Movies

# Stats by person

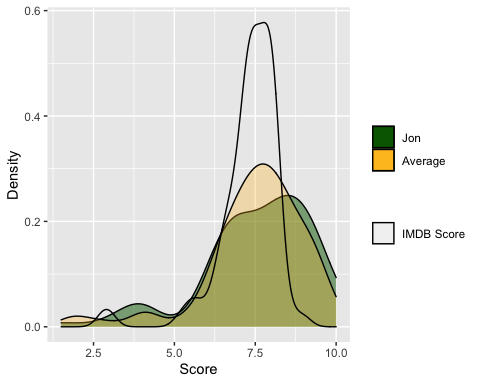
## Jon

Jon tended to rate movies highly cmpared to the rest of the reviewers as well as IMDB reviewers.

jon <- summary(movies$jon)  
average <- summary(movies$average)  
imdb <- summary(movies$imdb\_score)  
  
kable(rbind(jon,   
 average,   
 imdb))

|  | Min. | 1st Qu. | Median | Mean | 3rd Qu. | Max. |
| --- | --- | --- | --- | --- | --- | --- |
| jon | 1.5 | 6.75 | 8.0 | 7.565934 | 9.0 | 10 |
| average | 1.8 | 6.80 | 7.7 | 7.491209 | 8.5 | 10 |
| imdb | 2.8 | 7.00 | 7.4 | 7.297802 | 7.9 | 9 |

ggplot(movies) +  
 geom\_density(aes(x = jon, fill = "Jon"),  
 alpha = 0.5) +  
 geom\_density(aes(x = average, fill = "Average"),  
 alpha = 0.3) +  
 geom\_density(aes(x = imdb\_score, color = "IMDB Score")) +  
 scale\_fill\_manual(name = "", values = c("Jon" = "darkgreen", "Average" = "goldenrod1")) +  
 scale\_color\_manual(name = "", values = c("IMDB Score" = "black")) +  
 xlab("Score") +  
 ylab("Density")



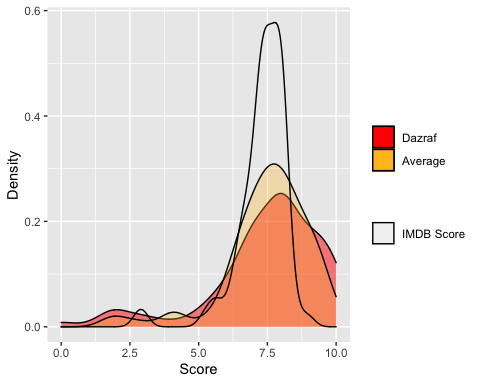
## Dazraf

Dazraf tended to have scores close to average and also rated movies more highly than IMDB users.

dazraf <- summary(movies$dazraf)  
average <- summary(movies$average)  
imdb <- summary(movies$imdb\_score)  
  
kable(rbind(dazraf,   
 average,   
 imdb))

|  | Min. | 1st Qu. | Median | Mean | 3rd Qu. | Max. |
| --- | --- | --- | --- | --- | --- | --- |
| dazraf | 0.0 | 7.0 | 8.0 | 7.456044 | 9.0 | 10 |
| average | 1.8 | 6.8 | 7.7 | 7.491209 | 8.5 | 10 |
| imdb | 2.8 | 7.0 | 7.4 | 7.297802 | 7.9 | 9 |

ggplot(movies) +  
 geom\_density(aes(x = dazraf, fill = "Dazraf"),  
 alpha = 0.5) +  
 geom\_density(aes(x = average, fill = "Average"),  
 alpha = 0.3) +  
 geom\_density(aes(x = imdb\_score, color = "IMDB Score")) +  
 scale\_fill\_manual(name = "", values = c("Dazraf" = "red", "Average" = "goldenrod1")) +  
 scale\_color\_manual(name = "", values = c("IMDB Score" = "black")) +  
 xlab("Score") +  
 ylab("Density")



