Quality, Analytics, and Big Data

Charles A. Liedtke, Ph.D.
Strategic Improvement Systems, LLC
www.strategicimprovementsystems.com
charles@sisliedtke.com
(952) 380-0778

January 27, 2016

Summary

The integration of quality, analytics, and big data guided by the strategic direction of the organization can potentially result in new sources of customer value and a new source of competitive advantage. This research report describes some of the advances related to quality, analytics, and big data and explores the benefits of integrating them in practice. Analytics and big data techniques can help improve product and service quality by generating new customer insights and enhancing decision making. Also, it is argued that analytics and big data initiatives can be improved if quality principles are applied. It is explained that analytics is not just about numbers and big data doesn't necessarily mean we need millions of data points. A new emergent definition of big data is introduced plus ten ideas for better integrated quality, analytics, and big data practices.

Key Words: Quality, Quality Superiority, Analytics, Big Data

Research Report Outline

I. Introduction to Quality, Analytics, and Big Data

Phenomenon of Interest

Many organizations strive to improve product and service quality and some have made *providing superior quality* a strategic intent (Liedtke et al., 2010). Quality dashboards and scorecards are now commonplace in organizations in order to make quality performance visible. This has been made easier because of the advances in real-time data collection and display technologies. The growing number of *analytics* and *big data* success stories have captured the imagination of many senior executives. They represent novel ways to measure, understand, and improve organizational performance including quality performance.

The pursuit of superior quality is not a new human endeavor and neither is *the collection and analysis of data*. These activities have occurred for thousands of years. This research report describes some of the advances related to quality, analytics, and big data and explores the benefits of integrating them in practice. Analytics and big data techniques can help improve product and service quality by generating new customer insights and enhancing decision making. Also, it is argued that analytics and big data initiatives can be improved if quality principles are applied. The integration of quality, analytics, and big data guided by the strategic direction of the organization can potentially result in new sources of customer value and a new source of competitive advantage. Four questions guided the research:

1) What are the data analysis trends in the U.S.?
2) What are the benefits of integrating quality, analytics, and big data?
3) How can product and service quality be improved by applying analytics and big data?
4) How can analytics and big data initiatives be improved by applying quality principles?

The three constructs are depicted in Figure 1.

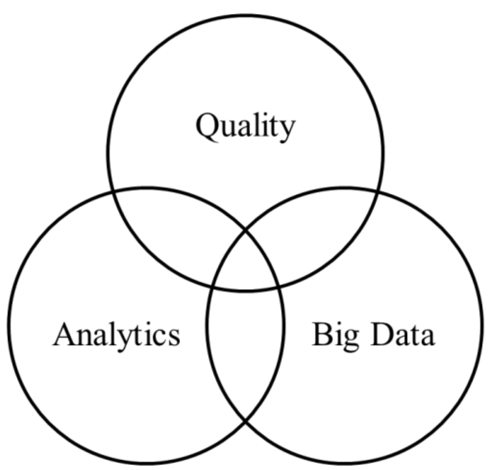

Figure 1. Quality, Analytics, and Big Data.

Quality principles help focus organizational attention, resources, and actions on improving product and service quality. *Quality principles* are defined for the purpose of this report as *principles for guiding quality-oriented organizational actions. Analytics* is defined here as *the collection and analysis of qualitative and quantitative data for decision making. Big data* will tentatively be characterized by the "Four V" framework consisting of data *volume*, *variety*, *velocity*, and *veracity* (see, e.g., Kelly & Hamm, 2013; Zikopoulos et al., 2015). A new emergent definition of big data will be presented later.

The research for this report began because of an invitation to present on the "*Trend of Data Analysis in the U.S.A.*" at the International Conference on Quality held in Tokyo, Japan in 2014. There were similar *trend* presentations on Europe and Asia. The positive response to the presentation and the realization that there were potential benefits associated with integrating quality, analytics, and big data led to an expansion of the original research project scope.

Quality Background

Statistical methods have long played an important role in quality improvement activities. Walter A. Shewhart's invention of the statistical control chart in 1924 helped workers understand the variation associated with measurable product characteristics (Shewhart, 1939, p. 23). This knowledge led to actions to reduce and control variation. Shewhart (1939, pp. 44-45) identified a three step mass production process consisting of *specification*, *production*, and *inspection*. He recommended that the steps be conducted sequentially and iteratively "in a dynamic scientific process of acquiring knowledge" within which statistical methods would play a vital role. Many of the data collection and analysis activities were very labor-intensive in the 1920s.

Quality principles have appeared often in the quality management literature. Kaoru Ishikawa (1985, p. 104) identified six principles that described Total Quality Control as practiced in Japan. Deming (1986, pp. 23-24) suggested *14 points* to serve as the basis for transforming American industry. For example, his 5th point is, "Improve constantly and forever the system of production and service, to improve quality and productivity, and thus constantly decrease costs." Deming (1994, p. 93) later offered a four-component *system of profound knowledge* for understanding and improving organizations: appreciation for a system, knowledge about variation, theory of knowledge, and psychology. Liker (2004, pp. 37-41) provided an executive summary of the fourteen Toyota Way principles. For example, Principle #1 is, "Base your management decisions on a long-term philosophy, even at the expense of short-term financial goals." Kano (2014) mentioned five principles for TQM in the 21st Century to enhance competitiveness including "Scientific way based on data and logic." Lastly, the Baldrige Performance Excellence Program (2015) suggested eleven core values including "Management by Fact." All of these sets of principles were intended to guide actions for achieving superior quality.

The Deming Prize was established in 1951 and is administered by the Japanese Union of Scientists and Engineers. It remains one of the most prestigious organizational quality awards in the world. Six categories comprise the Deming Prize Framework including "Collection and analysis of quality information and utilization of IT" (Deming Prize Committee - JUSE, 2015).

Statistical methods—including the statistical control chart—play a central role in the Japanese Society for Quality Control's (JSQC) "STANDARD: Guidelines for Daily Management" (JSQC, 2014). The Malcolm Baldrige National Quality Award is another prestigious organizational quality award that is administered by the U.S. Department of Commerce's National Institute of Standards and Technology. The Baldrige Excellence Framework is composed of seven categories including "Measurement, Analysis, and Knowledge Management" (Baldrige Performance Excellence Program, 2015). Big data is now explicitly mentioned in the Baldrige Award material (p. 45): "**Big data.** For all organizations, turning data into knowledge and knowledge into useful strategic insights is the real challenge of big data. . . . In 2015, the Criteria incorporate an enhanced focus on data analytics, data integrity, and cybersecurity." Evans (2015) discussed the various roles analytics plays in the Baldrige Award criteria and commented (p. 15) on what he termed *modern analytics*: "Analytical methods have been essential to quality assurance and quality management since the birth of the discipline; however, modern analytics opens many new opportunities for quality managers, particularly with applications of data mining and text analytics." The task of integrating quality, analytics, and big data practices is made easier because of the historical role statistical methods have played in organizational quality improvement activities.

Analytics Background

Analytics has become an organizational performance improvement approach in its own right and continues to receive extensive media attention. Recall that *analytics* is defined here as *the collection and analysis of qualitative and quantitative data for decision making*. We can see from this definition that analytics is *not just about numbers*, but can involve the collection and analysis of text data, pictures, videos, audio recordings, etc.

The modern version of analytics arguably became mainstream after the 2003 publication of the book *"Moneyball"* (Lewis, 2003) followed by the 2011 successful movie of the same title. The book chronicled how the Oakland Athletics Major League Baseball (MLB) team was able to succeed with a relatively low payroll—in part—through the use of statistical methods. Every pitch of every MLB game is measured and videotaped today allowing for even more advanced analytics.

This author has received a steady stream of questions from senior executives over the past few years as they attempted to understand analytics: "Can analytics help us achieve our strategic objectives?"; "How can analytics help us improve our product and service quality?"; "Who should lead our analytics team?"; "How much should we invest in analytics?"; and "How can we create a data-oriented culture?" Similar questions have been asked regarding big data. One reason terms like *analytics* and *big data* are confusing is because there are no globally-accepted standard definitions. Albright and Winston (2015, p. 1) commented in their textbook: "But regardless of what it is called, data analysis is currently a hot topic and promises to get even hotter in the future." They (2015, p. 3) elaborated further on **business analytics**:

> "A large amount of data already exists, and it will only increase in the future. Many companies already complain of swimming in a sea of data. However, enlightened companies are seeing this expansion as a source of competitive advantage. In fact,

> one of the hottest topics in today's business world is **business analytics**. This term has been created to encompass all of the types of analysis discussed in this book, so it isn't really new; we have been teaching it for years. The new aspect of business analytics is that it typically implies the analysis of very *large* data sets, the kind that companies currently encounter. By using quantitative methods to uncover the information in these data sets and then acting on this *information*—again guided by quantitative analysis—companies are able to gain advantages that their less enlightened competitors are not able to gain."

Organizations in numerous industries—from Major League Baseball to financial services to manufacturing—are now competing on analytics. Davenport and Harris (2007, pp. 34-35) identified four pillars and five stages of analytical competition. Hal Varian (Varian, 2009), Chief Economist for Google, described some of the skills that will be necessary in the future: "The ability to take data – to be able to understand it, to process it, to extract value from it, to visualize it, to communicate it's going to be a hugely important skill in the next decades, not only at the professional level but even at the educational level for elementary school kids, for high school kids, for college kids."

There continues to be a steady stream of new writings on analytics (see, e.g., Brown, 2013; Davenport, 2013; Hoffman, Lesser, & Ringo, 2012; McNeill, 2014). Nate Silver helped direct the media spotlight on analytics in 2008 by successfully predicting the winner of the presidential election in 49 out of 50 states and then he improved his performance in the 2012 presidential election with 50 out of 50 states (search for *Wikipedia Nate Silver*). Silver (2012) described some of the benefits and risks of analytics-based prediction and elaborated on the reasons why predictions sometimes fail in his award winning book *"Signal and the Noise."*

Information has played an important role in society and has contributed to economic progress for centuries (see, e.g., Chandler & Cortada, 2000; Gleick, 2011). The rapid advances in data collection and analysis capabilities through information technology have altered information economics (Evans & Wurster, 2000; Shapiro & Varian, 1999; Varian, Farrell, & Shapiro, 2004). Textbooks on information systems run the risk of becoming *behind the times* soon after publication because of the rapid changes that are occurring. *Information* is now commonly viewed as a valuable asset and potential source of competitive advantage. See Porter (1985) for a treatise on competitive advantage and Rainer and Cegielski (2011) for an introduction to information systems.

Big Data Background

The exact origin of the term *big data* is somewhat of a mystery. According to Diebold (2012, p. 5): "The term 'Big Data,' which spans computer science and statistics/econometrics, probably originated in lunch-table conversations at Silicon Graphics Inc. (SGI) in the mid 1990s, in which John Mashey figured prominently." Press (2014) traced the term "big data" to a 1997 NASA paper. Big data definitions that limit themselves to only the amount of data (volume) are not very useful because "big data" implies there might exist "bigger" data like *gigantic data* or *insanely gigantic data*. This idea is depicted in the original cartoon in Figure 2 (Liedtke, 2014).

Figure 2. Big Data and Gigantic Data.

Laney (2001) is credited with the "Three V's of Big Data" (Diebold, 2012). It is a framework consisting of three components: data *volume*, data *velocity*, and data *variety*. *Veracity* is a fourth "V" that has been mentioned in the literature (see, e.g., Zikopoulos et al., 2015, p. 8). There have been numerous examples of big data applications mentioned in the media and literature ranging from public health to airline ticket pricing to astronomy (see, e.g., Davenport, 2014; Mayer-Schonberger & Cukier, 2013). Some organizations routinely apply advanced analytics and big data techniques in their day-to-day operations such as Amazon (Stone, 2013), Facebook (Kirkpatrick, 2010), and Google (Schmidt & Rosenberg, 2014).

Komatsu started developing big data capabilities in the late 1990s (Asada, 2014). It is one of the global leaders in heavy equipment manufacturing. Komatsu has been attempting to create *Dantotsu* products for several years. *Dantotsu* means *unique and unrivaled* (Hasegawa, 2010, p. 150). This has evolved with the addition of *Dantotsu* services and solutions (Asada, 2014; Sakane, 2014). KOMTRAX is a telematics system for monitoring and improving the performance of Komatsu's construction machines. The system helps Komatsu customers achieve their goals related to productivity and safety through remote monitoring, reporting, and focused interventions. There were more than 350,000 Komatsu machines working with KOMTRAX as of August of 2014 (Asada, 2014). Some of the system's features include the Orbcomm satellite, GPS capabilities, web data delivery, and an Internet user interface. Komatsu also has KOMTRAX Plus which is a machine health monitoring system for mining machines.

Zillow—a big data company that functions in the real estate industry—has created a distinctive competitive advantage by making (1) massive amounts of residential home data and (2) a user-friendly calculator available to consumers on its website (Rascoff & Humphries, 2015). McLaren Applied Technologies has developed big data capabilities to help improve Formula 1 race car performance (Higginbotham, 2015). Big data is not always about the perceived sample size. The sensors on one car in a Formula 1 race generate an "extremely large" number of measurements that can be analyzed remotely in *near real-time* (Higginbotham, 2015). GE is spending $1bn a year as it transforms itself through its *industrial internet strategy* (Crooks, 2016). Mahindra & Mahindra's Farm Equipment Sector business won the Deming Prize in 2003 and the Japan Quality Medal in 2007. Data collection and analysis activities are occurring across the entire business to gain customer insights for developing innovative solutions and improving performance such as machine uptime (Sharma, 2014). Wachter (2015, pp. 116) discussed the growing use of big data applications in healthcare like the use of nanosensors and accelerometers. H. Ishikawa (2015) provided explanations of social media and big data and how the two can be integrated. IBM has made aggressive strategic moves in analytics, big data, and artificial intelligence. It is now marketing Watson—which is IBM's famous computer system that defeated two human champions in *Jeopardy!* (Waters, 2016).

There are now available numerous books on the technical aspects of analytics and big data. For example, Provost and Fawcett (2013) described data science tools and techniques and discussed the connection between data science and business strategy. Grus (2015) showcased tools, techniques, and detailed code associated with Python. Marz and Warren (2015) described some of the technical aspects of big data such as master datasets, batching, layering, data immutability, queries, and Lambda Architecture. White (2015) provided a detailed guide to Hadoop—which is an open-source software system. Lastly, Zikopoulos and his co-authors (2015) wrote one of the most comprehensive books on big data—employing both business and technical perspectives.

Analytics and big data have rapidly evolved along with other information technologies as part of the digital revolution. Steve Jobs, former CEO of Apple, announced his *digital hub strategy* in his keynote speech at the Macworld event in San Francisco in January of 2001 (Kahney, 2009). According to Kahney (2009, p. 180): ". . . he [Jobs] laid out Apple's vision—a vision that would inspire more than a decade's worth of innovation at Apple, and would shape almost everything the company did, from the iPod to its retail stores and even its advertising. The digital hub strategy is possibly the most important thing Jobs has laid out in a keynote speech." Much has been written on digital technology (see, e.g., Schmidt & Cohen, 2014; Wilson, 1995) and the evolution of the computer and Internet (Greengard, 2015; Isaacson, 2014).

We should be aware of the limitations of digital technology from a knowledge management perspective. Polanyi (1966, p. 4) made the insightful comment: "*we can know more than we can tell*." Nonaka and Takeuchi (1995, p. 8) described two kinds of knowledge: *explicit* and *tacit*. *Explicit* knowledge is formal and systematic and can be easily communicated and shared. *Tacit* knowledge is not easily visible or expressible. Arguably, analytics and big data techniques are better suited for explicit knowledge at this point in time, but this could change with practice and the advancement of analytical tools. It is useful to know the four conversion modes for creating knowledge described by Nonaka and Takeuchi (1995, pp. 62-70): socialization, externalization, combination, and internalization. More will be said about these knowledge conversion modes later.

