

Jeff Wu again suggested that statistics should be renamed 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. Moreover, both fields benefit from critical thinking and domain knowledge, as understanding the context and nuances of the data is essential for accurate analysis and modeling. 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. Data science is an interdisciplinary field focused on extracting knowledge from typically large data sets and applying the knowledge and insights from that data to solve problems in a wide range of application domains. Data science and data analysis are both important disciplines in the field of data management and analysis, but they differ in several key ways. 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. Data science also integrates domain knowledge from the underlying application domain (e.g., natural sciences, information technology, and medicine). Data science also integrates domain knowledge from the underlying application domain (e.g., natural sciences, information technology, and medicine). During the 1990s, popular terms for the process of finding patterns in datasets (which were increasingly large) included "knowledge discovery" and "data mining". Others argue that data science is distinct from statistics because it focuses on problems and techniques unique to digital data. He describes data science as an applied field growing out of traditional statistics. Vasant Dhar writes that statistics emphasizes quantitative data and description. 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 1996, the International Federation of Classification Societies became the first conference to specifically feature data science as a topic. In contrast, data science deals with quantitative and qualitative data (e.g., from images, text, sensors, transactions, customer information, etc.) and emphasizes prediction and action. After the 1985 lecture at the Chinese Academy of Sciences in Beijing, in 1997 C. Despite these differences, data science and data analysis are closely related fields and often require similar skill sets. During the 1990s, popular terms for the process of finding patterns in datasets (which were increasingly large) included "knowledge discovery" and "data mining". 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. Andrew Gelman of Columbia University has described statistics as a non-essential part of data science. As such, it incorporates skills from computer science, statistics, information science, mathematics, data visualization, information visualization, data sonification, data integration, graphic design, complex systems, communication and business. Data science is a "concept to unify statistics, data analysis, informatics, and their related methods" to "understand and analyze actual phenomena" with data. Turing Award winner Jim Gray imagined data science as a "fourth paradigm" of science (empirical, theoretical, computational, and now data-driven) and asserted that "everything about science is changing because of the impact of information technology" and the data deluge. 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.