



Gender Representation of 20 years Malaysian Films based on Andy Kirk Dynamic Visualization Framework

Nur Naelah 'Ariefah Binti Nordin

Faculty of Computer and Mathematical Sciences, Universiti Teknologi MARA, Shah Alam, Selangor, Malaysia
naelahnordin@gmail.com, 2019850146@uitm.edu.my

Nasiroh Omar

Faculty of Computer and Mathematical Sciences, Universiti Teknologi MARA, Shah Alam, Selangor, Malaysia
nasiroh@uitm.edu.my

Article Info

Article history:

Received may 20, 2021

Revised Aug. 25, 2021

Accepted Sept. 02, 2021

Keywords:

Dynamic Visualization

Malaysian Film

Gender Representation

Genre

Box Office Success

D3.js

ABSTRACT

The role of visualization in journalism evolved when dynamic visualization was introduced. Previously, data visualization in media were constrained with the static nature of the media itself. Nowadays, front-end technology advancement allows for dynamic visualization included in journalism or as a standalone itself. Dynamic visualization ables to provide a compelling story and also effective in summarizing large amounts of data into useful insights. Using Andy Kirk's framework of creating effective visualization, this paper aims to provide insightful and dynamic visualization on the evolution of Malaysian films in terms of its gender representation, box office success, budget, and genre. About 809 movies produced from the year 2000 until 2020 were sampled and processed prior to being visualized. The visualization shows that while male centric movies have higher box office earnings, female centric movies perform better in relation to their budget, implying that female centric movies can potentially earn better in box office sales if given a higher budget. The paper demonstrates the design, exploration, and implementation processes for an effective dynamic visualization.

Corresponding Author:

Nasiroh Omar,

Faculty of Computer and Mathematical Sciences,

Universiti Teknologi MARA, Shah Alam,

Selangor, Malaysia.

email: nasiroh@uitm.edu.my

1. Introduction

Dynamic or interactive visualization[1] differs in important ways from the traditional forms of information visualization. Politicians, television reporters and journalists have started to incorporate dynamic graphics to create the dynamic visualization. However, crafting a successful data stories based on the dynamic visualization requires a diverse set of techniques and skills[2]. Despite the popularity used of dynamic visualizations, limited research have been reported in the literature that describe the structured methodology for conveying the stories in a way of dynamic visualizations.

In this paper, the structured approach on designing and implementing dynamic visualization is presented by focusing on gender representation in the Malaysian films industries. The government of Malaysia realized the importance of cultural products such as films as an income-generating industry and also an effective medium to promote local culture and heritage to the world[3].

Therefore, an outstanding quality and a competitive film industry are important in order to help the industry withstand. It is crucial to develop effective strategies as to ensure that this industry remain sustain to contribute the Malaysian economic growth. One useful way for achieving this agenda is by gaining the details insight of the films progress. Dynamic visualization for films are existed for many countries[4] but limited for Malaysian films. In this paper, the Malaysian films box success are narrated based different aspects, mainly with gender representation. Additionally, this paper provides the fundamental design and implementation techniques for the huge information analysis based on dynamic visualization.

2. Research background

Gender representation in films have been given wide attention recently. In [5], the researchers have investigated the linkage between gender characters with cultural portrayed in popular American films between 2002-2014. Focused on female representation in animated short modern films (2007-2014), researchers in [6] addressed the issue of gender in-equality from the 700 films. They focused on examining the influences of gender, race and lesbian,gay,bisexual,transgender (LGBT) status to the viewers' impression. Issues on wars films in Western cinema and the connection with male spectatorship have been raised in [7]. According to the researchers in [6], from *The Big Parade* to *Saving Private Ryan*, the war film has always been privileged to male spectatorship while behind the scene was an elite woman warrior who worked on the theoretically ideas on characters and sophisticated scripts. Therefore, they offered a valuable set of methodologicals and theoretical frameworks from their research findings to achieve equal gender representation in the success of films production. Recently, gender inequality has been experienced by women directors in German cinema that received less public funds for films production[8]. The researchers found that, although with lower budgets, most female producers work significantly more often with female authors and directors and their films are often more successful than the male producers.

