

Shaping Regional Development Policy Through News Coverage: Potential and Limitation

Adiwan Aritenang
Insitut Teknologi Bandung
Bandung, Indonesia
a.aritenang@gmail.com

Guntur Budi Herwanto
Universitas Gadjah Mada
Yogyakarta, Indonesia
gunturbudi@ugm.ac.id

Novri Suhermi
Institut Teknologi Sepuluh Nopember
Surabaya, Indonesia
novri@statistika.its.ac.id

Imaduddin Amin
Pulse Lab Jakarta
Jakarta, Indonesia
imaduddin.amin@un.or.id

Mellyana Frederika
Pulse Lab Jakarta
Jakarta, Indonesia
mellyana.frederika@un.or.id

ABSTRACT

For public policy to address a problem timely and correctly, ability to respond to an actual, perceived or anticipated problem is critical. Big data holds tremendous potential in providing information for policy analyst that is more timely, accurate and detailed. This paper examines the use of big data to shape public policy, to look on the potential of worldwide news capture in The Global Data on Event Location and Tone (GDELDT) project inform policy making in Indonesia context. GDELDT data sets reveal topic of interests from the following neighbourhood countries: Australia, Singapore and Malaysia on selected Indonesian big islands namely Sumatera, Kalimantan, Jawa, Bali and Papua. We concluded that big data from news play a role in shaping foreign perceptions towards specific Indonesian regions which could be responded and anticipated by policy analyst in that regions.

KEYWORDS

GDELDT Data, Public Policy, Indonesia

1 INTRODUCTION

For public policy to address a problem timely and correctly, ability to respond to an actual, perceived or anticipated problem is critical. Policy analyst have been using large, high-dimensional data sets as evidence to policy making. In the advance of technology, there are new sources of digital data known as big data available for policy analysis.

Shintler and Kulkarni argue that big data holds tremendous potential for public policy analysis, new resource for helping to inform different points in the policy analysis process, from problem conceptualization to ongoing evaluation of existing policies and even empowering and engaging citizens and stakeholders in the process. Big data can be useful in producing information that is more timely, accurate and detailed than that gleaned from more traditional sources of data [3].

This paper examines the use of big data to shape public policy, to look on the potential of worldwide news capture in GDELDT project inform policy making. The Global Data on Event Location and Tone¹ database contains nearly a quarter of billions geocoded records on global events going back to 1979 and collects 100,000 news events every day. The news from more than 100 languages has

been translated into English and, uses natural-language processing turned the news into data points. It is one of the largest open-access spatio-temporal datasets with total archives span more than 215 years, GDELDT provides a wealth and unprecedented amount of information on global societal system and behavior [2].

We ask the following, question "How can we use worldwide news inform and advise us to shape our development policies". This is done by (i) understanding the real-world events from GDELDT data, (ii) understanding the connection between sources and events and its connection to the real-world, and (iii) examining how a specific regional development policy in Indonesia can be formulated based on the world news.

2 RESEARCH METHODOLOGY

2.1 Data

We downloaded GDELDT 1.0 Global Knowledge Graph, specifically the GKG Counts File that are available since April 1st, 2013. GDELDT GKG collects data globally from news article from the previous day, and pairs a set of person names, organization names, locations, counts, and themes that then being called a *nameset*. The data is available in CSV format, with each unique *nameset* per row.

We focused on events happened in Indonesia, it is about 242,500 events out of 42 millions event available from GDELDT 1.0. We limit the research to the potential of GDELDT data to policy making in Indonesia context.

2.2 Methodology

2.2.1 Descriptive Analysis. To understand GDELDT data reflection to real-world event, we conducted descriptive anlysis by visualizing the simple count of news by date as seen in Figure 1. The graph shows top news on Indonesia are on natural disaster such as earthquake, volcano eruption and flood, and news on terrorism. The datasets provide near-realtime insights into what is happening in different places in Indonesia.

