Assignment 6: GLMs week 1 (t-test and ANOVA)

Monisha Eadala

OVERVIEW

This exercise accompanies the lessons in Environmental Data Analytics on t-tests and ANOVAs.

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_A06_GLMs_Week1.Rmd") prior to submission.

The completed exercise is due on Tuesday, February 18 at 1:00 pm.

Set up your session

1. Check your working directory, load the `tidyverse`, `cowplot`, and `agricolae` packages, and import the NTL-LTER_Lake_Nutrients_PeterPaul_Processed.csv dataset.

2. Change the date column to a date format. Call up `head` of this column to verify.

```
#1
getwd()

## [1] "/Users/monishaeadala/Environmental_Data_Analytics_2020"

library(tidyverse)
library(cowplot)
library(agricolae)

PeterPaul.nutrients <- read.csv("./Data/Processed/NTL-
LTER_Lake_Nutrients_PeterPaul_Processed.csv")

#2
PeterPaul.nutrients$sampledate <- as.Date(PeterPaul.nutrients$sampledate,
format = "%Y-%m-%d") # Setting date to date format
class(PeterPaul.nutrients$sampledate) # Calling up head of the column to
verify

## [1] "Date"
```

Wrangle your data

3. Wrangle your dataset so that it contains only surface depths and only the years 1993-1996, inclusive. Set month as a factor.

```
class(PeterPaul.nutrients$month) # Checking class of month column

## [1] "integer"

PeterPaul.nutrients$month <- as.factor(PeterPaul.nutrients$month) # Setting
month as a factor

surface.depths <-
  PeterPaul.nutrients %>%
  filter(depth == 0) %>%
  filter(year4 == 1993 | year4 == 1994 | year4 == 1995 | year4 == 1996) #
Wrangling the dataset so that it contains only surface depths and only the
years 1993-1996, inclusive
```

Analysis

Peter Lake was manipulated with additions of nitrogen and phosphorus over the years 1993-1996 in an effort to assess the impacts of eutrophication in lakes. You are tasked with finding out if nutrients are significantly higher in Peter Lake than Paul Lake, and if these potential differences in nutrients vary seasonally (use month as a factor to represent seasonality). Run two separate tests for TN and TP.

4. Which application of the GLM will you use (t-test, one-way ANOVA, two-way ANOVA with main effects, or two-way ANOVA with interaction effects)? Justify your choice.

Answer: I will use two-way ANOVA with interaction effects, because there are two categorical explanatory variable with one categorical variable wiht two categories and than other with more than two catergories. T-test will require only one categorical explanatory variable with only two categories, and one-way ANOVA will require only one categorical explanatory variable with more than two categories; therefore, they are both not suitable. In this case, we have “month” and “lakename” as two categorical explanatory variables; therefore two-way ANOVA is the most suitable. A two-way ANOVA with interaction effects is safer to use here since we are not sure how the the variables “months” and “lakes” interact with each other. After running the two-way ANOVA with interaction effects, if the interaction is significant, we interpret pairwise differences for the interaction. If the interaction is not significant, we interpret differences for the main effects only.

