

# Zika misinformation tracking in social media



Presenter: Amira Ghenai- PhD  
Candidate U.Waterloo

Project with:

- Yelena Mejova
- Luis Fernandez Luque

Date: 19/09/2016

# Outline

- Zika outbreak
- Project Goal
- Data Description
- Location extraction
- Topic extraction
- Labeling
- Timeline

# Zika Outbreak

- The Zika virus infection across the Americas is considered a serious outbreak nowadays
- WHO has declared an international health alert
- PAHO & WHO send key messages to the population to minimize the risks (mosquito control, avoiding mosquito bites and pregnancy risks)

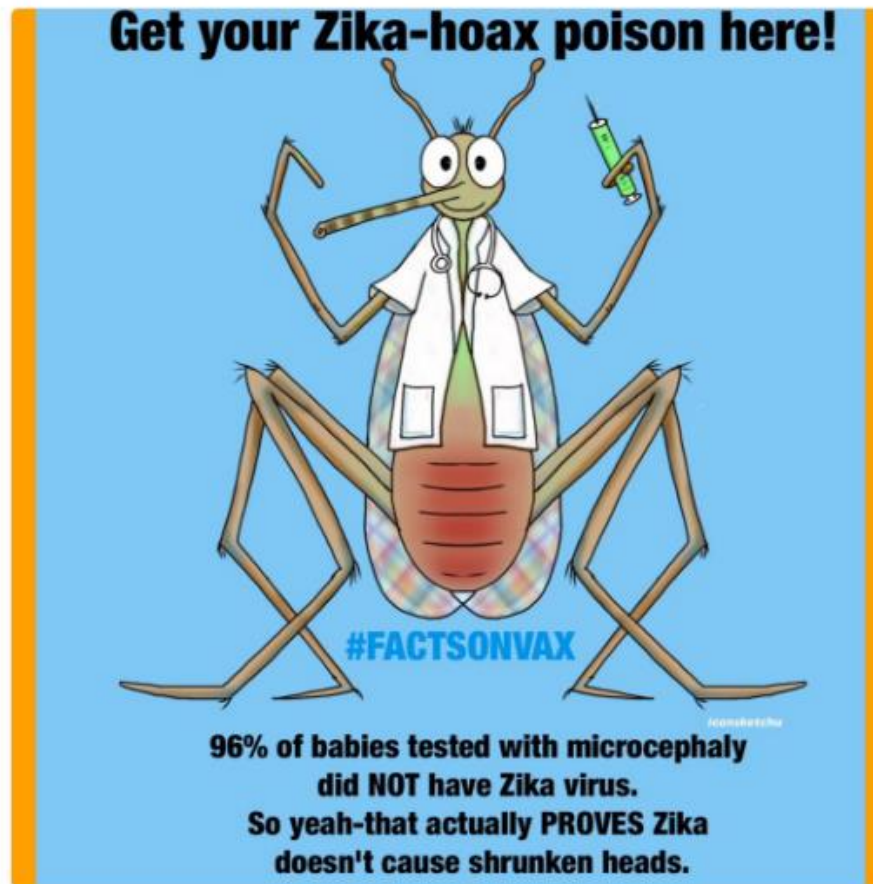
# Vaccines cause microcephaly in babies

in reply to Ex\_MisITech



**macesmom** @HealthyATV · Jun 15

@Ex\_MisITech #Zika is a sick joke on the uninformed. Mandated vaccines 2 pregnant moms causes **microcephaly** of course. 🐛



# Microcephaly is caused by genetically modified mosquitoes



**Peter Shinn** @prolifeunity · Sep 17

Think about it. **Mosquitoes** didn't affect babies until researchers started **genetically** modifying them.



**Zika Study: 'Prepare for Global Epidemic of Micro...**

A new study published in Lancet reveals a "striking magnitude of association between microcephaly and Zika virus infection."

[breitbart.com](http://breitbart.com)



# Fish can help stop Zika

**CORDS**

**CORDS** @CORDS\_HQ · Sep 17

A tiny **fish** to fight spread of **#zika** and other vector borne diseases - new research shows promise @APEIRCO



## **A New Weapon to Fight Mosquito-Borne Diseases: Tiny Fish**

Scientists hope the simple solution can help slow the spread of dengue, malaria, Zika, and other illnesses.

