



## How to Use This Booklet

---

Data visualization can make information more memorable, more persuasive, facilitate understanding and ultimately motivate action. And within human rights research, it can help investigators and researchers draw a bigger picture from individual human rights abuses by allowing them to identify patterns that may suggest the existence of abusive policies, unlawful orders, negligence, or other forms of culpable action or inaction by decision-makers.

The purpose of this activity is to explore some of the ideas and principles around designing effective data visualization for human rights advocacy.

This activity is broken into a series of six topics each with its own PDF guide.

In practice not every visualization process follows these exact steps in this same order. However, for a workshop setting, we present these as a way to walk through the topics. Each topic has a corresponding list of options and choices. Read through each topic and follow the instructions. Explore the options for each step as you progress.

---

The six steps are:

**Step 1:** Choose a human rights issue

**Step 2:** Discuss some kinds of data you might acquire

**Step 3:** Consider what question are you trying to answer with your data and visualization

**Step 4:** Choose a chart type for your visualization

**Step 5:** Consider some data and visualization hazards

**Step 6:** Consider some ways your charts can be improved

## Step 2

### What are some kinds of data you might acquire?

---

Based on your human rights issue, consider a few different kinds of data you might acquire. To help with your data selection, see our list of a few different data collection methods. You don't necessarily need to be limited by this list of methods.

Should you use primary or secondary data? Primary data is data you gather through direct observations, interviews, surveys, and other tools. Secondary data is gathered by someone else (e.g. researchers, government institutions, NGOs, etc.) or for some other purpose than the one currently being considered or often a combination of the two.

What are the challenges of working with these types of data, and what questions should we be asking when using this data?

- With primary data, you can gather the data that best fits with the human rights issue you are exploring. Plus, you know what you are getting—there is no mystery concerning the methods used to gather information, how it has been categorized, and what limits it contains.
- On the other hand, gathering primary data often requires expertise, and it can be expensive and time-consuming. Secondary data is often available online, is frequently presented in neat and clean formats, and may have been gathered by experts. It may be

---

standardized, making comparisons across space and time easy and reliable.

- However, secondary data may not be available for some issues or some locations, and it could contain hidden biases or assumptions. Looking “under the hood” at the data-gathering process is important. It is also crucial to acknowledge that some aspects of rights may not be quantifiable.

What kind of data is available?

Beyond the question of primary and secondary data, it is important to think about the type of data available to you. Some of the most prevalent types of data relevant to human rights include:

- Event-based data: data concerning specific events that have human rights relevance, such as extrajudicial executions, cases of torture, forced evictions, or illegal water shut-offs. These are specific events that are in themselves violations of human rights.
- Administrative or service data: information about the users of government or other services such as health care, water and sanitation services, electricity, or education. This data may be relevant when asking about a variety of dimensions of human rights norms.

- 
- Census, household, or other survey data: data drawn from randomized, representative samples allow you to assess the enjoyment or deprivation of some rights within an entire population, or—in the case of opinion or perception surveys—the views of an entire population about rights-related issues. There is a wide variety of survey data that is relevant to human rights, though it may not be presented as rights data per se (for example, data about child malnutrition or under-5 mortality is relevant to children’s rights and the right to health, though it may not be discussed as such; similarly, data about perceptions of corruption reveal a great deal about rights related to participation and civic life but may not be presented as such).
  - Big Data: large unstructured datasets created as a byproduct of recent technological innovations and/or used for purposes other than those for which the data was originally collected. Big Data can be used to create statistical predictions concerning human rights violations, potential victims, and possible perpetrators. These applications remain at the nascent phase.

What are some structural indicators, process indicators, or outcome indicators you can measure or have access to?

- 
- Structural indicators measure the existence of institutions, legal rules, and other mechanisms needed to implement human rights. Basic counts of things like laws, constitutional provisions, or accountability mechanisms might be useful structural indicators.
  - Process indicators measure the efforts of governments and other duty-bearers to implement human rights laws and make their enjoyment real. Some examples include opinion data about trust in courts, data about the proportion of complaints of police misconduct that are satisfactorily resolved, and the proportion of those accessing food aid who are members of a national minority group.
  - Outcome indicators reflect the attainment—or lack of enjoyment—of rights by all. Quantitative data about prevalence rates for preventable illnesses, literacy rates, and rates of access to potable water are some examples of outcome indicators. (See more in relation to [questions to interrogate human rights data](#).)

