Predicting which anime are animazing through quantitative analysis*

People are more likely to rate great anime, regardless of age

Allen Uy

April 15, 2024

This paper examines the top 10000 anime on MyAnimeList (MAL) by user score to understand the tendencies of highly rated anime. Using data provided by the MAL API, we discovered that the fraction of users that rate an anime to the total users that have engaged with the anime is a large indicator of user score. Our multiple linear regression model predicts the user score of an anime based on scoring fraction, date, and popularity and may indicate highly rated anime that have not yet gotten adequate exposure based on difference between the predicted and actual scores. This concept of scoring fraction is applicable to media forms beyond anime and may serve as the basis for recommendation algorithms.

Keywords: anime, regression, MyAnimeList, scoring fraction, rank, popularity, rating

1 Introduction (longer)

Anime refers to animation originating from Japan. (I want to introduce what anime is better). In 2021, half of Netflix's 222 million subscribers watched some anime on the platform and viewership saw an increase of 20% in the total hours over the previous year (Brzeski 2022). Anime is not much different compared to other forms of media, however anime and animation in general is still often regarded by many as 'childish', despite plenty of anime covering serious topics(Fiirgaard 2023).

MyAnimeList, herein referred to as MAL, is one of the largest and most popular anime databases and communities online. Users are able to keep track of anime they have watched or plan to watch, as well as rate and find similar anime. MAL is similar to IMDb in terms of functionality.

*Code and data are available at: https://github.com/varygx/AnimeAnalysis

The estimand of interest is the effect of

We conduct an analysis on the ratings of anime on MAL. Specifically we contrast

The subsequent sections follow a structured format. Section 2 outlines the source and variables of interest for our analysis. Section 3 details the construction and methodology of the statistical models used. Section 4 presents the key findings of our analysis, while Section 5 critically reviews the content, addresses the implications of the results, acknowledges model limitations, and suggests potential research directions.

2 Data

The data in this paper is sourced from the MAL API (*MyAnimeList API (Beta Ver.) (2)* 2024) and was gathered on April 12, 2024. The rankings of anime in the database are updated twice a day so the data gathered can be assumed to be up to date. Data analysis is performed in R (R Core Team 2023) with help from the following libraries: `tidyverse` (Wickham et al. 2019), `arrow` (Richardson et al. 2024), `rstanarm` (Goodrich et al. 2022), `modelsummary` (Arel-Bundock 2022), `testthat` (Wickham 2011), `here` (Müller 2020), `knitr` (Xie 2023), `kableExtra` (Zhu 2021), `dotenv` (Csárdi 2021), and `httr2` (Wickham 2023).

2.1 Measurement

In order to rate anime and add them to a list users must register an account on the MAL website. Each account can add an anime to their list once by choosing 1 of 5 statuses: watching, completed, on-hold, dropped, and plan to watch. It is also possible to remove an anime from the list if it was added by mistake. Every user that has an anime on their list is counted as a member of that anime, which is how the popularity rank is calculated. The user can also give a score to an anime, regardless of status, which is an integer from 1 to 10. The score of an anime on MAL is a weighted score calculated via the following:

$$\text{Weighted Score} = \left(\frac{v}{v+m}\right) \cdot S + \left(\frac{m}{v+m}\right) \cdot C$$

where:

Weighted Score is the calculated score for the anime/manga.

v is the number of users who have given a score for the anime/manga.

m is the minimum number of scored users required to get a calculated score.

S is the average score given by users to the anime/manga.

C is the mean score across the entire anime/manga database.

Table 1: Sample of Anime

id	title	rank	start_date	mean	popularity	num_list_users	num_scoring_users	fraction	days_since_start
52991	Sousou no Frieren	1	2023-09-29	9.39	295	681005	344608	0.5060286	196
5114	Fullmetal Alchemist: Brotherhood	2	2009-04-05	9.09	3	3333671	2113588	0.6340122	5486
9253	Steins;Gate	3	2011-04-06	9.07	13	2555269	1395401	0.5460877	4755
26017	Backkom Specials	9998	NA	5.90	15675	537	224	0.4171322	NA
38341	Bai Niao	9999	2017-06-13	5.90	14471	737	233	0.3161465	2495
31965	Balala Xiao Mo Xian: Qiji Wubu	10000	2013-04-19	5.90	12966	1128	291	0.2579787	4011

This equation requires a minimum number of scores to ensure a fair sample size, takes into account popularity, and normalizes the score compared to the mean score in the database. Note that for a sufficiently large number of scoring users the weighted score will be close to the actual mean score. In addition, MAL verifies whether the user has viewed 1/5 of the series upon completion and excludes scores by illegitimate accounts that try to sway votes. How this is done is not detailed. Due to the personal nature of these lists we assume that each user and list is associated with exactly one individual.

