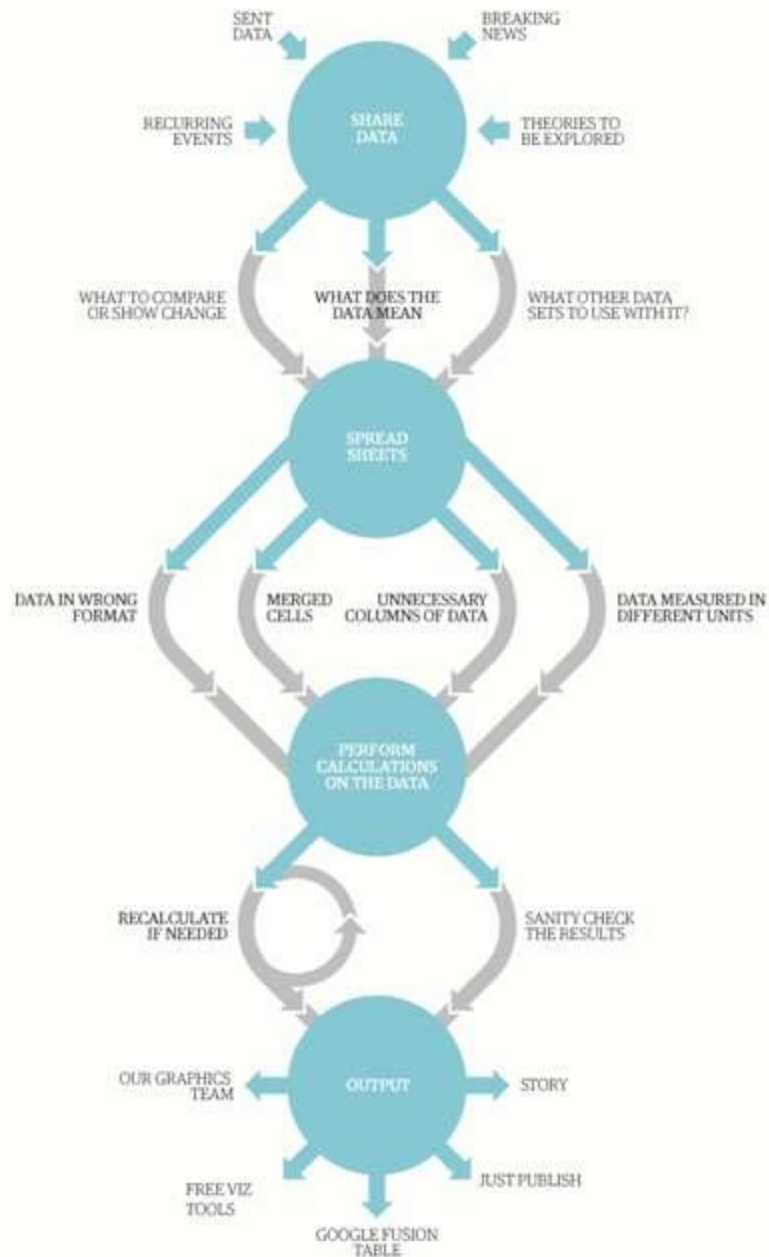
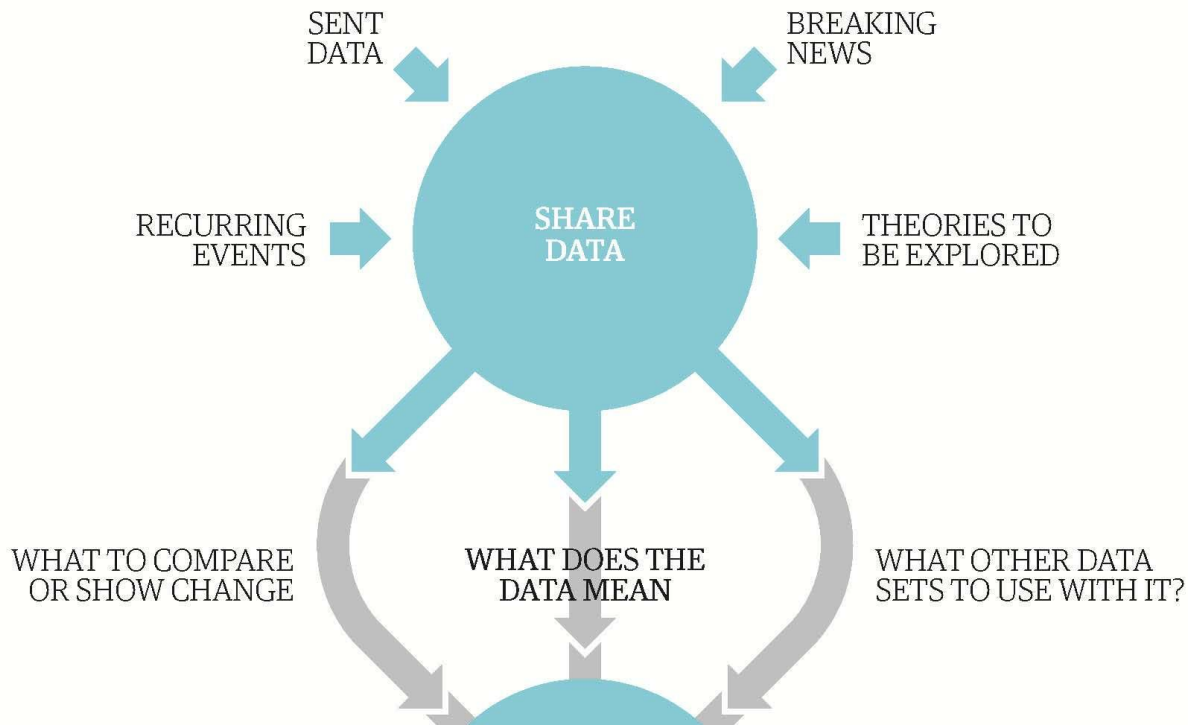


