

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. The term "data science" has been traced back to 1974, when Peter Naur proposed it as an alternative name to computer science. Data science is multifaceted and can be described as a science, a research paradigm, a research method, a discipline, a workflow, and a profession. It uses techniques and theories drawn from many fields within the context of mathematics, statistics, computer science, information science, and domain knowledge. 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. While both fields involve working with data, data science is more of an interdisciplinary field that involves the application of statistical, computational, and machine learning methods to extract insights from data and make predictions, while data analysis is more focused on the examination and interpretation of data to identify patterns and trends. Data science and data analysis are both important disciplines in the field of data management and analysis, but they differ in several key ways. However, the definition was still in flux. In 2015, the American Statistical Association identified database management, statistics and machine learning, and distributed and parallel systems as the three emerging foundational professional communities. A data scientist is a professional who creates programming code and combines it with statistical knowledge to create insights from data. 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. Many statisticians, including Nate Silver, have argued that data science is not a new field, but rather another name for statistics. The term "data science" has been traced back to 1974, when Peter Naur proposed it as an alternative name to computer science. Data analysis focuses on extracting insights and drawing conclusions from structured data, while data science involves a more comprehensive approach that combines statistical analysis, computational methods, and machine learning to extract insights, build predictive models, and drive data-driven decision-making. The term "data science" has been traced back to 1974, when Peter Naur proposed it as an alternative name to computer science. Data scientists are responsible for breaking down big data into usable information and creating software and algorithms that help companies and organizations determine optimal operations. 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. "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. A data scientist is a professional who creates programming code and combines it with statistical knowledge to create insights from data. In 1962, John Tukey described a field he called "data analysis", which resembles modern data science. Statistician Nathan Yau, drawing on Ben Fry, also links data science to human-computer interaction: users should be able to intuitively control and explore data. Jeff Wu again suggested that statistics should be renamed data science. In 2014, the American Statistical Association's Section on Statistical Learning and Data Mining changed its name to the Section on Statistical Learning and Data Science, reflecting the ascendant popularity of data science.