

Interpretabilidade em Portais de Dados Abertos Georreferenciados, Um Estudo de Caso

Sistemas de Governo Eletrônico(2024-1) – Profª Flávia Bernardini
Lucas Costa(G), Ruan Pablo(G), Ítalo Portinho(PG)

Sumário

- 1) Interpretabilidade
- 2) A INDE
- 3) O Dataset
- 4) O Cluster
- 5) Import/Export
- 6) Geometrias
- 7) Visualizações

Interpretabilidade

Raissa Barcellos, Flavia Bernardini, José Viterbo,
**Towards defining data interpretability in open data portals:
Challenges and research opportunities**, Information Systems,
Volume 106, 2022, 101961, ISSN 0306-4379,
<https://doi.org/10.1016/j.is.2021.101961>.

Interpretabilidade

- **Motivação: Buscar uma definição formal para a interpretabilidade de dados em portais de transparência**
- **Como?**
 - ◆ **Revisão sistemática de literatura**
 - ◆ **Aplicação de questionários a especialistas**
 - **Validar características identificadas**
 - **Traçar conceito único de interpretabilidade**

Interpretabilidade

- (G1) Compreensibilidade, simplicidade, clareza e redigibilidade;
- (G2) Confiabilidade e rastreabilidade;
- (G3) Estrutura e organização;
- (G4) Acurácia e correção;
- (G5) Completude;
- (G6) Concisão;
- (G7) Consistência e coerência;
- (G8) Informatividade;

Interpretabilidade

→ Definição encontrada:

“Interpretabilidade é a capacidade de um conjunto de dados, com total precisão, consistência, coerência e organização, de transmitir significado de forma clara e compreensível ao usuário.”

Interpretabilidade

	Dados.gov.br	Data.europa.eu	Data.gov
Dicionários de dados	✓	✓	✓
Rastreamento automático de dados	x	x	x
Rastreamento manual de dados	✓	✓	✓
Linguagem acessível ao cidadão(tradução)	x	x	x
Descrições claras sobre cada conjunto de dados	✓	✓	✓
Organização na interface	✓	✓	✓
Metadados estruturados para cada conjunto de dados	✓	✓	✓
Identificadores semânticos padronizados	x	✓	x
Dados em tempo real	x	x	x
Comunidade pública de veracidade de dados	x	x	✓
Equipe interna para avaliar qualidade e precisão dos dados	x	✓	x
Dados vinculados de outras fontes	x	✓	✓
Descrições de entidades de dados vinculados	x	✓	✓
Informações sobre o que está sendo realmente divulgado	✓	✓	✓
Análise semântica para recuperação de dados	x	x	x
Linguagem padronizada e unificada	x	✓	✓
Estratégia de similaridade entre vetores contextuais	x	x	x
Modelos ontológicos	✓	✓	✓
Vocabulário controlado	x	✓	✓
Metadados padronizados para cada conjunto de dados	x	✓	✓
Mecanismos para sugerir dados faltantes	✓	✓	✓
Medida de valor ou utilidade dos dados divulgados	✓	x	x
Pesquisas de satisfação	x	x	x

Conjunto integrado de tecnologias, políticas, mecanismos e procedimentos de coordenação e monitoramento, padrões e acordos, necessário para facilitar e ordenar a geração, o armazenamento, o acesso, o compartilhamento, a disseminação e o uso dos dados geoespaciais de origem federal, estadual, distrital e municipal

O Ecossistema de aplicações da INDE:

- Portal
- **Visualizador**
- Catálogo de Metadados(Geonetwork)
- Catálogo de Geosserviços(Geoserver)
- Integrações
- Aplicações administrativas: monitoramento, extrator de camadas, Shape INDE, Gestão INDE.

A INDE

	Visualizador.gov.br	Dados.gov.br	Data.europa.eu	Data.gov
Dicionários de dados	✓	✓	✓	✓
Rastreamento automático de dados	x	x	x	x
Rastreamento manual de dados	✓	✓	✓	✓
Linguagem acessível ao cidadão(tradução)	x	x	x	x
Descrições claras sobre cada conjunto de dados	✓	✓	✓	✓
Organização na interface	✓	✓	✓	✓
Metadados estruturados para cada conjunto de dados	✓	✓	✓	✓
Identificadores semânticos padronizados	x	x	✓	x
Dados em tempo real	x	x	x	x
Comunidade pública de veracidade de dados	x	x	x	✓
Equipe interna para avaliar qualidade e precisão dos dados	x	x	✓	x
Dados vinculados de outras fontes	✓	x	✓	✓
Descrições de entidades de dados vinculados	✓	x	✓	✓
Informações sobre o que está sendo realmente divulgado	✓	✓	✓	✓
Análise semântica para recuperação de dados	x	x	x	x
Linguagem padronizada e unificada	✓	x	✓	✓
Estratégia de similaridade entre vetores contextuais	x	x	x	x
Modelos ontológicos	✓	✓	✓	✓
Vocabulário controlado	✓	x	✓	✓
Metadados padronizados para cada conjunto de dados	✓	x	✓	✓
Mecanismos para sugerir dados faltantes	✓	✓	✓	✓
Medida de valor ou utilidade dos dados divulgados	✓	✓	x	x
Pesquisas de satisfação	x	x	x	x