[takepart.com](http://takepart.com)

# Project Goal

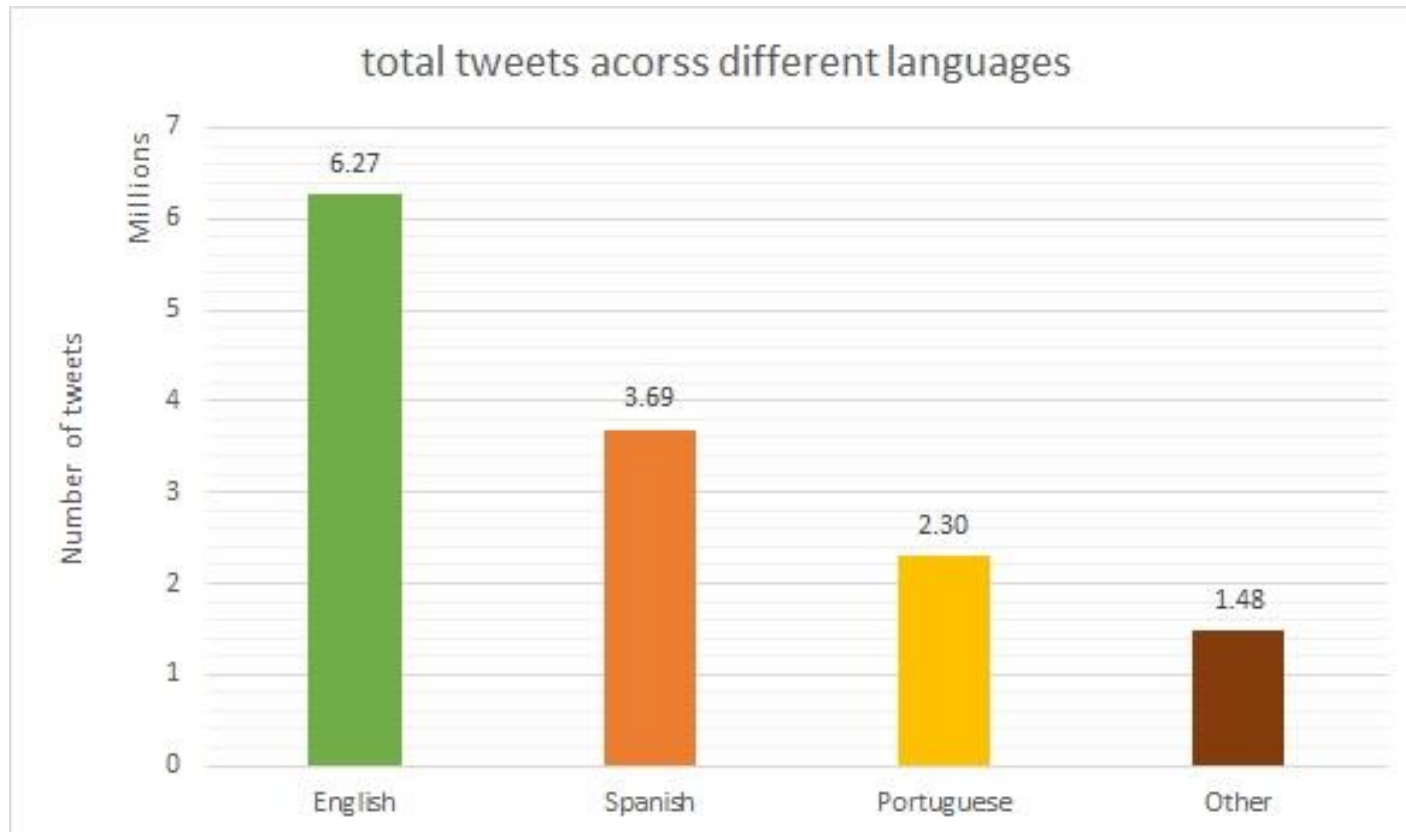
- The objective of this project is to study the feasibility of using social media monitoring as a tool to help the communication effort in the health crisis
- Monitor potential threats to the communication effort such as people spreading rumors and misinformation about Zika infection to the world

# Data Description

- Twitter data collection is happening on AIDR.
- Keywords related to Zika: microcephaly, Zika, Aedes, Zika fever...
- Period from 2016-01-13 to 2016-08-22
- All languages
- Total collected tweets ~ 13 million tweet



