

VISUALIZAÇÃO GEOGRÁFICA DAS ESCOLAS DE SANTA CATARINA

Joel João Medeiros Filho <joel.joao.medeiros@gmail.com>

Nayara Mara Honorato da Silva <nayaramarah@gmail.com>

Vinicius Rover <viniciusrover3@gmail.com>



RESUMO

O presente estudo do projeto mapa escolas SC, realizou-se no intuito de: gerar visualizações geográficas das escolas de Santa Catarina. Para isso foram aplicadas técnicas de limpeza, organização e aquisição de informações, que resultaram em um arquivo que pode ser utilizado para outros projetos. Os resultados se projetam como relevantes para a abertura dos estudos na área, mesmo com seus problemas, as informações obtidas com o mapa de calor corroboram com outras fontes de estudo de populações e distribuição populacionais regionais.

Palavras-Chave: Escolas, Visualização, Geográfica, Heatmap, Santa Catarina

A VISUALIZAÇÃO DE DADOS

Foi dito que "uma imagem vale mais que mil palavras". E hoje, na era do big data, quando as empresas são inundadas com informações de vários tipos de dados e de fontes on-premise e baseadas na nuvem, esse velho ditado nunca foi tão relevante.

Peneirar informações para entender o que importa é o que não importa está se tornando mais difícil. O visual torna a análise muito mais fácil e rápida e oferece a capacidade de ver de relance o que é importante. Além disso, a maioria das pessoas responde muito melhor aos recursos visuais do que ao texto - 90% das informações enviadas ao cérebro são visuais, e o cérebro processa os recursos visuais a 60.000 vezes a velocidade do texto¹. Esses pontos são um forte argumento para o uso da visualização de dados para analisar e transmitir informações.

A visualização de dados faz parte de muitas ferramentas de inteligência de negócios e é essencial para análises avançadas. Ela ajuda as pessoas a entender todas as informações ou dados gerados hoje. Com a visualização de dados, as informações são representadas em forma gráfica, como um gráfico de pizza ou outro tipo de apresentação visual.

A boa visualização de dados é essencial para analisar e tomar decisões com base nesses dados. Ela permite que as pessoas vejam e entendam com rapidez e facilidade padrões e relacionamentos e identifiquem tendências emergentes que podem passar despercebidas com apenas uma tabela ou planilha de números brutos. E na maioria dos casos, nenhum treinamento especializado é necessário para interpretar o que é apresentado nos gráficos, permitindo a compreensão universal.

Um gráfico bem projetado pode não apenas fornecer informações, mas também aumentar o impacto dessas informações com uma apresentação forte, atraindo a atenção e mantendo o interesse do jeito que nenhuma tabela ou planilha consegue fazer.

O TEMA ESCOLHIDO

A visualização de dados geoespaciais se concentra na relação entre os dados e sua localização física para criar insights. Quaisquer dados posicionais funcionam para análise espacial. A geovisualização sobrepõe variáveis em um mapa usando latitude e longitude para promover a percepção.

Os mapas são o foco principal das visualizações geoespaciais. Isso permite criar contexto usando formas e cores para alterar o foco visual. Eles ajudam a identificar problemas, rastrear mudanças, entender tendências e realizar previsões relacionadas a locais e horários específicos.

Destacam a conexão física entre os pontos de dados através de pontos, linhas ou regiões, que, conforme a informação, podem ter diferentes propriedades como tamanho, forma, espessura, textura e cor. Os mapas podem ser subdivididos de acordo com as propriedades dos dados (qualitativo ou quantitativo, discreto ou contínuo) e das variáveis geográficas (pontos, linhas, superfícies e volumes).

Existem dois tipos de dados geoespaciais, os chamados dados vetoriais, usados para mapear fenômenos ou objetos espacialmente explícitos que possuem localização ou dimensões bem definidas, representado a partir de formas geométricas (como pontos, linhas e polígonos).

