Exploring ratings on a undergraduate “General Education” program across different raters and different rubrics

Zixuan Jin

Department of Statistics and Data Science

Carnegie Mellon University

zixuanji@andrew.cmu.edu

Abstract

Dietrich College at Carnegie Mellon University has experimented with Freshman Statistics rating work on different rubrics. We are interested in analyzing the relationship among ratings, rubrics, raters, and other factors related to this experiment. The experiment provided 91 project papers randomly sampled from Fall and Spring sections for Freshman Statistics and some personal characteristics of students writing these papers. We created bar plots and percentage tables to see the relationship among ratings, rubrics, and raters. We also calculated exact agreement rates between any raters on rating different rubrics. Then, we used AIC and BIC values to determine a final model predicting ratings using all other variables related to this experiment. Moreover, we used the same criterion to choose a final model predicting raters using all other variables. We found that the artifacts are not of equal quality on each rubric, and raters are not interpreting the evidence in the artifacts in the same way. These ratings might not correctly reflect the quality of this course. For future studies, the program might need to include more raters and change the grading system to make raters more consistent in grading project papers.

1 Introduction

Dietrich College at Carnegie Mellon University is implementing a new "General Education" program for undergraduates. This program specifies a set of courses and experiences that all undergraduates must take. The college hopes to rate student work performed in each of the "General Education" classes each year (Junker 2021).

The associate dean in charge of the experiment recently asked us to access the rating work in Freshman Statistics on different rubrics, using raters from across the college. In this report, we will investigate the relationship among ratings, rubrics, raters, and some other variables related to the experiment. Specifically, we will

1. Illustrate the relationship among ratings, rubrics, and raters
2. Identify the agreement or disagreement among raters for each rubric
3. Develop a regression model to predict scores from other factors in this experiment
4. Develop a regression model to predict raters from other factors in this experiment

2 Data

The data provides ratings on 91 project papers (referred to as "artifacts" in the following) randomly sampled from Fall and Spring 2019 sections of Freshman Statistics separately. Three raters from three different departments were asked to rate these artifacts on seven rubrics, as shown in Table 1. We offer the rating scale for all rubrics in Table 2. The raters did not know which class or which students produced the artifacts that they rated. Thirteen of the 91 artifacts were rated by all three raters; each of the remaining 78 artifacts were graded by only one rater.

Table 1: Rubrics for rating Freshman Statistics projects

Short Name	Full Name	Description
RsrchQ	Research Question	Given a scenario, the student generates, critiques or evaluates a relevant empirical research question.
CritDes	Critique Design	Given an empirical research question, the student critiques or evaluates to what extent a study design convincingly answer that question.
InitEDA	Initial EDA	Given a data set, the student appropriately describes the data and provides initial Exploratory Data Analysis.
SelMeth	Select Method(s)	Given a data set and a research question, the student selects appropriate method(s) to analyze the data.
InterpRes	Interpret Results	The student appropriately interprets the results of the selected method(s).
VisOrg	Visual Organization	The student communicates in an organized, coherent and effective fashion with visual elements (charts, graphs, tables, etc.).
TxtOrg	Text Organization	The student communicates in an organized, coherent and effective fashion with text elements (words, sentences, paragraphs, section and subsection titles, etc.).

Table 2: Rating scale used for all rubrics

Rating	Meaning
1	Student does not generate any relevant evidence.
2	Student generates evidence with significant flaws.
3	Student generates competent evidence; no flaws, or only minor ones.
4	Student generates outstanding evidence; comprehensive and sophisticated.

The data are given in the file ratings.csv in the project 02 folder for our course on Canvas. A detailed description of the variables in the dataset is shown in Table 3.

Table 3: Variable definitions for data in the file ratings.csv

Variable Name	Values	Description
(X)	1, 2, 3, ...	Row number in the data set
Rater	1, 2 or 3	Which of the three raters gave a rating
(Sample)	1, 2, 3, ...	Sample number
(Overlap)	1, 2, ..., 13	Unique identifier for artifact seen by all 3 raters
Semester	Fall or Spring	Which semester the artifact came from
Sex	M or F	Sex or gender of student who created the artifact
RsrchQ	1, 2, 3 or 4	Rating on Research Question
CritDes	1, 2, 3 or 4	Rating on Critique Design
InitEDA	1, 2, 3 or 4	Rating on Initial EDA
SelMeth	1, 2, 3 or 4	Rating on Select Method(s)
InterpRes	1, 2, 3 or 4	Rating on Interpret Results
VisOrg	1, 2, 3 or 4	Rating on Visual Organization
TxtOrg	1, 2, 3 or 4	Rating on Text Organization
Artifact	(text labels)	Unique identifier for each artifact
Repeated	0 or 1	1 = this is one of the 13 artifacts seen by all 3 raters

The X variable here is the same as the row number in our data, and we did not include it in our data analyses. Since the Sample and Overlap variables are just identifiers for those artifacts, we did not include these two variables in our data analyses.

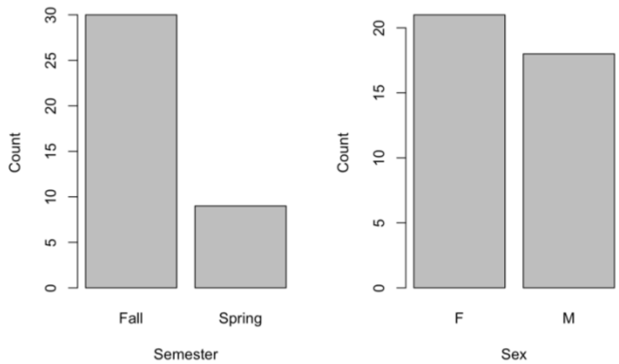
The file tall.csv provided in the project 02 folder contains the same data in the ratings.csv. But the tall.csv is organized so that each row has just one rating in the column labeled "Rating," and the rubric for that rating is listed in the column labeled "Rubric." We used files ratings.csv and tall.csv interchangeably for our research questions. The summary statistics of ratings for seven rubrics for those 13 artifacts seen by all raters are given in Table 4 below.

Table 4: Summary statistics for ratings for seven rubrics for those 13 artifacts

Variables	Min	1st Qu.	Median	Mean	3rd Qu.	Max	SD
RsrchQ	1	2	2	2.282	3	3	0.560
CritDes	1	1	2	1.718	2	3	0.724
InitEDA	1	2	2	2.385	3	3	0.544
SelMeth	1	2	2	2.051	2	3	0.510
InterpRes	1	2	3	2.513	3	4	0.601
VisOrg	1	2	2	2.282	3	3	0.605
TxtOrg	1	2	3	2.667	3	4	0.621

Looking at the summary statistics above, we can tell that the mean rating of these rubrics is around 2. For rating on critique design, the mean is less than the median, suggesting that the distribution is right-skewed. For rating on text organization and interpreting results, the mean is greater than the median, indicating that the distributions are slightly skewed to the left. We will explore the distributions in the following section. However, there are no rubrics with a mean substantially smaller or larger than the median. The summary statistics for the dummy variables in the data are given in Figure 1 below.

Figure 1: Histograms of dummy variables Semester and Sex for those 13 artifacts



In Figure 1, we see that most of the artifacts are from the Fall semester. Moreover, the number of female (sex 'F') students and the number of male (sex 'M') students who created those artifacts seen by all three raters are approximately the same. We also noticed some missing values for variables ratings and sex in the full dataset. In this case, we deleted these missing values when we used the complete dataset for our analyses.

3 Methods

3.1 The relationship among ratings, rubrics, and raters

We provided rating distributions and some numeric summaries to see if any rubric tends to get exceptionally high or low ratings. We included tables of counts, percentages, means, and standard deviations of the ratings for each rubric. We also offered rating distributions across three raters to see if there is any rater that tends to give exceptionally high or low ratings. We first analyzed the relationship among ratings, rubrics, and raters for those 13 artifacts seen by all three raters and then for all the 91 artifacts.

3.2 The agreement or disagreement among raters for each rubric

We fitted a random-intercept model for each rubric to see if raters agree on ratings for each rubric (See equations 1 and 2 below). Then, we calculated the intra-class correlation (ICC) as a measure of rater agreement, which will be the correlation between ratings on any two different artifacts by the same rater (See equation 3 below).

$$y_i = \alpha_{j[i]} + \epsilon_i, \epsilon_i \sim N(0, \sigma^2) \qquad (1)$$

$$\alpha_j = \beta_0 + \eta_j, \eta_j \sim N(0, \tau^2) \qquad (2)$$

$$ICC = \frac{\tau^2}{\tau^2 + \sigma^2} \qquad (3)$$

The j denotes each artifact, and i represents each observation in our data. In this case, j[i] means the artifact for rating i.

For our data, we fitted the following model (See equation 4 below) for each rubric to calculate ICC using estimates of σ^2 and τ^2 from the fitted models. The equations (1), (2), and (3) are equivalent versions of the random-intercept model. A high value of ICC means a high correlation between any two raters, which implies a high agreement between any two raters for a specific rubric.

$$Rating \sim 1 + (1 \mid Artifact) \qquad (4)$$

To compute exact agreement rates between raters, we also constructed tables of counts cross classifying the ratings that each pair of raters gives. A high value directly indicates a high agreement between two raters for a specific rubric. We calculated ICC and exact agreement rates for those 13 artifacts seen by all raters and for all artifacts separately.

3.3 Develop a regression model to predict scores from other factors in this experiment

3.3.1 Develop a regression model for each rubric

To explore how various factors in this experiment (Rater, Semester, Sex, Repeated, and Rubric) are related to ratings, we first added these factors as fixed effects to the equation (4) above for each rubric using all artifacts. We added one factor each time and used AIC and BIC to determine a final model for each rubric.

3.3.2 Develop a regression model examining interactions with rubrics

The models in 3.3.1 consider only one rubric at a time, which did not allow us to directly examine the interactions with all the rubrics. Then, we added interactions and new random effects to the "combined" model (See equations 5 and 6) using all the data as follows:

$$Rating \sim 1 + (0 + Rubric \mid Artifact) \qquad (5)$$

$$Rating_i = \beta_0 + \alpha_{k[i]j[i]} + \epsilon_i \qquad (6)$$

$$\begin{bmatrix} \alpha_{1j} \\ \alpha_{2j} \\ \alpha_{3j} \\ \alpha_{4j} \\ \alpha_{5j} \\ \alpha_{6j} \\ \alpha_{7j} \end{bmatrix} \sim N(\underline{0}, \Sigma) \qquad (7)$$

where

$$\underline{0} = (0, 0, 0, 0, 0, 0, 0)^T ,$$

$$\Sigma = \begin{bmatrix} {\tau_1}^2 & \tau_{12} & \cdots & \tau_{17} \\ \tau_{21} & {\tau_2}^2 & \cdots & \tau_{27} \\ \vdots & \vdots & \ddots & \vdots \\ \tau_{71} & \tau_{72} & \cdots & {\tau_7}^2 \end{bmatrix},$$

and

$$\epsilon_i \sim N(0, \sigma^2).$$

k[i] in equation 6 is the rubric k that $Rating_i$ is on, and j[i] in equation 6 is the artifact j that $Rating_i$ is on. The equations 5 and 6 above are equivalent versions of the combined model. The random effects show how much the scores/ratings vary across artifacts, from the prediction made by the fixed effects in section 3.3.1. At last, we chose our final model based on F-tests, AIC, and BIC values.

3.4 Develop a regression model to predict raters from other factors in this experiment

Before accessing the rating work, we want to see if the experimental design or data collection is reasonably free of bias. It is equivalent to show if these three raters grade students' work equally. In this case, we want to explore how various factors in this experiment (Semester, Rubric, Rating, and Sex) are related to raters. Then, we developed a multi-nominal regression model as we have three raters in this experiment. Then, we chose our final model based on AIC values, accompanied by a summary of the final model.

4 Results

4.1 The relationship among ratings, rubrics, and raters

4.1.1 The relationship among ratings, rubrics, and raters for 13 artifacts seen by all raters

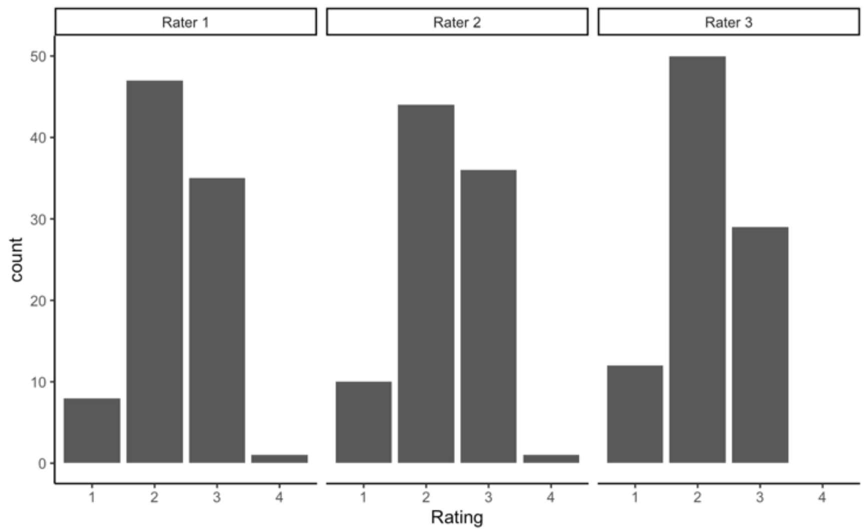
Looking at Table 5 below, we notice that most ratings are at 2 for most rubrics. Given the percentage tables for ratings on critique design, we see most counts are of rating 1. The ratings on critique design have exceptionally high percentages of ratings 1 compared to ratings on other rubrics. Given the percentages tables for ratings on text organization, we see most counts are of rating 3. The ratings on text organization have exceptionally high rates of rating 3 compared to ratings on other rubrics. In this case, we believe rating on critique design tends to get especially low ratings, and rating on text organization tends to get exceptionally high ratings. Bar plots of rating distributions for each rubric for 13 artifacts are in Appendix A (see page 30 below).

Table 5: Percentages of each rating score for each rubric for 13 artifacts seen by all raters

Rubrics	Rating = 1	Rating = 2	Rating = 3	Rating = 4
Rating on Research Question	5	62	33	0
Rating on Critique Design	44	41	15	0
Rating on Initial EDA	3	56	41	0
Rating on Select Method(s)	10	74	15	0
Rating on Interpret Results	3	46	49	3
Rating on Visual Organization	8	56	36	0
Rating on Text Organization	5	26	67	3

Given Figure 2 below, we can tell the rating patterns across the three raters are comparable. In this case, we believe that there is no rater giving exceptionally high or low ratings for those 13 artifacts.

Figure 2: Distributions of ratings across three raters for 13 artifacts seen by all three raters



4.1.2 The relationship among ratings, rubrics, and raters for all artifacts

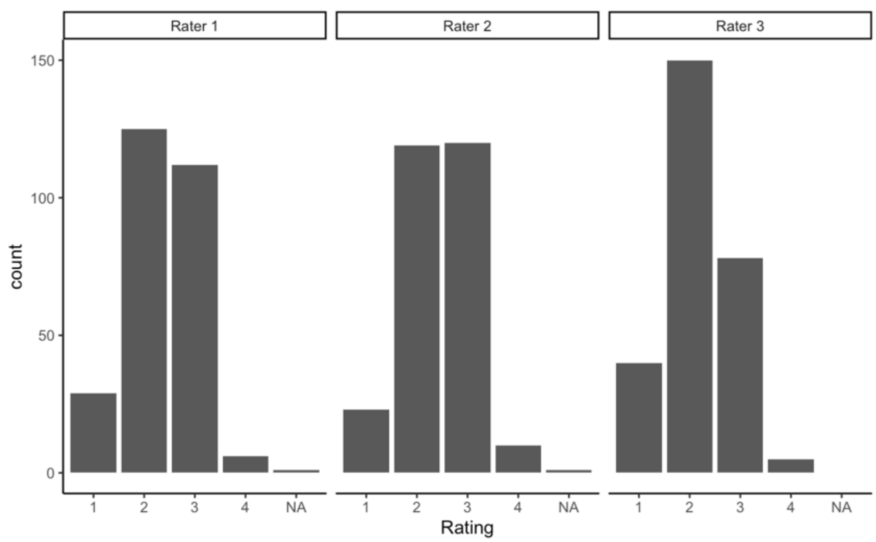
Looking at Tables 5 and 6, we notice that rating distributions for each rubric of those 13 artifacts seen by all three raters are indistinguishable from those ratings for each rubric of all artifacts. But the distribution of ratings on interpreting results for all artifacts is slightly different from that for those 13 artifacts. For all artifacts, the counts of rating 3 are higher for those 13 artifacts. Given Table 6, even though we notice some null values, we still find an exceptionally high percentage of rating 1 for ratings on critique design. However, we find high rates of rating 3 for interpreting results and text organization. Bar plots of rating distributions for each rubric for all artifacts are in Appendix A (see page 33 below).

Table 6: Percentages of each rating score for each rubric for all artifacts

Rubrics	Rating = 1	Rating = 2	Rating = 3	Rating = 4	NA
Rating on Research Question	5	56	38	1	0
Rating on Critique Design	40	33	24	2	1
Rating on Initial EDA	7	48	40	5	0
Rating on Select Method(s)	9	76	15	0	0
Rating on Interpret Results	5	42	52	1	0
Rating on Visual Organization	6	50	38	4	1
Rating on Text Organization	7	32	56	5	0

Given Figures 2 and 3, we find the distributions of ratings for all artifacts are comparable for those artifacts seen by all raters. However, we notice that Rater 3 tends to give a rating of 2 more frequently than other raters. In general, we can tell that the ratings for those 13 artifacts are representative of the whole set of 91 artifacts.

Figure 3: Distributions of ratings across three raters for all artifacts



4.2 The agreement or disagreement among raters for each rubric

Given Table 7, we notice that ICC values for all artifacts are comparable to ICC values for those 13 artifacts. We do not find an exceptionally high ICC value, indicating no correlation between raters for each rubric. Given low ICC values for ratings on research questions, we believe there is a low correlation between any two raters. Given the exact agreement rate of 0.85 between Rater 2 and Rater 3, we believe Rater 2 pretty much agrees with Rater 3 on rating on initial EDA (exploratory data analysis). Looking at the exact agreement rate of 0.92 on rating on selecting method(s), we find that Rater 1 and Rater 2 agree with each other for the most time. Given exact agreement rates of 0.77, we believe Raters 1 and 3 agree on rating on research questions and the visual organization. Moreover, we regard that Raters 2 and 3 agree on rating on the visual organization. In this case, we can tell raters agree on their ratings on rating on the visual organization.

Table 7: ICC values and exact agreement rates between any two raters for each rubric

Rubrics	ICC (all artifacts)	ICC (13 artifacts)	a12	a23	a13
Rating on Research Question	0.21	0.19	0.38	0.54	0.77
Rating on Critique Design	0.67	0.57	0.54	0.69	0.62
Rating on Initial EDA	0.69	0.49	0.69	0.85	0.54
Rating on Select Method(s)	0.47	0.52	0.92	0.69	0.62
Rating on Interpret Results	0.22	0.23	0.62	0.62	0.54
Rating on Visual Organization	0.66	0.59	0.54	0.77	0.77
Rating on Text Organization	0.19	0.14	0.69	0.54	0.62

4.3 Develop a regression model to predict scores from other factors in this experiment

4.3.1 Develop a regression model for each rubric on full dataset

Given AIC and BIC values, we do not need to add any fixed effect or interaction for predicting ratings on research questions, initial EDA, and text organization (Appendix C, see pages 67 to 83). The model is the same as the equation 4 above, with ratings treated as numeric values. By the same criterion, we should include Rater as a fixed effect in predicting ratings on critique design, interpreting results, and visual organization (Appendix C, see pages 67 to 83). Depending on AIC and BIC values, we need to include Sex and Semester in predicting ratings on selecting method(s) (Appendix C, see pages 74 to 77). Table 8 gives the full table of estimated coefficients for the final model for each rubric over all the data.

Table 8: Estimated coefficients of the final model for each rubric on all the data

	Research Question	Critique Design	Initial EDA	Select Method(s)	Interpret Results	Visual Organization	Text Organization
(Intercept)	2.36	1.69	2.44	2.11	2.70	2.38	2.59
Repeated	-	-	-	-	-	-	-
SemesterS19	-	-	-	-0.320	-	-	-
SemesterF19	-	-	-	-	-	-	-
SexF	-	-	-	-	-	-	-
SexM	-	-	-	0.111	-	-	-
Rater 1	-	-	-	-	-	-	-
Rater 2	-	0.427	-	-	-0.118	0.271	-
Rater 3	-	0.205	-	-	-0.565	-0.094	-
$\hat{\sigma}^2$	0.278	0.247	0.166	0.116	0.253	0.147	0.396
$\hat{\tau}^2$	0.073	0.435	0.365	0.090	0.062	0.291	0.094

Given Table 8, we realized that:

- The average ratings are 2.36, 2.44, and 2.69 for research question, initial EDA, and text organization, respectively.
- For ratings on critique design, Rater 2 tends to give 0.427 points more than Rater 1 does; Rater 3 tends to give 0.205 points more than Rater 1 does on average.
- For ratings on selecting method(s), ratings in Spring 2019 are 0.32 points lower than ratings in Fall 2019; Ratings for male students tend to have 0.111 points higher than that for female students.
- For ratings on interpreting results, Rater2 tends to give 0.118 points less than Rater 1 does; Rater 3 tends to give 0.565 points less than Rater 1 does on average.
- For ratings on visual organization, Rater 2 tends to give 0.271 points more than rater 1 does; Rater 3 tends to give 0.094 points less than Rater 1 does on average.

$\hat{\tau}^2$ shows the estimated deviation from each rubric's overall mean of ratings. The departures for ratings on critique design, initial EDA, and visual organization are higher than for ratings on other rubrics.

4.3.2 Develop a regression model examining interactions with rubric

Table 9: Estimated coefficients and standard errors of the fixed effects for model (6)

MODEL INFO:
Observations: 810
Dependent Variable: as.numeric(Rating)
Type: Mixed effects linear regression

MODEL FIT:
AIC = 1521.60, *BIC* = 1761.15

FIXED EFFECTS:

	Est.	S.E.	t val.	d.f.	p
(Intercept)	1.76	0.12	14.77	176.70	0.00
Rater2	0.37	0.14	2.70	301.50	0.01
Rater3	0.21	0.14	1.59	301.41	0.11
SemesterS19	-0.18	0.09	-2.01	86.93	0.05
RubricInitEDA	0.75	0.14	5.39	165.65	0.00
RubricInterpRes	1.01	0.14	7.44	166.13	0.00
RubricRsrchQ	0.75	0.13	5.95	164.89	0.00
RubricSelMeth	0.43	0.13	3.23	168.11	0.00
RubricTxtOrg	1.05	0.14	7.66	165.26	0.00
RubricVisOrg	0.68	0.14	4.84	167.41	0.00
Rater2:RubricInitEDA	-0.31	0.18	-1.76	279.65	0.08
Rater3:RubricInitEDA	-0.30	0.18	-1.68	278.65	0.09
Rater2:RubricInterpRes	-0.54	0.17	-3.10	277.74	0.00
Rater3:RubricInterpRes	-0.75	0.17	-4.33	276.80	0.00
Rater2:RubricRsrchQ	-0.50	0.16	-3.05	264.71	0.00
Rater3:RubricRsrchQ	-0.37	0.16	-2.25	263.47	0.03
Rater2:RubricSelMeth	-0.40	0.17	-2.36	274.62	0.02
Rater3:RubricSelMeth	-0.41	0.17	-2.46	273.42	0.01
Rater2:RubricTxtOrg	-0.58	0.17	-3.34	277.79	0.00
Rater3:RubricTxtOrg	-0.49	0.17	-2.78	276.79	0.01
Rater2:RubricVisOrg	-0.14	0.18	-0.81	282.66	0.42
Rater3:RubricVisOrg	-0.33	0.18	-1.87	281.73	0.06

Then, we added fixed effects, interactions, and new random effects to the combined model (see equation 5 above in 3.3.2), including interaction with the rubrics. Without considering new random effects, we used AIC and BIC values to choose our final model as follows (Appendix C, see pages 86 to 91):

$$Rating \sim Rater + Semester + Rubric + Rater \times Rubric + (0 + Rubric \mid Artifact) \quad (6)$$

Table 9 gives the full table of estimated coefficients and standard errors for the fixed effects for model (6), and Table 10 shows the estimated standard deviations for the random effects for the model (6).

