

Final Stats 133 Project

Disney Movies Through the Years: A Text-Mining Analysis of Language, Sentiment, and Themes in The Lion King, Up, and Zootopia

Authored by:

Xianya Fu, Yechen Cao, Qianping Wu

UCLA Department of Statistics

March 2025

Table of Contents

Table of Contents	1
1 Introduction	2
2 Data Processing	2
3 Word Analysis	3
3.1 Word Frequency and Correlation	3
3.2 Word Importance and Uniqueness	4
3.3 Correlation of Movies Based on Word Usage	6
4 Sentimental Analysis	6
4.1 Proportion of Positive/Negative Sentiment	7
4.2 Barchart for Sentiments	7
4.3 WordCloud	8
4.4 LLM: compare with ChatGPT and DeepSeek	9
5 Latent Dirichlet Allocation: LDA	11
5.1 Beta Score	11
5.2 Gamma Score	12
5.3 Assignments	13
6 Conclusion and Limitation	14
6.1 Conclusion	14
6.2 Limitation	15
6.2.1 Insufficient Data Cleaning	15
6.2.2 Chapter Separation	15
6.2.3 Character Research	15

1. Introduction

We grew up with Disney animated films and always have special feelings for Disney stories full of magic and wonder. Disney movies enriched our childhood and shaped our understanding of kindness, courage, and love. As time passes, we can see noticeable changes in the characters, stage designs, and even songs of Disney movies. This evolution reflects Disney's commitment to staying relevant to contemporary audiences, adapting narratives to suit their tastes [1]. As Disney movie enthusiasts with text analysis skills, we want to know how Disney's content style has changed. Our research question is: how have the language usage, sentiment expression, and thematic focus in Disney animated movies evolved? We will compare the similarities and differences between different movies through word, sentiment, and topic analysis. We believe that the comparative analysis can provide valuable insights into Disney's evolving narrative techniques and their impact on different audiences. From this study, we aim to enhance our appreciation of the art and depth of Disney films.

2. Data Processing

We chose to conduct a text-mining study of three productions from different periods: The Lion King (1994) [2], Up (2009) [3], and Zootopia (2016) [4]. We found the scripts of these three movies online, including environmental descriptions and third-person narration, which are about 100 pages for each. We also performed a data-cleaning process on these texts to prepare them for further analysis.

Firstly, we used the classic eight cleaning steps. It includes simple transforms, conversion to lowercase, removing numbers, removing punctuation, removing English stop words, removing other unique stop words, stripping whitespace, and stemming. Since movie scripts contain many character names and our subsequent analysis is not related to specific characters in the films, we have categorized these character names, along with some additional script-specific words that are meaningless to the plot development, as special stop words and remove them. We also noticed that movie scripts often use EXT. (Exterior) and INT. (Interior) to indicate the context for visualization. Each occurrence of EXT. or INT. signals the

beginning of a new scene, which effectively marks the transition to a new film segment. Therefore, we used these markers to divide the three movies into chapters. As a result, The Lion King was divided into 53 chapters, Up into 182 chapters, and Zootopia into 80 chapters.

3. Word Analysis

Our first analysis focuses on word analysis. This section examines how the vocabulary used in the three movies is related. More specifically, we analyze the correlation and differences among the three films by looking at word frequency, importance, and uniqueness.

3.1. Word Frequency and Correlation

We first present the top 15 most frequently occurring words for each film to gain a distinguished understanding of the three movies. Notably, the word house appears over 200 times in Up, highlighting the movie's strong connection to themes of home and family. In The Lion King, the word king naturally emerges as the most frequently occurring term. Apart from these words strongly tied to the storyline, the remaining words are simple, commonly used words in everyday life, possibly due to Disney's target audience primarily being teenagers. Using widely used words daily helps capture and maintain young audiences' interest. Beyond that, The Lion King contains a relatively higher number of emotion-describing words, which will probably be reflected more in subsequent sentiment analysis.

