

Kathleen McGrath

Professor Cohen

Computer Science Principles 1030

May 5, 2023

## Khan Academy “Data Analysis”

### Data Tools

Data tools are software programs or applications created to simplify data analysis operations and increase accuracy and efficiency. As they offer features for data manipulation, visualization, and statistical analysis, these tools are crucial for managing large and complex datasets. Learners are introduced to some basic data tools that are frequently used in the field in the Data Analysis lesson on Khan Academy.

### Big Data

In the context of data analysis, "big data" refers to extraordinarily massive and complicated datasets that are more complex than what can be handled by conventional methods for data processing and analysis. This lesson exposes learners to the methods used to evaluate large datasets while also examining the special opportunities and difficulties posed by big data. Big data is distinguished by its quantity, speed, and variety.

### Bias in Machine Learning

When referring to bias in the context of data analysis, it is meant to describe the systematic and repeatable mistakes or prejudices that might happen during the creation and use of machine learning models. This lesson examines the effects of bias in machine learning and offers tips on how to spot it and lessen its effects. At many points in the model-development process, machine learning bias might appear. The training data itself is one of the main sources of bias.

### Unit Test

The unit test evaluates the student's knowledge of fundamental ideas and data analysis abilities. It covers a wide range of subjects and might have inquiries about theoretical concepts, real-world applications, and data analytic techniques. The test tested our performance on specific data analysis tasks, our knowledge of key data analysis concepts such as data visualization, statistical analysis, data manipulation, and result interpretation, as well as our ability to comprehend the findings of statistical studies. Test questions cover subjects like correlation, probability, hypothesis testing, and descriptive statistics.

