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

Kim Myers

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] "C:/Users/Temp/Documents/Duke/S20/DataAnalytics/Environmental_Data_Analytics_2020/Assignments"
```

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

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

#2

peterpaul$sampledate <- as.Date(peterpaul$sampledate, format = "%Y-%m-%d")

head(peterpaul$sampledate)
```

```
## [1] "1991-05-20" "1991-05-20" "1991-05-20" "1991-05-20" "1991-05-20"
## [6] "1991-05-20"
```

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.

```
#3

dpth9396 <- peterpaul %>%
  filter(depth_id == 1&year4>1993&year4<1997)

dpth9396$month <- as.factor(dpth9396$month)
```

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 conduct a two-way ANOVA with interaction effects. I will see if there are significant differences in P and N both between lakes and between lakes in different seasons.

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 - nitrogen
aovn <- aov(tn_ug~lakename*month, data=dpth9396)
summary(aovn)
```

```
##                Df  Sum Sq Mean Sq F value   Pr(>F)
## lakename        1 2059290 2059290  24.578 5.68e-06 ***
## month           4  565328  141332   1.687    0.164
## lakename:month  4  308095   77024   0.919    0.458
## Residuals      63 5278475   83785
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
## 23 observations deleted due to missingness
```

```
TukeyHSD(aovn)
```

```
##   Tukey multiple comparisons of means
##     95% family-wise confidence level
##
## Fit: aov(formula = tn_ug ~ lakename * month, data = dpth9396)
##
## $lakename
##                           diff      lwr      upr   p adj
## Peter Lake-Paul Lake 335.9449 200.5311 471.3587 5.7e-06
##
## $month
##          diff        lwr      upr     p adj
## 6-5 166.70182 -151.03328 484.4369 0.5832892
## 7-5 238.16277  -93.70072 570.0263 0.2706950
```

