

Análise dos dados de vacinação da COVID-19 e seus estabelecimentos de saúde nas regiões norte e sudeste do Brasil

Disciplina: Tópicos Especiais S.I. 1 (MAI716)

Apresentação do Trabalho Final

Professores: Sérgio Serra e Jorge Zavaleta

Grupo: Alessandra Saraiva, Fabiane Carvalho e Massao Iwanaga

Índice

1. Introdução
2. Descrição do problema
3. Datasets
4. Dicionários de dados
5. Análise e seleção dos dados
6. Tratamento dos dados / data cleaning
7. Análises
8. Proveniência
9. Reprodutibilidade

1 - Introdução



Saúde

Desigualdade pesa mais que faixa etária nas mortes pela covid-19 no Brasil, diz estudo

Levantamento foi feito por pesquisadores brasileiros ligados ao Instituto de Estudos para Políticas de Saúde (IEPS) e foi publicado na revista científica Lancet

Roberta Jansen, O Estado de S.Paulo

13 de abril de 2021 | 00h01

...

Comparando os índices de vulnerabilidade sócioeconômica dos Estados aos registros de casos e mortes por covid, o trabalho mostra que o impacto foi maior nas regiões mais pobres, com menos recursos médicos, como Norte e Nordeste.

<https://saude.estadão.com.br/noticias/geral/desigualdade-pesou-mais-que-faixa-etaria-nas-mortes-por-covid-no-brasil-diz-estudo.70003679379#:~:text=RIO%20%2D%20As%20desigualdades%20sociais%20e,em%20diversos%20pa%C3%ADses%20da%20Europa.>



2 - Descrição do problema

Devido a recente campanha de vacinação contra a COVID-19, fazem-se necessárias análises para avaliar o perfil dos vacinados e dos estabelecimentos de saúde nas diferentes regiões do país devido às suas desigualdades. Por isso, serão analisadas a região norte, com aproximadamente 18,7 milhões de habitantes, pois foi muito afetada nos últimos meses por nova variante do coronavírus e também a região sudeste, por ser a mais populosa do país, com sua população em torno de 89 milhões de habitantes.


3- Datasets Escolhidos

 Campanha Nacional de Vacinação contra Covid-19

<https://opendatasus.saude.gov.br/dataset/covid-19-vacao>

Este dataset possui informações sobre os vacinados (idade, data de nascimento, sexo, raça e etc.), os estabelecimentos de saúde (nome, município, estado e etc.) onde ocorreu a vacinação, grupo de atendimento ao qual o vacinado pertence e da vacina (fabricante, lote e etc.)

Tamanho: 4.2 GB (7.958.949 registros)

 Cadastro Nacional de Estabelecimentos de Saúde (CNES) - Ativo

https://dados.gov.br/dataset/cnes_ativo

Este dataset possui informações (código, nome, endereço, tipo e etc.) dos estabelecimentos de saúde do país, públicos ou não.

Tamanho: 57 MB (354.804 registros)

4 - Dicionário de Dados

+ Campanha Nacional de Vacinação contra Covid-19 (recorte)

Ordem	Descrição	Categoria	Categoria
1	document_id	Identificador do documento	
2	paciente_id	Identificador do vacinado	
3	paciente_idade	Idade do vacinado	
4	paciente_dataNascimento	Data de nascimento do vacinado	
5	paciente_enumSexoBiologico	Sexo do vacinado	M: masculino; F: feminino
6	paciente_racaCor_codigo	Código da raça/cor do vacinado	1; 2; 3; 4; 99
7	paciente_racaCor_valor	Descrição da raça/cor do vacinado	1: branca; 2: preta; 3: parda; 4: amarela; 99: sem informação
8	paciente_endereco_colbgeMunicipio	Código IBGE do município de endereço do vacinado	
9	paciente_endereco_coPais	Código do país de endereço do vacinado	
10	paciente_endereco_nmMunicipio	Nome do município de endereço do vacinado	
11	paciente_endereco_nmPais	Nome do país de endereço do vacinado	
12	paciente_endereco_uf	Sigla da UF de endereço do vacinado	
13	paciente_endereco_cep	5 dígitos para anonimizado e 7 dígitos para identificado	
14	paciente_nacionalidade_enumNacionalidade	Nacionalidade do vacinado	

