**Thu thập dữ liệu**

* Import dữ liệu từ cơ sở dữ liệu
* Import dữ liệu thô: như file xlsx,csv,txt,...
* Thu Thập dữ liệu Web
* Thu Thập dữ liệu qua API

Các packages cần thiết:

* readr : đọc các file như csv, tsv,...
* readxl : đọc file excel
* DBI, RMySQL, RPostgres, RSQLite: đọc dữ liệu từ database
* httr và rvest: thu thập dữ liệu web Được sử dụng để trích xuất dữ liệu web. httrXử lý các yêu cầu HTTP cho API, trong khi rvestdùng để thu thập nội dung HTML từ web.
* jsonlite và xml2 : để Import dữ liệu ở định dạng JSON hoặc XML

**Chọn gói phù hợp:**

* **Tệp văn bản (CSV, TSV, v.v.): readr hoặc data.table::fread().**
* **Tệp Excel: readxl .**
* **Tệp phần mềm thống kê (SPSS, SAS, Stata): haven hoặc foreign.**
* **Cơ sở dữ liệu: DBI cùng với gói trình điều khiển cơ sở dữ liệu cụ thể (ví dụ: RMySQL).**
* **Dữ liệu web: httr dành cho API, rvestđể thu thập dữ liệu web.**
* **JSON/XML: jsonlite hoặc xml2.**