

# Using SAS® Text Analytics to Assess International Human Trafficking Patterns

Tom Sabo, Adam Pilz, SAS Institute Inc

## ABSTRACT

The US Department of State (DOS) and other humanitarian agencies have a vested interest in assessing and preventing human trafficking in its many forms. A subdivision within the DOS releases publicly facing Trafficking in Persons (TIP) reports for approximately 200 countries annually. These reports are entirely freeform text, though there is a richness of structure hidden within the text. How can decision-makers quickly tap this information for patterns in international human trafficking?

This paper showcases a strategy of applying SAS® Text Analytics to explore the TIP reports and apply new layers of structured information. Specifically, we identify common themes across the reports, use topic analysis to identify a structural similarity across reports, identifying source and destination countries involved in trafficking, and use a rule-building approach to extract these relationships from freeform text. We subsequently depict these trafficking relationships across multiple countries in SAS® Visual Analytics, using a geographic network diagram that covers the types of trafficking as well as whether the countries involved are invested in addressing the problem. This ultimately provides decision-makers with big-picture information about how to best combat human trafficking internationally.

## INTRODUCTION

Human trafficking is one of the most tragic human rights issues of our time. It splinters families, distorts global markets, undermines the rule of law, and spurs other transnational criminal activity. It threatens public safety and national security<sup>1</sup>. The International Labour Organization estimates that there are 20.9 million victims of human trafficking globally, and that forced labor and human trafficking generates 150 billion dollars in illicit gains annually. Of the 20.9 million victims, 26% are children, and 55% are women and girls<sup>2</sup>.

The U.S. Department of state produces the Trafficking in Persons (TIP) report annually. It assesses the state of human trafficking in approximately 200 countries. This report is the U.S. Government's principal diplomatic tool to engage foreign governments on human trafficking. It is also the world's most comprehensive resource of governmental anti-trafficking efforts and reflects the U.S. Government's commitment to global leadership on this key human rights and law enforcement issue. It is used by the U.S. Government and worldwide as a tool to engage in dialogs to advance anti-trafficking reforms, and examine where resources are most needed. Freeing victims, preventing trafficking, and bringing traffickers to justice are the ultimate goals of the report and of the U.S Government's anti-trafficking policy<sup>1</sup>. However, the insights in these reports are scattered across hundreds of free-form text documents. How can we make the data in these reports more accessible to the broad audience that it supports, and how can we better envision patterns in the data which can be used to combat human trafficking?

This paper showcases a combination of SAS technology to identify patterns in the reports ultimately making the information more accessible to the various stakeholders. In particular, we will show how SAS can be used to identify links between source and destination countries, and visually depict these geospatial patterns in a network diagram. In this process, we will apply text analytics and visualization capabilities, primarily from SAS Visual Text Analytics and SAS Visual Analytics, available through SAS Viya. We will answer the following questions.

- Can we assess overall themes in international trafficking from the reports?
- Can we identify more focused patterns in trafficking, such as how women and children are seeking and achieving refuge?
- Can we identify and geospatially visualize patterns in trafficking across countries, including who is being trafficked (men, women, children), what type of trafficking is occurring (labor or sex trafficking), and whether the countries in question are in cooperation to address the problem?

By the end of this paper, the reader will learn how to apply the full SAS analytics lifecycle to this problem<sup>3</sup>. In this case, the analytics lifecycle includes data acquisition, unstructured and structured data management, text analytics, and network visualization. The reader will also gain an understanding of some key features available in SAS Visual Text Analytics, including similarity scores and fact extraction. We will also highlight some functionality common to the SAS text analytics products, including capabilities available across SAS Visual Text Analytics, SAS Contextual Analysis, and SAS Text Miner.

## DATA ACQUISITION AND DATA MANAGEMENT

We obtained the data for each country narrative from the U.S. Department of State Trafficking in Persons report for 2017 using a script that accessed the following link: <https://www.state.gov/j/tip/rls/tiprpt/countries/2017/index.htm>. A slight modification to the script enabled us to obtain country narrative data from 2013-2016. Each report contains summary information about trafficking in the country, as well as several subsections. Subsections include recommendations, how the country prosecutes human traffickers, how the country protects victims, how the country prevents trafficking, and an overall trafficking profile.

The country level trafficking reports are several pages in length. When working with documents greater than a page or two, it is helpful to apply some level of tokenization prior to text analytics<sup>4</sup>. Longer documents are more likely to have multiple themes embedded within. Tokenization breaks the documents up into smaller chunks, while maintaining a reference for each chunk to the larger documents. Then, patterns that appear across chunks can be readily surfaced using the capabilities at our disposal. This makes our algorithms more likely to identify these discrete themes or topics within documents.

For this effort, we applied sentence level tokenization. The following is SAS code we used for sentence level tokenization. In this case, it accepted as input a SAS data set that contained a number of rows of data, each containing a country level narrative from the TIP reports from 2013-2017.

```
/*Define the library where the data set is stored*/
libname _mylib 'D:\data\SamplePDF';

/*Define the data set for which you desire tokenized sentences*/
%let dsn = _mylib.output_sas_data;

/*Define the text variable to parse*/
%let text_var = text;

/*Strip the data and create an index*/
data temp (compress=yes); set &dsn;
    doc_id = _n_;
run;

/*parse the data set*/
proc hptmine data=temp;
    doc_id doc_id;
    var &text_var;
    parse
        nostemming notagging nonoungroups shownumpunct
        entities = none
        outpos = position
        buildindex ;
    performance details ;
run;
```

```

proc sort data=position;
  by document sentence _start_;
run;

data sentenceSize (compress=yes);
  retain document start size;
  set position;
  by document sentence;
  if First.sentence then start=_start_+1;
  if Last.sentence then do;
    size=_end_ -start+2;
    output;
  end;
  keep document start size;
run;

/*Clean up*/
proc delete data=position; run;

data sentenceObs(compress=yes);
  length sentences $1000;
  merge sentenceSize(in=A ) temp (rename=(doc_id=document) );
  by document;
  if A then do;
    sentences=substrn(&text_var,start,size);
    output;
  end;
  keep sentences document;
run;

/*Clean up*/
proc delete data=sentenceSize; run;

data _mylib.output_sentences(compress=yes);
  set sentenceObs;
  by document;
  if first.document then sid = 1; else sid + 1;
run;

/*Clean up*/
proc delete data=sentenceObs; run;

/*view the data*/
proc print data=_mylib.output_sentences (obs=100);
run;

```

The output data set from the sentence tokenization procedure contained a row of data for each sentence in the original country level trafficking narratives, maintaining year, country, and sentence ID. This amounted to 63,648 rows of sentence level data. The following figure depicts a snapshot of the data. We took a 15,000 row sample of this sentence level data across all countries and years to use in the text analytics exercise described in the next section.

