

Estatística descritiva para *Data Science*

0.2 - Aula Prática: Obtendo e tratando a base de dados

Profa. Dra. Amanda Buosi Gazon Milani

2024-07-19

Conjunto de dados - Nascidos Vivos 2024 (DataSUS)

O conjunto de dados que será utilizado nesta disciplina foi obtido no site do OpenDataSUS e consiste da base de informações sobre nascidos vivos 2024 (parcial). Esse conjunto de dados foi obtido pelo link <https://opendatasus.saude.gov.br/dataset/sistema-de-informacao-sobre-nascidos-vivos-sinasc> (acesso e download em 08/07/2024), onde também encontramos seu respectivo dicionário de dados. O arquivo DNOPEN24.csv está disponível na plataforma do curso.

Inicialmente vamos importar (carregar) os dados no R, utilizando os códigos a seguir.

```
setwd <- "C:\\Users\\AmandaBGM\\Google Drive\\UFG\\Especialização_FEN_IME\\2024\\Scripts"
df.nv24 <- read.csv("DNOPEN24.csv", sep = ";")
head(df.nv24)
```

##	contador	ORIGEM	CODESTAB	CODMUNNASC	LOCNASC	IDADEMAE	ESTCIVMAE	ESMAE	
## 1	1	1	2516500	110001	1	24	1	3	
## 2	2	1	2516500	110001	1	29	2	5	
## 3	3	1	2494299	110002	1	20	5	4	
## 4	4	1	2494299	110002	1	40	1	5	
## 5	5	1	2494299	110002	1	27	2	4	
## 6	6	1	2494299	110002	1	19	1	4	
##	CODOCUPMAE	QTDFILVIVO	QTDFILMORT	CODMUNRES	GESTACAO	GRAVIDEZ	PARTO	CONSULTAS	
## 1	999992	1	0	110001	5	1	2	2	
## 2	999992	0	0	110001	5	1	2	4	
## 3	999992	0	0	110002	5	1	1	4	
## 4	252305	4	1	110110	5	1	2	4	
## 5	999992	3	0	110002	4	1	1	4	
## 6	521115	0	0	110002	5	1	2	4	
##	DTNASC	HORANASC	SEXO	APGAR1	APGAR5	RACACOR	PESO	IDANOMAL	DTCADASTRO
## 1	14022024	845	1	8	9	4	3120	2	5032024
## 2	17042024	850	1	8	9	4	3564	2	9052024
## 3	1012024	516	1	8	9	1	3240	2	3012024
## 4	1012024	2123	1	9	9	4	3960	2	3012024
## 5	1012024	1937	1	8	9	4	3610	2	3012024
## 6	1012024	2225	1	9	9	4	3724	2	3012024
##	CODANOMAL	NUMEROLOTE	VERSAOSIST	DTRECEBIM	DIFDATA	OPORT_DN	DTRECORIGA		
## 1		20240003	3.2.50	5032024	20	20	5032024		
## 2		20240005	3.2.50	9052024	22	22	9052024		
## 3		20240001	3.2.50	9012024	8	8	9012024		
## 4		20240001	3.2.50	9012024	8	8	9012024		
## 5		20240001	3.2.50	9012024	8	8	9012024		