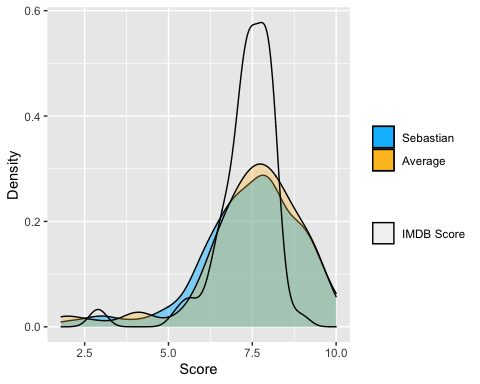
## Sebastian

Sebastian rate movies closest to both the average reviewer score and IMDB users. A true man of the people.

sebastian <- summary(movies$sebastian)  
average <- summary(movies$average)  
imdb <- summary(movies$imdb\_score)  
  
kable(rbind(sebastian,   
 average,   
 imdb))

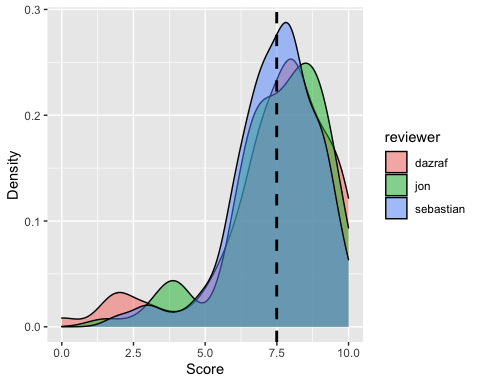
|  | Min. | 1st Qu. | Median | Mean | 3rd Qu. | Max. |
| --- | --- | --- | --- | --- | --- | --- |
| sebastian | 2.0 | 6.75 | 7.5 | 7.461538 | 8.5 | 10 |
| average | 1.8 | 6.80 | 7.7 | 7.491209 | 8.5 | 10 |
| imdb | 2.8 | 7.00 | 7.4 | 7.297802 | 7.9 | 9 |

ggplot(movies) +  
 geom\_density(aes(x = sebastian, fill = "Sebastian"),  
 alpha = 0.5) +  
 geom\_density(aes(x = average, fill = "Average"),  
 alpha = 0.3) +  
 geom\_density(aes(x = imdb\_score, color = "IMDB Score")) +  
 scale\_fill\_manual(name = "", values = c("Sebastian" = "deepskyblue", "Average" = "goldenrod1")) +  
 scale\_color\_manual(name = "", values = c("IMDB Score" = "black")) +  
 xlab("Score") +  
 ylab("Density")

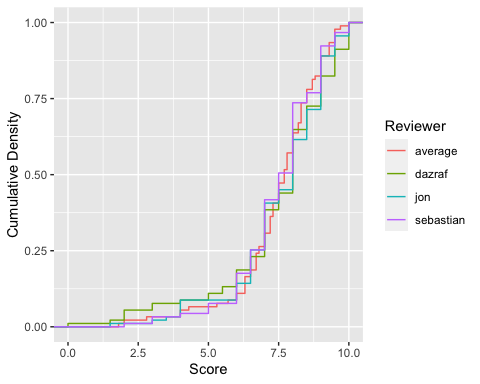


## The reviewers relative to each other

plot\_multi\_histogram <- function(df, feature, label\_column) {  
 plt <-  
 ggplot(df, aes(x = eval(parse(text = feature)), fill = eval(parse(text =  
 label\_column)))) +  
 # geom\_histogram(  
 # alpha = 0.7,  
 # position = "identity",  
 # aes(y = ..density..),  
 # color = "black"  
 # ) +  
 geom\_density(alpha = 0.5) +  
 geom\_vline(  
 aes(xintercept = mean(eval(  
 parse(text = feature)  
 ))),  
 color = "black",  
 linetype = "dashed",  
 size = 1  
 ) +  
 labs(x = "Score", y = "Density")  
 plt + guides(fill = guide\_legend(title = label\_column))  
}  
  
plot\_multi\_histogram(movies.long %>% filter(reviewer != "average"),  
 'reviewer\_score',  
 'reviewer')



ggplot(movies.long, aes(reviewer\_score, colour = reviewer)) +   
 stat\_ecdf()+  
 scale\_colour\_hue(name="Reviewer"#,   
 # labels=c('AAA','BBB', 'CCC')  
 ) +  
 labs(x = "Score", y = "Cumulative Density")



