

Extending the Armed Conflict Location and Event Data Project with SAS® Text Analytics

Tom Sabo, SAS Institute Inc., Arlington, VA

ABSTRACT

The Armed Conflict Location & Event Data Project (ACLED) is a comprehensive public collection of conflict data for developing states. As such, the data is instrumental in informing humanitarian and development work in crisis and conflict-affected regions. The ACLED project currently manually codes for eight types of events from a summarized event description, which is a time-consuming process. In addition, when determining the root causes of the conflict across developing states, these fixed event types are limiting. How can researchers get to a deeper level of analysis?

This paper showcases a repeatable combination of exploratory and classification-based text analytics provided by SAS® Contextual Analysis, applied to the publicly available ACLED for African states. We depict the steps necessary to determine (using semi-automated processes) the next deeper level of themes associated with each of the coded event types; for example, while there is a broad protests and riots event type, how many of those are associated individually with rallies, demonstrations, marches, strikes, and gatherings? We also explore how to use predictive models to highlight the areas that require aid next. We prepare the data for use in SAS® Visual Analytics and SAS® Visual Statistics using SAS® DS2 code and enable dynamic explorations of the new event sub-types. This ultimately provides the end-user analyst with additional layers of data that can be used to detect trends and deploy humanitarian aid where limited resources are needed the most.

INTRODUCTION

US Federal Government agencies including the Intelligence agencies, the US Department of State, and the US Agency for International development (USAID), can benefit from insight into events affecting developing and conflict-affected regions. The use cases are various. Intelligence agencies are informed proactively of declining conditions in areas where terrorist organizations might subsequently take root, and determine how escalating crises impact homeland security. Humanitarian agencies such as USAID are informed of where to distribute aid, and similarly, this insight can benefit the programming of the US Department of State.

There are a number of public and private data sets, as well as projects depicting worldwide events in these developing and conflict-affected regions. One such initiative is the Armed Conflict Location & Event Data Project (ACLED), a comprehensive public collection of conflict data for developing states. This data and analysis project produces information about the specific dates and locations of conflict events and protests, including the types of events, the groups involved, fatalities, and changes in territorial control. As of early 2016, ACLED has recorded over 100,000 individual events with ongoing data collection focused on Africa and ten countries in South and Southeast Asia. ACLED seeks to support research and work devoted to understanding, predicting, and reducing levels of political violence. ACLED data are coded by a range of experienced researchers who collect information primarily from secondary sources and code it according to the scoring methods outlined in the codebook¹. ACLED data for Africa extends as far back as 1997, and the data is being updated in near-real time.

While ACLED manually codes for eight event types, including general protests and violence against civilians, when determining the root causes of the conflict across developing states, these fixed event types are limiting. How can researchers get to a deeper level of analysis? Fortunately, ACLED provides a narrative associated with each event. Text Analytics techniques and subsequent visualization can generate a wealth of additional information against this and other event-based data sets that include a narrative or a reference to one, such as the GDELT project². In this paper, we showcase a repeatable combination of predictive and classification-based text analytics provided by SAS® Contextual Analysis, applied to the ACLED data set for African states. We depict the steps necessary to determine (using semi-automated text analytics processes shown later in this paper) the next deeper level of themes associated with each of the coded event types. For example, while there is a broad RIOTS_PROTESTS

event type, how many of them are associated individually with rallies, demonstrations, marches, strikes, and gatherings? Analysis of these deeper level themes can be used to answer the following questions:

- How do the various forms of protest take shape and how are they interconnected?
- What types of violence are occurring against civilians and are there any logical places for intervention?
- How can we use models, based on text and structured data, to determine areas that require aid or closer monitoring for an escalating situation?

The process to accomplish this builds upon two previously presented papers at SAS Global Forum, one in 2014 to define a framework for research analytics³, and second in 2015 to extend this framework for government spending⁴. Readers benefit from learning how to concisely apply SAS capabilities to achieve an end-to-end solution that is focused around text analytics and subsequent visualization.

The five-step process for generating and using the framework is as follows:

1. **Data acquisition and preparation for text analytics:** Data is acquired for our example use case through web interfaces and is converted into a SAS data set using SAS® Enterprise Guide®.
2. **Text analytics:** We use SAS® Contextual Analysis for modeling and rule-building techniques to generate hierarchical categorical data. This newly generated categorical data serves as additional structured information for exploring the ACLED data set at a deeper level of analysis.
3. **Data preparation for visual analysis:** Readers of the previous papers benefit from new scoring techniques depicted in this step. We use the DS2 code provided by SAS Contextual Analysis to score the ACLED data set against the newly generated categories, and we use additional code to augment this scoring, enabling hierarchical exploration of the data in the subsequent visualization steps.
4. **Ad hoc exploration and modeling:** This is accomplished with SAS® Visual Analytics and SAS® Visual Statistics.
5. **Interactive report generation and use:** This is also accomplished with SAS Visual Analytics.

STEP 1: DATA ACQUISITION AND PREPARATION FOR TEXT ANALYTICS

We obtained the ACLED Version 4 from <http://www.acleddata.com/data/acled-versions-1-5-data-1997-2014/>. You are encouraged to try these methods against the latest version of the data. The data spanned 25 data columns across 80K unique events. The data was provided as an XLSX file. We imported the data into a SAS data set through the SAS Enterprise Guide interface. In the process, we retained the original 80K events for scoring, but used a sample of 15K events for the text analytics work.

STEP 2: TEXT ANALYTICS – SAS CONTEXTUAL ANALYTICS

After registering the SAS data set using SAS® Management Console, we were able to select and load the data set within the SAS Contextual Analysis interface. This enabled us to select the NOTES column as the freeform text field to perform text analytics against, as well as specify the EVENT_TYPE as a nominal category variable. See Figure 1 below for an example snippet of the data, including the LATITUDE/LONGITUDE columns.

