

MODELAGEM DA SITUAÇÃO URBANO-RURAL POR REGRESSÃO LOGÍSTICA DO USO E COBERTURA DA TERRA: O CASO DO DISTRITO FEDERAL NO CENSO BRASILEIRO 2022

RESUMO

Até o Censo 2010, a situação urbano-rural era definida por limites administrativos e legislações municipais. Para o Censo 2022, o IBGE utilizou imagens orbitais de alta resolução na predefinição da tipologia, que foi ajustada em campos pelos pesquisadores, integrando o conhecimento empírico ao sensoriamento remoto. O presente estudo teve como objetivo caracterizar os setores urbanos e rurais da malha censitária frente às categorias de uso e cobertura da terra provenientes do MapBiomas. Tendo como área de análise o Distrito Federal, foi utilizada a técnica de regressão logística como a variável dependente a tipologia urbano-rural e como as variáveis independentes os percentuais de áreas de vegetação, agropecuária e não vegetada. O modelo, que teve acurácia de 97,3%, mostra que o percentual de área não vegetada tem o maior peso para a tipificação de um setor como urbano. Os resultados podem ser usados para prever a tipologia de áreas até a realização do próximo Censo, bem como para se planejar outras pesquisas amostrais em regiões que apresentem mudança no uso e cobertura da terra, usando apenas dados já disponíveis pelo MapBiomas.

Palavras-chave: Sensoriamento remoto. Assentamentos humanos. Ordenamento territorial. MapBiomas. Geoprocessamento.

MODELING THE URBAN-RURAL TYPOLOGY USING LOGISTIC REGRESSION OF LAND USE AND LAND COVER: THE CASE OF THE FEDERAL DISTRICT IN THE 2022 BRAZILIAN CENSUS

ABSTRACT

Until the 2010 Census, IBGE defined the urban-rural typology based on administrative boundaries and municipal legislation. For the 2022 Census, IBGE used high-resolution orbital images to predefine the typology, which was then refined in the field by researchers, integrating empirical knowledge with remote sensing. This study focused on the Federal District and aimed to characterize the urban and rural sectors of the census grid in relation to land use and land cover categories from MapBiomas. The method employed was logistic regression, with urban-rural typology as the dependent variable and the percentages of vegetated, agricultural, and non-vegetated areas as independent variables. With an accuracy of 97.3%, the model indicates that the percentage of non-vegetated area is the strongest predictor of a sector being classified as urban. The findings can be used to predict area typology until the next Census, as well as to support the planning of other sample surveys in regions undergoing land use and cover changes, relying solely on available MapBiomas data.

Keywords: Remote sensing. Human settlements. Territorial planning. MapBiomas. Geoprocessing.

Comentado [A1]: Se não der tempo de submeter novamente pra Caminhos UFU, precisa ver como fica o título do MapBiomas

Comentado [A2R1]: Verificar "tipologia" em todo o documento

Formatado: Realce

Comentado [A3]: Rever

Comentado [A4]: Rever

INTRODUÇÃO

A distinção entre territórios urbanos e territórios rurais serve para orientar políticas públicas, distribuir recursos, e planejar infraestruturas e serviços. Os estudos e debates sobre ordenamento territorial não são novos, porém o uso de dados de sensoriamento remoto levanta aspectos ainda pouco explorados neste campo. Até o Censo de 2010, o Instituto Brasileiro de Geografia e Estatística (IBGE) utilizava os limites administrativos e as legislações de ordenamento territorial para definir a situação urbano-rural dos setores censitários. No Censo de 2022, uma nova abordagem agregando imagens orbitais de alta resolução foi incorporada à metodologia existente (IBGE, 2024).

A situação urbano-rural é definida como a categorização dicotômica, urbana ou rural, que os setores censitários recebem do IBGE na elaboração de malha territorial para elaboração de coletas de campo, em especial o Censo Demográfico. No Censo 2022, a situação urbano-rural se dá pelo agrupamento de situações detalhadas, sendo a urbana composta por área urbana de alta densidade de edificações, área urbana de baixa densidade de edificações e núcleo urbano; enquanto a situação rural é composta de situação detalhadas como povoado, núcleo rural, lugarejo, área rural (exceto aglomerados) e massas de água (IBGE, 2024).

As definições sobre as situações são elaboradas com base em procedimentos analíticos de caráter morfológico-funcional que consideram aspectos políticos, econômicos, sociais e ambientais relacionados aos agregados estatísticos (IBGE, 2024). Segundo Souza *et al.* (2024),

O trabalho de revisão metodológica da Base Territorial que amparou o Censo de 2022 buscou compatibilizar demandas legais, conceituais e operacionais, baseada numa revisão de processos de trabalho e uso de novas tecnologias. Uma vez priorizada a análise morfológica, a classificação dos setores censitários necessitava de revisão, especialmente naquelas situações mais afetadas pela desatualização ou imprecisão da legislação local de perímetro ou zoneamento urbano, confrontadas com imagens recentes de alta resolução. As imagens também foram úteis na revisão dos limites de setores, sobretudo nos casos de linhas secas ou imaginárias, toleradas apenas em razão de limites político-administrativos ou nos casos de setores de aglomerados rurais onde não foi possível identificar elementos físicos estáveis, como rios, estradas etc. (Souza *et al.*, 2024, p. 8).

Sendo assim, o conhecimento empírico dos técnicos do IBGE foi integrado à interpretação das imagens orbitais pela primeira vez que, configurando um processo de validação cruzada entre observação de campo e dados de sensoriamento remoto. Por outro lado, tanto na nota técnica sobre a malha censitária IBGE (2024) quanto em artigo em que técnicos do IBGE descrevem os procedimentos (Souza *et al.*, 2024), não há detalhamento sobre o processamento das imagens de satélite nem a sistematização da extração de informações provenientes do sensoriamento remoto.

A saber, no contexto do Censo 2010, Pera e Bueno (2016) já haviam identificado que as informações contidas nos cadernos metodológicos do IBGE não eram suficientes para a compreensão do método usado naquele momento, e solicitaram mais informações ao setor técnico do instituto. As autoras concluíram que a elaboração das situações detalhadas possui caráter híbrido, pautando-se tanto em legislações urbanísticas quanto na realidade instaurada, porém não foi explicitado pelo IBGE como e em quais municípios a realidade instaurada foi utilizada. No caso do uso das imagens orbitais em 2022, a principal inovação parece residir na análise prévia das imagens em escritório, antes das vistorias presenciais nos territórios. Todavia, não foi descrito como é feita a validação no território das características coletadas pelas imagens.

