

Olha no Google! Usando o Google Trends para Analisar a PEB¹

Felipe Ferreira de Oliveira Rocha - Universidade Federal de Pernambuco (UFPE)

Resumo

O artigo possui objetivo duplo: a) discutir a importância de se incorporar dados empíricos ao realizar análises sobre a Política Externa Brasileira; b) demonstrar os limites e as possibilidades por trás da utilização de uma plataforma de acesso público e livre para a coleta de dados que podem servir como indicadores essenciais para se entender a inserção internacional do Brasil. Nesse sentido, buscamos estimular o uso consciente e cauteloso de ferramentas como o Google Trends em análises acadêmicas. Para isso, primeiro, demonstramos a situação de escassez de dados empíricos na análise da Política Externa Brasileira. Em segundo lugar, apresentamos as possibilidades e os limites por trás da coleta de dados sociais provenientes dos rastros virtuais de milhares de usuários. Por fim, simulamos, na prática, uma minipesquisa cujo intuito fosse explorar e descrever diferentes padrões da sociedade brasileira e argentina ao pesquisarem na internet sobre o Mercosul. Desse modo, esperamos que esse artigo sirva de porta de acesso inicial ao complicado e instigante mundo da análise de dados.

Palavras-chave: Análise de Política Externa; Política Externa Brasileira; Análise de Dados; Big Data; Google Trends.

Abstract

This paper has two main objectives: a) to discuss how important it is to use empirical data when analyzing Brazilian Foreign Policy; b) to demonstrate both the promises and perils of a free and public data extraction platform which its data can be employed as key indicators for analysts to better understand Brazilian international insertion. In this sense, we want to stimulate the conscious and cautious use of tools such as Google Trends in academic analysis. In order to do that, first, we comment on the scarcity of empirical data in the analysis of Brazilian Foreign Policy. Secondly, we present the limits and the possibilities of collecting data coming from millions of users. At last, we simulate, in practice, a mini-research so we can explore and describe different patterns of Brazilian and Argentinian society when they search for Mercosur on the internet. Hopefully, this paper may serve as a gateway to the thought-provoking world of data analysis.

Keywords: Angola: Foreign Policy Analysis; Brazilian Foreign Policy; Data Analysis; Big Data; Google Trends.

1 O presente trabalho foi realizado com apoio do CNPq, Conselho Nacional de Desenvolvimento Científico e Tecnológico – Brasil.

1. Introdução

Cada vez mais os brasileiros ampliam seu acesso às tecnologias de informação, garantem significativa presença nas mais diversas redes sociais e usam diariamente a internet como meio de interação, de pesquisa e de manifestação de sua opinião sobre os mais diversos aspectos da vida social. Acabam por produzir assombrosa quantidade de dados (*Big Data*) e seus rastros virtuais ficam acessíveis para uma pluralidade de fins, dentre as quais as pesquisas científicas (KAPLAN; HAENLEIN, 2010).

Apesar desse cenário de abundância de dados, o analista de Política Externa Brasileira (PEB), por vezes, se depara com uma colossal dificuldade de encontrar indicadores tradicionalmente utilizados pelos demais pesquisadores para realizar suas inferências causais e descritivas. Há, nesse caso, um cenário de escassez que faz com que algumas produções acabem por apresentar número insuficiente de evidências para sustentar determinadas conclusões e “saltos” lógico-argumentativos (CAMPOS, 2015).

Diante disso, esse artigo objetiva auxiliar os analistas de PEB a encontrar achados empíricos que sejam úteis ao estudo de variados tópicos, como opinião pública, saliência temática, interesse em Relações Internacionais (RI), dados de tendência, dados sobre processos decisórios da PEB, entre outros. Almejamos propor um meio relativamente fácil e de acesso livre para que, assim, os pesquisadores se beneficiem dos abundantes dados virtuais em seus trabalhos. Destarte, o presente artigo demonstra, na prática, como acessar e coletar dados do *Google Trends* (GT) e como os mesmos devem ou não ser usados, quais suas limitações e pressupostos, bem como casos em que podem ser aplicados.

Para tanto, o artigo se divide em três seções, desconsiderando essa introdução e as futuras considerações

finais. A próxima parte comenta brevemente sobre o uso da análise de dados na PEB, o seu cenário de escassez e de possibilidades. Em seguida, discutimos sobre a importância do *Google Trends* nas pesquisas científicas e quais são os seus pressupostos, limites e dificuldades. Por fim, na terceira seção, apresentamos um guia rápido sobre como acessar e coletar dados na plataforma, além de trazer uma possibilidade concreta de análise ao trabalharmos o interesse das sociedades brasileira e argentina para com o Mercosul.

Esperamos, assim, estimular, na medida do possível, o emprego de dados empíricos em análises feitas por pesquisadores das Relações Internacionais e demonstrar como coletar indicadores sociais provenientes de meios virtuais, como a internet, não precisa ser, todas as vezes, vista como algo difícil e/ou inacessível.

2. Peb: entre a escassez de dados e “big data”

Ao analisar o que chama de labirinto metodológico das Relações Internacionais, Campos (2015) identifica algumas armadilhas, desafios e dificuldades ao tentarmos conectar esforços teóricos com análise empírica na disciplina. A autora elenca uma série de problemas – como o uso exagerado de estudos de caso, as dificuldades da mensuração e da definição de causalidade – e chama atenção para a necessidade de produzir mais pensamento reflexivo, crítico, didático e educacional para que, assim, seja possível ampliar a preocupação metodológica nas RI.

Semelhantemente, Medeiros *et al.* (2016, p. 25) demonstrou, via coleta de dados bibliográficos e respectiva análise de conteúdo, que as RI sul-americanas tendem a um “desconhecimento metodológico generalizado”², uma vez que apenas 12,94% dos 7,857 artigos analisados mencionaram o emprego de algum tipo de

método. Sem sombra de dúvidas, esse é um cenário bastante assustador, principalmente se considerarmos que “o conteúdo da ciência é primariamente os métodos e as regras, não o assunto em si, já que podemos usar esses métodos para estudar praticamente qualquer coisa”³ (KING *et al.*, 1994, p. 9).

Naturalmente, não se trata de hierarquizar a ciência e considerar que metodologia seria supostamente mais importante do que teoria (ou vice-versa) ou que a lógica qualitativa seria menos relevante do que a quantitativa (ou vice-versa). Ao contrário, trata-se de alertar ao pesquisador social que ele precisa tentar equilibrar ao máximo as atividades analíticas com as reflexivas para que, desse modo, não cometa um dos seguintes erros: “(1) teorizar sem dedicar muita atenção aos testes; e (2) realizar testes empíricos sem dedicar muita atenção à teoria”⁴ (MEARSHEIMER; WALT, 2013, p. 437).

Nesse sentido, nas RI da América Latina, o uso da narrativa histórica, por vezes, predomina hegemônica tanto sobre a reflexão teórica, quanto sobre a modelagem de dados (CAMPOS, 2015). De forma análoga (e até mesmo mais intensa), os estudos da PEB parecem padecer de um mal duplo causador de um ciclo vicioso: os dados disponíveis para as pesquisas são raros e de difícil operacionalização, ao mesmo tempo em que alguns pesquisadores parecem ignorar as possibilidades que surgem de coleta e análise dos mais variados indicadores empíricos.

Quanto ao problema da escassez de dados em PEB, convém mencionar as conclusões a que chegam Waisbich *et al.* (2017, p. 191) ao mapearem todos os pedidos de informação recebidos pelo Itamaraty desde 2012 e sua respectiva resposta (seja recusando ou aceitando liberar dados). Conforme destacado pelas autoras, apesar dos avanços fomentados através de leis em prol da transpa-

rência institucional, o Itamaraty “não está suficientemente disposto ou capaz de se tornar mais transparente, algo que deveria ser monitorado de perto pelos órgãos de fiscalização estabelecidos no âmbito da FOI”⁵⁶

Definitivamente, esse deve ser um fator a ser considerado como explicativo da escassez de dados em estudos de PEB: o alto insulamento burocrático, a baixa democratização da política externa e a área cinzenta de disponibilização de informações. Entretanto, convém lembrar dois pontos necessários. Primeiro, o Itamaraty, desde a redemocratização, esforça-se para ampliar o acesso aos dados. Disponibiliza boa parte das conferências, reuniões, entrevistas, discursos, notas, documentos históricos em plataformas de acesso público facilitado (FARIA, 2012). Obviamente, ainda estamos longe de um modelo de transparência ideal, mas há já um número significativo de material para pesquisa qualitativa e quantitativa.

