Assessment of the Stretch Composition Program

Cyrus Davis, David Kaufman, Eugene Monforte, Max Wang

May 1, 2025

Abstract:

At a state university operating on the quarter system, incoming freshmen choose from three course sequences to fulfill their GE requirement in Written Communication. Prior to registration, the students receive recommendations from faculty, a placement exam, or an online self-placement survey. This study categorizes students into three groups based on whether they followed, shortened, or lengthened their recommended sequence. Our primary goals for this project are to analyze the data and assess whether parametric or nonparametric methods are appropriate based on the given statistical assumptions. From there, we will answer our client's questions based on the given circumstances to detect performance patterns across each group for a deeper analysis.

1) Introduction

A dataset provided from our client includes academic and demographic information for 124 first-time freshmen enrolled in either a 1-quarter course "Composition", 2-quarter course "Stretch Composition", or a 3-quarter course "Extended Stretch Composition". Students receive placement recommendations based on faculty advice, placement tests, or a self-placement survey. However, students do not always follow these recommendations, and the academic impact of these decisions remains a key question for the university.

To assess the effectiveness and meaning of the program, we are to make sure any row involving Column E (Followed Recommend = 9999) is excluded. This will ensure high accuracy of the results. The 15 questions assess a variety of information about the students containing demographic information, course sequence taken, recommendation alignment, portfolio rubric scores, and course grades. Our goals for this report are to clearly define research variables, goals and select appropriate statistical methods to address the university's key questions. Methods stated would range from a variety of ANOVA techniques (one-way, two-way, repeated measure, etc.), to t-test and chi-square testing. We will state the type of techniques used for this report (one-sample or two-sample, independent or dependent) on the cleared data.

2) Methods:

In our given dataset, we have 19 columns (A-S) for the Stretch Composition Program. Column E includes a section of data that has '9999' (denoted missing). Four students with this data classification were removed, reducing the sample size from 124 to 120 students. The following includes a description and subclassification (in brackets) of each column.

1) College [1 = Agriculture, 2 = Business, ..., 8 = Science, 9 = Undecided], (A)
2) Sex [1 = Female, 2 = Male], (B)
3) Underrepresented Minority [(URM); 1 = Yes, 2 = No], (C)
4) Sequence [1, 2, 3], (D)
   - 1 = 1-quarter course "Composition", ENG 110
   - 2 = 2-quarter sequence course "Stretch Composition", ENG 105/106

- 3 = 3-quarter sequence course “Extended Stretch Composition”, ENG 100/101/102

5) The course sequence the student completed after the recommendation (Followed Recommend), (E)

Sequence	Followed Recommend	Description
1	Yes	The student took a 1-quarter course, as recommended
	Shorter	Against the recommendation of a 2- or 3-quarter course, the student took a shorter 1-quarter course
	Longer	None
	9999	The student took a 1-quarter course, without any recommendation.
2	Yes	The student took a 2-quarter course, as recommended.
	Shorter	Against the recommendation of a 3-quarter course, the student took a shorter 2-quarter course.
	Longer	Against the recommendation of a 1-quarter course, the student took a longer 2-quarter course.
	9999	The student took a 2-quarter course, without any recommendation
3	Yes	The student took a 3-quarter course, as recommended
	Shorter	None
	Longer	Against the recommendation of a 1- or 2-quarter course, the student took a longer 3-quarter course.
	9999	None

Student’s Composition Assignment Score based on Six Portfolio Rubric Categories; Based on a 0.0 - 4.0 grading scale; (F-K):

6) Process Revision, (F)
7) Critical Reading, (G)
8) Rhetorical Analysis, (H)
9) Reseach, (I)
10) Style, (J)
11) and Grammer, (K)
12) Portfolio Total Score (Sum of Columns 6 through 9), (L)

Course Grade in Composition Class the student took; Based on a 0.0 - 4.0 grading scale; (M–R)

13) ENG 100 Grade (M)
14) ENG 101 Grade (N)
15) ENG 102 Grade (O)
16) ENG 105 Grade (P)
17) ENG 106 Grade (Q)
18) ENG 110 Grade (R)
19) Composition Average Grade (Average of Columns M through R); (S).

Multiple questions were conducted using parametric and nonparametric testing methods. We will analyze each question by verifying the normality assumption with the Shapiro-Wilk Test. If the data is normal (or nonnormal) we will use parametric (or nonparametric) techniques. Pairwise comparison post-hoc tests will also be conducted if applicable via Tukey's HSD test for parametric tests and Methods with Bonferroni corrections for nonparametric data. All questions involving comparing means (or medians) will begin by summarizing the dependent variable (outcome) with summary statistics that include a mean, standard deviation (SD), boxplots for all observations and for each level of the independent variable (factor). Some questions will include an (r x c) contingency table, where r (or c) represents the number of rows (or columns), based on the categorial variables involved.

Each subsection of our statistical results will have a section where it states "Q#X" for X = 1, 2, ..., 15. Each section will have a description stating the variables involved, the client's goal of each question, the stated method we will use for analysis and the procedure with statistical tables (if necessary). Prior to any of our statistical methods, we will test if the normality assumption holds for each of our client's questions. For any two-way ANOVA testing, two-way interactions plots will be conducted. All hypotheses testing will be two-tailed and conducted at the 5% significance level.

3) <u>Results</u>

Our first set of questions is to determine how well students evaluated their skills by examining their success related to the following basic questions listed below.

1) Portfolio Total Score
2) Comp AVF Grade

The basic questions:

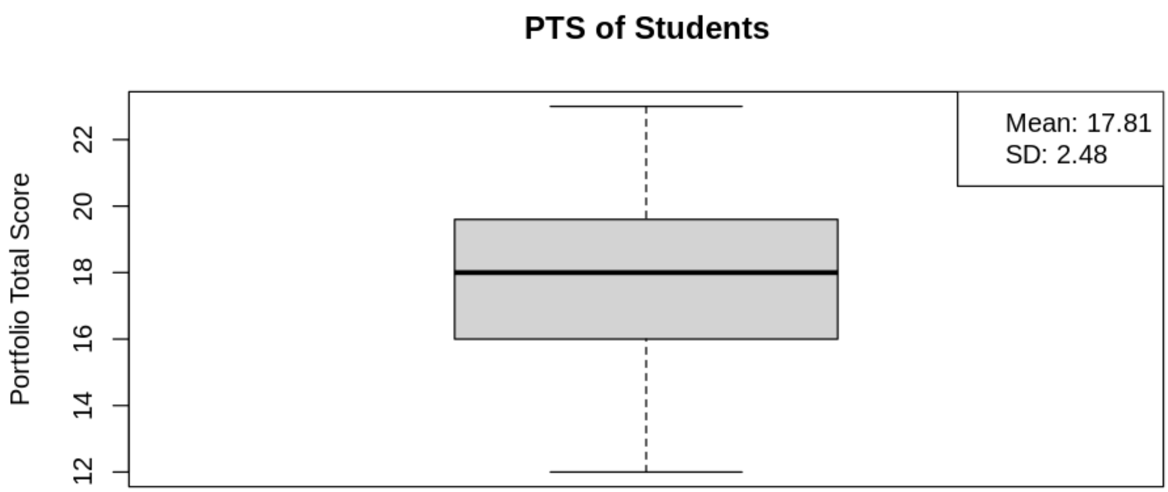
- How well do students perform when they follow our recommendation?
- How well do students perform when they take a shorter sequence than recommended?
- How well do students perform when they take a longer sequence than recommended?

3.1) Question 1. Our goal for the first question is to test the difference in the Portfolio Total Score (PTS) among the three Groups. Our variables involved in this question are the PTS (Column L) and the Group (Followed Recommend, Column E). Our dependent variable for this question is the PTS where our independent variable is the Followed Recommend. Since we are testing the difference in the PTS from all three Groups, we will conduct a one-way ANOVA test.

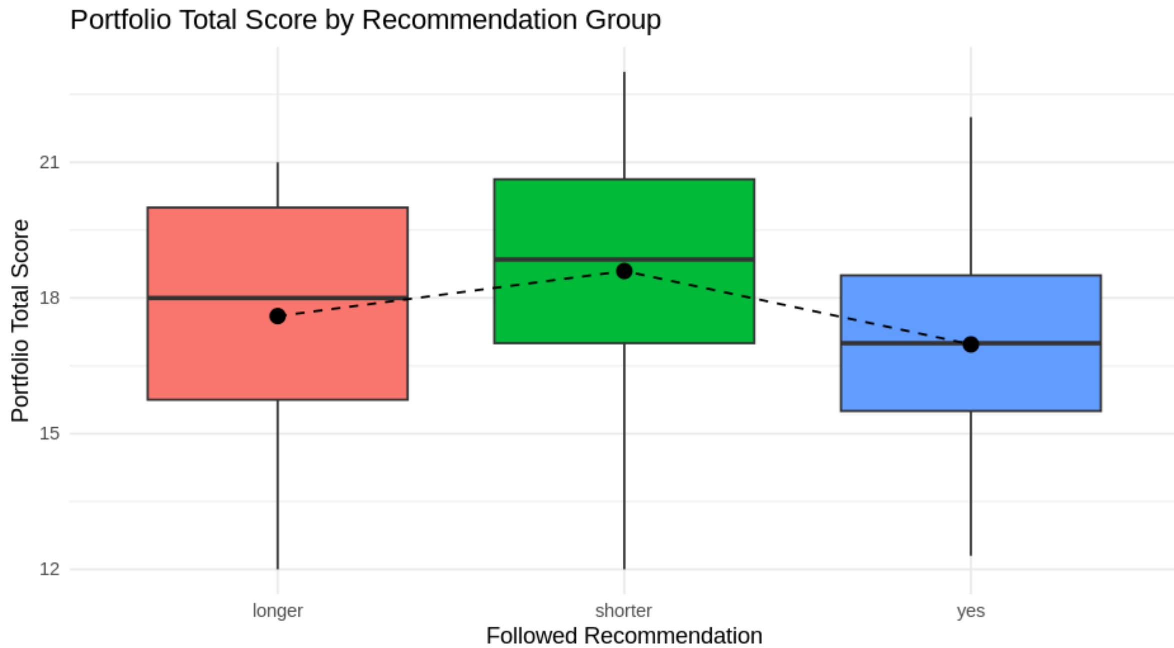
Listed below are a boxplot and summary statistics depicting all observations of PTS. The observations for the boxplot for all observations look ideally symmetric with a mean PTS of 17.81 and standard deviation of 2.48. There are also no outliers based on the Inter-Quartile Range (IQR) test.

Summary Statistics for All Observations:

Min.	1st Qu.	Median	3rd Qu.	Max	Mean	SD
12.00	16.00	18.00	19.55	23.00	17.81	2.48



For the boxplot of the individual boxplots relating to the three Recommendation groups (longer, shorter, yes) we get the following summary statistics and depictions. All three groups look primarily symmetric with no outliers outside of the IQR rule. The Shorter group has the highest range of PTS scores with a mean of 18.6 compared to longer (17.6) and shorter (17.0) but the longer group has a higher variation with an SD of 2.65.



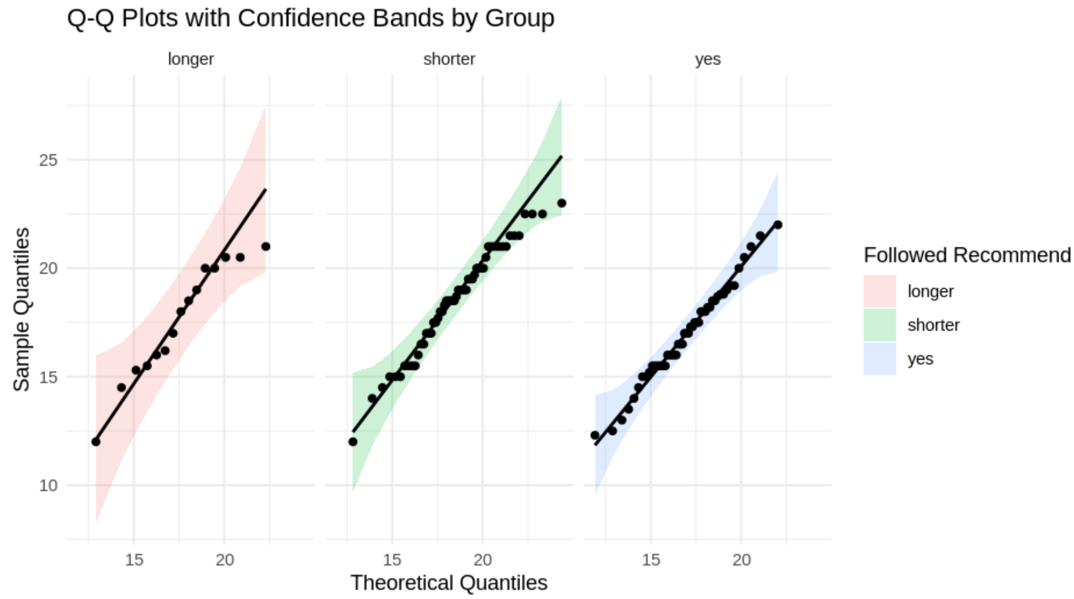
Summary Statistics for each Group:

Group Type	Min	1st Qu.	Median	3rd Qu.	Max	Mean	SD
Longer	12.0	15.8	18.0	20.0	21.0	17.6	2.65
Shorter	12.0	17.0	18.8	20.6	23.0	18.6	2.46
Yes	12.3	15.5	17.0	18.5	22.0	17.0	2.20

Prior to conducting the one-way ANOVA, we checked if the variables followed a normal distribution. To do this, we formed the following linear regression model:

$$y_{ij} = \mu + \alpha_i + \varepsilon_{ij} \ ;$$

where y_{ij} is the PTS scores, μ is the mean effect on the PTS, α_i represents the Group (i = 1, 2, 3) and ε_{ij} is the error, (j = 1, 2, ... , 40) of the 120 samples. From this we conducted a QQ-plot to determine if the data follows a normal distribution and a Shapiro-Wilk test. The QQ plots for all three groups, each look symmetric but have heavy tails. A Shapiro-Wilk test reveals each group has a p-value greater than the 0.05 level. Since it is greater than the significance level, we can assume the model is normal and proceed with the one-way ANOVA.



Shapiro-Wilk Test Group Q01 results:

Group	p-value
Longer	0.377
Shorter	0.265
yes	0.832

Given the data is normal, we constructed a one-way ANOVA with the following hypothesis:

- H_0: The mean Portfolio Total Score is equal across all three groups. ($\mu_1 = \mu_2 = \mu_3$)
- H_A: At least one of the groups is different. ($\mu_1 \neq \mu_2 \neq \mu_3$)

When conducting one-way ANOVA between the PTS and the three groups we got the following ANOVA results. The ANOVA reveals an F statistic of 6.129 with a p-value of 0.00294. Since the p-value is less than the 0.05 level, we reject the null hypothesis and conclude that at least one of the mean Portfolio Total Scores are different from the three groups.

ANOVA table for Q01:

	Df	Sum Sq	Mean Sq	F value	Pr(>F)
Group	2	69.6	34.79	6.129	0.00294
Residuals	117	664.2	5.68		

From there we conducted Tukey's HSD Test to compare the mean difference between each level of the Groups. We construct the following hypothesis:

- H_0: There is no difference between the means of the two groups. ($\mu_i = \mu_j$)
- H_A: There is a difference between the means of the two groups ($\mu_i \neq \mu_j$)

Of the comparisons, the subcategory comparison between “yes” and “shorter” has the highest difference between each level comparison. It's also the only significant comparison result. We can say that from the recommendation the difference between the “shorter” recommendation and “yes”, those who take the recommendation, is significant.

Tukey’s HSD test for Group

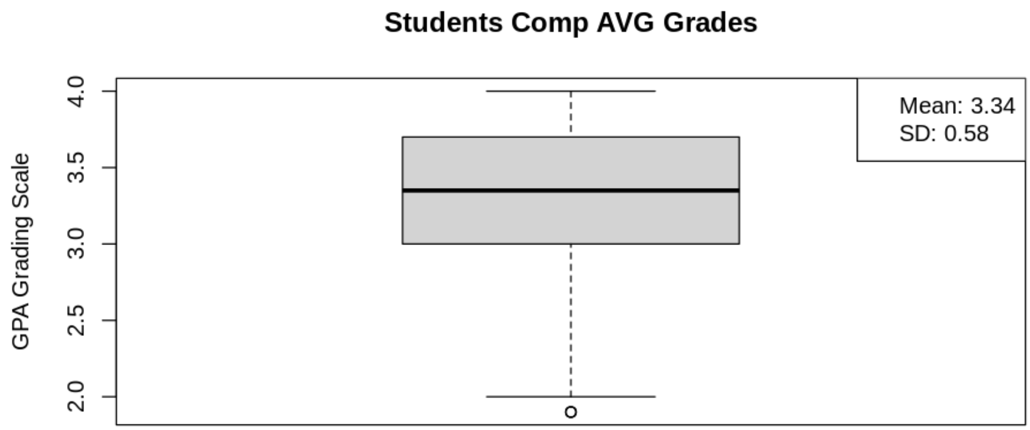
	diff	lwr	upr	p adj
Shorter-Longer	0.996	-0.648	2.641	0.325
Yes-Longer	-0.627	-2.296	1.042	0.647
Yes-Shorter	-1.623	-2.729	-0.517	0.002

3.1) Question 2. For the second question, our goal is to test the difference in Comp AVG Grades (Column S) with Group. Our variables are the Comp AVG Grade (Column L) and Group (Followed Recommend, Column E). Our dependent variable for this question is the Comp AVG Grade where our independent variable is the Followed Recommend. Since we are testing the difference in the Comp AVG Grades from all the three Groups, we will conduct a one-way ANOVA test.

Listed below is a boxplot and summary statistics depicting all observations of Comp AVG Grades. The boxplot for all observations looks ideally symmetric with a mean of 3.34 and standard deviation of 0.58. There is one outlier based on the Inter Quartile Range (IQR) test below the 2.0 GPA grading scale.

Summary Statistics for All Observations:

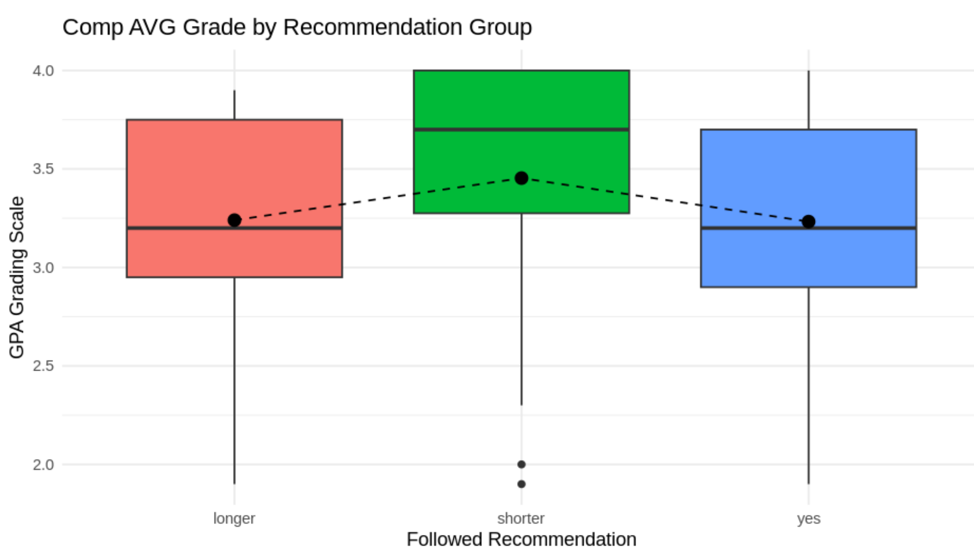
Min.	1st Qu.	Median	3rd Qu.	Max	Mean	SD
1.90	3.00	3.35	3.70	4.00	3.34	0.58



For each Group level based on the Comp AVG Grade, each graphical display looks skewed. The Longer and Yes categories of Group look skewed to the right compared to the boxplot of Shorter which is skewed to the left. The Shorter group has the better range in terms of comparing it between Longer and Yes levels especially since it has the highest mean Comp AVG Grade of 3.45, but has the lowest standard deviation of 0.544.

Summary Statistics for each Group:

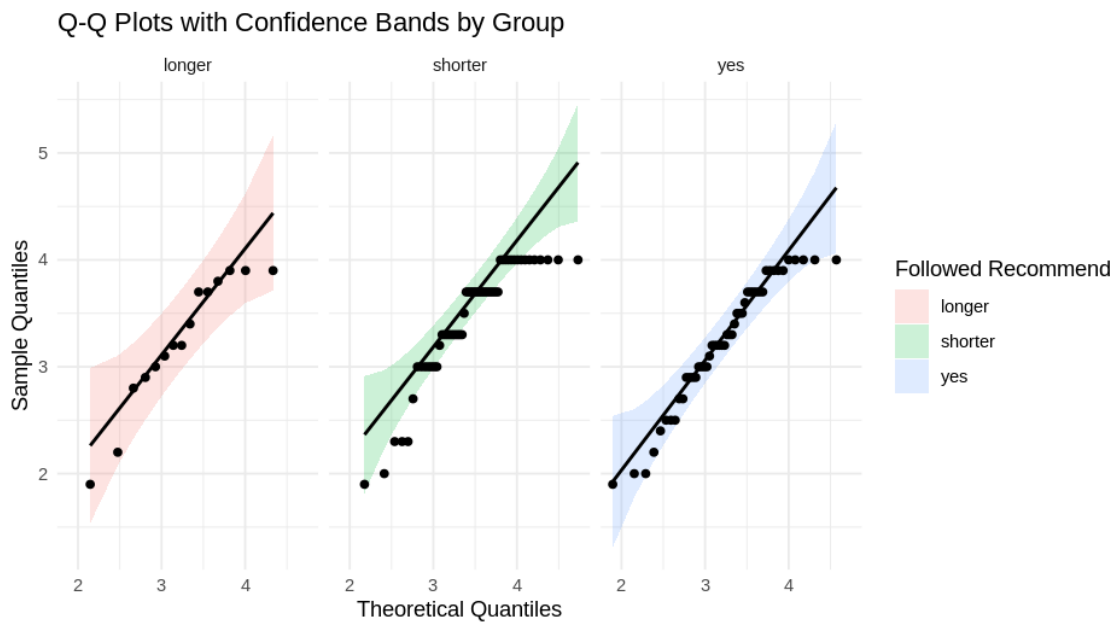
Group Type	Min	1st Qu.	Median	3rd Qu.	Max	Mean	SD
Longer	1.90	2.95	3.20	3.75	3.90	3.24	0.617
Shorter	1.90	3.27	3.70	4.00	4.00	3.45	0.544
Yes	1.90	2.90	3.20	3.70	4.00	3.23	0.583



Similar to question 1, we formed the following linear regression model:

$$y_{ij} = \mu + \alpha_i + \varepsilon_{ij};$$

where y_{ij} is the Comp AVG Grades, μ is the mean effect on the PTS, α_i represents the Group (i = 1, 2, 3) and ε_{ij} is the error, (j = 1, 2, ..., 40) of the 120 samples. We tested the normality assumption on the model between the variables to ensure that one-way ANOVA would provide correct results. Listed below are the QQ plots of each Group level. Unlike for question 1, for the Comp AVG Grades when comparing it to the Group levels, the Shorter and Yes levels fail to achieve a symmetric distribution based on the Confidence bands provided. The Shapiro-Wilk test confirms this. For both Shorter and Yes groups, they have p-values less than the 0.05 level which suggests the parametric one-way ANOVA technique would not be useful for this question.



Shapiro-Wilk Test Group Q02 results:

Group	p-value
Longer	0.0964
Shorter	0.00000669
Yes	0.0140

Since there is only one independent variable, we would use the Kruskal-Wallis test which does not assume normality, compares the median scores across 2 or more independent groups and ranks the data. There are 3 independent groups in the Followed Recommended (longer, shorter, and yes) so this method would work. We use the following hypotheses and find that from this test it yields a p-value that is greater than the 0.05 level. There is not a statistically significant difference in Comp AVG Grades among the groups. We can say that the median distributions of the Comp AVG Grades are the same for all recommendation groups.

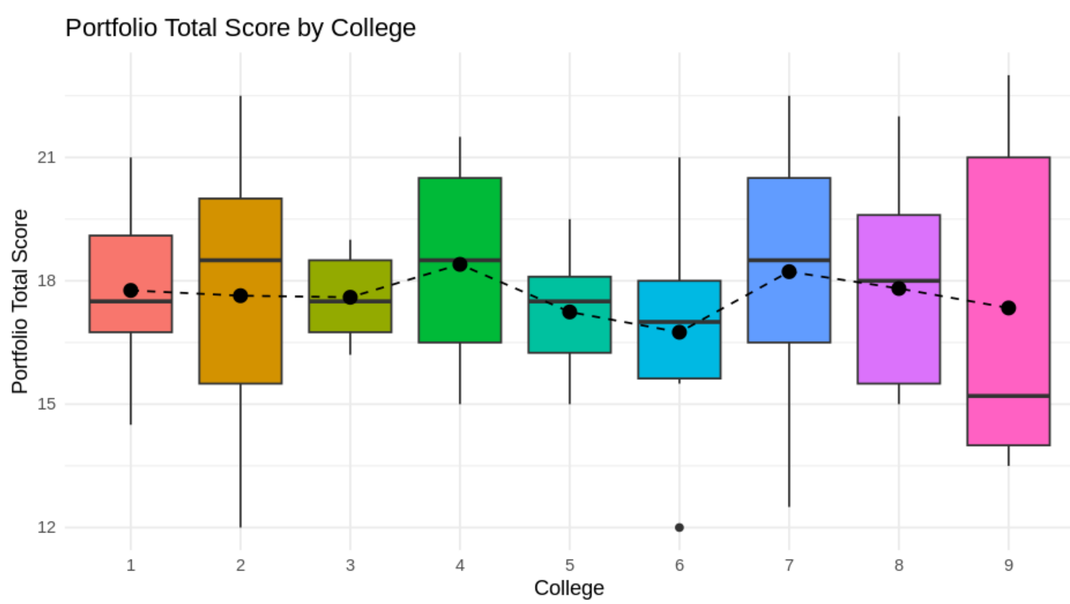
- H_0: The median distributions of Comp AVG Grades are the same across all groups.
- H_A: At least one median distribution group is different.

Kruskal Wallis Test for Q02:

Test statistic	df	p-value
5.8614	2	0.05336

The next set of questions (3 through 7) are to identify any "achievement gaps" in a student's information.

3.3) Question 3. We are asked to test if "there is any statistical significance in the Portfolio Total Score between Colleges?". We will use the PTS and College (Column A) columns from the dataset to test the difference in the Scores by the nine colleges. For this we will use one-way Anova. The PTS boxplot and summary statistics for all observations are shown in section 3.1 in question 1. The individual boxplots below display graphical depictions and summary statistics for all PTS based on the nine colleges. Most of the college's graphical depictions look skewed. The best college mean score is College 4 where it has a mean of 18.4 with a standard deviation of 2.13. For the best variation, College 3 has the best standard deviation of 1.09. College 6 has one outlier on the lower tail despite having most values on the lower values of between a PTS of 15 and 18.



Summary Statistics for each College:

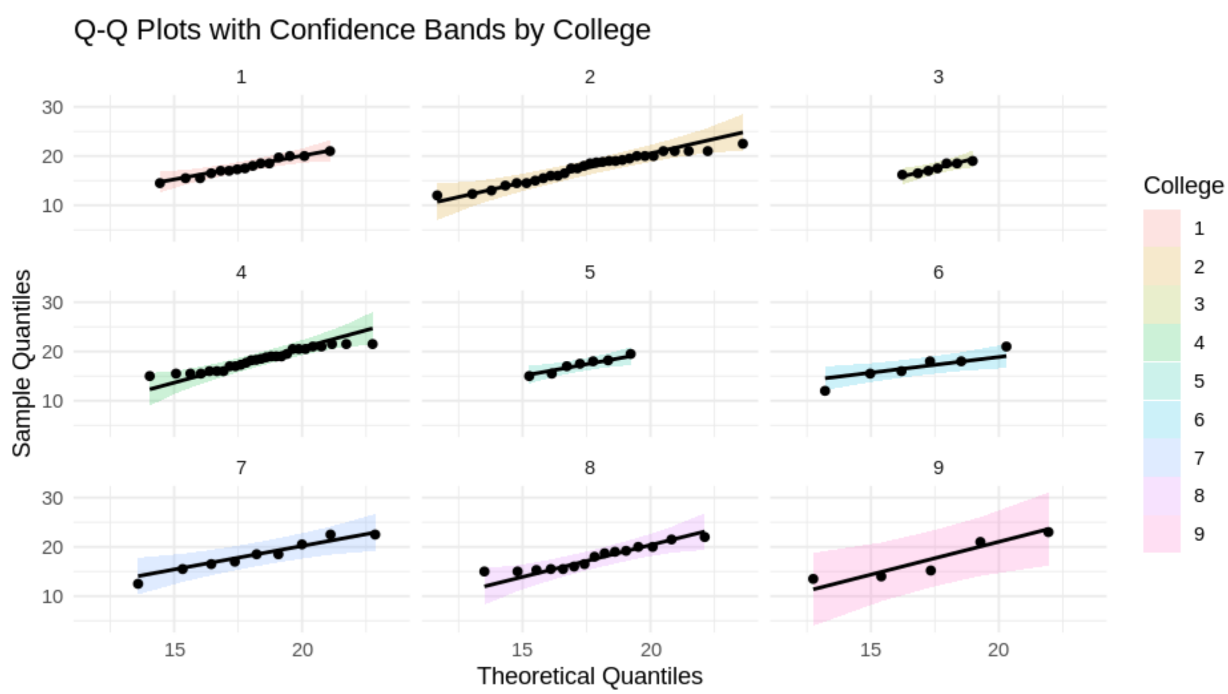
College #	Min	1st Qu.	Median	3rd Qu.	Max	Mean	SD
1	14.5	16.8	17.5	19.1	21.0	17.8	1.88
2	12.0	15.5	18.5	20.0	22.5	17.6	2.99
3	16.2	16.8	17.5	18.5	19.0	17.6	1.09
4	15.0	16.5	18.5	20.5	21.5	18.4	2.13
5	15.0	16.2	17.5	18.1	19.5	17.2	1.57

6	12.0	15.6	17.0	18.0	21.0	16.8	3.03
7	12.5	16.5	18.5	20.5	22.5	18.2	3.29
8	15.0	15.5	18.0	19.6	22.0	17.8	2.43
9	13.5	14.0	15.2	21.0	23.0	17.3	4.36

Before we conduct a one-way ANOVA, we would have to check if the data is normal. We formed the following linear regression model:

$$y_{ij} = \mu + \beta_i + \varepsilon_{ij} \ ;$$

where y_{ij} is the PTS Scores, μ is the mean effect on the PTS, β_i represents the Group (i = 1, 2, ..., 9) and ϵ_{ij} is the error of the 120 samples. The QQ plots between each labeled college from 1 through 9 are listed below. Each of these distributions looks roughly symmetric where each of the points are within the confidence bands with college 2 and 4 having heavy tails. Multiple Shapiro-Wilk Tests were conducted to test if any of the college types violate the normality assumption. Each of the p-values for the college types, have a p-value that is greater than the 0.05 level. Thus, we confirm the data is normal and conduct a one-way ANOVA. Listed below are the results for each of the QQ plot depictions and their p-values.



Shapiro-Wilk Test College Results:

College #	p-value
1	0.888
2	0.212

3	0.487
4	0.0668
5	0.805
6	0.881
7	0.745
8	0.745
9	0.176

Given the data is normal, we follow our one-way ANOVA statistical method with the following hypothesis:

- H_0: The mean Portfolio Total Scores are equal across all nine colleges.
$$(\mu_1 = \mu_2 = \cdots = \mu_9)$$
- H_A: At least one of the mean Portfolio Total Scores is different.
$$(\mu_1 \neq \mu_2 \neq \cdots \neq \mu_9)$$

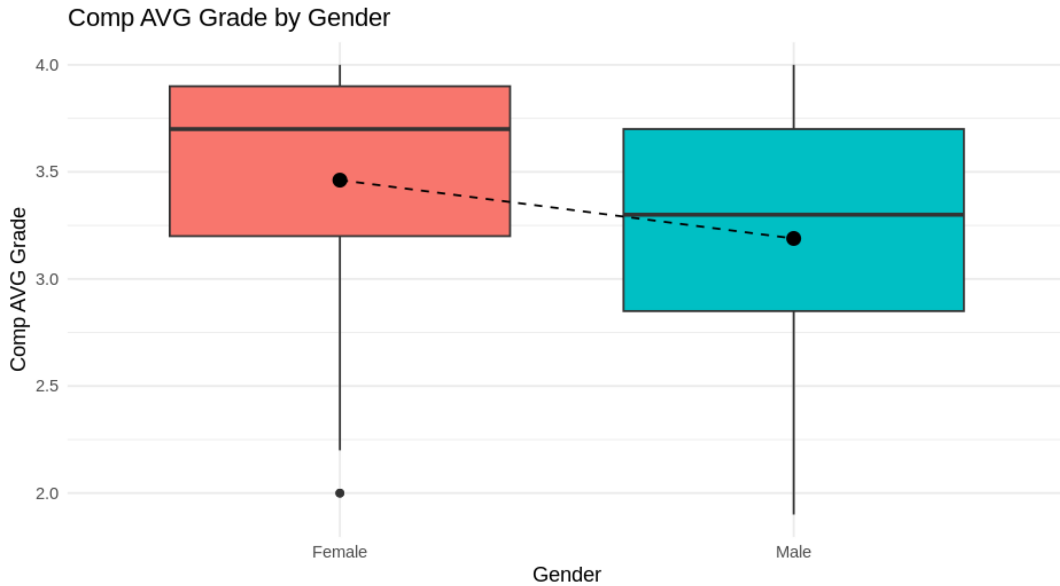
When conducting one-way ANOVA between the PTS and the 9 colleges we got the following ANOVA results. ANOVA reveals an F statistic of 0.433 with a p-value of 0.899. Since the p-value is greater than the 0.05 level, we do not reject the null hypothesis and say that the mean Portfolio Total Scores are the same for all nine colleges. Since there was no significance, we didn't continue with Tukey's HSD post hoc test. We can conclude the PTS scores are equal across all nine colleges.

ANOVA table for Q03:

	Df	Sum Sq	Mean Sq	F value	Pr(>F)
College	8	22.2	2.778	0.433	0.899
Residuals	111	711.5	6.410		

3.4) Question 4. We are asked to test if "there is any statistical significance in the Comp Average Grade between Gender?". We will use the Comp AVG Grade and Sex (Column B) columns from the dataset to test the difference in the Grades between the two genders. For these questions we will use Independent Two Sample T-Test since we have a numerical variable (Comp AVG Grade) and a categorical variable (Sex).

The Comp AVG Grade boxplot and summary statistics for all observations are shown in section 3.2 in question 2. The individual boxplots below display graphical depictions and summary statistics for all Comp AVG Grade based on the two genders. The graphical depictions for females skewed to the left with one outlier outside the IQR test unlike male distribution where it has a symmetrical shape. The Female distribution has a higher mean Comp AVG Grade with a 3.46 and standard deviation of 0.505 compared to males with a mean of 3.19 and standard deviation of 0.621.



Summary Statistics for Gender:

Group Type	Min	1st Qu.	Median	3rd Qu.	Max	Mean	SD
Female	2.0	3.2	3.7	3.9	4	3.46	0.505
Male	1.9	2.9	3.3	3.7	4	3.19	0.621

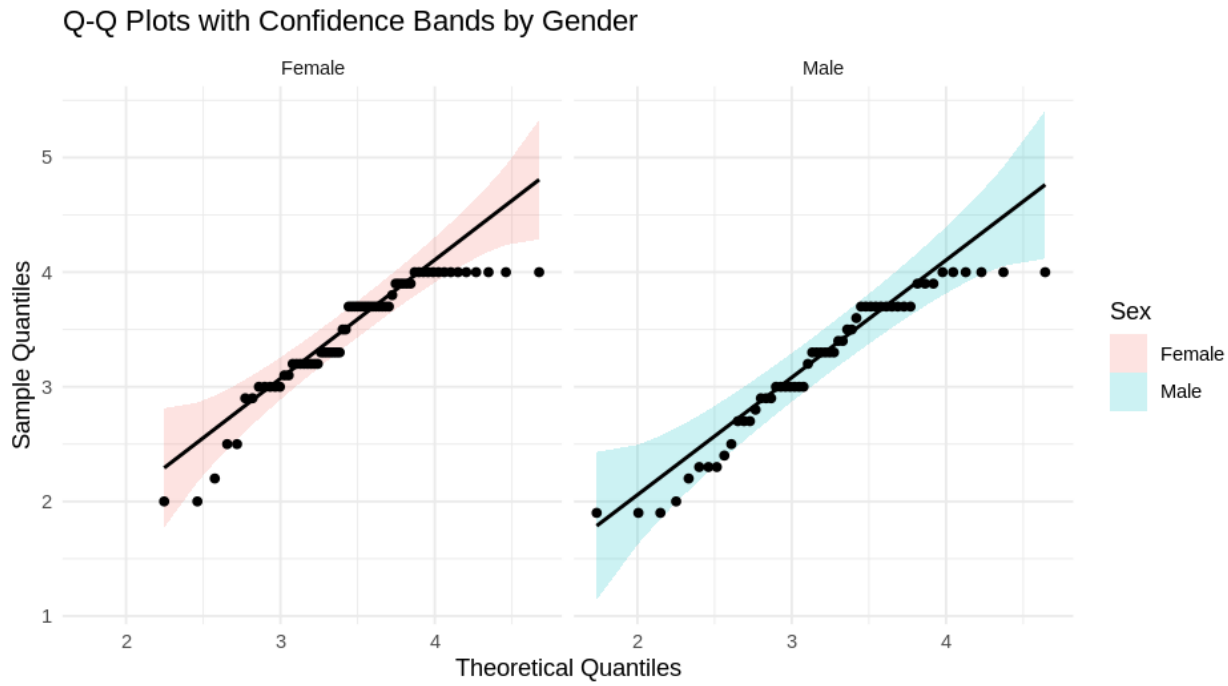
Before conducting an independent two sample t-test, we formed the following model:

$$y_{ij} = \mu + \tau_i + \varepsilon_{ij};$$

where y_{ij} is the Comp AVG Grades, μ is the mean effect of the model, τ_i is the gender coefficient (1 = Female, 0 = Male) and ε_{ij} is the error term for the equation. We generated a QQ plot from the model and noticed there is a pattern. The female distribution plot has points outside of the confidence bands of each of them. This plot also looks concave up as some points are on the same y-axis that shift to the right. The same can be said about the male QQ plot. Only two points are outside of the confidence bands, and the points have a similar pattern to the female distribution. A Shapiro-wilk test confirms that that data itself is not normal. Both female and male distributions have a p-value less than the 0.05 level so the parametric independent 2-sample t-test is not valid.