Analytics and big data applications sometimes involve the use of one or more algorithms. They are often managed in a secretive manner due to their proprietary nature and are analogous to a "secret sauce" used by a restaurant. Some algorithm archetypes include *search completion*, *preference learning*, *matching*, *geographic spread*, *network connections*, *topic trending*, *sentiment analysis*, and *missing puzzle piece*. A *search completion* algorithm completes our search request before we are finished typing based on an algorithm. If I start typing "che" into a popular search engine, then the first item that appears on the pop-up list is "cheap flights" even though I wanted "chess." A *preference learning* algorithm recommends an item like a song, movie, or book based on what we've selected in the past. An example of a *matching* algorithm is one that connects two people who are looking for a companion. A *geographic spread* algorithm predicts the geographic movement of something like a disease based on search results. A *network connection* algorithm predicts others in a larger network we might know and want to add as a connection. A *topic trending* algorithm provides a real-time counting and ranking of topics based on Internet "appearances." A *sentiment analysis* algorithm estimates the general feeling towards a phenomenon. Finally, a *missing puzzle piece* algorithm predicts what someone might need based on their selections to date. For example, if you purchase plane tickets to Jamaica and a snorkeling book, then an algorithm might predict you also want to purchase a snorkel, mask, and flippers.

Much has been written about algorithms (see, e.g., Luca, Kleinberg, & Mullainathan, 2016; MacCormick, 2012; Pasquale, 2015; Siegel, 2013). An algorithm is a model that produces outcomes. Some media articles on analytics and big data state that "an algorithm was used." Rarely are we provided with statistics on the performance of the algorithm. According to Box and Draper (1987, p. 424): "Essentially, all models are wrong, but some are useful. However the approximate nature of the model must always be borne in mind." An algorithm used for prediction is *predictive analytics*. Suppose we have customers *on contract* who will either renew their contract or not. An algorithm can be developed to predict what each customer will do based on certain characteristics. We can evaluate the performance of the algorithm if we compare actual and predicted outcomes. There are four possibilities – two where the algorithm predicted correctly and two where it was wrong. Figure 3 (Liedtke, 2015) shows the hypothetical performance of the algorithm. The use of an algorithm should lead to an evaluation of its performance and potential modifications.

Big data is not a panacea. Elizabeth von der Goltz, Senior Vice President at Bergdorf Goodman discussed the limitations of big data (Garrahan, 2015):

> "She [von der Goltz] does not believe her buyers' expertise can ever be replaced by that new retail obsession: big data. 'Our buyers are editors as well as discoverers of new fashion. It's different maybe if you're buying for a supermarket, but when you're talking about luxury, and especially jewelry, you have to know why a piece is going to retail at $85,000 from its weight, its feel, to the quality of the stones used.' Big data analysis also lacks that basic human element required in a luxury investment. 'Jewelry is an emotional purchase,' she says. 'You have to get in pieces that you think will move people.'"

More research needs to be conducted to identify the *conditions under which* big data will be useful.

Algorithm Performance
Correct Predictions = 65%
Incorrect Predictions = 35%

		Actual Customer Behavior	
		Customer Renews Contract	Customer Does Not Renew Contract
Predicted Contract Renewal (Algorithm)	Customer Will Renew Contract	**Correct 42%**	**Bad Surprise 21%**
	Customer Will Not Renew Contract	**Good Surprise 14%**	**Correct 23%**

Figure 3. Algorithm Performance.

So what is new with big data? Here are some clear differences from the mid 1990s when big data was discussed at Silicon Graphics (Diebold, 2012):

- More data from more sources of different types
- Data arriving more often in real-time
- Faster data processing and more data storage
- More devices and applications
- Mobile access to data and data summaries
- More analytical tools and techniques
- More *technically savvy* customers

This presents organizations with a great opportunity to improve quality, analytics, and big data practices through the integration of the three items.

II. Research Study

Purpose of the Research Study

The research study was initially launched to develop content for a presentation on the "*Trend of Data Analysis in the U.S.A.*" (Liedtke, 2014) at the International Conference on Quality (ICQ) held in Tokyo, Japan in October of 2014. The context of the research was *organizational performance measurement* and *improvement*. The original research activities included a review of the literature; internet searches; a search for best practices; a review of university curriculums; and a survey. The scope of the study expanded following the conference based on the positive feedback and the realization that there were potential benefits associated with integrating quality, analytics, and big data. This led to a more extensive review of the literature and search for best practices.

Literature Review: Timeline of Selected Events

The people who inhabited the Chaco Canyon region of the United States (northwestern New Mexico) most likely made celestial observations over a long period of time before 1100 A.D. They then created predictive models based on *naturally occurring cycles*. This allowed them to (1) strategically design their dwellings and (2) plan and conduct agricultural and ceremonial activities (Frazier, 2005; Zeilik, 1984). Zeilik (1984, p. 66) described the Sun Priest's role in Chaco Canyon:

> "Solar observation is invested in a religious officer, usually called the Sun Priest. He watches daily from a special spot within the pueblo or not far outside of it, carefully observing the position of sunrise or sunset relative to features on the horizon. He knows from experience which horizon points mark the summer and winter solstices and the times to plant crops. These he announces within the pueblo, usually ahead of time so that ritual and planting preparations can be carried out."

Today's data scientists can be thought of as the modern version of the Sun Priests because they are performing similar activities at a fundamental level.

Figure 4 depicts a timeline of subjectively chosen events related to quality, analytics, and big data to help illustrate how practices have evolved and to *weave together* interesting historical *threads*. It begins with the earliest known *pie chart* and ends with a visit to the IBM website.

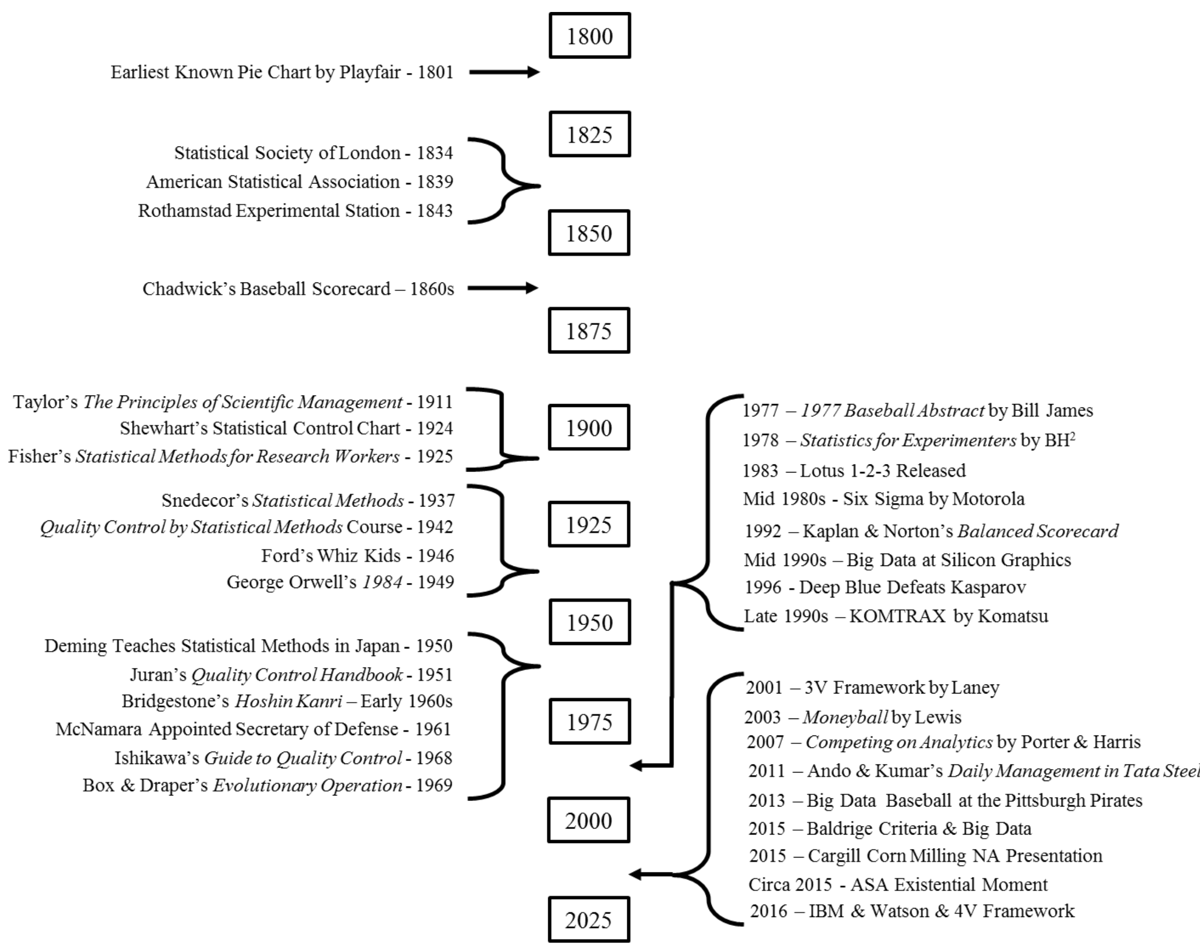

Figure 4. Timeline of Selected Events.

1801: Earliest Known Pie Chart by Playfair

The *pie chart* is a circular graphic that depicts the relative frequencies or relative proportions of multiple categories. Most people have seen a pie chart and know how to interpret one. According to Wikipedia (search for *Wikipedia pie chart*), the earliest known pie chart appeared in William Playfair's *"Statistical Breviary"* (book) of 1801 (see also Cleveland, 1985). This simple chart is still widely used to visually communicate data summaries.

1834 & 1839: The Formation of Two Early Statistical Societies

The Statistical Society of London was founded in 1834—which later became the Royal Statistical Society in 1887 (search for *Wikipedia Royal Statistical Society*). The American Statistical Association was founded in Boston in 1839 (search for *Wikipedia American Statistical Association*). A professional society consists of members with common interests who learn from each other and advance their field of interest. They typically share a paradigm (see Kuhn, 1970).

1843: The Founding of the Rothamstad Experimental Station

The Rothamstad Experimental Station (now named Rothamstad Research) was founded in 1843. It is the oldest agricultural research station in the world whose mission is "to deliver the knowledge and new practices to increase crop productivity and quality and to develop environmentally sustainable solutions for food and energy production" (see Rothamstad Research website, 2016). Many advances in statistical methods have occurred at the research station (Diggle & Chetwynd, 2011; Salsburg, 2001).

1860s: Invention of the Baseball Scorecard by Chadwick

Advanced analytical techniques have been developed and applied in Major League Baseball for several decades. The collection and analysis of baseball game data began as early as the 1850s (Schwarz, 2004; Thorn & Palmer, 1985). The invention of the baseball scorecard is attributed to Henry Chadwick (Schwarz, 2004, p. 6): "To spread the gospel, Chadwick invented his own personal scoring form in the hope that it would become standard." MLB baseball fans still use modern versions of the original scorecard at the same time every pitch is measured and videotaped.

1911: *The Principles of Scientific Management* by Taylor

Frederick Taylor and others advanced the practice of data collection and analysis in improving operational efficiency at the Midvale Steel Company and Bethlehem Steel Company. Taylor's book *"The Principles of Scientific Management"* was published in 1911 (Taylor, 1947). This work led to the development of best management practices for improving quality and productivity. Many of the concepts and techniques are still used in the name of *lean manufacturing* today.

1924: Invention of the Statistical Control Chart by Shewhart

Walter A. Shewhart invented the statistical control chart in 1924 (Shewhart, 1939, p. 23) while working at Bell Telephone Laboratories as a tool to help production workers understand and reduce variation in the measurable characteristics of telephones (Shewhart, 1931, 1939). This tool became one of the foundational elements of the statistical quality control movement. Shewhart had a great influence on generations of quality management thought leaders (see, e.g., Deming, 1986, 1994; Ishikawa, 1985; Juran, 2004). According to Ishikawa (1985, p. 14): "Modern quality control, or statistical quality control (SQC), as we know it today, began in the 1930s with the industrial application of the control chart invented by Dr. W. A. Shewhart of the Bell Laboratories."

1925: *Statistical Methods for Research Workers* by Fisher

Sir Ronald Fisher's classic book "*Statistical Methods for Research Workers*" was published in 1925. Three of the chapters included *Distributions*, *The Correlation Coefficient*, and *The Principles of Statistical Estimation*. The reference for the fourteenth reprinted edition is shown in the Reference section (Fisher, 1973). According to Diggle and Chetwynd (2011, p. 59): "The statistician R. A. Fisher (1890-1962) was employed at Rothamstad [Experimental Station] between 1919 and 1933. During this period he revolutionized the theory and practice of statistics, especially as it applies to agricultural experimentation . . ."

1937: *Statistical Methods* by Snedecor

The first edition of George Snedecor's book *"Statistical Methods"* was published in 1937. It became a highly respected "methods" text for the application of statistical methods in agriculture and industry. The sixth edition was co-written with William Cochran while Cochran was a graduate student at Iowa State University (Snedecor & Cochran, 1989).

1942: *Quality Control by Statistical Methods* Course

Statistical methods were used extensively by the U.S. for war efforts during World War II. According to Deming (1986, p. 487): "Statistical methods had taken fire in America around 1942, following a series of 10-day intensive courses for engineers initiated by Stanford University on a suggestion from this author. The War Department also gave courses at factories of suppliers." William McNamara—the future Secretary of Defense in the Kennedy administration—served on a statistics team in the war. Rosenzweig (2010, p. 88) noted on a timeline: "[McNamara] Serves in the army on an elite team, Statistical Control, that applies quantitative analysis to the war effort." Several team members became a group at Ford known as the Whiz Kids (Rosenzweig, 2010).

1946: Ford's Whiz Kids

McNamara and seven others who were on the Statistical Control team during the war joined the Ford Motor Company. They achieved improvements using modern management control systems (Rosenzweig, 2010). McNamara later became the first nonfamily Ford president in 1960.

1949: *1984* by Orwell

George Orwell's book *"1984"* was published in 1949 (Orwell, 1949). Winston Smith, who worked at the Ministry of Truth, was constantly watched by telescreens controlled by Big Brother.

1950: Deming Teaches Statistical Methods in Japan

W. Edwards Deming made several trips to Japan following the war to serve as a statistical consultant and teach top managers, engineers, and foremen. Deming (1986, p. 489) stated: "Over 400 engineers studied in eight-day courses in the summer of 1950 in Tokyo, Osaka, Nagoya, Hakata, given by this author, on the methods and philosophy of Shewhart." The prestigious Deming Prize was established in 1951 (see the JUSE website).

1951: *Quality Control Handbook* by Juran

The classic quality management book *"Quality Control Handbook"* was published in 1951 (Juran, 1951). Joseph M. Juran edited the book and wrote six of the fifteen chapters. He later made several trips to Japan—his first in 1953. According to Juran (2004, p. 247): "My invitation to lecture in Japan was a result of the publication of my *Quality Control Handbook*. [Ken-ichi] Koyangi told me he was interested in the *Handbook* because most of it dealt with matters beyond statistics—matters such as economics of quality, specification, organization, inspection, assurance, and supplier relations. He thought that Japan had reached a state of self-sufficiency in SQC but that to achieve quality required much more than application of statistics." Juran (2004, p. 244) identified three chief contributions to Japan's progress after the war—one of which related to training: "The lectures of two Americans, W. E. Deming and J. M. Juran, which provided the seed training courses in statistical methodology and managing for quality, respectively."

Early 1960s: *Hoshin Kanri* at Bridgestone

Statistical methods have a long history in Japanese quality efforts. According to Akao (1991, p. 3): "Japanese quality control in its present form is based on the statistical quality control (SQC) that was brought over from the United States after World War II. Later in Japan we developed total quality control (TQC) . . . The transition from SQC to TQC occurred during the years 1961 to 1965 in companies whose achievements in quality earned them the Deming Prize." *Hoshin kanri*—a strategic improvement system—began around 1962. According to Akao (1991): "In 1962 the Bridgestone Tire Company conceived the idea of systematizing *hoshin kanri* as part of TQC."

