

Dados abertos para quem?

Cássia Sampaio, Msc.

Dados abertos para quem?

~1 minuto

Respostas dados abertos para quem?

1. pesquisadores
2. formuladores de politicas publicas
3. estudantes
4. jornalistas
5. professores
6. pessoas curiosas
7. para todo mundo
8. agencias de pp
9. empresas
10. a oposicaaaaum
11. desafios do kaggle

Da onde surgiu essa idéia?

Idéia de dados abertos

~1942

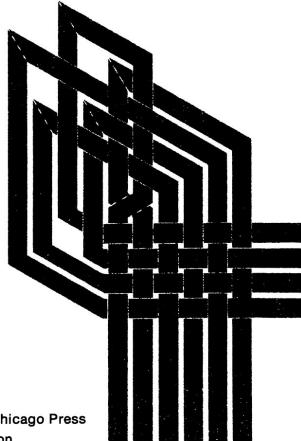
sociologia (científica)

Robert K.
Merton

Edited and
with an Introduction by
Norman W. Storer

The Sociology of Science

Theoretical and
Empirical
Investigations



Four sets of institutional imperatives—universalism, communism, disinterestedness, organized skepticism—are taken to comprise the ethos of modern science.

Idéia de dados abertos

~1942

sociologia (científica)

Normas Mertonianas

1. universalismo
2. comunismo
3. desinteresse
4. ceticismo organizado

Idéia de dados abertos

~1942

sociologia (científica)

Normas Mertonianas

1. **universalismo**: todos os cientistas podem contribuir para a ciência, independente de gênero, raça, nacionalidade ou cultura;
2. **comunismo**: os resultados científicos são propriedade comum de toda a sociedade;
3. **desinteresse**: cientistas agem em benefício de um interesse científico coletivo comum, ao invés de ganhos pessoais;
4. **ceticismo organizado**: as alegações científicas devem ser expostas a uma análise crítica contínua.

Porque "dados abertos"?

Termo "dados abertos"

~1995

geofísica

ON THE FULL AND OPEN EXCHANGE OF SCIENTIFIC DATA

and open exchange of scientific data and information. In May 1995, however, a proposal will be considered by the Congress of the World Meteorological Organization (WMO), originating in its working group on commercialization. This proposal would restrict the availability of environmental data, information, and relevant products, seriously affecting the ability of scientists to conduct research on global- or regional-scale problems. Current international practices guaranteeing the full and open exchange of scientific data should, if anything, be expanded, not restricted.

International programs for global change research and environmental monitoring crucially depend on the principle of full and open exchange (i.e., data and information are made available without restriction, on a non-discriminatory basis, for no more than the cost of reproduction and distribution; see Attachment 1, OECD [1994]). Commonly, observations collected for a specific,

Committee on Geophysical and Environmental Data
National Research Council
Washington, D.C.
1995

Termo "dados abertos" ~1995 geofísica

Contexto

1. tentativa de restrições de acesso, custo, propriedade e limitação dos dados
2. quanto mais dados, melhor (para a ciência entender fenômenos)
3. termo "open data"? ou seria "open exchange data"?

E os dados dos governos?

Dados públicos abertos

~2007

web 2.0

About the 2007 Workshop

Participants: Carl Malamud (Public.Resource.Org), Tim O'Reilly (O'Reilly Media), Greg Elin (Sunlight Foundation), Micah Sifry (Sunlight Foundation), Adrian Holovaty (EveryBlock), Daniel X. O'Neil (EveryBlock), Michal Migurski (Stamen Design), Shawn Allen (Stamen Design), Josh Tauberer (GovTrack.us), Lawrence Lessig (Stanford), Dan Newman (MapLight.Org), John Geraci (outside.in), Edwin Bender (Inst. for Money), Tom Steinberg (My Society), David Moore (Participatory Politics), Donny Shaw (Participatory Politics), JL Needham (Google), Joel Hardi (Public.Resource.Org), Ethan Zuckerman (Berkman), Greg Palmer (NewCo), Jamie Taylor (MetaWeb), Bradley Horowitz (Yahoo), Zack Exley (New Organizing Institute), Karl Fogel (Question Copyright), Michael Dale (Metavid), Joseph Lorenzo Hall (UC Berkeley), Marcia Hofmann (EFF), David Orban (Metasocial Web), Will Fitzpatrick (Omidyar Network), Aaron Swartz (Open Library).

The meeting was coordinated by Tim O'Reilly of O'Reilly Media and Carl Malamud of Public.Resource.Org, with sponsorship from the Sunlight Foundation, Google, and Yahoo.

Dados públicos abertos

~2007

web 2.0

Workshop (30 pessoas)

1. movimento de **código aberto** e **software livre** na web
2. dados devem ser de domínio público
3. definição de **dados públicos abertos** (open public data)
4. adoção pelo governo
5. práticas de dados abertos

Orientações para dados abertos ~2007

web 2.0

As stated in the note from the Sunlight Foundation's Board Chair, as of September 2020 the Sunlight Foundation is no longer active. This site is maintained as a static archive only.



SUNLIGHT FOUNDATION

To search, type and hit enter. Follow Us [Facebook](#) [Twitter](#)

ABOUT OUR WORK BLOG PRESS CONTACT

OPEN DATA POLICY GUIDELINES

Find the most up to date version of our guidelines at the [Open Data Policy Hub](#).

Download as [PDF](#) or [ODT](#)

Other Languages: [español](#)

Previous Versions: [2013](#), [2012](#)

What Data Should Be Public

1. Proactively release government information online
2. Reference and build on existing public accountability and access policies
3. Build on the values, goals and mission of the community and government
4. Create a culture

The Sunlight Foundation created this living set of open data guidelines to address: what data should be public, how to make data public, and how to implement policy.

