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

Kathleen Mason

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/kathleenmason/Documents/DUKE/Data Analytics/Environmental_Data_Analytics_2020"
```

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

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

#2
Nutrients$collectDate <- as.Date(Nutrients$sampledate , format = "%Y-%m-%d")
class(Nutrients$sampledate)
```

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

```
head(Nutrients$sampledate)
```

```
## [1] 1991-05-20 1991-05-20 1991-05-20 1991-05-20 1991-05-20 1991-05-20
## 490 Levels: 1991-05-20 1991-05-27 1991-05-28 1991-06-03 ... 2016-08-16
```

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.

```
Nutrients.Wrangled<- Nutrients%>%
  filter(year4==1993:1996)%>%
  filter(depth==0)
```

```
## Warning in year4 == 1993:1996: longer object length is not a multiple of shorter
## object length
```

```
Nutrients.Wrangled$month <- as.factor(Nutrients.Wrangled$month)
class(Nutrients.Wrangled$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: 2 two-way Anovas with interaction effects We are testing two categorical variable (month and lake name) as predictors for a continuous variable (TP, TN). Lake name and month might interact if one lake is effected more by sesonality than another.

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
N.W.test <- aov(data = Nutrients.Wrangled, tn_ug ~ lakename * month)
summary(N.W.test)
```

```
##                Df Sum Sq Mean Sq F value Pr(>F)
## lakename        1 106141  106141   5.276 0.0355 *
## month           4  57671   14418   0.717 0.5927
## lakename:month  3  15334    5111   0.254 0.8573
## Residuals      16 321872   20117
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
## 7 observations deleted due to missingness
```

```
N.W.test2 <- lm(data = Nutrients.Wrangled, tn_ug ~ lakename * month)
summary(N.W.test2)
```

```
##
## Call:
## lm(formula = tn_ug ~ lakename * month, data = Nutrients.Wrangled)
##
## Residuals:
##     Min      1Q  Median      3Q     Max
## -204.84  -89.37    0.00   45.60  336.13
##
## Coefficients: (1 not defined because of singularities)
##                            Estimate Std. Error t value Pr(>|t|)
## (Intercept)                  417.35     141.83   2.942  0.00956 **
## lakenamePeter Lake           -12.64     173.71  -0.073  0.94290
```

```
