

## **Análise comparativa entre os dados de arborização divulgados pelo IBGE e dados de cobertura vegetal extraídos de imagens do sensor WorldView II.**

Tarik de Sousa Araújo<sup>1,2</sup>

<sup>1</sup> Instituto Brasileiro de Geografia e Estatística - IBGE  
Av. Franklin Roosevelt, 166 - Centro - 20021120 – Rio de Janeiro - RJ, Brasil  
tarik.araujo@ibge.gov.br

<sup>2</sup> Universidade Federal do Rio Grande do Norte - UFRN  
Caixa Postal 1524 - Natal - RN, Brasil

**Abstract.** This paper shows a comparison between data from urban afforestation, produced by a proper methodology of the Brazilian Institute of Geography and Statistics (IBGE), during the 2010 Census preparatory stages and land cover data obtained from a WorldView II satellite image. The territorial clipping of this research was the city of Natal. Comparison of the data was done using basic statistical tools: histogram analysis, cross tabulation, scatterplot's analysis and simple linear regression. The results contradicted expectations, showing low correlation between urban afforestation data from IBGE and land cover data. We conclude that the result is due to systematic limitations resulting from the methodology of collecting data from IBGE. we recommend caution in using IBGE's data published about urban afforestation.

**Palavras-chave:** Arborização, WorldView II, Censo 2010, Regressão Linear.

### **1. Introdução**

A cada 10 anos, o IBGE – (Instituto Brasileiro de Geografia e Estatística) realiza o Censo Demográfico em todo o Brasil, com o objetivo de retratar as características da população.

No ultimo Censo, realizado em 2010, foram implementadas grandes inovações tecnológicas e metodológicas que possibilitaram uma significativa melhora no processo de coleta das informações, permitido um maior controle da coleta e ampliando a escala de confiabilidade da distribuição espacial dos dados.

Nesse sentido se destaca como uma das principais inovações a realização anterior ao Censo 2010, do CNEFE – (Cadastro Nacional de Endereços para Fins Estatísticos). Em termos práticos, o Cadastro Nacional de Endereços, significou o levantamento sistemático das geometrias (linhas) de todas as faces de quadras das áreas urbanas brasileiras e a sua disposição em um banco de dados, onde também foram associados a estas, atributos como: nome do logradouro, código postal, lista de domicílios, etc. Em áreas rurais o CNEFE adotou como padrão a geometria de pontos para cada domicilio individual.

A realização desse cadastro de endereços possibilitou a inclusão de novas variáveis ao levantamento censitário. Essas variáveis são resultantes de uma pesquisa realizada paralelamente a etapa preparatória do Censo 2010, conhecida como Pré-coleta. Elas não representam dados essencialmente demográficos e foram intituladas de Características Urbanísticas do Entorno dos Domicílios. Cobrem aspectos importantes da infraestrutura urbana, com destaque para as dimensões de circulação e meio ambiente como: iluminação pública, pavimentação, arborização, bueiro/boca de lobo, lixo acumulado, esgoto a céu aberto, meio/fio /guia, calçada e rampa para cadeirante IBGE (2010).

Nesta investigação foram utilizados os dados levantados acerca da arborização urbana, divulgados pelo IBGE, pois nosso objetivo é comparar os dados oriundos do Censo 2010, com dados de cobertura vegetal, obtidos a partir da classificação de imagens multiespectrais de alta resolução do sensor WorldView II, também adquiridas no mesmo ano. Assim,

buscaremos através de análises estatísticas, verificar a consistência dos dados sobre arborização urbana, produzidos pelo IBGE, tendo em vista a grande abrangência geográfica e o caráter inédito deste levantamento.

Devido à disponibilidade das imagens de sensoriamento remoto utilizadas nesse estudo, o recorte territorial aplicado será o Município de Natal, capital do Estado do Rio Grande do Norte como pode ser visualizado na Figura 1. Natal possui uma característica peculiar, todo o seu território é classificado como zona urbana, o que significa que em todos os setores censitários que o recobrem, foram coletadas informações a respeito de características urbanísticas.