The provisions are not ranked in order of priority and do not address every question one should consider when preparing a policy, but are a guide to answer the question of what an open data policy can and should do in striving to create a government data ecosystem where open data is the default. Setting the default to open means that the government and parties acting on its behalf will make public information available proactively and that they'll put that information within reach of the public (online), without barriers for its reuse and consumption. Setting the default to open is about living up to the potential of our information, about looking at comprehensive information management and making determinations that fall in the public interest.

- What Data Should Be Public
 - 1 Proactively Release Government Information Online

See examples

Most government information disclosure laws and systems currently in place, including

The Sunlight Foundation created this living set of open data guidelines to address: what data should be public, how to make data public, and how to implement policy.

**diretrizes de política de dados abertos
(open data policy guidelines)**

Orientações para dados abertos

atual

web 2.0+

A screenshot of a website titled "Open Data Policy Step-by-Step". The top navigation bar includes links for "Step-by-Step", "Generator", "Guidelines", "Collection", and "Why open data?". The main content area has a yellow header with the title and a sub-headline: "A step-by-step guide to creating an open data policy." Below this, there's a "Contents" sidebar with a numbered list from 1 to 6. The first item, "1. Where to begin", is highlighted with a yellow background. The main content area discusses the importance of public policy in helping governments make cultural shifts toward transparency and openness. It explains that open data is a cornerstone of open government and can help cities connect with residents. A "Read more about open data" button is present. The footer contains a "Build on what exists" section and a note about successful open data policy existing in a vacuum.

A product of the SUNLIGHT FOUNDATION

OPEN DATA POLICY HUB

Open Data Policy Step-by-Step

A step-by-step guide to creating an open data policy.

Contents:

- 1. Where to begin**
2. Set policy objectives for broad appeal
3. Draft policy language
4. Collect stakeholder feedback
5. Enact and implement policy
6. Announce policy

Public policy is an essential tool in helping governments make cultural shifts toward transparency and openness in the spirit of collaboration. As a cornerstone of open government, open data is one of the best tools data-driven cities can use to begin connecting with and collaborating with their residents.

We've developed this guide to help cities pass open data policies on their own. Passing an open data policy should be the first step in creating a sustainable, thriving open data program.

[Read more about open data](#)

1. Where to begin

Build on what exists — don't start from scratch

Successful open data policy does not exist in a vacuum, but rather [builds upon both formal and informal processes and structures already in place](#).

hub de política de dados
abertos
(open data policy hub)

Princípios para dados abertos (10)

atual

web 2.0+

1. **Completude:** conjuntos de dados devem ser os mais completos possíveis, refletindo o que foi registrado sobre um assunto.
2. **Precedência:** conjuntos de dados devem ser vinculados à entidade que os coletou.
3. **Temporalidade:** conjuntos de dados devem ser disponibilizados ao público em um tempo hábil. Sempre que possível, tão rapidamente quanto foi coletado.
4. **Acessibilidade:** conjuntos de dados publicados devem ser acessíveis (o quanto possível).
5. **Legibilidade por máquinas:** conjuntos de dados devem ser armazenados em formatos de arquivo que sejam facilmente processados por máquina (por exemplo, CSV, XML).
6. **Não discriminatório:** qualquer pessoa deve poder acessar os dados em qualquer momento, sem necessidade de identificação ou justificativa.
7. **Não proprietário:** o arquivo de conjunto de dados deve ser gratuito.
8. **Livre de licença:** licenças associadas aos dados devem ser concebidas para aumentar a abertura e minimizar restrições de utilização dos dados.
9. **Permanência:** informação publicada online, deve ser mantida online.
10. **Custos de uso:** conjuntos de dados publicados devem ser livres de custo de uso.

Uma linha do tempo mais completa

História (mais completa)

1942 - **Normas mertonianas:** Robert King Merton, pai da sociologia da ciência, diz que os resultados de pesquisa devem ser gratuitos e acessíveis para todas as pessoas

1957-58 - Movimento de dados abertos na ciência ([open science data movement](#))

1966 (julho) - Passa o Ato da liberdade de informação ([Freedom of Information Act, FOIA](#)) que garante que qualquer pessoa ou organização possa acessar dados do governo dos Estados Unidos por meio de um pedido de informação

1995 - Termo "dados abertos" ([open data](#)) aparece em um relatório sobre dados geofísicos e ambientais da agência americana chamado [Sobre a troca total e aberta de dados científicos](#) (On the full and open exchange of scientific data).

2004 (maio) - Criação da Fundação de conhecimento aberto ([Open Knowledge Foundation](#), OKF)

2007 (dezembro) - 30 pensadores e ativistas da Internet (web) realizam um encontro em [Sebastopol](#), São Francisco para definir dados públicos abertos.

2009 (janeiro) - Presidente Barack Obama assina o Memorando da transparência e governo aberto ([Memorandum on Transparency and Open Government](#)) que inicia a plataforma [data.gov](#) nos EUA.

2011 (setembro) - Início da parceria Governo aberto ([Open Government Partnership, OGP](#)). A proposta é que as entidades governamentais participantes assumam compromissos de transparência em relação aos dados públicos

2014 (fevereiro) - Lançamento oficial da [Open Knowledge Brasil](#) (iniciou em 2013).

E o Braseel?

Is Brazilian Open Government Data Actually Open Data?

Is Brazilian Open Government Data Actually Open Data? An Analysis of the Current Scenario

Kellyton dos Santos Brito, Pernambuco Rural Federal University and Pernambuco Federal University, Recife, Brazil

Marcos Antônio da Silva Costa, Brazilian Federal Service of Prosecution, Recife, Brazil

Vinicius Cardoso Garcia, Pernambuco Federal University, Recife, Brazil

Silvio Romero de Lemos Meira, Pernambuco Federal University, Recife, Brazil

10 Princípios analisados para
dados do Brasil, Recife e
Rio de Janeiro

2015

Is Brazilian Open Government Data Actually Open Data?