2.2 Top 10000 Ranked Anime

The MAL API is capable of fetching up to 500 anime based on ranking at a time for a given offset. Table 1 shows a small subset of the data gathered along with the main variables of interest. Relatively little cleaning was required as the API allows one to specify what fields should be returned. The genre was returned as a JSON array which was turned into a multi-hot encoding for data processing. The fraction column was calculated as the fraction of num_scoring_users to num_list_users, this is the fraction of users that gave a score to the anime. The days_since_start column was calculated as the integer day difference between the start date of the anime and the date the dataset was gathered.

2.2.1 Score Distribution

From the right skewed distribution of Figure 1, we observe that it is difficult for an anime to become top rated. This makes sense in the context of any subject that is reviewed as it requires a large group of people to consistently find greatness with no glaring flaws. It is also sensible that the mean score of the top 10000 anime is closer to 7 than to 5. Scores are inherently a subjective opinion and may take into account enjoyability, as well as how 'good' the piece of work is. Many people could enjoy a bad show even despite having many flaws. This matches the description MAL gives to each score number, where 1 is appalling, 5 is average, 7 is good, and 10 is a masterpiece. Seven is often seen as the average review score according to this forum discussion (see "What Is Closer to the Average Review Score for Video Games, 50/100 or 70/100?" (2019)).

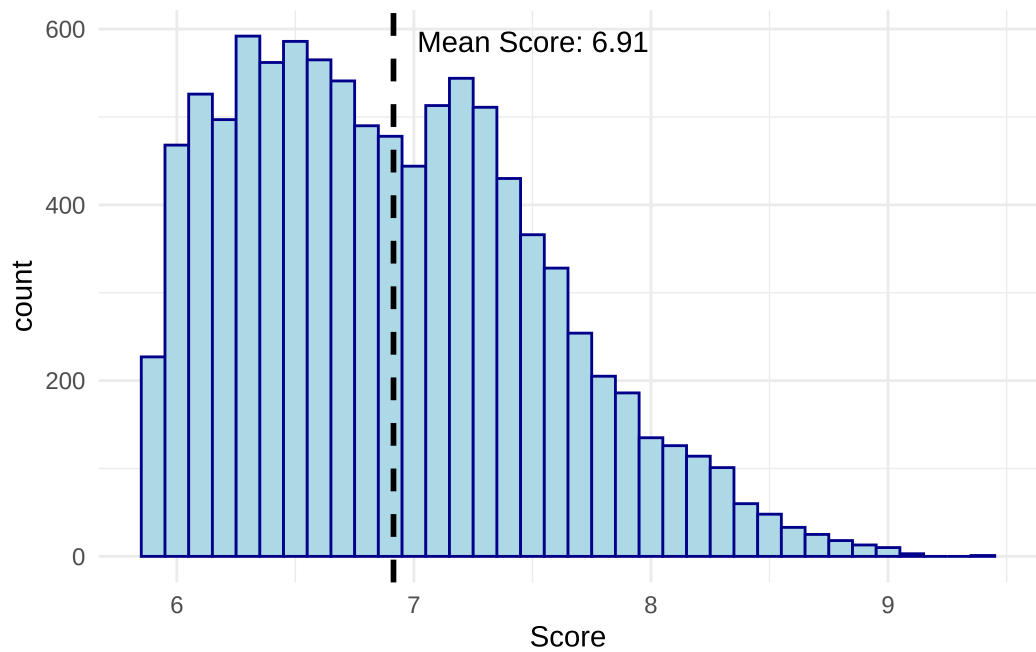


Figure 1: Distribution of Scores of Top 10000 Anime

2.2.2 Rank and Popularity

If scoring is subjective how much of a relationship is there between rating and popularity? From Figure 2 we see there is a moderately strong positive linear correlation between rank and popularity. This is also a sensible conclusion as a highly rated anime is more likely to attract new viewers.