O Dataset

- PolRoute-DS: conjuntos de dados sobre a criminalidade em São Paulo, com o objetivo de melhorar o roteamento de viaturas;
- time.csv, crime.csv, segment.csv, vertice.csv, district.csv e neighborhood.csv

O Cluster

- namenode
- datanode
- hive-server
- metastore
- metastore-postgresql

Import/Export

```
hive -f database_ddl_nofrag.hql;
```

```
hadoop fs -put crime.csv hdfs://namenode:8020/user/hive/warehouse/trabalho_egov.db/crime
```

Import/Export

```
CREATE TABLE IF NOT EXISTS crime(  
  id int,  
  total_feminicide int,  
  total_homicide int,  
  total_felony_murder int,  
  total_bodily_harm int,  
  total_theft_cellphone int,  
  total_armed_robbery_cellphone int,  
  total_theft_auto int,  
  total_armed_robbery_auto int,  
  segment_id int,  
  time_id int  
)  
  
row format delimited  
fields terminated by ','  
lines terminated by '\n'  
stored as textfile location 'hdfs://namenode:8020/user/hive/warehouse/trabalho_egov_nofrag.db/crime'  
;
```

Import/Export

```
CREATE TABLE IF NOT EXISTS districts_crimes_allyears_nogeom
  row format delimited
  fields terminated by ';'
  lines terminated by '\n'
  STORED as textfile
AS
select
  'id' as id,
  'name' as name,
  'year' as year,
  'feminicide' as feminicide,
  'homicide' as homicide,
  'felony_murder' as felony_murder,
  'bodily_harm' as bodily_harm,
  'theft_cellphone' as theft_cellphone,
  'robbery_cellphone' as robbery_cellphone,
  'theft_auto' as theft_auto,
  'armed_robbery_auto' as armed_robbery_auto,
  'criminal_index' as criminal_index
;
```

Import/Export

```
INSERT INTO districts_crimes_allyears_nogeom
select
  district_frag_geom.id as id,
  district_frag_geom.name as name,
  time_frag.year as year,
  sum(crime.total_femicide) as femicide,
  sum(crime.total_homicide) as homicide,
  sum(crime.total_felony_murder) as felony_murder,
  sum(crime.total_bodily_harm) as bodily_harm,
  sum(crime.total_theft_cellphone) as theft_cellphone,
  sum(crime.total_armed_robbery_cellphone) as robbery_cellphone,
  sum(crime.total_theft_auto) as theft_auto,
  sum(crime.total_armed_robbery_auto) as armed_robbery_auto,
  sum(crime.total_femicide) +
    sum(crime.total_homicide) +
    sum(crime.total_felony_murder) +
    sum(crime.total_bodily_harm) +
    sum(crime.total_theft_cellphone) +
    sum(crime.total_armed_robbery_cellphone) +
    sum(crime.total_theft_auto) +
    sum(crime.total_armed_robbery_auto) as criminal_index
from
  vertice, district_frag_geom, segment, crime, time_frag

where time_frag.id = crime.time_id
and crime.segment_id = segment.id
and (segment.start_vertice_id = vertice.id OR segment.start_vertice_id = vertice.id)
and vertice.district_id = district_frag_geom.id
and time_frag.year IN (2011, 2012, 2013, 2014, 2015, 2016, 2017, 2018)
group by district_frag_geom.id, district_frag_geom.name, time_frag.year
order by district_frag_geom.name ASC, time_frag.year DESC
;
```


Import/Export

```
INSERT INTO districts_crimes_2011_geom
select
    distrito.*
    , geometria.geometry
from
    (select
        district_frag_geom.id as id,
        district_frag_geom.name as name,
        time_frag.year as year,
        sum(crime.total_femicide) as femicide,
        sum(crime.total_homicide) as homicide,
        sum(crime.total_felony_murder) as felony_murder,
        sum(crime.total_bodily_harm) as bodily_harm,
        sum(crime.total_theft_cellphone) as theft_cellphone,
        sum(crime.total_armed_robbery_cellphone) as robbery_cellphone,
        sum(crime.total_theft_auto) as theft_auto,
        sum(crime.total_armed_robbery_auto) as armed_robbery_auto,
        sum(crime.total_femicide) +
            sum(crime.total_homicide) +
            sum(crime.total_felony_murder) +
            sum(crime.total_bodily_harm) +
            sum(crime.total_theft_cellphone) +
            sum(crime.total_armed_robbery_cellphone) +
            sum(crime.total_theft_auto) +
            sum(crime.total_armed_robbery_auto) as criminal_index
    from
        vertice, district_frag_geom, segment, crime, time_frag
```

```
where time_frag.id = crime.time_id
    and crime.segment_id = segment.id
    and (segment.start_vertice_id = vertice.id OR segment.start_vertice_id =
vertice.id)
    and vertice.district_id = district_frag_geom.id
    and time_frag.year = 2011
group by district_frag_geom.id, district_frag_geom.name, time_frag.year)
as distrito,
(select
    district_frag_geom.id,
    district_frag_geom.geometry
    from district_frag_geom)
as geometria
where distrito.id = geometria.id
;
```

