

Change or Continuity?: Topic Modeling of Thailand in Transition

Napon Jatusripitak
Northwestern University

Introduction

Do newspaper articles that are published before and after the occurrence of a regime transition differ systematically in terms of their topic coverage and content? Utilizing an original dataset containing all political news published from 2013 to 2015 by *The Nation*, an English-language daily newspaper in Thailand, this project explores the relationship between regime transition and the media. Specifically, I analyze Thai newspaper articles that were published under democratic rule (pre-May 2014 coup) to those that were published under military rule (post-May 2014 coup), using latent Dirichlet allocation (LDA) to examine topic variation amidst a regime transition.



Data Collection

All newspaper articles published by The Nation from 2013 to 2015 were collected from the Nexis Uni database via Relenium, which extracted relevant information such as the headline, content, and date. The search query used in this process of data collection is *thailand*. For the analysis, only articles that contain a set of keywords that are predictive of their political content were retained for the analysis. These keywords include: politic*, elect*, gov*, democ*, Yingluck, junta, Prayut, Pheu, and NCPO. This process yields a total of 21,371 articles where the average length is 573 words.

doc_id	text	title	date	wordcount
9611	the nation The junta has explained to foreign countri...	Junta explains reasons for coup to foreign nations	2014-05-26	Length: 373 words
9612	Agence France-Presse HONG KONG Young middle-cla...	Liquor companies look beyond China for new Asian d...	2014-05-26	Length: 483 words
9613	Usanee Mongkolporn The Nation @BT New Screen/bri...	@New screen multi head 32:NBTC to go ahead with s...	2014-05-26	Length: 408 words
9614	focus on growth @BT New Screen/briefs Text - no ind...	@New screen S/C head:Junta says no intention to hin...	2014-05-26	Length: 702 words
9615	The Straits Times Asia News Network @BT New Scree...	@New screen S/C head:Project to recruit Cambodian ...	2014-05-26	Length: 373 words
9616	@BT New Screen/briefs Text - no indent;"Hold" By Asi...	@New Subhead colour:CPF Pic	2014-05-26	Length: 1262 words
9617	Kwanchai Rungfapaisarn The Nation Major retailers in...	Retail giants set growth strategy under the AEC more ...	2014-05-26	Length: 598 words
9618	Panya Thiosangwan, Kornchanok Raksaseri The Natio...	Security offences to be tried in military courtsprayuth ...	2014-05-26	Length: 762 words
9619	eyeing sales growth but coup concerns linger Pichaya...	Thai Optical	2014-05-26	Length: 543 words
9620	Panya Thiosangwan, Kornchanok Raksaseri The Natio...	The Nation reporter taken into custodyNMG editor-in...	2014-05-26	Length: 565 words
9621	The US, the UN and the EU have all trotted out the sa...	Coup doesn't have to be #124public-relations disaster	2014-05-27	Length: 701 words
9622	Petchanet Pratrungkrai The Nation The Finance Minis...	Finance seeks Bt50-bn term loans to pay off rice farm...	2014-05-27	Length: 792 words
9623	legal action @BT New Screen/briefs Text - no indent;...	@HD2 B1 briefs head;PTT threatens	2014-05-27	Length: 411 words
9624	@BT New Screen/briefs Text - no indent; @BT New Sc...	@HD2 New Briefs head;Credit where credit's due, plea...	2014-05-27	Length: 540 words
9625	the heartThe life of Thailand's last executioner #124i...	@HD New Head 80 inside lead/cover;Bullet to	2014-05-27	Length: 970 words
9626	@BT New Screen Text Drop opinion;Shelter is recogni...	@HD New opinion Head 48:The difference between a ...	2014-05-27	Length: 931 words

Figure 1: Data structure

Pre-Processing

Using the tm package, I converted the data to a corpus object, transformed words to lower cases and removed stop words, numbers, punctuations, followed by stemming and removing white spaces. Then, I parsed the corpus to a document-term matrix with articles as documents in the rows and terms in the columns. Finally, sparse terms that appear in less than 0.01 of all documents are removed.

Topic Model (2013-2015)

Following Blei, Ng and Jordan (2003), I employ latent Dirichlet allocation (LDA) as an unsupervised method for learning the topics associated with the newspaper articles. This allows for documents to be represented as mixtures of topics characterized by a probabilistic distribution of words. For LDA, the dimensionality of the Dirichlet distribution, k, is assumed to be known and fixed. In this study, however, it is unclear what k should be. Hence, one challenge to the implementation of LDA is to select the optimal value of k. I implemented three topic models with k = 20, 30 and 40. Out of all three models, the model with k = 30 is most informative in terms of providing topic labels that are helpful for making an informed guess as to what each topic is about. The output of the model with 30 topics is parsed through R package LDAvis, a web-based interactive visualization of topics.

