



**Disciplina:** Visualização de Dados    **Entrega:** 13.12.2021

## **Trabalho Final**

**Professor(a):** Asla Medeiros de Sá    **Aluno:** Gilberto Pereira Ramos



## Sumário

<b>1. Introdução</b>	4
1.1 Definição das Dimensões:	4
1.2 Tabela – Definição dos indicadores	6
1.3 Metodologia da criação dos indicadores:	9
<b>2. Definições de termos</b>	11
<b>3. Fonte de dados e exemplos de Datasets</b>	13
3.1 Fonte dos dados	13
3.2 Exemplos de Datasets utilizados no projeto:	13
<b>4. Análise exploratória dos dados</b>	16
<b>5. Produto Final : Web App - Visualização com Altair, Streamlit e Geopandas</b>	19
5.1 Overview do Web App Streamlit	19
5.2 Renderização e Visualização de Mapas Choropleth no Streamlit	24
5.2.1 Exemplo da estrutura de um file TopoJson:	24
5.2.2 Visualização do file TopoJson 'br-states.json'	25
5.2.3 Resultado do Mapeamento para Web App Streamlit:	26
<b>6. Conclusão</b>	29
<b>7. Referências</b>	30
<b>8. Anexo (guia passo-a-passo p/ executar o app web via streamlit)</b>	30

# Índice de Progresso Social da cidade do Rio de Janeiro: Uma Análise exploratória de dados com visualização interativa.

## Resumo

Trabalho final da disciplina de Visualização de Dados do curso de Mestrado em Modelagem Matemática da Informação da Emap-FGV. Esse trabalho tem como objetivo aplicar técnicas de visualização da informação desenvolvidas ao longo do curso de terceiro trimestre de 2021. Para execução desse trabalho foi implementado código Python versão 3.8 usando as bibliotecas Altair, Streamlit e Geopandas. Foi utilizado também o padrão TopoJson para arquivos de topologia geográfica para permitir explorar mapas interativos do tipo ChoroPleth. O relatório está dividido da seguinte forma: Na seção 1 são apresentadas algumas motivações que levaram a escolha do tema bem como a definição das dimensões que serão utilizadas no trabalho; A seção 2 apresenta a definição dos termos que são utilizados no presente documento. Na seção 3 são apresentados alguns exemplos dos datasets utilizados bem como a fonte dos mesmos. Na seção 4 é apresentada a exploração de dados como objetivo de analisar a coerência semântica e qualidade dos dados. A seção 5 apresenta retratos da visualização interativa usando Altair e Streamlit e também a análise exploratória da relação entre algumas dimensões. Ao final do relatório é apresentada a conclusão, as referências e uma seção de anexo com o passo a passo para execução do web app Streamlit, que é o produto final desse projeto.

## 1. Introdução

O objetivo desse trabalho é a análise exploratória e reprodução de uma visualização interativa usando a biblioteca Altair do Python e publicação do app web usando a biblioteca Streamlit, bem como explorar o recurso de mapas Choropleth com Geopandas. O trabalho foi motivado pela possibilidade de descoberta de relações relevantes entre as dimensões encontradas na base ‘data-rio’, e com isso gerar insights e visualizações acerca dos problemas sociais que circundam nossa comunidade local no Rio de Janeiro. O trabalho discorre efetuando uma análise exploratória da base de dados ‘data.rio’ da prefeitura do rio de Janeiro. Essa análise é baseada na extração de algumas dimensões de questões sociais apontadas como indicadores da região metropolitana do Rio de Janeiro. O portal da prefeitura do Rio de Janeiro disponibiliza esses indicadores como resultado de pesquisas levantadas nos anos de 2016, 2018 e 2020. Essas features de questões sociais dizem respeito as dimensões analisadas para se chegar ao Indicador de Progresso Social (IPS) da região metropolitana do Rio de Janeiro. Não é objetivo desse trabalho analisar todas as features ou dimensões, mas sim utilizar algumas dimensões e propor uma visualização interativa ao final do trabalho usando a biblioteca Altair e geração do produto final, um web app desenvolvido com Streamlit.

### 1.1 Definição das Dimensões:

**IPS - ÍNDICE DE PROGRESSO SOCIAL DA CIDADE DO RIO DE JANEIRO:**

- O IPS é formado por 36 indicadores que estão dispostos em 12 componentes, igualmente distribuídos dentro de 3 dimensões. Os indicadores foram construídos com dados administrativos municipais, estaduais e federais e do Censo Demográfico. Ver esquema abaixo.

**ÍNDICE DE PROGRESSO SOCIAL**

**Necessidades Humanas Básicas**

**Nutrição e Cuidados Médicos Básicos**

Mortalidade na Infância  
Baixo Peso ao Nascer  
Mortalidade Materna  
Internações Infantis por crise respiratória aguda

**Água e Saneamento**

Acesso à Água Canalizada  
Acesso a Esgotamento Sanitário  
Acesso a Banheiro

**Moradia**

Pessoas Vivendo em Favelas Não-Urbanizadas  
Acesso a Energia Elétrica  
Adensamento Habitacional Excessivo

**Segurança Pessoal**

Taxas de Homicídios  
Roubos de Rua

**Fundamentos do Bem-Estar**

**Acesso ao Conhecimento Básico**

Alfabetização  
Qualidade do Ensino Fundamental, anos iniciais  
Qualidade do Ensino Fundamental, anos finais  
Abandono Escolar No Ensino Médio

**Acesso à Informação**

Acesso a Telefone Celular ou Fixo  
Acesso à internet

**Saúde e Bem-Estar**

Mortalidade por Doenças Crônicas  
Incidência de Dengue  
Mortalidade por Tuberculose e HIV

**Qualidade do Meio Ambiente**

Coleta Seletiva de Lixo  
Degradação de Áreas Verdes

**Oportunidades**

**Direitos Individuais**

Mobilidade Urbana  
Homicídios por Ação Policial  
Tempo Médio de Deslocamento  
Participação Política

**Liberdades Individuais**

Gravidez na Adolescência  
Trabalho Infantil  
Índice de Acesso à Cultura

**Tolerância e Inclusão**

Violência Contra a Mulher  
Homicídios de Jovens Negros  
Vulnerabilidade Famílias

**Acesso à Educação Superior**

Pessoas com Ensino Superior  
Negros e Indígenas com Ensino Superior  
Frequência ao Ensino Superior

## 1.2 Tabela – Definição dos indicadores

INDICADORES	DEFINIÇÃO	FONTE	ANO DOS DADOS		
			IPS 2016	IPS 2018	IPS 2020
<b>1. Mortalidade Infantil:</b>	Óbitos de crianças de 0 a 5 anos por 1.000 nascidos vivos. O número de óbitos foi obtido por meio do Sistema de Informações sobre Mortalidade do Sistema de Informações em Saúde da Secretaria Municipal de Saúde do Município do Rio de Janeiro (SMS-RJ). O número de nascidos vivos foi obtido por meio do Sistema de Informação de Nascidos Vivos, da mesma base de dados da SMS-RJ.	DataSus	2014	2017	2019
<b>2. Baixo Peso ao Nascer:</b>	Proporção de nascidos vivos que, na primeira pesagem (preferencialmente, durante a primeira hora de vida), tinham peso inferior a 2.500 quilogramas. O número de nascimentos abaixo do peso e o total de nascidos vivos foram obtidos pelo Sistema de Informação de Nascidos Vivos do Sistema de informação em Saúde da SMS-RJ.	DataSus	2014	2017	2019
<b>3. Mortalidade Materna:</b>	Óbitos de mulheres ocorridos até 42 dias após o término da gravidez, atribuídos a causas ligadas a gravidez, parto, e puerpério, em relação ao total de nascidos vivos. O número de óbitos de mulheres por essas causas foi obtido pelo Sistema de Informações sobre Mortalidade do Sistema de Informações em Saúde da SMS-RJ. O número de nascidos vivos foi obtido pelo Sistema de Informação de Nascidos Vivos, da mesma base de dados da SMS-RJ.	DataSus	2014	2017	2019
<b>4. Internações Infantis por Crise Respiratória:</b>	Taxa de internações de crianças por infecção respiratória aguda (pneumonias, sinusites, faringoamigdalites e resfriados comuns). O número de internações foi em 2016 fornecido pela Secretaria Municipal de Saúde do Rio; e em 2018, foi calculado pela Rio como Vamos, com base na mesma fonte. O número de crianças, para 2016, foi obtido pelo Censo Demográfico 2010.	SMS	2014	2014	2019
<b>5. Acesso à Água Canalizada:</b>	Proporção de domicílios cuja forma de abastecimento de água é a rede geral de distribuição. A forma de abastecimento de água e o número total de domicílios foram obtidos dos resultados do Censo Demográfico de 2010 – Questionário Básico.	Censo Demográfico	2010	2010	2010
<b>6. Acesso a Esgotamento Sanitário:</b>	Proporção de domicílios cujo esgoto do banheiro ou sanitário é lançado na rede geral de esgoto ou pluvial. O local de lançamento do esgoto e o número total de domicílios foram obtidos dos resultados do Censo Demográfico 2010 – Questionário Básico.	Censo Demográfico	2010	2010	2010
<b>7. Acesso a Banheiro:</b>	Proporção de domicílios com pelo menos um banheiro. O número de banheiros por domicílio, assim como número total de domicílios, foram obtidos dos resultados do Censo Demográfico 2010 – Questionário Básico.	Censo Demográfico	2010	2010	2010
<b>8. Pessoas Vivendo em Favelas NãoUrbanizadas:</b>	proporção da população total da RA vivendo em favelas não urbanizadas. Os dados sobre a população total das regiões e o número de habitantes das favelas vêm do Censo Demográfico 2010. A informação sobre quais favelas foram urbanizadas vem do Sistema de Assentamentos de Baixa Renda (SABREN) do Instituto Pereira Passos.	IPP	2013	2017	2019
<b>9. Acesso à Energia Elétrica:</b>	proporção de domicílios nos quais há energia elétrica de companhia distribuidora. A existência de energia elétrica formal e o número total de domicílios foram obtidos dos resultados do Censo Demográfico 2010 – Questionário Básico.	Censo Demográfico	2010	2010	2010
<b>10. Adensamento Habitacional Excessivo:</b>	proporção de domicílios com mais de três moradores por dormitório. O número de domicílios nessas condições foi obtido por meio dos resultados do Censo Demográfico 2010 – Questionário da Amostra, por Área de Ponderação, e depois reorganizado por RA.	Censo Demográfico	2010	2010	2010

