

Pacote R: esusnotifica

Introdução

O e-SUS Notifica (módulo notificação) é um sistema de informação em saúde voltado para a notificação de casos de síndrome gripal suspeitos de Covid-19 no território brasileiro. O objetivo do sistema é registrar os casos da doença que provocou uma pandemia, iniciada a partir de 2019, e inclui dados de identificação, dados clínicos e epidemiológicos e dados de encerramento da investigação.

Com a finalidade de realizar o tratamento da base de dados do e-SUS Notifica, foi criada uma biblioteca de funções que permitem o tratamento e padronização dos dados do sistema para análise de indicadores em saúde. O pacote tem ainda a função de auxiliar os gestores municipais, estaduais e nacionais na análise dos dados referentes à notificação de Covid-19.

Instalação do pacote

Para utilizar o pacote é necessário ter o R, a partir da versão XXXX. O pacote ainda não está disponível no *The Comprehensive R Archive Network (CRAN)*, portanto é necessário realizar a instalação diretamente do repositório de desenvolvimento no GitHub. Para isso, a instalação do pacote “devtools” é necessária, conforme o comando:

```
install.packages("devtools")
```

Após a instalação do pacote “devtools”, é necessário carregá-lo no R e realizar a instalação direta do GitHub do pacote “esusnotifica”.

```
library(devtools)  
install_github("ronaldoalves-ms/esusnotifica")
```

Funções do pacote

clean_esus

A função tem por objetivo realizar uma curadoria dos dados. A partir de uma extração de dados brutos da base do e-SUS Notifica é possível realizar a

padronização e transformação das variáveis de interesse para análises em saúde conforme o quadro 1. O objetivo é padronizar os dados de uma extração nominal do banco de dados do e-SUS Notifica.

Os dados do e-SUS Notifica devem ser importados para o ambiente R e depois tratados por meio da função, usando os comandos:

```
library(esusnotifica) #carrega pacote de limpeza dos dados
library(data.table) #carrega pacote para função importação de dados

dados_esus = fread(files = "arquivo_esus_notifica.csv") #cria data.frame
dados_esus_processados = clean_esus(dados_esus) #processa os dados
```

Dada a magnitude das notificações no e-SUS Notifica, sugerimos o uso da biblioteca `data.table` na importação, que pode comportar grandes volumes de dados. A função `fread` pertence a essa biblioteca.

Quadro 1- Transformações de variáveis do e-SUS Notifica (base nominal)

Variável	Nome da variável no arquivo	Transformação
NOME COMPLETO	nomeCompleto	A variável é convertida para a classe "character". São excluídos todos os acentos e todos os caracteres que não são alfabéticos, incluindo espaços em branco ao início ou final da variável. Todas as letras são convertidas em maiúsculas. Caso haja mais de um espaço em branco entre as palavras, a função reduz para um único espaço. Os campos vazios são substituídos por NA, indicando que o valor é ausente.
NOME DA MÃE	nomeMae	A variável é convertida para a classe "character". São excluídos todos os acentos e todos os caracteres que não são alfabéticos, incluindo espaços em branco ao início ou final da variável. Todas as letras são convertidas em maiúsculas. Caso haja mais de um espaço em branco entre as palavras, a função reduz para um único espaço. Os campos vazios são substituídos por NA, indicando que o valor é ausente.
NOME DO LOGRADOURO DE RESIDÊNCIA	logradouro	A variável é convertida para a classe "character". São excluídos todos os acentos e todos os caracteres que não são alfabéticos, incluindo espaços em branco ao início ou final da variável. Todas as letras são convertidas em maiúsculas. Caso haja mais de um espaço em branco entre as palavras, a função reduz para um único espaço. Os campos vazios são substituídos por NA, indicando que o valor é ausente.

