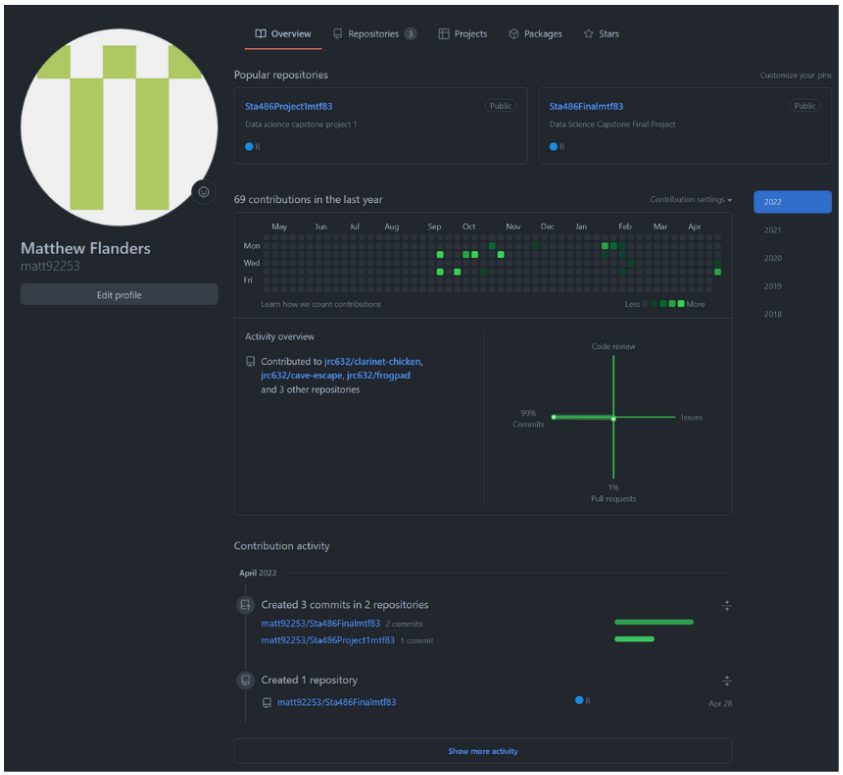
NAU Pre-Calculus Data Exploration: Covid, Exam Retakes, and Effects of Remote Learning

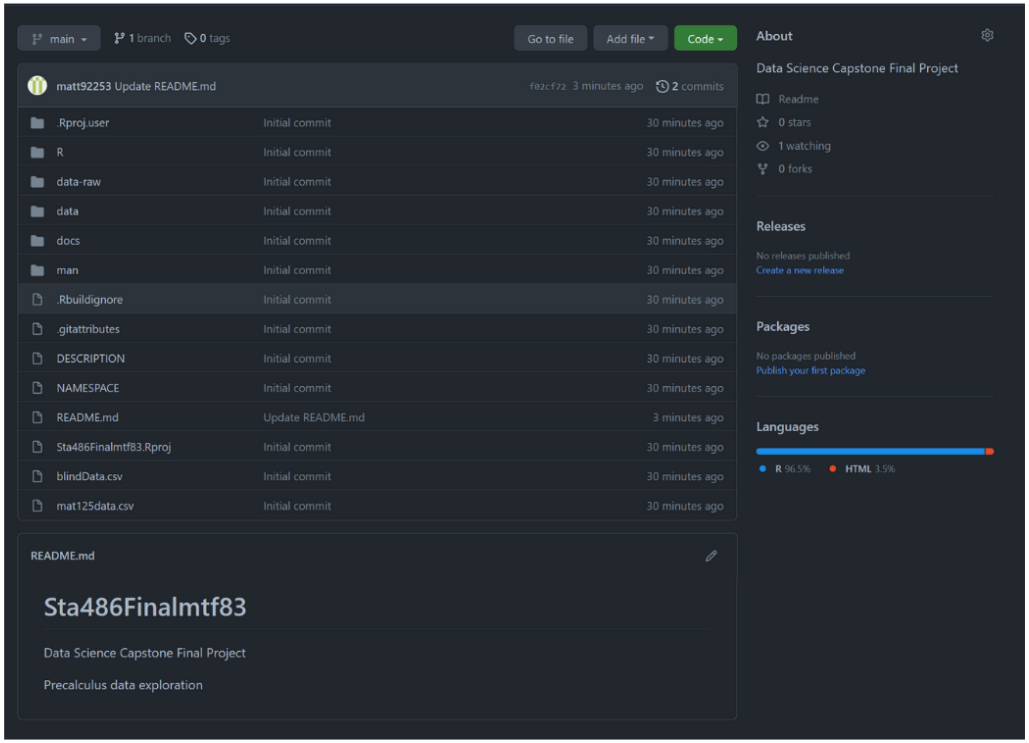
Matthew Flanders

mtf83@nau.edu

August 15, 2022

Github Repository





https://github.com/matt92253/Sta486Finalmtf83

Data Information

- Files:
  - 1 large file for Fall 2016 to Spring 2020
  - 19 additional files for Fall 2020 and Spring 2021
    - 9 files for Fall 2020
    - 10 files for Spring 2021
- Counts:
  - 5 years
  - 11 semesters
  - 15 instructors
  - 138 sections
  - 7,850 students
  - 95,760 tests scores
  - Many more quizzes and home works
- Additional data collected from NAU grade distribution web site.
- FERPA compliance training

Data Cleaning

Steps taken to prepare data:

1. Loading CSV files
2. Removing of N/A entries
3. Add missing information, such as section ids and professors
4. Creating new semester and year variables from dates
5. Creating new indicator variables for different module test types: honors code, learning aids, practice tests
6. String extractions for:
   - section ids
   - student information
   - professor information
7. Regular expressions for typo corrections and standardizing test names
8. Table pivots to match file formats
9. Binding of all files into one
10. Blinding all personal information for FERPA

Background Information

- From Fall of 2016 to Spring of 2018 students were allowed to take multiple attempts on tests to improve scores. Since Spring of 2018 students were only allowed one attempt on a test.

- Halfway through the Spring semester of 2020 NAU moved to NAU Flex due to Covid-19. In the Fall of 2021 students began to return to in person classes.

