EPPS6356 Data Visualization
Members: Kiran Jaura (coordinator), Marcus Sianan, Glen Cooper, Wen Si
Final Project Report

Abstract

This paper attempts to address the research question "How do relations between the ASEAN countries and China change over time?" To answer this question, we utilize the following data visualization methods while analyzing event data: Network igraphs, a geographic map, the R Shiny app, and ggplot2. We find that during the time period of interest (2012 to 2020), our data source (ICEWS) provides 1,517 events, which mainly capture interactions between China and Cambodia. We also observe that ICEWS shows variation in the number of events between 2012-2020 and that those events are mostly cooperation-based (rather than "conflict-based" or "neutral"). Our visualization tool provides the ability to identify trends, outliers, and breakdown of events through both a timeline of all years found in the data set as well as by event type, while simultaneously providing researchers the ability to investigate the data for themselves. This kind of visualization tool can be built to generalize to other geographical locations and time periods as suited to a researcher's needs. The current tool can be accessed here: https://t5904j-kjaura1.shinyapps.io/Storyboard/_w_529b15f0/Story.Rmd.

Introduction

ASEAN stands for the Association of Southeast Asian Nations and was established on August 8, 1967. ASEAN now has ten member states: Brunei, Cambodia, Indonesia, Laos, Malaysia, Myanmar, the Philippines, Singapore, Thailand, and Vietnam.

There are mixed relations between ASEAN and China. On the one hand, there have been territorial disputes in the South China Sea between China and ASEAN countries since the first armed clash in 1974 (International Crisis Group, 2021). Starting in 2012, confrontations between China and other claimant states in Southeast Asia have intensified through flotillas of coast guard, paramilitary and naval vessels (Huang, 2021). In 2018, tensions between China and both the Philippines and Vietnam cooled down even as China increased its military activity in the South China Sea by conducting naval exercises. Meanwhile, China continues to construct military and industrial outposts on artificial islands it has built in dispute waters (Center for Preventive Action, 2022).

On the other hand, based on ASEAN data, China has retained its place as ASEAN's largest trading partner since 2009. FDI flows from China to ASEAN reached USD 9.1 billion in 2019, accounting for 5.7% of total FDI flows to the region, which placed China as the fourth largest source of FDI among ASEAN's Dialogue Partners. The trade volume between ASEAN and China during the first half of 2020 increased by 2.2%, when ASEAN leapt to become China's largest trading partner for the first time (asean.org/). "Through promoting partnerships in traditional and non-traditional security, in law enforcement, economic, trade and infrastructure development, China is fashioning a new architecture of relations within the Southeast Asian region" (Chatterji, 2021, Conclusion, para. 1).

This project attempts to understand the relations between the ASEAN countries and China over time by using data visualization tools to visualize Integrated Crisis Early Warning System (ICEWS) event data.

The rest of the paper is organized as follows: A brief review of the visualization of event data in the academic literature, a discussion of igraphs produced via the igraph R package, a presentation of the R Shiny app, and an analysis of the data.

Literature

A number of scholars produce visualizations from event data. More recently, Brandt et al. (2022) and D'Orazio and Lin (2022) rely heavily on the visualization of event data to demonstrate predicted violence in Africa (the dependent variable is the natural logarithm of state-based fatalities). Specifically, the authors make forecasts at the grid-month level. The authors visualize their results via geographic heatmaps in which each data point placed on the map represents an observation of the dependent variable. It thus enables the viewer to easily see where predicted violence occurs in the African continent. In addition to heatmaps, Brandt et al. (2022) use bar graphs to visualize changes in CAMEO root codes (Gerner, Schrodt, and Yilmaz 2009) for South Sudan and Sudan over time by relying on the ICEWS event data. The authors also implement line plots to compare the results of various models, and they use another set of line plots to visualize actual and predicted state-based fatalities in the Lake Chad region and Mozambique-Tanzania border area.

Data and Methods

For event data, this paper relies solely on ICEWS (Boschee et al. 2015). The ICEWS data were pulled by using the UTDEventData R package (Kim et al., 2019). The UTDEventData R

package allows researchers to pull event data from several different data tables. For this paper, the ICEWS data are chosen because they meet the needs of this project, particularly that they cover the countries of interest (the ASEAN countries and China) over the time period of interest (2012 to 2020 is sought, but the data only address February 2012 to August 2020). After filtering the source and target countries to where it is only the ASEAN countries and China interacting with each other (and not the ASEAN countries interacting with each other), there are 1,517 total events included in our analysis.

ICEWS automatically assigns a CAMEO code to each action that takes place when a source actor engages with a target actor (the configuration is therefore source-action-target); thus, ICEWS relies on the classic event data setup of *who did what to whom?* For this project, we take the many CAMEO codes and convert them into root codes by taking the first two digits of each CAMEO code observed. Doing so produces 20 CAMEO root codes (01, 02, 03, 04, 05, 06, 07, 08, 09, 10, 11, 12, 13, 14, 15, 16, 17, 18, 19, 20). From those root codes, we classify the actions further and more broadly into neutral, cooperation, and conflict types. Root codes 01 and 02 represent neutral actions between a source actor and target actor. Root codes 03 through 08 represent cooperation-based actions between a source actor and target actor. Root codes 09 through 20 represent conflict-based actions between a source actor and target actor.

Visual Analysis: Charts, Figures and Tables

All visualizations for this paper are produced in R. Figure 1 is the geographic map of the ASEAN countries and China, which assists the reader with knowing where the countries are in relation to each other. To make the geographic map, this project uses the mapdata R package and the ggplot2 R package. The mapdata R package allows the researcher to extract longitude and latitude values for a particular region from the world map. The ggplot2 R package focuses on

creating the graph that best reveals the message in the data and produces publication-quality graphics (Wickham, 2016).

For the igraphs, the igraph R package is utilized. The colored bands (edges) that connect the countries (nodes or vertices) represent the actions. Neutral actions are shown in "slategray4," cooperation-based actions are shown in "lightskyblue," and conflict-based actions are shown in "magenta." Figure 2 is an animated set of igraphs. The magick R package is used to animate the collection of igraphs. Figures 3 through Figure 11 are the static images that comprise the animation. While the animation provides some benefit, viewing the static images makes it quicker and easier for the user to ascertain the types of interactions that are taking place between the ASEAN countries and China, as the static layout is more conducive to efficient visual comparison.

The static images in Figure 3 through Figure 11 show how the relationships between the ASEAN countries and China change over time. According to the static images, the majority of events that occur between the ASEAN countries and China are cooperative. The data set mainly captures the interactions between China and Cambodia. Although it may seem odd that for some of the years, ICEWS only registers events dealing with China and Cambodia, that is what is recorded in the data. As an exercise outside the scope of the research question, we expanded our data set by allowing interactions between the ASEAN countries (in addition to their interactions with just China). We found that most of them are interacting with each other over the same 2012 to 2020 time period that our research question covers. Ultimately, we cannot say why some years only show interactions between China and Cambodia. It is possible, however, that indeed none of the news sources from which ICEWS gets its data mentioned interactions between China and a country other than Cambodia for an entire year.

Although ICEWS mainly captures the interactions between China and Cambodia over time, for certain years, it captures some key events between the ASEAN countries and China. For instance, Figure 3 shows the relations between the ASEAN countries and China in 2012 and the graph captures the conflictual relationship between the Philippines and China, and Vietnam and China. In 2012, there was the Scarborough Shoal Incident, in which the Philippines dispatched a warship to confront Chinese fishing boats in the Scarborough Shoal. China reacted by quarantining fruit imports from the Philippines and warned against tourism to the country; meanwhile, Beijing maintained regular patrols that prevented Philippine fishermen from accessing the water (Council on Foreign Relations, 2020). Vietnam passed maritime law asserting its jurisdiction over the disputed Spratly and Paracel Island, demanding notification from any foreign naval ship passing through the area, and China issued a strong response (Council on Foreign Relations, 2020). 2012 marks the first time in history that ASEAN failed to issue a communique at the conclusion of its annual meeting, the ten member states reached an impasse over China's claim in the South China Sea, and disagreed over whether to include the territorial issue in the joint statement (Council on Foreign Relations, 2020).

In 2016, Beijing deployed missiles on Woody Island. In the same year, the Philippine President announced the establishment of a no-fishing zone and marine sanctuary in the Scarborough Shoal. Neither China nor the Philippines has relinquished the sovereignty claim but the leaders seem poised to assume a more reconciliatory approach (Council on Foreign Relations, 2020). This can be observed in the conflictual and neutral relations between China and the Philippines in Figure 7. As tension cools between China and both Vietnam and the Philippines, we observe more and more cooperative interactions between China and those two countries, as well as with the rest of the ASEAN countries.

