

FINDING STORIES IN DATA

Local voices. Global change.

Contents

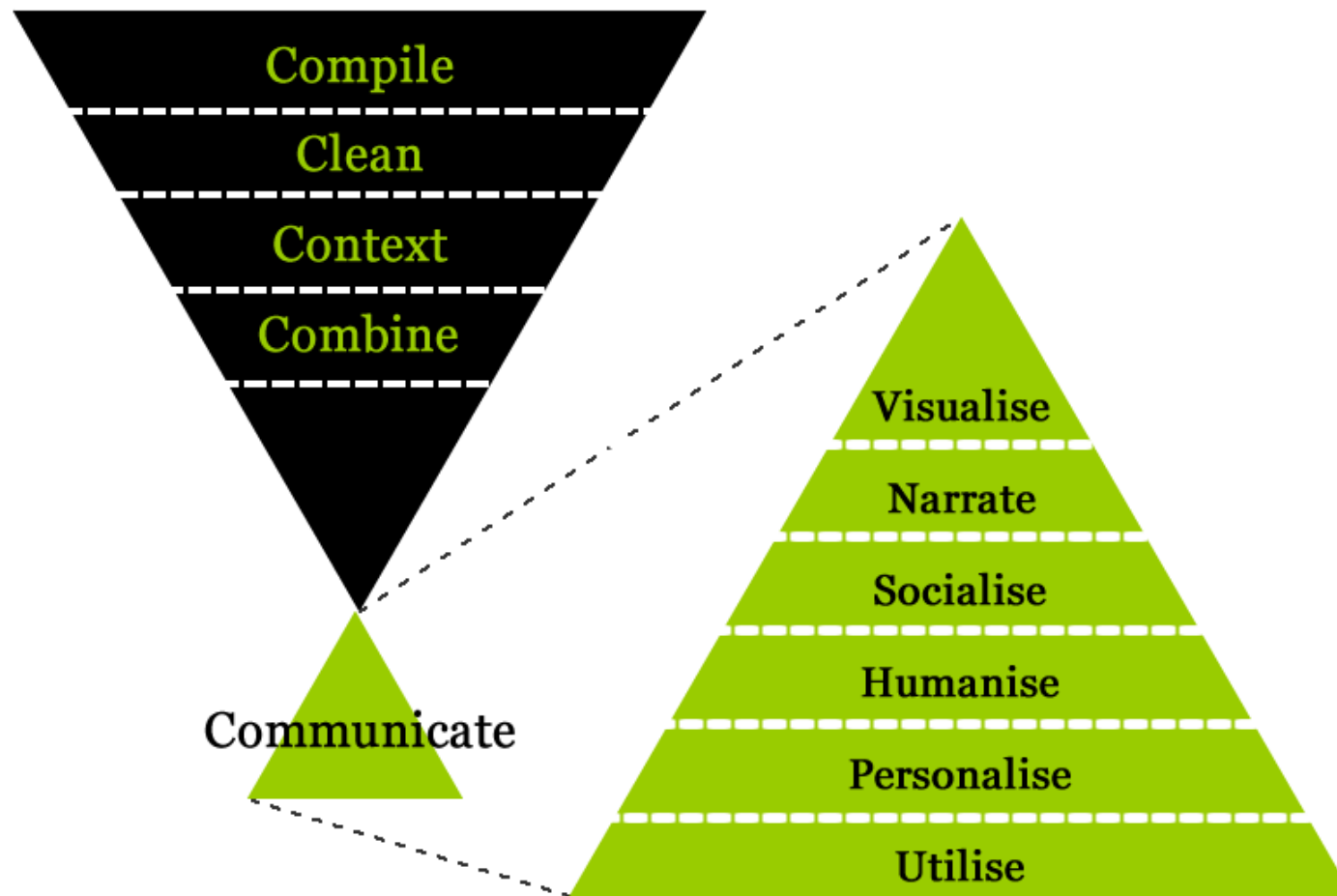
- The Data Journalism Process
- Data sources: recurring events
- Exercise
- Data sources: sent or leaked data
- Data sources: breaking news
- Data sources: theories to be explored
- Exercise
- Summary & References



The Data Journalism Process

The Inverted Pyramid of Data Journalism (complete)

