



INTERNEWS PAKISTAN DATA JOURNALISM TRAINING MANUAL

Thinking Like a Data Journalist

Disclaimer

This is an adaptation by the author of an original work produced for The World Bank's Sudan Evidence Base Programme found at <https://www.sudandata.org/learning/2>.

Eva Constantaras
evaconstantaras@gmail.com

Module 1: Introduction to Data Journalism

STUDENT WORKBOOK

This module introduces data and how it is used to create public service stories by journalists. To begin with, this module explains what open data is, and how it impacts policy. Subsequently, journalists will be introduced to several strategies for identifying story angles and then practice those skills through the analysis of a number of case studies that will demonstrate how data from various sectors have been used to create data-driven stories. Finally, we will look at investigative journalism and its relationship to data journalism.

At the end of this module, you will be able to:

- Define data and describe what a dataset is
- List the steps involved in generating public interest stories from data
- Identify ways to approach a data set for story ideas
- Identify opportunities to integrate sectoral data into a policy story

Contents

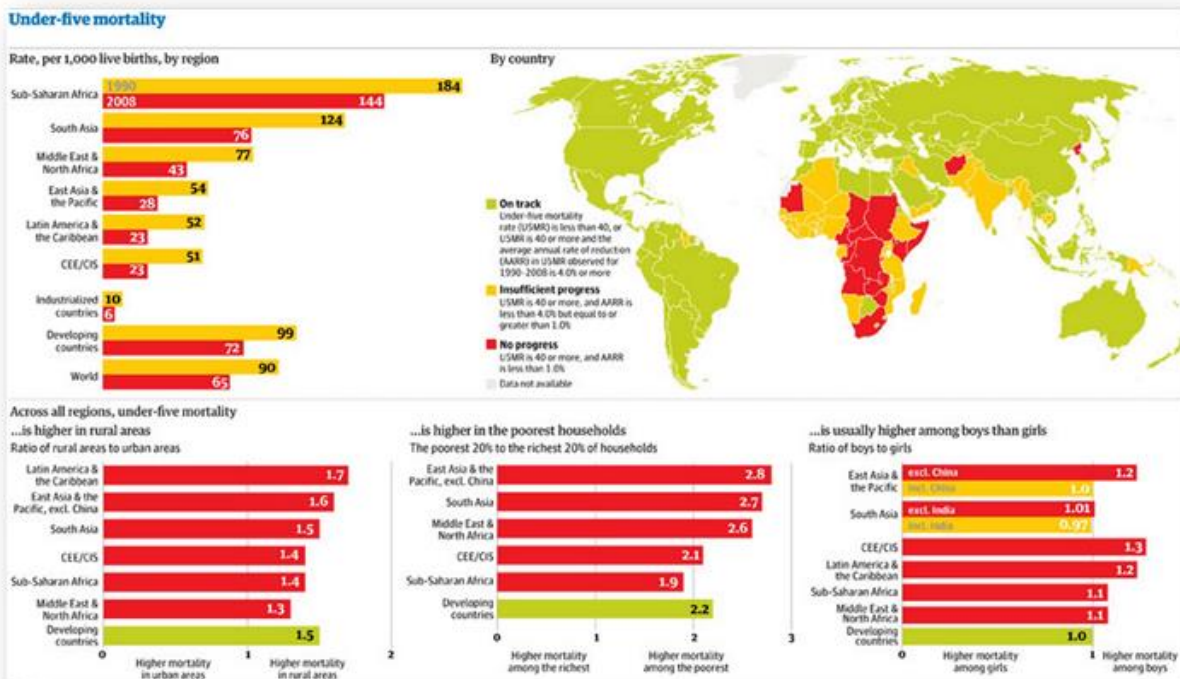
Lesson 1: How Data is Used for Public Interest Stories	3
Getting Started	3
Why is Data Journalism Important?	4
Source:	4
What is Data?	5
What is Open Data?	6
The Transparency Cycle	7
Data Journalism Process	8

Thinking Like a Data Journalist

Lesson 2: How to Find a Story in Data	9
Data lets us test hypotheses	9
Data shows us trends and contrasts	11
Data reveals startling outliers	13
Data shows us invisible connections	16
Lesson 3: Sector-Specific Data Stories	18
Sector-Specific Example: Governance and Security Indicators	19
Sector-Specific Example: Economic Indicators and Budgets.....	22
Sector-Specific Example: Education Indicators	24
Sector-Specific Example: Health Indicators.....	26
Sector-Specific Example: Human Rights and Equality Indicators	30
Sector-Specific Example: Environment and Agriculture	32
Exercise.....	34
Lesson 4: Introduction to Investigative Journalism.....	35
Objectives of Investigative Journalism.....	35
Comparison of Conventional and Investigative Journalism	36

Lesson 1: How Data is Used for Public Interest Stories

Getting Started

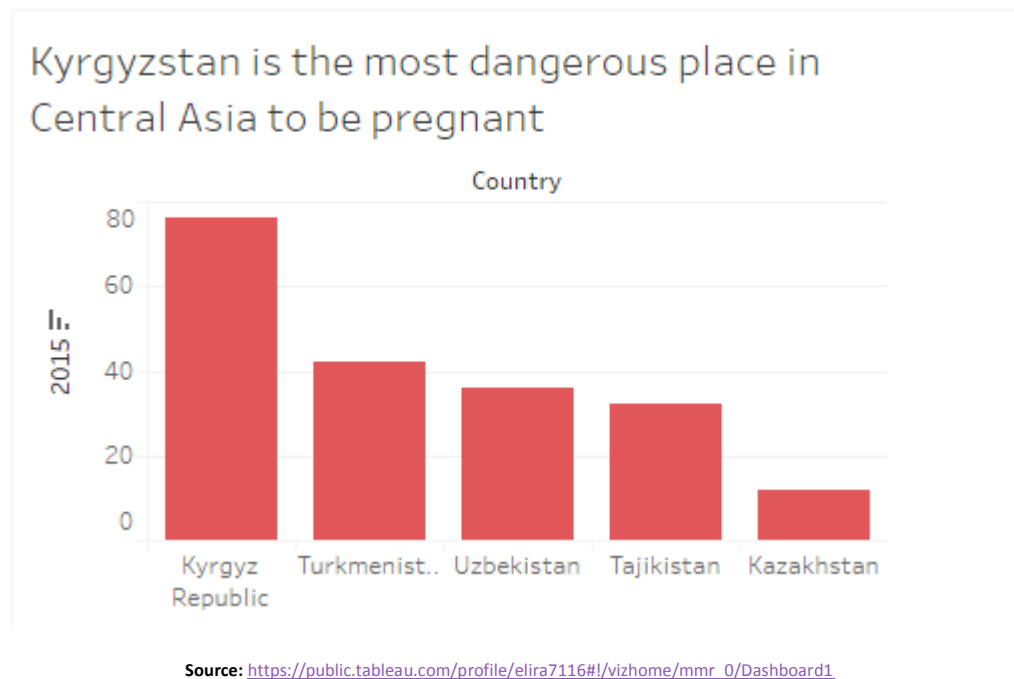


One of the Millennium Development Goals that countries across the world set out to achieve by 2015 was to reduce the under-five mortality rate by two-thirds between 1990 and 2015. Has the world achieved this goal? What about the countries in your region?

Data helps us answer this question. Not only that, it helps uncover insights that may help fill the gaps that remain in addressing this goal. Data journalists have used this data to tell engaging stories to help citizens and policymakers to make better decisions for the future. Here is an example from The Guardian: <http://www.theguardian.com/global-development/datablog/2010/sep/17/development-data-aid>

In this module, you will learn about how data from is used by journalists and civil society organizations to create public interest stories around a number of important public interest issues.

Why is Data Journalism Important?



Imagine you are at a press conference at the Ministry of Health and the press officers inform the room that the government's campaign to reduce maternal mortality has been successful.

A journalist carrying out her normal duties would copy down the information from the Ministry, call a civil society organization who works on maternal and child health issues for a quote and run the story. If the Ministry's information and the CSO's information are contradictory (say, the CSO says that maternal mortality has not been decreasing as quickly as it was supposed to), she would simply run both versions in the interest of balanced reporting.

If the journalist can access the data about maternal mortality rates in the country, she can determine whose version is true, the Ministry's or the CSOs and publish her findings along with a reference to the data she used to establish the facts.

This is one simple example of how data can enrich a journalist's reporting and make her stories more valuable to her readers. In the process of accessing and analyzing a data set, she may discover even more stories, like the top cause of death among mothers and children, and enable her to write more stories on the topic.

Data journalism enables journalists to go deeper into issues to find out the root causes for why people are getting sick or well, getting an education or not, finding jobs or not and put this information into context for citizens.

With better information about the root problems facing society, the audience can make better, more informed decisions for their families, communities, and government.

Thinking Like a Data Journalist

What is Data?



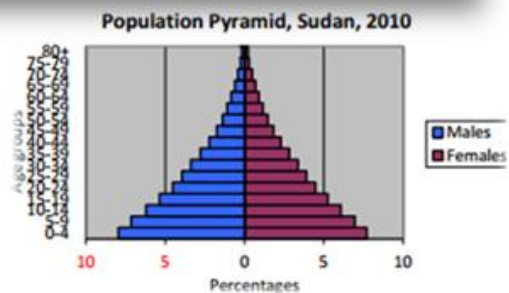
Data collection

Table 10: Population Distribution by Broad Age Groups in Sudan, 1980-2050

Year	Population by Age Group (%)			
	0-4	5-14	15-64	65+
1980	18.9	28.0	50.1	2.9
1985	18.2	28.3	50.6	2.9
1990	17.5	27.9	51.6	2.9
1995	17.4	27.1	52.6	3.0
2000	17.1	26.7	53.3	3.0
2005	16.6	26.6	53.8	3.0
2010	15.7	26.4	54.8	3.1

Dataset

Visualization

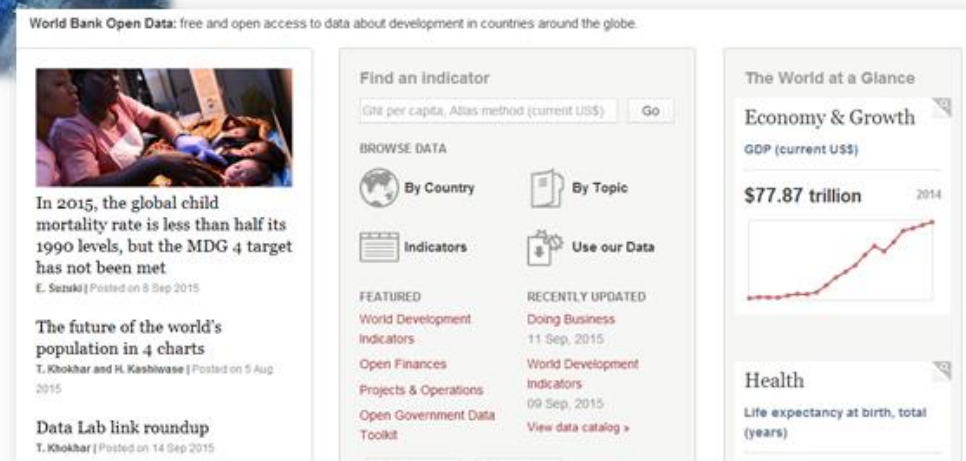


Data is a collection of distinct pieces of information, which are recorded and structured in a manner that makes it easy for analysis. Data is usually generated when something is measured or recorded. When many such pieces of information are recorded consistently - they can be easily analyzed individually or in groups. Data can be seen as a raw material that can be refined to produce meaningful information.