4 - Dicionário de Dados

✚ Cadastro Nacional de Estabelecimentos de Saúde (CNES) - Ativo (criado)

Ordem	Campo	Descrição
1	co_cnes	Código CNES
2	co_ibge	Código IBGE
3	no_fantasia	Nome Fantasia
4	ds_tipo_unidade	Tipo de estabelecimento de saúde: Posto de Saúde, Centro de Saúde/Unidade Básica de Saúde, Policlínica, Hospital Geral, Hospital Especializado, Unidade Mista, Pronto Socorro Geral, Pronto Socorro Especializado, Consultório Isolado, Unidade Móvel Fluvial, Clínica Especializada/Amb. Especializado, Unidade de Serviço de Apoio de Diagnose e Terapia, Unidade Móvel Terrestre, Unidade Móvel de Nível Pré-hospitalar na Área de Urgência e Emergência: Farmácia: Unidade de Vigilância em Saúde, Cooperativa, Centro de Parto Normal Isolado, Hospital /Dia- Isolado, Central de Regulação de Serviços de Saúde, Laboratório Central de Saúde Pública – LACEN, Secretaria de Saúde
5	tp_gestao	Tipo de gestão: Estadual (E), Municipal (M) ou Dupla (D)
6	no_logradouro	Nome do Logradouro
7	nu_endereco	Número do Endereço
8	no_bairro	Bairro
9	co_cep	CEP
10	uf	UF do estabelecimento
11	municipio	Município
12	nu_telefone	Número de Telefone

5 - Análise e seleção dos dados

**Campos
desconsiderados do
dataset da campanha
nacional de vacinação
contra Covid-19**

13	paciente_endereco_cep	5 dígitos para anonimizado e 7 dígitos para identificado
14	paciente_nacionalidade_enumNacionalidade	Nacionalidade do vacinado
15	estabelecimento_valor	Código do CNES do estabelecimento que realizou a vacinação
16	estabelecimento_razaoSocial	Nome/Razão Social do estabelecimento
17	estabelecimento_noFantasia	Nome fantasia do estabelecimento
18	estabelecimento_municipio_codigo	Código do município do estabelecimento
19	estabelecimento_municipio_nome	Nome do município do estabelecimento
20	estabelecimento_uf	Sigla da UF do estabelecimento
21	vacina_grupoAtendimento_codigo	Código do grupo de atendimento ao qual pertence o vacinado
22	vacina_grupoAtendimento_nome	Nome do grupo de atendimento ao qual pertence o vacinado
23	vacina_categoria_codigo	Código da categoria
24	vacina_categoria_nome	Descrição da categoria
25	vacina_lote	Número do lote da vacina
26	vacina_fabricante_nome	Nome do fabricante/fornecedor
27	vacina_fabricante_referencia	CNPJ do fabricante/fornecedor
28	vacina_dataAplicacao	Data de aplicação da vacina
29	vacina_descricao_dose	Descrição da dose
30	vacina_codigo	Código da vacina
31	vacina_nome	Nome da vacina/produto
32	sistema_origem	Nome do sistema de origem

5 - Análise e seleção dos dados

**Campos
desconsiderados do
dataset de cadastro
nacional de
estabelecimentos de
saúde (CNES) - Ativo**

Ordem	Campo	Descrição
1	co_cnes	Código CNES
2	co_ibge	Código IBGE
3	no_fantasia	Nome Fantasia
4	ds_tipo_unidade	Tipo de estabelecimento de saúde: Posto de Saúde, Centro de Saúde/Unidade Básica de Saúde, Policlínica, Hospital Geral, Hospital Especializado, Unidade Mista, Pronto Socorro Geral, Pronto Socorro Especializado, Consultório Isolado, Unidade Móvel Fluvial, Clínica Especializada/Amb. Especializado, Unidade de Serviço de Apoio de Diagnose e Terapia, Unidade Móvel Terrestre, Unidade Móvel de Nível Pré-hospitalar na Área de Urgência e Emergência: Farmácia: Unidade de Vigilância em Saúde, Cooperativa, Centro de Parto Normal Isolado, Hospital /Dia- Isolado, Central de Regulação de Serviços de Saúde, Laboratório Central de Saúde Pública – LACEN, Secretaria de Saúde
5	tp_gestao	Tipo de gestão: Estadual (E), Municipal (M) ou Dupla (D)
6	no_logradouro	Nome do Logradouro
7	nu_endereco	Número do Endereço
8	no_bairro	Bairro
9	co_cep	CEP
10	uf	UF do estabelecimento
11	municipio	Município
12	nu_telefone	Número de Telefone

