

Visual Analysis of African Conflicts

Thomas (McNair) Bledsoe, Amit Pandit, Rajkumar Selvam, Renee Sikes

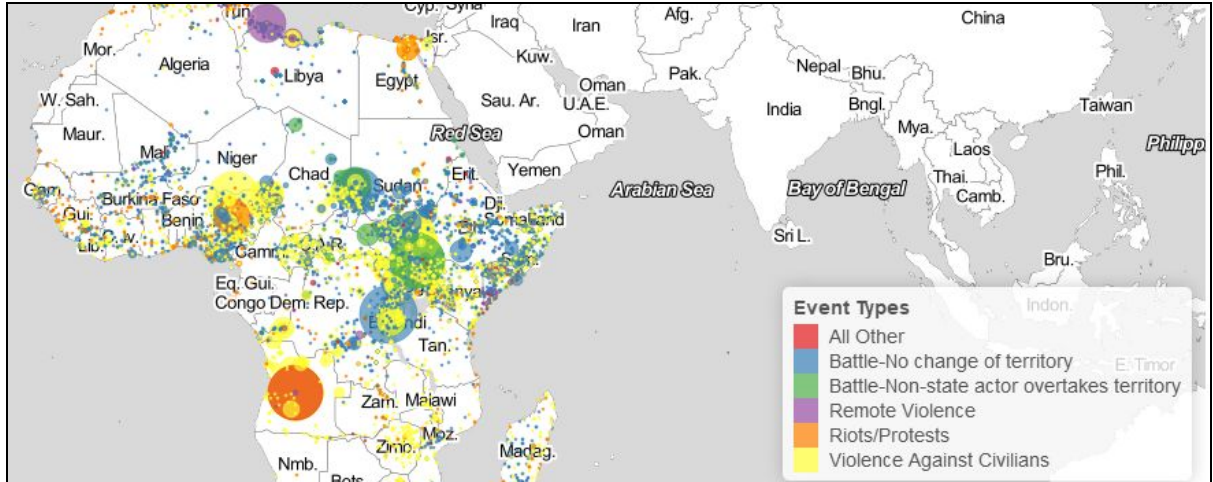


Fig. 1. Screen cap of the leaflet map used in the project application

Abstract—The world is constantly shifting borders. There are armed conflicts around the globe whether it is between governments or the people the governments control. With the excitement of the power of visual analysis within the tech community, easy to manage tools have become accessible to anyone willing to spend a few days learning them. The goal of this visual analytics project is to harness those tools to present a friendly UI that allows users to browse and understand conflicts that have occurred in Africa over a 20 year period. Using the R programming language and shiny, a dashboard was created that allows interactivity with data holding conflicts in Africa over time with varying degrees of granularity on a geographic medium..

1 INTRODUCTION

Data used within this project comes from the ACLED, Armed Conflict Location and Event Data Project. ACLED collects the dates, actors, types of violence, locations, and fatalities of all reported political violence and protest events across Africa, South Asia, Southeast Asia and the Middle East. Political violence and protest includes events that occur within civil wars and periods of instability, public protest and regime breakdown. ACLED's aim is to capture the forms, actors, dates and locations of political violence and protest as it occurs across states. The ACLED team conducts analysis to describe, explore and test conflict scenarios, and makes both data and analysis open to freely use by the public [1]. The goal of this project is to use data [2] from the ACLED to create a visual dashboard that allows users to interact with the data in a geographic medium with varying degrees of granularity. Inputs can be given to the system and the system will adjust the data displayed to allow for more detailed exploration. A link has been posted online for viewing [3].

2 RELATED WORK

2.1 ACLED

Most of this information is already supplied in the introduction. This inspiring project gives us the ability to see trends of conflict throughout the war torn areas of the world. The monthly reports that

come from the project also give the team plenty of ideas of how we can work with this valuable information. [1]

2.2 Research on Spatio-temporal Data Visualizations

2.2.1 Spatio-temporal Data

The ACLED data set is predominantly a time-series dataset, which collects information about events over time in specific geographic regions. There are many related research papers that explore ways to effectively visualize temporal data, spatial data, or both. We examined several related research papers to gain insights into other methods that have proven successful.

Many researches have been primarily focused on these issues of storage and management of time-referenced geographic data in geographic information systems. The referenced article [5] is one of such research, but it is neither focused on different approaches to internal data representation nor have they compared software packages according to the representational frameworks they incorporate. Instead it focuses more on the nature and inherent properties of spatio-temporal data. These concepts and the guidance to approach the spatio-temporal data has helped this project to a greater extent.

