

Narrative

Hanning Luo

2020/12/11

Brief background

Is China's diplomatic stance experiencing a radical turn? During the Trump administration, China-US relations have rapidly deteriorated. Most recently, the US declares to sanction the Vice Chairpersons of the Standing Committee of the National People's Congress. Several issues have contributed to the current situation. First, globalization and the rising competitiveness of China contribute to the inequality in America. President Trump believes that he needs to resist the impacts of China to maintain his social base of support. Meanwhile, he considers the rise of Chinese high-tech industries as a major threat. Second, the ideological conflicts between China and the US have been increasingly pronounced, as China pursues a higher political visibility to "export" its unique way of governance while the US retreats from many international agendas. Third, the conventional issues of human right violations in China still exist, and in recent years they have intensified in regions such as Xinjiang and Hong Kong. From the perspective of Beijing, these issues involve the integrity of the Chinese state's sovereignty, and they must not be intervened or challenged in any sense.

Under this situation, the spokespersons of China's Ministry of Foreign Affairs are increasingly bold and aggressive in their choice of languages on public occasions, including the Q&A sessions of their regular press conferences. The most typical case is the recent quarrel with Australia triggered by one of the spokespersons Zhao Lijian's controversial tweet. These diplomats are often called "Wolf Warriors", referring to a famous populist and nationalist movie in China.

Then, we need to ask: Are the discourses of Chinese diplomacy becoming increasingly aggressive? Are their comments toward western countries and political figures generally negative? What events are the most sensitive and infuriating for China? Are there any variations in sentiments of discourses between different spokespersons due to variations in their personalities? Figuring out these issue may be helpful for us to understand whether there are any significant changes in Chinese diplomatic policies.

Data scraping and text parsing

The first step is to collect the scripts of press conferences from the website of the Ministry of Foreign Affairs, and also parsing the texts to get a data frame of four variables: the date of the conference, the name of the spokesperson, the journalist's questions, and the spokesperson's answers. Web scraping was the most challenging task for me in this project. The reason is that the site was developed in a strange format, and simply using CSS Selectors is not sufficient to get the complete list of questions or answers. Meanwhile, some answers cross several paragraphs, so I have to identify them and concatenate them together. Therefore, I used the following technique: (1) Assign observations to be "questions" or "answers" according to whether there is or in the paragraph's html nodes. For answers, assign NA if there is a sign of bold. (2) Generate an empty vector as the vessel for concatenating the "answer" strings. (3) Write a loop to identify the answers between NAs. Add "answer" strings to the vessel, and concatenate eligible following "answer" strings to the last element of the vessel vector. Below is a simple illustration of how this loop works:

```
library(stringr)

test <- c("Hanning", NA, "Computational", "Tools", NA, "Social", "Science",
         NA, "Regular", "press", "conference")
vessel <- vector()

for (i in test) {
  if (is.na(i) == 1) {
    vessel <- append(vessel, "")
  } else {if (length(tail(vessel, 1)) > 0)} {
    vessel[length(vessel)] <- paste(tail(vessel, 1), i, collapse = "")
  } else {vessel[length(vessel)] <- i}}

str_trim(vessel)

## [1] "Computational Tools"      "Social Science"
## [3] "Regular press conference"
```

Using this method, I successfully scraped and parsed the texts of regular press conferences since 2018-11-07 to 2020-11-08. The final result of my data frame looks like the following:

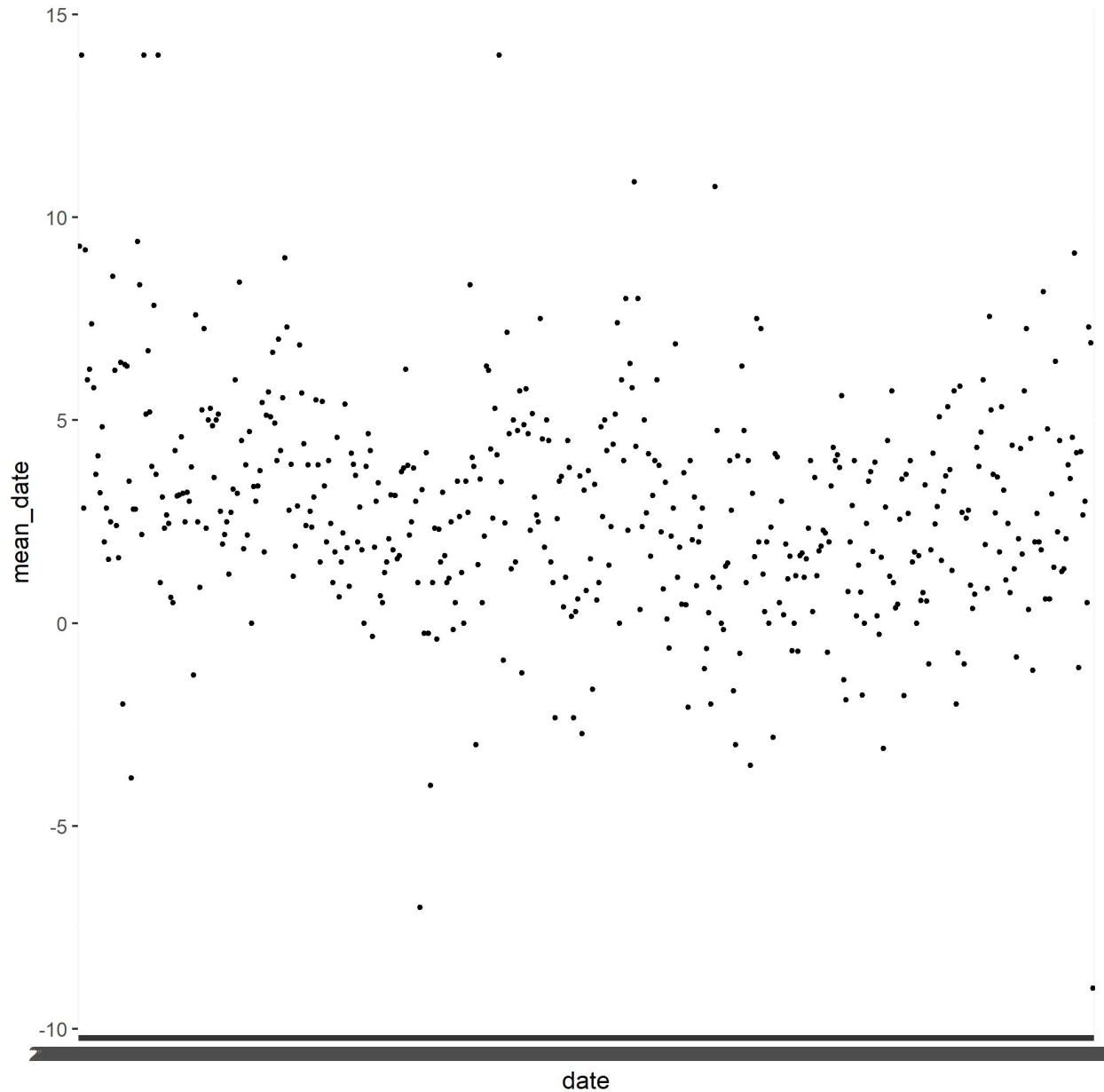
```
df <- read.csv("Data/all_texts_EN.csv", stringsAsFactors = F)
#Remove the name of the spokespersons in "answers" for a cleaner preview
df$answers <- gsub("^[:]+: ", "", df$answers)
head(df)
```

```
##      date      speaker
## 1 2020-12-08 Hua Chunying
## 2 2020-12-08 Hua Chunying
## 3 2020-12-08 Hua Chunying
## 4 2020-12-08 Hua Chunying
## 5 2020-12-08 Hua Chunying
## 6 2020-12-08 Hua Chunying
##
## 1
## 2 Bloomberg: First of all, I'd like you to clarify, if possible, on what sort of countermeasures will
## 3
## 4
## 5
## 6
##
## 1 Tibet affairs are purely internal affairs of China and no external forces have the right to interfere
## 2                               I have just made clear China's position on the egregious
## 3
## 4
## 5
## 6
```

