

Geoprocessamento

Prof. Diego Camargo

Aula 03 – Tratamento e estrutura de dados



Curso de Engenharia de Transportes
Departamento de Engenharia de Transportes

OBJETIVO DA AULA

1. Introduzir conceitos de tratamento de dados geográficos;
2. Tipos de dados e operações.

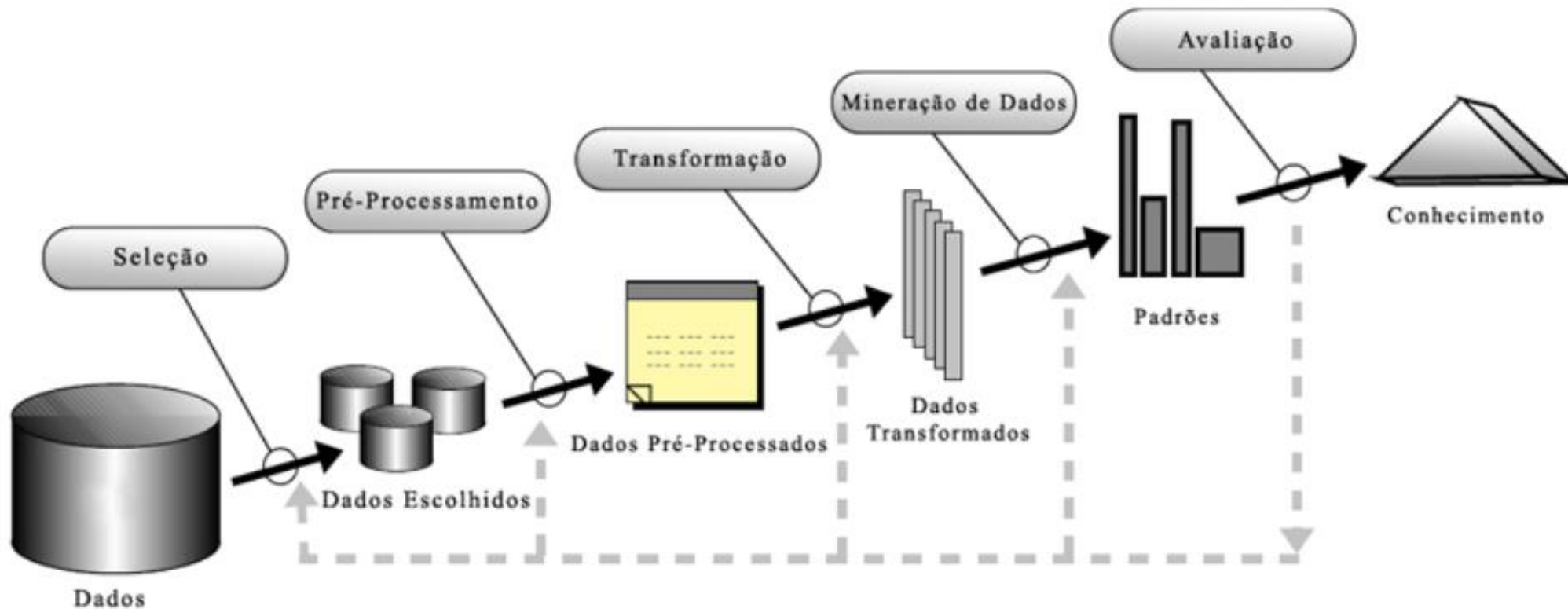
DADOS GEOGRÁFICOS

- Com o volume de dados armazenados crescendo diariamente, responder uma questão tornou-se crucial:
 - O que fazer com os dados armazenados?

