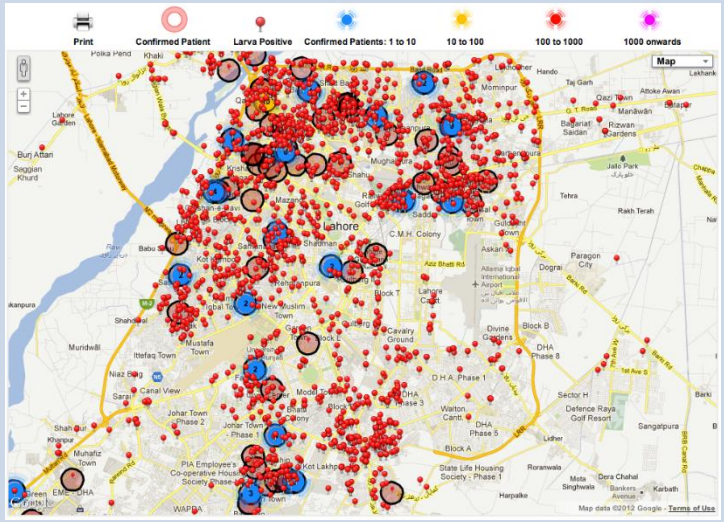
Data for Development

Lecture 24: CSE 490c



Announcements

- Homework 7, Due December 3
- Programming Assignment 4, Due December 10
- Readings for Wednesday / Friday posted

Topics

- Data Science for Development
- AI for Social Good

- Today
  - Strava Fitness App
  - Predicting Dengue fever based on helpline calls

Strava Fitness App

Fitness app Strava lights up staff at military bases

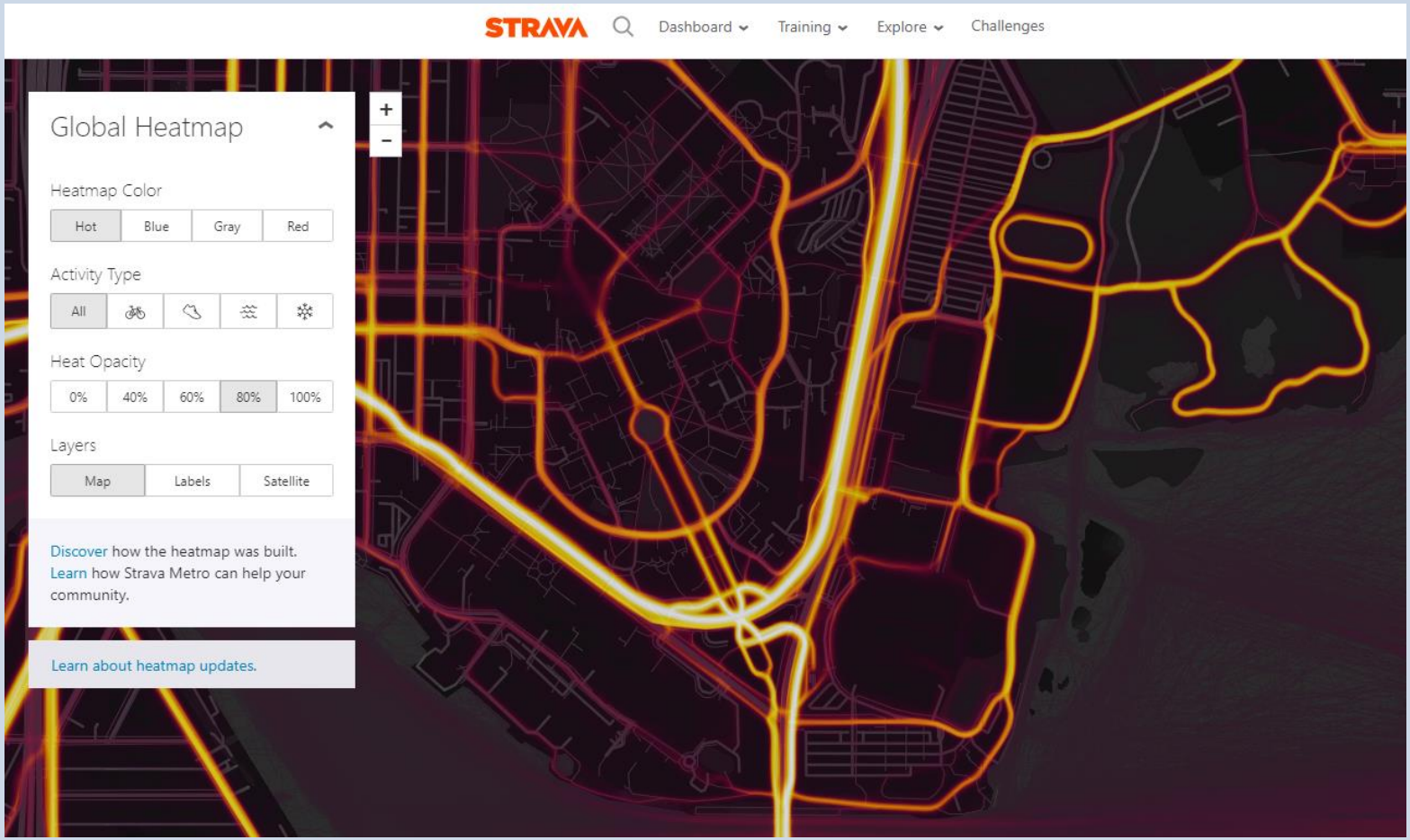
29 January 2018 Share

The movements of soldiers within Bagram air base - the largest US military facility in Afghanistan