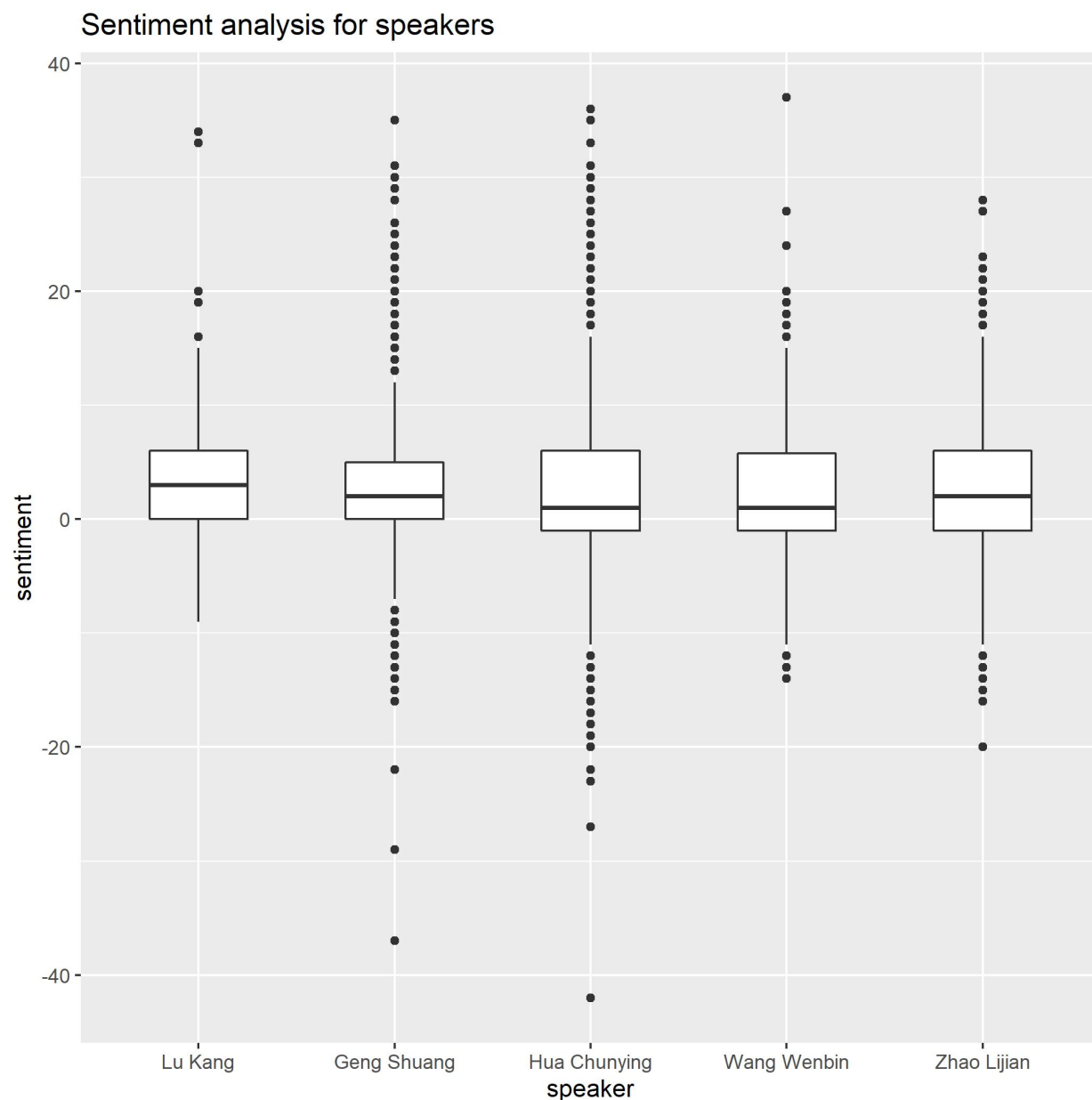
Analysis and visualization of some interesting findings