Segundo, é de interesse e parte do trabalho do pesquisador buscar alternativas criativas em contextos de escassez. Quer seja ou não estimulada, a ciência precisa avançar para que possa trilhar um caminho independente e crítico a partir de suas circunstâncias. Lógico, o pesquisador não é um Hércules que deve sozinho dá conta de tudo e é dever de uma sociedade que se quer minimamente democrática garantir altos níveis de *accountability* (seja horizontal, seja vertical) e isso, envolve, naturalmente, o acesso livre aos dados, registros e documentos governamentais (WEINGART, 1999). Ainda assim, é preciso reconhecer que, mesmo não estando em um cenário ideal, há no Brasil a possibilidade de encontrar diversos tipos de dados que são minimamente trabalháveis.

Por exemplo, há uma ampla gama de indicadores clássicos que podem ser úteis a determinados objetos de pesquisas provenientes de variáveis geográficas, demográficas, econômicas, discursivas, orçamentárias, estatísticas, etc. É possível saber o número de viagens presidenciais,

⁴ Citação original: “(1) theorizing that pays too little attention to testing; and (2) empirical tests that pay too little attention to theory”.

de pronunciamentos, obter índices de capacidades nacionais (como calculado pelo projeto Correlates of War (COW)), redes e tendências de pesquisas, tipos de projetos de lei em trâmite no Congresso Nacional, grau de interesse em RI, bem como a opinião pública (via *surveys* como Latinobarómetro, TRIP Around the World) acerca de determinados temas.

Enfim, há uma lista impossível de ser exaurida de possibilidades. Nesse aspecto, destacamos duas referências iniciais de obtenção de dados prontos para utilização: O Atlas da Política Externa Brasileira (MILANI *et al.*, 2015) e o livro Estatísticas para o estudo das Relações Internacionais (FUNAG, 2016). Essas duas referências trazem informações já produzidas que são úteis para um grandioso número de categorias temáticas⁷.

Para além dessas possibilidades de indicadores mais tradicionais à Política Externa Brasileira, existe um ascendente cenário ligado a produção e consumo de dados virtuais que são, na realidade, frutos direto da interação entre usuários em redes sociais que optam por se posicionar em relação a determinados temas, seja deixando comentários em postagens diversas, seja compartilhando notícias e/ou artigos de opinião, seja interagindo por meio de reações que demonstrem sentimentos negativos ou positivos (curtidas e descurtidas), etc. Fato é que esses usuários produzem um número extravagante de dados (*Big Data*) que ainda são, por sua vez, pouco explorados entre os cientistas sociais, em geral, e os analistas de política externa brasileira, em sentido estrito (MAGER, 2019).

Em resumo, *Big Data* pode ser definido como sendo “um termo que descreve grandes volumes de dados variáveis, complexos e de alta velocidade que requerem técnicas e tecnologias avançadas para permitir a captura, o armazenamento, a distribuição, o gerenciamento e a análise de informações”⁸ (TechAmerica Foundation’s Federal Big Data Commission, 2012 apud GANDOMI; HAIDER, 2015, p. 138). Nesse aspecto, *Big Data* são dados

cujas características prevaletentes podem ser resumidas em 3Vs: volume, variedade e velocidade. Quanto ao volume, chama atenção a magnitude assombrosa de dados que podem ser coletados. No caso da variedade, o grande destaque está na alta heterogeneidade nas estruturas dos bancos de dados obtidos. Por fim, o fator velocidade se refere à rápida criação e proliferação de novos indicadores, fato que, por sua vez, requer habilidades do pesquisador em acompanhar, de modo constante, as tendências de seu objeto de pesquisa que podem mudar em questões de dias, minutos ou até mesmo segundos (GANDOMI; HAIDER, 2015).

Desse modo, hoje é possível coletar milhares de dados do Twitter, do Facebook, do Youtube, do Spotify, por exemplo, e utilizá-los de formas criativas na investigação social (KAPLAN; HAENLEIN, 2010; NGAI *et al.*, 2015). Assim, é possível realizar análises textuais do que os usuários dizem, seja de forma quantitativa (Mineração Textual, Análise de Conteúdo, Análise de Sentimentos), seja de forma qualitativa (Análise de Discurso, identificação de metáforas, analogias, figuras de linguagem, etc), seja de forma mista (via métodos de triangulação e até mesmo aprendizado de máquina (machine learning) supervisionado ou não supervisionado).

Pode-se, também, coletar e descobrir padrões interativos de redes sociais, como *clusters* de usuários, de perfis, de temas, de locais, de posições político-ideológicas. Existe, naturalmente, uma curva de aprendizado que pode, em um primeiro momento, afastar alguns interessados em participar dessa nova forma de coleta de dados. Levando isso em consideração, o presente artigo busca demonstrar, por meio de uma plataforma fácil, livre e intuitiva, como buscar, monitorar, coletar e analisar dados temporais que demonstrem tendências e padrões de usuários em relação à temas específicos de PEB e RI.

Nesse aspecto, o Google Trends possui características e particularidades específicas que precisam ser objeto de

reflexão e crítica do pesquisador. A plataforma em si possui uma rotina de coleta dos dados muito fácil e intuitiva. Apesar disso ser, sem sombra de dúvidas, um aspecto muito positivo que faz com que a plataforma possua uma curva de aprendizado rápida e, de certa forma, até lúdica, é preciso ter em mente que os dados obtidos possuem seus pressupostos, limites e modos específicos de produção. Então, para evitar que os analistas interessados usem os dados de forma acrítica e equivocada, a próxima seção reúne as observações contidas em algumas publicações sobre quais são e como lidar com os limites dos dados de tendências contabilizados e fornecidos pelo Google. Apesar de não ser uma lista exhaustiva (já que depende muito também do próprio objeto de pesquisa trabalhado), inserimos vários problemas comuns a boa parte dos desenhos de pesquisas sociais.

3. Usando o 'google trends' para estudar a peb: questões teóricas

Conforme já reiteramos, a revolução informacional e a difusa e enorme quantidade de dados sendo produzidos diariamente em instâncias virtuais faz com que haja um terreno profícuo e inexplorado para os cientistas sociais, em geral, e os analistas de PEB, em específico. A obtenção e análise de indicadores nessa seara é de facilidade moderada.

Por um lado, o acesso, geralmente, feito via API (Interface de Programação de Aplicações) cria um sistema padronizado e parametrizado de coleta que assegura, em poucos passos, uma rotina de pesquisa rápida e segura, com informações que tendem a ser bem estruturadas em forma de banco de dados. Por outro lado, as melhores oportunidades de trabalhar com os arquivos obtidos ainda dependem de pacotes e de algoritmos computacionais que são mais bem aperfeiçoados em certas linguagens de programação (com destaque para

Python e R) que requerem um maior tempo e exercício de dedicação para o aprendizado, algo que, por sua vez, traz valorosos retornos ao pesquisador.

Não obstante, o GT (e algumas outras plataformas específicas) consegue facilitar a coleta de dados sociais ao fornecer uma plataforma interativa para a extração dos arquivos. Porém, antes de discutir como o pesquisador pode realizar suas análises, há certos limites e cuidados a serem levados em consideração. Para tanto, primeiro, convém demonstrar como os dados do GT estão sendo utilizados em publicações acadêmicas.

Em geral, os dados obtidos através do GT já são extensivamente utilizados e publicados em revistas de alto impacto científico nas mais diversas áreas. Desse modo, Yu *et al.* (2019) utiliza a plataforma para prever tendências em relação ao consumo de petróleo e gasolina. Nghiem *et al.* (2016) usam o GT para medir o interesse público em relação à tópicos de conservação ambiental. Por sua vez, Seifter *et al.* (2010) usam os dados dessa plataforma para realizar pesquisas epidemiológicas.

