



OBSERVATÓRIO BRASILEIRO DE POLÍTICAS PÚBLICAS
COM A POPULAÇÃO EM SITUAÇÃO DE RUA

METADADOS: DEZ PASSOS PARA COMEÇAR ESTUDOS E ANÁLISE DE DADOS

WELLINGTON MIGLIARI



Polos
DE CIDADANIA



DIREITO

PROEX
PRÓ-REITORIA
DE EXTENSÃO

UFMG

**METADADOS:
DEZ PASSOS PARA COMEÇAR
ESTUDOS E ANÁLISE DE DADOS**

WELLINGTON MIGLIARI

Belo Horizonte
2022

Projeto Gráfico: W. Migliari \LaTeX

Foto de Capa: Adrien Olichon

Revisão Textual: Raquel Araújo Monteiro e André Luiz Freitas Dias

© OBPP

METADADOS: DESCRIÇÃO E ENTENDIMENTO SOBRE A DESAGREGAÇÃO DE DADOS

Na página do Observatório Brasileiro de Políticas Públicas com a População em Situação de Rua, você encontrará dados sobre a população em situação de rua no Brasil entre os anos de 2012 e 2021. Esses dados foram coletados pelas prefeituras do país, posteriormente, transferidos ao Ministério da Cidadania e, por fim, tabulados pelo próprio Ministério. A seleção de dados que disponibilizamos foi realizada pelo Observatório e, por isso, não é possível encontrá-la no banco de dados do CECAD, Cadastro Único, CadÚnico. A seguir, veja algumas definições, conceitos e esclarecimentos que lhe ajudarão a entender de forma adequada o uso que fazemos da série histórica.

1. O que é uma série histórica?

Uma série histórica é um conjunto de dados coletados, tabulados e compilados em que seja suficiente notar tendências em um período de longo prazo. Esse tempo varia de instituição para instituição. Pode ser uma série histórica de cinco, dez ou mais anos. No caso da série oferecida pelo Ministério da Cidadania ao Observatório Brasileiro

de Políticas Públicas com a População em Situação de Rua sobre o fenômeno da população em situação de rua no Brasil, **temos dados de 2012 a 2021, isto é, tendências estatísticas da última década.** Para fins de definição, uma série histórica difere de uma amostra ou de um censo. Dados coletados em um período específico, representando

parte de uma população, podem ser definidos como amostras ou dados amostrais. Já os censos requerem tanto a inclusão de todos os eventos ou todas as pessoas de um país, estado ou município bem como a total cobertura geográfica do limite administrativo a ser estudado. Na prática, os censos incluem a realização de entrevistas ou registros obedecendo às marcações, por exemplo, “homens” e “mulheres”, “idade”, “renda” etc.¹ A série histórica que temos trabalhado, portanto, não é nem substitui um censo. Ela pode ser definida enquanto uma amostra, uma vez representa

parte de uma dada população em um município. Contudo, é importante ressaltar que o Ministério da Cidadania nos entregou tabelas de diferentes anos e o Observatório tratou esses dados de modo que pudessem ser lidos no formato de série histórica. Assim, essa amostra ou série histórica sobre os grupos socioeconomicamente vulneráveis, como a população em situação de rua ou indivíduos em situação de extrema pobreza, é fruto do trabalho da pesquisa social aplicada conduzida pelo Observatório Brasileiro de Políticas Públicas com a População em Situação de Rua.

“

Uma série histórica é um conjunto de dados coletados, tabulados e compilados em que seja suficiente notar tendências em um período de longo prazo.

”

¹Nem o Instituto Brasileiro de Geografia e Estatística, nem o CadÚnico realizam cadastros fora da classificação biológica binária “masculino” ou “feminino”.

Por que uma série histórica é tão importante? As séries históricas ajudam os poderes constitucionais e administrações públicas do país a desenharem políticas públicas, projetar orçamento, implementar benefícios sociais e efetuar o repasse de verbas do go-

verno federal às prefeituras, entre outras ações.² Por isso, a série histórica da que dispomos é um poderoso instrumento de medida e aferição de fenômenos sociais no Brasil afetados pela alta vulnerabilidade socioeconômica.

