Visualization tool for Gaming Data

M. Basting[1] and G. Charalampidis[1] and M. Zeilstra[1]

[1]TU Delft, Netherlands

Abstract

A Visualization was made for 12 steam groups corresponding to countries. The visualization consists of graphs showing the relative playtime of the top games per group, the individual rating and playtime per game, and, the relation between price and rating. For this the heat map, stacked bar chart, scatter plot, pie chart and bar chart idioms were used respectively.

1. Introduction

In this report, a visualization tool is presented which makes it possible to analyse various statistics about games, game-genres and specific groups. This was done based on a Visualization Design Study using Munzners 4-level analysis [Mun15]. In this section, the first two steps of the analysis will be discussed while in the method the visual encoding and interaction idioms together with the algorithm will be further explained.

The first step of Munzners 4-level Analysis is to determine the domain. The dataset chosen contains data for games and genres grouped by steam groups, roughly corresponding to 12 countries. There is one main situation block which consists of people looking to find information about playing times to ask questions about the popularity of certain games, games in general or specific genres. People in this context could either be game developers looking to develop a new game or video game players searching for a new game to play.

The second step is task and data abstraction. The dataset in question can be seen as a multidimensional table with the keys: country group, player, genre and game. The country group attribute refers to a large steam group linked to each country. An assumption has been made in which users would only join a country's group if they belong to the corresponding country. The keys have the following attributes: own game, playing time, price, positive-reviews and negative-reviews. Own game is a simple binary categorical attribute, whereas the other attributes are numerical. From the domain questions can be concluded that the main goal of the visualization tool is for users to analyse the data by exploring the data by looking at their favourite game, genre, or geographic location. A lower level goal of the user is then to do several searches through the dataset and query the individual items in order to compare, identify or summarize all possible categories or games in question. The main targets the user will be interested in are attributes such as playing time, but they could also be interesting to see outliers such as games which are played substantially more than others in their respective categories, leading to the following sub-questions.

1. What is the fraction of users that played a certain game given that they own it?
2. What is the fraction of positive reviews per genre or game?
3. What are the relative playing times for the top 5 games in general per country?
4. What is the playing time per genre or game?
5. What are the most played games in general and for each country group?
6. What is the distribution of positive ratings for different prices and genres of games?

2. Method

To answer the sub-questions first the dataset needed to be gathered since it did not exist prior to the project. This was done using python scripts making requests to the steam API and SteamSpy. Since the chosen dataset is also relatively large 12000 users and 5000 games which are about 50mb of data over 5 data files, we decided to make the entire project in 1 HTML file using JavaScript to hide content which is not necessary for the current "Page", thus preventing the need to re-download the dataset when changing pages and allowing for animations. It was decided to use 4 types of each fully responsive subpages each focusing on a different subset of the data but with similar idioms, the landing page containing aggregated data for the entire dataset and pages specific to a genre, group, and game. To directly navigate between pages for the different groups, genres and games, a search bar was made poly filling all possible items in our dataset, But clicking on graph marks will also navigate to the corresponding page.

2.1. Landing page

The landing page contains a heatmap, stacked bar chart and a normal bar chart. This page is mostly focused on answering questions 2 through 6.

2.1.1. Heat map

In order to visualize the possible correlation between game prices and game ratings and answer the research questions 2 and 6, a heatmap is used. The x-axis contains all the genres which were detected in the dataset, and the y-axis contains price points starting from $0 for free-to-play games to games costing $50 or more. A design decision was made to round game prices to intervals of $5, assuming there is no significant difference between similar continuous values. Additionally, a very small portion of the dataset included games which had games with higher price than $50 and as a result, they were merged in the same group. Each heatmap point illustrates proportionately the percentage of positive reviews with a darker colour of the selected colour spectrum, as well as a gray colour for undefined entries.