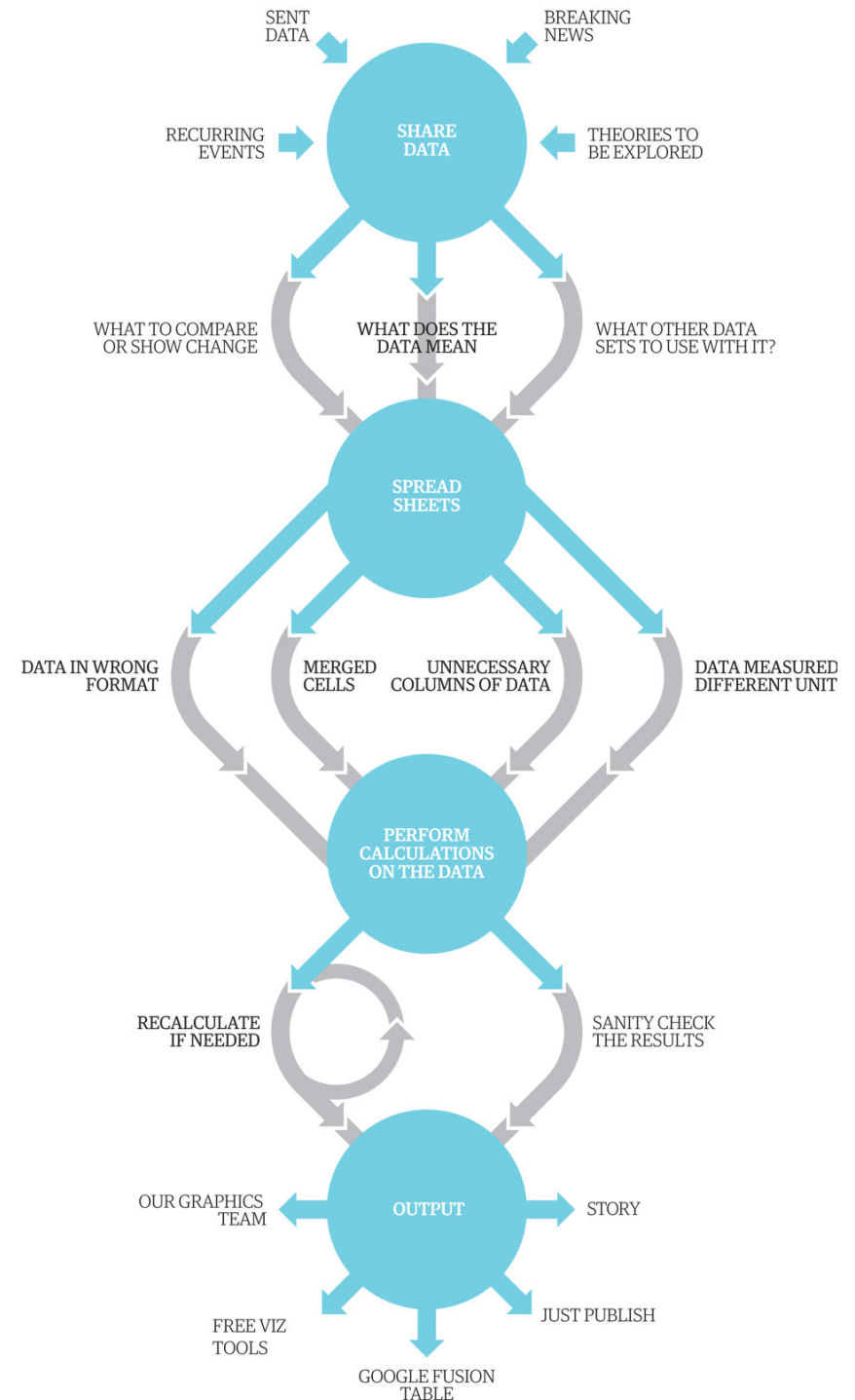
Paul Bradshaw, OnlineJournalismBlog.com



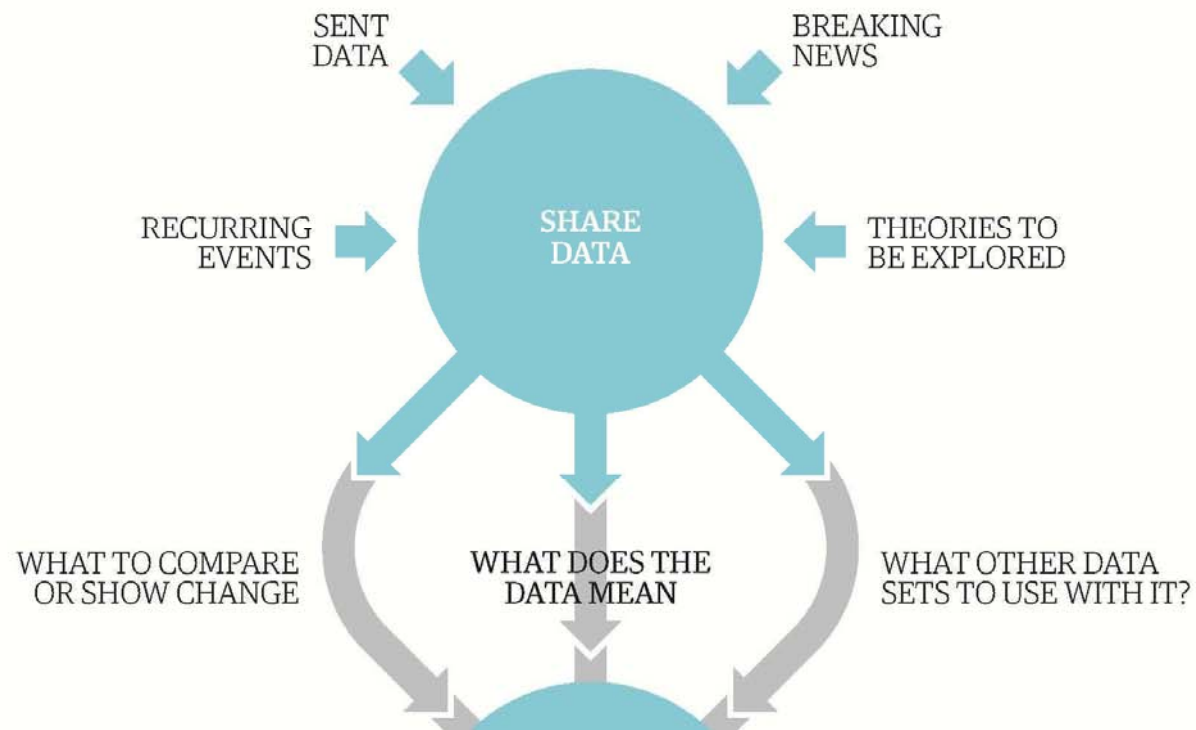
The Data Journalism Process

The Guardian:

1. Share Data
2. Spread Sheets
3. Perform Calculations
4. Output



The Data Journalism Process



Internews
Local voices. Global change.

Data sources

- **Data sources: recurring events**
- Data sources: sent or leaked data
- Data sources: breaking news
- Data sources: theories to be explored

