

Analysis of Domestic Film Box Office

Yihan Mo

Information Science and Technology, School of Information.
University of Arizona
Tucson, AZ

Haoyu Yang

MIS, Eller College of Management.
University of Arizona
Tucson, AZ

Abstract— Understanding its market is the first step in understanding an industry. In order to conduct market improvement, analytic the film industry is necessary before any idea proposed. In this article, some data collected and analyzed to draw a brief picture of the distribution and box office of the domestic film market.

Keywords—Market Analysis; Film; Box Office; Tableau;

I. INTRODUCTION

Market analysis is an important part of business analysis and also one of the core applications of business intelligence and analytics system. As shown in Wil Van der Aalst's book, marketing is becoming more data-driven [1]. The historical data could help people better understand the market and easily find potential opportunities. Through sufficient market analysis and research, we are able to provide a reliable basis for business decisions.

At present, people conduct market analysis in a variety of ways, involving multiple disciplines, including but not only including economics, statistics, psychology, and sociology. In addition, for different fields and businesses, the appropriate market analysis methods are not the same. Only by collecting and appropriate data and selecting the appropriate methods for analysis can we obtain credible results and provide useful support for business decisions.

Our semester research project is Social Media Analysis for Marketing Improvement. As part of this semester research project, in order to meet the needs of later comparative analysis, we will present the process of collecting data of domestic film market data from 2017 to 2019, and use Tableau for processing and data visualization, this process will present in the second section of this article. Then in the third section, the outcome will be analyzed, and we will propose our interpretation of the results. From market analysis, we hope to build a primitive understanding of characteristics and operation rules of relevant market segments. This article will provide theoretical support for further research.

Films are an important part of the entertainment industry in the United States. The United States is the birthplace of the which industry and also represents the highest development level of it. According to Bakker's report, film industry has

become a very important industry that combined technological advance with innovative content [2]. The domestic film industry in the United States has unique characteristics, from the film distribution system to the box office splitting model, these unique characteristics show on the domestic box office data. The driving force behind the development of the film industry is the box office of the film, however, the development of the film industry is not always smooth. There are a fixed number of cinema screens, and film releases are competing for those screens in real time [3]. In recent years, with the development of multimedia technology, video websites such as YouTube and Netflix have impacted the film industry. Which mainly reflected in the decline in the annual total box office growth rate. Among them, the local box office not only stopped increasing from 2018 to 2019, but it experienced negative growth.

In this context, it is even more important to conduct a market analysis of films. According to Finney's book, the film industry has been characterized as a high-risk industry, and with good reason, that sunk costs in film production and marketing are significant [3]. Therefore, it is necessary to better understand the film market in order to avoid risks.

In the past, film marketing mainly relied on billboards and TV commercials. Tripartite evaluations of films usually come from film critics. The angle of film critics' evaluation of films is usually from a professional perspective, focusing on the artistic and emotional expression of films. For most people, such film reviews do not provide a good reference for them to choose a film. However, the advent of social media has provided a platform for ordinary viewers to express their opinions, and people have turned to social media to understand how others feel about films that they may be interested in. In this context, the sentiment in social media may become an important factor affecting the box office of a film, and sentiment analysis of social media may credible provide support for the marketing strategy of a film.

Before performing sentiment analysis, we need to establish a basic understanding of the film market. Therefore, it is necessary to analyze the recent situation of the film market. This article will start from two perspectives, one of which is a basic overview of the film market from 2017 to 2019, introduce the

major film companies and their market share. Another perspective is the box office changes based on time series, including the box office distribution of all films in a natural year, and the box office change after a single film is released.

II. PROCESS

A. Data Collection

Data sources determine the credibility and range of an analysis. In this research, we used 3 data sets from two sources, these 3 datasets are all in csv format.

We collect datasets below are collected from Kaggle:

Dataset1: Box office data set for all films from 2017 to 2019

We are going to use the domestic_revenue, world_revenue, distributor, opening_theaters, MPAA and genres column in this dataset. In which columns, MPAA represent the film ratings of each film.

Dataset2: Top 10 Highest grossing Films from 1975 to 2018

We are going to use the Main_Genre, rank_in_year, studio, worldwide_gross and year column in this dataset. In this dataset, the worldwide_gross actually is the domestic_revenue, it needs to be declared before we conduct the next step of the process. In which columns, ranking_in_year represents the worldwide_gross ranking of the film in the year it was released, so it actually is the ranking of domestic revenue.

