

Topic Modeling for TV: A Reflection of American Culture Through Popular Sitcoms, 1950s-2000s



Chris Raffa, Professor Jon Chun, & Katherine Elkins
Kenyon College, Artificial Intelligence



© Can Stock Photo

Abstract

Using stylometry, and topic modeling, I have studied the American sitcom and how the culture reflected in that humor has changed and evolved over the last fifty years. I've selected two of the most popular series from each decade from the 1950s (*I Love Lucy*) to the 2000s (*The Office* [U.S.]), with popularity based on Nielsen ratings at the time and lasting cultural impact today. I then ran transcripts of every episode through these tools. I found that the content of comedy for a mass audience has remained surprisingly consistent over the last half-century, despite rapidly changing social norms and culture. The greatest change came from the 1970s to the 1980s, shifting focus away from the working class suburban lifestyle to the wealthier urban life, and in the liberality with which sex was discussed. Despite these changes, the American sitcom still held on to discussions of gatherings, such as Christmas, weddings, and birthdays, though this too saw some secularization in the 2000s.

Introduction

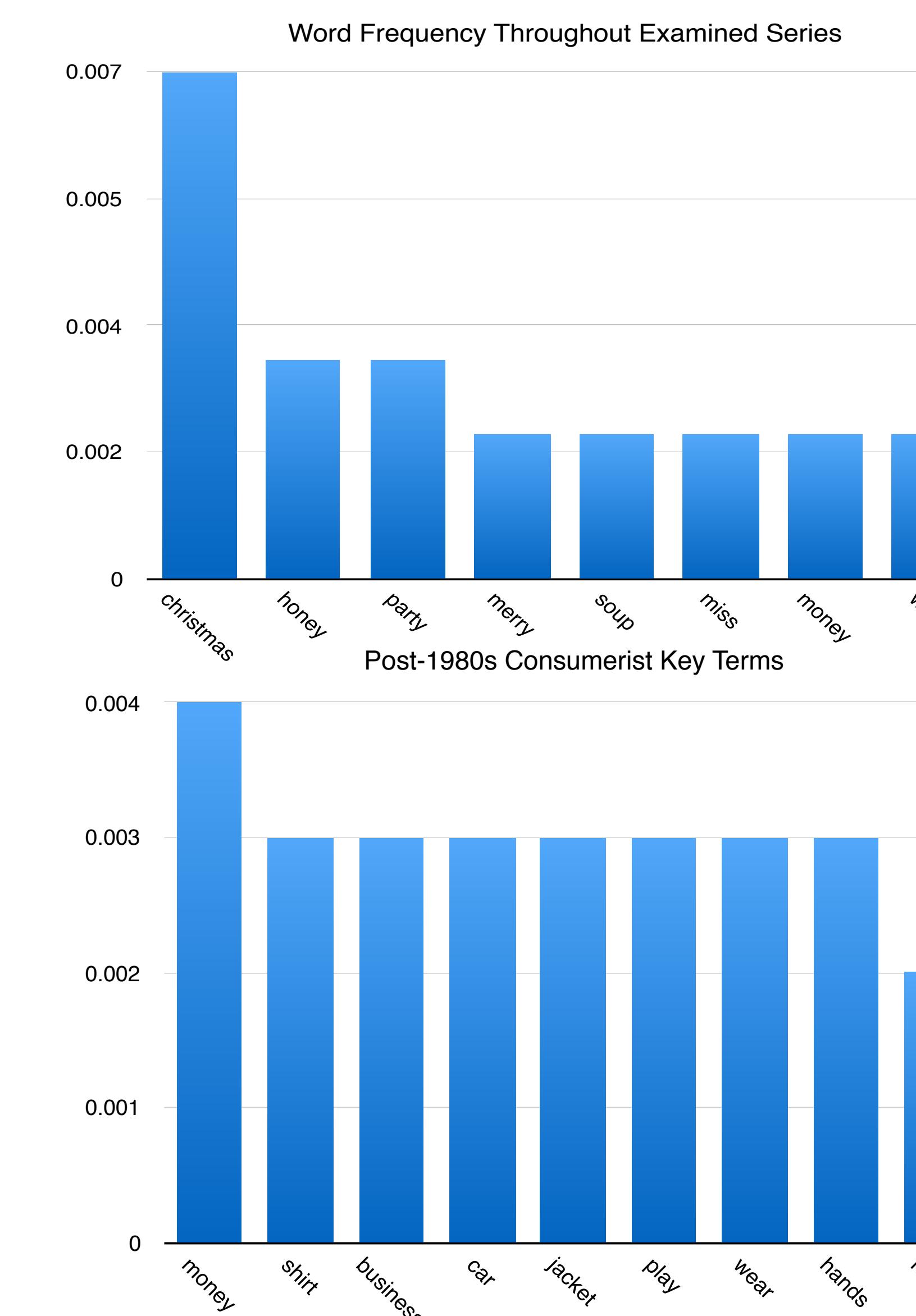
There is perhaps nothing so universally human about our ability to create comedy. As far as we can tell, we are the only species that is able to produce and utilize humor, as a means to forge social relationships, relate to one another, and cope with the world around us. Perhaps for this reason humor has presented a difficult problem for machine learning, as it involves numerous integrated cognitive behaviors that are difficult to teach machines. To be funny is not just a last-ditch attempt to appeal to your preferred sex. It requires theory of mind, linguistic mastery, and an awareness of self and culture, all working together with form, tone, and speed for optimal delivery. Some have suggested (Mulder & Nijholt, 2002) that it may be the final frontier for successful AGI.

