How students' performance on different rubrics affects students' work in Freshman Statistics and whether the new program is successful

Author: *Hongsheng Xie*
Department of Statistics and Data Science, Carnegie Mellon University
hongshex@andrew.cmu.edu

Abstract

In this paper, we will use two datasets containing ratings of students' works in Freshman Statistics to explore the correlation between each pair of variables, factors that affect students' performance significantly, and how are the various factors in this experiment (Rater, Semester, Sex, Repeated, Rubric) related to the ratings. Data from the process of a new "General Education" program in Dietrich College at Carnegie Mellon University is used and it contains freshmen's personal information and information about the rating of their works. By doing exploratory data analysis, multiple linear regression, linear mixed-effects model, cross-classifying tables, and interaction variables analysis, several factors that affect students' performance significantly are summed up. We find that CirtDes(Given an empirical research question, the student critiques or evaluates to what extent a study design convincingly answer that question) is a special rubrics in the performance testing, and factors interactions have some positive affects on the mixed-effects model, and the reason why factor interactions work is explained. The results for all four questions are stated in the conclusion part, and the limitations of the models and data, for example one of the seven rubrics has quite distinct distribution of rating, are discussed as well.

Introduction

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 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. With the dataset containing variables including rater, ratings, rubric, sex, etc., a statistical analysis can be developed to generate a professional conclusion that for each rubric, do the raters generally agree on their scores and besides rubric, what factors play a great role for students' performance. Using dataset containing data of artifact (# of project papers), sex, repeated (whether rated by all three raters or not), etc. from the experiment of Freshmen Statistics, this paper attempts to provide a mixed-effects model analysis to study how rubrics works with raters and the relationship between variables except rubrics, and build a variable set which could develop the most appropriate prediction model.

We will answer these following questions:

1. Is the distribution of ratings for each rubric 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?
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?
3. More generally, 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?

4. Is there anything else interesting to say about this data?

Data

The data used to study how different rubrics work for different raters and the correlation between other variables taken from the experiment on Freshmen Statistics on Dietrich College. 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 Table1. The rating scale for all rubrics is shown in Table 2.

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. *NOTE: These are* **not** *the rubrics used by instructors or TA's in Freshman Statistics. They are* **only** *approved to be used in this experiment.*

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. *NOTE: This is* **not** *the rating scale used by instructors or TA's in Freshman Statistics. It is* **only** *approved to be used in this experiment.*

We examine the numeric variables and develop a description table for min, max, mean, etc. of the numeric variables. After doing exploratory data analysis, the distribution bar plots provide information on the distributions for the quantitative variables. We find out that most of the distributions of ratings for each rubric are similar: a small portion of rating 1, most of the works are rated 2 or 3. However, in rubric CritDes, most of the works are rated 1.

Methods

For question one, we firstly check the missing values of two datasets, and do some value transformations on them. Then we do summary tables on different rubrics to find out the distribution of different rubrics in numeric form. Finally, we make bar plots on each rubric to show the distributions visually and compare the difference.

For question two, we use cross-classifying model to check if the raters generally agree on their scores or not. We compare the raters in pairs (rater 1 with rater 2, rater 1 with rater 3, and rater 2 with rater 3), and create tables of counts cross classifying the rating that each pair of raters gives.

Results

Question 1. Is the distribution of ratings for each rubric pretty much indistinguishable from the other rubrics?

From the bar plots for seven rubrics, we find that the distribution of CritDes is distinguish from the distributions of other rubrics. The rating distribution of CritDes has many works with rating 1, but all other rubrics has most of works with rating 2 or 3. Besides, SelfMeth is distinguish too. This rubric has most of its work with rating 2, and much less rating 3 then other rubrics.

Discussion

Reference

Appendix

Project 2

Hongsheng Xie

11/18/2021

```
library(lme4)
```

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

```
library(arm)
```

```
## Loading required package: MASS
```

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

```
## Working directory is C:/Users/danie/Desktop/CMU/36-617 applied regression analysis/Project 2
```

```
library(ggplot2)
```

```
ratings <- read.csv("C:/Users/danie/Desktop/CMU/36-617 applied regression analysis/Project 2/rat
ings.csv")
tall <- read.csv("C:/Users/danie/Desktop/CMU/36-617 applied regression analysis/Project 2/tall.c
sv")
```

```
sum(is.na(tall))
```

```
## [1] 2
```

```
sum(is.na(tall$Rating))
```

```
## [1] 2
```

```
sum(is.na(ratings))
```

```
## [1] 80
```

```
tall$Rating <- factor(tall$Rating,levels=1:4)
for (i in unique(tall$Rubric)) {
ratings[,i] <- factor(ratings[,i],levels=1:4)
}
sum(is.na(tall))
```

```
## [1] 2
```

```
sum(is.na(ratings))
```

```
## [1] 80
```

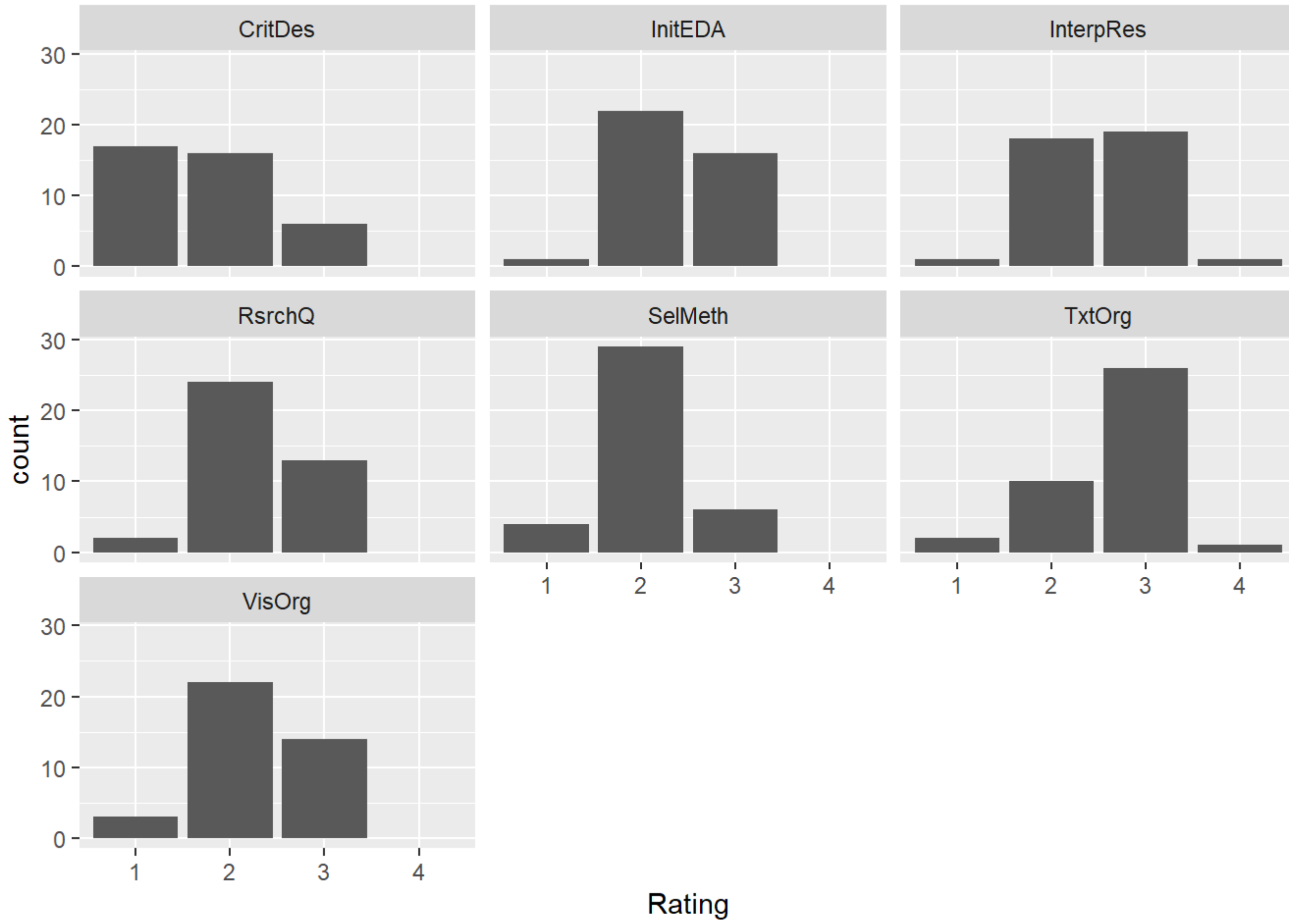
```
tall$Sex[nchar(tall$Sex)==0] <- "--"

ratings.13 <- ratings[grep("O",ratings$Artifact),]
tall.13 <- tall[grep("O",tall$Artifact),]
```

```
sum(is.na(tall))
```

```
## [1] 2
```