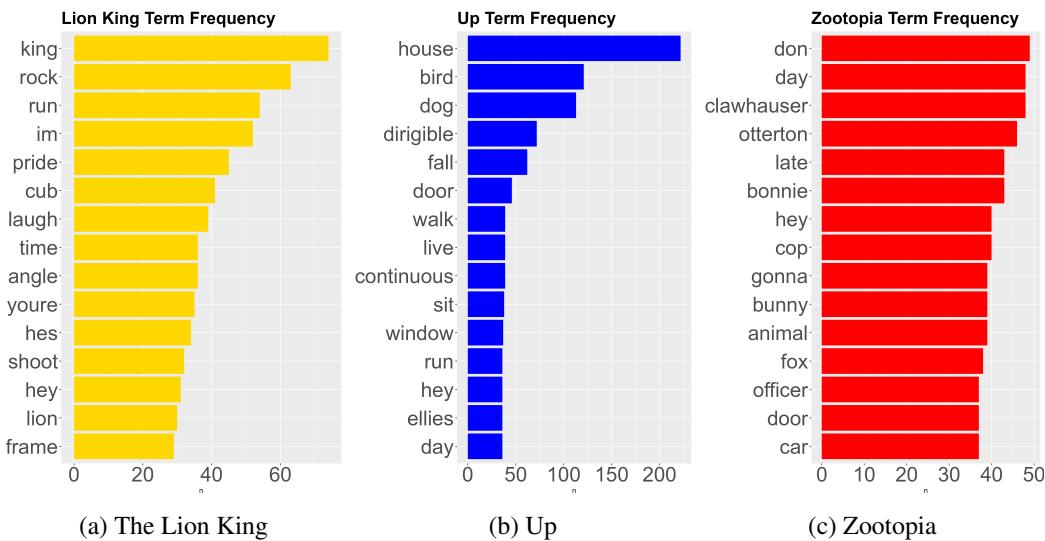


Figure 1: Comparison of Word Frequency Plots for Three Movies

The correlation plot further illustrates the relationships between these frequently used words. We set the lowest term frequency to 100 and the correlation threshold to 0.9. As shown, “house,” “dog,” “back,” and “bird” are closely linked, possibly reflecting Up’s storyline, while other words form another strong cluster. The high correlation among these frequently used, everyday words further demonstrates that all three Disney movies use daily life as a context to convey their stories.

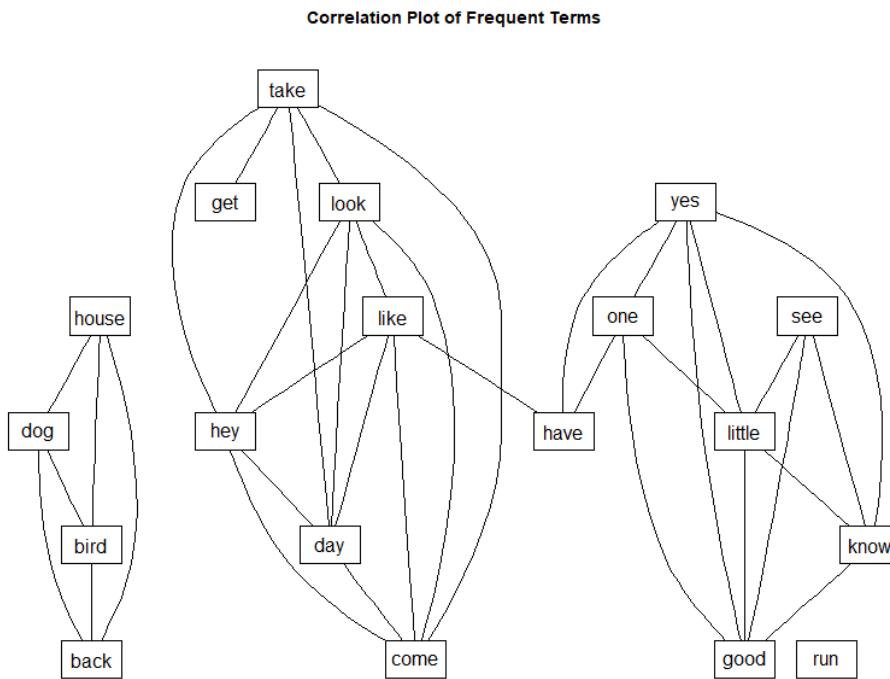


Figure 2: Word Correlation Plot with term frequency = 100 and correlation = 0.9

3.2. Word Importance and Uniqueness

We conducted a TF-IDF calculation to analyze the word usage in these movies better. In table 1, we list the top six words with the highest TF-IDF values. However, we can see that these words have higher TF-IDF values because they appear more frequently in a specific movie script. The IDF value 1.099 indicates that these words are uncommon in the other movie scripts. This observation suggests that the primary words used in each movie are distinct.

Movie	word	n	ft	idf	tf-idf
Up	house	222	0.034	1.099	0.038
Lion King	king	74	0.013	1.099	0.014
Zootopia	enter	84	0.012	1.099	0.013
Up	dirigible	72	0.011	1.099	0.012
Lion King	pride	45	0.008	1.099	0.009
Lion King	cub	41	0.007	1.099	0.008

Table 1: TF-IDF values for frequently used words in three movies.

