Movers

Exploring Refugee Movement with Social Media Data

The Pennsylvania State University

Geography
Mark Simpson
Xi Liu
Yanan Xin

Information Science & Technology
Feng Sun
Ying Xu

Political Science
Fridolin Linder

Co-instructors
Alan MacEachren
Alexander Savelyev

Understanding Spatial Data (Big and Small) with Visual Analytics
SpatialVA2016, GIScience 2016 Montreal, QC, Canada

PennState

Background

- Refugee crisis in Middle East and Europe
  - *2014, 59.5 million people*
  - *One in five are Syrians*
- Governments and NGOs need data to help
  - *When, where, how many?*

Background

- Yearly and monthly reports by UNHCR
  - *United Nations High Commissioner for Refugees*
  - *Data is aggregated and coarse*

The Potential

- Social media as data source
  - *Lots of data!*
  - *But difficult to find useful information*

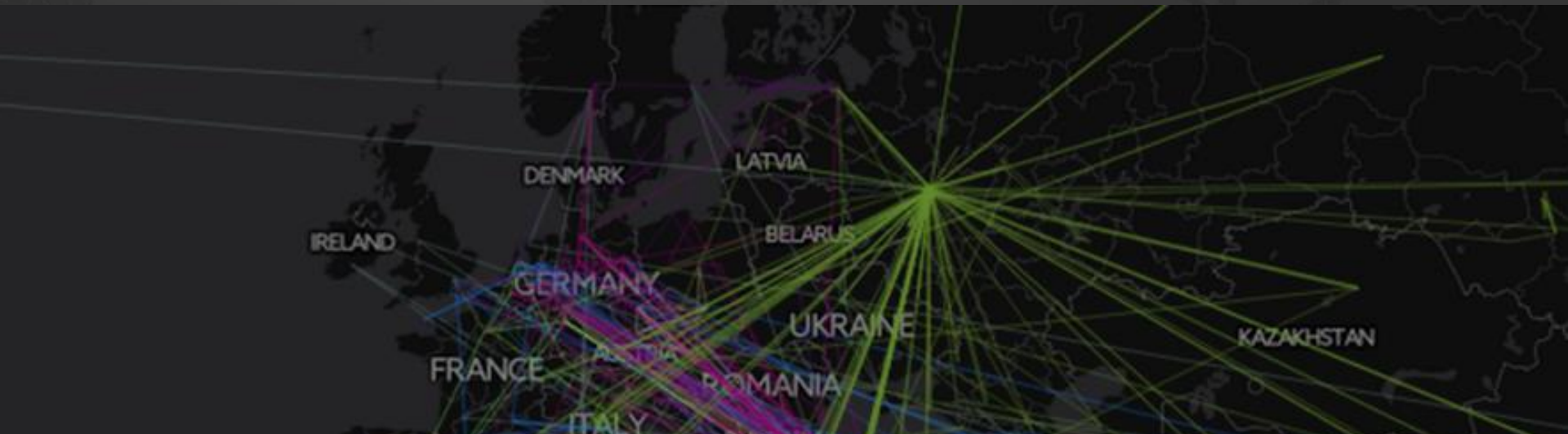
Refugees speaking to relatives in Syria from Berlin

http://www.independent.co.uk/voices/comment/surprised-that-syrian-refugees-have-smartphones-well-sorry-to-break-this-to-you-but-youre-an-idiot-10489719.html

Our Goals

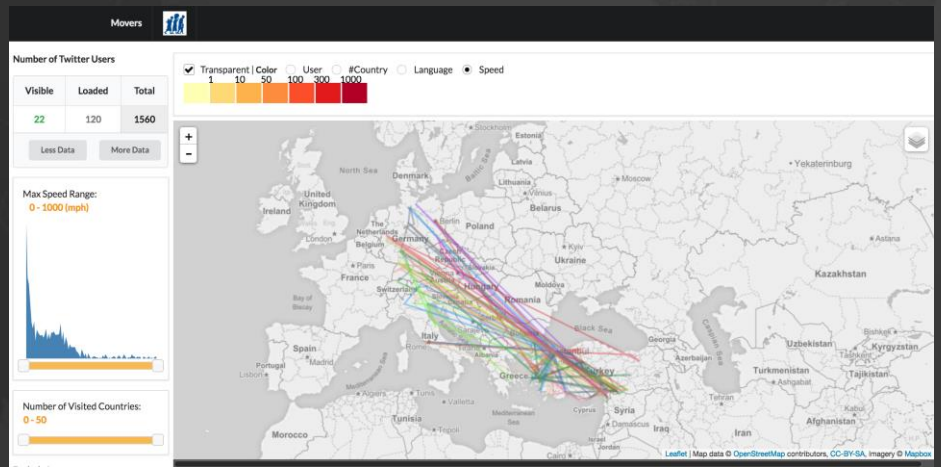
- Develop a visual analytics system to filter moving entities based on their properties