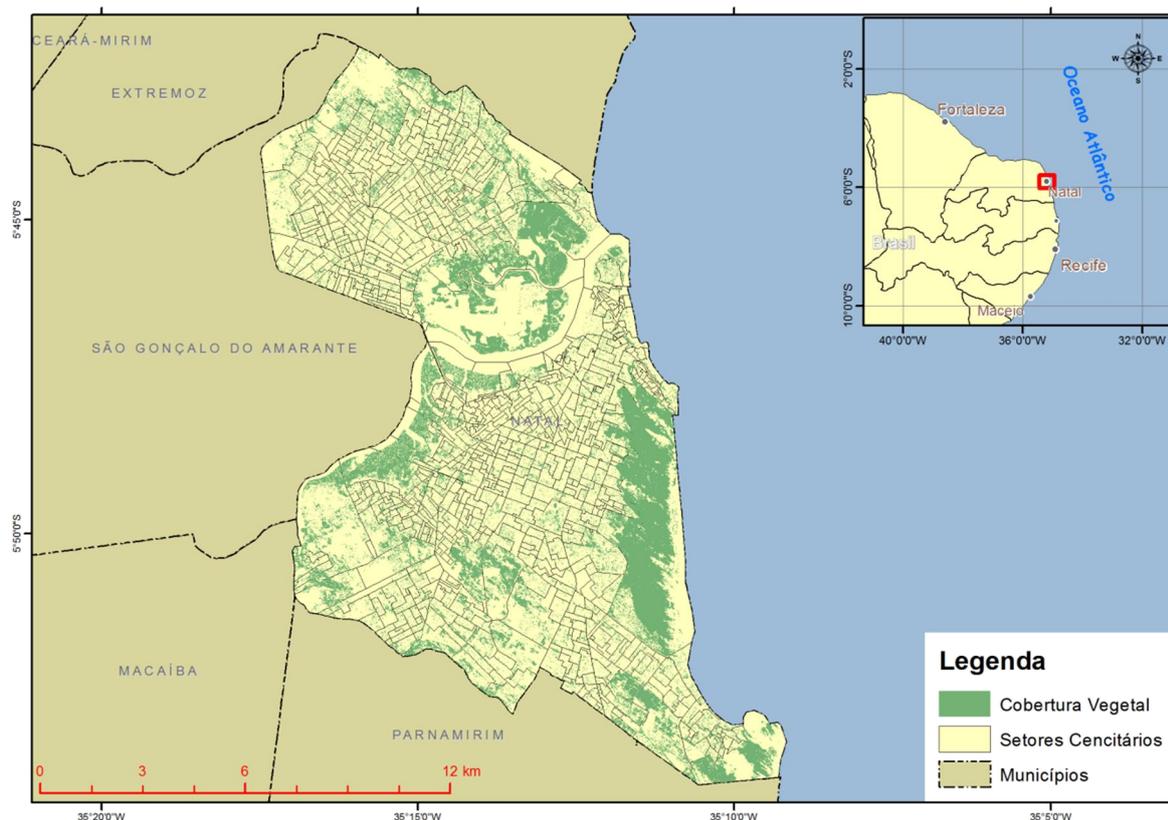


Figura 1. Localização da Área de estudo e distribuição da cobertura vegetal e dos setores censitário do Município de Natal.

## 2. Metodologia

Para fins explicativos, podemos dividir essa pesquisa em duas etapas: preparação e extração dos dados e a análise dos dados. Nessas fases de trabalho foram utilizados os softwares: ArcGis 10.2.1, Qgis 2.2.0, Envi 4.8, eCognition 8.7 e GeoDa 1.6.6.

