

MOBILE DATA ANALYTICS:

High-Impact Analysis for Social Good (and for Profit)

Dr. Joshua Blumenstock

University of Washington





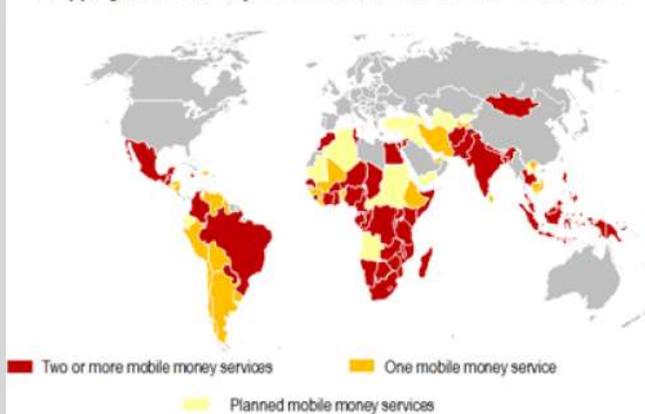
THIS TALK: OUTLINE

- **Mobile Data Overview**
- **Business Analytics**
 - Predicting customer churn
 - Optimizing referral marketing campaigns
- **Big Data for Social Good**
 - Detecting crises from call records
 - Models of migration and mobility
 - High-Resolution poverty maps

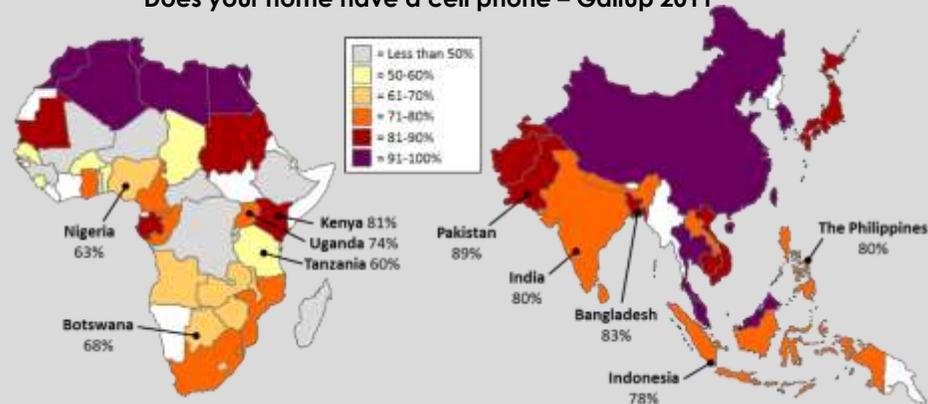
BACKGROUND AND MOTIVATION

- **“Mobile Phone Revolution” in developing countries**
 - 4 billion subscribers in developing countries
 - 1 new subscriber per second in Nigeria
- **Mobile Money and Value Added Services**
 - 2 billion unbanked phone owners
 - Mobile Money and Data rapidly expanding
 - Airtel (Africa): \$2B in mobile money in 2014 Q1

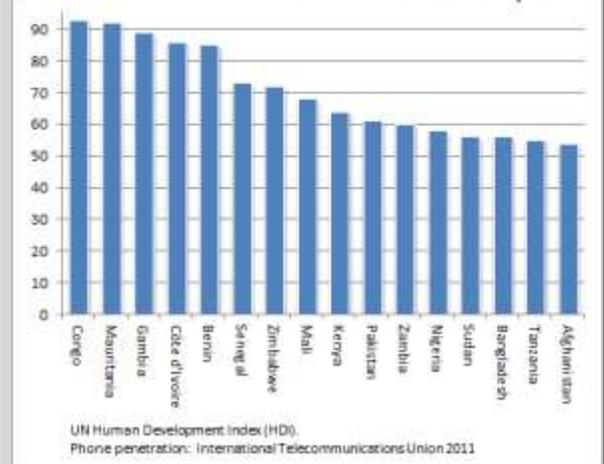
Mapping mobile money services for the unbanked – October 2013



Does your home have a cell phone – Gallup 2011



Mobile phone penetration (per 100 people) - countries with low human development



MOBILE DATA OVERVIEW

- **Transactional Data and Call Detail Records (CDR)**
 - **Call and SMS:** sender, receiver, time, duration
 - **Recharge and Voucher:** time, denomination, location
 - **Mobile Money:** deposits, withdrawals, transfers, payments, etc.
 - Including: HLR, VOU, HLR, PPS, MM, INDUMP, MSC/RCP/SSP, etc.
- **Example Call CDR:**

A-Party-ID	B-Party-ID	Date	Time	Duration	A-Party-Cell	...
979ae8gc7d	979f339b87	2014-01-04	22:00:11	42	2837	...

- **Note:** One month of CDR can be /1-100 terabytes

INFRASTRUCTURE AND ANALYTICS



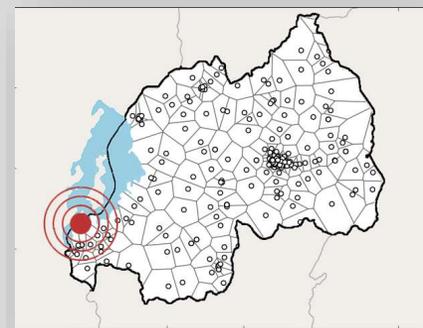
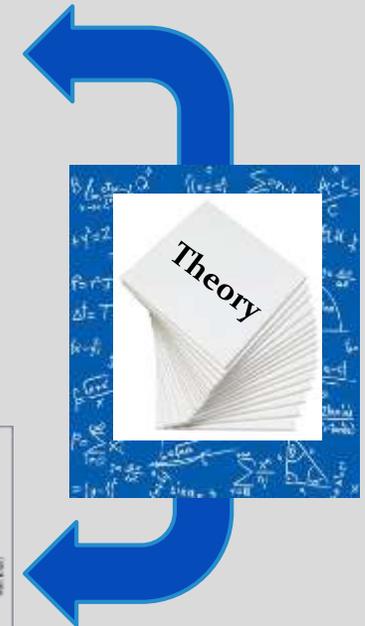
Data Acquisition:
PHP, python, JSON, XML, bash