There is a broad classification of spatio-temporal data according to the kind of changes occurring over time:

1. Existential changes, i.e. appearance and disappearance.
2. Changes of spatial properties: location, shape or/and size, orientation, altitude, height, gradient and volume.
3. Changes of thematic properties expressed through values of attributes: qualitative changes and changes of ordinal or numeric characteristics (increase and decrease).

Another research paper created a proprietary system that uses a T-pattern analysis method which identified temporal co-occurrence in events by identifying cases where the time in between two events are statistically significant, meaning they occur more often than would be expected by random chance [8]. Once the co-occurrences are extracted, they are visualized on a map with colored concentric rings representing highly associated events, based on user selections. The t-pattern results are also used to filter the rest of the visualizations in the tool, including a timeline, a word tag cloud and map views. Given the existing data set, which includes events that occur between two actors in a specific spatio-temporal location, this method would be an interesting way to explore the data because it would show an abstracted view of the associations within the data set, rather than just a visual representation of the summarized data points. While there would undoubtedly be insightful information that could be gathered from this approach, we deemed it too complex for the scope of our research tasks.

2.2.2 Information Extraction

There are several researches conducted about how the information can be extracted from the data in a much effective manner. For this project, we needed a typology referring more to the initial stage of data analysis and encompassing potential information needs. The information needs are nothing but questions that we would like to answer through visualizations. There are as many types of questions as components in the information. For example, our data set with fatalities by year contains two components, year (from event date) and the actual fatality count. Respectively, two types of questions are possible:

- On a given year, what is the total number of fatalities?
- For a given fatality count, on what date(s) was it occurred?

For each question type, according to Bertin [4], there are three levels of reading: elementary, intermediate and overall. The level of reading indicates whether a question refers to a single data element, to a group of elements or to the whole phenomenon characterized by all elements together. Overall, there are three distinguished characteristics in the spatio-temporal data: space (where), time (when) and objects (what). Accordingly, three basic kinds of questions are possible:

- * when + where → what: Describe the objects or set of objects that are present at a given location or set of locations at a given time or set of times.
- * when + what → where: Describe the location or set of locations occupied by a given object or set of objects at a given time or set of times.
- * where + what → when: Describe the times or set of times that a given object or set of objects occupied a given location or set of locations.

2.2.3 Different techniques and where to use it:

There are various computer-enabled techniques for representation and exploration of spatio-temporal data from the perspective of their applicability to different types of data. The below is the summary of the techniques into four categories, and their applicability:

1. “Universal” techniques, i.e. applicable to all data types: querying (lookup and filtering), map animation and map iteration.

2. Techniques suitable for data about existential changes: time labels, representation of the age by color, aggregation of data about events and space-time cube.

3. Techniques applicable to data about moving objects: trajectory lines, arrows, “tracing”, time labels, space-time cube and different animation modes, i.e. snapshot in time, movement history and time window.

4. Techniques for studying thematic (numeric) changes: change map, time-series graph and aggregation of attribute value

3 System Design

The project is designed in the R programming language using shiny dashboard as a layout and visualization tool. Code and resources were managed in a git repository, BitBucket. The goal of the project was to provide a visual exploration of the conflicts in Africa over a 20 year period. This was done using a map, dygraph, word cloud, data table, and parallel set. Each visualization has its own set of user interactions that supply more information to the user depending on the interaction. User inputs were supplied on the left of the dashboard so that the user can filter data further into areas of interest. These inputs filter all visualizations on the dashboard. An important note is that the dygraph’s time bar can be used to change the time frame filtered on all other visuals including its own. With inputs and visualization interaction, a user should easily be able to filter into information they want to find.

3.1 Tasks

Explore the prevalence of certain conflict event types on the African continent. Understand how the frequency of certain event types have fluctuated over the last 20 years. Understand the number and type of conflicts that occur in different regions of Africa. Summarize the number of fatalities associated with conflicts in certain regions of Africa

3.2 Data Cleaning and Prep

The data used by shiny requires to be UTF-8 encoded. Before loading the data it is important to ensure this is the encoding used on any csv file is in the UTF-8 format or else the program will be unable to run. In File, Save with Encoding... enables a user to set the default encoding that R Studio saves files with.

