

Project proposal

Andrei Teixeira and Rafael Gouveia
Visualização Avançada de Dados
Mestrado em Engenharia e Ciência de Dados (MECD)

Estrutura do doc:

- I. Autores
- II. Introdução
- III. Trabalhos relacionados
- IV. Requisitos de design
- V. Dados

I. Autores

Andrei Fokin Teixeira

- email: andrei.fokin.teixeira@gmail.com

- #UC: 2022135701

Rafael Santos Gouveia

- email: rafaelgou1110@gmail.com

- #UC: 2022130518

II. Introdução

A História da humanidade está repleta de eventos de grande impacto como guerras, crises econômicas e sanitárias ou alterações políticas das mais diversas naturezas e conforme o tempo passa, maior a frequência de eventos por conta dos avanços tecnológicos e sociais que aumenta a velocidade das interações e mais ampla é a reação em diferentes povos e países devido à maior inter-relação entre estes.

Ser capaz de entender como um evento impacta determinada estatística de uma população é interessante para uma melhor tomada de decisão de políticas no nível público e de investimentos no nível privado. Muitas vezes, ações são tomadas baseadas em ideias pré-concebidas sobre um determinado tema, em ideologias e/ou preconceitos, em teorias ou por mera repetição de ideias populares.

O mundo também está cada vez mais repleto de dados e existe um potencial enorme para buscar tais padrões e extrair valor das informações obtidas. E para tal, usar não apenas comunicação verbal, mas também comunicação visual em novos estudos e plataformas.

Nossa calculadora, "InfoCountry", é uma plataforma para visualizar a evolução de dados socioeconômicos dos países ao longo do tempo e permite a comparação entre um ou mais países, podendo-se escolher um deles como referência. Nesta versão, é disponibilizada uma lista dos principais eventos nacionais de 39 países europeus desde o ano de 1945 (ano do fim da Segunda Guerra Mundial).

A ferramenta permite explorar o comportamento e comparar a performance média de uma variável socioeconômica de um ou mais países em relação a um país de referência em dois períodos diferentes de tempo, tendo como marco separador um evento marcante para a história do país-referência, por exemplo, a Reunificação da Alemanha, a entrada/saída de um país na União Europeia ou a desintegração da antiga Iugoslávia. Logo, o principal papel é ser uma ferramenta de análise exploratória voltada para um maior potencial visual.

Por isso, este relatório se propõe a descrever com detalhes técnicos as características e a viabilidade da ferramenta a nível de código e de visualização.

III. Trabalhos relacionados

Quando se trata de desenvolvimento dos países, comumente comparações são feitas e exemplos de sucesso são destacados junto com possíveis respostas que as expliquem. No Brasil, um exemplo clássico de comparação é com a Coreia do Sul ao olhar renda *per capita*, apresentado a seguir na Imagem 1, cuja fonte é o site do Countryeconomy, que compara países apenas 2 a 2, mas que consegue informações a longo prazo:

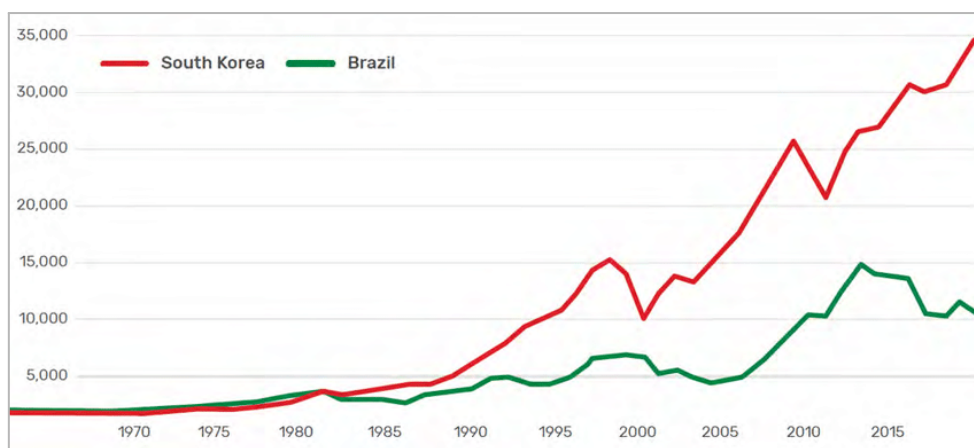


Imagem 1. Comparação do PIB per capita do Brasil e da Coreia do Sul (Fonte: [Countryeconomy](#))

