# PDM Video Games

## Steam Game Based Database

## The Team

Isaac Kim, Gabe Rippel,
Alex Denny, Will Poirier

## Game Ratings

Fig.1 shows the sampling distribution of every game's ratings. The distribution is normal as expected from the random rating generation. Each game and its ratings is a sample. It has a mean of 0.49 and a standard deviation of 0.085. Consider it a control group.

Fig.2 shows the distribution of ratings from Steam itself. The drastic skew to a maximum rating indicates that real world users tend to rate games far more positively. The mean of Steam's rating data is 0.83.

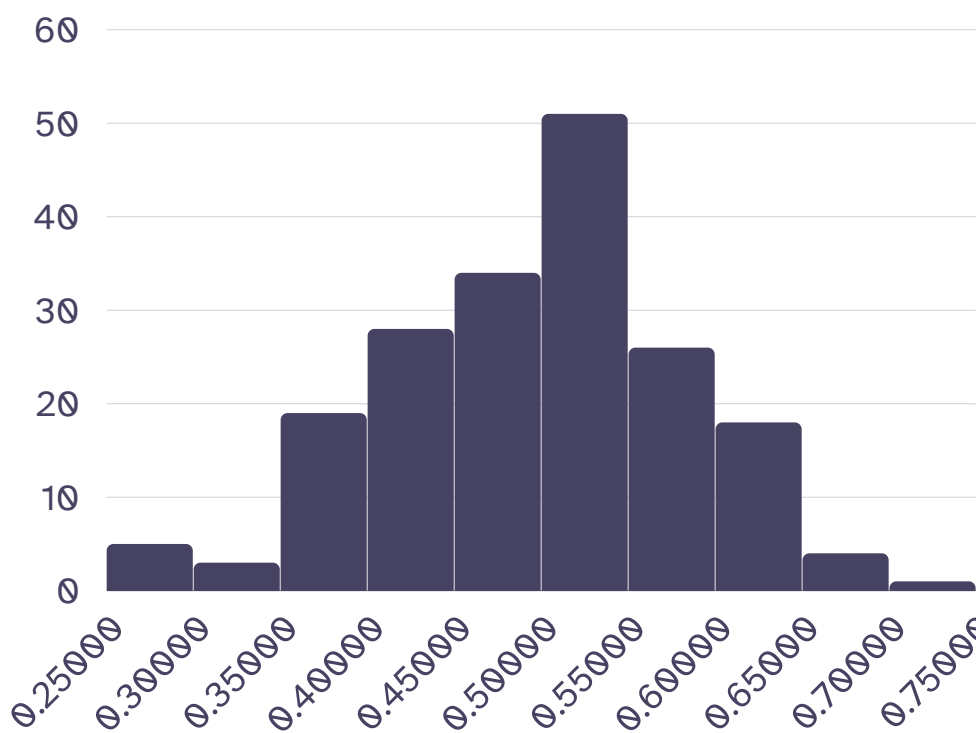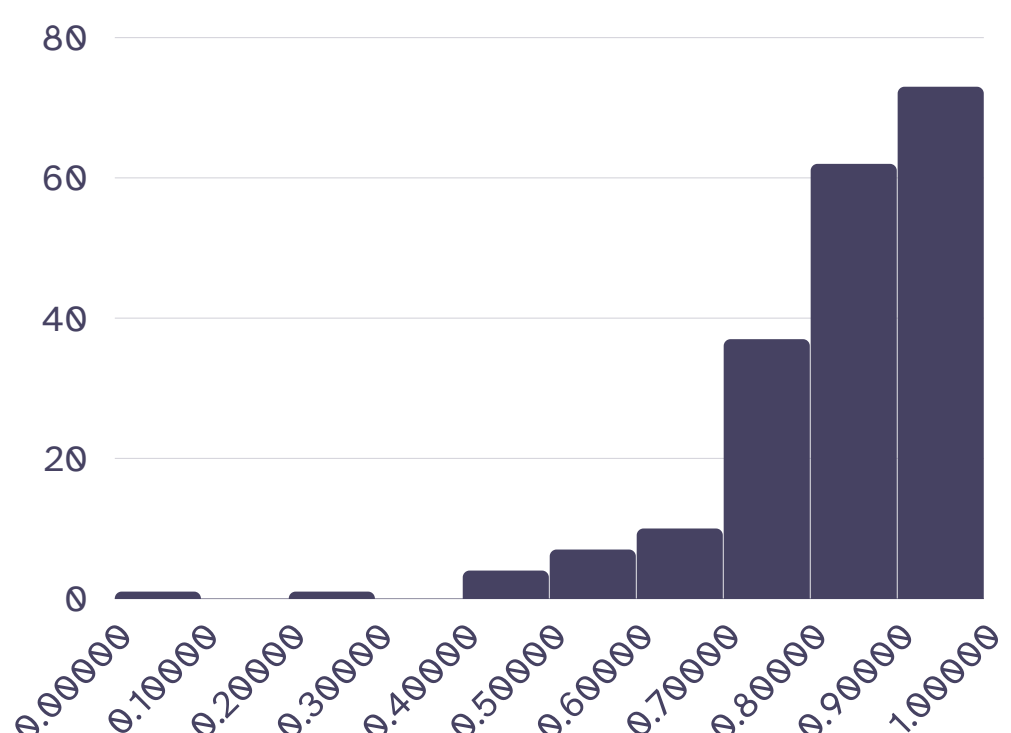Note: The ratings are measured from 1-5 but they've been normalized to 0-1.

