Music Lyrics Pilot Data

Gurdit Chahal, Adam Letcher, Carlos Castro, Noah Randolph

8/4/2018

Results from Pilot Testing of Experiment Site

```
dt_raw <- fread('pilot_results.csv')
```

Cleanup Columns

```
dt <- dt_raw[, .(hear_song = as.integer(as.character(factor(Check1, levels = c('A', 'B'),
                                                            labels = c(1, 0)))),
                 piano_playing = as.integer(as.character(factor(Check2,
                                                                levels = c('A', 'B'),
                                                                labels = c(1, 0)))),
                 q1 = as.factor(Q1),
                 q2 = as.factor(Q2),
                 q3 = as.factor(Q3),
                 q4 = as.factor(Q4),
                 q5 = as.factor(Q5),
                 age = factor(Q6, levels = c('A', 'B', 'C', 'D', 'E',
                                             'F', 'G', 'H', 'I', 'J'),
                              labels = c('<12', '12-17', '18-24', '25-34', '35-44',
                                         '45-54', '55-64', '65-74', '>75', 'decline')),
                 gender = factor(Q7, levels = c('A', 'B', 'C', 'D'),
                                 labels = c('male', 'female', 'other', 'decline')),
                 own_dog = as.integer(as.character(factor(Q8, levels = c('A', 'B'),
                                                          labels = c(1, 0)))),
                 education = factor(Q9, levels = c('A', 'B', 'C', 'D', 'E',
                                                   'F', 'G', 'H', 'I', 'J'),
                                    labels = c('none', '8th grade', 'some high school',
                                               'high school completed', 'some college',
                                               'vocational', 'associates', 'bachelors',
                                               'masters', 'doctorate')),
                 occupation = Q10,
                 native_english = as.integer(as.character(factor(Q11,
                                                                 levels = c('A', 'B'),
                                                                 labels = c(1, 0)))),
                 heard_lyrics = as.integer(as.character(factor(Q12, levels = c('A', 'B'),
                                                               labels = c(1, 0)))),
                 is_turk = as.integer(as.character(factor(isTurk,
                                                          levels = c('true', 'null'),
                                                          labels = c(1, 0)))),
                 time,
                 correct_count = correctCount,
                 assigned_lyrics = as.integer(
                   as.character(factor(lyrics, levels = c('true', 'false'),
                                       labels = c(1, 0))))]
```

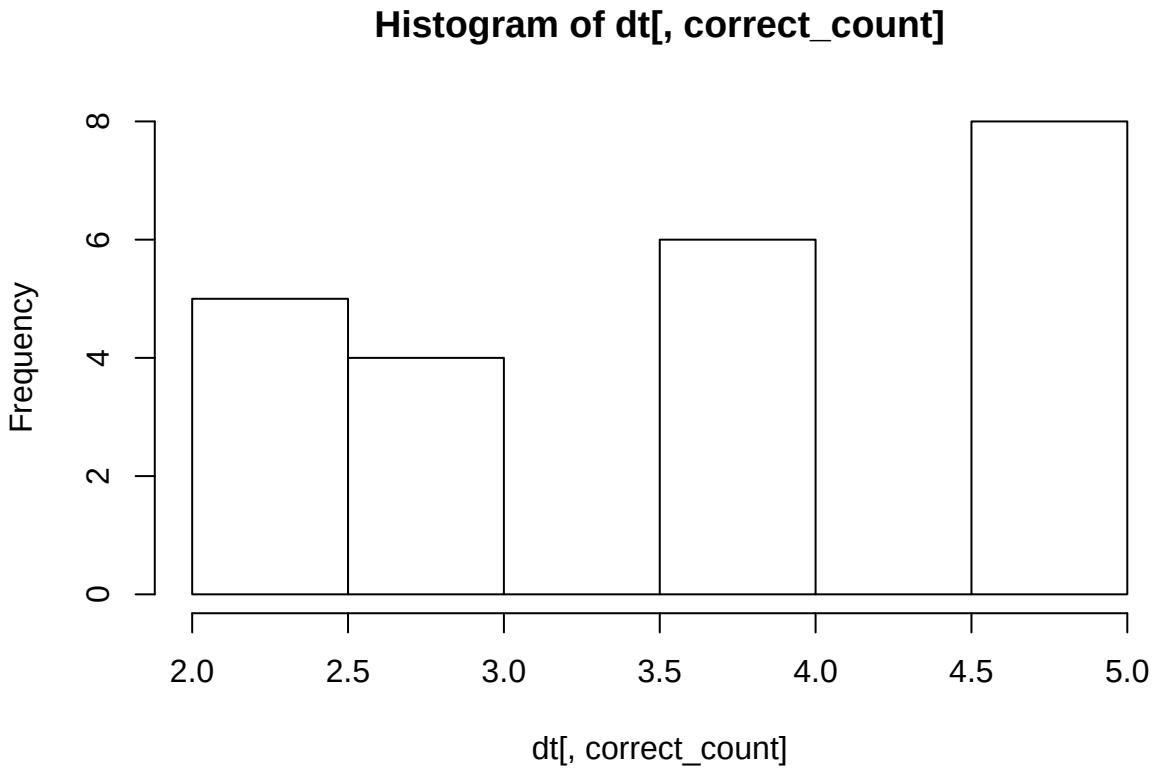
EDA

```
stargazer(dt, header=FALSE, type='latex')
```