Problem Statement

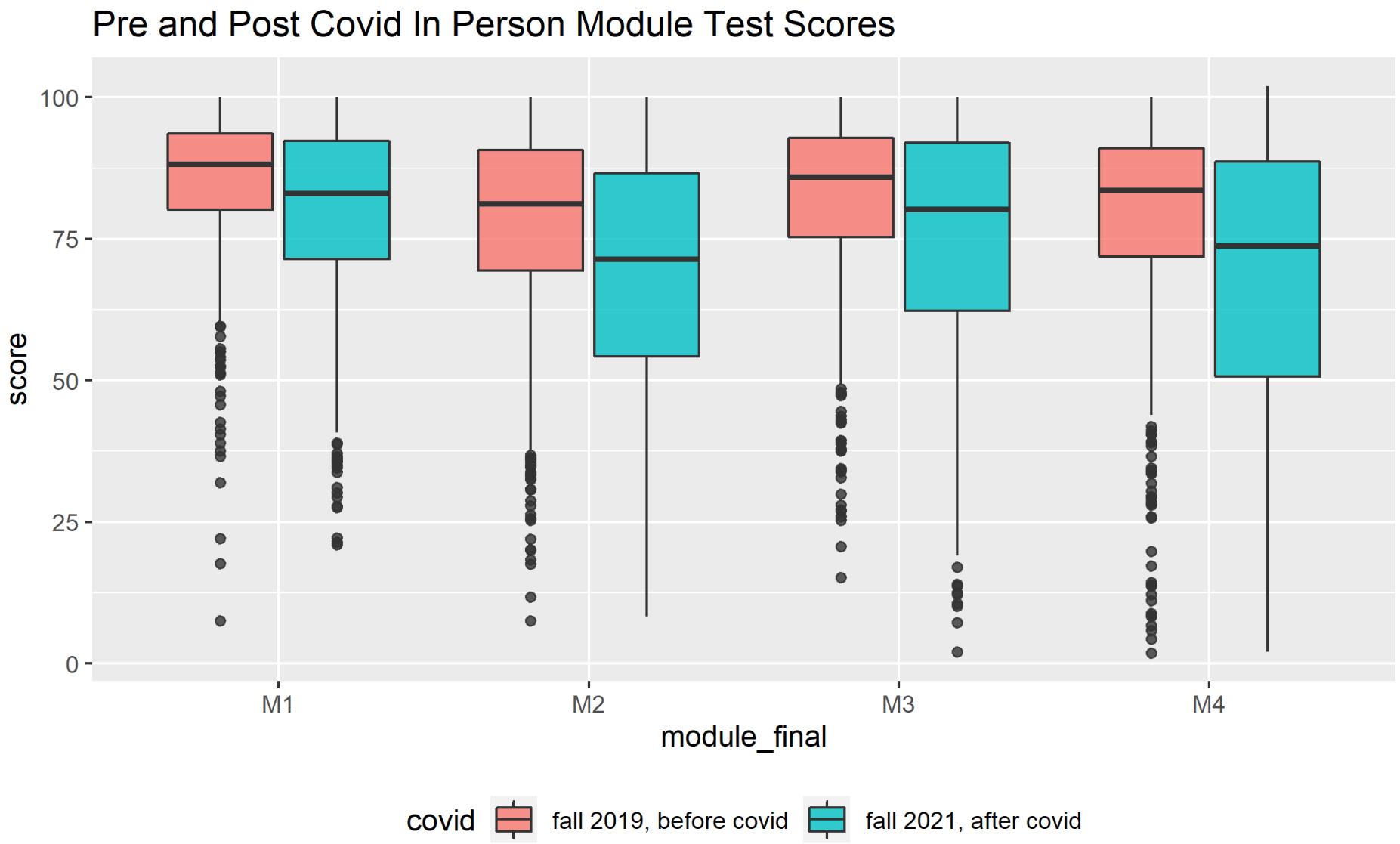
1. What were the effects of Covid-19 and NAU Flex on student test scores?
2. How did student test scores change during NAU Flex?
3. Do students score higher on tests when multiple attempts are allowed vs a single attempt?

How did Covid-19 effect test scores?

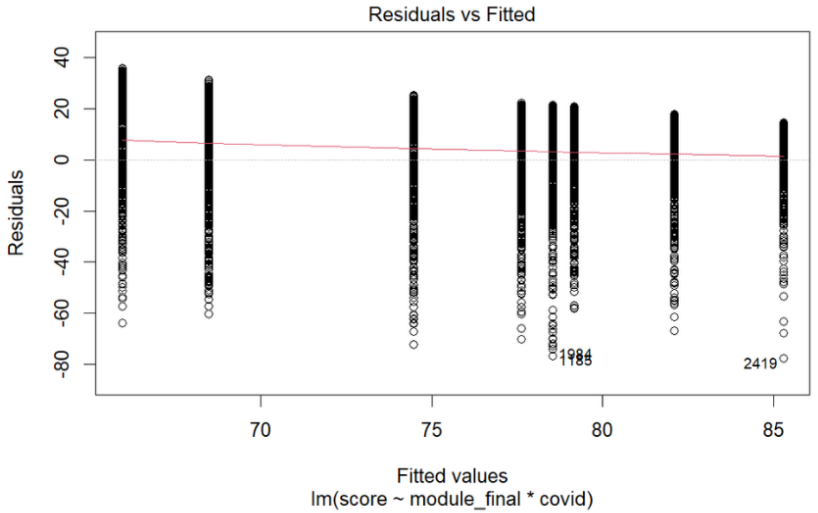
Did Covid-19 and NAU Flex have and effect on in person student test scores, and if so, how did it effect test scores?

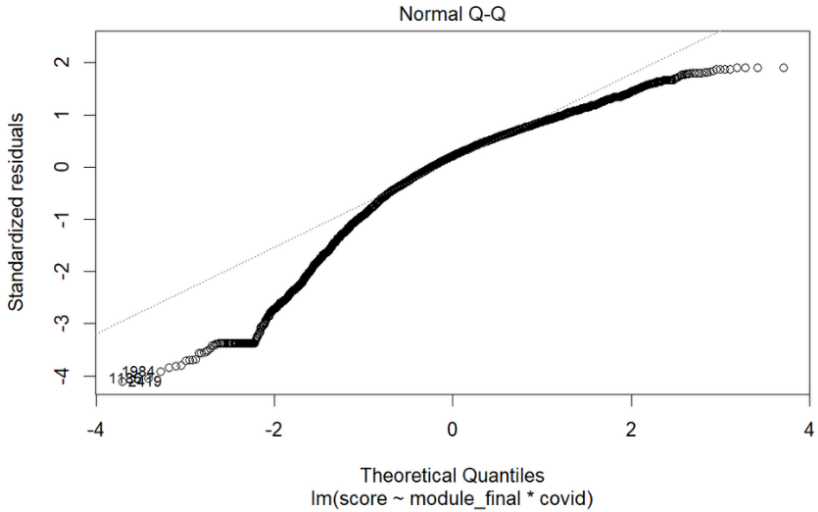
To answer this we will be performing visual analysis of the data, a two way ANOVA test with interaction, and contrasts.