We collect datasets below are collected from Box office Mojo by IMDbpro:

Dataset3: Domestic market box office data by day.

We are going to use the Date, Daily, to Date and Day column in this dataset.

B. Data Process

Before we start using Tableau to visualize data, we conduct cleanup and split operation for the datasets.

For the first step, we noticed that the dataset 1 contains duplicate records, to eliminate these duplications, we build a single sheet Excel workbook, paste the distinct title of films to it and use the VLOOKUP function to match the title with original records. After the treatment, we export the Excel sheet to a CSV file and replace the original dataset by the cleaned dataset.

For the second step, in the new dataset 1, we can observe that the format of the genre column is not well organized, for some films have more than one genre, it lists all of the genres in which column, so it induces some difficulty in the genre analysis. To make the genre analysis clear, we decide to use the first genre as a representation, similar to the Main_Genre in dataset 2. Same as the first step, we use Excel to conduct this treatment. We build a new column called Main_genres at the end and use the LEFT function to slice the genres column, find out the first genre, then store the result into the new column.

After we finished pre-process of data, we begin to visualize data, all the operations below based on Tableau Desktop software, version 2019.4.2.

C. Data Visualization Process

With the help of Tableau, we make 8 figures to present the data in 3 different datasets. We are going to briefly introduce the structure of these figures.

Fig.1, the Number of Films Distributed by Distributors from 2017 to 2019, this figure based on the Distributors column in dataset 1, we use the tree map to present the data. The area of each rectangle represents the number of films distributed by each distributor. This figure helps us build a direct view of who is the main player in the distributor market and how the market share separated by them.

Fig.2, the Genres of Films distributed by Distributors, this stacked bar plot based on the Distributor column and Main_genres column. It shows the genre distribute and the distributor's incline of different genres.

Fig.3, the Distributors with Top 10 Domestic Revenue, in the business world, market share may not mean revenue, this rule is obvious in the film industry. By comparing the information in this bar plot with the information in Fig.1, some business rules and characteristics in the film market may be revealed.

Fig.4 and Fig.5, the Genre Distribution by Number and by Revenue, these two bubbles plots are used to conduct comparative. By comparison, we can reveal the inherent box office difference determined by the genre of films.

Fig.6, the Office Box Change by Time Series, this bar plot base on the Day and Daily column form dataset 3. This figure reveals the changing tendency of the box office, we can observe significant cycle rules in this figure.

Fig.7 and Fig.8, these two pie plots are another group of figures used to compare. These two plots are both based on the Main_Genre column and year column in dataset2, the angle represents the count number of films and the data in these figures filtered by years. The difference between these two figures reveals the change of customer's favor.

III. RESULT

In order to better present the data, we use Tableau to conduct the visualize process, based on the data set we use, we conducted analysis in three sub areas.

The first part of the analysis is for film distributors. Film distributors are responsible for the promotion, marketing, and sales of films. To better understand film marketing strategies, we need to establish a basic understanding of the film distribution market. Our data presentation will help understand the market composition and competition of the film distribution market.

The second part analyzes the film types and box office distribution of the recent film market. In different social and cultural backgrounds, people have different preferences for genres in different periods. By analyzing the genre composition of the film market, we will likely find some opportunities. At the same time, by presenting the time series changes at the box

office, we will grasp the laws of people's viewing, which can also promote the efficiency of film marketing.

The third part analyzes the top 10 box office films every year from 1975 to 2018. Through the analysis, we hope to find some

characteristics of high box office films to support the improvement of marketing strategies.

