Homework 6: Analysis Twitter Data

Yuhao Jia, Zhonghua Yang, Zile Wu

figure 1. Gun Shot

1. Background

Gun violence has always been one of the persistent problems in American society. Under the intertwining of partisan politics, vote politics and money politics, the issue of gun control has not been effectively addressed for a long time. At the same time, gun violence is increasing, and more and more people are becoming victims of gun violence. The seriousness of the gun crime problem is of great concern.

In Philadelphia, for example, shootings accounted for more than 85 percent of the 562 homicides last year. Over the past five years, gun violence has caused the majority of homicides in Philadelphia. Although that percentage has actually declined since 2016, when 90 percent of murders were linked to shootings. But the number of guns in circulation in Philadelphia communities is at an all-time high. These guns join a sea of illegal and legal firearms that can be used in heated moments to turn what would otherwise be an altercation into a homicide.

At the same time, more and more teenagers are even becoming victims of various cases of gun violence. According to the nonprofit Gun Violence Archive, as of 8/18/22, 1,071 children and teens have been killed and 2,864 injured in gun-related violence across the United States so far this year. A growing number of student and parent groups are beginning to speak out against shootings on the Internet.

As an important channel for the expression of public opinion, we want to understand what is being said about shootings on social networks and to understand citizens' concerns and opinions about shootings. We used "**gun shot**" as a search term, using it as a direct reflection of the extent to which Philadelphians are discussing gun crimes.

2. Data Wrangling

Using the "rtweet" package in R, we got 10000 statements about "gun" on Twitter on December 19, 2022 at 8：13pm. We can start by looking at 10 of them：


```
[1] "A 12-year-old in Lakeland, Florida, unintentionally shot himself after finding a loaded gun at his friend's home
during a sleepover.\n\nGov. DeSantis says he plans to pass permitless carry in 2023, stripping safety training
requirements from gun ownership. https://t.co/BucmHE95XO"
[2] "This is Oleh Vasylyshyn. He's 25. Before the big war he had worked in the bookstore. Now serves in the Kyiv
territorial defence unit. 2d ago Oleh shot down a russian missile over Kyiv. W/ the trophy machine gun captured by his
comrades at the east 👍 \nSource @Nerodyk &amp; A.Kovaliov https://t.co/cZemkPU6Qj"
[3] "50 Cent joking about Megan thee Stallion being shot – despite the courts already stating that's an irrefutable fact
– is even more horrendous when you remember he currently has a show about the tragedy of gun violence in the hiphop
community. 🙄 https://t.co/m61Er8Xhvi"
[4] "@D4rkknight47 @Variety If i shot someone Infront of you. You've got me banged to rights. But if I shot someone
around the corner from where you were you can't say I'm guilty if I have no gun or other evidence."
[5] "+in the gun. Granted a lot of the shot missed because he cant see through tears. He turned his attention to the
Rick.)\n\nW...why...are you here I dont...\n\n(He understands other Ricks coming here but most of them wouldn't
help or intervene.)\n\nI dont understand why you-\n\n(He+"
[6] "@panicBoy @10KBC01 @Christa54196704 @irishnatl @chrissywaxes @ThisIsKyleR One guy shot 4 times while grabbing his
gun barrel. Another struck him with a skateboard before taking one to the chest. Another pointed a gun at him and got
shot in the arm.  That is very controlled and deliberate shooting."
[7] "@chrisWE95278948 @br4dm4l0n3y @TomFitton You see the gun you see the woman get shot, if you can't put that together
you might want to get your blood lead levels checked."
[8] "@nytimes lol guns are not the number one killer of children in the US. Your little graph is a JOKE. Pfyzer is
single handedly killing more children with their kill shot than any gun violence. Gun violence is only a problem when
they are taken away from law abiding citizens."
[9] "millie said she believe she's been shot, it's just a mystery as to who shot her. gun shot residue was on several
people, he maintained his innocence and, the prime witnesses have been unreliable, so why are we obligated to believe one
person's count of events when nothing adds up"
[10] "They said only a female dna was found on the gun that shot Megan … I think I figured it out . T lanez must be
trans…."
```

Figure 2. Import Raw Data

Of the ten tweets, all are related to "gun shots" except for the second one, which describes a veteran's World War II encounter and is not related to the gun violence we are looking for.

First, we need to convert all the tweets to text content. However, there are many special symbols and words in these text contents that we do not need, and these things will hinder our analysis of the text contents. Therefore, we need to process this data. For example, we found many emojis in the text. These emojis are not the focus of our attention, so we remove all of them in this step.

After this, we can transform the data frame into a corpus. A corpus (plural of corpus) is a collection of files containing (natural language) texts. They are used in packages that use the infrastructure provided by the text mining package `tm`. Specifically, each tweet is saved as a separate file. Now, let's see what data is in the corpus that needs to be removed. In the corpus, the following data are cleared (or rewritten) by us.

1. Turns some special symbols (@, /,], $) into spaces: these symbols do not convey a specific meaning and are commonly found in tweets, so they are removed.
2. Remove non-ASCII characters from the tweets: make sure we filter out tweets that contain only English text and numbers, and avoid data in other languages.
3. Remove numbers and look at the first entry of the corpus again: further remove the remaining numbers from the previous filter. Numbers are widely present in shooting-related tweets, but are not the focus of our attention.
4. Remove punctuation: the same as 1 to remove special symbols, which also do not convey meaning.

Finally, we processed the filtered words themselves, i.e., we further streamlined our corpus

by removing English stop words (e.g., a, to). In practice, we added some stop words: "by", "the", "http", "may", "ttp", "qhb...", "https", "http...", "for", "tco", "until", "while" are common conjunctions and beginnings in sentences. On top of this, we also remove whitespace characters, which do not correspond to a visible mark, but typically do occupy an area on a page. Finally, we also extract word stems by removing common word suffixes and endings to streamline our corpus.

A preview of the final corpus we obtained is as follows.

```
[1] "A yearold Lakeland Florida unintent shot find load gun friend home
sleepov Gov DeSanti say plan pass permitless carri strip safeti train"
[2] "requir gun ownership BucmHEXO"
```

Figuure 3. Processed data

3. Word Cloud

Figure 4. Word Cloud

Using the "wordcloud" package in r, we can make a word cloud of the above corpus. In addition to the most obvious "gun" and "shot", two personal names "kelsey" and "tori" appear. In a google search we can learn that these two names are related to Megan Thee Stallion and Tory Lanez Shooting Trial. The shooting case, which took place two years ago, recently went to trial and gained a lot of attention from citizens for a while because it happened between entertainers. Secondly, personal pronouns such as girl and man appear in the word cloud, which can reflect to some extent the group of people concerned in the shooting case more generally. Finally, adjectives such as "afraid" and "inconclus" appear in the word cloud. This reflects citizens' attitudes of concern about the potential problems of shooting time.