Table 10: Estimated standard deviations of the random effects for model (6)

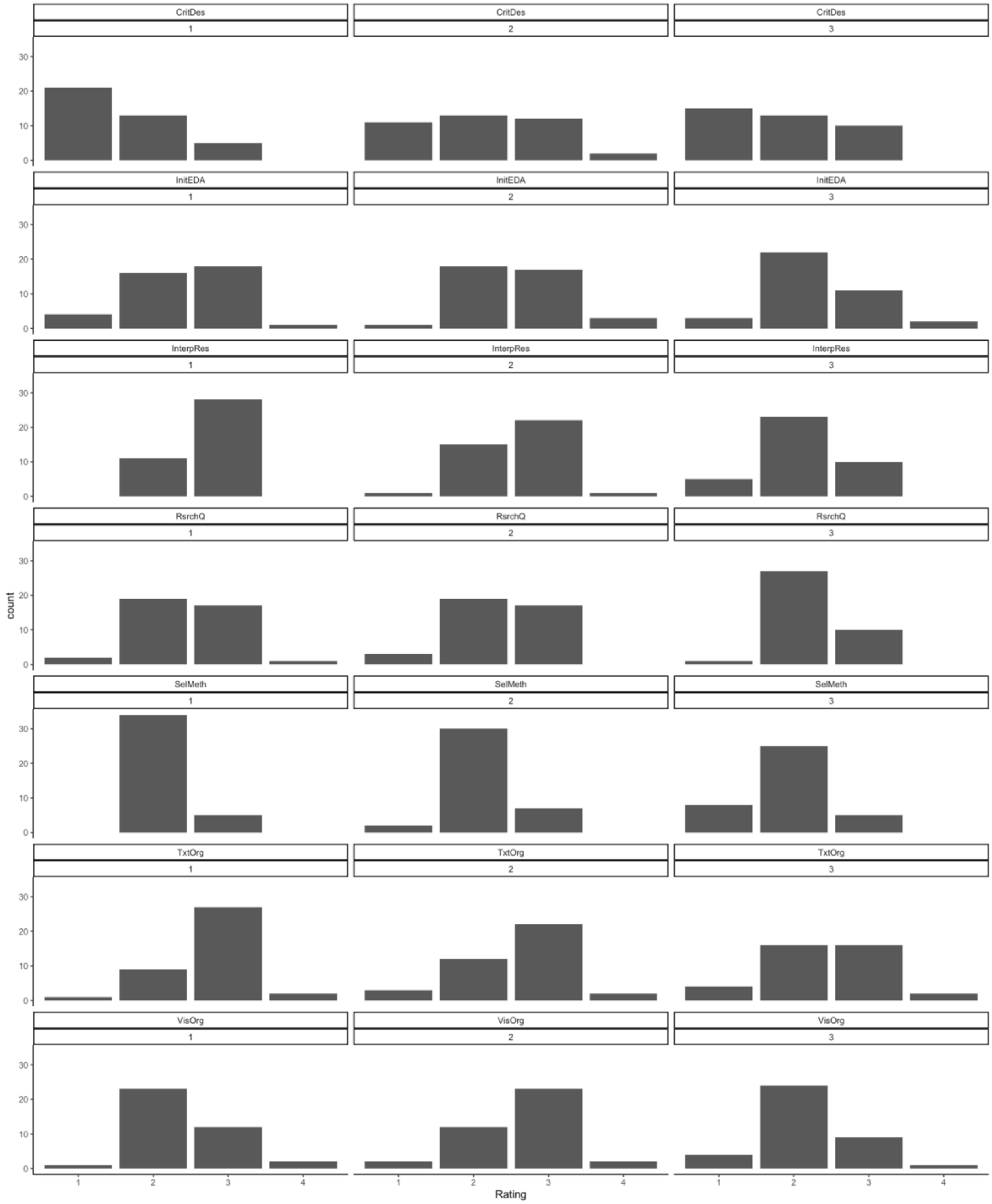
RANDOM EFFECTS:

Group	Parameter	Std. Dev.
Artifact	RubricCritDes	0.71
Artifact	RubricInitEDA	0.60
Artifact	RubricInterpRes	0.39
Artifact	RubricRsrchQ	0.42
Artifact	RubricSelMeth	0.26
Artifact	RubricTxtOrg	0.51
Artifact	RubricVisOrg	0.50
Residual		0.43

Grouping variables:

Group	# groups	ICC
Artifact	90	0.73

Figure 6: Facet graphs of ratings on each rubric given by each rater



Given the facet graphs in Figure 6, the distributions of ratings are different across these three raters for each rubric, especially for ratings on visual organization. We believe that these 3 raters have different ways of scoring the 7 rubrics. In this case, the interaction between raters and rubrics in model 6 makes sense. Table 10 shows that the coefficients are all statistically significant at 5% significance level except for Rater3 and the interaction term between Rater2 and ratings on visual organization. Given Table 10, we know:

- Rater 1's ratings on interpreting results or text organization are 1 point greater than the ratings on critique design.
- For rating on critique design, Rater 2 generally gives 0.37 points higher than Rater 1 does on average and Rater 3 generally gives 0.21 points higher than Rater 1 does. Rater 3 generally gives 0.17 points lower than Rater 2 does on average.
- For rating on initial EDA, Rater 2 gives 0.06 points higher than Rater 1 does on average, and Rater 3 gives 0.09 points less than Rater 1 does on average. Rater 3 gives 0.15 points higher than Rater 2 does on average.
- For rating on interpreting results, Rater 2 gives 0.17 points less than Rater 1 does, and Rater 3 gives 0.54 points lower than Rater 1 does on average. Rater 2 generally gives 0.37 points higher than Rater 3 does on average.
- For rating on the research question, Rater 2 gives 0.13 points less than Rater 2, and Rater 3 gives 0.16 points less than Rater 1 does on average. Rater 2 generally gives 0.03 points higher than Rater 2 does on average.
- For rating on selecting methods, Rater 2 gives 0.03 points less than Rater 1, and Rater 3 gives 0.2 points less than Rater 1 does on average. Rater 3 gives 0.17 points less than Rater 2 does on average.

- For rating on the text organization, Rater 2 gives 0.21 points less than Rater 1, and Rater 3 gives 0.28 points less than Rater 1 does on average. Rater 3 gives 0.07 point less than Rater 2 does on average.
- For rating on visual organization, Rater 2 gives 0.24 points less than Rater 1, and Rater 3 gives 0.12 points less than Rater 1 does on average. Rater 3 gives 0.12 point higher than Rater 1 does on average.
- Moreover, the overall ratings in Spring 2019 are 0.18 points less than ratings in Fall on average, holding others constant.

4.3.3 Considering new random effects in predicting ratings on the combined model

Considering adding new random effects to the model (6) above, we used AIC and BIC values to choose our final model as follows (Appendix C, see pages 93 to 101):

$$Rating \sim Rater + Semester + Rubric + Rater \times Rubric + (0 + Rubric \mid Artifact) + (0 + Rater \mid Artifact) \qquad (7)$$

Model (7) allows each rater's rating on each artifact to differ from what we would expect (from the fixed effects alone) by a small random effect that depends on the artifact. Table 11 below gives the full table of estimated coefficients and standard errors for the fixed effects for model (7). Table 12 shows the estimated standard deviations for the random effects for the model (7). Given Table 9 and random effects (Appendix C, see pages 93 to 102 below):

- Raters 1, 2, and 3 give the average ratings of 1.76, 2.13, and 1.96, respectively.
- The ratings on artifact 02 by Rater 1 are of the highest random errors, which leads to the overall average score of 1.93 on artifact 02. The ratings on artifact 01 by Rater 1 are of the lowest random errors, which leads to the overall average score of 1.55 on artifact 02.

- The ratings on artifact 04 by Rater 2 are of the highest random errors, which leads to the overall average score of 2.55 on artifact 02. The ratings on artifact 02 by Rater 2 are of the lowest random errors, which leads to the overall average score of 1.1 on artifact 02.
- The ratings on artifact 07 by Rater 3 are of the highest random errors, which leads to the overall average score of 2.44 on artifact 07. The ratings on artifact 02 by Rater 3 are of the lowest random errors, which leads to the overall average score of 1.34 on artifact 02.

As we mentioned above, the three raters have different scoring of the seven rubrics. The random effect of the rubric shows that there are different average scores on each rubric. Still, the rubric averages vary slightly from one artifact to the next, by a small random effect that depends on artifacts. Given Table 11 and random effects (Appendix C, see pages 93 to 102 below), we notice that:

- The average ratings are 1.76, 2.5, 2.75, 2.49, 2.17, 2.78, and 2.41 on critique design, initial EDA, interpret results, research questions, selecting method(s), text organization, and visual organization, respectively.
- The ratings for critique design on artifact 62 are of the highest random errors, which leads to the overall average score of 3.15 on artifact 62. The ratings for critique design on artifact 74 are of the highest random errors, which leads to the overall average score of 0.71 on artifact 74.
- The ratings for initial EDA on artifact 16 are of the highest random errors, which leads to the overall average score of 3.65 on artifact 16. The ratings for initial EDA on artifact 87 are of the highest random errors, which leads to the overall average score of 1.47 on artifact 87.

- The ratings for interpreting results on artifact 21 are of the highest random errors, which leads to the overall average score of 3.65 on artifact 21. The ratings for interpreting results on artifact 13 are of the highest random errors, which leads to the overall average score of 2.32 on artifact 13.
- The ratings for the research question on artifact 88 are of the highest random errors, which leads to the overall average score of 3.37 on artifact 88. The ratings for the research question on artifact 102 are of the highest random errors, which leads to the overall average score of 1.75 on artifact 102.
- The ratings for selecting method(s) on artifact 04 are of the highest random errors, which leads to the overall average score of 2.54 on artifact 04. The ratings for selecting method(s) on artifact 65 are of the highest random errors, which leads to the overall average score of 1.95 on artifact 65.
- The ratings for the text organization on artifact 21 are of the highest random errors, which leads to the overall average score of 3.7 on artifact 21. The scores for the text organization on artifact 102 are of the highest random errors, leading to an overall average score of 1.66 on artifact 102.
- The ratings for the visual organization on artifact 88 are of the highest random errors, which leads to the overall average score of 3.45 on artifact 88. The ratings for the visual organization on artifact 08 are of the highest random errors, which leads to the overall average score of 1.5 on artifact 08.

In all of this, the artifacts are not of equal quality on each rubric, so we should expect the average scores to vary from one artifact to the next. The interaction between rater and rubric suggests that the raters are not all interpreting the rubrics in the same way. The interaction between raters and artifacts indicates that the raters are not interpreting the evidence in the artifacts in the same way. These interactions suggest that the raters should be trained more to make the raters' ratings more like each other.

Table 11: Estimated coefficients and standard errors of the fixed effects for model (7)

MODEL INFO:
Observations: 810
Dependent Variable: as.numeric(Rating)
Type: Mixed effects linear regression

MODEL FIT:
AIC = 1484.57, *BIC* = 1752.30

FIXED EFFECTS:

	Est.	S.E.	t val.	d.f.	p
(Intercept)	1.76	0.12	15.04	143.15	0.00
Rater2	0.37	0.14	2.54	188.83	0.01
Rater3	0.20	0.13	1.45	194.57	0.15
SemesterS19	-0.16	0.09	-1.83	73.37	0.07
RubricInitEDA	0.74	0.13	5.64	163.63	0.00
RubricInterpRes	0.99	0.13	7.69	164.46	0.00
RubricRsrchQ	0.73	0.12	6.08	163.88	0.00
RubricSelMeth	0.41	0.13	3.26	167.51	0.00
RubricTxtOrg	1.02	0.13	7.74	163.30	0.00
RubricVisOrg	0.65	0.13	4.85	165.11	0.00
Rater2:RubricInitEDA	-0.30	0.16	-1.89	261.29	0.06
Rater3:RubricInitEDA	-0.29	0.16	-1.85	260.40	0.07
Rater2:RubricInterpRes	-0.51	0.16	-3.28	261.44	0.00
Rater3:RubricInterpRes	-0.71	0.16	-4.56	260.42	0.00
Rater2:RubricRsrchQ	-0.49	0.15	-3.25	255.05	0.00
Rater3:RubricRsrchQ	-0.32	0.15	-2.14	253.79	0.03
Rater2:RubricSelMeth	-0.39	0.15	-2.52	266.10	0.01
Rater3:RubricSelMeth	-0.39	0.15	-2.53	264.41	0.01
Rater2:RubricTxtOrg	-0.55	0.16	-3.46	262.42	0.00
Rater3:RubricTxtOrg	-0.44	0.16	-2.79	261.51	0.01
Rater2:RubricVisOrg	-0.10	0.16	-0.65	264.21	0.52
Rater3:RubricVisOrg	-0.28	0.16	-1.70	263.28	0.09

Table 12: Estimated standard deviations of the random effects for model (7)

RANDOM EFFECTS:

Group	Parameter	Std. Dev.
Artifact	RubricCritDes	0.70
Artifact	RubricInitEDA	0.56
Artifact	RubricInterpRes	0.32
Artifact	RubricRsrchQ	0.42
Artifact	RubricSelMeth	0.20
Artifact	RubricTxtOrg	0.50
Artifact	RubricVisOrg	0.48
Artifact.1	Rater1	0.11
Artifact.1	Rater2	0.33
Artifact.1	Rater3	0.31
Residual		0.37

Grouping variables:

Group	# groups	ICC
Artifact	90	0.77

4.4 Develop a regression model to predict raters from other factors in this experiment

To further test whether raters give ratings equally, we started with fitting the following full model as follows without considering any random effects:

$$Rater \sim Semester + Rubric + Rating + Sex \qquad (8)$$

We treated Rating as a numeric variable in this model (8) above. Then we use AIC to choose our final mode (9) shown below (Appendix D, see pages 102 to 104 below).

$$Rater \sim Rating + Sex \qquad (9)$$

Table 13 gives the full table of estimated coefficients and standard errors for model (9).

Table 13: Estimated coefficients and standard errors for model (9)

	Dependent variable:	
	2	3
	(1)	(2)
as.numeric(Rating)	1.184	0.669***
	(0.124)	(0.125)
SexF	16.119	0.0001
	(103.752)	(25.218)
SexM	19.864	0.0002
	(103.752)	(25.218)
Constant	0.038	14,884.530
	(103.752)	(25.221)
Akaike Inf. Crit.	1,771.164	1,771.164
Note:	*p<0.1; **p<0.05; ***p<0.01	

Keeping all other variables constant, if the rating increases one unit, the rating is 0.669 times more likely to come from Rater 3 than Rater 1. The coefficient is statistically significant at 1% significance level. Keeping all other variables constant, if the rating increases one unit, the rating is 1.184 times more likely to come from Rater 2 than Rater 1. However, the coefficient is not statistically significant. We showed that raters do not interpret the rubrics equally, and one rater tends to give higher ratings than others.

5 Discussion

In this study, we focused on examining the rating work on students' project papers in a Freshman Statistics course from three raters and across seven rubrics. We analyzed the relationship among various variables in the dataset to see if the experiment design is reasonably free of bias. We found that artifacts are not of equal quality on these rubrics. In addition, raters do not interpret the rubrics equally: one gives higher/lower ratings.

Specifically, we found that rating on critique design tends to get especially low ratings and ratings on text organization tend to get exceptionally high ratings. Moreover, we found that raters do not agree on ratings seven rubrics. Given section 4.3, we notice that the artifacts are not of equal quality on each rubric, and raters are not interpreting the evidence in the artifacts in the same way. Given section 4.4, it seems that Rater 3 is more likely to give high ratings. In all of this, raters are not consistent in ratings all the artifacts. Hence the ratings might not correctly reflect the results of this program or this experiment, which might lead to an incorrect decision of implementing the specific undergraduate course. Maybe raters should have rubrics on grading each topic in a paper to be more consistent in their ratings.

One of the limitations of this experiment is that we have only three raters. Three raters' ratings might not be representative, and they cannot truly reflect how students perform on those seven rubrics. In this case, it is better to include more raters in this experiment. Another limitation could be the range of the rating. Currently, the ratings take 1, 2, 3, and 4 levels. It might be hard for raters to give ratings if they find an artifact is not worth score 3 but is better than score 2. Changing the grading system might make this experiment more efficient.

We could include more raters for future studies to eliminate the limitations mentioned above. In addition, the program manager needs to generate detailed guidelines for grading these rubrics above. For example, the program needs to list out where raters should deduct their points to make raters more consistent in grading and reflect the accurate ratings of the program.

References

Junker, B. W. (2021). *Project 02 assignment sheet and data for 36-617: Applied Regression Analysis*. Department of Statistics and Data Science, Carnegie Mellon University, Pittsburgh PA. Accessed Nov 29, 2021 from *https://canvas.cmu.edu/courses/25337/files/folder/Project02*

R Core Team (2021). *R: A language and environment for statistical computing*. R Foundation for Statistical Computing, Vienna, Austria. URL *https://www.R-project.org/*.

RStudio Team (2021). RStudio: Integrated Development Environment for R. *RStudio, PBC, Boston, MA URL* http://www.rstudio.com/.

Sheather, S.J. (2009), *A Modern Approach to Regression with R*. New York: Springer Science + Business Media LLC.

Technical Appendices

Content **Page**

Read the data

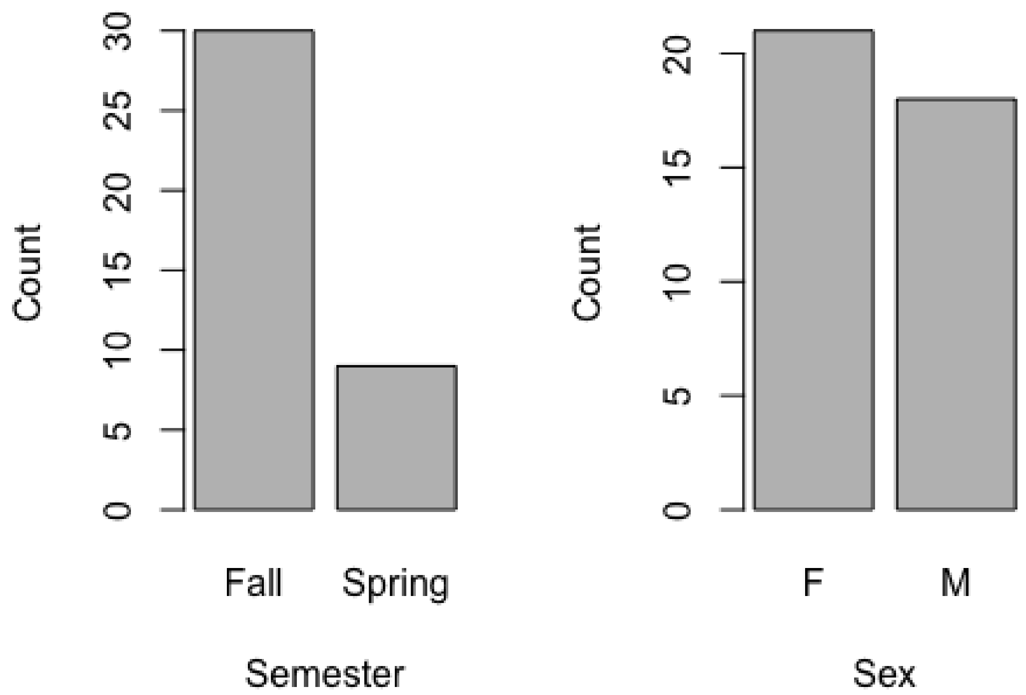
```
tall <- read.csv("/Users/rosahhh/Desktop/Fall 2021/36617 Applied Linear Model
/project02/tall.csv",
                 header = TRUE)
ratings <- read.csv("/Users/rosahhh/Desktop/Fall 2021/36617 Applied Linear Mo
del/project02/ratings.csv",
                    header = TRUE)
```

Appendix A

Make a subset of the data for just the 13 artifacts seen by all three raters and some EDA

```
ratings_sub <- ratings %>%
  filter(ratings$Repeated == 1)

summary(ratings_sub)

##        X              Rater       Sample          Overlap      Semester

##  Min.   :  1.00   Min.   :1   Min.   :  1.00   Min.   : 1   Length:39

##  1st Qu.: 23.50   1st Qu.:1   1st Qu.: 24.50   1st Qu.: 4   Class :charact
er
##  Median : 51.00   Median :2   Median : 52.00   Median : 7   Mode  :charact
er
##  Mean   : 53.46   Mean   :2   Mean   : 54.28   Mean   : 7

##  3rd Qu.: 81.50   3rd Qu.:3   3rd Qu.: 82.50   3rd Qu.:10

##  Max.   :109.00   Max.   :3   Max.   :110.00   Max.   :13

##      Sex                RsrchQ         CritDes         InitEDA
##  Length:39          Min.   :1.000   Min.   :1.000   Min.   :1.000
##  Class :character   1st Qu.:2.000   1st Qu.:1.000   1st Qu.:2.000
##  Mode  :character   Median :2.000   Median :2.000   Median :2.000
##                     Mean   :2.282   Mean   :1.718   Mean   :2.385
##                     3rd Qu.:3.000   3rd Qu.:2.000   3rd Qu.:3.000
##                     Max.   :3.000   Max.   :3.000   Max.   :3.000
##     SelMeth         InterpRes         VisOrg          TxtOrg
##  Min.   :1.000   Min.   :1.000   Min.   :1.000   Min.   :1.000
##  1st Qu.:2.000   1st Qu.:2.000   1st Qu.:2.000   1st Qu.:2.000
##  Median :2.000   Median :3.000   Median :2.000   Median :3.000
##  Mean   :2.051   Mean   :2.513   Mean   :2.282   Mean   :2.667
##  3rd Qu.:2.000   3rd Qu.:3.000   3rd Qu.:3.000   3rd Qu.:3.000
##  Max.   :3.000   Max.   :4.000   Max.   :3.000   Max.   :4.000
##    Artifact            Repeated
##  Length:39          Min.   :1
##  Class :character   1st Qu.:1
##  Mode  :character   Median :1
##                     Mean   :1
```

```
##                      3rd Qu.:1
##                      Max.   :1

round(sd(ratings_sub$RsrchQ), 3)

## [1] 0.56

round(sd(ratings_sub$CritDes), 3)

## [1] 0.724

round(sd(ratings_sub$InitEDA), 3)

## [1] 0.544

round(sd(ratings_sub$SelMeth), 3)

## [1] 0.51

round(sd(ratings_sub$InterpRes), 3)

## [1] 0.601

round(sd(ratings_sub$VisOrg), 3)

## [1] 0.605

round(sd(ratings_sub$TxtOrg), 3)

## [1] 0.621

par(mfrow=c(1,2))
ratings_sub$Semester <- as.factor(ratings_sub$Semester)
plot(ratings_sub$Semester, xlab="Semester", ylab="Count")
ratings_sub$Sex <- as.factor(ratings_sub$Sex)
plot(ratings_sub$Sex, xlab="Sex", ylab="Count")
```

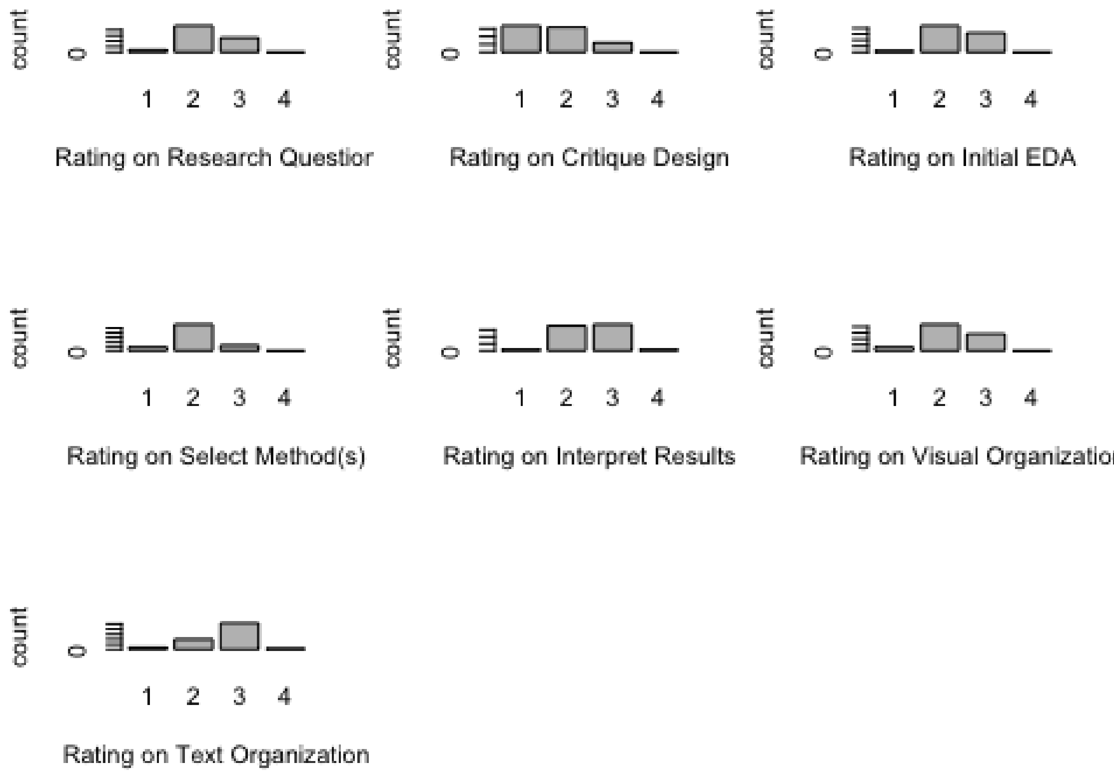


Make these rubric as categorical variables for those 13 artifacts

```
ratings_sub$RsrchQ <- factor(ratings_sub$RsrchQ, levels=1:4)
ratings_sub$CritDes <- factor(ratings_sub$CritDes, levels=1:4)
ratings_sub$InitEDA <- factor(ratings_sub$InitEDA, levels=1:4)
ratings_sub$SelMeth <- factor(ratings_sub$SelMeth, levels=1:4)
ratings_sub$InterpRes <- factor(ratings_sub$InterpRes, levels=1:4)
ratings_sub$VisOrg <- factor(ratings_sub$VisOrg, levels=1:4)
ratings_sub$TxtOrg <- factor(ratings_sub$TxtOrg, levels=1:4)
```

The distributions of ratings of each rubric

```
par(mfrow=c(3,3))
plot(ratings_sub$RsrchQ, xlab = "Rating on Research Question", ylab = "count"
)
plot(ratings_sub$CritDes, xlab = "Rating on Critique Design", ylab = "count")
plot(ratings_sub$InitEDA, xlab = "Rating on Initial EDA", ylab = "count")
plot(ratings_sub$SelMeth, xlab = "Rating on Select Method(s)", ylab = "count"
)
plot(ratings_sub$InterpRes, xlab = "Rating on Interpret Results", ylab = "cou
nt")
plot(ratings_sub$VisOrg, xlab = "Rating on Visual Organization", ylab = "coun
t")
plot(ratings_sub$TxtOrg, xlab = "Rating on Text Organization", ylab = "count"
)
```



Distributions of Ratings on each rubric

Calculating the percentage of ratings given each rubric

Frequency table and proportion table for each rubric of those 13 artifacts

```
RsrchQ <- table(ratings_sub$RsrchQ)
addmargins(RsrchQ)

##
##    1   2   3   4 Sum
##    2  24  13   0  39

round(prop.table(RsrchQ)*100,digits=0)

##
##   1  2  3  4
##   5 62 33  0

CritDes <- table(ratings_sub$CritDes)
addmargins(CritDes)

##
##    1   2   3   4 Sum
##   17  16   6   0  39
```

```
round(prop.table(CritDes)*100,digits=0)

##
##  1  2  3  4
## 44 41 15  0

InitEDA <- table(ratings_sub$InitEDA)
addmargins(InitEDA)

##
##   1   2   3   4 Sum
##   1  22  16   0  39

round(prop.table(InitEDA)*100,digits=0)

##
##  1  2  3  4
##  3 56 41  0

SelMeth <- table(ratings_sub$SelMeth)
addmargins(SelMeth)

##
##   1   2   3   4 Sum
##   4  29   6   0  39

round(prop.table(SelMeth)*100,digits=0)

##
##  1  2  3  4
## 10 74 15  0

InterpRes <- table(ratings_sub$InterpRes)
addmargins(InterpRes)

##
##   1   2   3   4 Sum
##   1  18  19   1  39

round(prop.table(InterpRes)*100,digits=0)

##
##  1  2  3  4
##  3 46 49  3

VisOrg<- table(ratings_sub$VisOrg)
addmargins(VisOrg)

##
##   1   2   3   4 Sum
##   3  22  14   0  39

round(prop.table(VisOrg)*100,digits=0)
```

```
##
##  1  2  3  4
##  8 56 36  0
```

```
TxtOrg <- table(ratings_sub$TxtOrg)
addmargins(TxtOrg)
```

