Anime Sources

Daniel Soto (ダニエル), Alnez (アルネズ), Devin Ruiz（デヴィン)

October 2021

Abstract

As the popularity of Japanese production increases worldwide, more fanatics like ourselves that over analyze their statistics grows. The three of us have a deep interest in Japanese animation as a whole and are interested in how the industry would develop. Anime is mainly Japanese animation, although recently China, Korea, and America have contributed to the industry in the past decade. Anime can either be original masterpieces or adaptations from different source materials; for instance: manga, light novel, visual novel, novel, game, book, 4-koma manga (online), etc. In other words there is a substantial amount of source material that can potentially be adapted into a anime by any studio. Due to the increase of popularity in the anime industry worldwide, studios must first ask themselves what type of product would sell in the market. With the information we had to work with, we have two main questions that we want to answer:

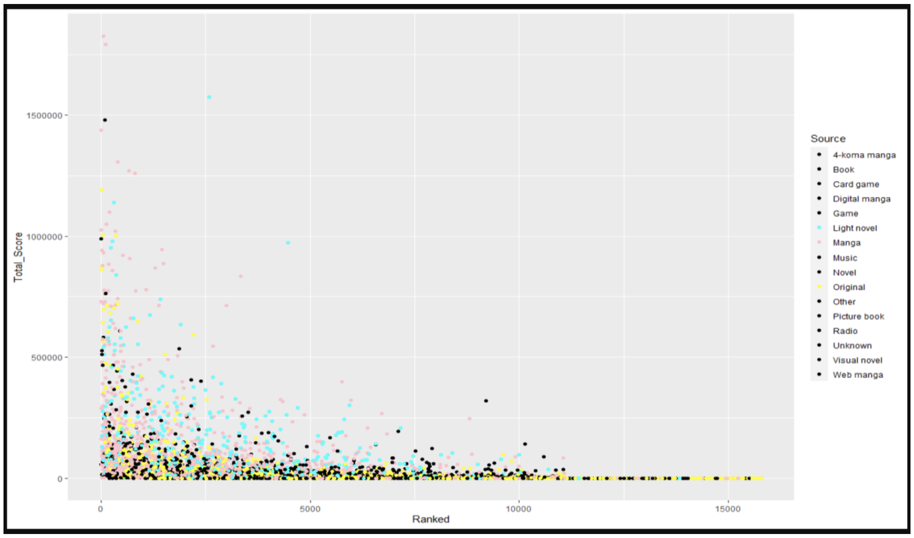
1. What anime performed best depending on the source?

2. What anime performed best depending on genre?

To tackle the problems above, the team had to be able to see trends over the years in the history of anime. Through this process, we will be able to analyze anime trends throughout the years. Given the data, we found an increase of adaptations from specific sources and a trend of popularity with different genres. The data set is based from the website My Anime List. We were able to analyze and attempt the use of a time series analysis method to find a correlation of any given show's genre or the source they were adapted from. Included in the data set are all the information from user participants.

1 Introduction

Within the years 2000-2020, the anime industry became unprecedentedly more popular through each passing year. For this reason anime studios have began to flood the market with high a quantity of works. Currently the industry is experiencing a high influx of anime works being produced in our present day. Due to the amount of anime throughout the year there has been some genres that have been over saturated and by consequence produced poorly made adaptations. In other words due to the over saturation of a particular anime genre, the industry is pumping out the shows quickly at the cost of the quality which is potentially costing the studios more revenue.



The graph above represents the lack of variety of adapted sources within the anime industry. There is a potential loss in profitable creative works. This further emphasizes the need for a larger variety of adapted source materials.

The data set contains high amount of data for over 17,000 anime works. After taking a look at our data, the team notice we will need to reformat some of the columns in order to be used. One example was the premier column in which we had the seasons and year into one. This column contains valuable data in which we can separate the column into two new columns respectively.

