Ministério da Educação

Instituto Nacional de Estudos e Pesquisas Educacionais Anísio Teixeira – INEP Diretoria de Estatísticas Educacionais

Microdados do Censo da Educação Básica 2018

Manual do Usuário

Instituto Nacional de Estudos e Pesquisas Educacionais Anísio Teixeira

Diretoria de Estatísticas Educacionais

Diretor: Carlos Eduardo Moreno Sampaio

Coordenação-Geral do Censo da Educação Básica

Coordenadora-Geral: Célia Cristina de Souza Gedeon Araújo

Coordenação-Geral de Controle de Qualidade e Tratamento das Informações Educacionais

Coordenador-Geral: Fábio Pereira Bravin

Coordenação de Sistematização da Informação

Coordenador: Clodoaldo de Oliveira Lemes
Maria das Dores Pereira Rosa
Márlei Afonso de Almeida
Maruska Pereira de Almeida
Melissa Riani Costa Machado
Palu Silveira Abe
Thaysa Guimarães Souza
Pedro Gonçalves Costa
Marcos Vinicius Rodrigues Lopes

Coordenação de Indicadores e Controle de Qualidade da Educação Básica

Coordenador: Vitor Passos Camargos Julio Cesar de Lima Filgueiras Laís Raiane Miguel Amaral Lana Torres Barreto Raphael Igor da Silva Corrêa Dias Vanessa Nespoli de Oliveira

Apresentação

O Instituto Nacional de Estudos e Pesquisas Educacionais Anísio Teixeira (Inep), autarquia vinculada ao Ministério da Educação, tem como missão institucional produzir e disseminar informações educacionais. As estatísticas e avaliações produzidas pelo Inep visam fornecer os subsídios para a formulação e implementação de políticas voltadas para a melhoria contínua da educação no país.

O Inep desenvolve uma série de ações de disseminação de dados que se encontram disponíveis no Centro de Informações e Biblioteca em Educação (Cibec), organizados para atender presencialmente aos usuários, e em seu site (http://portal.inep.gov.br).

O presente documento figura entre os instrumentos utilizados por este Instituto para garantir a transparência requerida pela sociedade e promover a participação dos diferentes agentes sociais envolvidos nas ações educativas.

Nesse contexto, os Microdados passaram a ser estruturados em formato CSV (Comma-Separated Values), e seus dados estão delimitados por *Pipe* (|), de modo a garantir que praticamente qualquer *software* estatístico, inclusive *open source*, consiga importar e carregar as bases de dados.

Devido à amplitude de nossas bases, os arquivos foram divididos por região geográfica (Norte, Nordeste, Sudeste, Sul e Centro-Oeste), tanto para as variáveis de Matrículas, quanto para as de Docentes.

Por meio dos Microdados é possível obter um amplo panorama da educação brasileira e, como importante ferramenta de transparência, é indubitavelmente um rico acervo sobre a educação básica do nosso país e uma fonte segura e eficaz de obtenção de dados, acessíveis aos pesquisadores, estudantes, gestores e sociedade em geral.

Sumário

- 1 DIRETÓRIOS
- 2 ORIENTAÇÕES PARA ABRIR OS ARQUIVOS CSV NO *SOFTWARE* R
- 3 ORIENTAÇÕES PARA ABRIR OS ARQUIVOS CSV NO SOFTWARE SPSS
- 4 ORIENTAÇÕES PARA ABRIR OS ARQUIVOS CSV NO *SOFTWARE* SAS

1 - Diretórios

DADOS

Estão disponíveis no site do Inep (http://portal.inep.gov.br/web/guest/microdados) os arquivos com os Microdados do Censo da Educação Básica 2018 (Escolas, Turmas, Matrículas e Docentes) em formato CSV delimitado por *Pipe* (|). Os arquivos encontram-se compactados (em formato .zip) pelo *software* 7zip e devem ser descompactados no diretório C:\ do seu disco rígido.

