

Figure 2: “Movies Critics Loved, But Audiences Really Didn’t”

I. Question 1: Analysis

a. The goal of the graphic

The goal the visualisation is to show the movies that critics loved but the audience did not appreciate. This was shown by a percentage measure so-called percentage gap or % gap, which is the difference between the rating given by Rotten Tomatoes’ critics and that of Rotten Tomatoes’ audience. The genre, the year of release, more importantly, the budget of each film were added to encourage further interrogations, possible normalisation and relative judgments

b. Dimensions visualised versus available dimensions

The original dataset has seven (7) rows and ninety-two (92) columns. This makes it a 7-dimensional data with 92 instances. The seven dimensions are as follows: *Film (name)*, *critics’ rating (%)*, *audience rating (%)*, *percentage gap (%)*, *the budget of the movie (\$ million)*, *year of release*, *genre*.

The film name, the genre and the year are categorical, the rests convey magnitude.

Contrary to the first 6 dimensions, the last dimension namely the genre may contain multiple dimensions. Thus, a movie can be a drama and action at the same time. This leads us to the first negative critique of the visualisation. In fact, the graphic only showed the first genre for films that have more than one genre. This was surely done for simplicity. Given that the genre was shown using colours, it would not be effective to shown all the possible combinations (More than 30) available using colour. Although this is reasonable because the primary goal of the graphic is to show the gap of each film, the target audience may be interested in querying based on the genre. This is in violation of the principle that good visualisation should convey correctness, and the full truth.

Besides not showing all the genres, the author did not include the actual ratings in the visualisation: only the gaps were displayed. One may argue that we are just interested in the divide, however, having the actual ratings could inform better interpretation of the data. For instance, the movie **Shaun the Sheep Movie** received a rating of 96% from critics and 80% from the audience. Its % gap is therefore 19, relatively ranking 22th (The exact ranking is from the dataset). This could be interpreted as not being “really” loved by the audience chiefly because of the title of the visualisation. Most films with a smaller gap were less appreciated. Thus, showing the critics’ and audience’ ratings should be considered.

Some movies present in the dataset were not visualised. Such example is *It Comes At Night*. The reason for this decision is unknown but represent a sub-representation of the dataset.

In sum, the graphic displayed less than 72% of the available dataset.

c. Important visual queries

The important visual queries identifies are listed next.

1. *What is the % gap of a given film X?* (X is the name of the film)

This query requires finding a particular movie fairly easily on the graphic. This is unfortunately not the case unless the year of release and/or the genre of the film is known by the user performing the query. Thus, grouping by genre with distinct colours and ordering by year is very effective. Finding a particular movie's % gap is however fairly difficult. Alphabetical ordering could make this easier.

2. *Which film or genre has the biggest or smallest % gap?*
3. *What is the budget used to make a given movie X?*

This query may be useful to estimate the resources allocated to the making of a particular movie in relation to how it was received by the audience and the critics. Thus subsequent queries could follow. Such examples are

- *What is the biggest budget movie with the largest review split between critics and audience?*
- *What is the smallest budget movie with the smaller review split between critics and audience?*

This was relatively well addressed except the use of area, which is hard to quantify for relative judgements. This will be further expanded when dealing with the visual channels.

4. *What is the genre of a given movie X?*
5. *What is the average % gap for a particular genre Y?*
6. *What is the % gap progression over the years?*
7. *What is the average % gap?*

d. The use of visual channels

Use of colour and spatial region

Colour was effectively used as an identity channel to categorise movies of the same genre. This is effective because colour hue is the second identity channel for categorical variables according to Muzner. This grouping greatly helps the user in answers all the query relating to the genres such as inter and intra genre queries.

Colour was also used together with spatial region to group the circle and the name of each movie. Thus, two channels were used to achieve the same. I may have been preferable to use just spatial region because using colour for the film name makes searching for a particular movie difficult. It requires constantly switching between different colours and makes the graphic unnecessarily colourful.

The choice of distinct colours was generally well done. Most categories were distinguishable except the comedy and action/adventure where the yellow interferes with the orange. This is because the two colours are not far apart. One of the categories could have used red which was not used. It is worth noticing that some genres were combined for simplicity. This does not compromise the ability to answer the main visual queries but may prevent further investigations for an interested user.