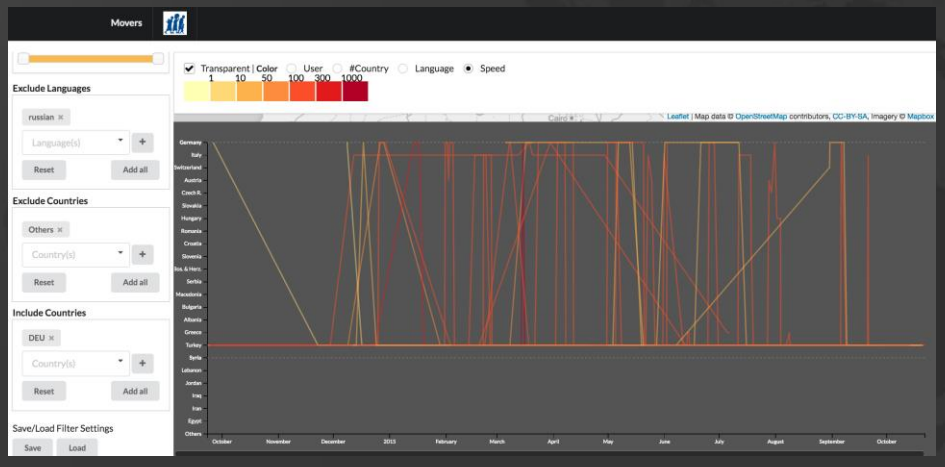
- Test if it is possible to retrieve activity information of refugees from social media.



Movers

- Web-based visual analytics system
- Intended to help identify movers
    - Refugees from twitter as initial target
    - Generalizable for different types of travelers
- Mostly focuses on subtractive filtering





