

Planning a Data Story to Break the Filter Bubble

The screenshot displays a news website interface with three main columns:

- Pajhwok Services:** Includes a 'Pajhwok Website Manager Vacancy' link, a 'Video Service' section featuring a video player with the title '80 dead, scores hurt as blasts rip through Kabul rally (Video)', and a 'Photo Service' section with a photo of people in a public square.
- Data Journalism Stories:** Contains three articles:
 - 'Over 10,000 people killed, wounded in past five months' by Ahmad Shah Erfanyar, dated Jan 04, 2017. It includes a scatter plot visualization of data points across a map of Afghanistan.
 - 'Education boosting vaccination drive across the country' by Ahmad Shah Erfanyar, dated Jan 03, 2017. It features a photo of a child receiving a vaccine.
 - 'Casualty toll hits nearly 1,400 in November' by Ahmad Shah Erfanyar, dated Dec 18, 2016. It includes a photo of a street scene with a red and white striped barrier.
- Advertisement:** Promotes a mobile app for 'PAJHWOK NEWS ON YOUR MOBILE' and includes 'Available on the App Store' and 'Get it on GOOGLE PLAY' buttons, along with a logo for the 'National Solidarity Programme'.

The key to success for any data journalist is organization. Unlike in many other kinds of journalism, how you decide to organize your information and narrate your story can make or break your story. The process we will follow for organizing a data journalism story consists of six steps:

1. Background
2. Hypothesis
3. Questions
4. Analysis
5. Interviews
6. Story Organization
7. Visualization

As we proceed through this course, we will cover the tasks and skills required for each step. In this lesson, we will cover the two planning stages: background, hypothesis and questions.

Background

The first step in looking for a public interest data topic is to search for unanswered questions about news issues affecting marginalized communities. There might be a lot of coverage of the topic, but not an in-depth story that helps affected populations understand the root of the issue.

In the case of Afghanistan, there was frequent coverage of horrific domestic violence cases and speculation about whether the new domestic violence law was doing enough to protect women. The journalist wanted to answer that question for the public.

The second step is to search for other data stories produced by other journalists on the same topic. This serves several purposes. It familiarizes you with how other journalists have approached the issue, where and what kind of data he or she used and what storytelling strategy was effective.

Case studies

Using advanced Google search techniques, they identified three similar data stories in the media:

[India is a Nation of Violent, Stressed Men](#), IndiaSpend, India

[Till Death Do Us Part](#), Post and Courier, USA

[Most Dangerous Transport System for Women](#), Global Post, Global

From these examples the journalists noted:

- The government's ability to provide services is key in determining whether or not victims of domestic violence survive their experience and whether justice is done.
- Rates of domestic violence and reporting vary widely across geographical regions
- Visualizations can be effective in showing the scope of the problem

Reports/Data

In the next stage, journalists identify reports and data related the subject of the investigation. These reports can be found through searches, through data requests to the government and CSOs or through the creation of data for the investigation.

For the domestic violence story, journalists identified the following key reports:

[Justice through the Eyes of Afghan Women: Cases of Violence against Women Addressed through Mediation and Court Adjudication UNAM?OCHA](#)

[An Update on Implementation of the Law on Elimination of Violence against Women in Afghanistan](#)

[AFGHANISTAN Ending Child Marriage and Domestic Violence](#)

[USIP Women's Access to Justice in Afghanistan](#)

[World Bank Gender Data Portal](#)

Methodology

To evaluate the source of the data, journalists answer the following questions (here with sample answers from the first report)

Who gathered the data?	United Nations Assistance Mission in Afghanistan
When was the data gathered?	Detailed information from 18 of Afghanistan's 34 provinces for the one-year period October 2012 to September 2013 with technical review by the UN Office of the High Commissioner for Human Rights
What time period does the data cover?	Two years (2012-2013)
How was the data gathered?	Field monitoring and analysis of police and court records

Important findings

Finally, journalists read the executive summary of the report and write down 3-5 interesting findings.

1. The number of reports of domestic violence is rising
2. The economic and social vulnerability of women remains constant
3. Most cases of domestic violence are registered through the police
4. Most cases of domestic violence are resolved through mediation

Never skip background research when exploring a new topic. Most likely, the topic has already been investigated and analyzed by many people and you want to make use of the important work they have already done and find fresh angles that are relevant to your audience.

Hypothesis

The most important stage of the data story process is the formulation of the hypothesis and questions. In scientific research, an experimenter develops a hypothesis with a suspected set of ideas and then builds an experiment to support the hypothesis. The same process is true for data journalism. The journalist, using his or her news nose, develops a hypothesis that can be proved or disproved with data. News articles have a habit of bringing up more questions than they answer, and following up with a hypothesis and investigation can uncover details that lead to a data journalism story.