<b>11. Taxa de Homicídios:</b>	número de óbitos por homicídio, latrocínio e lesão corporal seguida de morte por 100.000 habitantes. Número de óbitos foi obtido por meio dos boletins de divulgação mensal do Instituto de Segurança Pública do Estado do Rio de Janeiro (ISP), enquanto o número de habitantes foi obtido dos resultados do Censo Demográfico 2010.	ISP	2015	2017	2019
<b>12. Roubos de Rua:</b>	taxas de ocorrências de roubo a transeunte, roubo de aparelho celular e roubo em coletivo por 1.000 habitantes. Número de ocorrências foi obtido por meio dos boletins de divulgação mensal do ISP, enquanto o número de habitantes foi obtido dos resultados do Censo Demográfico 2010.	ISP	2015	2017	2019
<b>13. Alfabetização:</b>	Proporção de crianças alfabetizadas até o terceiro ano do Ensino Fundamental em escolas públicas. Crianças alfabetizadas são as que têm nota acima do nível 2 na Avaliação Nacional de Alfabetização (ANA) do Instituto Nacional de Estudos e Pesquisas Educacionais Anísio Teixeira (INEP), em relação ao número total de crianças da rede pública de ensino do Município do Rio de Janeiro que participaram da ANA.	ANA	2014	2016	2016
<b>14. Qualidade do Ensino Fundamental, anos iniciais:</b>	Nota média no Índice de Desenvolvimento da Educação Básica (IDEB) do INEP para os anos iniciais. A nota por RA foi agregada pela média das escolas públicas e ponderada pelo número de alunos.	IDEB	2013	2017	2019
<b>15. Qualidade do Ensino Fundamental, anos finais:</b>	Nota média no IDEB para os anos finais. A nota por RA agregada pela média das escolas públicas e ponderada pelo número de alunos de cada escola.	IDEB	2013	2017	2019
<b>16. Abandono Escolar no Ensino Médio:</b>	Proporção dos alunos do Ensino Médio em instituições públicas de ensino na cidade do Rio de Janeiro que abandonaram a escola. O número de alunos que abandonaram a escola e o número total de alunos foram obtidos por meio do Censo da Educação Básica do INEP	Censo Escolar INEP	2013	2016	2018
<b>17. Acesso a Telefone Celular ou Fixo:</b>	Proporção de domicílios com telefone celular ou fixo. O Número de domicílios que têm telefone, celular ou fixo, e o número total de domicílios foram obtidos dos resultados do Censo Demográfico 2010 – Questionário da Amostra, por Área de Ponderação, e depois reorganizados por RA.	Censo Demográfico	2010	2010	2010
<b>18. Acesso à internet:</b>	Proporção de domicílios com acesso à internet. Número de domicílios com microcomputador com acesso à internet, e o número total de domicílios foram obtidos dos resultados do Censo Demográfico 2010 – Questionário da Amostra, por Área de Ponderação, e depois reorganizados por RA.	Censo Demográfico	2010	2010	2010
<b>19. Mortalidade por Doenças Crônicas:</b>	Taxa de mortalidade por doenças crônicas não transmissíveis antes dos 70 anos por 100.000 habitantes. O número de mortes foi obtido por meio do Sistema de Informações sobre Mortalidade do Sistema de Informações em Saúde da SMS-RJ, enquanto o número de habitantes com menos de 70 anos foi obtido nos resultados do Censo Demográfico 2010.	DataSus	2015	2017	2019
<b>20. Incidência de Dengue:</b>	Taxa de incidência de dengue por 100.000 habitantes. O número de casos notificados de dengue foi obtido no Sistema Nacional de Agravos Notificáveis do Sistema de Informações em Saúde da SMS-RJ, enquanto o número de habitantes foi obtido dos resultados do Censo Demográfico 2010.	DataSus	2014	2017	2019
<b>21. Mortalidade por Tuberculose e HIV:</b>	Taxa de mortalidade por tuberculose ou causas associadas ao HIV por 100.000 habitantes. O número de óbitos por tuberculose ou causas associadas ao HIV foi obtido por meio do Sistema de Informações sobre Mortalidade do Sistema de Informações em Saúde da SMS-RJ, enquanto o número de habitantes foi obtido dos resultados do Censo Demográfico 2010.	DataSus	2014	2017	2019

<b>22. Coleta Seletiva de Lixo:</b>	Proporção do lixo potencialmente reciclável recolhido pela Coleta Seletiva Domiciliar. Os dados de potencial reciclado e resíduo sólido coletado foram fornecidos para a elaboração do IPS pela Companhia Municipal de Limpeza Urbana (COMLURB).	Comlurb	2015	2017	2019
<b>23. Degradação de Áreas Verdes:</b>	Diferença entre o tamanho da área coberta por vegetação arbóreo-arbustiva e gramíneo-lenhosa no ano de referência do IPS e o da área de mesmo tipo em 2013. O tamanho das áreas cobertas por vegetação no período foi calculado com base no Mapa de Uso do Solo do Instituto Pereira Passos	IPP	2015	2017	2019
<b>24. Mobilidade Urbana:</b>	Proporção da população que vive em um raio de 1 km de uma estação de acesso a um meio de transporte de massa (metrô, trem, BRT e VLT). O índice foi calculado pelo Instituto de Políticas de Transporte e Desenvolvimento (ITDP) com base nas estimativas populacionais do Censo Demográfico de 2010 e da geolocalização das estações dos meios de transporte de massa do Instituto Pereira Passos.	ITDP	2014	2017	2019
<b>25. Homicídios por Ação Policial:</b>	Taxa de homicídios por intervenção policial por 100.000 habitantes. O número de óbitos foi obtido por meio dos boletins de divulgação mensal do Instituto de Segurança Pública do Estado do Rio de Janeiro (ISP), enquanto o número de habitantes foi obtido nos resultados do Censo Demográfico 2010.	ISP	2015	2017	2019
<b>26. Tempo Médio de Deslocamento:</b>	Tempo médio, em minutos, gasto no deslocamento entre residência e atividade principal. A estimativa foi feita com base nos dados do Plano Diretor de Transporte Urbano (PDTU) de 2012 do governo do estado do Rio de Janeiro.	PDTU	2012	2012	2012
<b>27. Participação Política:</b>	Taxa de participação nas últimas eleições federais e municipais com dados disponíveis. O número de votantes, por zona eleitoral, nas duas eleições e o número de eleitores aptos para votar foram fornecidos pelo Tribunal Superior Eleitoral (TSE).	TSE	2014/2012	2014/2016	2016/2018
<b>28. Gravidez na Adolescência:</b>	Proporção de nascidos vivos com mães de até 19 anos. O número total de nascidos vivos e o número de nascidos vivos com mãe de até 19 anos foram obtidos por meio do Sistema de Informação de Nascidos Vivos do Sistema de informação em Saúde da SMS-RJ.	DataSus	2014	2017	2019
<b>29. Trabalho Infantil:</b>	Crianças ou adolescentes exercendo atividades laborais por 100.000 habitantes. Os dados foram obtidos por meio do Cadastro Único para programas sociais do governo federal, geolocalizado para o cálculo do IPS pela Secretaria Municipal do Desenvolvimento Social do município do Rio de Janeiro (SMDS). O número de habitantes foi obtido dos resultados do Censo Demográfico 2010.	CadÚnico	2015	2018	2018
<b>30. Índice de Acesso à Cultura:</b>	Índice calculado pelo IPP para a produção do IPS. O índice considera o número de equipamentos culturais municipais, o número de ações e eventos culturais da prefeitura, além da capacidade das salas privadas de cinema, todos em relação à população da Região Administrativa. As informações foram fornecidas pela Secretaria Municipal de Cultura do município do Rio de Janeiro.	SMC	2015	2017	2017
<b>31. Violência Contra a Mulher:</b>	Taxas de crimes de violência física e sexual contra mulheres por 100.000 habitantes. O número de ocorrências de crimes de violência física (homicídio doloso, tentativa de homicídio e lesão corporal dolosa) e violência sexual (estupro e tentativa de estupro) contra mulheres foi obtido por meio de dados trabalhados a partir da publicação Dossiê Mulher do ISP, enquanto o número de habitantes foi obtido dos resultados do Censo Demográfico 2010.	ISP	2014	2017	2019