De fato, as áreas da ciência médica e da tecnologia de informações são as que mais utilizam dados provenientes desse ambiente do Google. Schootman *et al.* (2015), por exemplo, debatem sobre a utilidade dessa plataforma para descobrir graus de interesse em relação à exames de câncer Nuti *et al.* (2014) fazem um artigo de revisão sistemática das publicações que usam os indicadores de tendências do Google para pesquisas de serviços de saúde (*health care*) e (JUN e colab., 2018) fazem algo semelhante em relação a tais dados na ciência como um todo.

A ciência política também não está alheia às vantagens fornecidas por essa ferramenta. Mellon (2013, 2014) tem se dedicado a discutir quando os dados do GT são alternativas boas e ruins para pesquisas interessadas em medir questões de saliência (*issue salience*) político-eleitorais.

Semelhantemente, Connor *et al.* (2019) propuseram utilizar, de forma complementar, dados de tendências do Google para medir tópicos relacionados à desigualdade de renda e viés racial nos EUA. Ademais, Walker (2019) utilizou dessa mesma natureza de dados como *proxy* para interesse em relação à liderança política.

Como se pode perceber, a inserção do GT nas ciências parece ser algo cada vez mais aceito. No entanto, ao empregar tais indicadores, cada um desses pesquisadores supracitados tem clareza acerca da natureza dos dados e fazem ressalvas quanto ao uso dos mesmos em seus objetos de pesquisa. Nesse sentido, embora isso seja algo particular à área em que pesquisadoras e pesquisadores estão inseridas e inseridos, há uma série de ressalvas e limites em comum que devem ser destacados.

Introduzido em 2006 e reformulado em 2012, o GT é hoje a maior plataforma de dados sobre tendências dos usuários em pesquisas na internet; estimando-se haver 5,5 bilhões de consultas diárias e 167 bilhões mensais (JUN; YOO; CHOI, 2018). Ele é capaz de fornecer os dados mais recentes sobre as tendências do tópico pesquisado e uma série temporal iniciando em 2004 (JUN; YOO; CHOI, 2018; NGHIEM *et al.*, 2016). Vale mencionar que, em certos casos, o ano de 2004 pode conter dados de baixa qualidade, ficando à cargo do pesquisador analisar e decidir se vale a pena ou não retirá-lo da amostra (MELLON, 2013).

O primeiro e principal ponto a se destacar em relação aos dados obtidos pelo GT está relacionado com a sua escala de construção. Os números são convertidos em uma escala que varia entre 0 e 100. Em resumo, não temos acesso aos valores absolutos das pesquisas, uma vez que eles são transformados para valores relativos (algo semelhante à escala percentual). Assim, quanto mais pessoas pesquisaram determinado tópico, mais próximo do 100 a observação fornecida será. E quanto menos consultas o tópico obteve, mais próximo do 0

será o valor apresentado (MELLON, 2013, 2014; NUTI *et al.*, 2014; SCHOOTMAN *et al.*, 2015; SEIFTER *et al.*, 2010; WALKER, 2019; YU *et al.*, 2019).

Convém ressaltar que o valor zero não significa necessariamente a ausência absoluta de interesse, mas um grau insignificativo de pesquisas se comparado aos valores maiores. Com efeito, essa escala pode trazer sérios problemas e complicações, uma vez que possui relativa carga de pressupostos matemáticos que precisam ser lembrados em caso dos dados serem aplicados em modelagem de testes estatísticos mais complexos (SEIFTER *et al.*, 2010).

Um segundo ponto de atenção é lembrar que os dados coletados são frutos de buscas de milhares de usuários e não feitos “sob encomenda” para utilização científica. Por isso, a sua natureza possui forte carga ambígua e um mesmo tópico pode conter uma dezena de significados que, devem, por sua vez, serem analisados e, se for o caso, retirados da fase de coletas (MELLON, 2013; NGHIEM *et al.*, 2016). Uma alternativa (não livre de problemas) é utilizar lógicas booleanas ao pesquisar o objeto de interesse na plataforma. Assim, é possível pesquisar pelo tópico Mercosul e Unasul sem incluir referências ao Parlasul do seguinte modo: Mercosul + Unasul – Parlasul.

O terceiro ponto crítico a se ter em mente está intrincado com a questão supramencionada e se refere ao alcance da representatividade dos dados coletados. Primeiro, é necessário entender que existem diferenças sociais, econômicas e culturais nos usos da internet dentro e fora dos diferentes países (MELLON, 2013). Nem todos têm acesso frequente à internet e um mesmo termo de pesquisa pode apresentar diferentes significados entre os usuários.

Desse modo, é necessário agir com cautela ao tentar fazer generalizações com base apenas na amostra disponibilizada pelo Google, já que a mesma pode não ser

representativa da população que está sendo pesquisada (SEIFTER *et al.*, 2010). Nesse sentido, generalizações podem ser feitas se houver outras fontes de dados que arregimentem os indicadores de tendência do Google. Em caso negativo, convém especificar que a pesquisa utiliza dados de forma exploratória ou com o intuito de gerar hipóteses futuras (GERRING, 2012; NGHIEM *et al.*, 2016).

Há também determinados pontos específicos que são próprios à natureza estatística dos dados coletados. Nesse aspecto, estudar um pouco sobre séries temporais, a questão da sazonalidade e da não-estacionariedade é importante para caso os dados sejam aplicados em testes estatísticos de correlação e/ou de causalção específicos (MELLON, 2014).

Por último e, talvez, a principal questão que precisa ser refletida pelo pesquisador diz respeito à necessidade dele buscar encontrar formas de validar os dados obtidos. Para tanto, é necessário tentar encontrar outras fontes de informações e, assim, fazer cruzamentos analíticos. Ademais, buscar realizar análises qualitativas, revisões históricas, reflexões teóricas. Acerca dessa questão, recomendamos a leitura do artigo do Mellon (2014). Nele, há uma seção detalhada com um *framework* pormenorizado de como garantir dados válidos e confiáveis. Cabe ressaltar, porém, que a validação dos dados depende também da criatividade do pesquisador e, acima de tudo, da disponibilização pública do trabalho para que, assim, seja possível garantir um grau adequado de reprodução e replicação do saber científico (JANZ, 2015; PARANHOS *et al.*, 2014).

Para assegurar níveis adequados de replicação, é preciso registrar e fornecer, do modo mais transparente possível, todo o tracejado por trás da obtenção dos resultados obtidos (KING *et al.*, 1994). Nesse sentido, o GT também facilita o trabalho do pesquisador, uma vez que os arquivos baixados por ele apresentam duas

características ideais. Primeiro, cada tipo de informação está reunido em um único arquivo de banco de dados. A priori, isso aumenta o nível de facilidade de replicação de informações específicas, já que cada uma delas foi disponibilizada de forma individual por um arquivo em formato CSV.

Segundo, cada um dos bancos de dados fornecidos pelo Google apresentam, em si, informações claras sobre o termo pesquisado, a categoria selecionada, o país e o intervalo temporal abrangido nos dados. Por essa razão, caso o analista decida modificar o banco original, recomendamos que os arquivos do Google sejam disponibilizados inalterados em conjunto com a versão reestruturada. Por fim, há dezenas de plataformas públicas e livres para o fornecimento dos dados da pesquisa, de todo modo, sugerimos verificar o Consórcio de Informações Sociais (CIS) da USP⁹ e o Projeto Dataverse¹⁰.

Haja vista, esperamos ter chamado a atenção do leitor para a necessidade de empregar dados do Google Trend de forma cautelosa. Como em toda e qualquer espécie de dados, há limites, pressupostos teóricos e matemáticos, desvantagens, e pontos a serem refletidos de forma crítica pelo pesquisador. Urge que se saiba o que se está utilizando e para que propósito, de modo que seja assegurada a validade e a confiabilidade da pesquisa realizada.