My first method was to create a dictionary of aggressive attitudes, including the more conventional diplomatic parlance such as “firmly oppose”, “strongly condemn”, and “taking necessary countermeasures”. However,

I found that this method is very inefficient and imprecise. The translations are inconsistent, while the spokespersons (and their translators) today have become increasingly creative in their use of adjectives, so I simply cannot cover all of them. Sentiment analysis follows a similar logic of identifying the attitudes behind the words, and I found that it can almost always reach a similar conclusion as the method of searching for keywords. Therefore, I use the lexicon “bing” to score the sentiments of every answer in my data set (4825 in total).



The general tendency of sentiments over time can hardly be observed from the scatter plot. The variations between speakers are also trivial at the first glance (all the box plots below follow a descending sequence according to the mean values).



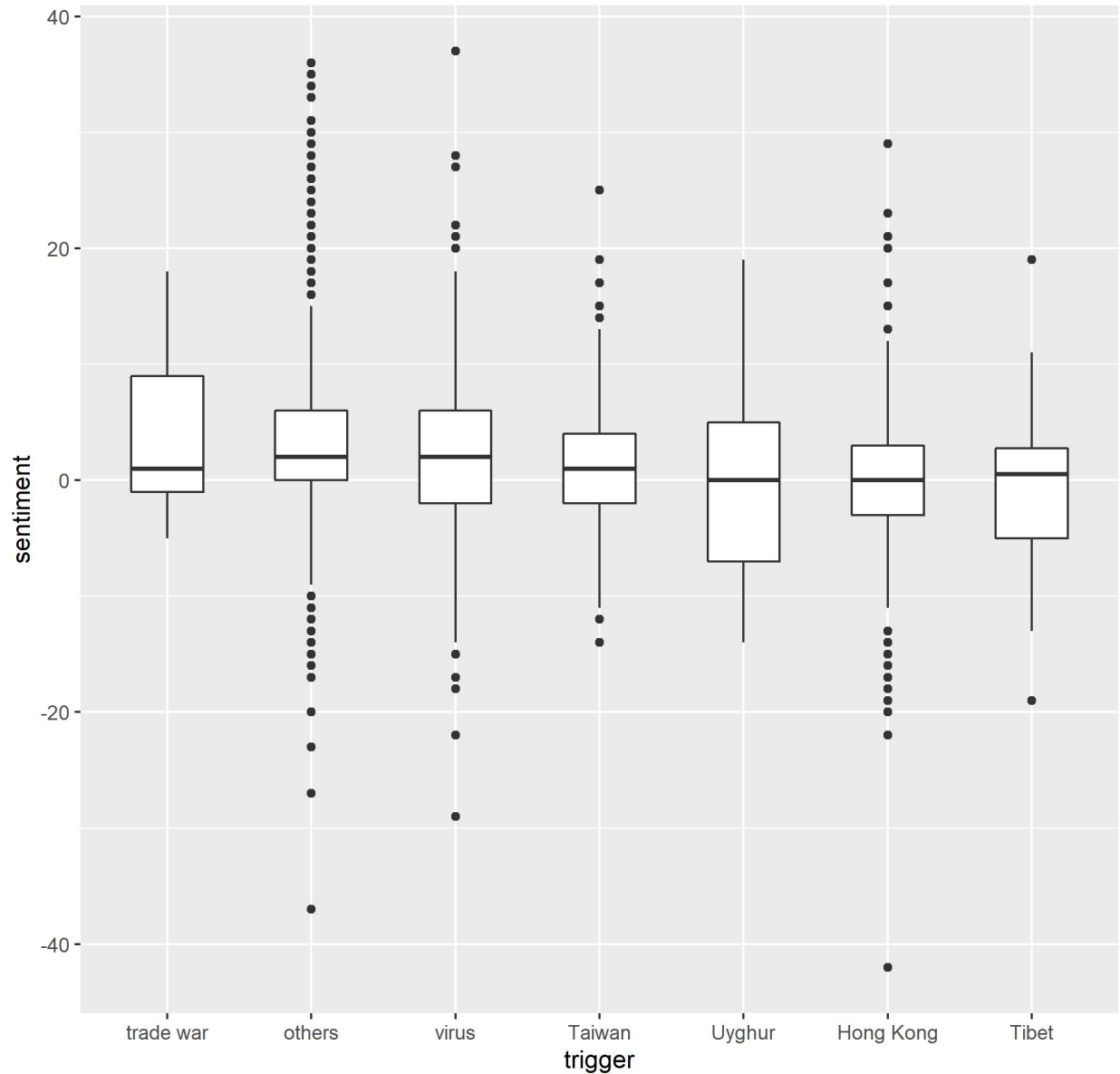
However, I conduct a simple t test to compare the average sentiment scores of different spokesperson, and it returns interesting results. There are five speakers in total within this two-year period, but two of them, Lu Kang and Geng Shuang, who are more senior and experienced, have left the post and transferred to other offices. The t test shows that compared to the most negative speaker Zhao Lijian, the two senior officials are significantly more positive in discourses, while the other two incumbents do not exhibit any significant differences. This probably shows that the ministry prefers the spokespersons to be less “cautious” and more aggressive in their personalities and discourses.

```
read.csv("Results/ttest speakers.csv", stringsAsFactors = F)
```