<b>32. Homicídios de Jovens Negros:</b>	Taxa de homicídios de jovens (até 26 anos) pretos e pardos por 100.000 habitantes. O número de óbitos de jovens negros foi calculado para a construção do IPS pelo ISP, enquanto o número de habitantes foi obtido nos resultados do Censo Demográfico 2010.	ISP	2014	2017	2019
<b>33. Vulnerabilidade Familiar:</b>	Proporção de domicílios com mulher com ensino fundamental incompleto como responsável e filho(s) menor(es) de 10 anos ou economicamente inativo(s). O número de domicílios com essas características e número total de domicílios foram obtidos dos resultados do Censo Demográfico 2010 – Questionário da Amostra, por Área de Ponderação, e depois reorganizados por RA.	Censo Demográfico	2010	2010	2010
<b>34. Pessoas com Ensino Superior:</b>	Proporção da população de 25 anos ou mais que completou Ensino Superior. Os dados de escolaridade e tamanho da população de 25 anos ou mais foram obtidos dos resultados do Censo Demográfico 2010 – Questionário da Amostra, por Área de Ponderação, e depois reorganizados por RA.	Censo Demográfico	2010	2010	2010
<b>35. Negros e Indígenas com Ensino Superior:</b>	Proporção da população que se autoidentifica como preta, parda ou indígena com idade igual ou superior a 25 anos que completou o Ensino Superior. Os dados de escolaridade desse grupo populacional e seu tamanho foram obtidos dos resultados do Censo Demográfico 2010 – Questionário da Amostra, por Área de Ponderação, e depois reorganizados por RA.	Censo Demográfico	2010	2010	2010
<b>36. Frequência ao Ensino Superior:</b>	Proporção da população com idade entre 18 e 24 anos que frequenta ou concluiu o Ensino Superior. Os dados de escolaridade desse grupo populacional e seu tamanho foram obtidos dos resultados do Censo Demográfico 2010 – Questionário da Amostra, por Área de Ponderação, e depois reorganizados por RA.	Censo Demográfico	2010	2010	2010

### 1.3 Metodologia da criação dos indicadores:

**1. Vetorização da série de dados:** Uma vez que os indicadores nem sempre apontam para uma mesma direção, é necessária a inversão de alguns indicadores para que valores mais altos reflitam sempre mais progresso social. Ao todo, foram invertidos 14 indicadores.

- 1. Normalização da série de dados:** Transformação da série em uma distribuição normal padrão, subtraindo-se média da série ( $\mu$ ) e dividindo pelo desvio padrão, onde  $x_{ij}$  é o valor da observação da variável  $j$  para a comunidade  $i$ .

$$z_{ij} = \frac{x_{ij} - \mu_j}{\sigma_j}$$

- 2. Análise de consistência interna:** Feita entre indicadores padronizados dentro de cada componente, por meio da medida de covariância de Alpha de Cronbach, buscando-se coeficientes de Alpha acima de 0.7

3. **Distribuição de pesos:** Calculou-se um peso  $w_j$  para cada vetor (variável)  $j$  pertencente a determinado componente, por meio de análise fatorial das variáveis de cada componente, utilizando-se somente a primeira carga fatorial atribuída a cada vetor.
5. **Teste de Kaiser-Meyer-Olkin:** A robustez da análise fatorial é avaliada pelo teste Kaiser-Meyer-Olkin (KMO) de adequação amostral, buscando-se resultados acima de 0.5.
6. **Padronização dos pesos da AF:** Os pesos  $w_j$  obtidos na análise fatorial para cada variável  $j$  são padronizados por componente de modo que a soma dos pesos seja igual a 1.
7. **Nota dos componentes:** A nota de cada observação em cada componente é dada pela multiplicação de  $z_{ij}$  pelo peso  $w_j$  atribuído a cada variável por meio da análise fatorial, de modo que a nota do componente é dada por:

$$c = \sum w_i z_i$$

onde  $C$  é a nota do componente,  $i$  é o número de observações por variável e  $n = 31$ , número de Regiões Administrativas.

8. **Nota das dimensões:** A nota de cada dimensão é calculada pela média simples das notas de cada componente, onde o número de componentes é sempre igual a 4, pois são 4 componentes por dimensão  $D$ .

$$d = \frac{1}{n} \sum_{i=1}^n c$$

9. **Nota do IPS:** A nota do IPS é calculada pela média simples das notas das dimensões, onde o

$$IPS = \frac{1}{3} \sum_{d=1}^3 D_d$$

OBS: Os testes de Alpha de Cronbach e KMO são realizados em componentes com três ou mais variáveis. Por esta razão os, testes foram realizados nos componentes de Segurança Pessoal, Acesso à Informação e Comunicação e Sustentabilidade dos Ecossistemas, sendo o peso das variáveis igual a 0.5. Os componentes de Direitos Individuais e Sustentabilidade e Ecossistemas foram os únicos com Alpha de Cronbach inferior a 0.7

## 2. Definições de termos

- **Python** é uma linguagem de programação de alto nível, interpretada e multi-paradigma. Lançada por Guido Van Rossum em 1991 é, atualmente, mantida de forma colaborativa e aberta. Para mais informações, consulte a página oficial da linguagem Python: <https://www.python.org/> e a comunidade Python Brasil: <http://wiki.python.org.br/> Fonte: <http://www.ufrgs.br>
- **Altair** é uma biblioteca de visualização estatística declarativa para Python, baseada em Vega e Vega-Lite. Fonte: <https://altair-viz.github.io/>
- **Vega** é uma gramática de visualização, uma linguagem declarativa para criar, salvar e compartilhar designs de visualização interativa. Com o Vega, você pode descrever a aparência visual e o comportamento interativo de uma visualização em formato JSON e gerar visualizações baseadas na web usando Canvas ou SVG. Fonte: <https://vega.github.io/vega/>
- **Vega-Lite** é uma gramática de alto nível de gráficos interativos. Ele fornece uma sintaxe JSON concisa e declarativa para criar uma gama expressiva de visualizações para análise e apresentação de dados. Fonte: <http://vega.github.io/vega-lite/>
- **Streamlit** é um framework de código aberto, criado justamente para ajudar cientistas de dados a colocarem em produção seus projetos sem a necessidade de conhecer ferramentas de front-end ou de deploy de aplicações. Por meio desse framework é possível transformar um projeto de ciência de dados e machine learning em uma aplicação interativa. Para essa aplicação é gerada uma URL pública que, ao ser compartilhada, permite que qualquer pessoa consiga acessar e usufruir sem necessariamente ter que conhecer o código que está por trás. Fonte: <http://www.alura.com.br>
- **GeoPandas** é uma biblioteca usada para analisar e trabalhar com dados geoespaciais em Python. Em uma de suas atualizações recentes, a biblioteca adicionou métodos e utilitários para oferecer suporte a visualizações interativas. Fonte: <https://towardsdatascience.com/>

- **JSON**, que significa JavaScript Object Notation, é uma formatação utilizada para estruturar dados em formato de texto e transmiti-los de um sistema para outro, como em aplicações cliente-servidor ou em aplicativos móveis. Uma maneira de utilização é por meio de requisições AJAX, em que a aplicação recupera os dados armazenados no servidor de hospedagem sem a necessidade de recarregar a página. A especificação JSON surgiu por volta do ano 2000, mas só passou a fazer parte da linguagem JavaScript após o lançamento da versão 5 do ECMAScript. Atualmente, esse formato é suportado por diversos tipos de linguagem de programação, além de ser uma alternativa mais leve que o modelo XML. Fonte: <https://rockcontent.com/br/blog/json/>
- **TopoJSON** é uma extensão do GeoJSON que codifica a topologia. Em vez de representar geometrias discretamente, as geometrias nos arquivos TopoJSON são costuradas a partir de segmentos de linha compartilhados chamados arcos. Fonte: <https://github.com/topojson/topojson>
- **Anaconda** é uma distribuição das linguagens de programação Python e R para computação científica (ciência de dados, aplicativos de aprendizado de máquina, processamento de dados em grande escala, análise preditiva, etc.), que visa simplificar o gerenciamento e implantação de pacotes. A distribuição inclui pacotes de ciência de dados adequados para Windows, Linux e macOS. É desenvolvido e mantido pela Anaconda, Inc., fundada por Peter Wang e Travis Oliphant em 2012. Como um produto Anaconda, Inc., também é conhecido como Anaconda Distribution ou Anaconda Individual Edition.  
Fonte: [https://en.wikipedia.org/wiki/Anaconda\\_\(Python\\_distribution\)](https://en.wikipedia.org/wiki/Anaconda_(Python_distribution))

### 3. Fonte de dados e exemplos de Datasets

#### 3.1 Fonte dos dados

Os dados aqui utilizados são públicos e estão disponibilizados nos links informados abaixo.

<https://www.rio.rj.gov.br/web/transparencia/dados-abertos>

<https://www.data.rio/>

<http://www.ssp.df.gov.br/> (Secretaria de Segurança Pública dos Estados e do DF.)

