Analysis on Ratings for Freshman Statistics Procjects

Zhuoheng Han
Department of Statistics and Data Science, Carnegie Mellon University
zhouhenh@andrew.cmu.edu

Abstract

In this paper, we explore the ratings for Freshman Statistics projects from Carnegie Mellon University Dietrich College. We used two data sets, ratings and tall, to help us analyze the ratings. Methods such as exploratory data analysis, intraclass correlation, 2-way table, and multilevel multinomial logit model are used. We find that distributions of ratings for each rubrics or each rater are not indistinguishable from the distributions of other rubrics or other raters; discover that ratings given by raters are mostly not agree with others' ratings... In order to improve our analysis, we can investigate the missing values by checking the student records.

Introduction

Dietrich College at Carnegie Mellon University is in the process of implementing a new "General Education" program for undergraduates. In order to determine whether the new program is successful, the college hopes to rate student work performed. In this paper, we are discussing how ratings are related to various factors in this experiment such as rater, semester, sex and 7 rating rubrics from ratings data set and tall data set. We address four research questions:

- Is the distribution of ratings for each rubrics pretty much indistinguishable from the other rubrics? Is the distribution of ratings given by each rater pretty much indistinguishable from the other raters?
- 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?
- How are the various factors in this experiment (Rater, Semester, Sex, Repeated, Rubric) related to the ratings? Do the factors interact in any interesting ways?
- Undecided (For two semesters, is the distribution of ratings indistinguishable?)

Data

The rating dataset contains total 15 columns and 117 rows. Each line of the dataset provides information for one rating piece. The definition of each variable is given below:

1. (X): Row number in the data set
2. Rater: Which of the three raters gave a rating
3. (Sample): Sample number
4. (Overlap): Unique identifier for artifact seen by all 3 raters Which semester the
5. Semester: Which semester the artifact came from
6. Sex: Sex of student who created the artifact Rating on Research
7. RsrchQ: Rating on Research Question
8. CritDes: Rating on Critique Design
9. InitEDA: Rating on Initial EDA
10. SelMeth: Rating on Select Method(s)
11. InterpRes: Rating on Interpret Results
12. VisOr: Rating on Visual Organization
13. TxtOrg : Rating on Text Organization
14. Artifact: Unique identifier for each artifact
15. Repeated: 1 = this is one of the 13 artifacts seen by all 3 raters

The tall data set contains the same data, but organized so that each row contains just one rating, in the column labelled Rating, and the rubric for that rating is listed in the column labelled Rubric. In this way, the dimension of the data set is 819×8.

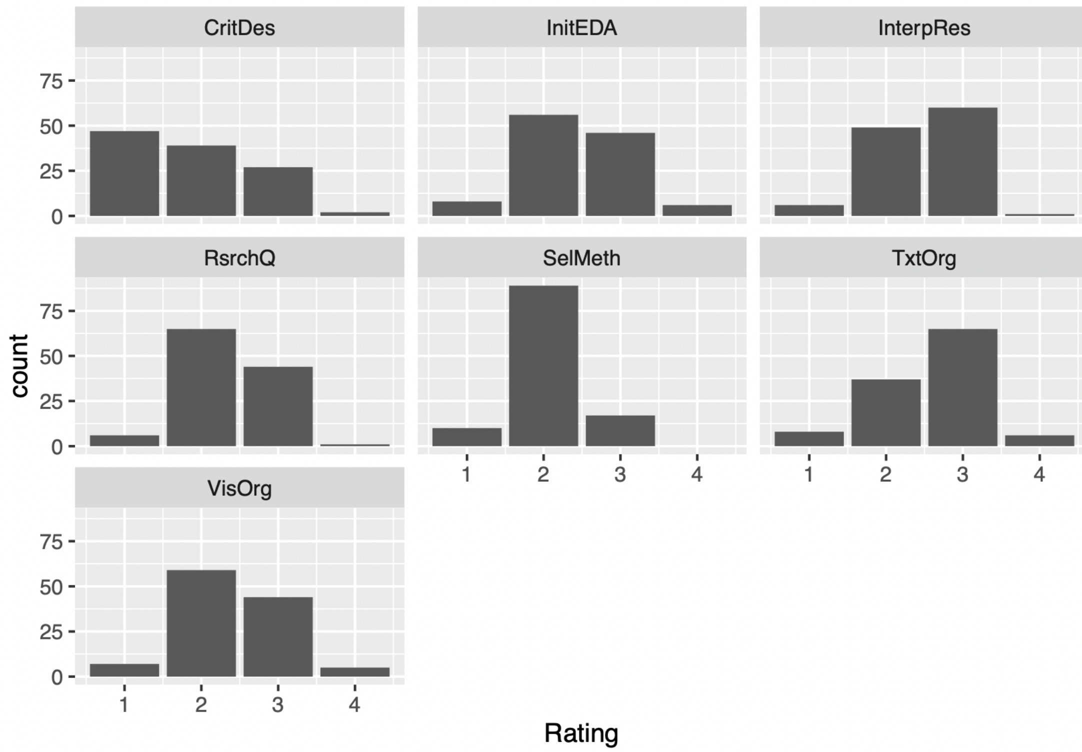
Variables X, Sample, and Overlap are unique, which means we can ignore these these variables when doing data analysis.
Below are the summary tables for category variables Rater, Semester, Sex, Repeated:

```
$Rater

 1  2  3
39 39 39

$Semester

  Fall Spring
    83     34

$Sex

--  F  M
 1 64 52

$Repeated

 0  1
78 39
```

Below are the summary statistics for numeric variables:

```
    RsrchQ        CritDes         InitEDA         SelMeth         InterpRes
 Min.   :1.00   Min.   :1.000   Min.   :1.000   Min.   :1.000   Min.   :1.000
 1st Qu.:2.00   1st Qu.:1.000   1st Qu.:2.000   1st Qu.:2.000   1st Qu.:2.000
 Median :2.00   Median :2.000   Median :2.000   Median :2.000   Median :3.000
 Mean   :2.35   Mean   :1.871   Mean   :2.436   Mean   :2.068   Mean   :2.487
 3rd Qu.:3.00   3rd Qu.:3.000   3rd Qu.:3.000   3rd Qu.:2.000   3rd Qu.:3.000
 Max.   :4.00   Max.   :4.000   Max.   :4.000   Max.   :3.000   Max.   :4.000
                NA's   :1
     VisOrg          TxtOrg
 Min.   :1.000   Min.   :1.000
 1st Qu.:2.000   1st Qu.:2.000
 Median :2.000   Median :3.000
 Mean   :2.414   Mean   :2.598
 3rd Qu.:3.000   3rd Qu.:3.000
 Max.   :4.000   Max.   :4.000
 NA's   :1
```

From the summary statistics, there are some NA's in the data sets, two missing rating values and one missing sex information. Since we are focusing on ratings and cannot find another group for sex, so we decide to drop those NA's. After removing missing values, we make bar plots of counts of each rating for each rubric. We can find that for critique design, the most ratings are 1, which are totally different with other rubrics. For other rubrics except critique design, the top 2 ratings are 2 and 3 and the count of ratings 1 and 4 are relatively low.



