

From spreadsheets to SQL to R

Although the programming language R might be new to you, it actually has a lot of similarities to the other tools you have explored in this program. In this reading, you will compare spreadsheet programs, SQL, and R to have a better sense of how to use each moving forward.



Spreadsheets, SQL, and R: a comparison

As a data analyst, there is a good chance you will work with SQL, R, and spreadsheets at some point in your career. Each tool has its own strengths and weaknesses, but they all make the data analysis process smoother and more efficient. There are two main things that all three have in common:

- **They all use filters:** for example, you can easily filter a dataset using any of these tools. In R, you can use the filter function. This performs the same task as a basic SELECT-FROM-WHERE SQL query. In a spreadsheet, you can create a filter using the menu options.

- **They all use functions:** In spreadsheets, you use functions in formulas, and in SQL, you include them in queries. In R, you will use functions in the code that is part of your analysis.

The table below presents key questions to explore a few more ways that these tools compare to each other. You can use this as a general guide as you begin to navigate R.

Key question	Spreadsheets	SQL	R
What is it?	A program that uses rows and columns to organize data and allows for analysis and manipulation through formulas, functions, and built-in features	A database programming language used to communicate with databases to conduct an analysis of data	A general purpose programming language used for statistical analysis, visualization, and other data analysis
What is a primary advantage?	Includes a variety of visualization tools and features	Allows users to manipulate and reorganize data as needed to aid analysis	Provides an accessible language to organize, modify, and clean data frames, and create insightful data visualizations
Which datasets does it work best with?	Smaller datasets	Larger datasets	Larger datasets
What is the source of the data?	Entered manually or imported from an external source	Accessed from an external database	Loaded with R when installed, imported from your computer, or loaded from external sources
Where is the data from my analysis usually stored?	In a spreadsheet file on your computer	Inside tables in the accessed database	In an R file on your computer
Do I use formulas and functions?	Yes	Yes	Yes

Key question	Spreadsheets	SQL	R
Can I create visualizations?	Yes	Yes, by using an additional tool like a database management system (DBMS) or a business intelligence (BI) tool	Yes