

My Little Pony Friendship Is Magic: Side Character Talk Comparison

Cathy Chen, Tanav Bansal, Yifan Lin

McGill University

Abstract

My Little Pony: Friendship is Magic is a popular animated series that takes place in the magical land of Equestria. While the six main ponies drive most episodes, side characters also play important roles in advancing plots. This study analyzes the dialogue patterns of three important side characters—Princess Celestia, Discord, and Sweetie Belle—across Seasons 1-8 to understand how their speech reflects their narrative roles and life stages. Using proportionate stratified sampling, we collected 300+ dialogue lines for each character and conducted manual topic annotation across six categories: Friendship & Relationships, Identity & Personal Growth, Authority & Leadership, History, Lore, & Magic, Nature & Animals, and Food & Drink. Our findings reveal that Princess Celestia focuses primarily on Authority & Leadership (39.9%) and world-building knowledge, Discord emphasizes Identity & Personal Growth (34.74%) reflecting his redemption arc, and Sweetie Belle centers on Friendship & Relationships (61.1%) consistent with childhood development. These distinct dialogue patterns demonstrate how side characters contribute unique narrative perspectives shaped by their social roles and life stages, offering meaningful educational value that the production company should consider when developing new storylines or merchandise.

Introduction

My Little Pony: Friendship is Magic is a popular animated series that takes place in the magical land of Equestria. The story focuses on the theme of friendship and follows Twilight Sparkle, a unicorn who moves to the small town of Ponyville. There, she meets important friends, goes on adventures, and together they defeat different enemies—helping her learn the magic of friendship.

While the six main ponies drive most episodes, the side characters also play important roles in advancing plots. For this reason, understanding how side characters talk can provide valuable insights for the production company aiming to maintain consistent characterization when developing new storylines or producing merchandise. To address this need, we analyzed the full transcript of Seasons 1-8 and analyzed the dialogue of three important side characters—Princess Celestia, Discord, and Sweetie Belle. For each character, we

calculated the proportion of their total dialogue that fell into each of six topic categories: (1) Friendship & Relationships, (2) Identity & Personal Growth, (3) Authority & Leadership, (4) History, Lore, & Magic, (5) Nature & Animals, and (6) Food & Drink. This allowed us to quantify how frequently each character discusses each topic and compare their engagement across themes. We also computed tf-idf scores for the top ten terms in each topic and used ChatGPT to generate topic summaries, further helping to interpret the major topics each character cares about. By combining lexical patterns with contextual knowledge of the show, we were able to justify why specific topics appear more frequently in the dialogue of some characters than others.

Through the rigorous data science process, we found that each side character contributes a distinct narrative perspective shaped by their social roles and life stages. Specifically, Princess Celestia pays most attention to her leadership role and national affairs, which aligns with her highest duty in Equestria, as she needs to concern a lot about the kingdom. Discord pays most attention to his identity, magic, as well as friendship development, which aligns with his redemptive self transformed from an antagonist to a protagonist. Sweetie Belle centers her dialogue on friendship building and personal growth, consistent with her stage of developing self-concepts and learning through peer support at a young age. These findings show that side characters are not merely accessories to the plot—they actively learn, change, and contribute to the show's themes alongside the main characters. Their distinct narrative perspectives offer meaningful educational value for viewers that the production company should carefully consider.

Data

The original dataset comes from the My Little Pony Transcript collection on Kaggle (<https://www.kaggle.com/datasets/liury123/my-little-pony-transcript>). It contains the full transcripts of all episodes from Seasons 1–8 of My Little Pony: Friendship Is Magic, along with My Little Pony: The Movie and the holiday special My Little Pony: Best Gift Ever. Each line in the dataset is organized by episode title, writer, speaker, and the spoken content, which makes it easy to extract dialogue for specific characters. Since the movie and the holiday special take place outside the main storyline and may introduce topics that differ from the core narrative

of the show, we filtered these out and kept only the 195 main episodes from Seasons 1-8. This gave us a more consistent dataset, grounded in the main plot of the show.

Some episode titles are split into Part 1 and Part 2, usually because they cover a major adventure that needs two episodes to tell. In these cases, the first few lines of Part 2 were clips repeated from Part 1. We removed all repeated lines at the beginning of Part 2. The recap section in Part 2 almost always appears within the first 10–20 lines. We found the last repeated line and removed anything before that line.