Figure 1.

Figure 2. For animated version, see the following:

https://t5904j-kjaura1.shinyapps.io/Storyboard/_w_529b15f0/Story.Rmd#section-igraph-and-map

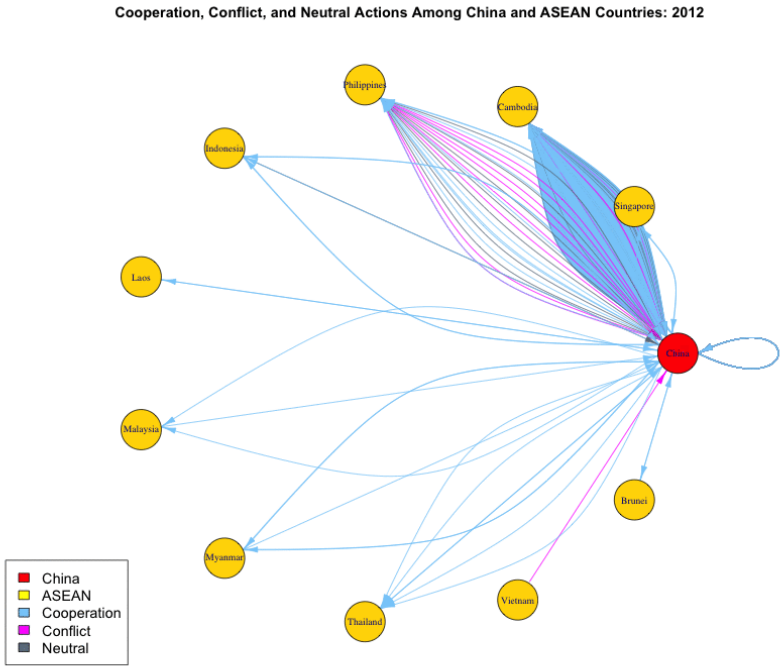


Figure 3.

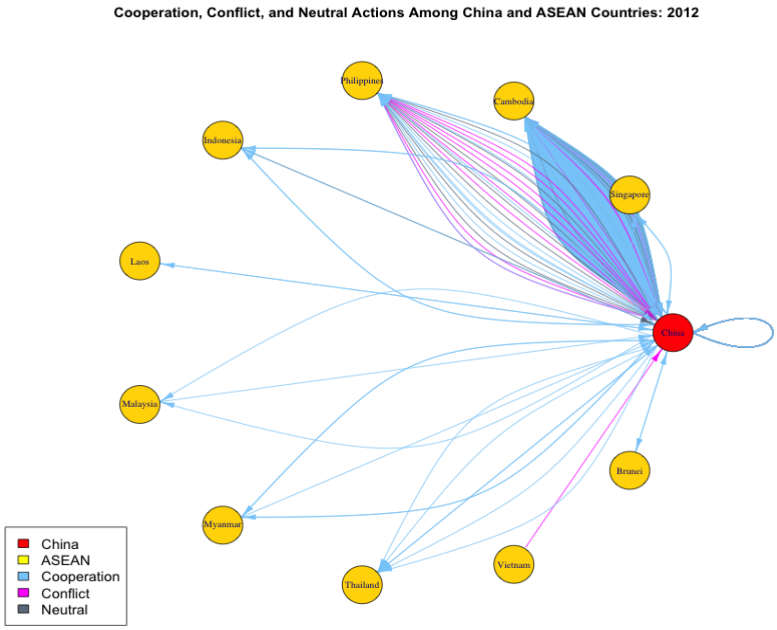


Figure 4.

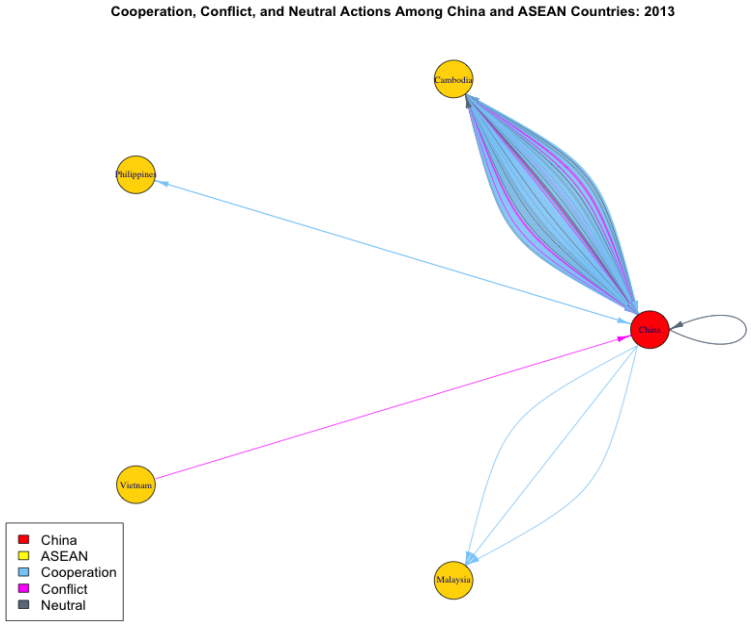


Figure 5.

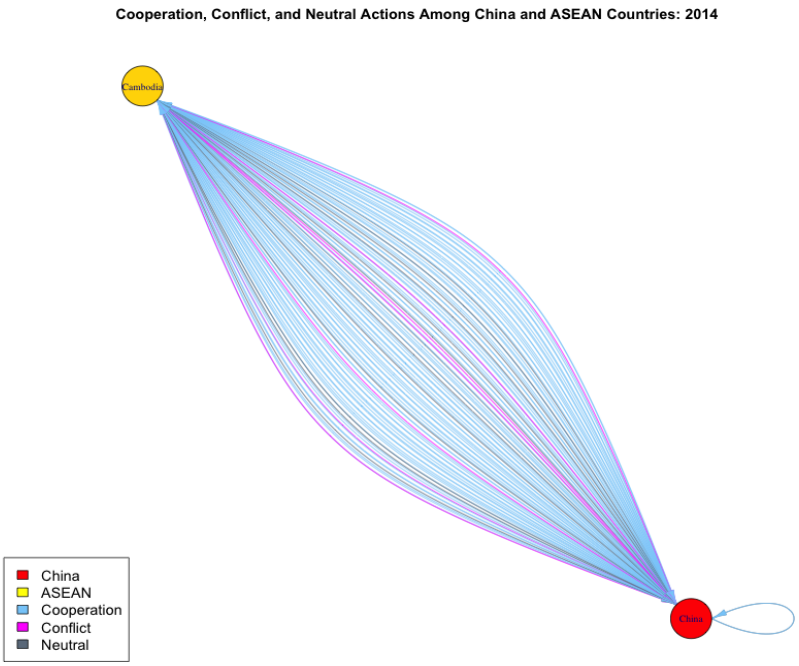


Figure 6.

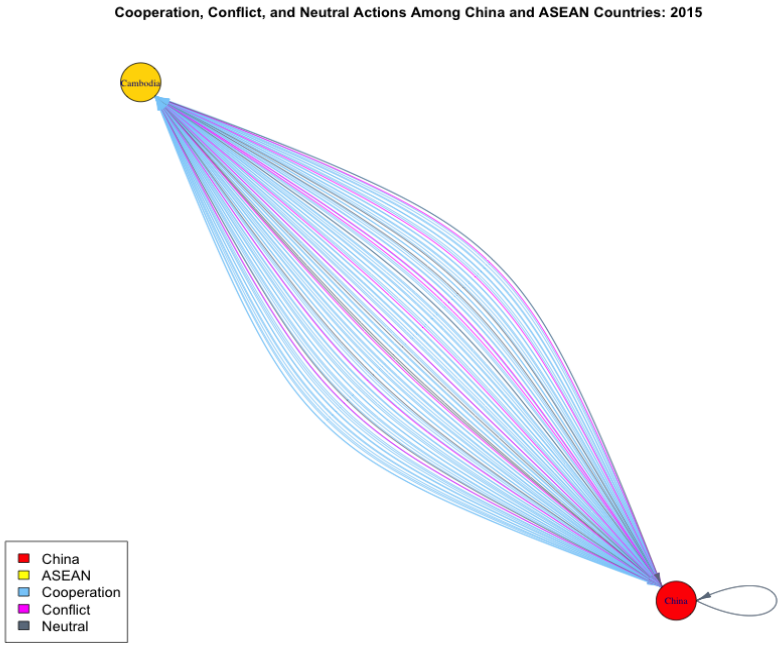


Figure 7.