Table 2. Analysis of Brazilian OGD repositories

	Brazil	Recife	Rio de Janeiro
Complete	Partially	Partially	Partially
Primary	Compliant	Compliant	Compliant
Timely	Not compliant	Compliant	Not compliant
Accessible	Compliant	Compliant	Compliant
Machine Processable	Compliant	Compliant	Compliant
Nonproprietary	Compliant	Compliant	Compliant
Nondiscriminatory	Compliant	Compliant	Not compliant
License-free	Compliant	Compliant	Not compliant
Permanence	N/A	N/A	N/A
Usage Costs	Compliant	Compliant	Compliant

Is Brazilian Open Government Data Actually Open Data?

Brazil's Portal

At the time of the analysis, the Brazilian Open Data Portal listed 240 datasets and 2,109 resources in 12 categories. The analysis is as follows:

- **Complete:** Partially compliant. Health data is not complete, containing just a few datasets, such as basic health units, indices of HIV incidence and death, dengue incidence, labor accidents, and diseases related to poor sanitation. In addition, there is little demographic information about the units, such as their code, name, address, neighborhood, city, phone, and position (latitude and longitude);
- **Primary:** Compliant. Datasets are primary and can be downloaded in bulk;
- **Timely:** Not compliant. Although the datasets' metadata includes last updated time, there are no update policies, so it is impossible to determine the update frequency and next time data will be updated;
- **Accessible:** Compliant. Data are available on the Internet without restriction;
- **Machine processable:** Compliant. The datasets are published mainly using comma-separated values (CSV) format. The repository uses CKAN software (CKAN, n.d.), so an application programming interface (API) is also provided and returns information according to JavaScript Object Notation (JSON) format. Both formats are machine processable;
- **Nonproprietary:** Compliant. Both formats to publish data, CSV and JSON, are non-proprietary. In addition, the data dictionary to be read by developers is also in a non-proprietary format—the OpenDocument Text (ODT) format;
- **Nondiscriminatory:** Compliant. Data are available to anyone, with no registration requirement;
- **License-free:** Compliant. Datasets are licensed under the Open Data Commons Open Database License (ODbL);
- **Usage Costs:** Compliant. There is no cost, besides that required for an Internet connection, to access data.

Is Brazilian Open Government Data Actually Open Data?

Rio de Janeiro's Portal

At the time of the analysis, the Rio de Janeiro official repository contained 46 datasets from three agencies, accessed only through an API. The analysis proceeded as follows:

- **Complete:** Partially compliant. The entire dataset is small. Only a few datasets are published, with little data. Considering the health domain, there is only one dataset, containing only demographic data, such as name, code, address, phone, e-mail contact, and latitude and longitude of the health unit. It is a bit smaller than the Recife repository because it does not include health unit specialties;
- **Primary:** Compliant. Although the API allows data filter, access to all data without filtering is allowed;
- **Timely:** Not compliant. Datasets' metadata do not show the last update time, so it is impossible to know if the data are updated;
- **Accessible:** Compliant. Data are available on the Internet without restriction. However, it is important to notice that there is a limitation because users have to implement the OAuth authentication protocol to gain access to data;
- **Machine processable:** Compliant. Despite this repository uses CKAN software, the repository is based only on API calls. The responses are based on public machine processable formats, such as JSON and Keyhole Markup Language (KML);
- **Nonproprietary:** Compliant. Both formats to publish data, JSON and KML, are nonproprietary;
- **Nondiscriminatory:** Not Compliant. Prior to accessing any data, users have to register on the site. Thus, anonymous access is impossible. In addition, they also state that they collect lots of data from users, such as IP addresses and ISP names, among others;
- **License-free:** Not compliant. Specific dataset licenses were not found, and the terms of use are protective, including the following: "The City may require the user to cancel the use, distribution and/or any service and any data set for any reason" (Rio de Janeiro City Hall, 2011c);
- **Usage Costs:** Compliant. There is also no cost, besides that required for an Internet connection, to access data.

Será que algo mudou?
(prática)

Plataforma de dados abertos (gov.br) 2023



Dados Abertos

7 respostas de busca

busca:
"violencia contra a mulher"

Conjuntos De Dados (7) Organizações (0) Temas (0) Reusos (0) Notícias (0) Discussões (0)

Violência Contra Mulher
17 Recursos 0 Reusos
22/05/2023 20:32:07

Registros De Violência Contra A Mulher No Estado Do Rio Grande Do Sul
2 Recursos 0 Reusos
05/10/2016 17:48:15

Violência Contra A Mulher No DF
3 Recursos 0 Reusos
03/01/2023 23:34:26

Central De Atendimento À Mulher – Ligue 180
1 Recursos 0 Reusos
14/10/2021 20:10:59

Ouvidoria Nacional De Direitos Humanos – ONDH
2 Recursos 0 Reusos
06/09/2021 13:50:44

3. Segurança Urbana - Diagnósticos E Estratégias Para Maceió
5 Recursos 0 Reusos
03/07/2022 04:01:22

COVID-19 - Leis Estaduais
41 Recursos 0 Reusos
04/08/2022 04:15:02

Plataforma de dados abertos (gov.br)

2023

The screenshot shows a grid of 12 dataset cards from the gov.br open data platform. Most cards have red error messages indicating they are unavailable for preview. Some cards also mention missing original files or manipulated data.

