

Happy Moments for Generations of American

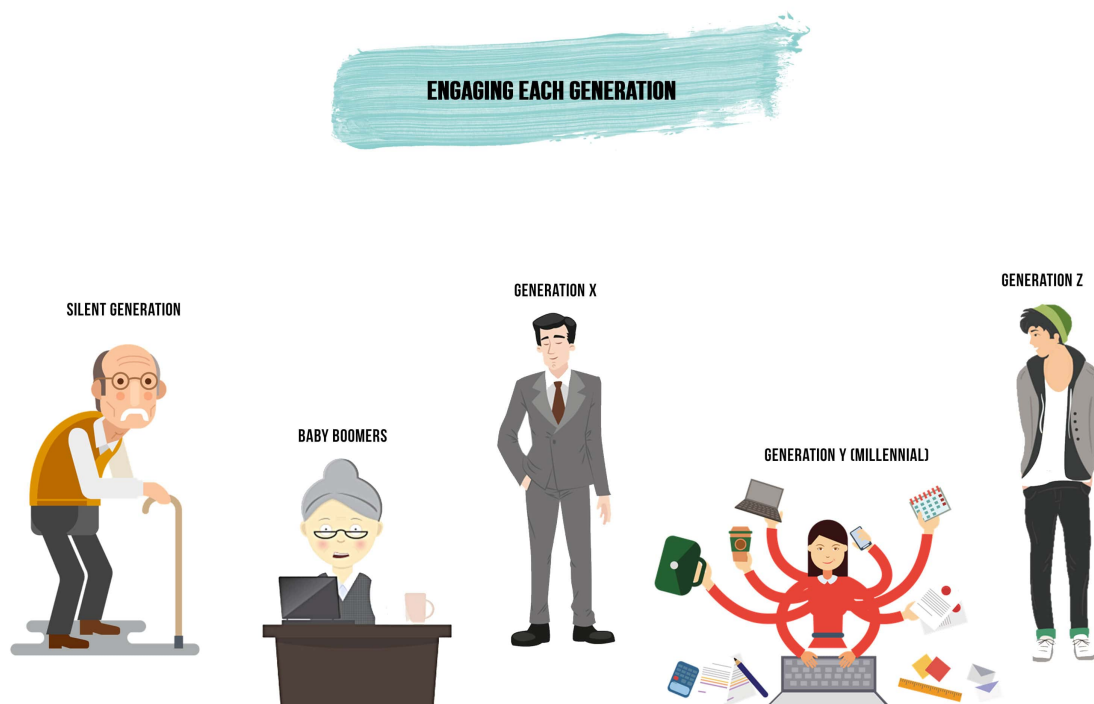


Figure 1:

There are countless moments of happiness scattered through our lives, literally there are countless moments of happiness scattered through every day. The touch of a loved one, the beauty of a sunrise or sunset, the smell of fresh air, a smile from a stranger. This project dives into a well-organized dataset (HappyDB) that can help explore some fundamentals of happiness.

In 20th century, USA has five generation groups: the Greatest Generation (or GI Generation), the Silent Generation, baby boomers, Generation X and millennial. About this project, I curiously thought about how different generations will responded this same question “What made you happy today?” or “What made you happy last 3 month?”. In order to look at happy moments for different generation groups, I will explore the texts using tools from text mining and natural language processing.

Part 1: Preparing Data and EDA

@ Why I only analysed USA for this dataset?

Reason 1: Almost 82% of participants are came from USA.

Reason 2: Different countries may have different definition about their happy moments.

@ Why I only analysed all generation groups except the Greatest Generation?

Reason : Because 99% of participants were ranged between 17 years old and around 90 years old.

The groupings below are based on studies by the US Census, Pew Research and demographers Neil Howe and William Strauss. Because of age range inside our dataset, this project only analysed four generation groups: Silent, Baby boomers, Generation X and Millennial.

Group 1 - The “Silent Generation”: Born 1925-1945 or 73-93 years old (in 2018). “Silent Generation” were named because they were more cautious than their parents.

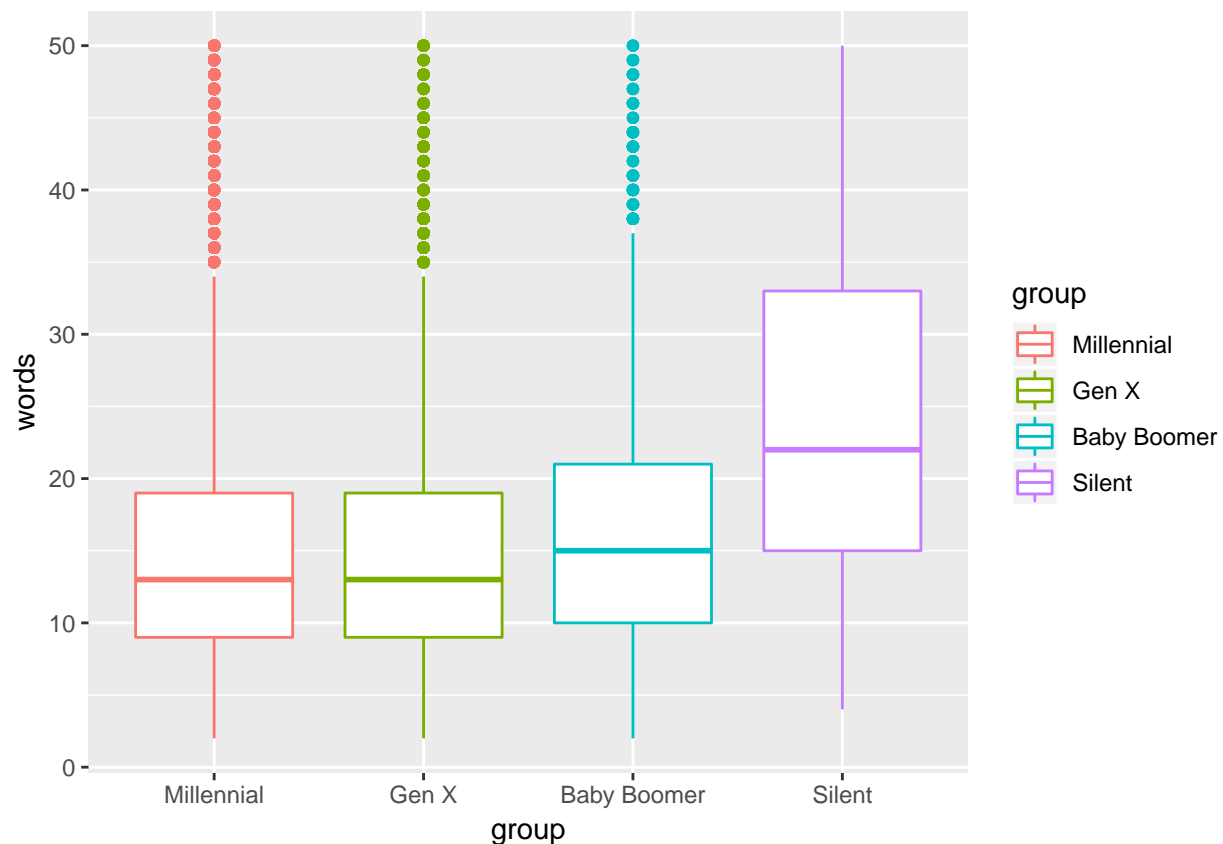
Group 2 - Baby Boomers: Born 1946-1964 or 54-72 years old (in 2018). Baby boomers were named for an uptick in the post-WWII birth rate.

