What is Data Analytics?

With the advent of the internet (and the internet of things), data is being produced at an incredible rate. Our challenge is to take the raw data we are collecting, turn it into useful information, and to use that information to make decisions.

Analytics has been defined in many ways. One way it has been defined is '*the scientific process of transforming data into insight for making better decisions*' [informs]

Data analytics is a multi-disciplinary field. It most often utilizes data visualization and statistics, but can also incorporate operations research, machine learning, and computer programming. The insights uncovered are used to drive decision-making, raise productivity and gain competitive advantage.

A few examples of successful analytics:

- **McKesson's Supply Chain Scenario Modeler Cures Pharmaceutical Distribution Network**
  McKesson is America's oldest and largest healthcare services company. IBM Research developed an innovative scenario modeling and analysis tool, supply chain scenario modeler (SCSM), for McKesson to optimize its end-to-end pharmaceutical supply chain policies. Through integrated operations research (OR) models, SCSM optimizes the distribution network, supply flow, inventory, and transportation policies, and quantifies the impacts of changes on financial, operational, and environmental metrics. The modeling work spawned a roadmap of projects with quantified opportunities, including a new air freight supply chain path, and provided new insights that have been critical to improving McKesson's performance as a pharmaceutical industry leader. A structured data model supporting the OR models has provided a basis for additional improvement projects. The model directly links OR modeling results to a detailed profit-and-loss statement by product category for the different supply chain paths that McKesson uses. Since this effort began in 2009, McKesson Pharmaceutical division has reduced its committed capital by more than $1 billion.

- **Optimizing Chevron's Refineries**
  Chevron has developed a software modeling tool that its seven company-owned refineries use to select the most profitable raw materials, evaluate product options, optimize refinery processes, and promote efficient capital investments. The tool is a linear program with distributive recursion mathematics, which Chevron uses in operations and strategic planning. Over the past 30-plus years, the company has continually improved this application of operations research, and its

complementary and supporting systems and business processes, and they are now deeply embedded into the fabric of Chevron's downstream business of reliably and efficiently supplying products to our customers. The value that these efforts bring to Chevron now approaches $1 billion annually. We estimate that the cumulative value to Chevron over the past three decades is approximately $10 billion.

- **Kroger Uses Simulation-Optimization to Improve Pharmacy Inventory Management**
  The Kroger Co. is the largest grocery retailer in the United States. It operates 2,422 supermarkets and 1,950 in-store pharmacies. Improving customer service is at the heart of Kroger's business strategy. Toward this end, Kroger's operations research team, in collaboration with faculty from Wright State University, developed an innovative simulation-optimization system for pharmacy inventory management. In pharmacy applications, traditional standard statistical distributions fall short of providing accurate pharmacy demand distributions. To overcome business resistance to complex formulas, this simulation-optimization approach uses empirical distributions to model demand, provides end users with a visual intuitive experience, and delivers optimal or near-optimal results in milliseconds through local search heuristics. The system was implemented in October 2011 in all Kroger pharmacies in the United States, and has reduced out-of-stocks by 1.6 million per year, ensuring greater patient access to medications. It has resulted in an increase in revenue of $80 million per year, a reduction in inventory of more than $120 million, and a reduction in labor cost equivalent to $10 million per year.

More success story examples:
https://www.informs.org/Impact/O.R.-Analytics-Success-Stories