### Preparação e extração dos dados

A metodologia de coleta do IBGE para os dados de arborização consiste em identificar se em uma face de quadra ou na sua face confrontante ou no canteiro central, se existia arborização, ou seja, existia ao menos uma árvore ao longo da calçada/passeio e ou em canteiro que dividia pistas de um mesmo logradouro, mesmo que apenas em parte. Considerou-se também a arborização quando existente em logradouros sem pavimentação e ou sem calçada/passeio IBGE (2010).

Embora a existência ou não de árvores seja contabilizada em relação às faces de quadra, é importante compreender que o IBGE não verifica o número de árvores. Quando uma face é identificada positivamente, todos os domicílios que estão atribuídos àquela face passam a terem o status de domicílios com arborização. A escala de divulgação é feita apenas por setor censitário, que são áreas de controle territorial da coleta, constituídas por áreas contíguas IBGE (2013). Para fins de geoprocessamento os setores censitários são representados por uma malha de polígonos que devem conservar suas relações topológicas e podem ser modelados como Geo-objetos.

Contido em cada setor censitário urbano estão um número  $n$  de faces de quadras e para cada face são associados um número  $x$  de domicílios, que são ainda classificados em Domicílios Particulares Permanentes: Próprios, Alugados ou Cedidos. Destacamos que o IBGE divulga como domicílios com arborização o somatório dos  $x$  domicílios atribuídos às  $n$  faces onde foram identificadas árvores.

Na Tabela 1, destacamos os dados divulgados pelo IBGE que foram utilizados em nossa pesquisa. O dado disponível por setor censitário está nos Resultados do Universo por Setor Censitário, IBGE (2011). E para ser trabalhado em um ambiente de informações geográficas precisa passar por alguns pré-processamentos.

Tabela 1. Dados divulgados pelo IBGE

Titulo	Formato	Ano
Malha de Setores RN	.SHP	2010
Resultados do universo agregados por Setor – entorno01_RN	.XLS	2010

A malha de setores censitários esta disponível em arquivos em formato *shapefile*, recortados por Unidades Federativas. Os atributos desses *shapefiles* contêm somente informações básicas de cada setor. Para utilizar os demais dados coletados no censo é necessário carregar a tabela alfanumérica específica do tema que se deseja trabalhar e realizar uma ligação (*Join*) com os atributos do *shapefile*. Essa ligação foi realizada utilizando o campo Geocódigo dos setores que representa uma identidade única para cada setor censitário.

Com os dados de arborização devidamente associados a seus respectivos polígonos de setores, fizemos um somatório dos valores de domicílios com arborização das três categorias (próprios, alugados ou cedidos) e dispoñdo também do valor total de domicílios em cada setor, calculamos a porcentagem de domicílios com arborização.

Para a obtenção da cobertura vegetal do Município de Natal, utilizamos uma cena WorldView II de 2010, com 8 (oito) bandas espectrais da região do visível e infravermelho próximo, com resolução de 1,85 m (multiespectral). A cena WorldView II foi convertida para radiância no topo da atmosfera por meio da Equação 01.

$$L_1 = \frac{ND_n}{CC_n} \quad (01)$$

Em que  $L_1$  é a radiância espectral aparente;  $ND_n$  é o número digital de uma determinada banda e  $CC_n$  é o coeficiente de calibração absoluta para uma determinada banda.

Posteriormente aplicou-se a correção atmosférica através do módulo Fast Line-of-sight Atmospheric Analysis of Spectral Hypercubes (FLAASH), transformando a imagem em reflectância de superfície.

Em seguida a imagem foi submetida ao segmentador de multiresolução com pesos iguais para todas as suas 8 bandas. Os objetos resultantes dessa segmentação foram classificados em

vegetação e não vegetação com a utilização do Índice de Vegetação de Diferença Normalizada (NDVI), que utilizou as bandas do infravermelho próximo (*NIR*) e do vermelho (*RED*) da imagem WorldView II (Equação 02).

$$NDVI = \frac{(NIR - RED)}{(NIR + RED)} \quad (02)$$

Em que, NDVI: Índice de Vegetação de Diferença Normalizada; *NIR* 1: reflectância na faixa do infravermelho próximo (0,772 – 0,890  $\mu\text{m}$ ); *RED*: reflectância na faixa do visível (0,629 – 0,689  $\mu\text{m}$ ).

