babble_analysis_raw_output

Read data

```
data_6 <- read.csv("./20190708-141738_the-maturity-of-baby-sounds-part-6_answers.csv", sep=';')
data_7_1 <- read.csv("./20190708-145538_the-maturity-of-baby-sounds-part-7_answers.csv", sep=';')
data_7_2 <- read.csv("./20190710-165105_the-maturity-of-baby-sounds-part-7_answers.csv", sep=';')

summary(data_7_1)
```

```
##     User.ID                                  Dataset.Name     Dataset.ID
##  Min.   : 93.0   the-maturity-of-baby-sounds-part-7:46518   Min.   :150
##  1st Qu.:726.0                                             1st Qu.:150
##  Median :729.0                                             Median :150
##  Mean   :709.5                                             Mean   :150
##  3rd Qu.:731.0                                             3rd Qu.:150
##  Max.   :733.0                                             Max.   :150
##
##                          Task.Name        Task.ID              Media.File
##  please-classify-this-sound-7:46518   Min.   :225   0101976887.wav:    8
##                                       1st Qu.:225   0238973764.wav:    8
##                                       Median :225   0316192601.wav:    8
##                                       Mean   :225   0336444621.wav:    8
##                                       3rd Qu.:225   0436113999.wav:    8
##                                       Max.   :225   0531354289.wav:    8
##                                                     (Other)       :46470
##  Media.Duration             Answer                          Timestamp
##  Min.   :0.0500                :   13   2019-05-07 07:44:41+00:00:    4
##  1st Qu.:0.4000   Canonical    : 1538   2019-05-07 07:57:33+00:00:    4
##  Median :0.4000   Crying       :  594   2019-05-07 09:38:13+00:00:    4
##  Mean   :0.3708   Junk         :19352   2019-05-08 09:48:06+00:00:    4
##  3rd Qu.:0.4000   Laughing     :  360   2019-05-08 09:53:53+00:00:    4
##  Max.   :0.5000   Non-canonical:24661   2019-05-08 09:56:22+00:00:    4
##                                         (Other)                  :46494
##  Decision.Time      Playback.Count   Bad.File.
##  Min.   :       0   Min.   : 0.000   Mode :logical
##  1st Qu.:     754   1st Qu.: 1.000   FALSE:46503
##  Median :    1087   Median : 1.000   TRUE :15
##  Mean   :    7136   Mean   : 1.586
##  3rd Qu.:    1941   3rd Qu.: 2.000
##  Max.   :15468262   Max.   :32.000
##
```

```
data_7_1 <- data_7_1 %>%
  mutate("Bad.File." = as.factor("Bad.File."))
summary(data_7_2)
```

```
##     User.ID                                  Dataset.Name    Dataset.ID
##  Min.   :482.0   the-maturity-of-baby-sounds-part-7:3971   Min.   :150
##  1st Qu.:743.0                                            1st Qu.:150
##  Median :743.0                                            Median :150
##  Mean   :707.9                                            Mean   :150
```

