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

Microdados do Censo da Educação Superior 2018

Manual do Usuário

Setembro de 2019

Instituto Nacional de Estudos e Pesquisas Educacionais Anísio Teixeira

Presidente: Alexandre Ribeiro Pereira Lopes

Diretoria de Estatísticas Educacionais

Diretor: Carlos Eduardo Moreno Sampaio

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

Marcos Vinicius Rodrigues Lopes Maria das Dores Pereira Rosa Márlei Afonso de Almeida Maruska Pereira de Almeida Melissa Riani Costa Machado

Palu Silveira Abe Pedro Gonçalves Costa

Coordenação de Indicadores e Controle de Qualidade da Educação Superior

Coordenador: Willians Kaizer dos Santos Maciel

Isabella Trevisol Macêdo Renan Carlos Dourado Simone Poch Vieira Palma Thaysa Guimarães Souza Zilá Ribeiro de Ávila

Coordenação-Geral do Censo da Educação Superior

Coordenadora-Geral: Laura Bernardes da Silva

Arthur Otavio Araújo Bruna Mota de Lima Douglas Pereira Silva Gleidilson Costa Alves

Janaina Ferreira Ma

Katia Cristina da Silva Vaz

Lana Torres Barreto

Lucas Rocha Soares de Assis

Luciana Ferreira Artuso

Luciana Vieira de Almeida

Nara Núbia Vieira

Patricia Carolina Santos Borges

Patrick Sampaio

Rafael Arlon da Luz

Suellen Mary Kock Fachinetto

Thaiane Lima de Paula

Viviane Pereira Gangá

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.

Entre as informações educacionais produzidas pelo Inep, destacam-se os dados coletados no Censo da Educação Superior, levantamento de âmbito nacional, realizado anualmente pela Diretoria de Estatísticas Educacionais em todas as Instituições de Educação Superior (IES), públicas e privadas do país.

O Inep desenvolve uma série de ações de disseminação de dados que se encontram divulgados de forma ativa no portal do instituto na internet (www.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 foram 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.

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 superior 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
- ${f 2}$ ORIENTAÇÕES PARA ABRIR OS ARQUIVOS CSV NO ${f SOFTWARE}$ 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 Microdados do Censo da Educação Superior 2018 (DM_IES, DM_CURSO, DM_DOCENTE, DM_ALUNO, DM_LOCAL_OFERTA e TB_AUX_CINE_BRASIL) em formato CSV delimitados por *Pipe* (|). Os arquivos encontram-se compactados (em formato .zip) pelo *software* 7-zip 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 Microdados, como orientações para abrir os arquivos nos *softwares* R, SPSS e SAS. É, portanto, uma espécie de "Manual do Usuário" que traz os Dicionários das Variáveis do Censo da Educação Superior 2018.

FILTROS DA EDUCAÇÃO SUPERIOR

É um documento que norteia o usuário na obtenção de dados em sintonia com os valores publicados nas Sinopses Estatísticas produzidas pelo INEP. Nele, são elencados, de forma simples, os principais filtros e funções que devem ser aplicados.

ANEXO I - Dicionários de Dados e Tabelas Auxiliares

Contém, em formato .xlsx (Excel), o Dicionário de Dados do Censo da Educação Superior 2018 e também uma tabela auxiliar com o código e o nome dos países:

1. DICIONÁRIO DE DADOS

- TABELA DE ALUNO
- TABELA DE CURSO
- TABELA DE IES
- TABELA DE LOCAL DE OFERTA
- TABELA DE DOCENTE
- TABELA AUXILIAR CINE BRASIL

2. TABELA CONTENDO O NOME DO PAÍS DE ORIGEM OU NATURALIZAÇÃO

ANEXO II – Questionários do Censo da Educação Superior

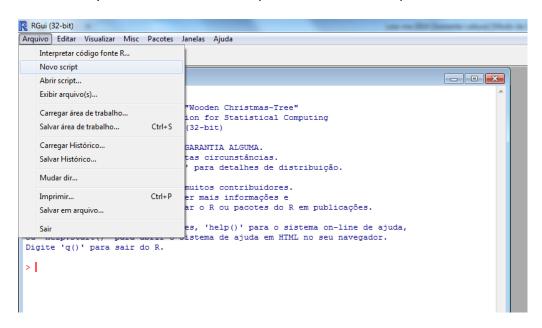
Contém, em formato .pdf (Portable Document Format), os seguintes questionários do Censo da Educação Superior 2018 e estão disponíveis para download na pasta anexos:

- MÓDULO IES
- MÓDULO CURSO
- MÓDULO DOCENTE
- MÓDULO ALUNO

2 - ORIENTAÇÕES PARA ABRIR OS ARQUIVOS CSV NO SOFTWARE R

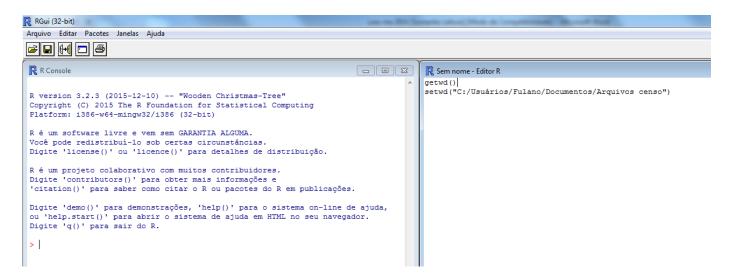
2.1 - Abertura de bases pequenas (IES, CURSO, DOCENTE e LOCAL DE OFERTA):

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

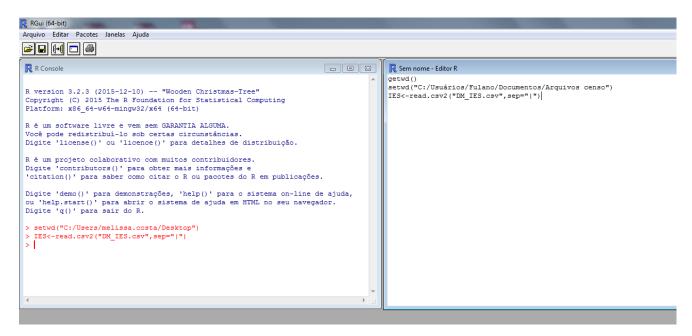
getwd() #Se o caminho não estiver correto, use a função a seguir para alterá-lo: setwd("C:/Usuários/Fulano/Documentos/Arquivos censo") #Exemplo de caminho. Lembre-se de usar barra simples ("/") ou barra invertida dupla ("\").



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

IES<-read.csv2("DM_IES.csv",sep="|")

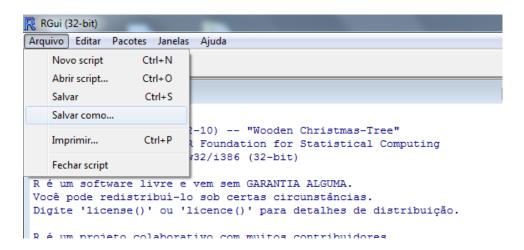
É importante que o *Pipe* (|) seja utilizado como delimitador.



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. IES[1:5,]

RGui (64-bit) Arquivo Editar Visualizar Misc Pacotes Janelas Ajuda - - X R Console R Sem nome - Edito getwd() setwd("C:/Usuá > IES<-read.csv2("DM IES.csv",sep="|" IES<-read.csv2 IES[1:5,] IES[1:5,] NO IES CO MANTENEDORA CO_IES FUNDACAO UNIVERSIDADE FEDERAL DE RONDONIA 1028 UNIVERSIDADE ESTADUAL DE MATO GROSSO DO SUL 729 1189 UNIVERSIDADE DO PLANALTO CATARINENSE 334 UNIVERSIDADE DO VALE DO PARAIBA UNIVERSIDADE FEDERAL DE PERNAMBUCO CO_CATEGORIA_ADMINISTRATIVA DS_CATEGORIA_ADMINISTRATIVA FEDERAL ESTADUAL 4 PARTICULAR EM SENTIDO ESTRITO PARTICULAR CONFESSIONAL FEDERAL CO_ORGANIZACAO_ACADEMICA NO_ORGANIZACAO_ACADEMICA CO_MUNICIPIO_IES UNIVERSIDADE INTUFESTDADE 5003702 3 UNIVERSIDADE 4209300 UNIVERSIDADE 3549904 UNIVERSIDADE 2611606 NO_MUNICIPIO_IES CO_UF SGL_UF NO_REGIAO IN_CAPITAL PORTO VELHO 11 RO Norte MS Centro_Oeste DOURADOS LAGES 42 4 SAO JOSE DO CAMPOS 35 SP Sudeste 0 RECIFE Nordeste 26 PE IN_ENTIDADE_BENEFICENTE QT_TEC_TOTAL QT_TEC_FUND_INCOMP_FEM