Premiered	Season	Year
Spring 1998	Spring	2009
Unknown	Winter	2021
Spring 1998	Spring	2011
Summer 2002	Fall	2011
Fall 2004	Spring	2015
Spring 2005	Spring	2019
Spring 2005	Spring	2011
Fall 2002	Unknown	NA
Spring 2004	Fall	2012
Spring 2004	Unknown	NA
Fall 2002	Fall	2017
Fall 1999	Winter	2017
Fall 2001	Spring	2006
Fall 2004	Fall	2008
Fall 2004	Unknown	NA
Fall 2004	Unknown	NA
Spring 2003	Summer	2017

Using the same process to separate the premier column, the team applied said process to separate the genre column. The genre column turned out to be the most difficulty to work with due to implication of multiple genres for each anime title.

Genres
Action, Adventure, Comedy, Drama, Sci-Fi, Space
Action, Drama, Mystery, Sci-Fi, Space
Action, Sci-Fi, Adventure, Comedy, Drama, Shounen
Action, Mystery, Police, Supernatural, Drama, Magic
Adventure, Fantasy, Shounen, Supernatural
Action, Sports, Comedy, Shounen
Comedy, Drama, Josei, Romance, Slice of Life
Slice of Life, Comedy, Sports, Shounen
Action, Cars, Sports, Drama, Seinen
Drama, Horror, Mystery, Police, Psychological, Seinen,...
Action, Adventure, Comedy, Super Power, Martial Arts...
Action, Adventure, Comedy, Super Power, Drama, Fan...
Action, Comedy, Sports, School, Shounen
Action, Shounen, Sports
Comedy, Romance, School, Shounen
Action, Adventure, Comedy, Ecchi, Sci-Fi, Shounen
Action, Sci-Fi, Psychological, Drama

Aside from the having multiple genres, not all titles would have the same amount of genres. As seen on the image above, the first title has six genres as its classification. While the second one has only five genres. The amount of genres for each title varies throughout the data-frame.For this reason when the team split the columns to individual genres it resulted the creation of multiple null values.

Genre2	Genre3	Genre4	Genre5	Genre6	Genre7	Genre8
Adventure	Comedy	Drama	Magic	Fantasy	Shounen	Not Available
Mystery	Super Power	Drama	Fantasy	Shounen	Not Available	Not Available
Not Available	Not Available	Not Available	Not Available	Not Available	Not Available	Not Available
Fantasy	Shounen	Super Power	Not Available	Not Available	Not Available	Not Available
Historical	Parody	Samurai	Sci-Fi	Shounen	Not Available	Not Available
Fantasy	Military	Mystery	Shounen	Super Power	Not Available	Not Available
Comedy	Historical	Parody	Samurai	Shounen	Not Available	Not Available
Space	Drama	Not Available	Not Available	Not Available	Not Available	Not Available
Historical	Parody	Samurai	Sci-Fi	Shounen	Not Available	Not Available
Shounen	Not Available	Not Available	Not Available	Not Available	Not Available	Not Available
Seinen	Slice of Life	Not Available	Not Available	Not Available	Not Available	Not Available
Historical	Parody	Samurai	Sci-Fi	Shounen	Not Available	Not Available
Historical	Parody	Samurai	Sci-Fi	Shounen	Not Available	Not Available
Supernatural	Drama	Romance	Not Available	Not Available	Not Available	Not Available
Comedy	Historical	Parody	Samurai	Shounen	Not Available	Not Available
School	Drama	Not Available	Not Available	Not Available	Not Available	Not Available
Supernatural	Vampire	Not Available	Not Available	Not Available	Not Available	Not Available

The image above is the result of splitting the genre column.

The data-frame also contained columns in which they represent the amount of participants in which the users rated the the title in a 10 point rating system.