2.2.2 Location Analysis. The next step is understanding the connection between sources and events and its connection to the real-world. We infer sources of newsroom by identifying server location and domain name such as .au that indicates an Australian newsroom and .my that indicates a Malaysian newsroom. We notice the limitation of this approach. First, there are newsrooms with

¹www.gdeltd.org

server located outside their country such as United States of America. Second, there are newsroom that use .com or .co instead of country-related domain name.

The data sets consists of event's location up to city level. However, our analysis reveal list of news without specific location and incorrect location. Based on this finding, we analyse news content to identify location of the events based on selected big islands in Indonesia namely Sumatera, Kalimantan, Java, Bali and Papua. These are big islands that came at the top of the news list. Hence, other big islands such as Sulawesi and islands in East and West of Nusa Tenggara, Maluku and North Maluku Province are not selected.

2.2.3 Topic Analysis. GDELT data sets has given 44 categories such as kill, arrest, wound, protest and more. This is a broad categorization and focused on crisis and violent events. We created different category by using topic analysis called Latent Dirichlet Allocation (LDA) to identify news topic and category automatically. We identified 9 different topic models with three different news corpus. In order to create the corpus, the news article is transformed into a corpus with a bag-of-words form. The Corpus became an input for LDA topic modeling. Based on this, we pre-processed the news through the following steps: a) elimination stopword b) dictionary formation c) forming document matrix and d) creation of corpus with bag-of-words form. We enhanced the above-mentioned topics by manually adding relevant keywords according to researchers' knowledge (keyword spotting) [1].

2.2.4 Analyse GDELT data sets with international visitors datasets. As the last step, we combined the information from GDELT with Indonesia Statistics (Badan Pusat Statistik) data sets on number of international visitor per country of origin in the year of 2014 and 2015.

3 RESULTS AND DISCUSSION

Each neighboring country has different topic of interests towards specific Indonesian island. In the past five years, Australia main interest is the Bali Nine case. It is the name given to group of nine Australian convicted for attempting to smuggle heroin out of Indonesia in April 2005 and the execution of two convicts happened in April 2015. Singapore is the only country of origin that reported haze events. We found reports on forest-related issues such as plantation, orang utan, elephant from newsroom located in Malaysia and Singapore. We noted that forest-related issues are appeared only on news about Sumatera and Kalimantan. Hence, there is specific characteristic of an island that attracts certain topic. For Papua, topic around human right is appeared.

The newsroom located in Australia has significant coverage on Bali while newsroom located in Singapore and Malaysia shows a large interest in Sumatera and Kalimantan. The bordering regions also exhibit the area of cross-culture and political dimensions. The following examples showcase result of topic modeling on understanding main interests of neighboring country: a) Malaysia main topics are on Rohingya refugee, Jakarta governor election, tourism, and also the flood disaster, b) Singapore main topics are on forest fire, mining, tourism and gay issues and c) Australia main topics are on volcano eruption.

We compare this to Indonesian-based news on the three countries. We found the following coverage: a) on Malaysia: palm oil and batik b) on Singapore: tax amnesty, and c) on Australia: refugee and tourism.

We use international visitors statistics to see the correlation between the number of international visitors with the bad news reported by the country of origin of the international visitors. One of the results of the analysis can be seen in Figure 6. During bad news, the number of international visitors from Singapore and Australia to Indonesia is reduced.

4 CONCLUSION

GDELT data sets reveal real-world events in Indonesia with some limitations. GDELT data set suffers from temporal bias and captures only events with news-value. Category with the biggest number of nameset is 'kill' and keywords such as poverty is not in the top list. Create a different category by automatically create new topics combined with domain expert input can be valuable to capture information on specific issues such as tourism, politics and human rights issues. A further study on the regional development policy can provide better context to the automatic topic.

GDELT data is rich resources on global news that provide potential new sources for national and local analysis. Issues that are important for Australia, Singapore and Malaysia newsroom on Indonesian islands can be different with national and local perspective and this can be used to understand the different perspective and to provide appropriate policy towards foreign perspective. A further study is required to understand the different perspective between national and foreign news and how that affect regional development in Indonesia context.