Research on gender representation in Malaysian films are mostly qualitative, such as the research in [9] that investigating the representation of Malay-Muslim women's identity in Malaysian films. The researchers further highlighted sentiment finding on the female Malay-Muslim character representation and its role in portraying Islamic values such as purity and obedience. They are often portrayed as struggling in a patriarchal world and throughout the years, these characters have become more multi-faceted, complex and progressive. While this type of character seems to be empowering, it is still boxed in a stereotypical profile of the demure obedient damsel in distress and does not offer a more diverse character. Apart from that, even when female characters are written to be career-driven, it is often implied that their morality is compromised and is in need of a patriarchal institution to guide them back to the truth, which implies that independence should be avoided and this is not an empowering message for young women [10]. Aside from the damsel in distress, female characters are also portrayed as the polar opposite, the evil conniving villainous mother or antagonistic lover [11] which implies that a movie only needs two female characters and if a movie has more female characters than male characters, then the storyline must be focused on storylines that does not conform with the usual female stereotypes. To the best of our knowledges, there exists very few recent studies on box office success of the Malaysian film industry[12] and gender representation in relation to the box office success on Malaysian films is difficult to be found the literature.

Quantitatively, generalized linear regression and random forest model has been used to measure female representation and impact in films. The variables Bechdel test and female cast ratio is compared and the results show that films with better female representation are more likely to succeed and efforts need to be made in terms of budget and improve the upstream routine by including women in film industry gatekeepers such as directors, producers, and executives in movie studios in order to produce films with quality female representation [13]. Another research implementing multiple variate regression analysis found that the underrepresentation of female in films are due to the upstream routine of film industry gatekeeper such as production budget, franchise familiarity, popularity of the actors and release dates, not a lack of interest or demand by the masses [14]. Therefore, a visualization to summarize 20 years of Malaysian movies would benefit the film industry as well as the audience to give perspective on gender representation in Malaysian cinema.

The technology of visualization has evolved to becoming dynamic and more engaging. This allows the audience to participate and explore the data themselves. Consequently, there has been an increasing emphasis on effective visualization design by a having a balance with the journalism

aspect of the story [15]. An emerging type of data journalism is scrollytelling which takes advantage of the scrolling nature of online longform articles [15]. The scrolling event triggers animation or visualization which enriches the narrative and allows for the author to determine which media is shown in between which paragraphs thus controlling the order of what the audience read and interact with. Scrollytelling is not commonly used for instant reporting, instead it's preferred in more in-depth stories, particularly longform investigative journalism with complex data to be presented in simpler form. In line with the emphasis on the importance of visualization in storytelling, the design process is also emphasized. There needs to be a balanced process in designing and editorial thinking. This sentiment is reiterated by Andy Krik in his framework of creating effective visualization [16] that breaks down the process in different phases. There is also a growing concern of the acceptance and usability of dynamic visualization across audiences with varying computer skills and may not be able to appreciate the visualization fully [17]. Research has shown general acceptance of visualization in online newspapers [18] however the level of complexity should be limited and not overwhelm the audience. This result is inline with the opinion of [15], [16] that the design of visualization should have balance with the journalism aspect of the story and the designer should be mindful of the audience overall reception within that domain. This paper demonstrates the design steps focusing on gender centric representation in association to other aspects such as genre and the films box office earnings. Previously, the implementation framework for dynamic visualization of P.Ramlee movies have been introduced by using the similar framework [19].

3. Methodology

This study has implemented Andy Kirk visualization approach [16], which consists of five important steps namely formulating brief, data acquisition and pre-processing, and lastly data visualization development.