1961: McNamara Appointed Secretary of Defense

William McNamara was appointed Secretary of Defense by President Kennedy in 1961. Rosenzweig (2010, p. 89) noted on a timeline: "[McNamara] starts to apply principles of modern management to the Pentagon, improving efficiency and instituting systems analysis as a basis for making decisions." McNamara later created the Vietnam Study Task Force (Rosenzweig, 2010) to write an analysis of the Vietnam War which later was published as the *"Pentagon Papers"* (Sheehan et al., 1971). McNamara (1996) eventually wrote an introspective book on the war.

1968: *Guide to Quality Control* by Ishikawa

One early leader in statistical quality control in Japan was Kaoru Ishikawa. His book "*Guide to Quality Control*" (Ishikawa, 1982) became a valuable quality resource for Japanese companies.

1969: *Evolutionary Operation* by Box and Draper

George E. P. Box and Norman Draper had their book *"Evolutionary Operation"* published in 1969. This book described how small-scale designed experiments could be conducted over time to improve and/or control manufacturing processes (Box & Draper, 1969).

1977: *1977 Baseball Abstract* by James

Baseball analytics pioneer Bill James (search for *Wikipedia Bill James)* published his first Baseball Abstract in 1977 (James, 1977). He has become one of the most famous baseball statisticians. *The Bill James Handbook* is now published annually.

1978: *Statistics for Experimenters* by Box, Hunter, & Hunter

George E. P. Box, J. Stuart Hunter, and William Hunter (1978) had their classic book "*Statistics for Experimenters"* published in 1978. This book covered a number of fundamental design of experiments topics including randomization, blocking, model building, and factorial designs.

1982: Lotus 1-2-3 Released

Lotus was founded in 1982 by Mitch Kapor and Jonathan Sachs with backing from Ben Rosen and the spreadsheet program Lotus 1-2-3 was released on January 26, 1983 (search for *Wikipedia Lotus 1-2-3*). Lotus 1-2-3 went on to become one of the leading desktop computer analytical tools used in organizations and contributed to the decentralization of analytical activities.

Mid 1980s: *Six Sigma* by Motorola

Six Sigma began as a formal improvement approach at Motorola during the middle 1980s (Schroeder et al., 2008). It emphasized the application of basic and advanced statistical methods and led to the widespread teaching and application of statistical methods in industry. GE became one of the most well-known organizations to launch a Six Sigma initiative in the mid-1990s. The GE Capital business unit was one of the first *service* organizations to deploy Six Sigma and the associated advanced statistical techniques.

1992: *Balanced Scorecard* by Kaplan and Norton

Kaplan and Norton (1992) had their classic paper on the *"Balanced Scorecard"* published in the Harvard Business Review in 1992 followed by a book with the same title in 1993. Their writings helped elevate performance measurement to the strategic level of firms and led to the creation of balanced scorecards in many Fortune 500 companies. Their four performance perspectives are *Financial*, *Customer*, *Internal/Business Process*, and *Learning and Growth*.

Mid 1990s: Big Data Discussed at Silicon Graphics

Diebold (2012) pinpointed the mid 1990s at Silicon Graphics as one of the first instances where the term *big data* was used. Press (2014) dated the first usage of the term *big data* to a 1997 NASA paper. Laney (2001) later described the "3V Framework": Volume (of data), Velocity (of data), and Variety (of data). Some authors now use a "4V Framework" which consists of *volume*, *velocity*, and *variety* as before, but adds *veracity* (see, e.g., Zikopoulos et al., 2015).

1996: Garry Kasparov Defeated in Chess by Deep Blue (IBM)

Deep Blue, a computer developed by IBM, defeated the then reigning Chess World Champion Garry Kasparov in a game of chess for the first time in 1996. Deep Blue eventually lost the 1996 match, but went on to win the 1997 match amidst some controversy (search for *Wikipedia Deep Blue versus Garry Kasparov*).

Late 1990s: KOMTRAX by Komatsu

Komatsu—a company that strives to produce Dantotsu products that emphasize Environment, Safety, and Information and Communication Technologies—started to develop big data capabilities in the late 1990s with its KOMTRAX (Komatsu Tracking System). The first machines equipped with KOMTRAX appeared in 2000 in the U.S. There were over 350,000 machines worldwide on the system as of August of 2014 (Asada, 2014).

2001: 3V Framework by Laney

Laney (2001) described the "3V Framework": Volume (of data), Velocity (of data), and Variety (of data). Some still use the 3V framework to define big data (see, e.g., H. Ishikawa, 2015).

2003: *Moneyball* by Lewis

Arguably, one of the major starting point events for the beginning of the modern analytics movement was the 2003 book on the Oakland Athletics titled *"Moneyball"* by Michael Lewis (Lewis, 2003) and the 2011 box office hit movie by the same title. All Major League Baseball teams use analytics today to identify the factors that affect team performance and to improve.

2007: *Competing on Analytics* by Davenport and Harris

The book by Davenport and Harris (2007) titled "*Competing on Analytics*" was published in 2007. They discussed how analytics can be a source of competitive advantage and identified four pillars for competing on analytics and they also identified five stages of analytical competition.

2011: *Daily Management the TQM Way in Tata Steel* by Ando and Kumar

The book by Ando and Kumar (2011) titled *"Daily Management The TQM Way"* was published in 2011. The authors discussed in depth the Daily Management System which emphasizes standard work, the use of statistical control charts, and the never-ending rotation of the Plan-Do-Check-Act (PDCA) cycle—which is a version of the scientific method. One of the chapters is on the "History of promotion of Daily Management at Tata Steel." Interestingly, Taylor and his colleagues applied some similar concepts (standard work) and techniques (observing work) at the Midvale Steel Company and the Bethlehem Steel Company.

2013: Big Data Baseball at the Pittsburgh Pirates

The Pittsburgh Pirates Major League Baseball team ended a streak of twenty losing seasons with a winning season in 2013 and they earned a trip to the post-season playoffs. The application of big data is credited in part (see, e.g., Sawchik, 2015) to the success of the Pirates—although many factors affect team performance such as talent, injuries, and strength of competition. The Pirates also had winning seasons in 2014 and 2015 and made the playoffs both those years.

2015: Baldrige Criteria & Big Data

The Malcolm Baldrige National Quality Award, administered by the U.S. National Institute of Standards and Technology, explicitly mentioned big data in the Baldrige Criteria: "**Big data**. For all organizations, turning data into knowledge and knowledge into useful strategic insights is the real challenge of big data. . . . In 2015, the Criteria incorporate an enhanced focus on data analytics, data integrity, and cybersecurity." (Baldrige Performance Excellence Program, 2015).

2015: Cargill Corn Milling North America Presentation

Cargill, one of the largest privately held companies in the world, was founded in Conover, Iowa in 1865 (Broehl, 1992). The company's early history was primarily in the agricultural industry, but it is now a highly diversified company that functions in multiple industries. Cargill has shifted its strategic intent from being more commodities-focused to focusing on providing customers with value-added products and services (Broehl, 2008). Cargill Corn Milling North America, one of Cargill's business units, uses big data techniques to provide value-added solutions to farmers. According to Muenzmaier (2015): "We will utilize data offered by farmers to analyze the environmental performance of corn production in the regions where we purchase corn for our operations. These data will, on a consolidated basis, form a baseline of economic performance among several environmental factors including greenhouse gas emissions, water use, and soil erosion, among other factors. Once a baseline is set, we will then utilize these data to determine whether or not we witness improvement among these environmental factors over time. Cargill and its customers will then be able to make claims based on the sustainable production of corn feedstocks used in product ingredients."

Circa 2015: ASA Existential Moment

The American Statistical Association (ASA) was founded in 1839 (see earlier timeline event). There has been an on-going discussion in one of the association's periodicals—the Amstat News—regarding the future role of statistics and statisticians and their potential relationship with data science and data scientists (see, e.g., Jones, 2015; Priestley, 2016; Speidel, 2014; van Dyke et al., 2015). An *existential moment* occurs when someone attempts to answer questions like, "Who am I?" and "What should I become?" The existentialist Sartre (1947) remarked: "Man is nothing else but what he makes of himself." Priestley (2016) commented: "I agree with Tommy Jones in his *Amstat News* article, 'The Identity of Statistics in Data Science,' when he says the '. . . conversation around data science betrays an anxiety about our identity.'" van Dyke et al. (2015) stated:

> "The rise of data science, including Big Data and data analytics, has recently attracted enormous attention in the popular press for its spectacular contributions in a wide range of scholarly disciplines and commercial endeavors. These successes are largely the fruit of the innovative and entrepreneurial spirit that characterize this burgeoning field. Nonetheless, its interdisciplinary nature means that a substantial collaborative effort is needed for it to realize its full potential for productivity and innovation. While there is not yet a consensus on what precisely constitutes data science, three professional communities, all within computer science and/or statistics, are emerging as foundational to data science: (i) **Database Management** enables transformation, conglomeration, and organization of data resources, (ii) **Statistics and Machine Learning** convert data into knowledge,, and (iii) **Distributed and Parallel Systems** provide the computational infrastructure to carry out data analysis."

Time will tell if this introspection results in a *paradigm change* for ASA (see Kuhn, 1970).

2016: Visit to the IBM Website: Watson, Cognitive Computing, & the 4V Framework

A visit to the IBM website reveals insights from one of the global leaders in analytics and big data (www.ibm.com/analytics). The IBM Analytics Technology Platform has six components: Advanced Analytics; Integration and Governance; Data and Content Management; Open Source; Enterprise Content Management; and Cloud Data Services. The infographic found on the website titled *"The FOUR V's of Big Data"* describes the big data dimensions IBM data scientists use: Volume – Scale of Data; Variety – Different Forms of Data; Velocity – Analysis of Streaming Data; and Veracity – Uncertainty of Data. Five IBM employees wrote one of the most comprehensive books on big data which was published in 2015 (Zikopoulos et al., 2015). On the website we also learn about Watson Analytics and the new IBM Cognos Analytics. The exciting aspect about IBM's Watson computer system is that "Watson learns from each interaction and gets smarter with time through its machine learning capabilities." (Zikopoulos et al., 2015, p. 15). Waters (2016) recently wrote an article on the importance of Watson in IBM's future success.

Postscript: Ironically, there was a January 20, 2016 article in the Financial Times (U.S. Edition) by Clover (2016) titled, *"China: When big data meets big brother."* According to Clover: "Critics say China's internet is fast becoming a laboratory where big data meets big brother . . ."

Timeline Observations

- Humans have made significant advances in analytics: from Playfair's pie chart to IBM's cognitive computing; from the Chaco Canyon Sun Priests to big data at Cargill; from Chadwick's baseball scorecard to big data baseball at the Pittsburgh Pirates; and from scientific management at Midvale Steel and Bethlehem Steel to TQM at Tata Steel
- Data collection and analysis activities are not new, but the science keeps evolving
- Statistical methods and analytics have a long history in quality improvement
- Quality, analytics, and big data can be a potential source of competitive advantage

Survey

The survey was completed by representatives of sixteen organizations of varying types including eight manufacturing organizations, three health care organizations, one service organization, and four government entities. Eight of the sixteen organizations have international operations and the organizations are geographically dispersed throughout the United States. Six survey questions in the context of *performance measurement* and *improvement* were asked:

1. What is your organization doing that is new in terms of data analysis?
2. How has data analysis changed in your organization over the past three years?
3. What data analysis trends are occurring in your industry?
4. Which data analysis tools are used most in your organization?
5. What is an example of how data analysis was used to improve performance?
6. What are the greatest challenges to effective data analysis in your organization?

Selected Responses to the Survey Questions

Question #1: What is your organization doing that is new in terms of data analysis?

Summary of the Responses

We are . . .

- establishing a data governance structure
- implementing real-time dashboards
- analyzing customers in more depth
- being more systematic in analyzing new product and service ideas
- trying to be more effective at predicting performance outcomes
- analyzing our overall cost structure
- doing more with data visualization
- studying population indicators

Selected Responses

"We are doing more in-depth analysis of customers (segmentation) and their locations, product offerings, transportation routes, and overall costs to determine how we will define our business, how we will approach customers, and how we will ensure profitability in the future."

"We have developed a systematic way to analyze new product or service ideas in order to prioritize promising ideas and provide future growth. This process enables us to prioritize new ideas in the pipeline."

"Our data governance project is developing the policies, processes, procedures, organization, and technologies required to leverage data as an enterprise asset."

"As a science-based institution, information is absolutely fundamental to our organization's success. Our goal is that data will function as an information foundation that is trustworthy, integrated, consistent, and readily accessible to staff and the public. This will help staff, decision makers, and the public have the best information as possible as they work together to address complex challenges . . ."

"We are using multi-variable statistical analytics to determine the key factors that most affect business performance and to optimize those factors for improved performance."

"We are looking at current data instead of relying on older data. Historically some of our data was up to a year removed from when the actual date the data point was derived."

"We are focused more now on optimization and real-time data collection and analysis."

"We are using predictive analytics more now to improve forecasts for both internal data and customer-owned data."

"Some Lean initiatives are shedding light on metrics for manufacturing performance. Not much is new in terms of overall company performance at the Tier 1 Level. We tried to promote and use SPC for the metrics and it didn't take root with the executives."

"Our Business Intelligence Group is adopting data visualization tools and techniques."

"We are implementing Tableau Software as a primary data analysis, visualization, and reporting tool for our entire organization. User support includes administration of Tableau Enterprise Server as well as program-specific customized reporting and in-house software training."

"We have collected data on individual patients for a long time, but now we are developing capabilities to collect and analyze data at the network population level."

Question #2: How has data analysis changed in your organization?

Summary of the Responses

We are . . .

- focusing more on the quality of our data
- tracking performance in more areas of our organization
- holding more people accountable with our performance metrics
- integrating performance measures into the way we report and do our work
- using more automated real-time online dashboards

- using analytics more to predict and influence business outcomes
- requiring a higher degree of statistical proof during decision making
- now requiring that all leaders in the organization have quantitative skills

Selected Responses

"We are in the process of moving the entire organization to a new ERP [Enterprise Resource Planning] system. This requires much cleaner starting data, better definitions of terms and the meaning of data, and aligns reporting across the business units that have implemented the system. The plan would be to be able to analyze customer information and dig deeper into profitability, specifically for customers served by multiple business units."

"We have more data now through the use of automation."

"There are now standardized on-line dashboards available to all clinical departments."

"Looking at real-time data is a change. Also, having all performance data on the intranet for all employees to view is relatively new in the last three years, e.g., any employee can see any physician's quality data."

"There is more emphasis on understanding costs and determining how we will meet future growth projections via organic growth, mergers & acquisitions, and new products & services."

"I would characterize data analysis in our organization as potentially more sophisticated, but less pervasive with Six Sigma, as a quality improvement discipline, on the decline and data visualization having emerged more prominently in the business intelligence arena."

"We initially used analytics to determine why something happened after the fact, now we are using analytics to be able to better predict and influence business outcomes."

"Due both to increased data collection skills used during improvement projects as well as an increase in our staff's familiarity with data analysis, we have been better able to show improvements using data."

"In the past three budget cycles (actually covering six years), we have integrated performance measures into the way we report and plan our work, and using these tools to report to stakeholders about progress on meeting goals and strategies, and expenditure of public funds."

"Within the last four years we defined four Critical Success Factors (CSFs) and we aligned our Tier 1 and Tier 2 metrics to the four CSFs. Each Tier 1 and Tier 2 metric has a defined target (desired performance level) and thresholds where formal action is required. We've integrated these metrics into our monthly performance review meetings. Action can be in the form of a corrective action led by a Six Sigma trained Green Belt or a Six Sigma project led by a Black Belt."