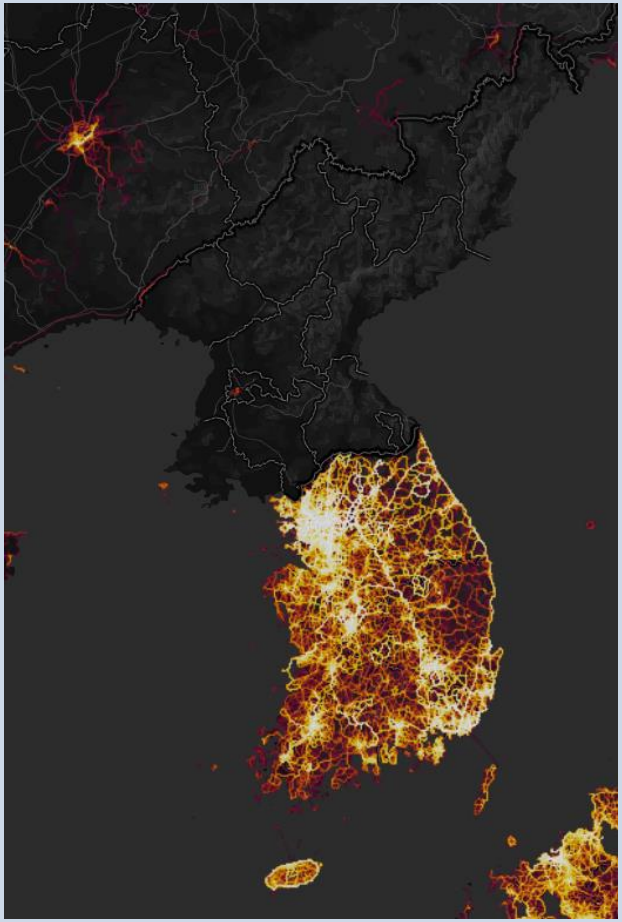
Security concerns have been raised after a fitness tracking firm showed the exercise routes of military personnel in bases around the world.

