

**Comparative Analysis of Students' Performance  
Between Online and on Campus in an Introductory  
Statistics Course**



**Kendal McDonald**

**Mentor: Jebessa Mijena**

**Department of Mathematics**

**Georgia College and State University 2017**

# Contents

<b>Acknowledgement.....</b>	<b>2</b>
<b>Abstract .....</b>	<b>3</b>
<b>Introduction.....</b>	<b>4</b>
<b>Methods.....</b>	<b>5</b>
<b>Exploratory Analysis.....</b>	<b>13</b>
<b>Future Research.....</b>	<b>29</b>
<b>References.....</b>	<b>30</b>

# **Acknowledgement**

This research was conducted under the approval of the Georgia College and State University Institutional Review Board. Also, a thank you to Jebessa Mijena for allowing me to conduct this research and mentoring me in approving my mathematical research and mathematical knowledge in the subject of statistics, probability and data science.

# Abstract

In this research, we compare students' performance in an online and on campus introductory statistics and probability course. MyStatLab is the learning management system used in both an online and on campus courses for homework and quizzes. The online data is produced by five summer courses between Summer 2014 to Summer 2017 and the on-campus data is produced from nine on campus courses from Spring 2014, Spring 2016, and Spring 2017. For homework, the research compares the scores made, and how early a student completed the homework in the online and on campus courses. For quizzes, we tested if the scores are same, if there is a difference in how early a student completed the quiz, and the number of attempts taken out of the five attempts granted in the online and on campus courses. We also analyzed the difference between the first attempt score and highest score to see if there is significant improvement in scores by taking the assignments again. In addition, we also modeled the final quiz average as a function of number of attempts and the number of days a student attempted the quiz before the due date which we found are correlated with the final quiz average. Also by using the Wilcoxon Rank Sum Test, we showed that there is no significant difference between online and on campus for both first attempt quiz scores and final quiz scores. However, for homework scores there was a significant difference between online and on campus.

# Introduction

Comparing on campus and online student performance in an introductory statistics course gives insight on which class would be better for a student to take given the performance of past students. The motivation behind this research is to see if there is a correlation between a student's performance and the number of days before the due date the student starts attempting the quiz or homework. Also for quizzes, this research will see if the correlation between a student's performance and the total number of attempts a student takes. Through linear regression models, the goal is to predict the final quiz average from a student's homework average or the student's first attempt quiz average. Also with linear regression models, the second goal is to predict the final quiz grade from the student's total number of attempts taken and the days before the quiz due date. The exploratory analysis of correlation and the linear regression models were all conducted in RStudio version 1.0.316. A student's performance was based on the work the student does in MyStatLab for both online and on campus. The online data was comprised of five sections which were from the semesters of summer 2014, summer 2015, summer 2016, and summer 2017 with summer 2017 having two sections during a semester. For on campus data, the data was comprised of nine sections from the semesters of spring 2014, spring 2016, and spring 2017 with each semester having three sections.

## Methods

Through MyStatLab, both on campus and online sections were able to take all quizzes and homework assignments online at their own convenience. For the online group consisted of five sections from the summer semester of the years 2014, 2015, 2016, and 2017. Each of the years had one section except for 2017 which had two sections. Each summer course was taken over eight conservative weeks. Summer 2014 section had eighteen students who took this class with no students withdrawing from the class; however, for homework and quizzes one student did not attempt majority of the assignments which was ignored in the data. From the excel sheet downloaded from MyStatLab line 14 will need to be deleted from the homework and quiz excel sheets. The summer 2015 section had twenty-two students with no one dropping the course, but one student did not complete a sufficient amount of the homework assignments or the quizzes. Therefore, to clean the data from summer 2015 the deletion of line 8 on the homework and quiz excel sheets were done. Summer 2016 section had twenty-four students with one student withdrawing from the course before the end of the eight-week course was over. Thus, the withdrawing student was the only student to not complete majority of the work which was line 10 on the homework and quiz from the excel sheets from MyStatLab which was deleted. For summer 2017, there were two sections the first section, labeled W04, was taken over the course of May and June while the second section, labeled W07, was taken over the course of June and July. The first section of summer 2017 had twenty-four students and the second section had twenty-six students. Between both sections of summer 2017 all students

completed majority of the homework and quiz assignments that no data clean up in these sections were needed.

On campus data was comprised of nine sections from spring 2014, spring 2016, and spring 2017. Each semester had three sections with all sections covering a span of sixteen weeks. For spring 2014, the first and second section, labeled section 11 and section 12 respectively, each had thirty-five students and the third section, labeled section 13, had thirty-three students. To clean the data for spring 2014 section 11, the deletion of lines 24, 34, and 37 are necessary for both homework and quiz excel sheets. Section 12 of spring 2014, for data cleaning the deletion of lines 6, 10, 30, 35, and 36 from the excel sheets for homework and quizzes. For section 13 of spring 2014, removing lines 10 and 28 will clean up the data since these two students either withdrew from the class or did not complete majority of the work after a certain period in the course. For spring 2016, the first section, labeled 2, had a total of thirty-three students and the second and the third sections, labeled 3 and 4 respectively, had thirty-seven students in the course. Section 2 of the semester spring 2016, only two students did not complete sufficient amount of the course work which would be students correlating to lines 13 and 43 thus these lines should be deleted. For section 3 for the semester spring 2016, lines 21 and 25 need to be deleted to clean the data since these students did not finish majority of the course work. For the last section, section 4, of spring 2016 there was only one student who did not complete the majority of the assignments which is student signified by line 24 was deleted for data cleaning. The next semester the on-campus data is comprised of is spring 2017 with first and third sections, labeled 2 and 13 respectively, have thirty students and the third section, labeled 3, has thirty-one students. To clean the data for section 2 of spring 2017 the

removal of line 31 on the excel sheets for both homework and quizzes since the student did not complete a sufficient amount of the course work. For spring 2017 section 3, the removal of lines 25 and 29 are need for cleaning of the data. For the last section, section 13, of spring 2017, to clean the data the removal of lines 8 and 15.

With these data clean ups completed, the next step in getting the data information needed is by creating columns for the quizzes in the order of days before due date, number of attempts, highest score on the quiz, and the attempt the highest score was achieved after each quiz. The only column needed after each homework should be labeled as days before due date. After the final quiz and final homework, for quiz the addition of first attempt quiz average and final quiz average should be added and for homework only final homework average should be added. The first attempt quiz average is comprised of all the first attempt quiz score a student makes on the quiz then minus the two lowest grades made on a quiz divided by four for online sections and seven for on campus sections except for the semester of spring 2016 which will be six instead of seven. The final quiz average is calculated the same way but instead of the first attempt scores being used the highest score on the quiz is used in the same method. The homework final average is calculated the same way with the two lowest homework grades are dropped then divided by seven except for spring 2016 semester for on campus and also by seven for online.

The days before due date column is generate by looking at the first attempt of the quiz and determining how many days before the due date. If a student starts their first attempt on the quiz the day the assignment is due, in the column of days before due date the student would get a zero. Also created an extra column by the first attempt on each quiz with the due

date of the assignment so in the days before the due date the subtraction of the due date column minus the first attempt day. From the data received from MyStatLab, may have an hour time instead of just the date the hour time will need to be removed for the subtraction method to work. Also, if a student did not attempt the assignment then there should be no data input in the days before due date column.