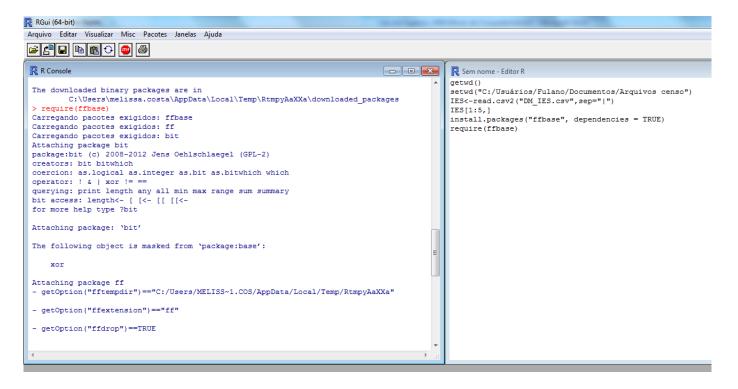
5) Sugerimos salvar o *script* criado para posterior utilização. Acesse a aba "Arquivo" → "Salvar como...".



2.2 - Abertura de bases maiores (ALUNO):

- 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 da base de ALUNO, que possui mais de 10 milhões de linhas (Brasil), faz-se necessário o uso de pacotes adicionais. Sugerimos o uso do pacote "ffbase" para trabalhar com essa base, tendo em vista que o mesmo faz uso do disco rígido em detrimento da memória RAM. Evidentemente, as bases de IES, CURSO, DOCENTE e LOCAL DE OFERTA 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"). Informações adicionais estão disponíveis na ajuda do pacote (https://cran.r-project.org/web/packages/ffbase/ffbase.pdf).
- 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.

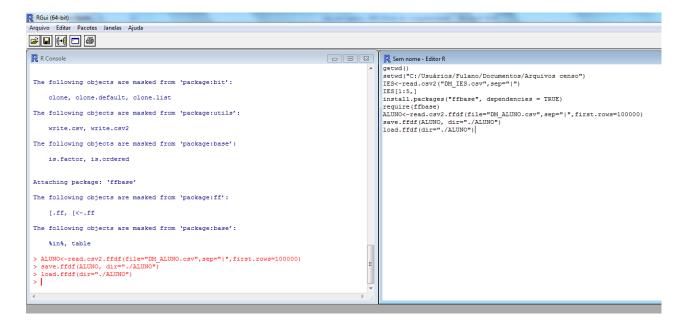
install.packages("ffbase", dependencies = TRUE) #Instala o pacote require(ffbase) #Carrega o pacote



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 ("\").

3) Caso você tenha aberto um novo script, será necessário certificar-se novamente de que o

4) Para abrir o arquivo "DM_ALUNO.csv", 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. ALUNO<-read.csv2.ffdf(file="DM_ALUNO.csv",sep="|",first.rows=1000000) É importante que o Pipe (|) seja utilizado como delimitador.