Strava Heatmam

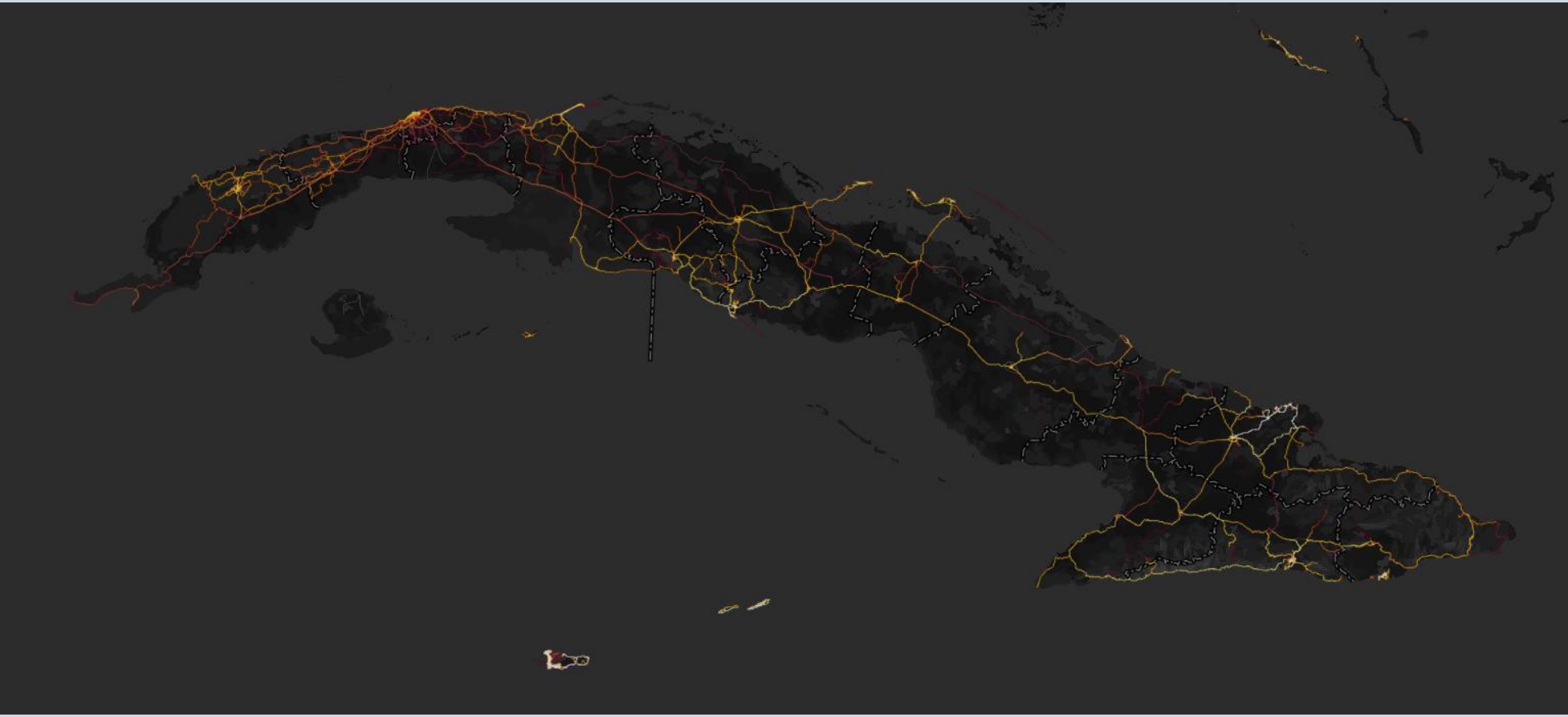
https://www.strava.com/heatmap



Korea

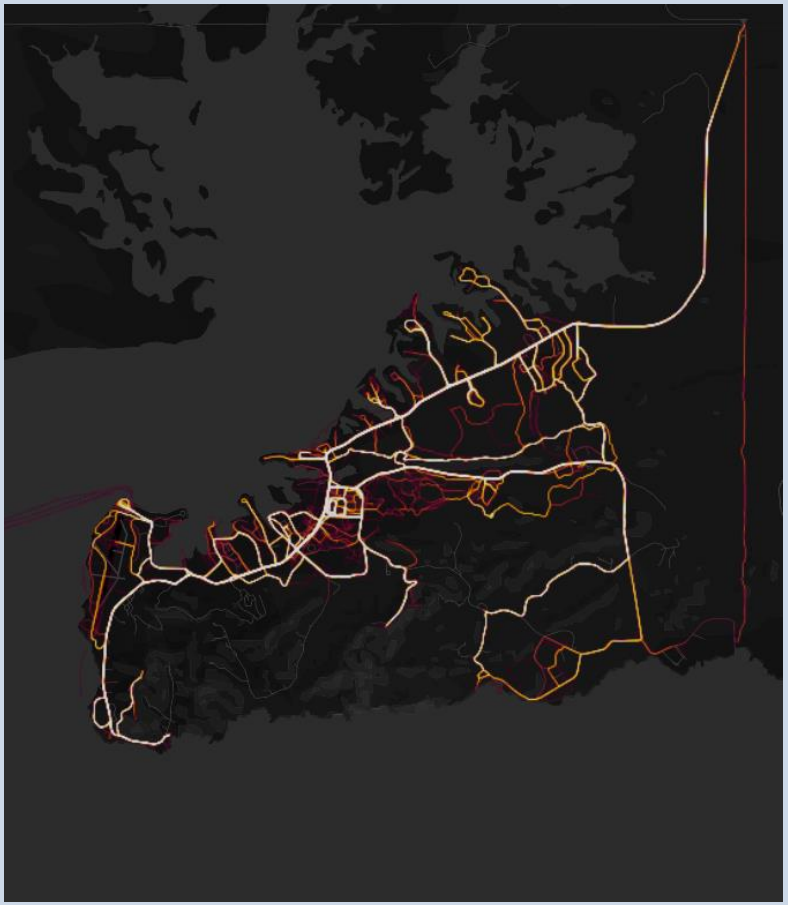


Cuba

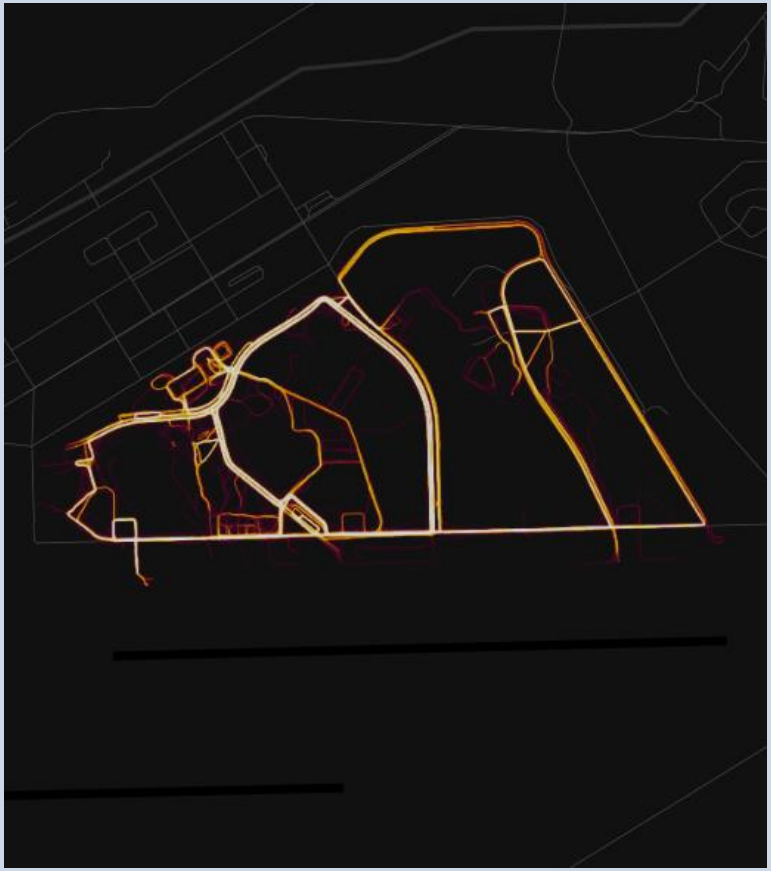


Eastern Cuba

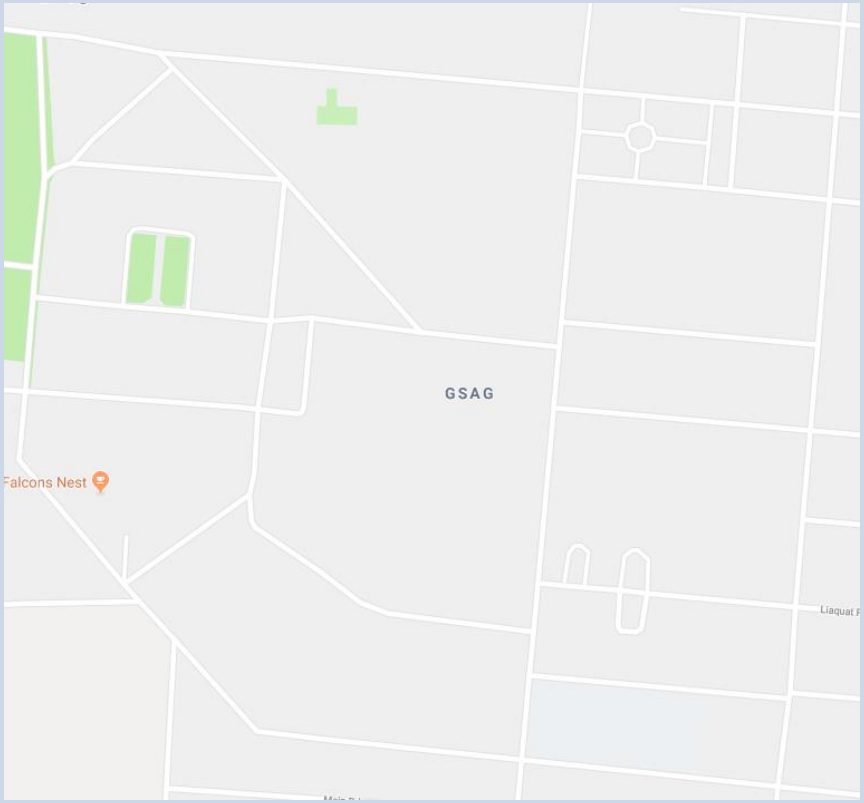
Guantanamo Bay



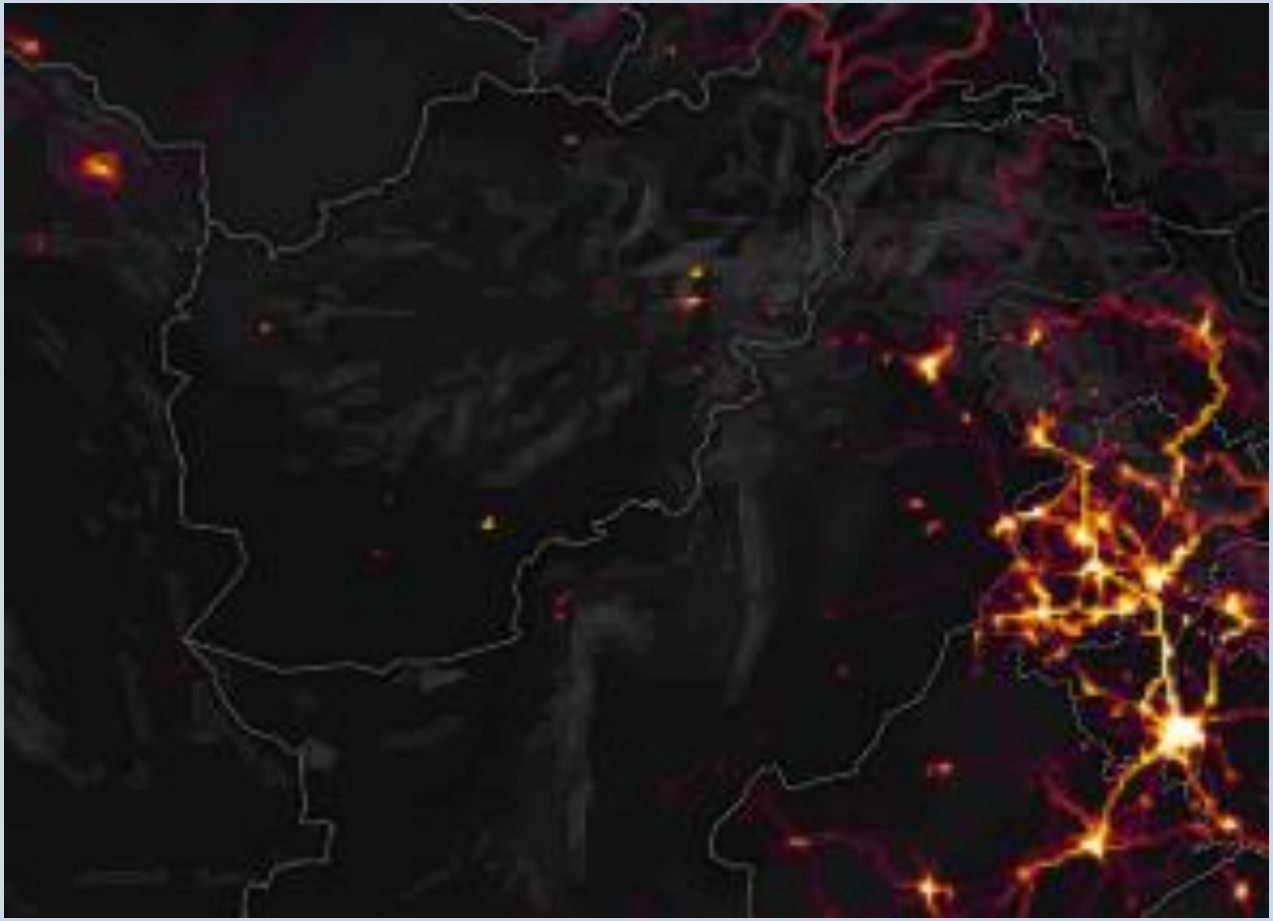
Eastern Iraq



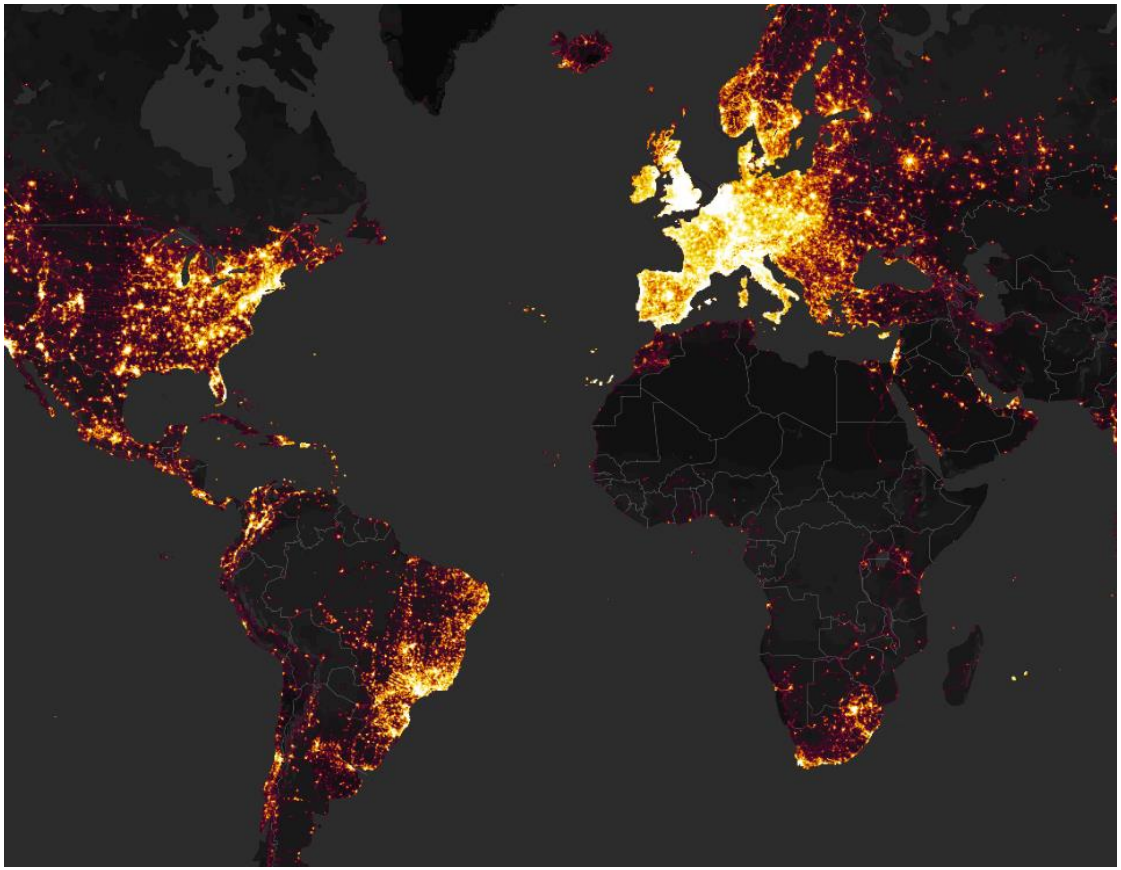
Risalpur, Pakistan



Afghanistan



Kandahar Airport



Strava App

- Business model
  - Social networking for exercise
  - Monthly subscription for various features
- Interesting data source
- Enormous amount of data
- Highly biased towards The West
- Proprietary

Data for Development