3.1 Formulating brief

Formulating brief phase is where the curiosities of the domain are discussed and the purpose of the artefact is determined. In this stage, it is important to identify the scopes of visualization through constructing research questions as the following:

1. How different is the proportion between male-centric and female-centric?
2. How different is the proportion between male-centric and female-centric in relation to movie genre?
3. How different is the proportion between male-centric and female-centric in relation to movie year?
4. How different is the proportion between male-centric and female-centric in relation to budget support?

Based on the research questions, the purpose of this project are:

1. To visualize the proportion of male-centric and female-centric movies in relation to their achievement for box office.
2. To visualize the proportion of male-centric and female-centric movies in relation to movie genre.
3. To visualize the proportion of male-centric and female-centric movies in relation to movie year.
4. To visualize the achievement (box office earnings) of male-centric and female-centric movies in relation to budget support.

The presentation of dynamic visualization was designed based on these objectives. Then, it is important to construct a purpose map as depicted in Table 1.

Table 1. The Purpose Map

		Explanatory		Exhibitory	Exploratory	
		Sequence/Drama	Annotate/Describe	Display	Manipulate/ Interrogate	Participate/ Contribute
Reading	Efficient/ Precision		X			
Feeling	Emotive/ Big Picture					

The output elements is explanatory, which can be presented with scrolling effect to allow for a specific storytelling can be narrated. Each movie is annotated and described for the audience's convenience to explore and interact with the visualization. In the next phase, data acquisition and data pre-processing is done to acquire the insights needed and discover new information that would be useful for the narrative.

3.2. Data Acquisition and Pre-processing

The data was collected by using web scrapping technique from the cached pages of FINAS websites in 2006 until recent. As the information is not available in the current website, researchers have to access the cached pages by using Waybackmachine, a software tool to retrieve snapshots of a website's previous version. Apart from that, this research also has collected the related data on Malaysian Consumer Price Index (CPI) from the World Bank website as to formulate inflation value based on the year of the movies production. Furthermore, the gender of actors has been identified based on the actors name provided in Wikipedia pages.

After data cleaning and filtering, three entity tables were created to locate the data for pre-processing. Table 2 is the first table that lists the the movie details that consists of 9 columns and 809 records.

Table 2. The variables of movie

Variable	Data type	Range	Description
Year	numerical	Integer	Represents the year the movie is produced Minimum: 2000 Maximum: 2020
Month	numerical	Integer	Represents the month the movie is produced.
Day	numerical	Integer	Represents the day of the month the movie is produced
Title	String		The title of the movie
Director	String		The director(s) of the movie
Cast	String		The cast(s) of the movie
Genre_final	Categorical		Category of movie genre. Categories consist of the following: Animation Historical Romance Documentary Action Comedy Horror Romantic Comedy Action Comedy Drama Horror Comedy Scifi/Fantasy Adventure Family Musical Thriller

Gross_final	numerical	{0, 48000000}	The box office earning of the movie in MYR, adjusted to inflation
Budget_final	numerical	{0, 20000000}	The budget of the movie in MYR, adjusted to inflation
Female_dist	numerical	{0,1}	Female to Male ratio in the cast based on Cast column with a maximum of 7 actors
Male_dist	numerical	{0,1}	Male to Female ratio in the cast based on Cast column with a maximum of 7 actors
Gender_Majority	categorical	{0:1}	0: female_centric 1: male_centric

The Gender_Majority was calculated based on a 50% cut-off point of the gender distribution. If the movie had above 50% male in their cast, the movie is labeled as *male_centric*, otherwise is *female_centric*. The second table consists of the actor details with two columns of 2241 records as presented in Table 3.

Table 3. Variables of actor

Variable	Data type	Range	Description
Name	numerical	Integer	Unique identifier of an actor
Gender	categorical	{0,1}	0 : Male 1 : Female

The CPI Malaysian index data was recorded in the table that consists of two variables as listed in Table 4. The number of records for CPI Malaysian index is 60.

