Summary of *50 years of Data Science*

In recent years many Data Science programs have sprung up at universities across the country. Data Science is described as a field of study that includes collecting, analyzing and interpreting large volumes of data with applications in many fields. Many feel that Data Science is just a relabeling of Statistics, while others feel that Statistics is not very important to Data Science. Many argue that Data Science is necessary for handling larger quantities of data than Statistics is capable of, but Statistics has always worked with large data sets. Computer Scientists claim that Data Scientists have special skills necessary to handle data sets too large to fit on a single processor, but these skills are merely for handling problems caused by multiple processors and networking, rather than skills to solve interesting scientific problems. Commercially, prominent successes by large IT companies have highlighted the value of "big data" and driven demand for people with training in advanced computing and database knowledge in addition to statistics. Finally, the author advocates for a new version of Data Science that would be an extension of statistics driven by intellectual, rather than commercial, needs.

References

[1] Donoho, David, "50 years of Data Science," *Tukey Centennial workshop*, Princeton, NJ, Sept. 2015