

Como ler os microdados do ENEM no R?

Perguntada 3 anos, 6 meses atrás Ativa 3 anos, 2 meses atrás Vista 4mil vezes

13 O INEP (Instituto Nacional de Estudos e Pesquisas Educacionais Anísio Teixeira) disponibiliza para download os microdados do ENEM (e outros) a partir deste [link](#). Estes microdados são os dados crus do ENEM, com informações na granularidade de alunos.

Ao fazer o download dos microdados do [ENEM 2014](#), por exemplo, você baixa um arquivo com extensão zip de aproximadamente 1,2Gb. Ao extrair o conteúdo deste arquivo há na estrutura de pastas toda a documentação deste conjunto de dados e também os microdados. Na pasta DADOS há o arquivo MICRODADOS_ENEM_2014.csv. Esse arquivo tem 6Gb de tamanho tal que comandos como:

```
dados <- read.csv('MICRODADOS_ENEM_2014.csv', header=T)
```

não funcionam na maioria dos PC's comuns em virtude da limitação da memória RAM.

Qual é a melhor alternativa para ler este conjunto de dados no R em um PC comum? Gostaria especialmente de métodos concretos para efetuar esta leitura com a ajuda de pacotes do R e TAMBÉM utilizando sistemas de gerenciamento de banco de dados.

OBS: Estou ciente com relação a pergunta [Estratégias para analisar bases de dados muito grandes em R \(que não caibam na memória RAM\)](#), tendo inclusive participado como um dos respondentes. ENTRETANTO, a pergunta teve um caráter mais genérico, não produzindo nas respostas nenhum código em R que pudesse resolver o problema específico desta questão.

r

csv

big-data

editada 13/04/17 às 12:59



Comunidade ♦

1

perguntada 24/09/16 às 6:41



Flavio Barros

1.609 8 23

4 Possível duplicata de [Estratégias para analisar bases de dados muito grandes em R \(que não caibam na memória RAM\)](#) – Molx 25/09/16 às 0:59

@Molx eu fui uma das pessoas que respondeu essa questão. Aquela era uma questão mais conceitual, com um caráter mais aberto. Esta em especial é bem mais específica e relacionada a um conjunto de dados em específico. Como resultado dessa questão também ficará documentado o código em R necessário para realizar essa tarefa. Muita gente vai no google especificamente para procurar uma solução para esse problema. – Flavio Barros 25/09/16 às 14:18

Também não acho que seja duplicata. – Carlos Cinelli 27/09/16 às 14:34

@CarlosCinelli acho que o pessoal quando ganha acesso a fila de análise nem lê direito a questão e já começa a querer "agir". Basta ler com calma para perceber que são questões diferentes.
– Flavio Barros 28/09/16 às 3:41

Usar o Hadoop e o MapReduce, afim de filtrar e reduzir a quantidade de dados, seria uma opcao que lhe ajuda ? – Gilberto Santos 31/01/17 às 10:41