```
## 8-5 285.63898  -43.30055 614.5785 0.1185014
## 9-5 266.02098 -275.90983 807.9518 0.6434643
## 7-6  71.46095 -182.00398 324.9259 0.9320784
## 8-6 118.93716 -130.68717 368.5615 0.6687450
## 9-6  99.31915 -398.47608 597.1144 0.9802383
## 8-7  47.47621 -219.90009 314.8525 0.9872059
## 9-7  27.85821 -479.07165 534.7881 0.9998689
## 9-8 -19.61800 -524.63851 485.4025 0.9999671
##
## $`lakename:month`
##                                       diff          lwr        upr     p adj
## Peter Lake:5-Paul Lake:5    77.7536000  -559.050490  714.55769 0.9999948
## Paul Lake:6-Paul Lake:5     49.4124615  -493.365713  592.19064 0.9999996
## Peter Lake:6-Paul Lake:5   328.3022727  -225.964543  882.56909 0.6408341
## Paul Lake:7-Paul Lake:5     40.2005556  -530.252634  610.65375 1.0000000
## Peter Lake:7-Paul Lake:5   485.1906667   -85.262523 1055.64386 0.1626068
## Paul Lake:8-Paul Lake:5     77.4781111  -492.975079  647.93130 0.9999869
## Peter Lake:8-Paul Lake:5   536.3934000   -25.215018 1098.00182 0.0730423
## Paul Lake:9-Paul Lake:5    173.5785000  -648.532045  995.68904 0.9994733
## Peter Lake:9-Paul Lake:5   356.5320000  -704.808150 1417.87215 0.9827936
## Paul Lake:6-Peter Lake:5   -28.3411385  -527.891366  471.20909 1.0000000
## Peter Lake:6-Peter Lake:5  250.5486727  -261.461080  762.55843 0.8408165
## Paul Lake:7-Peter Lake:5   -37.5530444  -567.042539  491.93645 1.0000000
## Peter Lake:7-Peter Lake:5  407.4370667  -122.052428  936.92656 0.2764506
## Paul Lake:8-Peter Lake:5    -0.2754889  -529.764984  529.21401 1.0000000
## Peter Lake:8-Peter Lake:5  458.6398000   -61.308562  978.58816 0.1293851
## Paul Lake:9-Peter Lake:5    95.8249000  -698.409342  890.05914 0.9999953
## Peter Lake:9-Peter Lake:5  278.7784000  -761.118324 1318.67512 0.9965966
## Peter Lake:6-Paul Lake:6   278.8898112  -110.009514  667.78914 0.3717975
## Paul Lake:7-Paul Lake:6     -9.2119060  -420.852311  402.42850 1.0000000
## Peter Lake:7-Paul Lake:6   435.7782051    24.137800  847.41861 0.0296212
## Paul Lake:8-Paul Lake:6     28.0656496  -383.574755  439.70605 1.0000000
## Peter Lake:8-Paul Lake:6   486.9809385    87.687817  886.27406 0.0060988
## Paul Lake:9-Paul Lake:6    124.1660385  -596.872608  845.20468 0.9999014
## Peter Lake:9-Paul Lake:6   307.1195385  -678.006796 1292.24587 0.9897455
## Paul Lake:7-Peter Lake:6  -288.1017172  -714.776511  138.57308 0.4581573
## Peter Lake:7-Peter Lake:6  156.8883939  -269.786400  583.56319 0.9687363
## Paul Lake:8-Peter Lake:6  -250.8241616  -677.498956  175.85063 0.6505638
## Peter Lake:8-Peter Lake:6  208.0911273  -206.684178  622.86643 0.8205392
## Paul Lake:9-Peter Lake:6  -154.7237727  -884.449945  575.00240 0.9994551
## Peter Lake:9-Peter Lake:6   28.2297273  -963.272896 1019.73235 1.0000000
## Peter Lake:7-Paul Lake:7   444.9901111    -2.510188  892.49041 0.0524809
## Paul Lake:8-Paul Lake:7     37.2775556  -410.222744  484.77785 0.9999998
## Peter Lake:8-Paul Lake:7   496.1928444    60.023506  932.36218 0.0139753
## Paul Lake:9-Paul Lake:7    133.3779444  -608.717349  875.47324 0.9998592
## Peter Lake:9-Paul Lake:7   316.3314444  -684.309645 1316.97253 0.9886889
## Paul Lake:8-Peter Lake:7  -407.7125556  -855.212855   39.78774 0.1038171
## Peter Lake:8-Peter Lake:7   51.2027333  -384.966605  487.37207 0.9999963
## Paul Lake:9-Peter Lake:7  -311.6121667 -1053.707460  430.48313 0.9297066
## Peter Lake:9-Peter Lake:7 -128.6586667 -1129.299756  871.98242 0.9999918
## Peter Lake:8-Paul Lake:8   458.9152889    22.745950  895.08463 0.0314352
## Paul Lake:9-Paul Lake:8     96.1003889  -645.994904  838.19568 0.9999913
## Peter Lake:9-Paul Lake:8   279.0538889  -721.587200 1279.69498 0.9954286
## Paul Lake:9-Peter Lake:8  -362.8149000 -1098.132925  372.50313 0.8342774
```

```
## Peter Lake:9-Peter Lake:8 -179.8614000 -1175.486713  815.76391 0.9998531
## Peter Lake:9-Paul Lake:9   182.9535000  -979.686382 1345.59338 0.9999539
```

```
nhsd <- HSD.test(aovn, c("lakename","month"),group=T)
nhsdl <- HSD.test(aovn, "lakename",group=T)
print(nhsdl) # c,abc,c,abc,c,ab,bc,a,abc,abc
```

```
## $statistics
##    MSerror Df     Mean       CV
##   83785.32 63 545.0658 53.10496
##
## $parameters
##    test   name.t ntr StudentizedRange alpha
##   Tukey lakename   2          2.82608  0.05
##
## $means
##               tn_ug       std  r     Min      Max      Q25      Q50      Q75
## Paul Lake  379.3943  70.30007 37 219.406  557.812 335.7370 381.5510 422.9340
## Peter Lake 715.3392 413.14065 36 312.133 2048.151 480.0263 631.2905 749.5362
##
## $comparison
## NULL
##
## $groups
##               tn_ug groups
## Peter Lake 715.3392      a
## Paul Lake  379.3943      b
##
## attr(,"class")
## [1] "group"
```