2.2.3 Correlation to Date

It may seem like the date an anime airs can be a factor for popularity as older anime could have longer exposure over time or newer anime has more coverage but this does not seem to be the case. Figure 3a shows people are less likely to watch very old anime and Figure 4a suggests date is generally irrelevant to popularity but modern anime does have more viewership compared to older anime by the density of the points.

We observe a weaker relationship between date and rating in Figure 3b which suggests viewers are still appreciative of older works. As the amount of anime produced has increased in modern times, Figure 4b suggests that for every ‘good’ anime there is a ‘bad’ anime to match it.

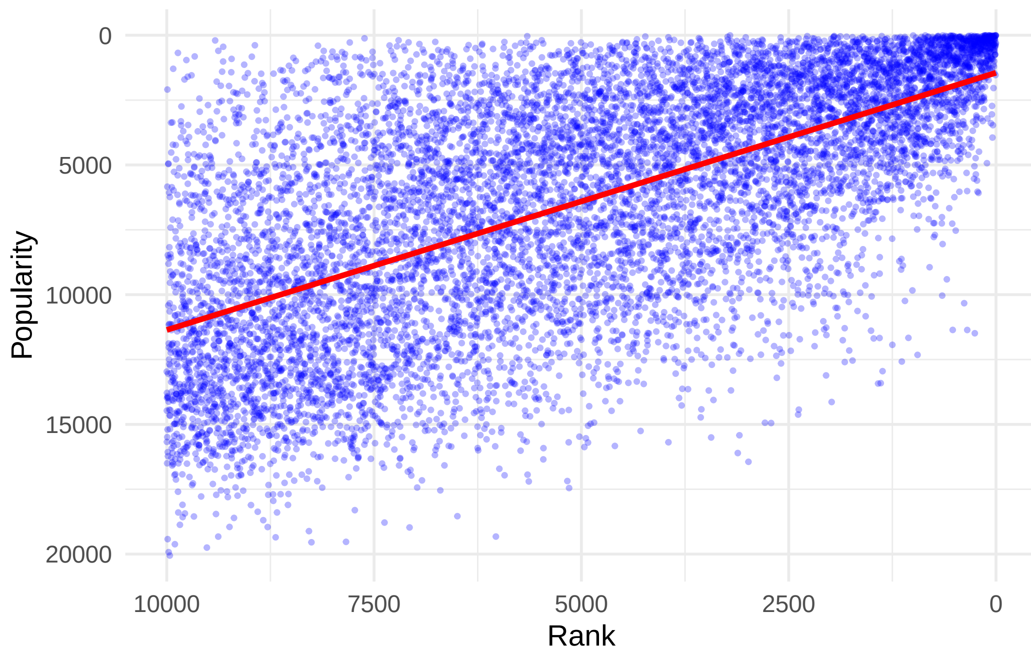
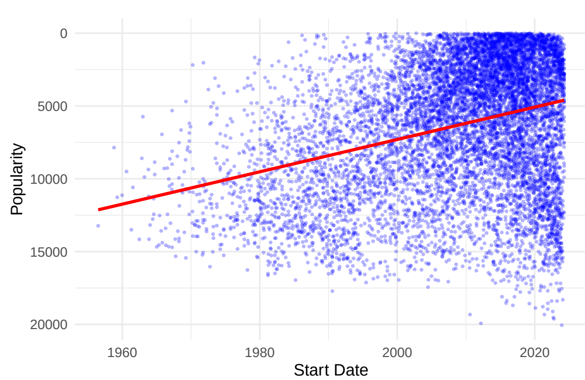