5 - Análise e seleção dos dados

Estados da região norte do Brasil

Amazonas (AM)	Acre (AC)	Amapá (AP)
Pará (PA)	Rondônia (RO)	Roraima (RR)
Tocantins (TO)		

Estados da região sudeste do Brasil

Espírito Santo (ES)	Minas Gerais (MG)
Rio de Janeiro (RJ)	São Paulo (SP)

```
UFs = ['RJ', 'SP', 'MG', 'ES', 'AM', 'AC', 'AP', 'PA', 'RO', 'RR', 'TO']
Norte = ['AM', 'AC', 'AP', 'PA', 'RO', 'RR', 'TO']
Sudeste = ['RJ', 'SP', 'MG', 'ES']
# Eliminando os estados que não são da região norte nem sudeste
df_vacinacao = df_vacinacao.loc[(df_vacinacao['estabelecimento_uf'].isin(UFs))]
```

6 - Tratamento dos dados / data cleaning (problemas encontrados)

- Pacientes com mais de 2 registros (ex: 1ª e 2º dose de um fabricante e 1ª e 2ª dose de outro fabricante);
- Pacientes com idade abaixo de 18 anos e acima de 120 anos;
- Pacientes com sexo com valor “I”;

6 - Tratamento dos dados (linhas duplicadas)

- Manter a primeira ocorrência (menor data) e eliminar demais linhas

```
# Mantém a primeira ocorrência e elimina as demais linhas de mesmo paciente_id
# O parâmetro inplace=True persiste o resultado da operação ao dataframe "df_vacinacao"
df_vacinacao.drop_duplicates(subset=['paciente_id'], keep='first', inplace=True)

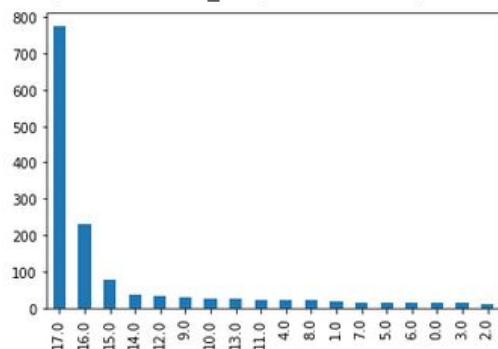
# O resultado esperado do comando abaixo é a ausência de valores "True", já que todos serão "False" (não duplicados)
df_vacinacao.duplicated(subset=['paciente_id']).value_counts()

False      3497840
dtype: int64
```

6 - Tratamento dos dados (pacientes com idade abaixo de 18 anos e acima de 120 anos)

```
df_idade_menor['paciente_idade'].value_counts().plot.bar()
```

```
<matplotlib.axes._subplots.AxesSubplot at 0x7f38a1d3ad90>
```



```
df_idade_menor['paciente_idade'].value_counts()
```

17.0	1665
16.0	445
15.0	147
14.0	81
12.0	61
13.0	57
9.0	51
10.0	50
4.0	43
11.0	42
7.0	41
8.0	37
6.0	35
5.0	34
0.0	30
1.0	29
3.0	28
2.0	21

```
Name: paciente_idade, dtype: int64
```

```
df_idade_maior['paciente_idade'].value_counts()
```

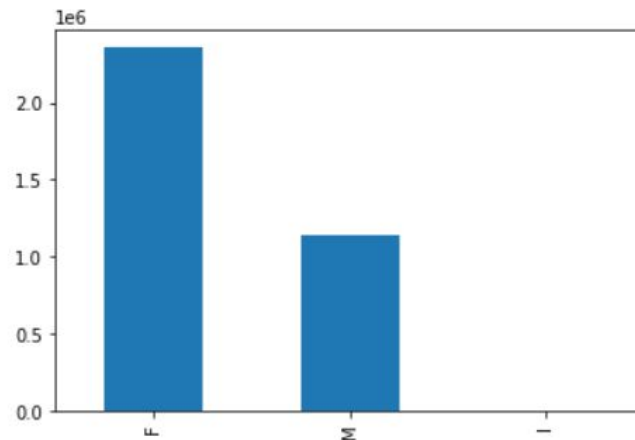
123.0	1
122.0	1
129.0	1
126.0	1
124.0	2
164.0	2
128.0	3
127.0	4
221.0	5
118.0	6
119.0	12
120.0	13
116.0	18
117.0	21
115.0	74
113.0	78
112.0	93
114.0	101
111.0	138
121.0	4250

```
Name: paciente_idade, dtype: int64
```