2. Qual a diferença de dados coletados, tabulados e compilados?

Os “dados coletados” são aqueles imediatos ou brutos, retirados dos cadastros municipais sobre a população em situação de rua, e visam à construção de um primeiro banco de dados. Depois da coleta, eles devem ser tabulados ou organizados em tabelas para que possam ser lidos por pes-

soas e máquinas – programas ou *softwares* de leitura de dados. É possível que a coleta e tabulação de dados ocorra em nível municipal para que, posteriormente, os dados sejam enviados ao Ministério da Cidadania.³ Por fim, **a compilação de dados é a fase em que ocorre a consolidação**

²O repasse de verbas do governo federal para as prefeituras obedece o Decreto No. 11.016 de 29 de março de 2022 que regulamenta o Cadastro Único ou CadÚnico. Em seu artigo 7, há claro comando para regular os atos administrativos municipais e, por isso, elemento juridicamente vinculante entre os entes municipais e a União quanto à alimentação das bases de dados administrativas locais consoante instruções sobre a coleta, disponibilidade, acesso e atualização dos cadastros. O elemento jurídico que dá solidez aos convênios é ato normativo próprio também previsto na Lei No. 14.129 de 29 de março de 2021 sobre interoperabilidade de dados, artigo 2, inciso III.

³Conforme notamos, parece não haver interoperabilidade de dados entre as bases administrativas dos municípios e a do CECAD, Ministério da Cidadania. A interoperabilidade resulta do compartilhamento e da integração das bases de dados dos entes locais com o CadÚnico. Quando um banco de dados municipal é interoperável com o CadÚnico, qualquer alteração local significa alteração imediata nas bases do governo federal. Esse tipo de integração e compartilhamento de informações aumenta a transparência dos registros, mitiga custos para a administração pública na gestão do CadÚnico e aumenta a efetivação de direitos fundamentais com a redução do risco de que alguém deixe de receber um repasse de benefício por desatualização de cadastro.

das tabelas com os registros sobre a população em situação de rua pelo Ministério da Cidadania em nível nacional. Contudo, nem sempre os dados compilados se encontram disponíveis em formato de série histórica, ou seja, informações decodificadas e organizadas em tabelas que possam ser facilmente lidas sem a necessidade de desagregá-las ano a ano. Então quer dizer que não há dados compilados na forma de série histórica na

página do CECAD? No portal do Ministério da Cidadania, [CECAD](#) ou [CadÚnico](#), encontramos muitas informações e séries históricas sobre diversos seguimentos populacionais vulneráveis, entretanto, o acesso é restrito para os dados sobre a quantidade de pessoas que se encontram em situação de rua. As séries históricas disponíveis no CECAD para esse seguimento são referentes a famílias. O que nos interessa são os totais por indivíduo.

“

As séries históricas disponíveis no CECAD para esse seguimento são referentes a famílias. O que nos interessa são os totais por indivíduo.

”

3. Depois de compilados, os dados já podem ser facilmente lidos sem precisar decodificá-los?

Não. Conforme exposto anteriormente, grandes bases de dados são compiladas por meio de

códigos. Isso pelo fato de se evitar erros de digitação, usar menos espaço no armazenamento

dessas informações e padronizar a leitura computacional por números no lugar de letras. De forma mais simples, as colunas possuem códigos de identificação em vez de nomes. A chave para cada código, como no caso do CadÚnico, está registrada no dicionário de dados. Com a ajuda de um *software*, deciframos cada código atribuindo-lhe o conceito definido pelo dicionário de dados. Exemplo: uma coluna chamada “CO_SEXO_PESSOA” foi criada para identificar se o cadastro de uma certa pessoa remete ao sexo “Masculino” ou “Feminino”. Para localizar o município, existe outra coluna ao lado de nome “CD_IBGE_CADASTRO” com o respectivo código do

município. Ao lermos as informações contidas na coluna “CO_SEXO_PESSOA” ao lado da coluna “CD_IBGE_CADASTRO”, vemos o número 1 ou 2 e o código do IBGE, que corresponde a essa linha, atribuído ao município. Assim, toda vez que tivermos o dígito 1 aparecer nessa coluna sobre o sexo dos entrevistados, deve-se decodificar como “Masculino” e, no caso do dígito 2, “Feminino”. Se essa linha dizer respeito a um cadastro feito na cidade de Belo Horizonte, na coluna “CD_IBGE_CADASTRO”, teremos 3106200. Na Tabela 1, é possível visualizar a forma como recebemos os dados compilados a partir dos quais iniciamos a etapa de **desagregação de dados**.