Table 4. Variables of CPI Malaysian index

Variable	Data type	Range	Description
Year	numerical	Integer	Represents the year
CPI	numerical	{-0.41,17.33}	Represents the CPI value of that year

A lot of movies in the original data are labeled as of more than one genres. Malaysian movies in particular like to instill drama in various forms of genres including action and horror. Some of these movies even infuse three genres together and when labeled as a string, creates an immense variations as displayed in Figure 1.

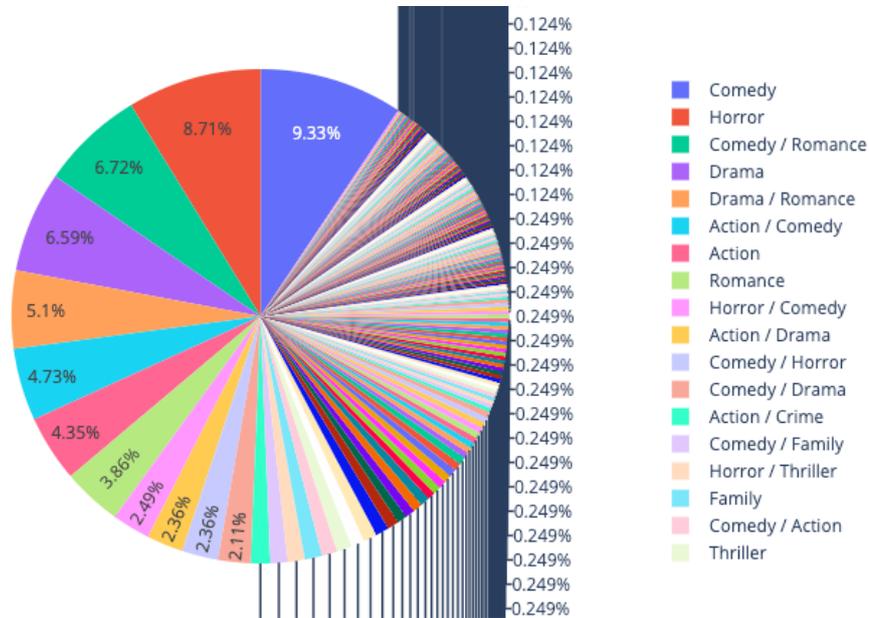


Figure 1. Pie chart of original genre distribution

The genre distribution from the original data depicted in Figure 4 shows that the movies with the same combination can be labeled as different categories according to the genre order. There exists numerous combinations of more than two genres, which are not labeled in the Figure 4. Therefore, it is essential to deconstruct these string of genres into different columns and limit the label with two genres only. Figure 2 shows the recategorization steps.

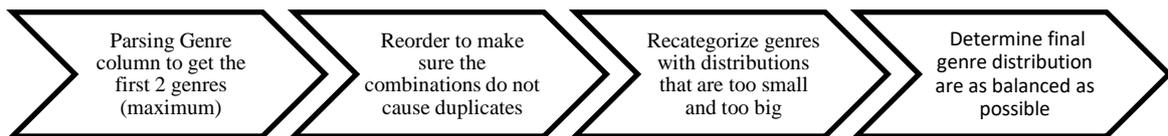


Figure 2. Steps to process genre of movies

The steps described above are to ensure that the genre distribution are balanced and can give insights. The genre final distribution are displayed in figure 3.