## month6                         12.71     173.71   0.073  0.94260
## month7                        -17.21     173.71  -0.099  0.92233
## month8                         36.31     173.71   0.209  0.83707
## month9                        -86.75     173.71  -0.499  0.62431
## lakenamePeter Lake:month6     114.56     208.77   0.549  0.59078
## lakenamePeter Lake:month7     169.34     212.75   0.796  0.43771
## lakenamePeter Lake:month8     165.37     212.75   0.777  0.44833
## lakenamePeter Lake:month9         NA         NA      NA       NA
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Residual standard error: 141.8 on 16 degrees of freedom
##   (7 observations deleted due to missingness)
## Multiple R-squared:  0.3576, Adjusted R-squared:  0.03635
## F-statistic: 1.113 on 8 and 16 DF,  p-value: 0.4047
```

```
TukeyHSD(N.W.test)
```

```
##   Tukey multiple comparisons of means
##     95% family-wise confidence level
##
## Fit: aov(formula = tn_ug ~ lakename * month, data = Nutrients.Wrangled)
##
## $lakename
##                           diff      lwr      upr    p adj
## Peter Lake-Paul Lake 135.7473 10.46586 261.0288 0.035454
##
## $month
##            diff       lwr      upr     p adj
## 6-5   86.258431 -207.9227 380.4395 0.8933665
## 7-5   95.688167 -211.5740 402.9503 0.8712982
## 8-5  146.555167 -160.7070 453.8173 0.5998934
## 9-5   12.176556 -384.4972 408.8503 0.9999802
## 7-6    9.429736 -225.2456 244.1051 0.9999420
## 8-6   60.296736 -174.3786 294.9721 0.9306602
## 9-6  -74.081875 -417.6114 269.4476 0.9620319
## 8-7   50.867000 -200.0115 301.7455 0.9694965
## 9-7  -83.511611 -438.3074 271.2841 0.9484941
## 9-8 -134.378611 -489.1744 220.4171 0.7728103
##
## $`lakename:month`
##                                 diff       lwr      upr     p adj
## Peter Lake:5-Paul Lake:5   -12.63800 -645.1982 619.9222 1.0000000
## Paul Lake:6-Paul Lake:5     12.70600 -619.8542 645.2662 1.0000000
## Peter Lake:6-Paul Lake:5   114.62517 -443.2405 672.4909 0.9984984
## Paul Lake:7-Paul Lake:5    -17.20650 -649.7667 615.3537 1.0000000
## Peter Lake:7-Paul Lake:5   139.49750 -437.9483 716.9433 0.9950423
## Paul Lake:8-Paul Lake:5     36.31000 -596.2502 668.8702 1.0000000
## Peter Lake:8-Paul Lake:5   189.03975 -388.4061 766.4856 0.9633334
## Paul Lake:9-Paul Lake:5    -86.74700 -719.3072 545.8132 0.9999411
## Peter Lake:9-Paul Lake:5         NA        NA       NA        NA
## Paul Lake:6-Peter Lake:5    25.34400 -491.1393 541.8273 1.0000000
## Peter Lake:6-Peter Lake:5  127.26317 -294.4437 548.9700 0.9777501
## Paul Lake:7-Peter Lake:5    -4.56850 -521.0518 511.9148 1.0000000
## Peter Lake:7-Peter Lake:5  152.13550 -295.1521 599.4231 0.9541212
```

```
