



ESCOLA
DE
DADOS

MÓDULO I

ADMIRÁVEL MUNDO NOVO

DO JORNALISMO DE DADOS

Natália Mazotte

Twitter: @nataliamazotte

E-mail: ncortezrj@gmail.com

Roteiro: Módulo I

- O que são dados?
- O que é jornalismo de dados e por que agora?
- O fluxo de trabalho com dados
- Incorporando novas competências à prática jornalística
- Por que investir nisso? Exemplos inspiradores e inovadores
- Ligando o radar para obter dados
 - Busca avançada no Google
 - Usando a Lei de Acesso à Informação
- Perguntas cruciais para compreender os dados



ESCOLA DE **DADOS**

JORNALISMO... DE DADOS?



É uma visualização bonitinha?



Números, muitos números na minha matéria!

O BOM E VELHO JORNALISMO... COM UMA NOVA FONTE

Jornalismo de dados

- Antes de mais nada: Jornalismo
- Dados não são o fim do jornalismo de dados, são o meio para o fim
- Herda os princípios do jornalismo tradicional
- **O diferencial:** Base de dados é mais uma fonte a ser entrevistada

**Dados para prover
contexto e melhor
entendimento...
não é de hoje!**

SPOTLIGHT

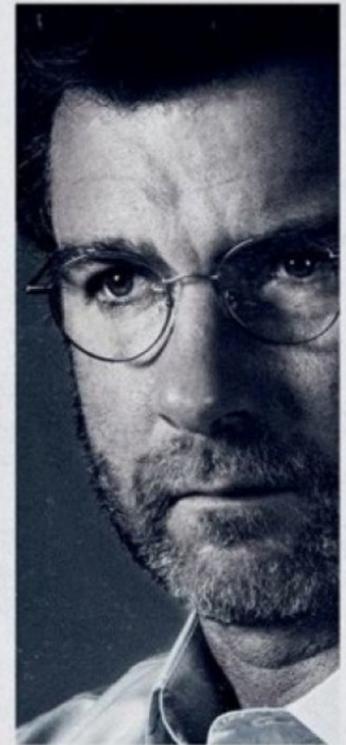
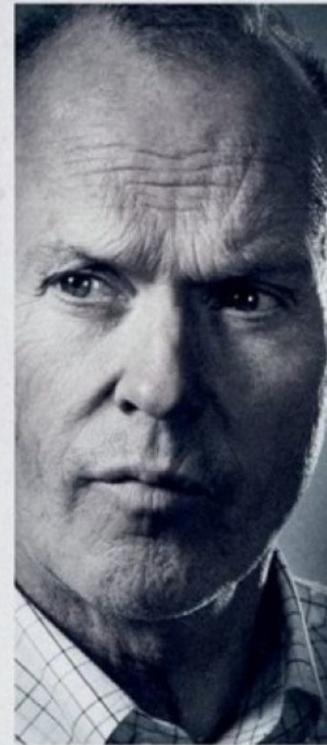
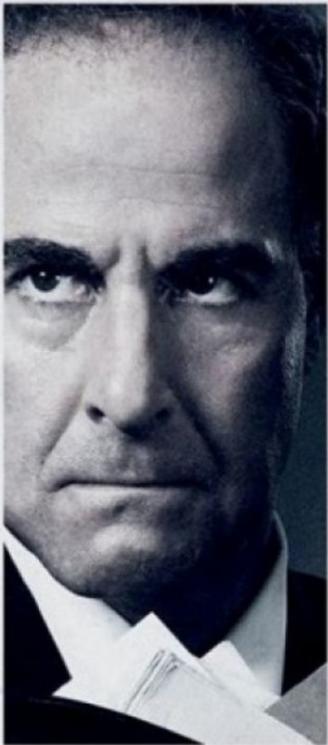
MARK
RUFFALO

