

# Explanability Study

Michael Amodeo, Krista Mar, Mona Iwamoto

August 14, 2017

## Data Import and Preperation

### Import data

In order to appropriately capture the instances where someone did not finish the survey, whether by clicking through to the final page or by stopping midway, I had to export from Qualtrics in a different format that allows us to see what was viewed and what was not. That also changed some other fields, particularly the multiple choice, multiple selection questions.

```
#Import all data
all_content = readLines("Explainability_Study_legacy_export.csv")

#Delete second and third rows of not useful information
skip_second = all_content[-c(2,3)]

#Create table and data.table
d <- read.csv(textConnection(skip_second), header = TRUE, stringsAsFactors = FALSE)
d <- data.table(d)

remove(all_content, skip_second)

# Create new data table without fields we are not using
dt <- d[, -c('ResponseSet', 'IPAddress', 'StartDate', 'EndDate', 'RecipientLastName',
            'RecipientFirstName', 'RecipientEmail', 'ExternalDataReference', 'Status',
            'Q_TotalDuration', 'Enter.Embedded.Data.Field.Name.Here...', 'LocationLatitude',
            'LocationLongitude', 'LocationAccuracy', 'Q3.5', 'Q4.5', 'Q6.5', 'Q7.5', 'Q8.1',
            'Q9.1', 'Q10.1', 'Q10.3')]

# Rename variables
old_names <- colnames(dt)

## Key to var names: tc = Twitter control group
#                   tt = Twitter treatment group
#                   rc = recidivism control group
#                   rt = recidivism treatment group

new_names <- c("ResponseID", "Finished", "First.Context", "random", "intro", "tweet",
              "tControl", "tcFair", "tcAcc", "tcSat", "tcUseful", "tcClear",
              "tcMeaningful", "tcReqInfo1", "tcReqInfo2", "tcReqInfo3", "tcReqInfo4", "tcReqInfo4_txt",
              "tTreat", "ttFair", "ttAcc", "ttSat", "ttUseful", "ttClear",
              "ttMeaningful", "ttReqInfo1", "ttReqInfo2", "ttReqInfo3", "ttReqInfo4", "ttReqInfo4_txt",
              "recidivism",
              "rControl", "rcFair", "rcAcc", "rcSat", "rcUseful", "rcClear",
              "rcMeaningful", "rcReqInfo1", "rcReqInfo2", "rcReqInfo3", "rcReqInfo4", "rcReqInfo4_txt",
              "rTreat", "rtFair", "rtAcc", "rtSat", "rtUseful", "rtClear",
              "rtMeaningful", "rtReqInfo1", "rtReqInfo2", "rtReqInfo3", "rtReqInfo4", "rtReqInfo4_txt",
              "ageGroup", "white", "black", "native", "asian", "pac_isle", "hispanic",
```

```

        'other', 'gender', 'socMed', 'educ', 'feedback')

setnames(dt, old_names, new_names)
colnames(dt)

## [1] "ResponseID"      "Finished"        "First.Context"   "random"
## [5] "intro"           "tweet"           "tControl"        "tcFair"
## [9] "tcAcc"           "tcSat"           "tcUseful"        "tcClear"
## [13] "tcMeaningful"    "tcReqInfo1"      "tcReqInfo2"      "tcReqInfo3"
## [17] "tcReqInfo4"      "tcReqInfo4_txt"  "tTreat"          "ttFair"
## [21] "ttAcc"           "ttSat"           "ttUseful"        "ttClear"
## [25] "ttMeaningful"    "ttReqInfo1"      "ttReqInfo2"      "ttReqInfo3"
## [29] "ttReqInfo4"      "ttReqInfo4_txt"  "recidivism"      "rControl"
## [33] "rcFair"          "rcAcc"           "rcSat"           "rcUseful"
## [37] "rcClear"         "rcMeaningful"    "rcReqInfo1"      "rcReqInfo2"
## [41] "rcReqInfo3"      "rcReqInfo4"      "rcReqInfo4_txt"  "rTreat"
## [45] "rtFair"          "rtAcc"           "rtSat"           "rtUseful"
## [49] "rtClear"         "rtMeaningful"    "rtReqInfo1"      "rtReqInfo2"
## [53] "rtReqInfo3"      "rtReqInfo4"      "rtReqInfo4_txt"  "ageGroup"
## [57] "white"           "black"           "native"          "asian"
## [61] "pac_isle"        "hispanic"        "other"           "gender"
## [65] "socMed"          "educ"            "feedback"

remove(old_names)

colnames(d)

## [1] "ResponseID"
## [2] "ResponseSet"
## [3] "IPAddress"
## [4] "StartDate"
## [5] "EndDate"
## [6] "RecipientLastName"
## [7] "RecipientFirstName"
## [8] "RecipientEmail"
## [9] "ExternalDataReference"
## [10] "Finished"
## [11] "Status"
## [12] "Q_TotalDuration"
## [13] "First.Context"
## [14] "Enter.Embedded.Data.Field.Name.Here..."
## [15] "random"
## [16] "Q1.1"
## [17] "Q2.1"
## [18] "Q3.1"
## [19] "Q3.2"
## [20] "Q3.3"
## [21] "Q3.4"
## [22] "Q3.5"
## [23] "Q3.6"
## [24] "Q3.7"
## [25] "Q3.8"
## [26] "Q3.9_1"
## [27] "Q3.9_2"
## [28] "Q3.9_3"

```

```

## [29] "Q3.9_4"
## [30] "Q3.9_4_TEXT"
## [31] "Q4.1"
## [32] "Q4.2"
## [33] "Q4.3"
## [34] "Q4.4"
## [35] "Q4.5"
## [36] "Q4.6"
## [37] "Q4.7"
## [38] "Q4.8"
## [39] "Q4.9_1"
## [40] "Q4.9_2"
## [41] "Q4.9_3"
## [42] "Q4.9_4"
## [43] "Q4.9_4_TEXT"
## [44] "Q8.1"
## [45] "Q5.1"
## [46] "Q6.1"
## [47] "Q6.2"
## [48] "Q6.3"
## [49] "Q6.4"
## [50] "Q6.5"
## [51] "Q6.6"
## [52] "Q6.7"
## [53] "Q6.8"
## [54] "Q6.9_1"
## [55] "Q6.9_2"
## [56] "Q6.9_3"
## [57] "Q6.9_4"
## [58] "Q6.9_4_TEXT"
## [59] "Q7.1"
## [60] "Q7.2"
## [61] "Q7.3"
## [62] "Q7.4"
## [63] "Q7.5"
## [64] "Q7.6"
## [65] "Q7.7"
## [66] "Q7.8"
## [67] "Q7.9_1"
## [68] "Q7.9_2"
## [69] "Q7.9_3"
## [70] "Q7.9_4"
## [71] "Q7.9_4_TEXT"
## [72] "Q9.1"
## [73] "Q9.2"
## [74] "Q9.3_1"
## [75] "Q9.3_2"
## [76] "Q9.3_3"
## [77] "Q9.3_4"
## [78] "Q9.3_5"
## [79] "Q9.3_6"
## [80] "Q9.3_7"
## [81] "Q9.4"
## [82] "Q9.5"

```

```
## [83] "Q9.6"
## [84] "Q10.1"
## [85] "Q10.2"
## [86] "Q10.3"
## [87] "LocationLatitude"
## [88] "LocationLongitude"
## [89] "LocationAccuracy"
```

```
#summary(dt)
```

## Data Cleanup

```
# Most Qualtrics questions were set so the extreme positive value was the first choice (1)
# Function to flip the scale to show more positive as larger number
flip <- function(originalScale) {
  x <- originalScale - 3      # 3 is median
  return(3 - x)
}

# For an unknown reason, question 7.2 Fairness for Recidivism Treatment
# the values were offset by 24. This was cross-checked with the text-based responses.

dt$rtFair <- dt$rtFair - 24    # Qualtrics weirdness

# For the Twitter fairness questions, the Qualtrics survey
# responses were reversed in two instances. All others
# are reversed using the flip function below

# Organize scales so larger values correlate to more fair, more accurate, etc.
dt[, tcAcc      := flip(dt[, tcAcc])]
dt[, tcSat      := flip(dt[, tcSat])]
dt[, tcUseful   := flip(dt[, tcUseful])]
dt[, tcClear    := flip(dt[, tcClear])]
dt[, tcMeaningful := flip(dt[, tcMeaningful])]

dt[, ttFair     := flip(dt[, ttFair])]
dt[, ttAcc      := flip(dt[, ttAcc])]
dt[, ttSat      := flip(dt[, ttSat])]
dt[, ttUseful   := flip(dt[, ttUseful])]
dt[, ttClear    := flip(dt[, ttClear])]
dt[, ttMeaningful := flip(dt[, ttMeaningful])]

dt[, rcAcc      := flip(dt[, rcAcc])]
dt[, rcSat      := flip(dt[, rcSat])]
dt[, rcUseful   := flip(dt[, rcUseful])]
dt[, rcClear    := flip(dt[, rcClear])]
dt[, rcMeaningful := flip(dt[, rcMeaningful])]

dt[, rtFair     := flip(dt[, rtFair])]
dt[, rtAcc      := flip(dt[, rtAcc])]
dt[, rtSat      := flip(dt[, rtSat])]
dt[, rtUseful   := flip(dt[, rtUseful])]
dt[, rtClear    := flip(dt[, rtClear])]
```

```
dt[, rtMeaningful := flip(dt[, rtMeaningful])]
```