We defined Trivial lines as those with fewer than 25 characters. Since 25 characters typically correspond to more than five words, any line exceeding this threshold is likely to contain enough meaningful content to be analyzed.

With the cleaned dataset, we used a proportionate stratified sampling technique to collect dialogue for each selected character from the cleaned dataset. Since each episode centers on a different theme and the amount of dialogue each character has varies widely across episodes, we first calculated how many lines the character speaks in each episode and computed the percentage of their total dialogue that comes from each episode. Based on these percentages, we allocated the sample size (300-310 lines) proportionally across episodes, and then randomly sampled that number of lines from each episode. This reduced the sampling biases toward specific topics that mainly appear in certain episodes while ignoring the others.

At the end of data collection, we created a CSV file for each selected character, containing more than 300 lines of dialogue randomly selected from the cleaned full dataset. We obtained 304 lines for Princess Celestia across 41 episodes, 308 lines for Discord across 17 episodes, and 319 lines for Sweetie Belle across 39 episodes. In addition to the default headers—character, title, and dialogue—each file included additional headers, with the following information recorded under each: *line_id* (i.e., the line’s index in the full dataset), the three coders’ names (each indicating that coder’s annotation), *topic* (i.e., the final decision based on majority agreement), and *notes* (to record justification).

Methods

Dataset Selection

We used a dataset created in 2015 that contains dialogue from Seasons 1–8 of the show. It is important to note that a final season (Season 9) was released in 2018. Due to time constraints that prevented us from manually collecting dialogue from Season 9, we decided to focus on the available dataset for Seasons 1–8. Despite this limitation, the dataset remains highly informative for our analysis.

Additionally, Season 9 shifts the narrative largely to the newer generation (i.e., the Young Six), and various side characters begin to take on the role of main characters in certain episodes. This change makes it more challenging to define which characters should be considered main or side. This further persuaded us to restrict our analysis to Seasons 1–8, as the changing story focus of Season 9 may contribute limited value to the dialogue topic analysis that we aim to conduct.

Character Selection

In *My Little Pony: Friendship is Magic*, the officially recognized main characters are Twilight Sparkle, Applejack, Fluttershy, Rarity, Pinkie Pie, Rainbow Dash, and Spike, with Starlight Glimmer later joining the main cast starting in Season 6.

For our project, we purposely chose characters who are not part of the main cast but still play essential roles in the storyline, so that our analysis better reflects how side characters in the show talk. Based on this, we selected: 1) Princess Celestia, the ruler of Equestria and Twilight’s teacher, who often guides the main characters in important moments; 2) Discord, a draconequus who first appears as an antagonist but then is influenced by the main characters and become one of their most supportive friends; 3) Sweetie Belle, Rarity’s little sister and a founding member of the Cutie Mark Crusaders (i.e., a group of foals whose goal is to discover their special talents through various adventures), who learns the magic of friendship alongside the elder ponies.

The three characters represent different groups in the show: royalty, a non-pony creature, and a foal. Looking into their dialogue gave us a balanced view of how various types of side characters speak.

Data Annotation

We randomly selected the dialogue lines using proportionate stratified sampling, but an isolated line often does not provide enough information to determine what the character is talking about. Thus, to support accurate annotation, we created a separate CSV file for each character that includes the previous and subsequent lines of their dialogue for every line. This allowed us to look up the contextual information for ambiguous lines to ensure the reliability of our annotation.

For each selected character, we randomly sampled 100 lines into a separate CSV file. Each group member independently performed open coding on one character’s dataset to identify meaningful themes in the dialogue. After coding, we compared our findings across characters to find the common topics.

Following discussion and agreement, we developed a typology comprising six topics. A codebook was created, containing detailed information for each defined topic, including explicit boundaries between topics (i.e., how the topics differ from one another), positive and negative examples for each topic, as well as some edge cases.

Then, we three coders annotated the 300-line datasets for all three characters. During the process, we iteratively refined the definitions of topics to eliminate ambiguity and overlap. After the first round of annotation, most dialogue lines fell into at least one of the six topics with majority agreement (i.e., two or more coders selecting the same topic(s)). We then jointly reviewed lines that did not reach majority agreement and collaboratively made the final decisions. By the end of this process, all dialogue fell within at least one topic, with some lines conveying richer information fitting into multiple topics.