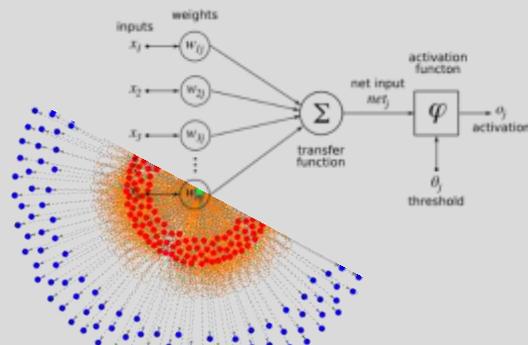
ETL and Storage
Hadoop/Hive, SQL, Spark/Spark, GraphLab



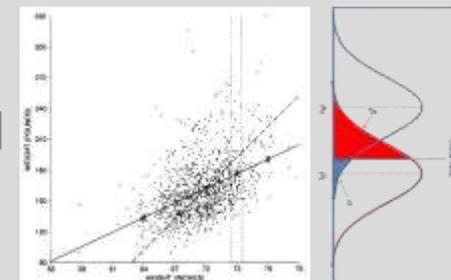
Preprocessing
Java, Scala, php/python



Visualization
Processing, D3, Adobe



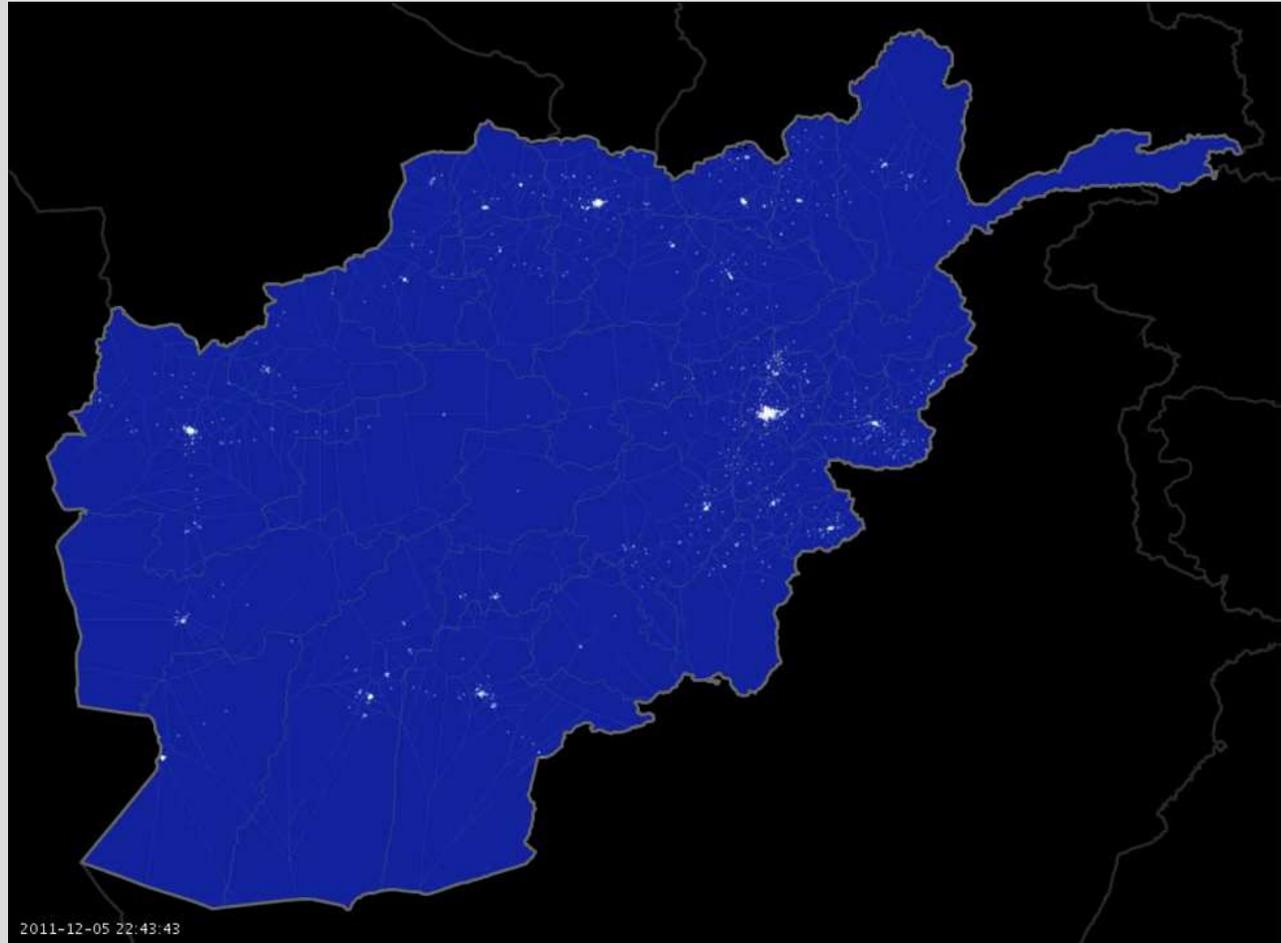
Machine Learning
Java, MatLab, Stata



Statistical Analysis
R, MatLab, Stata

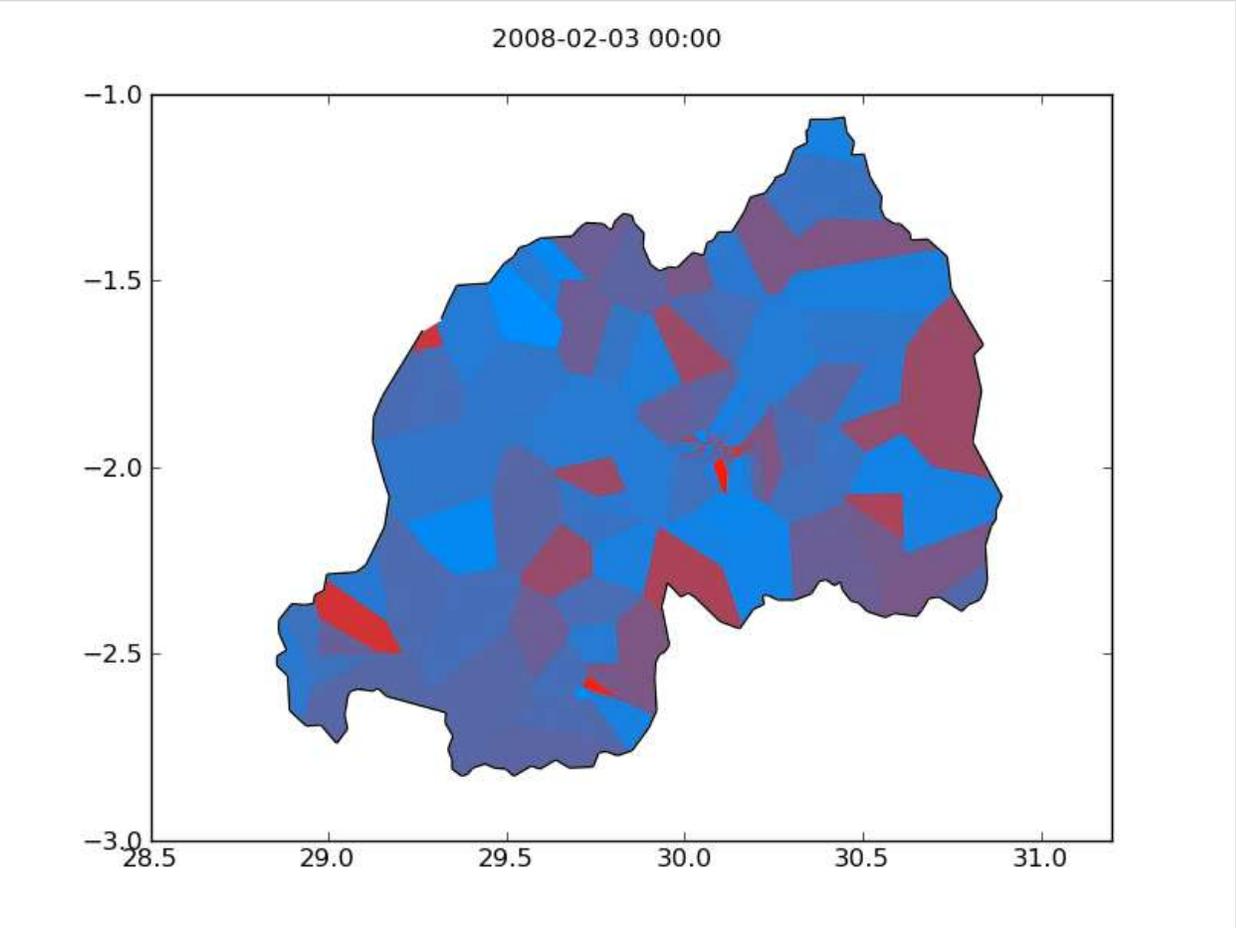
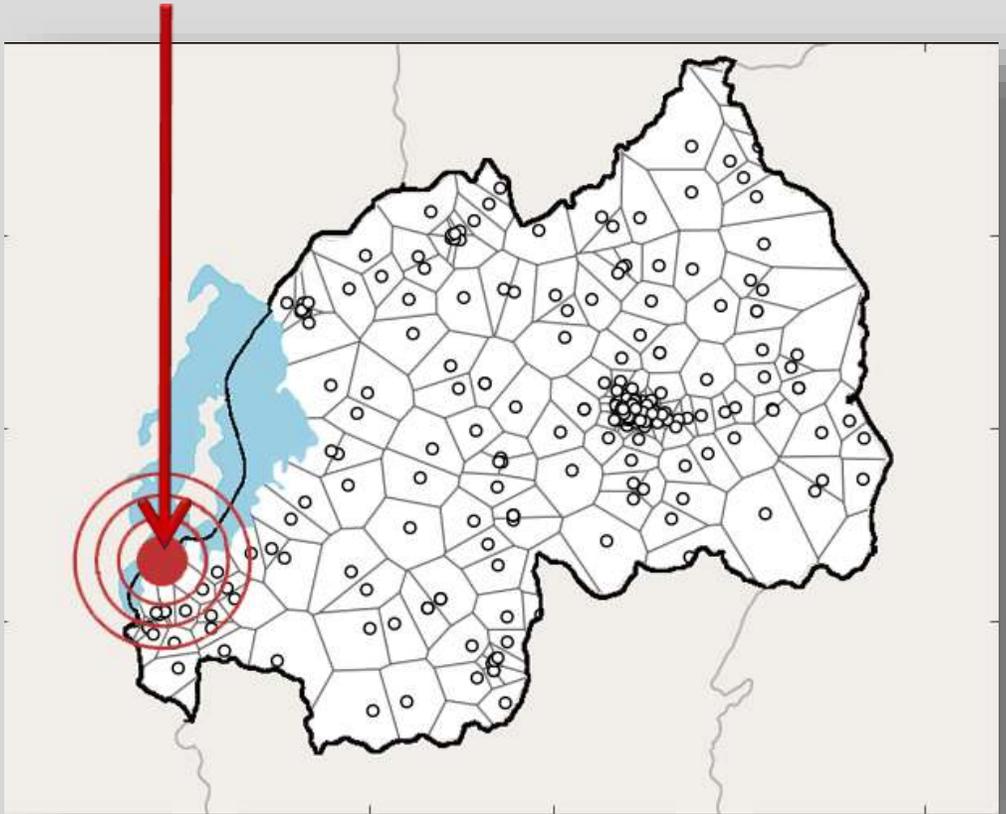


WHAT CAN WE LEARN FROM CDR?



DATA REFLECTS SOCIETY

Lake Kivu Earthquake





THIS TALK: OUTLINE

- ~~Mobile Data Overview~~
- **Business Analytics**
 - Predicting customer churn
 - Optimizing referral marketing campaigns
- **Big Data for Social Good**
 - Detecting crises from call records
 - Models of migration and mobility
 - High-Resolution poverty maps

BUSINESS ANALYTICS

Sample Project: **Machine Learning for Churn Prediction**

- *What are the “early warning signs” of churn?*
- *Which customers are most likely to become inactive?*

Input Data:

6 months of complete CDR

- Call CDR, SMS CDR, HLR, etc.
- Voucher activity (top up, transfer)
- Mobile Money records

Target Data:

“Churn” over 3 months

- “Total churn”: Complete inactivity
- “Relative churn”: % of inactivity

Feature Extraction:

>5,000 raw features

Based on Combinatorics:

- Calls vs. SMS vs. MM
- Outgoing vs. Incoming
- Duration vs. Count
- In-Network vs. Out-Network
- Domestic vs. International
- Time of Day
- Special Numbers

Early Warning Signs of Churn



BUSINESS ANALYTICS

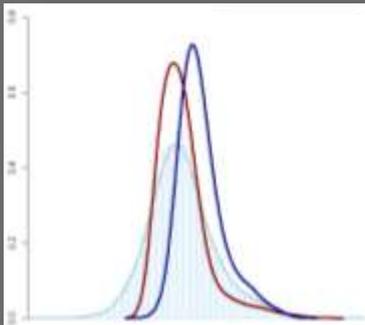
Sample Project: **Machine Learning for Churn Prediction**

- What are the “early warning signs” of churn?
- Which customers are most likely to become inactive?