Import/Export

```
hadoop fs -cat  
hdfs://namenode:8020/user/hive/warehouse/trabalho_egov_nofrag.db/districts_crimes_2011_geom/  
* > crimes_2011.csv
```

Geometrias

[illegible]

Geometrias

```
crimes_2011.geojson ×
datasets > {} crimes_2011.geojson > [ ] features > {} 0 > {} geometry > [ ] coordinates > [ ] 0 > [ ] 67
1 {
2   "type": "FeatureCollection",
3   "name": "crimes_2011",
4   "crs": { "type": "name", "properties": { "name": "urn:ogc:def:crs:OGC:1.3:CRS84" } },
5   "features": [
6     { "type": "Feature", "properties": { "id": "1", "name": "AGUA RASA", "year": "2011",
"femicide": "0", "homicide": "25", "felony_murder": "5", "bodily_harm": "0",
"theft_cellphone": "1580", "robbery_cellphone": "2255", "theft_auto": "3230",
"armed_robbery_auto": "1745", "criminal_index": "8840" }, "geometry": { "type":
"Polygon", "coordinates": [ [ [ -46.58146994582895, -23.572987279095102 ], [ -46.
581488672022701, -23.572968291371112 ], [ -46.581498033145991, -23.572950495404555 ], [
-46.581515554364906, -23.57292920917704 ], [ -46.581551723884637, -23.57288736743077 ], [
-46.581628583915659, -23.572798453830952 ], [ -46.581781912942283, -23.
572620872713635 ], [ -46.581805485382183, -23.572605507836535 ], [ -46.581882045069698,
-23.572451116524174 ], [ -46.581905336657613, -23.572326288711469 ], [ -46.
581787129685118, -23.57211650854391 ], [ -46.581467866270465, -23.571732301601884 ], [
-46.581404106720001, -23.571646411894488 ], [ -46.581343933999719, -23.
571537899612903 ], [ -46.581337209057509, -23.571517246614079 ], [ -46.581321264584709,
-23.571438132318551 ], [ -46.581304723088557, -23.571315359184428 ], [ -46.
581286686075387, -23.570942576019874 ], [ -46.581286280693377, -23.570809335280543 ], [
-46.581290763350708, -23.57069374892842 ], [ -46.581307141058318, -23.
570538055052523 ], [ -46.581345474686032, -23.57010526701059 ], [ -46.58134504502631,
-23.570073235676119 ], [ -46.581354905009533, -23.569967646624349 ], [ -46.
581387027313632, -23.569740391714753 ], [ -46.581413879689507, -23.5696334888823 ], [
-46.581434253226725, -23.569584825557385 ], [ -46.581467078937671, -23.
569533813370128 ], [ -46.58150310645361, -23.569492503054892 ], [ -46.581607628957862,
-23.569384661572823 ], [ -46.581728569471252, -23.569274600078675 ], [ -46.
581867381860036, -23.569162167242951 ], [ -46.58203952275494, -23.569040663368092 ], [
-46.582234774760074, -23.568913922862635 ], [ -46.582387330501881, -23.
568808629036948 ], [ -46.582585191884682, -23.568688720794711 ], [ -46.582634256870975,
-23.568673396368318 ], [ -46.582653696282811, -23.568668220002525 ], [ -46.
58305735836543, -23.568649310302036 ], [ -46.583184770386602, -23.56864039150763 ], [
Ln 6, Col 3428 (1064 selected) Spaces: 4 UTF-8 LF (1) JSON
```

Geometrias

Create a Virtual Layer

Layer name crimes_2012

Build a layer by referring to the current project's vector layer names and/or any configured embedded layer names as tables in the custom SQL query.

▼ Embedded layers

Local Name	Source	Provider
crimes_2012_csv	balho-02-visdados-2024-1/datasets/crimes_2012.csv	OGR data provider

Add Import Remove

Query

```
1 select
2   GeomFromEWKB(geometry) as geom,
3   id, name, year,
4   feminicide, homicide, felony_murder, bodily_harm,theft_cellphone, robbery_cellphone, theft_auto,
5   armed_robbery_auto, criminal_index
6 from crimes_2012_csv;
```

☐ Unique identifier column

► Geometry

Help

Test

✓ Add

✗ Close

Project Edit View Layer Settings Plugins Vector Raster Database Web Mesh Processing Help