Outro exemplo, na Imagem 2, cuja fonte é o site do Banco Mundial, comparando N países, mas com menos dados disponíveis, é o de Portugal e Espanha em relação a alguns países bálticos, como Lituânia e Estônia, onde o flip da renda *per capita* é mais recente e já aconteceu:

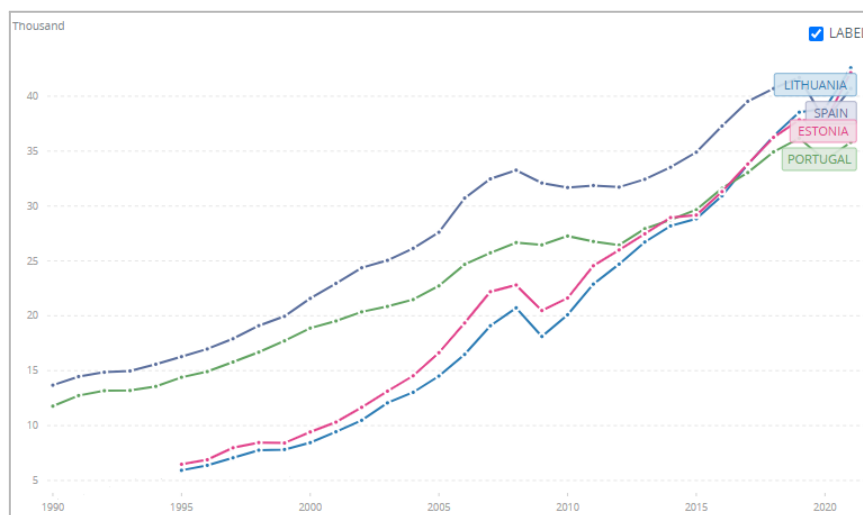


Imagem 2. Comparação do PIB per capita de Portugal, Espanha, Estônia e Lituânia (Fonte: [The World Bank Data](#))

Visualizações como estas são bem comuns de se encontrar e geralmente estão no formato de gráfico de linhas, o que ajuda muito no entendimento da tendência e mostra um pouco da aproximação e do distanciamento dos países. Interessante também seria haver uma visualização comparativa e que tivesse também uma indicação de um evento importante que pode explicar um começo de divergência/convergência de tendências entre séries temporais.

Quando se trata de comparações no mesmo instante do tempo, geralmente se observam os valores brutos das variáveis de todos os países, como por exemplo em mapas-mundi. É possível extrair muita informação dessa maneira, principalmente se variáveis estão ao redor de zero ou de um valor de referência que é conhecido e usado por um grupo de pessoas. A Imagem 3 ilustra a comparação com valores absolutos em torno de zero (quando as cores mudam):