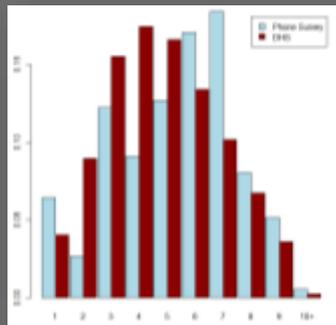
Machine Learning Models

- Principal Component Analysis
- Support Vector Machines
- Random Forest Classifiers
- Regression and Logistic Regression

Churn vs. Stay



Performance vs. Baseline



Predicting Churn

- Still a work in progress!

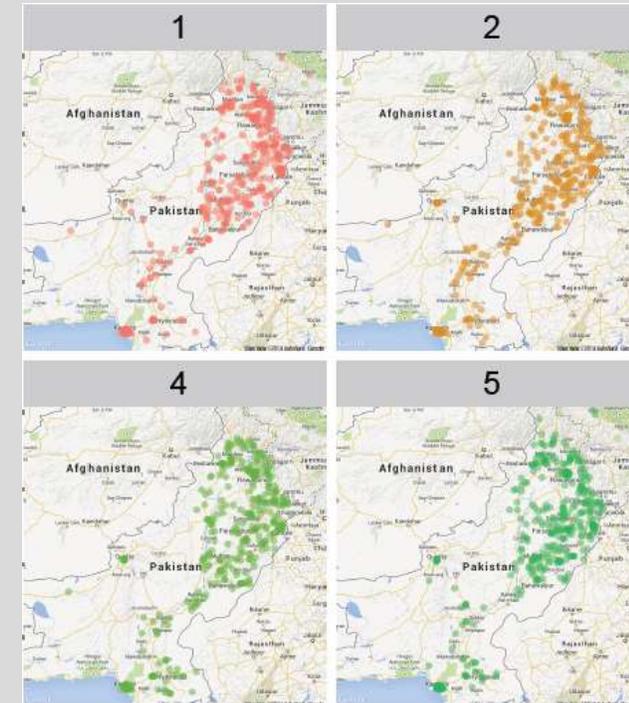


BUSINESS ANALYTICS



Sample Project: **Optimizing a Referral Incentive Program**

- Referral bonuses if friends become **active** MM users
- Product marketing optimization
 - What information makes people spread the word?
 - Where are the optimal “injection points” of new information?
 - What incentives create sustained use?
- 12 Different “Treatments” (A-B Testing)



Financial
Incentives



Product
Information



Low
Touch



Social
Incentives



Benefit
Information



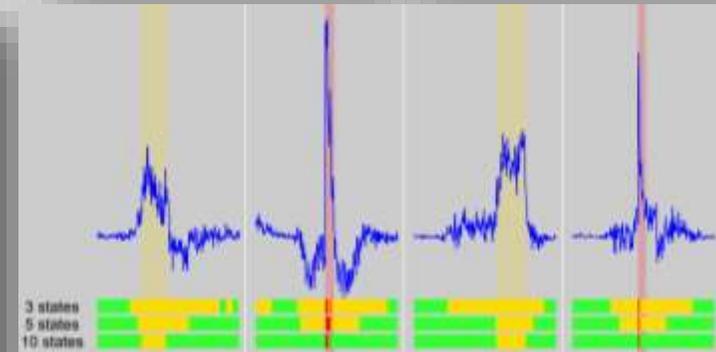
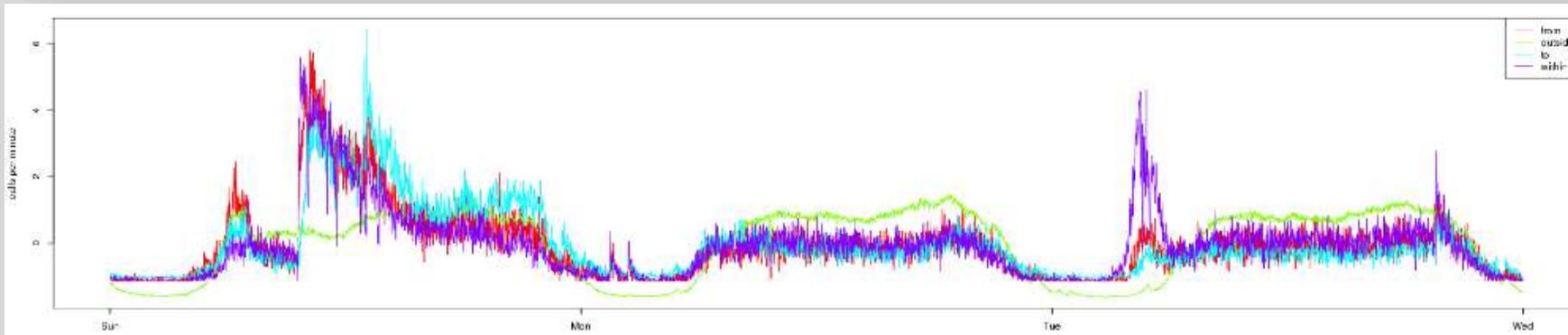
High
Touch