O outro tipo é dados no formato raster, que consistem em uma matriz em que os elementos representam células. As células dos dados raster possuem duas informações: das células para especificar sua posição na matriz e valores das células, onde são coloridos para facilitar a interpretação da variação dos valores no espaço.

Além disso possuem um componente fundamental que é o Sistema de Referência de Coordenadas. Esse componente define a referência espacial dos elementos geoespaciais na superfície da Terra.

Um conceito dos CRS é o *datum*, que é a relação do sistema de coordenadas (geográfica ou projetada) com a superfície da Terra. O outro conceito é que tipos de unidades estão sendo utilizadas para a representação geográfica, podendo assumir dois tipos - ângulos ou metros, que definem os:

- O Sistema de Coordenadas Geográficas, utiliza ângulos (graus) para representar feições na superfície da Terra através de dois valores: longitude e latitude

- O Sistema de Coordenadas Projetadas utiliza um Sistema Cartesiano de Coordenadas em uma superfície plana, traçando sistemas tridimensionais em superfícies bidimensionais. Essa conversão faz com que haja distorções de diversos tipos.

Muitos dados são adquiridos hoje em dia, principalmente com a evolução da tecnologia. Com isso, aplicações são criadas constantemente com diversos objetivos em mente. Os projetos PRODES e IPAM são um exemplo de aplicações, onde é realizado o monitoramento das taxas anuais de desmatamento da Amazônia através de satélites. É feito o geoprocessamento dos dados coletados e renderizados em um gráfico.