Vale notar que as discussões entre urbano e rural ou cidade e campo pautam-se geralmente em características sociais, como população, densidade populacional e relações de trabalho (Araújo, Soares, 2009; Santos, Silva, 2020). Entretanto, ao incluir as imagens orbitais e produtos oriundos de sensoriamento remoto, abre-se a possibilidade da inserção de aspectos ambientais, bem como as características de uso e cobertura da terra, na distinção entre urbano e rural. O projeto MapBiomias (Projeto Brasileiro de Mapeamento Anual de Uso e Cobertura do Solo) vem produzindo informações de uso e cobertura da terra desde 2015. Tais dados vêm amplamente sendo utilizados para apoiar o desenvolvimento de projetos governamentais ou iniciativas de apoio a gestão (Brito, 2017).

Nesse contexto, o presente trabalho tem como objetivo modelar a situação urbano-rural a partir de categorias de uso e cobertura da terra provenientes de sensoriamento remoto. O modelo consiste em

Comentado [A5]: Dei uma reestruturada na intro... Favor avaliar! Acredito que há espaço para agregar mais informações e deixá-la mais extensa.

Comentado [A6]: Isto não está adicionando nada aqui, mas vai servir para dizer que trabalhos futuros podem usar outras modelagens

Comentado [A7]: Brito, B. 2017. Potential trajectories of the upcoming forest trading mechanism in Pará State, Brazilian Amazon, PLoS ONE, 12: e0174154.

Comentado [A8R7]: Melhor: Assessment of TerraClass and MapBiomias data on legend and map agreement for the Brazilian Amazon biome Alana Kasahara NEVES1*, Thales Sehn KÖRTING1, Leila Maria Garcia FONSECA1, Maria Isabel Sobral ESCADA

uma regressão logística, que fornece a probabilidade de um setor censitário ser tipificado como urbano ou rural, a partir de categorias de usos e coberturas naturais ou antrópicos do MapBiomias (2024).

Do ponto de vista socioambiental, este estudo se justifica visto que a situação de áreas urbanas ou rurais influencia diretamente a formulação e implementação de políticas públicas, afetando a alocação de recursos, a infraestrutura e o acesso a serviços essenciais. A Agenda 2030, por meio do Objetivo de Desenvolvimento Sustentável 11 (ODS 11), enfatiza a necessidade de tornar as cidades e os assentamentos humanos inclusivos, seguros, resilientes e sustentáveis (Brasil, 2025). No entanto, a efetividade dessas metas depende de definições dos limites urbanos e rurais, uma vez que essa diferenciação influencia indicadores-chave, alocação de recursos e estratégias de planejamento territorial. A avaliação *ex-ante* dessas políticas possibilita a análise de seus potenciais impactos e a mitigação de distorções no desenvolvimento territorial, assegurando maior eficiência nos investimentos (Casa Civil; Ipea, 2018). A complexidade dessa discussão ganha ainda mais relevância diante aos desafios enfrentados pela sociedade contemporânea, como o acelerado processo de urbanização e os impactos das mudanças climáticas, que ampliam a vulnerabilidade e os riscos especialmente em área com menor capacidade de adaptação (Dodman *et al.*, 2023).

Do ponto de vista técnico-acadêmico, o ineditismo do uso de imagens de satélite para o delineamento da coleta de campo pelo IBGE justifica um estudo para discutir abordagens indiretas e por vezes mais baratas em termos de coleta e de processamento de dados. Em relação à aplicação prática, o estudo pode ajudar a prever a chance de uma área ser urbana a partir de dados de sensoriamento remoto já processados e disponíveis, auxiliando na elaboração e execução de pesquisas em campo até o próximo Censo.

MATERIAL E MÉTODOS

Este estudo adota uma abordagem quantitativa e um desenho não-experimental conforme definição de (Creswell; Creswell, 2017). Em termos de métodos de pesquisa, caracteriza-se como uma pesquisa descritiva, de acordo com Gil (2002), pois se concentra na descrição de um fenômeno e o estabelecimento da relação entre as variáveis. Das práticas de Sistemas de Informações Geográficas (SIG), vale-se de planimetrias e assinaturas, considerando o SIG um hiperespaço heurístico para se informar empiricamente sobre possíveis associações causas entre as variáveis ambientais (Xavier-da-Silva; Carvalho-Filho, 1993).

Área de estudo

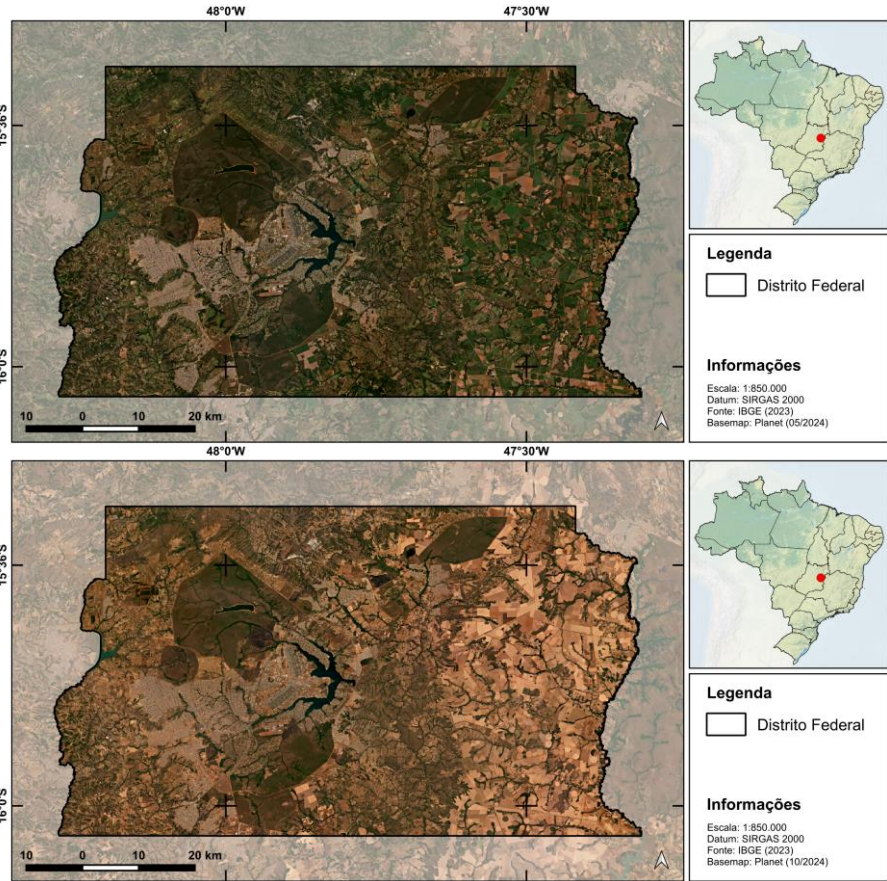
A área de estudo escolhida foi o Distrito Federal. De acordo com IBGE (2024) e Souza et al. (2024), a atualização da base cartográfica para o Censo Demográfico de 2022 foi conduzida de forma descentralizada, com participação ativa das Superintendências do IBGE e suas respectivas agências. Segundo os autores, esse modelo permitiu uma representação mais precisa das referências geográficas que subsidiaram as operações de campo dos censos e pesquisas amostrais, assegurando uma correta identificação das estruturas territoriais por meio da integração de registros administrativos, imagens de sensoriamento remoto, interpretação de legislação e informações coletadas em campo.

