***Desafio Sanofi***

* **DESCOBERTA DA BASE**

Para acessar as Bases do DATASUS entramos no link ([https://datasus.saude.gov.br/transferencia-de-arquivos/#](https://datasus.saude.gov.br/transferencia-de-arquivos/)).

Temos diversas Bases, elas estão divididas por assuntos. Ex:



Nosso primeiro passo foi extrair uma dessas Bases para começarmos a analisar os dados. (**Vou utilizar a base SIASUS como exemplo**).

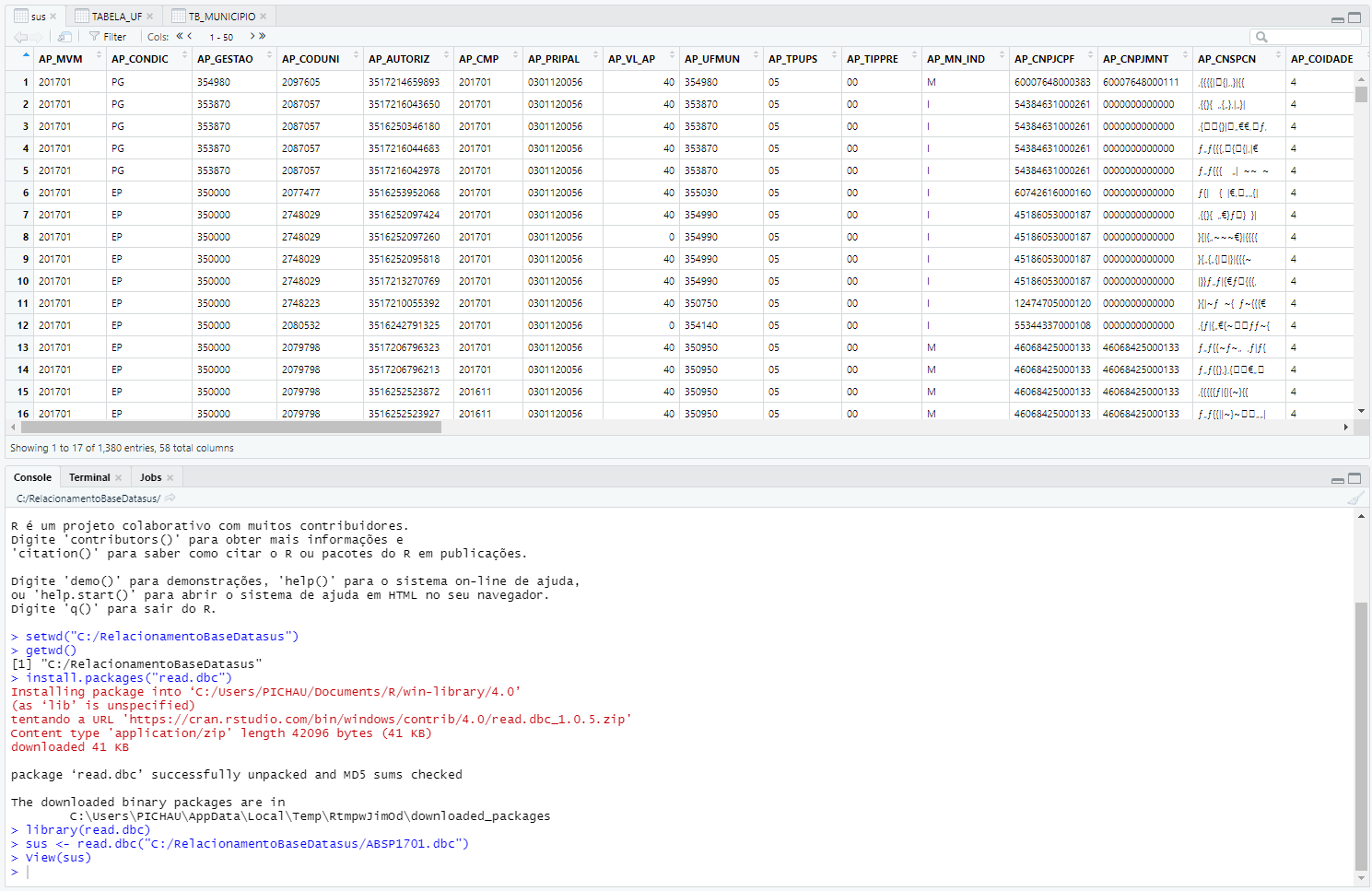
Para fazer o download de uma base do sistema DATASUS devemos passar alguns parâmetros como, Ano, Mês e UF. Essa base vem no formato .DBC e para analisá-la e transformá-la em arquivo .xlsx para posteriormente gerarmos insights no Power BI utilizamos a Linguagem R.

Segue abaixo os comandos no **RStudio** utilizados para leitura do arquivo do DATASUS.

**PASSO A PASSO:**

|  |  |
| --- | --- |
| **Comando** | **O que ele faz?** |
| setwd("C:/RelacionamentoBaseDatasus") | Seleciona o diretório que o arquivo será trabalhado |
| getwd() | valida o diretório |
| install.packages("read.dbc") | instala o pacote para leitura do arquivo .dbc |
| library(read.dbc) | chama a biblioteca read.dbc |
| sus <- read.dbc("C:/RelacionamentoBaseDatasus/ABSP1701.dbc") | seleciona o arquivo |
| View(sus) | Gera uma View do Arquivo |

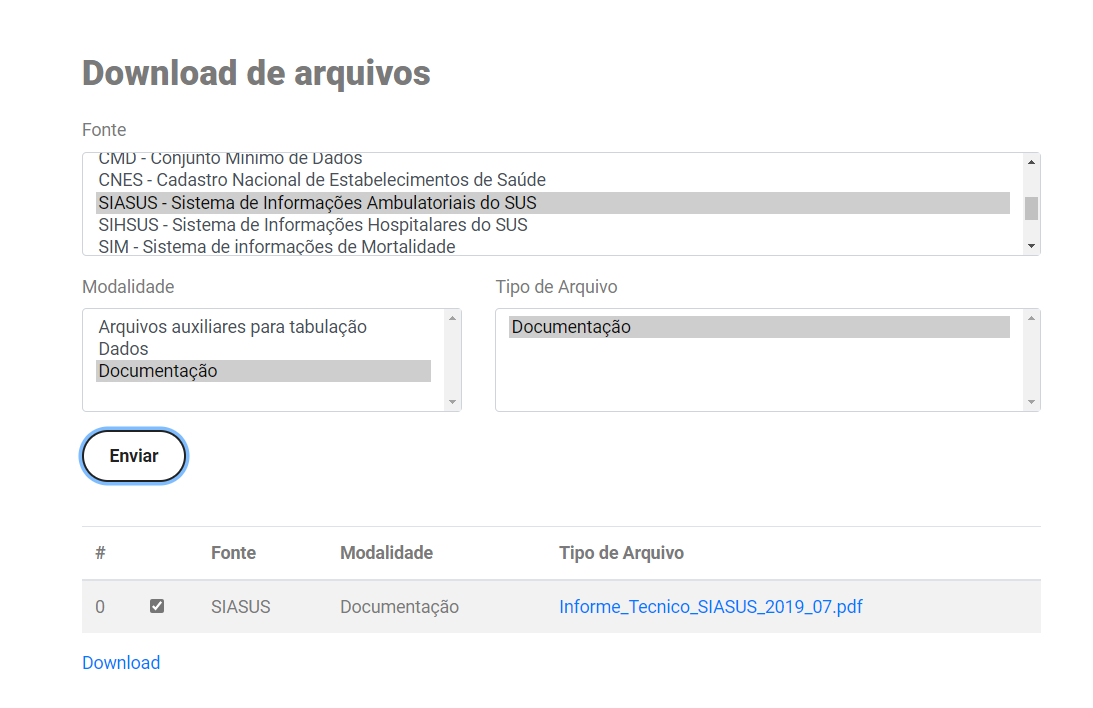
Seguindo o passo a passo acima conseguimos gerar uma massa de dados para analisar. Ex: (imagem abaixo)



* **MAPA DO ECOSSISTEMA**

1. Tratamento dos dados do sistema **- LAYOUT DO ARQUIVO DE APAC DE CIRURGIA BARIÁTRICA.**

Conforme podemos observar na imagem acima temos muitos campos que são Códigos/ Chaves que em um primeiro momento não conseguimos dar sentido a eles. Para um melhor entendimento fomos ler a documentação da base que baixamos. (no nosso exemplo foi a base (SIASUS)).

