

Liberia Media Analysis

By Nathan Danielsen



Problem

- How do we assess the impact and measure the effectiveness of USG government funded media capacity building projects?
- Current Solutions are:
 - Labor intensive (manual scoring)
 - Expensive (staff and training time)
 - Small samples (< 500)
 - Scored content is hard to access
 - Subjective
 - Not real time

Guideline/Topic	Media Coverage	
	n*	Comments
Choose a diet low in fat, saturated fat, and cholesterol.	55	Coverage concentrated on total fat intake and consuming low-fat diets to lose weight.
Balance the food you eat with physical activity. Maintain or improve your weight.	52	35 of 52 addressed weight loss/ maintenance in conjunction with exercise. 17 of 52 addressed weight loss/ maintenance.
Choose a diet with plenty of grain products, vegetables, and fruits.	26	9 of 26 only mentioned fruits and vegetables. Consumers were warned that most Americans do not consume enough vegetables and fruits.
Actual release of the new 1995 Dietary Guidelines for Americans	22	Discussion centered around the differences from the 1990 Dietary Guidelines document.
If you drink alcoholic beverages, do so in moderation.	18	Coverage concentrated on the health benefits of drinking alcohol in moderation.
Eat a variety of foods.	4	Articles appeared in January, close to the time when the Guidelines were released.
Choose a diet moderate in salt and sodium.	4	Articles appeared in January, close to the time when the Guidelines were released.
Choose a diet moderate in sugars.	4	Articles appeared in January, close to the time when the Guidelines were released.

*number of stories that reported the guideline or topic.

A Better Way?

- Automating Data Collection (content scraping)
- Modeling Media Coverage As a Baseline Assessment
- Reproducibility (Raw data is saved for further analysis)
- Reusable tools/ code
- More objective?



Research Questions

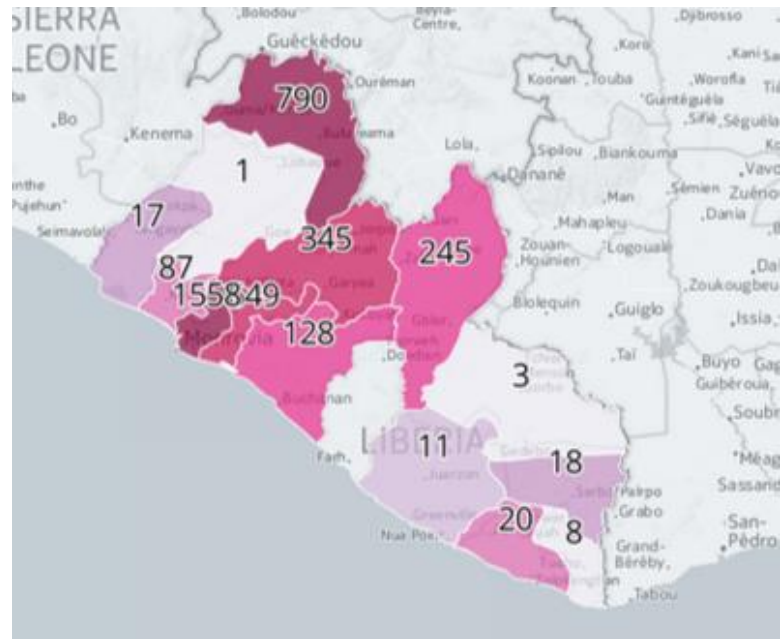
- Do local media outlets in Liberia adequately provide media coverage of the entire country?
- What are the predictors of media coverage in Liberia?
- Test Case
 - How did local media outlets in Liberia cover events before, during and after the Ebola crisis?
 - Was media coverage representative of the geographical scope and extent of the Ebola outbreak?

