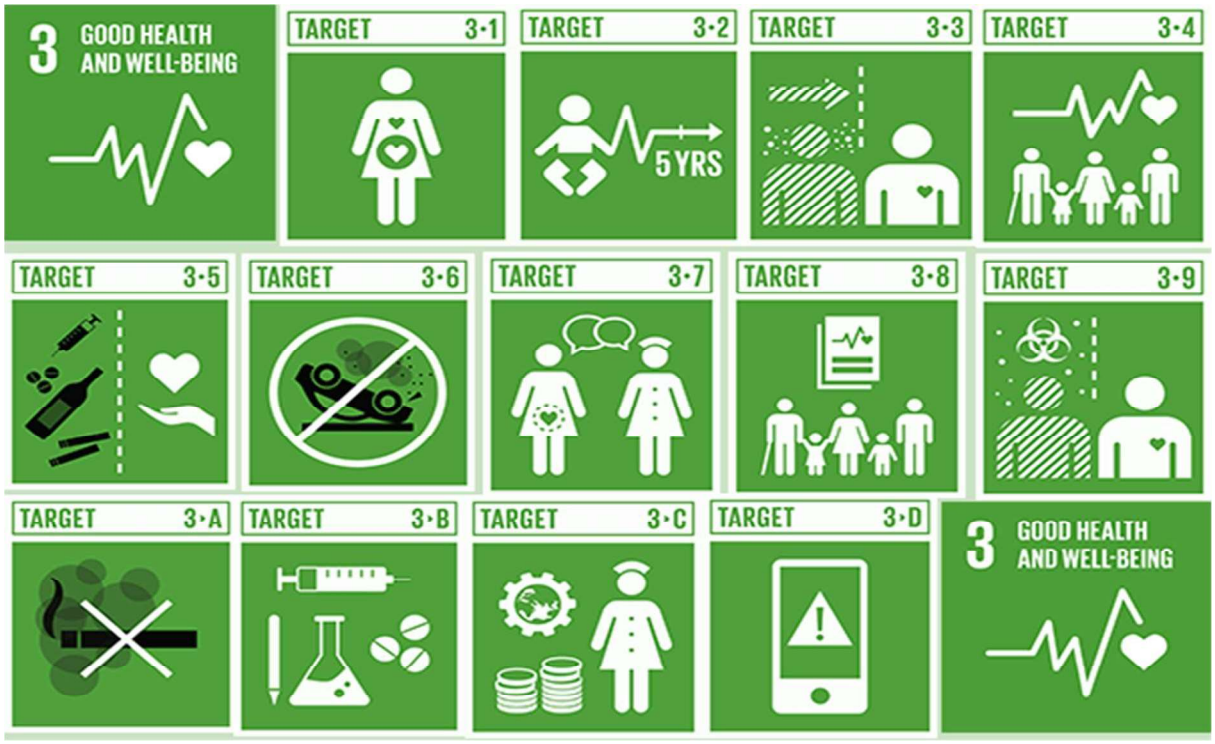
SUSTAINABLE DEVELOPMENT GOALS



Target 3D : Strengthen the capacity of all countries, in particular developing countries, for **early warning, risk reduction** and management of national and global health risks

BBC Sign in | News | Sport | Weather | Shop | Reel | Travel | Mo

NEWS

Home | Video | World | US & Canada | UK | Business | Tech | Science | Stories | Entert