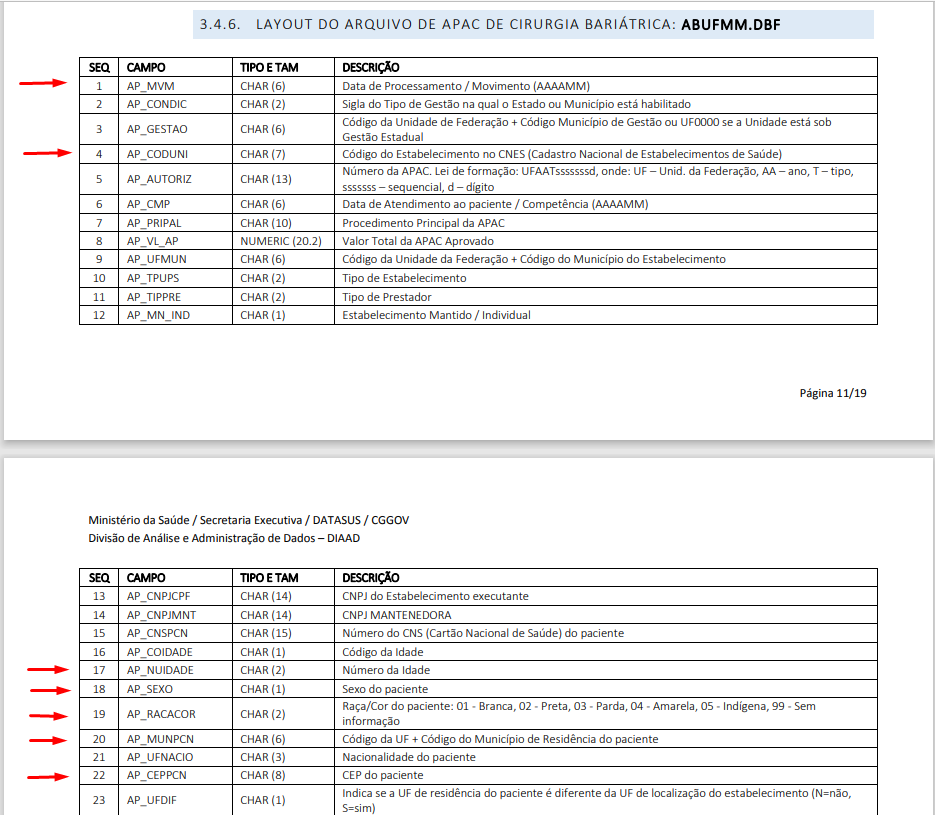


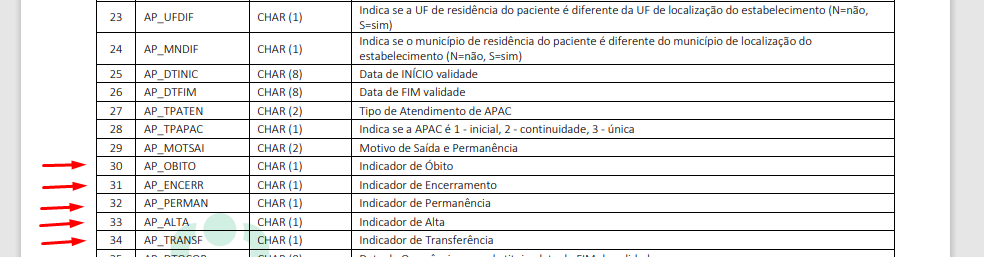
Analisando o documento do sistema conseguimos entender melhor esses dados e dar significado a eles.

**Imagem 1** - Print da documentação do DATASUS.

**Imagem 2** - Campo que vamos utilizar na nossa amostragem.

**Imagem 1:**





|  |  |
| --- | --- |
| **Nome do campo** | **O que ele significa?** |
| AP\_MVM | Data de Processamento / Movimento (AAAAMM) |
| AP\_CODUNI | Código do Estabelecimento no CNES (Cadastro Nacional de Estabelecimentos de Saúde) |
| AP\_NUIDADE | Número da Idade |
| AP\_SEXO | Sexo do paciente |
| AP\_RACACOR | Raça/Cor do paciente: 01 - Branca, 02 - Preta, 03 - Parda, 04 - Amarela, 05 - Indígena, 99 - Sem informação |
| AP\_MUNPCN | Código da UF + Código do Município de Residência do paciente |
| AP\_CEPPCN | CEP do paciente |
| AP\_OBITO | Indicador de Óbito |
| AP\_ENCERR | Indicador de Encerramento |
| AP\_PERMAN | Indicador de Permanência |
| AP\_ALTA | Indicador de Alta |
| AP\_TRANSF | Indicador de Transferência |

**Imagem 2:**

Com os campos escolhidos e com o entendimento do retorno de cada um deles fomos tratar os dados e para isso utilizamos a Linguagem SQL dentro do RStudio.

**Passo a Passo.**

|  |  |
| --- | --- |
| **Comando** | **O que ele faz?** |
| install.packages("sqldf") | Pacote para utilização do R com scripts de sql |
| library(sqldf) | Chama a biblioteca sqldf |

**O comando SQL utilizado para tratar os campos escolhidos:**

CIRURGIA\_BARIATRICA <- sqldf('

SELECT

"LAYOUT DO ARQUIVO DE APAC DE CIRURGIA BARIÁTRICA " AS TIPO\_DO\_DADO,

A.AP\_MVM DATA\_REGISTRO,

CASE

WHEN AP\_RACACOR = "01" THEN "Branca"

WHEN AP\_RACACOR = "02" THEN "Preta"

WHEN AP\_RACACOR = "03" THEN "Parda"

WHEN AP\_RACACOR = "04" THEN "Amarela"

WHEN AP\_RACACOR = "05" THEN "Indígena"

ELSE "Não Informado" END RACA,

A.AP\_CEPPCN as CEP,

CASE

WHEN AP\_SEXO = "F" THEN "Feminino"