```
#6 - phosphorus
aovp <- aov(tp_ug~lakename*month, data=dpth9396)
summary(aovp)
```

```
##                Df Sum Sq Mean Sq F value   Pr(>F)
## lakename        1   7845    7845  61.817 1.07e-11 ***
## month           4    761     190   1.500   0.2096
## lakename:month  4   1109     277   2.185   0.0774 .
## Residuals      85  10787     127
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
## 1 observation deleted due to missingness
```

```
TukeyHSD(aovp)
```

```
##   Tukey multiple comparisons of means
##     95% family-wise confidence level
##
## Fit: aov(formula = tp_ug ~ lakename * month, data = dpth9396)
##
## $lakename
##                          diff     lwr      upr p adj
## Peter Lake-Paul Lake 18.17559 13.5793 22.77189     0
##
## $month
```

```
##           diff        lwr       upr     p adj
## 6-5  5.3070072  -6.836241 17.450255 0.7409099
## 7-5 10.0728149  -2.070433 22.216063 0.1511904
## 8-5  5.7614293  -6.323856 17.846714 0.6741890
## 9-5  6.2830274  -9.540325 22.106380 0.8025642
## 7-6  4.7658077  -3.942567 13.474182 0.5491006
## 8-6  0.4544221  -8.172943  9.081787 0.9998934
## 9-6  0.9760202 -12.393938 14.345979 0.9996105
## 8-7 -4.3113856 -12.938750  4.315979 0.6338963
## 9-7 -3.7897875 -17.159746  9.580171 0.9328422
## 9-8  0.5215981 -12.795737 13.838934 0.9999673
##
## $`lakename:month`
##                                  diff         lwr        upr     p adj
## Peter Lake:5-Paul Lake:5    2.2465000 -22.3077988  26.800799 0.9999996
## Paul Lake:6-Paul Lake:5    -1.8121923 -22.7409798  19.116595 0.9999998
## Peter Lake:6-Paul Lake:5   12.9028077  -8.0259798  33.831595 0.5992351
## Paul Lake:7-Paul Lake:5    -2.9750385 -23.9038260  17.953749 0.9999830
## Peter Lake:7-Paul Lake:5   23.5972692   2.6684817  44.526057 0.0148889
## Paul Lake:8-Paul Lake:5    -3.6925000 -24.6212875  17.236288 0.9998938
## Peter Lake:8-Paul Lake:5   15.6487857  -5.1033844  36.400956 0.3105154
## Paul Lake:9-Paul Lake:5    -2.2842500 -28.1667536  23.598254 0.9999997
## Peter Lake:9-Paul Lake:5   15.2328333 -12.7234657  43.189132 0.7514159
## Paul Lake:6-Peter Lake:5   -4.0586923 -23.3206691  15.203284 0.9995410
## Peter Lake:6-Peter Lake:5  10.6563077  -8.6056691  29.918284 0.7348314
## Paul Lake:7-Peter Lake:5   -5.2215385 -24.4835152  14.040438 0.9966790
## Peter Lake:7-Peter Lake:5  21.3507692   2.0887925  40.612746 0.0180180
## Paul Lake:8-Peter Lake:5   -5.9390000 -25.2009768  13.322977 0.9914136
## Peter Lake:8-Peter Lake:5  13.4022857  -5.6676426  32.472214 0.4103379
## Paul Lake:9-Peter Lake:5   -4.5307500 -29.0850488  20.023549 0.9998457
## Peter Lake:9-Peter Lake:5  12.9863333 -13.7450014  39.717668 0.8546735
## Peter Lake:6-Paul Lake:6   14.7150000   0.3579702  29.072030 0.0399474
## Paul Lake:7-Paul Lake:6    -1.1628462 -15.5198760  13.194184 0.9999999
## Peter Lake:7-Paul Lake:6   25.4094615  11.0524317  39.766491 0.0000059
## Paul Lake:8-Paul Lake:6    -1.8803077 -16.2373375  12.476722 0.9999916
## Peter Lake:8-Paul Lake:6   17.4609780   3.3626546  31.559301 0.0046163
## Paul Lake:9-Paul Lake:6    -0.4720577 -21.4008452  20.456730 1.0000000
## Peter Lake:9-Paul Lake:6   17.0450256  -6.3999058  40.489957 0.3614600
## Paul Lake:7-Peter Lake:6  -15.8778462 -30.2348760  -1.520816 0.0184724
## Peter Lake:7-Peter Lake:6  10.6944615  -3.6625683  25.051491 0.3273443
## Paul Lake:8-Peter Lake:6  -16.5953077 -30.9523375  -2.238278 0.0111406
## Peter Lake:8-Peter Lake:6   2.7459780 -11.3523454  16.844301 0.9997586
## Paul Lake:9-Peter Lake:6  -15.1870577 -36.1158452   5.741730 0.3641468
## Peter Lake:9-Peter Lake:6   2.3300256 -21.1149058  25.774957 0.9999992
## Peter Lake:7-Paul Lake:7   26.5723077  12.2152779  40.929337 0.0000019
## Paul Lake:8-Paul Lake:7    -0.7174615 -15.0744913  13.639568 1.0000000
## Peter Lake:8-Paul Lake:7   18.6238242   4.5255008  32.722148 0.0018213
## Paul Lake:9-Paul Lake:7     0.6907885 -20.2379991  21.619576 1.0000000
## Peter Lake:9-Paul Lake:7   18.2078718  -5.2370597  41.652803 0.2713907
## Paul Lake:8-Peter Lake:7  -27.2897692 -41.6467990 -12.932739 0.0000010
## Peter Lake:8-Peter Lake:7  -7.9484835 -22.0468069   6.149840 0.7133014
## Paul Lake:9-Peter Lake:7  -25.8815192 -46.8103068  -4.952732 0.0047119
## Peter Lake:9-Peter Lake:7  -8.3644359 -31.8093674  15.080496 0.9764362
## Peter Lake:8-Paul Lake:8   19.3412857   5.2429623  33.439609 0.0010031
```