# Finding Stories in Data

Where to Begin?



# Sources of data story ideas



# Recurring Events: data releases

- Inequality studies: [Rich-Poor Gap Widens in Urban & Rural India](#)
- Exam Scores: [MySchool.KE](#)
- Government budget reports: [Twaweza Kenyan Budget Explorer](#)
- Crime statistics: Oman's first release of crime data: [Crime in Oman Visualized](#)
- Argentina Census: [Census Project Argentina](#)

# Brainstorm: What is the news hook?

- Health surveys?
- Crime statistics?
- School performance?
- Demographic information?
- Government budgetary reports?
- Agricultural research?
- What other kinds of press releases to you receive related to data? What is the news angle for digging into the data?

# Sent data ... sometimes you get lucky!

- [Offshore Leaks](#): A leaked hard drive of 2.5 million documents about offshore bank accounts to the International Consortium of Investigative Journalists
- [Wikileaks](#): The release of State Department cables in 2010 to a variety of international media outlets.
- Both databases are now public: open data push extends to leaked data

# Breaking News

## Data collection:

- Crowd-sourcing via Ushahidi-type platforms
- Citizen media contributions
- Mining data from social media

## Risks:

- Selection bias
- Verification process
- Context

# Breaking news: data enriches story

- Mother Jones: [What you Need to Know about the Horn of Africa Famine](#)
- Infoamazonia: [Hydroelectrics of Tapajos in a map, numbers and graphs](#)
- Opportunity: [Kenya Health Policy Project](#)



# Exercise: What Data is Missing?

## Education Sector Poses the Toughest Test

- List the data that is included in the story
- List possible comparisons between the figures cited
- List contextual data that is missing
- List questions that could be answered with complete data

# Theories to be explored: databases

[World Bank Data](#)

[United Nations Data](#)

[Africa Open Data](#)

[World Health Organization Data](#)

[Population Reference Bureau Data](#)

[Open Data Catalogue](#)

[International Development Data](#)

[Transparency International Corruption Index](#)

[Land Ownership Database](#)

[Gapminder World](#)

# Finding Data on the Web

# Data sources online

- Advanced search techniques to find data in the deep web
- Scraping databases in html format using tools like Scraperwiki
- Converting documents like PDFs into a more useable format
- Pulling information from Advanced Passenger Information Systems (APIs), which automatically updates a database with new data pulled from the web

# Advanced Google Searches

- Type OR to find variations of the same search: Kenya “cholera epidemic” OR “cholera outbreak” OR “cholera spread”
- Use quotes to search for phrases
- Narrow search to specific website: “site:url”  
Example “site:[www.nation.co.ke](http://www.nation.co.ke) education”
- Narrow search to filetype: “Filetype: [extension]” Example: “site:http://www.doh.gov.za/ filetype:pdf maternal”
- [Google alerts](#): get notified about your beat

# Data formats for download

- Portable Document Format (PDF): charts that contain data but are saved in a unified document with text
- Excel file (xls): data is saved as a table readable by Microsoft Excel
- Comma separated values: Plain text file with each data separated by a comma

# Exercise: Finding Education Data

- Returning to [Education Sector Poses the Toughest Test](#) let's see what data we can find online using advanced search techniques
- Remember the list of potential data sources as a place to start
- Try to find datasets that can be compared to produce new insight that proves or disproves your hypothesis. Try to find: the number of primary school teachers and students in 2010, 2011 and 2012.