## Paul Lake:8-Peter Lake:5    48.94800 -467.5353 565.4313 0.9999975
## Peter Lake:8-Peter Lake:5  201.67775 -245.6099 648.9654 0.8103960
## Paul Lake:9-Peter Lake:5   -74.10900 -590.5923 442.3743 0.9999142
## Peter Lake:9-Peter Lake:5         NA        NA       NA        NA
## Peter Lake:6-Paul Lake:6   101.91917 -319.7877 523.6260 0.9950268
## Paul Lake:7-Paul Lake:6    -29.91250 -546.3958 486.5708 1.0000000
## Peter Lake:7-Paul Lake:6   126.79150 -320.4961 574.0791 0.9851373
## Paul Lake:8-Paul Lake:6     23.60400 -492.8793 540.0873 1.0000000
## Peter Lake:8-Paul Lake:6   176.33375 -270.9539 623.6214 0.8982052
## Paul Lake:9-Paul Lake:6    -99.45300 -615.9363 417.0303 0.9990877
## Peter Lake:9-Paul Lake:6          NA        NA       NA        NA
## Paul Lake:7-Peter Lake:6  -131.83167 -553.5385 289.8752 0.9722734
## Peter Lake:7-Peter Lake:6   24.87233 -308.5162 358.2608 0.9999997
## Paul Lake:8-Peter Lake:6   -78.31517 -500.0220 343.3917 0.9993120
## Peter Lake:8-Peter Lake:6   74.41458 -258.9739 407.8031 0.9972083
## Paul Lake:9-Peter Lake:6  -201.37217 -623.0790 220.3347 0.7604932
## Peter Lake:9-Peter Lake:6         NA        NA       NA        NA
## Peter Lake:7-Paul Lake:7   156.70400 -290.5836 603.9916 0.9456930
## Paul Lake:8-Paul Lake:7     53.51650 -462.9668 569.9998 0.9999945
## Peter Lake:8-Paul Lake:7   206.24625 -241.0414 653.5339 0.7918011
## Paul Lake:9-Paul Lake:7    -69.54050 -586.0238 446.9428 0.9999494
## Peter Lake:9-Paul Lake:7          NA        NA       NA        NA
## Paul Lake:8-Peter Lake:7  -103.18750 -550.4751 344.1001 0.9964442
## Peter Lake:8-Peter Lake:7   49.54225 -315.6666 414.7511 0.9999462
## Paul Lake:9-Peter Lake:7  -226.24450 -673.5321 221.0431 0.7031933
## Peter Lake:9-Peter Lake:7         NA        NA       NA        NA
## Peter Lake:8-Paul Lake:8   152.72975 -294.5579 600.0174 0.9530788
## Paul Lake:9-Paul Lake:8   -123.05700 -639.5403 393.4263 0.9955097
## Peter Lake:9-Paul Lake:8          NA        NA       NA        NA
## Paul Lake:9-Peter Lake:8  -275.78675 -723.0744 171.5009 0.4676978
## Peter Lake:9-Peter Lake:8         NA        NA       NA        NA
## Peter Lake:9-Paul Lake:9          NA        NA       NA        NA
```

```
Nutrient.interaction <- with(Nutrients.Wrangled,interaction(lakename, month))
Nutrient.anova.2way5 <- aov(data = Nutrients.Wrangled, tn_ug ~ Nutrient.interaction)

Nutrient.groups <- HSD.test(Nutrient.anova.2way5, "Nutrient.interaction", group = TRUE)
Nutrient.groups
```

```