	year	country	sentence_id	sentence
51050	2017	Brunei	44	TRAFFICKING PROFILE...As reported in the last five years. Brunei is a destinati...
51051	2017	Brunei	45	Men and women from Indonesia, Bangladesh, China, the Philippines, Thailand,...
51052	2017	Brunei	46	Upon arrival, some are subjected to involuntary servitude, debt bondage, non-pa...
51053	2017	Brunei	47	Some migrants who transit Brunei become victims of sex or labor trafficking upo...
51054	2017	Brunei	48	Some Bruneian women and girls are subjected to sex trafficking domestically.
51055	2017	Brunei	49	Although it is illegal for employers to withhold wages of domestic workers for mor...
51056	2017	Brunei	50	Retention of migrant workers travel documents by employers or agencies remain...
51057	2017	Bulgaria	1	Office To Monitor and Combat Trafficking in Persons....2017 Trafficking in Perso...
51058	2017	Bulgaria	2	The government demonstrated significant efforts during the reporting period by i...
51059	2017	Bulgaria	3	Law enforcement continued to take action against public officials and police offic...
51060	2017	Bulgaria	4	However, the government did not demonstrate increasing efforts compared to th...
51061	2017	Bulgaria	5	Although the total number of investigations and prosecutions of traffickers increa...
51062	2017	Bulgaria	6	The governments capacity to shelter victims and provide specialized services re...
51063	2017	Bulgaria	7	Because the government has devoted sufficient resources to a written plan that, i...
51064	2017	Bulgaria	8	Therefore, Bulgaria remained on Tier 2 Watch List for the third consecutive year.
51065	2017	Bulgaria	9	RECOMMENDATIONS FOR BULGARIA... Enhance efforts to investigate, prosec...

**Figure 1: Sentence Level Country Narrative Data Used in Text Analytics**

## TEXT ANALYTICS

SAS now has a variety of capabilities in text analytics available in different solution packages. This includes capabilities in SAS Visual Text Analytics, available as a part of SAS Viya. This also includes capabilities present in SAS Text Miner, an add-on to SAS Enterprise Miner, and SAS Contextual Analysis, both available on any SAS 9 release. In this section on text analytics methods, we will show snapshots from individual solutions, and discuss which of the aforementioned SAS products also have the described capability.

### IDENTIFYING OVERALL HUMAN TRAFFICKING TRENDS AND PATTERNS

One of the questions previously identified is whether we can assess overall themes in international trafficking from the reports. This is a capability available through an unsupervised machine learning method in text clustering. SAS assesses all the sentences across TIP reports and identifies key sets of terms that tend to occur together. For example, the terms “forced”, “child”, “beg”, and “street” tend to co-occur in the data along with other terms. These are indicative of a pattern across country narratives where children are coerced into begging. The following snapshot takes results from the text clustering capabilities of SAS Text Miner, and depicts the cluster results along with example sentences associated with the text cluster.

**Documents Nearest to the Center of Cluster 1**

Terms: *forced child beg street border roma koranic teacher mali school* # docs: cluster / corpus **249 / 3187 (7.8%)**

tokenized_sentence	Distance to Cluster Seed
Some Ukrainian children are subjected to forced begging.	0.686142
Forced begging was on the rise in 2012.	0.69465
Children, particularly Roma, are subjected to forced begging.	0.741508
Ethnic Roma children may be subjected to forced begging on the street.	0.755781
Some victims are forced into street begging.	0.759511
Children, particularly Romani children, are recruited for forced begging in Poland.	0.763812
Georgian, Romani, and Kurdish children are subjected to forced begging or coerced into criminality.	0.765216
Children in street vending or begging are reportedly vulnerable to forced labor.	0.766867
In Dakar alone, approximately 30,200 talibes beg in the streets.	0.767311
Street children are sometimes coerced into criminality or forced to beg; begging ringmasters sometimes maim children to increase their earnings.	0.776786
Some street children may be subjected to forced begging or coerced into criminality.	0.777112
Some street children may be subjected to forced begging or coerced into criminality.	0.777112

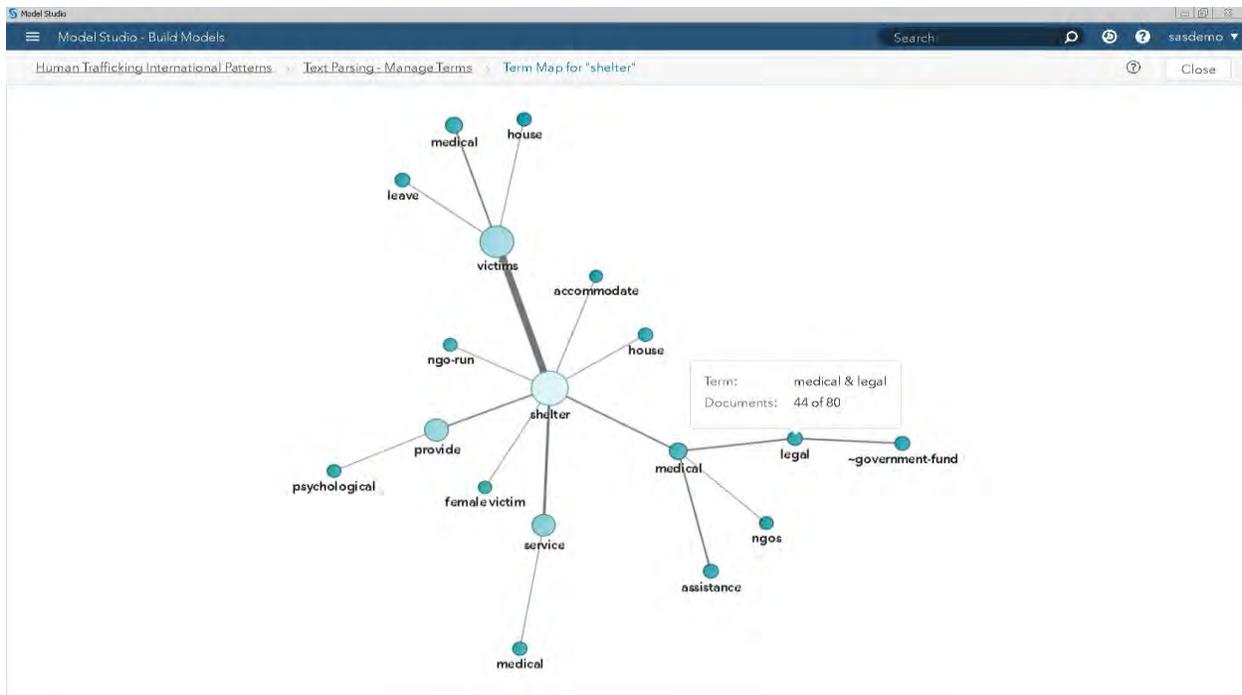
**Figure 2: Themes Report Derived from SAS Text Miner Text Clustering**

Similar results are available across all clusters and are indicative of a variety of themes in human trafficking. This includes where sex trafficking victims are typically exploited, groups who are subject to forced marriage and domestic servitude, what characteristics make individuals most vulnerable to human trafficking, and how debt bondage plays into human trafficking. Similar capabilities are available through the topics capability of both SAS Visual Text Analytics and SAS Contextual Analysis.

## IDENTIFYING FOCUSED PATTERNS RELATED TO HUMAN TRAFFICKING

A second method to identify themes in the data is through a term map. In this method of interactive exploration, the user selects a term from the full list of extracted terms across all trafficking reports, and selects to view a term map of related terms and phrases. The user is then presented with a visual depiction of other terms and phrases that tend to be connected to the source term or phrase.