Q1

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

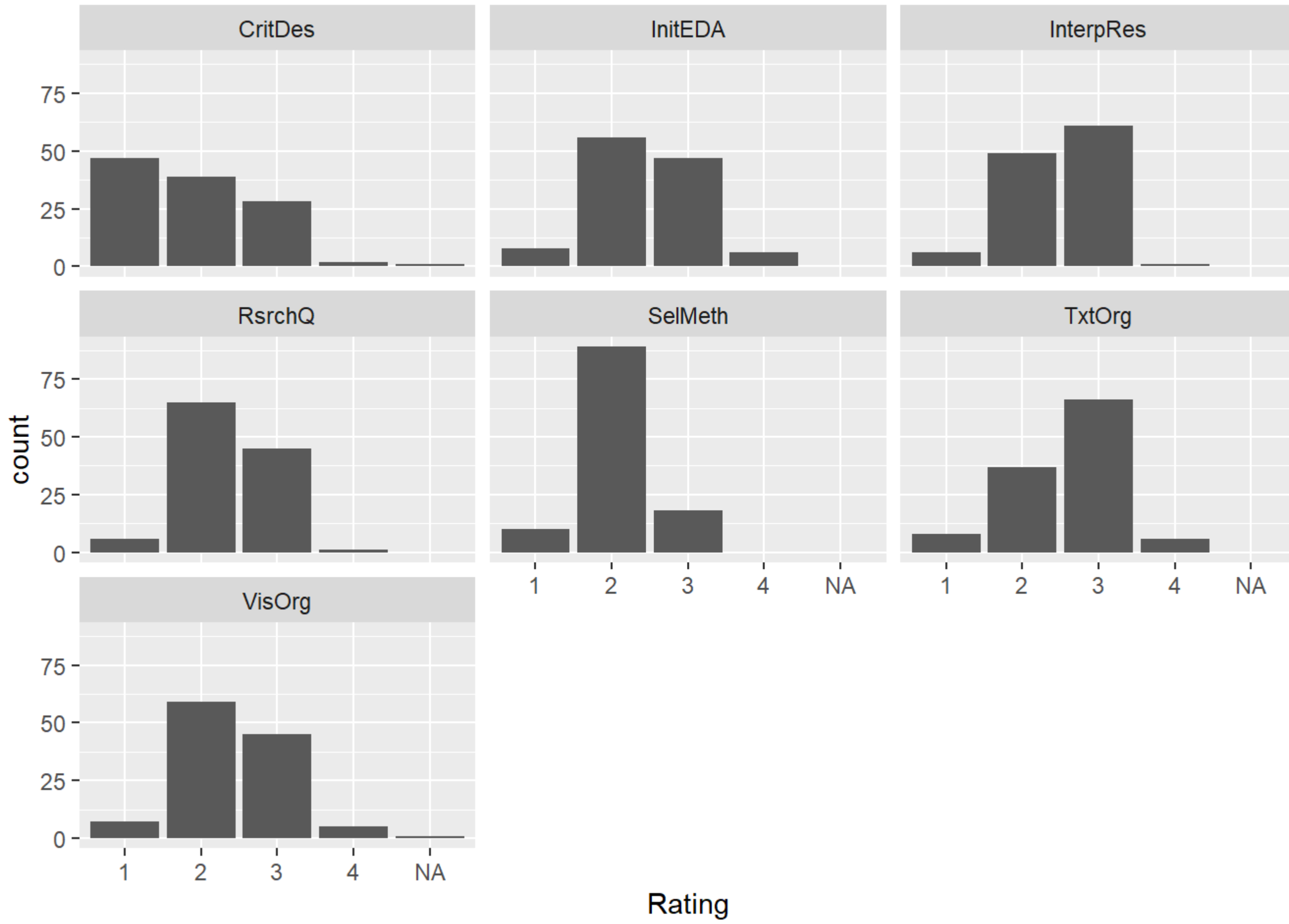



```
tmp <- data.frame(lapply(split(tall.13$Rating,tall.13$Rubric),summary))
row.names(tmp) <- paste("Rating",1:4)
tmp
```

```
##          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
```

#barplot for full dataset

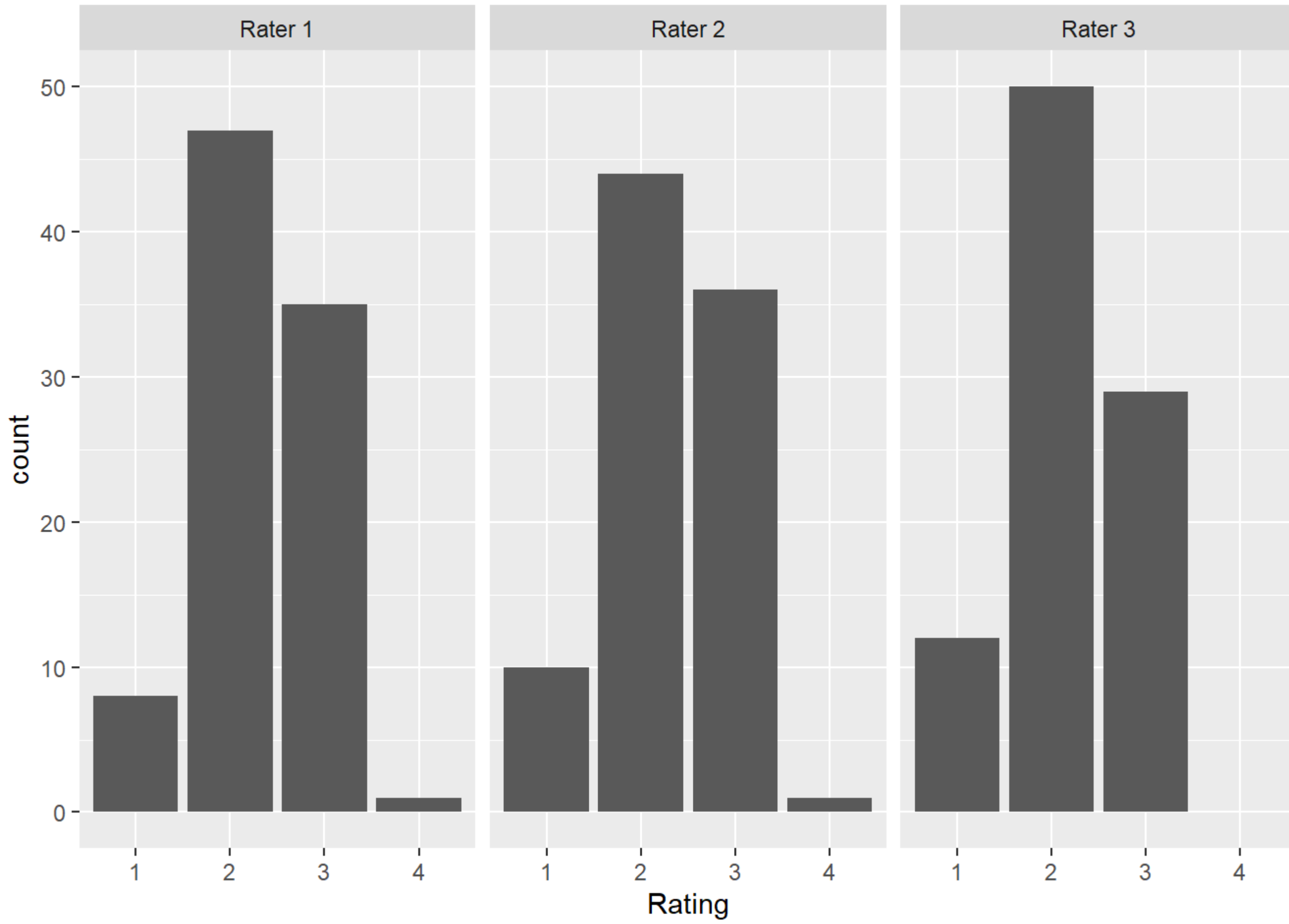
```
ggplot(tall,aes(x = Rating)) +
facet_wrap( ~ Rubric) +
geom_bar()
```




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

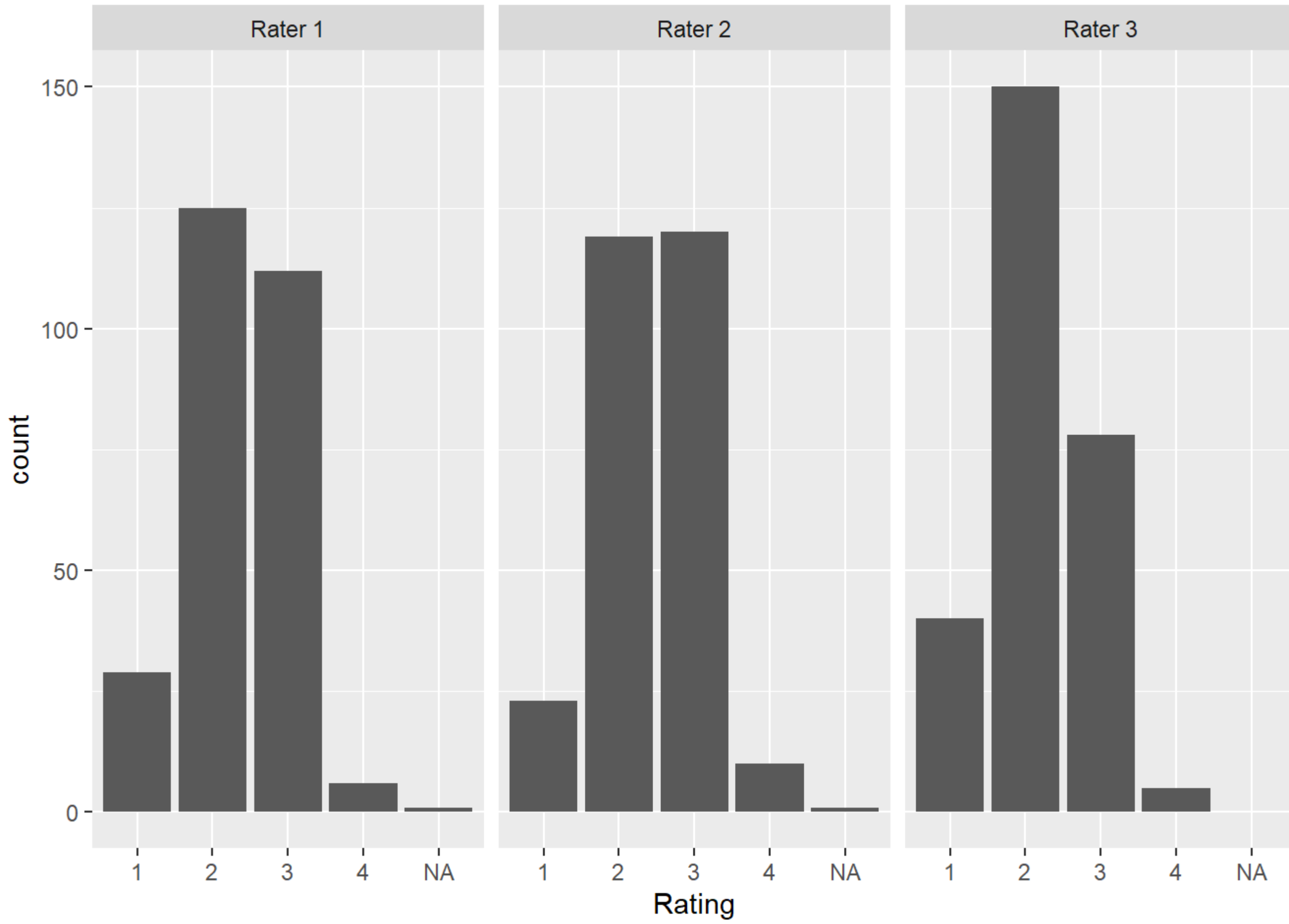
```
##
## Needed to make the title of each facet more human-readable...
rater.name <- function(x) { paste("Rater",x) }

##
## Barplots for reduced data...
g <- ggplot(tall.13,aes(x = Rating)) +
facet_wrap( ~ Rater, labeller=labeller(Rater=rater.name)) +
geom_bar()
g
```




```
##
## Corresponding table of counts...
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
```