```
##  3rd Qu.:746.0                                             3rd Qu.:150
##  Max.   :748.0                                             Max.   :150
##
##                       Task.Name       Task.ID              Media.File
##  please-classify-this-sound-7:3971   Min.   :225   0005427940.wav:   5
##                                      1st Qu.:225   0006968363.wav:   5
##                                      Median :225   0015242083.wav:   5
##                                      Mean   :225   0018504442.wav:   5
##                                      3rd Qu.:225   0028480000.wav:   5
##                                      Max.   :225   0037181090.wav:   5
##                                                    (Other)       :3941
##  Media.Duration             Answer                        Timestamp
##  Min.   :0.0700   Canonical    : 245   2019-07-09 15:35:24+00:00:   3
##  1st Qu.:0.4000   Crying       :  62   2019-07-09 16:24:59+00:00:   3
##  Median :0.4000   Junk         :1598   2019-07-09 16:25:03+00:00:   3
##  Mean   :0.3668   Laughing     :  42   2019-07-09 16:55:35+00:00:   3
##  3rd Qu.:0.4000   Non-canonical:2024   2019-07-09 16:55:52+00:00:   3
##  Max.   :0.5000                        2019-07-09 16:56:30+00:00:   3
##                                        (Other)                  :3953
##  Decision.Time     Playback.Count   Bad.File.
##  Min.   :    350   Min.   : 1.000   False:3969
##  1st Qu.:   1272   1st Qu.: 1.000   True :   2
##  Median :   1809   Median : 1.000
##  Mean   :  14061   Mean   : 1.805
##  3rd Qu.:   3102   3rd Qu.: 2.000
##  Max.   :4505917   Max.   :36.000
##
```

```
data_7 <- bind_rows(data_7_1, data_7_2)
```

```