## Automated Analysis

---

Automated data analysis is the use of computer algorithms to scan, process, and analyze data.

Software can be used to automate some kinds of data collection, and for cleaning, interpreting, and summarizing for reporting and analytical purposes. Depending on the task, software can perform operations at great speed and reduce manual labor.

Different software is available in the market, depending on the type of data automation and analysis required. Automated analysis is particularly useful for repetitive tasks, such as monthly reporting on a set of standardized internal data.

Machine learning can also be used to automate tasks such as recognizing images, generating categories, extracting proper nouns etc.; that is, analyzing qualitative data. Machine learning has been used to detect incidents of [hate crimes](#) reported in the news, and to uncover [mass graves in Mexico](#).

As with any other method of analysis, computer models should be rigorously reviewed. Biased data collection, modeling and weighting can produce biased results.

## Census

---

A census involves gathering information from every member of the population. A common use of census is by governments, when recording the country’s population statistics to inform policy and planning.

Because of the large size of most populations, and the resources required in conducting a census, this method is not commonly used unless absolutely necessary. Populations may be dispersed or inaccessible, making it impossible to achieve a complete census even if desired. Due to the time and labor involved in reaching the population, as well as entering the data and interpreting it, a census can be extremely expensive to conduct.

While a census provides the greatest accuracy of results as compared to other data collection methods, the increase in accuracy usually does not justify the added cost. An exception to this may be where the population size is very small.

The inclusion of minority, marginalized, and undocumented populations is also a challenge for censuses. These populations may be overlooked as a result of methodology, or by intent.

## Commercial Data

---

Commercial data is data collected by or processed by commercial organizations.

While commercial data can be easy to access, different data sets can vary in terms of the quality and integrity of data, and careful attention should be paid to the data production methodology, as well as the process of ground-proofing the data.

Proprietary data may also have restrictive licenses that limit the scope of how the data can be used or published.

An example of proprietary data would be a database of corporate ownership used to track the flow of revenues. High-quality satellite imagery and other digital geospatial shapefile data can also be licensed from commercial vendor.

## Crowdsourced Data

---

The term “crowdsourced data” refers to two main types of data: data collected and submitted from a wide variety of sources, and data analyzed or processed by a large group of “microtaskers.” [Microtasking](#) is where large tasks are split into smaller tasks that individuals can complete over the internet with little or no training. Microtasking is especially useful for tasks that are repetitive, but require human judgment and cannot be completed accurately by software, such as categorizing content, audio transcription, or identifying features in an image. Amnesty International used its extensive membership network [to help examine](#) thousands of satellite images of remote parts of Darfur where bombings and chemical weapons attacks are suspected to have taken place. Crowd-sourcing enables large amounts of data to be collected and processed with great speed, and at a relatively low cost. Volunteer-contributed data is valuable when information about human rights violations is dispersed and not easily gathered. For example, [Ushahidi](#) was developed in Kenya to allow victims of electoral violence to submit data about their experience to a single online hub. While such data can be crucial in emergency or otherwise insecure settings, it may also be prone to error, or skewed by accessibility. Crowdsourced data can sometimes be processed using statistical tools to improve the accuracy and quality of the data collected.

## Data Log

---

Data logs are computer generated data sets, collected through sensors. Data loggers are usually small, portable, battery powered devices, equipped with sensors, memory storage, and a microprocessor. They can be stand-alone devices, or can be used in conjunction with a computer. When used with a computer, the computer not only collects the data, but can also analyze it, and save the output.

Data loggers can be set to automatically gather data over a specified period—for instance, social media output around a particular incident.

Data logs have been used by human rights defenders to identify sources of hacking and internet surveillance. [OONI](#), the Open Observatory of Network Interference, a smartphone application, uses network data to detect censorship, surveillance, and traffic manipulation on the internet.

Data logs can also be connected to environmental sensors to track air, water, or soil contamination over time.