MICHAEL
KEATON

RACHEL
McADAMS

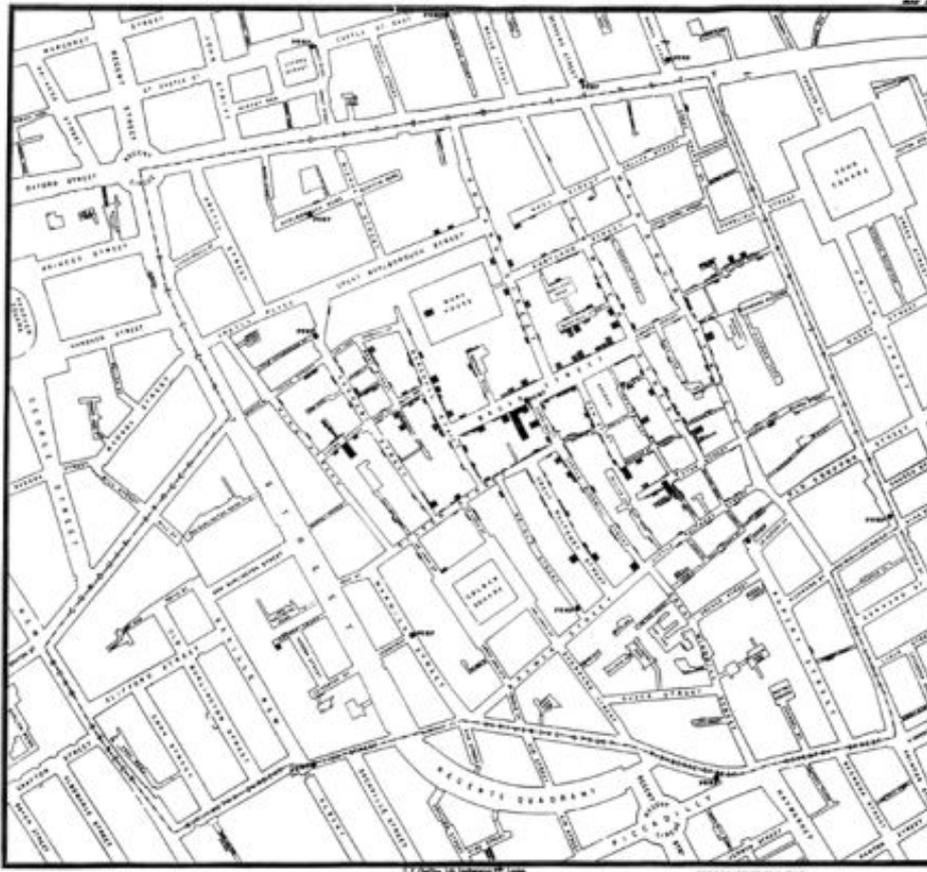
LIEV
SCHREIBER

AND STANLEY
TUCCI



BASED ON THE PULITZER PRIZE-WINNING INVESTIGATION

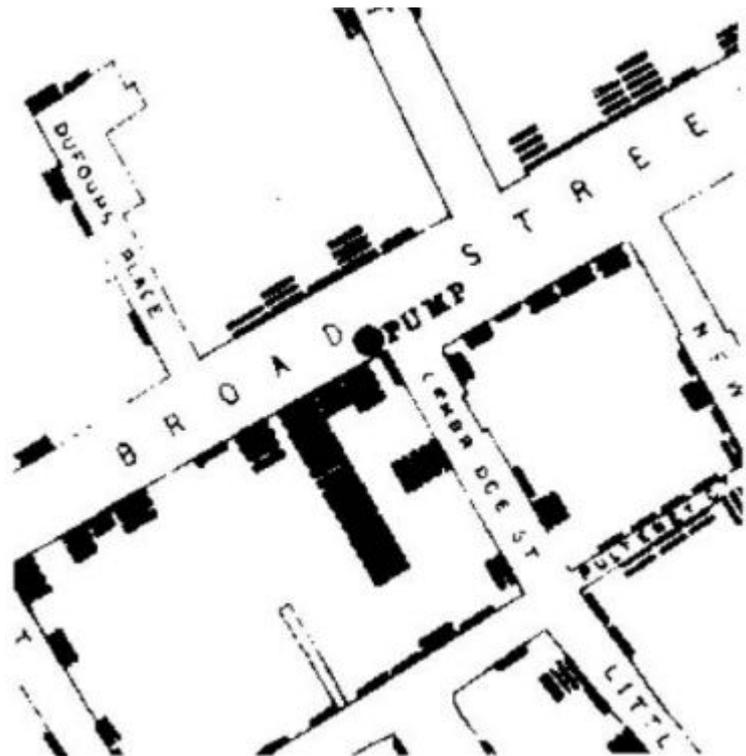