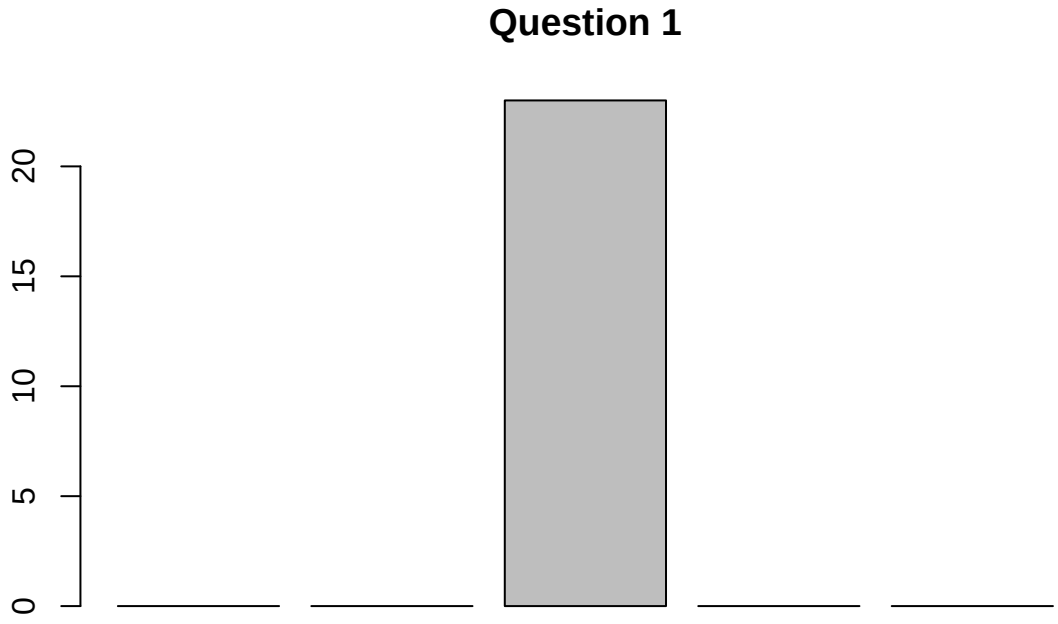
Table 1:

Statistic	N	Mean	St. Dev.	Min	Max
hear_song	23	1.000	0.000	1	1
piano_playing	23	1.000	0.000	1	1
own_dog	23	0.304	0.470	0	1
native_english	22	0.727	0.456	0	1
heard_lyrics	22	0.545	0.510	0	1
is_turk	23	0.217	0.422	0	1
time	23	−590.696	310.507	−1,478	−182
correct_count	23	3.739	1.176	2	5
assigned_lyrics	23	0.435	0.507	0	1

