

# Data Innovation in UNHCR

@rebecamoreno

Data Science Research

November 2017

# What is data innovation in UNHCR?

Application of data science techniques and artificial intelligence analyzing non-traditional data sets - including big data - for advocacy, preparedness and operational response.

# What are we doing? = Experiments

- ML: Social Media
- Text Clustering
- Google Trends
- **Project JETSON**
- Innovation fellows:
  - Use of Call Detail Records for mobility
  - UASC separation model

# Project JETSON

Prediction of arrivals

Case-study: Arrivals from Somalia into Ethiopia (Dollo, south)

# Why JETSON?

Help Somalia Operation **make evidence-based decisions** and **adequately plan/prepare** contingencies

Set a **precedent** for this type of predictive analytics work in **humanitarian sector** for coordination and compilation of data

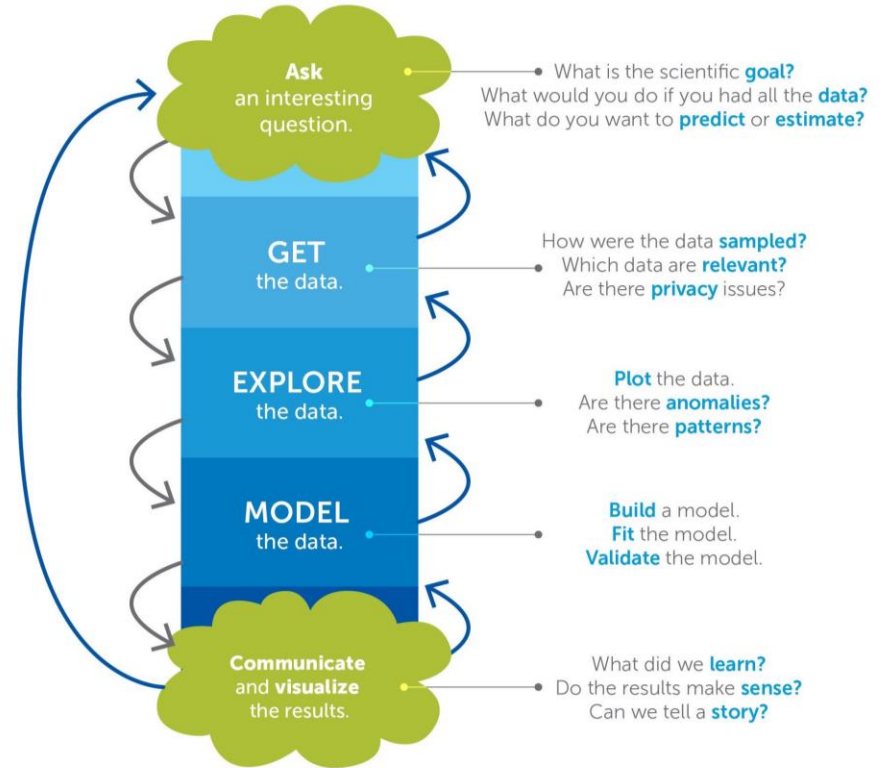
**Open data and model** for other data scientists, computer scientists, programmers can edit, add, and improve upon

# Methodology

## Applied Predictive Analytics:

- Process of discovering interesting and meaningful patterns in data
- Identify the 'best way' (the best fit) to predict a target variable
- Using Artificial Intelligence → Machine Learning (ML) support to find the best fit

## The Data Science Process



Derived from the work of Joe Blitzstein and Hanspeter Pfister, originally created for the Harvard data science course <http://ds100.org/>

# Partnerships Relevance

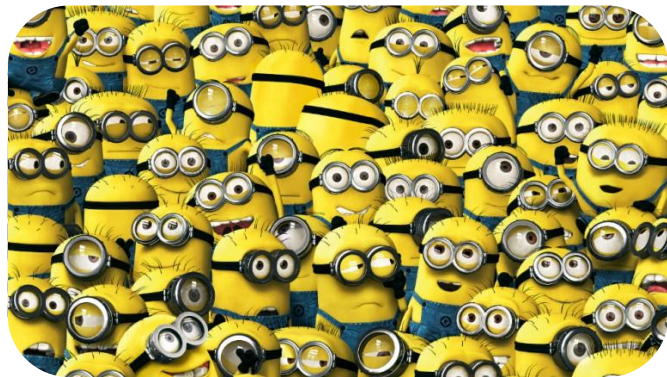
Non-traditional data sets require strong partnerships, for example:

*Machine-Learning (Social Media) + Clustering Text*

- UN Global Pulse

*Project JETSON:*

- Climate and weather data (WMO), big data/satellite (UNOSAT, GEO), market prices (FSNAU), violent conflict (ACLED), among others...