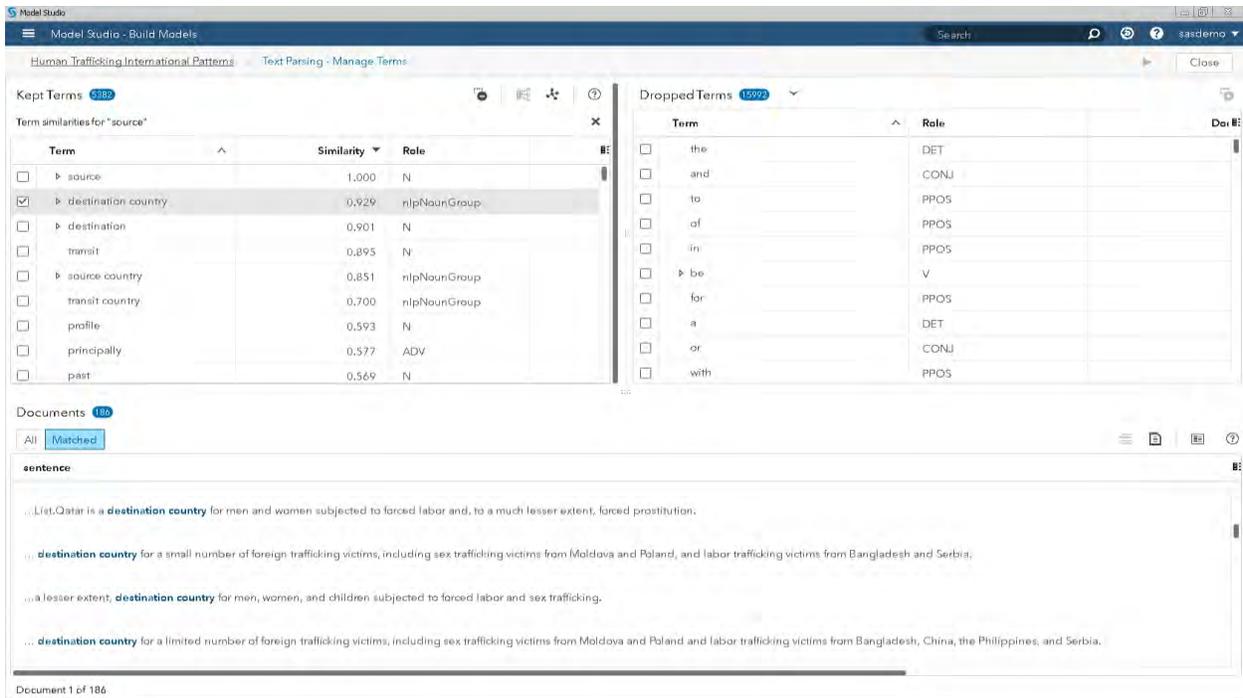
The example below depicts a term map surrounding the term “shelter”. This links other key terms and phrases, such as “provide” and “psychological”, indicating that shelters often provide psychological assistance. Another key linkage includes “female victim”, denoting who the shelters primarily serve. Finally, the term “medical” and “legal” tend to be associated with shelters, indicating other types of aid that are received at shelters. In the example below, in the 80 sentences across all reports that contain the term “shelter”, 44 of them also contain the terms “medical” and “legal”. This type of analysis provides quantitative data to advance anti-trafficking reforms, examine where resources are most needed, and can assist in determining where methods of providing assistance have been proven to be helpful. These methods can subsequently be implemented elsewhere.



**Figure 3: Term Map from SAS Visual Text Analytics Depicting Terms and Phrases Interconnected with the Term "Shelter"**

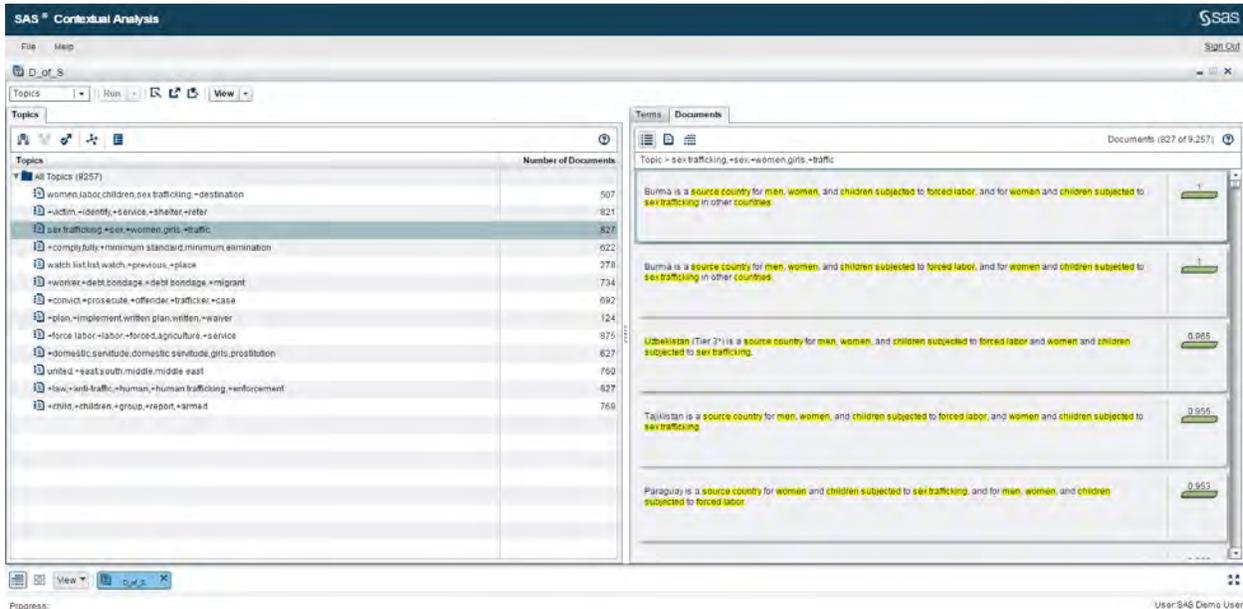
Similar capability is available from the Text Filter node of SAS Text Miner, as well as from the Terms panel of SAS Contextual Analytics.

SAS Visual Text Analytics includes a unique capability across the SAS Text Analytics products that can identify term and phrase similarities to terms of interest. This differs from the term map capabilities in that algorithms identify terms used in a similar context to the selected term. From the Terms node of SAS Visual Text Analytics, the user can select a term, such as "source" in the example below, and view terms and phrases used in a similar context. In this visualization, SAS identifies terms including "source country", "transit country", and "destination country" used in a similar context, indicating that there are connective patterns in the text between countries that are a source of human trafficking victims, and countries where these victims become involved in human trafficking. The visualization also shows these terms in the context of the sentences in which they appear.



