



MyAnimeList.net data analyses: part I

Petr Mandrik

14 May 2021

Dataset exploration

Input:

- myanimelist.net database at Kaggle: [anime-recommendation-database-2020](#)
- information about 17.562 anime and the preference from 325.772 different users

Tasks:

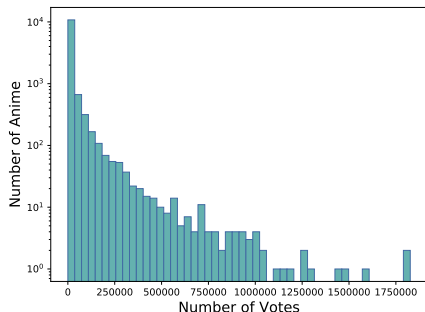
- Exploration of initial dataset, feature engineering

Technology solutions stack:

- Python, pandas, numpy, matplotlib, SQL

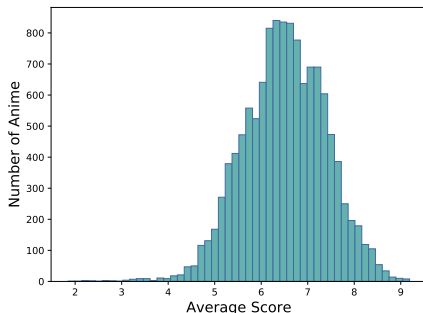
Some codes are available:

- github.com/pmandrik/AnimeRecommendationSystem2020
- kaggle.com/pmandrik/anime-data-exploration



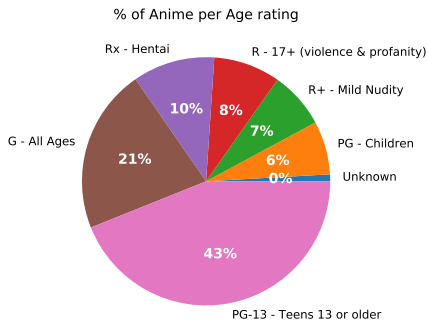
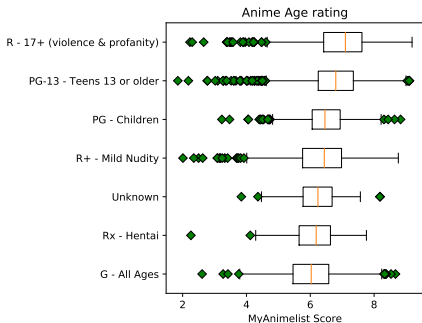
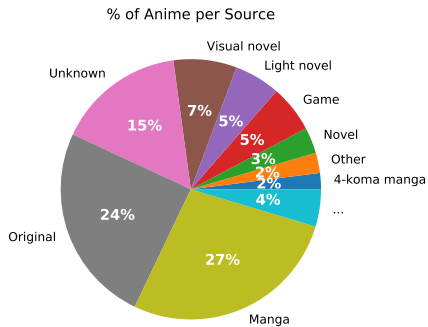
Top 10 Anime by N votes:

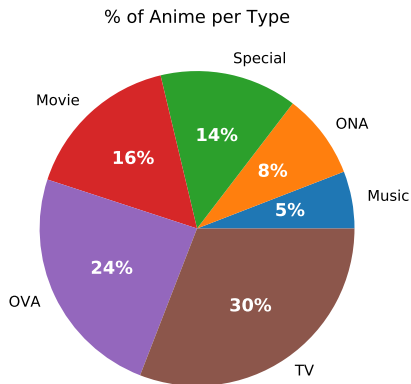
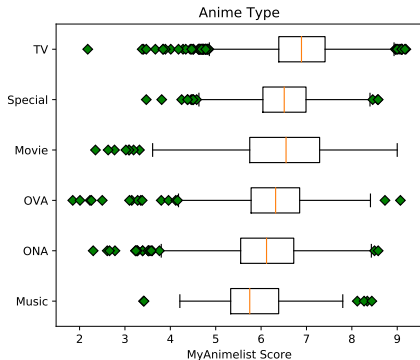
Name	<Score>	N votes
Death Note	8.63	1826691
Shingeki no Kyojin	8.48	1791099
Sword Art Online	7.25	1574372
One Punch Man	8.57	1478645
Fullmetal Alchemist: Brotherhood	9.19	1438767
Boku no Hero Academia	8.11	1305467
Naruto	7.91	1268593
Tokyo Ghoul	7.81	1260191
Kimi no Na wa.	8.96	1190759
No Game No Life	8.2	1137376