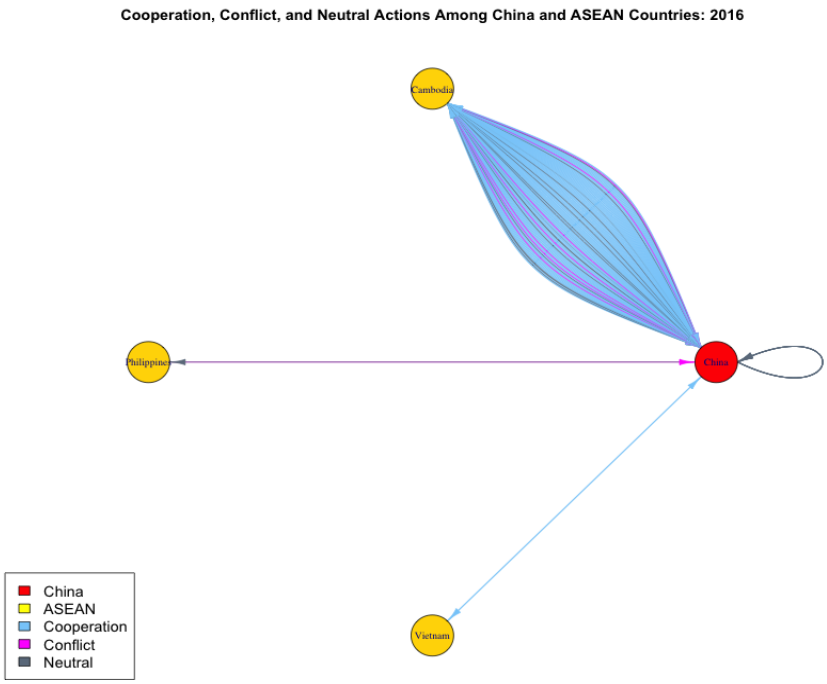


Figure 8.

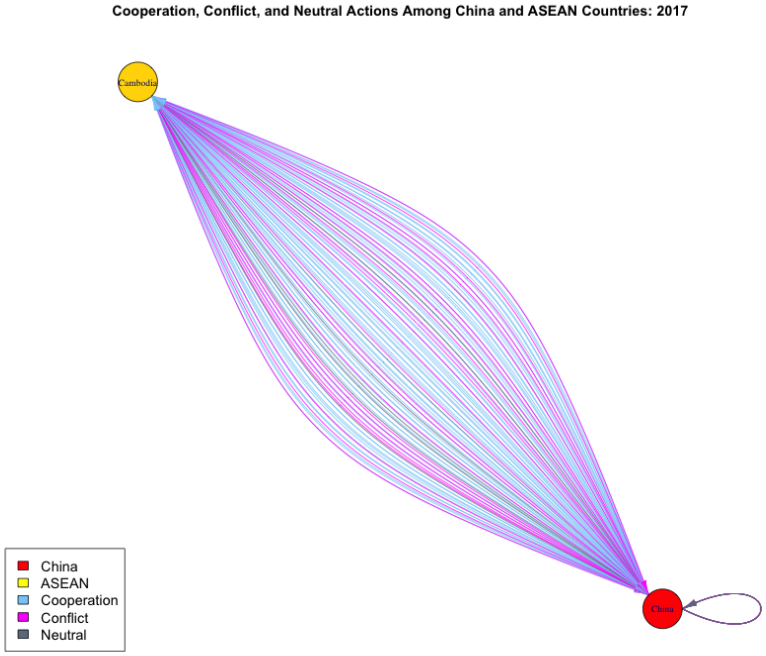


Figure 9.

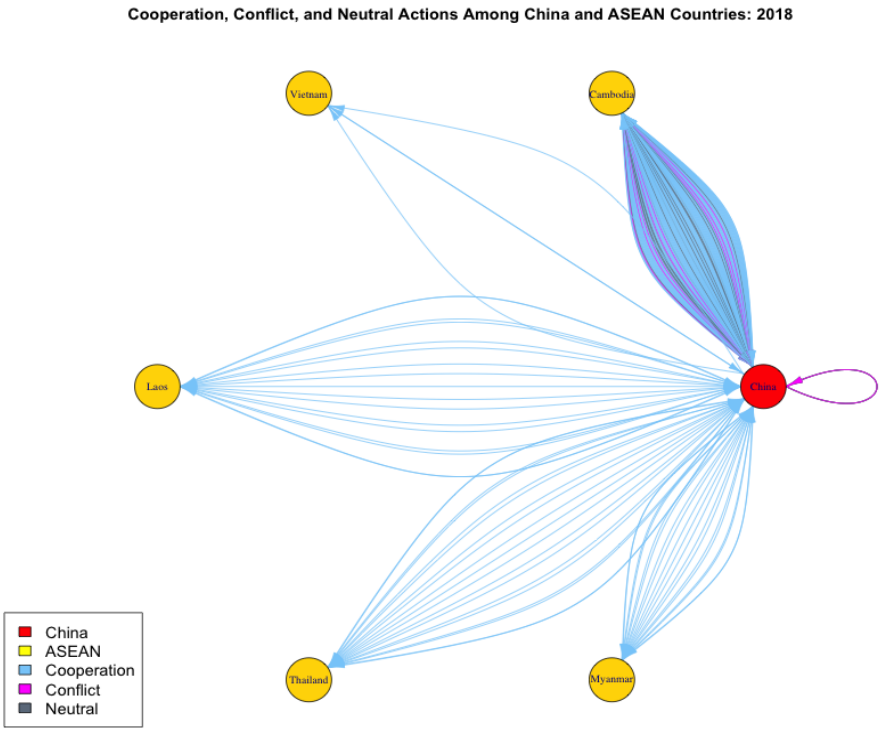


Figure 10.

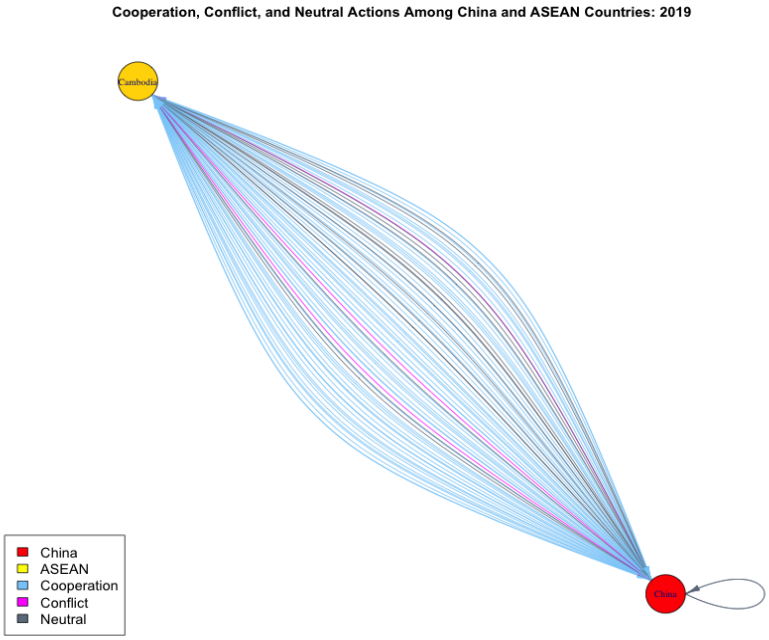
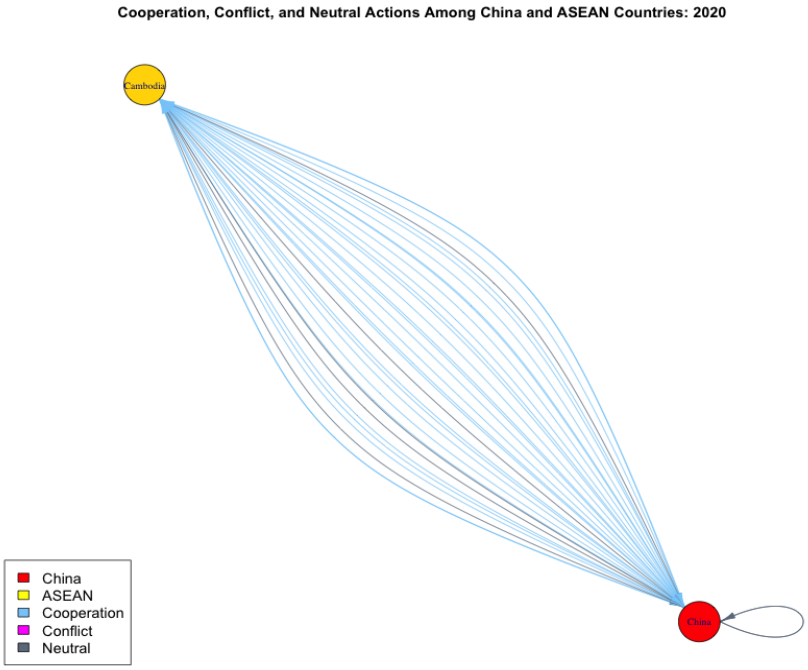


Figure 11.