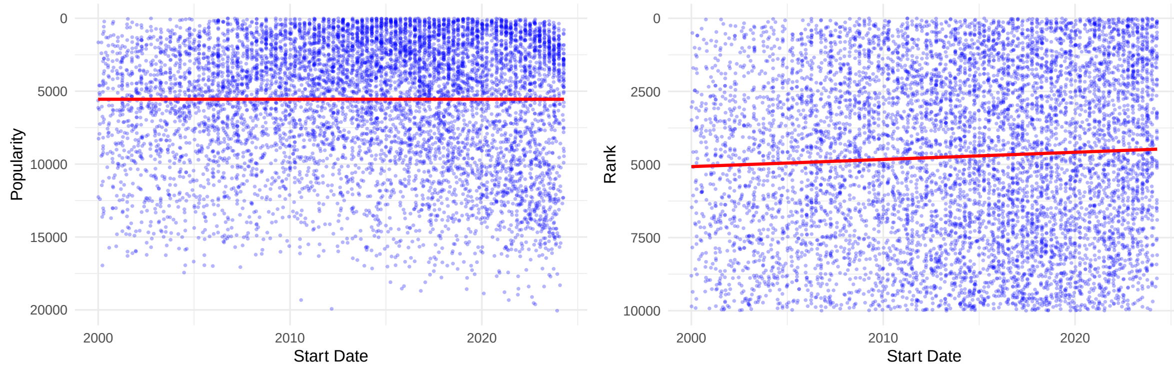
Figure 2: Rank vs Popularity for Top 10000 Anime



(a) Popularity plot

(b) Rank plot

Figure 3: Correlation to Date



(a) Popularity plot

(b) Rank plot

Figure 4: Correlation to Modern Dates

2.2.4 Scoring Fraction

On average, 39% of users end up rating an anime on their list according to Figure 5. There is a strong positive correlation between scoring fraction and popularity as seen in Figure 6a, which is sensible as fans of an anime or any media in general are likely express their opinion on that particular media. Surprisingly in Figure 6b there is a positive linear correlation between scoring fraction and rank. One might expect a normal distribution as people are more likely to voice their opinion on a particularly good or bad piece of work but not so much for average ones. Instead we see a linear relationship between scoring fraction and rank.

3 Model

Here we briefly describe the Bayesian analysis model used to investigate the user score of anime on MAL. Background details and diagnostics are included in Appendix B.

3.1 Model Set-up

From the preliminary analysis conducted in Section 2, we build a multiple linear regression model to predict the rating of an anime based on quantitative statistics. Define y_i as the user score of the anime i. Then fraction_i is the fraction of users who have rated the anime i to members of the anime i, popularity_i is the total members of anime i, and startDate_i is the number of days from the start date of anime i to the current date.

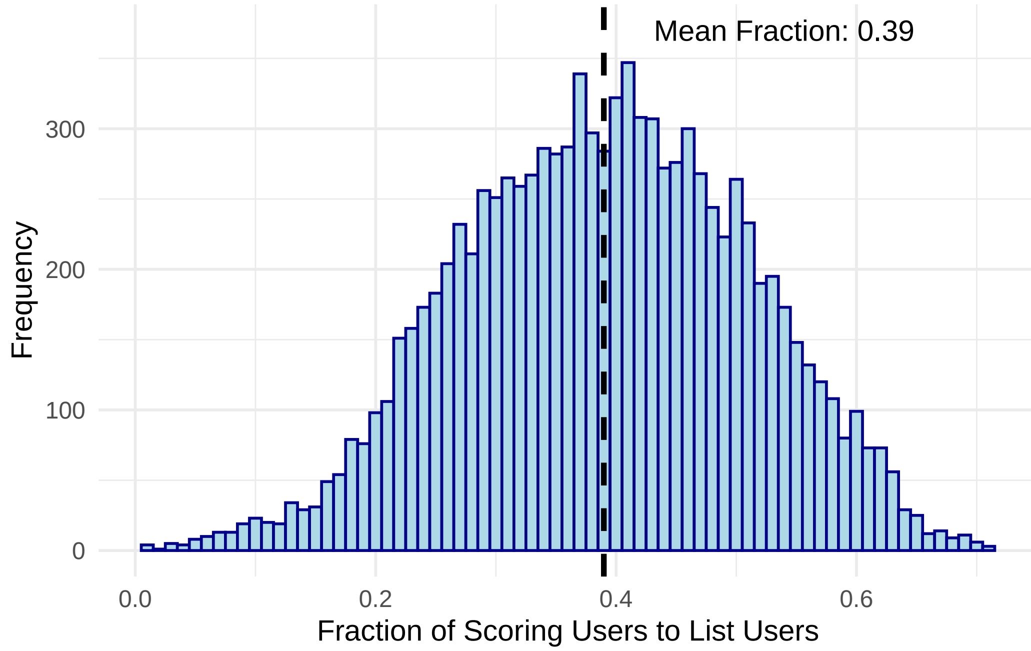
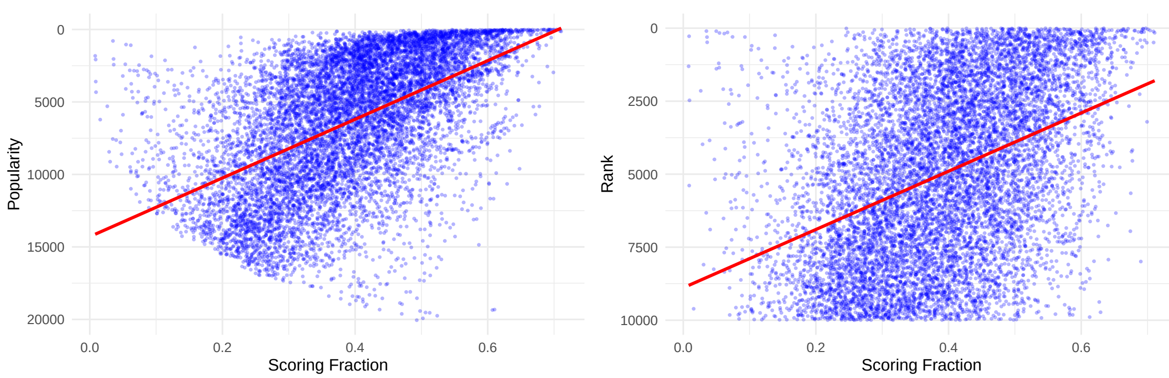