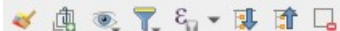
WKT

Browser

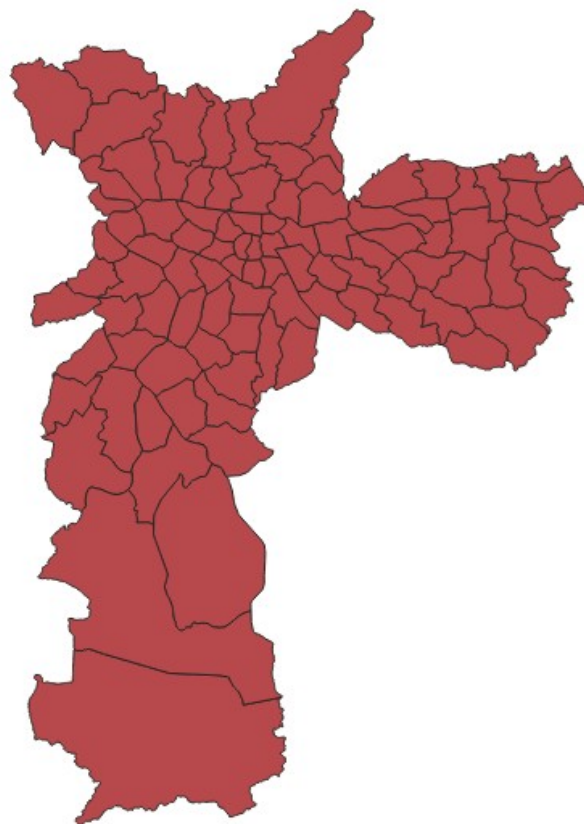


- ★ Favorites
- ▼ Spatial Bookmarks
 - Project Bookmarks
 - User Bookmarks
- ▼ Project Home
- ▼ Home

Layers



- ☐ crimes_2014
- ☐ crimes_2013
- ☒ crimes_2011
- ☐ crimes_2017
- ☐ crimes_2016
- ☐ crimes_2018
- ☐ crimes_2015
- ☐ crimes_2012



Q Type to locate (Ctrl+K)

Coordinate -23.596°, -46.086°



Scale 1:396672



Magnifier 100%



Rotation 0.0°



Render



EPSG:4326

Project Edit View Layer Settings Plugins Vector Raster Database Web Mesh Processing Help



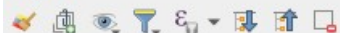
WKT

Browser



- ★ Favorites
- ▼ Spatial Bookmarks
 - Project Bookmarks
 - User Bookmarks
- ▼ Project Home
- ▼ Home

Layers



- ☐ crimes_2014
- ☐ crimes_2013
- ☒ crimes_2011
- ☐ crimes_2017
- ☐ crimes_2016
- ☐ crimes_2018
- ☐ crimes_2015
- ☐ crimes_2012

crimes_2011 — Features Total: 96, Filtered: 96, Selected: 0

Expression

- ☒ AGUA RASA
- ☐ ALTO DE PINHEIROS
- ☐ ARICANDUVA
- ☐ ARTUR ALVIM
- ☐ BARRA FUNDA
- ☐ BELA VISTA
- ☐ BOM RETIRO
- ☐ BRAS
- ☐ CAMBUCI
- ☐ CAMPO LIMPO
- ☐ CANGAIBA
- ☐ CAPAO REDONDO
- ☐ CARRAO
- ☐ CID ADEMAR
- ☐ CID LIDER
- ☐ CID TIRADENTES
- ☐ CONSOLACAO


id	1
name	AGUA RASA
year	2011
feminicide	0
homicide	25
felony_murder	5
bodily_harm	0
theft_cellphone	1580
robbery_cellphone	2255
theft_auto	3230
armed_robbery_auto	1745
criminal_index	8840

1 / 96

Show All Features

Visualizações

```
278
279 map.on('load', () => {
280   map.addSource('source_crimes', {
281     'type': 'geojson',
282     'data': layer_crimes
283   });
284   map.addLayer({
285     'id': 'layer_crimes',
286     'type': 'fill',
287     'source': 'source_crimes',
288     'layout': {},
289     'paint': {
290       'fill-color': [
291         'interpolate',
292         ['linear'],
293         ['get', MapCrimeType[crimeType]],
294         0, colors[0],
295         colorStops[1], colors[1],
296         colorStops[2], colors[2],
297         colorStops[3], colors[3],
298         colorStops[4], colors[4],
299         colorStops[5], colors[5],
300         colorStops[6], colors[6],
301         colorStops[7], colors[7],
302         colorStops[8], colors[8]
303       ],
304       'fill-opacity': 1
305     }
306   });
307 }
```



Ln 306, Col 8 Spaces: 2 UTF-8 LF Markdown

Visualizações

```
307
308 map.addSource('limites_dos_distritos', {
309   'type': 'geojson',
310   'data': layer_limite_distritos
311 });
312
313 map.addLayer({
314   'id': 'limites_dos_distritos',
315   'type': 'line',
316   'source': 'limites_dos_distritos',
317   'layout': {},
318   'paint': {
319     'line-color': '#000000'
320   }
321 });
322
```