"Improving performance measurement has increased management's ability to analyze data to support decision making. As such, there has been a growing interest in data quality. This results mainly from analysts highlighting uncertainty in an analysis, or the outright inability to conduct a reasonable analysis, owing to poor data quality. Poor data quality stems from inconsistent data collection methods and poor QA [Quality Assurance] in data systems. Ultimately, data analysis is changing the way we collect and use data."

"In the last couple of years we have put more focus on doing a better job of collecting, analyzing and displaying the analytical results from our continuous improvement projects."

"It has gone from a niche skillset that only a few people had to a critical skillset that all leaders must possess to be successful."

Question #3: What data analysis trends are occurring in your industry?

Summary of the Responses

- Better end-to-end integration of data across customer lifecycles
- Everyone is moving to enterprise-wide information systems
- More use of social media metrics
- Trying to correlate positive social media reviews with performance
- More uniformity in measuring performance ("apples to apples")
- More public reporting and data transparency
- Higher degree of statistical proof required during decision making
- Big Data (predictive analytics) is the new trend in health care

Selected Responses

"I know that our competitors are moving to ERP [Enterprise Resource Planning] systems (as we are), probably for the same reasons – replace old legacy systems, provide better access to data, and standardize data for business intelligence reporting and analysis."

"State agencies are looking for uniformity in reporting project status to the Governor's Office. Agreement on and standard use of a Statewide CI [Continuous Improvement] Project Tracker has recently been established and are gaining momentum."

"In state government work, there is an increasing demand by citizens for self-service data and transparency in reporting."

"We are using big data sets with a focus on visual exploration of the data. We are also working on a better end-to-end integration of the data across our product, services, and customer lifecycles."

"Big Data analysis tools are new to our industry – tools to analyze large amounts of data at once."

"Big Data is the new trend in health care."

"Predictive analytics for healthcare quality and transparency of data."

"More emphasis on regulatory mandated data collection vs. organizational priorities. More automation of data collection (i.e., Meaningful Use)."

"A much higher degree of statistical proof in decision making is required now in our organization."

Question #4: Which data analytic tools are used most in your organization?

Sampling of the Responses

- Time Series Plots
- Statistical Process Control
- "Top Box" Scores
- Process Capability Analysis
- Logistic Regression
- Multivariate Analysis
- Linear Programming
- Propensity Score Matching
- Structural Equation Modeling
- "What if" Scenario Modeling
- Software: Excel, Access, R, Tableau, Minitab, etc.

Selected Responses

"We use graphical data representations (e.g., bar chart, pie chart, Pareto chart) to show change and/or make comparisons in data from projects. We are working to mature our data analysis within our Office of Continuous Improvement to incorporate more proactive, predictive measurement of information and consistent project tracking, rather than reactively analyzing data, but this is a gradual growth process."

"Excel is the most commonly used data analytic tool."

"Trend analysis. Mean performance. Twelve month rolling averages."

"Simple SPC at best, but not much for analytical tools. There are pockets of visual trending analysis in departments."

"Averages and Top Box Scores."

Question #5: What is an example of how data analysis was used to improve performance?

Sampling of Responses

"We utilized cluster analysis to determine which product offers were most meaningful to our customers allowing us to improve both product margin and customer experience."

"We used a data driven method to reduce our emergency room 'left without being seen' metric from a high rate to an exceptionally low rate."

"Warranty claims analysis helped identify defect frequency which led to a failure mode analysis which led to a process change which led to a reduction in warranty payments."

"We used logistic regression to determine which attributes contributed most to the ability to resolve critical incidents within the targeted resolution time. As a result of the analysis, we were able to help a technology service provider significantly improve performance in incident resolution."

"Prior to our product campaign last year we spent a great deal of time analyzing the profitability of our customers. This required in-depth analysis of all our cost drivers, challenging assumptions, and questioning existing data. This work led to rationalizing customers and products helping us to increase overall margins while still maximizing the production at our facilities."

"One of our divisions is moving to an electronics records system . . . The new records system will provide just-in-time reporting . . . The new system will advise the appropriate staff of the need for attention to the spot; via the Web, it will advise the public of potential issues at the location."

"Clinical care for diabetic patients – standardized definitions of optimal patient results. Online tools are utilized to share performance results at the patient level for the entire diabetic population. Outliers are easily identifiable and standard work is rolled out to get the patients the care they need."

"One of our divisions use data extensively . . . data is collected, interpreted, and analyzed in consultation with a variety of stakeholders and as a result of these deliberations . . . regulations are changed."

Question #6: What are the greatest challenges to effective data analysis?

Summary of the Responses

- Getting accurate data
- Managing our data
- Turning our data into meaningful information
- Collecting more meaningful data and less meaningless data
- Interpreting our data
- Telling a compelling story based on our data
- Getting people to "believe" the data—data is questioned
- Making sure the appropriate people have access to the data
- Determining what to do about social media
- Deciding how to handle "instant feedback"
- Finding people with excellent quantitative skills
- Creating a data-oriented culture

Selected Responses

"Our data is not always 'connected' across the [customer] lifecycle, so there is a need to have better horizontal integration of the data."

"In many cases, leaders aren't using the data and metrics to drive priority-setting and/or improvement initiatives."

"Understanding of the power of analysis and acceptance of the data instead of gut for decisions."

"We have the ability to get data on almost anything. Being able to take this data and convert it quickly into meaningful information and understanding what the information is telling us is a challenge. Being able to cut through all the superfluous information and make good, data-driven decisions is a skill our leaders will need."

"Expertise in using the analysis to tell a compelling, and yet easily digestible, story."

"We don't sell many consumer products [directly] but our customers use our ingredients in their products. How social media will play into this and how we can use 'instant feedback' is something we will also need to understand."

"Funding for tools and expertise to fully realize associated capabilities."

"Master Data Management ensuring that we have the right data foundation and quality of the data to perform the analytics. We still have to spend a large amount of time on data cleanup prior to running the analytics."

"We lack meaningful data: We struggle with finding and collecting data which is meaningful to identifying issues, tracking trends, and effectively measuring performance."

"We have an overwhelming amount of unnecessary data. We often find our organization tracking data without asking 'Why?' resulting in the creation of unnecessary reports, sending of unnecessary information, and time spent on these activities which could be used elsewhere."

"Fostering a wider culture of using data to improve performance."

"Data standardization and integration to foster evaluation across program boundaries."

"The availability of accurate data at the customer and program levels."

"The quality of the data, the availability of the data, and the skill sets of the people performing analytics around our data."

"Belief in the data. Belief in the metric. Belief in the process."

"Team members continue to question the validity of the data."

"The amount of data is huge and the number of people who can effectively change it into meaningful information is small."

"We always hope the data that is available will help us tell the story of where processes are breaking down."

"We collect a lot of data in our organization, but finding the time and human resources to meet the demand for in-depth data analysis is an on-going challenge."

"We are unable to effectively quantify the uncertainty deriving from questionable data quality."

"Data accuracy is a challenge because data is being collected differently at nine facilities. We need to create a company-wide standard."

"Data that is collected and housed in different functions of the organization in different formats."

"One challenge is the availability of trained and knowledgeable data scientists."

"We have a limited understanding of how non-financial data can be used for decision support."

"We have low accessibility to the data that currently resides within the organization."

"For data analysis to be widely valued in our organization, we need decision-makers to ask questions that can only be properly answered using facts and data. Gut feel and intuition are still highly valued in this organization. When an organization is able to grow profitability without the use of deep data analysis, there is no burning platform for more data analysis."

"We don't have people with the necessary skills to be effective at data analysis."

"I think we are moving into an exciting period where we will be able to do a much better job of business intelligence with common systems across the company. Having standardized meanings for common terms (gross margin, net profit, etc.) will help immensely. Being able to see which customers are the most profitable company-wide will also help us with *go to market* strategies across business units. How new data sources such as Twitter and other social media will affect us is unknown, but will probably force a new paradigm."

"We are working on basic elementary data analysis for the most part and there are pockets of those interested in higher-level analysis, but it is a challenge with leadership decisions that may act contrary or contradictory towards long-term sustainability considerations."

"Some of the reluctance to data analysis seems to come from the fear that a lot more work will result because of the analysis."

"Too many companies jump into new software to see improved data. Companies need to focus their efforts on the simple day-to-day data of their four Key Performance Indicators (Safety/Quality/Delivery/Cost) to view company performance, and drive daily improvements to improve those metrics."

Limitations of the Study

The aim of the study was to learn as much as possible about quality, analytics, and big data within the time and budget constraints of the research project. The companies and universities that were studied and the sixteen organizations that participated in the survey were chosen by judgment and/or convenience. It is impossible to generalize the findings to "all organizations" either in the U.S. or any other country. Therefore, this report does not make universal conclusions. The findings in this report (1) summarized what was learned from these particular organizations and (2) provided a basis for developing emergent themes, an emergent theory, an emergent definition, and ideas for better practices. Further research is necessary to reach more definitive conclusions.

III. Emergent Themes

Emergent Themes

Eight themes emerged from the within-respondent and cross-respondent analyses of the participant responses:

- Quantitative Confusion
- Broader View of Data
- Fast-Paced Analytical Changes
- Data Governance and Management
- Data Integrity
- Data Interpretation
- Analytical Capabilities
- Data-Oriented Culture

Quantitative Confusion: Many respondents said they have become confused about the distinctions between quantitative-related terms used in the media like *data, information, statistics*, *statistician*, *analytics*, *business intelligence*, *data science*, *data scientist*, and *big data*. Most of the respondents reported that their organization is re-thinking the roles data collection and analysis play in efforts to improve organizational performance. Some respondents thought that—at a minimum—their organization should define some of those terms for their own purposes.

Broader View of Data

Most of the respondents stated that their organization now collects more types of data and so a broader view of data is required. Where once the common data types were continuous data (e.g., data values in decimal form) and discrete data (e.g., categorical data)—now they have voice data, pictures, video clips, search queries, social media comments, and big data. The "new" types of data require different analytical tools and organizational capabilities.

Fast-Paced Analytical Changes: Most of the respondents commented that the changes in the analytical practices that were occurring in their industry were occurring at a fast pace—especially changes related to enterprise-wide systems, mobile capabilities, real-time data collection, and social media usage. Some further commented that their organization needs to get better at detecting analytical capability advances in their industry.

Data Governance and Management: Several of the respondents reported that their organization was putting in place a data governance structure and developing enterprise-wide data management capabilities. Decisions are being made regarding who owns analytics, who owns what data, who sets data standards, who has data decisions rights, and whose budget data management falls under. Some organizations are deciding whether to have an Analytics Team.

Data Integrity: Most of the respondents reported that data integrity is becoming increasingly important because data is being used more often to make decisions and people need to trust the data. Some commented that it is important to understand the sources of data and to have an assessment of the credibility of the sources. Generally, data should be accurate, complete, and represent some sort of *truth*. Some respondents commented on the importance of measurement system analysis so that the risks associated with data are better known. Also, many respondents commented on the need for everyone in their organization to have critical thinking skills in order to critically evaluate and challenge the data.

Data Interpretation: Most of the respondents said that their organization needs to get better at interpreting data and creating compelling stories. "We now have all this data, but what does it mean?" Interpretation skills are important because the consumers of charts and graphs need to be able to quickly extract the key points. Also, if someone is presented with a correlation coefficient, then they need to know what it is and what information it conveys in order to use that information for decision making.

Analytical Capabilities: Most of the respondents commented that their organization is working to develop the analytical capabilities of employees. Each person/position might need a target *analytics knowledge and skill level* requiring a personal development plan. It was also mentioned that executives need analytical capabilities because they are the consumers of analytics and big data outputs; they make important decisions, and they direct and control analytical resources. Some respondents didn't think their organization was capable of evaluating analytical talent yet. It is difficult to assess organizational analytical capabilities without in-depth knowledge of analytics.

Data-Oriented Culture. Some respondents said that their organization is attempting to create a data-oriented culture. Some characteristics of a data-oriented culture could be the following: leaders ask for data, decisions are typically made based on data, performance dashboards are visible throughout the organization, employees are trained on data collection and analysis, analytics is incorporated into leadership development, and investments are made in developing analytical capabilities. You would eventually see data-oriented values, behaviors, artifacts, policies, practices, and rituals. An organization with a *data-oriented culture* would have employees who have *a pattern of shared basic assumptions* and a clear sense of *the correct way to perceive, think, and feel* related to the collection and analysis of data (see Schein, 2004). An organization structure can be changed quickly *towards analytics* through the creation of analytics positions and teams, but a data-oriented culture might take a long time to create.

IV. Emergent Theory & Ideas for Better Practices

The emergent theory from the research is that *the integration of quality, analytics, and big data guided by the strategic direction of the organization can result in new sources of customer value and a new source of competitive advantage* (or *competence* in the case of non-profit organizations). This is depicted in Figure 5. This could eventually contribute to the achievement of quality superiority and improved financial performance.

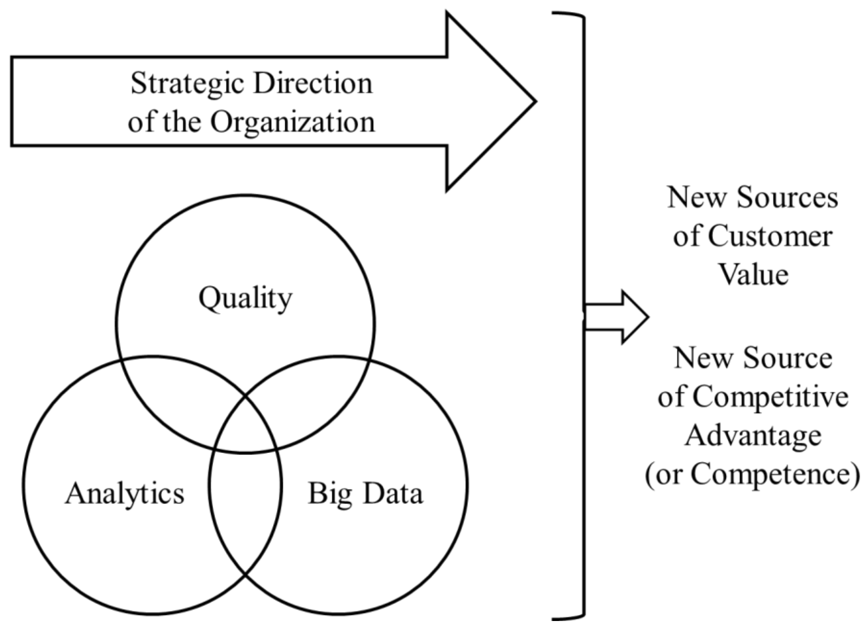

Figure 5. Emergent Theory.

Strategic direction relates to the mission, vision, values, priorities, performance measures, strategic objectives, and strategies. Ten ideas for better integrated quality, analytics, and big data practices will now be presented and discussed:

- Have superior quality as a strategic intent
- Model customer experiences
- Broaden your view of data
- Adopt system thinking
- Apply the scientific method
- Enhance the organization structure
- Develop well-rounded quality and analytics professionals
- Learn from variation
- Identify and eliminate waste
- Manage customer data with great care

1. Have Superior Quality as a Strategic Intent

Many organizations strive to achieve product and service *quality superiority*, but not all will succeed and so their quality will be worse than the competition. Deming stated (1986, p. 5): "The consumer is the most important part of the production line. Quality should be aimed at the needs of the consumer, present and future. Quality begins with intent, which is fixed by management." We know more about how to achieve quality superiority than preserving it once attained. "***Quality superiority is a strategic intent"*** was one of six factors identified as being important for preserving quality superiority (Liedtke et al., 2010). Analytics and big data techniques can be used to generate new insights into customers and identify quality improvement opportunities.

A general formula for "Value" is "Quality" divided by "Cost." If we borrow this definition for an information (or data) context, then a formula for the "Value of Information" could be the following (Liedtke, 2015):

$$V_{I,t} = Q_{I,t} / C_{I,t}$$

The value (V) of a "piece" of information (I) at a point in time (t) is the ratio of the quality (Q) of that piece of information (I) at that point in time (t) divided by the cost (C) of that piece of information (I) at that point in time (t). Quality and cost could be further defined:

$Q_{I,t}$ = f (Relevancy, Accuracy, Completeness, Exclusivity, Monetizability, etc.)
$C_{I,t}$ = f (Money, Time, Storage Space, Risks, etc.)