# Reviewers vs the reviewer who picked the movie

All reviewers took turns picking movies and this is how their picks stacked up.

summary(  
 tableby(  
 pick\_coded ~ kwt(dazraf) + kwt(sebastian) + kwt(jon) + kwt(average),  
 data = movies %>%  
 mutate(pick\_coded = ifelse(  
 pick == "Farzad",  
 "Dazraf pick",  
 ifelse(  
 pick == "Steven",  
 "Sebastian pick",  
 ifelse(pick == "Jon",  
 "Jon pick",  
 "")  
 )  
 )) %>%  
 filter(pick != "All"),  
 total = FALSE  
 ),  
 text = TRUE,  
 pfootnote = TRUE  
)

|  | Dazraf pick (N=29) | Jon pick (N=29) | Sebastian pick (N=29) | p value |
| --- | --- | --- | --- | --- |
| dazraf |  |  |  | 0.464 (1) |
| - Mean (SD) | 7.724 (1.962) | 7.190 (2.085) | 7.603 (1.754) |  |
| - Range | 2.000 - 10.000 | 1.500 - 10.000 | 2.000 - 10.000 |  |
| sebastian |  |  |  | 0.002 (1) |
| - Mean (SD) | 6.810 (1.644) | 7.655 (1.119) | 7.983 (1.257) |  |
| - Range | 2.000 - 10.000 | 5.000 - 9.500 | 3.000 - 9.500 |  |
| jon |  |  |  | 0.001 (1) |
| - Mean (SD) | 6.879 (1.840) | 8.328 (1.227) | 7.586 (1.452) |  |
| - Range | 1.500 - 10.000 | 4.000 - 10.000 | 3.500 - 10.000 |  |
| average |  |  |  | 0.130 (1) |
| - Mean (SD) | 7.134 (1.664) | 7.717 (1.277) | 7.724 (1.304) |  |
| - Range | 1.800 - 10.000 | 4.000 - 9.500 | 2.800 - 9.500 |  |

1. Kruskal-Wallis rank sum test

Knowing that the data are non-parametric, a Kruskal-Wallis test of significance was used to compare how each reviewer scored each others’ picks. On average, there was no significant difference between how the reviewers rated each others picks. However, on an individual basis, both Sebastian and Jon rated each reviewer’s picks signficantly different. Dazraf didn’t.

Given that Sebastian and Jon both rated Dazraf’s picks the lowest compared to each others’ picks, it can be assumed that only Dazraf tended to rate Dazraf’s picks well.

# Over the years

Here’s how the reviewers scored movies based on the decade the movie was released in.

kable(  
 movies %>%  
 group\_by(decade) %>%  
 summarise(  
 count\_movies = n(),  
 jon\_average = mean(jon),  
 jon\_range = paste0(min(jon), " - ", max(jon)),  
 dazraf\_average = mean(dazraf),  
 dazraf\_range = paste0(min(dazraf), " - ", max(dazraf)),  
 sebastian\_average = mean(sebastian),  
 sebastian\_range = paste0(min(sebastian), " - ", max(sebastian)),  
 avg\_average = mean(average),  
 avg\_range = paste0(min(average), " - ", max(average)),  
 )  
)