Tabela 1: Reprodução de parte da tabela de dados compilados pelo Ministério da Cidadania

CD_IBGE_CADASTRO	ID_FAMILIA	ID_PESSOA	CO_EST_CADASTRAL_FAM	FX_RENDA	MARC_PAB	CO_SEXO_PESSOA
5002704	1	1	3	1	1	1
4108304	2	2	3	1	1	1
3304557	4	3	3	1	1	1
3550308	5	4	3	1	1	1
...

Fonte: Ministério da Cidadania, Dados Compilados da Série Histórica 2012-2021

Por que se criar códigos em vez de inserir diretamente os pala-

avras ou frase? Não seria menos trabalhoso colocar os nomes das

variáveis e já identificar “Masculino” ou “Feminino” em cada linha ao lado do nome do município em vez dos códigos? Para uma leitura final sim, mas não para o processamento e análise desses dados. É importante lembrar que qualquer coleta, tabulação e compilação de dados de uma capital brasileira é volumosa. Além disso, é indubitavelmente mais difícil digitar milhares de vezes

“Masculino” em vez de somente o número 1. O banco de dados, que gerou a série histórica sobre a população em situação de rua no Brasil, tem quase um milhão de linhas e algo aproximado a 31 milhões de células. Portanto, em vez de escrever “Masculino” com 9 caracteres, digita-se 1 e, com isso, economizam-se milhões de toques de tecla além de milhares de horas de trabalho.

4. Como os dados são desagregados?

Os dados compilados são desagregados por meio de um *software* chamado R. Escrevemos um comando parecido ao seguinte sintagma: “na coluna **CO_SEXO_PESSOA**, decodifique todo número 1 como ‘Masculino’ e o 2 como ‘Feminino’. Dessa maneira, começamos a decodificar e a desagregar os dados, ou em outras palavras, **dar nomes aos números**. Na sequência, redigimos outro comando: “na coluna **CD_IBGE_CADASTRO**,

substituir 5002704 por ‘Campo Grande no Mato Grosso do Sul’”. Repetimos esse processo para todos os municípios do Brasil. Esse é o **primeiro algoritmo**. Naturalmente, dada a quantidade de municípios no Brasil, 5.570, tivemos que identificá-los com um **segundo algoritmo** para evitar a execução manual desse procedimento em toda a série histórica. Nesse segundo algoritmo, incluímos uma tabela conhecida como *shapefile* e nela encontra-

mos uma coluna com o código do IBGE para os municípios brasileiros, outra com o nome oficial do município e uma terceira com os polígonos correspondentes à geolocalização de cada cidade. Vejamos um exemplo. Em 2021, somente para o Município de Campo Grande, Mato Grosso do Sul, há 532 registros ou linhas. E como substituir o código do IBGE pelo nome da cidade? Basta unificarmos a tabela *shapefile* à tabela com os dados sobre a população em situação de rua pela coluna que ambas têm em comum, isto é, `CD_IBGE_CADASTRO`.

Com a nova tabela de dados desagregados e decodificados para todos os municípios, conseguimos ler todas as informações sem a ajuda de dicionários. Podemos ainda contar e georreferenciar o CadÚnico de cada município e, então, elaborar gráficos, mais tabelas e mapas. **Ressaltamos que não temos acesso à identidade ou a quaisquer outras informações pessoais a partir dos registros do CadÚnico na série histórica recebida**, que permanecem protegidas e sob sigilo no Ministério da Cidadania.

5. A partir da desagregação dos dados, quantas tabelas foram geradas?

A desagregação das tabelas fornecidas pelo Ministério da Cidadania gerou outros dados sobre o fenômeno da população de rua no Brasil. Essas informações adicionais são definidas como **metadados**, pois dimensionam o conjunto dos registros, dados criados

a partir da série histórica, algoritmos elaborados e novos *shapefiles* gerados. A Tabela 2 sistematiza em duas colunas esses metadados e, na primeira linha, referenciamos a 10 tabelas fornecidas pelo Ministério da Cidadania sobre a população em situação de rua no

Brasil sendo que cada uma delas corresponde a um ano da série histórica 2012-2021. A partir dos 363 algoritmos, geramos 708 novas tabelas já com os dados desagregados, códigos identificados e, portanto, informações possíveis

de serem visualizadas por qualquer pessoa ou máquina, em formato xlsx no Excel. Os 10 *shapefiles* nela referidos concernem à geolocalização dos municípios para a elaboração de mapas.