[https://pt.wikipedia.org/wiki/Lista\\_de\\_unidades\\_federativas\\_do\\_Brasil\\_por\\_IDH](https://pt.wikipedia.org/wiki/Lista_de_unidades_federativas_do_Brasil_por_IDH)

<http://hdr.undp.org/sites/default/files/hdr2020.pdf>

#### 3.2 Exemplos de Datasets utilizados no projeto:

**BASE\_IPS\_RJ\_2020.CSV**

	regiao	ano	mort_inf	baixo_peso	mort_mat	Intern_inf_CRA	ace_agua	ace_esgot_sanit	ace_ban	pop_fav_ao_urb	ace_energia_a
0	RIO DE JANEIRO	2020	2.79	0.09	80.63	8.96	0.98	0.90	0.99	0.14	0.94
1	PORTUARIA	2020	4.46	0.10	0.00	6.16	1.00	0.94	0.94	0.04	0.91
2	CENTRO	2020	3.16	0.10	90.99	14.30	1.00	0.92	0.92	0.00	0.97
3	RIO COMPRIDO	2020	2.36	0.10	31.55	5.07	0.99	0.93	0.96	0.09	0.91
4	BOTAFOGO	2020	1.05	0.08	33.11	2.73	1.00	0.99	0.99	0.02	0.99
5	COPACABANA	2020	1.66	0.08	27.98	2.73	1.00	0.99	0.99	0.02	0.97
6	LAGOA	2020	3.16	0.10	0.00	8.90	1.00	0.99	1.00	0.03	0.99
7	SÃO CRISTOVÃO	2020	3.26	0.09	90.42	8.48	0.99	0.93	0.97	0.05	0.89
8	TIJUCA	2020	1.81	0.09	0.00	5.50	0.97	0.96	1.00	0.01	0.97

**Tx Femicídio Brasil 2020 (brasil\_estados\_ok.csv)**

	estado	value
2	Amapá	5.0
26	Tocantins	9.0
22	Roraima	9.0
0	Acre	11.0
19	Rio Grande do Norte	13.0
25	Sergipe	14.0
21	Rondônia	14.0
3	Amazonas	16.0
6	Distrito Federal	17.0
7	Espírito Santo	26.0

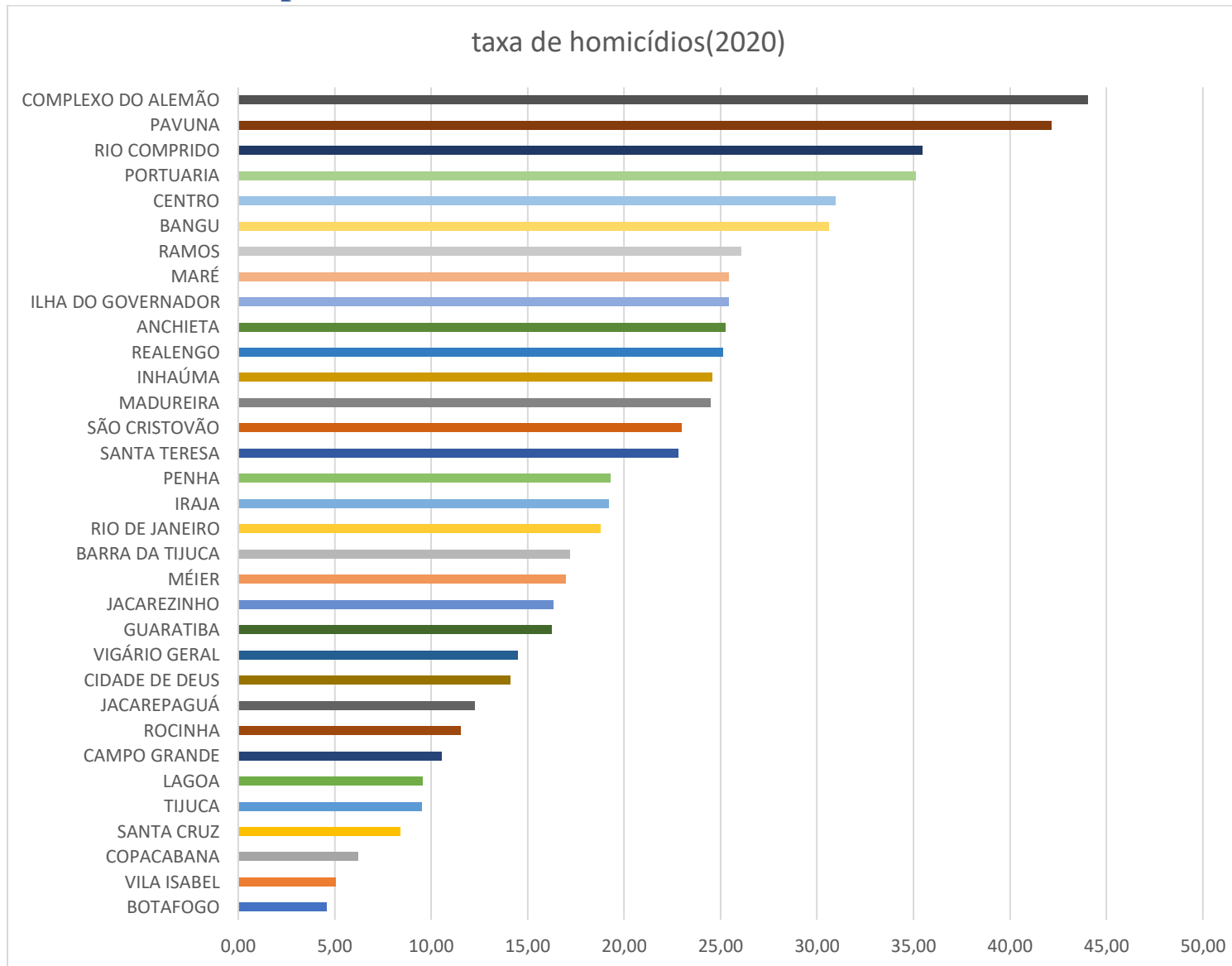
**IDH BRASIL 2020 (IDH\_Brasil\_2020\_ok.csv)**

	estado	value
1	Alagoas	0.683
9	Maranhão	0.687
17	Piauí	0.697
13	Pará	0.698
25	Sergipe	0.702
4	Bahia	0.714
0	Acre	0.719
14	Paraíba	0.722
21	Rondônia	0.725

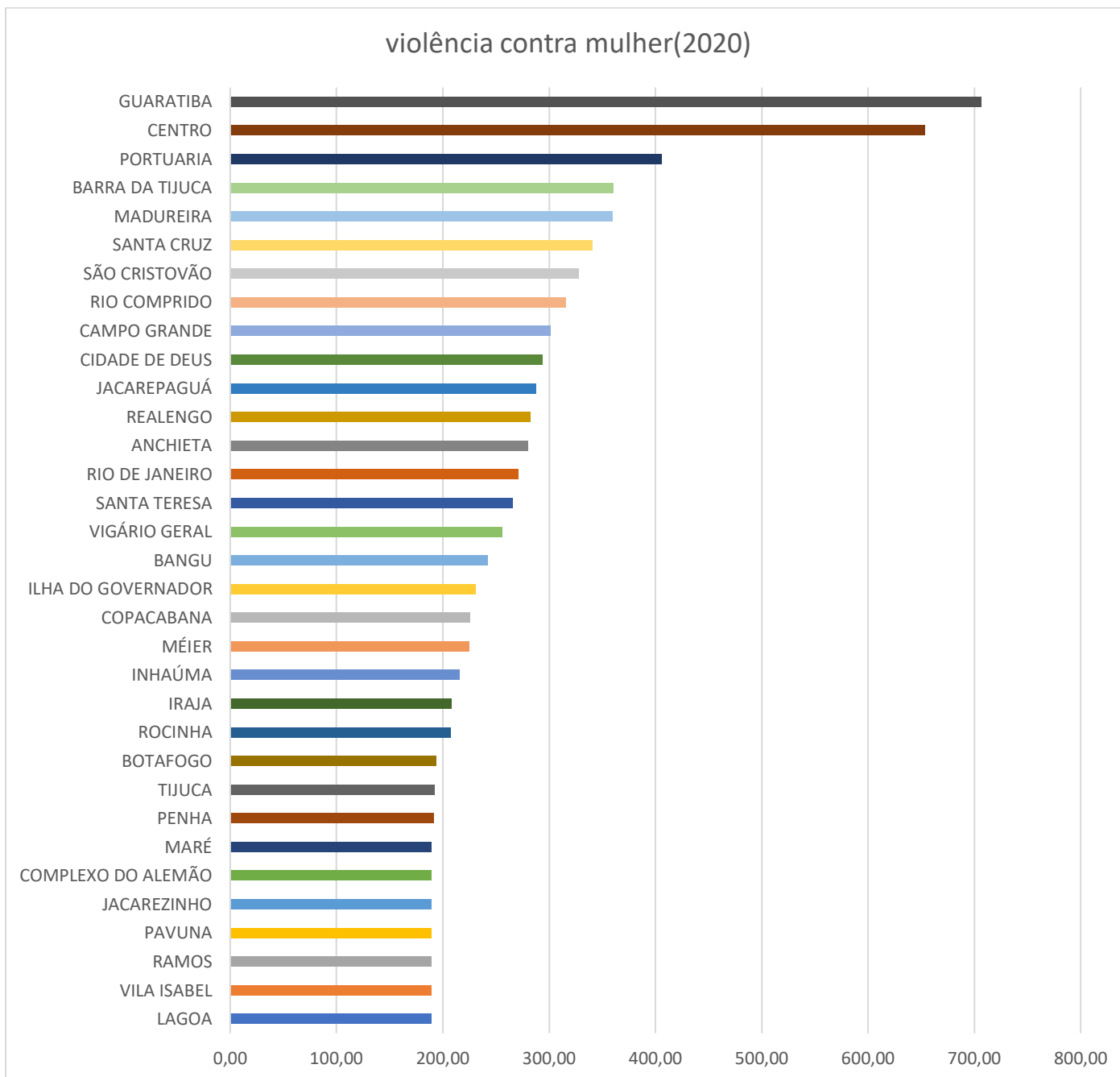
**IDH Mundo 2020 (IDH\_mundo\_2020\_hdr.csv)**

