

# Assignment 5: Data Visualization

*Yang Wang*

## 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., “Salk\_A05\_DataVisualization.Rmd”) prior to submission.

The completed exercise is due on Tuesday, February 11 at 1:00 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 (tidy and gathered) and the processed data file for the Niwot Ridge litter dataset.
2. Make sure R is reading dates as date format; if not change the format to date.

```
#1
getwd()

## [1] "C:/Users/26059/OneDrive/Desktop/ENV 872 R/Yang_ENV872"

library(tidyverse)

## -- Attaching packages ----- tidyverse 1.3.0 --

## v ggplot2 3.2.1    v purrr   0.3.3
## v tibble  2.1.3    v dplyr  0.8.3
## v tidyr   1.0.0    v stringr 1.4.0
## v readr   1.3.1    v forcats 0.4.0

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

#install.packages("cowplot")
library(cowplot)
```

```
##
## *****

## Note: As of version 1.0.0, cowplot does not change the

## default ggplot2 theme anymore. To recover the previous

## behavior, execute:
## theme_set(theme_cowplot())

## *****
```

```
library(viridis)
```

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

```
lake<- read.csv("./Data/Processed/NTL-LTER_Lake_Chemistry_Nutrients_PeterPaul_Processed.csv")
litter<- read.csv("./Data/Processed/NEON_NIWO_Litter_mass_trap_Processed.csv")

#2
class(lake$sampleddate)
```

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

```
lake$sampleddate <- as.Date(lake$sampleddate, format = "%Y-%m-%d")
class(litter$collectDate)
```

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

```
litter$collectDate <- as.Date(litter$collectDate, format = "%Y-%m-%d")
class(lake$sampleddate)
```

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

```
class(litter$collectDate)
```

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

## Define your theme

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

```
mytheme <- theme_classic(base_size = 14) + # for the size of label
  theme(axis.text = element_text(color = "black"),
        legend.position = "bottom", legend.text = element_text(size = 12), legend.title = element_text(s
  theme_set(mytheme)
```