However, GDELT database suffers from temporal bias. To avoid implementation of inappropriate or inequitable policies, it is important to understand the extent and nature of bias in the data, and if possible correct for it.

REFERENCES

- [1] David M Blei, Andrew Y Ng, and Michael I Jordan. 2003. Latent Dirichlet Allocation. *Journal of Machine Learning Research* (2003). <http://www.jmlr.org/papers/volume3/blei03a/blei03a.pdf>
- [2] Kalev Leetaru and Philip A. Schrodt. 2013. (2013). <http://data.gdelproject.org/documentation/ISA.2013.GDELT.pdf>
- [3] Laurie A. Schintler and Rajendra Kulkarni. 2014. Big Data for Policy Analysis: The Good, The Bad, and The Ugly. *Review of Policy Research* 31, 4 (jul 2014), 343–348. <https://doi.org/10.1111/ropr.12079>

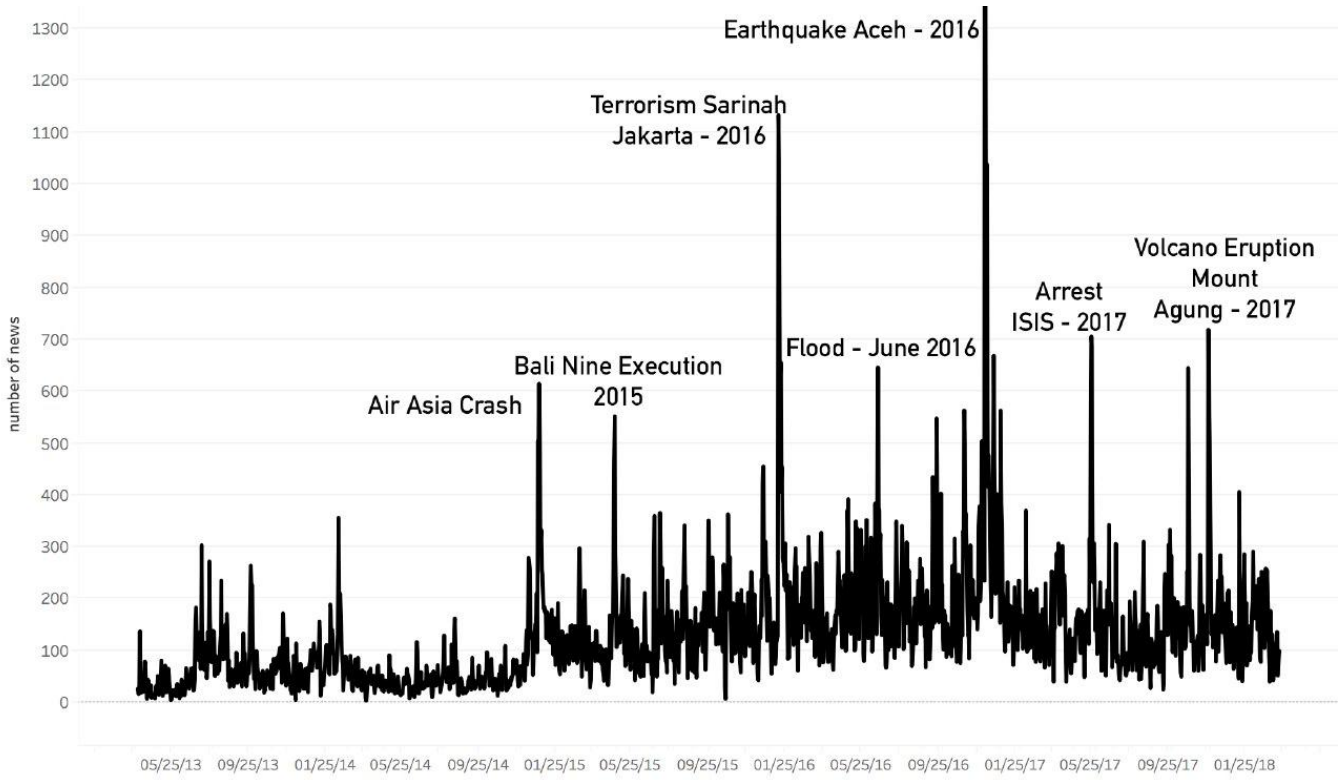


Figure 1: News Count by Date

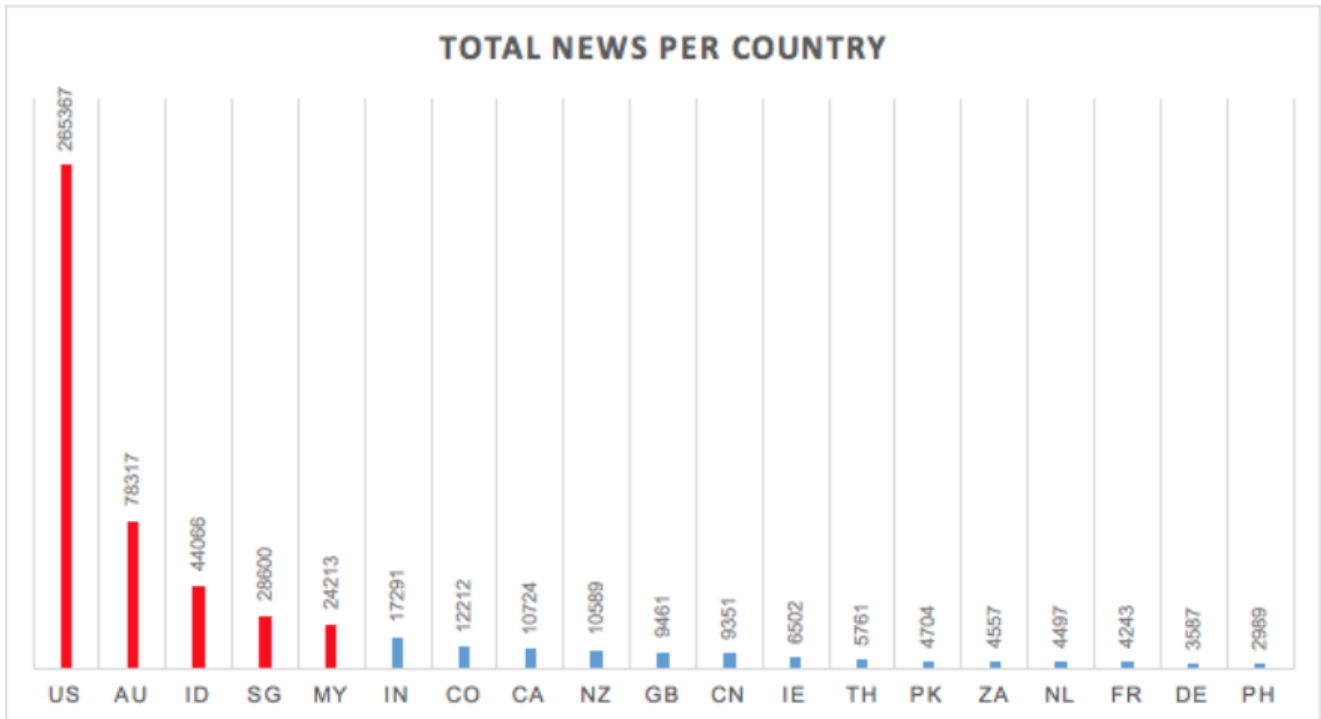


Figure 2: Total News Per Country

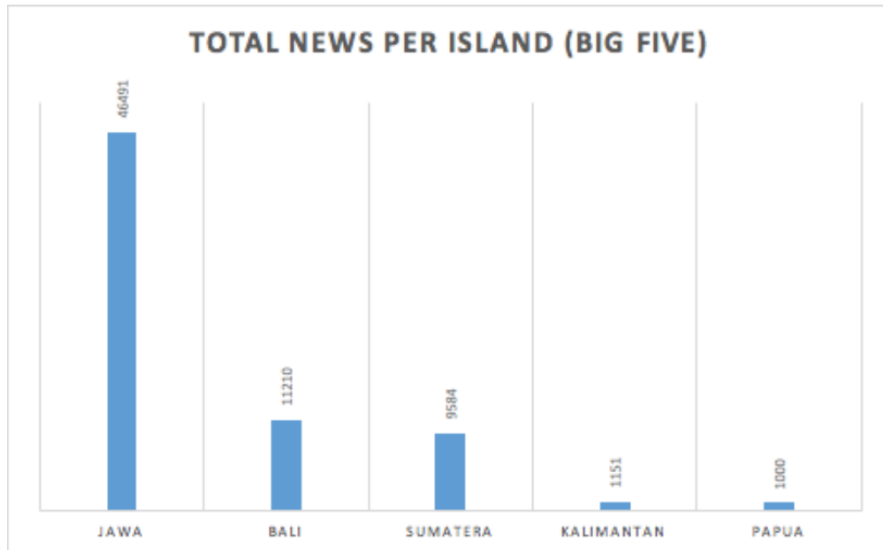


Figure 3: Total News Per Island (Top Five)