For instance, data is generated by surveys (such as census data), when something is voted on (such as elections results data), when something is registered (such as a birth records data), when something is purchased (such as sales records for an online store). Data is also generated by mobile devices, sensors, the internet, and satellites (such as GPS data) - and many other technologies.

In our daily lives, data is commonly found organized in tables. Contents of a single distinct table can be referred as a dataset. The analysis of a dataset can generate new knowledge and visual representations - which are valuable to make arguments and take decisions.

What is Open Data?



“Open data and content can be freely used, modified, and shared by anyone for any purpose.”¹

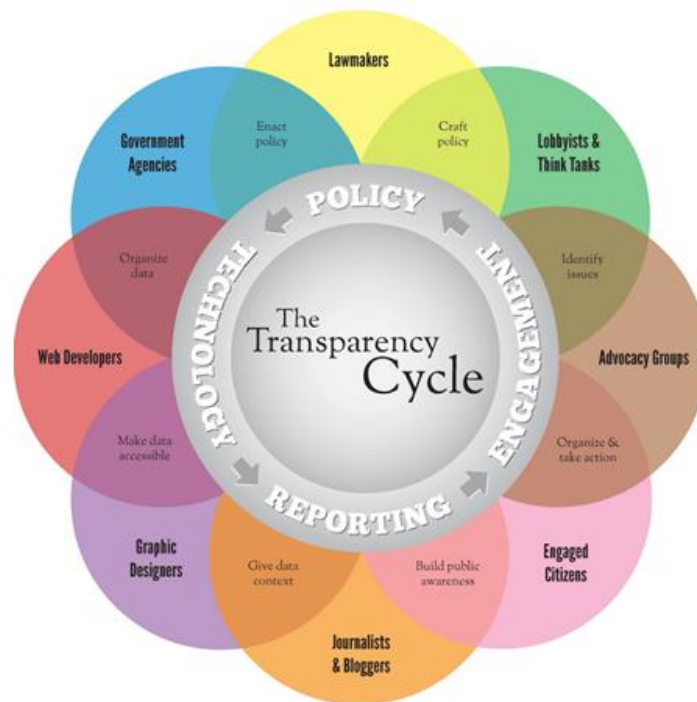
With a global push for open data many governments, international organizations are creating their own open data portals with the data they have aggregated and opened. These portals are a source of rich information for civic engagement and generating public interest stories. Although, these portals are by no means the only sources of data since much data lives on various government ministry websites.

Some important datasets that are (or could be) open come from people’s personal data that is aggregated and anonymized. Much statistical information ultimately comes from surveys of individuals, but the end results are heavily aggregated so that individuals can’t be identified. The open data community should remain cautious about the need to protect privacy and individual well-being in the data’s march towards public good and transparency.²

¹ <http://opendefinition.org/>

² <http://techpresident.com/news/wegov/24895/what-does-privacy-have-do-open-government>

The Transparency Cycle



To understand how open data can create civic impact - consider this "Cycle of Transparency"³ which suggests how variety of stakeholders can work together to make data from city, state, national, and even international governments more transparent.

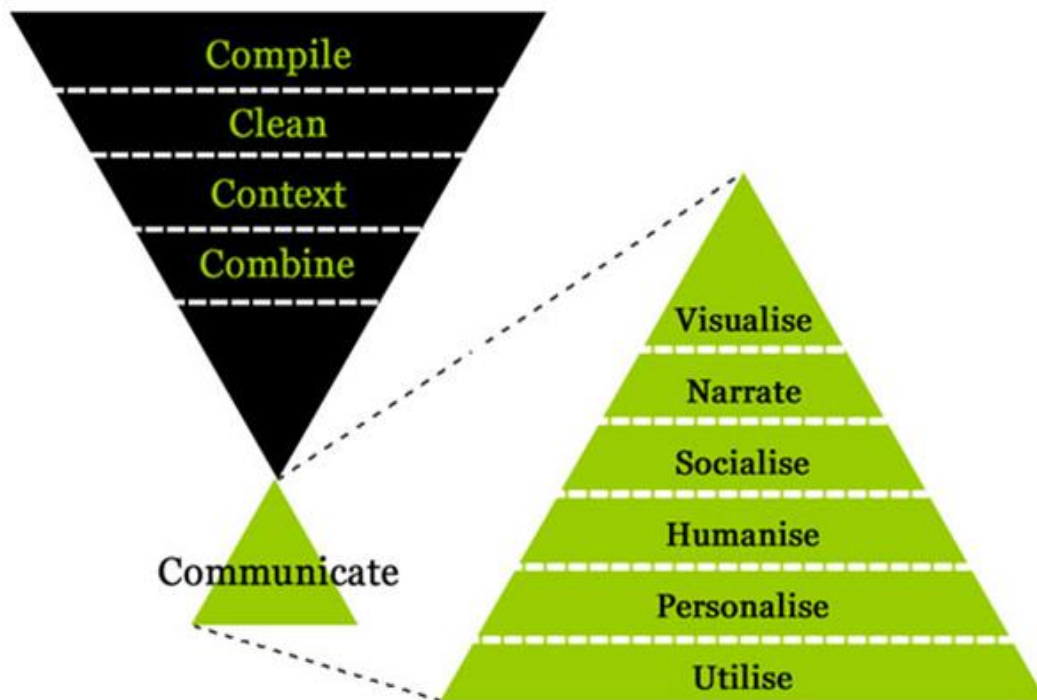
Lawmakers, lobbyists and think tanks (as well as citizens) all play a role in articulating new transparency policies and pushing them through. Once data is released, government agencies and web developers can build the necessary technology to organize the data and make it usable. But creating a dashboard is not enough. Data is useless unless it is analyzed and transformed into information.

With data being made easily accessible, journalists and bloggers can begin to dig into it, mix it up, identify relevant information and give the data context. As that critical context is provided, citizens absorb it and spread the information to others - both online and face-to-face - and make the data actionable.

Ultimately, informed citizen action creates greater public awareness; citizens become more effective, responsible advocates; holding government accountable.

³ <https://sunlightfoundation.com/blog/2010/03/10/introducing-the-cycle-of-transparency/>

Data Journalism Process



From child and maternal mortality to budgets, a number of issues of public interest can be understood and explained with data in a manner that's engaging and incisive. So how does dense data in tables, say in an Excel sheet or a report, turn into engaging narratives and visuals? This involves a process that is being evolved across newsrooms in various countries - including South Asia. CSOs and NGOs have also applied similar methods to work with data. Here are the steps a typical process⁴ may involve:

- **Compile:** Data journalism begins in one of two ways: either you have a question that needs data, or a dataset that needs questioning. Whichever it is, the compilation of data is what defines it as an act of data journalism.
- **Clean:** Having data is just the beginning. Being confident in the stories hidden within it means being able to trust the quality of the data – and that means cleaning it. Cleaning typically takes two forms: removing human error; and converting the data into a format that is consistent with other data you are using.
- **Context:** Like any source, data cannot always be trusted. It comes with its own histories, biases, and objectives. So like any source, you need to ask questions of it: who gathered it, when, and for what purpose? How was it gathered? Who can explain the data?

⁴ <http://onlinejournalismblog.com/2011/07/07/the-inverted-pyramid-of-data-journalism/>

- **Combine:** Good stories can be found in a single dataset, but often you will need to combine two together. After all, given the choice between a single-source story and a multiple-source one, which would you prefer?
- **Communicate:** In data journalism the all-too-obvious thing to do at this point is to visualize the results – on a map, in a chart, an infographic, or an animation. But there’s a lot more here to consider – from the classic narrative, to news apps, case studies and personalization.

Lesson 2: How to Find a Story in Data

When first looking at a dataset, it can be difficult to know where to look for a good story idea. There are a couple of common ways into a data story. These are four common ways to find the story you are looking for.

- Data lets us test hypotheses
- Data shows us trends and contrasts
- Data reveals startling outliers
- Data shows us invisible connections

Data lets us test hypotheses

The most common way into a data story is to develop a hypothesis and then use data to prove or disprove that hypothesis. A hypothesis is a theory you have due to your experience as a journalist. It uses your “news nose” to sniff out something you find suspicious or you expect is being explained incorrectly.

In this example, a journalist at IndiaSpend who wrote [Poverty Does Not = Crime](#) started with the following hypothesis, “Poverty is not the primary cause for crime rates in Indian cities.”

The journalist looked at data around changes in poverty level and changes in crime rates and discovered that poor communities did not have higher crime rates than rich communities. Also, neighborhoods that became poorer did not experience a corresponding increase in crime and communities that became richer did not necessarily experience a drop in crime. If he had found that there *was* indeed an increase in crime in poor neighborhoods, his hypothesis would be proven false but he would still have an interesting story.

The next think the journalist can develop new hypotheses about what might cause increased crime, things like police presence, public street lighting, presence of organized crime groups, etc.

Wednesday, February 19, 2014 | Stay Connected

Search in site...