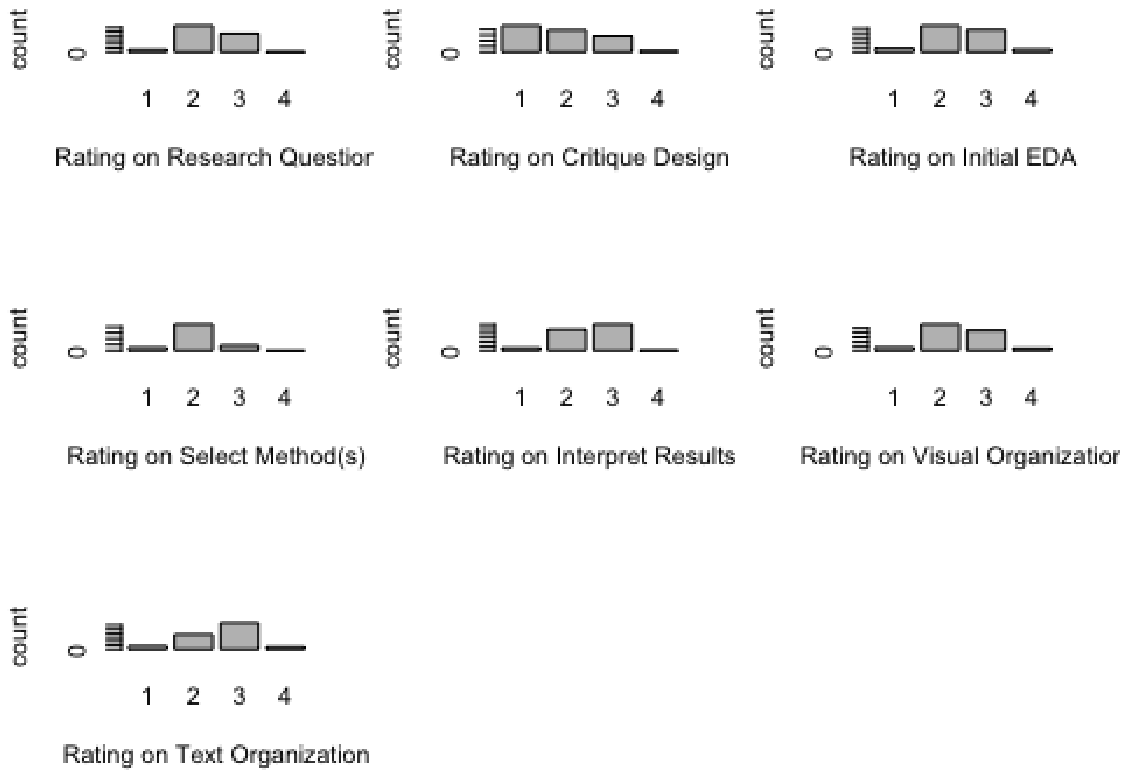
```
##
##   1   2   3   4 Sum
##   2  10  26   1  39
```

```
round(prop.table(TxtOrg)*100,digits=0)
```

```
##
##  1  2  3  4
##  5 26 67  3
```

Working on the full data

```
ratings$RsrchQ <- factor(ratings$RsrchQ, levels=1:4)
ratings$CritDes <- factor(ratings$CritDes, levels=1:4)
ratings$InitEDA <- factor(ratings$InitEDA, levels=1:4)
ratings$SelMeth <- factor(ratings$SelMeth, levels=1:4)
ratings$InterpRes <- factor(ratings$InterpRes, levels=1:4)
ratings$VisOrg <- factor(ratings$VisOrg, levels=1:4)
ratings$TxtOrg <- factor(ratings$TxtOrg, levels=1:4)

par(mfrow=c(3,3))
plot(ratings$RsrchQ, xlab = "Rating on Research Question", ylab = "count")
plot(ratings$CritDes, xlab = "Rating on Critique Design", ylab = "count")
plot(ratings$InitEDA, xlab = "Rating on Initial EDA", ylab = "count")
plot(ratings$SelMeth, xlab = "Rating on Select Method(s)", ylab = "count")
plot(ratings$InterpRes, xlab = "Rating on Interpret Results", ylab = "count")
plot(ratings$VisOrg, xlab = "Rating on Visual Organization", ylab = "count")
plot(ratings$TxtOrg, xlab = "Rating on Text Organization", ylab = "count")
```




```
RsrchQ_full <- table(ratings$RsrchQ, useNA = "always")
addmargins(RsrchQ_full)

##
##    1    2    3    4 <NA>  Sum
##    6   65   45    1    0  117

round(prop.table(RsrchQ_full)*100,digits=0)

##
##    1    2    3    4 <NA>
##    5   56   38    1    0

CritDes_full <- table(ratings$CritDes, useNA = "always")
addmargins(CritDes_full)

##
##    1    2    3    4 <NA>  Sum
##   47   39   28    2    1  117

round(prop.table(CritDes_full)*100,digits=0)

##
##    1    2    3    4 <NA>
##   40   33   24    2    1
```

```r
InitEDA_full <- table(ratings$InitEDA, useNA = "always")
addmargins(InitEDA_full)

##
##    1    2    3    4 <NA>  Sum
##    8   56   47    6    0  117

round(prop.table(InitEDA_full)*100,digits=0)

##
##    1    2    3    4 <NA>
##    7   48   40    5    0

SelMeth_full <- table(ratings$SelMeth, useNA = "always")
addmargins(SelMeth_full)

##
##    1    2    3    4 <NA>  Sum
##   10   89   18    0    0  117

round(prop.table(SelMeth_full)*100,digits=0)

##
##    1    2    3    4 <NA>
##    9   76   15    0    0

InterpRes_full <- table(ratings$InterpRes, useNA = "always")
addmargins(InterpRes_full)

##
##    1    2    3    4 <NA>  Sum
##    6   49   61    1    0  117

round(prop.table(InterpRes_full)*100,digits=0)

##
##    1    2    3    4 <NA>
##    5   42   52    1    0

VisOrg_full <- table(ratings$VisOrg, useNA = "always")
addmargins(VisOrg_full)

##
##    1    2    3    4 <NA>  Sum
##    7   59   45    5    1  117

round(prop.table(VisOrg_full)*100,digits=0)

##
##    1    2    3    4 <NA>
##    6   50   38    4    1
```

```
TxtOrg_full <- table(ratings$TxtOrg, useNA = "always")
addmargins(TxtOrg_full)
```

```
##
##    1    2    3    4 <NA>  Sum
##    8   37   66    6    0  117
```

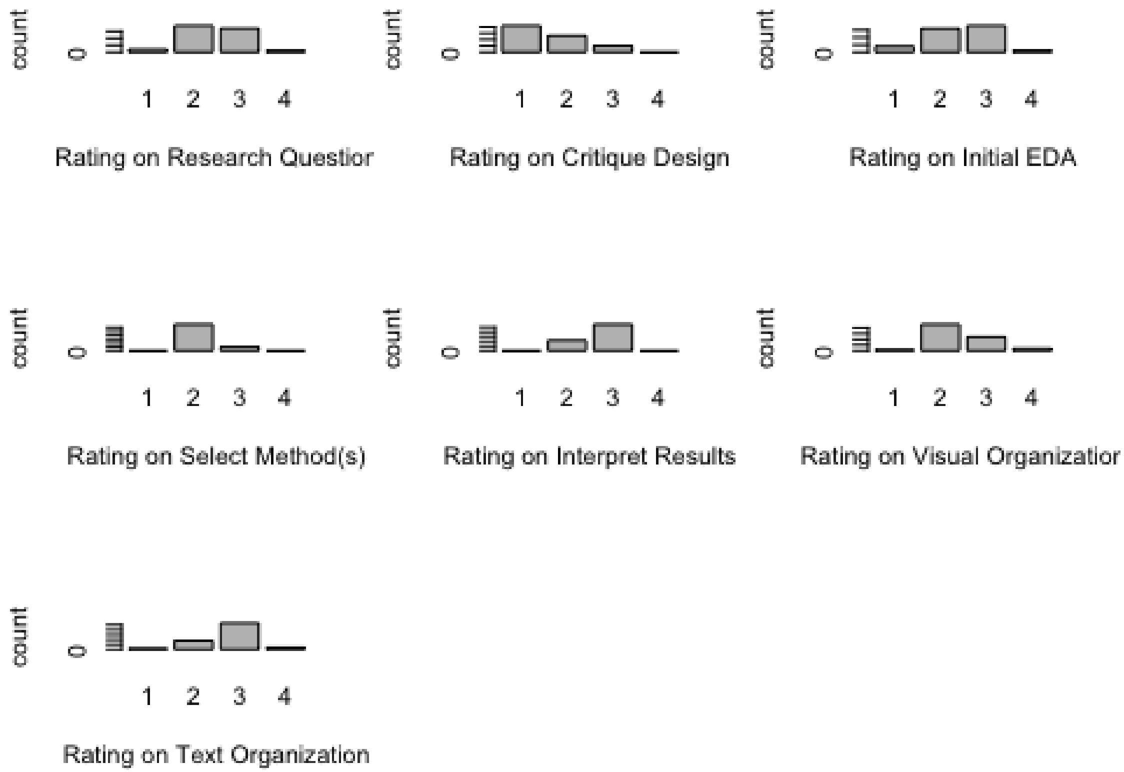
```
round(prop.table(TxtOrg_full)*100,digits=0)
```

```
##
##    1    2    3    4 <NA>
##    7   32   56    5    0
```

Rater 1

The distributions of how Rater 1 rates on different rubrics

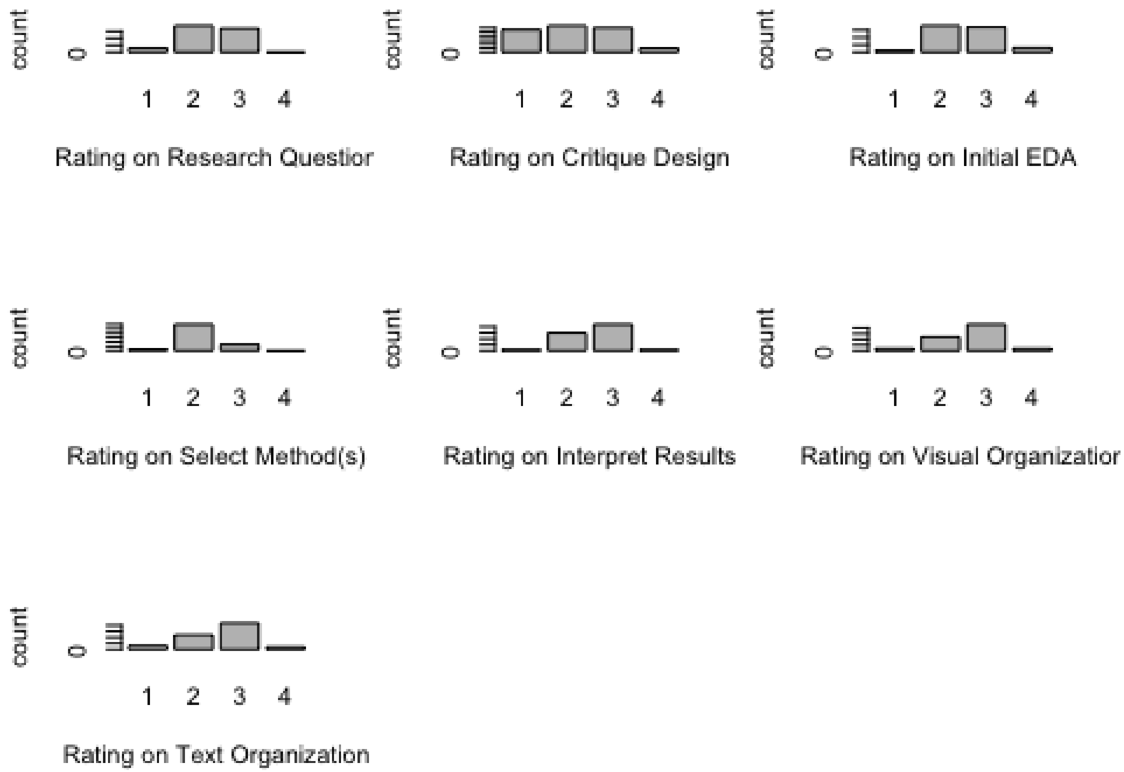
```
ratings_1 <- ratings %>%
  filter(ratings$Rater == 1)

par(mfrow=c(3,3))
plot(ratings_1$RsrchQ, xlab = "Rating on Research Question", ylab = "count")
plot(ratings_1$CritDes, xlab = "Rating on Critique Design", ylab = "count")
plot(ratings_1$InitEDA, xlab = "Rating on Initial EDA", ylab = "count")
plot(ratings_1$SelMeth, xlab = "Rating on Select Method(s)", ylab = "count")
plot(ratings_1$InterpRes, xlab = "Rating on Interpret Results", ylab = "count
")
plot(ratings_1$VisOrg, xlab = "Rating on Visual Organization", ylab = "count"
)
plot(ratings_1$TxtOrg, xlab = "Rating on Text Organization", ylab = "count")
```



Rater 2

The distributions of how Rater 2 rates on different rubrics

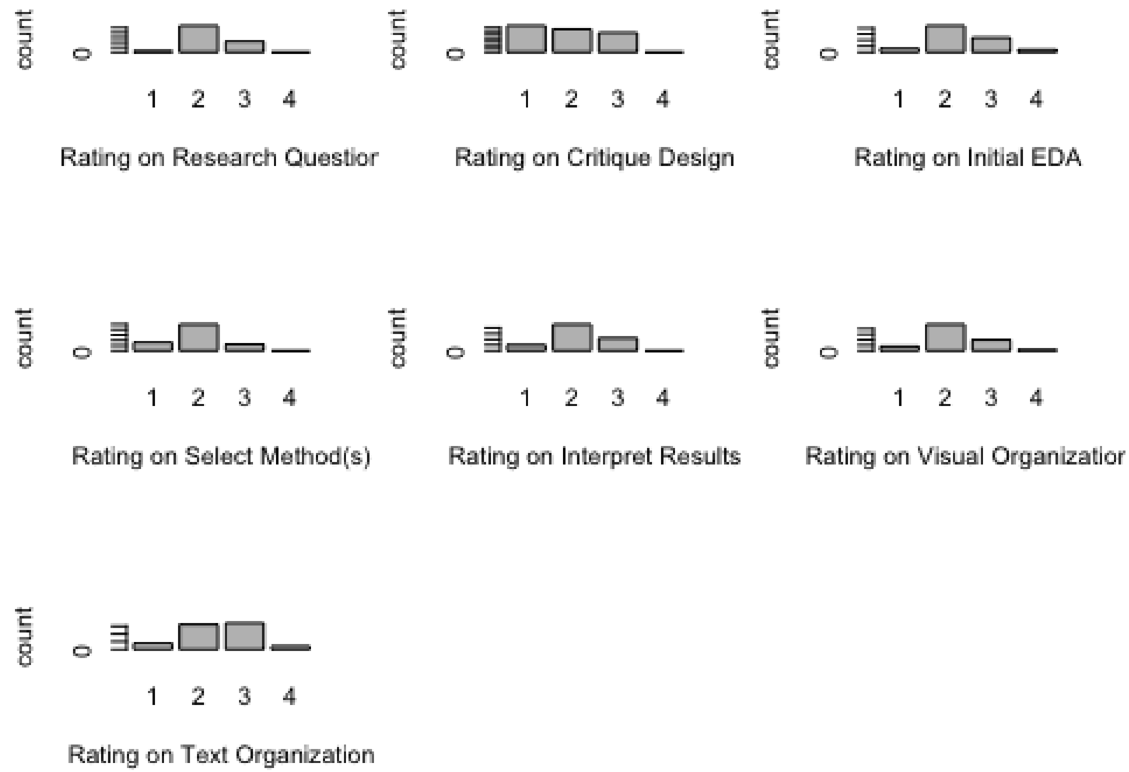
```
ratings_2 <- ratings %>%
  filter(ratings$Rater == 2)

par(mfrow=c(3,3))
plot(ratings_2$RsrchQ, xlab = "Rating on Research Question", ylab = "count")
plot(ratings_2$CritDes, xlab = "Rating on Critique Design", ylab = "count")
plot(ratings_2$InitEDA, xlab = "Rating on Initial EDA", ylab = "count")
plot(ratings_2$SelMeth, xlab = "Rating on Select Method(s)", ylab = "count")
plot(ratings_2$InterpRes, xlab = "Rating on Interpret Results", ylab = "count
")
plot(ratings_2$VisOrg, xlab = "Rating on Visual Organization", ylab = "count"
)
plot(ratings_2$TxtOrg, xlab = "Rating on Text Organization", ylab = "count")
```



Rater 3

The distributions of how Rater 3 rates on different rubrics

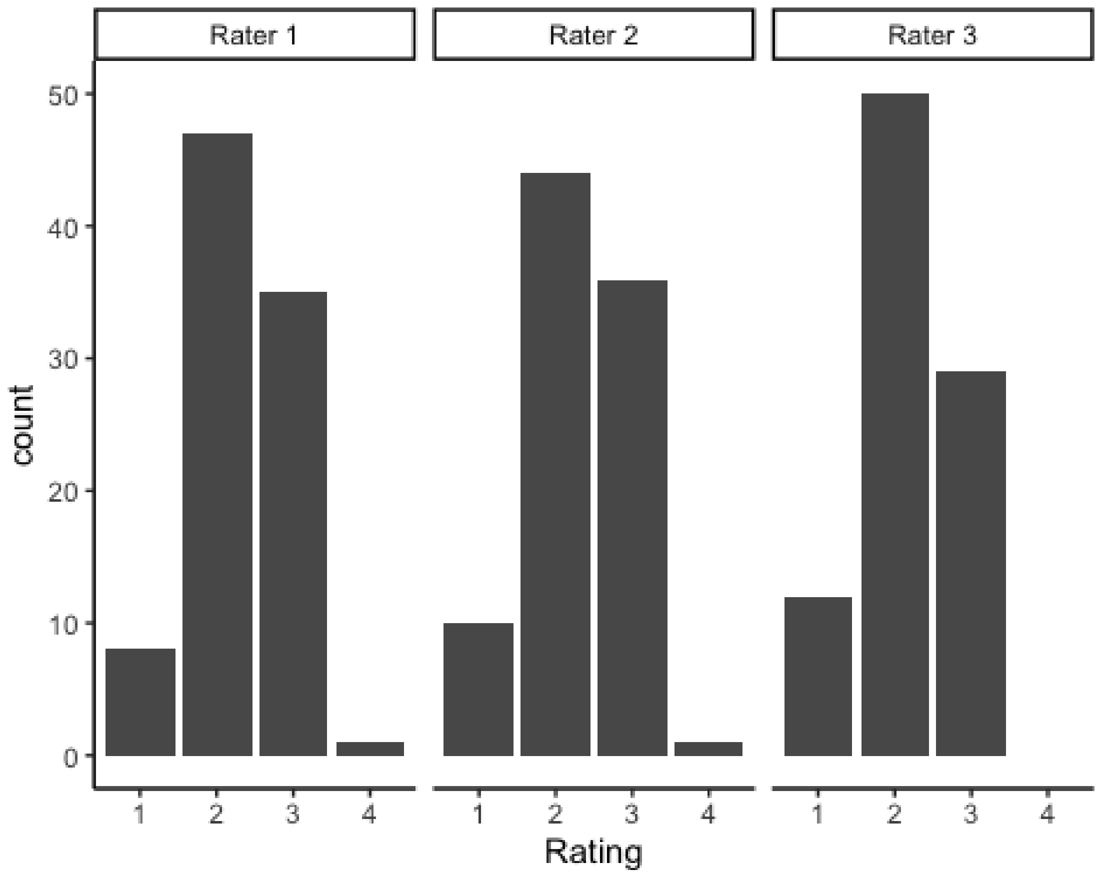
```
ratings_3 <- ratings %>%
  filter(ratings$Rater == 3)

par(mfrow=c(3,3))
plot(ratings_3$RsrchQ, xlab = "Rating on Research Question", ylab = "count")
plot(ratings_3$CritDes, xlab = "Rating on Critique Design", ylab = "count")
plot(ratings_3$InitEDA, xlab = "Rating on Initial EDA", ylab = "count")
plot(ratings_3$SelMeth, xlab = "Rating on Select Method(s)", ylab = "count")
plot(ratings_3$InterpRes, xlab = "Rating on Interpret Results", ylab = "count
")
plot(ratings_3$VisOrg, xlab = "Rating on Visual Organization", ylab = "count"
)
plot(ratings_3$TxtOrg, xlab = "Rating on Text Organization", ylab = "count")
```



Compare distributions across Raters using tall dataset

First, we focus on those 13 artifacts

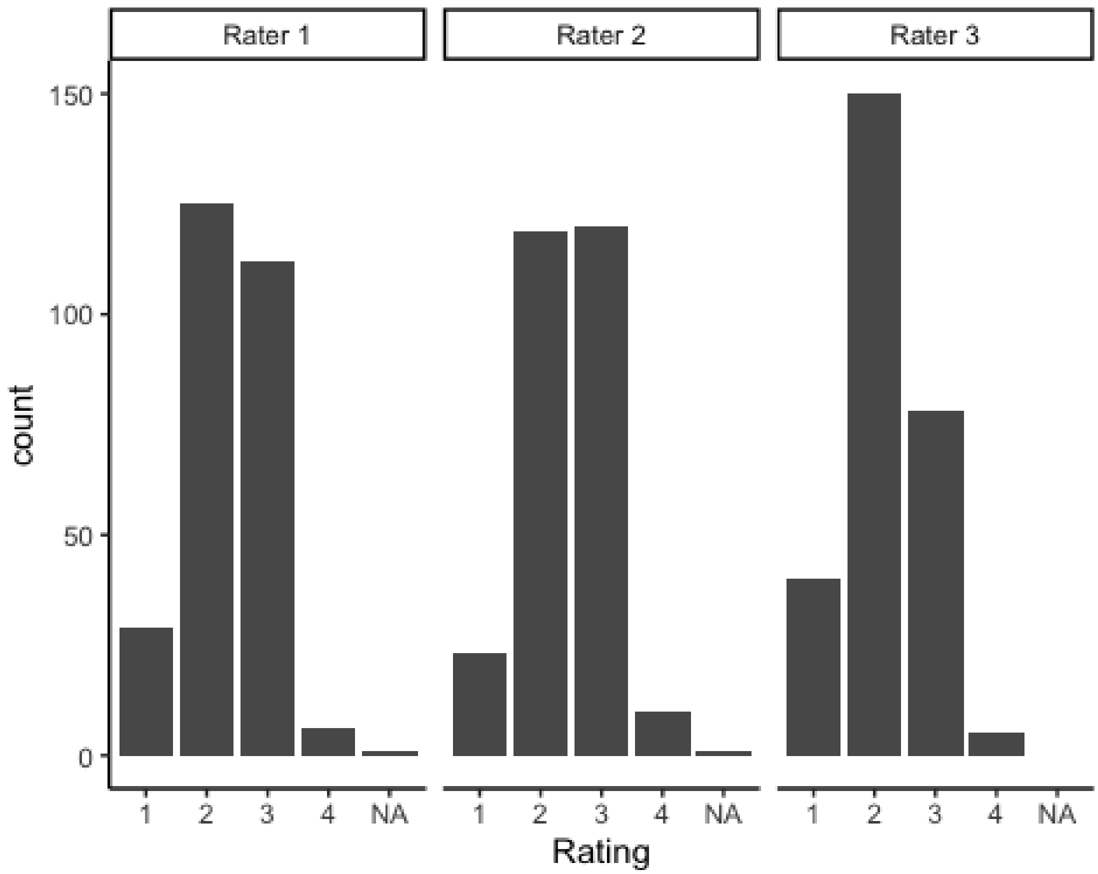
```
tall$Rating <- factor(tall$Rating,levels=1:4)
tall$Sex[nchar(tall$Sex)==0] <- "--"
tall_sub <- tall[grep("O",tall$Artifact),]

## Needed to make the title of each facet more human-readable...
rater.name <- function(x) { paste("Rater",x) }
## Barplots for reduced data...
g <- ggplot(tall_sub, aes(x = Rating)) +
  facet_wrap( ~ Rater, labeller=labeller(Rater=rater.name)) +
  geom_bar() + theme_classic()

g
```




```
## Corresponding table of counts...
tmp <- data.frame(lapply(split(tall_sub$Rating,tall_sub$Rater),summary))
row.names(tmp) <- paste("Rating",1:4)
names(tmp) <- paste("Rater",1:3)

tmp

##          Rater 1 Rater 2 Rater 3
## Rating 1       8      10      12
## Rating 2      47      44      50
## Rating 3      35      36      29
## Rating 4       1       1       0
```

Working on full data

```
## Barplots for full data...
g <- ggplot(tall,aes(x = Rating)) +
  facet_wrap( ~ Rater, labeller=labeller(Rater=rater.name)) +
  geom_bar() + theme_classic()
g
```




```
## Corresponding table of counts...
tmp0 <- lapply(split(tall$Rating,tall$Rater),summary)
tmp <- data.frame(matrix(0,nrow=5,ncol=3)) ## three raters...
names(tmp) <- names(tmp0)
row.names(tmp) <- c(paste("Rating",1:4),"<NA>")
for (i in names(tmp0)) {
  tmp[,i] <- tmp[,i] + c(tmp0[[i]],0)[1:5]
}
names(tmp) <- paste("Rater",1:3)
tmp

##          Rater 1 Rater 2 Rater 3
## Rating 1      29      23      40
## Rating 2     125     119     150
## Rating 3     112     120      78
## Rating 4       6      10       5
## <NA>           1       1       0

tall[apply(tall,1,function(x){any(is.na(x))}),]

##       X Rater Artifact Repeated Semester Sex  Rubric Rating
## 161 161     2       45        0      S19   F CritDes   <NA>
## 684 684     1      100        0      F19   F  VisOrg   <NA>
```

```
ratings[ratings$Sex=="--",]

##   X Rater Sample Overlap Semester Sex RsrchQ CritDes InitEDA SelMeth Inter
pRes
## 5 5     3      5      NA     Fall  --      3       3       3       3
   3
##   VisOrg TxtOrg Artifact Repeated
## 5      3      3        5        0
```

Appendix B

Investigate the rating consistency over these three raters on those 13 artifacts

Calculating ICC's as a measure of rater agreement

```
names(tall)

## [1] "X"        "Rater"    "Artifact" "Repeated" "Semester" "Sex"      "Rub
ric"
## [8] "Rating"
```

Group the ratings

```
common <- tall[grep("O",tall$Artifact),]
head(common)

##     X Rater Artifact Repeated Semester Sex Rubric Rating
## 1   1     3       O5        1      F19   M RsrchQ      3
## 2   2     3       O7        1      F19   F RsrchQ      3
## 3   3     3       O9        1      S19   F RsrchQ      2
## 4   4     3       O8        1      S19   M RsrchQ      2
## 10 10     3      O10        1      F19   F RsrchQ      2
## 11 11     3      O13        1      F19   M RsrchQ      2

dim(common)

## [1] 273   8
```

ICC calculations on each rubric for the 13 artifacts that all three raters saw

```
common$Rater <- as.factor(common$Rater)
common$Artifact <- as.factor(common$Artifact)
common$Semester <- as.factor(common$Semester)
common$Sex <- as.factor(common$Sex)
```

Ratings on Rating on Research Question

```
RsrchQ.ratings <- common[common$Rubric=="RsrchQ",]
RsrchQ_1 <- lmer(as.numeric(Rating) ~ 1 + (1|Rater), data=RsrchQ.ratings)

## boundary (singular) fit: see ?isSingular

summary(RsrchQ_1)

## Linear mixed model fit by REML ['lmerMod']
## Formula: as.numeric(Rating) ~ 1 + (1 | Rater)
```

```
##    Data: RsrchQ.ratings
##
## REML criterion at convergence: 67.4
##
## Scaled residuals:
##     Min      1Q  Median      3Q     Max
## -2.2912 -0.5041 -0.5041  1.2831  1.2831
##
## Random effects:
##  Groups   Name        Variance Std.Dev.
##  Rater    (Intercept) 0.0000   0.0000
##  Residual             0.3131   0.5595
## Number of obs: 39, groups:  Rater, 3
##
## Fixed effects:
##             Estimate Std. Error t value
## (Intercept)   2.2820     0.0896   25.47
## optimizer (nloptwrap) convergence code: 0 (OK)
## boundary (singular) fit: see ?isSingular
```

```
RsrchQ_ICC_1 <- (0.0000)/(0.0000+0.3131)
RsrchQ_ICC_1
```

```
## [1] 0
```

```
RsrchQ_2 <- lmer(as.numeric(Rating) ~ 1 + (1|Artifact), data=RsrchQ.ratings)
summary(RsrchQ_2)
```

```
## Linear mixed model fit by REML ['lmerMod']
## Formula: as.numeric(Rating) ~ 1 + (1 | Artifact)
##    Data: RsrchQ.ratings
##
## REML criterion at convergence: 66.2
##
## Scaled residuals:
##     Min      1Q  Median      3Q     Max
## -2.3025 -0.5987 -0.3276  0.9696  1.6472
##
## Random effects:
##  Groups   Name        Variance Std.Dev.
##  Artifact (Intercept) 0.05983  0.2446
##  Residual             0.25641  0.5064
## Number of obs: 39, groups:  Artifact, 13
##
## Fixed effects:
##             Estimate Std. Error t value
## (Intercept)   2.2821     0.1057   21.59
```

```
RsrchQ_ICC_2 <- (0.05983)/(0.05983+0.25641)
RsrchQ_ICC_2
```

```
## [1] 0.1891918
```