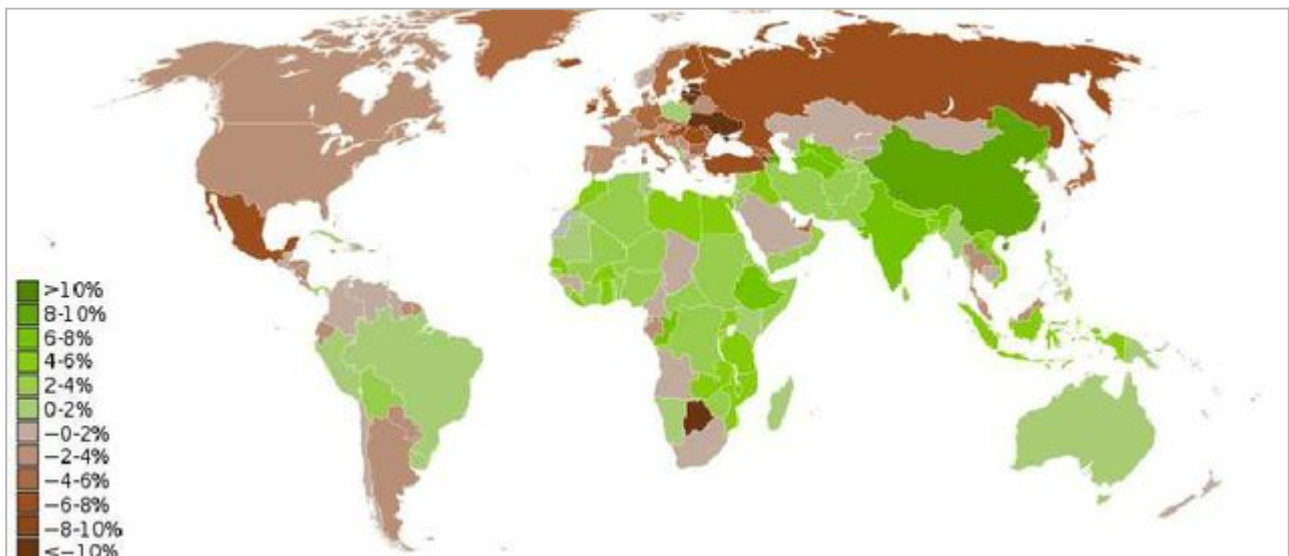


Imagem 3. Crescimento do PIB dos países em 2009 (Fonte: [ResearchGate](#))

Com relação à interação, um dos sites mais conhecidos é o Gapminder, que com sua ferramenta Gapminder Tools, é possível explorar a dinâmica dos dados ao longo do tempo para diferentes países e em diferentes tipos de gráficos. A dinâmica é muito interessante, mas muitas vezes a informação fica perdida pela alta performance visual. A Imagem 4 mostra a quantidade de possibilidades oferecidas pelo site:

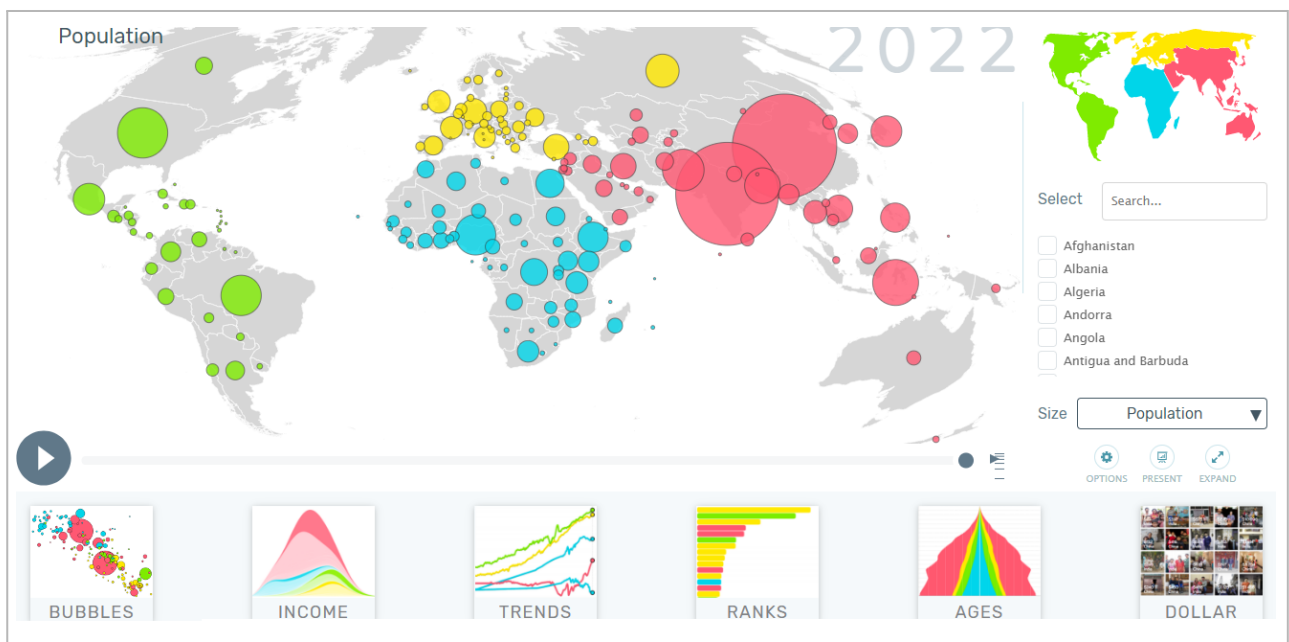


Imagem 4. Visualizações dinâmicas (Fonte: [Gapminder Tools](#))

Um tipo de comparação menos comum e que é um dos objetos deste trabalho, são comparações em termos percentuais entre poucos países e para isso, é necessário adotar um país como referência. Na nossa “calculadora”, intitulada "InfoCountry" será adicionada informação sobre os principais eventos dos países para que o utilizador tenha um tipo adicional de iteração: a iteração buscando relacionar os números com acontecimentos do mundo real.