Methods

Distribution of ratings by rubrics and raters

We use exploratory data analysis, mainly plotting bar plots. We make a subset of the data for just the 13 artifacts seen by all three raters, and make bar plots of counts of ratings for each rubric based on the subset of the data. We compare with the bar plots for the whole data set shown in the data section and determine whether these thirteen artifacts are representative of the whole set of 91 artifacts.

Agreement on scores

We treat each artifact as a cluster of three ratings, and fit the random-intercept model on repeated data set and calculate intraclass correlations. To find which raters might be contributing to disagreement, we make 2-way tables of counts for the ratings of each pair of raters on each rubric. Then, we can calculate the agreement, which equals to diagonal sum of the table / total sum.

Factors related to the ratings

We add fixed effects for rater, semester, sex to the random intercept models for the full data set and do variable selection. In order to explore interactions with rubric, we begin with the model Rating$\tilde{(}0+\text{Rubric}|\text{Artifact})$, and then add fixed effects (and possibly interactions) for all of the variables rater, semester, sex, repeated and/or rubric.

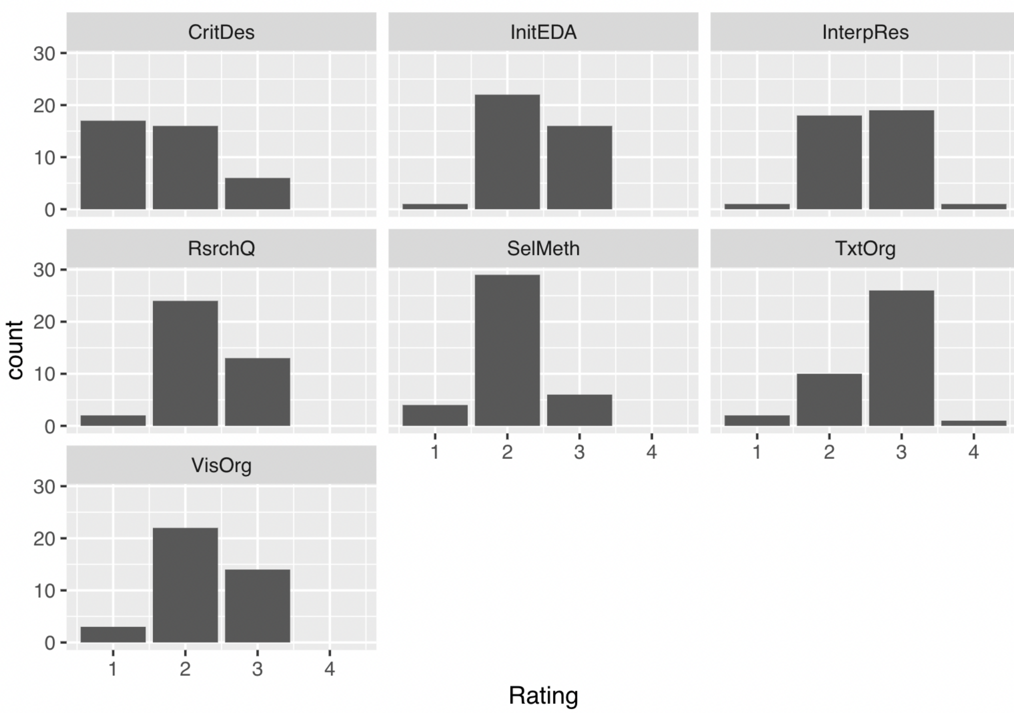
Undecided (Semester?)

Make plots and perform exploratory data analysis.

Results

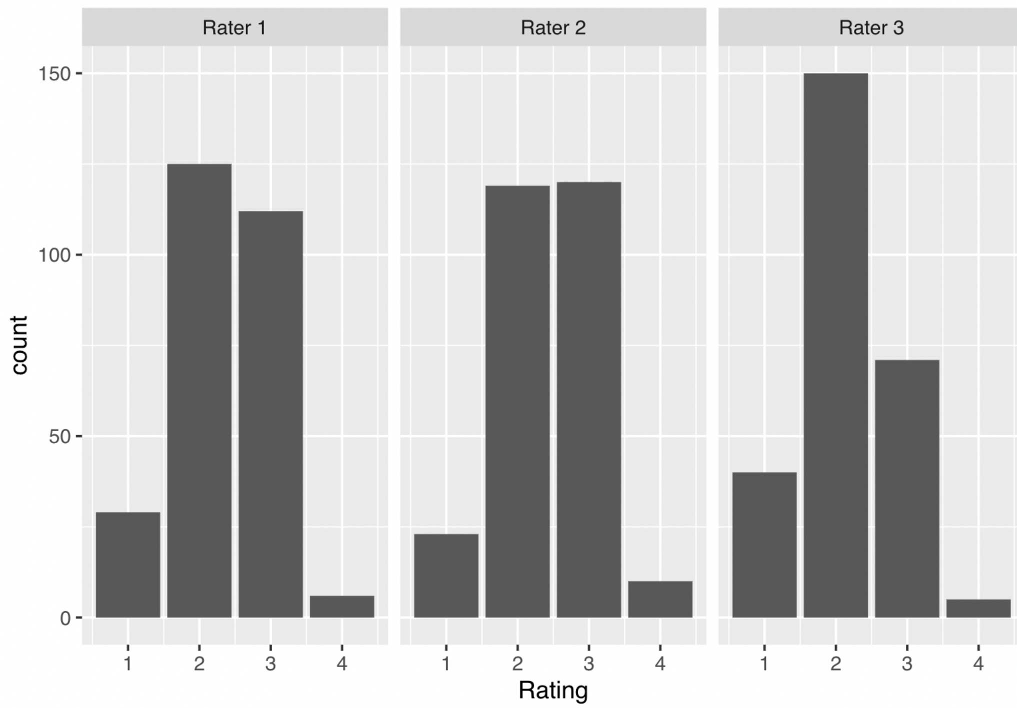
Distribution of ratings by rubrics and raters