Effects of Covid-19 visual



Effects of Covid-19 ANOVA Analysis






```
Analysis of Variance Table

Response: score
                  Df  Sum Sq Mean Sq  F value    Pr(>F)
module_final       3   76885   25628  71.3493 < 2.2e-16 ***
covid              1   93318   93318 259.7995 < 2.2e-16 ***
module_final:covid 3    6845    2282   6.3521 0.0002707 ***
Residuals       4833 1735981     359
---
Signif. codes:  0 ‘***’ 0.001 ‘**’ 0.01 ‘*’ 0.05 ‘.’ 0.1 ‘ ’ 1

Call:
lm(formula = score ~ module_final * covid, data = pre_post)

Residuals:
    Min      1Q  Median      3Q     Max
-77.796  -8.273   4.119  12.871  36.059

Coefficients:
                                                Estimate Std. Error t value Pr(>|t|)
(Intercept)                                      85.2959     0.7133 119.582  < 2e-16 ***
module_finalM2                                   -7.6669     1.0091  -7.598 3.59e-14 ***
module_finalM3                                   -3.1951     1.0218  -3.127  0.00178 **
module_finalM4                                   -6.7481     1.0302  -6.550 6.35e-11 ***
covidfall 2021, after covid                      -6.1225     1.0864  -5.636 1.84e-08 ***
module_finalM2:covidfall 2021, after covid       -3.0275     1.5399  -1.966  0.04935 *
module_finalM3:covidfall 2021, after covid       -1.5189     1.5534  -0.978  0.32822
module_finalM4:covidfall 2021, after covid       -6.4838     1.5518  -4.178 2.99e-05 ***
---
Signif. codes:  0 ‘***’ 0.001 ‘**’ 0.01 ‘*’ 0.05 ‘.’ 0.1 ‘ ’ 1

Residual standard error: 18.95 on 4833 degrees of freedom
Multiple R-squared:  0.09255,   Adjusted R-squared:  0.09123
F-statistic: 70.41 on 7 and 4833 DF,  p-value: < 2.2e-16
```

Effects of Covid-19 Contrasts Analysis

```
$emmeans
 covid                    emmean    SE   df lower.CL upper.CL
 fall 2019, before covid    80.9 0.363 4833     80.2     81.6
 fall 2021, after covid     72.0 0.413 4833     71.2     72.8

Results are averaged over the levels of: module_final
Confidence level used: 0.95

$contrasts
 contrast                                          estimate   SE   df t.ratio p.value
 fall 2019, before covid - fall 2021, after covid     8.88 0.55 4833  16.159  <.0001

Results are averaged over the levels of: module_final
```

```
$emmeans
 covid                    module_final emmean    SE   df lower.CL upper.CL
 fall 2019, before covid M1             85.3 0.713 4833     83.9     86.7
 fall 2021, after covid  M1             79.2 0.819 4833     77.6     80.8

Confidence level used: 0.95

$contrasts
 contrast                                              estimate   SE   df t.ratio p.value
 fall 2019, before covid M1 - fall 2021, after covid M1     6.12 1.09 4833   5.636  <.0001
```

```
$emmeans
 covid                    module_final emmean    SE   df lower.CL upper.CL
 fall 2019, before covid M2             77.6 0.714 4833     76.2     79.0
 fall 2021, after covid  M2             68.5 0.826 4833     66.9     70.1

Confidence level used: 0.95

$contrasts
 contrast                                              estimate   SE   df t.ratio p.value
 fall 2019, before covid M2 - fall 2021, after covid M2     9.15 1.09 4833   8.384  <.0001
```

```
$emmeans
 covid                    module_final emmean    SE   df lower.CL upper.CL
 fall 2019, before covid M3             82.1 0.732 4833     80.7     83.5
 fall 2021, after covid  M3             74.5 0.835 4833     72.8     76.1

Confidence level used: 0.95

$contrasts
 contrast                                              estimate   SE   df t.ratio p.value
 fall 2019, before covid M3 - fall 2021, after covid M3     7.64 1.11 4833   6.882  <.0001
```

