Analysis on rating work in Freshman Statistics

Yucheng Wang
Department of Statistics and Data Science,
Carnegie Mellon University
yw6@andrew.cmu.edu

Abstract

We addressed the problem of what factors could have the largest effect on the ratings for each artifact in a school and learned how Ratings related to other variables associated with fairness. We examined rating data for 91 artifacts taken from Dietrich College at Carnegie Mellon University. From the bar plots for the ratings of different rubrics and raters, it appeared that the ratings of different raters and rubrics are sometimes significantly different. Then, by fitting Linear Mixed-Effects Models, we checked the intraclass correlation, and confirmed such difference for some rubrics. Then we used anova table to select the fixed effects for the model of each rubric and found that most of the ratings of some rubrics are not related to raters, but some of them are related to raters or the different semesters, which showed a unfair pattern in these rubrics. In the end, we confirmed that biases entailed by sex are not exist in this experiment.

1 Introduction

The fairness of the ratings from different raters and different rubrics are always an important topics, only the ratings rated fairly could be useful in evaluating the works. Now, Dietrich College at Carnegie Mellon University is in the process of implementing a new "General Education "program for undergraduates. This program specifies a set of courses and experiences that all undergraduates must take, and in order to determine whether the new program is successful, the college hopes to rate student work performed in each of the "Gen Ed" courses each year. Recently the college has been experimenting with rating work in Freshman Statistics, using raters from across the college. We want to know the fairness in this rating process at CMU. In the next sections, we will answer the following four questions by statistical approach. The statistical method in this paper was referred to [Sheather(2009)]

1, Is the distribution of ratings for each rubrics pretty much indistinguishable from the other rubrics, or are there rubrics that tend to get especially high or low rating? Is the distribution of ratings given by each rater pretty much indistinguishable from the other raters, or are there raters that tend to give especially high or low ratings?

2, For each rubric, do the raters generally agree on their scores? If not, is there one rater who disagrees with the others? Or do they all disagree?

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 1: Rubrics for rating Freshman Statistics projects

3, More generally, how are the various factors in this experiement (Rater, Semester, Sex, Repeated, Rubric) related to the ratings? Do the factors interact in any interesting ways?

4, Is the ratings for each rubric related to Sex?

2 Data

In a recent experiment, 91 project papers—referred to as "artifacts"—were randomly sampled from a Fall and Spring section of Freshman Statistics. Three raters from three different departments were asked to rate these artifacts on seven rubrics, as shown in Table 1. The rating scale for all rubrics is shown 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 rated by only rater. The variables available for analysis are defined in Table 3.

3 Method

In this paper, we used different methods to solve the question mentioned in the introduction part.

1, We used Numerical summaries and graphs to find the visualize the data.

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.

Table 2: Rating scale used for all rubrics

Variable Name	Values	Description
(X)	$1, 2, 3, \ldots$	Row number in the data set
Rater	1,2 or 3	Which of the three raters gave a rating
(Sample)	$1, 2, 3, \ldots$	Sample number
(Overlap)	$1, 2, \ldots, 13$	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

Table 3: Variables in the Dataset

2, We evaluated the 13 ”common” artifacts that all 3 raters saw by comparing their intraclass correlation (ICC) by fitting random-intercept model for each rubrics.

3, To solve this question, we evaluated different random-intercept models by adding fixed effects for Rater, Semester, Sex and/or Repeated to the random intercept models using the full data set as well as using the dataset with 13 artifacts only. And we tried interactions and new random effects for the seven rubric specific models using all the data. In the end, we tried to add fixed effects, interactions, and new random effects to the “combined” model using all the data. By building this model, we found the relationship between ratings and the various factors.

4, To solve this problem, we visualized the data with respect to sex and found the difference between the Ratings of Male and Female.

4 Result

4.1 Question 1

By visualizing the ratings for each rubric of the 13 artifacts which was rated by all three raters(figure 1), we find that the distribution of the rubrics InitEDA, RsrchQ, VisOrg, SelMeth have the similar distribution with rating 2 as the highest frequency rating. The rubrics TxtOrg, InterpRes are similar with rating 3 as the highest frequency rating. Also, these rubrics stated above all have very low number of rating 1 and 4. However, rubric CritDes is significantly different to other rubric, it has rating 1 as the highest frequency rating, the distribution of the ratings in this rubric shows that it might be a totally different rubric comparing with other rubrics. Thus, we find that not all the distributions of the rubrics are identical, some of them are not the same, especially for the rubric CritDes, its distribution is very different from other rubrics. The rubric CritDes seems tend to give especially low rating to the artifacts.

According to figure 2, we find that the distribution of ratings for each rubric in the full dataset are similar to the distribution in figure 2.

Also, for the ratings distribution for different raters, we find that in the 13 artifacts data, the distribution for the three raters are similar. However, the rater 3 tend to give the greatest number of rating 2 and greatest number of rating 1.

According to figure 4, we find that the distributions are similar to the distributions in figure 3, however, rater 1 and rater 2 are more tend to have similar number of rating 2 and 3. And for rater 3, it seems he/she gave more rating 2(in percentage). But we can find that the distribution of the ratings for each raters are not that distinguishable, they all not tend to give especially high or low score, however, rater 3 was really likely to give the rating 2.

4.2 Question 2

In this part we fitted the random-intercept model for each rubrics, and compared the ICC for each rubrics. First, we examined the 13 ”common” artifacts, and then we examined the full dataset.