No caso do GT, é primordial saber as implicações da escala relativa para os testes estatísticos, os atributos por trás da natureza dos dados coletados (qual o perfil do usuário brasileiro de internet?), e a relevância em incorporar, na medida do possível, um desenho de pesquisa misto que triangule e compare dados de fontes diversas sobre o objeto de interesse. Então, tendo tomado ciência das principais questões e implicações teóricas por trás da utilização dos indicadores de tendência do Google, a

9 Link de acesso: <http://www.nadd.prp.usp.br/cis/>

10 Link de acesso: <https://dataverse.harvard.edu/>

próxima seção traz detalhadamente as formas de coleta de dados e exemplos que demonstram a aplicabilidade dos mesmos nas pesquisas de PEB.

4. Usando o 'google trends' para estudar a peb: exemplo prático

Cientes de quais são os limites, desafios e oportunidades ao utilizarmos uma ferramenta como o GT para analisar as Relações Internacionais, como um todo, e a Política Externa Brasileira, em particular, podemos partir para um exemplo de pesquisa na prática. Nesse aspecto, imaginemos que estamos interessados em explorar e descrever quais são as particularidades sociais, entre Brasil e Argentina, quanto ao interesse em relação ao Mercosul.

Como na maior parte das investigações acadêmicas, é necessário esquematizar um desenho de pesquisa que possua relevância científica, apresente um problema abordável, objetivos claros e forte revisão da literatura (KING *et al.*, 1994). Satisfeitas essas condições, é necessário refletir sobre modos de operacionalização. No nosso caso, desejamos saber quais semelhanças e diferenças entre Brasil e Argentina quanto ao tema do Mercosul. Como se nota, as unidades de análise seriam ambos os países e o objeto de estudo o Mercosul. Usaremos, então, esses atributos como parâmetros iniciais da coleta de dados. Vejamos como fazer isso no GT¹¹. Para tanto, é recomendável analisar as telas demonstradas por meio da figura 1.

Ao iniciar a página do GT será visto algo muito próximo à imagem mostrada no quadrante A (parte superior e esquerda) da figura 1. Na barra central de pesquisa digitamos o nosso termo de interesse, a palavra Mercosul. Como se percebe, a plataforma sugere uma série de possibilidades categóricas relacionadas

com essa busca. Após explorar todas as sugestões, selecionamos a opção que indica que estamos buscando coletar dados para o assunto “Mercado Comum do Sul – Bloco comercial”. Perceba, também, que estamos, primeiro, interessados nos resultados para o Brasil. Caso fosse diferente, poderíamos trocar o país através da caixa de seleção superior, próxima a tela de login do usuário.

Tendo selecionado os critérios iniciais da coleta de dados, podemos seguir para a próxima tela, demonstrada na figura 1 através do quadrante B (parte superior e direita). Aqui, é possível visualizar os principais resultados da busca, modificar parâmetros da mesma e coletar, de fato, os dados. Convém lembrar que essa fase é, de longe, a mais importante para se obter uma pesquisa com alto nível de validade e de representatividade dos dados.

Portanto, sugerimos que, antes de salvar os dados já disponíveis, o pesquisador interessado dedique uma parte considerável de seu tempo explorando as possibilidades dessa tela, verificando possíveis fatores *confounders* e refletindo se o que está sendo mostrado representa, de fato, o que se busca medir. Primeiramente, caso seja de interesse comparar países, já se torna possível, nessa mesma tela, inserir uma unidade de comparação. Basta lembrar de inserir o respectivo termo no idioma oficial do país desejado.

Na verdade, o pesquisador pode optar por realizar buscas de cada país individualmente e, depois, agregar os bancos de dados obtidos ou já inserir um segundo país para comparação. Não há melhor abordagem aqui, dependendo mais das preferências pessoais do analista e da quantidade de países a serem comparados (sendo possível, no máximo, cinco). Para facilitar a visualização das telas da plataforma, usamos a primeira abordagem. Sendo que os mesmos passos mostrados para obter dados do Brasil aqui foram aplica-

dos para a coleta da Argentina.

Como mencionado anteriormente, o quadrante B é a principal tela de coleta de dados. Aqui, é possível escolher o país dos quais os dados serão obtidos e até mesmo delimitar uma região específica dentro do país. Há também uma caixa de seleção que nos permite escolher o intervalo temporal dos dados, indo de 2004 até o momento atual da pesquisa. Nesse sentido, os dados são agregados de forma semanal. Pode-se, inclusive, selecionar uma categoria específica, algo fundamental para termos ambíguos e densamente associados com os mais diversos assuntos. Por fim, é também possível definir se o pesquisador deseja incluir resultados gerais ou pesquisas direcionadas com um meio específico, como imagens, notícias e até padrão de buscas no *Youtube*.

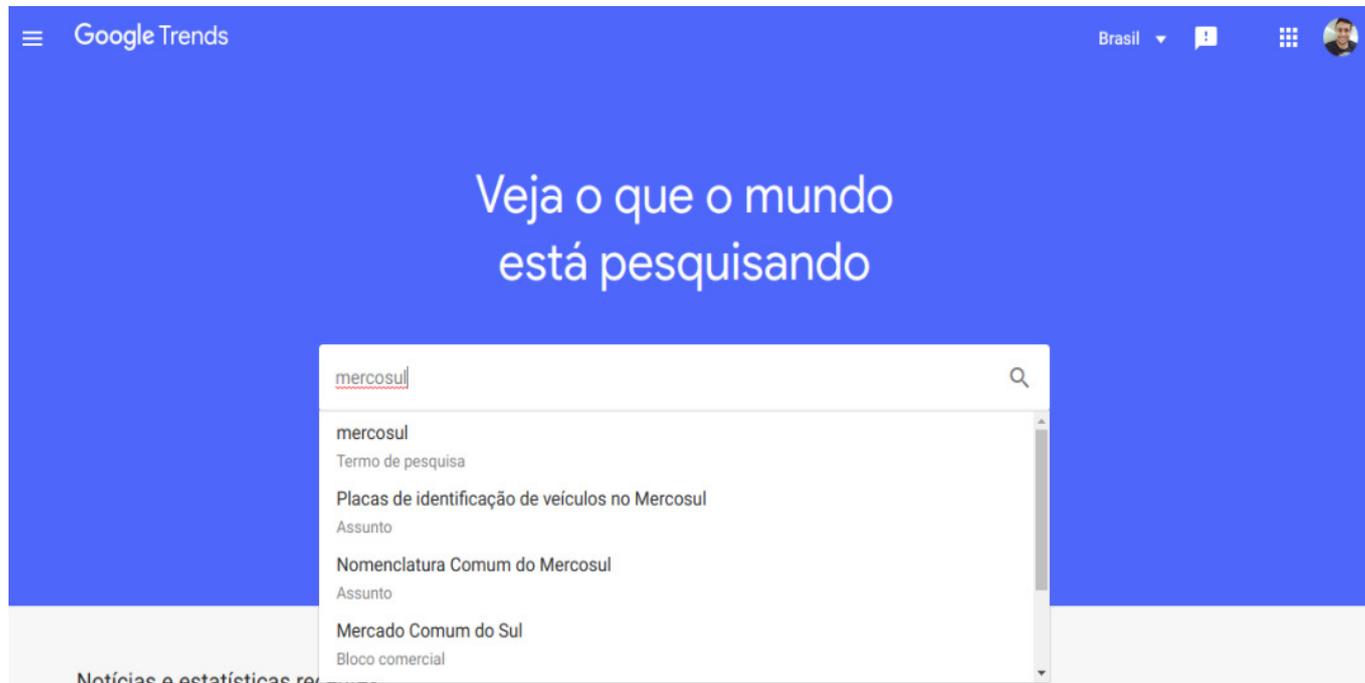
Definido cada um desses critérios, a plataforma já fornece um gráfico de evolução temporal. Provém daqui o primeiro banco de dados. Para obtê-lo, basta apenas pressionar o botão que equivale a uma seta para baixo. Em instantes, estará disponível no computador um arquivo no formato CSV com os indicadores desejados. Mais abaixo da tela, estão resultados por sub-região (quadrante C – esquerda inferior). Com eles, é possível descobrir quais unidades geográficas procuraram mais e menos o assunto da pesquisa. Esse é, a propósito, um ótimo indicador de quais regiões se interessam mais ou menos por assuntos relacionados à PEB e RI. A coleta dos dados se dá de forma semelhante à anterior.