From the initial dataset of 28 columns, we filtered columns that were not relevant, and kept a base of 15 columns. These columns included values such as the names of the two actors involved in each conflict, the administrative region of the country in which the event occurred, the country name, the date of the event, the type of event, the number of fatalities, latitude and longitude, the location name, i.e. town or city, some textual notes summarizing the event, and the year the event occurred. We also filtered records where Actor1 was made of government forces like ‘Military’ and ‘Police’ and filtered out the NA values, since these were not categorized as conflicts. This left us with 24,259 records. We also added couple more calculated columns for ease of computations. For example, we broke the number of fatalities into a custom range of fatalities: 1, 2, 3-5, 6-10, 11-20, 21-30, 31-40, 41-60, 61-100, 100-500, >500. These ranges were chosen after examining the distribution of the total number of fatalities per event. We also created a column that assigned each country to the appropriate region of Africa to which it belongs such as: Northern, Southern, Eastern, Western, and Central.

3.3 Visualizations and Code

All visualizations are supplied by the shiny library. There is a map, dygraph, word cloud, data table, and parallel set. User are able to interact with the visualization through inputs in the sidebar on the left as well as the dygraph time bar. An important feature in the code

is the ability to filter data according to user inputs. This was made possible by use of a reactive expression in the code. Data is initially filtered by the three inputs on the left of the dashboard. Further details can be found in the following sections. When a selection is changed within these list input controls, an event is broadcasted that causes the reactive expression to change what data is filtered. Any visual using this reactive expression will update its data frame used and then update its output. Another separate reactive expression is used to then filter on the time chosen within the dygraph visualization. Reactive expressions are the key to the functionality of this visual application.

3.3.1 Input

Input consists of three multi-select list input control and a time frame bar located below the dygraph as seen in fig. 2 and fig. 3. All four inputs filter the data for all visualizations seen in the main area of the dashboard. Multiple selections can be made in each select box creating a large number of possible combinations that allow for a customized exploration of the data. Event Types consists of Battle-No change in territory, Battle-Non-State actor overtakes territory, Remote violence, Riots/protests, Violence against civilians, and Other (Battle-Government regains territory, Strategic development, Headquarters or base established, Non-violent transfer of territory.) Regions consist of Northern, Western, Eastern, Southern, and Central Africa. Actors consists of the primary actors. This allows for users to analyze individual belligerents and their actions across the country. All list boxes use autocomplete for searches, which allows Actors to be useful in searching for individual belligerents. The time bar in the dygraph allows the user to choose a range of time on which to filter the data including a start date and end date. Fig. 2. shows the dygraph limiting the from data and leaving the to date at its limit. This allows a user to note spikes in fatalities in the dygraph and to limit displayed data to just that time to deeply analyze a period of time. For example, a user can use the notes description in the map popup to view what battles happened and where within the chosen time period to see what group or battles caused the spike. The user can then use that information to analyze the group further by selecting the actor in the list box.

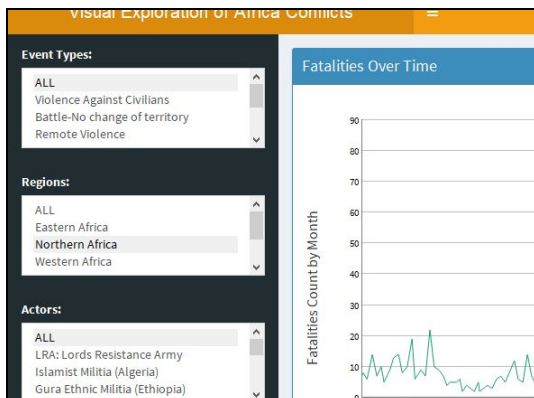


Fig. 2. User input as seen on the left of the dashboard.

3.3.2 Map

See fig. 1 for a screencapped map visual in use. The map has a few important features to be noted and can be considered the focal point of the project. It is the area where the user will gather the most valuable amount of data and most likely will focus on during use. From just one initial glance, the user may be able to tell what kind of major events have occurred. It plots the data points using the leaflet library, a d3 javascript tool made available to the shiny library.

Latitude and longitude are used to plot each individual event to the map. The provider tile, which is the map used for plotting, is a

provider called Stamen.TonerLight. There are many different maps provided for free that can be used for point plotting. TonerLight was chosen for its neutral grayscale color tone that would be useful for plotting colored markers onto the map. Circle markers, or the plotted points, are given a popup for when it is clicked, a radius, and a color from a palette. The radius is directly correlated with the number of fatalities in an event. The color palette is mapped to event types which is shown in a legend on the bottom left of the map. The color of the circle marker shows what event type the plotted point represents. The pop up shows notes, location, fatalities, and other information important in understanding individual events. The pop ups provide detailed data that can aid in understanding the data and finding new items to explore. The markers on the map change depending on the user chosen inputs. Zooming is supplied by default for a leaflet map.