Group 3 - Generation X: Born 1965-1980 or 38-53 years old (in 2018). Novelist Douglas Coupland used the term as the title of his first book, “Generation X: Tales for An Accelerated Culture,” published in 1991.

Group 4 - Millennials: Born 1981-2000 or 17-37 years old (in 2018). Howe and Strauss introduced the term millennials in 1991, the year their book, “Generations,” was published.

1.1 Overview of sentence length distribution for each generation

This step is used to analyse how many words each group approximately used to describe their happy moment. After deleting some outliers, we used ggplot to run the boxplot with four groups.



According to the boxplot, we could find that both Millennials and Gen X are no difference between their length of happy moments. Baby Boomer has slightly more words to describe their happy moment than first two groups. Significantly, Silent group are more likely to use the most words to describe their happy moment than the other three groups. This boxplot also shows that as age goes up, people are more likely give more details in one thing and more carefully organize their language to express their emotion. Basically, these results are consistent with the social cognition for elder people who have more patient and thorough grasp of a subject.

Part 2: Topic Modeling

2.1 What each generation focus on? (Wordcloud)

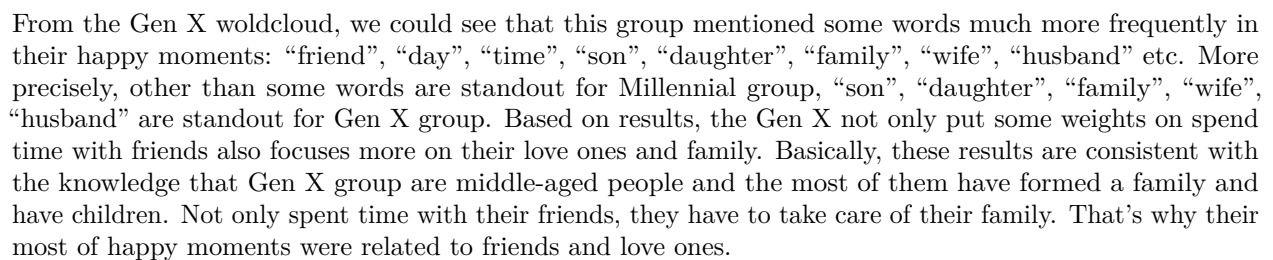
In this part, I would like to know these four groups focuses and get some inspiration on the topics these four groups concentrate on. Therefore, I counted the frequency of words in four groups seperated by using tidytext package and tried to find the most often used words in each group's happy moment. Below are the word clouds

2.1 (a) WordCloud for Millennials group



From the Millennials wordcloud, we could see that this group mentioned some words much more frequently in their happy moments: “friend”, “day”, “time”, “game”, “dinner”, “played”, “finally” etc. More precisely, “friend”, “day” and “time” are standout for Millennial group. Based on results, the Millennials put more weights on spend time with friends or play games with friends. Basically, these results are consistent with the knowledge that Millennials grew up with technology - computers, cell phones, internet, etc. which can easily make new friends and play with them. That’s why their most of happy moments were related to their friends.

2.1 (b) WordCloud for Gen X group



mentioned some words much more frequently than the Gen X group. For example, “home”, “wife”, “husband” etc. Based on the results, it can be concluded that the Gen X group is similar to the Gen X group. They mentioned that most of their happy moments were related to family and home life.



From the Silent worldcloud, we could see that this group mentioned some words much more frequently in their happy moments: “husband”, “favorite”, “received”, “found”, “friend”, “daughter”, “home” etc. More precisely, other than some words are standout for other three groups, “received” and “favorite” are standout for Silent group. Basically, these results are consistent with the knowledge that the Silent group are elder people and the most of time they will received favorite stuff from thier love ones and watch their favorite game or show. That’s why their most of happy moments came from friends, love ones and their own interesting.

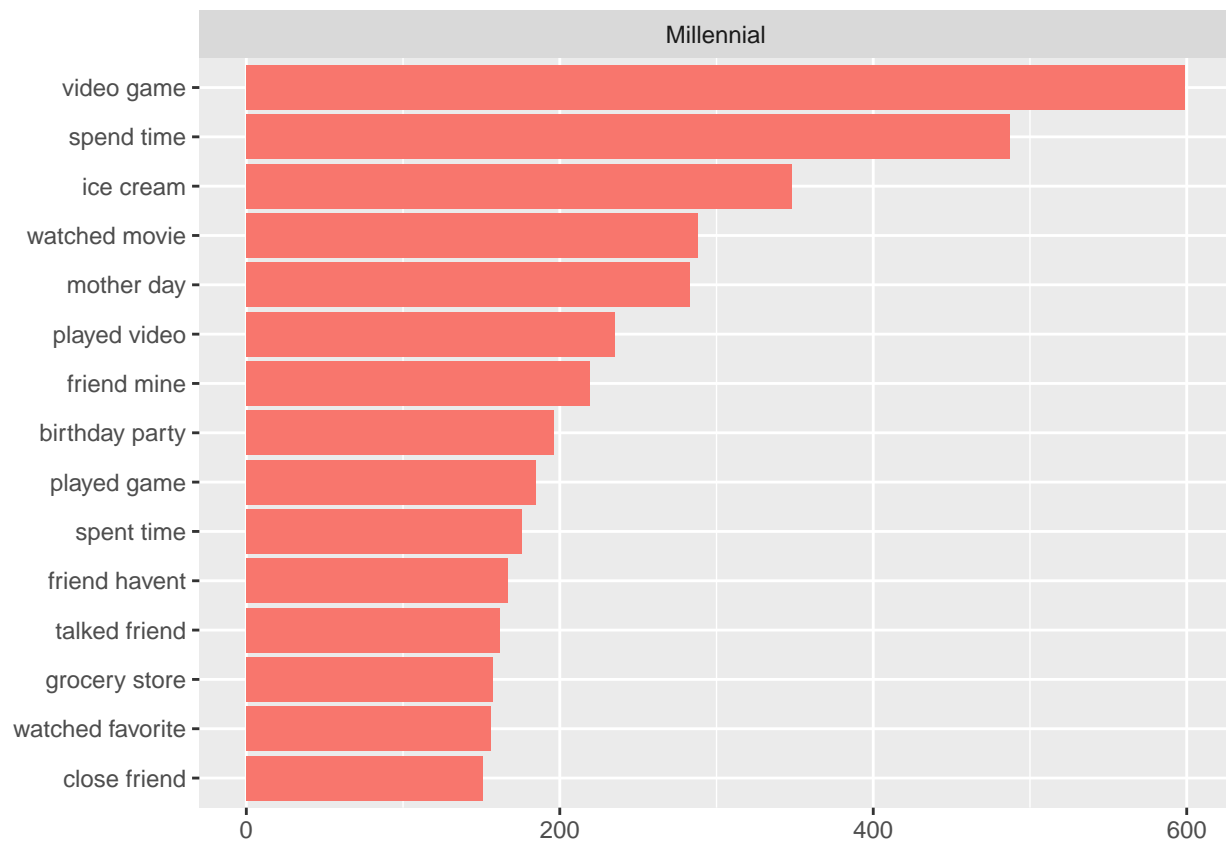
From the four wordcloud plots, we can see that four groups used some slightly different high frequency words to describe their happy moments, but most of the words are similar and related two categories - friends and family. Therefore, no evidence to show that some generations have significant different happy moments than others.

