

```
## [1] "C:/Users/nadia/Documents/Duke_/EDA-Spring2023"
```

```r
Lake_Nutrients <-
→   read.csv("./Data/Processed/NTL-LTER_Lake_Chemistry_Nutrients_PeterPaul_Processed.csv",
→   stringsAsFactors = TRUE) #loading in NTL-LTER processed data files for nutrients and
→   chemistry/physics
Niwot_Litter <- read.csv("./Data/Processed/NEON_NIWO_Litter_mass_trap_Processed.csv",
→   stringsAsFactors = TRUE) #loading in processed data file for the Niwot Ridge litter
→   dataset
```

```
#2
class(Lake_Nutrients$sampledate) #checking format of NTL-LTER dataset date column (it's
↪  factor)
```

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

```
class(Niwot_Litter$collectDate) #checking format of Niwot Ridge litter dataset date
↪  column (it's factor)
```

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

```
Lake_Nutrients$sampledate <- ymd(Lake_Nutrients$sampledate) #Changing date column in
↪  NTL-LTER dataset (which were originally factors) to be date objects
Niwot_Litter$collectDate <- ymd(Niwot_Litter$collectDate) #Changing date column in Niwot
↪  Ridge litter dataset (which were originally factors) to be date objects

class(Lake_Nutrients$sampledate) #checking format of NTL-LTER dataset date column (it's
↪  date)
```

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

```
class(Niwot_Litter$collectDate) #checking format of Niwot Ridge litter dataset date
↪  column (it's date)
```

```
## [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

```
#3
nadia_theme <- theme_classic() +
  theme(
    axis.title = element_text(color = "hotpink", #making axis titles hot pink
                              family = "mono",#making axis titles 'mono' font
                              face = 'bold', #making axis titles bold
                              size = 14), #making axis titles larger (size 14)
    plot.title = element_text(color = "navy", #making plot title navy
                              family = "mono", #making plot title 'mono' font
                              hjust = 0.5, #making plot title centered
                              face = 'bold'), #making plot title bold
```

```
     panel.background = element_rect(fill = 'lightgrey'), #making panel background light
     ↪  gray
     axis.line.x.bottom = element_line(size = 0), #making bottom axis line thinner
     axis.line.y.left = element_line(size = 0), #making left axis line thinner
     axis.ticks = element_line(size = 1), #making axis tick marks size 1

     panel.grid.major = element_line(linetype = 2, #adding grid lines and making them
     ↪  dashed
                                     linewidth = 1), #making grid lines size 1

     legend.background = element_rect(
       fill = 'lightgrey'),#making the legend background light grey
     legend.text = element_text(family = 'mono', #making the legend text 'mono' font
                                size = 12), #making the legend text size 12
     legend.position = 'bottom' #putting the legend on the bottom of the plot
     )

