Case Study 9.1 from Cabrera and McDougall's *Statistical Consulting.*

Introduction:

This study was conducted to explore if an experimental, learning-style specific teaching method provides an advantage over traditional classroom learning methods. This was done by assigning groups of students to take science class units using either traditional or experimental teaching methods and comparing the results of a number of tests afterwards. The specific research questions are as follows:

1. Do the students who were taught using the experimental teaching methods have a higher average score on the unit post-test than those who were taught using traditional methods?

2. Do the students who were taught using the experimental teaching methods have a higher score on a test measuring higher-learning performance than those who were taught using traditional methods?

3. Do the students who were taught using the experimental teaching methods feel better and more confident about studying the sciences than those who were taught using traditional methods?

4. Do student's learning style preferences influence how their post-test scores behave based on teaching method?

Study Design + Data:

This study was comprised of 59 elementary school students (ages 9-10 yrs) divided into 3 groups. Those 3 groups were taught 3 science units in succession over the course of 3 weeks. Before the study period began, each student's learning style preferences were measured using a LSI (Learning Style Inventory) test, scoring each student's preference for each of the following learning styles: Tactile, Auditory, Kinesthetic, and Visual.

At the beginning of each unit, each student was given a pre-test to establish the student's prior knowledge on the subject. At the end, the students were given three tests: An

official post-test (measured in %), A SDS (Semantic Differential Scale) test measuring the student's confidence regarding the learning sciences (scale of 1-60), and an HLT (Higher Learning Test) that required a worked-out solution to the given problem (scale of 0-4).

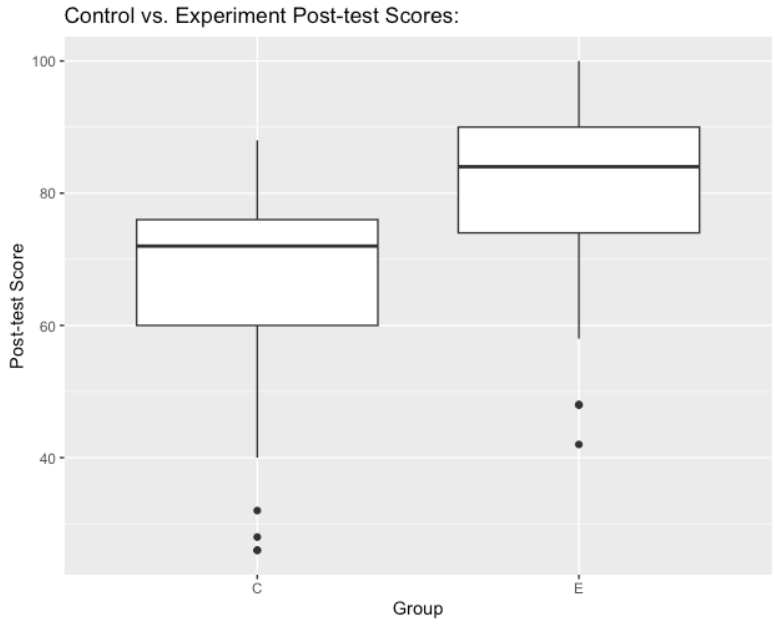
For each unit, two of the student groups were assigned to experimental teaching methods, while one was assigned the control (traditional) teaching method. Each of the three student groups participated in one unit where they were the control group.

Analysis:

H1: There will be a difference between Post-Test Score (PST) by GROUP.

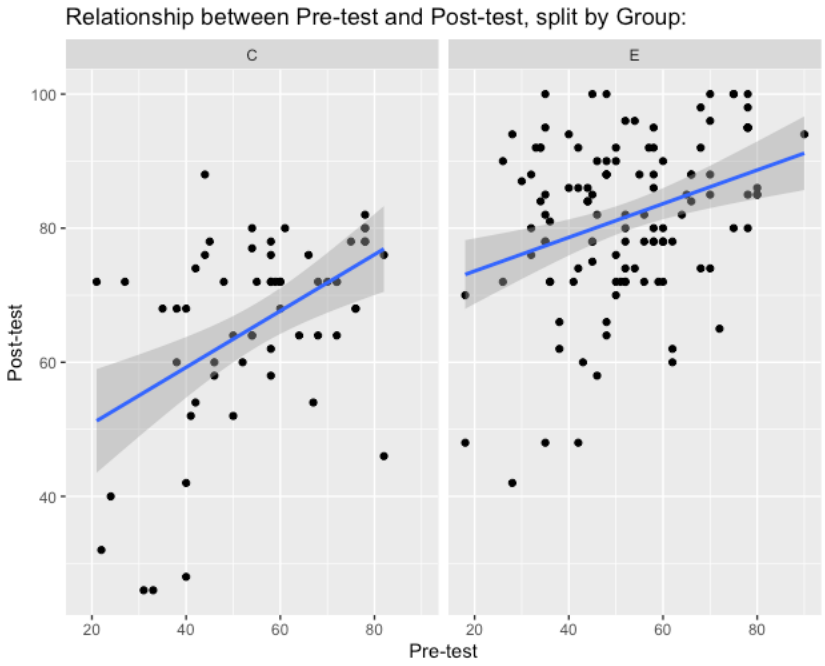
Yes; There was found a significant difference between PST by Group.

