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

Claire Mullaney

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/clairemullaney/Desktop/ENV 872/Environmental_Data_Analytics_2020"
```

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

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

#2
#Looking at initial column names and date format
head(Peter_Paul_Nutrients)
```

```
##   lakeid  lakename year4 daynum month sampledate depth_id depth tn_ug
## 1      L Paul Lake  1991    140     5 1991-05-20        1  0.00   538
## 2      L Paul Lake  1991    140     5 1991-05-20        2  0.85   285
## 3      L Paul Lake  1991    140     5 1991-05-20        3  1.75   399
## 4      L Paul Lake  1991    140     5 1991-05-20        4  3.00   453
## 5      L Paul Lake  1991    140     5 1991-05-20        5  4.00   363
## 6      L Paul Lake  1991    140     5 1991-05-20        6  6.00   583
##   tp_ug nh34 no23 po4 comments
## 1    25   NA   NA  NA       NA
## 2    14   NA   NA  NA       NA
## 3    14   NA   NA  NA       NA
## 4    14   NA   NA  NA       NA
## 5    13   NA   NA  NA       NA
## 6    37   NA   NA  NA       NA
```

```r
class(Peter_Paul_Nutrients$sampledate)
```

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

```r
#Changing column sampledate to a date format
Peter_Paul_Nutrients$sampledate <- as.Date(Peter_Paul_Nutrients$sampledate, format = "%Y-%m-%d")

#Verifying change
class(Peter_Paul_Nutrients$sampledate)
```

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

```r
head(Peter_Paul_Nutrients)
```

```
##   lakeid  lakename year4 daynum month sampledate depth_id depth tn_ug
## 1      L Paul Lake  1991    140     5 1991-05-20        1  0.00   538
## 2      L Paul Lake  1991    140     5 1991-05-20        2  0.85   285
## 3      L Paul Lake  1991    140     5 1991-05-20        3  1.75   399
## 4      L Paul Lake  1991    140     5 1991-05-20        4  3.00   453
## 5      L Paul Lake  1991    140     5 1991-05-20        5  4.00   363
## 6      L Paul Lake  1991    140     5 1991-05-20        6  6.00   583
##   tp_ug nh34 no23 po4 comments
## 1    25   NA   NA  NA       NA
## 2    14   NA   NA  NA       NA
## 3    14   NA   NA  NA       NA
## 4    14   NA   NA  NA       NA
## 5    13   NA   NA  NA       NA
## 6    37   NA   NA  NA       NA
```

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.

```r
Peter_Paul_Nutrients_wr <-
  filter(Peter_Paul_Nutrients,
         depth == 0 &
           year4 >= 1993 &
           year4 <= 1996)

Peter_Paul_Nutrients_wr$month <- as.factor(Peter_Paul_Nutrients_wr$month)
class(Peter_Paul_Nutrients_wr$month)
```

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

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 a two-way ANOVA with interaction effects. We have two continuous response variables (TN and TP), each of which is being examined separately as the effect of two categorical explanatory variables, lake name and month; when two categorical variables are used as explanatory variables for a continuous response variable, a two-way ANOVA must be used in the analysis. I will consider interaction effects to account for the possibility that one of the explanatory variables depends on or affects the other. For example, if one of the lakes was in the sun and one was in the shade, levels of nitrogen or phosphorus could potentially be partially predicted by the interaction between the lake and season variables.

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 Test for tn
##As linear model
tn.anova.2way.int <- lm(data = Peter_Paul_Nutrients_wr,
                        tn_ug ~ lakename * month)
summary(tn.anova.2way.int)
```

```
##
## Call:
## lm(formula = tn_ug ~ lakename * month, data = Peter_Paul_Nutrients_wr)
##
## 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
```

```
##As aov
tn.anova.2way.aov <- aov(data = Peter_Paul_Nutrients_wr,
                         tn_ug ~ lakename * month)
summary(tn.anova.2way.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
```

```r
##TukeyHSD post-hoc test for pairwise differences to confirm ANOVA results
TukeyHSD(tn.anova.2way.aov)
```