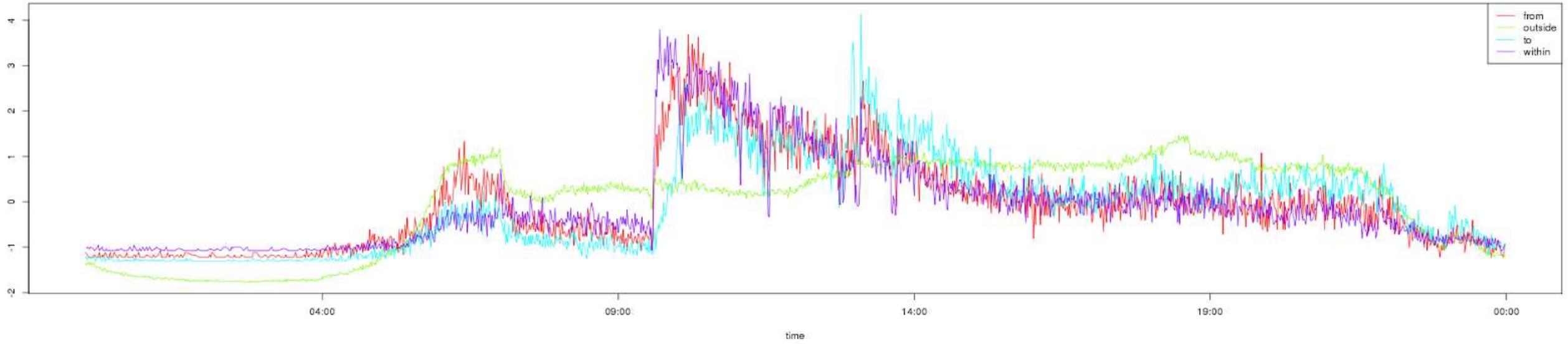
BIG DATA FOR SOCIAL GOOD

Sample Project: **Detecting Crises from Call Detail Records**

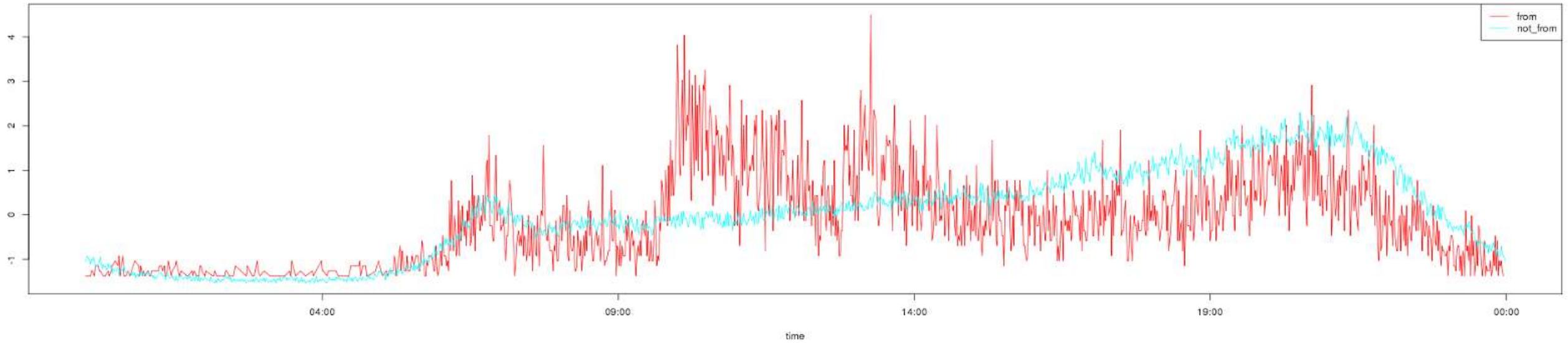
- How to identify the **time**, **location**, and **severity** of crisis events?
- Early Warning Systems
- Real-time assistance