**Figure 4: Visual Text Analytics Depiction of Term Similarity to the Term "source"**

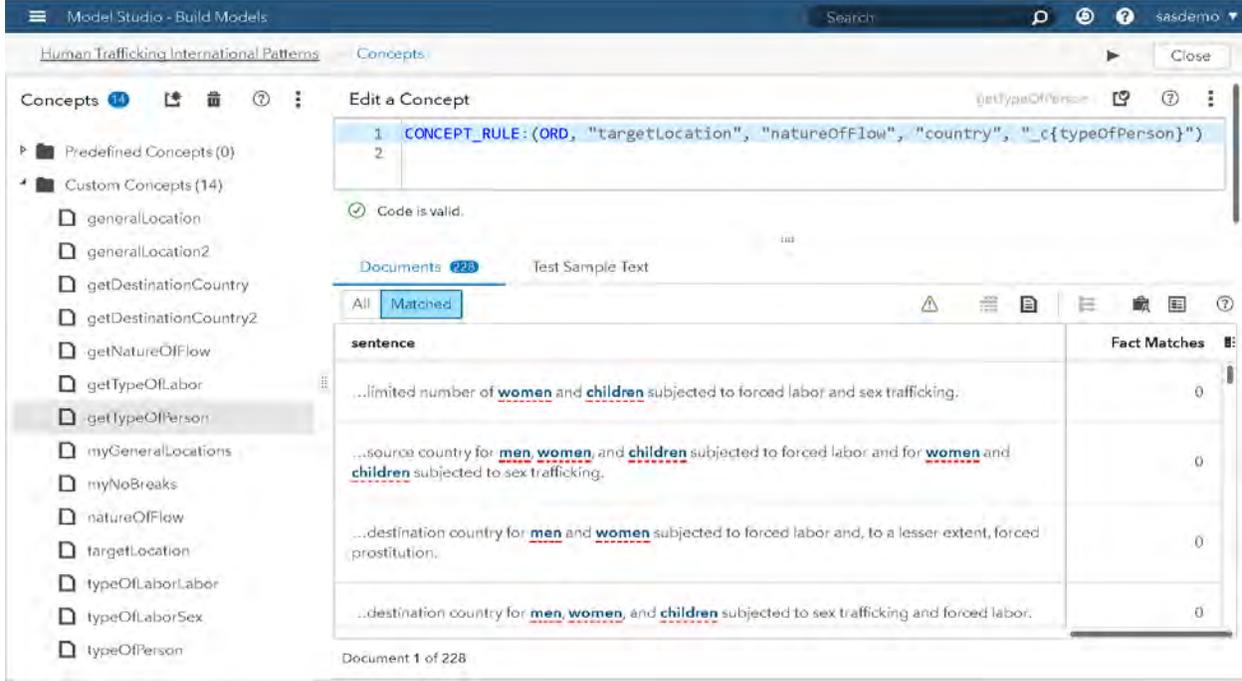
This connection between source, target, and transit countries is worth further exploration. To verify the depth of this pattern, we turn to the SAS Topics capability. In the example below taken from SAS Contextual Analysis, across 827 sentences, SAS identifies a theme (with no user input) between source countries and target countries. This theme also covers victims including men, women, and children, and the two forms of human trafficking, sex trafficking and labor trafficking.



**Figure 5: SAS Contextual Analysis Depicts a Topic Showing Network Connections in Human Trafficking**

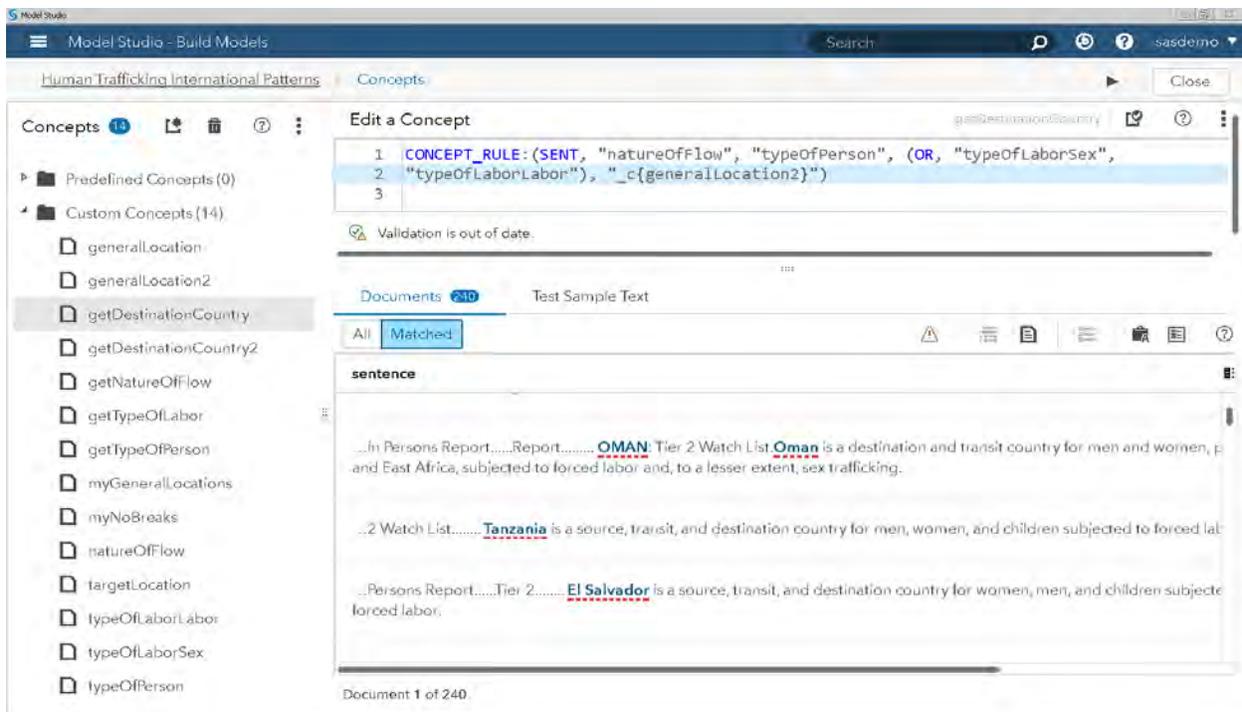
## EXTRACTING PATTERNS IN HUMAN TRAFFICKING FOR NETWORK VISUALIZATION

Now that we have used exploratory text analytics capabilities to identify a pattern of interest, analysts might be interested in geospatially depicting the interconnection between countries on a world map over time. To prepare for this activity, it is necessary to develop rules to extract these patterns or facts via extraction rules. SAS Visual Text Analytics and SAS Contextual Analysis provide the capability to use a SAS proprietary rule-writing language called LITI to define parameters for fact extraction. Through the SAS Visual Text Analytics interface using LITI, we define rules to extract the victims of trafficking in context, including men, women, and children. This is depicted in the following example visualization of the rule editor and tester below.



**Figure 6: LITI Rules in SAS Visual Text Analytics to Identify Victims of Human Trafficking**

These definitions build upon themselves, and some rule definitions, such as a list of country names, become helper definitions when writing rules to extract a larger pattern. A set of rules, along with some post-processing, enables us to extract the full pattern of source countries, destination countries, Boolean indicators indicating the victims of trafficking and types of trafficking, and finally a cooperation indicator derived from the text for each pair of countries to determine whether they are working together to address the trafficking problem. The following screenshot depicts a rule which extracts destination countries for human trafficking victims.



