Make Data Make Sense: The Importance of Visualization in Data Analytics

By Anthony Vanky

"The variables are many, but they are not helter-skelter; they are interrelated into an organic whole," said urbanist Jane Jacobs. Jacobs was referring to Hudson Street in 1960s Manhattan, but the same could be said of any street in any city today.[1] The key difference is that today, big data allows us to describe and analyze Jacobs' "interrelated variables" better than she could have imagined.

In the current big data paradigm, data is king. Every additional layer, dimension, and dataset promises to bring us closer to a "perfect" representation of reality. Visualization provides fertile opportunities for analysis, helping to situate data in social, economic, and political contexts; it also changes the way we understand those contexts. In James Corner's words, mapping (a specific application of visualization) "unfolds potential; it re-makes territory over and over again, each time with new and diverse consequences."[2] By adding context to data, visualization helps us move beyond a focus on finding specific answers and enables us to ask richer questions that ultimately lead to greater insights. Too often, data analysis is focused on achieving narrow outcomes. But how do we know if we are even asking the right questions?

The MIT Senseable City Lab studies how data can reveal underlying patterns of cities and their citizens. With an abundance of data, question-driven exploration can be challenging in general, but this is especially true in the complex environment of cities. The Lab's work is an example of how visualization can be used as part of a research methodology, as we explore how different representations of data can transform how we understand and communicate city dynamics. This article discusses two projects developed at the Senseable City Lab in an effort to show how visualization can be used to support data analysis and to instigate critical innovation in cities.

LIVE Singapore! Asking Better Questions

Singapore is perhaps the city that is most aggressively leveraging technology and computing to transform an urban area into a more livable community. Data is produced by and collected from an increasingly large and diverse set of infrastructures, forming an urban living laboratory. However, the current data ecosystem is segregated. And while segregated data is useful for individual operations, data contains much more value when it is combined from different sources and made accessible to users through creative new applications.

The LIVE Singapore! Project aims to close the feedback loop between citizens and the digital, real-time data collected across the multiple urban infrastructure networks that sustain urban life. The project seeks to develop an open platform for the collection, fusion, and distribution of real-time data that originate from a large number and variety of different sources.

With current capabilities, investigating this data is difficult. There are limited tools that allow for quick data visualization and intuitive analysis of discrete datasets from various spatial and temporal sources. The goal of the LIVE Singapore! project is not only to offer new capabilities, but to inform the analytical and modeling process — to make it question-driven, instead of answer-oriented. Through LIVE Singapore! the MIT Senseable City Lab is creating several visualization tools that will allow researchers and policymakers to quickly and intuitively survey the city's data. Overlaying these different types of data, within the specificities of their contexts, allows for a type of exploration that will shape questions for further analysis.

For example, Singapore's population is heavily reliant on public and shared transit. And due to its location on the equator, tropical rainstorms are often heavy, fast-moving, and short. By viewing the map in Figure 1 — showing rain data overlaid with real-time location and availability of taxis — researchers could quickly form new questions, test hypotheses and gain insight into how the country's transportation infrastructure might be impacted by heavy rains during the wet season. Through this exploratory analysis, we ultimately found that taxis tended to avoid areas with heavy rainstorms despite the ease of picking up a quick fare. We then confirmed this finding analytically and shared it with the Singaporean government for consideration in future policy decisions.

In another study, the Lab investigated correlations between electricity usage and ambient temperature to better understand (and demonstrate) how exhaust from air conditioning increases the urban heat island effect, thus creating a viscous cycle that drives further energy consumption.

Trash Track: Driving Individual Behavior

In Trash Track, the Lab studied refuse data from Seattle to explore the potential of making an otherwise invisible infrastructure — our waste infrastructure — visible. Due to complex sorting processes, regulatory environments,

and fragmented ecosystems of waste handers, it is difficult to generate a global understanding of the waste stream. While many industries thoroughly understand the processes by which their products are created, surprisingly little is known about what happens when we discard an object. As a response, Trash Track can be viewed as an inversion of the supply chain by examining the "removal chain."

A dramatic drop in the cost of sensors is creating a new paradigm in technological development — the Internet of Things is connecting objects of everyday life to the Internet, enabling these objects to contribute their data to tell their own stories. Researchers in the Lab leveraged the dramatic drop in the cost of sensors in order to gain insight into this otherwise opaque removal chain.

The researchers designed and placed small sensors — including GPS sensors that communicated over a cellular phone network — into the refuse of 500 volunteers from Seattle. Residents disposed of the tracked items as they normally would, allowing the researchers to record the movement of various types of household waste. While a bulk of the waste went to local disposal and recycling centers, over time, the waste footprint grew to span the breadth of the continent. From Washington to Georgia, trash traveled to the most efficient — measured by cost or by regulation — resting place.