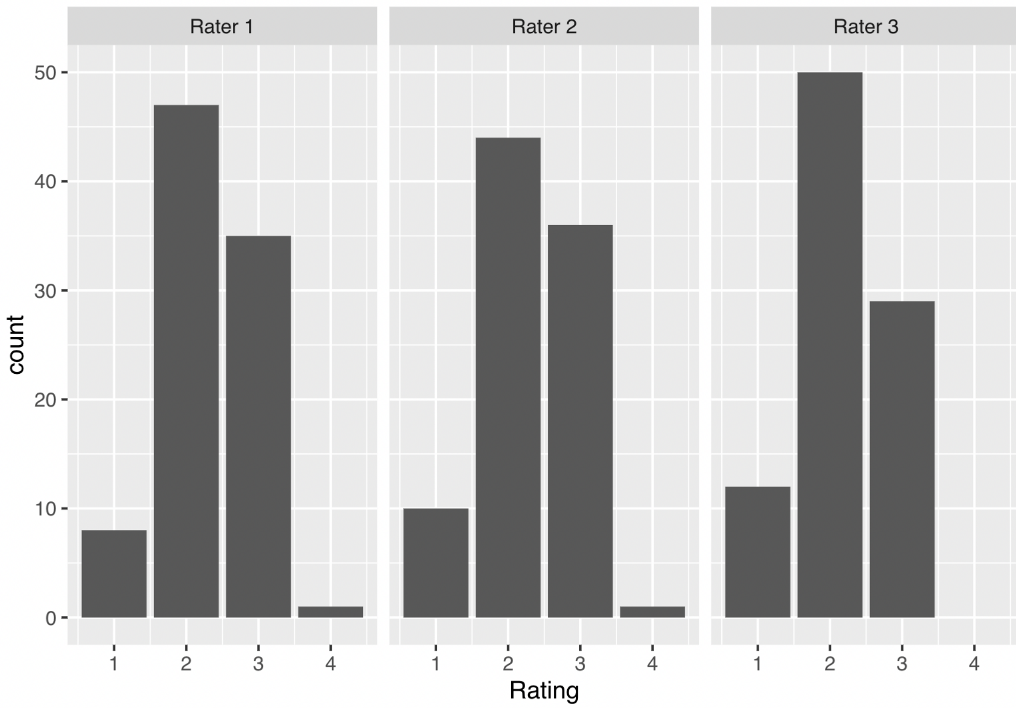
First of all, we filter out 13 artifacts rated by all three raters. Here are the summary statistics and bar plots of ratings for each rubric based on repeated artifacts. Compared with the summary statistics and bar plots in the Data Section, we can find distributions are quite similar for each rubric between the full data set and subset of data set.



For critique design, the most ratings are 1, which are totally different with other rubrics. For rubrics initial EDA, research question, and visual organization, the ratings distribution are quite similar, where counts of rating 2 are the most and counts of rating 3 are quite close to counts rating 2. For interpretation results and text organization, the ratings distribution are quite similar, where counts of rating 3 are the most and then counts of rating 2. As for select methods, the ratings are mostly 2. Counts of rating 1,3, and 4 are extremely fewer than rating 2. Thus, distributions of ratings for each rubrics are not indistinguishable from the distributions of other rubrics.

In order to find the distribution of ratings given by different raters, we also make bar plots based on full data set and repeated data set. Compared with two sets of bar plots, we can find distributions of ratings are quite similar for each rater.





For the rater 1 and 2, the most ratings they give are 2 and 3 and counts are close. For the rater 3, he/she gives rating 2 most, which is about one time more than rating 3. In this way, distributions of ratings for each rater are not indistinguishable from the distributions of other raters.

Agreement on scores

Since we want to find out whether raters generally agree on their scores for each rubrics, we focus on the 13 artifacts rated by all 3 raters. Below is a table including ICC (intraclass correlation), agreement12 (agreement rate on rater 1 and 2), agreement13 (agreement rate on rater 1 and 3), and agreement23 (agreement rate on rater 2 and 3) for each rubric.

	ICC	agreement12	agreement13	agreement23
RsrchQ	0.1891892	0.3846154	0.7692308	0.5384615
CritDes	0.5725594	0.5384615	0.6153846	0.6923077
InitEDA	0.4929577	0.6923077	0.5384615	0.8461538
SelMeth	0.5212766	0.9230769	0.6153846	0.6923077
InterpRes	0.2295720	0.6153846	0.5384615	0.6153846
VisOrg	0.5924529	0.5384615	0.7692308	0.7692308
TxtOrg	0.1428571	0.6923077	0.6153846	0.5384615

We can find that ICCs of rubric research question, interpret result, and text organization are relatively low and ICCs of rubric critical design, initial EDA, select results, and visual organization are relatively higher, but they are not very high.
For research questions,

Factors related to the ratings

To start with, we add fixed effects to the seven rubric-specific models using repeated data. Our seven models start with

$$as.numeric(Rating) \sim -1 + as.factor(Rater) + Semester + Sex + (1|Artifact).$$

We apply backward elimination to each model and here are the final models for each rubric. From the result below, we don't need to add any fixed effects or interactions to the models for each rubric when using 13 artifacts data set.

```
$RsrchQ
as.numeric(Rating) ~ (1 | Artifact)

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

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

$SelMeth
as.numeric(Rating) ~ (1 | Artifact)

$InterpRes
as.numeric(Rating) ~ (1 | Artifact)

$VisOrg
as.numeric(Rating) ~ (1 | Artifact)

$TxtOrg
as.numeric(Rating) ~ (1 | Artifact)
```

Then, we add fixed effects to the seven rubric-specific models using full data. Same procedure as above, we also start with the model

$$as.numeric(Rating) \sim -1 + as.factor(Rater) + Semester + Sex + (1|Artifact).$$

We apply backward elimination to each model and here are the final models for each rubric. From the result below, we don't need to add any fixed effects or interactions to the models for rubric research questions, initial EDA, and text organization. As for other rubrics, we need to examine each of these 4 models to see if the fixed effects make sense to us and if there are any interactions or additional random effects to consider.

```
$RsrchQ
as.numeric(Rating) ~ (1 | Artifact)

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

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

$SelMeth
as.numeric(Rating) ~ as.factor(Rater) + Semester + (1 | Artifact) -
    1

$InterpRes
as.numeric(Rating) ~ as.factor(Rater) + (1 | Artifact) - 1

$VisOrg
as.numeric(Rating) ~ as.factor(Rater) + (1 | Artifact) - 1

$TxtOrg
as.numeric(Rating) ~ (1 | Artifact)
```

We try to add interactions and new random effects for those 4 models using all data set. As for critical design, we applied ANOVA on fixed effects model, intercept model, and fixed interaction model to check which model performs the best. Based on the ANOVA result as shown below, we can find that

$$as.numeric(Rating) \sim -1 + as.factor(Rater) + (1|Artifact)$$

with the smallest p-value, which means this model performs best among those three.

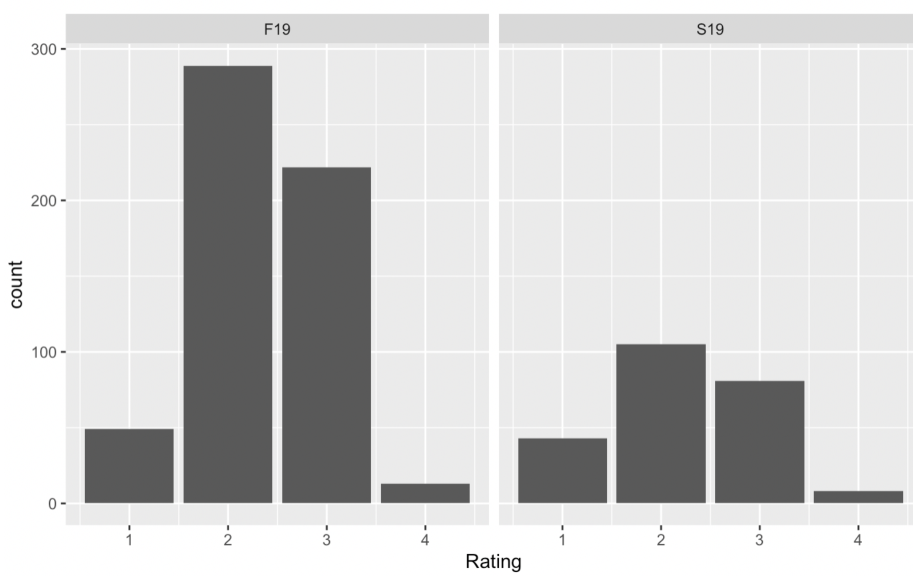
```
Data: tall[tall$Rubric == "CritDes", ]
Models:
tmp.single_intercept: as.numeric(Rating) ~ (1 | Artifact)
tmp: as.numeric(Rating) ~ as.factor(Rater) + (1 | Artifact) - 1
tmp.fixed_interactions: as.numeric(Rating) ~ as.factor(Rater) + (1 | Artifact) +
as.factor(Rater):Semester - 1
                       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 *
tmp.fixed_interactions    8 277.32 299.28 -130.66   261.32 2.3036  3    0.51183
---
Signif. codes:  0 ‘***’ 0.001 ‘**’ 0.01 ‘*’ 0.05 ‘.’ 0.1 ‘ ’ 1
```