## $statistics
##   MSerror Df    Mean       CV
##     20117 16 492.015 28.82726
##
## $parameters
##    test               name.t ntr StudentizedRange alpha
##   Tukey Nutrient.interaction   9         5.031007  0.05
##
## $means
##                 tn_ug       std r     Min     Max      Q25      Q50      Q75
## Paul Lake.5  417.3450        NA 1 417.345 417.345 417.3450 417.3450 417.3450
## Paul Lake.6  430.0510  10.06496 2 422.934 437.168 426.4925 430.0510 433.6095
## Paul Lake.7  400.1385  17.72080 2 387.608 412.669 393.8732 400.1385 406.4038
## Paul Lake.8  453.6550  64.48248 2 408.059 499.251 430.8570 453.6550 476.4530
## Paul Lake.9  330.5980 151.03659 2 223.799 437.397 277.1985 330.5980 383.9975
## Peter Lake.5 404.7070  44.72592 2 373.081 436.333 388.8940 404.7070 420.5200
```

```
## Peter Lake.6 531.9702 184.16157 6 388.163 868.105 415.4948 452.3175 583.8525
## Peter Lake.7 556.8425 148.65011 4 352.001 707.771 520.6047 583.7990 620.0367
## Peter Lake.8 606.3847 137.37816 4 465.333 754.679 504.0952 602.7635 705.0530
##
## $comparison
## NULL
##
## $groups
##                 tn_ug groups
## Peter Lake.8 606.3847      a
## Peter Lake.7 556.8425      a
## Peter Lake.6 531.9702      a
## Paul Lake.8  453.6550      a
## Paul Lake.6  430.0510      a
## Paul Lake.5  417.3450      a
## Peter Lake.5 404.7070      a
## Paul Lake.7  400.1385      a
## Paul Lake.9  330.5980      a
##
## attr(,"class")
## [1] "group"
```

```
#6
N.W.test.p <- aov(data = Nutrients.Wrangled, tp_ug ~ lakename * month)
summary(N.W.test.p)
```

```
##                Df Sum Sq Mean Sq F value  Pr(>F)
## lakename        1   2099  2099.5  10.581 0.00351 **
## month           4    649   162.2   0.817 0.52735
## lakename:month  3    197    65.7   0.331 0.80278
## Residuals      23   4564   198.4
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
```

```
N.W.test.p.lm <- lm(data = Nutrients.Wrangled, tp_ug ~ lakename * month)
summary(N.W.test2)
```

```
##
## Call:
## lm(formula = tn_ug ~ lakename * month, data = Nutrients.Wrangled)
##
## Residuals:
##     Min      1Q  Median      3Q     Max
## -204.84  -89.37    0.00   45.60  336.13
##
## Coefficients: (1 not defined because of singularities)
##                             Estimate Std. Error t value Pr(>|t|)
## (Intercept)                   417.35     141.83   2.942  0.00956 **
## lakenamePeter Lake            -12.64     173.71  -0.073  0.94290
## month6                         12.71     173.71   0.073  0.94260
## month7                        -17.21     173.71  -0.099  0.92233
## month8                         36.31     173.71   0.209  0.83707
## month9                        -86.75     173.71  -0.499  0.62431
## lakenamePeter Lake:month6     114.56     208.77   0.549  0.59078
## lakenamePeter Lake:month7     169.34     212.75   0.796  0.43771
```

```