## 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.

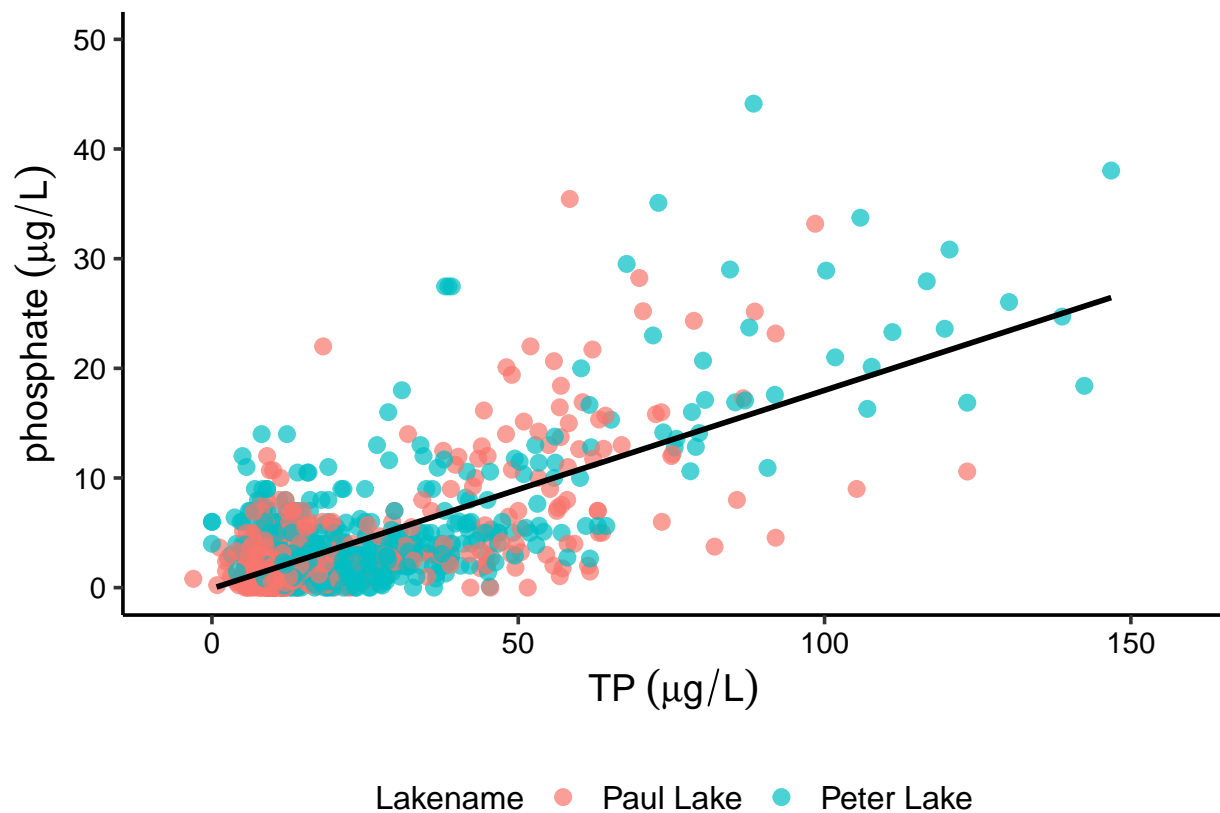
4. [NTL-LTER] Plot total phosphorus by phosphate, with separate aesthetics for Peter and Paul lakes. Add a line of best fit and color it black. Adjust your axes to hide extreme values.

```
tp.po4 <-  
  ggplot(lake, aes(x = tp_ug, y = po4, color = lakename)) +  
  geom_point(alpha = 0.7, size = 2.5) + xlab("tp(ug/l)") + ylab("po4(ug/l)") +  
  geom_smooth(method = lm, se = FALSE, color = "black") +  
  labs(x = expression(TP ~ (mu*g / L)), y = expression(phosphate ~ (mu*g / L)),  
       color = "Lakename") +  
  scale_y_continuous(limits = c(0, 50))  
print(tp.po4)
```

```
## Warning: Removed 21947 rows containing non-finite values (stat_smooth).
```

```
## Warning: Removed 21947 rows containing missing values (geom_point).
```

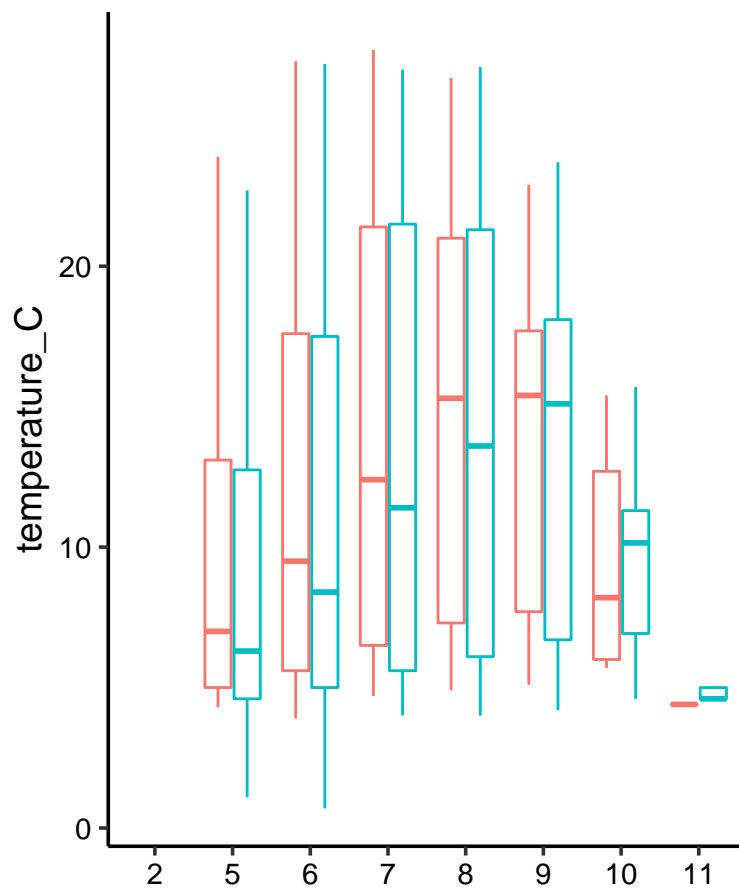
```
## Warning: Removed 2 rows containing missing values (geom_smooth).
```



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.