Nesse contexto, o Distrito Federal destaca-se como uma unidade federativa singular que não possui subdivisões municipais, apresentando uma estrutura administrativa com as competências legislativas reservadas aos Estados e municípios (Brasil, 1988). Essa característica favorece a adoção de abordagens mais homogêneas na definição sobre a situação urbano-rural. Dessa forma, para reduzir fontes de variância no modelo, optou-se por delimitar a área de estudo conforme os limites dos setores censitários localizados no Distrito Federal, conforme apresentado na Figura 1.

Figura 1 – Distrito Federal, localização da área de estudo

Comentado [A9]: Verificar se precisar estar em itálico

Comentado [A10]: Decida qual imagem gostou mais kkk:
1-maio/24
2-out/24



Fontes: IBGE, 2022. Elaboração: Os autores, 2025.

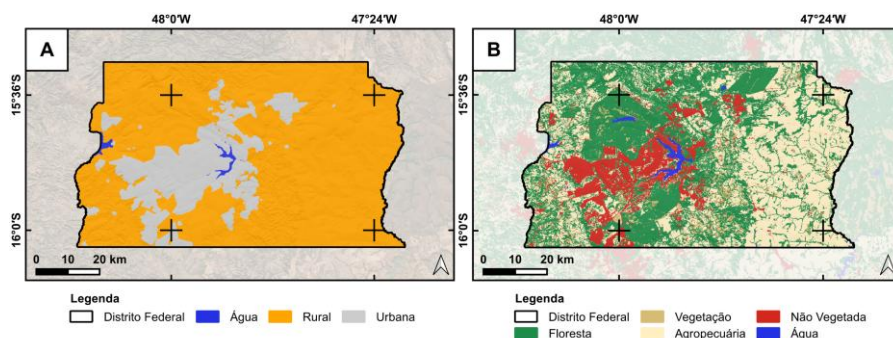
Com uma área de 5.761 km² e uma população estimada de 2.817.381 habitantes (IBGE, 2025), o Distrito Federal está localizado na região Centro-Oeste e é organizado em 35 regiões administrativas (SEGOV, 2025). Inserido no bioma Cerrado, o território do Distrito Federal apresenta uma dinâmica espacial marcada pela intensa urbanização e acelerado crescimento populacional, impulsionados pelo desenvolvimento da capital federal. A área urbana do Distrito Federal expandiu aproximadamente 2.983% entre 1960 e 2019, passando de 23 km² para 709 km², com um padrão de crescimento direcionado do centro para o sudoeste e, em menor escala, para o nordeste. A variação da taxa de crescimento do Distrito Federal ao longo dos 60 anos aponta para uma aceleração da urbanização, com destaque para o crescimento a partir de 2013, o que indica que o Distrito Federal está longe da estagnação ou redução do crescimento, conforme é esperado na maturidade de centros urbanos (Chelotti; Sano, 2023). Isto significa que se pode esperar mudanças de uso e cobertura da terra nos próximos anos e décadas.

Dados utilizados

A malha dos setores censitários do Distrito Federal, utilizada para levantamento do censo de 2022 (Figura 2A), foi obtida a partir do repositório oficial do IBGE (IBGE, 2022). Essa malha totalizou 5.418 setores censitários, dos quais 5.021 são urbanos, 393 rurais e 4 sem situação urbano-rural. Esses últimos, por estarem inseridos em massas de água e sem as respectivas situações censitárias, foram desconsiderados no modelo de análise. Vale destacar que essa malha foi atualizada em períodos anteriores à data de início do censo, sendo adotado neste estudo o ano de 2021 como base temporal para as investigações (Souza *et al.*, 2024).

Os dados de uso e cobertura da terra (Figura 2B) foram adquiridos a partir da plataforma MapBiomas para o ano de 2021, especificamente a Coleção 9 com 30 metros de resolução espacial (MapBiomas, 2024; Souza *et al.*, 2020). Ressalta-se que foram consideradas apenas as categorias correspondentes ao nível 1 de processamento da respectiva coleção, sendo: florestas, formação natural não florestal, agropecuária, área não vegetada e corpos d'água.

Figura 2 – (A) Malha dos setores censitários do Distrito Federal e (B) Uso e cobertura da terra em 2021 do Distrito Federal



Fontes: MapBiomas, 2024; IBGE, 2022. Elaboração: Os autores, 2025.

Métodos

Para o processamento dos dados geoespaciais, foi utilizado o SIG de código aberto QGIS 3.34.15 'Prizren' (QGIS, 2025). As bases de dados foram recortadas e ajustadas conforme os limites dos setores censitários do Distrito Federal. Além disso, para realização dos cálculos de área, os dados foram convertidos para o sistema de coordenadas Albers Equal Area Conic (Reference, 2025).

Os dados de uso e cobertura da terra foram convertidos do formato matricial para vetorial e, em seguida, realizaram-se a união das respectivas bases de dados e o cálculo de área das unidades vetoriais. Após a integração dos dados, estes foram exportados em formato tabular para tratamento em planilha eletrônica. Nesse ambiente, foram calculadas as estatísticas descritivas gerais e feitos testes de previsão do modelo.

Para a regressão logística, o conjunto de dados foi dividido em amostras de treinamento e validação, sendo 80% e 20% respectivamente, e as análises foram então conduzidas no software livre RStudio 2024.12.0+467 (Posit, 2024) utilizando a linguagem de programação R (versão 4.4.2) (R Core Team, 2024). A importação dos dados foi realizada com a biblioteca readxl (Wickham *et al.*, 2023), enquanto a análise exploratória foi conduzida com a biblioteca GGally (Schloerke *et al.*, 2024). A modelagem estatística baseou-se na biblioteca stats (RStudio Team, 2019), utilizando o modelo de regressão logística via glm().

Para avaliação da qualidade do modelo foram utilizadas as bibliotecas pscl (Jackman *et al.*, 2024) que fornece uma adaptação para o coeficiente de determinação (R^2) de McFadden. Diferentemente de um modelo linear, a regressão logística não tem uma única medida síntese da variação na variável dependente explicada pelo modelo, tal como o coeficiente de determinação (Fernandes *et al.*, 2020).

O R² de McFadden mede a redução do erro quando o modelo inclui variáveis independentes. Ele compara o desvio residual (erro no modelo ajustado) com o desvio nulo (erro ao prever a variável dependente sem variáveis independentes). Quanto maior o R² de McFadden, melhor o ajuste do modelo, pois indica que ele explica mais a variação da variável dependente.