## Consolidate each metric across treatments

```
# Create consolidated data table

dc <- data.table(tFair = rowSums(dt[, c('tcFair', 'ttFair')], na.rm=T))

dc[, complete := !is.na(dt[, random])]
dc[, tAssign := dt[, tTreat] == 1 | dt[, tControl] == 1]
dc[, tControl := !is.na(dt[, tControl])]
dc[, tTreat := !is.na(dt[, tTreat])]
dc[, rControl := !is.na(dt[, rControl])]
dc[, rTreat := !is.na(dt[, rTreat])]
dc[, rAssign := dt[, rTreat] == 1 | dt[, rControl] == 1]
dc[, tweet := !is.na(dt[, tweet])]
dc[, recidivism := !is.na(dt[, recidivism])]

dc[, tAcc := rowSums(dt[, c('tcAcc', 'ttAcc')], na.rm=T)]
dc[, tSat := rowSums(dt[, c('tcSat', 'ttSat')], na.rm=T) ]
dc[, tUseful := rowSums(dt[, c('tcUseful', 'ttUseful')], na.rm=T)]
dc[, tClear := rowSums(dt[, c('tcClear', 'ttClear')], na.rm=T)]
dc[, tMeaningful := rowSums(dt[, c('tcMeaningful', 'ttMeaningful')], na.rm=T)]

dc[, rFair := rowSums(dt[,c('rcFair', 'rtFair')], na.rm=T)]
dc[, rAcc := rowSums(dt[,c('rcAcc', 'rtAcc')], na.rm=T)]
dc[, rSat := rowSums(dt[,c('rcSat', 'rtSat')], na.rm=T) ]
dc[, rUseful := rowSums(dt[,c('rcUseful', 'rtUseful')], na.rm=T)]
dc[, rClear := rowSums(dt[,c('rcClear', 'rtClear')], na.rm=T)]
dc[, rMeaningful := rowSums(dt[,c('rcMeaningful', 'rtMeaningful')], na.rm=T)]

dc[, tReqInfo1 := rowSums(dt[,c('tcReqInfo1', 'ttReqInfo1')], na.rm=T)]
dc[, tReqInfo2 := rowSums(dt[,c('tcReqInfo2', 'ttReqInfo2')], na.rm=T)]
dc[, tReqInfo3 := rowSums(dt[,c('tcReqInfo3', 'ttReqInfo3')], na.rm=T)]
#dc[, tOther4 := rowSums(dt[,c('tcOther4', 'ttOther4')], na.rm=T)]

dc[, rReqInfo1 := rowSums(dt[,c('rcReqInfo1', 'rtReqInfo1')], na.rm=T)]
dc[, rReqInfo2 := rowSums(dt[,c('rcReqInfo2', 'rtReqInfo2')], na.rm=T)]
dc[, rReqInfo3 := rowSums(dt[,c('rcReqInfo3', 'rtReqInfo3')], na.rm=T)]
#dc[, rOther4 := rowSums(dt[,c('rcOther4', 'rtOther4')], na.rm=T)]

dc[, white := !is.na(dt[, white])]
dc[, black := !is.na(dt[, black])]
dc[, native := !is.na(dt[, native])]
dc[, asian := !is.na(dt[, asian])]
dc[, pac_isle := !is.na(dt[, pac_isle])]
dc[, hispanic := !is.na(dt[, hispanic])]
dc[, other := !is.na(dt[, other])]

dc[, female := (dt[, gender]==2)]
dc[, gender_nc := (dt[, gender]==3)]
```

```

dt1 <- dt[, c('ageGroup', 'socMed', 'educ', 'First.Context')]

dc <- cbind(dc, dt1)

# Converting tTreat and rTreat to binaries instead of logicals
(to.replace <- names(which(sapply(dc, is.logical))))
for (var in to.replace) dc[, var:= as.numeric(get(var)), with=FALSE]

#view dc
head(dc)

```

## Randomization Check

### Were the two contexts assigned equally?

The first randomization was to assign which context the respondent would see first. Did that work?

```
dc[, .N, by = First.Context]
```

```

##      First.Context    N
## 1:      Recidivism 320
## 2:         Twitter 321

```

320 received the 'Recidivism' context first. 321 received the 'Twitter' context first. This was a pretty even split. Next is worth checking if each context received similar assignment to treatment.

```
dc[, .N, by = .(tTreat, tAssign)]
```

```

##      tTreat tAssign    N
## 1:         0         1 315
## 2:         1         1 312
## 3:         0        NA   14

```

In this instance, we see that of those assigned to the Twitter context (tAssign), it was a pretty even split between treatment and control. However, those 14 that were not assigned to either treatment or control are indicative of attrition that we will need to review in greater detail.

```
dc[, .N, by = .(rTreat, rAssign)]
```

```

##      rTreat rAssign    N
## 1:         0         1 312
## 2:         1         1 316
## 3:         0        NA   13

```

Similarly, we see a pretty even split between recidivism context assignment, with another 13 instances of attrition. These could overlap with the other examples of attrition.

From the analysis above, it appears that 315 respondents were assigned to the Twitter-control group, 312 were assigned to the Twitter-treatment group, 312 were assigned to the recidivism-control group and 316 were assigned to the recidivism-treatment group.

### Was treatment assigned equally across contexts?

```
dc[tAssign == 1, .N, by = .(First.Context, tTreat)]
```

```

##      First.Context tTreat    N
## 1:      Recidivism         0 155

```

```
## 2:      Twitter      1 158
## 3:    Recidivism      1 154
## 4:      Twitter      0 160
```

```
dc[rAssign ==1, .N, by = .(First.Context, rTreat)]
```

```
##      First.Context rTreat    N
## 1:    Recidivism      0 155
## 2:      Twitter      0 157
## 3:    Recidivism      1 158
## 4:      Twitter      1 158
```

These two tables show that treatment and control were assigned roughly equally across contexts, regardless of order.

### Were all questions answered?

```
## Show number of responses for each question
apply(dt, 2, function(x) length(which(!is.na(x))))
```

```
##      ResponseID      Finished First.Context      random      intro
##           641           641           641           622           641
##      tweet      tControl      tcFair      tcAcc      tcSat
##           631           315           315           315           315
##      tcUseful      tcClear      tcMeaningful      tcReqInfo1      tcReqInfo2
##           315           315           315           84           173
##      tcReqInfo3      tcReqInfo4 tcReqInfo4_txt      tTreat      ttFair
##           145           9           641           312           312
##      ttAcc      ttSat      ttUseful      ttClear      ttMeaningful
##           312           312           311           311           311
##      ttReqInfo1      ttReqInfo2      ttReqInfo3      ttReqInfo4 ttReqInfo4_txt
##           93           144           129           15           641
##      recidivism      rControl      rcFair      rcAcc      rcSat
##           636           312           312           312           312
##      rcUseful      rcClear      rcMeaningful      rcReqInfo1      rcReqInfo2
##           310           310           310           90           176
##      rcReqInfo3      rcReqInfo4 rcReqInfo4_txt      rTreat      rtFair
##           175           12           641           316           316
##      rtAcc      rtSat      rtUseful      rtClear      rtMeaningful
##           316           316           315           315           315
##      rtReqInfo1      rtReqInfo2      rtReqInfo3      rtReqInfo4 rtReqInfo4_txt
##           89           165           154           15           641
##      ageGroup      white      black      native      asian
##           622           422           36           14           134
##      pac_isle      hispanic      other      gender      socMed
##           1           30           5           620           622
##      educ      feedback
##           622           637
```

From this, it appears that there were a couple instances of attrition in the middle of answering questions about a treatment. Note the drop from 312 to 311 between ttSat and ttUseful, or the drop from 316 to 315 between rtSat and rtUseful.

### Attrition effects

Out of 641 surveys, 622 were completed. Was either context more impacted than the other?

```
dc[, sum(complete)/.N, by = First.Context]
```

```
##      First.Context      V1
## 1:      Recidivism 0.9656250
## 2:      Twitter 0.9750779
```

Similar ratios completed the survey regardless of which context they started with. This does not seem indicative of a problem with the experiment, but we will need to be careful about how we calculate effects.

## Define Metrics

The metrics we evaluated were split into two groups. The first three asked respondents to rate the decision that was made with respect to fairness, accuracy, and their satisfaction with the decision. The second three asked specifically about the explanation itself. Respondents were asked if the explanation was useful, clear, and meaningful.

### Visual data exploration, grouped bars

#### Twitter Response Histograms

```
par(mfrow=c(2,3))
hist(dt$tcFair, col=rgb(0,0,1,1/4), breaks = seq(0.5, 5.5, 1),
     main= "Twitter Fairness", xlab="Fairness", ylim=c(0,175))
hist(dt$ttFair,col=rgb(1,0,0,1/4), breaks = seq(0.5, 5.5, 1), add=T)
#legend("topright", c("Control", "Treatment"), fill=c("blue", "red"))

hist(dt$tcAcc, col=rgb(0,0,1,1/4), breaks = seq(0.5, 5.5, 1),
     main= "Twitter Accuracy", xlab="Accuracy", ylim=c(0,175))
hist(dt$ttAcc,col=rgb(1,0,0,1/4), breaks = seq(0.5, 5.5, 1), add=T)

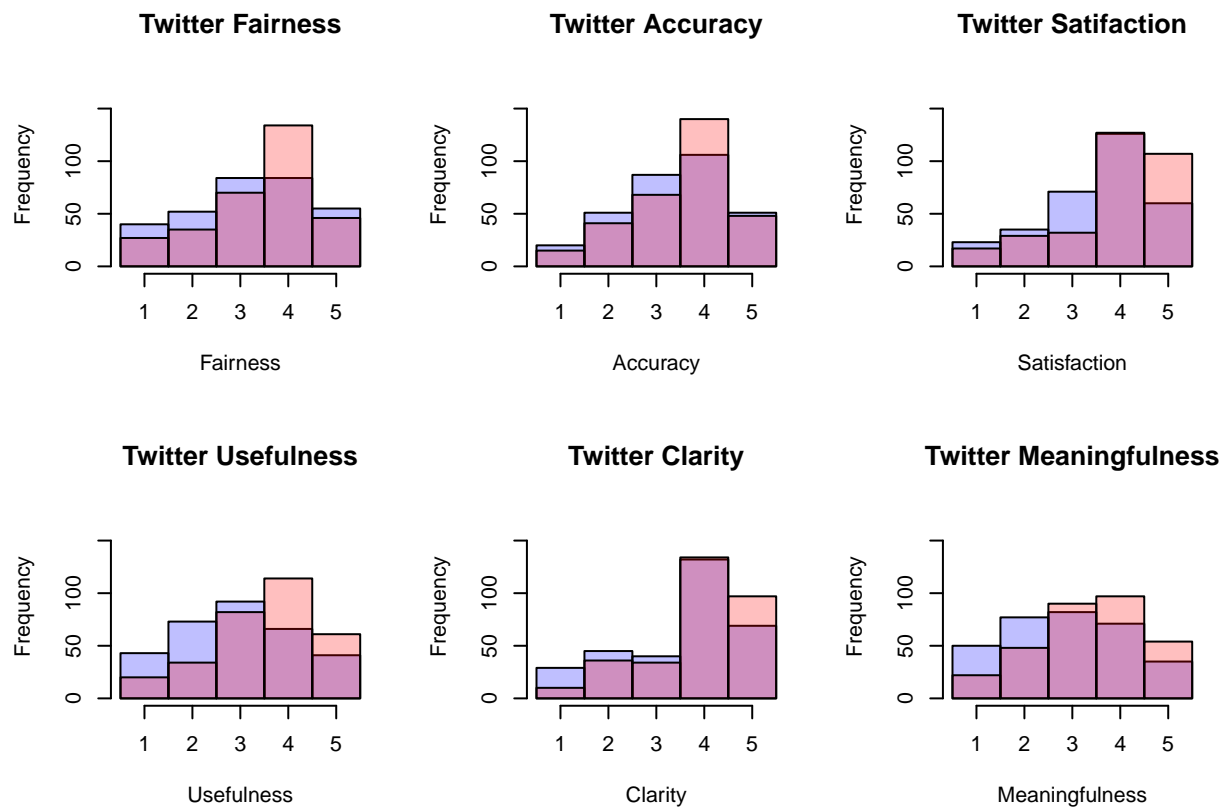
hist(dt$tcSat, col=rgb(0,0,1,1/4), breaks = seq(0.5, 5.5, 1),
     main= "Twitter Satisfaction", xlab="Satisfaction", ylim=c(0,175))
hist(dt$ttSat,col=rgb(1,0,0,1/4), breaks = seq(0.5, 5.5, 1), add=T)

hist(dt$tcUseful, col=rgb(0,0,1,1/4), breaks = seq(0.5, 5.5, 1),
     main= "Twitter Usefulness", xlab="Usefulness", ylim=c(0,175))
hist(dt$ttUseful,col=rgb(1,0,0,1/4), breaks = seq(0.5, 5.5, 1), add=T)

hist(dt$tcClear, col=rgb(0,0,1,1/4), breaks = seq(0.5, 5.5, 1),
     main= "Twitter Clarity", xlab="Clarity", ylim=c(0,175))
hist(dt$ttClear,col=rgb(1,0,0,1/4), breaks = seq(0.5, 5.5, 1), add=T)

hist(dt$tcMeaningful, col=rgb(0,0,1,1/4), breaks = seq(0.5, 5.5, 1),
     main= "Twitter Meaningfulness", xlab="Meaningfulness", ylim=c(0,175))
hist(dt$ttMeaningful,col=rgb(1,0,0,1/4), breaks = seq(0.5, 5.5, 1), add=T)
```