5. Run your test for TN. Include examination of groupings and consider interaction effects, if relevant.

6. Run your test for TP. Include examination of groupings and consider interaction effects, if relevant.

```
#5
surface.depths.TN.anova.2way <- aov(data = surface.depths, tn_ug ~ lakename *
month)
summary(surface.depths.TN.anova.2way) # Format as aov

##                 Df  Sum Sq Mean Sq F value   Pr(>F)
## lakename        1 2468595 2468595  36.414 2.91e-08 ***
```

```
## month            4  459542  114885   1.695    0.157
## lakename:month   4  288272   72068   1.063    0.379
## Residuals       97 6575834   67792
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
## 23 observations deleted due to missingness
```

```
surface.depths.TN.anova.2way2 <- lm(data = surface.depths, tn_ug ~ lakename *
month)
summary(surface.depths.TN.anova.2way2) # Format as lm; another option
```

```
##
## Call:
## lm(formula = tn_ug ~ lakename * month, data = surface.depths)
##
## Residuals:
##     Min      1Q  Median      3Q     Max
## -357.88 -118.10  -10.41   50.58 1353.86
##
## Coefficients:
##                             Estimate Std. Error t value Pr(>|t|)
## (Intercept)                   300.51     106.30   2.827   0.0057 **
## lakenamePeter Lake             84.43     144.86   0.583   0.5614
## month6                         23.61     123.64   0.191   0.8489
## month7                         53.12     127.05   0.418   0.6768
## month8                         36.00     127.05   0.283   0.7775
## month9                        105.82     184.11   0.575   0.5668
## lakenamePeter Lake:month6     200.49     170.90   1.173   0.2436
## lakenamePeter Lake:month7     271.82     176.18   1.543   0.1261
## lakenamePeter Lake:month8     325.05     174.20   1.866   0.0651 .
## lakenamePeter Lake:month9      59.70     278.35   0.214   0.8306
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Residual standard error: 260.4 on 97 degrees of freedom
##   (23 observations deleted due to missingness)
## Multiple R-squared:  0.3285, Adjusted R-squared:  0.2662
## F-statistic: 5.272 on 9 and 97 DF,  p-value: 7.729e-06
```

```
# Since the interaction is not significant for the above, we interpret
differences for the main effects only

#6
surface.depths.TP.anova.2way <- aov(data = surface.depths, tp_ug ~ lakename *
month)
summary(surface.depths.TP.anova.2way) # Format as aov
```

```
##                 Df Sum Sq Mean Sq F value Pr(>F)
## lakename         1  10228   10228  98.914 <2e-16 ***
## month            4    813     203   1.965 0.1043
## lakename:month   4   1014     254   2.452 0.0496 *
```

```
## Residuals      119  12305     103
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
## 1 observation deleted due to missingness
```

```
surface.depths.TP.anova.2way2 <- lm(data = surface.depths, tp_ug ~ lakename *
month)
summary(surface.depths.TP.anova.2way2) # Format as lm; another option
```

```
##
## Call:
## lm(formula = tp_ug ~ lakename * month, data = surface.depths)
##
## Residuals:
##     Min      1Q  Median      3Q     Max
## -17.384  -4.473  -0.693   1.939  32.489
##
## Coefficients:
##                            Estimate Std. Error t value Pr(>|t|)
## (Intercept)                 11.4740     4.1514   2.764  0.00662 **
## lakenamePeter Lake           4.3136     5.6574   0.762  0.44729
## month6                      -0.9179     4.8288  -0.190  0.84957
## month7                      -1.7271     4.7936  -0.360  0.71927
## month8                      -2.0872     4.7936  -0.435  0.66405
## month9                      -0.7380     6.1575  -0.120  0.90480
## lakenamePeter Lake:month6  13.4882     6.6207   2.037  0.04384 *
## lakenamePeter Lake:month7  20.3440     6.6207   3.073  0.00263 **
## lakenamePeter Lake:month8  12.7937     6.5722   1.947  0.05394 .
## lakenamePeter Lake:month9  11.1697     8.8622   1.260  0.21000
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Residual standard error: 10.17 on 119 degrees of freedom
##   (1 observation deleted due to missingness)
## Multiple R-squared:  0.4949, Adjusted R-squared:  0.4567
## F-statistic: 12.95 on 9 and 119 DF,  p-value: 3.24e-14
```

```
# Since the interaction is significant, we interpret pairwise differences for
the interaction
TukeyHSD(surface.depths.TP.anova.2way) # Runs a post-hoc test for pairwise
differences
```

```