First, I took a look at the distribution of Post-test scores based on group:



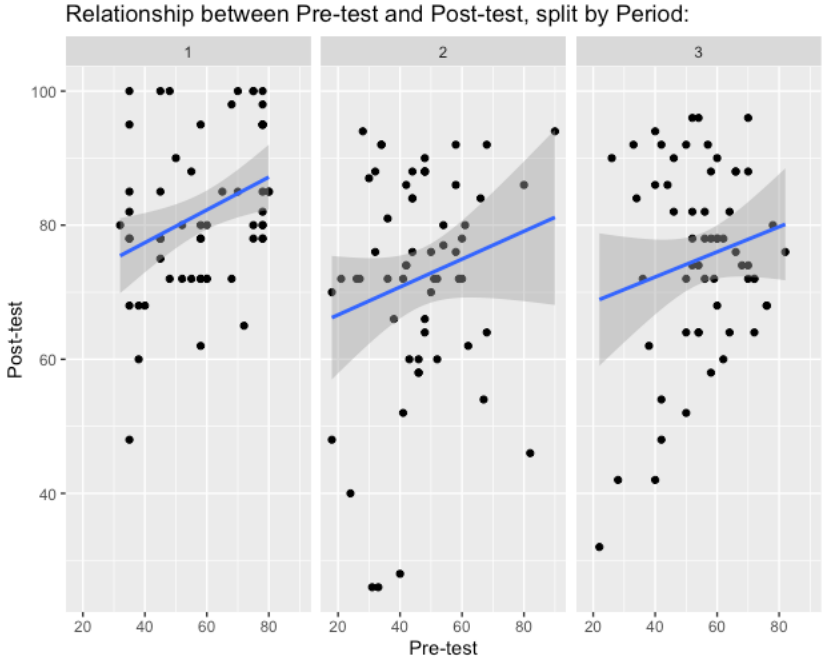
As we can see, the mean score is a bit higher for the experimental group than it is for the control group. The question is: is that difference significant?

Now, there are other factors than group that could potentially influence the post-test scores.



Pre-test score is one such variable. As shown in the above plot, as pre-test increases, post-test tends to increase as well. Additionally, It seems that between the two groups, the relationship between Post-test and pre-test changes. Pre-test will be included in our model as a baseline from which to compare post-test scores. Its interaction with Group will also be included.

Also, we can look at the effect of period (which class/subject the students were taught).



We can see a difference between the overall distribution of scores between Period, and the intercepts of the regression lines are in different places (but they seem approximately the same slope). We'll test for the difference between these Period and Period's effect on Group and Pre-test scores.

When I ran the test, I accounted for differences in scores between individuals and groups of students. If one class (group of students) does better than the others in general, we want our conclusion to account for that. And we want to test if individual students are learning better, not the student body as a whole, right?

I used a two-way ANCOVA test, where the two factors are Group and Period, while PRE was used as the covariate baseline. Interactions were tested for between the IVs, and the one found to be statistically significant was between Group and PRE.

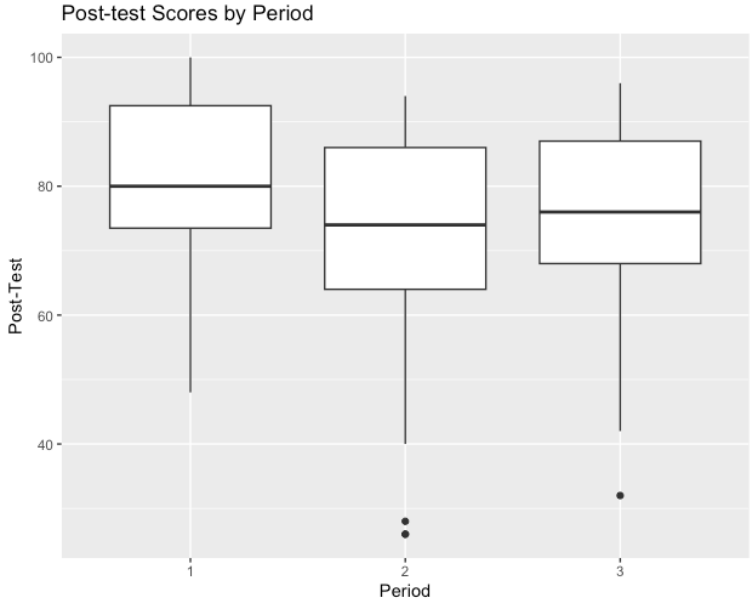
After removing the statistically insignificant interactions, the model is as follows:

Formula for lmer: PST ~ Group * PRE + Period + (1 | id) + (1 | class)

```
Type III Analysis of Variance Table with Satterthwaite's method
           Sum Sq Mean Sq NumDF  DenDF F value    Pr(>F)
Group     2130.18 2130.18     1 149.16 21.7470 6.854e-06 ***
PRE       1868.72 1868.72     1 169.85 19.0777 2.177e-05 ***
Period    1539.76  769.88     2 117.69  7.8597 0.0006253 ***
Group:PRE  401.72  401.72     1 151.60  4.1011 0.0446097 *
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
```

What does this all mean?

1. There is a significant difference in mean post-test score between the experimental and control groups. (Experimental is higher by approximately 16 points.)
2. There is a significant difference between post test scores of the 3 classes/subjects.