```
##   Tukey multiple comparisons of means
##     95% family-wise confidence level
##
## Fit: aov(formula = tn_ug ~ lakename * month, data = Peter_Paul_Nutrients_wr)
##
## $lakename
##                            diff      lwr      upr p adj
## Peter Lake-Paul Lake 303.796 203.8773 403.7146     0
##
## $month
##          diff       lwr      upr     p adj
## 6-5 132.58168 -104.4173 369.5807 0.5296645
## 7-5 196.50011  -47.8276 440.8278 0.1755245
## 8-5 208.77984  -32.7942 450.3539 0.1234174
## 9-5 160.08048 -220.7887 540.9497 0.7692917
## 7-6  63.91843 -123.8978 251.7346 0.8780820
## 8-6  76.19815 -108.0216 260.4179 0.7795574
## 9-6  27.49879 -319.8343 374.8318 0.9994702
## 8-7  12.27972 -181.2775 205.8370 0.9997797
## 9-7 -36.41964 -388.7941 315.9548 0.9984863
## 9-8 -48.69936 -399.1701 301.7714 0.9952106
##
## $`lakename:month`
##                                 diff         lwr       upr     p adj
## Peter Lake:5-Paul Lake:5    84.42736 -384.695091 553.54981 0.9998802
## Paul Lake:6-Paul Lake:5     23.61297 -376.795278 424.02122 1.0000000
## Peter Lake:6-Paul Lake:5   308.53119  -95.128061 712.19044 0.2949521
## Paul Lake:7-Paul Lake:5     53.12257 -358.325034 464.57018 0.9999929
## Peter Lake:7-Paul Lake:5   409.37327   -6.794730 825.54127 0.0577843
## Paul Lake:8-Paul Lake:5     35.99664 -375.450962 447.44425 0.9999998
## Peter Lake:8-Paul Lake:5   445.47177   38.159418 852.78411 0.0206524
## Paul Lake:9-Paul Lake:5    105.82450 -490.419726 702.06873 0.9998933
## Peter Lake:9-Paul Lake:5   249.95650 -438.527028 938.44003 0.9743614
## Paul Lake:6-Peter Lake:5   -60.81439 -439.493476 317.86470 0.9999541
## Peter Lake:6-Peter Lake:5  224.10383 -158.011173 606.21883 0.6694487
## Paul Lake:7-Peter Lake:5   -31.30479 -421.638257 359.02869 0.9999999
## Peter Lake:7-Peter Lake:5  324.94591  -70.360160 720.25198 0.2042224
## Paul Lake:8-Peter Lake:5   -48.43071 -438.764185 341.90276 0.9999950
## Peter Lake:8-Peter Lake:5  361.04441  -24.927657 747.01648 0.0870846
## Paul Lake:9-Peter Lake:5    21.39714 -560.477640 603.27193 1.0000000
## Peter Lake:9-Peter Lake:5  165.52914 -510.548261 841.60655 0.9985431
## Peter Lake:6-Paul Lake:6   284.91822   -8.787028 578.62346 0.0650344
## Paul Lake:7-Paul Lake:6     29.50960 -274.811140 333.83034 0.9999994
## Peter Lake:7-Paul Lake:6   385.76030   75.087182 696.43342 0.0043241
## Paul Lake:8-Paul Lake:6     12.38367 -291.937068 316.70441 1.0000000
## Peter Lake:8-Paul Lake:6   421.85880  123.152702 720.56489 0.0005774
## Paul Lake:9-Paul Lake:6     82.21153 -445.831232 610.25429 0.9999647
## Peter Lake:9-Paul Lake:6   226.34353 -403.998878 856.68594 0.9761624
## Paul Lake:7-Peter Lake:6  -255.40862 -563.994320  53.17709 0.1964898
## Peter Lake:7-Peter Lake:6  100.84208 -214.009961 415.69412 0.9891274
```

