Many of the world's biggest discoveries and decisions in science, technology, business, medicine, politics, and society as a whole, are now being made on the basis of analyzing massive data sets, but it is surprisingly easy to come to false conclusions from data analysis alone, and privacy of data connected to individuals can be a major concern. This course provides a broad introduction to big data: historical context and case studies; privacy issues; data analysis techniques including databases, data mining, and machine learning; sampling and statistical significance; data analysis tools including spreadsheets, SQL, Python, R; data visualization techniques and tools. Tools and techniques are hands-on but at a cursory level, providing a basis for future exploration and application.

**CS102 is aimed primarily at students who may not major in CS but want to learn about big data and apply that knowledge in their areas of study.** The only required background is high school AP computer science, CS106A, or other equivalent programming experience. Comfort with statistics and spreadsheets is helpful but not required.