

# SQL para jornalismo

*2º Domingo de Dados - 15º Congresso Internacional de Jornalismo Investigativo*  
13/08/2020

# Olá!



## Adriana Letícia Reis

Uma apaixonada por Dados! Eles são a minha matéria prima e estão na minha vida desde 2004. Trabalho há 15 anos na Vivo, com projetos de BI. Quer saber mais sobre minha trajetória profissional. Meu LinkedIn é [linkedin.com/in/reis -al](https://www.linkedin.com/in/reis-al).



## Ana Carolina Moreno

Jornalista desde 2006, trabalhando com dados desde 2017. Atualmente sou produtora sênior na TV Globo. Você pode ver minhas reportagens em [www.facebook.com/ anacarolinamoreno](https://www.facebook.com/anacarolinamoreno).

“

*Sem dados, você é apenas  
mais uma pessoa com opinião*

—William Edwards Deming

# A linguagem SQL

- Structured Query Language, ou linguagem de consulta estruturada para bancos de dados relacionais
- SQL é uma linguagem completa de manipulação de dados.
- Com SQL é possível fazer todas as operações necessárias para gerenciar um banco de dados relacional.
- As operações são, criação, exclusão e alteração de banco de dados. Além de, buscar, alterar, excluir e incluir dados em tabelas.

# Por que **jornalista** começa a usar **SQL**?

- Grande motivo: **tem base que não abre no Excel**
- No início: DB-Browser → Excel
- Jornalistas com conhecimento mais avançados e em projetos mais autônomos já abandonaram o Excel e fazem tudo com programação. Mas essa provavelmente não é a realidade da maioria.
- **OBS:** Caso você queira manter grandes bases de dados na **Amazon (AWS)** ou no **Google (BigQuery)**, a linguagem usada também é SQL.

## Glossário básico





| TERMO                          | SIGNIFICADO   |
|--------------------------------|---|
| <b>VARIÁVEL</b>                | É a coluna da sua tabela. Cada coluna é uma variável.   |
| <b>OBSERVAÇÃO</b>              | É a linha da sua tabela. Cada linha é uma observação.   |
| <b>DICIONÁRIO DE VARIÁVEIS</b> | É um documento (em geral XLS ou PDF) com uma tabela mostrando o nome de cada variável e os detalhes dela. Também chamado de “dicionário de dados”.  |
| <b>QUERY</b>                   | É a sua busca. No SQL, para fazer uma “query” você escreve os comandos, selecionando suas variáveis e filtros de valores de cada linha, e depois clica em “executar/run” para rodar o comando e chegar ao resultado da busca. |

# Mãos à obra: baixando nossas bases

- **Link:** [http://download.inep.gov.br/microdados/microdados\\_educacao\\_superior\\_2018.zip](http://download.inep.gov.br/microdados/microdados_educacao_superior_2018.zip)
- Baixar e extrair o zip

microdados\_ed\_superior\_2018

Nome

-  leia-me
-  dados
-  filtros
-  anexos

- Na pasta **dados** estão as bases (arquivos grandes!)
- Na pasta **anexos** está o dicionário (arquivo pequeno)
- Sempre leia o **leia-me** antes de mexer em um base! Leva cinco minutos e te ajuda a não cometer erros básicos.
- **BÔNUS:** na pasta **filtros** o Inep te dá uma “cola” em SQL para fazer filtros seguindo os critérios oficiais

# Mãos à obra: baixando nossas bases

dados

| Nome                                | Data de modificaç... | Tipo                 | Tamanho      |
|-------------------------------------|----------------------|----------------------|--------------|
| DM_LOCAL_OFERTA.CSV                 | 16/08/2020 10:35     | Arquivo de Valore... | 95.558 KB    |
| MD5_microdados_ed_superior_2018.txt | 16/08/2020 10:35     | Documento de Te...   | 1 KB         |
| TB_AUX_CINE_BRASIL.CSV              | 16/08/2020 10:35     | Arquivo de Valore... | 45 KB        |
| DM_ALUNO.CSV                        | 16/08/2020 10:35     | Arquivo de Valore... | 2.946.027 KB |
| DM_CURSO.CSV                        | 16/08/2020 10:35     | Arquivo de Valore... | 8.779 KB     |
| DM_DOCENTE.CSV                      | 16/08/2020 10:35     | Arquivo de Valore... | 48.288 KB    |
| DM_IES.CSV                          | 16/08/2020 10:35     | Arquivo de Valore... | 676 KB       |