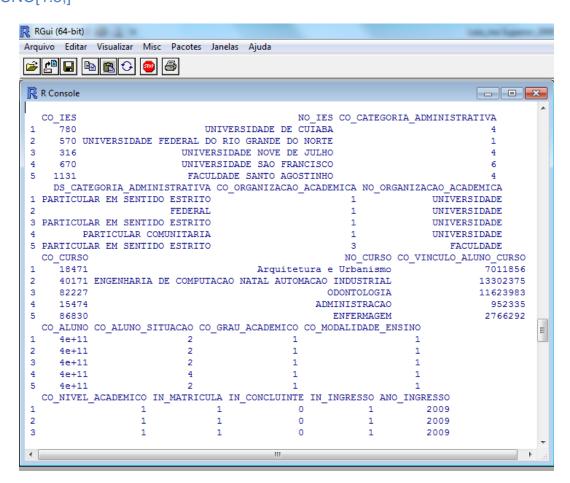


5) Bases grandes, quando carregadas pela primeira vez, podem demorar alguns minutos para concluir o processamento (dependendo das configurações do computador). Sugerimos salvar a base no formato 'ffdf' para que uma nova carga inicial não seja necessária numa próxima sessão.

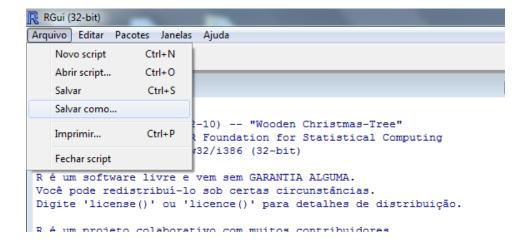
save.ffdf(ALUNO, dir="./ALUNO")

6) Dados salvos no formato 'ffdf' podem ser carregados em outro momento utilizando o comando abaixo. O carregamento da base com a função abaixo é praticamente instantâneo. load.ffdf(dir="./ALUNO")

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. ALUNO[1:5,]

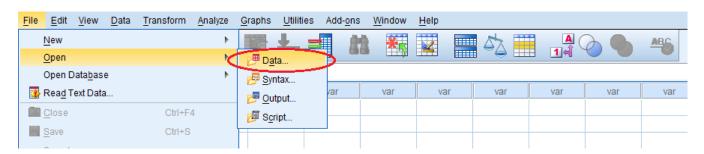


8) Sugerimos salvar o script criado para posterior utilização. Acesse a aba "Arquivo" → "Salvar como...".

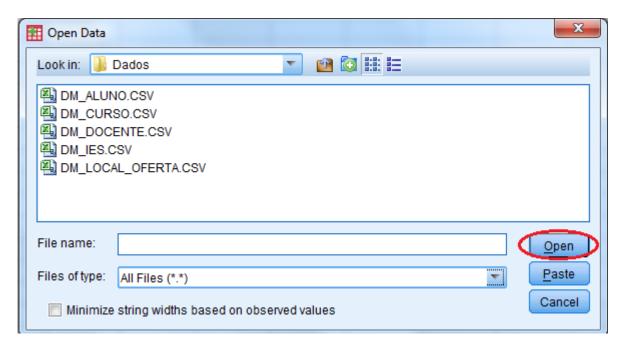


3 - ORIENTAÇÕES PARA ABRIR OS ARQUIVOS CSV NO SOFTWARE SPSS

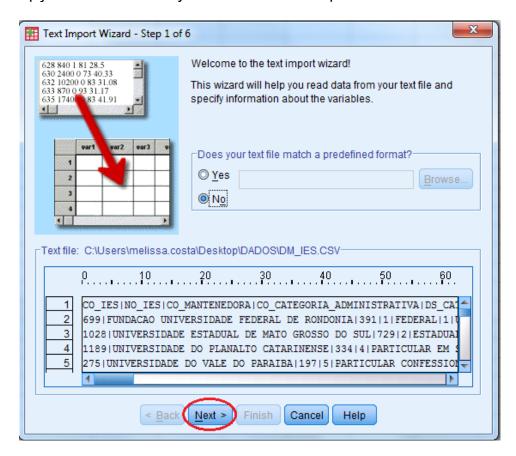
1) Abrir uma planilha no software SPSS e clicar na aba "File" → "Open" → "Data".