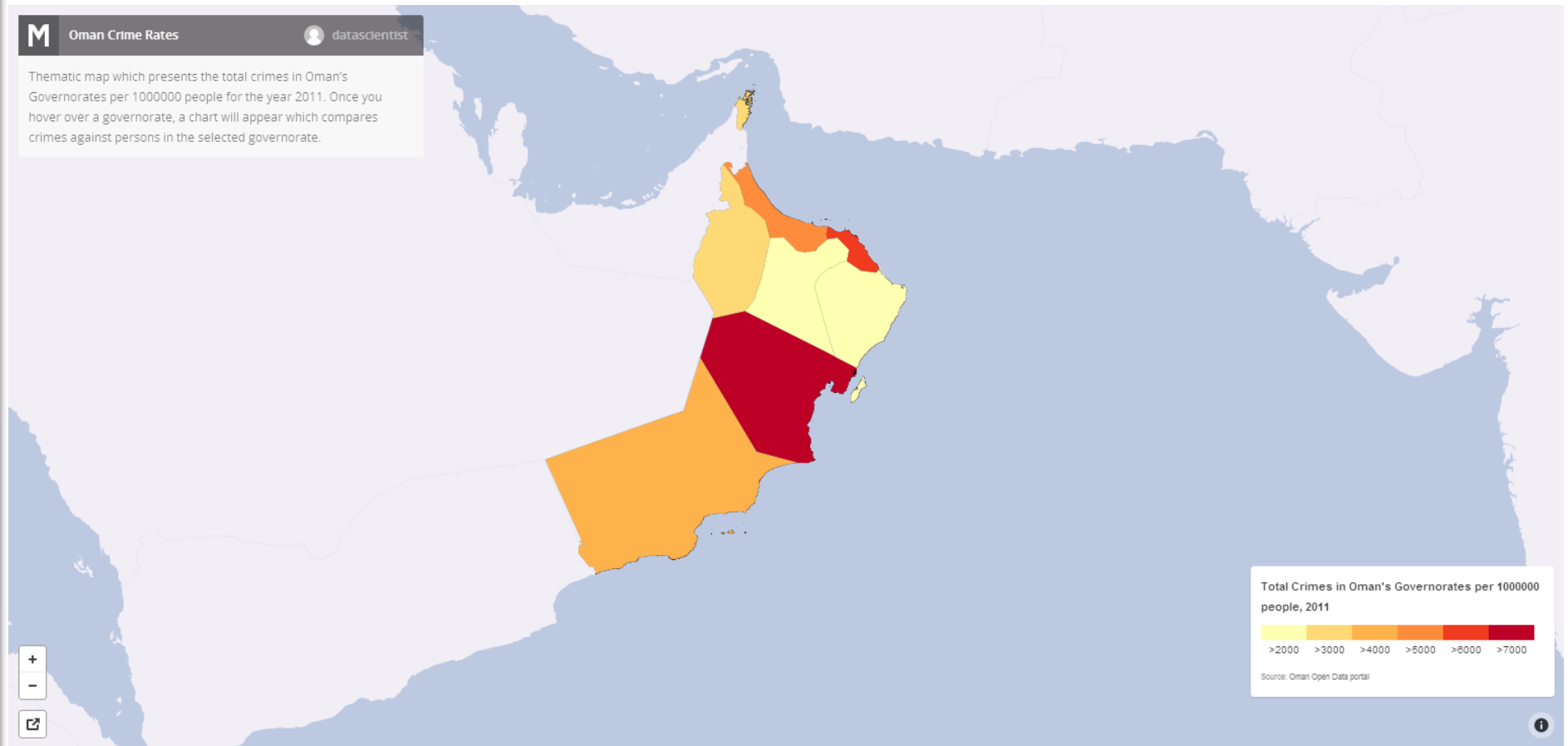
Data sources

- **Data sources: recurring events**
 - Governance & Security Indicators
 - Economic Indicators & Budgets
 - Education Indicators
 - Human rights & Equality Indicators
 - Environmental Indicators
 - Health Indicators

Data sources

- Data sources: recurring events
 - **Governance & Security Indicators**
 - Economic Indicators & Budgets
 - Education Indicators
 - Human rights & Equality Indicators
 - Environmental Indicators
 - Health Indicators

Data Sources: Recurring Events



Crime In Oman Visualized: For the first time, the government of Oman released crime data, which was visualized by a local blogger



Internews
Local voices. Global change.

Data Sources

- Data Sources: recurring events
 - Governance & Security Indicators
 - **Economic Indicators & Budgets**
 - Education Indicators
 - Human rights & Equality Indicators
 - Environmental Indicators
 - Health Indicators



Data Sources: Recurring Events

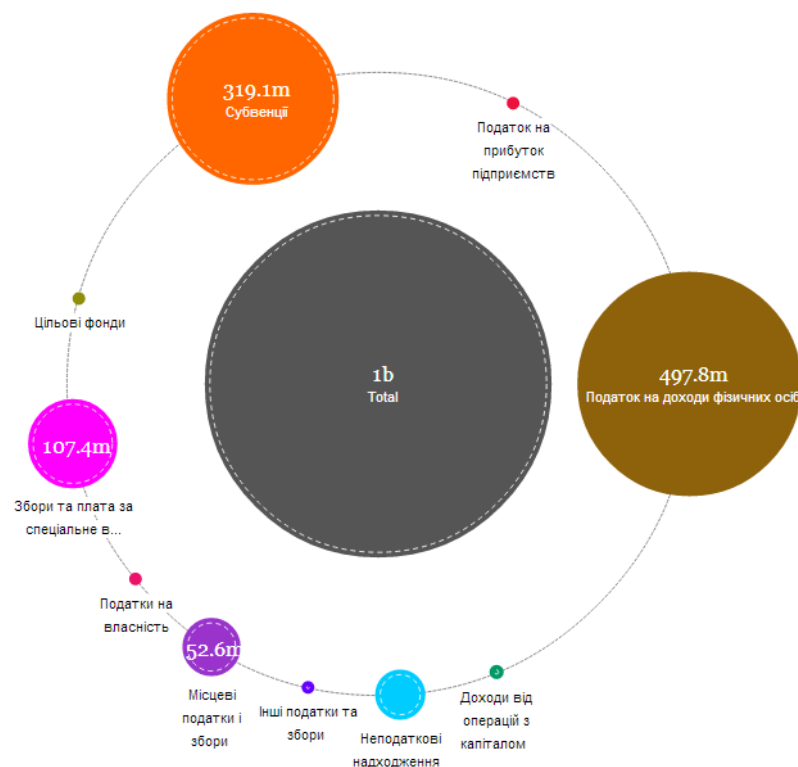
Ukraine 2013 budget: A Ukrainian newspaper created an interactive data visualization when the government released its annual budget so that citizens could evaluate how the government is spending its tax dollars



26 грудня 2012 року Сумська міська рада схвалила міський бюджет на 2013 рік, але при цьому велике занепокоєння викликає перспектива його виконання з причин:

- обсяг доходів, які враховуються при визначенні міжбюджетних трансфертів, затверджений в сумі 498651,1 тис. грн. (додаток 6 Закону України "Про Державний бюджет на 2013 рік"), що на 67696,6 тис. грн., або на 15,7% більше очікуваного надходження в 2012 році. При цьому, констатуємо, що невиконання затвердженого показника прогнозується в сумі 32183,3 тис. грн.