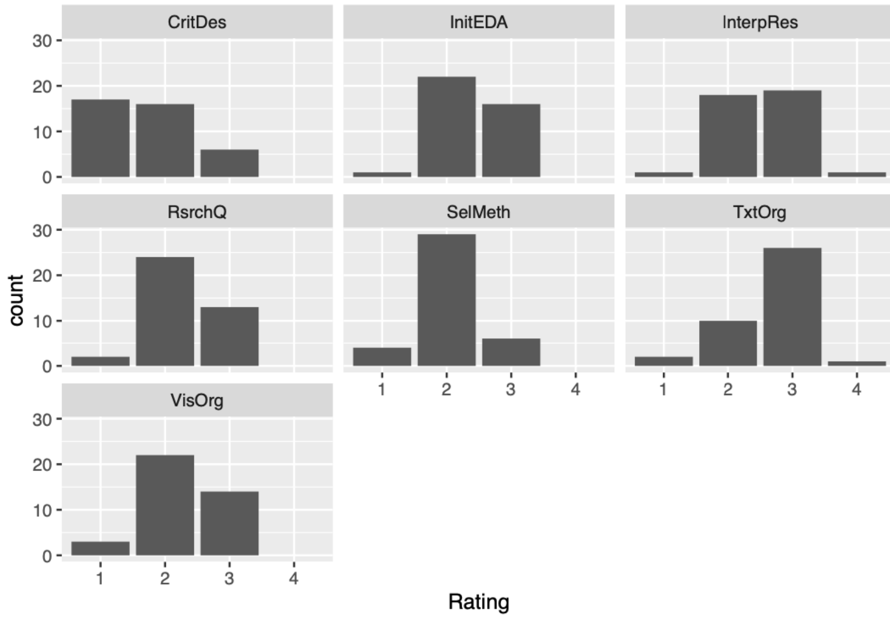


Figure 1: Distribution of ratings for each rubric for the 13 artifacts

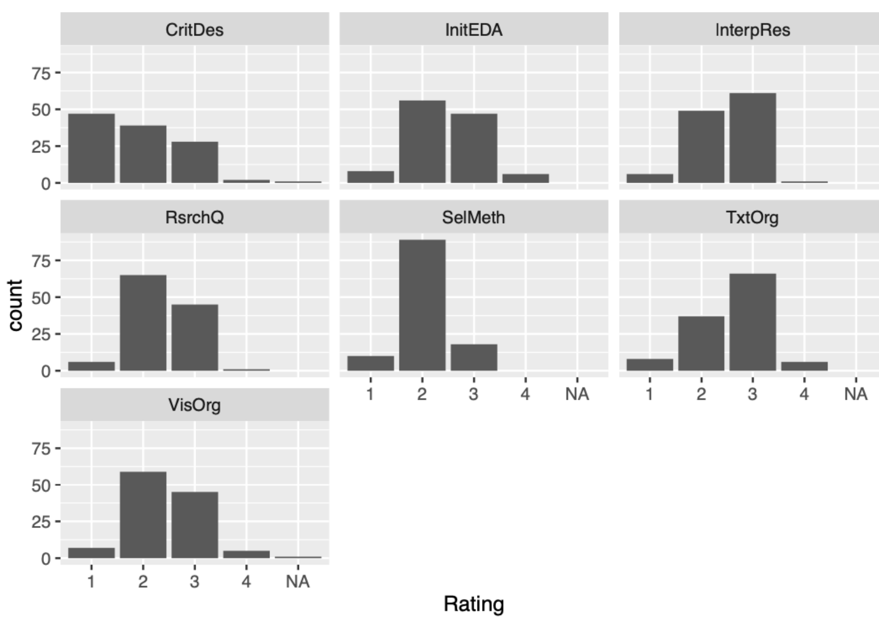


Figure 2: Distribution of ratings for each rubric for the full dataset

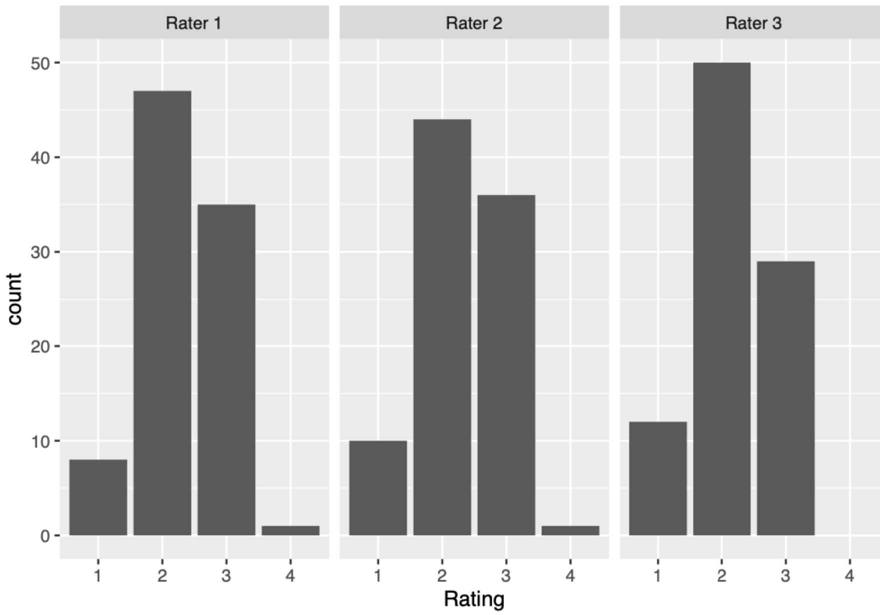


Figure 3: Distribution of ratings for each rater for the 13 artifacts

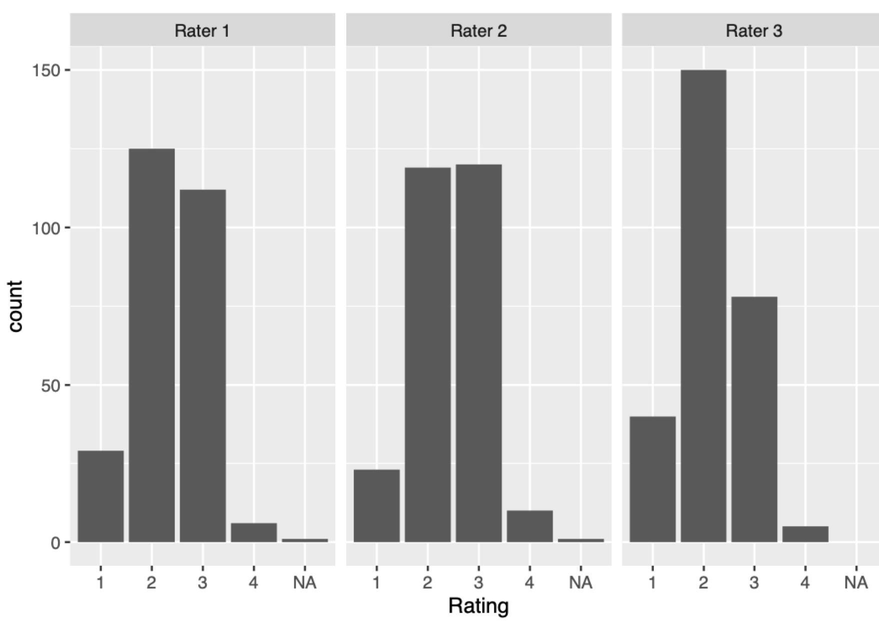


Figure 4: Distribution of ratings for each rater for the full dataset

According to figure 5 below, the three columns on the right are the pairwise ICCs for the ratings by different raters with respect to different rubrics. we find that most of the ICCs in the three columns are greater than 0.5, it shows strong intraclass correlation for most of the raters pairs, it means that the raters tend to give similar ratings for most of the raters. The lowest ICC was 0.38 in the three columns is the rubric RsrchQ for rater 1 and rater2. According to the ICC.common, we can see the overall ICC for all the three raters with respect to different rubrics. We find that the ICCs were very low for the rubrics TxtOrg, RsrchQ and InterpRes, which mean for these three rubrics, the three raters tend to give different ratings. And for the rest of the rubric, we can roughly say that the three raters do not usually disagree with each other.

	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

Figure 5: The ICC for each rubric, 13 artifacts

According to figure 6, we find that by using the full dataset, we can get a similar result comparing with figure 5 that the three raters roughly agree with each other for the rubrics CritDes, InitEDA and VisOrg. They usually disagree with each other for the rubrics InterpRes, TxtOrg and RsrchQ.

```
##              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
```

Figure 6: The ICC for each rubric, full data

4.3 Question 3

In this part, we found the most important factors related to the ratings. Firstly, by adding the fixed effects to the seven rubric-specific models using just the data from the 13 common artifacts that all three raters saw. For each rubrics, we tried to add fixed effects to our seven rubric-specific models. We start from the full model for each rubric contains all the variables(Rater, Semester and Sex), after using the ANOVA table for backward stepwise variables selection(details please refer to the appendix), we found that we do not need to add any fixed effects to the model for each rubric.

So, we failed to add any fixed effects to the model for each rubric by using the 13 artifacts. We then tried to add the fixd effects to the model fitted by the full dataset. By using the same procedure, we found that we should not add any fixed effects to the models for rubrics InitEDA, RsrchQ and TxtOrg. And we should add Rater as a fixed effect to the model for rubrics CritDes, InterpRes, SelMeth, it showed that different Raters could significantly influence the Ratings for these rubrics. In the end, we find that for rubric SelMeth, we should add the variables Rater and Semester to its model, it meant that for this rubric, different Raters and different semesters could be influential for the final ratings.

Then, we considered about the fixed effects of the interaction terms. For the models with intercept only, we did not need to examine the interaction terms. For the rubric SelMeth, we found our previous model make sense given the t value of each variable are greater than 1.96 that they are all significant. After adding the interaction between Semester and Rater, we find that there was no evidence that we should add this interaction to the model. For random effects, since we should only try the effects appeared in the fixed effect, we tried Rater and Semester as the random effects, after fitting these model, we found that we do not need to add any random effects to the model(details please refer to the appendix). Thus we obtained our final model, with random effects group by each artifact and fixed effects Rater and Semester. It showed that for each artifact rated by rubric SelMeth, different Raters and Semester are significant factors influencing the Ratings.

For the rubric CritDes, we did the similar thing, since Rater is the only fixed effect we included in the model, we only test the Rater as a random effect. And we also find we did not need to add Rater as a random effect here too. It showed that group by each artifact rated by rubric CritDes, different Raters is the only significant fixed effect influencing the Ratings, Different Raters could have different rating for the rubric CritDes for each artifact.

Similarly, for the rubric InterpRes, we did the similar thing, since Rater is the only fixed effect we included in the model, we only test the Rater as a random effect. And we also find we did not need to add Rater as a random effect here too. It showed that group by each artifact rated by rubric InterpRes, different Raters is the only significant fixed effect influencing the Ratings, Different Raters could have different rating for the rubric InterpRes for each artifact.

In summation, by using the full data set, we found that we do not need to add any random effects for each rubric. For fixed effects, rubrics InitEDA, RsrchQ and TxtOrg do not need any fixed effect. The fixed effect Rater could be significant for the rubrics SelMeth, CritDes, InterpRes. And the fixed effect Semester could be influential for the rubric SelMeth.

4.4 Question 4

Firstly, we plotted the histgrams just like we did in question 1.

According to the histgram by using the 13 artifacts, we find that the distributions of the Ratings for Male and Female are almost identical, no bias was found from this plot. Similarly, we also did not find any bias in Ratings for Male and Female populations.

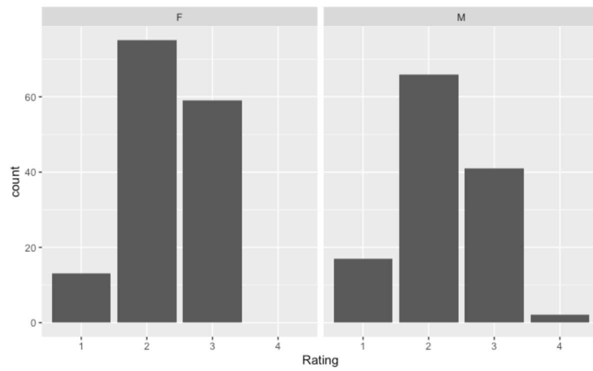


Figure 7: Histgrams of Sex and Ratings, 13 artifacts

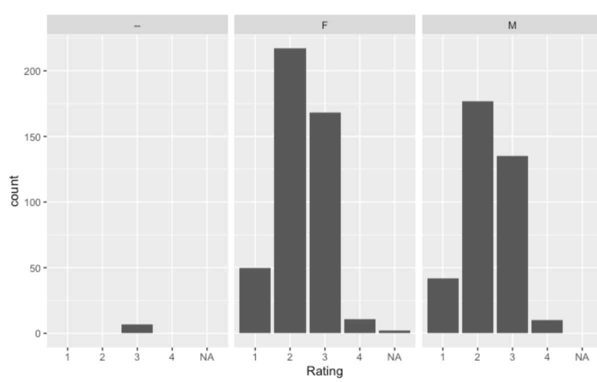


Figure 8: Histgrams of Sex and Ratings, full dataset

5 Discussion

According to the result of question 1, we found that most of the rubrics have the similar distribution of the Ratings, however, rubric CritDes seemed to be very different from other rubrics. It might because of the rule of rating for the rubric CritDes is different from the rest of the rubrics. It seemed to be a much more strict rubric that more than half of the artifacts got rating 1 in this rubric. I think it is a good thing to have rubrics with different distribution that if all the rubrics are the same, we do not need that many rubrics anymore. Different rubrics, which could evaluate the different aspects of an artifact are we really need. And for different raters, we can say they roughly agree with each other when rating, however, rater 3 seems to be more extreme, since he/she are more likely to give a lower rating(1,2).

According to the result of question 2, we find that for most of the artifacts and for most of the rubrics, raters are quite likely to give similar or identical ratings, and thus have high ICCs, but for some rubrics(InterpRes, RsrchQ, VisOrg), the ICCs are low, which means, for these rubrics, the raters do not usually agree with each others. For the same artifacts, it could be really weird to have raters disagree with each other and give significantly different ratings. From my perspective, it might because these specific rubrics are more subjective and thus different raters could have more different results.

For the result of question 3, we find that the for most of the rubrics(InitEDA, RsrchQ and TxtOrg), the influence of different Raters are not significant, that the fairness with respect to different raters are guaranteed for these rubrics. And for the rest of the rubrics, it seems the rating for each rubric are related to different raters, which is, it seems it is not so fair to use these rubric. Moreover, we found that different semesters is also a significant variable for the rubric SelMeth, it means that the fairness for this rubric is very poor, the rubric we want should be objective and uniform.

For the result of question 4, we find that sex is not a significant factor that influence the rating, which means that the bias of sex does not exist in the Rating process.

References

[Sheather(2009)] Sheather, S. J. (2009), *A Modern Approach to Regression*, Springer.

Technical Appendix

11/29/2021

```r
library(lme4)
```

```
## Loading required package: Matrix
```

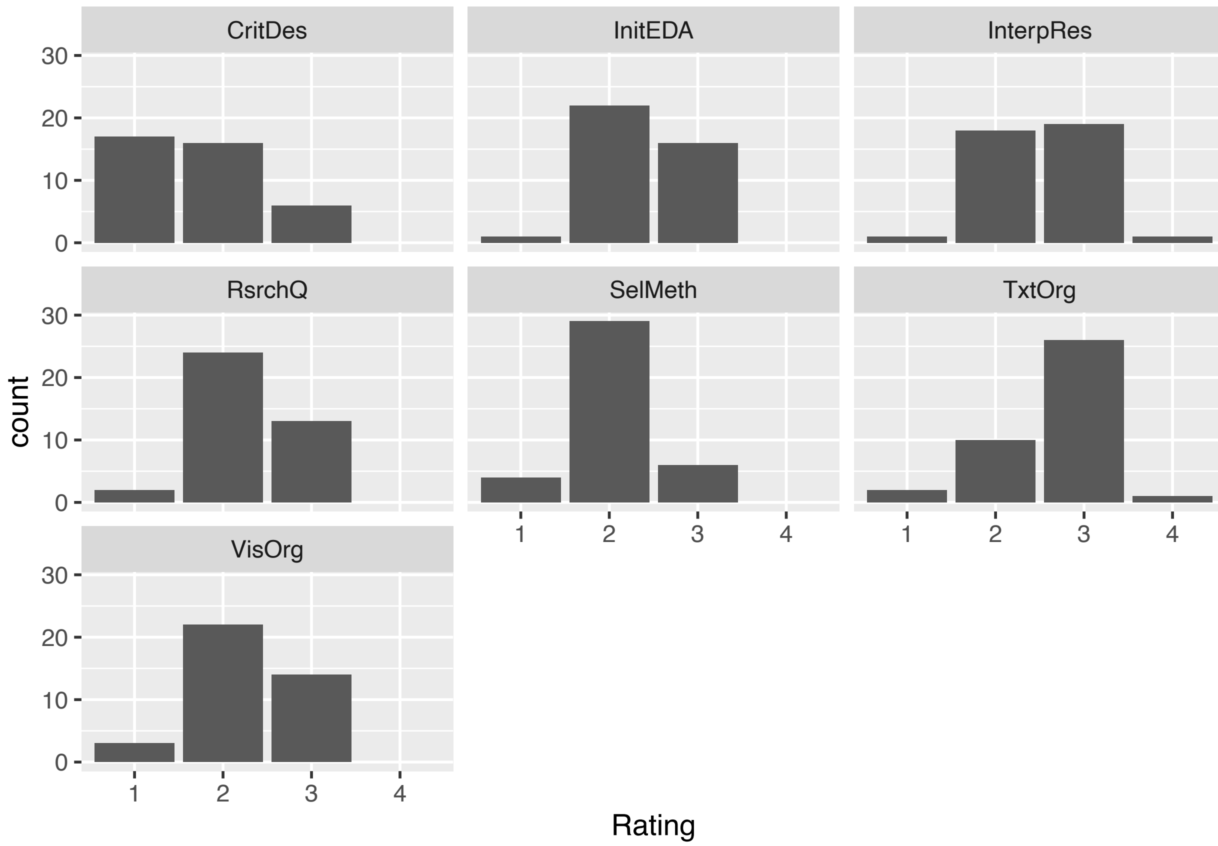
```r
library(arm)
```

```
## Loading required package: MASS

##
## arm (Version 1.12-2, built: 2021-10-15)

## Working directory is /Users/wyc/Downloads
```

```r
library(ggplot2)
ratings <- read.csv("/Users/wyc/ratings.csv",header=T)
tall <-  read.csv("/Users/wyc/tall.csv",header=T)
tall$Rating <- factor(tall$Rating,levels=1:4)
for (i in unique(tall$Rubric)) {
  ratings[,i] <- factor(ratings[,i],levels=1:4)
}
tall$Sex[nchar(tall$Sex)==0] <- "--"

##
## Extract the reduced data set with the 13 artifacts that all 3 raters saw...
ratings.13 <- ratings[grep("0",ratings$Artifact),]
tall.13 <-  tall[grep("0",tall$Artifact),]
```