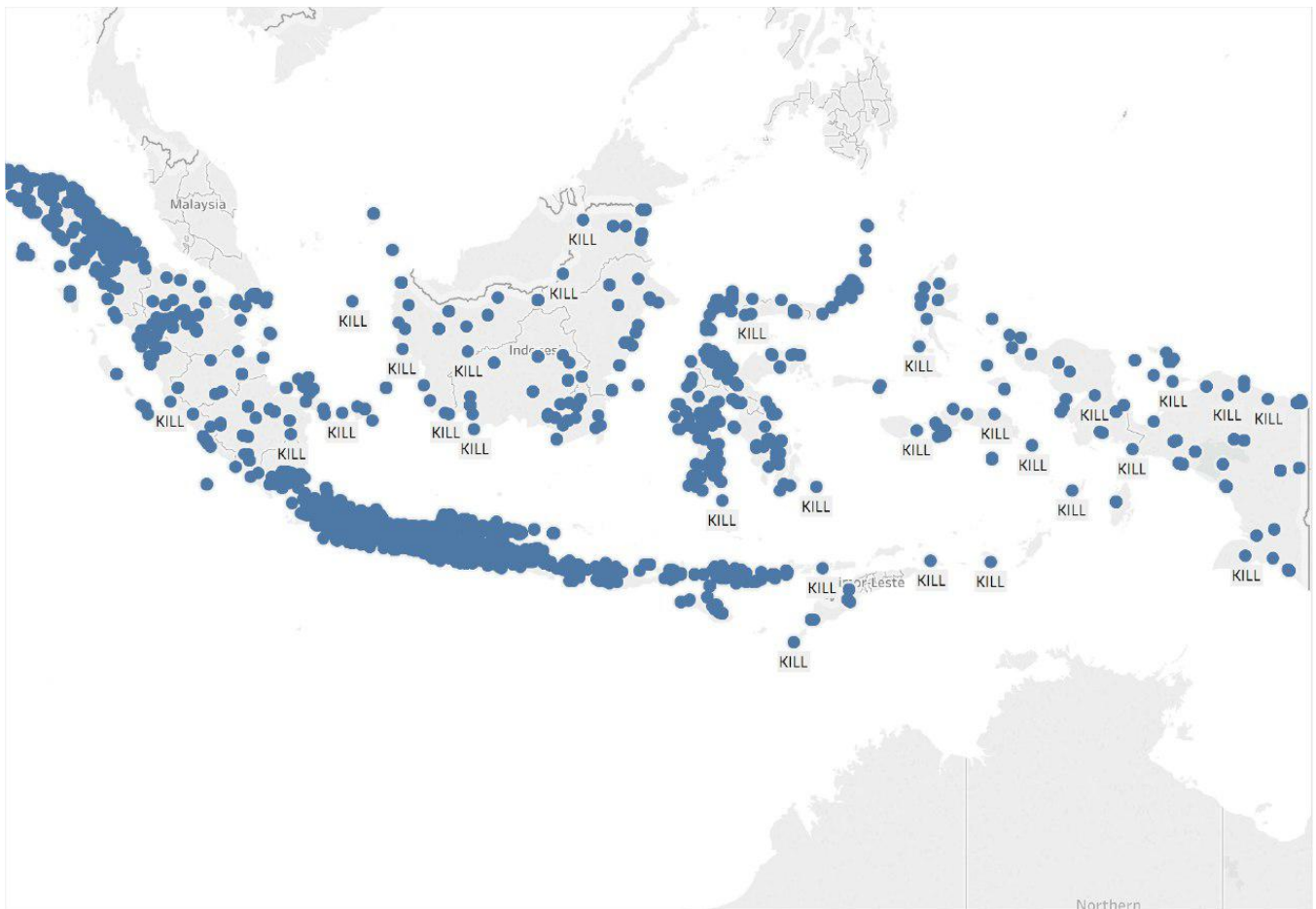


Figure 4: Indonesia News Distribution over Indonesia