Shapiro-Wilk Test Gender Q04 Results:

Group	p-value
Female	0.0000111
Male	0.00291



Since normality fails, we will use the nonparametric Mann-Whitney Test. This test is used when comparing two independent variables where the dependent variable is continuous such as the Comp AVG Grade. We formed the following hypothesis and got the following test statistics and results. Since the p-value is less than the 0.05 level, we reject the null hypothesis. We can conclude that the distributions of the Comp AVG Grades are different between males and females.

- **(H_0):** The distributions of Comp AVG Grade are the same for males and females.
- **(H_1):** The distributions of Comp AVG Grade are different between males and females.

Mann-Whitney Test Results:

Test Statistic (W)	p-value
2257.5	0.01268

3.5) Question 5. We are asked to test if "there is a difference in Gender (Column B) gaps depending on the Sequence (Column D) students took?". We will use Gender and Sequence (Group; refer to Question 1 and 2) columns from the dataset to test if the Gender Gaps depend on the Sequence. For this question, we will use a Chi-Square Test of Independence since both variables are categorical.

Since we are doing a Chi-Square Test for two categorical variables, we do not need to test for normality but will create a contingency table. It is also worth noting that for Gender and Group, they both fail the normality assumptions from the previous questions. A contingency table is created based on the different levels of each variable. Gender has 2 levels (Female and Male), and Sequence (1-quarter, 2-quarter, and 3-quarter). A (2 x 3) The contingency table was created from a 120-student sample.

Gender/Group	1-quarter	2-quarter	3-quarter	Total
Female	29	22	14	65
Male	28	16	11	55
Total	57	38	25	120

From there we conducted the Chi-Square Test of Independence with the hypothesis:

- H_0: There is no association between the Gender and Group
- H_A: There is an association between the Gender and Group

Chi-Square Test of Independence Q05 Results:

Test statistic	p-value
0.49502	0.7807

where we obtained the test statistic of 0.49502 with a p-value of 0.7807. Since the p-value is greater than the 0.05 level, we do not reject the null hypothesis and conclude that there is no association (difference) in the gender gaps on the sequences (Groups) a student takes.

3.6) Question 6. We are asked "Does the URM status (Column C; 1 = Yes, 2 = No) vary based on the Group (Column E)? Which URM status is more likely to follow the recommendation?". We will use URM (Underrepresented Minority; Column C) and Sequence (Column E) columns in the dataset to test if URM Status is related to the Group. For this question, we will also use a Chi-Square Test of Independence since both variables are categorical.

The two categorical variables for this section also have different levels. A contingency table is created based on the URM status which has 2 levels (Yes and No) and Group (shorter, longer, and yes). The following a (2 x 3) contingency table was created from the 120-student sample.

URM/Group	Longer	Shorter	Yes	Total
Yes	8	20	30	58
No	7	36	19	62
Total	15	56	49	120

From there we constructed the Chi-Square Test of Independence with the hypothesis:

- H_0: There is no association between the URM Status and Group
- H_A: There is an association between the URM Status and Group

Chi-Square Test of Independence Q06 Results:

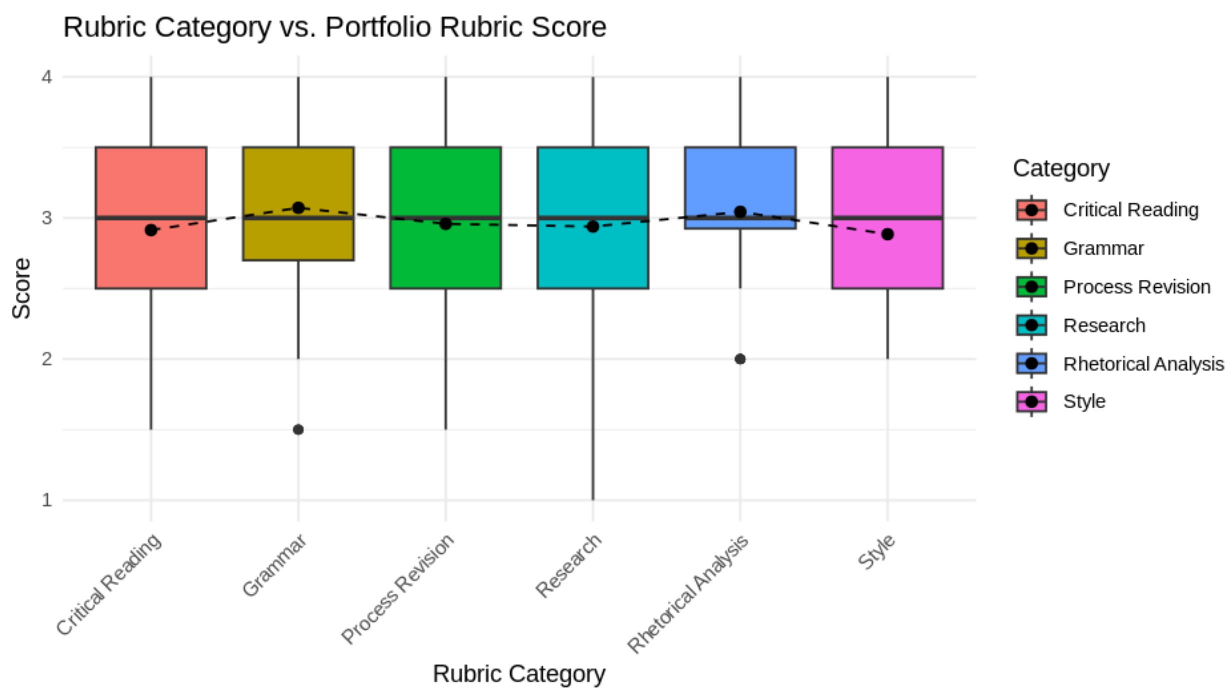
Test statistic	p-value

6.9819	0.03047

where we obtained the test statistic of 6.9819 with a p-value of 0.03047. Since the p-value is less than the 0.05 level, we reject the null hypothesis and conclude that there is an association (difference) between URM status and the Group a student takes.

3.7) Question 7. We are asked "Is there any statistical relationship between Gender (Column B) gaps and the 6 specific Portfolio Rubric Category Scores (process revision, critical reading, rhetorical analysis, research, style, grammar) (Columns F–K)? In other words, are there significant differences between males and females in areas like "critical reading" or "rhetorical analysis" for instance? The variables involved in this question are the Portfolio Rubric Category Score, Portfolio Rubric Category (Columns F-K) and Gender. Our goal for this question is to test the difference in the Scores from 6 Rubric Categories by the Genders. We will conduct this test via two-way ANOVA if normality is satisfied.

Prior to our testing of our two-way ANOVA, we depicted the boxplots for each of the Rubric Categories from the PTS values that include both Genders and for each of the Genders individually. For the boxplot for all genders, the Rubric Categories: Critical Reading, Process Revision, Research and Style are symmetrical. Grammer and Rhetorical Analysis have a skewed distribution with each graphical representation having only one outlier.



Summary Statistics for All Rubric Categories:

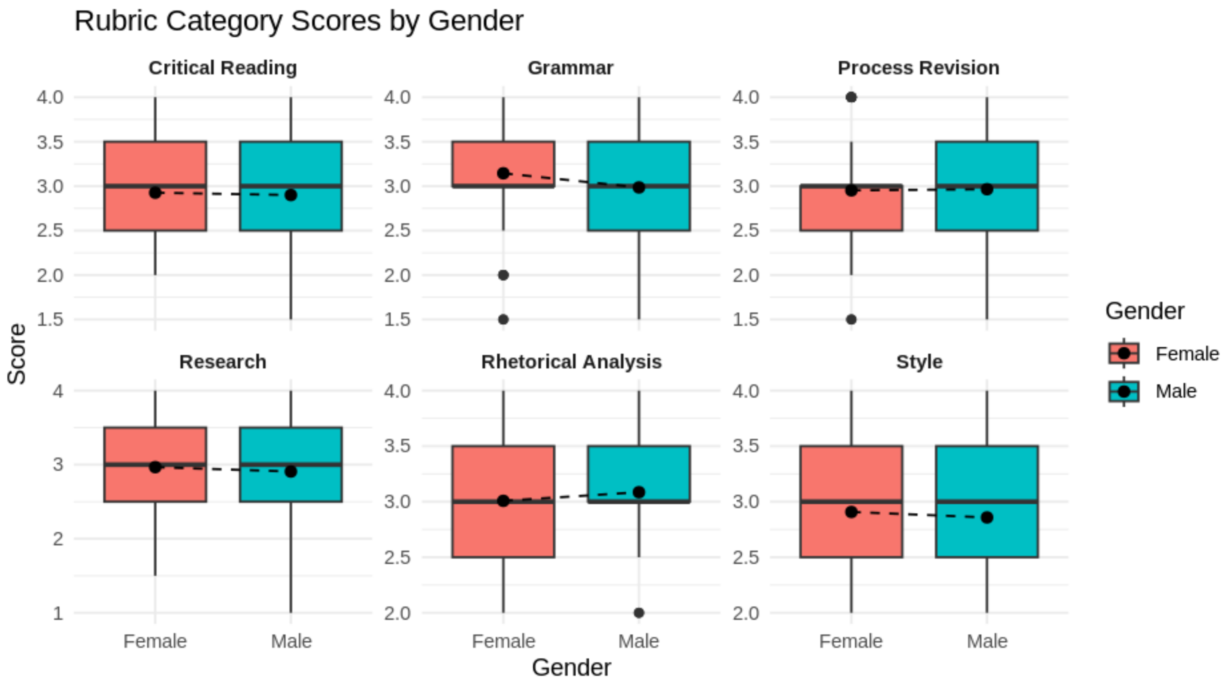
Rubric Category	Min	1st Qu.	Median	3rd Qu.	Max	Mean	SD
Critical Reading	1.5	2.5	3	3.5	4	2.91	0.583
Grammer	1.5	2.7	3	3.5	4	3.07	0.582

Process Revision	1.5	2.5	3	3.5	4	2.96	0.557
Research	1	2.5	3	3.5	4	2.94	0.621
Rhetorical Analysis	2	2.92	3	3.5	4	3.04	0.470
Style	2	2.5	3	3.5	4	2.88	0.564

For the boxplot of their respective genders, the Rubric Categories: Critical Reading, Research and Style are symmetrical. In these three categories, female students tend to have a higher score compared to the males in each of these categories. This trend is also the same for the Grammer Category however there is difference in the female and male graphical displays. Female students have a smaller SD but has 2 outliers with a higher mean score of 3.14, whereas the male students have a wider range on a symmetric graphical representation at a mean score of 2.99.

Other categories such as Process Revision and Rhetorical Analysis continue the skewed data representation. Female students have a skewed representation for Process Revision with a mean of 2.95, where most score at a range of 2.5 or 3.0 Score. One female student scored a 4.0 whereas another scored a 1.5. For male students, its primarily symmetric but has a wider SD compared to females. For Research, female students have a symmetrical representation and a higher SD. Males on the other hand is skewed but score higher with a mean score of 3.09 (compared to females, 3.01). Only one male student scored a 2.0 in the skewed representation of Rhetorical Analysis.

For Grammer, female students objectively score higher than males however the female students' graphical representation for Grammer is skewed to the right. The representation has two outliers outside of its lower tail whereas males do not. The male student representation has a wider range compared to Females but is generally symmetric. To further test it we created two-way interaction plots to see if there are any interactions. There appears to be some lines crossing each other in the respective categories which suggest there may be some interaction.

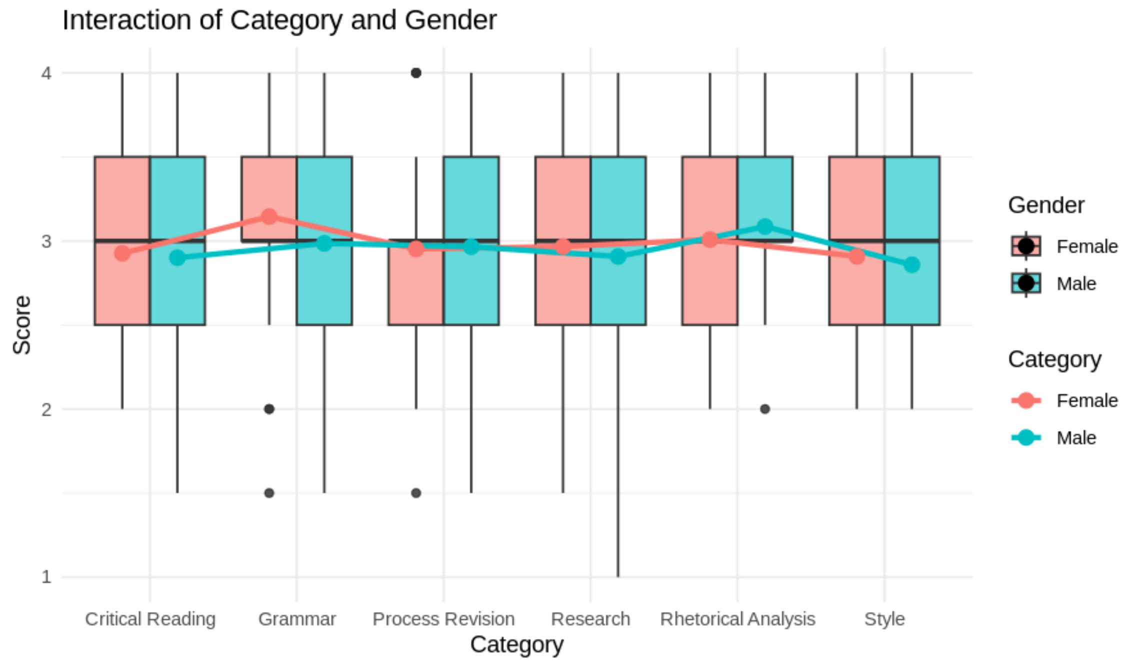


Female Summary Statistics for All Rubric Categories:

Category	Min	1st Qu.	Median	3rd Qu.	Max	Mean	SD
Critical Reading	2	2.5	3	3.5	4	2.93	0.579
Grammer	1.5	3	3	3.5	4	3.14	0.593
Process Revision	1.5	2.5	3	3	4	2.95	0.583
Research	1.5	2.5	3	3.5	4	2.97	0.622
Rhetorical Analysis	2	2.5	3	3.5	4	3.01	0.483
Style	2	2.5	3	3.5	4	2.91	0.583

Male Summary Statistics for All Rubric Categories:

Category	Min	1st Qu.	Median	3rd Qu.	Max	Mean	SD
Critical Reading	1.5	2.5	3	3.5	4	2.90	0.591
Grammer	1.5	2.5	3	3.5	4	2.99	0.561
Process Revision	1.5	2.5	3	3.5	4	2.97	0.531
Research	1	2.5	3	3.5	4	2.91	0.623
Rhetorical Analysis	2	3	3	3.5	4	3.09	0.455
Style	2	2.5	3	3.5	4	2.86	0.597

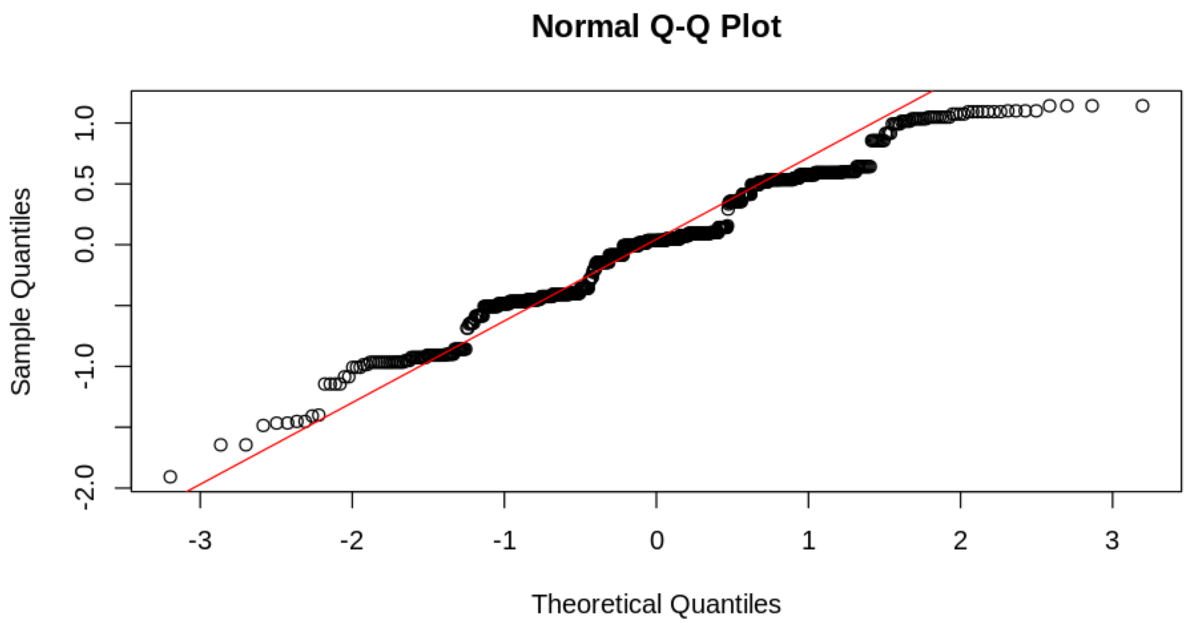


We constructed a model for this question to test normality. The following linear model is the one used for this project.

$$y_{ij} = \mu + \gamma_i + \tau_j + (\gamma\tau)_{ij} + \varepsilon_{ij}$$

Where y_{ij} is the Portfolio Rubric Score, μ is the mean effect on the model, γ_i is the Portfolio Rubric Category (i = 1 (Critical Reading), 2 (Grammer), 3 (Process Revision), 4 (Research), 5 (Rhetorical Analysis), and 6 (Style)), τ_j is the Gender (1 = Female, 0 = Male), $(\gamma\tau)_{ij}$ is the interaction effect between Rubric Category and Gender and ϵ_{ij} is the error term for this equation.

We generated a QQ plot to see if the data is normal. The representation generated for the data between the Portfolio Rubric Category Score, Portfolio Rubric Category and Gender, the data looks symmetrical with a very heavy upper tail. We continued with Shapiro-Wilk test which suggested the data is not normal. Due to this conclusion, we did not conduct the parametric two-way ANOVA testing and instead do nonparametric testing with Aligned Rank Transform (ART) ANOVA.



Shaprio-Wilk Test Q07 Results:

W test statistic	p-value
0.97692	3.093e-09

Aligned Rank Transform (ART) ANOVA is a nonparametric alternative to traditional ANOVA where it's useful to deal with ordinal, non-normally distributed data that violates homoscedasticity. For ART ANOVA, we ranked each of the variables and then conducted our analysis. Once we ranked them, we form the following hypotheses and obtained the results based on our equation:

$$y_{ij} = \gamma_i + \tau_j + (\gamma\tau)_{ij} + \varepsilon_{ij}\,;$$

where y_{ij} is the Score, γ_i is the Rubric Category, τ_j is the Gender, and $(\gamma\tau)_{ij}$ is the interaction effect between Rubric category and Gender and ϵ_{ij} is the error term.

- H_0: The factor in question is zero for every i and j has no linear relationship with the response variable
- H_A: The coefficient term in question is nonzero for at least one i or j has a linear relationship with the response variable

ART ANOVA Q07:

	df	Df.res	F value	Pr(>F)
Category	5	708	1.81517	0.10759
Gender	1	708	0.72462	0.39492
Category:Gender	5	708	0.89395	0.48457

Based on the following results, none of the p-values are significant as well as the interaction between category and Gender. We can conclude that there is no relationship between Rubric Category and Gender. For Critical Reading and Rhetorical Analysis, there are not significant comparisons when looking at the post-hoc test (see appendices).

Other questions our client would like to explore are the relationships in questions 8 and 9 which focus on the 6 Portfolio Rubric Category Scores (Columns F-K). These questions check to see if the students performed high or low on their Scores.

3.8) Question 8. We are asked "Are there significant gender differences in the six Portfolio Rubric Category Scores (Process Revision, Critical Reading, Rhetorical Analysis, Research, Style, Grammar)?" The variables used in question are the Portfolio Rubric Category Score (Column and the Portfolio Rubric Category). Our goal for this question is to test the difference in the Scores from six Rubric Categories for all students combined using one-way ANOVA.

The box plots for all the Rubric Categories are shown in the previous section 3.7. What we will show is the QQ plots for all Categories (Process Revision, Critical Reading, Rhetorical Analysis, Research, Style, Grammar) in their respective QQ plots. For the model, we constructed the following equation:

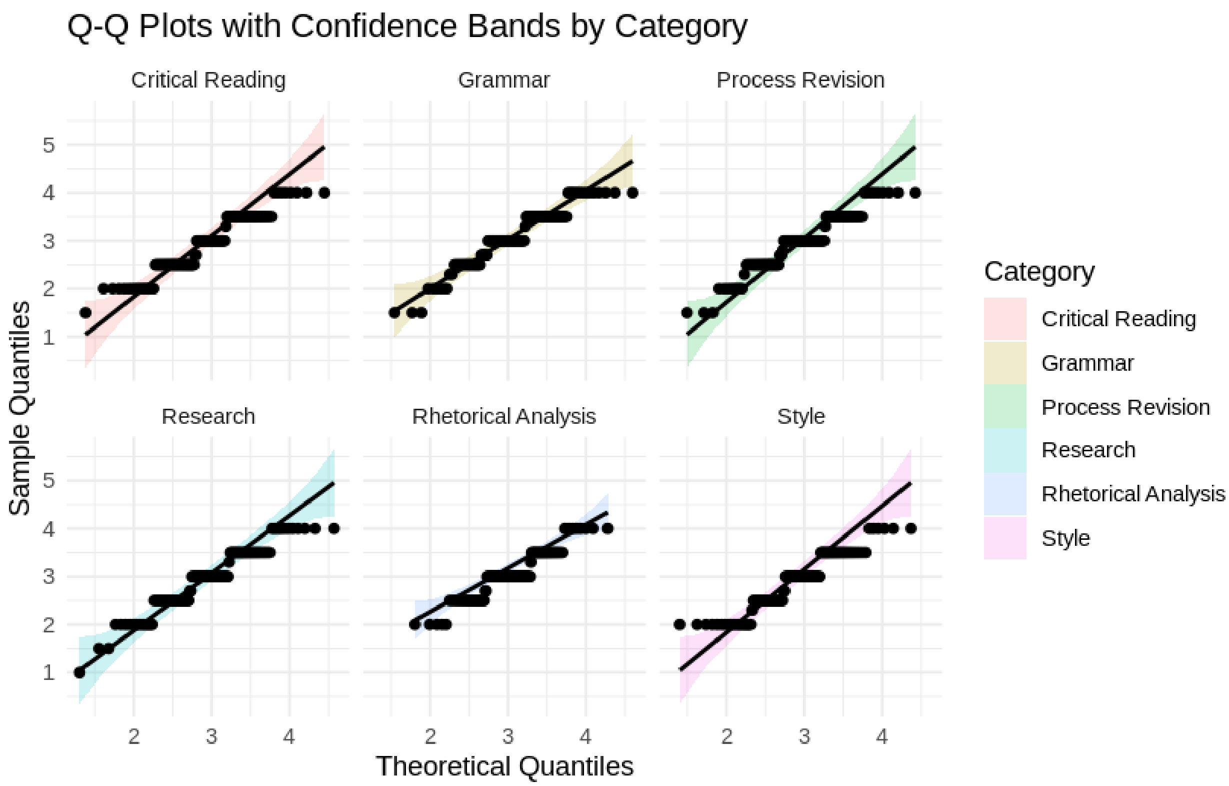
$$y_{ij} = \mu + \gamma_i + \varepsilon_i;$$

where $y_i$is the Score, μ is the mean of between the γ_i factor (Rubric Category i = 1, 2, ... , 6) and ε_i is the error term. Depicted below are the QQ plots for each of the Categories. Each graphical representation shows a staircase plot point where there are multiple points plotted on the same y-axis, and there is a jump to the next line. This depicts a pattern, and the Shapiro-Wilk test confirms that this is nonnormal. Because of this, we do not use the parametric one-way ANOVA for our procedure and conduct the nonparametric Kruskal-Wallis test.

Shapiro-Wilk Test Q08 Results:

Category	p-value
Critical Reading	0.00000557
Grammer	0.00000892
Process Revision	0.0000124
Research	0.0000322

Rhetorical Analysis	0.000000289
Style	0.00000101



For Kruskal-Wallis, we construct the following equation:

$$y_i = \gamma_i + \varepsilon_i;$$

where y_i is the Score, the factor γ_i (Rubric Category i = 1, 2, ... , 6) and ε_i is the error. We form the following hypotheses and obtain the Kruskal-Wallis results from our equation. The Kruskal-Wallis test fails to reject the null hypothesis. The median distributions of the Rubric Category Scores are the same across all categories. There does not appear to be any difference in the distributions.

- H_0: The median distributions of Rubric Category Scores are the same across all Categories.
- H_A: At least one median distribution Categories is different.

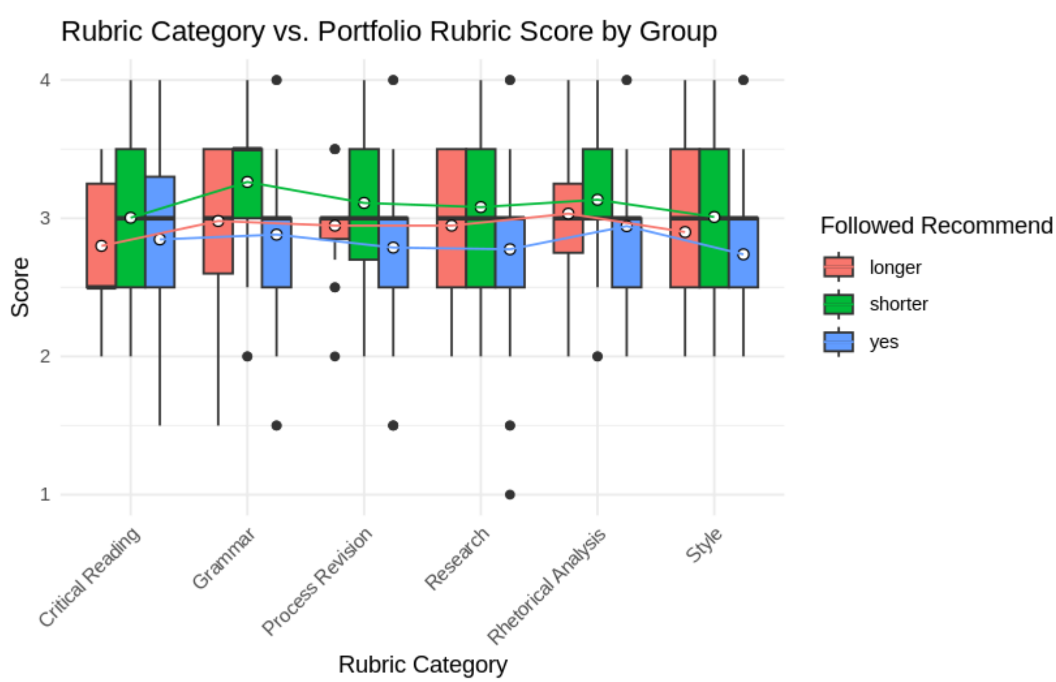
Kruskal Wallis Test for Q08:

Test statistic	df	p-value
10.091	5	0.0727

3.9) Question 9. Question 9 asks us to "Across all students, are there statistically significant differences between the six Portfolio Rubric Category Scores?" Our Goal for this question is to test the difference in the Scores from six Rubric Categories by three Groups. With that in mind, the variables: Portfolio Rubric Category Score, Portfolio Rubric Category, and Group (Sequence) will be used along with the two-way ANOVA technique. We will use two-way ANOVA (since there are 6 Rubric Categories and 3 Sequences) if the model is satisfied by normality.

For the boxplot and summary statistics for all Rubric Category Scores, refer to section 3.7. As for the individual boxplots for all the sequences there are notable outliers for all categories except for Critical Reading. There is also at least one symmetrical distribution for each of the groups in their respective Rubric Cateogry. These distributions with outliers look heavily skewed to either to the right or left except for Critical Reading's "longer" group. An example of a heavily skewed distribution is the Process Revision. For the "longer" group, the distribution is heavily skewed to the left where there are 3 outliers (one above the upper tail and two below the lower tail) and similarly of the "yes" group with 2 outliers (one above and below the upper tails).

In terms of two-way interactions between the Rubric Category and the Group, all the boxplots and interactions tend to be parallel with one another. We see the interaction lines connecting the means of each boxplot increase and decrease with one another. Since they are parallel, there does not seem to be any interaction between the groups and the respective categories.



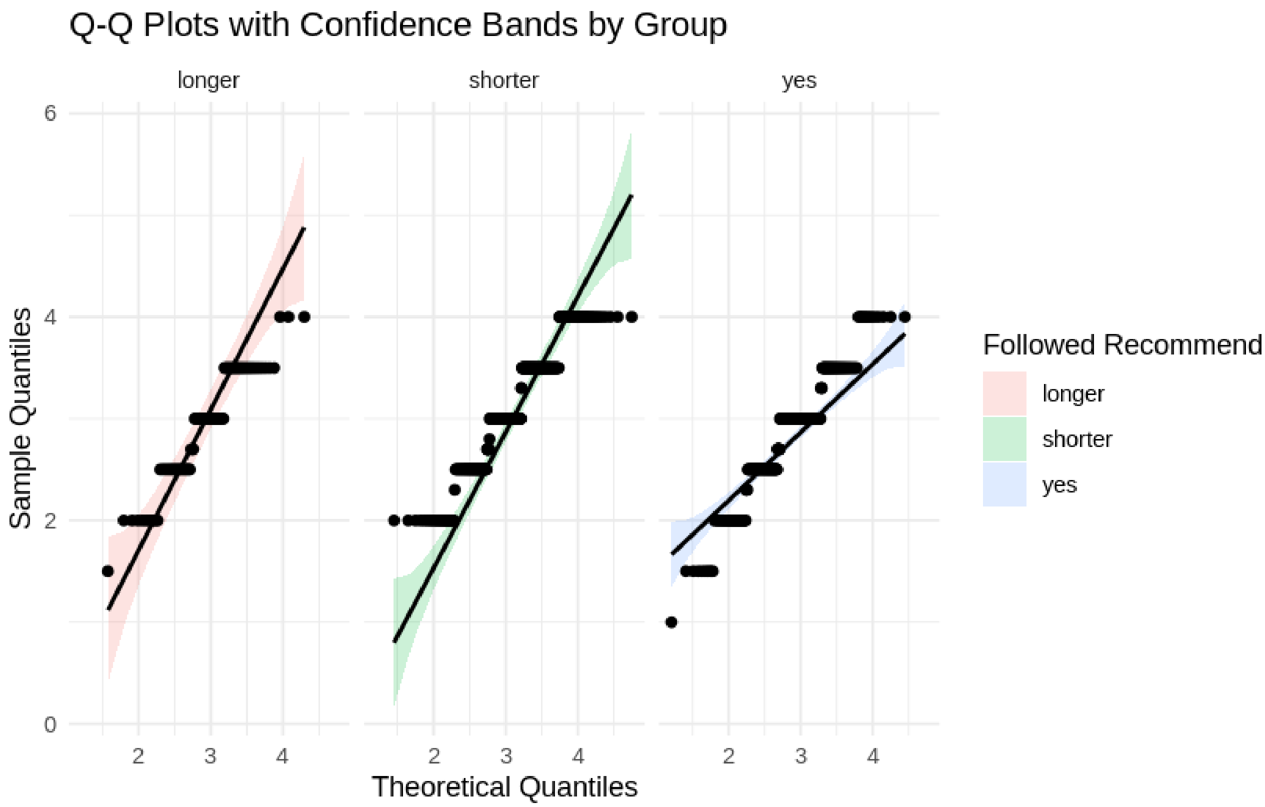
Summary Statistics for Rubric Category by Group:

Category	Group	Min	Q1	Median	Q3	Max	Mean	SD
Critical Reading	longer	2	2.5	2.5	3.25	3.5	2.80	0.528
Critical Reading	shorter	2	2.5	3	3.5	4	3.00	0.608
Critical Reading	yes	1.5	2.5	3	3.3	4	2.85	0.564
Grammer	longer	1.5	2.6	3	3.5	3.5	2.98	0.632
Grammer	shorter	2	3	3.5	3.5	4	3.26	0.552
Grammer	yes	1.5	2.5	3	3	4	2.88	0.540
Process Revision	longer	2	2.85	3	3	3.5	2.95	0.405
Process Revision	shorter	2	2.7	3	3.5	4	3.11	0.540
Process Revision	yes	1.5	2.5	3	3	4	2.97	0.575
Research	longer	2	2.5	3	3.5	3.5	2.95	0.485
Research	shorter	2	2.5	3	3.5	4	3.08	0.596
Research	yes	1	2.5	3	3	4	2.78	0.654
Rhetorical Analysis	longer	2	2.75	3	3.25	4	3.03	0.550
Rhetorical Analysis	shorter	2	3	3	3.5	4	3.13	0.475
Rhetorical Analysis	yes	2	2.5	3	3.5	4	2.94	0.425
Style	longer	2	2.5	3	3.5	4	2.90	0.660
Style	shorter	2	2.5	3	3.5	4	3.01	0.536
Style	yes	2	2.5	3	3	4	2.74	0.540

For testing for normality, we formed the linear model:

$$y_{ij} = \mu + \gamma_i + \alpha_j + (\gamma\alpha)_{ij} + \varepsilon_{ij};$$

where y_{ij} is the Scores, μ is the mean effect between the factors: γ_i (Rubric Category), α_j (Group), the interaction term between Category and Group $(\gamma\alpha)_{ij}$, and ϵ_{ij} is the error term. For the plots each of the points for the Groups (Followed Recommended) show a staircase pattern. The Shapiro Wilk test for each of the groups reveals the data is not normal for each of the groups. We do not conduct two-way ANOVA testing and instead ART Anova is done.



Shapiro Wilk Q09 Results:

Group	p-value
1	0.00340
2	0.00340
3	0.00340

For ART ANOVA, we ranked each of the variables and then conducted our analysis based on the formed following hypothesis and equation:

$$y_{ij} = \gamma_i + \alpha_j + (\gamma\alpha)_{ij} + \varepsilon_{ij} \ ;$$

Hypothesis:

- H_0: The factor in question is zero for every i and j has no linear relationship with the response variable
- H_A: The coefficient term in question is nonzero for at least one i or j has a linear relationship with the response variable

Based on the results of ART ANOVA, the Group variable is significant. The Cateogry variable and the interaction between Category and Group are not significant since they failed the 0.05

level. Post-hoc testing revealed the comparisons between “longer” and “shorter” groups with the comparison between “shorter” and “yes” groups are significant. From this we can conclude there is significance in the Scores between each of the groups. Post hoc test reveals the “shorter” group have a great significance in the Rubric Cateogry Scores.

ART ANOVA Q09 Results:

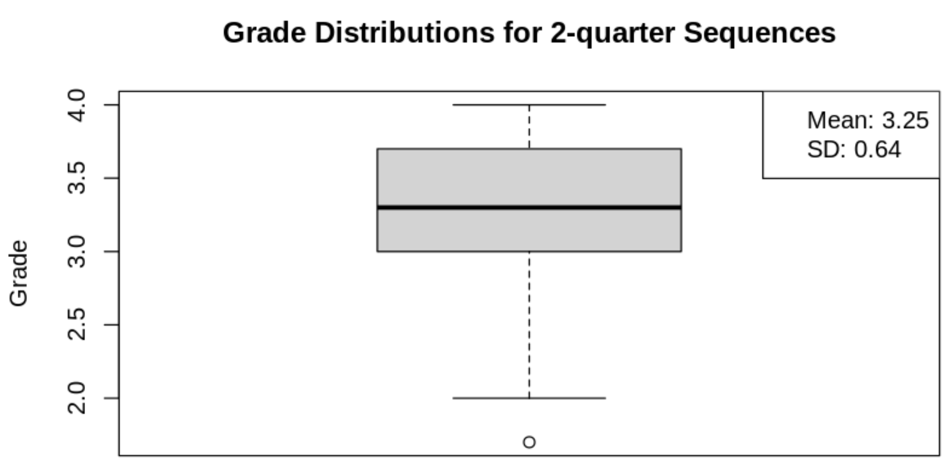
	Df	Df.res	F value	Pr(>F)
Category	5	702	1.47474	0.19582
Group	2	702	20.66223	1.9085e-09
Category:Group	10	702	0.34109	0.96968

Post Hoc Test Group Results:

Contrast	estimate	SE	df	t.ratio	p.value
Longer - shorter	-64.9	24.1	702	-2.687	0.0221
Longer - yes	38.9	24.5	702	1.587	0.3386
Shorter - yes	103.8	16.2	702	6.388	<0.001

For questions 10 through 13, our client asks if there is any evidence to suggest that students improve their Grades (Columns M–R) from one course to the next in a 2- and 3-quarter sequences.

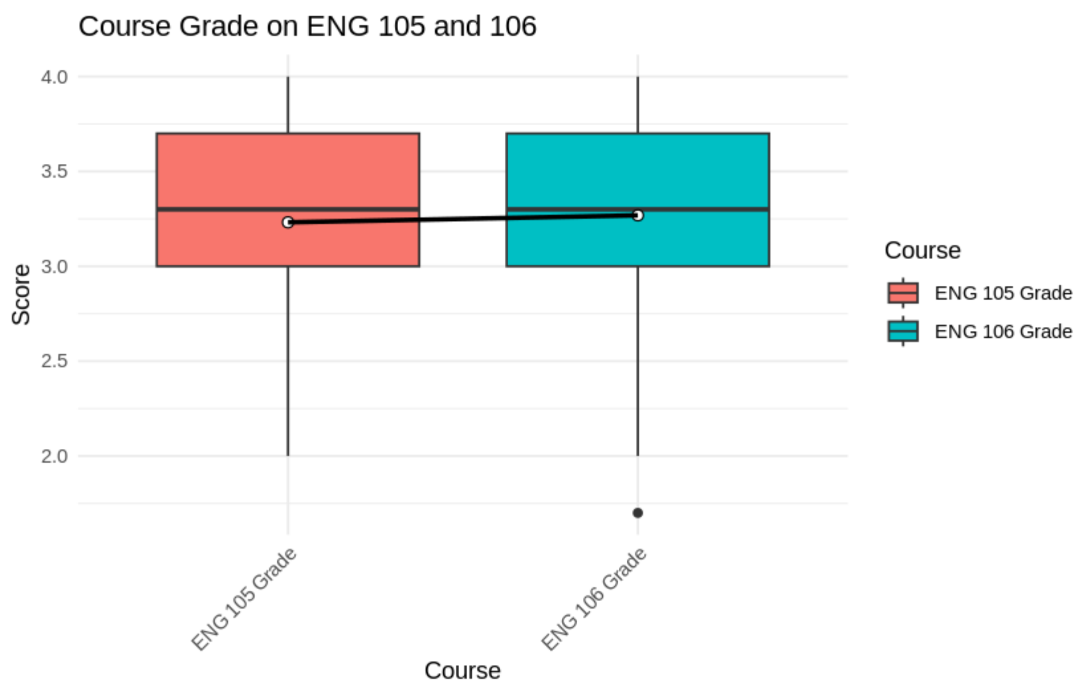
3.10) Question 10. Question 10 asks “Among students in the 2-quarter sequence, is there a significant improvement in grades from ENG 105 to ENG 106”? The variables used for this question are the ENG 105 and 106 Grades (Dependent variables) and the Followed Recommended (Course; independent variable). Our goal for this question is to test if Students’ Grades improved in the 2-quarter Sequences (all combined) via a two-sample t-test since we have two continuous variables.