Dataset Title	Description
dados de Minas Gerais	Violência contra Mulher + página não encontrada
Violência Contra Mulher	Registros De Violência Contra A Mulher No Estado Do Rio Grande Do Sul + outra página com csvs
Registros De Violência Contra A Mulher No Estado Do Rio Grande Do Sul	Violência contra a mulher no DF pdfs com gráficos (sem arquivo original)
Violência Contra A Mulher No DF	Central De Atendimento À Mulher – Ligue 180 + outra página com csvs
Central De Atendimento À Mulher – Ligue 180	Ouvintes Nacionais de Direitos Humanos – ONDH arquivo csv com download infinito
Ouvintes Nacionais de Direitos Humanos – ONDH	3. Segurança urbana - Diagnósticos e estratégias para Maceió pdfs com slides (sem arquivo original)
3. Segurança Urbana - Diagnósticos E Estratégias Para Maceió	Covid-19 - Leis Estaduais diversos pdfs com leis (sem arquivo de dados)
Covid-19 - Leis Estaduais	COVID-19 - Leis Estaduais
COVID-19 - Leis Estaduais	COVID-19 - Leis Estaduais
COVID-19 - Leis Estaduais	COVID-19 - Leis Estaduais
COVID-19 - Leis Estaduais	COVID-19 - Leis Estaduais

Dados de [minas gerais](#) podiam ser usados ou dados do [lige 180](#) com bastante manipulação e uniformização

Plataforma de dados abertos (gov.br)

2024

9 respostas de busca

busca:
"violencia contra a mulher"

sem metadados
dados de 2021
esgotou o tempo
de resposta do
servidor

sem metadados
sem informação sobre a
data dos dados
esgotou o tempo de
resposta do servidor

relatórios, sem
dados, sem data ou
arquivos de dados

Conjuntos De Dados (9) Organizações (0)

LIBERTAS QNÉ SERA
TAMEN

Violência Contra Mulheres SES

15 Recursos 0 Reusos
16/02/2024 00:04:09

sem metadados
dados de 2023
busquei pelo github

LIBERTA JÉ SERA
TAMEN

Violência Contra Mulher

17 Recursos 0 Reusos
23/08/2023 19:44:09

scussões (0)

Registros De Violência Contra A
Mulher No Estado Do Rio Grande
Do Sul

2 Recursos 0 Reusos
05/10/2016 14:48:15

Violência Contra A Mulher No DF

3 Recursos 0 Reusos
03/01/2023 20:34:26

sem metadados
dados de 2019
arquivos
separados

Central De Atendimento À
Mulher – Ligue 180

1 Recursos 0 Reusos
14/10/2021 17:10:59

DOS NOS UNIA
MDHC

Ouvidoria Nacional De Direitos
Humanos – ONDH

2 Recursos 0 Reusos
06/09/2021 10:50:44

[2ª Fase] 5. Oficinas Cidade
Mulher

3 Recursos 0 Reusos
16/01/2024 07:08:09

relatórios, sem data
ou arquivos de
dados

[1ª Fase] 3. Segurança Urbana -
Diagnósticos E Estratégias Para
Maceió

5 Recursos 0 Reusos
12/11/2023 03:24:18

COVID-19 - Leis Estaduais

41 Recursos 0 Reusos
04/06/2022 01:15:02

dados indisponíveis
dados de 2011

dados de 2020
erro 404 de página
não encontrada

outro assunto
pdfs com leis

Plataforma de dados abertos (gov.br)

2023

Conjuntos De Dados (7) Organizações (0) Temos (0) Reusos (0) Discussões (0)

dados de Minas Gerais
Violência Contra Mulher
+ página não encontrada

Registros De Violência Contra A Mulher No Estado Do Rio Grande Do Sul
Violência Contra A Mulher No DF
+ outra página com pdfs com gráficos (sem arquivo original)

Central De Atendimento À Mulher – Ligue 180
Violência Contra Mulher SES

arquivo csv com download infinito
Ouvintoria Nacional De Direitos Humanos – ONDH
pdfs com slides (sem arquivo original)

diversos pdfs com leis (sem arquivo de dados)
COVID-19 - Leis Estaduais

2024

sem metadados dados de 2021 esgotou o tempo de resposta do servidor
LIBER, TÉRUA TAMEN

sem metadados dados de 2023 busquei pelo github
LIBER, TÉRUA TAMEN

dados indisponíveis dados de 2011
dados indisponíveis dados de 2011

Violência Contra Mulher
Violência Contra Mulher SES

sem metadados sem informação sobre a data dos dados esgotou o tempo de resposta do servidor
Violência Contra A Mulher No DF

sem metadados dados de 2019 arquivos separados
Central De Atendimento À Mulher – Ligue 180

dados de 2020 erro 404 de página não encontrada
Ouvintoria Nacional De Direitos Humanos – ONDH

relatórios, sem dados, sem data ou arquivos de dados
[2ª Fase] 5. Oficinas Cidade Mulher

relatórios, sem data ou arquivos de dados
[1ª Fase] 3. Segurança Urbana - Diagnósticos E Estratégias Para Maceió

outro assunto pdfs com leis
COVID-19 - Leis Estaduais

Perguntas

2023

Qual o tipo mais comum de violência contra a mulher
no ano de 2022, em Minas Gerais?



dados de Minas Gerais de
2022

2024

Qual o tipo mais comum de violência contra a mulher
no ano de 2023, em Minas Gerais?



dados de Minas Gerais de
2023

Dados disponíveis

2023

colunas

1. **municipio_cod:** Código do Município do Evento
2. **municipio_fato:** Nome do Município do Evento
3. **data_fato:** data de ocorrência do evento
4. **mes:** mês de ocorrência do evento
5. **ano:** ano de ocorrência do evento
6. **risp:** Região Integrada de Segurança Pública
7. **rbmh:** Região Metropolitana de BH
8. **natureza_delito:** Natureza do Delito
9. **tentado_consumado:** Tentado ou consumado
10. **qtde_vitimas:** quantitativo de vítimas de violência doméstica

