Self-Evaluation of a Data Story on the Topic of Agreement between Players' and Critics' Ratings on Video Games

Gregory Walsh
Student ID 29785685
gw2g17@soton.ac.uk

***Abstract*—The ubiquity of quantitative consumer and critic ratings on the web, for products and services alike, gives data scientists the opportunity to evaluate how closely these two parties align in their opinions, and the means to highlight potential conflicts of interest. In this report, a self-evaluation of the visualisations used in a data story exploring the difference in ratings given to video games by critics and players is given.**

I. Data Story Summary

The author's data story, written in an informal style intended for a millennial audience, begins by posing to the reader, the question of whether or not video game critics represent the typical player. A scatter plot, contrasting average player and critic ratings, shows there are indeed differences. Specifically, a cohort of games with unusually high average critic ratings when compared to average player ratings.

Player ratings for this cohort of games, referred to in this report as the "favoured-by-critics cohort", are then analysed for peculiarities, however, none are found, suggesting a deeper underlying issue.

The prevalence of microtransactions[1] as a cause for the differences is then raised, and supporting evidence is presented. Specifically, the composition by genre and publisher of the favoured-by-critics cohort in comparison to all other games.

The narrative then concludes, stating that critics do not always represent players and that there is some evidence to suggest certain large publishers may be favoured with unduly high ratings from critics.

Throughout the article, Toulmin's model of argument was used as a framework for presenting claims. Within the actual article, the strength of the backing for each claim made was discussed.

II. Dataset Summary

Data for critic and player ratings were collected from Metacritic.com. For each game, their hallmark "Metascore" (an aggregation of reviews from trusted online critics) was used as a proxy for average critic scores. Granular player ratings data for each game was scraped with a python script, also from Metacritic.com.

Additional metadata about each game, such as genre, was sourced from IGDB.com, an online service for video game information, using their application programming interface (API). Fuzzy matching was used to combine the two data sources together into a single dataset.

III. Visualisations - Scatter Plot Contrasting Player and Critic Ratings

The data story begins by asking whether critics represent the players. To provide initial, high-level insight into this question, a scatter plot depicting the relationship between the average player rating and average critic rating for each game was presented (see Fig. 1).

During the design phase, it was found that games tended to sit above the identity line, which is labelled as "line of agreement" on the chart for a non-technical audience. Given that the Metascores will contain some bias due to the decisions made by Metacritic.com in their formulation, it would be dishonest to present the two sets of ratings as being equally calibrated and therefore perfectly comparable. To reduce the bias, player ratings were scaled by a constant factor (1.06), which sent the median displacement to the identity line to zero.

The games were then classified in an unbiased way, using a statistical approach to define the boundaries, by modelling the distances for all games to the identity line as a normal distribution and defining a 95% confidence interval.

The ability to filter games by genre and publisher was added to support exploration of the data, should the reader be curious having read the latter part of the article which includes analysis of the ratings broken down by these dimensions.

A. Justifications

Readers first needed to be made aware the extent to which player and critic ratings are correlated, and in which regions there exists unusual densities of data. Scatter plots provide a robust way to illustrate both of these kinds of information simultaneously. The option to categorise points by colour was useful for communicating the existence of the three classes.