DADOS GEOGRÁFICOS

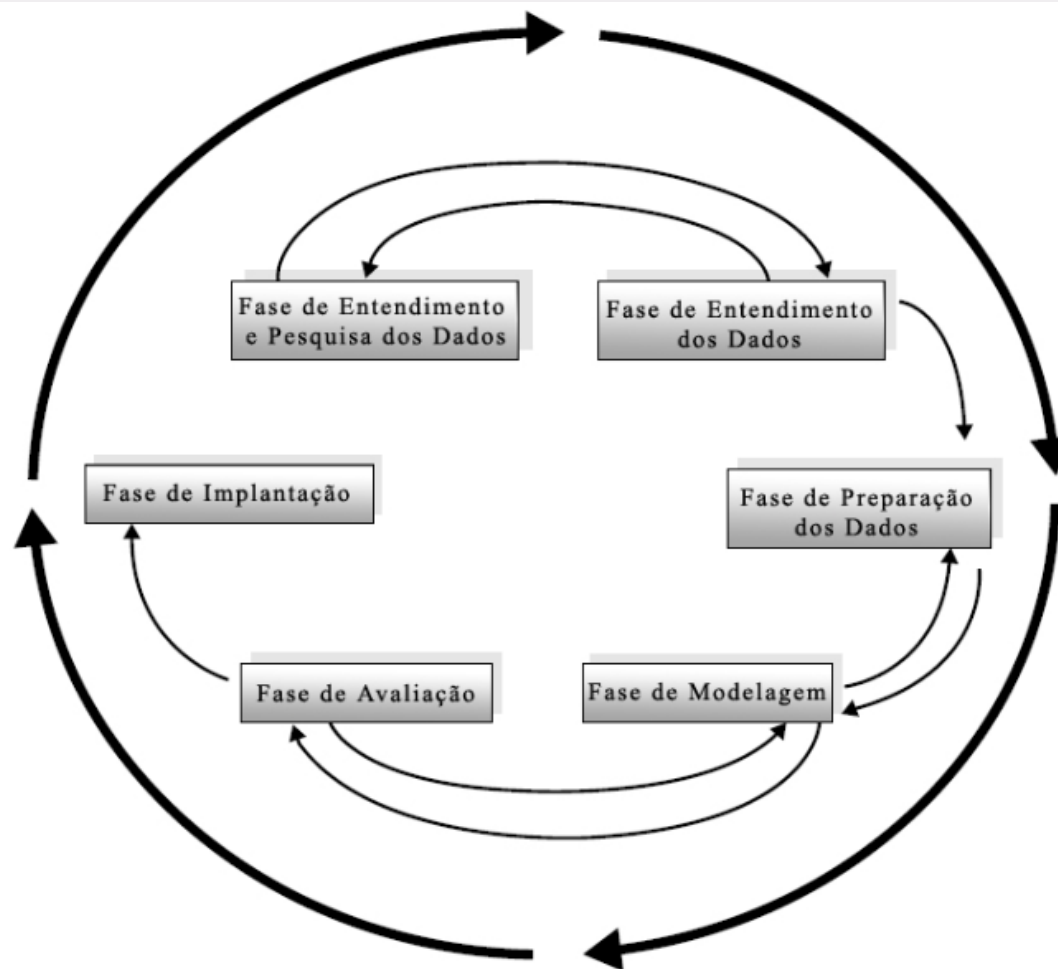
- Georreferenciar ou geocodificar dados tornou-se uma tarefa indispensável nos dias atuais;
- Contudo, envolve mais tarefas ao processo de tratamento de dados e deixa as operações mais complexas.

PROCESSO DE TRATAMENTO



PROCESSO DE TRATAMENTO

- CRISP-DM (*Cross-Industry Standard Process of Data Mining*)



PROCESSO DE TRATAMENTO

- **Entendimento dos Dados:** As fontes fornecedoras dos dados podem vir de diversos locais e possuírem diversos formatos. Após definir os objetivos, é necessário conhecer os dados visando;
 1. Descrever de forma clara o problema;
 2. Identificar os dados relevantes para o problema em questão;
 3. Certificar-se de que as variáveis relevantes para o projeto não são interdependentes.

PROCESSO DE TRATAMENTO

- **Preparação dos Dados:** Devido às diversas origens possíveis, é comum que os dados não estejam preparados para que os métodos de Mineração de Dados sejam aplicados diretamente. Dependendo da qualidade desses dados, algumas ações podem ser necessárias. Este processo de limpeza dos dados geralmente envolve filtrar, combinar e preencher valores vazios.

