

Others argue that data science is distinct from statistics because it focuses on problems and techniques unique to digital data. A data scientist is a professional who creates programming code and combines it with statistical knowledge to create insights from data. It uses techniques and theories drawn from many fields within the context of mathematics, statistics, computer science, information science, and domain knowledge. Data science, on the other hand, is a more complex and iterative process that involves working with larger, more complex datasets that often require advanced computational and statistical methods to analyze. A data scientist is a professional who creates programming code and combines it with statistical knowledge to create insights from data. However, data science is different from computer science and information science. Data science is an interdisciplinary academic field that uses statistics, scientific computing, scientific methods, processes, algorithms and systems to extract or extrapolate knowledge and insights from noisy, structured, and unstructured data. He reasoned that a new name would help statistics shed inaccurate stereotypes, such as being synonymous with accounting or limited to describing 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. A decade later, they reaffirmed it, stating that "the job is more in demand than ever with employers". 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. "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. Data science is an interdisciplinary academic field that uses statistics, scientific computing, scientific methods, processes, algorithms and systems to extract or extrapolate knowledge and insights from noisy, structured, and unstructured data. Data science also integrates domain knowledge from the underlying application domain (e.g., natural sciences, information technology, and medicine). There is still no consensus on the definition of data science, and it is considered by some to be a buzzword. Both fields play vital roles in leveraging the power of data to understand patterns, make informed decisions, and solve complex problems across various domains. Data analysts typically use statistical methods to test these hypotheses and draw conclusions from the data. 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. In 1996, the International Federation of Classification Societies became the first conference to specifically feature data science as a topic. Andrew Gelman of Columbia University has described statistics as a non-essential part of data science. Data science and data analysis are both important disciplines in the field of data management and analysis, but they differ in several key ways. Though it was used by the National Science Board in their 2005 report "Long-Lived Digital Data Collections: Enabling Research and Education in the 21st Century", it referred broadly to any key role in managing a digital data collection. 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. Both fields play vital roles in leveraging the power of data to understand patterns, make informed decisions, and solve complex problems across various domains. Others argue that data science is distinct from statistics because it focuses on problems and techniques unique to digital data.