Figure 5: Distribution of Scoring Fraction of Top 10000 Anime



(a) Popularity plot

(b) Rank plot

Figure 6: Correlation to Scoring Fraction

$$
\begin{aligned}
y_i|\mu_i, \sigma &\sim \text{Normal}(\mu_i, \sigma) \\
\mu_i &= \beta_0 + \beta_1 \times \text{fraction} + \beta_2 \times \text{popularity} + \beta_3 \times \text{startDate} \\
\beta_0 &\sim \text{Normal}(7, 1.6) \\
\beta_1 &\sim \text{Normal}(1, 13) \\
\beta_2 &\sim \text{Normal}(0, 3.6 \times 10^{-4}) \\
\beta_3 &\sim \text{Normal}(0, 6.3 \times 10^{-6}) \\
\sigma &\sim \text{Exponential}(1.6)
\end{aligned}
$$

We run the model in R (R Core Team 2023) using the `rstanarm` package of Goodrich et al. (2022). We use Normal priors for the coefficients centered around sensible values which are detailed in the following section. We apply autoscaling in `rstanarm` as we are unsure of the likely range of these coefficients.

3.1.1 Model Justification

We expect the intercept of the model to be close to the mean score of anime so we center the prior distribution around it. We expect a positive relationship between the score of the anime and scoring fraction as seen in Section 2.2.4. This is likely to be the most important factor in our model and since the parameter is a fraction its coefficient should be relatively large, which is why we chose to center around 1. We expect a very minor to no relationship between the score of the anime and the start date of an anime as seen in Section 2.2.3. As such this coefficient is expected to be close to zero. Finally, we expect a minor positive relationship between the score of an anime and its popularity as seen in Section 2.2.2. This coefficient will likely be close to zero and smaller than start date, as an anime can have millions of members, but should be a better indicator than start date.

4 Results

Our results are summarized in Table 2. Our model observes a score intercept of 6.02, which means when an anime has been out for 0 days, has 0 members and thus 0 scoring fraction, it will likely have a user score of 6.34. This is slightly nonsensical as an anime had to have a positive scoring fraction and the minimum amount of scoring users to be ranked within the top 10000. When considering only the mean scoring fraction in Section 2.2.4 of 0.39, our model predicts a user score of $6.34 + 1.546 \times 0.39 = 6.94$, which is close to the mean user score of 6.91 seen in Section 2.2.1. The start date coefficient is extremely small indicating the date of an anime has no impact on its score. For a 10 year old anime the model expects the score to decrease by $-1.605 \times 10^{-5} \times (10 \times 365) = -0.06$, which can be considered a negligible amount. Finally, for every 1 million members of an anime, its score increases by $7.776 \times 10^{-7} \times 10^6 = 0.78$.