[HOME](#)
[ABOUT](#)
[INVESTIGATIONS](#)
[FACT CHECK](#)
[SECTORS](#)
[STATES](#)
[DATA BLOG](#)
[VIDEOS](#)
[VIZNOMICS](#)

Home » Latest Reports » Poverty Does Not = Crime

Poverty Does Not = Crime

Pranav Garimella & Prachi Salve, February 14, 2014

Follow @GarimellaPranav Follow @PrachiSalve11

Share 1 2 11

Why do people commit crime? And what are rising crimes linked to? Inequality, disparity, need, feeling of lawlessness or desperation that arises from a combination of some of the above? We continue with our Dostoevskian (sounding) journey.

In an [earlier report](#), we found some correlation between economic growth and crimes involving property. Now we look at other economic indicators like unemployment and poverty and explore a possible (and quick) correlation between them and rising crime (against property) across the country.

We will try and answer the following questions: First, do rising and falling unemployment rates affect crime? Second, do poverty numbers translate into more crime, state-wise?

The answer, actually, is not simple. Possibly there is little correlation. The inverse might be more

Most Read **Most Popular** **Most Mailed**

- India Demonetizes Currency After 36 Years
- Delhi 'Power' Story: Can Tariffs Be Actually Reduced?
- The Frightening Gap Between Private & Government Run Schools
- Why Congress & BJP Need Allies To Form A Govt
- Guess Which Indian State Has The Largest Appetite?
- Does Mumbai Have A Solution To India's Public Transport Problem?
- Forget Delhi, More Shocking Power Stories In Other States

Recent Posts

- India, Philippines Among The World's Five Most Dangerous Countries For Journalists
- How Rupee Comes To Government & How It Is Spent
- Budget 2014: Costing India's Flagship

[Putin's Friend Profits in Purge of Schoolbooks](#) is an investigation into who benefited from the abrupt change in schoolbook standards in Russia, in which most public schools were forced to get entirely new books. The journalist from *the New York Times* began with a simple hypothesis: "The reason behind the nationwide change in schoolbooks was political and economic, not educational."

The journalist discovered that a close friend who owned a publishing business benefited hugely from the canceling of the contract in schoolbooks due to new content standards that much of the educational community objected to. The new publisher was the only company whose books met those standards. When he started the story, the journalist did not know exactly who had benefited from the new policy, but financial data about who won the new contract pointed him to the reason why the schoolbooks were changed.

EUROPE | Putin's Friend Profits in Purge of Schoolbooks

Putin's Friend Profits in Purge of Schoolbooks

By JO BECKER and STEVEN LEE MYERS NOV. 1, 2014



President Vladimir V. Putin of Russia, left, and Arkady R. Rotenberg, the chairman of a publishing house that has benefited from government maneuvers. Mr. Rotenberg, a judo sparring partner from Mr. Putin's youth, has gone from a modest trader to a billionaire during Mr. Putin's tenure. Sasha Mordovets/Getty Images

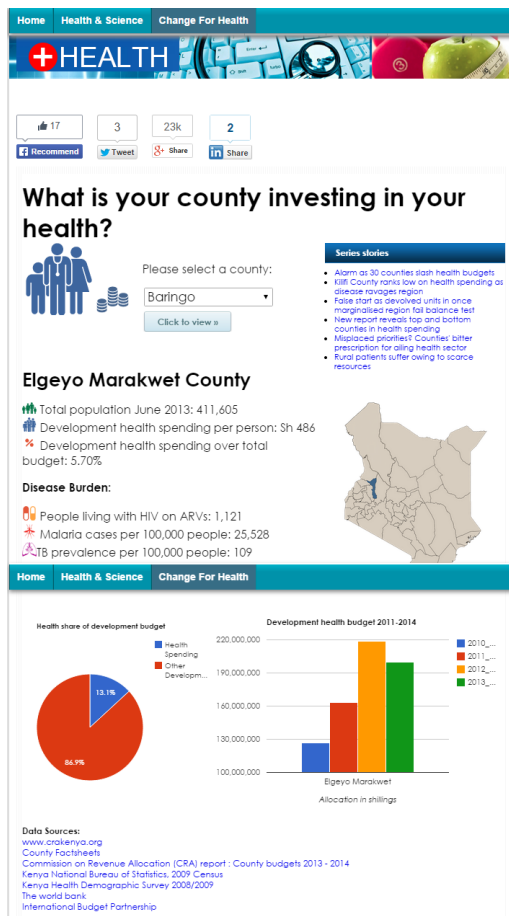
[Email](#)[Share](#)

MOSCOW — The purge began in late winter. One by one, hundreds of textbooks that Russian schoolchildren had relied upon for

Activity: Can you think of a hypothesis about public service delivery that you could prove or disprove with data?

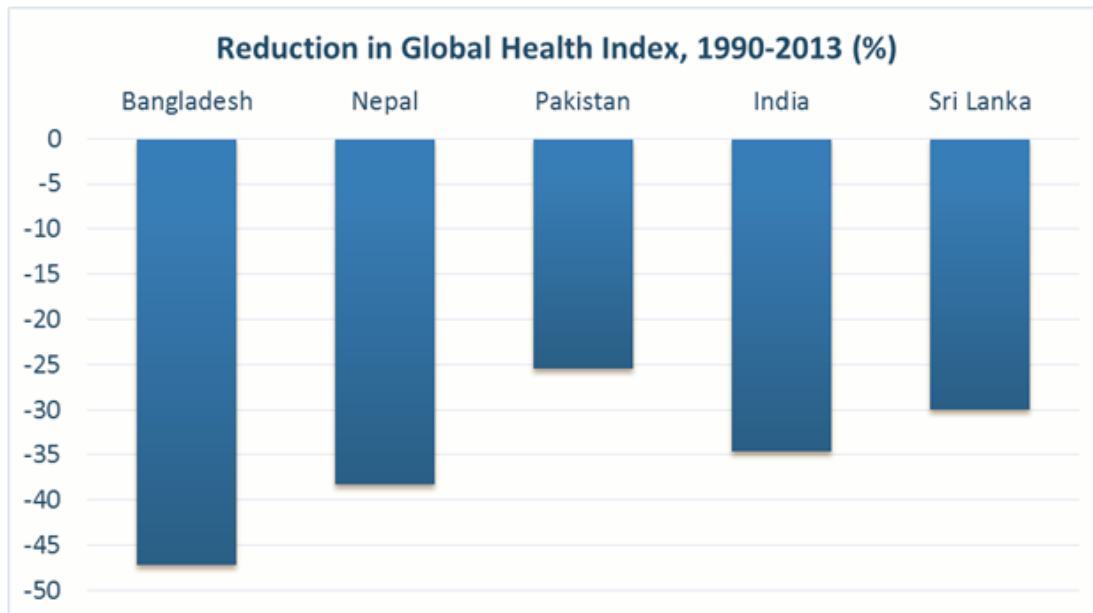
Data shows us trends and contrasts

One of the biggest challenges when looking at a data set is deciding what is “normal.” What is an expected number of children dying under the age of five? Or the price of a bag of rice? Or the amount the Ministry of Education spends on each child? One of the ways to answer this question and find a story is through comparisons. By ranking countries, supermarkets, schools or provinces, we can compare and contrast those with high numbers with those with low numbers.



In The Standard of Kenya's [Change for Health](#) project, the journalists investigated how much each county was spending per person for healthcare. He found that some states were spending very little per person despite having high rates of disease in the county. He found that more than half of states had reduced spending on health compared to the year before. By doing a comparison, he forced county leaders to explain why they weren't doing more to keep their citizens healthy.

In Pakistan's [Dawn](#), story [Bangladesh: No longer the 'hungry' man of South Asia](#) Murtaza Haider demonstrates that Pakistan made the least progress in the region to reducing hunger in the country from 1990 to 2013. Again, we only know what is good progress or bad progress based on a comparison with other countries.

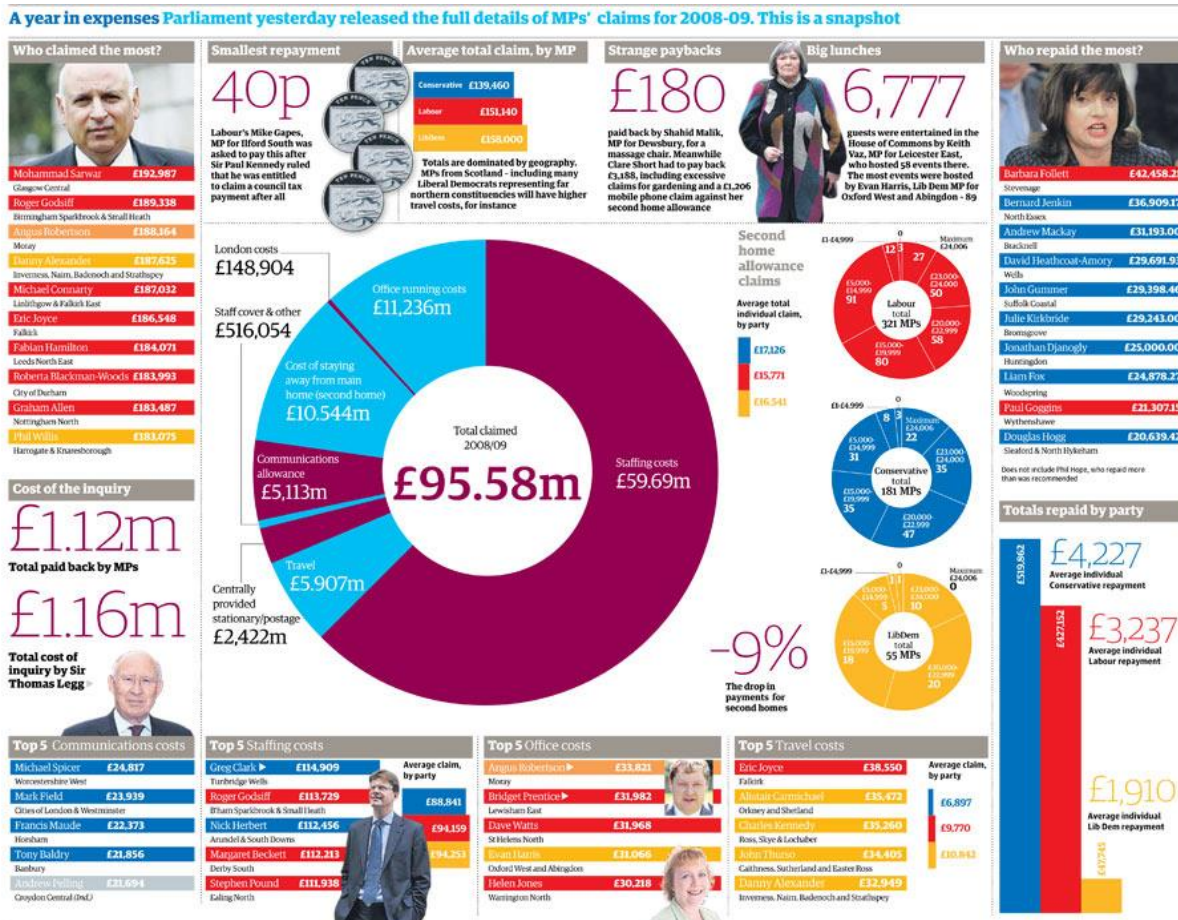


Pakistan shows the least progress in reducing hunger amongst other South Asian countries. Source: Calculations and graph by Murtaza Haider, 2013. Data obtained from GHI, 2013.

