**Pré-processamento Dados Datasus**

**Tutorial para fazer o download dos dados:**

Ir no site do datasus <https://datasus.saude.gov.br/>

Selecionar Tabwin ou:

<https://datasus.saude.gov.br/transferencia-de-arquivos/>

Selecionar:

Interface gráfica do usuário, Aplicativo, Word

Descrição gerada automaticamente

Se for baixar vários meses para agilizar pode selecionar o arquivo que arquivo que foi solicitado o download e copiar o link do arquivo:

<ftp://ftp.datasus.gov.br/dissemin/publicos/SIHSUS/200801_/Dados/RDPA2003.dbc>

Deletar a informação final e seguir o link que vai abrir a pasta q contém todos os arquivos desse tipo:

<ftp://ftp.datasus.gov.br/dissemin/publicos/SIHSUS/200801_/Dados/>

Colocar na janela de pesquisa: RDPA20\* para selecionar os dados de hospitalização do Pará para o ano de 2020. Selecione os meses.

Copie os arquivos de meses selecionados para uma pasta

Abra o programa TABWIN (Fazer o download do programa no site do datasus):

Interface gráfica do usuário, Texto, Aplicativo, Word

Descrição gerada automaticamente

Na aba de arquivos expandidos selecionar a pasta de destino

Selecionar na aba de arquivos comprimidos os arquivos com a extensão dbc que deseja expandir para dbf

Marcar a seta “Expande”

Verificar na pasta de destino se os arquivos expandidos já estão lá.

Interface gráfica do usuário, Tabela

Descrição gerada automaticamente

**Limpeza dos dados**

1. Limpar dados q não são internação de 2019

No banco de janeiro e seguintes de 2019, identificar variável dt\_saida. Selecionar o ano da data e excluir todas as linhas com ano de 2018.

1. Na banco de 2020 selecionar casos com ano de 2019 e agregar ao banco de 2019.
2. Juntar todos os meses em um banco só.
3. Calcular a idade correta a partir da variável data de nascimento: Selecionar o ano da data de nascimento e subtrair de 2019 para o cálculo da idade correta.
4. Filtrar as crianças com idade de 0 a 5 anos.
5. Esse será o banco inicial de 2019
6. Repetir o procedimento para o ano de 2010.