## Document Review

---

Document review is a data collection method that involves gathering information by reviewing and analyzing existing documents on the subject. These may be internal organizational documents, such as a previous research study, or they may be external documents, such as public court records, government publications, or other third party research.

Document review can help provide background information required to define the scope of more detailed research. Review of public records can establish evidence of dates, times, and chains of command.

“Meta-analysis” is a research technique that uses a structured approach to analyze previous studies.” Using statistical techniques, the review can derive quantitative results.

Human rights research sometimes draws on events or names of victims from news accounts—coverage, however, may be inconsistent or unrepresentative, making the matching and de-duplicating of names from news accounts complicated and time-consuming.

## Event-Based Data

---

Event-based data collection involves gathering information related to specific events, including when, why, and how they occurred. For example, gathering data related to human rights violations or data about fatalities due to violence are examples of event-based data. Data may be collected from “convenience samples” such as case files, crowd-sourcing platforms, news accounts, or court records; or from random samples, such as household surveys that collect data on incidents of human rights violations.

The data collected may identify perpetrators, victims, and circumstances in each case, and include other useful information such as timelines, locations, and demographics of those involved. Such data can be analyzed to identify useful patterns and trends, and accordingly develop measures for improvement. Event-based data can also help answer the questions of why an event occurred, or what the statistical predictors are for such events that may occur in the future. Automated analysis and visualization may help reveal trends, though the insights gained may depend largely on the judgment and skills of the researcher.

In the human rights context, event-based data is rarely captured through random sampling, raising the risk of statistical bias. Analysts should carefully consider these issues before choosing a visualization for their data.

## Geospatial Data

---

Geospatial data includes data regarding the size, shape, or location of the subject of inquiry. Data can be labeled by city and state; by unit in a survey (such as a household or health facility); or by geographic coordinates such as longitude, latitude, and altitude. Geographic Information Systems (GIS) software can be used help collect, visualize, manipulate, and analyze this data. For example, a study on the right to water might include geospatial data about the location of water points such as wells and hand pumps, as well as their distance from key locations such as households and schools. Geospatial data can be linked with other data to add another dimension to the data gathered. Visualization can help spot trends and patterns in data, for example, mapping respondents of a survey to show areas where greater focus is required. Large-scale human rights events, such as attacks on a civilian population, or the razing of villages, can sometimes be captured using geospatial data captured by satellites. Before- and after-event analysis can shed light on the event, as well as clues about the perpetrators. Modern mobile phones come with geographic sensors and can be used to collect location data, though these may have limited accuracy compared to dedicated GPS receivers. Digital shape files are increasingly available, including from crowdsourced projects like OpenStreetMap, as well as high quality [digital satellite imagery](#).

## Government Data

---

Governments are sharing great quantities of data online, and making them accessible via Freedom of Information or other “sunshine” requests. And while there are inherent difficulties in getting a complete or unbiased data set of anything, it is particularly challenging when it is in a government’s self-interest to hide abuses and obstruct accountability. Marginalized groups may be excluded or hidden from, or under-counted by, the available information as a result of implicit bias, or even by design.

Data about economic and social rights may seem easier to gather since there is a plethora of official data in most countries about education, housing, water, and other core rights. This data is not designed to assess rights, however, meaning that it is at best proxy data for rights fulfillment. Even when there are flaws in the data collection or the data itself, however, the results can sometimes be useful to researchers and rights advocates. For instance, if the methodology for gathering data is consistent year after year, one may be able to draw certain types of conclusions about trends in respect for rights over time. If the data in question was collected by a government agency, it may be strategic for activists to lobby the government using its own data despite the flaws it contains—since such a strategy makes the conclusions that much harder to refute.

## Interviews

---

Much data in the human rights field is gathered through interviews. Interviews can be conducted in person, over the telephone, or via the internet. Interviews can be structured, semi-structured, or unstructured (informal), and usually gather mostly qualitative data. Structured interviews involve pre-set questions, administered in the same order for each subject (similar to a questionnaire). Unstructured interviews ask each subject different questions, with the order varying, and some questions depending on previous answers or the subjects’ knowledge and experiences. Interviews often take place as a conversation, allowing the researcher to probe further where new information comes to light or more in-depth exploration is required. In-person interviews can also respond to non-verbal cues such as facial expressions, tone of voice, and body language. In-person interviews can be time consuming to conduct and interpret. Information collected may not be comparable over a number of participants, especially when an unstructured approach is used. With structured or semi-structured interview, quantitative data can be drawn from the interview text (*e.g.*, “50% of participants reported that they experienced discrimination in the past year”). Secondary quantitative data can be used to contextualize interviews; examples might include a timeline of events or a map describing where the interviews took place.