Given the nature of humor and the mass market demands of the American sitcom, in these scripts there is bound to be a reflection of the culture that produced it. Comedy has always been a place of both relation and transgression. We laugh because we understand, and we laugh when there is a violation of some invisible law, whether logical in the form of absurdity, or social in the form of taboo. From the conflicts displayed in the show, and how the characters feel and laugh about them, we can discern how America felt and laughed. These shows were created to capture the national eye, and thus could not be too transgressive without being accessible, in the way that a darker comedy of HBO or Netflix fare today may be.

Natural language processing tools have already done work in identifying critical linguistic patterns and the implications of those patterns. This includes work on language analysis to understand the composition of humorous statements (Yang et. al. 2015), in which AI was able to successfully identify comedy writing from more serious work. Study utilizing stylometry has been done on literature, such as its use on James Joyce's *Ulysses* (Byrnes, 2010) to detect cliché density in narration regarding the novel's various characters. There has been a recent push toward using these tools for the study of television, such as stylometry used to determine the personality traits of Sheldon Cooper of *The Big Bang Theory* (Zyl & Botha, 2016) and to predict humor responses based on delivery (Bertero & Fung, 2016). This experiment hopes to gain similar results on the conventions of humor and varying social norms as displayed through humor.

Due to the limited amount of data, I unfortunately was only able to pick one show from the 1950s and 1960s, with *I Love Lucy* for the former and *Bewitched* for the latter. I would've rather picked a series like *Beverly Hillbillies*, *The Andy Griffith Show*, or *Leave It To Beaver*, which had higher Nielsen ratings than *Bewitched*, but the data was simply not available.

My picks for the 1970s-2000s are as follows:
1970s: *M*A*S*H* (11 seasons, spent most its life around #4) and *All in the Family* (9 seasons, #1 for five years)
1980s: *Cheers* (11 seasons, entered top 5 in season 4 and peaked at #1) and *Family Ties* (entered top 5 in season 3, peaked at #2)
1990s: *Seinfeld* (9 seasons, entered top 3 in season 5 and hit #1 twice, including its last season) and *Friends* (10 seasons, in top 5 for every season but one, peaked at #1)
2000s: *The Office* (U.S.) and *How I Met Your Mother*. At this point Nielsen ratings become irrelevant, but the former had a viewership of 5-9 million, and the latter between 8-12. Both ran for 9 seasons.



Materials

Springfield Springfield! Movie and TV Script Database. Python. Microsoft R Visual Code Learner. Python/Anaconda. Jupyter notebook, LDA over topic directory topic modelling program.

Methodology

Professor Chun and I developed a program to scrape the data from the Springfield Springfield! Database entitled Script Scraper.py. Once a base url was entered, the program would scrape all the text from a prescribed number of related page links, and classify them into their own separate text files. From there, I used a topic modelling program Professor Chun helped develop in order to find, based on term, what subject matter appeared most in each of these shows, by individual series and by decade. I then took the first twenty terms that appeared for each show, and each decade, disqualifying character names (e.g. Michael, Lucy) and terms specific to the series' premise (e.g. lieutenant for *M*A*S*H*, bar for *Cheers*).