Por fim, conforme indicado pelo quadrante D (direita inferior), ainda é possível obter quais assuntos estavam mais correlacionados com o termo da pesquisa e também quais consultas os usuários realizaram a partir do tópico de interesse inicial. Apesar de sutil, a diferença é importante. No primeiro caso, podemos obter os assuntos que estavam mais associados com

o termo buscado, ao passo que, no segundo, podemos verificar quais foram as consultas feitas a partir do tópico inicial. Ambos os resultados, mostram duas tendências: a) os assuntos relacionados que são mais frequentes; e, b) os assuntos com uma aumento de interesse repentino. Dados que, por sua vez, são fundamentais para os interessados em determinar padrões de estabilidade e instabilidade em relação ao termo pesquisado. Novamente, a coleta de dados é feita como as das vezes anteriores, bastando pressionar o botão indicador de seta pra baixo.

Figura 1 – Principais Telas do GT

QUADRANTE A



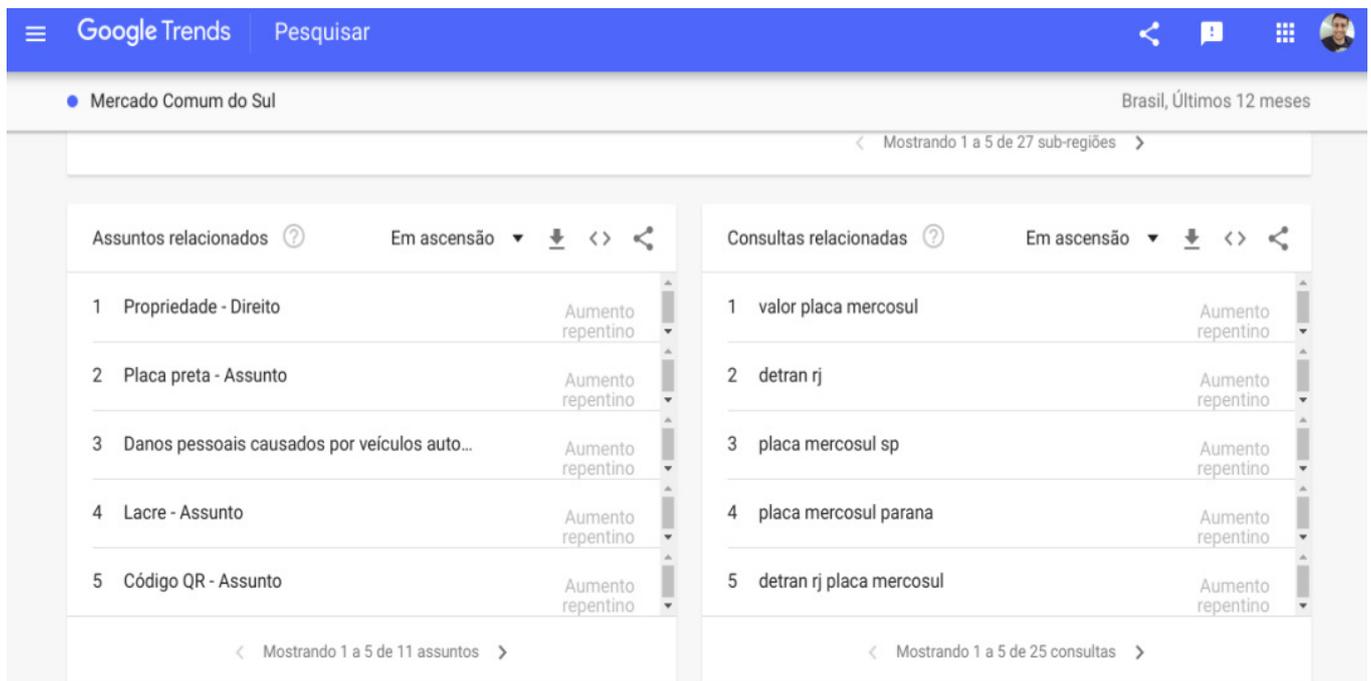
QUADRANTE B



QUADRANTE C



QUADRANTE D



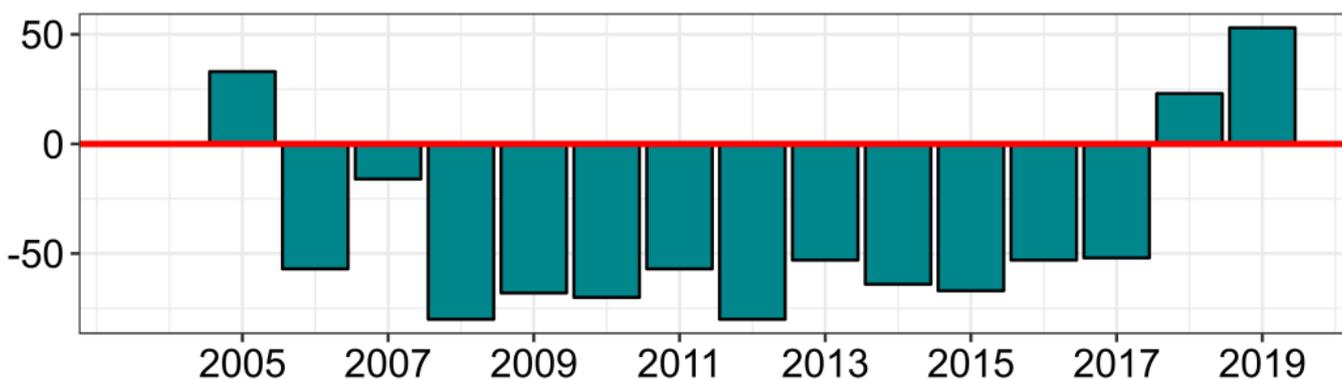
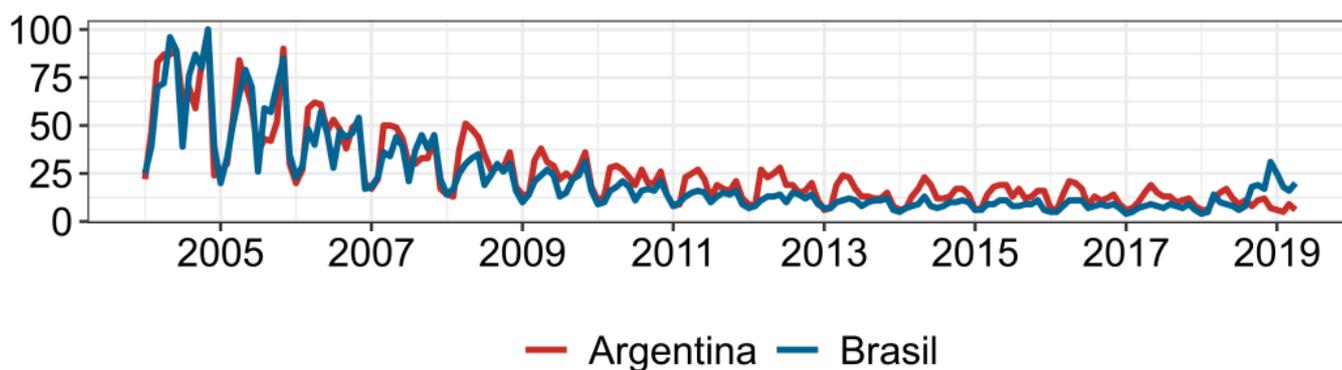
Fonte: Google Trends (<https://www.google.com/trends>).