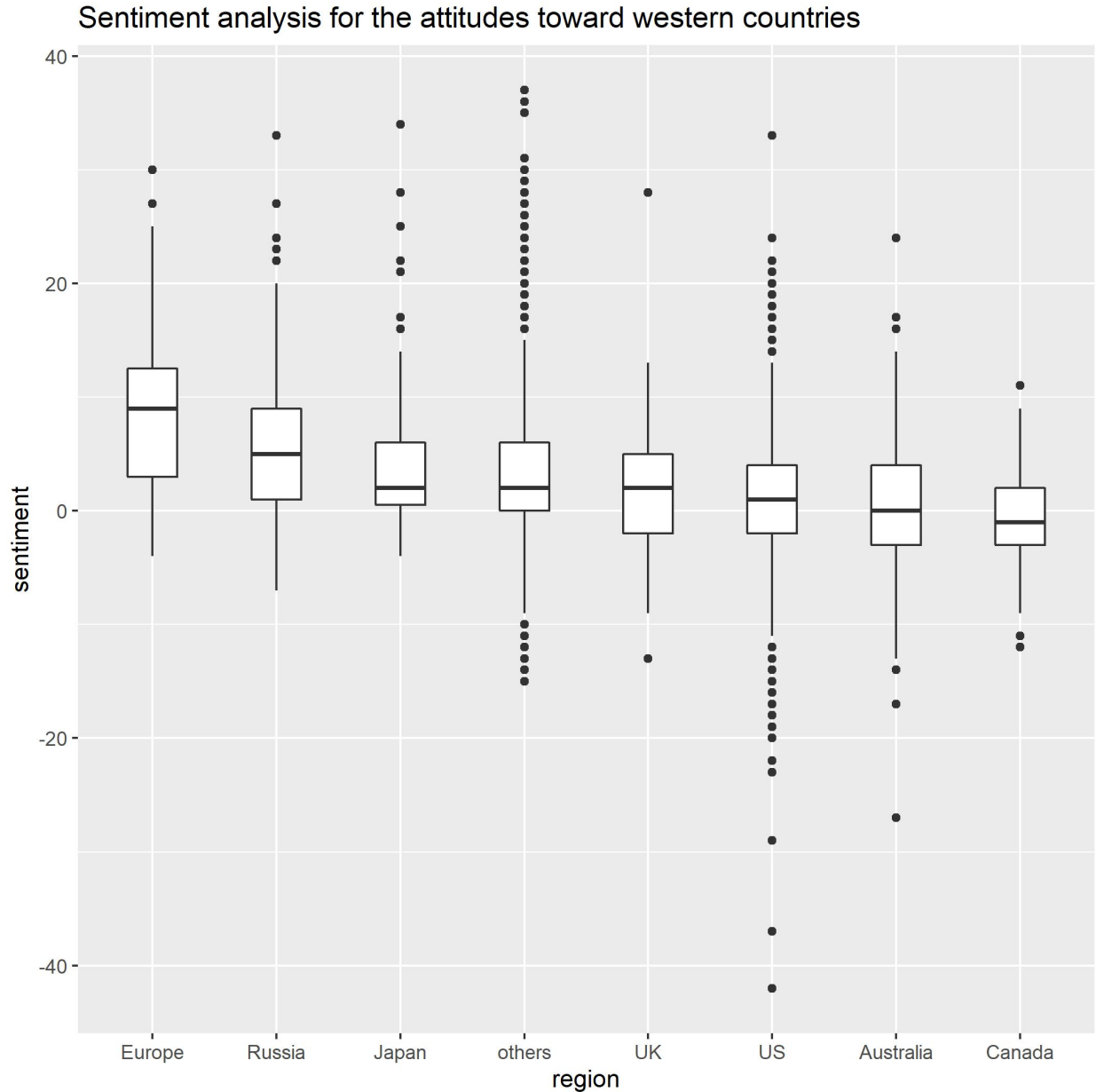
##	X	.y.	group1	group2	p	p.signif
## 1	1	sentiment	Lu Kang	Zhao Lijian	1.310146e-06	****
## 2	2	sentiment	Geng Shuang	Zhao Lijian	1.119873e-03	**
## 3	3	sentiment	Hua Chunying	Zhao Lijian	2.032595e-01	ns
## 4	4	sentiment	Wang Wenbin	Zhao Lijian	3.201573e-01	ns