1. Microtransactions are small payments made in video games for virtual items.

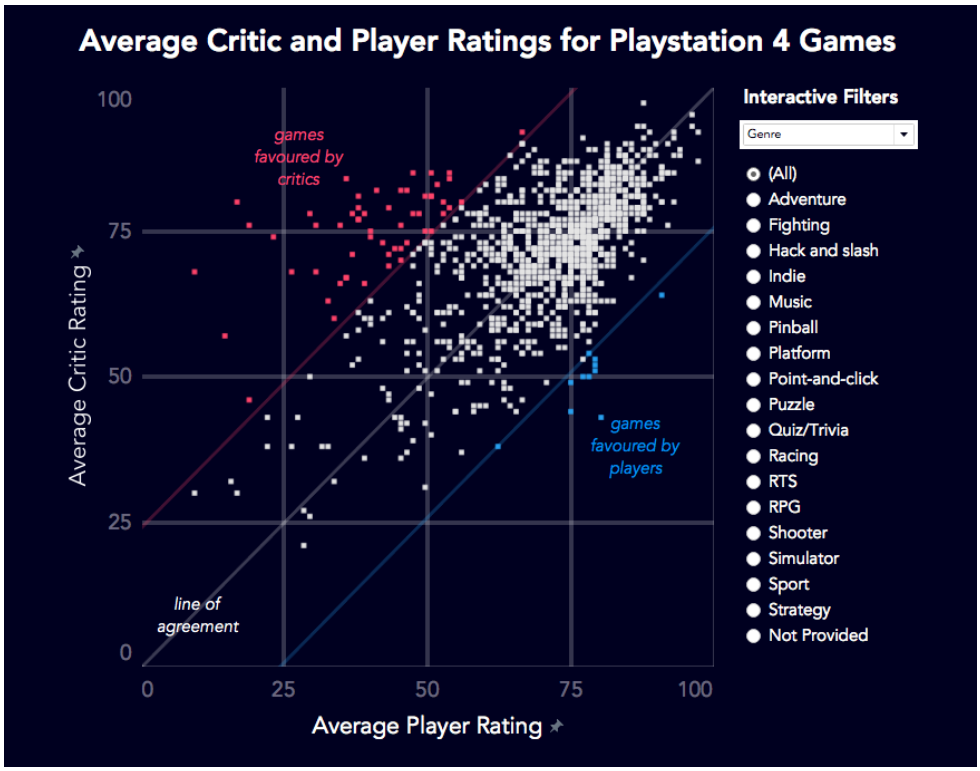

Fig. 1. *Scatter plot contrasting player and critic ratings. Each point represents an individual game.*

The scatter plot was set over the full possible range of ratings. This was important to avoid confusing (or worse, deliberately misleading) the reader, which Tufte quantifies with the "lie factor" [1].

Supplementary visual cues, such as reference lines can be helpful in focusing attention on key parts of a graphic [2]. This principle was used to clarify the boundaries of the classes. Research suggests that presenting information simultaneously in both visual and written form helps with memory retention [3]. To take advantage of this effect, labels were added to the identity line and classification regions. As a result, no legend was required reducing chart junk.

Bright colours are known to attract attention, which evolutionary theory suggests is a result of the benefit gained when searching for food, particularly ripe fruit [4]. Additionally, exposure to specific colours may affect our ability to memorise information [5]. To take advantage of these effects, bright colours were selected for the outlier classes. Since the story is targeted at an English-speaking audience, the cultural connotations of warning and danger associated with the colour red were deemed most appropriate for the favoured-by-critics cohort, because its existence should be alarming to the average player. A light grey shade was chosen for the general-agreement cohort so the emphasis lay on the outlier cohorts.

B. Strengths, Weaknesses and Improvements

The choice of red for the favoured-by-critics cohort is slightly contentious. It could be argued that shading this cohort red may induce negative preconceptions in the reader's mind, due to the cultural implications of the colour, before any further analysis has been discussed, thereby reflecting the author's bias.

On the positive side, the ability to hover over individual data points and see more information is useful to the curious reader who may wish to see more detail about specific games.

The plot could be further improved by including some of the statistical information used to define the classes to help the reader get a better understanding how the cohorts were created.

IV. Visualisations - Comparison of Ratings Breakdown for Review-Bombed and Non-Review-Bombed Games

After discussing the features of the scatter plot, it was necessary to assess if the games in the favoured-by-critics cohort lie far from the line because of unusual player ratings. Specifically, whether the games in the favoured-by-critics cohort had on the whole been subject to significant review-bombing[1]. A typical review-bombed game, as shown in Fig. 2 tends to have many positive and negative reviews.

To illustrate to the reader that the games in the favoured-by-critics cohort did not on average look like a review bombed game, and therefore these reviews were genuine reflections of typical players, histograms illustrating the breakdown of ratings for a typical review bombed game, a genuine low scoring game and the average for the favoured-by-critics cohort were shown.

By inspecting these figures, the reader can see that the average favoured-by-critics game looks very similar to the low

1. Review-bombing is the act of collaboratively posting negative user reviews for a product online with the intention of reducing the average rating of that product.