Score.10	Score.9	Score.8	Score.7	Score.6	Score.5	Score.4	Score.3
229170.0	182126.0	131625.0	62330.0	20688.0	8904.0	3184.0	1357.0
30043.0	49201.0	49505.0	22632.0	5805.0	1877.0	577.0	221.0
50229.0	75651.0	86142.0	49432.0	15376.0	5838.0	1965.0	664.0
2182.0	4806.0	10128.0	11618.0	5709.0	2920.0	1083.0	353.0
312.0	529.0	1242.0	1713.0	1068.0	634.0	265.0	83.0
9226.0	14904.0	22811.0	16734.0	6206.0	2621.0	795.0	336.0
11829.0	16309.0	20008.0	13062.0	5574.0	3148.0	1339.0	484.0
1123.0	1777.0	3102.0	3075.0	1286.0	602.0	218.0	88.0
10948.0	15820.0	22379.0	12912.0	3874.0	1236.0	369.0	97.0
77350.0	60652.0	43459.0	22045.0	8861.0	4381.0	2086.0	882.0
216866.0	234481.0	345563.0	286175.0	108155.0	46886.0	15477.0	6098.0
292445.0	166186.0	141755.0	85424.0	35342.0	19019.0	8201.0	3675.0
9576.0	12719.0	19791.0	16284.0	6624.0	3086.0	1071.0	443.0
69.0	61.0	152.0	260.0	236.0	143.0	88.0	40.0
17907.0	27623.0	38112.0	27327.0	11034.0	5396.0	1897.0	780.0
3589.0	6664.0	12349.0	12702.0	6033.0	3079.0	1521.0	643.0
6856.0	8309.0	10851.0	8636.0	4429.0	2462.0	1456.0	684.0

These columns are important due to the representation of the amount of users who have watched an anime work. The team then decided to have the total participants of a anime title in a new column. This new column would be called "Total Participants" and it will be a culmination of all other score columns added together in our data-frame for a particular anime work.

Total.Participants
1438767
288274
989905
1026866
161812
728435
162866
58655
113662
940843
131218
93327
289115
527385
97673
1190759
127875

We aim to use these new columns in conjunction with the rest of the data-frame in order to analyze and identify trends to answer the questions mention in our abstract.

2 Data Description

The purpose of our data is to see a correlation with the genre of an anime and the sources they were adapted from throughout the years. All of the data we worked with was attained from the website MyAnimeList which contains 17,562 unique anime works. The data set also contains a variety info for each anime work which includes name, score, genres, type of media, episodes, premiered, producers, studio, source material, score, rating, and many more. However, after analyzing the provided data we notice some discrepancies in score, premiered, and genres where there were missing values. There were also anime works in our data that did not contain enough information to provided the needed analysis we are striving to achieve. It also turned out that the data contained anime works which haven't premiered yet, normally this isn't an issue but we saw these works were being given ratings and receiving score from the public audience. We concluded that there was a high likelihood that these ratings and scores of unreleased anime works were being rated with bias from users who likely read the original source material. Another possibility is the pre-premier of the anime to selected audience for feedback.

After some consideration, we revised some of the columns and entries into a new data-frame that contained five significant columns: genre, year, ranked, source, and total participants. This new data-frame gave us the necessary anime works needed to do our research without much difficulty from its previous version. We also ran into the issue in which plotting genres cause havoc in our graphs due to high amount genre classification each anime contained. To counteract with this issue without losing important data we generated a new data-frame that only contained one specific genre. This lead to the creation of 40 different data-frames, which allowed the team to work with the data more efficiently. After making the cuts and reclassifying some columns, we finally established a data-frame to begin our analysis. The team mainly used the years provided for each show to categorize the total participants score to any specific genre and its adapted source. We noticed the total amount of user participation each anime generated (what score they gave to any given show) was correlated to the ranking the anime was placed overall. We tried to implement random sampling algorithms into our analysis, but due to the nature of our data, but unfortunately they weren't effective. In more detail, the majority of our data was categorical not numeric. Thus, numeric random sampling functions just didn't work even if we changed the categorical variables to numeric.