# The process (in a nutshell)



A UNHCR operation gives our team challenge/question

1



Harass people to give you data

2



Clean years of data

3



Use AI-based software to get formulas

4



5

Make many calculations  
(implement formulas in R)

Website + Draft  
Report  
predictions  
for 1 month



7

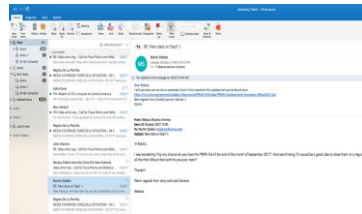
Run correlations  
and do graphs

6

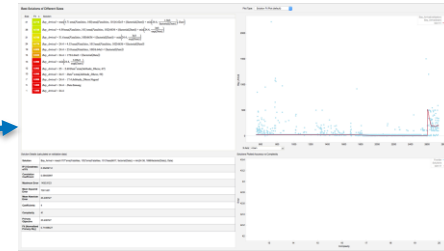




1



2

[illegible]

4

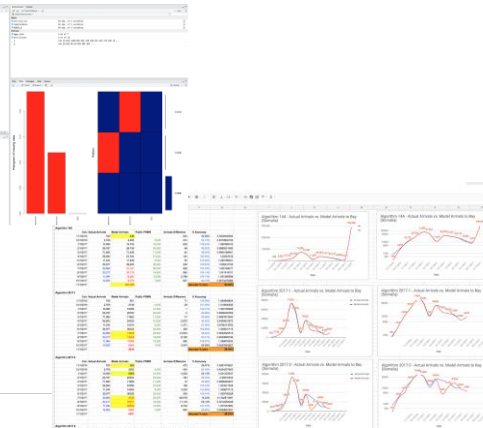


5

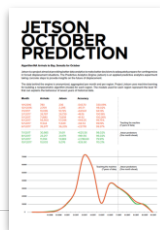
```

A <- 1.51214905518997<"current.long(t-13)","Sooi_CurrentRegion"
B <- 0.736056514680045<"before.long(t-12):(t-13)","Bay_ForeRegion"
C <- 0.70189390495416<"before.long(t-13)","Sooi_ForeRegion"
D <- future.long(t-7),"Samag_FutureRegion"|"Conflicts.long(t-15)","Gaijindud_Conflict"
E <- median.before.long(t-13):(t-15),"Bay_ForeRegion"
XA <- rivers.long(t-6),"Shabelle_River_SouwarStation","Shabelle_River"
XB <- 1.81928081648004
XC <- Pain.long(t-15),"Jubbada_Moose_Pain"|"rain.long(t-16),"Jubbada_Moose_rain"
XD <- median.before.long(t-6),"Geogony_Gallied_rain"
YE <- not(MOTA)
if(is(MOTA))XD <- 1
else if(!MOTA)&&(XD <= 0)
else(XD <- 1)
Z <- Fatalities.long(t-1),"Jubbada_Moose_fatalities"
K <- less(XA, XB)
if (is.na(XA) && is.na(XB))(Z <- 0)
else if(is.na(XA) && !is.na(XB))
else if(!is.na(XA) || XA < XB)(Z<=1)
else (Z <- 0)
FIN <- sum(A, B, C, D, E, X, XC, XD, Z, na.rm=TRUE)