**Figure 7: Concept Extraction in SAS Visual Text Analytics to Capture Human Trafficking Patterns**

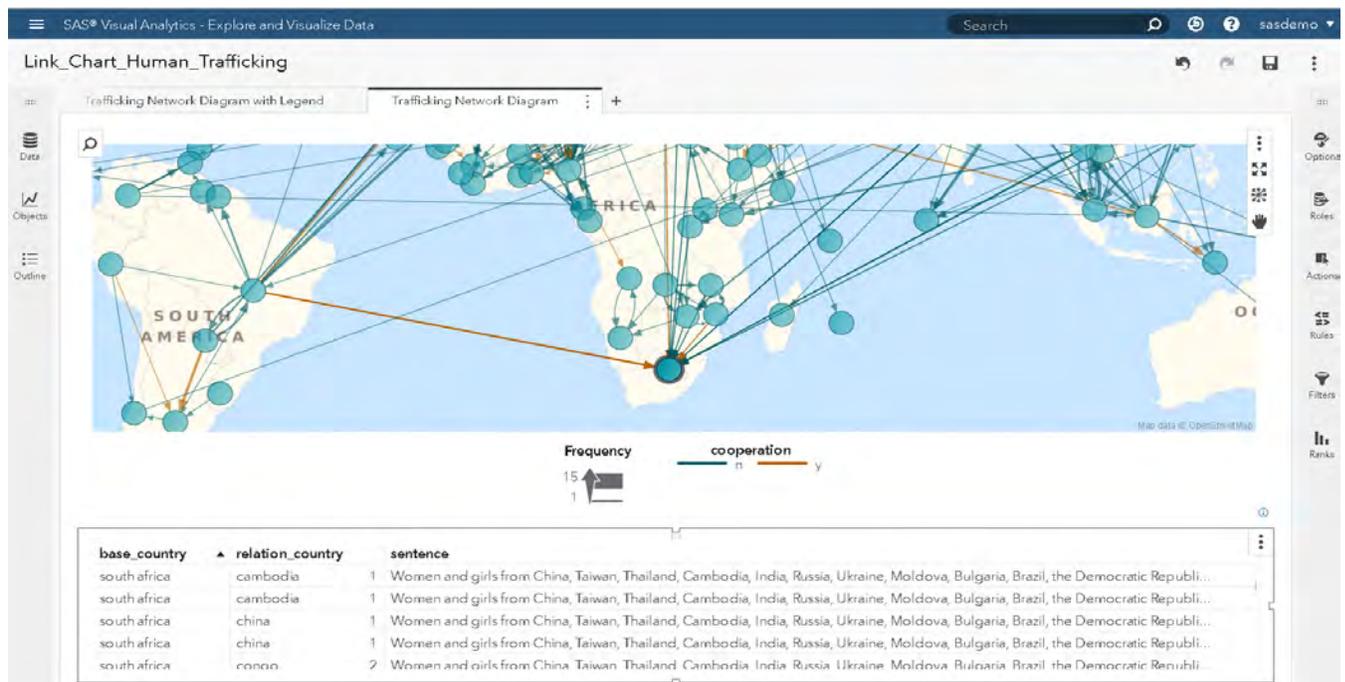
The purpose of SAS Visual Text Analytics is to develop and score these rules against source data to generate an additional data set used for visualization and interpretation of the data. In this case, we score the rules developed above to extract source/destination country patterns against the full 63,648 rows of sentence level data. In prior SAS Global Forum submissions, we've explored the output of a text analytics exercise in dashboard format, such as in assessing consumer financial complaints<sup>5</sup>. These past use cases had the benefit of additional structured data in conjunction with the free-text field, such as a user complaint in context of structured geographical information, date of claim, type of claim, and whether a user who submitted the complaint received some form of monetary compensation. In this case, we develop a visualization using only structured data we generated from the unstructured reports, namely, the connection between the countries, including victim information, year of the report, type of trafficking, and cooperation indicator. Consider that we applied automated analysis to turn reams of text into visualization-ready structured data. Consider also that these processes could be immediately deployed for new sets of these reports as they emerge in 2018, 2019 and beyond! A snapshot of this data generated from Visual Text Analytics after postprocessing is depicted below.

year	men	women	children	forced_labor	sex_trafficking	base_country	relation_country	cooperation	sentence	
2600	2017	0	1	1	0	1	afghanistan	iran	y	Afghan women and girls are subjected to sex trafficking...
2601	2017	0	1	1	0	1	afghanistan	pakistan	n	Afghan women and girls are subjected to sex trafficking...
2602	2017	1	0	1	1	0	afghanistan	turkey	n	Afghan boys and men are subjected to forced labor and...
2603	2017	1	0	1	1	0	afghanistan	greece	n	Afghan boys and men are subjected to forced labor and...
2604	2017	1	0	1	1	0	afghanistan	pakistan	n	Afghan boys and men are subjected to forced labor and...
2605	2017	1	0	1	1	0	afghanistan	iran	y	Afghan boys and men are subjected to forced labor and...
2606	2017	0	1	1	0	1	afghanistan	china	n	There were reports of women and girls from the Philippin...
2607	2017	0	1	1	0	1	afghanistan	sri lanka	n	There were reports of women and girls from the Philippin...
2608	2017	0	1	1	0	1	afghanistan	tajikistan	n	There were reports of women and girls from the Philippin...
2609	2017	0	1	1	0	1	afghanistan	iran	y	There were reports of women and girls from the Philippin...
2610	2017	0	1	1	0	1	afghanistan	iran	y	There were reports of women and girls from the Philippin...
2611	2017	0	1	1	0	1	afghanistan	iran	y	There were reports of women and girls from the Philippin...
2612	2017	0	0	0	0	1	albania	united kingdom	n	Albanian victims are subjected to sex trafficking in countr...
2613	2017	0	0	0	0	1	albania	italy	y	Albanian victims are subjected to sex trafficking in countr...
2614	2017	0	0	0	0	1	albania	greece	n	Albanian victims are subjected to sex trafficking in countr...
2615	2017	0	0	1	1	0	albania	united kingdom	n	NGOs report an increase in the number of Albanian child...
2616	2017	0	0	0	1	1	albania	nigeria	n	Foreign victims from European countries, Philippines, an...