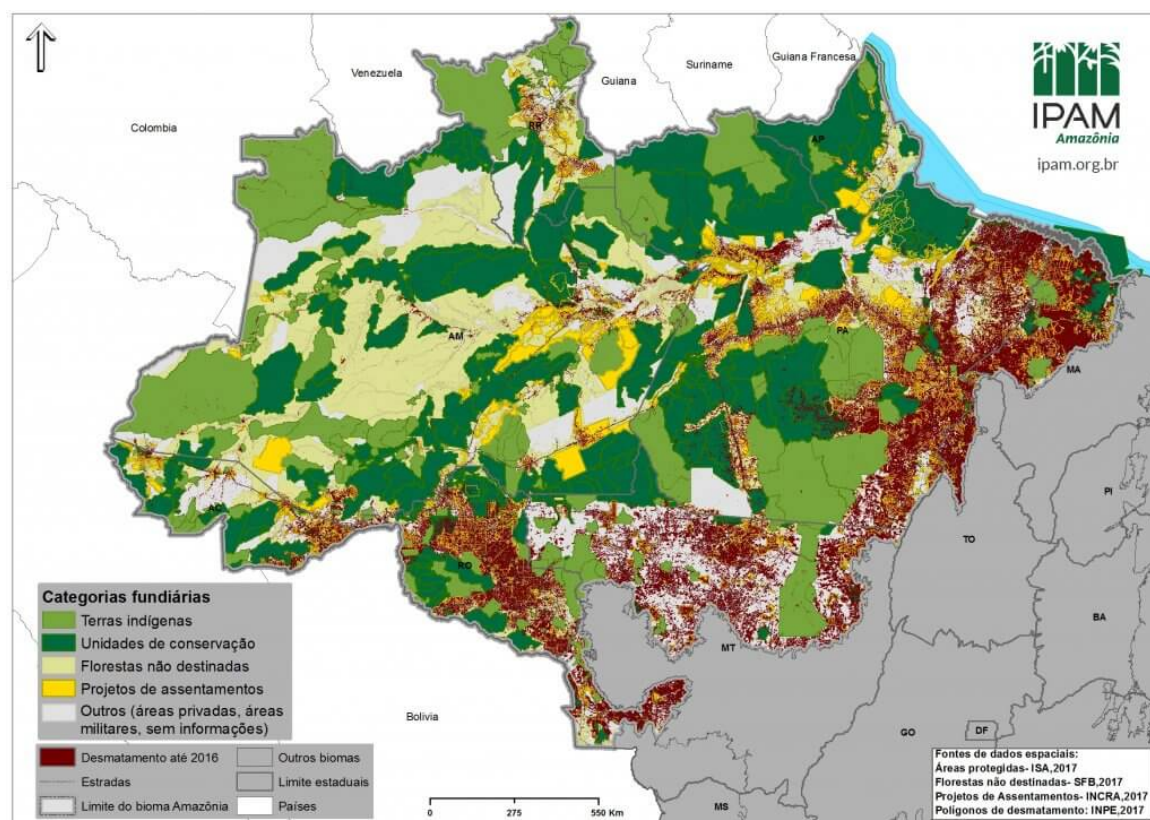


Figura 1 - Mapa do desmatamento gerado pelo projeto IPAM.

Uma das aplicações mais famosas é o GPS, que funciona a partir de uma rede de 24 satélites na órbita próxima da Terra. Estes, por sua vez, trocam sinais com o seu dispositivo e, a partir disso, são capazes de dizer onde você está na superfície da Terra, que pode ser usado tanto em terra, alto-mar e áreas da aviação.

Cada empresa cria seu próprio mapa, assim mostrando os locais, e utiliza os graus de latitude e longitude para saber a localização do dispositivo, estabelecimento e entre

outras coisas Um muito conhecido é o Google maps.