Results

The terms brought up were much more consistent than originally hypothesized, and it is unclear how much of this necessarily has to do with American cultural fascinations or the conventions of sitcom. Two topics consistently brought up were of get-togethers, and of women, in a broad or romantic context. In every decade of television analyzed, there was a high quantity of references to holidays and significant gatherings: "Christmas" (consistently having high frequency), "birthday," "party," "wedding," "thanksgiving;" and numerous references to marriage ("married," "wife," "husband,"). It should be noted however, that the 2000s were the first decade in television studied here where "thanksgiving" and "Halloween" had a higher frequency than "Christmas." With these were the associated pleasures, "food," "eat," "play," and "drink," appearing often. References to drinking were and are plentiful.
Women were always a topic of discussion, though the terms with which they were discussed changed over the years. Starting in the 1950s and '60s with "madame," "miss," to the '70s and '80s' discussions of "women," "girls," "lady." It can be assumed that these discussions were contextualized by the pursuit of these women, as results show associations with terms like "date," "couple," "hot," and later in the 2000s, "sex," "naked," and "pants."
One of the most significant cultural shifts detected from the topics pulled out was the shift between the 1950s-70s to the 1980s-2000s in terms of setting and focus. Each of the popular sitcoms pre-1980s analyzed here (*M*A*S*H* being the exception) took place in suburban settings, and focused on domestic issues, usually regarding family. Terms like "mommy," "daddy," "child," "father," suggest emphasis on family, while terms like "street," "tree," "car," suggest a suburban setting. Families in these sitcoms were always working class; "class," appearing particularly high in the 1970s, as well as "loan," and "business."
With the 1980s and afterward, we see a shift away from the working class family and toward well-off urban residents. The most popular sitcoms in the 1990s and 2000s take place away from families, instead either in the workplace or with friends. "Apartment," "coffee," a higher use of "business," and "city," suggests the move to an urban setting. Of the four popular sitcoms chosen from this era, three take place in New York City. There is also a shift toward a focus on the protagonist's parents, "parents," "dad" and "mom" appearing highly in many series. The working class struggles have changed to consumerist wealth and spending, wanting and needing; "buy," "money," ("gotta" and "wanna" were high terms in the 1980s) and a shift to aesthetic emphasis and materialism; "hair," "ring," "shirt," "suit," "beautiful."



Conclusion

On the whole, what works in popular sitcoms has remained constant for the American audience throughout the last fifty years. There is an American value in rituals; holidays, weddings, and birthdays especially. Perhaps this is because our day-to-day lives involve these gatherings, or because we value these moments of levity and company with others. Romance, particularly the pursuit of women, has also been a consistent theme in these sitcoms. I believe this to be the result of primarily male-written, male-leading sitcoms for whom having love reciprocated or consummated is a prevalent concern and a good objective for an episode. Not to mention, both of these objectives are outlined by invisible social norms, and the violation of these norms is good for comedy.
American culture's increased comfort engaging in and talking about sex is reflected in popular television. What may have been implied before the 1980s, becomes much more explicit with characters talking about it directly, using the term "sex" and being "naked" on the show—a far cry from the separate twin beds married couples used to sleep in. Notable too is the secularization of culture in America, becoming most apparent in the 2000s, as "thanksgiving" became more used than "Christmas." I imagine this is of the "War on Christmas" effect, the push toward representation of religious diversity and secularization. Not everyone in America celebrates Christmas, but everyone in America celebrates Thanksgiving.
The most significant cultural change reflected in popular television comes from the change in culture following the economic boom and consumerist greed of the 1980s. Following the television of the 1980s, we've seen a move away from the working class families of the Bunkers and the Ricardos, and toward the upper class lives of the cast of *Seinfeld* or *Friends*, living in spacious apartments with specialized careers. Sitcoms have moved from the relatability of the working class family to wish fulfillment, from the stationary domestic life to the upward mobility of being twenty-something in an urban area, signifying a cultural shift away from the experience of the wealthy over the common person.

References

- Dario Bertero, Pascale Fung, "Predicting humor responses in dialogues from TV sitcoms", *Spoken Language Technology Workshop (SLT) 2016 IEEE*, 2016.
- Binsted, Kim and Ritchie, Graeme. *Computational Rules for Generating Punning Riddles*. In: HUMOR, the International Journal of Humor Research, Mouton de Gruyter, volume 10-1 pages 25-76, 1997.
- Mulder, M. P., and A. Nijholt. "Humour Research: State of the Art." *CiteSeerX*, Sep. 2002, pp. 1-24.
- Zyl, Maryka van, and Yolande Botha. "Stylometry and Characterisation in The Big Bang Theory." *Literator- Journal of Literary Criticism, Comparative Linguistics, and Literary Studies*, vol. 37, no. 2, 22 Nov. 2016, pp. 1-11.
- The Springfield! Springfield! TV Scripts Database