Narrative

Jichun Zhang

2020/12/10

Brief substantive background/goal

Cross-strait relations will be one of the dominant factors shaping East Asian politics in the forthcoming decades. As the Chinese People's Liberation Army appears to be more and more aggressive in its neighboring waters and air, a concurrent reflection of "wolf-warrior diplomacy" and assertiveness. To understand how the Chinese government perceives cross-strait relations and how the balance of power shifts between the two is of paramount importance. This project attempts to depict this relational dynamics by incorporating both subjective and objective dimensions.

The first part of this project will present how the Taiwan Affairs Office, a party institution, describes the cross-strait relations by examining how often the Office uses "Taiwan Independence Advocates" (Taidu) and "Taiwan Compatriot" (Taibao) in its news headlines respectively. The assumption is that changes in news frequency reflect changes in attitudes of the party in dealing with the island. To do this, I will web scrapping all related news from the website.

I will then turn to introduce objective factors by comparing states' military expenditure. To make the results more informative, I plan to include four East Asian entities, namely, China, Taiwan, South Korea and Japan and a global superpower: The United States. States' military expenditure, in this project, will be presented as a share of Gross Domestic Product (GDP) to make the results comparable and clear.

Collecting data

To measure the Office's attitude towards Taiwan, I used the `rvest` package to scrap all related news links and texts from the website. In fact, I used the package twice, one for "Taiwan Independence Advocates" and one for "Taiwan Compatriots." I first used the built-in search function of the website to filter all relevant information and then created a function to scrap all URLs from all pages. Unfortunately, Selector Gadget failed to work in all cases. When it failed to gather wanted information and returned NA values, I inspected the base code of the website and used `html_nodes()` to get the information needed. For each subject, I collected information on news title, news date, and news link. Also, I label the status of the title in question. ("Independence Advocates" vs. "Compatriots") Finally, I stored the information into separated csv files.

To compare the military expenditure of selected stakeholders, I used the excel spreadsheet provided by the Stockholm International Peace Research Institute. I used the readxl package to load the data frame.

Cleaning/ pre-processing data

To collect the date information, I first subset all URLs by using the `stringr` package and then used `lubridate`package for the sake of converting it to a year-month-date format. After that , I used the full_join function to create a large dataset. The next step was to count the daily total frequency and daily status frequency. After all, it was how often the Taiwan Affairs Office used those keywords that matters. To count their daily frequency gave me a reasonable amount of data points to make an informative graph. I then used the `distinct()` function to filter all repeated news links. The ultimate data frame contains 934 news titles.

For the excel spreadsheet, I first tried to use functions, such as select() and filter() in the `tidyverse` package to perform all initial data cleansing tasks but failed. I then turned to base R functions to subset redundant columns and assign a new column as the name column for rows. Also, the original dataset was in wide formate, so I used the gather() function to turn the data frame into long format. With a new `name` column for rows and appropriate format of the data frame, I was able to identify states I was interested in, namely China, South Korea, Japan, and the United States. I picked all relevant information out for future plotting.

Analysis and visualization

For the first part of the project, I first visualized the results and highlight the data points representing daily news frequency>1, and then created a logged version to highlight its pattern of using those two keywords. Finally, I arranged the four plots into the same graph for comparison.