	EVENT_TYPE	NOTES	LATITUDE	LONGITUDE
1	Violence against civilians	4 January: 16 citizens were murdered in the village of Benachour (Blida) by masked men a few hund...	36.4686	2.8289
2	Violence against civilians	19 January: Explosion of a car bomb in the main street of Algiers??? working-class district of Beloui...	36.766	3.05
3	Violence against civilians	10 February: 25 people massacred in Hammam Melouane (Blida) by an armed group.	36.4831	3.0478
4	Violence against civilians	23 March: 32 people, including 14 women, were savagely mutilated and killed in Ouled Antar (Ksar...	35.8856	2.7508
5	Violence against civilians	11 April: 23 citizens including 12 women and 6 children were savagely massacred in the.Menaa villa...	36.574125	2.912142
6	Violence against civilians	22 April: 42 people massacred in El Omaria (MTdT).	36.2644	3.0342
7	Violence against civilians	26 May: a family of 8 people was assassinated by an armed group in the village of Djebabra.(MTdT...	36.5897	2.4475
8	Violence against civilians	6 July: 27 people were killed by an armed group in Ann Boucif (MTdT).	36.2675	2.75
9	Violence against civilians	12 July: 33 citizens were killed by an armed group in the Fetha doyar close to Ksar Boukhari.(MTdT ...	36.2675	2.75
10	Violence against civilians	21 July: Massacre of 21 citizens in the M????ghita doyar, in ChrTa (Blida) by an armed group.	36.4256	2.8767
11	Violence against civilians	25 July: 13 citizens of the Sidi-Salem doyar near Omaria (MTdT) were savagely assassinated by a...	36.2644	3.0342
12	Battle-No change of territory	3 August: Massacre of 19 young citizens in Amroussa (Blida) by an armed group. It was discovered t...	36.4686	2.8289
13	Violence against civilians	5 August: 9 citizens massacred in the Benamor area of Oued Slama (Blida) by armed men.	36.5314	3.1269
14	Violence against civilians	5 September: Massacre of over 70 citizens in the working-class area of Sidi-Youcef in BTni- Messou...	36.7833	2.9833
15	Violence against civilians	12 September: 9 people massacred in LarbaG by an armed group.	35.0961	0.2078
16	Violence against civilians	18 December: 31 citizens from Larba?????6 (Blida), including women and children were massacred...	36.4686	2.8289
17	Violence against civilians	23 Dec 1997: Carnage in two villages located between the Wilayas of Taret and Tissemsilt.(Sidi-Ant...	34.9167	1.5833
18	Violence against civilians	24 December 1997: Massacre of 26 citizens, including women and children, in the village of Zouabri...	35.3758	1.3131
19	Violence against civilians	Horrible carnage on the first day of Ramadan in three villages located a few kilometres away from R...	36.3875	1.2803
20	Violence against civilians	12 January Adelaide Advertiser- 26 villagers from three large families, including 11 children and four...	36.38	3.9014
21	Violence against civilians	12 January 1998 Adelaide Advertiser 20 civilians were killed by GIA gunmen in Tablat in Medea pro...	36.4119	3.3128
22	Violence against civilians	19 April 1998 Reuters News One peasant and his 10-year-old daughter were killed on Thursday in...	36.2675	2.75

Figure 1: Snippet of ACLED Data Set Including EVENT_TYPE, NOTES, LATITUDE, and LONGITUDE

Defining a category variable tells SAS Contextual Analysis to model textual rules against the narrative (NOTES). These textual rules differentiate terms that appear in the narrative for each category from the others. In the context of this data set, it highlights the different ways that events are expressed in that narrative, such as the various forms of protests (demonstrations, sit-ins, riots) and forms of violence against civilians (village bombings, kidnappings, or massacres). A researcher could similarly manually generate categorical taxonomy to capture these instances, but this would require subject matter expertise to define a taxonomy to cover all events. The power of this approach is that it automatically generates a taxonomy that fits and describes each data set. Figure 2 depicts generating a project against this data, using the EVENT_TYPE as a categorical variable and the NOTES as the text variable.

Figure 2: Defining a New Project in SAS Contextual Analytics That Includes a Category Variable

While SAS Contextual Analysis includes a number of capabilities in exploratory text analytics, including term exploration and topic generation, we focus in this paper on its capability to generate textual rules against nominal categorical data, and the subsequent scoring and augmentation of data sets using these rules. For a further study of terms and topic exploration in the context of a research-oriented data set,

please see SAS Global Forum paper 061-2014, “Uncovering Trends in Research using Text Analytics with Examples from Nanotechnology and Aerospace Engineering.”³

In Figure 3, we examine the textual rules that SAS Contextual Analysis generated against the NOTES narrative for the EVENT_TYPE, RIOTS_PROTESTS. Again, these are certain terms and phrases that appear in the narrative that don’t tend to appear for the other event types. These can be used to auto-classify new narratives as particular events, characterize the various forms of the event, and subsequently be used to visually interact with the data in a more finely tuned manner. The terms that are generated here, including “demonstrate”, “demonstration”, “gather”, “march”, “protest”, “rally”, “riot”, and “sit-in” differentiate the various forms of protests that are taking shape in the region. In addition, it can identify groups, such as the Muslim Brotherhood, that are almost exclusively mentioned in the context of a protest.

Each rule identified in SAS Contextual Analysis is defined by a colored bar, including blue, yellow, and red components. The blue component of the bar defines cases where a rule correctly matches for the given event type, or true positives. The yellow component of the bar defines cases where the rule matched for a different event type, or false positives. The red component of the bar is primarily applicable to the EVENT_TYPE level, and defines cases where SAS Contextual Analysis is unable to define a consistent rule to differentiate these events from the other event types. In Figure 3, for example, what this means is that the sum total of true positives from the rules generated by SAS Contextual Analysis for RIOTS_PROTESTS cannot differentiate all the cases of RIOTS_PROTESTS from the other event types. This is likely because a number of these events contain key themes that are too similar to the other event types, such as VIOLENCE AGAINST CIVILIANS. For an individual rule, the red component simply depicts the terms that make up the rule, such as “riot” does not cover all cases of the parent event type, RIOTS_PROTESTS.