As cenas do WorldView II apresentam duas bandas na região do infravermelho próximo (*NIR* 1 e *NIR* 2), optou-se por selecionar somente a primeira, que neste caso corresponde a comprimentos de onda 0,772 – 0,890  $\mu\text{m}$ .

De acordo com Ponzoni e Shimabikuro (2007), o NDVI é uma ferramenta muito utilizada para monitoramento da vegetação. Sua aplicação faz com que a vegetação seja evidenciada com valores próximos a +1 e representada com tons mais claros na imagem, e o contrário, com tons mais escuros para as áreas com ausência de vegetação. O intervalo de valores de NDVI onde os objetos assumiram a classe de vegetação foram estabelecidos de forma empírica, priorizado à vegetação arbustiva e arbórea, assim, objetos que apresentavam médias de NDVI com valores entre 0,6 a 1, foram classificados.

Na Figura 2 é possível observar que os resultados dessa metodologia de classificação cumpriram com o objetivo de identificar e quantificar a cobertura vegetal do Município de Natal.

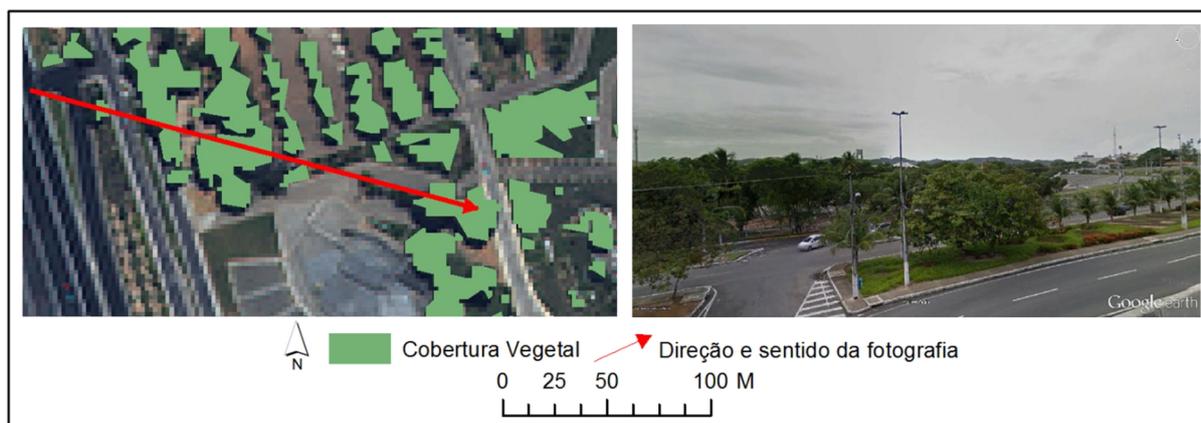


Figura 2. A esquerda composição colorida das bandas do visível da imagem WorldView II, sobreposta pelos objetos classificados como cobertura vegetal e uma seta (em vermelho), que representa a direção e o sentido da fotografia a direita, oriunda do Google Street View.

Os objetos classificados foram transformados em *shapefile* para serem trabalhados em ambiente de sistema de informações geográficas, onde foram dissolvidos e recortados obedecendo aos limites dos setores censitários, de forma que para cada setor censitário passou a existir um objeto correspondente à sua cobertura vegetal, porém, em uma camada diferente. Os valores de área dos objetos correspondentes à cobertura vegetal por setor foram calculados e transferidos para a tabela de atributos da camada principal de setores censitários através da ferramenta *Spatial Join*.

Dispondo do valor total da área de cada setor censitário e da parcela de área com cobertura vegetal, foi possível calcularmos a porcentagem de área vegetada contida em cada polígono de setor.