2.2 Relationships between words in four generations (bigrams)

So far we've considered words as individual units. However, many interesting text analyses are based on the relationships between words, whether examining which words tend to follow others immediately, or that tend to co-occur within the same documents. Pairs of consecutive words might capture structure that isn't present when one is just counting single words, and may provide context that makes tokens more understandable.

In this section, I'll explore some of the methods tidytext offers for calculating and visualizing relationships between words in each generation's happy moments dataset.

2.2(a) Analyzing bigrams for Millennial group

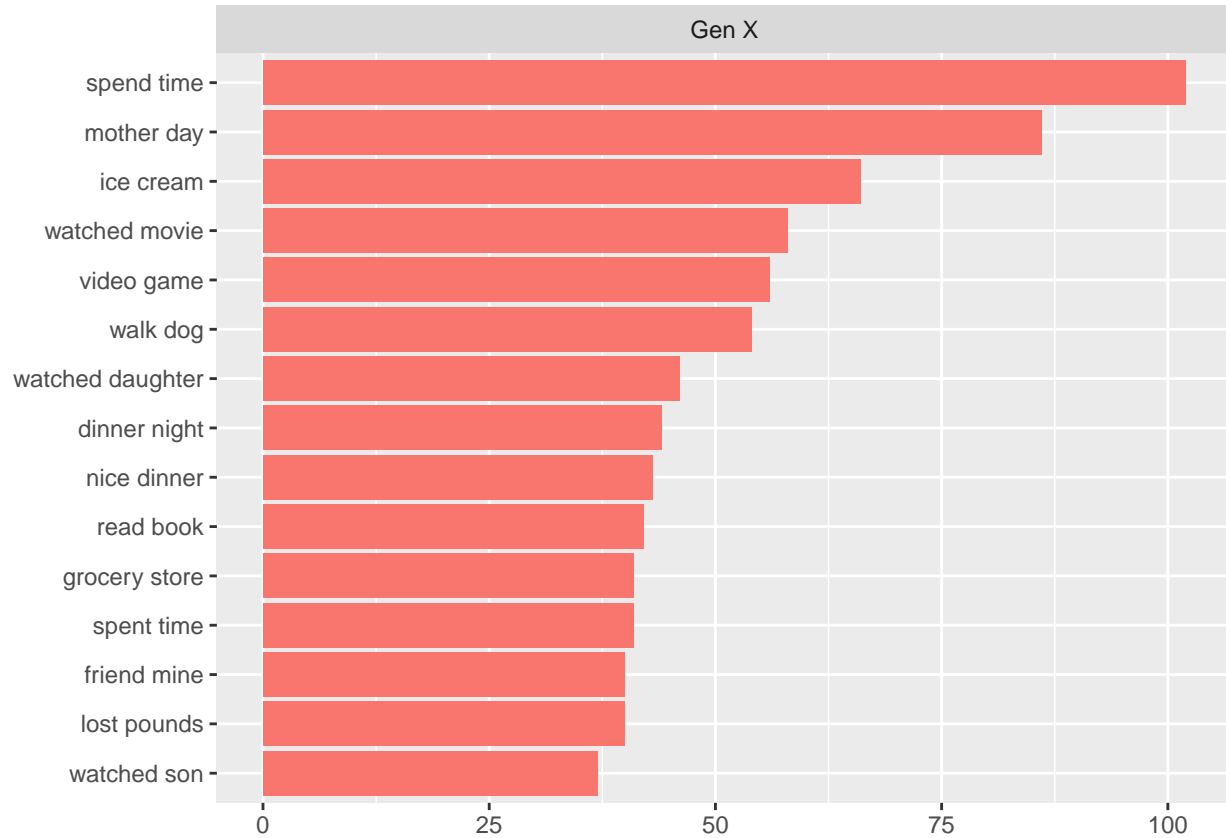


Visualizing a network of bigrams for Millennial group

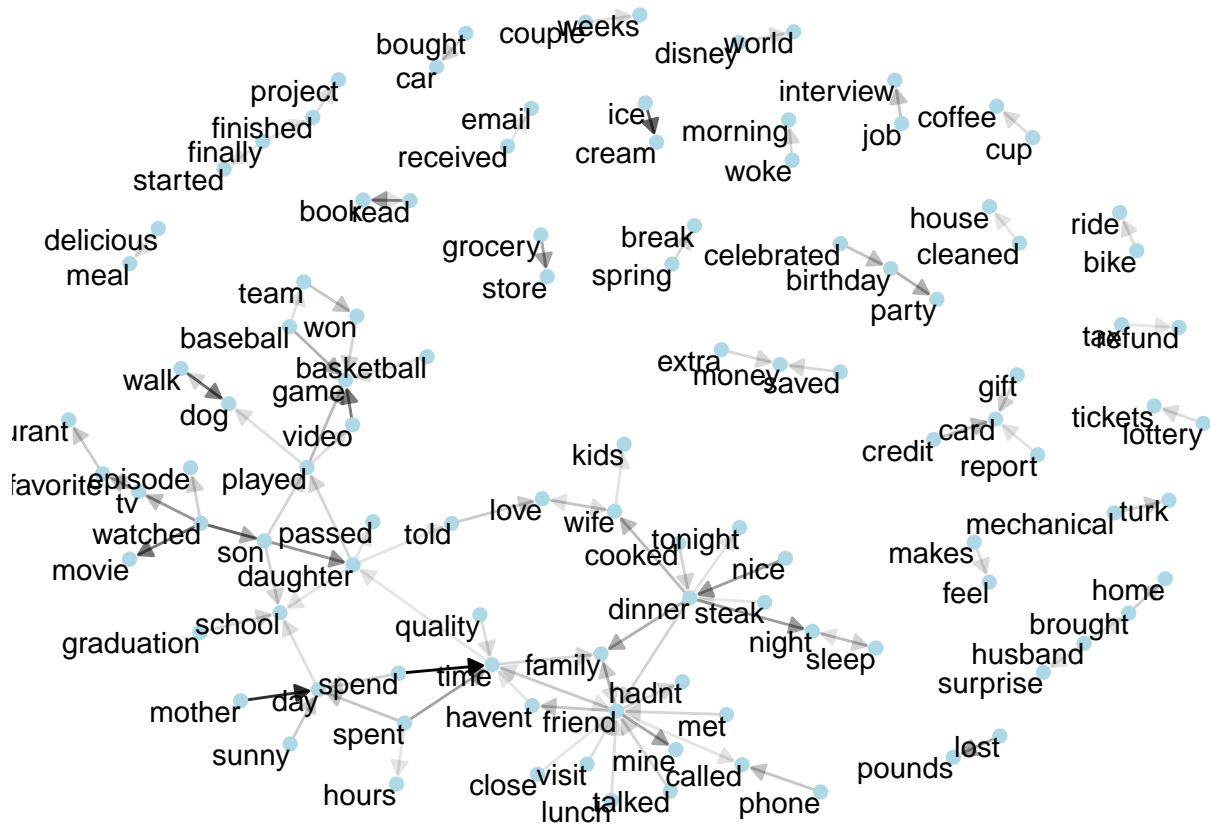


From the Millennials bigram graph, we can see that “video game”, “spend time”, “Ice cream” are the most common pairs in Millennials group. In network bigrams, we can visualize some details of the text structure. For example, we see that key words (which we discovered in section 2.1 (a)) such as “friend”, “game”, “family”, “day”, “time” etc. form common centers of nodes. We also see pairs or triplets along the outside that form common short phrases (“ice cream”, “grocery store”).

