Impact of Anime Creation Sources on Anime Quality

A Multivariate Analysis with Bangumi Data

Hanwen Zhang, Jiahui Gu, Yihong Xiao

December 17, 2025

Introduction & Data Source

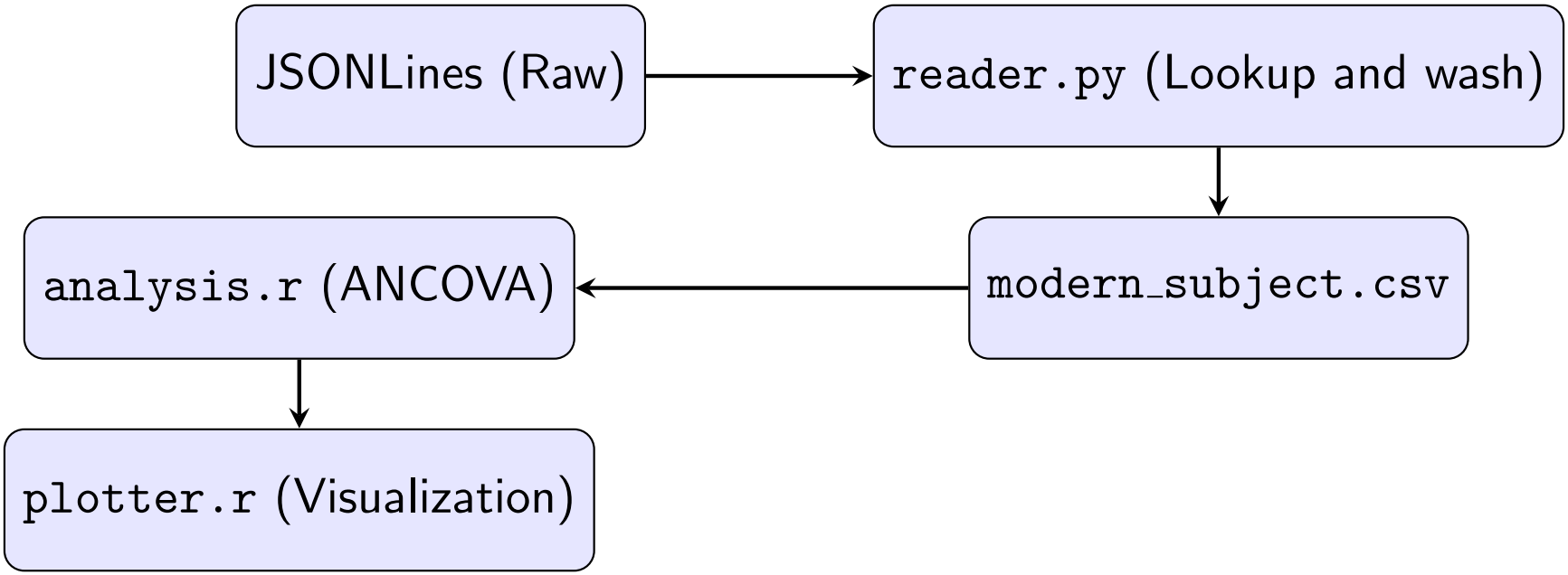
Thesis

This project will analyze the Impact of different anime creation sources on anime quality, with the relation-mapping of different genere adaptation.

Data Source: Bangumi Archive

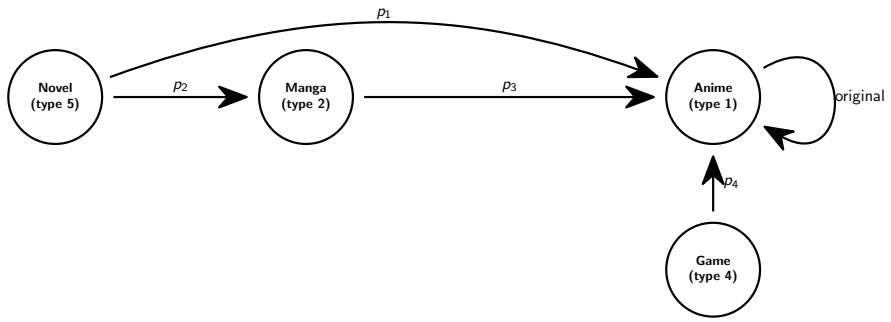
- **Core Datasets:** `Subject.jsonlines` (Metadata with 580000 rows and 6 snapshots) and `Subject-relations.jsonlines` (Adaptation relations with 800000 rows).
- **Scale:** Modern dataset (1980–2025) containing thousands of entries.
- **Key Features Extracted:**
  - `score` (User rating) & `id` (generated by Bangumi server).
  - `played_amount, dropped_amount` (Popularity).
  - `type` (Novel, Manga, Anime, Game).
  - `relation_type`: adaptation.