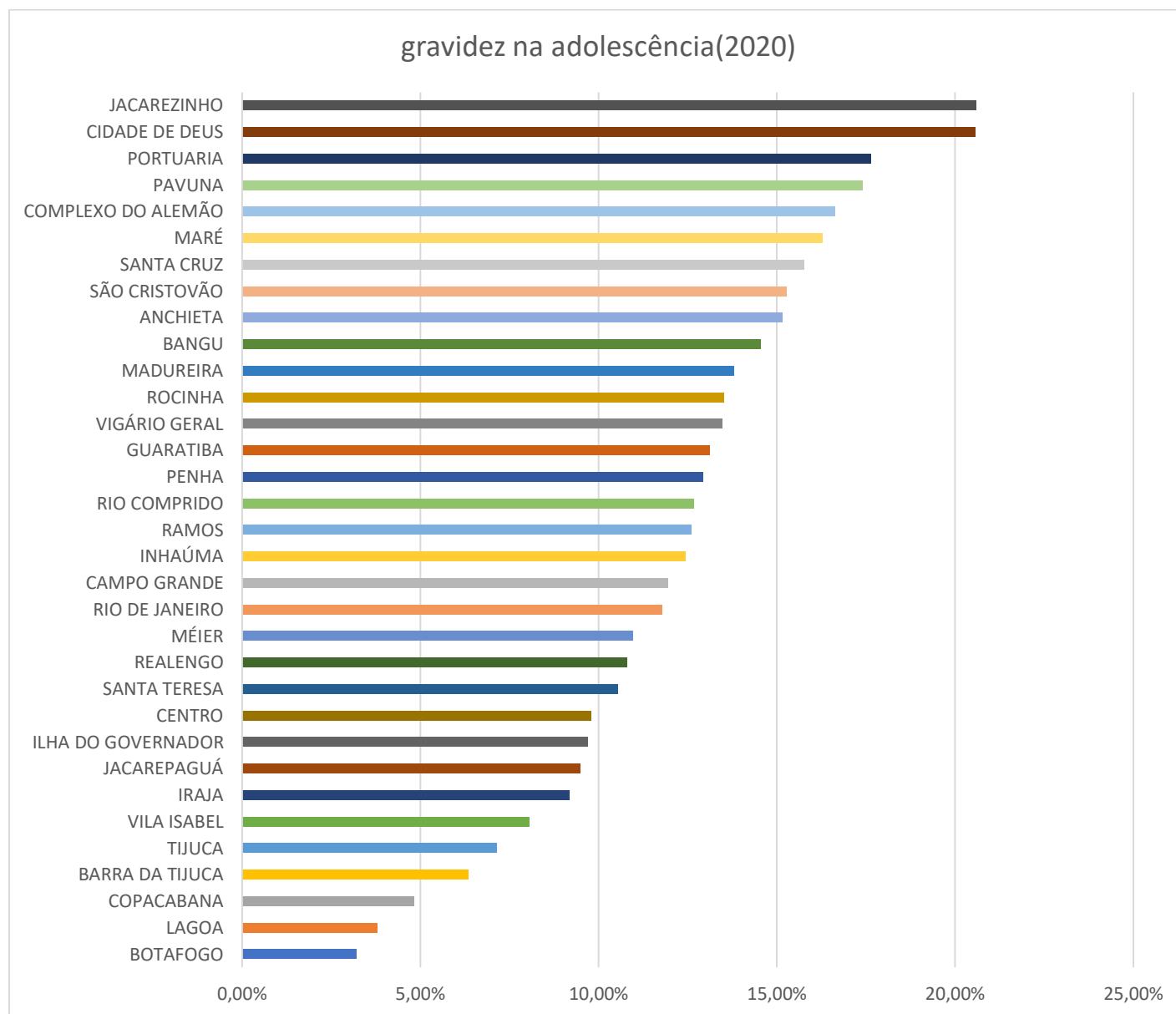
	<b>Countrie</b>	<b>value</b>
<b>189</b>	Niger	0.394
<b>188</b>	Central African Republic	0.397
<b>187</b>	Chad	0.398
<b>186</b>	South Sudan	0.432
<b>185</b>	Burundi	0.433
<b>184</b>	Mali	0.434
<b>182</b>	Burkina Faso	0.452
<b>183</b>	Sierra Leone	0.452
<b>181</b>	Mozambique	0.456
<b>180</b>	Eritrea	0.459

