

Module 5a: Data Analytical Thinking – making the case for data-driven initiatives

Welcome.

In the previous module, we focused on how to observe and talk to humans to understand the problem better. Here, we turn to looking at how quantitative methods – analyzing data – are also essential to understanding the problem.

By the time you finish this module you will be able to:

1. Describe how data can be used to help solve public problems
2. Define a process for applying data analytical thinking, and
3. Identify key risks and mitigations when using data to solve public problems.

In an effort to reduce the city's murder rate, Mitch Landrieu, then Mayor of New Orleans created a unit in city government dubbed the innovation team or i-Team to tackle the problem.

Using data on murder rate, crime rate, educational attainment, unemployment rate, and recidivism rate organized by neighborhood and dating back to 1960, the team uncovered a significant correlation between unemployment and violent crime (and thus recidivism).

The data showed that a majority of murders in New Orleans were committed by a small and identifiable set of people in a few neighborhoods as the result of petty disputes.

Although New Orleans still suffers from one of the highest homicide rates in the nation, the Team's work led to a decline in the murder rate by 20% between 2012 and 2013, and the number continues to drop. In 2018, murders in New Orleans were at their lowest level in almost fifty years.

Similarly, in Boston, the city was facing the problem of subpar response times by first responders. The median response time for the most serious cases rose from 5.5 minutes in 2009 to 6.8 minutes in 2015 attributed to an increase in call volume.

On average, Boston EMS responds to more than 126,000 incidents a year across the city, marking a 20 percent increase over the last 10 years. In response, Mayor Walsh launched a data-driven initiative to change the way ambulances are deployed and improve ambulance logistics and utilization. As in New Orleans, public officials looked at the available data to understand the problem and its root causes. They analyzed types of emergencies, routes and response times using 911 calls.

Adding the element of quantitative research to policy making practices has allowed these cities to improve outcomes dramatically. These two examples are among thousands of recent stories of how governments are turning to quantitative research and mining newly available data to improve how services are delivered.

They are representative of a major push in recent years toward embracing evidence-based decision-making, including the Commission on Evidence-based Policymaking established by

Congress. In 2018, this Commission recommended changes to strengthen government's ability to make use of its own data while also ensuring stringent privacy protections.

The adoption of evidence-based practices – and, in particular successes like those in New Orleans and Boston – reflect a new way of working. But they depend on more than just using numbers. In both cases, responsible officials did more than look at the data.

They combined human-centered design and participatory approaches with data-driven approaches to develop a more robust evidence base to define the problem.

In Boston, the analysts rode along with the paramedics. The observations and interviews they did together with the data modeling led to a better diagnosis and the creation of an Emergency Management System Community Assistance Team that works with substance abuse and outreach efforts to connect people in crisis to social services.

In New Orleans, the iTeam sat down with law enforcement, social services agencies, educators and, above all, residents themselves. The combination of careful data analysis and human-centered design activities, during which officials talked to and worked with the public, improved the impact of the city's social services on violent crime. Municipal agencies were able to institute various programs to train and hire ex-offenders in an effort to reduce the likelihood of the formerly incarcerated reoffending.

The key learning is that people are still crucial in data analytical processes: the people who collect the data; the people who interpret the data; and the people whose insights into the data matter at least as much as insights from machines and algorithms.

Too many projects are handed off to data scientists and statisticians. While data is extraordinarily helpful and we can make inferences from the right data that lead to genuine insights, collaboration among people remains core to the process.

And, as with any social science problem, defining problems effectively requires us to use both quantitative and qualitative techniques.

Let's get started.

We have always collected information. But digitization has made it dramatically easier to collect, analyze and use information by more and more diverse people with or without extensive technical skills. This is improving our ability to solve problems in a number of ways.

When New York City data scientist Ben Wellington explains his work in analyzing taxi cab data, he uses it to tell a story that reveals an important insight about “rush hour” in Manhattan.

In New York City, the City collects data about each taxi trip, including start and end time and location, fare and tip. Just after 5 am cabs are going an average of 24 miles an hour. But by 8:30 in the morning they are going an average of 11 mph and the speed stays that way all day.

So there is in fact no “rush hour” - there is a “rush day”.

Similarly, if we want to tell an insightful story about the conditions of refugees, we can look at mobile phone data.

One of the hardest problems with aiding refugees is gaining real-time insight.

Enter technology. We no longer have to rely on professional inspectors slowly collecting information face-to-face.

The UN World Food Programme, which provides food assistance to eighty million people each year, together with Nielsen is conducting mobile phone surveys in 15 countries (with plans to expand to thirty), asking people by voice and text about their consumption habits.

The mobile Vulnerability Analysis and Mapping project uses using text messages, to ask people questions like: what did you eat today? Formerly blank maps are now filled in with information provided quickly and directly by the most affected people, making it easier to prioritize the allocation of resources.

With data we can also spot mistakes, outliers and rare events. It is not necessary for data to be a representative sample of the population. If we are looking to generalize about a group based on a smaller group then, yes, we might need a representative sample. But if the goal is simply to measure and spot problems in a given group than nonrepresentative data may be wholly adequate to the task.

By analyzing some of the New York City taxi data, Wellington was able to identify a radical disparity between two different types of cabs. One taxi company's on-board computer was calculating drivers' tips based on the base fare of the ride while others were calculating the tip based on the gross, including fees and tolls, resulting in inequality between drivers.

Ultimately, the goal of using data is to craft better policies, design more effective services and solve problems by virtue of having a more precise definition of the problem. With data, we can save money.

For example, data can help to target scarce enforcement resources more effectively. Chicago has more than 15,000 food establishments but only three dozen inspectors. Chicago's city government used its data on restaurant inspections to create an algorithm to predict food-safety violations. This project increased the effectiveness of its inspections by 25%.

With data, we can also formulate a plan of attack that is more likely to solve the problem because it addresses the root cause.

For example, travel patterns of rats are virtually unpredictable, making rat infestations in large cities difficult to tackle. In 2011, Chicago saw a peak in its rodent problem after receiving over 25,000 rodent complaints via its 311 hotline in that year alone.

This call center data generates a novel database that offers a deeper understanding of both the day-to-day patterns and long-term trends of the city and its neighborhoods.

In search of a new strategy, the City of Chicago partnered with then Carnegie Mellon University's Event and Pattern Detection Lab computer scientist, Daniel Neill (now at NYU).

Together, they gathered 12 years of 311 call data including information on not only rat sightings, but also other related factors such as overflowing trash bins, food poisoning cases, tree debris, and building vacancies.

They quickly discovered 311 calls related to food and shelter were the strongest predictors of rat infestations then built a model to predict spikes in rodent complaints days before the infestation would actually happen.

The model was shared with Chicago's sanitation department and deployed in 2013 for a trial period. After running Neill's model that year, the City of Chicago claimed it saw this method to be 20% more effective than traditional baiting methods for catching rats. In July 2013, Mayor Rahm Emanuel announced that as a result of the City's increase in preventive rodent baiting efforts in 2012, resident requests for rodent control services dropped 15% in 2013.

Now that we have discussed the case for using data analytical thinking we turn in our next module to how this new way of working can be more effectively applied.