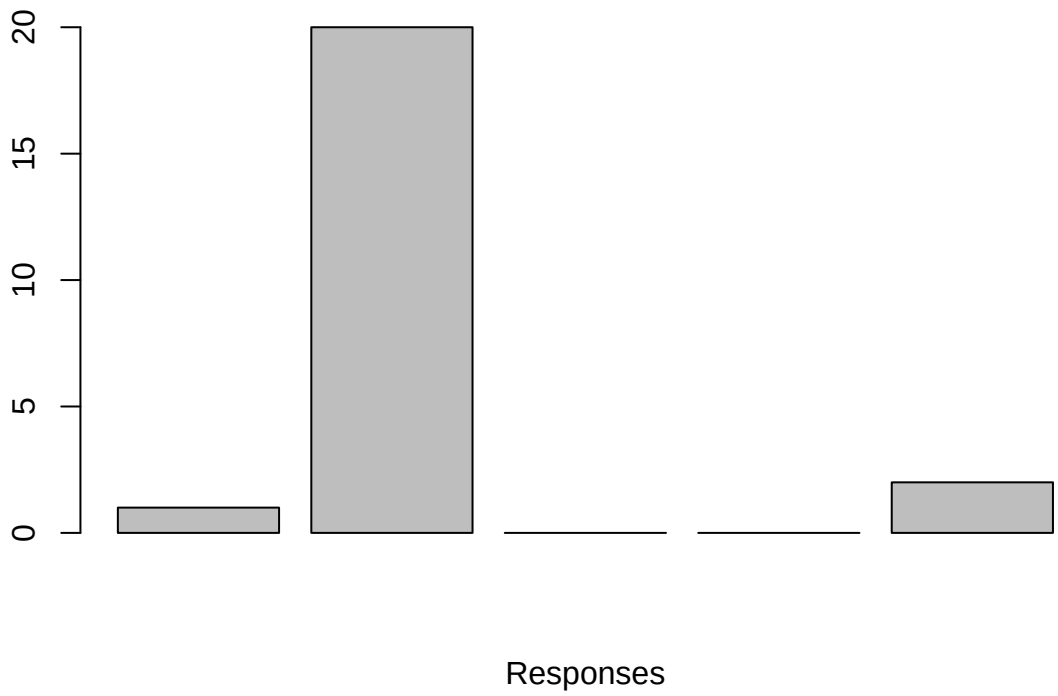
```
hist(dt[, correct_count])
```



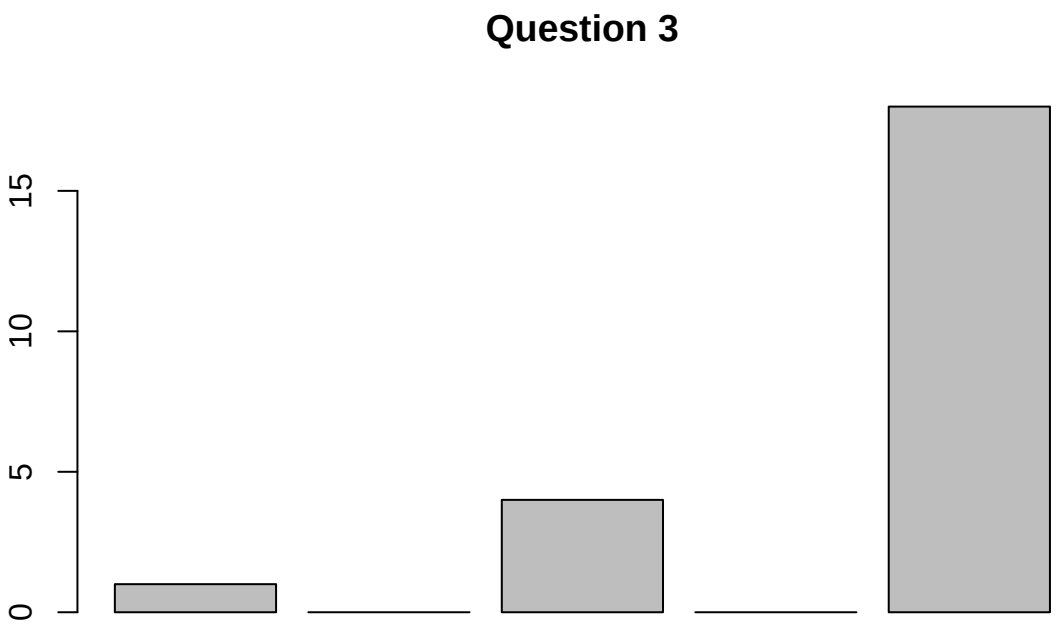

```
q1_counts <- c(sum(dt[,q1=='A']), sum(dt[,q1=='B']), sum(dt[,q1=='C']),
               sum(dt[,q1=='D']), sum(dt[,q1=='E']))
barplot(q1_counts, main="Question 1",
    xlab="Responses")
```