3.3.3 Dygraph

The dygraph is another major feature within the application.

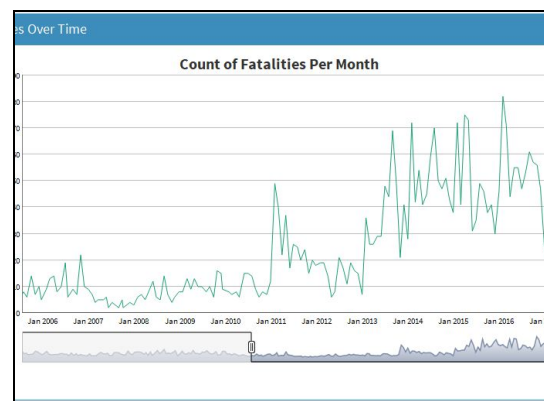


Fig. 3. Dygraph with time frame input bar.

The dygraph initializes by displaying the count of fatalities per month over the 20 year period of data, for all events in the dataset. The user inputs will allow the users to change the granularity of the views to show more targeted information, such as the fatalities over time within a specific region, or associated with a specific event type such as Violence Against Civilians.

The dygraph also includes a timeline slider which allows the user to change the time range of the visuals. Any interaction with this timeline, or by brushing within the graph will change the outputs of the other associated visualizations.

3.3.4 Word Cloud

Due to computational requirements to produce each word cloud, the word cloud feature is executable on demand by a button.

The word cloud is generated using the term frequency, inverse document frequency (TFIDF) method of text analytics using the Notes as in corpus. The notes included in the corpus will adjust depending on the other user provided inputs, previously described. The word cloud adds a new dimension to the analysis providing a more in depth understanding of what is happening during the events within the selected input ranges. It guides the user toward an understanding of what is happening, not just when and where.

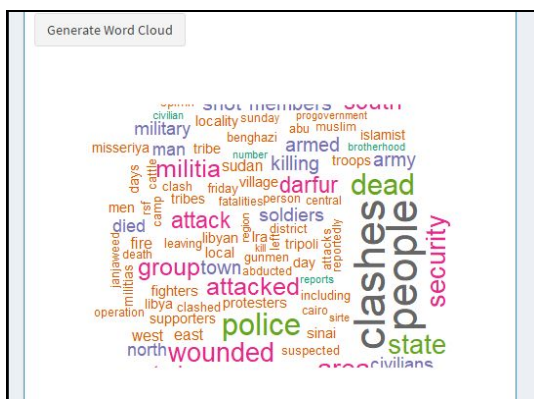


Fig. 4. Generated WordCloud as seen on the dashboard.

3.3.5 Data table

The data tables helps the user to understand the question of who is involved in the events.

Search: <input type="text"/>		
ACTOR1	ACTOR2	Count
Militia (Pro-Government)	Civilians (Sudan)	151
LRA: Lords Resistance Army	Civilians (Sudan)	131
Rioters (Egypt)	Rioters (Egypt)	94
Janjaweed	Civilians (Sudan)	87

Fig. 5. Data Table on the top left of the dashboard.

The data table dynamically filters based on other user inputs and displays the two actors involved in the conflicts along with a summarized count of the number of times those two actors registered a conflict, within the given parameters. This visualization is useful to identify who the major players are within a given set of criteria.

3.3.6 Parallel Set

The parallel sets visualization allows us to explore the magnitude of interactions between categorical variables in our dataset.

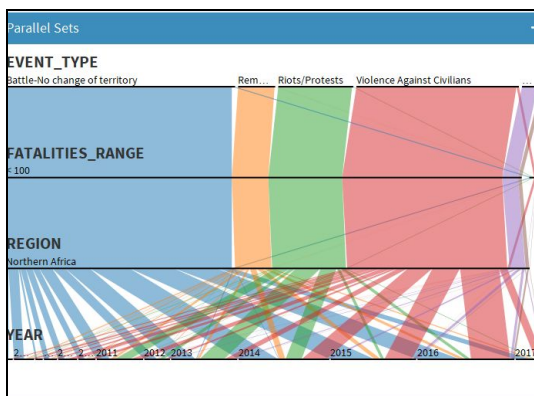


Fig. 6. Parallel Set as seen on the bottom right of the dashboard.