Then, from the summary statistic of this model, we can find all coefficients are significant.


```
Data: tall[tall$Rubric == "CritDes", ]
Models:
tmp.single_intercept: as.numeric(Rating) ~ (1 | Artifact)
tmp: as.numeric(Rating) ~ as.factor(Rater) + (1 | Artifact) - 1
tmp.fixed_interactions: as.numeric(Rating) ~ as.factor(Rater) + (1 | Artifact) +
as.factor(Rater):Semester - 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 *
tmp.fixed_interactions    8 277.32 299.28 -130.66   261.32 2.3036  3    0.51183
---
Signif. codes:  0 ‘***’ 0.001 ‘**’ 0.01 ‘*’ 0.05 ‘.’ 0.1 ‘ ’ 1
```


Undecided

There is indistinguishable distribution of ratings for two semesters.



Discussion

There are 2 NA's in ratings and 1 missing sex in our data sets. In this paper, we dropped those 2 ratings, which might cause the imbalance of data set since two students will only have 6 rubrics rated. In addition, we also dropped all data related to the freshman with the missing sex. The best way to address the missing values is to investigate the records on those students. In this way, we can fill out those NA's and get the full data set.

Appendix

```r
library(lme4)
library(arm)
library(ggplot2)
library(tidyverse)
library(LMERConvenienceFunctions)
library(RLRsim)
```

Import the dataset and delete variables that are not expected to be useful for analysis. Then, we drop NA's.

```r
ratings = read.csv("~/Desktop/ratings.csv")
tall = read.csv("~/Desktop/tall.csv")
```

```r
dim(ratings)
```

```
## [1] 117  15
```

```r
dim(tall)
```

```
## [1] 819   8
```

```r
ratings = ratings[-c(1,3,4)]
```

See the summary statistics of the numeric variable.

```r
summary(ratings[- which(names(ratings) %in% c("Rater", "Semester", "Sex", "Artifact", "Repeated"))])
```

```
##      RsrchQ        CritDes         InitEDA         SelMeth         InterpRes    
##  Min.   :1.00   Min.   :1.000   Min.   :1.000   Min.   :1.000   Min.   :1.000  
##  1st Qu.:2.00   1st Qu.:1.000   1st Qu.:2.000   1st Qu.:2.000   1st Qu.:2.000  
##  Median :2.00   Median :2.000   Median :2.000   Median :2.000   Median :3.000  
##  Mean   :2.35   Mean   :1.871   Mean   :2.436   Mean   :2.068   Mean   :2.487  
##  3rd Qu.:3.00   3rd Qu.:3.000   3rd Qu.:3.000   3rd Qu.:2.000   3rd Qu.:3.000  
##  Max.   :4.00   Max.   :4.000   Max.   :4.000   Max.   :3.000   Max.   :4.000  
##                 NA's   :1                                                      
##      VisOrg          TxtOrg     
##  Min.   :1.000   Min.   :1.000  
##  1st Qu.:2.000   1st Qu.:2.000  
##  Median :2.000   Median :3.000  
##  Mean   :2.414   Mean   :2.598  
##  3rd Qu.:3.000   3rd Qu.:3.000  
##  Max.   :4.000   Max.   :4.000  
##  NA's   :1
```

```r
apply(ratings[which(names(ratings) %in% c("Rater", "Semester", "Sex", "Repeated"))], 2, table)
```

```
## $Rater
##
##  1  2  3
## 39 39 39
##
## $Semester
##
##   Fall Spring
##     83     34
##
## $Sex
##
## --  F  M
##  1 64 52
##
## $Repeated
##
##  0  1
## 78 39
```

```r
tall[is.na(tall["Rating"]),]
```

```
##       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
```

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

```
##   Rater Semester Sex RsrchQ CritDes InitEDA SelMeth InterpRes VisOrg TxtOrg
## 5     3     Fall  --      3       3       3       3         3      3      3
##   Artifact Repeated
## 5        5        0
```

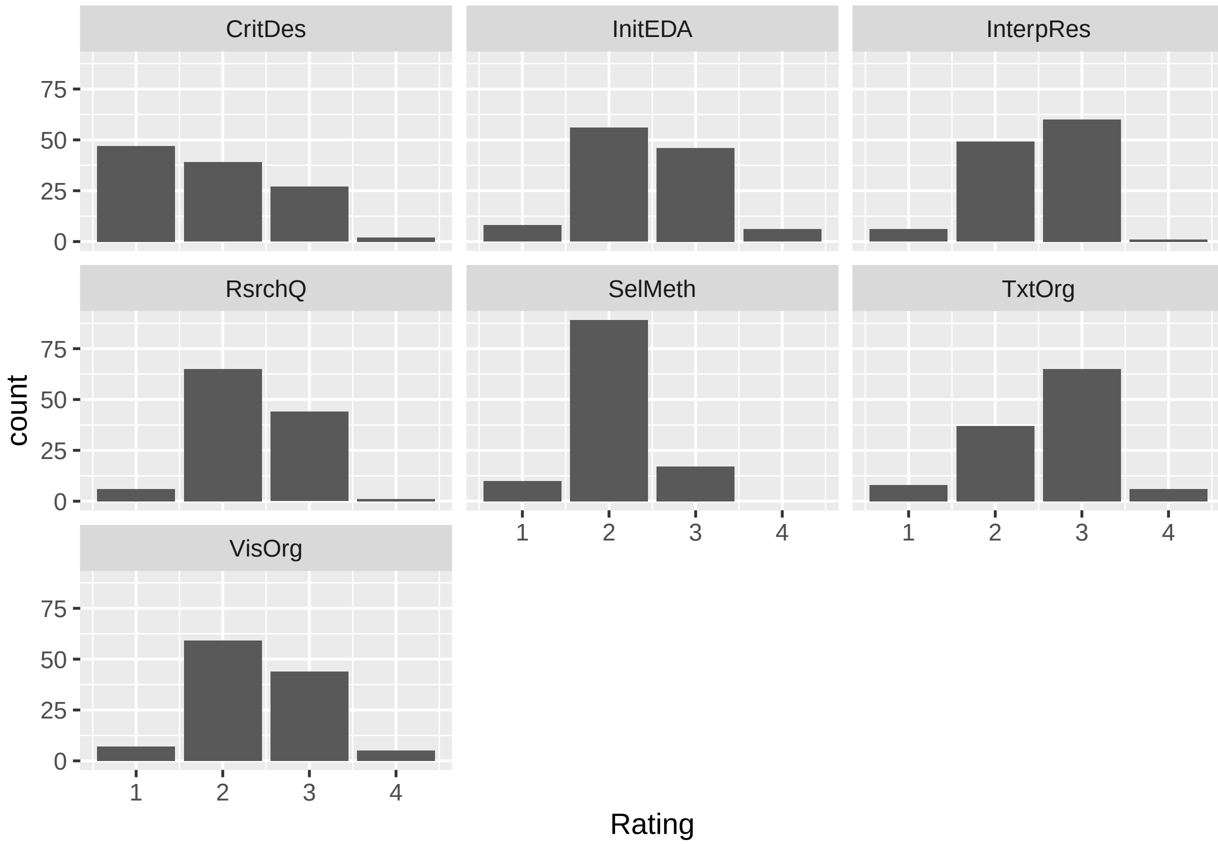
```r
ratings <- ratings %>% drop_na()
ratings <- ratings[-c(5),]
```

```r
tall <- tall %>% drop_na()
tall <- tall[-c(5,122,238,355,472,589,705),]
```

```r
ratings.repeat = ratings[ratings$Repeated == 1,]
```

```r
tall.repeat <- tall[grep("O",tall$Artifact),]
```

```r
ggplot(tall,aes(x = Rating)) +
  facet_wrap( ~ Rubric) +
  geom_bar()
```