##   Tukey multiple comparisons of means
##     95% family-wise confidence level
##
## Fit: aov(formula = tp_ug ~ lakename * month, data = surface.depths)
##
## $lakename
##                            diff      lwr      upr p adj
## Peter Lake-Paul Lake 17.80939 14.26365 21.35513     0
##
```

```
## $month
##           diff         lwr       upr     p adj
## 6-5  6.3451786  -2.8038335 15.494191 0.3119085
## 7-5  8.8661326  -0.2828796 18.015145 0.0622967
## 8-5  4.8191843  -4.2626118 13.900980 0.5839528
## 9-5  5.4951391  -6.7194172 17.709695 0.7243206
## 7-6  2.5209540  -4.2125367  9.254445 0.8376355
## 8-6 -1.5259943  -8.1678685  5.115880 0.9688094
## 9-6 -0.8500395 -11.3776631  9.677584 0.9994372
## 8-7 -4.0469483 -10.6888225  2.594926 0.4453729
## 9-7 -3.3709935 -13.8986170  7.156630 0.9012092
## 9-8  0.6759548  -9.7933076 11.145217 0.9997679
##
## $`lakename:month`
##                                   diff         lwr         upr     p adj
## Peter Lake:5-Paul Lake:5    4.3135714 -13.9293175  22.5564604 0.9989515
## Paul Lake:6-Paul Lake:5    -0.9178824 -16.4886641  14.6528993 1.0000000
## Peter Lake:6-Paul Lake:5   16.8838889   1.4263507  32.3414270 0.0206973
## Paul Lake:7-Paul Lake:5    -1.7271111 -17.1846493  13.7304270 0.9999981
## Peter Lake:7-Paul Lake:5   22.9304706   7.3596889  38.5012523 0.0002415
## Paul Lake:8-Paul Lake:5    -2.0872222 -17.5447604  13.3703159 0.9999902
## Peter Lake:8-Paul Lake:5   15.0200000  -0.3355071  30.3755071 0.0607728
## Paul Lake:9-Paul Lake:5    -0.7380000 -20.5935673  19.1175673 1.0000000
## Peter Lake:9-Paul Lake:5   14.7452500  -6.4208558  35.9113558 0.4316694
## Paul Lake:6-Peter Lake:5   -5.2314538 -19.9572479   9.4943403 0.9787107
## Peter Lake:6-Peter Lake:5  12.5703175  -2.0356832  27.1763181 0.1571717
## Paul Lake:7-Peter Lake:5   -6.0406825 -20.6466832   8.5653181 0.9437275
## Peter Lake:7-Peter Lake:5  18.6168992   3.8911050  33.3426933 0.0032014
## Paul Lake:8-Peter Lake:5   -6.4007937 -21.0067943   8.2052070 0.9208652
## Peter Lake:8-Peter Lake:5  10.7064286  -3.7915495  25.2044066 0.3464892
## Paul Lake:9-Peter Lake:5   -5.0515714 -24.2516579  14.1485150 0.9975850
## Peter Lake:9-Peter Lake:5  10.4316786 -10.1207861  30.9841433 0.8273658
## Peter Lake:6-Paul Lake:6   17.8017712   6.7120688  28.8914737 0.0000401
## Paul Lake:7-Paul Lake:6    -0.8092288 -11.8989312  10.2804737 1.0000000
## Peter Lake:7-Paul Lake:6   23.8483529  12.6013419  35.0953640 0.0000000
## Paul Lake:8-Paul Lake:6    -1.1693399 -12.2590423   9.9203626 0.9999989
## Peter Lake:8-Paul Lake:6   15.9378824   4.9908457  26.8849190 0.0003006
## Paul Lake:9-Paul Lake:6     0.1798824 -16.5021309  16.8618956 1.0000000
## Peter Lake:9-Paul Lake:6   15.6631324  -2.5591082  33.8853729 0.1584032
## Paul Lake:7-Peter Lake:6  -18.6110000 -29.5411300  -7.6808700 0.0000101
## Peter Lake:7-Peter Lake:6   6.0465817  -5.0431207  17.1362841 0.7595330
## Paul Lake:8-Peter Lake:6  -18.9711111 -29.9012412  -8.0409811 0.0000062
## Peter Lake:8-Peter Lake:6  -1.8638889 -12.6492426   8.9214648 0.9999197
## Paul Lake:9-Peter Lake:6  -17.6218889 -34.1982518  -1.0455259 0.0276305
## Peter Lake:9-Peter Lake:6  -2.1386389 -20.2642090  15.9869312 0.9999970
## Peter Lake:7-Paul Lake:7   24.6575817  13.5678793  35.7472841 0.0000000
## Paul Lake:8-Paul Lake:7    -0.3601111 -11.2902412  10.5700189 1.0000000
## Peter Lake:8-Paul Lake:7   16.7471111   5.9617574  27.5324648 0.0000827
## Paul Lake:9-Paul Lake:7     0.9891111 -15.5872518  17.5654741 1.0000000
## Peter Lake:9-Paul Lake:7   16.4723611  -1.6532090  34.5979312 0.1087387
```

```
## Paul Lake:8-Peter Lake:7  -25.0176928 -36.1073952 -13.9279904 0.0000000
## Peter Lake:8-Peter Lake:7  -7.9104706 -18.8575073   3.0365661 0.3778093
## Paul Lake:9-Peter Lake:7  -23.6684706 -40.3504838  -6.9864574 0.0004851
## Peter Lake:9-Peter Lake:7  -8.1852206 -26.4074611  10.0370199 0.9089776
## Peter Lake:8-Paul Lake:8   17.1072222   6.3218685  27.8925759 0.0000523
## Paul Lake:9-Paul Lake:8     1.3492222 -15.2271407  17.9255852 0.9999999
## Peter Lake:9-Paul Lake:8   16.8324722  -1.2930979  34.9580424 0.0926020
## Paul Lake:9-Peter Lake:8  -15.7580000 -32.2392597   0.7232597 0.0735733
## Peter Lake:9-Peter Lake:8  -0.2747500 -18.3133864  17.7638864 1.0000000
## Peter Lake:9-Paul Lake:9   15.4832500  -6.5132124  37.4797124 0.4163366
```

```
lake.month.interaction <- with(surface.depths, interaction(lakename, month))
surface.depths.anova.2way3 <- aov(data = surface.depths, tp_ug ~
lake.month.interaction)

lake.month.groups <- HSD.test(surface.depths.anova.2way3,
"lake.month.interaction", group = TRUE)
lake.month.groups
```