```

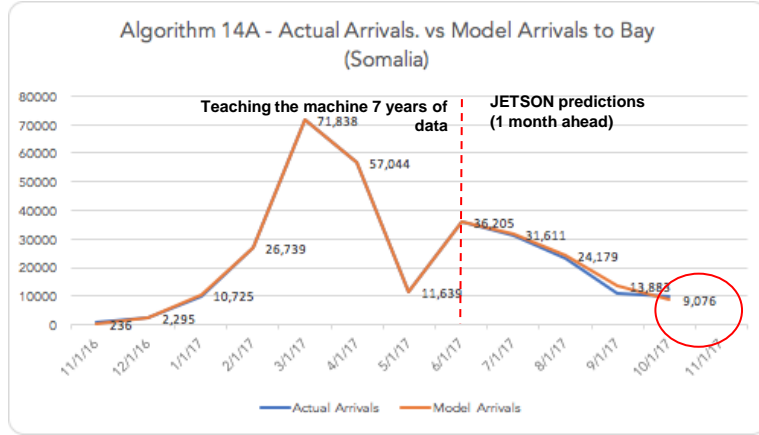


6



7

# Results: Bay Area (IDPs)

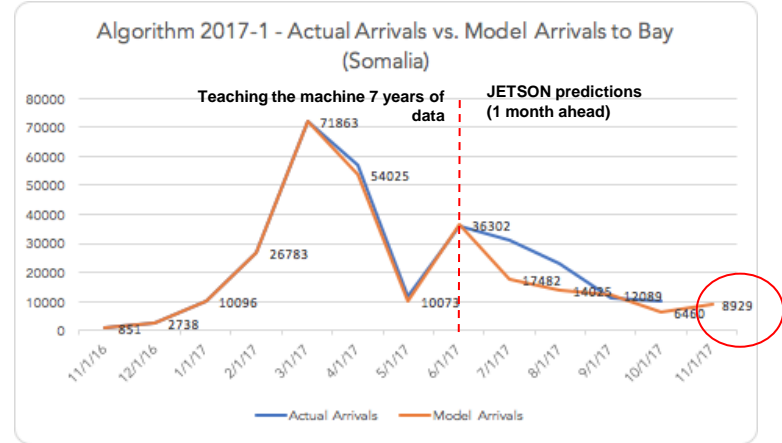


Predicting OCTOBER

Time prior: 1 week

JETSON: 9,076

Actual Arrivals: 10,003



Predicting NOVEMBER

Time prior: 2 weeks

JETSON: 8,929\*

Actual Arrivals: ???