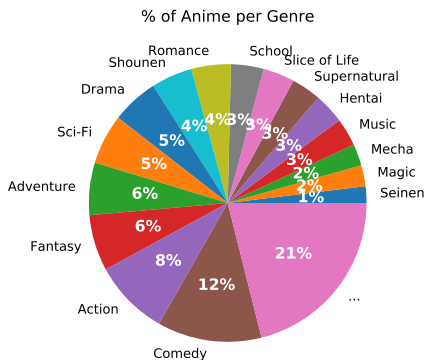
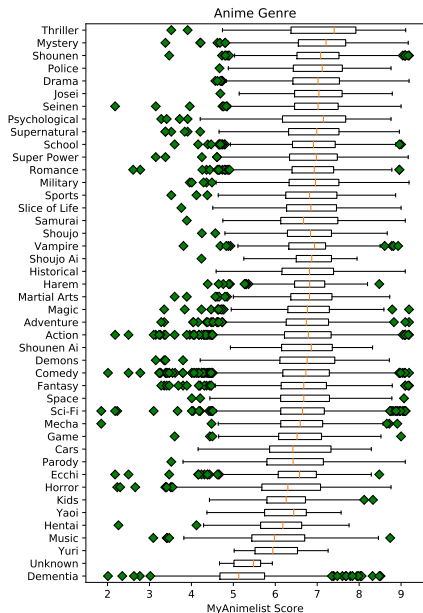
Top 10 Anime by <Score>:

Name	<Score>	N votes
Fullmetal Alchemist: Brotherhood	9.19	1438767
Shingeki no Kyojin: The Final Season	9.17	288274
Steins;Gate	9.11	989905
Shingeki no Kyojin Season 3 Part 2	9.1	728435
Gintama ^o	9.1	161812
Hunter x Hunter (2011)	9.1	1026866
Gintama'	9.08	162866
Ginga Eiyuu Densetsu	9.07	58655
Gintama': Enchousen	9.04	113662
Koe no Katachi	9.0	940843



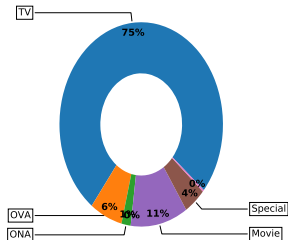


- Adoption of Light Novel, Manga and other works receive better score in average than Original anime
- well rated R-17+ anime are only 8% of the total number of films

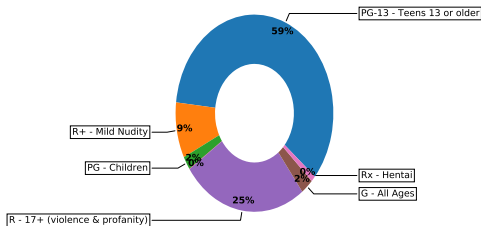


- * Anime contain different genres
- looks like it is really hard to create a good 'Dementia' but Experiments Lain still exist