Os dados de frequência e de avanço repentino dos assuntos são bastante interessantes. Ao pesquisar pelo Mercosul, por exemplo, é de se esperar que outros tópicos e categorias estejam correlacionados com o bloco sul-americano. Nesse sentido, o GT dá conta de mostrar quais são esses termos que estão, de forma constante e densa, associados com o termo da pesquisa, mas vai mais além: fornece também assuntos e tópicos que, embora não comumente associado com a busca, de forma repentina começou a chamar atenção dos usuários por algum motivo. Como se nota, esses indicadores são bem importantes para os investigadores interessados em estudar momentos de crise ou de choques exógenos que possam interferir no objeto e, por consequência, no interesse da sociedade, em geral.

Finalizado o *download* de todos os indicadores de interesse, o próximo passo é realizar a análise dos dados. Esse é um ponto particular que depende mais dos objetivos do desenho do pesquisador e da ferramenta que

prefere utilizar. Em geral, há uma diversidade enorme de instrumentais aqui. Dentre eles, destacamos os gratuitos e de acesso livre, como: LibreOffice Calc, PSPP, R + RStudio, Python + PyCharm.

A partir daqui, as formas de procedimento se multiplicam, uma vez que cada investigador de acordo com seu objeto e objetivos de pesquisa poderá direcionar a utilização dos dados obtidos da forma que melhor convém. Portanto, vejamos quais seriam os resultados que podemos obter ao utilizar os dados baixados na nossa pesquisa simulada sobre especificidades sociais de brasileiros e argentinos em relação ao Mercosul. Os resultados apresentados, a partir de agora, foram gerados graças ao R e ao RStudio com pacotes contidos no Tidyverse. Dessa maneira, primeiramente, vejamos o quanto e quando ambas as sociedades se interessaram por pesquisar pelo Mercosul na internet, ou seja, em quais momentos o bloco sul-americano se tornou mais e menos saliente. Para tanto, segue o gráfico 1:



Fonte: Elaboração do autor com dados do Google Trends.

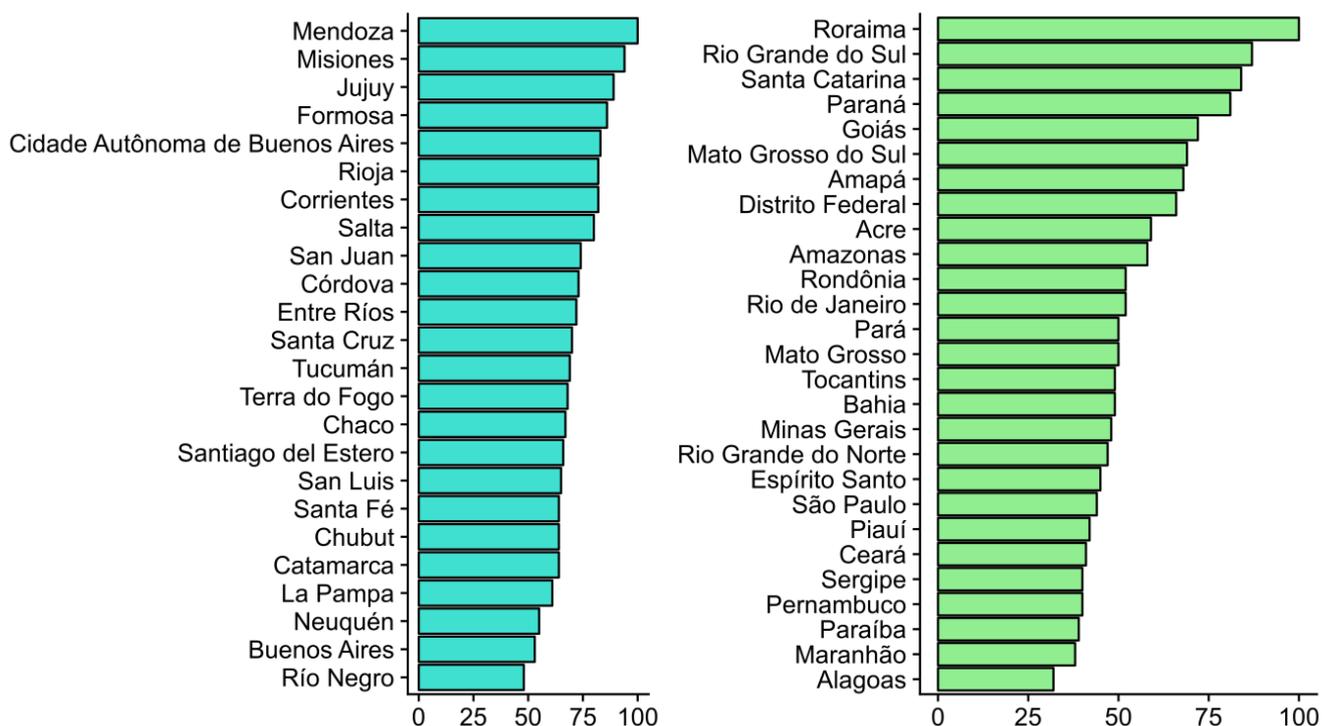
O gráfico 1 condensa duas figuras. Na parte superior, vemos, em escala proporcional específica do GT, duas séries temporais. Em azul, temos os valores para o Brasil e, em vermelho, temos os valores indicativos da Argentina. Os dados aparentam mostrar que ambos os países possuem mais proximidade do que distanciamento ao pesquisarem pelo Mercosul. Em geral, o grau de procura é baixo (Argentina: média (25,6), mediana (19), desvio padrão (19,8); Brasil: média (22,3), mediana (14), desvio padrão (20,3)). Há, porém, períodos de grande pesquisa, com destaque para os anos de 2004 e 2005, em que a escala proporcional variou entre 80 e 100. Outra questão interessante é que as duas séries temporais praticamente seguem juntas, apresentando alto grau associativo.

Considerando que ambas as linhas temporais seguem de modo muito próximo, convém verificar quando elas se equivalem perfeitamente ou se distanciam mais. Isso está informado na parte inferior do gráfico 1. Ele foi construído a partir da seguinte lógica: agregamos através da soma todos os valores para cada ano (sendo

que o valor máximo passa de 100 (na escala do Google) para 1200) e diminuimos os valores do Brasil pelos da Argentina. No gráfico 1, a linha em vermelho demarcada no valor zero do eixo Y simboliza equivalência total de valores. Resultados positivos significam que o Brasil buscou mais do que a Argentina, ao passo que resultados negativos apontam para situação reversa. Em 2004, os valores foram iguais. Com exceção de 2005, 2018 e 2019, todas as barras demonstram que a Argentina tendeu a pesquisar mais pelo Mercosul do que o Brasil. É válido, entretanto, lembrar que essa prioridade é residual, uma vez que a escala resultante vai apenas de -50 a 50, sendo que os valores possíveis de incongruência poderiam variar de -1200 a 1200.

A plataforma não apenas permite coletarmos a saliência agregada do país, mas também a frequência relativa em que regiões (e até mesmo subregiões) buscaram pelo tópico de pesquisa. Nesse aspecto, o gráfico 2 demonstra e compara as regiões brasileiras e argentinas e ordena as que mais e menos se interessaram por pesquisar na plataforma do Google pelo Mercosul.

Gráfico 2: Frequência Proporcional de Buscas pelo Mercosul por Regiões



Fonte: Elaboração do autor com dados do Google Trends.

Em azul, estão as subdivisões geográficas referentes à Argentina, ao passo que, em verde, estão os Estados brasileiros. No primeiro caso, percebemos que Mendoza, Misiones e Jujuy foram os locais que mais pesquisaram sobre o Mercosul, ao passo que Neuquén, Buenos Aires e Rio Negro os que menos procuraram. Em termos descritivos, a média de pesquisas da Argentina foi 72 (mediana de 69,5) e desvio padrão de 12,8.

Quanto ao Brasil, Roraima foi o Estado que mais pesquisou sobre o Mercosul, seguido pelo Rio Grande do Sul, Santa Catarina e Paraná. Em paralelo, Pernambuco, Paraíba, Maranhão e Alagoas foram os que menos apresentaram interesses de pesquisa em relação ao bloco. Em termos descritivos, a média de pesquisas do Brasil foi 55,6 (mediana de 50) e desvio padrão de 17,2). Números que mostram que o Brasil tende a se interessar menos do que a Argentina em pesquisar pelo Mercosul.