Question 1, Is the distribution of ratings for each rubrics pretty much indistinguishable from the other rubrics, or are there rubrics that tend to get especially high or low ratings? Is the distribution of ratings given by each rater pretty much indistinguishable from the other raters, or are there raters that tend to give especially high or low ratings?

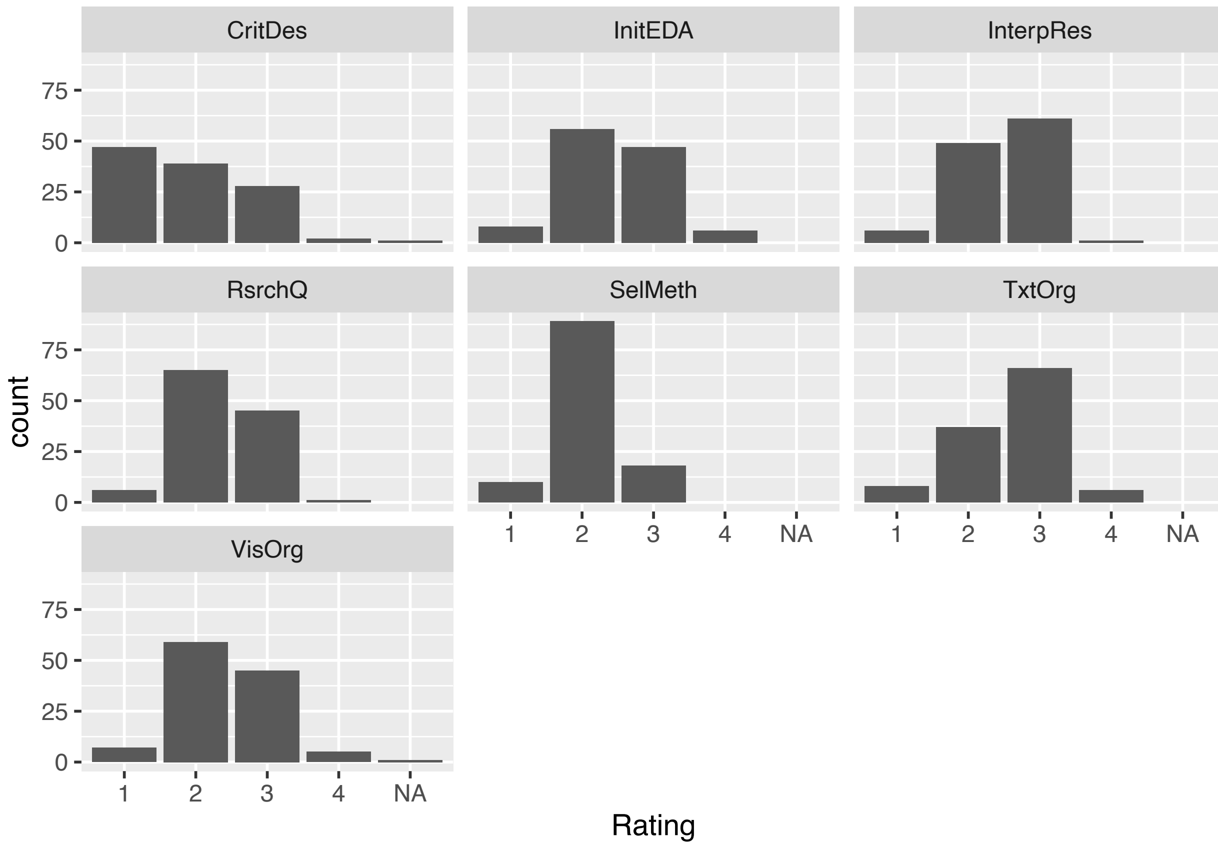
```r
g <- ggplot(tall.13,aes(x = Rating)) +
  facet_wrap( ~ Rubric) +
  geom_bar()

g
```




```
tmp <- data.frame(lapply(split(tall.13$Rating,tall.13$Rubric),summary))
row.names(tmp) <- paste("Rating",1:4)

tmp
```

```
##          CritDes InitEDA InterpRes RsrchQ SelMeth TxtOrg VisOrg
## Rating 1      17       1         1      2       4      2      3
## Rating 2      16      22        18     24      29     10     22
## Rating 3       6      16        19     13       6     26     14
## Rating 4       0       0         1      0       0      1      0
```

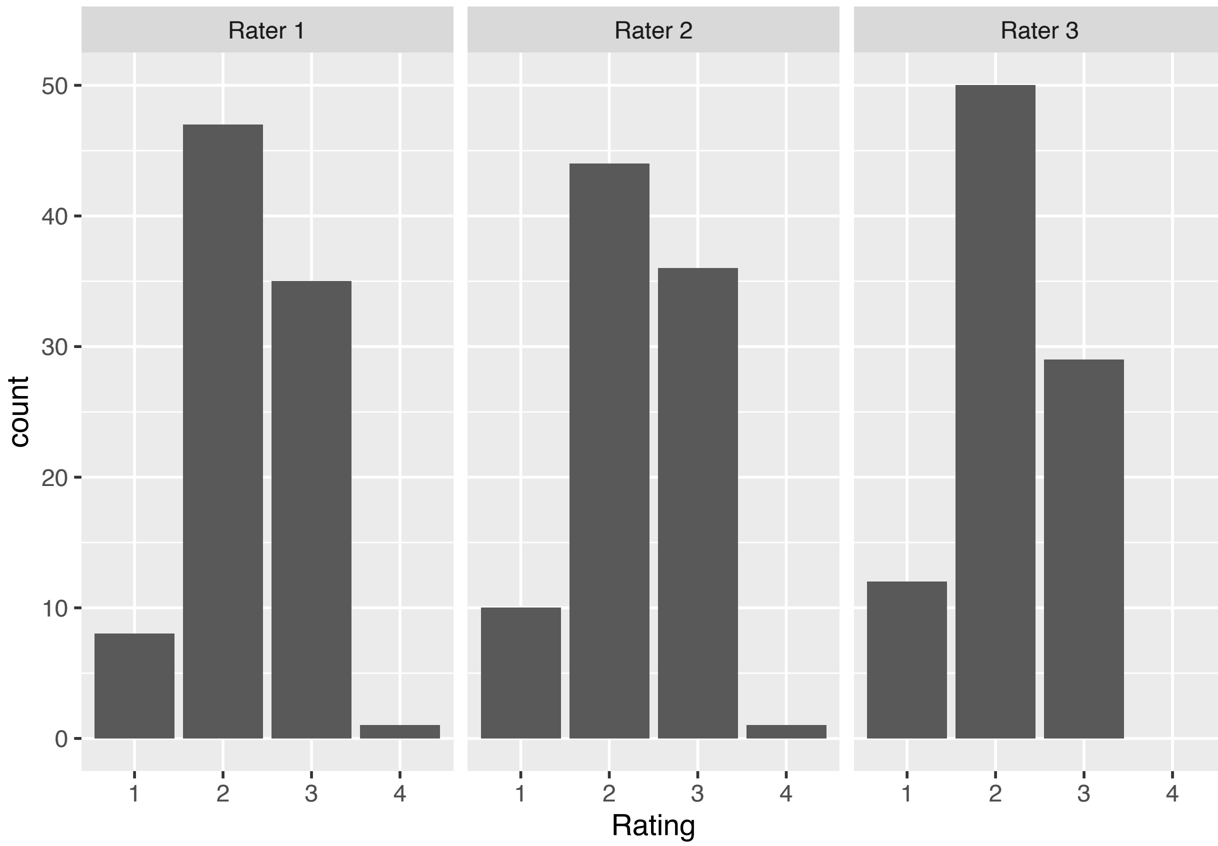
```
g <- ggplot(tall,aes(x = Rating)) +
  facet_wrap( ~ Rubric) +
  geom_bar()

g
```




```
tmp0 <- lapply(split(tall$Rating,tall$Rubric),summary)
tmp <- data.frame(matrix(0,nrow=5,ncol=7))
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]
}

tmp
```

```
##          CritDes InitEDA InterpRes RsrchQ SelMeth TxtOrg VisOrg
## Rating 1      47       8         6      6      10      8      7
## Rating 2      39      56        49     65      89     37     59
## Rating 3      28      47        61     45      18     66     45
## Rating 4       2       6         1      1       0      6      5
## <NA>           1       0         0      0       0      0      1
```