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




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

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

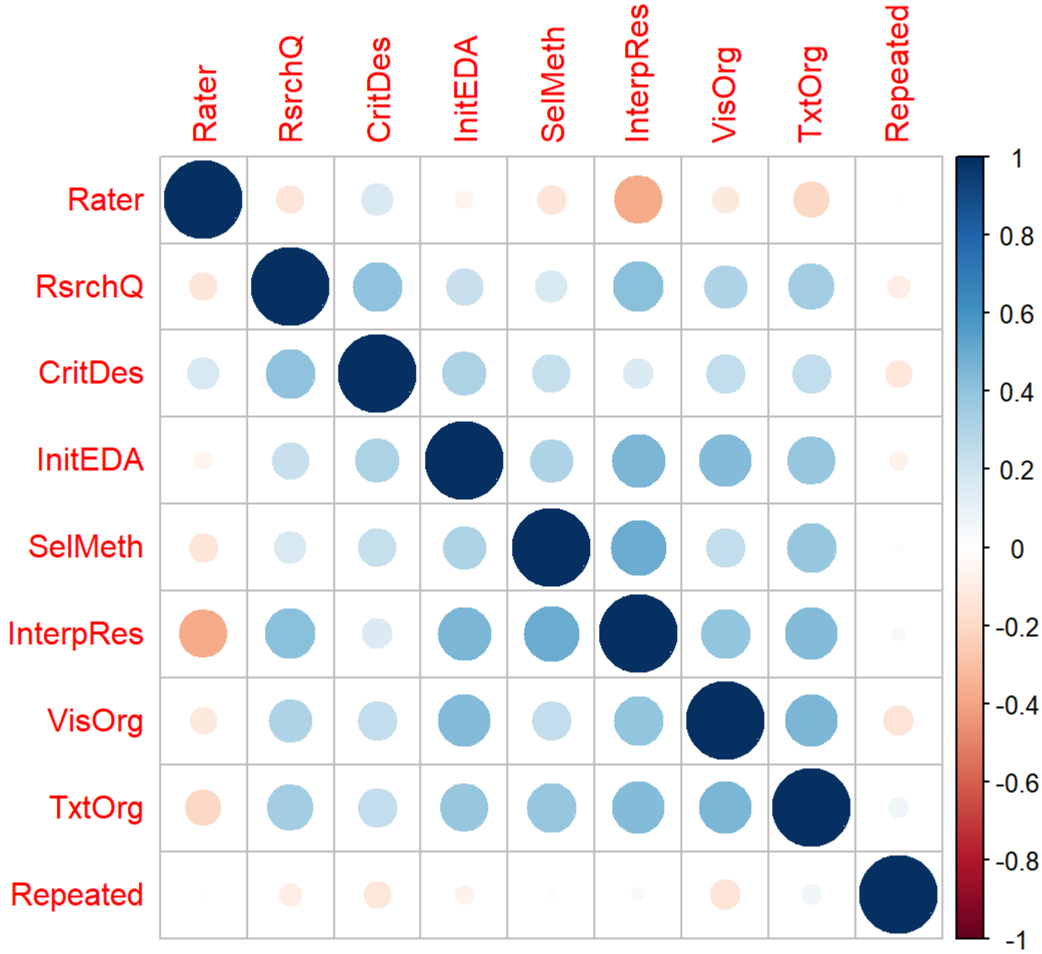
```
tall[apply(tall,1,function(x){any(is.na(x))}),]
```

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

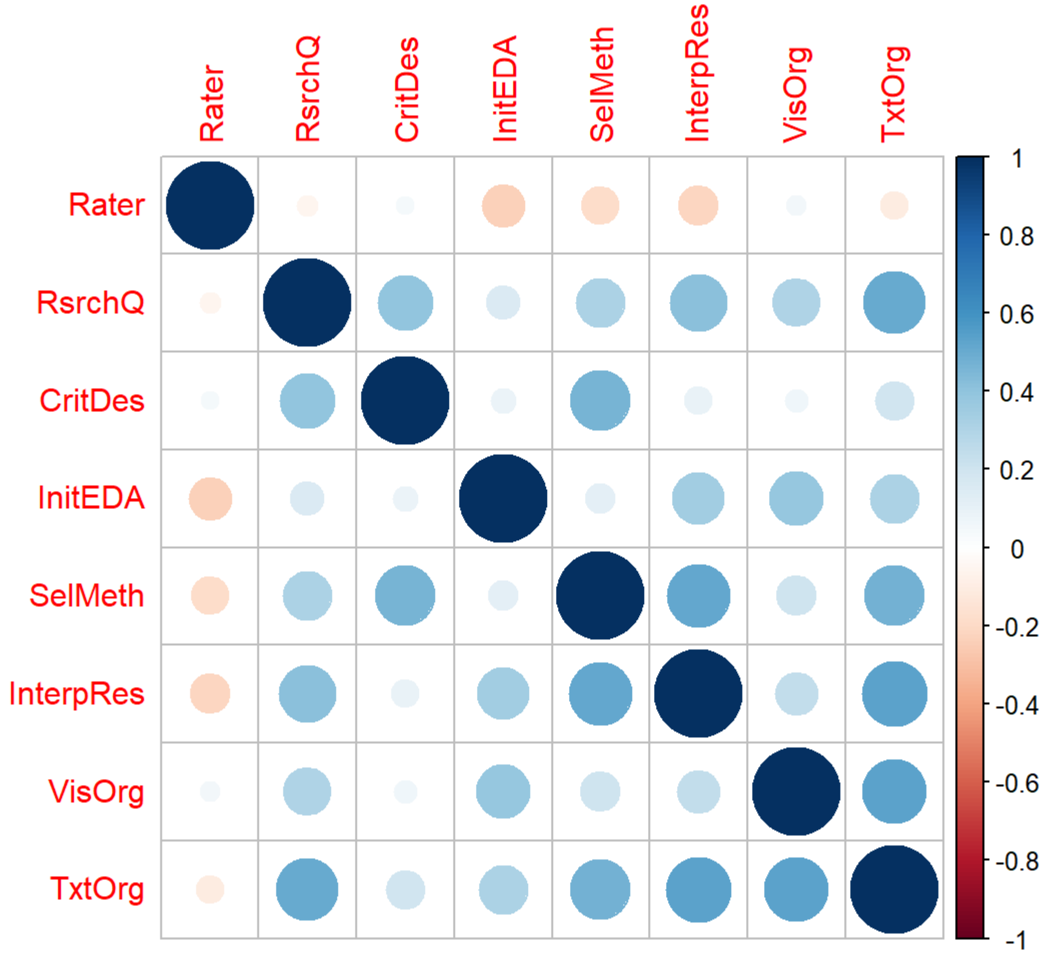
```
ratings[ratings$Sex=="--",]
```

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

```
ratings <- read.csv("C:/Users/danie/Desktop/CMU/36-617 applied regression analysis/Project 2/rat
ings.csv")

rate <- ratings[-c(44,99),-c(1,3,4)]
corrplot::corrplot(cor(rate[,-c(2,3,11)]))
```




```
data = rate[rate$Repeated == 1,]
corrplot::corrplot(cor(data[,-c(2,3,11,12)]))
```




```
# From the correlation tables, the correlation between each two factors is not quite strong. May
be considering interactions will help us.
```

```
library(psych)
```

```
##
## Attaching package: 'psych'
```

```
## The following objects are masked from 'package:ggplot2':
##
##     %+%, alpha
```

```
## The following objects are masked from 'package:arm':
##
##     logit, rescale, sim
```

