# AFRICA'S VOICES

## Machine learning for targeted communication in emergency

AFRICA'S VOICES

www.africasvoices.org

Twitter: @africas_voices
Facebook: Africa's Voices/ Voix d'Afrique

**“…like listening at the keyhole of a giant conversation”**

- Rob Burnet, CEO Well Told Story

# Approach

- Engage citizens in discussions
- Use digital methods to hear about their opinion
- Social media, SMS
- Analyse the messages
- Deliver insights and recommendations to partners

# UNICEF Somalia

- Gather data on beliefs, knowledge and practices of the Somali population – health, education, gender differences

- Data can be triangulated with other data sources to inform UNICEF's programmes – eg. health promotion

- Gain evidence of Africa's Voices approach as a remote monitoring tool for UNICEF in fragile, insecure regions

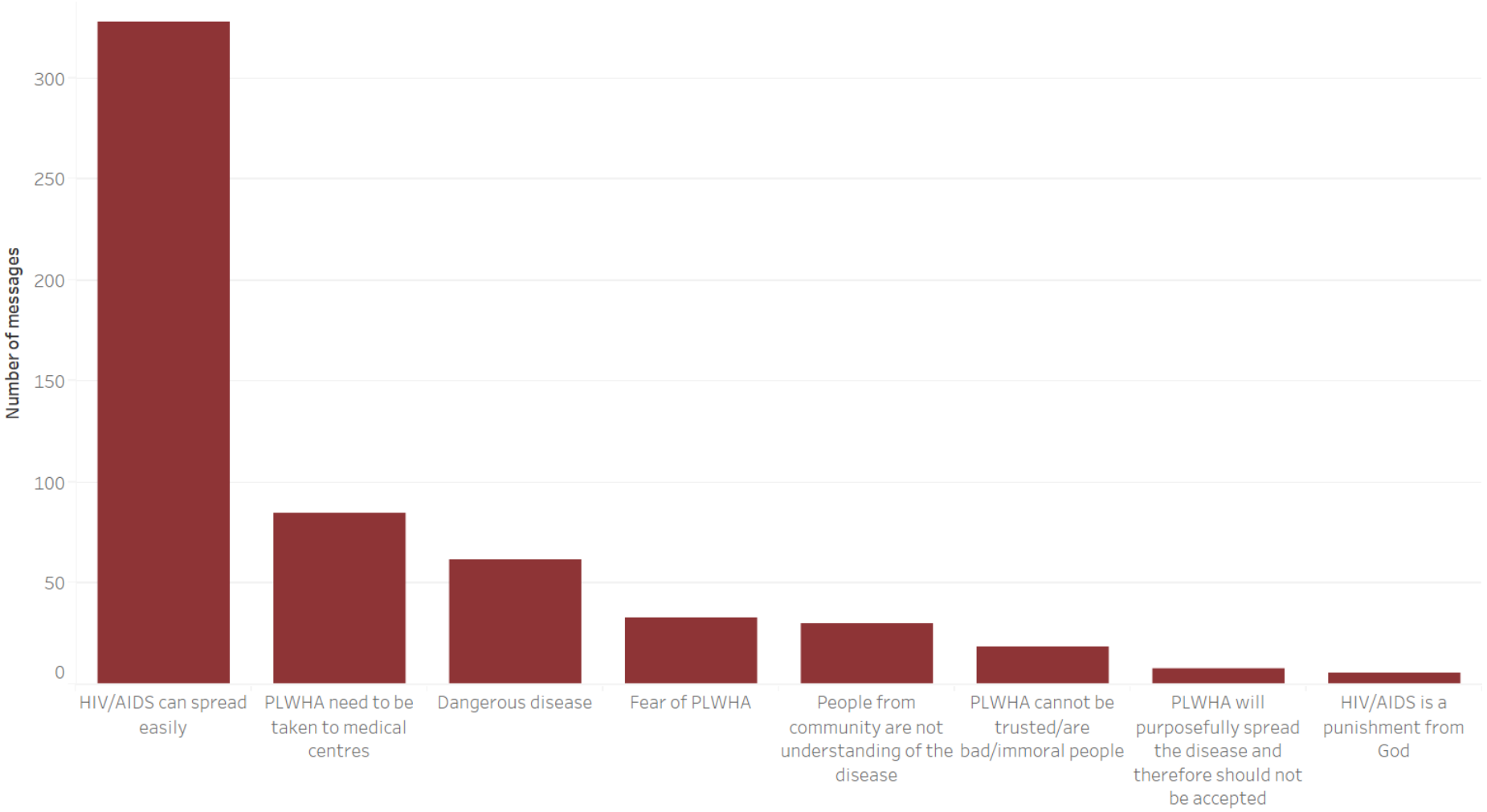
**Example topic**

- Research on acceptance of people living with HIV/AIDS across Somalia

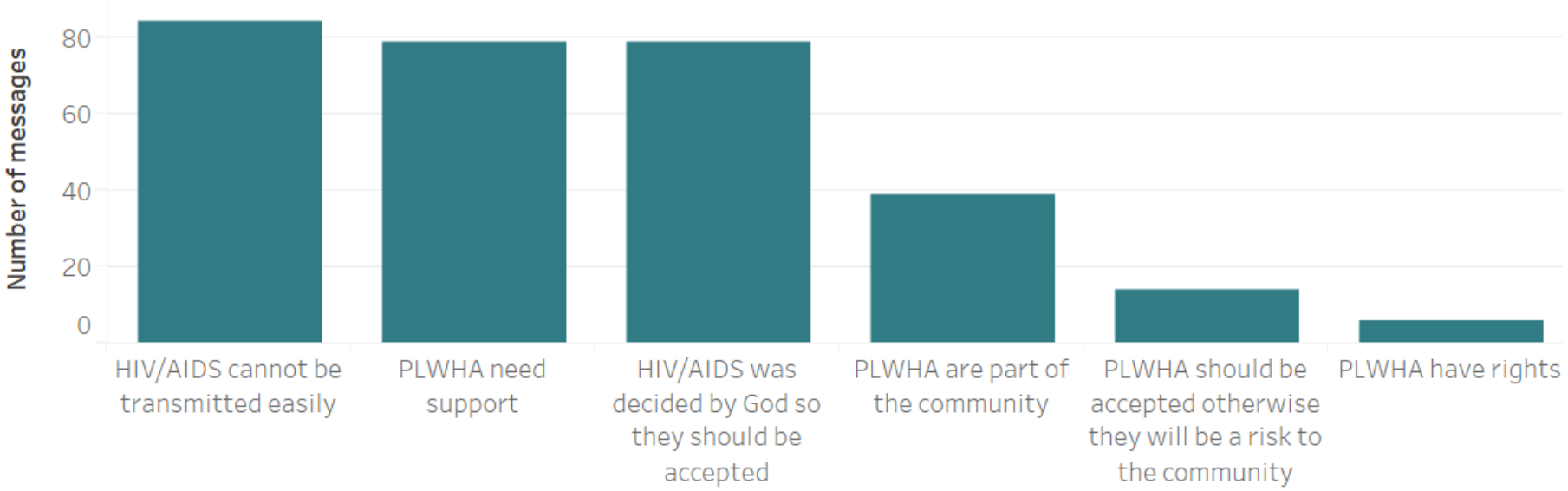
# UNICEF Somalia

- First point of engagement: **generating discussion** on radio through interactive shows on 26 radio stations in Somaliland, Puntland and South Central
- Invite listeners to **contribute their opinion via SMS** – gathered through an online SMS platform (RapidPro)
- Audience receives **SMS survey** asking for demographic information and information about the main topic – HIV/AIDS testing
- Participation: ~8600 messages over a week, ~3900 for analysis

