COMMUNICATING DATA FOR IMPACT

Peter Speyer
Chief Data and Technology Officer
Institute for Health Metrics and Evaluation (IHME)

Brian Pagels
Chief Impact Officer
Forum One

Nam-ho Park
Managing Director, West Coast
Forum One

Table of Contents

Introduction

The year was 1854. On the Crimean peninsula, British and French soldiers were engaged with Russia in a battle over religion and territory. Soldiers were dying in droves.

Across the Black Sea, a young yet accomplished British nurse volunteered to serve in the military hospitals in Turkey, where injured soldiers were transported for care. When she arrived, the sanitary conditions in the hospitals were horrendous—as was the manner in which the hospitals were collecting data about their patients.
The young nurse took notice, and improved both during her tenure. Then she did something truly remarkable.

Florence Nightingale
Source: **Wikipedia**

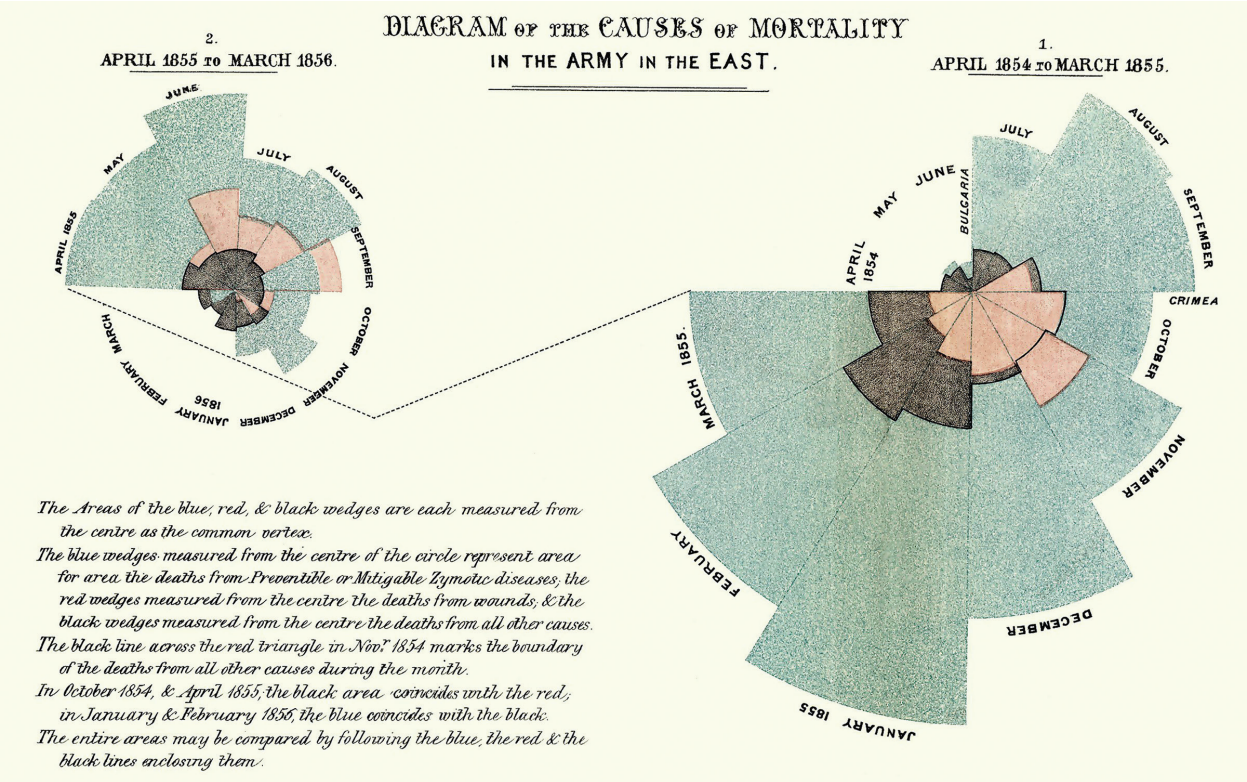
In the mid-19th century, data storytelling and public policy converged. Florence Nightingale – Britain's beloved "Lady with the Lamp" and the young nurse from the story – saved many more lives with her data analysis and storytelling than she could have saved alone as a nurse. Nightingale believed it was her religious imperative to study statistics. "To understand God's thoughts," she said, "we must study statistics for these are the measure of His purpose."