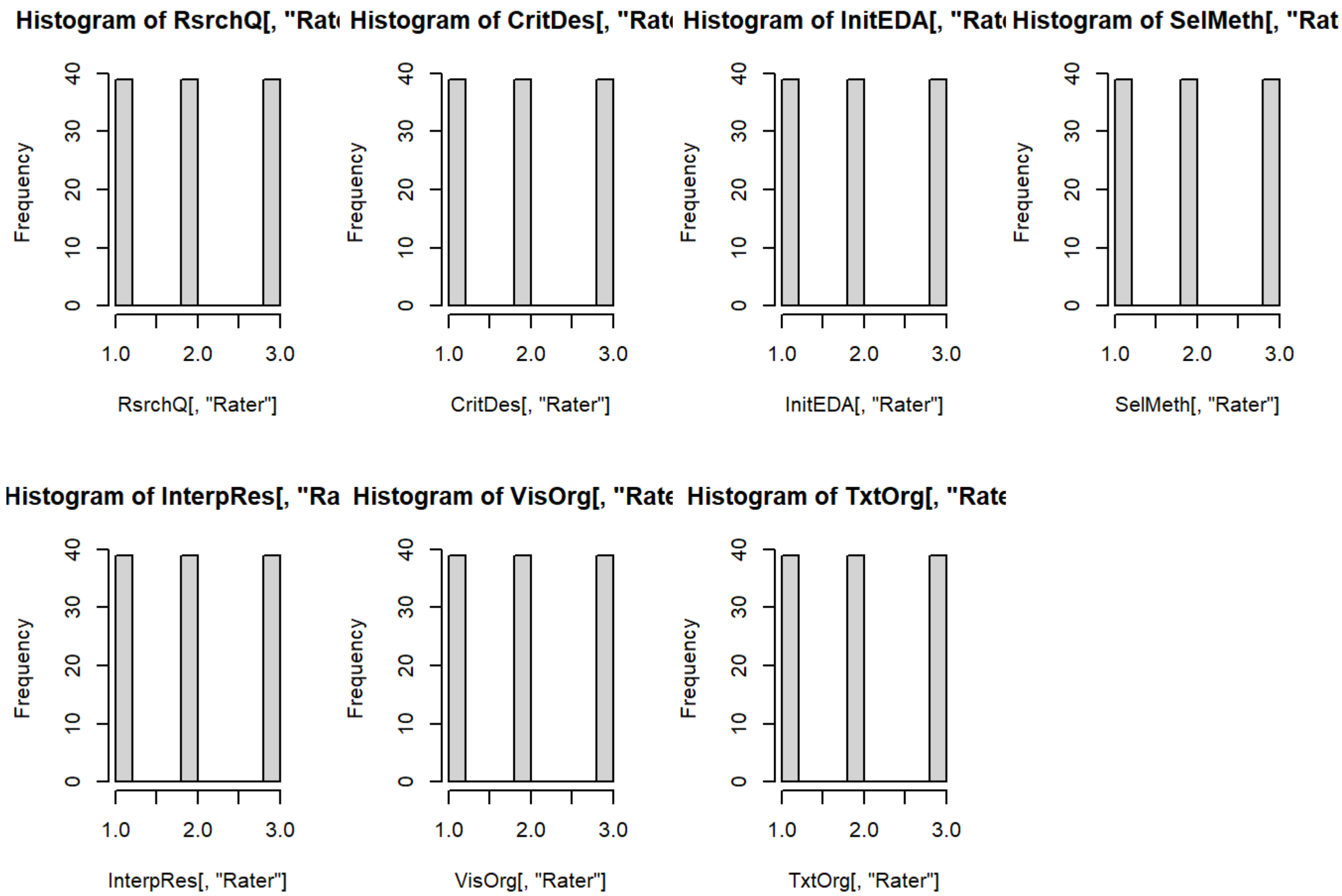
```
tall <- read.csv("C:/Users/danie/Desktop/CMU/36-617 applied regression analysis/Project 2/tall.c
sv")
describe(ratings)
```

```
##           vars   n  mean    sd median trimmed   mad min max range  skew
## X           1 117 59.00 33.92     59   59.00 43.00   1 117   116  0.00
## Rater       2 117  2.00  0.82      2    2.00  1.48   1   3     2  0.00
## Sample      3 117 59.89 34.09     60   59.98 43.00   1 118   117 -0.01
## Overlap     4  39  7.00  3.79      7    7.00  4.45   1  13    12  0.00
## Semester*   5 117  1.29  0.46      1    1.24  0.00   1   2     1  0.91
## Sex*        6 117  2.44  0.52      2    2.43  0.00   1   3     2  0.07
## RsrchQ      7 117  2.35  0.59      2    2.37  0.00   1   4     3 -0.03
## CritDes     8 116  1.87  0.84      2    1.82  1.48   1   4     3  0.42
## InitEDA     9 117  2.44  0.70      2    2.44  1.48   1   4     3  0.08
## SelMeth    10 117  2.07  0.49      2    2.07  0.00   1   3     2  0.17
## InterpRes  11 117  2.49  0.61      3    2.54  0.00   1   4     3 -0.52
## VisOrg     12 116  2.41  0.67      2    2.41  0.00   1   4     3  0.14
## TxtOrg     13 117  2.60  0.70      3    2.64  0.00   1   4     3 -0.50
## Artifact*  14 117 54.67 28.45     59   56.25 35.58   1  91    90 -0.35
## Repeated   15 117  0.33  0.47      0    0.29  0.00   0   1     1  0.70
##           kurtosis   se
## X            -1.23 3.14
## Rater        -1.53 0.08
## Sample       -1.22 3.15
## Overlap      -1.30 0.61
## Semester*    -1.18 0.04
## Sex*         -1.56 0.05
## RsrchQ       -0.47 0.05
## CritDes      -1.03 0.08
## InitEDA      -0.26 0.06
## SelMeth       1.06 0.04
## InterpRes    -0.44 0.06
## VisOrg       -0.20 0.06
## TxtOrg       -0.01 0.06
## Artifact*    -1.29 2.63
## Repeated     -1.53 0.04
```

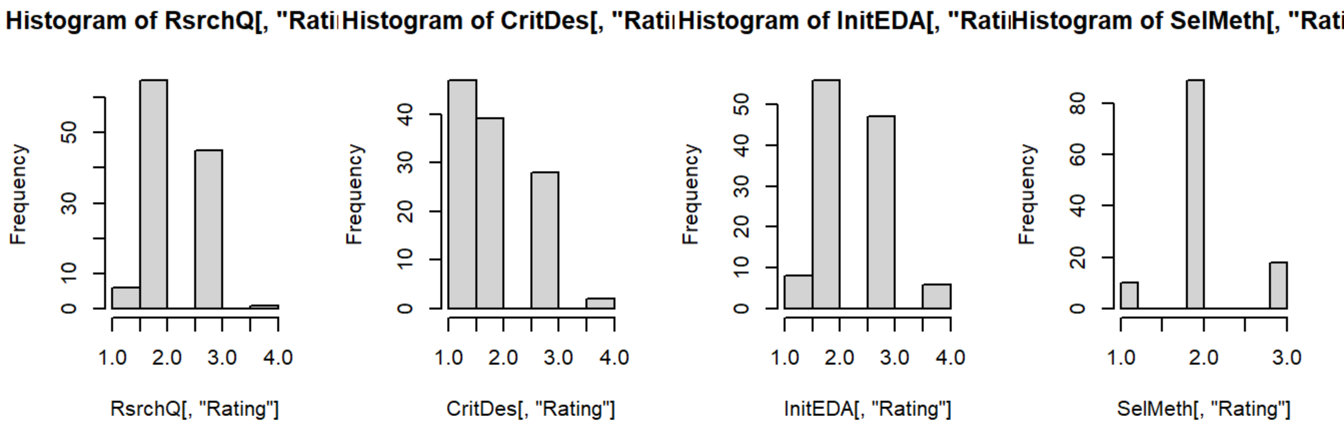
```
summary(ratings)
```

```
##        X           Rater       Sample         Overlap     Semester
##  Min.   :  1   Min.   :1   Min.   :  1.00   Min.   : 1   Length:117
##  1st Qu.: 30   1st Qu.:1   1st Qu.: 31.00   1st Qu.: 4   Class :character
##  Median : 59   Median :2   Median : 60.00   Median : 7   Mode  :character
##  Mean   : 59   Mean   :2   Mean   : 59.89   Mean   : 7
##  3rd Qu.: 88   3rd Qu.:3   3rd Qu.: 89.00   3rd Qu.:10
##  Max.   :117   Max.   :3   Max.   :118.00   Max.   :13
##                                             NA's   :78
##      Sex                RsrchQ       CritDes         InitEDA
##  Length:117         Min.   :1.00   Min.   :1.000   Min.   :1.000
##  Class :character   1st Qu.:2.00   1st Qu.:1.000   1st Qu.:2.000
##  Mode  :character   Median :2.00   Median :2.000   Median :2.000
##                     Mean   :2.35   Mean   :1.871   Mean   :2.436
##                     3rd Qu.:3.00   3rd Qu.:3.000   3rd Qu.:3.000
##                     Max.   :4.00   Max.   :4.000   Max.   :4.000
##                                    NA's   :1
##     SelMeth        InterpRes         VisOrg          TxtOrg
##  Min.   :1.000   Min.   :1.000   Min.   :1.000   Min.   :1.000
##  1st Qu.:2.000   1st Qu.:2.000   1st Qu.:2.000   1st Qu.:2.000
##  Median :2.000   Median :3.000   Median :2.000   Median :3.000
##  Mean   :2.068   Mean   :2.487   Mean   :2.414   Mean   :2.598
##  3rd Qu.:2.000   3rd Qu.:3.000   3rd Qu.:3.000   3rd Qu.:3.000
##  Max.   :3.000   Max.   :4.000   Max.   :4.000   Max.   :4.000
##                                  NA's   :1
##    Artifact            Repeated
##  Length:117         Min.   :0.0000
##  Class :character   1st Qu.:0.0000
##  Mode  :character   Median :0.0000
##                     Mean   :0.3333
##                     3rd Qu.:1.0000
##                     Max.   :1.0000
##
```

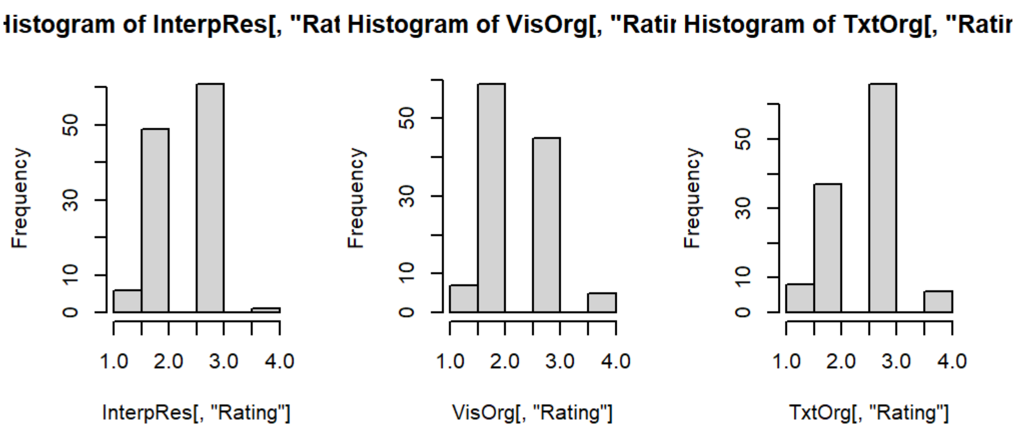
```
RsrchQ = tall[tall[,"Rubric"] == "RsrchQ",]
CritDes = tall[tall[,"Rubric"] == "CritDes",]
InitEDA = tall[tall[,"Rubric"] == "InitEDA",]
SelMeth = tall[tall[,"Rubric"] == "SelMeth",]
InterpRes = tall[tall[,"Rubric"] == "InterpRes",]
VisOrg = tall[tall[,"Rubric"] == "VisOrg",]
TxtOrg = tall[tall[,"Rubric"] == "TxtOrg",]

par(mfrow = c(2,4))
hist(RsrchQ[,"Rater"])
hist(CritDes[,"Rater"])
hist(InitEDA[,"Rater"])
hist(SelMeth[,"Rater"])
hist(InterpRes[,"Rater"])
hist(VisOrg[,"Rater"])
hist(TxtOrg[,"Rater"])
# The distributions of Rater for all Rubrics are the same.
```




```
par(mfrow = c(2,4))
hist(RsrchQ[,"Rating"])
hist(CritDes[,"Rating"])
hist(InitEDA[,"Rating"])
hist(SelMeth[,"Rating"])
hist(InterpRes[,"Rating"])
hist(VisOrg[,"Rating"])
hist(TxtOrg[,"Rating"])

# All the distributions have a similar trend except CritDes.
```


