

Vincent Dufour

Professor Blanche Cohen

Computer Science Principles

May 5th 2023

Final Project Khan Academy Summary

Data Tools - This section starts by talking about how data is stored, in text files, in csv files which are then used in spreadsheets like Excel or Google Sheets, or a more complex form would be using SQL. Now that the data is stored, we can use mathematical tools available to start finding information like average across a column of values, sums, maxes or mins, etc. Once we use these, we can use graphs to start visualizing how our data interacts across an x and y axis.

Big Data - Next, we learn about the size of data and different types of data. The world runs on data and it takes up a lot of space. We also review sources of data, how it's gathered and where it's stored. Lastly, we went over the biggest problems data's faced with: storage, processing power, and making sure we use data responsibly.

Bias in Machine Learning - We take a detour and talk here about how machine learning's only interaction with the world is its data banks and how that may harbor unwanted biases. Neural networks learn by making its own connections from a huge database, reinforced by positive or negative reinforcement when the neural network makes accurate connections. We then reviewed how the way data is collected introduces multiple avenues for biases to seep through.