Com efeito, o GT permite ir ainda mais além, sendo possível descobrir, para cada uma dessas unidades regionais, quais foram as respectivas subunidades responsáveis por mais ou menos pesquisas em relação ao termo de interesse. Por exemplo, no caso brasileiro, verificamos que, em Alagoas, apenas uma cidade esteve interessada em pesquisar pelo Mercosul que foi a capital Maceió (recebendo, portanto, valor 100 na escala relativa do GT). De modo análogo, em Roraima, a única cidade através da qual partiram buscas pelo bloco sul-americano foi Boa Vista (valor 100), também capital do estado. Por outro lado, no Rio Grande do Sul, o quadro de cidades interessadas é bem mais diversificado. Uruguaiana (100) foi a unidade territorial que mais pontuou em buscas no GT. Em seguida, está Gravataí (72), depois Bento Gonçalves (52) e Santa Cruz do Sul (50). A capital, Porto Alegre, apresentou um valor relativamente baixo de interesse, 43.

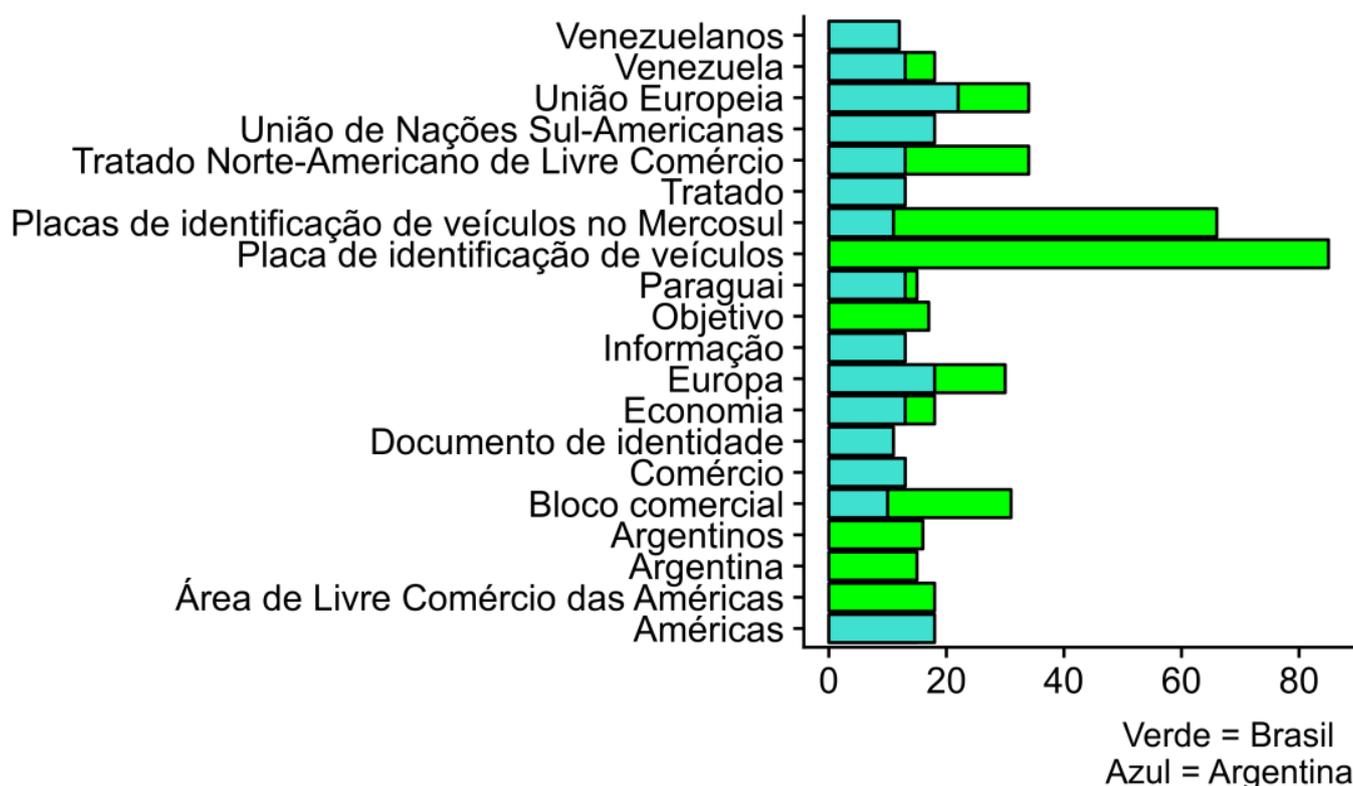
No caso da Argentina, em Mendoza, as cidades mais interessadas em pesquisar sobre o Mercosul foram

Maipú (100), própria Mendoza (100), Las Heras (89), San Martín (81), Godoy Cruz (79) e San Rafael (76). Em Rio Negro, unidade regional mais desinteressada conforme demonstra o gráfico 2, a cidade com um padrão de busca mais alto foi Bariloche (100), também a única cidade interessada dessa região.

Da forma como se apresentam, esses dados podem servir de excelentes indicadores para o pesquisador que está interessado em descrever e/ou explicar a variação geográfica e demográfica por trás do conhecimento ou desconhecimento em relação à assuntos que envolvem a seara internacional. Nesse quesito, o GT serve de plataforma apropriada para que tentemos acumular indicadores sobre quais foram as unidades regionais mais e menos interessadas em temas específicos e quais são as semelhanças e as diferenças que prevalecem em cada uma das unidades. Seria, portanto, mais uma forma de testar se distância geográfica, nível de escolaridade, índice de desenvolvimento humano, índice de desigualdades sociais e de renda (a exemplo do coeficiente de GINI), partidos políticos e tendências ideológicas, número de cursos em relações internacionais nas cidades, entre outras dezenas de variáveis possíveis explicam ou não o maior ou menor interesse de uma determinada região em pesquisar pelo Mercosul na internet.

Para além disso, os indicadores de tendência do Google dão conta de registrar e contabilizar quais foram os assuntos (e também as consultas) que mais estiveram correlacionados com o termo de interesse, no nosso caso o Mercosul. Isso, por sua vez, torna possível saber camadas temáticas de interesse em relação ao bloco. Nesse sentido, segue o gráfico 3.

Gráfico 3: Assuntos Relacionadas com o Mercosul por País



Fonte: Elaboração do autor com dados do Google Trends.

Como se percebe, há assuntos relacionados ao Mercosul que são comuns nas pesquisas tanto da sociedade brasileira (em barras verdes) da sociedade argentina (em barras azuis), por exemplo: Venezuela, tratado Norte-Americano de Livre Comércio (NAFTA), placas de identificação de veículos no Mercosul, Paraguai, Europa e União Europeia, economia e bloco comercial. Simultaneamente, também há assuntos relacionados apenas com as buscas brasileiras, como saber acerca do objetivo do Mercosul, da Argentina e da Área de Livre Comércio das Américas. E, por fim, assuntos relacionados apenas com buscas argentinas, tais quais os venezuelanos, União de Nações Sul-Americanas (UNASUL), Américas, comércio e identidade.

Para além dessa dimensão mais associada com os assuntos relacionados ao Mercosul que são mais procurados, também é possível acrescentar dados sobre as pesquisas relacionadas ao projeto sul-americano que tiveram, seja por qual razão, um crescimento repentino de buscas.

Exemplifiquemos com o caso da sociedade brasileira. Segundo o Google, consultar por placas do Mercosul sofreu um elevado aumento. O motivo, como suspeitamos, está na implementação da lei brasileira em relação a esse tópico. Semelhantemente, procurar pela relação entre a Venezuela e o Mercosul também apresentou um padrão inconsistente com os parâmetros normais. Novamente, o provável motivo está nos novos rumos que a Política Externa Brasileira tem tomado perante a crise cada vez mais aprofundada do regime venezuelano, fato que faz com que assuntos relacionados com a Venezuela se torne uma consulta muito mais discutida, necessária e saliente entre os usuários brasileiros.

