

Data science and data analysis are both important disciplines in the field of data management and analysis, but they differ in several key ways. In 1985, in a lecture given to the Chinese Academy of Sciences in Beijing, C. Data science is a "concept to unify statistics, data analysis, informatics, and their related methods" to "understand and analyze actual phenomena" with data. A decade later, they reaffirmed it, stating that "the job is more in demand than ever with employers". The modern conception of data science as an independent discipline is sometimes attributed to William S. Data scientists often work with unstructured data such as text or images and use machine learning algorithms to build predictive models and make data-driven decisions. There is still no consensus on the definition of data science, and it is considered by some to be a buzzword. During the 1990s, popular terms for the process of finding patterns in datasets (which were increasingly large) included "knowledge discovery" and "data mining". It uses techniques and theories drawn from many fields within the context of mathematics, statistics, computer science, information science, and domain knowledge. Both fields require a solid foundation in statistics, programming, and data visualization, as well as the ability to communicate findings effectively to both technical and non-technical audiences. Others argue that data science is distinct from statistics because it focuses on problems and techniques unique to digital data. The field encompasses preparing data for analysis, formulating data science problems, analyzing data, developing data-driven solutions, and presenting findings to inform high-level decisions in a broad range of application domains. Later, attendees at a 1992 statistics symposium at the University of Montpellier II acknowledged the emergence of a new discipline focused on data of various origins and forms, combining established concepts and principles of statistics and data analysis with computing. In a 2001 paper, he advocated an expansion of statistics beyond theory into technical areas; because this would significantly change the field, it warranted a new name. They work at the intersection of mathematics, computer science, and domain expertise to solve complex problems and uncover hidden patterns in large datasets. This can involve tasks such as data cleaning, data visualization, and exploratory data analysis to gain insights into the data and develop hypotheses about relationships between variables. "Data science" became more widely used in the next few years: in 2002, the Committee on Data for Science and Technology launched the Data Science Journal. Despite these differences, data science and data analysis are closely related fields and often require similar skill sets. Data analysts typically use statistical methods to test these hypotheses and draw conclusions from the data. However, the definition was still in flux. Data science is multifaceted and can be described as a science, a research paradigm, a research method, a discipline, a workflow, and a profession. Data scientists often work with unstructured data such as text or images and use machine learning algorithms to build predictive models and make data-driven decisions. In 1996, the International Federation of Classification Societies became the first conference to specifically feature data science as a topic. In 1962, John Tukey described a field he called "data analysis", which resembles modern data science.