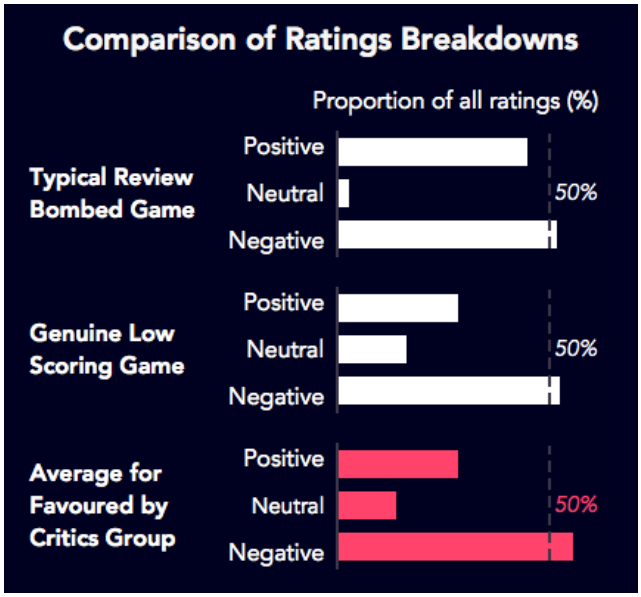

Fig. 2. *Histograms of ratings breakdown for typical games in three cohorts.*

Fig. 3. *A histogram showing the ratings breakdown for a product, as displayed in the review section of Amazon.co.uk*

scoring game, and quite different from the review bombed game. They can then see that the player ratings for these games are genuine.

To reduce bias resulting from including many games with few reviews, which could sway the proportions of ratings, only games with greater than 50 reviews were included.

A. *Justifications*

In this visualisation, it was necessary to communicate information encoded in three variables: group (categorical), rating level (ordinal), and proportion of total ratings (continuous). For this reason, the data was organised into sub figures. The effects of different groupings on the perception of entities are described by the "Gestalt Laws of Grouping" [6]. The arrangement of data illustrated in Fig. 2 was chosen to emphasise the signature distribution of rating levels for each group, rather than promote comparison of examples at specific levels.

Furthermore, this arrangement is likely to be more familiar to the reader, thereby reducing cognitive load, since each sub-figure is reminiscent of those commonly used on retail websites to depict the distributions of review scores (see Fig. 3).

A marked axis was not included since precise comparisons were not required. Instead, reference lines were used to indicate all sub-figures were plotted to the same scale.

B. *Strengths, Weaknesses and Improvements*

It is the author's opinion that the choice of arrangement helps to minimise the cognitive load imposed on the reader resulting from the combination of ordinal, categorical and continuous variables.

The ordinal values “positive”, “neutral”, and “negative” are not closely associated with ratings (unlike star ratings) and this likely makes the visualisation more difficult to comprehend than would be ideal. Unfortunately, this level of granularity was not available in the dataset, but would have been utilised if it were.

V. Visualisations - Doughnut Charts Contrasting Genre Mixes

At this point in the story, the author presents evidence that critics favour games from certain genres, and that this information may be useful in understanding why players give low ratings to games in the favoured-by-critics cohort.

The supporting visualisation, as depicted in Fig. 4, shows the genre mix for the favoured-by-critics cohort alongside the genre mix for all games. Its purpose is to help readers to see which genres were unusually prevalent in the favoured-by-critics cohort. The mix for all games was included as a baseline against which the mix of the favoured-by-critics cohort could be compared, as a way of reducing bias.

To emphasise the key data, only the genres which made up a significant proportion of the mix in the favoured-by-critics cohort were shown individually, and the remaining genres were grouped together.

A. *Justifications*

A pair of doughnut charts was used to create, in the reader’s mind, the perception of two entities with distinct compositions, and invite the reader to make comparisons between them.

Because it was not the purpose of these charts to facilitate the reader to make precise measurements of individual genres, the issue that charts which display magnitudes using areas or angles are less effective for comparing relative sizes than charts

Furthermore, since the number of genres in each chart is very small, the effect that pie and doughnut charts lose clarity when the data is very granular was also not an issue.

