Analysis of Anime Popularity:

by MyAnimeList Fan-Service - Milestone 3

Aaron J. Brown

Spring 2023

Introduction.

With the increasing popularity of anime, we are seeing the emergence of more anime titles in more recent years. In contribution to anime's constant uprising, we see more platforms in which to consume our animated favorites; some of these platforms also allow users to subscribe, watch, favorite, and rate new and old titles to their liking. The avaliabilty of data like this has allowed programmers to determine the most popular, most viewed, or most favorited anime titles, and serve them right back to their audience as recommendations!

About the Data.

The data used in this analysis was scrapped from an anime rating and listing site (MyAnimeList) that provides details of almost all available anime. Though the site does not allow users to actually watch any of the programs directly from the site, it provides a space for anime fans to compile all of their favorite anime into a list, as well as ranking, sharing, commenting, and a host of other features.

The datasets used provide details of over 18,000 separate anime productions, including series, movies, OVAs, ONAs, music, and others. Information such as number of episodes, duration, score, popularity, members, etc. can be extracted from the data. The three datasets contain data from three years: 2023, 2022, and 2021.

The problem statement you addressed.

This analysis will help shed light on what titles and genres are most liked and consumed, and what contributes to the popularity of those titles. Visualizations will be used to display the relationship between popularity and related variables.

How you addressed this problem statement.

The approach is to utilize, primarily: scores, popularity, members, and favorites to determine the relationship between these variables and the most popular anime, and will determine if there are any relationships between an anime's popularity and the amount of members, as well as the different means to which fans interact with anime on MyAnimeList. As MyAnimeList allows only registered users to "score", and un-registered users can "favorite" there is expected to be a difference with the relationship between scores and favorites as they relate to popularity, possibly due

to a more dedicated fan-service willing to go through the motions of creating an account.

GGPlot and Patchwork are used to produce visualizations of the various relationships mentioned above.

Analysis.

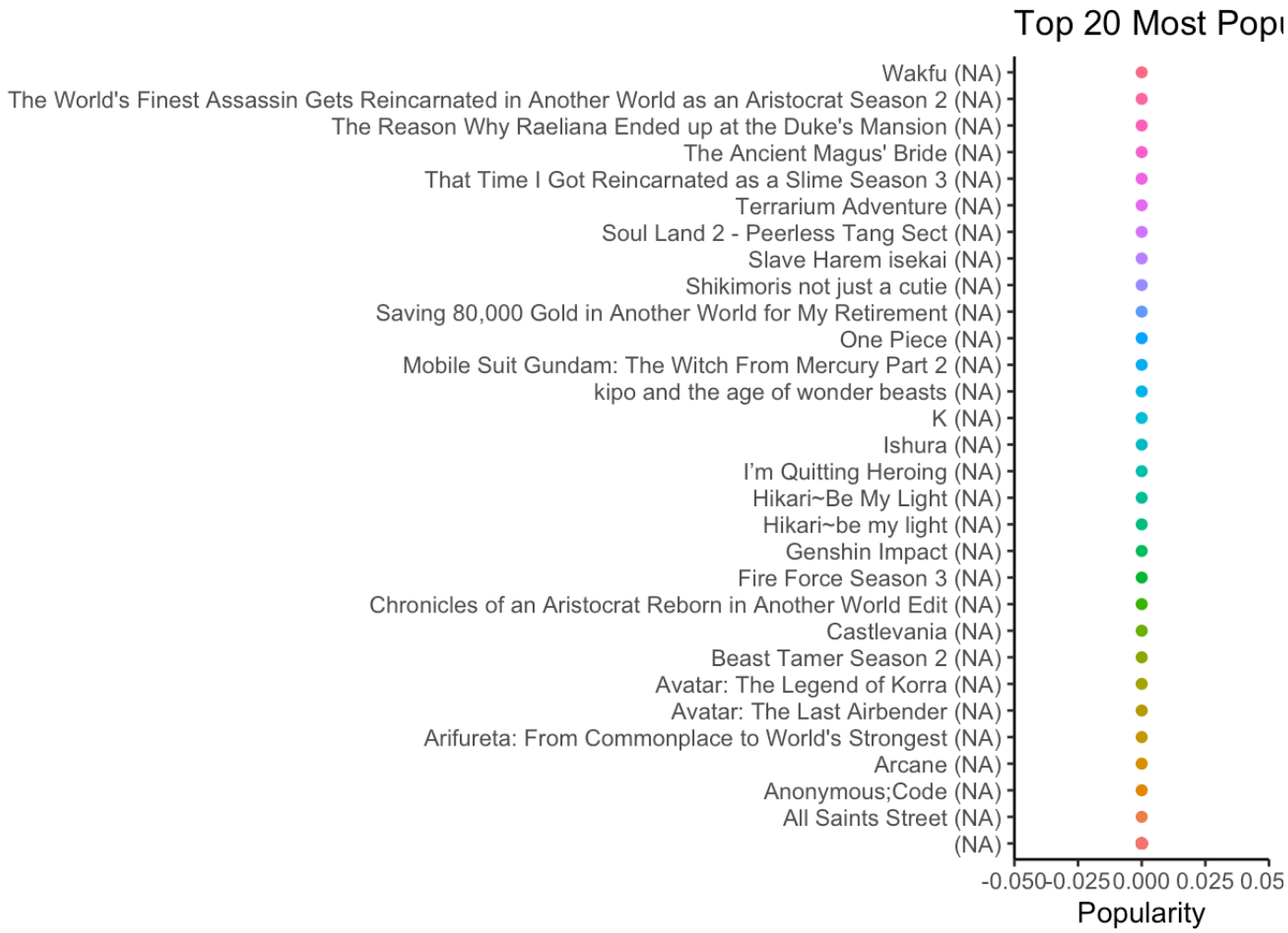
```
anime_data23 <- read.csv("/Users/aaronbrown/Documents/ClassWork/DSC 520 - Statistics for Data Science/Project/anime_data_2023.csv")
#View(anime_data23)

anime_tidy23 <- anime_data23%>%
  select(-c(url, images, trailer, approved, titles, title_japanese,  producers, studios, genres, themes, demographics, title_synonyms, aired, synopsis, background, broadcast, licensors, explicit_genres))
head(anime_tidy23)
```

```
##   mal_id                                    title          title_english type
## 1      1                          Cowboy Bebop           Cowboy Bebop   TV
## 2      8                        Bouken Ou Beet Beet the Vandel Buster   TV
## 3     18                Initial D Fourth Stage                           TV
## 4     23                      Ring ni Kakero 1                           TV
## 5     28                      Yakitate!! Japan       Yakitate!! Japan   TV
## 6     45 Rurouni Kenshin: Meiji Kenkaku Romantan        Rurouni Kenshin   TV
##     source episodes          status airing      duration
## 1 Original       26 Finished Airing  False 24 min per ep
## 2    Manga       52 Finished Airing  False 23 min per ep
## 3    Manga       24 Finished Airing  False 27 min per ep
## 4    Manga       12 Finished Airing  False 25 min per ep
## 5    Manga       69 Finished Airing  False 24 min per ep
## 6    Manga       94 Finished Airing  False 25 min per ep
##                          rating score scored_by rank popularity members
## 1 R - 17+ (violence & profanity)  8.75    889879   39         43 1727016
## 2                  PG - Children  6.94      6351 4337       5037   14809
## 3      PG-13 - Teens 13 or older  8.15     94664  402       1269  167698
## 4                  PG - Children  6.39      1329 7311       8178    4467
## 5      PG-13 - Teens 13 or older  7.92     40054  688       1953   94169
## 6      PG-13 - Teens 13 or older  8.29    227129  246        419  469746
##   favorites season year
## 1     76617 spring 1998
## 2        14   fall 2004
## 3      1201 spring 2004
## 4         3   fall 2004
## 5       891   fall 2004
## 6      9802 winter 1996
```