2.2 (b) Analyzing bigrams for Gen X group

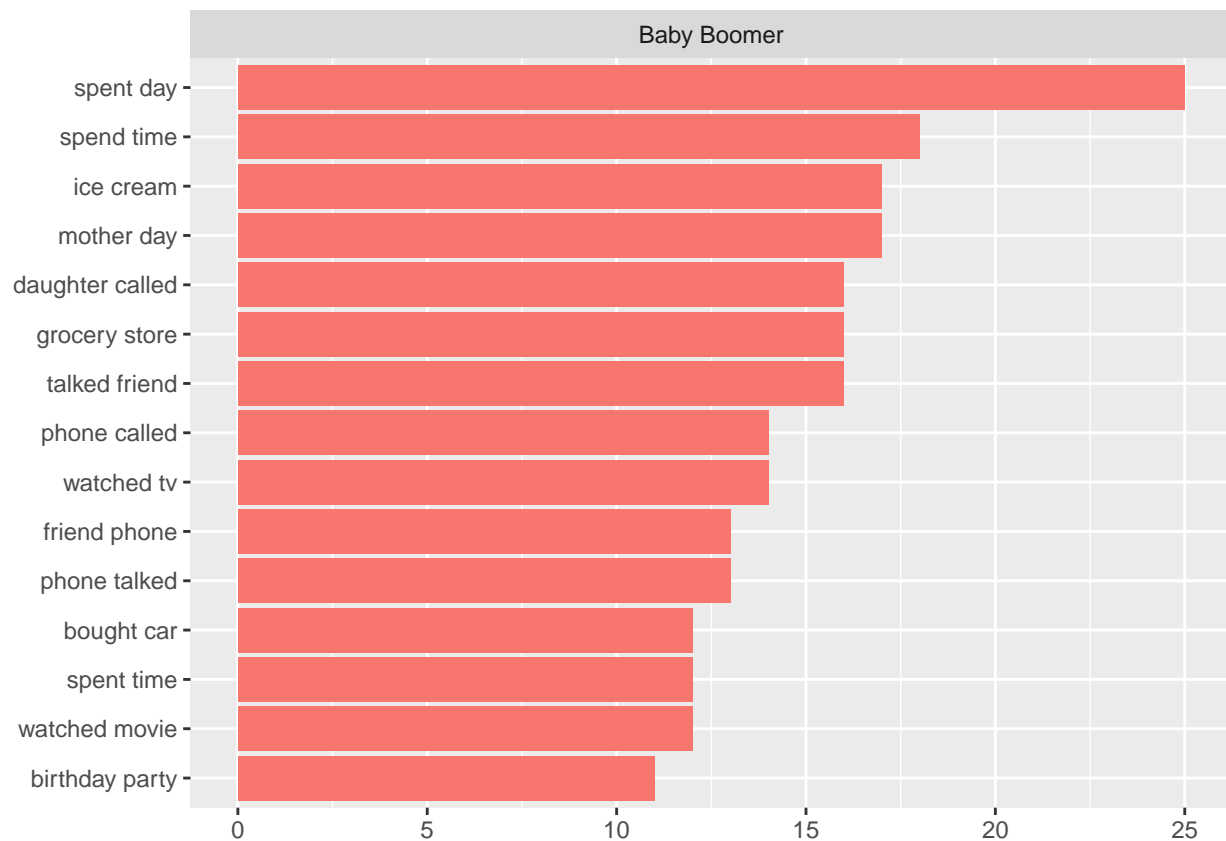


Visualizing a network of bigrams for Gen X group

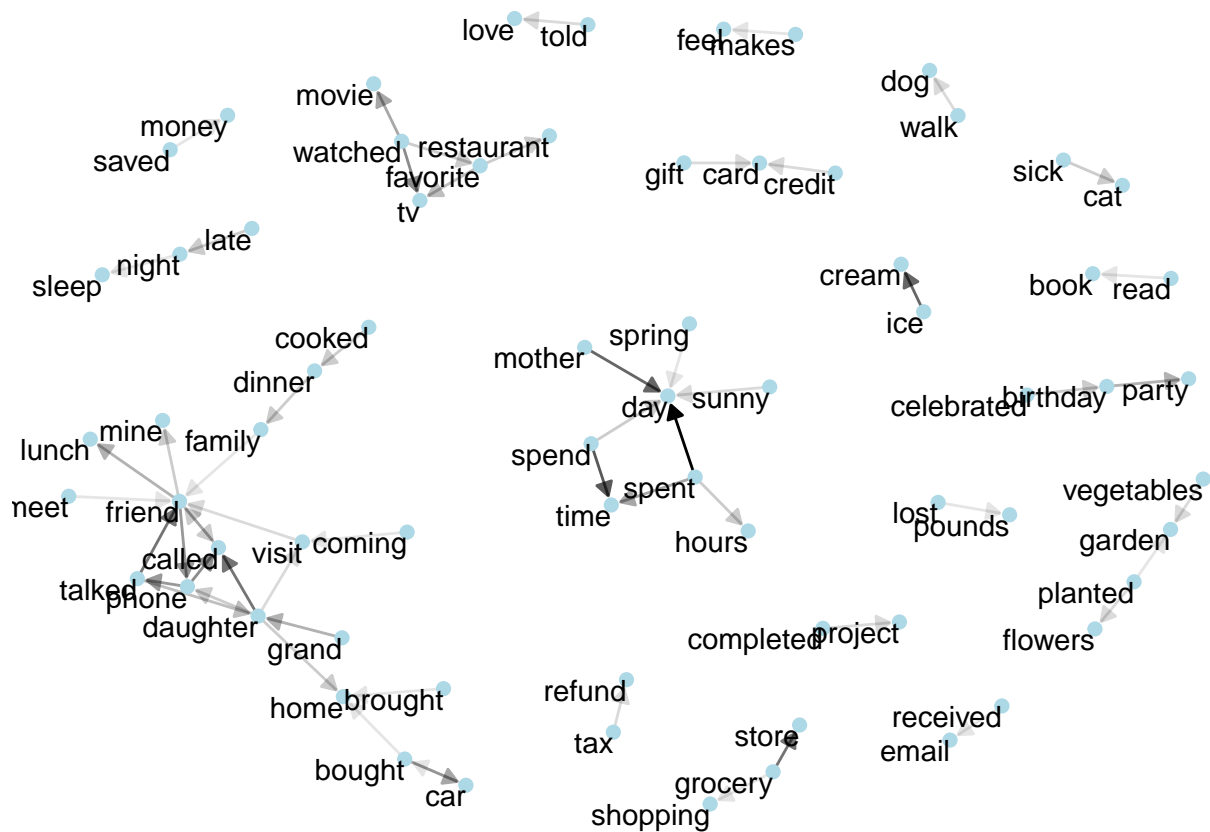


From the Gen X bigram graph, we can see that “spend time”, “mother day”, “Ice cream” are the most common pairs in Gen X group. In network bigrams, we see that key words such as “friend”, “game”, “family”, “son”, “daughter”, “time” etc. form common centers of nodes. We also see pairs or triplets along the outside that form common short phrases (“ice cream”, “read book”).