Following her service in the Crimean War, Nightingale partnered with the accomplished statistician William Farr to analyze the hospital mortality data. The numbers were clear: more soldiers died from diseases spread through poor sanitary conditions than from combat wounds. Nightingale and Farr published these findings in the behemoth 830-page Notes on Matters Affecting Health, Efficiency, and Hospital Administration of the British Army. Farr would have been willing to stop there, but Nightingale was concerned that the dense report and data tables alone wouldn't move Queen Victoria or the British Army to improve sanitary conditions in military hospitals.

"We do not want impressions, we want facts…You complain that your report would be dry. The dryer [sic] the better. Statistics should be the driest [sic] of all reading," Farr warned her.

Nightingale was aware of emerging data visualization techniques. Scottish economist and Renaissance man William Playfair invented the line graph and bar chart in the late 18th century, and then the pie chart and circle graph in the early 19th century. Nightingale knew she could leverage these methods to illustrate her findings.

So she ignored Farr's advice. Instead, Nightingale created polar area diagrams – often referred to as Nightingale's "Rose" or "Coxcombs" – which used colored wedges to represent causes of death in the army during each month of the war.



Florence Nightingale, "Diagram of the causes of mortality in the army in the East" was published in Notes on Matters Affecting the Health, Efficiency, and Hospital Administration of the British Army and sent to Queen Victoria in 1858. Source: Wikipedia

Nightingale said her intention was "to affect thro' the Eyes what we fail to convey to the public through their word-proof ears."

Her report and graphics had a significant impact on military hospital practices and conditions. By considering her audiences and selecting the best data visualizations to affect change, Nightingale became one of the earliest effective data communicators.

Just like Florence Nightingale, data owners who want to make an impact must consider their intended audiences and tailor the data communication strategy accordingly. Today, more data are being collected, scrubbed, analyzed, and shared than ever before. But the sheer amount of data creates its share of problems for those who wish to improve conditions in the world.

What data are available to highlight or help solve societal issues? Which data and sources should we trust? And once we've identified the right data, how do we effectively communicate them to those in a position to influence public policy?

Challenges in Communicating Data

Communicating data and maximizing impact is about supplying the right audience with the right amount of data in the right format. And in order to increase the likelihood that a target audience will pay attention, one must consider the appropriate timing and channel(s) for delivery of the data.

Nightingale's and Farr's 830-page tome was a massive achievement and found attentive readers. However, to drive change, Nightingale targeted very specific audiences: the Queen and the leadership of the Royal Army. And she needed to provide them with the data in a way that would capture their attention. Nightingale's "coxcomb" visualizations combined those data in easy-to-understand graphs, and delivered a clear message.

"Communicating data and maximizing impact is about supplying the right audience with the right amount of data in the right format."

Any individual or organization providing data to audiences has the opportunity to decide how to share their data. The larger the dataset, the more options there are to create impact (and the more decisions must be made to provide audiences with the right data.).

For some audiences, access to all available data is crucial. However, this also means users must navigate vast amounts of information, which puts most of the burden on them. For other audiences, like news consumers, fewer, punchier numbers are best. Vast amounts of data can be condensed to one key illustrative number or visual via aggregation and filtering, and the onus of picking and creating this one number is on the datapoint provider.

Those are two extremes. The best ways to provide the data to a given audience will probably be somewhere in between, ranging from API access to the full database, sophisticated or basic query tools, interactive visualizations, reports, infographics, and many more.

Identifying the Right Audience

To maximize impact, data providers must identify audiences they need to address, and then provide them with the right amount and granularity of data.

They must then choose the appropriate data products and tools for those audiences, which range from complex interactive visualizations or query tools to simple graphs or illustrations. These data products can then be packaged and delivered as part of scientific books, policy reports, press releases, blog posts, email newsletters, social media outlets, and more.

The table on the following page provides an overview of key audiences, along with relevant data and tools for these audiences. For each of the audience "types" below, we selected a role that best represents that audience's data-related needs.