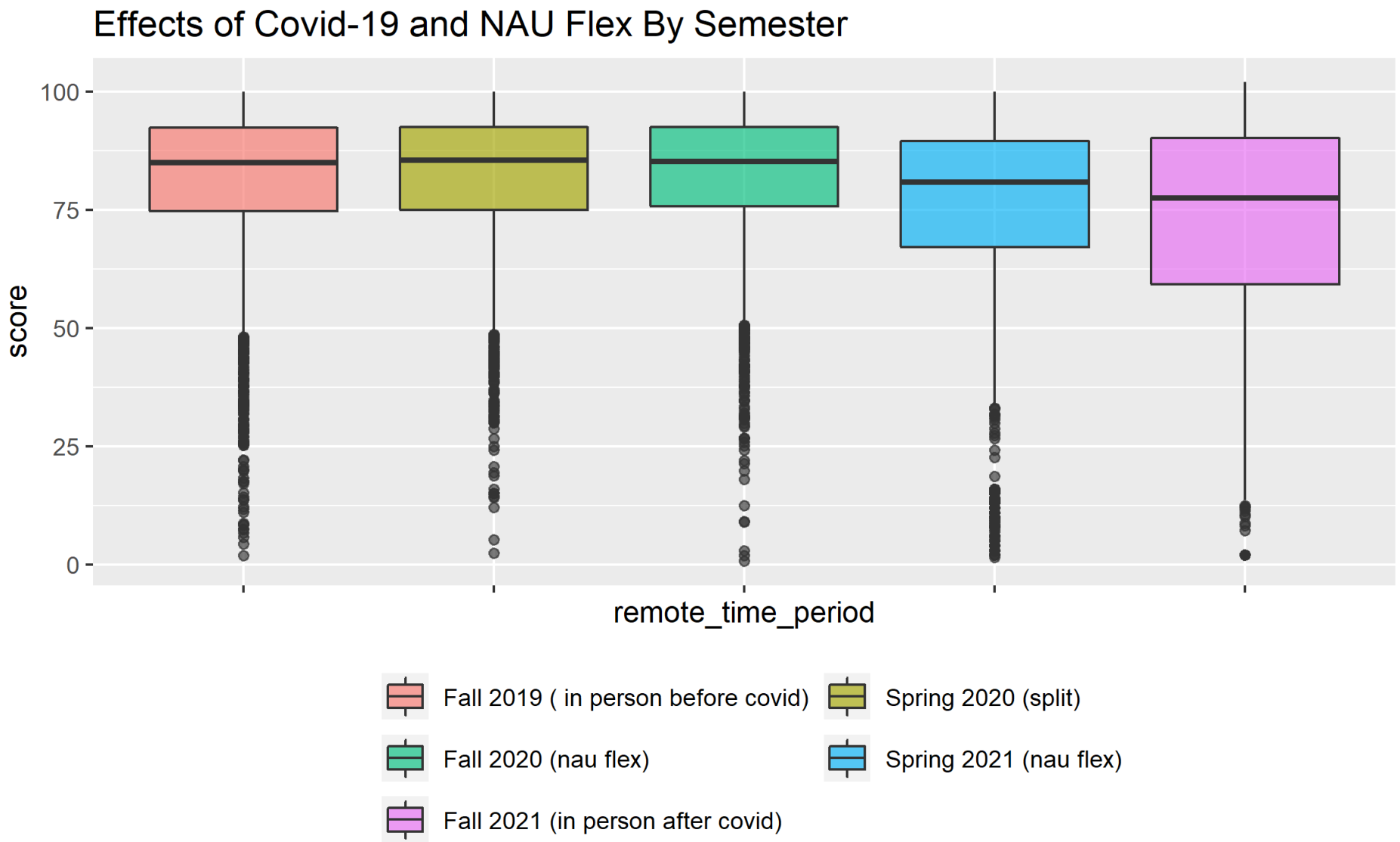
```
$emmeans
 covid                    module_final emmean    SE   df lower.CL upper.CL
 fall 2019, before covid M4             78.5 0.743 4833     77.1     80.0
 fall 2021, after covid  M4             65.9 0.822 4833     64.3     67.6

Confidence level used: 0.95

$contrasts
 contrast                                              estimate   SE   df t.ratio p.value
 fall 2019, before covid M4 - fall 2021, after covid M4     12.6 1.11 4833  11.377  <.0001
```

How did student test scores change over time

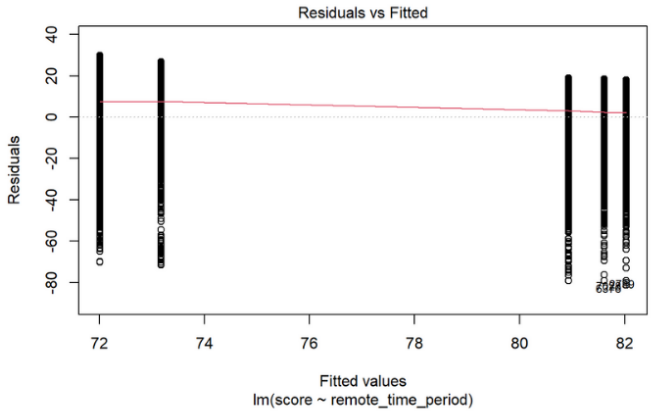
How did student test scores change over time from learning in person before Covid, to NAU Flex, and then back to in person?

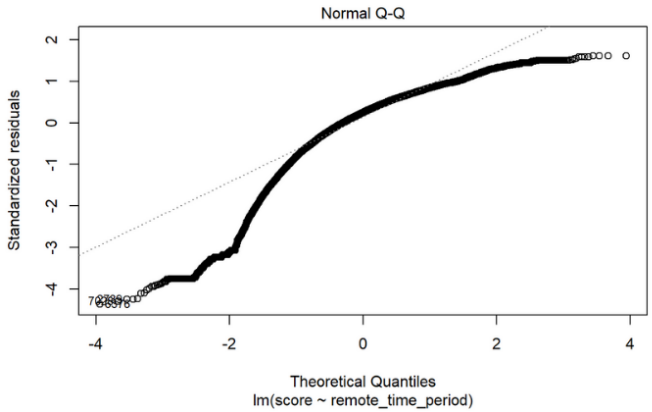
To answer this we will be performing visual analysis of the data, a one way ANOVA test, and contrasts.