2.1.2. Stacked bar chart and bar chart

The stacked bar chart and normal bar chart are linked very closely and are used to answer questions 3, 4 and 5. The stacked bar chart is made such to have the groups as labels on the x-axis and the percentage of playtime of the top N on the y-axis. This N can be adjusted to different values to reduce clutter occurring at higher numbers which show more detail. The glyphs are simple bars which are subdivided by coloured bars. The hue of a bar is consistent per genre which allows for easy comparison of genres at first glance, the vertical position shows at which position of the top N for the group a genre is, while the length shows the percentage of total playtime of the top N games. Hovering over a mark shows a tooltip with the genre name and absolute playtime in minutes. It also highlights marks representing the same genre by applying a lightness filter over the marks across all visible graphs. This allows for an efficient comparison of the position in the group top N and the global top 10. The bar chart is a simple bar chart consisting of the genres on the x-axis and playtime in minutes, with the marks using length to represent their value. The goal of this graph is to provide context for the stacked bar chart. Of course, clicking on any mark in either graph takes the user to the corresponding genre page.

2.2. Genre, Game and Group pages

The group, genre and game pages are all very similar, containing a simple bar chart displaying the top games in the collection in case of genre and group, and number of players in case of a game page. Besides this, the genre page also contains a stacked bar chart which is exactly the same as the landing page version, but now contains the top N games in the selected genre. The game and genre pages also have pie charts, and the genre page contains a scatter plot.

2.2.1. scatter plot

For each genre page, a connected scatter plot is rendered. Along its x-axis, the price of games are displayed in intervals of $5, and on its y-axis the percentage of positive ratings are shown. The scatter plot is featured in order for the viewer to get a more in-depth look at whether the price of a game in a given genre has an impact on how well it is received by the public, and help in answering research question 6. The scatter plot is connected in an attempt to potentially identify a trend in terms of positive reviews as the price increases.

2.2.2. Pie chart

These pie charts are connected to research question 1 and 2 respectively. The genre and game page contain a pie chart that shows the fraction of positive to negative reviews in a clear way. A percentage score could also have been used instead, but using a pie-chart additionally shows a direct representation of the relative size of the fraction. Showing this pie-chart of a genre makes it possible to get a rough estimate of the user satisfaction of the genre as a whole. In addition, the individual game page contains a pie-chart which shows the fraction of players that own and have played the game against the fraction that have only bought the game.

3. Evaluation

One of the initial goals was to visualize what fraction of people who own a game have played it in the past. This task is achieved through the use of pie charts in each game's individual page by showing exactly this statistic at an easily accessible position. In terms of how highly games and genres are rated, meaningful insight is provided by the heatmap by showing it per price and genre, the scatter plots showing the rating per price on the genre page, and the pie charts which show the rating per game. In the cases where more than a game belongs to more than one genre it was counted for all genres since there is no way to decide which genre is "best", besides this also a design decision was made to group games into price ranges and display their mean value along with the number of games of that group. There is an argument to be made in favour of keeping the price values as continuous for each game, however due to the wildly vary sample size per genre, grouping them was deemed as the better option to keep the graphs less cluttered. In terms of individual games, a simple pie chart has been deemed sufficient to visualize how well received a game is. Another research goal of this project was to discover and compare the most popular games for the selected country groups. This goal has been achieved by the use of a stacked bar chart on the landing page and on the respective genre pages these bar chart achieve their goal overall well, however the different games can sometimes get as high as 20+ games, this means that no existing categorical colour scheme can handle this in a satisfactory manner, leaving the hovering animation the best option to differentiate between games. Additionally, one of the primary focus of this project was to explore the play time of the selected genres and games, which is addressed with the use of bar charts on the landing page, and respective genre and game pages. Finally, the distribution of positive ratings for different prices and genre of games is covered by the heatmap on the landing page.

4. Discussion/Conclusion

Overall, the primary goals of the project were achieved considering the limited data set, API restrictions, as well as the given time frame. Nevertheless, it is important to discuss the relative emptiness of the group pages, due to lack of time, we could not fully finish the page, and it currently only consists of a bar chart showing the top games from that group. In the future, this could be improved with graphs showing, for example, the average rating this group gives, or the average price of the games of the group. The implementation is available at `https:`github.com/MikaZeilstra/ `DataVisualization`.

References

[Mun15] MUNZNER T.: *Visualization Analysis and Design*. AK Peters Visualization Series. CRC Press, 2015. URL: https://books.google.de/books?id=NfkYCwAAQBAJ. 1