We also created comparative plots of word usage proportions for each pair of the three movies. Amid the abundance of everyday words, the three films still show subtle differences. The Lion King features more words such as "run", "laugh", and "father", which align with its themes of leadership, family, and destiny. And Up has more adventurous words like "jump" and "climb".

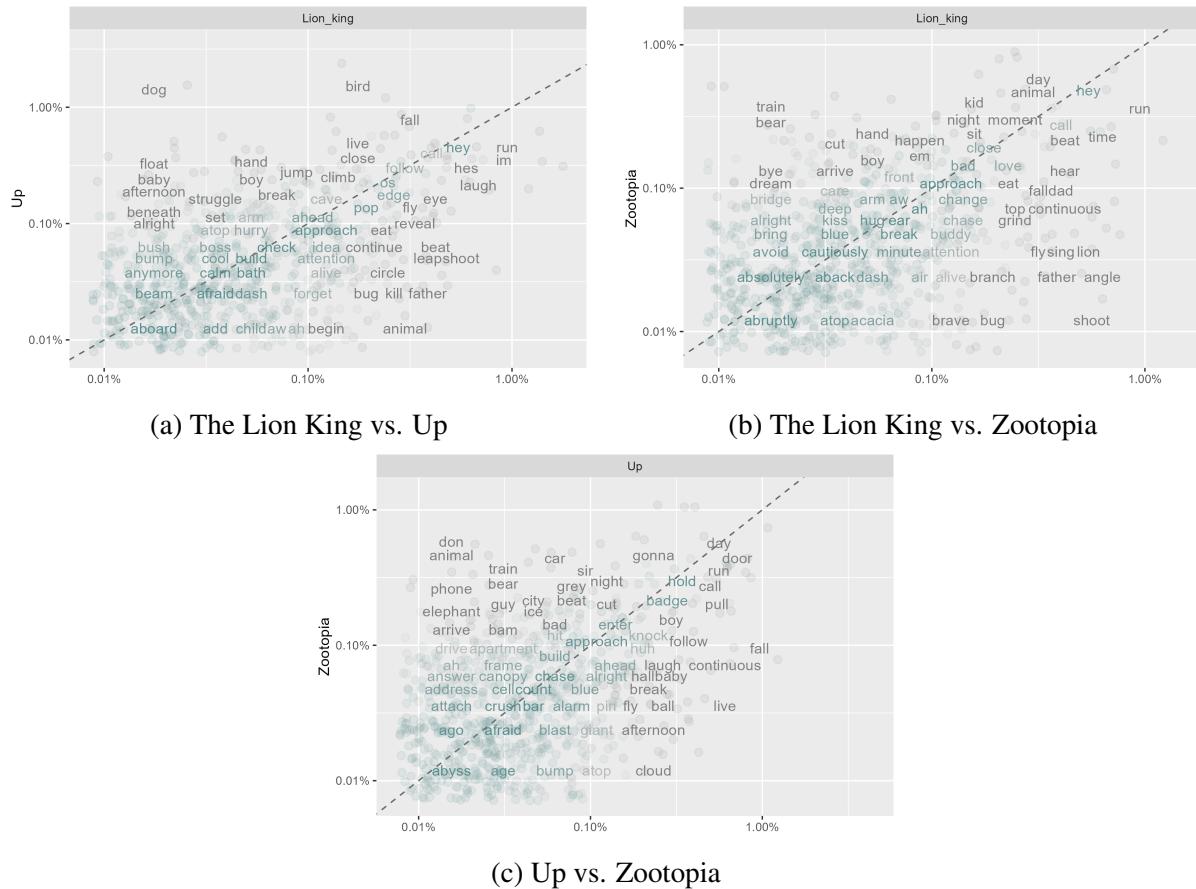


Figure 3: Comparison of different movies' unique words

3.3. Correlation of Movies Based on Word Usage

Finally, we calculated the correlation coefficients for each pairwise comparison of the movies. Based on the coefficients in table 2, we observe that the overall correlation between the three movies is relatively low. This suggests that their vocabularies and linguistic styles differ significantly. Among the comparisons, The Lion King and Up exhibit the lowest correlation at 0.3668 (95% CI: 0.3007 - 0.4294). This indicates a substantial difference in their word usage, likely due to their contrasting themes. The correlation between The Lion King and Zootopia is 0.4897 (95% CI: 0.4330 - 0.5426), suggesting a slightly stronger relationship but still a notable distinction in vocabulary. Lastly, Up and Zootopia have a correlation of 0.4530 (95% CI: 0.3943 - 0.5080), which is higher than The Lion King vs. Up but still indicates considerable linguistic differences. Overall, the low correlation values emphasize the distinct storytelling approaches of each film, with unique linguistic choices that reflect their respective narratives and themes.