- Bases e variáveis que vamos usar na oficina:

| DM_ALUNO         | DM_CURSO    | DM_DOCENTE                  | DM_IES                    |
|------------------|-------------|-----------------------------|---------------------------|
| CO_UF_NASCIMENTO | NO_CURSO    | NU_ANO_CENSO                | IN_CAPITAL                |
|                  | IN_GRAUITO  | TP_CATEGORIA_ADMINISTRATIVA | VL_RECEITA_PROPRIA        |
|                  | IN_SITUACAO | TP_SITUACAO                 | VL_DESPESA_PESSOA_DOCENTE |
|                  | IN_CAPITAL  | TP_SEXO                     | VL_DESPESA_PESQUISA       |
|                  |             | TP_COR_RACA                 | TP_ORGANIZACAO_ACADEMICA  |
|                  |             |                             | SG_IES                    |



| DOCENTE (DM_DOCENTE) |                             |   |      |     |   |  |                                       |        |        |   |   |   |   |   |   |   |   |
|----------------------|-----------------------------|---|------|-----|---|--|---------------------------------------|--------|--------|---|---|---|---|---|---|---|---|
| OR<br>D              | NOME DA VARIÁVEL            | DESCRIÇÃO DA VARIÁVEL   | TIPO | TAM | DESCRIÇÃO DAS CATEGORIAS  | OBSERVAÇÃO   | Inclusão no ano<br>("s"=sim; "-"=não) |        |        |   |   |   |   |   |   |   |   |
|                      |                             |   |      |     |   |  | 0<br>9                                | 1<br>0 | 1<br>1 | 2 | 3 | 4 | 5 | 6 | 7 | 8 |   |
| DADOS DA IES         |                             |   |      |     |   |  |                                       |        |        |   |   |   |   |   |   |   |   |
| 1                    | NU_AND_CENSO                | Ano de referência do Censo Superior   | Num  | 4   |   | * Variável criada a partir do ano de referência do Censo   | s                                     | s      | s      | s | s | s | s | s | s | s | s |
| 2                    | CO_IES                      | Código único de identificação da IES  | Num  | 8   |   |  | s                                     | s      | s      | s | s | s | s | s | s | s | s |
| 3                    | TP_CATEGORIA_ADMINISTRATIVA | Tipo da Categoria Administrativa da IES   | Num  | 1   | 1. Pública Federal<br>2. Pública Estadual<br>3. Pública Municipal<br>4. Privada com fins lucrativos<br>5. Privada sem fins lucrativos<br>6. Privada - Particular em sentido estrito<br>7. Especial<br>8. Privada comunitária<br>9. Privada confessional | * As categorias 6, 8 e 9 somente no ano de 2009.<br>* Em 2012 foi criada a nova opção "7" Especial.  | s                                     | s      | s      | s | s | s | s | s | s | s | s |
| 4                    | TP_ORGANIZACAO_ACADEMICA    | Tipo da Organização Acadêmica da IES  | Num  | 1   | 1. Universidade<br>2. Centro Universitário<br>3. Faculdade<br>4. Instituto Federal de Educação, Ciência e Tecnologia<br>5. Centro Federal de Educação Tecnológica   |  | s                                     | s      | s      | s | s | s | s | s | s | s | s |
| DADOS DO DOCENTE     |                             |   |      |     |   |  |                                       |        |        |   |   |   |   |   |   |   |   |
| 5                    | CO_DOCENTE_IES              | Código gerado pelo Inep para o vínculo do docente à IES                               | Num  | 13  |   |  | s                                     | s      | s      | s | s | s | s | s | s | s | s |
| 6                    | ID_DOCENTE                  | Código de identificação gerado pelo Inep para o docente no Censo da Educação Superior | Char | 32  |   |  | s                                     | s      | s      | s | s | s | s | s | s | s | s |
| 7                    | TP_SITUACAO                 | Tipo da situação do docente na IES  | Num  | 1   | 1. Em exercício<br>2. Afastado para qualificação<br>3. Afastado para exercício em outros órgãos/entidades<br>4. Afastado por outros motivos<br>5. Afastado para tratamento de saúde<br>6. Falecido  | * Em 2010 a opção 5 é "Falecido" e nos demais anos posteriores é "Afastado para tratamento de saúde" | s                                     | s      | s      | s | s | s | s | s | s | s | s |
| 8                    | TP_ESCOLARIDADE             | Tipo do grau de escolaridade do docente   | Num  | 1   | 1. Sem graduação<br>2. Graduação<br>3. Especialização<br>4. Mestrado<br>5. Doutorado  |  | s                                     | s      | s      | s | s | s | s | s | s | s | s |



# O comando **Select**

- ❖ Com o **select \*** você busca todas as variáveis (colunas) da tabela. Após o **from** deve-se colocar o nome da tabela.

```
select * from nome_da_tabela
```

- ❖ Você pode selecionar somente algumas variáveis, para isso, é necessário separar cada variável por **vírgula**.