Rating on Critique Design

```
CritDes.ratings <- common[common$Rubric=="CritDes",]
CritDes_1 <- lmer(as.numeric(Rating) ~ 1 + (1|Rater), data=CritDes.ratings)

## boundary (singular) fit: see ?isSingular

summary(CritDes_1)

## Linear mixed model fit by REML ['lmerMod']
## Formula: as.numeric(Rating) ~ 1 + (1 | Rater)
##    Data: CritDes.ratings
##
## REML criterion at convergence: 86.9
##
## Scaled residuals:
##     Min      1Q  Median      3Q     Max
## -0.9922 -0.9922  0.3898  0.3898  1.7717
##
## Random effects:
##  Groups   Name        Variance Std.Dev.
##  Rater    (Intercept) 0.0000   0.0000
##  Residual             0.5236   0.7236
## Number of obs: 39, groups:  Rater, 3
##
## Fixed effects:
##             Estimate Std. Error t value
## (Intercept)   1.7179     0.1159   14.83
## optimizer (nloptwrap) convergence code: 0 (OK)
## boundary (singular) fit: see ?isSingular

CritDes_ICC_1 <- (0.0000)/(0.0000+0.5236)
CritDes_ICC_1

## [1] 0

CritDes_2 <- lmer(as.numeric(Rating) ~ 1 + (1|Artifact), data=CritDes.ratings
)
summary(CritDes_2)

## Linear mixed model fit by REML ['lmerMod']
## Formula: as.numeric(Rating) ~ 1 + (1 | Artifact)
##    Data: CritDes.ratings
##
## REML criterion at convergence: 75.1
##
## Scaled residuals:
##     Min      1Q  Median      3Q     Max
## -1.9647 -0.4386 -0.2978  0.5318  2.1987
##
```

```
## Random effects:
##  Groups   Name        Variance Std.Dev.
##  Artifact (Intercept) 0.3091   0.5560
##  Residual             0.2308   0.4804
## Number of obs: 39, groups:  Artifact, 13
##
## Fixed effects:
##             Estimate Std. Error t value
## (Intercept)   1.7179     0.1723   9.969
```

```
CritDes_ICC_2 <- (0.3091)/(0.3091+0.2308)
CritDes_ICC_2
```

```
## [1] 0.5725134
```

Rating on Initial EDA

```
InitEDA.ratings <- common[common$Rubric=="InitEDA",]
InitEDA_1 <- lmer(as.numeric(Rating) ~ 1 + (1|Rater), data=InitEDA.ratings)
summary(InitEDA_1)
```

```
## Linear mixed model fit by REML ['lmerMod']
## Formula: as.numeric(Rating) ~ 1 + (1 | Rater)
##    Data: InitEDA.ratings
##
## REML criterion at convergence: 65.2
##
## Scaled residuals:
##     Min      1Q  Median      3Q     Max
## -2.5616 -0.7083 -0.6965  1.1215  1.1451
##
## Random effects:
##  Groups   Name        Variance  Std.Dev.
##  Rater    (Intercept) 0.0009862 0.0314
##  Residual             0.2948718 0.5430
## Number of obs: 39, groups:  Rater, 3
##
## Fixed effects:
##             Estimate Std. Error t value
## (Intercept)  2.38462    0.08882   26.85
```

```
InitEDA_ICC_1 <- (0.0009862)/(0.0009862+0.2948718)
InitEDA_ICC_1
```

```
## [1] 0.003333356
```

```
InitEDA_2 <- lmer(as.numeric(Rating) ~ 1 + (1|Artifact), data=InitEDA.ratings
)
summary(InitEDA_2)
```

```
## Linear mixed model fit by REML ['lmerMod']
## Formula: as.numeric(Rating) ~ 1 + (1 | Artifact)
```

```
##    Data: InitEDA.ratings
##
## REML criterion at convergence: 56.8
##
## Scaled residuals:
##     Min      1Q  Median      3Q     Max
## -2.1670 -0.2504 -0.2504  0.4006  1.6663
##
## Random effects:
##  Groups   Name        Variance Std.Dev.
##  Artifact (Intercept) 0.1496   0.3867
##  Residual             0.1538   0.3922
## Number of obs: 39, groups:  Artifact, 13
##
## Fixed effects:
##             Estimate Std. Error t value
## (Intercept)   2.3846     0.1243   19.18
```

```
InitEDA_ICC_2 <- (0.1496)/(0.1496+0.1538)
InitEDA_ICC_2
```

```
## [1] 0.4930784
```

Rating on Select Method(s)

```
SelMeth.ratings <- common[common$Rubric=="SelMeth",]
SelMeth_1 <- lmer(as.numeric(Rating) ~ 1 + (1|Rater), data=SelMeth.ratings)
```

```
## boundary (singular) fit: see ?isSingular
```

```
summary(SelMeth_1)
```

```
## Linear mixed model fit by REML ['lmerMod']
## Formula: as.numeric(Rating) ~ 1 + (1 | Rater)
##    Data: SelMeth.ratings
##
## REML criterion at convergence: 60.4
##
## Scaled residuals:
##     Min      1Q  Median      3Q     Max
## -2.0599 -0.1005 -0.1005 -0.1005  1.8590
##
## Random effects:
##  Groups   Name        Variance Std.Dev.
##  Rater    (Intercept) 0.0000   0.0000
##  Residual             0.2605   0.5104
## Number of obs: 39, groups:  Rater, 3
##
## Fixed effects:
##             Estimate Std. Error t value
## (Intercept)  2.05128    0.08172    25.1
```

```
## optimizer (nloptwrap) convergence code: 0 (OK)
## boundary (singular) fit: see ?isSingular

SelMeth_ICC_1 <- (0.0000)/(0.0000+0.2605)
SelMeth_ICC_1

## [1] 0

SelMeth_2 <- lmer(as.numeric(Rating) ~ 1 + (1|Artifact), data=SelMeth.ratings
)
summary(SelMeth_2)

## Linear mixed model fit by REML ['lmerMod']
## Formula: as.numeric(Rating) ~ 1 + (1 | Artifact)
##    Data: SelMeth.ratings
##
## REML criterion at convergence: 50.9
##
## Scaled residuals:
##      Min       1Q   Median       3Q      Max
## -2.11366 -0.03357 -0.03357  0.62101  2.04652
##
## Random effects:
##  Groups   Name        Variance Std.Dev.
##  Artifact (Intercept) 0.1396   0.3736
##  Residual             0.1282   0.3581
## Number of obs: 39, groups:  Artifact, 13
##
## Fixed effects:
##             Estimate Std. Error t value
## (Intercept)   2.0513     0.1184   17.32

SelMeth_ICC_2 <- (0.1396)/(0.1396+0.1282)
SelMeth_ICC_2

## [1] 0.5212845
```

Rating on Interpret Results

```
InterpRes.ratings <- common[common$Rubric=="InterpRes",]
InterpRes_1 <- lmer(as.numeric(Rating) ~ 1 + (1|Rater), data=InterpRes.rating
s)
summary(InterpRes_1)

## Linear mixed model fit by REML ['lmerMod']
## Formula: as.numeric(Rating) ~ 1 + (1 | Rater)
##    Data: InterpRes.ratings
##
## REML criterion at convergence: 72.8
##
## Scaled residuals:
##     Min      1Q  Median      3Q     Max
```

```
## -2.4822 -0.8773  0.7917  0.7917  2.4608
##
## Random effects:
##  Groups   Name        Variance Std.Dev.
##  Rater    (Intercept) 0.003945 0.06281
##  Residual             0.358974 0.59914
## Number of obs: 39, groups:  Rater, 3
##
## Fixed effects:
##             Estimate Std. Error t value
## (Intercept)   2.5128     0.1026    24.5
```

```
InterpRes_ICC_1 <- (0.003945)/(0.003945+0.358974)
InterpRes_ICC_1
```

```
## [1] 0.01087019
```

```
InterpRes_2 <- lmer(as.numeric(Rating) ~ 1 + (1|Artifact), data=InterpRes.rat
ings)
summary(InterpRes_2)
```

```
## Linear mixed model fit by REML ['lmerMod']
## Formula: as.numeric(Rating) ~ 1 + (1 | Artifact)
##    Data: InterpRes.ratings
##
## REML criterion at convergence: 71.1
##
## Scaled residuals:
##     Min      1Q  Median      3Q     Max
## -2.0965 -0.8061  0.4844  0.7806  2.6635
##
## Random effects:
##  Groups   Name        Variance Std.Dev.
##  Artifact (Intercept) 0.08405  0.2899
##  Residual             0.28205  0.5311
## Number of obs: 39, groups:  Artifact, 13
##
## Fixed effects:
##             Estimate Std. Error t value
## (Intercept)    2.513      0.117   21.47
```

```
InterpRes_ICC_2 <- (0.08405)/(0.08405+0.28205)
InterpRes_ICC_2
```

```
## [1] 0.2295821
```

Rating on Visual Organization

```
VisOrg.ratings <- common[common$Rubric=="VisOrg",]
VisOrg_1 <- lmer(as.numeric(Rating) ~ 1 + (1|Rater), data=VisOrg.ratings)
```

```
## boundary (singular) fit: see ?isSingular
```

```
summary(VisOrg_1)

## Linear mixed model fit by REML ['lmerMod']
## Formula: as.numeric(Rating) ~ 1 + (1 | Rater)
##    Data: VisOrg.ratings
##
## REML criterion at convergence: 73.3
##
## Scaled residuals:
##     Min      1Q  Median      3Q     Max
## -2.1200 -0.4664 -0.4664  1.1872  1.1872
##
## Random effects:
##  Groups   Name        Variance Std.Dev.
##  Rater    (Intercept) 0.0000   0.0000
##  Residual             0.3657   0.6047
## Number of obs: 39, groups:  Rater, 3
##
## Fixed effects:
##             Estimate Std. Error t value
## (Intercept)  2.28205    0.09684   23.57
## optimizer (nloptwrap) convergence code: 0 (OK)
## boundary (singular) fit: see ?isSingular

VisOrg_ICC_1 <- (0.0000)/(0.0000+0.3657)
VisOrg_ICC_1

## [1] 0

VisOrg_2 <- lmer(as.numeric(Rating) ~ 1 + (1|Artifact), data=VisOrg.ratings)
summary(VisOrg_2)

## Linear mixed model fit by REML ['lmerMod']
## Formula: as.numeric(Rating) ~ 1 + (1 | Artifact)
##    Data: VisOrg.ratings
##
## REML criterion at convergence: 60.5
##
## Scaled residuals:
##     Min      1Q  Median      3Q     Max
## -1.5168 -0.7176 -0.1341  0.3414  1.7241
##
## Random effects:
##  Groups   Name        Variance Std.Dev.
##  Artifact (Intercept) 0.2236   0.4729
##  Residual             0.1538   0.3922
## Number of obs: 39, groups:  Artifact, 13
##
## Fixed effects:
##             Estimate Std. Error t value
## (Intercept)   2.2821     0.1454   15.69
```

```
VisOrg_ICC_2 <- (0.2236)/(0.2236+0.1538)
VisOrg_ICC_2

## [1] 0.5924748
```

Rating on Text Organization

```
TxtOrg.ratings <- common[common$Rubric=="TxtOrg",]
TxtOrg_1 <- lmer(as.numeric(Rating) ~ 1 + (1|Rater), data=TxtOrg.ratings)

## boundary (singular) fit: see ?isSingular

summary(TxtOrg_1)

## Linear mixed model fit by REML ['lmerMod']
## Formula: as.numeric(Rating) ~ 1 + (1 | Rater)
##    Data: TxtOrg.ratings
##
## REML criterion at convergence: 75.3
##
## Scaled residuals:
##     Min      1Q  Median      3Q     Max
## -2.6827 -1.0731  0.5365  0.5365  2.1462
##
## Random effects:
##  Groups   Name        Variance Std.Dev.
##  Rater    (Intercept) 0.000    0.0000
##  Residual             0.386    0.6213
## Number of obs: 39, groups:  Rater, 3
##
## Fixed effects:
##             Estimate Std. Error t value
## (Intercept)  2.66667    0.09948   26.81
## optimizer (nloptwrap) convergence code: 0 (OK)
## boundary (singular) fit: see ?isSingular

TxtOrg_ICC_1 <- (0.000)/(0.000+0.386)
TxtOrg_ICC_1

## [1] 0

TxtOrg_2 <- lmer(as.numeric(Rating) ~ 1 + (1|Artifact), data=TxtOrg.ratings)
summary(TxtOrg_2)

## Linear mixed model fit by REML ['lmerMod']
## Formula: as.numeric(Rating) ~ 1 + (1 | Artifact)
##    Data: TxtOrg.ratings
##
## REML criterion at convergence: 74.6
##
## Scaled residuals:
##     Min      1Q  Median      3Q     Max
## -2.6943 -0.7698  0.3849  0.3849  2.5019
```

```
##
## Random effects:
##  Groups   Name        Variance Std.Dev.
##  Artifact (Intercept) 0.05556  0.2357
##  Residual             0.33333  0.5774
## Number of obs: 39, groups:  Artifact, 13
##
## Fixed effects:
##             Estimate Std. Error t value
## (Intercept)   2.6667     0.1132   23.55
```

```
TxtOrg_ICC_2 <- (0.05556)/(0.05556+0.33333)
TxtOrg_ICC_2
```

```
## [1] 0.1428682
```

Calculations of exact agreement between any two raters and on each rubric

Cross-classifying the ratings that each pair of raters gives on the subset of 13 artifacts seen by every rater

```
repeated <- ratings[ratings$Repeated==1,]
```

Rating on Research Question

```
raters_1_and_2_on_CritDes <- data.frame(r1=repeated$RsrchQ[repeated$Rater==1]
,
                                        r2=repeated$RsrchQ[repeated$Rater==2],
                                        a1=repeated$Artifact[repeated$Rater==1
],
                                        a2=repeated$Artifact[repeated$Rater==2
])
```

```
r1 <- factor(raters_1_and_2_on_CritDes$r1,levels=1:4)
r2 <- factor(raters_1_and_2_on_CritDes$r2,levels=1:4)
```

```
(t12 <- table(r1,r2))
```

```
##    r2
## r1  1 2 3 4
##   1 0 0 0 0
##   2 1 4 3 0
##   3 1 3 1 0
##   4 0 0 0 0
```

```
(4+1)/(1+4+3+1+3+1)
```

```
## [1] 0.3846154
```

```
raters_1_and_3_on_RsrchQ <- data.frame(r1=repeated$RsrchQ[repeated$Rater==1],
                                       r3=repeated$RsrchQ[repeated$Rater==3],
                                       a1=repeated$Artifact[repeated$Rater==1
],
```

```
                                             a3=repeated$Artifact[repeated$Rater==3
])

r1 <- factor(raters_1_and_3_on_RsrchQ$r1,levels=1:4)
r3 <- factor(raters_1_and_3_on_RsrchQ$r3,levels=1:4)

(t13 <- table(r1,r3))

##     r3
## r1   1 2 3 4
##    1 0 0 0 0
##    2 0 7 1 0
##    3 0 2 3 0
##    4 0 0 0 0

(7+3)/(7+3+2+3)

## [1] 0.6666667

raters_2_and_3_on_RsrchQ <- data.frame(r2=repeated$RsrchQ[repeated$Rater==2],
                                       r3=repeated$RsrchQ[repeated$Rater==3],
                                       a2=repeated$Artifact[repeated$Rater==2
],
                                       a3=repeated$Artifact[repeated$Rater==3
])

r2 <- factor(raters_2_and_3_on_RsrchQ$r2,levels=1:4)
r3 <- factor(raters_2_and_3_on_RsrchQ$r3,levels=1:4)

(t23 <- table(r2,r3))

##     r3
## r2   1 2 3 4
##    1 0 2 0 0
##    2 0 5 2 0
##    3 0 2 2 0
##    4 0 0 0 0

(5+2)/(2+5+2+2+2)

## [1] 0.5384615
```

Rating on Critique Design

```
raters_1_and_2_on_CritDes <- data.frame(r1=repeated$CritDes[repeated$Rater==1
],
                                        r2=repeated$CritDes[repeated$Rater==2]
,
                                        a1=repeated$Artifact[repeated$Rater==1
],
                                        a2=repeated$Artifact[repeated$Rater==2
])
```

```
r1 <- factor(raters_1_and_2_on_CritDes$r1,levels=1:4)
r2 <- factor(raters_1_and_2_on_CritDes$r2,levels=1:4)

(t12 <- table(r1,r2))

##     r2
## r1   1 2 3 4
##    1 3 2 1 0
##    2 2 3 1 0
##    3 0 0 1 0
##    4 0 0 0 0

(3+3+1)/(3+2+1+2+3+1+1)

## [1] 0.5384615

raters_1_and_3_on_CritDes <- data.frame(r1=repeated$CritDes[repeated$Rater==1
],
                                       r3=repeated$CritDes[repeated$Rater==3]
,
                                       a1=repeated$Artifact[repeated$Rater==1
],
                                       a3=repeated$Artifact[repeated$Rater==3
])

r1 <- factor(raters_1_and_3_on_CritDes$r1,levels=1:4)
r3 <- factor(raters_1_and_3_on_CritDes$r3,levels=1:4)

(t13 <- table(r1,r3))

##     r3
## r1   1 2 3 4
##    1 4 2 0 0
##    2 2 3 1 0
##    3 0 0 1 0
##    4 0 0 0 0

(4+3+1)/(4+2+2+3+1+1)

## [1] 0.6153846

raters_2_and_3_on_CritDes <- data.frame(r2=repeated$CritDes[repeated$Rater==2
],
                                       r3=repeated$CritDes[repeated$Rater==3]
,
                                       a2=repeated$Artifact[repeated$Rater==2
],
                                       a3=repeated$Artifact[repeated$Rater==3
])

r2 <- factor(raters_2_and_3_on_CritDes$r2,levels=1:4)
r3 <- factor(raters_2_and_3_on_CritDes$r3,levels=1:4)
```

```
(t23 <- table(r2,r3))

##     r3
## r2   1 2 3 4
##    1 5 0 0 0
##    2 1 3 1 0
##    3 0 2 1 0
##    4 0 0 0 0

(5+3+1)/(5+1+3+1+2+1)

## [1] 0.6923077
```

Rating on Rating on Initial EDA

```
raters_1_and_2_on_InitEDA <- data.frame(r1=repeated$InitEDA[repeated$Rater==1
],
                                        r2=repeated$InitEDA[repeated$Rater==2]
,
                                        a1=repeated$Artifact[repeated$Rater==1
],
                                        a2=repeated$Artifact[repeated$Rater==2
])

r1 <- factor(raters_1_and_2_on_InitEDA$r1,levels=1:4)
r2 <- factor(raters_1_and_2_on_InitEDA$r2,levels=1:4)

(t12 <- table(r1,r2))

##     r2
## r1   1 2 3 4
##    1 0 1 0 0
##    2 0 4 0 0
##    3 0 3 5 0
##    4 0 0 0 0

(4+5)/(1+4+3+5)

## [1] 0.6923077

raters_1_and_3_on_InitEDA <- data.frame(r1=repeated$InitEDA[repeated$Rater==1
],
                                        r3=repeated$InitEDA[repeated$Rater==3]
,
                                        a1=repeated$Artifact[repeated$Rater==1
],
                                        a3=repeated$Artifact[repeated$Rater==3
])

r1 <- factor(raters_1_and_3_on_InitEDA$r1,levels=1:4)
r3 <- factor(raters_1_and_3_on_InitEDA$r3,levels=1:4)

(t13 <- table(r1,r3))
```

```
##    r3
## r1  1 2 3 4
##   1 0 1 0 0
##   2 0 4 0 0
##   3 0 5 3 0
##   4 0 0 0 0
```

```
(4+3)/(1+4+5+3)
```

```
## [1] 0.5384615
```

```
raters_2_and_3_on_InitEDA <- data.frame(r2=repeated$InitEDA[repeated$Rater==2
],
                                       r3=repeated$InitEDA[repeated$Rater==3]
,
                                       a2=repeated$Artifact[repeated$Rater==2
],
                                       a3=repeated$Artifact[repeated$Rater==3
])

r2 <- factor(raters_2_and_3_on_InitEDA$r2,levels=1:4)
r3 <- factor(raters_2_and_3_on_InitEDA$r3,levels=1:4)

(t23 <- table(r2,r3))
```

```
##    r3
## r2  1 2 3 4
##   1 0 0 0 0
##   2 0 8 0 0
##   3 0 2 3 0
##   4 0 0 0 0
```

```
(8+3)/(8+2+3)
```

```
## [1] 0.8461538
```

Rating on Select Method(s)

```
raters_1_and_2_on_SelMeth <- data.frame(r1=repeated$SelMeth[repeated$Rater==1
],
                                       r2=repeated$SelMeth[repeated$Rater==2]
,
                                       a1=repeated$Artifact[repeated$Rater==1
],
                                       a2=repeated$Artifact[repeated$Rater==2
])

r1 <- factor(raters_1_and_2_on_SelMeth$r1,levels=1:4)
r2 <- factor(raters_1_and_2_on_SelMeth$r2,levels=1:4)

(t12 <- table(r1,r2))
```

```
##    r2
## r1   1  2  3  4
##   1  0  0  0  0
##   2  1 10  0  0
##   3  0  0  2  0
##   4  0  0  0  0

(10+2)/(1+10+2)

## [1] 0.9230769

raters_1_and_3_on_SelMeth <- data.frame(r1=repeated$SelMeth[repeated$Rater==1
],
                                       r3=repeated$SelMeth[repeated$Rater==3]
,
                                       a1=repeated$Artifact[repeated$Rater==1
],
                                       a3=repeated$Artifact[repeated$Rater==3
])

r1 <- factor(raters_1_and_3_on_SelMeth$r1,levels=1:4)
r3 <- factor(raters_1_and_3_on_SelMeth$r3,levels=1:4)

(t13 <- table(r1,r3))

##    r3
## r1  1 2 3 4
##   1 0 0 0 0
##   2 3 7 1 0
##   3 0 1 1 0
##   4 0 0 0 0

(7+1)/(3+7+1+1+1)

## [1] 0.6153846

raters_2_and_3_on_SelMeth <- data.frame(r2=repeated$SelMeth[repeated$Rater==2
],
                                       r3=repeated$SelMeth[repeated$Rater==3]
,
                                       a2=repeated$Artifact[repeated$Rater==2
],
                                       a3=repeated$Artifact[repeated$Rater==3
])

r2 <- factor(raters_2_and_3_on_SelMeth$r2,levels=1:4)
r3 <- factor(raters_2_and_3_on_SelMeth$r3,levels=1:4)

(t23 <- table(r2,r3))

##    r3
## r2  1 2 3 4
```

```
##   1 1 0 0 0
##   2 2 7 1 0
##   3 0 1 1 0
##   4 0 0 0 0
```

```
(1+7+1)/(1+2+7+1+1+1)
```

```
## [1] 0.6923077
```

Rating on Interpret Results

```
raters_1_and_2_on_InterpRes <- data.frame(r1=repeated$InterpRes[repeated$Rate
r==1],
                                              r2=repeated$InterpRes[repeated$Rater==
2],
                                              a1=repeated$Artifact[repeated$Rater==1
],
                                              a2=repeated$Artifact[repeated$Rater==2
])
```

```
r1 <- factor(raters_1_and_2_on_InterpRes$r1,levels=1:4)
r2 <- factor(raters_1_and_2_on_InterpRes$r2,levels=1:4)
```

```
(t12 <- table(r1,r2))
```

```
##    r2
## r1  1 2 3 4
##   1 0 0 0 0
##   2 0 3 1 1
##   3 0 3 5 0
##   4 0 0 0 0
```

```
(3+5)/(3+1+1+3+5)
```

```
## [1] 0.6153846
```

```
raters_1_and_3_on_InterpRes <- data.frame(r1=repeated$InterpRes[repeated$Rate
r==1],
                                              r3=repeated$InterpRes[repeated$Rater==
3],
                                              a1=repeated$Artifact[repeated$Rater==1
],
                                              a3=repeated$Artifact[repeated$Rater==3
])
```

```
r1 <- factor(raters_1_and_3_on_InterpRes$r1,levels=1:4)
r3 <- factor(raters_1_and_3_on_InterpRes$r3,levels=1:4)
```

```
(t13 <- table(r1,r3))
```

```
##    r3
## r1  1 2 3 4
##   1 0 0 0 0
```

```
##   2 1 3 1 0
##   3 0 4 4 0
##   4 0 0 0 0
```

```
(3+4)/(1+3+1+4+4)
```

```
## [1] 0.5384615
```

```
raters_2_and_3_on_InterpRes <- data.frame(r2=repeated$InterpRes[repeated$Rate
r==2],
                                          r3=repeated$InterpRes[repeated$Rater==
3],
                                          a2=repeated$Artifact[repeated$Rater==2
],
                                          a3=repeated$Artifact[repeated$Rater==3
])
```

```
r2 <- factor(raters_2_and_3_on_InterpRes$r2,levels=1:4)
r3 <- factor(raters_2_and_3_on_InterpRes$r3,levels=1:4)
```

```
(t23 <- table(r2,r3))
```

```
##    r3
## r2  1 2 3 4
##   1 0 0 0 0
##   2 1 4 1 0
##   3 0 2 4 0
##   4 0 1 0 0
```

```
(4+4)/(1+4+1+2+4+1)
```

```
## [1] 0.6153846
```

Rating on Visual Organization

```
raters_1_and_2_on_VisOrg <- data.frame(r1=repeated$VisOrg[repeated$Rater==1],
                                       r2=repeated$VisOrg[repeated$Rater==2],
                                       a1=repeated$Artifact[repeated$Rater==1
],
                                       a2=repeated$Artifact[repeated$Rater==2
])
```

```
r1 <- factor(raters_1_and_2_on_VisOrg$r1,levels=1:4)
r2 <- factor(raters_1_and_2_on_VisOrg$r2,levels=1:4)
```

```
(t12 <- table(r1,r2))
```

```
##    r2
## r1  1 2 3 4
##   1 1 0 0 0
##   2 0 4 5 0
##   3 0 1 2 0
##   4 0 0 0 0
```

```
(1+4+2)/(1+4+5+1+2)

## [1] 0.5384615

raters_1_and_3_on_VisOrg <- data.frame(r1=repeated$VisOrg[repeated$Rater==1],
                                       r3=repeated$VisOrg[repeated$Rater==3],
                                       a1=repeated$Artifact[repeated$Rater==1
],
                                       a3=repeated$Artifact[repeated$Rater==3
])

r1 <- factor(raters_1_and_3_on_VisOrg$r1,levels=1:4)
r3 <- factor(raters_1_and_3_on_VisOrg$r3,levels=1:4)

(t13 <- table(r1,r3))

##    r3
## r1  1 2 3 4
##   1 1 0 0 0
##   2 0 7 2 0
##   3 0 1 2 0
##   4 0 0 0 0

(1+7+2)/(1+7+2+1+2)

## [1] 0.7692308

raters_2_and_3_on_VisOrg <- data.frame(r2=repeated$VisOrg[repeated$Rater==2],
                                       r3=repeated$VisOrg[repeated$Rater==3],
                                       a2=repeated$Artifact[repeated$Rater==2
],
                                       a3=repeated$Artifact[repeated$Rater==3
])

r2 <- factor(raters_2_and_3_on_VisOrg$r2,levels=1:4)
r3 <- factor(raters_2_and_3_on_VisOrg$r3,levels=1:4)

(t23 <- table(r2,r3))

##    r3
## r2  1 2 3 4
##   1 1 0 0 0
##   2 0 5 0 0
##   3 0 3 4 0
##   4 0 0 0 0

(1+5+4)/(1+5+3+4)

## [1] 0.7692308
```