Histogram of RsrchQ[, "Rati
Histogram of CritDes[, "Rati
Histogram of InitEDA[, "Rati
Histogram of SelMeth[, "Rati
Frequency
RsrchQ[, "Rating"]
CritDes[, "Rating"]
InitEDA[, "Rating"]
SelMeth[, "Rating"]

Q2


listogram of InterpRes[, "Rat
Histogram of VisOrg[, "Ratir
Histogram of TxtOrg[, "Ratir
Frequency
InterpRes[, "Rating"]
VisOrg[, "Rating"]
TxtOrg[, "Rating"]

```
##
## useful preliminaries
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
##
## Now add in ICC's calculated from all the data...
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)

agreement.results
```

```
##           ICC.alldata ICC.common  a12 a23 a13
## CritDes     0.6730647  0.5725594    0   0   0
## InitEDA     0.6867210  0.4929577    0   0   0
## InterpRes   0.2200285  0.2295720    0   0   0
## RsrchQ      0.2096214  0.1891892    0   0   0
## SelMeth     0.4719014  0.5212766    0   0   0
## TxtOrg      0.1879927  0.1428571    0   0   0
## VisOrg      0.6607372  0.5924529    0   0   0
```

```
round(agreement.results,2)
```

```
##           ICC.alldata ICC.common  a12 a23 a13
## CritDes          0.67       0.57    0   0   0
## InitEDA          0.69       0.49    0   0   0
## InterpRes        0.22       0.23    0   0   0
## RsrchQ           0.21       0.19    0   0   0
## SelMeth          0.47       0.52    0   0   0
## TxtOrg           0.19       0.14    0   0   0
## VisOrg           0.66       0.59    0   0   0
```

```
library(arm)
library(lme4)
tall <- read.csv("C:/Users/danie/Desktop/CMU/36-617 applied regression analysis/Project 2/tall.c
sv")
common <- tall[grep("0",tall$Artifact),]
```

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

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

```
RsrchQ_Artifact <- lmer(Rating ~ 1 + (1|Artifact), data=RsrchQ.ratings)
```

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

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

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

```
CritDes_Artifact <- lmer(Rating ~ 1 + (1|Artifact), data=CritDes.ratings)
```

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

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

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

```
InitEDA_Artifact <- lmer(Rating ~ 1 + (1|Artifact), data=InitEDA.ratings)
```

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

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

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

```
SelMeth_Artifact <- lmer(Rating ~ 1 + (1|Artifact), data=SelMeth.ratings)
```

```
## Warning in checkConv(attr(opt, "derivs"), opt$par, ctrl = control$checkConv, : Model is nearl
y unidentifiable: very large eigenvalue
##  - Rescale variables?
```

```
InterpRes.ratings <- common[common$Rubric=="InterpRes",]
InterpRes_Rater <- lmer(Rating ~ 1 + (1|Rater), data=InterpRes.ratings)
```

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

```
InterpRes_Artifact <- lmer(Rating ~ 1 + (1|Artifact), data=InterpRes.ratings)
```

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

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

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

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

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

```
TxtOrg_Artifact <- lmer(Rating ~ 1 + (1|Artifact), data=TxtOrg.ratings)
```

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

```
#summary(RsrchQ_Rater)
#summary(CritDes_Rater)
#summary(InitEDA_Rater)
#summary(SelMeth_Rater)
#summary(InterpRes_Rater)
#summary(VisOrg_Rater)
#summary(TxtOrg_Rater)

#summary(RsrchQ_Artifact)
#summary(CritDes_Artifact)
#summary(InitEDA_Artifact)
#summary(SelMeth_Artifact)
#summary(InterpRes_Artifact)
#summary(VisOrg_Artifact)
#summary(TxtOrg_Artifact)
```

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

RsrchQ rubric

```
raters_1_and_2_on_RsrchQ <- data.frame(r1=repeated$RsrchQ[repeated$Rater==1],
                                       r2=repeated$RsrchQ[repeated$Rater==2],
                                       a1=repeated$Artifact[repeated$Rater==1],
                                       a2=repeated$Artifact[repeated$Rater==2]
)

r1 <- factor(raters_1_and_2_on_RsrchQ$r1,levels=1:4)
r2 <- factor(raters_1_and_2_on_RsrchQ$r2,levels=1:4)
(t12 <- table(r1,r2))
```

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

```
#For the artifacts which are rated by all three raters, we find that the rater 1 and rater 2 for
the rubric RsrchQ have the same rate in 5/13 of the cases. From the rubric RsrchQ, rater 1 and 2
do not always disagree with each other.

agreement.results[4,3] = round(5/13,2)
```

```
raters_1_and_3_on_RsrchQ <- data.frame(r1=repeated$RsrchQ[repeated$Rater==1],
                                       r3=repeated$RsrchQ[repeated$Rater==3],
                                       a1=repeated$Artifact[repeated$Rater==1],
                                       a3=repeated$Artifact[repeated$Rater==3]
)

r1 <- factor(raters_1_and_3_on_RsrchQ$r1,levels=1:4)
r3 <- factor(raters_1_and_3_on_RsrchQ$r3,levels=1:4)
(t13 <- table(r1,r3))
```

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

```
#For the artifacts which are rated by all three raters, we find that the rater 1 and rater 3 for
the rubric RsrchQ have the same rate in 10/13 of the cases. From the rubric RsrchQ, rater 1 and
 3 do not always disagree with each other.

agreement.results[4,5] = round(10/13,2)
```

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

r2 <- factor(raters_2_and_3_on_RsrchQ$r2,levels=1:4)
r3 <- factor(raters_2_and_3_on_RsrchQ$r3,levels=1:4)
(t23 <- table(r2,r3))
```

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

```
#For the artifacts which are rated by all three raters, we find that the rater 1 and rater 2 for
the rubric RsrchQ have the same rate in 7/13 of the cases. From the rubric RsrchQ, rater 2 and 3
do not always disagree with each other.
agreement.results[4,4] = round(7/13,2)
```

CritDes rubric

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

r1 <- factor(raters_1_and_2_on_CritDes$r1,levels=1:4)
r2 <- factor(raters_1_and_2_on_CritDes$r2,levels=1:4)
(t12 <- table(r1,r2))
```

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

```
#For the artifacts which are rated by all three raters, we find that the rater 1 and rater 2 for the rubric CritDes have the same rate in 7/13 of the cases. From the rubric CritDes, rater 1 and 2 do not always disagree with each other.
agreement.results[1,3] = round(7/13,2)
```

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

r1 <- factor(raters_1_and_3_on_CritDes$r1,levels=1:4)
r3 <- factor(raters_1_and_3_on_CritDes$r3,levels=1:4)
(t13 <- table(r1,r3))
```

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

```
#For the artifacts which are rated by all three raters, we find that the rater 1 and rater 3 for the rubric CritDes have the same rate in 8/13 of the cases. From the rubric CritDes, rater 1 and 3 do not always disagree with each other.
agreement.results[1,5] = round(8/13,2)
```

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

r2 <- factor(raters_2_and_3_on_CritDes$r2,levels=1:4)
r3 <- factor(raters_2_and_3_on_CritDes$r3,levels=1:4)
(t23 <- table(r2,r3))
```

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

```
#For the artifacts which are rated by all three raters, we find that the rater 3 and rater 2 for the rubric CritDes have the same rate in 9/13 of the cases. From the rubric CritDes, rater 3 and 2 do not always disagree with each other.
agreement.results[1,4] = round(9/13,2)
```

InitEDA rubric

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

r1 <- factor(raters_1_and_2_on_InitEDA$r1,levels=1:4)
r2 <- factor(raters_1_and_2_on_InitEDA$r2,levels=1:4)
(t12 <- table(r1,r2))
```

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

```
#For the artifacts which are rated by all three raters, we find that the rater 1 and rater 2 for the rubric InitEDA have the same rate in 9/13 of the cases. From the rubric InitEDA, rater 1 and 2 do not always disagree with each other.
agreement.results[2,3] = round(9/13,2)
```

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

r1 <- factor(raters_1_and_3_on_InitEDA$r1,levels=1:4)
r3 <- factor(raters_1_and_3_on_InitEDA$r3,levels=1:4)
(t13 <- table(r1,r3))
```

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

```
#For the artifacts which are rated by all three raters, we find that the rater 1 and rater 3 for
the rubric InitEDA have the same rate in 7/13 of the cases. From the rubric InitEDA, rater 1 and
3 do not always disagree with each other.
agreement.results[2,5] = round(7/13,2)
```

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

r2 <- factor(raters_2_and_3_on_InitEDA$r2,levels=1:4)
r3 <- factor(raters_2_and_3_on_InitEDA$r3,levels=1:4)
(t23 <- table(r2,r3))
```

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

```
#For the artifacts which are rated by all three raters, we find that the rater 2 and rater 3 for
the rubric InitEDA have the same rate in 11/13 of the cases. From the rubric InitEDA, rater 2 an
d 3 do not always disagree with each other.
agreement.results[2,4] = round(11/13,2)
```

SelMeth rubric

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

r1 <- factor(raters_1_and_2_on_SelMeth$r1,levels=1:4)
r2 <- factor(raters_1_and_2_on_SelMeth$r2,levels=1:4)
(t12 <- table(r1,r2))
```

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

```
#For the artifacts which are rated by all three raters, we find that the rater 1 and rater 2 for the rubric SelMeth have the same rate in 12/13 of the cases. From the rubric SelMeth, rater 1 and 2 do not always disagree with each other.
agreement.results[5,3] = round(12/13,2)
```

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

r1 <- factor(raters_1_and_3_on_SelMeth$r1,levels=1:4)
r3 <- factor(raters_1_and_3_on_SelMeth$r3,levels=1:4)
(t13 <- table(r1,r3))
```

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

```
#For the artifacts which are rated by all three raters, we find that the rater 1 and rater 3 for the rubric SelMeth have the same rate in 8/13 of the cases. From the rubric SelMeth, rater 1 and 3 do not always disagree with each other.
agreement.results[5,5] = round(8/13,2)
```

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

r2 <- factor(raters_2_and_3_on_SelMeth$r2,levels=1:4)
r3 <- factor(raters_2_and_3_on_SelMeth$r3,levels=1:4)
(t23 <- table(r2,r3))
```