Type	Description	Audience examples	Data	Products & tools
The Casual User	Interested individuals take in the data. If the data trigger action, these individuals may move into a more active role. They typically have little data and domain knowledge.	Anyone with a casual interest in the topic.	Specific data points, trends, developments	Infographics, declarative/narrative visualizations, illustrative diagrams
The Data Actor	Data actors act on and leverage the data to drive change. They have significant domain knowledge, but often only limited time.	Policy and decision makers in governments, nonprofits, and corporations	Curated datasets, e.g. by topic, country/region	Press releases, reports/briefs, limited interactive visualizations, search data tools
	Other data actors may also be looking for interesting stories or data to back up a story. They need to review information quickly, and often need a specific figure.	Journalists, bloggers, advocates	See above	See above
The Analyst	Analysts use data to create deeper understanding, while informing data actors and consumers. They have a deep domain knowledge, extensive data knowledge, and will review and condense large amounts of data for a given topic into reports, presentations, apps, etc. .	Domain experts at int'l, national, local levels; often staffers for decisions makers	Comprehensive datasets showing global trends, data by topic, country etc.	Query tools, exploratory visualizations
		Web/software developers, "data geeks"	Full database	Application programming interface (API)
The Researcher	Researchers work in the trenches to collect, analyze, and synthesize data for the groups above; they could have done this data collection / analysis themselves.	Researchers, academics, analysts, modelers	Full database; source data & methods (input for dataset)	Query tools, exploratory visualizations, data catalogue, data repository, visuals to explain the methods

Understanding Audience Needs

When releasing data for public consumption, data providers should be prepared to make a few upfront decisions. One fundamental question must be asked: "How much data should I present to the user?"

The above table provides a high-level illustration of the range of data needs, related products, and desired levels of interactivity, all based on audience characteristics.

To see this in action, consider a policy maker interested in health. Perhaps she's looking to improve health outcomes and lower the cost of the delivery of health care services for individuals with complex needs. Her data format and delivery needs likely differ greatly from those of a web developer looking to create a map that highlights differences in health care delivery costs across the country (which this policy maker may in fact find useful).

"One fundamental question must be asked: "How much data should I present to the user?"

With the big picture already considered, let's take a deeper dive into the audiences highlighted above. We will look at each of the audience groups in turn, and some overlap between these audiences will emerge. We'll also provide concrete examples from a recently published study of health patterns around the world, the Global Burden of Disease study.

Global Burden of Disease (GBD)

Health data are collected in many places, including censuses, surveys, vital registration systems and registries run by governments, health records, claims data, administrative data, and scientific literature. To understand the state of health around the world, we must identify, access, compile, analyze, and evaluate all these data comprehensively—a seemingly insurmountable challenge.

In the early 1990s, Professors **Christopher Murray** and **Alan Lopez** accepted that challenge. Called Global Burden of Disease (GBD), their approach measures the impact of diseases, injuries, and risk factors that shorten lives or create health loss through short- or long-term disabilities.

Three metrics are used to measure health loss:

- Years of Life Lost (YLL) due to premature mortality
- Years Lived with Disability (YLDs)
- Disability-Adjusted Life Year (DALY)

In 2012, the Institute for Health Metrics and Evaluation (IHME), under the leadership of Chris Murray, published the results of the **GBD 2010 Study** —a complete revision done in collaboration with 488 co-authors from 50 countries around the world.

Based on all health data available to the researchers, GBD 2010 covers 187 countries and provides data for 291 diseases and injuries, and 67 risk factors. 3 different metrics (YLLs, YLDs and DALYs) are available as numbers, rates, or percentage, by age (20 age groups) and sex, for 1990, 2005 and 2010, with uncertainty intervals.

The results dataset has more than 1 billion data points. There were a number of significant challenges in creating this dataset, from identifying and accessing data, to cleaning and preparing data for analysis, to the actual analysis and evaluation. IHME built out a 3,000-node computer cluster to facilitate the analysis, and even then, it took about 1 week to run the analytics from beginning to end.

