sqldf Cheat Sheet - SQL for R Users

Setup

install.packages("sqldf")

library(sqldf)

Basic Querying

Select columns: SELECT column1, column2 FROM df

Filter rows: SELECT * FROM df WHERE column1 = 'value'

Sort rows: SELECT * FROM df ORDER BY column1 DESC

Limit output: SELECT * FROM df LIMIT 10

Distinct values: SELECT DISTINCT column1 FROM df

Rename columns: SELECT column1 AS new_name FROM df

Aggregation

Count rows: SELECT COUNT(*) FROM df

Group + summarize: SELECT column1, COUNT(*) FROM df GROUP BY column1

Average, sum, etc.: SELECT column1, AVG(column2) FROM df GROUP BY column1

Filter groups: ... GROUP BY column1 HAVING COUNT(*) > 100

Joins

INNER JOIN: SELECT * FROM df1 INNER JOIN df2 ON df1.key = df2.key

LEFT JOIN: SELECT * FROM df1 LEFT JOIN df2 ON df1.key = df2.key

Note: sqldf only supports INNER and LEFT JOINs (no RIGHT/FULL joins).

Conditional Logic

Multiple filters: WHERE column1 = 'X' AND column2 > 10

IN operator: WHERE column1 IN ('A', 'B', 'C')

LIKE pattern matching: WHERE column1 LIKE '%text%'

CASE WHEN: SELECT CASE WHEN col > 10 THEN 'High' ELSE 'Low' END AS label

Useful Patterns

Top 5 categories:

SELECT category, COUNT(*) AS n FROM df GROUP BY category ORDER BY n DESC LIMIT 5

Average by group:

SELECT group_var, AVG(value) FROM df GROUP BY group_var

sqldf Cheat Sheet - SQL for R Users

Join example:

```
SELECT a.id, b.label FROM df1 a INNER JOIN df2 b ON a.id = b.id
```

Tips

- sqldf() runs SQL on data frames in your R environment.

- Tables = your data frames (hospital, counties, etc.).

- Use AS to rename columns.

- Check case sensitivity (R is case-sensitive).