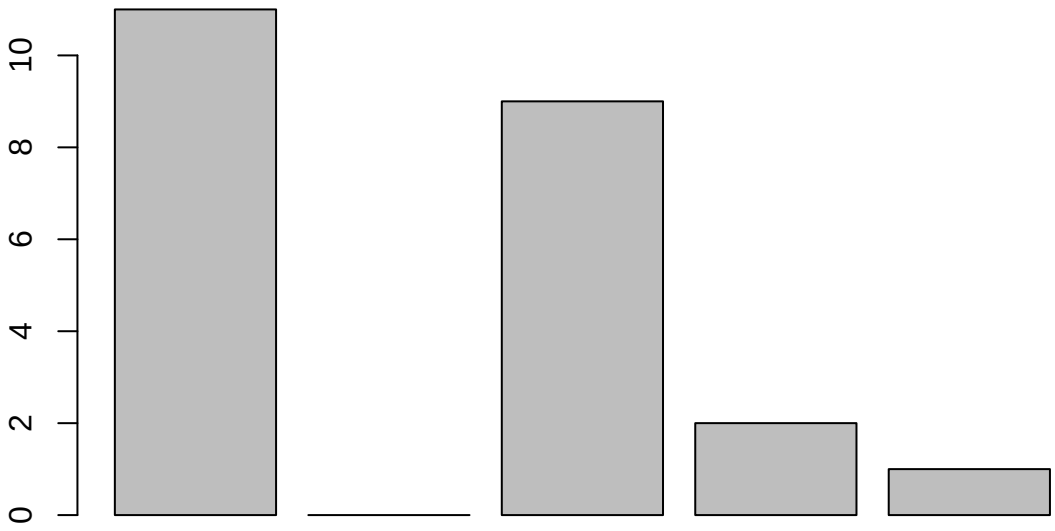

```
q2_counts <- c(sum(dt[,q2=='A']), sum(dt[,q2=='B']), sum(dt[,q2=='C']),
               sum(dt[,q2=='D']), sum(dt[,q2=='E']))
barplot(q2_counts, main="Question 2",
    xlab="Responses")
```



