

A3 131-Project:

1) Project Objective: **What genres of anime do Males and Females like the most?**

2) Metadata of the Anime 2023 dataset:

- How many unique users?

116,125

- Male and Female count?

Anonymous count: 506907

Male count: 126984

Female count: 96485

Non-Binary count: 914

- value ranges:

Popularity column: 1 - 56505

my_score column: 0 – 10

Nontrivial information:

- Male and female users like exactly the same anime genres

Male top 10:

genre,users-ratings

Comedy,12443717

Action,10506464

Romance,7422955

Drama,6718771

Fantasy,6232419

School,6205281

Supernatural,5899197

Shounen,5862779

Sci-Fi,5428070

Adventure,5195149

Female top 10:

genre,users-ratings

Comedy,4674081

Action,3482950

Drama,3104768

Romance,3058903

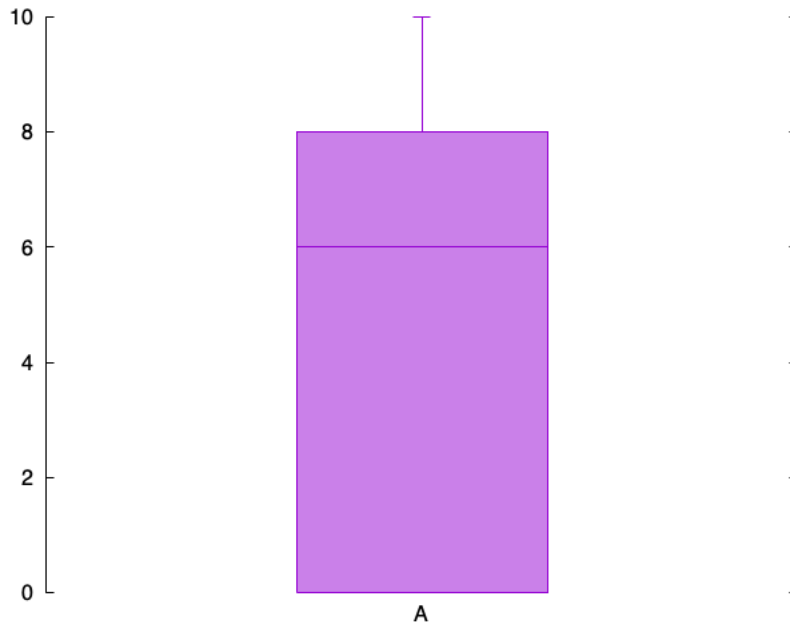
Fantasy,2477082

Supernatural,2408702

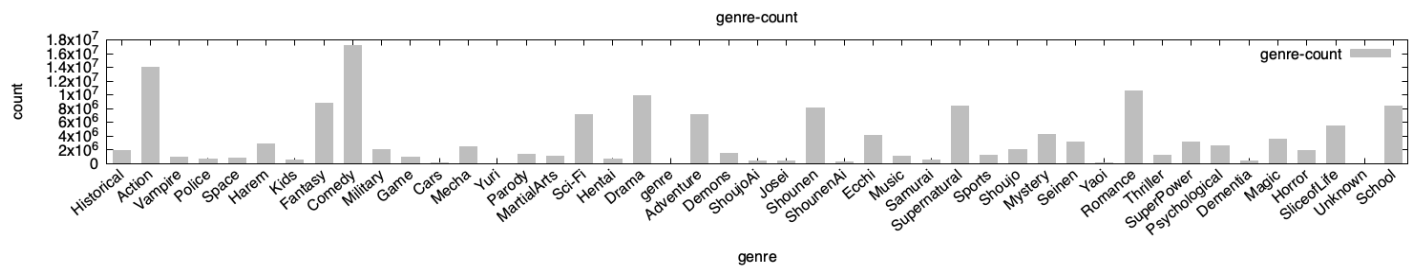
School,2163059
Shounen,2134648
Adventure,1970463
Sci-Fi,1658090

3) Plots:

User-score distribution (box plot):



Genre-count (Histogram):



4) Summary:

I found that male and female users like the same genres of anime. As a side note there are also significantly fewer female users than male. I also found out that the median rating that a user gives a 6 which means that most of the anime that people watch end up being something they don't like. It can also be inferred from the box plot that 25% of the anime are rated a score of 0.

5) Tools used:

I used the Kaggle api via command line to download the dataset. Then, I used awk and sed to clean the data, but I didn't have to do much as it was mostly already clean. I also wrote scripts to collect data to plot using gnuplot.