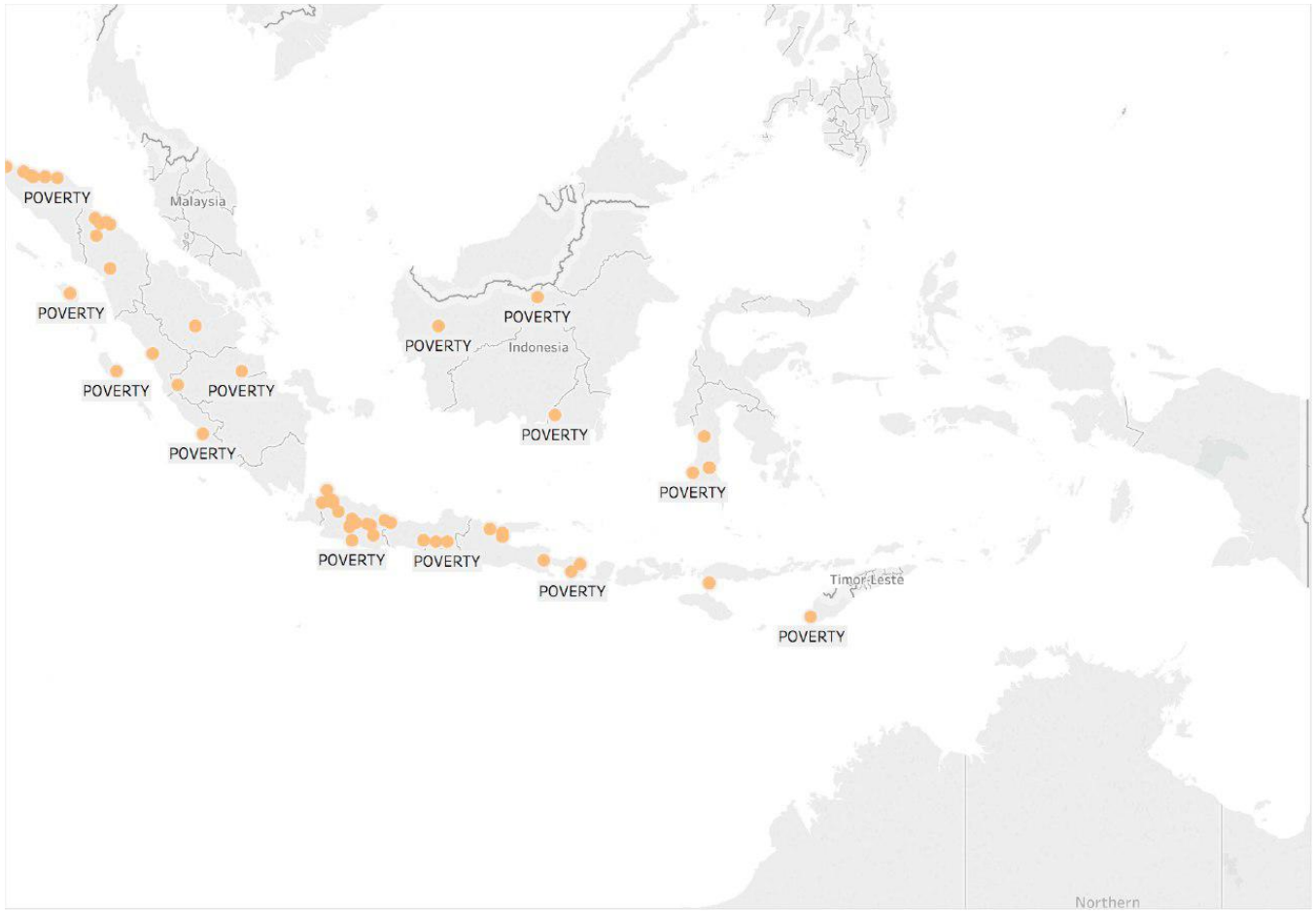


Figure 5: Indonesia News Distribution over Indonesia