## 6	20240001	3.2.50	9012024	8	8	9012024	
##	NATURALMAE	CODMUNNATU	CODUFNATU	ESCMAC2010	SERIESCMAE	DTNASCMAE	RACACORMAE
## 1	811	110028	11	2	5	11012000	4
## 2	811	110001	11	5	NA	22111994	4
## 3	811	110002	11	3	3	24062003	1
## 4	812	120040	12	5	NA	2061983	4
## 5	811	110002	11	3	3	15051996	4
## 6	811	110020	11	3	3	9122004	4
##	QTDGESTANT	QTDPARTNOR	QTDPARTCES	IDADEPAI	DTULTMENST	SEMAGESTAC	TPMETESTIM
## 1	1	0	1	NA	17052023	38	8
## 2	0	0	0	41	17072023	39	8
## 3	0	0	0	NA	5042023	38	8
## 4	5	1	3	NA	NA	38	1
## 5	3	3	0	NA	22042023	36	8
## 6	0	0	0	NA	13032023	41	8
##	CONSPRENAT	MESPRENAT	TPAPRESENT	STTRABPART	STCESPARTO	TPNASCASSI	TPFUNCRESP
## 1	2	3	1	2	2	1	2
## 2	8	5	1	2	1	1	2
## 3	10	2	1	2	3	2	2
## 4	7	4	1	2	2	1	2
## 5	10	1	1	2	3	2	2
## 6	8	3	1	2	2	1	2
##	TPDOCRESP	DTDECLARAC	ESCMACAGR1	STDNEPIDEM	STDNNNOVA	CODPAISRES	TPROBSON
## 1	3	14022024	3	0	1	1	5
## 2	3	17042024	8	0	1	1	2
## 3	3	1012024	6	1	1	1	1
## 4	3	1012024	8	1	1	1	5
## 5	3	1012024	6	1	1	1	10
## 6	3	1012024	6	1	1	1	1
##	PARIDADE	KOTELCHUCK					
## 1	1	2					
## 2	0	2					
## 3	0	5					
## 4	1	2					
## 5	1	5					
## 6	0	5					

```
dim(df.nv24)
```

```
## [1] 779927      62
```

```
# Verificando se é um dataframe
is.data.frame(df.nv24)
```

```
## [1] TRUE
```

```
# Caso não seja (apagar # para rodar:)
#df.nv24 <- as.data.frame(df.nv24)
```

Note que o conjunto de dados contém 779.927 observações e 62 variáveis. A descrição completa deste conjunto de dados pode ser obtida no dicionário no site do OpenDataSUS e a seguir encontramos o detalhamento de parte das variáveis, que utilizaremos na disciplina. São elas:

Tabela 1: Variáveis do conjunto de dados

Nome da variável	Descrição	Níveis
LOCNASC	Local de ocorrência do nascimento	1: Hospital 2: Outro Estab Saúde 3: Domicílio 4: Aldeia Indígena 5: Outros 9: Ignorado
IDADEMAE	Idade da mãe em anos	
ESTCIVMAE	Estado civil da mãe	1: Solteira 2: Casada 3: Viúva 4: Separada judicialmente 5: União consensual 9: Ignorado
QTDFILVIVO	Número de filhos vivos	
QTDFILMORT	Número de filhos mortos	
GESTACAO	Semanas de gestação, conforme classificação	1: Menos de 22 semanas 2: 22 a 27 semanas 3: 28 a 31 semanas 4: 32 a 36 semanas 5: 37 a 41 semanas 6: 42 semanas e mais 9: Ignorado
GRAVIDEZ	Tipo de gravidez, conforme classificação	1: Única 2: Dupla 3: Tripla e mais 9: Ignorado
PARTO	Tipo de parto	1: Vaginal 2: Cesáreo 9: Ignorado
CONSULTAS	Número de consultas pré-natal, conforme classificação	1: Nenhuma 2: de 1 a 3 3: de 4 a 6 4: 7 e mais 9: Ignorado
DTNASC	Data do nascimento, no formato ddmmaaaa	
SEXO	Sexo	0: Ignorado, não informado 1: Masculino 2: Feminino
APGAR1	Apgar no primeiro minuto, 00 a 10	
APGAR5	Apgar no quinto minuto, 00 a 10	
RACACOR	Raça/cor	1: Branca 2: Preta 3: Amarela 4: Parda 5: Indígena
PESO	Peso ao nascer, em gramas	
IDANOMAL	Anomalia congênita	1: Sim 2: Não 9: Ignorado

Continua na próxima página

Tabela 1 – Continuação da tabela

Nome da variável	Descrição	Níveis
CODUFNATU	Código da UF de naturalidade da mãe	Tabela do IBGE*
ESMAE2010	Escolaridade da mãe - 2010	0: Sem escolaridade 1: Fundamental I (1ª a 4ª série) 2: Fundamental II (5ª a 8ª série) 3: Médio (antigo 2º Grau) 4: Superior incompleto 5: Superior completo 9: Ignorado
RACACORMAE	Raça/cor da mãe	1: Branca 2: Preta 3: Amarela 4: Parda 5: Indígena
QTDGESTANT	Número de gestações anteriores	
QTDPARTNOR	Número de partos vaginais	
QTDPARTCES	Número de partos cesáreos	
IDADEPAI	Idade do pai, em anos	
SEMANAGESTAC	Número de semanas de gestação	
CONSPRENAT	Número de consultas pré-natal	
STTRABPART	Trabalho de parto induzido?	1: Sim 2: Não 3: Não se aplica 9: Ignorado
STCESPARTO	Cesárea ocorreu antes do trabalho de parto iniciar?	1: Sim 2: Não 3: Não se aplica 9: Ignorado
TPNASCASSI	Nascimento foi assistido por?	1: Médico 2: Enfermeira/obstetiz 3: Parteira 4: Outros 9: Ignorado

