Global Affair Exploration

Jie Rong
School of Information Science
University of Pittsburgh
135 N Bellefield Avenue
Pittsburgh, PA, USA 15260
jir18@pitt.edu

Hongyuan Cui
School of Information Science
University of Pittsburgh
135 N Bellefield Avenue
Pittsburgh, PA, USA 15260
hoc27@pitt.edu

Wenxing Li
School of Information Science
University of Pittsburgh
135 N Bellefield Avenue
Pittsburgh, PA, USA 15260
wel85@pitt.edu

ABSTRACT

Political events and crises happen constantly in our world. In the highly connected world, both of them has influence on international environment and even on our daily life to some extent. Hence, it is significant for us to know more about crises and events. How many crises happen? Where these crises occurred? How serious and what accounts for these crises? We intend to find some clues to answer these kind of questions from political events. However, there are billions of event and crisis records, using traditional way to read them one by one is very inefficient. Therefore, we used information visualization tools to explore these huge quantities of data.

In this project, we designed and generated four kinds of visualizations to help analyze crises and events data from aspects of district(international/domestic), time, political sector, target and source, overview and details. We processed the data, designed visualization models, added interactions, made evaluation and made analysis. As a result, events and crises indeed have some potential relation according to our observation and analysis.

Keywords

visualization, d3.js, crisis, event, choropleth map, flow map, tile map, line chart, bar chart

1. INTRODUCTION

Our event data comes from Integrated Crisis Early Warning System (ICEWS) which is a global database of society. It contains information of event like date, the source country, source sectors, the target country, target sectors, CAMEO code, intensity, longitude and latitude of events. ICEWS project came from a desire to better understand global human society and especially the connection between communicative discourse and physical societal-scale behavior [3]. It is a political dataset. The basic objective of data visualization is to provide an efficient graphical display for summarizing and reasoning about quantitative information [1]. It is believed that any objective analysis in political science must be multidimensional and combine various sources of quantitative information; however, human capabilities for perception of large massifs of numerical information are limited [1]. Hence, methods and approaches for visualization of quantitative and qualitative data (and, especially multivariate data) is an extremely important topic [1]. News articles provide summaries of current events and include information such as what happened, when, where, and why [2]. Since both spatial and temporal information is provided, events can be analyzed in a wide variety of ways and can enable both predictive as well as retrospective analysis [2]. ICEWS as an immense amount of data is nearly impossible to understand at the corpus level, requiring immense amounts of time in order to understand the meaning of, and relationships in between events [7]. The prominence of different actors, and the geographic distribution of different stories can be answered at a greater scale [7]. However, reading the individual news stories, or even the meta data of them, is still impossible [7]. Therefore, we need to visualize data by aggregation into a map which is more reasonable to help understand events happened in the world.

Our Crisis data comes from Ground Truth Data Set(GTDS) which is a time-series dataset. In this dataset, it contains information about the number of crises for each country on each month. Ground Truth Data Set is a collection of data which lists, for the EOIs supported, whether or not the EOI did occur in any given country for any given month, historically speaking [13].

In order to visualize a dataset whose actors are countries of the world, we choose global map as the base. Map animation is particularly attractive to earth system scientists who typically study large spatio-temporal data sets [11]. In addition to the "visual variables" of static maps, animated maps are composed of three basic design elements or "dynamic variables"–scene duration, rate of change between scenes, and scene order [11]. Also, both datasets are time-series data, we need to display time as an important factor to help users to understand trend of events and crisis. Visualizing time series is useful to support discovery of relations and patterns in financial, genomic, medical and other applications [12].

In the Section 3, we discuss how to deal with data. There are two datasets and in the event dataset, we need to divide it into two parts. One is for international events, and the other is for domestic events. In the Section 4, we explain how the base map Choropleth-Flow Map works, and how users can interact with it. In Section 5, we demonstrate the line chart, which is also used as a slider. In the Section 6, we illustrate the visualization of domestic events, and give examples to analyze it. In the Section 7, we demonstrate the design of tile maps for crisis and event, and make comparisons of two maps.