US & Canada

Hundreds of US children treated for respiratory virus

8 September 2014

ALJAZEERA News ▾ Middle East Documentaries ▾ Shows ▾ Investigatior

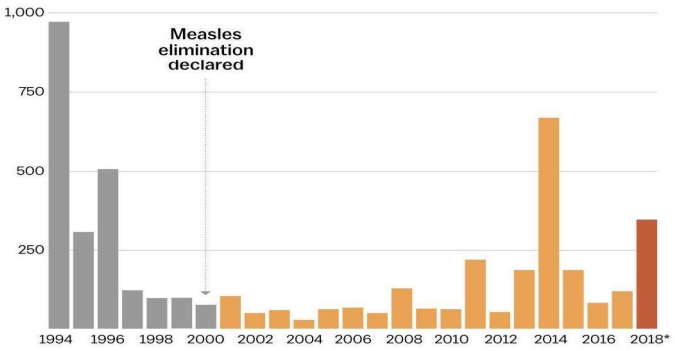
UNITED STATES 13 August 2016

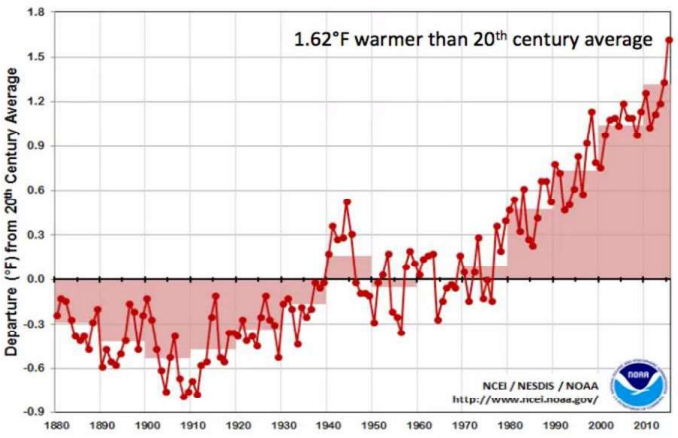
US declares emergency in Puerto Rico over Zika

Move following report of 1,914 new cases in the past week will allow territory access to federal funds and personnel.

Measles outbreaks are preventable — but they keep happening

Cases of measles reported per year in the US





- More travel & trade
- Urbanization
- Pervasive poverty
- A warming climate

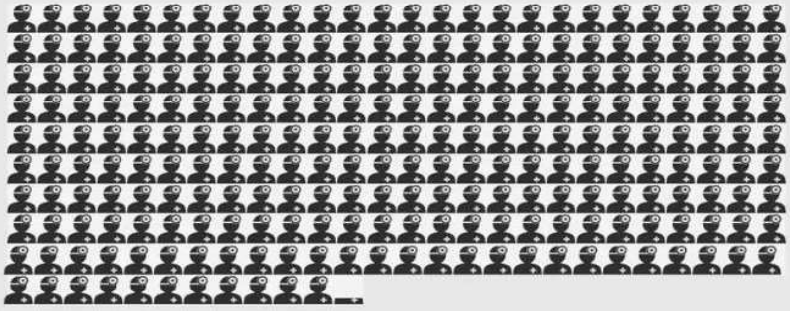
Number of doctors per 100,000 population