# Data Description



# Location Extraction

- Location is important
- To extract the exact country name from tweets is not a trivial task.
- The explained method gives a very high coverage for the tweets to locate by country name

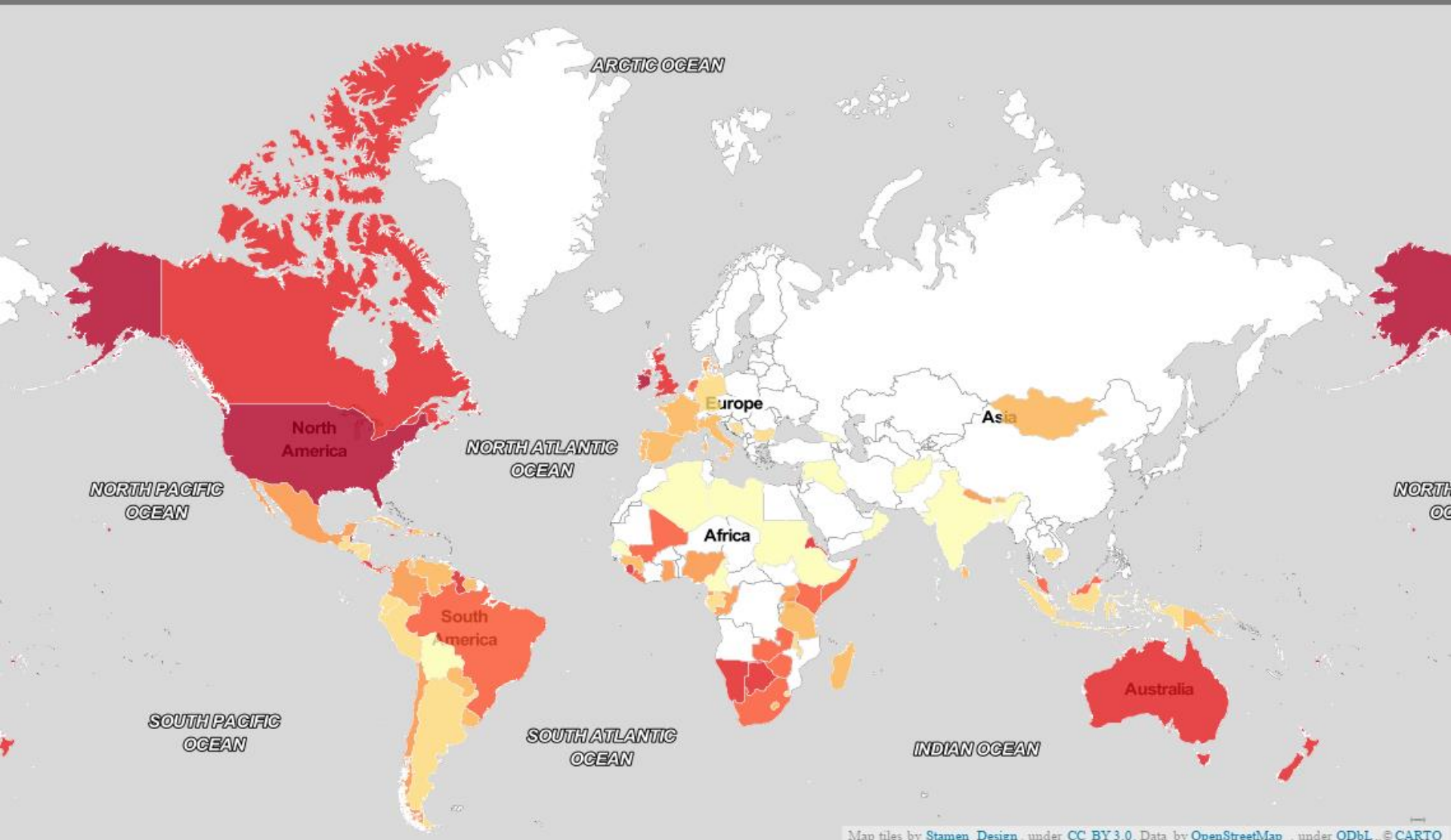
# Location Extraction

- Very high coverage for the tweets to locate by country name

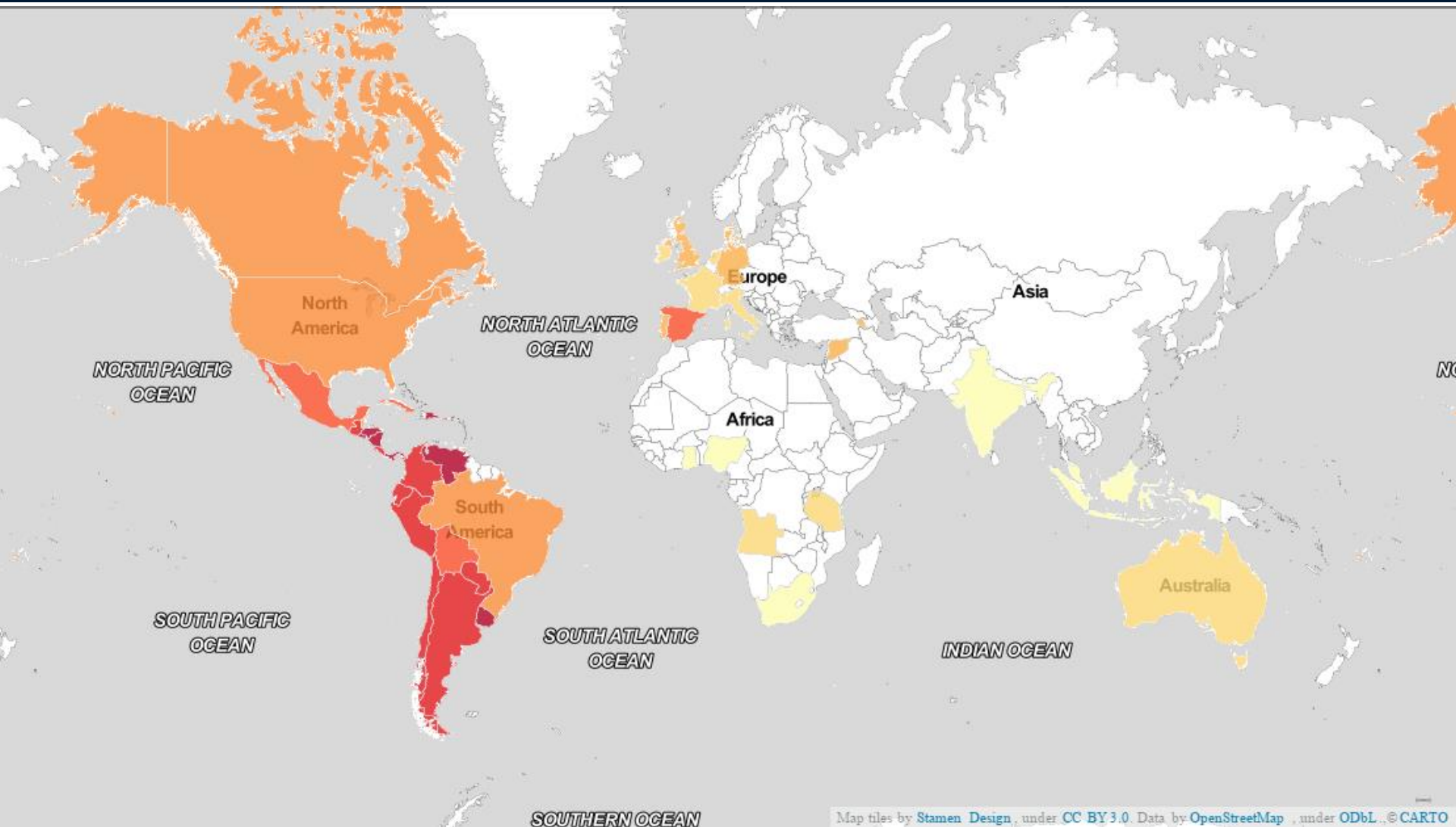
| Language   | Coverage Percentage |
|------------|---------------------|
| English    | 68%                 |
| Spanish    | 63%                 |
| Portuguese | 64%                 |

- English is spread in more than one location
- Spanish: Most tweets come from the southern Americas
- Portuguese: Most tweets are located in Brazil

# Location Extraction – English Map



# Location Extraction – Spanish Map



# Topic Extraction

- Automatic topic extraction
  - Latent Dirichlet allocation – LDA
- Preprocessing:
  - Remove stop words/ highly frequent words
  - Remove Twitter special characters
  - Lower case
  - Tokenization
  - Stemming