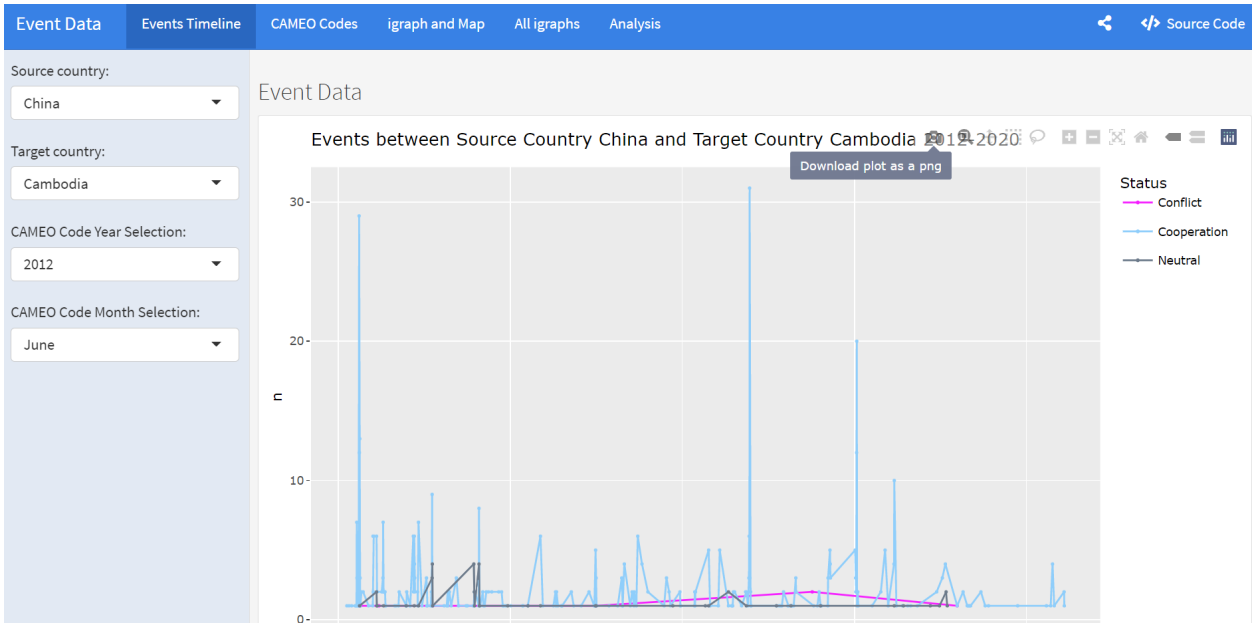


In addition to our igraphs and geographic map, we incorporated interactive visualizations via a couple of R Shiny applications. Through the R Shiny applications, we hoped to provide researchers the tools necessary to guide their own research. The first of our Shiny apps was an overall timeline that showed what all of the data looked like for 2012-2020 in terms of the status of "cooperation," "neutral," or "conflict." These statuses were colored to match the igraphs to ensure users of the dashboard understood the shared connection of data between the visualization from one perspective of the igraph for each year and the second perspective of the line graph showing events for all years in the data. This application limits the user to look at these events through one source country and one target country through a filter on the left hand side of the dashboard (see Figure 12). Changing the country specifications will also change the title of the visualization so that the source country and target country that the user is filtered to will be

reflected in the title. This is especially useful because a researcher could send the image with the title and be assured that the image would not be taken out of context because the title explicitly references the countries involved and the time period. Because the visualization uses plotly, the image can be downloaded as a png file and shared (see Figure 12).

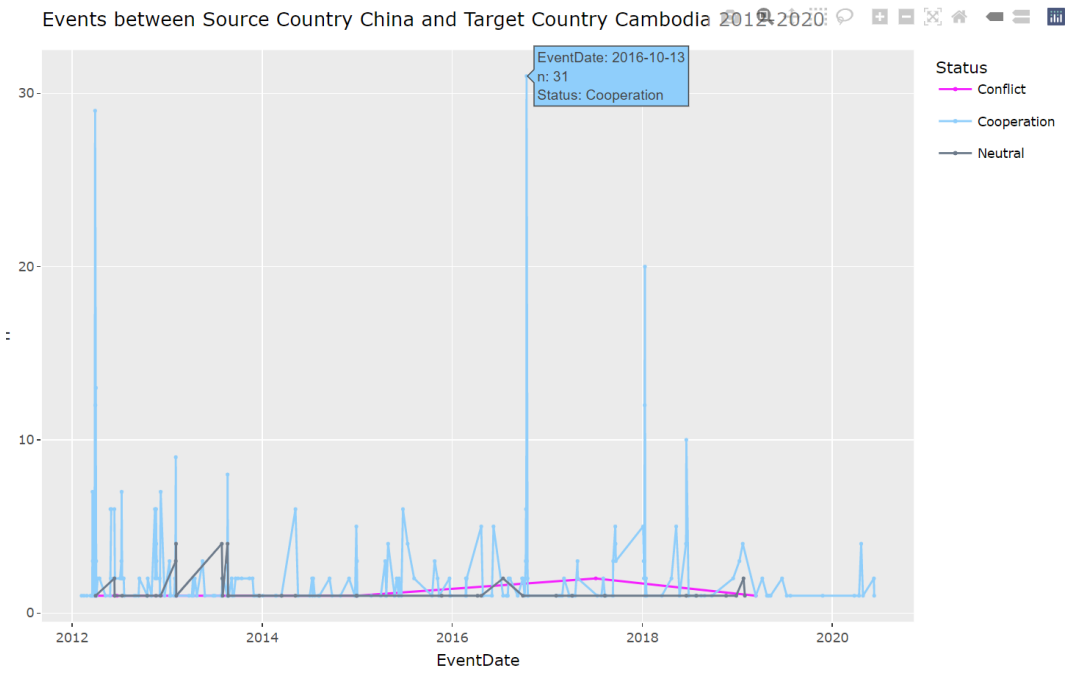
The plotly features in the visualization include: isolating the line of interest by double clicking on the line you want to see in the legend, zooming in, zooming out, panning, resetting axes, and, most importantly, hovering over a data point and reading the specific information for each data point. In our visual, the hover feature shows the date of the event and an aggregate of the number of events that happened for that event status type that day. These events are by day and are plotted within the year and month for which they occurred (shown on x-axis). Plotly guides the user on the capabilities of the visual once he or she clicks on a plotly feature.

Figure 12.



The filtering ability along with the plotly capability to hover on a point for details go hand in hand to make this a valuable tool for those investigating event data. One could hover over a point of interest in the visualization and get details on dates. Looking at this data, we see a high peak for number of events in the mid-October time range. Hovering on this point, we find it is for October 13, 2016, when we see a total of 31 events, as seen in Figure 13.

Figure 13.



Using the events found on this date, one could search for the significance of that date. In doing so, the results would show that China's president was visiting the king of Cambodia during this time. Both countries were showing events of "cooperation" during this time. From looking through the various relations between countries in the dashboard, it becomes clear that in this data set, most relationships are in fact related to China and Cambodia and most of the relations

they have are of the “cooperation” status type. This fact is important when analyzing this event data and will be referenced again throughout this paper.

As noted, some details are provided in the “Events Timeline” section of the dashboard when the user hovers over a data point, however this section is meant to be an overall macro view of the events to answer our research question: “How do relations between the ASEAN countries and China change over time?” To inquire further into the relevant details of a question like this one, we built a second R Shiny application within our dashboard to assist researchers in contextualizing the events that are taking place.

