



Justin Carlou "Jus" Lim | DigHum100 Dr. Adam Anderson | Summer 2021

Project Description

For this project, I will be exploring the representation of gender in the television show, *The Simpson's*.

Dataset

For this project, I use Prashant Banerjee's "The Simpson's Dataset" (2020) posted on Kaggle. This dataset was originally web-scraped by Tod Schneider. The dataset is split into four datasets: the character dataset, the episode dataset, the locations dataset, and the script lines dataset. Because the character dataset and the script dataset had information about characters, speaking lines, and gender, I focused my analysis on these two datasets. For simplicity's sake, I merged these two tables into one and dropped the irrelevant columns.

```
list(s_script.columns)

['id',
 'episode_id',
 'number',
 'raw_text',
 'timestamp_in_ms',
 'speaking_line',
 'character_id',
 'location_id',
 'raw_character_text',
 'raw_location_text',
 'spoken_words',
 'normalized_text',
 'word_count']
```

Note

** expressions of gender in this project are confined to a binary because that is what is provided in the dataset. Further analysis in including nonbinary genders would be fruitful to explore in the future.

```
list(s_char.columns)

['id', 'name', 'normalized_name', 'gender']
```

	episode_id	normalized_name	char_id	gender	line	timestamp_in_ms	raw_location_text
0	32	miss hoover	464	f	No, actually, it was a little of both. Sometim...	848000	Springfield Elementary School
1	32	miss hoover	464	f	I don't know. Although I'd sure like to talk t...	856000	Springfield Elementary School
2	49	miss hoover	464	f	Thank you, Ralph. Very graphic. Lisa Simpson? ...	784000	Springfield Elementary School
3	49	miss hoover	464	f	Dear God!	803000	Springfield Elementary School
4	51	miss hoover	464	f	I question the educational value of this assem...	226000	Springfield Elementary School

Questions for Analysis

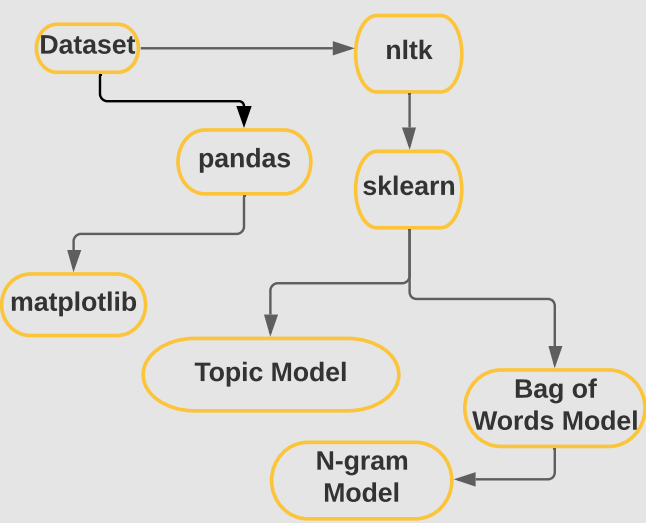
1. How has the inclusion of female characters changed over time?
2. How does dialogue between male characters differ from dialogue between female characters?

Hypotheses

1. Because *The Simpson's* is a show which follows pop culture closely, I do believe that the preportion of female characters will change, but only slightly.
2. I also believe that conversations between folks of the same sex will be stereotypically gendered.



Tools & Methods



Workflow

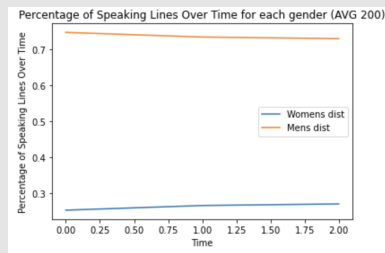
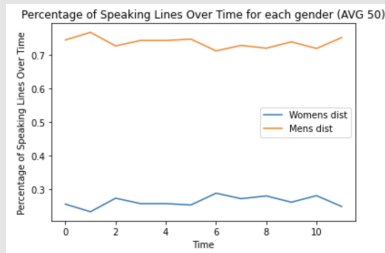
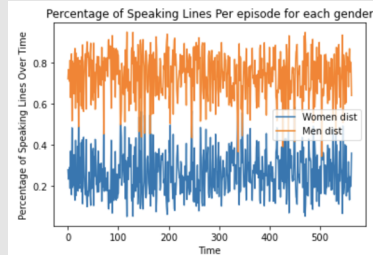
In order to track representation of women over time, I used methods from the pandas library to find the distribution female speaking roles to male speaking roles over time, then used matplotlib to visualize the results.

To compare dialogue between male and female characters, I created a word cloud to give a visual representation of common phrases, then created an N-gram model to see differences in actual phrases. Moreover, I supplemented the analysis with a topic model out of each group's dialogue, then visualized the topic network.