Data Analysis

Based on the annotated dataset, we calculated the percentage distribution of how often each character mentioned each topic. We then computed the term frequency-inverse document frequency (tf-idf) for each topic to identify the ten most commonly appeared terms within each category. We also used ChatGPT to generate a representative summary for each topic category, highlighting the specific issues that characters most frequently focus on within those broader themes. Additionally, we conducted an exploratory analysis to examine which characters mentioned the tf-idf terms to better interpret the motivations behind their speech pattern.

Results

Topic Selection

Through open coding, we developed a typology of six distinct topics present in the dialogue of the three selected characters:

The **Friendship & Relationships** category involves interpersonal connections, such as friendships, kinship, mentorship, or other social bonds. It includes discussions of relationship problems and lessons learned from major events, comments on how characters relate to one another, and emotional interactions tied to those relationships (e.g., expressing concern for others, reacting to changes in relationships, appreciating someone's support or commitment).

The **Identity & Personal Growth** category focuses on a character's self-development, including their identity, values, beliefs, and personal growth through major events. It also includes lines about a character's internal experiences, such as their mental or physical feelings. The emphasis should be on the self rather than on relationships with others.

The **Authority & Leadership** category reflects leadership roles, responsibility, or authority, whether referring to the speaker or to others. It includes decisions about events that matter to the ponies involved and crisis-management dialogue (e.g., posing questions to assess whether a situation requires one's protective action). Declarations made toward the enemy also fall into this category, as they implicitly affirm the speaker's sense of duty.

The **History, Lore, & Magic** category provides world-building information, including Equestrian history, legends, artifacts, magical spells, and other forms of lore. It focuses on semantic knowledge about the world rather than personal opinions or emotional reactions.

The **Nature & Animals** category is about the natural world or non-pony creatures that cannot speak human-readable language. Non-pony creatures that can talk are treated as characters rather than animals, and therefore do not fall into this category.

The **Food & Drink** category refers to discussions about food or drinks, such as comments about them, invitations to share a meal, or cooking-related conversations.

Topic Characterization

Most Frequent Words Analysis Details for tf-idf scores for the top-scoring terms in each category are illustrated in Figure 1.

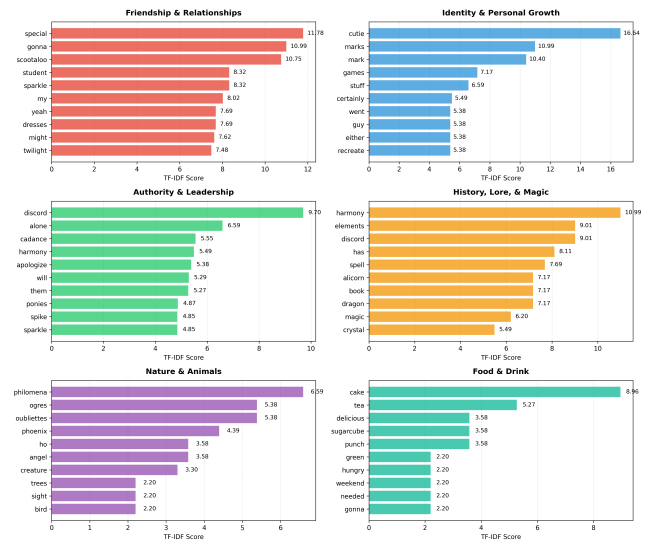


Figure 1: Scores of the top ten most frequent words by topic.

In **Friendship & Relationships**, the top ten frequent words are 'special', 'gonna', 'scootaloo', 'student', 'sparkle', 'my', 'yeah', 'dresses', 'might', and 'twilight'. 'Scootaloo' is the name of Sweetie Belle's best friend. The words 'twilight' and 'sparkle' form the name of the main character, Twilight Sparkle.

In **Identity & Personal Growth**, the top ten frequent words are 'cutie', 'marks', 'mark', 'games', 'stuff', 'certainly', 'went', 'guy', 'either', and 'recreate'. 'Cutie', 'marks', and 'mark' together refer to cutie marks, which are symbols located on ponies' haunches that represent one's most identifiable trait.

In **Authority & Leadership**, the top ten frequent words are 'discord', 'alone', 'cadance', 'harmony', 'apologize', 'will', 'the', 'ponies', 'spike', 'sparkle'. 'Cadance', 'sparkle', and 'spike' refer to names of Princess Cadance (ruler of the Crystal Empire), Twilight, and Twilight's baby dragon assistant, respectively.