<https://ide.emplasa.sp.gov.br/geoserver/ows?>

[service=WMS&version=1.1.0&request=GetFeature&typeName=emplasa:LIMITES_EMPLASA_DISTrito_UT_MSP&styles=&tilled=true&srsName=EPSG:4326&format=application/geojson&transparent=true](https://ide.emplasa.sp.gov.br/geoserver/ows?service=WMS&version=1.1.0&request=GetFeature&typeName=emplasa:LIMITES_EMPLASA_DISTrito_UT_MSP&styles=&tilled=true&srsName=EPSG:4326&format=application/geojson&transparent=true)

Visualizações

```
code > src > data > {} limite_dos_distritos.geojson > ...
```

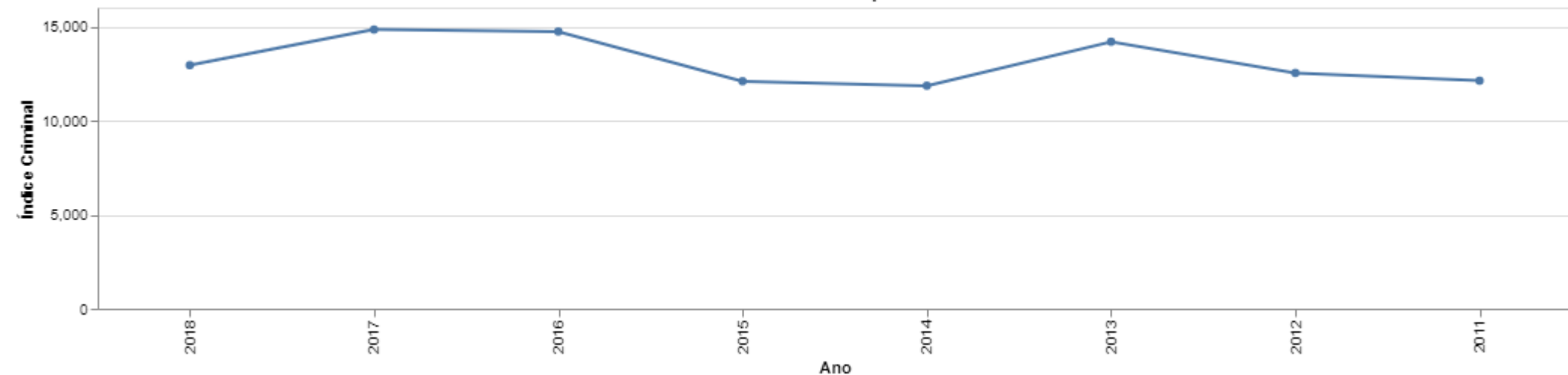
```
1  {"type":"FeatureCollection","totalFeatures":"unknown","features":[{"type":"Feature",
  "id":"LIMITES_EMPLASA_DISTRITO UIT MSP.1","geometry":{"type":"MultiPolygon",
  "coordinates":[[[-46.650216364621954,-23.5928369338366],[-46.65021630168885,-23.
  592836122140017],[-46.65020187673034,-23.592650072089512],[-46.650195369578434,-23.
  592607699236908],[-46.650170559338946,-23.592551253137906],[-46.65012914769444,-23.
  592483605369104],[-46.65005664350705,-23.592421254574102],[-46.649951250872135,-23.
  592369457312508],[-46.64984065006414,-23.59234778591901],[-46.64973292176294,-23.
  592352090947394],[-46.649619137013644,-23.592381383628297],[-46.649508511851245,-23.
  592404773028008],[-46.64941867400715,-23.59241367382591],[-46.64932897112214,-23.
  5924176968225],[-46.64895653882324,-23.592418039740302],[-46.64873059714875,-23.
  59231686207311],[-46.64860374256515,-23.592217702472198],[-46.648730868317145,-23.
  591860381684985],[-46.648997894874945,-23.591156711369365],[-46.649055536501436,-23.
  591017846683155],[-46.649114700378334,-23.590875353727263],[-46.64935265365935,-23.
  59025254115574],[-46.64953518672055,-23.59023624450174],[-46.649514747343346,-23.
  590137490407944],[-46.649398191042735,-23.590012475012642],[-46.64820957509484,-23.
  5887375755742],[-46.64812714385073,-23.58865681710399],[-46.64813109181203,-23.
  588549315551894],[-46.648307548030736,-23.58754490757426],[-46.64833838971733,-23.
  58739413673806],[-46.64815450219184,-23.586972299682852],[-46.64802761362384,-23.
  586681698872734],[-46.64781920182194,-23.58620241834793],[-46.64778283083505,-23.
  58616153605642],[-46.64771171422854,-23.58655083016003],[-46.64767640224634,-23.
  585664737361107],[-46.64767640224634,-23.585664737361107],[-46.64767640224634,-23.
  584729766726575],[-46.64775140224634,-23.584729766726575],[-46.64775140224634,-23.
  584436922819783],[-46.64792902985903,-23.58427203535067],[-46.64812928940853,-23.
  58398907061626],[-46.64816379505824,-23.583966941658364],[-46.64824781605394,-23.
  583853623818158],[-46.648462085299634,-23.58356283435455],[-46.64856634377583,-23.
  58344605071775],[-46.64865398236903,-23.58334786507113],[-46.64881540056104,-23.
  583124459762132],[-46.64906541282664,-23.582785339666927],[-46.649155651940944,-23.
  582652269485823],[-46.64921403318213,-23.582588528409524],[-46.64926796898343,-23.
  58252971083211],[-46.64932469103774,-23.582035335834703],[-46.64934431055604,-23.
  581864817457298],[-46.64936154684524,-23.581715274505797],[-46.64937996819523,-23.
  581605677723882],[-46.64939838952053,-23.581496080939985],[-46.64944536877834,-23.
  5814529594693881],[-46.649650672328543,-23.5812644864566821],[-46.64966703324164,-23.
  5812644864566821],[-46.64966703324164,-23.5812644864566821]]]]]}
```