Changes in Test Scores Over Time Visual



Changes in Test Scores ANOVA and Contrasts






```
Call:
lm(formula = score ~ remote_time_period, data = remote_learning)

Residuals:
    Min      1Q  Median      3Q     Max
-81.325  -7.183   4.555  12.515  29.988

Coefficients:
                                                Estimate Std. Error t value Pr(>|t|)
(Intercept)                                      80.9271     0.3570 226.697   <2e-16 ***
remote_time_periodSpring 2020 (split)             0.6848     0.5115   1.339   0.1807
remote_time_periodFall 2020 (nau flex)            1.0975     0.5140   2.135   0.0328 *
remote_time_periodSpring 2021 (nau flex)         -7.7474     0.5283 -14.665   <2e-16 ***
remote_time_periodFall 2021 (in person after covid) -8.9150     0.5409 -16.483   <2e-16 ***
---
Signif. codes:  0 ‘***’ 0.001 ‘**’ 0.01 ‘*’ 0.05 ‘.’ 0.1 ‘ ’ 1

Residual standard error: 18.66 on 12271 degrees of freedom
Multiple R-squared:  0.05028,   Adjusted R-squared:  0.04997
F-statistic: 162.4 on 4 and 12271 DF,  p-value: < 2.2e-16

Analysis of Variance Table

Response: score
                    Df  Sum Sq Mean Sq F value    Pr(>F)
remote_time_period     4  226190   56547  162.42 < 2.2e-16 ***
Residuals          12271 4272249     348
---
Signif. codes:  0 ‘***’ 0.001 ‘**’ 0.01 ‘*’ 0.05 ‘.’ 0.1 ‘ ’ 1


$emmeans
 remote_time_period                 emmean    SE    df lower.CL upper.CL
 Fall 2019 ( in person before covid)   80.9 0.357 12271     80.2     81.6
 Spring 2020 (split)                   81.6 0.366 12271     80.9     82.3
 Fall 2020 (nau flex)                  82.0 0.370 12271     81.3     82.7
 Spring 2021 (nau flex)                73.2 0.389 12271     72.4     73.9
 Fall 2021 (in person after covid)     72.0 0.406 12271     71.2     72.8

Confidence level used: 0.95

$contrasts
 contrast                                                          estimate    SE    df t.ratio p.value
 Fall 2019 ( in person before covid) - Spring 2020 (split)           -0.685 0.512 12271  -1.339  0.6669
 Fall 2019 ( in person before covid) - Fall 2020 (nau flex)          -1.097 0.514 12271  -2.135  0.2053
 Fall 2019 ( in person before covid) - Spring 2021 (nau flex)         7.747 0.528 12271  14.665  <.0001
 Fall 2019 ( in person before covid) - Fall 2021 (in person after covid)  8.915 0.541 12271  16.483  <.0001
 Spring 2020 (split) - Fall 2020 (nau flex)                          -0.413 0.521 12271  -0.793  0.9328
 Spring 2020 (split) - Spring 2021 (nau flex)                         8.432 0.535 12271  15.771  <.0001
 Spring 2020 (split) - Fall 2021 (in person after covid)              9.600 0.547 12271  17.547  <.0001
 Fall 2020 (nau flex) - Spring 2021 (nau flex)                        8.845 0.537 12271  16.469  <.0001
 Fall 2020 (nau flex) - Fall 2021 (in person after covid)            10.012 0.549 12271  18.223  <.0001
 Spring 2021 (nau flex) - Fall 2021 (in person after covid)           1.168 0.563 12271   2.075  0.2311

P value adjustment: tukey method for comparing a family of 5 estimates
```

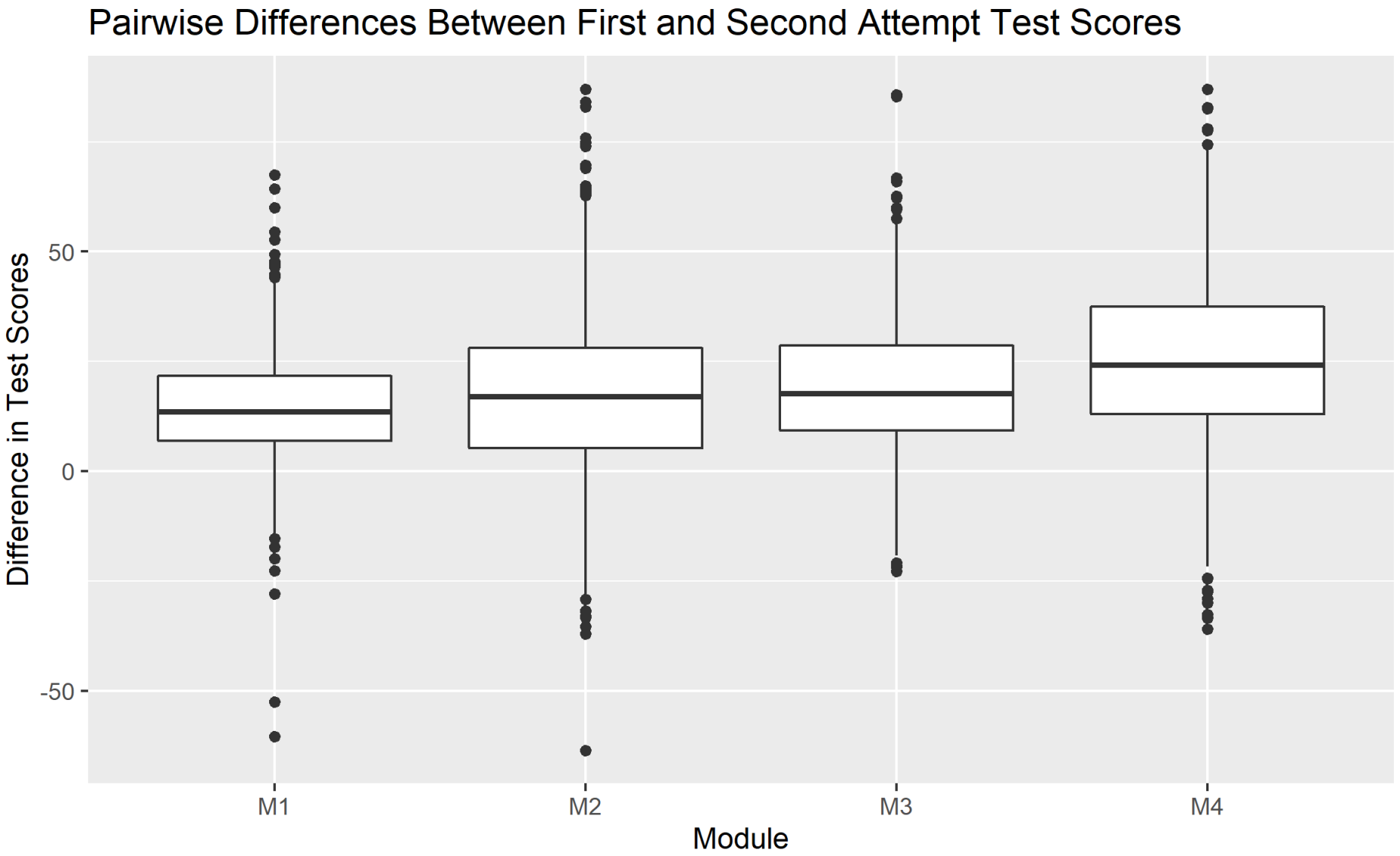
Second Attempts and Score Improvement

From Fall of 2016 to Spring of 2018 students were allowed to take multiple attempts on tests to improve scores. Since Spring of 2018 students were only allowed one attempt on a test.