Comparison	p-value	95% CI	Estimated Correlation
Lion King vs. Up	< 2.2e-16	(0.3007, 0.4294)	0.3668
Lion King vs. Zootopia	< 2.2e-16	(0.4330, 0.5426)	0.4897
Up vs. Zootopia	< 2.2e-16	(0.3943, 0.5080)	0.4530

Table 2: Correlation between three Disney movies

4. Sentimental Analysis

This section will do the sentiment analysis for three films. We assess the balance between positive and negative sentiment, comparing their emotions, and identifying key themes that contribute to their emotional depth through sentiment analysis. Moreover, we compare sentiment analysis results generated by Large Language Models (LLMs), specifically ChatGPT and DeepSeek, to evaluate their consistency and methodological differences.

4.1. Proportion of Positive/Negative Sentiment

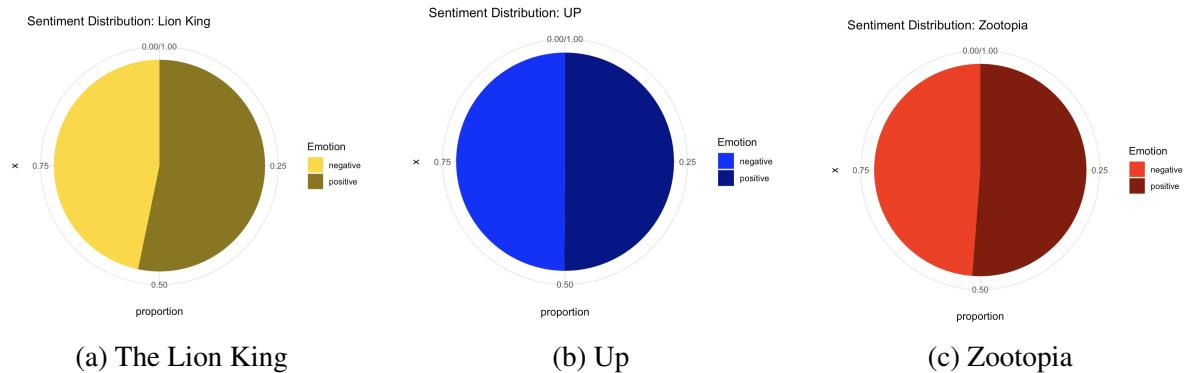


Figure 4: Proportion of Positive/Negative Sentiment for Three Films

All three movies have a nearly equal balance of positive and negative sentiment. The negative sentiment likely comes from emotional moments or heavy themes in each movie. The positive sentiment reflects uplifting messages, engaging stories, and satisfying conclusions. The Lion King has the highest positive proportion of these films, and UP has the most equal balance between positive and negative sentiment.