## 4. Análise exploratória dos dados



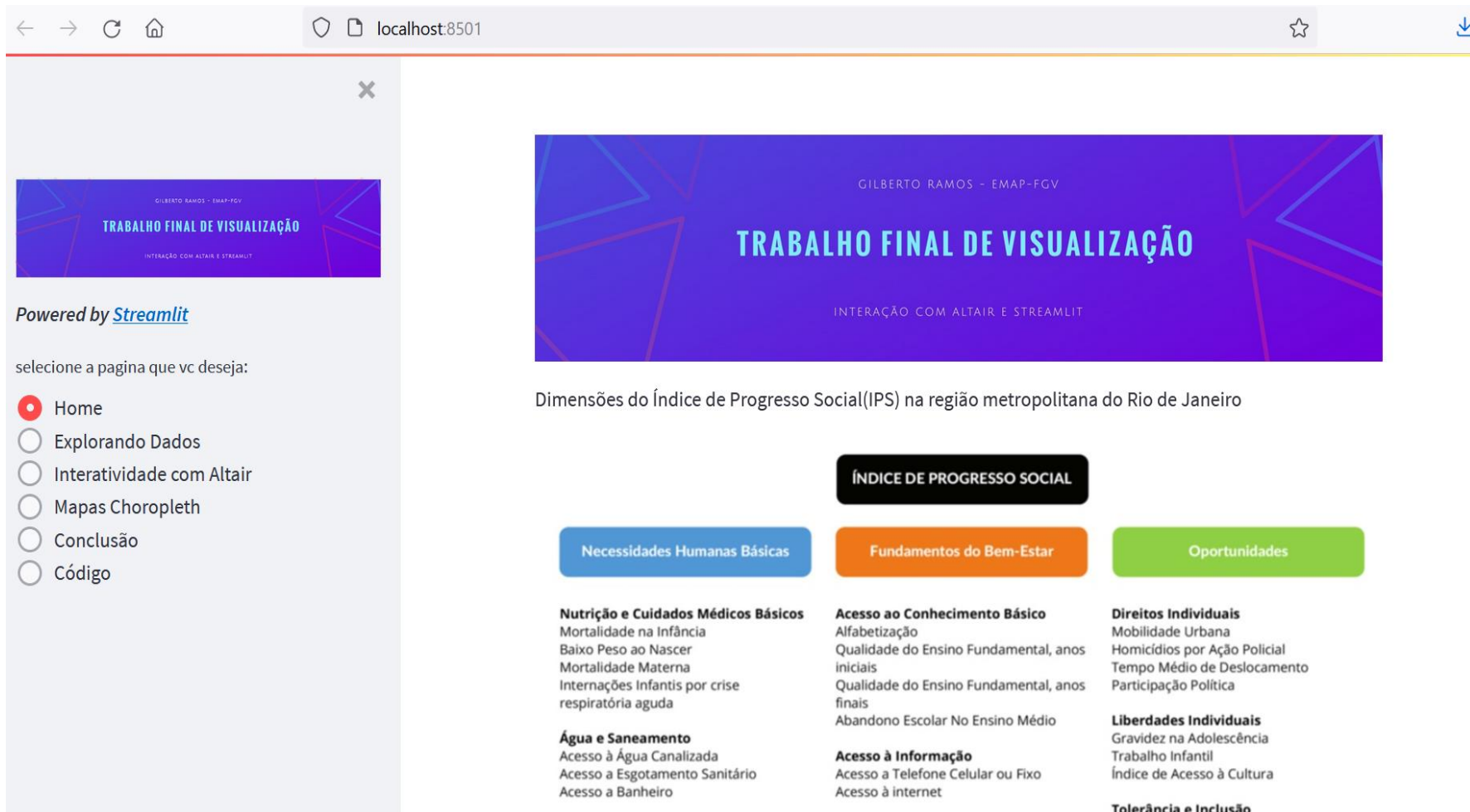






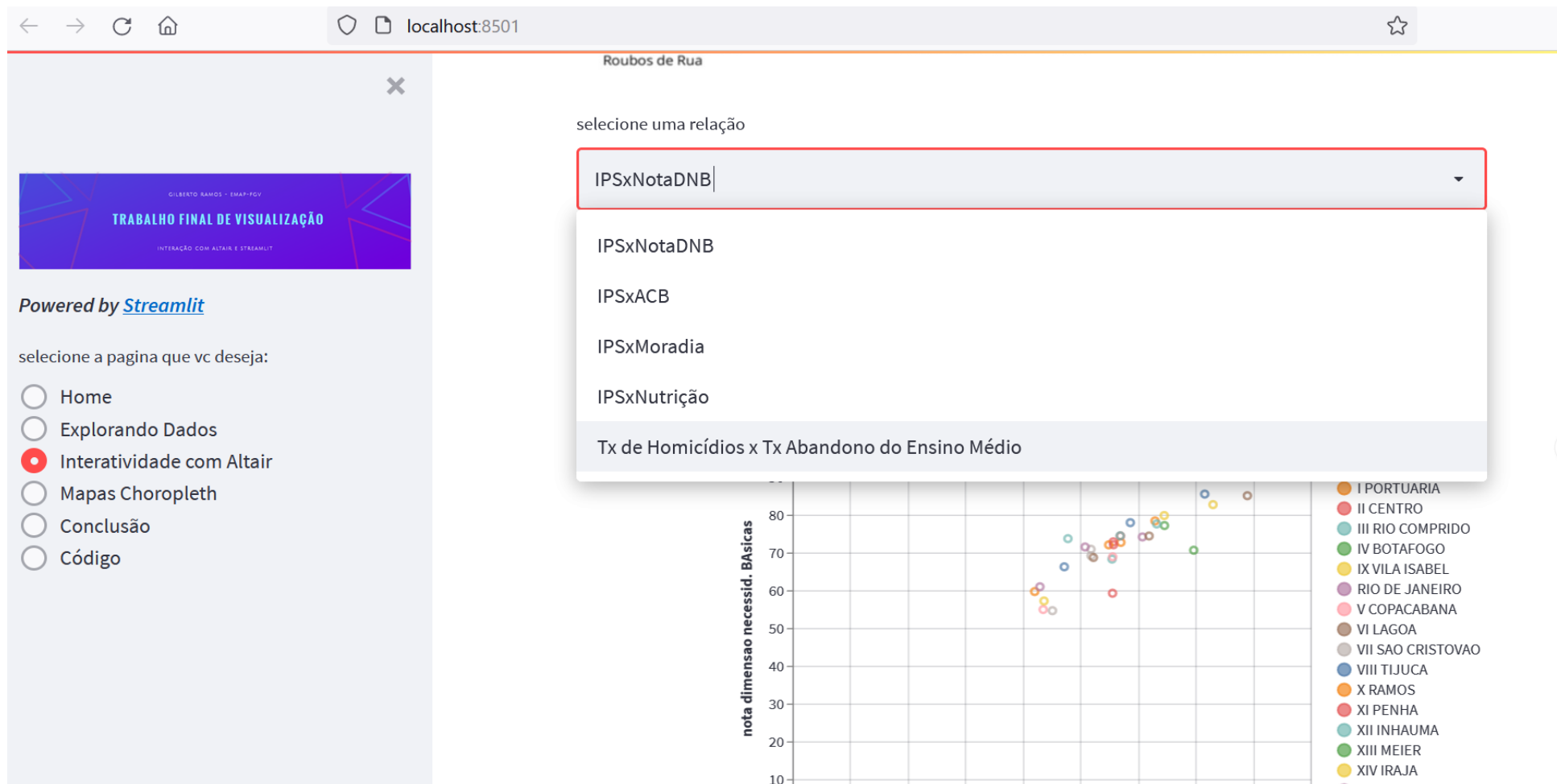
## 5. Produto Final : Web App - Visualização com Altair, Streamlit e Geopandas

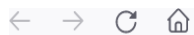
### 5.1 Overview do Web App Streamlit



The screenshot shows a web browser window displaying a Streamlit application. The browser's address bar shows 'localhost:8501'. The application has a purple header with the text 'GILBERTO RAMOS - EMAP-FGV' and 'TRABALHO FINAL DE VISUALIZAÇÃO' in large white letters, with 'INTERAÇÃO COM ALTAIR E STREAMLIT' in smaller text below. A sidebar on the left contains a menu with the following items: 'Home' (selected), 'Explorando Dados', 'Interatividade com Altair', 'Mapas Choropleth', 'Conclusão', and 'Código'. Below the menu, it says 'selecione a pagina que vc deseja:'. The main content area has the title 'Dimensões do Índice de Progresso Social(IPS) na região metropolitana do Rio de Janeiro' and a table of indicators categorized under 'ÍNDICE DE PROGRESSO SOCIAL'.

ÍNDICE DE PROGRESSO SOCIAL		
Necessidades Humanas Básicas	Fundamentos do Bem-Estar	Oportunidades
<b>Nutrição e Cuidados Médicos Básicos</b> Mortalidade na Infância Baixo Peso ao Nascer Mortalidade Materna Internações Infantis por crise respiratória aguda	<b>Acesso ao Conhecimento Básico</b> Alfabetização Qualidade do Ensino Fundamental, anos iniciais Qualidade do Ensino Fundamental, anos finais Abandono Escolar No Ensino Médio	<b>Direitos Individuais</b> Mobilidade Urbana Homicídios por Ação Policial Tempo Médio de Deslocamento Participação Política
<b>Água e Saneamento</b> Acesso à Água Canalizada Acesso a Esgotamento Sanitário Acesso a Banheiro	<b>Acesso à Informação</b> Acesso a Telefone Celular ou Fixo Acesso à internet	<b>Liberdades Individuais</b> Gravidez na Adolescência Trabalho Infantil Índice de Acesso à Cultura
		<b>Tolerância e Inclusão</b>





localhost:8501



selecione uma relação

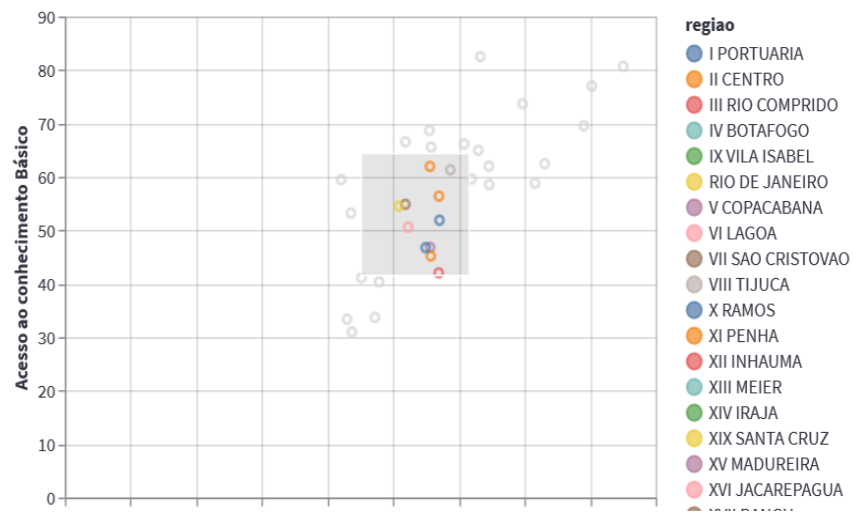
IPSxACB

# Índice de Progresso Social X Acesso ao Conhecimento Básico.\*

Powered by [Streamlit](#)

selecione a pagina que vc deseja:

- ☐ Home
- ☐ Explorando Dados
- ☒ Interatividade com Altair
- ☐ Mapas Choropleth
- ☐ Conclusão
- ☐ Código

