Sentiment Analysis of Twitter Response to Major World Events

Juan Pablo Arrivillaga
Kyle Ferris
Dong Hyuk Kim
Kaustav Saha

Problem Description

What is the problem?

- Social media makes available unprecedented amounts of unstructured content concerning the sentiments and opinions of populations.
- How can we make sense of this?

Why is important?

- Understanding how the populations of different states react differently to events will help to increase understanding of cultural and political differences, and potentially aid in policy decisions and political campaigning.

Problem Description

Who cares about it?

- Users could include politicians and political activists.
    - Great utility to decision-makers and the authorities.

Why does it remain unsolved?

- The performance on social media data is still unsatisfactory due to the distinct data characteristics
    - Social media posts are always short and unstructured.
    - It is labor intensive and time consuming to obtain ground truth for training data.

Objectives

- Monitor the most talked about messages at any given time to chronicle the civilian response to major events.
- Reduce Dimensionality.
- Compare generated topics from different populations.
- Perform a visualization of the emotions obtained on a world map to understand the sentiments from different parts of the world.

Related Work

Hughes and Palen (2009)

- Surveyed the adoption and use of Twitter during mass convergence and emergency events.

Cheong and Lee (2011)

- Assigned responses into seven distinct categories: Fear/Anxiety, Shock, Response, Need for information and Updates, Threat Assessment, Casualties Assessment, and Response and Law Enforcement.

Jungherr (2009)

- Detailed the role of Twitter in social activism and looked into case studies whereby Twitter was an instrumental tool in disseminating information on terrorist attacks, political dissent, and acts of oppression.

Data-Extraction

- Paris Attack - Nov 12/13
- Tweets over the period of November 17, 2015 to November 26, 2015
- Divided our final data (60,998 tweets) into two subsets
  - Tweets from 39 major cities (4,673 tweets)
  - Tweets from around 1000 cities around the world (56,325 tweets)
- Data cleaning
  - Removed whitespace
  - Ignored case
  - Removed punctuation

Models

Latent Dirichlet Allocation (LDA)

- Unsupervised machine learning technique that identifies latent topic information.
- A member of a family of models known as probabilistic topic models.

Pachinko Allocation Model (PAM)

- Captures correlations using a Directed Acyclic Graph (DAG), where topic nodes occupy the interior levels and the leaves are words.

Approach

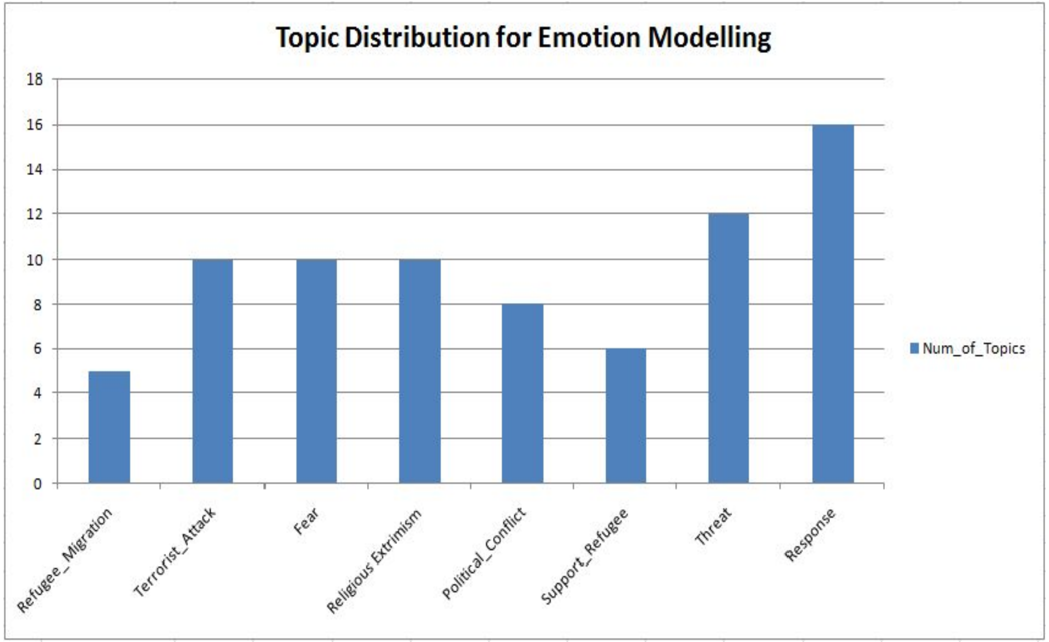
Analyses

After performing topic modeling on the corpus with a total topic count of 400, we manually went through all 400 generated topics and selected those that were coherent and related to one of our topics of interest.