Tabela 2: Metadados da série histórica 2012-2021 gerados com a desagregação dos dados fornecidos pelo Ministério da Cidadania

Tabelas/Algoritmos	Número de Documentos
Tabelas Utilizadas	10
Tabelas Geradas	708
Algoritmos/Códigos em R	363
Shapefiles	10

Fonte: Observatório Brasileiro de Políticas Públicas, 2022

6. Agora que já sei sobre como os dados foram desagregados, por onde começo?

O início do trabalho com dados pode se dar de diferentes maneiras, embora sugerimos que qualquer estudo comece pela compreensão do conjunto de eventos a ser estudado ou analisado. Por exemplo, se o destaque a ser dado

por uma pesquisa ou consulta for gênero, é importante conhecer os totais de “homens” e “mulheres” para que, posteriormente, informações como a proporção entre um grupo e outro possam ser observadas. Essa etapa pa-

rece dispensável, contudo revela o quanto uma série histórica pode ser significativa para qualquer pesquisa de tendência ou mesmo o levantamento de problemas sobre

o porquê, por exemplo, de mulheres representarem sempre algo em torno de 10% das pessoas em situação de rua.

Tabela 3: Metadados da Série Histórica 2012-2021 Medidos por Coluna, Linha e Células Processadas

Anos da Série Histórica 2012-2021 e o Acumulado ou Média	Total de Linhas Processadas para todos os Municípios	Número de colunas ou variáveis como "sexo", "cidade" etc	Linhas Processadas Relacionadas às Capitais Brasileiras	Total de Células Processadas (Quantidade de linhas vezes colunas)	Representação Percentual das Capitais no Banco de Dados
2012	12775	32	9039	408800	70.76%
2013	22922	32	15549	733504	67.83%
2014	37419	32	24283	1197408	64.89%
2015	52440	32	33235	1678080	61.05%
2016	73874	32	45825	2363968	62.03%
2017	101302	32	59337	3241664	58.57%
2018	138332	32	77263	4426624	55.85%
2019	174766	32	94427	5592512	54.03%
2020	194824	32	103869	6234368	53.31%
2021	158057	31	85634	4899767	54.18%
Acumulado ou Média	808654	319	462827	30776695	60.25%

Metadados do Observatório Brasileiro de Políticas Públicas com a População em Situação de Rua, 2022

Outro ponto relevante para se começar uma pesquisa com uma série histórica é saber qual o tamanho do banco de dados consultado. A Tabela 3 mostra os totais ano a ano de linhas, colunas e células processadas referentes à série histórica 2012-2021. Caso consideremos todas as células processadas pelo programa R, ao longo da década, para todos os municípios, chegaremos ao número de quase 31 milhões de células. A coluna “Representação Percentual das Capitais no Banco de Dados” mostra o impacto das capitais na série histórica. Evidenciada na Tabela, a média percentual de 60,25% é

uma informação estatística relevante para estudos que envolvam o direito urbanístico e o planejamento de políticas públicas. É importante frisar que, dos 5.570 municípios, nem todas as prefeituras coletam e tabulam dados para o Ministério da Cidadania. Portanto, os dados compilados correspondem à maior cobertura possível de cadastros sobre o fenômeno, mas não necessariamente de todos os municípios. Os dados selecionados, desagregados e apresentados pelo Observatório Brasileiro de Políticas Públicas com a População em Situação de Rua são inéditos.

7. As capitais possuem mais cadastros sobre pessoas em situação de rua?

Sim. Todas as capitais brasileiras, incluindo o Distrito Federal, possuem mais pessoas em situação de rua que os demais municípios segundo os dados provenientes da série histórica 2012-

2021. A Tabela 4 mostra tanto o número de municípios participantes do CadÚnico ao longo de uma década quanto o peso percentual das capitais na série histórica. Já a Tabela 5 resume os números ge-

rais para a última década no Brasil bem como os totais relativos às capitais brasileiras. A ordem dos dados está do maior para o

menor quantitativo e tem como base o ano de 2021.