The next column beside the days before due date is the number of attempts. The number of attempts column is calculated by looking at the total number of attempts a student takes a certain quiz. If a student did not attempt the quiz, the student does not receive a zero and the box in the column should just remain blank. Also, if a number is negative which means the assignment was completed after due date either delete it or remember to ignore the negatives when completing the analysis.

The following column after number of attempts is the highest score. The highest score is when the student made the highest grade on the quiz. With the highest score column, all is needed is inputting the highest score the student made. Once again, if a student did not attempt the quiz no grade would be given and the box would be left blank in the column.

Next column is the highest score attempt. This column signifies the attempt the student achieved the highest score on the quiz. If a student made a hundred on their fourth attempt, the number recorded in the column would be four since it was on the fourth attempt the student made their highest score. Also, a student who does not attempt a quiz shall not receive a number in the box thus the box would remain empty.

For finding the first attempt quiz average, final quiz average and homework average, if a student did not attempt the quiz a zero must be given. The best way to find the averages would

be copying the columns first score attempt, the first column under the quiz number, and the highest score columns then pasting them in a new excel sheet. Once the data is put into the new excel sheet, for every blank put a zero. The reason behind putting the zeros in is because if there were no zeros then an error would occur in the calculation of the data. Once all zeros are in place, calculating the first attempt quiz average is done by filling in the column with the equation  $=((\text{sum}(\text{all the columns relating to first attempt quiz average})-\text{small}(\text{same columns},1)-\text{small}(\text{same columns},2)))/7$  for on campus course except for the spring 2016 semesters which would be divide by six and for online semesters would be divided by four. The same equation layout will be used for the final quiz average which is based off of the highest score a student makes on a quiz except the columns for the sum and small parts of the equation would now be with the highest scores columns. The homework average is calculated the same way as either quiz grades however for online it will be divided by seven instead of four and spring 2016 will still be divided by six.

Once all the data is filled in, the combining of all sections will be combined for both online and on campus with quiz and homework being in the same excel sheet. For online data, label the top row with the same variables: first attempt score, days before due date for quiz, number of attempts, highest score, highest score attempt, first attempt quiz average, final quiz average, homework score, days before for homework, and homework quiz average. Once this is labeled, a copy of each semesters first quiz data columns correlating to the columns on the combined excel sheet is put into the correct order. The first quiz from each semester covers lines 2 through 111, quiz two covers lines 112 through 221, quiz three covers lines 222-331, quiz four covers lines 332 through 441, quiz five covers lines 442 through 551, and quiz six

covers lines 552 through 661. The first attempt quiz average, final quiz average, and homework average will have the ranges based on the semester and section. Summer 2014 averages range from lines 2 through 18, summer 2015 averages range from lines 19 through 39, summer 2016 range from lines 40 through 62, summer 2017 W04 averages range from lines 63 through 86, and summer 2017 W07 averages from lines 87 through 111. The homework scores and days before the due date are based off of the homework assignment. Therefore, homework assignment one ranges from lines 2 through 111, assignment two ranges from 112 through 221, assignment three ranges from lines 222 through 331, assignment four ranges from lines 332 through 441, assignment five ranges from lines 442 through 551, assignment six ranges from lines 552 through 661, assignment seven ranges from lines 662 through 771, assignment eight ranges through 772 through 881, and assignment nine ranges from lines 882 through 991. Now for on campus data is combined using the same headings as the first row of the online combined data. The same steps are taken for on campus as the online data but the row numbers are different. For the quiz data consisting of first score on quiz, days before due date, number of attempts, highest score, and highest score attempt, the first quiz ranges from lines 2 through 285, quiz two ranges from lines 286 through 569, quiz three ranges from lines 570 through 853, quiz four ranges from lines 854 through 1137, quiz five ranges from lines 1138 through 1421, quiz six ranges from lines 1422 through 1705, quiz seven ranges from lines 1706 through 1989, quiz eight ranges from lines 1990 through 2273, and quiz nine ranges from lines 2274 through 2450. The first attempt quiz average, final quiz average, and homework averages range based on semester with the sections in sequential order. Therefore, spring 2014 section 11 ranges from lines 2 through 32, spring 2014 section 12 ranges from lines 33 through 62,

spring 2014 section 13 ranges from lines 63 through 92, spring 2016 section 2 ranges from lines 92 through 128, spring 2016 section 3 ranges from lines 129 through 163, spring 2016 section 4 ranges from lines 164 through 199, spring 2017 section 2 ranges from lines 200 through 228, spring 2017 section 3 ranges from lines 229 through 257, and spring 2017 section 13 ranges from lines 258 through 285. The homework score and days before follow the same line ranges as the quiz data.

Quiz averages, homework average, days before due date, and number of attempts will be separated into individual excel documents for analyzation in R. For quiz averages, the headings will be first attempt quiz average, first attempt class type, final quiz average, final class type. Looking at the combined online data excel document copy first attempt quiz average and paste under the new document then in the next column labeled first attempt class type write online for each of the first attempt quiz average. The following steps continue for both the final quiz average for online. Also, the steps taken for online will be the same for on campus data, will be directly below the online data, but the class type would be on campus. For homework average, the labels for the new excel document will be homework average and class type. The steps taken for homework average is similar to the quiz average by copying online homework average and pasting the data into the column labeled homework average then adding the on-campus homework average below online homework average. For the class type column, put online next to the online homework averages and on campus next to the on-campus homework averages. For days before due date, the labels for the new excel document are quiz days before due date, quiz class type, homework days before due date, and homework class type. The steps to inputting the data in the new document is similar to the previous excel

documents with copying both online and on campus days before for both quiz and homework then adding the class type in the respective columns. The next excel document created is for the number of attempts with the first row being labeled as number of attempts and class type. The steps are exactly like the other excel documents with combining the online and on campus number of attempts then adding the number attempts for online and on campus.

For individual quiz comparisons, the combining of certain quizzes are need before comparing online and on campus quizzes. Since online sections only had six total quizzes, on campus nine or eight quizzes must be composed down to six quizzes. The composition of multiple quizzes was done based on the chapters covered. Start a new excel sheet with the first-row headings being quiz one online, quiz one on campus, quiz two online, quiz two on campus, quiz three online, quiz three on campus, quiz four online, quiz four on campus, quiz five online, quiz five on campus, quiz six online, and quiz six on campus. Quiz one covers chapter one which both online and on campuses only covered on the first quiz. However, for the second quiz online sections' quiz two covers chapters two and three, thus for the on-campus data combining quiz two and three by creating a new excel sheet with quiz two in one column and quiz three in the next column. Then take the average of the two quizzes. After taking the average of the two quizzes, if a student did not take one of the quizzes that students' combined average must be ignored. Once all the incomplete quiz averages are ignored, then add to the excel sheet that already contains quiz one for both online and on campus. However, for the online quiz two, summer 2014 section must be ignored because of the quiz chapters not matching up. Quiz three covers chapters four and five, the same steps were taken for quiz two including ignoring online section summer 2014. However, for quiz four only covering chapter six

and quiz five covering only chapter seven, the only steps need to be taken is adding quiz four and quiz five data for on campus and online with ignoring summer 2014 section is the only steps needed. However, for quiz six chapters eight and nine are covered, so the combining of the two quizzes will take the same steps as quiz two and three except summer 2014 section can be added with the online data since the chapters align.