The Casual User

The general—or perhaps interested and informed—public sits at the top of the table. They don't necessarily have domain expertise with an issue, but might be passionate about one or more topics. If they're not initially interested, data providers must pique their interest, or even motivate them to act.

If, say, a member of this audience wants to donate to the international relief organization with the best track record, he'll be looking for key data points and trends that give him answers and better insight into the problem at hand. These data points are best delivered as numbers, or as infographics telling a simple data story.

If they're not initially interested, data providers must pique their interest, or even motivate them to act.

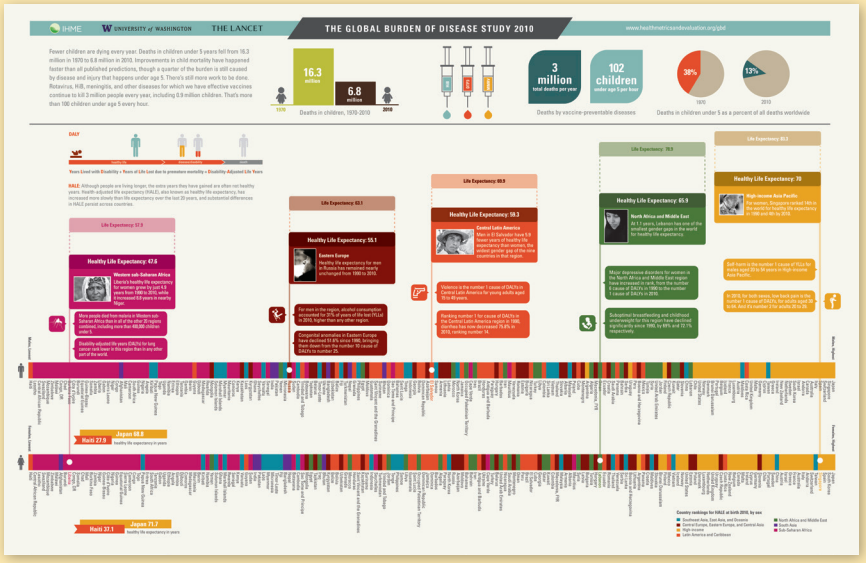
Selecting one number to focus on could require tremendous effort from the data provider. Despite the challenge, and although the user experience isn't exactly interactive, boiling a complex issue down to one number can create a powerful and elegant statement.

If the data story requires more than one number, data providers and analysts have many options for visualizing their data. These techniques include: tables, bar charts, tree maps, dashboards, cartograms, etc.

Infographics may include one or more of these graphical approaches to represent data, and can be either static or interactive.

GBD Results for Data Consumers

As we've discussed, data geared toward general public should be limited to key data points, trends, or stories. While some visualizations aimed at data actors may interest the general public, IHME created a higher-level visualization with key summaries per country, called GBD Insight. And a printed infographic, **the GBD 2010 Poster**, appeared in Lancet, providing some key metrics, a country ranking by life expectancy, and more details on a few example countries.



GBD 2010 Poster.
Source: IHME

The Data Actor

This is a crucial audience that includes legislators, ministers, nonprofit leaders, business executives, etc. Since they're in the position of influencing and often creating/refining policy, their data needs are more complex than those of the general public. However, policy makers are not likely to spend hours wading through reams of data to find what they need. Instead, they'll look for packaged briefs with corresponding figures they can use in a speech, or maybe a curated dataset or visualization that only requires selecting a few filters or settings to get the information they're looking for.

In recent years, innovators - inspired by the visual storytelling of leading online news media - have made the traditional paper or PDF briefs more interactive and accessible. Modern briefs will include text, pictures, videos, statistics or interactive graphs, and other contacts. The objective, however, remains the same: provide the most relevant data in a way that makes best use of a policy or decision makers' time.

Policy makers are not likely to spend hours wading through reams of data to find what they need.