## lakenamePeter Lake:month8   165.37     212.75   0.777  0.44833
## lakenamePeter Lake:month9       NA         NA      NA       NA
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Residual standard error: 141.8 on 16 degrees of freedom
##   (7 observations deleted due to missingness)
## Multiple R-squared:  0.3576, Adjusted R-squared:  0.03635
## F-statistic: 1.113 on 8 and 16 DF,  p-value: 0.4047
```

```
TukeyHSD(N.W.test.p)
```

```
##   Tukey multiple comparisons of means
##     95% family-wise confidence level
##
## Fit: aov(formula = tp_ug ~ lakename * month, data = Nutrients.Wrangled)
##
## $lakename
##                           diff      lwr      upr     p adj
## Peter Lake-Paul Lake 18.01535 6.558264 29.47244 0.0035062
##
## $month
##          diff       lwr       upr     p adj
## 6-5  7.020072 -20.73949 34.779638 0.9427514
## 7-5 13.537183 -14.22238 41.296749 0.6084546
## 8-5  3.657961 -24.10161 31.417527 0.9947785
## 9-5  4.759735 -33.25162 42.771086 0.9957085
## 7-6  6.517111 -13.11187 26.146088 0.8608344
## 8-6 -3.362111 -22.99109 16.266866 0.9859033
## 9-6 -2.260337 -34.81131 30.290640 0.9995728
## 8-7 -9.879222 -29.50820  9.749755 0.5802043
## 9-7 -8.777448 -41.32842 23.773529 0.9288093
## 9-8  1.101774 -31.44920 33.652751 0.9999755
##
## $`lakename:month`
##                                    diff       lwr      upr     p adj
## Peter Lake:5-Paul Lake:5    1.638000 -58.56751 61.84351 1.0000000
## Paul Lake:6-Paul Lake:5    -3.439000 -63.64451 56.76651 1.0000000
## Peter Lake:6-Paul Lake:5   13.986000 -38.56568 66.53768 0.9936223
## Paul Lake:7-Paul Lake:5    -0.472000 -60.67751 59.73351 1.0000000
## Peter Lake:7-Paul Lake:5   21.517429 -31.03425 74.06910 0.9054101
## Paul Lake:8-Paul Lake:5    -7.076500 -67.28201 53.12901 0.9999909
## Peter Lake:8-Paul Lake:5   10.702571 -41.84910 63.25425 0.9991437
## Paul Lake:9-Paul Lake:5    -6.158500 -66.36401 54.04701 0.9999973
## Peter Lake:9-Paul Lake:5          NA        NA       NA        NA
## Paul Lake:6-Peter Lake:5   -5.077000 -54.23459 44.08059 0.9999970
## Peter Lake:6-Peter Lake:5  12.348000 -27.06576 51.76176 0.9805580
## Paul Lake:7-Peter Lake:5   -2.110000 -51.26759 47.04759 1.0000000
## Peter Lake:7-Peter Lake:5  19.879429 -19.53433 59.29319 0.7516639
## Paul Lake:8-Peter Lake:5   -8.714500 -57.87209 40.44309 0.9997168
## Peter Lake:8-Peter Lake:5   9.064571 -30.34919 48.47833 0.9978170
## Paul Lake:9-Peter Lake:5   -7.796500 -56.95409 41.36109 0.9998862
## Peter Lake:9-Peter Lake:5         NA        NA       NA        NA
## Peter Lake:6-Paul Lake:6   17.425000 -21.98876 56.83876 0.8609745
## Paul Lake:7-Paul Lake:6     2.967000 -46.19059 52.12459 1.0000000
```