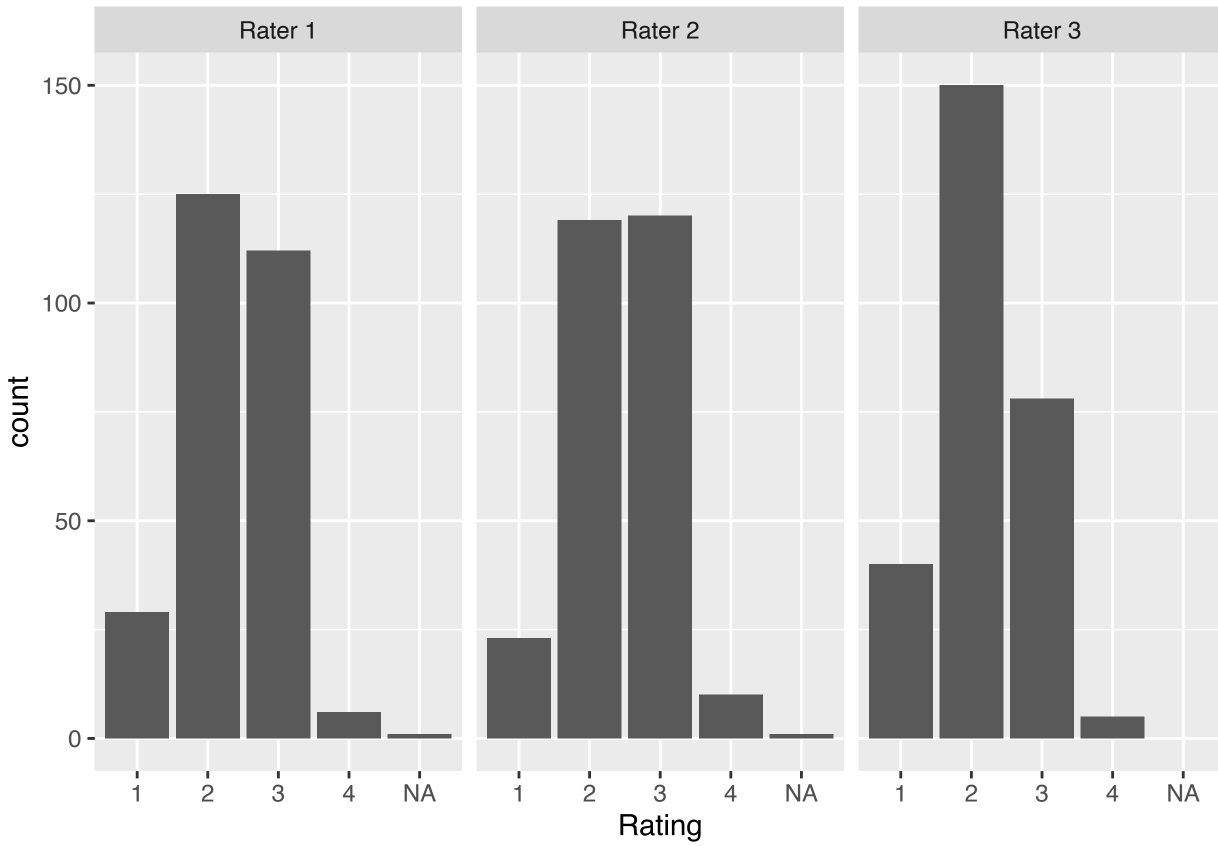
```
rater.name <- function(x) { paste("Rater",x) }


g <- ggplot(tall.13,aes(x = Rating)) +
  facet_wrap( ~ Rater, labeller=labeller(Rater=rater.name)) +
  geom_bar()

g
```




```
tmp <- data.frame(lapply(split(tall.13$Rating,tall.13$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
```

```
g <- ggplot(tall,aes(x = Rating)) +
  facet_wrap( ~ Rater, labeller=labeller(Rater=rater.name)) +
  geom_bar()

g
```




```
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 InterpRes
## 5 5     3      5      NA     Fall  --      3       3       3       3         3
##   VisOrg TxtOrg Artifact Repeated
## 5      3      3        5        0
```

Question 2 For each rubric, do the raters generally agree on their scores? If not, is there one rater who disagrees with the others? Or do they all disagree?

```
Rubric.names <- sort(unique(tall$Rubric))



ICC.vec <- NULL
for (i in Rubric.names) {

  tmp <- lmer(as.numeric(Rating) ~ 1 + (1|Artifact), data=tall.13[tall.13$Rubric==i,])
  sig2 <- summary(tmp)$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.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.13[ratings.13$Rater==1,i],levels=1:4),
                    r2=factor(ratings.13[ratings.13$Rater==2,i],levels=1:4),
                    a1=ratings.13[ratings.13$Rater==1,"Artifact"],
                    a2=ratings.13[ratings.13$Rater==2,"Artifact"])
  if(any(r12[,3]!=r12[,4])) { stop(paste("Rater 1-2 Artifact mismatch on rubric",i)) }
  a12 <- mean(r12[,1]==r12[,2])
  r12 <- table(r12[,1:2])

  r23 <- data.frame(r2=factor(ratings.13[ratings.13$Rater==2,i],levels=1:4),
                    r3=factor(ratings.13[ratings.13$Rater==3,i],levels=1:4),
                    a2=ratings.13[ratings.13$Rater==2,"Artifact"],
                    a3=ratings.13[ratings.13$Rater==3,"Artifact"])
  if(any(r23[,3]!=r23[,4])) { stop(paste("Rater 2-3 Artifact mismatch on rubric",i)) }
  a23 <- mean(r23[,1]==r23[,2])
  r23 <- table(r23[,1:2])

  r13 <- data.frame(r1=factor(ratings.13[ratings.13$Rater==1,i],levels=1:4),
                    r3=factor(ratings.13[ratings.13$Rater==3,i],levels=1:4),
                    a1=ratings.13[ratings.13$Rater==1,"Artifact"],
                    a3=ratings.13[ratings.13$Rater==3,"Artifact"])
  if(any(r13[,3]!=r13[,4])) { stop(paste("Rater 1-3 Artifact mismatch on rubric",i)) }
  a13 <- mean(r13[,1]==r13[,2])
  r13 <- table(r13[,1:2])

  agreement.results[i,2:4] <- c(a12,a23,a13)

  agreement.tables[[i]] <- list(r12,r23,r13)

}
```

```r
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
```

```r
##

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
```

Question 3 More generally, how are the various factors in this experiement (Rater, Semester, Sex, Repeated, Rubric) related to the ratings? Do the factors interact in any interesting ways?

Question 3 (i): Adding fixed effects to the seven rubric-specific models using just the data from the 13 common artifacts that al three raters saw

```r
library(LMERConvenienceFunctions)
library(RLRsim)

tmp <- lmer(as.numeric(Rating) ~ -1 + as.factor(Rater) +
              Semester + Sex + (1|Artifact),
            data=tall.13[tall.13$Rubric=="RsrchQ",],REML=FALSE)
```

```r
tmp.back_elim <- fitLMER.fnc(tmp,set.REML.FALSE = TRUE,log.file.name = FALSE)
```

```
## ======================================================
## ===               backfitting fixed effects        ===
## ======================================================
## processing model terms of interaction level 1
##   iteration 1
##     p-value for term "Semester" = 0.7355 >= 0.05
##     not part of higher-order interaction
##     removing term
##   iteration 2
##     p-value for term "Sex" = 0.279 >= 0.05
##     not part of higher-order interaction
##     removing term
## pruning random effects structure ...
##   nothing to prune
## ======================================================
## ===             forwardfitting random effects      ===
## ======================================================
##  ===       random slopes       ===
## ======================================================
## ===             re-backfitting fixed effects       ===
## ======================================================
## processing model terms of interaction level 1
##   all terms of interaction level 1 significant
## resetting REML to TRUE
## pruning random effects structure ...
##   nothing to prune
```

```r
formula(tmp.back_elim)
```

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

```r
tmp.int_only <- update(tmp.back_elim, . ~ . + 1 - as.factor(Rater))

anova(tmp.int_only,tmp.back_elim)
```

```
## Data: tall.13[tall.13$Rubric == "RsrchQ", ]
## Models:
## tmp.int_only: as.numeric(Rating) ~ (1 | Artifact)
## tmp.back_elim: as.numeric(Rating) ~ as.factor(Rater) + (1 | Artifact) - 1
##               npar    AIC    BIC  logLik deviance  Chisq Df Pr(>Chisq)
## tmp.int_only     3 69.457 74.447 -31.728   63.457
## tmp.back_elim    5 72.018 80.335 -31.009   62.018 1.4391  2      0.487
```

```r
anova(tmp.int_only,tmp.back_elim)$"Pr(>Chisq)"[2]
```

```
## [1] 0.4869707
```

```r
Rubric.names <- sort(unique(tall$Rubric))

model.formula.13 <- as.list(rep(NA,7))
names(model.formula.13) <- Rubric.names


for (i in Rubric.names) {
```

```r
  rubric.data <- tall.13[tall.13$Rubric==i,]
  tmp <- lmer(as.numeric(Rating) ~ -1 + as.factor(Rater) +
              Semester + Sex + (1|Artifact),
            data=rubric.data,REML=FALSE)


  tmp.back_elim <- fitLMER.fnc(tmp,set.REML.FALSE = TRUE,log.file.name = FALSE)


  tmp.single_intercept <- update(tmp.back_elim, . ~ . + 1 - as.factor(Rater))
  pval <- anova(tmp.single_intercept,tmp.back_elim)$"Pr(>Chisq)"[2]


  if (pval<=0.05) {
    tmp_final <- tmp.back_elim
  } else {
    tmp_final <- tmp.single_intercept
  }


  model.formula.13[[i]] <- formula(tmp_final)

}
```

```
## ======================================================
## ===              backfitting fixed effects         ===
## ======================================================
## processing model terms of interaction level 1
##   iteration 1
##     p-value for term "Sex" = 0.2229 >= 0.05
##     not part of higher-order interaction
##     removing term
##   iteration 2
##     p-value for term "Semester" = 0.1826 >= 0.05
##     not part of higher-order interaction
##     removing term
## pruning random effects structure ...
##   nothing to prune
## ======================================================
## ===            forwardfitting random effects       ===
## ======================================================
##  ===       random slopes       ===
## ======================================================
## ===            re-backfitting fixed effects        ===
## ======================================================
## processing model terms of interaction level 1
##   all terms of interaction level 1 significant
## resetting REML to TRUE
## pruning random effects structure ...
##   nothing to prune
## ======================================================
## ===              backfitting fixed effects         ===
```

```
## ========================================================
## processing model terms of interaction level 1
##   iteration 1
##     p-value for term "Semester" = 0.8137 >= 0.05
##     not part of higher-order interaction
##     removing term
##   iteration 2
##     p-value for term "Sex" = 0.6429 >= 0.05
##     not part of higher-order interaction
##     removing term
## pruning random effects structure ...
##   nothing to prune
## ========================================================
## ===            forwardfitting random effects       ===
## ========================================================
##  ===       random slopes       ===
## ========================================================
## ===            re-backfitting fixed effects        ===
## ========================================================
## processing model terms of interaction level 1
##   all terms of interaction level 1 significant
## resetting REML to TRUE
## pruning random effects structure ...
##   nothing to prune
## ========================================================
## ===              backfitting fixed effects         ===
## ========================================================
## processing model terms of interaction level 1
##   iteration 1
##     p-value for term "Semester" = 0.8294 >= 0.05
##     not part of higher-order interaction
##     removing term
##   iteration 2
##     p-value for term "Sex" = 0.2947 >= 0.05
##     not part of higher-order interaction
##     removing term
## pruning random effects structure ...
##   nothing to prune
## ========================================================
## ===            forwardfitting random effects       ===
## ========================================================
##  ===       random slopes       ===
## ========================================================
## ===            re-backfitting fixed effects        ===
## ========================================================
## processing model terms of interaction level 1
##   all terms of interaction level 1 significant
## resetting REML to TRUE
## pruning random effects structure ...
##   nothing to prune
## ========================================================
## ===              backfitting fixed effects         ===
## ========================================================
## processing model terms of interaction level 1
```

```
##   iteration 1
##     p-value for term "Semester" = 0.7355 >= 0.05
##     not part of higher-order interaction
##     removing term
##   iteration 2
##     p-value for term "Sex" = 0.279 >= 0.05
##     not part of higher-order interaction
##     removing term
## pruning random effects structure ...
##   nothing to prune
## =======================================================
## ===            forwardfitting random effects       ===
## =======================================================
##  ===       random slopes       ===
## =======================================================
## ===            re-backfitting fixed effects        ===
## =======================================================
## processing model terms of interaction level 1
##   all terms of interaction level 1 significant
## resetting REML to TRUE
## pruning random effects structure ...
##   nothing to prune
## =======================================================
## ===              backfitting fixed effects         ===
## =======================================================
## processing model terms of interaction level 1
##   iteration 1
##     p-value for term "Sex" = 0.9383 >= 0.05
##     not part of higher-order interaction
##     removing term
##   iteration 2
##     p-value for term "Semester" = 0.4287 >= 0.05
##     not part of higher-order interaction
##     removing term
## pruning random effects structure ...
##   nothing to prune
## =======================================================
## ===            forwardfitting random effects       ===
## =======================================================
##  ===       random slopes       ===
## =======================================================
## ===            re-backfitting fixed effects        ===
## =======================================================
## processing model terms of interaction level 1
##   all terms of interaction level 1 significant
## resetting REML to TRUE
## pruning random effects structure ...
##   nothing to prune
## =======================================================
## ===              backfitting fixed effects         ===
## =======================================================
## processing model terms of interaction level 1
##   iteration 1
##     p-value for term "Semester" = 0.5358 >= 0.05
```

```
##     not part of higher-order interaction
##     removing term
##   iteration 2
##     p-value for term "Sex" = 0.1319 >= 0.05
##     not part of higher-order interaction
##     removing term
## pruning random effects structure ...
##   nothing to prune
## ======================================================
## ===            forwardfitting random effects       ===
## ======================================================
##  ===       random slopes       ===
## ======================================================
## ===            re-backfitting fixed effects        ===
## ======================================================
## processing model terms of interaction level 1
##   all terms of interaction level 1 significant
## resetting REML to TRUE
## pruning random effects structure ...
##   nothing to prune
## ======================================================
## ===              backfitting fixed effects         ===
## ======================================================
## processing model terms of interaction level 1
##   iteration 1
##     p-value for term "Semester" = 0.1922 >= 0.05
##     not part of higher-order interaction
##     removing term
##   iteration 2
##     p-value for term "Sex" = 0.1078 >= 0.05
##     not part of higher-order interaction
##     removing term
## pruning random effects structure ...
##   nothing to prune
## ======================================================
## ===            forwardfitting random effects       ===
## ======================================================
##  ===       random slopes       ===
## ======================================================
## ===            re-backfitting fixed effects        ===
## ======================================================
## processing model terms of interaction level 1
##   all terms of interaction level 1 significant
## resetting REML to TRUE
## pruning random effects structure ...
##   nothing to prune
```

```
model.formula.13
```

```
## $CritDes
## as.numeric(Rating) ~ (1 | Artifact)
##
## $InitEDA
## as.numeric(Rating) ~ (1 | Artifact)
##
```

```
## $InterpRes
## as.numeric(Rating) ~ (1 | Artifact)
##
## $RsrchQ
## as.numeric(Rating) ~ (1 | Artifact)
##
## $SelMeth
## as.numeric(Rating) ~ (1 | Artifact)
##
## $TxtOrg
## as.numeric(Rating) ~ (1 | Artifact)
##
## $VisOrg
## as.numeric(Rating) ~ (1 | Artifact)
```