PROCESSO DE TRATAMENTO

- **Modelagem:** É nesta fase que as técnicas (algoritmos) de mineração serão aplicadas. A escolha da(s) técnica(s) depende dos objetivos desejados.

PROCESSO DE TRATAMENTO

- **Avaliação:** Considerada uma fase crítica do processo de mineração, nesta etapa é necessária a participação de especialistas nos dados, conhecedores do negócio e tomadores de decisão. Diversas ferramentas gráficas são utilizadas para a visualização e análise dos resultados (modelos).

TIPOS DE DADOS

- **Dados temáticos:** distribuição espacial de uma grandeza geográfica, como mapas de vegetação e declividade;
- **Dados cadastrais:** cada um de seus elementos é um objeto geográfico, que possui atributos;
- **Modelo Numérico de Terreno (MNT):** é utilizado para denotar a representação quantitativa de uma grandeza que varia continuamente no espaço;

TIPOS DE DADOS

- **Imagens:** Obtidas por satélites, fotografias aéreas ou scanners. Representam formas de captura indireta de informação espacial. Armazenadas como matrizes, cada elemento da imagem (pixel) tem valor proporcional à energia eletromagnética refletida pela superfície terrestre;

TIPO DE ARMAZENAMENTO

Atualmente os arquivos geoespaciais são armazenados nos seguintes formatos do tipo vetorial:

- 1. Shapefile;**
- 2. Geopackage;**
- 3. GeoJSON;**
- 4. CSV.**

Arquivos matriciais (tipo raster)

- 1. GeoTIFF;**
- 2. JPEG 2000;**
- 3. ENVI.**

SHAPEFILE

- **Formato de Arquivo:** O Shapefile é um formato de arquivo geoespacial desenvolvido pela Esri para armazenar informações geográficas, como pontos, linhas e polígonos.

SHAPEFILE

- **Extensões de Arquivo:** Os arquivos Shapefile geralmente têm três extensões de arquivo associadas (obrigatórios): .shp (contém geometrias), .shx (contém índices) e .dbf (contém atributos). Há arquivos que não são obrigatórios, mas a depender do uso, serão gerados pelo SIG, são: .prj (SRC), .idx (indexação AutoCAD), .sbn (índice espacial), .cpg (código de arquivos)

SHAPEFILE

- **Geometrias:** Os Shapefiles podem armazenar diferentes tipos de geometrias, incluindo pontos, linhas e polígonos, permitindo representar objetos geográficos do mundo real.
- **Atributos:** Além das informações geométricas, os Shapefiles também podem conter atributos associados a cada geometria. Esses atributos são armazenados em um arquivo DBF (dBase) e podem conter dados alfanuméricos.

SHAPEFILE

- **Projeção:** Shapefiles podem incluir informações de projeção que definem como as coordenadas geográficas se relacionam com as posições na superfície da Terra. Isso é crucial para a correta representação espacial (prj).
- **Índices:** O arquivo de índice (.shx) ajuda a otimizar a recuperação de dados espaciais, permitindo um acesso mais rápido às geometrias armazenadas no arquivo principal (.shp).

SHAPEFILE

- **Limitações:** Shapefiles têm algumas limitações, como a capacidade de armazenar apenas um tipo de geometria por arquivo e um limite de tamanho de arquivo de 2 GB.
- **Suporte:** Shapefiles são amplamente suportados por sistemas de informações geográficas (SIG) e software de visualização, tornando-os um formato popular para compartilhar dados geoespaciais.

GEOPACKAGE

- **Formato de Arquivo:** O Geopackage é um formato de arquivo geoespacial que combina informações geográficas e atributos em um único arquivo, tornando-o mais flexível e eficiente.
- **Extensão de Arquivo:** Os arquivos Geopackage geralmente têm a extensão .gpkg.
- **Geometrias e Atributos:** O Geopackage suporta vários tipos de geometrias, incluindo pontos, linhas e polígonos, e permite a inclusão de atributos alfanuméricos para cada geometria.