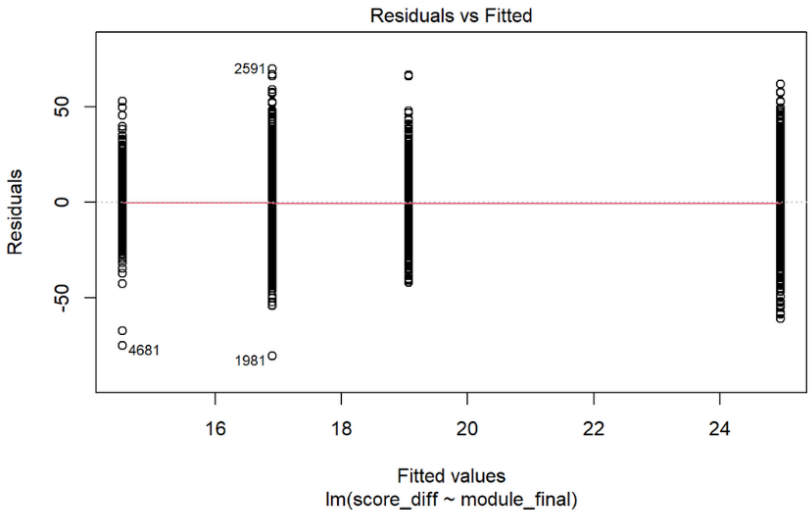
When second attempts were allowed did test scores improve, and if so by how much?

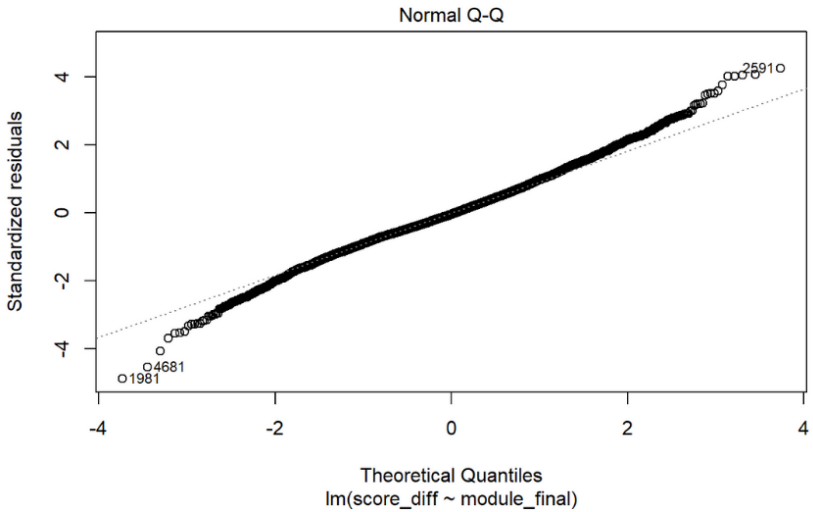
To answer this we will be preforming visual analysis of the data, a one way ANOVA test, and contrasts.

Changes in Test Scores Visual



Changes in Test Scores ANOVA and Contrasts






```
Call:
lm(formula = score_diff ~ module_final, data = pairwise_comp)

Residuals:
    Min      1Q  Median      3Q     Max
-80.504 -10.288  -0.816  10.045  70.096

Coefficients:
              Estimate Std. Error t value Pr(>|t|)
(Intercept)     14.5307     0.5083  28.586  < 2e-16 ***
module_finalM2   2.3734     0.6582   3.606 0.000314 ***
module_finalM3   4.5365     0.6957   6.521 7.66e-11 ***
module_finalM4  10.4239     0.6677  15.611  < 2e-16 ***
---
Signif. codes:  0 ‘***’ 0.001 ‘**’ 0.01 ‘*’ 0.05 ‘.’ 0.1 ‘ ’ 1

Residual standard error: 16.51 on 5272 degrees of freedom
Multiple R-squared:  0.05234,   Adjusted R-squared:  0.0518
F-statistic: 97.06 on 3 and 5272 DF,  p-value: < 2.2e-16

Analysis of Variance Table

Response: score_diff
              Df  Sum Sq Mean Sq F value    Pr(>F)
module_final    3   79374 26457.9  97.061 < 2.2e-16 ***
Residuals    5272 1437089   272.6
---
Signif. codes:  0 ‘***’ 0.001 ‘**’ 0.01 ‘*’ 0.05 ‘.’ 0.1 ‘ ’ 1


$emmeans
 module_final emmean    SE   df lower.CL upper.CL
 M1             14.5 0.508 5272     13.5     15.5
 M2             16.9 0.418 5272     16.1     17.7
 M3             19.1 0.475 5272     18.1     20.0
 M4             25.0 0.433 5272     24.1     25.8

Confidence level used: 0.95

$contrasts
 contrast estimate    SE   df t.ratio p.value
 M1 - M2     -2.37 0.658 5272  -3.606  0.0018
 M1 - M3     -4.54 0.696 5272  -6.521  <.0001
 M1 - M4    -10.42 0.668 5272 -15.611  <.0001
 M2 - M3     -2.16 0.633 5272  -3.418  0.0036
 M2 - M4     -8.05 0.602 5272 -13.374  <.0001
 M3 - M4     -5.89 0.643 5272  -9.160  <.0001

P value adjustment: tukey method for comparing a family of 4 estimates
```