LEIA-ME

Este presente instrumento traz instruções fundamentais para a correta extração e manipulação dos dados educacionais. É importante enfatizar que os Microdados estão estruturados em CSV (Comma-Separated Values), que é basicamente uma implementação de arquivos de texto separados por um delimitador (pipe), sendo compatível com praticamente qualquer sistema moderno, incluindo: softwares estatísticos ou de manipulação de dados, servidores de dados e planilhas eletrônicas (Excel, Calc, etc), conforme formatos adotados ou recomendados pelos Padrões de Interoperabilidade de Governo Eletrônico – ePING.

Como sugestão, disponibilizamos neste documento algumas orientações para abrir os arquivos nos principais softwares estatísticos e de acesso à base de dados utilizados pelos pesquisadores, que permitem o acesso, manipulação, análise e apresentação de informações, tais como o R (que é open source, composto por linguagem e ambiente de desenvolvimento integrado para cálculos estatísticos e relatórios), IBM SPSS e SAS (Statistical Analysis System).

FILTROS DA EDUCAÇÃO BÁSICA

É um documento que norteia o usuário na obtenção de dados corretos e em sintonia com os valores publicados nas Sinopses Estatísticas. Nele, são elencados, de forma simples, os principais filtros e funções que devem ser aplicados nas diferentes etapas e modalidades de ensino, considerando as quatro tabelas principais: Escola, Turma, Matrícula e Docente.

ANEXO I - Dicionário de Dados e Tabelas Auxiliares

- Contém, em formato .xlsx (Excel), os seguintes Dicionários das Variáveis do Censo da Educação Básica 2018:
 - TABELA DE ESCOLA
 - TABELA DE TURMA
 - TABELA DE MATRÍCULA
 - TABELA DE DOCENTE
 - TABELA DE ETAPA DE ENSINO
- Contém, em formato .xlsx (Excel), os seguintes Tabelas Auxiliares do Censo da Educação Básica 2018:
 - TABELA DE LÍNGUAS INDÍGENAS
 - TABELA DE CURSOS DA EDUCAÇÃO PROFISSIONAL TÉCNICA
 - TABELA DE TIPO DE ATIVIDADE COMPLEMENTAR POR CATEGORIA/ÁREA
 - TABELA DE PAÍSES
 - TABELA DE CURSOS DE FORMAÇÃO SUPERIOR
 - TABELA DE IES

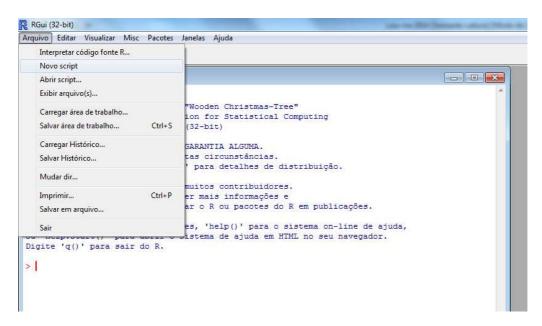
ANEXO II - Questionários do Censo da Educação Básica

- Contém, em formato PDF os questionários do Censo da Educação Básica.

2 - ORIENTAÇÕES PARA ABRIR OS ARQUIVOS CSV NO SOFTWARE R

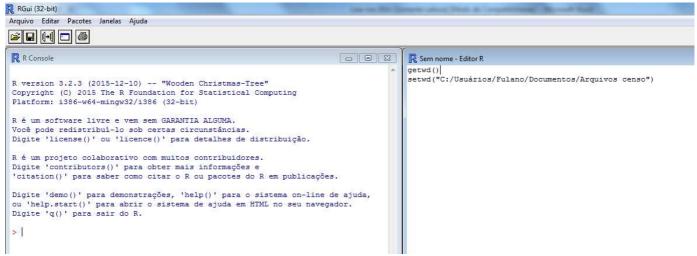
2.1 - Abertura de bases pequenas (ESCOLAS e TURMAS):

1) Abra um Novo *script* clicando na aba "Arquivo" "Nove script".



2) Certifique-se de que o Diretório de Trabalho no R corresponde ao local onde estão presentes os arquivos. Copie o comando abaixo e cole na nova janela aberta. Atenção: Para rodar cada comando, posicione o cursor do *mouse* na linha do comando e pressione Ctrl+R. Alternativamente, o comando pode ser copiado e colado diretamente no console.

