Assignment 5: Data Visualization

David Amanfu

OVERVIEW

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

Directions

1. Change "Student Name" on line 3 (above) with your name.
2. Work through the steps, **creating code and output** that fulfill each instruction.
3. Be sure to **answer the questions** in this assignment document.
4. When you have completed the assignment, **Knit** the text and code into a single PDF file.
5. After Knitting, submit the completed exercise (PDF file) to the dropbox in Sakai. Add your last name into the file name (e.g., "Fay_A05_DataVisualization.Rmd") prior to submission.

The completed exercise is due on Monday, February 14 at 7:00 pm.

Set up your session

1. Set up your session. Verify your working directory and load the tidyverse and cowplot packages. Upload the NTL-LTER processed data files for nutrients and chemistry/physics for Peter and Paul Lakes (use the tidy

$$`NTL-LTER_Lake_Chemistry_Nutrients_PeterPaul_Processed.csv`$$

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.

```
#1a
library(tidyverse)
```

```
## -- Attaching packages --------------------------------------- tidyverse 1.3.1 --

## v ggplot2 3.3.5     v purrr   0.3.4
## v tibble  3.1.6     v dplyr   1.0.8
## v tidyr   1.2.0     v stringr 1.4.0
## v readr   2.1.2     v forcats 0.5.1

## Warning: package 'tidyr' was built under R version 4.0.5
```

```
## Warning: package 'dplyr' was built under R version 4.0.5
```

```
## -- Conflicts --------------------------------------- tidyverse_conflicts() --
## x dplyr::filter() masks stats::filter()
## x dplyr::lag()    masks stats::lag()
```

```
library(cowplot)
library(ggplot2)
getwd()
```

```
## [1] "/Users/davidamanfu/Desktop/Duke MPP/Environ Data /Environmental_Data_Analytics_2022/Assignments
```

```
knitr::opts_knit$set(root.dir = "~/Desktop/Duke MPP/Environ Data /Environmental_Data_Analytics_2022/")
```

```
#1b
getwd()
```

```
## [1] "/Users/davidamanfu/Desktop/Duke MPP/Environ Data /Environmental_Data_Analytics_2022"
```

```
NTL_LTER_Lake_Nutrients <- read.csv("Data/Processed/NTL-LTER_Lake_Chemistry_Nutrients_PeterPaul_Processe
NTL_LTER_Lake_ChemistryPhysics <- read.csv("Data/Processed/NTL-LTER_Lake_ChemistryPhysics_PeterPaul_Pro
NEON_NIWO_Litter <- read.csv("Data/Processed/NEON_NIWO_Litter_mass_trap_Processed.csv", stringsAsFactor

#2
class(NEON_NIWO_Litter$collectDate)
```

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

```
class(NTL_LTER_Lake_ChemistryPhysics$sampledate)
```

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

```
class(NTL_LTER_Lake_Nutrients$sampledate)
```

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

```
NTL_LTER_Lake_ChemistryPhysics$sampledate <- as.Date.factor(NTL_LTER_Lake_ChemistryPhysics$sampledate, 
NTL_LTER_Lake_Nutrients$sampledate <- as.Date(NTL_LTER_Lake_Nutrients$sampledate, tryFormats ="%Y-%m-%d
NEON_NIWO_Litter$collectDate <- as.Date(NEON_NIWO_Litter$collectDate, tryFormats ="%Y-%m-%d")
class(NEON_NIWO_Litter$collectDate)
```

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

```
class(NTL_LTER_Lake_ChemistryPhysics$sampledate)
```

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

```
class(NTL_LTER_Lake_Nutrients$sampledate)
```

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

Define your theme

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

```
#3
library(ggthemes)
```

```
##
## Attaching package: 'ggthemes'

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

```
library(hrbrthemes)
library(extrafont)
```

```
## Registering fonts with R
```

```
#remotes::install_version("Rttf2pt1", version = "1.3.8")
library(extrafontdb)
#font_import()
#fonts()