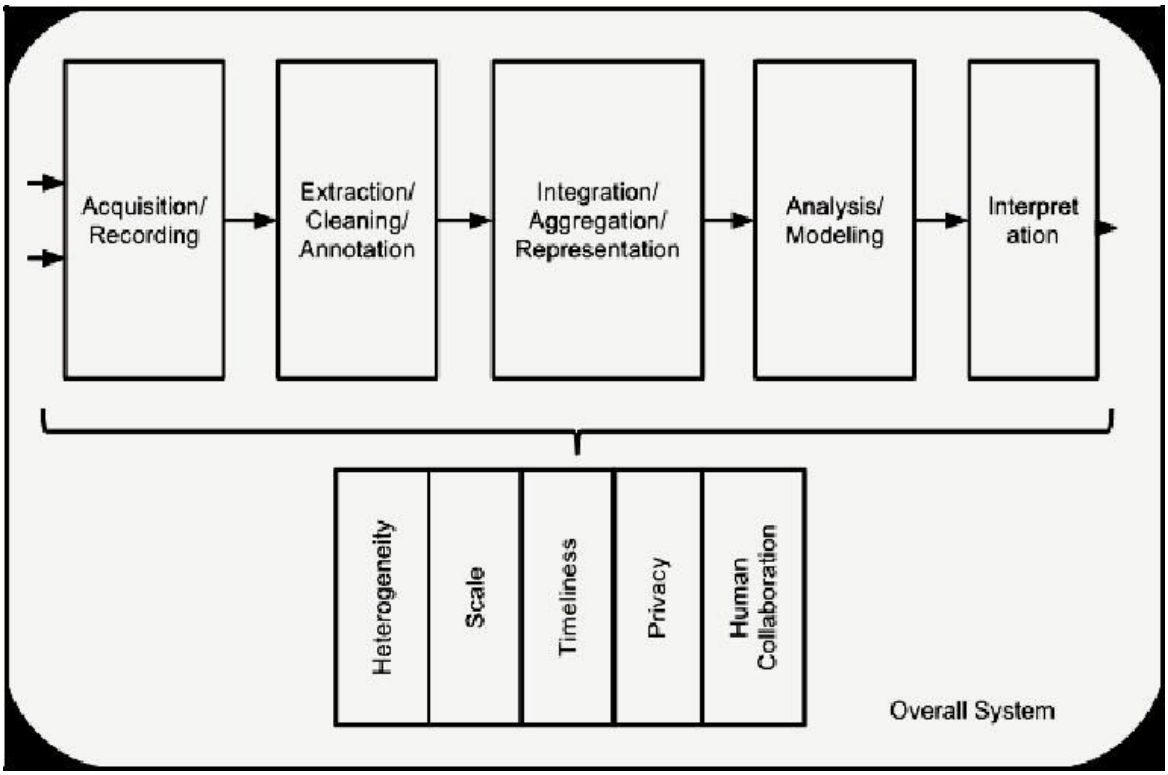
Challenges

Previously, massive sets of data have been limited to scientific communities. The intensive growth of the Internet and social media generate enormous volumes of data. This tremendous size of data is known as 'Big Data'. Enterprises become interested in leveraging the data they have obtained. The aims, generally, are to identify potential customers, predict trends, product/service recommendations, and fraud detection, including the integration of various data analytics to solve the overall organization problems. Big corporations such as Google, Microsoft, Amazon, and Netflix can gain insights and make greater revenues from the data they collect from users.

However, the central focus of this challenge is finding the methodology that can handle not only the size of data but can also process the data with speed.

Another challenge is the structure of data. Enterprises produce data in diverse forms, both structured and unstructured formats. Working with unstructured data such as emails, documents (text, PDF), data streams, and network graphs, geospatial locations can be complicated. However, the data is valuable with so much information in it and to make the most out of it, various data sources should be integrated. Not all businesses, however, have the capability to tackle the unstructured data. Inevitably, industries are forced to come up with cutting-edge technology, raising analytics to novel levels. Other challenges still remain such as cloud computing, security, etc.

The following figure depicts the major steps in big data analysis:



(Big) Data analysis Pipeline
Source: http://www.cra.org/ccc/files/docs/init/bigdatawhitepaper.pdf

Risks in analytics include privacy (e.g. social media networks, such as Facebook, accommodate a lot of information, which is revealed to many others), security, making decisions on incomplete or inaccurate data, using only the data that supports intuitive decisions, drawing the wrong conclusions from the data, etc.

Analytics Tools are comprised of:
- Data mining
- Statistical Analysis
- Predictive Analysis

- Correlation
- Regression
- Forecasting
- Process Modeling
- Operations Research
- Optimizations
- Simulations
- Machine learning
- Visualizations

Techniques in data analysis such as statistics, machine learning, visualizations, and operation research will be further discussed in later chapters.

Related Topics:

Business Analytics (BA)
BA applies the statistical and quantitative tools to gain insights of business performance. Many people believe that analytics date back to the early 19th century when Henry Ford attempted to measure the time of each component in the assembly line. Nevertheless, the explosion of business analytics was in the late 1960s with the introduction of computers for decision support systems (DSS) and planning. Business intelligence (BI) is believed to progress from the DSS and became the center of attention in the late 1980s. According to Thomas Davenport (a professor of information system and a research director at International Institute for Analytic), business analytics is the subset of BI, while BA emphasizes statistics, prediction, and optimization, BI covers a broader spectrum including querying, reporting, OLAP, alerts tools, and BA.

Business Analytics consist of three phases:

· *Descriptive analytics* is the first stage of BA, which concerns answering the questions regarding what happened and why. Past or historical data such as scorecards and reports (e.g. sales, marketing, finance) are analyzed to understand performance (either success or failure).

· *Predictive analytics* applies various algorithms and techniques to the data for making predictions. This stage tries to answer questions of what or when something will happen.

· *Prescriptive analytics* extends to cover questions such as why something will happen. In this stage, the alternative decisions and implications are suggested. It is a continuous stage, where new data can re-predict and re-prescribe to better the prediction accuracy and improve decision options.