In **History, Lore, & Magic**, the top ten frequent words are 'harmony', 'elements', 'discord', 'has', 'spell', 'alicorn', 'book', 'dragon', 'magic', and 'crystal'. 'Harmony' and 'elements' form the concept called Elements of Harmony, the six powerful artifacts that are frequently used to defeat enemies. 'Alicorn' refers to princess ponies who have both a unicorn horn and Pegasus wings. 'Crystal' refers to the Crystal Empire, a territory of Equestria ruled by Princess Cadance.

In **Nature & Animals**, the top ten frequent words are 'philomena', 'ogres', 'oubliettes', 'phoenix', 'ho', 'angel', 'creature', 'trees', 'sight', 'bird'. 'Philomena' is the name of Princess Celestia's pet phoenix/bird, and 'angel' is Fluttershy's pet bunny.

In **Food & Drink**, the top ten frequent words are 'cake', 'tea', 'delicious', 'sugarcube', 'punch', 'green', 'hungry', 'weekend', 'needed', and 'gonna'.

LLM summaries The **Friendship & Relationships** category centers on emotionally charged interactions where characters navigate misunderstandings, hurt feelings, and the desire to feel appreciated by important others. Twilight appears as the most prominent figure, often acting as the mediator, problem-solver, or source of emotional tension as others seek her guidance. Other characters like Fluttershy and Rarity also contribute to the emotional landscape. Sibling figures additionally bring emotional tension through family closeness.

The **Identity & Personal Growth** category centers on characters struggling to understand themselves, with cutie marks serving as the strongest symbol of self-exploration. Characters grapple with uncertainty, insecurity, and fear of not living up to expectations, often expressing their doubts through reflective, vulnerable dialogue. Apologies and efforts to improve are common, creating a tone that is introspective and focused on personal transformation.

The **Authority & Leadership** category centers on responsibility, decision-making, and the obligations of ruling or protecting Equestria. Twilight and the princesses are mentioned or in the dialogue most prominently, often navigating expectations tied to royal duty, strategic planning, and high-stakes threats. The dialogue frequently uses directive language, reflecting urgent problem-solving and the weight of leadership.

The **History, Lore, & Magic** category centers on large-scale worldbuilding, focusing on ancient magical systems, national history, legendary artifacts, and recurring threats that have shaped Equestria's past and present. The Elements of Harmony and the history of Equestria provide the backbone for many discussions. Characters such as Twilight, the princesses, Discord, and Tirek (i.e., a villain who tries to take advantage of Discord to steal magic and rule over Equestria) appear frequently as key figures connected to ancient lore or past conflicts, often speaking in a serious, prophetic tone about dangers, truth, and magical balance.

The **Nature & Animals** category focuses on wildlife and curious encounters with both ordinary animals and fantastical creatures. Philomena the phoenix stands out as the central figure. Fluttershy, as an animal expert, frequently appears in interactions to indicate care for animals or explain their behaviors. Some dialogue also blends natural elements with fantasy—such as ogres and oubliettes—which naturally overlaps with the History, Lore, & Magic category. This reflects a limitation of our typology, but it can be explained by the show's world-building, where magical forces or beings often appear throughout the natural environment of Equestria.

The **Food & Drink** category focuses on cozy moments centered around tea and cakes, often appearing during parties, banquets, and shared family/friend gatherings. The dialogue often involves polite tension between characters, which seemingly overlaps with Friendship & Relationship, but the focus is on food (e.g., thanks for sharing cakes). Fluttershy is the most prominent character in interactions, often appearing in tender hosting roles over snacks and tea.

Topic Engagement

Figure 2 illustrates the detailed percentage of engagement that each character allocates to each topic.