### Recidivism Responses Histogram

```
par(mfrow=c(2,3))
hist(dt$rcFair, col=rgb(0,0,1,1/4), breaks = seq(0.5, 5.5, 1), main= "Recidivism Fairness", xlab="Fairness", add=T)
hist(dt$rtFair,col=rgb(1,0,0,1/4), breaks = seq(0.5, 5.5, 1), add=T)
#legend("topright", c("Control", "Treatment"), fill=c("blue", "red"))

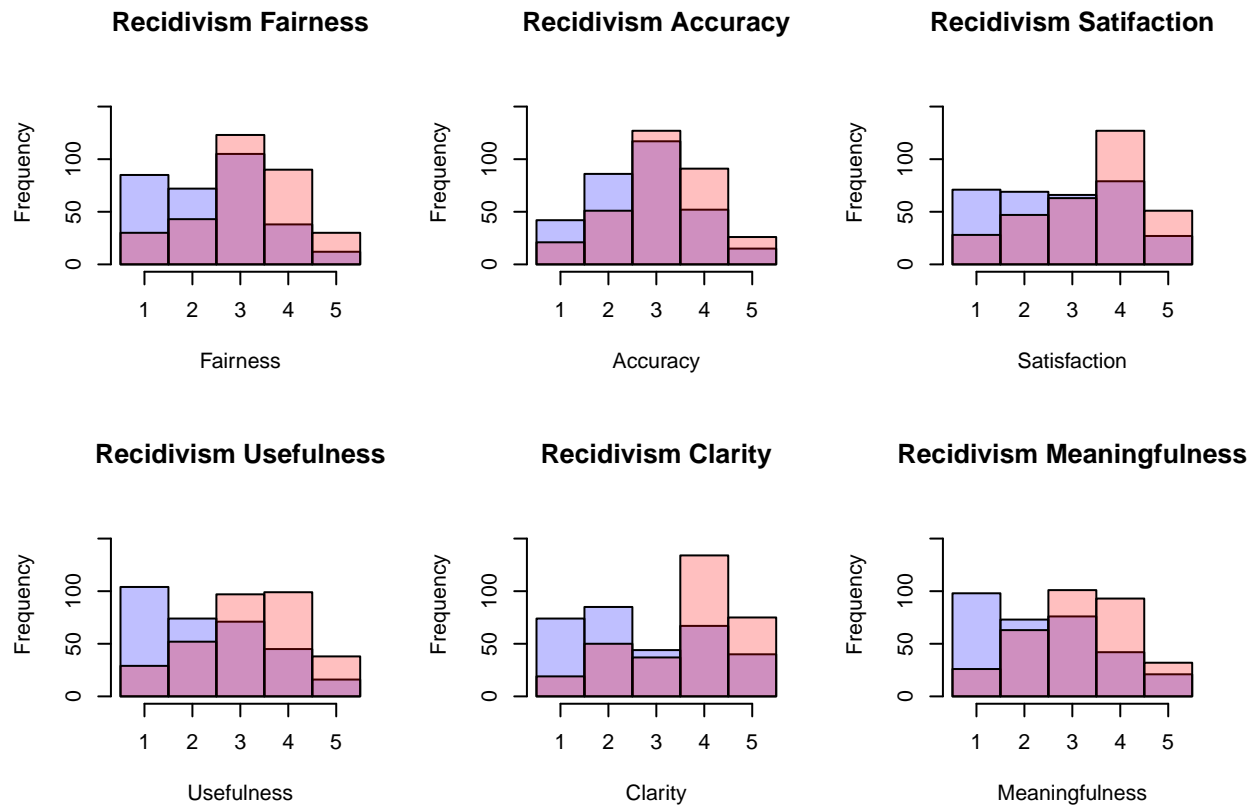
hist(dt$rcAcc, col=rgb(0,0,1,1/4), breaks = seq(0.5, 5.5, 1), main= "Recidivism Accuracy", xlab="Accuracy", add=T)
hist(dt$rtAcc,col=rgb(1,0,0,1/4), breaks = seq(0.5, 5.5, 1), add=T)

hist(dt$rcSat, col=rgb(0,0,1,1/4), breaks = seq(0.5, 5.5, 1), main= "Recidivism Satisfaction", xlab="Satisfaction", add=T)
hist(dt$rtSat,col=rgb(1,0,0,1/4), breaks = seq(0.5, 5.5, 1), add=T)

hist(dt$rcUseful, col=rgb(0,0,1,1/4), breaks = seq(0.5, 5.5, 1), main= "Recidivism Usefulness", xlab="Usefulness", add=T)
hist(dt$rtUseful,col=rgb(1,0,0,1/4), breaks = seq(0.5, 5.5, 1), add=T)

hist(dt$rcClear, col=rgb(0,0,1,1/4), breaks = seq(0.5, 5.5, 1), main= "Recidivism Clarity", xlab="Clarity", add=T)
hist(dt$rtClear,col=rgb(1,0,0,1/4), breaks = seq(0.5, 5.5, 1), add=T)

hist(dt$rcMeaningful, col=rgb(0,0,1,1/4), breaks = seq(0.5, 5.5, 1), main= "Recidivism Meaningfulness", xlab="Meaningfulness", add=T)
hist(dt$rtMeaningful,col=rgb(1,0,0,1/4), breaks = seq(0.5, 5.5, 1), add=T)
```



### Twitter Histograms Greyscale

```
par(mfrow=c(2,3))
hist(dt$tcFair, col=rgb(0.1,0.1,0.1,0.5), breaks = seq(0.5, 5.5, 1), main= "Twitter Fairness", xlab="Fairness", add=T)
hist(dt$ttFair,col=rgb(0.8,0.8,0.8,0.5), breaks = seq(0.5, 5.5, 1), add=T)
#legend("topright", c("Control", "Treatment"), fill=c("blue", "red"))

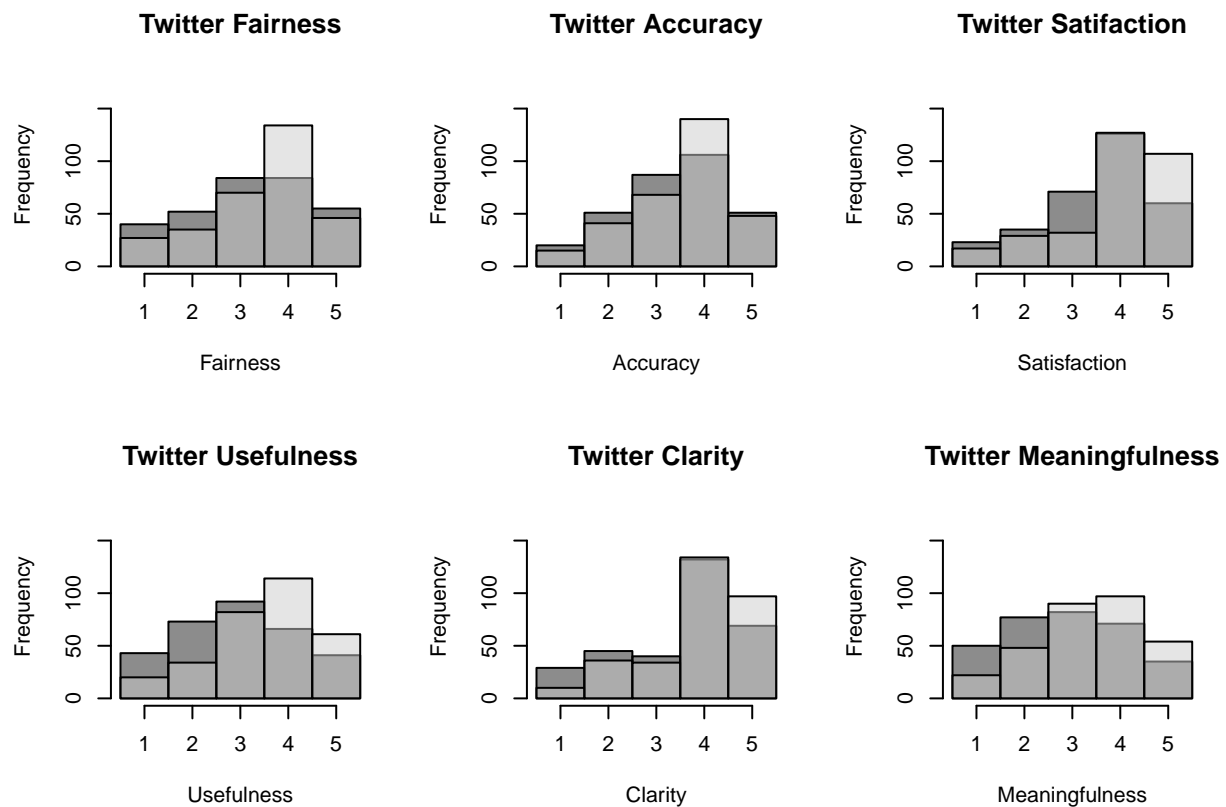
hist(dt$tcAcc, col=rgb(0.1,0.1,0.1,0.5), breaks = seq(0.5, 5.5, 1), main= "Twitter Accuracy", xlab="Accuracy", add=T)
hist(dt$ttAcc,col=rgb(0.8,0.8,0.8,0.5), breaks = seq(0.5, 5.5, 1), add=T)

hist(dt$tcSat, col=rgb(0.1,0.1,0.1,0.5), breaks = seq(0.5, 5.5, 1), main= "Twitter Satisfaction", xlab="Satisfaction", add=T)
hist(dt$ttSat,col=rgb(0.8,0.8,0.8,0.5), breaks = seq(0.5, 5.5, 1), add=T)

hist(dt$tcUseful, col=rgb(0.1,0.1,0.1,0.5), breaks = seq(0.5, 5.5, 1), main= "Twitter Usefulness", xlab="Usefulness", add=T)
hist(dt$ttUseful,col=rgb(0.8,0.8,0.8,0.5), breaks = seq(0.5, 5.5, 1), add=T)

hist(dt$tcClear, col=rgb(0.1,0.1,0.1,0.5), breaks = seq(0.5, 5.5, 1), main= "Twitter Clarity", xlab="Clarity", add=T)
hist(dt$ttClear,col=rgb(0.8,0.8,0.8,0.5), breaks = seq(0.5, 5.5, 1), add=T)

hist(dt$tcMeaningful, col=rgb(0.1,0.1,0.1,0.5), breaks = seq(0.5, 5.5, 1), main= "Twitter Meaningfulness", xlab="Meaningfulness", add=T)
hist(dt$ttMeaningful,col=rgb(0.8,0.8,0.8,0.5), breaks = seq(0.5, 5.5, 1), add=T)
```