Summary Statistics for 2-quarter Sequences:

Min.	1st Qu.	Median	3rd Qu.	Max	Mean	SD
1.7	3.0	3.3	3.7	4	3.25	0.642

Based upon the boxplot for all 2-quarter sequences, we have a symmetrical distribution with one outlier outside of the lower tail. Overall, the plot does not have anything else out of the ordinary. For the individual Courses, ENG 105 and ENG 106, we see the individual distributions for each of the courses (see next page). The ENG 106 has a higher mean grade of 3.27 which is slightly higher than the ENG 105 mean grade of 3.23. The individual box plot reveals where the lower outlier comes from. ENG 106 has the lower outlier compared to ENG 105 distribution which does not.



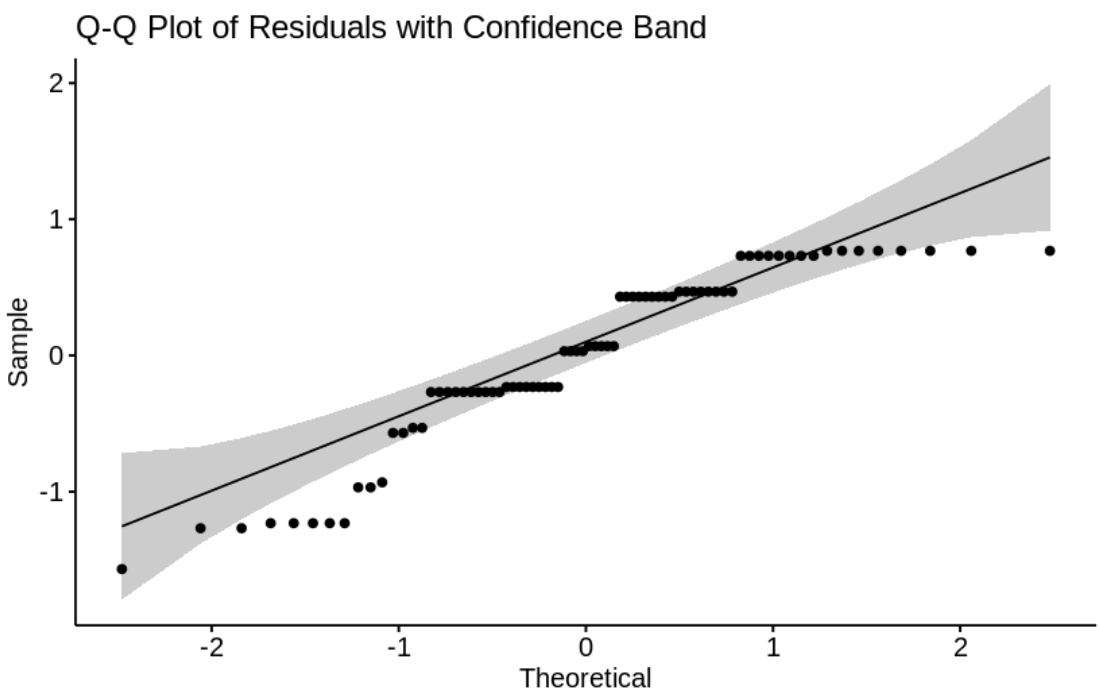
Summary Statistics for ENG 105 and 106:

2-quarter Grade	Min	1st Qu.	Median	3rd Qu.	Max	Mean	SD
ENG 105	2.0	3.0	3.3	3.7	4	3.23	0.658
ENG 106	1.7	3.0	3.3	3.7	4	3.27	0.634

For this question, we formed the linear model:

$$y_i = \mu + \alpha_i + \epsilon_i;$$

where y_i is the Grades for the courses: ENG 105 and ENG 106, μ is the mean effect between the Grade and the Course (α_i , i = 1 (Eng105), and 2 (ENG 106) and ϵ_i is the error term. A generated QQ plot reveals the data has a staircase pattern with two heavy tails. A Shapiro-Wilk test confirms this model does not have a normal distribution. The p-value is less than the 0.05 level so we consider a nonparametric Wilcoxon test to compare the two Grades.



Shapiro Wilk Q10 Results:

V test statistic	p-value
0.94044	0.04329

We conducted the nonparametric Wilcoxon test by comparing the medians of the ENG 105 and 106 Grades. This test is two-tailed and ignores the normality assumption. The following hypothesis is written for the comparison between ENG 105 and 106. From this test, we obtained a test statistic of 184 with a p-value of 0.9126. We do not reject the null hypothesis which states the median difference between ENG 105 and 106 grades is zero. There does not appear to be any grade improvement in the 2-quarter sequences.

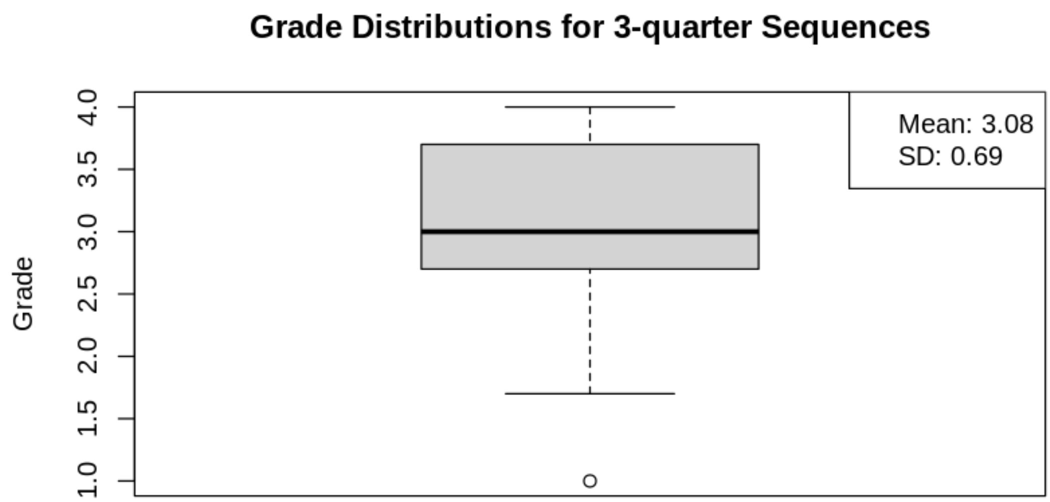
- H_0: The median difference between ENG 105 and ENG 106 grades is zero.
- H_A: The median difference between ENG 105 and ENG 106 grades is not zero.

Wilcoxon Test Q10 Results:

V test statistic	p-value

184	0.9126

3.11) Question 11. Among students in the 3-quarter sequence, is there a significant grade trend across ENG 100, ENG 101, and ENG 102? The variables used for this question are the ENG 101, ENG 101, and 102 Grades (Dependent variables) and the Followed Recommended (Course; independent variable). Our goal for this question is to test if Students' Grades improved in the 2-quarter Sequence (all combined) via a one-way repeated measure ANOVA. One-way repeated measure ANOVA is used if there is a student who took 2 quarter courses and got the same grades for each course. An example would be when a student scores a "3.0" in ENG 100, a "2.7" in ENG 101, and "3.0" in ENG 102. This student has two repeated grade measurements.

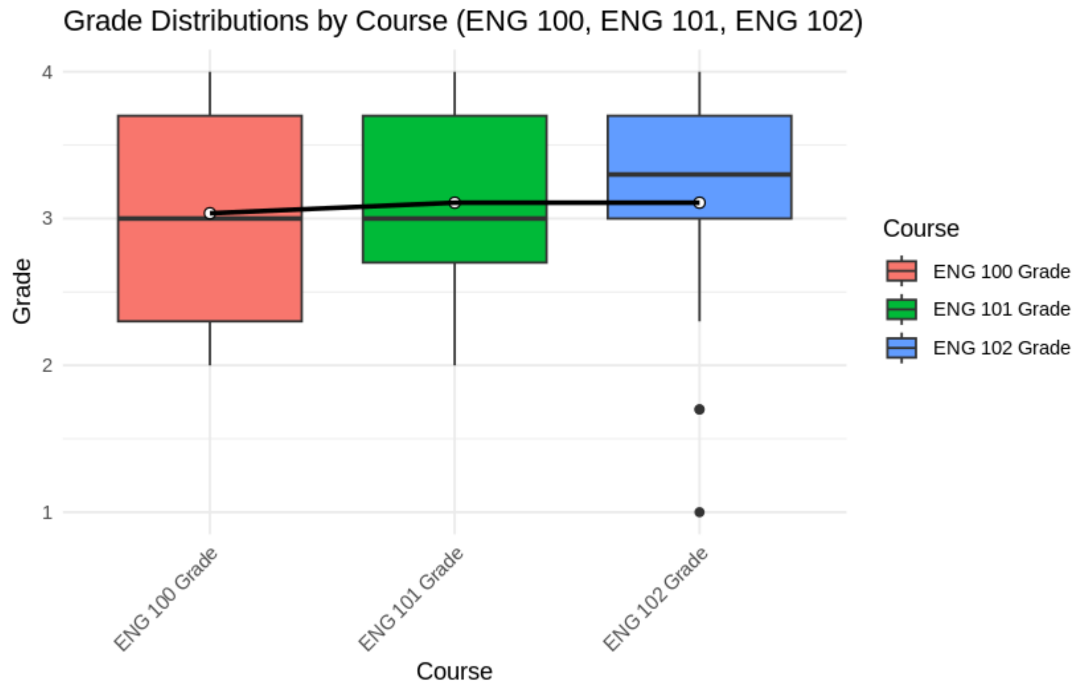


Summary Statistics for 3-Quarter Sequence Grades:

Min.	1st Qu.	Median	3rd Qu.	Max	Mean	SD
1.0	2.7	3.0	3.7	4	3.08	0.691

For the overall distribution for all 3-quarter sequences, the boxplot distribution looks skewed to the right. The overall mean Grade for the 3-quarter sequences is 3.08 and there is one outlier outside of the lower tail. For a deeper analysis we constructed individual boxplots that divide the 3-quarter sequences by their respective Courses (shown on the next page).

Of the distributions for ENG 100, 101, and 102, the ENG 101 and ENG 102 scores are higher in terms of the mean score. Both have a mean Grade of 3.11 whereas ENG 100 has a mean of 3.0. In terms of distributions, ENG 100 and ENG 102 have a symmetrical shape. The Boxplot for ENG 102, however, has two lower outliers where ENG 100 and 101 do not. For ENG 101 the distribution is skewed to the right which is the only outlier in terms of the three distributions of 3-quarter courses. It could be this shaping is what changes the distribution for the overall distribution for the 3-quarter sequences.



Summary Statistics for ENG 100, 101 and 102 Grades:

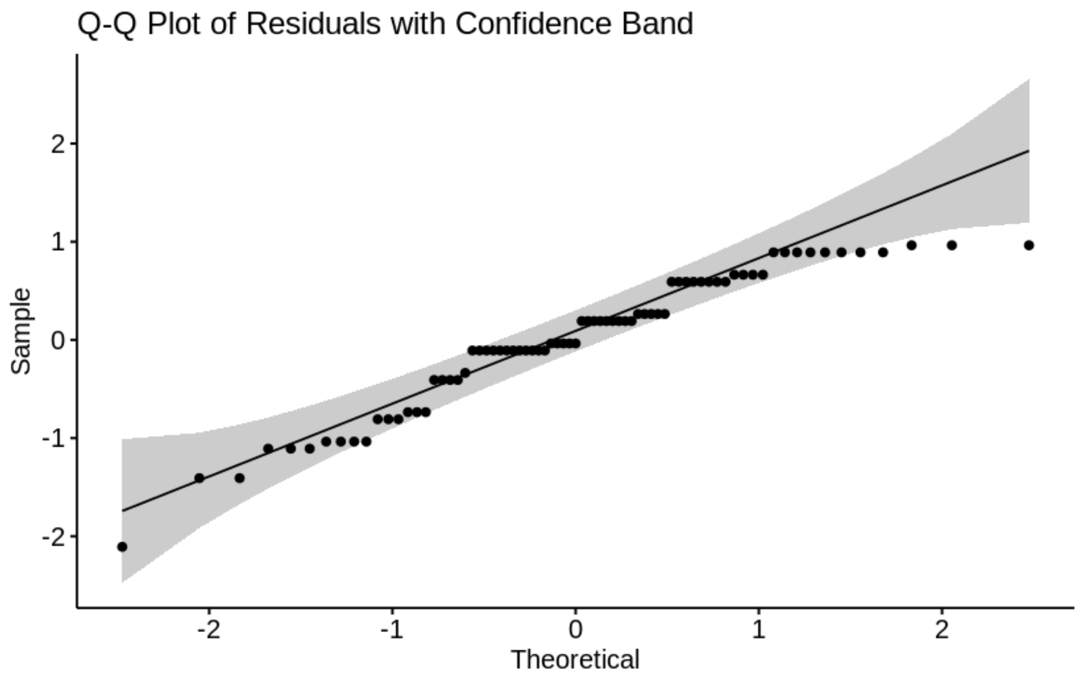
3-quarter Grade	Min	1st Qu.	Median	3rd Qu.	Max	Mean	SD
ENG 100	2.0	2.3	3.0	3.7	4.0	3.04	0.675
ENG 101	2.0	2.7	3.0	3.7	4.0	3.11	0.636
ENG 102	1.0	3.0	3.3	3.7	4.0	3.11	0.781

For the question, we will test normality with the following formed model:

$$y_i = \mu + \alpha_i + \epsilon_i;$$

where y_i is the Grades for the courses: ENG 101, ENG 101 and ENG 102, μ is the mean effect between the Grade and the Course (α_i where , i = 1 (Eng100), and 2 (ENG 101), and 3 (ENG 102)) and ϵ_i is the error term.

Based on the boxplot, the distribution looks like the previous QQ distributions. The points are approximately close to the drawn QQ line. As they increase, there is an upper tail where the points out of the normality line. The Shapiro Wilk test confirms that the model is not normal since the p-value is less than 0.05 level. We consider the nonparametric Friedman test as we are comparing 3 or more repeated samples in the sample. We have 3 categories (ENG 100, 101, and 102) so this is a suitable case.



Shapiro Wilk Q11 Results:

W test statistic	p-value
0.94006	0.001476

We conducted the Friedman test with the following hypothesis:

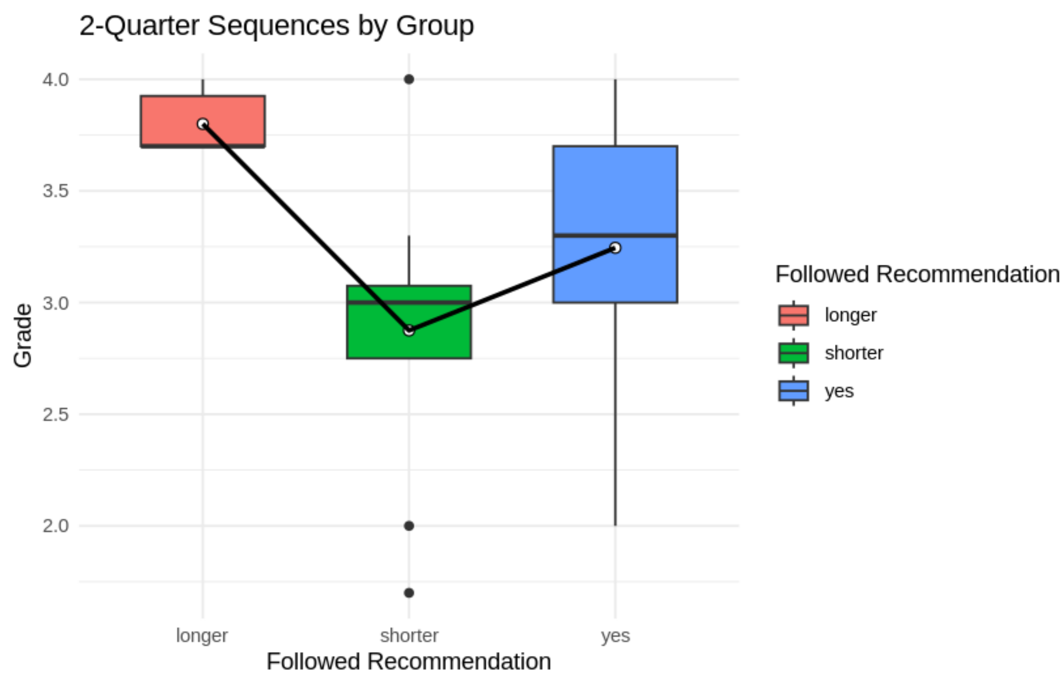
- H_0: There is no significant difference between the dependent Courses
- H_A: There is a significant difference between the dependent Courses

From the test, the calculated chi-squared (X^2) test statistics is 0.025974 with a p-value of 0.9871. This p-value is insignificant at the 0.05 level, so we do not reject the null hypothesis that there is no significant difference in the Grade variable across the different courses. There does not appear to be any grade improvement in the 3-quarter sequence by the followed recommendation group.

Friedman Test Q11 Results:

X^2 test statistic	df	p-value
0.025974	2	0.9871

3.12) Question 12. Question 12 asks "Does grade improvement from ENG 105 to ENG 106 in the 2-quarter sequence differ between Groups (Followed, Shorter, Longer)"? The variables used for this question are the ENG 105 and ENG 106 Grades (Dependent variables), a newly created independent variable Course (i = 1 (ENG 105) and 2 (ENG 106)), and the independent variable Followed Recommended (Group). Our goal for this question is to test if their Grades are improved in the 2-quarter Sequence by three Groups via two-way repeated measure ANOVA. Two-way repeated measure ANOVA is like one-way ANOVA except there are different levels for each of the independent variables.

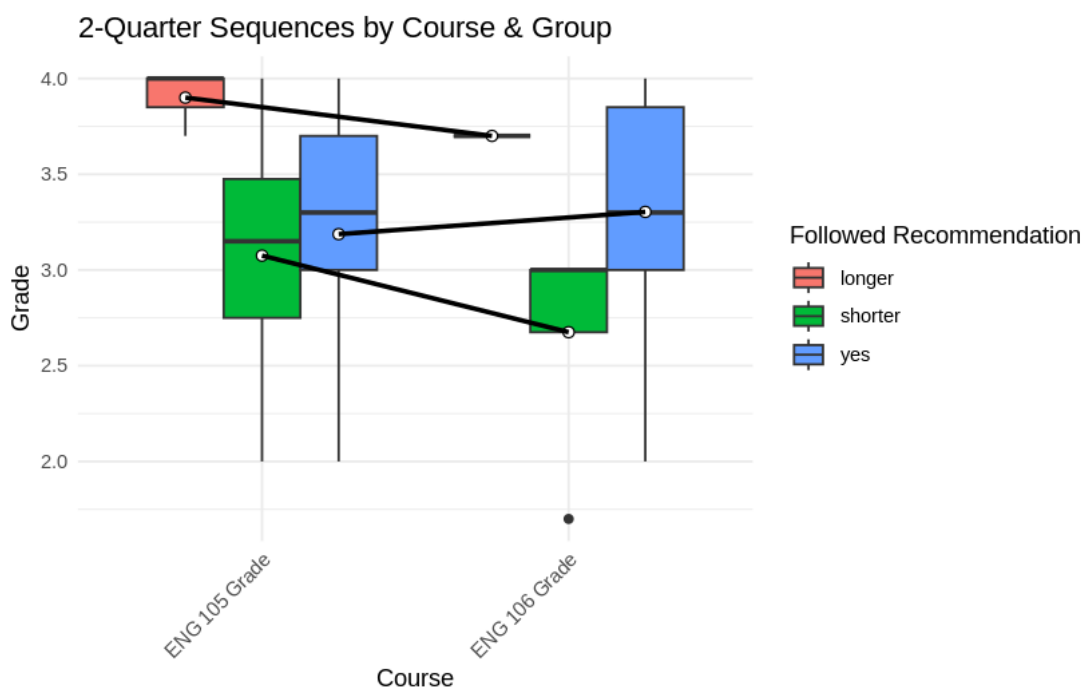


Summary Statistics for each Group:

Group Type	Min	1st Qu.	Median	3rd Qu.	Max	Mean	SD
Longer	3.7	3.7	3.7	3.9	4.0	3.80	0.155
Shorter	1.7	2.8	3.0	3.1	4.0	2.88	0.723
Yes	2.0	3.0	3.3	3.7	4.0	3.25	0.630

For the boxplots for this series, the “longer” group has the highest mean when compared to the mean score of the “yes” group and “shorter” Group. The “longer” Group has the lowest variation in the SD. When comparing distributions, this distribution is very skewed to the right despite having scores that are very high. As the lowest scoring group, the “shorter” group’s distribution is also skewed with 2 leftward outliers outside of the lower tail. For the “yes” group, it is symmetrical with no outliers.

For a deeper analysis, we separated the 2-quarter distribution by the course (ENG 105 and ENG 106). From the separation, the “longer” and “shorter” groups have higher grades in the ENG 105 course compared to the “yes” group that scored higher in the ENG 106 course. As for the distributions, ENG 105 still has a skewed distribution for the “longer” group while the “shorter” and “yes groups” have symmetrical distributions. For ENG 106, the is distribution is a single point (the students had the same grades). For the "shorter" and "yes" groups, the "shorter" distribution is skewed to the left with one lower outlier whereas the "yes" group's distribution is symmetrical.



Summary Statistics by Course (ENG 105 and 106) by Group:

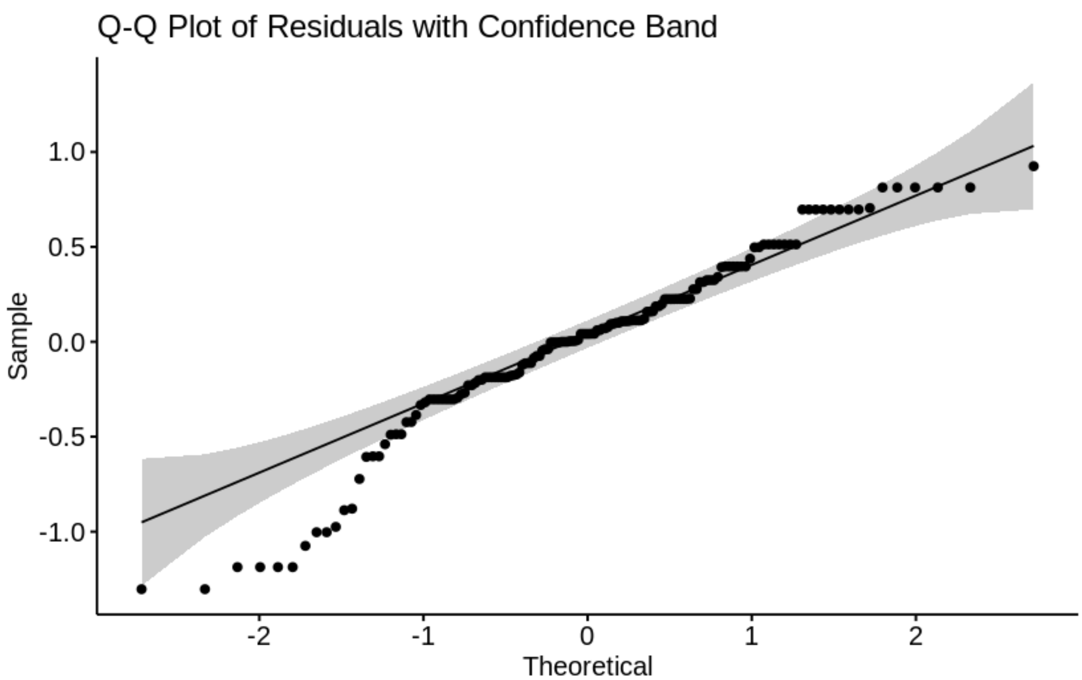
Course	Group	Min	Q1	Median	Q3	Max	Mean	SD
ENG 105	longer	3.7	3.9	4.0	4.0	4.0	3.90	0.173
ENG 105	shorter	2.0	2.8	3.2	3.5	4.0	3.08	0.830
ENG 105	yes	2.0	3.0	3.3	3.7	4.0	3.19	0.642
ENG 106	longer	3.7	3.7	3.7	3.7	3.7	3.70	0.000
ENG 106	shorter	1.7	2.7	3.0	3.0	3.0	2.68	0.650
ENG 106	yes	2.0	3.0	3.3	3.9	4.0	3.30	0.622

We conducted a two-way ANOVA and formed the linear regression model below between the Grades (the outcome) between the 2-quarter Courses, the Group and the interaction between the Courses and Groups.

$$y_{ij} = \mu + \alpha_i + \gamma_j + (\alpha\gamma)_{ij} + \epsilon_{ij};$$

y_{ij} is the 2-quarter course Grades, μ is the mean effect between the terms: Course (α_i, i = 1 (ENG 105), and 2 (ENG 106)), Group (γ_j , j = 1 (Longer), 2(shorter), and 3 (yes), and the interaction between Course and Grade $(\alpha\gamma)_{ij}$ with the error term: ϵ_{ij}.

A constructed QQ plot was generated to check the model's normality. The distribution looks normal but there is a heavy lower tail and the upper tail. A Shapiro-Wilk test was conducted, and the p-value is less than the 0.05 level. The data appears to be nonnormal, so the parametric Two-way repeated measure ANOVA was to be conducted. We employed the nonparametric ART ANOVA technique.



Shapiro Wilk Q12 Results:

W test statistic	p-value
0.95513	7.998e-05

For ART ANOVA, we ranked each of the variables and then conducted our analysis based on the formed following hypotheses and equation:

$$y_{ij} = \alpha_i + \gamma_j + (\alpha\gamma)_{ij} + \epsilon_{ij} \,;$$

Hypothesis:

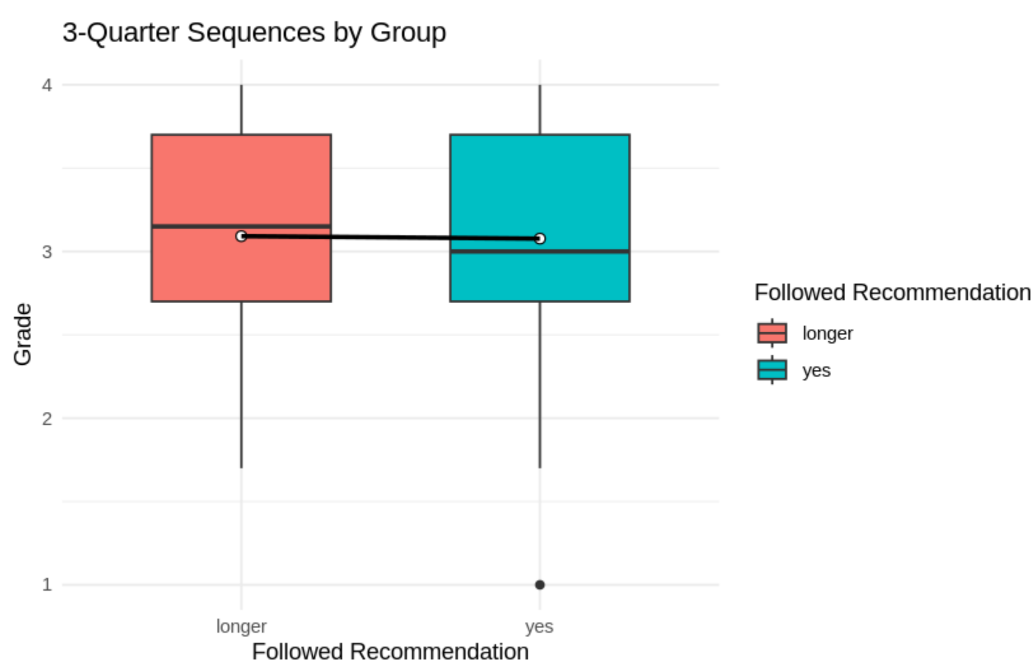
- H_0: The factor in question is zero for every i and j has no linear relationship with the response variable
- H_A: The coefficient term in question is nonzero for at least one i or j has a linear relationship with the response variable

Based on the results of ART ANOVA, none of the variables along with the interaction between the Course and Group have significance. All fail to reject the null hypotheses as they are greater than the 0.05 level. No post-hoc testing was conducted since there is no significance. As such we can conclude grade improvement in the 2-quarter sequence does not differ between the 3 Groups (Followed, Shorter, Longer)

ART ANOVA Q12 Results:

	Df	Df.res	F value	Pr(>F)
Course	1	35	0.0016218	0.96811
Followed Recommended	2	35	2.12653626	0.12983
Course:Followed Recommend	2	35	0.9977645	0.37895

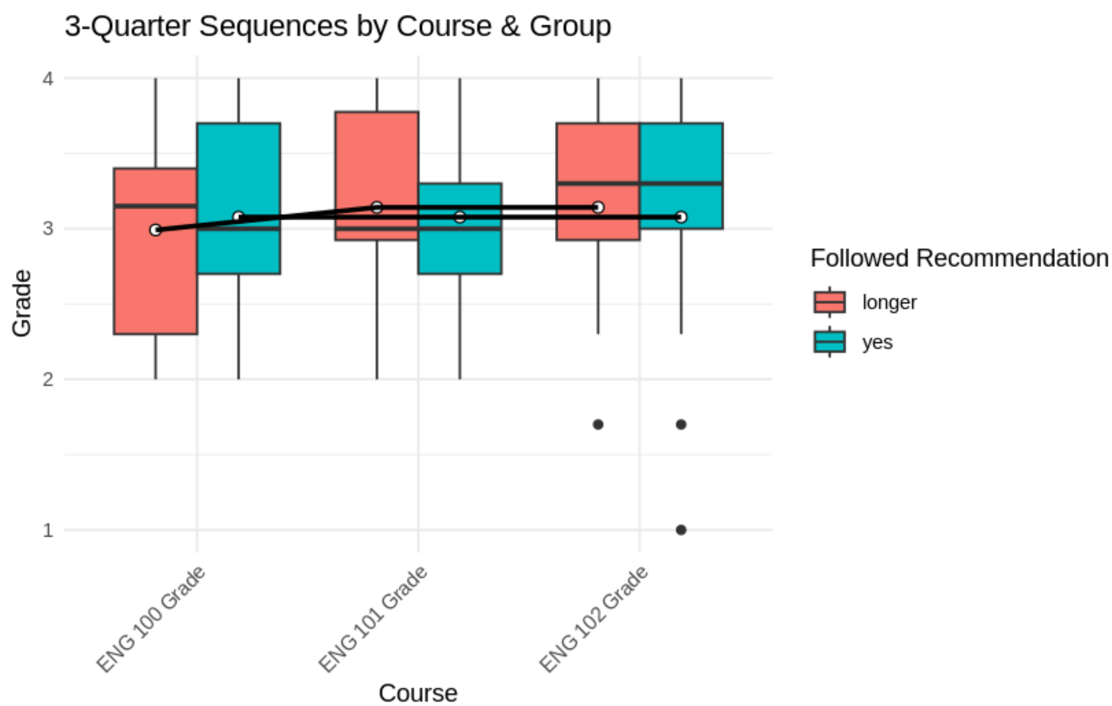
3.13) Question 13. Question 13 asks "Does the pattern of grades across ENG 100, ENG 101, and ENG 102 in the 3-quarter sequence differ between Groups"? The variables used for this question are the ENG 101, ENG 101, and 102 Grades (Dependent variables), a newly created independent variable Course (i = 1 (ENG 100), 2 (ENG 101), and 3 (ENG 102)), and the independent variable Followed Recommended (Group). Our goal for this question is to test if their Grades are improved in the 2-quarter Sequence by three Groups via a two-way repeated measure ANOVA. Two-way repeated measure ANOVA is like the one-way except there are different levels for each of the independent variables.



Summary Statistics for each Group:

Group Type	Min	1st Qu.	Median	3rd Qu.	Max	Mean	SD
Longer	1.7	2.7	3.2	3.7	4.0	3.09	0.668
Yes	1.0	2.7	3.0	3.7	4.0	3.08	0.721

For the individual boxplots separating the 3-quarter sequences by group, there are only two groups from the 'Followed Recommend' Column of the 3-quarter sequences. Of the two, the "longer" group has a higher mean grade of 3.09 with a lower SD. The "longer group" is symmetrical whereas the "yes" group is skewed to the right with a lower outlier. Deeper analysis between the Courses reveals some interactions. The lines intersect which would suggest there may be some interaction. The distributions between the "longer" and "yes" group are skewed for ENG 100. ENG 101 has one "longer" group skewed distribution and a symmetrical "yes" group distribution. For ENG 102, both are symmetrical with at least one or two lower outliers. Grades for ENG 101 and ENG 102 from the "longer" group contains the highest mean grade of 3.14.



Summary Statistics by Course (ENG 100, 101, 102) by Group:

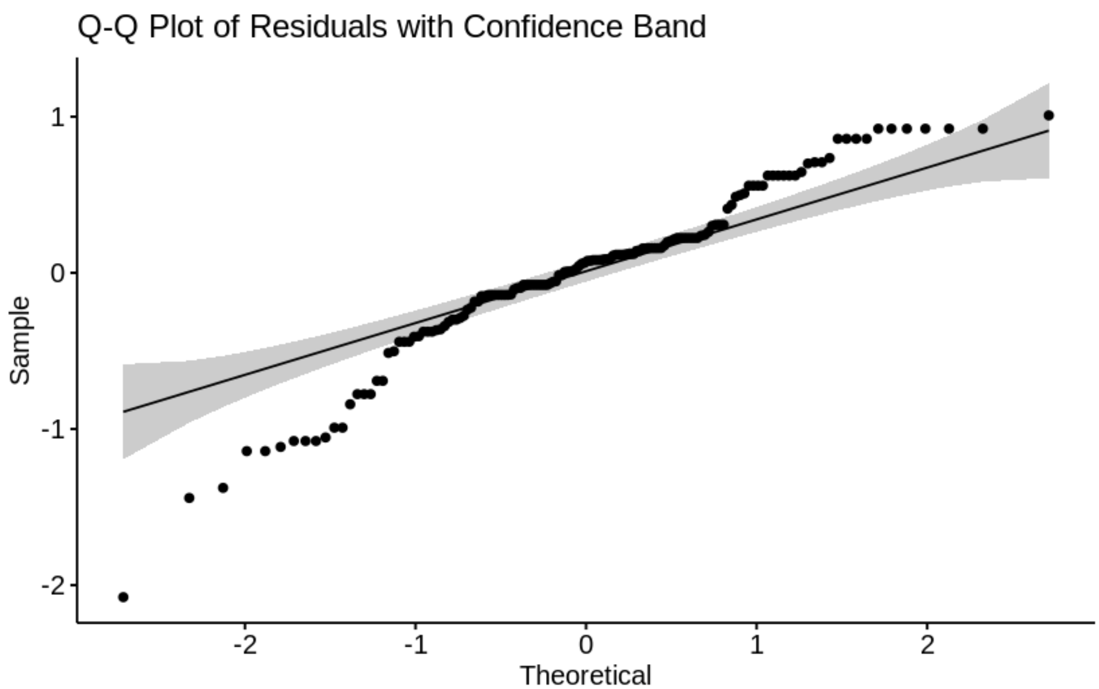
Course	Group	Min	Q1	Median	Q3	Max	Mean	SD
ENG 100	longer	2.0	2.3	3.2	3.4	4.0	2.99	0.689
ENG 100	yes	2.0	2.7	3.0	3.7	4.0	3.08	0.687
ENG 101	longer	2.0	2.9	3.0	3.8	4.0	3.14	0.703
ENG 101	yes	2.0	2.7	3.0	3.3	4.0	3.08	0.596
ENG 102	longer	1.7	2.9	3.3	3.7	4.0	3.14	0.658
ENG 102	yes	1.0	3.0	3.3	3.7	4.0	3.08	0.906

To conduct a two-way ANOVA, we formed the linear regression model between the Grades (the outcome) between the 3-quarter Courses, the Group and the interaction between the Courses and Groups.

$$y_{ij} = \mu + \alpha_i + \gamma_j + (\alpha\gamma)_{ij} + \epsilon_{ij}$$

y_{ij} is the Grades, μ is the mean effect on the variables: α_i (Course; i = 1 (ENG 100), 2 (ENG 101), 3 (ENG 102)), γ_j (Group, j = 1(longer), 2(shorter), and 3(yes), and the interaction term $(\alpha\gamma)_{ij}$between Group and Course. ϵ_{ij} is the error term for the equation.

A constructed QQ plot was generated to check the model's normality. The distribution looks normal but there is a heavy lower tail and the upper tail. A Shapiro-Wilk test was conducted, and the p-value is less than the 0.05 level. The data appears to be nonnormal, so the parametric Two-way repeated measure ANOVA was not conducted. We conducted the nonparametric ART ANOVA technique.



Shapiro Wilk Q13 Results:

W test statistic	p-value
0.95128	4.189e-05

For ART ANOVA, we ranked each of the variables and then conducted our analysis based on the formed following hypothesis and equation:

$$y_{ij} = \alpha_i + \gamma_j + (\alpha\gamma)_{ij} + \epsilon_{ij} ;$$

- H_0: The factor in question is zero for every i and j has no linear relationship with the 3-quarter grades
- H_A: The coefficient term in question is nonzero for at least one i or j has a linear relationship with the 3-quarter grades

Based on the results of ART ANOVA, none of the variables along with the interaction between the Course and Group have significance. All fail to reject the null hypotheses as they are greater than the 0.05 level. All the coefficients had no significant relationships with the 3-quarter grades, as such no post-hoc testing was conducted. From this we can conclude grade improvement in the 3-quarter sequence does not significantly differ between the 3 Groups (Followed, Shorter, Longer)

ART ANOVA Q13 Results:

	Df	**Df.res**	**F value**	**Pr(>F)**
Followed Recommended	1	23	0.065521	0.80025
Course	2	46	0.576165	0.56605
Course:Followed Recommend	2	46	0.136262	0.87297

For the last set, our client asks us to do a series of tests for the Portfolio Total Score (Column L) and the Comp AVG Grade (Column S) between the three sequences in Column E?

3.14) Question 14. Question 14 asks "Is there a linear relationship between Portfolio Total Score and Comp Average Grade, and does this relationship differ by Group (Followed, Shorter, Longer)"? The variables used for this question are Portfolio Total Score (our dependent variable), Comp AVG Grade and the Group (Followed Recommended). We will find the linear regression equations for each of the three Groups by forming a multiple linear regression model with interactions between the Comp AVG Grade and Group independent variables.