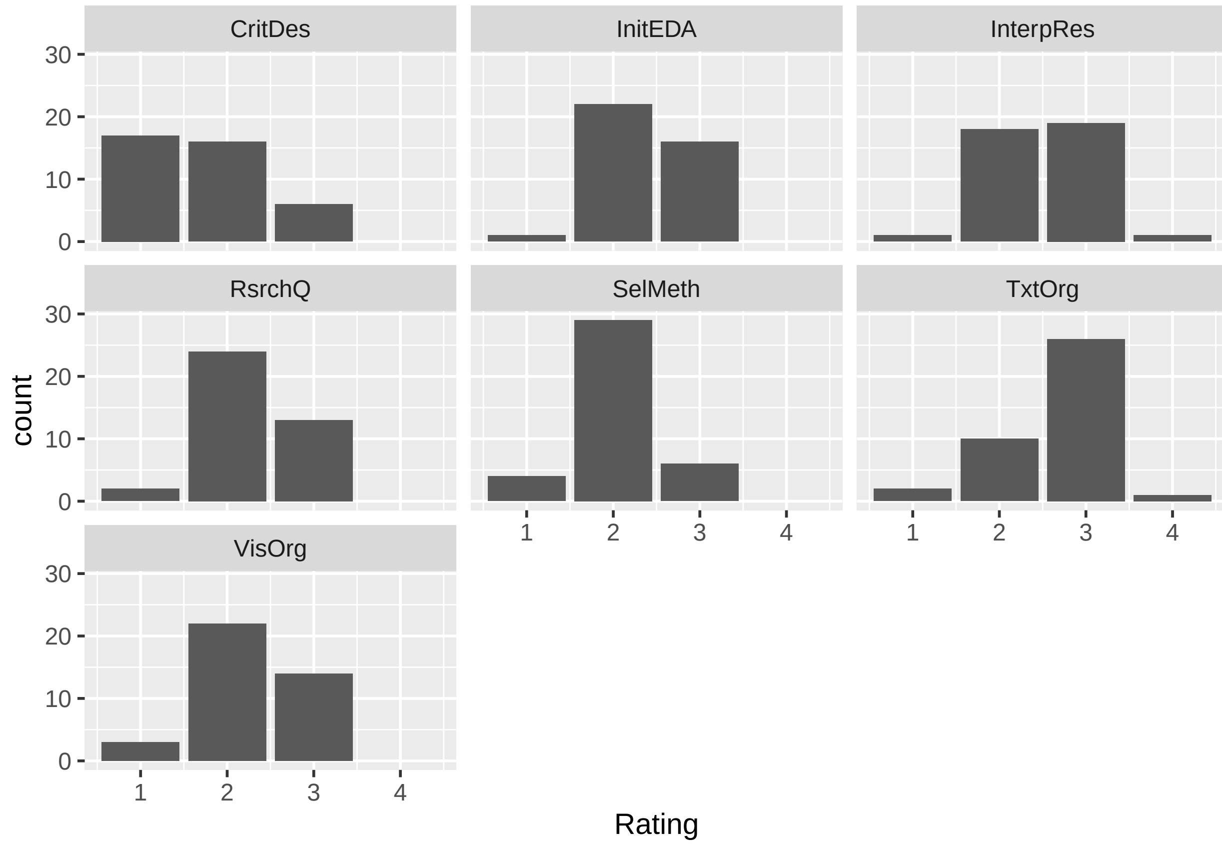


Question 1

```
summary(ratings.repeat[- which(names(ratings.repeat) %in% c("Rater", "Semester", "Sex", "Artifact", "Rep
```

```
##       RsrchQ          CritDes          InitEDA          SelMeth     
##  Min.   :1.000   Min.   :1.000   Min.   :1.000   Min.   :1.000  
##  1st Qu.:2.000   1st Qu.:1.000   1st Qu.:2.000   1st Qu.:2.000  
##  Median :2.000   Median :2.000   Median :2.000   Median :2.000  
##  Mean   :2.282   Mean   :1.718   Mean   :2.385   Mean   :2.051  
##  3rd Qu.:3.000   3rd Qu.:2.000   3rd Qu.:3.000   3rd Qu.:2.000  
##  Max.   :3.000   Max.   :3.000   Max.   :3.000   Max.   :3.000  
##    InterpRes         VisOrg          TxtOrg     
##  Min.   :1.000   Min.   :1.000   Min.   :1.000  
##  1st Qu.:2.000   1st Qu.:2.000   1st Qu.:2.000  
##  Median :3.000   Median :2.000   Median :3.000  
##  Mean   :2.513   Mean   :2.282   Mean   :2.667  
##  3rd Qu.:3.000   3rd Qu.:3.000   3rd Qu.:3.000  
##  Max.   :4.000   Max.   :3.000   Max.   :4.000
```

```
ggplot(tall.repeat,aes(x = Rating)) +
  facet_wrap( ~ Rubric) +
  geom_bar()
```