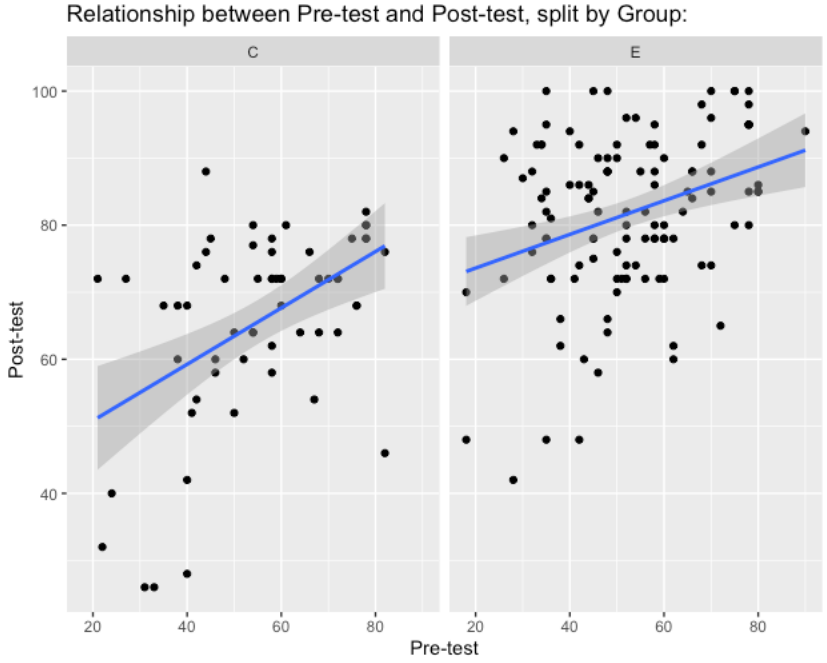


```
Least Squares Means table:

                  Estimate Std. Error    df  t value      lower      upper  Pr(>|t|)
GroupC - GroupE  -16.820578   1.592037 113.5 -10.5654 -19.974537 -13.666618 < 2.2e-16 ***
Period1 - Period2  6.552416   1.952975 121.9   3.3551   2.686273  10.418559 0.0010581 **
Period1 - Period3  6.491788   1.848068 114.0   3.5127   2.830788  10.152788 0.0006367 ***
Period2 - Period3 -0.060628   1.897607 117.7  -0.0319  -3.818520   3.697265 0.9745664
---
```

After running a pairwise test on an ANOVA testing the difference on our model, we can conclude that Period 1 has a significantly higher average score than the other two groups.

3. Post-test scores are positively correlated with pre-test scores.
4. The amount that post-test and pre-test scores are correlated changes in the two groups.



It appears that the experimental teaching method flattens out the final difference between the students who scored lower on the pre-test and those who scored higher.

H2: There will be a difference between HLT by GROUP.

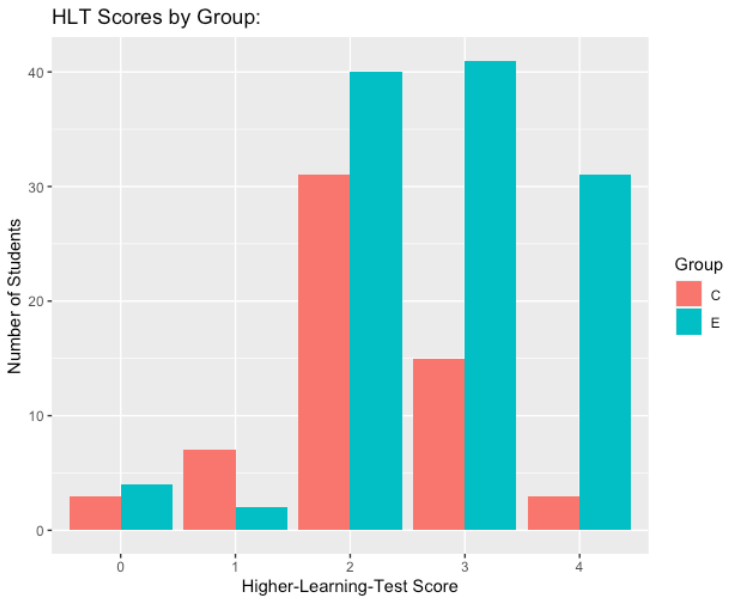
Initially, there is not a significant difference between HLT by Group.

A two-way anova on an ordinal logistic regression model was used, where HLT is the response, while period and class are the predictors. Again, id was used to account for random effect. The output from the test is as follows:

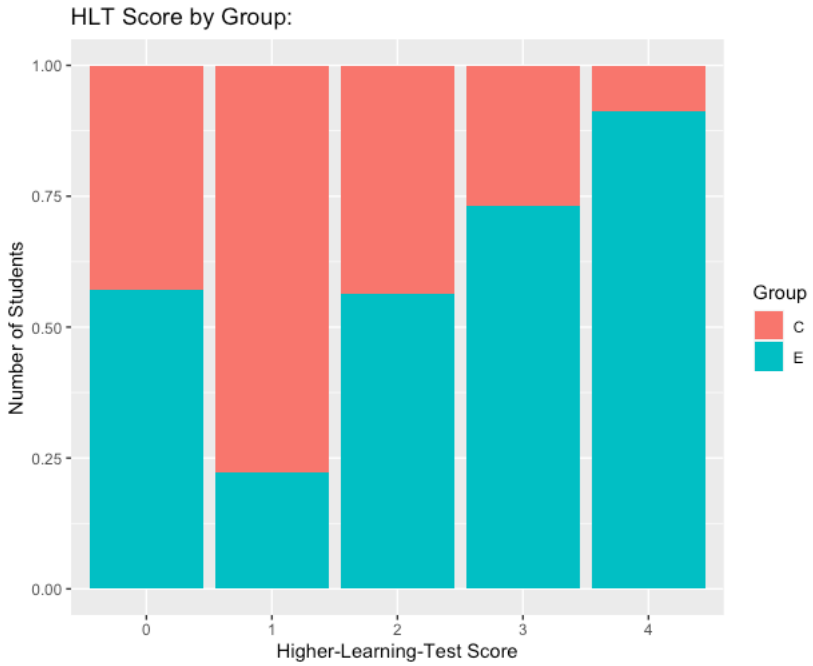
Formula for clmm: HLT ~ Group * Period + (1 | id)

```
Response: HLT
      LR Chisq Df Pr(>Chisq)
g1     24.7763  1  6.438e-07 ***
u2      1.8526  1    0.17348
u3      0.0023  1    0.96211
g1:u2   0.2300  1    0.63155
g1:u3   3.9918  1    0.04572 *
```

This output indicates that, overall, there is a significant difference between HLT between the two groups. The presence of an interaction indicates there is some sort of relationship between the two groups and if the unit was unit 3. If we look at scores by group, we can see that the Experimental group seems to score higher overall, indicated by the first p-value of ~6.4e-7:



Below is another way to visualize, based on proportion of total students who scored in each category:



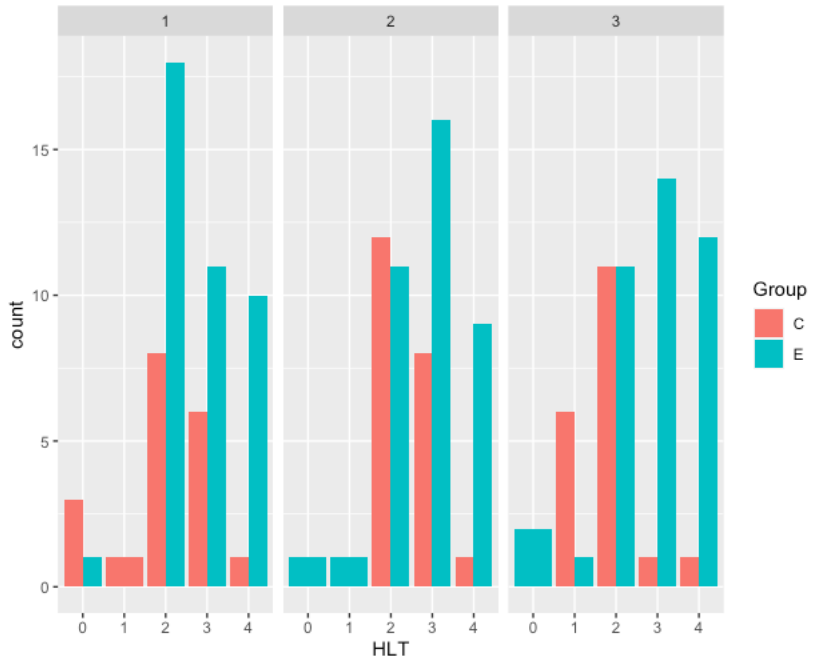
Here's what I found running a couple pairwise tests:

```
Period = 1:
 contrast estimate    SE  df z.ratio p.value
 C - E      -1.021 0.715 Inf  -1.429  0.1530

Period = 2:
 contrast estimate    SE  df z.ratio p.value
 C - E      -0.672 0.616 Inf  -1.091  0.2751

Period = 3:
 contrast estimate    SE  df z.ratio p.value
 C - E      -3.156 0.847 Inf  -3.725  0.0002
```

It appears we have one difference here (hence the interaction). It appears that the Experimental group scores significantly higher only during Period 3. In the following graph, we can see that in Period 3, there is a higher frequency of high-scoring experimental group compared to control, while in the other 2 periods the red control lines are larger on the right sides of the graphs, indicating no significant difference.



H3: There will be a difference between SDS by GROUP.

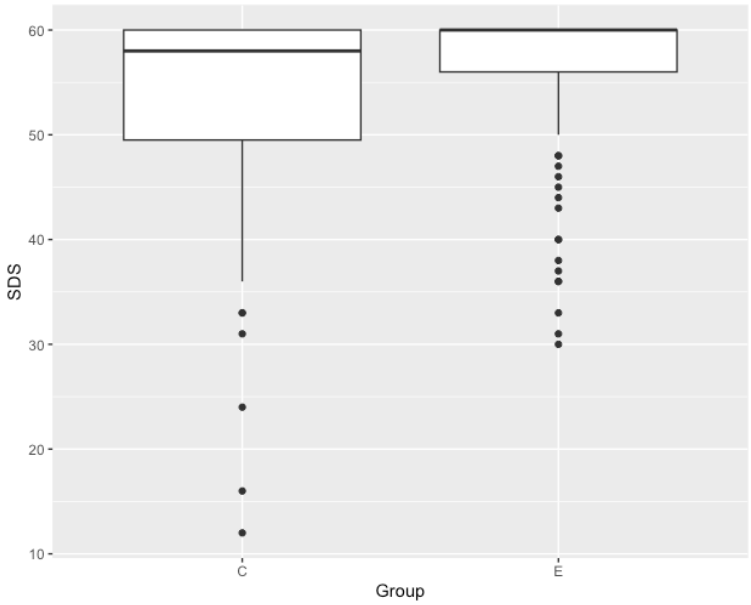
Yes, there is a significant difference in SDS by GROUP.