Table 2: Explanatory models of score based on scoring fraction, popularity, and start date

	(1)
(Intercept)	6.335
	(2.143×10^{-2})
fraction	1.546
	(5.106×10^{-2})
days_since_start	-1.602×10^{-5}
	(1.270×10^{-6})
num_list_users	7.776×10^{-7}
	(2.374×10^{-8})
Num.Obs.	9821
R2	0.276
R2 Adj.	0.275
Log.Lik.	−7977.065
ELPD	−7982.2
ELPD s.e.	69.0
LOOIC	15 964.5
LOOIC s.e.	138.0
WAIC	15 964.5
RMSE	0.55

Our model has an R2 of 0.276 meaning 27.6% of variance in user score is explained by effectively only scoring fraction and popularity, as only the oldest anime are affected by the start date coefficient. This is relatively impressive, considering the subjective nature of reviews, that user score can possibly be predicted purely by numbers.

Table 3 shows the predictions of our model compared to the actual scores of the top 5 and bottom 5 anime in the dataset. Some observations that should be noted are:

1. The score intercept is higher than the score of lower ranked anime
2. The model harsh on highly ranked anime that are not yet popular

5 Discussion

5.1 Key Findings

From the visualizations in Section 2 we discover 3 key findings:

1. There exists a positive relationship between rank and popularity.
2. Except for the oldest anime, anime is popular regardless of age.

Table 3: Comparing predictions of model to actual scores

title	rank	popularity	num_list_users	fraction	start_date	days_since_start	mean	predicted
Sousou no Frieren	1	295	681005	0.5060286	2023-09-29	196	9.39	7.643890
Fullmetal Alchemist: Brotherhood	2	3	3333671	0.6340122	2009-04-05	5486	9.09	9.819707
Steins;Gate	3	13	2555269	0.5460877	2011-04-06	4755	9.07	9.090224
Gintama°	4	341	628518	0.3982623	2015-04-08	3292	9.06	7.386914
Shingeki no Kyojin Season 3 Part 2	5	21	2265274	0.6973669	2019-04-29	1810	9.05	9.145720
Zou no Inai Doubutsuen	9995	16500	452	0.2831858	1982-03-20	15364	5.90	6.527286
Akai Koudan Zillion Recaps	9996	13984	854	0.2177986	1987-07-21	13415	5.90	6.457752
Atagoal wa Neko no Mori	9997	13785	897	0.2441472	2006-10-14	6390	5.90	6.611038
Bai Niao	9999	14471	737	0.3161465	2017-06-13	2495	5.90	6.784590
Balala Xiao Mo Xian: Qiji Wubu	10000	12966	1128	0.2579787	2013-04-19	4011	5.90	6.670703

3. The fraction of people that score an anime to the total members of an anime is a strong indicatowwwwwwwwwwr of rating.

Through a multiple linear regression model built in Section 3, we find that the base anime is given a score of 6.34. Every percentage point in the scoring fraction improves the score by 0.01546. For every day an anime has been released its score decreases by -1.605×10^{-5}, which is only relevant for the oldest anime on the database. For every 1 million members of an anime its score increases by 0.7776.

The correlation found in Section 2.2.4 is surprising as it is contrary to human psychology. People are more likely to remember negative experiences so one might expect an increase in scoring fraction for low quality anime Harting (2022). However we find the opposite on MAL, suggesting people are less likely to rate low quality anime, perhaps as a way to forget about the show.

5.2 Applications

When comparing the predicted scores of the model and the actual user scores on MAL we notice that the model performs poorly for highly rated but not yet popular anime in terms of total member count. This matches the rank-popularity relation discussed in Section 2.2.2. This observation leads to the potential usage of the model as an indicator of which anime are not popular enough relative to their ranking.

For instance, if we look at Sousou no Frieren in Table 2 we see it is by far the highest rated anime on MAL. One could assume that the main reason the anime is not yet popular in terms of total member count is due to airing only recently. The difference between the predicted score and the actual score can indicate an anime should get more viewership relative to its quality. This scoring could be used in a recommendation system to highlight anime that may be deserving of viewership.

5.3 Weaknesses and Next steps

There is some noise regarding the relationship between rating and popularity. How much of popularity is due to rating and how much of rating is due to popularity are areas of interest. We assume that each user on MAL is a unique individual due to the personal nature of lists. It would not generally serve much purpose for a person to keep track of separate lists. However in the past there have been issues of people attempting to artificially lower or raise the scores of some anime "MAL Address Changes Made to Combat "Illegitimate" (Duplicate) Accounts & Vote Brigading" (2020). Since then MAL has attempted to police this issue and we do not believe it should be a concern in this dataset.

