Star Trek is a Science Fiction media franchise spanning several TV shows, movies, books, and video games. This analysis focused on four Star Trek TV shows that ran in sequence from 1987 through 2005, with significant overlaps in the years they each ran. The original set of TV shows ran with an episodic formula, with almost each episode having a self contained plot. As with TV shows of this nature, there is a core set of characters, along some recurring characters and one-off characters. There were also many directors involved, each directing a varying number of episodes. What is interesting is that the audiences are not likely to favour all core cast members equally, neither are they likely to prefer the work of all directors equally.

Checkout report.pdf for the full analysis.

This work attempts to analyze this and test this hypothesis, while also trying to identify the most and least liked characters and directors.
The entire work was done using R v4.1.1.

Data

- IMDb Ratings and Episode information from https://datasets.imdbws.com (collected on Nov 23, 2021). This formed the source for the tabular data used in the analysis.
- Script Data from www.chakoteya.net/StarTrek/. This was the source for the script of each episode in text form. The text was used to calculate a proxy for each character's relative screentime in each episode, the proxy being proportion of words spoken by a chraracter in an episode. This proxy variable was then merged with the tabular data.
- Four TV Shows:
  - Star Trek: The Next Generation (1987–1994)
  - Star Trek: Deep Space 9 (1993–1999)
  - Star Trek: Voyager (1995–2001)
  - Star Trek: Enterprise (2001–2005)
- 614 Episodes in all
- Relevant Variables:
  - Name of Director(s)
  - Name of writer(s)
  - Episode Rating
  - Proxy Character Screentimes for main characters (33 in all)

Methodology ⚒

Data Extraction and Cleaning

This step involved cleaning up the script data, creating the proxy variable, and accurately matching it with the IMDb source data. This ste also involved cleaning up inconcsistencies between data from different shows and extracting relevant items from different datasets to create our final tabular dataset.
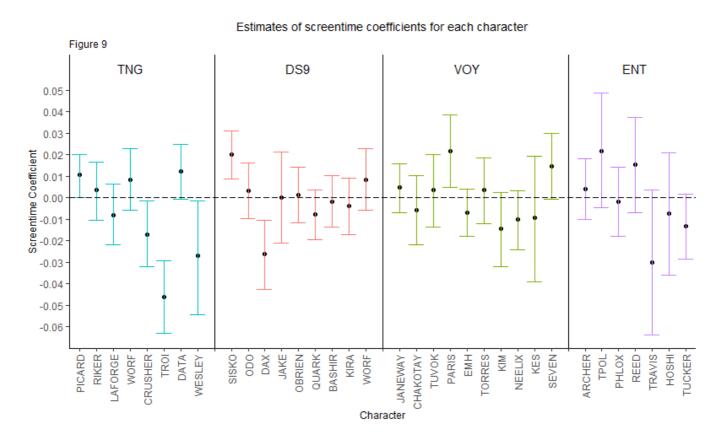
Statistical Modeling

The main idea behind our statistical methodology is that we create a proxy screentime variable for each character and quantify the character's likeability as the regression coefficient of their screentime on IMDb ratings. We use an heirarchical regression methodology as this allows us to adjust for general quality of shows and seasons while also allowing for the "director" variable. We cannot treat the "director" variable as a regular

one hot encoded variable because of the large number of directors involved, with many of them directing only a single episode. We formulated our regression as:

- Target variable: IMDB rating
- Unit of observation: Episode
- Fixed Effects: Character Screentimes
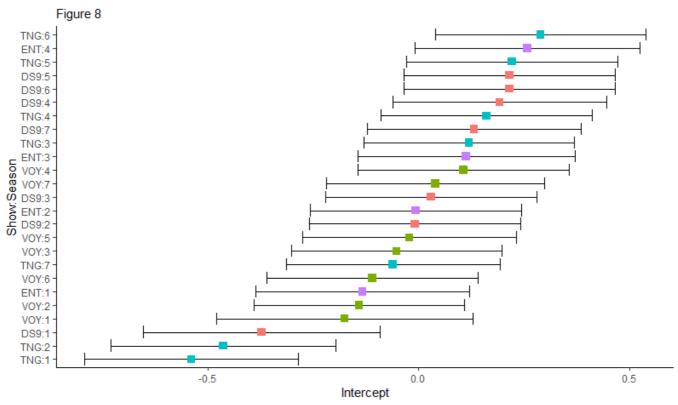- Random Effects: Show:Season combination, Name of director(s)

Results 📊

Influence of Characters

Figure 9

- The figure illustrates the estimated average change in episode rating when a character's screentime is increased by 1 percentage point. This increase is at the expense of the screentime of "other" characters i.e. any character not included in the main character list. Corresponding 95% confidence intervals are also illustrated.
- Most characters were not observed to have a significant impact on episode ratings.
- Screentimes of Sisko from Star Trek: Deep Space 9 and Paris from Star Trek: Voyager have the strongest positive impact on episode ratings.
- Screentimes of Dr. Crusher, Troi, and Wesley from Star Trek: The Next Generation, and Dax from Star Trek: Deep Space 9 have the strongest negative impact on the ratings.

Differences across Shows and Seasons

Random Effects for Show:Seasons

Figure 8

- We observed significant difference in quality across Show:Seasons. Season 1 of The Next Generations was the worst while Season 6 of The Next Generations was the best

We found no evidence for difference in quality across directors. This is likely due to the small number of episodes per director.

Conclusions

We observe that audiences appear to show a clear preference for some characters over the others. We can also extend this analysis to other TV shows, one that I have in mind is the popular sitcom Friends, which I feel could be even better suited to this type of analysis.

Limitaions

There are a few limitations to this study that need to be taken into consideration.

- While the hierarchical structure of the model helps in taking care of dependence in residuals within seasons, it is still possible that a few of the multi-part episodes have non independent residuals. This is a violation of the assumptions of linear regression.
- The screentime is calculated as proportion of words spoken by characters. The source data is not fully clean as a few whitespace issues may have fused words together and reduced counts. Incidence of this issue is not very common and is completely at random, and thus should only have a minor influence on the results.
- IMDb calculates ratings as a weighted mean from individual user ratings, with the weight calculation algorithm being a secret. While the weighting scheme is in place to prevent manipulation, it leaves an uncertainty as to exactly what our target variable is.

- Characters change and evolve through a show's natural progression, and thus their effect on episode ratings is likely to be variable between seasons. This effect can be modeled using a random slopes model however the multifold increase in number of variables makes the modeling difficult. Thus we are forced to assume that each character has a uniform linear effect on episode ratings.
- We did not perform post-hoc analysis on our regression estimates. Due to the large number of statistical tests done, we are likelier to commit Type I errors and misrepresent the confidence bands.