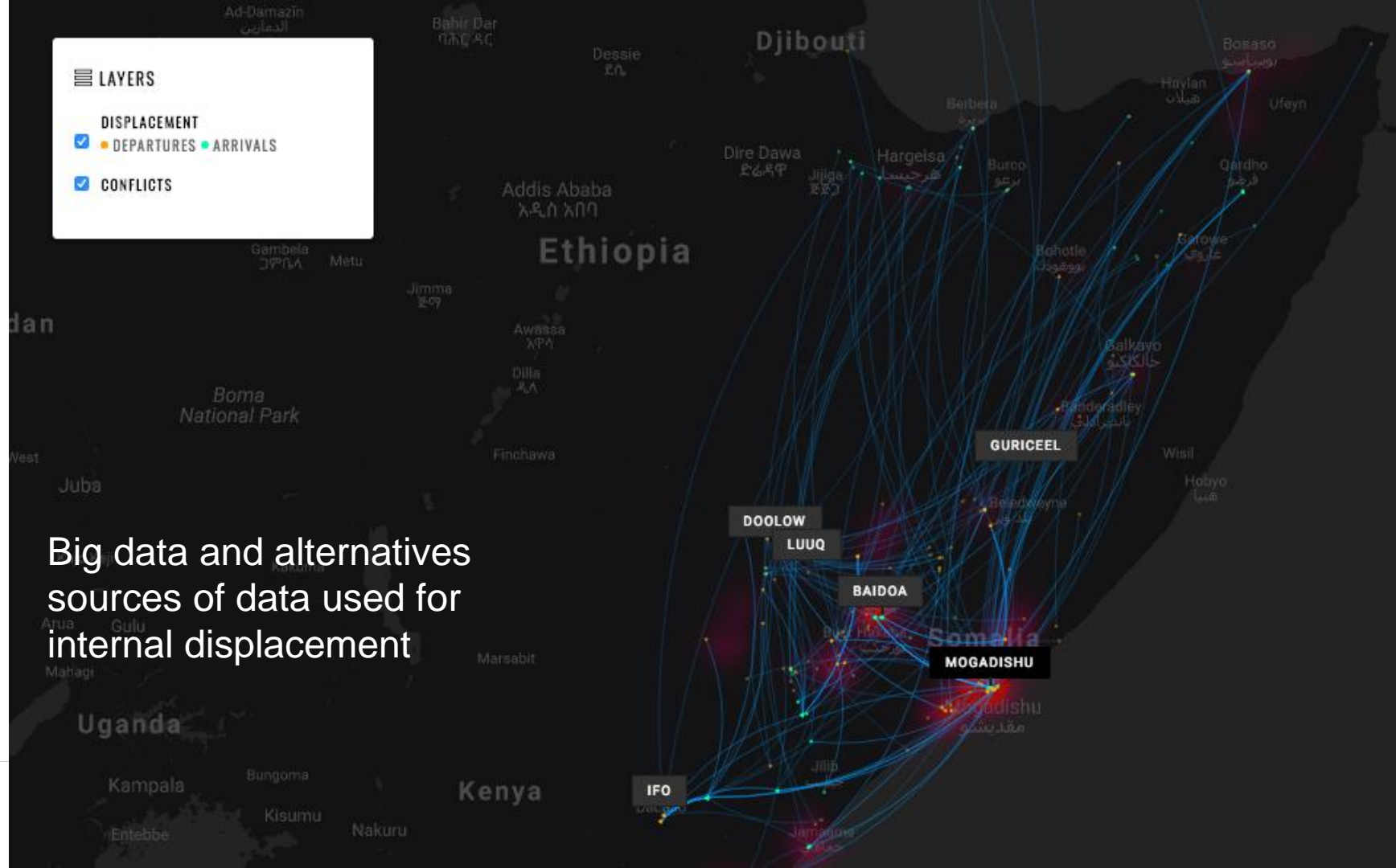
## LAYERS

### DISPLACEMENT

☒ DEPARTURES ☒ ARRIVALS

☒ CONFLICTS

Big data and alternatives  
sources of data used for  
internal displacement



# Other Experiments

# Machine-Learning: Social

Social media is one out of many types of big data sources giving insights on situations.

For example: Globally  
200B tweets/year = 6,000  
tweets/second

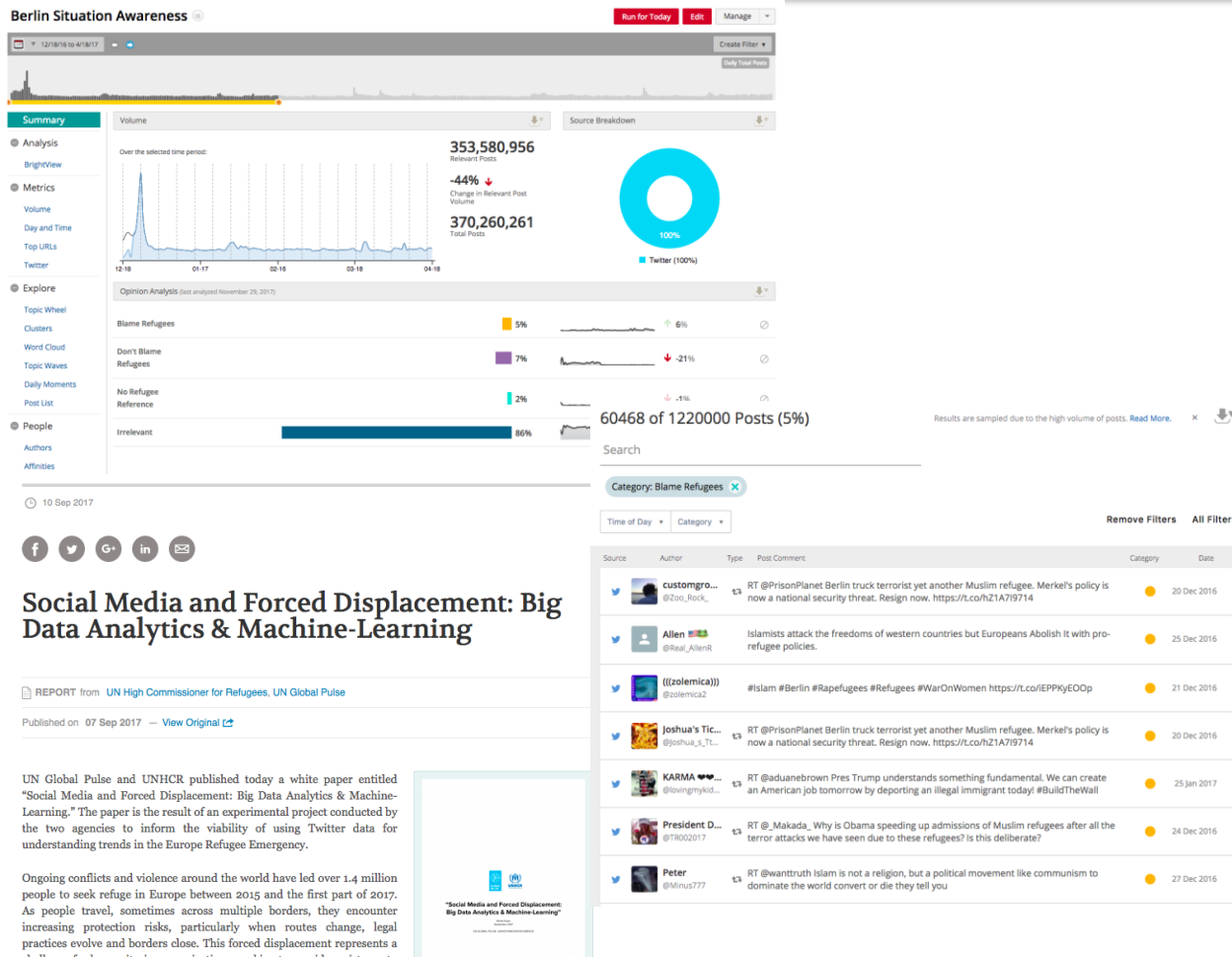
**What is this used for:**  
Tracking human rights violations,  
violent incidents, and xenophobia.