Three fields of interest were unable to be retrieved without an API call for a specific id, namely related anime, recommendations, and statistics regarding the exact numbers of member status. When attempting to get this data, rate limiting was encountered at roughly 250 calls per 10 minutes, with no clear way to raise the API call limit. Due to this the relationships between anime could not investigated thoroughly. Analysis of a graph structure could prove interesting, especially in the context of recommendations and series of anime.

A variable that would likely be helpful in our model is drop rate, which can be calculated as the fraction of users that label an anime as dropped compared to the total members of that anime. As mentioned previously this variable is capable of being retrieved from the API but would take a long time to gather. We believe that higher ranked anime are likely to correspond to lower drop rate.

An analysis on scoring fraction should be done on IMDb if applicable to verify whether the assumptions made in this paper apply to other forms of media beyond anime.

Appendix

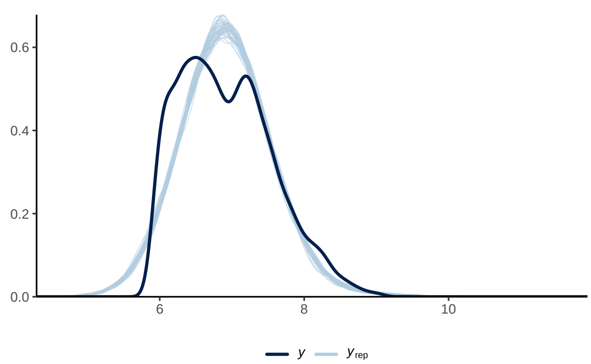
A Additional data details

B Model details

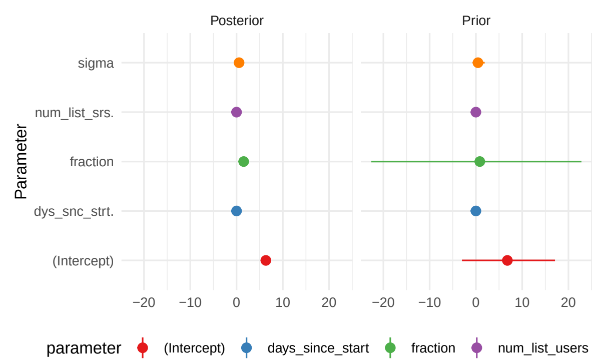
B.1 Posterior predictive check

In Figure 7a we implement a posterior predictive check. This compares the actual scores of anime with simulations from the posterior distribution. The solid black line is the score distribution seen in Section 2.2.1 and we see that the model fits the data relatively well for only being capable of making a normally distributed posterior.

In Figure 7b we compare the posterior with the prior. This show how much the estimates change once data are taken into account. We see that the posterior coefficients are close the prior, indicating a good choice of priors.



(a) Posterior prediction check

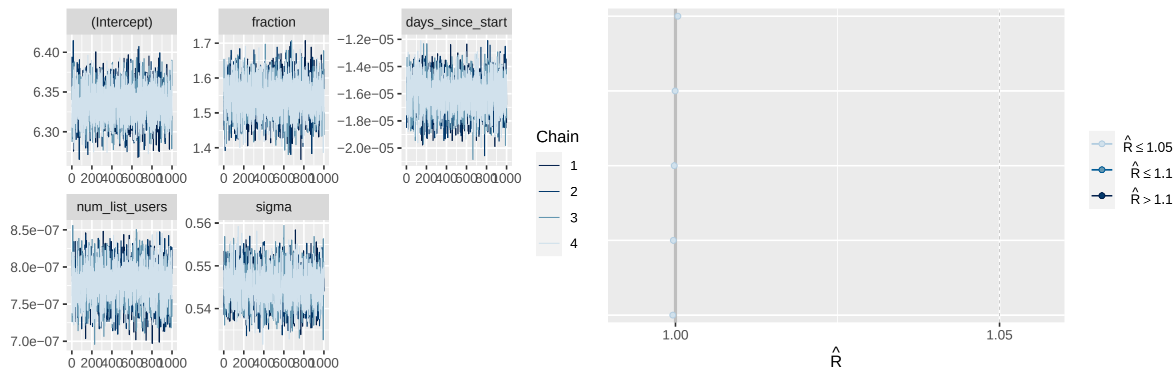


(b) Comparing the posterior with the prior

Figure 7: Examining how the model fits, and is affected by, the data

B.2 Diagnostics

Figure 8a and Figure 8b are trace and Rhat plots of our model. We see horizontal lines that bounce around and have overlap between the chains in the trace plot and values close to 1 in the Rhat plot. This suggests our model ran into no problems.



