
Exercício - Manipulação de Dados II

Marcelo Prudente

10/02/2020

1 Leitura dos dados do Censo Escolar

1.1 Exercício

Na prática, a análise de dados envolve uma série encadeada de operações. Vamos utilizar os dados do Censo Escolar para aplicar a manipulação de dados.

Primeiros passos:

- baixar os dados do Censo Escolar (verifique a pasta do github) para o seu computador.
- crie um novo projeto do R chamado **exercicio_r**
- abrir o arquivo dicionário de dados - ele proverá informações valiosas para o exercício

Segundo passo: análise dos dados.

1. Importe os dados das matrículas e para o seu ambiente **R**
 - Qual função utilizar? Lembre-se: as bibliotecas devem ser chamadas ao início de cada seção.
2. Verifique a estrutura de dados.
3. Faça os seguintes sumários estatísticos - salve essas informações em objetos:
 - quantos alunos estudam no Estado de referência?
 - quais os 10 municípios com mais alunos?
 - informe o número de alunos por tipo de dependência, localização e categoria de escola.
 - quantos alunos estudam em tempo integral? Há prevalência de idade ou de etapa de ensino? Ordene os dados se necessário.
 - agrupe os alunos em blocos de idade de 4 anos - utilize a função **cut** e conte o total de alunos por esses blocos em cada município.
4. Agora, vamos olhar especificamente para os **alunos do ensino médio** (obs: use o *filter()*).
 - apresente o percentual de alunos que cursam ensino integral de acordo com a idade (antes você deve criar uma variável booleana que indique se o aluno estuda em regime integral ou não. Um aluno em tempo integral passa mais de 7h na escola por dia). Arredonde os valores com dois dígitos.
 - quantas turmas apresentam algum aluno com necessidade especial. Qual o percentual dessas turmas no total de turmas do município?
 - quantos alunos em média tem cada turma?
5. Importe dos dados das Escolas e indique:

- O percentual e o total de escolas com água filtrada, energia elétrica, esgoto sanitário, biblioteca e acesso à internet. Há diferenças entre as escolas particulares e públicas nesse quesito? Ainda, indique se há distinção nessas variáveis entre escolas federais, municipais e estaduais?

6. Importe os dados dos docentes.

- crie uma coluna com a data de nascimento completa de nascimento do docente - a data deve ter formato de data.
- indique o percentual de docentes por tipo de escolaridade e sexo. Antes, verifique se há docentes duplicados para não viesar a informação.
- Caso haja docentes duplicados, filtre o banco apenas pelos casos de código do docente duplicado e aponte a média de vínculos deles.
- Quais as principais formações dos professores (utilize o “código do curso 3 da escolaridade superior”). Mescle o objeto obtido com as informações do Anexo 5 também presente no link do one-drive. + Filtre os dados obtidos anteriormente para encontrar apenas os professores de matemática - salve em um novo objeto.
- Identifique as 5 Instituições de Ensino que mais formaram professores nesse Estado. Qual o percentual de docentes que saíram dessas IES? Identifique-as pelo nome.

1.2 Análise dos dados do IDEB

7. Por fim, importe a planilha *ideb_anos_iniciais_2017.xlsx*. Esses dados estão em uma planilha de excel e devem ser baixados com o pacote *readxl*. Também lembre de pular algumas linhas, pois os títulos das variáveis não estão evidentes.

- Com essas informações, selecione apenas as colunas referentes às informações municipais, à rede e ao ideb de 2017.
- Quais os dez municípios com melhor IDEB? Em média nesses locais as escolas públicas tem melhores equipamentos do que os 10 municípios com pior IDEB?