## Exploratory Analysis

The exploratory analysis consisted of using the Wilcoxon Rank Sum Test which is used due to the outliers produced in the analysis while also due to the lack of normality in the data. The Wilcoxon Rank Sum Test formulas can be seen in figure 1 with the  $W$  standing Wilcoxon statistic for the sum of the  $n$  signed ranks.

$$\mu_W = \frac{n_2(n_1 + n_2 + 1)}{2}$$

$$\text{Var}(W) = \frac{n_1 n_2 (n_1 + n_2 + 1)}{12}$$

$$Z = \frac{W - \frac{n_2(n_1 + n_2 + 1)}{2}}{\sqrt{\frac{n_1 n_2 (n_1 + n_2 + 1)}{12}}}$$

Figure 1: Wilcoxon Rank Sum Test Formulas

The first comparison analyzed is the final quiz average. Comparing the final quiz average, the summaries (figure 2) do not provide any true comparison due to the total number

of students varying for online and on campus. By figure 3, we can see the final quiz averages are very close in average with on campus having a higher final average by 0.33 percent.

	Final Online	Final on Campus
1	Min. : 29.77	Min. : 38.18
2	1st Qu.: 85.11	1st Qu.: 82.03
3	Median : 89.47	Median : 90.42
4	Mean : 87.02	Mean : 87.35
5	3rd Qu.: 94.37	3rd Qu.: 95.14
6	Max. :100.00	Max. :100.00

Figure 2: Summary of Online and On Campus Final Quiz Average

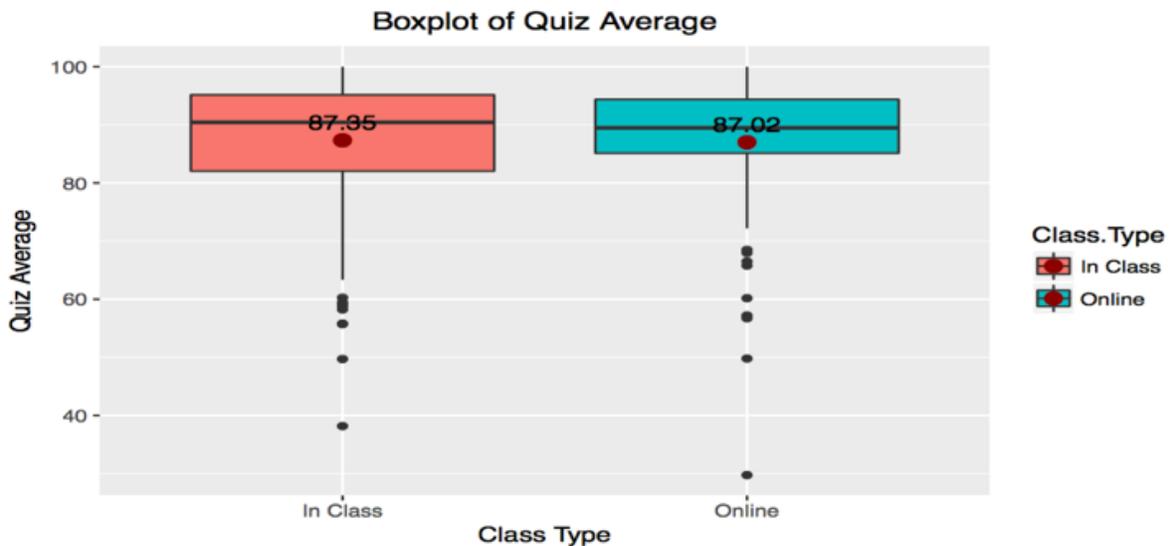


Figure 3: Boxplot of Online and On Campus Final Quiz Average

By the Wilcoxon Rank Sum Test, the calculated p-value is 0.7615 which is greater than 0.05 therefore the final quiz average for online and on campus are not significantly different.

Similarly to the final quiz average, the first attempt quiz average behaved very similar as shown in figure 4 and figure 5. The p-value calculated for the first attempt quiz average was 0.8280 which shows online and on campus are not significantly different.

1st ATT. Online	1st ATT. on Ca
Min. :23.12	Min. :12.87
1st Qu.:57.03	1st Qu.:54.83
Median :67.84	Median :68.28
Mean :65.08	Mean :64.21
3rd Qu.:76.93	3rd Qu.:76.00
Max. :93.42	Max. :97.13

Figure 4: Summary of First Attempt Quiz Average

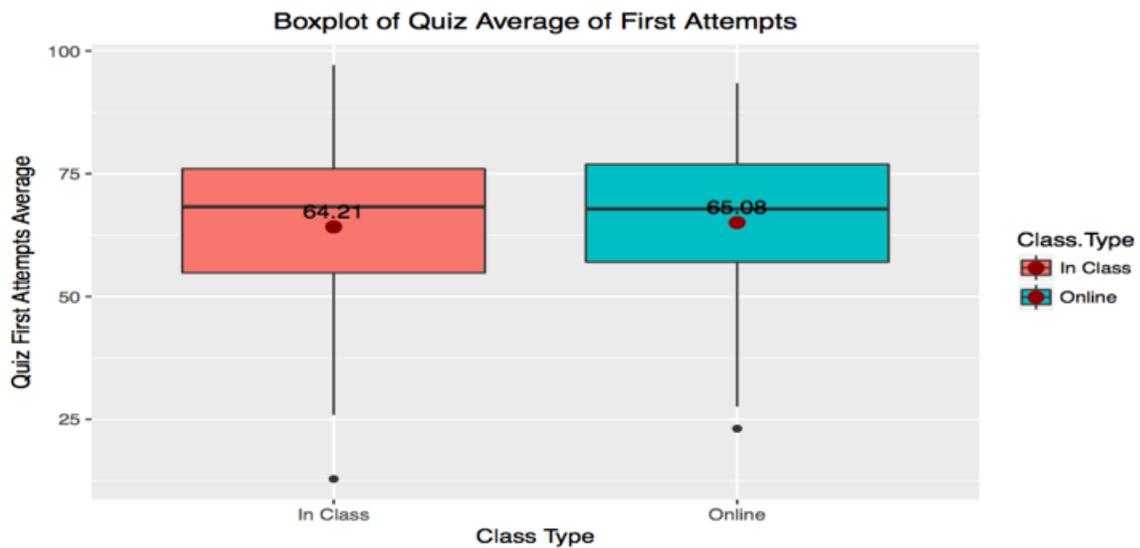


Figure 5: Boxplot for First Attempt Quiz

Differently from the first and final quiz average, the homework averages vary by a difference of 3.19 percent with on campus having the higher average as shown in figure 6. Looking at figure 7, the boxplot shows there are many outliers in both online and on campus. These outliers would cause a significant issue in calculating the p-value if a t-test was used instead of the Wilcoxon Rank Sum Test. The p-value calculated is 0.001404 which is significant less than 0.05 which means the homework averages for online and on campus are significantly different.