2. RELATED WORK

Political event data have had a long presence in the quantitative study of international politics, dating back to the early efforts of Edward Azar's COPDAB [Azar, 1980] and Charles McClelland's WEIS [McClelland, 1976] as well as a variety of more specialized efforts such as Leng's BCOW [Leng, 1987]

ICEWS is one such system that visualizes spatio-temporal data related to political events [2]. ICEWS provides individual interfaces for visualizing different facets of data [4]. European Media Monitor News Explorer [5] provides an interface for exploring current news4. It automatically determines what is being reported in the news, location of events, people involved and what they talked about [2].

The Terrorism and Extreme Violence in United States (TEVUS) database and portal [4] provides an interactive, coordinated interface for visualizing and querying terrorist events based on the data collected from four open-source database [2].

Healthmap [6] provides an interface for visualizing spread and outbreak of diseases over time and location. The system geocodes disease related news articles and also provides timelines for individual diseases [2].

3. DATA PROCESS

3.1 Event Dataset(ICEWS)

Event data set is a very large dataset which contains nearly 1013400 events and 20 features for each month. This dataset is simplified in the table 3.1 below(S=Source, T=Target). Here we divided it into two parts: international events and domestic events. However, some rows have empty values at Target Country or Source Country, we just removed them all. Some countries like "Occupied Palestinian Territory" or "United Nation" cannot be found in the world map, thus we need to remove rows whose target country or source country are these association names.

Table 3.1

Date	SCountry	TCountry	TSectors	CAMEO	Intensity
2005-01-01	Israel	Israel	Executive Office	57	-10
2005-01-01	India	Protestors	Pakistan	112	-2

International events can be identified by the target country and source country which are different while the target country and source country of domestic events are the same. Intensity ranging from -10 to 10 works as political indicator for a certain event. Here are several steps to get the international dataset for each month Firstly, we removed all positive events which are not related to crises, because our goal is to explore potential factors resulting in crises. Secondly, we cleaned the data by removing rows containing blank values in source countries or target countries and rows whose target country or source target name could not be found in the map. Thirdly, we filtered events whose target country and source country were different. Finally, in order to analyze relation among countries on a certain month, we aggregated the dataset by year, month, source country , target country, and sum up intensity so as to form an edge like international event datasets. The summed intensity worked like a weight of edge. We found that on each month, there are nearly 300 connections among countries. This international event dataset was used to draw arcs in Section 4.

Domestic events are based on sectors of target countries, which will probably lead to crises of a country as well. First of all, we cleaned the data as international event data extraction did and then filtered events whose target country and source country were the same. We aggregated the dataset by month, target sectors, target country, and sum up intensity. Then we could get an dataset which displays summed intensity of each sector in each month. This domestic event dataset was used to draw the bar chart in the section 6.

Besides showing intensity between each pair of countries or for each sector, we also need datasets to show the overview of the intensity of a given country. We aggregated the dataset by month and target country to let the dataset can be used to draw the line chart in the section 5.

3.2 Crisis Dataset(GTDS)

This is a ground truth dataset that contains the number of 5 types of crisis in each month with date and country. The dataset is like table 3.2 below:

Table 3.2

Country	Year	Month	ins	...	ic
Iraq	2005	1	1	...	0

Here we summed all types of crisis to get crisis number of each country in each month. That was used to fill the color of each country area in the section 4.

Also, we aggregated this dataset by year, month to get an overview of worldwide crises along with time to draw the the line chart in the section 5. That worked as an overview of the world, and provided a better interaction with users.

4. CHOROPLETH-FLOW MAP

The world map is the best choice for our visualization, because all actors in events data sets are countries. If we represent each country as its real shape and real position in a global map, it would be easier for users to find the country they want. Of course, the global map has some drawbacks. One is that some districts like "Pacific Ocean" or "Antarctica" occupy large area but do not contain any useful information. Nevertheless, in order to provide users with best experience of interaction, we chose to keep the original world map.

4.1 Worldwide Crises Choropleth Map

We planned to use the choropleth map with geographic country boundaries to visualize crises density distribution around the world. Crises quantities of countries are reflected by color progression. The reason why we used color was that hue is the first caught element in a graph, and quantities of crises are the most important value to differentiate different countries. As is shown in the Fig 4.1. Users can easily find out which countries have more crises than others on a certain month. Also users can compare the crisis quantity of a country with its neighbors to detect the possible war zone. For example, we can see the color of countries in middle east is all dark, so a conflict might arise in that zone on that month. Of course, when we move the slider which is the line chart below, we can clearly see the transition of color change along with time so that detection of the crisis trend is simple. This can help us identify a certain war period or a peace period.