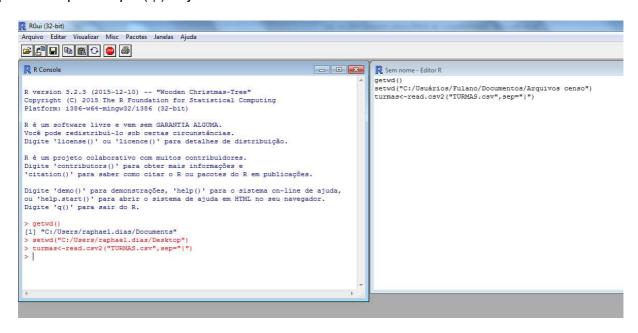
getwd() #Se o caminho não estiver correto, use a função a seguir para alterá-lo: setwd("C:/Usuários/Fulano/Documentos/Arquivos censo") #Exemplo de caminho. Lembre-se de usar barra simples ("/") ou barra invertida dupla ("\").



3) Para abrir, por exemplo, a base "TURMAS.csv", copie o comando abaixo e cole no *script*. Rode o comando pressionando Ctrl+R.

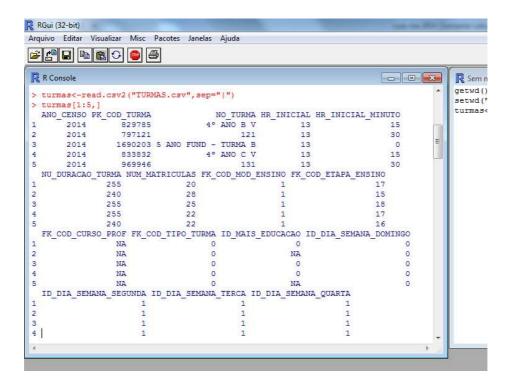
turmas<-read.csv2("TURMAS.csv",sep="|")

É importante que o *Pipe* (|) seja utilizado como delimitador.

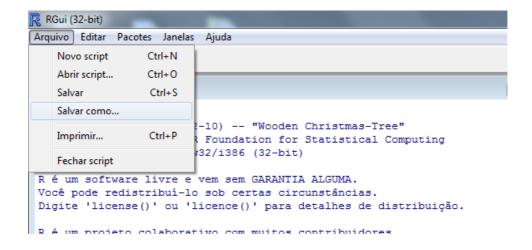


4) Após esses procedimentos, a base de dados já estará aberta, permitindo sua análise. Para conferir se a base foi lida corretamente, é possível visualizar algumas linhas copiando o comando abaixo e colando no *script*. Rode o comando pressionando Ctrl+R. Nesse exemplo, são visualizadas as cinco primeiras linhas de cada coluna.

turmas[1:5,]

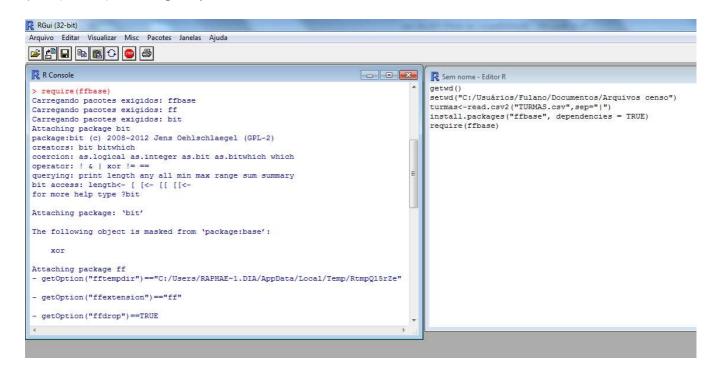


Sugerimos salvar o script criado para posterior utilização. Acesse a aba "Arquivo" "Salvar como...".



