

Exploratory Data Analysis for AVA Dataset

Jens Laufer

9 Juli 2018

Overview

Check for missing files

There are 255508 files, but 255530 images in dataset.

Date entries without image file were deleted.

Summary

The ordataset has 255530 rows and 14 columns.

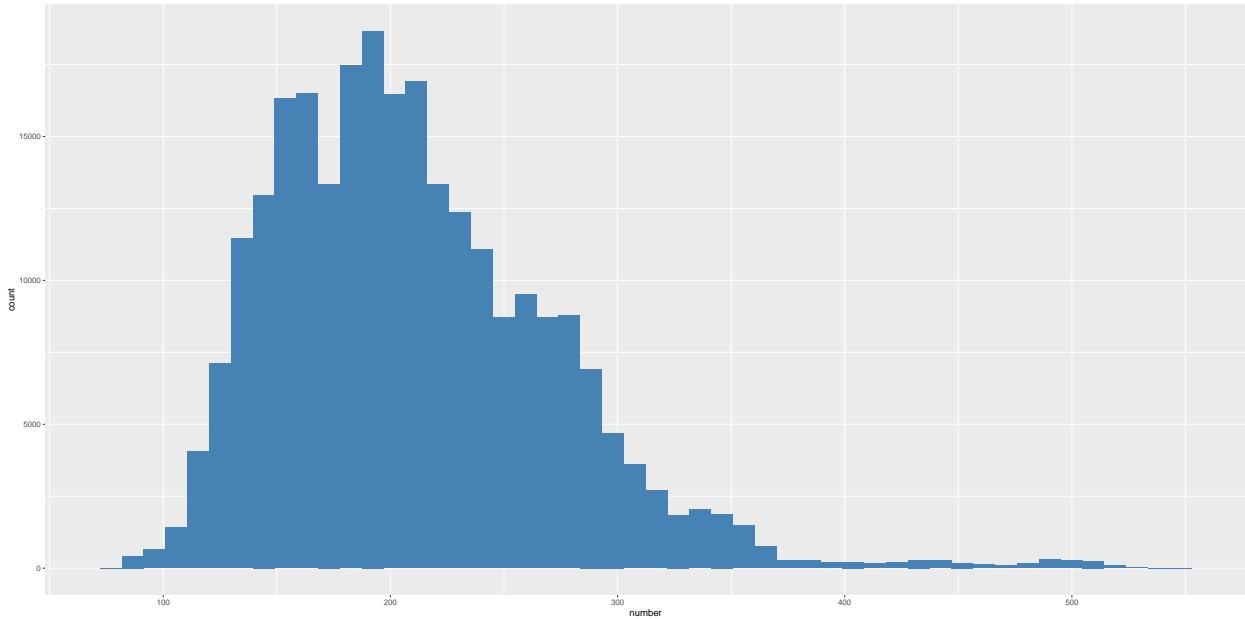
Columns: image.id, 1, 2, 3, 4, 5, 6, 7, 8, 9, 10, semantic.tag.id1, semantic.tag.id2, challenge.id

Sample rows

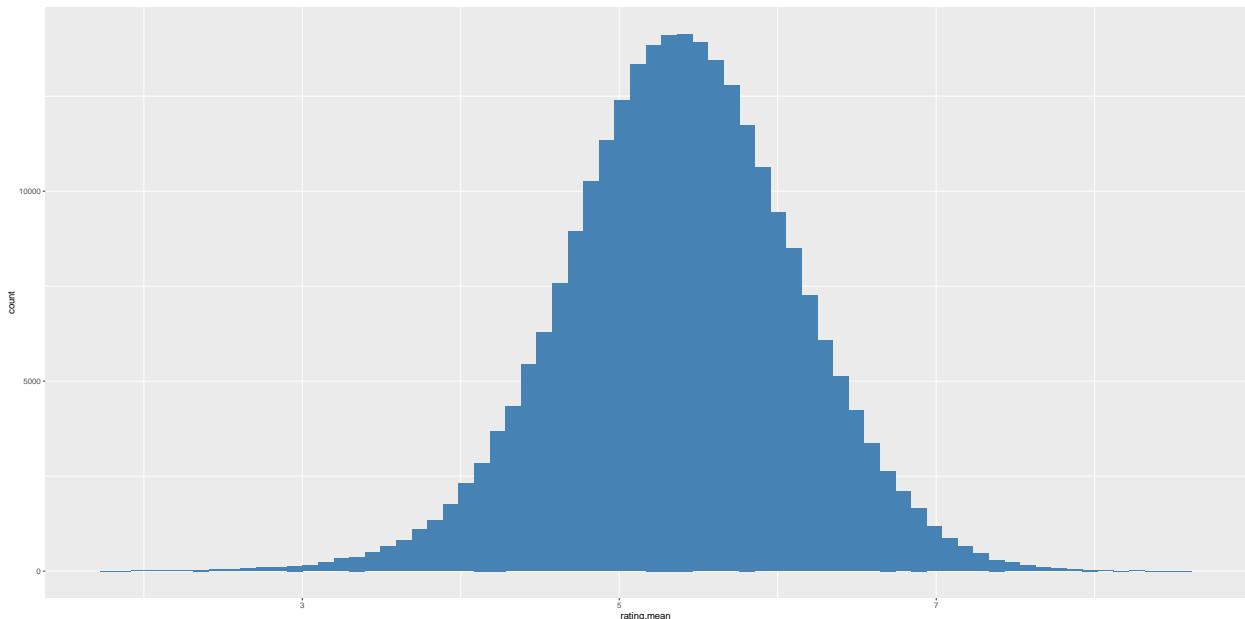
```
## # A tibble: 6 x 14
##   image.id `1`   `2`   `3`   `4`   `5`   `6`   `7`   `8`   `9`   `10`
##   <int> <int> <int> <int> <int> <int> <int> <int> <int> <int>
## 1   958297    0     2     2    19    39    41     7     6     3     3
## 2   958296    1     7     7    25    32    24    14     7     1     3
## 3   958285    2     3    18    37    41    15     2     1     0     0
## 4   958284    1     0     3    17    17    27    24    22     5     9
## 5   958282    0     1     1     8    44    48    10     2     3     3
## 6   958279    1     0     4    12    35    49    12     5     1     2
## # ... with 3 more variables: semantic.tag.id1 <int>,
## #   semantic.tag.id2 <int>, challenge.id <int>
```