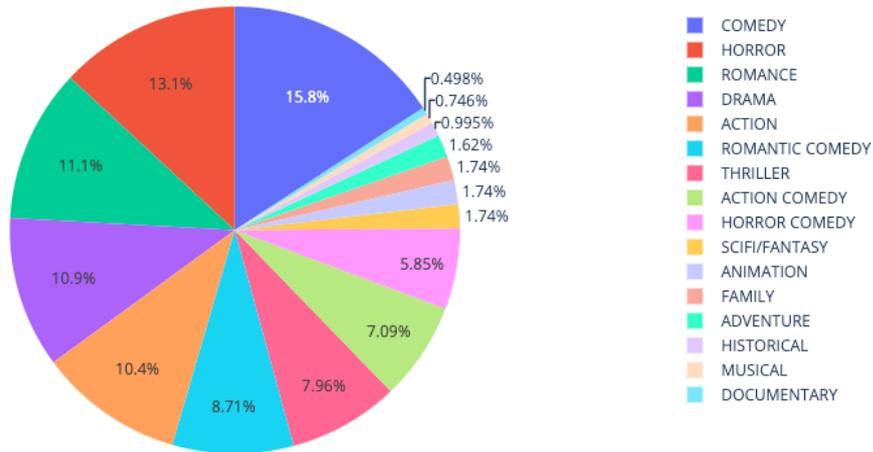


Figure 3. Pie chart of processed genre distribution

The distribution of genre were deducted with 16 unique labels. Furthermore, the following formula in Equation (1) and equation (2) were used to calculate the gender distribution.

$$Distribution_{male} = \frac{Male_{total}}{Cast_{total}} \quad (1)$$

$$Distribution_{female} = 1 - Distribution_{male} \quad (2)$$

Based on the male and femail distributions, the gender centric was determined by looking at which distribution takes up **more than 50%** of the cast. Figure 4 summarized the steps.

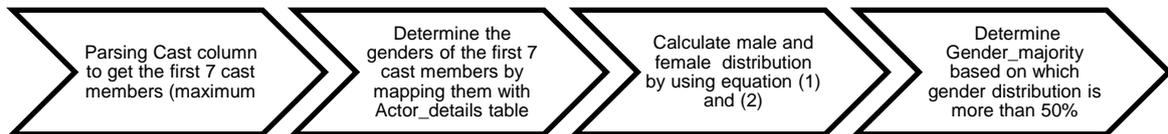


Figure 4. Steps to determine gender_majority of a movie

3.3 Dynamic visualization development

To develop the dynamic visualization, D3.js Javascript library[20] has been used. D3.js is a rapid tool useful for sophisticated data visualization on the web. Developed by Mike Bostock, D3 has full capabilities of web standards, including CSS3, HTML and Scalable Vector Graphics(SVG).

4. Results and Discussion

As depicted in Figure 5, the visualization consists of circles representing a movie. By hovering on each circles, a more details information about the movie will be displayed. Red circles are a collection of majority female cast while the yellow representing majority male cast. The size of the circles represents the box office sales of the movie.

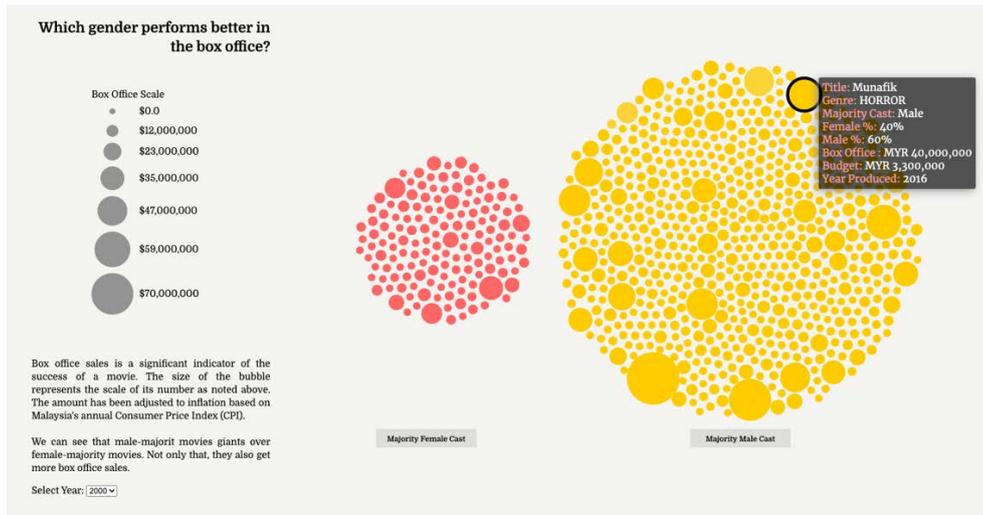


Figure 5. Visualization on gender centric in correspondence to the films' box office earnings