(a) Trace plot (b) Rhat plot

Figure 8: Checking the convergence of the MCMC algorithm

C Datasheet

Link to datasheet in repo here

D R package

TODO: create package

References

Arel-Bundock, Vincent. 2022. "modelsummary: Data and Model Summaries in R." *Journal of Statistical Software* 103 (1): 1–23. https://doi.org/10.18637/jss.v103.i01.

Brzeski, Patrick. 2022. "Netflix's Head of Anime Says Half of Global Subscribers Watch Japanese Animation, Bullish on Growth (Exclusive)." https://www.hollywoodreporter.com/business/business-news/netflix-anime-kohei-obara-interview-1235124676/.

Csárdi, Gábor. 2021. *Dotenv: Load Environment Variables from '.env'*. https://CRAN.R-project.org/package=dotenv.

Fiirgaard, Milie. 2023. "Why Are Adults Not Allowed to Enjoy Animation?" *Meer*. https://www.meer.com/en/71903-why-are-adults-not-allowed-to-enjoy-animation.

Goodrich, Ben, Jonah Gabry, Imad Ali, and Sam Brilleman. 2022. "Rstanarm: Bayesian Applied Regression Modeling via Stan." https://mc-stan.org/rstanarm/.

Harting, Caroline. 2022. "How Can We Break the Cycle of Focusing on Negative Experiences?" *Columbia News*. https://news.columbia.edu/news/how-can-we-break-cycle-focusing-negative-experiences#:~:text=However%2C%20research%20shows%20that%20humans,impediments%2C%20like%20depression%20or%20anxiety.

"MAL Address Changes Made to Combat "Illegitimate" (Duplicate) Accounts & Vote Brigading." 2020. MyAnimeList Co.,Ltd. https://www.reddit.com/r/anime/comments/f2ourd/mal_address_changes_made_to_combat_illegitimate/.

Müller, Kirill. 2020. *Here: A Simpler Way to Find Your Files*. https://CRAN.R-project.org/package=here.

MyAnimeList API (Beta Ver.) (2). 2024. MyAnimeList Co.,Ltd.

R Core Team. 2023. *R: A Language and Environment for Statistical Computing*. Vienna, Austria: R Foundation for Statistical Computing. https://www.R-project.org/.

Richardson, Neal, Ian Cook, Nic Crane, Dewey Dunnington, Romain François, Jonathan Keane, Dragoș Moldovan-Grünfeld, Jeroen Ooms, Jacob Wujciak-Jens, and Apache Arrow. 2024. *Arrow: Integration to 'Apache' 'Arrow'*. https://CRAN.R-project.org/package=arrow.

"What Is Closer to the Average Review Score for Video Games, 50/100 or 70/100?" 2019. *GameFAQs*. https://gamefaqs.gamespot.com/boards/691087-playstation-4/77877326.

Wickham, Hadley. 2011. "Testthat: Get Started with Testing." *The R Journal* 3: 5–10. https://journal.r-project.org/archive/2011-1/RJournal_2011-1_Wickham.pdf.

———. 2023. *Httr2: Perform HTTP Requests and Process the Responses*. https://CRAN.R-project.org/package=httr2.

Wickham, Hadley, Mara Averick, Jennifer Bryan, Winston Chang, Lucy D'Agostino McGowan, Romain François, Garrett Grolemund, et al. 2019. "Welcome to the tidyverse." *Journal of Open Source Software* 4 (43): 1686. https://doi.org/10.21105/joss.01686.

Xie, Yihui. 2023. *Knitr: A General-Purpose Package for Dynamic Report Generation in r*. https://yihui.org/knitr/.

Zhu, Hao. 2021. *kableExtra: Construct Complex Table with 'Kable' and Pipe Syntax*. https://CRAN.R-project.org/package=kableExtra.