AmanfuTheme <- theme_classic(base_size = 14) +
  theme(axis.text = element_text(color = "black"),
        legend.position = "right", legend.direction = "vertical",) #alternative: legend.position + lege
AmanfuTheme2 <- theme_ipsum()+
  theme(legend.position = "bottom",
        legend.key = element_rect(fill = "white", colour = "black"),legend.direction = "horizontal",
        legend.title = element_text(face = "bold"))
AmanfuTheme3 <- theme_fivethirtyeight()+theme(legend.position = "right", legend.direction = "vertical")
theme_set(AmanfuTheme2)
```

Create graphs

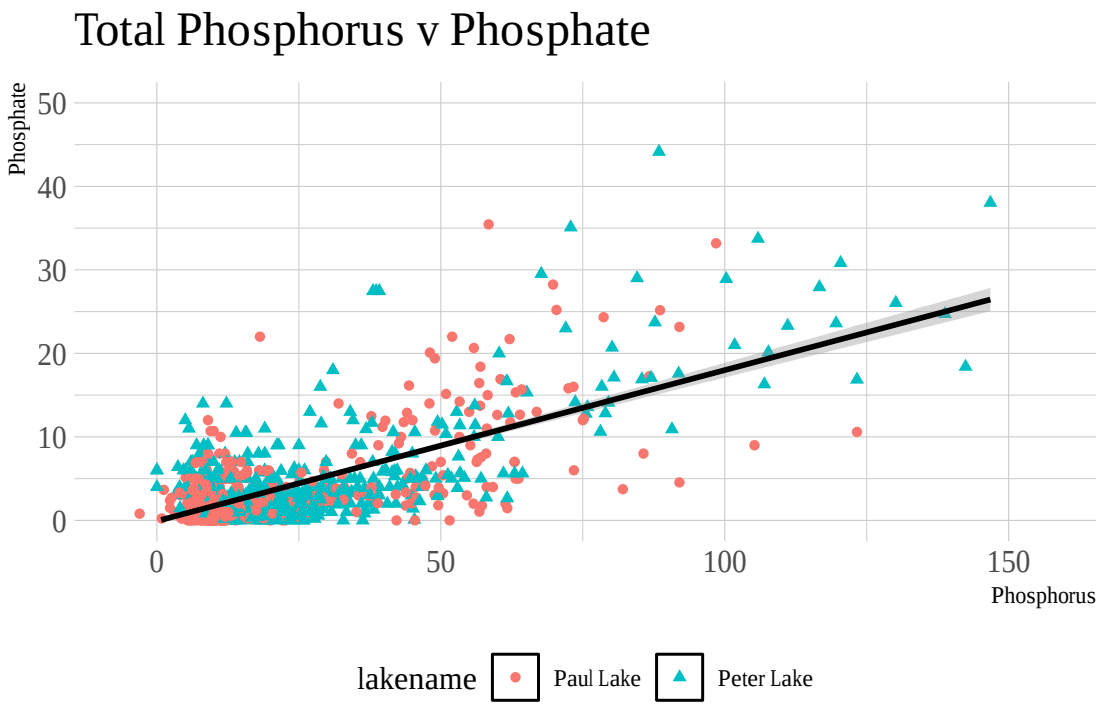
For numbers 4-7, create ggplot graphs and adjust aesthetics to follow best practices for data visualization. Ensure your theme, color palettes, axes, and additional aesthetics are edited accordingly.

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 `ylim()`).