The luminance and saturation of most colours are identical, this is not good for a colour blind person. However, given the dark background, bright colours were required to avoid interference. This fact was well used except for Horror movies which were represented by a grey colour. This was a poor choice because of the large interference from the black background.

Use of position on common scale

According to Muzner, position on common scale should be the first choice as magnitude channel for ordered attributes. Given that the year and the % gap are ordered attributes, this principle was effectively applied. Regarding the efficacy, the % gap is one of the most important dimension, it was therefore well applied.

Showing the % gaps per year helps one query on the progression of this measure over the years.

Use of size and area

The budget of each movie was represented by circular area. Although this helps the user to answer part of the query 3, it is not very effective for two reasons. Firstly, it is very hard to compare two movies with relatively close but different budgets. This is true for movies of very small budget. Secondly, relative judgement is the primary focus however, one may be interested in the approximate budget values. The graphic fails to show any figure on the key of the circles, furthermore, it is extremely difficult for human being to accurately estimate area: we tend to underestimate it as opposed to length. It may be more effective to show the budget on a bar chart, next to the main visualisation.

Use of space

The film names were mapped out on the 2D central space provided by the combination of 1D year ordering and the % gap scale. This was an effective choice because, the 92 film instances better fit on a larger space. Showing the movies on a common line based on their year of release is very effective because it allows one to find a particular movie while searching if the movie year of release is known. The circle objects and the film names very well placed on the 2D area to avoid any overlap that may lead to confusion in querying a movie name.

Use of depth

The budget key has some element of depth to it. This should have not been the case because the aim was to show relative size however, the depth could have been interpreted otherwise. The small circles in the key seem farther away. It may have been a way of giving more importance to the movies with bigger budgets. This is not very intuitive.

e. Use of design heuristics for visualization

The visualisation did not make use of any 3D representation. This was a good decision because there was no reason to do so.

The budget is a 1D attribute, however, it was represented with an area of 2D (circles). This was done to take advantage of the 2D space used for the %gap. An alternative could have been a separate bar chart with movie names as columns and budget as height. The problem with this comes in being able to show all the 92 instances on a 1D dimension. It was attempted in the synthesis.

f. Scene gist

The graphic does not have any visual/graphic to draw someone's attention at first glance. Although, colour is fairly quickly perceived, one may not be able to know that it is about movie ratings until the title and subtitles are read. The colours however makes the visualisation lively. The gist was somehow attempted in the fairly catchy titles and axis' labels. After realising that it is about films, then the dark background together with the various colours may be linked to a cinema or film. It may be better to add a graphic to convey the idea of film at first glance.

g. Pop-out and attention

The Last Jedi's circle was not filled. It was also connected to a clear text by a dotted line. This was a very effective way of making it pop-out. It helps one precisely answer one of the important visual queries regarding the biggest budget movie with the biggest divide between critics and audience' opinion, in a split of a second.

h. Title, axis and labels

The years were shown on the y-axis. This is a counter-intuitive design decision because it is more natural to display timeline horizontally. Thus, answering the query on %gap progression over the years may not be as quick as if one was progressing from left to right.

The ticks shown on the % gap axis oddly kept to just 3 values. This makes estimating a given movies % gap harder. Also, contrary to the dataset which shows up to 45% gap, the axis only shows up to 40% gap. This confirms that the full dataset was not plotted.

The y-axis does not have any label. It is better to assume that the user does not know that it is about years. Also, the axis's label is biased because it only represents the films that were not liked by the users. This may mislead the user in his/her interpretation.

The title does not do justice to the dataset. It suggest that a great divide in the critics' and audience' opinions implies that the audience did not appreciate the film. This is not the case as demonstrated above on page 1. The following title would be both catchy and more accurate: Movies Loved by Critics, But Not so much by the Audience. This subtle nuance do

General conclusive analysis

In general, the design is effective because it allows us to answer most of the visual queries. However, is it far from perfection because of the problems highlighted above. The major problem concerns the

ability to quickly search for a movie name, the ability to represent all the dataset, and the budget only represented by area. Most of these issues with other improvements are addressed in the synthesis below.