A two-way ancova was used, where Group and Period were the factors, PRE was the covariate, and SDS was the response. Id and class were used to account for random effect, as in earlier cases. No significant interactions were found, so the interaction terms were removed from the model. The output is as follows:

Formula for lmer: SDS ~ Group + PRE + Period + (1 | id) + (1 | class)

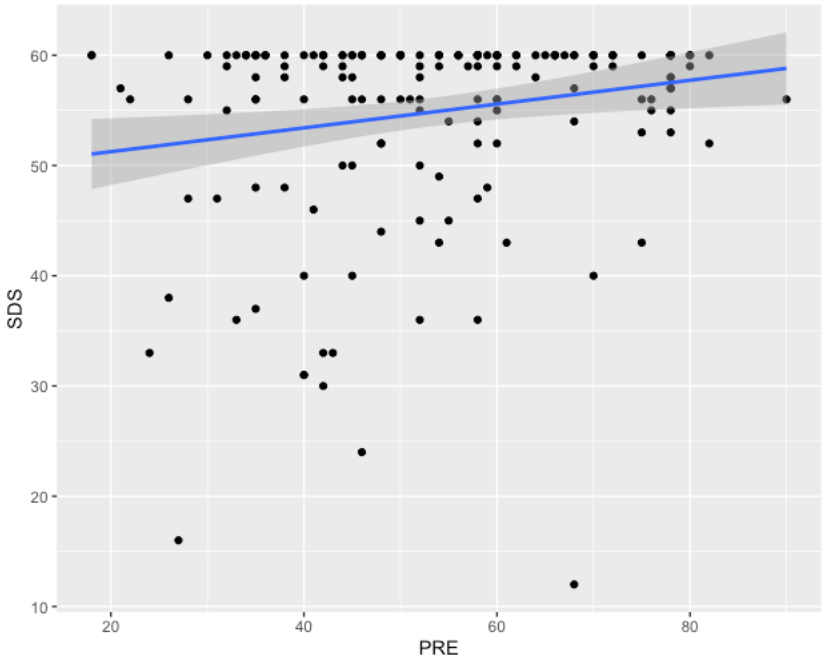
```
Type III Analysis of Variance Table with Satterthwaite's method
       Sum Sq Mean Sq NumDF  DenDF F value   Pr(>F)
Group  457.84  457.84     1 114.80 11.2400 0.001085 **
PRE    125.67  125.67     1 169.47  3.0853 0.080810 .
Period 432.96  216.48     2 118.61  5.3146 0.006160 **
---
Signif. codes:  0 ‘***’ 0.001 ‘**’ 0.01 ‘*’ 0.05 ‘.’ 0.1 ‘ ’ 1
```

1. There is a significant difference between mean SDS scores by Group:



From the plot, we can see that a large percentage of the students in all scored highly on this test, but nearly all of the students in the experimental group scored the maximum (60). This is a test measuring how students feel about the learning sciences, so, we can conclude that the Experimental group felt better than the control group.

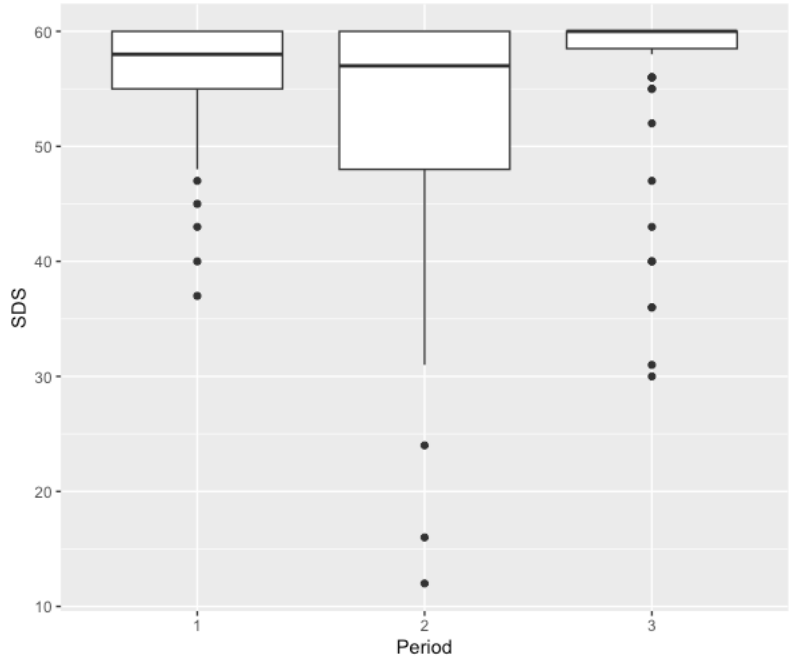
2. There is not a significant correlation between PRE and SDS:



As we can see, though there are less low SDS values as we move to the right, there is not a huge correlation between these two variables. Many students from all values of PRE scored highly on the SDS score. This leads us to conclude that the student's competence in the sciences didn't make

them feel any more or less confident than other students at the end of each unit.