theme_set(nadia_theme) #setting my theme as the default theme
```
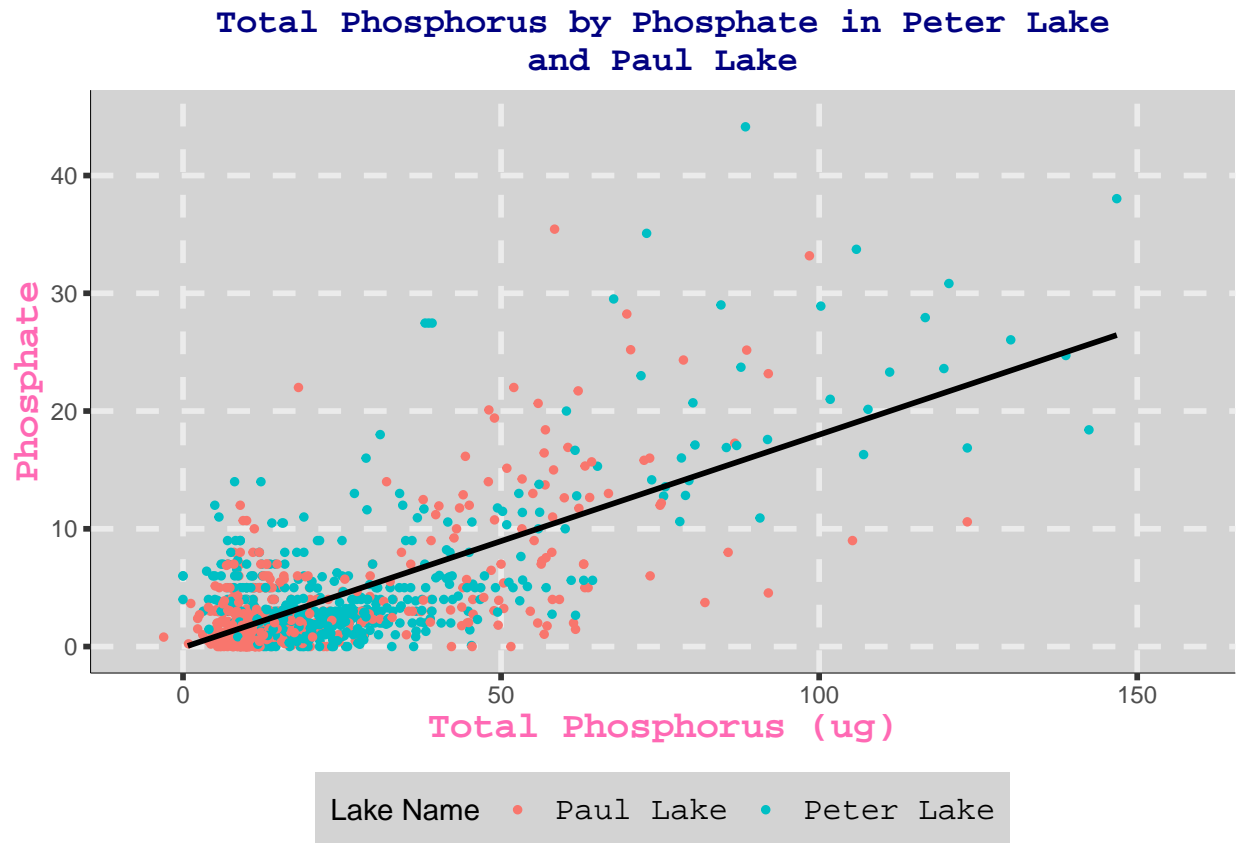
Create graphs

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
TPxPO4_plot <- ggplot(Lake_Nutrients, aes(x = tp_ug, y = po4, color = lakename)) +
↪  #assigning my x and y axes and color
  geom_point(size = 1) + #making point size 1
  geom_smooth(
    method = 'lm', #adding a best fit line
    se = FALSE, #no standard error
    color = 'black') + #making best fit line black
  ylim(0,45) + #creating a y axis limit to exclude extreme points
  labs(
    x = 'Total Phosphorus (ug)', #adding x axis label
    y = 'Phosphate', #adding y axis label
    title = 'Total Phosphorus by Phosphate in Peter Lake\nand Paul Lake', #adding title
    color = 'Lake Name') #renaming legend title
TPxPO4_plot #showing plot
```

Total Phosphorus by Phosphate in Peter Lake and Paul Lake

Phosphate (y-axis), Total Phosphorus (ug) (x-axis)