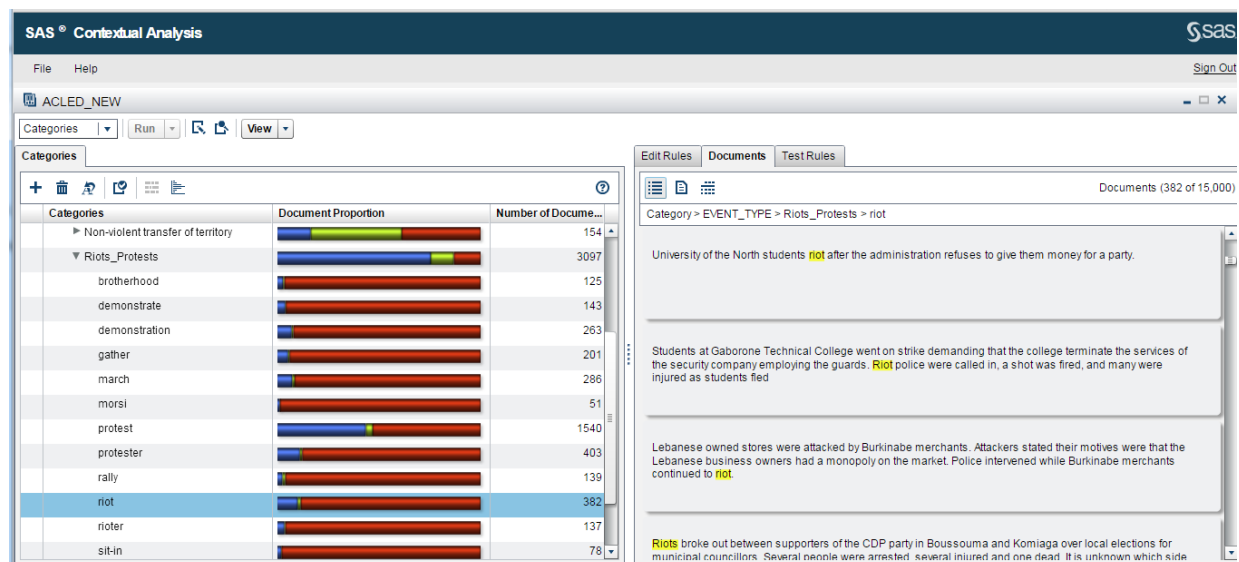


Figure 3: SAS Contextual Analysis Characterizes Various Forms of Protest in the Data Set

Similarly, SAS Contextual Analysis generates rules that characterize violence against civilians. This is shown in Figure 4 below. These include cases such as abduction and kidnapping. The automatically generated rules can be refined manually with subject matter expertise. Perhaps it makes sense to combine rules defining abduction and kidnapping into one group. This rule might be one which an overseeing organization can benefit from in terms of intervention. The rules can include phrases such as “bomb” and “village” – indicating that this was a frequent enough trend in the data to warrant its own rule. Some rules tend to be governed by a particular term in the absence of others, such as “attack” where “force”, “soldier”, “police”, or “position” is not mentioned, thus differentiating attacks against civilians from militaristic clashes. Furthermore, this automated rule generation identifies extreme forms of violence, such as massacres in the data set.

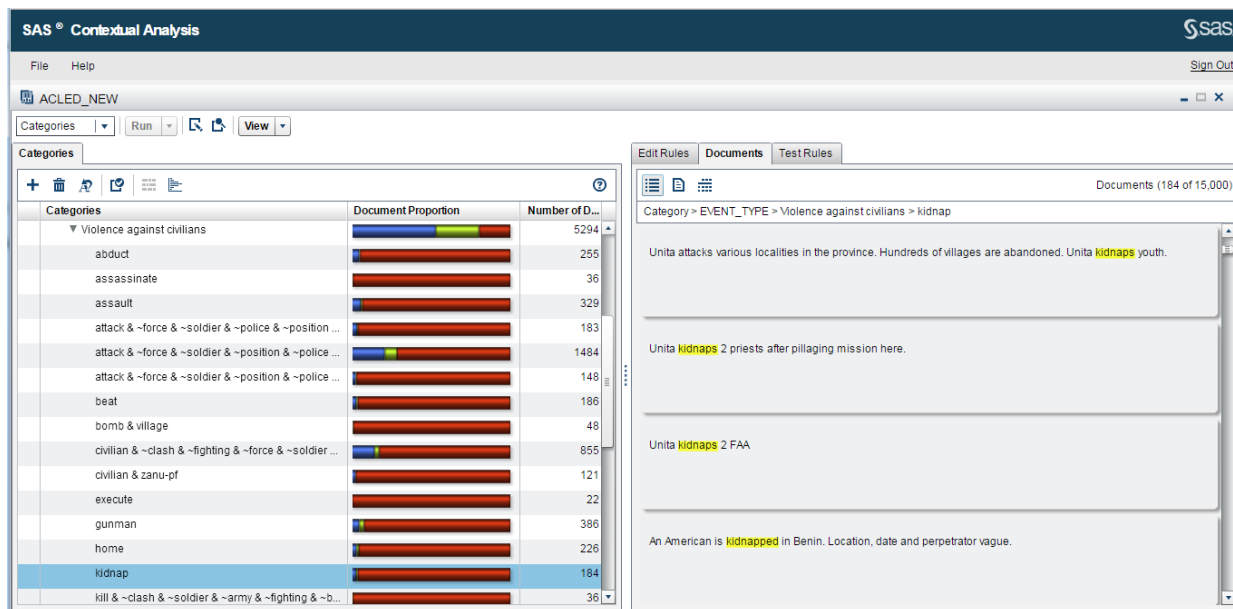


Figure 4: SAS Contextual Analysis Characterizes Various Forms of Violence against Civilians in the Data Set