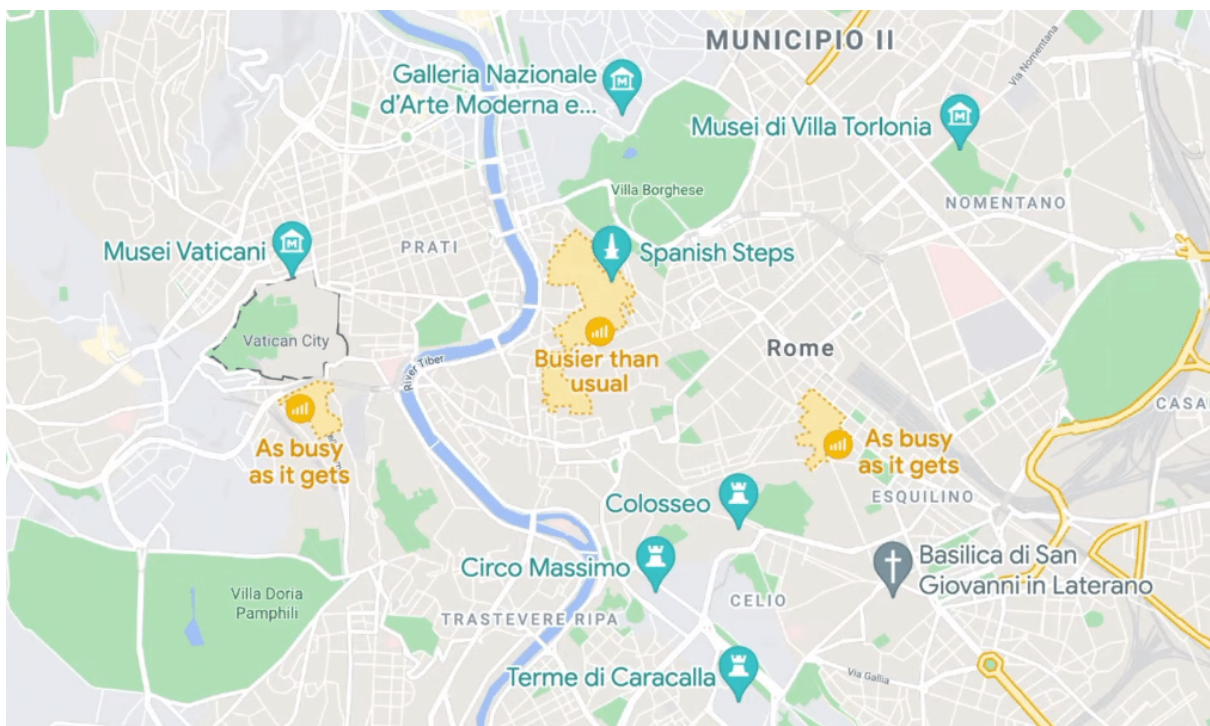


Figura 2 - Imagem tirada do Google Maps

BASE DE DADOS E OS DADOS

Os dados apresentados são de diferentes fontes, já que não existia consistentemente nenhuma base de dados já organizada. Nesse sentido, a construção de uma nova base de dados foi necessária, realizando requisições a um end-point API para adquirir as coordenadas dos endereços da base de dados principais que contém as informações das escolas. Então nesse sentido temos: os dados das informações das escolas, os dados requisitados pela API e ainda o geojson que define os limites do estado de SC.

GEOJSON

O arquivo geojson é um padrão aberto de organização e processamento de dados pensado para comportar as necessidades de explicitar todas as estruturas de dados comuns para as representações de mapas, marcações, linhas e agrupamentos desses e outros elementos, detalhado e explicitado pela norma RFC 7946.

No presente projeto o arquivo GeoJSON foi utilizado para a definição do contorno das fronteiras de Santa Catarina, foi obtido a partir da query do site *overpass turbo* que utiliza dados abertos do *Open Street Map*. ele contém linhas que determinam o entorno do estado utilizando as coordenadas desses determinados pontos.

XLXS

Os arquivos xlsx foram obtidos através do portal de cadastro da Unidade Escolar, que fornece a busca de todas as instituições de ensino no estado de Santa Catarina, no âmbito: federal, estadual, municipal, privado e público. A obtenção dos dados se deu complexamente, haja visto que o site não atualiza quando novas requisições foram efetuadas. Os dados relevantes para o trabalho foram os dados que continham informações de endereço, utilizados na busca das coordenadas geográficas pela API.

API

Os dados de coordenadas foram obtidos a partir da API do site MapQuest que contém um plano gratuito de 15000 requisições, que foram suficientes para enviar os endereços de 100 em 100 utilizando requisições batch. Os dados são salvos em uma estrutura binária python utilizando a biblioteca pickle do python, eles são integrados e salvos na estrutura final chamada 'dataframe.csv'.

METODOLOGIA

Neste trabalho, utilizamos múltiplas tecnologias para o desenvolvimento de nossos resultados. Inicialmente, nossa base de dados consistia em 4 planilhas que continham informações sobre escolas separadas por dependência. O primeiro passo foi juntar essas planilhas em uma só, para isso, utilizamos a biblioteca Pandas do Python.

Além disso, a biblioteca Pandas também foi utilizada para remover algumas colunas vazias que existiam nas nossas planilhas. Para facilitar o entendimento e evitar possíveis conflitos, renomeamos algumas colunas, como a conversão de "N°" para "num".

Para utilizar o API, precisamos criar um dado para cada tupla da nossa base de dados, contendo o endereço completo de cada escola. Para isso, concatenamos as colunas que possuíam os dados de endereço separados, adicionando algumas marcações e textos.

Dado os requisitos de uso da API MapQuest que realizou a busca das coordenadas, foi necessário separar os endereços completos a cada 100. Para isso, criamos uma função que realiza esse processo.

Também criamos funções para limpar strings de espaços vazios utilizando *regex*. A própria consumação da API foi feita por uma função que obtém através da biblioteca *requests*, resultados em formato JSON. Todos os resultados obtidos foram concatenados e, em seguida, processados por uma função que extrai, de cada requisição, as coordenadas dos 100 respectivos endereços, adicionando todos a uma lista com seu endereço, latitude e longitude.

Para evitar que o processo de consumo da API seja repetido várias vezes, criamos um binário utilizando o serializador de objetos Python Pickle. Se esse binário não for apagado, o processo de consumo da API não será realizado.

Até aqui todos os processos foram somente de aquisição e tratamento dos dados, que geraram o arquivo `dataframe.csv`, que contém o endereço completo, coordenadas, nome e outras informações das escolas.