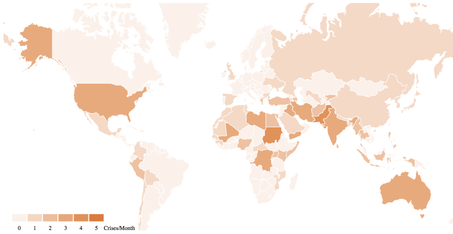

Fig 4.1

4.2 International Event Flow Map

We can classify events data according to their source country and target country as domestic event or international event. If the source country and target country of an event are the same, then this event belongs to domestic event. Otherwise, it belongs to international event. For international event, we used flow map to encode the data.

International events data shows event records happening worldwide during 2005 to 2014. There are 20 variables and over one hundred thousand records within a month. It will be a messy to display all of them in the map. So firstly, we filtered out the event records with positive event intensity. This is because we only cared about the events which have negative effect on crisis generation. Secondly, we cut off irrelevant variables, like event sentence and department. Then, we aggregated data. We combined records which have the same source country and target country, and summed their intensity into one record. This is because we did not want thousands of arcs between two countries. When the map is full of arcs, it is hard to distinguish them and get valuable information.

Event data has directivity. Each record has its source country and target country. We drew an arc from the source to the target in the world map to show the relation between the initiator of the event and the recipient. In the map, it will be displayed in the way that an arc is drawn from a source location and ends in the target location. To represent directivity, we made two ends of arcs different. For each arc, the end which connects to the source country is narrow, and the end which connects to the target country is wide (Fig 4.2) .

Fig 4.2

We used line width and color intensity to represent event intensity. The higher the event intensity is, the thicker the arc is. Moreover, the color intensity also decreases along with the event intensity (Fig 4.3).

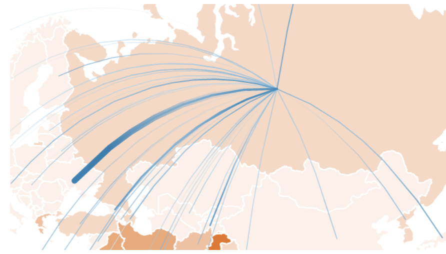

Fig 4.3

We drew the event flow map upon the crisis map, and used the same slider of the crisis map to change time(Fig 4.4). The reason is, as is mentioned before, the map is used to help users to find the relation between events and crisis. If we develop two maps, one shows event information and the other shows crisis information. Users need to switch their view back and forth between two maps. It is time consuming and inefficiency. Displaying them in the same map is more intuitive.

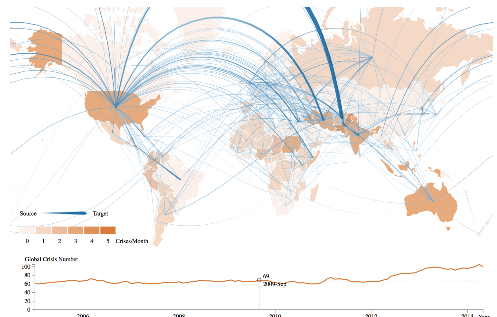

Fig 4.4

We designed a dashboard to display international event details after users select a country (Fig 4.5). The dashboard will show top 3 countries with highest event intensity, which are the target of selected events. The other top 3 countries with highest event intensity are the source of selected events. The aim of this design is to help users quickly find countries which have strongest connections to the country selected.

Fig 4.5

To illustrate how the choropleth map, flow map and dashboard work together, we choose a time point and see what happened during that time. According to our observation of the line chart (we will talk about it in the next section), we choose 2011, March. This is the peak of the global crisis number from 2005 to 2012. As is shown in Fig 4.6, there are two obvious arcs initiating from the United States. The choropleth map also shows intensive color in some countries, which means they are with high crisis rate. We can look up our dashboard to find the name of those countries (Fig 4.7). Then, by clicking the country with most arcs (in this example, the United State), we can see event information of this country. It shows that the United State initiated events, which influenced Libya and Pakistan most (Fig 4.8). The United States was also influenced by other countries (Fig 4.9).