Fim da tabela

* A tabela de códigos UF do IBGE pode ser encontrada em <https://www.ibge.gov.br/explica/codigos-dos-municipios.php>

Selecionando algumas variáveis de interesse para construção do dataframe a ser usado na aula

A seguir, vamos construir um novo dataframe, contendo somente as variáveis descritas anteriormente. Para isso, construímos um vetor auxiliar contendo as variáveis desejadas e, em seguida, definimos o novo dataframe (df) contendo apenas estas variáveis.

```
vars_selecionadas <- c('LOCNASC', 'IDADEMAE', 'ESTCIVMAE', 'QTDFILVIVO', 'QTDFILMORT',
  'GESTACAO', 'GRAVIDEZ', 'PARTO', 'CONSULTAS', 'DTNASC',
  'SEXO', 'APGAR1', 'APGAR5', 'RACACOR', 'PESO',
  'IDANOMAL', 'CODUFNATU', 'ESMAE2010', 'RACACORMAE', 'QTDGESTANT',
  'QTDPARTNOR', 'QTDPARTCES', 'IDADEPAI', 'SEMAGESTAC', 'CONSPRENAT',
  'STTRABPART', 'STCESPARTO', 'TPNASCASSI')
```

```
df <- df.nv24[,vars_selecionadas]
dim(df)
```

```
## [1] 779927      28
```

O dataframe `df` contém as 779.927 observações e as 28 variáveis definidas na Tabela 1. Entretanto, a base vem com os códigos (números) dos níveis das variáveis e, para trabalharmos corretamente com ela, precisamos primeiramente tratar tais variáveis, pois não se tratam de variáveis numéricas, e sim categóricas, cujos fatores estão representados por números. A seguir, faremos o tratamento necessário na base de dados.

Transformando as variáveis categóricas em fatores e adicionando os respectivos labels

• LOCALNASC (Local de ocorrência do nascimento)

```
table(df$LOCNASC)
```

```
##
##      1      2      3      4      5      9
## 768765 4829 4113 1987  202   31
```

```
df$LOCNASC <- factor(df$LOCNASC, levels = c(1,2,3,4,5,9),
  labels = c("Hospital",
    "Outro Estab Saúde",
    "Domicílio",
    "Aldeia Indígena",
    "Outros",
    "Ignorado"),
  ordered = FALSE)
```

```
table(df$LOCNASC)
```

```
##
##      Hospital Outro Estab Saúde      Domicílio Aldeia Indígena
##      768765           4829           4113           1987
##      Outros           Ignorado
##      202           31
```

• ESTCIVMAE (Estado civil da mãe)

```
table(df$ESTCIVMAE)
```

```
##
##      1      2      3      4      5      9
## 403848 242908  1286 13273 112982  2192
```

```
df$ESTCIVMAE <- factor(df$ESTCIVMAE, levels = c(1,2,3,4,5,9),
  labels = c("Solteira",
             "Casada",
             "Viúva",
             "Separada judicialmente",
             "União consensual",
             "Ignorado"),
  ordered = FALSE)
table(df$ESTCIVMAE)
```

```
##
##           Solteira           Casada           Viúva
##           403848           242908           1286
## Separada judicialmente      União consensual      Ignorado
##           13273           112982           2192
```

• GESTACAO (Semanas de gestação)

```
table(df$GESTACAO)
```

```
##
##      1      2      3      4      5      6      9
##    474   4540   8881  89511 658364  11245    52
```

```
df$GESTACAO <- factor(df$GESTACAO, levels = c(1,2,3,4,5,6,9),
  labels = c("Menos de 22 semanas",
             "22 a 27 semanas",
             "28 a 31 semanas",
             "32 a 36 semanas",
             "37 a 41 semanas",
             "42 semanas e mais",
             "Ignorado"),
  ordered = TRUE)
table(df$GESTACAO)
```

```
##
## Menos de 22 semanas      22 a 27 semanas      28 a 31 semanas      32 a 36 semanas
##              474              4540              8881              89511
##      37 a 41 semanas  42 semanas e mais      Ignorado
##              658364              11245              52
```