2024

colunas

1. **municipio_cod:** Código do Município do Evento
2. **municipio_fato:** Nome do Município do Evento
3. **data_fato:** data de ocorrência do evento
4. **mes:** mês de ocorrência do evento
5. **ano:** ano de ocorrência do evento
6. **risp:** Região Integrada de Segurança Pública
7. **rbmh:** Região Metropolitana de BH
8. **natureza_delito:** Natureza do Delito
9. **tentado_consumado:** Tentado ou consumado
10. **qtde_vitimas:** quantitativo de vítimas de violência doméstica

Exploração dos dados

2023

Informações sobre cada coluna

```
1 df.info()

<class 'pandas.core.frame.DataFrame'>
RangeIndex: 101872 entries, 0 to 101871
Data columns (total 10 columns):
 #   Column           Non-Null Count  Dtype  
--- 
 0   municipio_cod    101872 non-null   int64  
 1   municipio_fato   101872 non-null   object  
 2   data_fato         101872 non-null   datetime64[ns]
 3   mes               101872 non-null   int64  
 4   ano               101872 non-null   int64  
 5   risp              101872 non-null   object  
 6   rmbh              101872 non-null   object  
 7   natureza_delito  101872 non-null   object  
 8   tentado_consumado 101872 non-null   object  
 9   qtde_vitimas     101872 non-null   int64  
dtypes: datetime64[ns](1), int64(4), object(5)
memory usage: 7.8+ MB
```

2024

```
1 df2.info()

<class 'pandas.core.frame.DataFrame'>
RangeIndex: 61536 entries, 0 to 61535
Data columns (total 10 columns):
 #   Column           Non-Null Count  Dtype  
--- 
 0   municipio_cod    61536 non-null   int64  
 1   municipio_fato   61536 non-null   object  
 2   data_fato         61536 non-null   datetime64[ns]
 3   mes               61536 non-null   int64  
 4   ano               61536 non-null   int64  
 5   risp              61536 non-null   object  
 6   rmbh              61536 non-null   object  
 7   natureza_delito  61536 non-null   object  
 8   tentado_consumado 61536 non-null   object  
 9   qtde_vitimas     61536 non-null   int64  
dtypes: datetime64[ns](1), int64(4), object(5)
memory usage: 4.7+ MB
```

Exploração dos dados

2023

2024

Medidas descritivas

1	df.describe().T
	count mean std min 25% 50% 75% max
municipio_cod	101872.0 313692.296107 2127.476060 310010.0 311860.0 313670.0 315430.0 317220.0
mes	101872.0 6.544438 3.457818 1.0 4.0 7.0 10.0 12.0
ano	101872.0 2022.000000 0.000000 2022.0 2022.0 2022.0 2022.0 2022.0
qtde_vitimas	101872.0 1.374872 1.295465 1.0 1.0 1.0 1.0 30.0

1	df2.describe().T
	count mean std min 25% 50% 75% max
municipio_cod	61536.0 313681.674353 2124.333117 310010.0 311860.0 313670.0 315400.0 317220.0
mes	61536.0 3.927701 1.995426 1.0 2.0 4.0 6.0 7.0
ano	61536.0 2023.000000 0.000000 2023.0 2023.0 2023.0 2023.0 2023.0
qtde_vitimas	61536.0 1.393314 1.403380 1.0 1.0 1.0 1.0 32.0

Exploração dos dados

2023

rmbh: Região Metropolitana de BH

```
In [140]: 1 df['rmbh'].value_counts()
```

```
Out[140]: 3) Interior de MG      84348
           2) RMBH (sem BH)      13154
           1) Belo Horizonte      4370
Name: rmbh, dtype: int64
```

```
In [239]: 1 df['rmbh'].value_counts(normalize=True)
```

```
Out[239]: 3) Interior de MG      0.827980
           2) RMBH (sem BH)      0.129123
           1) Belo Horizonte      0.042897
Name: rmbh, dtype: float64
```

2024

```
1 df2['rmbh'].value_counts()
3) Interior de MG      50484
2) RMBH (sem BH)      8347
1) Belo Horizonte      2705
Name: rmbh, dtype: int64
```

```
1 df2['rmbh'].value_counts(normalize=True)
```

```
3) Interior de MG      0.820398
2) RMBH (sem BH)      0.135644
1) Belo Horizonte      0.043958
Name: rmbh, dtype: float64
```

Exploração dos dados

2023

natureza_delito: Natureza do Delito

```
In [142]: 1 len(df['natureza_delito'].unique())  
Out[142]: 188
```

```
In [146]: 1 df['natureza_delito'].value_counts(normalize=True).head(15)  
  
Out[146]: AMEACA 0.284642  
VIAS DE FATO / AGRESSAO 0.197846  
LESAO CORPORAL 0.165639  
DESCUMPRIMENTO DE MEDIDA PROTETIVA DE URGENCIA 0.054676  
ATEND. DENUNCIA INFRACOES CONTRA MULHER(VIOLENCIA 0.033807  
DANO 0.030195  
ATRITO VERBAL 0.020693  
PERSEGUICAO 0.020163  
FURTO 0.018582  
INJURIA 0.017885  
OUTRAS ACOES DEFESA SOCIAL 0.015500  
OUTROS INFRACOES C/ A PESSOA 0.015097  
VIOLENCIA PSICOLOGICA 0.013753  
DIFAMACAO 0.010631  
ESTUPRO DE VULNERAVEL 0.010582  
Name: natureza_delito, dtype: float64
```