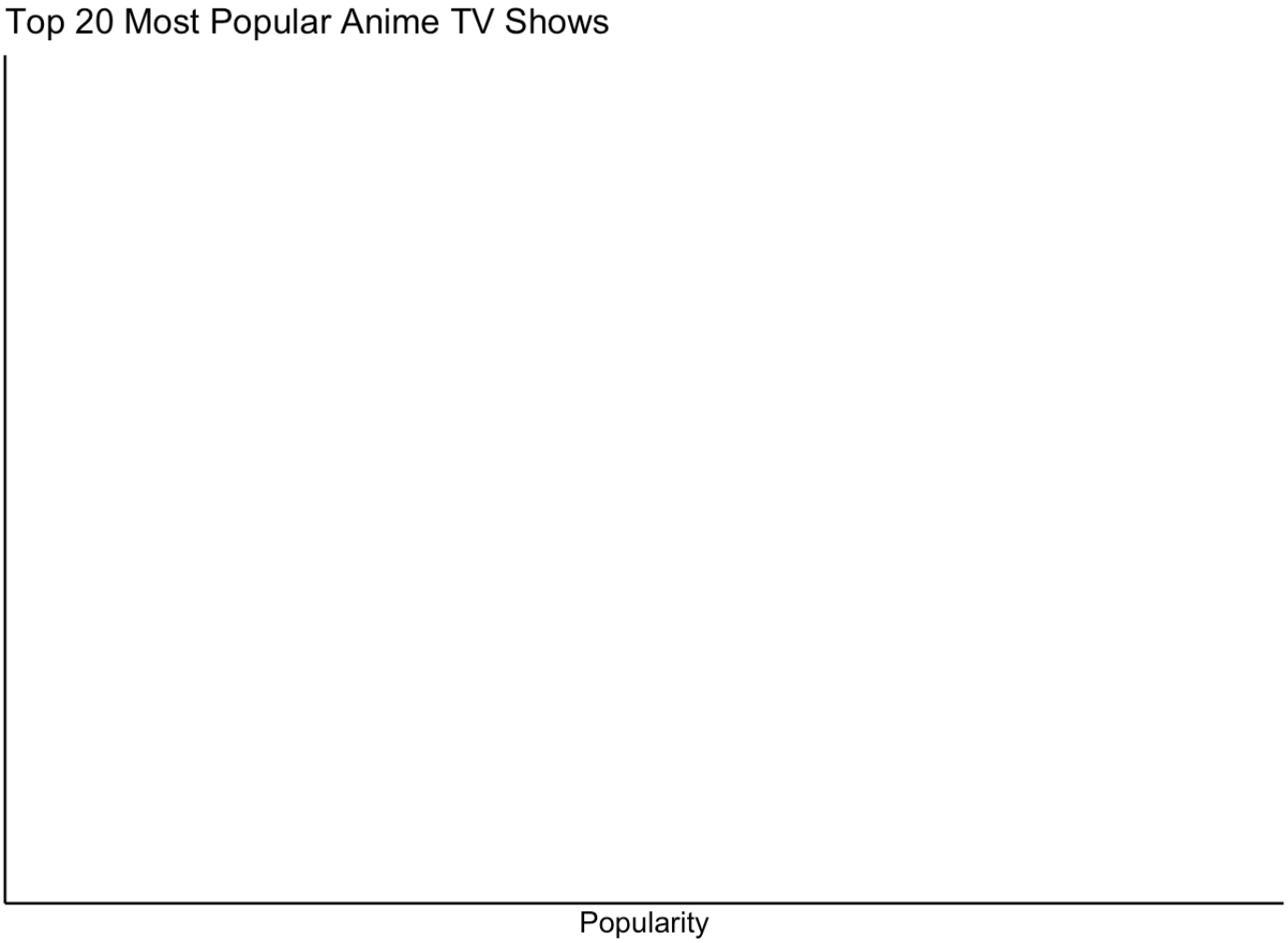
```
#View(anime_tidy23)
```

```
unique_anime <- anime_tidy23%>%
distinct()%>%
filter(!is.na(title_english))
#View(unique_anime)
```

```
tv_anime <- unique_anime%>%filter(type=="TV")
tv_plot <-tv_anime%>%
  mutate(graph_name=paste0(title_english," (",year,")"))%>%
  top_n(-20,wt=popularity) %>%
  ggplot(aes(reorder(graph_name,desc(popularity)),popularity,colour=title_english))+
  geom_point(show.legend = FALSE)+
  geom_segment(aes(x=graph_name,xend=graph_name,y=0,yend=popularity),show.legend = FA
LSE)+
  coord_flip()+
  theme_classic()+
  labs(x="",y="Popularity",title = "Top 20 Most Popular Anime TV Shows")
tv_plot
```