Tabela 4: Peso (%) das Capitais na Série Histórica 2012-2021

O Cadastro Único para Programas do Governo Federal (CadÚnico) com a População em Situação de Rua no Brasil

Ano	Municípios Participantes na Coleta de Dados	Percentual das Capitais no CadÚnico
2012	607	10.90%
2013	916	16.45%
2014	1122	20.14%
2015	1267	22.75%
2016	1426	25.60%
2017	2092	37.56%
2018	2754	49.44%
2019	3208	57.59%
2020	3435	61.67%
2021	2965	53.23%

Metadados produzidos pelo Observatório Brasileiro de Políticas Públicas com a População em Situação de Rua

8. Agora que conheço melhor como os dados foram gerados, os totais para Brasil e para os municípios, como posso me aprofundar no uso de dados sobre o fenômeno da população em situação de rua?

A página do Observatório Brasileiro de Políticas Públicas com a População em Situação de Rua possui gráficos, tabelas e mapas interativos de fácil consulta. Basta navegar pelos eixos norteadores e transversais para ver como fizemos a seleção de dados.

Nos gráficos e mapas, ao passar o cursor sobre os polígonos dos municípios, aparecerão as respectivas etiquetas com informações quali-quantitativas. Já as tabelas requerem digitação do nome do município a ser consultado. Todos os dados da página são **in-**

Tabela 5: Totais de Pessoas em Situação de Rua, Série Histórica 2012-2021, Brasil e Capitais

Brasil	2012	2013	2014	2015	2016	2017	2018	2019	2020	2021
Total	12775	22922	37419	52440	73874	101302	138332	174766	194824	158057
Capital	2012	2013	2014	2015	2016	2017	2018	2019	2020	2021
São Paulo	3842	7883	13185	18608	25095	31336	38887	44372	48134	37200
Belo Horizonte	2324	3034	3839	4753	6347	8035	9700	11578	11858	9147
Rio de Janeiro	125	265	476	994	1839	3460	5659	7667	8728	8693
Distrito Federal	88	150	433	755	1651	2445	3327	4602	5280	5281
Salvador	105	274	358	491	918	1262	1840	2680	3328	4259
Fortaleza	548	713	1015	1278	1772	2265	2976	3941	4478	4182
Curitiba	373	792	1068	1338	1634	1892	2492	3227	3683	2782
Porto Alegre	775	1125	1610	1781	1979	2128	2440	3029	3208	2048
Boa Vista	5	7	7	6	4	25	1029	1695	2484	1923
Florianópolis	0	48	203	493	756	952	1131	1561	1720	1314
Recife	36	54	116	202	283	449	628	859	1003	1087
Goiânia	46	61	148	301	518	857	1160	1367	1430	1035
Manaus	30	58	86	117	152	276	448	899	959	805
São Luís	0	43	111	169	207	327	626	911	1066	770
Natal	128	151	192	279	438	595	750	855	898	736
Cuiabá	4	62	260	318	383	460	632	722	814	724
Aracaju	83	129	217	243	296	414	591	744	785	643
Campo Grande	10	33	54	112	209	309	503	603	655	532
Vitória	113	160	213	268	331	383	512	626	681	522
Teresina	0	35	61	113	173	258	355	501	529	502
Maceió	356	366	372	401	555	694	821	913	958	480
Belém	4	33	71	97	141	220	296	416	458	420
João Pessoa	12	30	44	52	59	93	125	178	203	160
Rio Branco	31	37	47	53	56	114	176	213	215	136
Porto Velho	1	3	3	4	11	22	48	109	134	112
Palmas	0	3	8	5	5	46	82	123	133	99
Macapá	0	0	86	4	13	20	29	36	47	42

Fonte: Ministério da Cidadania, Série Histórica 2012-2021
Totais Calculados pelo Observatório Brasileiro de Políticas Públicas com a População em Situação de Rua

teroperáveis, ou seja, há compartilhamento e integração de informações entre as tabelas geradas e as recebidas. Isso por razão de os algoritmos estarem programados para fazer o recorte e a análise da série histórica sempre

com base nos dados recebidos do Ministério da Cidadania. Portanto, um mesmo dado pode ser encontrado em lugares diferentes com uma ênfase distinta.