2.2 (c) Analyzing bigrams for Baby Boomer group

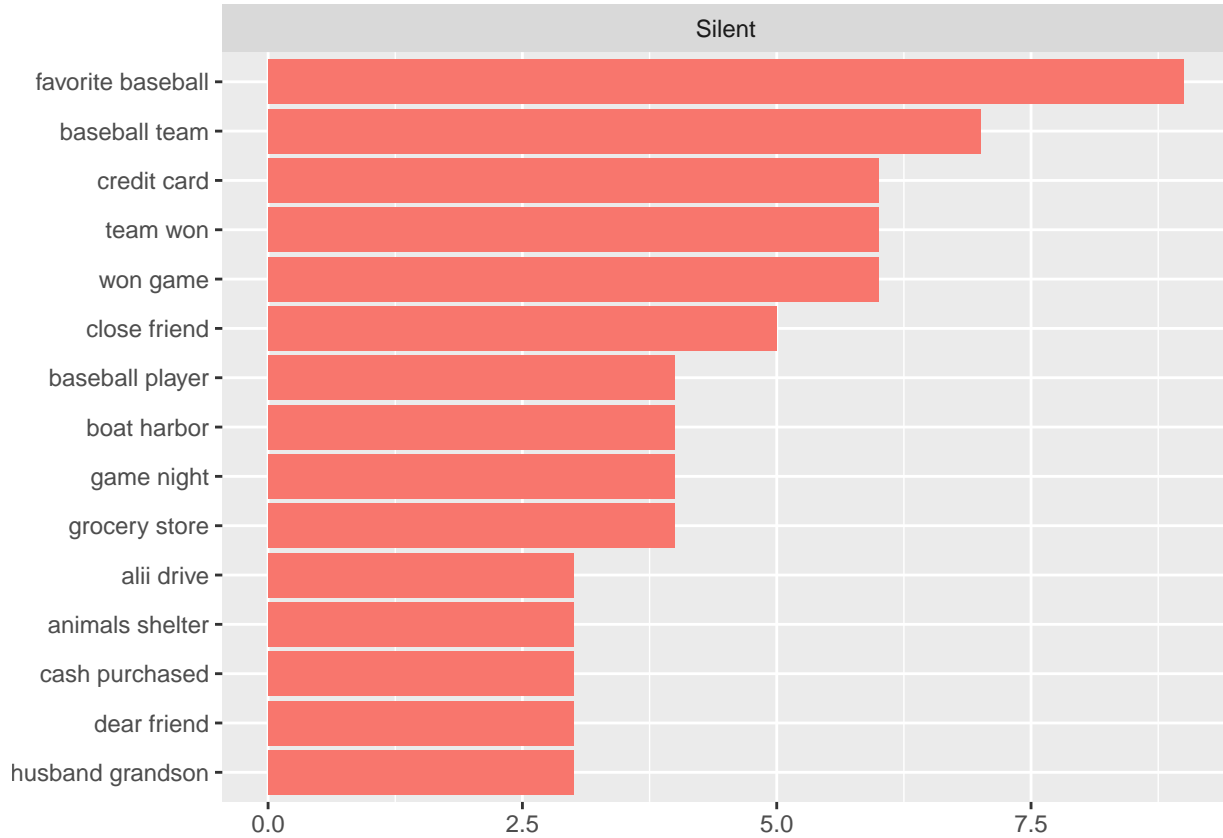


Visualizing a network of bigrams for Baby Boomer group

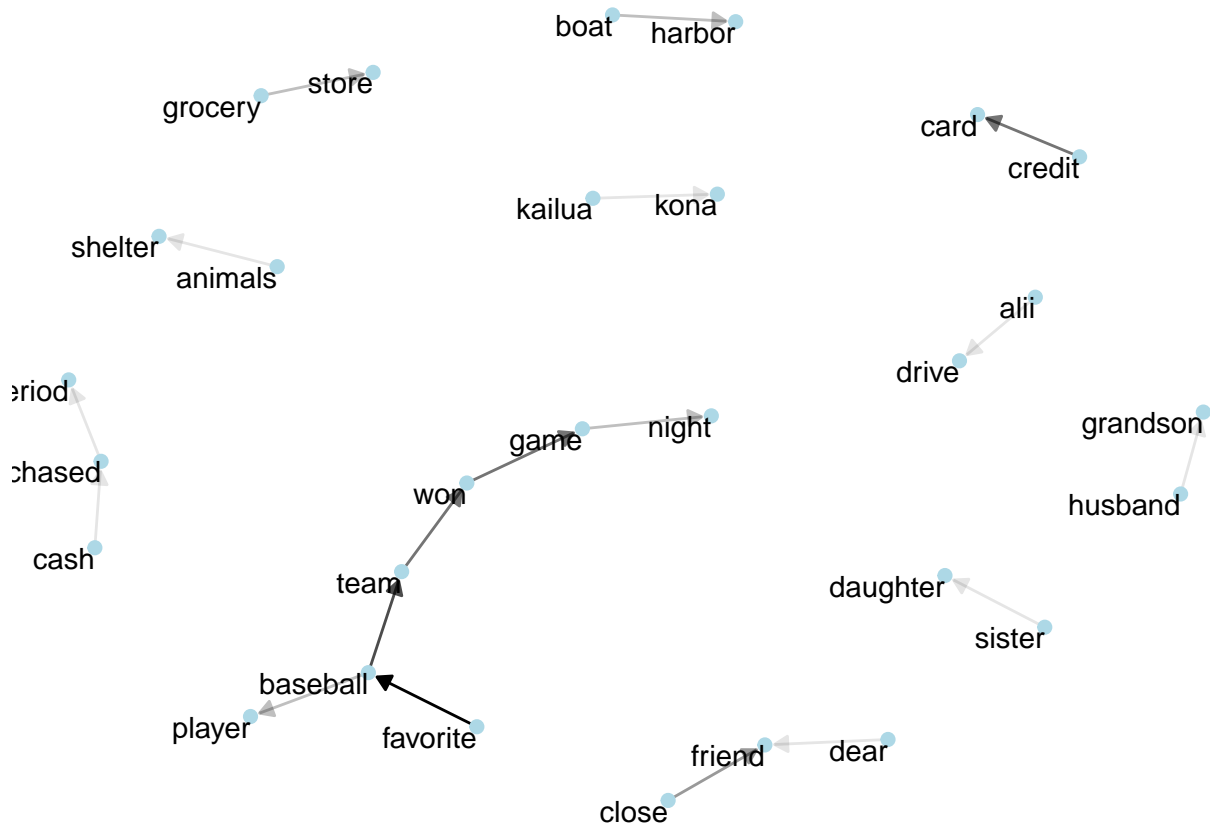


From the Baby Boomers bigram graph, we can see that “spent day”, “spend time”, “Ice cream” are the most common pairs in the Baby boomers group. In network bigrams, we see that key words such as “friend”, “favorite”, “day”, “daughter”, “home” etc. form common centers of nodes. We also see pairs or triplets along the outside that form common short phrases (“ice cream”, “dog walk”).

2.2 (d) Analyzing bigrams for Silent group



Visualizing a network of bigrams for Silent group



From the Silent bigram graph, we can see that “favorite baseball”, “baseball team”, “credit card” are the most common pairs in the Silent group. In network bigrams, we see that key words such as “baseball”, “friend”, “game” form common centers of nodes. We also see pairs or triplets along the outside that form common short phrases (“credit card”).

Above all, Millennials, Gen X, and Baby Boomers have similar common pairs inside their happy moments which has “spent time”, “ice cream” and “mother day” etc. However, Silent’s common pairs of happy moments kind of self-related interesting, such as “baseball team”, “game night” and “credit card” etc. From analyzing bigrams, we concluded that Silent’s happy moments separated it from other three generations.