Para ilustrar o interesse deste trabalho, foi criada uma visualização simples em Tableau e olhando apenas uma variável: crescimento anual do PIB (%). O Mockup 1 mostra a comparação da evolução da variável de interesse para Alemanha e Suíça em relação à Áustria e tomando como marco o ano de 1995, quando este último entrou na União Europeia. Na parte 1, é a comparação entre países e na parte 2 uma visualização individual livre (podem ser os mesmos países acima ou não):

IV. Requisitos de design

De maneira geral, gostávamos de ter uma visualização em que seja possível:

- escolher diferentes países para consulta;
- escolher até 5 países (para não ter demasiados dados);
- observar eventos nacionais que são marcos para os países selecionados;
- ter um scroll que ajuda a ver os anos de alguns eventos para os países selecionados;
- um botão para escolher como referência um dos países selecionados ou a média dos valores de suas variáveis;
- escolher diferentes variáveis socioeconômicas para consulta;
- escolher até 3 dessas variáveis;
- ter um mapa para ver os países de maneira comparada;
- no mapa comparar em relação à média do país de referência ou em relação à média de todos os selecionados;
- ver dois mapas, um com a média da variável antes do evento-marco e outro com a média posterior ao evento, independente de haver um país de referência ou não;
- escolher diferentes momentos do tempo para cada mapa para poder comparar ou então, ter um campo onde se digita o número X de anos e os mapas já calculam marco+X e marco-X;
- ver em um outro gráfico a evolução individual da variável escolhida → para isso gráfico de linhas

Logo, temos três grandes blocos de visualização:

- eventos;
- países comparados;
- países individualizados.

De maneira específica, assim poderia ser:

- específicos para países:

>> botão para tornar um país como referência (seletor).

- específicos para os eventos:

>> scroll para selecionar o intervalo de tempo dos eventos (seletor).

>> tabela para mostrar eventos dos países (resultado).

- específicos para o comparador de países (PARTE 1):

>> dropdown para países (seletor);

>> scroll para selecionar o intervalo de tempo (seletor);

>> checkboxes para variáveis (seletor);

>> mapa para mostrar a comparação das médias (resultado);

>> gráfico de barras horizontal para comparar a média da variável (resultado);

>> gráfico de barras horizontal para comparar a diferença média entre países, tomando tendo um como referência (resultado).

- específicos para a informação de países (PARTE 2):

>> checkboxes para variáveis (seletor);

- >> dropdown para países (seletor);
- >> scroll para selecionar o intervalo de tempo (seletor);
- >> gráfico de linhas para mostrar a evolução individual de cada país (resultado).

**INFOCOUNTRY**

Your country information app

Year event



Countries to compare

(Valores múltiplos)

country eve... event (Sheet12)

Austria	2002 European floods: Heavy rains resulted in destruc...	2002
	2004 Austrian presidential election: Heinz Fischer of t...	2004
Germany	2006 FIFA World Cup: The 2006 FIFA World Cup was he...	2006
	Alliance '90/The Greens was established from the mer...	1993
	Eschede train disaster	1998
	European debt crisis: Greece requested a loan from th...	2010
	Expo 2000: A world's fair was held in Hanover.	2000
	German federal election, 2009: Elections were held to ...	2009

Country reference

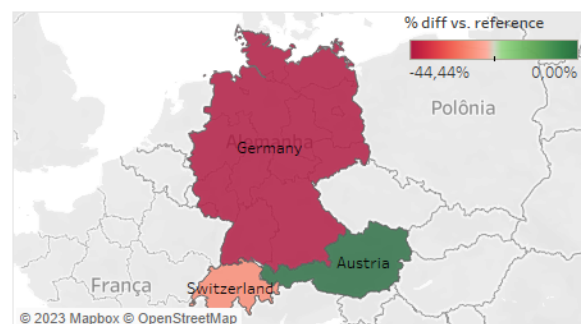
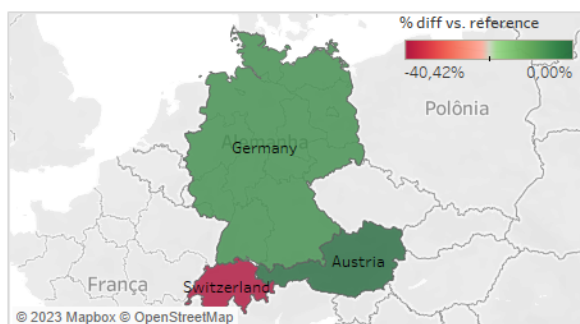
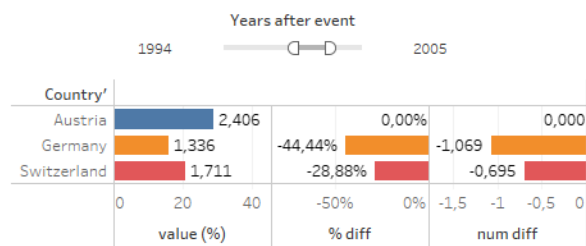
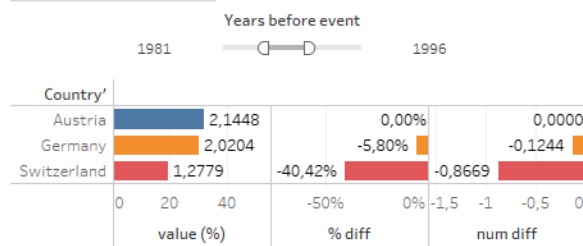
Austria
Germany
Switzerland