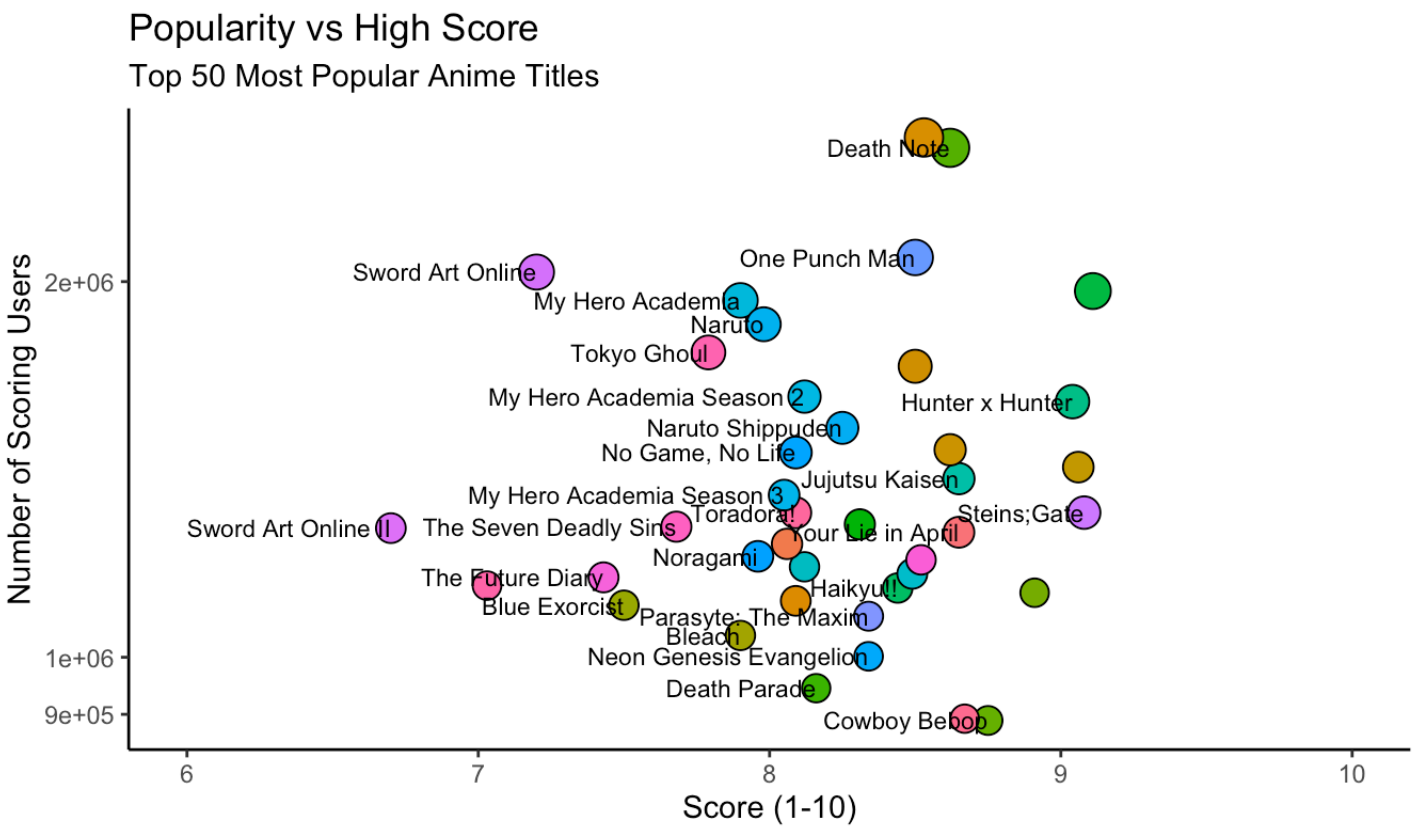
```
tv_anime_airing <- anime_tidy23%>%filter(type=="TV",airing=="TRUE")
#View(tv_anime_airing)
tv_plot_airing <-tv_anime_airing%>%
  mutate(graph_name=paste0(title," (",year,")"))%>%
  ggplot(aes(reorder(graph_name,desc(popularity)),popularity,colour=title))+
  geom_point(show.legend = FALSE)+
  geom_segment(aes(x=graph_name,xend=graph_name,y=0,yend=popularity),show.legend = FA
LSE)+
  coord_flip()+
  theme_classic()+
  labs(x="",y="Popularity",title = "Top 20 Most Popular Anime TV Shows")+
  theme(axis.text.y.left = element_text(size = 12))
tv_plot_airing
```



```
tv_anime %>%
  filter(popularity <= 50) %>%
  mutate(title_english = str_replace(title_english, "Code Geass: Lelouch of the Rebel
lion", "Code Geass")) %>%
  ggplot(aes(score, scored_by)) +
  geom_point(shape=21,aes(fill=title_english,size=members)) +
  geom_text(aes(label = title_english ), check_overlap = T, show.legend = F, size = 3
, hjust = 1) +
  xlim(c(6, 10)) +
  scale_y_log10()+
  labs(title = "Popularity vs High Score",
       subtitle = "Top 50 Most Popular Anime Titles",
       y = "Number of Scoring Users",
       x = "Score (1-10)") +
  theme_classic()+
  theme(legend.position = 'none',aspect.ratio = 0.5)
```

```
## Warning: Removed 47 rows containing missing values (`geom_point()`).
```

```
## Warning: Removed 47 rows containing missing values (`geom_text()`).
```