Question 3 (ii): Adding fixed effects to the seven rubric-specific models using all the data

```
Rubric.names <- sort(unique(tall$Rubric))


tall[c(161,684),]
```

```
##       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),]


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!="--",]

model.formula.alldata <- as.list(rep(NA,7))
names(model.formula.alldata) <- Rubric.names


for (i in Rubric.names) {


  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)
```

```r
  tmp.back_elim <- fitLMER.fnc(tmp,set.REML.FALSE = TRUE,log.file.name = FALSE)


  tmp.single_intercept <- update(tmp.back_elim, . ~ . + 1 - as.factor(Rater))
  pval <- anova(tmp.single_intercept,tmp.back_elim)$"Pr(>Chisq)"[2]


  if (pval<=0.05) {
    tmp_final <- tmp.back_elim
  } else {
    tmp_final <- tmp.single_intercept
  }


  model.formula.alldata[[i]] <- formula(tmp_final)

}
```

```
## Warning in fitLMER.fnc(tmp, set.REML.FALSE = TRUE, log.file.name = FALSE): Argument "ran.effects" is
## TRUE
```

```
## ======================================================
## ===              backfitting fixed effects         ===
## ======================================================
## processing model terms of interaction level 1
##   iteration 1
##     p-value for term "Semester" = 0.7154 >= 0.05
##     not part of higher-order interaction
##     removing term
##   iteration 2
##     p-value for term "Sex" = 0.5297 >= 0.05
##     not part of higher-order interaction
##     removing term
## pruning random effects structure ...
##   nothing to prune
## ======================================================
## ===            forwardfitting random effects       ===
## ======================================================
##  ===       random slopes       ===
## ======================================================
## ===            re-backfitting fixed effects        ===
## ======================================================
## processing model terms of interaction level 1
##   all terms of interaction level 1 significant
## resetting REML to TRUE
## pruning random effects structure ...
##   nothing to prune
```

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

```
## Warning in fitLMER.fnc(tmp, set.REML.FALSE = TRUE, log.file.name = FALSE): Argument "ran.effects" is
## TRUE
```

```
## ======================================================
## ===              backfitting fixed effects         ===
## ======================================================
```

```
## processing model terms of interaction level 1
##   iteration 1
##     p-value for term "Semester" = 0.8802 >= 0.05
##     not part of higher-order interaction
##     removing term
##   iteration 2
##     p-value for term "Sex" = 0.7402 >= 0.05
##     not part of higher-order interaction
##     removing term
## pruning random effects structure ...
##   nothing to prune
## ========================================================
## ===            forwardfitting random effects       ===
## ========================================================
##  ===       random slopes       ===
## ========================================================
## ===            re-backfitting fixed effects        ===
## ========================================================
## processing model terms of interaction level 1
##   all terms of interaction level 1 significant
## resetting REML to TRUE
## pruning random effects structure ...
##   nothing to prune

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

## Warning in fitLMER.fnc(tmp, set.REML.FALSE = TRUE, log.file.name = FALSE): Argument "ran.effects" is
## TRUE

## ========================================================
## ===              backfitting fixed effects         ===
## ========================================================
## processing model terms of interaction level 1
##   iteration 1
##     p-value for term "Sex" = 0.608 >= 0.05
##     not part of higher-order interaction
##     removing term
##   iteration 2
##     p-value for term "Semester" = 0.5312 >= 0.05
##     not part of higher-order interaction
##     removing term
## pruning random effects structure ...
##   nothing to prune
## ========================================================
## ===            forwardfitting random effects       ===
## ========================================================
##  ===       random slopes       ===
## ========================================================
## ===            re-backfitting fixed effects        ===
## ========================================================
## processing model terms of interaction level 1
##   all terms of interaction level 1 significant
## resetting REML to TRUE
## pruning random effects structure ...
##   nothing to prune
```

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

## Warning in fitLMER.fnc(tmp, set.REML.FALSE = TRUE, log.file.name = FALSE): Argument "ran.effects" is
## TRUE

## ======================================================
## ===              backfitting fixed effects         ===
## ======================================================
## processing model terms of interaction level 1
##   iteration 1
##     p-value for term "Sex" = 0.6166 >= 0.05
##     not part of higher-order interaction
##     removing term
##   iteration 2
##     p-value for term "Semester" = 0.3987 >= 0.05
##     not part of higher-order interaction
##     removing term
## pruning random effects structure ...
##   nothing to prune
## ======================================================
## ===            forwardfitting random effects       ===
## ======================================================
##  ===       random slopes       ===
## ======================================================
## ===            re-backfitting fixed effects        ===
## ======================================================
## processing model terms of interaction level 1
##   all terms of interaction level 1 significant
## resetting REML to TRUE
## pruning random effects structure ...
##   nothing to prune

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

## Warning in fitLMER.fnc(tmp, set.REML.FALSE = TRUE, log.file.name = FALSE): Argument "ran.effects" is
## TRUE

## ======================================================
## ===              backfitting fixed effects         ===
## ======================================================
## processing model terms of interaction level 1
##   iteration 1
##     p-value for term "Sex" = 0.1935 >= 0.05
##     not part of higher-order interaction
##     removing term
## pruning random effects structure ...
##   nothing to prune
## ======================================================
## ===            forwardfitting random effects       ===
## ======================================================
##  ===       random slopes       ===
## ======================================================
## ===            re-backfitting fixed effects        ===
## ======================================================
## processing model terms of interaction level 1
##   all terms of interaction level 1 significant
```

```
## resetting REML to TRUE
## pruning random effects structure ...
##   nothing to prune

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

## Warning in fitLMER.fnc(tmp, set.REML.FALSE = TRUE, log.file.name = FALSE): Argument "ran.effects" is
## TRUE

## =========================================================
## ===              backfitting fixed effects          ===
## =========================================================
## processing model terms of interaction level 1
##   iteration 1
##     p-value for term "Sex" = 0.5041 >= 0.05
##     not part of higher-order interaction
##     removing term
##   iteration 2
##     p-value for term "Semester" = 0.205 >= 0.05
##     not part of higher-order interaction
##     removing term
## pruning random effects structure ...
##   nothing to prune
## =========================================================
## ===            forwardfitting random effects        ===
## =========================================================
##  ===       random slopes       ===
## =========================================================
## ===            re-backfitting fixed effects         ===
## =========================================================
## processing model terms of interaction level 1
##   all terms of interaction level 1 significant
## resetting REML to TRUE
## pruning random effects structure ...
##   nothing to prune

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

## Warning in fitLMER.fnc(tmp, set.REML.FALSE = TRUE, log.file.name = FALSE): Argument "ran.effects" is
## TRUE

## =========================================================
## ===              backfitting fixed effects          ===
## =========================================================
## processing model terms of interaction level 1
##   iteration 1
##     p-value for term "Semester" = 0.2158 >= 0.05
##     not part of higher-order interaction
##     removing term
##   iteration 2
##     p-value for term "Sex" = 0.3523 >= 0.05
##     not part of higher-order interaction
##     removing term
## pruning random effects structure ...
##   nothing to prune
## =========================================================
## ===            forwardfitting random effects        ===
```

```
## ======================================================
## ===       random slopes       ===
## ======================================================
## ===            re-backfitting fixed effects        ===
## ======================================================
## processing model terms of interaction level 1
##   all terms of interaction level 1 significant
## resetting REML to TRUE
## pruning random effects structure ...
##   nothing to prune

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

```r
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
```

Question 3 (iii): Trying interactions and new random effects for the seven rubric specific models using all the data

```r
fla <- formula(model.formula.alldata[["SelMeth"]])
tmp <- lmer(fla,data=tall.nonmissing[tall.nonmissing$Rubric=="SelMeth",])
round(summary(tmp)$coef,2)  ## fixed effects and their t-values
```

```
##                    Estimate Std. Error t value
## as.factor(Rater)1     2.25       0.08   29.99
## as.factor(Rater)2     2.23       0.07   29.99
## as.factor(Rater)3     2.03       0.08   27.03
## SemesterS19          -0.36       0.10   -3.66
```

```r
tmp.single_intercept <- update(tmp, . ~ . + 1 - as.factor(Rater))
anova(tmp.single_intercept,tmp)
```