Доходи бюджету м. Суми в 2013 році

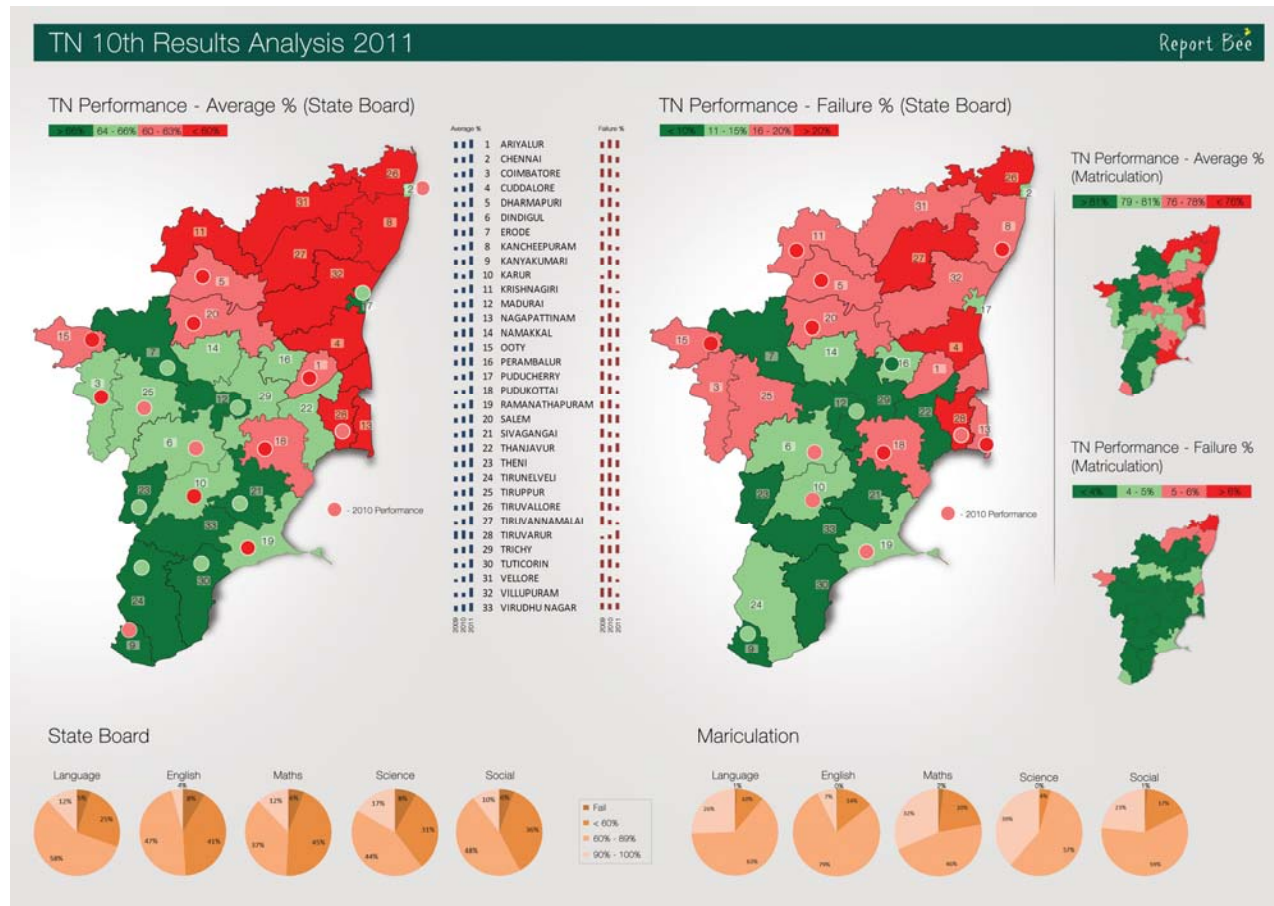


Data Sources

- Data Sources: recurring events
 - Governance & Security Indicators
 - Economic Indicators & Budgets
 - **Education Indicators**
 - Human rights & Equality Indicators
 - Environmental Indicators
 - Health Indicators



Data sources: Recurring Events



2011 Tamilnadu School Performance: India's Reporter Bee visualized the performance of different school districts based on annual performance reports.