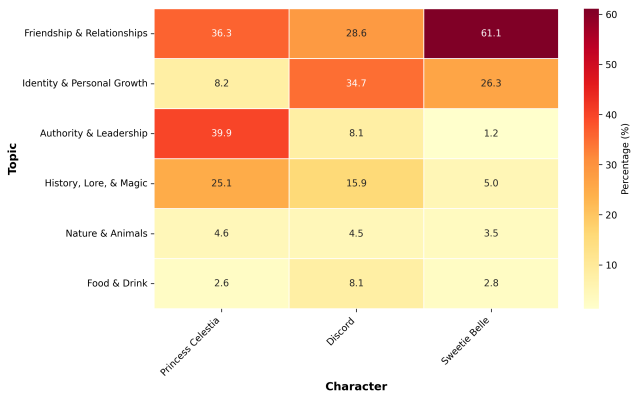


Figure 2: Topic Distribution Across Characters

For **Princess Celestia**, the data show that Authority & Leadership is the most prominent category at 39.9%. Friendship & Relationships is also substantial at 36.3%, followed by History, Lore, & Magic at 25.1%. The remaining categories play comparatively minor roles: Identity & Personal Growth (8.2%), Nature & Animals (4.6%), and Food & Drink (2.6%).

When examining the data for **Discord**, we observed that Identity & Personal Growth stands out as his strongest area of engagement at 34.74%. Friendship & Relationships accounts for 28.57%, and History, Lore, & Magic contributes 15.91%. Smaller portions come from both Authority & Leadership and Food & Drink, each at 8.12%. The remaining 4.5% is covered by Nature & Animals.

Looking at **Sweetie Belle**, Friendship & Relationships dominates her dialogue at 61.1%, while Identity & Personal Growth accounts for 26.3%. The remaining topics appear in much smaller proportions: History, Lore, & Magic contributes 5.02%, followed by Nature & Animals (3.45%) and Food & Drink (2.82%). Authority & Leadership is minimal for her, representing just 1.25% of her dialogue.

Discussion

This project aimed to investigate how three side characters with different narrative roles in My Little Pony: Friendship Is Magic distribute their dialogue across a set of topic categories. Our findings support the following points:

Big Point 1. Each character’s topic emphasis presents the story from a different narrative angle.

Princess Celestia spends most of her time talking about Authority & Leadership and History, Lore, & Magic (65%). This aligns with her role as Equestria’s ruler who cares much about the political matters and protecting her subjects. Our word frequency analysis revealed that highly mentioned words in dialogue of these two topics primarily focus on royalty references (e.g., Princess Cadance, Princess Twilight Sparkle, and Spike), magical artifacts (e.g., Elements

of Harmony), and Equestrian legends (e.g., history of Crystal Empire, the evil Tirek, dragon lands, alicorns). Princess Celestia provides most of these explanations, often briefing the main ponies before assigning them tasks vital to the kingdom's safety. She also devotes meaningful attention to Friendship & Relationships, mostly as a guide to friendship issues by telling the gorgeous truth, mostly towards her student, Twilight Sparkle. Overall, Princess Celestia's narrative offers world-building knowledge as well as valuable lessons on leadership.

Discord shows a very different distribution. One third of his dialogue falls under Identity & Personal Growth. At first, this seemed surprising because our word-frequency analysis showed that dialogue in this topic largely involves cutie-mark-related discussions, and Discord does not even have a cutie mark. However, we realized that Discord's entire redemption arc is essentially one long identity crisis. He constantly questions who he is without causing chaos, and much of his History, Lore, & Magic dialogue also reflects how deeply his chaotic powers are tied to his self-concept. He also engages heavily in Friendship & Relationships. Although initially dismisses the value of friendship, he gradually learns meaningful lessons from the main ponies. His emphasis on this topic largely lies on Fluttershy, the friend who changes him a lot and whom he cares most about. Overall, Discord's narrative shows how an antagonist can transform into a kind and courageous individual, ultimately forming genuine friendships.

Sweetie Belle focuses heavily on Friendship & Relationships and Identity & Personal Growth (in total 87.4%). This pattern makes sense given her age—she is still learning about herself and navigating friendships. Notably, Scootaloo's name is one of the most frequently mentioned words in the Friendship & Relationships category, always coming from Sweetie Belle. As best friends, the two offer each other consistent peer support. She sometimes talked about her sister Rarity, but the frequency is significantly lower compared to mentions to her friends in CMC. Overall, Sweetie Belle's narrative reflects the life journey of a child, emphasizing rich social learning and personal development.

Big Point 2. The differences in how much attention each character gives to each topic reflect perspectives shaped by their social roles and life stages, mirroring real-world dynamics.