```
## Peter Lake:7-Paul Lake:6   24.956429 -14.45733 64.37019 0.4779259
## Paul Lake:8-Paul Lake:6    -3.637500 -52.79509 45.52009 0.9999998
## Peter Lake:8-Paul Lake:6   14.141571 -25.27219 53.55533 0.9547531
## Paul Lake:9-Paul Lake:6    -2.719500 -51.87709 46.43809 1.0000000
## Peter Lake:9-Paul Lake:6          NA        NA       NA        NA
## Paul Lake:7-Peter Lake:6  -14.458000 -53.87176 24.95576 0.9484839
## Peter Lake:7-Peter Lake:6   7.531429 -18.74441 33.80727 0.9892840
## Paul Lake:8-Peter Lake:6  -21.062500 -60.47626 18.35126 0.6904416
## Peter Lake:8-Peter Lake:6  -3.283429 -29.55927 22.99241 0.9999847
## Paul Lake:9-Peter Lake:6  -20.144500 -59.55826 19.26926 0.7383124
## Peter Lake:9-Peter Lake:6         NA        NA       NA        NA
## Peter Lake:7-Paul Lake:7   21.989429 -17.42433 61.40319 0.6402523
## Paul Lake:8-Paul Lake:7    -6.604500 -55.76209 42.55309 0.9999716
## Peter Lake:8-Paul Lake:7   11.174571 -28.23919 50.58833 0.9900595
## Paul Lake:9-Paul Lake:7    -5.686500 -54.84409 43.47109 0.9999921
## Peter Lake:9-Paul Lake:7          NA        NA       NA        NA
## Paul Lake:8-Peter Lake:7  -28.593929 -68.00769 10.81983 0.3033257
## Peter Lake:8-Peter Lake:7 -10.814857 -37.09070 15.46098 0.9028046
## Paul Lake:9-Peter Lake:7  -27.675929 -67.08969 11.73783 0.3433151
## Peter Lake:9-Peter Lake:7         NA        NA       NA        NA
## Peter Lake:8-Paul Lake:8   17.779071 -21.63469 57.19283 0.8470660
## Paul Lake:9-Paul Lake:8     0.918000 -48.23959 50.07559 1.0000000
## Peter Lake:9-Paul Lake:8          NA        NA       NA        NA
## Paul Lake:9-Peter Lake:8  -16.861071 -56.27483 22.55269 0.8816516
## Peter Lake:9-Peter Lake:8         NA        NA       NA        NA
## Peter Lake:9-Paul Lake:9          NA        NA       NA        NA
```

```
Nutrient.interaction.p <- with(Nutrients.Wrangled,interaction(lakename, month))
Nutrient.anova.2way5.p <- aov(data = Nutrients.Wrangled, tp_ug ~ Nutrient.interaction.p)

Nutrient.groups.p <- HSD.test(Nutrient.anova.2way5.p, "Nutrient.interaction.p", group = TRUE)
Nutrient.groups.p
```

```
## $statistics
##    MSerror Df     Mean       CV
##   198.4229 23 26.22831 53.70634
##
## $parameters
##    test                 name.t ntr StudentizedRange alpha
##   Tukey Nutrient.interaction.p   9         4.825996  0.05
##
## $means
##                 tp_ug        std r    Min    Max      Q25     Q50      Q75
## Paul Lake.5  17.09000         NA 1 17.090 17.090 17.09000 17.0900 17.09000
## Paul Lake.6  13.65100  4.3076945 2 10.605 16.697 12.12800 13.6510 15.17400
## Paul Lake.7  16.61800  7.2761288 2 11.473 