II. Question 2: Synthesis/Redesign

Because the graphic fairly answers the important queries, the improvement option was chosen. It was redesigned using the Tableau software and the necessary changes were made. **Figure 1** shows the redesign and the changes are explained next.

a. Improvements

Additional dimensions and data instances

The actual rating of the critics and that of the audience were added to the visualisation for completeness and allow better interpretation of the dataset. They were added as numerical text below the name of the movie. The critics' rating on the left and that of the audience on the right, separated by a vs to convey the idea of comparison. They were added with slightly different colour intensity matching the colour contrast in the title to facilitate identification. This is based on humans' ability to recognise patterns and colours easily to convey. It might require learning but once learned, it is fairly quickly recognised.

All the 92 instances of film available in the dataset were visualised

Axis switching

As explained above, the year axis and % gap axis were switched to allow for a more familiar and natural progression of time from left to right instead of bottom to top progression.

The axis were consequently renamed with less biased labels. This is to convey a partial truth without manipulating the data.

All first genres added

Instead of adding all the possible combinations of genre, the first genre was chosen for the films with multiple genres. However, the genres were not aggregated, each genre had its own colour. This introduces an additional channel to choose distinct colours. This was done successfully by carefully using variation of saturation, luminance and the distinct colour channels. Furthermore, the grey colour was removed to avoid interference with the black background. The red channel was introduced and used as a mean to make the biggest budget movie with high % gap pop-out in addition to its size. This is because we are very sensitive to red.

Budget bar chart

One of the biggest improvement was the addition of a bar chart to effectively represent the film's budget. This decision was to solve two very important issues. Firstly, it facilitates the search of a given movie's name. This is because the bar charts is ordered alphabetically based on the film names. This can speed up locating a particular film if its genre, its % gap and its year of release are unknown. So, a query by name. Secondly, it gives exact budget values and a better relative comparison of movies' budgets since humans can exactly interpret length. The area was not removed because it might be useful to answer quickly queries tightly linked to the budget and the % gap without having to switch between two visuals.

The colour grouping for the circles was also used for the bar chart to facilitate correspondence and genre grouping.

This additional feature introduced another challenge. Being able to represent all 92 instances with their names without overlap. This was dealt with by leaving out some names. This does not particularly affect the primary queries, also the movie names are ordered alphabetically.

Use of Occlusion

Occlusion was used to be able to represent all the movies, this was effectively used at the expense of some film names not appearing to avoid overcrowding.

Gist: graphics

To improve the gist, two pictures were added to the two visualisation next to the two major titles. The first picture is the rating of The Last Jedi from Rotten Tomatoes. It attracts any movie lover because of the popularity of Star Wars and tells the user that the graphic is about movie ratings. Besides, it helps emphasis that The Last Jedi was not loved by users even with a large budget.

The second picture is to convey complement the title of the bar chart.

Other improvements: better colour, axis labelling, title, size key with figures

Axis were labelled correctly and partially. The size key now has figures, the colours are brighter to minimise interference from the dark background. The colourful film names were removed to ease the search of names.

b. Why it is a better design and Strengths

In addition to keeping most of the positive aspects of the initial visualisation, it added many improvements. This design is better because it shows more data dimension and every instance in a similar way with overcrowding the visuals. It also has a better gist and added more channels such as occlusion. It solved the axis problems. It allows faster query and answer process. The exact budget values, the exact critic and audience ratings could be queried. Also, the titles and the axis labels were more accurate and impartial.

c. Weaknesses

The cost of showing all the instances was to have some movie names not shown. Also, searching for a film name was made faster but was not solved entirely. The biggest 3 sizes of the area key are not very different. This is not very problematic because the bar char gives the exact values.

d. Trade-offs

Effectiveness was chosen over completeness when it come about showing all the genres for a given movie. This is to allow colours to be distinguishable because the colour channel has a limited number of distinct bins.

The names of the movies could not be ordered in the 2D space because ordering is dictated by the year of release and the %gap. The price to pay was being able to show all the instances at the expense of relatively slow search process of movie name.