Rating on Text Organization

```
raters_1_and_2_on_TxtOrg <- data.frame(r1=repeated$TxtOrg[repeated$Rater==1],
                                       r2=repeated$TxtOrg[repeated$Rater==2],
```

```
                                               a1=repeated$Artifact[repeated$Rater==1
],
                                               a2=repeated$Artifact[repeated$Rater==2
])

r1 <- factor(raters_1_and_2_on_TxtOrg$r1,levels=1:4)
r2 <- factor(raters_1_and_2_on_TxtOrg$r2,levels=1:4)

(t12 <- table(r1,r2))

##     r2
## r1   1 2 3 4
##    1 0 0 0 0
##    2 0 2 2 0
##    3 0 1 7 0
##    4 1 0 0 0

(2+7)/(2+2+1+7+1)

## [1] 0.6923077

raters_1_and_3_on_TxtOrg <- data.frame(r1=repeated$TxtOrg[repeated$Rater==1],
                                       r3=repeated$TxtOrg[repeated$Rater==3],
                                       a1=repeated$Artifact[repeated$Rater==1
],
                                       a3=repeated$Artifact[repeated$Rater==3
])

r1 <- factor(raters_1_and_3_on_TxtOrg$r1,levels=1:4)
r3 <- factor(raters_1_and_3_on_TxtOrg$r3,levels=1:4)

(t13 <- table(r1,r3))

##     r3
## r1   1 2 3 4
##    1 0 0 0 0
##    2 1 1 2 0
##    3 0 1 7 0
##    4 0 1 0 0

(1+7)/(1+1+2+1+7+1)

## [1] 0.6153846

raters_2_and_3_on_TxtOrg <- data.frame(r2=repeated$TxtOrg[repeated$Rater==2],
                                       r3=repeated$TxtOrg[repeated$Rater==3],
                                       a2=repeated$Artifact[repeated$Rater==2
],
                                       a3=repeated$Artifact[repeated$Rater==3
])

r2 <- factor(raters_2_and_3_on_TxtOrg$r2,levels=1:4)
r3 <- factor(raters_2_and_3_on_TxtOrg$r3,levels=1:4)
```

```r
(t23 <- table(r2,r3))

##    r3
## r2  1 2 3 4
##   1 0 1 0 0
##   2 1 0 2 0
##   3 0 2 7 0
##   4 0 0 0 0

(7)/(1+1+2+2+7)

## [1] 0.5384615

Rubric.names <- sort(unique(tall$Rubric))
ICC.vec <- NULL
for (i in Rubric.names) {
  tmp <- lmer(as.numeric(Rating) ~ 1 + (1|Artifact), data=tall_sub[tall_sub$R
ubric==i,])
  sig2 <- summary(tmp)$sigma^2
  tau2 <- attr(summary(tmp)$varcor[[1]],"stddev")^2
  ICC <- tau2 / (tau2 + sig2)
  ICC.vec <- c(ICC.vec,ICC)
}
names(ICC.vec) <- Rubric.names
agreement.results <- cbind(ICC.common=ICC.vec,"    a12"=0,a23=0,a13=0)
agreement.tables <- as.list(rep(NA,7))
names(agreement.tables) <- Rubric.names
for (i in Rubric.names) {
  r12 <- data.frame(r1=factor(ratings_sub[ratings_sub$Rater==1,i],levels=1:4)
,
                    r2=factor(ratings_sub[ratings_sub$Rater==2,i],levels=1:4)
,
                    a1=ratings_sub[ratings_sub$Rater==1,"Artifact"],
                    a2=ratings_sub[ratings_sub$Rater==2,"Artifact"])
  if(any(r12[,3]!=r12[,4])) { stop(paste("Rater 1-2 Artifact mismatch on rubr
ic",i)) }
  a12 <- mean(r12[,1]==r12[,2])
  r12 <- table(r12[,1:2]) ## print this to see how much agreement there is am
ong raters 1-2
  r23 <- data.frame(r2=factor(ratings_sub[ratings_sub$Rater==2,i],levels=1:4)
,
                    r3=factor(ratings_sub[ratings_sub$Rater==3,i],levels=1:4)
,
                    a2=ratings_sub[ratings_sub$Rater==2,"Artifact"],
                    a3=ratings_sub[ratings_sub$Rater==3,"Artifact"])
  if(any(r23[,3]!=r23[,4])) { stop(paste("Rater 2-3 Artifact mismatch on rubr
ic",i)) }
  a23 <- mean(r23[,1]==r23[,2])
  r23 <- table(r23[,1:2]) ## print this to see how much agreement there is am
ong raters 2-3
  r13 <- data.frame(r1=factor(ratings_sub[ratings_sub$Rater==1,i],levels=1:4)
```

```
,
                    r3=factor(ratings_sub[ratings_sub$Rater==3,i],levels=1:4)
,
                    a1=ratings_sub[ratings_sub$Rater==1,"Artifact"],
                    a3=ratings_sub[ratings_sub$Rater==3,"Artifact"])
  if(any(r13[,3]!=r13[,4])) { stop(paste("Rater 1-3 Artifact mismatch on rubr
ic",i)) }
  a13 <- mean(r13[,1]==r13[,2])
  r13 <- table(r13[,1:2]) ## print this to see how much agreement there is am
ong raters 1-3
  agreement.results[i,2:4] <- c(a12,a23,a13)

  agreement.tables[[i]] <- list(r12,r23,r13)
}
round(agreement.results,2)

##           ICC.common     a12  a23  a13
## CritDes         0.57    0.54 0.69 0.62
## InitEDA         0.49    0.69 0.85 0.54
## InterpRes       0.23    0.62 0.62 0.54
## RsrchQ          0.19    0.38 0.54 0.77
## SelMeth         0.52    0.92 0.69 0.62
## TxtOrg          0.14    0.69 0.54 0.62
## VisOrg          0.59    0.54 0.77 0.77
```

Working on full data

```
ratings$Rater <- as.factor(ratings$Rater)
tall$Artifact <- as.factor(tall$Artifact)
tall$Rater <- as.factor(tall$Rater)
tall$Sex <- as.factor(tall$Sex)
tall$Semester <- as.factor(tall$Semester)
tall$Rubric <- as.factor(tall$Rubric)
tall$Repeated <- as.factor(tall$Repeated)

Rubric.names <- sort(unique(tall$Rubric))
tall[c(161,684),] ## just to check that these are the rows with missing ratin
gs...

##       X Rater Artifact Repeated Semester Sex  Rubric Rating
## 161 161     2       45        0      S19   F CritDes   <NA>
## 684 684     1      100        0      F19   F  VisOrg   <NA>

tall.nonmissing <- tall[-c(161,684),] ## now delete them...
tall.nonmissing[tall.nonmissing$Sex=="--",]

##       X Rater Artifact Repeated Semester Sex    Rubric Rating
## 5     5     3        5        0      F19  --    RsrchQ      3
## 122 122     3        5        0      F19  --   CritDes      3
## 239 239     3        5        0      F19  --   InitEDA      3
## 356 356     3        5        0      F19  --   SelMeth      3
## 473 473     3        5        0      F19  -- InterpRes      3
```

```
## 590 590     3        5        0       F19  --     VisOrg       3
## 707 707     3        5        0       F19  --     TxtOrg       3
```

```
tall.nonmissing <- tall.nonmissing[tall.nonmissing$Sex!="--",] ## eliminate t
hem
model.formula.alldata <- as.list(rep(NA,7))
names(model.formula.alldata) <- Rubric.names
```

Ratings on Rating on Research Question

```
RsrchQ.ratings_full <- tall.nonmissing[tall.nonmissing$Rubric=="RsrchQ",]
RsrchQ_3 <- lmer(as.numeric(Rating) ~ 1 + (1|Artifact), data=RsrchQ.ratings_f
ull)
summary(RsrchQ_3)
```

```
## Linear mixed model fit by REML ['lmerMod']
## Formula: as.numeric(Rating) ~ 1 + (1 | Artifact)
##    Data: RsrchQ.ratings_full
## 
## REML criterion at convergence: 209.1
## 
## Scaled residuals: 
##     Min      1Q  Median      3Q     Max 
## -2.2694 -0.5285 -0.3736  0.9743  2.4770 
## 
## Random effects:
##  Groups   Name        Variance Std.Dev.
##  Artifact (Intercept) 0.07276  0.2697  
##  Residual             0.27825  0.5275  
## Number of obs: 116, groups:  Artifact, 90
## 
## Fixed effects:
##             Estimate Std. Error t value
## (Intercept)  2.35169    0.05794   40.59
```

```
RsrchQ_ICC_3 <- (0.07372)/(0.07372+0.27797)
RsrchQ_ICC_3
```

```
## [1] 0.2096164
```

```
RsrchQ_ICC_2
```

```
## [1] 0.1891918
```

Rating on Critique Design

```
CritDes.ratings_full <- tall.nonmissing[tall.nonmissing$Rubric=="CritDes",]
CritDes_3 <- lmer(as.numeric(Rating) ~ 1 + (1|Artifact), data=CritDes.ratings
_full)
summary(CritDes_3)
```

```
## Linear mixed model fit by REML ['lmerMod']
## Formula: as.numeric(Rating) ~ 1 + (1 | Artifact)
##    Data: CritDes.ratings_full
```

```
##
## REML criterion at convergence: 274.7
##
## Scaled residuals:
##      Min       1Q   Median       3Q      Max
## -2.00615 -0.60064  0.02999  0.67713  2.06614
##
## Random effects:
##  Groups   Name        Variance Std.Dev.
##  Artifact (Intercept) 0.4909   0.7006
##  Residual             0.2412   0.4911
## Number of obs: 115, groups:  Artifact, 89
##
## Fixed effects:
##             Estimate Std. Error t value
## (Intercept)  1.89533    0.08889   21.32
```

```
CritDes_ICC_3 <- (0.4963)/(0.4963+0.2411)
CritDes_ICC_3
```

```
## [1] 0.6730404
```

```
CritDes_ICC_2
```

```
## [1] 0.5725134
```

Rating on Initial EDA

```
InitEDA.ratings_full <- tall.nonmissing[tall.nonmissing$Rubric=="InitEDA",]
InitEDA_3 <- lmer(as.numeric(Rating) ~ 1 + (1|Artifact), data=InitEDA.ratings
_full)
summary(InitEDA_3)
```

```
## Linear mixed model fit by REML ['lmerMod']
## Formula: as.numeric(Rating) ~ 1 + (1 | Artifact)
##    Data: InitEDA.ratings_full
##
## REML criterion at convergence: 239
##
## Scaled residuals:
##     Min      1Q  Median      3Q     Max
## -1.8889 -0.3391 -0.1427  0.4276  1.6035
##
## Random effects:
##  Groups   Name        Variance Std.Dev.
##  Artifact (Intercept) 0.3651   0.6042
##  Residual             0.1655   0.4068
## Number of obs: 116, groups:  Artifact, 90
##
## Fixed effects:
##             Estimate Std. Error t value
## (Intercept)  2.44226    0.07537    32.4
```

```
InitEDA_ICC_3 <- (0.3628)/(0.3628+0.1655)
InitEDA_ICC_3

## [1] 0.686731

InitEDA_ICC_2

## [1] 0.4930784
```

Rating on Select Method(s)

```
SelMeth.ratings_full <- tall.nonmissing[tall.nonmissing$Rubric=="SelMeth",]
SelMeth_3 <- lmer(as.numeric(Rating) ~ 1 + (1|Artifact), data=SelMeth.ratings
_full)
summary(SelMeth_3)

## Linear mixed model fit by REML ['lmerMod']
## Formula: as.numeric(Rating) ~ 1 + (1 | Artifact)
##    Data: SelMeth.ratings_full
##
## REML criterion at convergence: 153.6
##
## Scaled residuals:
##      Min       1Q   Median       3Q      Max
## -2.21774 -0.09510 -0.09510 -0.04934  2.11906
##
## Random effects:
##  Groups   Name        Variance Std.Dev.
##  Artifact (Intercept) 0.1060   0.3256
##  Residual             0.1227   0.3502
## Number of obs: 116, groups:  Artifact, 90
##
## Fixed effects:
##             Estimate Std. Error t value
## (Intercept)   2.0621     0.0485   42.52

SelMeth_ICC_3 <- (0.1108)/(0.1108+0.1240)
SelMeth_ICC_3

## [1] 0.471891

SelMeth_ICC_2

## [1] 0.5212845
```

Rating on Interpret Results

```
InterpRes.ratings_full <- tall.nonmissing[tall.nonmissing$Rubric=="InterpRes"
,]
InterpRes_3 <- lmer(as.numeric(Rating) ~ 1 + (1|Artifact), data=InterpRes.rat
ings_full)
summary(InterpRes_3)
```

```
## Linear mixed model fit by REML ['lmerMod']
## Formula: as.numeric(Rating) ~ 1 + (1 | Artifact)
##    Data: InterpRes.ratings_full
##
## REML criterion at convergence: 216.3
##
## Scaled residuals:
##     Min      1Q  Median      3Q     Max
## -2.1331 -0.6911  0.5205  0.7508  2.6562
##
## Random effects:
##  Groups   Name        Variance Std.Dev.
##  Artifact (Intercept) 0.08286  0.2879
##  Residual             0.29166  0.5401
## Number of obs: 116, groups:  Artifact, 90
##
## Fixed effects:
##             Estimate Std. Error t value
## (Intercept)    2.479      0.060   41.32

InterpRes_ICC_3 <- (0.08219)/(0.08219+0.29136)
InterpRes_ICC_3

## [1] 0.2200241

InterpRes_ICC_2

## [1] 0.2295821
```

Rating on Visual Organization

```
VisOrg.ratings_full <- tall.nonmissing[tall.nonmissing$Rubric=="VisOrg",]
VisOrg_3 <- lmer(as.numeric(Rating) ~ 1 + (1|Artifact), data=VisOrg.ratings_f
ull)
summary(VisOrg_3)

## Linear mixed model fit by REML ['lmerMod']
## Formula: as.numeric(Rating) ~ 1 + (1 | Artifact)
##    Data: VisOrg.ratings_full
##
## REML criterion at convergence: 224.7
##
## Scaled residuals:
##     Min      1Q  Median      3Q     Max
## -1.5890 -0.3728 -0.1605  0.4763  1.6335
##
## Random effects:
##  Groups   Name        Variance Std.Dev.
##  Artifact (Intercept) 0.3106   0.5573
##  Residual             0.1589   0.3987
## Number of obs: 115, groups:  Artifact, 89
##
```

```
## Fixed effects:
##             Estimate Std. Error t value
## (Intercept)  2.43901    0.07114   34.29
```

```
VisOrg_ICC_3 <- (0.3092)/(0.3092+0.1588)
VisOrg_ICC_3
```

```
## [1] 0.6606838
```

```
VisOrg_ICC_2
```

```
## [1] 0.5924748
```

Rating on Text Organization

```
TxtOrg.ratings_full <- tall.nonmissing[tall.nonmissing$Rubric=="TxtOrg",]
TxtOrg_3 <- lmer(as.numeric(Rating) ~ 1 + (1|Artifact), data=TxtOrg.ratings_f
ull)
summary(TxtOrg_3)
```

```
## Linear mixed model fit by REML ['lmerMod']
## Formula: as.numeric(Rating) ~ 1 + (1 | Artifact)
##    Data: TxtOrg.ratings_full
##
## REML criterion at convergence: 247.5
##
## Scaled residuals:
##     Min      1Q  Median      3Q     Max
## -2.3557 -0.7550  0.3834  0.5302  2.4132
##
## Random effects:
##  Groups   Name        Variance Std.Dev.
##  Artifact (Intercept) 0.09371  0.3061
##  Residual             0.39573  0.6291
## Number of obs: 116, groups:  Artifact, 90
##
## Fixed effects:
##             Estimate Std. Error t value
## (Intercept)  2.58745    0.06821   37.93
```

```
TxtOrg_ICC_3 <- (0.09145)/(0.09145+0.39503)
TxtOrg_ICC_3
```

```
## [1] 0.1879831
```

```
TxtOrg_ICC_2
```

```
## [1] 0.1428682
```

```
if (F) { print(agreement.tables) }
ICC.vec <- NULL
for (i in Rubric.names) {
  tmp <- lmer(as.numeric(Rating) ~ 1 + (1|Artifact), data=tall[tall$Rubric==i
```

```
,])
  sig2 <- summary(tmp)$sigma^2
  tau2 <- attr(summary(tmp)$varcor[[1]],"stddev")^2
  ICC <- tau2 / (tau2 + sig2)
  ICC.vec <- c(ICC.vec,ICC)
}
names(ICC.vec) <- Rubric.names
agreement.results <- cbind(ICC.alldata=ICC.vec,agreement.results)
round(agreement.results,2)

##            ICC.alldata ICC.common     a12  a23  a13
## CritDes           0.67       0.57    0.54 0.69 0.62
## InitEDA           0.69       0.49    0.69 0.85 0.54
## InterpRes         0.22       0.23    0.62 0.62 0.54
## RsrchQ            0.21       0.19    0.38 0.54 0.77
## SelMeth           0.47       0.52    0.92 0.69 0.62
## TxtOrg            0.19       0.14    0.69 0.54 0.62
## VisOrg            0.66       0.59    0.54 0.77 0.77
```

Appendix C

Exploring the relationship of various factors in this experiment (Rater, Semester, Sex, Repeated, Rubric) and the ratings

Rating on Research Question Variable Selection

```
RsrchQ_3_1 <- update(RsrchQ_3, . ~ . + Sex)
anova(RsrchQ_3, RsrchQ_3_1)

## refitting model(s) with ML (instead of REML)

## Data: RsrchQ.ratings_full
## Models:
## RsrchQ_3: as.numeric(Rating) ~ 1 + (1 | Artifact)
## RsrchQ_3_1: as.numeric(Rating) ~ (1 | Artifact) + Sex
##             npar    AIC    BIC  logLik deviance  Chisq Df Pr(>Chisq)
## RsrchQ_3       3 211.21 219.47 -102.61   205.21
## RsrchQ_3_1     4 212.60 223.61 -102.30   204.60 0.6148  1      0.433

RsrchQ_3_2 <- update(RsrchQ_3, . ~ . + Semester)
anova(RsrchQ_3, RsrchQ_3_2)

## refitting model(s) with ML (instead of REML)

## Data: RsrchQ.ratings_full
## Models:
## RsrchQ_3: as.numeric(Rating) ~ 1 + (1 | Artifact)
## RsrchQ_3_2: as.numeric(Rating) ~ (1 | Artifact) + Semester
##             npar    AIC    BIC  logLik deviance Chisq Df Pr(>Chisq)
## RsrchQ_3       3 211.21 219.47 -102.61   205.21
## RsrchQ_3_2     4 212.47 223.48 -102.23   204.47 0.744  1     0.3884
```

```
RsrchQ_3_3 <- update(RsrchQ_3, . ~ . + Rater)
anova(RsrchQ_3, RsrchQ_3_3)

## refitting model(s) with ML (instead of REML)

## Data: RsrchQ.ratings_full
## Models:
## RsrchQ_3: as.numeric(Rating) ~ 1 + (1 | Artifact)
## RsrchQ_3_3: as.numeric(Rating) ~ (1 | Artifact) + Rater
##            npar    AIC    BIC  logLik deviance  Chisq Df Pr(>Chisq)
## RsrchQ_3      3 211.21 219.47 -102.61   205.21
## RsrchQ_3_3    5 213.04 226.81 -101.52   203.04 2.1767  2     0.3368

summary(RsrchQ_3)

## Linear mixed model fit by REML ['lmerMod']
## Formula: as.numeric(Rating) ~ 1 + (1 | Artifact)
##    Data: RsrchQ.ratings_full
##
## REML criterion at convergence: 209.1
##
## Scaled residuals:
##     Min      1Q  Median      3Q     Max
## -2.2694 -0.5285 -0.3736  0.9743  2.4770
##
## Random effects:
##  Groups   Name        Variance Std.Dev.
##  Artifact (Intercept) 0.07276  0.2697
##  Residual             0.27825  0.5275
## Number of obs: 116, groups:  Artifact, 90
##
## Fixed effects:
##             Estimate Std. Error t value
## (Intercept)  2.35169    0.05794   40.59
```

Rating on Critique Design Variable Selection

```
CritDes_3_1 <- update(CritDes_3, . ~ . + Sex)
anova(CritDes_3, CritDes_3_1)

## refitting model(s) with ML (instead of REML)

## Data: CritDes.ratings_full
## Models:
## CritDes_3: as.numeric(Rating) ~ 1 + (1 | Artifact)
## CritDes_3_1: as.numeric(Rating) ~ (1 | Artifact) + Sex
##             npar    AIC    BIC  logLik deviance Chisq Df Pr(>Chisq)
## CritDes_3      3 277.68 285.91 -135.84   271.68
## CritDes_3_1    4 279.15 290.13 -135.58   271.15 0.524  1     0.4692

CritDes_3_2 <- update(CritDes_3, . ~ . + Semester)
anova(CritDes_3, CritDes_3_2)
```

```
## refitting model(s) with ML (instead of REML)

## Data: CritDes.ratings_full
## Models:
## CritDes_3: as.numeric(Rating) ~ 1 + (1 | Artifact)
## CritDes_3_2: as.numeric(Rating) ~ (1 | Artifact) + Semester
##             npar    AIC    BIC  logLik deviance  Chisq Df Pr(>Chisq)
## CritDes_3      3 277.68 285.91 -135.84   271.68
## CritDes_3_2    4 279.48 290.46 -135.74   271.48 0.1942  1     0.6595

CritDes_3_3 <- update(CritDes_3, . ~ . + Rater)
anova(CritDes_3, CritDes_3_3)

## refitting model(s) with ML (instead of REML)

## Data: CritDes.ratings_full
## Models:
## CritDes_3: as.numeric(Rating) ~ 1 + (1 | Artifact)
## CritDes_3_3: as.numeric(Rating) ~ (1 | Artifact) + Rater
##             npar    AIC    BIC  logLik deviance  Chisq Df Pr(>Chisq)
## CritDes_3      3 277.68 285.91 -135.84   271.68
## CritDes_3_3    5 273.62 287.35 -131.81   263.62 8.0535  2    0.01783 *
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

CritDes_3_4 <- update(CritDes_3, . ~ . + Sex + Rater)
anova(CritDes_3, CritDes_3_3, CritDes_3_4)

## refitting model(s) with ML (instead of REML)

## Data: CritDes.ratings_full
## Models:
## CritDes_3: as.numeric(Rating) ~ 1 + (1 | Artifact)
## CritDes_3_3: as.numeric(Rating) ~ (1 | Artifact) + Rater
## CritDes_3_4: as.numeric(Rating) ~ (1 | Artifact) + Sex + Rater
##             npar    AIC    BIC  logLik deviance  Chisq Df Pr(>Chisq)
## CritDes_3      3 277.68 285.91 -135.84   271.68
## CritDes_3_3    5 273.62 287.35 -131.81   263.62 8.0535  2    0.01783 *
## CritDes_3_4    6 275.23 291.70 -131.61   263.23 0.3960  1    0.52919
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

CritDes_3_5 <- update(CritDes_3, . ~ . + Sex * Rater)
anova(CritDes_3, CritDes_3_3, CritDes_3_4, CritDes_3_5)

## refitting model(s) with ML (instead of REML)

## Data: CritDes.ratings_full
## Models:
## CritDes_3: as.numeric(Rating) ~ 1 + (1 | Artifact)
## CritDes_3_3: as.numeric(Rating) ~ (1 | Artifact) + Rater
## CritDes_3_4: as.numeric(Rating) ~ (1 | Artifact) + Sex + Rater
```

```
## CritDes_3_5: as.numeric(Rating) ~ (1 | Artifact) + Sex + Rater + Sex:Rater
##             npar    AIC    BIC  logLik deviance  Chisq Df Pr(>Chisq)
## CritDes_3      3 277.68 285.91 -135.84   271.68
## CritDes_3_3    5 273.62 287.35 -131.81   263.62 8.0535  2    0.01783 *
## CritDes_3_4    6 275.23 291.70 -131.61   263.23 0.3960  1    0.52919
## CritDes_3_5    8 278.63 300.59 -131.32   262.63 0.5957  2    0.74241
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

CritDes_3_6 <- update(CritDes_3, . ~ . + Semester + Rater)
anova(CritDes_3, CritDes_3_3, CritDes_3_6)

## refitting model(s) with ML (instead of REML)

## Data: CritDes.ratings_full
## Models:
## CritDes_3: as.numeric(Rating) ~ 1 + (1 | Artifact)
## CritDes_3_3: as.numeric(Rating) ~ (1 | Artifact) + Rater
## CritDes_3_6: as.numeric(Rating) ~ (1 | Artifact) + Semester + Rater
##             npar    AIC    BIC  logLik deviance  Chisq Df Pr(>Chisq)
## CritDes_3      3 277.68 285.91 -135.84   271.68
## CritDes_3_3    5 273.62 287.35 -131.81   263.62 8.0535  2    0.01783 *
## CritDes_3_6    6 275.49 291.96 -131.75   263.49 0.1325  1    0.71589
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

CritDes_3_7 <- update(CritDes_3, . ~ . + Semester * Rater)
anova(CritDes_3, CritDes_3_3, CritDes_3_6, CritDes_3_7)

## refitting model(s) with ML (instead of REML)

## Data: CritDes.ratings_full
## Models:
## CritDes_3: as.numeric(Rating) ~ 1 + (1 | Artifact)
## CritDes_3_3: as.numeric(Rating) ~ (1 | Artifact) + Rater
## CritDes_3_6: as.numeric(Rating) ~ (1 | Artifact) + Semester + Rater
## CritDes_3_7: as.numeric(Rating) ~ (1 | Artifact) + Semester + Rater + Seme
ster:Rater
##             npar    AIC    BIC  logLik deviance  Chisq Df Pr(>Chisq)
## CritDes_3      3 277.68 285.91 -135.84   271.68
## CritDes_3_3    5 273.62 287.35 -131.81   263.62 8.0535  2    0.01783 *
## CritDes_3_6    6 275.49 291.96 -131.75   263.49 0.1325  1    0.71589
## CritDes_3_7    8 277.32 299.28 -130.66   261.32 2.1711  2    0.33771
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

formula(CritDes_3_3)

## as.numeric(Rating) ~ (1 | Artifact) + Rater

summary(CritDes_3_3)
```

```
## Linear mixed model fit by REML ['lmerMod']
## Formula: as.numeric(Rating) ~ (1 | Artifact) + Rater
##    Data: CritDes.ratings_full
##
## REML criterion at convergence: 271
##
## Scaled residuals:
##      Min       1Q   Median       3Q      Max
## -1.55495 -0.50027 -0.08228  0.64663  1.60935
##
## Random effects:
##  Groups   Name        Variance Std.Dev.
##  Artifact (Intercept) 0.4349   0.6595
##  Residual             0.2473   0.4972
## Number of obs: 115, groups:  Artifact, 89
##
## Fixed effects:
##             Estimate Std. Error t value
## (Intercept)   1.6863     0.1207  13.976
## Rater2        0.4266     0.1491   2.861
## Rater3        0.2045     0.1491   1.371
##
## Correlation of Fixed Effects:
##        (Intr) Rater2
## Rater2 -0.610
## Rater3 -0.610  0.496
```

```
ranef(CritDes_3_3)
```

```
## $Artifact
##     (Intercept)
## 100  0.83753019
## 101 -0.43756481
## 102 -0.43756481
## 103  0.19998269
## 104 -0.43756481
## 105 -0.43756481
## 106  0.19998269
## 107 -0.43756481
## 111 -0.43756481
## 112 -0.43756481
## 113 -0.43756481
## 114 -0.43756481
## 115 -0.43756481
## 116 -0.43756481
## 117 -0.43756481
## 118 -0.43756481
## 13   0.06962462
## 15   0.70717212
## 16   0.70717212
```

```
## 17   0.06962462
## 21   0.70717212
## 22   0.70717212
## 23  -0.56792288
## 24   0.06962462
## 25   0.70717212
## 26  -0.56792288
## 27   0.06962462
## 28  -0.56792288
## 32   0.70717212
## 33   0.06962462
## 34   0.70717212
## 35  -0.56792288
## 36   0.06962462
## 37   0.70717212
## 38   0.06962462
## 39  -0.56792288
## 40   0.06962462
## 46  -0.07196897
## 47   0.56557853
## 48   0.56557853
## 49  -0.70951647
## 53   1.20312602
## 54  -0.70951647
## 55  -0.07196897
## 56   0.56557853
## 57  -0.70951647
## 6   -0.56792288
## 61  -0.07196897
## 62   1.20312602
## 63   0.56557853
## 64   0.56557853
## 65   0.56557853
## 66   0.56557853
## 67  -0.70951647
## 68   0.56557853
## 7   -0.56792288
## 72  -0.07196897
## 73  -0.70951647
## 74  -0.70951647
## 75  -0.07196897
## 76  -0.07196897
## 77  -0.07196897
## 78   0.56557853
## 79  -0.07196897
## 8   -0.56792288
## 84   0.19998269
## 85   0.83753019
## 86   0.19998269
## 87   0.19998269
```

```
## 88   0.83753019
## 9   -0.56792288
## 92  -0.43756481
## 93  -0.43756481
## 94   0.83753019
## 95   0.19998269
## 96   0.19998269
## O1  -0.47358754
## O10 -0.47358754
## O11 -0.75381650
## O12 -0.19335859
## O13  0.08687037
## O2  -0.47358754
## O3   0.36709933
## O4   0.08687037
## O5   0.92755724
## O6  -0.75381650
## O7   0.36709933
## O8   0.08687037
## O9  -0.75381650
##
## with conditional variances for "Artifact"
```

```
qt(p=.05/2, df=116-3, lower.tail=FALSE)
```

```
## [1] 1.98118
```

```
CritDes_ICC_4 <- (0.4401)/(0.4401+0.2475)
CritDes_ICC_4
```

```
## [1] 0.6400524
```

```
CritDes_ICC_3
```

```
## [1] 0.6730404
```