It is shown the Figure 8 that male has centric the Malaysian movie films from the 2000-2020 years. Additionally, most of the box office movies has been brought by the male-centric movies. Furthermore, the cricles were grouped in their respective genres as illustrated in the following figure 6.

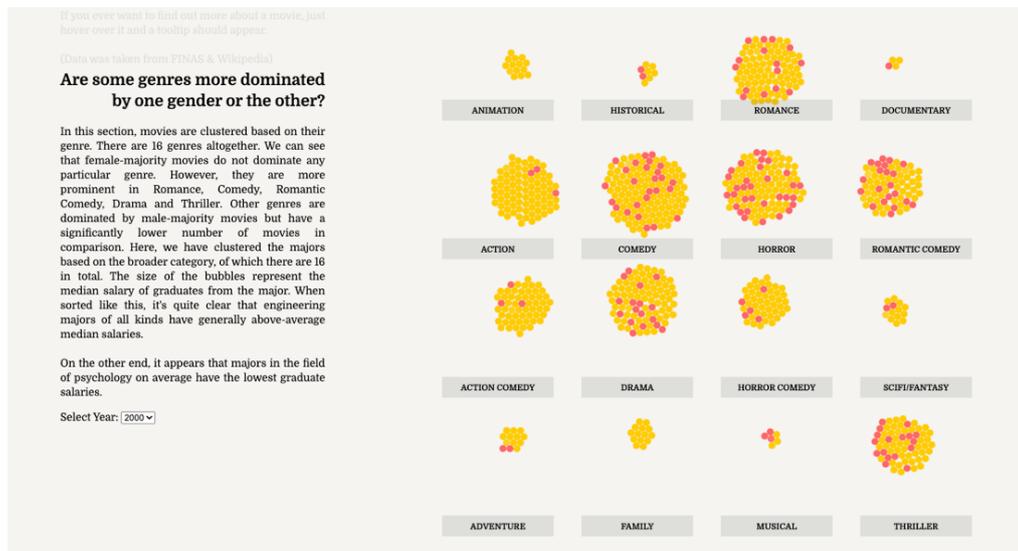


Figure 6. Bubble chart of movies grouped by genre

Family and animation movies are 100% male-centric. Moderate female-centric movies can be seen in romance, comedy, horror, romantic comedy, drama and thriller but the proportions are very few in action, action comedy and horror comedy.

Next, the evolution of movie distribution between the male-centric and female-centric can be analyzed in a barplot by year as presented in Figure 7. The circle size indicates the box office achievement.