Time (t) can be an important dimension in valuing information. For example, the announcement of a future acquisition (information) is probably more valuable to an investor a few minutes after the announcement than one year later. We could alternatively think of Value, Quality, and Cost as being either *cumulative* or *predicted*. An organization could strive to achieve *information* (or *data*) *quality superiority* in addition to achieving product and service quality superiority. As the quality of information increases—assuming the cost of information remains constant—then the value of information will increase. Some have suggested that big data is sometimes *messy* (see, e.g., Mayer-Schonberger & Cukier, 2013). Focusing quality improvement concepts, methods, and tools on improving information quality (e.g., increasing accuracy) could increase the value of information and decision quality. Quality controls can be implemented to minimize data errors at their source whenever possible.

Six Sigma projects led by Black Belts or Master Black Belts could be launched to increase the information (data) quality and/or decrease information costs. Those improvement experts and their cross-functional teams could attack known information quality and cost issues. They could use the DMAIC method and the accompanying statistical tools or the QC Story (see, e.g., Hosotani, 1992; Kume, 1985). A portfolio of such projects would potentially increase the value of information; improve decision quality, and improve product and service quality. They could also help improve the effectiveness and efficiency of analytics and big data initiatives. More valuable information might reveal new sources of customer value and result in a competitive advantage.

2. Model Customer Experiences

Customers represent a critical stakeholder group for many organizations because they judge product and service quality and they are the primary source of revenue. For a non-profit organization or government entity—customers can be critical to the organization's mission. Customers base their product and service quality judgments on their experiences. For example, suppose someone wants to purchase a new laptop computer. Figure 6 (Liedtke, 2015) depicts some of their likely experiences.

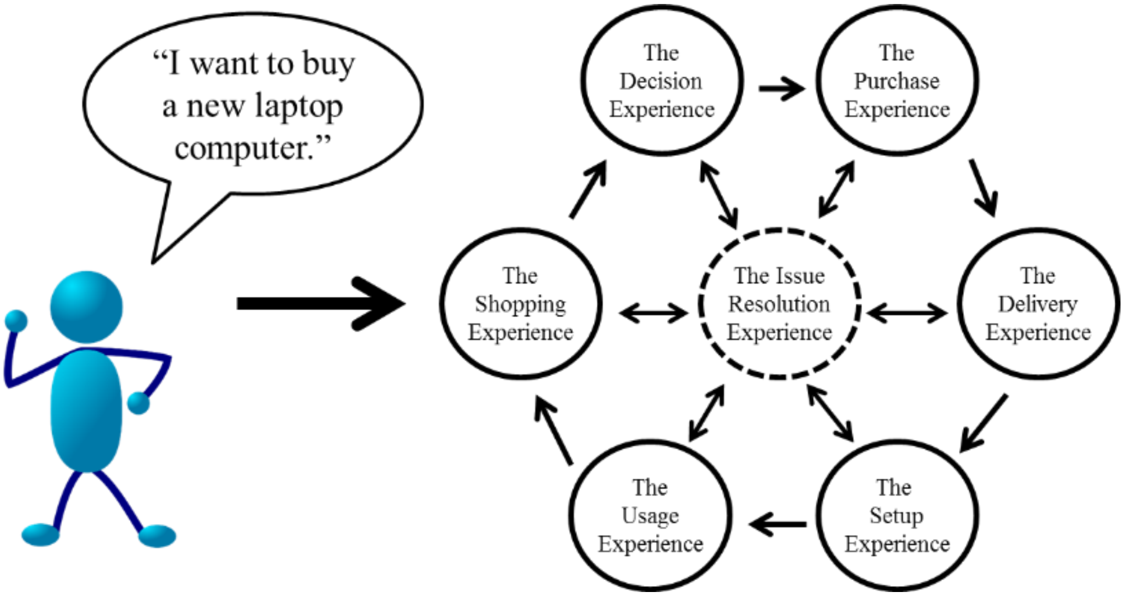

Figure 6. An Example of a Customer Experience Model.

The *shopping experience* might consist of on-line browsing and/or store visits resulting in a list of options. Easy website navigation might be important to on-line shoppers. Website analytics and clickstream analysis could be useful in better understanding the experiences of on-line shoppers. The *decision experience* would involve the person choosing one of the options. The *purchase experience* would involve the person (now customer) placing the order on-line or paying for the product at a store counter. An intuitive on-line payment process or not having to wait in line at a store might be important. The *delivery experience* would involve waiting for the delivery of the product and the actual delivery itself. On-line status updates and late delivery notifications might be useful to the customer. Data on customer delivery inquiries and queries could be analyzed. The *setup experience* would involve the customer unpacking the laptop computer and attempting to get it up-and-running. Instructions and product support contact information might be important to the customer. The *usage experience* would involve the customer using the laptop computer. Access to product support resources might be important to the customer. The customer could encounter an issue at any time during the process and so there might be an *issue resolution experience*. Short wait times and issue resolution status updates might be important to the customer.

The laptop computer company can develop analytics and big data capabilities for each customer experience (across all customers) in order to gain new customer insights; evaluate company performance on each customer experience; and identify quality improvement opportunities. The company could create a dashboard for each customer experience that is periodically reviewed. The company could also develop an improvement plan and create a project portfolio for each customer experience. We would expect that if a customer had *all positive experiences*, then they would be more likely to judge product and service quality favorably. Analytics and big data techniques applied to each customer experience (across all customers) might lead to the identification of new customer segments and reveal new opportunities for identifying new sources of customer value.

3. Broaden Your View of *Data*

Those organizations attempting to achieve product and service quality superiority can benefit from a broadened view of data. *Variety* is one of the dimensions of big data (Zikopoulos et al., 2015). No longer do organizations only possess customer focus group and/or customer survey data. Now organizations potentially have voice data, pictures, on-line videos, search query information, sensor data, wearables data, social media comments, *likes*, on-line product reviews, and tweets. Each data type can be weighted based on data quality dimensions like data source credibility, relevancy, accuracy, completeness, and monetizability.

If the laptop computer company (mentioned above) launches a new laptop computer model in the market, then they will want to understand its performance: "Do customers like our new product?" There will be multiple data sources available to the company—some examples are depicted in Figure 7 (Liedtke, 2015). Analysis of the data might reveal new customer insights and quality improvement opportunities. Big data techniques could potentially be used if there was a large amount of structured and unstructured data from a variety of sources arriving in real-time.

The analytical activities used to analyze all of the data would be analogous to some of the activities a jury carries out during a trial. A jury is presented with various types of evidence during a trial and then they must reach a verdict. The evidence could include testimonies, forensic lab results, pathology reports, pictures, surveillance videos, public records, phone logs, etc. Some of the pieces of evidence might be more credible and/or compelling than others. Each jury member must "weigh and consider" all the evidence during the trial to form an opinion. The jury members are given instructions from the judge to guide their deliberations. Jury members might be persuaded to change their opinion during deliberations. In the case of the laptop computer company that wanted to know if customers liked the new laptop computer model, the quality and analytics professionals—as part of a cross-functional team—could "weigh and consider" the various *types of data* evidence and "deliberate" in order to form a collective answer to the question.

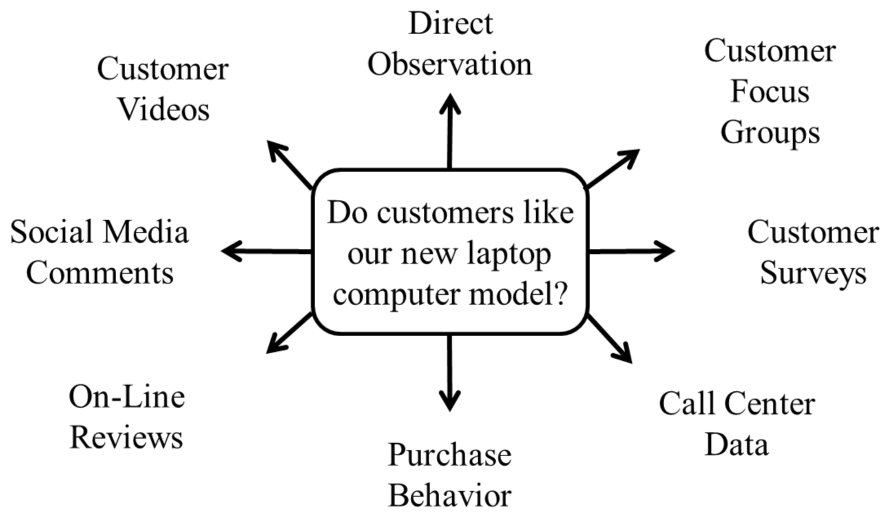

Figure 7. Potential Data Sources.

A definition of big data that only considers the *volume* of data is not very useful because "big" is relative and it suggests that "bigger" data exists thus requiring a different adjective. One cause of confusion in the minds of practitioners is that so many big data definitions exist (see, e.g., Press, 2014). It was mentioned earlier that Laney (2001) is credited with the "3V" framework (*volume*, *velocity*, & *variety*) and IBM currently uses the "4V" framework which includes *veracity* (see Zikopoulos et al., 2015). It can be argued that neither of these frameworks is comprehensive enough to adequately describe the phenomenon known as *big data*. What follows (Liedtke, 2015) is an emergent definition of big data based upon a review of the literature and a study of the practices of some leading big data companies:

> ***Big data*** is a relatively large amount of data consisting of multiple types from multiple sources possibly arriving in real-time of varying degrees of accuracy requiring exploratory data analysis and integrative analytical methods.

Although not perfect, this is a broader definition than the "V" frameworks. Here is the same definition with the addition of descriptors to show how it differs from the "4V" framework:

Big data is a . . .

relatively large amount of data	**Volume**
consisting of multiple types	**Variety**
from multiple sources	Sources
possibly arriving in real-time	**Velocity**
of varying degrees of accuracy	**Veracity**
requiring exploratory data analysis	Exploration
and integrative analytical methods.	Integration/Synthesis

It should be noted that big data can be used for solving problems and discovering problems.

If *big data* exists, then we can imagine other data sizes—both smaller and larger. This is depicted in Figure 8 (Liedtke, 2014). This author coined the term *zero data* to refer to items such as the organization's mission, vision, values, strategic objectives, etc. The keys to success with zero data might be related to wisdom, experience, intuition, passion, insight, and judgment.

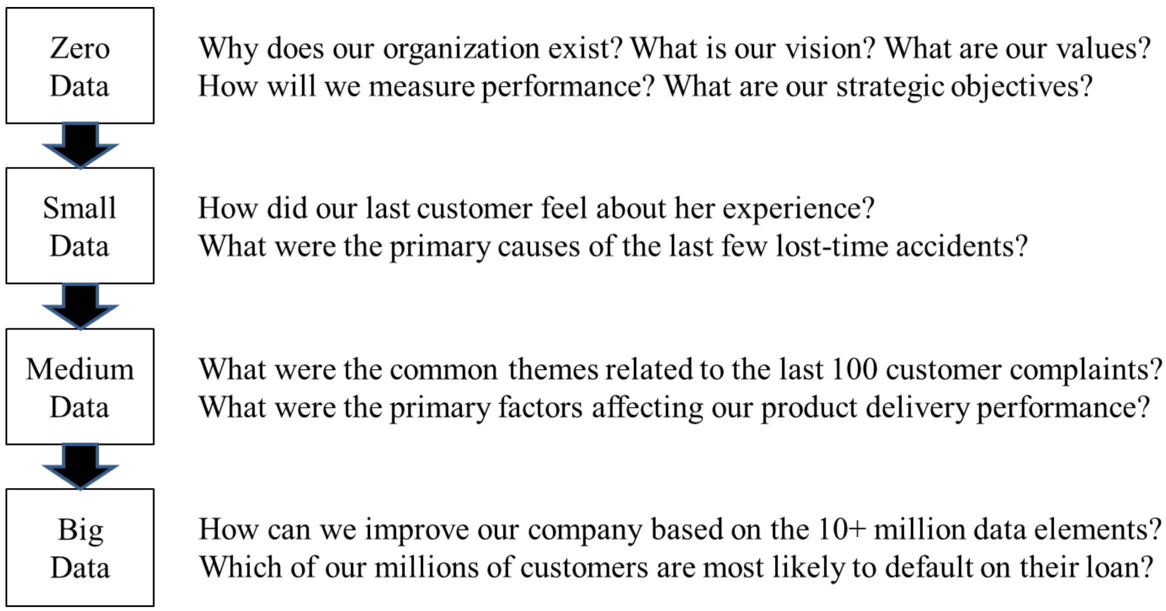

Figure 8. Zero Data to Big Data.

One could argue that we should start with zero data because little else matters from a strategic perspective. Small data is a situation where the phenomenon of interest is small in scale and scope. "How did our last customer feel about her experience?" This is not a big data situation and yet it is important to the company's relationship with that customer. Data collection and analysis could involve a customer history review and interview. Here is a different example: "What were some of the causes of our last five lost-time accidents?" This is probably not a big data situation. Root cause analysis and the case study method (Yin, 2014) could be useful in this situation. Medium data could involve a situation where we have hundreds to tens of thousands of data values. "What were the common themes related to the last 100 complaints?" Qualitative research tools might be useful in this situation (see Denzin & Lincoln, 2011). "What were the primary factors affecting our product delivery performance in our Western Region last month?" We might use regression or multivariate analysis techniques (see, e.g., Johnson & Wichern, 2014) to answer that question.

An organization not having zero, small, and medium data capabilities should be careful in developing only big data capabilities. It is possible to become overly-focused on information technology. Porter (2001) commented: "Even well-established, well-run companies have been thrown off track by the Internet. Forgetting what they stand for or what makes them unique, they have rushed to implement hot Internet applications and copy the offerings of dot-coms." Combining zero data with the appropriate information technology appears to be important.

4. Adopt System Thinking

Deming (1994, p. 95) offered a definition of a system: "A system is a network of interdependent components that work together to accomplish the aim of the system." If the aim of an organization (system) is to achieve quality superiority and analytics and big data techniques are used, then some component systems could be customers, analytics team members, information systems, databases, information, sensors, satellites, code, algorithms, servers, mobile devices, etc. The components would have to interact in special ways in order to achieve quality superiority. Figure 9 (Liedtke, 2014) depicts a system analytics view of an organization which bears some resemblance to a diagram Deming (1986, p. 4) introduced as *Production Viewed as a System*.