ELSE "Masculino" END SEXO,

A.AP\_NUIDADE as IDADE,

A.AP\_OBITO AS INDICADOR\_OBITO,

A.AP\_ENCERR AS INDICADOR\_ENCERRAMENTO,

A.AP\_PERMAN AS INDICADOR\_PERMANENCIA,

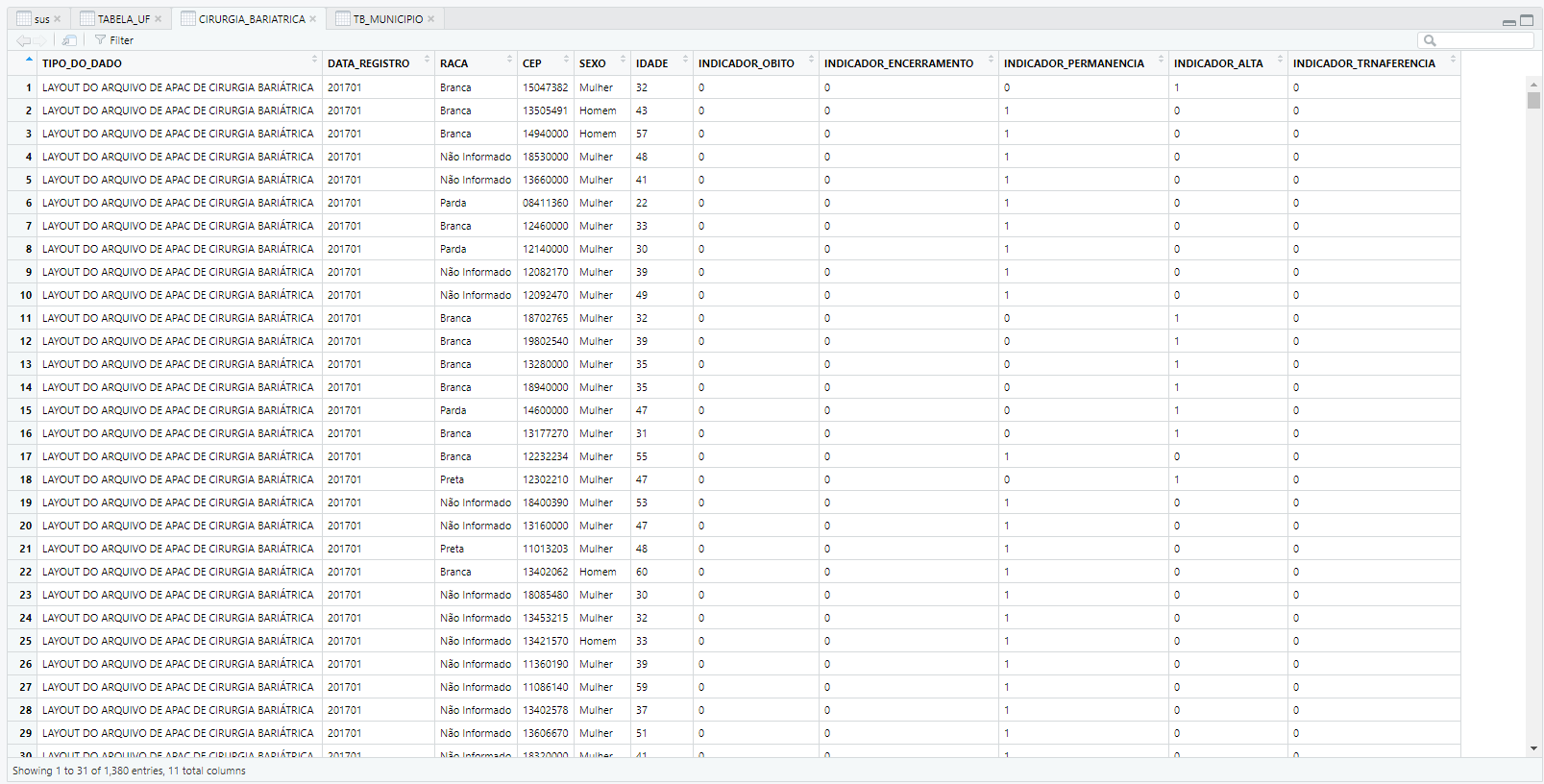
A.AP\_ALTA AS INDICADOR\_ALTA,

A.AP\_TRANSF AS INDICADOR\_TRNAFERENCIA

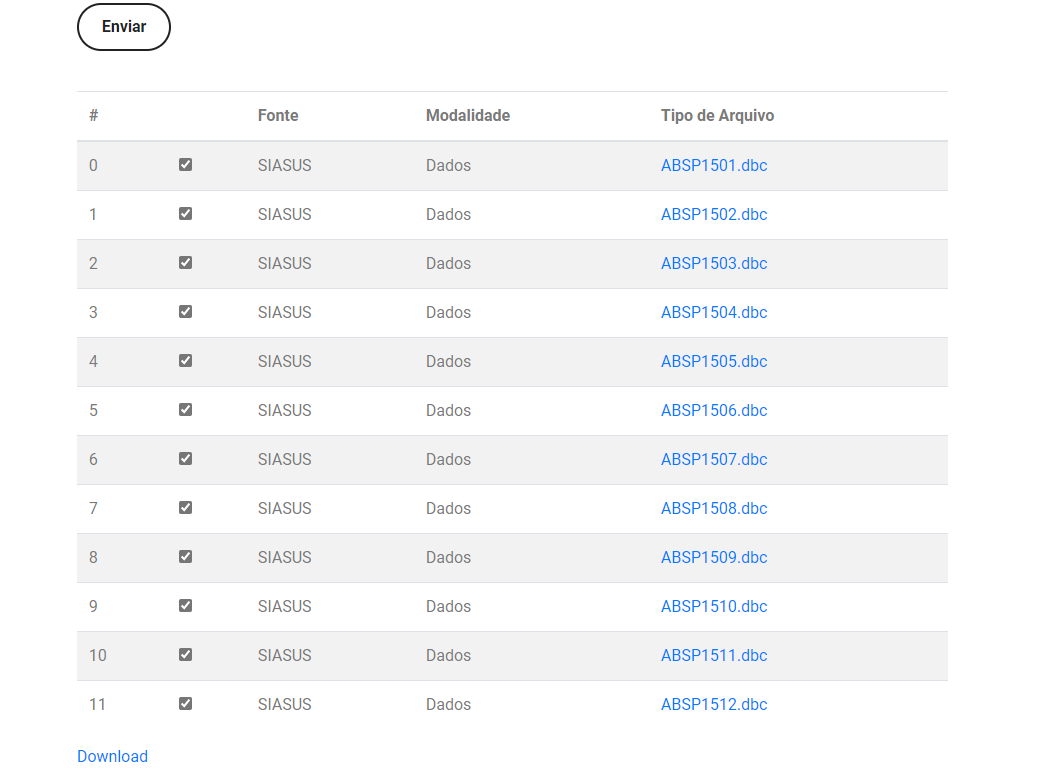
FROM sus as A

')

**Retorno do comando:**



Após os passos acima conseguimos dar sentido a alguns dados e temos os campos escolhidos da base de **Cirurgia Bariátrica do SIASUS** tratados.



Olhando a imagem acima podemos perceber que quando escolhemos a opção para baixar vários meses do sistema do SIASUS ele gera vários arquivos, para resolver esse problema executamos o comando abaixo:

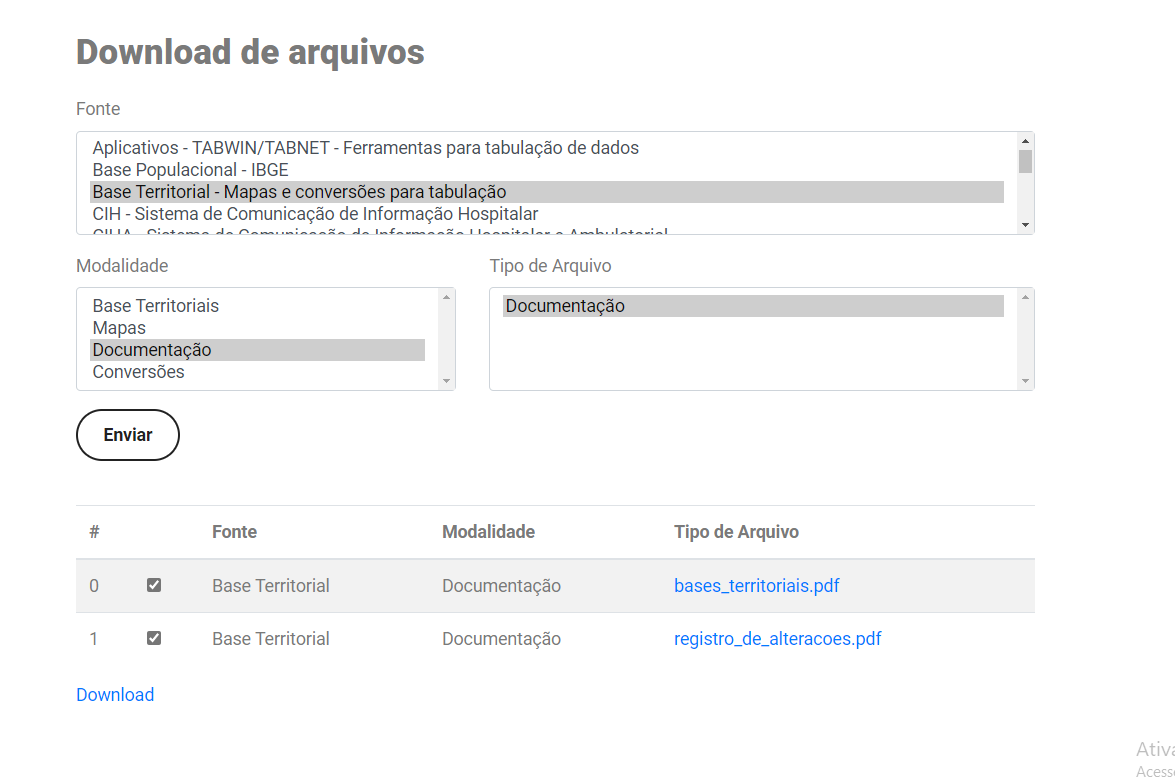
unionall <- sqldf('select \* from sus01 union select \* from sus02 union select \* from sus03 union select \* from sus04 union select \* from sus05 union select \* from sus06 union select \* from sus07 union select \* from sus08 union select \* from sus09 union select \* from sus10 union select \* from sus11 union select \* from sus12')

O comando acima serve para unificar os nossos dados e desta forma temos uma amostra com todos os meses escolhidos.

