INDE Infraestrutura Nacional de Dados Espaciais
IDEs Inteligentes para governança e políticas públicas
IV Simpósio Brasileiro de Infraestrutura de Dados Espaciais - SBIDE
15 a 18 de outubro de 2024
USP Universidade de São Paulo · UFPR Universidade Federal do Paraná · uff Universidade Federal Fluminense · IBGE Instituto Brasileiro de Geografia e Estatística

GIS, Hadoop e Hive: Um Estudo de Caso da Criminalidade na Cidade de São Paulo

Italo L. F. Portinho
Instituto de Computação – Universidade Federal Fluminense (UFF)
italoleite@id.uff.br, italons@gmail.com
https://github.com/italoportinho

Este estudo consiste em uma análise detalhada de subconjutos de um dataset composto pelo registro de determinados crimes na cidade de São Paulo. O dataset em questão é o PolRoute-DS, e pode ser obtido no endereço https://osf.io/mxrgu/. O dataset no formato csv foi importado para um datawarehouse (Apache Hive), para possibilitar rodar consultas em paralelo de forma eficiente em um ambiente distribuído. O objetivo é extrair sub-datasets com os crimes por ano e distrito, com e sem sua geometria, e também rodar uma consulta para descobrir o ano com mais crimes e deste ano extrair um dataset com os crimes por dia e mês para poder fazer uma *HeatMatrix*. Foi utilizado o software Quantum GIS para converter a geometria do formato EWKB(Extended Weel Know Binary) para Lat/Long na projeção EPSG:4326, e exportar o dataset para o formato GeoJSON. O dataset com a geometria dos distritos foi separado por ano devido ao seu tamanho, e somente foram considerados os anos de 2011 a 2018, por apresentarem registros mais completos. Também foi extraído pelo Visualizador da INDE, a camada “Distritos do município de São Paulo” da IDE-SP, uma camada com limites para fins de melhor visualização.

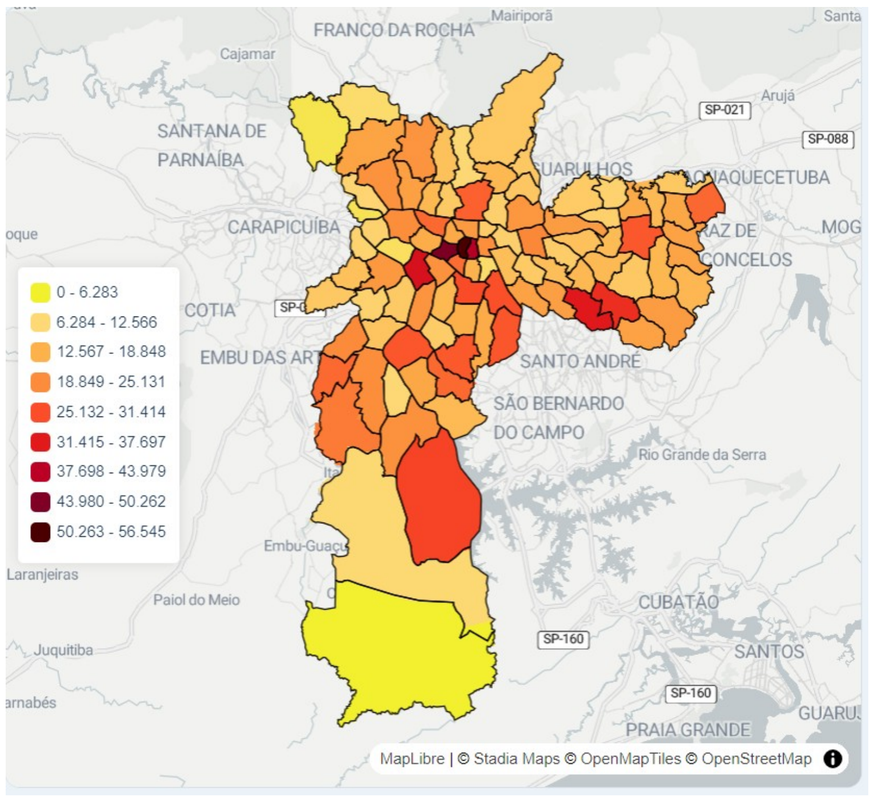

Figura 1: Índice de criminalidade nos distritos de São Paulo no ano de 2017. Foram gerados datasets separados para os anos de 2011 a 2018 e comparados pelo índice criminal, sendo 2017 o ano mais violento. Nesse ano, República é o distrito com maior índice criminal.

