Assignment 4 – Data Visualization
Abhinethri Tumkur Umesh (1213187714)

1) Who are the audience?
This visualization has been targeted for movie lovers and makers to get information such as production company, genre, IMDB rating, budget, gross and Oscar information for top movies.

2) Where is the data source?
Source: https://www.behance.net/gallery/44767671/100-Greatest-Movies-Data-Visualization

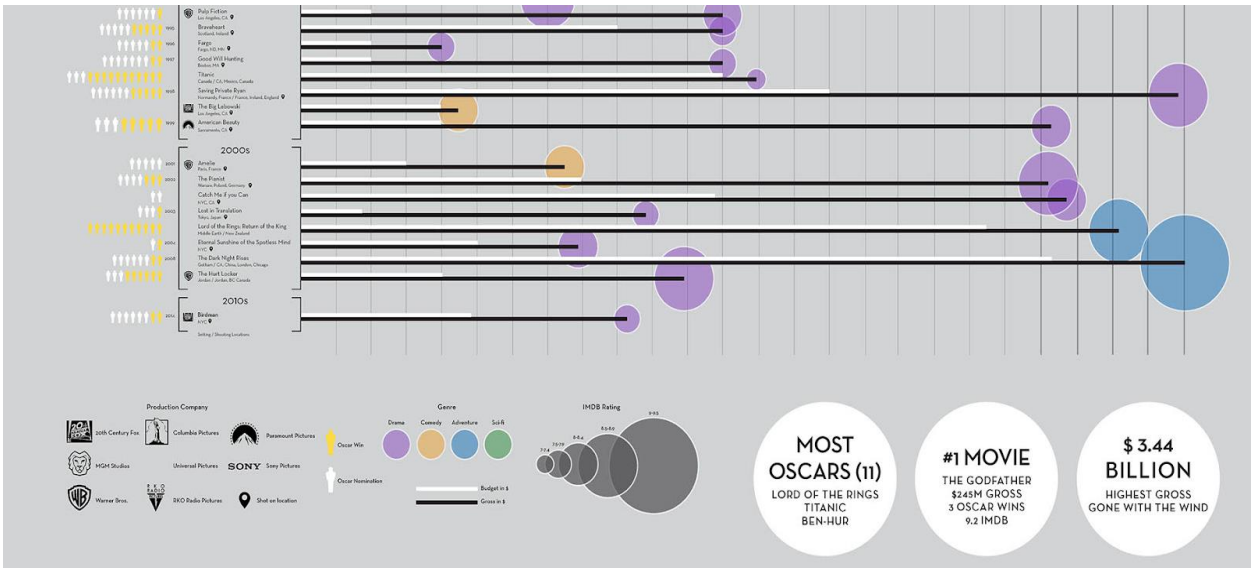
3) Strengths and weaknesses?
Strengths: The topic is great as there are huge number of audiences in this field and very less number of visualizations. The information which user will be able to capture in few seconds, apart from knowing top movies are number of Oscars won, total income and an option to compare the top movies based on the ratings which is represented through varying bubble size. The combination of Genre and IMDB rating in the form of bar graphs and bubble chart is effective and the color combination for genre and Oscar info is appealing.

Weakness: Too much of information has been squeezed in a single page, author has mainly cluttered the representation of the left part which has info about year, location and Oscar info so viewer must zoom in to check this info. Gross represented in the black bars are overshadowing the white budget bars. Some information could have been consolidated and added it to summary page. Also, this visualization is difficult to scale when new movies have to be added.

4) What are your opinions?
I believe author has failed in the execution of a great idea. This visualization should have helped movie lovers to pick movies to watch from the top list based on their interest and the movie makers to strategize their upcoming projects, but it has not succeeded. She could have pulled much more information from IMDB database, researched and referred good visualization for representation. On hovering, extra info could have been displayed or information could have been exposed through navigation. One more way is to divide the single page into multiple sections and to come up with catchy visualizations through infographics.