STEP 3: DATA PREPARATION FOR VISUAL ANALYSIS

To prepare the data for visual analysis, we benefit from the out-of-the-box DS2 code. The predecessor to SAS Contextual Analysis, SAS Enterprise Content Categorization, required modification of a Java or Python script to invoke scoring of a categorical taxonomy. Now, users of SAS Contextual Analysis can invoke SAS directly to score data. We obtain the code through a drop-down list in the SAS Contextual Analysis interface after we've generated and refined the category rules as described above. The code is accessed from the "View" drop-down under "Category Code". Figure 5 below depicts a snippet of the DS2 code that is available for scoring.

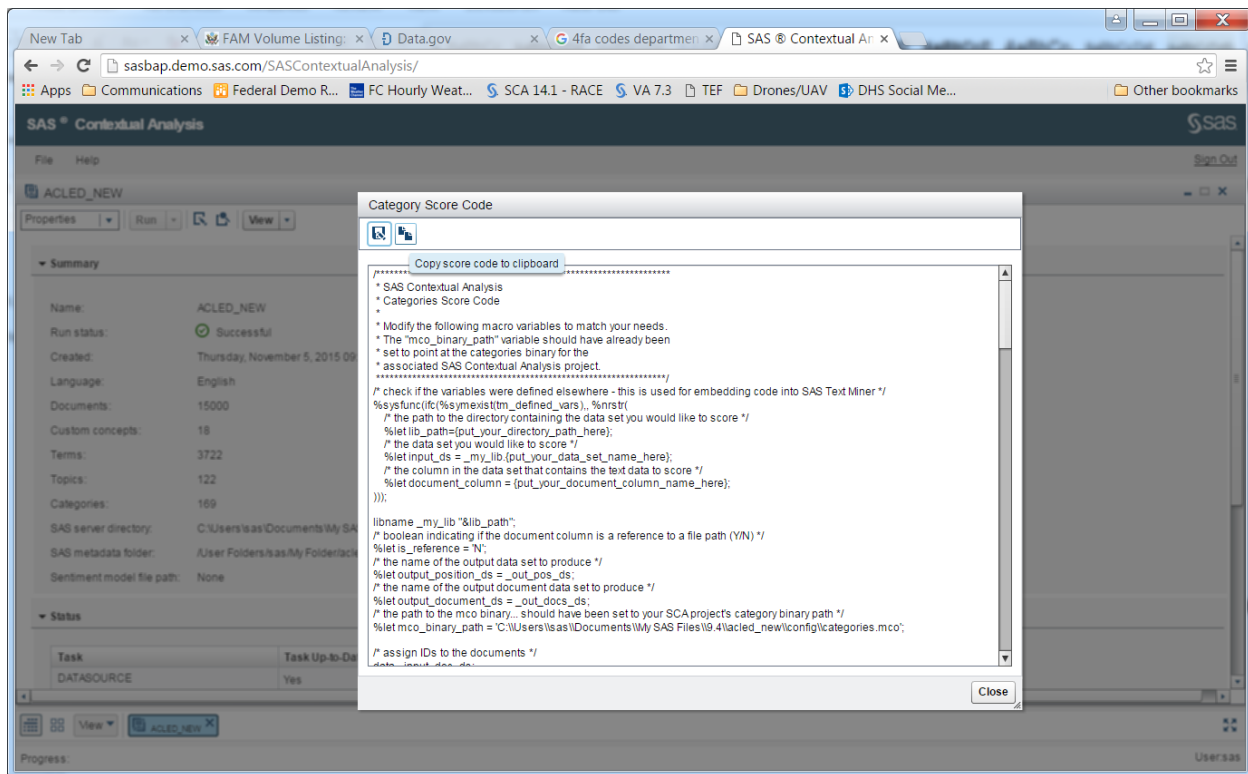


Figure 5: Accessing the DS2 Code for Scoring Categories

This code can be modified to point to a new data set, so in our case, we point to the original 80K events that we took a 15K sample from. Thus, we can explore our category hierarchy against the full data set. In addition, we can reshape the data to produce a categorical hierarchy using additional code, which we can further use to explore the data in SAS Visual Analytics. We would like a data set that includes at one level, the various EVENT_TYPES, and at a second level of depth, the different rules associated with each event type that SAS Contextual Analysis generated. This enables us to explore these rules against the rest of our structured data, including the geospatial coordinates of the events, and the number of fatalities.

To accomplish this, we offer the following additional code that can be run immediately following and in conjunction with the code supplied out-of-the-box:

```
data &input_ds; set &input_ds;
    document_id = _n_;
run;

/*****
*      Remove the 'Top/' text from the hierarchy strings
*****/

data &output_position_ds; set &output_position_ds (drop= term column_name );
    full_path = tranwrd(full_path, 'Top/' , '');
run;

/*****
```

```

* create the offsets to grab all text that triggered a match to the category
*****/

proc sql noprint;
    create table temp_table as
    select *, min(start_offset) as min_offset, max(end_offset) as
        max_offset
    from &output_position_ds
    group by document_id, full_path ;
quit;

/*****
* Remove dupes where a doc maps to the same lowest level in the hierarchy
*****/

proc sort data= temp_table out= &output_position_ds nodupkey;
    by document_id full_path;
run;

proc delete data=temp_table;
run;

/*****
* Create hierarchies from the extraction paths
*****/

/* Find the maximum category depth in any of the hierarchies and make it a
macro variable */
proc sql noprint;
    select max(count(full_path, '/')) + 1 as maximum into
        :max_number_of_levels
    from &output_position_ds;
quit;

data &output_position_ds (drop = ExpressionID start stop position length i
level_count ); set &output_position_ds;

    /* Find the depth of the hierarchy for the current observation */
    level_count = count(full_path, '/') + 1;

    /* Set an array to create the hierarchical structure and capture the
named levels */
    array category_level_{&max_number_of_levels} $45;

    /* Set the regex and other initial variables */
    ExpressionID = prxparse("/[a-zA-Z0-9_&~\ - ]+/" );
    start = 1;
    stop = length(full_path);

    /* Call the regex for the first time */
    call prxnext(ExpressionID, start, stop, full_path, position, length);

    /* Extract and Iterate */
    do i = 1 to level_count while (position > 0);
        category_level_{i} = substr(full_path, position, length);
        call prxnext(ExpressionID, start, stop, full_path, position,
            length);
    end;