4.2. Barchart for Sentiments

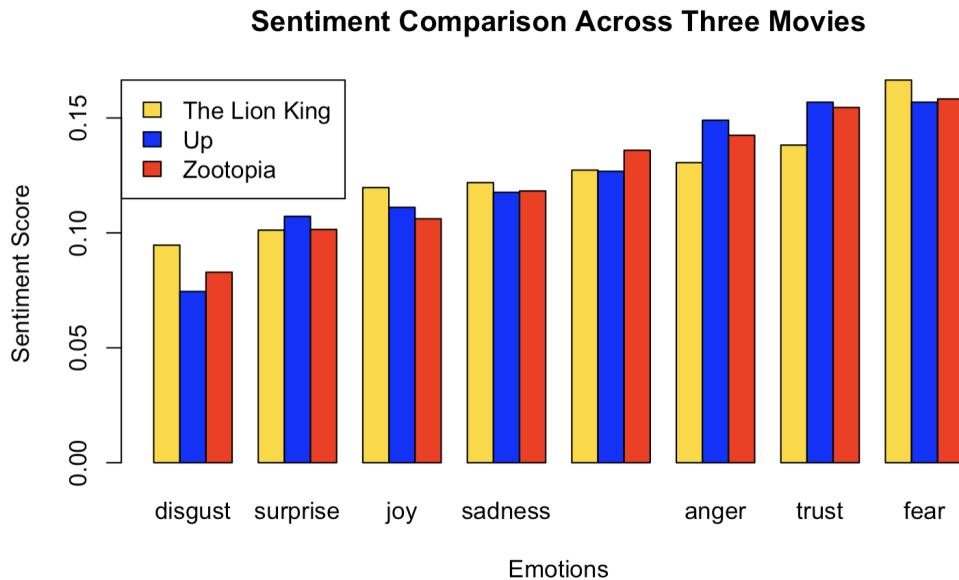


Figure 5: Sentiment Comparison Across Three Movies

From above, the Lion King has the highest scores of fear, sadness, joy, and disgust, and Up has the highest scores of trust, anger, and surprise. The Lion King has a mix of big emotions. There is a lot of fear, trust, and sadness, but the joyful moments help balance it out. Up is the most emotional one, with a lot of trust and surprise, and some anger that make the story go deep. Zootopia feels more real-world because it deals with anger, trust, and fear, especially around social issues like bias and discrimination.

4.3. WordCloud



Figure 6: WordCloud for Three Films

Each WordCloud effectively captures the essence of the film, highlighting the themes, and settings of the story. The dominant words in The Lion King include king, pride, rock, cub, and laugh, reflecting themes of royalty, family, and the savanna setting. Significant words In Up like house, bird, dog, and dirigible highlight key elements of the movie, such as Carl's house. The words emphasize adventure, companionship, and exploration. The major words in Zootopia, including bellwether, cop, officer, and train, suggest a focus on law enforcement, mystery, and city life, and words like predator and savage, hint at the film's themes of prejudice and social dynamics.



Figure 7: Positive vs Negative WordCloud for Three Films

The Lion King has a heavy emotional loss (death, danger) but is balanced with themes of courage (pride, bravery). Up has Emotional challenges (struggle, screaming) but is filled with hope and companionship. Zootopia has strong conflicts (crime and fear), as well as justice and trust.

4.4. LLM: compare with ChatGPT and DeepSeek

When we ask ChatGPT to do the sentiment analysis on these three movies, it only shows this table with three column names: movie, polarity score, and subjectivity. But the result was different compared to the result I got, so I asked, “Are you sure?” twice. Then, it gave me three different results, including the first answer.

Movie Script Sentiment Analysis			
	Movie	Polarity	Subjectivity
1	The Lion King	0.7043195452008929	1.0
2	Up	0.23815855687978316	1.0
3	Zootopia	0.13252766927083334	1.0

Verified Sentiment Analysis			
	Movie	Verified Polarity	Verified Subjectivity
1	The Lion King	0.351953125	0.6
2	Up	0.2534835233126249	1.0
3	Zootopia	0.14328369140625	1.0

Verified Sentiment Analysis			
	Movie	Verified Polarity	Verified Subjectivity
1	The Lion King	0.351953125	0.6
2	Up	0.2534835233126249	1.0
3	Zootopia	0.14328369140625	1.0

Figure 8: Three Answers from ChatGPT

The same thing happened on DeepSeek, it also gave three different answers when I questioned the results.

Polarity Score:

If we were to assign a polarity score (on a scale of -1 to +1, where -1 is fully negative, 0 is neutral, and +1 is fully positive), *The Lion King* would likely score around **+0.6 to +0.7**. This reflects its overall positive tone, balanced by the presence of significant negative moments.

Figure 9: First Answer from DeepSeek

Revised Polarity Score:

On a scale of -1 to +1, *The Lion King* would likely score around **+0.3 to +0.4**. This reflects its **slightly positive** polarity, acknowledging the significant negative moments but recognizing the uplifting resolution and themes.

Figure 10: Second Answer from DeepSeek

Revised Polarity Score:

On a scale of -1 to +1, *The Lion King* would likely score around **-0.2 to -0.1**. This reflects its **slightly negative** polarity, acknowledging the significant negative moments and their emotional weight, while recognizing the positive resolution and themes.

Figure 11: Third Answer from DeepSeek

However, it has a higher requirement for the input file. I gave the same PDFs, and the chat read them all, but DeepSeek can only read the Lion King script, and it shows that if it didn't read 100 percent of all files, it would not do the sentiment analysis.