Data Processing Pipeline



Methodology: The Adaptation Flow

1. Relation Mapping
Paths to anime (e.g., Novel $\rightarrow$ Manga $\rightarrow$ Anime).



2. Statistical Model
Filter the mean anime score (s_{anime}) adjusted by one-year time interval.
Setting: If $s_{anime} \geq s_{source}$, the adaptation is "successful or qualified".

Result: Descriptive Statistics

Table: Mean Scores by Source (Stable Anime)

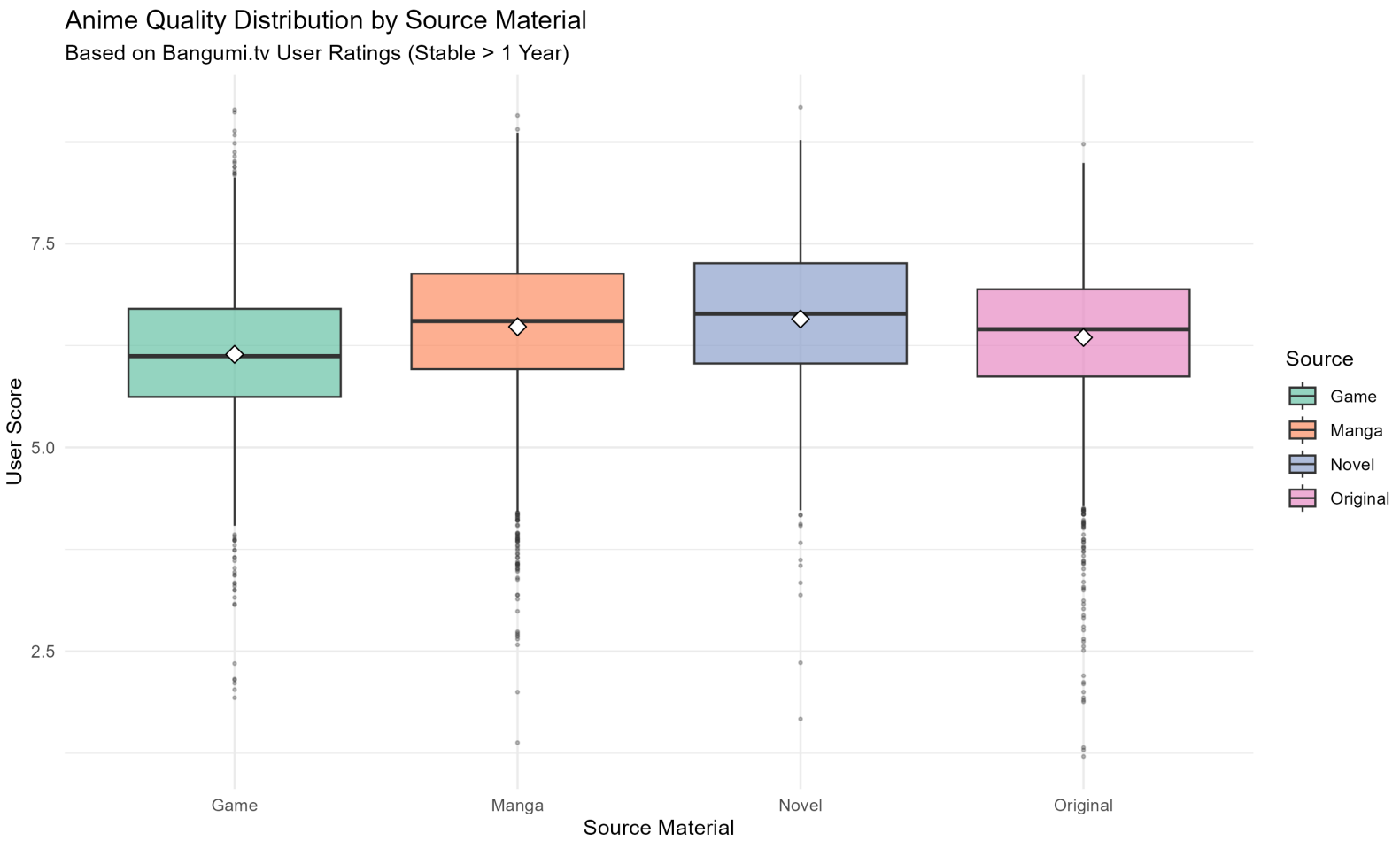
Source Category	Mean Score	N (Count)
Novel	**6.59**	1,880
Manga	6.49	**11,401**
Original	6.36	8,310
Game	6.15	4,885