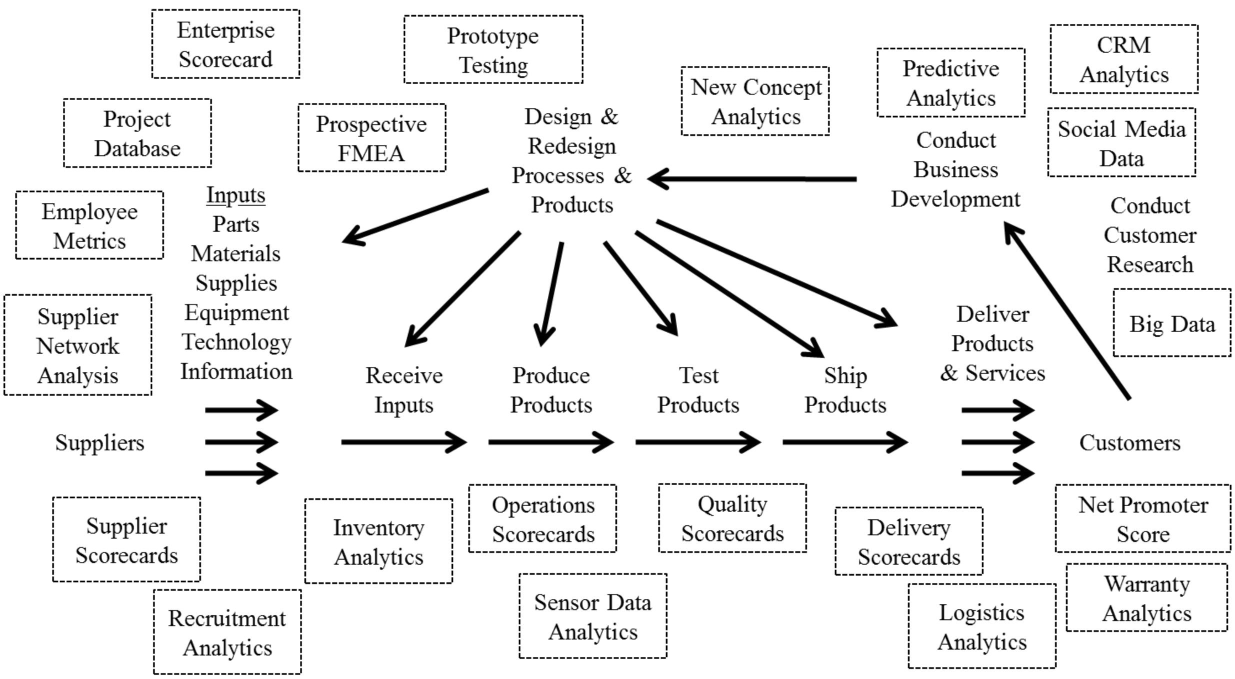

Figure 9. System Analytics View of an Organization.

The interior arrows and non-boxed descriptors represent a revised version of Deming's diagram. The boxed items represent a hypothetical inventory of analytical practices. We see that analytical activities can occur throughout the entire system. This diagram is useful from an improvement perspective because it shows the major organizational processes and their flow. Perhaps most importantly, the analytics system is aimed at customers. This view could also be used for planning and resource allocation. Analytics and big data techniques could help identify quality improvement opportunities. If we had access to all of the analytics system data, then big data techniques could be applied because there would be *a large amount of data consisting of multiple types from multiple sources possibly arriving in real-time of varying degrees of accuracy requiring exploratory data analysis and integrative analytical methods*. Big data analysis might identify new sources of customer value and lead to a new competitive advantage.

5. Apply the Scientific Method

The scientific method has played an important role in quality improvement activities for decades (see, e.g., Box, Hunter, & Hunter, 1978; Deming, 1986, 1994; Ishikawa, 1985; Shewhart, 1939). The Plan-Do-Check-Act (PDCA) cycle (Ishikawa, 1985, p. 59) and the Plan-Do-Study-Act (PDSA) cycle (Deming, 1994, p. 132) are versions of the scientific method that are widely used to improve quality. If someone has an idea for change, then the PDSA cycle involves planning the change (Plan), implementing the change (Do), studying the adherence to the plan and the results of the change (Study), and then taking action on what was learned (Act). Figure 10 (Liedtke, 2014) depicts two levels of the PDSA cycle that can be used to help integrate quality, analytics, and big data activities. The boxes with dashed-line borders represent the "larger" PDSA cycle comprised mostly of analytics activities and decision making and then there is an interior PDSA cycle that occurs within the Act Step of the larger PDSA cycle. Someone might begin with a complex

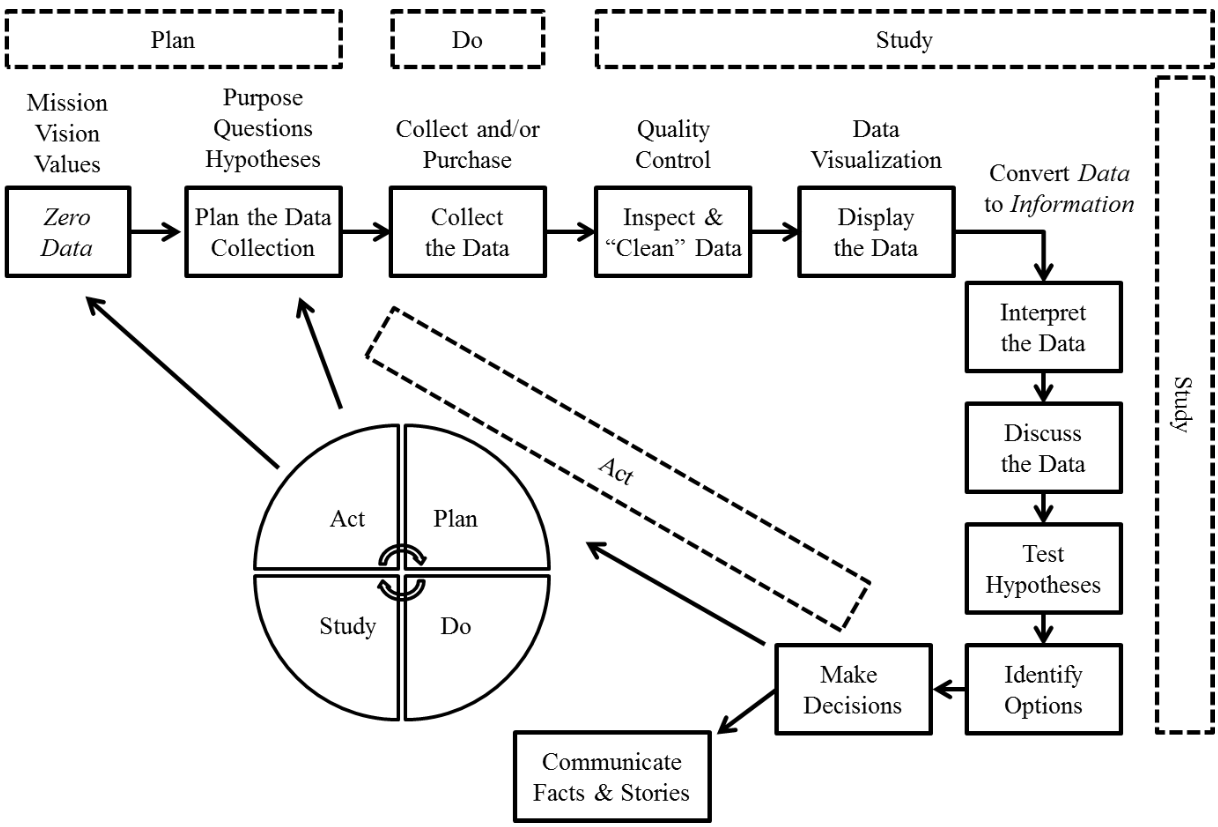

Figure 10. Emergent PDSA-Based Analytics Model.

problem to solve as is the case with a Six Sigma project. This would result in the implementation of solutions. Alternatively, the starting point could be "let's see what we find" from data mining. That represents problem/opportunity discovery and could result in the creation of hypotheses.

Zero data could be a good starting point in order to remind people of the organization's mission, vision, values, etc. The next step involves planning the data collection which clarifies the purpose of the data collection and formalizes the questions you would like to answer and/or the hypotheses you would like to test. The data is then collected—potentially a large volume of data consisting of multiple data types from multiple sources. We then inspect and "clean" (delete/modify) the data. This step can inform us of our decision risk. We then summarize and display the data (see, e.g., Cleveland, 1985; Tufte, 2001; Tukey, 1977; Yau, 2013). Data visualization is one area where there have been significant advances in analytics software (e.g., business intelligence software) and mobile dashboard/scorecard display technologies. A current popular term is *infographics*.

The data is *converted* to information at some point. The data is interpreted and discussed and then hypotheses might be tested. Options might be identified for "what to do" and decisions are made. This might lead to two activities: the communication of facts and stories and/or the implementation of an *idea for change* using the PDSA cycle. This is a simplified model, but possibly a good starting point. Quality professionals might consider adding information system symbols to their company's actual model. Some symbols are depicted in Figure 11 (Liedtke, 2015).

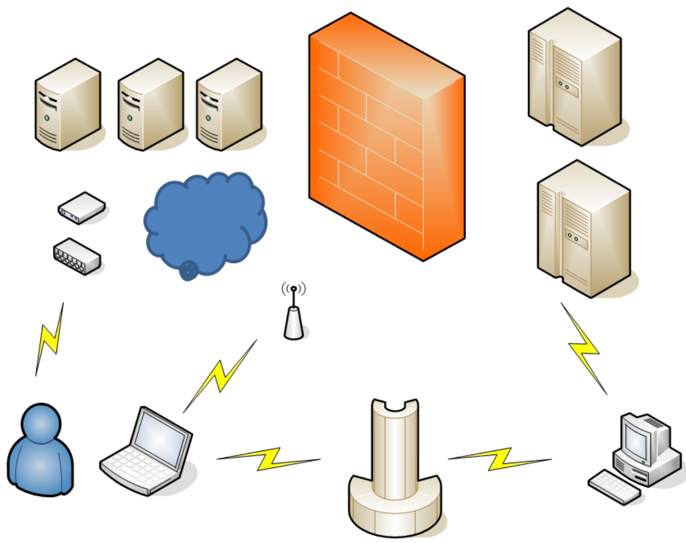

Figure 11. Some Information System Symbols.

Nonaka and Takeuchi (1995, p. 62) identified four knowledge conversion modes related to tacit and explicit knowledge: socialization (tacit to tacit); externalization (tacit to explicit); combination (explicit to explicit); and internalization (explicit to tacit). There now exist relatively new data sources that can help organizations create knowledge. For example, conversations on social media can create new tacit knowledge through the sharing of experiences (socialization). Also, organization members can "gain a better feeling" about what customers think of their company's products by watching customer-posted product usage videos—here creating tacit knowledge from explicit knowledge (internalization). An organization can potentially identify new sources of customer value and gain a competitive advantage by rigorously applying PDSA to analytics.

6. Enhance the Organization Structure

Organization structure can be used to integrate quality, analytics, and big data. Senior executives must decide where the quality and analytics professionals will reside in the organization structure: "Should they be centrally located on an analytics team or should they be distributed throughout the organization?" The positioning of these professionals will affect their roles and responsibilities and their effectiveness in helping the organization achieve superior quality. An organization that is just beginning its analytics journey might consider forming a temporary Analytics Team. Members could include representatives from the major organizational units such as Sales, Marketing, Engineering, Operations, Human Resources, Information Technology, Quality, Finance, and Risk. A more permanent organizational unit is depicted in Figure 12 (Liedtke, 2015). The Office of Strategic Improvement could oversee five areas: Strategy, Quality, Analytics, Improvement, and Innovation. The head of the office could be a member of the Leadership Team. A more detailed view of the office is depicted in Figure 13 (Liedtke, 2015).

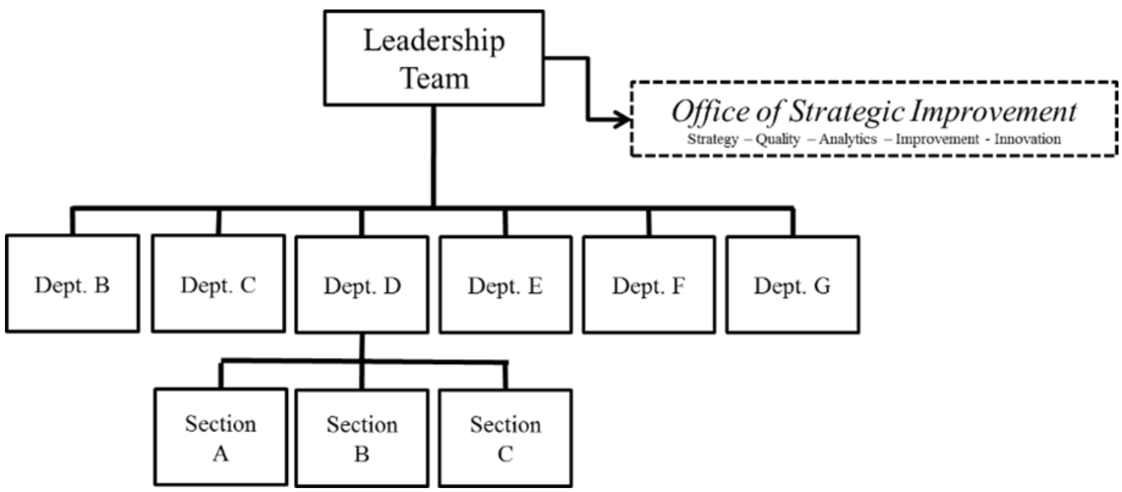

Figure 12. The Office of Strategic Improvement.

Strategy could encompass strategic planning and/or *hoshin kanri*. Quality could oversee the quality management system including dashboards/scorecards and quality reviews. Improvement could oversee improvement activities such as Rapid Tests of Change, Lean projects, and Six Sigma projects. Innovation could oversee new product and new service development activities and Design for Six Sigma projects. Lastly, Analytics could oversee both qualitative and quantitative analytical resources. Analytics could itself have seven knowledge domains: Qualitative Tools, Traditional Voice of the Customer, Basic Quality Tools, Strategy Tools, Statistical Methods, the Internet of Things, and Big Data. Data governance would an important responsibility of the office—clarifying who has various data decision rights and who develops enterprise-wide analytics standards.

The structure might need to be flexible to successfully accommodate emergent strategies (see, e.g., Mintzberg, 1994) and strategic issues. There could be multiple A Teams (analytics teams) executing quality, analytics, and/or big data projects at any point in time.

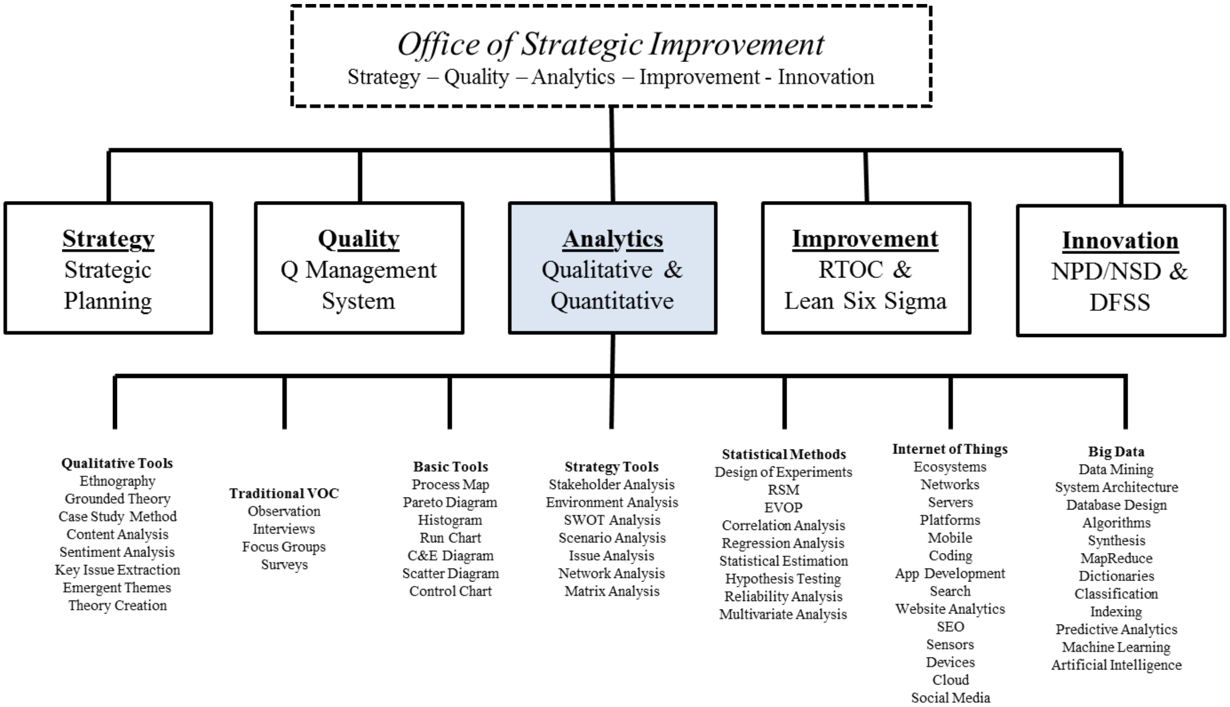

Figure 13. Five Areas in the Office of Strategic Improvement.