Famous Images from the Ebola Outbreak



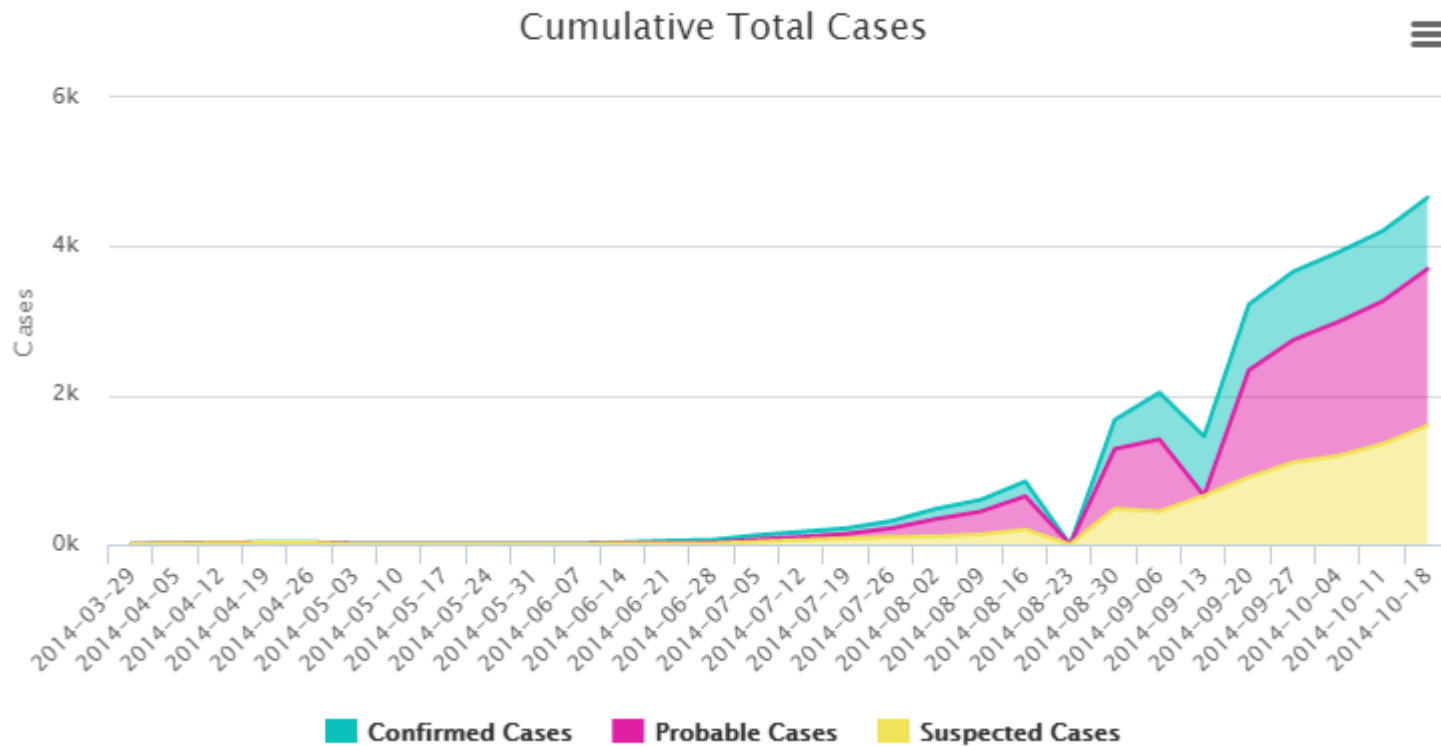
Hypothesis

I hypothesize that timely and accurate media coverage of significant events in Liberia is determined by population density, availability (and functionality) of telecommunications infrastructure, levels of literacy, and ethnic and linguistic divisions.



Source: <http://ebolainliberia.org/>

Visualization (Not Mine)



Source: <http://ebolainliberia.org/>

Data Collection Plan

- Scrape all the online media content “stories” from 9 local media outlets.
- Demographic and Health Survey Data
 - Accurate and representative data on population, health, HIV, and nutrition
 - Fairly Tidy
- Ebola Time Series by Region Data from HDX
 - Humanitarian Data Exchange (HDX)
 - Fairly Tidy

Media Content Tasks

- Build Scraper (Nearly Done)
- Data Extraction and Cleaning (Nearly Done)
 - *Title*
 - *Post Date*
 - *Author*
 - *Story Lead (First ~4 sentences)*
 - *Entire Story*
- Feature Extraction (Partially done)
 - *Named Entity Recognition (Names, Locations, Organizations, etc)*
 - *Story Category*
 - *Keywords*
 - *Sentiment*
 - *LDA*

Modeling Plan

- Time Series of Media Coverage (# stories per day, week, month covering a region, story category etc.)
- Determine predictors of Media Coverage with DHS data using logistical regression and clustering
- Model sentiment by publication and author for regions and/or story category.

Modeling Test Cases

- As the Ebola outbreak starts, does news coverage follow casualty counts?
 - Is there a correlation?
- Since the Ebola outbreak has receded, has media coverage returned to the pre-outbreak model?

Project Status

- Web Scraper Issues
 - Collection Rules:
 - False Positives (Redirects from a page that does not exist to the homepage. No 404 error and scraper collects it)
- Media Content Feature Extraction
 - Extraction Rules:
 - Need to customize scraper to each media outlet
- Modeling
 - Haven't started yet.