For this linear model formed the following equation:

$$y_{ij} = \mu + \kappa + \alpha_j + (\kappa\alpha)_j + \epsilon_j;$$

where $y_{ij}$is the PTS scores, μ is the mean effect of the main effects: κ (Comp AVG Grade) and α_j (Group; j = 1 (longer), 2 (shorter), and 3 (yes), $(\kappa\alpha)_j$ is the interaction effect between the Comp AVG Grade and Group and ϵ_j is the error term of the model.

The following metrics were derived:

Summary Statistics of Full Model:

Coefficients	Estimate	Std. Error	T value	Pr(>|t|)
Intercept	11.504	3.385	3.398	0.000935
Comp AVG Grade	1.882	1.028	1.831	0.69696
Group:shorter	7.653	3.961	1.932	0.055832
Group:yes	4.187	3.897	1.074	0.28914
Comp AVG Grade: Group'Shorter'	-2.044	1.184	-1.726	0.087052
Comp AVG Grade: Group'yes'	-1.485	1.184	-1.254	0.21347

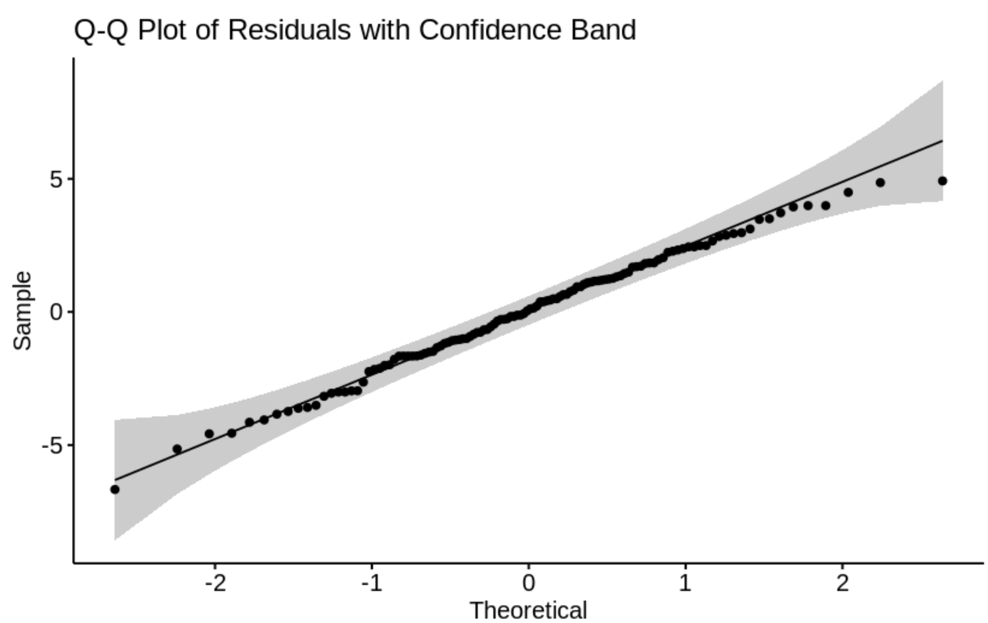
The model was:

$$Y_0 = 11.504 + 1.882\kappa + 7.653\alpha_2 + 4.187\alpha_3 - 2.044(\kappa\alpha)_2 - 1.485(\kappa\alpha)_3;$$

With the following metrics:

Residual std error	**R^2**	**R (correlation)**	**Adjusted R-squared**	**MSE**
2.374 on 114 df	0.1247	0.353	0.08627	5.636

Overall, the model has a weak relation between the variables and the interaction term. We conducted a Shapiro-Wilk test and it confirmed the model is normal due to having a p-value that is greater than 0.05.



Shapiro Wilk Q14 Results:

W test statistic	p-value
0.99587	0.7743

We conducted a one-way ANOVA test to form the best model of the combined equation (the following includes the error metrics). We found that the Group variable is statistically significant hence we created a model based on the Group types (longer, shorter, yes)

ANOVA Table of full model:

	Df	Sum sq	Mean sq	F value	Pr(>F)
Comp AVG Grade	1	14.14	14.1385	2.5095	0.1159
Group	2	60.52	30.2609	5.3710	0.0059
Comp AVG Grade: Group	2	16.81	8.4062	1.4920	0.2293
Residuals	114	642.29	5.6341		

Based upon previous results, we formed the following models:

$$y_{longer} = 11.509 + 1.882\kappa + 4.857\alpha_1 + \epsilon_1,$$

$$y_{shorter} = 11.509 + 1.882\kappa + 7.653\alpha_2 - 2.044(\kappa\alpha)_2 + \epsilon_2,$$

$$y_{yes} = 11.509 + 1.882\kappa + 4.187\alpha_3 - 1.485(\kappa\alpha)_3 + \epsilon_3;$$

As these are subsets of the overall model, we can conclude that these models are normal since the overall model is normal. As such continued testing the three models:

3.15) Question 15. Question 15 asks, "What is the correlation between Portfolio Total Score and Comp Average Grade for each Group"? The variables in question were the Portfolio Total Score (dependent variable), Comp AVG Grade, and Group (followed Recommended). Our Goal for this question would be to compute the correlation coefficients and assess their statistical significance by the parametric Pearson's Correlation coefficient test.

The following table displays the Groups (longer, shorter and yes) with each correlation coefficient (r) and a p-value to test their significance. None of the groups, appears to be statistically significant as the p-values are greater than the 0.05 level. From this the "longer" group has the best correlation followed by the "yes" group and "shorter" group.

Pearson's Coefficient Correlation Test Results:

Group	correlation	p-value
longer	0.438	0.102

shorter	-0.0359	0.793
yes	0.105	0.473

4) Conclusion:

From our analysis of the Stretch Composition program, we uncovered many revelations from the students' grades by conducting through statistical testing. By examining the Students' Portfolio Total Score, we uncovered students tend to take the shorter recommendation more compared to the longer recommendation. By ANOVA we can tell that "group" is significant but through a Tukey's HSD post-hoc test, we uncovered that the "shorter" recommendation is the one students would most likely take. As for comparing the Group with students Comp AVG Grades, students do in fact follow a shorter recommendation but with the Kruskal Wallis test, this is insignificant and would suggest that the median distributions of the Grades are the same.

As for any achievement grasp in student information, we compared the PTS scores with the different colleges at the university. While College 4 has the highest PTS (18.4) compared to the highest variability, College 9 has the highest. However one-way ANOVA reveals that College is insignificant so there is not a statistical difference that is significant. From the Comp AVG Grades, the distribution between genders reveals female students tend to score higher and the distributions are statistically significant. A Chi-Square test of Independence reveals the comparisons between Gender and Group to have no association whereas the URM status and Group reveals an association. When comparing the Rubric Category Scores, there was also no significance despite using nonparametric techniques. From this we can collude that at this university, female students have a different distribution for males and URM status is related to the recommendation a student takes.

Continuing analyzing the 6 Portfolio Rubric Categroy Scores reveal to be no differences from the 6 categories for all the students combined. The Kruskal Wallis test reveals no significance, but comparing the Portfolios Rubric Category score, the Category and the three groups reveals students taking the shorter group to be significant. ART ANOVA confirms this as students prefer to take a shorter recommendation route compared to a longer route. This reveals that there is a difference between the groups when comparing the Rubric Category Scores.

We next examined the 2-quarter and 3-quarter sequences to see if there are grade improvements. There is not statistical significance when comparing the 2-sequence and 3-sequence courses to their respective courses. Similarly, it can be concluded when comparing them by the three recommendation groups. There is not an improvement in the 2-quarter sequence or 3-quarter sequence.

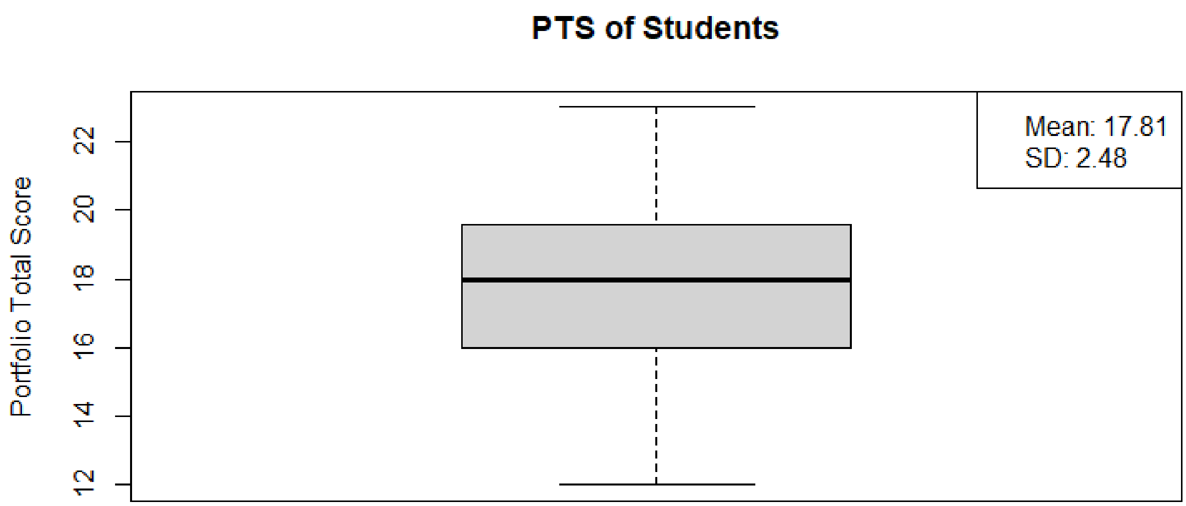
We then formed a model to explore the relationship between the PTS, the Comp AVG Grade and the three Groups. From the overall model, we found out that the model was not entirely strong, therefore we continued to examine models based on the three groups. Since the overall model was normal the subset models were normal by default. Of the groups, the group with the highest correlation is the longer group model. This would be the best model to study when comparing the dependent variable to the independent variables.

5) Appendices:

```
library(readr)
edu <- read_csv("C:/Users/david/OneDrive/Documents/STA classes/STA 5900/Group Project 3/Education-1.csv")
View(edu)

library(dplyr)

# Step 1: Exclude rows where Followed Recommend is 9999
edu_group <- edu %>%
  filter(`Followed Recommend` != 9999) %>%
  mutate(group = as.factor(`Followed Recommend`))

View(edu_group)
```

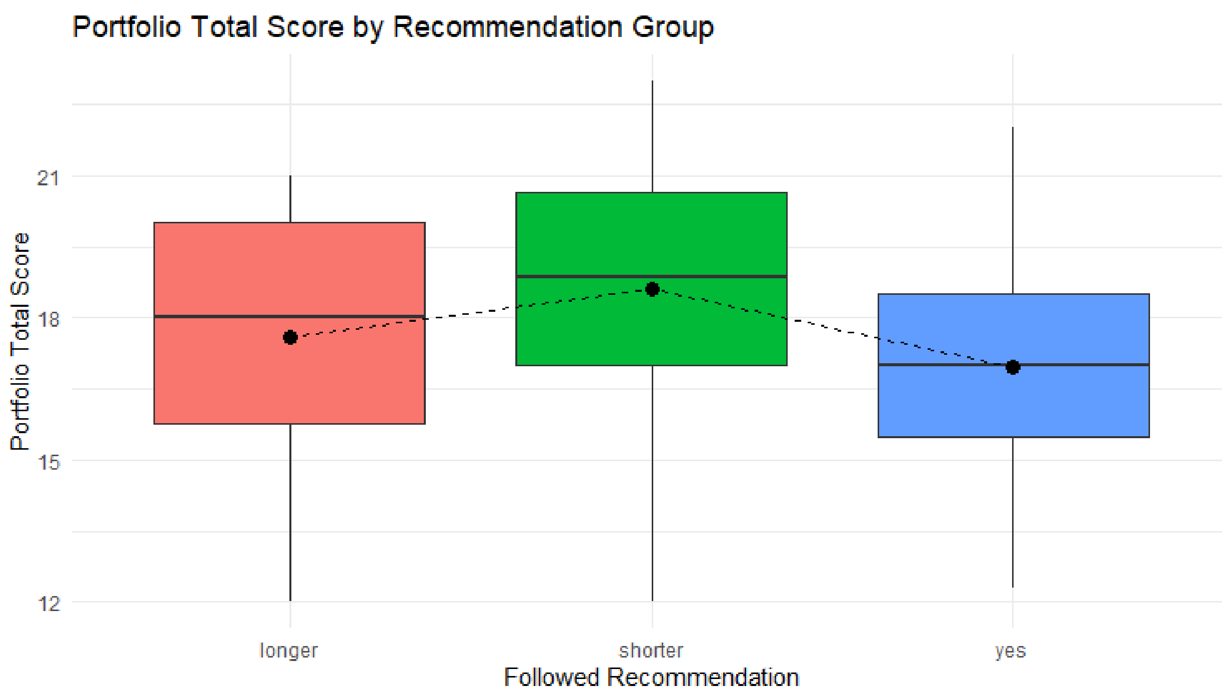
```
# Q#01; Variables involved: Portfolio Total Score, Group; Method: One Way Anova

# Define Portfolio Total Score vector (PTS)
edu_group$`Followed Recommend` <- as.factor(edu_group$`Followed Recommend`)

# Summary statistics (overall)
summary(edu_group$`Portfolio Total Score`)
sd(edu_group$`Portfolio Total Score`)

# Boxplot with mean and SD displayed in the legend
boxplot(edu_group$`Portfolio Total Score`,
        ylab = "Portfolio Total Score",
        main = "PTS of Students")
legend("topright",
       legend = c(paste("Mean:", round(mean(edu_group$`Portfolio Total Score`), 2)),
                  paste("SD:", round(sd(edu_group$`Portfolio Total Score`), 2))))

# Group-level summary stats
edu_group %>%
  group_by(group) %>%
  summarise(
    Min = min(`Portfolio Total Score`, na.rm = TRUE),
    Q1 = quantile(`Portfolio Total Score`, 0.25, na.rm = TRUE),
    Median = median(`Portfolio Total Score`, na.rm = TRUE),
    Q3 = quantile(`Portfolio Total Score`, 0.75, na.rm = TRUE),
    Max = max(`Portfolio Total Score`, na.rm = TRUE),
    Mean_PTS = mean(`Portfolio Total Score`, na.rm = TRUE),
    SD_PTS = sd(`Portfolio Total Score`, na.rm = TRUE),
    n = n()
  )
```

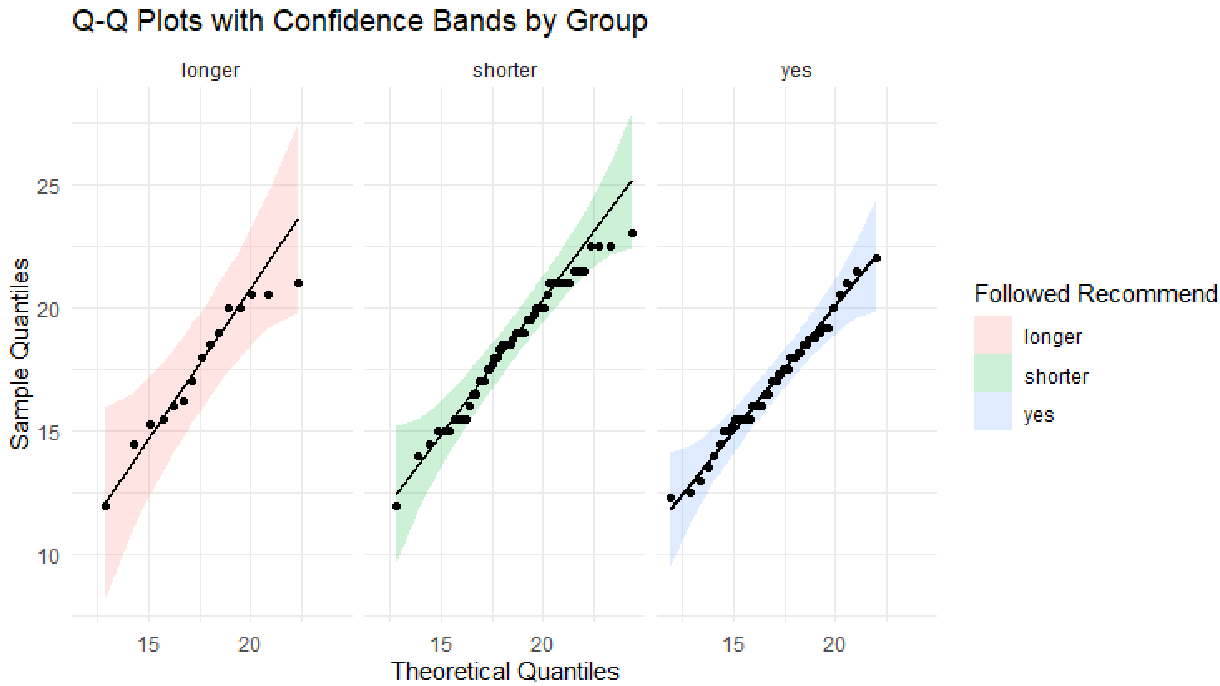
```
>  summary(edu_group$`Portfolio Total Score`)
   Min. 1st Qu.  Median    Mean 3rd Qu.    Max. 
  12.00   16.00   18.00   17.81   19.55   23.00 
>  sd(edu_group$`Portfolio Total Score`)
[1] 2.483153
```




```
>  # Group-level summary stats
>  edu_group %>%
+    group_by(group) %>%
+    summarise(
+      Min = min(`Portfolio Total Score`, na.rm = TRUE),
+      Q1 = quantile(`Portfolio Total Score`, 0.25, na.rm = TRUE),
+      Median = median(`Portfolio Total Score`, na.rm = TRUE),
+      Q3 = quantile(`Portfolio Total Score`, 0.75, na.rm = TRUE),
+      Max = max(`Portfolio Total Score`, na.rm = TRUE),
+      Mean_PTS = mean(`Portfolio Total Score`, na.rm = TRUE),
+      SD_PTS = sd(`Portfolio Total Score`, na.rm = TRUE),
+      n = n()
+    )
# A tibble: 3 × 9
  group     Min    Q1 Median    Q3   Max Mean_PTS SD_PTS     n
  <fct>   <dbl> <dbl>  <dbl> <dbl> <dbl>    <dbl>  <dbl> <int>
1 longer  12     15.8   18    20      21     17.6   2.65    15
2 shorter 12     17     18.8  20.6    23     18.6   2.46    56
3 yes     12.3   15.5   17    18.5    22     17.0   2.20    49
```

```
# Enhanced boxplot with group means overlaid as red dots and lines
ggplot(edu_group, aes(x = group, y = `Portfolio Total Score`, fill = group)) +
  geom_boxplot() +
  stat_summary(fun = mean, geom = "point", color = "black", size = 3) +  # Red dot for group mean
  stat_summary(fun = mean, geom = "line", aes(group = 1), color = "black", linetype = "dashed") +  # Line across means
  labs(title = "Portfolio Total Score by Recommendation Group",
       x = "Followed Recommendation",
       y = "Portfolio Total Score") +
  theme_minimal() +
  theme(legend.position = "none")

# Shapiro-Wilk test for normality per group
ggplot(edu_group, aes(sample = `Portfolio Total Score`)) +
  stat_qq_band(aes(fill = `Followed Recommend`), alpha = 0.2) +  # Confidence band
  stat_qq_point() +  # Q-Q points
  stat_qq_line() +   # Q-Q line
  facet_wrap(~ `Followed Recommend`) +
  labs(title = "Q-Q Plots with Confidence Bands by Group",
       x = "Theoretical Quantiles",
       y = "Sample Quantiles") +
  theme_minimal()

edu_group %>%
  group_by(group) %>%
  summarise(p_value = shapiro.test(`Portfolio Total Score`)$p.value)

# One-Way ANOVA
anova1 <- aov(`Portfolio Total Score` ~ group, data = edu_group)
summary(anova1)

# Post-hoc (if ANOVA is significant)
TukeyHSD(anova1)
```

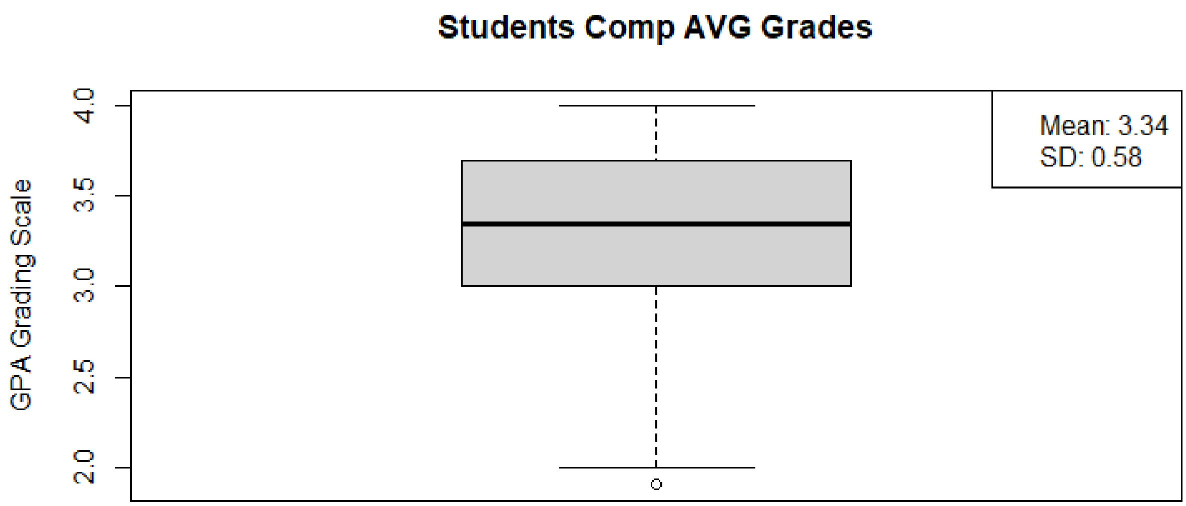




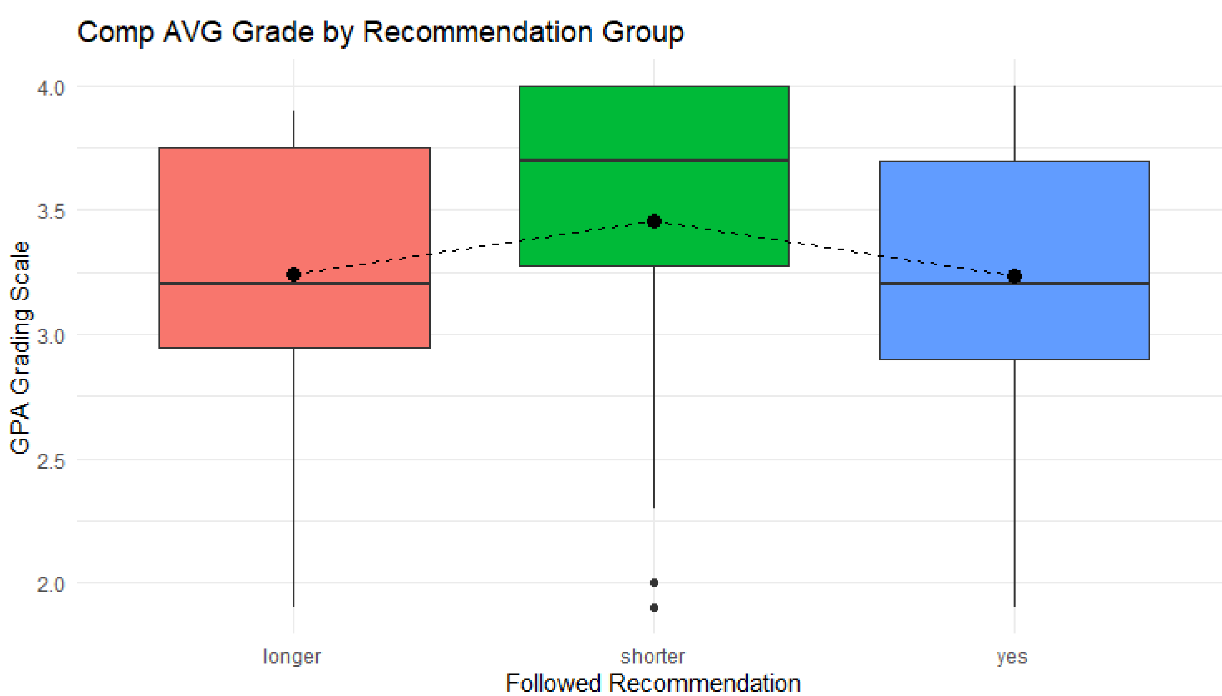

```
>  # Shapiro-Wilk test for normality per group
>  ggplot(edu_group, aes(sample = `Portfolio Total Score`)) +
+    stat_qq_band(aes(fill = `Followed Recommend`), alpha = 0.2) +  # Confidence
band
+    stat_qq_point() +  # Q-Q points
+    stat_qq_line() +   # Q-Q line
+    facet_wrap(~ `Followed Recommend`) +
+    labs(title = "Q-Q Plots with Confidence Bands by Group",
+         x = "Theoretical Quantiles",
+         y = "Sample Quantiles") +
+    theme_minimal()
>  edu_group %>%
+    group_by(group) %>%
+    summarise(p_value = shapiro.test(`Portfolio Total Score`)$p.value)
# A tibble: 3 × 2
  group   p_value
  <fct>     <dbl>
1 longer    0.377
2 shorter   0.265
3 yes       0.832
>  edu_group %>%
+    group_by(group) %>%
+    summarise(p_value = shapiro.test(`Portfolio Total Score`)$p.value)
# A tibble: 3 × 2
  group   p_value
  <fct>     <dbl>
1 longer    0.377
2 shorter   0.265
3 yes       0.832
```

```
> # One-Way ANOVA
> anova1 <- aov(`Portfolio Total Score` ~ group, data = edu_group)
> summary(anova1)
             Df Sum Sq Mean Sq F value  Pr(>F)
group         2   69.6   34.79   6.129 0.00294 **
Residuals   117  664.2    5.68
---
Signif. codes:  0 ‘***’ 0.001 ‘**’ 0.01 ‘*’ 0.05 ‘.’ 0.1 ‘ ’ 1
> # Post-hoc (if ANOVA is significant)
> TukeyHSD(anova1)
  Tukey multiple comparisons of means
    95% family-wise confidence level

Fit: aov(formula = `Portfolio Total Score` ~ group, data = edu_group)

$group
                     diff        lwr       upr     p adj
shorter-longer  0.9964286 -0.6479519  2.640809 0.3246162
yes-longer     -0.6265306 -2.2955412  1.042480 0.6469595
yes-shorter    -1.6229592 -2.7293673 -0.516551 0.0020097
```

```
# Q#02; Variables Involved: Comp AVG Grade, Group; Method: One-Way Anova

# Define Comp AVG Grade (CG)
edu_group$`Comp AVG Grade`

# Summary statistics (overall)
summary(edu_group$`Comp AVG Grade`)
sd(edu_group$`Comp AVG Grade`)

# Boxplot with mean and SD displayed in the legend
boxplot(edu_group$`Comp AVG Grade`,
        ylab = "GPA Grading Scale",
        main = "Students Comp AVG Grades")
legend("topright",
       legend = c(paste("Mean:", round(mean(edu_group$`Comp AVG Grade`), 2)),
                  paste("SD:", round(sd(edu_group$`Comp AVG Grade`), 2))))

# Group-level summary stats
edu_group %>%
  group_by(group) %>%
  summarise(
    Min = min(`Comp AVG Grade`, na.rm = TRUE),
    Q1 = quantile(`Comp AVG Grade`, 0.25, na.rm = TRUE),
    Median = median(`Comp AVG Grade`, na.rm = TRUE),
    Q3 = quantile(`Comp AVG Grade`, 0.75, na.rm = TRUE),
    Max = max(`Comp AVG Grade`, na.rm = TRUE),
    Mean = mean(`Comp AVG Grade`, na.rm = TRUE),
    SD = sd(`Comp AVG Grade`, na.rm = TRUE),
    n = n()
  )

# Enhanced boxplot with group means overlaid as red dots and lines
ggplot(edu_group, aes(x = group, y = `Comp AVG Grade`, fill = group)) +
  geom_boxplot() +
  stat_summary(fun = mean, geom = "point", color = "black", size = 3) +  # Red dot for group mean
  stat_summary(fun = mean, geom = "line", aes(group = 1), color = "black", linetype = "dashed") +  # Line across means
  labs(title = "Comp AVG Grade by Recommendation Group",
       x = "Followed Recommendation",
       y = "GPA Grading Scale") +
  theme_minimal() +
  theme(legend.position = "none")
```



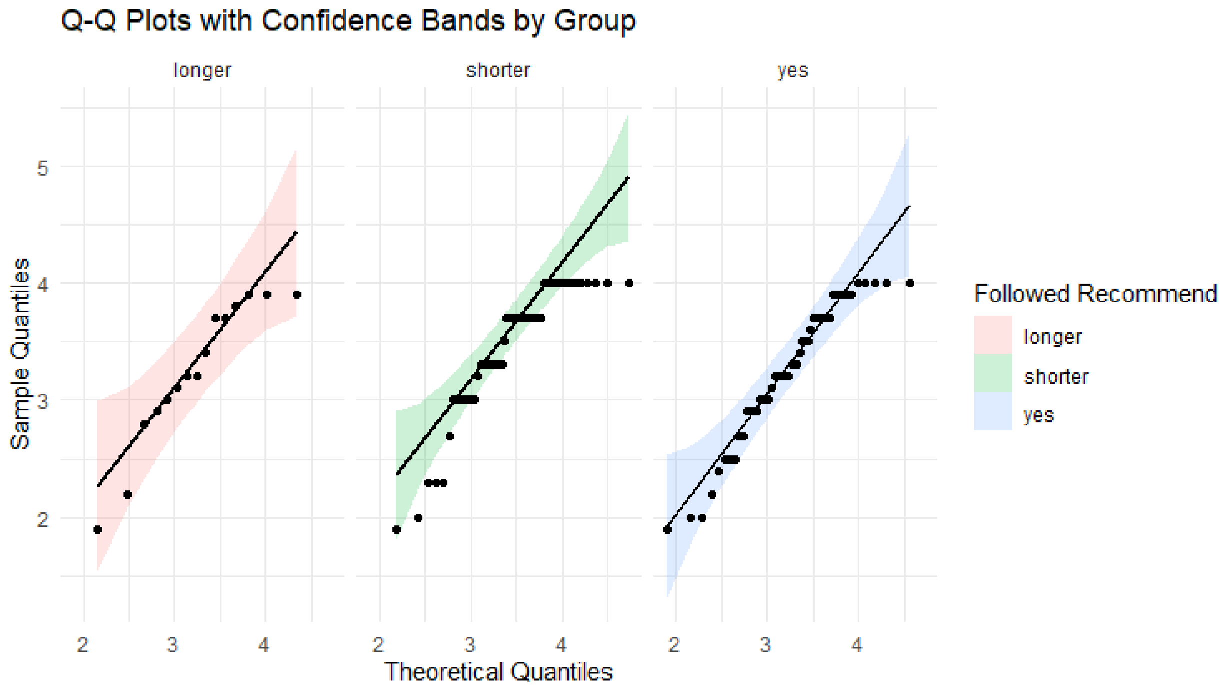

```
>  # Summary statistics (overall)
>  summary(edu_group$`Comp AVG Grade`)
   Min. 1st Qu.  Median    Mean 3rd Qu.    Max. 
  1.900   3.000   3.350   3.337   3.700   4.000 
>  sd(edu_group$`Comp AVG Grade`)
[1] 0.5751044
```




```
>  # Group-level summary stats
>  edu_group %>%
+    group_by(group) %>%
```

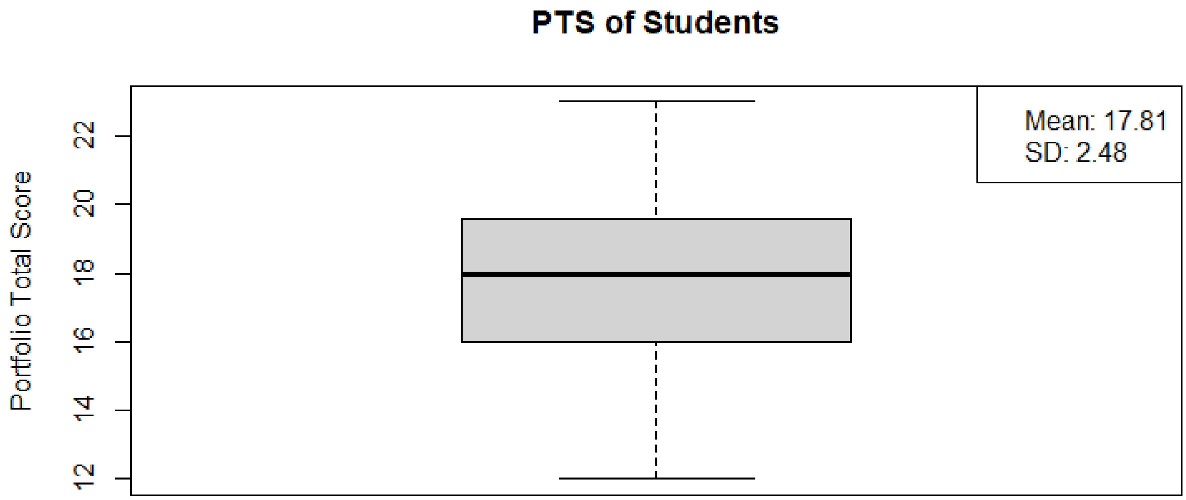
```
+     summarise(
+       Min = min(`Comp AVG Grade`, na.rm = TRUE),
+       Q1 = quantile(`Comp AVG Grade`, 0.25, na.rm = TRUE),
+       Median = median(`Comp AVG Grade`, na.rm = TRUE),
+       Q3 = quantile(`Comp AVG Grade`, 0.75, na.rm = TRUE),
+       Max = max(`Comp AVG Grade`, na.rm = TRUE),
+       Mean = mean(`Comp AVG Grade`, na.rm = TRUE),
+       SD = sd(`Comp AVG Grade`, na.rm = TRUE),
+       n = n()
+     )
# A tibble: 3 × 9
  group     Min    Q1 Median    Q3   Max  Mean    SD     n
  <fct>   <dbl> <dbl>  <dbl> <dbl> <dbl> <dbl> <dbl> <int>
1 longer    1.9  2.95    3.2  3.75   3.9  3.24 0.617    15
2 shorter   1.9  3.27    3.7  4      4    3.45 0.544    56
3 yes       1.9  2.9     3.2  3.7    4    3.23 0.583    49
```

```
# QQ plot
ggplot(edu_group, aes(sample = `Comp AVG Grade`)) +
  stat_qq_band(aes(fill = `Followed Recommend`), alpha = 0.2) +  # Confidence band
  stat_qq_point() +  # Q-Q points
  stat_qq_line() +   # Q-Q line
  facet_wrap(~ `Followed Recommend`) +
  labs(title = "Q-Q Plots with Confidence Bands by Group",
       x = "Theoretical Quantiles",
       y = "Sample Quantiles") +
  theme_minimal()

# Shapiro-Wilk test for normality per group
edu_group %>%
  group_by(group) %>%
  summarise(p_value = shapiro.test(`Comp AVG Grade`)$p.value)

# nonparametric kruskal wallis test
kruskal.test(`Comp AVG Grade` ~ group, data = edu_group)
```

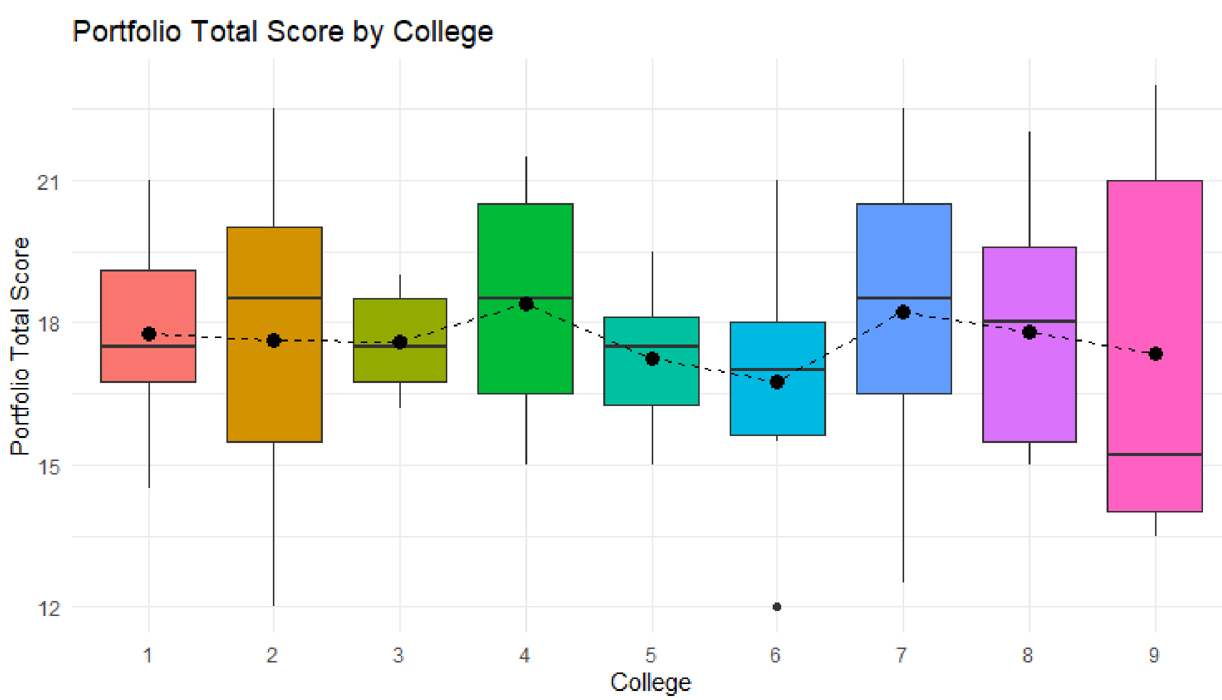



```
> # Shapiro-Wilk test for normality per group
> edu_group %>%
+   group_by(group) %>%
+   summarise(p_value = shapiro.test(`Comp AVG Grade`)$p.value)
# A tibble: 3 × 2

  group      p_value
  <fct>        <dbl>
1 longer   0.0964
2 shorter 0.00000669
3 yes     0.0140
> # nonparametric kruskal wallis test
> kruskal.test(`Comp AVG Grade` ~ group, data = edu_group)

	Kruskal-Wallis rank sum test

data:  Comp AVG Grade by group
Kruskal-Wallis chi-squared = 5.8614, df = 2, p-value = 0.05336
```

```r
# Q#03; Variables Involved: Portfolio Total Score, College; Method: One-Way Anova

# Define Comp AVG Grade (CG) and College
edu_group$`Portfolio Total Score`
edu_group$College <- as.factor(edu_group$College)

# Summary statistics (overall)
summary(edu_group$`Portfolio Total Score`)
sd(edu_group$`Portfolio Total Score`)

# Boxplot with mean and SD displayed in the legend
boxplot(edu_group$`Portfolio Total Score`,
        ylab = "Portfolio Total Score",
        main = "PTS of Students")
legend("topright",
       legend = c(paste("Mean:", round(mean(edu_group$`Portfolio Total Score`), 2)),
                  paste("SD:", round(sd(edu_group$`Portfolio Total Score`), 2))))

# boxplot with group means overlaid as dots and lines
ggplot(edu_group, aes(x = College, y = `Portfolio Total Score`, fill = College)) +
  geom_boxplot() +
  stat_summary(fun = mean, geom = "point", color = "black", size = 3) +
  stat_summary(fun = mean, aes(group = 1), geom = "line", color = "black", linetype = "dashed") +
  labs(title = "Portfolio Total Score by College",
       x = "College",
       y = "Portfolio Total Score") +
  theme_minimal() +
  theme(legend.position = "none")

# Group-level summary stats
edu_group %>%
  group_by(College) %>%
  summarise(
    Min = min(`Portfolio Total Score`, na.rm = TRUE),
    Q1 = quantile(`Portfolio Total Score`, 0.25, na.rm = TRUE),
    Median = median(`Portfolio Total Score`, na.rm = TRUE),
    Q3 = quantile(`Portfolio Total Score`, 0.75, na.rm = TRUE),
    Max = max(`Portfolio Total Score`, na.rm = TRUE),
    Mean_PTS = mean(`Portfolio Total Score`, na.rm = TRUE),
    SD_PTS = sd(`Portfolio Total Score`, na.rm = TRUE),
    n = n()
  )
```