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

```
#For the artifacts which are rated by all three raters, we find that the rater 2 and rater 3 for the rubric SelMeth have the same rate in 8/13 of the cases. From the rubric SelMeth, rater 2 and 3 do not always disagree with each other.
agreement.results[5,4] = round(8/13,2)
```

InterpRes rubric

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

r1 <- factor(raters_1_and_2_on_InterpRes$r1,levels=1:4)
r2 <- factor(raters_1_and_2_on_InterpRes$r2,levels=1:4)
(t12 <- table(r1,r2))
```

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

```
#For the artifacts which are rated by all three raters, we find that the rater 1 and rater 2 for the rubric InterpRes have the same rate in 8/13 of the cases. From the rubric InterpRes, rater 1 and 2 do not always disagree with each other.
agreement.results[3,3] = round(8/13,2)
```

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

r1 <- factor(raters_1_and_3_on_InterpRes$r1,levels=1:4)
r3 <- factor(raters_1_and_3_on_InterpRes$r3,levels=1:4)
(t13 <- table(r1,r3))
```

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

```
#For the artifacts which are rated by all three raters, we find that the rater 1 and rater 3 for
the rubric InterpRes have the same rate in 7/13 of the cases. From the rubric InterpRes, rater 1
and 3 do not always disagree with each other.
agreement.results[3,5] = round(7/13,2)
```

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

r2 <- factor(raters_2_and_3_on_InterpRes$r2,levels=1:4)
r3 <- factor(raters_2_and_3_on_InterpRes$r3,levels=1:4)
(t23 <- table(r2,r3))
```

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

```
#For the artifacts which are rated by all three raters, we find that the rater 2 and rater 3 for
the rubric InterpRes have the same rate in 8/13 of the cases. From the rubric InterpRes, rater 2
and 3 do not always disagree with each other.
agreement.results[3,4] = round(8/13,2)
```

VisOrg rubric

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

r1 <- factor(raters_1_and_2_on_VisOrg$r1,levels=1:4)
r2 <- factor(raters_1_and_2_on_VisOrg$r2,levels=1:4)
(t12 <- table(r1,r2))
```

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

```
#For the artifacts which are rated by all three raters, we find that the rater 1 and rater 2 for the rubric VisOrg have the same rate in 6/13 of the cases. From the rubric VisOrg, rater 1 and 2 do not always disagree with each other.
agreement.results[7,3] = round(6/13,2)
```

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

r1 <- factor(raters_1_and_3_on_VisOrg$r1,levels=1:4)
r3 <- factor(raters_1_and_3_on_VisOrg$r3,levels=1:4)
(t13 <- table(r1,r3))
```

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

```
#For the artifacts which are rated by all three raters, we find that the rater 1 and rater 3 for the rubric VisOrg have the same rate in 9/13 of the cases. From the rubric VisOrg, rater 1 and 3 do not always disagree with each other.
agreement.results[7,5] = round(9/13,2)
```

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

r2 <- factor(raters_2_and_3_on_VisOrg$r2,levels=1:4)
r3 <- factor(raters_2_and_3_on_VisOrg$r3,levels=1:4)
(t23 <- table(r2,r3))
```

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

```
#For the artifacts which are rated by all three raters, we find that the rater 2 and rater 3 for the rubric VisOrg have the same rate in 9/13 of the cases. From the rubric VisOrg, rater 2 and 3 do not always disagree with each other.
agreement.results[7,4] = round(9/13,2)
```

TxtOrg rubric

```
raters_1_and_2_on_TxtOrg <- data.frame(r1=repeated$TxtOrg[repeated$Rater==1],
                                       r2=repeated$TxtOrg[repeated$Rater==2],
                                       a1=repeated$Artifact[repeated$Rater==1],
                                       a2=repeated$Artifact[repeated$Rater==2]
)

r1 <- factor(raters_1_and_2_on_TxtOrg$r1,levels=1:4)
r2 <- factor(raters_1_and_2_on_TxtOrg$r2,levels=1:4)
(t12 <- table(r1,r2))
```

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

```
#For the artifacts which are rated by all three raters, we find that the rater 1 and rater 2 for the rubric TxtOrg have the same rate in 9/13 of the cases. From the rubric TxtOrg, rater 1 and 2 do not always disagree with each other.
agreement.results[6,3] = round(9/13,2)
```

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

r1 <- factor(raters_1_and_3_on_TxtOrg$r1,levels=1:4)
r3 <- factor(raters_1_and_3_on_TxtOrg$r3,levels=1:4)
(t13 <- table(r1,r3))
```

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

```
#For the artifacts which are rated by all three raters, we find that the rater 1 and rater 3 for
the rubric TxtOrg have the same rate in 8/13 of the cases. From the rubric TxtOrg, rater 1 and 3
do not always disagree with each other.
agreement.results[6,5] = round(8/13,2)
```

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

r2 <- factor(raters_2_and_3_on_TxtOrg$r2,levels=1:4)
r3 <- factor(raters_2_and_3_on_TxtOrg$r3,levels=1:4)
(t23 <- table(r2,r3))
```

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

```
#For the artifacts which are rated by all three raters, we find that the rater 2 and rater 3 for
the rubric TxtOrg have the same rate in 7/13 of the cases. From the rubric TxtOrg, rater 2 and 3
do not always disagree with each other.
agreement.results[6,4] = round(7/13,2)
```

q3

```
tall$Rater = as.factor(tall$Rater)
```

fixed effects

```
m0 <- lmer(Rating ~ (0 + Rubric | Artifact),data = tall)
```

```
## Warning in checkConv(attr(opt, "derivs"), opt$par, ctrl = control$checkConv, :
## Model failed to converge with max|grad| = 0.00236116 (tol = 0.002, component 1)
```

```
m1 <- lmer(Rating ~ 1 + Rater + (0 + Rubric | Artifact),data = tall)
```

```
## Warning in checkConv(attr(opt, "derivs"), opt$par, ctrl = control$checkConv, :
## Model failed to converge with max|grad| = 0.00429684 (tol = 0.002, component 1)
```

```
m2 <- lmer(Rating ~ 1 + Semester + (0+Rubric | Artifact),data = tall)
```

```
## Warning in checkConv(attr(opt, "derivs"), opt$par, ctrl = control$checkConv, :
## Model failed to converge with max|grad| = 0.00221712 (tol = 0.002, component 1)
```

```
m3 <- lmer(Rating ~ 1 + Sex + (0+Rubric | Artifact),data = tall)
```

```
## Warning in checkConv(attr(opt, "derivs"), opt$par, ctrl = control$checkConv, :
## Model failed to converge with max|grad| = 0.00300248 (tol = 0.002, component 1)
```

```
m4 <- lmer(Rating ~ 1 + Repeated + (0+Rubric | Artifact),data = tall)
```

```
## Warning in checkConv(attr(opt, "derivs"), opt$par, ctrl = control$checkConv, :
## Model failed to converge with max|grad| = 0.00462524 (tol = 0.002, component 1)
```

```
m5 = lmer(Rating ~ 1 + Rater + Semester + (0+Rubric | Artifact),data = tall)
```

```
## Warning in checkConv(attr(opt, "derivs"), opt$par, ctrl = control$checkConv, :
## Model failed to converge with max|grad| = 0.0245363 (tol = 0.002, component 1)
```

```
m6 = lmer(Rating ~ 1 + Rater + Sex + (0+Rubric | Artifact),data = tall)
```

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

```
m7 = lmer(Rating ~ 1 + Rater + Repeated + (0+Rubric | Artifact),data = tall)
```

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

```
m8 = lmer(Rating ~ 1 + Semester + Sex + (0+Rubric | Artifact),data = tall)
```

```
## Warning in checkConv(attr(opt, "derivs"), opt$par, ctrl = control$checkConv, :
## Model failed to converge with max|grad| = 0.00281508 (tol = 0.002, component 1)
```

```
m9 = lmer(Rating ~ 1 + Semester + Repeated +(0+Rubric | Artifact),data = tall)
```

```
## Warning in checkConv(attr(opt, "derivs"), opt$par, ctrl = control$checkConv, :
## Model failed to converge with max|grad| = 0.017182 (tol = 0.002, component 1)
```

```
m10 = lmer(Rating ~ 1 + Sex + Repeated + (0+Rubric | Artifact),data = tall)
```

```
## 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 2 negative eigenvalues
```

```
m11 = lmer(Rating ~ 1 + Rater + Semester+ Sex + (0+Rubric | Artifact),data = tall)
```

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

```
m12 = lmer(Rating ~ 1 + Rater + Semester + Repeated + (0+Rubric | Artifact),data = tall)
```

```
## Warning in checkConv(attr(opt, "derivs"), opt$par, ctrl = control$checkConv, :
## Model failed to converge with max|grad| = 0.0040295 (tol = 0.002, component 1)
```

```
m13 = lmer(Rating ~ 1 + Rater + Sex + Repeated + (0+Rubric | Artifact),data = tall)
```

```
## Warning in checkConv(attr(opt, "derivs"), opt$par, ctrl = control$checkConv, :
## Model failed to converge with max|grad| = 0.204177 (tol = 0.002, component 1)
```

```
m14 = lmer(Rating ~ 1 + Repeated + Sex + Semester+(0+Rubric | Artifact),data = tall)
m15 = lmer(Rating ~ 1 + Rater + Repeated + Sex + Semester + (0+Rubric | Artifact),data = tall)

anova(m0,m1,m2,m3,m4,m5,m6,m7,m8,m9,m10,m11,m12,m13,m14,m15)
```