3. There is a significant difference in SDS by period.



As we can see, there is a difference between the three periods. Which ones are significantly different?

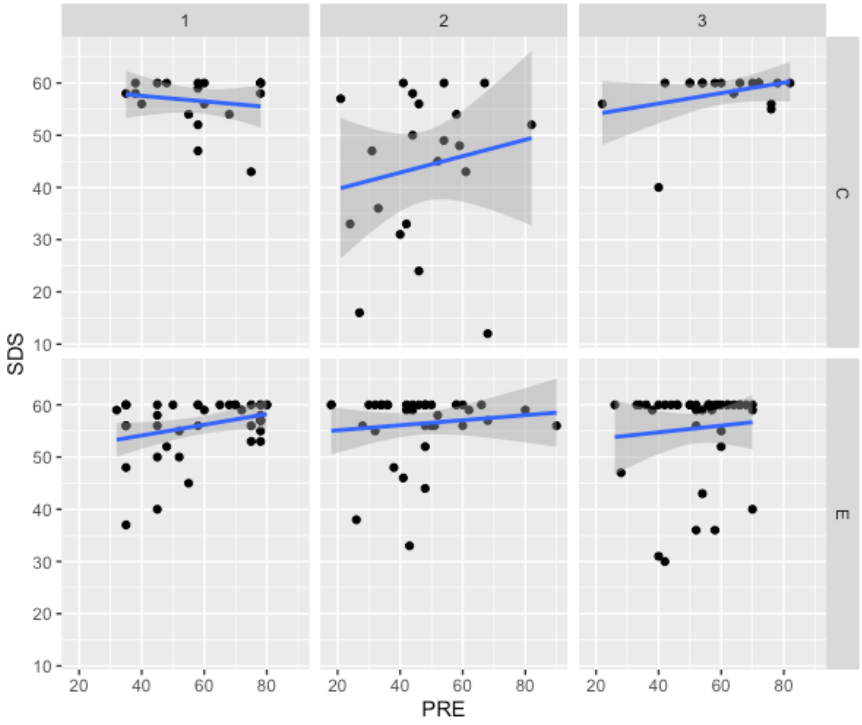
```
Least Squares Means table:

                 Estimate Std. Error    df t value    lower    upper Pr(>|t|)
GroupC - GroupE   -3.43625    1.02495 114.8 -3.3526 -5.46651 -1.40598 0.001085 **
Period1 - Period2  3.32678    1.26093 122.9  2.6384  0.83084  5.82272 0.009408 **
Period1 - Period3 -0.38555    1.18253 114.8 -0.3260 -2.72797  1.95687 0.744991
Period2 - Period3 -3.71233    1.21924 118.7 -3.0448 -6.12660 -1.29805 0.002868 **
---
```

According to the pairwise analysis, Period two is significantly lower than the other two, meaning students were less confident after unit 2 about the sciences than they were after the other two units. Potentially unit 2 was a harder subject to grasp? We don't have that specific information, but could lead us to believe that.

4. What about interactions?

As mentioned earlier, no significant interactions were found.



If we look at this plot, we can see that, potentially there is an interaction betseen Period and PRE. If we look at boxes C1 and C2 in the above plot, we can see that the slope of PRE vs SDS changes based on Period.

Below are the results from the ANOVA with interactions:

```
Type III Analysis of Variance Table with Satterthwaite's method
                 Sum Sq Mean Sq NumDF  DenDF F value Pr(>F)
Group            12.885  12.885     1 138.93  0.3109 0.5780
PRE              87.899  87.899     1 161.83  2.1212 0.1472
Period          109.444  54.722     2 139.75  1.3206 0.2703
Group:PRE         1.229   1.229     1 140.67  0.0297 0.8635
Group:Period     57.753  28.877     2 149.67  0.6969 0.4998
PRE:Period       37.041  18.521     2 141.00  0.4469 0.6405
Group:PRE:Period 45.883  22.942     2 152.43  0.5536 0.5760
```

None of the interactions in the above output have a p-value of 0.05 or less. Even if we just tested the interaction between Period and PRE (like we mentioned above):

```
Type III Analysis of Variance Table with Satterthwaite's method
           Sum Sq Mean Sq NumDF  DenDF F value   Pr(>F)
Group      479.00  479.00     1 114.14 11.6140 0.000905 ***
PRE        144.21  144.21     1 167.82  3.4966 0.063236 .
Period      65.70   32.85     2 142.22  0.7965 0.452893
PRE:Period  29.41   14.71     2 143.79  0.3565 0.700710
---
Signif. codes:  0 ‘***’ 0.001 ‘**’ 0.01 ‘*’ 0.05 ‘.’ 0.1 ‘ ’ 1
```

Still, the interaction is not significant. Therefore, though there are significant differences between Period and Group, as one variable changes (Group, Period, PRE), there is no significant change in the effect of another variable.

H4: There will be an interaction between PST and PREF by GROUP.