Such a structure has the potential to integrate quality, analytics, and big data guided by the strategic direction of the organization. The structure could also become a source of competitive advantage.

7. Develop Well-Rounded Quality and Analytics Professionals

Quality and analytics professionals will most likely need to be well-rounded in their knowledge domains if an organization is to succeed at integrating quality, analytics, and big data. Much has been written about so-called *quants* (see, e.g., Davenport & Kim, 2013; Granville, 2014; Gutierrez, 2014; Patterson, 2010; Shan et al., 2015), but not explicitly from a *quality superiority* perspective. Some of the potential knowledge domains are depicted in Figure 14 (Liedtke, 2014). A quality or analytics professional doesn't necessarily need to become an expert in all the domains. The specific roles and responsibilities will vary by company. Those individuals who are experts in two or three knowledge domains and have at least some basic knowledge in each of the other domains would potentially be valuable Analytics Team members.

Organizations who possess such well-rounded quality and analytics professionals and deploy them wisely could gain a competitive advantage over competitors. Analytics Team members (skilled in quality, analytics, and big data) who are focused on customers might identify new sources of customer value through the use of integrative analytical techniques. Each quality and analytics professional could have their own personal development plan. What might be especially important is for the Analytics Team to have all the knowledge domains *covered* by at least one member of the team.

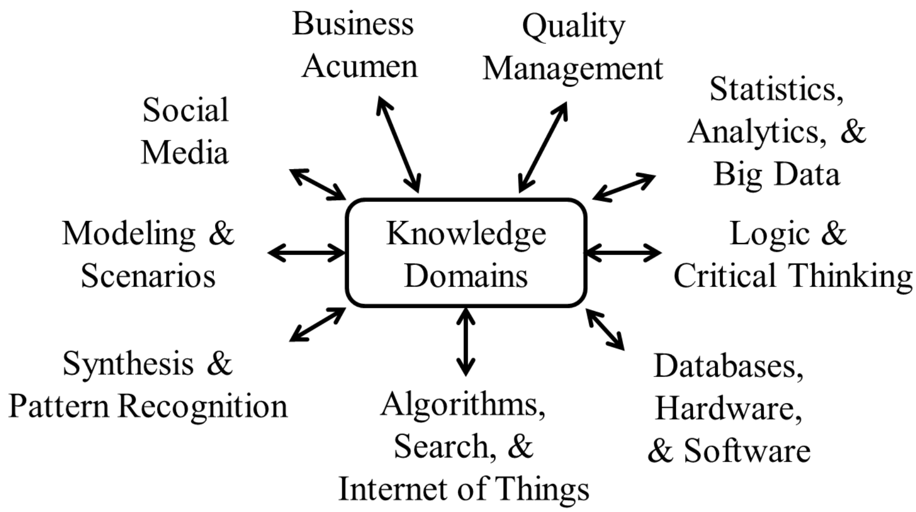

Figure 14. Potential Knowledge Domains.

8. Learn From Variation

Analytics and big data activities generate a large amount of data of various types from a variety of sources sometimes arriving in real-time. Decision quality is at risk if organization members make poor decisions due to inaccurate data, incomplete data, misinterpretation of the data, false assumptions, and/or false conclusions. A lot of bad decisions can be made in a short amount of time given the *new pace of data*. It seems that this is a key question: "What is the data telling us and what should we do?" This is especially important if the veracity of the data is questionable and events are unfolding quickly. Suppose that we observe the following for our company:

- We see a "significant drop" in visits to our website
- Our competitor's new product is *trending*
- Our sentiment analysis reveals "quite a bit more" negativity towards our product
- We think there has been a "shift up" in the number of critical blogs about our product

How should our company respond?

There are many guides available to senior executives for how organizations should interpret and respond to data/information. Shewhart's statistical control chart (1931, 1939) has helped people learn from variation since the 1920s and is still widely used today (see, e.g., Ando & Kumar, 2011). Various statistical methods—such as regression and multivariate analysis—can be used to identify the factors that affect the variation seen in performance metrics. Ansoff (1984) described how organizations can develop capabilities to appropriately respond to weak and strong signals. Kahneman (2011) discussed how humans think and provided suggestions related to *fast thinking* and *slow thinking*. Mudd (2015) described a process for solving complex problems quickly through High Efficiency Analytic Decision-Making (HEAD). Some decision making situations require much time and effort to "get it right." Allison and Zelikow (1999) described various analytical techniques that were used to explain the Cuban Missile Crisis. For each decision situation, the challenge is to move forward as fast as possible with the appropriate level of caution.

9. Identify and Eliminate Waste

Toyota has led the way in popularizing *lean manufacturing* which aims to identify and eliminate waste (see, e.g., Liker, 2004). Ohno (1988, p. 19-20) defined "present capacity" as "work + waste" and identified seven types of waste in a manufacturing company based on his work at Toyota: waste of overproduction, waste of time on hand (waiting), waste in transportation, waste of processing itself, waste of stock on hand (inventory), waste of movement, and waste of making defective products.

Analytics and big data techniques can be applied to help an organization identify and eliminate waste. For example, the strategic placement of sensors, algorithms to predict equipment failures, and the development of models to optimize processes and networks. Alternatively, lean concepts, tools, and techniques can be applied to reduce waste associated with analytics and big data initiatives. Here are potential examples of waste associated with data and information:

- Unnecessary data/information is collected
- Unnecessary data/information is stored
- Unnecessary reports are generated
- Unnecessary notifications are sent
- Reports sent to people who don't need them
- Data entry errors
- Coding errors
- Inefficient code
- Algorithm errors
- Inefficient algorithms
- Bad data from faulty sensors
- Inefficient routing of information
- Information stored in multiple systems
- Data/information that is hard to find
- Data/information that takes time (multiple steps) to access
- Lost data due to inadequate backup procedures
- Slow information system performance due to information clutter

Lean projects could be launched to identify and eliminate *data-related* waste. Tools such as the value stream map and spaghetti diagram could be used to understand the current situation with respect to data/information flows. An organization could apply inventory control practices on bits and bytes and develop Just-In-Time information capabilities where it made sense. An analysis could be conducted to find appropriate data uses of kanbans and andons. The 5S framework could be used to manage and improve data storage items like folders and shared drives. A "Sixth S" could be added to the 5S framework—for Security (of data/information). Organization-wide standards could be developed to improve the effectiveness and efficiency associated with data capture, storage, transmission, and retrieval. Standards for file naming, folder maintenance, picture naming, and video storage could be created where it made sense. Achieving information quality superiority could potentially create a new source of competitive advantage.

10. Manage Customer Data with Great Care

With data and information comes responsibility. You will have plenty of result options to choose from if you type "data breach" into your search engine. Ishikawa (1985, p. 104) identified **"Respect for humanity as a management philosophy"** as an important quality-oriented principle. **"Valuing People"** is one of the Baldrige Criteria core values and concepts. **"Respect for people"** is a prevalent theme in descriptions about Toyota (see, e.g., Liker, 2004). The careful care and management of customer data and information would demonstrate respect for customers.

It is customers who judge product and service quality. How an organization collects and uses a customer's information could affect that customer's satisfaction, loyalty, and willingness to recommend the organization. One of the goals of Komatsu in 2014 was *"Continuous Enhancement of Corporate Value"* – where *corporate value* was defined as *the total sum of trust given to Komatsu by all stakeholders* (Sakane, 2014). Arguably, an organization that managed customer data with great care could increase customer trust and thus contribute to the enhancement of *corporate value* as defined by Komatsu. Here are some <u>sample</u> guiding principles that would demonstrate a **"Manage Customer Data with Great Care"** intent:

- We will respect a customer's data privacy preferences
- We will safeguard and protect customer data
- We will only share customer data when it is authorized and appropriate
- We will work aggressively to prevent errors associated with customer data
- We will correct any customer data error that we discover as quickly as possible

Customers might be more likely to establish a long-term relationship with a company if those principles are evident in the operations of the company. Customers whose data is managed with great care might be more likely to be open and honest with an organization—this could lead to the identification of new sources of customer value and a potential source of competitive advantage.

V. Conclusion

The pursuit of superior quality is not a new human endeavor and neither is *the collection and analysis of data*. These activities have occurred for thousands of years. What is new are the multitude of information technology advances over the past twenty years combined with the convergence of phenomena like the Internet, smartphones, tablets, Wi-Fi, mobility, apps, social media, the cloud, machine learning, cognitive computing, etc. We can assume that technology will continue to advance. Organizations will have access to even more data of different types from new sources arriving in real-time. There is the potential for a greater *digital divide* between individuals and organizations. Some organizations will have to rapidly develop their analytical capabilities in order to survive because of analytics-based competition.

It has been argued that an organization can benefit from integrating quality, analytics, and big data. Analytics and big data can be used to improve product and service quality. Additionally, the application of quality principles can potentially improve analytics and big data initiatives. The <u>emergent theory</u> from the research is that *the integration of quality, analytics, and big data guided* *by the strategic direction of the organization can result in new sources of customer value and a new source of competitive advantage* (or *competence* in the case of non-profit organizations). This could also contribute to the achievement of quality superiority and improved financials.

It was shown that analytics is not just about numbers and big data doesn't necessarily mean we need millions of data points. Zero data, small data, and large data might be as important to organizations as big data. A new definition of big data was suggested: ***Big data*** *is a relatively large amount of data consisting of multiple types from multiple sources possibly arriving in real-time of varying degrees of accuracy requiring exploratory data analysis and integrative analytical methods.* While this is not a definition you would use in normal conversation, it is more comprehensive than the 3V and 4V big data frameworks that appear in the literature.

Ten ideas for better practices were offered beginning with **have superior quality as a strategic intent** and ending with **manage customer data with great care**. Further research is needed, but these ideas provide practitioners with some direction if they want to start integrating quality, analytics, and big data in the attempt to achieve quality superiority.

Perhaps the greatest benefits of integrating quality, analytics, and big data relate to the acquisition of more and better insights into customers; more and better data on process and system performance (causes); and increased decision quality. Non-profit organizations have a better chance to contribute to society through better service to customers and more efficient processes. For-profit companies have the opportunity to identify new sources of customer value and gain a quality <u>and</u> analytics-based competitive advantage.

It is up to the leaders of organizations to make sure George Orwell's dreadful depiction of the future does not become reality. Great care should be taken with customer data and information. We should also keep in mind that "going digital" is not a panacea because "*we can know more than we can tell"* (Polanyi, 1966).

An organization can get started by declaring superior quality as a strategic intent and creating an analytics vision, structure, and roadmap. Best wishes on your journey.

Acknowledgements

I would like to thank the sixteen people who responded to the survey. Their responses were insightful and played an important role in shaping the findings of this report. I am grateful to Dr. Hiroshi Osada for encouraging me to explore *the trend of data analysis in the U.S.* and Dr. Noriaki Kano for sharing his thoughts on *Quality in the 21st Century*. I am also grateful to my Global Quality Futures Workshop friends who gave me feedback on my *Analytics* presentations in Tokyo and Winnipeg. Lastly, I would like to thank Laura Liedtke for her invaluable editorial comments.

References

Akao, Y. (1991). TQM and *Hoshin Kanri*. In Y. Akao (Ed.), Hoshin Kanri: Policy Deployment for Successful TQM (pp. 1-14). New York, NY: Productivity Press. (Originally published in Japan, 1988)

Albright, S. C., & Winston, W. L. (2015). Business Analytics: Data Analysis and Decision Making (5th Ed.). Stamford, CT: Cengage Learning.

Allison, G., & Zelikow, P. (1999). Essence of Decision: Explaining the Cuban Missile Crisis (2nd Ed.). New York, NY: Longman.

Ando, Y., & Kumar, P. (2011). Daily Management the TQM Way: The Key to Success in Tata Steel. Madras, India: Productivity & Quality Publishing Pvt. Ltd.

Ansoff, H. I. (1984). Implanting Strategic Management. New York, NY: Prentice Hall.

Asada, H. (2014). Telematics and Big Data Analysis in Komatsu. Presentation at the International Conference on Quality 2014 Pre-Conference Seminar, Tokyo, Japan, October 19, 2014.

Baldrige Performance Excellence Program. (2015). 2015-2016 Baldrige Excellence Framework: A Systems Approach to Improving Your Organization's Performance. Gaithersburg, MD: U.S. Department of Commerce, National Institute of Standards and Technology. http/www.nist.gov/baldrige.

Box, G. E. P., & Draper, N. R. (1969). Evolutionary Operation: A Statistical Method for Process Improvement. New York, NY: John Wiley & Sons.

Box, G. E. P., Hunter, W. G., & Hunter, J. S. (1978). Statistics for Experimenters: An Introduction to Design, Data Analysis, and Model Building. New York, NY: John Wiley & Sons.

Box, G. E. P., & Draper, N. R. (1987). Empirical Model-Building and Response Surfaces. New York, NY: John Wiley & Sons.

Broehl, W. G. (1992). Cargill: Trading the World's Grain. Hanover, NH: University Press of New England.

Broehl, W. G. (2008). Cargill: From Commodities to Customers. Hanover, NH: University Press of New England.

Brown, M. G. (2013). Killer Analytics: Top 20 Metrics Missing from Your Balance Sheet. Hoboken, NJ: John Wiley & Sons.

Chandler, A. D., & Cortada, J. W. (Eds.) (2000). A Nation Transformed by Information: How Information Has Shaped the United States from the Colonial Times to the Present. New York, NY: Oxford University Press.

Cleveland, W. S. (1985). The Elements of Graphing Data. Monterey, CA: Wadsworth Advanced Books and Software.

Clover, C. (2016). China: When big data meets big brother. Financial Times (U.S. Edition), January 20, 2016, p. 7.

Crooks, E. (2016). General Electric: Post-industrial revolution. Financial Times (U.S. Edition), January 13, 2016, p. 7.

Davenport, T. H., & Harris, J. G. (2007). Competing on Analytics: The New Science of Winning. Boston, MA: Harvard Business School Press.

Davenport, T. H. (Ed.) (2013). Enterprise Analytics: Optimize Performance, Process, and Decisions through Big Data. Upper Saddle River, NJ: FT Press.

Davenport, T. H., & Kim, J. (2013). Keeping Up with the Quants: Your Guide to Understanding and Using Analytics. Boston, MA: Harvard Business Review Press.

Davenport, T. H. (2014). Big Data @ Work: Dispelling the Myths, Uncovering the Opportunities. Boston, MA: Harvard Business Review Press.

Deming, W. E. (1986). Out of the Crisis. Cambridge, MA: Massachusetts Institute of Technology Center for Advanced Engineering Study.

Deming, W. E. (1994). The New Economics: For Industry, Government, Education (2nd Ed.). Cambridge, MA: Massachusetts Institute of Technology Center for Advanced Engineering Study.

Deming Prize Committee – Union of Japanese Scientists and Engineers. (2015). The Application Guide for The Deming Prize, The Deming Grand Prize, 2015. Tokyo, Japan: Union of Japanese Scientists and Engineers.

Denzin, N. K., & Lincoln, Y. S. (Eds.) (2011). Handbook of Qualitative Research (4th Ed.). Thousand Oaks, CA: SAGE Publications.

Diebold, F. X. (2012). A Personal Perspective on the Origin(s) and Development of "Big Data": The Phenomenon, the Term, and the Discipline. Draft paper dated November 26, 2012.

Diggle, P. J., & Chetwynd, A. G. (2011). Statistics and Scientific Method: An Introduction for Students and Researchers. New York, NY: Oxford University Press.

Evans, J. R. (2015). Modern Analytics and the Future of Quality and Performance Excellence. Quality Management Journal, 22(4), 6-17.

Evans, P., & Wurster, T. S. (2000). Blown to Bits: How the New Economics of Information Transforms Strategy. Boston, MA: Harvard Business School Press.