### Recidivism Histogram Greyscale

```
par(mfrow=c(2,3))
hist(dt$rcFair, col=rgb(0.1,0.1,0.1,0.5), breaks = seq(0.5, 5.5, 1), main= "Recidivism Fairness", xlab="Fairness")
hist(dt$rtFair,col=rgb(0.8,0.8,0.8,0.5), breaks = seq(0.5, 5.5, 1), add=T)
#legend("topright", c("Control", "Treatment"), fill=c("blue", "red"))

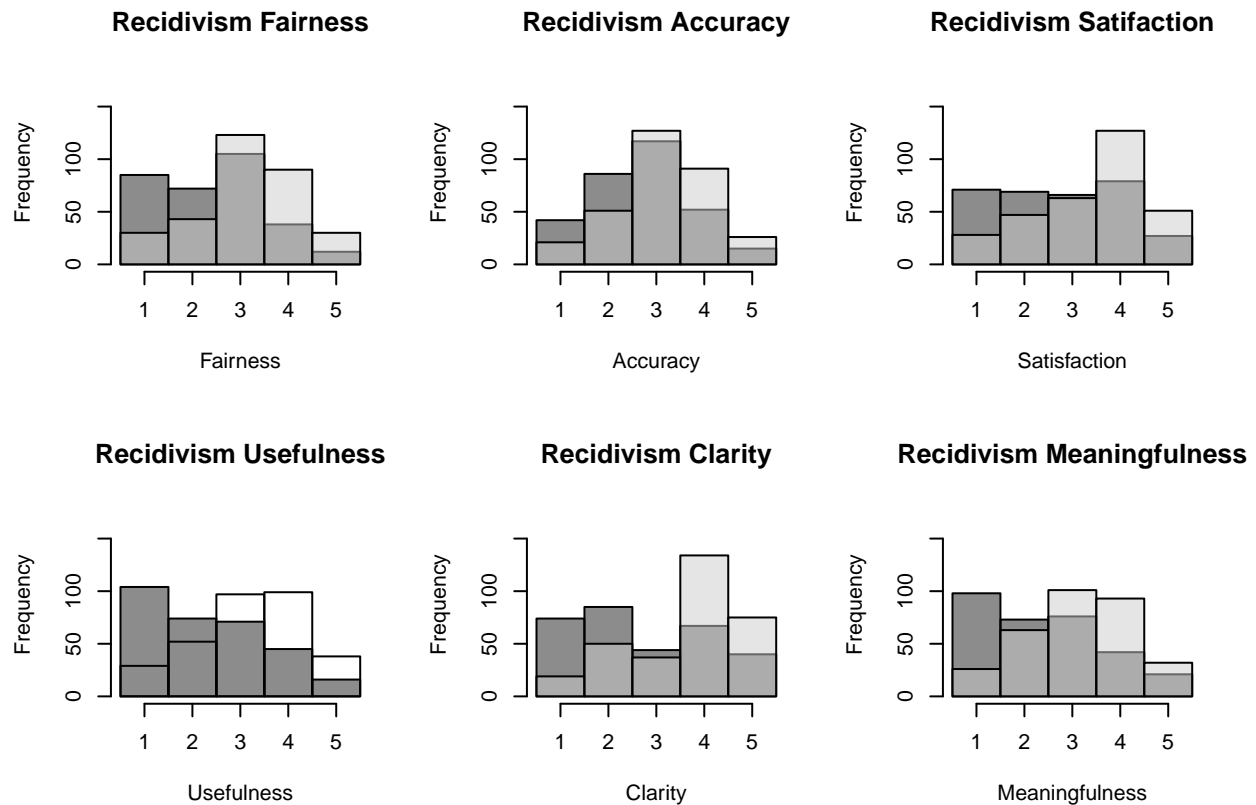
hist(dt$rcAcc, col=rgb(0.1,0.1,0.1,0.5), breaks = seq(0.5, 5.5, 1), main= "Recidivism Accuracy", xlab="Accuracy")
hist(dt$rtAcc,col=rgb(0.8,0.8,0.8,0.5), breaks = seq(0.5, 5.5, 1), add=T)

hist(dt$rcSat, col=rgb(0.1,0.1,0.1,0.5), breaks = seq(0.5, 5.5, 1), main= "Recidivism Satisfaction", xlab="Satisfaction")
#legend(4,9, Treat(df),lwd=4, col=c())
hist(dt$rtSat,col=rgb(0.8,0.8,0.8,0.5), breaks = seq(0.5, 5.5, 1), add=T)

hist(dt$rcUseful, col=rgb(0.1,0.1,0.1,0.5), breaks = seq(0.5, 5.5, 1), main= "Recidivism Usefulness", xlab="Usefulness")
hist(dt$rtUseful,col=rgb(0.8,0.8,0.8,0.5), breaks = seq(0.5, 5.5, 1), add=T)

hist(dt$rcClear, col=rgb(0.1,0.1,0.1,0.5), breaks = seq(0.5, 5.5, 1), main= "Recidivism Clarity", xlab="Clarity")
hist(dt$rtClear,col=rgb(0.8,0.8,0.8,0.5), breaks = seq(0.5, 5.5, 1), add=T)

hist(dt$rcMeaningful, col=rgb(0.1,0.1,0.1,0.5), breaks = seq(0.5, 5.5, 1), main= "Recidivism Meaningfulness", xlab="Meaningfulness")
hist(dt$rtMeaningful,col=rgb(0.8,0.8,0.8,0.5), breaks = seq(0.5, 5.5, 1), add=T)
```



###What Additional Information did respondents want

```
tcsumReqInfo1<-sum(dt$tcReqInfo1, na.rm=TRUE)
#tcsumother1
tcsumReqInfo2<-sum(dt$tcReqInfo2, na.rm=TRUE)
#tcsumother2
tcsumReqInfo3<-sum(dt$tcReqInfo3, na.rm=TRUE)
#tcsumother3

ttsumReqInfo1<-sum(dt$ttReqInfo1, na.rm=TRUE)
#ttsumother1
ttsumReqInfo2<-sum(dt$ttReqInfo2, na.rm=TRUE)
#ttsumother2
ttsumReqInfo3<-sum(dt$ttReqInfo3, na.rm=TRUE)
#ttsumother3

rcsumReqInfo1<-sum(dt$rcReqInfo1, na.rm=TRUE)
rcsumReqInfo2<-sum(dt$rcReqInfo2, na.rm=TRUE)
rcsumReqInfo3<-sum(dt$rcReqInfo3, na.rm=TRUE)

rtsumReqInfo1<-sum(dt$rtReqInfo1, na.rm=TRUE)
rtsumReqInfo2<-sum(dt$rtReqInfo2, na.rm=TRUE)
rtsumReqInfo3<-sum(dt$rtReqInfo3, na.rm=TRUE)
```

```
Number <- c("Other examples","Relative importance","Algorithm detail")
tc <- c(tcsumReqInfo1,tcsumReqInfo2,tcsumReqInfo3)
```

```

tt <- c(ttsumReqInfo1,ttsumReqInfo2,ttsumReqInfo3)
rc <- c(rcsumReqInfo1,rcsumReqInfo2,rcsumReqInfo3)
rt <- c(rtsumReqInfo1,rtsumReqInfo2,rtsumReqInfo3)
nyx <- data.frame(Number,tc,tt, rc, rt)

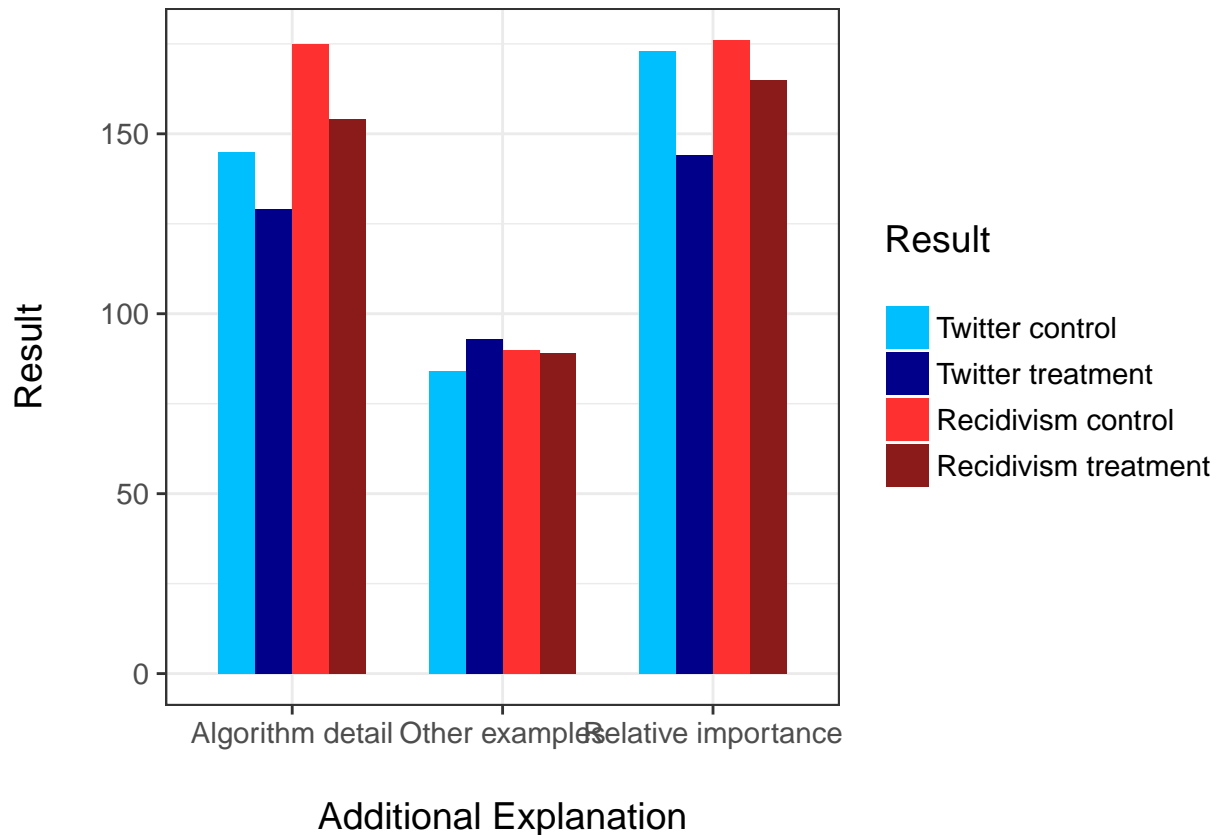
# load needed libraries
library(reshape2)

##
## Attaching package: 'reshape2'
## The following objects are masked from 'package:data.table':
##
##      dcast, melt
library(ggplot2)

# reshape your data into long format
nyxlong <- melt(nyx, id=c("Number"))

# make the plot
ggplot(nyxlong) +
  geom_bar(aes(x = Number, y = value, fill = variable),
    stat="identity", position = "dodge", width = 0.7) +
  scale_fill_manual("Result\n", values = c("deepskyblue","blue4", "firebrick1","firebrick4"),
    labels = c("Twitter control", "Twitter treatment","Recidivism control", "Recidivism
  labs(x="\nAdditional Explanation",y="Result\n") +
  theme_bw(base_size = 14)