**Atenção:** antes do comando **from** **nunca** tem vírgula.

```
select variável_10, variável_5, variável_2  
from nome_da_tabela
```

- ❖ Você pode dar outro nome para a variável. Basta colocar **AS** após a variável.

```
select variável_10 AS novo_nome  
from nome_da_tabela
```

```
-- Este é um comentário de uma linha
```




```
/*  
    Este é um comentário  
    de várias linhas  
*/
```

```
SELECT * FROM DM_DOCENTE
```

```
SELECT NU_ANO_CENSO, TP_ORGANIZACAO_ACADEMICA  
FROM DM_CURSO
```

```
SELECT  
    UF AS CODIGO_UF,  
    NOME_UF AS UF  
FROM IBGE_RELATORIO_DTB_BRASIL_MUNICIPIO
```

# Os tipos de dados

- ❖ **VARCHAR**(n) - tamanho variável – máximo de n caracteres
- ❖ **CHAR**(n) - tamanho fixo - n caracteres
- ❖ **INTEGER** 
- ❖ **DECIMAL** [(precision, scale)] - precision é o número total de dígitos total e scale é o número de dígitos depois do ponto
- ❖ **DOUBLE PRECISION** | **FLOAT** | **REAL** 
- ❖ **DATE** | **TIME** | **TIMESTAMP**
- ❖ **BLOB** – Binary Large Object.  | **IMAGE**
- ❖ **BOOLEAN**

# Funções para tratamento das variáveis

- ❖ **COALESCE** quando a variável está com valor nulo e você não quer exibir como nulo, ou seja, você quer preencher com algum texto ou valor.

**select coalesce( nome\_da\_variavel, 'valor novo' ) from nome\_da\_tabela**

```
SELECT DISTINCT
    CO_UF_NASCIMENTO AS CO_UF_NASCIMENTO_sem_tratamento,
    COALESCE(CO_UF_NASCIMENTO, 'Sem UF') AS CODIGO_UF_NASCIMENTO_com_tratamento
FROM DM_ALUNO
```

| CO_UF_NASCIMENTO_sem_tratamento | CODIGO_UF_NASCIMENTO_com_tratamento |
|---------------------------------|-------------------------------------|
| NULL                            | Sem UF                              |
| 31                              | 31                                  |
| 51                              | 51                                  |
| 52                              | 52                                  |

- ❖ **LOWER** transforma os caracteres da variável para minúsculo.

**select lower( nome\_da\_variavel) from nome\_da\_tabela**

- ❖ **UPPER** transforma os caracteres da variável para maiúsculo.

**select upper( nome\_da\_variavel) from nome\_da\_tabela**

# Funções para tratamento das variáveis

- ❖ **DISTINCT** Exibe somente uma vez os valores repetidos. **Atenção**, o **distinct** deve ser utilizado sempre **após** o **select**.

## select distinct

nome\_da\_variável\_1, nome\_da\_variável\_2  
from nome\_da\_tabela

```
SELECT DISTINCT
    CODIGO_UF,
    NOME_UF AS DESCRICAO_UF
FROM IBGE_RELATORIO_DTB_BRASIL_MUNICIPIO
```

### sem distinct

| CODIGO_UF | DESCRICAO_UF |
|-----------|--------------|
| 11        | Rondônia     |
| 11        | Rondônia     |
| 11        | Rondônia     |
| 12        | Acre         |
| 12        | Acre         |
| 12        | Acre         |

### com distinct

| CODIGO_UF | DESCRICAO_UF |
|-----------|--------------|
| 11        | Rondônia     |
| 12        | Acre         |
| 13        | Amazonas     |
| 14        | Roraima      |
| 15        | Pará         |
| 16        | Amapá        |

- ❖ **SUBSTR** com a função substring, você seleciona partes do texto, podendo indicar a posição inicial do texto e a quantidade de caracteres.

select substr(nome\_da\_variável, posição\_inicial,  
quantidade\_caracteres)  
from nome\_da\_tabela

```
SELECT
    NO_CURSO,
    SUBSTR(NO_CURSO, 3, 4) AS SIGLA_CURSO
FROM DM_CURSO
```

| NO_CURSO            | SIGLA_CURSO |
|---------------------|-------------|
| DIREITO             | REIT        |
| CIÊNCIAS ECONÔMICAS | ÊNCI        |
| ENGENHARIA CIVIL    | GENH        |
| PEDAGOGIA           | DAGO        |
| SERVIÇO SOCIAL      | RVIÇ        |

# Funções para tratamento das variáveis

❖ **CASE** você pode transformar o dado no valor que deseja ver. A partir de uma condição, é possível transformar o valor da variável no momento da exibição (visualização).

**Atenção:** Todo **case** finaliza com **end**. Cada **condição** deve ser acompanhada de um **operador lógico**. O **else** não é obrigatório