Below are the summary statistics of rubrics for each rater

```
summary(ratings.repeat[ratings.repeat$Rater == 1,][- which(names(ratings.repeat) %in% c("Semester", "Se
```

```
##      Rater       RsrchQ          CritDes          InitEDA          SelMeth
##  Min.   :1   Min.   :2.000   Min.   :1.000   Min.   :1.000   Min.   :2.000
##  1st Qu.:1   1st Qu.:2.000   1st Qu.:1.000   1st Qu.:2.000   1st Qu.:2.000
##  Median :1   Median :2.000   Median :2.000   Median :3.000   Median :2.000
##  Mean   :1   Mean   :2.385   Mean   :1.615   Mean   :2.538   Mean   :2.154
##  3rd Qu.:1   3rd Qu.:3.000   3rd Qu.:2.000   3rd Qu.:3.000   3rd Qu.:2.000
##  Max.   :1   Max.   :3.000   Max.   :3.000   Max.   :3.000   Max.   :3.000
##    InterpRes         VisOrg          TxtOrg
##  Min.   :2.000   Min.   :1.000   Min.   :2.000
##  1st Qu.:2.000   1st Qu.:2.000   1st Qu.:2.000
##  Median :3.000   Median :2.000   Median :3.000
##  Mean   :2.615   Mean   :2.154   Mean   :2.769
##  3rd Qu.:3.000   3rd Qu.:2.000   3rd Qu.:3.000
##  Max.   :3.000   Max.   :3.000   Max.   :4.000
```

```
summary(ratings.repeat[ratings.repeat$Rater == 2,][- which(names(ratings.repeat) %in% c("Semester", "Se
```

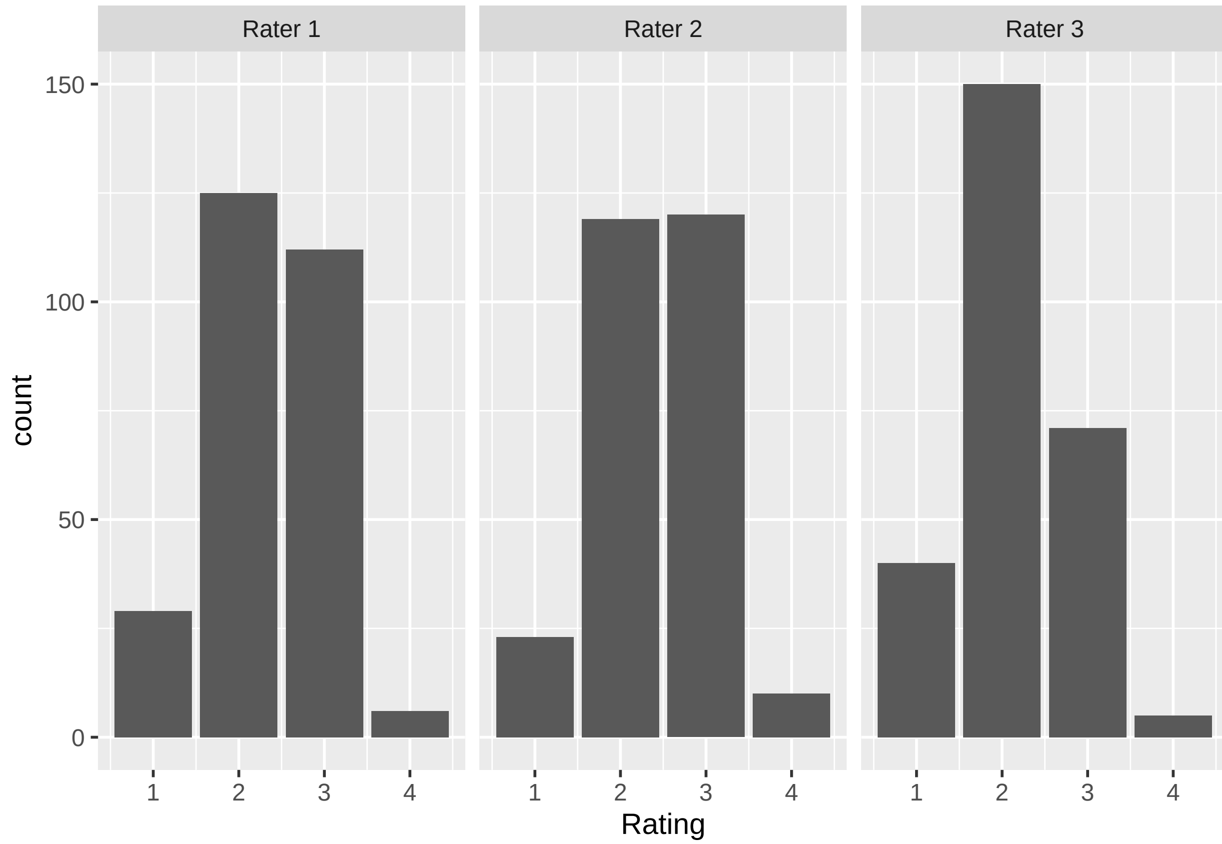
```
##      Rater       RsrchQ          CritDes          InitEDA          SelMeth
##  Min.   :2   Min.   :1.000   Min.   :1.000   Min.   :2.000   Min.   :1.000
##  1st Qu.:2   1st Qu.:2.000   1st Qu.:1.000   1st Qu.:2.000   1st Qu.:2.000
##  Median :2   Median :2.000   Median :2.000   Median :2.000   Median :2.000
```

```
##  Mean   :2   Mean   :2.154   Mean   :1.846   Mean   :2.385   Mean   :2.077
##  3rd Qu.:2   3rd Qu.:3.000   3rd Qu.:2.000   3rd Qu.:3.000   3rd Qu.:2.000
##  Max.   :2   Max.   :3.000   Max.   :3.000   Max.   :3.000   Max.   :3.000
##    InterpRes         VisOrg          TxtOrg
##  Min.   :2.000   Min.   :1.000   Min.   :1.000
##  1st Qu.:2.000   1st Qu.:2.000   1st Qu.:2.000
##  Median :3.000   Median :3.000   Median :3.000
##  Mean   :2.615   Mean   :2.462   Mean   :2.615
##  3rd Qu.:3.000   3rd Qu.:3.000   3rd Qu.:3.000
##  Max.   :4.000   Max.   :3.000   Max.   :3.000
```