1. **Relacionamento entre (Base Territorial) e (SIASUS)**

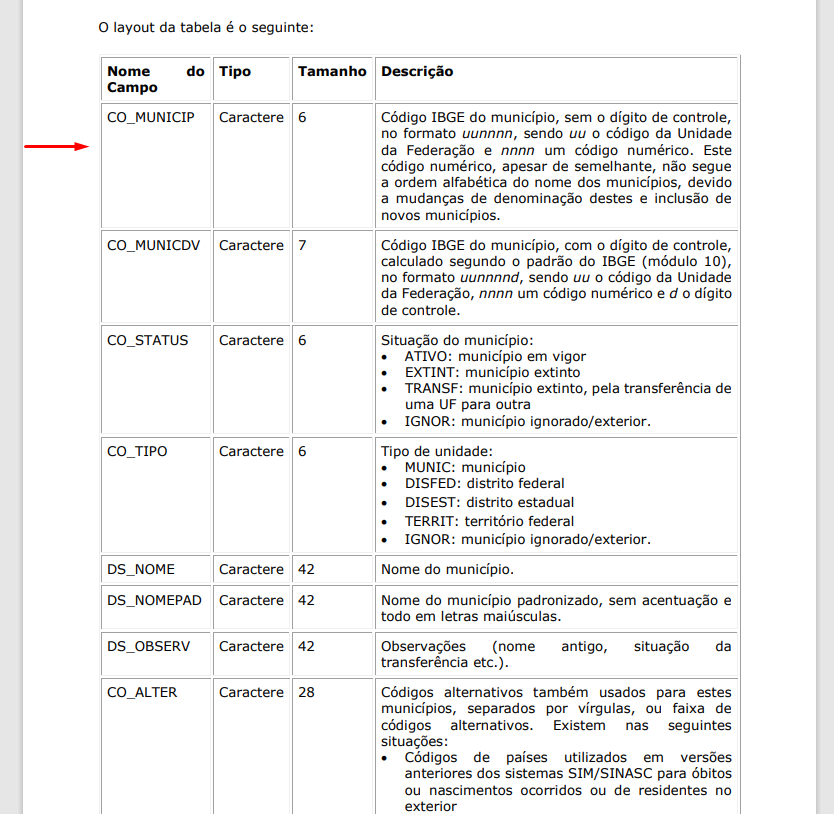
Seguindo com a análise das bases do DATASUS encontramos um relacionamento entre uma tabela da Base Territorial e alguns sistemas do SIASUS.

Download documentação (Base Territorial)

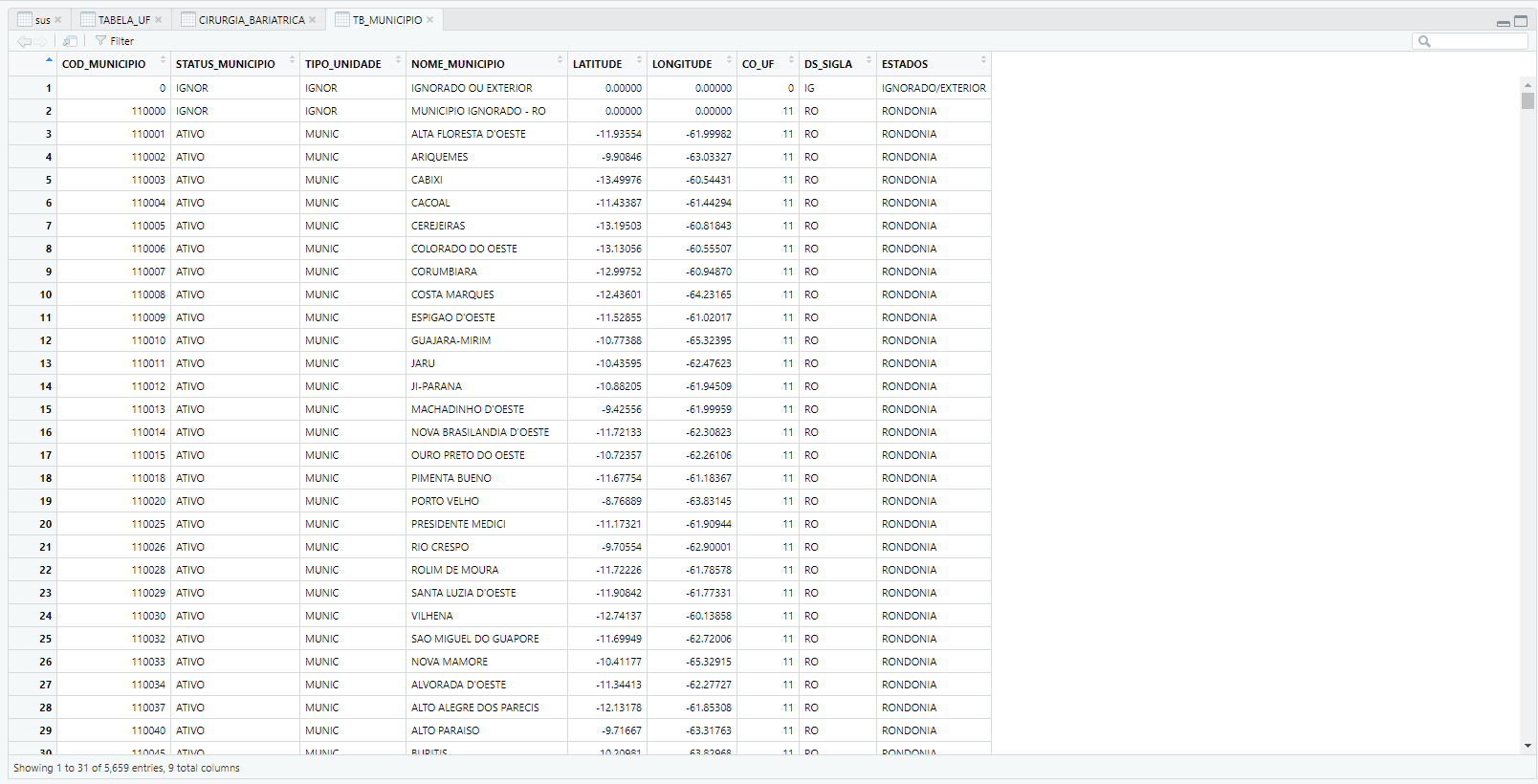


Encontramos na tabela **TB\_MUNICIP** um relacionamento do campo (**CO\_MUNICIP**) com o campo (**AP\_UFMUN**) de algumas bases do sistema SIASUS.

Essa tabela tem as informações dos municípios, desta forma conseguimos chegar no município do hospital em que o paciente foi atendido.



Essa tabela já vem no formato .xlsx, fizemos um tratamento apenas nos nomes das colunas e escolhemos as que vamos utilizar e importamos para dentro do RStudio.



**Para fecharmos esse relacionamento entre esses sistemas utilizamos o comando SQL conforme abaixo:**

CIRURGIA\_LOCAL\_HOSPITAL <- sqldf('

SELECT

"LAYOUT DO ARQUIVO DE APAC DE CIRURGIA BARIÁTRICA " AS TIPO\_DO\_DADO,

A.AP\_MVM DATA\_REGISTRO,

CASE

WHEN AP\_RACACOR = "01" THEN "Branca"

WHEN AP\_RACACOR = "02" THEN "Preta"

WHEN AP\_RACACOR = "03" THEN "Parda"

WHEN AP\_RACACOR = "04" THEN "Amarela"

WHEN AP\_RACACOR = "05" THEN "Indígena"

ELSE "Não Informado" END RACA,

A.AP\_CEPPCN as CEP,

CASE

WHEN AP\_SEXO = "F" THEN "Feminino"