```
## Paul Lake:8-Peter Lake:6  -272.53454 -581.120248  36.05116 0.1316086
## Peter Lake:8-Peter Lake:6  136.94058 -166.109506 439.99066 0.9029804
## Paul Lake:9-Peter Lake:6  -202.70669 -733.218875 327.80550 0.9642843
## Peter Lake:9-Peter Lake:6  -58.57469 -690.987190 573.83782 0.9999996
## Peter Lake:7-Paul Lake:7   356.25070   31.473618 681.02778 0.0200027
## Paul Lake:8-Paul Lake:7    -17.12593 -335.831873 301.58002 1.0000000
## Peter Lake:8-Paul Lake:7   392.34920   79.000035 705.69836 0.0038467
## Paul Lake:9-Paul Lake:7     52.70193 -483.760115 589.16397 0.9999994
## Peter Lake:9-Paul Lake:7   196.83393 -440.577960 834.24582 0.9916222
## Paul Lake:8-Peter Lake:7  -373.37663 -698.153706 -48.59955 0.0116944
## Peter Lake:8-Peter Lake:7   36.09850 -283.423597 355.62059 0.9999978
## Paul Lake:9-Peter Lake:7  -303.54877 -843.639684 236.54215 0.7209271
## Peter Lake:9-Peter Lake:7 -159.41677 -799.885807 481.05227 0.9983429
## Peter Lake:8-Paul Lake:8   409.47512   96.125963 722.82428 0.0020552
## Paul Lake:9-Paul Lake:8     69.82786 -466.634186 606.28990 0.9999924
## Peter Lake:9-Paul Lake:8   213.95986 -423.452032 851.37175 0.9849047
## Paul Lake:9-Peter Lake:8  -339.64727 -872.944314 193.64978 0.5579223
## Peter Lake:9-Peter Lake:8 -195.51527 -830.265716 439.23518 0.9917740
## Peter Lake:9-Paul Lake:9   144.13200 -625.615985 913.87999 0.9998333
```

```r
###No significant interactions; only lakename has significant differences
##No need to form groups

#6 Test for tp
##As linear model
tp.anova.2way.int <- lm(data = Peter_Paul_Nutrients_wr,
                        tp_ug ~ lakename * month)
summary(tp.anova.2way.int)
```

```
##
## Call:
## lm(formula = tp_ug ~ lakename * month, data = Peter_Paul_Nutrients_wr)
##
## 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
```

```r
##As aov
tp.anova.2way.aov <- aov(data = Peter_Paul_Nutrients_wr,
                         tp_ug ~ lakename * month)
summary(tp.anova.2way.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
```

```r
##TukeyHSD post-hoc test for pairwise differences to confirm ANOVA results
TukeyHSD(tp.anova.2way.aov)
```

```
##   Tukey multiple comparisons of means
##     95% family-wise confidence level
##
## Fit: aov(formula = tp_ug ~ lakename * month, data = Peter_Paul_Nutrients_wr)
##
## $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
```

```r
##Some of the interactions, along with lakenames, are significantly different from
##each other; creating groups to more easily interpret statistical differences

##Creating an interaction as a separate variable and formatting as an aov
##to group interaction data
tp.interaction <- with(Peter_Paul_Nutrients_wr,
                       interaction(lakename, month))

tp.anova.2way.aov2 <- aov(data = Peter_Paul_Nutrients_wr,
                          tp_ug ~ tp.interaction)
summary(tp.anova.2way.aov2)
```

```
##                 Df Sum Sq Mean Sq F value   Pr(>F)
## tp.interaction   9  12055  1339.5   12.95 3.24e-14 ***
## Residuals      119  12305   103.4
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
## 1 observation deleted due to missingness
```

