**IBM Data Science Coursera Certificate – Tools for Data Science (Course 2)**

**Week 2**

**Notes:**

**Categories of Data Science Tools**

A picture containing text, screenshot, font, number

Description automatically generated

* Python is an open-source platform while R is a free software.

**Summary:**

* You should select a language to learn depending on your needs, the problems you are trying to solve, and whom you are solving them for.
* The popular languages are Python, R, SQL, Scala, Java, C++, and Julia.
* For data science, you can use Python's scientific computing libraries like Pandas, NumPy, SciPy, and Matplotlib.
* Python can also be used for Natural Language Processing (NLP) using the Natural Language Toolkit (NLTK).
* Python is open source, and R is free software.
* R language’s array-oriented syntax makes it easier to translate from math to code for learners with no or minimal programming background.
* SQL is different from other software development languages because it is a non-procedural language.
* SQL was designed for managing data in relational databases.
* If you learn SQL and use it with one database, you can apply your SQL knowledge with many other databases easily.
* Data science tools built with Java include Weka, Java-ML, Apache MLlib, and Deeplearning4.
* For data science, popular program built with Scala is Apache Spark which includes Shark, MLlib, GraphX, and Spark Streaming.
* Programs built for Data Science with JavaScript include TensorFlow.js and R-js.
* One great application of Julia for Data Science is JuliaDB.