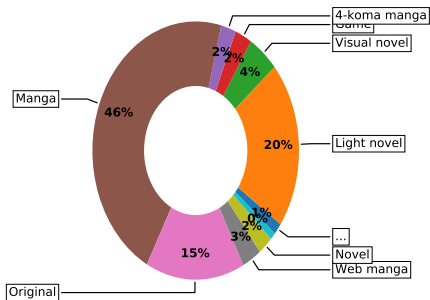
% of User Votes per Anime Type



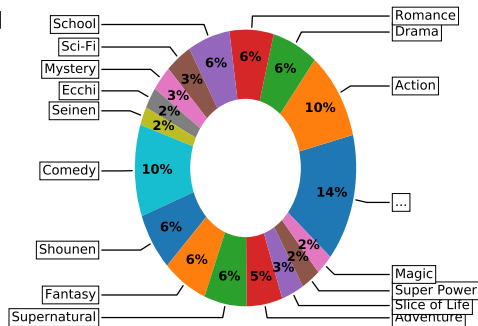
% of User Votes Anime Age rating

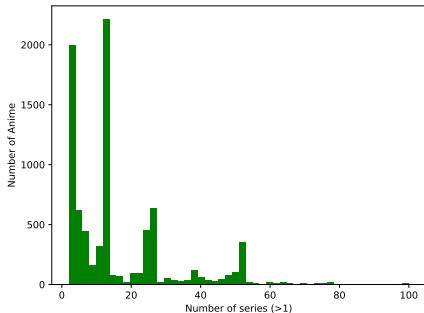


% of User Votes per Anime Source

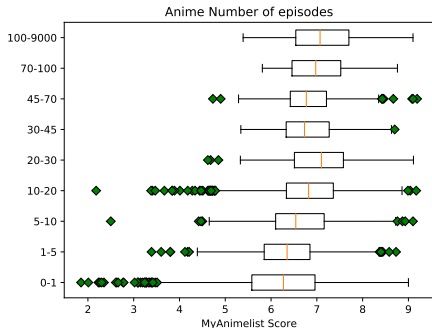


% of User Votes per Anime Genre

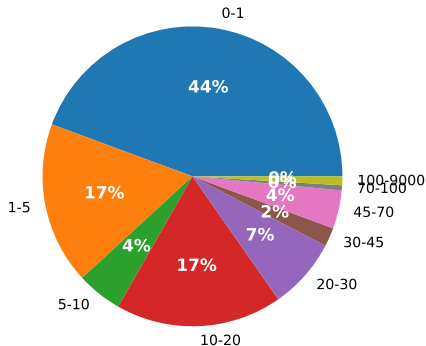


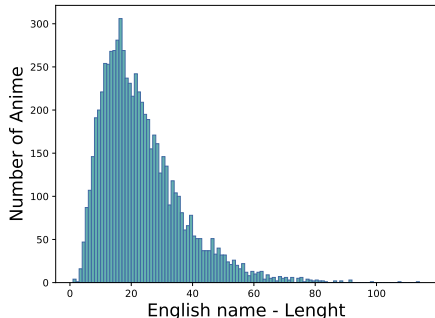
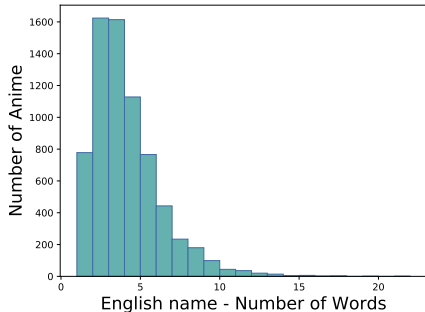


The number of episodes is grouped based on the visible structures in the distribution. Obtained groups are one-hot encoded.



% of Anime per Number of episodes



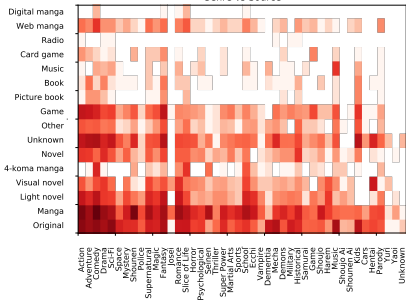


Longest English Anime names are:

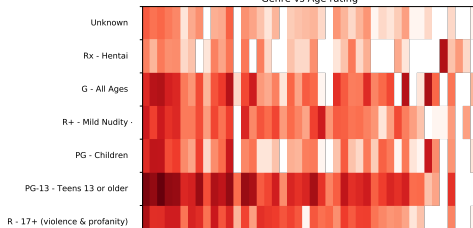
- “Code:Realize ~Guardian of Rebirth~ OVA, Code:Realize ~Guardian of Rebirth~ Chapter 13:Set a thief to catch a thief”
- “Miss Kobayashi’s Dragon Maid Episode 14:Valentine’s, and Then Hot Springs! - Please Don’t Get Your Hopes Up”
- “Is It Wrong to Try to Pick Up Girls in a Dungeon?:Is It Wrong to Expect a Hot Spring in a Dungeon?”