Colour coding was used to help the user identify the same genre across sub-figures. Since the colours blue and red had already been used to represent previous concepts, in order to make the visualisation more accessible to people with colour blindness, the yellow and green shades were kept separated, and the ordering of the genres in both sub-figures was kept consistent.

Fig. 4. *A pair of doughnut charts showing the relative compositions of genres for all games and for games in the favoured-by-critic cohort.*

B. Strengths, Weaknesses and Improvements

It is the author's feeling that the visualisation shown in Fig. 4 neatly conveys to the reader that these are two distinct entities, with unique compositions, and that they ought to be compared.

Unfortunately, when using sub-figures, particularly when the axes are unaligned, additional mental effort is required to compare the magnitudes of a genre across sub-figures. Since exact evaluations are not required to understand the main message, this issue is minor, but if a more precise evaluation were required, a Sankey diagram could be considered as an alternative. However, such a decision would mean the loss of the perceived wholeness of each group, as even though the data being presented would be identical the composition would change (another result of the Gestalt laws of grouping).

VI. Visualisations - Stacked Line Chart showing Favoured-Publisher Count Over Time

After showing that shooters and sports games were the most prevalent genres in the favoured-by-critics cohort, and under the hypothesis that this may be related to the growing incidence of microtransactions in the games of those genres, the article proposes to investigate if any particular publishers are favoured by the critics. This proposal is made principally because the publishers which are publicly traded companies are pushing microtransactions as part of their annual growth plans according to their annual statements [9][10].

The visualisation shows the total number of games in the favoured-by-critics cohort, broken down by publisher for the years 2013-2017 inclusive (Fig. 5).

It enables the reader to understand the composition of the favoured-by-critics cohort by publisher, thereby providing insight into the publishers which critics may favour. Additionally, and of equal importance, a reader can see the trends over the time period that microtransactions have become more invasive in video games. [11]

To emphasise the most important information, only the top five publishers by total count of games in the favoured-by-critics cohort were shown as individual series. The remaining publishers with games in this cohort were grouped together as a single series.

To reduce the bias of the data being presented, games with a very small number of reviews that fell into favoured-by-critics cohort were excluded to prevent the inflation of the counts. Additionally, the axes display the full range of the data.

A. Justifications

A stacked line chart was used since it effectively shows both the composition of the favoured-by-critics cohort over time and the overall trend.

Because the number of sub-series is small, and the variations from year to year in each sub-series are simple to follow, the downsides of using a stacked chart, which are that it can be difficult to follow the trend of a sub-series with large variance, and it can be difficult to compare the values of different points, were lessened to a large degree.

Written labels were added to each series so the user could identify the publisher illustrated, without requiring a legend. To help cement the message of the visualisation in the mind of the reader, harmonious shades of colour were used, as it has been shown that this kind of colouring can help improve the recall of information from visualisations at a later date [12]. The shades were based on the red colour associated with the favoured-by-critics cohort, as a way to bring to mind the connection of this visualisation with that class.

B. Strengths, Weaknesses and Improvements

This visualisation has the unfortunate downside that it makes it difficult to gauge whether the prevalence of publishers in the favoured-by-critics cohort is substantially different from the general population, due to the lack of context. To mitigate this problem, the visualisation could be paired with another chart which illustrates the publisher mix in the general population.

On the positive side, the visualisation enables the reader to view the values of all relevant points and get a sense for the overall trend over time. Additionally, the author feels the graph is visually striking and clear in its message, which should help to drive home the main claims of the article.

