



# Desenvolvimento de Software para Análise Exploratória de Dados Espaciais

Regiane Leandro Pizon  
Maurílio de Abreu Monteiro

**Agência financiadora:** CNPq

**Resumo:** Este trabalho avalia as abordagens utilizadas para a exploração de dados espaciais e propõe um software eficiente que viabiliza a aplicação do procedimento em problemas com grandes volumes de dados. O método proposto combina o uso de técnicas de otimização combinatória com um esquema de simplificação dos gráficos, representando áreas espaciais através do uso de plano cartesiano.

**Palavras chave:** *software*; dados espaciais; grandes volumes de dados.

## 1. INTRODUÇÃO

Compreender a distribuição espacial de dados oriundos de fenômenos ocorridos no espaço, constitui hoje um grande desafio para o esclarecimento de questões centrais em diversas áreas do conhecimento. Tais estudos vêm se tornando cada vez mais comuns com o uso de sistemas de informação geográfica (SIG), Segundo Neves (2003) estes sistemas necessitam de uma base de dados e de uma base geográfica, para permitir a visualização do padrão espacial do fenômeno, entretanto, as bases geográficas para geração de mapas possui um acervo limitado, dificultando a análise espaciais de determinados fenômenos.

O objetivo deste trabalho é apresentar um software alternativo para o procedimento de análise aplicável a grandes volumes de dados oriundos de forma a permitir sua utilização como ferramenta de desenvolvimento de novas representações geográficas, utilizando de planos cartesianos para a análise exploratória de dados espaciais

## 2. MATERIAL E MÉTODOS

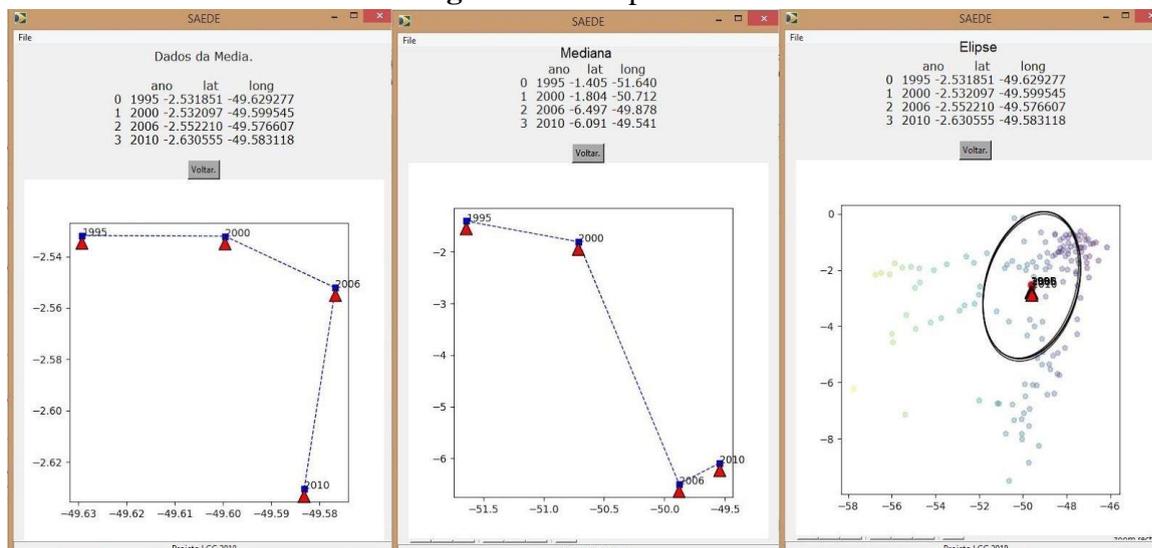
O software SAEDE (*Software* Análise Exploratória de Dados Espaciais), desenvolvido na linguagem *Python*, com o principal objetivo de apresentar uma ferramenta de fácil acesso, com uma interface amigável, simplificando a entrada de inputs e rapidez em seu processamento. Para a utilização desta ferramenta é necessário que o usuário utilize um arquivo .csv com informações da latitude e longitude como obrigatoriedade, além disso é opcionalmente possível passar qualquer variável de agrupamento e uma variável de ponderação no *input* do sistema.

O SAEDE possibilita a análise exploratória de dados através de planos cartesianos, oferecendo a Média, Mediana, Centróides e Elipses. Partindo de um pressuposto que uma região pode ser representada pelo seu centróide.

Foi desenvolvida uma função que calcula espacialmente o centro dos dados, representado pelo ponto médio (latitude, longitude), de uma determinada amostra.

Opcionalmente, é possível passar qualquer variável de agrupamento e uma variável de ponderação. Assim é possível verificar o viés ou tendência de concentração geográfica do fenômeno. Para construir estes gráficos, como pode ser observado na figura 1, utilizam-se as coordenadas dos centróides das áreas, aproximando-as para um espaçamento regular de forma a montar uma matriz. Calcula-se, então, as médias e as medianas do indicador ao longo das linhas e colunas desta matriz (CARVALHO, 2017). Esta técnica permite identificar a flutuação das medidas ao longo de duas direções, sugerindo a presença de valores discrepantes quando a diferença entre estas é grande, e a tendência ao longo de uma direção quando os valores variam suavemente.

**Figura 1- Exemplo de Anomalia**



### 3. RESULTADOS E DISCUSSÃO

Os principais resultados encontrados foram: (1) Facilidade de input, visto que a ferramenta necessita apenas de uma base de dados que tenha obrigatoriamente a Latitude e Longitude; (2) Análise exploratória de dados sem a necessidade de um conhecimento específico na área; (3) Com a utilização dos planos cartesianos para representar o mapa, foram superados problemas como a demora no processamento desses dados espaciais.

### 4. CONCLUSÃO

O software SAEPE mostrou-se eficaz para auxiliar a análise de dados espaciais com representação geográfica poligonal, através do uso de um procedimento de classificação dos objetos espaciais, devido a facilidade de acessar os resultados gerados dinamicamente pela mesma, possibilitando apoiar variados tipos de pesquisas que necessitem extrair informações de dados espaciais de forma otimizada.

### 5. REFERÊNCIAS E CITAÇÕES

CARVALHO, Alexandre Ywata de, et al. *Clusterização espacial e não espacial: um estudo aplicado à agropecuária brasileira*. No. 2279. Texto para Discussão, 2017.

NEVES, Marcos Corrêa. *Procedimentos eficientes para regionalização de unidades socioeconômicas em bancos de dados geográficos*. 129 p. Diss. Tese (Doutorado em Sensoriamento Remoto) – Instituto Nacional de Pesquisas Espaciais (INPE), São José dos Campos, 2003.