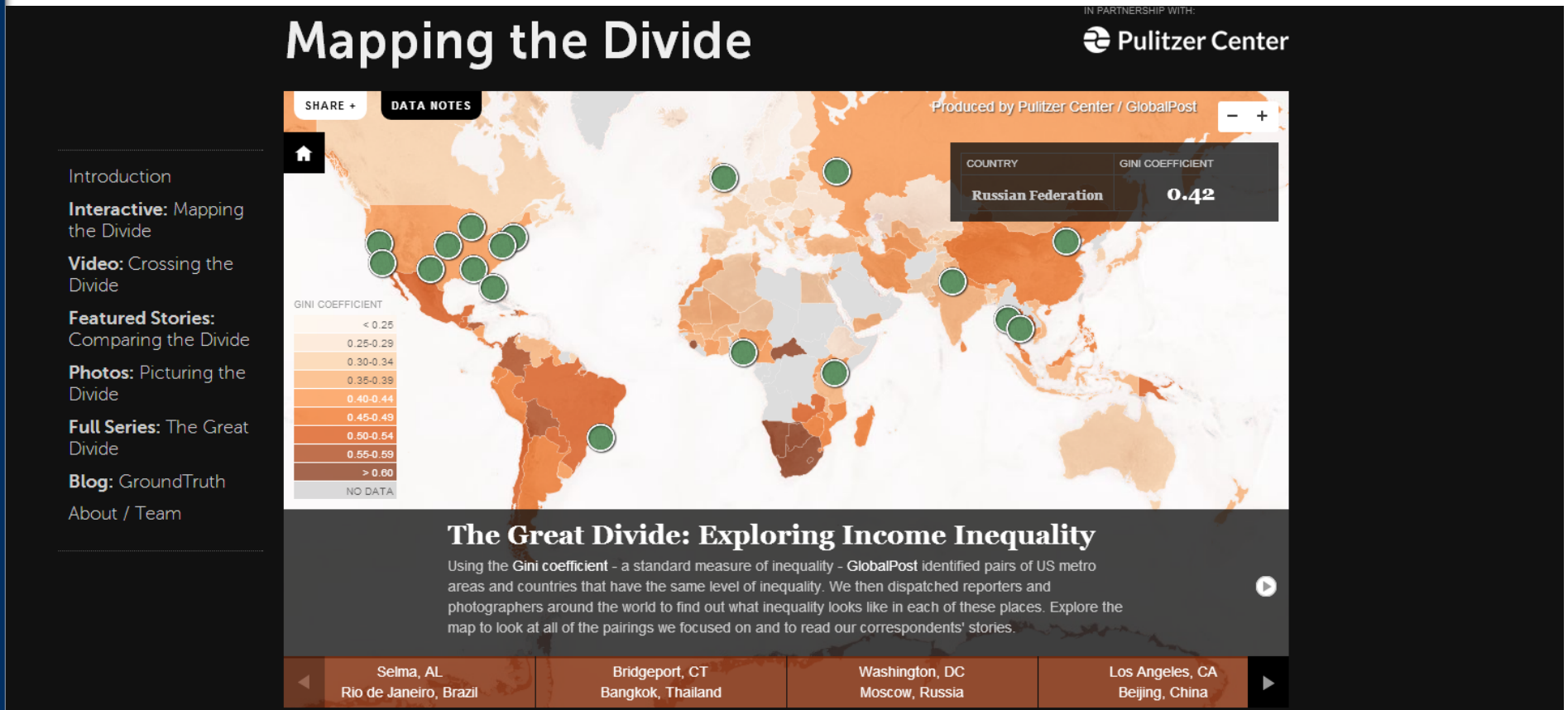
Internews
Local voices. Global change.

Data Sources

- Data Sources: recurring events
 - Governance & Security Indicators
 - Economic Indicators & Budgets
 - Education Indicators
 - **Human rights & Equality Indicators**
 - Environmental Indicators
 - Health Indicators



Data sources: Recurring Events



[The Great Divide Global Income Inequality](#): The Global Post maps annual data on global inequality by country and compares cities with similar levels of inequality.



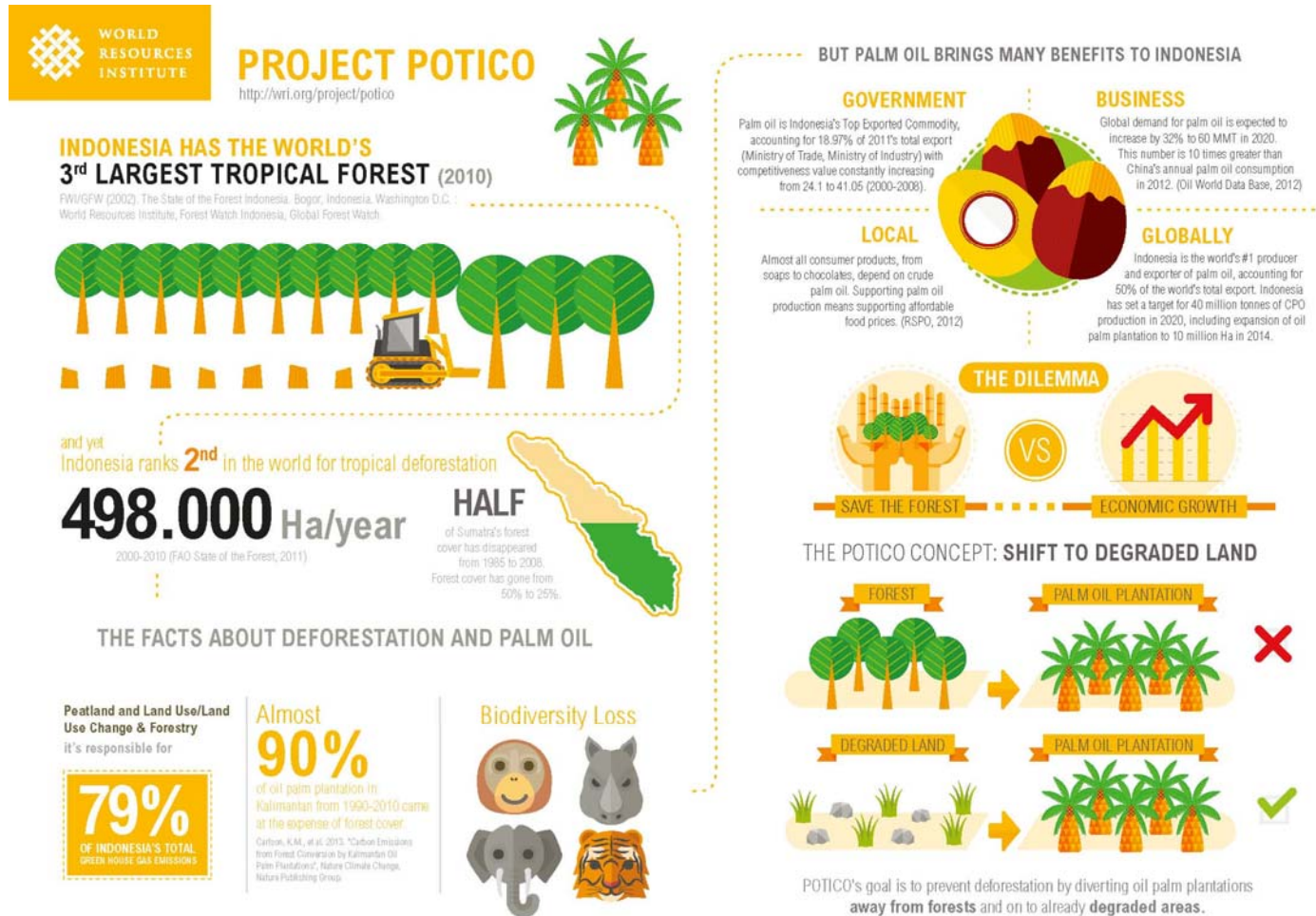
Internews
Local voices. Global change.