Shortest: “9”, “2”, “X”, “K”, “CP”, “Fly”, “Bug”, “For”, “Dot”, “She”

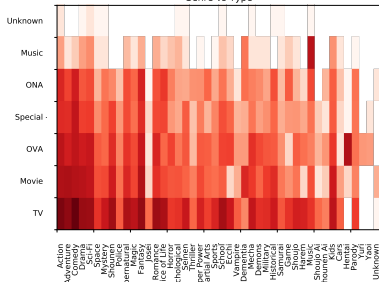
Genre vs Source



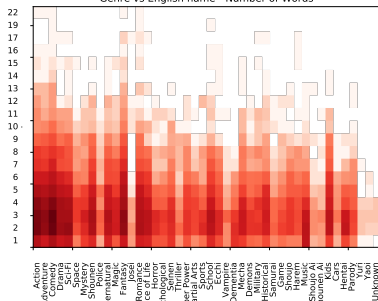
Genre vs Age rating



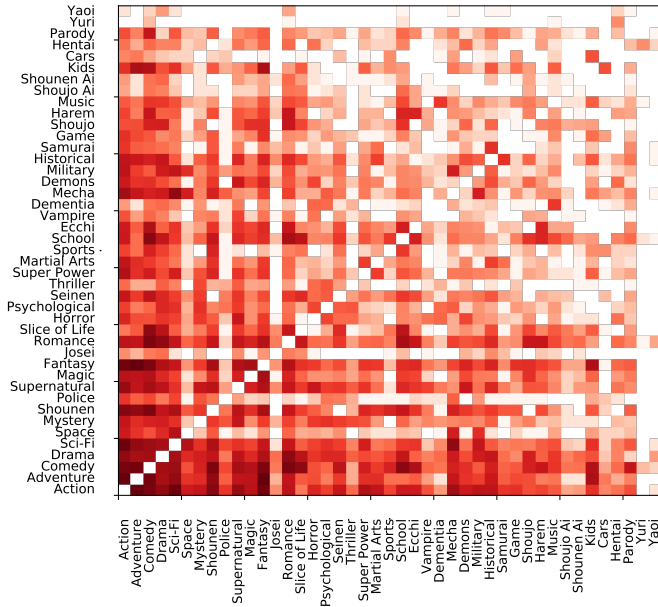
Genre vs Type



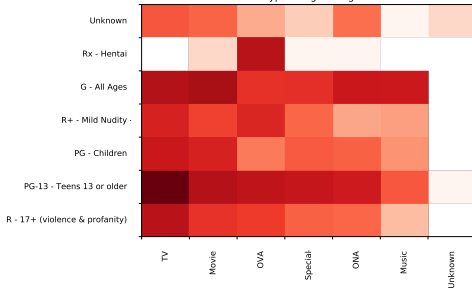
Genre vs English name - Number of Words



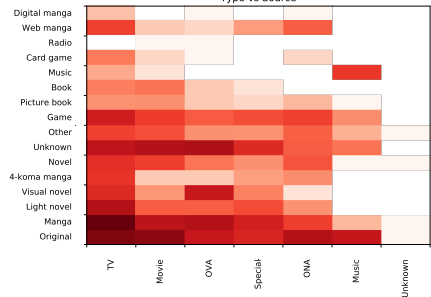
Genre vs Genre



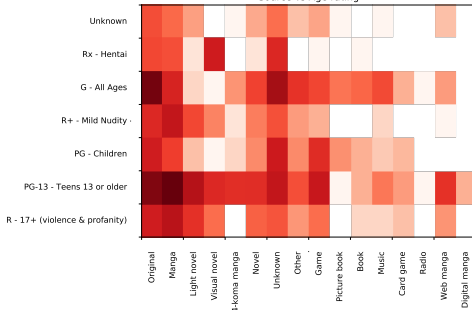
Type vs Age rating



Type vs Source



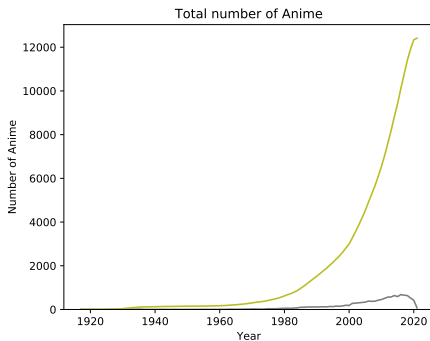
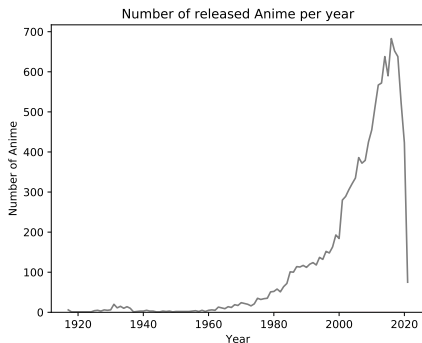
Source vs Age rating