21.763 14.04550 16.6180 19.19050
## Paul Lake.8  10.01350  0.7092281 2  9.512 10.515  9.76275 10.0135 10.26425
## Paul Lake.9  10.93150  1.0458109 2 10.192 11.671 10.56175 10.9315 11.30125
## Peter Lake.5 18.72800  0.2743574 2 18.534 18.922 18.63100 18.7280 18.82500
## Peter Lake.6 31.07600 17.3070751 7 10.974 53.388 18.76900 25.1410 45.24550
## Peter Lake.7 38.60743 18.7907570 7 19.617 66.447 23.51900 30.7330 53.20850
## Peter Lake.8 27.79257  9.7876790 7 18.787 48.694 22.73850 25.6030 27.99350
##
## $comparison
## NULL
```

```
##
## $groups
##                   tp_ug groups
## Peter Lake.7 38.60743      a
## Peter Lake.6 31.07600      a
## Peter Lake.8 27.79257      a
## Peter Lake.5 18.72800      a
## Paul Lake.5  17.09000      a
## Paul Lake.7  16.61800      a
## Paul Lake.6  13.65100      a
## Paul Lake.9  10.93150      a
## Paul Lake.8  10.01350      a
##
## attr(,"class")
## [1] "group"
```

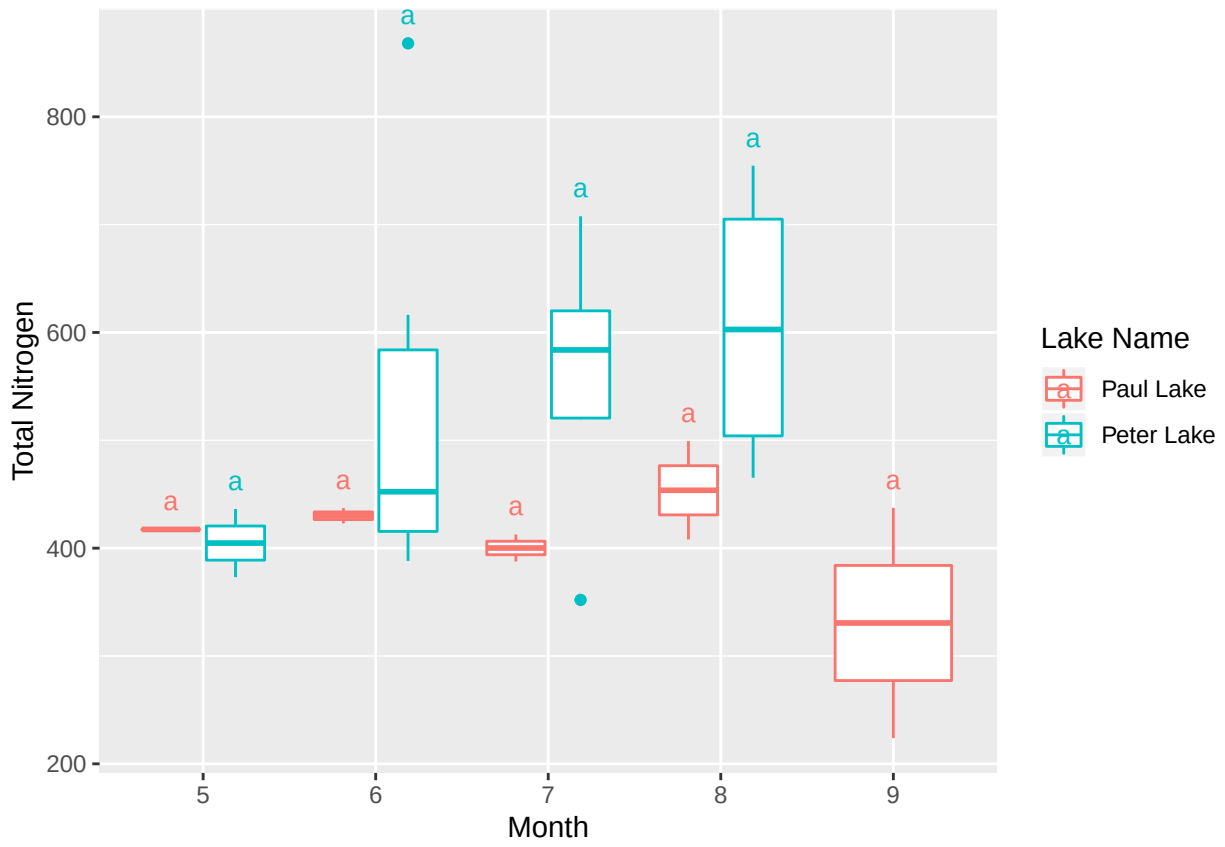
7. Create two plots, with TN (plot 1) or TP (plot 2) as the response variable and month and lake as the predictor variables. Hint: you may use some of the code you used for your visualization assignment. Assign groupings with letters, as determined from your tests. Adjust your axes, aesthetics, and color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

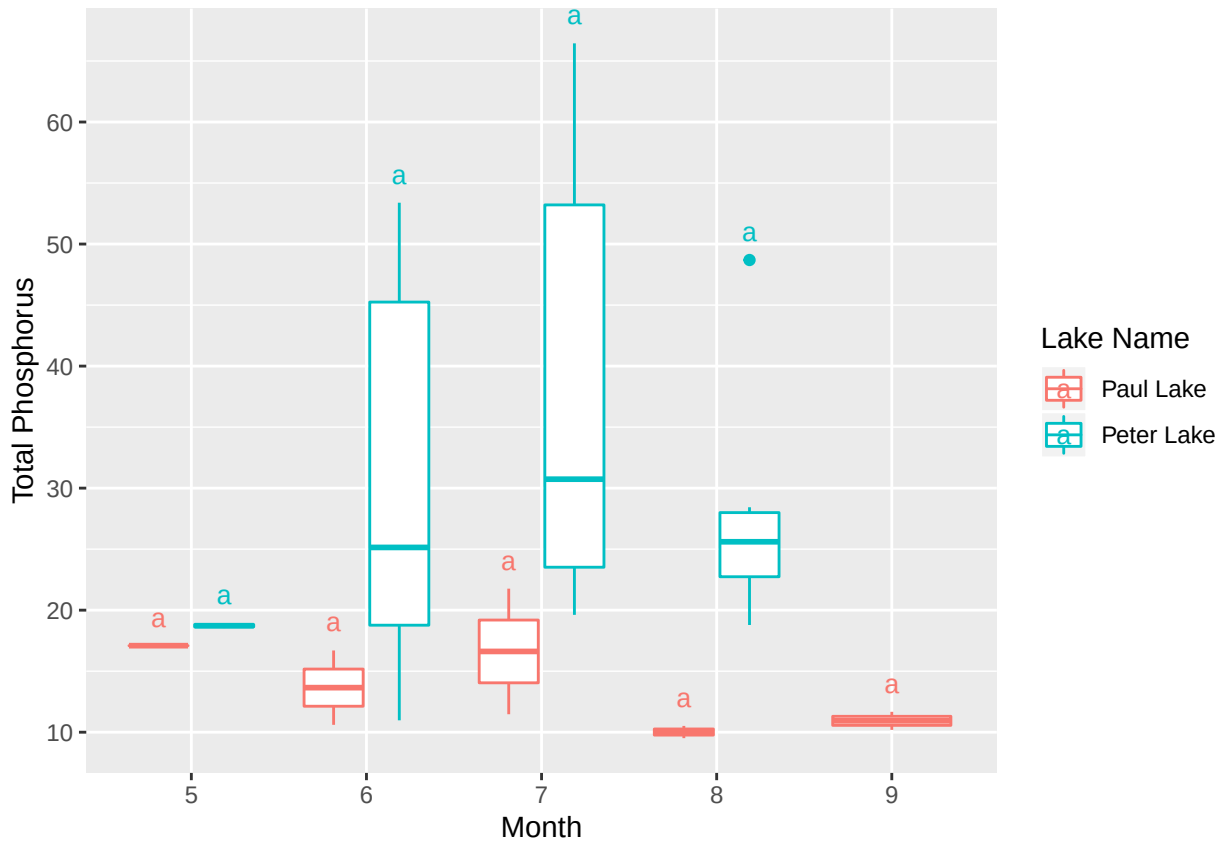
```
#7
Nutrient.anova.plot.n <- ggplot(Nutrients.Wrangled, aes(y = tn_ug, x = month, color = lakename)) +
  geom_boxplot()+
  stat_summary(geom = "text", fun.y = max, vjust = -1, size = 3.5,
               position = position_dodge(.75),
               label = c("a", "a", "a","a","a","a","a","a","a")) +
  labs(x = "Month", y = "Total Nitrogen", color= "Lake Name")
print(Nutrient.anova.plot.n)
```

```
## Warning: Removed 7 rows containing non-finite values (stat_boxplot).
```

```
## Warning: Removed 7 rows containing non-finite values (stat_summary).
```




```
Nutrient.anova.plot.p <- ggplot(Nutrients.Wrangled, aes(y = tp_ug, x = month, color = lakename)) +
  geom_boxplot()+
  stat_summary(geom = "text", fun.y = max, vjust = -1, size = 3.5,
               position = position_dodge(.75),
               label = c("a", "a", "a","a","a","a","a","a","a")) +
  labs(x = "Month", y = "Total Phosphorus", color= "Lake Name")
print(Nutrient.anova.plot.p)
```




```
#8
library(cowplot)
Nutrient.cowplot<-plot_grid(Nutrient.anova.plot.n, Nutrient.anova.plot.p, nrow = 2, align = 'v', rel_he
```

```
## Warning: Removed 7 rows containing non-finite values (stat_boxplot).
```

```
## Warning: Removed 7 rows containing non-finite values (stat_summary).
```