As expected, all three side characters engage heavily with Friendship & Relationships, and all mention frequently the princess of friendship, Twilight Sparkle. This reflects the show's consistent storytelling emphasis that everybody cares about friendship and learns their own lessons from it, with Twilight being an important guide to them. However, there is an interesting age gap between the characters' topic emphasis. Sweetie Belle spends most of her dialogue navigating friend dynamics and managing her sisterhood with Rarity. In contrast, Princess Celestia and Discord emphasize this topic less. This is possibly because adults' concerns extend beyond interpersonal dynamics. This mirrors real-world patterns in which people at different life stages tend to prioritize

different types of issues.

Authority & Leadership shows the largest difference across characters' topic emphasis. Princess Celestia dominates this category, followed by Discord, and Sweetie Belle rarely pays attention to this topic. This makes sense because the topic is almost entirely reserved for characters with institutional power. Celestia handles royal duty and high-stakes threats, whereas Discord occasionally offers ideas during crises but rarely carries actual leadership responsibilities. Sweetie Belle's small presence (mostly coordinating CMC activities) is fundamentally different in scale from making decisions about protecting kingdoms. A similar pattern appears in History, Lore, & Magic, where adult concerns far outweigh those of a child character. Gaps in these emphases reveal that the more mature a character is—and the more responsibility their social role carries—the more likely they are to engage with deeper topics, such as politics, leadership, and history.

Identity & Personal Growth show another distinct pattern of comparison between characters' topic emphasis. Here, Discord pays attention most, followed by Sweetie Belle, with Princess Celestia engaging the least. This topic is characterized by self-reflection, and Discord's highest emphasis among the three characters aligns with his redemption arc. Sweetie Belle's emphasis reflects the identity exploration typical for children, especially in relation to cutie mark development. Celestia's limited emphasis aligns with her established role as someone who already knows who she is and what her duties are. Taken together, these patterns reveal that individuals who are undergoing transformation or uncertainty pay more attention to identity-related dialogue.

Across all three characters, Nature & Animals and Food & Drink receive the least attention. These dialogues tend to be less central to the storyline and instead help make the world feel more vivid by showing small moments of daily life and casual interactions. Notably, Discord devotes nearly 10% of his dialogue to these two topics, whereas the other two each allocate only about 5%. This difference can be explained by the nature of Discord's interactions—his closest friend, Fluttershy, is a dominant character associated with animals and tea parties. Thus, Discord chat about these topics more, influenced by Fluttershy. This pattern indirectly reinforces Discord's friendship-building arc.

Conclusion

The two big points we made support the conclusion that each side character contributes a distinct narrative perspective shaped by their social roles and life stages. These differences are not random: they reflect deliberate writing choices that lend authority to adult figures, foster identity exploration in characters in transition, and feature friendship-focused dialogue in children.

For the production company, these topic engagement patterns highlight how dialogue functions as a tool for maintaining consistent characterization while also supporting the show's educational goals. Our findings on each character's unique speech profile can guide the development of new episodes featuring these supporting characters, ensuring their behavior remains coherent with their established

roles. Furthermore, understanding how each character's role aligns with their speech profile provides insights for designing interactive products that appeal to diverse audiences and allow them to learn from a character's unique perspective. For example, young children can relate to Sweetie Belle's peer friendship work, while adult audiences may appreciate Celestia's authority and leadership themes. All audience can connect with Discord's identity journey as evidence that personal transformation is normal. By aligning new creative work with these dialogue patterns, the production team can continue to expand the My Little Pony universe while preserving the integrity and appeal of each character.

Future research could expand this analysis in several directions. First, our analysis used only dialogue from the first eight seasons of the show. Incorporating Season 9 dialogue may complete side characters' development arcs and capture any new characters or dynamics introduced in the final season. Second, while our study focused on individual character patterns with few peripheral evidence supporting interpretation, network analysis examining character interactions could reveal which character pairs or groups generate the most dialogue together. This would be valuable for production companies interested in fan engagement, given current trends around "shipping" (making character couples). Knowing which character combinations produce the most interaction could inform decisions about merchandise, spin-off content, or storyline innovation.

Group Member Contributions

All three members collaboratively defined and reframed the question, designed the methodology, annotated the data, synthesized the key findings, and reviewed the final report.

Cathy Chen provided contextual information about the show, wrote the Introduction, Data, and Methods sections.

Tanav Bansal developed the scripts for data collection and data analysis and wrote the Results section.

Yifan Lin developed the scripts for data collection and visualization and wrote the Discussion section.