Lake Name • Paul Lake • Peter Lake

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
Temp_plot <- Lake_Nutrients %>% #creating plot variable
  ggplot(
    aes(x = factor(month, levels = 1:12, labels = month.abb), #assigning x axis (make
    ↪  factor with month abbreviations)
             y = temperature_C, #assigning y axis (Temperature)
             color = lakename)) + #color by lakename
  geom_boxplot() + #make boxplots
  scale_x_discrete( #discrete x axis variables
    name = "Month",
    drop = FALSE) +
  labs(x = 'Month', #x axis label
       y = 'Temperature (C)', #y axis label
       title = 'Temperature per Month in Peter Lake and Paul Lake') + #title
  theme(legend.position="none") #remove legend
Temp_plot #show temperature x month boxplot
```

Temperature per Month in Peter Lake and Paul Lake
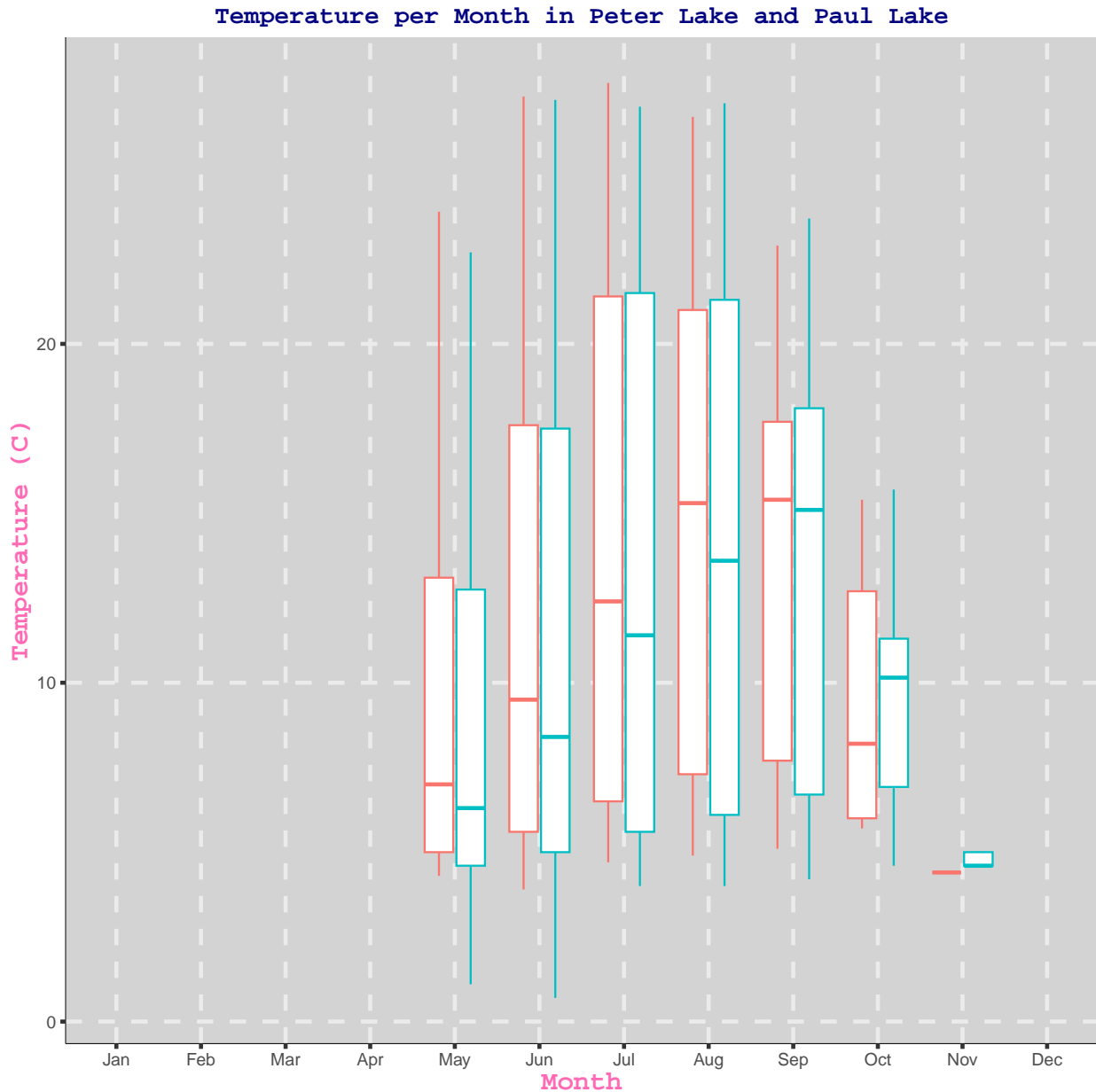

```
TP_plot <- Lake_Nutrients %>% #creating plot variable
  ggplot(
    aes(x = factor(month, levels = 1:12, labels = month.abb), #assigning x axis (make
    ↪  factor with month abbreviations)
        y = tp_ug, #assigning y axis (TP)
        color = lakename)) + #color by lakename
  geom_boxplot() + #make boxplots
  scale_x_discrete( #discrete x axis variables
    name = "Month",
    drop = FALSE) +
  labs(x = 'Month', #x axis label
       y = 'Total Phosphorus (ug)', #y axis label
       title = 'Total Phosphorus per Month in Peter Lake and Paul Lake') + #title
```