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

```
## Data: tall
## Models:
## m0: Rating ~ (0 + Rubric | Artifact)
## m2: Rating ~ 1 + Semester + (0 + Rubric | Artifact)
## m4: Rating ~ 1 + Repeated + (0 + Rubric | Artifact)
## m1: Rating ~ 1 + Rater + (0 + Rubric | Artifact)
## m3: Rating ~ 1 + Sex + (0 + Rubric | Artifact)
## m9: Rating ~ 1 + Semester + Repeated + (0 + Rubric | Artifact)
## m5: Rating ~ 1 + Rater + Semester + (0 + Rubric | Artifact)
## m7: Rating ~ 1 + Rater + Repeated + (0 + Rubric | Artifact)
## m8: Rating ~ 1 + Semester + Sex + (0 + Rubric | Artifact)
## m10: Rating ~ 1 + Sex + Repeated + (0 + Rubric | Artifact)
## m6: Rating ~ 1 + Rater + Sex + (0 + Rubric | Artifact)
## m12: Rating ~ 1 + Rater + Semester + Repeated + (0 + Rubric | Artifact)
## m14: Rating ~ 1 + Repeated + Sex + Semester + (0 + Rubric | Artifact)
## m11: Rating ~ 1 + Rater + Semester + Sex + (0 + Rubric | Artifact)
## m13: Rating ~ 1 + Rater + Sex + Repeated + (0 + Rubric | Artifact)
## m15: Rating ~ 1 + Rater + Repeated + Sex + Semester + (0 + Rubric | Artifact)
##     npar    AIC    BIC  logLik deviance   Chisq Df Pr(>Chisq)    
## m0    30 1537.2 1678.3 -738.58   1477.2                          
## m2    31 1535.1 1681.0 -736.57   1473.1  4.0182  1  0.0450117 *  
## m4    31 1538.1 1684.0 -738.05   1476.1  0.0000  0               
## m1    32 1529.9 1680.5 -732.94   1465.9 10.2233  1  0.0013868 ** 
## m3    32 1536.9 1687.5 -736.43   1472.9  0.0000  0               
## m9    32 1535.5 1686.1 -735.77   1471.5  1.3198  0               
## m5    33 1527.5 1682.8 -730.75   1461.5 10.0451  1  0.0015276 ** 
## m7    33 1530.9 1686.2 -732.46   1464.9  0.0000  0               
## m8    33 1535.6 1690.9 -734.82   1469.6  0.0000  0               
## m10   33 1537.9 1693.2 -735.95   1471.9  0.0000  0               
## m6    34 1528.2 1688.1 -730.07   1460.2 11.7439  1  0.0006104 ***
## m12   34 1528.0 1688.0 -729.99   1460.0  0.1615  0               
## m14   34 1536.2 1696.2 -734.09   1468.2  0.0000  0               
## m11   35 1526.7 1691.3 -728.32   1456.7 11.5357  1  0.0006827 ***
## m13   35 1529.3 1694.0 -729.64   1459.3  0.0000  0               
## m15   36 1527.2 1696.6 -727.61   1455.2  4.0457  1  0.0442828 *  
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
```

In this ANOVA, we did not consider the model with interaction of two or more variables. According to the result above, the best model is m11(AIC = 1526.7), we find that the best fix effect variables are Rater, Semester and Sex.

```
summary(m11)
```

```
## Linear mixed model fit by REML ['lmerMod']
## Formula: Rating ~ 1 + Rater + Semester + Sex + (0 + Rubric | Artifact)
##    Data: tall
## 
## REML criterion at convergence: 1476.1
## 
## Scaled residuals: 
##     Min      1Q  Median      3Q     Max 
## -3.1991 -0.4964 -0.0597  0.5344  3.6774 
## 
## Random effects:
##  Groups   Name            Variance Std.Dev. Corr                         
##  Artifact RubricCritDes   0.65059  0.8066                                
##           RubricInitEDA   0.38494  0.6204    0.27                        
##           RubricInterpRes 0.22696  0.4764   -0.02  0.78                  
##           RubricRsrchQ    0.18048  0.4248    0.40  0.51  0.73            
##           RubricSelMeth   0.08057  0.2839    0.58  0.33  0.31  0.21      
##           RubricTxtOrg    0.38105  0.6173    0.02  0.67  0.78  0.62  0.15
##           RubricVisOrg    0.29122  0.5396    0.16  0.77  0.73  0.58  0.19  0.78
##  Residual                 0.19067  0.4367                                
## Number of obs: 817, groups:  Artifact, 91
## 
## Fixed effects:
##             Estimate Std. Error t value
## (Intercept)  3.17375    0.37780   8.401
## Rater2       0.00266    0.05494   0.048
## Rater3      -0.17375    0.05510  -3.153
## SemesterS19 -0.16929    0.08687  -1.949
## SexF        -0.83571    0.38056  -2.196
## SexM        -0.82337    0.37961  -2.169
## 
## Correlation of Fixed Effects:
##             (Intr) Rater2 Rater3 SmsS19 SexF 
## Rater2      -0.073                           
## Rater3      -0.146  0.499                    
## SemesterS19  0.000  0.008  0.000             
## SexF        -0.983  0.002  0.077 -0.099      
## SexM        -0.984 -0.004  0.069 -0.035  0.978
## optimizer (nloptwrap) convergence code: 0 (OK)
## boundary (singular) fit: see ?isSingular
```

```
# According to p-values for the parameters, we find that only Rater2 is apparently not significa
nt. Rater, Semester, Sex and Repeated are all close related to rating.
```

some of interactions

```
m16 = lmer(Rating ~ 1 + Rater * Semester + (0+Rubric | Artifact),data = tall)
```

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

```
m17 = lmer(Rating ~ 1 + Rater * Sex + (0+Rubric | Artifact),data = tall)
```

```
## fixed-effect model matrix is rank deficient so dropping 2 columns / coefficients
```

```
## Warning in checkConv(attr(opt, "derivs"), opt$par, ctrl = control$checkConv, :
## Model failed to converge with max|grad| = 0.00217225 (tol = 0.002, component 1)
```

```
m18 = lmer(Rating ~ 1 + Rater * Repeated + (0+Rubric | Artifact),data = tall)
```

```
## Warning in checkConv(attr(opt, "derivs"), opt$par, ctrl = control$checkConv, :
## Model failed to converge with max|grad| = 0.0052721 (tol = 0.002, component 1)
```

```
m19 = lmer(Rating ~ 1 + Repeated + Rater + Semester + Rater*Semester + (0+Rubric | Artifact),dat
a = tall)
```

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

```
m20 = lmer(Rating ~ 1 + Repeated + Rater + Semester + Rater*Rubric + Rater*Semester + Semester*R
epeated + (0+Rubric | Artifact),data = tall)
```

```
## 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
```

```
m21 = lmer(Rating ~ 1 + Repeated + Rater + Semester + Semester*Repeated+(0+Rubric | Artifact),da
ta = tall)
```

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

```
anova(m16,m17,m18,m19,m20,m21)
```

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

```
## Data: tall
## Models:
## m16: Rating ~ 1 + Rater * Semester + (0 + Rubric | Artifact)
## m18: Rating ~ 1 + Rater * Repeated + (0 + Rubric | Artifact)
## m21: Rating ~ 1 + Repeated + Rater + Semester + Semester * Repeated + (0 + Rubric | Artifact)
## m17: Rating ~ 1 + Rater * Sex + (0 + Rubric | Artifact)
## m19: Rating ~ 1 + Repeated + Rater + Semester + Rater * Semester + (0 + Rubric | Artifact)
## m20: Rating ~ 1 + Repeated + Rater + Semester + Rater * Rubric + Rater * Semester + Semester
* Repeated + (0 + Rubric | Artifact)
##     npar    AIC    BIC  logLik deviance   Chisq Df Pr(>Chisq)
## m16   35 1530.3 1695.0 -730.16   1460.3
## m18   35 1532.9 1697.6 -731.45   1462.9  0.0000  0
## m21   35 1530.0 1694.7 -729.99   1460.0  2.9050  0
## m17   36 1528.6 1698.0 -728.30   1456.6  3.3772  1    0.06611 .
## m19   36 1530.8 1700.2 -729.41   1458.8  0.0000  0
## m20   55 1471.7 1730.5 -680.86   1361.7 97.1069 19  1.783e-12 ***
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
```

In the interaction we considered, m20 performs the best(AIC = 1471.7). The model performs much better than all of the models which don’t consider variable interactions. It means interaction needs to be considered.

```
summary(m20)
```

```
## Linear mixed model fit by REML ['lmerMod']
## Formula: Rating ~ 1 + Repeated + Rater + Semester + Rater * Rubric + Rater *
##     Semester + Semester * Repeated + (0 + Rubric | Artifact)
##    Data: tall
##
## REML criterion at convergence: 1438.2
##
## Scaled residuals:
##     Min      1Q  Median      3Q     Max
## -2.9739 -0.5261 -0.0533  0.4824  3.5836
##
## Random effects:
##  Groups   Name            Variance Std.Dev. Corr
##  Artifact RubricCritDes   0.50154  0.7082
##           RubricInitEDA   0.35102  0.5925   0.45
##           RubricInterpRes 0.15803  0.3975   0.37 0.82
##           RubricRsrchQ    0.18470  0.4298   0.64 0.46 0.74
##           RubricSelMeth   0.07309  0.2703   0.44 0.62 0.77 0.43
##           RubricTxtOrg    0.25699  0.5069   0.43 0.64 0.73 0.56 0.65
##           RubricVisOrg    0.24743  0.4974   0.35 0.73 0.69 0.53 0.39 0.80
##  Residual                 0.18754  0.4331
## Number of obs: 817, groups:  Artifact, 91
##
## Fixed effects:
##                          Estimate Std. Error t value
## (Intercept)              1.768431   0.121053  14.609
## Repeated                -0.080393   0.114541  -0.702
## Rater2                   0.378470   0.136853   2.766
## Rater3                   0.257412   0.136429   1.887
## SemesterS19             -0.139462   0.116157  -1.201
## RubricInitEDA            0.745704   0.136609   5.459
## RubricInterpRes          1.012990   0.134601   7.526
## RubricRsrchQ             0.750968   0.124249   6.044
## RubricSelMeth            0.429639   0.130574   3.290
## RubricTxtOrg             1.048033   0.134903   7.769
## RubricVisOrg             0.683659   0.139190   4.912
## Rater2:RubricInitEDA    -0.309500   0.172901  -1.790
## Rater3:RubricInitEDA    -0.305290   0.172443  -1.770
## Rater2:RubricInterpRes -0.539666   0.170489  -3.165
## Rater3:RubricInterpRes -0.752291   0.170100  -4.423
## Rater2:RubricRsrchQ     -0.504381   0.162031  -3.113
## Rater3:RubricRsrchQ     -0.374885   0.161477  -2.322
## Rater2:RubricSelMeth    -0.399095   0.165313  -2.414
## Rater3:RubricSelMeth    -0.401004   0.164880  -2.432
## Rater2:RubricTxtOrg     -0.584054   0.171320  -3.409
## Rater3:RubricTxtOrg     -0.496174   0.170882  -2.904
## Rater2:RubricVisOrg     -0.146940   0.174656  -0.841
## Rater3:RubricVisOrg     -0.340358   0.174265  -1.953
## Rater2:SemesterS19      -0.044235   0.125612  -0.352
## Rater3:SemesterS19      -0.128576   0.125604  -1.024
## Repeated:SemesterS19    -0.005197   0.229407  -0.023
```