Activity: What surprises do you think you would discover by comparing public service delivery in different provinces or administrative divisions?

Data reveals startling outliers

Sometimes, especially when investigating corruption, the way into a story is to find a number in your dataset that is surprisingly low or surprisingly high, an outlier, in other words. An outlier can indicate, for example, that a government agency is overpaying for a specific item and an official is getting a kick-back and pocketing the rest of the money. It can indicate that someone found a loophole in a system and is exploiting it.



In, [The Guardian MP Expenses](#) the journalist demonstrated that some Members of Parliament were charging an outrageous amount for expenses while on the job, for example, on phone bills or on renovating their offices. The journalist investigated those who had unusually high expenditures, published that information and many Members of Parliament were forced to return the money and some were not reelected.

In, Romania, [EurActiv Romania](#) used data journalism skills to uncover corruption in awarding of government contracts through the public procurement portal. They didn't know what a "normal" procurement looked like, but what they looked for in the data were outliers.

Journalism project uncovers public procurement fraud in Romania

Home | Central Europe | News

By EurActiv.ro

Nov 10, 2014 (updated: Feb 11, 2016)

1987 de zile de la trimiterea în judecată
Dosarul Buldotancul - Mugurel Surupaceanu - prejudiciu 200.000 euro
 TERMEN DE JUDECATĂ | 10 DECEMBRIE 2014

In anul 2007 Complexul Energetic Turceni a cumparat de la o companie controlata de fostul deputat PSD, Mugurel Surupaceanu un tractor forestier vechi la un pret supraevaluat de patru ori. Procurorii DNA au aratat ca Mugurel Surupaceanu impreuna cu directorul Complexului Energetic Turceni, Dumitru Cristea au trucat procedura de achizitie a utilajului.

[citește tot articolul](#)

1929 de zile de la trimiterea în judecată
Dosarul "Mită la Electrica Nord": 1 milion de euro pentru întrerupătoare contrafăcute
 TERMEN DE JUDECATĂ | 4 DECEMBRIE 2014

Doi directori ai ELECTRICA TRANSILVANIA NORD au fost trimisi în judecată de DNA Cluj pentru că în

Proiect finantat de Comisia Europeana

COMBATAREA CRIMINALITĂȚII ÎN DOMENIUL ACIȚIILOR PUBLICE

Platforma inițiată de Freedom House și Asociația Actorii Europeeni, cu sprijinul HotNews.ro în cadrul programului „Combaterea fraudei în achiziții publice”, derulat cu sprijinul financiar al Comisiei Europene prin „Programul de prevenire și luptă contra criminalității” – Direcția Generală Afaceri Interne.

Întreaga răspundere asupra corectitudinii și coerenței informațiilor prezentate revine inițiatorilor.

Această secțiune nu reprezintă poziția oficială a Comisiei Europene sau a vreunui dintre partenerii implicați în derularea proiectului. Foto

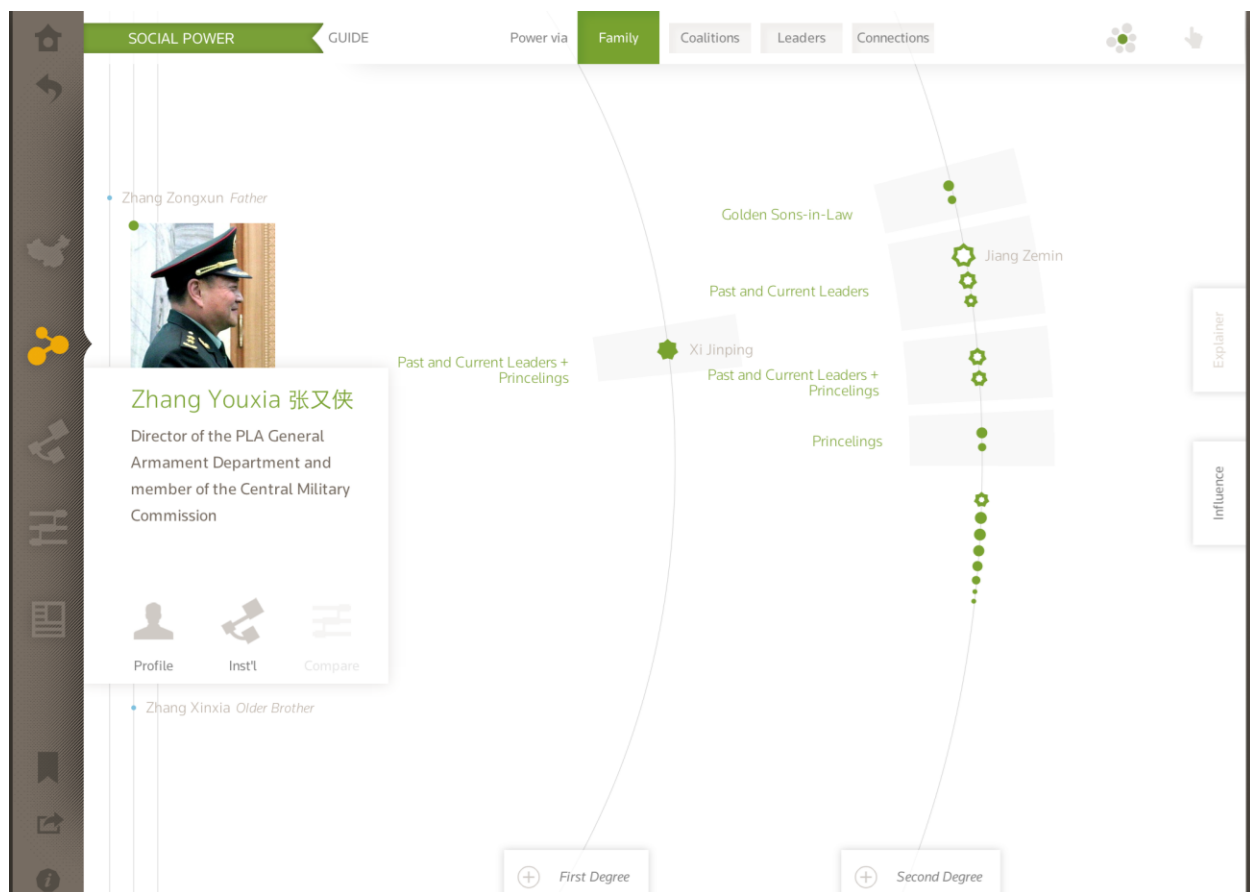
Source: <https://www.euractiv.com/section/central-europe/news/journalism-project-uncovers-public-procurement-fraud-in-romania/>

- A ministry that awarded an unusually high number of contracts not through public bidding by claiming exceptions.
- A company that received many small contracts for the same project to avoid going through procurement procedures.
- Projects that received unusually small numbers of bids and were not well advertised.

Activity: What kind of data is available in Pakistan that you could dig into to look for outliers, even if you are not an expert in that specific sector?

Data shows us invisible connections

By crossing databases, we can often uncover invisible power relationships that govern illicit political, economic and social relationships. [Connected China](#): tracks and visualizes the people, institutions and relationships that form China's elite power structure using data from Reuter's network of sources. Many civil society organizations are beginning to investigate relationships between politicians and the private sector and revealing corruption in awarding of mining contracts and other government procurements. By working with these groups, journalists can expose these relationships and enable citizens to demand change.



In [this example](#) from El Universal in Mexico, journalists map out all the legitimate business owned by the head of a major drug cartel. The visualization points to all the legitimate businesses that might have been used to launder money and provide a cover for drug trafficking activities.



Activity: What political, business or family networks do you think influence the functioning of government and the delivery of public services in Pakistan? Where could you find data to begin to uncover these links?

Lesson 3: Sector-Specific Data Stories

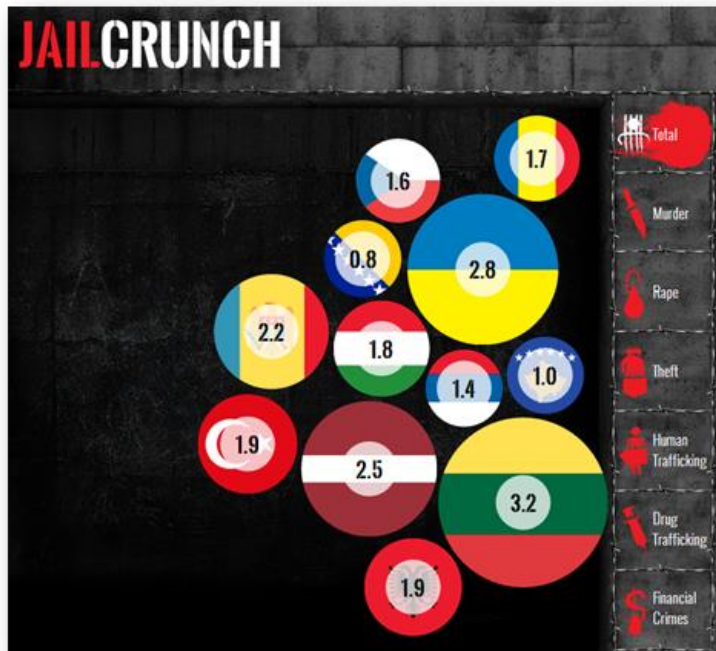
Examples

- Governance and Security
- Economy and Budget
- Education
- Health and Equality
- Human Rights and Equality
- Environment

Generally, data for data-driven decision making falls into a few broad sectors, for instance: governance, economy, education, health etc.