As shown in figure 6, we used 4 variables in the parallel sets, including the event type, the calculated column summarizing the fatalities range, the region, and the year. This visualization uses the default interactive parameters that allow the variables to be reordered by dragging and dropping the rows. It also allows the user to resort the ranges alphabetically or by size.

4 ANALYSIS

Due to the size variation of the data set, there are many ways a user can explore the data depending on the types of questions they are interested in answering. While it is not possible to provide a comprehensive list of interesting findings within the dataset, we will highlight a few findings of interest.

The default initialization of the dashboard already provides users with some interesting questions that could be explored further. An initial review of the map shows that the majority of conflicts are concentrated in the central and eastern regions of Africa, followed closely by the west. The north and the south appear to be much more stable with fewer incidents, and those that do occur are concentrated within specific regions of the countries. We can also see that there are several large events resulting in high fatality counts, given by the radius of the circle. By clicking on each of these markers we can see that there have been a high number of fatalities associated with riots and protests in Angola, remote violence initiated by NATO in Libya, and conflicts between Muslims and Christians in Nigeria.

The regional concentration of the conflicts is also further supported with the parallel sets, which shows the largest blocks on conflicts in Eastern Africa. We can also see that the most common type of event is Violence against civilians followed by battles in which there was no change in territory. Also, from the parallel sets we can see that the majority of events result in a total fatality count of less than 100.

From the actors table, we can see that Al Shabaab and Boko Haram are the most active Actors aggressing against civilians. We can also see that the biggest hotspots of activity are Somalia, Nigeria, and Uganda.

We can see from the digraph that during the periods of 1997 to 2010 the number of fatalities appeared to be fairly steady, after which there was a sharp increase in the number of fatalities.

We were interested in further exploring the events that were contributing to the high spikes, in particular during the period around 2016. By adjusting the timeline of the dygraph, we can focus the rest of the visualization on that timeframe. After filtering the time range, the data table shows us that the top actor of that time was Al Shabaab with 266 recorded incidents with the Nigerian civilians. The word cloud tells us that there were attacks that resulted in killings. Other frequent words such as dead and civilians also support these findings.

Knowing the Al Shabaab appeared to be a very active player on the African continent contributing to a large number of events and fatalities, we wished to further understand their activities throughout the 20 year period of available data. After resetting the visualization to the default view, we then selected Al Shabab from the Actors input list. This shows a count of monthly fatalities per month ranging from about 2 to about 37 over the last 20 years with notable spikes consistently around January and then again in the summer months such as May and July. The 2007 and 2009 years were comparatively quiet. From the data table, we can also see that most of the events occurred between Al Shabaab and the Somali civilians, with 812 events. There were also 320 clashes with the Somalia military forces. We can also see that the group did not operate exclusively within Somalia borders, extending their reach into Kenya with 92 events and 63 conflicts with Ethiopian military. This can also be seen in the map in figure 7.

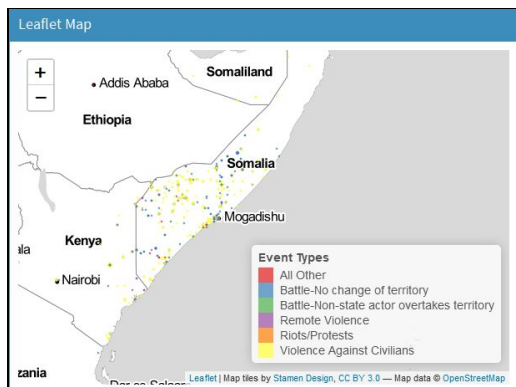


Fig. 7 Leaflet map displaying conflicts involving Al Shabaab

The parallel sets shows us again that most the events were violence against civilians followed by battles with no change of territory and then remote violence. The size of the markers on the map indicate that most of the individual events had a small number of casualties, with all events falling in our bucket of less than 100. We can also see that most of the violence against civilians occurred between 2012 and 2016, while the battles for territory occurred in earlier years between 2008 and 2012. This suggests there may have been a change in focus or tactics from the militant group.