```r
##Groups
tp.groups <- HSD.test(tp.anova.2way.aov2, trt = "tp.interaction",
                      group = TRUE)
```

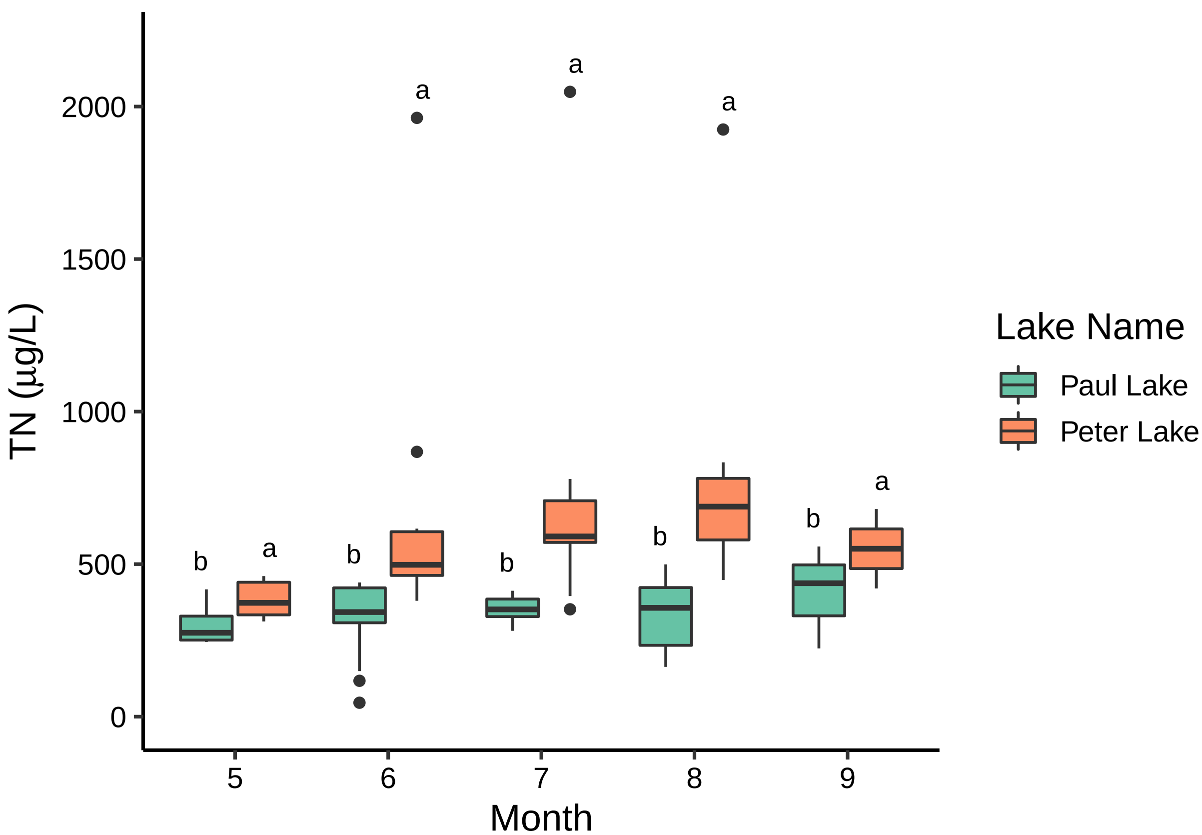
```r
tp.groups
```

```
## $statistics
##    MSerror  Df     Mean      CV
##   103.4055 119 19.07347 53.3141
##
## $parameters
##    test         name.t ntr StudentizedRange alpha
##   Tukey tp.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.