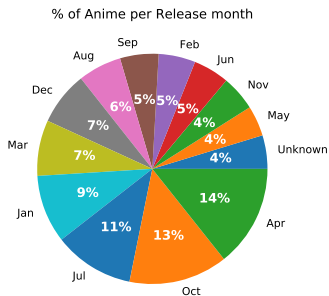
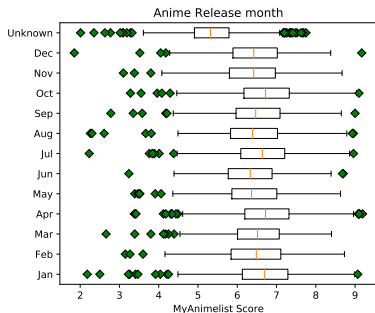
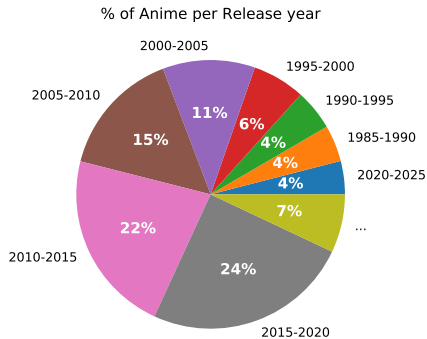
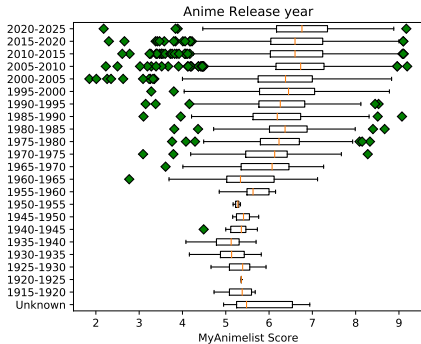


- most obvious correlations for Hentai Anime genre
- Manga is a universal source, while there is no Space, Police and Cars genre based on Visual Novel

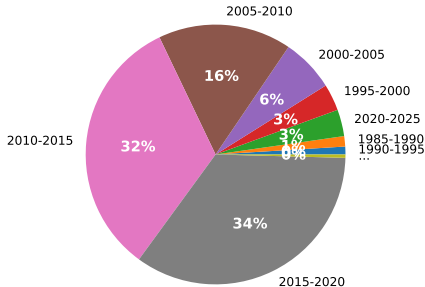
Evolution through years

For the following I will use the release date of the first episode:

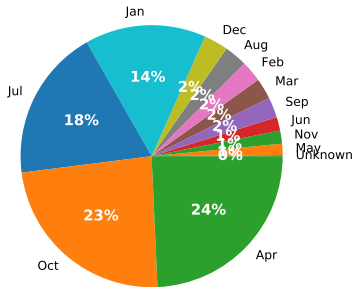




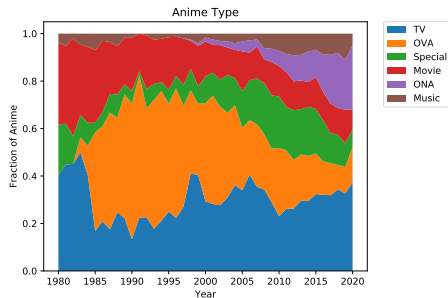
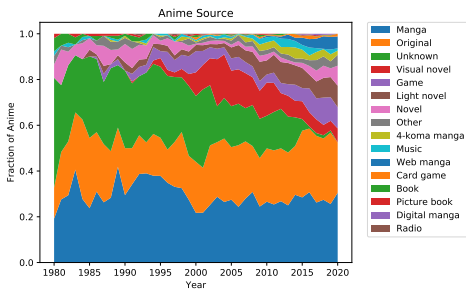
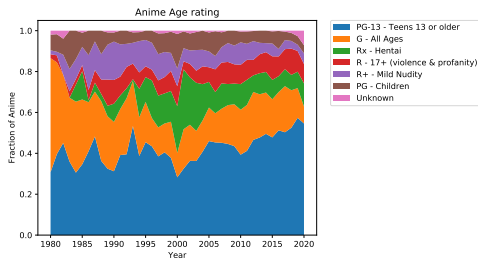
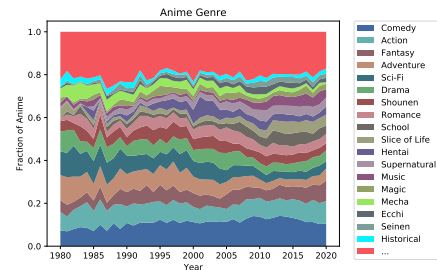
% of User Votes per Anime Release year