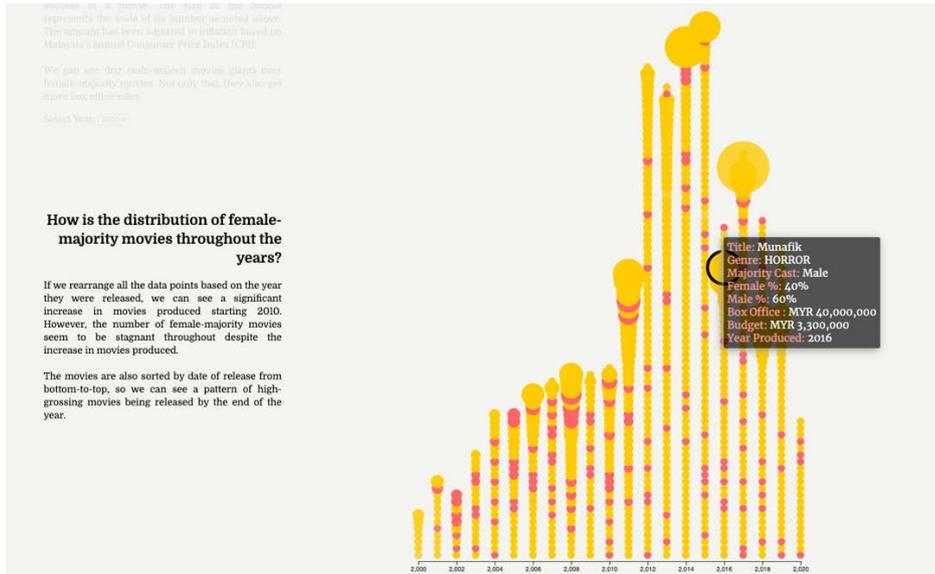


Figure 7. Distribution of movies produced by year

It is apparent that number of movies increased from the year 2000 until 2018 and drastically decreased in 2020. Male-centric movie in 2018 has been the most outperformed movie in the 20 years.

Finally, a scatterplot of movie budget against box office sales was plotted as presentd in Figure 8 and the trend lines for each gender representation was calculated in python programming using Ordinary Least Squares (OLS) method based on Equation (3) and Equation (4).

$$box_office_{male} = budget_{male} * 0.561884 + 2607780, R^2 = 0.070334 \quad (3)$$

$$box_office_{female} = budget_{female} * 0.999765 + 360517, R^2 = 0.382614 \quad (4)$$

This trendline gives insight on the achievement of male-centric and female-centric movies in terms of box office earnings based on the fund support.



Figure 8. Scatterplot showing correlation between budget and box office sales

It can be seen that male-centric movies have a larger range of box office earnings as well as budget, however their trendline suggest that their achievement is not highly correlated. A male-centric movie that has been provided with a big budget does not significantly achieved high earnings of box offices. Female-centric movies however shows a higher chance of performing well when given with a bigger budget.

5. Conclusion

The design and development of this dynamic visualization allows a clear view of the Malaysian film industry and where it stands in terms of representing gender in films. The design process adopted technique from Andy Kirk's established guide on visualization and was developed using D3.js, a javascript library for visualization on HTML canvas. The visualization gives access for people in the Malaysian film industry as well as Malaysian avid movie-goers to further analyze Malaysian movies and explore how progressive the movies have become over the years. This project provides a framework in design and implementation of a dynamic visualization that can be implemented in various domains.

Acknowledgements

The authors gratefully acknowledge Universiti Teknologi MARA (UiTM) for giving the authors an opportunity, support, and facilities to accomplish this study The visualization is publicly available on www.naelahnordin.com/visualizing-malay-cinema.

References

- [1] P. Godfrey, J. Gryz, and P. Lasek, "Interactive visualization of large data sets," *IEEE Trans. Knowl. Data Eng.*, vol. 28, no. 8, pp. 2142–2157, 2016.
- [2] J. Liutvinavičienė and O. Kurasova, "Multi-level Massive Data Visualization: Methodology and Use Cases," *Balt. J. Mod. Comput.*, vol. 6, no. 4, pp. 321–334, 2018.
- [3] H. T. Hisham, "The impact of globalization on the Malaysia film industry," *Int. J. Innov. Technol. Explor. Eng.*, 2019.
- [4] D. Polson, J. Kerr, R. Quagliata, and C. Wong, "Decoding the Marvel Cinematic Universe. An interactive data visualisation tool for exploring categories and connections between films, characters, artefacts and attributes," *Decod. Marvel Cinematic Universe. An Interact. data Vis. tool Explor. Categ. Connect. between Film. characters, artefacts Attrib.*, 2017.
- [5] J. Steinke and P. M. P. Tavares, "Cultural representations of gender and STEM: portrayals