```



Not to include in report, but text outputs of Other q 4

### Twitter control

again, whether there's any oversight into the decision or any appeals (though in this specific case it was clearly valid) Explanation of why they feel it's right to limit free speech. Freedom of speech infringement, you can not like the guy but twitter shouldn't silence his viewpoint, individuals should block him. When you are a platform for social interaction you shouldn't be allowed to restrict who can say what. None, I am against the restriction of the freedom of speech. none, moderation should be done by humans Statistics on pattern of behavior of banned individual, whether wrong people have ever been flagged What conduct rule in particular was deemed violated by the algorithm.

### Twitter treatment

A contextual analysis of the actual subject of the comment an explanation of the 70% threshold Definition of the characteristics Explanation of Sentiment Analysis Explanation on what is sentiment analysis and how it was judged I am perfectly satisfied with the explanation exactly as it is. I'm satisfied with the information. It should be stacked. All of the bars should add together. The way it's set up now, it could be at 69% threshold for all criteria and still would not have any action taken against it. more detailed explanation of the graph No other information required. nothing, it is ridiculous Proper spelling and explanation of sentiment decisions What "sentiment analysis" means. Is this thing reacting to everything it views as expressing a negative or argumentative attitude? when does using offensive vocabulary mean you are guilty, then pretty much all of us would be in jail and not got out of college Why Twitter feels the need to implement an algorithm like this at all.

## Recidivism control

A human isn't simple enough, there are going to be many factors that won't be considered. Again - ridiculous just as previous answer - doesn't take "person" into account, just numbers. An explanation of why and how they use an algorithm in court—and who allowed it. basically, it needs tons more information to make a decision like that none past success percentage. Statistics on accuracy the specific information the algorithm uses to make the assessment what happened to innocent until proven guilty? you can not make assessment of someone based on algorithm when they did not do anything wrong whether there's any human oversight into the decision

## Recidivism treatment

A possibility of a human psychologist or social worker weighing in on the algorithm results too. An explanation of how level of criminal personality was determined. biases of those who wrote the algorithm. Definitions of each. How it can assume that everyone is the same based on answers. human evaluation. I want to review the source code, the questions, the regression test results (when it was tested on repeat offenders). I would like to know what information the algorithm considered when making the decision. I'm already satisfied with the explanation. More detailed breakdown of what is meant in this context by phrases like "criminal personality". NONE. Numbers records showing other uses that turned out to be accurate and the percentage of accuracy overall.

## Demographic

```
#ageGroup', 'race', 'gender', 'socMed', 'educ', 'feedback', 'duration'
# par(mfrow=c(2,2))
# hist(dt$gender, main = "Gender")
# hist(dt$ageGroup, main = "Age")
# hist(dt$socMed, main= "Social Media Usage")
# hist(dt$edu, main = "Educational level")
#barplot(prop.table(table(dt$race)))
```

Interesting? Our MTers are mostly college students?

## Regression Models

### Twitter Moderation

Create linear models for each question for both Twitter and recidivism. The models subset the data to look only at those respondents that were assigned to either treatment or control for that context. In this way, someone who attrited in teh first context will not count against the second context.

```
mtFair <- ivreg(tFair ~ tTreat, data = dc[tAssign == 1])
mtAcc <- ivreg(tAcc ~ tTreat, data = dc[tAssign == 1])
mtSat <- ivreg(tSat ~ tTreat, data = dc[tAssign == 1])
mtUseful <- ivreg(tUseful ~ tTreat, data = dc[tAssign == 1])
mtClear <- ivreg(tClear ~ tTreat, data = dc[tAssign == 1])
mtMeaningful <- ivreg(tMeaningful ~ tTreat, data = dc[tAssign == 1])

stargazer(mtFair, mtAcc, mtSat, mtUseful, mtClear, mtMeaningful,
  type = 'text',
  covariate.labels = c("Explanation"),
  dep.var.labels = c("Fairness", "Accuracy", "Satisfaction", "Usefulness"),
```

```

    "Clarity", "Meaningfulness"),
  dep.var.caption = "Twitter Moderation")

```

```

##
## =====
##                                     Twitter Moderation
##                                     -----
##                               Fairness Accuracy Satisfaction Usefulness Clarity  Meaningfulness
##                               (1)      (2)      (3)      (4)      (5)      (6)
## -----
## Explanation                    0.242**   0.157*   0.367***   0.545***   0.332***   0.467***
##                               (0.096)   (0.087)   (0.091)   (0.094)   (0.093)   (0.096)
##
## Constant                      3.197***  3.371***  3.524***  2.965***  3.530***  2.886***
##                               (0.068)   (0.061)   (0.064)   (0.067)   (0.066)   (0.068)
##
## -----
## Observations                   627      627      627      627      627      627
## R2                            0.010     0.005     0.025     0.050     0.020     0.036
## Adjusted R2                   0.008     0.004     0.024     0.049     0.018     0.035
## Residual Std. Error (df = 625) 1.203     1.091     1.139     1.183     1.170     1.202
## =====
## Note:                                                                    *p<0.1; **p<0.05; ***p<0.01

```

```
dc[(tAssign == 1 & tSat == 0), .N]
```

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

```
dc[(tAssign == 1 & tMeaningful == 0), .N]
```

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

This shows that 1 person dropped out between seeing the treatment and responding in the Twitter context. This is probably not affecting our last few metrics, but they are all statistically significant by a large margin anyway. This represents our intent to treat effect.

However, they get through all of the first three questions, so those responses are not affected by attrition.

```

mrFair <- lm(rFair ~ rTreat, data = dc[rAssign == 1])
mrAcc <- lm(rAcc ~ rTreat, data = dc[rAssign == 1])
mrSat <- lm(rSat ~ rTreat, data = dc[rAssign == 1])
mrUseful <- lm(rUseful ~ rTreat, data = dc[rAssign == 1])
mrClear <- lm(rClear ~ rTreat, data = dc[rAssign == 1])
mrMeaningful <- lm(rMeaningful ~ rTreat, data = dc[rAssign == 1])

```

```

stargazer(mrFair, mrAcc, mrSat, mrUseful, mrClear, mrMeaningful,
  type = 'text',
  covariate.labels = c("Explanation"),
  dep.var.labels = c("Fairness", "Accuracy", "Satisfaction", "Usefulness",
    "Clarity", "Meaningfulness"),
  dep.var.caption = "Recidivism Risk Assessment")

```

```

##
## =====
##                                     Recidivism Risk Assessment
##                                     -----
##                               Fairness Accuracy Satisfaction Usefulness Clarity  Meaningfulness
##

```



```
##              (1)      (2)      (3)      (4)      (5)      (6)
## -----
## Explanation      0.726***  0.440***  0.649***  0.872***  0.906***  0.736***
##                  (0.088)  (0.082)  (0.099)  (0.095)  (0.103)  (0.095)
##
## Constant         2.423***  2.718***  2.750***  2.324***  2.705***  2.388***
##                  (0.062)  (0.058)  (0.070)  (0.068)  (0.073)  (0.067)
## -----
## Observations      628      628      628      628      628      628
## R2                0.098      0.044      0.064      0.118      0.109      0.088
## Adjusted R2       0.097      0.042      0.063      0.117      0.108      0.086
## Residual Std. Error (df = 626)  1.102      1.029      1.239      1.192      1.295      1.189
## F Statistic (df = 1; 626)      68.079***  28.723***  43.073***  84.045***  76.767***  60.139***
## =====
## Note:                                                     *p<0.1; **p<0.05; ***p<0.01
```

```
dc[(rAssign == 1 & rSat == 0), .N]
```

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

```
dc[(rAssign == 1 & rMeaningful == 0), .N]
```

```
## [1] 3
```

This shows that 3 people dropped out between seeing the treatment and responding in the recidivism context. This could be throwing off the last few metrics, but those are all statistically significant by a large margin. This represents our intent to treat effect.