Tokenization is skipped for long lines for performance reasons. This can be configured via
editor.maxTokenizationLineLength.

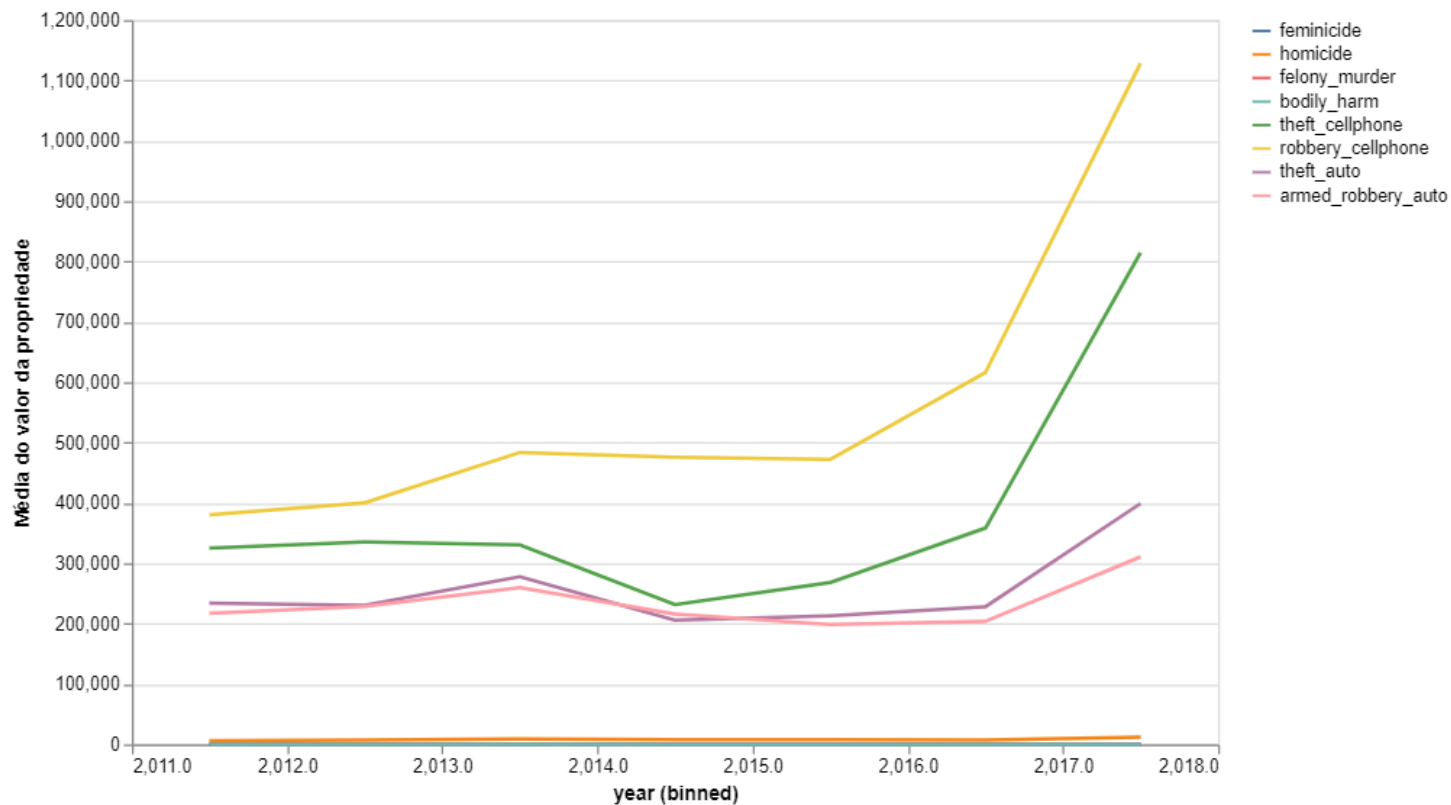
Ln 1, Col 1 Spaces: 4 UTF-8 LF JSON

Visualizações

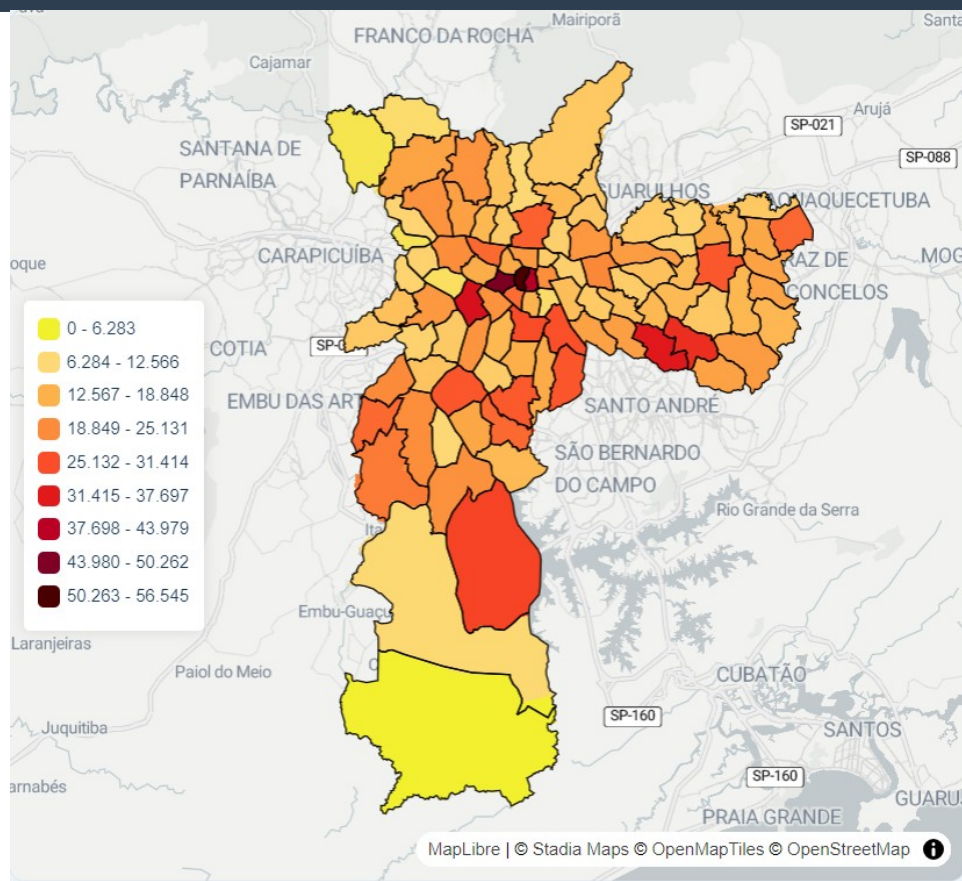
Índice Criminal por ano



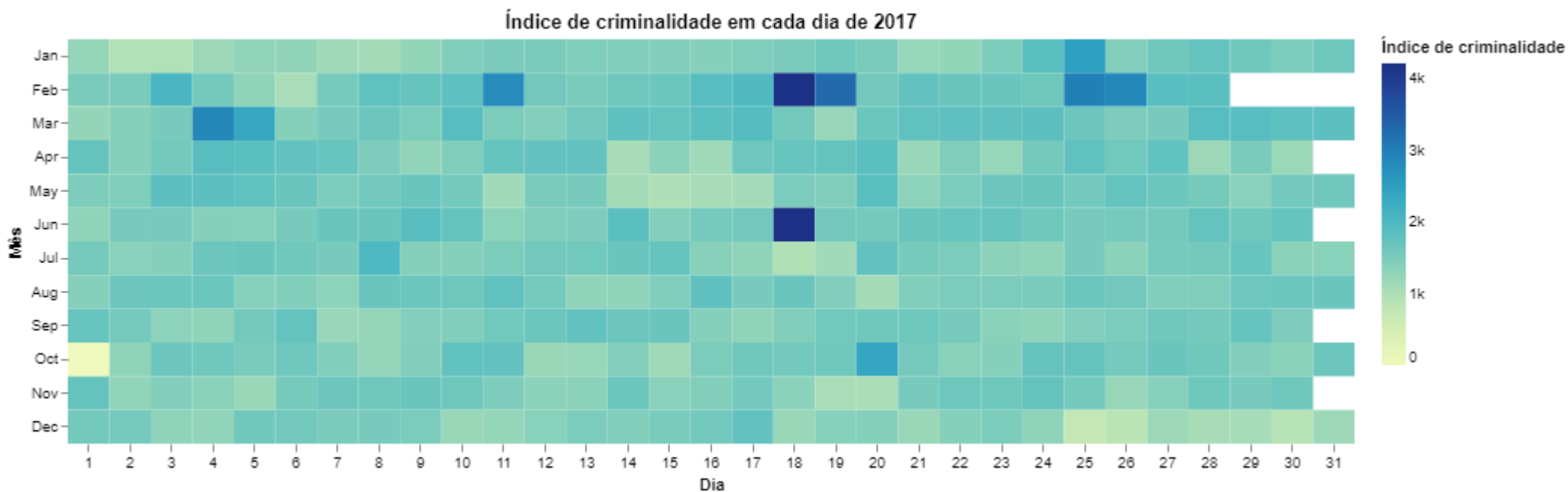
Visualizações



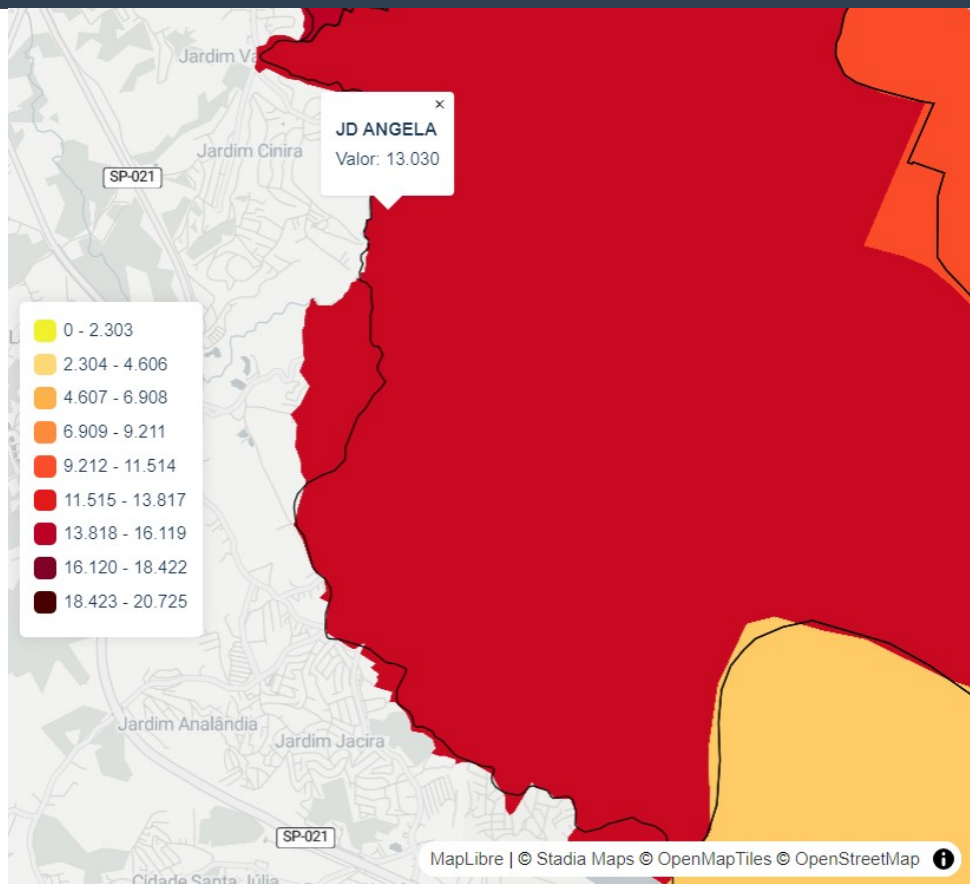
Visualizações



Visualizações



Visualizações



Conclusão

- Lucas Natan: lucasnatan@id.uff.br
- Ruan Pablo: ruanpablo@id.uff.br
- Italo Portinho: italons@gmail.com, italoleite@id.uff.br

Referências