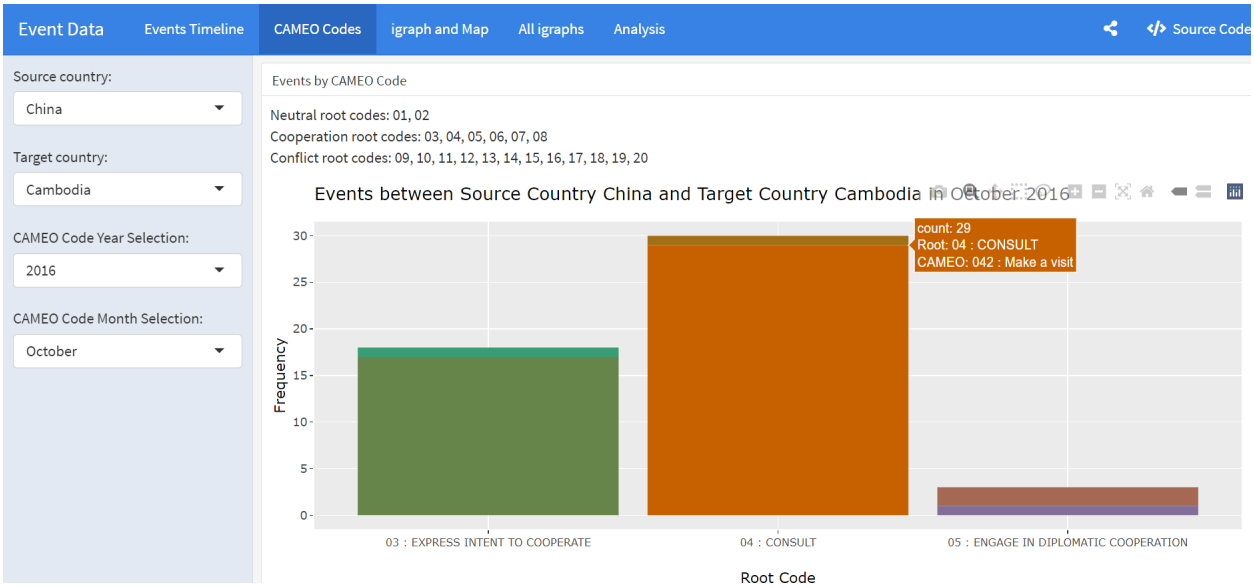
The second shiny app utilizes the same “Source Country” and “Target Country” filters but adds in the filter of “CAMEO Code Year Selection” and “CAMEO Code Month Selection.” Note that both of these additional filters contain the label “CAMEO Code” at the beginning. This was done to ensure that the user did not mistakenly think that changing either of these filters would affect a section of the dashboard outside of the “CAMEO Codes” section. Ideally, a user would only see the relevant filters to a section of the dashboard when looking at a specific visualization. However, due to limited space on the dashboard and our desire to allow more space for the visualization, the decision was made to keep one side options bar that would apply to the R Shiny applications and designate the ones that only apply to the “CAMEO Code” section with labels that begin with “CAMEO Code.” It is uncertain how clear the labels and filtering capabilities are for a user. The filters in the side bar are an area that needs to be tested with users to understand how they interpret these filters.

For the second app on the dashboard, the user is able to drill into the type of events that are taking place within the “cooperation,” “conflict,” or “neutral” status events from the “Events Timeline” section of the dashboard. Using the example of October 13, 2016, for China as source

country and Cambodia as target country, we can filter to 2016 in “CAMEO Code Year Selection” and October in “CAMEO Code Month Selection” and navigate to the “CAMEO Codes” section to see what kind of events were happening at that time. This part of the dashboard addresses the micro perspective on what event data looks like when broken into subparts of root codes. Here, one can see the events grouped together by Root Code and notes at the top of the dashboard explaining which root codes belong to each of the three event status types that were found on the “Events Timeline.” Like the first Shiny app, this one is also a plotly graph and has the capabilities that come with it.

While hovering on a bar within this second section, details for the Root Code, CAMEO codes, and aggregation of the number of events within this subcategories are provided. Continuing with the example of the October 13, 2016 date, a search for China and Cambodia on this day shows a meeting between China’s President Xi Jinping with King Norodom Sihamoni of Cambodia (FMPRC 2016). The meeting and consultations that were happening at the time between the two is reflected in the relatively frequent counts found in the “Make a visit” CAMEO code for October 2016, as seen in Figure 14.

Figure 14.



With user experience in mind, the R Shiny apps were shaped to incorporate capabilities to assist research into event data. While our use case looked specifically at China and ASEAN countries, the application of this app to other event data contexts are endless.The advantage of these Shiny apps is that they empower the user to do their own analysis and also reduce the amount of time spent digging through the data. The visualizations found in the R Shiny apps facilitate the discovery of trends and anomalies that would either require a large time commitment without the application or may not have been discovered if it were not for the interactivity provided through the Shiny apps, which display a large data set in an easy to digest and user-friendly interactive visualization.

Data Analysis

The primary variable of concern in our study was the CAMEO event code between China and the ASEAN countries. From our filtered data set from ICEWS, the event type classifications we assigned assigned a value to each (i.e., Cooperative = 1, Neutral = 0, or Conflictual = -1) and added this field to the data set. The detailed daily data was summarized by year and ASEAN country as averages. These averages were necessary to align with the country-level demographic and investment data.

The main purpose of obtaining the summarized event data was to compare the actions directed at the ASEAN countries by China to those ASEAN countries' demographics. Data sources included the following: the Economist Intelligence Unit (Economist Intelligence Unit. 2022), Global Entrepreneurship (Gem Global Entrepreneurship Monitor. 2022), UN's International Telecommunications Union (ITU. 2022), and the World Bank (World Bank. 2022). Table 1 provides a description of each demographic element and that element's source.

Table 1.

Variable	Source
Democracy Index	The Economist Intelligence Unit
Commercial and professional infrastructure	GEM - Global Entrepreneurship Monitor
Internal market dynamics	GEM - Global Entrepreneurship Monitor
Internal market openness	GEM - Global Entrepreneurship Monitor
Percentage of Individuals using the Internet	UN's International Telecommunication Union (ITU)
Access to electricity (percent of population)	World Bank
Armed forces personnel (percent of total labor force)	World Bank
GDP growth (annual percent)	World Bank
Gross capital formation (percent of GDP)	World Bank
Military expenditure (percent of GDP)	World Bank
Trade percent of GDP)	World Bank
GINI index	World Bank
Bachelor's or equivalent, population 25+, female percent) (cumulative)	UNESCO / World Bank
Bachelor's or equivalent, population 25+, male (percent) (cumulative)	UNESCO / World Bank
Proportion of seats held by women in national parliaments (percent)	Inter-Parliamentary Union (IPU) / World Bank
Proportion of women in ministerial level positions (percent)	Inter-Parliamentary Union (IPU) / World Bank
Strength of legal rights index (0=weak to 12=strong)	World Bank, Doing Business project

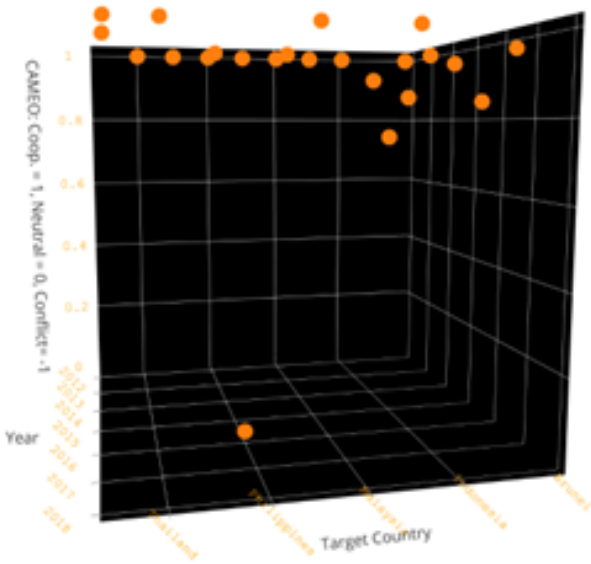
All available data was annual. Some variables had missing periods requiring a degree of extrapolation and stylizing. Only the following ASEAN country data was available given the time constraints we were under: Indonesia, Malaysia, Singapore, and Vietnam. Additionally, this information was also available only between 2012 through 2018.

In addition to comparing actions to national demographics, the impact of China's investment level was also a possible factor in action types. Data on Chinese investment in other countries was found in the China Global Investment Tracker portal (Center for Preventive Action, 2022). Data available from this portal included elements for year, month, investor, quantity in millions, share size, transaction party, sector, subsector, country, and region. However, only the amount of investment in ASEAN countries by China for the years 2012 to 2020 was extracted from this data.