I found two interesting common pairs inside three groups’ bigram analysis except Silent group.

(a) Why “mother day” is top 5 common pairs in Millennials, Gen X, and Baby Boomers?

Because HappyDB was collected in period 3/28/2017 - 6/16/2017, and 2017’s Mother’s Day at Sunday, May 14. That’s the reason so many happy moments were related to “mother day”. Unfortunately, 2017’s Father’s day at Sunday, June 18, which is just out of the data collecting period. Otherwise, we will probably see some “father day” common pairs related to their happy moments.

(b) Why “ice cream” is top 3 common pairs in Millennials, Gen X, and Baby Boomers?

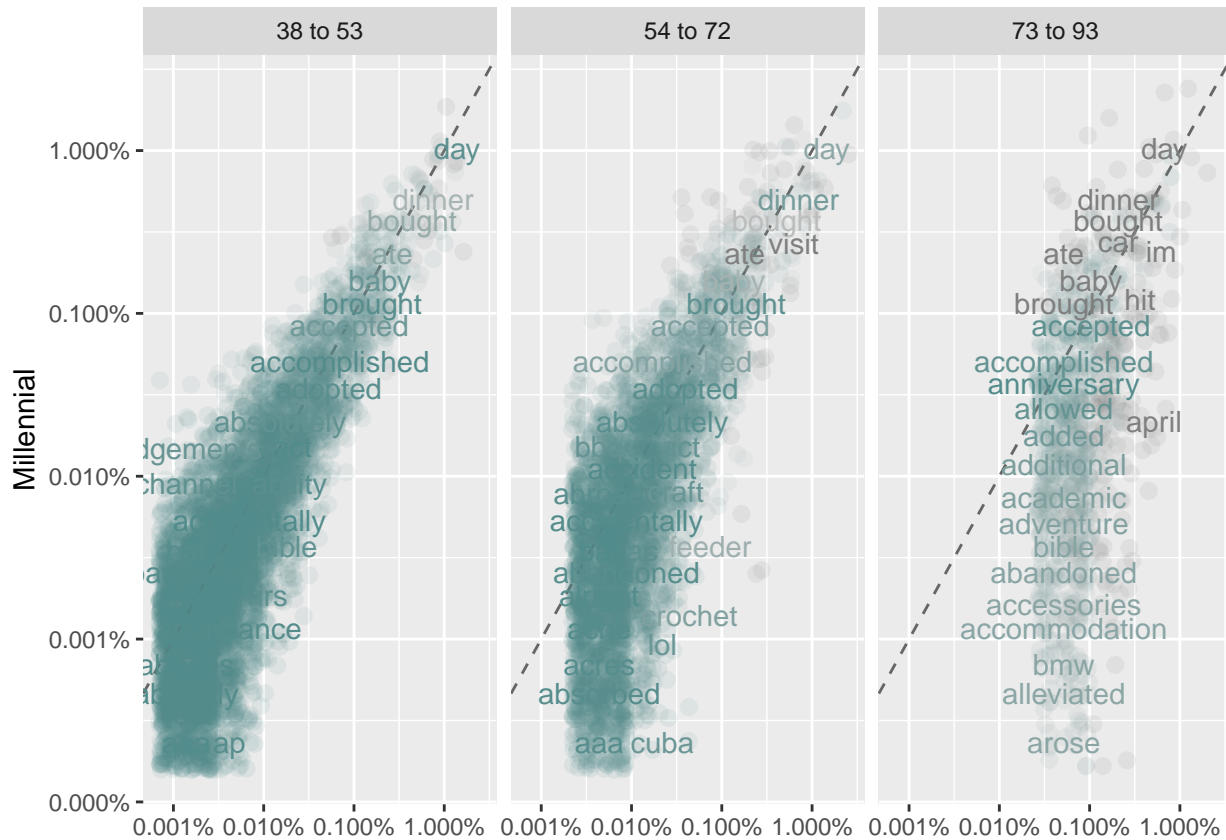
In fact, American people love ice cream. But why eat ice cream became peoples’ happy moment? Some papers have given a deep explanation to this question. “It’s true ... ice cream does make us happy! It was found that the Orbitofrontal cortex (OFC), a part of the brain which plays a role in emotional processing and decision making, is activated when eating ice cream.” (Penn State)