A. Distributors

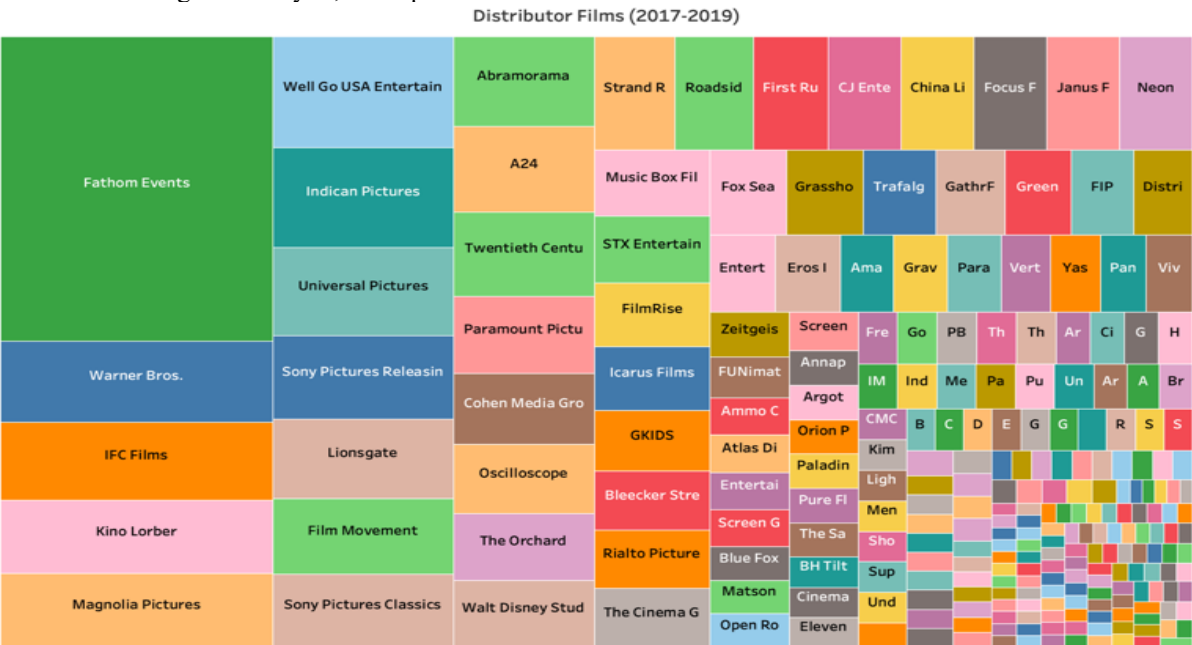


Fig.1 Number of Films Distributed by Distributors from 2017 to 2019

Form the tree maps of distributors, we can observe that, the long tail effect in the distributors is significant, the big scale distributors occupied most of the market share, the top 20 distributors distribute approximately 50% of the films. For some

small companies, they may only distribute a few films. According to the data present, we can make a conclusion that the competition of small companies is fierce.