ELSE "Masculino" END SEXO,

A.AP\_NUIDADE as IDADE,

A.AP\_OBITO AS INDICADOR\_OBITO,

A.AP\_ENCERR AS INDICADOR\_ENCERRAMENTO,

A.AP\_PERMAN AS INDICADOR\_PERMANENCIA,

A.AP\_ALTA AS INDICADOR\_ALTA,

A.AP\_TRANSF AS INDICADOR\_TRNAFERENCIA,

B.COD\_MUNICIP as CODIGO\_HOSPITAL,

B.STATUS as STATUS\_MUNICIPIO,

B.CO\_TIPO AS TIPO\_UNIDADE,

B.DS\_NOMEPAD AS MUNICIPIO\_HOSPITAL,

B.NU\_LATITUD AS LATITUDE,

B.NU\_LONGIT AS LONGITUDE,

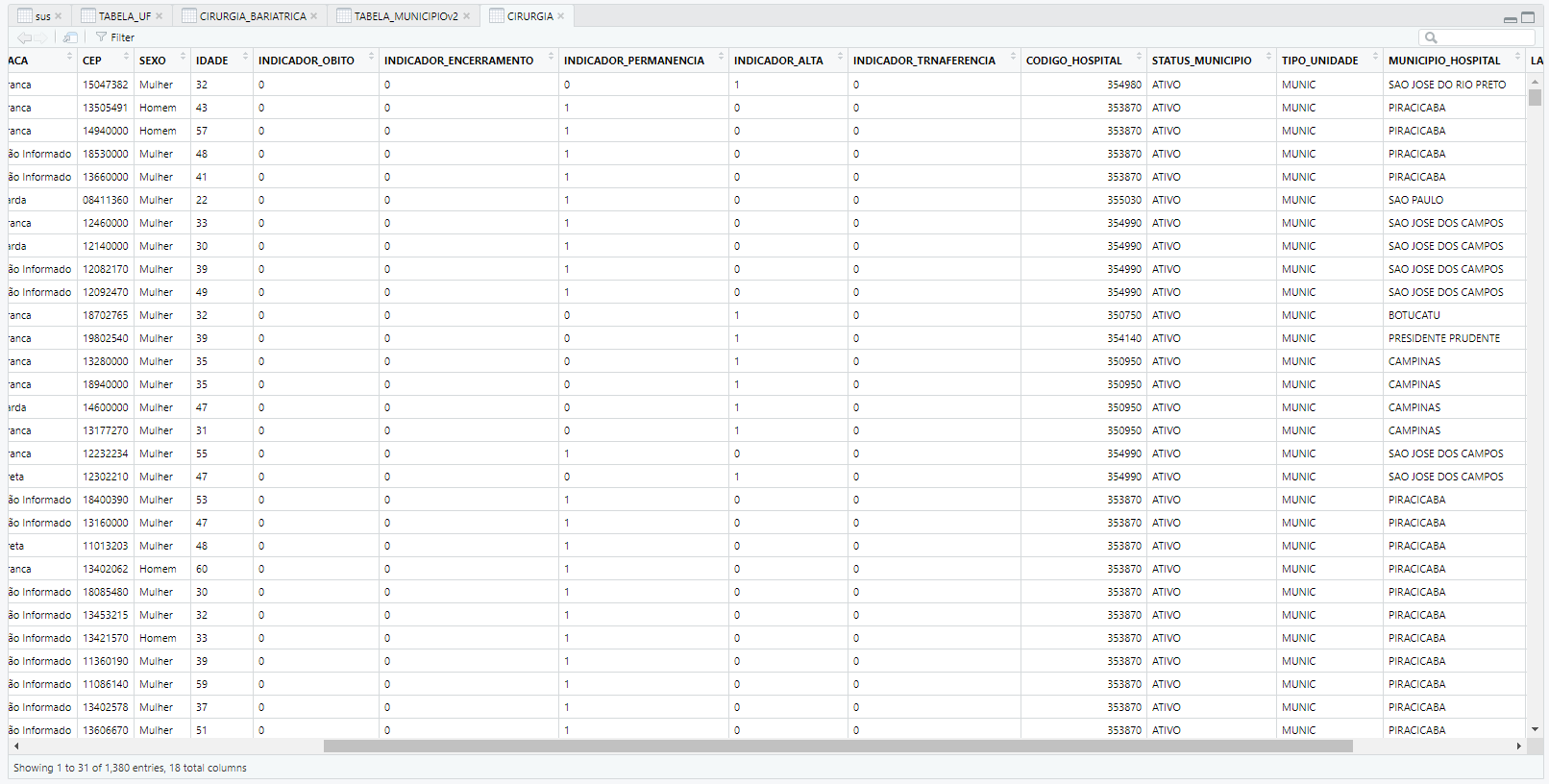
B.CO\_UF

FROM sus as A

INNER JOIN TABELA\_MUNICIPIOv2 as B on a.AP\_UFMUN = b.COD\_MUNICIP

')

**RESULTADO DO COMANDO:**



1. **Relacionamento entre (SIASUS) X (TB\_MUNICIP) X (TB\_UF)**

Explorando o sistema de **Base Territorial** encontramos um outro relacionamento entre a Tabela **TB\_MUNICIP** com a tabela **TB\_UF,** o retorno desta junção são os municípios que se encontram na tabela **TB\_MUNICIP** e seus devidos Estados que se encontram na tabela **TB\_UF.**

A tabela **TB\_UF** já vem no formato .xlsx, fizemos um tratamento apenas nos nomes das colunas e escolhemos as que vamos utilizar e importamos para dentro do RStudio.

**Para fecharmos esse relacionamento entre esses sistemas utilizamos o comando SQL conforme abaixo:**

TB\_MUNICIPIO <- sqldf('select

B.COD\_MUNICIP as CODIGO\_HOSPITAL,

B.STATUS as STATUS\_MUNICIPIO,

B.CO\_TIPO AS TIPO\_UNIDADE,

B.DS\_NOMEPAD AS MUNICIPIO\_HOSPITAL,

B.NU\_LATITUD AS LATITUDE,

B.NU\_LONGIT AS LONGITUDE,

B.CO\_UF,

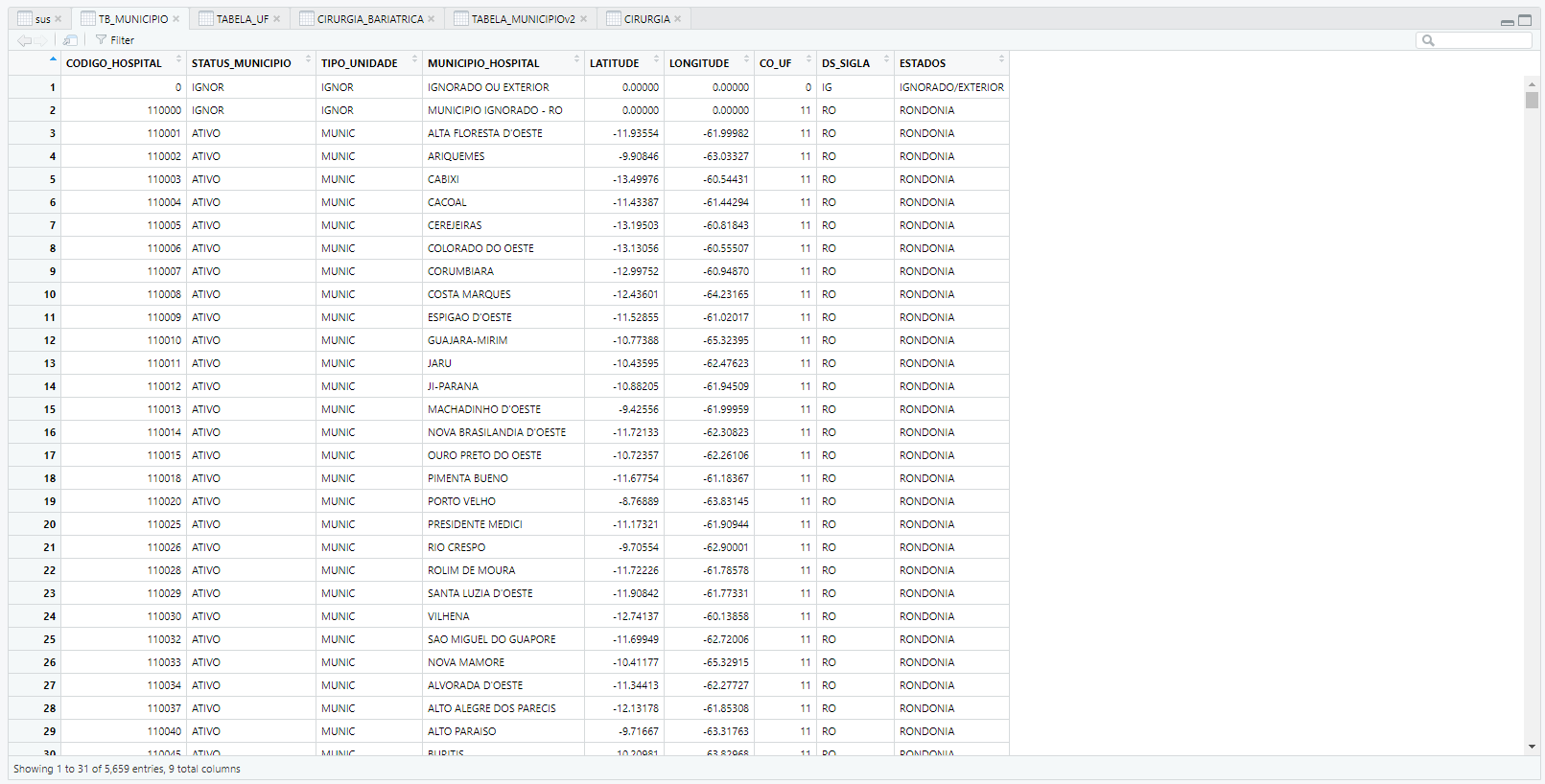
C.DS\_SIGLA,

C.DS\_NOMEPAD AS ESTADOS

FROM TABELA\_MUNICIPIOv2 as B

inner join TABELA\_UF as C on B.CO\_UF = C.CO\_UF ')

**RESULTADO DO COMANDO:**



Com isso podemos deixar o relacionamento entre (**Base Territorial**) e (**SIASUS**) mais completo fazendo mais um join com a tabela **TB\_UF.**

CIRURGIA\_LOCAL\_HOSPITAL\_UF <- sqldf('

