**5 Concepts That Will Help Your Team Be More Data-Driven**

by [Thomas C. Redman](https://hbr.org/search?term=thomas%20c.%20redman)

November 01, 2018, Updated November 02, 2018

Koron/Getty Images

Summary.    Data is invading every nook and cranny of every team, department, and company in every industry, everywhere. Developing the talent needed to take full advantage must be a high priority. Indeed, everyone must be able to contribute to improving data quality,...

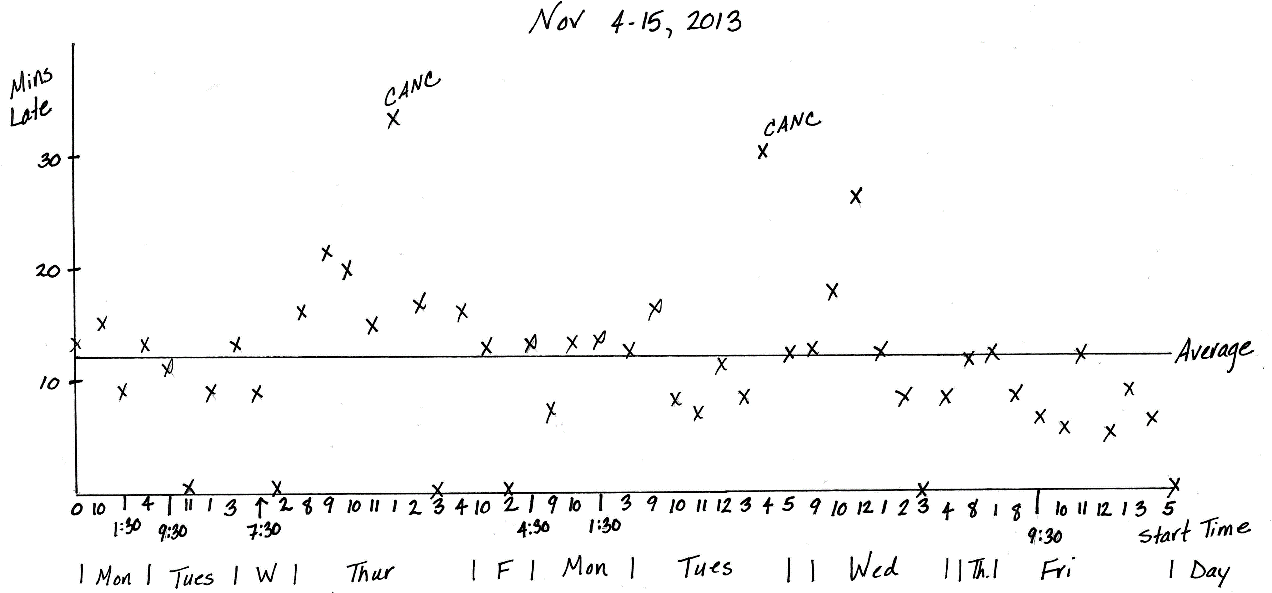
I’ve spent my career helping companies address their data and data quality opportunities. Overall, I rate progress as “[slower than hoped](https://hbr.org/2018/10/5-ways-your-data-strategy-can-fail).” While there are many contributing factors, one of the most important is the sheer lack of analytic talent, up and down the organization chart. In turn, this lack of talent makes it harder for companies to leverage their data, to take full advantage of their data scientists, and to get in front of data quality issues. Lack of talent breeds fear, exacerbating difficulties in adopting a data-driven culture. And so forth, in a vicious cycle.

Still, progress in the data space is inexorable and smart companies know they must address their talent gaps. It will take decades for the public education systems to churn out enough people with the needed skills — far too long for companies to wait. Fortunately managers, aided by a senior data scientist engaged for a few hours a week, can introduce five powerful “tools” that will help their existing teams start to use analytics more powerfully to solve important business problems. To be sure, these are not the only tools you’ll need — for example, I haven’t included A/B testing, understanding variation, or visualization here.  Nor is my intent to make people experts.  Rather, based on my experiences working with companies on their data strategy, these five concepts offer the biggest near-term bang for the buck.

The first is learning to **think like a data scientist**. We don’t speak about this often enough, but it is really hard to acquire good data, analyze it properly, follow the clues those analyses offer, explore the implications, and present results in a fair, compelling way.  This is the essence of data science. You can’t read about this in a book — you simply have to experience the work to appreciate it.  To give your team some hands-on practice, charge them with selecting a topic of their own interest (such as “whether meetings start on time”) and then have them complete [the exercise described in this article](https://hbr.org/2013/11/how-to-start-thinking-like-a-data-scientist). The first step will lead to a picture similar to the one below, and the rest of the exercise involves exploring the implications of that picture.