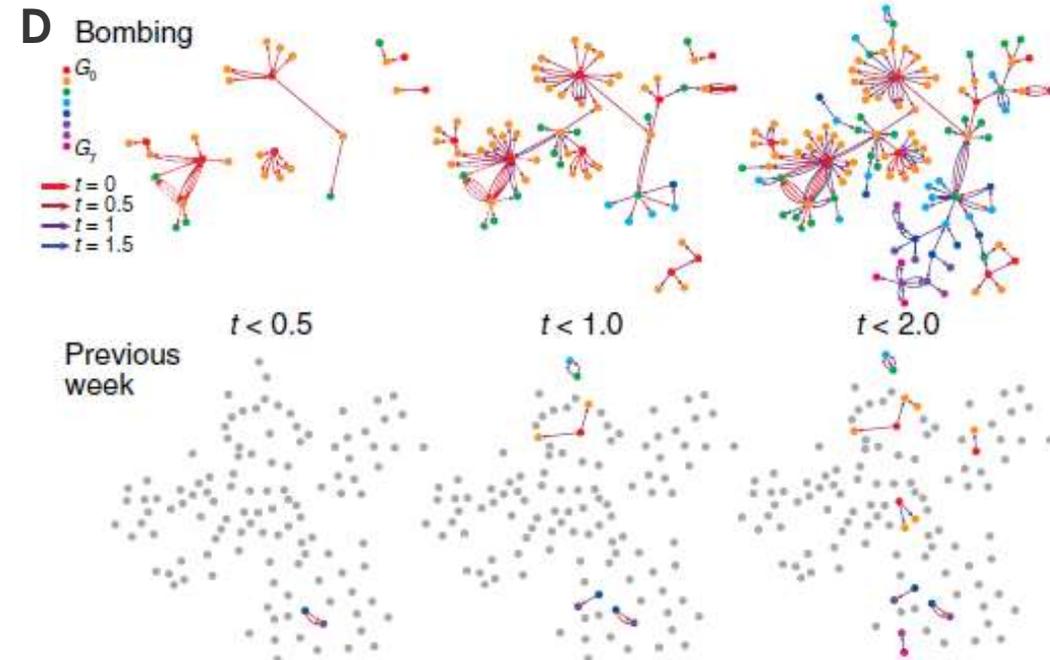
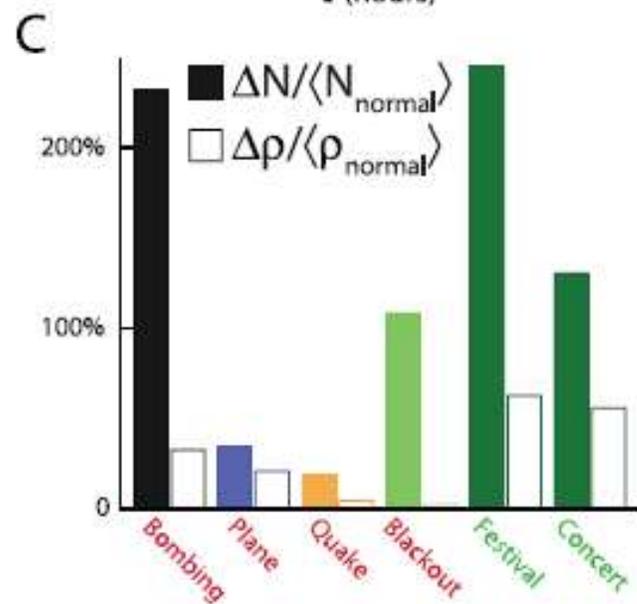
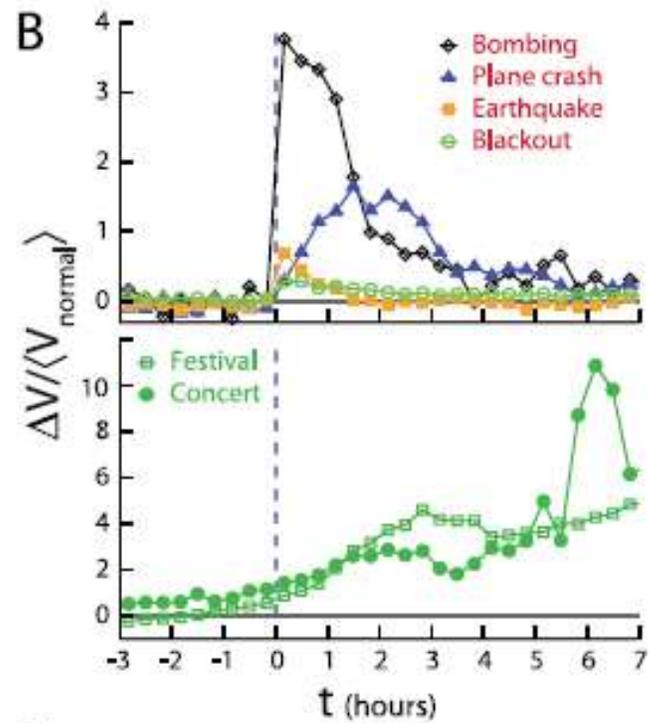
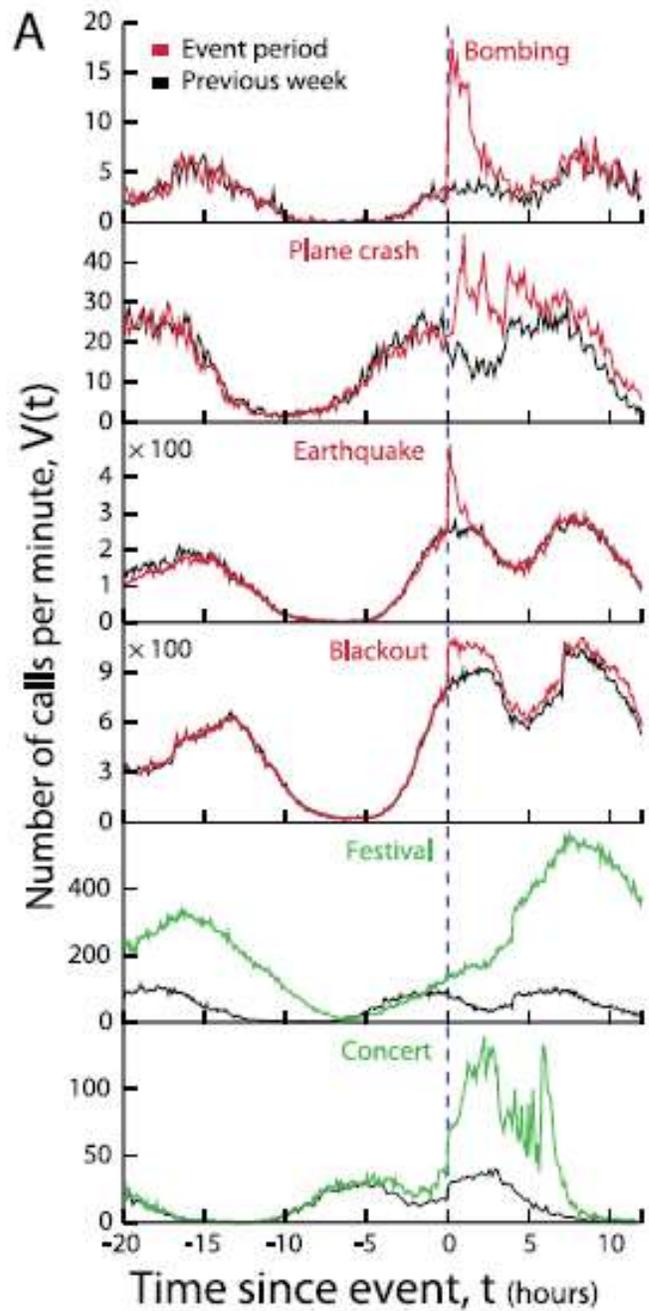