The first analysis was to evaluate the CAMEO classification (i.e., Cooperative = 1, Neutral = 0, or Conflictual = -1) codes for the ASEAN countries and years available. There were 23 data points available for this analysis. The values for the CAMEO classification variable

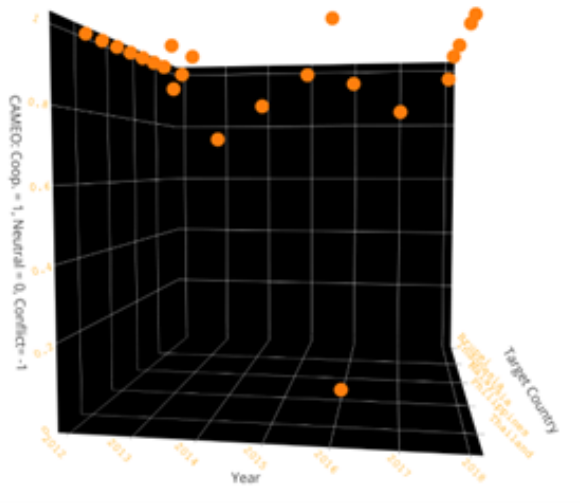
ranged from 0 to 1, with a mean of 0.92 and a standard deviation of 0.21. The values were further examined using a 3-dimensional rotatable scatter plot placing the Target Country on the x-axis, the year along the y-axis, and the average CAMEO classification value for a given country and year combination along the z-axis. Each combination was represented by an orange point set against a contrasting black background. As seen in the chart below, most countries, except for the Philippines in 2016, which had an average CAMEO code of zero, had very high CAMEO codes. Generally, 0.90 or above:

Figure 15.



A rotation toward the y-axis year values indicates that the high CAMEO classification remained consistent throughout the available years:

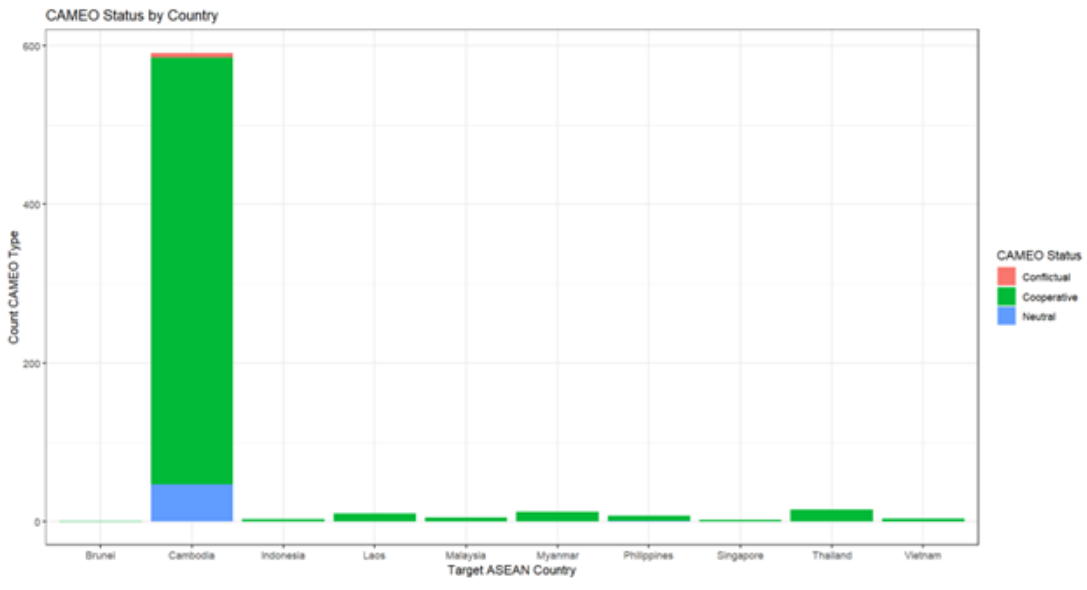
Figure 16.



Unfortunately, the data set does not appear to contain significant variability, which will limit its applicability to regression-type analysis.

A stacked bar chart was utilized to evaluate the spread of CAMEO classifications between target countries in more detail. Using primary colors (i.e., red, blue, and green) for CAMEO codes and placing the count of those codes on the y-axis and the target country label on the x-axis, Figure 17 was developed.

Figure 17.



This chart confirms the earlier chart's conclusion that most CAMEO codes in this data set are of a cooperative nature (in the visualization, there is an abundance of green shading). However, one obvious limitation demonstrated by this graphic is the concentration of available data in one target country, Cambodia. This graphical analysis is confirmed by Table 2.

Table 2.

		Count				Percent			
Country	Year(s)	Cooperative	Neutral	Conflict	Totals	Cooperative	Neutral	Conflict	Totals
Brunei	2012	1	0	0	1	100%	0%	0%	100%
Cambodia	2012 to 2020	568	50	7	625	91%	8%	1%	100%
Indonesia	2012	3	0	0	3	100%	0%	0%	100%
Laos	2012 and 2018	10	0	0	10	100%	0%	0%	100%
Malaysia	2012 and 2013	5	0	0	5	100%	0%	0%	100%
Myanmar	2012 and 2018	12	0	0	12	100%	0%	0%	100%
Philippines	2012 and 2013	6	1	0	7	86%	14%	0%	100%
Singapore	2012	2	0	0	2	100%	0%	0%	100%
Thailand	2012 and 2018	15	0	0	15	100%	0%	0%	100%
Vietnam	2016 and 2018	4	0	0	4	100%	0%	0%	100%
	Totals	626	51	7	684	92%	7%	1%	100%

Note that there is a high concentration of cooperation within the single country of Cambodia. This concentration will again limit the applicability of regression-type analysis on this data set.

The available demographic data are summarized in Table 3. While there is some variability in the data, the country and year count limitations are likely to limit this information's applicability to regression analysis.

Table 3.

Country		Democracy Index	Commercial and professional infrastructure	Country internal market dynamics	Country internal market openness	Percentage of individuals using the Internet	Access to electricity (% of population)	Armed forces personnel (% of total labor force)	GDP growth (annual %)	Gross capital formation (% of GDP)
Indonesia	Years: 2012 - 2018									
	Mean	6.8	3.1	3.9	2.9	23.8	97.3	0.5	5.2	34.2
	Standard Deviation	0.27	0.38	0.21	0.32	9.56	0.82	0.01	0.41	0.51
Malaysia	Years: 2012 - 2018									
	Mean	6.5	3.2	3.7	2.9	61.0	100.0	0.9	5.2	25.3
	Standard Deviation	0.05	0.17	0.18	0.12	24.75	0.07	0.04	0.59	0.83
Singapore	Years: 2012 - 2018									
	Mean	7.6	2.9	3.3	2.6	70.9	100.0	4.2	3.7	27.9
	Standard Deviation	0.28	0.47	0.08	0.53	27.83	0.00	0.89	0.75	1.74
Vietnam	Years: 2012 - 2018									
	Mean	3.2	2.8	3.7	2.6	40.0	99.8	1.0	6.2	26.9
	Standard Deviation	0.23	0.06	0.29	0.14	16.34	0.29	0.02	0.70	0.43

Country		Military expenditure (% of GDP)	Trade (% of GDP)	Bachelor's or equivalent, population 25+, female (%)	Bachelor's or equivalent, population 25+, male (%)	GINI	Proportion of seats held by women in national parliaments (%)	Proportion of women in ministerial level positions (%)	Strength of legal rights index (0=weak to 12=strong)
Indonesia	Years: 2012 - 2018								
	Mean	0.8	44.9	8.4	9.0	39.1	18.3	18.7	5.0
	Standard Deviation	0.08	3.68	0.66	0.53	0.77	1.26	6.60	1.00
Malaysia	Years: 2012 - 2018								
	Mean	1.3	137.6	12.3	10.4	41.5	10.9	8.6	7.0
	Standard Deviation	0.21	5.73	na	na	1.07	1.32	4.49	0.00
Singapore	Years: 2012 - 2018								
	Mean	3.1	318.7	26.5	31.0	39.6	23.9	5.5	8.0
	Standard Deviation	0.05	51.65	2.02	2.05	2.99	0.79	5.57	0.00
Vietnam	Years: 2012 - 2018								
	Mean	2.3	172.3	na	na	35.2	25.4	7.0	7.3
	Standard Deviation	0.10	8.90	na	na	0.33	1.26	2.66	0.49

Finally, Table 4 below summarizes the available information by country relating to China's capital investments made in the ASEAN countries:

Table 4.