	Homework Average Online	Homework Average on Campus
1	Min. : 42.64	Min. : 54.10
2	1st Qu.: 90.54	1st Qu.: 93.91
3	Median : 95.19	Median : 97.42
4	Mean : 91.93	Mean : 95.12
5	3rd Qu.: 98.50	3rd Qu.: 99.54
6	Max. :100.00	Max. :100.00

Figure 6: Summary of Homework Average

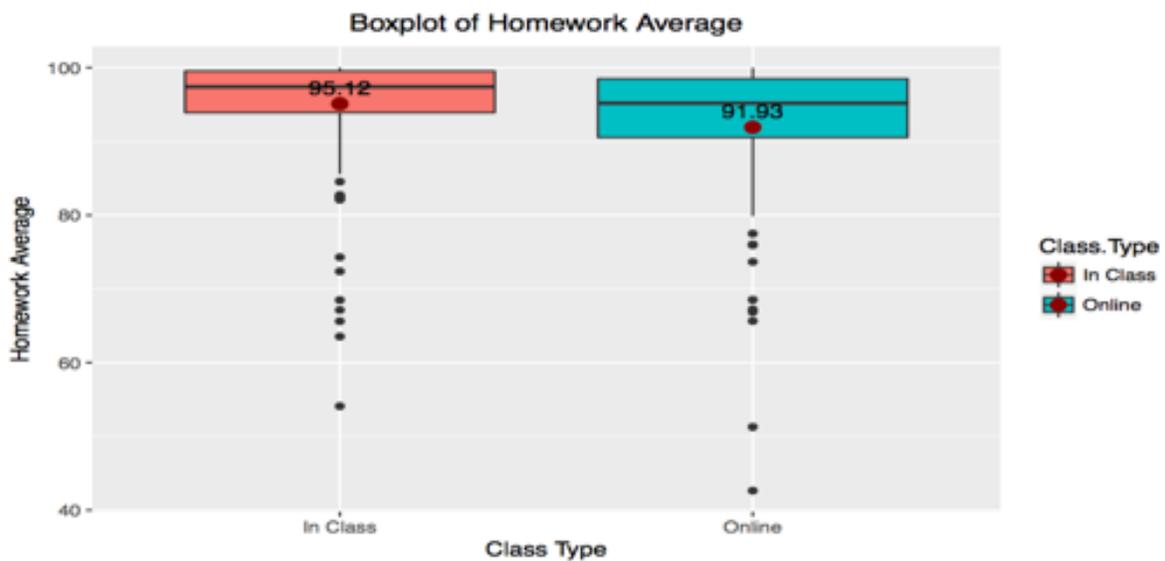


Figure 7: Boxplot of Homework Average

The next exploration was through the days before due date. However, since online is an eight-week course and on campus is a sixteen-week course, this causes on campus to have more days to attempt the quiz or homework. For quizzes, on campus were given a maximum of eleven days while online had a maximum of four days. For homework, on campus were given a maximum of fourteen days while online had a max of five days. This is a data limitation in the comparison aspect of the exploratory analysis.

For number of attempts, looking at figure 8 the number of attempts for quiz and highest score attempts for both environments have around two attempts or figure 9 for only the overall quiz. However, for the p-value calculated for number attempts for a quiz gave a 0.04245 which means the data are significantly different. However, if more data in the future is added to this research than the p-value could possibly change to where the two data collections are proven to be not significantly different. Since highest score attempt was a subpart of the total quiz attempts, the p-value was not calculated but can simply be calculated using the data.

	Quiz online	Quiz on Campus	HS Online	HS on Campus
1	Min. :1.000	Min. :1.000	Min. :1.000	Min. :1.000
2	1st Qu.:2.000	1st Qu.:1.000	1st Qu.:2.000	1st Qu.:1.000
3	Median :2.000	Median :2.000	Median :2.000	Median :2.000
4	Mean :2.243	Mean :2.118	Mean :2.175	Mean :2.055
5	3rd Qu.:3.000	3rd Qu.:3.000	3rd Qu.:3.000	3rd Qu.:3.000
6	Max. :5.000	Max. :8.000	Max. :5.000	Max. :8.000

Figure 8: Summary of Number of Attempt

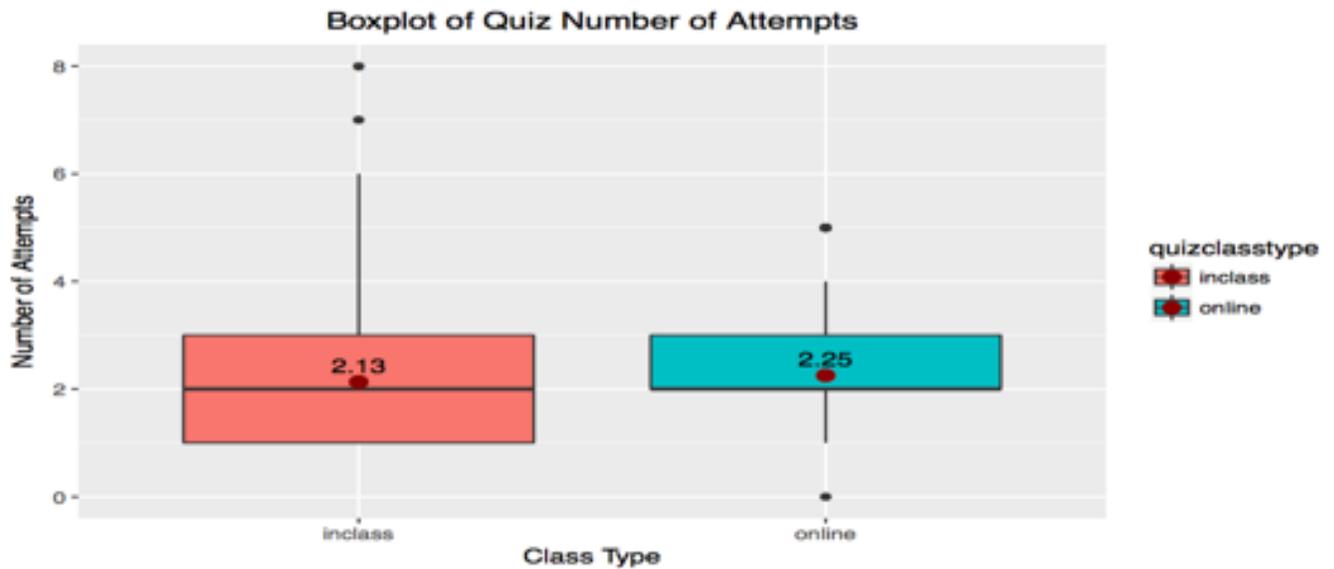


Figure 9: Boxplot of Number of Attempts

The next exploratory analysis of data was the individual quiz comparison to see if online or on campus performed better. The total number of students for online is 110 while for on campus the total number of students is 284. The first quiz comparison produced a p-value of .0792 which means the online and on campus data is not significantly different. Online only had 102 students attempt the quiz while on campus had 265 students attempt the quiz. However, looking at summary of the two data sets in figure 10, the minimum is vastly different. Also looking at figure 11, there are many outliers in the on-campus data. If more data is added, the p-value probably would not change drastically to go below 0.05 since the material covered on quiz one discusses sample versus population and the three main components of statistics.