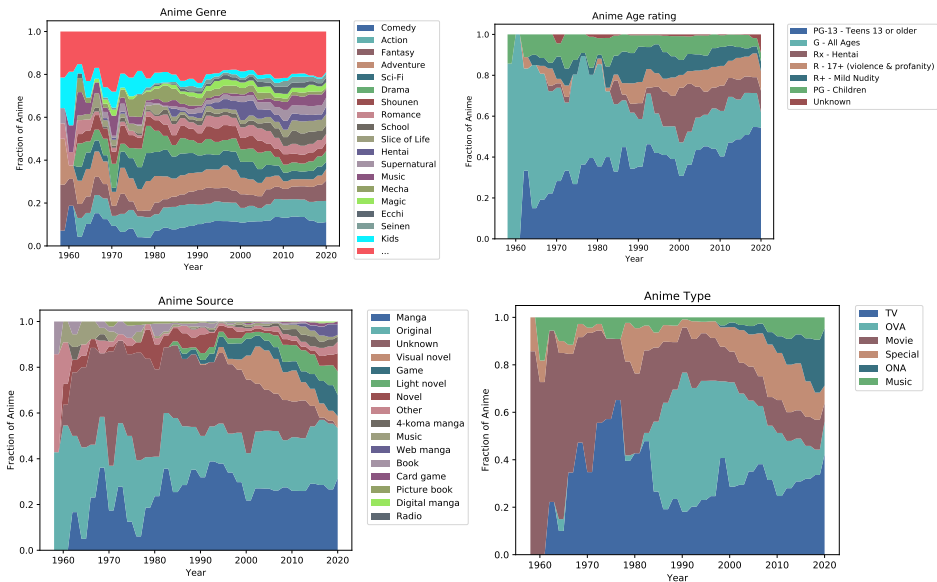
% of User Votes per Anime Release month



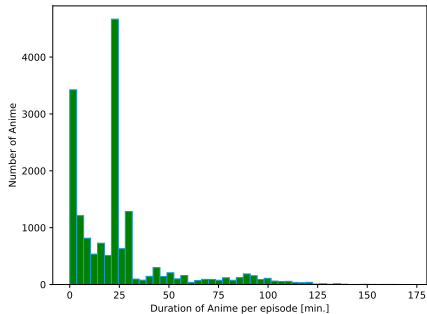
- while good Anime may be released at any time, only Apr.+Oct.+Jul.+Jan. have $\approx 45\%$ of all available Anime with $\approx 80\%$ of User Votes!
- Anime released before 2010 gain $\approx 30\%$ of User Votes only!



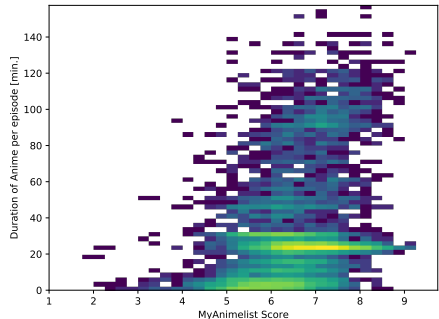
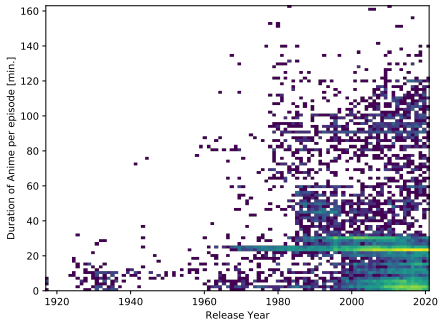
More School and Slice of life and less Sci-fi and Adventures