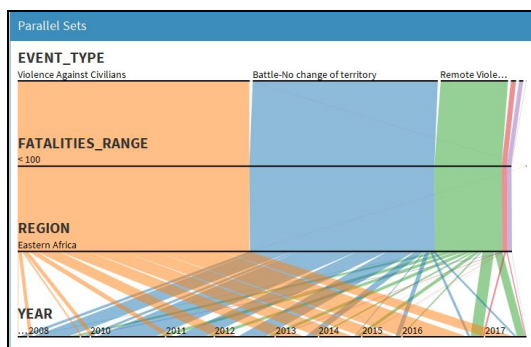


Fig. 8 Parallel sets showing distribution of events, fatalities, regions, and years of conflicts involving Al Shabaab.

5 TEAM MEMBER CONTRIBUTION

All team members contributed a fair level of technical and organizational value. Amit Pandit joined a week after forming and shortly convinced the team to use an R shiny application and gave an overview of its initial use. Renee supplied the data source and the team collectively collaborated on the tasks to be performed on the data set. Renee supplied excellent organizational and decision making skills. She supplied quick changes in the application when there were changes in vision of layout and visual usage. Rajkumar and Renee both polished and cleaned the final version of the application to be presentable at the end of the cycle in bugs as well as visual appeal. Rajkumar created the code repository for team use and instructed on its use as well as forming code structure and major features. Amit supplied multiple ideas as well as working code samples for the team to consider or use. He also did data set cleaning for use in visuals. McNair contributed to organization, decisions, code features, bug fixes, and reports. Overall, the team worked very hard to pull their share for the final product.

6 CONCLUSION AND FUTURE IMPROVEMENTS

The visualization dashboard provides an intuitive way for users to explore the details of conflicts recorded on the continent of Africa

during the periods of 1997-2017. Through various visualization techniques we are able to provide the user with a multi-dimensional way to interactively explore the data from different levels of granularity. The variety of visualizations gives a user freedom to explore the data from different perspectives. The user can freely explore and interact with the data until they settle on a question of interest that requires more research. Then using the input tools they can target the visualizations to explore that specific question.

We believe that additional data cleansing could improve the performance of the application. Due to the large size of the dataset, there is a noticeable delay in updating visualizations based on the user input, which could lead to a negative experience and reduce a user's incentive to spend more time exploring the data. Some ways we would consider improving performance is reducing the categories of actors. For example, each country's military is listed as a separate actor and sometimes the bounded within specific years. More data processing could be employed to reduce these to a single category, such as "Local Military" rather than individually as "Military Forces of Somalia (1997-)" and "Military Forces of Somalia (2012-)", or a single category for "Civilians". Some granularity would be lost, but the gains in performance might be worth it.

For each of the target tasks there are multiple ways the data could be represented. It would be interesting to create several versions of the application and then perform some user-validation testing to decide which visualizations are the most effective. For example, while the actors table is intended to provide an understanding of the parties involved in the conflict it could be interesting to instead display the results as a network graph, which would show the interactions amongst actors in context of all the other events.

We would also like to incorporate more interactivity between the visualizations. For example, if a user clicks on one of the events in the legend it could filter the results to just that event. Or clicking on an actor from the table could highlight all of the related markers on the map.

7 REFERENCES

- [1] <https://www.acleddata.com/about-acled/>
- [2] <https://www.kaggle.com/jboysen/african-conflicts/data>
- [3] <https://amipan.shinyapps.io/visanalyticsproj/>
- [4] J. Bertin, *Semiology of Graphics Diagrams Networks Maps*, The University of Wisconsin Press, Madison, 1983.
- [5] Andrienko, N., Andrienko, G., & Gatalsky, P. (2003). Exploratory spatio-temporal visualization: an analytical review. *Journal of Visual Languages and Computing*, 14(6), 503–541. doi:10.1016/S1045-926X(03)00046-6
- [6] Karpi, Andre, et al. "Spatio-temporal data processing systems and methods." U.S. Patent No. 9,734,220. 15 Aug. 2017.
- [7] "Parallel Sets: Visual Analysis of Categorical Data - Robert Kosara." https://kosara.net/papers/2005/Bendix_InfoVis_2005.pdf. Accessed 15 Apr. 2018.
- [8] Anthony C. Robinson, Donna J. Peuquet, Scott Pezanowski, Franklin A. Hardisty & Brian Swedberg (2016) Design and evaluation of a geovisual analytics system for uncovering patterns in spatio-temporal event data, *Cartography and Geographic Information Science*, 44:3, 216-228, DOI: 10.1080/15230406.2016.1139467