This lesson explores examples of how data has been harnessed to improve policy in other countries and brainstorms local applications.

Sector-Specific Example: Governance and Security Indicators



Governance data covers a wide variety of data sets including on topics such as voting records, campaign financing, parliamentary expenses, transparency ratings, and procurement. Security data can include everything from monthly crime statistics to prison populations and judicial sentences.

Take a look at this interactive story based on data received from freedom of information requests to prison authorities across Eastern Europe: <https://reportingproject.net/JailCrunch/>

Next, try to answer the following questions.

Data Questions:

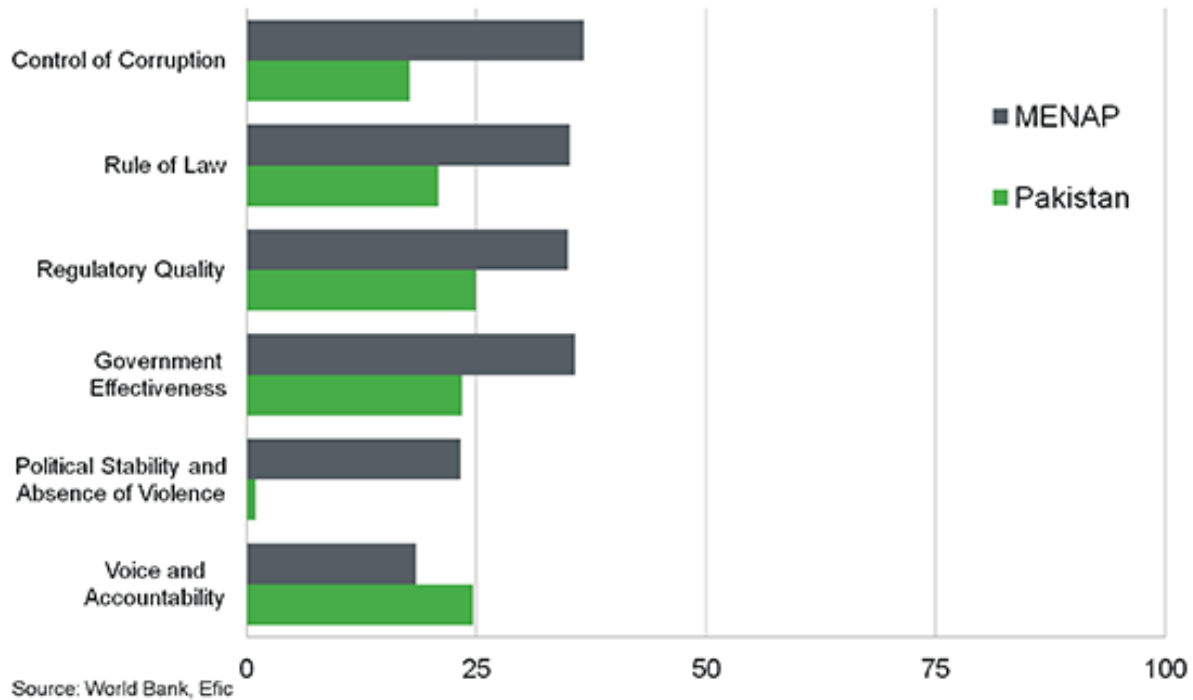
- Which country has the highest number of inmates?
- Which country has the highest percentage of inmates?
- What is the difference between the total number and percentage?
- What is the most common crime in Lithuania? How do you know?
- How does length of prison sentence influence this data?
- How does the effectiveness of the justice system influence this data?

Public Service Question:

What would your news angle be for the below governance data?

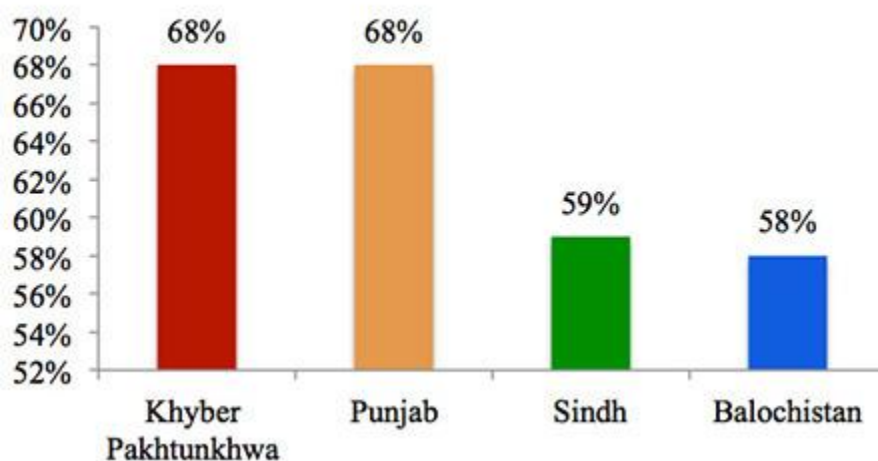
Thinking Like a Data Journalist

Governance indicators



Source: <http://www.efic.gov.au/education-and-tools/country-profiles/asia/pakistan/>

Governance Assessment Scores



Source: <http://www.samaa.tv/pakistan/2016/03/punjab-kp-on-top-in-quality-of-governance/>

Global Data Source If you are writing a crime or governance story about your country, consider mentioning how you compares to other countries on these issues. Check out:

Thinking Like a Data Journalist

- [Transparency International Corruption Perception Index](#) to see how corruption in your country compares to other countries
- [United Nations Office of Drugs and Crime \(UNODC\) Crime and Criminal Justice Statistics](#) to compare crime and imprisonment rates.
- [Stockholm International Peace Research Institute](#) for data on movement of weapons around the world

Pakistan Data Sources

[Official Gateway to the Government of Pakistan](#)

[Pakistan Establishment Division](#)

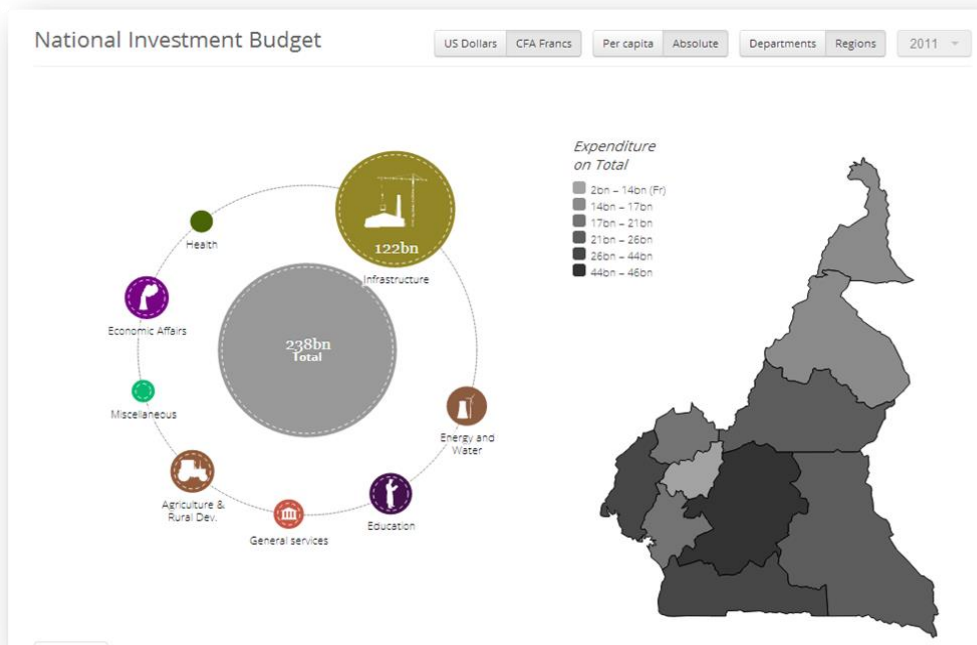
[Pakistan Bureau of Statistics](#)

[Ministry of Planning, Development and Reform](#)

[Ministry of Interior and Narcotics Control](#) (no data)

[Open Parliament](#)

Sector-Specific Example: Economic Indicators and Budgets



Source: <http://cameroon.openspending.org/en/pib.html/~total>

Economic data can be crucial for citizens to understand not only the economic outlook of the country but also how to make personal economic decisions that lead to economic growth and stability. Many economic reports come out regularly including government budgets and spending reports, inflation rates, economic growth reports, trade data and business environment surveys.

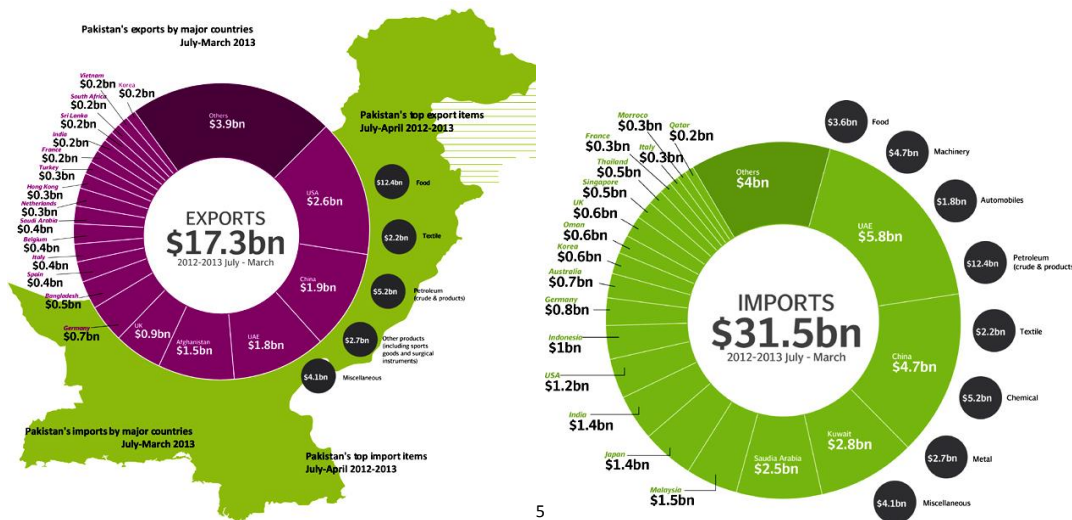
Take a look at the Cameroon Budget Enquirer, which visualizes national budget data for Cameroon using interactive charts: <http://cameroon.openspending.org/en/pib.html/~total>

Data Questions:

- Which ministry has the biggest budget?
- Which region under that ministry has the biggest budget?
- What about calculated per capita?
- What is the difference?