Variável	Nome da variável no arquivo	Transformação
DATA DE NASCIMENTO	dataNascimento	A data é padronizada no formato “YYYY-MM-DD”. Os valores inferiores à 01 de janeiro de 2020 e superiores ao dia de tratamento dos dados são substituídos por NA, indicando que o valor é ausente.
DATA DO INICIO DOS SINTOMAS	dataInicioSintomas	A data é padronizada no formato “YYYY-MM-DD”. Os valores inferiores à 01 de janeiro de 2020 e superiores ao dia de tratamento dos dados são substituídos por NA, indicando que o valor é ausente. Para as datas que possuem NA, ocorre a substituição em ordem pelas variáveis: “dataTeste”, “dataTesteSorologico”, “dataNotificacao” e “dataRegistro”. Caso a data seja preenchida por alguma das variáveis consideradas, não há substituição pela próxima.
DATA DE NOTIFICAÇÃO	dataNotificacao	A data é padronizada no formato “YYYY-MM-DD”. Os valores inferiores à 01 de janeiro de 2020 e superiores ao dia de tratamento dos dados são substituídos por NA, indicando que o valor é ausente.
DATA DO TESTE	dataTeste	A data é padronizada no formato “YYYY-MM-DD”. Os valores inferiores à 01 de janeiro de 2020 e superiores ao dia de tratamento dos dados são substituídos por NA, indicando que o valor é ausente.
DATA DO TESTE SOROLÓGICO	dataTesteSorologico	A data é padronizada no formato “YYYY-MM-DD”. Os valores inferiores à 01 de janeiro de 2020 e superiores ao dia de tratamento dos dados são substituídos por NA, indicando que o valor é ausente.
DATA DE ENCERRAMENTO	dataEncerramento	A data é padronizada no formato “YYYY-MM-DD”. Os valores inferiores à 01 de janeiro de 2020 e superiores ao dia de tratamento dos dados são substituídos por NA, indicando que o valor é ausente.
DATA DA CRIAÇÃO DO REGISTRO	dataRegistro	A variável “_created_at”, que se refere à data de criação do registro no banco de dados é transformada em “dataRegistro”. A data é padronizada no formato “YYYY-MM-DD”. Os valores inferiores à 01 de janeiro de 2020 e superiores ao dia de tratamento dos dados são substituídos por NA, indicando que o valor é ausente.
DATA DA ATUALIZAÇÃO DO REGISTRO	dataAtualizacao	A variável “_updated_at”, que se refere à data de atualização do registro no banco de dados é transformada em “dataAtualizacao”. A data é padronizada no formato “YYYY-MM-DD”. Os valores inferiores à 01 de janeiro de 2020 e superiores ao dia de tratamento dos dados são substituídos por NA, indicando que o valor é ausente.
SEXO	sexo	A variável é convertida para a classe “character”. São removidos todos os acentos e todas as letras são transformadas em maiúsculas. Todos os campos com “MASCULINO” são substituídos por 1, os campos com “FEMININO” por 2 e os campos vazios ou com “INDEFINIDO” são substituídos por NA.

Variável	Nome da variável no arquivo	Transformação
RACA/COR	racaCor	A variável é convertida para a classe "character". São removidos todos os acentos e todas as letras são transformadas em maiúsculas. Os campos são substituídos conforme a sequência: "BRANCA" = 1, "PARDA" = 2, "PRETA" = 3, "AMARELA" = 4, "INDIGENA" = 5, "IGNORADO" = 9. Os campos vazios são substituídos por NA.
ETNIA	etnia	A variável é convertida para a classe "character". São excluídos todos os acentos e todos os caracteres que não são alfabéticos, incluindo espaços em branco ao início ou final da variável. Todas as letras são convertidas em maiúsculas. Os campos vazios são substituídos por NA.
COMUNIDADE TRADICIONAL (I)	contemComunidadeTradicional	A variável é convertida para a classe "character". São excluídos todos os acentos e todos os caracteres que não são alfabéticos, incluindo espaços em branco ao início ou final da variável. Todas as letras são convertidas em maiúsculas. Todos os campos que contem "SIM" são substituídos por 1 e os que contém "NÃO" são substituídos por 0. Os campos vazios são substituídos por NA.
COMUNIDADE TRADICIONAL (II)	comunidadeTradicional	A variável é convertida para a classe "character". São excluídos todos os acentos e todos os caracteres que não são alfabéticos, incluindo espaços em branco ao início ou final da variável. Todas as letras são convertidas em maiúsculas. Caso haja mais de um espaço em branco entre as palavras, a função reduz para um único espaço. Os 39 valores restantes são substituídos por caracteres numéricos, conforme o dicionário de dados [incluir dicionário de dados]. Os campos vazios são substituídos por NA, indicando que o valor é ausente.
PROFISSIONAL DA SAÚDE	profissionalSaude	A variável é convertida para a classe "character". São excluídos todos os acentos e todos os caracteres que não são alfabéticos, incluindo espaços em branco ao início ou final da variável. Todas as letras são convertidas em maiúsculas. Todos os campos que contem "SIM" são substituídos por 1 e os que contém "NÃO" são substituídos por 0. Os campos vazios são substituídos por NA.
PROFISSIONAL DA SEGURANÇA	profissionalSeguranca	A variável é convertida para a classe "character". São excluídos todos os acentos e todos os caracteres que não são alfabéticos, incluindo espaços em branco ao início ou final da variável. Todas as letras são convertidas em maiúsculas. Todos os campos que contem "SIM" são substituídos por 1 e os que contém "NÃO" são substituídos por 0. Os campos vazios são substituídos por NA.
ESTRANGEIRO	estrangeiro	A variável é convertida para a classe "character". São excluídos todos os acentos e todos os caracteres que não são alfabéticos, incluindo espaços em branco ao início ou final da variável. Todas as letras são convertidas