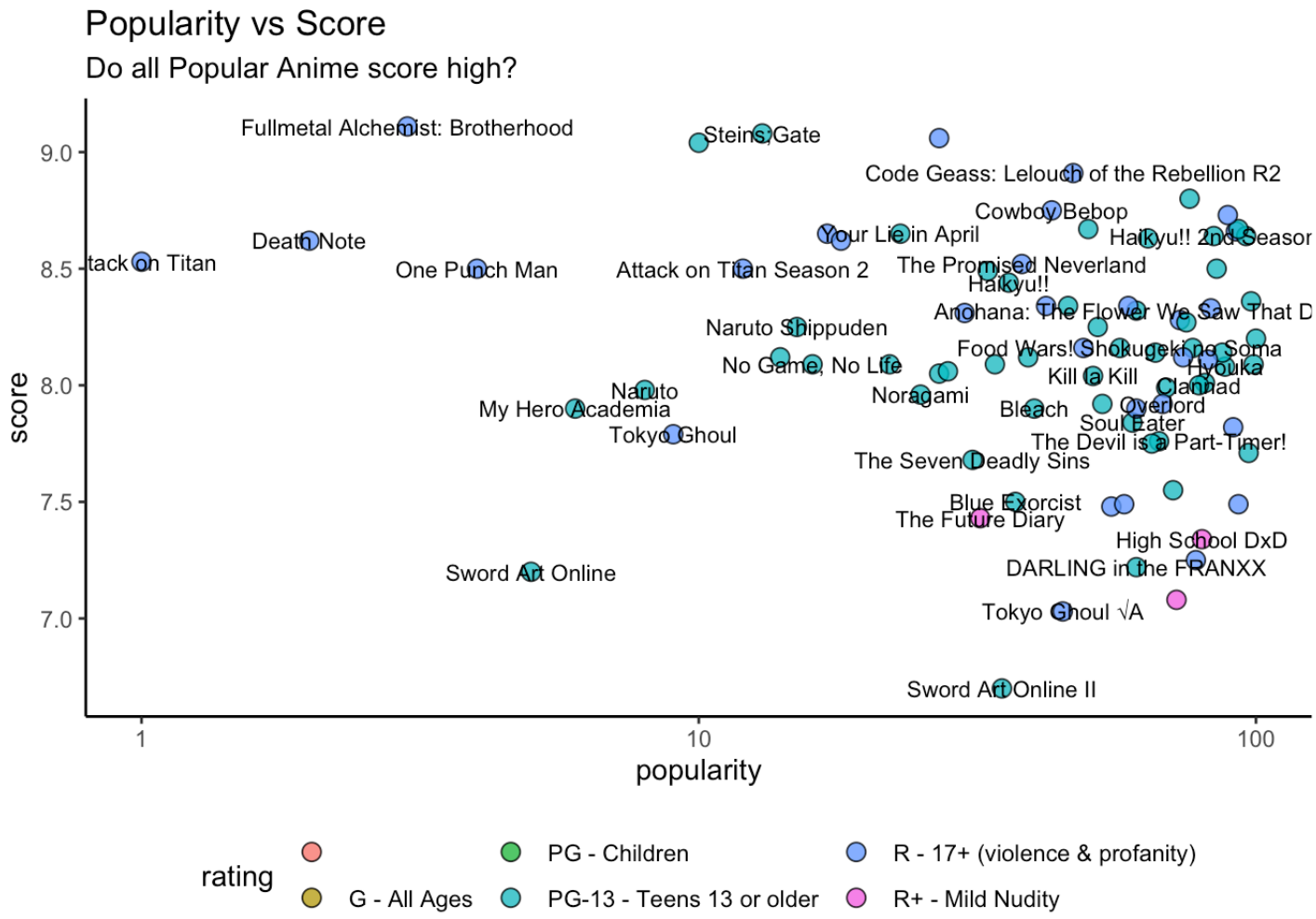


```
tv_anime%>%
  filter(popularity<=100) %>%
  ggplot(aes(popularity,score,fill=rating))+
  geom_point(shape=21,size=3,alpha=0.8)+
  geom_text(aes(label=title_english),size=3,check_overlap = TRUE)+
  scale_x_log10()+
  theme_classic()+
  scale_color_brewer(palette = 'Set1')+
  theme(legend.position = "bottom")+
  labs(title = "Popularity vs Score",subtitle = "Do all Popular Anime score high?")
```

```
## Warning: Transformation introduced infinite values in continuous x-axis
## Transformation introduced infinite values in continuous x-axis
```

```
## Warning: Removed 47 rows containing missing values (`geom_point()`).
```