## Difference in Order

We also discussed looking at the difference in responses depending on the order of contexts.

```
otFair <- lm(tFair ~ First.Context + tTreat + rTreat + tTreat*rTreat, data = dc[tAssign == 1])
otAcc <- lm(tAcc ~ First.Context+ tTreat + rTreat + tTreat*rTreat, data = dc[tAssign == 1])
otSat <- lm(tSat ~ First.Context+ tTreat + rTreat + tTreat*rTreat, data = dc[tAssign == 1])
otUseful <- lm(tUseful ~ First.Context+ tTreat + rTreat + tTreat*rTreat, data = dc[tAssign == 1])
otClear <- lm(tClear ~ First.Context+ tTreat + rTreat + tTreat*rTreat, data = dc[tAssign == 1])
otMeaningful <- lm(tMeaningful ~ First.Context+ tTreat + rTreat + tTreat*rTreat, data = dc[tAssign == 1])

library(stargazer)
stargazer(otFair, otAcc, otSat, otUseful, otClear, otMeaningful,
  type = 'text',
  dep.var.labels = c("Fairness", "Accuracy", "Satisfaction", "Usefulness",
    "Clarity", "Meaningfulness"),
  dep.var.caption = "Twitter Moderation")
```

```
##
## =====
##              Twitter Moderation
## -----
##              Fairness Accuracy Satisfaction Usefulness Clarity Meaningfulness
##              (1)      (2)      (3)      (4)      (5)      (6)
## -----
## First.ContextTwitter      0.102      0.153*      0.160*      0.012      0.018      -0.112
##                  (0.096)  (0.087)  (0.091)  (0.095)  (0.094)  (0.096)
```

```
##
## tTreat          0.268**   0.224*   0.517***   0.656***   0.448***   0.629***
##                (0.136)  (0.123)  (0.128)   (0.134)   (0.132)   (0.136)
##
## rTreat          -0.050   0.110   0.123   0.108   0.124   0.254*
##                (0.136)  (0.123)  (0.128)   (0.133)   (0.132)   (0.135)
##
## tTreat:rTreat   -0.050  -0.133  -0.298  -0.223  -0.231  -0.324*
##                (0.192)  (0.174)  (0.182)   (0.189)   (0.187)   (0.192)
##
## Constant        3.170*** 3.239*** 3.381*** 2.905*** 3.459*** 2.816***
##                (0.107)  (0.097)  (0.101)   (0.106)   (0.105)   (0.107)
##
## -----
## Observations    627     627     627     627     627     627
## R2              0.013   0.011   0.035   0.053   0.022   0.044
## Adjusted R2     0.007   0.005   0.028   0.047   0.016   0.038
## Residual Std. Error (df = 622) 1.204   1.090   1.136   1.184   1.171   1.200
## F Statistic (df = 4; 622)    2.040*   1.801   5.585*** 8.635*** 3.541*** 7.214***
## =====
## Note:                                                    *p<0.1; **p<0.05; ***p<0.01
```

In the case of Twitter moderation, there does not seem to be a difference based on order of context.

```
orFair <- lm(rFair ~ First.Context + rTreat + tTreat + rTreat*tTreat, data = dc[rAssign == 1])
orAcc <- lm(rAcc ~ First.Context + rTreat + tTreat + rTreat*tTreat, data = dc[rAssign == 1])
orSat <- lm(rSat ~ First.Context + rTreat + tTreat + rTreat*tTreat, data = dc[rAssign == 1])
orUseful <- lm(rUseful ~ First.Context + rTreat + tTreat + rTreat*tTreat, data = dc[rAssign == 1])
orClear <- lm(rClear ~ First.Context + rTreat + tTreat + rTreat*tTreat, data = dc[rAssign == 1])
orMeaningful <- lm(rMeaningful ~ First.Context + rTreat + tTreat + rTreat*tTreat, data = dc[rAssign == 1])

library(stargazer)
stargazer(orFair, orAcc, orSat, orUseful, orClear, orMeaningful,
  type = 'text',
  dep.var.labels = c("Fairness", "Accuracy", "Satisfaction", "Usefulness",
    "Clarity", "Meaningfulness"),
  dep.var.caption = "Recidivism Risk Assessment")
```

```
##
## =====
##                               Recidivism Risk Assessment
## -----
##      Fairness  Accuracy  Satisfaction  Usefulness  Clarity  Meaningfulness
##      (1)      (2)      (3)      (4)      (5)      (6)
## -----
## First.ContextTwitter  -0.063   -0.043   -0.140   -0.191**  -0.263**   -0.110
##                      (0.088)  (0.082)  (0.099)  (0.095)  (0.103)  (0.095)
##
## rTreat                0.608***  0.430***  0.658***  0.862***  0.885***  0.716***
##                      (0.124)  (0.116)  (0.139)  (0.134)  (0.145)  (0.134)
##
## tTreat                0.012   0.019   0.201   0.159   0.100   -0.020
##                      (0.125)  (0.117)  (0.140)  (0.135)  (0.146)  (0.135)
##
## rTreat:tTreat         0.236   0.020   -0.020   0.019   0.039   0.040
```

```
##              (0.176) (0.165) (0.197) (0.190) (0.206) (0.190)
##
## Constant      2.449*** 2.730*** 2.721*** 2.341*** 2.788*** 2.453***
##              (0.098) (0.092) (0.110) (0.106) (0.115) (0.106)
##
## -----
## Observations      628      628      628      628      628      628
## R2                0.105    0.045    0.073    0.128    0.120    0.090
## Adjusted R2       0.099    0.038    0.067    0.123    0.115    0.084
## Residual Std. Error (df = 623) 1.101    1.031    1.236    1.188    1.290    1.190
## F Statistic (df = 4; 623)    18.199*** 7.255*** 12.252*** 22.951*** 21.312*** 15.347***
## =====
## Note:                                     *p<0.1; **p<0.05; ***p<0.01
```

## Other Factors

```
otFair <- lm(tFair ~ tTreat + ageGroup + educ + socMed + black + asian + hispanic + other + pac_is)
otAcc <- lm(tAcc ~ tTreat + ageGroup + educ + socMed + black + asian + hispanic + other + pac_is)
otSat <- lm(tSat ~ tTreat + ageGroup + educ + socMed + black + asian + hispanic + other + pac_is)
otUseful <- lm(tUseful ~ tTreat + ageGroup + educ + socMed + black + asian + hispanic + other + pac_is)
otClear <- lm(tClear ~ tTreat + ageGroup + educ + socMed + black + asian + hispanic + other + pac_is)
otMeaningful <- lm(tMeaningful ~ tTreat + ageGroup + educ + socMed + black + asian + hispanic + other + pac_is)
```

```
library(stargazer)
stargazer(otFair, otAcc, otSat, otUseful, otClear, otMeaningful,
  type = 'text',
  covariate.labels = c("Explanation", "Age Group", "Education", "Social Media"),
  dep.var.labels = c("Fairness", "Accuracy", "Satisfaction", "Usefulness",
    "Clarity", "Meaningfulness"),
  dep.var.caption = "Twitter Moderation")
```

```
##
## =====
##
##              Twitter Moderation
##              -----
##              Fairness  Accuracy Satisfaction Usefulness Clarity  Meaningfulness
##              (1)      (2)      (3)      (4)      (5)      (6)
## -----
## Explanation      0.219** 0.148* 0.347*** 0.555*** 0.350*** 0.489***
##              (0.096) (0.088) (0.092) (0.094) (0.093) (0.096)
##
## Age Group        0.016    0.036    0.008    -0.025    0.004    0.025
##              (0.043) (0.039) (0.041) (0.042) (0.042) (0.043)
##
## Education        0.031    0.004    0.022    0.037    0.004    0.026
##              (0.037) (0.034) (0.036) (0.037) (0.036) (0.037)
##
## Social Media     0.022    0.041    -0.097*   -0.077   -0.142** -0.102*
##              (0.062) (0.056) (0.059) (0.061) (0.060) (0.061)
##
## black            -0.203   -0.174   -0.116   -0.236   -0.328   -0.182
##              (0.207) (0.188) (0.197) (0.203) (0.200) (0.205)
##
```

```
## asian                -0.130   -0.031    0.054    0.078    0.057    0.227*
##                    (0.127) (0.115) (0.121) (0.124) (0.123) (0.126)
##
## hispanic             -0.429*   -0.252   -0.135   -0.406*   -0.157   -0.548**
##                    (0.230) (0.209) (0.219) (0.226) (0.223) (0.228)
##
## other                -1.579***  -0.680   -0.459   -0.208   -0.464   -0.637
##                    (0.538) (0.489) (0.513) (0.528) (0.520) (0.534)
##
## pac_isle             0.714     0.591    1.093   -2.649**  -1.945*  -2.630**
##                    (1.199) (1.090) (1.143) (1.175) (1.159) (1.189)
##
## female               0.265***  0.269***  0.224**   0.117    0.188*   0.025
##                    (0.102) (0.093) (0.097) (0.100) (0.099) (0.101)
##
## gender_nc            0.600     0.596    0.154    0.405    0.117   -0.378
##                    (1.199) (1.090) (1.143) (1.175) (1.159) (1.189)
##
## Constant             2.666***  3.088***  3.250***  2.613***  3.602***  2.552***
##                    (0.539) (0.490) (0.514) (0.528) (0.521) (0.535)
##
## -----
## Observations          620      620      620      620      620      620
## R2                    0.044    0.030    0.042    0.077    0.048    0.073
## Adjusted R2           0.027    0.012    0.025    0.060    0.031    0.056
## Residual Std. Error (df = 608) 1.193    1.085    1.137    1.170    1.154    1.183
## F Statistic (df = 11; 608) 2.549***  1.699*   2.442***  4.603***  2.780***  4.347***
## =====
## Note:                                     *p<0.1; **p<0.05; ***p<0.01
```

Women were statistically significantly more likely than men to agree with the Twitter decision. Pacific Islanders did not like the explanation.

```
orFair <- lm(rFair ~ rTreat + ageGroup + educ + socMed + socMed^2 + black + asian + hispanic + other)
orAcc <-  lm(rAcc ~ rTreat + ageGroup + educ + socMed + socMed^2 + black + asian + hispanic + other)
orSat <-  lm(rSat ~ rTreat + ageGroup + educ + socMed + socMed^2 + black + asian + hispanic + other)
orUseful <- lm(rUseful ~ rTreat + ageGroup + educ + socMed + socMed^2 + black + asian + hispanic + other)
orClear <- lm(rClear ~ rTreat + ageGroup + educ + socMed + socMed^2 + black + asian + hispanic + other)
orMeaningful <- lm(rMeaningful ~ rTreat + ageGroup + educ + socMed + socMed^2 + black + asian + hispanic + other)

library(stargazer)
stargazer(otFair, otAcc, otSat, otUseful, otClear, otMeaningful,
  type = 'text',
  covariate.labels = c("Explanation", "Age Group", "Education", "Social Media", "Social Media2"),
  dep.var.labels = c("Fairness", "Accuracy", "Satisfaction", "Usefulness",
    "Clarity", "Meaningfulness"),
  dep.var.caption = "Recidivism Risk Assessment")
```

```
##
## =====
##                                     Recidivism Risk Assessment
##                                     -----
##                                     Fairness  Accuracy Satisfaction Usefulness Clarity  Meaningfulness
##                                     (1)      (2)      (3)      (4)      (5)      (6)
## -----
```