A validação do modelo com ponto de corte em 50% e construção da matriz de confusão foram realizadas com a biblioteca caret (Kuhn *et al.*, 2024) enquanto a performance do modelo foi avaliada com a biblioteca pROC (Robin *et al.*, 2023) por meio da curva ROC (*Receiver Operating Characteristic*) e AUC (Área Sob a Curva). Por fim, a visualização dos resultados por meio de gráficos foi desenvolvida com a biblioteca ggplot2 (Wickham *et al.*, 2024).

RESULTADOS E DISCUSSÃO

Esta seção apresenta a análise descritiva dos dados, a elaboração do modelo e discute os resultados com base na literatura.

Análise descritiva

Setores Censitários do Distrito Federal (IBGE – bruto)

Numero, tamanho (mínimo, máximo, médio, desvio padrão, etc)

No Distrito Federal, os setores urbanos correspondem a 16,6% da área.

Setores Censitários do Distrito Federal associados ao MapBiomass

Em relação a cada categoria de uso e cobertura da terra, encontram-se em setores urbanos 8,9% das florestas, 7,7% das áreas vegetadas, 5,0% das áreas de agropecuária e 83,2% das áreas não-vegetadas (Tabela 1). Estes resultados estão em concordância com Sudhira, Ramachandra, Jagadish (2004), que indicam que áreas construídas são geralmente o principal indício de urbanização de áreas urbanas.

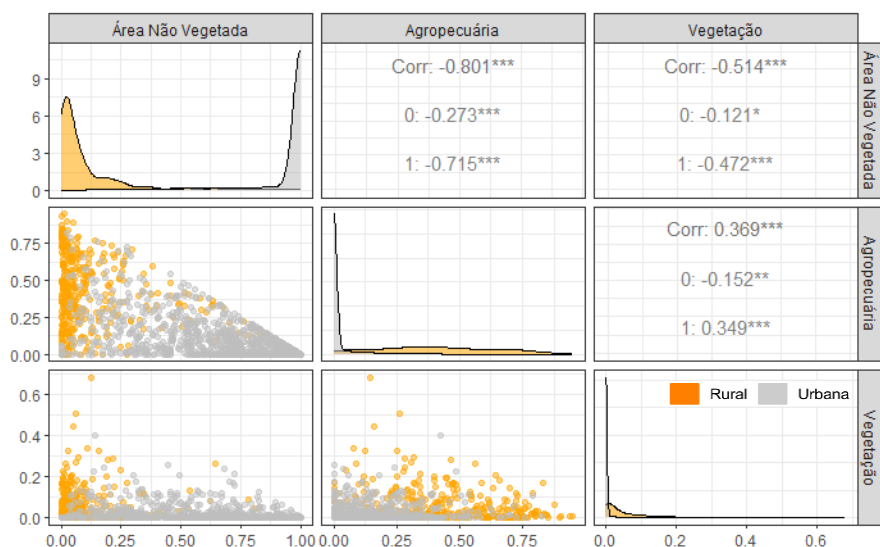
Tabela 1 - Situação e área de cobertura da terra em hectares

Situação	Setores (un)	Área (ha)	Floresta (ha)	Vegetação (ha)	Agropecuária (ha)	Não-vegetada (ha)
Rural	393 (7,3%)	475.155 (83,4%)	225.501 (91,1%)	25.405 (92,3%)	212.265 (95,0%)	11.983 (16,8%)
Urbana	5021 (92,7%)	94.472 (16,6%)	21.990 (8,9%)	2.133 (7,7%)	11.068 (5,0%)	59.281 (83,2%)
TOTAL	5.414	569.627	247.492	27.538	223.333	71.264

Fontes: MapBiomass, 2024; IBGE, 2022. Elaboração: Os autores, 2025.

A matriz de correlação e dispersão apresentada na Figura 3 permite uma análise detalhada das relações entre as três variáveis, evidenciando seus padrões de distribuição. Na figura, são indicados os níveis de significância, sendo **** para p-valor menor que 0,001, *** menor que 0,01, e **, menor que 0,1.

Figura 3 - Matriz de dispersão e correlação entre as variáveis por situação



De modo geral, as correlações entre áreas não vegetadas e as vegetações mostram-se negativas e moderadas (R de -0,514**), indicando, conforme esperado, que os setores com maior cobertura vegetal apresentam, proporcionalmente, menores extensões de áreas não vegetadas. Ao segmentar as análises por categorias de ocupação (0 = rural, 1 = urbano), observa-se que essa relação se intensifica nos setores censitários urbanos (R de -0,472**), onde a conversão de vegetação e superfícies não vegetadas é mais evidente. Já nos setores rurais, essa correlação é fraca e pouco significativa (R de -0,121*), sugerindo que outros fatores podem influenciar a variação entre essas categorias.

As correlações entre áreas não vegetadas e agropecuária apresentam-se negativas e fortes (R de 0,801***), evidenciando que o aumento de uma dessas categorias tende a reduzir a outra. Por outro lado, a relação entre agropecuária e vegetação se mostra positiva, variando de fracas a moderadas (R de 0,369**), sugerindo que em alguns setores pode haver coexistência entre essas duas categorias. No contexto urbano, essa relação permanece positiva e moderada (R de 0,349**), reforçando a possibilidade de setores que ainda mantêm remanescentes vegetais junto a áreas de uso agropecuário. No entanto, nos setores rurais, a correlação se torna negativa e fraca (R de -0,152**), indicando que a expansão da agropecuária pode estar associada à conversão direta de vegetação em áreas produtivas, refletindo padrões típicos de uso da terra.

Quanto às distribuições das categorias analisadas, observa-se que as áreas não vegetadas apresentam distribuição bimodal, indicando a presença de dois grupos distintos de setores: um com baixa e outro com alta proporção dessa categoria. Já a agropecuária exibe uma distribuição dispersa, com setores censitários variando desde pequenas até grandes proporções de cobertura, refletindo a diversidade de uso nesta categoria. Por fim, a vegetação apresenta uma alta concentração de valores próximos a zero, indicando que a maioria desses setores possui cobertura vegetal reduzida, reforçando a tendência de fragmentação da vegetação remanescente dentro da área de estudo.

Modelo de Regressão Logística

A regressão logística é uma técnica estatística que modela a relação entre uma variável dependente binária (rural ou urbano) e variáveis independentes. No caso em questão, as quatro variáveis - floresta, vegetação, agropecuária e não vegetada - somam 100% da área, resultando em colinearidade perfeita e inviabilizando a regressão. Para resolver essa limitação, uma das variáveis foi excluída, sendo

escolhida a variável “floresta”. Esta decisão foi tomada de forma dedutiva, considerando que, entre as quatro variáveis, a cobertura florestal é a menos suscetível a ser confundida especificamente com áreas urbanas. Em outras palavras, a priori, é menos provável que uma floresta seja interpretada como cidade, diferentemente de outras formas de uso e cobertura da terra que podem gerar maior ambiguidade na distinção entre urbano e rural. Vale ressaltar que a área de floresta no Distrito Federal são 247.492 hectares, correspondendo a 43,4% do território.

