

Both fields play vital roles in leveraging the power of data to understand patterns, make informed decisions, and solve complex problems across various domains. Jeff Wu again suggested that statistics should be renamed 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. For instance, a data scientist might develop a recommendation system for an e-commerce platform by analyzing user behavior patterns and using machine learning algorithms to predict user preferences. 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. 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. Others argue that data science is distinct from statistics because it focuses on problems and techniques unique to digital data. While data analysis focuses on extracting insights from existing data, data science goes beyond that by incorporating the development and implementation of predictive models to make informed decisions. 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. They work at the intersection of mathematics, computer science, and domain expertise to solve complex problems and uncover hidden patterns in large datasets. Vasant Dhar writes that statistics emphasizes quantitative data and description. 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. Data science also integrates domain knowledge from the underlying application domain (e.g., natural sciences, information technology, and medicine). However, the definition was still in flux. Jeff Wu again suggested that statistics should be renamed data science. In 2003, Columbia University launched The Journal of Data Science. In 1996, the International Federation of Classification Societies became the first conference to specifically feature data science as a topic. 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. In 1962, John Tukey described a field he called "data analysis", which resembles modern data science. The modern conception of data science as an independent discipline is sometimes attributed to William S. 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. 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 used the term "data science" for the first time as an alternative name for statistics. Data science is multifaceted and can be described as a science, a research paradigm, a research method, a discipline, a workflow, and a profession.