Change here too

Austria
Germany
Switzerland

COUNTRIES COMPARATOR

Indicators

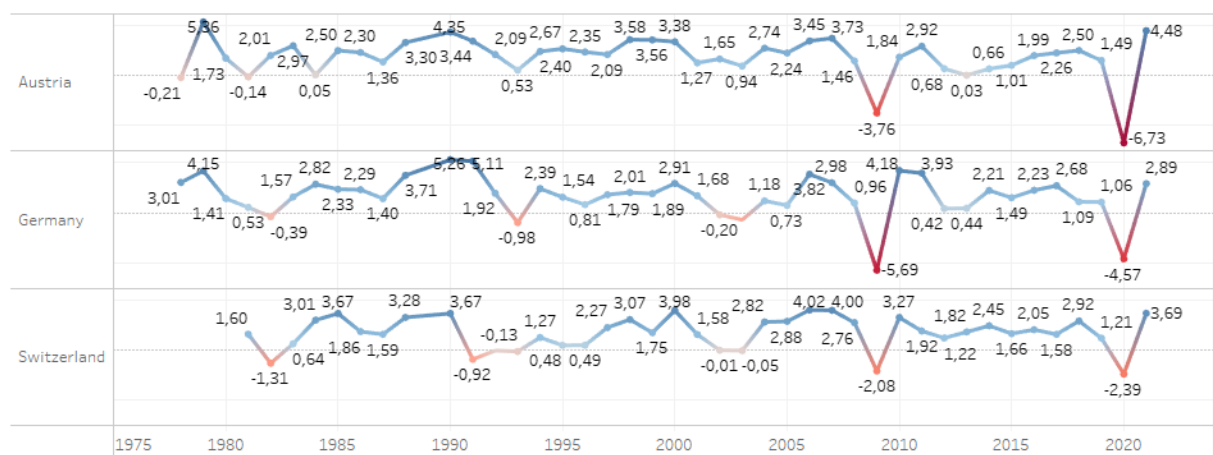
☐ (Tudo)☒ GDP growth (annual ...