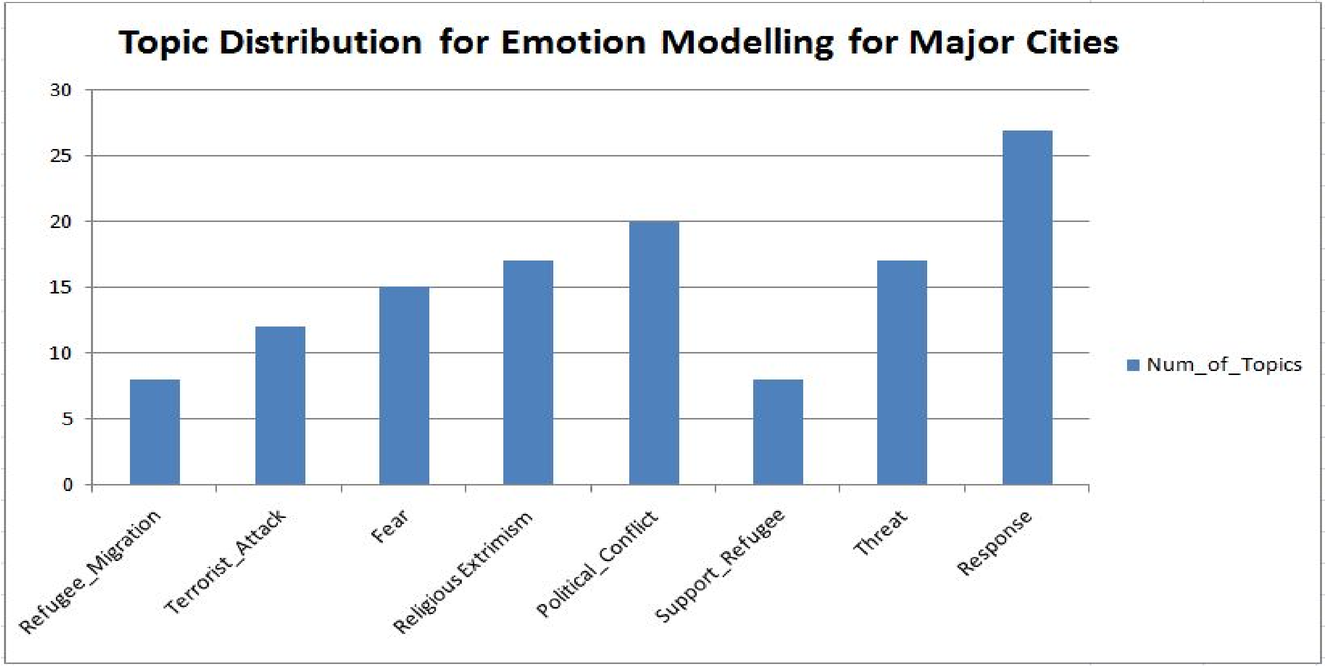
For Pachinko we had
200 sub-topics grouped into
20 super-topics.

Fear	Anxiety, Anxiousness, Catastrophic, Concern, Disaster, Fear, Insecure, Panic, Scared, Terror, Trouble, Warning, Worry
Political Conflict	Assad,Syria,Russia,Turkey,Policy,EU,Emergency,Embargo, Sanction,Xenophobia,France,USA
Refugee Migration	Open,Border,Ship,Boat,Refugees,Racist,Child,Women,Sleep,Food, Transport
Religious Extremism	Bigotry,Shame,Muslim,Execute,Trust,Intolerance,Hindu,Christian, Islam,Hate,Foul Play
Response	Breaking News, Call, Incident, Phone, Report, Situation, Unconfirmed,Act, Asap, Escape, Evacuate, Flee, Help, Hide, Run
Support Refugee	Welcome,Refugee,Commitment,Homeless,Support,Obama, Republican,Asylum,Trump
Terrorist Attack	Attack, Bomb, Bullet, Collapse, Crash, Explosion, Fire, Gun, Hijack, Hit, Hostage, Plane, Responsibility,Rifle, Shoot, Struck, Suicide, Terrorist
Threat	Threats, Accident, Shock, Aback, Floor, God, Bless, Omg, Shock, Stun, Sudden

Larger Sample- This process resulted in a total of 77 relevant topics out of 400. Then, these topics were manually classified into one of our eight topic groups.



