**Summary: “50 Years of Data Science”**

In the first ten pages of his article “50 Years of Data Science,” statistician David Donoho dispels popular misconceptions about data science among statisticians and in the media. A lot of statisticians consider data science to be a rebranding of statistics; however, on the opposite end of the spectrum, there are data analysts who consider statistics practically irrelevant to data science. Both are extreme stances, yet the notion of data science propagated by the media seems to reinforce statisticians’ views; the media sees data science merely as a way of handling large data sets, but statisticians have been handling similarly large data sets such as census data for years.

By investigating the literary origins of data science, Donoho attempts to identify the essence of data science and prescribe a direction for the evolution of data science as a field. Donoho most often refers to the article “The Future of Data Analysis” written by statistician James Tukey in 1962, in which Tukey coins the term “data analysis” in an effort to expand the field of statistics. According to Donoho, Tukey’s conception of the four major influences that act on data analysis today are still relevant. Tukey believed that the “formal theories of statistics,” “accelerating developments in computers and display devices,” “the challenge … of more and ever larger bodies of data”, and “the emphasis on quantification in an ever wider variety of disciplines” were the foundation of a new field: data science.

In his article “Data Science: An Action Plan for Expanding the Technical Areas of the field of Statistics,” statistician William Cleveland refine the idea of data science by suggesting that data science offers a flexibility that statistics does not. By their standard, academic statisticians focus too much on theoretical statistics, while truly learning from the data requires using techniques which theoretical statisticians typically do not, such as multidisciplinary investigations, models and methods for data, computing with data, pedagogy, and tool evaluation. Donoho’s own perspective seems to be in line with both Tukey and Cleveland, though Donoho imposes the additional constraint that data science is responsible for contributing to the evolution of science as a whole, and data scientists have the responsibility to use data science techniques even to predict the impact of data science itself on other areas of science.