- [1] Raissa Barcellos, Flavia Bernardini, José Viterbo, Towards defining data interpretability in open data portals: Challenges and research opportunities, *Information Systems*, Volume 106, 2022, 101961, ISSN 0306-4379, <https://doi.org/10.1016/j.is.2021.101961>.
- [2] (2008, November 27). *Decreto 666/2008*. Planalto.gov.br. Retrieved July 4, 2024, from https://www.planalto.gov.br/ccivil_03/_ato2007-2010/2008/decreto/d6666.htm
- [3] (2008, November 27). *Catálogo de Metadados INDE*. INDE. Retrieved July 4, 2024, from <https://inde.gov.br/CatalogoMetadados>
- [4] (2008, November 27). *Catálogo de Geosserviços*. INDE. Retrieved July 4, 2024, from <https://inde.gov.br/CatalogoGeoservicos>
- [5] Cunha Sá, B., Muller, G., Banni, M., Santos, W., Lage, M., Rosseti, I., Frota, Y. and de Oliveira, D. 2022. PolRoute-DS: a Crime Dataset for Optimization-based Police Patrol Routing. *Journal of Information and Data Management*. 13, 1 (Aug. 2022). DOI:<https://doi.org/10.5753/jidm.2022.2355>.
- [6] Özsu, M. Tamer, Valduriez, P. (2011) “Principles of Distributed Database Systems - Third Edition”, p.81-98. Springer
- [7] Apache Hadoop. (2024, June 28). In *Wikipedia*. https://en.wikipedia.org/wiki/Apache_Hadoop
- [8] Apache Hive. (2024, June 28). In *Wikipedia*. https://en.wikipedia.org/wiki/Apache_Hive

FIM

- **Muito Obrigado!!!**