Data Sources

- Data Sources: recurring events
 - Governance & Security Indicators
 - Economic Indicators & Budgets
 - Education Indicators
 - Human rights & Equality Indicators
 - **Environmental Indicators**
 - Health Indicators



Data sources: Recurring Events



Project Potico: World Resource Institute explores deforestation in Indonesia



Internews
Local voices. Global change.

Data Sources

- Data Sources: recurring events
 - Governance & Security Indicators
 - Economic Indicators & Budgets
 - Education Indicators
 - Human rights & Equality Indicators
 - Environmental Indicators
 - **Health Indicators**

Sources of Data: Recurring Events

Child Injury: The Center for Disease Control in the United States created this infographic for journalists to draw attention to a major health risk for children

INJURY

The #1 killer of children in the US



In 2005, injuries that resulted in death, hospitalization or an ER visit cost nearly \$11.5 billion in medical expenses.

SOURCES: Web-based Injury Statistics Query and Reporting System (WISQARS), CDC, 2009.

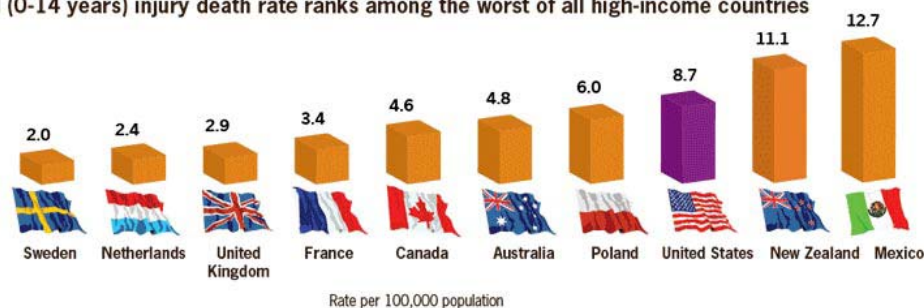
National Health Interview Survey, 2009 data release, CDC, National Center for Health Statistics.

Global view

The US child (0-14 years) injury death rate ranks among the worst of all high-income countries

The US child injury death rate is 4 times that of countries with the lowest rates.

SOURCE: Global Health Observatory Data Repository. Mortality and burden of disease: disease and injury country estimates, 2008, by sex and age.



Internews
Local voices. Global change.

Contents

- The Data Journalism Process
- Data sources: recurring events
- **Exercise**
- Data sources: sent or leaked data
- Data sources: breaking news
- Data sources: theories to be explored
- Exercise
- Summary & References

Exercise

How could these recurring reports be used in your reporting?

- Security, crime and violence data by year
- Unemployment data by age group
- Maternal and child mortality reports
- Other reports, press releases and data sent to your media?

Contents

- The Data Journalism Process
- Data sources: recurring events
- Exercise
- **Data sources: sent or leaked data**
- Data sources: breaking news
- Data sources: theories to be explored
- Exercise
- Summary & References