# Topic Extraction

- Run LDA for:
  - English
  - Spanish
  - Portuguese

# Topic Extraction

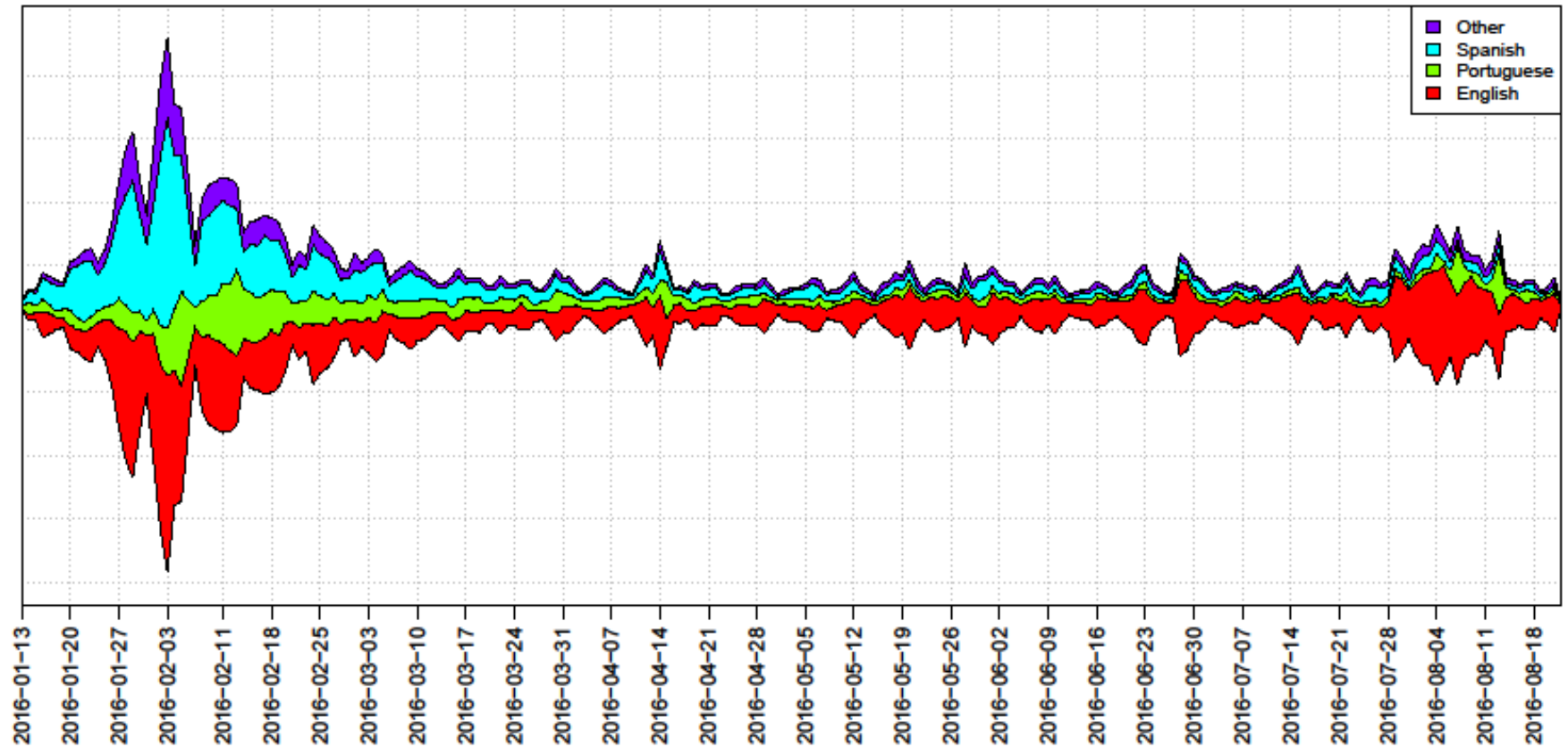
- Example of English top 5 topics:
  - women\_pregnant\_travel\_cdc\_warn
  - case\_first\_confirm\_report\_transmit
  - birth\_caus\_babi\_microcephali\_link
  - spam\_just\_like\_blood\_look (weird!!)
  - mosquito\_help\_fight\_control\_can