```



```

        end;
run;

/*****
* Join the tables
*****/

proc sql;
    create table output_scored as
    select *
    from &input_ds as a
    left join
    &output_position_ds as b
    on a.document_id = b.document_id;
quit;

/*****
* Add 'Unclassified' tags to the entries that didn't make it into a category
*****/

%let max_number_of_levelsC = &max_number_of_levels;

data output_scored (drop= i j);
    set output_scored;
    if category_level_1 = "" then do;
        array cat{&max_number_of_levels} $ 30 category_level_1-
            category_level_&max_number_of_levelsC;
        do i = 1 to &max_number_of_levels;
            if cat[i] = '' then cat[i] = 'Uncategorized';
        end;
    end;
    else do;
        array cat1{&max_number_of_levels} $ 30 category_level_1-
            category_level_&max_number_of_levelsC;
        do j = 1 to &max_number_of_levels;
            if cat1[j] = '' then cat1[j] = 'Other';
        end;
    end;
end;

run;

/*****
* Extract all text that triggered a match to the category
*****/

data output_scored (drop = snippet_length max_offset min_offset start_offset
end_offset ); set output_scored;

    /*Nothing to extract if doc was uncategorized*/
    if category_level_1 not in ('Uncategorized') then do;

        /*calculate length*/
        snippet_length = max_offset - min_offset;

        /*extract the snippets*/
        extracted_snippet = substr(&document_column, min_offset,
            snippet_length + 2) ;
    end;
else;

```



```

run;

/*****
* View the data
*****/

proc print data=output_scored (obs=100);
run;

/*****
*      Save the data - Replace with your own library
*****/

libname outputlib 'c:\data\output';
data outputlib.cbm_maintenance_scored; set output_scored;
run;

/* delete intermediate data sets */
proc sql noprint;
    drop table _tmp_doc_cat_ds;
    drop table _tmp_transposed_ds;
    drop table _tmp_missing_ds;
    drop table _input_doc_ds;
quit;

```

STEP 4: AD HOC EXPLORATION AND MODELING

We load the data into SAS Visual Analytics for exploration, modeling, and reporting. SAS Visual Statistics, a set of predictive capabilities available within SAS Visual Analytics, provides interactive decision trees that illuminate trends in the data. This is just one example of the kind of models which we can apply to the original structured data results, as well as the new structured data (rules) generated from the textual modeling exercise in the previous step.

In this decision tree, we set the EVENT_TYPE as a target and use the textual rules, latitude, longitude, and the year associated with each event as input to the model. The resulting bins generated from the decision tree help characterize the data in meaningful ways. See Figure 6 below. From this model, in one area defined by a particular latitude and longitude boundary, the primary event prior to the year 2011 is VIOLENCE AGAINST CIVILIANS. After the year 2011, the primary event is RIOTS_PROTESTS, with a substantial proportionate decrease in VIOLENCE_AGAINST_CIVILIANS. In a different region bounded by latitude and longitude, the opposite trend is present. Prior to 1999, the primary event was RIOTS_PROTESTS, while after 1999 the primary event is VIOLENCE_AGAINST_CIVILIANS. This type of interplay can assist with understanding the trends and factors influencing events in a region, and how an overseeing or humanitarian organization can best respond.