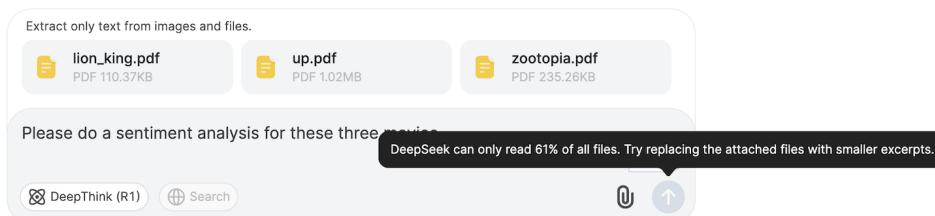


Figure 12: Put All Three Files Into DeepSeek

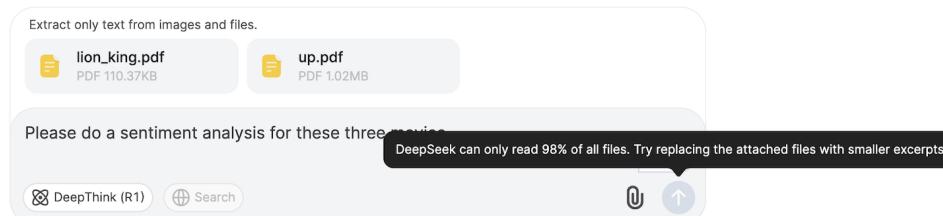


Figure 13: Only "The Lion King" and "Up" Into DeepSeek

5. Latent Dirichlet Allocation: LDA

In this section, we will perform Latent Dirichlet Allocation (LDA) to discover hidden topics in the dataset. LDA helps us to find the mix of words associated with each topic while determining the mix of topics each paper describes^[5]. This process not only gives us an idea of the central themes in the dataset, but also provides insight into the distribution of these themes across different documents.

5.1. Beta Score

We will begin by examining the beta scores since they provide a quantitative measure of the importance of each term in defining the various topics.

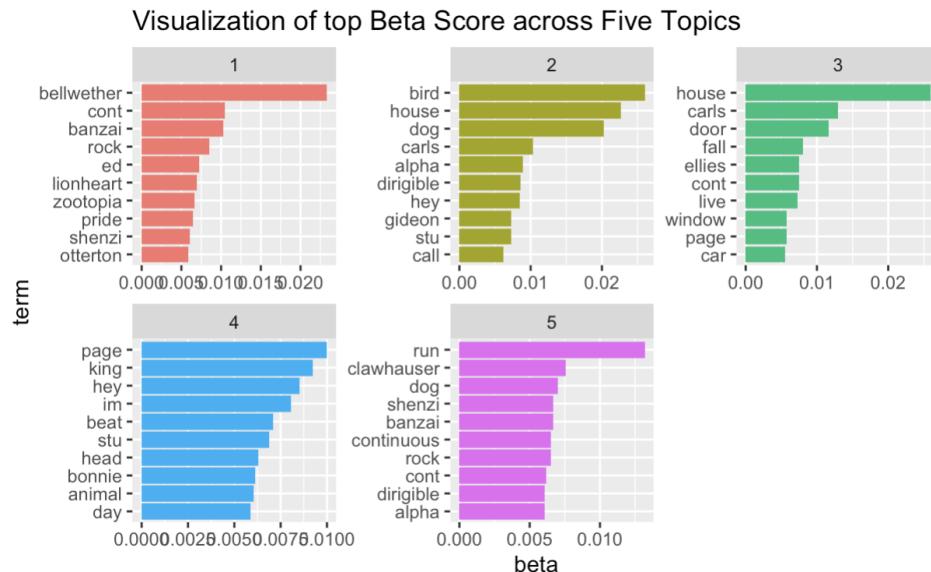


Figure 14: Visualization of top Beta Score

This diagram shows the key terms for each topic identified by Latent Dirichlet Allocation (LDA). Topics are shown in different colors, highlighting the key terms that define each theme. For example, Topic 1 (highlighted in red) contains terms such as “rock,” “lionheart,” and “pride” which can be associated with leadership. Topic 2 (green) contains terms such as “bird,” “house” and “dog,” which reflect topics in which animals or the environment may be central. Topic 3 (teal) also emphasizes terms such as “house,” “carls” and “door,” which may indicate a focus on place or living space. Topic 4 (blue) and 5 (pink)

included words such as “page,” “king” and “run,” which further indicated a focus on characters, actions, or places that were related to thematic elements. The prominence of these terms in each topic reflects underlying patterns in the data and provides insights for further exploration of topics and associations.

5.2. Gamma Score

Next, we move on to the analysis of gamma scores. Gamma scores reflect the degree of association between a specific document and a particular topic, and provide insight into how well the document fits the topic.