LIBERIA: **1.4**

SIERRA LEONE: **2.2**

GUINEA: **10**

UNITED STATES: **245.2**



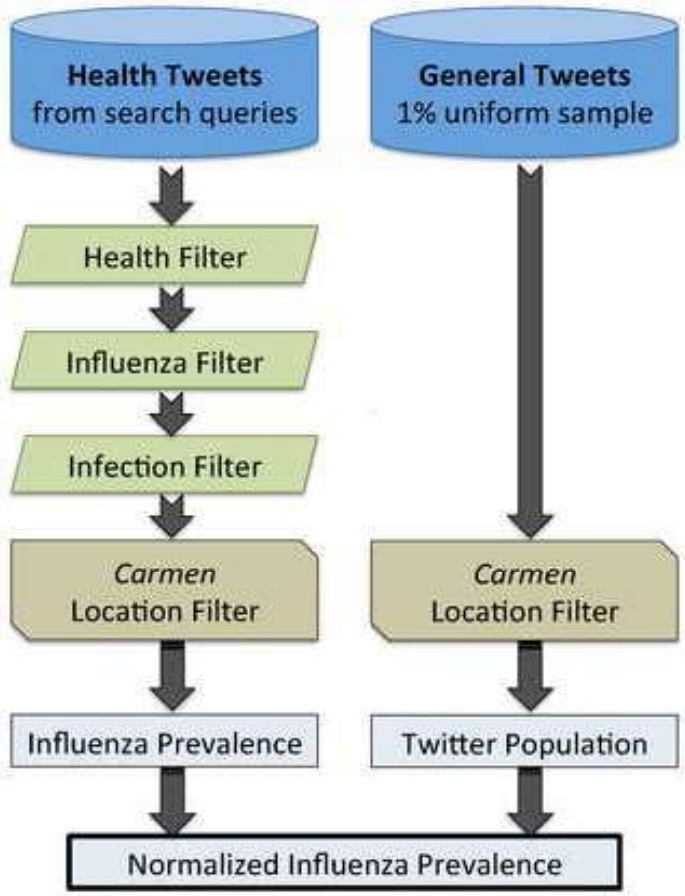
Vox

SOURCE: WHO

BACKGROUND

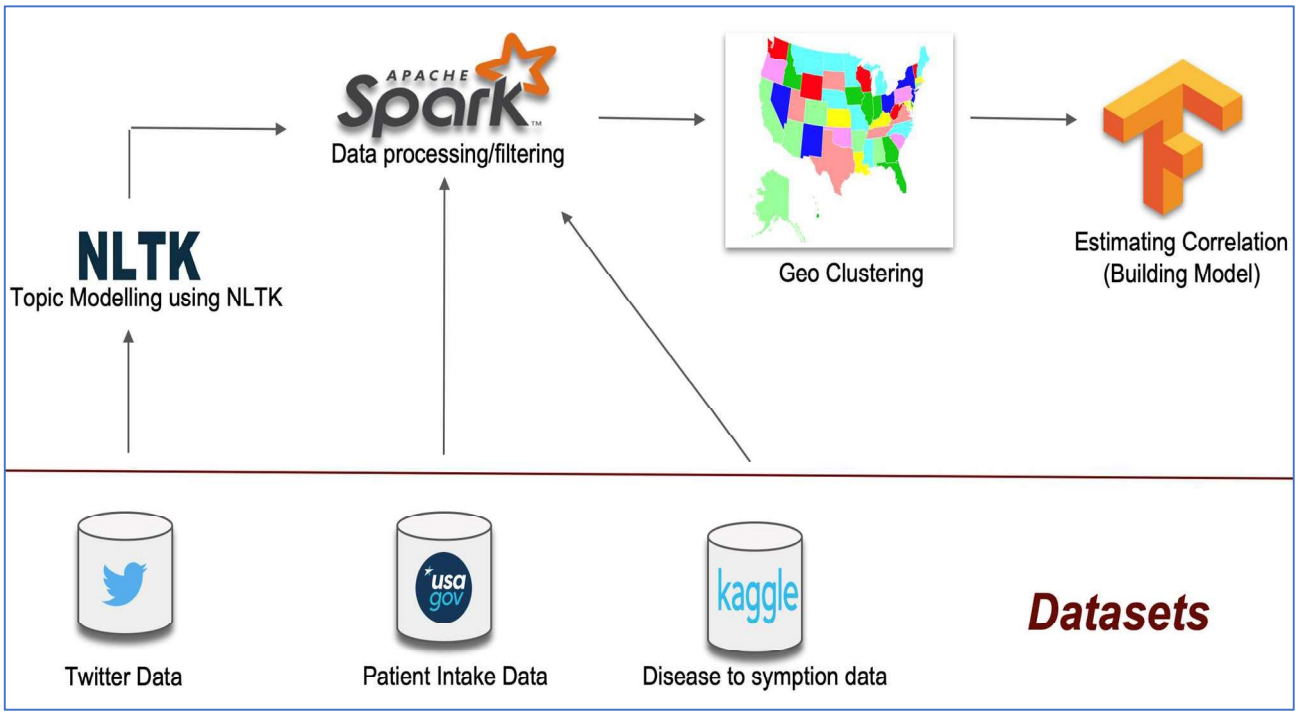
What do people tweet when they are sick? A preliminary comparison of symptom reports and twitter timelines.
Ashlynn R.Daughton, Michael J Paul, Romi Chunara

Towards real time measurement of public epidemic awareness: Monitoring influenza awareness through twitter
Michael C Smith, David Broniatowski