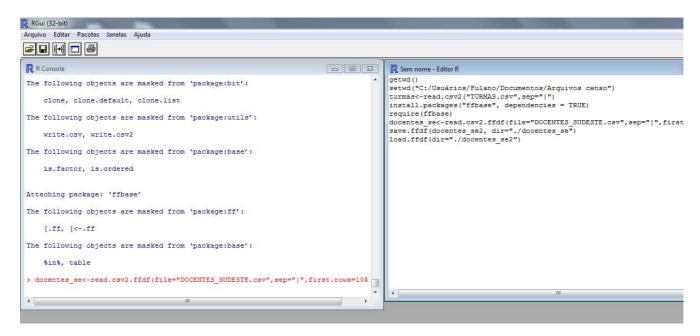
2.2 - Abertura de bases maiores (MATRÍCULAS e DOCENTES):

- 1) O software R, como padrão, trabalha com as bases de dados utilizando a memória RAM do computador. Entretanto, o carregamento de bases muito grandes (por exemplo, mais de 1 milhão de linhas) utilizando a leitura tradicional (como "read.table" ou "read.csv") pode sobrecarregar o computador, ou mesmo resultar em erro por falta de memória. Dessa forma, para a leitura das bases de MATRICULAS e DOCENTES, que possuem mais de 10 milhões de linhas (Brasil), faz-se necessário o uso de pacotes adicionais. Sugerimos o uso do pacote "ffbase" para trabalhar com essas bases, tendo em vista que o mesmo faz uso do disco rígido em detrimento da memória RAM. Evidentemente, as bases de ESCOLAS e TURMAS também podem ser carregadas utilizando esse pacote, sendo recomendável quando houver necessidade de cruzar informações entre as bases. O pacote "ffbase" armazena a base de dados no R como um objeto da classe "ffdf" - diferentemente da leitura tradicional, que gera um objeto da classe "data.frame". O objeto "ffdf" também permite a aplicação de algumas funções – não todas – que são utilizadas com objetos da classe "data.frame" (por exemplo, "table", "merge" e "transform"). Para aplicação de filtros nas bases "ffdf", recomendamos o uso da função "ffwhich" (veja a ajuda da função para maiores informações: "?ffwhich"). Além disso, para concatenar as bases de cada região (as bases Docentes e Matrículas estão separadas por região), uma abaixo da outra, é necessário utilizar a função "ffdfappend" (para mais informações: "?ffdfappend"). Informações adicionais estão disponíveis na ajuda do pacote (https://cran.r-project.org/web/packages/ffbase/ffbase.pdf).
- 2) Ainda utilizando o mesmo script, instale o pacote necessário para a manipulação de bases maiores (ffbase) copiando e colando os comandos abaixo. Rode os comandos pressionando Ctrl+R. install.packages("ffbase", dependencies = TRUE) #Instala o pacote require(ffbase) #Carrega o pacote



- 3) Caso você tenha aberto um novo *script*, será necessário certificar-se novamente de que o Diretório de Trabalho no R corresponde ao local onde estão presentes os arquivos por meio dos mesmos comandos já citados anteriormente:
 - getwd() #Se o caminho não estiver correto, use a função a seguir para alterá-lo: setwd("C:/Usuários/Fulano/Documentos/Arquivos censo") #Exemplo de caminho. Lembre-se de usar barra simples ("/") ou barra invertida dupla ("\\").
- 4) Para abrir o arquivo "DOCENTES_SUDESTE.csv", por exemplo, copie o comando abaixo e cole no *script*. Rode o comando pressionando Ctrl+R. O atributo "first.rows" determina a quantidade de linhas utilizadas para que o programa perceba o tipo/categoria das variáveis.

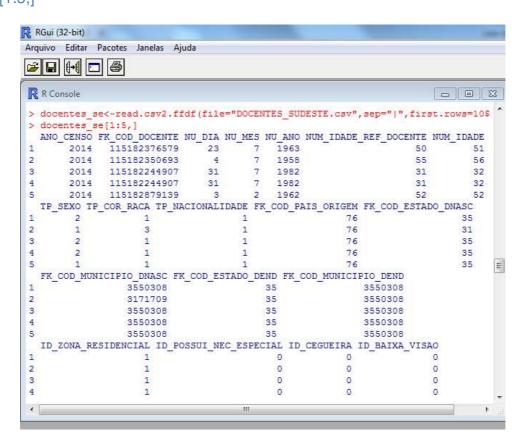
docentes_se<-read.csv2.ffdf(file="DOCENTES_SUDESTE.csv",sep="|",first.rows=100000) É importante que o *Pipe* (|) seja utilizado como delimitador.