5) How would you design it differently?
I would have created interactive visualization based on years and represented the information through layers. User would be provided with an option of choosing a range of 10 years (E.g.: 2000-2010) and a summary option. Each page would be having 2 sections, first section represents genre, budget and gross through bar chart and bubbles (as represented by author) and 2nd section will have clustered bubble chart (link), where each outer bubble represents a movie and inner bubbles represents the nominations and the colors tell whether Oscar has been won for that particular nomination. Finally, the summary page will have a pie chart which represents the percentage of top movies belonging to a particular production, a Google map chart representing location wise movie contribution and a bubble chart with varying sizes (link) to show top 100 movies which have won the highest number of Oscar awards.

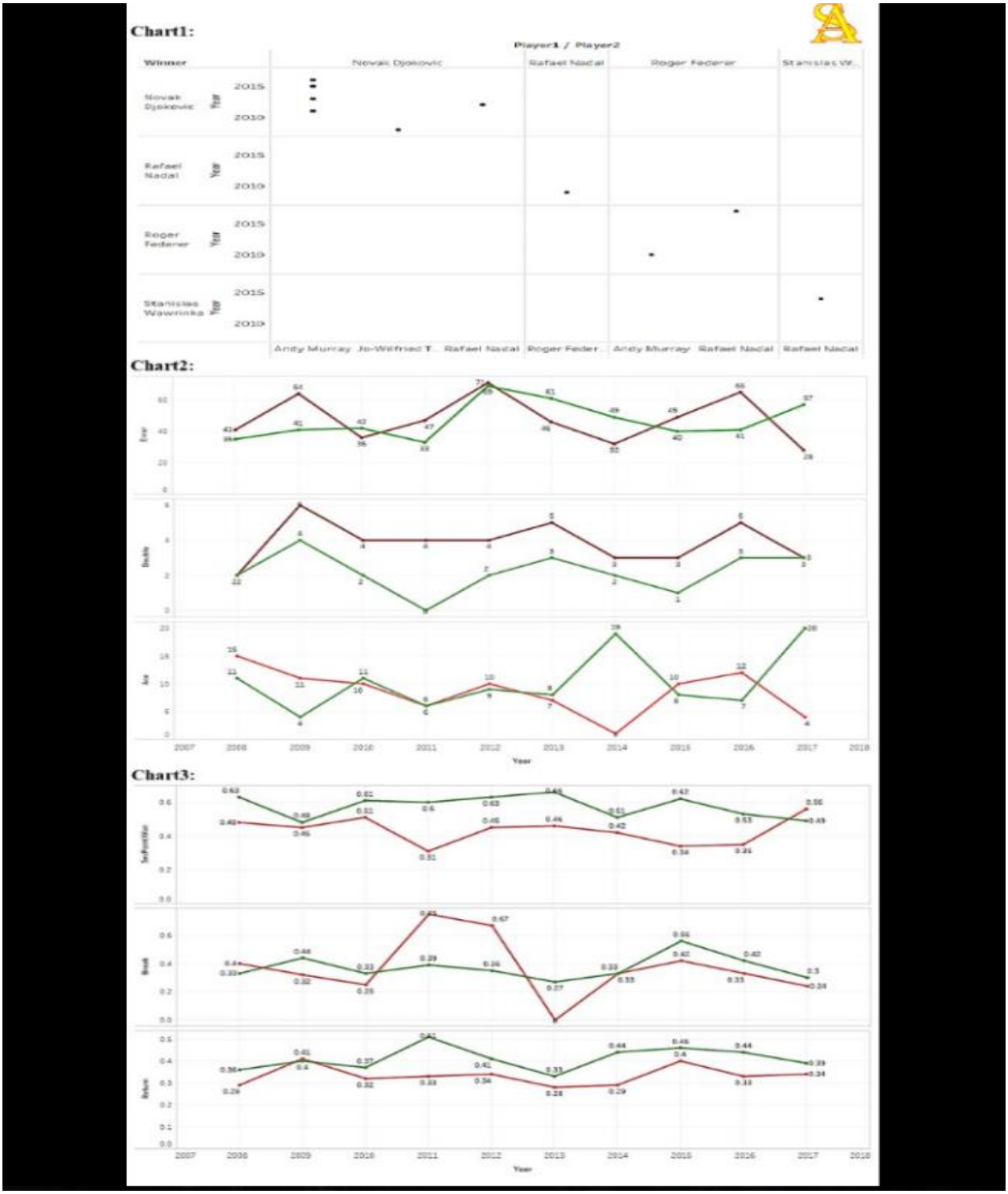
Visualization (Since lot of info has been represented in one image, the picture is not clear. Please refer the source for clear image):