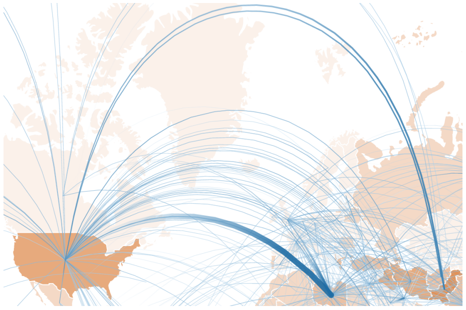

Fig 4.6

Fig 4.7

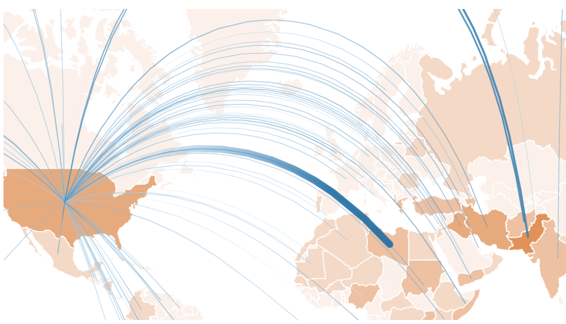

Fig 4.8

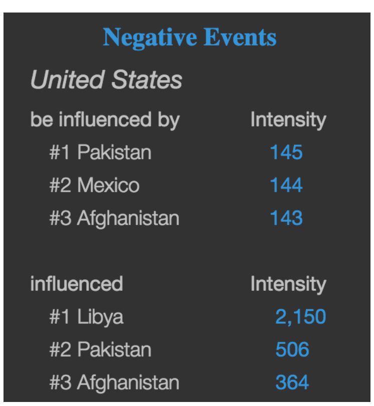

Fig 4.9

4.3 Interactive Design for Map

To make data much easier to be manipulated, we added several interactions to our map.

· Filter: It will be a little messy if users only care about one kind of data (event or crisis), because we put two visual encoding into a single map. By using switch buttons, users can turn on, or turn off the display of event arcs or map color (Fig 4.10). This design enables users to see the result more clearly if they are only interested in one kind of data.

Also, when users click on one country, arcs which are not connected to this country will be filtered out (Fig 4.11). Sometimes, there are over thousands of arcs in the map, and most of arcs are not related to the user's interest. So this function is very useful when the user has particular interest in one country.

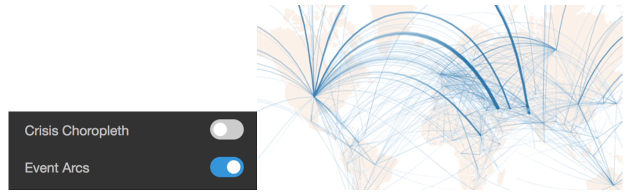

Fig 4.10

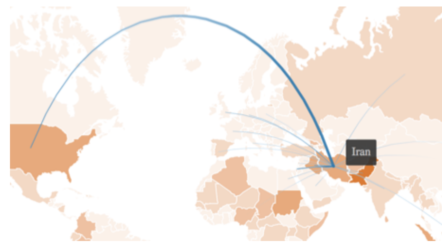

Fig 4.11

· Selection: Users can select a country by clicking it. Only arcs which are related to the selected country will be displayed, and others will be hidden. This function highlights the information that users have interest in.

Since the map data are very complex, so we hid a part of data that is not very important, and display it when users select it. For example, we hide country names. When users mouse over the country area, the country name will display. Also, we hide detailed information of arcs, like their source country name, target country name, and their intensity. When users mouse over an arc they are interested in, the information will display on tooltip (Fig 4.12).

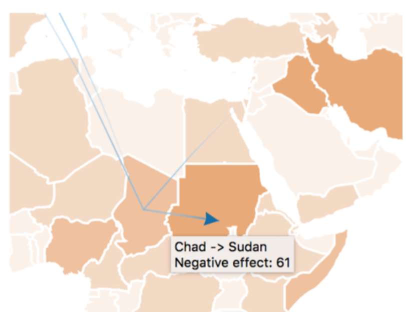

Fig 4.12

· Organization: The dashboard on the left is designed to display switch buttons and result data. It has three main parts. Two switch buttons are on the top. Crisis data is displayed in the middle, a list of top 5 countries with highest crisis rate is shown in this part. Event data is displayed in the bottom, there are two tables which show event details of the selected country.