There is a significant difference between PST by Group, as well as by PREF. There are significant interactions between GROUP and PREF as well.

To answer this problem, another ancova was run. This time, it was just a 1-way ancova with 5 covariates: PRE and the 4 attributes of PREF. Period wasn't used in this current analysis for the sake of simplicity.

Here is the output from the model:

Formula: PST ~ Group + t + k + a + v + Group:t + Group:k + Group:a + Group:v +
PRE + (1 | id) + (1 | class)

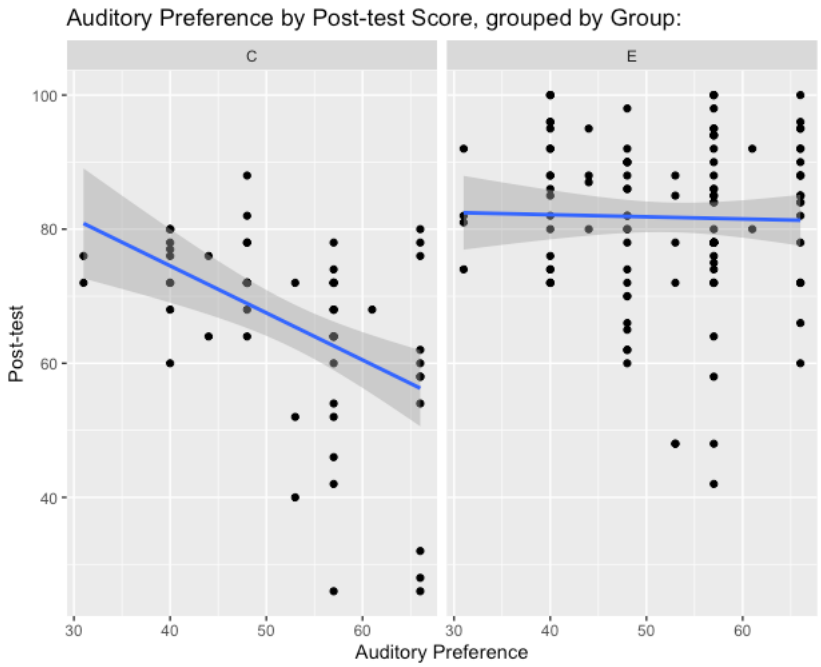
```
Type III Analysis of Variance Table with Satterthwaite's method
         Sum Sq Mean Sq NumDF   DenDF F value    Pr(>F)
Group   1462.13 1462.13     1 112.832 16.3200 9.795e-05 ***
t        543.86  543.86     1  61.180  6.0705 0.0165721 *
k         23.73   23.73     1  61.675  0.2648 0.6086670
a       1043.48 1043.48     1  62.535 11.6472 0.0011327 **
v         62.60   62.60     1  61.625  0.6987 0.4064468
PRE     2552.32 2552.32     1 165.712 28.4886 3.062e-07 ***
Group:t  239.73  239.73     1 113.488  2.6758 0.1046561
Group:k  337.21  337.21     1 111.908  3.7638 0.0548883 .
Group:a 1090.93 1090.93     1 113.587 12.1768 0.0006897 ***
Group:v  514.11  514.11     1 112.828  5.7384 0.0182433 *
---
Signif. codes:  0 ‘***’ 0.001 ‘**’ 0.01 ‘*’ 0.05 ‘.’ 0.1 ‘ ’ 1
```

As we can see, there's still a significant difference in PST based on PRE and Group, just like in H1. There also is a difference in PST based on t and a, while a and v change the effect Group has on PST. After removing k and its interaction with Group from the model, we get the following output:

```
Type III Analysis of Variance Table with Satterthwaite's method
         Sum Sq Mean Sq NumDF   DenDF F value    Pr(>F)
Group   1275.00 1275.00     1 113.700 13.8691 0.0003066 ***
t        777.47  777.47     1  62.220  8.4571 0.0050349 **
a       1037.27 1037.27     1  63.461 11.2832 0.0013269 **
v         58.53   58.53     1  62.782  0.6367 0.4279275
PRE     2780.35 2780.35     1 167.947 30.2439 1.397e-07 ***
Group:t  619.72  619.72     1 114.467  6.7411 0.0106568 *
Group:a 1321.14 1321.14     1 114.685 14.3710 0.0002410 ***
Group:v  545.37  545.37     1 113.766  5.9324 0.0164168 *
---
Signif. codes:  0 ‘***’ 0.001 ‘**’ 0.01 ‘*’ 0.05 ‘.’ 0.1 ‘ ’ 1
```