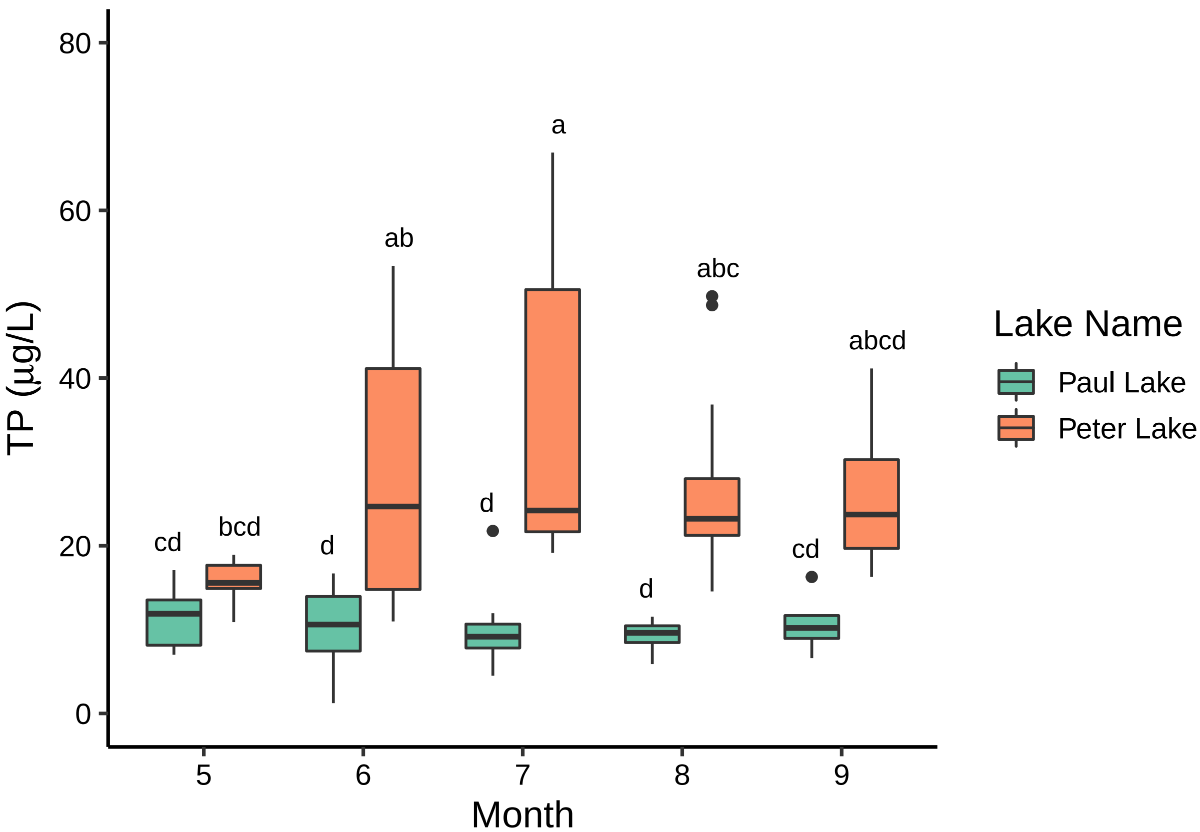
```r
#7
#Defining a new theme
theme_def <- theme_classic(base_size = 14) +
  theme(axis.text = element_text(color = "black"),
```

```
        legend.position = "right")

#Setting new theme as the default theme
theme_set(theme_def)

##Plot 1 (tn)
##Groups were not created for TN, but within each month, the lake with the larger
##TN value is labelled "a" and the lake with the smaller TN value is labelled "b"
##(because the lakename significantly predicted TN concentrations)
tn.anova.plot <- ggplot(Peter_Paul_Nutrients_wr, aes(y = tn_ug, x = month,
                                                     fill = lakename)) +
  geom_boxplot() +
  labs(y = expression(paste("TN (", mu, "g/L)")),
       x = "Month", fill = "Lake Name") +
  scale_fill_manual(values = c("#66c2a5", "#fc8d62")) +
  ylim(0, 2200) +
    stat_summary(geom = "text", position=position_dodge(width=0.9),
                 fun.y = max, vjust = -1,
                 size = 3.5,
                 label = c("a", "b", "a", "b", "a",
                           "b", "a", "b", "a", "b"))

print(tn.anova.plot)
```




```
##Plot 2 (tp)
tp.anova.plot <- ggplot(Peter_Paul_Nutrients_wr,
                        aes(y = tp_ug, x = month,
```

```r
                             fill = lakename)) +
  geom_boxplot() +
  labs(y = expression(paste("TP (", mu, "g/L)")),
       x = "Month", fill = "Lake Name") +
  scale_fill_manual(values = c("#66c2a5", "#fc8d62")) +
  ylim(0, 80) +
  stat_summary(geom = "text", position=position_dodge(width=0.9),
               fun.y = max, vjust = -1,
               size = 3.5,
               label = c("bcd", "cd", "ab", "d", "a",
                         "d", "abc", "d", "abcd", "cd"))

print(tp.anova.plot)
```




```r
#8
##Cowplot
tp_tn_plot <- plot_grid(tn.anova.plot +
                          theme(legend.position = "none",
                                axis.title.x = element_blank(),
                                axis.text.x=element_blank(),
                                axis.ticks.x=element_blank(),
                                plot.margin = margin(t = 0, r = 0,
                                                     b = 0, l = 0,
                                                     unit = "cm")),
                        tp.anova.plot +
                          theme(legend.position="none",
                                plot.margin = margin(t = 0, r = 0,
```

```
                                              b = 0, l = 0,
                                              unit = "cm")),
                        nrow = 2, align = "hv")

legend <- get_legend(tn.anova.plot +
                       guides(color = guide_legend(nrow = 1)) +
                       theme(legend.position = "top"))


plot_grid(legend, tp_tn_plot, ncol = 1, rel_heights = c(.1, 1))
```