Para a aquisição do mapa de calor, utilizamos a biblioteca Folium, a qual é uma biblioteca Python que junta as forças com a biblioteca leaflet.js, para criar mapas interessantes com muitas utilidades.

Para tal, criamos um mapa com foco na localização media de Santa Catarina, carregamos os arquivos das bordas do estado de Santa Catarina, para que se demarcasse os limites e carregamos o HeatMap do folium com os dados do arquivo `dataframe.csv`.

RESULTADOS

Os resultados bem como todo o processo podem ser encontrados no link do GitHub do projeto, sendo o resultado do projeto podendo ser acessado no seguinte endereço: https://medeirosjoel.github.io/mapa-Escolas-SC/mapas/mapa_cru.html, abaixo uma snapshot do site.

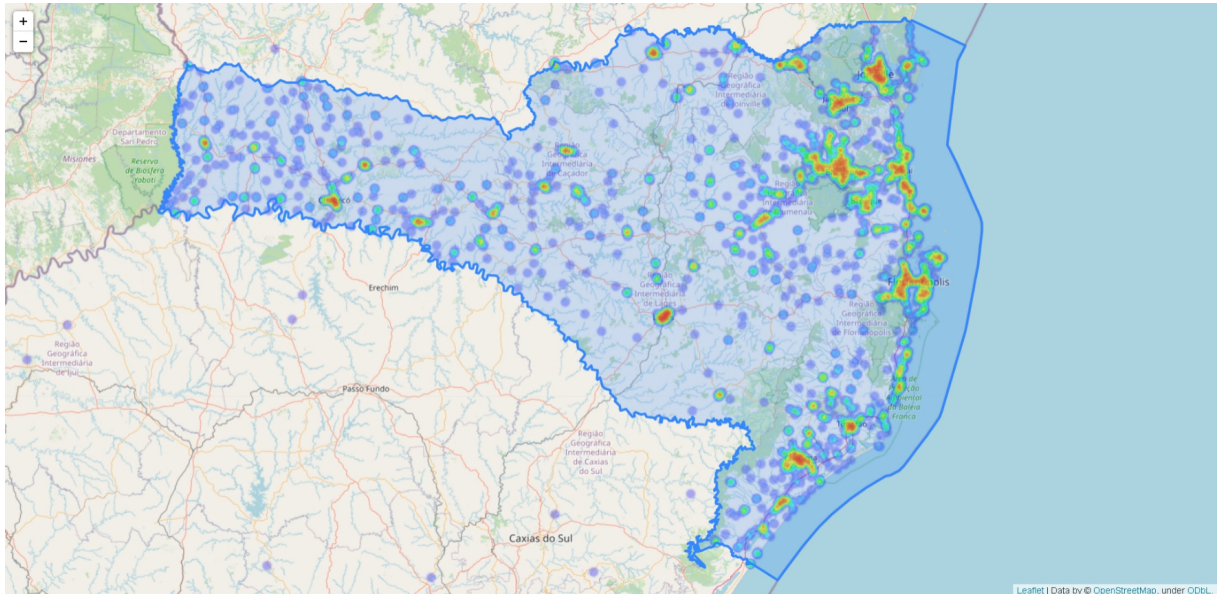


Figura 3 - Snapshot resultado final.

Como pode ser observado, o mapa nos revela alguns *insights*, como, por exemplo, a área litorânea contém mais escolas, existe uma relativa boa distribuição de escolas pelo estado. O mapa foi distribuído para focar no Estado de Santa Catarina contornando suas fronteiras, e focalizando o centro do estado para melhor visualização.

Podem se colocar em dúvidas quanto a precisão e a assertividade do mapa, haja visto que a recolha das coordenadas geográficas se deu por uma API que tem certo nível de imprecisão. Como pode ser visto alguns pontos fora de Santa Catarina, erros realizados pela API, infelizmente dado o orçamento zero do projeto, outras fontes mais precisas não puderam ser utilizadas.