INDIVIDUAL INFORMATION

Years

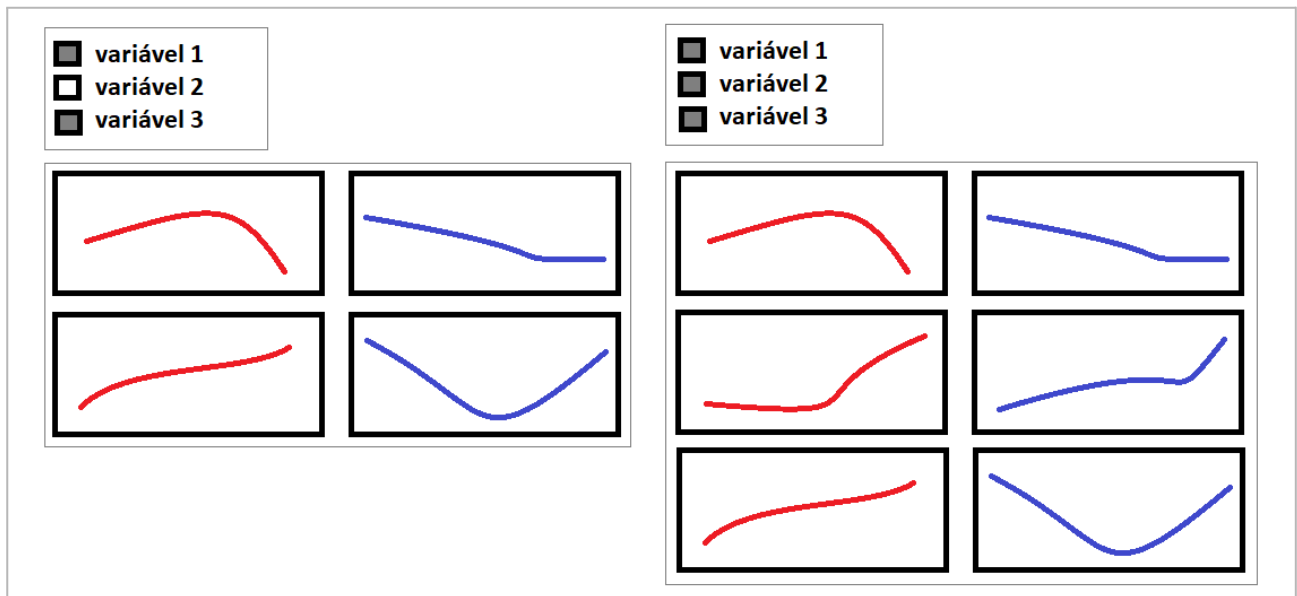


Time serie (individual values)



Mockup 1. Comparação de uma variável para três países (Fonte:Autores)

Se for tecnicamente viável, outro requisito será o seguinte: ao clicar em uma nova variável, mostrar um segundo par de gráficos e assim por diante (todos submetidos ao mesmo intervalo de tempo do Mockup 1). A seguir, o Mockup 2 ilustra a proposta para duas ou mais variáveis:



Mockup 2. Ideia visual da proposta, para duas ou mais variáveis

Durante a fase de código, serão descobertas as potencialidades da ferramenta, permitindo uma previsão mais precisa dos requisitos.

<Mockup 1 disponibilizado em documento Tableau>

V. Dados

Para iniciar o projeto, extraem-se quatro variáveis de interesse com frequência anual: crescimento do PIB (%), PIB *per capita* (US\$), taxa de inflação (%) e expectativa de vida. São extraídos do site do Banco Mundial, em formato tabular e pronto para serem trabalhados posteriormente em Python enquanto dataframe.

Já os dados dos eventos são obtidos de duas maneiras: (i) a partir de Web Scraping realizado em páginas da Wikipédia onde o dado estava devidamente tabelado [válido para 25 países]; e (ii) criados manualmente a partir de páginas da Wikipédia em que a informação estava disponível, mas não em formato tabular [válido para 2 países].

Para países cujos dados não estão organizados na Wikipédia, será estudado a viabilidade de buscar sites relacionados a eventos históricos dos demais países e, em caso positivo, para que países (por exemplo, Países Baixos, Dinamarca, Chéquia e Ucrânia).

A seguir, a Tabela 1 mostra um resumo sobre os dados de origem:

DADOS	variáveis de interesse	eventos tabulados	eventos não-tabulados	fora da Wikipédia
fonte	site do Banco Mundial	páginas da Wikipédia, cujos dados estavam tabulados	páginas da Wikipédia, cujos dados não estavam tabulados	páginas suporte: ex xxx, xxx, xxx, xxx, xxxxx, xxxxx
tipo	quantitativas	temporal (ano) + qualitativa (evento)	temporal (ano) + qualitativas (evento)	
estrutura	tabular (excel)	tabular (HTML)	escrito listado → tabular (excel)	
países	todos os 39 países de interesse	Portugal , Espanha , França , Suíça , Alemanha , Áustria , Eslovênia , Itália , Bélgica , Reino Unido , Irlanda , Islândia , Suécia , Finlândia , Rússia , Estônia , Letônia , Polônia , Romênia , Sérvia , Bulgária , Croácia , Albânia , Chipre e Malta (25)	Grécia e Luxemburgo (2)	Dinamarca, Noruega, Lituânia, Belarus, Ucrânia, Moldávia, Eslováquia, Chéquia, Montenegro, Bósnia, Macedônia do Norte e Países Baixos (12)
recursos/extração	download para o PC e upload para Jupyter Notebook	técnicas de web scraping das URL das páginas; no código Python	extração manual para um Excel e upload para Jupyter Notebook	extração manual para um Excel e upload para Jupyter Notebook
código	Python			
visual	Pyplot/Dash			

Tabela 1. Características dos dados a serem usados