“ Todos os dados da página são **interoperáveis**, ou seja, há compartilhamento e integração de informações entre as tabelas geradas e as recebidas. ”

9. Posso baixar os dados que visualizo nas tabelas, gráficos e mapas?

Sim. Em cada gráfico, tabela e mapa, é possível ver um link chamado “Obter dados” logo abaixo da imagem. Ao clicar nele, os arquivos em forma de tabela de Excel com os dados desagregados começam a ser baixados. Outra ferramenta útil é o link “Download da imagem”. Ela permite

que o usuário faça o *download* das imagens de interesse em formato PNG. É importante citar tanto o Ministério da Cidadania, dados agregados, quanto o Observatório Brasileiro de Políticas Públicas com a População em Situação de Rua, dados desagregados.

10. O que são os eixos norteadores e transversais que vejo no site? Por que são importantes?

Os eixos norteadores apresentam os dados da série histórica em seis facetas: i) Cidadania, Assistência Social e Saúde; ii) Moradia; iii) Infâncias, Adolescências e Juventudes; vi) Mulheres, Gêneros e Famílias; v) Justiça e iv) Trabalho e Tecnologias Sociais. Caso exista interesse em saber, por exemplo, o número de mulheres em situação de rua, é possível encontrar esse quantitativo e outros dados para todas as capitais e municípios do país no **Eixo Norteador Mulheres, Gêneros e Famílias**. Se além dos totais para cada município, busca-se o total de mulheres pretas e pardas em situação de rua, esses dados estão disponíveis no **Eixo Transversal Violências estruturais: racismo, gênero e capital**. Em síntese, os eixos norteadores organizam os dados em totais e georreferenciamento temático, às vezes em *ranking* de capitais; por outro lado, os eixos transversais

cruzam os dados da série histórica a partir de dois ou mais temas. O objetivo do cruzamento de dados da série histórica é aproximar o universo dos números das teses defendidas no mundo acadêmico ou recorrentes na opinião pública. Esse acercamento também nos auxilia na produção de conhecimento com maior precisão e, portanto, damos um passo além de formulações hipotéticas puramente qualitativas ou quantitativas. Muito embora inúmeros estudos qualitativos e quantitativos produzidos no Brasil tenham relevo nacional e internacional, destaque para institutos, universidades e terceiro setor empenhados em compreender fenômenos sociais, políticos e jurídicos, é cada vez maior a necessidade de se formar novas linhas de pesquisa cujo enfoque supere o isolamento disciplinar e possibilite análises quali-quantitativas. Exemplo disso é a tese do racismo estrutural. Se

essa contribuição é de fato importante, ao analisarmos e cruzarmos os dados no **Eixo Transversal Violências estruturais: racismo, gênero e capital**, os números deverão dialogar com as reflexões e estudos sobre racismo no país. E, com base nos dados, isso se verifica. Segundo a série histórica do Ministério da Cidadania, quando desagregada pelo Observatório Brasileiro de Políticas Públicas com a População em Situação de Rua, podemos identificar não apenas os totais de homens entre os cadastros de indivíduos em situação de rua, como também notar que 83,6% dos homens nessa situação de vulnerabilidade em Belo Horizonte em 2021, por exemplo, eram pretos ou pardos. Já entre as mulheres, esse percentual era de 83,4% para o mesmo ano. Isso nos possibilita a formulação de outros questionamentos e hipóteses, ou seja, parece existir uma variável para além de

uma suposta conjuntura determinada pela condição de pobreza, concentração de renda e desemprego em um dado período da vida dessas pessoas afetadas. O caso concreto é que já temos, por exemplo, base suficiente para testar a hipótese de que o fenômeno de pessoas em situação de rua é mais causado por fatores intergeracionais associados ao racismo do que perda de renda, uso de drogas ilícitas, laços familiares fragilizados ou outros arquétipos que mais explicam uma parte da equação do que a resolvem por inteiro. Não dispensamos as variáveis renda, uso abusivo de substâncias, conflitos familiares etc, mas é preciso entender que o fenômeno da população em situação de rua é gestado no tempo e a pobreza passada de progenitores para filhos parece ser mais estruturante entre pretos e pardos do que entre brancos.