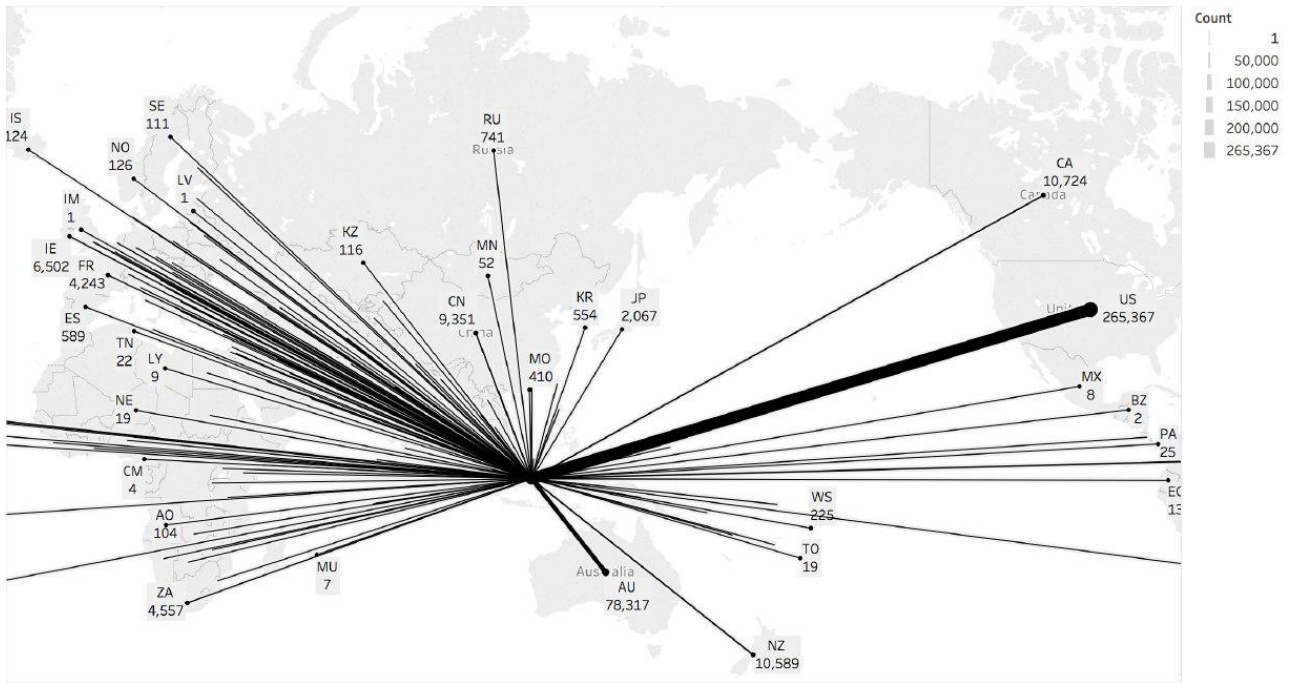


Figure 6: Indonesia News From Around the World