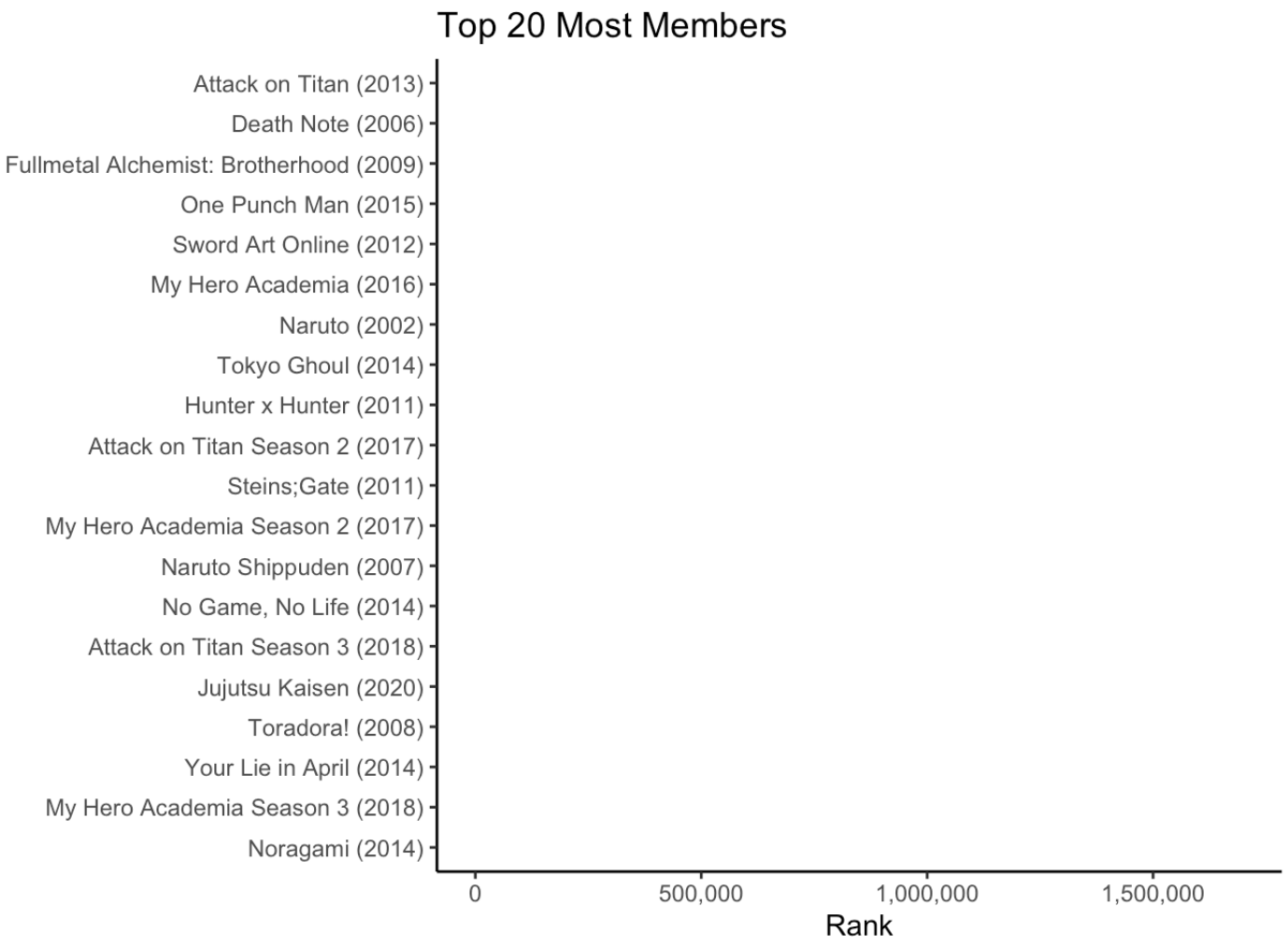
```
## Warning: Removed 47 rows containing missing values (`geom_text()`).
```



```
most_members <- tv_anime%>%
  top_n(20,wt=members) %>%
  mutate(graph_name=paste0(title_english," (",year,")"),graph_name=fct_reorder(graph_
name,members))%>%
  ggplot(aes(graph_name,members,fill=graph_name))+
  geom_bar(stat = 'identity',width=0.5,show.legend = FALSE,color='red')+
  coord_flip()+
  theme_classic()+
  scale_y_continuous(limits = c(0,1700000),labels = comma)+
  geom_text(aes(label=comma(members)),size=3)+
  labs(x="",y="Rank",title = "Top 20 Most Members")
most_members
```

```
## Warning: Removed 20 rows containing missing values (`position_stack()`).
```