```r
#4
tppo4 <-ggplot(NTL_LTER_Lake_Nutrients,aes(x= tp_ug, y=po4)) +
  geom_point(aes(shape = lakename, color = lakename, fill = lakename)) +
  ylim(0,50) +
  labs(title = "Total Phosphorus v Phosphate", x = "Phosphorus", y= "Phosphate") +
  geom_smooth(method = "lm",color = "black")
tppo4
```

```
## `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).
```



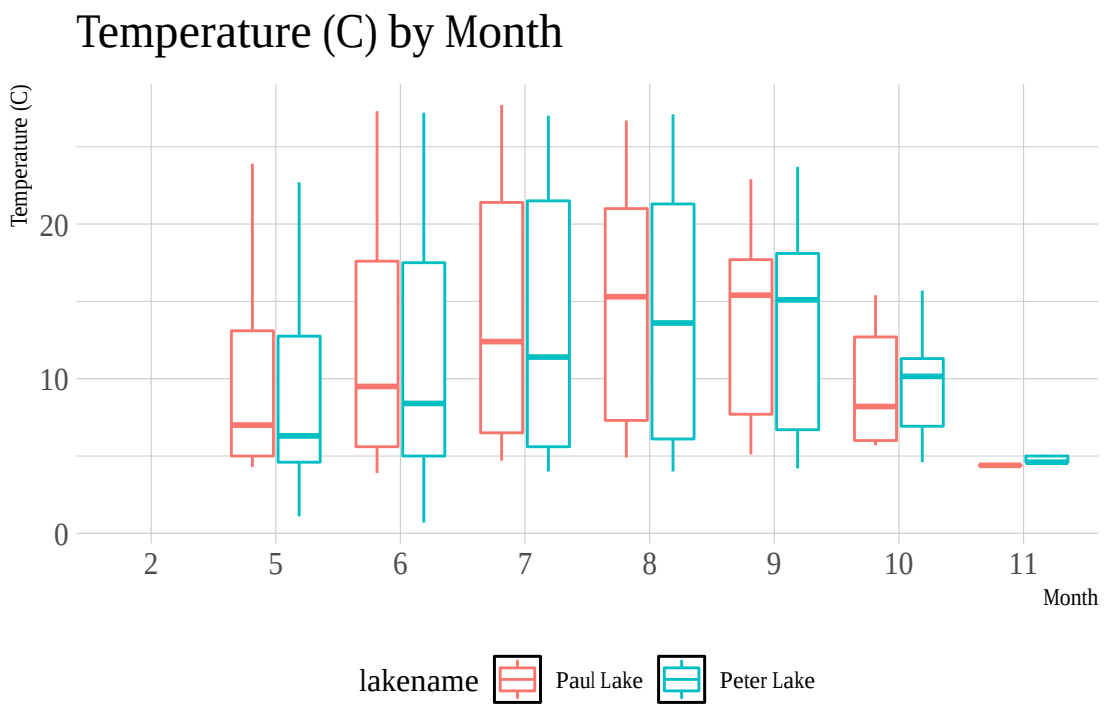
5.

$$NTL - LTER$$

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.

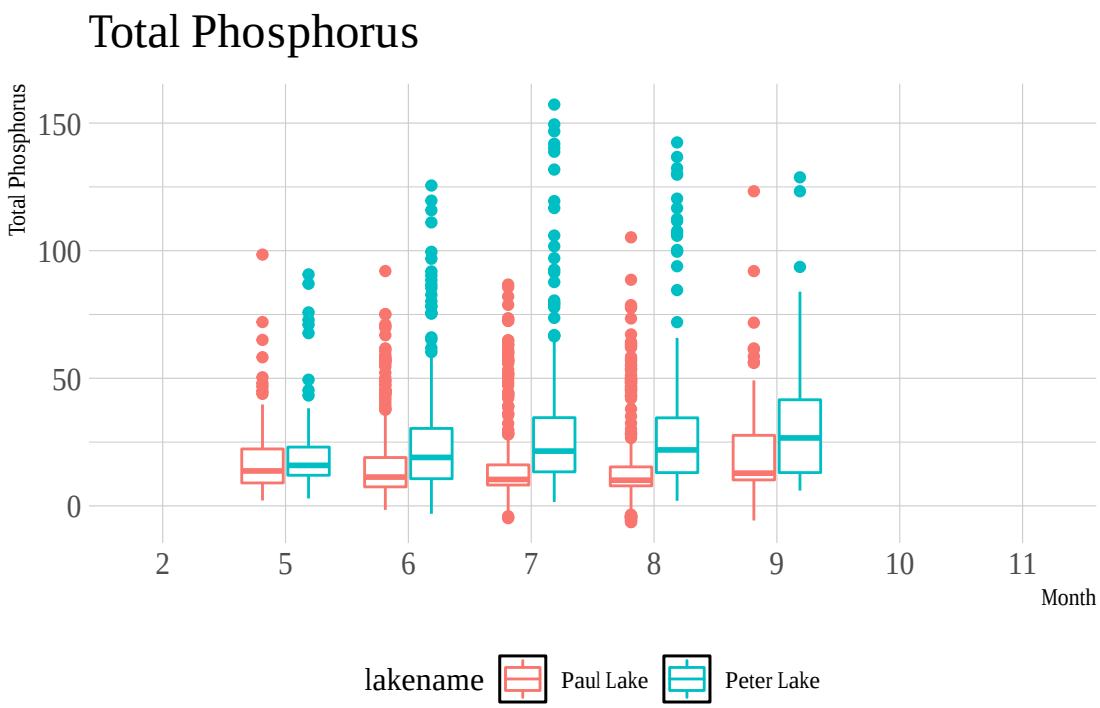
```r
#5
tempplot <- ggplot(NTL_LTER_Lake_Nutrients,aes(x=factor(month))) +
  geom_boxplot(aes(y=temperature_C,color=lakename)) +
  labs(title = "Temperature (C) by Month", x = "Month", y= "Temperature (C)")
tp_plot <- ggplot(NTL_LTER_Lake_Nutrients,aes(x= factor(month))) +
  geom_boxplot(aes(y=tp_ug, color = lakename)) +
  #geom_violin(aes(y=tp_ug,color=lakename))+
  labs(title = "Total Phosphorus", x = "Month", y= "Total Phosphorus")
tn_plot <- ggplot(NTL_LTER_Lake_Nutrients,aes(x= factor(month))) +
  geom_boxplot(aes(y=tn_ug, color = lakename)) +
  labs(title = "Total Nitrates", x = "Month", y= "Total Nitrates")