Variável	Nome da variável no arquivo	Transformação
		em maiúsculas. Todos os campos que contem "SIM" são substituídos por 1 e os que contém "NÃO" são substituídos por 0. Os campos vazios são substituídos por NA.
ESTADO DE RESIDENCIA	estado	A variável é convertida para a classe "character". São excluídos todos os acentos e todos os caracteres que não são alfabéticos, incluindo espaços em branco ao início ou final da variável. Todas as letras são convertidas em maiúsculas. Caso haja mais de um espaço em branco entre as palavras, a função reduz para um único espaço. Os 27 valores restantes são substituídos por caracteres numéricos, referentes ao código do estado no IBGE. Os campos vazios são substituídos por NA, indicando que o valor é ausente.
ESTADO DE NOTIFICAÇÃO	estadoNotificacao	A variável é convertida para a classe "character". São excluídos todos os acentos e todos os caracteres que não são alfabéticos, incluindo espaços em branco ao início ou final da variável. Todas as letras são convertidas em maiúsculas. Caso haja mais de um espaço em branco entre as palavras, a função reduz para um único espaço. Os 27 valores restantes são substituídos por caracteres numéricos, referentes ao código do estado no IBGE. Os campos vazios são substituídos por NA, indicando que o valor é ausente.
MUNICÍPIO DE RESIDENCIA	municipio	A variável é convertida para a classe "character". São excluídos todos os acentos e todos os caracteres que não são alfabéticos, incluindo espaços em branco ao início ou final da variável. Todas as letras são convertidas em maiúsculas. Caso haja mais de um espaço em branco entre as palavras, a função reduz para um único espaço. Os campos vazios são substituídos por NA, indicando que o valor é ausente.
MUNICÍPIO DE NOTIFICAÇÃO	municipioNotificacao	A variável é convertida para a classe "character". São excluídos todos os acentos e todos os caracteres que não são alfabéticos, incluindo espaços em branco ao início ou final da variável. Todas as letras são convertidas em maiúsculas. Caso haja mais de um espaço em branco entre as palavras, a função reduz para um único espaço. Os campos vazios são substituídos por NA, indicando que o valor é ausente.
CBO	cbo	A variável é convertida para a classe "character". São excluídos todos os acentos e todos os caracteres que não são alfanuméricos, incluindo espaços em branco ao final ou início do CBO. Todas as letras são convertidas em maiúsculas. Os campos vazios são substituídos por NA, indicando que o valor é ausente.
PAÍS DE ORIGEM	paisOrigem	A variável é convertida para a classe "character". São excluídos todos os acentos e todos os caracteres que não são alfabéticos, incluindo espaços

Variável	Nome da variável no arquivo	Transformação
		em branco ao início ou final da variável. Todas as letras são convertidas em maiúsculas. Caso haja mais de um espaço em branco entre as palavras, a função reduz para um único espaço. Os campos vazios são substituídos por NA, indicando que o valor é ausente.
RESULTADO DO TESTE (PCR/TESTE RAPIDO)	resultadoTeste	A variável é convertida para a classe "character". São removidos todos os acentos e todas as letras são transformadas em maiúsculas. Caso haja mais de um espaço em branco entre as palavras, a função reduz para um único espaço. Todos os campos com "POSITIVO" são substituídos por 1, os campos com "NEGATIVO" por 2, "INCONCLUSIVO OU INDETERMINADO" por "9" e vazios substituídos por NA.
RESULTADO DO TESTE SOROLÓGICO IGA	resultadoTesteSorologicoIga	A variável é convertida para a classe "character". São removidos todos os acentos e todas as letras são transformadas em maiúsculas. Caso haja mais de um espaço em branco entre as palavras, a função reduz para um único espaço. Todos os campos com "REAGENTE" são substituídos por 1, os campos com "NÃO REAGENTE" por 2, "INCONCLUSIVO OU INDETERMINADO" por "9" e vazios substituídos por NA.
RESULTADO DO TESTE SOROLÓGICO IGG	resultadoTesteSorologicoIgg	A variável é convertida para a classe "character". São removidos todos os acentos e todas as letras são transformadas em maiúsculas. Caso haja mais de um espaço em branco entre as palavras, a função reduz para um único espaço. Todos os campos com "REAGENTE" são substituídos por 1, os campos com "NÃO REAGENTE" por 2, "INCONCLUSIVO OU INDETERMINADO" por "9" e vazios substituídos por NA.
RESULTADO DO TESTE SOROLÓGICO IGM	resultadoTesteSorologicoIgm	A variável é convertida para a classe "character". São removidos todos os acentos e todas as letras são transformadas em maiúsculas. Caso haja mais de um espaço em branco entre as palavras, a função reduz para um único espaço. Todos os campos com "REAGENTE" são substituídos por 1, os campos com "NÃO REAGENTE" por 2, "INCONCLUSIVO OU INDETERMINADO" por "9" e vazios substituídos por NA.
RESULTADO DO TESTE SOROLÓGICO ANTICORPOS TOTAIS	resultadoTesteSorologicoTotais	A variável é convertida para a classe "character". São removidos todos os acentos e todas as letras são transformadas em maiúsculas. Caso haja mais de um espaço em branco entre as palavras, a função reduz para um único espaço. Todos os campos com "REAGENTE" são substituídos por 1, os campos com "NÃO REAGENTE" por 2, "INCONCLUSIVO OU INDETERMINADO" por "9" e vazios substituídos por NA.