# Most common reasons/beliefs for lack of acceptance of people living with HIV/AIDS



# Most common reasons/beliefs for acceptance of people living with HIV/AIDS

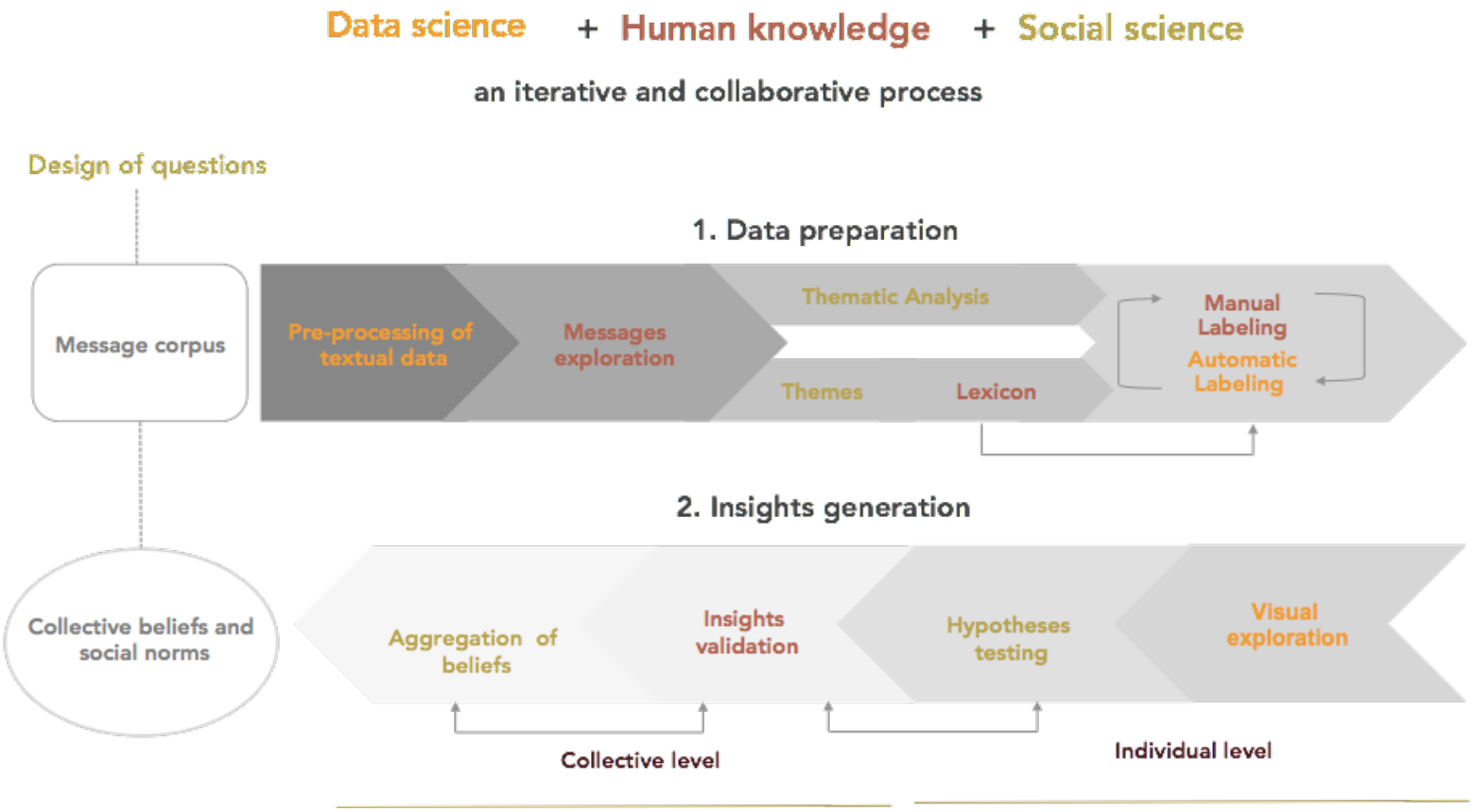


# Text analysis

## Challenge

From the SMS messages to structured data and graphs

- Making sense of unstructured text
- Analysing local language data
- Low-resource language: computational resources unavailable or for specific purpose
- Time


Data science + Human knowledge + Social science
an iterative and collaborative process
Design of questions
1. Data preparation
Message corpus
Pre-processing of textual data
Messages exploration
Thematic Analysis
Themes
Lexicon
Manual Labeling
Automatic Labeling
2. Insights generation
Collective beliefs and social norms
Aggregation of beliefs
Insights validation
Hypotheses testing
Visual exploration
Collective level
Individual level

# Text analysis

- Machine learning technique: classification
- Cycle: manual labeling – training data – classification – evaluation

**Challenges of manual labeling**

- Manual work and human expertise – going through the messages to find emerging themes
- Consolidate those into well-defined labels
- Ensure solid shared understanding of labels among research assistants
- Identifying labels with low message counts, revisions

# Text analysis

## Challenges of machine learning

- SMS data: short, messy
- Complex ideas/labels
- High amount of irrelevant messages – noise or not related to the question

# Text analysis

- Binary classification using logistic regression for each code
- Features: bag of words, word sequences, character sequences
- Some knowledge of the language: spelling, keywords
- No grammatical parsing, tagging
- **https://github.com/guyemerson/somali**
- Results: mixed
- Depends on complexity of ideas, language around it

# Real-time classification

**Next challenge**: can we use labeling in the interaction itself? How?

**Context**

Ongoing drought in Somalia, contributing to the spread of diseases such as cholera

Cooperating with UNICEF Somalia to understand the water, sanitation and hygiene situation, accessibility of clean water, water treatment practice practices

Especially barriers – eg. barrier to access to clean water

# Real-time classification

- Have you treated your water recently?
- Not at all.
- Why not?
- I don’t have any aquatabs left

Label as: no access to chemicals

Send targeted information message based on the response

Info: boiling water for 10 minutes reduces risk of cholera

Or: chlorinated water unhealthy
Info: it’s actually safe to drink

# Real-time classification

What we need to do

- Collect and label data
- Train a classifier
- Deploy it online
- Receives a text message, runs the classifier, returns the label(s)

# Real-time classification

- Piloting and testing

**Open questions and challenges**

- Ongoing validation
- Threshold for certainty of prediction
- Effect –  stir engagement?
- Effect of errors?
- How we make use of the targeting to make the communication more efficient?


AFRICA'S VOICES
Questions?
www.africasvoices.org
Twitter: @africas_voices
Facebook: Africa's Voices Foundation