**select**

**case**

**when** nome\_da\_variável\_1 = 'x' **then** 'aqui é um X'

**when** nome\_da\_variável\_1 = 'y' **then** 'aqui é um y'

**when** nome\_da\_variável\_1 = 'z' **and** nome\_da\_variável\_2 = 'zz' **then** 'aqui é um z com zz'

**else** 'o else não é obrigatório'

**end**

**from** nome\_da\_tabela

```
SELECT
    TP_SITUACAO,
    CASE
        WHEN TP_SITUACAO = 1 THEN 'Em exercício'
        WHEN TP_SITUACAO = 2 THEN 'Afastado para qualificação'
        WHEN TP_SITUACAO = 3 THEN 'Afastado para exercício em outros órgãos/entidades'
        WHEN TP_SITUACAO = 4 THEN 'Afastado por outros motivos'
        WHEN TP_SITUACAO = 5 THEN 'Afastado para tratamento de saúde'
        WHEN TP_SITUACAO = 6 THEN 'Falecido'
        WHEN TP_SITUACAO IS NULL THEN 'Sem preenchimento'
        ELSE 'Valor não previsto'
    END AS SITUACAO_DOCENTE
FROM DM_DOCENTE
```

| TP_SITUACAO | SITUACAO_DOCENTE                             |
|-------------|--|
| 1           | Em exercício                                 |
| 3           | Afastado para exercício em outros órgãos/... |
| 2           | Afastado para qualificação                   |
| 4           | Afastado por outros motivos                  |
| 5           | Afastado para tratamento de saúde            |

# Os operadores lógicos

- ❖ **Comparação:** =, <>, <, >, <=, >=
- ❖ **Lógicos:** and, or
- ❖ **Range da valores:** between
- ❖ **Conjuntos:** in, not in
- ❖ **Match parcial de texto:** like, not like
- ❖ **Nulabilidade:** is null, is not null

# Os comandos **Where** e **Order By**

- ❖ O **where** funciona como um filtro, ou seja, você escolhe quais observações (linhas) deseja visualizar.
- ❖ Ele é sempre utilizado **após** o nome da tabela
- ❖ Quando você deseja aplicar mais de um filtro, deve-se acrescentar os **operadores lógicos** e depois o nome da variável (coluna)

**select \* from** nome\_da\_tabela **where** nome\_da\_variável = 'x'

```
SELECT
    NO_CURSO
    , TP_NIVEL_ACADEMICO
FROM DM_CURSO
WHERE IN_GRATUITO = 1
```

```
SELECT * FROM DM_IES WHERE IN_CAPITAL <> 0

SELECT * FROM DM_IES WHERE IN_CAPITAL != 0
```

- ❖ O **order by** serve para ordenar por uma ou mais variável, e ele é sempre utilizado **no final** de todos os comando

**select \* from** nome\_da\_tabela **order by** nome\_da\_variável **DESC**

```
SELECT NO_CURSO, IN_GRATUITO FROM DM_CURSO
WHERE IN_CAPITAL = 1 AND TP_SITUACAO = 1
ORDER BY NO_CURSO ASC, IN_GRATUITO DESC
```

```
SELECT NO_CURSO, IN_GRATUITO FROM DM_CURSO
WHERE IN_CAPITAL = 1 AND TP_SITUACAO = 1
ORDER BY 1 ASC, 2 DESC
```



# O operador Like

- ❖ Com o **like** você pode buscar um caracter ou uma sequência de caracteres em uma variável. Esta busca pode ser no início, meio ou fim da variável.
- ❖ O **único** operador **aceito** pelo like é o **%**. E ele tem que estar **dentro** de “
- ❖ **% no início** a busca é feita por tudo que **termina com**.
- ❖ **% no final** a busca é feita por tudo que **inicia com**.

## contém

```
SELECT
    NO_CURSO,
    REPLACE(NO_CURSO, 'Ç', 'C') AS NOME_CURSO
from DM_CURSO
WHERE NO_CURSO LIKE '%Ç%'
```

## inicia com

```
SELECT
    DISTINCT NO_CURSO
FROM DM_CURSO
WHERE NO_CURSO LIKE 'ADMINISTRAÇÃO%'
ORDER BY 1
```

## termina com

```
SELECT
    DISTINCT NO_CURSO
FROM DM_CURSO
WHERE NO_CURSO LIKE '%SOCIAL'
ORDER BY 1
```

# Funções de agrupamento Group By

❖ São funções para fazer **contas** e **agrupar** os dados (observações).

❖ **Count**: conta quantas observações (linhas/registros).

```
SELECT COUNT(*) FROM DM_DOCENTE
```

```
SELECT
    TP_SEXO,
    COUNT(*) AS QTDE_DOCENTES
FROM DM_DOCENTE
GROUP BY TP_SEXO
```