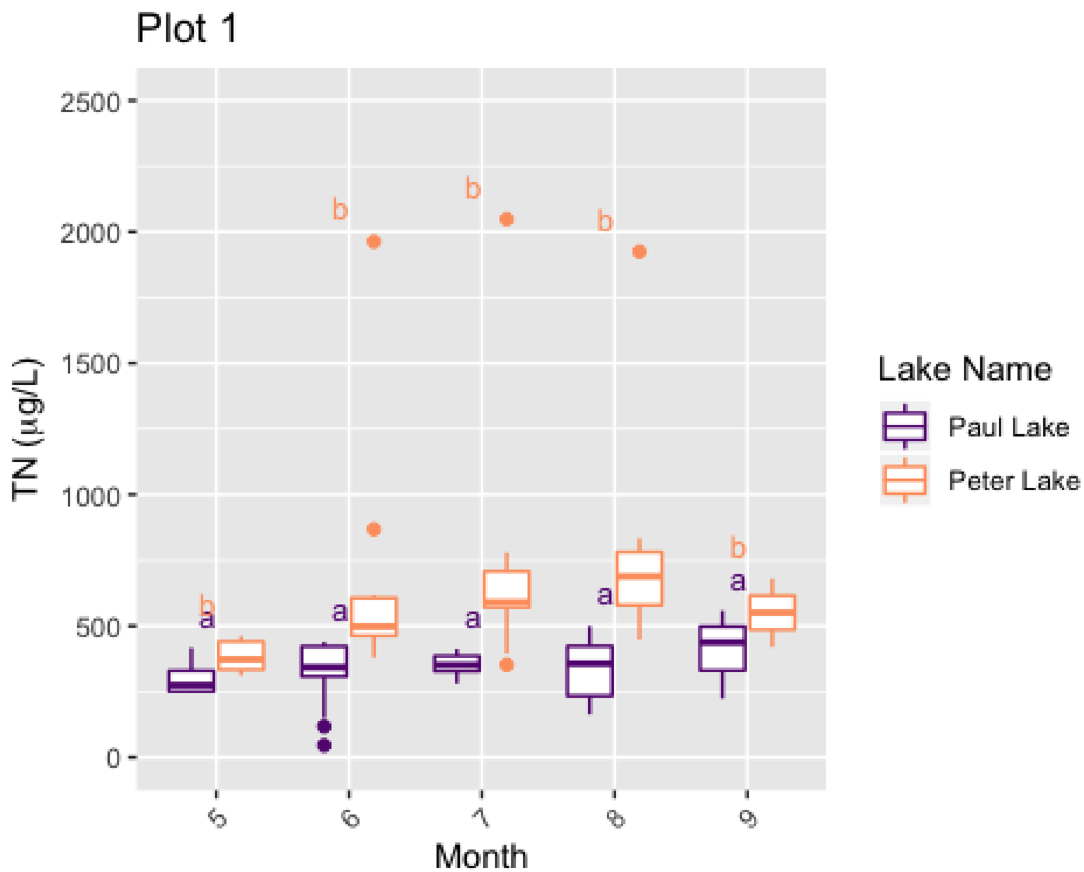
```
## $statistics
##    MSerror  Df     Mean      CV
##   103.4055 119 19.07347 53.3141
## 
## $parameters
##    test                 name.t ntr StudentizedRange alpha
##   Tukey lake.month.interaction  10         4.560262  0.05
## 
## $means
##                  tp_ug       std  r    Min    Max     Q25     Q50      Q75
## Paul Lake.5  11.474000  3.928545  6  7.001 17.090  8.1395 11.8885 13.53675
## Paul Lake.6  10.556118  4.416821 17  1.222 16.697  7.4430 10.6050 13.94600
## Paul Lake.7   9.746889  3.525120 18  4.501 21.763  7.8065  9.1555 10.65700
## Paul Lake.8   9.386778  1.478062 18  5.879 11.542  8.4495  9.6090 10.45050
## Paul Lake.9  10.736000  3.615978  5  6.592 16.281  8.9440 10.1920 11.67100
## Peter Lake.5 15.787571  2.719954  7 10.887 18.922 14.8915 15.5730 17.67400
## Peter Lake.6 28.357889 15.588507 18 10.974 53.388 14.7790 24.6840 41.13000
## Peter Lake.7 34.404471 18.285568 17 19.149 66.893 21.6640 24.2070 50.54900
## Peter Lake.8 26.494000  9.829596 19 14.551 49.757 21.2425 23.2250 27.99350
## Peter Lake.9 26.219250 10.814803  4 16.281 41.145 19.6845 23.7255 30.26025
## 
## $comparison
## NULL
## 
## $groups
##                  tp_ug groups
## Peter Lake.7 34.404471      a
## Peter Lake.6 28.357889     ab
## Peter Lake.8 26.494000    abc
## Peter Lake.9 26.219250   abcd
## Peter Lake.5 15.787571    bcd
```

```
## Paul Lake.5  11.474000     cd
## Paul Lake.9  10.736000     cd
## Paul Lake.6  10.556118      d
## Paul Lake.7   9.746889      d
## Paul Lake.8   9.386778      d
##
## attr(,"class")
## [1] "group"
```