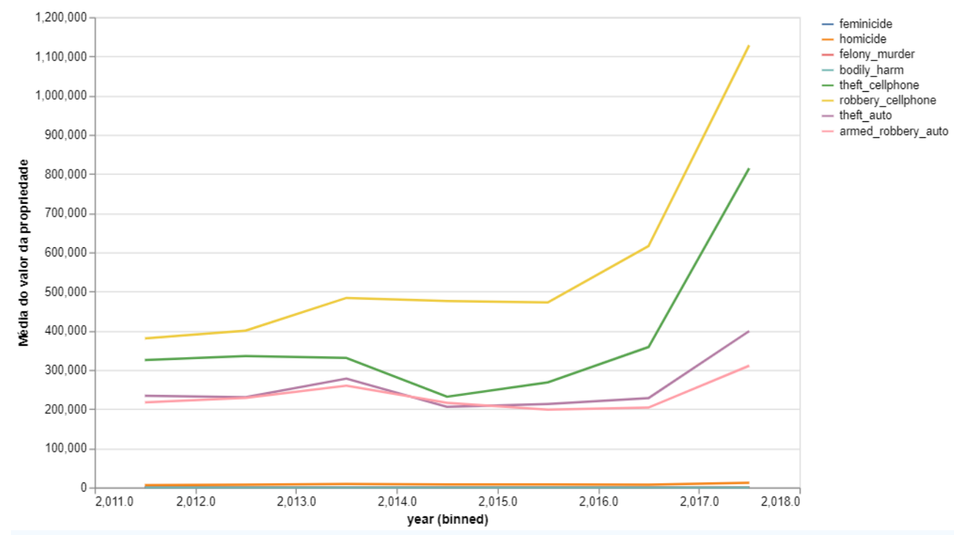

Figura 2: Crimes mais prevalentes no período considerado. Delitos relacionados à subtração de telefones celulares são de longe os mais presentes.

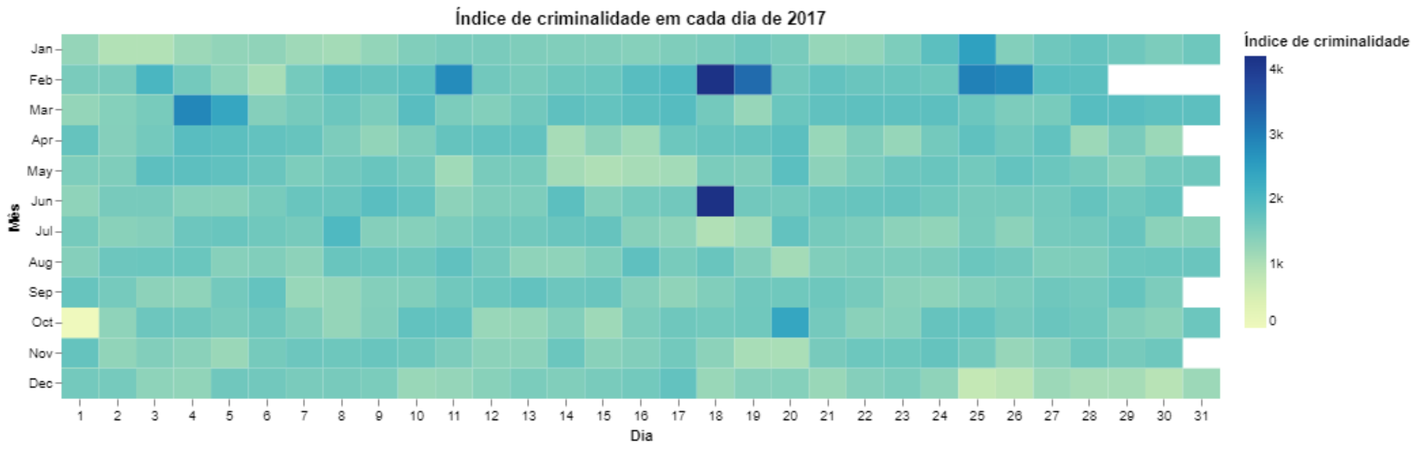

Figura 3: Matriz de calor da criminalidade em 2017. Os finais de semana de Carnaval e da realização da parada do orgulho LGBTQIA+ foram os mais violentos.

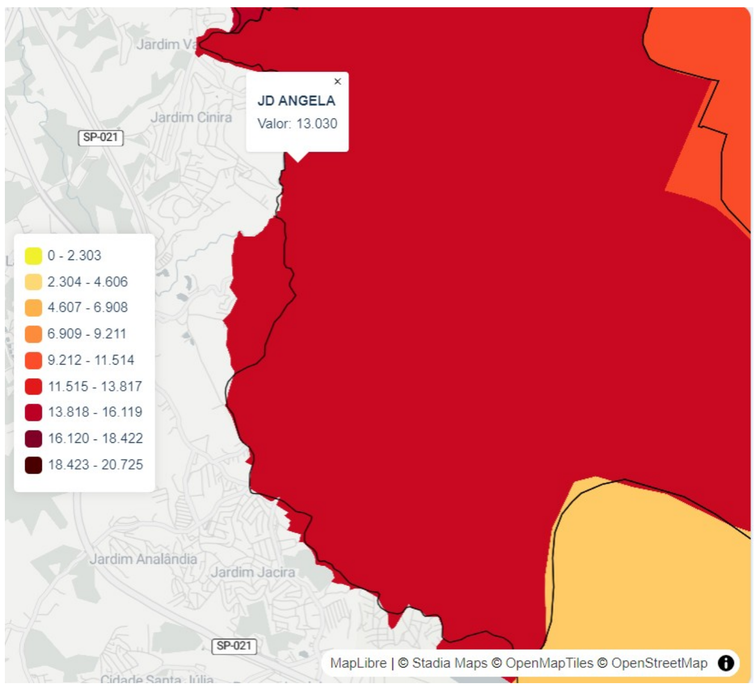

Figura 4: São percebidas imperfeições na geometria do dataset quando comparada com a obtida no Visualizador da INDE.