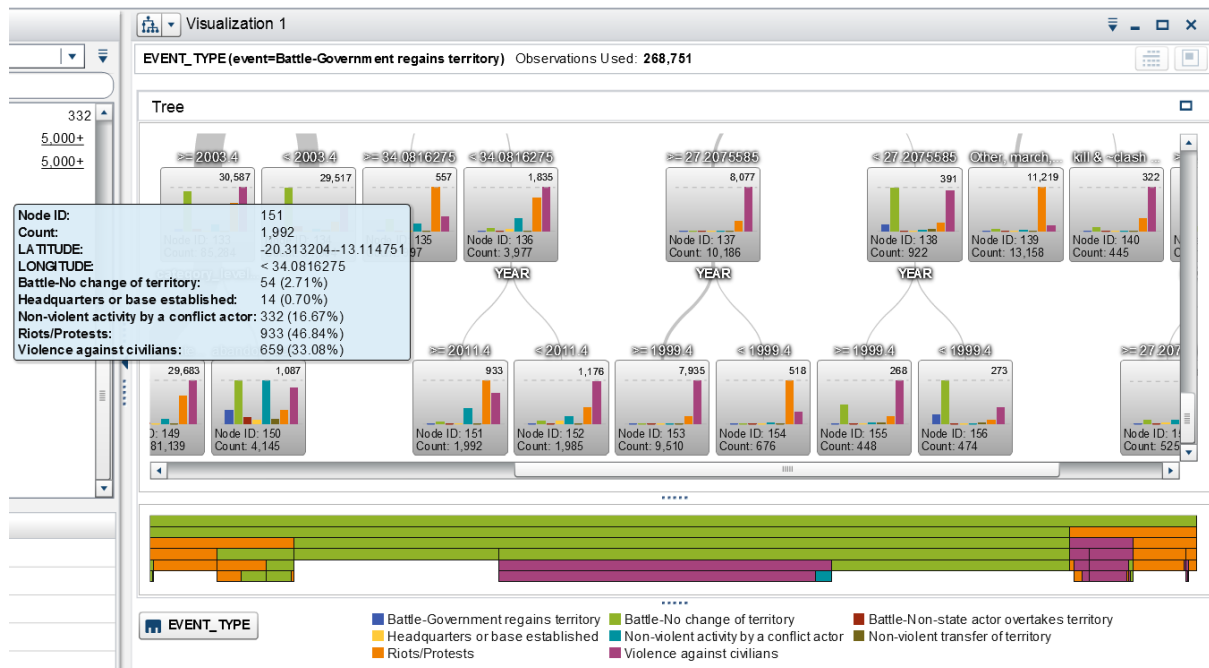


Figure 6: Decision Tree of Event Type by Rules, Location, and Timeframe

STEP 5: INTERACTIVE REPORT GENERATION AND USE

SAS Visual Analytics enables the generation of interactive reports against the available structured and unstructured data. The report benefits greatly from interactive filtering against the time component, geospatial components, and key terms and phrases that are associated with the events and event subtypes. Figure 7 below depicts an example of an interactive dashboard against this data to enable analysts to interactively explore trends in events. The dashboard enables exploration by timeframe using the slider at the top of the dashboard. It also enables exploration by the taxonomy modeled against the EVENT_TYPE, including the newly generated variations of violence against civilians such as kidnapping, looting, massacres, or village bombings. It also enables geospatial interaction with the data. Hence, as an analyst, I can explore actionable information such as recent cases of kidnapping or abduction, and subsequently drill down into a particular geographic area, such as Tripoli. This interaction is specifically depicted in Figure 7 below. Each selection of this dashboard influences the visualization results, such that we ultimately in this interaction drill down to those specific events in Tripoli to better determine their cause. Recall that prior to the text analytics exercise from the previous step, there was no way to differentiate cases of kidnapping and abduction from other variations of violence against civilians.

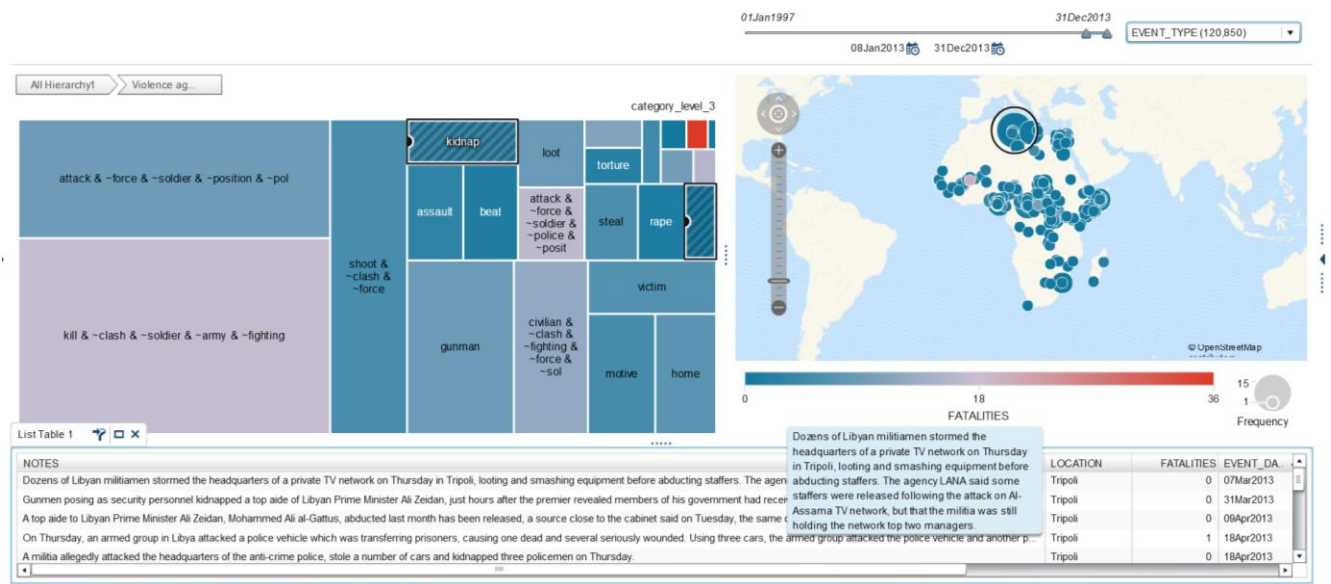


Figure 7: SAS Visual Analytics Dashboard Enabling Interactive Exploration of Violence against Civilians

Similarly, we can explore trends related to the other event types. In exploring the EVENT_TYPE RIOTS_PROTESTS, Figure 8 depicts how certain subtypes are correlated with stronger levels of fatalities than others. As seen below, the sit-in subtype is associated with violence to clear out the Muslim brotherhood from particular sit-ins in 2013. The Muslim Brotherhood is associated with a number of protests and warrants its own automatically generated rule. We can geospatially depict how these events play out, and drill down into a particular area. It is also a simple matter to add other drill-down capability to this visualization, including the structured actor data. This would enable analyses to determine what groups are involved with events, and subsequently visualize what actions these groups are taking, or what is happening to these groups from other actors. All of these analyses benefit from the finer granularity that the generated textual subtypes add.