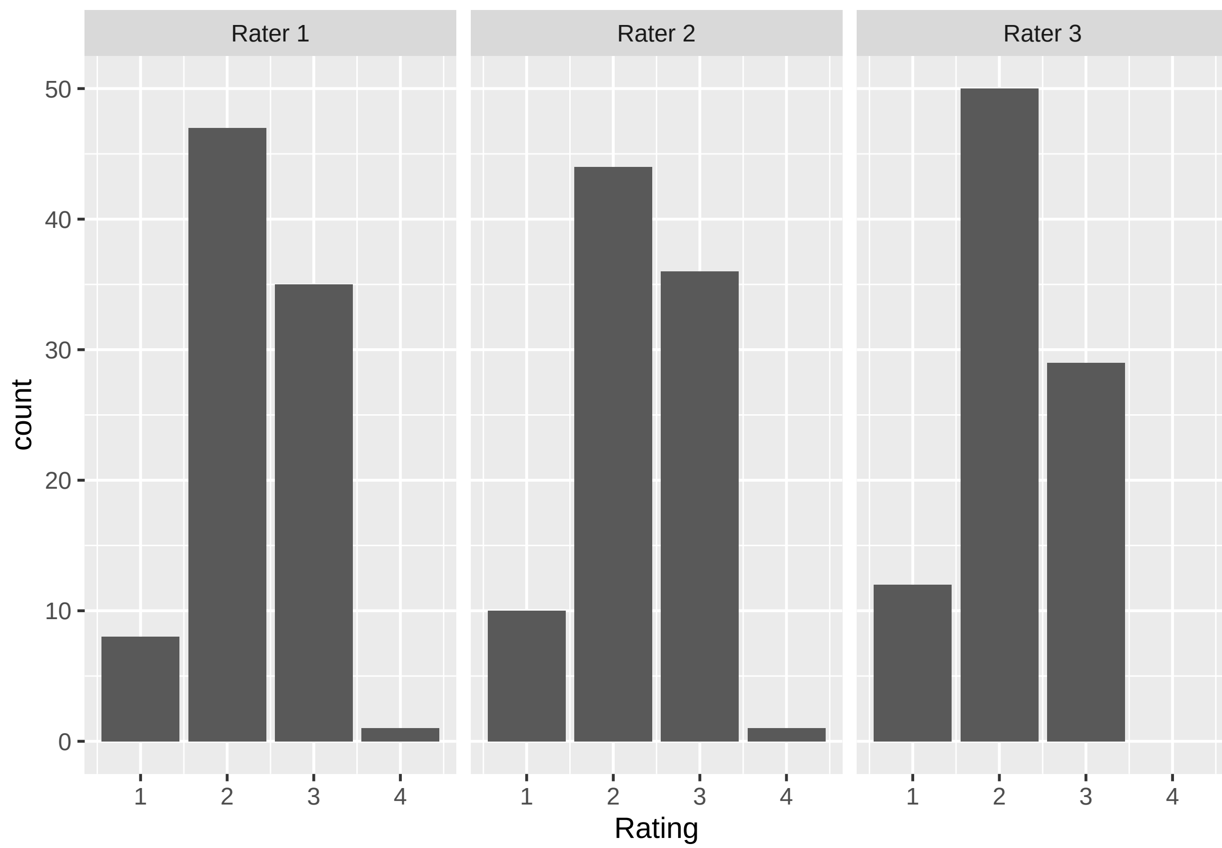
```
summary(ratings.repeat[ratings.repeat$Rater == 3,][- which(names(ratings.repeat) %in% c("Semester", "Se
```

```
##      Rater       RsrchQ          CritDes          InitEDA          SelMeth
##  Min.   :3   Min.   :2.000   Min.   :1.000   Min.   :2.000   Min.   :1.000
##  1st Qu.:3   1st Qu.:2.000   1st Qu.:1.000   1st Qu.:2.000   1st Qu.:2.000
##  Median :3   Median :2.000   Median :2.000   Median :2.000   Median :2.000
##  Mean   :3   Mean   :2.308   Mean   :1.692   Mean   :2.231   Mean   :1.923
##  3rd Qu.:3   3rd Qu.:3.000   3rd Qu.:2.000   3rd Qu.:2.000   3rd Qu.:2.000
##  Max.   :3   Max.   :3.000   Max.   :3.000   Max.   :3.000   Max.   :3.000
##    InterpRes         VisOrg          TxtOrg
##  Min.   :1.000   Min.   :1.000   Min.   :1.000
##  1st Qu.:2.000   1st Qu.:2.000   1st Qu.:2.000
##  Median :2.000   Median :2.000   Median :3.000
##  Mean   :2.308   Mean   :2.231   Mean   :2.615
##  3rd Qu.:3.000   3rd Qu.:3.000   3rd Qu.:3.000
##  Max.   :3.000   Max.   :3.000   Max.   :3.000
```

```
rater.name <- function(x) { paste("Rater",x) }
ggplot(tall,aes(x = Rating)) +
  facet_wrap( ~ Rater, labeller=labeller(Rater=rater.name)) +
  geom_bar()
```




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



Question 2

Then, we can calculate ICC for each rubric as follow.

```
Rubric.names <- c("RsrchQ","CritDes","InitEDA","SelMeth","InterpRes","VisOrg","TxtOrg")
ICC.vec <- NULL
for (i in Rubric.names) {
  tmp <- lmer(as.numeric(Rating) ~ 1+(1|Artifact), data=tall.repeat[tall.repeat$Rubric==i,])
  sig2 <- summary(tmp)$sigma^2
  tau2 <- attr(summary(tmp)$varcor[[1]],"stddev")^2
  ICC <- tau2/(tau2+sig2)
  ICC.vec <- append(ICC.vec,ICC)
}
names(ICC.vec) <- Rubric.names
```

```
agreement.results <- cbind(ICC=ICC.vec, agreement12=0, agreement13=0, agreement23=0)
agreement.tables <- as.list(rep(NA,7))
for (i in Rubric.names) {
  tmp.data = data.frame(r1=ratings.repeat[ratings.repeat$Rater==1, i],
                        r2=ratings.repeat[ratings.repeat$Rater==2, i],
                        r3=ratings.repeat[ratings.repeat$Rater==3, i],
                        a1=ratings.repeat[ratings.repeat$Rater==1, "Artifact"],
                        a2=ratings.repeat[ratings.repeat$Rater==2, "Artifact"],
                        a3=ratings.repeat[ratings.repeat$Rater==3, "Artifact"])

  t1 <- factor(tmp.data$r1,levels=1:4)
  t2 <- factor(tmp.data$r2,levels=1:4)
```

```
  t3 <- factor(tmp.data$r3,levels=1:4)
  t12 <- table(t1,t2)
  agreement12 <- (t12[1,1]+t12[2,2]+t12[3,3]+t12[4,4])/sum(t12)
  t13 <- table(t1,t3)
  agreement13 <- (t13[1,1]+t13[2,2]+t13[3,3]+t13[4,4])/sum(t13)
  t23 <- table(t2,t3)
  agreement23 <- (t23[1,1]+t23[2,2]+t23[3,3]+t23[4,4])/sum(t23)
  agreement.results[i,2:4] <- c(agreement12, agreement13, agreement23)
}
agreement.results
```

```
##                  ICC agreement12 agreement13 agreement23
## RsrchQ    0.1891892   0.3846154   0.7692308   0.5384615
## CritDes   0.5725594   0.5384615   0.6153846   0.6923077
## InitEDA   0.4929577   0.6923077   0.5384615   0.8461538
## SelMeth   0.5212766   0.9230769   0.6153846   0.6923077
## InterpRes 0.2295720   0.6153846   0.5384615   0.6153846
## VisOrg    0.5924529   0.5384615   0.7692308   0.7692308
## TxtOrg    0.1428571   0.6923077   0.6153846   0.5384615
```