2.3 Compare Word frequencies across different generations

A common task in text mining is to look at word frequencies, then to compare frequencies across different generations. We can do this intuitively and smoothly using tidy data principles. Also we will quantify how

similar and different these sets of word frequencies are using a correlation test

2.3 (a) Comparing the word frequencies between Millennial VS (Gen_X, Baby Boomer, Silent)

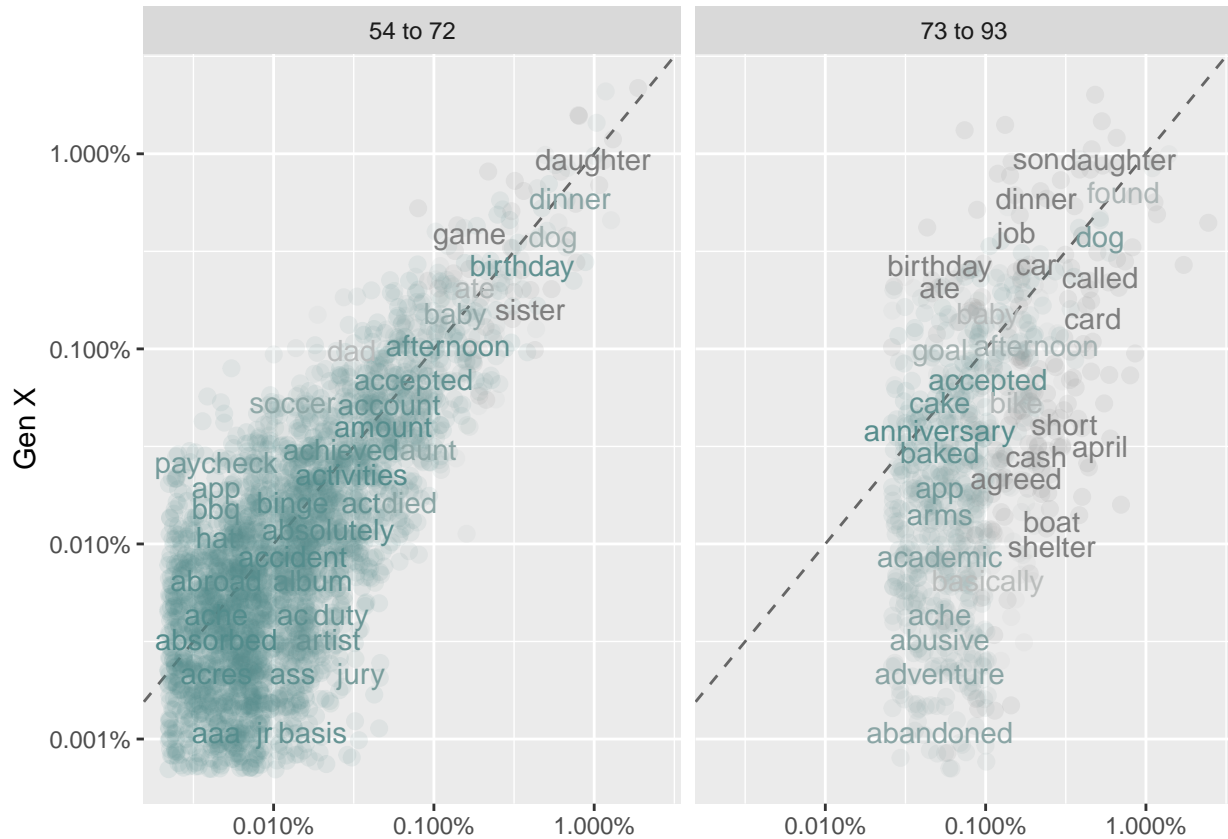


Words that are close to the line in these plots have similar frequencies in both sets of texts. For example, in both Millennial and Gen_X texts (“day”, “dinner”, “bought” at the upper frequency end), in both Millennial and Baby Boomers texts (“dinner”, “day”, “visit” at the high frequency end), in both Millennial and Silent texts (“dinner”, “day”, “baby” at the high frequency end).

Words that are far from the line are words that are found more in one set of texts than another. For example, in the Millennial-Baby Boomers panel, words like “cuba” are found in Baby Boomers’ texts but not much in the Millennial texts. Also words like “arose” and “bmw” are found in the Silent texts but not in the Millennial texts.

Overall, notice that the words in the Millennial-Gen X panel are closer to the zero-slope line than in the Millennial-Baby Boomers panel and the Millennial-Silent panel. Also notice that the words extend to lower frequencies in the Millennial-Gen X panel; there is empty space in the Millennial-Baby Boomers panel and the Millennial-Silent panel at low frequency. These characteristics indicate that Millennial and Gen X use more similar words.

2.3 (b) Comparing the word frequencies between Gen_X VS (Baby Boomer and Silent)



Words that are close to the line in these plots have similar frequencies in both sets of texts. For example, in both Gen_X and Baby Boomers texts (“daughter”, “dinner”, “birthday” at the upper frequency end), in both Gen_X and Silent texts (“daughter”, “dog”, “found” at the high frequency end).

Words that are far from the line are words that are found more in one set of texts than another. For example, in the Gen X-Baby Boomers panel, words like “competition”, “paycheck” are found in Gen_X’ texts but not much in the Baby Boomers’ texts. Also words like “abandoned” and “adventure” are found in the Silent texts but not in the Gen_X texts.

Overall, notice that the words in the Gen X-Baby Boomer panel are closer to the zero-slope line than in the Gen_X-Silent panel. Also Gen X and Baby Boomers use more similar words.

2.3 (c) Comparing the word frequencies between Baby Boomer and Silent

“achievement”, “affection”, “boonding”, “enjoy_the_moment”, “exercise”, “leisure”, “nature” total 7 categories.