**How is it done:**

We train software (machine) to detect keywords. And then the machine learns to read, separate, and classify the data

**Where have we done this?**

- Europe (Xenophobia, [White Paper](#))
- Venezuela (Human Rights)
- Syria (Intention of return)
- Angola (Arrivals proxy)



# Text clustering

Text is unstructured qualitative data into a network to see patterns and topics linked to each other.

## What is this used for:

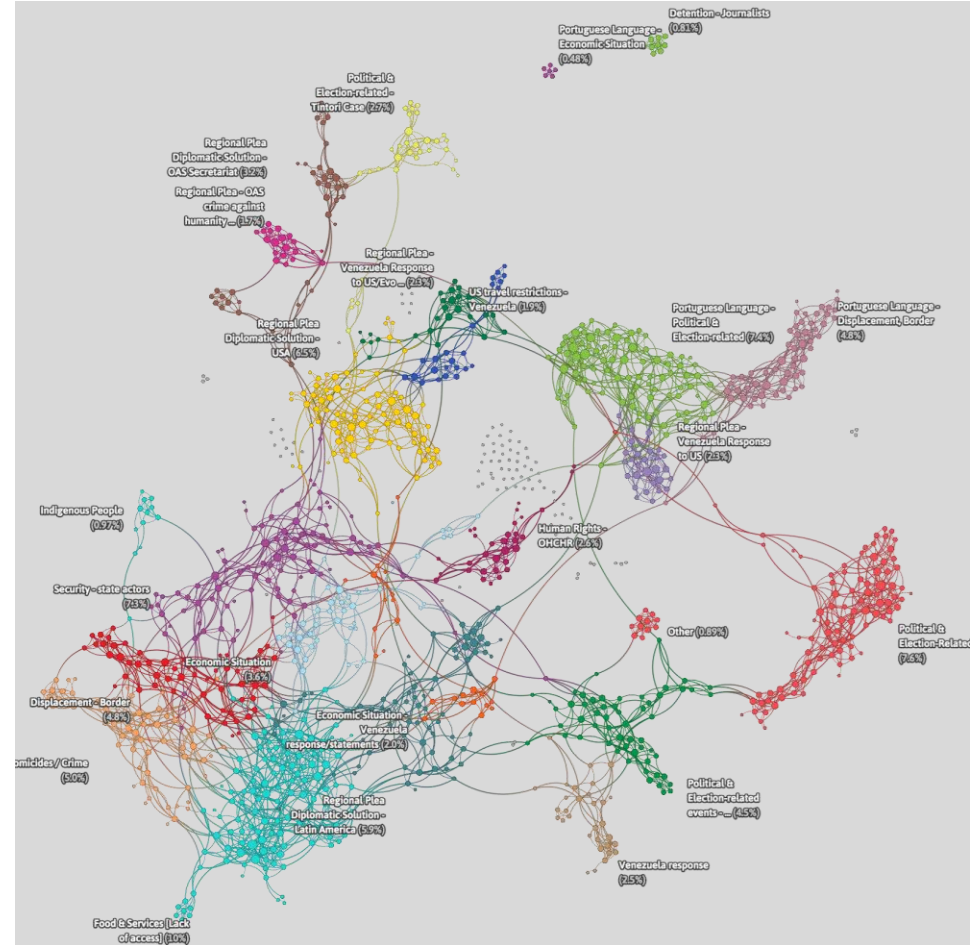
Provide insights of a particular situation where NO or partial access to full data/picture is available.