tempplot
```

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



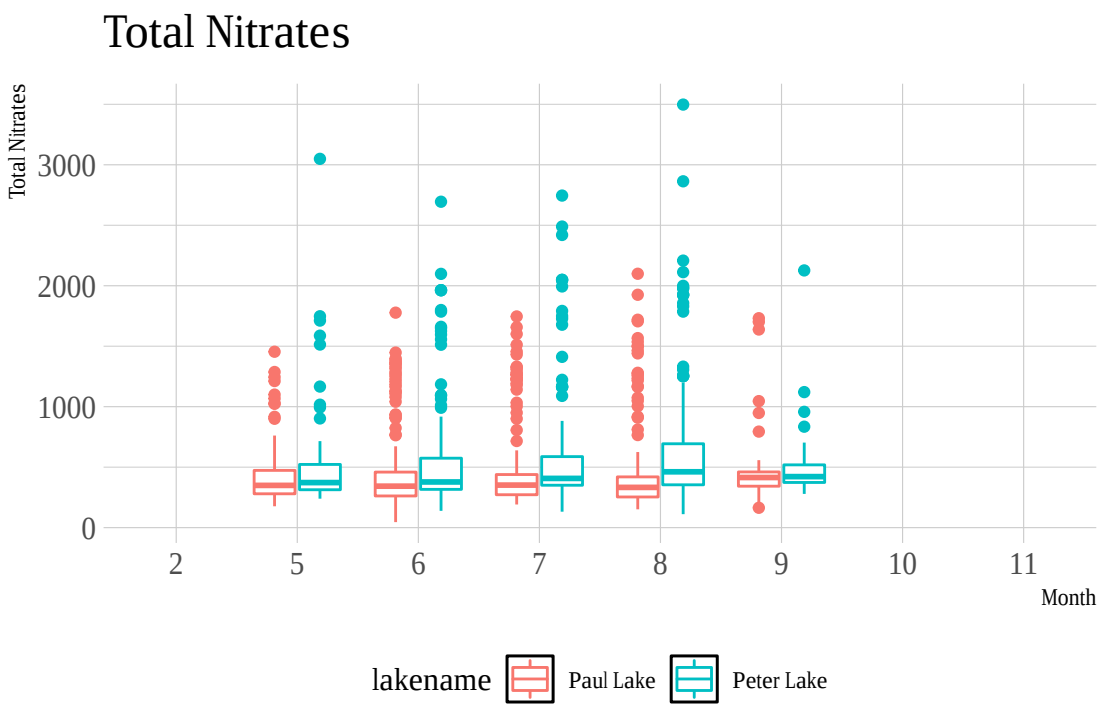