```
## Paul Lake:9-Paul Lake:8     1.4082500 -19.5205375  22.337038 1.0000000
## Peter Lake:9-Paul Lake:8   18.9253333  -4.5195982  42.370265 0.2234770
## Paul Lake:9-Peter Lake:8  -17.9330357 -38.6852058   2.819134 0.1505667
## Peter Lake:9-Peter Lake:8  -0.4159524 -23.7033573  22.871453 1.0000000
## Peter Lake:9-Paul Lake:9   17.5170833 -10.4392157  45.473382 0.5767472
```

```
phsd <- HSD.test(aovp, c("lakename","month"),group=T)
phsdl <- nhsdl <- HSD.test(aovp, "lakename",group=T)
print(phsdl) # bc,bc,c,abc,c,a,c,ab,c,abc
```

```
## $statistics
##    MSerror Df     Mean       CV
##   126.9058 85 19.43009 57.97835
##
## $parameters
##    test   name.t ntr StudentizedRange alpha
##   Tukey lakename   2         2.811835  0.05
##
## $means
##               tp_ug       std  r    Min    Max     Q25    Q50      Q75
## Paul Lake  10.24664  3.735219 47  1.222 21.763  7.8600  9.931 11.22300
## Peter Lake 28.42223 15.989260 48 10.887 66.893 16.3395 22.402 37.28625
##
## $comparison
## NULL
##
## $groups
##               tp_ug groups
## Peter Lake 28.42223      a
## Paul Lake  10.24664      b
##
## 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.

```
library(cowplot)
mytheme <- theme_light(base_size = 12) +
  theme(axis.text = element_text(color = "black"),
        legend.position = "top")

theme_set(mytheme)

#7
n <- ggplot(dpth9396) + geom_boxplot(aes(x=month, y=tn_ug, fill=lakename),size=0.2) +
  labs(x="",y=expression("N " *mu~ "g liter"^-1),fill="") +
  scale_color_brewer(palette = "Dark2") +
  stat_summary(aes(x=month,y=tn_ug,fill=lakename),geom = "text", fun.y = max, vjust = -1,
  size = 3.5,label = c("c","abc","c","abc","c","ab","bc","a","abc","abc"),position=position_dodge(width=
```