## How is it done:

Structure text in a spreadsheet and use software that automatically categorizes by taxonomy similar words. Different languages (English, Spanish, Portuguese, French)

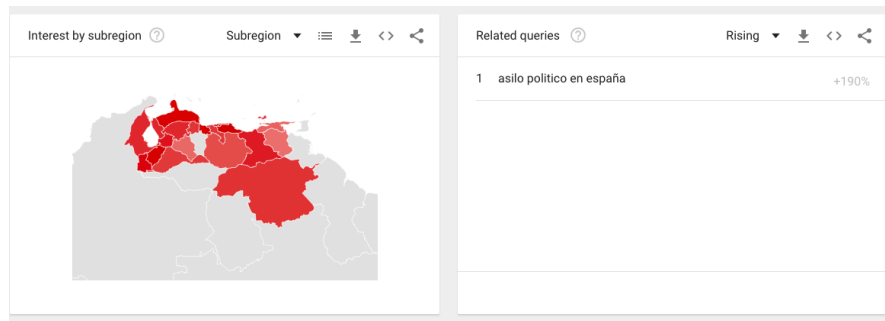
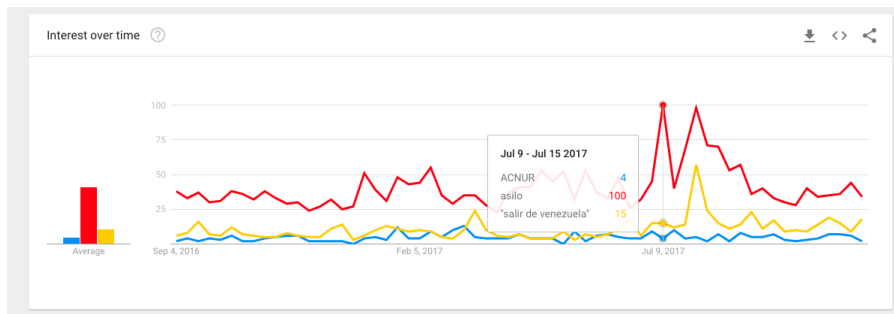
## Where have we done this?

- Venezuela (media/news on displacement/situation)
- Ethiopia (focus groups discussion notes)
- Lebanon (test, TBD)



Venezuela, news on displacement (Oct 2017)

# Google Trends



What is it used for:  
Geo-located searches/queries as a  
proxy of intention of movement.

Importance of partnerships\*

Where have we done this?

- **MENA:** intention of return (neighboring countries Syria)
- **Venezuela:** intention of fleeing ([Link](#))



*"Data is King"*

[innovation@unhcr.org](mailto:innovation@unhcr.org)

**REACH US:**

[YouTube.com/unhcrinnovation](https://www.youtube.com/unhcrinnovation)

[@unhcrinnovation](https://www.instagram.com/unhcrinnovation)

[facebook.com/unhcrinnovation](https://www.facebook.com/unhcrinnovation)



# Experimentation in data innovation 101

- Funding
- Capacity-building support
  - Particularly in the field
- Failure tolerance
  - with the models, functions, or training the machine until the “robot” (artificial intelligence) understands it. Humans are smart. Machines are not (yet)
  - We use data to test hypotheses that help us explain a complex world

# Innovation Fellows Experiments

Use of Call Detail Records (CDR, mobile phones) to analyze human mobility

- Daniel Macguire, Class 2016
- Asuka Imai, Class 2016

UASC Separation Model: degree of separation of unaccompanied and separated children in displacement situations

- Janis Riedel, Class 2017