Building a Hypothesis for Data and Investigative Stories¹

1. **A hypothesis gives you something to verify, instead of trying to uncover a secret.** People do not give up their secrets without a very good reason. They are much more likely to offer confirmation of information that is already in your possession, simply because most people hate to lie. A hypothesis enables you to ask them to confirm something, rather than to advance information. It also puts you in the position of someone who is open to discovering that there is more to the story than he or she thought at first, because you are willing to accept that there are facts beyond what you suspected at the start.
2. **A hypothesis increases your chances of discovering secrets.** A lot of what we call “secrets” are simply facts that no one ever asked about. A hypothesis has the psychological effect of making you more sensitive to the material, so you can ask those questions. As the French investigator Edwy Plenel said, “If you want to find something, you have to be looking for it.” We would add that if you’re really looking for something, you’ll find more than you were looking for.
3. **Hypotheses makes it easier to manage your project.** Having defined what you’re looking for, and where to start looking for it, you can estimate how much time the initial steps of the investigation will require. This is the first step to treating an investigation as a project that you can manage. We’ll return to this point at the end of this chapter.
4. **Hypotheses are a tool that you can use again and again.** When you can work in a methodical way, your career will change. More important, you will change. You will no longer need someone to tell you what to do. You will see what needs to be done to combat some of the chaos and suffering in this world, and you will be able to do it. Isn’t that why you became a journalist in the first place?
5. **A hypothesis virtually guarantees that you will deliver a story, not just a mass of data.** Editors want to know that at the end of a specific period of time – a specific investment of resources – there will be a story to publish. A hypothesis hugely increases the likelihood of that outcome. It enables you to predict a minimum and maximum positive result for your work, as well as a worst case.
 - The worst case is that verification of the hypothesis will quickly show there is no story, and the project can be ended without wasting significant resources.
 - The minimum positive outcome is that the initial hypothesis is true, and can be quickly verified
 - The maximum is that if this hypothesis is true, others must logically follow, and either a series of related stories or one big story will result.

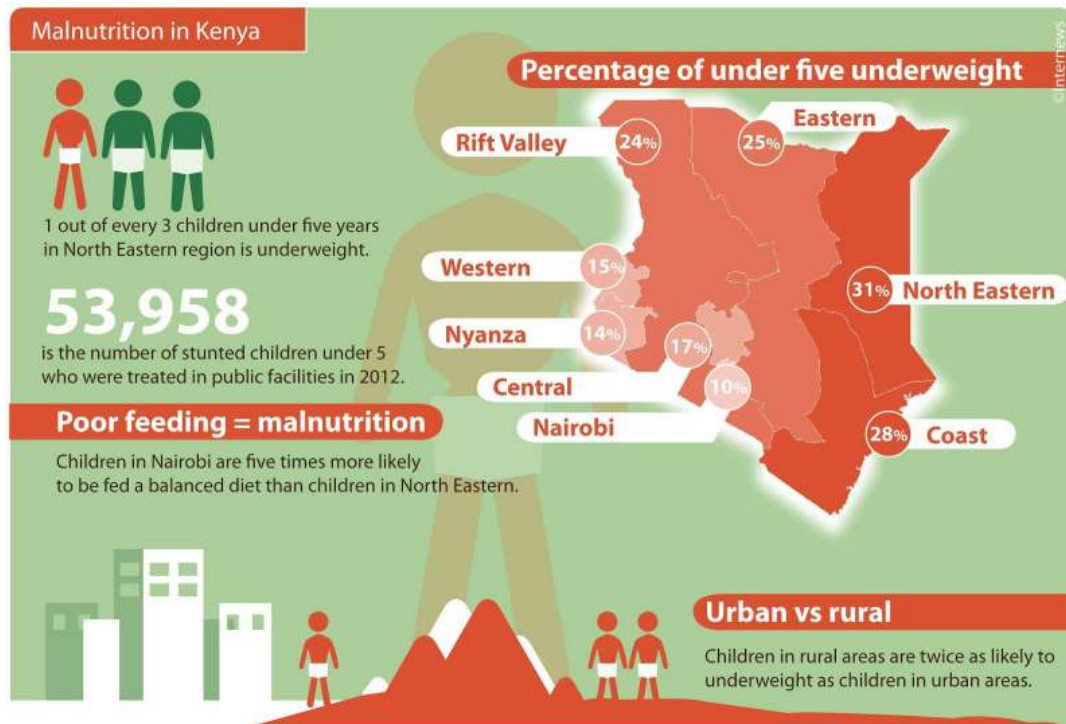
¹ <http://unesdoc.unesco.org/images/0019/001930/193078e.pdf>

Tips for a strong hypothesis:

- Posits a theory that can either be proven or disproven with data
- Is specific about what is being measured
- Measures the problem, causes, impact and solutions
- The data is available
- The topic is important to the public

Below is an example of how to transform a weak hypothesis into a strong hypothesis:

1. Children in this country are dying of malnutrition.
2. Most under-five deaths in the country are caused by malnutrition. *(Add an indicator that can be measured "under five cause of death.")*
3. Most under-five deaths in the country are caused by malnutrition and victims live in the poorest provinces. *(Add a geographical measure and an economic measure)*
4. Most under-five deaths in the country are caused by malnutrition and victims live in the poorest provinces despite a donor funded feeding program that was supposed to cut malnutrition rates in half over five years. *(Add a measure of whether the solution to the problem is working or not)*



Five common mistakes when formulating a hypothesis

1. The hypothesis is too simple and measures only the general problem.
Examples: Women are not getting jobs after graduating from university.
2. The topic is too broad. It would require measuring too many indicators. The topic is better for a book than a single story. Additionally, topics that are too broad become too general and don't help solve specific parts of the problem.
Examples: Kyrgyzstan could improve mother and child health by investing more in healthcare.
3. One half or both halves of the hypothesis cannot be proved with data.
Examples: People get sick after going to the hospital because doctors are dispensing counterfeit medicine.
4. The hypothesis has already been proven true and is common knowledge.
Examples: Women in rural areas are more likely to die during childbirth because of lack of access to healthcare facilities.
5. The hypothesis is too narrow: it only measures how one factor influences a trend and discounts other data sources that might also contribute to the trend. There is a high risk of confusing causation with correlation.

Examples: Girls drop out after primary school because of lack of access to toilets.

Hypothesis

After completing the background section for their investigation the Afghan journalists developed the following hypothesis:

Government programs cannot keep up with the increased demand for domestic violence services.

As you can see this hypothesis posits two theories that can be proven or disproven with data:

- Domestic violence reporting rates are increasing
- Government services to respond to these reports are inadequate.

Questions

Once you have a strong hypothesis, you should develop at least five questions that can be answered with data to prove or disprove your hypothesis. All the questions should be able to be answered with a number. Other types of questions, such as interview questions, will be developed after analysis.

The questions you write should:

- Problem questions: how big is the problem? How expensive is it? Is the problem getting better or worse?
- Impact questions: Who is affected by the problem? How? Are some groups more affected than others?
- Cause questions: What are the causes of the problem? What factors have made it worse?
- Solution questions: What is the solution to the problem? How can we measure effectiveness?

Hypothesis: Most under-five deaths in the country are caused by malnutrition and victims live in the poorest provinces despite a donor funded feeding program that was supposed to cut malnutrition rates in half over five years.

Question Type	Questions
Problem questions	1. What are the national rates of under-five child death over the last 10 years? And in neighboring countries?

	2. What percentage of those deaths each year are attributed to malnutrition?
Impact questions	3. What are the rates of malnutrition by province? 4. What are the rates of poverty by province?
Cause Questions	5. What are the rates of stunting by province? 6. What is the level of food insecurity in each region?
Solution Questions	7. What were malnutrition rates when the donor funded feeding program started? 8. What were the rates after five years? 9. How many children were served by the program? 10. How many would have to have been served to meet the goal? 11. What was the cost per child?

Questions

Remember the hypothesis for the domestic violence story:

Government programs cannot keep up with the increased demand for domestic violence services.

These are the questions the journalists wrote:

Problem:

1. What percentage of women have experienced domestic violence in Afghanistan?
2. What percentage of women who have experienced domestic violence in the last 12 months report it?
3. Are the number of cases being reported each year going up or down?

4. What is the trend in budgeting for domestic violence cases per woman?
5. What are the different bodies tasked with providing domestic violence services and what are their budgets?

Impact:

6. What is the age breakdown of women registering domestic violence cases under the new law?
7. What are the education levels of women reporting domestic violence?
8. What is the rate of women experiencing domestic violence in Afghanistan compared to the region?
9. What are the most common types of domestic violence reported by women?
10. What is the economic and employment status of women reporting domestic violence?

Cause:

11. Across the country, who decides whether to investigate and prosecute cases of domestic violence?
12. What percentage of cases that are registered go to court? What happens to the rest of the cases?
13. How long is the resolution time for cases that go to mediation?
14. What kinds of violence are being prosecuted under the new law?
15. How much does it take to process a mediation case vs a court case?

Solution:

16. What percentage of cases that make it to court result in a conviction?
17. What percentage of cases settled in mediation result in a woman returning to her family?
18. What percentage of women returned to her family end up reporting a new case?
19. What would the required budget be to assign a prosecutor to evaluate all reported domestic violence cases?
20. What are the government's benchmarks for yearly progress on domestic violence services?

Stories

[Cases of violence against women: Is mediation the best option?](#)

Other data units following this methodology:

<http://datajournalismpakistan.org/>

<http://www.odecanet.org/stories/>

<http://www.nation.co.ke/newsplex/>

<http://www.indiaspend.com>