Rating on Initial EDA Variable Selection

```
InitEDA_3_1 <- update(InitEDA_3, . ~ . + Sex)
anova(InitEDA_3, InitEDA_3_1)
```

```
## refitting model(s) with ML (instead of REML)
```

```
## Data: InitEDA.ratings_full
## Models:
## InitEDA_3: as.numeric(Rating) ~ 1 + (1 | Artifact)
## InitEDA_3_1: as.numeric(Rating) ~ (1 | Artifact) + Sex
##             npar    AIC    BIC  logLik deviance  Chisq Df Pr(>Chisq)
## InitEDA_3      3 241.64 249.90 -117.82   235.64
## InitEDA_3_1    4 243.56 254.58 -117.78   235.56 0.0796  1     0.7779
```

```
InitEDA_3_2 <- update(InitEDA_3, . ~ . + Semester)
anova(InitEDA_3, InitEDA_3_2)
```

```
## refitting model(s) with ML (instead of REML)

## Data: InitEDA.ratings_full
## Models:
## InitEDA_3: as.numeric(Rating) ~ 1 + (1 | Artifact)
## InitEDA_3_2: as.numeric(Rating) ~ (1 | Artifact) + Semester
##             npar    AIC    BIC  logLik deviance  Chisq Df Pr(>Chisq)
## InitEDA_3      3 241.64 249.90 -117.82   235.64
## InitEDA_3_2    4 243.62 254.64 -117.81   235.62 0.0211  1     0.8846

InitEDA_3_3 <- update(InitEDA_3, . ~ . + Rater)
anova(InitEDA_3, InitEDA_3_3)

## refitting model(s) with ML (instead of REML)

## Data: InitEDA.ratings_full
## Models:
## InitEDA_3: as.numeric(Rating) ~ 1 + (1 | Artifact)
## InitEDA_3_3: as.numeric(Rating) ~ (1 | Artifact) + Rater
##             npar    AIC    BIC  logLik deviance  Chisq Df Pr(>Chisq)
## InitEDA_3      3 241.64 249.90 -117.82   235.64
## InitEDA_3_3    5 242.20 255.96 -116.10   232.20 3.4472  2     0.1784

summary(InitEDA_3)

## Linear mixed model fit by REML ['lmerMod']
## Formula: as.numeric(Rating) ~ 1 + (1 | Artifact)
##    Data: InitEDA.ratings_full
##
## REML criterion at convergence: 239
##
## Scaled residuals:
##     Min      1Q  Median      3Q     Max
## -1.8889 -0.3391 -0.1427  0.4276  1.6035
##
## Random effects:
##  Groups   Name        Variance Std.Dev.
##  Artifact (Intercept) 0.3651   0.6042
##  Residual             0.1655   0.4068
## Number of obs: 116, groups:  Artifact, 90
##
## Fixed effects:
##             Estimate Std. Error t value
## (Intercept)  2.44226    0.07537    32.4
```

Rating on Select Method(s) Variable Selection

```
SelMeth_3_1 <- update(SelMeth_3, . ~ . + Sex)
anova(SelMeth_3, SelMeth_3_1)

## refitting model(s) with ML (instead of REML)
```

```
## Data: SelMeth.ratings_full
## Models:
## SelMeth_3: as.numeric(Rating) ~ 1 + (1 | Artifact)
## SelMeth_3_1: as.numeric(Rating) ~ (1 | Artifact) + Sex
##             npar    AIC    BIC  logLik deviance  Chisq Df Pr(>Chisq)
## SelMeth_3      3 155.37 163.63 -74.687   149.37
## SelMeth_3_1    4 153.01 164.03 -72.505   145.01 4.3633  1    0.03672 *
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

SelMeth_3_2 <- update(SelMeth_3_1, . ~ . + Semester)
anova(SelMeth_3, SelMeth_3_1, SelMeth_3_2)

## refitting model(s) with ML (instead of REML)

## Data: SelMeth.ratings_full
## Models:
## SelMeth_3: as.numeric(Rating) ~ 1 + (1 | Artifact)
## SelMeth_3_1: as.numeric(Rating) ~ (1 | Artifact) + Sex
## SelMeth_3_2: as.numeric(Rating) ~ (1 | Artifact) + Sex + Semester
##             npar    AIC    BIC  logLik deviance  Chisq Df Pr(>Chisq)
## SelMeth_3      3 155.37 163.63 -74.687   149.37
## SelMeth_3_1    4 153.01 164.03 -72.505   145.01 4.3633  1   0.036720 *
## SelMeth_3_2    5 145.72 159.48 -67.858   135.72 9.2948  1   0.002298 **
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

SelMeth_3_3 <- update(SelMeth_3, . ~ . + Sex*Semester)
anova(SelMeth_3, SelMeth_3_2, SelMeth_3_3)

## refitting model(s) with ML (instead of REML)

## Data: SelMeth.ratings_full
## Models:
## SelMeth_3: as.numeric(Rating) ~ 1 + (1 | Artifact)
## SelMeth_3_2: as.numeric(Rating) ~ (1 | Artifact) + Sex + Semester
## SelMeth_3_3: as.numeric(Rating) ~ (1 | Artifact) + Sex + Semester + Sex:Se
mester
##             npar    AIC    BIC  logLik deviance   Chisq Df Pr(>Chisq)
## SelMeth_3      3 155.37 163.63 -74.687   149.37
## SelMeth_3_2    5 145.72 159.48 -67.858   135.72 13.6582  2   0.001082 **
## SelMeth_3_3    6 143.11 159.63 -65.553   131.11  4.6102  1   0.031783 *
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

SelMeth_3_4 <- update(SelMeth_3_2, . ~ . + Rater)
anova(SelMeth_3, SelMeth_3_2, SelMeth_3_3, SelMeth_3_4)

## refitting model(s) with ML (instead of REML)

## Data: SelMeth.ratings_full
## Models:
```

```
## SelMeth_3: as.numeric(Rating) ~ 1 + (1 | Artifact)
## SelMeth_3_2: as.numeric(Rating) ~ (1 | Artifact) + Sex + Semester
## SelMeth_3_3: as.numeric(Rating) ~ (1 | Artifact) + Sex + Semester + Sex:Se
mester
## SelMeth_3_4: as.numeric(Rating) ~ (1 | Artifact) + Sex + Semester + Rater
##             npar    AIC    BIC  logLik deviance   Chisq Df Pr(>Chisq)
## SelMeth_3      3 155.37 163.63 -74.687   149.37
## SelMeth_3_2    5 145.72 159.48 -67.858   135.72 13.6582  2   0.001082 **
## SelMeth_3_3    6 143.11 159.63 -65.553   131.11  4.6102  1   0.031783 *
## SelMeth_3_4    7 142.35 161.63 -64.178   128.35  2.7502  1   0.097244 .
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

SelMeth_3_5 <- update(SelMeth_3_4, . ~ . + Sex*Rater)
anova(SelMeth_3, SelMeth_3_4, SelMeth_3_2, SelMeth_3_5)

## refitting model(s) with ML (instead of REML)

## Data: SelMeth.ratings_full
## Models:
## SelMeth_3: as.numeric(Rating) ~ 1 + (1 | Artifact)
## SelMeth_3_2: as.numeric(Rating) ~ (1 | Artifact) + Sex + Semester
## SelMeth_3_4: as.numeric(Rating) ~ (1 | Artifact) + Sex + Semester + Rater
## SelMeth_3_5: as.numeric(Rating) ~ (1 | Artifact) + Sex + Semester + Rater
+ Sex:Rater
##             npar    AIC    BIC  logLik deviance   Chisq Df Pr(>Chisq)
## SelMeth_3      3 155.37 163.63 -74.687   149.37
## SelMeth_3_2    5 145.72 159.48 -67.858   135.72 13.6582  2   0.001082 **
## SelMeth_3_4    7 142.35 161.63 -64.178   128.35  7.3603  2   0.025219 *
## SelMeth_3_5    9 144.31 169.10 -63.157   126.31  2.0408  2   0.360453
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

SelMeth_3_6 <- update(SelMeth_3_4, . ~ . + Semester*Rater)
anova(SelMeth_3, SelMeth_3_2, SelMeth_3_4, SelMeth_3_6)

## refitting model(s) with ML (instead of REML)

## Data: SelMeth.ratings_full
## Models:
## SelMeth_3: as.numeric(Rating) ~ 1 + (1 | Artifact)
## SelMeth_3_2: as.numeric(Rating) ~ (1 | Artifact) + Sex + Semester
## SelMeth_3_4: as.numeric(Rating) ~ (1 | Artifact) + Sex + Semester + Rater
## SelMeth_3_6: as.numeric(Rating) ~ (1 | Artifact) + Sex + Semester + Rater
+ Semester:Rater
##             npar    AIC    BIC  logLik deviance   Chisq Df Pr(>Chisq)
## SelMeth_3      3 155.37 163.63 -74.687   149.37
## SelMeth_3_2    5 145.72 159.48 -67.858   135.72 13.6582  2   0.001082 **
## SelMeth_3_4    7 142.35 161.63 -64.178   128.35  7.3603  2   0.025219 *
## SelMeth_3_6    9 143.63 168.42 -62.817   125.63  2.7219  2   0.256412
```

```
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
```

```
summary(SelMeth_3_2)
```

```
## Linear mixed model fit by REML ['lmerMod']
## Formula: as.numeric(Rating) ~ (1 | Artifact) + Sex + Semester
##    Data: SelMeth.ratings_full
##
## REML criterion at convergence: 145.7
##
## Scaled residuals:
##      Min       1Q   Median       3Q      Max
## -2.45094 -0.36643 -0.09728  0.34827  2.15811
##
## Random effects:
##  Groups   Name        Variance Std.Dev.
##  Artifact (Intercept) 0.08971  0.2995
##  Residual             0.11560  0.3400
## Number of obs: 116, groups:  Artifact, 90
##
## Fixed effects:
##             Estimate Std. Error t value
## (Intercept)  2.11008    0.07603  27.754
## SexM         0.11119    0.09658   1.151
## SemesterS19 -0.32038    0.10430  -3.072
##
## Correlation of Fixed Effects:
##             (Intr) SexM
## SexM        -0.691
## SemesterS19 -0.587  0.299
```

```
qt(p=.05/2, df=117-7, lower.tail=FALSE)
```

```
## [1] 1.981765
```

```
SelMeth_ICC_4 <- (0.0850)/(0.0850+0.1057)
SelMeth_ICC_4
```

```
## [1] 0.4457263
```

```
SelMeth_ICC_3
```

```
## [1] 0.471891
```

Rating on Interpret Results Variable Selection

```
InterpRes_3_1 <- update(InterpRes_3, . ~ . + Sex)
anova(InterpRes_3, InterpRes_3_1)
```

```
## refitting model(s) with ML (instead of REML)
```

```
## Data: InterpRes.ratings_full
## Models:
## InterpRes_3: as.numeric(Rating) ~ 1 + (1 | Artifact)
## InterpRes_3_1: as.numeric(Rating) ~ (1 | Artifact) + Sex
##               npar    AIC    BIC  logLik deviance  Chisq Df Pr(>Chisq)
## InterpRes_3      3 218.53 226.79 -106.26   212.53
## InterpRes_3_1    4 220.31 231.32 -106.15   212.31 0.2209  1     0.6384
```

```
InterpRes_3_2 <- update(InterpRes_3, . ~ . + Semester)
anova(InterpRes_3, InterpRes_3_2)
```

```
## refitting model(s) with ML (instead of REML)

## Data: InterpRes.ratings_full
## Models:
## InterpRes_3: as.numeric(Rating) ~ 1 + (1 | Artifact)
## InterpRes_3_2: as.numeric(Rating) ~ (1 | Artifact) + Semester
##               npar    AIC    BIC  logLik deviance Chisq Df Pr(>Chisq)
## InterpRes_3      3 218.53 226.79 -106.26   212.53
## InterpRes_3_2    4 220.25 231.26 -106.12   212.25 0.279  1     0.5974
```

```
InterpRes_3_3 <- update(InterpRes_3, . ~ . + Rater)
anova(InterpRes_3, InterpRes_3_3)
```

```
## refitting model(s) with ML (instead of REML)

## Data: InterpRes.ratings_full
## Models:
## InterpRes_3: as.numeric(Rating) ~ 1 + (1 | Artifact)
## InterpRes_3_3: as.numeric(Rating) ~ (1 | Artifact) + Rater
##               npar    AIC    BIC   logLik deviance  Chisq Df Pr(>Chisq)

## InterpRes_3      3 218.53 226.79 -106.263   212.53

## InterpRes_3_3    5 200.66 214.43  -95.331   190.66 21.864  2  1.787e-05 **
*
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
```

```
summary(InterpRes_3_3)
```

```
## Linear mixed model fit by REML ['lmerMod']
## Formula: as.numeric(Rating) ~ (1 | Artifact) + Rater
##    Data: InterpRes.ratings_full
##
## REML criterion at convergence: 199.7
##
## Scaled residuals:
##     Min      1Q  Median      3Q     Max
## -2.5317 -0.7627  0.2635  0.6614  2.6535
##
## Random effects:
```

```
##  Groups   Name        Variance Std.Dev.
##  Artifact (Intercept) 0.06224  0.2495
##  Residual             0.25250  0.5025
## Number of obs: 116, groups:  Artifact, 90
##
## Fixed effects:
##             Estimate Std. Error t value
## (Intercept)  2.70421    0.08912  30.342
## Rater2      -0.11847    0.12213  -0.970
## Rater3      -0.56503    0.12287  -4.599
##
## Correlation of Fixed Effects:
##        (Intr) Rater2
## Rater2 -0.685
## Rater3 -0.680  0.497
```

```
qt(p=.05/2, df=117-3, lower.tail=FALSE)
```

```
## [1] 1.980992
```

```
InterpRes_ICC_4 <- (0.06471)/(0.06471+0.25381)
InterpRes_ICC_4
```

```
## [1] 0.2031584
```

```
InterpRes_ICC_3
```

```
## [1] 0.2200241
```

Rating on Visual Organization Variable Selection

```
VisOrg_3_1 <- update(VisOrg_3, . ~ . + Sex)
anova(VisOrg_3, VisOrg_3_1)
```

```
## refitting model(s) with ML (instead of REML)
```

```
## Data: VisOrg.ratings_full
## Models:
## VisOrg_3: as.numeric(Rating) ~ 1 + (1 | Artifact)
## VisOrg_3_1: as.numeric(Rating) ~ (1 | Artifact) + Sex
##            npar    AIC    BIC logLik deviance  Chisq Df Pr(>Chisq)
## VisOrg_3      3 227.21 235.44 -110.6   221.21
## VisOrg_3_1    4 228.41 239.39 -110.2   220.41 0.7969  1      0.372
```

```
VisOrg_3_2 <- update(VisOrg_3, . ~ . + Semester)
anova(VisOrg_3, VisOrg_3_2)
```

```
## refitting model(s) with ML (instead of REML)
```

```
## Data: VisOrg.ratings_full
## Models:
## VisOrg_3: as.numeric(Rating) ~ 1 + (1 | Artifact)
## VisOrg_3_2: as.numeric(Rating) ~ (1 | Artifact) + Semester
```

```
##            npar    AIC    BIC  logLik deviance  Chisq Df Pr(>Chisq)
## VisOrg_3      3 227.21 235.44 -110.60   221.21
## VisOrg_3_2    4 227.73 238.71 -109.86   219.73 1.4787  1      0.224
```

```
VisOrg_3_3 <- update(VisOrg_3, . ~ . + Rater)
anova(VisOrg_3, VisOrg_3_3)
```

```
## refitting model(s) with ML (instead of REML)
```

```
## Data: VisOrg.ratings_full
## Models:
## VisOrg_3: as.numeric(Rating) ~ 1 + (1 | Artifact)
## VisOrg_3_3: as.numeric(Rating) ~ (1 | Artifact) + Rater
##            npar    AIC    BIC  logLik deviance  Chisq Df Pr(>Chisq)
## VisOrg_3      3 227.21 235.44 -110.60   221.21
## VisOrg_3_3    5 220.82 234.54 -105.41   210.82 10.392  2   0.005539 **
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
```

```
summary(VisOrg_3_3)
```

```
## Linear mixed model fit by REML ['lmerMod']
## Formula: as.numeric(Rating) ~ (1 | Artifact) + Rater
##    Data: VisOrg.ratings_full
##
## REML criterion at convergence: 219.6
##
## Scaled residuals:
##     Min      1Q  Median      3Q     Max
## -1.5004 -0.3365 -0.2483  0.3841  1.8552
##
## Random effects:
##  Groups   Name        Variance Std.Dev.
##  Artifact (Intercept) 0.2907   0.5392
##  Residual             0.1467   0.3830
## Number of obs: 115, groups:  Artifact, 89
##
## Fixed effects:
##             Estimate Std. Error t value
## (Intercept)  2.37794    0.09658  24.620
## Rater2       0.27097    0.11669   2.322
## Rater3      -0.09440    0.11714  -0.806
##
## Correlation of Fixed Effects:
##        (Intr) Rater2
## Rater2 -0.612
## Rater3 -0.606  0.502
```

```
qt(p=.05/2, df=116-3, lower.tail=FALSE)
```

```
## [1] 1.98118
```

```
VisOrg_ICC_4 <- (0.2937)/(0.2937+0.1454)
VisOrg_ICC_4

## [1] 0.6688681

VisOrg_ICC_3

## [1] 0.6606838
```

Rating on Text Organization Variable Selection

```
TxtOrg_3_1 <- update(TxtOrg_3, . ~ . + Sex)
anova(TxtOrg_3, TxtOrg_3_1)

## refitting model(s) with ML (instead of REML)

## Data: TxtOrg.ratings_full
## Models:
## TxtOrg_3: as.numeric(Rating) ~ 1 + (1 | Artifact)
## TxtOrg_3_1: as.numeric(Rating) ~ (1 | Artifact) + Sex
##            npar    AIC    BIC  logLik deviance  Chisq Df Pr(>Chisq)
## TxtOrg_3      3 250.00 258.26 -122.00   244.00
## TxtOrg_3_1    4 251.88 262.90 -121.94   243.88 0.1098  1     0.7404

TxtOrg_3_1 <- update(TxtOrg_3, . ~ . + Semester)
anova(TxtOrg_3, TxtOrg_3_1)

## refitting model(s) with ML (instead of REML)

## Data: TxtOrg.ratings_full
## Models:
## TxtOrg_3: as.numeric(Rating) ~ 1 + (1 | Artifact)
## TxtOrg_3_1: as.numeric(Rating) ~ (1 | Artifact) + Semester
##            npar    AIC    BIC  logLik deviance  Chisq Df Pr(>Chisq)
## TxtOrg_3      3 250.00 258.26 -122.00   244.00
## TxtOrg_3_1    4 250.56 261.57 -121.28   242.56 1.4382  1     0.2304

TxtOrg_3_3 <- update(TxtOrg_3, . ~ . + Rater)
anova(TxtOrg_3, TxtOrg_3_3)

## refitting model(s) with ML (instead of REML)

## Data: TxtOrg.ratings_full
## Models:
## TxtOrg_3: as.numeric(Rating) ~ 1 + (1 | Artifact)
## TxtOrg_3_3: as.numeric(Rating) ~ (1 | Artifact) + Rater
##            npar    AIC    BIC  logLik deviance  Chisq Df Pr(>Chisq)
## TxtOrg_3      3 250.00 258.26 -122.00   244.00
## TxtOrg_3_3    5 249.09 262.86 -119.55   239.09 4.9021  2     0.0862 .
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

summary(TxtOrg_3)
```

```
## Linear mixed model fit by REML ['lmerMod']
## Formula: as.numeric(Rating) ~ 1 + (1 | Artifact)
##    Data: TxtOrg.ratings_full
##
## REML criterion at convergence: 247.5
##
## Scaled residuals:
##     Min      1Q  Median      3Q     Max
## -2.3557 -0.7550  0.3834  0.5302  2.4132
##
## Random effects:
##  Groups   Name        Variance Std.Dev.
##  Artifact (Intercept) 0.09371  0.3061
##  Residual             0.39573  0.6291
## Number of obs: 116, groups:  Artifact, 90
##
## Fixed effects:
##             Estimate Std. Error t value
## (Intercept)  2.58745    0.06821   37.93
```

Adding fixed effects to the seven rubric-specific models using all the data

```
for (i in Rubric.names) {
  ## fit each base model
  rubric.data <- tall.nonmissing[tall.nonmissing$Rubric==i,]
  tmp <- lmer(as.numeric(Rating) ~ -1 + as.factor(Rater) +
  Semester + Sex + (1|Artifact),
  data=rubric.data,REML=FALSE)
  ## do backwards elimination
  tmp.back_elim <- fitLMER.fnc(tmp,set.REML.FALSE = TRUE,log.file.name = FALS
E)
  ## check to see if the raters are significantly different from one another
  tmp.single_intercept <- update(tmp.back_elim, . ~ . + 1 - as.factor(Rater))
  pval <- anova(tmp.single_intercept,tmp.back_elim)$"Pr(>Chisq)"[2]
  ## choose the best model
  if (pval<=0.05) {
    tmp_final <- tmp.back_elim
  } else {
  tmp_final <- tmp.single_intercept
  }
  ## and add to list...
  model.formula.alldata[[i]] <- formula(tmp_final)
}

## see what "final models" we got...
model.formula.alldata
```

```
## $CritDes
## as.numeric(Rating) ~ as.factor(Rater) + (1 | Artifact) - 1
##
## $InitEDA
```

```
## as.numeric(Rating) ~ (1 | Artifact)
##
## $InterpRes
## as.numeric(Rating) ~ as.factor(Rater) + (1 | Artifact) - 1
##
## $RsrchQ
## as.numeric(Rating) ~ (1 | Artifact)
##
## $SelMeth
## as.numeric(Rating) ~ as.factor(Rater) + Semester + (1 | Artifact) -
##     1
##
## $TxtOrg
## as.numeric(Rating) ~ (1 | Artifact)
##
## $VisOrg
## as.numeric(Rating) ~ as.factor(Rater) + (1 | Artifact) - 1
```

Trying to add fixed effects, interactions, and new random effects to the "combined" model Rating ~ 1 + (0 + Rubric|Artifact), using all the data.