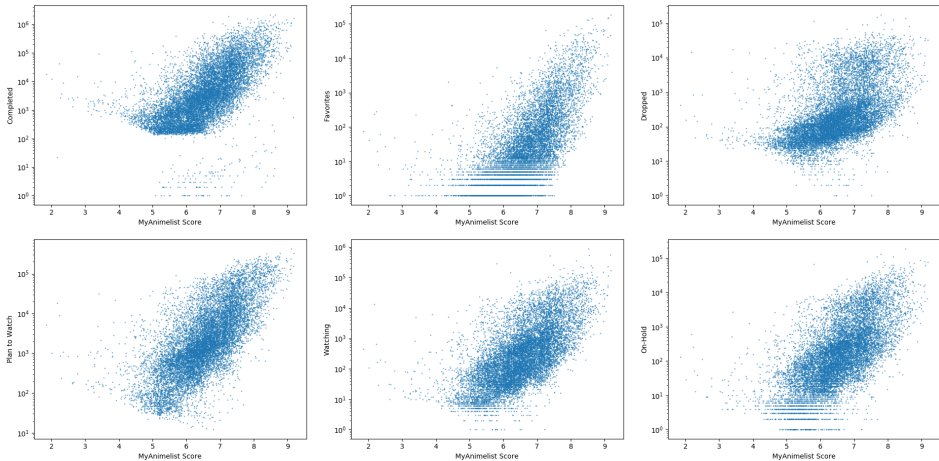
From 1958 with averaging over 2 years.



- 24 minutes duration of Anime per episode is established since 1970
- a lot of short Anime released recently

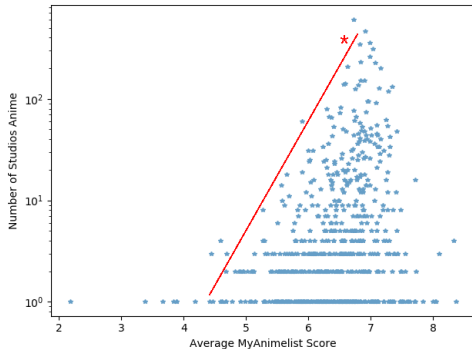


User List Status vs average Anime Score



Producers, Licensors, Studios

Total number of studios = 624

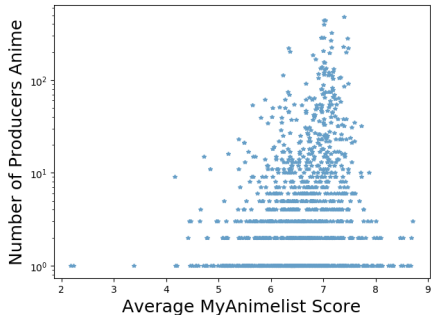


* minimal required average score
for Studio as a function of N of
Anima: $S = 0.018 \cdot N + 4.1$

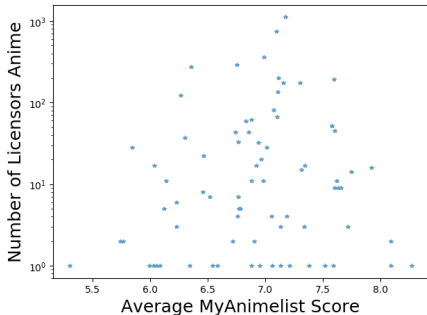
Top 10 Studio by number of
associated Anime

Studio	N
Toei Animation	604
Sunrise	464
Madhouse	355
J.C.Staff	346
Production I.G	312
Studio Deen	262
Studio Pierrot	232
TMS Entertainment	226
OLM	207
A-1 Pictures	201