	Highest Score Online	Highest Score on Campus
1	Min. : 86.67	Min. : 65.83
2	1st Qu.: 95.00	1st Qu.: 95.00
3	Median :100.00	Median :100.00
4	Mean : 97.99	Mean : 96.24
5	3rd Qu.:100.00	3rd Qu.:100.00
6	Max. :100.00	Max. :100.00

Figure 10: Summary of Quiz One

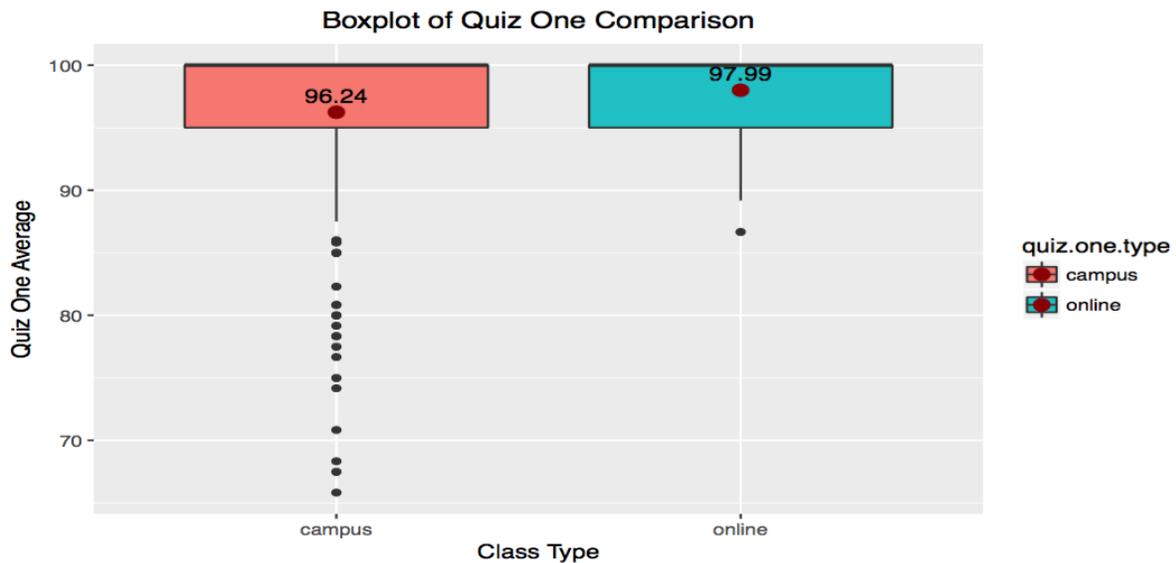


Figure 11: Boxplot for Quiz One

The second quiz comparisons produced a p-value of 0.6995. With the means of the on campus and online only differing by 0.56 percent as shown in figure 12. Similar to quiz one on campus has many outliers, but in quiz two online has many outliers as well as seen in figure 13. The on-campus data had 278 students take the second quiz while only 82 students took the second quiz in the online data. The second quiz covers two chapters with the topics of study

focusing on understanding graphical and numerical data, regression, correlation, and contingency of the data.

	Highest Score Online	Highest Score on Campus
1	Min. : 10.06	Min. : 21.46
2	1st Qu.: 85.35	1st Qu.: 85.05
3	Median : 92.42	Median : 91.86
4	Mean : 88.68	Mean : 89.06
5	3rd Qu.: 96.52	3rd Qu.: 96.94
6	Max. :100.00	Max. :100.00

Figure 12: Summary of Quiz Two

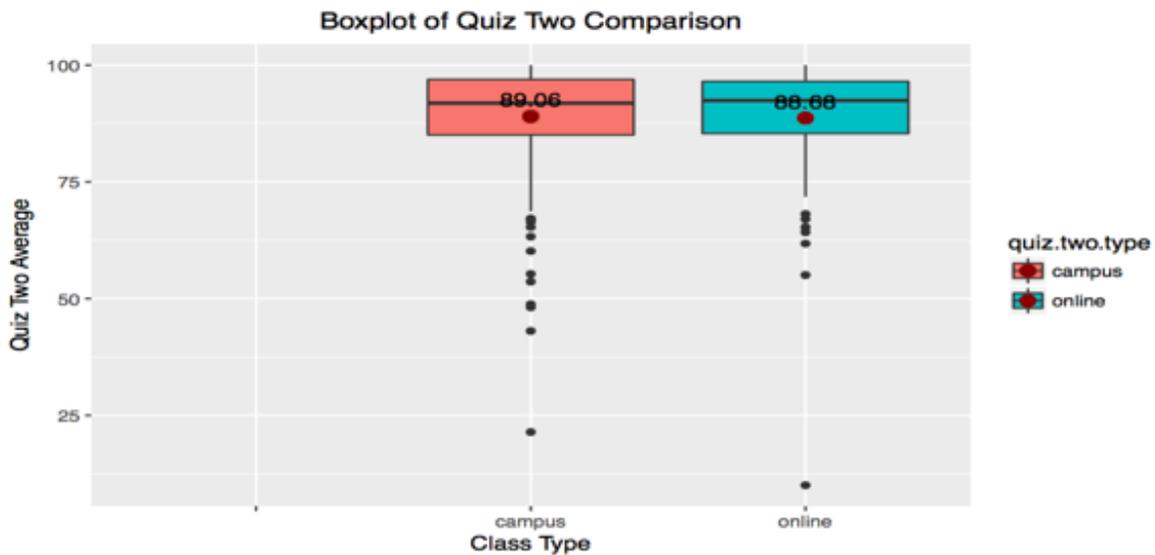


Figure 13: Boxplot of Quiz Two

For online there was only fifty-eight students while for on campus there was 139 students. This may be because the topics covered could be considered one of the hardest topics in the introductory statistics course. Quiz three produced the smallest p-value of  $8.611 \times 10^{-5}$  which is

significantly less than 0.05. Looking at the summary of the quiz (figure 14) the mean of the on campus and online differ 6.24. By looking at the boxplot of the quiz (figure 15) the overall box for online is lower in the percentage. The third quiz covers chapter four and five which covers determining the best ways to sample, experiment, and gather data and also how to find the probability and how to apply the probability rules.