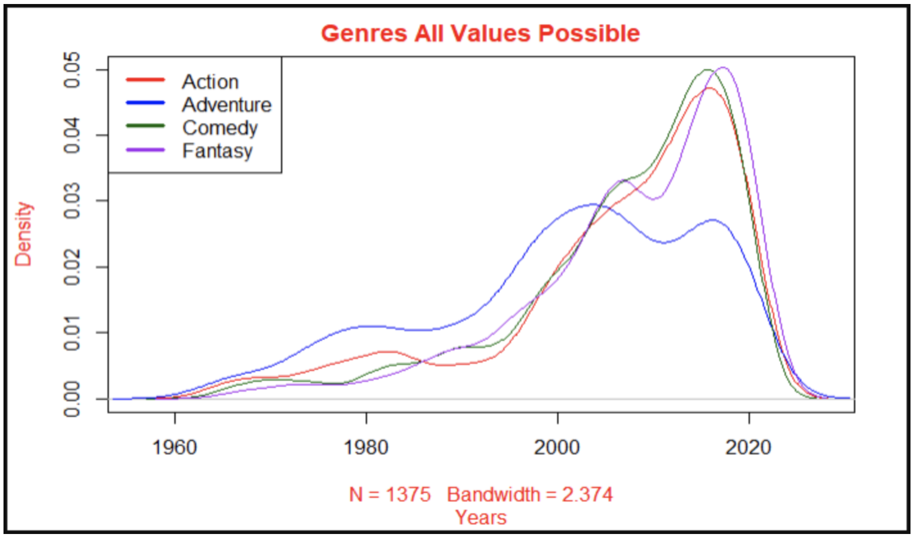
3 Graphical Analysis

In order to observe our data and find possible correlations we used plots and graphs in our data-frame. We revised our data several times and generated plots through RStudio to demonstrate our graphical analysis. In consequence to the way the data-frame is formatted, the most accurate types of graphs for our data are bar and density graphs. The team noticed that the earliest year recorded for any particular show was 1963, unfortunately this also reduced the amount of shows calculated in the data. That's not to say that the shows not included are *only* older than 1963, but it's more likely the case of poor documentation of the shows aired date in MyAnimeList.

```
Only_Comedy  6029 obs. of 5 variables
```

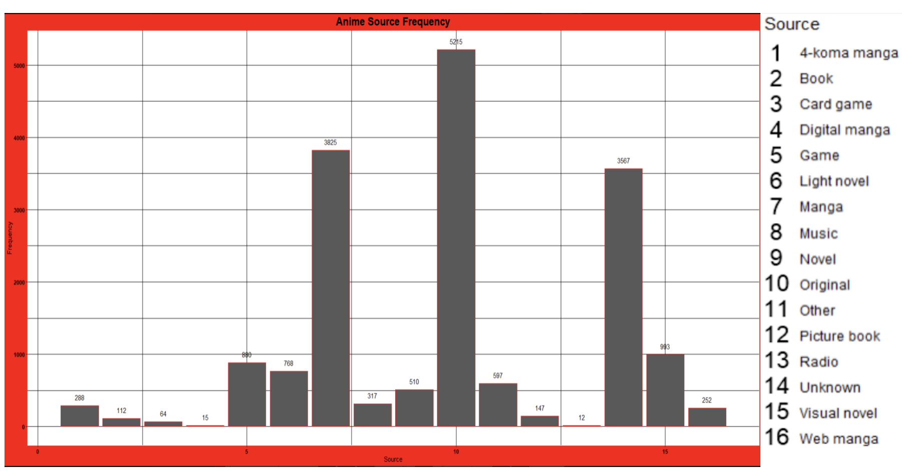
```
Only_Comedy  2297 obs. of 5 variables
```

The graph below shows a density chart of a specific genre throughout the years, the team produced several of these graphs from the 40 different genre specific data-frames. We created them to see the frequency of a genre throughout the years so that we could attain the necessary information to produce an educated guess in what types of genres would be better to create in the coming years.