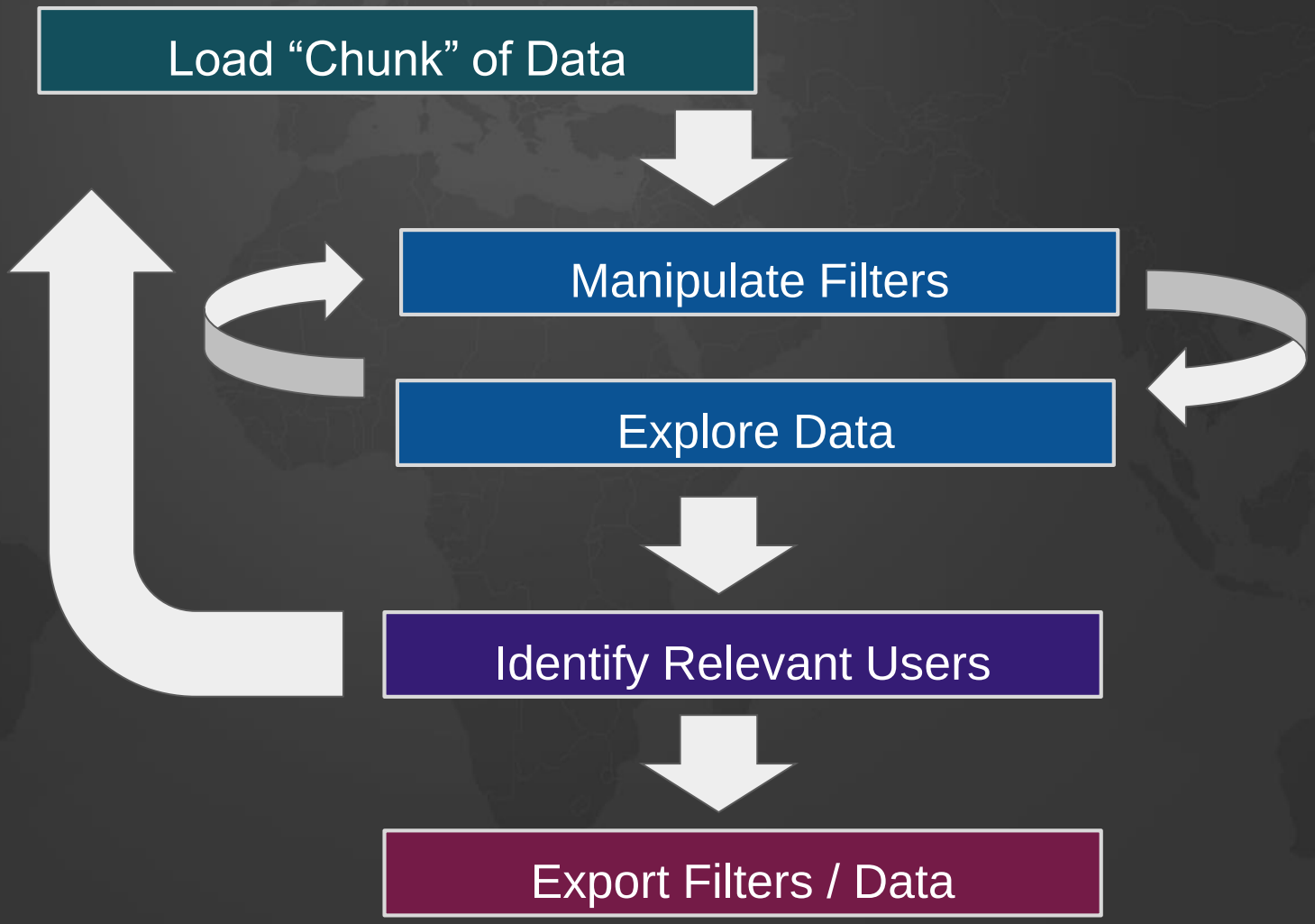
Data preprocessing

- 3 TB raw twitter data for 2015
- Pre-filtered with country bounding boxes
- Reduced to about 1.5 million tweets and 1,560 users
- Features stored in JSON format

Intended Workflow

“Overview first, zoom and filter, then details-on-demand.”

- Shneiderman, 1996



System Building

Data Collection

GeoVISTA Crawlers

- **“Raw” Tweets**
- **Sector 2- Europe**
- **All of 2015**
- **3TB of data**

Data Preprocessing

Python

- **1,560 users**
- **Users who tweeted at least once in Syria and Germany**
- **All tweets of these users**
- **~1.5m tweets**

Feature Generation

Python, PostGIS

- **Primary language**
- **Max. travel speed**
- **Countries visited**
- **Text features**