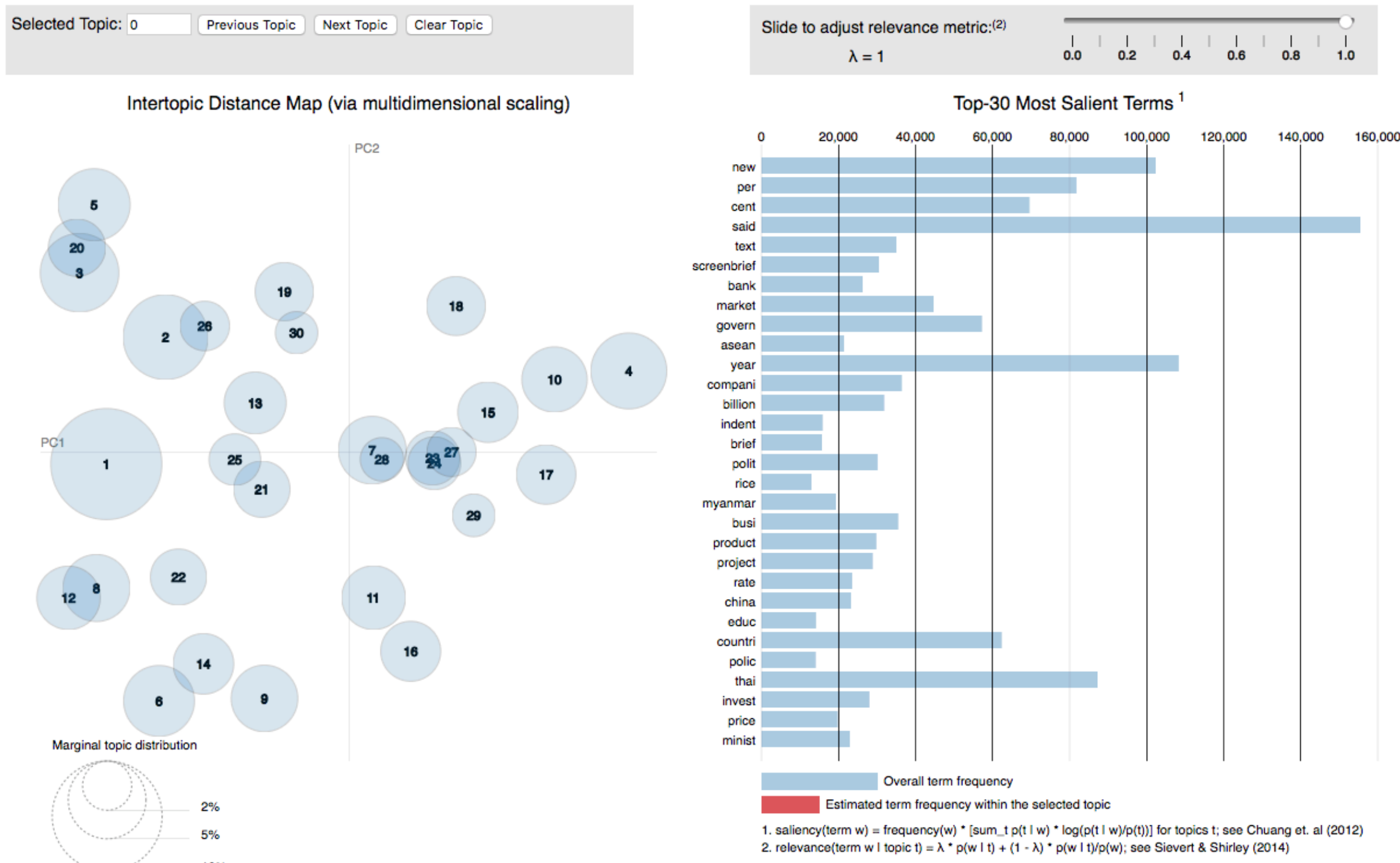


Figure 2: LDA visualization. URL: <https://naponjatusripitak.github.io/polisci490/Project/index.html>

Topic Variation Over Time

For the purpose of this study, it is important that we can observe the topic variation over time. To do so, I group the data by topic and month and plot the average probability of topics. Higher values indicate greater association of documents, and words in the documents, with a given topic. For visualization, I select 6 topics which provide insights into the relationship between regime transition and the media.