Variável	Nome da variável no arquivo	Transformação
ESTADO DO TESTE (PCR/TESTE RAPIDO)	estadoTeste	<p>A variável é convertida para a classe "character". São removidos todos os acentos e todas as letras são transformadas em maiúsculas. Caso haja mais de um espaço em branco entre as palavras, a função reduz para um único espaço. Todos os campos com "SOLICITADO" são substituídos por 1, os campos com "COLETADO" por 2, os campos com "CONCLUIDO" por 3, "EXAME NÃO SOLICITADO" por 4 e vazios substituídos por NA.</p> <p>Após o primeiro tratamento da base, se o resultado do teste (resultadoTeste) estiver vazio, mas a data do teste (dataTeste) estiver preenchida, a função substitui o valor por 2, equivalente a "COLETADO". Caso o resultado permaneça vazio, é substituído por NA. Se o campo estiver preenchido, retornará os valores já definidos.</p>
TIPO DE TESTE	tipoTeste	<p>A variável é convertida para a classe "character". São removidos todos os acentos e todas as letras são transformadas em maiúsculas. Caso haja mais de um espaço em branco entre as palavras, a função reduz para um único espaço. Todos os campos com "RT PCR" são substituídos por 1, os campos com "TESTE RAPIDO ANTICORPO" por 2, os campos com "TESTE RAPIDO ANTIGENO" por 3. Os valores "TESTE SOROLOGICO", "ENZIMAIMUNOENSAIO ELISA IGM", "ENZIMAIMUNOENSAIO ELISA", "IMUNOENSAIO POR ELETROQUIMIOLUMINESCENCIA ECLIA IGG", "IMUNOENSAIO POR ELETROQUIMIOLUMINESCENCIA ECLIA" e "QUIMIOLUMINESCENCIA CLIA" são substituídos por 4.</p> <p>Os campos vazios são substituídos por NA.</p>
CLASSIFICAÇÃO FINAL	classificacaoFinal	<p>Na primeira parte da curadoria, a variável é convertida para a classe "character". São removidos todos os acentos e todas as letras são transformadas em maiúsculas. Caso haja mais de um espaço em branco entre as palavras, a função reduz para um único espaço. Os campos com "CONFIRMACAO LABORATORIAL" ou "CONFIRMADO LABORATORIAL" são substituídos por 1, os campos com "CONFIRMACAO CLINICO EPIDEMIOLOGICO" ou "CONFIRMADO CLINICO EPIDEMIOLOGICO" são substituídos por 2, os campos com "CONFIRMACAO CLINICO IMAGEM" ou "CONFIRMADO CLINICO IMAGEM" são substituídos por 3, "CONFIRMACAO CLINICO", "CONFIRMADO CLINICO" ou "CONFIRMADO POR CRITERIO CLINICO" são substituídos por 4, "SINDROME GRIPAL NAO</p>