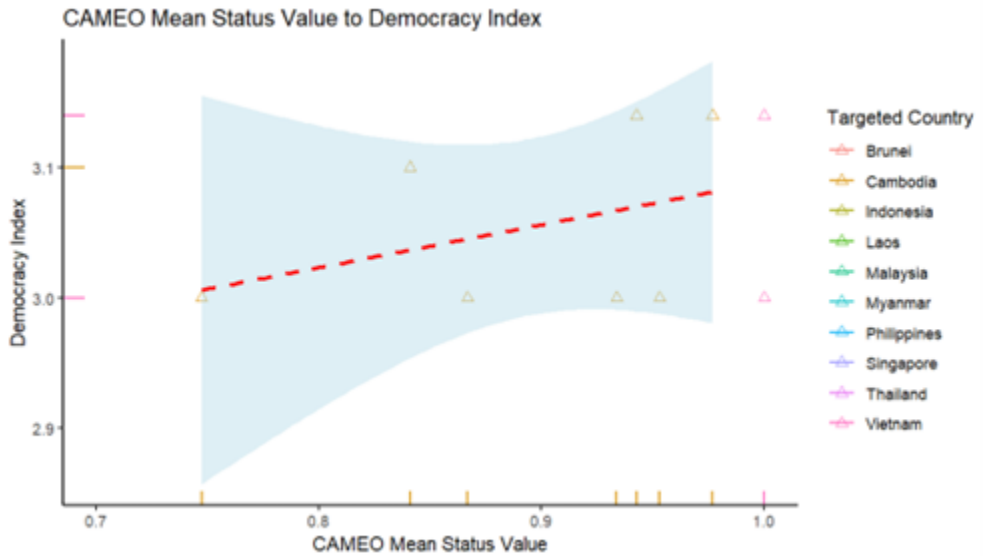
Country	Year(s)	Total China Investment ($ mil)	Average Annual
Brunei	2014	3,440	3,440
Cambodia	Various 2012 to 2020	5,800	829
Indonesia	Various 2012 to 2020	25,920	2,880
Laos	Various 2012 to 2020	12,620	1,803
Malaysia	Various 2012 to 2020	17,310	1,923
Myanmar	Various 2013 to 2020	3,390	848
Philippines	2017 and 2019	5,250	2,625
Singapore	Various 2012 to 2020	29,420	3,269
Thailand	Various 2012 to 2020	4,160	594
Vietnam	Various 2012 to 2020	6,590	824

Regression Results

As noted above, there are several concerns that the available data lacks sufficient quantity and variation to support regression analysis. To quickly test this conclusion, three separate scatter plots compare the average CAMEO status value to the Democracy Index, Military Expenditures, and the amount of investment made by China.

The hypothesis for a relationship between the Democracy Index and the CAMEO status value is that China, a single-party government, is likely to be more sympathetic to countries with a lower Democracy Index than countries with a higher index. The scatter plot in Figure 18 below uses all available target countries and years and includes a regression line indicating the direction of the relationship and shaded confidence intervals representing the possible range of the relationship:

Figure 18.

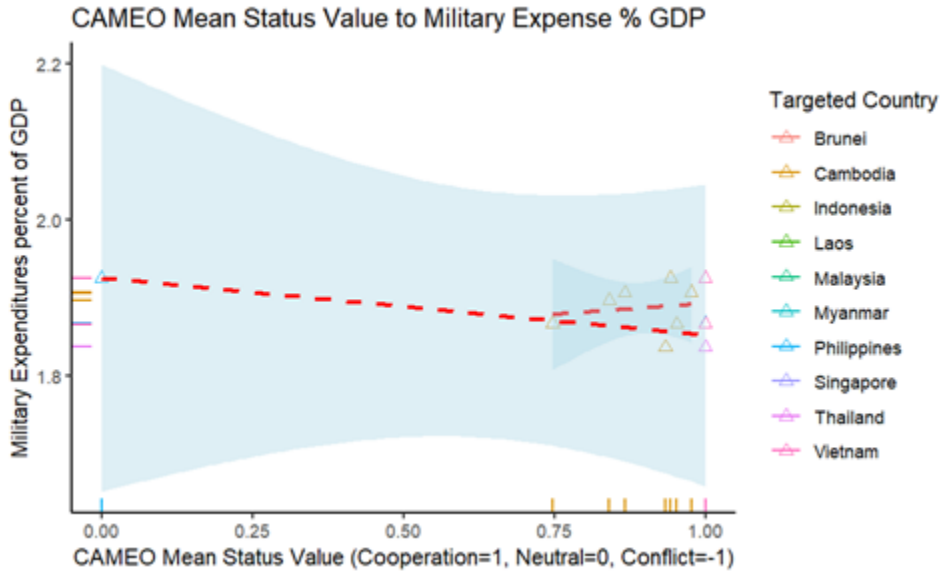


As this plot indicates, the relationship between the Democracy Index and the CAMEO codes appears to be a positive relationship. The higher the Democracy Index level (i.e., greater levels

of democracy), the greater the CAMEO code is likely to be cooperative. This is the opposite of what was expected. Additionally, the confidence intervals indicate a wide band suggesting that the relationship cannot be relied upon.

The hypothesis for a relationship between the Military Expenditures and the CAMEO status value is that China is likely to be more aggressive towards countries with higher military spending than countries with lower military expenditures. The scatter plot in Figure 19 uses all available target countries and years and includes a regression line indicating the direction of the relationship and shaded confidence intervals representing the possible range of the relationship:

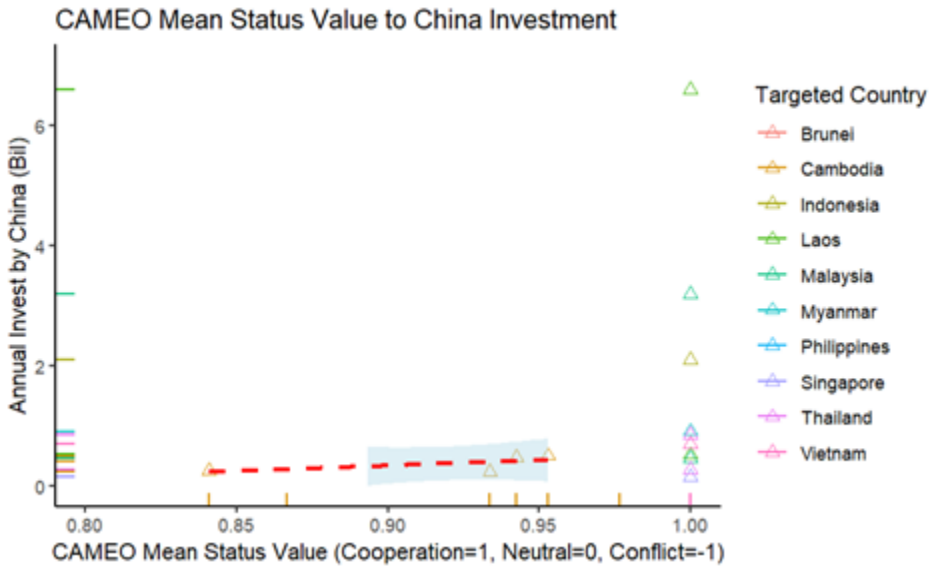
Figure 19.



Here the regression line appears to be downward sloping, suggesting that the hypothesis may have some validity; again, the confidence bands give little indication that this relationship can be relied upon.

The hypothesis for a relationship between the investment China makes within a country and the CAMEO status value is that China is likely to be more conciliatory towards countries where China has made investments than those it has not. The scatter plot in Figure 20 uses all available target countries and years and includes a regression line indicating the direction of the relationship and shaded confidence intervals representing the possible range of the relationship:

Figure 20.

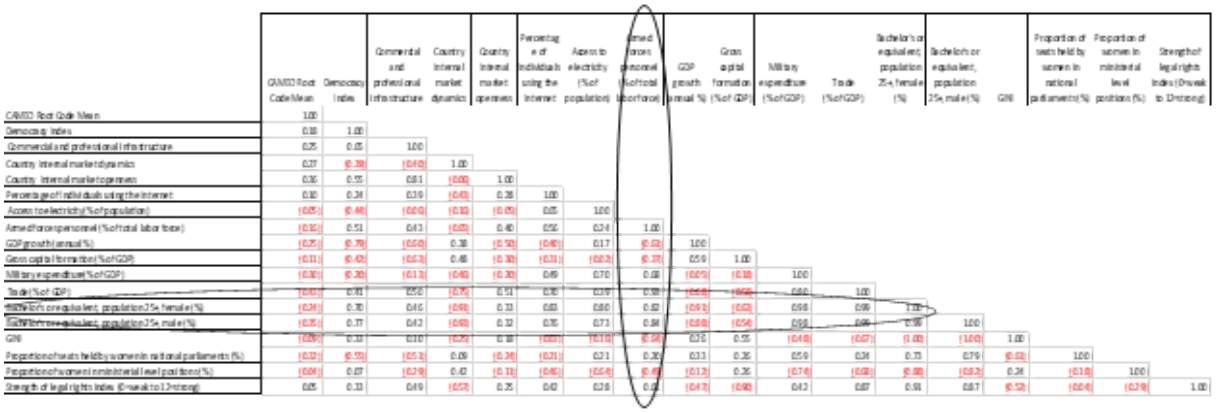


Here no relationship appears to exist. Notice how China might invest a significant amount in a country or nothing at all, and still, the CAMEO code generally represents a cooperative message. Therefore, the hypothesis does not appear valid.