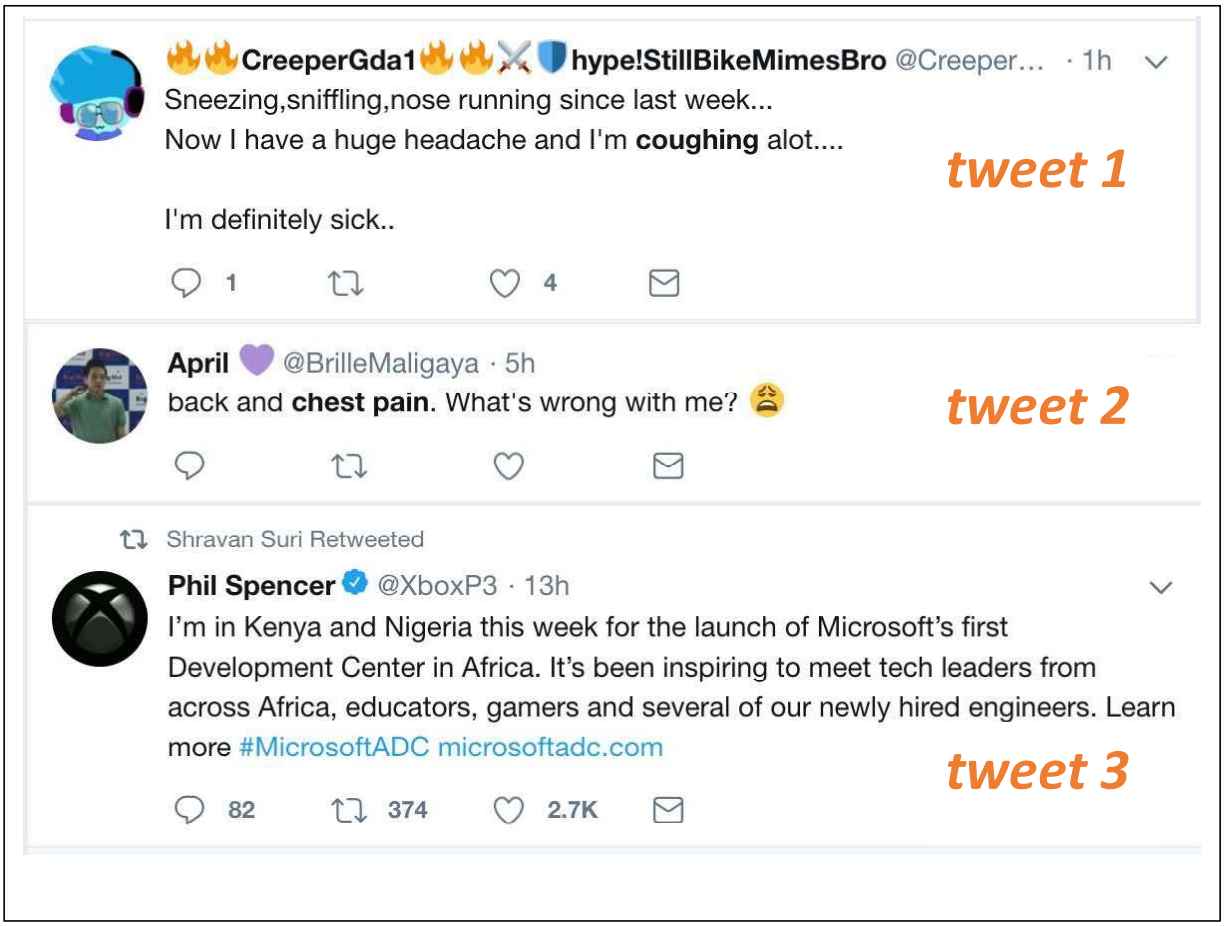
- The increased use of social networking platforms entails more widely shared personal and immediate information.
- Twitter provides easy access to data. Since each tweet is 280 characters only, text processing is simplified.

METHOD



- Twitter data for the year 2018 for USA obtained by tweepy.
- Patient intake data for hospitals across Texas. Includes columns for probable diagnosis.
- Health data corpus consists of text from medical articles. We use this to perform 'Topic modeling' to filter out health related tweets.
- Disease to symptom data is used to link symptoms to disease