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

## Data: tall.nonmissing[tall.nonmissing$Rubric == "SelMeth", ]
```

```
## Models:
## tmp.single_intercept: as.numeric(Rating) ~ Semester + (1 | Artifact)
## tmp: as.numeric(Rating) ~ as.factor(Rater) + Semester + (1 | Artifact) - 1
##                       npar    AIC    BIC  logLik deviance  Chisq Df Pr(>Chisq)
## tmp.single_intercept    4 145.07 156.08 -68.534   137.07
## tmp                     6 142.05 158.58 -65.027   130.05 7.0146  2    0.02998 *
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
```

```
tmp.fixed_interactions <- update(tmp, . ~ . + as.factor(Rater)*Semester - Semester)

anova(tmp,tmp.fixed_interactions)
```

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

```
## Data: tall.nonmissing[tall.nonmissing$Rubric == "SelMeth", ]
## Models:
## tmp: as.numeric(Rating) ~ as.factor(Rater) + Semester + (1 | Artifact) - 1
## tmp.fixed_interactions: as.numeric(Rating) ~ as.factor(Rater) + (1 | Artifact) + as.factor(Rater):Se
##                         npar    AIC    BIC  logLik deviance Chisq Df Pr(>Chisq)
## tmp                        6 142.05 158.58 -65.027   130.05
## tmp.fixed_interactions     8 143.46 165.49 -63.731   127.46 2.592  2     0.2736
```

```
m0 <- tmp                                        ## Null hypothesis
mA <- update(m0, . ~ . + (Semester|Artifact))    ## Alternative hypotheses
```

```
## Error: number of observations (=116) <= number of random effects (=180) for term (Semester | Artifac
```

```
m  <- update(mA, . ~ . - (1|Artifact))           ## Model with only the new R.E.
```

```
## Error in h(simpleError(msg, call)): error in evaluating the argument 'object' in selecting a method
```

```
exactRLRT(m0=m0,mA=mA,m=m)
```

```
## Error in exactRLRT(m0 = m0, mA = mA, m = m): object 'm' not found
```

```
m0 <- tmp                                        ## Null hypothesis
mA <- update(m0, . ~ . + (as.factor(Rater)|Artifact))   ## Alternative hypotheses
```

```
## Error: number of observations (=116) <= number of random effects (=270) for term (as.factor(Rater) |
```

```
m  <- update(mA, . ~ . - (1|Artifact))           ## Model with only the new R.E.
```

```
## Error in h(simpleError(msg, call)): error in evaluating the argument 'object' in selecting a method
```

```
exactRLRT(m0=m0,mA=mA,m=m)
```

```
## Error in exactRLRT(m0 = m0, mA = mA, m = m): object 'm' not found
```

```
summary(tmp)
```

```
## Linear mixed model fit by REML ['lmerMod']
## Formula: as.numeric(Rating) ~ as.factor(Rater) + Semester + (1 | Artifact) -
##     1
##    Data: tall.nonmissing[tall.nonmissing$Rubric == "SelMeth", ]
##
## REML criterion at convergence: 143.6
##
## Scaled residuals:
##     Min      1Q  Median      3Q     Max
## -2.0480 -0.3923 -0.0551  0.2674  2.5827
```

```
##
## Random effects:
##  Groups   Name        Variance Std.Dev.
##  Artifact (Intercept) 0.08973  0.2996
##  Residual             0.10842  0.3293
## Number of obs: 116, groups:  Artifact, 90
##
## Fixed effects:
##                     Estimate Std. Error t value
## as.factor(Rater)1  2.25037    0.07503  29.992
## as.factor(Rater)2  2.22653    0.07424  29.991
## as.factor(Rater)3  2.03316    0.07521  27.033
## SemesterS19       -0.35860    0.09796  -3.661
##
## Correlation of Fixed Effects:
##             a.(R)1 a.(R)2 a.(R)3
## as.fctr(R)2  0.285
## as.fctr(R)3  0.287  0.280
## SemesterS19 -0.413 -0.391 -0.394
```

```r
fla <- formula(model.formula.alldata[["CritDes"]])
tmp <- lmer(fla,data=tall.nonmissing[tall.nonmissing$Rubric=="CritDes",])
round(summary(tmp)$coef,2)
```

```
##                     Estimate Std. Error t value
## as.factor(Rater)1     1.69       0.12   13.98
## as.factor(Rater)2     2.11       0.12   17.34
## as.factor(Rater)3     1.89       0.12   15.51
```

```r
tmp.single_intercept <- update(tmp, . ~ . + 1 - as.factor(Rater))
anova(tmp.single_intercept,tmp)
```

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

## Data: tall.nonmissing[tall.nonmissing$Rubric == "CritDes", ]
## Models:
## tmp.single_intercept: as.numeric(Rating) ~ (1 | Artifact)
## tmp: as.numeric(Rating) ~ as.factor(Rater) + (1 | Artifact) - 1
##                      npar    AIC    BIC  logLik deviance  Chisq Df Pr(>Chisq)
## tmp.single_intercept    3 277.68 285.91 -135.84   271.68
## tmp                     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
```

```r
m0 <- tmp                                          ## Null hypothesis
mA <- update(m0, . ~ . + (as.factor(Rater)|Artifact))   ## Alternative hypotheses
```

```
## Error: number of observations (=115) <= number of random effects (=267) for term (as.factor(Rater) |
```

```r
m  <- update(mA, . ~ . - (1|Artifact))             ## Model with only the new R.E.
```

```
## Error in h(simpleError(msg, call)): error in evaluating the argument 'object' in selecting a method
```

```r
exactRLRT(m0=m0,mA=mA,m=m)
```

```
## Error in exactRLRT(m0 = m0, mA = mA, m = m): object 'm' not found
```

```r
summary(tmp)
```

```
## Linear mixed model fit by REML ['lmerMod']
## Formula: as.numeric(Rating) ~ as.factor(Rater) + (1 | Artifact) - 1
##    Data: tall.nonmissing[tall.nonmissing$Rubric == "CritDes", ]
##
## 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
## as.factor(Rater)1   1.6863     0.1207   13.98
## as.factor(Rater)2   2.1129     0.1219   17.34
## as.factor(Rater)3   1.8908     0.1219   15.51
##
## Correlation of Fixed Effects:
##             a.(R)1 a.(R)2
## as.fctr(R)2 0.244
## as.fctr(R)3 0.244  0.246
```

```r
fla <- formula(model.formula.alldata[["InterpRes"]])
tmp <- lmer(fla,data=tall.nonmissing[tall.nonmissing$Rubric=="InterpRes",])
round(summary(tmp)$coef,2)
```

```
##                   Estimate Std. Error t value
## as.factor(Rater)1     2.70       0.09   30.34
## as.factor(Rater)2     2.59       0.09   29.01
## as.factor(Rater)3     2.14       0.09   23.70
```

```r
tmp.single_intercept <- update(tmp, . ~ . + 1 - as.factor(Rater))
anova(tmp.single_intercept,tmp)
```

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

```
## Data: tall.nonmissing[tall.nonmissing$Rubric == "InterpRes", ]
## Models:
## tmp.single_intercept: as.numeric(Rating) ~ (1 | Artifact)
## tmp: as.numeric(Rating) ~ as.factor(Rater) + (1 | Artifact) - 1
##                      npar    AIC    BIC   logLik deviance  Chisq Df Pr(>Chisq)
## tmp.single_intercept    3 218.53 226.79 -106.263   212.53
## tmp                     5 200.66 214.43  -95.331   190.66 21.864  2  1.787e-05
##
## tmp.single_intercept
## tmp                  ***
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
```

```r
m0 <- tmp                                   ## Null hypothesis
mA <- update(m0, . ~ . + (as.factor(Rater)|Artifact))   ## Alternative hypotheses
```

```
## Error: number of observations (=116) <= number of random effects (=270) for term (as.factor(Rater) |
```

```r
m  <- update(mA, . ~ . - (1|Artifact))          ## Model with only the new R.E.
```

```
## Error in h(simpleError(msg, call)): error in evaluating the argument 'object' in selecting a method f
```

```r
exactRLRT(m0=m0,mA=mA,m=m)
```

```
## Error in exactRLRT(m0 = m0, mA = mA, m = m): object 'm' not found
```

```r
summary(tmp)
```

```
## Linear mixed model fit by REML ['lmerMod']
## Formula: as.numeric(Rating) ~ as.factor(Rater) + (1 | Artifact) - 1
##    Data: tall.nonmissing[tall.nonmissing$Rubric == "InterpRes", ]
##
## 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
## as.factor(Rater)1  2.70421    0.08912   30.34
## as.factor(Rater)2  2.58574    0.08912   29.01
## as.factor(Rater)3  2.13918    0.09027   23.70
##
## Correlation of Fixed Effects:
##             a.(R)1 a.(R)2
## as.fctr(R)2 0.061
## as.fctr(R)3 0.062  0.062
```

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

```r
comb.0 <- lmer(as.numeric(Rating) ~ 1 + (0 + Rubric | Artifact),
               data=tall.nonmissing)
```

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

```r
summary(comb.0)
```

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

```r
comb.full <- update(comb.0, . ~ . + as.factor(Rater) + Semester +
                      Sex + Repeated + Rubric)

summary(comb.full)
```

```
## Linear mixed model fit by REML ['lmerMod']
## Formula: as.numeric(Rating) ~ (0 + Rubric | Artifact) + as.factor(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
## as.factor(Rater)2  0.001977   0.054887   0.036
## as.factor(Rater)3 -0.174867   0.055045  -3.177
## SemesterS19        -0.175017   0.087850  -1.992
```

```
## SexM               0.010506   0.081271   0.129
## Repeated          -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) a.(R)2 a.(R)3 SmsS19 SexM   Repetd RbIEDA RbrcIR RbrcRQ
## as.fctr(R)2 -0.245
## as.fctr(R)3 -0.237  0.499
## SemesterS19 -0.361  0.008  0.000
## SexM        -0.398 -0.026 -0.035  0.302
## Repeated    -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
## as.fctr(R)2
## as.fctr(R)3
## SemesterS19
## SexM
## Repeated
## RubrcIntEDA
## RbrcIntrpRs
## RubrcRsrchQ
## RubricSlMth
## RubrcTxtOrg  0.725
## RubricVsOrg  0.680  0.750
```

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

```
## Warning in fitLMER.fnc(comb.full, log.file.name = FALSE): Argument "ran.effects" is empty, which mea
## TRUE

## =========================================================
## ===              backfitting fixed effects            ===
## =========================================================
## processing model terms of interaction level 1
##   iteration 1
##     p-value for term "Sex" = 0.887 >= 0.05
##     not part of higher-order interaction

## boundary (singular) fit: see ?isSingular

##     removing term
##   iteration 2
##     p-value for term "Repeated" = 0.0919 >= 0.05
##     not part of higher-order interaction

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