2024

```
1 len(df2['natureza_delito'].unique())  
167
```

```
1 df2['natureza_delito'].value_counts(normalize=True).head(15)  
  
AMEACA 0.280892  
VIAS DE FATO / AGRESSAO 0.186801  
LESAO CORPORAL 0.158931  
DESCUMPRIMENTO DE MEDIDA PROTETIVA DE URGENCIA 0.065474  
ATEND. DENUNCIA INFRACOES CONTRA MULHER(VIOLENCIA 0.036401  
DANO 0.029170  
PERSEGUICAO 0.025384  
ATRITO VERBAL 0.020703  
INJURIA 0.019663  
FURTO 0.017502  
VIOLENCIA PSICOLOGICA 0.016023  
OUTRAS ACOES DEFESA SOCIAL 0.015389  
OUTROS INFRACOES C/ A PESSOA 0.015048  
DIFAMACAO 0.011652  
ESTUPRO DE VULNERAVEL 0.010010  
Name: natureza_delito, dtype: float64
```

Exploração dos dados

2023

`tentado_consumado: delito tentado ou consumado`

```
In [148]: 1 df['tentado_consumado'].value_counts()
```

```
Out[148]: CONSUMADO    100948  
TENTADO        924  
Name: tentado_consumado, dtype: int64
```

```
In [149]: 1 df['tentado_consumado'].value_counts(normalize=True)
```

```
Out[149]: CONSUMADO    0.99093  
TENTADO      0.00907  
Name: tentado_consumado, dtype: float64
```

2024

```
1 df2['tentado_consumado'].value_counts()
```

```
CONSUMADO    60951  
TENTADO      585  
Name: tentado_consumado, dtype: int64
```

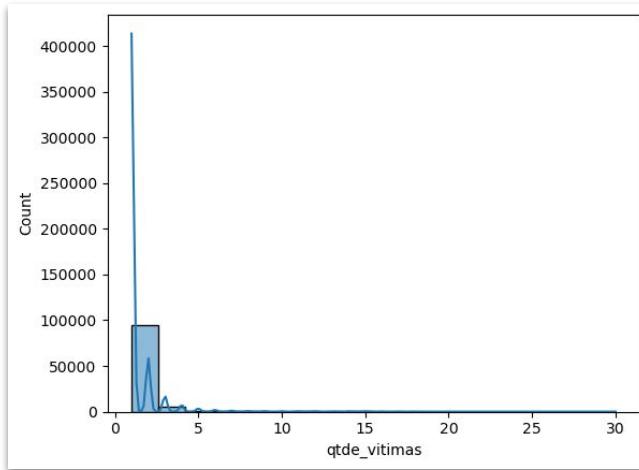
```
1 df2['tentado_consumado'].value_counts(normalize=True)
```

```
CONSUMADO    0.990493  
TENTADO      0.009507  
Name: tentado_consumado, dtype: float64
```