- GRAVIDEZ (Tipo de gravidez)

```
table(df$GRAVIDEZ)
```

```
##
##      1      2      3      9
## 760603 18191  302   16
```

```
df$GRAVIDEZ <- factor(df$GRAVIDEZ, levels= c(1,2,3,9),
                      labels = c("Única",
                                "Dupla",
                                "Tripla e mais",
                                "Ignorado"),
                      ordered = TRUE)
```

```
table(df$GRAVIDEZ)
```

```
##
##      Única      Dupla Tripla e mais      Ignorado
##      760603      18191      302      16
```

- CONSULTAS (Número de consultas de pré-natal)

```
table(df$CONSULTAS)
```

```
##
##      1      2      3      4      9
## 11492 33462 122689 609067  2448
```

```
df$CONSULTAS <- factor(df$CONSULTAS, levels = c(1,2,3,4,9),
                      labels = c("Nenhuma",
                                "de 1 a 3",
                                "de 4 a 6",
                                "7 e mais",
                                "Ignorado"),
                      ordered=TRUE)
```

```
table(df$CONSULTAS)
```

```
##
## Nenhuma de 1 a 3 de 4 a 6 7 e mais Ignorado
##   11492   33462  122689  609067    2448
```

- SEXO (Sexo)

```
table(df$SEXO)
```

```
##
##      0      1      2
##  129 399587 380211
```

```
df$SEXO <- factor(df$SEXO, levels = c(0,1,2),
                  labels = c("Ignorado, não informado",
                            "Masculino",
                            "Feminino"),
                  ordered = FALSE)
```

```
table(df$SEX0)
```

```
##
## Ignorado, não informado      Masculino      Feminino
##           129           399587           380211
```

• RACACOR (Raça/Cor)

```
table(df$RACACOR)
```

```
##
##      1      2      3      4      5
## 258580 62226 3475 434158 8501
```

```
df$RACACOR <- factor(df$RACACOR, levels = c(1,2,3,4,5),
                     labels = c("Branca",
                                "Preta",
                                "Amarela",
                                "Parda",
                                "Indígena"),
                     ordered = FALSE)
```

```
table(df$RACACOR)
```

```
##
##   Branca   Preta  Amarela   Parda Indígena
##   258580   62226    3475   434158    8501
```

Exercício:

Continue o processo de formatação das seguintes variáveis categóricas, transformando-as em fator e adicionando os respectivos labels, de acordo com o descrito na Tabela 1:

- PARTO (Tipo de parto)
- IDANOMAL (Anomalia congênita)
- CODUFNATU (Código UF de naturalidade da mãe - Código UF do IBGE)
- ESCMAE2010 (Escolaridade da mãe - 2010)
- RACACORMAE (Raça/cor da mãe)
- STTRABPART (Trabalho parto induzido?)
- STCESPARTO (Cesárea ocorreu antes do trabalho de parto iniciar?)
- TPNASCASSI (Nascimento foi assistido por...)

Formatando variável de data

- DTNASC (Data do nascimento)

Observação: A variável DTNASC veio como número inteiro, no formato ddmmaaaa, e o padrão em programação é aaaa-mm-dd. Vamos utilizar o pacote `lubridate` para transformá-la no formato adequado.

```
# install.packages("lubridate") # Remover o # para instalar, caso necessário
library(lubridate)
```

```
##
## Attaching package: 'lubridate'

## The following objects are masked from 'package:base':
##
##   date, intersect, setdiff, union
```

```
df$DTNASC <- dmy(df$DTNASC)
head(df$DTNASC)
```

```
## [1] "2024-02-14" "2024-04-17" "2024-01-01" "2024-01-01" "2024-01-01"
## [6] "2024-01-01"
```