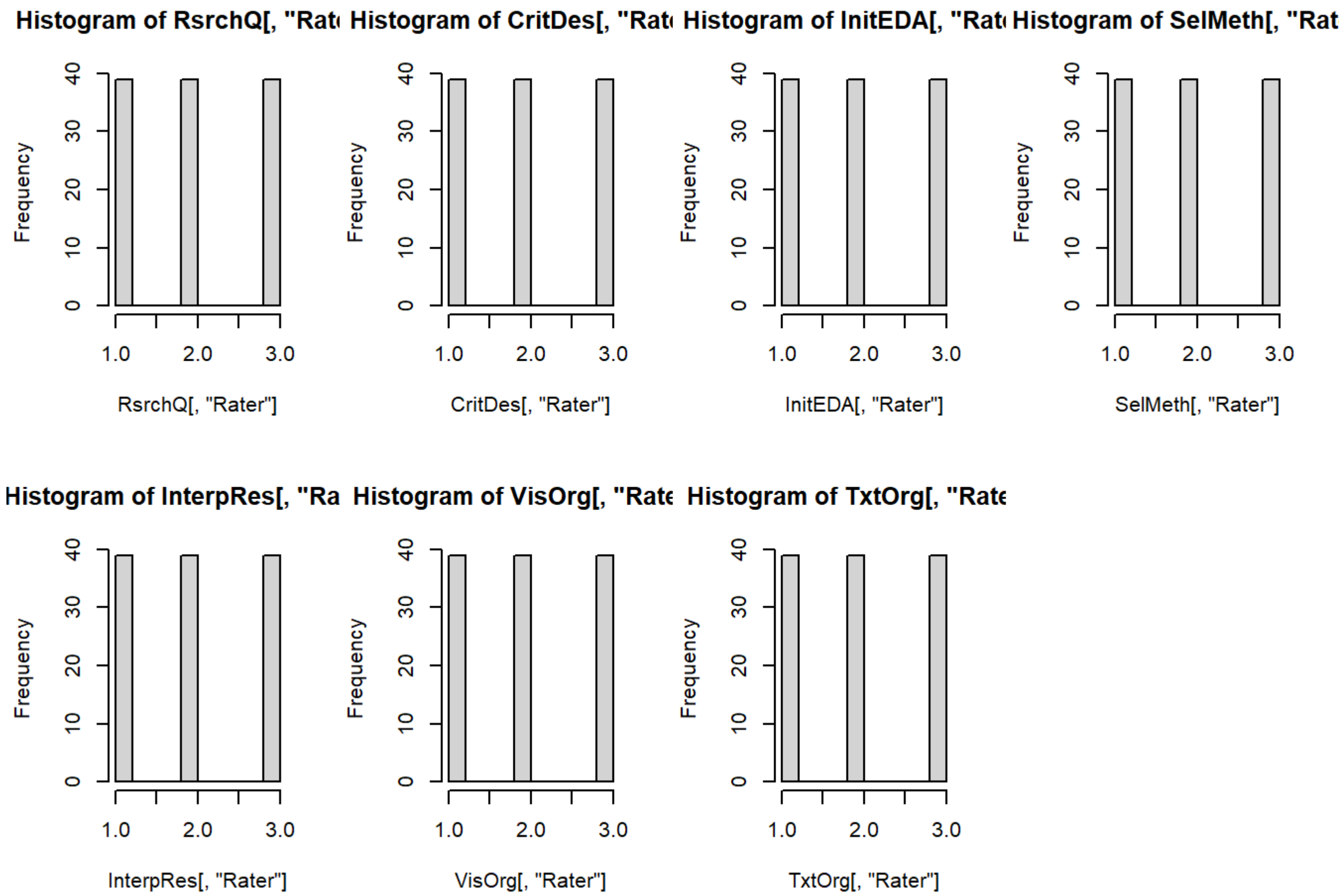
```
##
## Correlation matrix not shown by default, as p = 26 > 12.
## Use print(x, correlation=TRUE)  or
##     vcov(x)        if you need it
```

```
## optimizer (nloptwrap) convergence code: 0 (OK)
## unable to evaluate scaled gradient
## Model failed to converge: degenerate  Hessian with 1 negative eigenvalues
```

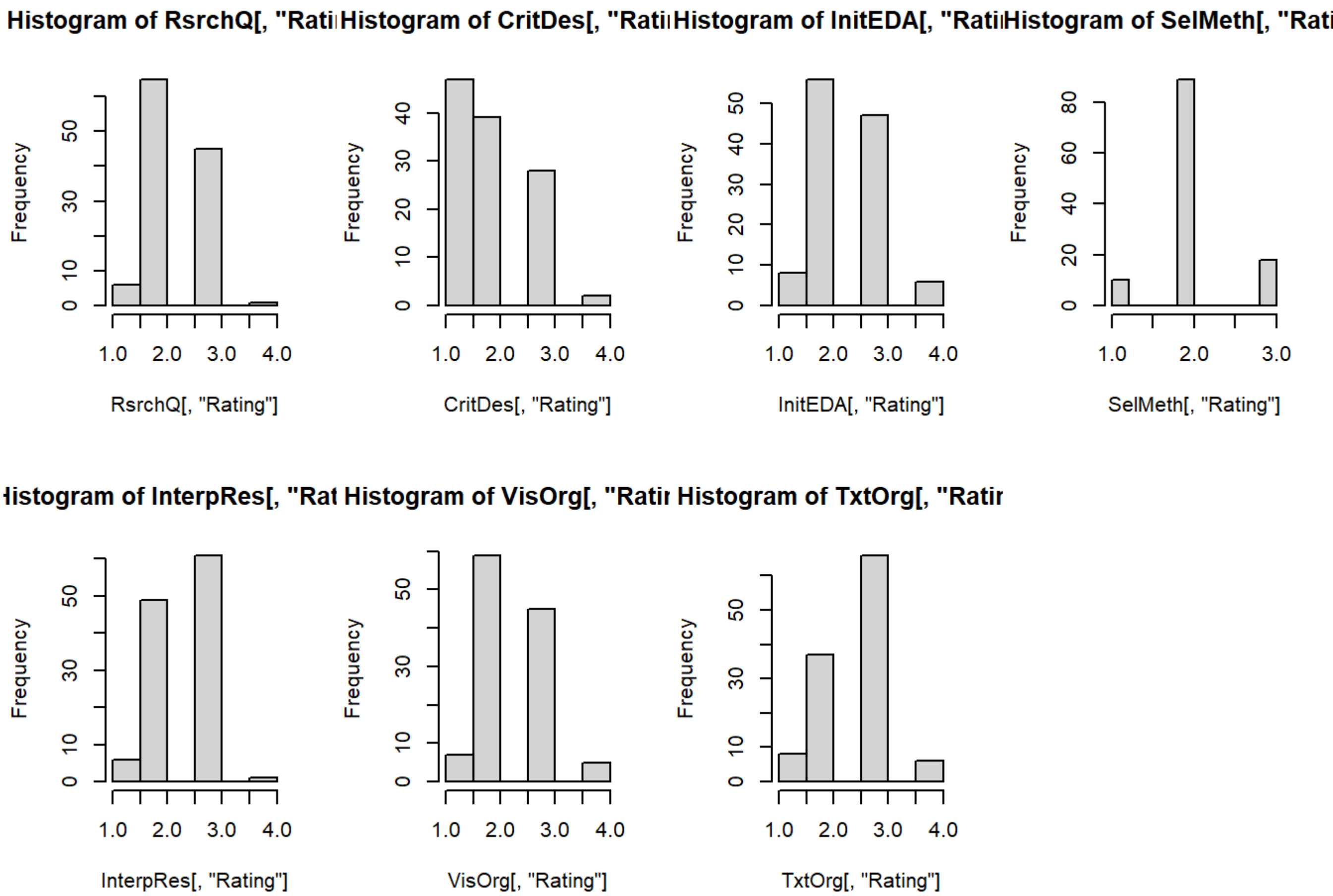
```
# It seems repeated, Rater:Semester, and Repeated:Semester are not significant. And we can then
 check more interactions to find out which interactions are significant and not significant.
```

q4

```
tall <- read.csv("C:/Users/danie/Desktop/CMU/36-617 applied regression analysis/Project 2/tall.c
sv")
par(mfrow = c(2,4))
RsrchQ = tall[tall[,"Rubric"] == "RsrchQ",]
CritDes = tall[tall[,"Rubric"] == "CritDes",]
InitEDA = tall[tall[,"Rubric"] == "InitEDA",]
SelMeth = tall[tall[,"Rubric"] == "SelMeth",]
InterpRes = tall[tall[,"Rubric"] == "InterpRes",]
VisOrg = tall[tall[,"Rubric"] == "VisOrg",]
TxtOrg = tall[tall[,"Rubric"] == "TxtOrg",]

hist(RsrchQ[,"Rater"])
hist(CritDes[,"Rater"])
hist(InitEDA[,"Rater"])
hist(SelMeth[,"Rater"])
hist(InterpRes[,"Rater"])
hist(VisOrg[,"Rater"])
hist(TxtOrg[,"Rater"])
# The raters for all rubrics have the same distribution
```




```
par(mfrow = c(2,4))
hist(RsrchQ[,"Rating"])
hist(CritDes[,"Rating"])
hist(InitEDA[,"Rating"])
hist(SelMeth[,"Rating"])
hist(InterpRes[,"Rating"])
hist(VisOrg[,"Rating"])
hist(TxtOrg[,"Rating"])
# It is interesting that only rubric CritDes has value 1 as majority, and all other rubrics all
 have a little value 1.
```


Histogram of RsrchQ[, "Rati
Histogram of CritDes[, "Rati
Histogram of InitEDA[, "Rati
Histogram of SelMeth[, "Rati
Frequency
RsrchQ[, "Rating"]
CritDes[, "Rating"]
InitEDA[, "Rating"]
SelMeth[, "Rating"]
istogram of InterpRes[, "Rat
Histogram of VisOrg[, "Ratir
Histogram of TxtOrg[, "Ratir
InterpRes[, "Rating"]
VisOrg[, "Rating"]
TxtOrg[, "Rating"]