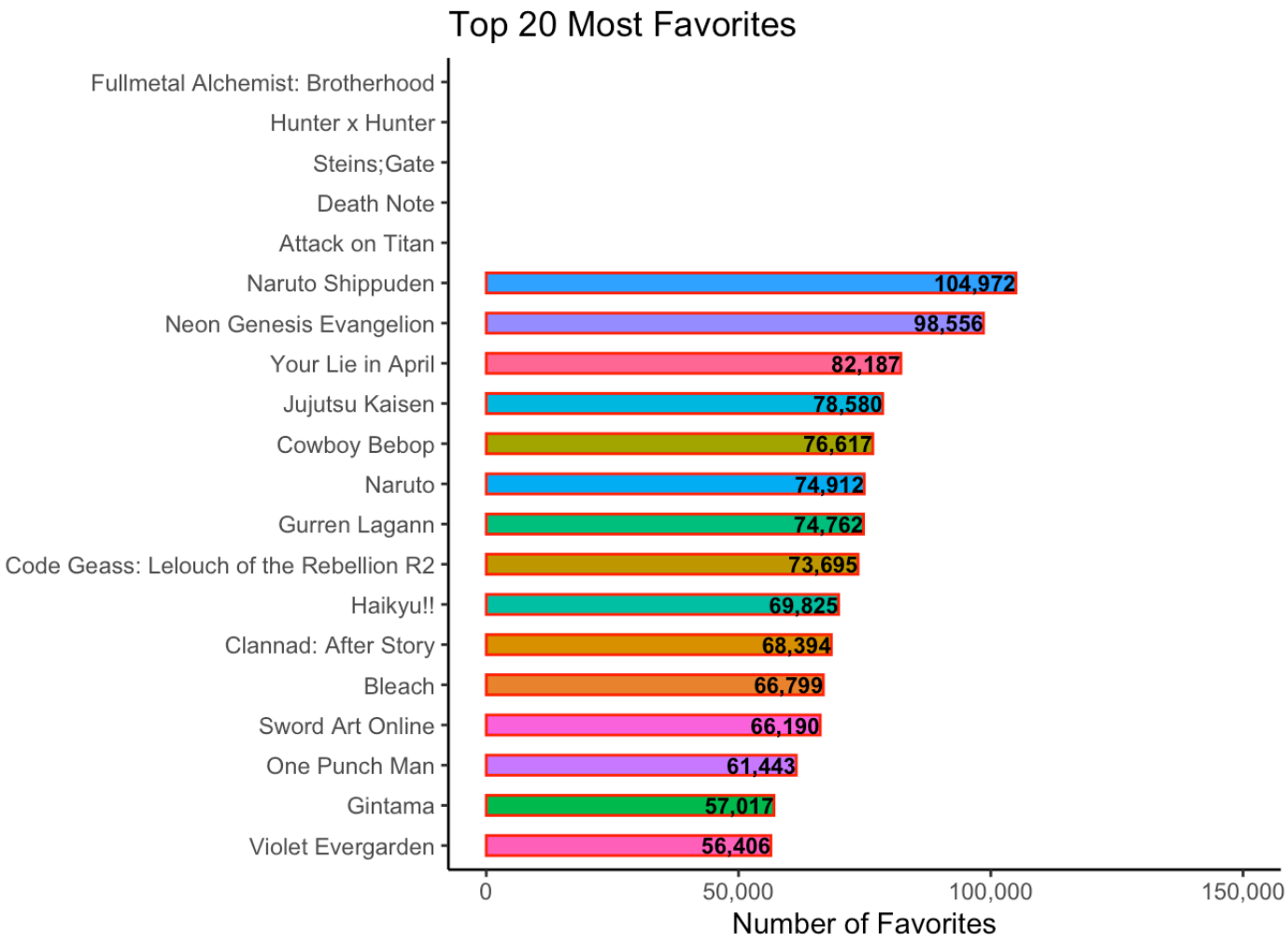
```
## Warning: Removed 20 rows containing missing values (`geom_text()`).
```



```
most_favorite<-  tv_anime%>%top_n(20,wt=favorites) %>%  mutate(title_english=fct_reor
der(title_english,favorites))%>%
  ggplot(aes(title_english,favorites,fill=title_english))+geom_bar(stat = 'identity',
width=0.5,show.legend = FALSE,color='red')+
  coord_flip()+ theme_classic()+ scale_y_continuous(limits = c(0,150000),labels = com
ma)+ geom_text(aes(label=comma(favorites)),size=3,hjust=1,fontface='bold')+
  labs(x="",y="Number of Favorites",title = "Top 20 Most Favorites")
most_favorite
```

```
## Warning: Removed 5 rows containing missing values (`position_stack()`).
```

```
## Warning: Removed 5 rows containing missing values (`geom_text()`).
```



```
most_scored<- tv_anime%>% top_n(20,wt=scored_by) %>%  mutate(title_english=fct_reorde
r(title_english,scored_by))%>%
  ggplot(aes(title_english,scored_by,fill= title_english))+ geom_bar(stat = 'identity
',width=0.5,color='red',show.legend = FALSE)+
  coord_flip()+ theme_classic()+scale_y_continuous(limits = c(0,1500000),labels = com
ma)+ geom_text(aes(label=comma(scored_by)),size=3,hjust=1,fontface='bold')+
  labs(x="",y="Number of Favorites",title = "Top 20 Most Scored")
most_favorite + most_members+ most_scored + plot_layout(widths = 2)
```

```
## Warning: Removed 5 rows containing missing values (`position_stack()`).
```