```
    theme(legend.position="none") #remove legend
TP_plot #show tp x month boxplot
```

Total Phosphorus per Month in Peter Lake and Paul Lake
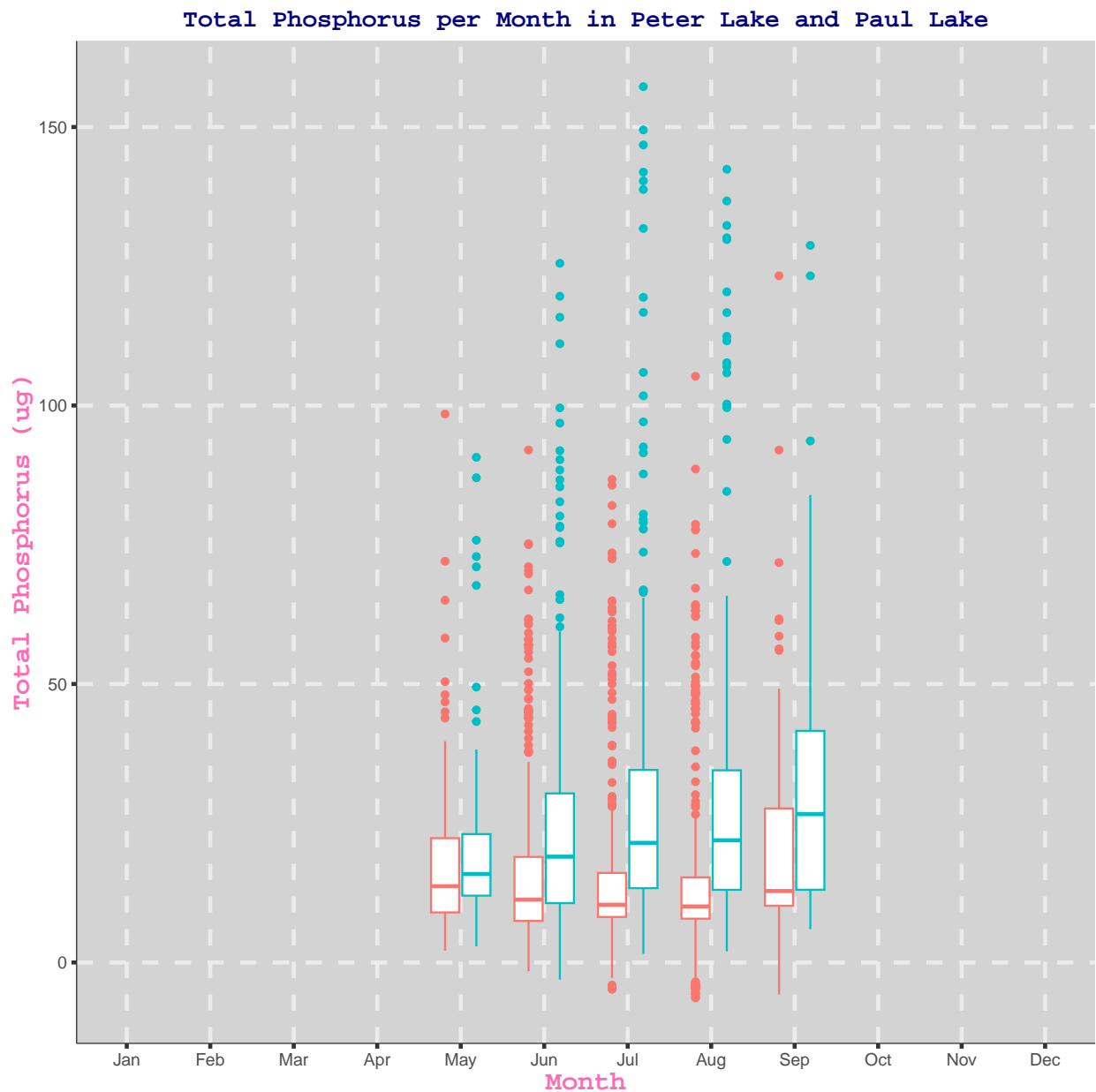


```
TN_plot <- Lake_Nutrients %>% #creating plot variable
  ggplot(
    aes(x = factor(month, levels = 1:12, labels = month.abb), #assigning x axis (make
    ↪   factor with month abbreviations)
        y = tn_ug, #assigning y axis (TN)
        color = lakename)) + #color by lakename
  geom_boxplot() + #make boxplots
  scale_x_discrete( #discrete x axis variables
    name = "Month",
    drop = FALSE) +
```