```
##     removing term
## pruning random effects structure ...
##   nothing to prune
## ======================================================
## ===            forwardfitting random effects       ===
## ======================================================
##  ===      random slopes      ===
## ======================================================
## ===            re-backfitting fixed effects        ===
## ======================================================
## processing model terms of interaction level 1
##   all terms of interaction level 1 significant
## resetting REML to TRUE
```

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

```
## pruning random effects structure ...
##   nothing to prune
```

```
summary(comb.back_elim)
```

```
## Linear mixed model fit by REML ['lmerMod']
## Formula: as.numeric(Rating) ~ (0 + Rubric | Artifact) + as.factor(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
## as.factor(Rater)2  0.0003231  0.0547446   0.006
## as.factor(Rater)3 -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) a.(R)2 a.(R)3 SmsS19 RbIEDA RbrcIR RbrcRQ RbrcSM RbrcTO
## as.fctr(R)2 -0.281
## as.fctr(R)3 -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
```

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

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

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

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

```
summary(comb.inter.u)
```

```
## Linear mixed model fit by REML ['lmerMod']
## Formula: as.numeric(Rating) ~ (0 + Rubric | Artifact) + as.factor(Rater) +
##     Semester + Rubric + as.factor(Rater):Semester + as.factor(Rater):Rubric +
##     Semester:Rubric + as.factor(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
## as.factor(Rater)2                            0.302995   0.155107   1.953
## as.factor(Rater)3                            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
## as.factor(Rater)2:SemesterS19                0.268014   0.303883   0.882
## as.factor(Rater)3:SemesterS19               -0.072789   0.301026  -0.242
## as.factor(Rater)2:RubricInitEDA             -0.325018   0.204108  -1.592
## as.factor(Rater)3:RubricInitEDA             -0.374190   0.205354  -1.822
## as.factor(Rater)2:RubricInterpRes           -0.469281   0.201051  -2.334
## as.factor(Rater)3:RubricInterpRes           -0.711515   0.202316  -3.517
## as.factor(Rater)2:RubricRsrchQ              -0.447050   0.189326  -2.361
## as.factor(Rater)3:RubricRsrchQ              -0.474411   0.190681  -2.488
## as.factor(Rater)2:RubricSelMeth             -0.301450   0.193678  -1.556
## as.factor(Rater)3:RubricSelMeth             -0.365656   0.194970  -1.875
## as.factor(Rater)2:RubricTxtOrg              -0.449164   0.200927  -2.235
## as.factor(Rater)3:RubricTxtOrg              -0.407754   0.202209  -2.016
## as.factor(Rater)2:RubricVisOrg               0.009042   0.205059   0.044
## as.factor(Rater)3: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
## as.factor(Rater)2:SemesterS19:RubricInitEDA    0.020326   0.392376   0.052
## as.factor(Rater)3:SemesterS19:RubricInitEDA    0.252422   0.389961   0.647
## as.factor(Rater)2:SemesterS19:RubricInterpRes -0.266618   0.385390  -0.692
## as.factor(Rater)3:SemesterS19:RubricInterpRes -0.152392   0.383354  -0.398
## as.factor(Rater)2:SemesterS19:RubricRsrchQ    -0.217348   0.360414  -0.603
## as.factor(Rater)3:SemesterS19:RubricRsrchQ     0.354319   0.357388   0.991
## as.factor(Rater)2:SemesterS19:RubricSelMeth   -0.401035   0.370200  -1.083
## as.factor(Rater)3:SemesterS19:RubricSelMeth   -0.192670   0.367887  -0.524
## as.factor(Rater)2:SemesterS19:RubricTxtOrg    -0.542267   0.385011  -1.408
## as.factor(Rater)3:SemesterS19:RubricTxtOrg    -0.316395   0.382614  -0.827
## as.factor(Rater)2:SemesterS19:RubricVisOrg    -0.603626   0.392909  -1.536
## as.factor(Rater)3: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
```

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

```
## Warning in fitLMER.fnc(comb.inter.u, log.file.name = FALSE): Argument "ran.effects" is empty, which r
## TRUE
```

```
## ======================================================
## ===              backfitting fixed effects         ===
## ======================================================
## processing model terms of interaction level 3
##   iteration 1
##     p-value for term "as.factor(Rater):Semester:Rubric" = 0.5526 >= 0.05
##     not part of higher-order interaction
## boundary (singular) fit: see ?isSingular
##     removing term
## processing model terms of interaction level 2
##   iteration 2
##     p-value for term "as.factor(Rater):Semester" = 0.598 >= 0.05
##     not part of higher-order interaction
## boundary (singular) fit: see ?isSingular
##     removing term
##   iteration 3
##     p-value for term "Semester:Rubric" = 0.0761 >= 0.05
##     not part of higher-order interaction
## boundary (singular) fit: see ?isSingular
##     removing term
## processing model terms of interaction level 1
##   all terms of interaction level 1 significant
## pruning random effects structure ...
##   nothing to prune
## ======================================================
## ===            forwardfitting random effects       ===
## ======================================================
##  ===       random slopes       ===
## ======================================================
## ===            re-backfitting fixed effects        ===
## ======================================================
## processing model terms of interaction level 2
##   all terms of interaction level 2 significant
## processing model terms of interaction level 1
##   all terms of interaction level 1 significant
## resetting REML to TRUE
## boundary (singular) fit: see ?isSingular
## pruning random effects structure ...
##   nothing to prune
```

```
summary(comb.inter_elim)
```

```
## Linear mixed model fit by REML ['lmerMod']
## Formula: as.numeric(Rating) ~ (0 + Rubric | Artifact) + as.factor(Rater) +
##     Semester + Rubric + as.factor(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
## as.factor(Rater)2                  0.36537    0.13296   2.748
## as.factor(Rater)3                  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
## as.factor(Rater)2:RubricInitEDA   -0.30843    0.17249  -1.788
## as.factor(Rater)3:RubricInitEDA   -0.29522    0.17282  -1.708
## as.factor(Rater)2:RubricInterpRes -0.53674    0.17008  -3.156
## as.factor(Rater)3:RubricInterpRes -0.75247    0.17049  -4.414
## as.factor(Rater)2:RubricRsrchQ    -0.50157    0.16151  -3.106
## as.factor(Rater)3:RubricRsrchQ    -0.37068    0.16179  -2.291
## as.factor(Rater)2:RubricSelMeth   -0.39602    0.16467  -2.405
## as.factor(Rater)3:RubricSelMeth   -0.41324    0.16504  -2.504
## as.factor(Rater)2:RubricTxtOrg    -0.58380    0.17141  -3.406
## as.factor(Rater)3:RubricTxtOrg    -0.48649    0.17177  -2.832
## as.factor(Rater)2:RubricVisOrg    -0.14444    0.17442  -0.828
## as.factor(Rater)3: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
```

```r
formula(comb.inter.u)
```

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

```r
formula(comb.inter_elim)
```

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

```
##     Semester + Rubric + as.factor(Rater):Rubric
```

```r
formula(comb.back_elim)
```

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

```r
summary(comb.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
```

```r
summary(comb.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
```

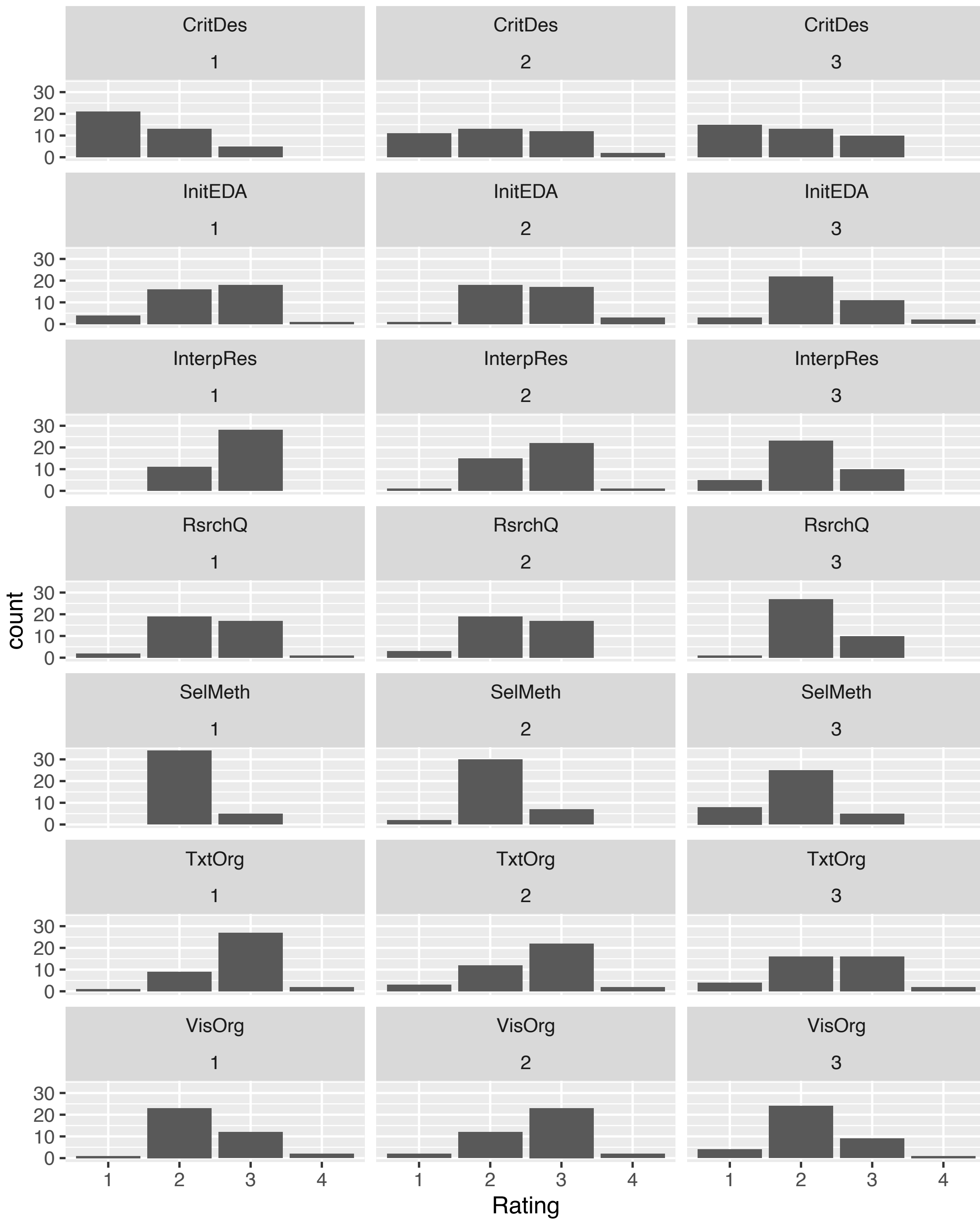
```r
summary(comb.back_elim)$varcor
```

```
##  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
```

```r
anova(comb.back_elim,comb.inter_elim,comb.inter.u)
```

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

## Data: tall.nonmissing
## Models:
## comb.back_elim: as.numeric(Rating) ~ (0 + Rubric | Artifact) + as.factor(Rater) + Semester + Rubric
## comb.inter_elim: as.numeric(Rating) ~ (0 + Rubric | Artifact) + as.factor(Rater) + Semester + Rubric
## comb.inter.u: as.numeric(Rating) ~ (0 + Rubric | Artifact) + as.factor(Rater) + Semester + Rubric + a
##                 npar    AIC    BIC  logLik deviance  Chisq Df Pr(>Chisq)
## comb.back_elim    39 1464.0 1647.2 -693.02   1386.0
## comb.inter_elim   51 1454.5 1694.1 -676.26   1352.5 33.526 12   0.000801 ***
## comb.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
```