## Measurement

---

Measurement is the process of quantifying concepts, phenomena, events, or physical properties using standard units.

In the human rights context, measurement often takes place through:

- counting specific observable events (so-called “events-based data,” such as the enumeration of extrajudicial killings);
- coding qualitative information about the status of rights in a specific place for a given period to produce a numerical score (so-called “standards-based measures,” such as the [Political Terror Scale](#) project);
- collecting survey-based measures of people’s experiences or opinions about rights-related issues; and
- using socio-economic and administrative data concerning outcomes or services, and conducting forensic investigations, including physical and chemical measurements. The Physicians for Human Rights [international forensic program](#) has mobilized forensic scientists and other experts worldwide to conduct autopsies and other forensic investigations; DNA analysis has also been used for human rights advocacy, particularly in [exonerations from the death penalty](#).

## Perception & Opinion Surveys

---

Perception and opinion surveys are useful in trying to collect data about people’s subjective experiences, thoughts, feelings, or beliefs regarding a particular issue. They are used to assess needs, analyze trends, develop solutions, and establish baselines. Because individuals differ vastly in the way they experience and view the world, these surveys are useful in collecting data that can otherwise not be directly quantified. This is especially important in the human rights context, where subjective experiences of discrimination, marginalization, or stigma are particularly relevant. An example is the [Afrobarometer](#), a comparative series of public perception surveys measuring citizen attitudes to democracy and governance, markets, and civil society in more than 30 countries in Africa. Transparency International’s [Corruption Perception Index](#) is also drawn from survey data. Many countries collect survey information about individuals’ experiences of discrimination based on race, ethnicity, sexual orientation, or other axes of discrimination. Perception and opinion surveys can be conducted in person, over the telephone, or online. However, when administering such a survey, it is essential that the confidentiality of subjects’ responses be guaranteed so the participants feel comfortable expressing their true feelings. Survey results may be biased if individuals do not feel comfortable expressing themselves.

## Policy Analysis

---

Related to document review is policy analysis. Policy commitments, contents and interpretations, and processes can all be examined.

Studying changes in policy over time can identify laws that successfully protect and fulfill human rights, as well as laws that are harmful.

The existence and implementation of policies on a particular subject can be examined with respect to international human rights norms and commitments.

Policies can also be compared across states and mapped against different effects and outcomes.

The [UN Treaty database](#) is one starting point for state obligations. Comparative databases like the [Global Gender Equality Constitutional Database](#) also facilitate tracking of government commitments.

## Random Sample

---

Unlike a census, where data is collected from every member of the population, a random sample involves collecting data from a small, representative group from within the population. In a purely random sample, every unit of the population has an equal chance of being selected, removing bias from the selection procedure.

To conduct a random sample, a population is first defined, as well as a target sample size. Units of the population are then chosen at random.

Because the selection is random, the sample is assumed to be representative of the population, and the information collected can be used to develop inferences about the whole population. Random samples can effectively be used to gain insights into a very large population, where a census is impossible to conduct. However, conducting a truly random sample may be challenging where the population is large, dispersed, or hidden. Furthermore, random sampling may not be feasible or ethical when dealing with populations that are marginalized or criminalized. If not conducted properly, random sampling may be subject to bias.

The [Human Rights Data Analysis Group](#) has used random sampling to conduct a number of large-scale, post-conflict body counts.

## Socio-Economic & Administrative Statistics

---

Socio-economic and administrative data describes humans and their activities. This could include demographic data (age, sex, race, etc.), economic data (income levels, employment rates, occupation, etc.), administrative data (collected by health, water, or sanitation service providers), housing related data, migration trends, public health indicators, and so on. This data is often drawn from government statistics and administrative records, survey statistics, or census results.