**Figure 8: Visualization-Ready Data Generated by Scoring Rules from SAS Visual Text Analytics**

## NETWORK VISUALIZATION

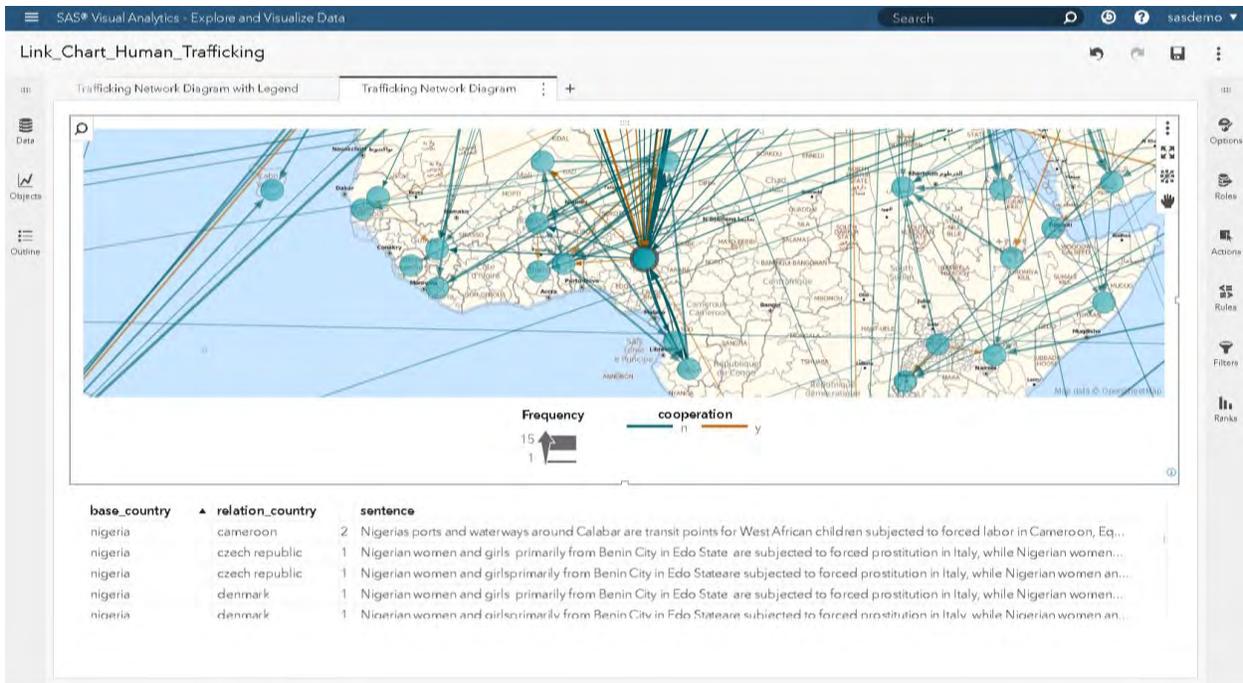
We load the data generated from the text analytics exercise into SAS Visual Analytics. The following visualizations were accomplished on SAS Viya, but similar visualizations are available on SAS 9. Once the data is loaded and the option to create a new report is selected with that data, we select the Network Analysis object for our geospatial visualization. We select the option under “Network Display” to enable a Map background, which leverages OpenStreetMap by default. We convert the base country and relation country from a categorical data variable to a geography data variable based on the country name. These are set as source and target “roles” for the Network Analysis object. The link width is set to the frequency of connections between source and target countries, enabling thicker lines for relationships that span multiple years. The link color is set to the cooperation\_indicator, highlighting links that involve cooperation between source and target countries in orange. Finally, the directionality of the links is assigned under the Link Direction option of the “Network Display” to “Source”, to show the links from source country to destination countries. The resulting diagram, initially centered around South Africa, is shown below.



**Figure 9: Network Analysis Diagram Showing Patterns of Trafficking in the Southern Hemisphere**

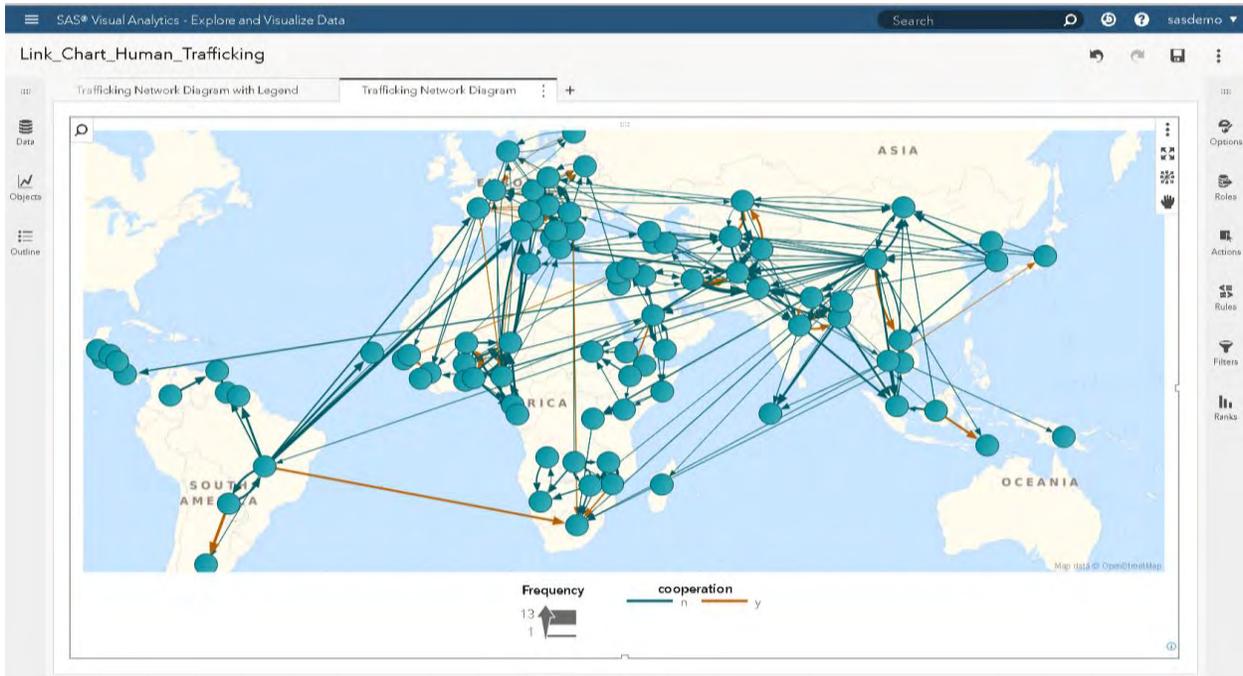
This visualization displays the interconnection between all countries across the TIP reports from 2013-2017. It highlights groups of countries involved in trafficking with each other, such as the various countries in the south of Africa as well as South America. It also highlights countries that serve as hubs for larger international trafficking patterns. For each node selected, SAS Visual Analytics displays the text associated with those connections. In this case, it highlights lines from the TIP reports identifying victims of human trafficking in South Africa from source countries including China, Taiwan, Thailand, Cambodia, India, Russia, and Brazil. From here, the text can be assessed to verify the authenticity of the links. Some links, including the link between Brazil and South Africa, are depicted in orange. This shows that SAS identified a relationship in the text between those two countries indicating that they were working together to address the trafficking problem.

Connections between Nigeria and other African countries, as well as to countries in Europe and Asia are particularly strong as shown in the diagram below. This might warrant an analysis of other circumstantial evidence surrounding Nigeria, and we will explore this further in the discussion section of this paper.