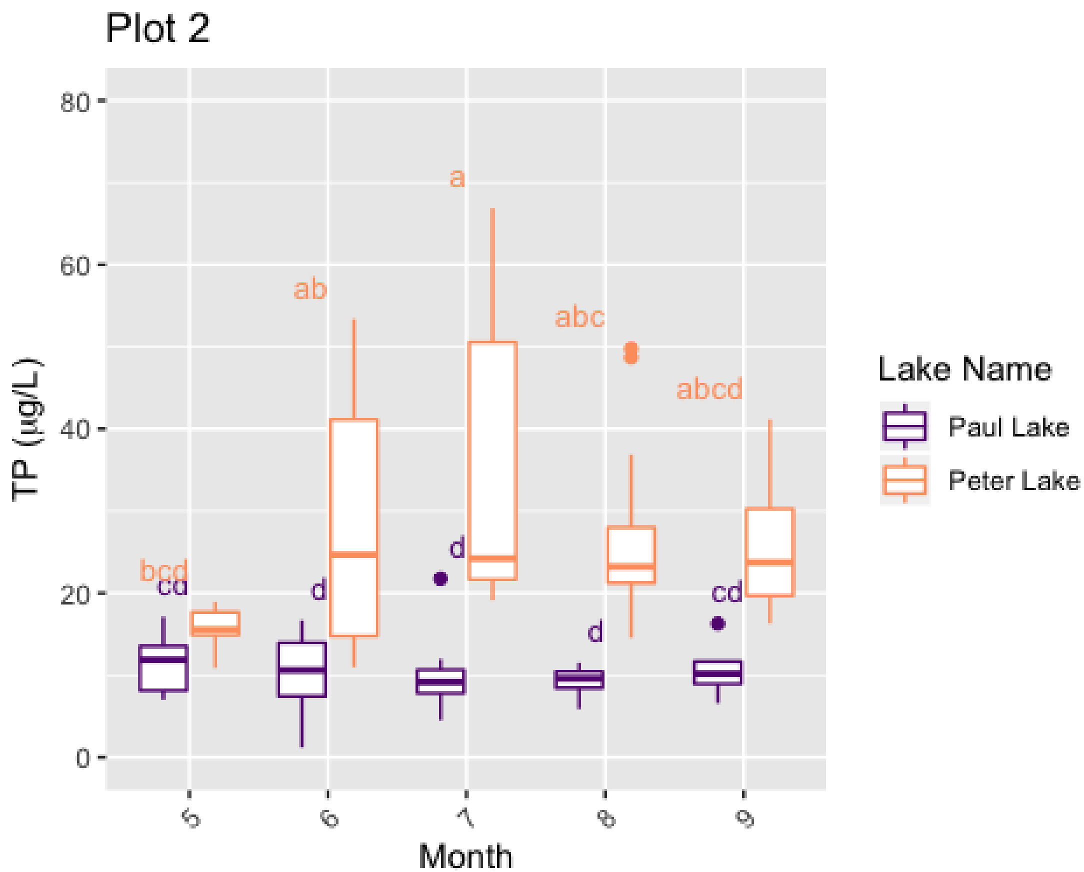
7. Create two plots, with TN (plot 1) or TP (plot 2) as the response variable and month and lake as the predictor variables. Hint: you may use some of the code you used for your visualization assignment. Assign groupings with letters, as determined from your tests. Adjust your axes, aesthetics, and color palettes in accordance with best data visualization practices.