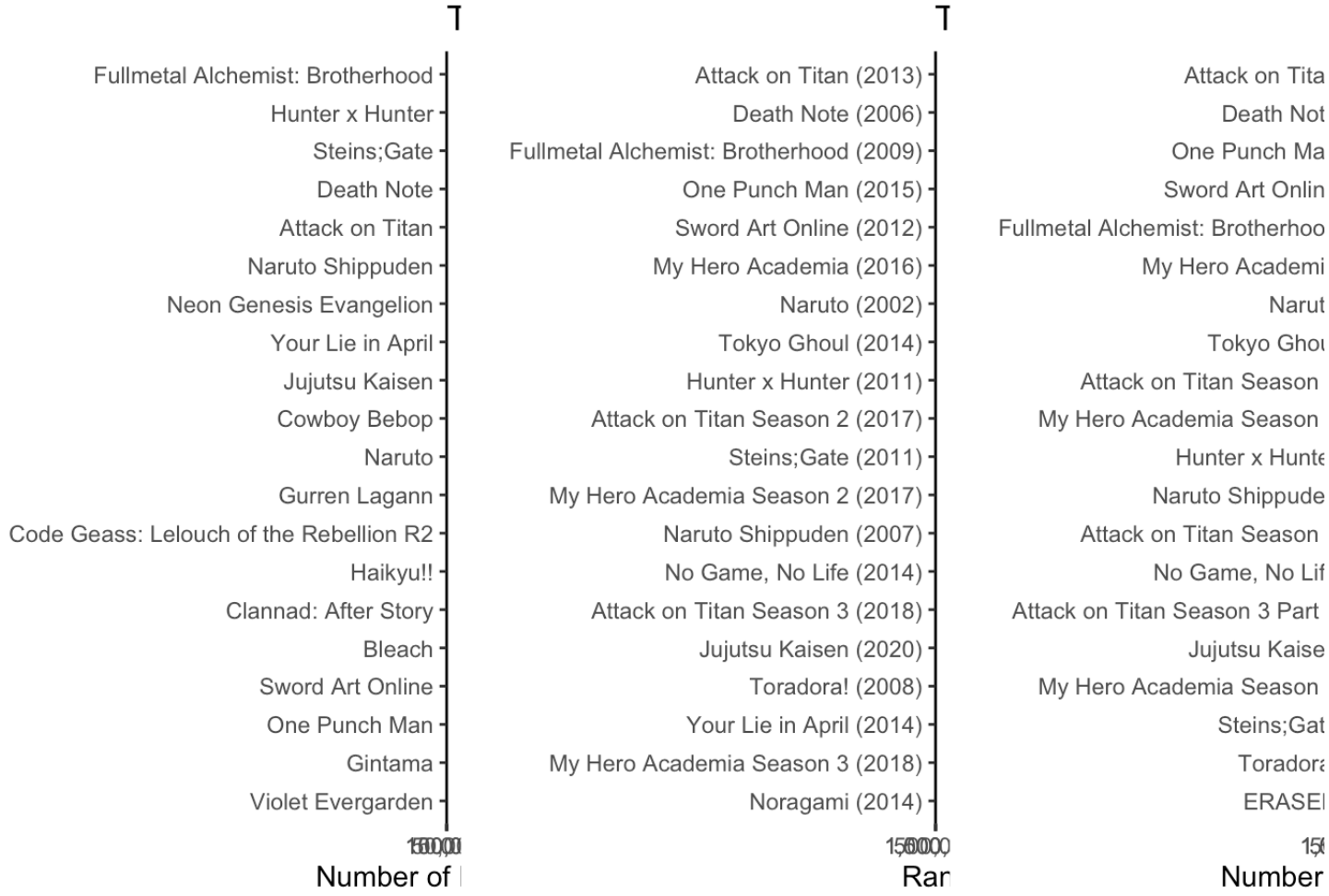
```
## Warning: Removed 5 rows containing missing values (`geom_text()`).
```

```
## Warning: Removed 20 rows containing missing values (`position_stack()`).
```

```
## Warning: Removed 20 rows containing missing values (`geom_text()`).
```