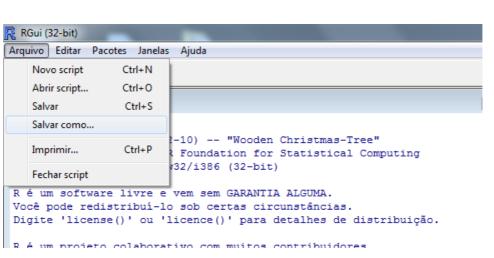
- 5) Bases grandes, quando carregadas pela primeira vez, podem demorar alguns minutos para concluir o processamento (dependendo das configurações do computador). Sugerimos salvar a base no formato 'ffdf' para que uma nova carga inicial não seja necessária numa próxima sessão.
 - save.ffdf(docentes_se, dir="./docentes_se")
- 6) Dados salvos no formato 'ffdf' podem ser carregados em um outro momento utilizando o comando abaixo. O carregamento da base com a função abaixo é praticamente instantâneo.

load.ffdf(dir="./docentes se")

7) Após esses procedimentos, a base de dados já estará aberta, permitindo sua análise. Para conferir se a base foi lida corretamente, é possível visualizar algumas linhas copiando o comando abaixo e colando no script. Rode o comando pressionando Ctrl+R. Nesse exemplo, são visualizadas as cinco primeiras linhas de cada coluna. docentes se[1:5,]

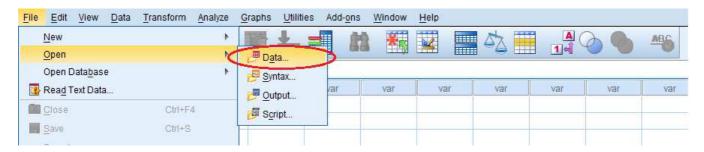


8) Sugerimos salvar o script criado para posterior utilização. Acesse a aba "Arquivo" "Salvar como...".

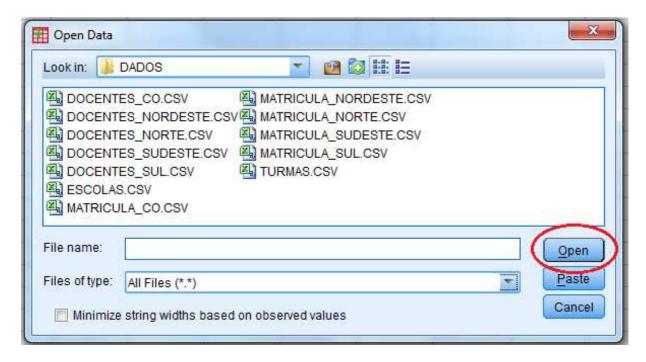


3 - ORIENTAÇÕES PARA ABRIR OS ARQUIVOS CSV NO SOFTWARE SPSS

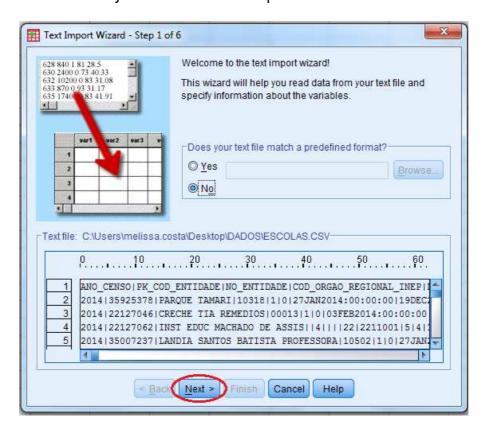
Abrir uma planilha no software SPSS e clicar na aba "File" "Oper" "Datar.