8. Combine your plots with cowplot, with a common legend at the top and the two graphs stacked vertically. Your x axes should be formatted with the same breaks, such that you can remove the title and text of the top legend and retain just the bottom legend.

```
#7
plot1 <- ggplot(surface.depths, aes(x = as.factor(month), y = tn_ug, color  =
lakename)) +
  geom_boxplot() + # makes boxplots of nitrogen content along the months for
both the lakes
  theme(axis.text.x = element_text(angle = 45, hjust = 1), legend.position =
"right") + # Adjusts the theme
  ggtitle("Plot 1") + # Names the plot "Plot 1"
  ylim(0,2500) + # Setting the y-axis limit as 2500 so that we can view the
grouping letters without getting cut-off
  scale_color_viridis_d(option = "magma", begin = 0.3, end = 0.8) + # Adjusts
the color palette
  stat_summary(geom = "text", fun.y = max, vjust = -1, hjust = 1, size = 3.5,
# Places the letters right above the maximum points
               label = c("a", "b", "a", "b", "a", "b",
                         "a", "b", "a", "b"), show.legend = FALSE) + #
Assigns groupings with letters
  labs (x = "Month", y = expression(paste("TN (", mu, "g/L)")), color = "Lake
Name") # Labels the y-axis appropriately with its units.
print(plot1)

## Warning: Removed 23 rows containing non-finite values (stat_boxplot).

## Warning: Removed 23 rows containing non-finite values (stat_summary).
```