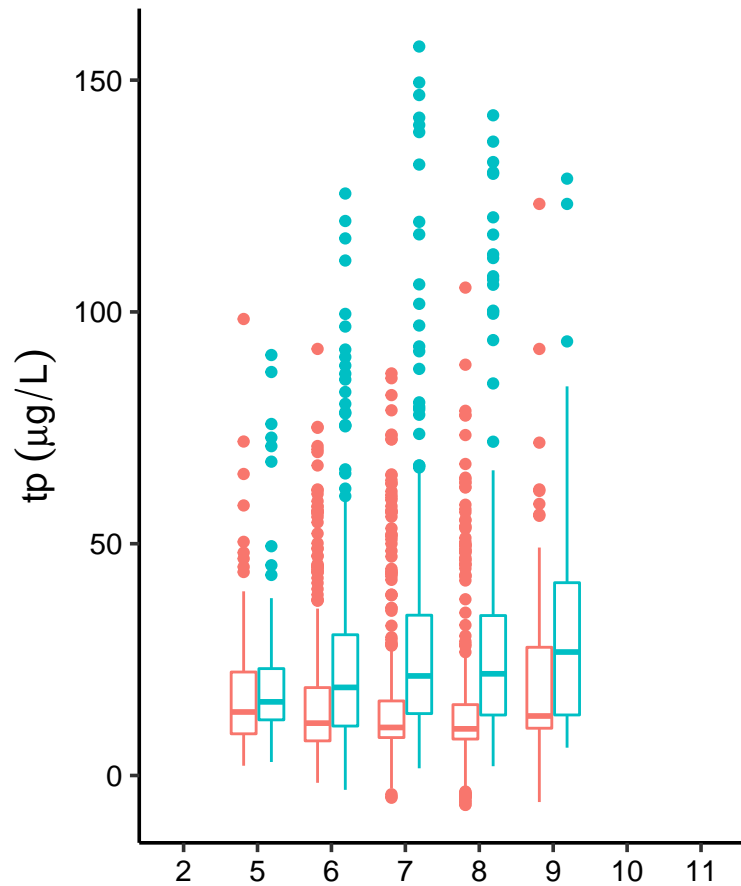
```
tem.plot <-  
  ggplot(lake, aes(x = as.factor(month), y = temperature_C, color = lakename)) +  
  geom_boxplot() +  
  labs(x = "", y = "temperature_C", color = "") +  
  theme(legend.position = "none")  
tem.plot
```