Use heatmap to see topic allocation among the 4 generations:

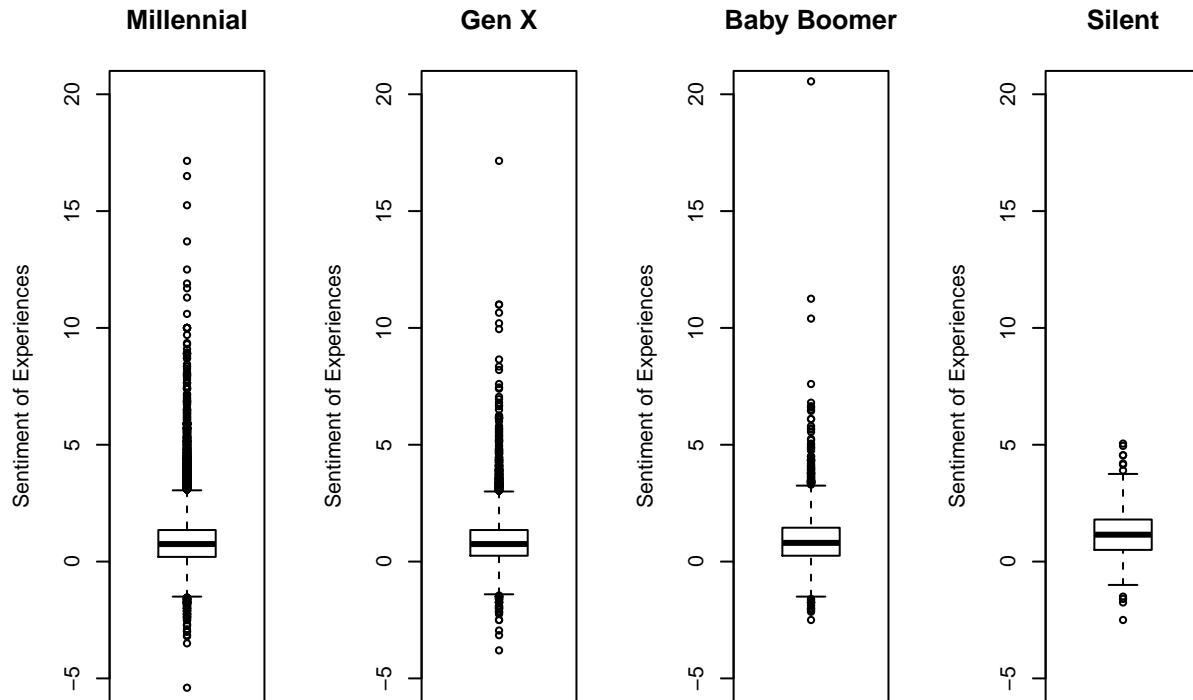


- From the heatmap, we see that the Millennials’ happy moments talk the most about “achievement”, then “affection” and “boonding”.
- Gen X, Baby Boomer, and Silent are talk the most about “affection”, then “achievement” and “enjoy_the_moment”.
- “exercise”, “nature”, “leisure” are three least mentioned categories.
- By the Hierarchical Clustering / Dendrograms: Baby Boomer and Silent are more similar to each other than they are to Gen X or Millennial.

Above all, heatmap and Dendrograms provided some consistent results with our previous analysis. Generations are next to each other have similar happy moments.

Part 3: Sentiment Analysis

Despite knowing every happy moments are relatively positive, sentiment analysis still necessary. Sentiment analysis weighs the emotional intensity of text, also measured the sentiment of these happy experiences to decide how their intensities vary. This created a spectrum of happy experiences that broken down by 4 generation groups.



There's little change in the spread of happy experiences across generations. (a) Surprisingly, happy moments are definitely positive. But the bottom quartile does have negative sentiment. (b) Some happy moments are extremely positive, and few are strikingly negative. (c) There is no significant difference in the range of happy moments analysed by four generations.

Conclusion

By analyzing the happy moments in the four generations, we could get the following results

- Millennial: Use the least words to describe their happy moments. “friend”, “day” and “time” are standouts for Millennials. Their most happy moments were related to their friends. From bigram, “video game”, “spend time”, “Ice cream” are the most common pairs in Millennials’ happy moments. Also, Millennial and Gen X use more similar words.
- Gen X: Use least words to describe their happy moments. More precisely, other than some words are standouts for the Millennial group, “son”, “daughter”, “family”, “wife”, “husband” are standouts for the Gen X group. The Gen X not only put some weights on spend time with friends also focuses more on their loved ones and family. That’s why their most happy moments were related to friends and loved ones. From bigram, “spend time”, “mother day”, “Ice cream” are the most common pairs in Gen X happy moments. Gen X and Baby boomers use the most similar words by all group comparison.
- Baby Boomer: Use slightly more words to describe their happy moments than Millennial and Gen X. The Baby Boomer group shares the similar words with the Gen X group. They have stable social relationships and family members. That’s why their most happy moments were also related to friends and loved ones. “spent day”, “spend time”, “Ice cream” are the most common pairs in the Baby boomers group.
- Silent: Use the most words to describe their happy moments. “received” and “favorite” are standouts for the Silent group. Many happy moments were related to their own interests. “favorite baseball”, “baseball team”, “credit card” are the most common pairs in the Silent group. Silent and Millennial use the least similar words.

Above all, first we can see that four groups used some slightly different high frequency words to describe their happy moments, but most of the words were similar and related to two categories - friends and family. However,

Silent have comparatively different happy moments than others. Silent's common pairs of happy moments kind of self-related interesting. Generations are next to each other have similar happy moments. but as age goes up, generations will diverged from each other and unconsciously use different words or thought to define their happy moments.



Figure 2:

References

HappyBD dataset: Akari Asai, Sara Evensen, Behzad Golshan, Alon Halevy, Vivian Li, Andrei Lopatenko, Daniela Stepanov, Yoshihiko Suhara, Wang-Chiew Tan, Yinzhan Xu, "HappyDB: A Corpus of 100,000 Crowdsourced Happy Moments", LREC '18, May 2018. (to appear)

Text Mining with R <https://www.tidytextmining.com>

image https://dunked.cdn.speedyrails.net/assets/prod/90143/880x0_p194cd5jk41r4vcqq1t3k8d63gf3.png

Penn State <https://sites.psu.edu/siowfa16/2016/12/02/ice-cream-happiness/>

CNN <https://www-m.cnn.com/2013/11/06/us/baby-boomer-generation-fast-facts/index.html?r=https%3A%2F%2Fwww.google.com%2F>

Sentiment Analysis <https://medium.freecodecamp.org/a-data-scientists-guide-to-happiness-findings-from-the-happy-experience>

Moments of Happiness https://www.huffingtonpost.com/susan-pearse/moments-of-happiness_b_7766466.html