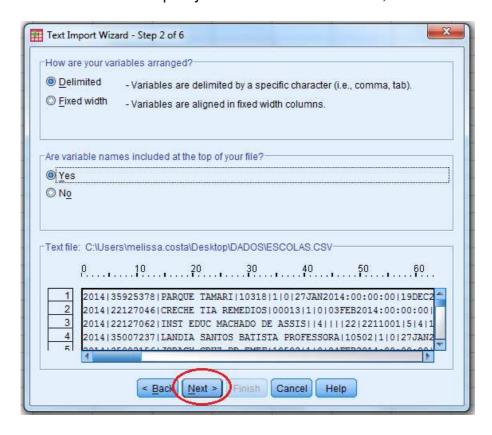
2) Em seguida, selecionar em "Files of type" a opção "All Files (*.*)", para que os arquivos em formato CSV possam ser visualizados. Selecionar o arquivo CSV desejado e clicar em "Open". No exemplo, os arquivos estão na pasta DADOS.



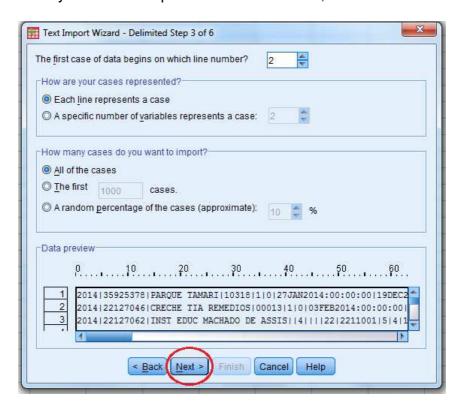
3) Manter a opção "No" em "Does your text file match a predefined format?" e clicar em "Next".



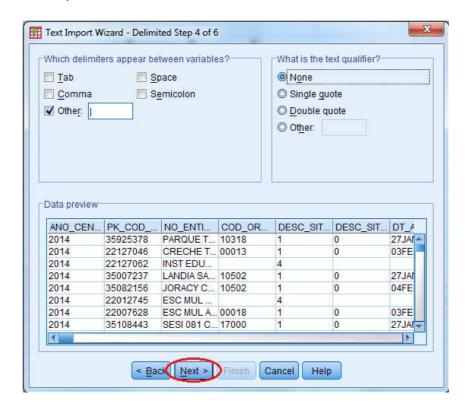
4) Manter a opção "Delimited" em "How are your variables arranged?" e alterar para a opção "Yes" em "Are variable names included at the top of your file?". Posteriormente, clicar em "Next".



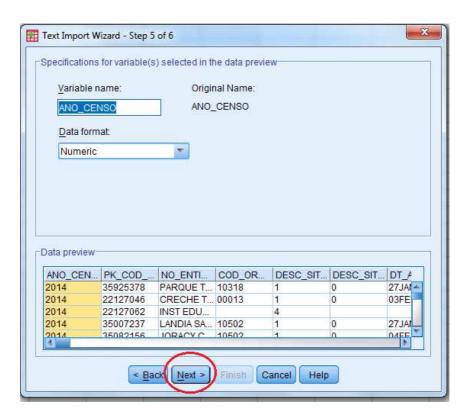
5) Manter a opção "2" em "The first case of data begins on which line number?". Manter, também, as opções "Each line represents a case" em "How are your cases represented?" e "All of the cases" em "How many cases do you want to import?". Posteriormente, clicar em "Next".



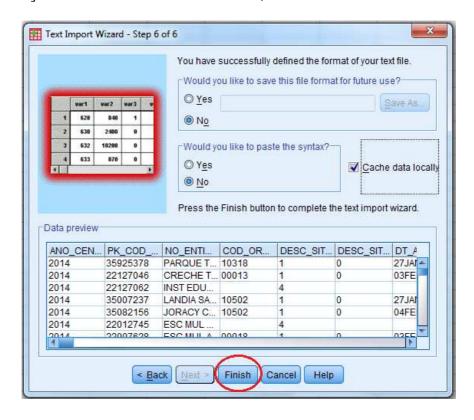
6) Desmarcar a opção "Space" e marcar a opção "Other" em "Which delimiters appear between variables?", utilizando o *Pipe* (|) como delimitador. Manter a opção "None" em "What is the text qualifer?". Posteriormente, clicar em "Next".