· Navigation: The map can be zoomed in and out to see a larger view of crisis or event information. Also, in the crisis part of dashboard, top 5 countries are links. Those links can navigate to domestic events visual encoding of selected country.

· Sorting: The dashboard displays statistical information of crisis and event data. Those information is sorted and filtered, the dashboard will only show top countries. This is because the space of dashboard is limited, and displaying all of the data is unnecessary. The top countries' data affect the overall picture most.

· Coordination: By associating the line chart with the map, when users select a time point, not only the line chart will show its statistical data, but also the choropleth map and flow map will update data according to that time point.

4.4 Pre-attentive Design for Map

We applied pre-attentive knowledge to our designs.

For color, we chose blue and orange to distinguish event and crisis data. Information related to event is blue, and information related to crisis is orange. Blue and orange are complementary colors, when they are placed next to each other, they create the strongest contrast. In this way, users can distinguish two types of data effortlessly. Also, since the number of distinct data components that people can process in short-term memory are limited, there is only four color in our map encoding:

· Dark grey for background;

· Light grey for texts;

· Orange for crisis information;

· Blue for negative information;

For forms, we used line width to emphasize events with high intensity. We also arranged event arcs in descending order, which means arcs with wider line width and higher color intensity are closer to the user while arcs which are thin and light are far from the user. This makes important messages can be seen first and clearly.

5. LINE CHART

Both event data and crisis data are time-oriented data, thus we need a slider to let users change the time. However, we did not choose traditional slider bar, instead, we used a line chart.

Generally, a time slider bar only has horizontal axis, which displays the time, and a pin which can be dragged to select time. The problem of this traditional slider bar is that if users do not have a particular time to look at, they just want to get some valuable information from the data, they will find it difficult to explore. For example, a user wants to find a month with peak crisis rate, and see what events happened on that month. By using the traditional slider, the user need another graph or table to help him/her find the month, and then uses slider bar to update the data on that month. This process is very tedious when the user has much information to explore.

Therefore, we integrated the line chart and the slide bar (Fig 5.1). The horizontal axis of the line chart stands for months from 2005 to 2014. The vertical axis stands for global crisis number of each month. The data line shows worldwide crisis trends from 2005 to 2014. To update the map's data, after users click the line chart, the map will get that month's data according to the vertical coordinate of clicked location.

Since there are 120 months in the horizontal axis of line chart, months are too close to each other. This arrangement increases the difficulty of selecting time and looking for corresponding crisis value. Here is our solution. When users mouse over the line chart, there will be a vertical dotted line and a horizontal dotted line. They intersect at the mouse pointer (Fig 5.2). Those lines will help user locate the x and y coordinate, and compare crisis value with others. Also, there are texts to display crisis number and date.

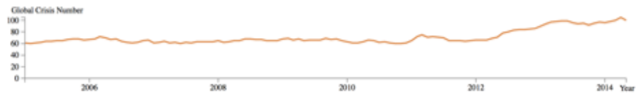

Fig 5.1

Fig 5.2

6. DOMESTIC EVENTS

In order to have an insight of countries with high crisis rate, we choose the top 5 countries with the most crises per year to make the visualization of domestic events for these countries. From 2005 to 2014, there are 21 countries in total. On the webpage of the Choropleth-Flow Map, each time a user chooses a point of date on the line chart, the top 5 countries with highest crisis rate will be listed on the manipulation panel (Figure 6.1). Users can click any country's name to jump to the webpage of that country (Figure 6.3) to learn detailed information from the visualization of domestic events.

Fig 6.1

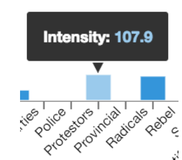

Fig 6.2

Since we intended to present the negative aspects of domestic events, which are more potentially relevant to the occurrence of crisis, we screened out events with positive intensity, so events in this section refer particularly to domestic events with negative intensity. The visualization of each country is composed of a line chart and a series of bar charts. We used the absolute value of the negative intensity to display the political indicator of events.