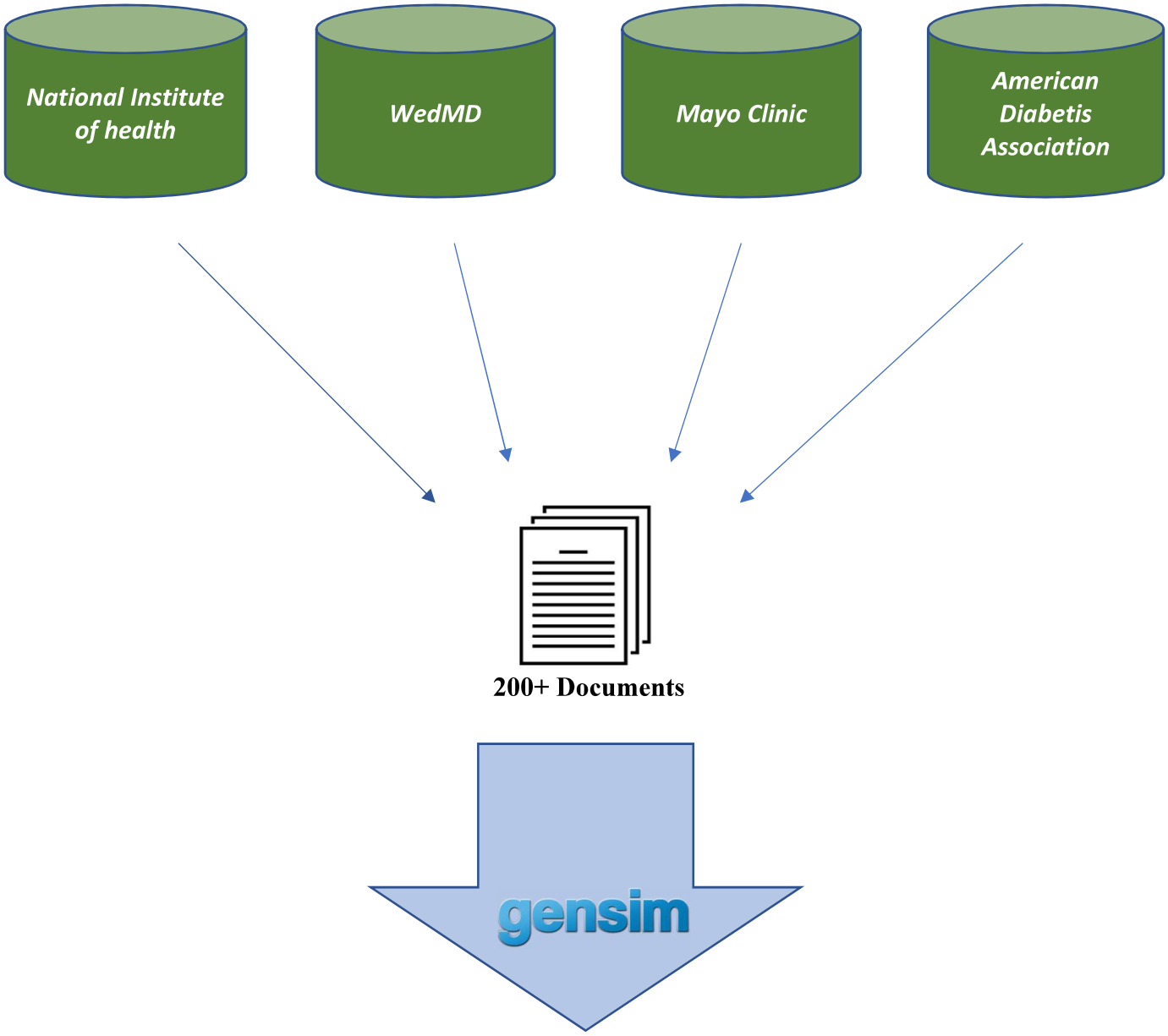
METHOD



tweet 1, tweet 2

- Topic modelling is a type of statistical modelling for discovering the abstract topics that are occurs in text
- LDA is a technique for topic modelling that builds a topic per document model and words per topic model, modeled as dirichlet distributions

METHOD



Health Corpus

RESULT

- Construct symptom count matrix from health filtered tweets by aggregating tweets into week buckets. Then, count the occurrence of symptoms, associated, with current disease.

influenza_count_matrix

state	week	dehydration	cough	chest pain	runny nose	cold sweat
TX	1	14	23	37	37	16
TX	2	16	15	50	37	3
TX	3	0	16	10	25	11
TX	4	21	30	58	11	7
TX	5	5	19	22	3	8
TX	6	17	16	49	21	13

- Use Linear Regression, with count matrix as input to predict weekly hospital intake data.
- For example, for influenza, we can see that “Runny Nose” has the highest correlation and hence, best signifies, the probable sick cases.

	Coefficient	P-value
Running Nose	291.65	$1.17e^{-6}$
Cough	116.82	0.018
Cold sweat	93.67	0.045
Dehydration	-21.07	0.65
Chest Pain	-37.2	0.75

RESULT

Probable Influenza Outbreak predicted around week 42, in advance,