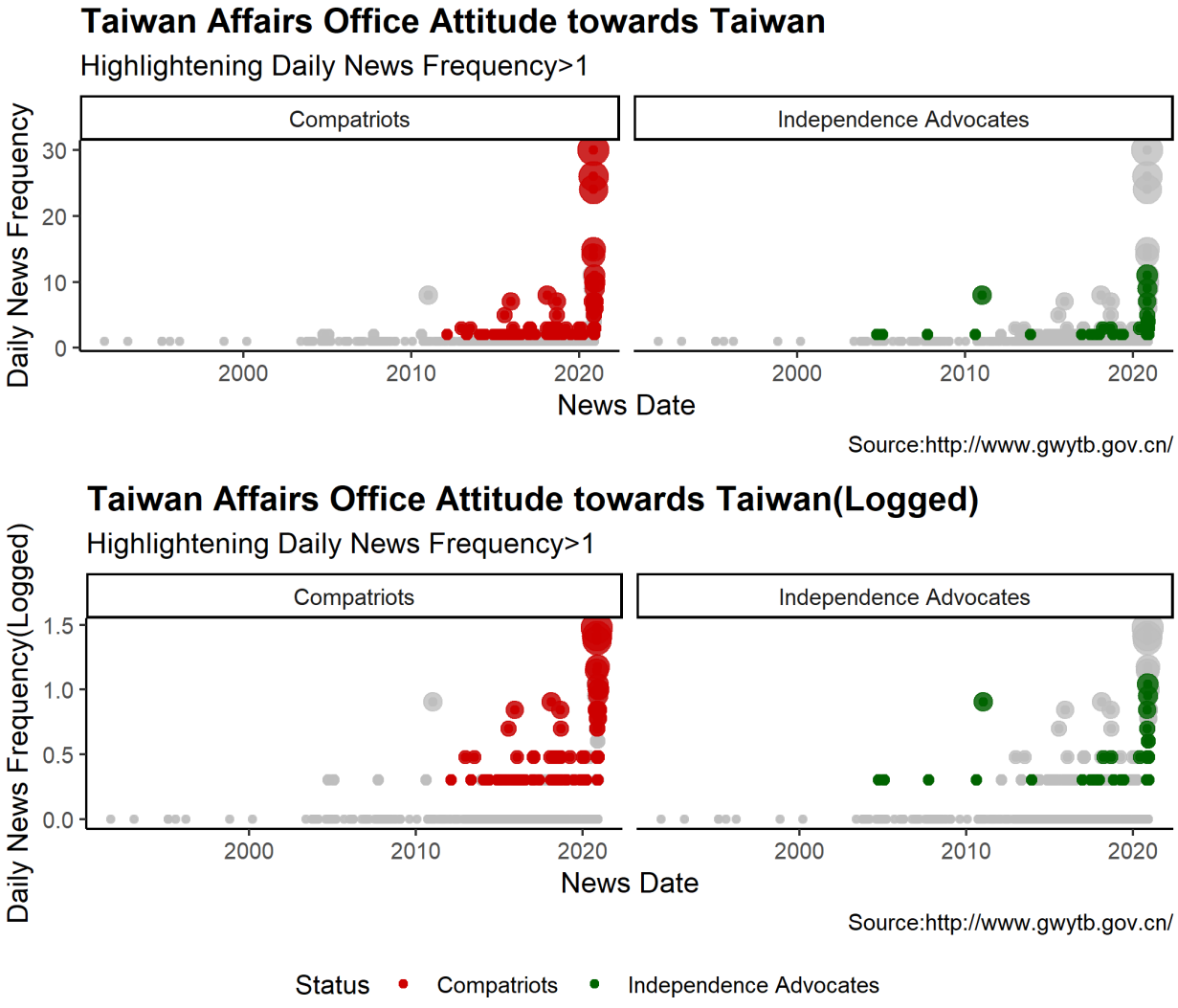
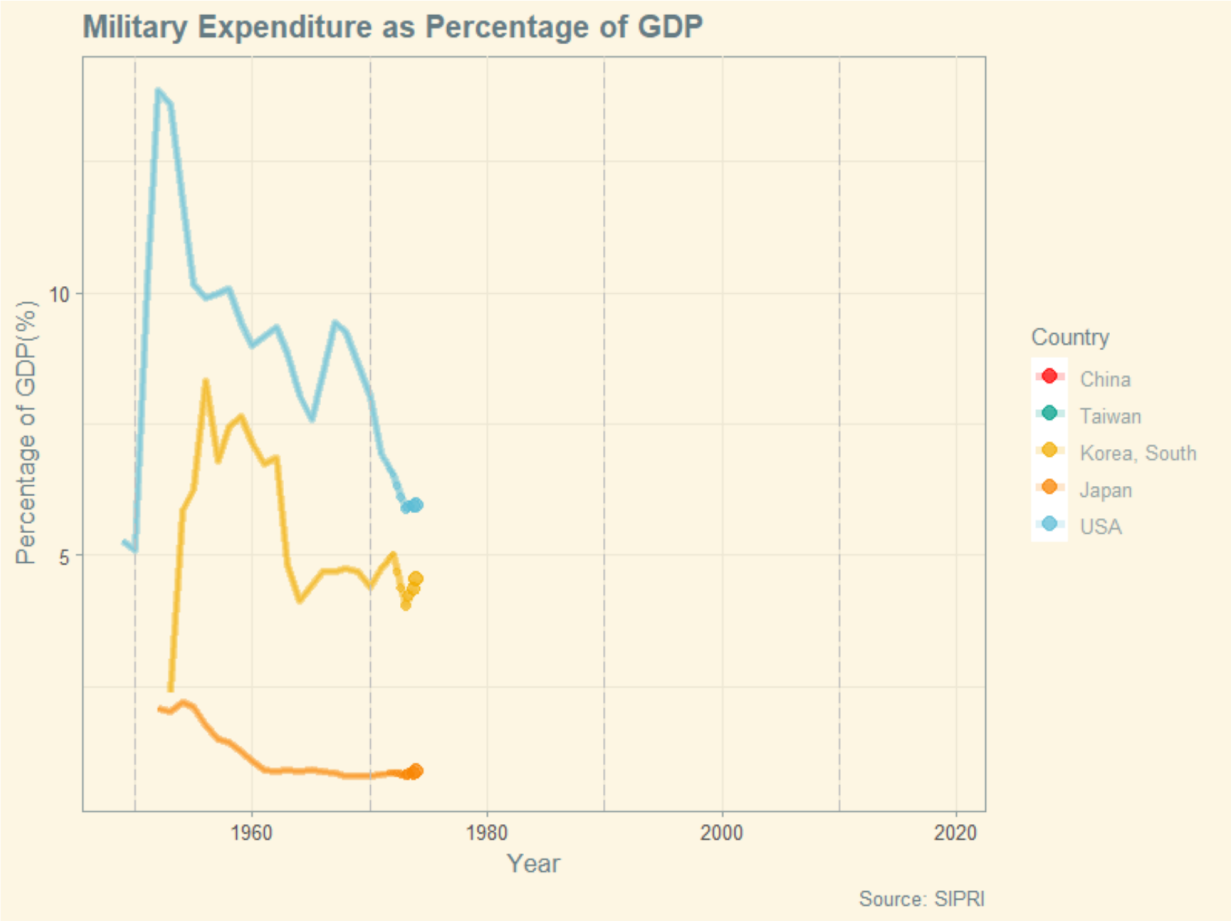