	Highest Score Online	Highest Score on Campus
1	Min. : 29.11	Min. : 51.44
2	1st Qu.: 75.60	1st Qu.: 81.39
3	Median : 80.70	Median : 90.19
4	Mean : 80.37	Mean : 86.61
5	3rd Qu.: 87.71	3rd Qu.: 95.29
6	Max. :100.00	Max. :100.00

Figure 14: Summary of Quiz Three

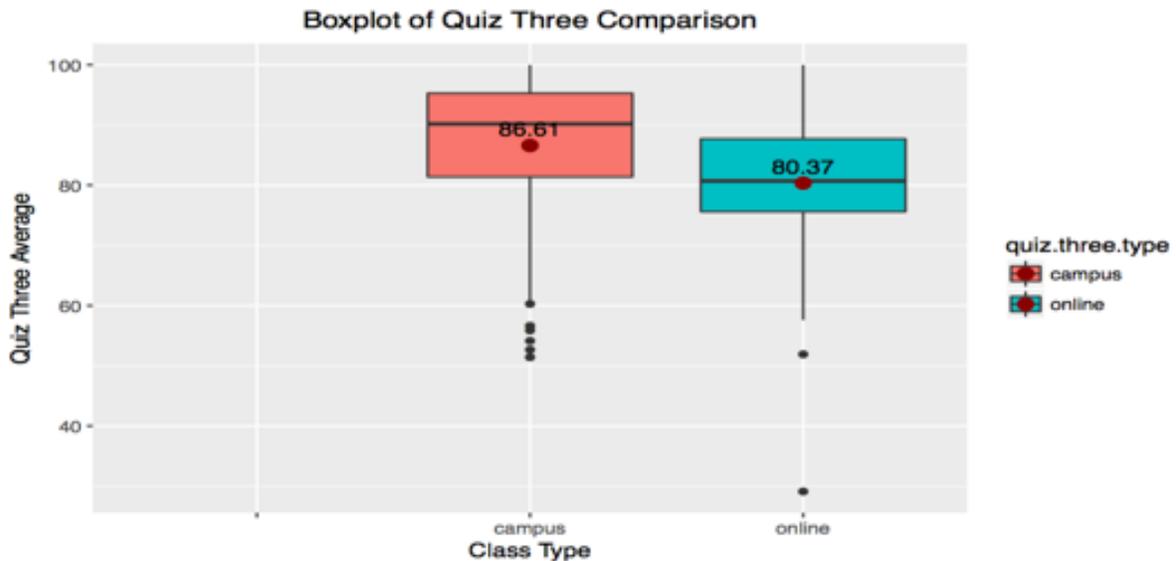


Figure 15: Boxplot of Quiz Three

Quiz four covered only chapter six which focuses on probability distribution of bell-shaped distribution and observations with two possible outcomes. Figure 16, the summary, shows the means of online and on campus having a difference of 3.03 percent. The boxplot in figure 17 shows the small data set of online generates closer first and third quartile then on campus. However, the p-value when comparing the two is 0.6051. This may due to the online data only having 23 students who completed the quiz and on campus data containing 175 students. The online data for quiz created a data limitation by not allowing the use of t-test since the n value was not greater than thirty which is another reason the Wilcoxon Rank Sum Test was used.

	Highest Score Online	Highest Score on Campus
1	Min. : 42.59	Min. : 3.33
2	1st Qu.: 80.78	1st Qu.: 74.00
3	Median : 86.11	Median : 84.67
4	Mean : 84.01	Mean : 80.98
5	3rd Qu.: 93.98	3rd Qu.: 94.44
6	Max. :100.00	Max. :100.00

Figure 16: Summary of Quiz Four

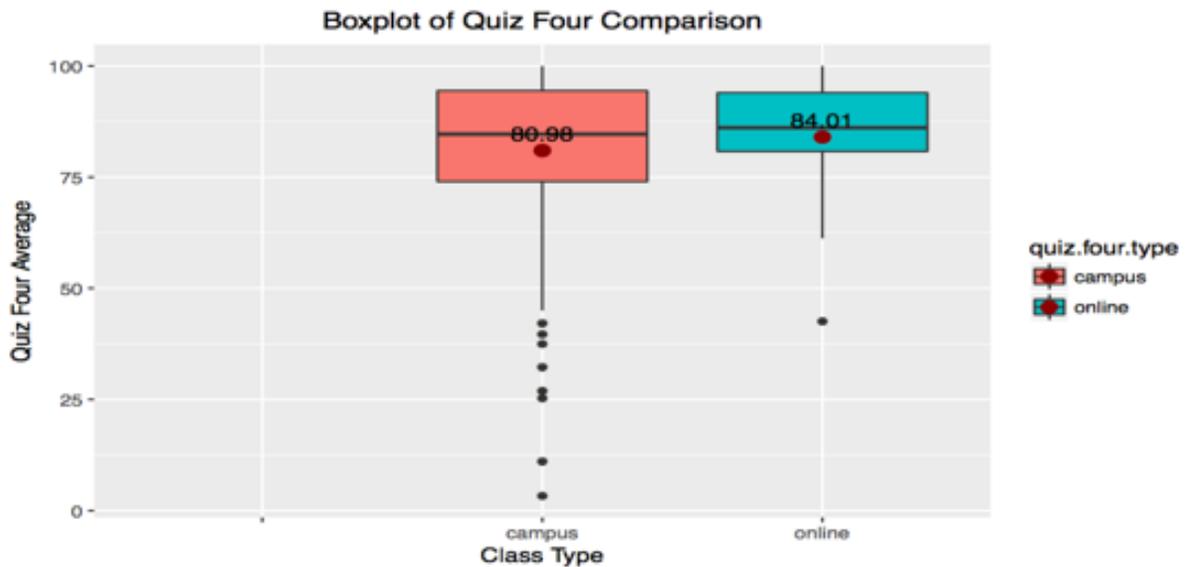


Figure 17: Boxplot of Quiz Four

For quiz five, online had eighty-five students take the quiz and on campus had 222 students take the quiz. Since five only covers chapter seven which covers sample distributions, the mean of the online and on campus only differ by .04 percent with online having the higher mean as seen in figure 18. Similar to quiz four online has closer first and third quartile as seen in the boxplot of quiz five (figure 19). The p-value when comparing online and on campus is 0.2925 which means the two are not significantly different.

	Highest Score Online	Highest Score On Campus
1	Min. : 24.62	Min. : 29.50
2	1st Qu.: 87.33	1st Qu.: 85.13
3	Median : 91.81	Median : 93.39
4	Mean : 89.32	Mean : 89.28
5	3rd Qu.: 96.00	3rd Qu.: 98.43
6	Max. :100.00	Max. :100.00