```
  labs(x = 'Month', #x axis label
       y = 'Total Nitrogen (ug)', #y axis label
       title = 'Total Nitrogen per Month in Peter Lake and Paul Lake') + #title
  theme(legend.position="none") #remove legend
TN_plot #show tn x month boxplot
```


Total Nitrogen per Month in Peter Lake and Paul Lake
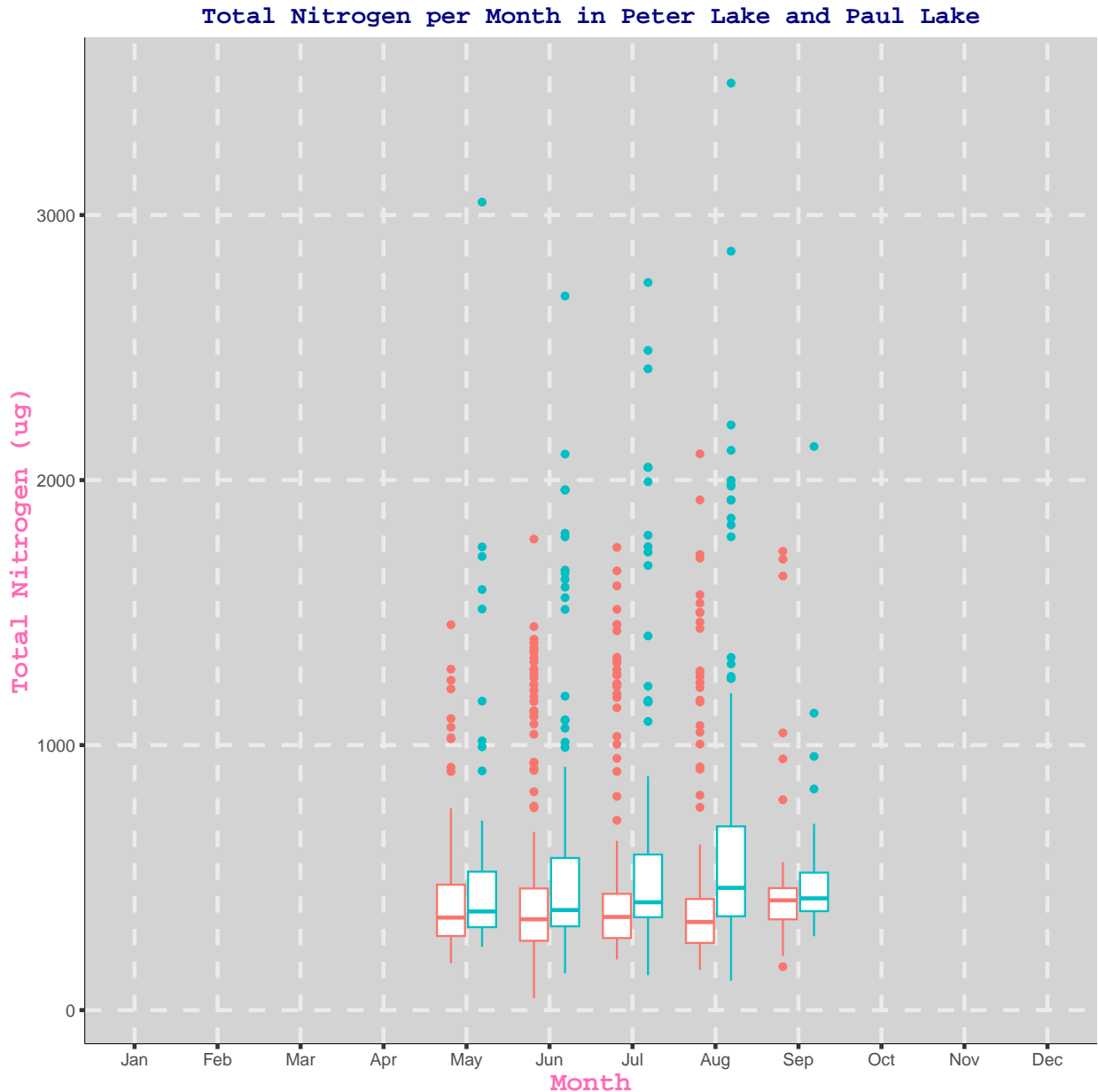
```
legend <- get_legend(TN_plot + labs(color='Lake Name') + theme(legend.position = 'right',
↪    legend.box.margin = margin(0,0,0,12))) #extracting legend
plotgrid <- plot_grid(Temp_plot, TP_plot, TN_plot, nrow = 3, align = 'v', rel_widths =
↪    c(6, .4)) #creating a cowplot that combines the three graphs
plotgrid_l <- plot_grid(plotgrid, legend, rel_widths = c(3, 0.75)) #adding legend to
↪    cowplot
plotgrid_l #show cowplot
```

Temperature per Month in Peter Lake and Paul Lake

Total Phosphorus per Month in Peter Lake and Paul Lake

Total Nitrogen per Month in Peter Lake and Paul Lake
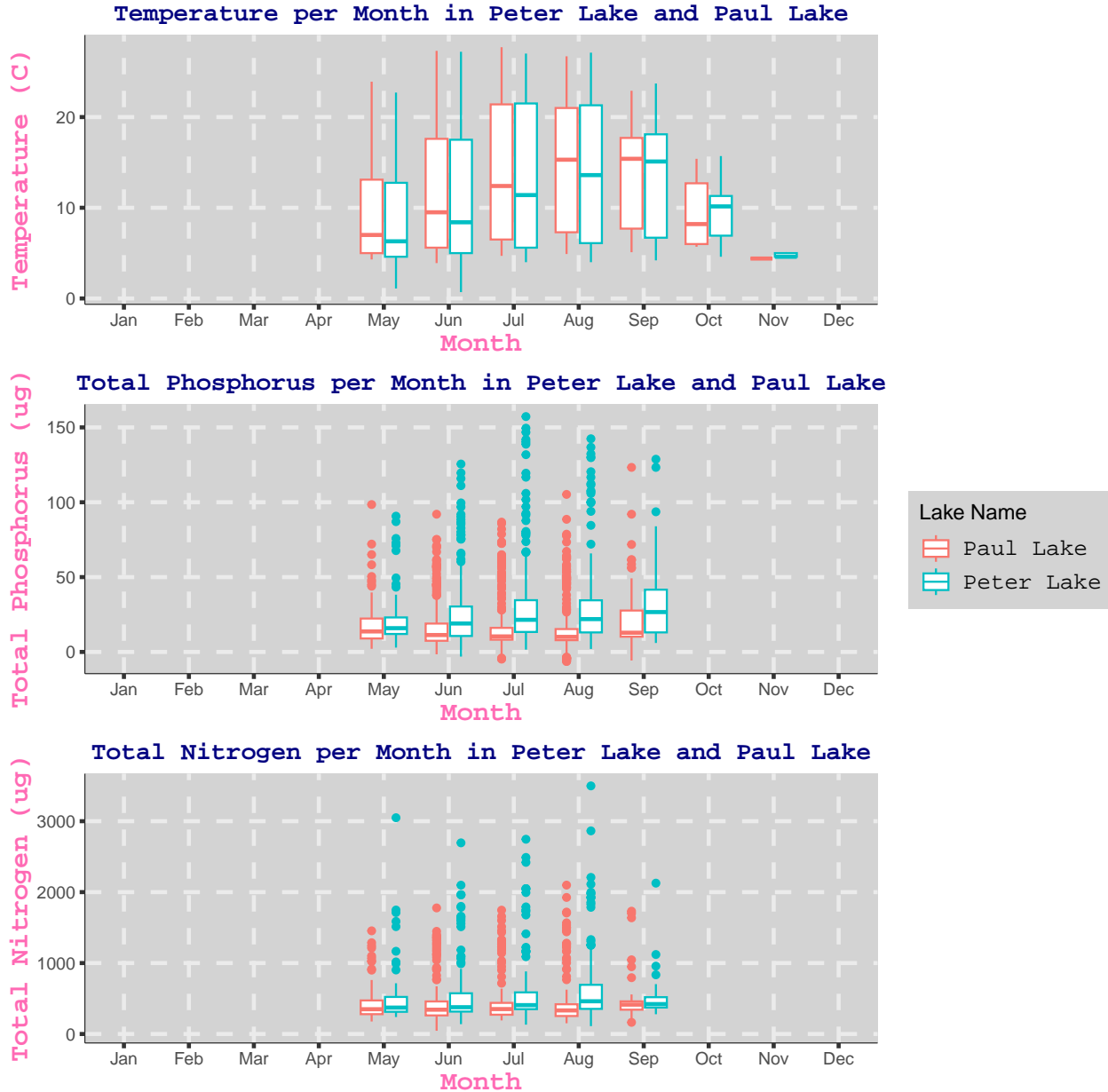
Question: What do you observe about the variables of interest over seasons and between lakes?

Answer: The lake temperatures increase into the summer months and then decrease in Oct and Nov. The temperatures seem similar between both lakes. Total phosphorus and total nitrogen appear to be similar within themselves throughout all seasons, but there may be in increase over the year in total phosphorus in Peter Lake. Peter Lake appears to have higher total phosphorus and total nitrogen levels, however we don't know if that is statistically significant.

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
NW_needles <- #assigning data to plot
  ggplot(
    subset(Niwot_Litter, functionalGroup == "Needles"), #subsetting the data we want
    ↪ (just needles in functiongroup variable)
    aes(x = factor(collectDate), #assigning x axis (as factor to see all dates)
        y = dryMass, #assigning y axis
        color=nlcdClass)) + #seperating data by NLCD class
  geom_boxplot() + #make boxplot
  labs(x = 'Collection Date', #adding x axis label
       y = 'Dry Mass', #adding y axis label
       title = 'Dry Mass of Needles at Every Collection Date', #adding title
       color = 'NLCD Class') + #changing legend name
  theme(axis.text.x = element_text(angle = 90, vjust = 0.5, hjust=1)) #making x axis tick
  ↪ labels vertical
NW_needles #showing figure
```