```r
tp_plot
```

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




```
tn_plot
```

```
## Warning: Removed 21583 rows containing non-finite values (stat_boxplot).
```



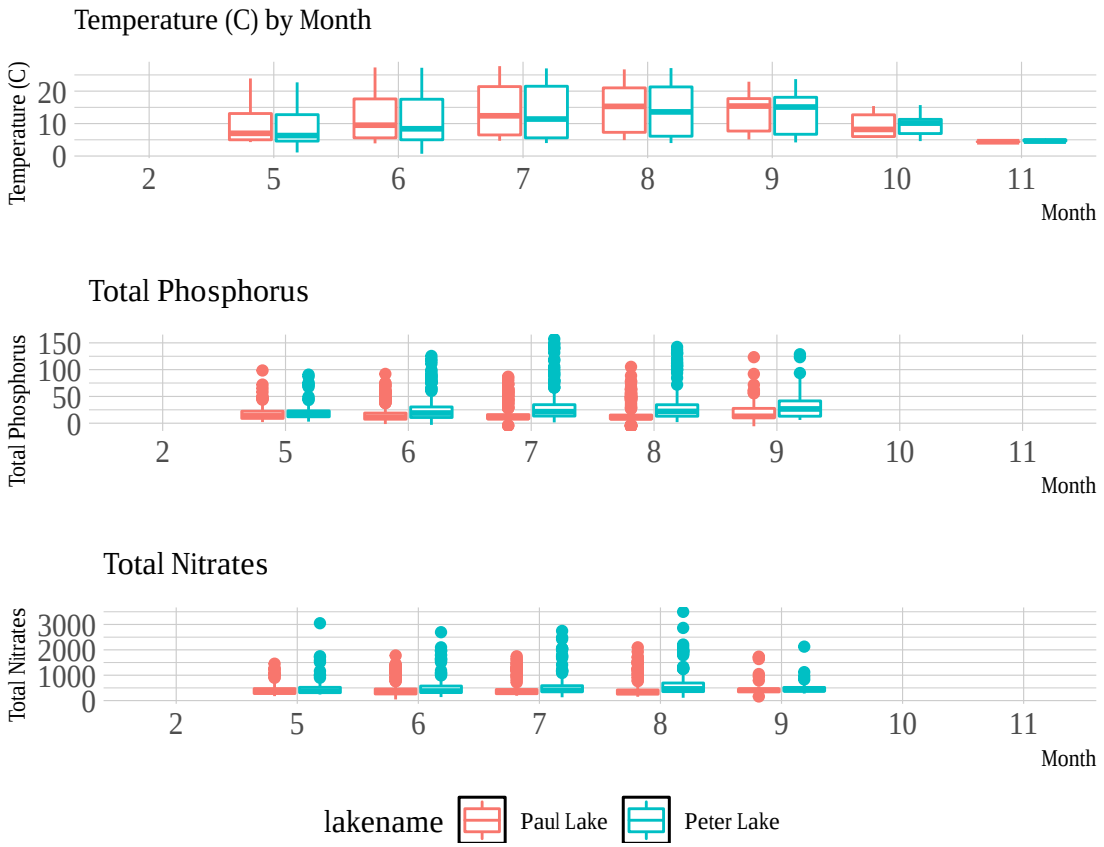