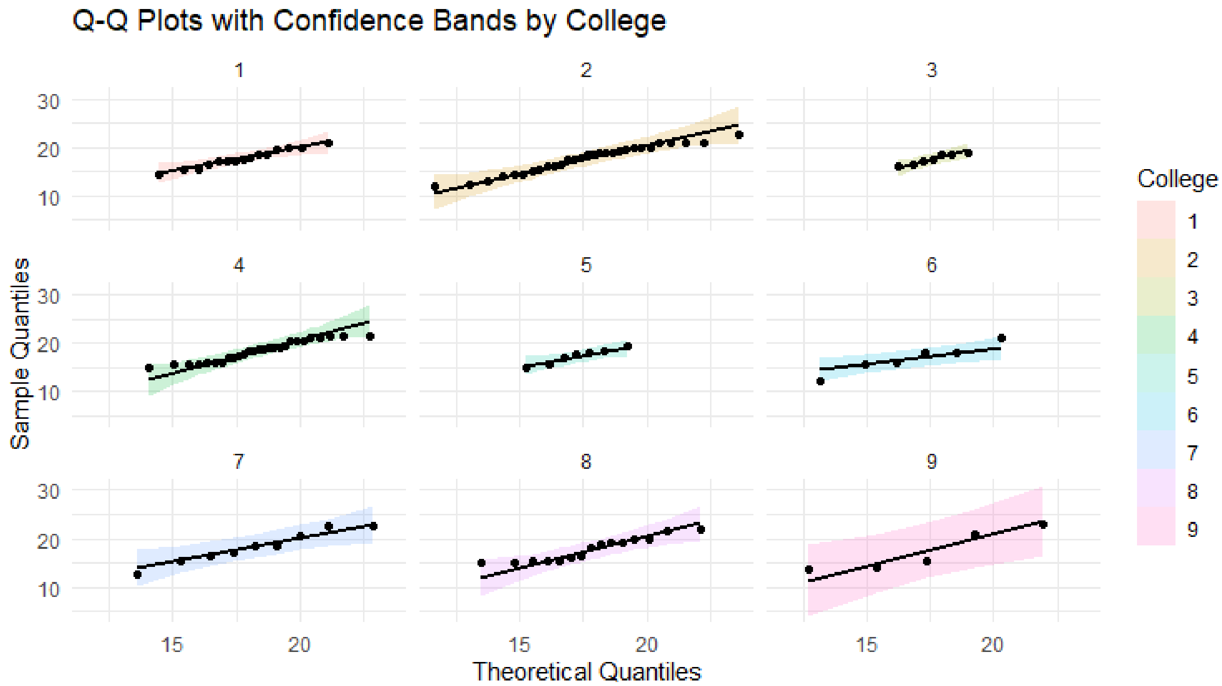



```
> summary(edu_group$`Portfolio Total Score`)
   Min. 1st Qu.  Median    Mean 3rd Qu.    Max. 
  12.00   16.00   18.00   17.81   19.55   23.00 
> sd(edu_group$`Portfolio Total Score`)
[1] 2.483153
```



```
> # Group-level summary stats
> edu_group %>%
+   group_by(College) %>%
+   summarise(
+     Min = min(`Portfolio Total Score`, na.rm = TRUE),
+     Q1 = quantile(`Portfolio Total Score`, 0.25, na.rm = TRUE),
+     Median = median(`Portfolio Total Score`, na.rm = TRUE),
+     Q3 = quantile(`Portfolio Total Score`, 0.75, na.rm = TRUE),
+     Max = max(`Portfolio Total Score`, na.rm = TRUE),
+     Mean_PTS = mean(`Portfolio Total Score`, na.rm = TRUE),
+     SD_PTS = sd(`Portfolio Total Score`, na.rm = TRUE),
+     n = n()
+   )
# A tibble: 9 × 9
  College   Min    Q1 Median    Q3   Max Mean_PTS SD_PTS     n
  <fct>   <dbl> <dbl>  <dbl> <dbl> <dbl>    <dbl>  <dbl> <int>
1 1        14.5  16.8   17.5  19.1  21       17.8   1.88    15
2 2        12    15.5   18.5  20    22.5     17.6   2.88    29
3 3        16.2  16.8   17.5  18.5  19       17.6   1.09     7
4 4        15    16.5   18.5  20.5  21.5     18.4   2.13    27
5 5        15    16.2   17.5  18.1  19.5     17.2   1.57     7
6 6        12    15.6   17    18    21       16.8   3.03     6
7 7        12.5  16.5   18.5  20.5  22.5     18.2   3.29     9
8 8        15    15.5   18    19.6  22       17.8   2.43    15
9 9        13.5  14     15.2  21    23       17.3   4.36     5
```

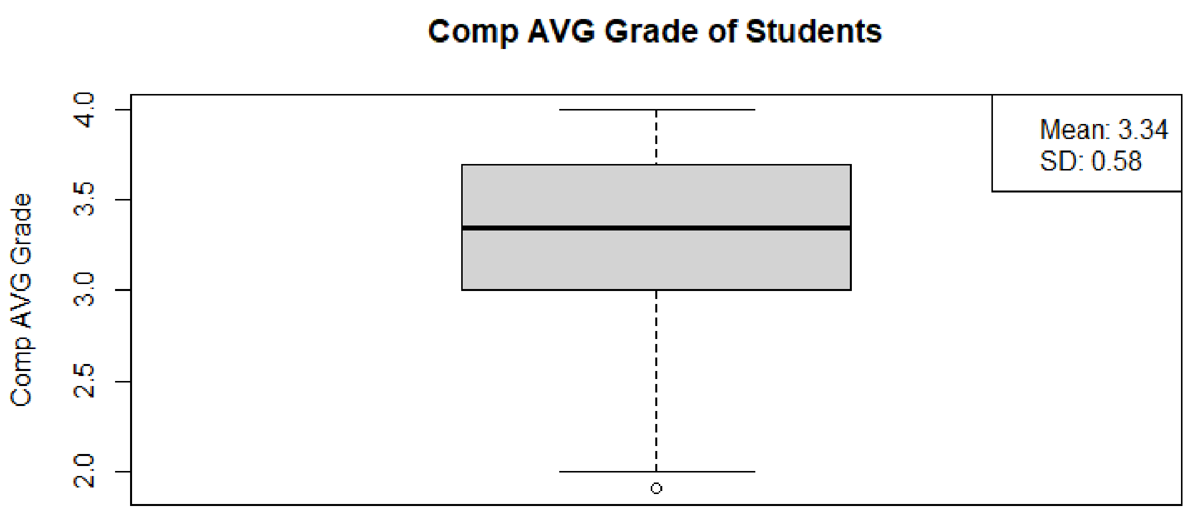
```
library(qqplotr)

# QQ plot
ggplot(edu_group, aes(sample = `Portfolio Total Score`)) +
  stat_qq_band(aes(fill = `College`), alpha = 0.2) +  # Confidence band
  stat_qq_point() +  # Q-Q points
  stat_qq_line() +   # Q-Q line
  facet_wrap(~ `College`) +
  labs(title = "Q-Q Plots with Confidence Bands by College",
       x = "Theoretical Quantiles",
       y = "Sample Quantiles") +
  theme_minimal()

# Shapiro-Wilk test for normality per College
edu_group %>%
  group_by(College) %>%
  summarise(p_value = shapiro.test(`Portfolio Total Score`)$p.value)

# One-Way ANOVA
anova3 <- aov(`Portfolio Total Score` ~ College, data = edu_group)
summary(anova3)

# Post-hoc (if ANOVA is significant)
TukeyHSD(anova3)
```



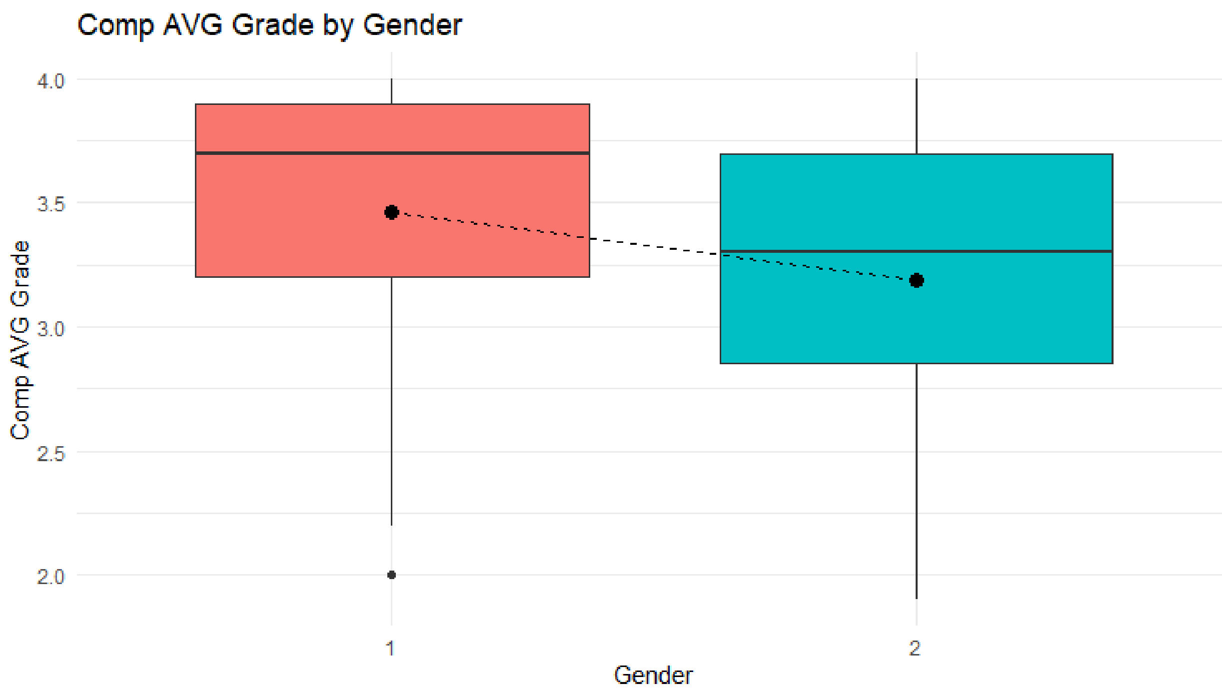

```
>  # Shapiro-Wilk test for normality per College
>  edu_group %>%
+    group_by(College) %>%
+    summarise(p_value = shapiro.test(`Portfolio Total Score`)$p.value)
# A tibble: 9 × 2
  College p_value
  <fct>     <dbl>
1 1         0.888
2 2         0.212
3 3         0.487
4 4         0.0668
5 5         0.805
6 6         0.881
7 7         0.745
8 8         0.0992
9 9         0.176
>  # One-Way ANOVA
>  anova3 <- aov(`Portfolio Total Score` ~ College, data = edu_group)
>  summary(anova3)
             Df Sum Sq Mean Sq F value Pr(>F)
College       8   22.2   2.778   0.433  0.899
Residuals   111  711.5   6.410
>  # Post-hoc (if ANOVA is significant)
>  TukeyHSD(anova3)
  Tukey multiple comparisons of means
    95% family-wise confidence level

Fit: aov(formula = `Portfolio Total Score` ~ College, data = edu_group)

$College
           diff       lwr      upr     p adj
2-1 -0.12873563 -2.676396 2.418925 1.0000000
```

3-1	-0.16666667	-3.833377	3.500043	1.0000000
4-1	0.63333333	-1.946294	3.212961	0.9972849
5-1	-0.52380952	-4.190520	3.142901	0.9999500
6-1	-1.01666667	-4.886108	2.852775	0.9956554
7-1	0.45555556	-2.921970	3.833081	0.9999678
8-1	0.04666667	-2.878356	2.971689	1.0000000
9-1	-0.42666667	-4.563274	3.709940	0.9999960
3-2	-0.03793103	-3.411294	3.335432	1.0000000
4-2	0.76206897	-1.380195	2.904333	0.9691164
5-2	-0.39507389	-3.768437	2.978289	0.9999892
6-2	-0.88793103	-4.480617	2.704755	0.9971559
7-2	0.58429119	-2.472258	3.640841	0.9995524
8-2	0.17540230	-2.372258	2.723063	0.9999998
9-2	-0.29793103	-4.176891	3.581029	0.9999996
4-3	0.80000000	-2.597570	4.197570	0.9979769
5-3	-0.35714286	-4.638938	3.924652	0.9999992
6-3	-0.85000000	-5.306633	3.606633	0.9995599
7-3	0.62222222	-3.414693	4.659137	0.9999108
8-3	0.21333333	-3.453377	3.880043	1.0000000
9-3	-0.26000000	-4.950471	4.430471	1.0000000
5-4	-1.15714286	-4.554713	2.240427	0.9763061
6-4	-1.65000000	-5.265425	1.965425	0.8780655
7-4	-0.17777778	-3.261023	2.905467	1.0000000
8-4	-0.58666667	-3.166294	1.992961	0.9984194
9-4	-1.06000000	-4.960030	2.840030	0.9945257
6-5	-0.49285714	-4.949491	3.963776	0.9999931
7-5	0.97936508	-3.057550	5.016280	0.9975027
8-5	0.57047619	-3.096234	4.237186	0.9999043
9-5	0.09714286	-4.593329	4.787614	1.0000000
7-6	1.47222222	-2.749685	5.694129	0.9726239
8-6	1.06333333	-2.806108	4.932775	0.9941016
9-6	0.59000000	-4.260602	5.440602	0.9999855
8-7	-0.40888889	-3.786414	2.968637	0.9999860
9-7	-0.88222222	-5.350268	3.585824	0.9994320
9-8	-0.47333333	-4.609940	3.663274	0.9999910

```
# Q#04; Variables Involved: Comp AVG Grade, Gender; Method: independent 2-sample t-test

# Define Comp AVG Grade (CG)
edu_group$`Comp AVG Grade`
edu_group$Sex <- as.factor(edu_group$Sex, levels = c(1, 2), labels = c("Female", "Male"))

# Summary statistics (overall)
summary(edu_group$`Comp AVG Grade`)
sd(edu_group$`Comp AVG Grade`)

# Boxplot with mean and SD displayed in the legend
boxplot(edu_group$`Comp AVG Grade`,
        ylab = "Comp AVG Grade",
        main = "Comp AVG Grade of Students")
legend("topright",
       legend = c(paste("Mean:", round(mean(edu_group$`Comp AVG Grade`), 2)),
                  paste("SD:", round(sd(edu_group$`Comp AVG Grade`), 2))))

# Enhanced boxplot with Gender means overlaid as red dots and lines
ggplot(edu_group, aes(x = Sex, y = `Comp AVG Grade`, fill = Sex)) +
  geom_boxplot() +
  stat_summary(fun = mean, geom = "point", color = "black", size = 3) +  # Red dot for group mean
  stat_summary(fun = mean, geom = "line", aes(group = 1), color = "black", linetype = "dashed") +  # Line across means
  labs(title = "Comp AVG Grade by Gender",
       x = "Gender",
       y = "Comp AVG Grade") +
  theme_minimal() +
  theme(legend.position = "none")

# Group-level summary stats
edu_group %>%
  group_by(Sex) %>%
  summarise(
    Min = min(`Comp AVG Grade`, na.rm = TRUE),
    Q1 = quantile(`Comp AVG Grade`, 0.25, na.rm = TRUE),
    Median = median(`Comp AVG Grade`, na.rm = TRUE),
    Q3 = quantile(`Comp AVG Grade`, 0.75, na.rm = TRUE),
    Max = max(`Comp AVG Grade`, na.rm = TRUE),
    Mean = mean(`Comp AVG Grade`, na.rm = TRUE),
    SD = sd(`Comp AVG Grade`, na.rm = TRUE),
    n = n()
  )
```

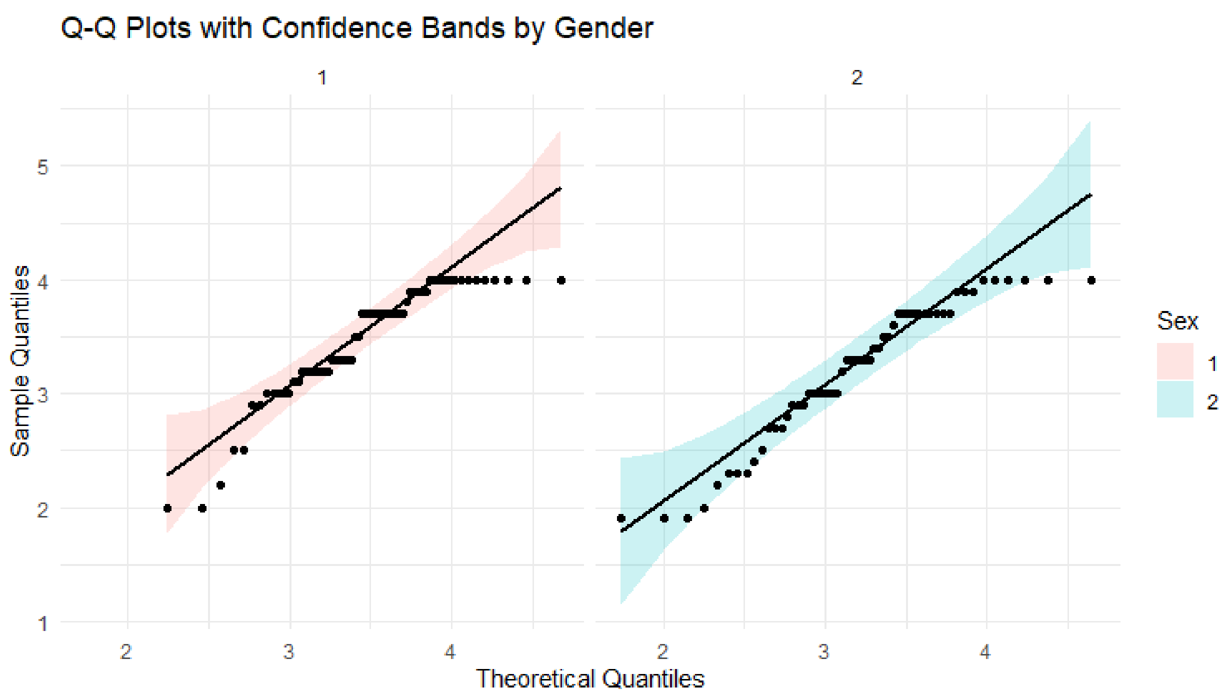



```
> summary(edu_group$`Comp AVG Grade`)
   Min. 1st Qu.  Median    Mean 3rd Qu.    Max. 
  1.900   3.000   3.350   3.337   3.700   4.000 
> sd(edu_group$`Comp AVG Grade`)
[1] 0.5751044
```




```
> # Group-level summary stats
> edu_group %>%
+   group_by(Sex) %>%
+   summarise(
+     Min = min(`Comp AVG Grade`, na.rm = TRUE),
+     Q1 = quantile(`Comp AVG Grade`, 0.25, na.rm = TRUE),
+     Median = median(`Comp AVG Grade`, na.rm = TRUE),
+     Q3 = quantile(`Comp AVG Grade`, 0.75, na.rm = TRUE),
+     Max = max(`Comp AVG Grade`, na.rm = TRUE),
+     Mean = mean(`Comp AVG Grade`, na.rm = TRUE),
+     SD = sd(`Comp AVG Grade`, na.rm = TRUE),
+     n = n()
+   )
# A tibble: 2 × 9
  Sex     Min    Q1 Median    Q3   Max  Mean    SD     n
  <fct> <dbl> <dbl>  <dbl> <dbl> <dbl> <dbl> <dbl> <int>
1 1       2    3.2     3.7   3.9     4  3.46 0.505    65
2 2       1.9  2.85    3.3   3.7     4  3.19 0.621    55
```

```
# qq plot
ggplot(edu_group, aes(sample = `Comp AVG Grade`)) +
  stat_qq_band(aes(fill = `Sex`), alpha = 0.2) +  # Confidence band
  stat_qq_point() +  # Q-Q points
  stat_qq_line() +   # Q-Q line
  facet_wrap(~ `Sex`) +
  labs(title = "Q-Q Plots with Confidence Bands by Gender",
       x = "Theoretical Quantiles",
       y = "Sample Quantiles") +
  theme_minimal()

# Shapiro-Wilk test for normality per group
edu_group %>%
  group_by(Sex) %>%
  summarise(p_value = shapiro.test(`Comp AVG Grade`)$p.value)

# fails normality
wilcox.test(`Comp AVG Grade` ~ Sex, data = edu_group)
```




```
> # Shapiro-Wilk test for normality per group
> edu_group %>%
+   group_by(Sex) %>%
+   summarise(p_value = shapiro.test(`Comp AVG Grade`)$p.value)
# A tibble: 2 × 2
  Sex     p_value
  <fct>     <dbl>
1 1     0.0000111
2 2     0.00291
> # fails normality
```

```
> wilcox.test(`Comp AVG Grade` ~ Sex, data = edu_group)

	Wilcoxon rank sum test with continuity correction

data:  Comp AVG Grade by Sex
W = 2257.5, p-value = 0.01268
alternative hypothesis: true location shift is not equal to 0
```

```
# Q05: Variables Involved: Gender, Sequence; Method: Chi-Squared Test of Independence

# Recode variables if needed
edu_group$Sex <- as.factor(edu_group$Sex, levels = c(1, 2), labels = c("Female", "Male"))
edu_group$Sequence <- factor(edu_group$Sequence)

# create table
gender_seq <- table(edu_group$Sex, edu_group$Sequence)
gender_seq

# test
chisq.test(gender_seq)
```

```
> # create table
> gender_seq <- table(edu_group$Sex, edu_group$Sequence)
> gender_seq

     1  2  3
  1 29 22 14
  2 28 16 11
> # test
> chisq.test(gender_seq)

	Pearson's Chi-squared test

data:  gender_seq
X-squared = 0.49502, df = 2, p-value = 0.7807
```

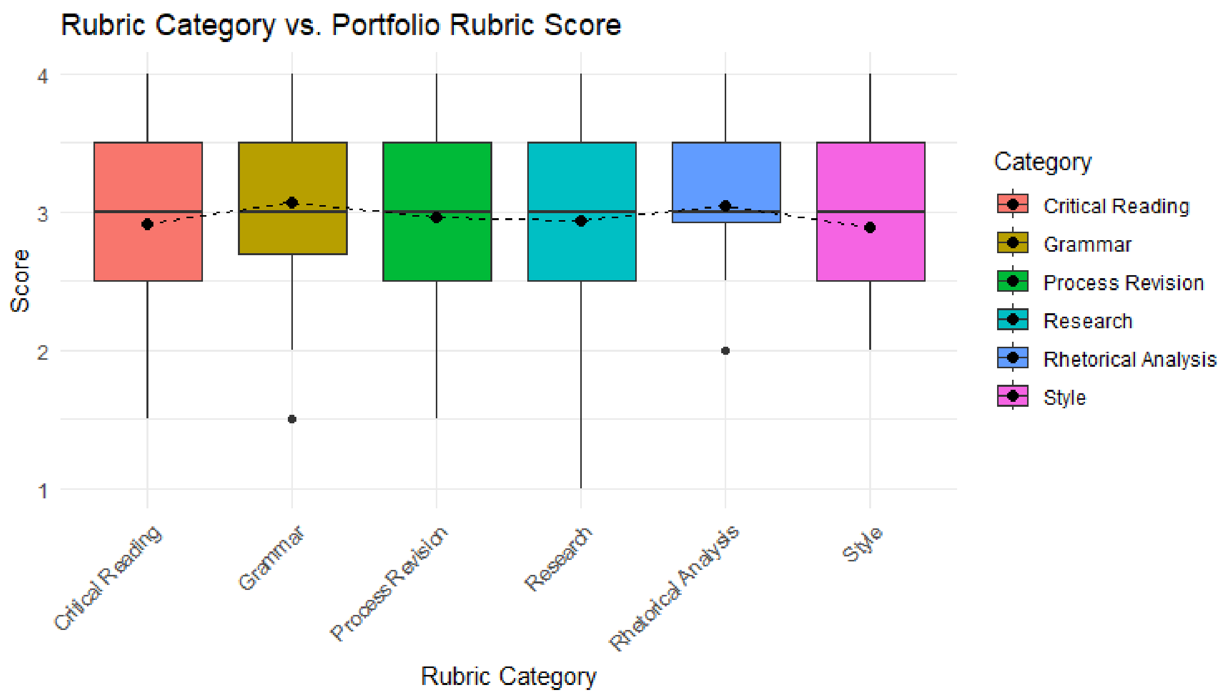
```
# Q06: Variables Involved: URM, Group; Method: Chi-Squared Test of Independence

# variables
edu_group$URM <- factor(edu_group$URM, labels = c("Not URM", "URM"))  # Assuming 0 = Not URM, 1 = URM
edu_group$group <- factor(edu_group$`Followed Recommend`)

# Create contingency table
urm_group <- table(edu_group$URM, edu_group$group)
urm_group

# Perform Chi-Squared Test of Independence
chisq.test(urm_group)
```

```
> # Create contingency table
> urm_group <- table(edu_group$URM, edu_group$group)
> urm_group

          longer shorter yes
  Not URM      8      20  30
```

```
  URM            7       36  19
> # Perform Chi-Squared Test of Independence
> chisq.test(urm_group)

	Pearson's Chi-squared test

data:  urm_group
X-squared = 6.9819, df = 2, p-value = 0.03047
```

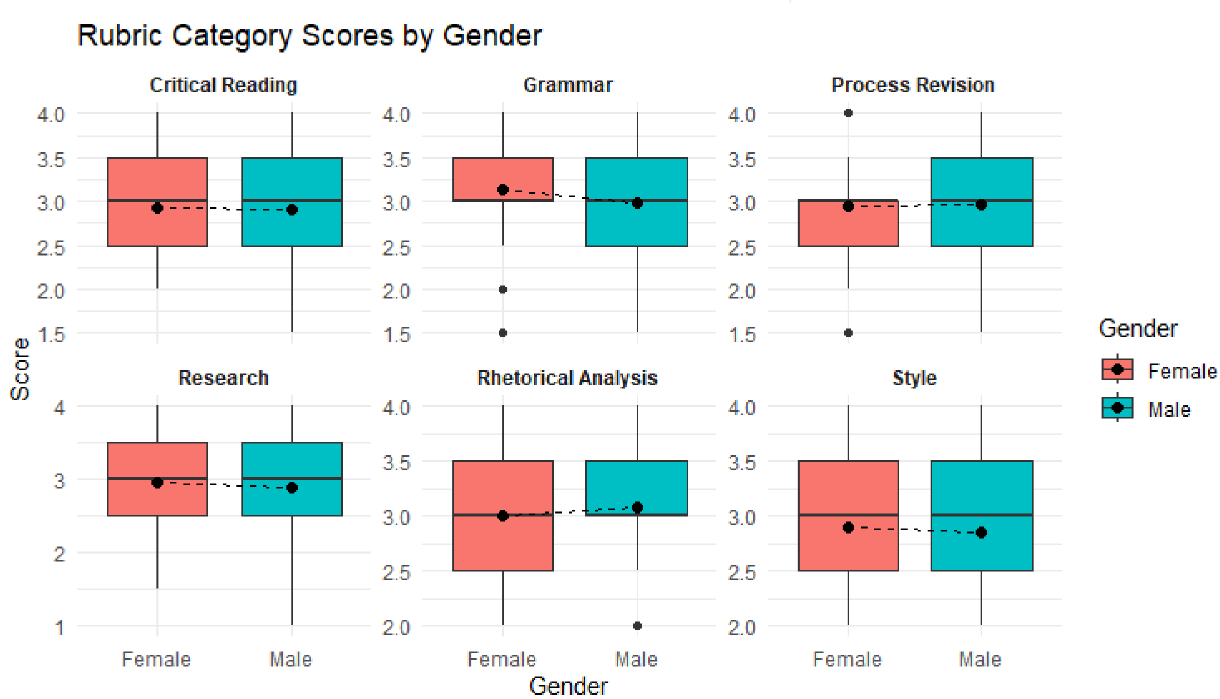
```
# Q07: Variables Involved: Portfolio Rubric Category Score, Gender, Rubric Category
# Method: Two-way Anova

library(tidyr)
library(dplyr)

test <- edu_group %>%
  pivot_longer(cols = c('Process Revision', 'Critical Reading', 'Rhetorical Analysis',
                        'Research', 'Style', 'Grammar'),
               names_to = "Category",
               values_to = "Score") %>%
  mutate(Gender = ifelse(Sex == 1, "Female", "Male"))

mod1.2way <- aov(Score ~ Category*Sex, data=test)
mod1.2way

library(ggplot2)

# Boxplot of Score by Rubric Category
ggplot(test, aes(x = Category, y = Score, fill = Category)) +
  geom_boxplot() +
  stat_summary(fun = mean, geom = "point", color = "black", size = 2) +  # Red dot for group mean
  stat_summary(fun = mean, geom = "line", aes(group = 1), color = "black", linetype = "dashed") +  # Line across means
  labs(title = "Rubric Category vs. Portfolio Rubric Score",
       x = "Rubric Category",
       y = "Score") +
  theme_minimal() +
  theme(axis.text.x = element_text(angle = 45, hjust = 1))

# Summary statistics whole gender
test %>%
  group_by(Category) %>%
  summarise(
    n = n(),
    Min = min(Score, na.rm = TRUE),
    Q1 = quantile(Score, 0.25, na.rm = TRUE),
    Median = median(Score, na.rm = TRUE),
    Q3 = quantile(Score, 0.75, na.rm = TRUE),
    Max = max(Score, na.rm = TRUE),
    Mean = mean(Score, na.rm = TRUE),
    SD = sd(Score, na.rm = TRUE)
  ) %>%
  arrange(Category)
```

```
> mod1.2way <- aov(Score ~ Category*Sex, data=test)
> mod1.2way
Call:
   aov(formula = Score ~ Category * Sex, data = test)

Terms:
                 Category       Sex Category:Sex Residuals
Sum of Squares    3.25578   0.20421      0.93247 226.37815
Deg. of Freedom         5         1            5       708

Residual standard error: 0.5654584
Estimated effects may be unbalanced
```

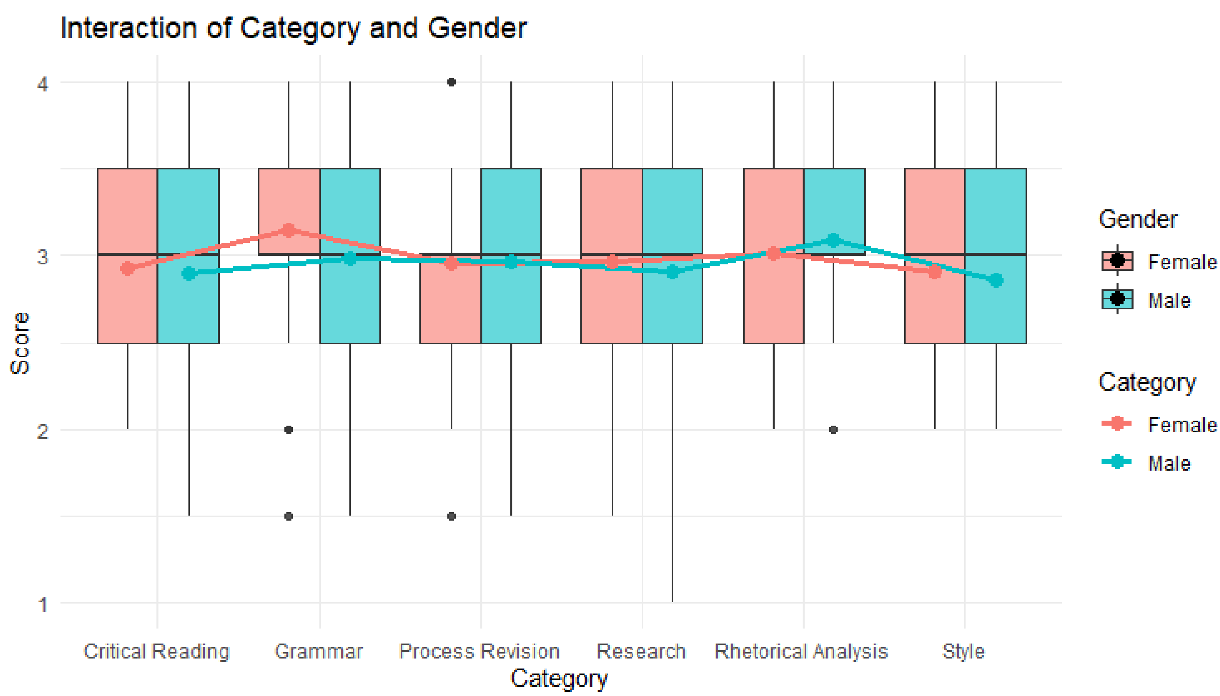



```
> test %>%
+   group_by(Category) %>%
+   summarise(
+     n = n(),
+     Min = min(Score, na.rm = TRUE),
+     Q1 = quantile(Score, 0.25, na.rm = TRUE),
+     Median = median(Score, na.rm = TRUE),
+     Q3 = quantile(Score, 0.75, na.rm = TRUE),
+     Max = max(Score, na.rm = TRUE),
+     Mean = mean(Score, na.rm = TRUE),
+     SD = sd(Score, na.rm = TRUE)
+   ) %>%
+   arrange(Category)
# A tibble: 6 × 9
  Category                n   Min    Q1 Median    Q3   Max  Mean    SD
  <chr>               <int> <dbl> <dbl>  <dbl> <dbl> <dbl> <dbl> <dbl>
1 Critical Reading      120   1.5  2.5       3   3.5     4  2.91 0.583
2 Grammar               120   1.5  2.7       3   3.5     4  3.07 0.582
3 Process Revision      120   1.5  2.5       3   3.5     4  2.96 0.557
4 Research              120   1    2.5       3   3.5     4  2.94 0.621
5 Rhetorical Analysis   120   2    2.92      3   3.5     4  3.04 0.470
6 Style                 120   2    2.5       3   3.5     4  2.88 0.564
```

```r
# boxplot by gender
ggplot(test, aes(x = Gender, y = Score, fill = Gender)) +
  geom_boxplot() +
  stat_summary(fun = mean, geom = "point", color = "black", size = 2) +  # Red dot for group mean
  stat_summary(fun = mean, geom = "line", aes(group = 1), color = "black", linetype = "dashed") +  # Line across means
  facet_wrap(~ Category, scales = "free_y") +
  labs(
    title = "Rubric Category Scores by Gender",
    x = "Gender",
    y = "Score"
  ) +
  theme_minimal() +
  theme(strip.text = element_text(face = "bold"))

# Summary statistics by gender
test %>%
  group_by(Category, Gender) %>%
  summarise(
    n = n(),
    Min = min(Score, na.rm = TRUE),
    Q1 = quantile(Score, 0.25, na.rm = TRUE),
    Median = median(Score, na.rm = TRUE),
    Q3 = quantile(Score, 0.75, na.rm = TRUE),
    Max = max(Score, na.rm = TRUE),
    Mean = mean(Score, na.rm = TRUE),
    SD = sd(Score, na.rm = TRUE)
  ) %>%
  arrange(Category, Gender)

# two-way interaction
ggplot(test, aes(x = Category, y = Score, fill = as.factor(Gender))) +
  geom_boxplot(position = position_dodge(width = 0.75), alpha = 0.6) +  # Boxplots with fill and transparency
  stat_summary(aes(group = Gender, color = as.factor(Gender)), fun = mean, geom = "line",
               position = position_dodge(width = 0.75), size = 1.2) +  # Mean interaction lines
  stat_summary(aes(group = Gender, color = as.factor(Gender)), fun = mean, geom = "point",
               position = position_dodge(width = 0.75), size = 3) +  # Mean points
  labs(title = "Interaction of Category and Gender", x = "Category", y = "Score", fill = "Gender", color = "Category") +
  theme_minimal()

# Function to compute mean and sd
compute_stats <- function(df, group_vars) {
  df %>%
    group_by(across(all_of(group_vars))) %>%
    summarise(
      Mean = mean(Score),
      SD = sd(Score),
      .groups = 'drop'
    )
}
```