-
- of female STEM characters in popular films 2002-2014,” *Int. J. Gender, Sci. Technol.*, vol. 9, no. 3, pp. 244–277, 2018.
- [6] M. Sap, M. C. Prasettio, A. Holtzman, H. Rashkin, and Y. Choi, “Connotation frames of power and agency in modern films,” in *Proceedings of the 2017 conference on empirical methods in natural language processing*, 2017, pp. 2329–2334.
- [7] J. Walsh, “Elite women warriors and dog soldiers: Gender adaptations in modern war films,” in *Gender and warfare in the twentieth century*, Manchester University Press, 2018.
- [8] S. Loist and E. Prommer, “Gendered production culture in the German film industry,” *Media Ind.*, vol. 6, no. 1, pp. 95–115, 2019.
- [9] F. K. H. Basri, “Representations of Gender in Malaysian Malay Cinema: Implications for Human Security,” *Asian Cine.*, vol. 19, no. 2, pp. 135–149, Dec. 2012, doi: 10.1386/ac.19.2.135_1.
- [10] Hanita Mohd Mokhtar-Ritchie, “Negotiating melodrama and the Malay woman: Female representation and the melodramatic mode in Malaysian-Malay films from the early 1990s-2009,” 2011.
- [11] F. Ibrahim, N. Yunus, M. Ilias, A. H. A. Mokhtar, and others, “Portrayal of women’s images in television dramas: A Malaysian case,” in *SHS Web of Conferences*, 2017, vol. 33, p. 63.
- [12] L. A. B. S. Ghazali and R. Islam, “Critical determinants of box office success for the Malaysian film industry,” *Int. J. Bus. Syst. Res.*, vol. 15, no. 4, pp. 491–509, 2021.
- [13] L. Yang, Z. Xu, and J. Luo, “Measuring Women Representation and Impact in Films over Time,” *arXiv*, vol. 1, no. 4. arXiv, pp. 1–14, Jan. 2020, doi: 10.1145/3411213.
- [14] A. M. Lindner, M. Lindquist, and J. Arnold, “Million Dollar Maybe? The Effect of Female Presence in Movies on Box Office Returns,” *Sociol. Inq.*, vol. 85, no. 3, pp. 407–428, Aug. 2015, doi: 10.1111/soin.12081.
- [15] D. Seyser and M. Zeiller, “Scrollytelling-an analysis of visual storytelling in online journalism,” *Inf. Vis. - Biomed. Vis. Vis. Built Rural Environ. Geom. Model. Imaging, IV 2018*, pp. 401–406, Dec. 2018, doi: 10.1109/IV.2018.00075.
- [16] A. Kirk, *Data visualisation: A handbook for data driven design*. Sage, 2016.
- [17] M. François, F. Osiurak, A. Fort, P. Crave, and J. Navarro, “Usability and acceptance of truck dashboards designed by drivers: Two participatory design approaches compared to a user-centered design,” *Int. J. Ind. Ergon.*, 2021, doi: 10.1016/j.ergon.2020.103073.
- [18] S. Zwinger, J. Langer, and M. Zeiller, “Acceptance and Usability of Interactive Infographics in Online Newspapers,” *2017 21st Int. Conf. Inf. Vis.*, pp. 176–181, Jul. 2017, doi: 10.1109/IV.2017.65.
- [19] J. Jawaheer, A. N. Nur Afiqah, O. Nasiroh, A. Sharifah, and M. Suraya, “Reminiscing the legendary P. Ramlee through data visualization technique/Jawaheer Jaffar...[et al.],” *Math. Sci. Informatics J.*, vol. 1, no. 2, pp. 59–69, 2020.
- [20] C. Smatt, R. Pratt, T. Abegaz, and A. Dobbs, “Towards Effective Customer Data Visualization: Data-Driven Documents (D3.JS) VS Google Charts.,” *QRBD*, p. 207, 2020.