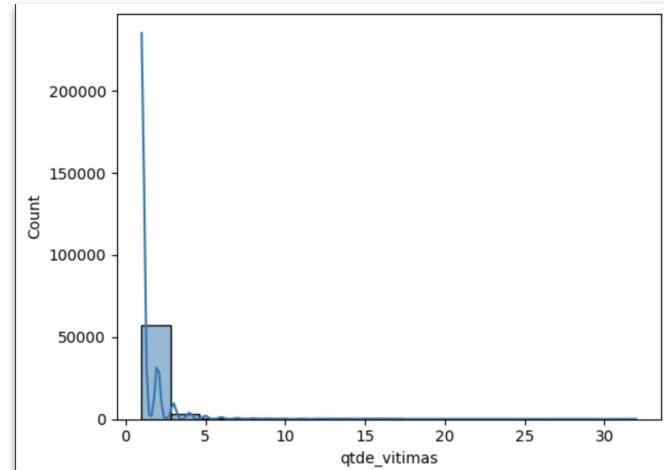
Exploração dos dados

2023

Histograma da quantidade de vítimas

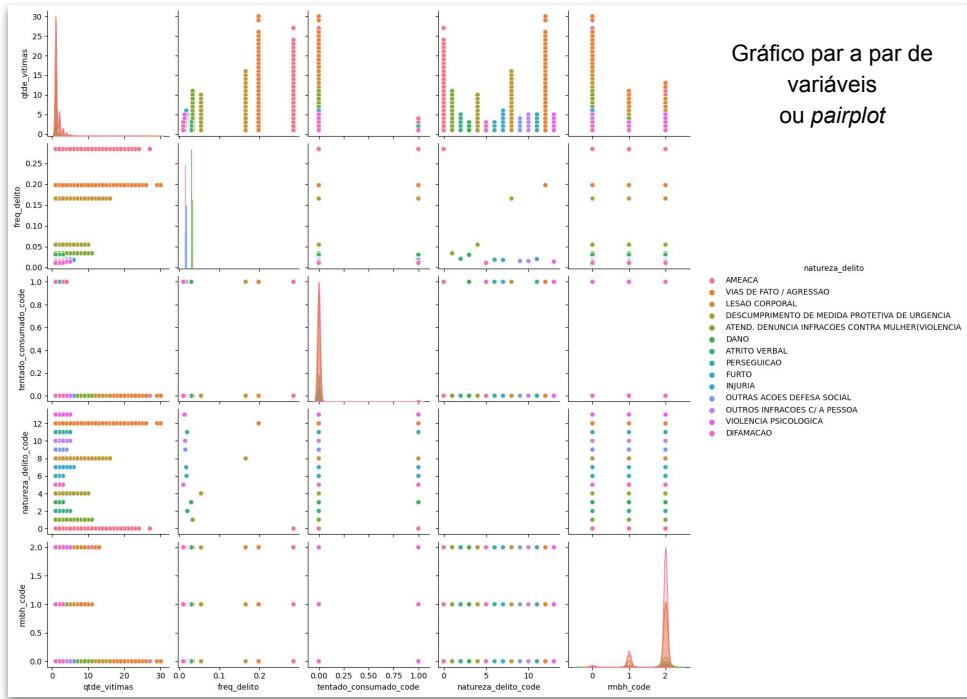


2024

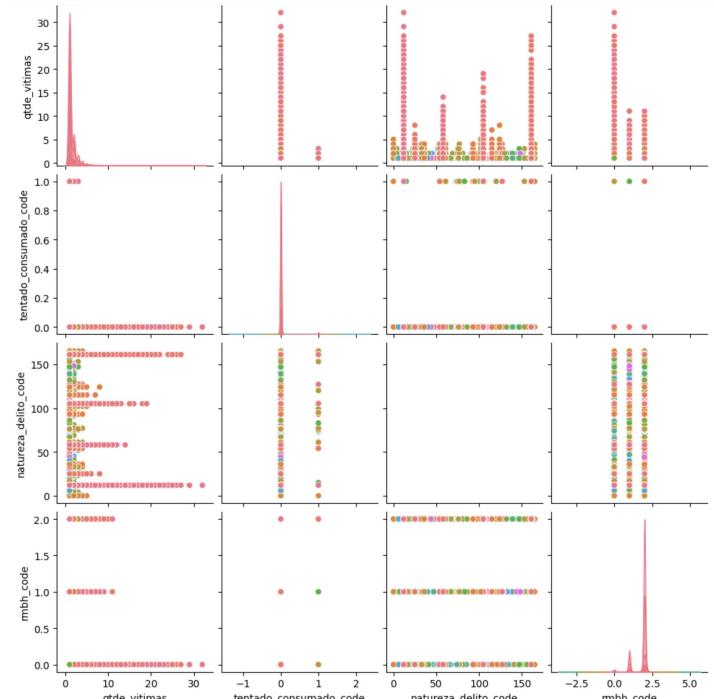


Exploração dos dados

2023



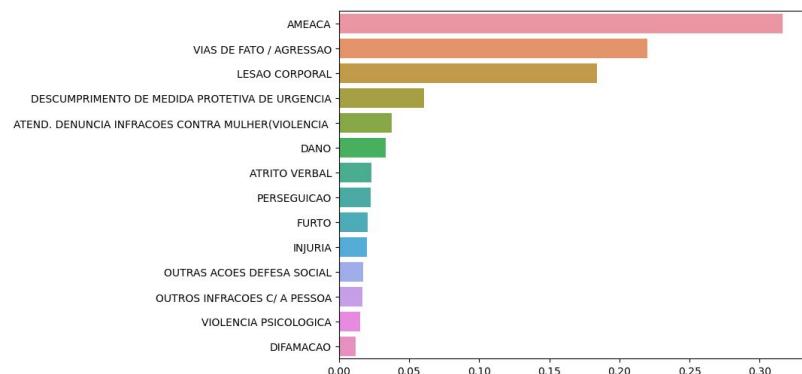
2024



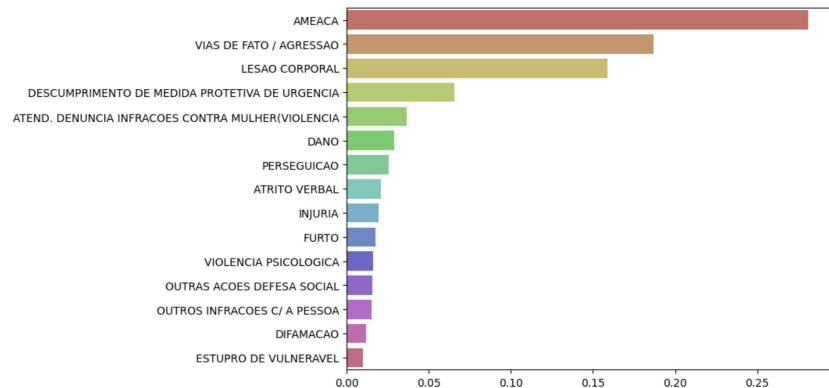
Apresentação de resultados

2023

Gráfico para análise



2024



Apresentação de resultados

2023

Gráfico para apresentação com técnicas de visualização



2024



Voltemos as respostas

Definição da OKF de dados abertos

The Open Definition sets out principles that define “openness” in relation to data and content.

It makes precise the meaning of “open” in the terms “open data” and “open content” and thereby ensures quality and encourages compatibility between different pools of open material.

It can be summed up in the statement that:

“Open means anyone can freely access, use, modify, and share for any purpose (subject, at most, to requirements that preserve provenance and openness).”

Put most succinctly:

“Open data and content can be freely used, modified, and shared by **anyone** for any purpose”

E é mesmo para
qualquer pessoa?