Interface/ Visualization

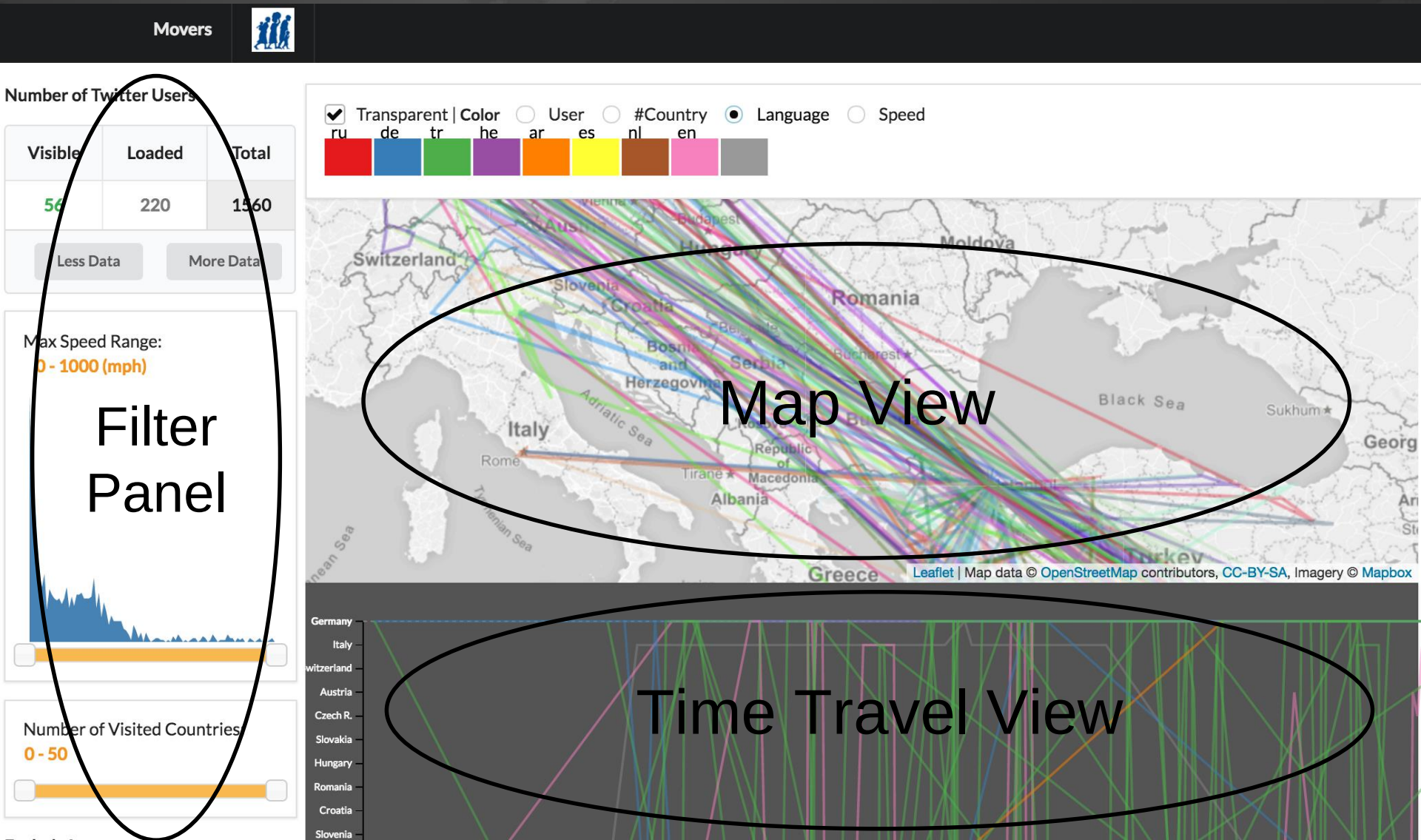
Javascript

D3

Leaflet

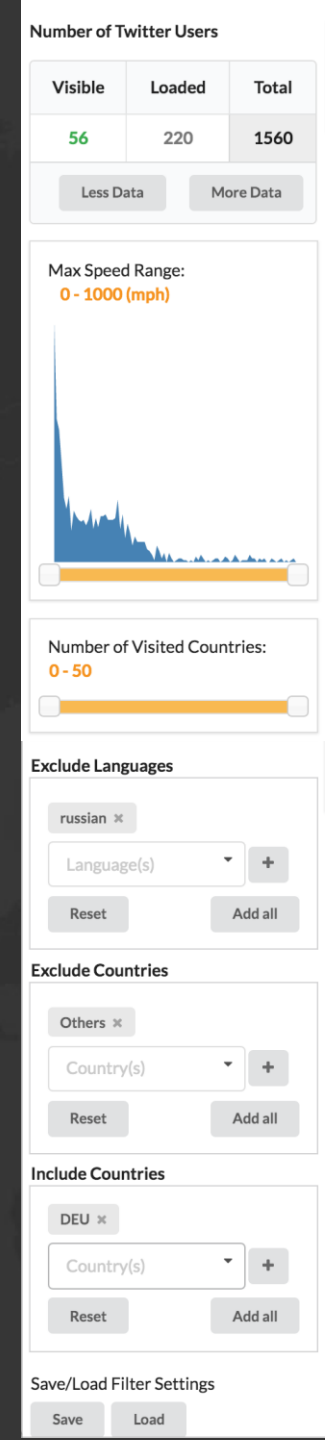
MongoDB

Interface Overview



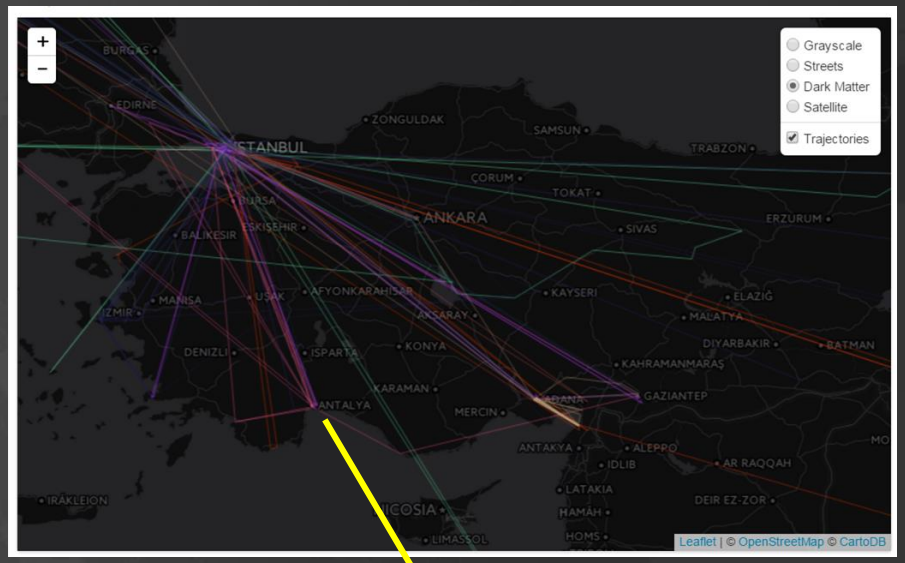
Filter Panel

- Starting with a small portion of movers
- Exclude uses by the following filters:
  - Maximum speed of user
  - Number of countries visited
  - Language used for tweets
  - Countries passed by
- User of the system can save filter settings, or load settings from past sessions



Map View

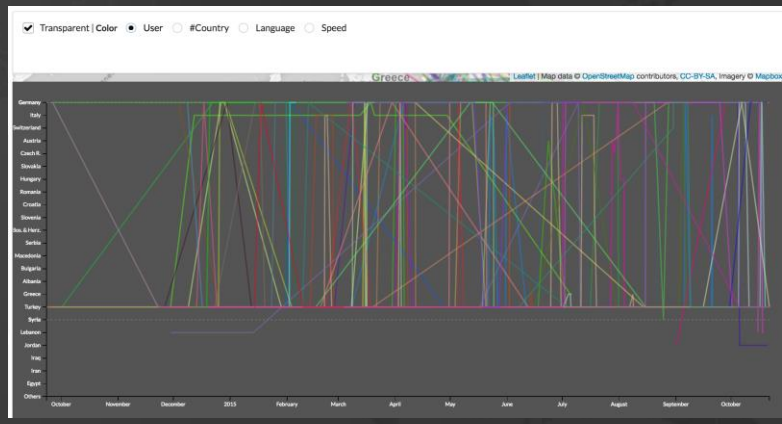