Haja vista, esses são alguns poucos exemplos de indicadores possíveis de ser coletados através da plataforma disponibilizada pelo Google. Na verdade, há centenas de possibilidades de análise que dependem diretamente dos objetivos do pesquisador, de seu tema de pesquisa, da natureza das buscas e de sua profundidade analí-

tica. Esperamos, portanto, ter fornecido um primeiro e rápido caminho de acesso¹² para a maior utilização de dados empíricos na pesquisa de Relações Internacionais, em geral, e PEB, em particular.

6. CONSIDERAÇÕES FINAIS

Conforme demonstrado, o artigo buscou cumprir com dois objetivos principais. Primeiro, debater sobre a relevância de incorporarmos dados empíricos nas análises de Política Externa Brasileira. Segundo, comentar sobre vantagens e desvantagens ao se utilizar o Google Trends para coletar dados e inferir determinadas características do objeto em estudo. No primeiro caso, demonstramos que a escassez de dados no subcampo da Política Externa Brasileira tem dupla motivação. Por um lado, não se pode negar que a própria natureza da PEB apresenta diversos constrangimentos e dificuldades ao analista encontrar e incorporar dados que são facilmente disponibilizados em outras áreas do conhecimento. Por outro lado, não se pode negar que há uma importante gama de evidências disponíveis que são subutilizadas.

Como forma de minimizar, ainda que parcialmente, essa subutilização de dados, o segundo objetivo desse artigo foi demonstrar, na prática, como é possível empregar dados do Google Trends em pesquisas sociais e, mais especificamente, na PEB. Como em toda base de dados, antes de coletar e utilizar os indicadores fornecidos, é preciso que o usuário tenha em mente quais são os riscos, pressupostos e benefícios de sua utilização. Portanto, elencamos e comentamos acerca dos principais e só então demonstramos, através da simulação de uma pesquisa sobre o Mercosul na sociedade brasileira e argentina, como utilizar dados dessa base.

Tendo chegado, então, à leitura final desse artigo, esperamos que o leitor tenha ficado, se não satisfeito, ao menos interessado em acessar à plataforma do GT e refletir se ela tem ou não algo a acrescentar em seus trabalhos analíticos. Conforme amplamente discutido, os indicadores coletados dessa ferramenta possuem sérios limites e *trade-offs* que precisam ser levados em consideração. Com efeito, essa não é a única (e talvez não seja a melhor) plataforma de coleta de dados virtuais, nem se propõe a ser isso. No entanto, eles já são satisfatoriamente usados em muitas áreas do fazer científico e são uma boa forma de complementar outros achados empíricos, bem como meios de explorar, descrever e construir novas hipóteses. De todo modo, acreditamos que ela pode servir de porta de acesso inicial ao complicado, mas instigante mundo da análise de dados.

12 Para detalhes mais técnicos e extensos em relação ao funcionamento e aos atributos do Google Trends sugerimos acessar o seguinte link: <https://support.google.com/trends/?hl=pt-BR#topic=6248052>.

REFERÊNCIAS BIBLIOGRÁFICAS

- CAMPOS, Cinthia Regina. (2015), “O labirinto metodológico das Relações Internacionais: dilemas e potenciais saídas”, *Revista Política Hoje*, v. 24, no. 2: 47–72.
- CONNOR, Paul et al. (2019), “Income inequality and white-on-black racial bias in the united states: evidence from project implicit and Google Trends”, *Psychological Science*, v. 30, no. 2: 205–202.
- FARIA, Carlos Aurélio Pimenta de. (2012), “O Itamaraty e a política externa brasileira: do insulamento à busca de coordenação dos atores governamentais e de cooperação com os agentes societários”, *Contexto Internacional*, v. 34, no. 1: 311–355.
- FUNAG. (2016), *Estatísticas para o estudo das Relações Internacionais*. Brasília, Funag.
- GANDOMI, Amir; Haider, Murtaza. (2015), “Beyond the hype: Big data concepts, methods, and analytics”, *International Journal of Information Management*, v. 35, no. 2: 137–144.
- GERRING, John. (2012), “Mere description”, *British Journal of Political Science*, v. 42, no. 4: 721–746.
- JANZ, Nicole. (2015), “Bringing the gold standard into the classroom: replication in university teaching”, *International Studies Perspectives*, v. 17, no. 4: 1–16.
- JUN, Seung-Pyo; YOO, Hyoung Sun; CHOI, San. (2018), “Ten years of research change using Google Trends: from the perspective of big data utilizations and applications”, *Technological Forecasting and Social Change*, v. 130, no. SN: 69–87.
- KAPLAN, Andreas M.; HAENLEIN, Michael. (2010), “Users of the world, unite! The challenges and opportunities of social media”, *Business Horizons*, v. 53, no. 1: 59–68.
- KING, Gary; KEOHANE, Robert O.; VERBA, Sidney. (1994), *Designing social inquiry: scientific inference in qualitative research*. Princeton, Princeton University Press.
- MAGER, Astrid. (2019), “The politics of big data: big data, big brother?”, *Information, Communication & Society*, v. SN, no. SN: p. 1–3.
- MEARSHEIMER, John J.; WALT, Stephen M. (2013), “Leaving theory behind: why simplistic hypothesis testing is bad for International Relations”, *European Journal of International Relations*, v. 19, no. 3: 427–457.
- MEDEIROS, Marcelo de Almeida et al.. (2016), “What does the field of International Relations look like in South America?”, *Revista Brasileira de Política Internacional*, v. 59, no. 1: 1–31.
- MELLON, Jonathan. (2014), “Internet search data and issue salience: the properties of Google Trends as a measure of issue salience”, *Journal of Elections, Public Opinion and Parties*, v. 24, no. 1: 45–72.
- MELLON, Jonathan. (2013), “Where and when can we use Google Trends to measure issue salience?”, *PS: Political Science & Politics*, v. 46, no. 2: 280–290.
-

- MILANI, Carlos et al.. (2015), Atlas da política externa brasileira. Ciudad Autónoma de Buenos Aires/Rio de Janeiro, CLACSO/EDUERJ.
- NGAI, Eric W.T.; TAO, Spencer S.C.; MOON, Karen K.L. (2015), “Social media research: theories, constructs, and conceptual frameworks”, *International Journal of Information Management*, v. 35, no. 1: 33–44.
- NGHIEM, Le T. P. et al.. (2016), “Analysis of the capacity of Google Trends to measure interest in conservation topics and the role of online news”, *PLOS ONE*, v. 11, no. 3: 1–12 .
- NUTI, Sudhakar V. et al.. (2014), “The use of Google Trends in health care research: a systematic review”, *PLOS ONE*, v. 9, no. 10: 1–49.
- PARANHOS, Ranulfo et al.. (2014), “A importância da replicabilidade na Ciência Política: o caso do SIGOBR”, *Revista Política Hoje*, v. 22, no. 2: 213–229.
- SCHOOTMAN, M. et al.. (2015), “The utility of Google Trends data to examine interest in cancer screening”, *BMJ Open*, v. 5, no. 6: 1–8.
- SEIFTER, Ari et al.. (2010), “The utility of “Google Trends” for epidemiological research: Lyme disease as an example”, *Geospatial Health*, v. 4, no. 2: 1-35.
- WAISBICH, Laura Trajber et al.. (2017), “The transparency frontier in Brazilian foreign policy”, *Contexto Internacional*, v. 39, no. 1: 179–200.
- WALKER, Finley W. (2019), Google trends data as a proxy for interest in leadership. 183f. Tese (Doutorado em Educação de Liderança Organizacional) – Faculdade de Educação, Southeastern University.
- WEINGART, Peter. (1999), “Scientific expertise and political accountability: paradoxes of science in politics”, *Science and Public Policy*, v. 26, no. 3: 151–161.
- YU, Lean et al.. (2019), “Online big data-driven oil consumption forecasting with Google Trends”, *International Journal of Forecasting*, v. 35, no. 1: 213–223.
-