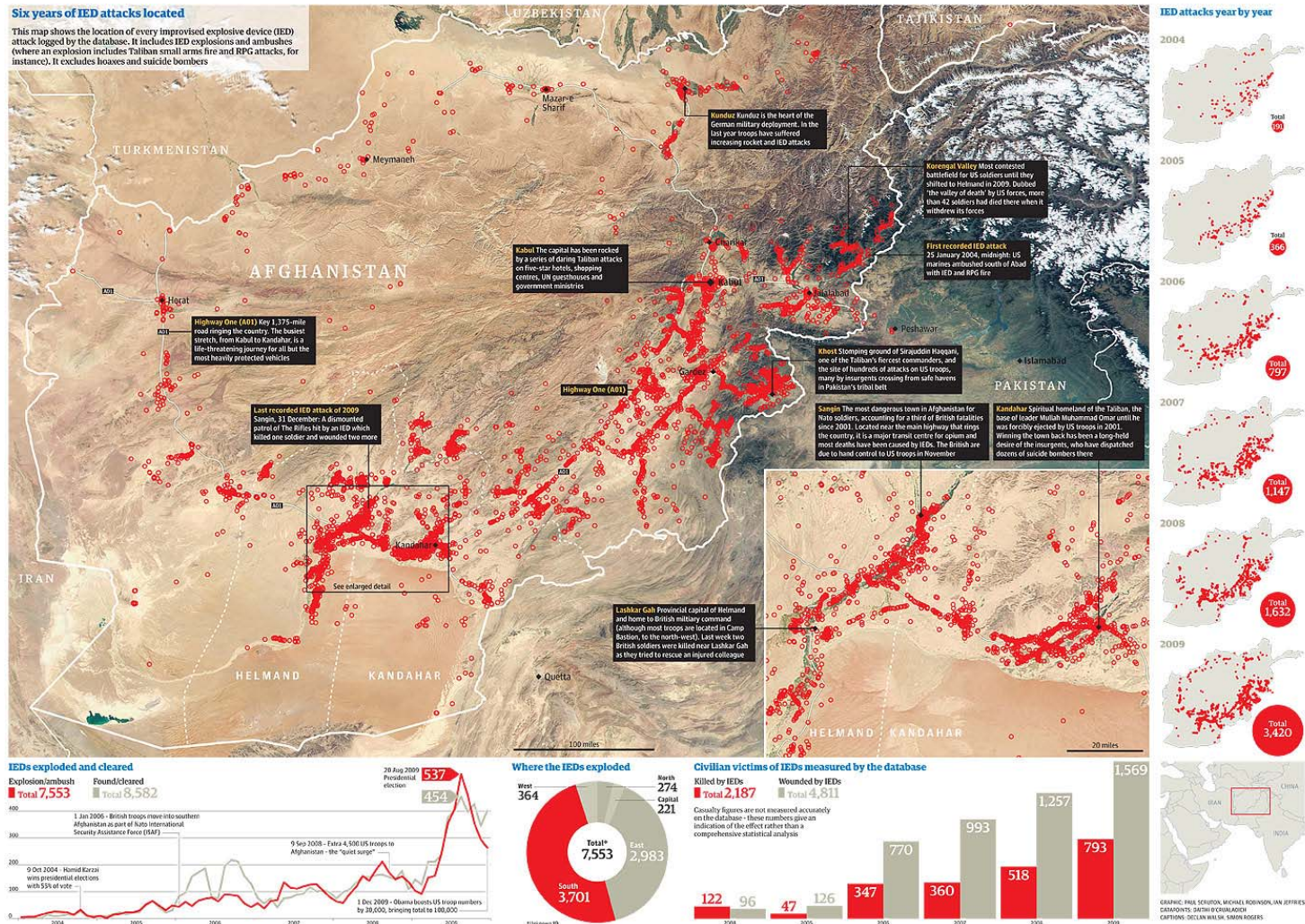
Data Sources: Sent or Leaked Data

Wikileaks Afghan War Diary

- The **Afghan War documents leak** is the disclosure of a collection of internal [U.S. military](#) logs of the [War in Afghanistan](#), also called the **Afghan War Diary**, which were published by [WikiLeaks](#) on 25 July 2010. The logs consist of 91,731 documents, covering the period between January 2004 and December 2009.
- The documents revealed information on the deaths of civilians, increased [Taliban](#) attacks, and involvement by [Pakistan](#) and [Iran](#) in the [insurgency](#).
- The New York Times, the Guardian and Der Spiegel received advanced copies of the logs.



Data Sources: Sent or Leaked Data



Afghanistan The War Logs: The Guardian's site explaining the human cost of war.

Internews
Local voices. Global change.

Data sources: sent or leaked data



Offshore Leaks: A leaked hard drive of 2.5 million documents about offshore bank accounts to the International Consortium of Investigative Journalists

Contents

- The Data Journalism Process
- Data sources: recurring events
- Exercise
- Data sources: sent or leaked data
- **Data sources: breaking news**
- Data sources: theories to be explored
- Exercise
- Summary & References



Data Sources: Breaking News

Data collection:

- Crowd-sourcing via Ushahidi-type platforms

- Citizen media contributions

- Sensor data for environmental catastrophes

Risks:

- Selection bias

- Verification process

- Context



Internews
Local voices. Global change.

Data Sources: Breaking News

Data collection:

- Crowd-sourcing via Ushahidi-type platforms

- Citizen media contributions

- Sensor data for environmental catastrophes

Risks:

- Selection bias

- Verification process

- Context



Internews
Local voices. Global change.

Data sources: breaking news

Westgate Nairobi

THE WESTGATE ATTACK
A SOCIAL MEDIA PERSPECTIVE

EVENT

GLOBAL CONTEXT

ATTACKERS

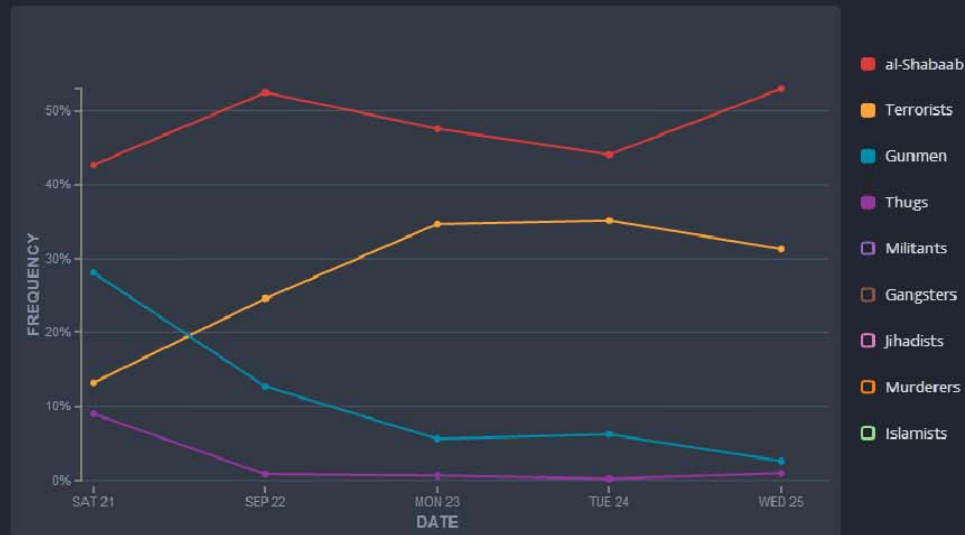
KEY ACTORS

WITNESSES

WORD FREQUENCY

This multiple line graph depicts the usage dynamics of the most common words used to describe the attackers by Twitter users in the online conversation.

Check the box for the words you're interested in to visualize their frequency throughout the four days of the attack. For easier comparison among words, check multiple boxes at the same time. Additionally, you can hover over a word to hide unrelated line segments.