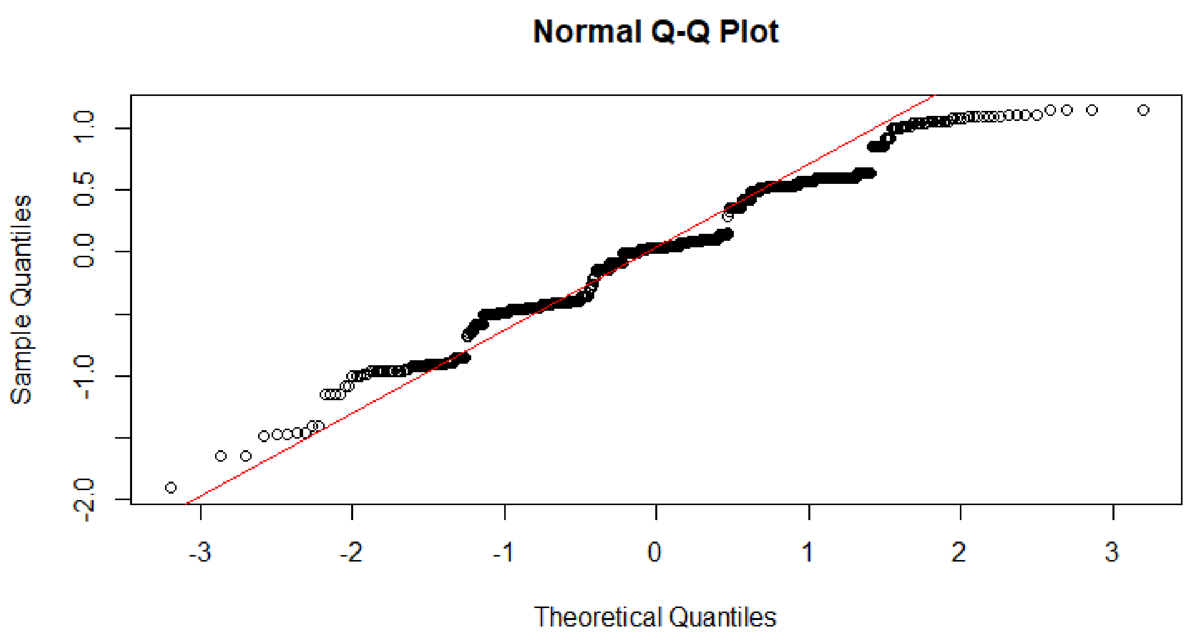
```
`summarise()` has grouped output by 'Category'. You can override using the
`.groups` argument.
# A tibble: 12 × 10
# Groups:   Category [6]
   Category           Gender     n   Min    Q1 Median    Q3   Max  Mean    SD
   <fct>              <fct>  <int> <dbl> <dbl>  <dbl> <dbl> <dbl> <dbl> <dbl>
 1 Critical Reading   Female    65   2     2.5      3   3.5     4  2.93 0.579
 2 Critical Reading   Male      55   1.5   2.5      3   3.5     4  2.9  0.591
 3 Grammar            Female    65   1.5   3        3   3.5     4  3.14 0.593
 4 Grammar            Male      55   1.5   2.5      3   3.5     4  2.99 0.561
 5 Process Revision   Female    65   1.5   2.5      3   3       4  2.95 0.583
 6 Process Revision   Male      55   1.5   2.5      3   3.5     4  2.97 0.531
 7 Research           Female    65   1.5   2.5      3   3.5     4  2.97 0.622
 8 Research           Male      55   1     2.5      3   3.5     4  2.91 0.623
 9 Rhetorical Analysis Female   65   2     2.5      3   3.5     4  3.01 0.483
10 Rhetorical Analysis Male     55   2     3        3   3.5     4  3.09 0.455
11 Style              Female    65   2     2.5      3   3.5     4  2.91 0.538
12 Style              Male      55   2     2.5      3   3.5     4  2.86 0.597
```




```
>  # Compute statistics for two-way interactions
>  stats_SG <- compute_stats(test, c("Category", "Gender"))
>  stats_SG
# A tibble: 12 × 4
   Category           Gender  Mean    SD
   <fct>              <fct>  <dbl> <dbl>
 1 Critical Reading   Female  2.93 0.579
 2 Critical Reading   Male    2.9  0.591
 3 Grammar            Female  3.14 0.593
 4 Grammar            Male    2.99 0.561
 5 Process Revision   Female  2.95 0.583
 6 Process Revision   Male    2.97 0.531
```

```
 7 Research              Female  2.97 0.622
 8 Research              Male    2.91 0.623
 9 Rhetorical Analysis  Female  3.01 0.483
10 Rhetorical Analysis  Male    3.09 0.455
11 Style                 Female  2.91 0.538
12 Style                 Male    2.86 0.597
```

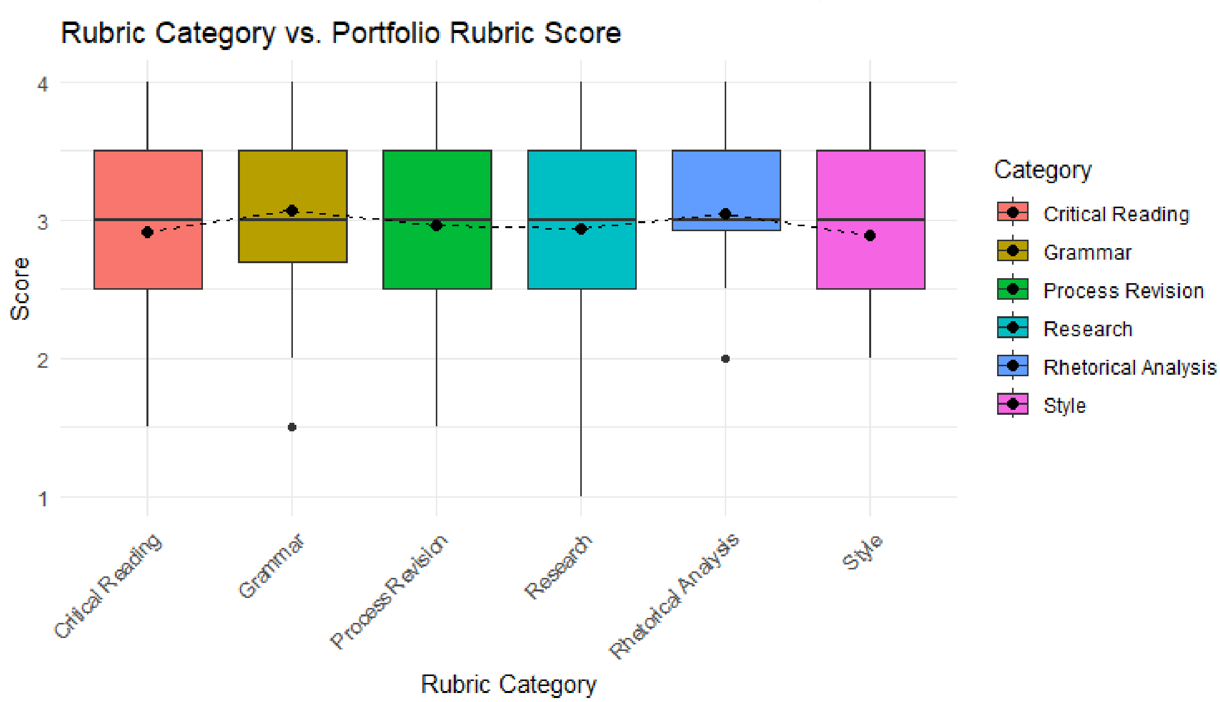
```
# residuals
resid1 = mod1.2way$residuals

qqnorm(resid1) # Q-Q plot
qqline(resid1, col = "red") # Adds the reference line

# shapiro test
shapiro.test(resid1)

# nonparametric
# Now run ART ANOVA
mod_art <- art(Score ~ Category * Gender, data = test)

# Type III ANOVA table
anova(mod_art)

# post hoc-test
art.con(mod_art, "Gender", adjust = "bonferroni")
art.con(mod_art, "Category", adjust = "bonferroni")

test$Gender <- factor(test$Gender, levels = c("Female", "Male"))
```



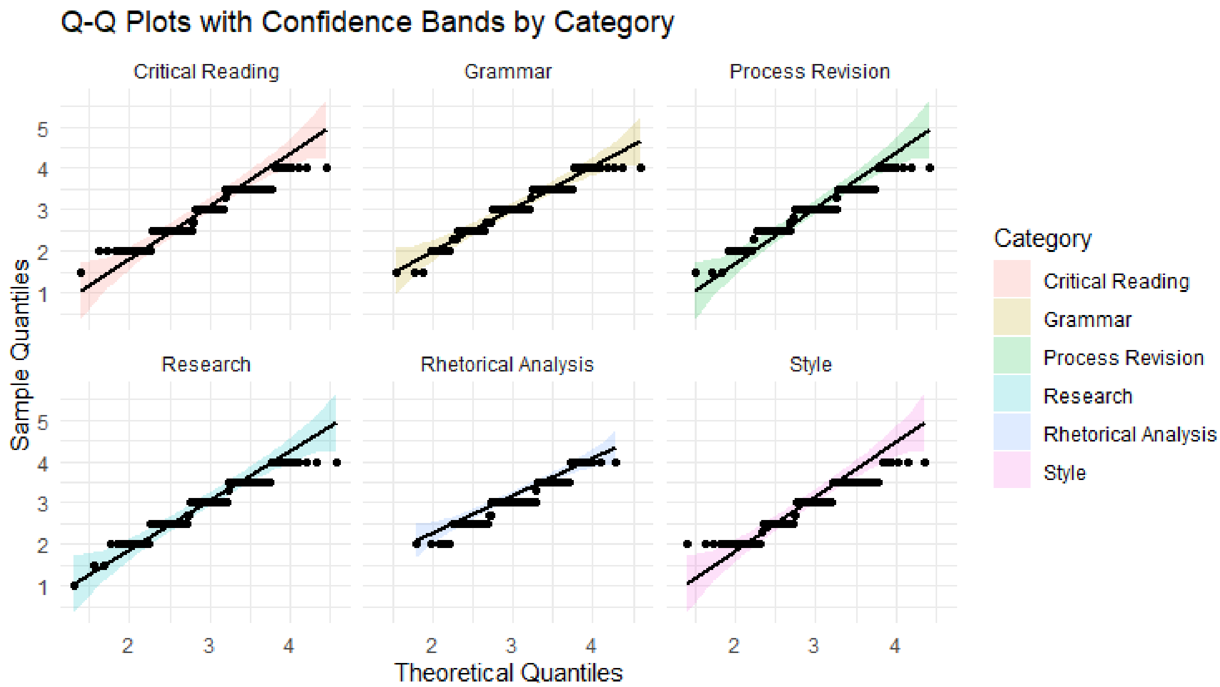
```
>  # shapiro test
>  shapiro.test(resid1)

	Shapiro-Wilk normality test

data:  resid1
W = 0.97692, p-value = 3.093e-09

>  # nonparametric
>  # Now run ART ANOVA
>  mod_art <- art(Score ~ Category * Gender, data = test)
>  # Type III ANOVA table
>  anova(mod_art)
Analysis of Variance of Aligned Rank Transformed Data

Table Type: Anova Table (Type III tests)
Model: No Repeated Measures (lm)
Response: art(Score)

                  Df Df.res F value  Pr(>F)
1 Category         5    708 1.81517 0.10759
2 Gender           1    708 0.72462 0.39492
3 Category:Gender  5    708 0.89395 0.48457
---
Signif. codes:   0 ‘***’ 0.001 ‘**’ 0.01 ‘*’ 0.05 ‘.’ 0.1 ‘ ’ 1
>  # post hoc-test
>  art.con(mod_art, "Gender", adjust = "bonferroni")
NOTE: Results may be misleading due to involvement in interactions
 contrast      estimate   SE  df t.ratio p.value
 Female - Male     13.2 15.6 708   0.851  0.3949

Results are averaged over the levels of: Category

>  art.con(mod_art, "Category", adjust = "bonferroni")
NOTE: Results may be misleading due to involvement in interactions
 contrast                               estimate   SE  df t.ratio p.value
 Critical Reading - Grammar               -54.92 26.7 708  -2.060  0.5969
 Critical Reading - Process Revision      -18.84 26.7 708  -0.707  1.0000
 Critical Reading - Research              -10.69 26.7 708  -0.401  1.0000
 Critical Reading - Rhetorical Analysis   -42.94 26.7 708  -1.610  1.0000
 Critical Reading - Style                  11.55 26.7 708   0.433  1.0000
 Grammar - Process Revision                36.08 26.7 708   1.353  1.0000
 Grammar - Research                        44.22 26.7 708   1.659  1.0000
 Grammar - Rhetorical Analysis             11.98 26.7 708   0.449  1.0000
 Grammar - Style                           66.47 26.7 708   2.493  0.1934
 Process Revision - Research                8.15 26.7 708   0.305  1.0000
 Process Revision - Rhetorical Analysis   -24.10 26.7 708  -0.904  1.0000
 Process Revision - Style                  30.39 26.7 708   1.140  1.0000
 Research - Rhetorical Analysis           -32.25 26.7 708  -1.209  1.0000
 Research - Style                          22.25 26.7 708   0.834  1.0000
 Rhetorical Analysis - Style               54.49 26.7 708   2.044  0.6202
```

```
Results are averaged over the levels of: Gender
P value adjustment: bonferroni method for 15 tests
```

```r
# Q08: Variables: Portfolio Rubric Score, Portfolio Rubric Category;
# Method: One-way ANOVA

# Boxplot of Score by Rubric Category
ggplot(test, aes(x = Category, y = Score, fill = Category)) +
  geom_boxplot() +
  stat_summary(fun = mean, geom = "point", color = "black", size = 2) +  # Red dot for group mean
  stat_summary(fun = mean, geom = "line", aes(group = 1), color = "black", linetype = "dashed") +  # Line across means
  labs(title = "Rubric Category vs. Portfolio Rubric Score",
       x = "Rubric Category",
       y = "Score") +
  theme_minimal() +
  theme(axis.text.x = element_text(angle = 45, hjust = 1))

# Summary statistics whole gender
test %>%
  group_by(Category) %>%
  summarise(
    n = n(),
    Min = min(Score, na.rm = TRUE),
    Q1 = quantile(Score, 0.25, na.rm = TRUE),
    Median = median(Score, na.rm = TRUE),
    Q3 = quantile(Score, 0.75, na.rm = TRUE),
    Max = max(Score, na.rm = TRUE),
    Mean = mean(Score, na.rm = TRUE),
    SD = sd(Score, na.rm = TRUE)
  ) %>%
  arrange(Category)


# qq plots of all categories
ggplot(test, aes(sample = Score)) +
  stat_qq_band(aes(fill = Category), alpha = 0.2) + # Confidence band
  stat_qq_point() + # Points
  stat_qq_line() + # Q-Q line
  facet_wrap(~ Category) +
  labs(title = "Q-Q Plots with Confidence Bands by Category",
       x = "Theoretical Quantiles",
       y = "Sample Quantiles") +
  theme_minimal()
```



```
> # Summary statistics whole gender
> test %>%
+   group_by(Category) %>%
+   summarise(
+     n = n(),
+     Min = min(Score, na.rm = TRUE),
+     Q1 = quantile(Score, 0.25, na.rm = TRUE),
+     Median = median(Score, na.rm = TRUE),
+     Q3 = quantile(Score, 0.75, na.rm = TRUE),
+     Max = max(Score, na.rm = TRUE),
+     Mean = mean(Score, na.rm = TRUE),
+     SD = sd(Score, na.rm = TRUE)
+   ) %>%
+   arrange(Category)
# A tibble: 6 × 9
  Category                n   Min    Q1 Median    Q3   Max  Mean    SD
  <fct>               <int> <dbl> <dbl>  <dbl> <dbl> <dbl> <dbl> <dbl>
1 Critical Reading      120   1.5  2.5       3   3.5     4  2.91 0.583
2 Grammar               120   1.5  2.7       3   3.5     4  3.07 0.582
3 Process Revision      120   1.5  2.5       3   3.5     4  2.96 0.557
4 Research              120   1    2.5       3   3.5     4  2.94 0.621
5 Rhetorical Analysis   120   2    2.92      3   3.5     4  3.04 0.470
6 Style                 120   2    2.5       3   3.5     4  2.88 0.564
```

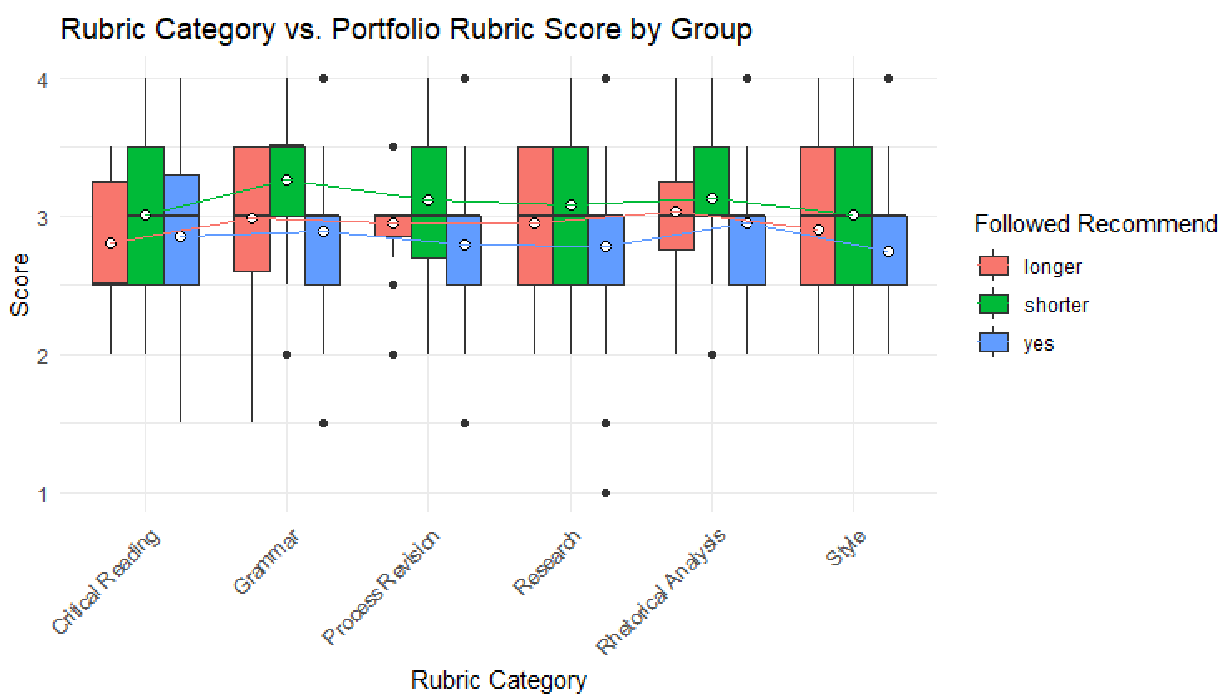



```
# shapiro test
test %>%
  group_by(Category) %>%
  summarise(p_value = shapiro.test(Score)$p.value)

test$Category <- factor(test$Category)

# nonparametric kruskal wallis test
mod_art8 <- art(Score ~ Category, data = test)
kruskal.test(Score ~ Category, data = test)

# Type III ANOVA table
anova(mod_art8)

# post hoc-test
art.con(mod_art8, "Category", adjust = "bonferroni")
```

```
> # shapiro test
> test %>%
+   group_by(Category) %>%
+   summarise(p_value = shapiro.test(Score)$p.value)
# A tibble: 6 × 2
  Category              p_value
  <fct>                   <dbl>
1 Critical Reading    0.00000557
2 Grammar             0.00000892
3 Process Revision    0.0000124
4 Research            0.0000322
5 Rhetorical Analysis 0.000000289
6 Style               0.00000101
> test$Category <- factor(test$Category)
> # nonparametric kruskal wallis test
```

```
> mod_art8 <- art(Score ~ Category, data = test)
> kruskal.test(Score ~ Category, data = test)

	Kruskal-Wallis rank sum test

data:  Score by Category
Kruskal-Wallis chi-squared = 10.091, df = 5, p-value = 0.0727

> # Type III ANOVA table
> anova(mod_art8)
Analysis of Variance of Aligned Rank Transformed Data

Table Type: Anova Table (Type III tests)
Model: No Repeated Measures (lm)
Response: art(Score)

           Df Df.res F value   Pr(>F)  
1 Category  5    714  2.0327 0.072096 .
---
Signif. codes:  0 ‘***’ 0.001 ‘**’ 0.01 ‘*’ 0.05 ‘.’ 0.1 ‘ ’ 1
> # post hoc-test
> art.con(mod_art8, "Category", adjust = "bonferroni")
 contrast                               estimate   SE  df t.ratio p.value
 Critical Reading - Grammar               -59.05 25.9 714  -2.277  0.3464
 Critical Reading - Process Revision      -16.68 25.9 714  -0.643  1.0000
 Critical Reading - Research              -12.83 25.9 714  -0.495  1.0000
 Critical Reading - Rhetorical Analysis   -43.75 25.9 714  -1.687  1.0000
 Critical Reading - Style                   9.47 25.9 714   0.365  1.0000
 Grammar - Process Revision                42.37 25.9 714   1.634  1.0000
 Grammar - Research                        46.22 25.9 714   1.782  1.0000
 Grammar - Rhetorical Analysis             15.30 25.9 714   0.590  1.0000
 Grammar - Style                           68.53 25.9 714   2.642  0.1264
 Process Revision - Research                3.85 25.9 714   0.148  1.0000
 Process Revision - Rhetorical Analysis   -27.07 25.9 714  -1.044  1.0000
 Process Revision - Style                  26.15 25.9 714   1.008  1.0000
 Research - Rhetorical Analysis           -30.92 25.9 714  -1.192  1.0000
 Research - Style                          22.30 25.9 714   0.860  1.0000
 Rhetorical Analysis - Style               53.22 25.9 714   2.052  0.6082

P value adjustment: bonferroni method for 15 tests
```

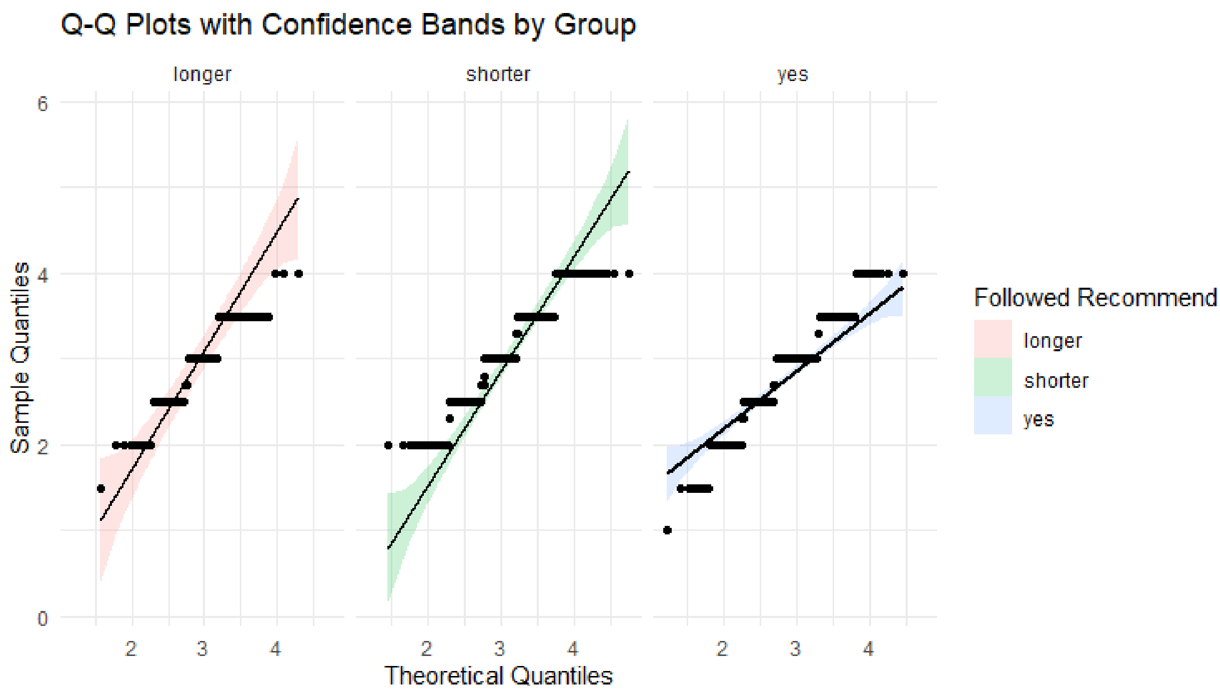
```r
# Q09: Variables: Porfolio Rubric Category Score, Group, Rubric Category
# Method: Two-Way Anova

test$`Followed Recommend` <- factor(test$`Followed Recommend`)

# Boxplot of Score by Rubric Category by Group (`Followed Recommend`)
ggplot(test, aes(x = Category, y = Score, fill = `Followed Recommend`)) +
  geom_boxplot(position = position_dodge(width = 0.75)) +
  stat_summary(
    fun = mean, geom = "point", aes(group = `Followed Recommend`), shape = 21, color = "black",
    fill = "white", size = 2, position = position_dodge(width = 0.75)) +
  stat_summary( fun = mean, geom = "line", aes(group = `Followed Recommend`, color = `Followed Recommend`),
                position = position_dodge(width = 0.75) ) +
  labs(
    title = "Rubric Category vs. Portfolio Rubric Score by Group",
    x = "Rubric Category",
    y = "Score") +
  theme_minimal() +
  theme(axis.text.x = element_text(angle = 45, hjust = 1))

# Summary statistics whole sequence
test %>%
  group_by(Category, `Followed Recommend`) %>%
  summarise(
    Min = min(Score, na.rm = TRUE),
    Q1 = quantile(Score, 0.25, na.rm = TRUE),
    Median = median(Score, na.rm = TRUE),
    Q3 = quantile(Score, 0.75, na.rm = TRUE),
    Max = max(Score, na.rm = TRUE),
    Mean = mean(Score, na.rm = TRUE),
    SD = sd(Score, na.rm = TRUE)
  ) %>%
  arrange(Category)


mod9.2way <- aov(Score ~ Category*`Followed Recommend`,test)
resid9 <- mod9.2way$residuals
```




```r
>  test %>%
+    group_by(Category, `Followed Recommend`) %>%
+    summarise(
+      Min = min(Score, na.rm = TRUE),
```

```
+       Q1 = quantile(Score, 0.25, na.rm = TRUE),
+       Median = median(Score, na.rm = TRUE),
+       Q3 = quantile(Score, 0.75, na.rm = TRUE),
+       Max = max(Score, na.rm = TRUE),
+       Mean = mean(Score, na.rm = TRUE),
+       SD = sd(Score, na.rm = TRUE)
+     ) %>%
+     arrange(Category)
`summarise()` has grouped output by 'Category'. You can override using the
`.groups` argument.
# A tibble: 18 × 9
# Groups:   Category [6]
   Category           `Followed Recommend`   Min    Q1 Median    Q3   Max  Mean    SD
   <fct>              <fct>                <dbl> <dbl>  <dbl> <dbl> <dbl> <dbl> <dbl>
 1 Critical Reading   longer                 2    2.5     2.5  3.25   3.5  2.8  0.528
 2 Critical Reading   shorter                2    2.5     3    3.5    4    3.00 0.608
 3 Critical Reading   yes                    1.5  2.5     3    3.3    4    2.85 0.564
 4 Grammar            longer                 1.5  2.6     3    3.5    3.5  2.98 0.632
 5 Grammar            shorter                2    3       3.5  3.5    4    3.26 0.552
 6 Grammar            yes                    1.5  2.5     3    3      4    2.88 0.540
 7 Process Revision   longer                 2    2.85    3    3      3.5  2.95 0.405
 8 Process Revision   shorter                2    2.7     3    3.5    4    3.11 0.540
 9 Process Revision   yes                    1.5  2.5     3    3      4    2.79 0.575
10 Research           longer                 2    2.5     3    3.5    3.5  2.95 0.485
11 Research           shorter                2    2.5     3    3.5    4    3.08 0.596
12 Research           yes                    1    2.5     3    3      4    2.78 0.654
13 Rhetorical Analysis longer                2    2.75    3    3.25   4    3.03 0.550
14 Rhetorical Analysis shorter               2    3       3    3.5    4    3.13 0.475
15 Rhetorical Analysis yes                   2    2.5     3    3      4    2.94 0.425
16 Style              longer                 2    2.5     3    3.5    4    2.9  0.660
17 Style              shorter                2    2.5     3    3.5    4    3.01 0.536
18 Style              yes                    2    2.5     3    3      4    2.74 0.540
```

```
# qq plots of all categories by Group
ggplot(test, aes(sample = Score)) +
  stat_qq_band(aes(fill = `Followed Recommend`), alpha = 0.2) + # Confidence band
  stat_qq_point() + # Points
  stat_qq_line() + # Q-Q line
  facet_wrap(~ `Followed Recommend`) +
  labs(title = "Q-Q Plots with Confidence Bands by Group",
       x = "Theoretical Quantiles",
       y = "Sample Quantiles") +
  theme_minimal()

#shaprio test
test %>%
  group_by(`Followed Recommend`) %>%
  summarise(p_value = shapiro.test(resid9)$p.value)

# Rename variable to avoid issues
test$Followed_Recommend <- test$`Followed Recommend`

# nonparametric
# Now run ART ANOVA
mod_art9 <- art(Score ~ Category*Followed_Recommend,test)

# Type III ANOVA table
anova(mod_art9)

# post hoc-test
art.con(mod_art9, "Category", adjust = "bonferroni")
art.con(mod_art9, "Followed_Recommend", adjust = "bonferroni")
```

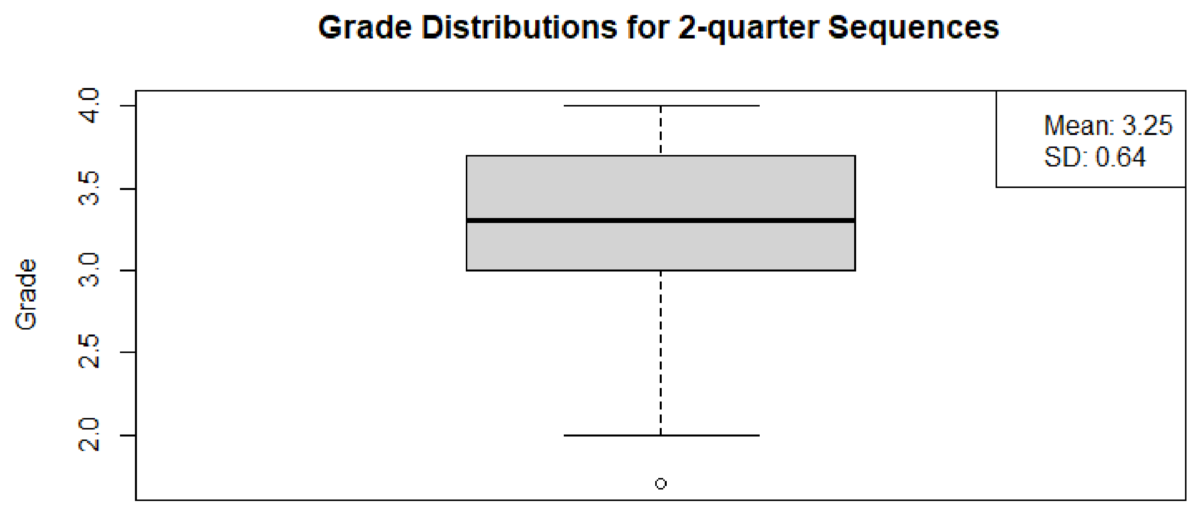



```
> #shaprio test
> test %>%
```

```
+    group_by(`Followed Recommend`) %>%
+    summarise(p_value = shapiro.test(resid9)$p.value)
# A tibble: 3 × 2
  `Followed Recommend` p_value
  <fct>                  <dbl>
1 longer               0.00340
2 shorter              0.00340
3 yes                  0.00340
>  # Rename variable to avoid issues
>  test$Followed_Recommend <- test$`Followed Recommend`
>  # nonparametric
>  # Now run ART ANOVA
>  mod_art9 <- art(Score ~ Category*Followed_Recommend,test)
>  # Type III ANOVA table
>  anova(mod_art9)
Analysis of Variance of Aligned Rank Transformed Data


Table Type: Anova Table (Type III tests)
Model: No Repeated Measures (lm)
Response: art(Score)


                              Df Df.res  F value     Pr(>F)
1 Category                     5    702  1.47474    0.19582
2 Followed_Recommend           2    702 20.66223 1.9085e-09 ***
3 Category:Followed_Recommend 10    702  0.34109    0.96968
---
Signif. codes:   0 ‘***’ 0.001 ‘**’ 0.01 ‘*’ 0.05 ‘.’ 0.1 ‘ ’ 1
>  # post hoc-test
>  art.con(mod_art9, "Category", adjust = "bonferroni")
NOTE: Results may be misleading due to involvement in interactions
 contrast                                 estimate   SE  df t.ratio p.value
 Critical Reading - Grammar                 -62.22 31.9 702  -1.953  0.7686
 Critical Reading - Process Revision        -13.79 31.9 702  -0.433  1.0000
 Critical Reading - Research                 -9.64 31.9 702  -0.303  1.0000
 Critical Reading - Rhetorical Analysis     -39.45 31.9 702  -1.238  1.0000
 Critical Reading - Style                    12.33 31.9 702   0.387  1.0000
 Grammar - Process Revision                  48.43 31.9 702   1.520  1.0000
 Grammar - Research                          52.58 31.9 702   1.650  1.0000
 Grammar - Rhetorical Analysis               22.77 31.9 702   0.715  1.0000
 Grammar - Style                             74.55 31.9 702   2.340  0.2937
 Process Revision - Research                  4.15 31.9 702   0.130  1.0000
 Process Revision - Rhetorical Analysis     -25.66 31.9 702  -0.805  1.0000
 Process Revision - Style                    26.12 31.9 702   0.820  1.0000
 Research - Rhetorical Analysis             -29.81 31.9 702  -0.935  1.0000
 Research - Style                            21.97 31.9 702   0.690  1.0000
 Rhetorical Analysis - Style                 51.78 31.9 702   1.625  1.0000

Results are averaged over the levels of: Followed_Recommend
P value adjustment: bonferroni method for 15 tests
>  art.con(mod_art9, "Followed_Recommend", adjust = "bonferroni")
NOTE: Results may be misleading due to involvement in interactions
 contrast         estimate   SE  df t.ratio p.value
 longer - shorter    -64.9 24.1 702  -2.687  0.0221
 longer - yes         38.9 24.5 702   1.587  0.3386
 shorter - yes       103.8 16.2 702   6.388  <.0001
```

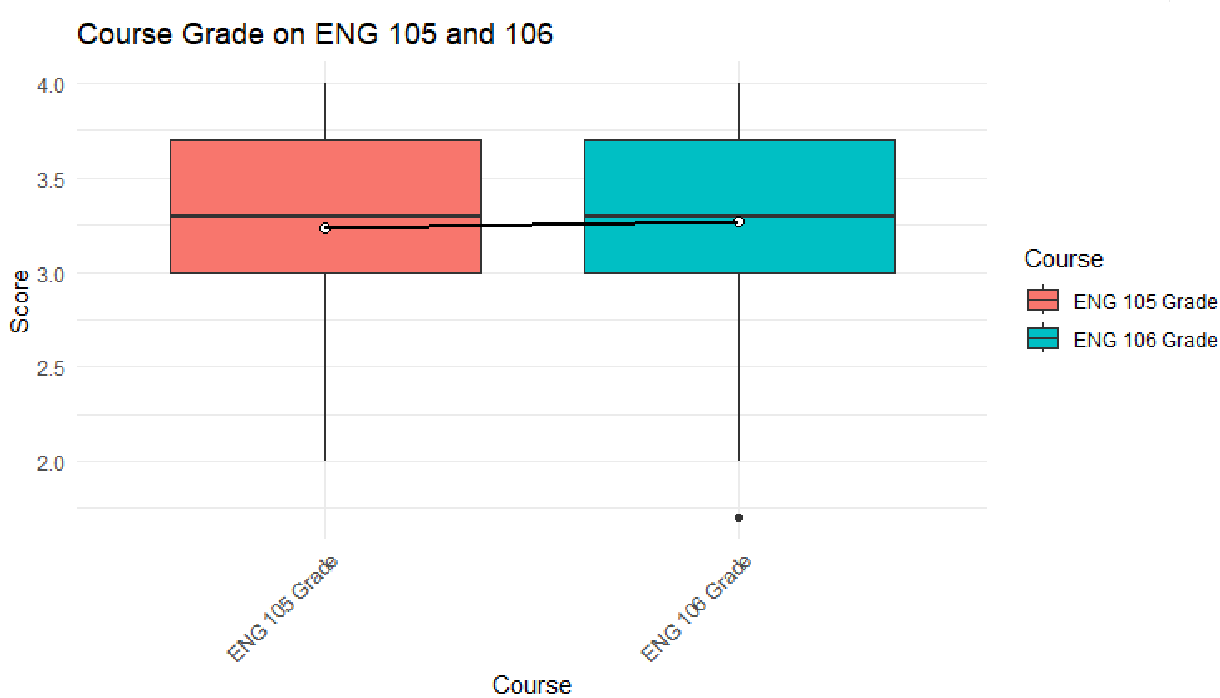
```
Results are averaged over the levels of: Category
P value adjustment: bonferroni method for 3 tests
```

```
# Q10: Variables Involved: Eng105 & 106 Grade, Course,
# Method: Dependent 2 Sample T-Test

test2 = edu_group

# Add Student ID
test2 <- test2 %>%
  mutate(Student_ID = row_number())

# Pivot longer
test2 <- test2 %>%
  pivot_longer(cols = c('ENG 105 Grade', 'ENG 106 Grade'),
               names_to = "Course",
               values_to = "Grade") %>%
  filter(!is.na(Grade))

# Boxplot with mean and SD displayed in the legend (Both ENG 105 and 106)
boxplot(test2$Grade,
        ylab = "Grade",
        main = "Grade Distributions for 2-quarter Sequences")
legend("topright",
       legend = c(paste("Mean:", round(mean(test2$Grade), 2)),
                  paste("SD:", round(sd(test2$Grade), 2))))

# grade distribution for all 2 quarter sequence

# Summary statistics
test2 %>%
  summarise(
    Min = min(Grade, na.rm = TRUE),
    Q1 = quantile(Grade, 0.25, na.rm = TRUE),
    Median = median(Grade, na.rm = TRUE),
    Q3 = quantile(Grade, 0.75, na.rm = TRUE),
    Max = max(Grade, na.rm = TRUE),
    Mean = mean(Grade, na.rm = TRUE),
    SD = sd(Grade, na.rm = TRUE)
  )
```




```
> # Summary statistics
> test2 %>%
+   summarise(
+     Min = min(Grade, na.rm = TRUE),
+     Q1 = quantile(Grade, 0.25, na.rm = TRUE),
+     Median = median(Grade, na.rm = TRUE),
+     Q3 = quantile(Grade, 0.75, na.rm = TRUE),
+     Max = max(Grade, na.rm = TRUE),
+     Mean = mean(Grade, na.rm = TRUE),
+     SD = sd(Grade, na.rm = TRUE)
+   )
# A tibble: 1 × 7
    Min    Q1 Median    Q3   Max  Mean    SD
  <dbl> <dbl>  <dbl> <dbl> <dbl> <dbl> <dbl>
1   1.7     3    3.3   3.7     4  3.25 0.642
```