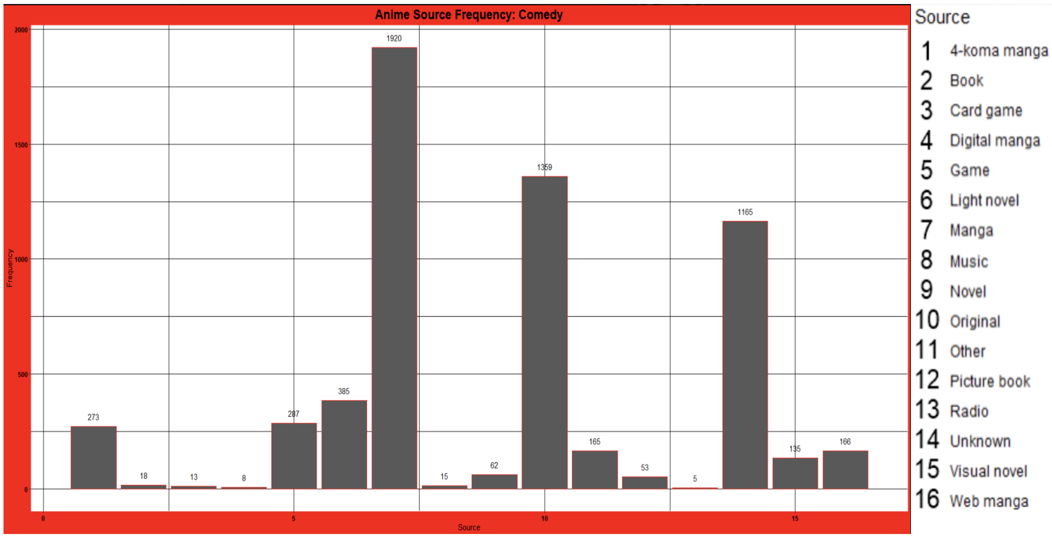


We interpreted the graphs as probabilities of specific genres throughout the years. Take the action genre in the above graph. At the start of 2000, the frequency of action genre began to steadily rise up to 2020. Now the reason why after 2020 the genres frequency dips is due to not many animes haven't released yet passed 2020.

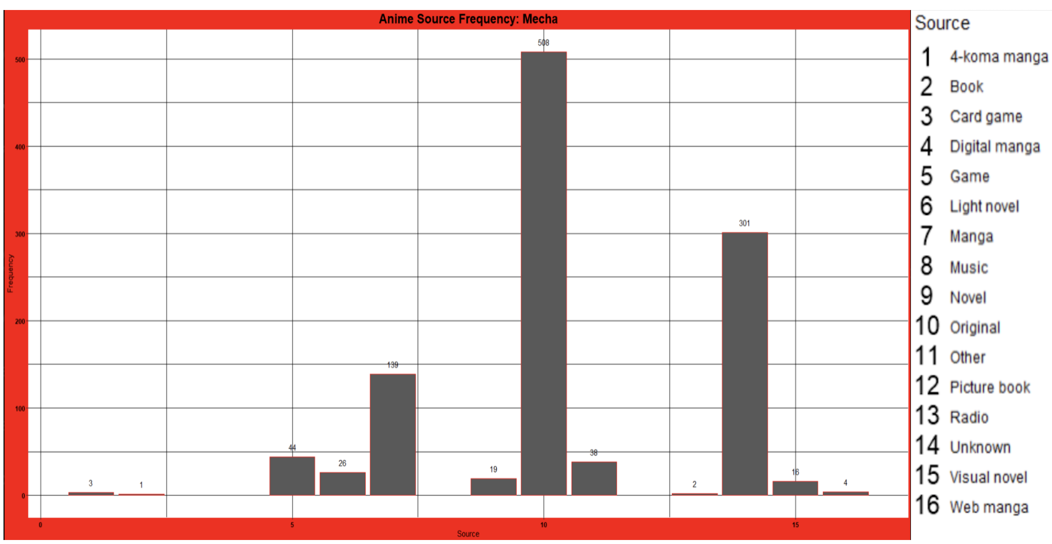
The graph below represents the frequency of adapted sources overall, without the filtered out years below 1963. We created it to get a broad view into how common the adapted sources were in our data set.



The graph below represents the frequency of adapted sources for the data-frame containing only comedy. We created it to get a broad view into how common the adapted sources were in a data-frame we have of a specific genre (large).



The graph below represents the frequency of adapted sources for the data-frame containing only mecha. We created it to get a broad view into how common the adapted sources were in a data-frame we have of a specific genre (small).