## Warning in bind_rows_(x, .id): Unequal factor levels: coercing to character

## Warning in bind_rows_(x, .id): binding character and factor vector,
## coercing into character vector

## Warning in bind_rows_(x, .id): binding character and factor vector,
## coercing into character vector

## Warning in bind_rows_(x, .id): Unequal factor levels: coercing to character

## Warning in bind_rows_(x, .id): binding character and factor vector,
## coercing into character vector

## Warning in bind_rows_(x, .id): binding character and factor vector,
## coercing into character vector

## Warning in bind_rows_(x, .id): Unequal factor levels: coercing to character

## Warning in bind_rows_(x, .id): binding character and factor vector,
## coercing into character vector

## Warning in bind_rows_(x, .id): binding character and factor vector,
## coercing into character vector

## Warning in bind_rows_(x, .id): Unequal factor levels: coercing to character

## Warning in bind_rows_(x, .id): binding character and factor vector,
## coercing into character vector
```

```
## Warning in bind_rows_(x, .id): binding character and factor vector,
## coercing into character vector
```

Analysis

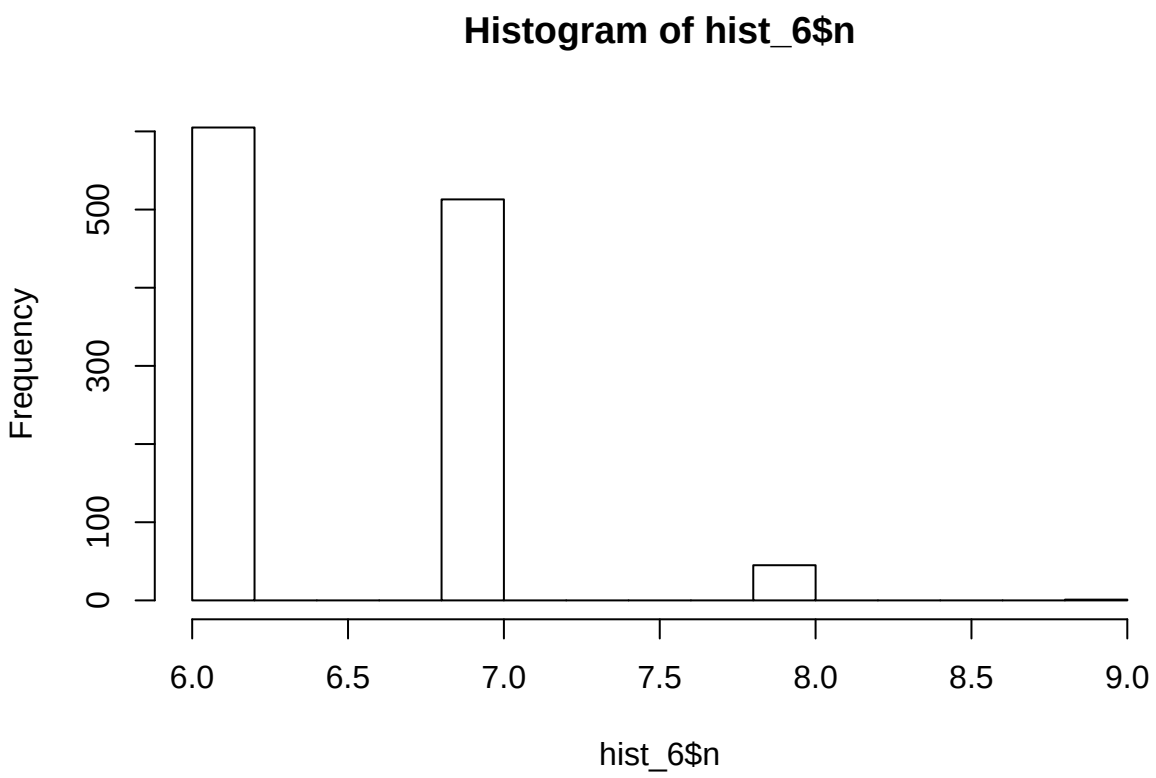
```
hist_6 <- data_6 %>%
  group_by(Media.File) %>%
  count()

hist_6 %>%
  group_by(n) %>%
  tally()
```

```
## Using `n` as weighting variable
```

```
## # A tibble: 4 x 2
##       n    nn
##   <int> <int>
## 1     6  3630
## 2     7  3591
## 3     8   360
## 4     9     9
```

```
# Histogram of number of annotations
hist(hist_6$n)
```




```
hist_7 <- data_7 %>%
  group_by(Media.File) %>%
  count()
```

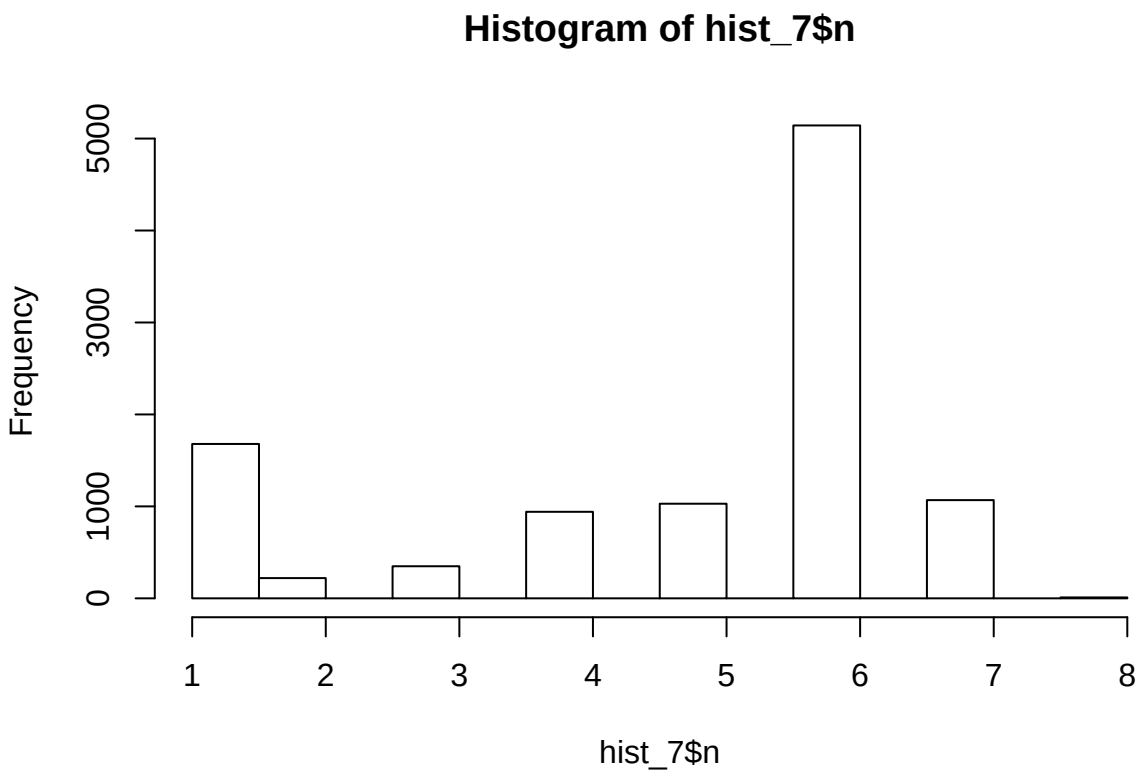
```
todo_7 <- hist_7 %>%
  filter(n <= 2)

hist_7 %>%
  group_by(n) %>%
  tally()
```

```
## Using `n` as weighting variable
```

```
## # A tibble: 8 x 2
##       n    nn
##   <int> <int>
## 1     1  1679
## 2     2   440
## 3     3  1047
## 4     4  3764
## 5     5  5145
## 6     6 30858
## 7     7  7476
## 8     8    80
```

```
hist(hist_7$n)
```




```
nrow(hist_7)
```

```
## [1] 10439
```