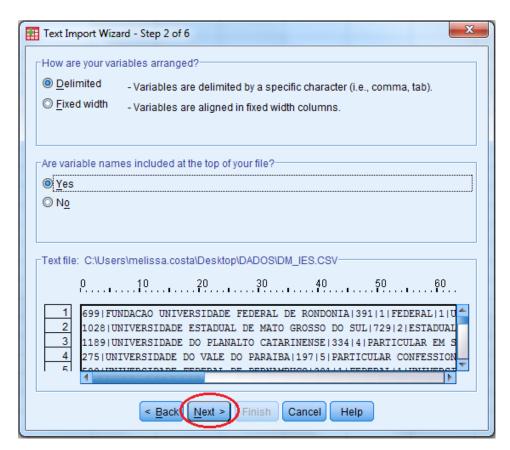
2) Em seguida, selecionar em "Files of type" a opção "All Files (*.*)", para que os arquivos em formato CSV possam ser visualizados. Selecionar o arquivo CSV desejado e clicar em "Open". No exemplo, os arquivos estão na pasta Dados.



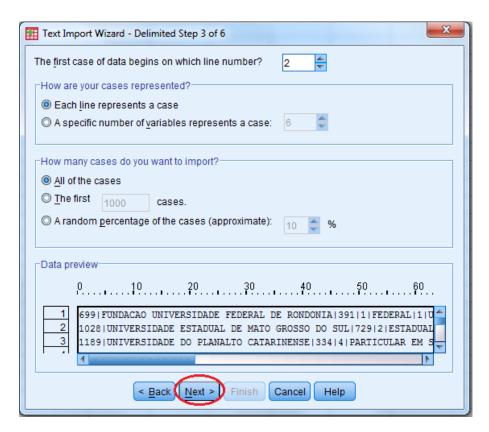
3) Manter a opção "No" em "Does your text file match a predefined format?" e clicar em "Next".



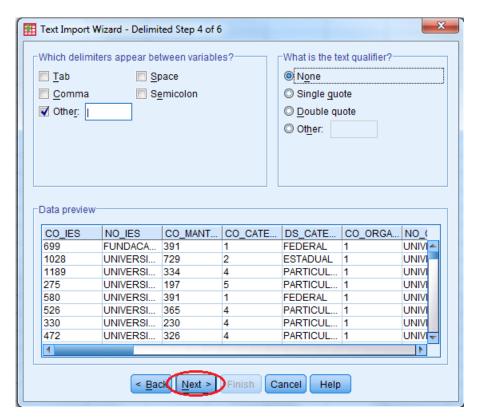
4) Manter a opção "Delimited" em "How are your variables arranged?" e alterar para a opção "Yes" em "Are variable names included at the top of your file?". Posteriormente, clicar em "Next".



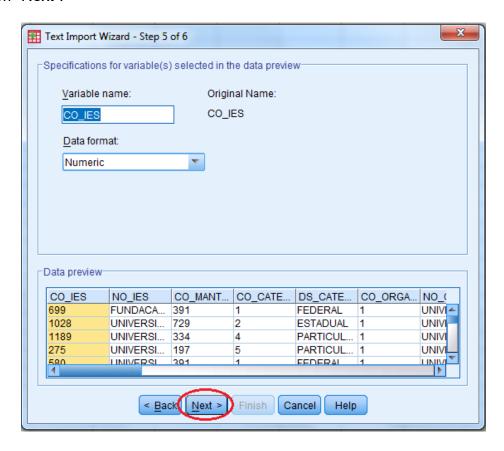
5) Manter a opção "2" em "The first case of data begins on which line number?". Manter, também, as opções "Each line represents a case" em "How are your cases represented?" e "All of the cases" em "How many cases do you want to import?". Posteriormente, clicar em "Next".