- Interesting facts
  - Identification of military bases
  - Earthquake identification for call data records
- Economic Studies
  - Populations and demographic data from call data records
  - Migration
- Policy, government and commercial interventions

Data Sources

- Where are there massive data sets?

- Satellite image data
- Mobile Operators
  - Call Data Records
- Technology monopolies
- Proprietary Applications

Dengue Fever

- Viral disease spread by Aedes mosquitos
- Primarily urban
- Increasing prevalence
- No cure
- Many people recover on there own, some cases require hospitalization
- Vector control to limit infections
- Importance of early detection, localization, monitoring

Challenges in Dengue monitoring

- Only a small percentage of cases reach hospitals
- By the time data is reported from hospitals, outbreak is well established
- Alternate approaches
  - Internet search queries
  - Online media reports
  - Environmental parameters
  - Telephone hotlines

Forecasting dengue fever outbreaks based on telephone hotline

RESEARCH ARTICLE

PUBLIC HEALTH

Fine-grained dengue forecasting using telephone triage services

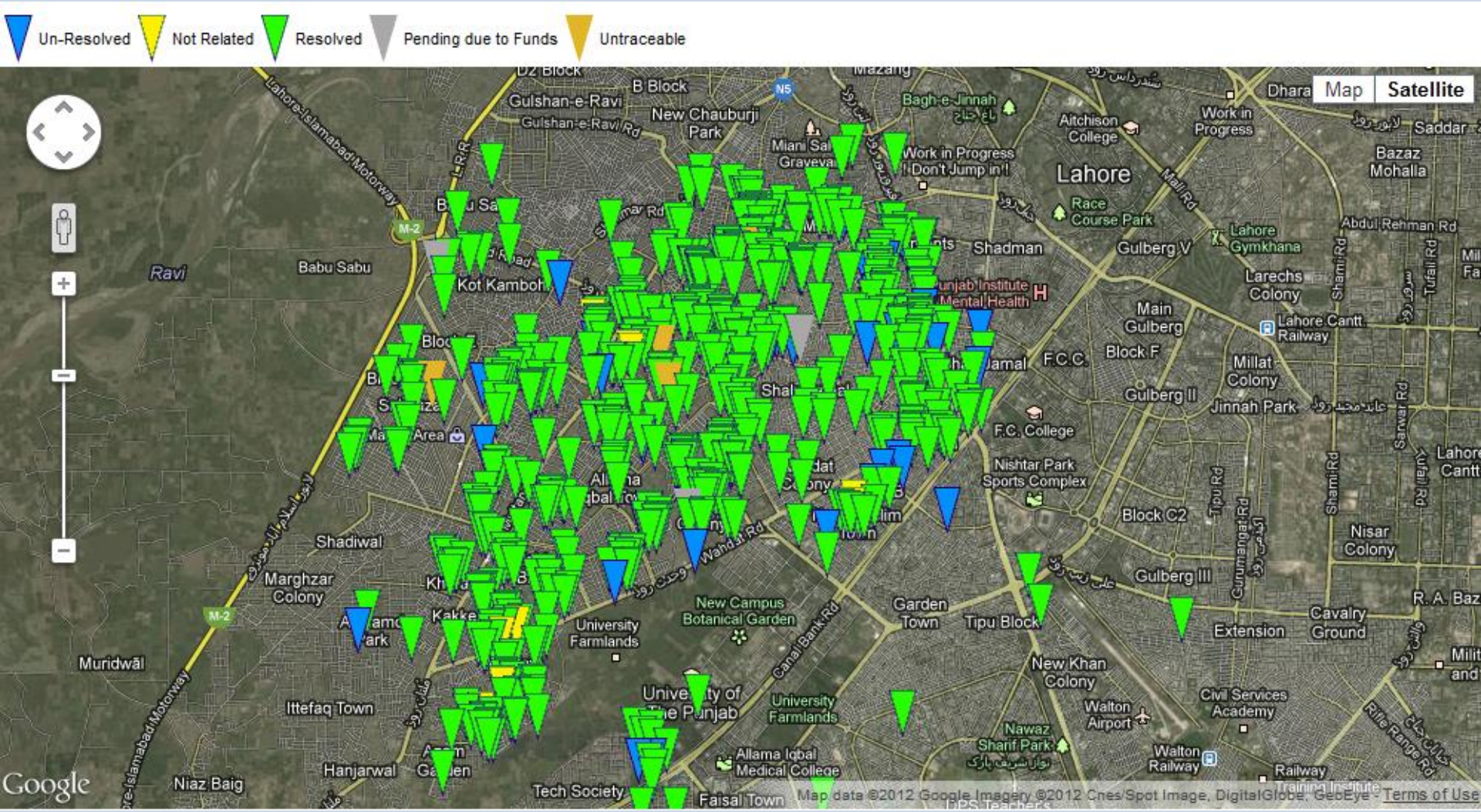
Nabeel Abdur Rehman,[1,2] Shankar Kalyanaraman,[3,4] Talal Ahmad,[3,4] Fahad Pervaiz,[5] Umar Saif,[1,6] Lakshminarayanan Subramanian[3,4]*