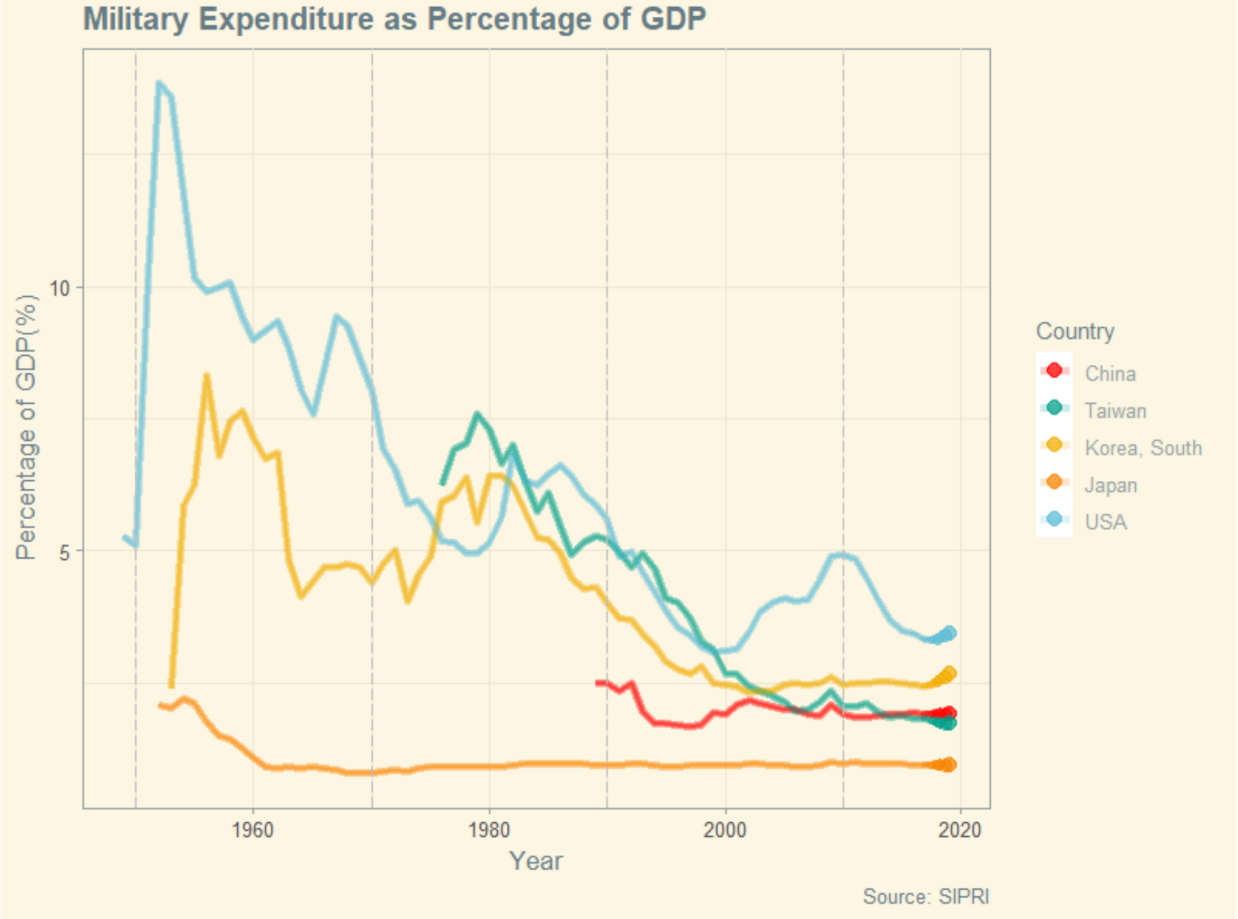
Figure 1: Taiwan Affairs Office Attitude towards Taiwan