**Figure 10: Network Analysis Diagram Highlighting Patterns of Trafficking Surrounding Nigeria**

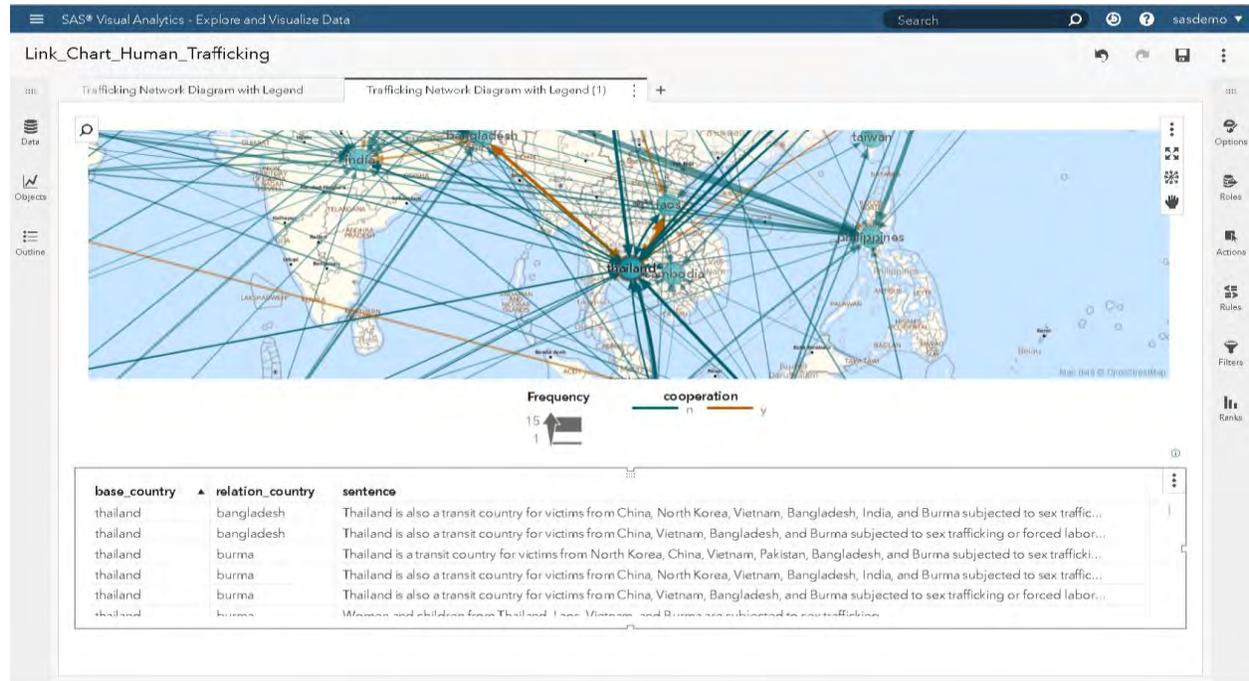
Filters can be applied that showcase certain aspects of trafficking, such as labor trafficking only, or sex trafficking only. In the visualization below, only the patterns of trafficking extracted from the TIP reports that mention children are shown.



**Figure 11: Network Analysis Diagram Depicting International Patterns of Child Trafficking**

Finally, in considering visualization and interconnectedness between countries, the single links available in the TIP reports play into a much broader picture. Reports might mention connections such as “Nigeria is a source country for trafficking in other countries including...”. These single node-to-node connections

become much more insightful when seen in the context of all the other node-to-node connections. This is particularly illuminating when countries are specifically called out as transit countries, and these connections in turn reveal second-degree and third-degree connections between source and destination countries. The following visualization reveals Thailand cited as a transit country for a variety of source and destination countries, revealing a larger pattern of international human trafficking.



**Figure 12: Network Analysis Diagram Depicting Thailand as a Transit Country for International Trafficking**

## CONCLUSION AND DISCUSSION

In this paper, we showed how SAS could be used to obtain TIP reports from the US Department of State, identify patterns across those reports, and visually depict those patterns. We used the text analysis and visualization capabilities of SAS to answer three questions. First, we identified general trends in the TIP reports. Second, we identified focused themes, including themes around victims seeking shelter internationally. Finally, we extracted a geospatial pattern across all trafficking reports between source and target countries. We then depicted this visually in a network analysis diagram. The network analysis diagram included controls for filtering on trafficking victims, trafficking type, and the year of the report. These results enhance the ability of the U.S. Government and foreign nations worldwide to engage in dialogs advancing anti-trafficking reforms, and to examine where resources are most needed.

The analysis effort to identify the source and destination countries took approximately three days of dedicated effort. Contrast this with a manual effort to extract the same information from the reports. If we approximate 30 minutes per report to identify all the relevant connections that occurred in the data, with approximately 1000 reports, this would require 3 months of effort, or 30 times the time investment. Also, consider that the automated rules can score reports in upcoming years for connections, including 2018, 2019 and beyond at little extra time investment.

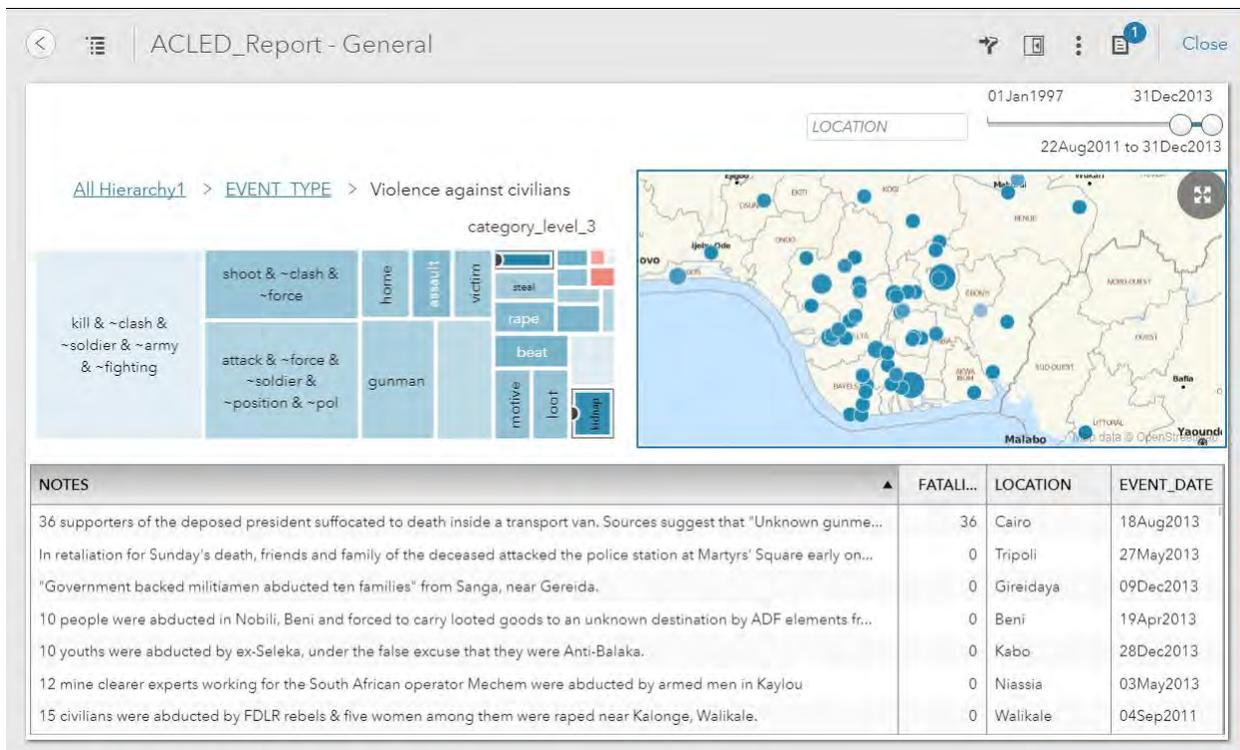
Analysis relies on the quality of the underlying data, and all analysis is fraught with challenges involving precision and recall. Precision in this case involves extracting only correct links, including getting the directionality of the connection correct. Recall involves extracting all of the links. Precision, in this analysis, can be improved by developing additional rules to ensure directionality accuracy in the links. Recall in this data set was influenced by factors including generality of information in the TIP reports. For example, the United States does not feature in any of the network links, as the United States is discussed

in general terms in the reports, indicating that the United States is a source country and destination country for trafficking with a variety of foreign nations. This means that SAS is unable to extract a specific pattern related to the United States since specific countries it is connected to are not called out in the reports directly. Such insight provides additional feedback to the analysts developing these reports in terms of where specific patterns need to be built out upon the more general patterns. Such work can enhance understanding of the international patterns between several degrees of source and destination countries.

