Choosing the right tool for the job

As a data analyst, you will usually have to decide which program or solution is right for the particular project you are working on. In this reading, you will learn more about how to choose which tool you need and when.

Depending on which phase of the data analysis process you're in, you will need to use different tools. For example, if you are focusing on creating complex and eye-catching visualizations, then the visualization tools we discussed earlier are the best choice. But if you are focusing on organizing, cleaning, and analyzing data, then you will probably be choosing between spreadsheets and databases using queries. Spreadsheets and databases both offer ways to store, manage, and use data. The basic content for both tools are sets of values. Yet, there are some key differences, too:

Spreadsheets	Databases
Software applications	Data stores - accessed using a query language (e.g. SQL)
Structure data in a row and column format	Structure data using rules and relationships
Organize information in cells	Organize information in complex collections
Provide access to a limited amount of data	Provide access to huge amounts of data
Manual data entry	Strict and consistent data entry
Generally one user at a time	Multiple users
Controlled by the user	Controlled by a database management system

You don't have to choose one or the other because each serves its own purpose. Generally, data analysts work with a combination of the two, as both tools are very useful in data analytics. For example, you can store data in a database, then export it to a spreadsheet for analysis. Or, if you are collecting information in a spreadsheet, and it becomes too much for that particular platform, you can import it into a database. And, later in this course, you will learn about programming languages like R that give you even greater control of your data, its analysis, and the visualizations you create.

As you continue learning about these important tools, you will gain the knowledge to choose the right tool for any data job.

Mark as completed