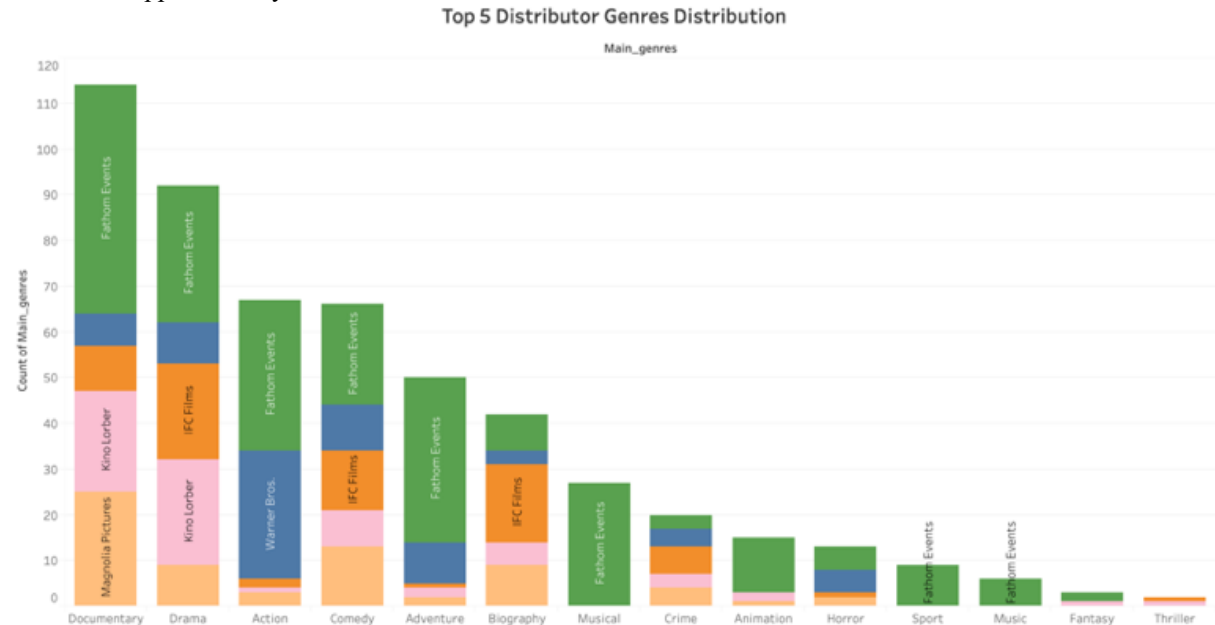


Fig.2 Genres of Films distributed by Distributors

However, film distributors are different from other areas of the film industry. The number of films released does not fully represent the company's volume and profit. We can see in Fig.1 that many distributors with a large number of released films are

not familiar to most people. Fig. 2 can always be partially explained this phenomenon. Film distributors like Kino Lorber and Fathom Event do not publish new, high-cost films. Their

main service targets are relatively niche films, including foreign language films, re-released classic films, etc.

On the other hand, we can see that, from 2017 to 2019, the circulation of documentaries has exceeded that of all other genres. This type of film is usually unable to obtain high box office. With regard to this phenomenon, we can say that unlike

other industries, the number of films in the film industry is not a key factor.

The best example of this is that Disney got 40% more in revenue by 32 films than Warner Bros, which released 75 films during the same period.

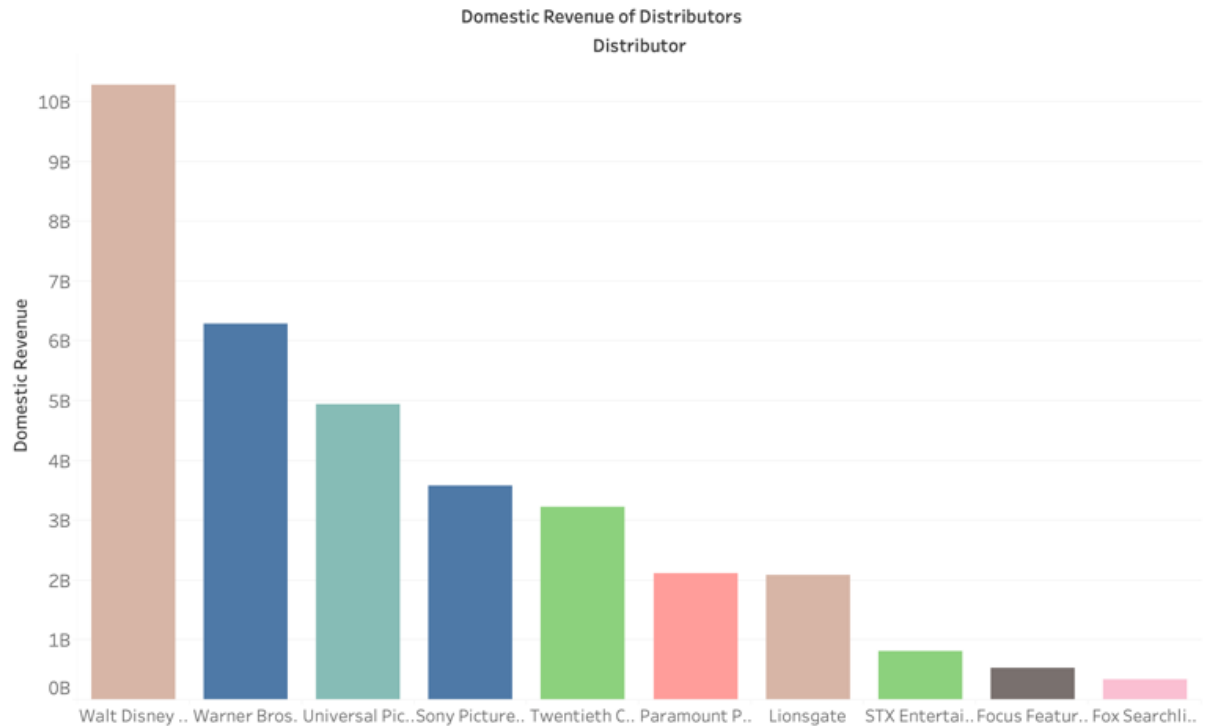


Fig.3 Distributors with Top 10 Domestic Revenue

B. Genre and Box office Distribution

For film revenue, there are also differences between genres, by comparing Fig. 4 and Fig. 5. It is not difficult to find the high-level average box office for action films, while the box office for documentaries and drama films are lower.