GEOPACKAGE

- **Projeção:** Os Geopackages podem incluir informações de projeção, permitindo que os dados sejam interpretados corretamente no contexto espacial.
- **Armazenamento Único:** Diferentemente dos Shapefiles, o Geopackage armazena todas as informações (geometrias, atributos e metadados) em um único arquivo, facilitando a gestão e o compartilhamento.

GEOPACKAGE

- **Estrutura de Banco de Dados:** Internamente, o Geopackage utiliza uma estrutura de banco de dados SQLite, o que permite maior flexibilidade e complexidade nas estruturas de dados.
- **Índices e Consultas:** O Geopackage suporta índices que melhoram o desempenho das consultas espaciais, tornando mais rápido o acesso às informações geoespaciais.

GEOPACKAGE

- **Tamanhos de Arquivo:** Os Geopackages podem lidar com arquivos maiores do que os Shapefiles, reduzindo a necessidade de dividir os dados em vários arquivos menores.
- **Suporte:** O formato Geopackage é amplamente suportado por SIG e outras ferramentas de geoprocessamento. É, também, um padrão do Consórcio Geoespacial Aberto (OGC), o que significa que é reconhecido internacionalmente e segue especificações rigorosas.

GEO-OBJETO E OBJETO NÃO-ESPACIAL

- Um geo-objeto é um elemento único que possui atributos não-espaciais e está associado a localizações geográficas.
- Um objeto não-espacial é um objeto que não possui localizações espaciais associadas.

GEO-OBJETO E OBJETO NÃO-ESPACIAL

Em muitas situações é conveniente permitir a associação de informações não espaciais a um banco de dados georeferenciado. Por exemplo, considere uma aplicação de cadastro urbano em uma prefeitura que já dispõe de um sistema para cálculo do IPTU baseado num cadastro alfanumérico de lotes. É POSSÍVEL associar o cadastro alfanumérico a dados georeferenciados contendo a localização geográfica e as dimensões destas propriedades. Para englobar estas entidades, introduz-se a noção de objeto não-espacial.

GEO-OBJETO E OBJETO NÃO-ESPACIAL

Ou seja, a noção de objeto não-espacial engloba qualquer tipo de informação que não seja georeferenciada e que se queira agregar a um SIG.

Exemplos:

1. Dados do censo demográfico;
2. Matrizes Origem/Destino;
3. Entre outros.

PROCESSO DE TRATAMENTO

Com uma visão inicial dos dados definida, é necessário **explorá-los**, buscando, além de mais conhecimento sobre os mesmos, encontrarmos valores que possam comprometer sua qualidade, tais como: valores em branco ou nulo, valores viciados, variáveis duplicadas, entre outras.

PROCESSO DE TRATAMENTO

Além das avaliações dos dados, são necessários tratamentos de erros (grosseiros, por exemplo) ou falhas do banco de dados. Para isso é importante conhecimento pleno dos campos do banco de dados, sua estrutura e suas propriedades.

ANÁLISE EXPLORATÓRIA

A finalidade da **Análise Exploratória de Dados (AED)** é examinar os dados previamente à aplicação de qualquer técnica estatística. Desta forma o analista consegue um entendimento básico de seus dados e das relações existentes entre as variáveis analisadas.

ANÁLISE EXPLORATÓRIA

A AED vai além do uso descritivo da estatística, procura olhar de forma mais profunda os dados, sem resumir muito a quantidade de informações.

Várias técnicas estatísticas de análise de dispersão (Quartiles, Variância) e de medida central (média, mediana, moda e faixa de valores) combinadas com gráficos (Histogramas, Frequência, Barra, BoxPlot, Dispersão) são usadas para a exploração dos dados.