Variável	Nome da variável no arquivo	Transformação
		<p>ESPECIFICADA" são substituídos por 5 e "DESCARTADO" é substituído por 6. Os campos vazios são substituídos por NA.</p> <p>Na segunda parte da curadoria, os campos que tem as variáveis "resultadoTeste" ou "resultadoTesteSorologicoIgA", ou "resultadoTesteSorologicoIgG" ou "resultadoTesteSorologicoIgM" ou "resultadoTesteSorologicoTotais" igual a 1, recebem o valor 1, equivalente a "CONFIRMADO LABORATORIAL". Caso as variáveis "resultadoTeste" e "resultadoTesteSorologicoIgA", e "resultadoTesteSorologicoIgG" e "resultadoTesteSorologicoIgM" e "resultadoTesteSorologicoTotais" sejam diferentes de 1, mas a variável "classificacaoFinal" seja igual a 1, o campo é substituído por 2. Se o campo não atender esses critérios, ele será mantido conforme o primeiro tratamento.</p>
EVOLUÇÃO CASO	evolucaoCaso	A variável é convertida para a classe "character". São removidos todos os acentos e todas as letras são transformadas em maiúsculas. Caso haja mais de um espaço em branco entre as palavras, a função reduz para um único espaço. Todos os campos com "CURA" são substituídos por 1, os campos com "OBITO" por 2, os campos com "EM TRATAMENTO DOMICILIAR" por 3, os campos com "INTERNADO" por 4, os campos com "INTERNADO EM UTI" por 5, os campos com "IGNORADO" por 6, os campos com "CANCELADO" por 9 e os vazios substituídos por NA.
SINTOMAS	sintomas	A variável é convertida para a classe "character". São removidos todos os acentos e todas as letras são transformadas em maiúsculas. Caso haja mais de um espaço em branco entre as palavras, a função reduz para um único espaço. Todos os campos vazios são substituídos por NA.
CONDICOES	condicoes	A variável é convertida para a classe "character". São removidos todos os acentos e todas as letras são transformadas em maiúsculas. Caso haja mais de um espaço em branco entre as palavras, a função reduz para um único espaço. Todos os campos vazios são substituídos por NA.
IDADE	idade	A variável idade é substituída por NA e então recalculada com base no intervalo entre a data de nascimento e a data de início de sintomas, em anos. Todos os campos que possuem idade igual a 0, ou seja, em que não há diferença entre a data de nascimento e a data de início de sintomas, recebe o valor NA. Após isso, os dados são transformados em inteiros.

Variável	Nome da variável no arquivo	Transformação
FAIXA ETARIA	faixaEtaria	A variável é criada com base na idade calculada, incluindo um intervalo de menores de 1 ano, entre 2 e 5 anos e após isso intervalos de 5 anos, até a idade de 100 anos.
TESTES	testes	Esse campo incorpora todos os testes disponíveis anteriormente em outras variáveis e inclui todos os atributos que descrevem um teste. O campo inclui os atributos: “tipoTeste”, “codigoEstadoTeste”, “estadoTeste”, “codigoTipoTeste”, “fabricanteTeste”, “loteTeste”, “resultadoTeste”, “codigoResultadoTeste”, “iso”, “dataColetaTeste”. Todos os valores são criados para cada teste realizado, contando até no máximo 3 testes. São criados os campos “numeroTotalTestes” para calcular quantos testes foram registrados em cada notificação e “resultadosPositivo”, que conta o número de testes que tiveram valores positivos.

Após o tratamento dos dados, o *data.frame* passa por um processo de transformação das variáveis da classe “*factor*”, que são transformadas para a classe mais adequada por meio da função “*droplevels*”.

`clean_esus_open`

A função tem por objetivo realizar uma curadoria dos dados abertos, ou seja, que não possuam dados pessoais. A partir de uma extração de dados brutos da base do e-SUS Notifica, como do portal OpenDataSUS, é possível realizar a padronização e transformação das variáveis de interesse para análises em saúde.

Os dados do e-SUS Notifica devem ser importados para o ambiente R e depois tratados por meio da função, usando os comandos:

```
library(esusnotifica) #carrega pacote de limpeza dos dados
library(data.table) #carrega pacote para função importação de dados

dados_esus = fread(files = "open_esus_notifica.csv") #cria um data.frame
dados_esus_processados = clean_esus(dados_esus) #processa os dados
```

Quadro 2- Transformações de variáveis do e-SUS Notifica (base aberta)