While the data itself is powerful in providing the first characterization of this removal chain, the map in Figure 2 reveals the complexity of waste infrastructure. Individuals can visually trace objects as they travel from city to city, state to state, or in one case from Seattle to the Midwest, only to come back to the Pacific Coast. Without the visual map, some of these paths would remain hidden from view and awareness.

Implicit in the act of visualizing is the desire to transform abstract data into relatable stories that engage and inform conversation. The true power of this visualization is how it can incite change. By seeing her own personal waste map, one volunteer decided to swear off bottled water after seeing how quick decisions made out of convenience led to long-lived negative consequences — plastic bottles sitting for eternity in a landfill.

The Promise of Visualization

Visualization allows for more equitable conversations about data — therein lies its power and vulnerability. Opinions and beliefs can be changed and swayed by

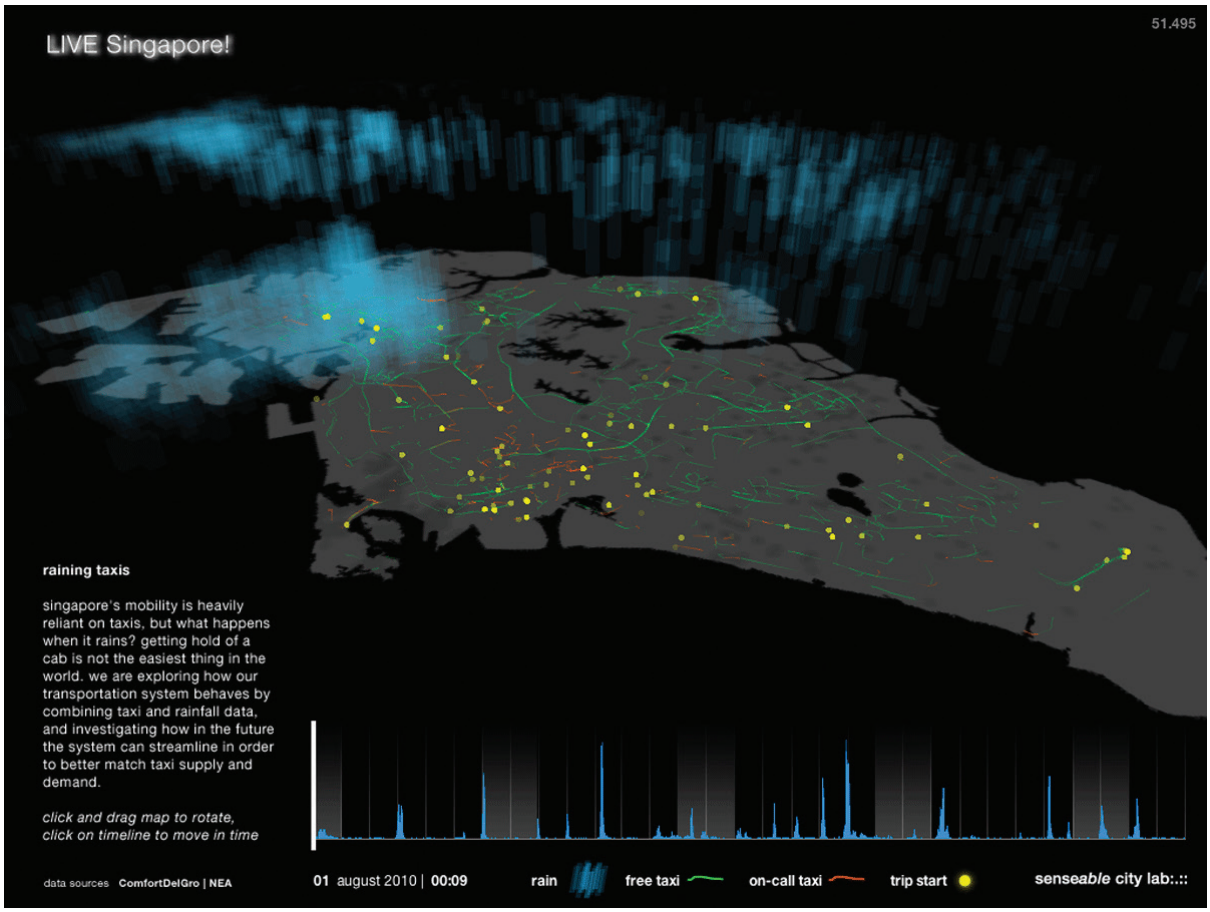

Figure 1 | The LIVE Singapore! project shows how visualizing different types of data in context can lead to innovative ideas about city operations.