- Contribution: Show that hotline calls can be used for localized prediction of dengue cases when combined with an environmental model
- Calls provide a two week leading indicator

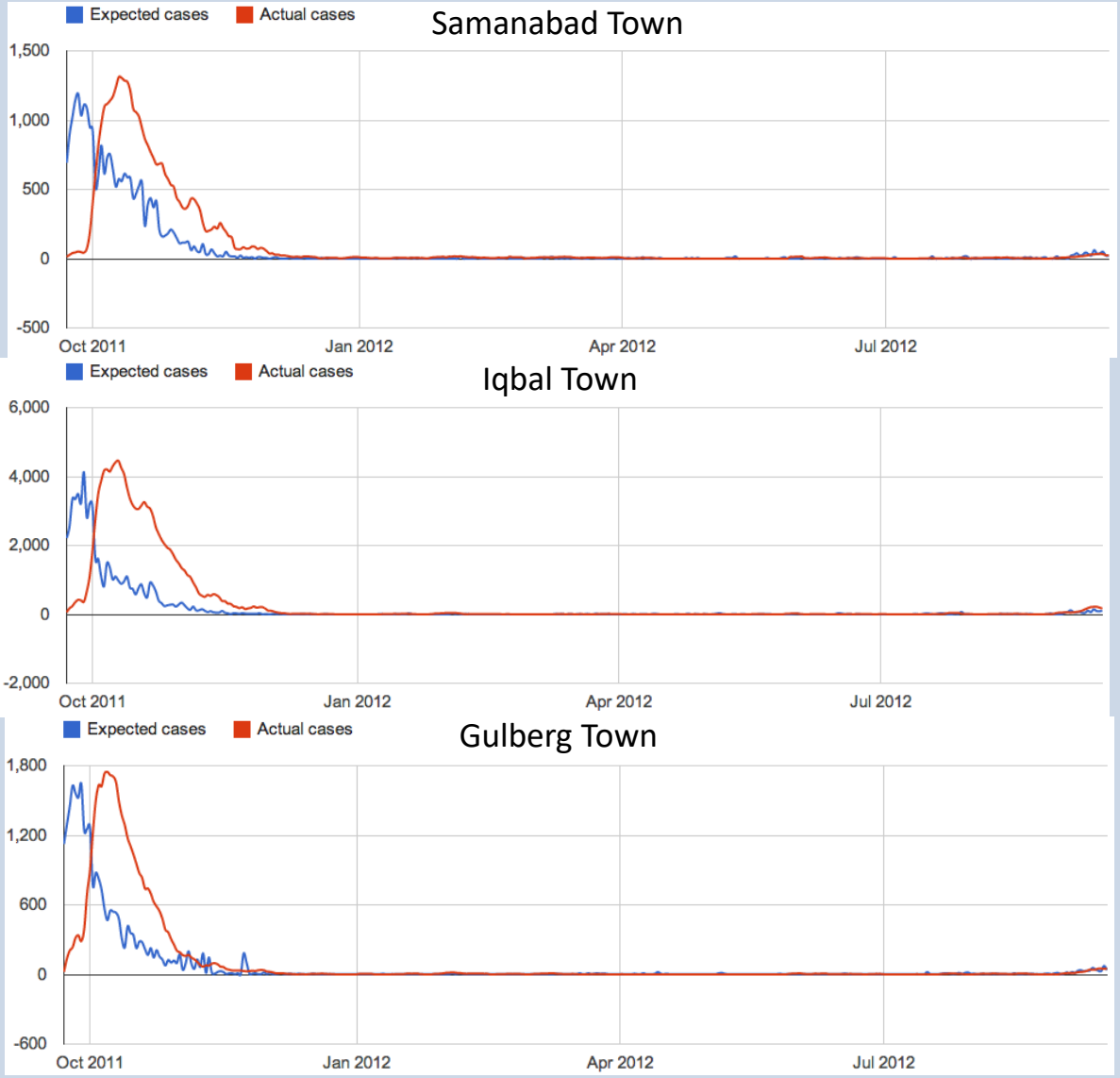
Setting: Punjab, Pakistan

- Dengue epidemic in 2011
  - 21,000 reported patients, at least 350 deaths
- Disease strikes during summer monsoon
- Big push for dengue control and monitoring
- Helpline introduced in 2012
  - Calls for information on dengue symptoms
  - Trained operators record caller information and make referrals if necessary
  - 300,000 calls received in 2012-2013