Public Service Question:

What would your news angle be for the below import/export data?



Data Sources

If you are writing an economic or budget story about your country, consider mentioning how you compares to other countries on these issues. Check out:

- [International Budget Survey](#) to see how your country's budget transparency compares to other countries.
- [World Trade Statistics](#) to investigate import and export data.
- [Doing Business Report](#): World Bank Data on the national regulation of business
- [Enterprise Survey](#): experience of business in the country
- [Aleph](#): investigative journalism database of companies and documents

Pakistan Data Sources

[Ministry of Finance](#)

[Federal Board of Revenue](#)

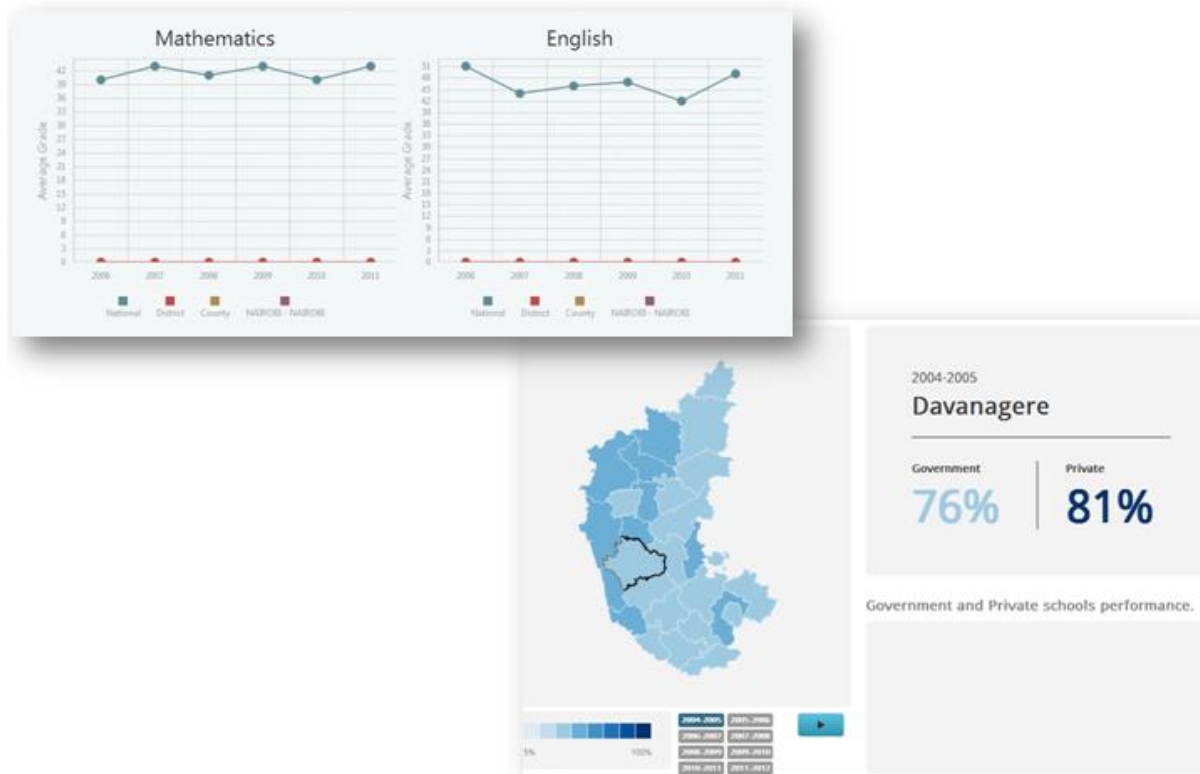
[Open Corporates Pakistan](#)

[Privatization Commission](#)

[Ministry of Industries & Production](#)

⁵ <http://tribune.com.pk/story/579954/feeling-the-pinch-as-trade-liberalises-large-players-look-forward-while-small-players-unsure/>

Sector-Specific Example: Education Indicators



Because of the critical role that education plays in economic and social mobility, education data is a rich source of stories that can impact the lives of citizens. Basic education data often includes the percentage of school age children in school in at primary, secondary and tertiary levels, ratio of girls to boys, completion rates and literacy rates. Often, detailed data is available for individual schools such as test score results by subject and the number of students transitioning on to the next level of education. This data can be important for parents enrolling their children in schools and governments interested in equalizing access to education among their constituency.

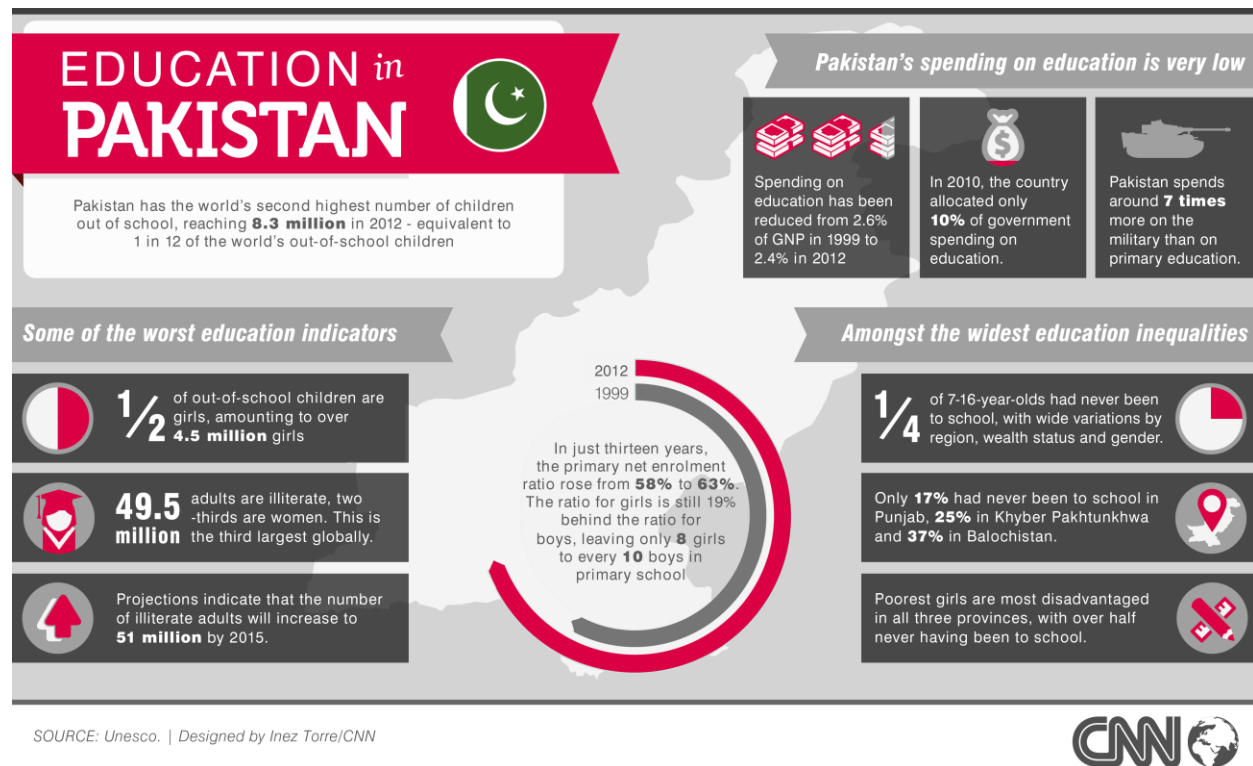
Here are some examples of different approaches to displaying education data by CSOs and media.

- Nine Years of SLCC in Karnataka, India
 - <http://sslc.klp.org.in/>
- KCPE Trends Kenya
 - <http://apps.openinstitute.com/kcpetrends/-trends>

- The Opportunity Gap, US
 - <http://projects.propublica.org/schools/>

Public Service Question:

What would your story angle be for the below education data? How about a local story using data like [this](#)?



This [story](#) from CNN provides an overview for an international audience of the education challenges facing Pakistan using data. [This](#) is a local data resource just about Punjab.

Data Questions:

- What kinds of data does each site provide?
- Which site is easiest to use?
- Which lets you make regional comparisons?
- Which lets you see change over time?

Global Data Sources

If you are writing an education story about your country, consider mentioning how your country compares to other countries on these issues. Check out:

- [UNESCO Education Statistics](#): global rankings on education, literacy, investment in education.
- [Global Competitiveness Index](#) by the World Economic Forum ranks countries by preparedness for the market