Jornalistas e dados: caso antigo



1854
Londres

Epidemia de
cólera mata
500+ em 10
dias

Mapa da Cólera, John Snow



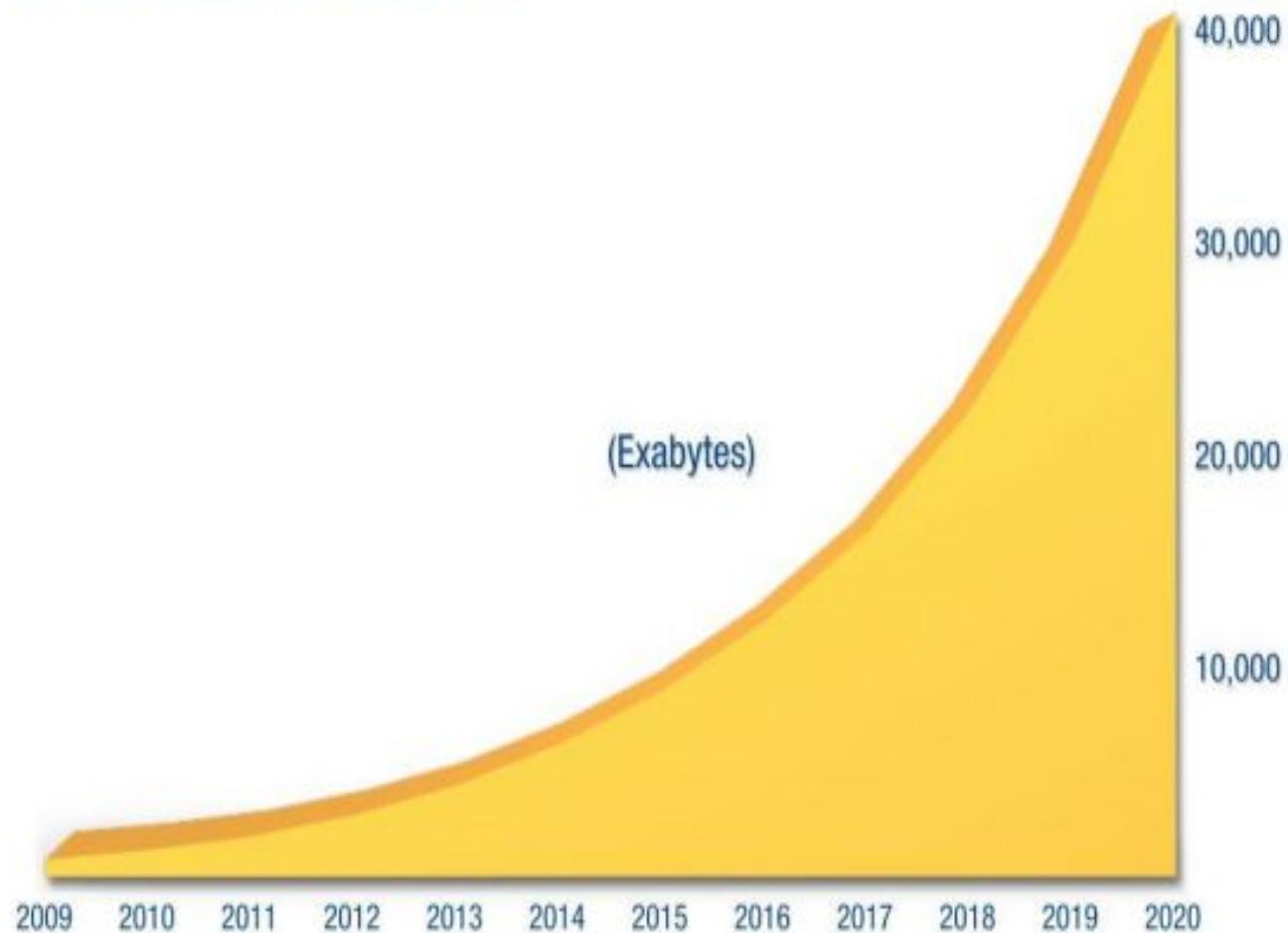
Método empírico - análise de dados!

E agora?