*Fig.1*

*Fig.2*

## Top 10 Games
### By Appearances in Collection

Minecraft - 12
Quake - 11
Full Spectrum Warrior - 11
Vampire: The Masquerade - Bloodlines - 10
Fortnite - 9
Peggle Nights - 9
Half-Life Opposing Force - 7
Peggle Extreme - 7
Cossacks: Art of War - 7
Bad 2 Bad: Apocalypse - 7

| SUM of genre_p... genre_name_2 genre_name_1 | 1990's | 2.5D | 2D | 2D Fighter | 2D Platformer | 3D | 3D Platformer | 4X | Abstract | Action |
|---|---|---|---|---|---|---|---|---|---|---|
| 1980's | | | | | | | | | | |
| 1990's | 22 | 3 | | | | | | | | 19 |
| 2.5D | 3 | 3 | | | | | | | | 3 |
| 2D | | | 15 | | | | | | | 9 |
| 3D | | | | | | 12 | | | | 8 |
| 3D Platformer | | | | | | | 3 | | | 3 |
| 4X | | | | | | | | 9 | | 4 |
| Action | 19 | 3 | 9 | | | 8 | 3 | 4 | | 123 |
| Action RPG | | | | | | | | | | 13 |
| Action-Adventure | 2 | | | | | | 2 | | | 5 |
| Addictive | | | | 3 | | | | 2 | | 5 |
| Adventure | 9 | 2 | 2 | | | 2 | 3 | | | 52 |
| Aliens | 5 | | | | | | | | | 14 |
| Arcade | | | | 6 | | | 2 | | | 8 |
| Arena Shooter | 4 | | | | | | | | | 7 |
| Assassin | | | | | | | | | | 5 |

*Fig.3*

## Related Games

Fig.3 is an excerpt from a table containing a matrix which outlines genres that are related to each other by how many times they appear in the same game. As you can see, the column for "Action" contains a lot of related genres. "Action" is found in the same games as "Adventure" 52 times. The full table is in the report.

## Is Expensive Better?

Fig.4 shows a scatterplot of every game in the database with rating on the Y-axis and price on the X-axis. The line of best fit is shown as well.

Surprisingly there is no meaningful skew towards more expensive games being more highly rated or otherwise. If anything, cheaper games seem to typically be rated just slightly higher than more expensive games.
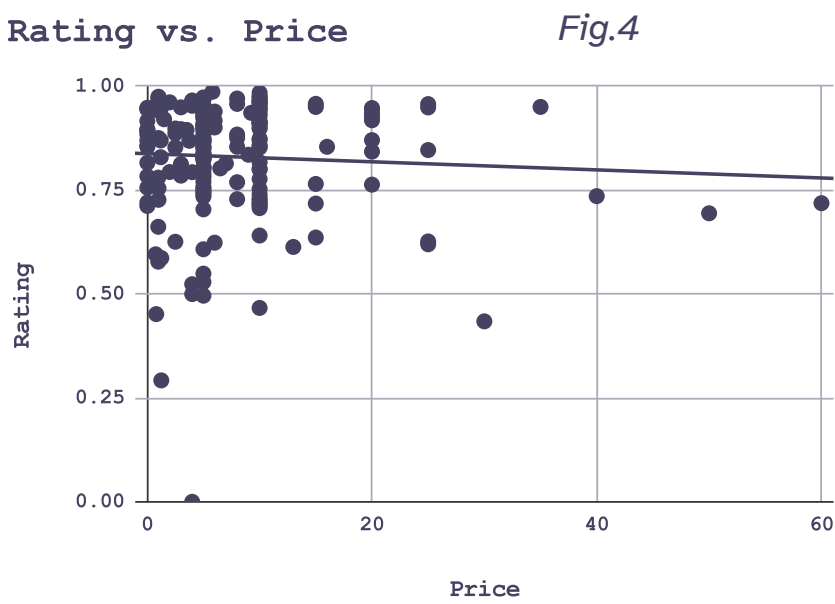
**Rating vs. Price** *Fig.4*

## Expensive Games In Collections

In the database, games that are in more expensive did not turn out to be in more collections. This is because the games in each collection were generated randomly. In the real world, we predict that more expensive games would show up generally less frequently.
A quick look through the list of prices would indicate that a lot of the more expensive games tend to be games that you've never heard of.

## Methods

The data was gathered mostly through joining tables together and outputting them into CSV files for further processing in Excel. The related genres table was created via self-joining game_genre tables together and operating on that. SQL outputted name1, name2, and the pair's coincidence count. That data was exported to a CSV which was then turned into a pivot table in Excel which arranged it into an N by N matrix.