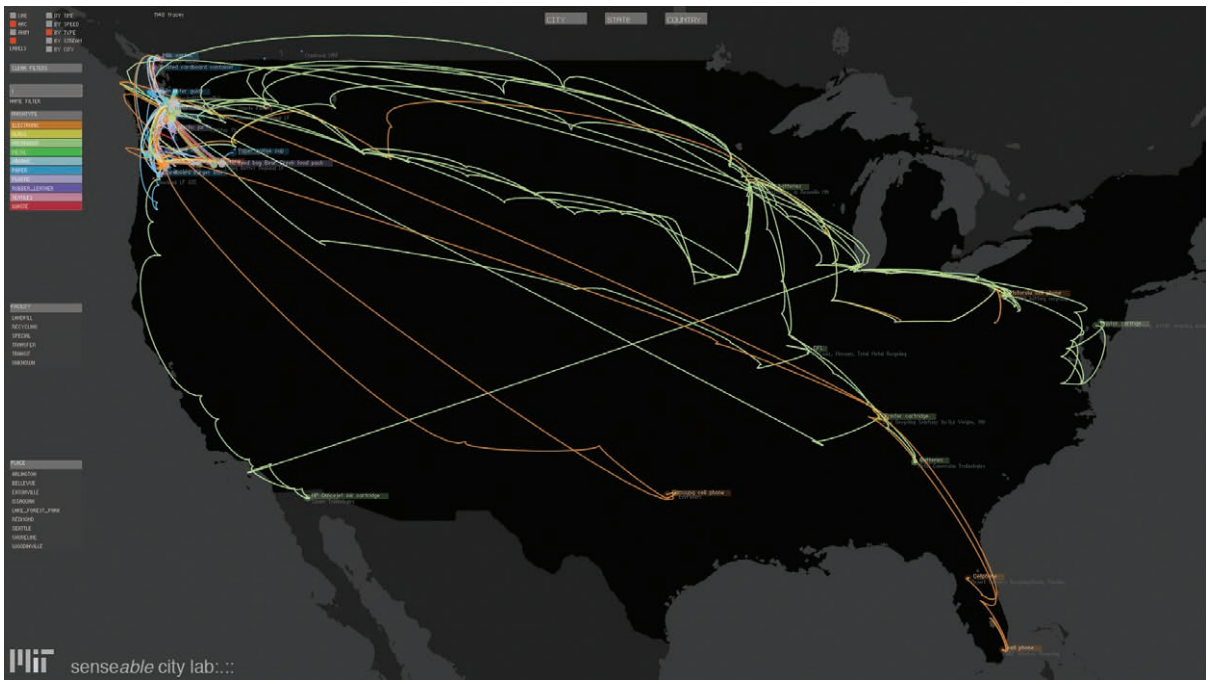

Figure 2 | The Trash Track project visualizes the "removal chain" of waste infrastructure.

these representations, but it is important to keep in mind that the audience is often at a disadvantage, lacking expert knowledge about the topic at hand. Further, as each layer, dimension, and dataset bring a more robust representation of reality, these representations are still imperfect, rife with problems and biases.

The simulated reality is never perfect, but representations of simulated reality — of data — still impact decisions about society, sometimes in negative ways. In the 1970s, New York City burned because of imprecise models that were intended to save the city millions of dollars per year by making the fire department more efficient. The promise of a streamlined, non-partisan technocracy — informed and supported by analysts and modelers — led to lost livelihoods for many of its citizens. However, the possibility of negative outcomes shouldn't prevent us from exploring the opportunities to be had from data. In a way, the current data paradigm is like the weather — despite general distrust of the weatherman, we all listen to the forecast because it's better than nothing.

In some ways, research and design tests the many potential futures that lay ahead of us. Herbert Simon, the noted scientist, believed that design was a unique process as it is concerned not with how things are, but with how the world could be. But the testing of these visions requires dialogue with a broader audience. Without accessible means to do so, these potential futures are left to be forgotten, consensus always out of reach.

By evolving how we use and share data, a new generation of informed decision-makers will be able to transform our communities. Armed with accessible visual representations of abstract data, we will understand more about the world in which we live. Q

Anthony Vanky *is a researcher and former partner strategist at the Massachusetts Institute of Technology (MIT) Senseable City Lab and a Ph.D. candidate in Urban Studies and Planning at MIT. His research considers the use of new, pervasive sensing technologies in measuring and evaluating the built environment. He has widely presented on topics of technology and urbanism, including at EmTech, Harvard University, Skanska Worldwide Research & Development, the Fraunhofer Institute, and the New York Center for Architecture. Trained as an urban designer and architect, he has worked on projects across the U.S., and his design work has been exhibited widely, including at the Venice Biennale, the Dutch Design Week, the Gwangju Design Biennale, and New Orleans DesCours. He previously served in national leadership capacities on each of the five governing organization of the architecture profession in the U.S., including the American Institute of Architects (AIA), and has launched two non-profit advocacy organizations.*

REFERENCES

[1] Jacobs, J. (1969). *The economy of cities*. New York: Random House.
[2] Corner, J. (1999). The Agency of Mapping: Speculation, Critique and Invention. In D. Cosgrove (Ed.), *Mappings* (pp. 213-252). London: Reaktion.
[3] Simon, Herbert A. 1996. *The Sciences of the Artificial - 3rd Edition*. The MIT Press.