RESUMO

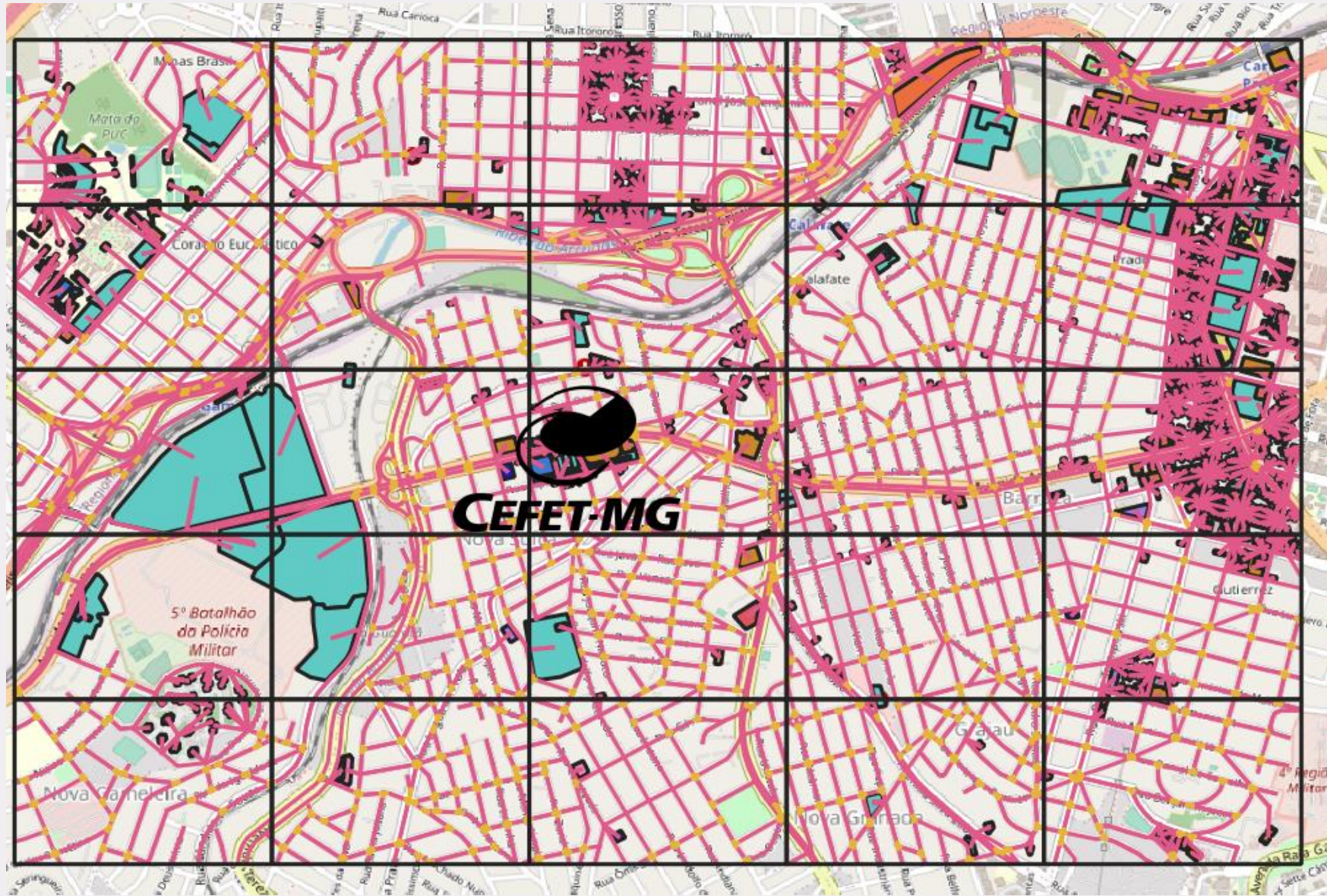
As representações do universo podem tomar duas formas, sejam:

1. Representação vetorial;
2. Representação matricial.

RESUMO

- Na representação vetorial, a representação de um elemento ou objeto é uma tentativa de reproduzi-lo o mais exatamente possível. Qualquer entidade ou elemento gráfico de um mapa é reduzido a três formas básicas: pontos, linhas, áreas ou polígonos.