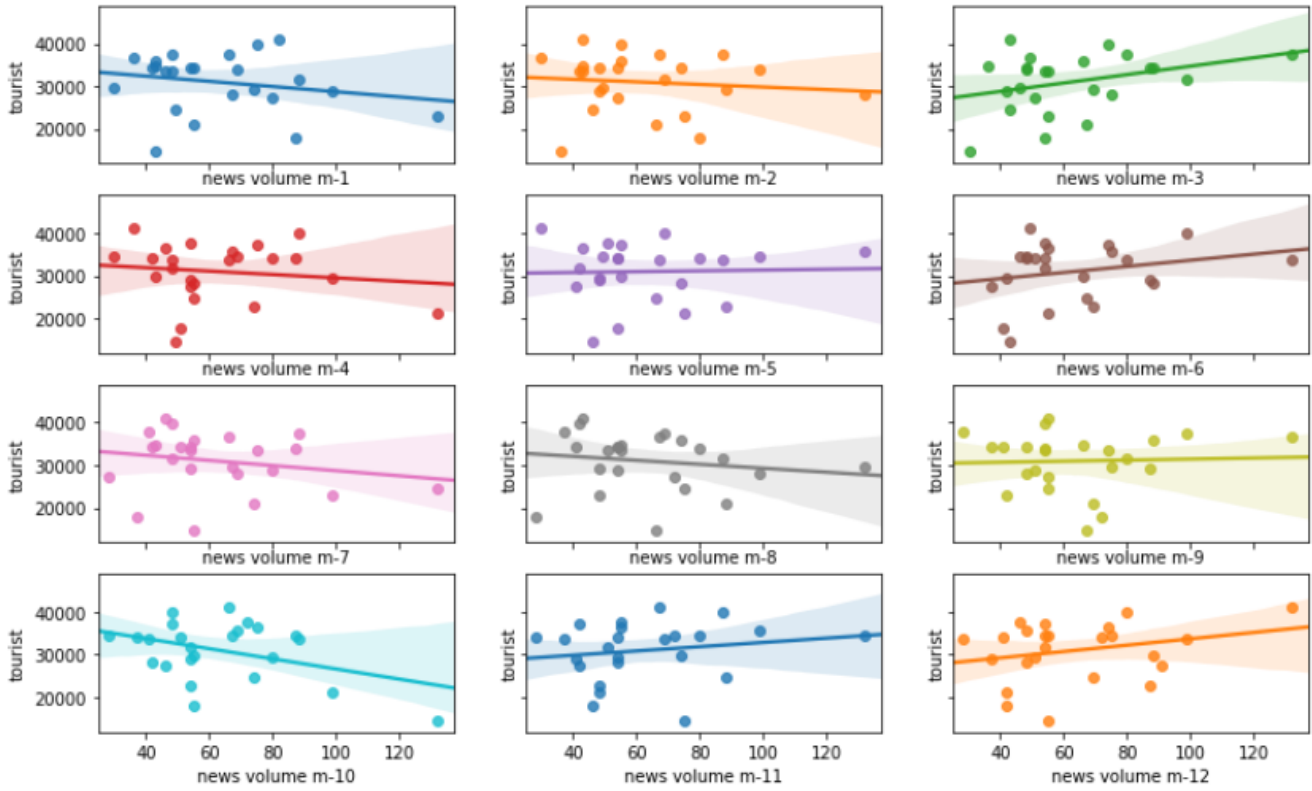


Figure 7: Bad News and Tourist Correlation