Health Hotline

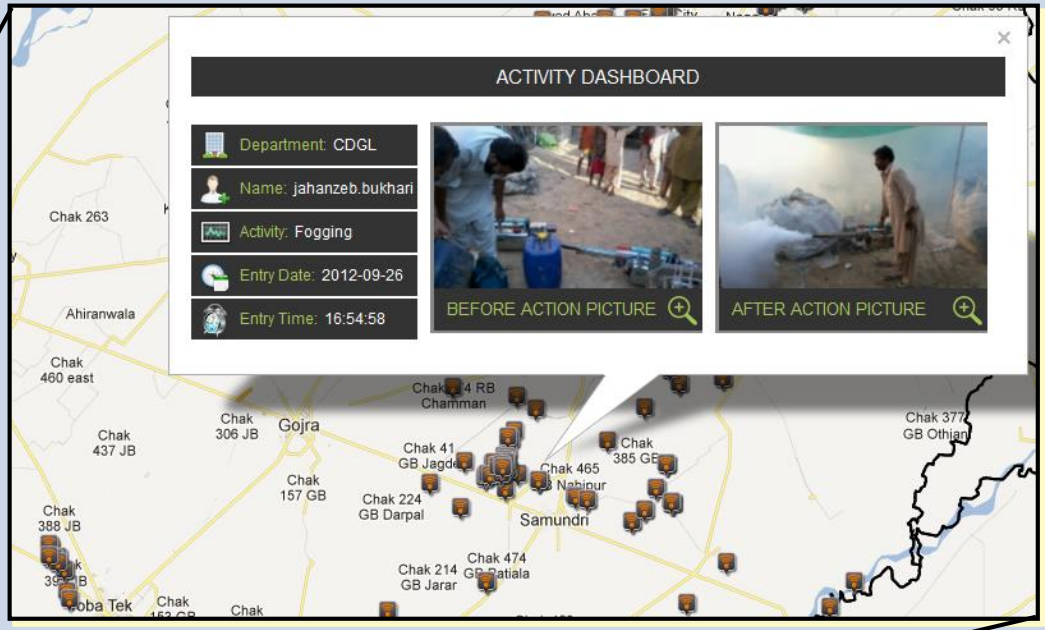


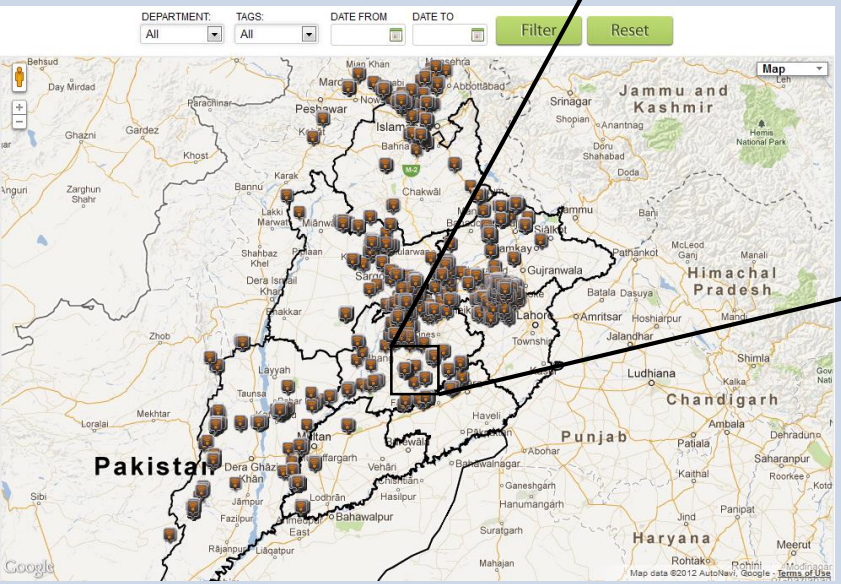
Health Hotline as an Early Predictor



Gives expected number of patient cases 11 days early

Preventive Activities





Real-time System: Automatically geo-tag pictures of an activity that is then uploaded to a centralized system

Goal: Predict cases at a local level based on call data

- How does calls to hotline predict suspected dengue cases
  - Comparing two measured quantities
- Are other variables needed?
  - Weather information
  - Awareness campaigns for hotline
- Localize across 10 towns of Lahore