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



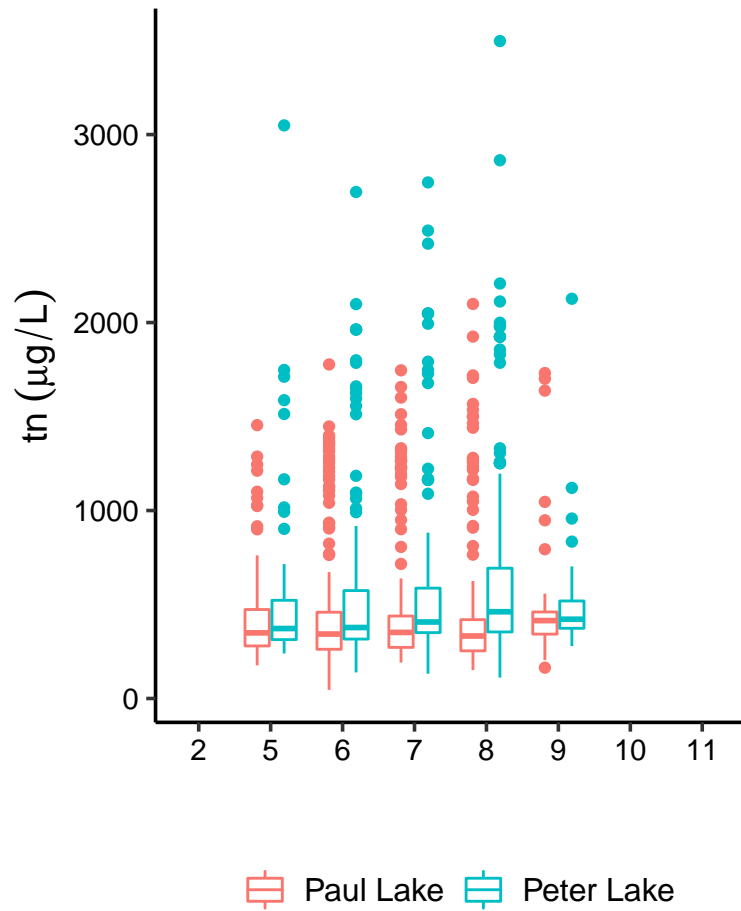
```
tp.plot <-  
  ggplot(lake, aes(x = as.factor(month), y = tp_ug, color = lakename)) +  
  geom_boxplot() +  
  labs(x = "", y = expression(tp ~ (mu*g / L)), color = "") +  
  theme(legend.position = "none")  
tp.plot
```

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



```
tn.plot <-
  ggplot(lake, aes(x = as.factor(month), y = tn_ug, color = lakename)) +
  geom_boxplot() +
  labs(x = "", y = expression(tn ~ (mu*g / L)), color = "")
tn.plot
```

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



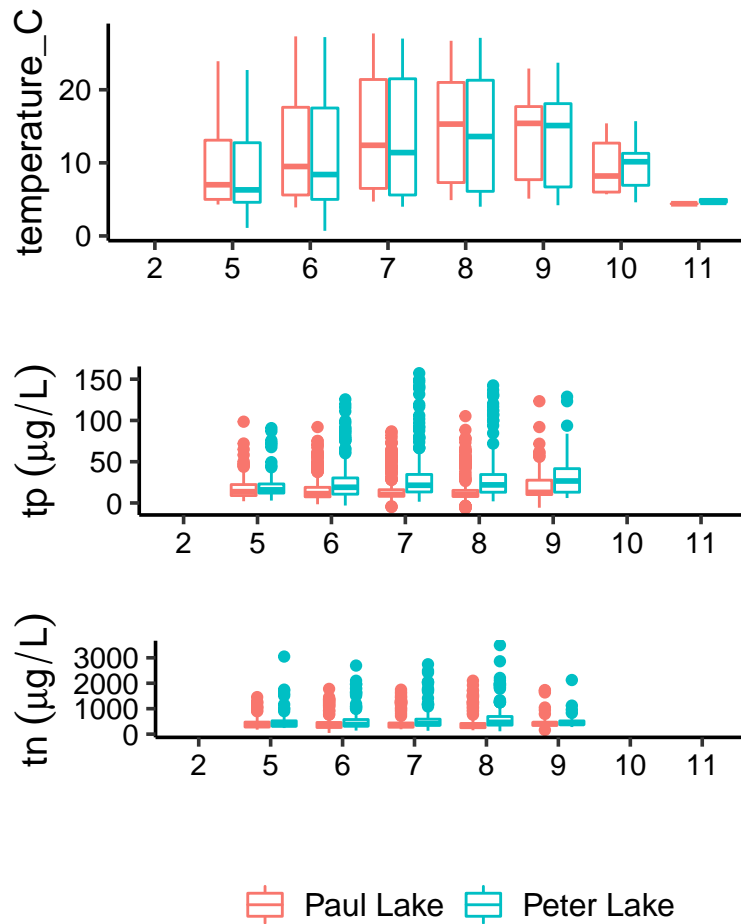
```
plot_grid(tem.plot, tp.plot, tn.plot, nrow = 3, align = 'h', rel_heights = c(1.25, 1))
```

```
## 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).
```