6 - Tratamento dos dados (pacientes com sexo com valor “I”)

```
# Quantitativo Total de Vacinados por Sexo Biológico  
df_vacinacao['paciente_enumSexoBiologico'].value_counts()
```

F	2357778
M	1140056
I	6



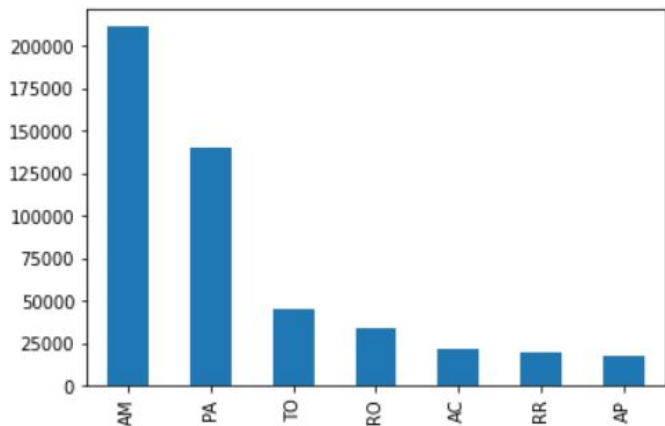
7 - Análises

- ➡ Perfil dos vacinados (sexo, idade, raça, grupo) e quantitativos por estado nas duas regiões;
- ➡ Quantitativo dos estabelecimentos de saúde e seus tipos nos estados nas duas regiões por total de população;
- ➡ Percentual de vacinados que tomaram a vacina em município diferente do da sua residência mas no mesmo estado.

Quantitativos de vacinados por estados na região norte / sudeste

```
# Total de vacinados por UF na região norte  
df_norte['paciente_endereco_uf'].value_counts()
```

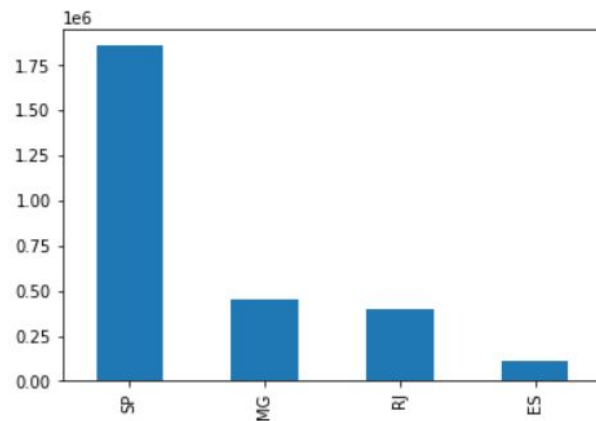
AM	211314
PA	139671
TO	45647
RO	33705
AC	21271
RR	19496
AP	17193



Região norte

```
# Total de vacinados por UF na região sudeste  
df_sudeste['paciente_endereco_uf'].value_counts()
```

SP	1857518
MG	452495
RJ	395395
ES	110396



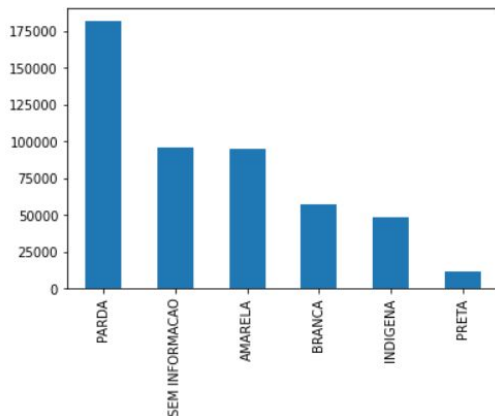
Região sudeste

Perfil dos vacinados e quantitativos por estados na região norte / sudeste (por raça)

```
# Quantitativo de Vacinados por Raça  
df_norte['paciente_racaCor_valor'].value_counts()
```