```
temptptn <- plot_grid(
  tempplot+theme(legend.position = "none",plot.margin=margin(10,30,0,30), plot.title = element_text(size
  tp_plot+theme(legend.position = "none", plot.margin=margin(10,30,0,30), plot.title = element_text(size
  tn_plot+theme(legend.position = "none", plot.margin=margin(10,30,10,30), plot.title = element_text(si
  get_legend(tempplot+theme(legend.direction = "horizontal")),
  align="hv",nrow = 4,rel_heights = c(1.5,1.5,1.5,.25),axis="tb")
```

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

```
temptptn
```



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

> Answer: So it appears that there's a lot more variation in Phosophorous and Nitrates(?) in Peter lake than there is in Paul lake over the course of the seasons. It is also notable that phosphorous levels seem to follow temperature patterns, while nitrates do not. there's nothing particularly interesting or notable about the temperature plot; it appears that these lakes are either next to each other or at worst in similar climates. We've also got a lot of potential outliers in our measurements, which I'm not exactly sure what that might suggest – perhaps too small a sample size? Or perhaps too largely bucketed/clustered x values. Perhaps weekly might give us smoother data.

6. *NiwotRidge*

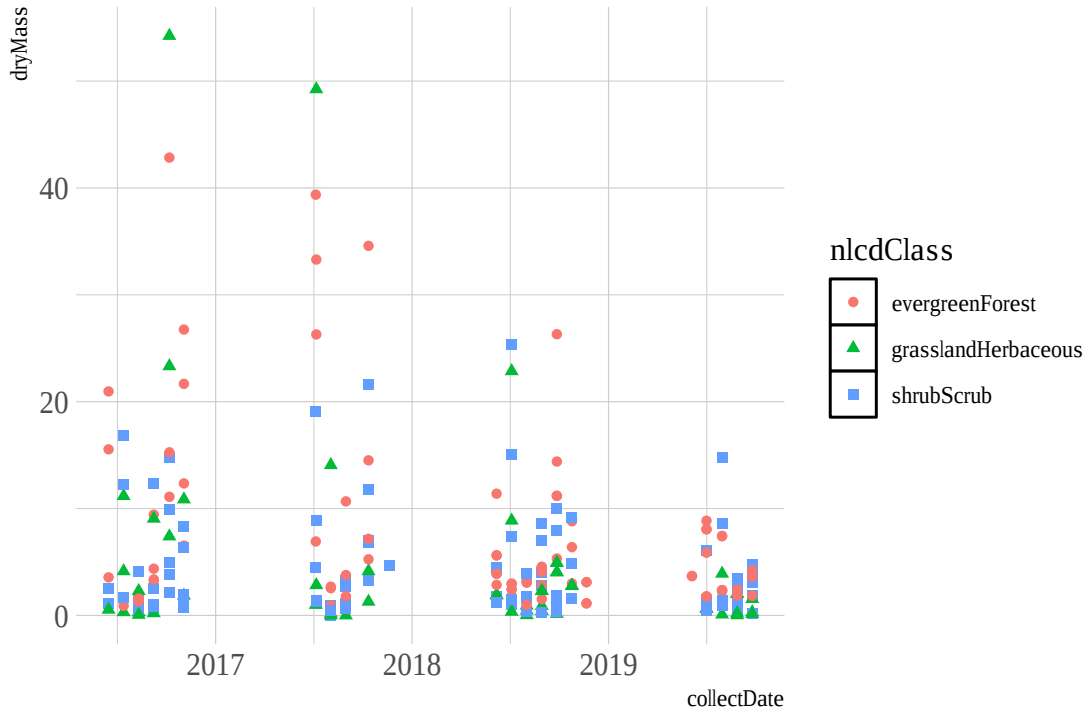