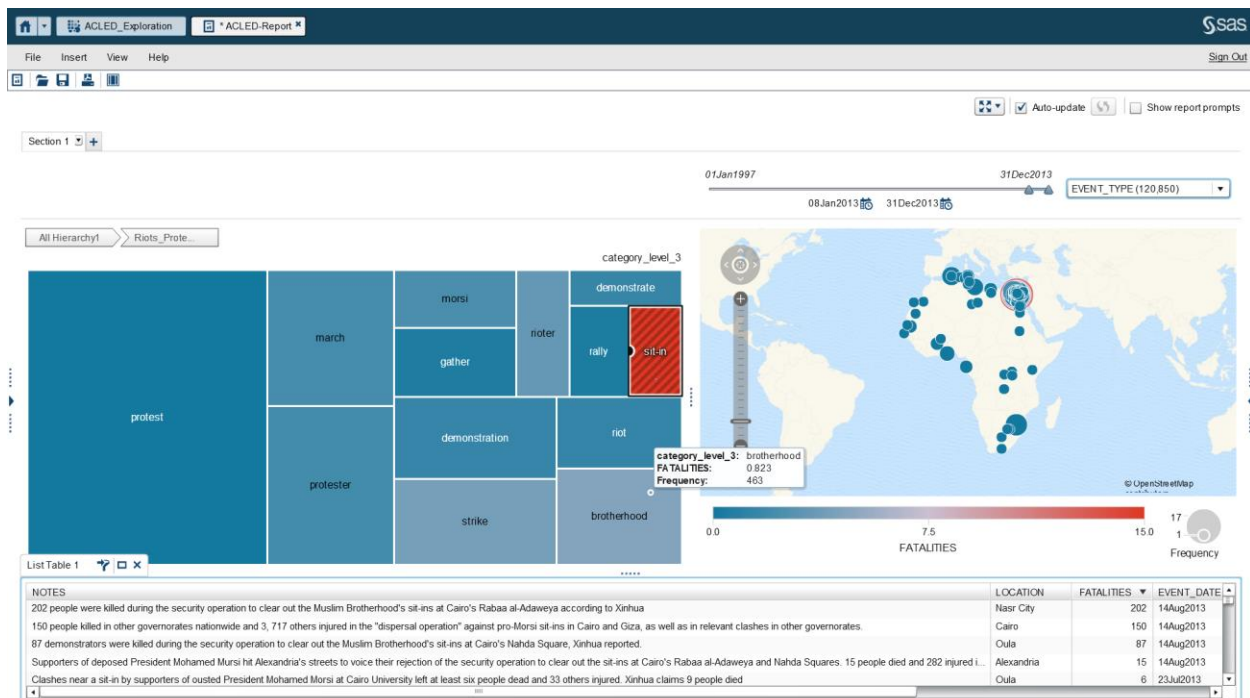


Figure 8: SAS Visual Analytics Dashboard Enabling Interactive Exploration of Riots and Protests

Figure 9 below shows an identical visualization to Figure 8 except that the treemap is replaced with a phrase cloud. The phrase cloud is a different way to enable the interactive dashboard experience, where the size of the phrase is synonymous to the size of the box in the treemap. It can be particularly useful when there are so many sub-terms to display at a hierarchical exploration level that they don't fit in the treemap. Either method enables subsequent drill down into the hierarchy of sub-categories and other structured data.

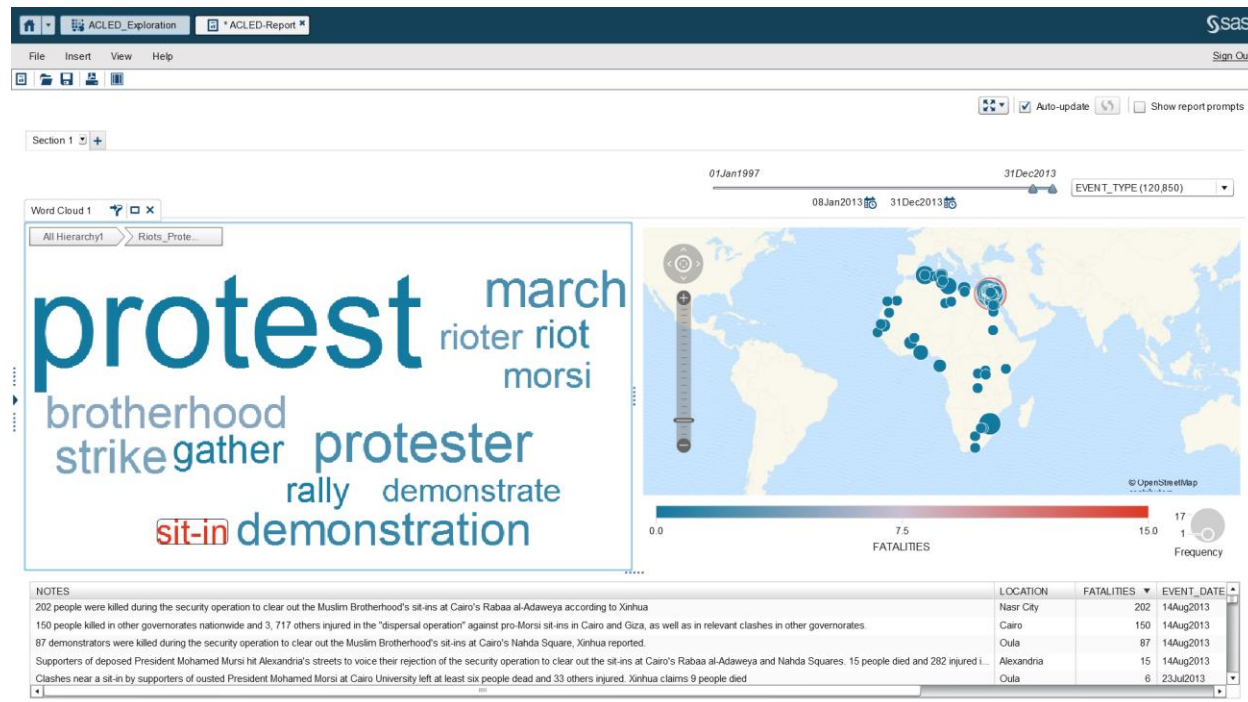


Figure 9: SAS Visual Analytics Dashboard Replacing a Treemap with a Phrase Cloud

CONCLUSION

In summary, we showcased in this paper how a repeatable combination of exploratory and classification-based text analytics applied to the publicly available ACLED for African states could generate actionable information for intelligence and humanitarian oriented organizations. Specifically, we showed how various forms of protest are related to violence and particular organizations, how subsets of violence against civilians could lead to actionable information through interactive dashboards, and how modeling not only produces textual taxonomy for refined exploration, but can also identify key shifts in trends across particular geographic areas over time.