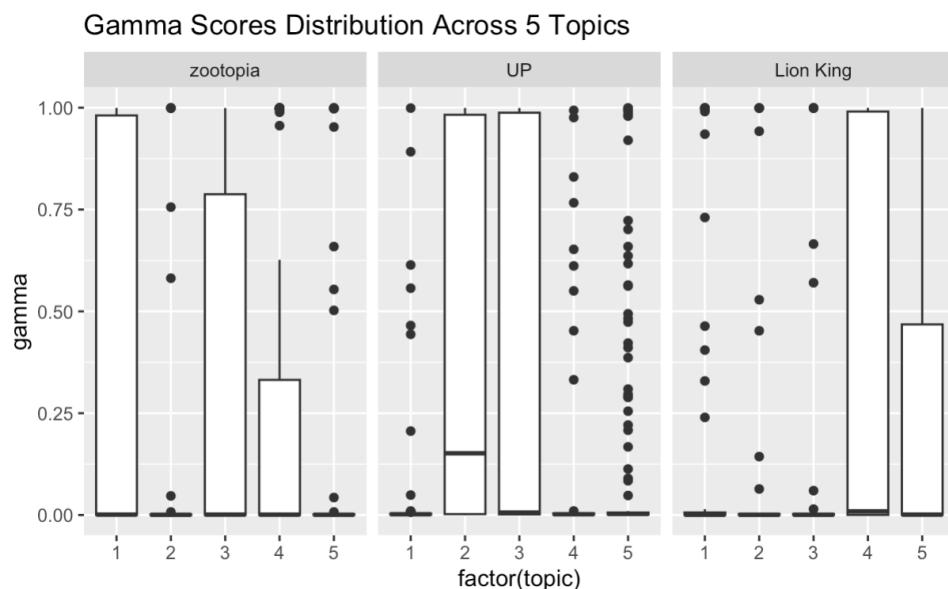


Figure 15: Gamma Score Distribution

The graph “Distribution of gamma scores across 5 subjects” shows the distribution of gamma scores for three movies. For Zootopia, gamma scores are most concentrated around Topic 1 and 3, with a significant portion of the data clustered near the upper range of Topic 1. UP shows a similar concentration of high gamma scores for Topic 2 and 3, with a wider spread of values in the other subjects. The Lion King shows a strong concentration of gamma scores for Topic 4, with less variability in the other subjects. The boxplots show the central tendency and spread of gamma scores, and show how well each movie fits each topic.

This table provides additional information to complement the previous Gamma Score distribution graph. It shows the percentage of chapters related to each topic in three movies.

Title	Topic	n	total_chapters	Percentage
Lion King	4	20	52	0.3846154
Up	2	77	181	0.4254144
Zootopia	1	22	77	0.2857143
Zootopia	3	22	77	0.2857143

Table 3: Distribution of topics in different films

The table shows:

- The Lion King has the highest correlation with Topic 4, accounting for about 38.5% of the movie's chapters.
- Up has a higher association with Topic 2, with 42.5% of the chapters related to this topic.
- Zootopia, on the other hand, is divided into Topic 1 and Topic 3, with 28.6% and 28.6% of the chapters related to these two themes, respectively.

The table provides a clearer picture of the thematic focus of each movie, highlights the distribution of different topics across chapters, and gives us a better understanding of the relevance of topics in the narrative structure of each movie. Overall, the movies seem to be more clearly associated with particular topics, indicating a strong thematic coherence within these specific categories.

5.3. Assignments

The “Assignment Distribution Matrix” visualizes the percentage of words from Zootopia, Up and Lion King assigned to each of these movies. The color intensity indicates the ratio of assignments, with darker red indicating a higher percentage. UP is the most dominant in terms of the percentage of words assigned back to its own movie. Zootopia’s words are more evenly distributed among the three movies, with a slight preference for itself. Lion King is closely associated with its own, but also overlaps with UP and Zootopia, showing some flexibility in the thematic distribution.