7) Manter as opções marcadas em "Specifications for variable(s) selected in the data preview" e clicar em "Next".

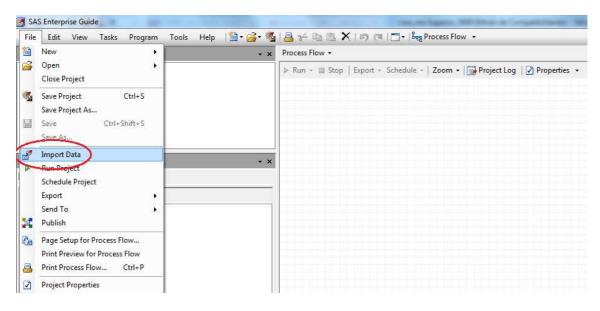


8) Na última etapa de "Text Import Wizard", é possível optar entre "Yes" e "No" em "Would you like to save this file format for future use?" e em "Would you like to paste the syntax?". Para finalizar o processo de importação de dados no *software* SPSS, basta clicar em "Finish".

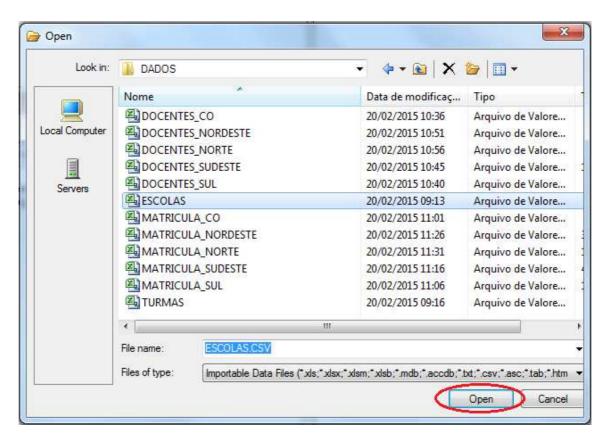


4 - ORIENTAÇÕES PARA ABRIR OS ARQUIVOS CSV NO SOFTWARE SAS

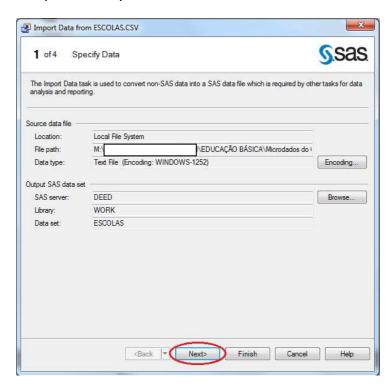
1) Abrir o software SAS e clicar na aba "File" "Import Data".



2) Em seguida, selecionar o arquivo CSV desejado e clicar em "Open". No exemplo, os arquivos estão na pasta DADOS.



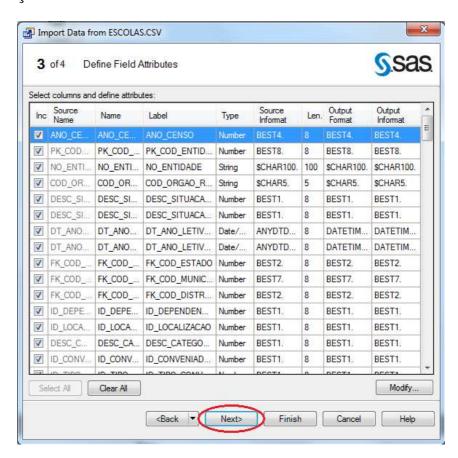
3) Na etapa seguinte, mantêm-se as opções marcadas em "Source data file" e pode-se optar pelo local onde será o *Output* da planilha. Após, clicar em "Next".



4) Manter a opção "Delimited fields" em "Text format" marcada e manter o *Pipe* (|) como delimitador. Manter as demais configurações, como a caixa de "File contains field names on record number:" marcada e o número 1 preenchendo o campo à frente, e manter em "Data records start at record number:" o número 2 preenchendo o campo à frente. Posteriormente, clicar em "Next".



5) Manter as configurações e clicar em "Next".



6) Para finalizar o processo de importação de dados no software SAS, basta clicar em "Finish".