This methodology is widely applicable. Similar analyses are possible wherever there is enough data, approximately more than 500 rows of information, wherever there is a target variable associated with the text. Tip lines, surveys, and medical encounters are just a few of the data sets where we have applied these text and visualization analyses. We have recently applied similar methods to offshore pipeline safety analyses using unsupervised learning methodology, armed forces logistics for identifying key flaws in maintenance, determining whether individuals enrolled in special-ops trainings are more or less likely to drop out based on free-form comments, and a variety of call center data analyses including data related to consumer finance protection. For more areas of text analytics and subsequent visualization applications, please see the paper, "Text Analytics in Government: Using Automated Analysis to Unlock the Hidden Secrets of Unstructured Data."⁵

One area for future work lies in exploring concept or fact extraction cross-cutting the category and subcategory extraction. The work listed in this paper focuses on category and sub-category extraction. Concept extraction involves identifying people, places, and things, while fact extraction identifies how these concepts interact. Concept or fact extraction would enable the detection of particular events in the

The taxonomy developed from this exercise could also be used in auto-coding events. Based on a description entered, the individual recording the event could be presented with a set of options based on how the event fits with the models developed from the existing data. They would be free to select the suggested event or override it. Such auto-coding or semi-automated coding also had broad applicability to any organization seeking to retain a knowledge repository, and tag/organize that knowledge repository with metadata in a consistent manner.

The screenshot displays the SAS Viya user interface for analyzing Boko Haram fatalities. At the top, a navigation bar contains 'Discover', 'Visualize', and 'Dashboards'. Below this, a filter bar allows users to refine their search by 'actor1' (Boko Haram) and 'FATALITIES'. A date range selector is set to 'January 1st 1997, 00:00:00.000 to January 8th 2014, 02:24:49 679'. On the left sidebar, various fields are listed under 'Selected Fields' and 'Available Fields', including administrative regions (ADMIN1), source organizations (SOURCE), and specific actors (ACTOR1, ALLY_ACTOR_1). The main area features a bar chart showing the count of fatalities over time, followed by a detailed table of incidents. Each row in the table provides information on the date and time of the event, the location (ADMIN1), the number of fatalities, the categories involved (category_level_1), the reporting source (SOURCE), and a brief description of the incident (NOTES).

Finally, many data sources are not as structured as ACLED, and for a demonstration of tokenization on a data set of large documents and subsequent analysis, I recommend looking at “Star Wars and the Art of Data Science: An Analytical Approach to Understanding Large Amounts of Unstructured Data”⁶.

1. The Armed Conflict Location & Event Data Project. Available at <http://www.acleddata.com/>.
2. The GDELT Project. Available at <http://www.gdeltproject.org/about.html>.
3. Sabo, Tom. 2014. "Uncovering Trends in Research Using Text Analytics with Examples from Nanotechnology and Aerospace Engineering." *Proceedings of the SAS Global Forum 2014 Conference*. Cary, NC: SAS Institute Inc. Available <http://support.sas.com/resources/papers/proceedings14/SAS061-2014.pdf>.

4. Sabo, Tom. 2015. "Show Me the Money! Text Analytics for Decision-Making in Government Spending." *Proceedings of the SAS Global Forum 2015 Conference*. Cary, NC: SAS Institute Inc. Available <http://support.sas.com/resources/papers/proceedings15/SAS1661-2015.pdf>.
5. Sabo, Tom. 2014. SAS Institute white paper. "Text Analytics in Government: Using Automated Analysis to Unlock the Hidden Secrets of Unstructured Data." Available http://www.sas.com/en_us/whitepapers/text-analytics-in-government-106931.html.
6. Osborne, Mary, and Adam Maness. 2014. "Star Wars and the Art of Data Science: An Analytical Approach to Understanding Large Amounts of Unstructured Data." *Proceedings of the SAS Global Forum 2014 Conference*. Cary, NC: SAS Institute Inc. Available <http://support.sas.com/resources/papers/proceedings14/SAS286-2014.pdf>.

ACKNOWLEDGMENTS

Thanks go out to Brian Major for originally obtaining the data and formatting it appropriately for analysis before the ACLED interface was subsequently improved. Thanks also to Brian Major for providing the integration with Elastic Search and Kibana as an alternative means to visualize the text analytics results. Thanks to Adam Pilz for providing the DS2 code used to format scored categories in a hierarchical fashion suited for visualization. Thanks to Emily McRae for reviewing this paper and suggesting clarifications on the concepts presented within.

RECOMMENDED READING

- Chakraborty, Goutam., M. Pagolu, and S. Garla. 2013. *Text Mining and Analysis: Practical Methods, Examples, and Case Studies Using SAS®*. Cary, NC: SAS Institute Inc.

CONTACT INFORMATION

Your comments and questions are valued and encouraged. Contact the author:

Tom Sabo, Principal Solutions Architect
SAS Institute Inc.
1530 Wilson Blvd.
Arlington, VA 22209
+1 (703) 310-5717
Email: tom.sabo@sas.com

SAS and all other SAS Institute Inc. product or service names are registered trademarks or trademarks of SAS Institute Inc. in the USA and other countries. ® indicates USA registration.

Other brand and product names are trademarks of their respective companies.