Total number of Producers = 1208



Total number of Licensors = 77



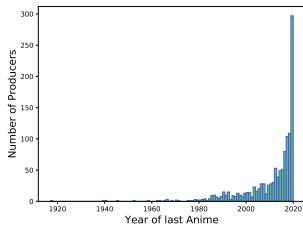
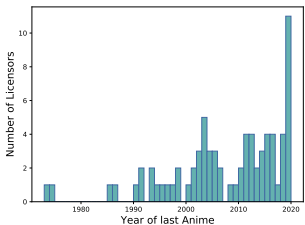
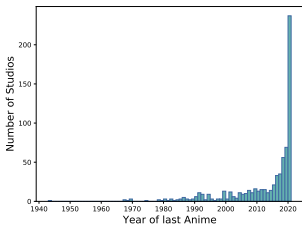
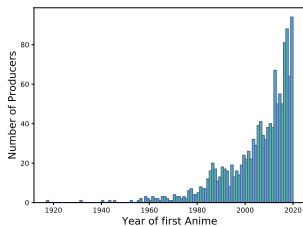
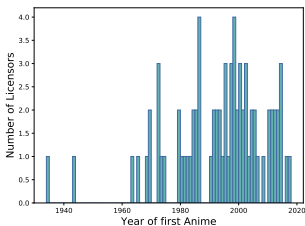
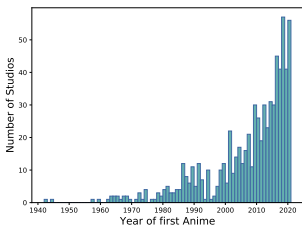
Top 5 Producers by number of Anime

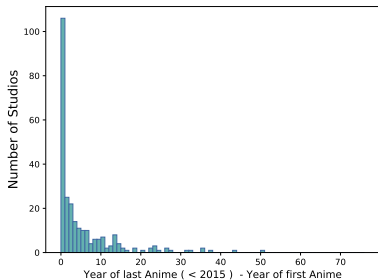
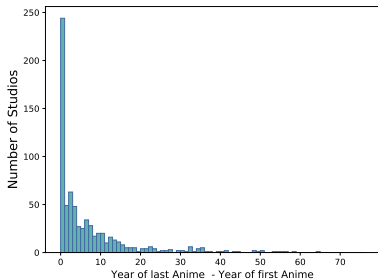
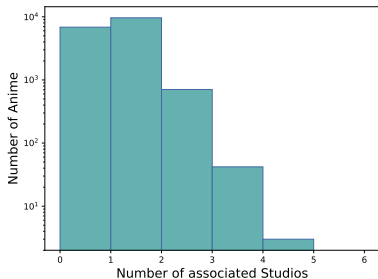
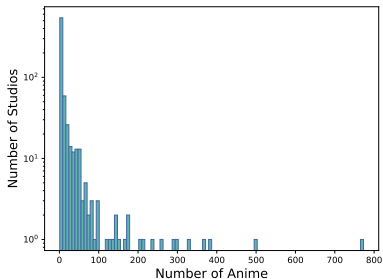
Producer	N
Aniplex	476
Lantis	436
TV Tokyo	436
Bandai Visual	393
Movic	323

Top 5 Licensors by number of Anime

Licensor	N
Funimation	1116
Sentai Filmworks	743
Discotek Media	360
ADV Films	292
Media Blasters	271

There is no much information about Studios but we can estimate the activity dates using first and last release year of associated Anime:

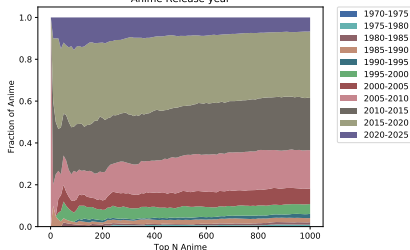




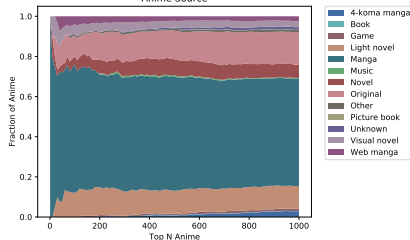
- More than 1 Animation Studio for most of the Anime
- First year is a really hard for Studio to survive

Features for Top-tier Anime

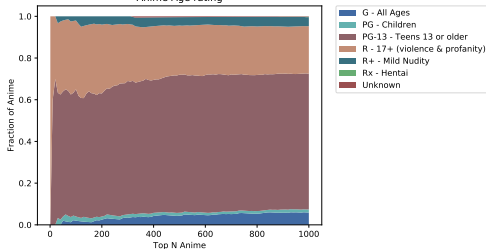
Anime Release year



Anime Source



Anime Age rating



Anime Type

