# R

In this notebook, we perform the following data engineering tasks:

1. Connect to a PostgreSQL database and extract data from a table using **dbConnect()** and **dbGetQuery()**.
2. Perform data cleaning and transformation using **filter()** and **mutate()** from the **dplyr** package.
3. Aggregate data using **group\_by()** and **summarize()** from **dplyr**.
4. Export data to a file using **write.csv()**.
5. Load data from a file using **fread()** from the **data.table** package.
6. Merge data using **full\_join()** from **dplyr**.
7. Load merged data to a database table using **dbWriteTable()**.
8. Disconnect from the database using **dbDisconnect()**.
9. Visualize data using **ggplot2**.

Of course, this is just a small subset of the tasks that a data engineer might perform, but it provides a starting point for understanding the kind of work that a data engineer might do.

# Python

In this notebook, we perform the following data engineering tasks:

1. Connect to a PostgreSQL database and extract data from a table using **psycopg2** and **pd.read\_sql\_query()**.
2. Perform data cleaning and transformation using **dropna()** and **np.where()**.
3. Aggregate data using **groupby()** and **agg()**.
4. Export data to a file using **to\_csv()**.
5. Load data from a file using **pd.read\_csv()**.
6. Merge data using **pd.merge()**.
7. Load merged data to a database table using **csv.writer()** and **cur.copy\_from()**.
8. Disconnect from the database using **cur.close()** and **conn.close()**.
9. Visualize data using **matplotlib.pyplot**.