Variável	Nome da variável no arquivo	Transformação
DATA DO INICIO DOS SINTOMAS	dataInicioSintomas	<p>A data é padronizada no formato “YYYY-MM-DD”. Os valores inferiores à 01 de janeiro de 2020 e superiores ao dia de tratamento dos dados são substituídos por NA, indicando que o valor é ausente.</p> <p>Para as datas que possuem NA, ocorre a substituição em ordem pelas variáveis: “dataTeste”, “dataTesteSorologico”, “dataNotificacao” e “dataRegistro”. Caso a data seja preenchida por alguma das variáveis consideradas, não há substituição pela próxima.</p>
DATA DE NOTIFICAÇÃO	dataNotificacao	A data é padronizada no formato “YYYY-MM-DD”. Os valores inferiores à 01 de janeiro de 2020 e superiores ao dia de tratamento dos dados são substituídos por NA, indicando que o valor é ausente.
DATA DO TESTE	dataTeste	A data é padronizada no formato “YYYY-MM-DD”. Os valores inferiores à 01 de janeiro de 2020 e superiores ao dia de tratamento dos dados são substituídos por NA, indicando que o valor é ausente.
DATA DO TESTE SOROLÓGICO	dataTesteSorologico	A data é padronizada no formato “YYYY-MM-DD”. Os valores inferiores à 01 de janeiro de 2020 e superiores ao dia de tratamento dos dados são substituídos por NA, indicando que o valor é ausente.
DATA DE ENCERRAMENTO	dataEncerramento	A data é padronizada no formato “YYYY-MM-DD”. Os valores inferiores à 01 de janeiro de 2020 e superiores ao dia de tratamento dos dados são substituídos por NA, indicando que o valor é ausente.
DATA DA CRIAÇÃO DO REGISTRO	dataRegistro	A data é padronizada no formato “YYYY-MM-DD”. Os valores inferiores à 01 de janeiro de 2020 e superiores ao dia de tratamento dos dados são substituídos por NA, indicando que o valor é ausente.
DATA DA ATUALIZAÇÃO DO REGISTRO	dataAtualizacao	A data é padronizada no formato “YYYY-MM-DD”. Os valores inferiores à 01 de janeiro de 2020 e superiores ao dia de tratamento dos dados são substituídos por NA, indicando que o valor é ausente.
SEXO	sexo	A variável é convertida para a classe “character”. São removidos todos os acentos e todas as letras são transformadas em maiúsculas. Todos os campos com “MASCULINO” são substituídos por 1, os campos com “FEMININO” por 2 e os campos vazios ou com “INDEFINIDO” são substituídos por NA.
RACA/COR	racaCor	A variável é convertida para a classe “character”. São removidos todos os acentos e todas as letras são transformadas em maiúsculas. Os campos são substituídos conforme a sequência: “BRANCA” = 1, “PARDA” = 2, “PRETA” = 3, “AMARELA” = 4, “INDIGENA” = 5, “IGNORADO” = 9. Os campos vazios são substituídos por NA.
IDADE	idade	Os dados são transformados em valores numéricos inteiros.

Variável	Nome da variável no arquivo	Transformação
PROFISSIONAL DA SAÚDE	profissionalSaude	São excluídos todos os acentos e todos os caracteres que não são alfabéticos, incluindo espaços em branco ao início ou final da variável. Todas as letras são convertidas em maiúsculas. Todos os campos que contem "SIM" são substituídos por 1 e os que contém "NÃO" são substituídos por 0. Os campos vazios são substituídos por NA.
PROFISSIONAL DA SEGURANÇA	profissionalSeguranca	São excluídos todos os acentos e todos os caracteres que não são alfabéticos, incluindo espaços em branco ao início ou final da variável. Todas as letras são convertidas em maiúsculas. Todos os campos que contem "SIM" são substituídos por 1 e os que contém "NÃO" são substituídos por 0. Os campos vazios são substituídos por NA.
ESTRANGEIRO	estrangeiro	São excluídos todos os acentos e todos os caracteres que não são alfabéticos, incluindo espaços em branco ao início ou final da variável. Todas as letras são convertidas em maiúsculas. Todos os campos que contem "SIM" são substituídos por 1 e os que contém "NÃO" são substituídos por 0. Os campos vazios são substituídos por NA.
ESTADO DE RESIDENCIA	estado	A variável é convertida para a classe "character". São excluídos todos os acentos e todos os caracteres que não são alfabéticos, incluindo espaços em branco ao início ou final da variável. Todas as letras são convertidas em maiúsculas. Caso haja mais de um espaço em branco entre as palavras, a função reduz para um único espaço. Os 27 valores restantes são substituídos por caracteres numéricos, referentes ao código do estado no IBGE. Os campos vazios são substituídos por NA, indicando que o valor é ausente.
ESTADO DE NOTIFICAÇÃO	estadoNotificacao	A variável é convertida para a classe "character". São excluídos todos os acentos e todos os caracteres que não são alfabéticos, incluindo espaços em branco ao início ou final da variável. Todas as letras são convertidas em maiúsculas. Caso haja mais de um espaço em branco entre as palavras, a função reduz para um único espaço. Os 27 valores restantes são substituídos por caracteres numéricos, referentes ao código do estado no IBGE. Os campos vazios são substituídos por NA, indicando que o valor é ausente.
MUNICÍPIO DE RESIDENCIA	municipio	A variável é convertida para a classe "character". São excluídos todos os acentos e todos os caracteres que não são alfabéticos, incluindo espaços em branco ao início ou final da variável. Todas as letras são convertidas em maiúsculas. Caso haja mais de um espaço em branco entre as palavras, a função reduz para um único espaço. Os campos vazios são substituídos por NA, indicando que o valor é ausente.