Figure 18: Summary of Quiz Five

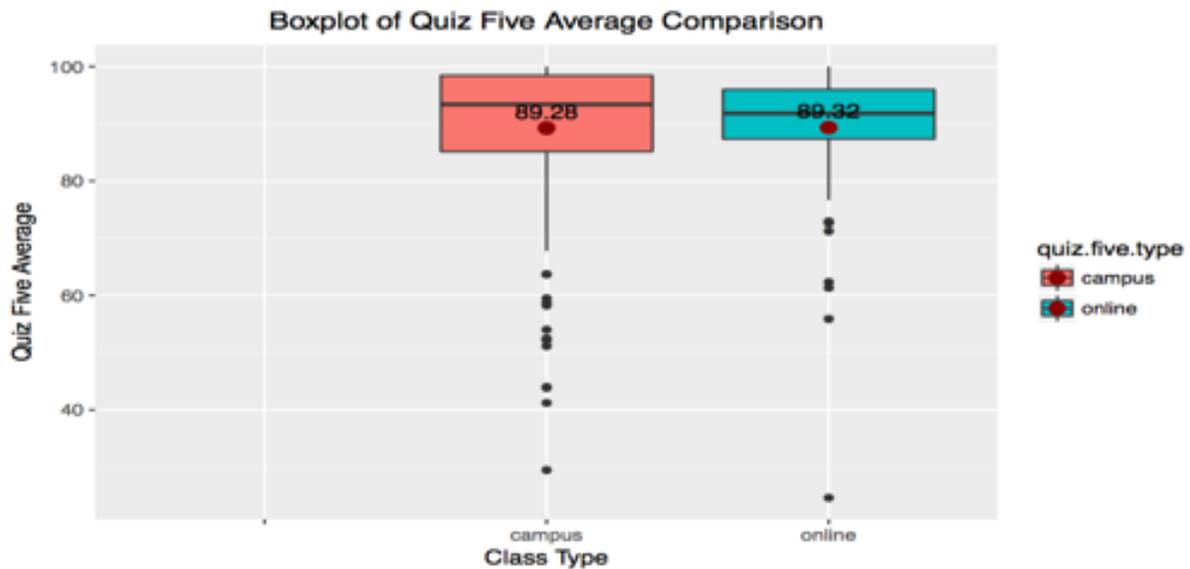


Figure 19: Boxplot of Quiz Five

For quiz six, the online data had only forty students who attempted the quiz while on campus only had sixty-six students. An assumption of why on campus students did not attempt or the student's scores being ignored is because this was the last quiz for online and the last two quizzes for on campus and the students know that their lowest two quiz grades will be dropped. Also, the chapters eight and nine which covers confidence intervals and significance test which could be considered two of the hardest chapters to understand. By looking at the summary (figure 20), we notice the mean of the online and on campus differ by 5.33 percent. The boxplot in figure 21 shows the online first and third quartile are actually farther apart than on campus. Since this quiz covers the hardest topics, the p-value is 0.06949 which is greater than 0.05 which means not significantly different, but if addition of more data could make the p-value drop lower than 0.05 which then makes online and on campus significantly different.

	Highest Score Online	Highest Score on Campus
1	Min. : 40.19	Min. : 21.25
2	1st Qu.: 64.52	1st Qu.: 73.34
3	Median : 79.12	Median : 86.07
4	Mean : 75.90	Mean : 81.23
5	3rd Qu.: 88.66	3rd Qu.: 90.86
6	Max. :100.00	Max. :100.00

Figure 20: Summary for Quiz Six

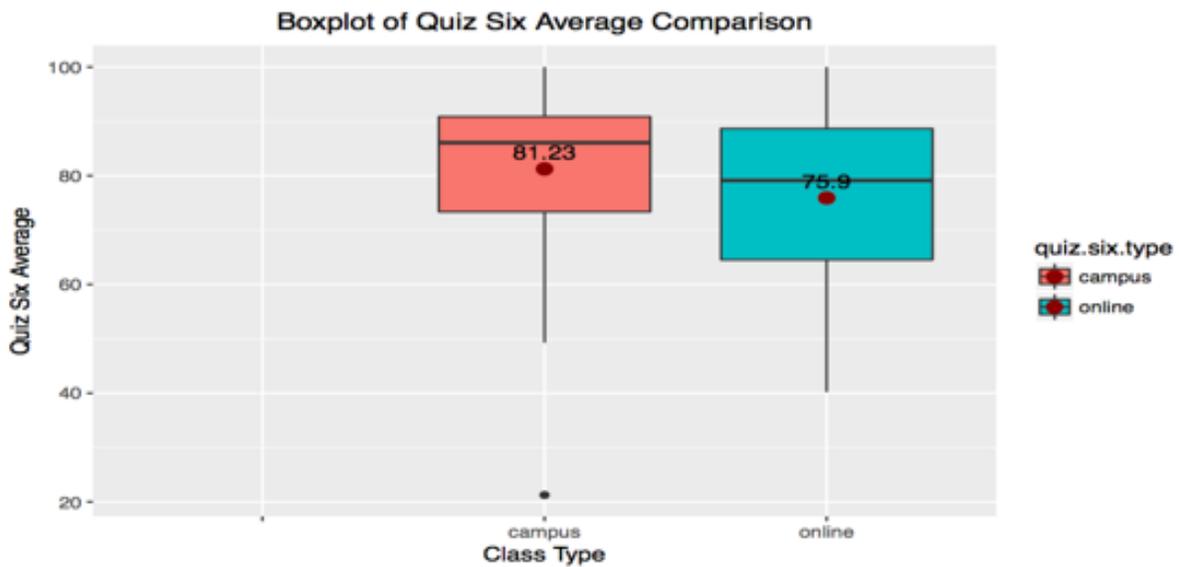


Figure 21: Boxplot for Quiz Six

The following exploratory analysis deals with linear regression models and scatterplot matrices. The first scatterplot matrix as shown in figure 22 shows the quiz average data is left skewed with homework correlation being highly correlated and the first attempt quiz average was correlated as well but not highly correlated. For the homework average, the on-campus data can better predict the final quiz average then the online data. This is possibly due to the

resources like the learning center and the professor office hours an on-campus student has to achieve a higher homework grade which correlates to a higher homework average.

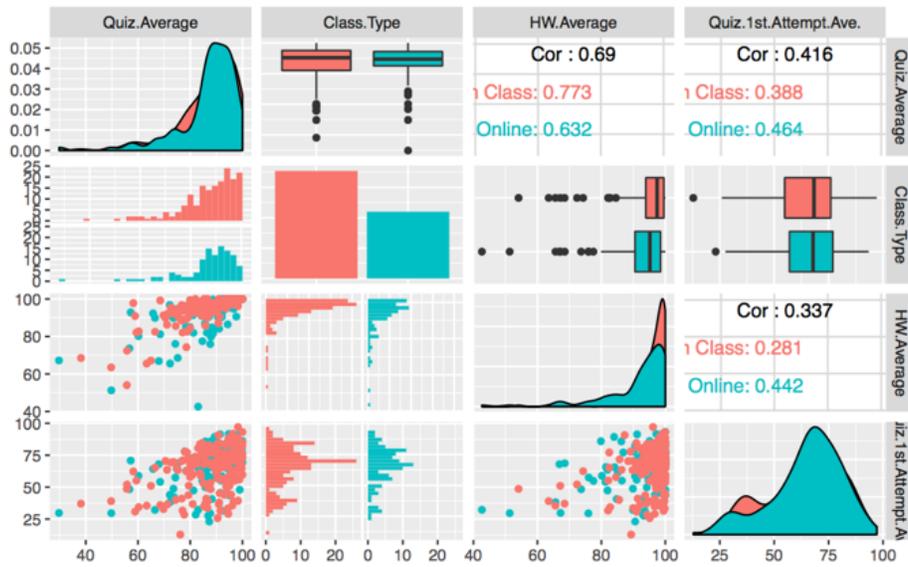


Figure 22: Scatterplot Matrix for Final Quiz Average

Since the final quiz grade can be predicted by both homework average and first attempt quiz average, the formula can be seen as final quiz average= slope estimate\*homework average/first attempt quiz average + intercept. For online, the final quiz average can be predicted by  $0.7059 * (\text{student's homework average}) + 22.1215$  and  $0.3290 * (\text{student's first attempt quiz average}) + 65.6095$ . For on campus, the final quiz average can be predicted by  $1.1642 * (\text{student's homework average}) - 23.3941$  and  $.2525 * (\text{student's first attempt quiz average}) + 71.1371$ .