| decade | count\_movies | jon\_average | jon\_range | dazraf\_average | dazraf\_range | sebastian\_average | sebastian\_range | avg\_average | avg\_range |
| --- | --- | --- | --- | --- | --- | --- | --- | --- | --- |
| 1920 | 1 | 8.000000 | 8 - 8 | 8.500000 | 8.5 - 8.5 | 7.000000 | 7 - 7 | 7.800000 | 7.8 - 7.8 |
| 1940 | 3 | 8.333333 | 6.5 - 9.5 | 7.833333 | 6.5 - 10 | 7.166667 | 6 - 8 | 7.766667 | 6.3 - 9 |
| 1950 | 5 | 8.000000 | 6.5 - 9 | 6.600000 | 5.5 - 8.5 | 8.100000 | 7 - 9 | 7.560000 | 6.7 - 8.5 |
| 1960 | 3 | 8.833333 | 8 - 9.5 | 9.333333 | 8 - 10 | 8.833333 | 7.5 - 10 | 9.000000 | 8.2 - 9.5 |
| 1970 | 7 | 7.928571 | 6.5 - 10 | 7.214286 | 2 - 10 | 8.071429 | 6 - 10 | 7.742857 | 5.3 - 10 |
| 1980 | 6 | 7.000000 | 6 - 8.5 | 8.000000 | 7 - 9.5 | 6.666667 | 5 - 9 | 7.233333 | 6.5 - 8.7 |
| 1990 | 16 | 7.843750 | 6 - 10 | 7.718750 | 5 - 9.5 | 7.531250 | 4 - 9 | 7.687500 | 6.3 - 9.5 |
| 2000 | 15 | 7.133333 | 4 - 10 | 7.000000 | 1.5 - 10 | 7.066667 | 5 - 9.5 | 7.073333 | 4 - 9.5 |
| 2010 | 21 | 7.428571 | 1.5 - 10 | 7.619048 | 2 - 10 | 7.428571 | 2 - 10 | 7.485714 | 1.8 - 9.7 |
| 2020 | 14 | 7.357143 | 3 - 9.5 | 7.035714 | 0 - 9.5 | 7.464286 | 3 - 9 | 7.278571 | 2 - 9 |

The 60s had the highest scores, with all three reviewers rating 60s movies highest of all decades from which their movies were selected. However only 3 movies they watched were released in the 1960s. The majority of their picks came from 1990 through the present day and their more even scores tend to reflect that.

# Genres

It should be noted that movies could be labelled with multiple genres and most of the movies that the reviewers picked were categorized with multiple genres. There were between 1 and 6 labelled genres per movie.

kable(movies.genres %>%  
 count(genre) %>%  
 arrange(desc(n)))

| genre | n |
| --- | --- |
| Drama | 59 |
| Comedy | 36 |
| Crime | 27 |
| Thriller | 27 |
| Action | 17 |
| Mystery | 17 |
| Romance | 16 |
| Sci-Fi | 11 |
| Horror | 9 |
| Adventure | 8 |
| Fantasy | 7 |
| Biography | 6 |
| Animation | 5 |
| War | 5 |
| Family | 4 |
| Music | 4 |
| Documentary | 3 |
| Film-Noir | 2 |
| History | 2 |
| Musical | 2 |
| Western | 1 |

kable(movies.genres.long %>%  
 group\_by(reviewer, genre) %>%  
 summarise(mean\_score = mean(reviewer\_score)) %>%   
 filter(reviewer == "average") %>%   
 arrange(desc(mean\_score)) #%>%   
 # pivot\_wider(names\_from = genre, values\_from = mean\_score)  
)

`summarise()` has grouped output by 'reviewer'. You can override using the `.groups` argument.

| reviewer | genre | mean\_score |
| --- | --- | --- |
| average | War | 8.400000 |
| average | History | 8.350000 |
| average | Biography | 8.133333 |
| average | Sci-Fi | 8.000000 |
| average | Animation | 7.880000 |
| average | Mystery | 7.817647 |
| average | Documentary | 7.800000 |
| average | Film-Noir | 7.750000 |
| average | Thriller | 7.659259 |
| average | Drama | 7.555932 |
| average | Crime | 7.503704 |
| average | Romance | 7.493750 |
| average | Adventure | 7.475000 |
| average | Horror | 7.366667 |
| average | Action | 7.300000 |
| average | Comedy | 7.019444 |
| average | Western | 7.000000 |
| average | Music | 6.725000 |
| average | Fantasy | 6.228571 |
| average | Family | 6.150000 |
| average | Musical | 5.400000 |

Most movies were categorized as eihter a comedy or drama with the crime, thriller, and action genres rounding out the top 5 genres most represented in their selection.

By rating, however, war was the highest rated genre. History, biography, sci-fi, and animation were also highly scored by the reviewers.