Variável	Nome da variável no arquivo	Transformação
MUNICÍPIO DE NOTIFICAÇÃO	municipioNotificacao	A variável é convertida para a classe "character". São excluídos todos os acentos e todos os caracteres que não são alfabéticos, incluindo espaços em branco ao início ou final da variável. Todas as letras são convertidas em maiúsculas. Caso haja mais de um espaço em branco entre as palavras, a função reduz para um único espaço. Os campos vazios são substituídos por NA, indicando que o valor é ausente.
CBO	cbo	A variável é convertida para a classe "character". São excluídos todos os acentos e todos os caracteres que não são alfanuméricos, incluindo espaços em branco ao final ou início do CBO. Todas as letras são convertidas em maiúsculas. Os campos vazios são substituídos por NA, indicando que o valor é ausente.
PAÍS DE ORIGEM	paisOrigem	A variável é convertida para a classe "character". São excluídos todos os acentos e todos os caracteres que não são alfabéticos, incluindo espaços em branco ao início ou final da variável. Todas as letras são convertidas em maiúsculas. Caso haja mais de um espaço em branco entre as palavras, a função reduz para um único espaço. Os campos vazios são substituídos por NA, indicando que o valor é ausente.
RESULTADO DO TESTE	resultadoTeste	A variável é convertida para a classe "character". São removidos todos os acentos e todas as letras são transformadas em maiúsculas. Caso haja mais de um espaço em branco entre as palavras, a função reduz para um único espaço. Todos os campos com "POSITIVO" são substituídos por 1, os campos com "NEGATIVO" por 2, "INCONCLUSIVO OU INDETERMINADO" por "9" e vazios substituídos por NA.
RESULTADO DO TESTE SOROLÓGICO IGA	resultadoTesteSorologicolga	A variável é convertida para a classe "character". São removidos todos os acentos e todas as letras são transformadas em maiúsculas. Caso haja mais de um espaço em branco entre as palavras, a função reduz para um único espaço. Todos os campos com "REAGENTE" são substituídos por 1, os campos com "NÃO REAGENTE" por 2, "INCONCLUSIVO OU INDETERMINADO" por "9" e vazios substituídos por NA.
RESULTADO DO TESTE SOROLÓGICO IGG	resultadoTesteSorologicolgG	A variável é convertida para a classe "character". São removidos todos os acentos e todas as letras são transformadas em maiúsculas. Caso haja mais de um espaço em branco entre as palavras, a função reduz para um único espaço. Todos os campos com "REAGENTE" são substituídos por 1, os campos com "NÃO REAGENTE" por 2, "INCONCLUSIVO OU INDETERMINADO" por "9" e vazios substituídos por NA.

Variável	Nome da variável no arquivo	Transformação
RESULTADO DO TESTE SOROLÓGICO IGM	resultadoTesteSorologicoIgm	A variável é convertida para a classe "character". São removidos todos os acentos e todas as letras são transformadas em maiúsculas. Caso haja mais de um espaço em branco entre as palavras, a função reduz para um único espaço. Todos os campos com "REAGENTE" são substituídos por 1, os campos com "NÃO REAGENTE" por 2, "INCONCLUSIVO OU INDETERMINADO" por "9" e vazios substituídos por NA.
RESULTADO DO TESTE SOROLÓGICO TOTAIS	resultadoTesteSorologicoTotais	A variável é convertida para a classe "character". São removidos todos os acentos e todas as letras são transformadas em maiúsculas. Caso haja mais de um espaço em branco entre as palavras, a função reduz para um único espaço. Todos os campos com "REAGENTE" são substituídos por 1, os campos com "NÃO REAGENTE" por 2, "INCONCLUSIVO OU INDETERMINADO" por "9" e vazios substituídos por NA.
ESTADO DO TESTE	estadoTeste	<p>A variável é convertida para a classe "character". São removidos todos os acentos e todas as letras são transformadas em maiúsculas. Caso haja mais de um espaço em branco entre as palavras, a função reduz para um único espaço. Todos os campos com "SOLICITADO" são substituídos por 1, os campos com "COLETADO" por 2, os campos com "CONCLUÍDO" por 3, "EXAME NÃO SOLICITADO" por 4 e vazios substituídos por NA.</p> <p>Após o primeiro tratamento da base, se o resultado do teste (resultadoTeste) estiver vazio, mas a data do teste (dataTeste) estiver preenchida, a função substitui o valor por 2, equivalente a "COLETADO". Caso o resultado permaneça vazio, é substituído por NA. Se o campo estiver preenchido, retornará os valores já definidos.</p>
TIPO DE TESTE	tipoTeste	A variável é convertida para a classe "character". São removidos todos os acentos e todas as letras são transformadas em maiúsculas. Caso haja mais de um espaço em branco entre as palavras, a função reduz para um único espaço. Todos os campos com "RT PCR" são substituídos por 1, os campos com "TESTE RÁPIDO ANTICORPO" por 2, os campos com "TESTE RÁPIDO ANTIGENO" por 3. Os valores "TESTE SOROLÓGICO", "ENZIMAIMUNOENSAIO ELISA IGM", "ENZIMAIMUNOENSAIO ELISA", "IMUNOENSAIO POR ELETROQUIMIOLUMINESCENCIA ECLIA IGG", "IMUNOENSAIO POR ELETROQUIMIOLUMINESCENCIA ECLIA" e "QUIMIOLUMINESCENCIA CLIA" são substituídos por 4.