Dry Mass of Needles at Every Collection Date
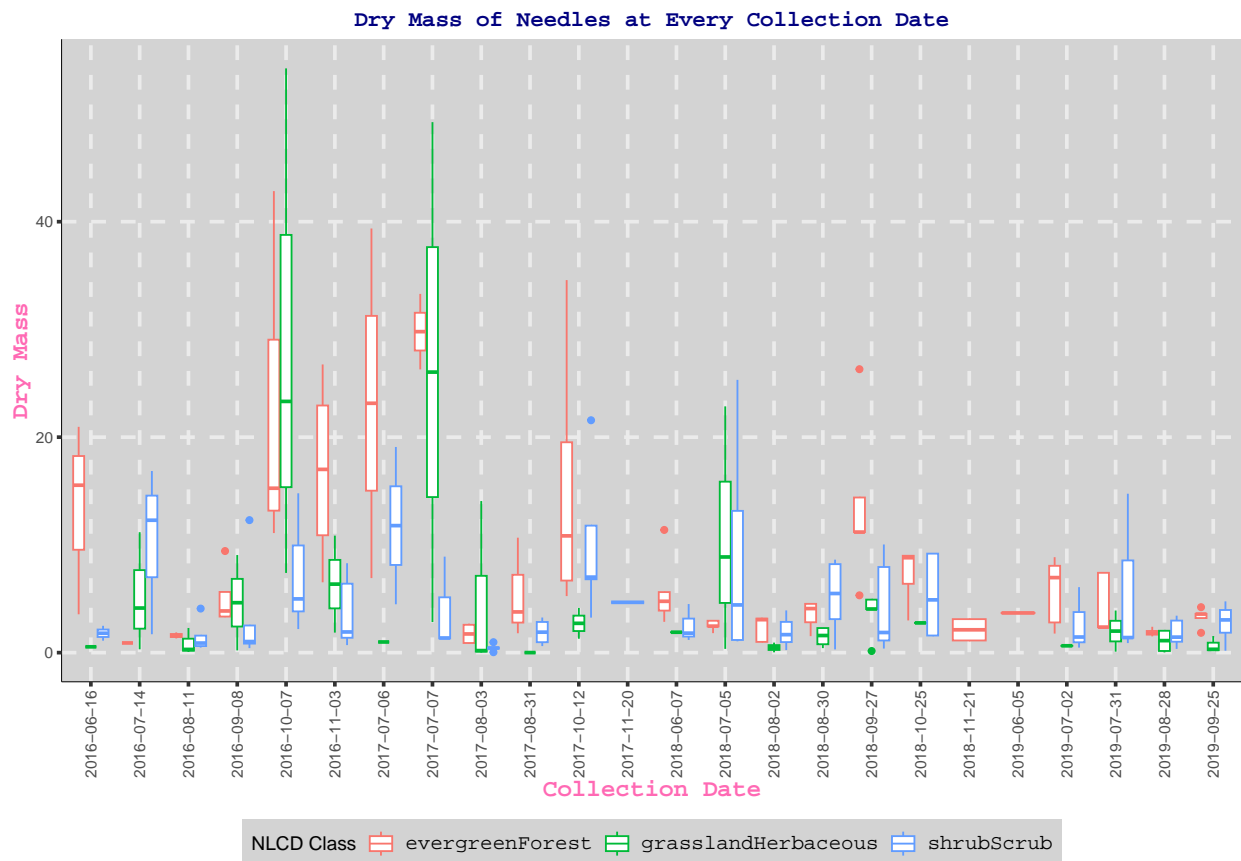
```
#7
NW_needles_facets <- #assigning data to plot
  ggplot(
    subset(Niwot_Litter, functionalGroup == "Needles"), #subsetting the data we want
    ↪ (just needles in functiongroup variable)
```