Pakistan Data Sources

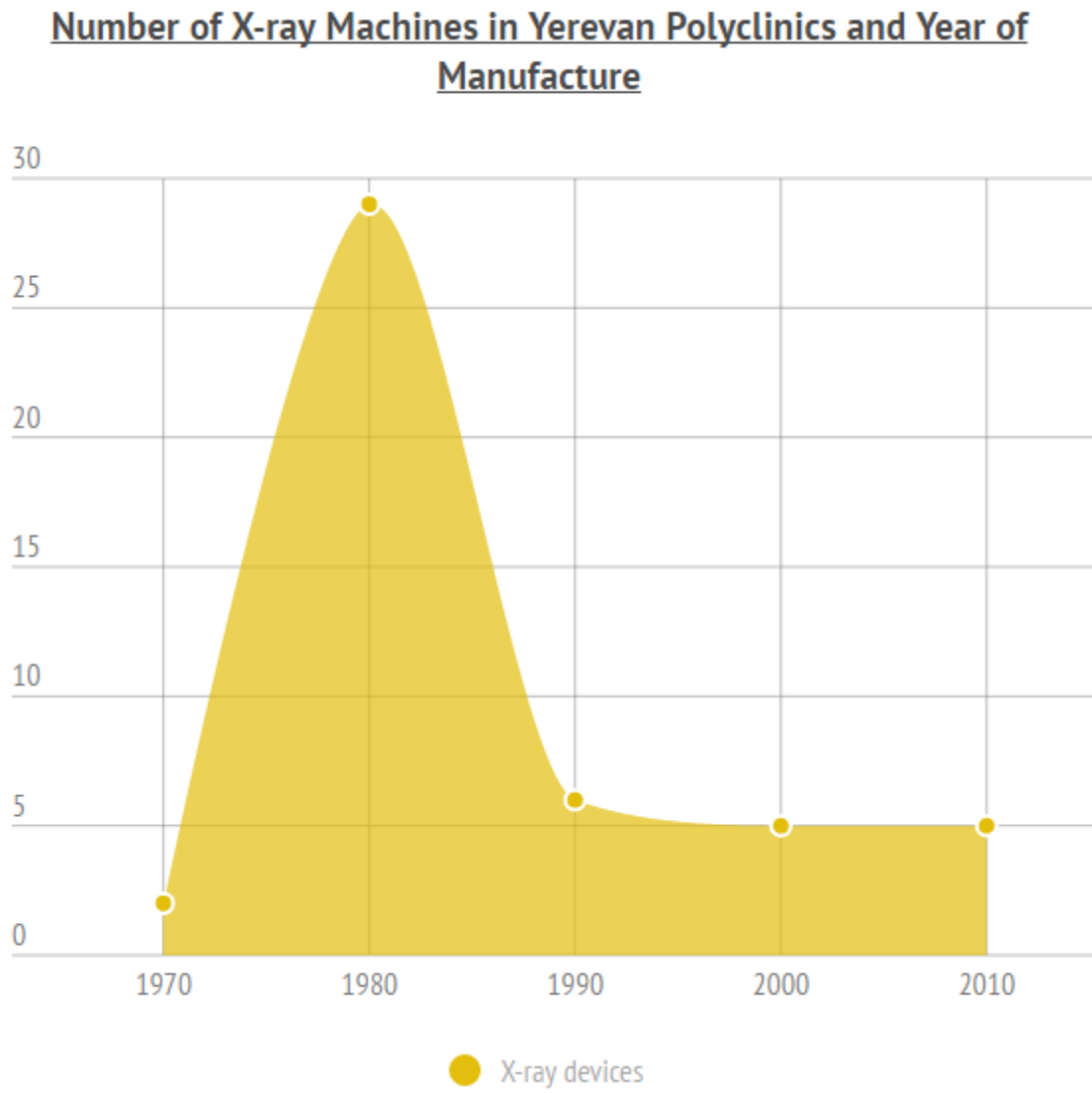
[Ministry of Education](#)

[Pakistan Data Portal Education](#)

Sector-Specific Example: Health Indicators

Public health indicators are a rich source of data to evaluate government provision of basic services. Good places to start are with data such as the number of doctors, nurses and hospital beds per 1,000 people, leading causes of death, rates of child death and maternal death, and spending per capita on health. This data can help identify where healthcare shortages are most severe and their impact. Deep dives into a specific area, such as malnutrition among children, can help you identify causes such as lack of breastfeeding. Above, the [micronutrient infographic](#) explains how supplying one missing nutrient can have a huge health impact.

Take a look at [this example](#) that uses data on the manufacturing of x-ray machines at free health clinics in Yerevan, Armenia to determine how many are expired and could be leaking dangerous radiation.

**Data Questions:**

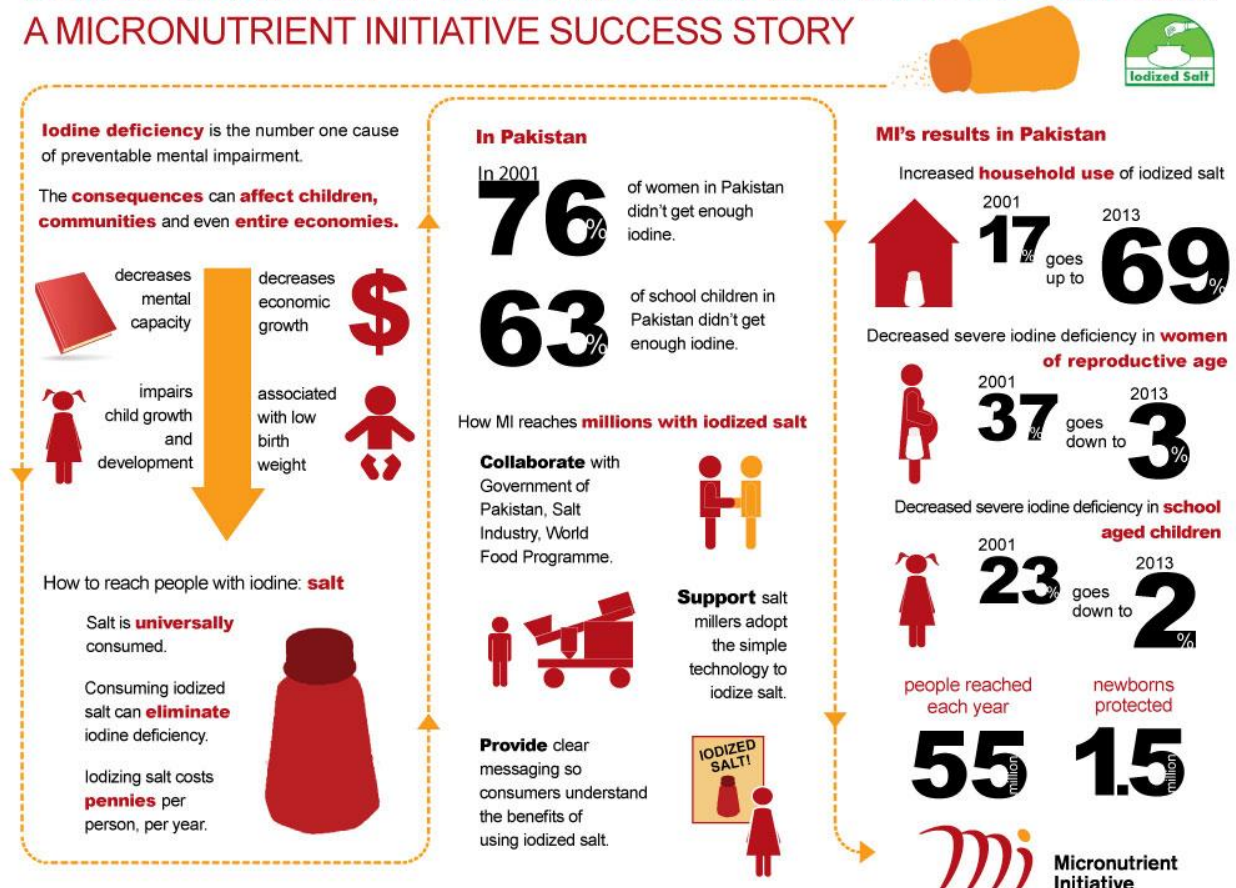
- Where does the data come from?
- What does the map add to understanding the issue?
- What additional information is available on a county level?

Public Service Question

Thinking Like a Data Journalist

What story angles do you see in the below data?

HOW IODIZING SALT IS HELPING WOMEN & KIDS IN PAKISTAN A MICRONUTRIENT INITIATIVE SUCCESS STORY



Data Sources

If you are writing a health story about your country, consider mentioning how you compares to other countries on these issues. Check out:

- [World Health Organization Statistics](#): global data on disease, health systems and health financing
- [Global Health Data Exchange](#): data that allows you to make easy comparisons between health outcomes in neighboring countries
- [GDB Data Tool](#): Global mortality data by cause of death, gender, age group

Pakistan Data Sources

Thinking Like a Data Journalist

[Pakistan Health Information System](#)

[Drug Regulatory Authority of Pakistan](#) (no data)

Sector-Specific Example: Human Rights and Equality Indicators



Human rights and equality data often address the people and issues that are most neglected by society and government. This includes data about the disabled, ethnic or religious minorities, women and the elderly. Since these topics and hard-to-reach groups are so rarely discussed openly, these types of data can open a window into the world of marginalized groups and make their challenges concrete and solutions more tangible.

Refer to this example, which maps deaths of migrants who have tried to reach Europe - access <http://www.themigrantsfiles.com/>, and search for countries that are at war.

Data Questions:

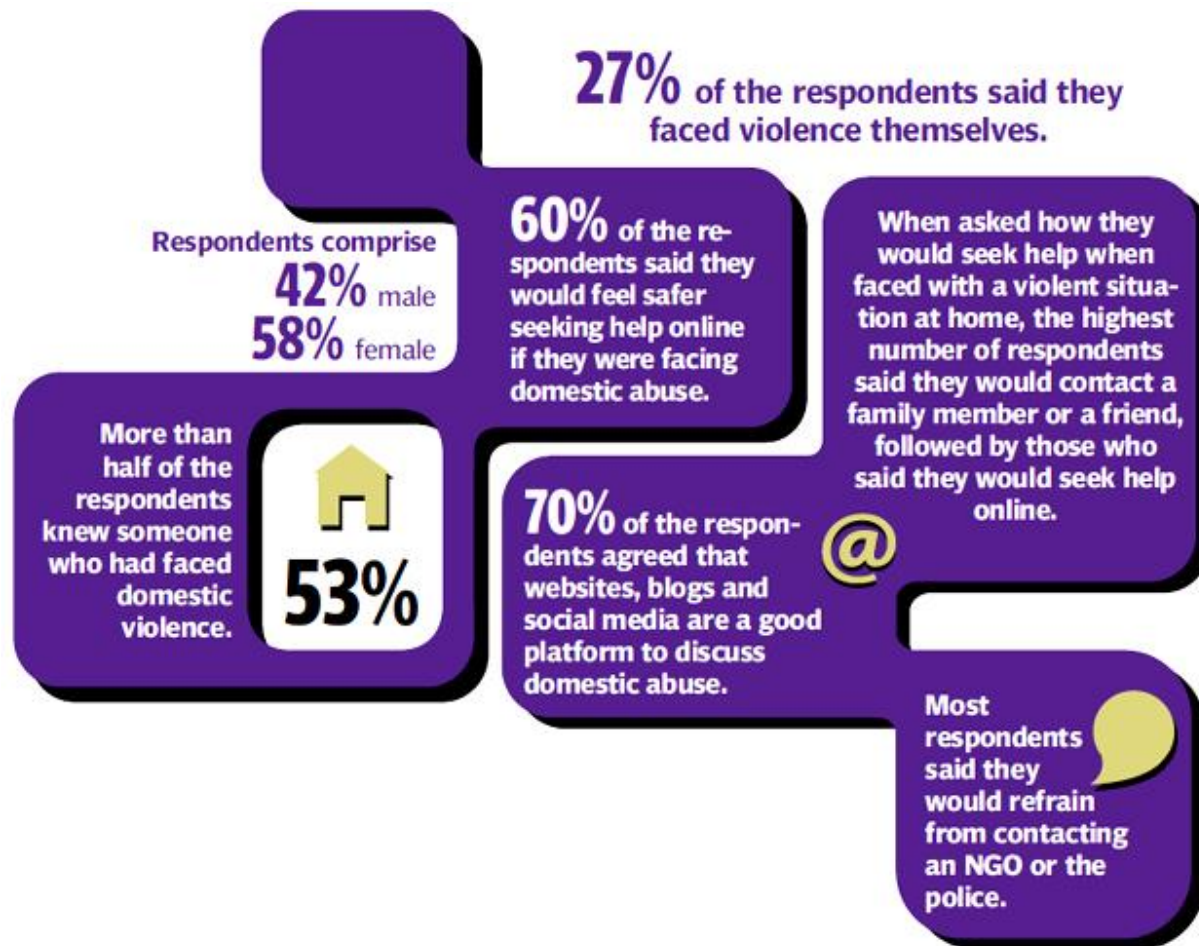
How much money does the European Union spend on deporting migrants?