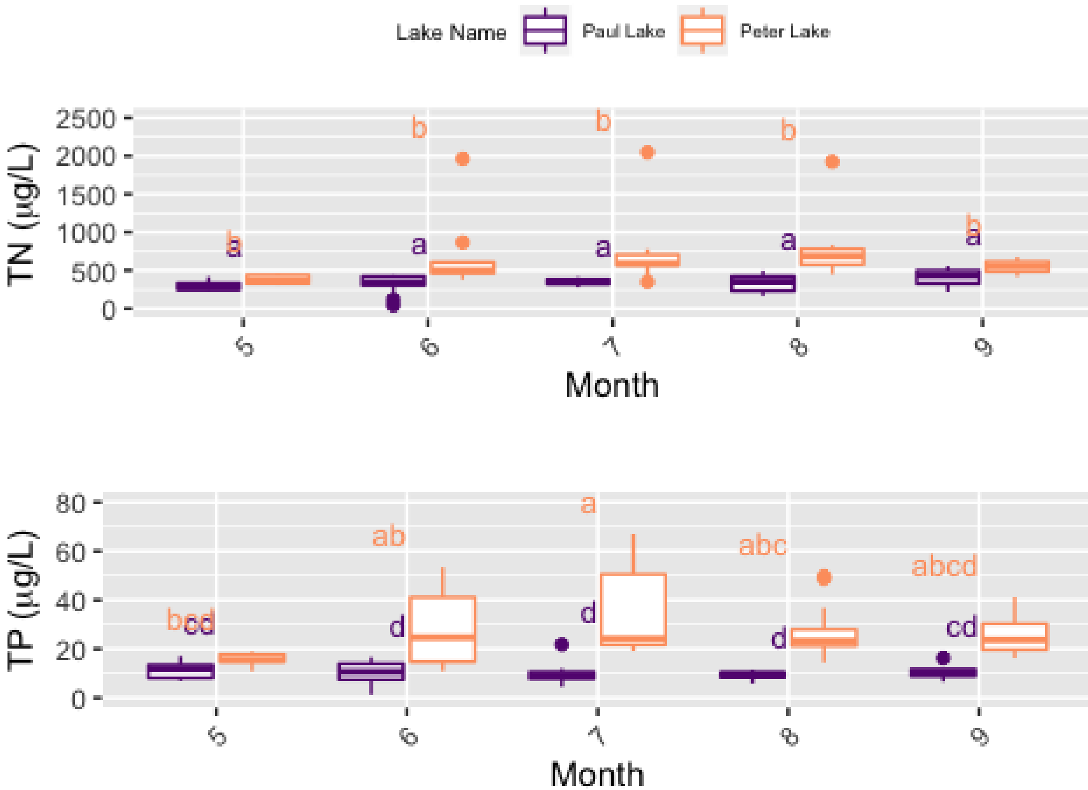



```
plot2 <- ggplot(surface.depths, aes(x = as.factor(month), y = tp_ug, color =
lakename)) +
  geom_boxplot() + # makes boxplots of phosphorus content along the months
for both the lakes
  theme(axis.text.x = element_text(angle = 45, hjust = 1), legend.position =
"right") + # Adjusts the theme
  ggtitle("Plot 2") + # To insert the main title of the plot
  ylim(0,80) + # Setting the y-axis limit as 80 so that we can view the
grouping letters without getting cut-off
  scale_color_viridis_d(option = "magma", begin = 0.3, end = 0.8) + # Adjusts
color palette
  stat_summary(geom = "text", fun.y = max, vjust = -1, hjust = 1, size = 3.5,
# Places the letters right above the maximum points
               label = c("cd", "bcd", "d", "ab", "d", "a",
                         "d", "abc", "cd", "abcd"), show.legend = FALSE) + #
Assigns groupings with letters, as determined from my test
  labs (x = "Month", y = expression(paste("TP (", mu, "g/L)")), color = "Lake
Name") # Labels the y-axis appropriately with its units.
print(plot2)

## Warning: Removed 1 rows containing non-finite values (stat_boxplot).

## Warning: Removed 1 rows containing non-finite values (stat_summary).
```




```
#8
library(cowplot)
plot1 = plot1 + theme(legend.position="top", legend.title = element_text(size
= 7),
  legend.text = element_text(size = 6)) + ggtitle("Plot 1 & Plot 2") # Takes
the legend position to the top, minimizes the size of the legend, and renames
the plot as "Plot 1 & Plot 2"
plot2 = plot2 + theme(legend.position = "none") + ggtitle("") # Removes the
legend and title from plot2
plot_grid(plot1, plot2,nrow = 2, align = 'h', rel_heights = c(1.3, 1)) #
Combines the two plots

## Warning: Removed 23 rows containing non-finite values (stat_boxplot).

## Warning: Removed 23 rows containing non-finite values (stat_summary).

## Warning: Removed 1 rows containing non-finite values (stat_boxplot).

## Warning: Removed 1 rows containing non-finite values (stat_summary).

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

Plot 1 & Plot 2



Lake Name
Paul Lake
Peter Lake
TN (µg/L)
2500
2000
1500
1000
500
0
Month
5
6
7
8
9
TP (µg/L)
80
60
40
20
0
Month