Ao final da etapa de preparação e extração de dados, agregamos a malha de setores censitários do Município de Natal, os atributos que serão analisados na etapa posterior.

### Análise dos dados

No intuito de verificar a consistência e a representatividade dos dados de arborização produzidos pelo IBGE submetemos os resultados apurados no Município de Natal a comparações com dados de cobertura vegetal extraídos de imagens de sensoriamento remoto.

Para a realização das comparações foram utilizadas ferramentas estadísticas clássicas implementadas no software GeoDa 1.6.6, como: análise de histogramas, tabulação cruzada, análises de *Scatterplot's* e regressão linear simples.

Optamos em comparar os dados de porcentagem de área vegetada com a porcentagem de domicílios arborizados. Essas variáveis foram escolhidas por acreditarmos existir uma relação direta de dependência entre ela.

### 3. Resultados e Discussão

Os padrões de dispersão da frequência de conjuntos de dados nos permite fazer uma aproximação inicial da natureza dos fenômenos que buscamos compreender.

Ao compararmos os histogramas das variáveis: porcentagem de área vegetada com a porcentagem de domicílios com arborização (Figura 3). Ficou evidente que as duas variáveis não possuíam uma relação de dependência, pois o histograma de porcentagem de área vegetada apresenta uma grande ocorrência de porcentagens entre 0 e 10%, enquanto a distribuição da frequência das porcentagem de domicílios com arborização aparenta uma certa homogeneidade.

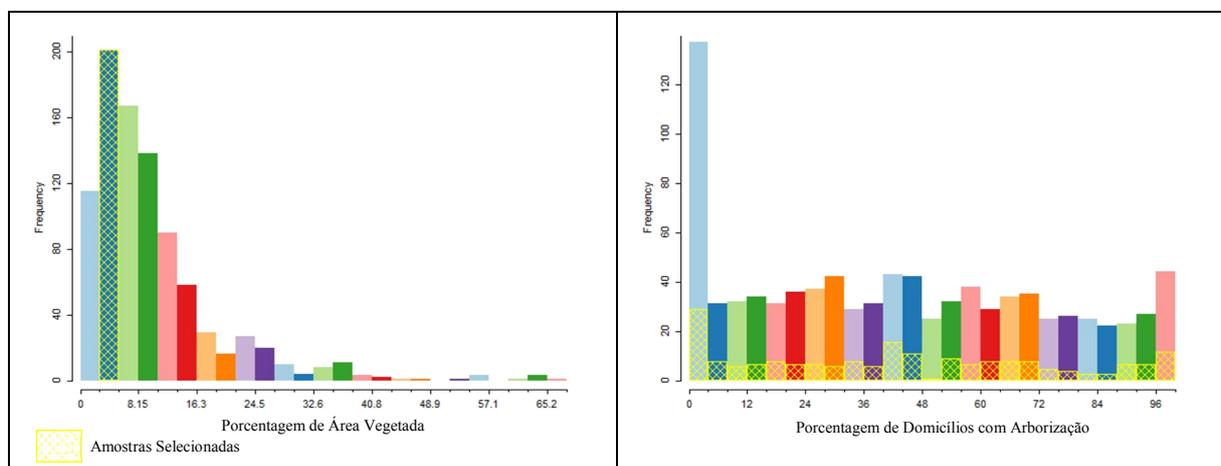


Figura 3. Comparação entre os histogramas das variáveis: porcentagem de área vegetada com a porcentagem de domicílios com arborização.

A falta de correlação entre as duas variáveis expostas na Figura 3, fica mais evidente quando fazemos a seleção das amostras condidas no intervalo de maior frequência das porcentagens de área vegetada. Setores que possuem entre 2,7 e 5,4% de área vegetada aparecem distribuídos homogeneamente em todos os intervalos de frequência das porcentagens de domicílios com arborização. Inclusive, apresentando ocorrências na faixa de 96 a 100% de domicílios com arborização. Para entender melhor o que estava exposto nos

histogramas, selecionamos um setor que se encaixa na condição acima e representamos a sua condição na Figura 4. Nela destacamos um logradouro onde se observa a existência de apenas uma árvore situada em espaço público.