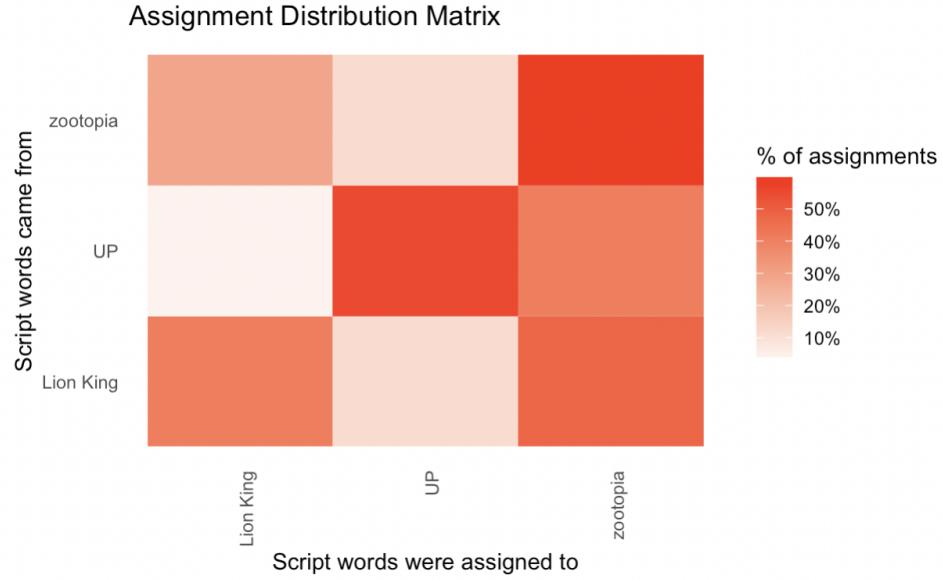


Figure 16: Assignments Distribution Matrix

6. Conclusion and Limitation

6.1. Conclusion

Word analysis shows that while all three movies are well-known animated films, their commonly used words differ significantly. Key terms such as house in Up, king in The Lion King, and enter in Zootopia highlight the distinct narrative elements of each film.

Additionally, the TF-IDF analysis confirms that certain words are more common in one movie while being uncommon in others. Correlation analysis further supports this finding and shows relatively low similarity in vocabulary between each pair of movies. All in all, despite being Disney productions, the word choices in each film are according to their specific plots, settings, and characters.

The sentiment analysis shows Disney's way of showing emotions has changed over time—starting with big, dramatic moments in The Lion King (1994), shifting to deep, personal sadness and nostalgia in Up (2009), and then mixing humor with serious social issues in Zootopia(2016). From the Large Language Models (LLM), we can conclude that sentiment scores are not absolute, and different Large Language Models apply unique methods and weightings, which can lead to different conclusions.

By analyzing the LDA, we found different thematic patterns. UP has the most

concentrated thematic content, with a clear dominance of Topic 2, suggesting that the text is strongly aligned with the assigned theme. Zootopia, on the other hand, has a more even distribution of thematic content, with text spread across three topics, indicating a broader exploration of topics throughout the story. Meanwhile, The Lion King is primarily aligned with its central theme (topic 4), but also shows some thematic overlap with UP and Zootopia, suggesting that the movie has flexible thematic connections that contribute to its complexity. These insights highlight how the various Disney films use their themes to create unique story structures, while at the same time sharing certain common themes.

6.2. Limitation

6.2.1. *Insufficient Data Cleaning*

Some words, such as “cont,” were not properly cleaned up during the preprocessing phase. This oversight may lead to inconsistencies or inaccuracies in the analysis, affecting the quality of the results.

6.2.2. *Chapter Separation*

Chapters were divided based on the Ext and Int factors, which may not be the most accurate method for segmentation. This approach could lead to inconsistencies in how chapters are grouped, potentially affecting the overall analysis of the content.

6.2.3. *Character Research*

The analysis lacks a detailed study of the characters in the movie. Character-level research could provide valuable insights into how specific characters relate to different themes and stylistic figures. Such an in-depth study would improve the understanding of narrative and character dynamics within the context of thematic model analysis. This is one of the directions in which we can dive into the future.

References

- [1] Protiti Mondal and Sraya Banerjee. From classic disney animations to contemporary live action remakes: Change in narrative in three disney fairy tale movies. *Research Journal of Berhampur University*, page 151, 2023.
- [2] Irene Mecchi. Scripts the lion king. Scripts.
<https://www.scripts.com/script-pdf-body.php?id=167>.
- [3] Bob Peterson and Pete Docter. Up. Indieground Films. <https://indiegroundfilms.wordpress.com/wp-content/uploads/2014/01/up-signed.pdf>.
- [4] Jared Bush et al. Original screenplay by zootopia. Cinefile.
<http://www.cinefile.biz/script/zootropolis.pdf>.
- [5] Julia Silge. *6 Topic Modeling | Text Mining with R*. Tidytextmining.com, 2017. Accessed 14 Mar. 2025. <https://tidytextmining.com/topicmodeling>.