Citations:

1. https://www.sas.com/content/dam/SAS/en_us/doc/whitepaper1/data-visualization-techniques-106006.pdf
2. http://jackcanfield.com/blog/visualize-and-affirm-your-desired-outcomes-a-step-by-step-guide/
3. https://hbr.org/2016/06/visualizations-that-really-work

Bonus (Refer Static Visulaization.png for clear picture):



Please find the visualization in the below link:
The above image is from Assignment 1.

1) Critics:
This visualization has too many graphs to tell a simple story, so viewers won't be able to capture the essence within few seconds and it takes time to understand the notion behind the above graph. The 1st graph which gives information about players doesn't tell any story and it shouldn't have been included. Out of remaining 6 graphs (For 6 attributes), the top two graphs should have been chosen after proper analysis.

2) Methods to Improve:
Since the data was less, I should have told the story with respect to 1 winner. I should have chosen the best attribute which contributed for the winning and compared with losers to find one good pattern. I should have used bar graphs for comparison, with better colors to represent the bars for losers and winners (Blue for winners and red for losers).

BONUS QUESTIONS:

Q1) How do you compare word clouds?

The comparison includes the comparing of relative frequency of terms present in more than 1 document. (Even the difference between the usage of words will be plotted).

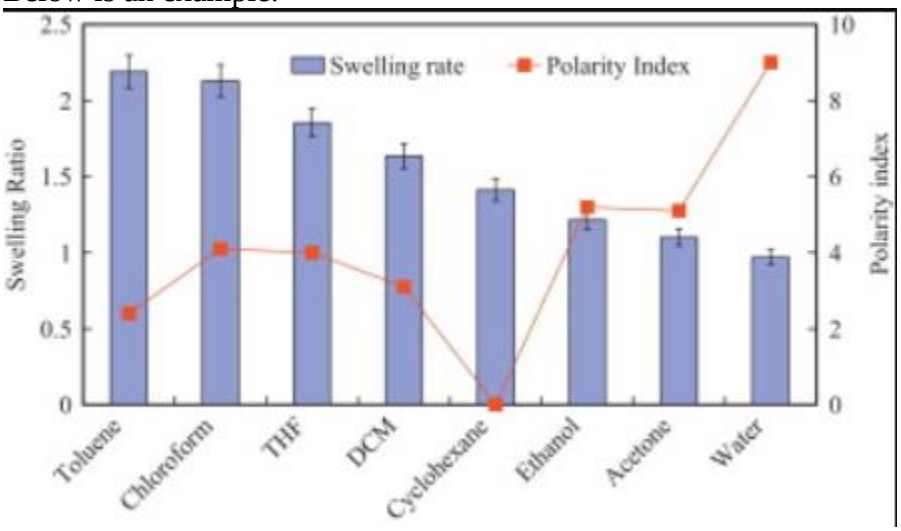
Example:

In a speech by 'XYZ' in the year 2008, word "success" was used 20 times. In next year, word success was used 11 times by 'ABC'. In the word cloud, "success" is mentioned with frequency of 9 on 'XYZ' side only because it has been cancelled out. Thus, 'XYZ' used the word 'success' more number of times.

Reference: https://www.youtube.com/watch?v=eRdVCWff8AA

Q2) What types of visualizations are good for showing polarities?

Below is an example:



Q3) What are both candidates first debate sentiment scores?

Debate Sentiment score for Trump: 15.4%

Debate Sentiment score for Clinton: 19.1%.