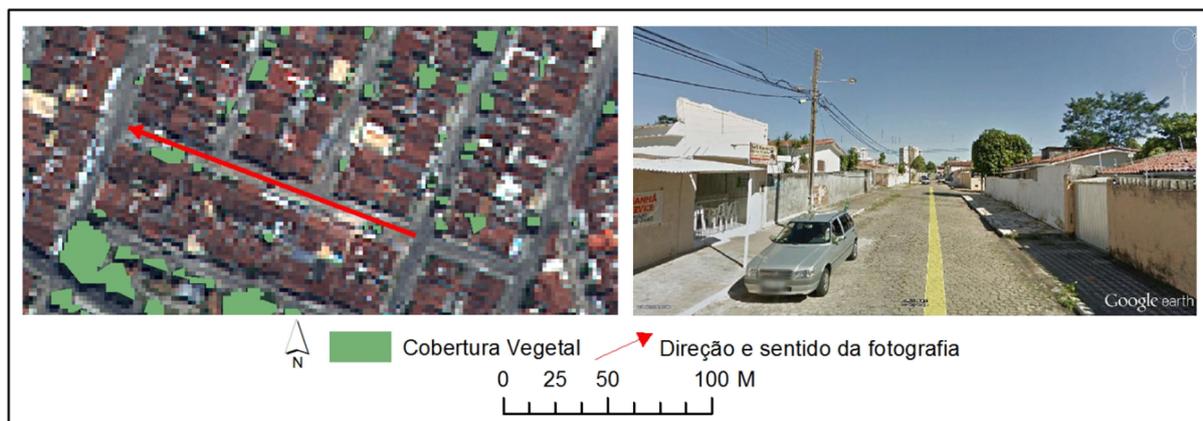


Figura 4. Destaque de setor censitário que possui 5,4% de área vegetada, porém 98,5% de seus domicílios foram classificados como possuindo arborização.

Buscando compreender melhor o comportamento das duas variáveis em análise construímos um *Scatterplot* e realizamos uma regressão linear simples. Como pode ser visualizado na Figura 5, as duas variáveis não possuem indícios de relação dependência ou correlação. O  $R^2$  da regressão apresentou valores ínfimos e a constante foi de 42,8.

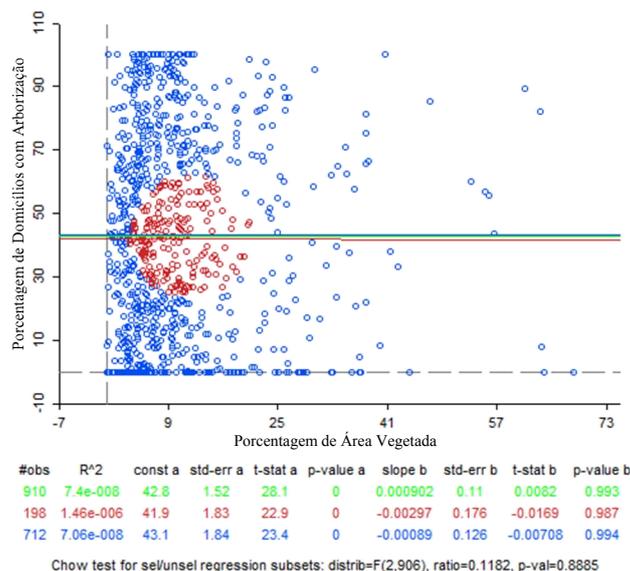


Figura 5. Scatterplot e regressão linear entre as variáveis: porcentagem de área vegetada e porcentagem de domicílios com arborização.

Esses resultados contrariaram o que era logicamente esperado: maiores proporções de cobertura vegetal se traduzissem em maiores proporções de domicílios com arborização e que o contrario também fosse verdadeiro. A não correlação dessas duas variáveis coloca em duvida a confiabilidade e a representatividade dos dados envolvidos na análise.