The outcome reveals two things: first, the fact that this party institution had mentioned "compatriots" far more often than "independence advocates" may imply that the party leadership still considered the Taiwanese as having a shared Chinese identity and hence preferred peaceful integration over military occupation. Second, the office gradually increased its mentioning of "independence advocates" in recent years. This may

imply that China is becoming more aggressive, but still prefer a gradual and limited way of escalating the conflict.

To present how military expenditure changes, I first used the ggplot2 to create a static graph and then used gganimate to make an animation. I changed during, end_pause, and add several vertical lines to improve this visualization. This graph shows that states, expect Japan for its constitutional restraints, gradually decreased their military expenditures. The United States spent 3.4 percent of its GDP on military expenditure, higher than any other states in the graph. In comparison, The Chinese government had spent around 2 percent of its GDP since 2000. However, although both Taiwan and China spent roughly 2 percent in 2019, China apparently spent much more than the island because the size of the two economies is significantly different.Since gif. files cannot be included in a PDF file, I listed two static graphs below; I also used the 'wesanderson' package to improve its appearance:





Future Work

The next step of this research is first to include other potentially relevant dichotomies, so that I may have more data points at my disposal. The limitation of focusing on one single source for information gathering is obvious. It can depict a general trend, but its explanatory and persuasive power at the micro-level is questionable. For instance, from the first paragraph, we see that the term"Independence Advocates" had been mentioned more than five times in 2011. This was a result of information update rather than instant diplomatic responses to external "stimulus." The office simply pasted comments made years ago on its website. This outlier would be highly misleading if we rely on this graph alone without understanding the underlying logic of the comments made by the Office. In order words, the graph itself shows no causal relations among those variables.

There are two ways to make improvements and enhance its explanatory power. First, future studies should diversity the sources they collect information from, so that outliers can be identified and additional scrutiny applied. For instance, they can use the same methods to collect data from state media and other censored content. Second, this graph can be used as a part of mixed methods research. A causal mechanism, for example, can be supplied by information collected from interviews and field studies.

Future research can also run a statistic model to make predictions and specify the factors leading to the change of attitudes. In this scenario, what made the Chinese government take a more proactive approach in dealing with the island is worth investigation.