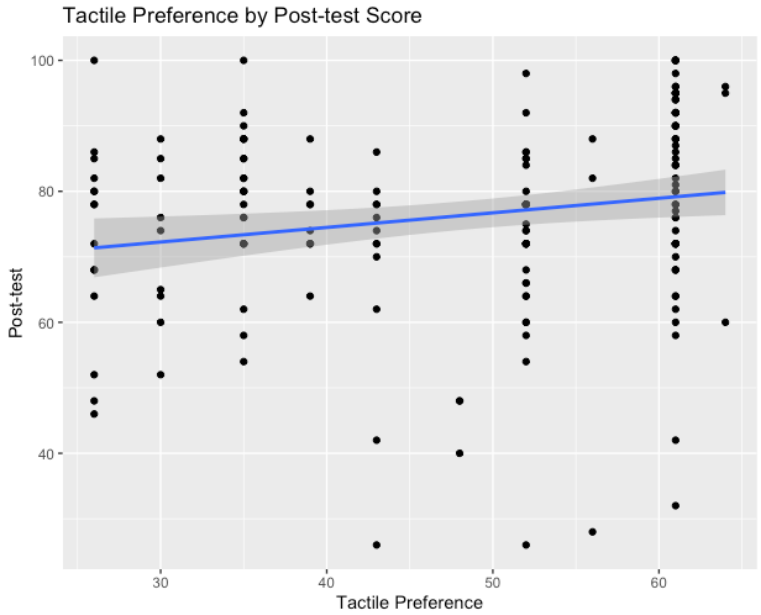
Additionally, the interaction between Group and t has become significant here. We already illustrated the difference in PST by Group and by PRE, so I won't worry about illustrating that here.

1. A and PST are correlated, and a changes the effect of Group on PST. A stands for Auditory is a measure of a student's preference to auditory learning methods.

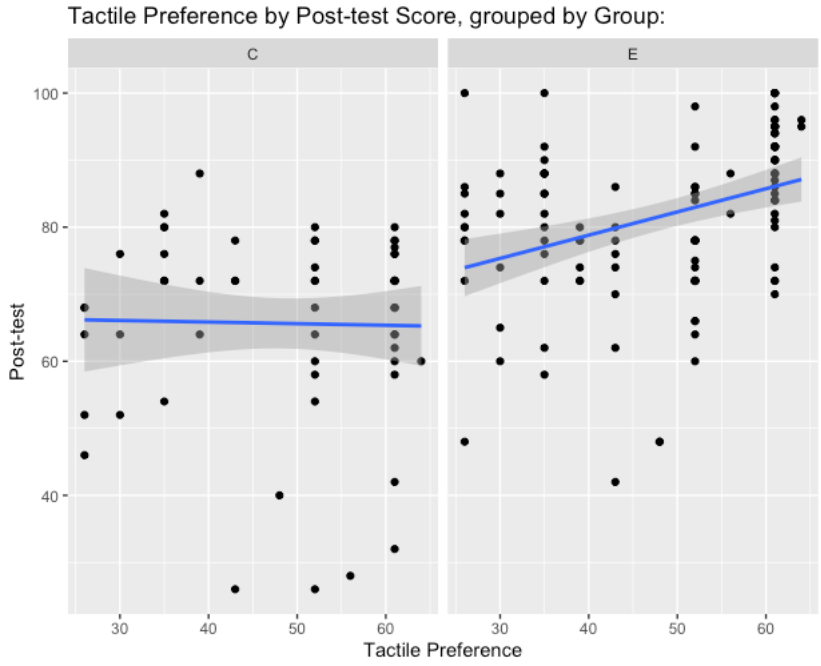


For the control group, as auditory preference increases, then PST scores decrease. The new experimental group seems to have remedied that problem. The scores are roughly equal across the auditory preference scale.

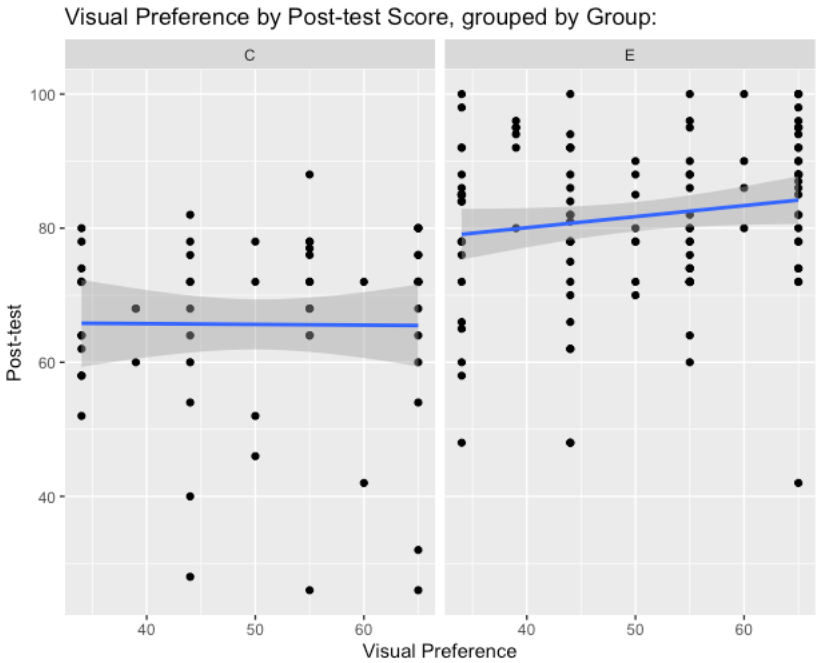
2. T and PST are correlated, and group does change the effect of T on PST. T stands for tactile, and is a measure of preference towards tactile learning methods. The slight positive correlation is shown below:



The interaction is also significant; it appears that potentially the new experimental teaching method causes those with tactile learning preferences to do slightly better on the post-test



3. There is an interaction between v and Group. V is a measure of a student's preference towards visual learning methods.



It looks like visual preference's slope increases slightly in the experimental category, while the mean is definitely higher.