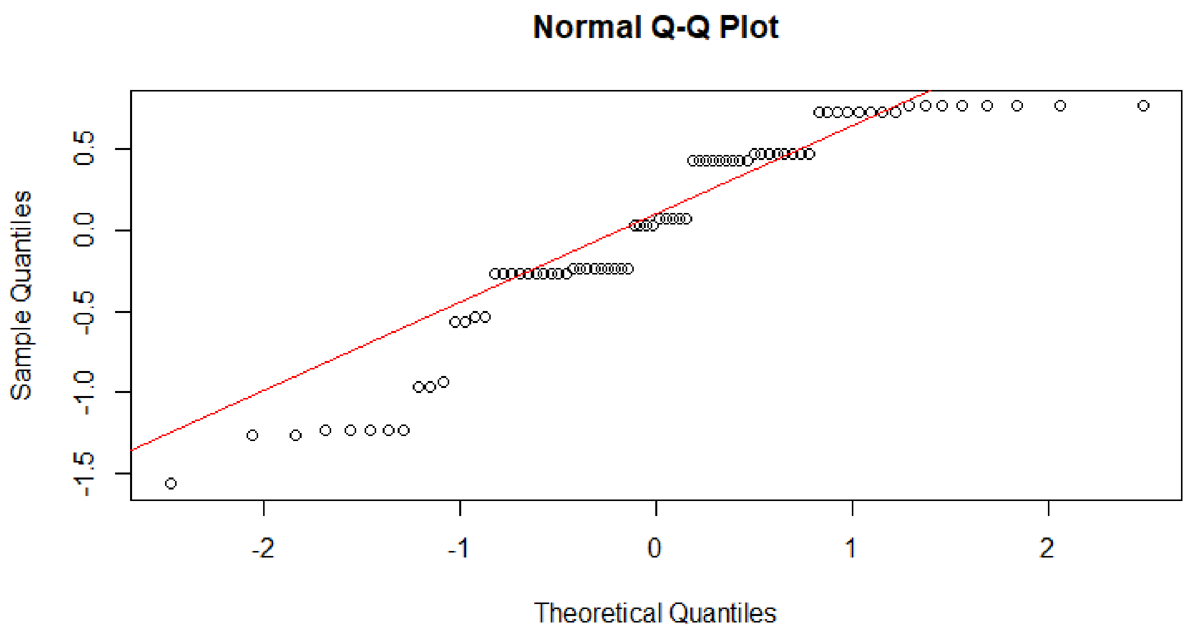
```r
# Boxplot of Course Grade by Course (ENG 105 or 106)
ggplot(test2, aes(x = Course, y = Grade, fill = Course)) +
  geom_boxplot(position = position_dodge(width = 0.75)) +
  stat_summary(
    fun = mean, geom = "point", aes(group = Course), shape = 21, color = "black",
    fill = "white", size = 2, position = position_dodge(width = 0.75)) +
  stat_summary(
    fun = mean, geom = "line", aes(group = 1),
    color = "black", size = 1, position = position_dodge(width = 0.75)
  ) +
  labs(
    title = "Course Grade on ENG 105 and 106",
    x = "Course",
    y = "Score") +
  theme_minimal() +
  theme(axis.text.x = element_text(angle = 45, hjust = 1))


# Summary statistics
test2 %>%
  group_by(Course) %>%
  summarise(
    Min = min(Grade, na.rm = TRUE),
    Q1 = quantile(Grade, 0.25, na.rm = TRUE),
    Median = median(Grade, na.rm = TRUE),
    Q3 = quantile(Grade, 0.75, na.rm = TRUE),
    Max = max(Grade, na.rm = TRUE),
    Mean = mean(Grade, na.rm = TRUE),
    SD = sd(Grade, na.rm = TRUE)
  )
```

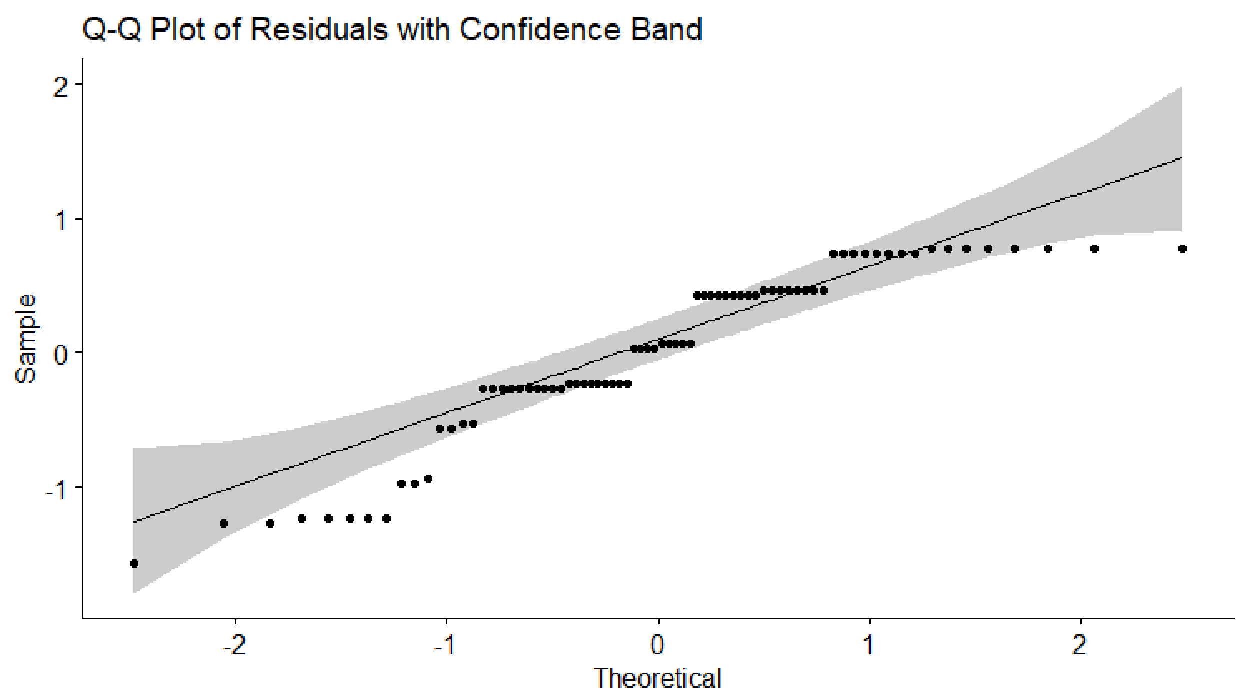



```
> test2 %>%
+   group_by(Course) %>%
```

```
+     summarise(
+       Min = min(Grade, na.rm = TRUE),
+       Q1 = quantile(Grade, 0.25, na.rm = TRUE),
+       Median = median(Grade, na.rm = TRUE),
+       Q3 = quantile(Grade, 0.75, na.rm = TRUE),
+       Max = max(Grade, na.rm = TRUE),
+       Mean = mean(Grade, na.rm = TRUE),
+       SD = sd(Grade, na.rm = TRUE)
+     )
# A tibble: 2 × 8
  Course          Min    Q1 Median    Q3   Max  Mean    SD
  <chr>         <dbl> <dbl>  <dbl> <dbl> <dbl> <dbl> <dbl>
1 ENG 105 Grade   2       3    3.3   3.7     4  3.23 0.658
2 ENG 106 Grade   1.7     3    3.3   3.7     4  3.27 0.634
```

```r
# create line for residuals
line10 <- aov(Grade ~ Course, data=test2)
resid10 <- line10$residuals

qqnorm(resid10) # Q-Q plot
qqline(resid10, col = "red") # Adds the reference line

# ggplot of QQ plot
library(ggpubr)

ggqqplot(resid10, conf.int = TRUE) + ggtitle("Q-Q Plot of Residuals with Confidence Band")

# shapiro test
shapiro.test(resid10) # not normal

# conduct nonparametric wilcoxn test
# Pivot wider for paired t-test
test2_wide <- test2 %>%
  pivot_wider(names_from = Course, values_from = Grade)

# Normality of difference
test2_wide <- test2_wide %>%
  mutate(Diff = `ENG 106 Grade` - `ENG 105 Grade`)

shapiro.test(test2_wide$Diff)

# nonparametric;  wilcoxon test

# Create a new dataset by pivoting the data to wide format
test2_wide <- test2 %>%
  pivot_wider(names_from = Course, values_from = Grade)

# Compute the difference between ENG 106 and ENG 105 grades
test2_wide <- test2_wide %>%
  mutate(Diff = `ENG 106 Grade` - `ENG 105 Grade`)

# Check normality of the differences (Shapiro-Wilk test)
shapiro.test(test2_wide$Diff)  # Perform Shapiro-Wilk test on the differences

# Now perform the Wilcoxon signed-rank test
wilcox.test(test2_wide$`ENG 105 Grade`, test2_wide$`ENG 106 Grade`, paired = TRUE)
```

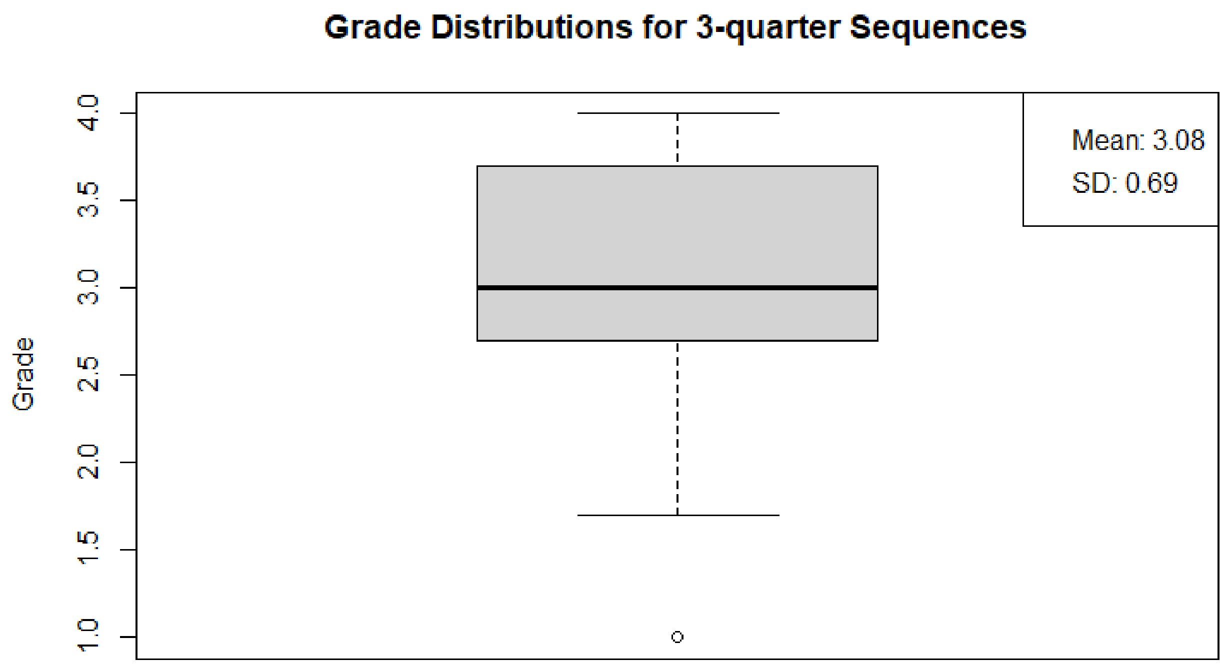





```
>  # shapiro test
>  shapiro.test(resid10) # not normal

        Shapiro-Wilk normality test

data:  resid10
W = 0.90144, p-value = 2.203e-05
```

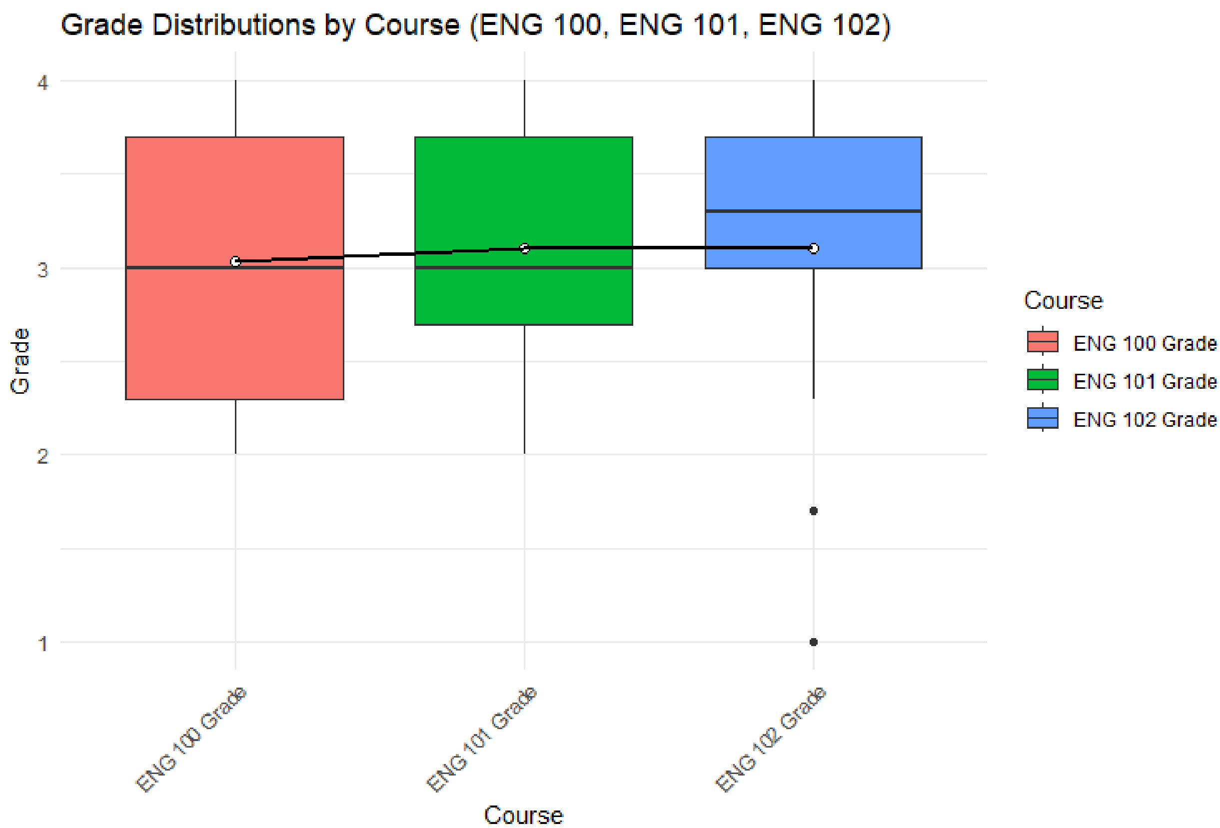
```
>  # conduct nonparametric wilcoxn test
>  # Pivot wider for paired t-test
>  test2_wide <- test2 %>%
+    pivot_wider(names_from = Course, values_from = Grade)
>  # Normality of difference
>  test2_wide <- test2_wide %>%
+    mutate(Diff = `ENG 106 Grade` - `ENG 105 Grade`)
>  shapiro.test(test2_wide$Diff)

	Shapiro-Wilk normality test

data:  test2_wide$Diff
W = 0.94044, p-value = 0.04329

>  # Create a new dataset by pivoting the data to wide format
>  test2_wide <- test2 %>%
+    pivot_wider(names_from = Course, values_from = Grade)
>  # Compute the difference between ENG 106 and ENG 105 grades
>  test2_wide <- test2_wide %>%
+    mutate(Diff = `ENG 106 Grade` - `ENG 105 Grade`)
>  # Check normality of the differences (Shapiro-Wilk test)
>  shapiro.test(test2_wide$Diff)  # Perform Shapiro-Wilk test on the differences

	Shapiro-Wilk normality test

data:  test2_wide$Diff
W = 0.94044, p-value = 0.04329

>  # Now perform the Wilcoxon signed-rank test
>  wilcox.test(test2_wide$`ENG 105 Grade`, test2_wide$`ENG 106 Grade`, paired =
TRUE)

	Wilcoxon signed rank test with continuity correction

data:  test2_wide$`ENG 105 Grade` and test2_wide$`ENG 106 Grade`
V = 184, p-value = 0.9126
alternative hypothesis: true location shift is not equal to 0
```

```r
# Q11: Variables: ENG 100, 101, and 102 Grade
# Method: Repeated Measures One Way ANOVA (nonparametric alternative: Friedman Test)

test3 = edu_group

# Add Student ID
test3 <- test3 %>%
  mutate(Student_ID = row_number())

# Pivot longer: Course and Grade
test3 <- test3 %>%
  pivot_longer(cols = c('ENG 100 Grade', 'ENG 101 Grade', 'ENG 102 Grade'),
               names_to = "Course",
               values_to = "Grade") %>%
  filter(!is.na(Grade))  # remove NA grades

# Boxplot with mean and SD displayed in the legend (Both ENG 105 and 106)
boxplot(test3$Grade,
        ylab = "Grade",
        main = "Grade Distributions for 3-quarter Sequences")
legend("topright",
       legend = c(paste("Mean:", round(mean(test3$Grade), 2)),
                  paste("SD:", round(sd(test3$Grade), 2))))

# Summary statistics
test3 %>%
  summarise(
    Min = min(Grade, na.rm = TRUE),
    Q1 = quantile(Grade, 0.25, na.rm = TRUE),
    Median = median(Grade, na.rm = TRUE),
    Q3 = quantile(Grade, 0.75, na.rm = TRUE),
    Max = max(Grade, na.rm = TRUE),
    Mean = mean(Grade, na.rm = TRUE),
    SD = sd(Grade, na.rm = TRUE)
  )
```



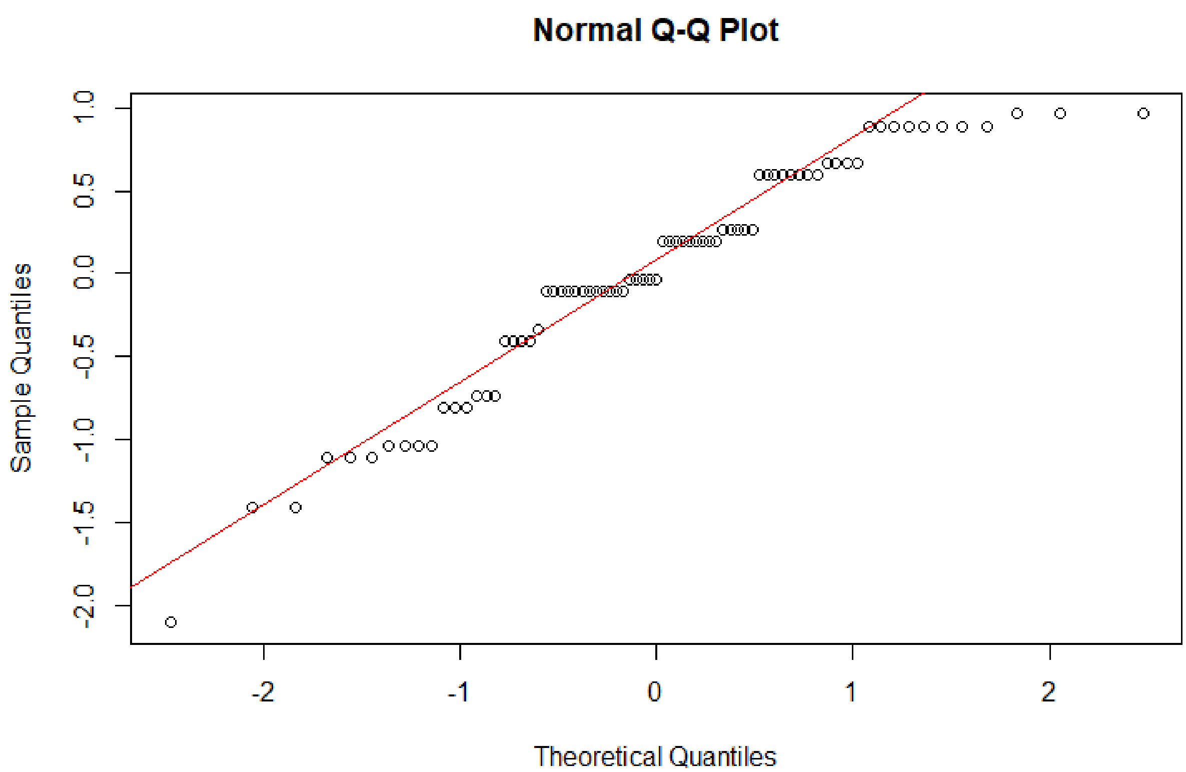

```
>  # Summary statistics
>  test3 %>%
+    summarise(
+      Min = min(Grade, na.rm = TRUE),
+      Q1 = quantile(Grade, 0.25, na.rm = TRUE),
+      Median = median(Grade, na.rm = TRUE),
+      Q3 = quantile(Grade, 0.75, na.rm = TRUE),
+      Max = max(Grade, na.rm = TRUE),
+      Mean = mean(Grade, na.rm = TRUE),
+      SD = sd(Grade, na.rm = TRUE)
+    )
# A tibble: 1 × 7
    Min     Q1 Median     Q3    Max   Mean     SD
  <dbl>  <dbl>  <dbl>  <dbl>  <dbl>  <dbl>  <dbl>
1     1    2.7      3    3.7      4   3.08  0.691
```

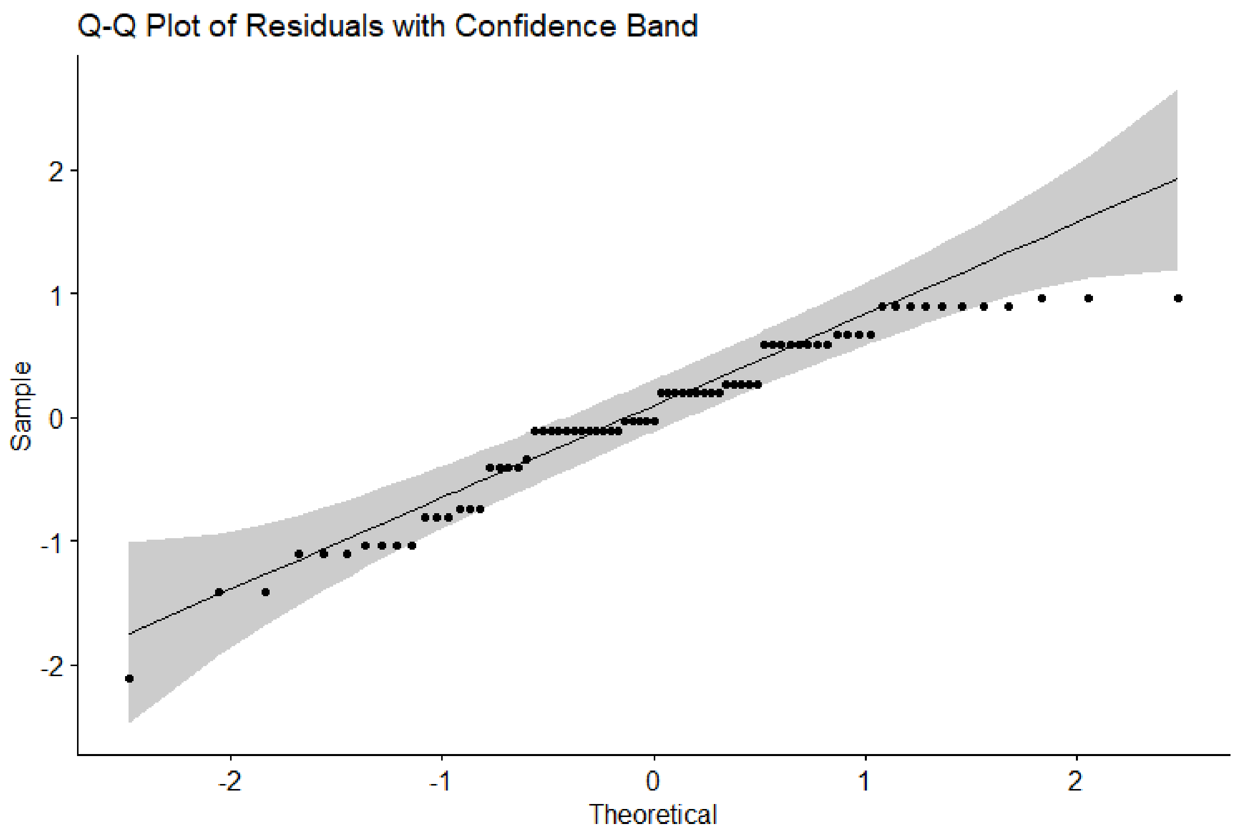
```r
# Boxplot by individual courses
ggplot(test3, aes(x = Course, y = Grade, fill = Course)) +
  geom_boxplot(position = position_dodge(width = 0.75)) +
  stat_summary(
    fun = mean, geom = "point", aes(group = Course), shape = 21, color = "black",
    fill = "white", size = 2, position = position_dodge(width = 0.75)
  ) +
  stat_summary(
    fun = mean, geom = "line", aes(group = 1),
    color = "black", size = 1, position = position_dodge(width = 0.75)
  ) +
  labs(
    title = "Grade Distributions by Course (ENG 100, ENG 101, ENG 102)",
    x = "Course",
    y = "Grade"
  ) +
  theme_minimal() +
  theme(axis.text.x = element_text(angle = 45, hjust = 1))

# Summary statistics
test3 %>%
  group_by(Course) %>%
  summarise(
    Min = min(Grade, na.rm = TRUE),
    Q1 = quantile(Grade, 0.25, na.rm = TRUE),
    Median = median(Grade, na.rm = TRUE),
    Q3 = quantile(Grade, 0.75, na.rm = TRUE),
    Max = max(Grade, na.rm = TRUE),
    Mean = mean(Grade, na.rm = TRUE),
    SD = sd(Grade, na.rm = TRUE)
  )
```




```
> # Summary statistics
> test3 %>%
+   group_by(Course) %>%
+   summarise(
+     Min = min(Grade, na.rm = TRUE),
+     Q1 = quantile(Grade, 0.25, na.rm = TRUE),
+     Median = median(Grade, na.rm = TRUE),
+     Q3 = quantile(Grade, 0.75, na.rm = TRUE),
+     Max = max(Grade, na.rm = TRUE),
+     Mean = mean(Grade, na.rm = TRUE),
+     SD = sd(Grade, na.rm = TRUE)
+   )
# A tibble: 3 × 8
  Course          Min    Q1 Median    Q3   Max  Mean    SD
  <chr>         <dbl> <dbl>  <dbl> <dbl> <dbl> <dbl> <dbl>
1 ENG 100 Grade     2   2.3    3     3.7     4  3.04 0.675
2 ENG 101 Grade     2   2.7    3     3.7     4  3.11 0.636
3 ENG 102 Grade     1   3      3.3   3.7     4  3.11 0.781
```

```
# from the equation
model_lm <- lm(Grade ~ Course, data = test3)

# Extract residuals
resid11 <- residuals(model_lm)

# Shapiro-Wilk test for normality of residuals
shapiro.test(resid11)

# Q-Q plot of residuals
qqnorm(resid11)
qqline(resid11, col = "red")

# ggplot of QQ plot
library(ggpubr)

ggqqplot(resid11, conf.int = TRUE) + ggtitle("Q-Q Plot of Residuals with Confidence Band")


# Nonparametric alternative: Friedman Test

# Pivot wider for Friedman test (each row = one student, columns = ENG 100, 101, 102 grades)
test3_wide <- test3 %>%
  pivot_wider(names_from = Course, values_from = Grade)

# Conduct Friedman Test
friedman.test(as.matrix(test3_wide[, c("ENG 100 Grade", "ENG 101 Grade", "ENG 102 Grade")]))
```

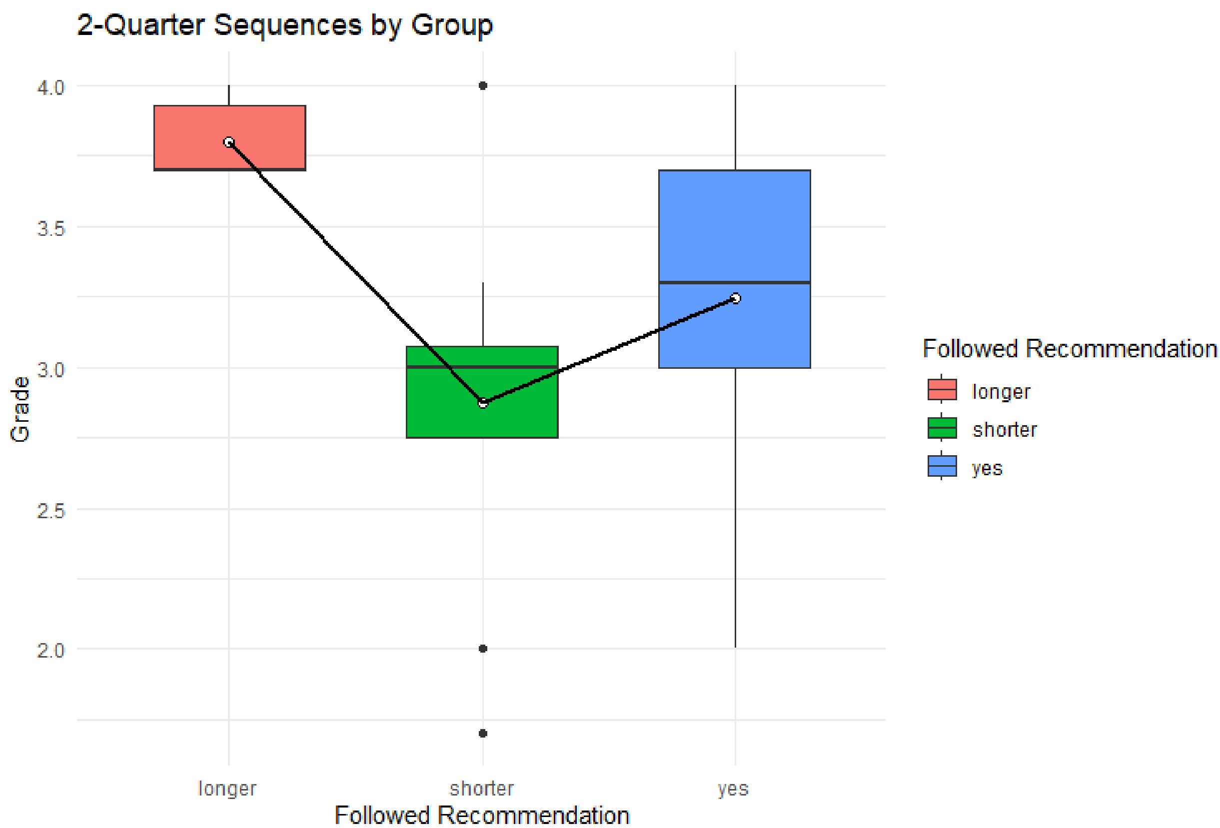





```
>  # Shapiro-Wilk test for normality of residuals
>  shapiro.test(resid11)

      Shapiro-Wilk normality test

data:  resid11
W = 0.94006, p-value = 0.001476

>  # Q-Q plot of residuals
>  qqnorm(resid11)
>  qqline(resid11, col = "red")
>  # ggplot of QQ plot
>  library(ggpubr)
>  ggqqplot(resid11, conf.int = TRUE) + ggtitle("Q-Q Plot of Residuals with
Confidence Band")
>  # Pivot wider for Friedman test (each row = one student, columns = ENG 100,
101, 102 grades)
>  test3_wide <- test3 %>%
+    pivot_wider(names_from = Course, values_from = Grade)
>  # Conduct Friedman Test
>  friedman.test(as.matrix(test3_wide[, c("ENG 100 Grade", "ENG 101 Grade", "ENG
102 Grade")]))

      Friedman rank sum test
```

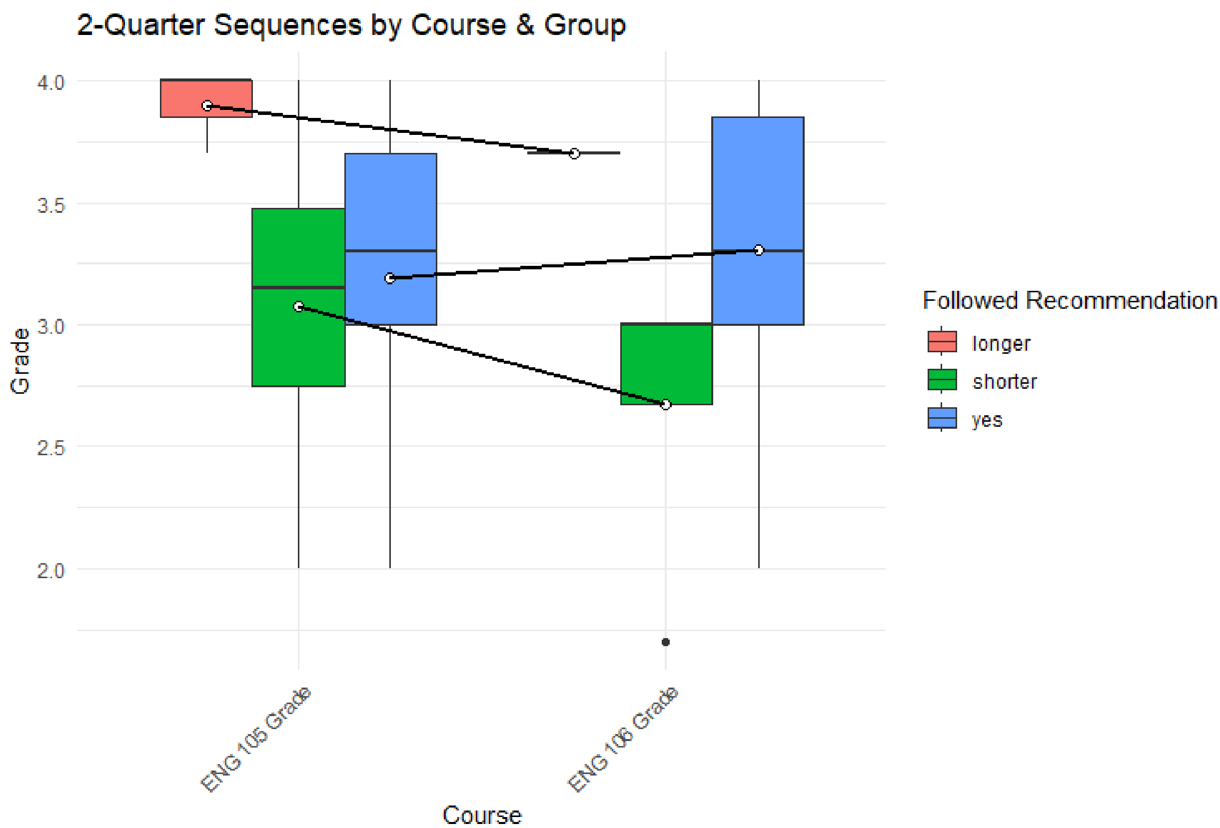
```
data:  as.matrix(test3_wide[, c("ENG 100 Grade", "ENG 101 Grade", "ENG 102
Grade")])
Friedman chi-squared = 0.025974, df = 2, p-value = 0.9871
```

```r
# Q12: Variables: Eng 105 and 106 Grade, Course, Group;
# Method: Repeated Measure Two-Way Anova

library(nlme)

# rename followed recommed
test2 <- test2 %>%
  rename(Followed_Recommend = `Followed Recommend`)

# Compute mean grades by Followed_Recommend
mean_grades <- test2 %>%
  group_by(Followed_Recommend) %>%
  summarize(Grade = mean(Grade, na.rm = TRUE)) %>%
  mutate(Followed_Recommend = as.factor(Followed_Recommend))

# Plot
ggplot(test2, aes(x = Followed_Recommend, y = Grade, fill = Followed_Recommend)) +
  geom_boxplot(width = 0.6) +
  stat_summary(
    fun = mean, geom = "point", shape = 21, color = "black", fill = "white", size = 2
  ) +
  # Line connecting means
  geom_line(data = mean_grades, aes(x = as.numeric(Followed_Recommend), y = Grade),
            color = "black", size = 1, group = 1) +
  labs(
    title = "2-Quarter Sequences by Group",
    x = "Followed Recommendation",
    y = "Grade",
    fill = "Followed Recommendation"
  ) +
  theme_minimal() +
  theme(axis.text.x = element_text(angle = 0, hjust = 0.5))

# Summary statistics
test2 %>%
  group_by(Followed_Recommend) %>%
  summarise(
    Min = min(Grade, na.rm = TRUE),
    Q1 = quantile(Grade, 0.25, na.rm = TRUE),
    Median = median(Grade, na.rm = TRUE),
    Q3 = quantile(Grade, 0.75, na.rm = TRUE),
    Max = max(Grade, na.rm = TRUE),
    Mean = mean(Grade, na.rm = TRUE),
    SD = sd(Grade, na.rm = TRUE)
  )
```



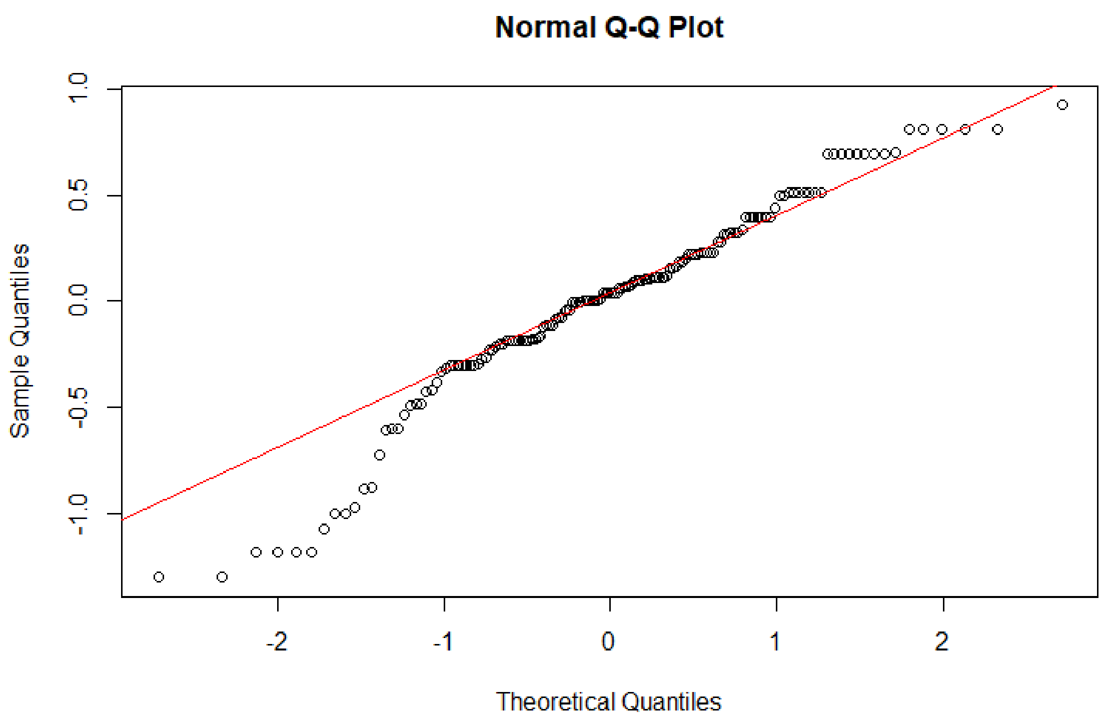

```
>  test2 %>%
+    group_by(Followed_Recommend) %>%
+    summarise(
+      Min = min(Grade, na.rm = TRUE),
+      Q1 = quantile(Grade, 0.25, na.rm = TRUE),
+      Median = median(Grade, na.rm = TRUE),
+      Q3 = quantile(Grade, 0.75, na.rm = TRUE),
+      Max = max(Grade, na.rm = TRUE),
+      Mean = mean(Grade, na.rm = TRUE),
+      SD = sd(Grade, na.rm = TRUE)
+    )
# A tibble: 3 × 8
  Followed_Recommend   Min    Q1 Median    Q3   Max  Mean    SD
  <fct>              <dbl> <dbl>  <dbl> <dbl> <dbl> <dbl> <dbl>
1 longer               3.7  3.7     3.7  3.92     4  3.8  0.155
2 shorter              1.7  2.75    3    3.08     4  2.88 0.723
3 yes                  2    3       3.3  3.7      4  3.25 0.630
```