What are the sources for the 'counting the dead' data?

What period of time is covered?

Public Service Questions:

What story angle would you pursue using the data summary below?



In this story by the [Express Tribune](#), journalists overcame the scarcity of domestic violence data by conducting their own survey to measure the prevalence of domestic violence and potential solutions.

Global Data Sources

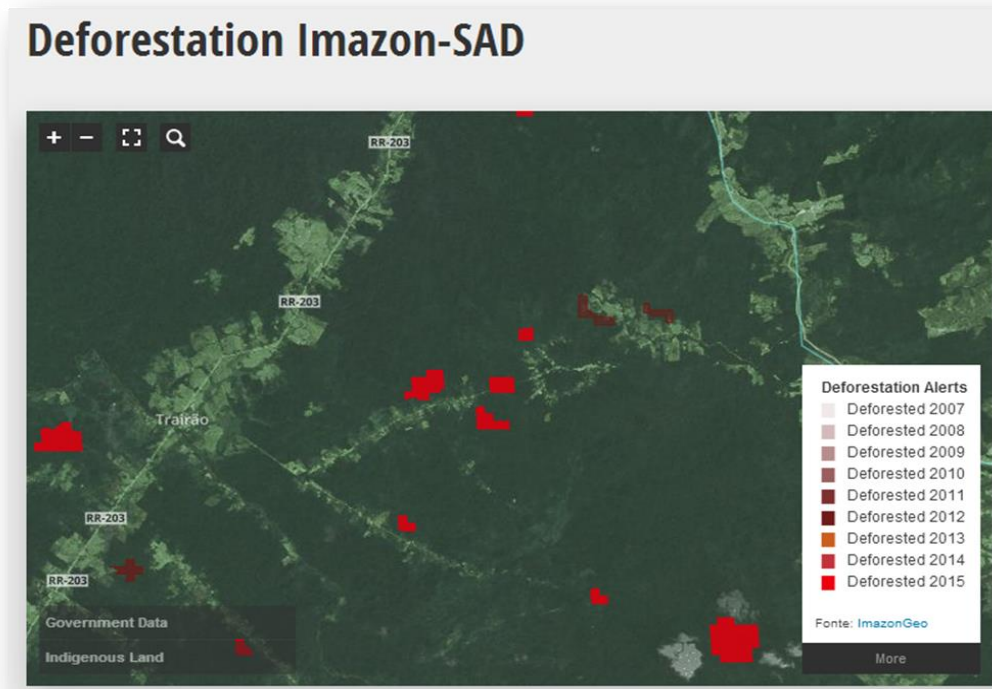
If you are writing a human rights story about your country, consider mentioning how you compares to other countries on these issues. Check out:

- [Universal Human Rights Index Database](#): global data on various human rights indicators
- [Freedom House Freedom in the World Index](#): data comparing different measures of freedom

Pakistan Data Sources

[Human Rights Commission of Pakistan](#)

Sector-Specific Example: Environment and Agriculture



Environmental and agricultural data can be a polarizing topic. On the one hand, exploitation of natural resources can lead to economic growth in a country but on the other hand, often local communities are negatively impacted and precious natural resources are destroyed. Analyzing this data can help citizens weight the costs and benefits of current environmental policy and advocate for changes.

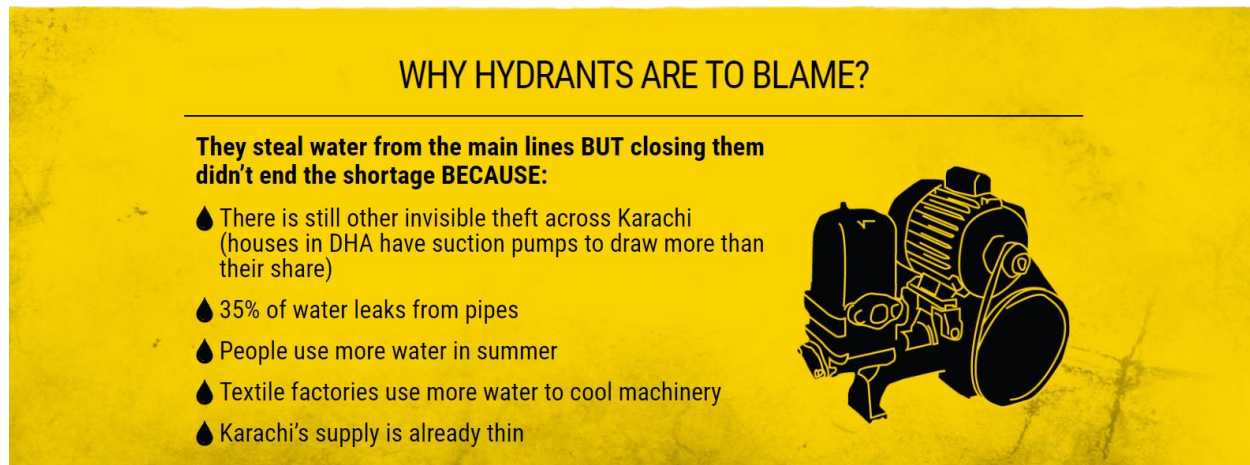
For example: InfoAmazonia a platform that mashes up geo-based environmental data and environmental media coverage: <http://infoamazonia.org/maps/>

Data Questions:

- What kind of data is displayed on the map?
- What is the effect of showing news stories along with the data?

Public Service Questions:

What other story angles do you see to cover water issues in your region?



In [Aqua Final](#) by the Express Tribune, journalists examine the causes, impact and solutions to Karachi's water shortage, a topic which touches not only on environmental issues but also on governance and health challenges facing the city.

Global Data Sources

If you are writing an environmental story about your country, consider mentioning how you compares to other countries on these issues. Check out:

- [Environmental Performance Index](#): measures how countries are doing on a variety of priority environmental topics
- [World Air Quality Index](#): up-to-date air quality data from around the world
- [Resource Project](#): Publically available mining contracts
-

Pakistan Data Sources

[Environmental Protection Agency](#)

[Bureau of Statistics Mining Data](#)

[Ministry of Water and Power](#)

[Ministry of Petroleum and Natural Resources](#)

[Ministry of National Food Security and Research](#)

Exercise

Develop one story idea using one of these ways into a story:

- Hypothesis
- Trends and contrasts
- Outliers
- Invisible connections

Choose one for each of these sectors and write a story idea for each:

- Governance and Security
- Education
- Economic Indicators and Budgets
- Health
- Human Rights
- Environment and Agriculture

Lesson 4: Introduction to Investigative Journalism

Often, once you begin to access, analyze, and tell stories about data from a specific sector - you become quite the expert on the topic. This is when the line between data journalism and investigative journalism begins to blur. The goal of both data and investigative journalism is to explain not just what is happening, but why it is happening and outline potential solutions.

Objectives of Investigative Journalism⁶

- An original, proactive process that digs deeply into an issue or topic of public interest
- Producing new information or putting known information together to produce new insights
- Multi-sourced, using more resources and demanding team-working and time
- Revealing secrets or uncovering issues surrounded by silence
- Looking beyond individuals at fault to the systems and processes that allow abuses to happen
- Bearing witness, and investigating ideas as well as facts and events
- Providing nuanced context and explaining not only what, but why
- Not always about bad news, and not necessarily requiring undercover techniques – though it often is, and sometimes does.

⁶ http://www.investigative-journalism-africa.info/?page_id=31

Comparison of Conventional and Investigative Journalism

Source relations	
The good faith of sources is presumed, often without verification.	The good faith of sources cannot be presumed; any source may provide false information; no information may be used without verification.
Official sources offer information to the reporter freely, to promote themselves and their goals.	Official information is hidden from the reporter, because its revelation may compromise the interests of authorities or institutions.
The reporter must accept the official version of a story, though he or she may contrast it to commentaries and statements from other sources.	The reporter may explicitly challenge or deny the official version of a story, based on information from independent sources.
The reporter disposes of less information than most or all of his sources.	The reporter disposes of more information than any one of his sources taken individually, and of more information than most of them taken together.
Sources are nearly always identified.	Sources often cannot be identified for the sake of their security.

Take a look at this comparison of conventional and investigative journalism:

<http://unesdoc.unesco.org/images/0019/001930/193078e.pdf> (page 9)

You may notice that the sections under investigative journalism describing research and outcomes sound very similar to what we have discussed about data journalism. The big difference is really around the treatment of sources. In traditional investigative journalism the source of the information is often a secret or involves leaked information. Often, in investigative data journalism, the source is official data, either from the government or private sector, hiding in plain sight. Often, official quarterly spending reports from government or private companies, procurement data and tax records can yield incriminating findings that may or may not be legal. In either case, often the journalist is safer in publishing a data story because the source of information is official. The journalist just reveals something that no one has noticed before, as in the example of Kenya county spending on health. As more data becomes available, evidence-based decision making based on open records law should become easier and more common tools for accountability.