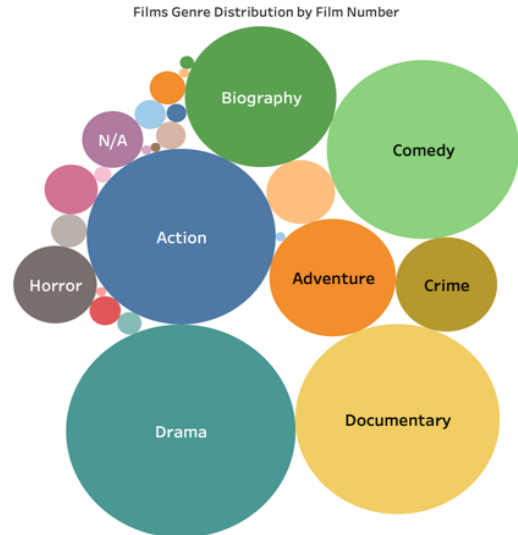


Fig.4 Genre Distribution by Number

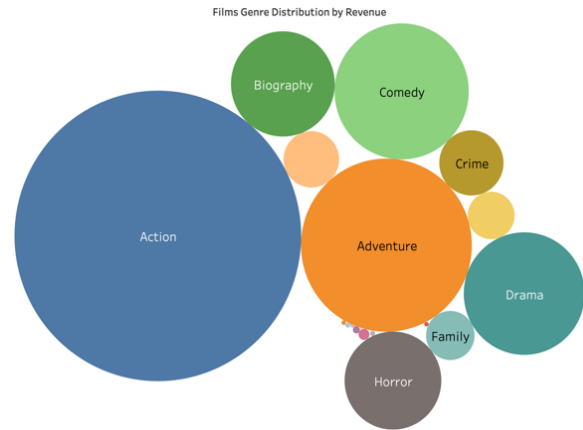


Fig.5 Genre Distribution by Revenue

Unlike the huge difference of other indicators of films from different genres, the daily box office variation tendency by the released time of the films in different genres shows a strong consistency. We all know that because of the nature of the billing system, American films are usually released on Friday. We selected the box office of five different genre high box office films from 2017 to 2019. By comparing the daily box office changes, we found that the box office changes took a

period of 7 days and fluctuated cyclically after the first or second day of the film. After the first weekend of release, Saturday was the highest peak for people to watch films. This

reveals the law of people's viewing, which can help to improve marketing efficiency.