Examples of applications include differentiating banking customers based on

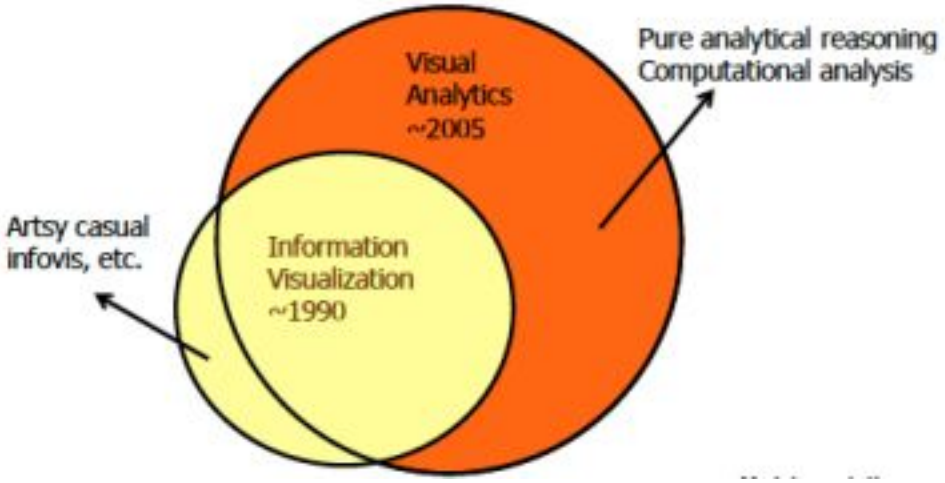
credit risk, offering products to match customer characteristics, customer loyalty programs for gambling businesses, inventory optimizations, etc.

Business analytics have also been applied in various domains including financial, marketing, pricing, retail sales, risk & credit, supply chain, and transportation.

Visual Analytics (VA)
VA is defined as "the science of analytical reasoning facilitated by interactive visual interfaces"[Stasko, J]. It is not the area but the umbrella idea. Visual analytics "combined automated analysis techniques with interactive visualizations for an effective understanding, reasoning, and decision making on the basis of very large and complex data sets" (Keim, et al. 2008)

In fact, information visualization tools do not always include data analysis algorithms. A paper authored by Shneiderman proposed the integration of computational analysis approaches (e.g. data mining) with information visualization (e.g. Infovis). Therefore, the contribution of each approach results in a more powerful approach for effective understanding, user learning, and decision making especially for very large and complex data. Additionally, the numbers of situations where better analysis for large data sets are rising such as law enforcement, security, and business intelligence also help to promote the idea of visual analytics. The main components for visual analytic consist of interactive visualization, analytical reasoning, and computational analysis.

The following figure shows the overlap between visual analytics and information visualization. The information visualization principles will be further discussed in a later chapter.



Source: Visual Analytics by Stasko, Georgia Tech.

Visual analytics have been employed in:

- Scientific Research
- Regulatory and Legal Communities
- Intelligence Analysis
- DOE and DOD
- Market Assessments
- Capability Analysis – Resumes
- Medical and Pharmaceutical Communities
- National Security and Law Enforcement
- Information Assurance, Web Analytics
- Technology Scanning, Asset and Intellectual Property Management

References

Defining the big data architecture framework (2013). University of Amsterdam.
Retrieved from:
http://bigdatawg.nist.gov/_uploadfiles/M0055_v1_7606723276.pdf

D. Keim, G. Andrienko, J.-D. Fekete, C. Gorg, J. Kohlhammer, and G. Melancon (2008), *Visual Analytics: Definition, Process, and Challenges*, in Information Visualization: Human-Centered Issues and Perspectives, (Editors: A. Kerren, J. Stasko, J.D. Fekete, C. North), Springer, pp. 1-18.

Ferire, J. (2013). *Massive data analysis: course overview*.
Retrieved from:
http://vgc.poly.edu/~juliana/courses/cs9223/Lectures/intro.pdf

Leek, J. (2013) *Data Analysis*. John Hopkins. Coursera.
Retrieved from:
https://www.coursera.org/browse/data-science/data-analysis

Naone, E. (2011) "*The New Big Data*". Technology Review, MIT.
Retrieved from:
https://www.technologyreview.com/2011/08/22/192225/the-new-big-data/

Stasko, J. *Visual Analytics*. Georgia Technological University.
Retrieved from:
https://www.cc.gatech.edu/~stasko/papers/computer13-intell.pdf

What is Analytics? (March 2017) INFORMS- Institute of operations research and the management sciences. Retrieved from:
https://www.informs.org/Explore/Operations-Research-Analytics