Another issue that I want to study is the triggers of responses with strong negative sentiments. First, what events are considered the most serious? The following box plot shows that compared to socioeconomic issues such as the trade war and the pandemic, questions about sovereignty (from China's perspective) and human rights (from America's perspective) are more likely to get a furious answer.

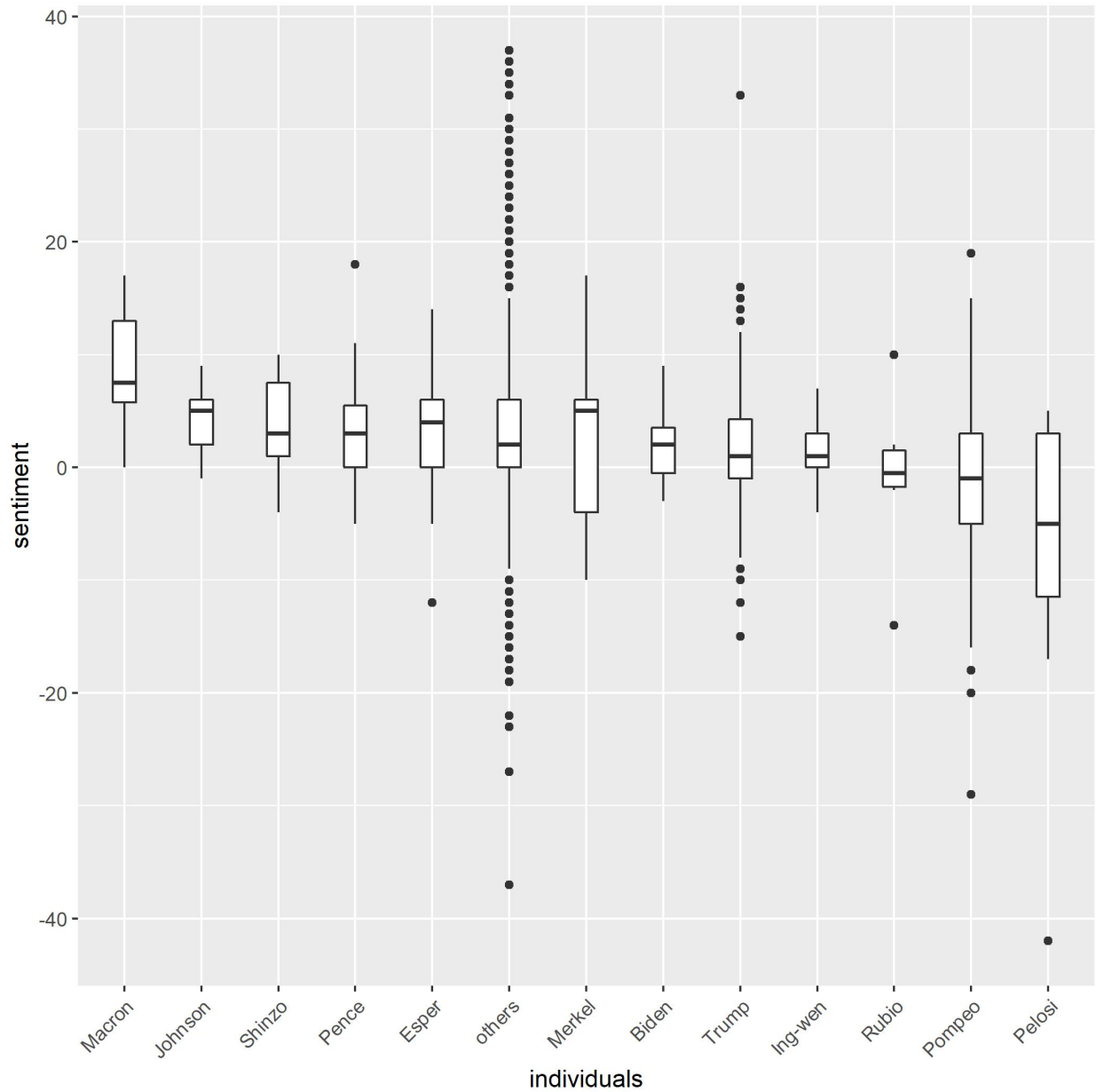
Sentiment analysis for sensitive political issues



The next graph show the average sentiments when the spokespersons refer to certain countries or regions. Surprisingly, the US ranks only the 3rd most hostile, while Canada and Australia trigger more negative sentiments in average. A possible explanation, aside from the relevant specific issues, is that the Chinese diplomats may consider the relations with these two countries as less important strategically.



Finally, I want to visualize the sentiments when speakers are responding to certain political figures (for two cases I use first names to better identify the figure). It turns out unexpected that the most infuriating politician is not Trump or the political leader of Taiwan, Ing-wen Tsai, but Nancy Pelosi. One probable reason is that again, the Chinese diplomats may consider political leaders as more important figures and choose to use more cautious language in their responses. On the other hand, as a Democrat (and also due to her personal experiences), Pelosi is highly concerned with human right issues in China and thus thus frequently involves the most dangerous trigger. Pompeo ranks the 2nd due to his similar attacks against China on issues of ideology and sovereignty. Another interesting finding according to these two graphs is that China is showing more positive attitudes toward its relations with Europe and Japan within the two-year period.



Research implications

I do not major in IR, and this research is totally out of personal interests. Though not familiar with the relevant theories of diplomacy, I would still like to conclude one implication from my findings. The Chinese diplomatic discourses may not have become significantly more aggressive, and when dealing with strategically more important relations the diplomats could be more cautious. However, the changing spokespersons and their personalities indicate the China has the full potentials to make a radical nationalist turn. If the international society, especially the US, continues to press and challenge China in sensitive issues regarding human rights (or sovereignty) and ideology, this process of radicalization may be accelerated.