Examing interactions with Rubric

```
model.01 <- lmer(as.numeric(Rating) ~ (0 + Rubric | Artifact), data=tall.nonm
issing)

## boundary (singular) fit: see ?isSingular

summary(model.01)

## Linear mixed model fit by REML ['lmerMod']
## Formula: as.numeric(Rating) ~ (0 + Rubric | Artifact)
##    Data: tall.nonmissing
##
## REML criterion at convergence: 1471.7
##
## Scaled residuals:
##     Min      1Q  Median      3Q     Max
## -3.0218 -0.4940 -0.0753  0.5271  3.7759
##
## Random effects:
##  Groups   Name            Variance Std.Dev. Corr
##  Artifact RubricCritDes   0.64070  0.8004
##           RubricInitEDA   0.38288  0.6188   0.26
##           RubricInterpRes 0.25658  0.5065   0.00 0.79
##           RubricRsrchQ    0.17398  0.4171   0.38 0.50 0.74
##           RubricSelMeth   0.09619  0.3102   0.56 0.37 0.41 0.26
##           RubricTxtOrg    0.40425  0.6358   0.03 0.69 0.80 0.64 0.24
##           RubricVisOrg    0.31878  0.5646   0.17 0.78 0.76 0.60 0.29 0.79
##  Residual                 0.19477  0.4413
## Number of obs: 810, groups:  Artifact, 90
##
```

```
## Fixed effects:
##             Estimate Std. Error t value
## (Intercept)  2.23210    0.04013   55.63
## optimizer (nloptwrap) convergence code: 0 (OK)
## boundary (singular) fit: see ?isSingular
```

Adding all factors available in the dataset

```
model.full <- update(model.01, . ~ . + Rater + Semester + Sex + Repeated + Ru
bric)
summary(model.full)
```

```
## Linear mixed model fit by REML ['lmerMod']
## Formula: as.numeric(Rating) ~ (0 + Rubric | Artifact) + Rater + Semester +

##     Sex + Repeated + Rubric
##    Data: tall.nonmissing
##
## REML criterion at convergence: 1429.6
##
## Scaled residuals:
##     Min      1Q  Median      3Q     Max
## -3.1091 -0.5065 -0.0178  0.5242  3.7932
##
## Random effects:
##  Groups   Name              Variance Std.Dev. Corr
##  Artifact RubricCritDes     0.55311  0.7437
##           RubricInitEDA     0.35239  0.5936   0.47
##           RubricInterpRes   0.17512  0.4185   0.23 0.75
##           RubricRsrchQ      0.16997  0.4123   0.58 0.44 0.71
##           RubricSelMeth     0.06816  0.2611   0.39 0.60 0.74 0.41
##           RubricTxtOrg      0.26339  0.5132   0.34 0.62 0.70 0.56 0.67
##           RubricVisOrg      0.25809  0.5080   0.35 0.73 0.68 0.52 0.41 0.76
##  Residual                   0.18916  0.4349
## Number of obs: 810, groups:  Artifact, 90
##
## Fixed effects:
##                  Estimate Std. Error t value
## (Intercept)      2.013748   0.109103  18.457
## Rater2           0.001977   0.054887   0.036
## Rater3          -0.174867   0.055045  -3.177
## SemesterS19     -0.175017   0.087850  -1.992
## SexM             0.010506   0.081271   0.129
## Repeated1       -0.073586   0.098522  -0.747
## RubricInitEDA    0.547054   0.095710   5.716
## RubricInterpRes  0.587091   0.100893   5.819
## RubricRsrchQ     0.460875   0.087516   5.266
## RubricSelMeth    0.164863   0.094265   1.749
## RubricTxtOrg     0.692880   0.099523   6.962
## RubricVisOrg     0.530182   0.099136   5.348
##
```

```
## Correlation of Fixed Effects:
##              (Intr) Rater2 Rater3 SmsS19 SexM   Reptd1 RbIEDA RbrcIR RbrcRQ
## Rater2       -0.245
## Rater3       -0.237  0.499
## SemesterS19 -0.361  0.008  0.000
## SexM        -0.398 -0.026 -0.035  0.302
## Repeated1   -0.154  0.001 -0.003  0.079  0.009
## RubrcIntEDA -0.552 -0.001  0.000 -0.001  0.000  0.007
## RbrcIntrpRs -0.660 -0.001  0.000 -0.001  0.000 -0.009  0.734
## RubrcRsrchQ -0.626 -0.001  0.000 -0.001  0.000 -0.039  0.585  0.756
## RubricSlMth -0.689 -0.001  0.000 -0.001  0.000 -0.088  0.659  0.777  0.689
## RubrcTxtOrg -0.611 -0.001  0.000 -0.001  0.000  0.005  0.674  0.751  0.682
## RubricVsOrg -0.607 -0.001 -0.001 -0.002 -0.001 -0.021  0.715  0.745  0.668
##              RbrcSM RbrcTO
## Rater2
## Rater3
## SemesterS19
## SexM
## Repeated1
## RubrcIntEDA
## RbrcIntrpRs
## RubrcRsrchQ
## RubricSlMth
## RubrcTxtOrg  0.725
## RubricVsOrg  0.680  0.750
```

Using fnc to choose factor

```
model.back_elim <- fitLMER.fnc(model.full, log.file.name = FALSE)

## Warning in fitLMER.fnc(model.full, log.file.name = FALSE): Argument "ran.e
ffects" is empty, which means you will not be forward-fitting the random effe
ct structure of your model. You could just as well run function "bfFixefLMER_
F.fnc" or "bfFixefLMER_t.fnc".
## TRUE

## boundary (singular) fit: see ?isSingular
## boundary (singular) fit: see ?isSingular
## boundary (singular) fit: see ?isSingular

summary(model.back_elim)

## Linear mixed model fit by REML ['lmerMod']
## Formula: as.numeric(Rating) ~ (0 + Rubric | Artifact) + Rater + Semester +

##     Rubric
##    Data: tall.nonmissing
##
## REML criterion at convergence: 1424.1
##
## Scaled residuals:
##     Min      1Q  Median      3Q     Max
```

```
## -3.1200 -0.5125 -0.0173  0.5302  3.7752
##
## Random effects:
##  Groups   Name             Variance Std.Dev. Corr
##  Artifact RubricCritDes    0.55495  0.7449
##           RubricInitEDA    0.35064  0.5921   0.47
##           RubricInterpRes  0.16892  0.4110   0.23 0.75
##           RubricRsrchQ     0.16777  0.4096   0.59 0.44 0.70
##           RubricSelMeth    0.06499  0.2549   0.40 0.60 0.74 0.40
##           RubricTxtOrg     0.25615  0.5061   0.33 0.61 0.69 0.55 0.66
##           RubricVisOrg     0.25894  0.5089   0.35 0.73 0.68 0.52 0.41 0.75
##  Residual                  0.18934  0.4351
## Number of obs: 810, groups:  Artifact, 90
##
## Fixed effects:
##                    Estimate Std. Error t value
## (Intercept)       2.0084130  0.0987610  20.336
## Rater2            0.0003231  0.0547446   0.006
## Rater3           -0.1771062  0.0548892  -3.227
## SemesterS19      -0.1730357  0.0826927  -2.093
## RubricInitEDA     0.5474747  0.0957148   5.720
## RubricInterpRes   0.5864544  0.1008618   5.814
## RubricRsrchQ      0.4584082  0.0874179   5.244
## RubricSelMeth     0.1590770  0.0937771   1.696
## RubricTxtOrg      0.6930033  0.0995479   6.962
## RubricVisOrg      0.5289027  0.0990973   5.337
##
## Correlation of Fixed Effects:
##             (Intr) Rater2 Rater3 SmsS19 RbIEDA RbrcIR RbrcRQ RbrcSM RbrcTO
## Rater2      -0.281
## Rater3      -0.277  0.499
## SemesterS19 -0.264  0.017  0.011
## RubrcIntEDA -0.610 -0.001  0.000 -0.002
## RbrcIntrpRs -0.735 -0.001  0.000  0.000  0.734
## RubrcRsrchQ -0.701 -0.001  0.000  0.002  0.586  0.756
## RubricSlMth -0.782  0.000  0.000  0.006  0.662  0.779  0.688
## RubrcTxtOrg -0.679 -0.001  0.000 -0.001  0.674  0.751  0.682  0.728
## RubricVsOrg -0.675 -0.001 -0.001  0.000  0.715  0.745  0.667  0.681  0.750
## optimizer (nloptwrap) convergence code: 0 (OK)
## boundary (singular) fit: see ?isSingular
```

Considering all interactions among factors available

```
model.inter <- update(model.back_elim, . ~ . + Rater*Semester*Rubric)

## Warning in checkConv(attr(opt, "derivs"), opt$par, ctrl = control$checkCon
v, :
## Model failed to converge with max|grad| = 0.00431172 (tol = 0.002, compone
nt 1)
```

```r
ss <- getME(model.inter,c("theta","fixef"))
model.inter.u<- update(model.inter,start=ss,
                       control=lmerControl(optimizer="bobyqa",
                                           optCtrl=list(maxfun=2e5)))
```

```
## boundary (singular) fit: see ?isSingular
```

```r
summary(model.inter.u)
```

```
## Linear mixed model fit by REML ['lmerMod']
## Formula: as.numeric(Rating) ~ (0 + Rubric | Artifact) + Rater + Semester +

##     Rubric + Rater:Semester + Rater:Rubric + Semester:Rubric +
##     Rater:Semester:Rubric
##    Data: tall.nonmissing
## Control: lmerControl(optimizer = "bobyqa", optCtrl = list(maxfun = 2e+05))
##
## REML criterion at convergence: 1424.4
##
## Scaled residuals:
##     Min      1Q  Median      3Q     Max
## -2.9141 -0.5141 -0.0653  0.5023  3.6609
##
## Random effects:
##  Groups   Name            Variance Std.Dev. Corr
##  Artifact RubricCritDes   0.48550  0.6968
##           RubricInitEDA   0.35257  0.5938   0.42
##           RubricInterpRes 0.14619  0.3824   0.32 0.80
##           RubricRsrchQ    0.16444  0.4055   0.66 0.43 0.72
##           RubricSelMeth   0.06297  0.2509   0.45 0.64 0.78 0.49
##           RubricTxtOrg    0.25441  0.5044   0.44 0.65 0.67 0.60 0.62
##           RubricVisOrg    0.25527  0.5052   0.35 0.73 0.68 0.57 0.35 0.76
##  Residual                 0.18839  0.4340
## Number of obs: 810, groups:  Artifact, 90
##
## Fixed effects:
##                                   Estimate Std. Error t value
## (Intercept)                       1.739538   0.136568  12.738
## Rater2                            0.302995   0.155107   1.953
## Rater3                            0.237851   0.155863   1.526
## SemesterS19                      -0.129077   0.250318  -0.516
## RubricInitEDA                     0.765215   0.165241   4.631
## RubricInterpRes                   0.979228   0.162160   6.039
## RubricRsrchQ                      0.710427   0.147386   4.820
## RubricSelMeth                     0.462750   0.155274   2.980
## RubricTxtOrg                      1.011251   0.160899   6.285
## RubricVisOrg                      0.647869   0.166603   3.889
## Rater2:SemesterS19                0.268014   0.303883   0.882
## Rater3:SemesterS19               -0.072789   0.301026  -0.242
## Rater2:RubricInitEDA             -0.325018   0.204108  -1.592
```

```
## Rater3:RubricInitEDA                 -0.374190   0.205354  -1.822
## Rater2:RubricInterpRes               -0.469281   0.201051  -2.334
## Rater3:RubricInterpRes               -0.711515   0.202316  -3.517
## Rater2:RubricRsrchQ                  -0.447050   0.189326  -2.361
## Rater3:RubricRsrchQ                  -0.474411   0.190681  -2.488
## Rater2:RubricSelMeth                 -0.301450   0.193678  -1.556
## Rater3:RubricSelMeth                 -0.365656   0.194970  -1.875
## Rater2:RubricTxtOrg                  -0.449164   0.200927  -2.235
## Rater3:RubricTxtOrg                  -0.407754   0.202209  -2.016
## Rater2:RubricVisOrg                   0.009042   0.205059   0.044
## Rater3:RubricVisOrg                  -0.287443   0.206299  -1.393
## SemesterS19:RubricInitEDA            -0.050212   0.301475  -0.167
## SemesterS19:RubricInterpRes           0.127813   0.295706   0.432
## SemesterS19:RubricRsrchQ              0.133874   0.267750   0.500
## SemesterS19:RubricSelMeth            -0.089616   0.282837  -0.317
## SemesterS19:RubricTxtOrg              0.166097   0.293176   0.567
## SemesterS19:RubricVisOrg              0.146845   0.302496   0.485
## Rater2:SemesterS19:RubricInitEDA      0.020326   0.392376   0.052
## Rater3:SemesterS19:RubricInitEDA      0.252422   0.389961   0.647
## Rater2:SemesterS19:RubricInterpRes -0.266618   0.385390  -0.692
## Rater3:SemesterS19:RubricInterpRes -0.152392   0.383354  -0.398
## Rater2:SemesterS19:RubricRsrchQ     -0.217348   0.360414  -0.603
## Rater3:SemesterS19:RubricRsrchQ      0.354319   0.357388   0.991
## Rater2:SemesterS19:RubricSelMeth    -0.401035   0.370200  -1.083
## Rater3:SemesterS19:RubricSelMeth    -0.192670   0.367887  -0.524
## Rater2:SemesterS19:RubricTxtOrg     -0.542267   0.385011  -1.408
## Rater3:SemesterS19:RubricTxtOrg     -0.316395   0.382614  -0.827
## Rater2:SemesterS19:RubricVisOrg     -0.603626   0.392909  -1.536
## Rater3:SemesterS19:RubricVisOrg     -0.186749   0.390759  -0.478

##
## Correlation matrix not shown by default, as p = 42 > 12.
## Use print(x, correlation=TRUE)  or
##     vcov(x)        if you need it

## optimizer (bobyqa) convergence code: 0 (OK)
## boundary (singular) fit: see ?isSingular
```

Using fnc function to choose out final model

```
model.inter_elim <- fitLMER.fnc(model.inter.u, log.file.name = FALSE)

## Warning in fitLMER.fnc(model.inter.u, log.file.name = FALSE): Argument "ra
n.effects" is empty, which means you will not be forward-fitting the random e
ffect structure of your model. You could just as well run function "bfFixefLM
ER_F.fnc" or "bfFixefLMER_t.fnc".
## TRUE

## boundary (singular) fit: see ?isSingular
## boundary (singular) fit: see ?isSingular
## boundary (singular) fit: see ?isSingular
## boundary (singular) fit: see ?isSingular
```

Initial interaction model

```
formula(model.inter.u)

## as.numeric(Rating) ~ (0 + Rubric | Artifact) + Rater + Semester +
##     Rubric + Rater:Semester + Rater:Rubric + Semester:Rubric +
##     Rater:Semester:Rubric
```

Final interaction model

```
formula(model.inter_elim)

## as.numeric(Rating) ~ (0 + Rubric | Artifact) + Rater + Semester +
##     Rubric + Rater:Rubric
```

Final model without interaction

```
formula(model.back_elim)

## as.numeric(Rating) ~ (0 + Rubric | Artifact) + Rater + Semester +
##     Rubric

summary(model.inter.u)$varcor

##  Groups   Name            Std.Dev. Corr
##  Artifact RubricCritDes   0.69678
##           RubricInitEDA   0.59378  0.416
##           RubricInterpRes 0.38235  0.324 0.800
##           RubricRsrchQ    0.40551  0.655 0.430 0.723
##           RubricSelMeth   0.25094  0.446 0.639 0.784 0.488
##           RubricTxtOrg    0.50439  0.436 0.649 0.667 0.604 0.622
##           RubricVisOrg    0.50524  0.349 0.727 0.675 0.567 0.346 0.757
##  Residual                 0.43404

summary(model.inter_elim)$varcor

##  Groups   Name            Std.Dev. Corr
##  Artifact RubricCritDes   0.70956
##           RubricInitEDA   0.59565  0.445
##           RubricInterpRes 0.38977  0.354 0.815
##           RubricRsrchQ    0.42371  0.631 0.440 0.716
##           RubricSelMeth   0.25937  0.424 0.601 0.737 0.364
##           RubricTxtOrg    0.51058  0.417 0.637 0.675 0.547 0.636
##           RubricVisOrg    0.50489  0.339 0.715 0.677 0.512 0.376 0.772
##  Residual                 0.43034

summary(model.back_elim)$varcor

##  Groups   Name            Std.Dev. Corr
##  Artifact RubricCritDes   0.74495
##           RubricInitEDA   0.59215  0.467
##           RubricInterpRes 0.41100  0.230 0.749
##           RubricRsrchQ    0.40960  0.588 0.436 0.704
##           RubricSelMeth   0.25493  0.399 0.603 0.736 0.397
##           RubricTxtOrg    0.50612  0.335 0.614 0.691 0.551 0.656
```

```
##           RubricVisOrg    0.50886  0.350 0.731 0.679 0.516 0.414 0.752
##  Residual                0.43513
```

```
anova(model.back_elim, model.inter_elim, model.inter.u)
```

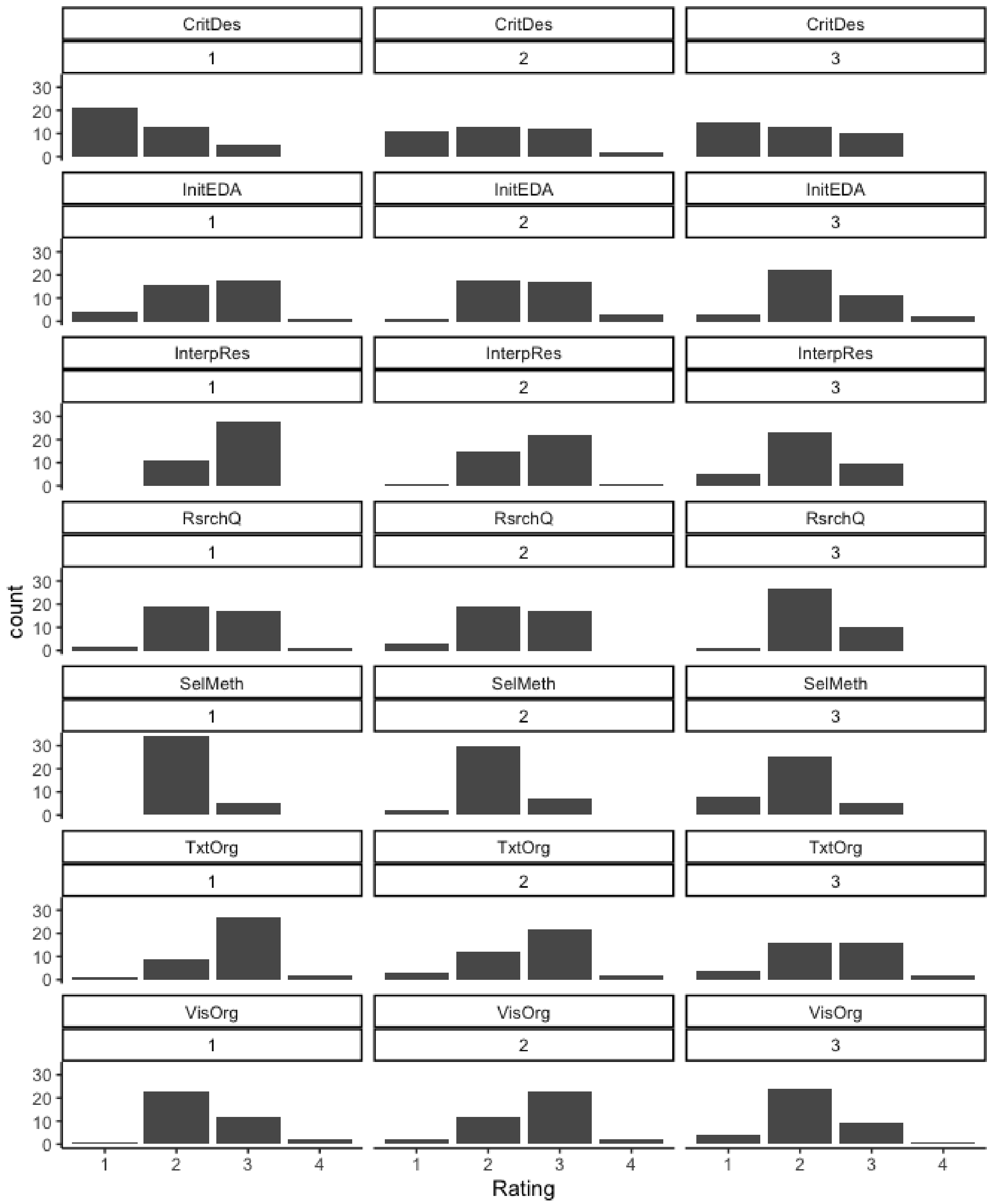
```
## refitting model(s) with ML (instead of REML)
```

```
## Data: tall.nonmissing
## Models:
## model.back_elim: as.numeric(Rating) ~ (0 + Rubric | Artifact) + Rater + Se
mester + Rubric
## model.inter_elim: as.numeric(Rating) ~ (0 + Rubric | Artifact) + Rater + S
emester + Rubric + Rater:Rubric
## model.inter.u: as.numeric(Rating) ~ (0 + Rubric | Artifact) + Rater + Seme
ster + Rubric + Rater:Semester + Rater:Rubric + Semester:Rubric + Rater:Semes
ter:Rubric
##                  npar    AIC    BIC  logLik deviance  Chisq Df Pr(>Chisq)

## model.back_elim    39 1464.0 1647.2 -693.02   1386.0

## model.inter_elim   51 1454.5 1694.1 -676.26   1352.5 33.526 12   0.000801
***
## model.inter.u      71 1471.4 1804.8 -664.68   1329.4 23.161 20   0.280962

## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
```

```
summary(model.inter_elim)
```

```
## Linear mixed model fit by REML ['lmerMod']
## Formula: as.numeric(Rating) ~ (0 + Rubric | Artifact) + Rater + Semester +

##     Rubric + Rater:Rubric
##    Data: tall.nonmissing
## Control: lmerControl(optimizer = "bobyqa", optCtrl = list(maxfun = 2e+05))
##
## REML criterion at convergence: 1419.6
##
## Scaled residuals:
##     Min      1Q  Median      3Q     Max
## -2.9280 -0.5122 -0.0447  0.4827  3.5854
##
## Random effects:
##  Groups   Name            Variance Std.Dev. Corr
##  Artifact RubricCritDes   0.50348  0.7096
##           RubricInitEDA   0.35480  0.5956   0.44
##           RubricInterpRes 0.15192  0.3898   0.35 0.82
##           RubricRsrchQ    0.17953  0.4237   0.63 0.44 0.72
##           RubricSelMeth   0.06727  0.2594   0.42 0.60 0.74 0.36
##           RubricTxtOrg    0.26069  0.5106   0.42 0.64 0.67 0.55 0.64
##           RubricVisOrg    0.25491  0.5049   0.34 0.71 0.68 0.51 0.38 0.77
```

```
##  Residual                 0.18519  0.4303
## Number of obs: 810, groups:  Artifact, 90
##
## Fixed effects:
##                          Estimate Std. Error t value
## (Intercept)               1.75945    0.11785  14.929
## Rater2                    0.36537    0.13296   2.748
## Rater3                    0.21421    0.13297   1.611
## SemesterS19              -0.17780    0.08228  -2.161
## RubricInitEDA             0.74625    0.13676   5.457
## RubricInterpRes           1.01453    0.13479   7.527
## RubricRsrchQ              0.74926    0.12419   6.033
## RubricSelMeth             0.42672    0.13040   3.272
## RubricTxtOrg              1.04967    0.13551   7.746
## RubricVisOrg              0.68354    0.13947   4.901
## Rater2:RubricInitEDA     -0.30843    0.17249  -1.788
## Rater3:RubricInitEDA     -0.29522    0.17282  -1.708
## Rater2:RubricInterpRes -0.53674    0.17008  -3.156
## Rater3:RubricInterpRes -0.75247    0.17049  -4.414
## Rater2:RubricRsrchQ      -0.50157    0.16151  -3.106
## Rater3:RubricRsrchQ      -0.37068    0.16179  -2.291
## Rater2:RubricSelMeth     -0.39602    0.16467  -2.405
## Rater3:RubricSelMeth     -0.41324    0.16504  -2.504
## Rater2:RubricTxtOrg      -0.58380    0.17141  -3.406
## Rater3:RubricTxtOrg      -0.48649    0.17177  -2.832
## Rater2:RubricVisOrg      -0.14444    0.17442  -0.828
## Rater3:RubricVisOrg      -0.33380    0.17481  -1.910

##
## Correlation matrix not shown by default, as p = 22 > 12.
## Use print(x, correlation=TRUE)  or
##     vcov(x)        if you need it

## optimizer (bobyqa) convergence code: 0 (OK)
## boundary (singular) fit: see ?isSingular

g <- ggplot(tall.nonmissing, aes(x=Rating)) +
  geom_bar() +
  facet_wrap( ~ Rubric + Rater, nrow=7) + theme_classic()

g
```




```
comb.final <- update(model.inter_elim, . ~ . - 1)

## boundary (singular) fit: see ?isSingular

summary(comb.final)$coef

##                            Estimate Std. Error    t value
## Rater1                    1.7594506 0.11785119 14.9294260
## Rater2                    2.1248217 0.11862140 17.9126339
## Rater3                    1.9736615 0.11875721 16.6192987
```

```
## SemesterS19              -0.1777990 0.08227894 -2.1609296
## RubricInitEDA             0.7462464 0.13675930  5.4566408
## RubricInterpRes           1.0145336 0.13478610  7.5269899
## RubricRsrchQ              0.7492624 0.12419095  6.0331480
## RubricSelMeth             0.4267187 0.13039726  3.2724519
## RubricTxtOrg              1.0496708 0.13551079  7.7460307
## RubricVisOrg              0.6835366 0.13947304  4.9008509
## Rater2:RubricInitEDA     -0.3084280 0.17249388 -1.7880516
## Rater3:RubricInitEDA     -0.2952155 0.17282356 -1.7081903
## Rater2:RubricInterpRes -0.5367420 0.17007959 -3.1558283
## Rater3:RubricInterpRes -0.7524696 0.17048675 -4.4136545
## Rater2:RubricRsrchQ      -0.5015692 0.16150820 -3.1055341
## Rater3:RubricRsrchQ      -0.3706755 0.16179323 -2.2910446
## Rater2:RubricSelMeth     -0.3960248 0.16466832 -2.4049850
## Rater3:RubricSelMeth     -0.4132363 0.16503758 -2.5038922
## Rater2:RubricTxtOrg      -0.5838002 0.17140824 -3.4059053
## Rater3:RubricTxtOrg      -0.4864857 0.17177159 -2.8321659
## Rater2:RubricVisOrg      -0.1444389 0.17442082 -0.8281058
## Rater3:RubricVisOrg      -0.3338016 0.17480762 -1.9095370
```