Conta de banco, título de eleitor, imposto de renda, dados eleitorais, sistemas educacionais, troca de mensagens, redes sociais (twitter, facebook, instagram, Snapchat, Vine, Okcupid, Tinder, Secret, Google+, Whatsapp, Telegram), sistemas de pagamentos (cartão de crédito, débito, financiamento), pesquisa de imóveis, compra e venda de mercadorias (Mercado Livre, eBay, Zap Móveis, Submarino), atividades governamentais (diário oficial, contratos, pregões eletrônicos, índices de performance, dados sobre programas sociais), notícias, entretenimento (Youtube, DailyMotion)

The Digital Universe: 50-fold Growth from the Beginning of 2010 to the End of 2020



Source: IDC's Digital Universe Study, sponsored by EMC, December 2012

Contexto: Big Data

BRASIL Acesso à informação Faltam **21** dias para a Copa Participe Serviços Legislação Canais

Ir para o conteúdo **1** Ir para o menu **2** Ir para a busca **3** Ir para o rodapé **4**

Dados Abertos

GOVERNO FEDERAL

Catálogo | Sobre o portal | Perguntas frequentes | Contatos

Pesquisa ...

em 231 conjuntos de dados com 2056 recursos [\(o que é isto?\)](#)

Dados em destaque

Instituições de Ensino Superior

Dados cadastrais das Instituições de Ensino Superior no Brasil.

Agências, Gerências e Superintendências de atendimento do ...

Postos e delegacias do Departamento de Polícia Rodoviária Federal

Informações sobre os postos e delegacias do Departamento de Polícia Rodoviária Federal - DPRF.

+ Acessados + Recentes

Conjunto de dados	Mês	Totais
Lista de Eleitores Filiados aos Partidos Políticos	609	1657
Microdados do Censo Escolar	450	863
Destinação do Lixo Domiciliar	437	1072
Obras do PAC - Programa de Aceleração do Crescimento	422	1187
População Economicamente Ativa por sexo	302	786