- Display trajectories interactively
- Slicing GeoJSON data into vector tiles on the fly
- Automatically simplify trajectories based on the zoom level

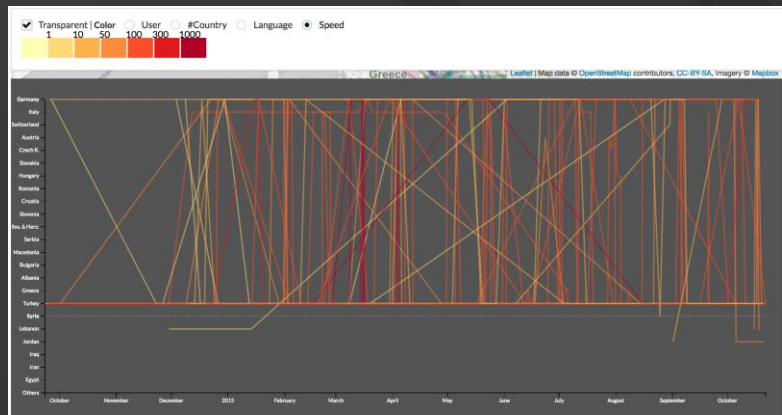


Tweet at an airport
(Satellite view)

Time Travel View

- Temporal focus
- Countries can be re-arranged
  - *Default: refugees “climb” to Germany*
- Flexible symbology implemented
- Implemented with D3
- Connected with Map View





Demo

Current and Future Work

- Expending the dataset
- Fusing heterogeneous data
- Incorporating social networks
- Active learning

Thank you!