Calls



Text Messages

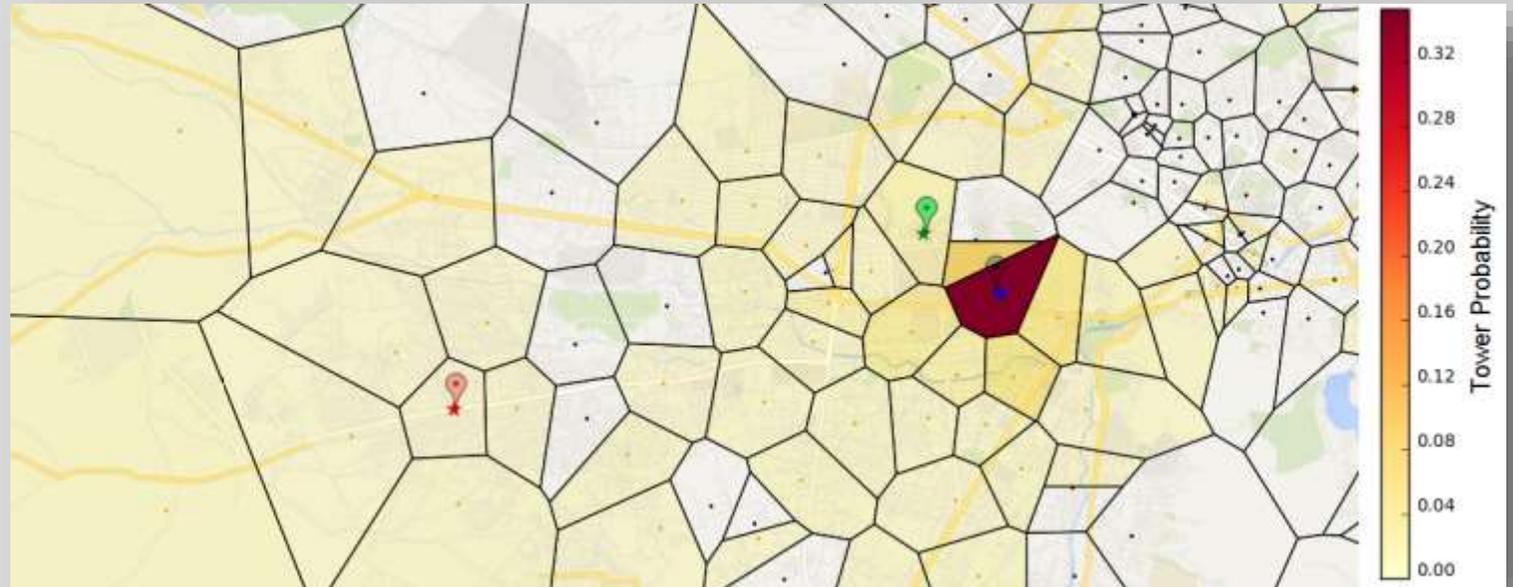
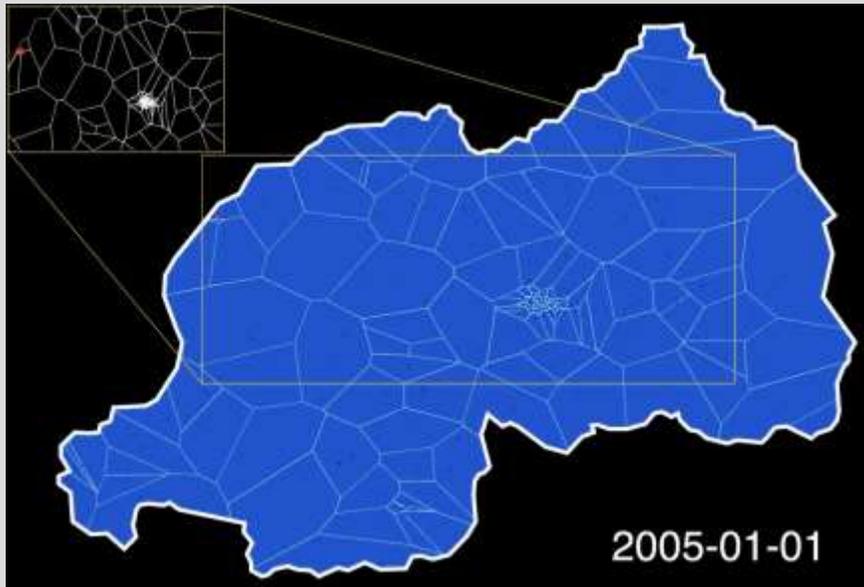




BIG DATA FOR SOCIAL GOOD

Sample Project: **Measuring Migration and Mobility**

- Where do people move after crisis events?
- How does migration affect the economy?



BIG DATA FOR SOCIAL GOOD

Sample Project: **High-Resolution Development Indicators**

- **Problem:** Unreliable data on key development indicators
 - Household surveys require significant investment and infrastructure
 - Angola's last national census was in 1970
- **Solution:** Predict poverty and vulnerability directly from CDR
 - At regional level (e.g. political district)
 - At individual level
 - Detect changes over time

BILL & MELINDA
GATES foundation



USAID
FROM THE AMERICAN PEOPLE

DFID
Department for
International
Development



BIG DATA FOR SOCIAL GOOD

Sample Project: High-Resolution Development Indicators

Data Assembly and Linkage

Passive Data

- Call Records
- SMS Records
- Networks
- Event Logs
- News Articles

```

1|11200756|2260849|1|15346499|11|08743747|08448569|01/01/2008 01:01:25|38|0|NBSC2I|SSFDJ01|635105001287432|35623200
0|11223401|2323394|1|15602242|08633583|03219934|01/01/2008 01:04:23|5|2|NBSC2I|GRINRM|113|5873|11111111
5|11223402|15635992|22|08660728|08110333|01/01/2008 01:04:28|0|NBSC2I|635105001279454|355510011936690|10-00104-1
5|11223403|15636006|22|08431310|08110333|01/01/2008 01:04:28|0|NBSC1I|635105000131310|357969004390900|10-00103-0
5|11223404|15636020|22|03119506|08110333|01/01/2008 01:04:28|0|NBSC1I|635105001514790|352323016463600|10-00103-1
5|11223405|15636025|22|08520255|08110333|01/01/2008 01:04:28|0|NBSC1I|635105000695432|357077005452950|10-00103-0
2|11223406|2310291|1|15547505|08618091|08847090|01/01/2008 01:03:48|40|0|NBSC1I|635105000786626|113|5873|11111111
2|11223407|2290567|1|15469250|08823243|08841803|01/01/2008 01:02:47|101|0|NBSC1I|635105000784476|113|5873|11111111
5|11223408|15636091|22|08476273|08110333|01/01/2008 01:04:28|0|NBSC1I|635105000176273|359333006069290|10-00103-0
5|11223409|15636100|22|08754900|08110333|01/01/2008 01:04:28|0|NBSC2I|635105001368709|356402014233580|10-00104-1
2|11223410|2316505|1|15573404|08467584|03140554|01/01/2008 01:03:52|35|0|NBSC1I|635105001535857|113|5873|11111111
5|11223411|15636109|22|0032498545307|0032495002530|01/01/2008 01:04:28|0|NBSC1I|206102200548388|356036017797950|
2|11223412|2296485|1|15493712|08442497|08427583|01/01/2008 01:03:01|87|0|NBSC1I|635105001654671|113|5873|11111111
4|11223413|2310426|1|15549853|11|08640680|08533203|01/01/2008 01:03:50|37|0|NBSC1I|NBSC10|63510500033203|353153010
    
```

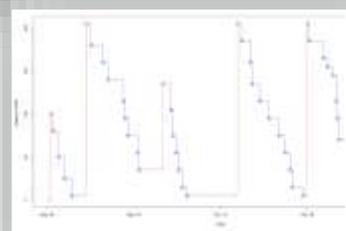
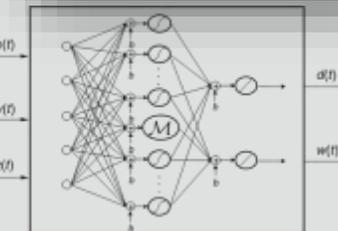
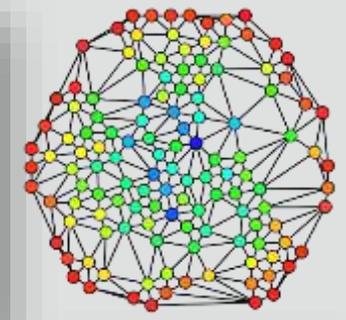
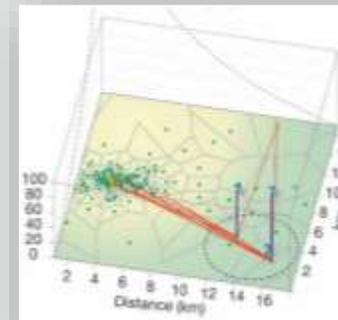
+