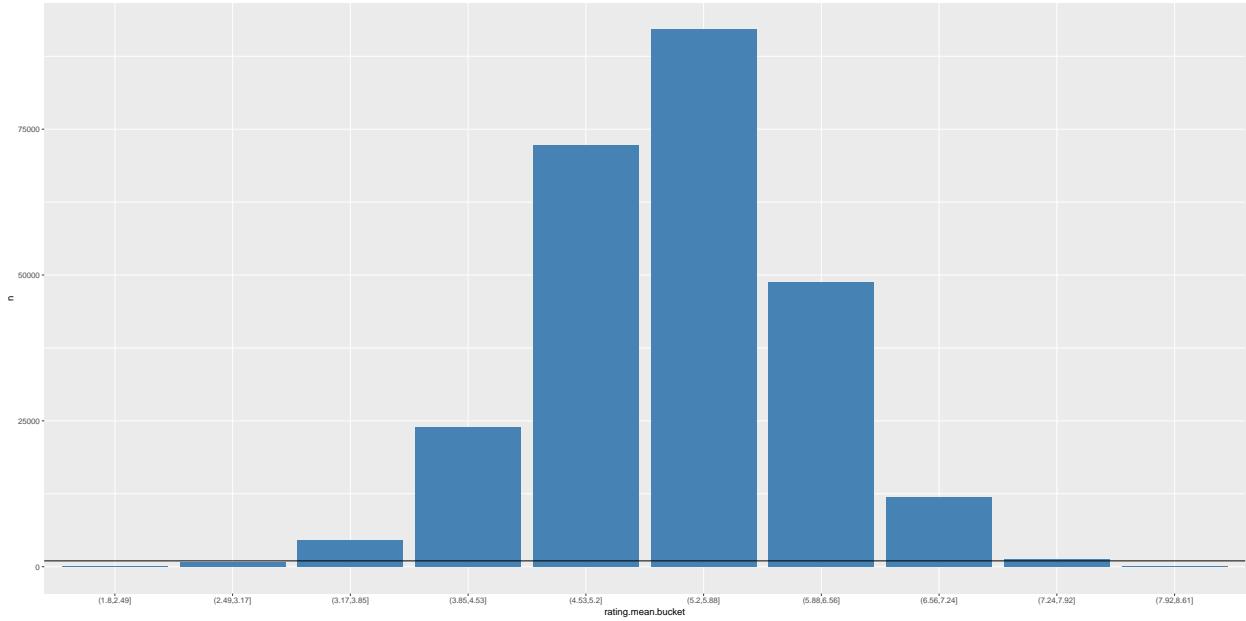
Univariate Analysis

Number Of Ratings per Image



Distribution of Ratings





```
##  
##  (1.8,2.49] (2.49,3.17] (3.17,3.85] (3.85,4.53] (4.53,5.2] (5.2,5.88]  
##    104        738      4516     23877      72318      92071  
## (5.88,6.56] (6.56,7.24] (7.24,7.92] (7.92,8.61]  
##    48699     11829      1287       69
```

Best rated Images



Worst rated Images