❖ **Sum**: soma valores

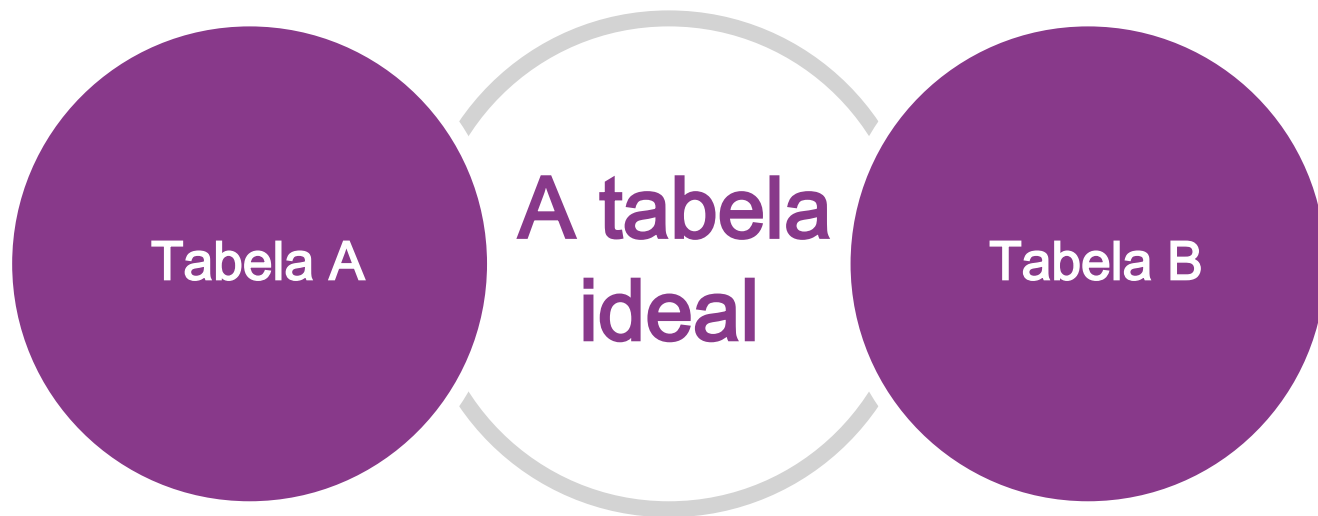
❖ **Max**: Valor máximo

❖ **Min**: Valor mínimo

❖ **AVG**: média

```
SELECT
    SG_IES,
    CASE
        WHEN TP_CATEGORIA_ADMINISTRATIVA IN (1,2,3) THEN 'Pública'
        WHEN TP_CATEGORIA_ADMINISTRATIVA IN (4,5,6,8,9) THEN 'Privada'
        WHEN TP_CATEGORIA_ADMINISTRATIVA = 7 THEN 'Especial'
    END AS CATEGORIA_ADMINISTRATIVA,
    SUM(CAST(VL_RECEITA_PROPRIA AS FLOAT)) AS VL_RECEITA_PROPRIA,
    SUM(CAST(VL_DESPESA_PESSOAL_DOCENTE AS FLOAT)) AS VL_DESPESA_PESSOAL_DOCENTE,
    SUM(CAST(VL_DESPESA_PESQUISA AS FLOAT)) AS VL_DESPESA_PESQUISA
FROM DM_IES
WHERE TP_ORGANIZACAO_ACADEMICA = 1
AND SG_IES IS NOT NULL
GROUP BY SG_IES, CATEGORIA_ADMINISTRATIVA
```

# Os comandos **Join**

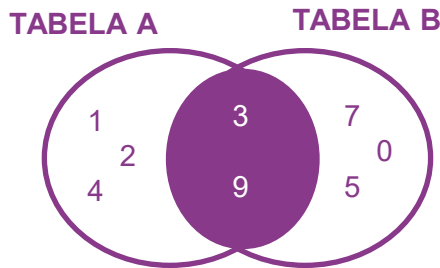


# O comando **INNER JOIN**

## TABELA A

Planilha DM\_DOCENTES

Tem dados de cada docente, porém, a UF e MUNICIPIO que nasceu o docente estão em códigos do IBGE. Para descobrir a UF e o nome do município, precisamos “juntar-conectar” com a planilha IBGE\_RELATORIO\_DTB\_B RASIL\_MUNICIPIO do IBGE.



```
select *  
from tabela_ as A  
inner join tabela_b as B  
on A.variavel_chave = B.variavel_chave
```

## TABELA B

Planilha  
IBGE\_RELATORIO\_DTB\_B  
RASIL\_MUNICIPIO

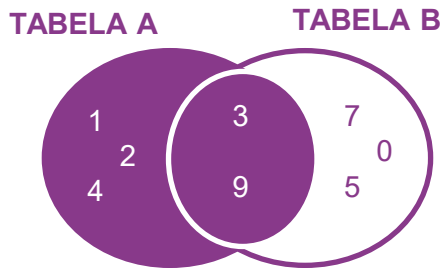
Tem os códigos dos municípios e UFs, bem como o nome de cada município e a sigla da UF.