There are several different trafficking-related use cases, including drug trafficking and weapons trafficking. These tend to tie together with the human trafficking element. Other organizations who could potentially benefit from a trafficking solution include federal, state, and local law enforcement agencies. A solution that assesses and prioritizes trafficking-related leads could be set up from a law enforcement perspective, but could also address victim assistance. Regarding data that would assist law enforcement, search engines for classified ads become a repository of data that plays into human trafficking, particularly sex trafficking. They can be assessed to identify geographically where recruitment ads are spiking, where there are similar or emerging patterns in ads, and can ultimately assist law enforcement in identifying networks of trafficking-based organizations. There is a trafficking related pattern to data from financial organizations as well, including the major banks. The Financial Crimes Enforcement Network (FinCEN) has published guidelines to banks on recognizing activity that might be associated with human smuggling and human trafficking<sup>6</sup>.

As mentioned, there are different sources of data that support the use case to assess patterns of international human trafficking. For example, to identify why Nigeria has a number of trafficking connections to a variety of countries in Africa, Europe, and Asia, we can examine data sources such as the Armed Conflict Location and Event Data project<sup>7</sup> (ACLED) to look for connections. In addition, we can apply machine learning and auto-categorization to the ACLED data as prescribed in a previous SAS Global Forum paper published in 2016<sup>8</sup>.

In the screenshot below, we used a categorical hierarchy developed with machine learning against the ACLED data to explore themes in violence against civilians in Nigeria and the surrounding regions. The visualization depicts specific recorded instances of abduction and kidnaping, and drills down to the event text describing what happened. There is significant event traffic in Nigeria, depicting a destabilizing force that contributes to the vulnerability of its citizens to human trafficking. This analysis adds to the current evidence that many Nigerians seek work abroad due to extreme poverty, and are subsequently exploited for forced labor and prostitution. The data available from the TIP reports and the ACLED project is further reinforced by an exposé by CNN, where individuals who have sought work abroad as migrants from Niger and Nigeria among other African countries are sold at a slave auction<sup>9</sup>.



**Figure 13: Visualization Depicting Kidnaping and Abduction Events in Nigeria and the Surrounding Countries Using Data from the ACLED Project**

In summary, the analytics and visualizations presented here are an effort to show how data related to human trafficking can be transformed into actionable information. By taking advantage of data and analytics, data scientists and researchers are able to shine light on the problem, and thereby help international government, law enforcement, and victims advocacy groups find better ways to address it<sup>10</sup>.

## REFERENCES

1. U.S. Department of State. 2017. "Trafficking in Persons Report." Accessed February 2, 2018. <https://www.state.gov/j/tip/rls/tiprpt/>.
2. The Polaris Project. 2018. "The Facts." Accessed February 2, 2018. <https://polarisproject.org/human-trafficking/facts>.
3. Figallo-Monge, Manuel. "Pedal-to-the-Metal Analytics with SAS® Studio, SAS® Visual Analytics, SAS® Visual Statistics, and SAS® Contextual Analysis" Proceedings of the SAS Global Forum 2016 Conference. Cary NC: SAS Institute Inc. Available <http://support.sas.com/resources/papers/proceedings16/SAS6560-2016.pdf>.
4. Albright, Russ. Cox, James. Jin, Ning. 2016. "Getting More from the Singular Value Decomposition (SVD): Enhance Your Models with Document, Sentence, and Term Representations" Proceedings of the SAS Global Forum 2016 Conference. Cary NC: SAS Institute Inc. Available <https://support.sas.com/resources/papers/proceedings16/SAS6241-2016.pdf>.
5. Sabo, Tom. 2017. "Applying Text Analytics and Machine Learning to Assess Consumer Financial Complaints." *Proceedings of the SAS Global Forum 2017 Conference*. Cary NC: SAS Institute Inc. Available <http://support.sas.com/resources/papers/proceedings17/SAS0282-2017.pdf>.

6. United States Department of the Treasury, Financial Crimes Enforcement Network. 2014. "Advisory Information". Accessed February 8, 2018. <https://www.fincen.gov/resources/advisories/fincen-advisory-fin-2014-a008>.
7. ACLED Data; Bringing Clarity to Crisis. 2018. "About". Accessed February 12, 2018. <http://www.acleddata.com/>.
8. Sabo, Tom. 2016. "Extending the Armed Conflict Location and Event Data Project with SAS® Text Analytics." Proceedings of the SAS Global Forum 2016 Conference. Cary NC: SAS Institute Inc. Available <https://support.sas.com/resources/papers/proceedings16/SAS6380-2016.pdf>.
9. CNN. 2017. "Migrants being sold as slaves." Accessed February 12, 2018. <http://www.cnn.com/videos/world/2017/11/13/libya-migrant-slave-auction-lon-orig-md-ejk.cnn>.
10. SAS. 2017. "Analytics tackles the scourge of human trafficking." Accessed February 12, 2018. [https://www.sas.com/en\\_us/insights/articles/analytics/analytics-tackles-human-trafficking.html](https://www.sas.com/en_us/insights/articles/analytics/analytics-tackles-human-trafficking.html).

## ACKNOWLEDGMENTS

Thanks to Emily McRae and John Dillman for assisting with the visualizations, and thanks also to Emily for her review of this paper. Thanks to Mary Beth Ainsworth for her insight into the human trafficking problem nationally and internationally. Also, thanks to Dr. James R. Van Scotter for providing awareness into how local and state law enforcement can use data to identify human trafficking trends within the U.S.

## RECOMMENDED READING

- 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](http://www.sas.com/en_us/whitepapers/text-analytics-in-government-106931.html).
- Chakraborty, G., M. Pagolu, S. Garla. 2013. *Text Mining and Analysis; Practical Methods, Examples, and Case Studies Using SAS®*. SAS Institute Inc.

- 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>
- 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>.
- Reamy, Tom. 2016. *Deep Text; Using Text Analytics to Conquer Information Overload, Get Real Value from Social Media, and Add Big(ger) Text to Big Data*. Medford NJ: Information Today, Inc.

## CONTACT INFORMATION

Your comments and questions are valued and encouraged. Contact the authors at:

Tom Sabo, Principal Solutions Architect  
1530 Wilson Blvd.  
Arlington, VA 22209  
SAS Federal LLC  
+1 (703) 310-5717  
[tom.sabo@sas.com](mailto:tom.sabo@sas.com)  
@mrTomSab

Adam Pilz, Senior Solutions Architect  
121 W Trade St.  
Charlotte, NC 28202  
SAS Institute Inc  
+1 (919) 348-6039  
[adam.pilz@sas.com](mailto:adam.pilz@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.