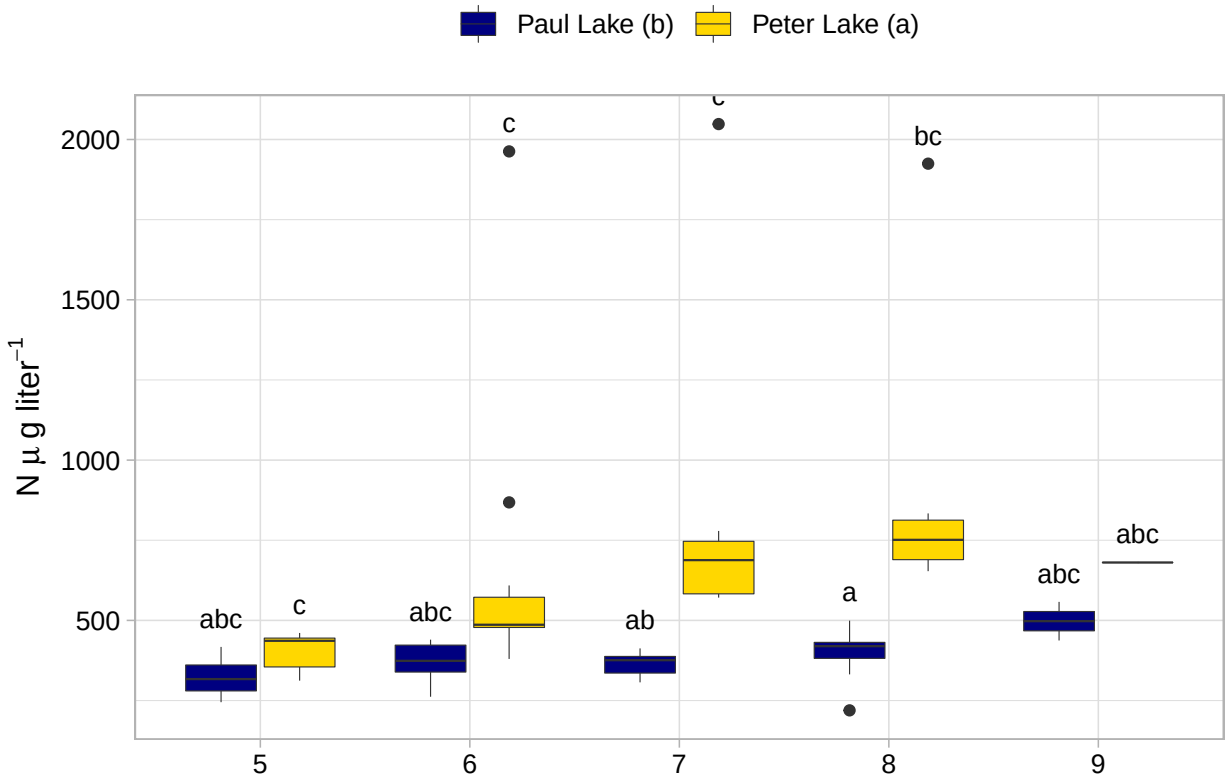
```
  scale_fill_manual(values=c("navy","gold"),labels=c("Paul Lake (b)","Peter Lake (a)"))
```

```
## Warning: Ignoring unknown aesthetics: fill
```

```
print(n)
```