Variável	Nome da variável no arquivo	Transformação
		Os campos vazios são substituídos por NA.
CLASSIFICAÇÃO FINAL	classificacaoFinal	<p>Na primeira parte da curadoria, a variável é convertida para a classe “character”. São removidos todos os acentos e todas as letras são transformadas em maiúsculas. Caso haja mais de um espaço em branco entre as palavras, a função reduz para um único espaço. Os campos com "CONFIRMACAO LABORATORIAL" ou "CONFIRMADO LABORATORIAL" são substituídos por 1, os campos com "CONFIRMACAO CLINICO EPIDEMIOLOGICO" ou "CONFIRMADO CLINICO EPIDEMIOLOGICO" são substituídos por 2, os campos com "CONFIRMACAO CLINICO IMAGEM" ou "CONFIRMADO CLINICO IMAGEM" são substituídos por 3, "CONFIRMACAO CLINICO", "CONFIRMADO CLINICO" ou "CONFIRMADO POR CRITERIO CLINICO" são substituídos por 4, "SINDROME GRIPAL NAO ESPECIFICADA" são substituídos por 5 e “DESCARTADO” é substituído por 6. Os campos vazios são substituídos por NA.</p> <p>Na segunda parte da curadoria, os campos que tem as variáveis “resultadoTeste” ou “resultadoTesteSorologicoIgA”, ou “resultadoTesteSorologicoIgG” ou “resultadoTesteSorologicoIgM” ou “resultadoTesteSorologicoTotais” igual a 1, recebem o valor 1, equivalente a “CONFIRMADO LABORATORIAL”. Caso as variáveis “resultadoTeste” e “resultadoTesteSorologicoIgA”, e “resultadoTesteSorologicoIgG” e “resultadoTesteSorologicoIgM” e “resultadoTesteSorologicoTotais” sejam diferentes de 1, mas a variável “classificacaoFinal” seja igual a 1, o campo é substituído por 2. Se o campo não atender esses critérios, ele será mantido conforme o primeiro tratamento.</p>
EVOLUCAO CASO	evolucaoCaso	A variável é convertida para a classe “character”. São removidos todos os acentos e todas as letras são transformadas em maiúsculas. Caso haja mais de um espaço em branco entre as palavras, a função reduz para um único espaço. Todos os campos com “CURA” são substituídos por 1, os campos com “OBITO” por 2, os campos com “EM TRATAMENTO DOMICILIAR” por 3, os campos com “INTERNADO” por 4, os campos com “INTERNADO EM UTI” por 5, os campos com “IGNORADO” por 6, os campos com “CANCELADO” por 9 e os vazios substituídos por NA.
SINTOMAS	sintomas	A variável é convertida para a classe “character”. São removidos todos os acentos e todas as letras são transformadas em maiúsculas. Caso

Variável	Nome da variável no arquivo	Transformação
		haja mais de um espaço em branco entre as palavras, a função reduz para um único espaço. Todos os campos vazios são substituídos por NA.
CONDICOES	condicoes	A variável é convertida para a classe "character". São removidos todos os acentos e todas as letras são transformadas em maiúsculas. Caso haja mais de um espaço em branco entre as palavras, a função reduz para um único espaço. Todos os campos vazios são substituídos por NA.
ANO/SEMANA EPIDEMIOLOGICA	anoEpiSintomas semEpiSintomas	Usa a variável dataInicioSintomas para criar duas variáveis referentes ao ano epidemiológico e à semana epidemiológica em que os casos registraram os primeiros sintomas
FAIXA-ETARIA	faixaEtaria	A variável é criada com base na idade calculada, incluindo um intervalo de menores de 1 ano, entre 2 e 5 anos e após isso intervalos de 5 anos, até a idade de 100 anos.

Após o tratamento dos dados, o *data.frame* passa por um processo de transformação das variáveis da classe "factor", que são transformadas para a classe mais adequada por meio da função "droplevels".