The prediction of a student's final quiz average can be determined by the number of attempts and the days before due date. The scatterplot matrix for the online final quiz average, figure 23, being predicted by number of attempts and days before due date for online is left

skewed which lacks normality therefore a logthrim transformation must be applied to the data.

The formula to get the response variable is  $\log(101 - \text{quiz average})$ .

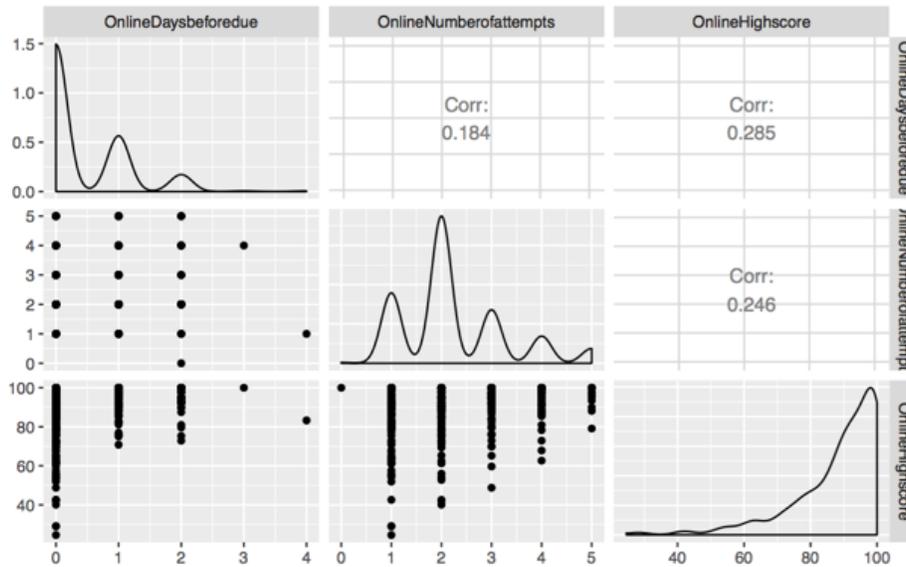


Figure 23: Scatterplot Matrix for Online Final Quiz Average

To predict the final quiz average for online the formulated by  $0.1353 * (\text{days before due date and number of attempts}) - 0.2855 * (\text{number of attempts}) - 0.8149 * (\text{days before due date}) + 2.7693$ .

The reason for having a data section with both days before due date and number of attempts together is because the two are correlated since a student can attempt the quiz more times generally instead of having limited time to complete the assignment multiple times. The negatives in the formula does not mean there is a negative correlation. For on campus final quiz grade prediction form the days before due date and number of attempts, is calculated the same way since in figure 24, the quiz average is left skewed. The same logarithim transformation as the online final quiz grade, the formula is slightly different by the coefficients since the formula is  $0.0668 * (\text{days before due date and number of attempts}) - 0.4723 * (\text{number of attempts}) - 0.3483 * (\text{days before due date}) + 3.0677$ .

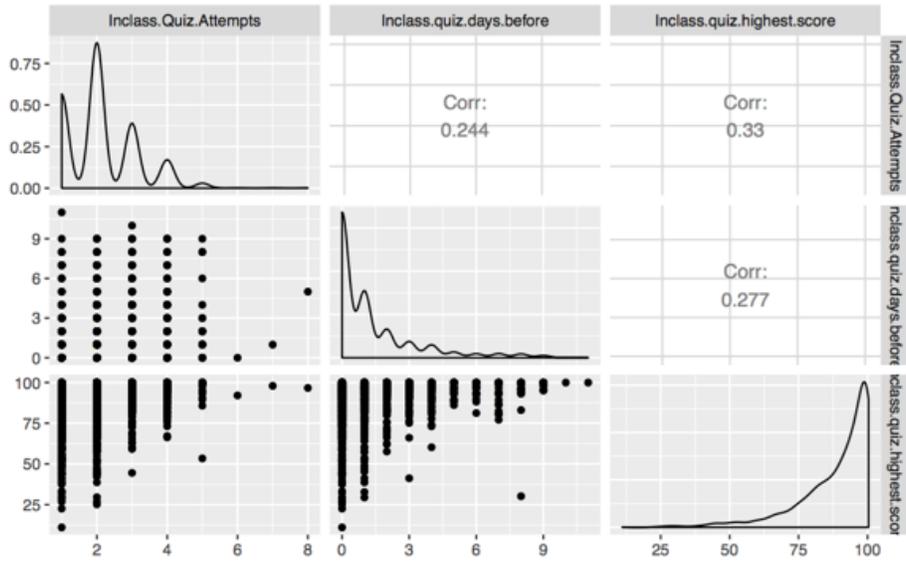


Figure 24: Scatterplot Matrix for On Campus

### Final Quiz Average

The data had only two major data limitations which did not allow further exploratory analysis is the days before due date between online and in class due to the length of the semesters varying greatly. The other limitation is the data sizes for the quiz comparisons being small in size which could be improved by adding more data. Other than these two data limitations, the comparisons of online and on campus are comparable; however, if a student should choose to take the class online or on campus, on campus would be the better choice since the homework average is higher and quiz comparisons are slightly higher for all but one quiz and the student has several resources to receive help from.

## Further Research

Collecting more data would be the first step in the further research. Since all the data in this research was from one statistics teacher the expansion of getting data from other teachers and the addition of more semesters for online and on campus. After the collection of more data, the addition of more variables like how long a student took on taking the assignment, the attendance of the student, number of times a student logs on into MyStatLab, and for homework if a student took the assignment in one session or multiple sessions. Adding the variable of how long a student took taking the assignment would help judge if a student actually truly attempted the first attempt or if the student looked at the first attempt as a view of what the quiz will cover. Analyzing this data will help determine if the first attempt average is accurate. For on campus, the attendance of a student would help analyze if a student who attends class receives a higher quiz and homework grade over students who less frequently attend class. Analyzing the number of times a student login to MyStatLab would help determine if a student only log on the days the assignment are dues or the day the student first attempt the assignment. Looking at the homework for if a student takes the homework in one session or multiple sessions will help understand if a student takes breaks or focuses on the statistics course. Adding these variables and more data would better enhance the continuing research to get a more accurate comparison of the environments based on the student's performance.

## References

Hastie, Trevor, Robert Tibshirani, and J. H Friedman. 2001. *The Elements of Statistical Learning*.

New York: Springer.

Hogg, Tanis, and Zimmerman. *Probability and Statistical Inference*, Ninth Edition.

Wickham, H. ggplot2. *Elegant Graphics for Data Analysis*.