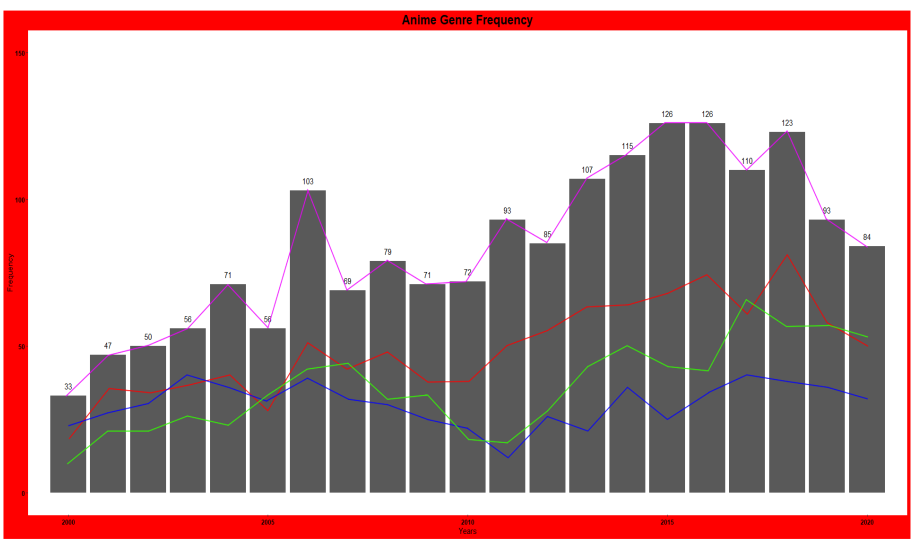
From these graphs, we observe that studios tend to adapt from manga or original. This is consistent throughout in our data-frame, regardless of the era or genre. As time passes, the other sources tend to increase slightly.

4 Analysis

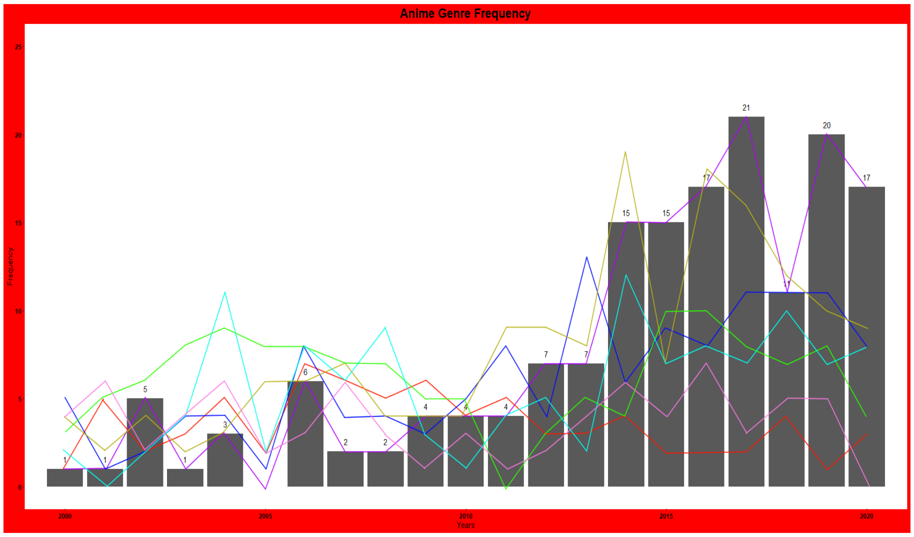
After we set up our data-frame, we began to take a look at anime's prominent years of growth which were from 2000-2020. From here, we took a look at all genres and observed the amount of animes that released per genre for each year. Once we had the tables set up in Excel, the team took the average of each genre from 2000-2009 and 2010-2020. The table below demonstrate a portion of said Excel table.