O modelo de regressão logística foi desenvolvido com a amostra de treinamento. Os resultados foram significativos ao nível de 1%, e revelam padrões na relação entre as categorias de uso e cobertura da terra e as situações dos setores censitários. A qualidade do modelo foi avaliada utilizando o R^2 de McFadden. O desvio residual foi 587,4 e o desvio nulo foi 2.211,6, resultando em $R^2 = 0,734$, o que demonstra que o modelo ajustado reduz a incerteza da variável resposta.

Conforme é apresentado na Tabela 2, o percentual de áreas não vegetadas (PercNãoVegetada) apresentou um impacto positivo significativo no modelo, indicando que os setores censitários com maior cobertura dessa categoria têm maior probabilidade de serem classificados como urbanos. Em contrapartida, o percentual de áreas de agropecuária (PercAgropecuária) e de vegetação (PercVegetação) tiveram efeitos negativos, sugerindo que os setores com maior cobertura têm menor probabilidade de serem do tipo urbano. A análise dos módulos dos coeficientes indica que em ordem de importância, as variáveis são área não vegetada ($\beta = 7,1277$, $p < 0,001$), vegetação ($\beta = -5,4206$, $p = 0,005$) e agropecuária ($\beta = -3,6621$, $p < 0,001$).

Tabela 2 - Coeficientes do modelo

Informações	Estimativa	Erro padrão	Valor z	Valor p
Intercepto	-0,2673	0,2566	-1,042	0,29742
PercNãoVegetada	7,1277	0,4708	15,138	< 2e-16***
PercAgropecuária	-3,6621	0,5445	-6,726	1,75e-11***
PercVegetação	-5,4206	1,9420	-2,791	0,00525**

Significância: *** 0,001; ** 0,01

Fonte: Os autores, 2025.

A Equação 1 apresenta o modelo ajustado, onde p é a probabilidade de um setor ser urbano e \ln é o logaritmo neperiano. O intercepto não foi incluído na equação porque sua estimativa não foi significativa ao nível de 1%. Para calcular a probabilidade de um setor ser urbano, primeiramente multiplica-se cada variável pelos seus respectivos coeficientes e soma-se esses valores. Em seguida, aplica-se a exponenciação usando o número de Euler (e), o que gera uma razão entre a probabilidade de ser urbano e a probabilidade de não ser. Por fim, esta razão é dividida pela soma dela mesma com uma unidade. Exemplos de resultados de cálculos são mostrados na Tabela 3.

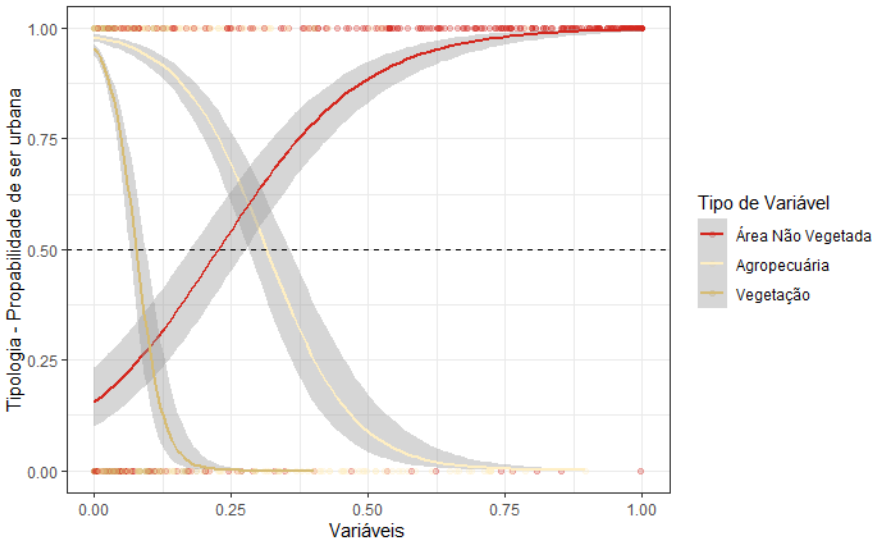
$$\ln \left(\frac{p}{1-p} \right) = 7,1277 \times \text{PercNãoVegetada} - 3,6621 \times \text{PercAgropecuária} - 5,4206 \times \text{PercVegetação} \quad (1)$$

Em relação à validação, o modelo confirmou uma alta capacidade preditiva, com uma acurácia geral de 97,3% na amostra de teste. Considerando o ponto de corte do modelo em 50%, o índice Kappa foi de 0,8216, indicando alto grau de concordância e reforçando a confiabilidade das previsões. Vale destacar que, perante a totalidade dos 5.414 setores censitários do Distrito Federal, o modelo aqui proposto tipifica cerca de 129 setores de forma distinta das definições do IBGE, ou seja, 97,6% apresentam êxito nas tipificações.

A análise das métricas de desempenho evidenciou um equilíbrio na identificação correta da situação. A sensibilidade de 85,0% demonstra a capacidade do modelo em classificar corretamente setores urbanos, enquanto a especificidade de 98,4% confirma sua eficácia na identificação de setores rurais. A acurácia balanceada de 91,7% sugere que o modelo mantém um desempenho satisfatório mesmo em cenários com possíveis desequilíbrios entre as situações. Além disso, os valores preditivos positivos (82,2%) e negativos (98,7%) indicam um baixo índice de erro nas previsões, minimizando tanto o número de falsos positivos quanto o de falsos negativos. A estatística de área sob a curva ROC (AUC) de 0,9852 reforça a capacidade discriminativa do modelo.

A Figura 4, elaborada com base na aplicação do modelo na amostra de validação, ilustra a distribuição das proporções das variáveis utilizadas no modelo (eixo horizontal) em relação ao tipo de setor censitário (eixo vertical), onde o valor zero (0) representa setores rurais e um (1) representa setores urbanos. Observa-se que os setores urbanos tendem a apresentar valores mais elevados de área não vegetada, o que corrobora o impacto positivo dessa variável no modelo. Em contraste, os setores rurais demonstram proporções mais altas de agropecuária e vegetação, alinhando-se com os coeficientes negativos dessas variáveis.

Figura 4 - Distribuição das proporções das variáveis preditoras em relação ao tipo de setor censitário



Fontes: Os autores, 2025.