Criando novas variáveis

- Criando a variável MES.NASC (mês de nascimento) a partir da variável DTNASC (Data de nascimento):

```
df$MES.NASC <- month(df$DTNASC)
head(df$MES.NASC)
```

```
## [1] 2 4 1 1 1 1
```

```
## Adicionando os labels das categorias da variável:
```

```
df$MES.NASC <- factor(df$MES.NASC, levels = 1:5, # 1:12,
                      labels = c("Jan", "Fev", "Mar", "Abr", "Mai"),
                      # , "Jun", "Jul", "Ago", "Set", "Out", "Nov", "Dez"),
                      ordered = TRUE)
```

```
table(df$MES.NASC)
```

```
##
##      Jan      Fev      Mar      Abr      Mai
## 200999 194103 204597 160205 20023
```

- Criando SEMAGESTAC_cat a partir de SEMAGESTAC p/ comparar com a variável GESTACAO:

```
table(df$GESTACAO) #Obs.: Essa variável já veio categorizada no dataframe
```

```
##
## Menos de 22 semanas      22 a 27 semanas      28 a 31 semanas      32 a 36 semanas
##           474              4540              8881              89511
##      37 a 41 semanas  42 semanas e mais      Ignorado
##           658364              11245              52
```

```
table(df$SEMAGESTAC)
```

```
##
##      19      20      21      22      23      24      25      26      27      28      29
##    108    132    234    384    485    672    808    997    1194    1440    1708
##      30      31      32      33      34      35      36      37      38      39      40
##   2361   3372   5367   8091  13570  20735  41748  97061 182587 215045 124419
##       41      42      43      44      45
##  39252   7372   2221   1111   541
```

```
# Reproduzindo os mesmos intervalos de classes da variável GESTACAO:
df$SEMAGESTAC_cat <- cut(df$SEMAGESTAC, breaks=c(0,21,27,31,36,41,50) )
table(df$SEMAGESTAC_cat)
```

```
##
## (0,21] (21,27] (27,31] (31,36] (36,41] (41,50]
##      474    4540    8881    89511  658364    11245
```

```
# Podemos ainda adicionar os rótulos das classes:
df$SEMAGESTAC_cat <- cut(df$SEMAGESTAC, breaks=c(0,21,27,31,36,41,50),
                          labels = c("Menos de 22 semanas",
                                      "22 a 27 semanas",
                                      "28 a 31 semanas",
                                      "32 a 36 semanas",
                                      "37 a 41 semanas",
                                      "42 semanas e mais"),
                          ordered = TRUE)
table(df$SEMAGESTAC_cat)
```

```
##
## Menos de 22 semanas      22 a 27 semanas      28 a 31 semanas      32 a 36 semanas
##           474              4540              8881              89511
##      37 a 41 semanas  42 semanas e mais
##           658364              11245
```

- Criando CONSPRENAT_cat a partir de CONSPRENAT p/ comparar com a variável CONSULTAS:

```
table(df$CONSULTAS) #Obs.: Essa variável já veio categorizada no dataframe
```

```
##
## Nenhuma de 1 a 3 de 4 a 6 7 e mais Ignorado
##    11492    33462    122689    609067    2448
```

```
table(df$CONSPRENAT)
```

```
##
```

```
##      0      1      2      3      4      5      6      7      8      9     10
##  4640   6843  10558  15974  24760  37658  59815  79505  103054  102543  119726
##     11     12     13     14     15     16     17     18     19     20     21
##  62602  56451  28046  24612  10966  6001   4002   2890   1721   1759   654
##     22     23     24     25     26     27     28     29     30     31     32
##    423    248    200    139    55    39    25    27    33    7    10
##     33     34     35     36     37     38     39     40     41     42     46
##     10      4     16     10     48    112    135    71    21     1     1
##     50     60     64     65     68     69     70     79     80     99
##      1      2      1      1      1      1      2      1      2    2418
```

Reproduzindo os mesmos intervalos de classes da variável CONSULTAS:

```
df$CONSPRENAT_cat <- cut(df$CONSPRENAT, breaks=c(-0.01,0,3,6,98,99))
table(df$CONSPRENAT_cat)
```

```
##
## (-0.01,0]      (0,3]      (3,6]      (6,98]      (98,99]
##      4640      33375      122233      606179      2418
```

Podemos ainda adicionar os rótulos das classes:

```
df$CONSPRENAT_cat <- cut(df$CONSPRENAT, breaks=c(-0.01,0,3,6,98,99),
                          labels = c("Nenhuma", "1 a 3", "4 a 6", "7 ou mais","Ignorado"),
                          ordered = TRUE)
table(df$CONSPRENAT_cat)
```

```
##
##   Nenhuma      1 a 3      4 a 6 7 ou mais  Ignorado
##      4640      33375      122233      606179      2418
```

Obs.: Podemos observar uma certa inconsistência entre as duas variáveis.