# O comando **LEFT JOIN**

## TABELA A

Planilha DM\_DOCENTES

Tem dados de cada docente, porém, a UF e MUNICIPIO que nasceu o docente estão em códigos do IBGE. Para descobrir a UF e o nome do município, precisamos “juntar-conectar” com a planilha IBGE\_RELATORIO\_DTB\_BRASIL\_MUNICIPIO do IBGE.



```
select *  
from tabela_a as A  
left join tabela_b as B  
on A.variavel_chave = B.variavel_chave
```

## TABELA B











Planilha  
IBGE\_RELATORIO\_DTB\_B  
RASIL\_MUNICIPIO









Tem os códigos dos municípios e UFs, bem como o nome de cada município e a sigla da UF






# Uma forma **técnica** de visualizar o **relacionamento** das tabelas



















Software: DB Designer

| LOCAL_OFERTA  |                 |              |   |
|---|-----------------|--------------|---|
|            | NU_ANO_CENSO    | integer      |  |
|   | NO_LOCAL_OFERTA | varchar(200) |  |
|            | CO_LOCAL_OFERTA | integer(8)   |  |
|            | CO_CURSO        | integer(8)   |  |
|            | CO_CURSO_POLO   | integer(8)   |  |
|  Add field |                 |              |   |

| CURSO   |                 |              |   |
|---|-----------------|--------------|---|
|            | CO_CURSO        | integer      |  |
|   | NO_CURSO        | varchar(255) |  |
|   | CO_IES          | integer(8)   |  |
|   | NU_ANO_CENSO    | integer(8)   |  |
|   | CO_LOCAL_OFERTA | integer(8)   |  |
|   | CO_IES          | integer(8)   |  |
|  Add field |                 |              |   |

| IES   |              |              |   |
|---|--------------|--------------|---|
|            | CO_IES       | integer      |  |
|   | NO_IES       | varchar(200) |  |
|   | NU_ANO_CENSO | integer(8)   |  |
|  Add field |              |              |   |

| ALUNO   |               |            |   |
|---|---------------|------------|---|
|            | ID_ALUNO      | integer    |  |
|   | CO_CURSO      | integer    |  |
|   | CO_IES        | integer    |  |
|   | TP_COR_RACA   | integer    |  |
|   | TP_SEXO       | text       |  |
|   | IN_MATRICULA  | integer(1) |  |
|   | NU_ANO_CENSO  | integer(4) |  |
|   | CO_CURSO_POLO | integer(8) |  |
|  Add field |               |            |   |

| DOCENTE   |                             |            |   |
|---|-----------------------------|------------|---|
|              | CO_DOCENTE_IES              | integer(8) |  |
|   | CO_IES                      | integer(8) |  |
|   | TP_CATEGORIA_ADMINISTRATIVA | integer(1) |  |
|   | NU_ANO_CENSO                | integer(8) |  |
|  Add field |                             |            |   |

## Exercício prático: perfil racial - docentes

- **ALERTA:** As bases também têm seus limites para representar a realidade de fato
- No caso racial, um dos problemas é a subnotificação: a alta porcentagem de pessoas que não quiseram declarar a cor ou raça, ou de linhas em que esse campo ficou em branco.
- Recomendação: faça a reportagem, mas inclua a ressalva de que os instrumentos existentes pra buscar esses dados são limitados.
- Proposta: atualizar os dados dessa reportagem:  
<https://g1.globo.com/educacao/guia-de-carreiras/noticia/2018/11/20/negros-representam- apenas-16-dos-professores-universitarios.ghtml>

# Exercício prático: perfil racial - docentes

- **Proposta:** atualizar os dados dessa reportagem:  
<https://g1.globo.com/educacao/guia-de-carreiras/noticia/2018/11/20/negros-representam-apenas-16-dos-professores-universitarios.ghtml>



## Negros representam apenas 16% dos professores universitários

No Dia da Consciência Negra, levantamento do G1 mostra ainda que, quanto maior o grau de escolaridade do professor, maior a desigualdade racial.

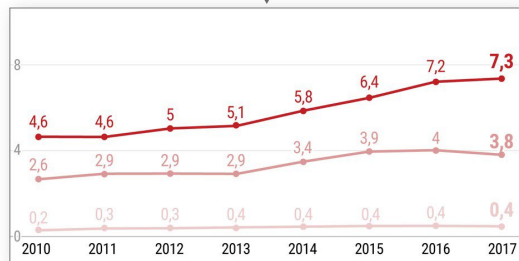
Por Ana Carolina Moreno, G1  
20/11/2018 10h44 - Atualizado há um ano



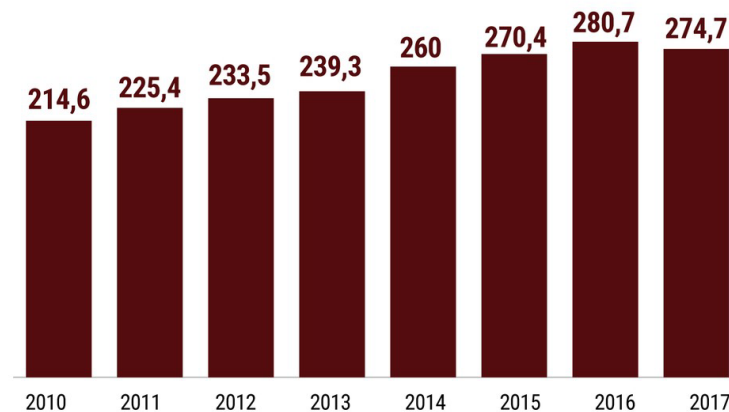


## Professores que declararam cor ou raça

Entre 2010 e 2017, n° de docentes negros cresceu 47,7%



## Total de professores



Fonte: : Inep/microdados do Censo da Educação Superior



Infográfico elaborado em: 18/11/2018

# Exercício prático: perfil racial - docentes

- Abaixo estão os dados de 2010 a 2017.
- Como atualizamos a tabela para incluir os dados de 2018 ?