Data is often available over time, displaying changes and trends in various indicators, which can also serve as a tool to show progress or decline. This is particularly useful for showing the progressive realization of economic, social, and cultural rights.

Often socio-economic and administrative data can be disaggregated by axes of discrimination to show disproportionate impact of government action or inaction on particular populations.

Much of this data is published online and may be easy to access, providing a range of useful information that would otherwise be labor-intensive to gather.

In choosing your data set, remember that socio-economic and administrative data may be out of date, or biased in its methodology.

## Streaming Data

---

Data streaming is the real-time transfer of data at a high speed, from one place to another through a network. Data streams are transmitted through the internet and can be processed and visualized to provide real-time insights.

A number of websites and social media services provide interfaces to access real-time data, which may be monitored by NGOs, for instance, during large-scale events like elections.

Other types of live, streaming data include weather and climate data, internet searches, and flight tracking data, as well as CCTV video and some satellite imagery.

## Survey of Experts

---

Data based on expert judgment involves information collected from well-informed and highly reputed experts in a given subject area. Experts may include activists, lawyers, or researchers. Data is usually gathered through interviews, but a self-administered questionnaire may also be used.

The sample of experts is usually small, but the data gathered can provide a variety of useful information. Data is relatively easy to gather, due to the small sample size and the general willingness of experts in the area to help with further studies. However, data may be biased or unrepresentative due to the small sample.

The [Human Rights Measurement Initiative](#) uses a calibrated survey of experts to assess civil and political rights across different countries.

## United Nations Indicators

---

The United Nations collects and publishes a variety of social and administrative indicators that can be accessed via the web. These indicators cover a wide range of subject areas and are compiled by the [Statistics Division](#), and the Department of Economic and Social Affairs of the United Nations Secretariat, from a variety of governmental and non-governmental sources. The indicators cover areas such as population size, growth and distribution, health, life expectancy and mortality rates, education and literacy rates, income levels, poverty and employment, and so on. Some of these data sets are updated regularly and can be a useful tool for researchers, serving as a starting point to develop a baseline and context for further research.

The Sustainable Development Goals will be monitored using [232 indicators](#). Many efforts are underway to track progress against these indicators; watch the [SDG Knowledge Hub](#) for the latest information. The World Bank has created [dashboards for its indicators associated with the SDGs](#).

As UN data is often drawn from a range of national statistics offices, these may use different methodologies or may be fraught with different types of bias. Some data, like the [UNICEF MICS](#), use the same methodology year over year and are more directly comparable.

## Additional Resources

---

[DatNav: A Guide to Navigate and Integrate Digital Data into Human Rights Research](#), published by the Engine Room;

[Verification Handbook](#), a definitive guide to verifying digital content for emergency coverage;

[HURIDOCS Manuals](#), guidelines for human rights documentation and data standards;

[Kobotoolbox](#), a web-based tool to create field surveys;

[Trifacta Wrangler](#), a desktop tool for data cleaning and processing; and

[Dataproofer](#), a tool that automates the process of checking a dataset for errors or potential mistakes.

## About this Booklet

---

This activity is the result of a research collaboration between the Center for Human Rights and Global Justice and Tandon School of Engineering at New York University, and was funded with a grant from the John D. and Catherine T. MacArthur Foundation.

The booklet was authored by John Emerson and Margaret Satterthwaite with help from contributors Brianne Cuffe and Sidra Mahfooz.

It was inspired by Shiqing He and Eytan Adar's [Vizit cards, The Data Visualisation Catalogue](#) by Severino Ribecca, and Tamara Munzner's [Nested Model for Visualization Design and Validation](#).

For more information about data visualization and human rights along with links to resources, research and tools, visit our project page at <http://visualizingrights.org>.

Cover illustrations by [John R. Holmes](#) for the Open Society Justice Initiative report [Presumption of Guilt: The Global Overuse of Pretrial Detention](#) (2014.)

Version 0.2, published December 2017. The text of this booklet is licensed under a [Creative Commons Attribution-Share Alike 4.0 License](#). Images are not covered by this license.

Please send suggestions, comments, or feedback to [john@backspace.com](mailto:john@backspace.com).