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

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



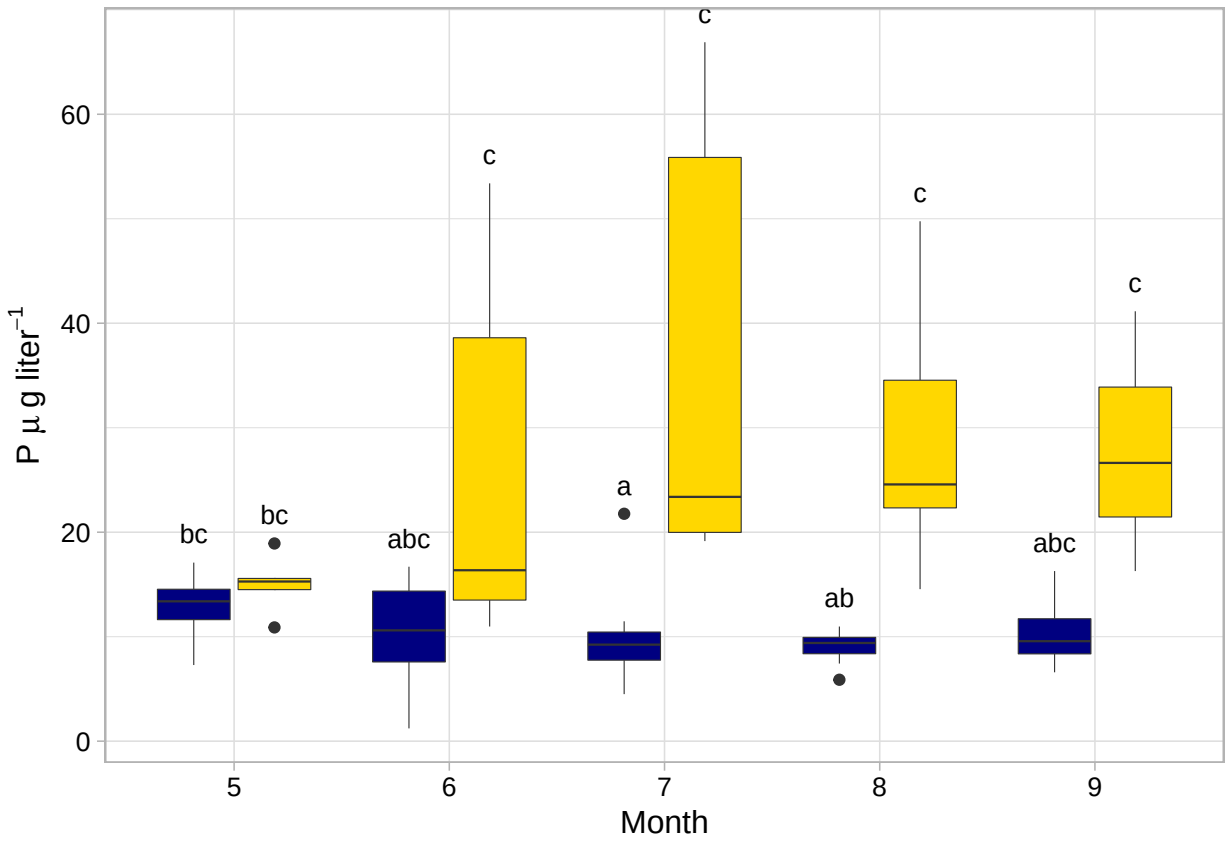

```
p <- ggplot(dpth9396) + geom_boxplot(aes(x=month, y=tp_ug, fill=lakename), size=0.2) +
  labs(x="Month",y=expression("P " *mu~ "g liter"^-1),fill="") +
  scale_color_brewer(palette = "Dark2") +
  theme(legend.position="none") +
  stat_summary(aes(x=month,y=tp_ug,fill=lakename),geom = "text", fun.y = max, vjust = -1,
  size = 3.5,label = c("bc","bc","c","abc","c","a","c","ab","c","abc"),position=position_dodge(width=0.
  scale_fill_manual(values=c("navy","gold"))
```

```
## Warning: Ignoring unknown aesthetics: fill
```

```
print(p)
```