Versão final do dataframe

Conferindo o dataframe final, temos:

```
str(df)
```

```
## 'data.frame':      779927 obs. of  31 variables:
## $ LOCNASC      : Factor w/ 6 levels "Hospital","Outro Estab Saúde",...: 1 1 1 1 1 1 1 1 1 1 ...
## $ IDADEMAE     : int   24 29 20 40 27 19 17 29 15 23 ...
## $ ESTCIVMAE    : Factor w/ 6 levels "Solteira","Casada",...: 1 2 5 1 2 1 1 1 5 2 ...
## $ QTDFILVIVO   : int    1 0 0 4 3 0 0 0 0 1 ...
## $ QTDFILMORT   : int    0 0 0 1 0 0 1 0 0 0 ...
## $ GESTACAO     : Ord.factor w/ 7 levels "Menos de 22 semanas"<...: 5 5 5 5 4 5 5 5 5 5 ...
## $ GRAVIDEZ     : Ord.factor w/ 4 levels "Única"<"Dupla"<...: 1 1 1 1 1 1 1 1 1 1 ...
## $ PARTO       : Factor w/ 3 levels "Vaginal","Cesáreo",...: 2 2 1 2 1 2 2 2 1 2 ...
## $ CONSULTAS   : Ord.factor w/ 5 levels "Nenhuma"<"de 1 a 3"<...: 2 4 4 4 4 4 4 4 4 4 ...
## $ DTNASC      : Date, format: "2024-02-14" "2024-04-17" ...
## $ SEXO        : Factor w/ 3 levels "Ignorado, não informado",...: 2 2 2 2 2 2 3 3 3 3 ...
## $ APGAR1      : int    8 8 8 9 8 9 8 8 9 9 ...
```

```
## $ APGAR5 : int 9 9 9 9 9 9 9 9 10 10 ...
## $ RACACOR : Factor w/ 5 levels "Branca","Preta",...: 4 4 1 4 4 4 4 2 4 1 ...
## $ PESO : int 3120 3564 3240 3960 3610 3724 3552 3764 3225 3430 ...
## $ IDANOMAL : Factor w/ 3 levels "Sim","Não","Ignorado": 2 2 2 2 2 2 2 2 2 2 ...
## $ CODUFNATU : Factor w/ 28 levels "RO","AC","AM",...: 1 1 1 2 1 1 1 1 2 1 ...
## $ ESCMAE2010 : Ord.factor w/ 7 levels "Sem escolaridade"<...: 3 6 4 6 4 4 4 6 3 4 ...
## $ RACACORMAE : Factor w/ 5 levels "Branca","Preta",...: 4 4 1 4 4 4 4 2 4 1 ...
## $ QTDGESTANT : int 1 0 0 5 3 0 1 0 0 1 ...
## $ QTDPARTNOR : int 0 0 0 1 3 0 0 0 0 0 ...
## $ QTDPARTCES : int 1 0 0 3 0 0 0 0 0 1 ...
## $ IDADEPAI : int NA 41 NA NA NA NA NA NA NA NA ...
## $ SEMAGESTAC : int 38 39 38 38 36 41 40 40 40 39 ...
## $ CONSPRENAT : int 2 8 10 7 10 8 10 10 12 9 ...
## $ STTRABPART : Factor w/ 3 levels "Sim","Não","Ignorado": 2 2 2 2 2 2 1 1 2 2 ...
## $ STCESPARTO : Factor w/ 4 levels "Sim","Não","Não se aplica",...: 2 1 3 2 3 2 2 2 3 1 ...
## $ TPNASCASSI : Factor w/ 5 levels "Médico","Enfermeira/obstettriz",...: 1 1 2 1 2 1 1 1 1 1 ...
## $ MES.NASC : Ord.factor w/ 5 levels "Jan"<"Fev"<"Mar"<...: 2 4 1 1 1 1 1 1 1 1 ...
## $ SEMAGESTAC_cat: Ord.factor w/ 6 levels "Menos de 22 semanas"<...: 5 5 5 5 4 5 5 5 5 5 ...
## $ CONSPRENAT_cat: Ord.factor w/ 5 levels "Nenhuma"<"1 a 3"<...: 2 4 4 4 4 4 4 4 4 4 ...
```