Conclusions

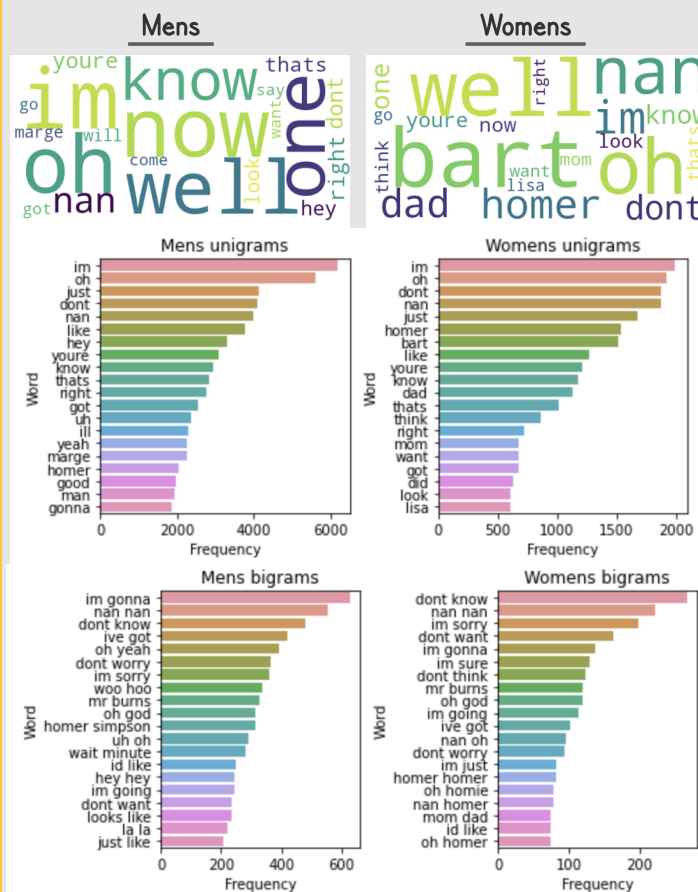
1. Women are highly underrepresented in the tv show.
2. Those that are more prominent mostly function as supporting roles for the men characters (like being a wife, or being a teacher who plays a maternal role).

Gender Distribution Over Time



I plotted the gender distribution of speaking roles for each episode in the top graph. As it resulted in overplotting, I took the average of an increasing group of data points (from 50 - 200) to show clearer trends. As we can see from the bottom graph, there seems to be an increase in female speaking roles over time, but it is small and insignificant.

Comparing Dialogue



Based on these word clouds, I found that most of the women speaking characters are maternal figures to Bart Simpson, the son (explanation in jupyter notebook). Moreover, Lisa Simpson, the daughter, makes up a huge majority of womens speaking roles. Both the unigram and bigram analyses support the main points found from the word cloud.

Womens

Topic #1:
youre, want, did, look, okay, make, love, theres, kids, say, yes,
home, school, help, great, honey, weve, boy, remember, talk
Topic #2:
oh, like, know, little, really, time, ive, gonna, maybe, simpson,
mean, better, wait, god, isnt, youll, baby, night, things, bad
Topic #3:
im, think, right, mom, lisa, going, ill, hes, come, sure, hey, didnt,
yeah, doing, let, father, marge, said, big, wont
Topic #4:
dont, homer, bart, good, sorry, people, need, whats, tell, day, youve,
theyre, new, house, maggie, thought, nice, lot, shes, ooh
Topic #5:
nan, just, dad, thats, got, homie, stop, mr, way, man, thing, lets,
thank, uh, guess, family, feel, id, doesnt, believe

Mens

Topic #1:
oh, dont, like, youre, ill, uh, good, man, little, come, say, simpson,
make, dad, lisa, let, life, help, mean, place
Topic #2:
know, marge, gonna, want, think, boy, day, sorry, old, god, son, work,
guys, money, mom, guy, school, youll, thanks, away
Topic #3:
im, just, nan, thats, got, yeah, homer, look, okay, ive, did, hes,
really, theres, thing, sir, sure, kids, new, ya
Topic #4:
right, time, yes, way, lets, love, going, whats, tell, mr, great, didnt,
id, big, home, theyre, said, stop, youve, hell
Topic #5:
hey, bart, need, people, wait, better, huh, gotta, thank, maybe,
springfield, thought, family, hello, long, aw, whoa, ooh, isnt, em

In visualizing the model, and going through each groups' topics, a main difference I noticed was that womens' dialogue is very centered around the topic of family and mother-daughter bonding, while mens' dialogue focuses mostly on Homer and his friend group or Bart and his friend group.

Works cited:

- Banerjee, P. (2020). *The Simpsons Dataset*. Kaggle: <https://www.kaggle.com/prashant111/the-simpsons-dataset>
- Cowles, D. L., Austin, M., & Houston, G. T. (1994). Feminist Criticism. In *The critical experience: literary reading, writing, and criticism* (pp. 214-234). essay, Kendall/Hunt.
- Jessup, J. B. (2021). *Topic Modeling* [ipython notebook]. Retrieved from University of California, Berkeley Digital Humanities 101: Practicing the Digital Humanities.