Fisher, R. A. (1973). Statistical Methods for Research Workers (14th Ed. Reprinted) (First published in 1925). New York, NY: Hafner Publishing Company.

Frazier, K. (2005). People of Chaco: A Canyon and Its Culture (3rd Ed.). New York, NY: W. W. Norton & Company.

Garrahan, R. (2015). Mastermind behind Bergdorf's ambitious plan. Financial Times, January 18, 2015.

Gleick, J. (2011). The Information: A History, A Theory, A Flood. New York, NY: Pantheon Books.

Granville, V. (2014). Developing Analytic Talent: Becoming a Data Scientist. Indianapolis, IN: John Wiley & Sons.

Greengard, S. (2015). The Internet of Things. Cambridge, MA: The MIT Press.

Grus, J. (2015). Data Science from Scratch: First Principles with Python. Sebastropol, CA: O'Reilly Media.

Gutierrez, S. (2014). Data Scientists at Work. New York, NY: Apress Media.

Hasegawa, Y. (2010). Rediscovering Japanese Business Leadership: 15 Japanese Managers and the Companies They're Leading to New Growth. Singapore: John Wiley & Sons (Asia) Pte. Ltd.

Higginbotham, S. (2015). Formula 1(01001011). Fortune, December 1, 2015, pp. 76-78.

Hoffmann, C., Lesser, E., & Ringo, T. (2012). Calculating Success: How the New Workplace Analytics Will Revitalize Your Organization. Boston, MA: Harvard Business Review Press.

Hosotani, K. (1992). The QC Problem-Solving Approach: Solving Workplace Problems the Japanese Way. Tokyo, Japan: 3A Corporation. (Originally published in Japan, 1989.)

Ishikawa, H. (2015). Social Big Data Mining. Boca Raton, FL: CRC Press – Taylor & Francis Group.

Ishikawa, K. (1982). Guide to Quality Control. Tokyo, Japan: Asian Productivity Organization. (Originally published in Japanese, 1968.)

Ishikawa, K. (1985). What Is Total Quality Control? The Japanese Way. Englewood Cliffs, NJ: Prentice-Hall. (Originally published in Japan, 1981.)

Isaacson, W. (2014). The Innovators: How a Group of Hackers, Geniuses, and Geeks Created the Digital Revolution. New York, NY: Simon & Schuster.

James, B. (1977). 1977 Baseball Abstract. Lawrence, KS: Bill James.

Japanese Society for Quality Control. (2014). STANDARD: Guidelines for Daily Management. Japan: Japanese Society for Quality Control.

Johnson, R. A., & Wichern, D. W. (2014). Applied Multivariate Statistical Analysis (6th Ed.). Upper Saddle River, NJ: Pearson Prentice Hall.

Jones, T. W. (2015). The Identity of Statistics in Data Science. Amstat News, November 2015, Issue #461, pp. 25-26.

Juran, J. M. (Ed.) (1951). Quality Control Handbook. New York, NY: McGraw-Hill.

Juran, J. M. (2004). Architect of Quality: The Autobiography of Dr. Joseph M. Juran. New York, NY: McGraw-Hill.

Kahneman, D. (2011). Thinking, Fast and Slow. New York, NY: Farrar, Straus and Giroux.

Kahney, L. (2009). Inside Steve's Brain (Expanded Edition). New York: NY: Portfolio – Penguin Group.

Kano, N. (2014). TQM Promotion and Implementation in the 21st Century. Presentation at the International Conference on Quality 2014 Pre-Conference Seminar, Tokyo, Japan, October 19, 2014.

Kaplan, R. S., & Norton, D. P. (1992). The Balanced Scorecard: Measures That Drive Performance. Harvard Business Review, 70(1), 71-79.

Kelly, J. E., & Hamm, S. (2013). Smart Machines: IBM's Watson and the Era of Cognitive Computing. New York, NY: Columbia University Press.

Kirkpatrick, D. (2010). The Facebook Effect: The Inside Story of the Company That is Connecting the World. New York, NY: Simon & Schuster Paperbacks.

Kuhn, T. S. (1970). The Structure of Scientific Revolutions (2nd Ed.). Chicago, IL: The University of Chicago Press.

Kume, H. (Ed.) (1985). Statistical Methods for Quality Improvement. Tokyo, Japan: The Association for Overseas Technical Scholarship. (Originally published in Japan, 1985.)

Laney, D. (2001). 3D Data Management: Controlling Data Volume, Velocity, and Variety. Stamford, CT: META Group Research Note, February 6, 2001.

Lewis, M. (2003). Moneyball: The Art of Winning an Unfair Game. New York, NY: W. W. Norton & Company.

Liedtke, C. A., Schroeder, R. G., Linderman, K., & Rickard, J. (2010). Top This: How to Maintain Your Edge After Gaining a Quality Advantage. Quality Progress, 43(7), pp. 24-29.

Liedtke, C. A. (2014). Trend of Data Analysis in the U.S.A. Presentation at the International Conference on Quality 2014 Pre-Conference Seminar, Tokyo, Japan, October 19, 2014.

Liedtke, C. A. (2015). Quality, Analytics, and the Big Data Revolution. Presentation at the Global Quality Futures Workshop, Winnipeg, Canada, August 26, 2015.

Liker, J. K. (2004). The Toyota Way: 14 Management Principles from the World's Greatest Manufacturer. New York, NY: McGraw-Hill.

Luca, M., Kleinberg, J., & Mullainathan, S. (2016). Algorithms Need Managers, Too. Harvard Business Review, 94(1/2), 96-101.

MacCormick, J. (2012). 9 Algorithms That Changed the Future: The Ingenious Ideas That Drive Today's Computers. Princeton, NJ: Princeton University Press.

Marz, N., & Warren, J. (2015). Big Data: Principles and Best Practices of Scalable Real-Time Data Systems. Shelter Island, NY: Manning Publications.

Mayer-Schonberger, V., & Cukier, K. (2013). Big Data: A Revolution That Will Transform How We Live, Work, and Think. Boston, MA: Houghton Mifflin Harcourt.

McNamara, R. S. (1996). In Retrospect: The Tragedy and Lessons of Vietnam. New York, NY: Vintage Books.

McNeil, D. (Ed.) (2014). Analytics in Healthcare and the Life Sciences: Strategies, Implementation Methods, and Best Practices. Upper Saddle River, NJ: Pearson.

Mintzberg, H. (1994). The Rise and Fall of Strategic Planning: Reconceiving Roles for Planning, Plans, Planners. New York, NY: The Free Press.

Mudd, P. (2015). The HEAD Game: High Efficiency Analytic Decision-Making and the Art of Solving Complex Problems Quickly. New York, NY: Liveright Publishing Corporation.

Muenzmaier. M. (2015). Cargill and Sustainability – Enterprise and Business Approaches. Presentation at the Advanced Strategic Improvement Practices Conference, Chanhassen, MN, October 7, 2015.

Nonaka, I., & Takeuchi, H. (1995). The Knowledge-Creating Company: How Japanese Companies Create the Dynamics of Innovation. New York, NY: Oxford University Press.

Ohno, T. (1988). Toyota Production System: Beyond Large-Scale Production. Portland, OR: Productivity Press.

Orwell, G. (1949). Nineteen Eighty-Four. London, United Kingdom: Secker and Warburg.

Pasquale, F. (2015). The Black Box Society: The Secret Algorithms That Control Money and Information. Cambridge, MA: Harvard University Press.

Patterson, S. (2010). The Quants: How a New Breed of Math Whizzes Conquered Wall Street and Nearly Destroyed It. New York, NY: Crown Business.

Polanyi, M. (1966). The Tacit Dimension. Chicago, IL: The University of Chicago Press.

Porter, M. E. (1985). Competitive Advantage: Creating and Sustaining Superior Performance. New York, NY: The Free Press.

Porter, M. E. (2001). Strategy and the Internet. Harvard Business Review, 79(3), pp. 62-78.

Press, G. (2014). 12 Big Data Definitions: What's Yours? Forbes / Tech. www.forbes.com.

Priestley, J. L. (2016). Data Science: The Evolution or the Extinction of Statistics? Amstat News, January 2016, Issue #463, pp. 26-27.

Provost, F., & Fawcett, T. (2013). Data Science for Business: What You Need to Know About Data Mining and Data-Analytic Thinking. Sebastropol, CA: O'Reilly Media.

Rainer, R. K., & Cegielski, C. G. Introduction to Information Systems: Supporting and Transforming Business (3rd. Ed.). Hoboken, NJ: John Wiley & Sons.

Rascoff, S., & Humphries, S. (2015). Zillow Talk: The New Rules of Real Estate. New York, NY: Grand Central Publishing.

Rosenzweig, P. (2010). Robert S. McNamara and the Evolution of Modern Management. Harvard Business Review, 88(12), 86-93.

Sakane, M. (2014). Innovation through Quality – Creating New Value for the World! (Komatsu). Presentation at the International Conference on Quality 2014, Tokyo, Japan, October 19, 2014.

Salsburg, D. (2001). The Lady Tasting Tea: How Statistics Revolutionized Science in the Twentieth Century. New York, NY: W. H. Freeman and Company.

Sartre, J. P. (1947). Existentialism. New York, NY: The Philosophical Library.

Sawchik, T. (2015). Big Data Baseball: Math, Miracles, and the End of a 20-Year Losing Streak. New York, NY: Flatiron Books.

Schein, E. H. (2004). Organizational Culture and Leadership (3rd Ed). San Francisco, CA: Jossey-Bass.

Schmidt, E., & Cohen, J. (2014). The New Digital Age: Transforming Nations, Businesses, and Our Lives. New York, NY: Vintage Books.

Schmidt, E., & Rosenberg, J. (2014). How Google Works. New York, NY: Grand Central Publishing.

Schroeder, R. G., Linderman, K., Liedtke, C., & Choo, A. S. (2008). Six Sigma: Definition and Underlying Theory, Journal of Operations Management, 26 (2008), 536-554.

Schwarz, A. (2004). The Numbers Game: Baseball's Lifelong Fascination With Statistics. New York, NY: Thomas Dunne Books – St. Martin's Press.

Shan, C., Wang, H., Chen, W., & Song, M. The Data Science Handbook: Advice and Insights from 25 Amazing Data Scientists. The Data Science Bookshelf.

Shapiro, C., & Varian, H. R. (1999). Information Rules: A Strategic Guide to the Network Economy. Boston, MA: Harvard Business Review Press.

Sharma, A. (2014). Data Analysis for Innovation: A Case Study by Mahindra & Mahindra Ltd. India. Presentation at the International Conference on Quality 2014 Pre-Conference Seminar, Tokyo, Japan, October 19, 2014.

Sheehan, N., Smith, H., Kenworthy, E. W., Butterfield, F. (1971). The Pentagon Papers. New York, NY: Bantam Books.

Shewhart, W. A. (1931). Economic Control of Quality of Manufactured Product. New York, NY: D. Van Nostrand Company.

Shewhart, W. A. (1939). Statistical Method from the Viewpoint of Quality Control. Washington, D.C.: Graduate School of the Department of Agriculture.

Siegel, E. (2013). Predictive Analytics: The Power to Predict Who Will Click, Buy, Lie, or Die. Hoboken, NJ: John Wiley & Sons.

Silver, N. (2012). The Signal and the Noise: Why So Many Predictions Fail – but Some Don't. New York, NY: The Penguin Press.

Snedecor, G. W., & Cochran, W. G. (1989). Statistical Methods (8th Ed.) (First published in 1937 by Snedecor, G. W.). Ames, IA: Iowa State University Press.

Speidel, T. (2014). Time to Embrace a New Identify? Amstat News, October 2014, Issue #448, pp. 19-21.

Stone, B. (2013). The Everything Store: Jeff Bezos and the Age of Amazon. New York, NY: Back Bay Books.

Taylor, F. W. (1947). The Principles of Scientific Management (First published in 1911). New York, NY: W. W. Norton & Company.

Thorn, J., & Palmer, P. (1985). The Hidden Game of Baseball: A Revolutionary Approach to Baseball and Its Statistics (2nd Ed.). Chicago, IL: The University of Chicago Press.

Tufte, E. R. (2001). The Visual Display of Quantitative Information (2nd Ed.). Cheshire, CT: Graphics Press.

Tukey, J. W. (1977). Exploratory Data Analysis. Reading, MA: Addison-Wesley Publishing.

Van Dyke, D., Fuentes, M., Jordan, M. L., Newton, M., Ray, B. K., Lang, D. T., & Wickham, H. (2015). ASA Statement on the Role of Statistics in Data Science. Amstat News, October 2015, Issue #460, p. 9.

Varian, H. (2009). Hal Varian on how the web challenges managers. McKinsey Quarterly, January of 2009.

Varian, H. R., Farrell, J., & Shapiro, C. (2004). The Economics of Information Technology: An Introduction. Cambridge, United Kingdom: Cambridge University Press.

Wachter, R. (2015). The Digital Doctor: Hope, Hype, and Harm at the Dawn of Medicine's Computer Age. New York, NY: McGraw-Hill Education.

Waters, R. (2016). Artificial intelligence: Can Watson save IBM? Financial Times (U.S. Edition), January 6, 2016, p. 5.

White, T. (2015). Hadoop: The Definitive Guide: Storage and Analysis at Internet Scale (4th Ed.). Sebastropol, CA: O'Reilly Media.

Wilson, F. A. (1995). Understanding Digital Technology. London, United Kingdom: Bernard Babani LTD.

Yau, N. (2013). Data Points: Visualization That Means Something. Indianapolis, IN: John Wiley & Sons.

Yin, R. K. (2014). Case Study Research: Design and Methods (5th Ed.). Los Angeles, CA: SAGE Publications.

Zeilik, M. (1984). Archaeoastronomy at Chaco Canyon: The Historic-Prehistoric Connection. In D. G. Noble (Ed.), New Light on Chaco Canyon (pp. 65-72). Santa Fe, NM: School of American Research Press.

Zikopoulos, P., deRoos, D., Bienko, C., Buglio, R., & Andrews, M. (2015). Big Data Beyond the Hype: A Guide to Conversations for Today's Data Center. New York, NY: McGraw-Hill Education.

On-Line Publications

FiveThirtyEight: www.fivethirtyeight.com
Information Management: www.information-management.com
Information Week: www.informationweek.com

Analytics/Big Data Companies

Amazon	www.amazon.com	Apple	www.apple.com
Cognizant	www.cognizant.com	EMC	www.emc.com
Facebook	www.facebook.com	Gartner	www.gartner.com
GE	www.ge.com	Google	www.google.com
Komatsu	www.komatsu.com	HP	www.hp.com
IBM	www.ibm.com	Microsoft	www.microsoft.com
Oracle	www.oracle.com	SAP	www.sap.com
SAS	www.sas.com	Teradata	www.teradata.com
Thinknum	www.thinknum.com	Yahoo	www.yahoo.com
Zillow	www.zillow.com		

University Programs

New York University – Master of Science in Data Science
Northwestern University – Master of Science in Analytics
Stanford University – Master of Science in Statistics: Data Science
University of Chicago – Master of Science in Analytics
University of San Francisco – Master of Science in Analytics
Wharton – MBA Specialization in Operations, Information and Decisions

Information on the Author

SIS
Strategic Improvement Systems

Charles A. Liedtke, Ph.D., is the owner of Strategic Improvement Systems (SIS) which is located in the Minneapolis, Minnesota area. Charles advises leaders of organizations on how to develop and achieve strategic objectives. He consults and delivers customized training in the areas of strategy, quality, analytics, improvement, and innovation. Charles earned a Ph.D. in business and a Ph.D. minor in statistics from the University of Wisconsin-Madison; an M.B.A. from UW-Madison; an M.S. degree in statistics from Iowa State University, and a B.S. degree in economics from South Dakota State University.

Contact Information for Charles:

Electronic mail address:	charles@sisliedtke.com
Office phone:	952-380-0778
Website:	www.strategicimprovementsystems.com