| <b>DOCENTES POR<br/>COR/RAÇA</b>                       | <b>2010</b>   | <b>%</b>   | <b>2011</b>   | <b>%</b>   | <b>2012</b>   | <b>%</b>   | <b>2013</b>   | <b>%</b>   | <b>2014</b>   | <b>%</b>   | <b>2015</b>   | <b>%</b>   | <b>2016</b>   | <b>%</b>   | <b>2017</b>   | <b>%</b>   |
|--|---------------|------------|---------------|------------|---------------|------------|---------------|------------|---------------|------------|---------------|------------|---------------|------------|---------------|------------|
| <b>"Não quis declarar"<br/>+<br/>"Sem informações"</b> | 152245        | 41,5       | 152794        | 40,4       | 145438        | 38,4       | 144359        | 37,6       | 136587        | 34,4       | 130851        | 32,6       | 116841        | 29,4       | 117242        | 29,9       |
| <b>1- Branca</b>                                       | 169561        | 46,2       | 175903        | 46,5       | 180052        | 47,5       | 181420        | 47,3       | 195881        | 49,4       | 202401        | 50,4       | 213196        | 53,6       | 208280        | 53,1       |
| <b>2- Preta</b>  | 4647          | 1,3        | 4636          | 1,2        | 5035          | 1,3        | 5154          | 1,3        | 5858          | 1,5        | 6460          | 1,6        | 7212          | 1,8        | 7361          | 1,9        |
| <b>3- Parda</b>  | 37488         | 10,2       | 41639         | 11,0       | 45110         | 11,9       | 49418         | 12,9       | 54336         | 13,7       | 57150         | 14,2       | 55856         | 14,1       | 54878         | 14         |
| <b>4- Amarela</b>                                      | 2662          | 0,7        | 2920          | 0,8        | 2927          | 0,8        | 2915          | 0,8        | 3483          | 0,9        | 3954          | 1,0        | 4019          | 1,0        | 3805          | 1          |
| <b>5- Indígena</b>                                     | 279           | 0,1        | 365           | 0,1        | 377           | 0,1        | 417           | 0,1        | 450           | 0,1        | 483           | 0,1        | 487           | 0,1        | 470           | 0,1        |
| <b>Total</b>   | <b>366882</b> | <b>100</b> | <b>378257</b> | <b>100</b> | <b>378939</b> | <b>100</b> | <b>383683</b> | <b>100</b> | <b>396595</b> | <b>100</b> | <b>401299</b> | <b>100</b> | <b>397611</b> | <b>100</b> | <b>392036</b> | <b>100</b> |

## Exercício prático: o que diz o dicionário

- Base: DM\_DOCENTE
- Variável: TP\_COR\_RACA
- Correspondência no dicionário:

|    |             |                             |     |   |   |
|----|-------------|-----------------------------|-----|---|---|
| 15 | TP_COR_RACA | Tipo da cor/raça do docente | Num | 1 | 0. Docente não quis declarar cor/raça<br>1. Branca<br>2. Preta<br>3. Parda<br>4. Amarela<br>5. Indígena<br>9. Não dispõe da informação (Não resposta) |
|----|-------------|-----------------------------|-----|---|---|

## Exercício prático: resposta simples

- Código simples:

```
SELECT TP_COR_RACA, count(TP_COR_RACA)
FROM DM_DOCENTE
GROUP BY TP_COR_RACA
```

- Resultado:

|   | TP_COR_RACA | count(TP_COR_RACA) |
|---|-------------|--------------------|
| 1 | 0           | 116943             |
| 2 | 1           | 210430             |
| 3 | 2           | 8020               |
| 4 | 3           | 57229              |
| 5 | 4           | 3883               |
| 6 | 5           | 499                |
| 7 | 6           | 889                |

# Exercício prático: resposta mais avançada

## ▪ Código mais elaborado:

```
SELECT
  TP_COR_RACA,
  CASE
    WHEN TP_COR_RACA= 0 THEN '0- Docente não quis declarar'
    WHEN TP_COR_RACA= 1 THEN '1- Branca'
    WHEN TP_COR_RACA= 2 THEN '2- Preta'
    WHEN TP_COR_RACA= 3 THEN '3- Parda'
    WHEN TP_COR_RACA= 4 THEN '4- Amarela'
    WHEN TP_COR_RACA= 5 THEN '5- Indígena'
    WHEN TP_COR_RACA= 9 THEN '9- Não foi informado'
  ELSE 'Código não identificado'
  END AS COR_RACA_DOCENTE_2018,
  COUNT(*) AS QTDE_DOCENTES
FROM DM_DOCENTE
GROUP BY TP_COR_RACA
```