## Explanation	0.219**	0.148*	0.347***	0.555***	0.350***	0.489***
##	(0.096)	(0.088)	(0.092)	(0.094)	(0.093)	(0.096)
##						
## Age Group	0.016	0.036	0.008	-0.025	0.004	0.025
##	(0.043)	(0.039)	(0.041)	(0.042)	(0.042)	(0.043)
##						
## Education	0.031	0.004	0.022	0.037	0.004	0.026
##	(0.037)	(0.034)	(0.036)	(0.037)	(0.036)	(0.037)
##						
## Social Media	0.022	0.041	-0.097*	-0.077	-0.142**	-0.102*
##	(0.062)	(0.056)	(0.059)	(0.061)	(0.060)	(0.061)
##						
## Social Media2	-0.203	-0.174	-0.116	-0.236	-0.328	-0.182
##	(0.207)	(0.188)	(0.197)	(0.203)	(0.200)	(0.205)
##						
## asian	-0.130	-0.031	0.054	0.078	0.057	0.227*
##	(0.127)	(0.115)	(0.121)	(0.124)	(0.123)	(0.126)
##						
## hispanic	-0.429*	-0.252	-0.135	-0.406*	-0.157	-0.548**
##	(0.230)	(0.209)	(0.219)	(0.226)	(0.223)	(0.228)
##						
## other	-1.579***	-0.680	-0.459	-0.208	-0.464	-0.637
##	(0.538)	(0.489)	(0.513)	(0.528)	(0.520)	(0.534)
##						
## pac_isle	0.714	0.591	1.093	-2.649**	-1.945*	-2.630**
##	(1.199)	(1.090)	(1.143)	(1.175)	(1.159)	(1.189)
##						
## female	0.265***	0.269***	0.224**	0.117	0.188*	0.025
##	(0.102)	(0.093)	(0.097)	(0.100)	(0.099)	(0.101)
##						
## gender_nc	0.600	0.596	0.154	0.405	0.117	-0.378
##	(1.199)	(1.090)	(1.143)	(1.175)	(1.159)	(1.189)
##						
## Constant	2.666***	3.088***	3.250***	2.613***	3.602***	2.552***
##	(0.539)	(0.490)	(0.514)	(0.528)	(0.521)	(0.535)
##						

## -----						
## Observations	620	620	620	620	620	620
## R2	0.044	0.030	0.042	0.077	0.048	0.073
## Adjusted R2	0.027	0.012	0.025	0.060	0.031	0.056
## Residual Std. Error (df = 608)	1.193	1.085	1.137	1.170	1.154	1.183
## F Statistic (df = 11; 608)	2.549***	1.699*	2.442***	4.603***	2.780***	4.347***
## =====						
## Note:						*p<0.1; **p<0.05; ***p<0.01

Again, Pacific Islanders rated the explanation worse than others.

## Data Checks

```
d[is.na(random)]
```

##	ResponseID	ResponseSet	IPAddress
##	1: R_2v69taLV3v9B0JK	Default Response Set	141.211.120.79

```

## 2: R_1FOGRrdExvqmcfB Default Response Set 108.247.233.242
## 3: R_3n7z1iEjtdM6j4r Default Response Set 59.97.52.72
## 4: R_1mQbHGvzBLwjXT3 Default Response Set 73.178.168.46
## 5: R_2Cdm5f3dzUP0QCx Default Response Set 71.83.244.72
## 6: R_ZfPoGNloc70vd3X Default Response Set 117.216.30.200
## 7: R_zUXTmV9teb5ve0x Default Response Set 50.88.107.50
## 8: R_5iL3KkeDppmQH9n Default Response Set 68.101.202.229
## 9: R_1lyMJTGEG3A0pXI Default Response Set 182.65.174.207
## 10: R_2qxBAAwrhLPEnsW Default Response Set 73.188.87.107
## 11: R_z73qJd4VX50kZGN Default Response Set 14.195.229.229
## 12: R_RKASQRmCF05750B Default Response Set 74.109.193.29
## 13: R_2qrChwva1dZ019j Default Response Set 75.129.219.187
## 14: R_RhlTsv6w4kYfqxz Default Response Set 69.11.148.23
## 15: R_1PdZqzrj9445W2n Default Response Set 66.68.202.189
## 16: R_SOXL0vUKRIeYRMt Default Response Set 24.186.101.183
## 17: R_1rHbD3PtawzKeC2 Default Response Set 198.24.30.133
## 18: R_2UbwResocVbp6aA Default Response Set 157.50.15.145
## 19: R_ZJmNWyfdlFJsJd3 Default Response Set 162.95.216.224
##      StartDate      EndDate RecipientLastName
## 1: 2017-08-04 17:38:59 2017-08-04 17:39:06 NA
## 2: 2017-08-04 17:39:20 2017-08-04 17:39:25 NA
## 3: 2017-08-04 17:40:24 2017-08-04 17:40:45 NA
## 4: 2017-08-04 17:41:35 2017-08-04 17:42:54 NA
## 5: 2017-08-04 17:43:07 2017-08-04 17:43:17 NA
## 6: 2017-08-04 17:50:04 2017-08-04 17:50:51 NA
## 7: 2017-08-04 17:51:09 2017-08-04 17:52:43 NA
## 8: 2017-08-04 18:05:04 2017-08-04 18:06:09 NA
## 9: 2017-08-04 17:49:20 2017-08-04 18:09:16 NA
## 10: 2017-08-04 18:11:07 2017-08-04 18:11:32 NA
## 11: 2017-08-04 17:39:56 2017-08-04 18:13:05 NA
## 12: 2017-08-04 18:15:45 2017-08-04 18:16:41 NA
## 13: 2017-08-04 18:17:58 2017-08-04 18:18:16 NA
## 14: 2017-08-04 18:24:19 2017-08-04 18:24:45 NA
## 15: 2017-08-04 18:26:22 2017-08-04 18:26:37 NA
## 16: 2017-08-04 18:27:37 2017-08-04 18:28:15 NA
## 17: 2017-08-04 18:28:37 2017-08-04 18:29:04 NA
## 18: 2017-08-04 17:55:07 2017-08-04 18:37:43 NA
## 19: 2017-08-04 18:07:35 2017-08-04 18:38:05 NA
##      RecipientFirstName RecipientEmail ExternalDataReference Finished
## 1:      NA      NA      NA      0
## 2:      NA      NA      NA      0
## 3:      NA      NA      NA      0
## 4:      NA      NA      NA      0
## 5:      NA      NA      NA      0
## 6:      NA      NA      NA      0
## 7:      NA      NA      NA      0
## 8:      NA      NA      NA      0
## 9:      NA      NA      NA      0
## 10:     NA      NA      NA      0
## 11:     NA      NA      NA      0
## 12:     NA      NA      NA      0
## 13:     NA      NA      NA      0
## 14:     NA      NA      NA      0
## 15:     NA      NA      NA      0

```

## 16:	NA	NA	NA	0
## 17:	NA	NA	NA	0
## 18:	NA	NA	NA	0
## 19:	NA	NA	NA	0

##	Status	Q_TotalDuration	First.Context
## 1:	0	7	Recidivism
## 2:	0	4	Twitter
## 3:	0	20	Twitter
## 4:	0	78	Twitter
## 5:	0	9	Recidivism
## 6:	0	46	Recidivism
## 7:	0	93	Twitter
## 8:	0	64	Twitter
## 9:	0	1195	Recidivism
## 10:	0	24	Recidivism
## 11:	0	1989	Recidivism
## 12:	0	56	Recidivism
## 13:	0	18	Recidivism
## 14:	0	25	Recidivism
## 15:	0	15	Twitter
## 16:	0	37	Recidivism
## 17:	0	26	Twitter
## 18:	0	2555	Twitter
## 19:	0	1829	Recidivism

##	Enter.Embedded.Data.Field.Name.Here...	random	Q1.1	Q2.1	Q3.1	Q3.2	Q3.3
## 1:		NA	NA	1	NA	NA	NA
## 2:		NA	NA	1	NA	NA	NA
## 3:		NA	NA	1	1	NA	NA
## 4:		NA	NA	1	1	1	3
## 5:		NA	NA	1	NA	NA	NA
## 6:		NA	NA	1	NA	NA	NA
## 7:		NA	NA	1	1	1	5
## 8:		NA	NA	1	1	NA	NA
## 9:		NA	NA	1	NA	NA	NA
## 10:		NA	NA	1	1	NA	NA
## 11:		NA	NA	1	1	NA	NA
## 12:		NA	NA	1	NA	NA	NA
## 13:		NA	NA	1	NA	NA	NA
## 14:		NA	NA	1	NA	NA	NA
## 15:		NA	NA	1	1	NA	NA
## 16:		NA	NA	1	NA	NA	NA
## 17:		NA	NA	1	1	1	2
## 18:		NA	NA	1	1	NA	NA
## 19:		NA	NA	1	NA	NA	NA

##	Q3.4	Q3.5	Q3.6	Q3.7	Q3.8	Q3.9_1	Q3.9_2	Q3.9_3	Q3.9_4	Q3.9_4_TEXT	Q4.1
## 1:	NA	NA	NA	NA	NA	NA	NA	NA	NA		NA
## 2:	NA	NA	NA	NA	NA	NA	NA	NA	NA		NA
## 3:	NA	NA	NA	NA	NA	NA	NA	NA	NA		NA
## 4:	3	1	4	3	4	NA	NA	1	NA		NA
## 5:	NA	NA	NA	NA	NA	NA	NA	NA	NA		NA
## 6:	NA	NA	NA	NA	NA	NA	NA	NA	NA		NA
## 7:	5	1	5	5	5	NA	1	NA	NA		NA
## 8:	NA	NA	NA	NA	NA	NA	NA	NA	NA		1
## 9:	NA	NA	NA	NA	NA	NA	NA	NA	NA		NA