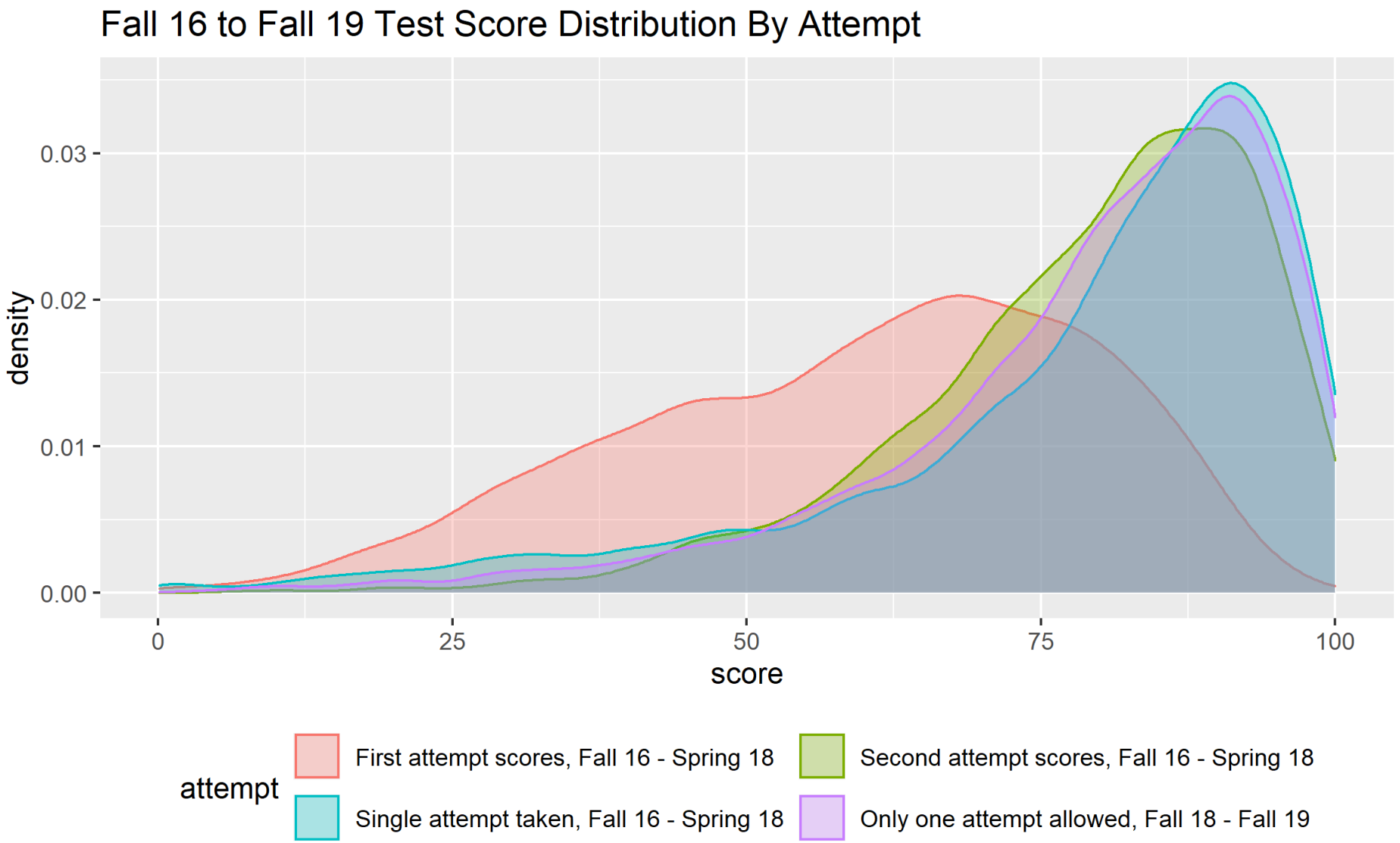
Further Analysis of Second Attempts

From the previous slides we did find that students on average did improve on their first attempt scores when they were allowed to take a second attempt.

When bringing these results to Ellie Blair it confirmed her intuition that students scored higher on a second attempt when compared to the first attempt. We were asked to look further into the data and see if allowing multiple attempts on tests improved scores more than if students were only allowed a single attempt.

To answer this we will be preforming visual analysis to look into the different behaviours between single and multiple test attempts.

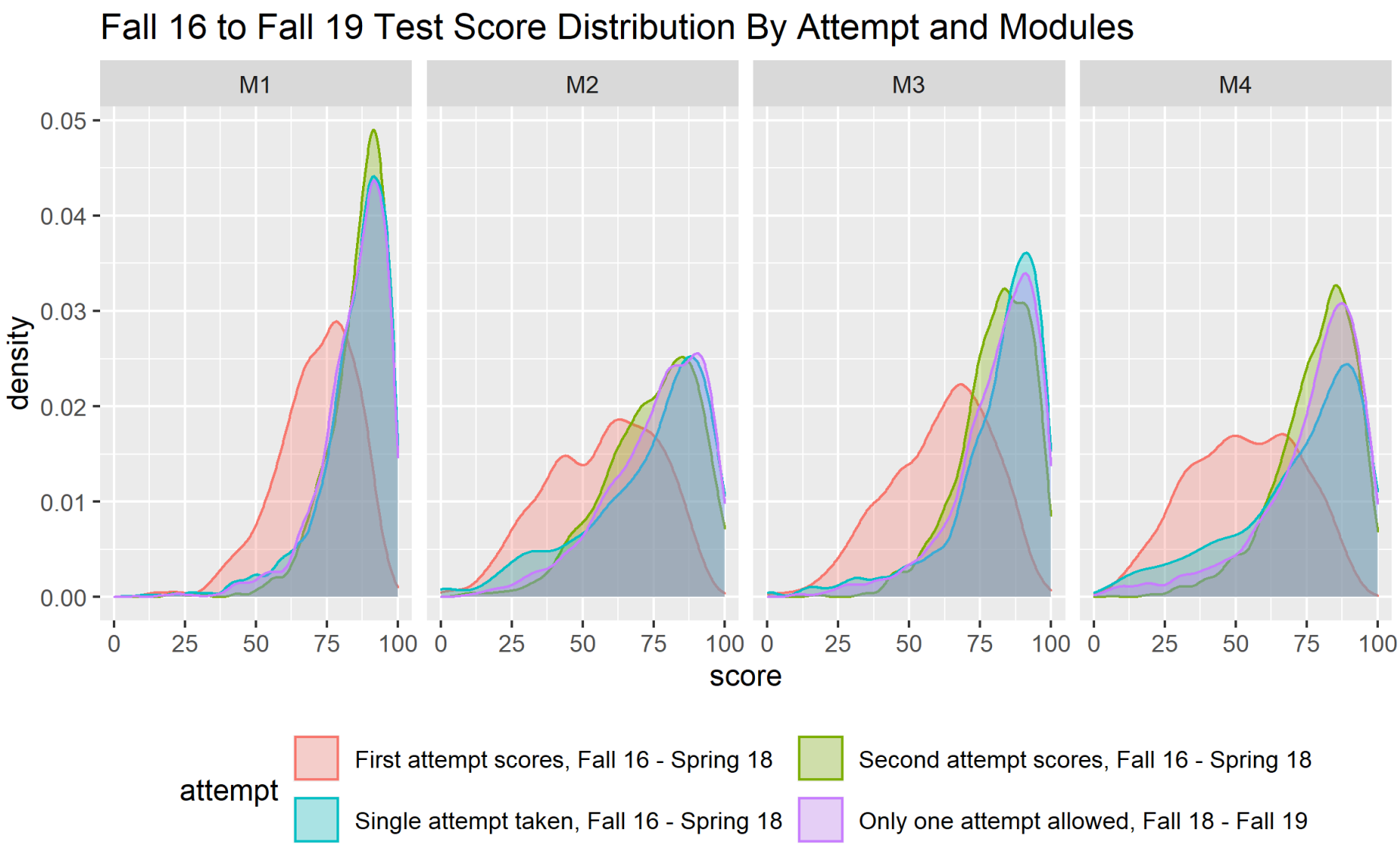
Further Analysis of Second Attempts Visual