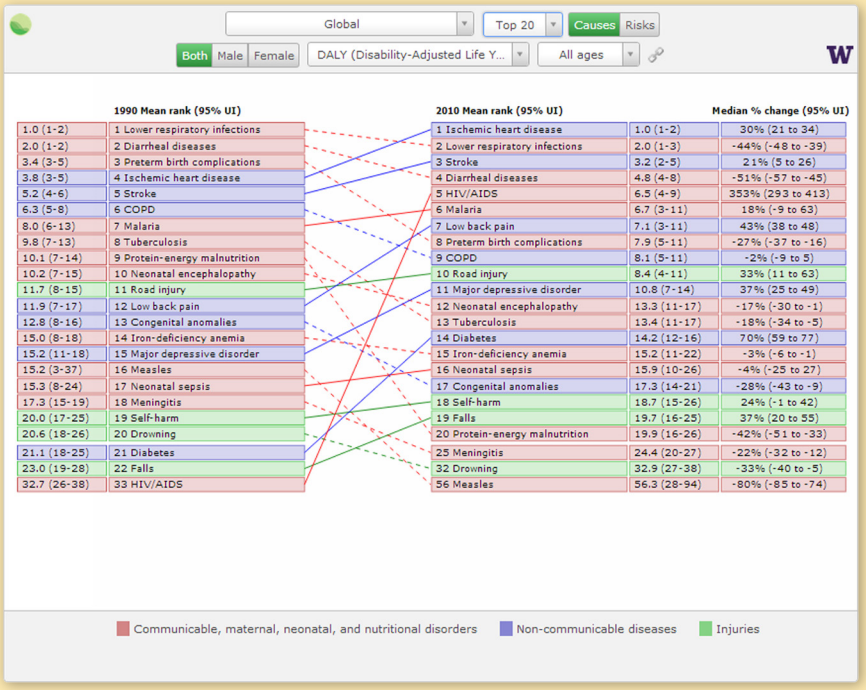
Most journalists, advocates, and bloggers often need data at about the same level as policy makers, but are **more likely to use an interactive tool** to find a unique data story to highlight.

Providers might approach interactivity in a few different ways. For instance, interactive visualizations allow users to review a specific story and further drill into the dataset. Alternatively, data query tools can allow users to find individual data points or create small tables. They can provide access to the data with curated filters or guides to point users in the right direction and allow them to easily retrieve relevant data.

GBD Results for Data Actors

IHME has created several tools that allow exploration of data at a higher, global level to give decision makers access to the specific data-supported conclusions they need to affect change. Most notably, **GBD Cause Patterns** shows a condensed view of 21 broad cause groups, and allows review of patterns by age, geography, sex, and chronology. **GBD Arrow Diagram** allows a quick comparison of disease and risk factor rankings in 1990 and 2010. It helps users look at the big picture and identify key priorities, based on size or growth of burden. Users can then drill down by country, age, and sex to see what populations are most affected. More examples of visualizations can be found on the **GBD Visualizations page**.

In addition, a **query tool** makes it easy for users to find individual data points or small tables with the information they choose. The information is then displayed as a simple visual and made available for download.



GDP Arrow Diagram.
Source: IHME

But people still appreciate paper. In addition to the online tools, IHME created a **GBD Policy Report** that was provided in printed form and for download online. Given the tremendous demand for this kind of information, the report was followed up with **regional editions** in collaboration with the World Bank, as well as a report for EU and EFTA in collaboration with the European Union.

The Analyst

Staff who work for data actors at NGOs, in government, or for corporations are often domain experts in a particular field. Their work requires data analysis in order to plan or evaluate programs and policies.

Analysts need access to filtered datasets for the topic or geographic region in which they operate. Some of this group may wish to export a portion of a dataset to run their own analyses, while others are content to use interactive tools provided through the web interface.

Analysts need access to filtered datasets for the topic or geographic region in which they operate.

Web or software developers and "data geeks" also need full access to the database, but they often seek Application Programming Interfaces (APIs) to more easily integrate data into existing applications, or to create entirely new applications with the data. APIs allow developers to tap into a database and use it to create their own applications, or combine datasets via multiple APIs, while ensuring the applications are refreshed dynamically when the source is updated. APIs also free them from the burden and cost of storing large datasets.

Developers sometimes serve as data promoters of sorts—their products or services may reach any of the above audiences.