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

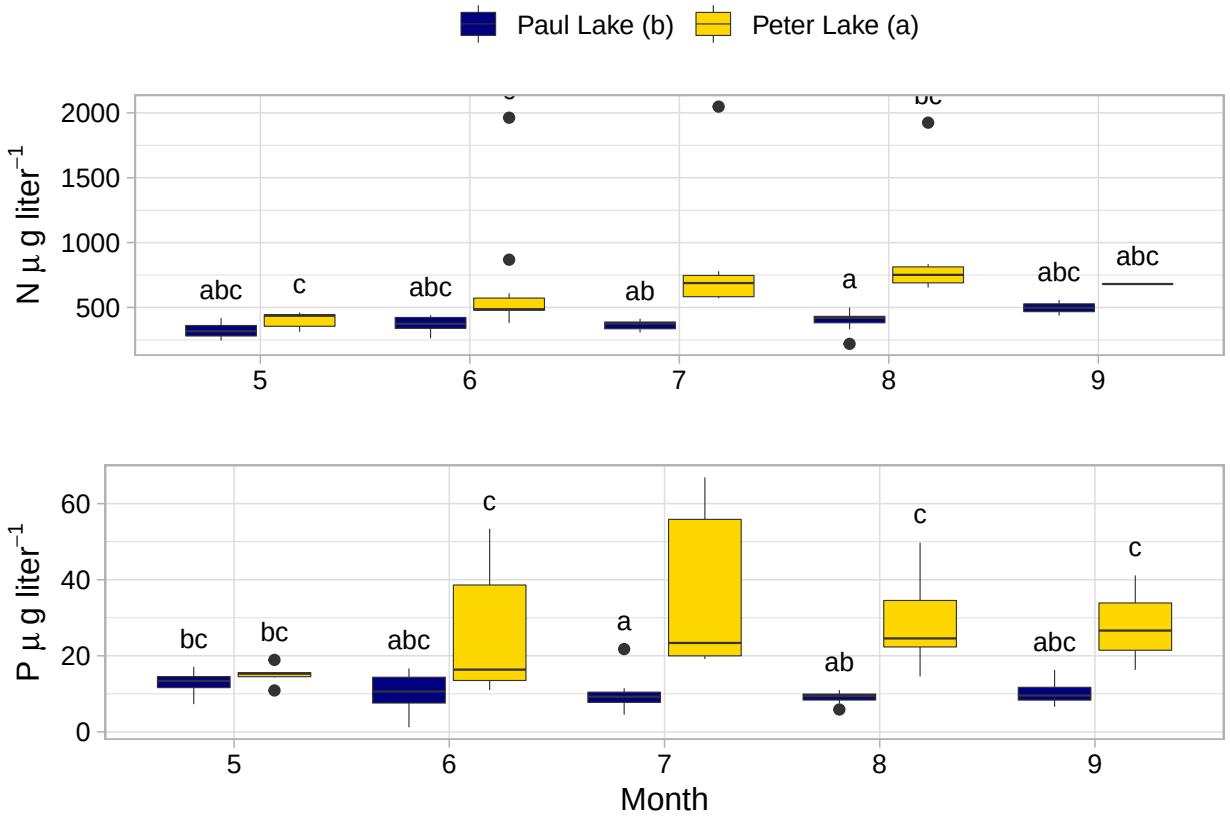
```
## Warning: Removed 1 rows containing non-finite values (stat_summary).
```




```
#8
np <- plot_grid(n, p, nrow = 2, align = 'h', rel_heights = c(1.25, 1))
```

```
## 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.
```

```
print(np)
```


Paul Lake (b)
Peter Lake (a)
N µ g liter−1
2000
1500
1000
500
abc
c
abc
ab
a
bc
abc
5
6
7
8
9
P µ g liter−1
60
40
20
0
bc
bc
abc
c
a
ab
c
abc
c
Month