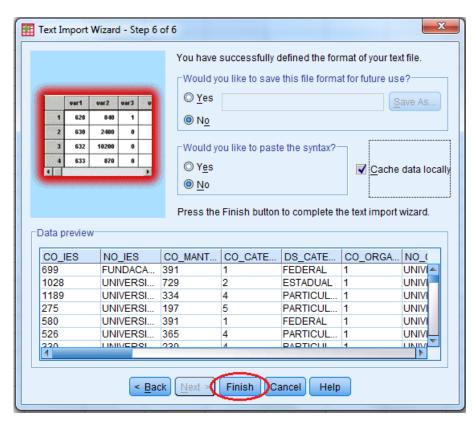
6) Desmarcar a opção "Space" e marcar a opção "Other" em "Which delimiters appear between variables?", utilizando o *Pipe* (|) como delimitador. Manter a opção "None" em "What is the text qualifer?". Posteriormente, clicar em "Next".



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

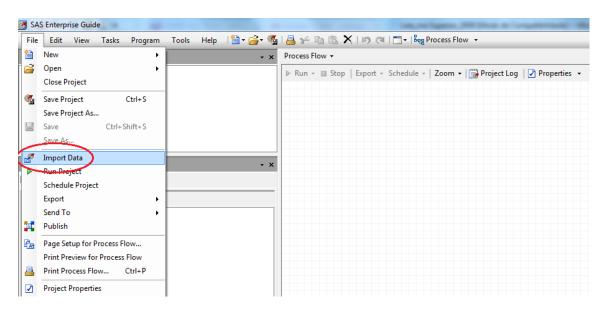


8) Na última etapa de "Text Import Wizard", é possível optar entre "Yes" e "No" em "Would you like to save this file format for future use?" e em "Would you like to paste the syntax?". Para finalizar o processo de importação de dados no *software* SPSS, basta clicar em "Finish".

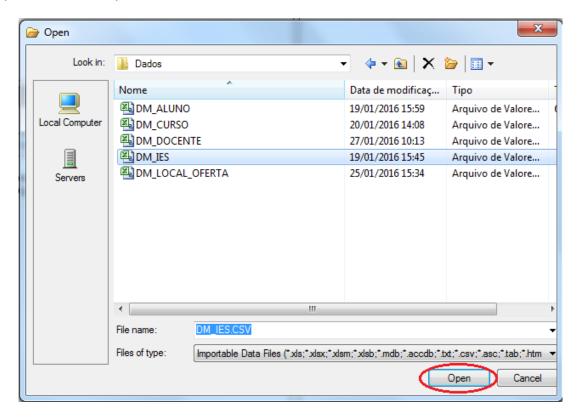


4 - ORIENTAÇÕES PARA ABRIR OS ARQUIVOS CSV NO SOFTWARE SAS

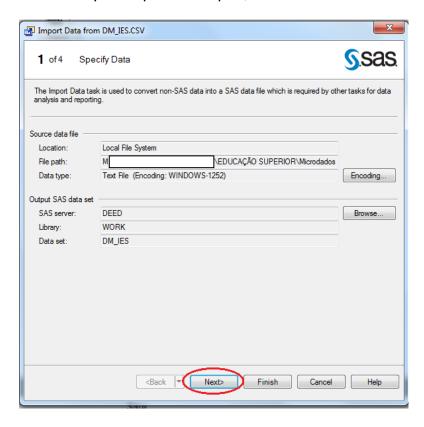
1) Abrir o software SAS e clicar na aba "File" → "Import Data".



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



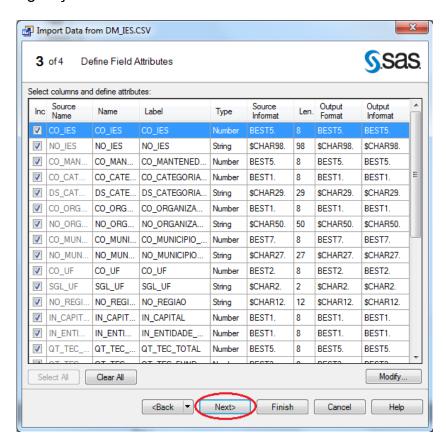
3) Na etapa seguinte, mantêm-se as opções marcadas em "Source data file" e pode-se optar pelo local onde será o Output da planilha. Após, clicar em "Next".



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".