Years	Action	Adventure	Comedy	Fantasy	Slice of Life	Shounen	Sci-Fi	School	Magic	Romance	Mecha	Supernatural	Seinen	Kids	Sports	Shoujo	Harem
2000	18	23	33	10	3	16	21	2	6	7	13	8	1	6	1	5	2
2001	35	27	47	21	10	25	29	7	10	14	13	7	3	7	10	6	3
2002	34	30	50	21	11	27	37	11	13	23	13	14	5	13	6	6	3
2003	37	40	56	26	14	35	46	9	15	22	18	16	5	11	7	5	7
2004	40	36	71	23	16	38	39	17	14	31	20	17	11	14	7	11	9
2005	28	31	56	33	20	28	21	12	19	36	14	22	8	13	8	9	9
2006	51	39	103	42	22	34	35	28	25	53	14	24	14	24	6	16	11
2007	42	32	69	44	19	20	28	32	15	32	16	18	15	21	9	9	8
2008	48	30	79	32	17	23	30	23	11	37	6	28	6	20	3	16	13
2009	38	25	71	33	24	20	21	23	16	34	5	18	11	27	8	11	9
2010	38	22	72	18	19	18	13	24	11	24	5	21	17	16	6	7	14
2011	50	12	93	17	24	25	19	34	11	28	5	27	17	16	8	12	14
2012	55	26	85	28	36	30	23	44	14	38	10	29	22	21	11	11	8
2013	63	21	107	43	42	29	19	48	16	33	12	26	20	23	16	9	18
2014	64	36	115	50	47	38	33	52	22	39	18	35	24	31	9	14	15
2015	68	25	126	43	51	35	28	56	18	38	13	33	24	27	11	10	18
2016	74	34	126	42	56	33	37	57	26	33	14	39	19	34	18	17	8
2017	61	40	110	66	53	34	33	54	27	35	12	34	19	40	10	9	7
2018	81	38	123	57	57	31	37	36	28	34	10	32	24	25	18	7	8
2019	58	36	93	57	42	44	28	51	18	31	5	26	15	19	13	9	10
2020	50	32	84	53	51	25	17	30	18	25	3	23	18	23	11	3	1
Avg 2000-2009	37.1	31.3	31.3	28.5	15.6	26.6	30.7	16.4	14.4	28.9	13.2	17.2	7.9	15.6	6.5	9.4	7.4
Avg 2010-2020	60.18181818	31.2	103.0909091	43.09090909	43.45454545	26.8	26.09090909	44.18181818	19	32.54545455	9.727272727	30.4	19.90909091	25	11.90909091	9.818181818	11

After we calculated the averages of all genres, we began to plot said averages based on our previous groups popularity. The first one below represents the averages of action (Red), adventure (Blue), comedy (Purple), and fantasy (Green).



Since the last graph focuses on the top genres in our data set. The next plot focuses on a group with less entries. The group has the following genres: martial arts (Red), demons (Blue), military (Green), psychological (Light Blue), space (Pink), game (Dark Yellow), and music (Purple) in which music is the highest frequency for this group. Although the plots may seem similar, the highest value in this group only reaches to lower part of the previous plot.



After observing the remaining groups, we notice the same pattern as the density graphs. The majority of the genres had a steady increase from 2010 to 2020. The increase for all genres could be due to the anime industry becoming popular worldwide. Another possible factor could be external events. For example, a possibility of the music genre's increase could be due to the popularity of Idols music such as Kpop (Korean pop) and Jpop (Japanese pop). Idols music is essentially boy/girl bands from the 90s. Although the possibilities could be endless, we know that there has been an increase in popularity of anime worldwide, and since the age of the internet there has been more investment in the anime industry.

5 Conclusion

We tried to implement the time series analysis method, but unfortunately due to the nature of our data-frame made it unworkable. Instead we used the concept of time series analysis within our data-frames, without the implementation of time series algorithms. Our analysis led us to many findings up to this date. We were able to see the common trends in different adaptations. Based on our findings we could see that the manga and visual novel popularity was significantly adapted more than the other source materials. Because of all the pre-existing unknowns in the data set we knew we had to change their values to null or the final observation of the data would be obsolete.

After we observed the data pertaining from the adapted shows, we moved on to analyzing the genre. After we learned how to manipulate the data to a workable state, we used the concept of time series analysis in the separation of the multiple data-frames. We were able to see a common trend with many of the genres. We expected to see action, adventure, and comedy in the top genres; however, we didn't expect to see the drastic increase of popularity in all genres in the early 2010s. Anime had sharp spikes in popularity in the years between 2010 to 2020. We assume these spikes in popularity were caused by an increase interest in Japanese animation worldwide. Such that as the interests increased so did the production of anime.

Currently we have been unable to accurately address the industry problem we set out to fix, this is something that we are working on; this leads us to continue our investigation so that we can see whether a particular adaptation or high frequency genre will be profitable for the industry.

6 Acknowledgements

Data from: https://myanimelist.net/
Users preference provided by: Jikan API
ありがとうございます。