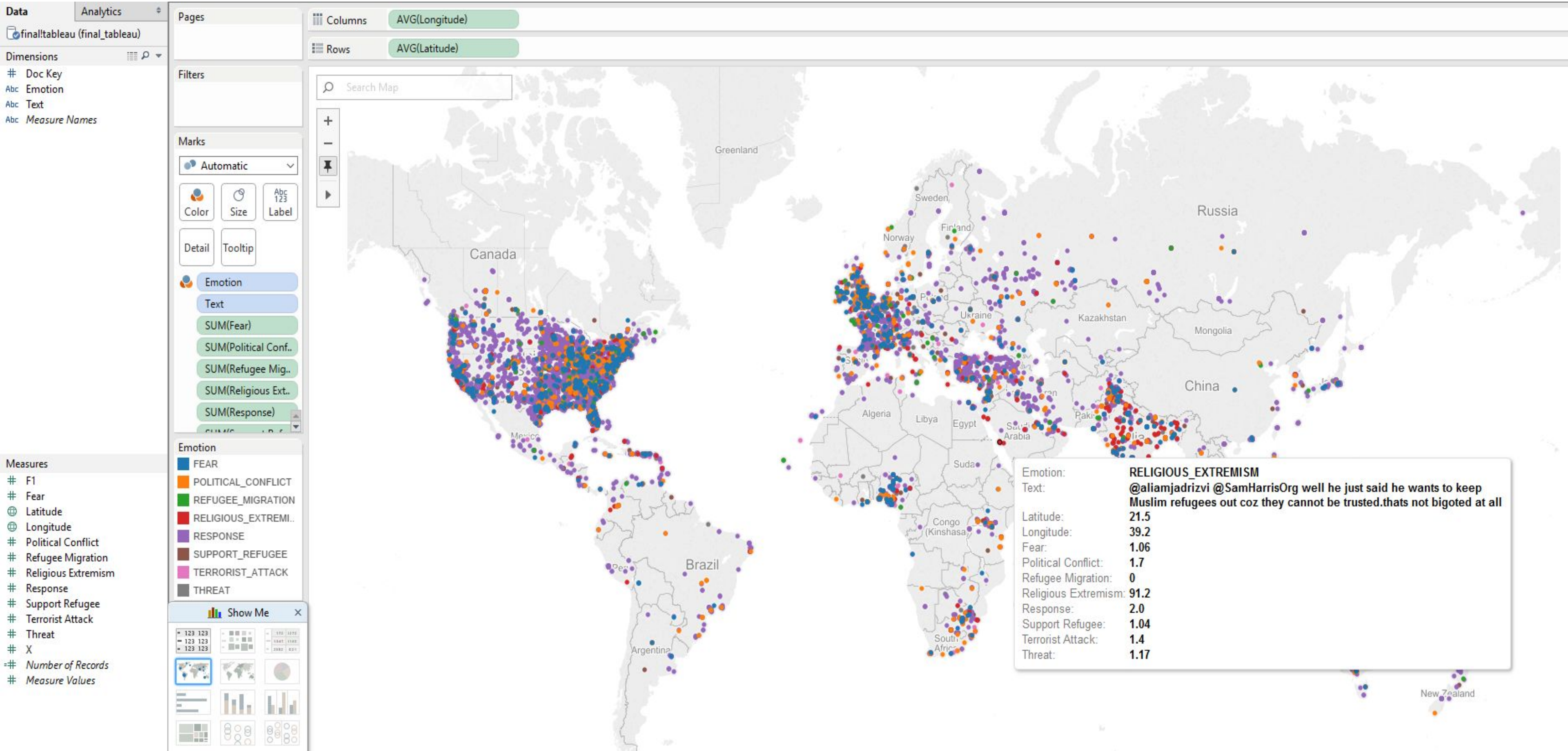
Topic Distribution for Major Cities

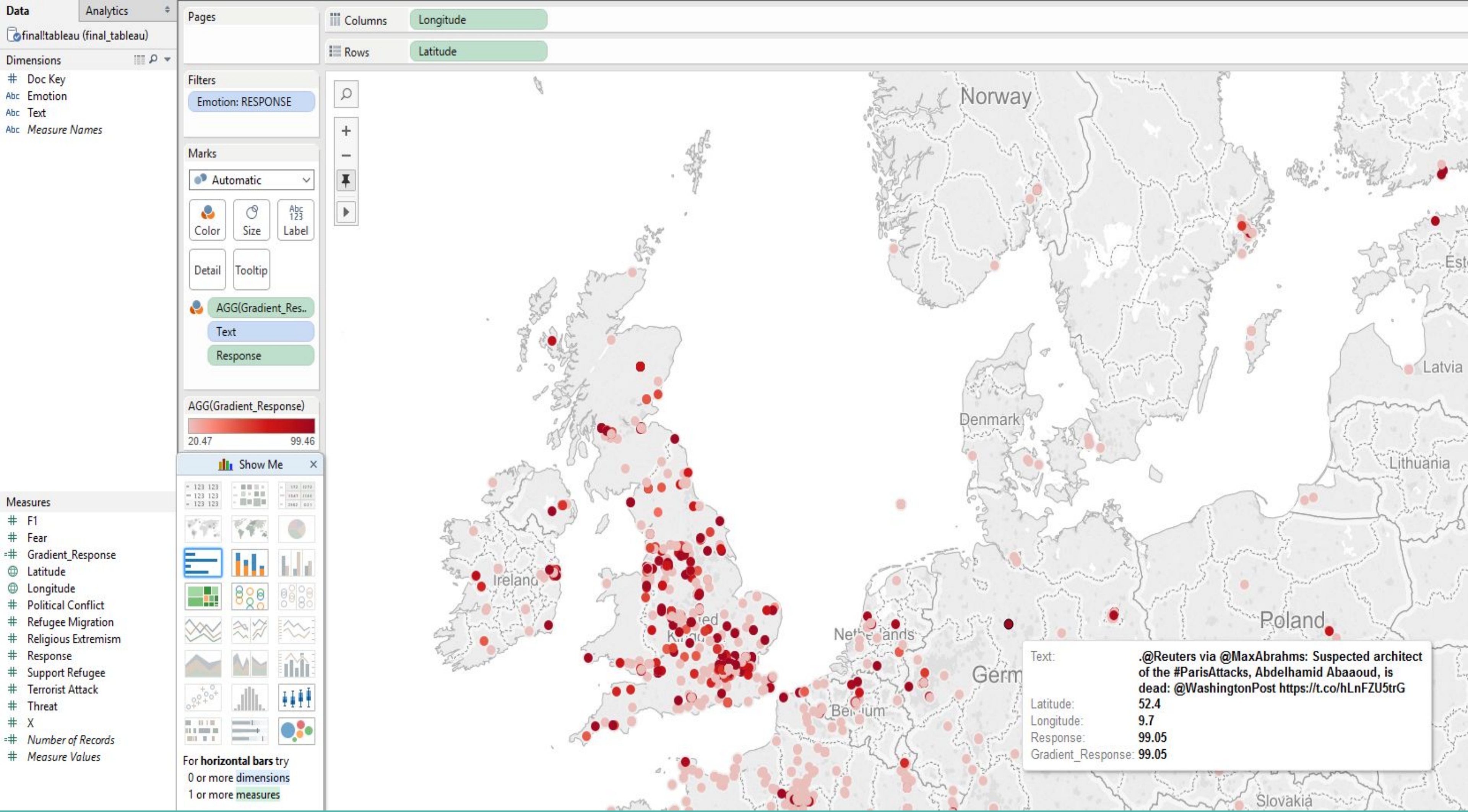


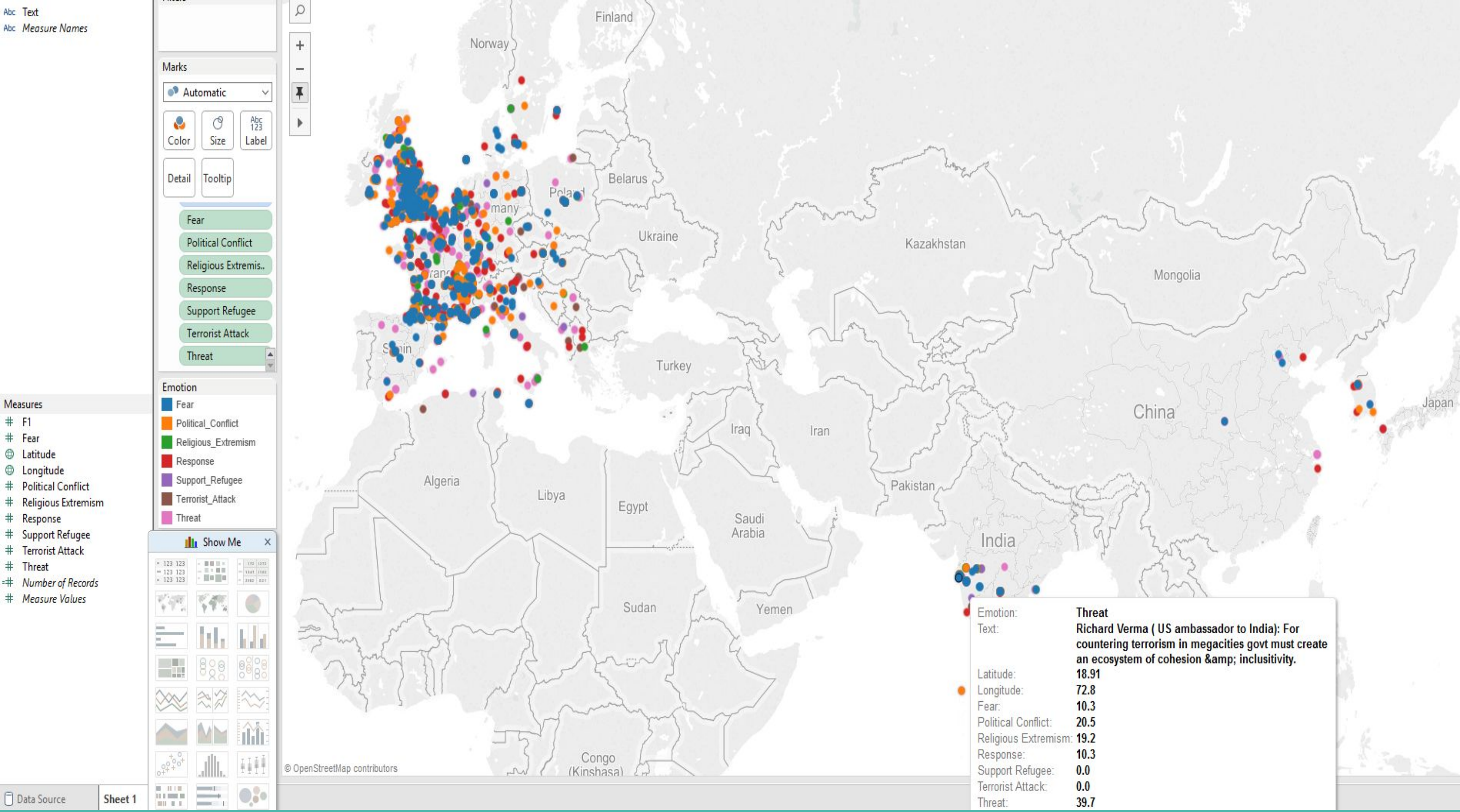
Total of

124 topics

Results


Data
Analytics
final!tableau (final_tableau)
Dimensions
Doc Key
Emotion
Text
Measure Names
Measures
F1
Fear
Latitude
Longitude
Political Conflict
Refugee Migration
Religious Extremism
Response
Support Refugee
Terrorist Attack
Threat
X
Number of Records
Measure Values
Pages
Filters
Marks
Automatic
Color
Size
Label
Detail
Tooltip
Emotion
Text
SUM(Fear)
SUM(Political Conf..
SUM(Refugee Mig..
SUM(Religious Ext..
SUM(Response)
Emotion
FEAR
POLITICAL_CONFLICT
REFUGEE_MIGRATION
RELIGIOUS_EXTREMI..
RESPONSE
SUPPORT_REFUGEE
TERRORIST_ATTACK
THREAT
Show Me
Columns
AVG(Longitude)
Rows
AVG(Latitude)
Search Map
Greenland
Canada
Russia
Sweden
Finland
Norway
Ukraine
Kazakhstan
Mongolia
China
Algeria
Libya
Egypt
Brazil
Argentina
New Zealand
Emotion: RELIGIOUS_EXTREMISM
Text: @aliamjadrizvi @SamHarrisOrg well he just said he wants to keep Muslim refugees out coz they cannot be trusted.thats not bigoted at all
Latitude: 21.5
Longitude: 39.2
Fear: 1.06
Political Conflict: 1.7
Refugee Migration: 0
Religious Extremism: 91.2
Response: 2.0
Support Refugee: 1.04
Terrorist Attack: 1.4
Threat: 1.17