CONSIDERAÇÕES FINAIS

O aprendizado adquirido com o projeto foi excelente para a construção de nossa formação acadêmica e profissional, além da contribuição que traz a comunidade, de ao menos sumarizar as informações, das escolas em uma fonte única.

Novas iterações com o projeto renderiam sem dúvidas frutos ainda mais interessantes, principalmente no que tange a correlacionar essa visualização e esses dados a outras fontes de dados como notas do Enem das escolas, por exemplo.

Os problemas de localizações fora do mapa de Santa Catarina, por exemplo, pode ser tratado melhor utilizando a API de *geocoding* do Google, a qual é mais precisa e exata, porém é paga.

Realizar o projeto, tendo que formar uma base de dados no processo foi um desafio, que trouxe uma complexidade que sem dúvidas afetou a execução do projeto, apesar disso apresentamos um resultado que pode ser muito interessante para inclusive novos projetos.

No fim, tanto pelo processo quanto pelo resultado, os autores ficam satisfeitos com o resultado, percebendo as melhorias possíveis e as possibilidades interessantes que o projeto pode obter, no âmbito educacional regional do estado, que detém de bons índices, mas ainda possui muitas discrepâncias regionais.

REFERÊNCIAS BIBLIOGRÁFICAS

Tema Escolhido

Domingues, Augusto, et al. "Dados geoespaciais: Conceitos e técnicas para coleta, armazenamento, tratamento e visualização." *Sociedade Brasileira de Computação* (2020);

SILVA, Fernando Rodrigues da; GONÇALVES-SOUZA, Thiago; PATERNO, Gustavo Brant; PROVETE, Diogo Borges; VANCINE, Maurício Humberto. **Análises Ecológicas no R**. 2022.

Disponível em:

https://paternogbc.github.io/livro_aer/cap15.html#exemplos-de-aplica%C3%A7%C3%B5es-de-an%C3%A1lises-geoespaciais-para-dados-ecol%C3%B3gicos. Acesso em: 11 dez. 2022;

SIGNORETTI, Wagner Gonçalves. **MODELO PARA ANÁLISE VISUAL DE DADOS MULTIVARIADOS NO CONTEXTO GEOESPACIAL**. 2018. 92 f. Dissertação (Mestrado) - Curso de Ciência da Computação, Universidade Católica do Rio Grande do Sul, Porto Alegre, 2018

Base de Dados

"RFC 7946: The GeoJSON Format." <https://www.rfc-editor.org/rfc/rfc7946>, Acessado 10 de dezembro. 2022;

"overpass turbo" Overpass-turbo, <https://overpass-turbo.eu/s/1oVy>. Acessado 10 de dezembro de 2022;

"Cadastro de Unidade Escolar" Serieweb, <http://serieweb.sed.sc.gov.br/cadueportal.aspx>. Acessado 10 de dezembro de 2022;

"MapQuest Developer Network | Mapping, Geocoding, Routes, Traffic | MapQuest Developer Network" Developer, <https://developer.mapquest.com/>. Acessado 10 de dezembro de 2022;

Metodologia

"Batch Geocode - GET | MapQuest API Documentation" Developer, <https://developer.mapquest.com/documentation/geocoding-api/batch/get/>. Acessado 10 de dezembro de 2022;

"pickle — Serialização de objetos Python — documentação Python 3.11.1" Docs, <https://docs.python.org/pt-br/3/library/pickle.html>. Acessado 10 de dezembro de 2022;

Story, Rob . "Folium — Folium 0.12.1 documentation" Python-visualization,
<https://python-visualization.github.io/folium/>. Acessado 10 de dezembro de 2022;

Resultados

Medeiros Filho, Joel João, et al. ""MedeirosJoel/mapa-Escolas-SC" Mapa Escolas SC,
<https://github.com/MedeirosJoel/mapa-Escolas-SC>. Acessado 10 de dezembro de 2022.