## ▪ Resultado:

|   | TP_COR_RACA | COR_RACA_DOCENTE_2018        | QTDE_DOCENTES |
|---|-------------|------------------------------|---------------|
| 1 | 0           | 0- Docente não quis declarar | 116943        |
| 2 | 1           | 1- Branca                    | 210430        |
| 3 | 2           | 2- Preta                     | 8020          |
| 4 | 3           | 3- Parda                     | 57229         |
| 5 | 4           | 4- Amarela                   | 3883          |
| 6 | 5           | 5- Indígena                  | 499           |
| 7 | 6           | Código não identificado      | 889           |

**EXPECTATIVA X REALIDADE**

# Exercício prático: resposta simples

| <b>DOCENTES POR<br/>COR/RAÇA</b>                    | <b>2010</b>   | <b>%</b>   | <b>2011</b>   | <b>%</b>   | <b>2012</b>   | <b>%</b>   | <b>2013</b>   | <b>%</b>   | <b>2014</b>   | <b>%</b>   | <b>2015</b>   | <b>%</b>   | <b>2016</b>   | <b>%</b>   | <b>2017</b>   | <b>%</b>   | <b>2018</b>   | <b>%</b>    |
|---|---------------|------------|---------------|------------|---------------|------------|---------------|------------|---------------|------------|---------------|------------|---------------|------------|---------------|------------|---------------|-------------|
| <b>0- Não quis declarar +<br/>"Sem informações"</b> | 152245        | 41,5       | 152794        | 40,4       | 145438        | 38,4       | 144359        | 37,6       | 136587        | 34,4       | 130851        | 32,6       | 116841        | 29,4       | 117242        | 29,9       | <b>117832</b> | <b>29,6</b> |
| <b>1- Branca</b>                                    | 169561        | 46,2       | 175903        | 46,5       | 180052        | 47,5       | 181420        | 47,3       | 195881        | 49,4       | 202401        | 50,4       | 213196        | 53,6       | 208280        | 53,1       | <b>210430</b> | <b>52,9</b> |
| <b>2- Preta</b>                                     | 4647          | 1,3        | 4636          | 1,2        | 5035          | 1,3        | 5154          | 1,3        | 5858          | 1,5        | 6460          | 1,6        | 7212          | 1,8        | 7361          | 1,9        | <b>8020</b>   | <b>2,0</b>  |
| <b>3- Parda</b>                                     | 37488         | 10,2       | 41639         | 11,0       | 45110         | 11,9       | 49418         | 12,9       | 54336         | 13,7       | 57150         | 14,2       | 55856         | 14,1       | 54878         | 14         | <b>57229</b>  | <b>14,4</b> |
| <b>4- Amarela</b>                                   | 2662          | 0,7        | 2920          | 0,8        | 2927          | 0,8        | 2915          | 0,8        | 3483          | 0,9        | 3954          | 1,0        | 4019          | 1,0        | 3805          | 1          | <b>3883</b>   | <b>1,0</b>  |
| <b>5- Indígena</b>                                  | 279           | 0,1        | 365           | 0,1        | 377           | 0,1        | 417           | 0,1        | 450           | 0,1        | 483           | 0,1        | 487           | 0,1        | 470           | 0,1        | <b>499</b>    | <b>0,1</b>  |
| <b>Total</b>  | <b>366882</b> | <b>100</b> | <b>378257</b> | <b>100</b> | <b>378939</b> | <b>100</b> | <b>383683</b> | <b>100</b> | <b>396595</b> | <b>100</b> | <b>401299</b> | <b>100</b> | <b>397611</b> | <b>100</b> | <b>392036</b> | <b>100</b> | <b>397893</b> | <b>100</b>  |

- Link para a planilha: [https://docs.google.com/spreadsheets/d/1\\_-QhFwGnl8hfYXEIdD4rASev12LXIOI9juqJCipEGaiU/edit?usp=sharing](https://docs.google.com/spreadsheets/d/1_-QhFwGnl8hfYXEIdD4rASev12LXIOI9juqJCipEGaiU/edit?usp=sharing)
- **Dica de pauta:** em breve sai o Censo da Educação Superior de 2019 ;-)

# Links úteis

- ❖ **SQLite:** <https://www.sqlite.org/index.html>
- ❖ **Download do SQLite** (Procurar pelo seu sistema operacional):  
<https://www.sqlite.org/download.html>
- ❖ **Instalar o SQLite:** [https://www.youtube.com/watch?v=gsZmB\\_MMKyc](https://www.youtube.com/watch?v=gsZmB_MMKyc)
- ❖ **DB Browser:** <https://sqlitebrowser.org/>
- ❖ **Funções SQL utilizadas no DB Browser:** [https://www.sqlite.org/lang\\_corefunc.html](https://www.sqlite.org/lang_corefunc.html)
- ❖ **DB Designer** (Procurar pelo seu sistema operacional): <https://www.dbdesigner.net/>