

Data science is an interdisciplinary field[8] 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.[9] 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. 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.[10][11] 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.[12][13] 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.[14]

Relationship to statistics

Many statisticians, including Nate Silver, have argued that data science is not a new field, but rather another name for statistics.[15] Others argue that data science is distinct from statistics because it focuses on problems and techniques unique to digital data.[16] Vasant Dhar writes that statistics emphasizes quantitative data and description. In contrast, data science deals with quantitative and qualitative data (e.g. from images, text, sensors, transactions or customer information, etc) and emphasizes prediction and action.[17] Andrew Gelman of Columbia University has described statistics as a nonessential part of data science.[18]

Stanford professor David Donoho writes that data science is not distinguished from statistics by the size of datasets or use of computing and that many graduate programs misleadingly advertise their analytics and statistics training as the essence of a data-science program. He describes data science as an applied field growing out of traditional statistics.[19]