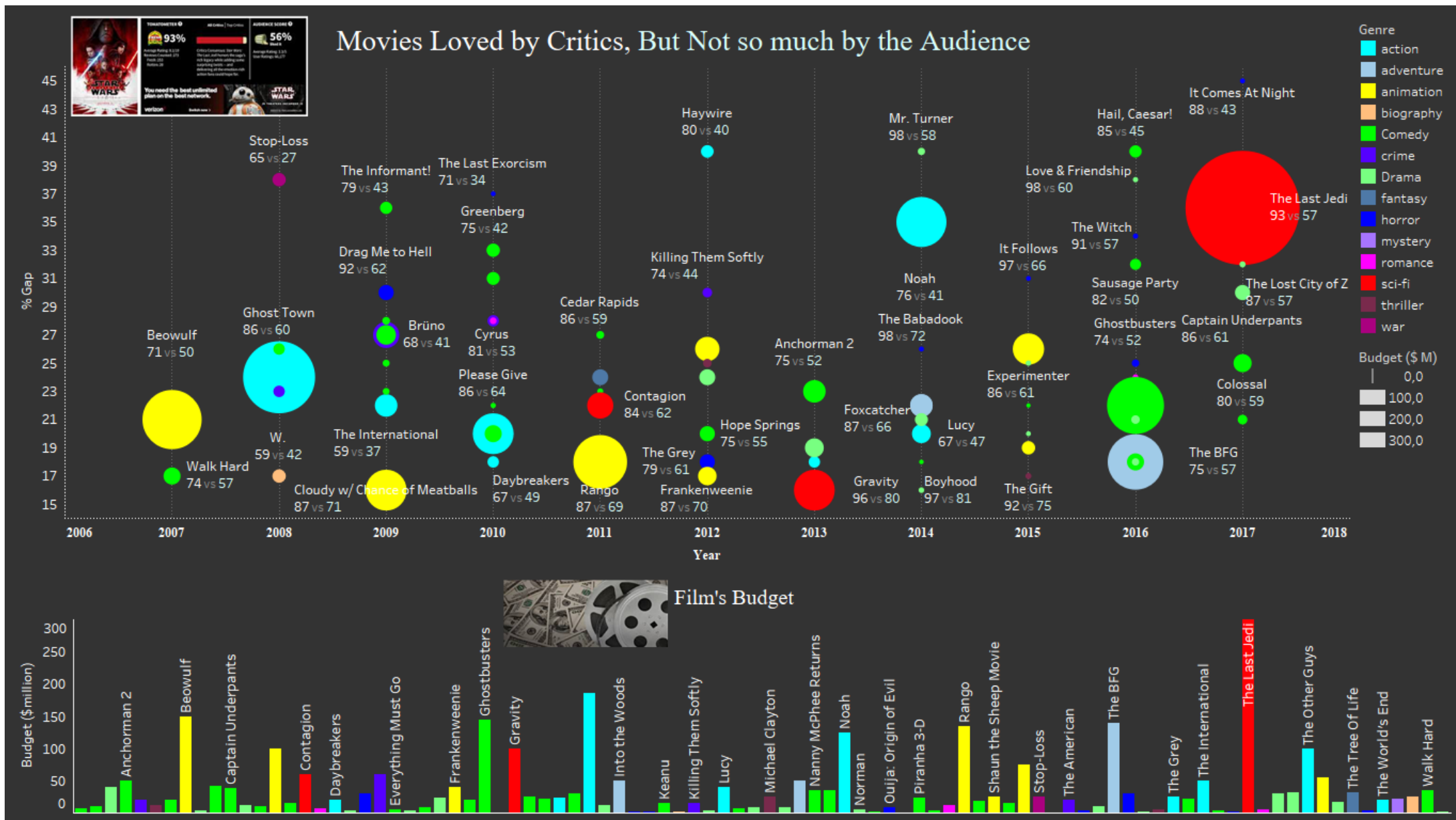


Figure 1: Improvement of "Movies Critics Loved, But Audiences Really Didn't" by David McCandless et al.