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

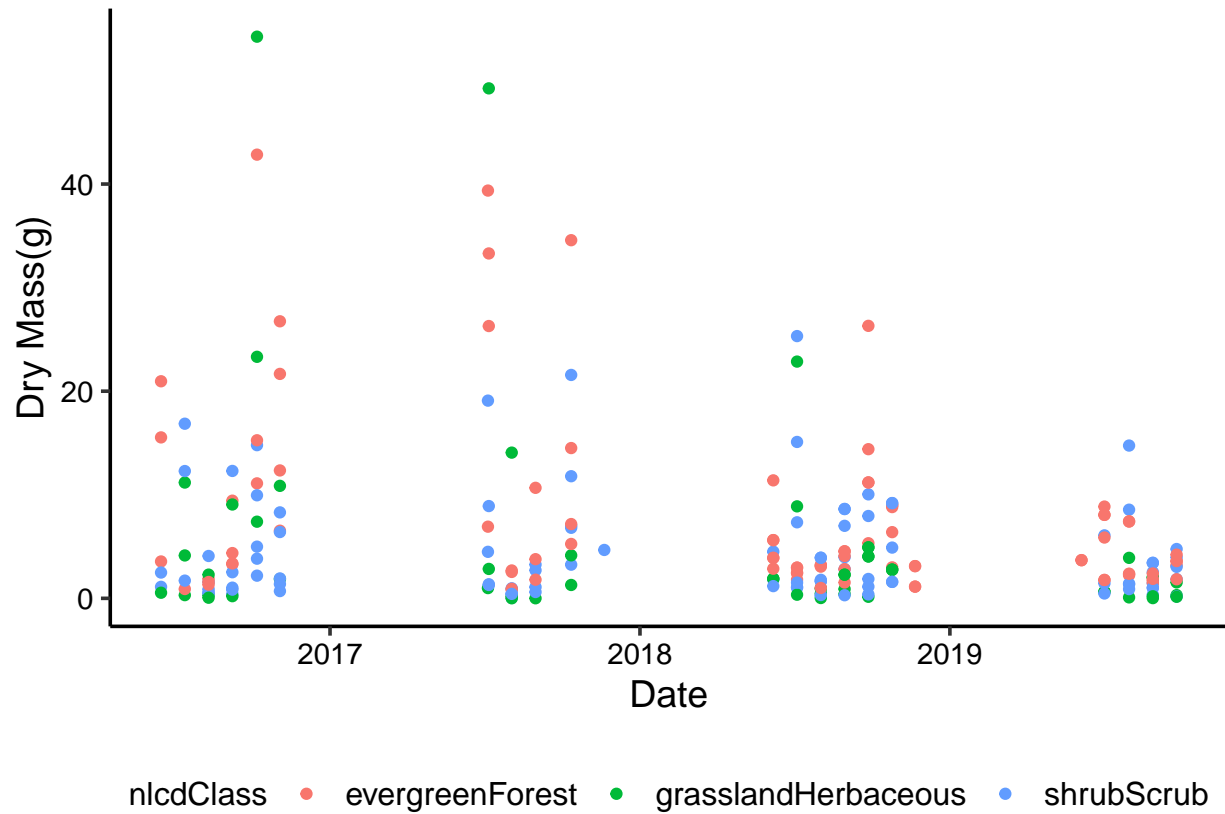


Question: What do you observe about the variables of interest over seasons and between lakes?

Answer: Temperature increases in summer and decreases in the fall in both lakes. During spring and summer, Peter lake has higher temperatures than Paul lake. In winter, Peter lake has higher temperatures than Paul lake. Total potassium and total nitrate show an increase in Peter lake from May to November, but do not show big fluctuation in Paul lake.