9

```
    aes(x = factor(collectDate), #assigning x axis (as factor to see all dates)
        y = dryMass)) + #assigning y axis
  geom_boxplot() + #make boxplot
  labs(x = 'Collection Date', #adding x axis label
       y = 'Dry Mass', #adding y axis label
       title = 'Dry Mass of Needles at Every Collection Date') + #adding title
  theme(axis.text.x = element_text(angle = 90, vjust = 0.5, hjust=1)) + #making x axis
  ↪    tick labels vertical
  facet_wrap(vars(nlcdClass)) #faceting the figure by NLCD class
NW_needles_facets #showing figure
```


Dry Mass of Needles at Every Collection Date
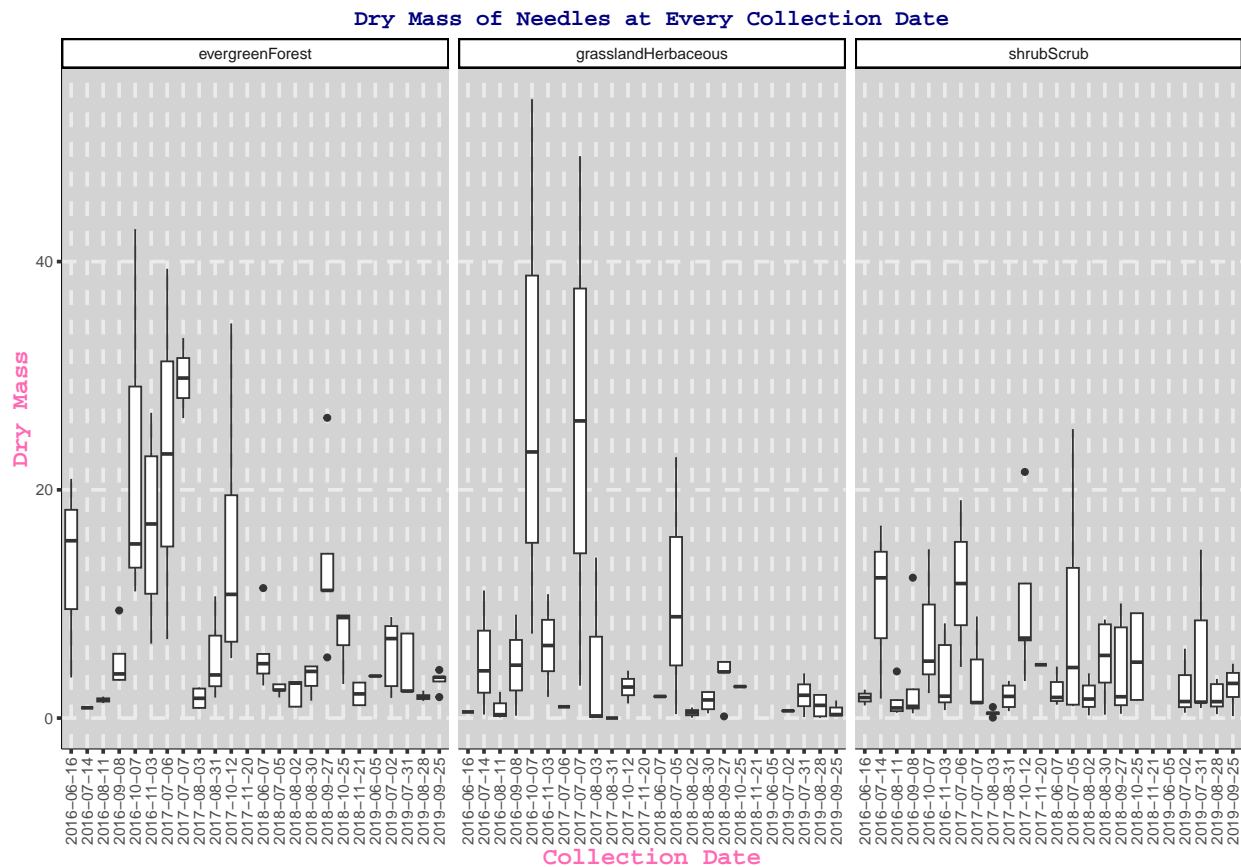
Question: Which of these plots (6 vs. 7) do you think is more effective, and why?

Answer: I think plot 7 is more effective because you can more easily compare the same NLCD class Needle dry mass between dates. In plot 6 it is visually difficult to compare across different dates. However, in plot 6 it is easier to compare NLCD classes within a single date, so it matters on what your question is. Ultimately, however, I think that plot 7 is more clear and less crowded/chaotic looking.