SELECT

"LAYOUT DO ARQUIVO DE APAC DE CIRURGIA BARIÁTRICA " AS TIPO\_DO\_DADO,

A.AP\_MVM DATA\_REGISTRO,

CASE

WHEN AP\_RACACOR = "01" THEN "Branca"

WHEN AP\_RACACOR = "02" THEN "Preta"

WHEN AP\_RACACOR = "03" THEN "Parda"

WHEN AP\_RACACOR = "04" THEN "Amarela"

WHEN AP\_RACACOR = "05" THEN "Indígena"

ELSE "Não Informado" END RACA,

A.AP\_CEPPCN as CEP,

CASE

WHEN AP\_SEXO = "F" THEN "Feminino"

ELSE "Masculino" END SEXO,

A.AP\_NUIDADE as IDADE,

A.AP\_OBITO AS INDICADOR\_OBITO,

A.AP\_ENCERR AS INDICADOR\_ENCERRAMENTO,

A.AP\_PERMAN AS INDICADOR\_PERMANENCIA,

A.AP\_ALTA AS INDICADOR\_ALTA,

A.AP\_TRANSF AS INDICADOR\_TRNAFERENCIA,

B.COD\_MUNICIP as CODIGO\_HOSPITAL,

B.STATUS as STATUS\_MUNICIPIO,

B.CO\_TIPO AS TIPO\_UNIDADE,

B.DS\_NOMEPAD AS MUNICIPIO\_HOSPITAL,

B.NU\_LATITUD AS LATITUDE,

B.NU\_LONGIT AS LONGITUDE,

B.CO\_UF

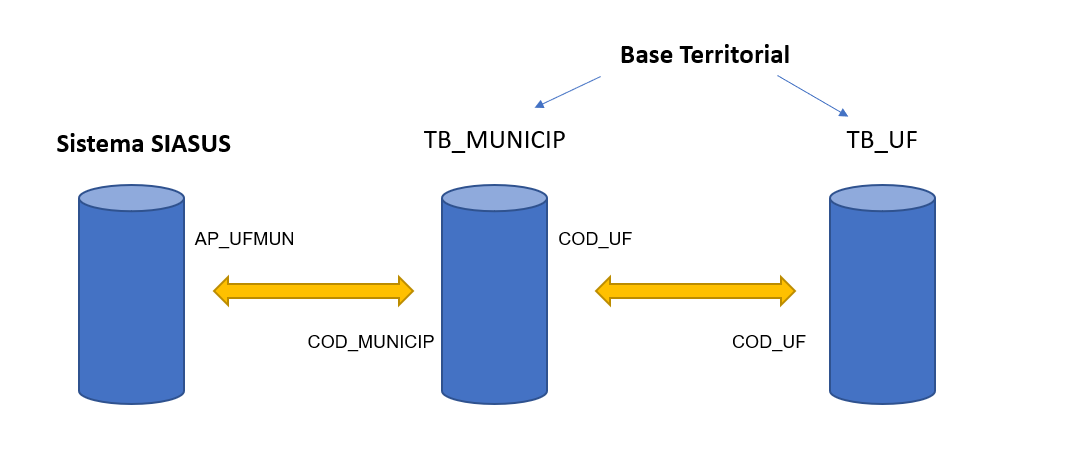
FROM sus as A

INNER JOIN TABELA\_MUNICIPIOv2 as B on a.AP\_UFMUN = b.COD\_MUNICIP

INNER JOIN TABELA\_UF as C on B.CO\_UF = C.CO\_UF

')

Imagem de relacionamento citado acima:

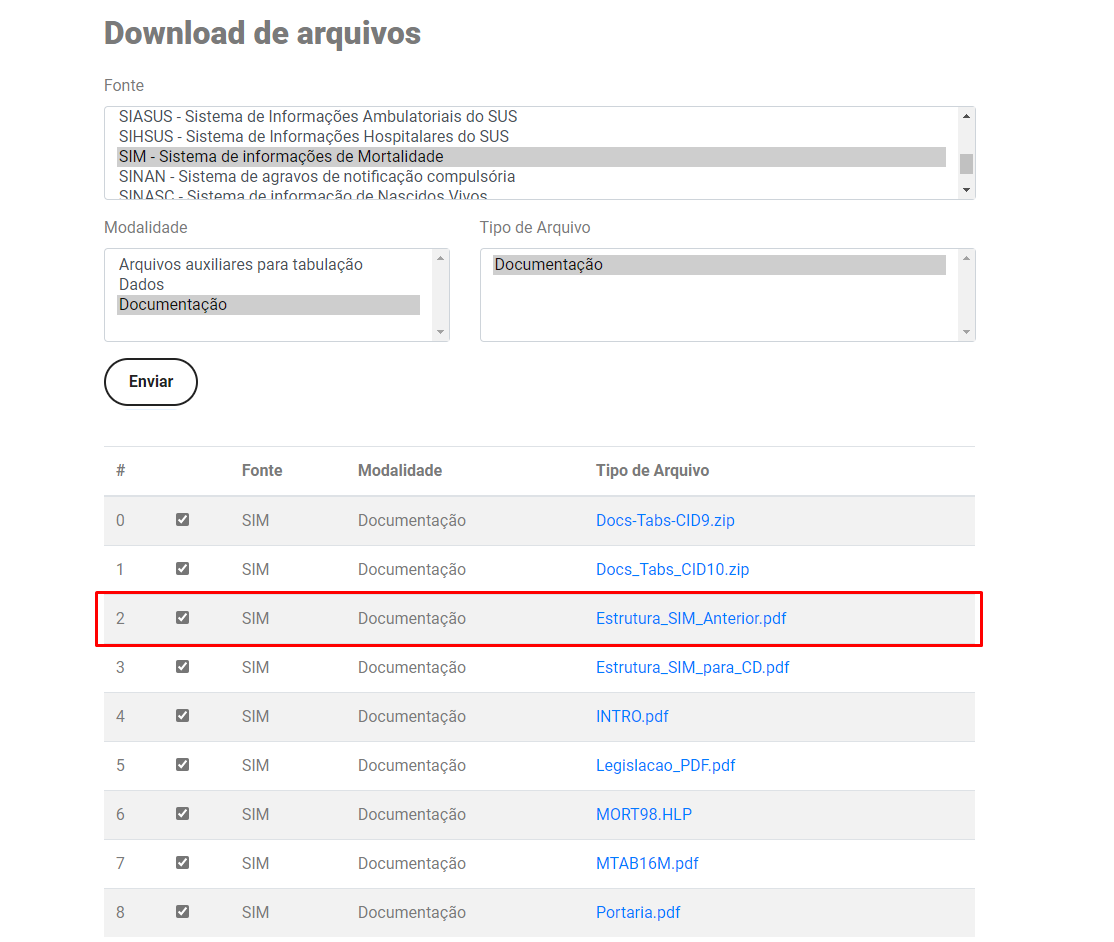


* Para extrair os dados dos outros sistemas do SIASUS que tem a mesma modelagem utilizamos o mesmo script alterando somente a fonte.
* **MAPA DO ECOSSISTEMA –> SIM-DO (MORTALIDADE)**

O passo a passo é bem parecido com o que está descrito acima.

Primeiramente baixamos um arquivo do sistema SIM-DO e lemos esse arquivo utilizando a linguagem R, como ele contém diversos campos que em um primeiro momento não fazem sentido para um melhor entendimento do relacionamento e a descrição de cada campo baixamos o documento do sistema e começamos a analisar.