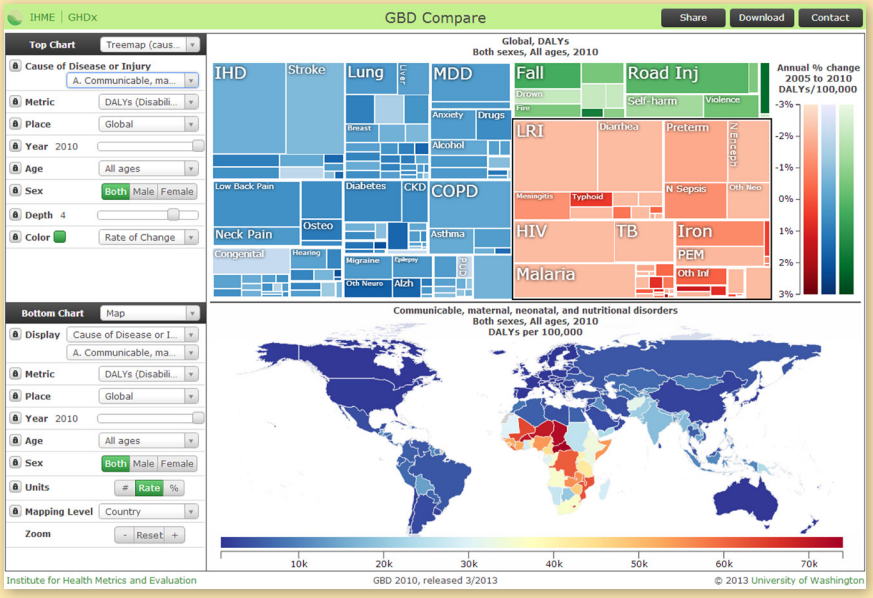
GBD Results for Data Analysts

Experts and analysts understand global health, specific diseases, injuries, and risk factors. They also understand the metrics used to measure them.

Their real interest lies in reviewing the data to find patterns and trends, and to answer questions. Then, they can use the data to plan interventions and programs. Just like researchers, they want lots of detail, but also more intuitive ways to interact with the data.

To satisfy this level of need, IHME built a exploratory data visualization for data experts and data analysts. Called **GBD Compare**, it's IHME's flagship tool for interaction with GBD data. The visualization allows users to review the data in 5 different chart types: treemaps, maps, age and time plots, and stacked bar charts.

There's a 2-panel view with linkage between panels, e.g. clicking on country in the map will set the other chart to that country. Some charts have many different view options; for instance, a treemap illustrating diseases and injuries can filter by trend data, uncertainty bounds, or risk factor attribution.



GBD Compare.
Source: **IHME**

The visuals are complex and require the user to have some prior knowledge of burden of disease measurement. They're powerful for exploring trends, from high-level to great detail. Users can export screenshots as JPGs; they can download the underlying data as a CSV; or they can share a permalink with the specific settings they're using via email or social media.

The Researcher

Researchers and academics almost always want full access to a database in order to analyze the data and create models for their research. They'll often look for the "download" option before spending time exploring an interactive tool. Providing downloadable files in machine readable format for easy importing into statistical analysis tools is a useful mechcanism here.

Researchers may also require background information on underlying data and methodology, including survey instruments, published papers, references to underlying data used in analysis, and similar documentation.

The researcher will often look for the "download" option before spending time exploring an interactive tool.

GBD Results for Researchers

Researchers and other data experts want to review and use the results at the most comprehensive level. They're comfortable working with databases, and would rather manipulate the data themselves. In addition, they need to understand the methods and data used to generate the results. Fully detailed accounts of the GBD methods are available across 8 papers and appendices (together far more than 1000 pages), accessible on the **Lancet website**.

To provide researchers with information about the more than 20,000 datasets used for GBD, the IHME team catalogued the data in the **Global Health Data Exchange (GHDx)** with a link to the data provider, as well as extensive metadata like geography, years, summary, keywords, publisher, and more.

Full results of GBD 2010 are available for download in CSV files in the **GHDx**. Given the size of the results database, the data have to be downloaded by disease, injury, risk factor, or country.

Mortality Visualization.
Source: **IHME**

IHME created the **Mortality Visualization**, which illustrates the starting point for these results. Here, users can review trends and find citations and other metadata for all data points. See also the **COD Visualization** for all data points used for estimating causes of death.

