

Carregamento e Análise de Dados em R: formatos Parquet e JSON

Dados Utilizados:

1. Dados de Táxi de Nova York (Formato Parquet)

- Fonte: NYC Taxi & Limousine Commission
- URL: https://d37ci6vzurychx.cloudfront.net/trip-data/yellow_tripdata_2023-01.parquet
- Características: formato colunar Parquet, altamente eficiente para grandes volumes de dados. Contém informações detalhadas sobre corridas de táxi amarelo e inclui dados temporais, geográficos e financeiros.
- Estrutura dos dados: 19 colunas incluindo data/hora de início e fim da corrida, localização de partida e destino (coordenadas geográficas), distância percorrida, tarifas e valores de pagamento e informações sobre taxas, impostos e gorjetas.
- Processo de Carregamento:
 - 1) Verificação de existência do arquivo local
 - 2) Download caso não exista
 - 3) Leitura com `read_parquet()`
 - 4) Conversão para data frame com `as.data.frame()`

2. Dados de Usuários (Formato JSON)

- Fonte: JSONPlaceholder API de teste
- URL: <https://jsonplaceholder.typicode.com/users>
- Características: formato JSON leve, amplamente utilizado em APIs web. Dados de usuários fictícios para testes e estrutura aninhada com objetos dentro de objetos.
- Estrutura dos dados: Informações básicas (ID, nome, nome de usuário, e-mail), dados de contato (telefone, website), endereço completo (rua, cidade, CEP, coordenadas geográficas) e informações da empresa (nome e frase de negócio).
- Processo de Carregamento:
 - 1) Verificação de existência do arquivo local
 - 2) Download caso não exista
 - 3) Leitura e conversão com `fromJSON()`

Pacotes Utilizados:

- arrow: leitura de arquivos Parquet
- jsonlite: processamento de arquivos JSON
- tsibble: criação e manipulação de séries temporais
- dplyr: manipulação e transformação de dados

Conclusões:

- Parquet: eficiente para grandes volumes de dados
- JSON: flexível e amplamente utilizado em APIs web