Referências

- Video the story of open data: https://www.youtube.com/watch?v=J4A8OTPSkWw&ab_channel=Socrata
- Post Open data 101: The history and principles of open data – Part 1: <https://apolitical.co/solution-articles/en/open-data-101-the-history-and-principles-of-open-data-part-1>
- Post Open data 101 part 2: <https://apolitical.co/solution-articles/en/open-data-101-the-possibilities-and-limitations-of-open-data-part-2>
- Definicao de dados abertos OKF: <https://opendefinition.org/>
- Wikipedia sobre dados abertos: https://pt.wikipedia.org/wiki/Dados_abertos
- Wikipedia sobre open knowledge brasil: https://pt.wikipedia.org/wiki/Open_Knowledge
- Post Brasil se torna o primeiro país com um representante da Open Knowledge fora da Europa:
<https://ok.org.br/noticia/brasil-se-torna-o-primeiro-pais-com-um-representante-da-open-knowledge-fora-da-europa/>
- Post 10 Years of Open Data: <https://www.opendatasoft.com/en/blog/open-data-anniversary-ten-years-after-the-sebastopol-meeting/>
- Artigo On the Full and Open Exchange of Scientific Data, Committee of geophysical and environmental data: <https://nap.nationalacademies.org/read/18769/chapter/1>
- The key concepts of Open Science white paper: <https://books.openedition.org/oep/pdf/1647>
- Post Scielo The Open Data movement: international consolidation: <https://blog.scielo.org/en/2014/07/14/the-open-data-movement-international-consolidation/>
- Post data.gov US Open Data: A History: <https://data.gov/blog/open-data-history/>
- Post A brief history of Open Data Paris Tech Review: <https://www.paristechreview.com/2013/03/29/brief-history-open-data/>
- Wikipedia sobre normas Mertonianas: https://en.wikipedia.org/wiki/Mertonian_norms#CITEREFMerton1973
- Site The Annotated 8 Principles of Open Government Data: <https://opengovdata.org/>
- Sunlight foundation: <https://sunlightfoundation.com/opendataguidelines/>
- Open data policy hub: <https://opendatapolicyhub.sunlightfoundation.com/>
- The normative structure of science: https://law.unimelb.edu.au/_data/assets/pdf_file/0005/3609203/1c-Merton-The-Normative-Structure-of-Science.pdf
- Artigo Is Brazilian Open Government Data Actually Open Data?
https://www.researchgate.net/publication/276532378_Is_Brazilian_Open_Government_Data_Actually_Open_Data
- O que é FOIA: https://wikilai.fiquemsabendo.com.br/wiki/O_que_%C3%A9_FOIA_-_Freedom_of_Information_Act
- Github com material: <https://github.com/cassiasamp/open-data-day-2024>
- Portal dados abertos: <https://dados.gov.br/home>
- Portal da transparéncia (MG): <https://github.com/transparencia-mg>

Sugestões de material

Livros para começar

Livros em inglês

[Python Data Science Handbook](#)

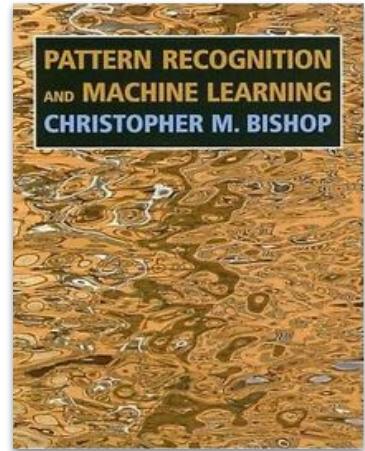
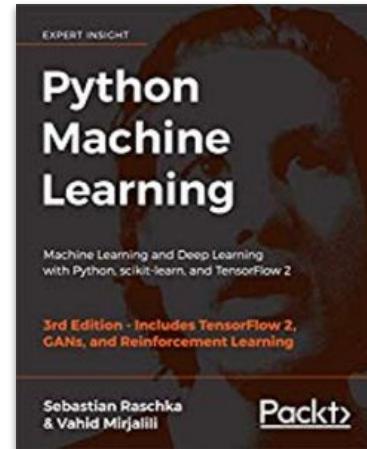
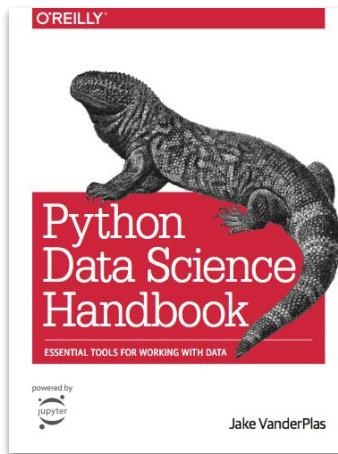
[Python Machine Learning Raschka](#)

[Pattern Recognition do Bishop](#)

Githubs em português

[Github de data science das PyLadies](#)

[Github do Fernando Masanori](#)



Cursos online

Coursera

[Data Science Courses](#)

Udacity

[Data Science Courses na Udacity](#)

Datacamp

[Data Science for Everyone](#)

Udemy

[Data Science da Hekima](#)

Repositórios de dados

Em inglês

[Kaggle Datasets](#)

[Repositório da UCI](#)

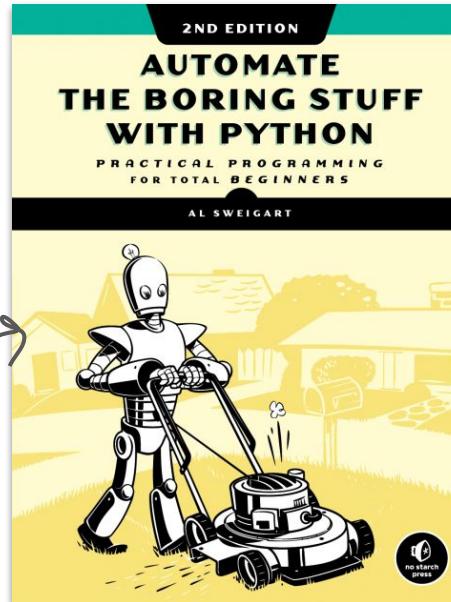
Em português

[Portal de Dados Abertos](#)

Como encontrar dados?

Geralmente fazendo pesquisas do tipo
assunto + dataset em um buscador

Para **coletar dados** automaticamente



[site do livro \(em inglês\)](#)

Contatos

Github: /cassiasamp

Linkedin: /in/cassiasamp

Twitter: @cassiasamp