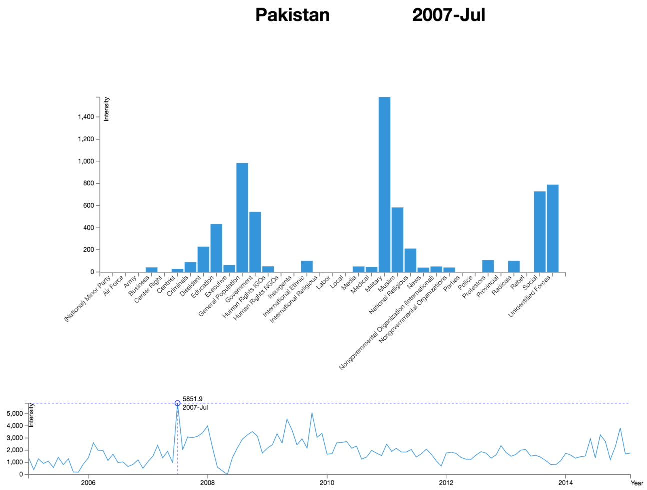

Fig 6.3

6.1 The Line Chart of Domestic Events

The line chart provides an overview of the trend of domestic events in a certain country from 2005 to 2014. The peak and valley are clearly shown on the line. Meanwhile, the line chart is also the time slider of bar charts. Users can select any point of the line to see the related data (the date and specific intensity), and the bar chart of corresponding date will be displayed. Meanwhile, the selected date will be listed on the top of the webpage.

For example, on the Figure 6.3, the selected country is Pakistan, and selected date is July 2007, a peak point. In the real world, Pakistan troops stormed the militant-occupied Red Mosque complex in Islamabad on July 10,2007 and following a week-long siege [14]. The line chart reflects the events in the country distinctly and explicitly.

6.2 The Bar Chart of Domestic Events

The bar chart provides detailed information of domestic events. It shows the distribution of events among different sectors in a certain month which is selected on the line chart. The horizontal axis represents the target sector such as "army" and "education". By moving mouse over a bar, the tooltip (Figure 6.2) will appear to show the event intensity of that target sector. From the distribution of target sectors, we can know which aspect of the country may have potential problem or crisis. The intensity indicates how severe they are affected by the event. Let us see the same example in 7.1. From the bar chart on the Figure 6.3, the "Military" and "General Population" are the top 2 affected sectors. In the history, this event led to the death of up to 50 militants, eight soldiers [14] and three security personnel [15], which is consistent with the bar chart.

7. TILE MAP

In order to observe the variation trend in the dimension of time, we designed tile maps to visualize global crises and events respectively. Both event data and crisis data are time-oriented data from 2005 to 2014. We counted the sum of crises happened in the world by month; aggregated the dataset of events only by month, and summed up intensity. It was arranged as one month per unit, 12 months per row (per year), and ten years listed in column. Crisis quantities and event intensity are reflected by color gradient. For global crises, the sum increases as the color turns from green to red. For global events, negative intensity increases as the red turns to yellow; positive intensity increases as light green turns dark. By moving mouse on a box, the tooltip will appear to show the sum of crises/event intensity in the selected month. Out of the aim of finding potential relations between events and crises, we put two tile maps side by side in one web page (Figure 7.1).

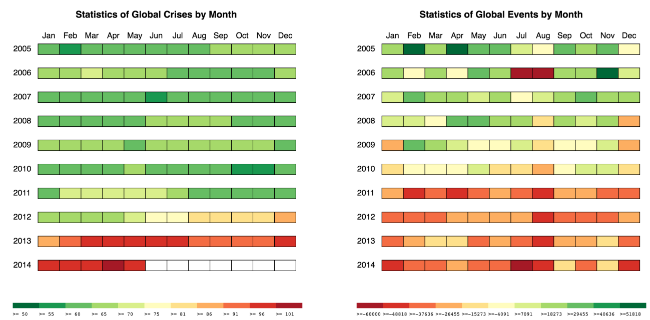

Fig 7.1

From figure 7.1, we found two maps have consistent variation trends. The sum of global crises grows along with the time. correspondingly, the intensity of global events drops constantly, which means the influence of global events becomes more and more severe and negative year by year. The variation of the events map is about two years ahead of the crises map. Negative events seem to have delayed influence on crises, but there are exceptions, such as July and August in 2006. A reasonable cause of exceptions might be that only persistent negative events can affect crises. This is just assumptions from this visualization, however, the result of observation and comparison provides a direction of future analysis work. Further works need to be done if we want to figure out the exact correlation between crisis and event.