PARDA	181759
SEM INFORMACAO	95489
AMARELA	94752
BRANCA	56984
INDIGENA	48284
PRETA	11029

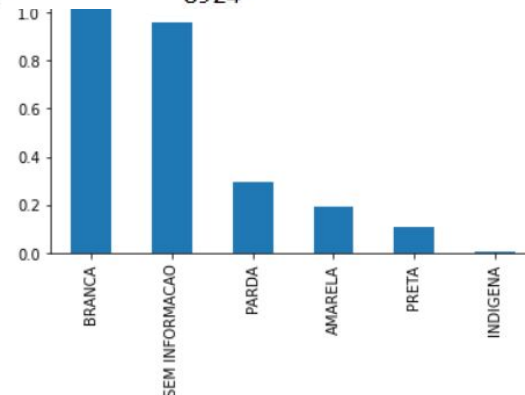
Região norte



```
# Quantitativo de Vacinados por Raça  
df_sudeste['paciente_racaCor_valor'].value_counts()
```

BRANCA	1247233
SEM INFORMACAO	955954
PARDA	296829
AMARELA	195292
PRETA	111572
INDIGENA	8924

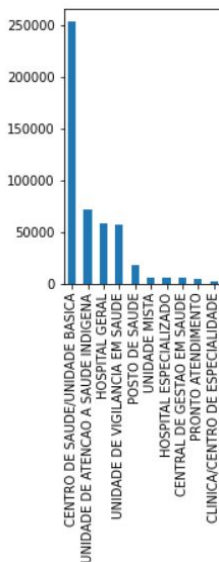
Região sudeste



Perfil dos estabelecimentos de saúde na região norte / sudeste (Top 10)

Quantitativo de Estabelecimentos de saúde por tipo de unidade
df_norte['ds_tipo_unidade'].value_counts()

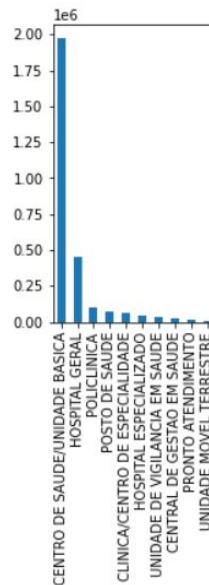
CENTRO DE SAUDE/UNIDADE BASICA	252957
UNIDADE DE ATENCAO A SAUDE INDIGENA	71406
HOSPITAL GERAL	58340
UNIDADE DE VIGILANCIA EM SAUDE	56656
POSTO DE SAUDE	17964
UNIDADE MISTA	5592
HOSPITAL ESPECIALIZADO	5092
CENTRAL DE GESTAO EM SAUDE	4990
PRONTO ATENDIMENTO	4148
CLINICA/CENTRO DE ESPECIALIDADE	2466



Região norte

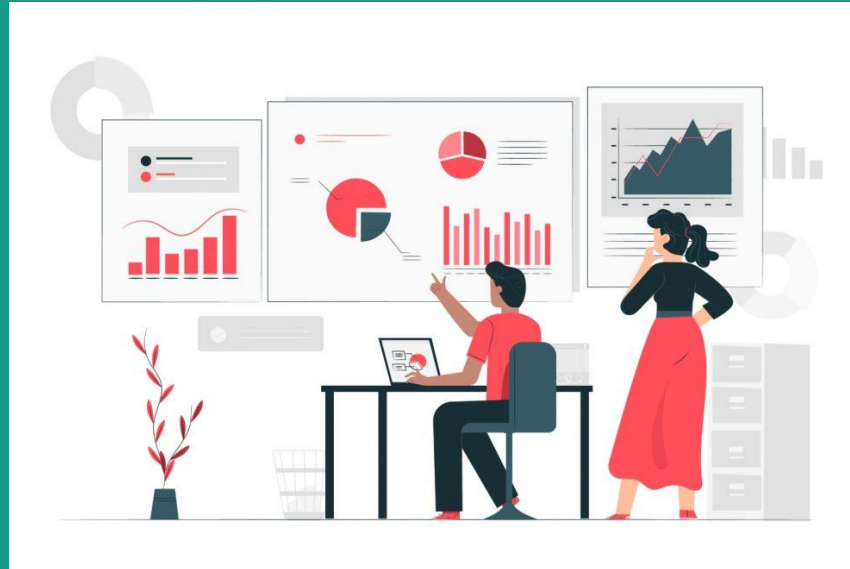
Quantitativo de Estabelecimentos de saúde por tipo de unidade
df_sudeste['ds_tipo_unidade'].value_counts()