Survey Data

- Phone-based
- Face-to-face
- Census data



Behavioral Machine Learning

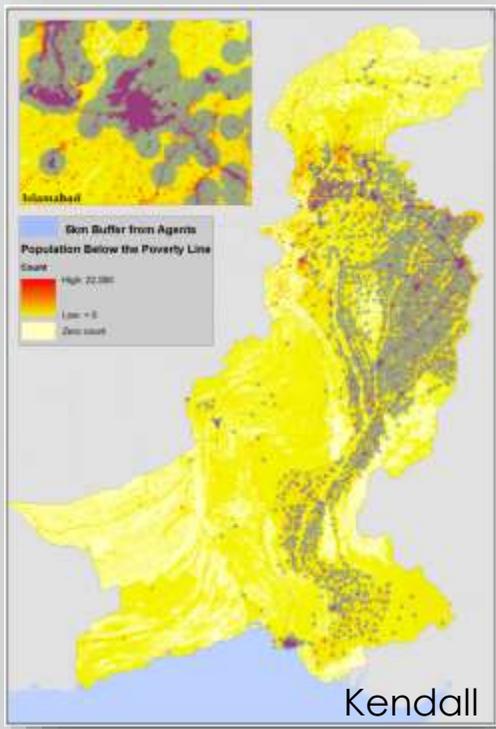


$$U_{it} = \underbrace{u_i(x_{it} + \tau_{ijt}) + \gamma u_j(x_{jt} - \tau_{ijt})}_{\text{single period utility}} + E \sum_{s=t+1}^{\infty} \delta^{s-t} \underbrace{[u_i(x_{is} + \tau_{ijs}) + \gamma u_j(x_{js} - \tau_{ijs})]}_{\text{continuation value of relationship}}$$

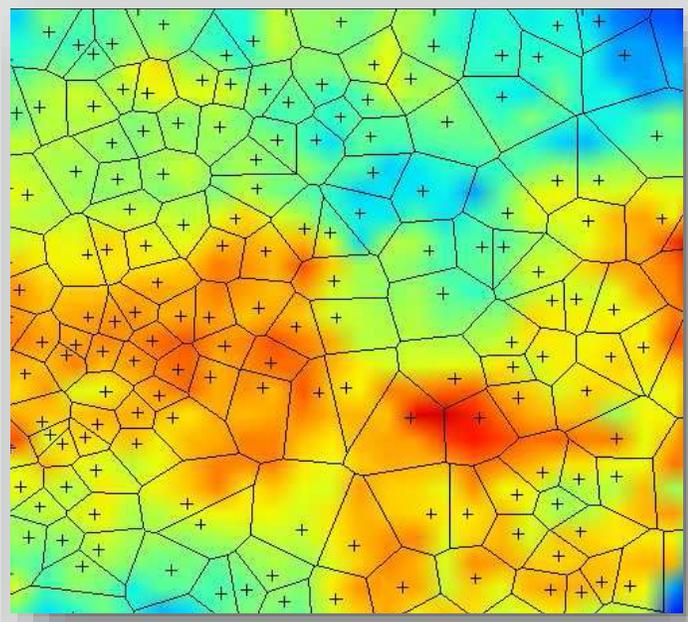
BIG DATA FOR SOCIAL GOOD

Sample Project: High-Resolution Development Indicators

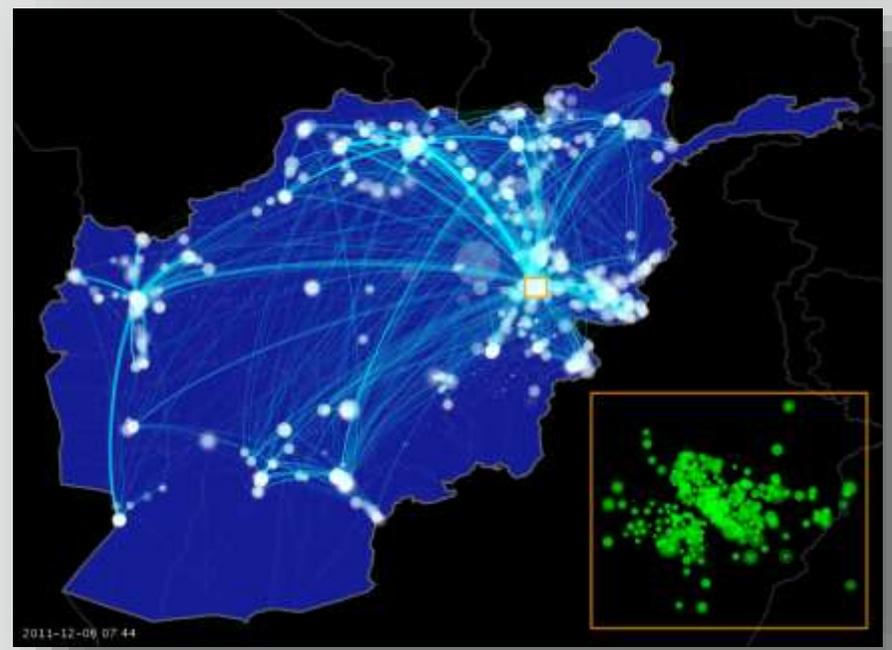
High-Resolution Poverty Maps



Advanced Demographic Segmentation



Risk and Vulnerability Assessment



SUMMARY

With Big Data comes Big Responsibility...

...and Big Opportunity!

- For profit
 - Churn detection and customer profiling, marketing and pricing, behavioral segmentation, product design and optimization
- For solving important societal problems
 - Mapping poverty and vulnerability, crisis detection and early response, impact evaluation, basic science research
- Many possible opportunities for partnership!



Thanks for your attention

Comments, Questions, Feedback are welcome!

joshblum@uw.edu

<http://www.jblumenstock.com>

<http://datalab.ischool.uw.edu>