Random Effects

The fixed-effects terms we have to work with are: Rater, Semester, Rater:Rubric

```
m0 <- model.inter_elim
mA <- lmer(as.numeric(Rating) ~ (0 + Rubric | Artifact) +
             (0 + Rater | Artifact) + Rater +
             Semester + Rubric + Rater:Rubric, data=tall.nonmissing)

## Warning in checkConv(attr(opt, "derivs"), opt$par, ctrl = control$checkCon
v, :
## Model failed to converge with max|grad| = 0.00347545 (tol = 0.002, compone
nt 1)

ss_new <- getME(mA, c("theta","fixef"))
mA.u <- update(mA, start=ss_new,
             control=lmerControl(optimizer="bobyqa",
                                 optCtrl=list(maxfun=2e5)))

## boundary (singular) fit: see ?isSingular

anova(m0,mA.u)

## refitting model(s) with ML (instead of REML)

## Warning in commonArgs(par, fn, control, environment()): maxfun < 10 *
## length(par)^2 is not recommended.

## Data: tall.nonmissing
## Models:
## m0: as.numeric(Rating) ~ (0 + Rubric | Artifact) + Rater + Semester + Rubr
ic + Rater:Rubric
```

```
## mA.u: as.numeric(Rating) ~ (0 + Rubric | Artifact) + (0 + Rater | Artifact
) + Rater + Semester + Rubric + Rater:Rubric
##      npar    AIC    BIC  logLik deviance  Chisq Df Pr(>Chisq)
## m0     51 1454.5 1694.1 -676.26   1352.5
## mA.u   57 1415.9 1683.6 -650.94   1301.9 50.647  6  3.487e-09 ***
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

m0 <- model.inter_elim
mA.1 <- lmer(as.numeric(Rating) ~ (0 + Rubric | Artifact) +
             (0 + Semester | Artifact) + Rater +
             Semester + Rubric + Rater:Rubric, data=tall.nonmissing)

## Warning in checkConv(attr(opt, "derivs"), opt$par, ctrl = control$checkCon
v, :
## unable to evaluate scaled gradient

## Warning in checkConv(attr(opt, "derivs"), opt$par, ctrl = control$checkCon
v, :
## Model failed to converge: degenerate Hessian with 1 negative eigenvalues

ss_1 <- getME(mA.1, c("theta","fixef"))
mA_1.u<- update(mA.1, start=ss_1,
             control=lmerControl(optimizer="bobyqa",
                                 optCtrl=list(maxfun=2e5)))

## boundary (singular) fit: see ?isSingular

anova(m0,mA_1.u)

## refitting model(s) with ML (instead of REML)

## Data: tall.nonmissing
## Models:
## m0: as.numeric(Rating) ~ (0 + Rubric | Artifact) + Rater + Semester + Rubr
ic + Rater:Rubric
## mA_1.u: as.numeric(Rating) ~ (0 + Rubric | Artifact) + (0 + Semester | Art
ifact) + Rater + Semester + Rubric + Rater:Rubric
##        npar    AIC    BIC  logLik deviance  Chisq Df Pr(>Chisq)
## m0       51 1454.5 1694.1 -676.26   1352.5
## mA_1.u   54 1458.4 1712.0 -675.18   1350.4 2.1534  3     0.5412

m0 <- model.inter_elim
#mA.2 <- lmer(as.numeric(Rating) ~ (0 + Rubric | Artifact) +
             #(0 + Rater | Artifact) +
             #(0 + Rater:Rubric | Artifact) + Rater +
             #Semester + Rubric + Rater:Rubric, data=tall.nonmissing)
#ss_2 <- getME(mA.2, c("theta","fixef"))
#mA_2.u<- update(mA.2, start=ss_2,
             #control=lmerControl(optimizer="bobyqa",
                                 #optCtrl=list(maxfun=2e5)))
#anova(m0,mA_2.u)
```

```r
m.final <- mA.u
formula(m.final)
```

```
## as.numeric(Rating) ~ (0 + Rubric | Artifact) + (0 + Rater | Artifact) +
##     Rater + Semester + Rubric + Rater:Rubric
```

```r
summary(m.final)$varcor
```

```
##  Groups     Name            Std.Dev. Corr

##  Artifact   RubricCritDes   0.70459

##             RubricInitEDA   0.56379   0.318

##             RubricInterpRes 0.31947   0.142  0.674

##             RubricRsrchQ    0.42309   0.500  0.194  0.538

##             RubricSelMeth   0.19556   0.145  0.226  0.376 -0.240

##             RubricTxtOrg    0.50025   0.268  0.437  0.364  0.305  0.213

##             RubricVisOrg    0.48200   0.175  0.504  0.445  0.276 -0.161  0
.537
##  Artifact.1 Rater1          0.11319

##             Rater2          0.33428  -0.486

##             Rater3          0.30681   0.332  0.663

##  Residual                   0.36700
```

```r
summary(m.final)$coef
```

```
##                          Estimate Std. Error    t value
## (Intercept)             1.7575529 0.11404064 15.4116374
## Rater2                  0.3660533 0.13918297  2.6300146
## Rater3                  0.1959108 0.12966717  1.5108747
## SemesterS19            -0.1591787 0.07647482 -2.0814529
## RubricInitEDA           0.7394943 0.12995953  5.6901889
## RubricInterpRes         0.9915204 0.12770698  7.7640272
## RubricRsrchQ            0.7261884 0.11793010  6.1577864
## RubricSelMeth           0.4106835 0.12470431  3.2932583
## RubricTxtOrg            1.0157846 0.12999569  7.8139862
## RubricVisOrg            0.6542579 0.13352924  4.8997349
## Rater2:RubricInitEDA   -0.2998068 0.15609073 -1.9207213
## Rater3:RubricInitEDA   -0.2947320 0.15635198 -1.8850546
## Rater2:RubricInterpRes -0.5132359 0.15348542 -3.3438739
## Rater3:RubricInterpRes -0.7148521 0.15364022 -4.6527671
## Rater2:RubricRsrchQ    -0.4874141 0.14722216 -3.3107390
## Rater3:RubricRsrchQ    -0.3223831 0.14726586 -2.1891228
```

```
## Rater2:RubricSelMeth   -0.3863783 0.15030982 -2.5705457
## Rater3:RubricSelMeth   -0.3871620 0.14961502 -2.5877213
## Rater2:RubricTxtOrg    -0.5510472 0.15646169 -3.5219304
## Rater3:RubricTxtOrg    -0.4448976 0.15673248 -2.8385794
## Rater2:RubricVisOrg    -0.1049082 0.15861058 -0.6614199
## Rater3:RubricVisOrg    -0.2752251 0.15884843 -1.7326269
```

```
#ranef(m.final)
```

Random Effects Interpretation

```
ranef(m.final)
```

```
## $Artifact
##     RubricCritDes RubricInitEDA RubricInterpRes RubricRsrchQ RubricSelMeth
## 100   0.799228673  -0.260988369    -0.121676228 -0.165034190   0.232354784
## 101  -0.496490764   0.434193612    -0.166210415 -0.741422230   0.032250292
## 102  -0.770664167  -0.332739283    -0.232342287 -0.744261730   0.048828636
## 103   0.139648095   0.330202549     0.098797859 -0.281574593   0.271532454
## 104  -0.576655609   0.308512109     0.091185912 -0.260833483   0.073411542
## 105  -0.590082567  -0.482706425    -0.340028946 -0.404191148  -0.097451509
## 106   0.204314871  -1.028870286    -0.399151748  0.233396176  -0.136302675
## 107  -0.559956940  -0.401607193     0.057199570  0.183121789  -0.102318462
## 111  -0.461607128  -0.351068551    -0.241457556 -0.311652955  -0.066934237
## 112  -0.499271908   0.322584692     0.271608864  0.197897696  -0.103855394
## 113  -0.586350275  -0.804675420    -0.145322135 -0.131181956   0.004207552
## 114  -0.448180170   0.440149982     0.189757302 -0.168295289   0.103928815
## 115  -0.370796470   0.454222566     0.370180254  0.290435890  -0.073338121
## 116  -0.588697641  -0.354290917    -0.287550594 -0.351711998  -0.123701161
## 117  -0.672325649  -0.136978546     0.011776569 -0.194758247  -0.104392128
## 118  -0.654140682  -0.212121408    -0.029484230 -0.208944243  -0.027846653
## 13    0.388197379  -0.746411759    -0.434665448  0.050616864  -0.158680799
## 15    0.688612229   0.476177747    -0.046951155 -0.110359616  -0.052077440
## 16    0.682576507   1.153679071     0.314118154 -0.008304766   0.186388494
## 17    0.335745272  -0.087909834     0.067055556  0.549516509  -0.199114447
## 21    0.775973601   1.090383767     0.451614191  0.434676277   0.085151211
## 22    0.730598872   0.381879048     0.248951560  0.430928063   0.090324859
## 23   -0.272033080  -0.723565319    -0.326997170 -0.010010469  -0.176362310
## 24    0.049022533  -0.202394477    -0.150306090 -0.130272458   0.068968855
## 25    1.160394027   0.007477134    -0.288579520  0.162956858  -0.130355206
## 26   -0.863342534  -0.484026614    -0.160029490 -0.518461324   0.112869618
## 27    0.101603791   0.386497668     0.006469889 -0.172815148   0.092085620
## 28   -0.298838715  -0.592039401    -0.413401631 -0.405027066  -0.059307297
## 32    0.660692132   0.404027001     0.250529389  0.407859940   0.001493709
## 33   -0.083939358   0.150249705    -0.059830213 -0.452412235   0.169485623
## 34    0.465668830  -0.311735141    -0.060528981 -0.201142929   0.261663110
## 35   -0.743814878  -0.254864707    -0.085368806 -0.283326943  -0.098084544
## 36    0.049022533  -0.202394477    -0.150306090 -0.130272458   0.068968855
## 37    0.775844451  -0.157010302    -0.206882791 -0.021680678   0.102468095
## 38   -0.016992960  -0.209480532    -0.141945507 -0.174735163  -0.064585303
## 39   -0.527929785   0.248584181     0.207325381  0.140179534  -0.087373911
```

```
## 40    0.105495038   0.357263660    0.013252642 -0.194209730   0.047362612
## 45    0.014601341  -0.241826696   -0.172301424 -0.089848602   0.049753233
## 46    0.149535107   0.324246478   -0.013518911 -0.141163939   0.032064865
## 47    1.130946855  -0.177600615   -0.049034191  0.677203770  -0.148552857
## 48    0.536953397   0.660646690    0.224713323  0.134798535   0.241963084
## 49   -0.903829078   0.215286173    0.273265528  0.159422600  -0.189826594
## 53    1.141796395   0.106587242    0.017126292  0.239034193   0.117988021
## 54   -0.662488872   0.383775402   -0.091947931 -0.662636684   0.147396816
## 55   -0.148659507  -0.372297306    0.070588895  0.317366869  -0.133642140
## 56    0.687800358  -0.264808495   -0.275984234 -0.060685484  -0.062077360
## 57   -0.769733405  -0.326341484   -0.169612570 -0.256473512  -0.084352519
## 6    -0.673908138  -0.277012659   -0.086946635 -0.260258819  -0.009253394
## 61    0.001518443   0.332803900   -0.079464547 -0.172098289   0.015974580
## 62    1.385258649   0.997751003    0.303308032  0.483020215  -0.052900537
## 63    0.682894456   0.348634536    0.207003063  0.445176456  -0.018991243
## 64    0.550876762  -0.162658004   -0.076557789  0.054793675   0.062445404
## 65    0.831989858  -0.452063533   -0.411592337  0.253172622  -0.217001863
## 66    0.741068501   0.910036095    0.371867282  0.382464767  -0.040193312
## 67   -0.816114529   0.144942293    0.292945018  0.146922587  -0.188060107
## 68    0.630996043  -0.239557956    0.050750990  0.488210985  -0.041945810
## 7    -0.621326880   0.311879486    0.069829344 -0.302801510   0.013863372
## 72   -0.056276144   0.416368054    0.192064480 -0.072216244   0.022351234
## 73   -0.746468532  -0.931295991   -0.323409444 -0.186584678  -0.017227328
## 74   -1.048632338   0.085996840    0.117171519 -0.493823693   0.092898704
## 75   -0.041414974   0.337819580    0.148253534 -0.088796302   0.098107195
## 76    0.037334578  -0.325732471   -0.216037130 -0.141563773  -0.005213499
## 77   -0.168476673  -0.233610896   -0.010453738 -0.072616078  -0.014927129
## 78    0.416167329   0.062425282    0.074618197  0.131301600   0.084772124
## 79   -0.309461737   0.018263322    0.132110101  0.023569006   0.051556226
## 8    -0.607316678  -0.208782045   -0.035855058 -0.212294372   0.006564336
## 84    0.308462408  -0.084143664    0.160677746  0.445184878  -0.061763649
## 85    1.073414727   0.376334712    0.315627886  0.876534368  -0.132002472
## 86    0.326647375  -0.159286525    0.119416947  0.430998881   0.014781826
## 87    0.204314871  -1.028870286   -0.399151748  0.233396176  -0.136302675
## 88    1.088136531   0.644185725    0.316297923  0.538713376  -0.077309214
## 9    -0.580511043  -0.340307963    0.050549402  0.182722224  -0.110490677
## 92   -0.540387048  -0.348334547    0.068417123  0.221414943  -0.052022638
## 93   -0.392341717  -0.163424028    0.178206065  0.352246290   0.028790458
## 94    1.037146165   1.033181368    0.338396676  0.394285942  -0.006568291
## 95    0.139648095   0.330202549    0.098797859 -0.281574593   0.271532454
## 96    0.239287075   0.379994841    0.224037299  0.314943441  -0.067568202
## O1   -0.501833140   0.422473459    0.120695991 -0.008570911  -0.092442214
## O10  -0.441631392  -0.001275801    0.155483871  0.049611438   0.036656231
## O11  -0.767153991  -0.330031584    0.328442232  0.512571448   0.013750224
## O12  -0.354433725  -0.488344262   -0.316182189 -0.338361615  -0.015411893
## O13   0.030661437   0.083195491   -0.316854087 -0.367370039  -0.071076508
## O2   -0.115974416   0.329802975    0.171336271  0.140197430  -0.072093513
## O3    0.283478568  -0.135245342   -0.013010805  0.231069958  -0.039724358
## O4   -0.111622024   0.044733113    0.203146398 -0.216878169   0.365020082
## O5    0.878786098  -0.426299951   -0.026132105  0.189673112   0.249959395
```

```
## O6   -0.897741426  -0.773066212    -0.419223258 -0.412716902  -0.112091336
## O7    0.137952287   0.206767701    -0.005695597 -0.013643972  -0.042674507
## O8    0.358945188  -0.209375016    -0.396308076 -0.311053040   0.029112249
## O9   -0.799409045   0.585053489     0.349322460 -0.190683108   0.074742585
##       RubricTxtOrg RubricVisOrg       Rater1       Rater2       Rater3
## 100 -0.4896647007  -0.44078110  0.034931681 -0.050181185  0.031420290
## 101  0.2892550790   0.46518535 -0.031025012  0.044569050 -0.027906326
## 102 -1.1193729140  -0.43426182 -0.071721529  0.103031722 -0.064511961
## 103  0.1090938838  -0.23982673  0.041761517 -0.059992600  0.037563579
## 104  0.0889396145  -0.11667184 -0.025621764  0.036807003 -0.023046222
## 105 -0.5321515025  -0.35607711 -0.059280009  0.085158829 -0.053321084
## 106  0.0172387877  -0.33466186 -0.022656653  0.032547465 -0.020379168
## 107 -0.5131583247  -0.30986085 -0.011090647  0.015932293 -0.009975796
## 111 -0.4110644686  -0.25279967 -0.035973311  0.051677539 -0.032357213
## 112  0.2083928112   0.41589210  0.019755134 -0.028379282  0.017769315
## 113 -0.4728809942  -0.30644317 -0.016211878  0.023289210 -0.014582233
## 114  0.2100266485  -0.01339440 -0.002315065  0.003325712 -0.002082351
## 115  0.3294798452   0.51916954  0.043061833 -0.061860572  0.038733186
## 116  0.1298215184   0.27758401 -0.030964676  0.044482374 -0.027852055
## 117  0.2258047162   0.84082836  0.010806620 -0.015524274  0.009720321
## 118  0.1275831078   0.31603234 -0.008672796  0.012458924 -0.007800992
## 13  -0.7207934746  -0.62611376 -0.065711508 -0.387578549 -0.536731220
## 15   0.4109028996   0.90082171  0.013940054  0.082221001  0.113862282
## 16   0.8982898139   0.58516719  0.020550481  0.121210512  0.167856209
## 17  -0.1152731275   0.03069094 -0.017997889 -0.106154856 -0.147006653
## 21   0.9174747130   0.59118946  0.043490361  0.256514138  0.355229016
## 22   0.2103534096  -0.12225601  0.018243494  0.107603483  0.149012758
## 23  -0.6709548949  -0.56002192 -0.056906376 -0.335644262 -0.464810952
## 24   0.2457853922  -0.13974498 -0.007463872 -0.044023288 -0.060964863
## 25  -0.0007488536   0.07477069 -0.038403769 -0.226512487 -0.313681765
## 26  -0.4701734729  -0.47160845  0.015710223  0.092661783  0.128321012
## 27   0.2990310484  -0.05305422 -0.006284131 -0.037064961 -0.051328749
## 28  -0.6273778968  -0.51251337 -0.068980251 -0.406858196 -0.563430294
## 32   0.2598276764   0.36096246  0.027325151  0.161168760  0.223191675
## 33  -0.4040044992  -0.39747871  0.018950836  0.111775516  0.154790323
## 34  -0.5023507961  -0.50586271  0.030224847  0.178271700  0.246876374
## 35  -0.2593031418   0.26602910 -0.004562553 -0.026910775 -0.037266905
## 36   0.2457853922  -0.13974498 -0.007463872 -0.044023288 -0.060964863
## 37   0.2587087098  -0.15230600 -0.005402999 -0.031867881 -0.044131665
## 38  -0.2463798242   0.25346808 -0.002501680 -0.014755369 -0.020433708
## 39  -0.2363468532  -0.12447635  0.010475411  0.061785900  0.085563098
## 40  -0.2426084348  -0.14305964 -0.010403596 -0.061362320 -0.084976511
## 45   0.2993372465  -0.21571483  0.013970237 -0.084812494 -0.051593070
## 46  -0.1862047454  -0.21909187  0.017906533 -0.108709516 -0.066130088
## 47  -0.6836739937  -0.67088428  0.060716457 -0.368606058 -0.224230148
## 48   0.6088799375  -0.35393914 -0.057927422  0.351673992  0.213930047
## 49   0.2597371161   0.72493327 -0.028117122  0.170697404  0.103838511
## 53   0.0732047156  -0.06269668 -0.066378356  0.402979117  0.245139941
## 54   0.3347177409  -0.11283191  0.048106191 -0.292049871 -0.177659549
## 55  -0.3649781373   0.05818993 -0.010299283  0.062526344  0.038035977
```

```
## 56  -0.2393219894   0.11176427  0.019212814 -0.116639869 -0.070954274
## 57   0.2764721031   0.22689279  0.019216113 -0.116659897 -0.070966457
## 6   -0.3087774086  -0.21718938 -0.013644209 -0.080476053 -0.111445822
## 61   0.8802654695   0.38760340  0.008571567 -0.052037481 -0.031655400
## 62   0.4045706333   0.81899037 -0.054479976  0.330744747  0.201198385
## 63   0.2956494724   0.26737555 -0.036658837  0.222553658  0.135383667
## 64   0.2363767211   0.18588506 -0.001775727  0.010780336  0.006557886
## 65   0.3136001697   0.16069448  0.010823870 -0.065711082 -0.039973314
## 66  -0.1911406680   0.26542869 -0.032407248  0.196742509  0.119682249
## 67  -0.8635673421   0.06976299 -0.003067147  0.018620471  0.011327190
## 68   0.2430473917   0.18123371 -0.034952632  0.212195387  0.129082532
## 7   -0.2555317524  -0.13049862 -0.012464469 -0.073517726 -0.101809708
## 72  -0.2053782403   0.24593314 -0.010258600  0.062279362  0.037885733
## 73   0.1793500066  -0.33824195  0.034056705 -0.206756261 -0.125773807
## 74  -0.4514211259  -0.05709925 -0.034017026  0.206515373  0.125627270
## 75  -0.3067324995  -0.28153477  0.018590795 -0.112863629 -0.068657115
## 76  -0.2956410695  -0.35370872  0.035327747 -0.214472686 -0.130467856
## 77  -0.3148145645   0.11131629  0.007162614 -0.043483808 -0.026452036
## 78   0.0614056512  -0.04917676 -0.063397810  0.384884393  0.234132573
## 79   0.0496065868  -0.03565685 -0.060417263  0.366789669  0.223125204
## 8   -0.2460155113  -0.16365856 -0.002778204 -0.016386360 -0.022692358
## 84   0.2938871345   0.42395176  0.044951790 -0.064575594  0.040433161
## 85   0.2775952083   0.41740812  0.054493427 -0.078282653  0.049015656
## 86   0.1956655261  -0.10084426  0.025472374 -0.036592396  0.022911849
## 87   0.0172387877  -0.33466186 -0.022656653  0.032547465 -0.020379168
## 88   0.4756722623   1.03771365  0.071382059 -0.102544056  0.064206615
## 9   -0.2895925094  -0.21116711  0.009295671  0.054827573  0.075926984
## 92   0.0505930878  -0.20099574 -0.002254730  0.003239037 -0.002028080
## 93   0.7354315343   0.01114681  0.029887886 -0.042935510  0.026883506
## 94   0.3159417350   0.50173202  0.031126392 -0.044714688  0.027997515
## 95   0.1090938838  -0.23982673  0.041761517 -0.059992600  0.037563579
## 96   0.2323782156   0.41276615  0.024175539 -0.034729425  0.021745374
## O1  -0.2365686011  -0.25844094 -0.212424307  0.271858297 -0.224070730
## O10  0.1214243684   0.01135396  0.108528474 -0.367903064 -0.112463271
## O11  0.3056934427  -0.26212810  0.052727240  0.052398829  0.174414349
## O12 -0.3628175318  -0.45355834  0.058312625 -0.016895575  0.118720745
## O13  0.2746761740   0.14798982 -0.008187209  0.048581018  0.029123030
## O2   0.1740473426   0.34223594  0.172186297 -1.028585706 -0.619300082
## O3   0.2431181829   0.13200223 -0.038549903  0.159659019  0.068663798
## O4   0.1612094788  -0.27831324 -0.089506832  0.418551218  0.206842594
## O5  -0.0424756266  -0.48867178  0.030694448  0.059839562  0.130604388
## O6  -0.0545801763  -0.18317293 -0.050150491  0.116032753 -0.001517470
## O7   0.1229856572   0.22421027  0.147797222  0.139029998  0.481116754
## O8  -0.5684887464  -0.90875574 -0.039474840 -0.248709854 -0.338166891
## O9   0.3976623625   0.55917406  0.048179883  0.035297611  0.146903740
##
## with conditional variances for "Artifact"
```

Interpretation

```
## Rater 1
## the highest value + beta on rater 1
0.17 + 1.76

## [1] 1.93

## the lowest value + beta on rater 1
-0.21 + 1.76

## [1] 1.55

## Rater 2
## the highest value + beta on rater 2
0.42 + 1.76 + 0.37

## [1] 2.55

## the lowest value + beta on rater 2
-1.03 + 1.76 + 0.37

## [1] 1.1

## Rater 3
## the highest value + beta on rater 3
0.48 + 1.76 + 0.20

## [1] 2.44

## the lowest value + beta on rater 3
-0.62 + 1.76 + 0.20

## [1] 1.34

## Rubrics: Critique Design
## the highest value + beta on critique design
1.39 + 1.76 #62

## [1] 3.15

## the lowest value + beta on critique design
-1.05 + 1.76 #74

## [1] 0.71

## Rubrics: Initial EDA
## the highest value + beta on initial EDA
1.15 + 1.76 + 0.74 #16

## [1] 3.65

## the lowest value + beta on initial EDA
-1.03 + 1.76 + 0.74 #87
```

```
## [1] 1.47

## Rubrics: Interpret Results
## the highest value + beta on Interpret Results
0.45 + 1.76 + 0.99 #21

## [1] 3.2

## the lowest value + beta on Interpret Results
-0.43 + 1.76 + 0.99 #13

## [1] 2.32

## Rubrics: Research Question
## the highest value + beta on Research Question
0.88 + 1.76 + 0.73 #88

## [1] 3.37

## the lowest value + beta on Research Question
-0.74 + 1.76 + 0.73 #102

## [1] 1.75

## Rubrics: Select Method(s)
## the highest value + beta on Select Method(s)
0.37 + 1.76 + 0.41 #04

## [1] 2.54

## the lowest value + beta on Select Method(s)
-0.22 + 1.76 + 0.41 #65

## [1] 1.95

## Rubrics: Text Organization
## the highest value + beta on Text Organization
0.92 + 1.76 + 1.02 #21

## [1] 3.7

## the lowest value + beta on Text Organization
-1.12 + 1.76 + 1.02 #102

## [1] 1.66

## Rubrics: Visual Organization
## the highest value + beta on Visual Organization
1.04 + 1.76 + 0.65 #88

## [1] 3.45

## the lowest value + beta on Visual Organization
-0.91 + 1.76 + 0.65 #08
```

```
## [1] 1.5
```

Appendix D

Can we use data available to predict the rater?

```
library(nnet)
new_tall <- tall
new_tall <- new_tall[complete.cases(new_tall), ]
full.model <- multinom(Rater ~ Semester + Rubric + as.numeric(Rating) + Sex,
data = new_tall)

## # weights:  36 (22 variable)
## initial  value 897.566240
## iter  10 value 878.661695
## iter  20 value 876.276137
## iter  30 value 876.074000
## final  value 876.070045
## converged

step.model <- full.model %>% stepAIC(trace = FALSE)

## # weights:  33 (20 variable)
## initial  value 897.566240
## iter  10 value 879.348472
## iter  20 value 876.410067
## iter  30 value 876.185576
## final  value 876.184091
## converged
## # weights:  18 (10 variable)
## initial  value 897.566240
## iter  10 value 878.871235
## iter  20 value 877.485042
## final  value 877.483498
## converged
## # weights:  33 (20 variable)
## initial  value 897.566240
## iter  10 value 890.028630
## iter  20 value 888.509167
## iter  30 value 888.505180
## final  value 888.505128
## converged
## # weights:  30 (18 variable)
## initial  value 897.566240
## iter  10 value 886.903560
## iter  20 value 886.474937
## final  value 886.474647
## converged
## # weights:  18 (10 variable)
## initial  value 897.566240
## iter  10 value 878.871235
```

```
## iter  20 value 877.485042
## final  value 877.483498
## converged
## # weights:  15 (8 variable)
## initial  value 897.566240
## iter  10 value 878.007639
## final  value 877.581811
## converged
## # weights:  15 (8 variable)
## initial  value 897.566240
## iter  10 value 888.655863
## final  value 888.522335
## converged
## # weights:  12 (6 variable)
## initial  value 897.566240
## iter  10 value 887.673841
## final  value 887.673829
## converged
## # weights:  15 (8 variable)
## initial  value 897.566240
## iter  10 value 878.007639
## final  value 877.581811
## converged
## # weights:  12 (6 variable)
## initial  value 897.566240
## iter  10 value 888.687950
## final  value 888.631020
## converged
## # weights:  9 (4 variable)
## initial  value 897.566240
## final  value 888.160682
## converged
```

```
coef(step.model)
```

```
##   (Intercept) as.numeric(Rating)      SexF      SexM
## 2   -3.275218          0.1689930  2.779994  2.988891
## 3    9.608078         -0.4025093 -8.835171 -8.583841
```

```
library(stargazer)
```

```
##
## Please cite as:
```

```
##  Hlavac, Marek (2018). stargazer: Well-Formatted Regression and Summary St
atistics Tables.
```

```
##  R package version 5.2.2. https://CRAN.R-project.org/package=stargazer
```

```
stargazer(step.model, type="text")
```

```
##
## =================================================
##                        Dependent variable:
##                     ----------------------------
##                           2             3
##                          (1)           (2)
## -------------------------------------------------
## as.numeric(Rating)     0.169        -0.403***
##                       (0.124)        (0.125)
##
## SexF                   2.780         -8.835
##                      (103.752)      (25.218)
##
## SexM                   2.989         -8.584
##                      (103.752)      (25.218)
##
## Constant               -3.275         9.608
##                      (103.752)      (25.221)
##
## -------------------------------------------------
## Akaike Inf. Crit.    1,771.164      1,771.164
## =================================================
## Note:                *p<0.1; **p<0.05; ***p<0.01
```

```
step.model.ratio = exp(coef(step.model))
stargazer(step.model, type="text", coef=list(step.model.ratio), p.auto=FALSE)
```

```
##
## =================================================
##                        Dependent variable:
##                     ----------------------------
##                           2             3
##                          (1)           (2)
## -------------------------------------------------
## as.numeric(Rating)     1.184         0.669***
##                       (0.124)        (0.125)
##
## SexF                   16.119         0.0001
##                      (103.752)      (25.218)
##
## SexM                   19.864         0.0002
##                      (103.752)      (25.218)
##
## Constant               0.038       14,884.530
##                      (103.752)      (25.221)
##
## -------------------------------------------------
## Akaike Inf. Crit.    1,771.164      1,771.164
## =================================================
## Note:                *p<0.1; **p<0.05; ***p<0.01
```