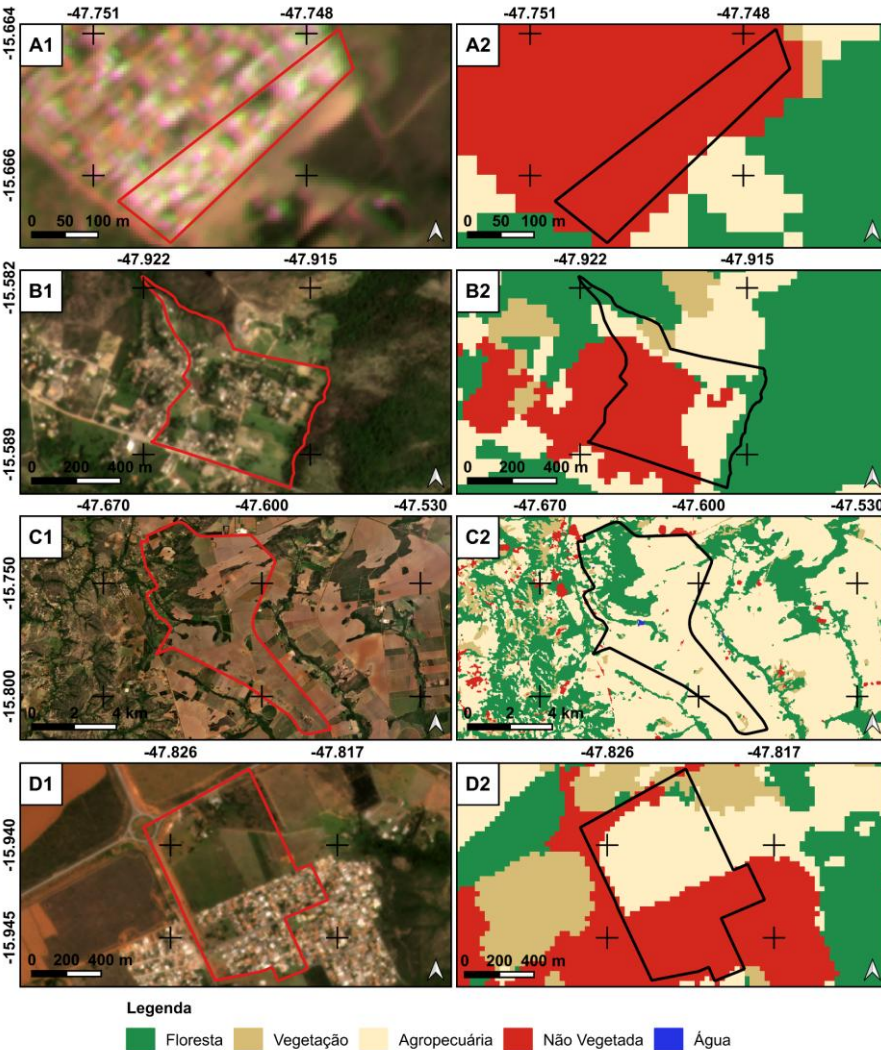
A seguir, a Tabela 3 e a Figura 5 ilustram quatro setores, evidenciando sua ocupação, os respectivos valores de categorias de uso e cobertura da terra e a probabilidade de ser urbana conforme determinada pelo modelo. Observa-se que os setores A e B tiveram suas situações definidas como rural no Censo 2022, no entanto, o modelo indica probabilidades de 99,9% e 89,0% de serem urbanas respectivamente. As imagens de satélites utilizadas na ilustração, identificadas como A1, B1, C1 e D1, foram capturadas em novembro de 2021 e fornecem uma visão detalhada da ocupação nas respectivas datas (Planet, 2025). Imagens de dezembro de 2021 não foram empregadas devido à presença de nuvens, o que poderia comprometer as análises.

Tabela 3 - Informações dos setores censitários demonstrados na Figura 5 e suas respectivas probabilidades de ser urbano conforme estimado pelo modelo

ID	Situação	Identificação	Não vegetada	Agropecuária	Vegetação	Prob. Urb.
A	Rural	530010805100441	98,1	1,9	0	99,9
B	Rural	530010805390021	49,5	34,2	3,5	89,0
C	Rural	530010805110393	0,5	74,4	2,1	5,7
D	Urbana	530010805400051	50,0	45,6	1,6	85,9

Fontes: Os autores, 2025.

Figura 5 - Representação visual de setores censitários analisados destacando a situação em novembro de 2021 e as categorias de uso e cobertura da terra.



Fontes: Planet, 2025; MapBiomias, 2024; IBGE, 2022. Elaboração: Os autores, 2025.

Discussão

Não há consenso sobre a definição dos conceitos e a mensuração dos limites e da expansão dos assentamentos humanos (Dadashpoor; Shahhossein, 2024), o que dificulta comparações e a formulação de políticas públicas uniformes, influenciando inclusive no ranking de maiores cidades (Taubenböck *et al.*, 2019). Em países como Espanha, Portugal, Itália e Grécia, por exemplo, são

Comentado [A11]: Inserir nova figura e legenda. Manter a cor da linha do polígono nas duas imagens. Azul deve funcionar. Inserir coordenadas na figura da direita, apesar de redundante.

consideradas áreas rurais aquelas localidades que, embora apresentem assentamentos humanos com pelo menos 10 mil habitantes vivendo em habitações contínuas, mantêm uma distância dos centros metropolitanos e de suas áreas de influência direta. Por outro lado, países como França, Argentina, Bolívia, México adotam critérios mais rígidos classificando como rurais as comunidades que possuem menos 2,5 mil habitantes (Araújo).

O Brasil adota a situação urbano-rural como padrão para a elaboração de estatísticas e políticas públicas. Essas definições são importantes para se entender os processos de urbanização. Em 2010 no Brasil, 84,4% da população vivia em setores censitários classificados como urbanos, e esse número passou para 85,9% no Censo de 2022 (IBGE, 2024).

Por outro lado, diversos autores analisam e criticam a formulação dicotômica empregada no Brasil, apontando que pode existir um pensamento contínuo entre rural e urbano, e como a situação urbano-rural mascara aspectos da diversidade de formas de se habitar o território (Veiga, 2002; Araújo, Soares, 2009; Santos, Silva, 2020).

EXPLICAR COMO O MODELO CONVERSA OU É OPOSTO DESTE DEBATE

Este trabalho inova ao trazer a questão ambiental na distinção entre urbano e rural. O que é uma cidade? Características antrópicas. Impacto ambiental para além do social.

Corte 50%

Discussão conceitual

Estudos frequentemente utilizam técnicas de sensoriamento remoto para delimitar áreas urbanas, como por mapeamento de ilhas de calor (Chen et al., 2005) e pela análise de luz noturna (Zheng et al., 2023).

A partir das estatísticas apresentadas e dos casos exemplificados, o modelo se mostra pertinente para auxiliar a elaboração de pesquisas, visando principalmente a superação de fragilidades apontadas por Pera e Bueno (2016), como os aspectos operacionais e financeiros da estrutura do IBGE para fazer o censo decenalmente.