VETORIAL



Conclusão

Desenvolvimento

Introdução

RESUMO

- A representação matricial consiste no uso de uma malha quadriculada regular sobre a qual se constrói, célula a célula, o elemento que está sendo representado. A cada célula, atribui-se um código referente ao atributo estudado, de tal forma que o computador saiba a que elemento ou objeto pertence determinada célula.

MATRICIAL

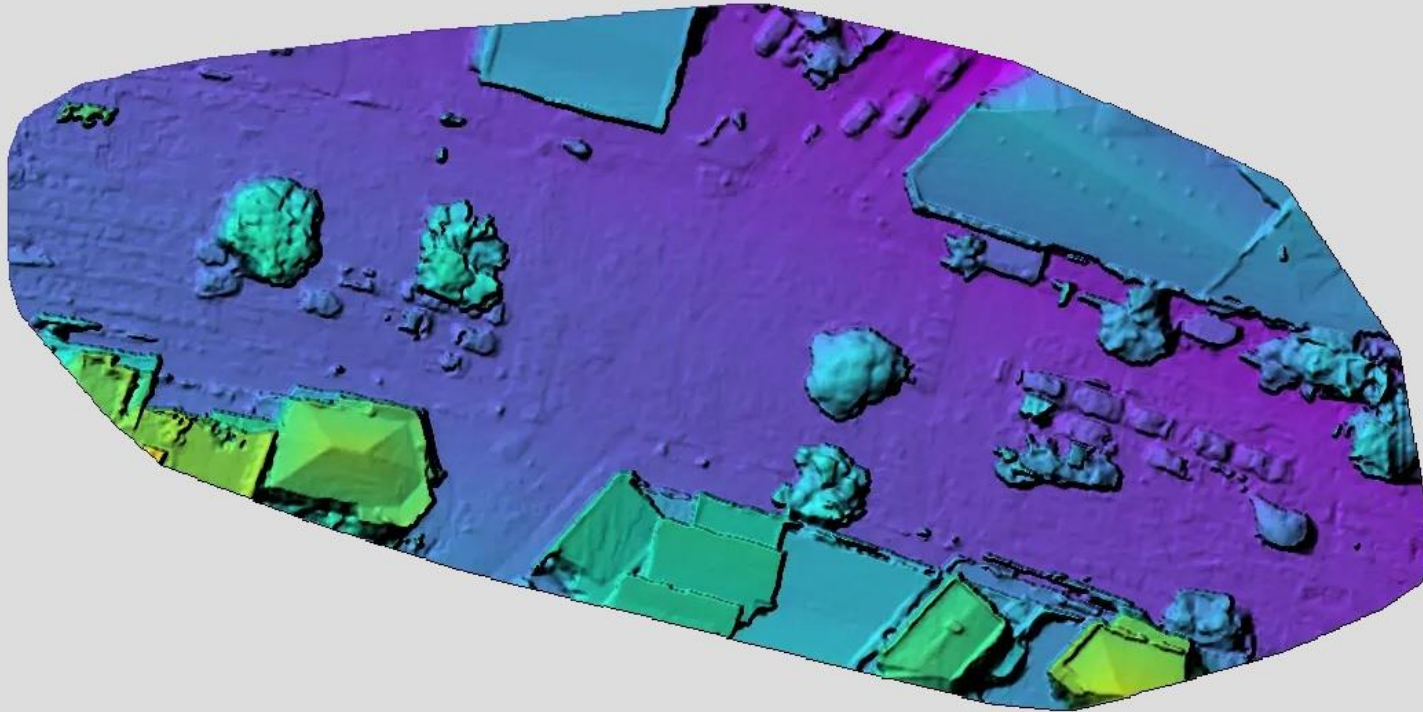


Conclusão

Desenvolvimento

Introdução

MATRICIAL



30 m
100 ft

-19.92909 ; -43.97579

Conclusão

Desenvolvimento

Introdução