# Labelling

- Each topic comes with a set of words that best describe it, and we also extracted most related tweets associated with this topic
- Topic classification of LDA tweets: (manual)
  - Spam / hashtag
  - Misuse
  - Joke
  - Reporting of a specific case involving zika
  - General zika information
  - Advise about zika
  - Misinformation about zika

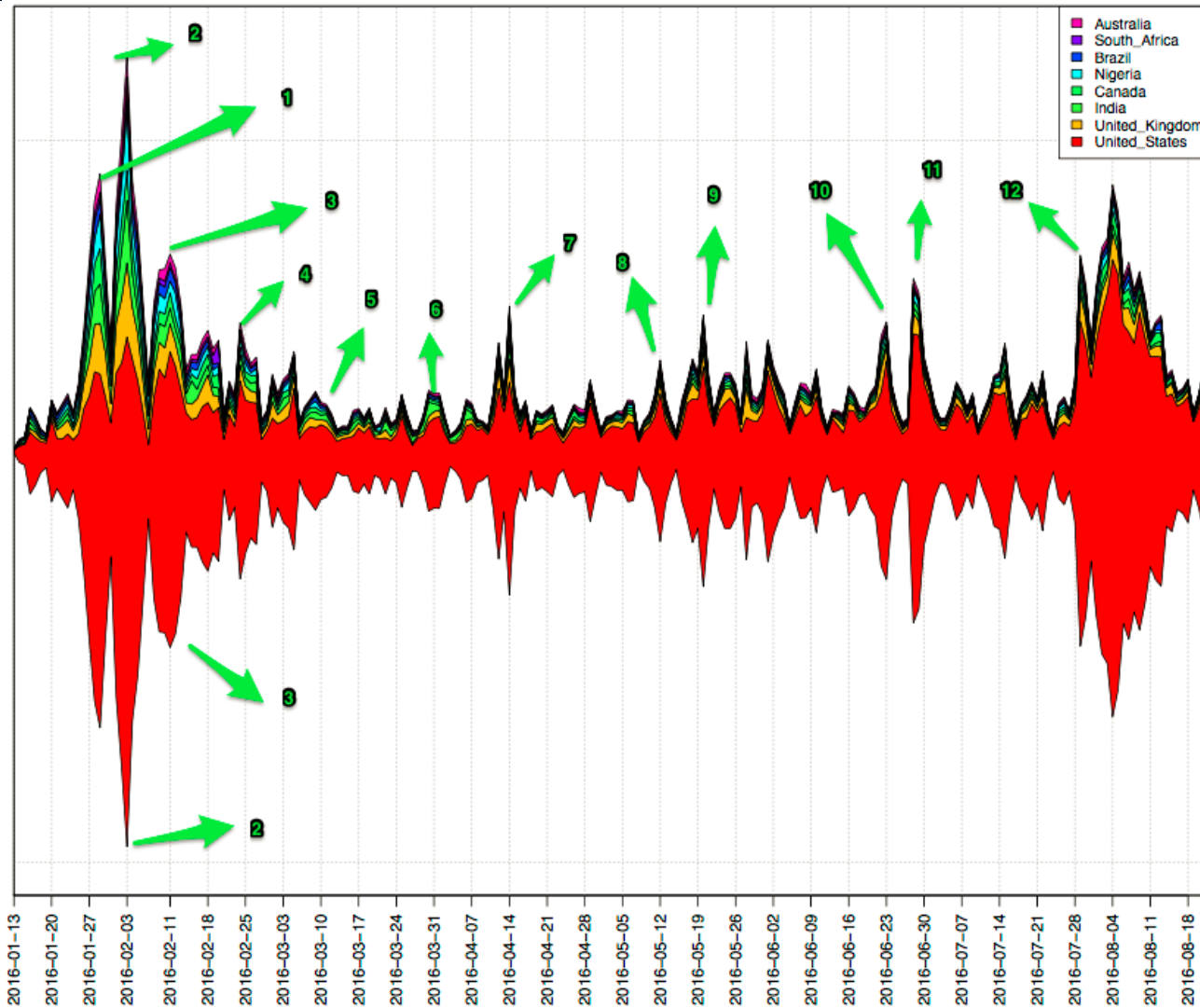
# Timeline – Language Volume



# Timeline – LDA Topics / English



# Timeline – Country distributions



# Future plans

- Improve LDA results
- Find better way to extract rumors and misinformation from the dataset