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

g
```



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

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

```r
anova(m0,mA)
```

```
## 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) + as.factor(Rater) + Semester + Rubric + as.factor(I
## mA: as.numeric(Rating) ~ (0 + Rubric | Artifact) + (0 + as.factor(Rater) | Artifact) + as.factor(Rate
##    npar    AIC    BIC  logLik deviance  Chisq Df Pr(>Chisq)
## m0   51 1454.5 1694.1 -676.26   1352.5
## mA   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
```

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

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

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

```r
anova(m0,mA)
```

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

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

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

```
## Error: number of observations (=810) <= number of random effects (=1890) for term (0 + as.factor(Rate
```

```r
comb.final <- lmer(as.numeric(Rating) ~ (0 + Rubric | Artifact) +
           (0 + as.factor(Rater) | Artifact) + as.factor(Rater) +
           Semester + Rubric + as.factor(Rater):Rubric, data=tall.nonmissing)
```

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

```r
formula(comb.final)
```

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

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

```
##  Groups     Name               Std.Dev. Corr
##  Artifact   RubricCritDes      0.70456
##             RubricInitEDA      0.56385   0.318
##             RubricInterpRes    0.31953   0.142  0.674
##             RubricRsrchQ       0.42309   0.500  0.194  0.538
##             RubricSelMeth      0.19564   0.145  0.227  0.376 -0.240
##             RubricTxtOrg       0.50029   0.268  0.437  0.364  0.305  0.213
##             RubricVisOrg       0.48201   0.175  0.504  0.445  0.276 -0.160
##  Artifact.1 as.factor(Rater)1 0.11309
##             as.factor(Rater)2 0.33421  -0.488
##             as.factor(Rater)3 0.30670   0.330  0.663
##  Residual                      0.36700
##
##
##
##
##
##
##
##   0.537
##
##
##
##
```

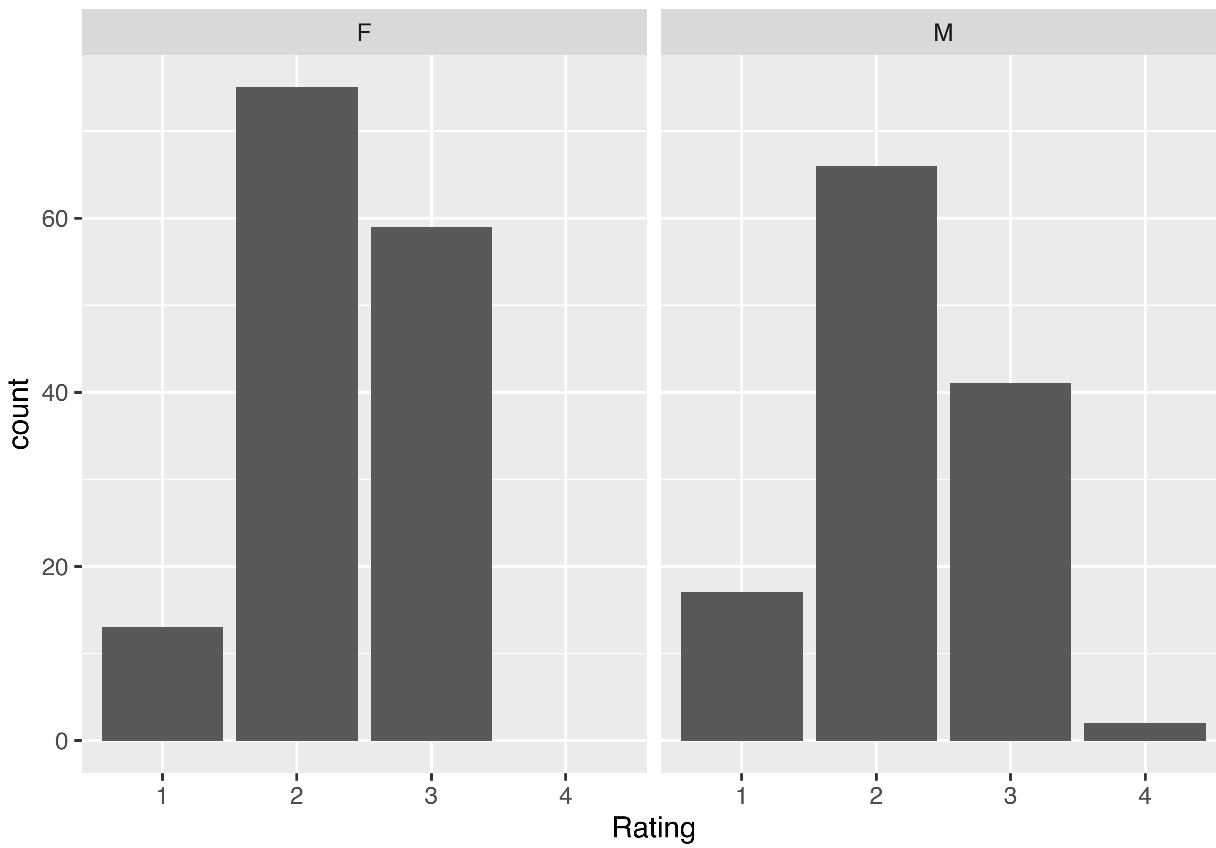
```r
summary(comb.final)$coef
```

```
##                                    Estimate Std. Error    t value
## (Intercept)                       1.7575675 0.11403884 15.4120075
## as.factor(Rater)2                 0.3660512 0.13918262  2.6300063
## as.factor(Rater)3                 0.1958650 0.12967617  1.5104163
## SemesterS19                      -0.1591929 0.07647446 -2.0816477
## RubricInitEDA                     0.7394806 0.12996198  5.6899761
## RubricInterpRes                   0.9915166 0.12771096  7.7637555
## RubricRsrchQ                      0.7261861 0.11792862  6.1578445
## RubricSelMeth                     0.4106681 0.12470221  3.2931906
## RubricTxtOrg                      1.0157886 0.12999521  7.8140465
## RubricVisOrg                      0.6542550 0.13353206  4.8996095
## as.factor(Rater)2:RubricInitEDA  -0.2997977 0.15609303 -1.9206348
## as.factor(Rater)3:RubricInitEDA  -0.2946987 0.15635429 -1.8848136
## as.factor(Rater)2:RubricInterpRes -0.5132368 0.15349003 -3.3437796
## as.factor(Rater)3:RubricInterpRes -0.7148456 0.15364513 -4.6525755
## as.factor(Rater)2:RubricRsrchQ   -0.4874143 0.14722200 -3.3107438
## as.factor(Rater)3:RubricRsrchQ   -0.3223763 0.14726598 -2.1890751
## as.factor(Rater)2:RubricSelMeth  -0.3863680 0.15031029 -2.5704694
## as.factor(Rater)3:RubricSelMeth  -0.3871301 0.14961676 -2.5874779
```

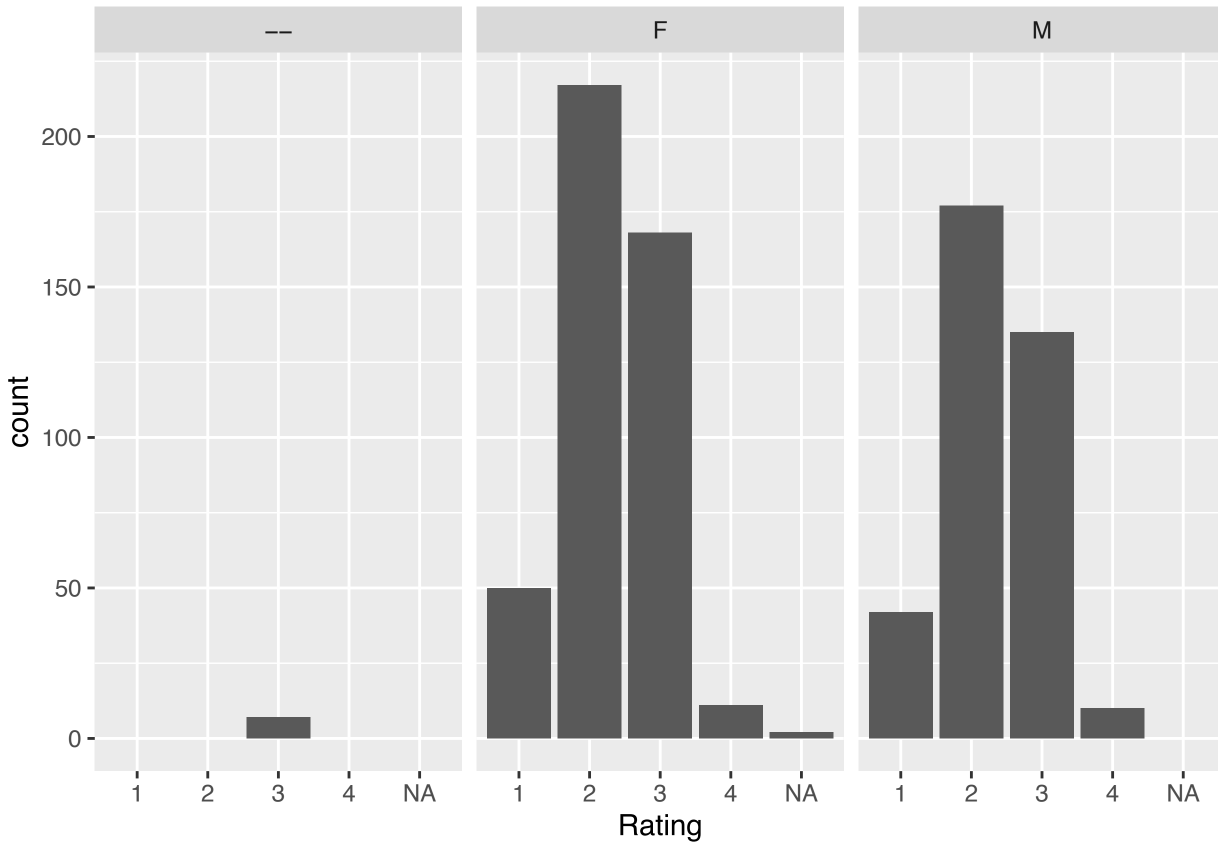
```
## as.factor(Rater)2:RubricTxtOrg    -0.5510564 0.15646236 -3.5219741
## as.factor(Rater)3:RubricTxtOrg    -0.4448931 0.15673326 -2.8385369
## as.factor(Rater)2:RubricVisOrg    -0.1049122 0.15861363 -0.6614326
## as.factor(Rater)3:RubricVisOrg    -0.2752225 0.15885162 -1.7325758
```

Question 4 Is the sex a influential factor in this experiment?

```r
g <- ggplot(tall.13,aes(x = Rating)) +
  facet_wrap( ~ Sex) +
  geom_bar()

g
```




```r
##
## Barplots for full data set
g <- ggplot(tall,aes(x = Rating)) +
  facet_wrap( ~ Sex) +
  geom_bar()

g
```


––
F
M
count
200
150
100
50
0
1
2
3
4
NA
Rating