**Imagem do documento utilizado para análise:**

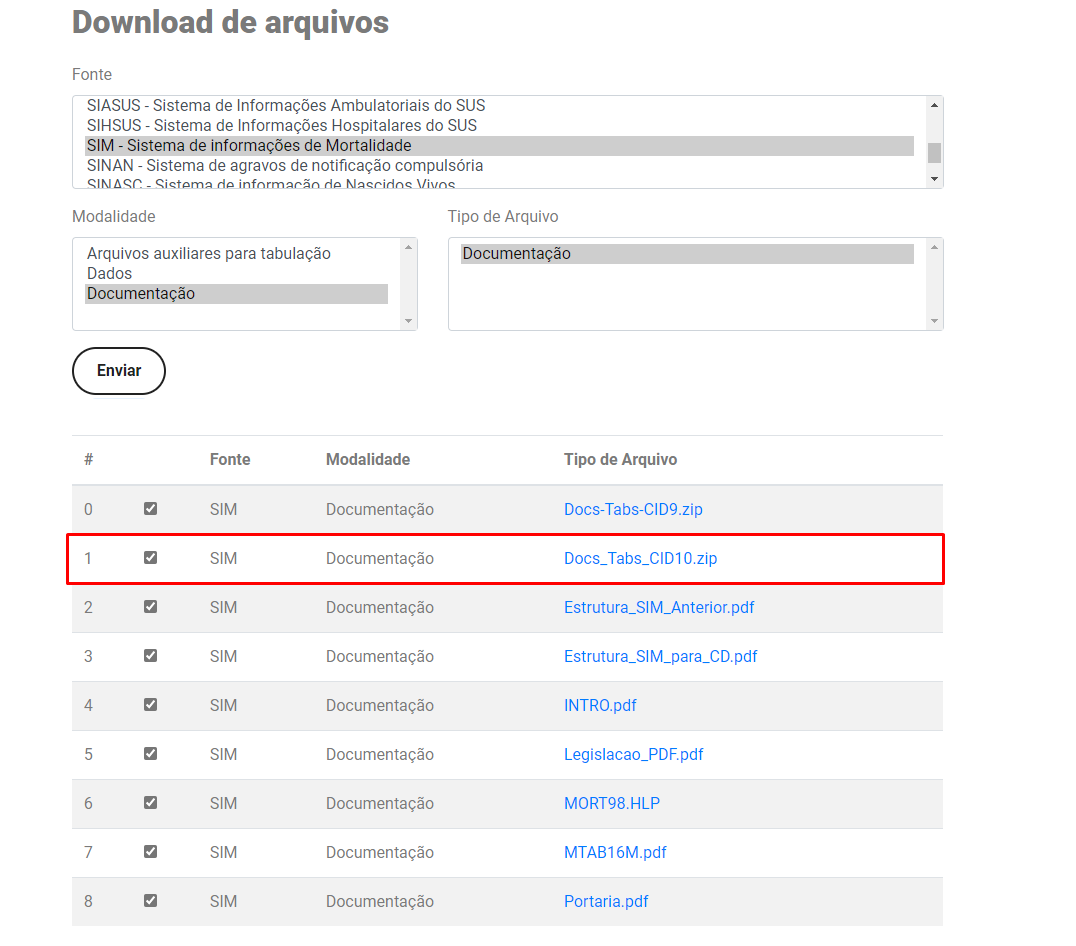


Campos como, TIPOBITO, SEXO, RACACOR, ESTCIV, ESC, LOCOCOR, CIRCOBITO, ACIDTRAB, TPPOS, IDADEMAE, ESCMAE e OBITOGRAV tratamos utilizando um (**CASE**) no SQL dentro do RSTUDIO. Campos como **CODMUNRES, CODMUNOCOR**, **CAUSABAS** utilizamos para fazer relacionamentos com outras tabelas.

A Base **SIM-DO** tem dois relacionamentos com a tabela **TB\_MUNICIP**, os campos são **CODMUNRES** (CÓDIGO\_MUNICIPIO\_RESIDENCIA) e **CODMUNOCOR** (CÓDIGO\_ MUNICIPIO\_DE\_OCORRENCIA\_DO\_OBITO) que correspondem ao campo **COD\_MUNICIP**.

Encontramos também um relacionamento da Base **SIM-DO** com uma tabela chamada (**CID10**) que existe dentro do arquivo zip **Docs\_Tabs\_CID10.zip.** O relacionamento é do campo **CAUSABAS** do **SIM-DO** com o campo **CID10** da tabela **(CID10)**. Com o retorno desse relacionamento conseguimos pegar a descrição da causa do óbito do paciente.

Imagem do arquivo zip:



Essa tabela (**CID10**) vem no formato **.BDF** e para analisá-la utilizamos o Excel.

Conforme já havíamos comentado existe um relacionamento da tabela (**TB\_MUNICIP**) com a tabela (**TB\_UF**), com isso utilizamos essa junção para pegar o Estado e a Sigla dos municípios de Residência e ocorrência do óbito.

**Para fecharmos esse relacionamento entre esses sistemas utilizamos o comando SQL abaixo:**

MORTALIDADE <- sqldf('

SELECT

"MORTALIDADE " AS TIPO\_DO\_DADO,

CASE

WHEN TIPOBITO = "1" THEN "ÓBITO FETAL"

ELSE "ÓBITO NÃO FETAL" END TIPO\_OBITO,

DTOBITO AS DATA\_OBITO,

HORAOBITO AS HORA\_OBITO,

DTNASC AS DATA\_DE\_NASCIMENTO,

CASE

WHEN SEXO = "1" THEN "MASCULINO"

WHEN SEXO = "2" THEN "FEMININO"

ELSE "IGNORADO" END SEXO,

CASE

WHEN RACACOR = "1" THEN "Branca"

WHEN RACACOR = "2" THEN "Preta"

WHEN RACACOR = "3" THEN "Amarela"

WHEN RACACOR = "4" THEN "Parda"

WHEN RACACOR = "5" THEN "Indígena"

ELSE "Não Informado" END RACA,

CASE

WHEN ESTCIV = "1" THEN "SOLTEIRO"

WHEN ESTCIV = "2" THEN "CASADO"

WHEN ESTCIV = "3" THEN "VIÚVO"

WHEN ESTCIV = "4" THEN "SEPARADO JUDICIALMENTE"

ELSE "IGNORADO" END ESTADO\_CIVIL,

CASE

WHEN ESC = "1" THEN "NENHUMA"

WHEN ESC = "2" THEN "1 A 3 ANOS"

WHEN ESC = "3" THEN "4 A 7 ANOS"

WHEN ESC = "4" THEN "8 A 11 ANOS"

WHEN ESC = "5" THEN "12 E MAIS"

ELSE "IGNORADO" END ANOS\_DE\_ESCOLARIDADE,

CASE

WHEN LOCOCOR = "1" THEN "HOSPITAL"

WHEN LOCOCOR = "2" THEN "OUTRO ESTAB SAÚDE"

WHEN LOCOCOR = "3" THEN "DOMICÍLIO"

WHEN LOCOCOR = "4" THEN "VIA PÚBLICA"

WHEN LOCOCOR = "5" THEN "OUTROS"

ELSE "IGNORADO" END LOCAL\_DO\_OBITO,

CASE

WHEN CIRCOBITO = "1" THEN "ACIDENTE"

WHEN CIRCOBITO = "2" THEN "SUICÍDIO"

WHEN CIRCOBITO = "3" THEN "HOMICÍDIO"

WHEN CIRCOBITO = "4" THEN "OUTROS"

ELSE "IGNORADO" END TIPO\_DE\_ACIDENTE,

CASE

WHEN ACIDTRAB = "1" THEN "SIM"

WHEN ACIDTRAB = "2" THEN "NÃO"

ELSE "IGNORADO" END ACIDENTE\_DE\_TRABALHO,

CASE

WHEN TPPOS = "1" THEN "SIM"

ELSE "NÃO" END OBITO\_INVESTIGADO,

F.DESCR AS CAUSA\_DO\_OBITO,

CASE

WHEN IDADEMAE IS NULL THEN "NÃO INFORMADO"

ELSE IDADEMAE END IDADE\_MAE,

CASE

WHEN ESCMAE = "1" THEN "NENHUMA"

WHEN ESCMAE = "2" THEN "1 A 3 ANOS"

WHEN ESCMAE = "3" THEN "4 A 7 ANOS"

WHEN ESCMAE = "4" THEN "8 A 11 ANOS"

WHEN ESCMAE = "5" THEN "12 E MAIS"

ELSE "IGNORADO" END ANOS\_DE\_ESCOLARIDADE\_DA\_MAE,

CASE

WHEN OBITOGRAV = "1" THEN "SIM"

WHEN OBITOGRAV = "2" THEN "NÃO"

ELSE "IGNORADO" END MORTE\_DURANTE\_GRAVIDEZ,

B.DS\_NOMEPAD AS MUNICIPIO\_RESIDENCIA\_FALECIDO,

B.NU\_LATITUD AS MUNICIPIO\_RESIDENCIA\_FALECIDO\_LATITUDE,

B.NU\_LONGIT AS MUNICIPIO\_RESIDENCIA\_FALECIDO\_LONGITUDE,

D.DS\_NOMEPAD AS ESTADO\_RESIDENCIA\_FALECIDO,

C.DS\_NOMEPAD AS MUNICIPIO\_DE\_OCORRENCIA\_DO\_OBITO,

C.NU\_LATITUD AS MUNICIPIO\_DE\_OCORRENCIA\_DO\_OBITO\_LATITUDE,

C.NU\_LONGIT AS MUNICIPIO\_DE\_OCORRENCIA\_DO\_OBITO\_LONGITUDE,

E.DS\_NOMEPAD AS ESTADO\_DE\_OCORRENCIA\_DO\_OBITO

FROM unionall AS A

INNER JOIN TABELA\_MUNICIPIOv2 as B on A.CODMUNRES = B.COD\_MUNICIP

INNER JOIN TABELA\_MUNICIPIOv2 as C on A.CODMUNOCOR = C.COD\_MUNICIP

INNER JOIN TABELA\_UF as D on B.CO\_UF = D.CO\_UF

INNER JOIN TABELA\_UF as E on C.CO\_UF = E.CO\_UF

INNER JOIN CID10\_CAUSA\_DE\_OBITO AS F ON A.CAUSABAS = F.CID10

')