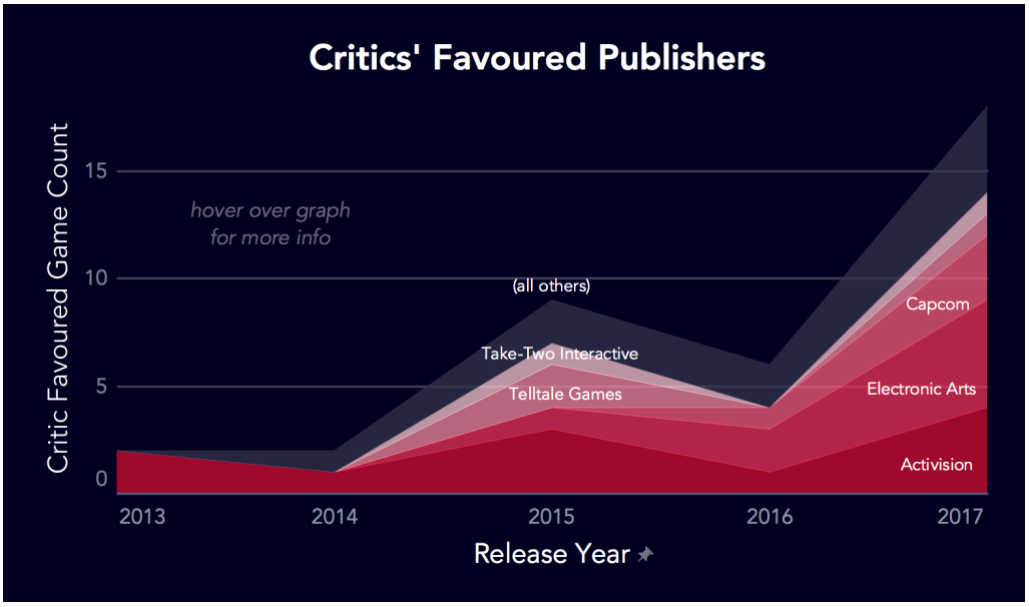

Fig. 5. *Stacked line chart showing the composition of the favoured-by-critic cohort by publisher and year of game release.*

VII. Conclusions

By collecting critic and player ratings, and combining them with publisher and genre metadata, it has been possible to present an original piece of analysis on the relationship that exists between video game players, critics and publishers. The claims presented on the role of microtransactions in player rating may only have circumstantial evidence as backing, but they provide a stepping stone to further discussion on the topic.

In future work, in order to describe with more confidence the relationships that exist between critics and publishers, the ratings given by each critic to each game would be required. This information is available online, and could easily be scraped. Because of the network-like nature of the data (since many games are rated by critics, and many games are published by publishers), graph visualisation tools such as Gephi may be useful in uncovering latent relationships.

References

[1] E. Tufte and P. Graves-Morris, *The visual display of quantitative information*. Cheshire: Graphics Press, 1983.

[2] N. Yau, *Data points: Visualization that means something*. John Wiley & Sons, 2013, pp. 202-203.

[3] Z. W. Pylyshyn, "What the mind's eye tells the mind's brain: A critique of mental imagery." *Psychological Bulletin*, vol. 80(1), pp. 1–24. 1973.

[4] A. Bompas, G. Kendall and P. Sumner, "Spotting fruit versus picking fruit as the selective advantage of human colour vision," *i-Perception*, vol. 4(2), pp. 84-94, April 2013.

[5] R. Mehta and R.J. Zhu, "Blue or red? Exploring the effect of color on cognitive task performances," *Science*, vol. *323*(5918), pp.1226-1229, 2009.

[6] W. Weiten, *Psychology: Themes and variations*, 7th edition. Belmont CA: Thomson Wadsworth, 2007.

[7] W.S. Cleveland and R, McGill, "Graphical perception: Theory, experimentation, and application to the development of graphical methods" *Journ. Amer. Stat. Assoc.* vol. 79(387), pp. 531-554, 1984.

[8] J. Heer and M. Bostock, "Crowdsourcing graphical perception: using mechanical turk to assess visualization design." *ACM Proc. SIGCHI Conf. Hum. Fact. Comp. Sys*, pp. 203-212, April 2010.

[9] Electronic Arts Inc., "Electronic Arts Inc. Fiscal Year 2017 Proxy Statement and Annual Report," 2017. Available: https://goo.gl/Rtcrqp

[10] Activision Blizzard, "Activision Blizzard 2016 Annual Report," 2016. Available: goo.gl/NEVHWL

[11] J. Švelch, *The Discourses of Microtransactions' Acceptance and Rejection in Mainstream Video Games*. The Evolution and Social Impact of Video Game Economics. London: Lexington Books pp.101, 2017.

[12] T. Sanocki, N. Sulman, "Color relations increase the capacity of visual short-term memory," *Perception*, vol. 40(6), pp. 635-648, 2011.