É importante mencionar que esse modelo não considera aspectos de vizinhança, ou seja, não incorpora a chance de ser urbano em relação à situação de setores próximos. Este aspecto é incomum em estudos sobre espalhamento urbano, que geralmente avaliam contiguidade e formato da mancha urbana (Chetty, 2023). A vantagem de não usar a característica espacial é poder prever áreas urbanas a partir de fragmentos não-vegetados que surjam não adjacentes às áreas não-vegetadas.

Outra inovação do modelo é não usar a população ou a densidade populacional, dependendo apenas dos dados de sensoriamento remoto. Considerando que a grande motivação de fundo desse campo de estudo é compreender os processos e impactos de assentamentos humanos, sabendo que são multidimensionais (Dadashpoor; Shahhossein, 2024), se ancorar em trabalho de reconhecimento e validação de campo já realizado pelo IBGE e poder fazer previsões sem demandar novos esforços de campo é uma grande vantagem analítica.

CONSIDERAÇÕES FINAIS

Este trabalho caracterizou a situação urbano-rural de setores censitários do Censo 2022 do Distrito Federal por meio de regressão logística, usando como variáveis independentes as categorias de uso e cobertura da terra 'não vegetada', 'agropecuária' e 'vegetação', provenientes do nível 1 de categorização do MapBiomas. As análises realizadas demonstraram que a utilização de dados de sensoriamento remoto permite distinguir áreas urbanas e rurais com uma acurácia de 97,3%. O estudo reforça a viabilidade dessa abordagem para o monitoramento territorial, destacando a forte relação entre áreas não vegetadas e a situação urbana. A escolha da regressão logística se mostrou adequada para elaboração do modelo de previsão.

A inédita metodologia implementada pelo IBGE para o Censo 2022, que considerou imagens de alta resolução para definição da situação urbano-rural e foi a campo validar os tipos de cada setor censitário, abre novas potencialidades para a elaboração de estudos, pois vai além dos tradicionais limites administrativos e fornece uma malha validada por características dos territórios e conhecimento dos

Comentado [A12]: Em Araújo e Soares estas informações estão em outra citação direta

recenseadores. Pode-se dizer que o trabalho aqui apresentado traduziu essas características e conhecimentos em um modelo que usa dados disponíveis de sensoriamento remoto. Isso pode ser usado para prever a situação de áreas até a realização do próximo Censo, bem como para se planejar outras pesquisas amostrais em regiões que apresentam mudança no uso e cobertura da terra.

O uso dos dados de resolução espacial de 30 metros da Coleção 9 do MapBiomas permitiu ao modelo distinguir áreas urbanas e rurais por meio da identificação de padrões de uso e cobertura da terra. Embora resoluções mais altas possam no futuro oferecer maior detalhamento, os produtos de 30 metros foram adequados para detectar mudanças na paisagem.

Na perspectiva de pesquisas futuras, há oportunidades ao se avaliar aspectos espaciais e temporais. Um fator que pode ser melhor avaliado é o tamanho dos setores censitários, que tendem a ser menores quanto mais é a população e consequentemente a área não vegetada. Assim, estudos podem se valer de outros recortes espaciais para análise do modelo. Novas pesquisas também podem explorar outros *datasets* sobre uso e cobertura da terra além do MapBiomas, em especial com detalhamento para assentamentos humanos. O trabalho também tem potencial de ser ampliado para outras regiões do território brasileiro. Sobre aspectos temporais, a disponibilidade histórica e consistente desses dados possibilita análises temporais para o entendimento da situação urbano-rural, na construção das cidades brasileiras e para previsões futuras.

REFERÊNCIAS

BRASIL. **Constituição da República Federativa do Brasil de 1988**. 1988. Disponível em: https://www.planalto.gov.br/ccivil_03/constituicao/constituicao.html. Acesso em: 19 fev. 2025.

Código de campo alterado

BRASIL. **Objetivo de Desenvolvimento Sustentável 11: Cidades e comunidades sustentáveis**. Disponível em: <https://brasil.un.org/pt-br/sdgs/11>. Acesso em: 19 fev. 2025.

Código de campo alterado

CASA CIVIL; IPEA. **Avaliação de Políticas Públicas: Guia Prático de Análise Ex Ante**. 1. ed. Brasília: Ipea, 2018. Disponível em: <https://repositorio.ipea.gov.br/handle/11058/8285>. Acesso em: 17 fev. 2025.

Código de campo alterado

CHELOTTI, G. B.; SANO, E. Sessenta anos de evolução urbana do Distrito Federal analisadas por meio de métricas de paisagem. **Revista Gestão & Sustentabilidade Ambiental**, v. 12, p. 21, 2023. <https://doi.org/10.59306/rgsa.v12e12023e10279>

CHETTRY, V. A Critical review of urban sprawl studies. **Journal of Geovisualization and Spatial Analysis**, v. 7, n. 28, p. 13, 2023. <https://doi.org/10.1007/s41651-023-00158-w>

CRESWELL, J. W.; CRESWELL, J. D. **Research Design: Qualitative, Quantitative, and Mixed Methods Approaches**. 6ª ed. Los Angeles: SAGE Publications, 2017.

DADASHPOOR, H.; SHAHHOSSEIN, G. Defining urban sprawl: A systematic review of 130 definitions. **Habitat International**, v. 146, p. 14, 2024. <https://doi.org/10.1016/j.habitatint.2024.103039>

DODMAN, D. *et al.* Cities, Settlements and Key Infrastructure. In: WORKING GROUP II CONTRIBUTION TO THE SIXTH ASSESSMENT REPORT OF THE INTERGOVERNMENTAL PANEL ON CLIMATE CHANGE (Coord.) **Climate Change 2022: Impacts, Adaptation, and Vulnerability**. Cambridge: Cambridge University Press, 2023. p. 907–1040. <https://doi.org/10.1017/9781009325844.008>

FERNANDES, A. A. T. *et al.* Leia este artigo se você quiser aprender regressão logística. **Revista de Sociologia e Política**, v. 28, n. 74, p. e006, 2020.

GIL, A. C. **Como elaborar projetos de pesquisa**. 4ª ed. São Paulo: Ed. Atlas, 2002.

IBGE. **Distrito Federal - Cidades e Estados**. Disponível em: <https://www.ibge.gov.br/cidades-e-estados/df.html>. Acesso em: 19 fev. 2025.