A few regression models were developed to explore other relationships between the CAMEO codes and demographic and investment data. These models confirm that the data set's limitations prohibit any meaningful interpretation. However, these models were developed so that future analysis can be performed once a sufficient data set can be obtained.

Before any regression analysis, the potential correlation between the independent variables should be evaluated. Multicollinearity can cause wide confidence intervals leading to the acceptance of zero null hypotheses, statistically insignificant coefficients, and larger R-squared values (Damodar 2009, 327). That is, models can become overly sensitive to model specifications when multicollinearity is present, and multicollinearity can reduce the precision of the estimators, creating difficulty in evaluating their statistical significance. To evaluate the potential for multicollinearity, a correlation matrix was prepared. Generally, a correlation of 70% to 80% should be of concern. Table 5 shows the correlation matrix between the demographic data and the mean CAMEO root code:

Table 5.



From this analysis, three variables raised concerns. These are armed forces personnel as a percentage of the total labor force, Bachelor's or equivalent for females, and Bachelor's or equivalent for males. These variables are highly correlated with other variables and will be removed during regression efforts.

The first model tests the relationship between the CAMEO codes and the other variables, excluding the highly correlated variables noted above. The final model report is found in Table 6.

Table 6.

Coefficients	Estimate	Std. Error	t value	Significance
Intercept	-41.7	24.2	-1.7	None
Democracy Index	-24.8	12.3	-2.0	At 0.1
Commercial and professional infrastructure	16.6	7.2	2.3	At 0.05
Internal market dynamics	11.1	4.6	2.4	At 0.05
Internal market openness	12.6	8.3	1.5	None
Investment by China	0.0	0.0	-0.3	None

Adjusted R-Squared: 0.29

Several variables not reported in Table 6 were dropped because of singularity. Therefore, Table 6 only shows those variables that produced estimates. Note that the Democracy index is reporting a slight relationship. Additionally, the Internal market dynamics and openness are reporting a slightly larger relationship; however, given the data set, it is difficult to consider these relationships meaningful.

A second regression model was computed using only the three variables analyzed above using scatter plot relationships. This model reports as found in Table 7.

Table 7.

Coefficients	Estimate	Std. Error	t value	Significance
Intercept	-6.9	7.7	-0.9	None
Democracy Index	6.8	4.6	1.5	None
MILXPND	-5.4	7.6	-0.7	None
Invest_by_China_mils	0.0	0.0	-0.2	None

Adjusted R-squared: 0.07

Here none of the variables report any significant relationship and the model has an extremely low adjusted R-squared value.

Finally, three machine learning models were computed on the data. These models included a linear regression forward stepwise, a backward stepwise, and the best subset. Unfortunately,

none of these models produced any satisfactory results. However, perhaps with additional data, they might bear fruit at a future date.

Conclusion and Implications

The clear implication of this study is that more data is required. It is possible that with a larger data set containing additional country and demographic information, the two regression models and three machine learning models developed here might be able to identify meaningful relationships. In the future, consideration should be given to replacing the mean CAMEO root code (i.e., 01 to 20) with the average event classification score (i.e., 1 to -1). Additionally, a probit model (Damodar 2009, 566) may be more appropriate to use the classification score rather than a simple linear regression model to detect when communications switch between cooperation and conflict.

Finally, Exponential Random Graph Models (ERGM) is an established method for modeling network features and has been applied in political (Ingold and Leifeld 2014) and sociological (Lubbers and Snijders 2007) networks. These models test the likelihood of an observed interactive network process occurring given the characteristics of the nodes (e.g., country characteristics) within the network (Robins et al., 2007). Preliminary evaluations suggested that an ERGM might be appropriate in the context of this current analysis. However, given the time-intensive nature of data reconfiguration and model building required to implement this approach, it was deemed inappropriate for this current analysis. Still, it should remain an approach for future investigation. While our investigation did not yield concrete modeling results, we foresee our visualization tool becoming a beneficial tool for other event data research questions and one that may potentially yield another researcher valuable insights and results.

Synergy Report

Kiran Jaura (coordinator): Coordinated meetings with team and with instructor. Created flex dashboard and two shiny apps. Incorporated code from the team with their respective portions into flex dashboard for integration.

Glen Cooper: Completed the analysis and modeling section of this report. Efforts included gathering demographic information about the ASEAN countries and the amount of investment China made within those countries. That information was analyzed and compared to the UTD Event Data CAMEO codes to investigate potential relationships. Approach included both graphical and regression analyses.

Wen Si: Coded geographic map of the ASEAN countries and China.

Marcus Sianan: Provided subject matter expertise for the event data (in general and as it applies to ICEWS) and for the CAMEO ontology that was utilized for the project. Taught the other members how to pull data with the UTDEventData R package and interpret the generated dataframes. Coded the igraphs (static and animated versions) via the igraph R package, and consulted with Wen Si on the geographic map of the ASEAN countries and China. For the final report, he wrote the Literature and Data and Methods sections, and worked with Wen Si to describe the results from the igraphs (as written about in the Visual Analysis: Charts, Figures and Tables section).

References

Association of Southeast Asian Nations. https://asean.org

Boschee, E., J. Lautenschlager, S. O'Brien, S. Shellman, J. Starz, and M. Ward. 2015. "ICEWS" Coded Event Data." Harvard Dataverse V30. Accessed 20 February 2022. doi:10.7910/DVN/28075.

Brandt, P. T., D'Orazio, V., Khan, L., Li, Y. F., Osorio, J., & Sianan, M. (2022). Conflict Forecasting with Event Data and Spatio-Temporal Graph Convolutional Networks. *International Interactions*, 1-23.

Center for Preventive Action (2022). Territorial Disputes in the South China Sea. Global Conflict Tracker. https://www.cfr.org/global-conflict-tracker/conflict/territorial-disputes-south-china-sea

Chatterji, R. (2021). China's Relationship with ASEAN: An Explainer.

Council on Foreign Relations (2020). *China's Maritime Disputes.* https://www.cfr.org/timeline/chinas-maritime-disputes

D'Orazio, V., & Lin, Y. (2022). Forecasting conflict in Africa with automated machine learning systems. *International Interactions*, 1-24.

Economist Intelligence Unit. 2022. "Democracy Index 2018." 2021. https://www.eiu.com/n/campaigns/democracy-index-2018 (December 9, 2022).

Gem Global Entrepreneurship Monitor. 2022. https://www.gemconsortium.org/report/gem-2018-2019-global-report (December 9, 2022).

Gerner, D. J., P. A. Schrodt, and €O. Yilmaz. 2009. "Conflict and Mediation Event Observations (CAMEO): An Event Data Framework for a Post Cold War World." In International Conflict Mediation: New Approaches and Findings, edited by J. Bercovitch and S. Gartner, 287–304. New York: Routledge.

Huang, C. H. (2022). *Power and Restraint in China's Rise*. Columbia University Press.

ITU. 2022. "Committed to Connecting the World." ITU. https://www.itu.int/en/Pages/default.aspx (December 9, 2022).

Kim, HyoungAh, Vito D'Orazio, Patrick T. Brandt, Jared Looper, Sayeed Salam, Latifur Khan, and Michael Shoemate. 2019. "UTDEventData: An R package to access political event data." Journal of Open Source Software. 4(36), 1322, doi:doi:10.21105/joss.01322

The Recent History of the South China Sea: A Timeline (2021). International Crisis Group. https://www.crisisgroup.org/asia/south-east-asia/south-china-sea/recent-history-south-china-sea-timeline

World Bank. 2022. Baffes, John, Marie Caitriona Hyland, and Brian Blankespoor. 2022. "World Bank Open Data." Data. https://data.worldbank.org/ (December 9, 2022).

Wickham, H. "ggplot2: elegant graphics for data analysis. data." (2016).

Xi Jinping Meets With King Norodom Sihamoni of Cambodia. www.fmprc.gov.cn/mfa_eng/topics_665678/2016zt/XJPDJPZMJLGJXGSFWBCXZYDGAJXDJZGJLDRDBCHW/201610/t20161017_704254.html.