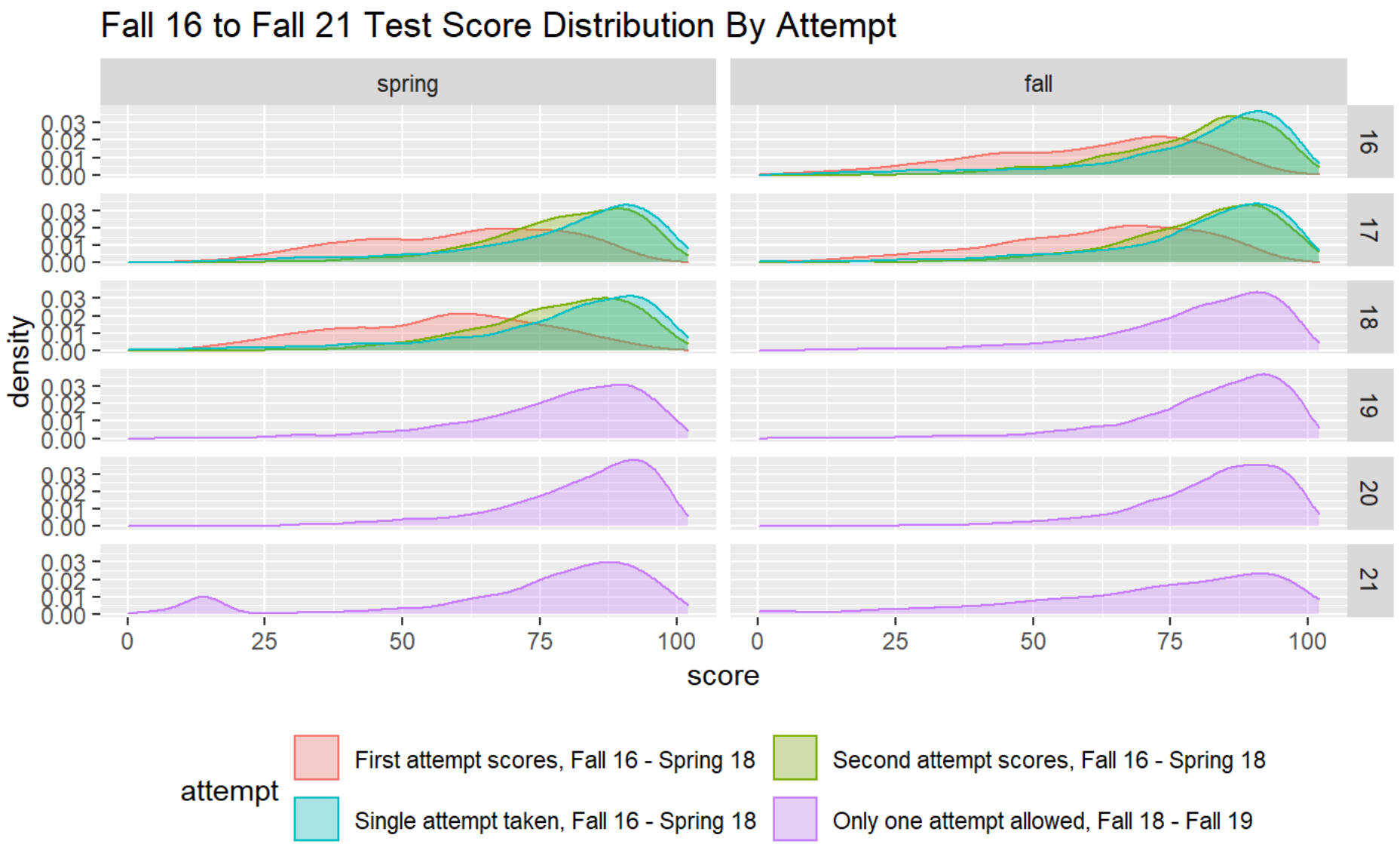


Further Analysis of Second Attempts

From the previous slide we can see that allowing multiple attempts does not improve test scores any more than if only a single attempt was taken or allowed.

Now we wanted to look deeper and see if certain modules benefited from multiple attempts.


Fall 16 to Fall 19 Test Score Distribution By Attempt and Modules
M1
M2
M3
M4
density
0.05
0.04
0.03
0.02
0.01
0.00
0
25
50
75
100
score
attempt
First attempt scores, Fall 16 - Spring 18
Second attempt scores, Fall 16 - Spring 18
Single attempt taken, Fall 16 - Spring 18
Only one attempt allowed, Fall 18 - Fall 19


Fall 16 to Fall 21 Test Score Distribution By Attempt
spring
fall
16
17
18
19
20
21
density
0.03
0.02
0.01
0.00
0
25
50
75
100
score
attempt
First attempt scores, Fall 16 - Spring 18
Second attempt scores, Fall 16 - Spring 18
Single attempt taken, Fall 16 - Spring 18
Only one attempt allowed, Fall 18 - Fall 19

1. What were the effects of Covid-19 on student test scores?
   We have seen that in person test scores have dropped since remote learning.

2. How did student test scores change during remote learning?
   We have seen scores improve slightly 2020 and then fall in 2021.

3. Does offering a second attempt improve test scores enough to make giving a second attempt worth the effort
   Over all module test given it does not seem that multiple attempts on test improve scores.

- Dashboard application done in R shiny
- Include homework or quizzes to measure their effectiveness
- What are some of the differences between instructors
- Are there differences between sections that instructor teach
- How do graduate teaching assistants compare to professors
- Other courses

Thank you,
Questions?

mtf83@nau.edu