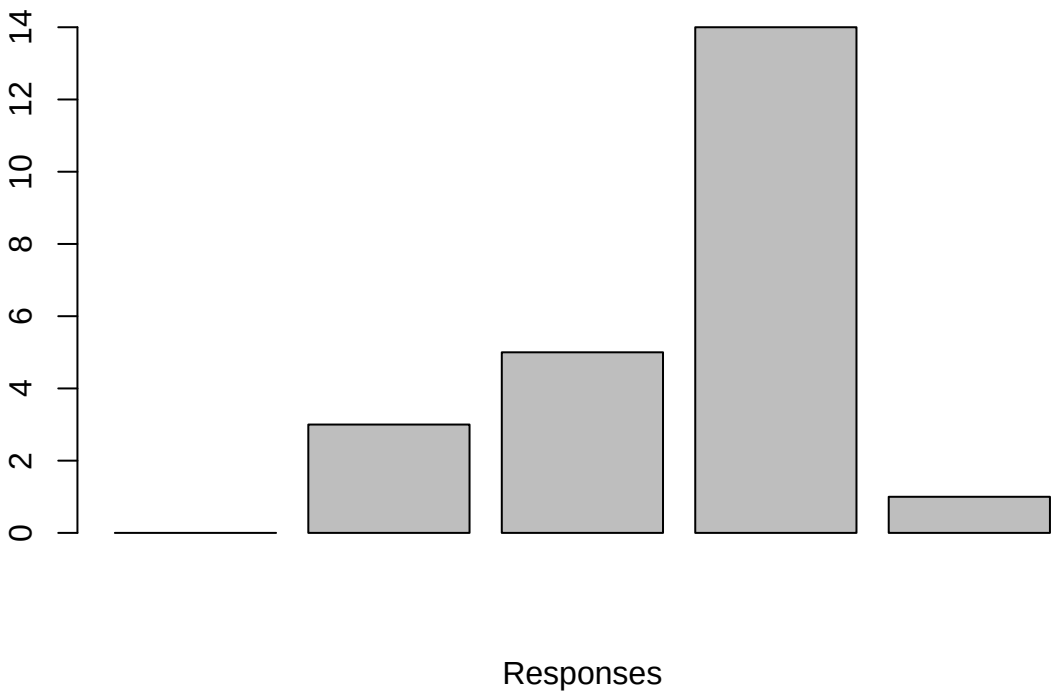
```
q3_counts <- c(sum(dt[,q3=='A']), sum(dt[,q3=='B']), sum(dt[,q3=='C']),
               sum(dt[,q3=='D']), sum(dt[,q3=='E']))
barplot(q3_counts, main="Question 3",
    xlab="Responses")
```




```
q4_counts <- c(sum(dt[,q4=='A']), sum(dt[,q4=='B']), sum(dt[,q4=='C']),
               sum(dt[,q4=='D']), sum(dt[,q4=='E']))
barplot(q4_counts, main="Question 4",
    xlab="Responses")
```




```
q5_counts <- c(sum(dt[,q5=='A']), sum(dt[,q5=='B']), sum(dt[,q5=='C']),
               sum(dt[,q5=='D']), sum(dt[,q5=='E']))
barplot(q5_counts, main="Question 5",
    xlab="Responses")
```



Regression

```
fit_pilot <- lm(correct_count ~ assigned_lyrics, dt)
summary(fit_pilot)
```

```
##
## Call:
## lm(formula = correct_count ~ assigned_lyrics, data = dt)
##
## Residuals:
##     Min      1Q  Median      3Q     Max
## -1.9000 -0.9000  0.3846  1.1000  1.3846
##
## Coefficients:
##                 Estimate Std. Error t value Pr(>|t|)
## (Intercept)       3.6154     0.3314  10.910 4.12e-10 ***
## assigned_lyrics   0.2846     0.5025   0.566    0.577
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Residual standard error: 1.195 on 21 degrees of freedom
## Multiple R-squared:  0.01504,    Adjusted R-squared:  -0.03186
## F-statistic: 0.3207 on 1 and 21 DF,  p-value: 0.5772
```

```
fit_pilot_with_covariates <- lm(correct_count ~ assigned_lyrics + age + education + is_turk + time, dt)
summary(fit_pilot_with_covariates)
```

```
##
## Call:
## lm(formula = correct_count ~ assigned_lyrics + age + education +
##     is_turk + time, data = dt)
##
## Residuals:
##     Min      1Q  Median      3Q     Max
## -1.8876 -0.5862  0.0000  0.7499  1.1607
##
## Coefficients: (1 not defined because of singularities)
##                       Estimate Std. Error t value Pr(>|t|)
## (Intercept)           1.634767   1.809226   0.904    0.386
## assigned_lyrics      -0.140392   0.653337  -0.215    0.834
## age25-34             -0.120515   0.789640  -0.153    0.881
## age35-44             -1.328552   1.224127  -1.085    0.301
## age45-54             -3.320230   2.015538  -1.647    0.128
## age55-64              0.998057   2.079189   0.480    0.641
## age65-74             -1.341487   1.144418  -1.172    0.266
## educationassociates   1.301135   1.759127   0.740    0.475
## educationbachelors    1.880305   1.585688   1.186    0.261
## educationmasters      2.650071   1.711015   1.549    0.150
## educationdoctorate          NA         NA      NA       NA
## is_turk               0.205208   1.223104   0.168    0.870
## time                 -0.001077   0.001221  -0.882    0.397
##
## Residual standard error: 1.234 on 11 degrees of freedom
## Multiple R-squared:  0.4496, Adjusted R-squared:  -0.1008
```

```
## F-statistic: 0.8168 on 11 and 11 DF,  p-value: 0.6285
```