According to the crises map (Figure 7.1), it is easy to find the darkest red box represents the month April 2014, which has most quantities of crises in the world from January 2005 to May 2014. This month is a peak on the line chart on the Choropleth-Flow map (Figure 4.4), which is consistent with the result of crises tile map.

8. CONTRIBUTION

There are three team members in our team. We worked together to find the best way of designing all visualizations, and finished coding parts separately. Here is the contribution of each member.

Hongyuan Cui

- Provided ideas for all visualizations;
- Data processing;
- Completed the worldwide crisis choropleth map;
- Completed the Line Chart as a slider;
- Completed final paper.

Jie Rong

- Provided ideas for all visualizations;
- Completed the international events flow map;
- Completed the dashboard;
- Completed the coding of web page;
- Completed final paper.

Wenxing Li

- Provided ideas for all visualizations;
- Wrote the progressive paper;
- Completed domestic events visualization;
- Completed tile maps for crisis and event;

- Completed final paper.

9. FUTURE WORK

In this project, we designed and accomplished choropleth map, flow map, line chart and bar chart. We used these visual encoding to visualize events and crises data and make connections between them. However, our designs still have room for improvement and the dataset we used also has more aspects can be analysis than we already have done. So we present some ideas for our future work in this section.

First of all, we can change 2D map to 3D, which solves the problem that some arcs are interrupted at the edge of the map. Secondly, we can develop a text visualization to process event description message. For example, the word cloud, by which the frequency of word can be displayed apparently, thus users can learn the keywords of the event. Usually the most frequently appearing words reflect the most critical aspect of a crisis or event. Thirdly, we can use spiral graph to observe whether the event data and crisis data have periodicity.We have this propose is because we find that many peak value of crisis number appear around the March.

10. REFERENCES

[1] Zinovyev, A. (2010). Data visualization in political and social sciences. *arXiv preprint arXiv:1008.1188*.

[2] Saraf, P., Self, N., & Ramakrishnan, N. Who, When, Where and Why? Visualizing Civil Unrest Events.

[3] http://www.gdeltproject.org/about.html

[4] G. LaFree and W. Braniff. Project fact sheet: Terrorism and extremist violence in the U.S. database and portal. 2014.

[5] R. Steinberger, B. Pouliquen, and E. van der Goot. An introduction to the europe media monitor family of applications. In *Proceedings of the 2009 SIGIR Workshop*, SIGIR-CLIR, pages 1–8, July 2009.

[6] C.C.Freifeld,K.D.Mandl,B.Y.Reis,andJ.S.Brownstein.Health map: global infectious disease monitoring through automated classification and visualization of internet media reports. *Journal of the American Medical* Informatics Association, 15(2):150–157, 2008.

[7] Wu, L., Morstatter, F., Baird, D., Sampson, J., Shu, K., & Liu, H. Integrating Interactive and Computational Approaches for News Understanding.

[8] Arva, B., Beieler, J., Fisher, B., Lara, G., Schrodt, P. A., Song, W., ... & Stehle, S. (2013, July). Improving forecasts of international events of interest. In *EPSA 2013 Annual General Conference Paper* (Vol. 78).

[9] Ward, M. D., Beger, A., Cutler, J., Dickenson, M., Dorff, C., & Radford, B. (2013). Comparing GDELT and ICEWS event data. *Analysis*, *21*, 267-297.

[10] Leetaru, K., & Schrodt, P. A. (2013, April). Gdelt: Global data on events, location, and tone, 1979–2012. In *ISA Annual Convention* (Vol. 2, No. 4).

[11] DiBiase, D., MacEachren, A. M., Krygier, J. B., & Reeves, C. (1992). Animation and the role of map design in scientific visualization. *Cartography and geographic information systems*, *19*(4), 201-214.

[12] Aris, A., Shneiderman, B., Plaisant, C., Shmueli, G., & Jank, W. (2005, September). Representing unevenly-spaced time series data for visualization and interactive exploration. In *IFIP Conference on Human-Computer Interaction* (pp. 835-846). Springer Berlin Heidelberg.

[13] https://dataverse.harvard.edu/dataset.xhtml?persistentId=doi:10.7910/DVN/28119

[14] http://news.bbc.co.uk/2/hi/south_asia/6286500.stm

[15] http://www.nytimes.com/2007/07/10/world/asia/10pakistan.html