## 10:	NA	NA	NA	NA	NA	NA	NA	NA	NA	NA	NA	NA
## 11:	NA	NA	NA	NA	NA	NA	NA	NA	NA	NA	NA	NA
## 12:	NA	NA	NA	NA	NA	NA	NA	NA	NA	NA	NA	NA
## 13:	NA	NA	NA	NA	NA	NA	NA	NA	NA	NA	NA	NA
## 14:	NA	NA	NA	NA	NA	NA	NA	NA	NA	NA	NA	NA
## 15:	NA	NA	NA	NA	NA	NA	NA	NA	NA	NA	NA	NA
## 16:	NA	NA	NA	NA	NA	NA	NA	NA	NA	NA	NA	NA
## 17:	3	1	2	2	2	NA	NA	1	NA	NA	NA	NA
## 18:	NA	NA	NA	NA	NA	NA	NA	NA	NA	NA	1	NA
## 19:	NA	NA	NA	NA	NA	NA	NA	NA	NA	NA	NA	NA
##	Q4.2	Q4.3	Q4.4	Q4.5	Q4.6	Q4.7	Q4.8	Q4.9_1	Q4.9_2	Q4.9_3	Q4.9_4	
## 1:	NA	NA	NA	NA	NA	NA	NA	NA	NA	NA	NA	NA
## 2:	NA	NA	NA	NA	NA	NA	NA	NA	NA	NA	NA	NA
## 3:	NA	NA	NA	NA	NA	NA	NA	NA	NA	NA	NA	NA
## 4:	NA	NA	NA	NA	NA	NA	NA	NA	NA	NA	NA	NA
## 5:	NA	NA	NA	NA	NA	NA	NA	NA	NA	NA	NA	NA
## 6:	NA	NA	NA	NA	NA	NA	NA	NA	NA	NA	NA	NA
## 7:	NA	NA	NA	NA	NA	NA	NA	NA	NA	NA	NA	NA
## 8:	2	2	1	NA	NA	NA	NA	NA	NA	NA	NA	NA
## 9:	NA	NA	NA	NA	NA	NA	NA	NA	NA	NA	NA	NA
## 10:	NA	NA	NA	NA	NA	NA	NA	NA	NA	NA	NA	NA
## 11:	NA	NA	NA	NA	NA	NA	NA	NA	NA	NA	NA	NA
## 12:	NA	NA	NA	NA	NA	NA	NA	NA	NA	NA	NA	NA
## 13:	NA	NA	NA	NA	NA	NA	NA	NA	NA	NA	NA	NA
## 14:	NA	NA	NA	NA	NA	NA	NA	NA	NA	NA	NA	NA
## 15:	NA	NA	NA	NA	NA	NA	NA	NA	NA	NA	NA	NA
## 16:	NA	NA	NA	NA	NA	NA	NA	NA	NA	NA	NA	NA
## 17:	NA	NA	NA	NA	NA	NA	NA	NA	NA	NA	NA	NA
## 18:	2	2	2	1	4	3	4	NA	1	NA	NA	NA
## 19:	NA	NA	NA	NA	NA	NA	NA	NA	NA	NA	NA	NA
##	Q4.9_4_TEXT	Q8.1	Q5.1	Q6.1	Q6.2	Q6.3	Q6.4	Q6.5	Q6.6	Q6.7	Q6.8	Q6.9_1
## 1:		NA	1	NA	NA	NA	NA	NA	NA	NA	NA	NA
## 2:		NA	NA	NA	NA	NA	NA	NA	NA	NA	NA	NA
## 3:		NA	NA	NA	NA	NA	NA	NA	NA	NA	NA	NA
## 4:		1	NA	NA	NA	NA	NA	NA	NA	NA	NA	NA
## 5:		NA	1	NA	NA	NA	NA	NA	NA	NA	NA	NA
## 6:		NA	1	NA	NA	NA	NA	NA	NA	NA	NA	NA
## 7:		1	1	NA	NA	NA	NA	NA	NA	NA	NA	NA
## 8:		NA	NA	NA	NA	NA	NA	NA	NA	NA	NA	NA
## 9:		NA	1	NA	NA	NA	NA	NA	NA	NA	NA	NA
## 10:		1	1	NA	NA	NA	NA	NA	NA	NA	NA	NA
## 11:		1	1	1	4	3	1	1	2	2	2	NA
## 12:		NA	1	NA	NA	NA	NA	NA	NA	NA	NA	NA
## 13:		NA	1	NA	NA	NA	NA	NA	NA	NA	NA	NA
## 14:		NA	1	NA	NA	NA	NA	NA	NA	NA	NA	NA
## 15:		NA	NA	NA	NA	NA	NA	NA	NA	NA	NA	NA
## 16:		NA	1	1	5	1	1	NA	NA	NA	NA	NA
## 17:		1	1	NA	NA	NA	NA	NA	NA	NA	NA	NA
## 18:		1	1	1	3	2	2	NA	NA	NA	NA	NA
## 19:		NA	1	NA	NA	NA	NA	NA	NA	NA	NA	NA
##	Q6.9_2	Q6.9_3	Q6.9_4	Q6.9_4_TEXT	Q7.1	Q7.2	Q7.3	Q7.4	Q7.5	Q7.6	Q7.7	
## 1:	NA	NA	NA		NA	NA	NA	NA	NA	NA	NA	NA
## 2:	NA	NA	NA		NA	NA	NA	NA	NA	NA	NA	NA
## 3:	NA	NA	NA		NA	NA	NA	NA	NA	NA	NA	NA



## 4:	NA	NA	NA		NA	NA	NA	NA	NA	NA	NA
## 5:	NA	NA	NA		NA	NA	NA	NA	NA	NA	NA
## 6:	NA	NA	NA		NA	NA	NA	NA	NA	NA	NA
## 7:	NA	NA	NA		NA	NA	NA	NA	NA	NA	NA
## 8:	NA	NA	NA		NA	NA	NA	NA	NA	NA	NA
## 9:	NA	NA	NA		1	27	3	3	NA	NA	NA
## 10:	NA	NA	NA		1	28	4	4	1	2	2
## 11:	1	NA	NA		NA	NA	NA	NA	NA	NA	NA
## 12:	NA	NA	NA		NA	NA	NA	NA	NA	NA	NA
## 13:	NA	NA	NA		NA	NA	NA	NA	NA	NA	NA
## 14:	NA	NA	NA		NA	NA	NA	NA	NA	NA	NA
## 15:	NA	NA	NA		NA	NA	NA	NA	NA	NA	NA
## 16:	NA	NA	NA		NA	NA	NA	NA	NA	NA	NA
## 17:	NA	NA	NA		1	26	3	3	1	3	2
## 18:	NA	NA	NA		NA	NA	NA	NA	NA	NA	NA
## 19:	NA	NA	NA		NA	NA	NA	NA	NA	NA	NA
##	Q7.8	Q7.9_1	Q7.9_2	Q7.9_3	Q7.9_4	Q7.9_4_TEXT	Q9.1	Q9.2	Q9.3_1	Q9.3_2	
## 1:	NA	NA	NA	NA	NA		NA	NA	NA	NA	
## 2:	NA	NA	NA	NA	NA		NA	NA	NA	NA	
## 3:	NA	NA	NA	NA	NA		NA	NA	NA	NA	
## 4:	NA	NA	NA	NA	NA		NA	NA	NA	NA	
## 5:	NA	NA	NA	NA	NA		NA	NA	NA	NA	
## 6:	NA	NA	NA	NA	NA		NA	NA	NA	NA	
## 7:	NA	NA	NA	NA	NA		NA	NA	NA	NA	
## 8:	NA	NA	NA	NA	NA		NA	NA	NA	NA	
## 9:	NA	NA	NA	NA	NA		NA	NA	NA	NA	
## 10:	3	1	NA	1	NA		NA	NA	NA	NA	
## 11:	NA	NA	NA	NA	NA		NA	NA	NA	NA	
## 12:	NA	NA	NA	NA	NA		NA	NA	NA	NA	
## 13:	NA	NA	NA	NA	NA		NA	NA	NA	NA	
## 14:	NA	NA	NA	NA	NA		NA	NA	NA	NA	
## 15:	NA	NA	NA	NA	NA		NA	NA	NA	NA	
## 16:	NA	NA	NA	NA	NA		NA	NA	NA	NA	
## 17:	2	NA	NA	1	NA		NA	NA	NA	NA	
## 18:	NA	NA	NA	NA	NA		NA	NA	NA	NA	
## 19:	NA	NA	NA	NA	NA		NA	NA	NA	NA	
##	Q9.3_3	Q9.3_4	Q9.3_5	Q9.3_6	Q9.3_7	Q9.4	Q9.5	Q9.6	Q10.1	Q10.2	Q10.3
## 1:	NA	NA	NA	NA	NA	NA	NA	NA	NA		NA
## 2:	NA	NA	NA	NA	NA	NA	NA	NA	NA		NA
## 3:	NA	NA	NA	NA	NA	NA	NA	NA	NA		NA
## 4:	NA	NA	NA	NA	NA	NA	NA	NA	NA		NA
## 5:	NA	NA	NA	NA	NA	NA	NA	NA	NA		NA
## 6:	NA	NA	NA	NA	NA	NA	NA	NA	NA		NA
## 7:	NA	NA	NA	NA	NA	NA	NA	NA	NA		NA
## 8:	NA	NA	NA	NA	NA	NA	NA	NA	NA		NA
## 9:	NA	NA	NA	NA	NA	NA	NA	NA	NA		NA
## 10:	NA	NA	NA	NA	NA	NA	NA	NA	NA		NA
## 11:	NA	NA	NA	NA	NA	NA	NA	NA	NA		NA
## 12:	NA	NA	NA	NA	NA	NA	NA	NA	NA		NA
## 13:	NA	NA	NA	NA	NA	NA	NA	NA	NA		NA
## 14:	NA	NA	NA	NA	NA	NA	NA	NA	NA		NA
## 15:	NA	NA	NA	NA	NA	NA	NA	NA	NA		NA
## 16:	NA	NA	NA	NA	NA	NA	NA	NA	NA		NA
## 17:	NA	NA	NA	NA	NA	NA	NA	NA	NA		NA

## 18:	NA	NA	NA	NA	NA	NA	NA	NA	NA	NA
## 19:	NA	NA	NA	NA	NA	NA	NA	NA	NA	NA
##	Location	Latitude	Location	Longitude	Location	Accuracy				
## 1:		NA		NA		-1				
## 2:		NA		NA		-1				
## 3:		NA		NA		-1				
## 4:		NA		NA		-1				
## 5:		NA		NA		-1				
## 6:		NA		NA		-1				
## 7:		NA		NA		-1				
## 8:		NA		NA		-1				
## 9:		NA		NA		-1				
## 10:		NA		NA		-1				
## 11:		NA		NA		-1				
## 12:		NA		NA		-1				
## 13:		NA		NA		-1				
## 14:		NA		NA		-1				
## 15:		NA		NA		-1				
## 16:		NA		NA		-1				
## 17:		NA		NA		-1				
## 18:		NA		NA		-1				
## 19:		NA		NA		-1				