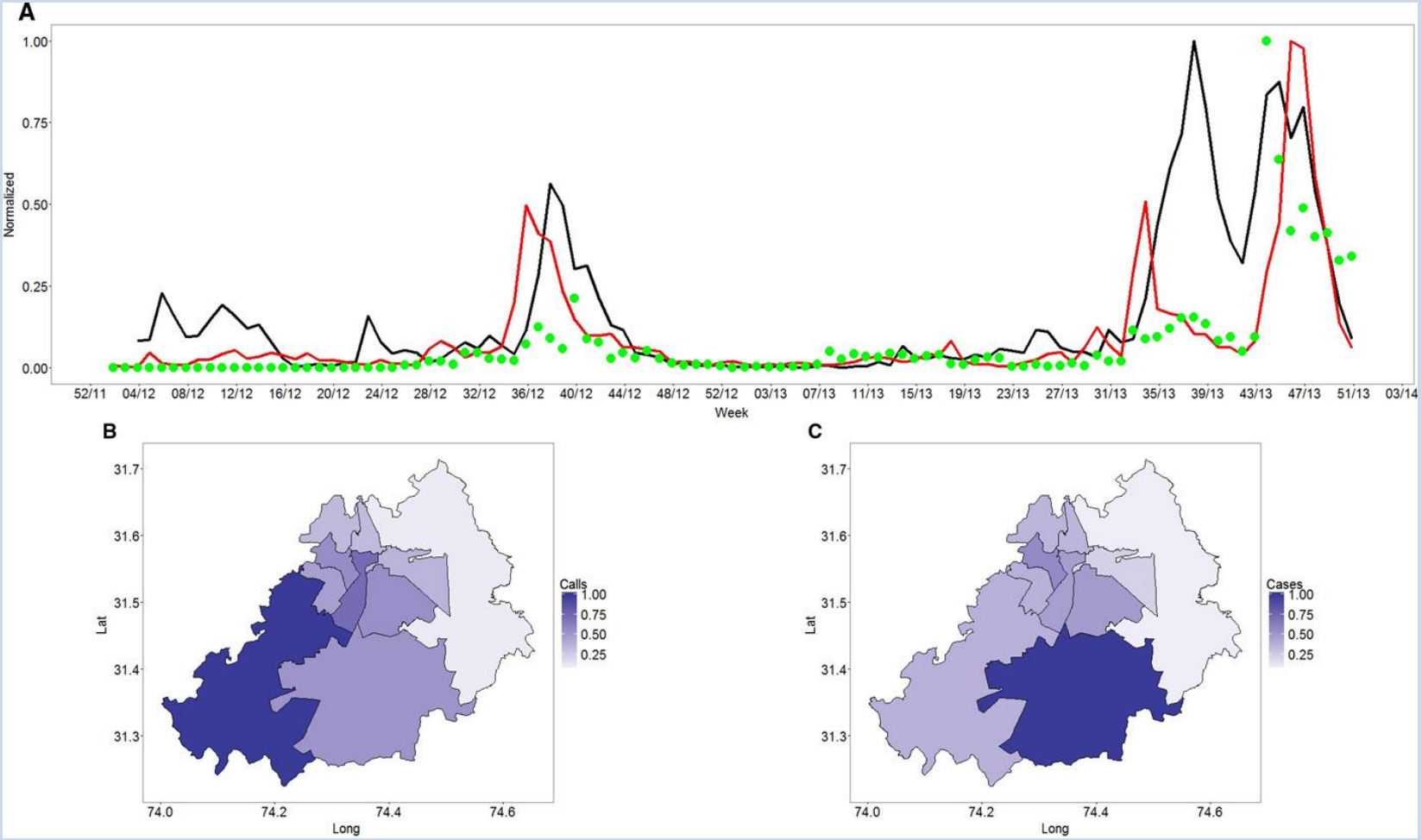


Fig. 1. Trends in call volume and suspected dengue cases measured during 2012 and 2013. (**A**) Time series of calls (red), suspected dengue cases (black), and awareness campaigns (green points). Scale normalized by dividing by individual maximum values. The *x*-axis label is in week of the year. (**B**) Density map of calls across towns in Lahore. (**C**) Density map of cases across towns in Lahore. The lightest shade represents the least number, and the darkest shade represents the highest number. The legend is normalized by the maximum value. Lat, latitude; long, longitude.

Nabeel Abdur Rehman et al. Sci Adv 2016;2:e1501215

ScienceAdvances
AAAS

Published by AAAS

Variables

Variable	Meaning
S(w,t)	Suspected cases week w, town t
C(w,t)	Calls week w, town t
A(w,t)	Awareness campaigns week w, town t
H(w)	Humidity week w
T(w)	Temperature week w
R(w)	Rainfall week w

Predict log S(w + 2, t) as a function of C(w,t), A(w,t), H(w), T(w), R(w)

Methodology

- Linear regression
- Random forest learning algorithm

- Estimates achieved good fit based on root mean square error
- Most important term, number of calls

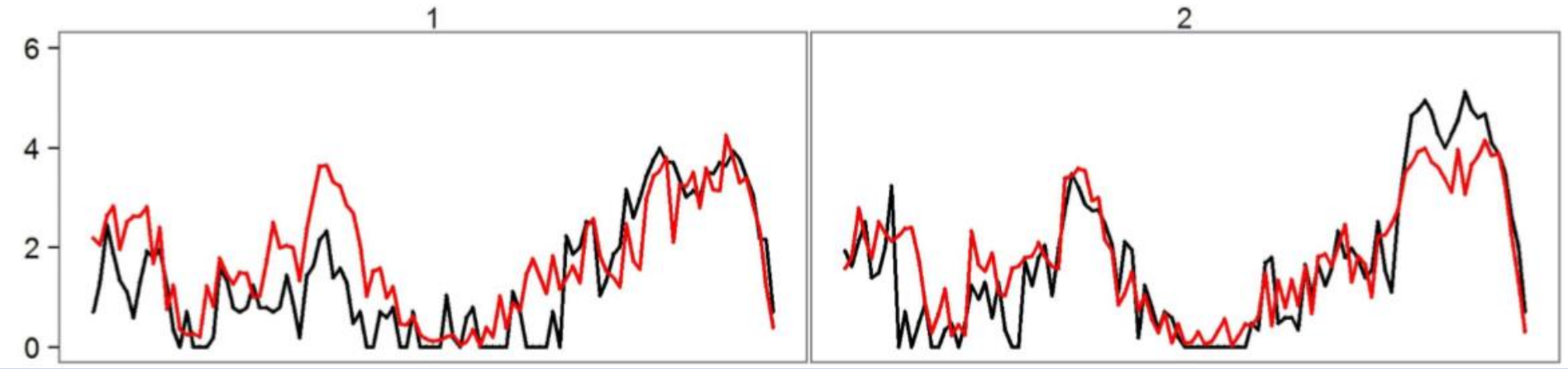


Fig. 2. Town-wise predictions of log-suspected cases from the ensemble model based on calls and weather data. Suspected dengue cases (black) and predictions from the model (red).

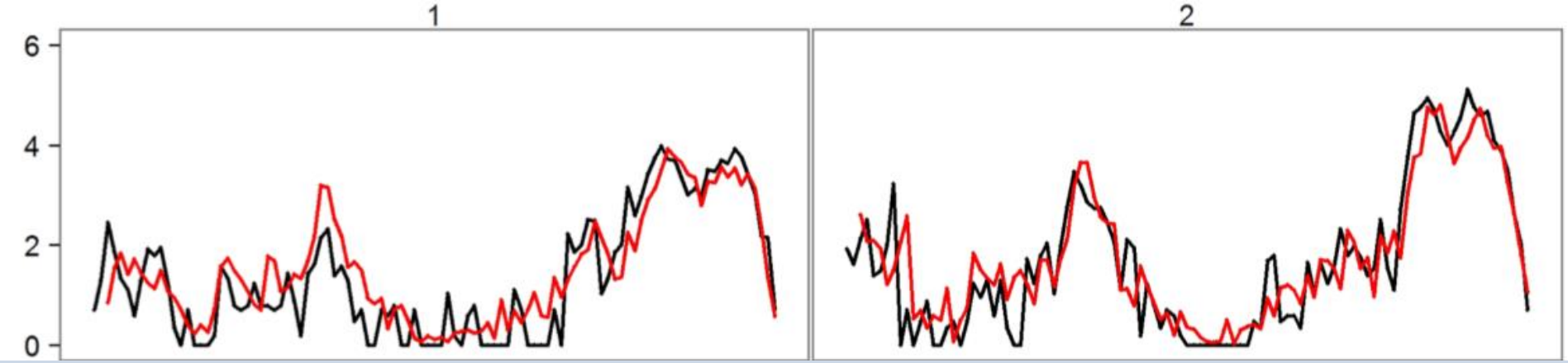


Fig. 3. Town-wise predictions of log-suspected cases from the ensemble model based on calls, cases, and weather data. Suspected dengue cases (black) and predictions from the model (red).

Validation and Deployment

- Validation done through standard methods of generating models for randomly constructed subsets and confirming on held out data
- Validation and cross-correlations as a supplement to the journal article
- Model deployed as part of dengue surveillance system, accessible through a web api