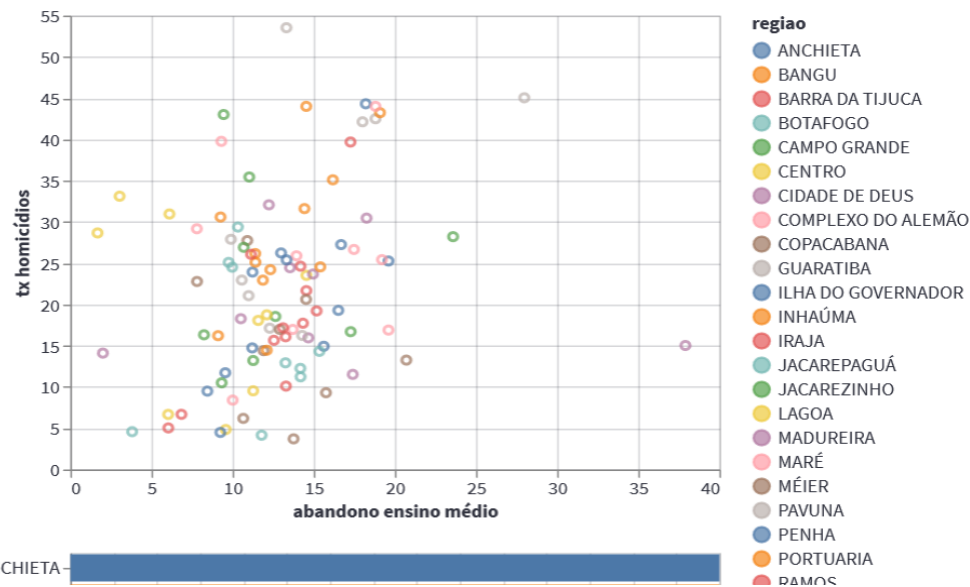


localhost:8501



Tx de Homicídios x Tx Abandono do Ensino Médio

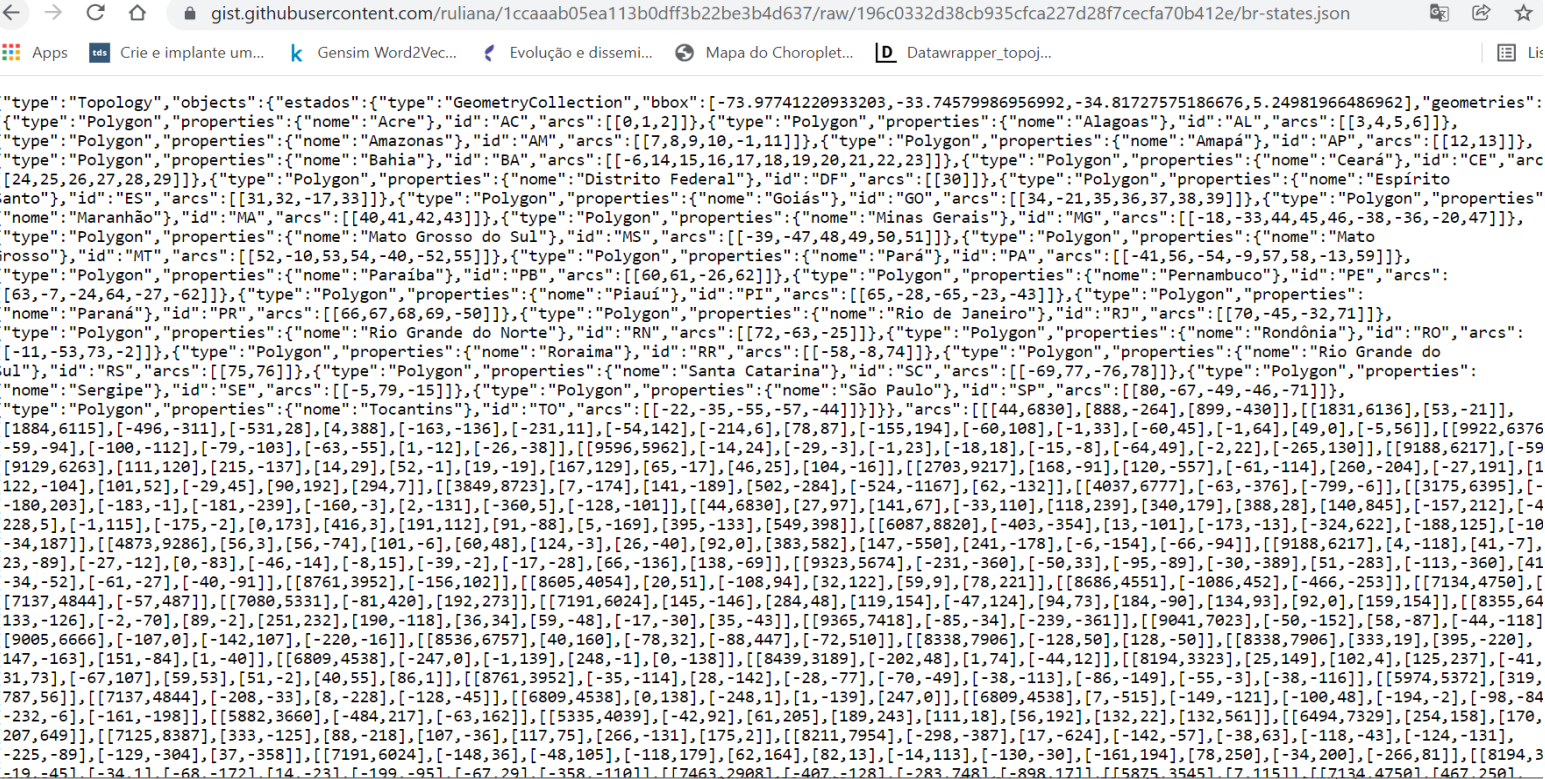
## Tx de Homicídios x Tx Abandono do Ensino Médio





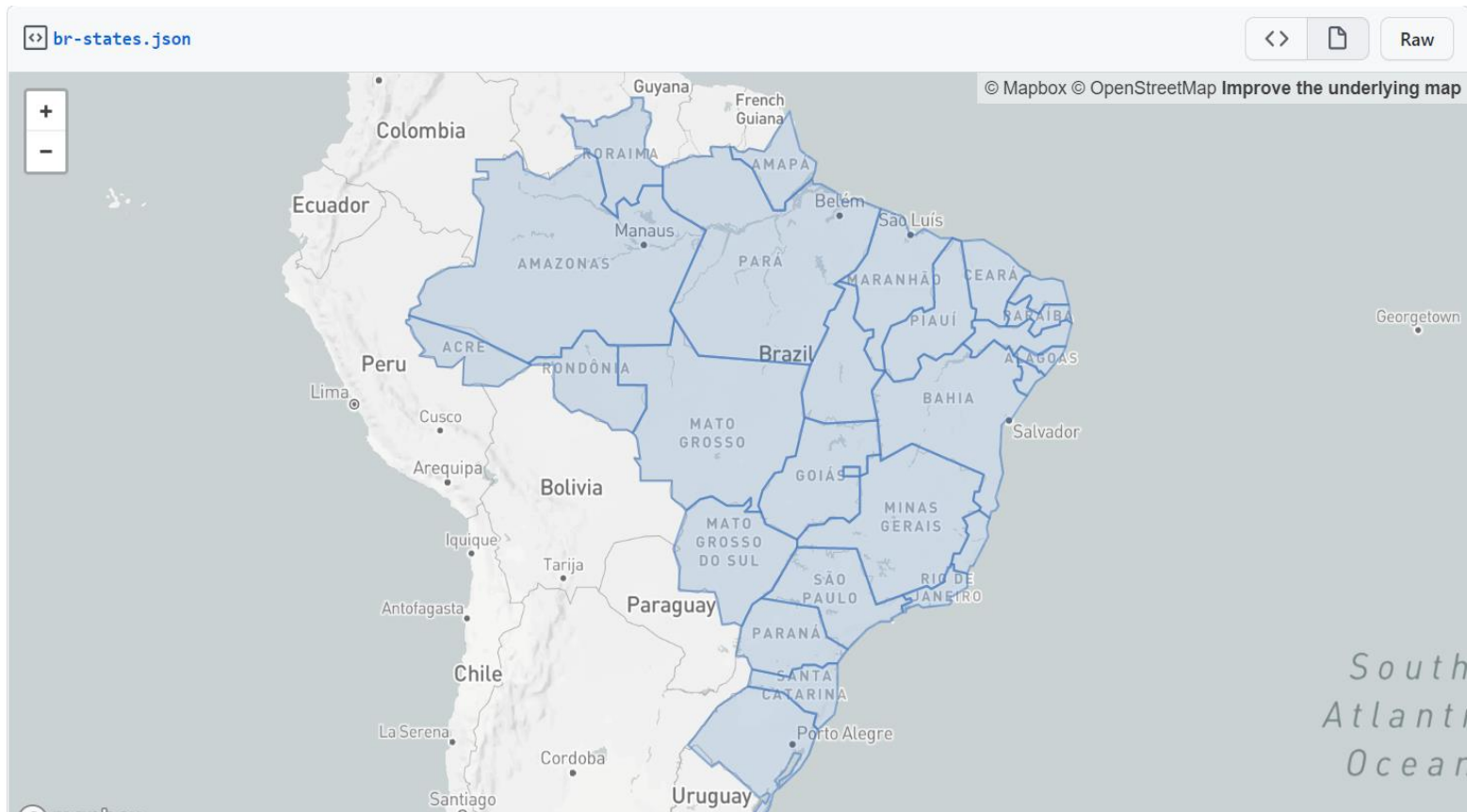
## 5.2 Renderização e Visualização de Mapas Choropleth no Streamlit

### 5.2.1 Exemplo da estrutura de um file TopoJson:





## 5.2.2 Visualização do file TopoJson 'br-states.json'

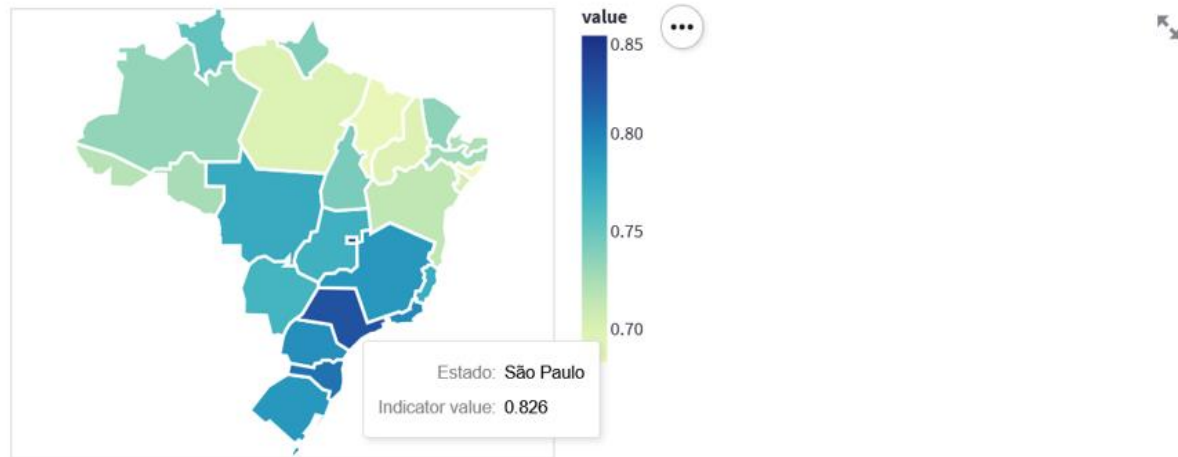


### 5.2.3 Resultado do Mapeamento para Web App Streamlit:



IDH-Brasil

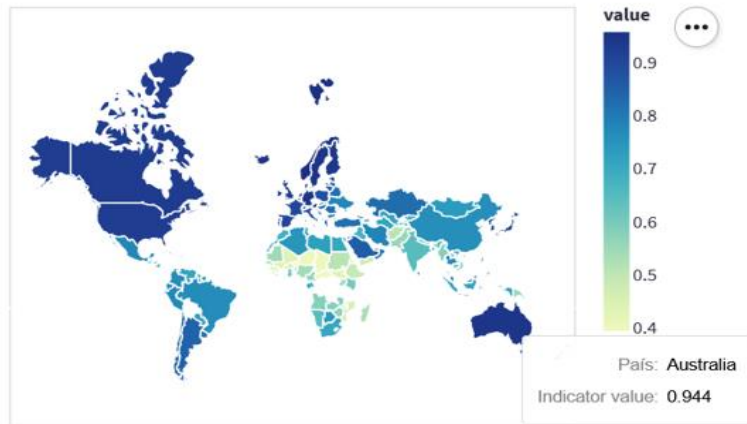
## IDH Brasil em 2020



Fonte: [https://pt.wikipedia.org/wiki/Lista\\_de\\_unidades\\_federativas\\_do\\_Brasil\\_por\\_IDH](https://pt.wikipedia.org/wiki/Lista_de_unidades_federativas_do_Brasil_por_IDH)