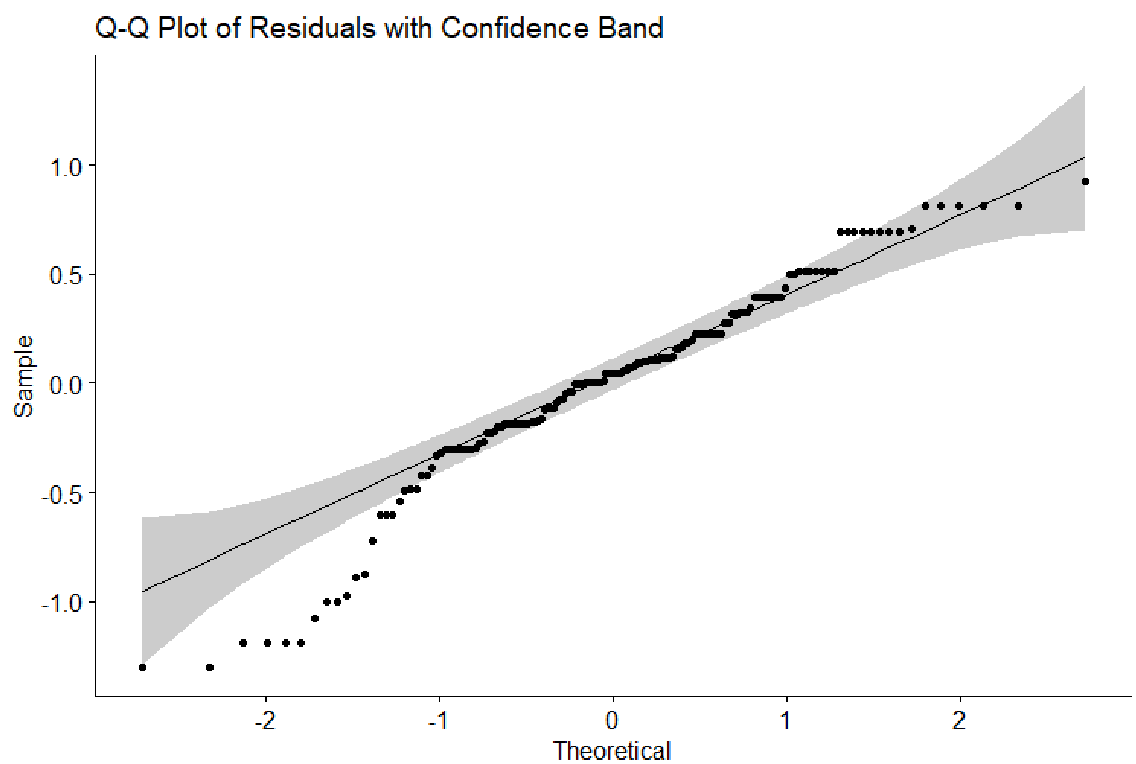
```r
# Course by Grade
ggplot(test2, aes(x = Course, y = Grade, fill = Followed_Recommend)) +
  geom_boxplot(position = position_dodge(width = 0.75)) +
  stat_summary(
    fun = mean, geom = "point", aes(group = Followed_Recommend),
    shape = 21, color = "black", fill = "white", size = 2,
    position = position_dodge(width = 0.75)
  ) +
  stat_summary(
    fun = mean, geom = "line", aes(group = Followed_Recommend),
    color = "black", size = 1,
    position = position_dodge(width = 0.75)
  ) +
  labs(
    title = "2-Quarter Sequences by Course & Group",
    x = "Course",
    y = "Grade",
    fill = "Followed Recommendation"
  ) +
  theme_minimal() +
  theme(axis.text.x = element_text(angle = 45, hjust = 1))


# Summary statistics
test2 %>%
  group_by(Course, Followed_Recommend) %>%
  summarise(
    Min = min(Grade, na.rm = TRUE),
    Q1 = quantile(Grade, 0.25, na.rm = TRUE),
    Median = median(Grade, na.rm = TRUE),
    Q3 = quantile(Grade, 0.75, na.rm = TRUE),
    Max = max(Grade, na.rm = TRUE),
    Mean = mean(Grade, na.rm = TRUE),
    SD = sd(Grade, na.rm = TRUE)
  )
```




```
> # Summary statistics
> test2 %>%
+   group_by(Course, Followed_Recommend) %>%
+   summarise(
+     Min = min(Grade, na.rm = TRUE),
+     Q1 = quantile(Grade, 0.25, na.rm = TRUE),
+     Median = median(Grade, na.rm = TRUE),
+     Q3 = quantile(Grade, 0.75, na.rm = TRUE),
+     Max = max(Grade, na.rm = TRUE),
+     Mean = mean(Grade, na.rm = TRUE),
+     SD = sd(Grade, na.rm = TRUE)
+   )
`summarise()` has grouped output by 'Course'. You can override using the
`.groups` argument.
# A tibble: 6 × 9
# Groups:   Course [2]
  Course         Followed_Recommend   Min    Q1 Median    Q3   Max  Mean    SD
  <chr>          <fct>              <dbl> <dbl>  <dbl> <dbl> <dbl> <dbl> <dbl>
1 ENG 105 Grade longer                3.7  3.85   4     4      4    3.9  0.173
2 ENG 105 Grade shorter               2    2.75   3.15  3.47   4    3.08 0.830
3 ENG 105 Grade yes                   2    3      3.3   3.7    4    3.19 0.642
4 ENG 106 Grade longer                3.7  3.7    3.7   3.7    3.7  3.7  0
5 ENG 106 Grade shorter               1.7  2.68   3     3      3    2.68 0.65
6 ENG 106 Grade yes                   2    3      3.3   3.85   4    3.30 0.622
```

```
# Rename the column to remove the space
colnames(test2)[colnames(test2) == "Followed Recommend"] <- "Followed_Recommend"

# set as factor
test2$Followed_Recommend <- as.factor(test2$Followed_Recommend)

# linear model
m.mod3 <- lme(Grade ~ Course * Followed_Recommend,  # interaction between Course and Group
              random = ~1 | Student_ID,
              data = test2)

# Residuals and Shapiro-Wilk Test
resid12 <- m.mod3$residuals
shapiro.test(resid12)

qqnorm(resid12)
qqline(resid12, col = "red")

library(ggpubr)
library(ggplot2)

# place as numerical; fix error
resid12 <- as.numeric(resid12)

ggqqplot(resid12, conf.int = TRUE) + ggtitle("Q-Q Plot of Residuals with Confidence Band")

# not normal; ART ANOVA
# fit the ART model

test2$Course <- as.factor(test2$Course)
Course <- test2$Course

art_model12 <- art(Grade ~ Course * Followed_Recommend + (1|Student_ID), data = test2)

# run ANOVA on the aligned ranks
anova(art_model12)

art.con(art_model12, "Course", adjust = "bonferroni")
art.con(art_model12, "Followed_Recommend", adjust = "bonferroni")
```

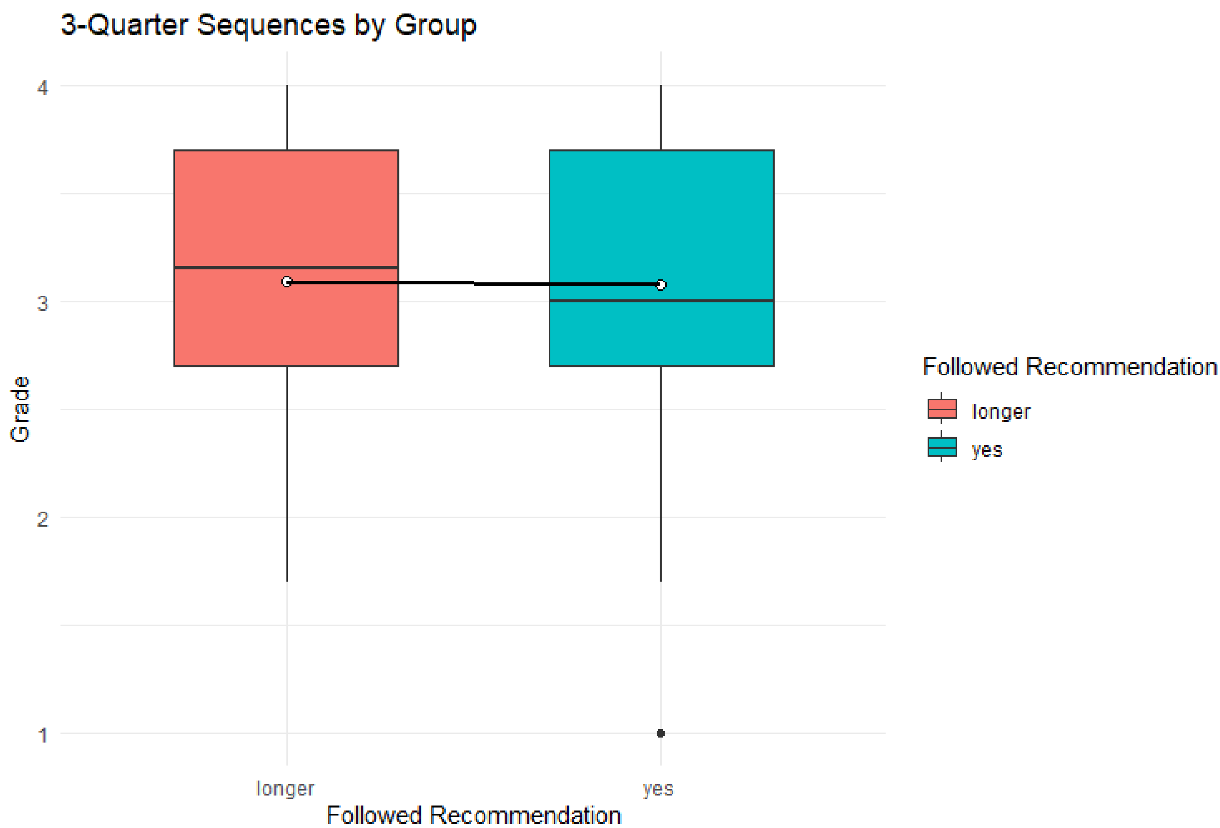

Normal Q-Q Plot
Sample Quantiles
Theoretical Quantiles
1.0
0.5
0.0
-0.5
-1.0
-2
-1
0
1
2


Q-Q Plot of Residuals with Confidence Band
Sample
Theoretical
1.0
0.5
0.0
-0.5
-1.0
-2
-1
0
1
2

```
>  # Residuals and Shapiro-Wilk Test
>  resid12 <- m.mod3$residuals
>  shapiro.test(resid12)

	Shapiro-Wilk normality test

data:  resid12
W = 0.95513, p-value = 7.998e-05

>  qqnorm(resid12)
>  qqline(resid12, col = "red")
>  library(ggpubr)
>  library(ggplot2)
>  # place as numerical; fix error
>  resid12 <- as.numeric(resid12)
>  ggqqplot(resid12, conf.int = TRUE) + ggtitle("Q-Q Plot of Residuals with
Confidence Band")
>  test2$Course <- as.factor(test2$Course)
>  Course <- test2$Course
>  art_model12 <- art(Grade ~ Course * Followed_Recommend + (1|Student_ID), data
= test2)
>  # run ANOVA on the aligned ranks
>  anova(art_model12)
Analysis of Variance of Aligned Rank Transformed Data

Table Type: Analysis of Deviance Table (Type III Wald F tests with Kenward-Roger
df)
Model: Mixed Effects (lmer)
Response: art(Grade)

                                    F Df Df.res  Pr(>F)
1 Course                    0.0016218  1     35 0.96811
2 Followed_Recommend        2.1653626  2     35 0.12983
3 Course:Followed_Recommend 0.9977645  2     35 0.37895
---
Signif. codes:   0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
>  art.con(art_model12, "Course", adjust = "bonferroni")
NOTE: Results may be misleading due to involvement in interactions
 contrast                      estimate   SE df t.ratio p.value
 ENG 105 Grade - ENG 106 Grade    0.193 4.78 35   0.040  0.9681

Results are averaged over the levels of: Followed_Recommend
Degrees-of-freedom method: kenward-roger
>  art.con(art_model12, "Followed_Recommend", adjust = "bonferroni")
NOTE: Results may be misleading due to involvement in interactions
 contrast         estimate   SE df t.ratio p.value
 longer - shorter     31.4 15.2 35   2.069  0.1379
 longer - yes         19.8 12.0 35   1.647  0.3256
 shorter - yes       -11.6 10.5 35  -1.100  0.8360

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

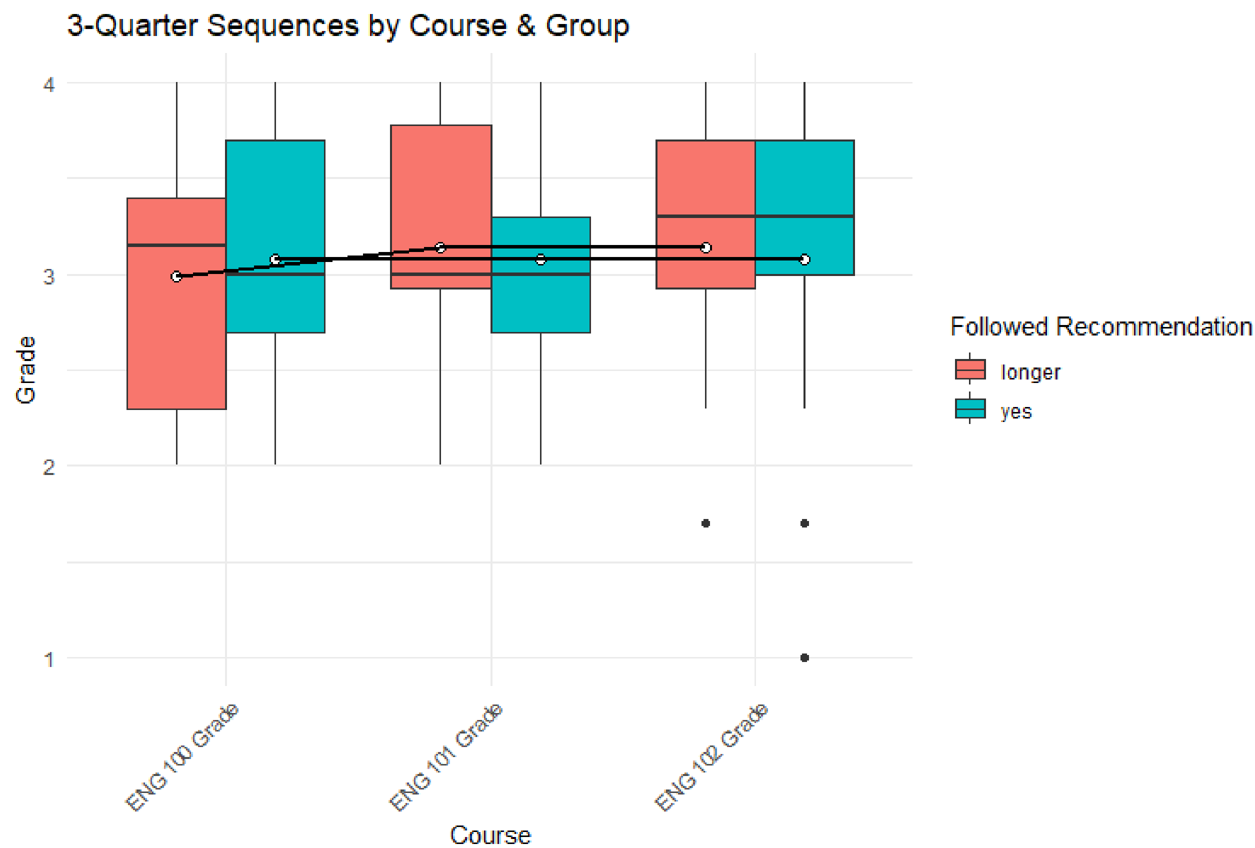
```
Degrees-of-freedom method: kenward-roger
P value adjustment: bonferroni method for 3 tests
```

```r
# Q13 Variables: Eng100, 101, 102 Grade, Course, Group;
# Method: Repeated Measure Two Way Anova

# Rename the column to remove the space
colnames(test3)[colnames(test3) == "Followed Recommend"] <- "Followed_Recommend"

# Compute mean grades by Followed_Recommend
mean_grades <- test3 %>%
  group_by(Followed_Recommend) %>%
  summarize(Grade = mean(Grade, na.rm = TRUE))

# Plot
ggplot(test3, aes(x = Followed_Recommend, y = Grade, fill = Followed_Recommend)) +
  geom_boxplot(width = 0.6) +
  stat_summary(
    fun = mean, geom = "point", shape = 21, color = "black", fill = "white", size = 2
  ) +
  geom_line(data = mean_grades,
            aes(x = Followed_Recommend, y = Grade, group = 1),
            color = "black", size = 1) +
  labs(
    title = "3-Quarter Sequences by Group",
    x = "Followed Recommendation",
    y = "Grade",
    fill = "Followed Recommendation"
  ) +
  theme_minimal() +
  theme(axis.text.x = element_text(angle = 0, hjust = 0.5))


# Summary statistics
test3 %>%
  group_by(Followed_Recommend) %>%
  summarise(
    Min = min(Grade, na.rm = TRUE),
    Q1 = quantile(Grade, 0.25, na.rm = TRUE),
    Median = median(Grade, na.rm = TRUE),
    Q3 = quantile(Grade, 0.75, na.rm = TRUE),
    Max = max(Grade, na.rm = TRUE),
    Mean = mean(Grade, na.rm = TRUE),
    SD = sd(Grade, na.rm = TRUE)
  )
```



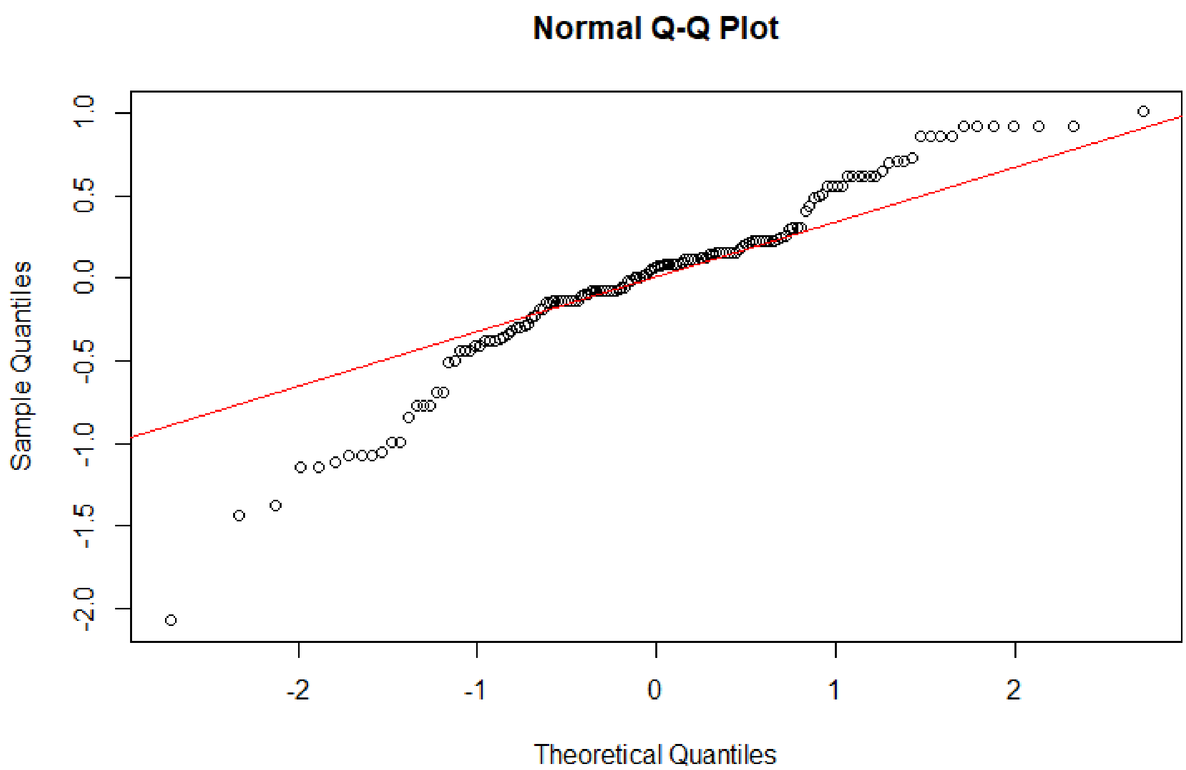

```
> # Summary statistics
> test3 %>%
+   group_by(Followed_Recommend) %>%
+   summarise(
+     Min = min(Grade, na.rm = TRUE),
+     Q1 = quantile(Grade, 0.25, na.rm = TRUE),
+     Median = median(Grade, na.rm = TRUE),
+     Q3 = quantile(Grade, 0.75, na.rm = TRUE),
+     Max = max(Grade, na.rm = TRUE),
+     Mean = mean(Grade, na.rm = TRUE),
+     SD = sd(Grade, na.rm = TRUE)
+   )
# A tibble: 2 × 8
  Followed_Recommend   Min    Q1 Median    Q3   Max  Mean    SD
  <fct>              <dbl> <dbl>  <dbl> <dbl> <dbl> <dbl> <dbl>
1 longer               1.7   2.7   3.15   3.7     4  3.09 0.668
2 yes                  1     2.7   3      3.7     4  3.08 0.721
```

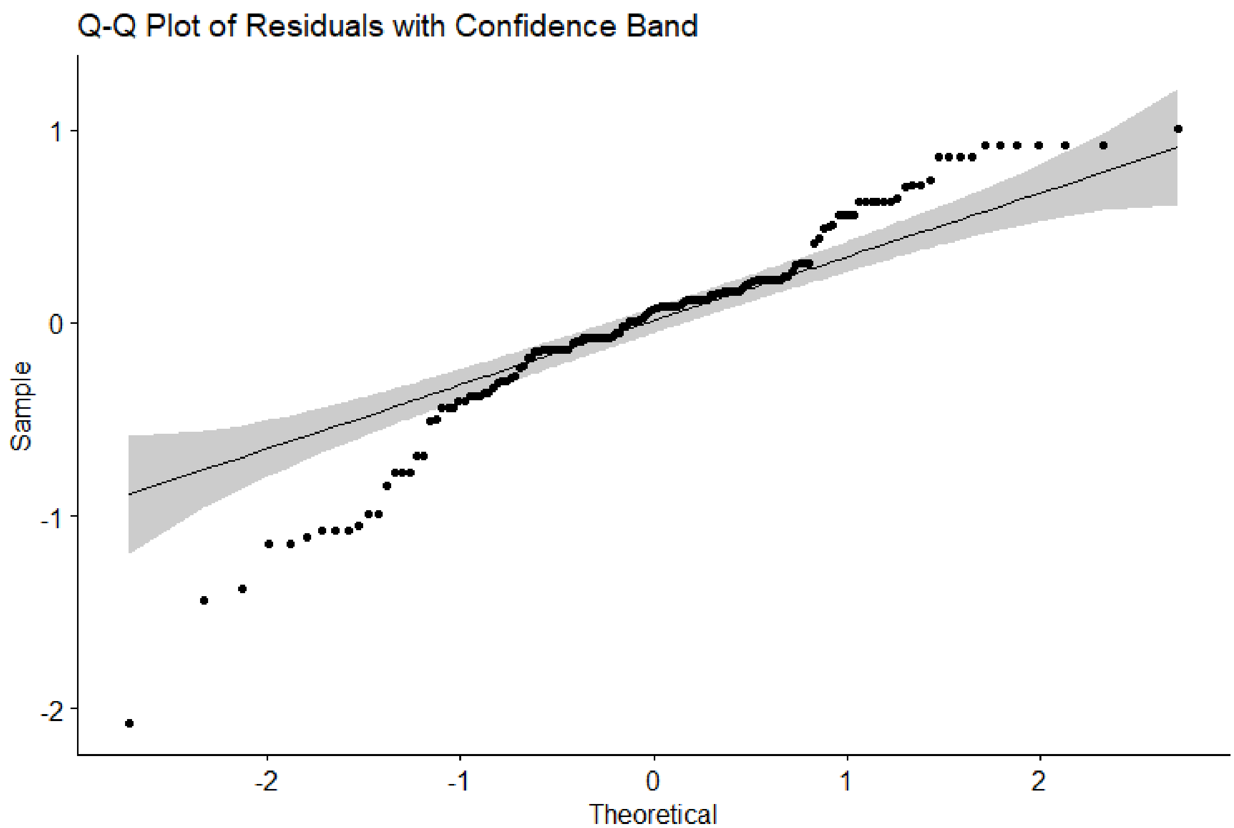
```r
# Course by Grade
ggplot(test3, aes(x = Course, y = Grade, fill = Followed_Recommend)) +
  geom_boxplot(position = position_dodge(width = 0.75)) +
  stat_summary(
    fun = mean, geom = "point", aes(group = Followed_Recommend),
    shape = 21, color = "black", fill = "white", size = 2,
    position = position_dodge(width = 0.75)
  ) +
  stat_summary(
    fun = mean, geom = "line", aes(group = Followed_Recommend),
    color = "black", size = 1,
    position = position_dodge(width = 0.75)
  ) +
  labs(
    title = "3-Quarter Sequences by Course & Group",
    x = "Course",
    y = "Grade",
    fill = "Followed Recommendation"
  ) +
  theme_minimal() +
  theme(axis.text.x = element_text(angle = 45, hjust = 1))


# Summary statistics
test3 %>%
  group_by(Course, Followed_Recommend) %>%
  summarise(
    Min = min(Grade, na.rm = TRUE),
    Q1 = quantile(Grade, 0.25, na.rm = TRUE),
    Median = median(Grade, na.rm = TRUE),
    Q3 = quantile(Grade, 0.75, na.rm = TRUE),
    Max = max(Grade, na.rm = TRUE),
    Mean = mean(Grade, na.rm = TRUE),
    SD = sd(Grade, na.rm = TRUE)
  )
```




```
> # Summary statistics
> test3 %>%
+   group_by(Course, Followed_Recommend) %>%
+   summarise(
+     Min = min(Grade, na.rm = TRUE),
+     Q1 = quantile(Grade, 0.25, na.rm = TRUE),
+     Median = median(Grade, na.rm = TRUE),
+     Q3 = quantile(Grade, 0.75, na.rm = TRUE),
+     Max = max(Grade, na.rm = TRUE),
+     Mean = mean(Grade, na.rm = TRUE),
+     SD = sd(Grade, na.rm = TRUE)
+   )
`summarise()` has grouped output by 'Course'. You can override using the
`.groups` argument.
# A tibble: 6 × 9
# Groups:   Course [3]
  Course        Followed_Recommend   Min    Q1 Median    Q3   Max  Mean    SD
  <chr>         <fct>              <dbl> <dbl>  <dbl> <dbl> <dbl> <dbl> <dbl>
1 ENG 100 Grade longer               2    2.3    3.15  3.4      4  2.99 0.689
2 ENG 100 Grade yes                  2    2.7    3     3.7      4  3.08 0.687
3 ENG 101 Grade longer               2    2.92   3     3.78     4  3.14 0.703
4 ENG 101 Grade yes                  2    2.7    3     3.3      4  3.08 0.596
5 ENG 102 Grade longer               1.7  2.92   3.3   3.7      4  3.14 0.658
6 ENG 102 Grade yes                  1    3      3.3   3.7      4  3.08 0.906
```

```r
# set as factor
test3$Followed_Recommend <- as.factor(test3$Followed_Recommend)
test3$Course <- as.factor(test3$Course)

rm.mod4 <- lme(Grade ~ Course * Followed_Recommend, # Fixed effects: main + interaction
                random = ~1 | Student_ID, # Random intercept for repeated measures
               data = test3)

resid13 <- rm.mod4$residuals

qqnorm(resid13)
qqline(resid13, col = "red")

library(ggpubr)
library(ggplot2)

# place as numerical; fix error
resid13 <- as.numeric(resid13)

ggqqplot(resid13, conf.int = TRUE) + ggtitle("Q-Q Plot of Residuals with Confidence Band")

shapiro.test(resid13)

# non parametric non normal
rm.mod4 <- art(Grade ~ Followed_Recommend*Course + (1|Student_ID), test3)

anova(rm.mod4)
```

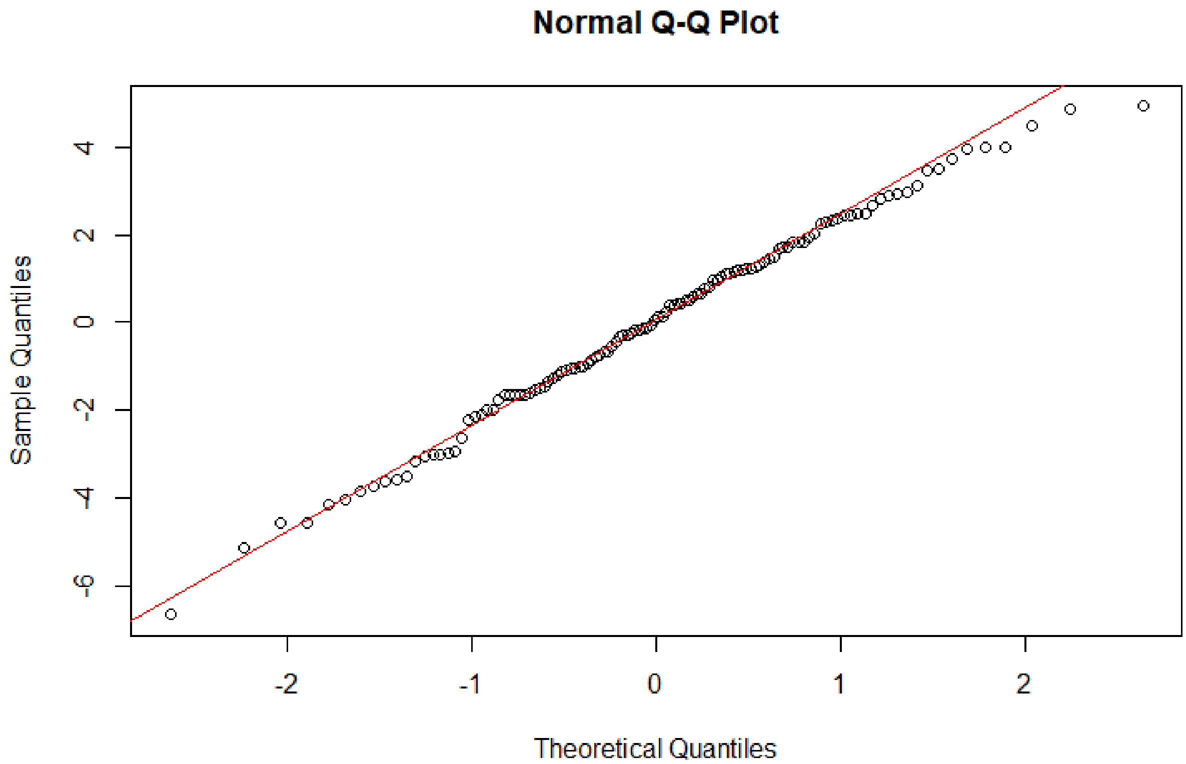




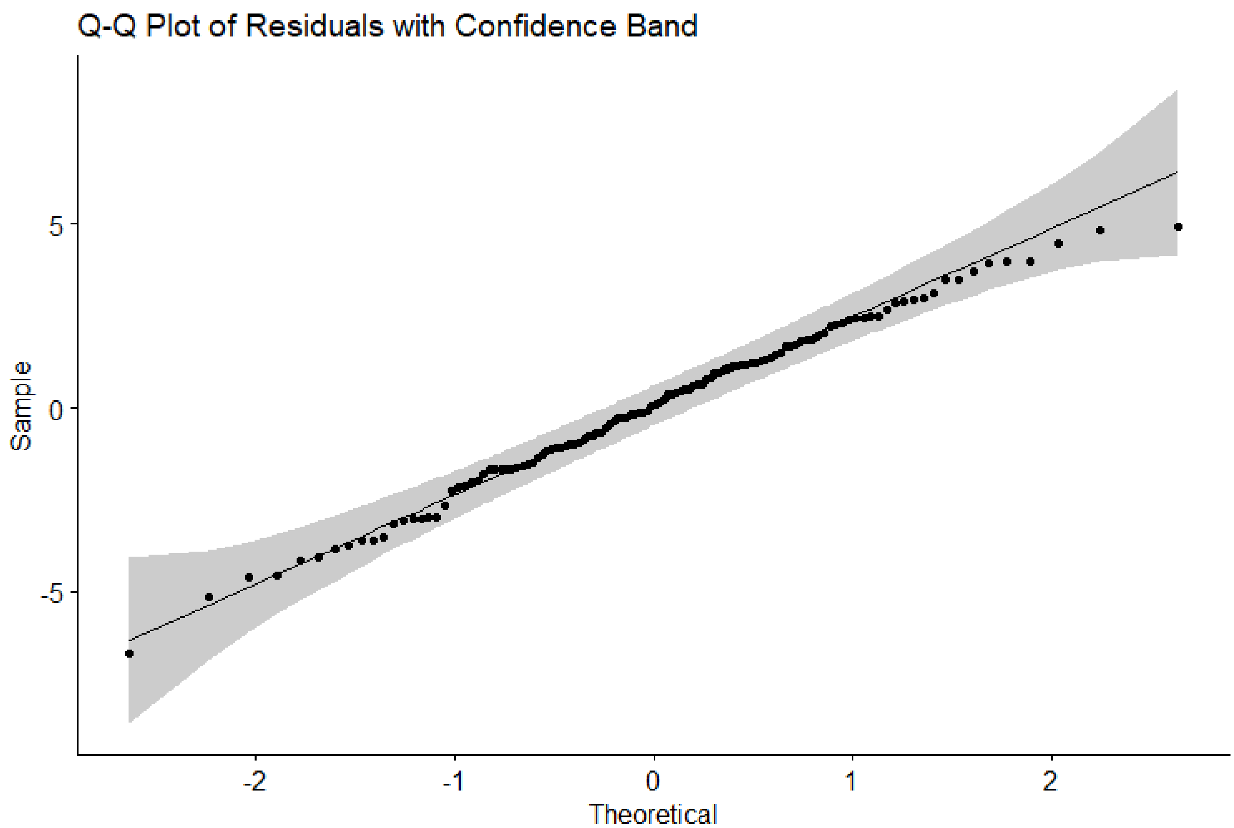

```
>  shapiro.test(resid13)

      Shapiro-Wilk normality test

data:  resid13
W = 0.95128, p-value = 4.189e-05

>  # non parametric non normal
>  rm.mod4 <- art(Grade ~ Followed_Recommend*Course + (1|Student_ID), test3)
>  anova(rm.mod4)
Analysis of Variance of Aligned Rank Transformed Data

Table Type: Analysis of Deviance Table (Type III Wald F tests with Kenward-Roger
df)
Model: Mixed Effects (lmer)
Response: art(Grade)

                                  F Df Df.res  Pr(>F)
1 Followed_Recommend        0.065521  1     23 0.80025
2 Course                    0.576165  2     46 0.56605
3 Followed_Recommend:Course 0.136262  2     46 0.87297
---
Signif. codes:   0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
```

```r
# Q14) Portfolio Total Score, Comp AVG Grade, Group
# Method: Multiple Linear Regression with Interactions (parametric)

# Fit a multiple linear regression model with interaction
model14 <- lm(`Portfolio Total Score` ~ `Comp AVG Grade` * `Followed Recommend`, data = edu_group)

# Summary of the model
summary(model14)

# residual test
resid14 <- model14$residuals

qqnorm(resid14)
qqline(resid14, col = "red")

library(ggpubr)
library(ggplot2)

# place as numerical; fix error
resid13 <- as.numeric(resid14)

ggqqplot(resid14, conf.int = TRUE) + ggtitle("Q-Q Plot of Residuals with Confidence Band")

# shapiro test
shapiro.test(model14$residuals)

# Best Subsets Regression
# library(leaps)

# regsubsets(`Portfolio Total Score` ~ `Comp AVG Grade` * `Followed Recommend`, data = edu_group)

# anova of model
anova(model14)
```

```
> # shapiro test
> shapiro.test(model14$residuals)

	Shapiro-Wilk normality test

data:  model14$residuals
W = 0.99257, p-value = 0.7743
```




```
> # anova of model
> anova(model14)
Analysis of Variance Table

Response: Portfolio Total Score
                                  Df Sum Sq Mean Sq F value Pr(>F)
`Comp AVG Grade`                   1  14.14 14.1385  2.5095 0.1159
`Followed Recommend`               2  60.52 30.2609  5.3710 0.0059 **
`Comp AVG Grade`:`Followed Recommend`  2  16.81  8.4062  1.4920 0.2293
Residuals                        114 642.29  5.6341
---
Signif. codes:  0 ‘***’ 0.001 ‘**’ 0.01 ‘*’ 0.05 ‘.’ 0.1 ‘ ’ 1
```




```
# Q15) Variables: Portfolio Total Score, Comp AVG Grade, Group
# Method: Correlation coefficients (parametric)

# Compute correlations by Group
edu_group %>%
  group_by(`Followed Recommend`) %>%
  summarize(
    correlation = cor(`Portfolio Total Score`, `Comp AVG Grade`, method = "pearson"),
    p_value = cor.test(`Portfolio Total Score`, `Comp AVG Grade`, method = "pearson")$p.value
  )
```

```
>  # Compute correlations by Group
>  edu_group %>%
+    group_by(`Followed Recommend`) %>%
+    summarize(
+      correlation = cor(`Portfolio Total Score`, `Comp AVG Grade`, method =
"pearson"),
+      p_value = cor.test(`Portfolio Total Score`, `Comp AVG Grade`, method =
"pearson")$p.value
+    )
# A tibble: 3 × 3
  `Followed Recommend` correlation p_value
  <fct>                      <dbl>   <dbl>
1 longer                     0.438   0.102
2 shorter                   -0.0359  0.793
3 yes                        0.105   0.473
```