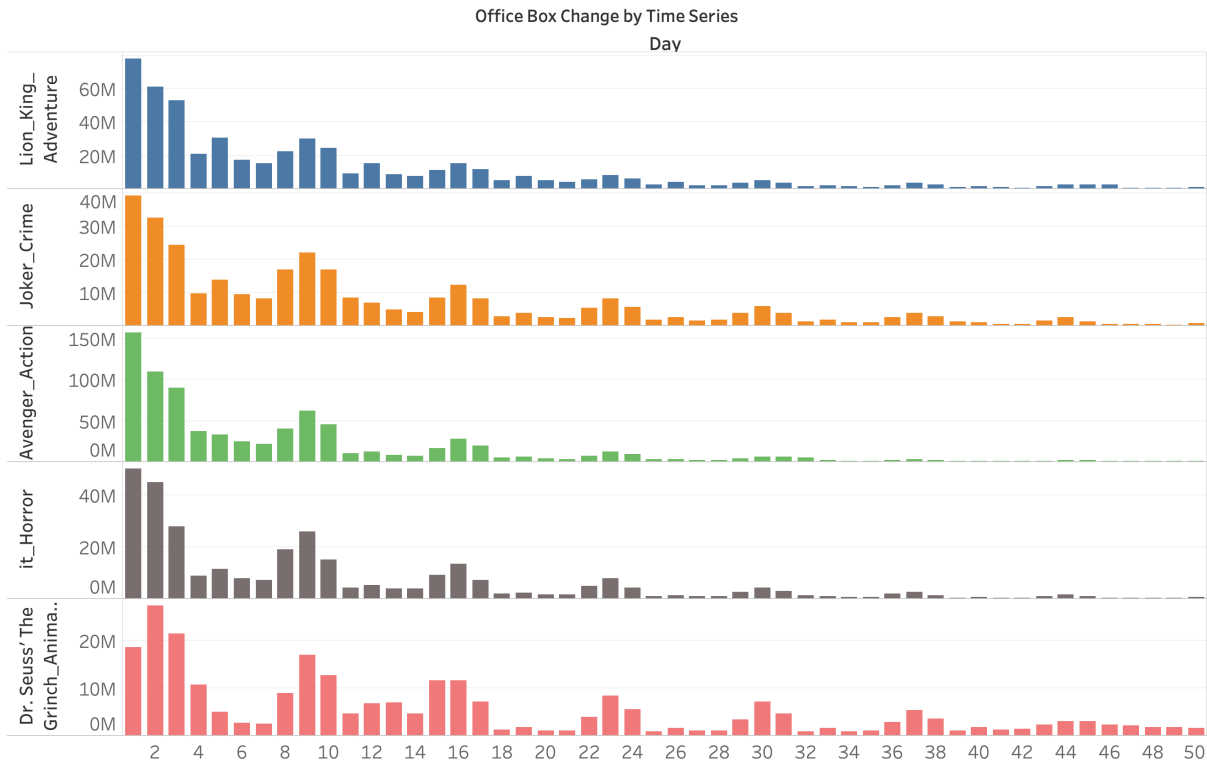


Fig.6 Comparison of 5 Typical Films from Different Genre

C. History Record of High Box Office Films

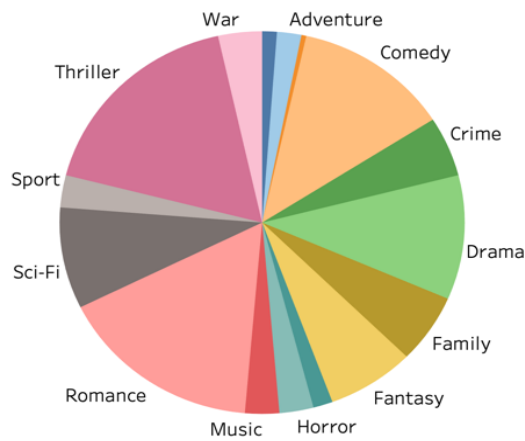


Fig.7 Top Selling Film Genre Distribution (1975-2000)

Analyzing historical data helps us predict the future. According to the statistics of the top 10 box office films from 1975 to 2018, we found that people's preferences for films have changed a lot. People's preferences for Romance fil, drama film, and thriller films are no longer, and action films, fantasy films, science fiction films and animation have become new choices.

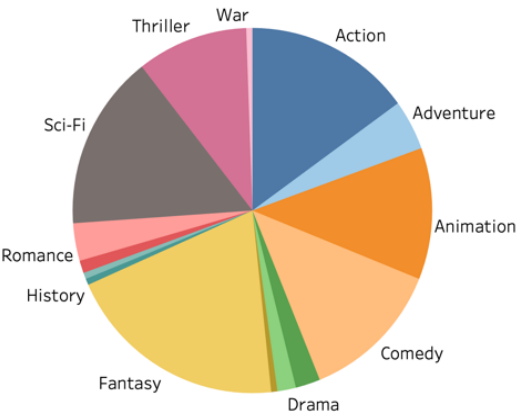


Fig.8 Top Selling Film Genre Distribution (2001-2018)

In recent years, the advancement of the film industry has greatly improved the visual effects of action films, science fiction films and anime films. We can see a phenomenon that people's purpose of watching films has changed from "listening to stories" to enjoying the pleasure brought by the audiovisual experience. This is a trend that deserves our attention.

REFERENCES

- [1] Van Der Aalst W. "Data Science in Action". Springer, Berlin, Heidelberg.
https://doi.org/10.1007/978-3-662-49851-4_1
- [2] Bakker. G, "The Decline and Fall of the European Film Industry: Sunk Costs, Market Size, and Market Structure, 1890–1927". The Economic History Review, 2005. <https://doi.org/10.1111/j.1468-0289.2005.00306.x>
- [3] Finney. A, "The International Film Business". London and New York: Routledge, 2015, p. 6, and p.8.
- [4] Peter R. K and Mary M. K, "Visual Cues: Practical Data Visualization World Wide Web" IEEE Computer Society Press, Los Almitos, CA, 1