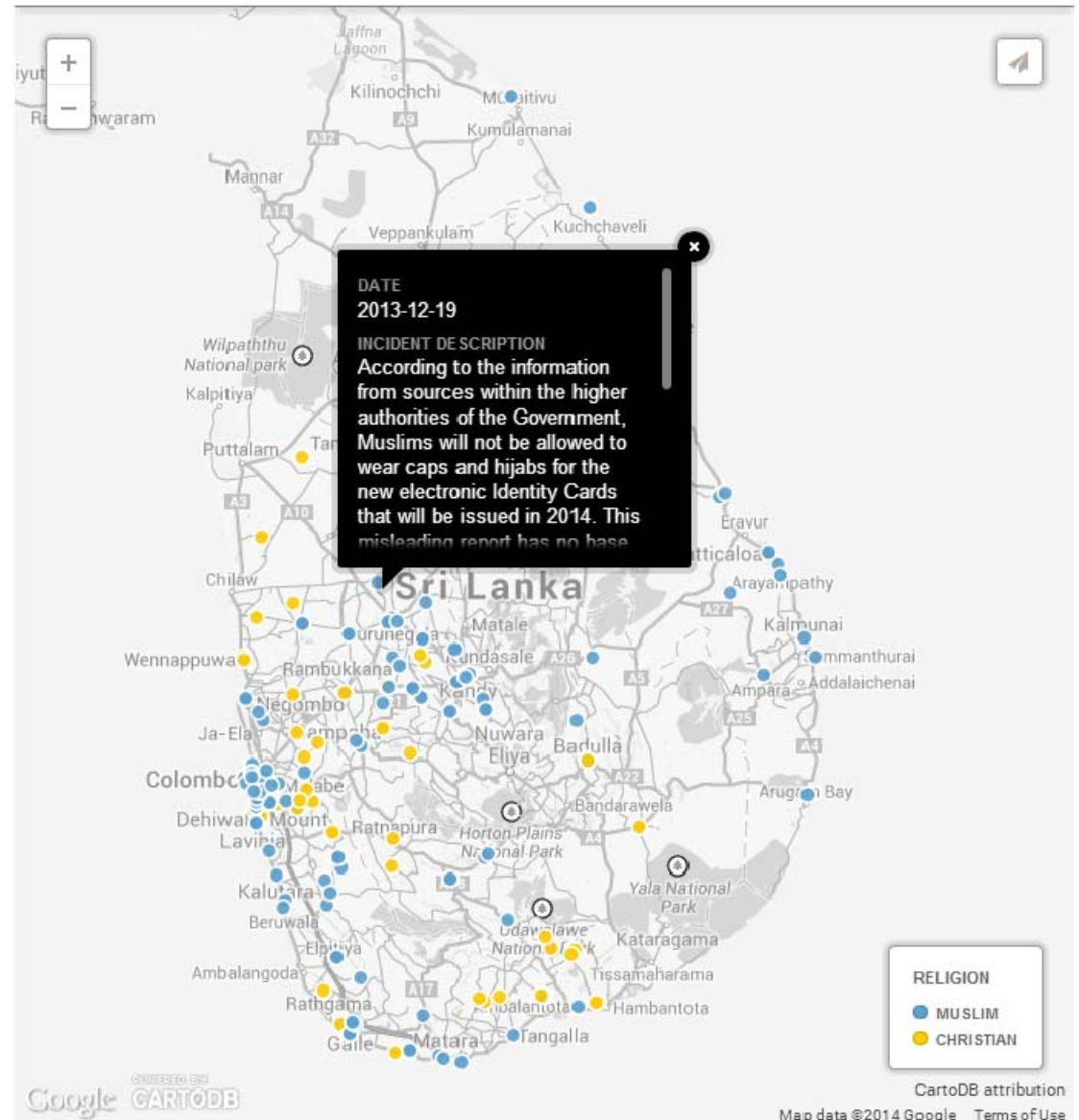
File: Internews-Investigate.html Tue May 20 2014 10:42:17 GMT-0100 (E. Africa Standard Time)

The Westgate Attack: A Social Media Perspective: An analysis of the events of Westgate driven by analysis of social media content.

Muslims Killed in Sri Lanka Mob Attack, Al Jazeera.

Attacks on religious minorities in 2013

Sources: The National Christian Evangelical Alliance and the Secretariat for Muslims



Create your own custom maps with [CartoDB](#)

Source: Al Jazeera and agencies

Data Sources: Theories to be Explored

Global and open data catalogs

[World Health Organization](#)

[United Nations](#)

[World Bank](#)

[DataCatalogs.org](#)

[The Guardian's world government data portal](#)

[Google's public data directory](#)

[The data hub](#)

[DBPedia Datasets](#)

[Factual](#)

[Free GIS data](#)

[List of open data resources](#)

[Energy data repositories](#)

[World Research Institute](#)

[Data wrangling](#)

[Quora thread: "Where can I find large datasets open to the public?"](#)

[Directory of APIs](#)

[Infochimps](#)

[Datamarket](#)

[Offshore Leaks](#)

[Investigative Dashboard](#)

[Open Corporates](#)

[Natural Earth data](#)

[UNEP Data](#)

[Transparency International](#)

[Corruption Index](#)

[Land Ownership Database](#)

[Gapminder World](#)



Internews
Local voices. Global change.

Data Sources: Theories to be explored

Sri Lankan Data Sources

Government

[Ministry of Education](#)

[Department of Census and Statistics](#)

[Sri Lankan Open Data](#)

[Central Bank of Sri Lanka](#)

[Sri Lanka Bureau of Foreign Employment](#)

[Sri Lanka Police](#)

Civil Society

[Center for Policy Alternatives](#)

[LIRNE Asia](#)



Internews
Local voices. Global change.

Exercise

Divide into two groups and read:

- [Sri Lanka solving China's unemployment!](#) *Daily FT*, June 19, 2013
- [Sri Lanka can be integral part of Asian transformation: ADB Chief](#), Lanka Business Online, June 19, 2014

Exercise

Answer

1. What is the journalist's hypothesis?

- What topics does the journalist cover in the story?
- What is the most important economic finding according to the story?
- What does he or she want the reader to understand after reading the story?

Exercise

Answer

2. What is the journalist's evidence for the story?

- What data does he or she use to prove his hypothesis?
- What experts does he or she quote?
- Does the author use information from the national budget or development indicators to put the data into context?

Exercise

Answer

3. What information and data is missing?

- What questions does the story not answer?
- What other data about the budget, education, economics, etc. would add to the story?
- What would your data angle for the story be?

Contents

- The Data Journalism Process
- Data sources: recurring events
- Exercise
- Data sources: sent or leaked data
- Data sources: breaking news
- Data sources: theories to be explored
- Exercise
- **Summary & References**

Summary & Resources

- Data sources include recurring events, sent or leaked data, breaking news and theories to be explored.
- Data from these stories can be used to enrich traditional reporting, find new angles on old stories, uncover corruption or engage citizens in the reporting process.
- Data sources are all around us: NGOs, research institutions, government and ordinary people.

Summary & Resources