**RESULTADO DO COMANDO:**

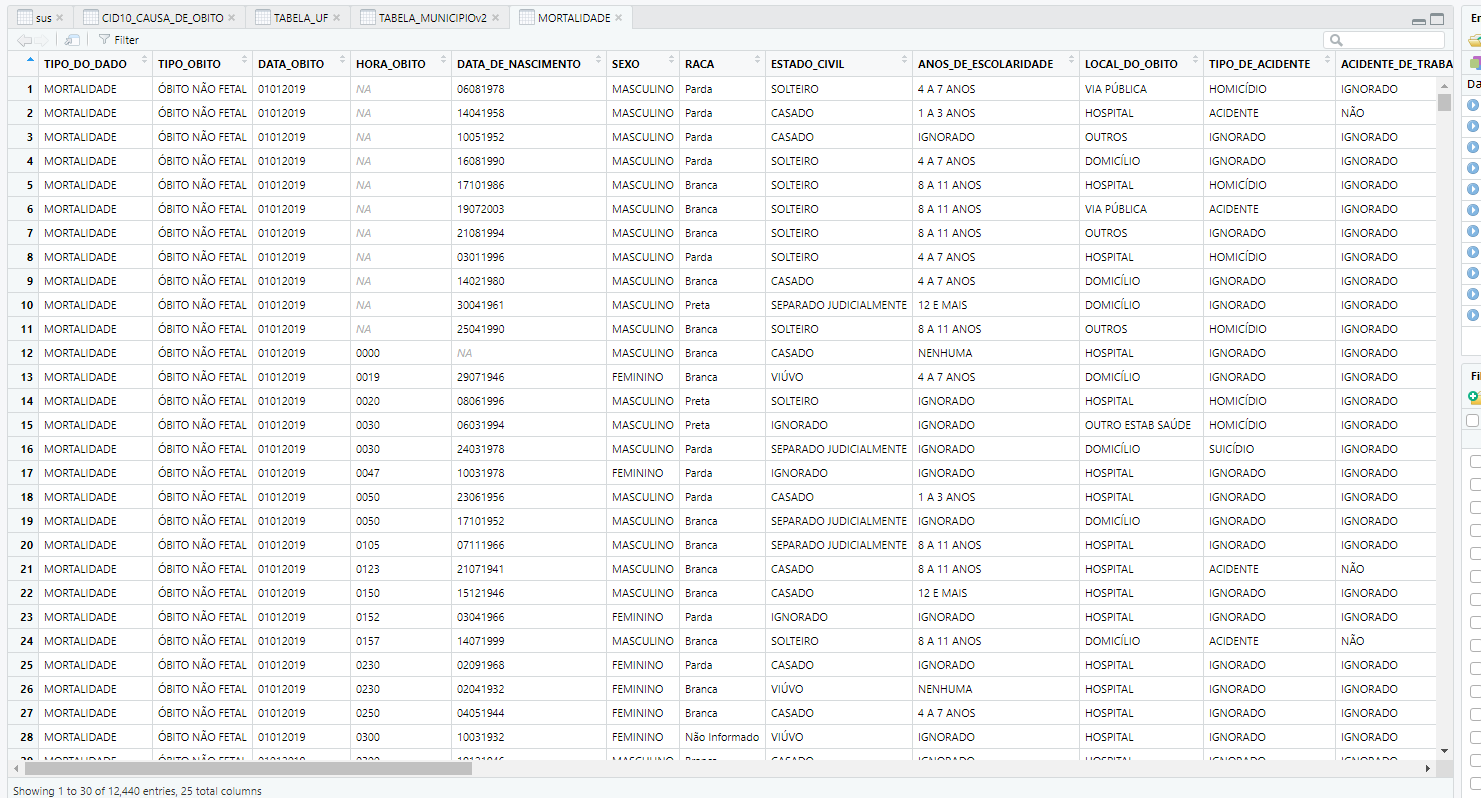
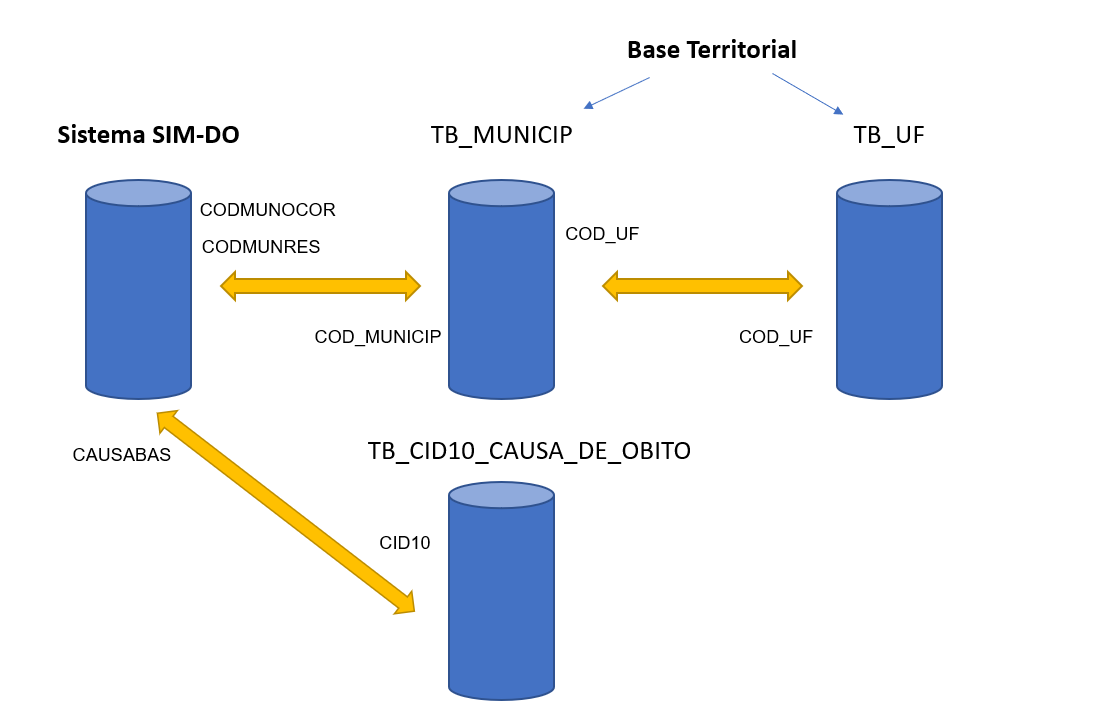


Imagem de relacionamento citado acima:



**Lista de Bases extraídas:**

|  |  |
| --- | --- |
| **Bases Extraídas - DATASUS** | |
| CIRURGIA\_BARIATRICA | AB - APAC de Acompanhamento a Cirurgia Bariátrica - A Partir de Jan/2008 até Mar/2013 |
| CIRURGIA\_QUIMIOTERAPIA | AQ - APAC de Quimioterapia - A Partir de Jan/2008 |
| CIRURGIA\_RADIOTERAPIA | AR - APAC de Radioterapia - A Partir de Jan/2008 |
| TB\_CONFECAO\_FISTULA\_ARTERIOVENOSA | ACF - APAC Confeção de Fístula Arteriovenosa - A Partir de Jun/2014 |
| TB\_LAUDOS\_DIVERSOS | AD - APAC de Laudos Diversos - A Partir de Jan/2008 |
| TB\_MEDICAMENTOS | AM - APAC de Medicamentos - A Partir de Jan/2008 |
| TABELA\_MUNICIPIO | Base Territorial |
| TB\_NEFROLOGIA | AN - APAC de Nefrologia - A Partir de Jan/2008 até Out/2014 |
| TB\_POS\_CIRURGIA\_BARIATRICA | ABO - APAC Acompanhamento Pós Cirurgia Bariátrica - A Partir de Abr/2013 |
| TB\_PSICOSSOCIAL | PS - Psicossocial - A Partir de Jan/2013 |
| TB\_TRATAMENTO\_DIALITICO | ATD - APAC Tratamento Dialítico - A Partir de Jun/2014 |
| TABELA\_UF | Base Territorial |
| Mortalidade | DO - Declarações de Óbito - 1979 a 2019 |
| CID10\_CAUSA\_DE\_OBITO | Causa do óbito |

Com isso temos os relacionamentos montados entre os sistemas citados e vamos partir para o Power BI para exibição desses indicadores.