Data
Analytics
final!tableau (final_tableau)
Dimensions
Doc Key
Emotion
Text
Measure Names
Measures
F1
Fear
Gradient_Response
Latitude
Longitude
Political Conflict
Refugee Migration
Religious Extremism
Response
Support Refugee
Terrorist Attack
Threat
X
Number of Records
Measure Values
Pages
Filters
Emotion: RESPONSE
Marks
Automatic
Color
Size
Label
Detail
Tooltip
AGG(Gradient_Res..
Text
Response
AGG(Gradient_Response)
20.47
99.46
Show Me
For horizontal bars try
0 or more dimensions
1 or more measures
Columns
Longitude
Rows
Latitude
Norway
Denmark
Latvia
Lithuania
Ireland
Netherlands
Poland
Slovakia
Text: .@Reuters via @MaxAbrahms: Suspected architect of the #ParisAttacks, Abdelhamid Abaaoud, is dead: @WashingtonPost https://t.co/hLnFZU5trG
Latitude: 52.4
Longitude: 9.7
Response: 99.05
Gradient_Response: 99.05


Abc Text
Abc Measure Names
Marks
Automatic
Color
Size
Label
Detail
Tooltip
Fear
Political Conflict
Religious Extremis..
Response
Support Refugee
Terrorist Attack
Threat
Emotion
Fear
Political_Conflict
Religious_Extremism
Response
Support_Refugee
Terrorist_Attack
Threat
Measures
F1
Fear
Latitude
Longitude
Political Conflict
Religious Extremism
Response
Support Refugee
Terrorist Attack
Threat
Number of Records
Measure Values
Show Me
Data Source
Sheet 1
Finland
Norway
Belarus
Poland
Germany
Ukraine
Kazakhstan
Mongolia
France
Spain
Turkey
Iraq
Iran
China
Japan
Algeria
Libya
Egypt
Saudi Arabia
Pakistan
India
Sudan
Yemen
Congo (Kinshasa)
© OpenStreetMap contributors
Emotion: Threat
Text: Richard Verma (US ambassador to India): For countering terrorism in megacities govt must create an ecosystem of cohesion & inclusivity.
Latitude: 18.91
Longitude: 72.8
Fear: 10.3
Political Conflict: 20.5
Religious Extremism: 19.2
Response: 10.3
Support Refugee: 0.0
Terrorist Attack: 0.0
Threat: 39.7

Limitations

- Tweets need to be Geo-Tagged.
- Contents written in different languages but in the same (Latin) Alphabet.

Evaluation

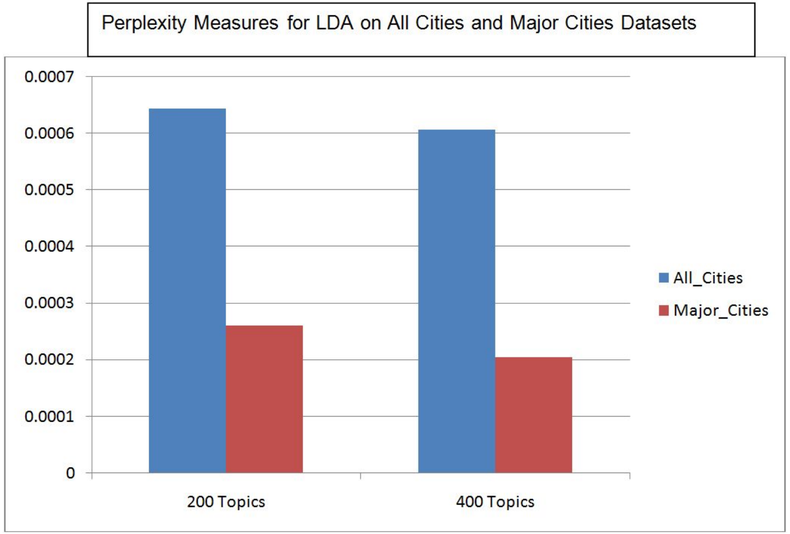
As there are distinct differences in super-topics discussed by different regions, we consider our project to be a success.

LDA Emotions:

-UK: Political Conflict
-Middle East: Religious Extremism
-Eastern Europe: Refugee migration

Pachinko Allocation Emotions:

-UK : Political Conflict and Fear
-India: Threat



Conclusion and Recommendations

- Distinct difference in Opinion/Emotion in different regions of the world.
    - Topic modelling can provide distinction of tweets across geographic areas
- Varying Gradient of an Emotion in a Region.

- Expand work to Languages other than English.
- Scalability of Pachinko.
- Spatial clustering.
- Human-in-the-loop validation of topics - AWS Mechanical Turk?


Questions & Answers