Como os dados de cobertura vegetal podem ser facilmente validados (Figuras 2 e 5), precisamos voltar nossa atenção para os dados de domicílios com arborização divulgados pelo IBGE.

#### 4. Conclusões

A presença de arborização urbana apresenta múltiplos benefícios como: amenizar a poluição atmosférica, influenciar o balanço hídrico, reduzir a velocidade do vento e principalmente, contribuir para a regularização do microclima urbano ao minimizar as temperaturas das áreas circunvizinhas pelo sombreamento e pelo consumo de energia para a evapotranspiração, Lima e Amorim (2011).

Em cidades como Natal, onde segundo Silva et al. (2008), o índice de radiação incidente apresenta altos valores durante todo o ano, a presença da cobertura vegetal, representa um importante fator atenuante dos efeitos da insolação e contribui para o conforto ambiental dos seus habitantes.

Tendo em vista esses fatores ambientais, reconhecemos que com as inovações apresentadas no Censo 2010, o IBGE realizou um grande esforço ao tentar produzir dados que quantificassem a arborização urbana em todas as cidades brasileiras.

No entanto, os experimentos realizados nesse estudo compararam os dados produzidos pelo IBGE com dados extraído de imagens de sensoriamento remoto de alta resolução espacial e verificaram que a metodologia utilizada pelo IBGE, não representa adequadamente a realidade da arborização dispersa na malha urbana de Natal.

Acreditamos que o problema dos dados divulgados pelo IBGE, é sistemático, ou seja, está diretamente relacionado com a metodologia da coleta e com a preparação dos dados. Pois, o IBGE, considerou que para uma face de quadra fosse classificada como arborizada bastaria possuir apenas uma árvore em seu entorno. Após a face ser classificada, todos os domicílios associados a ela passam a ser contabilizados como arborizados.

Levando em consideração a falta de padrões urbanísticos que determinem tamanhos uniformes de faces de quadra e o próprio adensamento urbano com a presença da verticalização. Consideramos que o dado do IBGE se apresenta superdimensionado.

Recomentamos cautela aos usuários que dispõem apenas do dado do IBGE e desejam realizar estudos acerca do tema arborização. Se possível, busquem dados de sensoriamento remoto que auxiliem em seus diagnósticos.

Para estudos futuros esperamos que o IBGE disponibilize a sociedade a geometria das faces de quadra dos setores censitários urbanos, pois com elas será possível refinar ainda mais a escala de estudos similares a este.

#### Referencias Bibliográficas

IBGE – Instituto Brasileiro de Geografia e Estatística. **Metodologia do Censo Demográfico 2010**. Rio de Janeiro, 2013.

IBGE – Instituto Brasileiro de Geografia e Estatística. **Censo Demográfico 2010: Características urbanísticas do entorno dos domicílios**. Rio de Janeiro, 2010.

IBGE – Instituto Brasileiro de Geografia e Estatística. Base de informações do **Censo Demográfico 2010: Resultados do Universo por setor censitário**. Rio de Janeiro, 2011.

Lima, V.; Amorim, M. C. C. T. A utilização de informações de temperatura da superfície, do NDVI e de temperatura do ar na análise de qualidade ambiental urbana. In: Simpósio Brasileiro de Sensoriamento Remoto, 15., 2011, Curitiba. **Anais...** Curitiba: INPE, 2011, p.1029.

Ponzoni, F. J.; Shimabukuro, Y. E. **Sensoriamento remoto no estudo da vegetação**. São José dos Campos: A. Silva Vieira Ed., 2007.

Silva, F. R.; Oliveira, H. S. M.; Marinho, G. S. Análise das componentes global e difusa da radiação solar em Natal-RN entre 2007 e 2008. In: Congresso Brasileiro de Energia Solar, 2., e Conferência Regional Latino-Americana, 3., 2008, Florianópolis. **Anais...** Florianópolis: ISES, 2008.