What's Your Problem?

Tim Chartier

Data analytics is an emerging, if not already present, field of business. I work in data science collaborating with businesses, professional sports teams, and college organizations at Davidson College, where I teach. What lessons can we learn about problem solving from this relatively new field? In this article, we look at the importance of having a clear vision of the problem you wish to solve or question to answer.

At first glance, data science may seem like a field where one need not define a precise question. While this may be true at times, it is generally far from the rule. For example, at the beginning of the 2013–2014 season, the NBA installed SportVU cameras in all of its arenas. The technology captures *x*-*y* coordinates of every player on the court and all three spatial positions of the ball. Such information is captured 24 times per second for the entire game. This leads to a massive amount of data containing tremendous information on the game. With such raw data, as it is often called, new insights emerge from Kirk Goldberry's heat maps showing the frequency of shots to new work introduced at the 2015 MIT Sloan Sports analytics meeting on identifying defensive match-ups from such data.

From such examples it can appear that large data means big insights. While NBA teams have analysts, many note the continued difficulty of mining such raw data. Goldsberry, for example, approached the creation of heat maps from the view of his field—cartography. His question grew from his training: How could the data give spatial insight on shooting?

If one does not have a question in mind, then tackling a large problem isn't like looking for a needle in a haystack. While difficult, one can devise methods to look for a needle given the knowledge of the goal. Without a definite goal, one sifts through the haystack hoping to discover something. Yet searching for an easily seen large gem versus a stealthy needle is quite different. As such, defining the goal can define an appropriate process.

Yet choosing a process can also be guided by careful and continued focus on the goal. I illustrate with an example from my research group. We were recently posed an open problem from a professional sports organization. I teach at a highly selective liberal arts college where many students desire independent research projects.

For this problem, the data set was large and could answer many questions. Our job was to select relevant portions of the data and analyze them to come to conclusions about the posed question.

After a week, my students were moving forward but at a much slower rate than I anticipated. In the first of our two weekly meetings, I learned that the students were integrating the data into a database. I thought, "Oh, of course. Why I didn't think of that? It is the natural step that everyone takes." The students noted that such work would give flexibility to future problems.

The next week we had made progress, but again at a much slower rate than I originally anticipated. Such a miscalculation regarding the time needed for research isn't uncommon. As Einstein said, "If we knew what we were doing, we wouldn't call it research." As such, we may not know the exact process of solution until the problem is solved. Yet the difference between my

Published online in Wiley Online Library (wileyonlinelibrary.com).
DOI 10.1002/jcaf.22085

expectation and the actual solution time made me pause. Why did I originally have a different expectation? I reflected on the reasons for my initial impression.

My impression was connected to our posed question. We needed only portions of the data. Once identified, we'd look only at those portions and have no need for the overall data set. Said another way, we'd extract what we needed and focus our work on that smaller data set. Creating a database, while a common and wise step with our data, was unnecessary for our problem.

While many—and possibly everyone—have approached a similar problem in a particular way, they were not tackling your problem in your context. While learning from others can save time, you must also remember that they were most probably not tackling your problem. As such, keep your question in focus. My group inadvertently began working on methods that were ideal solutions to other questions, not ours.

In applied math, it can be easy to move to similar problems. Sometimes the new questions are more important to the intended goal. Other times, a blurring of one's focus causes such a change in trajectory. Such changes happen most naturally when you or your group enters challenging stages of the work.

How do you keep focus? Ideally, you create checkpoints where you reflect on what your goals are in the project and compare those to your current processes. In a sense, you stop and ask, "What's my problem?" Once answered, you can compare this with your current path toward solution.

One last point: What if you get to a point where you can't describe your question or possibly your current steps toward solution. For that, I turn to one last quote from Einstein: "If you can't explain it simply, you don't understand it well enough." Taking the time to explain what you are doing and your intended goals is worth the time. It can keep you on task, save time and resources, and in the end help you mine through data or tasks for intended outcomes.

Tim Chartier is Associate Professor of Mathematics and Computer Science at Davidson College in North Carolina.

DOI 10.1002/jcaf