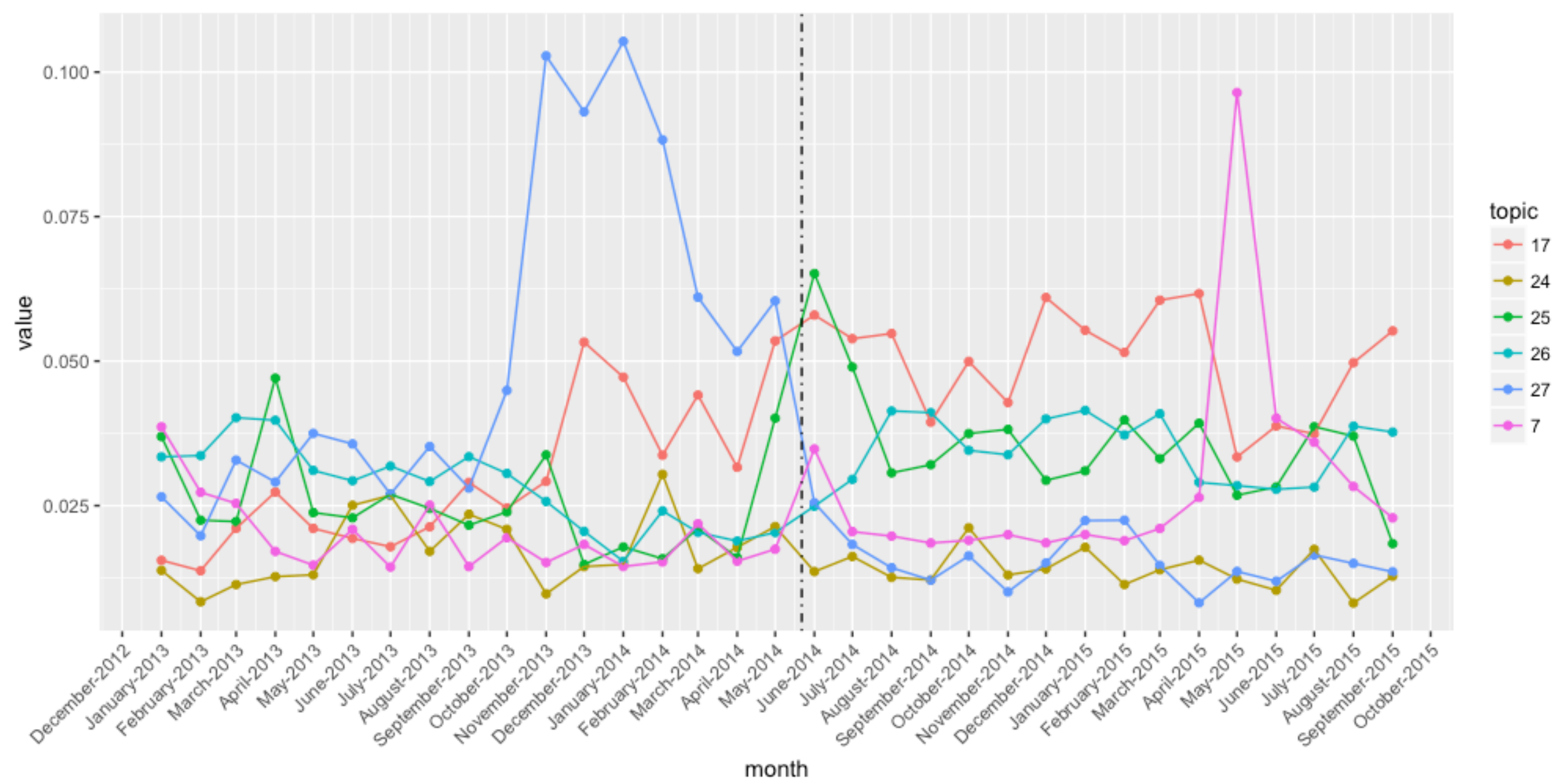


Figure 3: Based on informed guesses: Human Rights (7), Reform (17), Agricultural Policies (24), Military (25), Government (26), and Political Conflict (27). The dotted vertical line represents the May 2014 coup which removed the Pheu Thai government and installed a military government under the authority of the National Council for Peace and Order (NCPO).