Initial Finding: Manga or Novel have the highest raw mean, while Game adaptations has the lowest average quality.

Result Plot: Popularity vs Quality

- Positive correlation across all sources.
- **Selection Bias:** High-popularity titles tend to be better-funded adaptations.

Result Plot: Quality by Source



- Novels show a higher "upper bond" for quality.
- Original anime show the widest variance (highest "risk").

Result Plot: Temporal Stability Using Linear Mixed Model

```
> summary(lmm_model)
Linear mixed model fit by REML ['lmerMod']
Formula: score ~ Source_Category * Snapshot_ID + log_played + (1 | Snapshot_ID)
   Data: df_combined

REML criterion at convergence: 65473.5

Scaled residuals:
    Min      1Q  Median      3Q     Max
-6.2400 -0.5452  0.0947  0.6791  2.7861

Random effects:
 Groups      Name        Variance  Std.Dev.
 Snapshot_ID (Intercept) 9.821e-05 0.00991
 Residual                6.924e-01 0.83213
Number of obs: 26476, groups:  Snapshot_ID, 3
5609   0.026428  -1.347
log_played                                  0.307591   0.003982  77.252
Source_CategoryGame:Snapshot_ID25_09_30   0.008736   0.036913   0.237
Source_CategoryManga:Snapshot_ID25_09_30 -0.004759   0.029548  -0.161
Source_CategoryNovel:Snapshot_ID25_09_30 -0.005609   0.026428  -1.347
log_played                                  0.307591   0.003982  77.252
Source_CategoryGame:Snapshot_ID25_09_30   0.008736   0.036913   0.237
Source_CategoryManga:Snapshot_ID25_09_30 -0.004759   0.029548  -0.161
Source_CategoryNovel:Snapshot_ID25_09_30 -0.005845   0.052618  -0.111
Source_CategoryGame:Snapshot_ID25_11_26   0.006179   0.036811   0.168
Source_CategoryManga:Snapshot_ID25_11_26 -0.005457   0.029432  -0.185
Source_CategoryNovel:Snapshot_ID25_11_26 -0.016440   0.052330  -0.314

Correlation matrix not shown by default, as p = 13 > 12.
```

- **Monitoring:** No significant interaction between *Source* and *Snapshot Date*.

$$\text{Score}_{ij} = \beta_0 + \beta_1 \text{Source}_i + \beta_2 \log(\text{played\_amount})_{ij} + \gamma_j(\text{Snapshot}_j) + (\text{Source}_i \times \text{Snapshot}_j) + u_j + \epsilon_{ij}$$

- Quality perceptions are stable over the 6-month data's time-snapshot interval(crossed two seasons).

Citation & Discussion

Project Repository

ACGN-Biased-Multivariate-Analysis
`https://github.com/CarlXerophilem/ACGN-Biased-Multivariate-Analysis`

Data Sources

`https://github.com/bangumi/Archive/releases/tag/archive`

Future Works:

- Complete tracking of adaptation paths across multiple ancestors.
- Segmentation of different animation subgenres (TV / Movie / OVA).
- Introduction of time-entropy related models to observe market trends (include js script for auto updating)...