- 448 IBGE. **Malha de Setores Censitários de 2022**. Disponível em:
449 [https://www.ibge.gov.br/geociencias/organizacao-do-territorio/malhas-territoriais/26565-malhas-de-](https://www.ibge.gov.br/geociencias/organizacao-do-territorio/malhas-territoriais/26565-malhas-de-setores-censitarios-divisoes-intramunicipais.html)
450 [setores-censitarios-divisoes-intramunicipais.html](https://www.ibge.gov.br/geociencias/organizacao-do-territorio/malhas-territoriais/26565-malhas-de-setores-censitarios-divisoes-intramunicipais.html). Acesso em: 18 fev. 2025.
- 451 IBGE. **Malha de Setores Censitários**. Rio de Janeiro: IBGE, 2024. Disponível em:
452 <https://biblioteca.ibge.gov.br/index.php/biblioteca-catalogo?view=detalhes&id=2102138>. Acesso em:
453 18 fev. 2025.
- 454 JACKMAN, S. *et al.* **PscI: Political Science Computational Laboratory**. Disponível em:
455 <https://cran.r-project.org/web/packages/pscl/index.html>. Acesso em: 20 fev. 2025.
- 456 KUHN, M. *et al.* **Caret: Classification and Regression Training**. Disponível em: [https://cran.r-](https://cran.r-project.org/web/packages/caret/index.html)
457 [project.org/web/packages/caret/index.html](https://cran.r-project.org/web/packages/caret/index.html). Acesso em: 20 fev. 2025.
- 458 MAPBIOMAS. **Collection 9 of the Annual Land Cover and Land Use Maps of Brazil (1985-2023)**..
459 Disponível em:
460 <https://data.mapbiomas.org/dataset.xhtml?persistentId=doi:10.58053/MapBiomias/XXUKA8>. Acesso
461 em: 15 fev. 2025.
- 462 PERA, C. K. L.; BUENO, L. M. de M. Revendo o uso de dados do IBGE para pesquisa e
463 planejamento territorial: reflexões quanto à classificação da situação urbana e rural. **Cadernos**
464 **Metrópole**, v. 18, n. 37, p. 722–742, 2016. <https://doi.org/10.1590/2236-9996.2016-3705>
- 465 PLANET. **Norway's International Climate and Forests Initiative Satellite Data Program (NICFI**
466 **Program)**. Disponível em: <https://www.planet.com/nicfi/>. Acesso em: 21 fev. 2025.
- 467 POSIT, S. **RStudio: Integrated Development Environment for R**. Disponível em:
468 <https://posit.co/download/rstudio-desktop/>. Acesso em: 22 fev. 2025.
- 469 QGIS. **QGIS Geographic Information System**. Disponível em: <https://qgis.org/>. Acesso em: 22 fev.
470 2025.
- 471 R CORE TEAM. **The R Project for Statistical Computing**. Disponível em: [https://www.r-](https://www.r-project.org/about.html)
472 [project.org/about.html](https://www.r-project.org/about.html). Acesso em: 19 fev. 2025.
- 473 REFERENCE, S. **ESRI:102033 - South_America_Albers_Equal_Area_Conic**. Disponível em:
474 <https://spatialreference.org/ref/esri/102033/>. Acesso em: 21 fev. 2025.
- 475 ROBIN, X. *et al.* **pROC: Display and Analyze ROC Curves**. Disponível em: [https://cran.r-](https://cran.r-project.org/web/packages/pROC/index.html)
476 [project.org/web/packages/pROC/index.html](https://cran.r-project.org/web/packages/pROC/index.html). Acesso em: 20 fev. 2025.
- 477 RSTUDIO TEAM. **stats (version 3.6.2)**. Disponível em:
478 <https://www.rdocumentation.org/packages/stats/versions/3.6.2>. Acesso em: 20 fev. 2025.
- 479 SCHLOERKE, B. *et al.* **GGally: Extension to “ggplot2”**. Disponível em: [https://cran.r-](https://cran.r-project.org/web/packages/GGally/index.html)
480 [project.org/web/packages/GGally/index.html](https://cran.r-project.org/web/packages/GGally/index.html). Acesso em: 20 fev. 2025.
- 481 SEGOV. **Administrações Regionais**. Disponível em:
482 <https://segov.df.gov.br/category/administracoes-regionais/>. Acesso em: 19 fev. 2025.
- 483 SOUZA, C. M. *et al.* Reconstructing Three Decades of Land Use and Land Cover Changes in
484 Brazilian Biomes with Landsat Archive and Earth Engine. **Remote Sensing**, v. 12, n. 2735, p. 27,
485 2020. <https://doi.org/10.3390/rs12172735>
- 486 SOUZA, A. L. de *et al.* Revisão metodológica da tipologia urbano-rural no Censo Demográfico 2022.
487 **Ciência & Saúde Coletiva**, v. 49, n. 11, p. 12, 2024. [https://doi.org/10.1590/1413-](https://doi.org/10.1590/1413-812320242911.03062024)
488 [812320242911.03062024](https://doi.org/10.1590/1413-812320242911.03062024)

Comentado [A13]: <https://biblioteca.ibge.gov.br/visualizacao/livros/liv102138.pdf>

- 489 SUDHIRA, H. S.; RAMACHANDRA, T. V.; JAGADISH, K. S. Urban sprawl: metrics, dynamics and
490 modelling using GIS. **International Journal of Applied Earth Observation and Geoinformation**, v.
491 5, n. 1, p. 29–39, 2004. <https://doi.org/10.1016/j.jag.2003.08.002>
- 492 TAUBENBÖCK, H. *et al.* A new ranking of the world's largest cities—Do administrative units obscure
493 morphological realities? **Remote Sensing of Environment**, v. 232, n. 111353, p. 14, 2019.
494 <https://doi.org/10.1016/j.rse.2019.111353>
- 495 UN-HABITAT. **World Cities Report 2022: Envisaging the Future of Cities**. Disponível em:
496 <https://unhabitat.org/world-cities-report-2022-envisaging-the-future-of-cities>. Acesso em: 17 fev. 2025.
- 497 WICKHAM, H. *et al.* **Ggplot2: Create Elegant Data Visualisations Using the Grammar of**
498 **Graphics**. Disponível em: <https://cran.r-project.org/web/packages/ggplot2/index.html>. Acesso em: 20
499 fev. 2025.
- 500 WICKHAM, H. *et al.* **Readxl: Read Excel Files**. Disponível em: [https://cran.r-](https://cran.r-project.org/web/packages/readxl/index.html)
501 [project.org/web/packages/readxl/index.html](https://cran.r-project.org/web/packages/readxl/index.html). Acesso em: 19 fev. 2025.
- 502 XAVIER-DA-SILVA, J.; CARVALHO-FILHO, L. M. Sistemas de Informação Geográfica: uma proposta
503 metodológica. In: **IV CONFERÊNCIA LATINOAMERICANA SOBRE SISTEMAS DE INFORMAÇÃO**
504 **GEOGRÁFICA**, 1993, São Paulo. **Anais [...]**. São Paulo, 2019. p. 609–629.
- 505 ZHENG, Q. *et al.* Nighttime light remote sensing for urban applications: Progress, challenges, and
506 prospects. **ISPRS Journal of Photogrammetry and Remote Sensing**, v. 202, p. 125–141, 2023.
507 <https://doi.org/10.1016/j.isprsjprs.2023.05.028>.