Exportando a base final

Essa será a base que utilizaremos para as análises da disciplina. Vamos exportá-la em um arquivo csv (Dataframe_AulaAmanda.csv), que está disponível na plataforma do curso.

```
setwd <- "C:\\Users\\AmandaBGM\\Google Drive\\UFG\\Especialização_FEN_IME\\2024\\Scripts"
write.csv2(df, "Dataframe_AulaAmanda.csv", row.names = F)
```

Podemos importar para conferir se a exportação ocorreu com sucesso:

```
setwd <- "C:\\Users\\AmandaBGM\\Google Drive\\UFG\\Especialização_FEN_IME\\2024\\Scripts"

# 2 opções de importação (.csv com separador ponto-e-vírgula):

# dados <- read.csv2(file = "Dataframe_AulaAmanda.csv", header = TRUE)
# ou:
dados <- read.csv(file = "Dataframe_AulaAmanda.csv", sep=';', header = TRUE)

head(dados)
```

```
##   LOCNASC IDADEMAE      ESTCIVMAE QTDFILVIVO QTDFILMORT      GESTACAO
## 1 Hospital      24      Solteira           1           0 37 a 41 semanas
## 2 Hospital      29      Casada           0           0 37 a 41 semanas
## 3 Hospital      20 União consensual       0           0 37 a 41 semanas
## 4 Hospital      40      Solteira           4           1 37 a 41 semanas
## 5 Hospital      27      Casada           3           0 32 a 36 semanas
## 6 Hospital      19      Solteira           0           0 37 a 41 semanas
```

##	GRAVIDEZ	PARTO	CONSULTAS	DTNASC	SEXO	APGAR1	APGAR5	RACACOR	PESO
## 1	Única	Cesáreo	de 1 a 3	2024-02-14	Masculino	8	9	Parda	3120
## 2	Única	Cesáreo	7 e mais	2024-04-17	Masculino	8	9	Parda	3564
## 3	Única	Vaginal	7 e mais	2024-01-01	Masculino	8	9	Branca	3240
## 4	Única	Cesáreo	7 e mais	2024-01-01	Masculino	9	9	Parda	3960
## 5	Única	Vaginal	7 e mais	2024-01-01	Masculino	8	9	Parda	3610
## 6	Única	Cesáreo	7 e mais	2024-01-01	Masculino	9	9	Parda	3724
##	IDANOMAL	CODUFNATU			ESCMAE2010	RACACORMAE	QTDGESTANT		
## 1	Não	RO	Fundamental II (5ª a 8ª série)			Parda		1	
## 2	Não	RO	Superior completo			Parda		0	
## 3	Não	RO	Médio (antigo 2º grau)			Branca		0	
## 4	Não	AC	Superior completo			Parda		5	
## 5	Não	RO	Médio (antigo 2º grau)			Parda		3	
## 6	Não	RO	Médio (antigo 2º grau)			Parda		0	
##	QTDPARTNOR	QTDPARTCES	IDADEPAI	SEMAGESTAC	CONSPRENAT	STTRABPART	STCESPARTO		
## 1	0	1	NA	38	2	Não	Não		
## 2	0	0	41	39	8	Não	Sim		
## 3	0	0	NA	38	10	Não	Não se aplica		
## 4	1	3	NA	38	7	Não	Não		
## 5	3	0	NA	36	10	Não	Não se aplica		
## 6	0	0	NA	41	8	Não	Não		
##	TPNASCASSI	MES.NASC	SEMAGESTAC_cat	CONSPRENAT_cat					
## 1	Médico	Fev 37 a 41 semanas		1 a 3					
## 2	Médico	Abr 37 a 41 semanas		7 ou mais					
## 3	Enfermeira/obstetriz	Jan 37 a 41 semanas		7 ou mais					
## 4	Médico	Jan 37 a 41 semanas		7 ou mais					
## 5	Enfermeira/obstetriz	Jan 32 a 36 semanas		7 ou mais					
## 6	Médico	Jan 37 a 41 semanas		7 ou mais					

```
dim(dados)
```

```
## [1] 779927      31
```