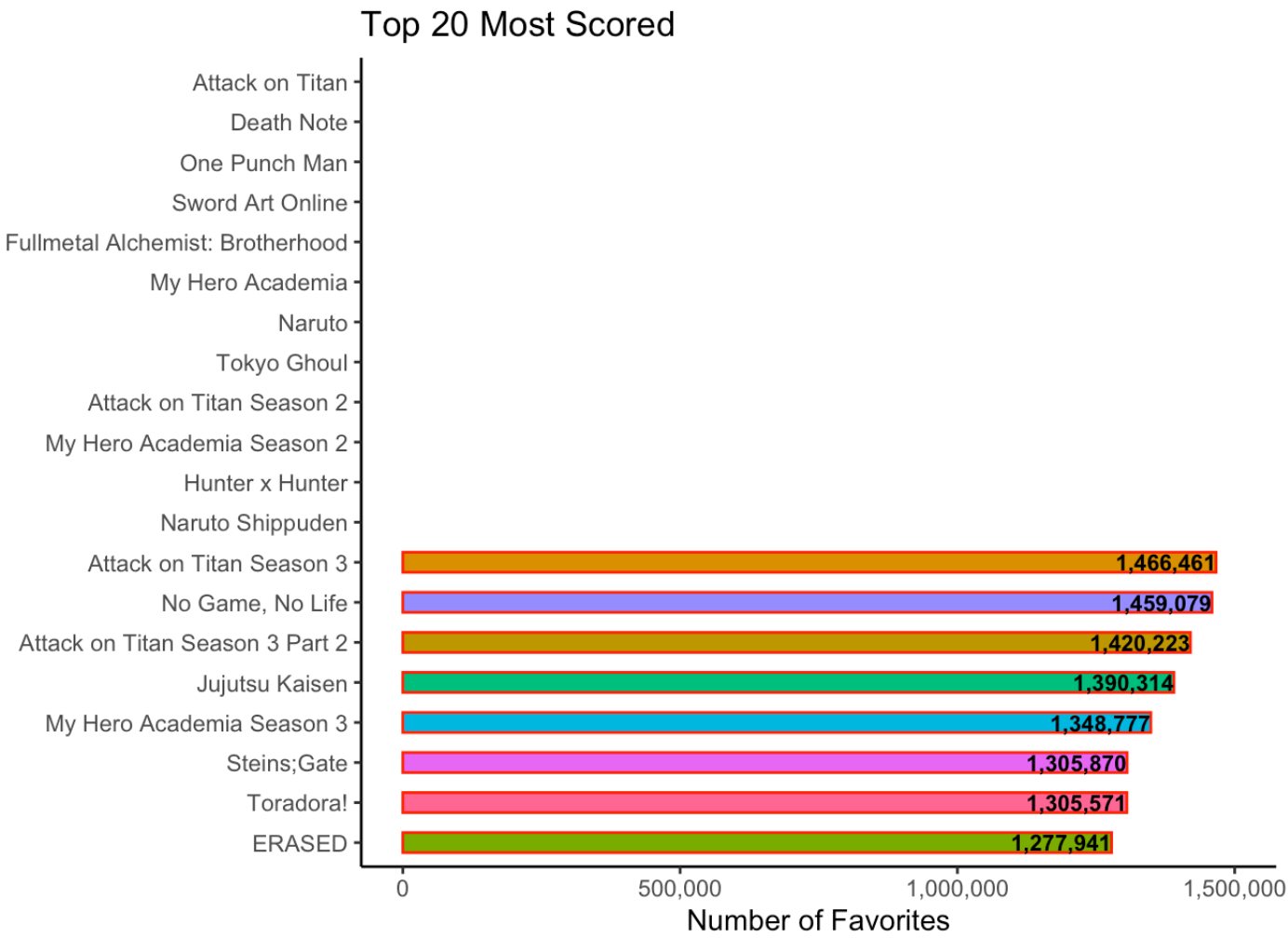
```
## Warning: Removed 12 rows containing missing values (`position_stack()`).
```

```
## Warning: Removed 12 rows containing missing values (`geom_text()`).
```




```
most_scored
```

```
## Warning: Removed 12 rows containing missing values (`position_stack()`).
## Removed 12 rows containing missing values (`geom_text()`).
```




```
# POPULARITY x SCORE
model1 = lm(popularity ~ score, data = tv_anime)
anova_table1 = anova(model1)
anova_table1
```

```
## Analysis of Variance Table
##
## Response: popularity
##              Df     Sum Sq    Mean Sq F value    Pr(>F)
## score         1 1.8576e+10 1.8576e+10  1530.6 < 2.2e-16 ***
## Residuals 3395 4.1204e+10 1.2137e+07
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
```

```
# POPULARITY x FAVORITES
model2 = lm(popularity ~ favorites, data = tv_anime)
anova_table2 = anova(model2)
anova_table2
```

```
## Analysis of Variance Table
##
## Response: popularity
##             Df     Sum Sq    Mean Sq F value    Pr(>F)
## favorites    1 1.7122e+10 1.7122e+10  250.96 < 2.2e-16 ***
## Residuals 5686 3.8794e+11 6.8226e+07
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
```

```
# POPULARITY x MEMBERS
model3 = lm(popularity ~ members, data = tv_anime)
anova_table3 = anova(model3)
anova_table3
```

```
## Analysis of Variance Table
##
## Response: popularity
##             Df     Sum Sq    Mean Sq F value    Pr(>F)
## members      1 7.3945e+10 7.3945e+10  1269.8 < 2.2e-16 ***
## Residuals 5686 3.3111e+11 5.8233e+07
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
```

```
# POPULARITY x SCORED_BY
model4 = lm(popularity ~ scored_by, data = tv_anime)
anova_table4 = anova(model4)
anova_table4
```

```
## Analysis of Variance Table
##
## Response: popularity
##             Df     Sum Sq    Mean Sq F value    Pr(>F)
## scored_by    1 1.0499e+10 1.0499e+10  723.26 < 2.2e-16 ***
## Residuals 3395 4.9282e+10 1.4516e+07
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
```

Implications.

From the analysis, we can see that the most “popular” titles are not necessarily the most-scored titles, or even the highest-scored titles. For the most part though, many of the top 20 most popular titles appear in both the “top members”, “top scored”, and “top favorite” graphs, but are not generally seen in the same order, as expected. We can see clustering to the upper right on the “Popularity vs Score” graph, suggesting that popularity and score have a positive relationship. An obvious relationship can be observed between the “members” varible and

“scores”, as titles with more members would have more scorers; the ranking by these variables produce a more similar graph than other variables included in this study.

Limitations.

A large portion of this data remained unusable due to excessive missing values. Acquiring a revised dataset with less missing values, or a method for filling in these missing values with predicted values could really help improve the statistical backbone of this analysis. Combining this data with data from other websites similar to MyAnimeList could also assist with the large quantity of missing values. The formatting for several variables like genre, producer, and studio caused errors when attempting to apply them to the analysis, separating the entries, making genre, producer, and studio more versatile would allow for a more detailed and explanatory analysis.

Concluding Remarks.

Through this analysis, we have observed how popularity coincides with score, favorites, members, and scored_by. The “popularity” of an anime title does not seem to result from either one of our dependent variables, directly, though we did see that there are similarities between the most members and the most scored, which makes sense. Though the analysis is unable to outline a relationship between popularity and our independent variables, we still ended up with several recommendations for anime to watch!

Resources

www.MyAnimelist.com