Question 3

i) Adding fixed effects to the seven rubric-specific models using repeated data

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

for (i in Rubric.names) {
  rubric.data <- tall.repeat[tall.repeat$Rubric==i,]
  tmp <- lmer(as.numeric(Rating) ~ -1+as.factor(Rater)+Semester+Sex+(1|Artifact),
             data=tall.repeat[tall.repeat$Rubric==i,], REML=FALSE)
  tmp.back_elim <- fitLMER.fnc(tmp, set.REML.FALSE = TRUE, log.file.name = FALSE)
  tmp.single_intercept <-  lmer(as.numeric(Rating) ~ (1|Artifact),
                                  data=tall.repeat[tall.repeat$Rubric==i,],REML=FALSE)
  pval <- anova(tmp.single_intercept,tmp.back_elim)$"Pr(>Chisq)"[2]
  if (pval<=0.05) {
    tmp_final <- tmp.back_elim
  }
  else {
    tmp_final <- tmp.single_intercept
  }

model.formula.repeat[[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" = 0.7355 >= 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

## 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" = 0.2229 >= 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
```

```
## 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" = 0.8137 >= 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

## 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" = 0.9383 >= 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

## 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" = 0.8294 >= 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

## 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" = 0.1922 >= 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

## 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" = 0.5358 >= 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
```

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

```r
model.formula.repeat
```

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

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

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

for (i in Rubric.names) {
  tmp <- lmer(as.numeric(Rating) ~ -1+as.factor(Rater)+Semester+Sex+(1|Artifact),
            data=tall[tall$Rubric==i,], REML=FALSE)
  tmp.back_elim <- fitLMER.fnc(tmp, set.REML.FALSE = TRUE, log.file.name = FALSE)
  tmp.single_intercept <-  lmer(as.numeric(Rating) ~ (1|Artifact),
                              data=tall[tall$Rubric==i,],REML=FALSE)
  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[[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 "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 "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" = 0.1935 >= 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" = 0.608 >= 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 "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)

## 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)
```

```
model.formula
```

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

iii) Trying interactions and new random effects for the 4 rubric specific models using all data set.

CritDes

```
tmp <- lmer(model.formula[["CritDes"]],data=tall[tall$Rubric=="CritDes",])
tmp.single_intercept <- update(tmp, . ~ . + 1 - as.factor(Rater))
tmp.fixed_interactions <- update(tmp, . ~ . + as.factor(Rater)*Semester - Semester)
anova(tmp,tmp.single_intercept,tmp.fixed_interactions)
```

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

## Data: tall[tall$Rubric == "CritDes", ]
## Models:
## tmp.single_intercept: as.numeric(Rating) ~ (1 | Artifact)
## tmp: as.numeric(Rating) ~ as.factor(Rater) + (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.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
## tmp.fixed_interactions    8 277.32 299.28 -130.66   261.32 2.3036  3    0.51183
##
## tmp.single_intercept
## tmp                    *
```

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

```
summary(tmp)
```

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

SelMeth

```
tmp <- lmer(model.formula[["SelMeth"]],data=tall[tall$Rubric=="SelMeth",])
tmp.single_intercept <- update(tmp, . ~ . + 1 - as.factor(Rater))
tmp.fixed_interactions <- update(tmp, . ~ . + as.factor(Rater)*Semester - Semester)
anova(tmp,tmp.single_intercept,tmp.fixed_interactions)
```

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

```
## Data: tall[tall$Rubric == "SelMeth", ]
## Models:
## tmp.single_intercept: as.numeric(Rating) ~ Semester + (1 | Artifact)
## 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.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
## tmp.fixed_interactions    8 143.46 165.49 -63.731   127.46 2.5920  2    0.27362
##
```

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

```
summary(tmp)
```

```
## Linear mixed model fit by REML ['lmerMod']
## Formula: as.numeric(Rating) ~ as.factor(Rater) + Semester + (1 | Artifact) -
##     1
##    Data: tall[tall$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
```

InterpRes

```
tmp <- lmer(model.formula[["InterpRes"]],data=tall[tall$Rubric=="InterpRes",])
tmp.single_intercept <- update(tmp, . ~ . + 1 - as.factor(Rater))
tmp.fixed_interactions <- update(tmp, . ~ . + as.factor(Rater)*Semester - Semester)
anova(tmp,tmp.single_intercept,tmp.fixed_interactions)
```

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

```
## Data: tall[tall$Rubric == "InterpRes", ]
## Models:
## tmp.single_intercept: as.numeric(Rating) ~ (1 | Artifact)
## tmp: as.numeric(Rating) ~ as.factor(Rater) + (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
## 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
## tmp.fixed_interactions     8 203.96 225.99  -93.982   187.96  2.697  3
##                         Pr(>Chisq)
## tmp.single_intercept
## tmp                      1.787e-05 ***
## tmp.fixed_interactions      0.4407
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
```

```
summary(tmp)
```

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

VisOrg

```
tmp <- lmer(model.formula[["VisOrg"]],data=tall[tall$Rubric=="VisOrg",])
tmp.single_intercept <- update(tmp, . ~ . + 1 - as.factor(Rater))
tmp.fixed_interactions <- update(tmp, . ~ . + as.factor(Rater)*Semester - Semester)
anova(tmp,tmp.single_intercept,tmp.fixed_interactions)
```

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

```
## Data: tall[tall$Rubric == "VisOrg", ]
## Models:
## tmp.single_intercept: as.numeric(Rating) ~ (1 | Artifact)
```

```
## tmp: as.numeric(Rating) ~ as.factor(Rater) + (1 | Artifact) - 1
## tmp.fixed_interactions: as.numeric(Rating) ~ as.factor(Rater) + (1 | Artifact) + as.factor(Rater):Sem
##                        npar    AIC    BIC  logLik deviance   Chisq Df
## tmp.single_intercept      3 227.21 235.44 -110.60   221.21
## tmp                       5 220.82 234.54 -105.41   210.82 10.3920  2
## tmp.fixed_interactions    8 223.12 245.08 -103.56   207.12  3.6986  3
##                        Pr(>Chisq)
## tmp.single_intercept
## tmp                      0.005539 **
## tmp.fixed_interactions   0.295899
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
```

```
summary(tmp)
```

```
## Linear mixed model fit by REML ['lmerMod']
## Formula: as.numeric(Rating) ~ as.factor(Rater) + (1 | Artifact) - 1
##    Data: tall[tall$Rubric == "VisOrg", ]
##
## 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
## as.factor(Rater)1  2.37794    0.09658   24.62
## as.factor(Rater)2  2.64891    0.09564   27.70
## as.factor(Rater)3  2.28355    0.09658   23.64
##
## Correlation of Fixed Effects:
##             a.(R)1 a.(R)2
## as.fctr(R)2 0.263
## as.fctr(R)3 0.265  0.263
```

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

Question 4

```
Fall = tall[tall$Semester == 'F19',]
Spring = tall[tall$Semester == 'S19',]
```

```
summary(Fall$Rating)
```

```
##    Min. 1st Qu.  Median    Mean 3rd Qu.    Max.
##   1.000   2.000   2.000   2.347   3.000   4.000
```

```
summary(Spring$Rating)
```

```
##    Min. 1st Qu.  Median    Mean 3rd Qu.    Max.
##   1.000   2.000   2.000   2.228   3.000   4.000
```

```
ggplot(tall,aes(x = Rating)) +
  facet_wrap( ~ Semester) +
  geom_bar()
```