**How Often Do Meetings Start Late?**

Give employees hands-on experience with data by asking them to collect and plot data on a familiar topic. Meeting start times are just one example.



**Source:** Thomas C. Redman

Charge that senior scientist you’ve engaged with helping people in completing the exercise, teaching them how to interpret some basic statistics, tables, and graphics, such as a time-series plot and Pareto chart. As they gain experience, encourage your team to apply what they’ve learned in their work everyday. Be sure to make time for people to show others what they’re learning, say by devoting fifteen minutes to the topic in each staff meeting.  Most critically, lead by example — do this work yourself, present your results, and freely discuss the challenges you faced in doing the work.

As you and your team dive into data, you’ll certainly encounter quality issues, which is why **pro-actively managing data quality** is the next important skill to learn.  Poor data is the norm — fouling operations, adding cost, and breeding mistrust in analytics.  Fortunately, virtually everyone can make a positive impact here.  The first step is to make a simple measurement using the [Friday Afternoon Measurement method](https://www.youtube.com/watch?v=X8iacfMX1nw)  (the technique acquired this name because so many teams end up using it on Friday afternoon).

**Insight Center**

* [**Scaling Your Team’s Data Skills**](https://hbr.org/insight-center/scaling-your-teams-data-skills)

Help your employees be more data-savvy.

To do so, instruct your team members to assemble 10-15 critical data attributes for the last 100 units of work completed by their departments — essentially the last 100 data records.  Then, they should work through each record, marking obvious errors.  They should then count up the error-free records.  The number, which can range from 0 to 100 represents the percent of data created correctly, their Data Quality (DQ) score.  DQ can also be interpreted as the fraction of time the work is completed correctly, the first time.  Most managers are surprised by the results — they expect to score in the high 90s, [but DQ = 54 is the median score.](https://hbr.org/2017/09/only-3-of-companies-data-meets-basic-quality-standards)

FAM can also point out which data attributes have the biggest error rates, suggesting where improvements can be made, using root cause analysis, described next. Charge each member of your team with making one such improvement.

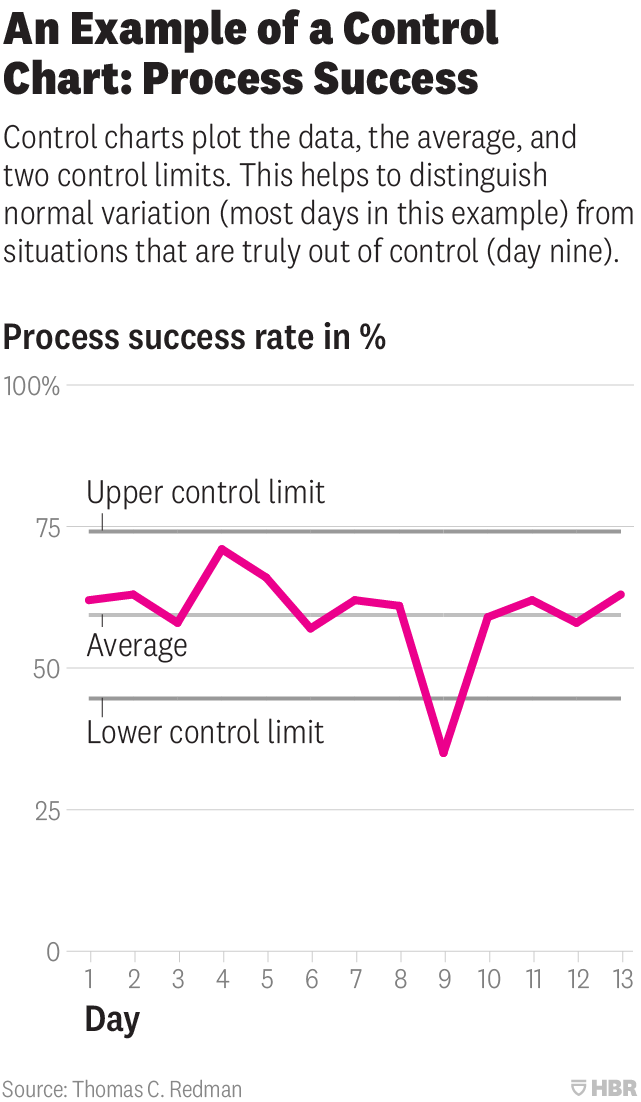
The third skill is **conducting a root cause analysis (RCA)** and its pre-requisite, **understanding the distinction between correlation and causation**. Studying the numbers can point to where most errors occur or demonstrate that two (or more) variables go up and down in tandem, but it cannot fully describe *why* this is.  [For example, studies show that the numbers of live births and storks in the countryside were highly correlated](https://priceonomics.com/do-storks-deliver-babies/). But storks do not bring babies!