To see if the reviewers were unique in their love of these genre specific films, we also looked at how IMDB users rated their selections by genre.

kable(movies.genres.long %>%  
 group\_by(genre) %>%  
 summarise(mean\_score = mean(imdb\_score)) %>%   
 # filter(reviewer == "average") %>%   
 arrange(desc(mean\_score))  
)

| genre | mean\_score |
| --- | --- |
| History | 8.100000 |
| Film-Noir | 8.000000 |
| Animation | 7.840000 |
| War | 7.640000 |
| Biography | 7.566667 |
| Sci-Fi | 7.490909 |
| Mystery | 7.488235 |
| Drama | 7.461017 |
| Romance | 7.400000 |
| Action | 7.370588 |
| Documentary | 7.366667 |
| Crime | 7.337037 |
| Thriller | 7.307407 |
| Adventure | 7.287500 |
| Horror | 7.011111 |
| Comedy | 6.897222 |
| Music | 6.825000 |
| Fantasy | 6.585714 |
| Western | 6.500000 |
| Family | 6.375000 |
| Musical | 5.300000 |

With the notable exception of film noir, the highest reviewer-rated genres were reflective of general tastes. This indicates that it’s more likely than not that the reviewers selected well-loved movies from those genres knowing that they were already highly rated by the public as well as critics.

For a more granular view, here are the top 5 highest rated genres by reviewer.

## Jon’s top 5 genres

kable(  
 movies.genres.long %>%  
 group\_by(reviewer, genre) %>%  
 summarise(mean\_score = mean(reviewer\_score)) %>%  
 filter(reviewer == "jon") %>%  
 arrange(desc(mean\_score)) %>%  
 slice(1:5)  
)

`summarise()` has grouped output by 'reviewer'. You can override using the `.groups` argument.

| reviewer | genre | mean\_score |
| --- | --- | --- |
| jon | Film-Noir | 9.000000 |
| jon | Animation | 8.900000 |
| jon | War | 8.600000 |
| jon | History | 8.500000 |
| jon | Sci-Fi | 8.318182 |

Jon’s favorite film(s) were of the film noir genre. His other highly rated movies were reflected in the average ratings.

## Dazraf’s top 5 genres

kable(  
 movies.genres.long %>%  
 group\_by(reviewer, genre) %>%  
 summarise(mean\_score = mean(reviewer\_score)) %>%  
 filter(reviewer == "dazraf") %>%  
 arrange(desc(mean\_score)) %>%  
 slice(1:5)  
)

`summarise()` has grouped output by 'reviewer'. You can override using the `.groups` argument.

| reviewer | genre | mean\_score |
| --- | --- | --- |
| dazraf | History | 8.500000 |
| dazraf | Biography | 8.250000 |
| dazraf | War | 8.200000 |
| dazraf | Sci-Fi | 8.090909 |
| dazraf | Documentary | 7.833333 |

Dazraf’s favorite films were reflective of average ratings with the exception of documentaries.

## Sebastian’s top 5 genres

kable(  
 movies.genres.long %>%  
 group\_by(reviewer, genre) %>%  
 summarise(mean\_score = mean(reviewer\_score)) %>%  
 filter(reviewer == "sebastian") %>%  
 arrange(desc(mean\_score)) %>%  
 slice(1:5)  
)

`summarise()` has grouped output by 'reviewer'. You can override using the `.groups` argument.

| reviewer | genre | mean\_score |
| --- | --- | --- |
| sebastian | War | 8.400000 |
| sebastian | Biography | 8.166667 |
| sebastian | Film-Noir | 8.000000 |
| sebastian | History | 8.000000 |
| sebastian | Mystery | 7.764706 |

Sebastian’s top 5 highest rated genres deviated the most from the average, with film noir and mystery films making his top 5.

It seems that Dazraf might particularly dislike the film noir genre given its appearance in both Jon and Sebastian’s top 5 and absence in the average top 5.

It should be noted that after the top 5, the individual reviewers’ preferences were the next most highly rated with the mystery, documentary, and film noir genres making up the 6th-8th top rated genres respectively.