   Plot a subset of the litter dataset by displaying only the "Needles" functional group. Plot the dry mass of needle litter by date and separate by NLCD class with a color aesthetic. (no need to adjust the name of each land use)

7. *NiwotRidge*

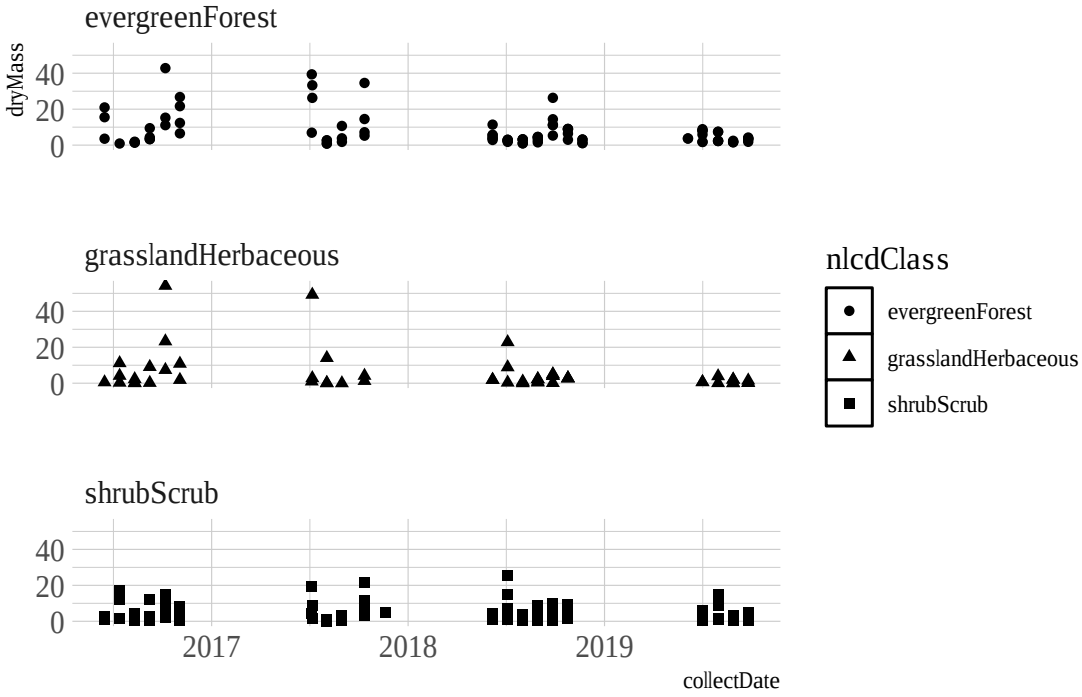
   Now, plot the same plot but with NLCD classes separated into three facets rather than separated by color.

```
#6
Niwot_needle <- ggplot(subset(NEON_NIWO_Litter, functionalGroup=="Needles"))+
  geom_point(aes(y=dryMass,x=collectDate, color = nlcdClass, shape=nlcdClass))+
```

```
  theme(legend.position = "right",legend.direction = "vertical")
Niwot_needle
```




```
#7
Niwot_needle_facet <- ggplot(subset(NEON_NIWO_Litter, functionalGroup=="Needles"))+
  geom_point(aes(y=dryMass,x=collectDate,shape=nlcdClass))+
  facet_wrap(vars(nlcdClass),nrow=3)+
  theme(legend.position = "right",legend.direction = "vertical")
Niwot_needle_facet
```



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

Answer: For information that is actually quite clustered, it can sometimes help to plot with facets rather than on a single plot. I think Plot 7 is more effective in this case because we have better clarity on the different classes of needles collected. Plot 6 is super crowded and not a lot of information can be gleaned from it.