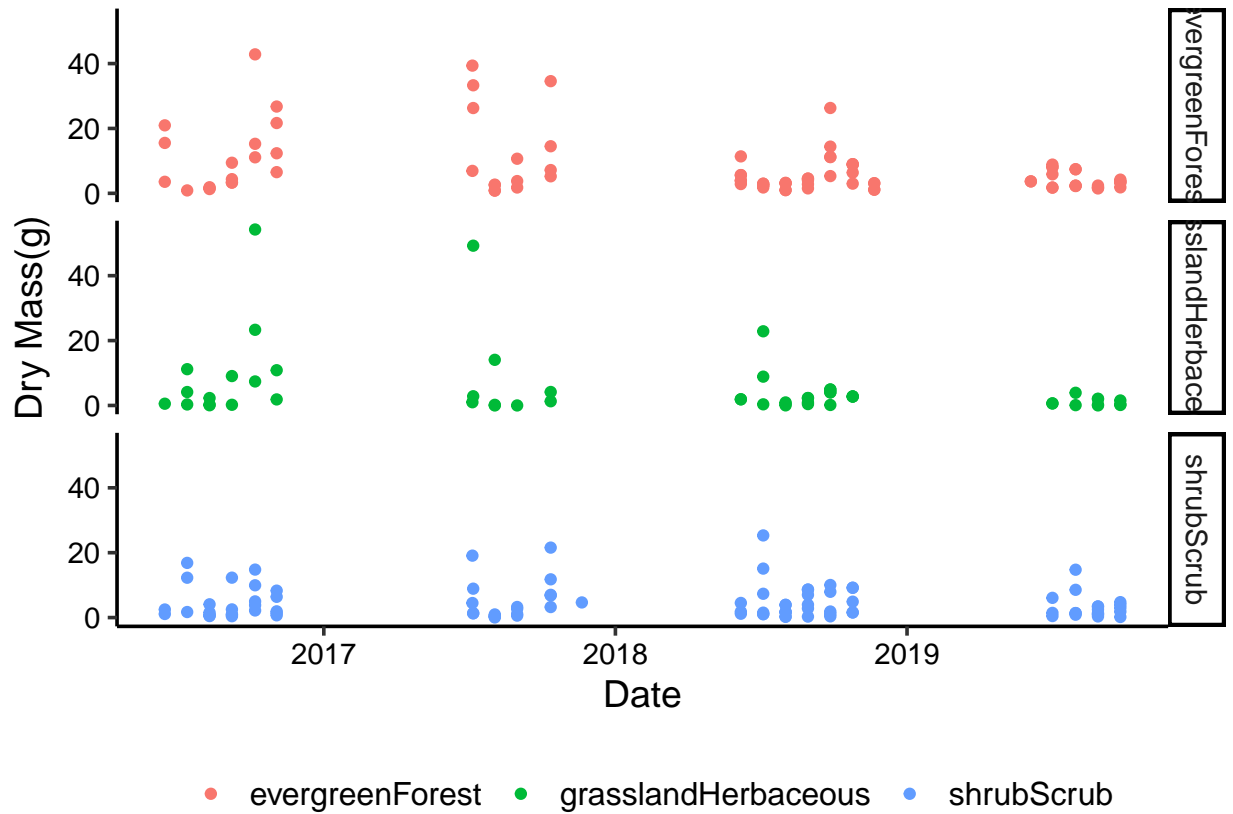
6. [Niwot Ridge] Plot a subset of the litter dataset by displaying only the “Needles” functional group. Plot the dry mass of needle litter by date and separate by NLCD class with a color aesthetic. (no need to adjust the name of each land use)
7. [Niwot Ridge] Now, plot the same plot but with NLCD classes separated into three facets rather than separated by color.

```
needle.plot <-
  ggplot(subset(litter, functionalGroup=="Needles")) +
  geom_point(aes(x = collectDate, y = dryMass, colour = nlcdClass)) +
  labs(x = "Date", y = "Dry Mass(g)")
print(needle.plot)
```



```
facet.plot <-
  ggplot(subset(litter, functionalGroup == "Needles"), aes(x = collectDate, y = dryMass, color = nlcdClass)) +
  labs(x = "Date", y = expression("Dry Mass(g)"), color = "") +
  geom_point() +
  facet_grid(nlcdClass~.)
facet.plot
```





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

Answer: The plot with facet is more effective because it can show each class and their relative distribution very clearly. The plot without facet has three colors mix together making it hard to show distribution and comparison.