Sentiment Analysis

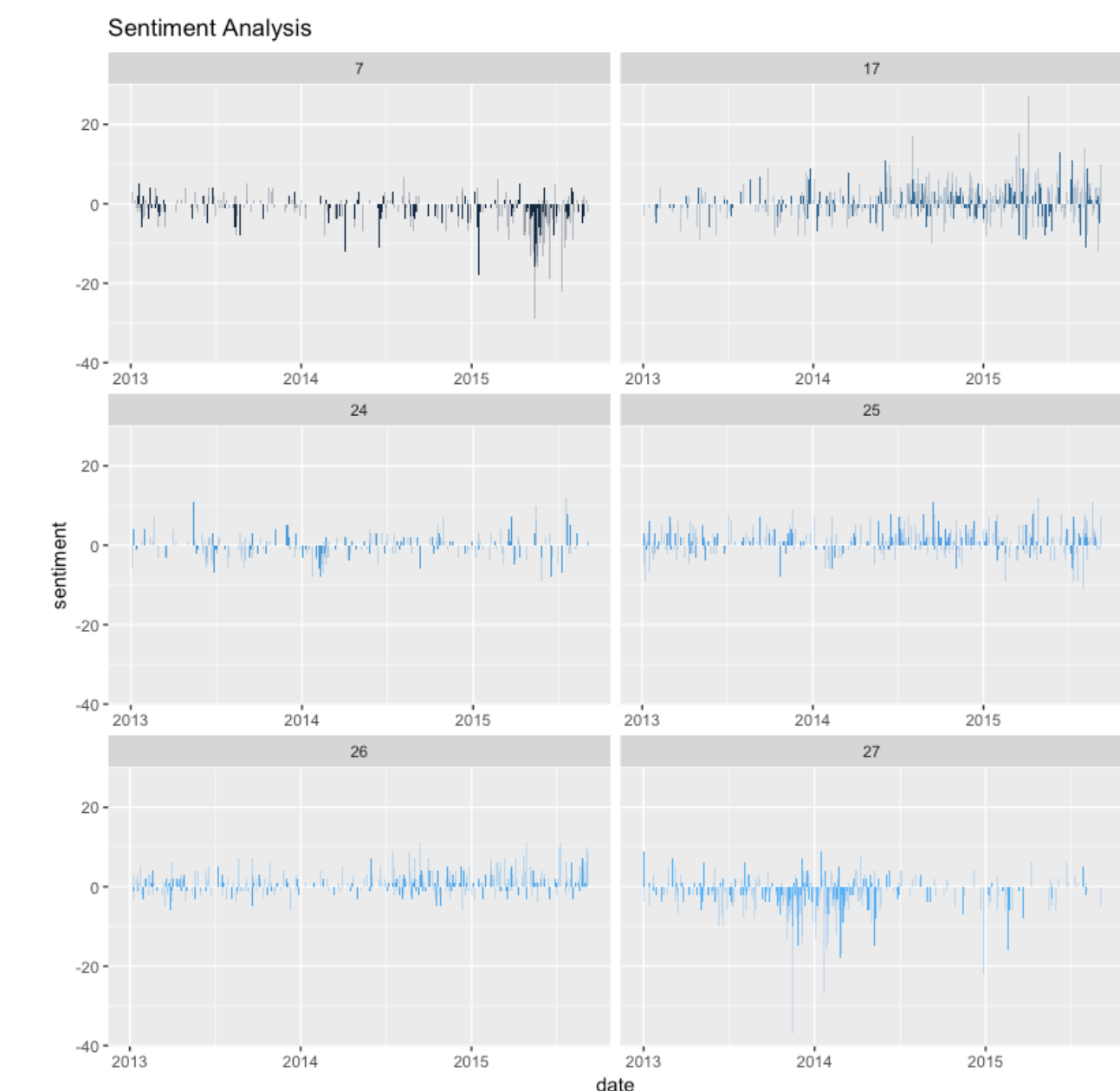


Figure 4: Sentiment Analysis

In addition to the topic model, I also conduct sentiment analysis using the labels generated by LDA. Each article is labeled as belonging to the topic with which it exhibits the highest probability.

On average, sentiments after the coup are more positive than prior to the coup which is an interesting finding, considering that military rule is conventionally associated with negative sentiments. This is most clearly illustrated in the plot for Political Conflict (27). On the other hand, Human Rights (7), which peaked in terms of negative sentiment in 2015, seems to contradict this overall trend. However, upon a closer look, the sentiment analysis seems to be picking up the 2015 Rohingya Refugee Crisis rather than the violation of civil rights by the military government.

Conclusions

What did you conclude? What additional work needs to be done?

- I did something great, get excited.
- More work for us!

References

Just as with regular \LaTeX documents, you can include a .bib file in the folder, and cite references on your poster