IDH-Mundo

## IDH Mundo em 2020



Fonte: <http://hdr.undp.org/sites/default/files/hdr2020.pdf>

## 6. Conclusão

Este trabalho vislumbra a análise exploratória de dados sociais da região metropolitana do Rio de Janeiro e também apresenta ao final uma exploração extra de dados do IDH 2020 para Brasil e Mundo. Este trabalho me ajudou a compreender um pouco da utilização da biblioteca **Altair** utilizando a linguagem Python, bem como explorar os recursos de renderização web da biblioteca **Streamlit**. Consegui aplicar os recursos básicos aos dados coletados da base da prefeitura do Rio de modo a interagir com a biblioteca e os métodos de visualização, acredito poder expandir esse conhecimento para explorar mais aspectos da visualização interativa aplicados a qualquer dataset minimamente qualificado. Consegui também aplicar o recurso de mapas cartográficos **Choropleth**, esse que me deu mais trabalho e em determinado ponto do projeto se tornou um desafio conseguir renderizá-lo no Streamlit.

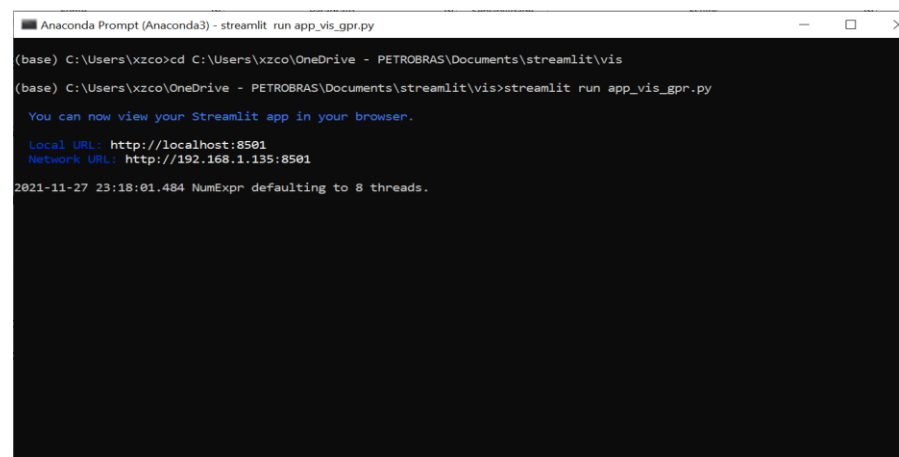
Isso deve-se ao fato de que não é simples a instalação da biblioteca **geopandas**, e também precisei entender um pouco sobre o formato de arquivo cartográfico de topologia **topojson**. A integração desse formato de arquivo é muito sensível ao seu dataset, os nomes de municípios, estados e países devem estar precisamente alinhados como o conteúdo topográfico do arquivo **topojson**. Uma outra dificuldade encontrada é que não é fácil encontrar arquivos topojson com mapas de municípios ou bairros do Rio de Janeiro. Encontrei alguns modelos semelhantes no site o IBGE, mas o formato era geojson, que não funcionava para interatividade que eu queria mostrar. Enfim só consegui fazer funcionar para mapas do Brasil e do Mundo, e nesse caso usei a dimensão IDH para ficar coerente com minha abordagem de aspectos sociais e de desenvolvimento humano. O produto final desse trabalho é um Web App Streamlit que nomeei de **app\_vis\_gpr.py**, o passo para execução desse app é descrito no capítulo 8 – Anexo. Os arquivos utilizados nesse trabalho juntamente com código fonte **.py** podem ser encontrados no meu diretório github: <https://github.com/w3gilberto/VIS>

## 7. Referências

- [1] <https://streamlit.io/>
- [2] <https://docs.streamlit.io/>
- [3] <https://altair-viz.github.io/>
- [4] <https://geopandas.org/en/stable/>

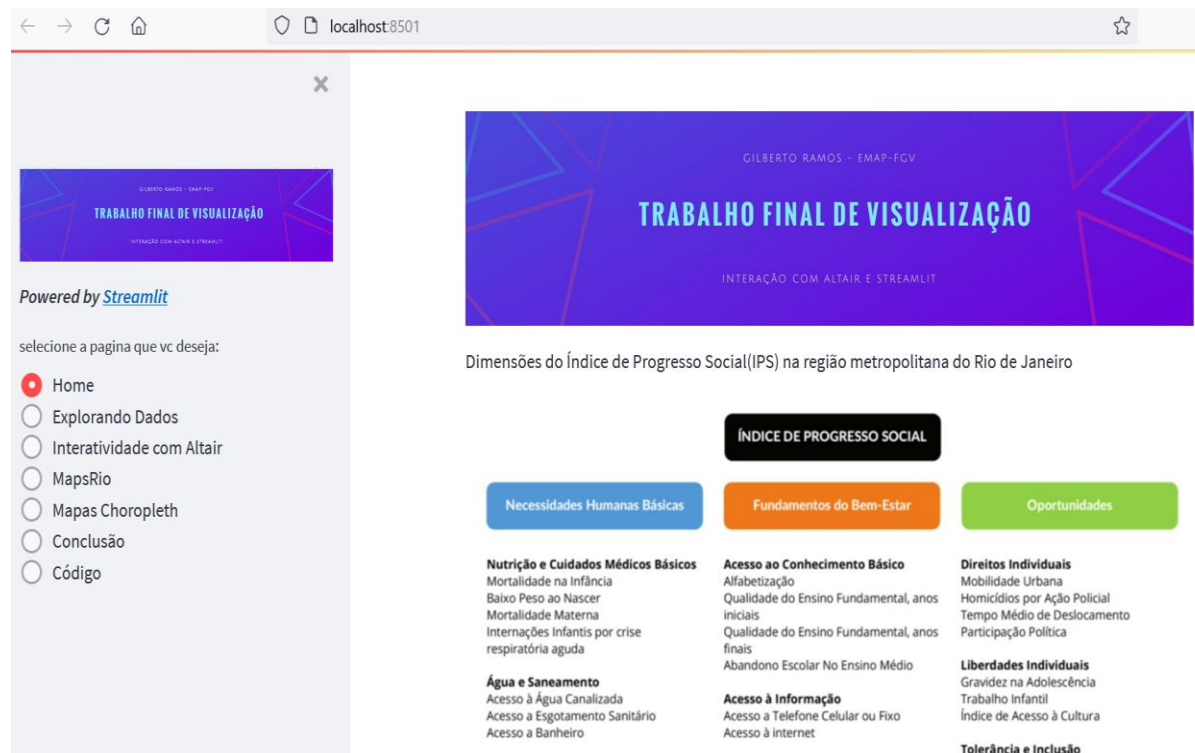
## 8. Anexo (guia passo-a-passo p/ executar o app web via streamlit)

- 1-No **Anaconda** prompt Instalar com comando **pip install** as bibliotecas **Streamlit**, **Altair**, **Vega Datasets** e **Geopandas**.
- 2-Todos os arquivos necessários devem estar no mesma pasta (.py e .csv)
- 3-Abrir anaconda prompt>>>executar linha de comando > **streamlit run app\_vis\_gpr.py**



```
Anaconda Prompt (Anaconda3) - streamlit run app_vis_gpr.py
(base) C:\Users\xzco>cd C:\Users\xzco\OneDrive - PETROBRAS\Documents\streamlit\vis
(base) C:\Users\xzco\OneDrive - PETROBRAS\Documents\streamlit\vis>streamlit run app_vis_gpr.py
You can now view your Streamlit app in your browser.
Local URL: http://localhost:8501
Network URL: http://192.168.1.135:8501
2021-11-27 23:18:01.484 NumExpr defaulting to 8 threads.
```

Nesse momento será aberto o Browser com a tela do app web gerada pelo Streamlit:



**Para funcionar a renderização de mapas Choropleth no Streamlit:**

Deve-se instalar via prompt anaconda as seguintes bibliotecas nas versões compatíveis com a sua versão python (no meu caso versão 3.8):

Libraries: Descartes; shapely; pyproj; fiona; gdal e geopandas

Fonte: site [unofficial windows Binaries for Python Extension Packages]: <https://www.lfd.uci.edu/~gohlke/pythonlibs/>

No meu caso na versão python 3.8 precisei baixar os seguintes files de pacotes, nessa ordem:

- 1- jsonobject-0.9.10-cp38-cp38-win\_amd64.whl
- 2- descartes-1.1.0-py2.py3-none-any.whl
- 3- pyproj-3.3.0-cp38-cp38-win\_amd64.whl
- 4- Shapely-1.8.0-cp38-cp38-win\_amd64.whl
- 5- Fiona-1.8.20-cp38-cp38-win\_amd64.whl
- 6- GDAL-3.3.3-cp38-cp38-win\_amd64.whl
- 7- geopandas-0.6.2-py2.py3-none-any.whl