CENTRO DE SAUDE/UNIDADE BASICA	1970468
HOSPITAL GERAL	451351
POLICLINICA	104206
POSTO DE SAUDE	70595
CLINICA/CENTRO DE ESPECIALIDADE	59941
HOSPITAL ESPECIALIZADO	43231
UNIDADE DE VIGILANCIA EM SAUDE	33221
CENTRAL DE GESTAO EM SAUDE	26397
PRONTO ATENDIMENTO	16781
UNIDADE MOVEL TERRESTRE	8750



Região sudeste

Análises em andamento



Análises em andamento

Quantitativo de estabelecimentos de saúde na Região Norte

```
df_cnes.query('uf in ["AM", "AC", "AP", "PA", "RO", "RR", "TO"]').count()
```

Total: 18361 Pop:18.672.591 milhoes

Quantitativo de estabelecimentos de saúde na Região Sudeste

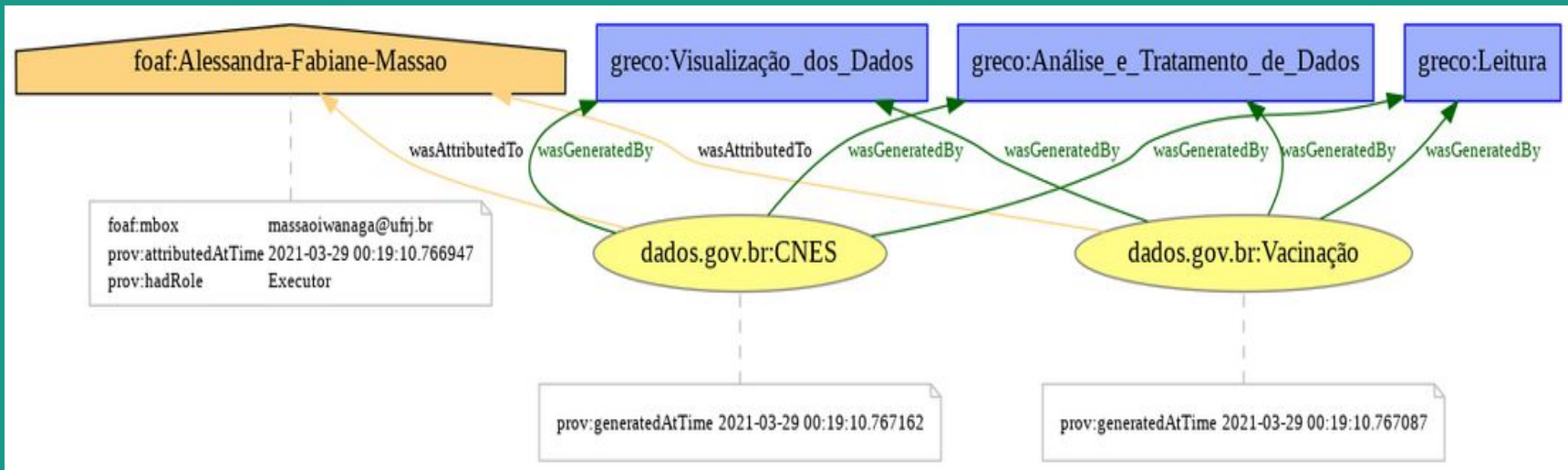
```
df_cnes.query('uf in ["ES", "MG", "RJ", "SP"]').count()
```

Total: 164896 Pop:89.012.240 milhões

89012240 ----- 164896

18672591 ----- x ---> 34.591 seria o "ideal" na região Norte

8 - Proveniência



9 - Reprodutibilidade

- ➡ Disponibilização de Notebook e datasets utilizados;
- ➡ Identificação do ambiente computacional usado no experimento;
- ➡ Inclusão de comentários descrevendo as ações realizadas;
- ➡ Código auto verificável.



Dúvidas?





Obrigado!

alessandrapsaraiva@gmail.com
famcrj@gmail.com
massaoiwanaga@ufrj.br



PPGI PROGRAMA
DE PÓS-GRADUAÇÃO
EM INFORMÁTICA
UNIVERSIDADE FEDERAL DO RIO DE JANEIRO