Thus, look to the numbers to understand correlation and to the real-world phenomena to understand causation. Root cause analysis is a structured approach for getting to the real reasons things go wrong — the root causes.  It is important because, too often, managers and teams often accept easy explanations and don’t dig deep enough. And problems remain. RCA can enable them to develop a clearer picture and take actions that are more likely to solve the problem.

To develop this skill with your team, start by discussing [“how to explore cause and effect like a data scientist”](https://hbr.org/2014/02/how-to-explore-cause-and-effect-like-a-data-scientist?autocomplete=true) with your staff.  Then, the next time you find yourself tempted to accept someone’s intuitive reasoning as to why something went wrong, seize the opportunity to conduct a solid root cause analysis. There are many formal means to do so. [“The five whys,”](https://en.wikipedia.org/wiki/5_Whys) which forces you to make sure you’ve gotten to the root cause, and [fishbone diagrams](https://whatis.techtarget.com/definition/fishbone-diagram), which graphically represent multiple causes, are probably the best known. Have your data scientist pick one, and follow it! Over time, seek to make root cause analysis your standard for all important issues.

The fourth skill stems from the desire all managers have to “be in control.”  My working definition of control is “the managerial act of comparing process to standards and acting on the difference.”  But even the simplest process varies.  How can one distinguish normal day-in, day-out variation from situations that are truly out of control? Fortunately, **understanding and applying control charts** provides a powerful way to do just that.

Control charts feature a plot of the data, the average, and two “control limits,” (an upper control limit and a lower control limit).  Basic as they are, they reveal so much!  For example, in the Figure below:



1. Since day 9 falls outside the control limits, a manager can be certain this process is out of control. They should initiate a root cause analysis to figure out why.
2. There is an uptick at day 4 that looks encouraging. But a manager should not get too excited — the uptick was more likely due to random variation and was not sustained.
3. It is clear enough that that this process only succeeds 60% of the time. If this is not good enough, the manager must make fundamental changes.

Engage your data scientist in helping you and your team try control charts on a few important processes. Learn as you go, understanding key terms, determining which control charts to use, and striving first to get processes under control — your confidence will grow, as will your ability to manage your team!

Finally, all managers and their teams should learn to **understand and apply regression analysis**. Regression provides a powerful means to explore the numerical relationships between variables.  To help illustrate this, consider “umbrella sales.”  There are dozens of factors that could increase sales (e.g., rain) or decrease sales (e.g., a competitor’s price cut).  Regression provides a way to determine which variables are most important and their impact on sales.  For example, an analysis may yield:

Monthly sales = 200 + 5\*(days of rain) – 10\*(competitor price cut in $) + error term

Meaning that:

* Absent other factors, monthly sales are about 200 units.
* A day of rain is associated with the sale of five more umbrellas,
* A competitor cutting its price by one dollar is associated with ten fewer umbrellas sold

The model is not perfect — hence the error term. For example, suppose you sold 250 umbrellas in a month when there were 15 rainy days and a competitor cut its price by $2.  Based on the formula, one would expect umbrellas sales to be 200 + 5\*15 – 10\*2 = 255 units.  So the error term in that case is 5 umbrellas.

Like all analyses, the more variables, the more complex the analysis, so start by focusing on one independent (e.g., explanatory) variable.  In parallel, read [“A Refresher in Regression Analysis,”](https://hbr.org/2015/11/a-refresher-on-regression-analysis) which uses umbrella sales as an example to explain the terms and underlying concepts. Charge your data scientist with helping your team do the work, and making sure team members don’t get bogged down in details.  Only then should you move onto two, three, or more variables and more complex regression models.

These five tools are powerful, even elegant, in their own ways.  They provide far greater capabilities than the steps described here, which aim only to get you started. You’re certain to take some false steps along the way, but press on.  Work with your data scientist to learn even more. As your team grows more confident in using analytics, the business benefits you gain will more than justify the effort.