Communicating Data: Cheat Sheet

The table below maps our audience types from above to their attributes as they pertain to data, along with the extent to which these attributes are present in the audiences. As an example, data consumers represent a large audiences size with limited attention span and rudimentary data manipulation skills and domain expertise. Data experts on the other hand represent a small audience, but they are willing to engage with the data (attention span) and have in-depth data skills and topic expertise.

	Audience size	Ability to drive change	Attention	Data manipulation skills	Topic expertise	Example
The interested individual	Large	Low	Low	Low	Low	Infographic
The data actor	Small	High	Low	Medium	Medium	Narrative visualization / briefing
The analyst	Medium	Medium	High	High	High	Detailed topical visualization
The researcher	Small	Medium	High	High	High	Exploratory visualization with all data available

Conclusion: Making Data Impactful

Today, data are being collected at an unprecedented scale, and they exist in many different shapes and formats. The tools for data analysis are becoming ever more sophisticated, and there are more and more powerful tools to present data and engage users. Data are being looked to as an important aspect of communication, as input for decision making, and as a means to evaluate the effectiveness and performance of organizations, programs, projects, policies, etc. But many people and organizations struggle with unlocking the power of data and communicating it in the most effective way.

Florence Nightingale.
Source: Wikipedia

Ultimately, data providers want to inspire action among their audiences. They understand we need to use these data to solve our current set of problems, and increasing the distribution and understanding of data can accelerate the pace of positive change.

Depending on their domain expertise and data skills, individuals in organizations and governments require data in relevant formats to inform these decisions. Despite more data, better tools and more data skills, the fundamental challenge for communicating data has remained the same since Florence Nightingale's time: supplying the right audience with the right amount of data in the right format.

Data curation and decisions about the data delivery mechanism are as important as the data themselves.

Careful curation of the data tailored to the audience it serves determines how that data is purposed for positive social change, whether it be saving lives in hospitals, making informed choices for the health of your family, or crafting public policy that will affect large populations. How we shed light on the data does influence its impact.

GBD Study Conclusion

The GBD Study provides the most comprehensive dataset ever compiled on global health outcomes. Experts have likened the study to the Human Genome project in its importance, and the response to the publication of the GBD Study has been overwhelming. Analysts have worked with or requested custom datasets. Domain experts have engaged with and commented on data reviewed in the visualization tools. Policy makers have requested conversations and workshops to interpret the data. Journalists have used the data for health coverage and for sparking Twitter conversations about specific data points; Wired Magazine turned one of the visuals into a beautifully rendered **infographic**.

Part of the reason for this response was that IHME took a hard look at the GBD results, evaluated the needs of different audiences, and created a suite of tools to share the results of the study. Analyzing the world's health data to generate the GBD results was a monumental effort, but the impact of the study wouldn't have been the same without tools that made the data easy to access and explore for different audiences.

About the Authors

Peter Speyer
Chief Data and Technology Officer
Institute for Health Metrics and Evaluation (IHME)
Peter and his team fuel IHME's research with data from organizations around the world, provide state-of-the-art data management and computational infrastructure, and share IHME's research results with all the different audiences mentioned in this paper.
speyer@uw.edu

Brian Pagels
Chief Impact Officer,
Forum One
Brian helps foundations, government, and nonprofits explore opportunities to manage, visualize, and share data more effectively. He is particularly passionate about domestic health and environmental issues, and has managed data-focused projects for the Robert Wood Johnson Foundation, the AARP, and the EPA, among others.
bpagels@forumone.com

Nam-ho Park
Managing Director, West Coast
Forum One
Nam-ho works in Seattle with organizations focused on global health and development, furthering their missions and influence through digital communications strategy and web technologies.
npark@forumone.com

Special Thanks

Reviewers

Kevin Merritt, President & CEO, Socrata

Noah Illinsky, Visualization & UX Expert, ComplexDiagrams.com

William Heisel, Director of Communications, Institute for Health Metrics and Evaluation

Kurt Voelker, Chief Technology Officer, Forum One

Jim Cashel, Chairman, Forum One

Editor

Kristina Bjoran, Forum One

Designer

Risa Willmeth, Forum One