1154-AE-2852 **Daniel Kaplan***. *Stats for data science.*

The canonical topics of college-level statistics are not the result of thoughtful examination of conceptual roots of the field. They are historically contingent, reflecting the needs, resources, and understanding of the era in which they were developed. Almost all of the topics of introductory statistics stem from the period from 1830 to 1925. They reflect the establishment of sociology, early genetics, and experiments on the bench-top or in agricultural field stations. Methods were tailored to very small amounts of data and calculation by hand.

Needless to say, needs and opportunities are different today. Data are abundant and multivariate; hypotheses are investigated by the dozens (nutrition research) or hundreds of thousands (genomics); an underlying methodology is machine learning; data are used to inform decisions, necessitating responsible inference about causation by adjusting for covariates. This new situation has resulted in a merging of components of computer science and statistics into "data science." In my presentation, I'll examine the appropriate statistical underpinnings for a meaningful engagement with data science, pointing out how they differ and sometimes contradict the topics of the century-old canon.

(Received September 17, 2019)