Contexto: Dados abertos e acesso à informação



+tableau++public

SQL



CARTO^{DB}

Geospatial database on the cloud

StoryMap ^{JS}

Refine^{OPEN} 


OpenOffice.org™

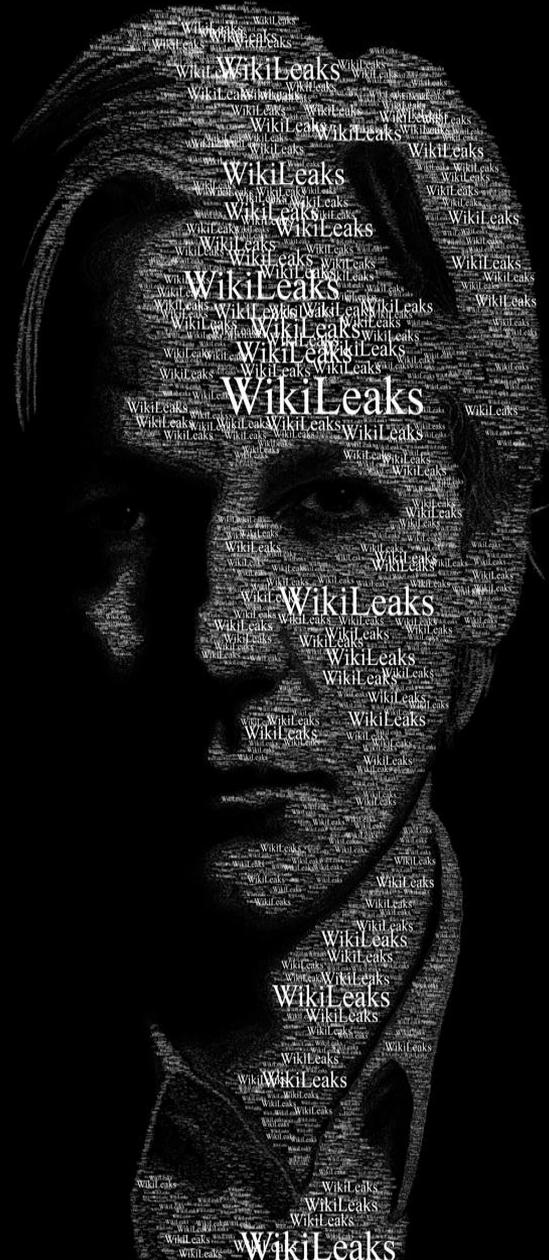
4th edition

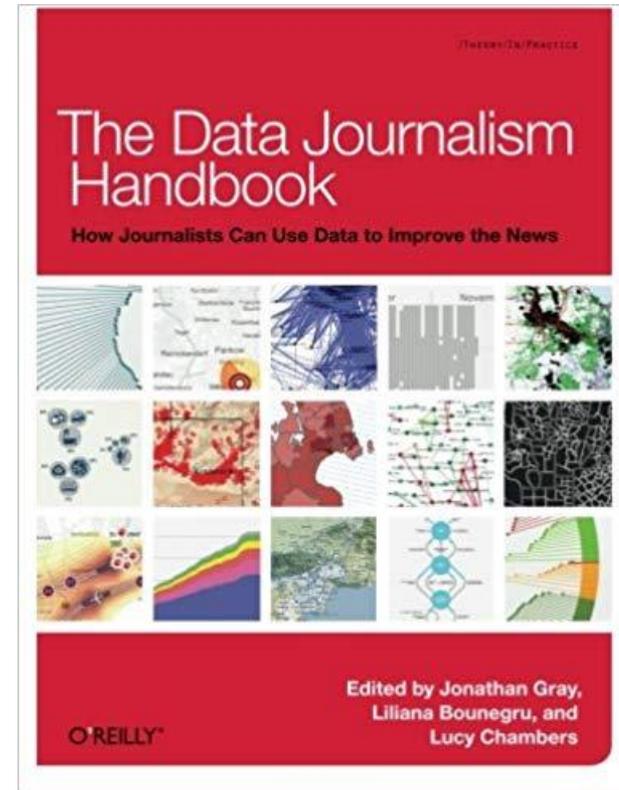
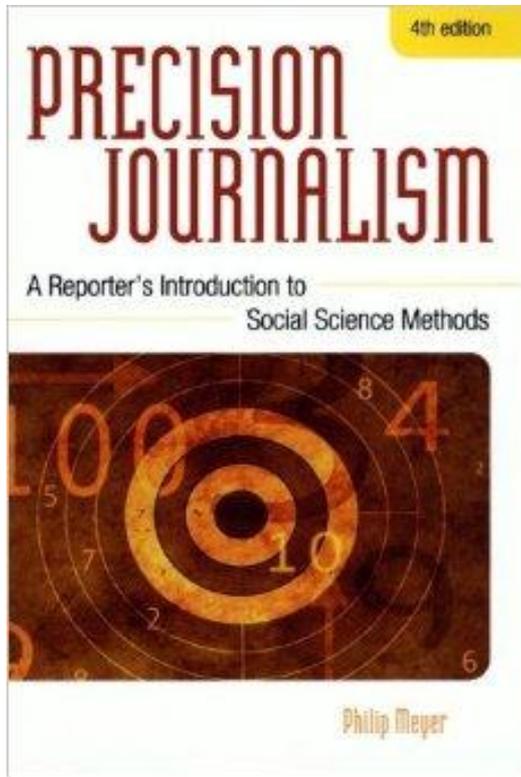
PRECISION JOURNALISM

A Reporter's Introduction to
Social Science Methods



Philip Meyer





{ Stanford
Computational
Journalism Lab }



Jornalismo de dados

Definições iniciais

“Talvez sejam as novas possibilidades que se abrem quando se combina o tradicional "**faro jornalístico**" e a habilidade de contar uma história envolvente com a **escala e o alcance** absolutos da informação digital agora disponível.” (Data Journalism Handbook)

Jornalismo de dados

Definições iniciais

“Quando a informação era escassa, a maior parte de nossos esforços estava voltada a caçar e reunir dados. Agora que a **informação é abundante, processá-la** tornou-se mais importante.” (Philip Meyer)

Jornalismo de dados

Definições iniciais

“É um novo conjunto de competências para buscar, entender e visualizar fontes digitais em um momento em que os **conhecimentos básicos** do jornalismo tradicional já **não são suficientes**. Não se trata da substituição do jornalismo tradicional, mas de um acréscimo a ele” (Data Journalism Handbook)



ESCOLA
DE
DADOS

EXEMPLOS DE JORNALISMO GUIADO POR DADOS

GOVERNO FEDERAL
BRASIL
PAÍS RICO E PAÍS SEM POBREZA

FIES
1 MILHÃO
DE CONTRATOS FIRMADOS.
MAIS ACESSO AO ENSINO SUPERIOR.

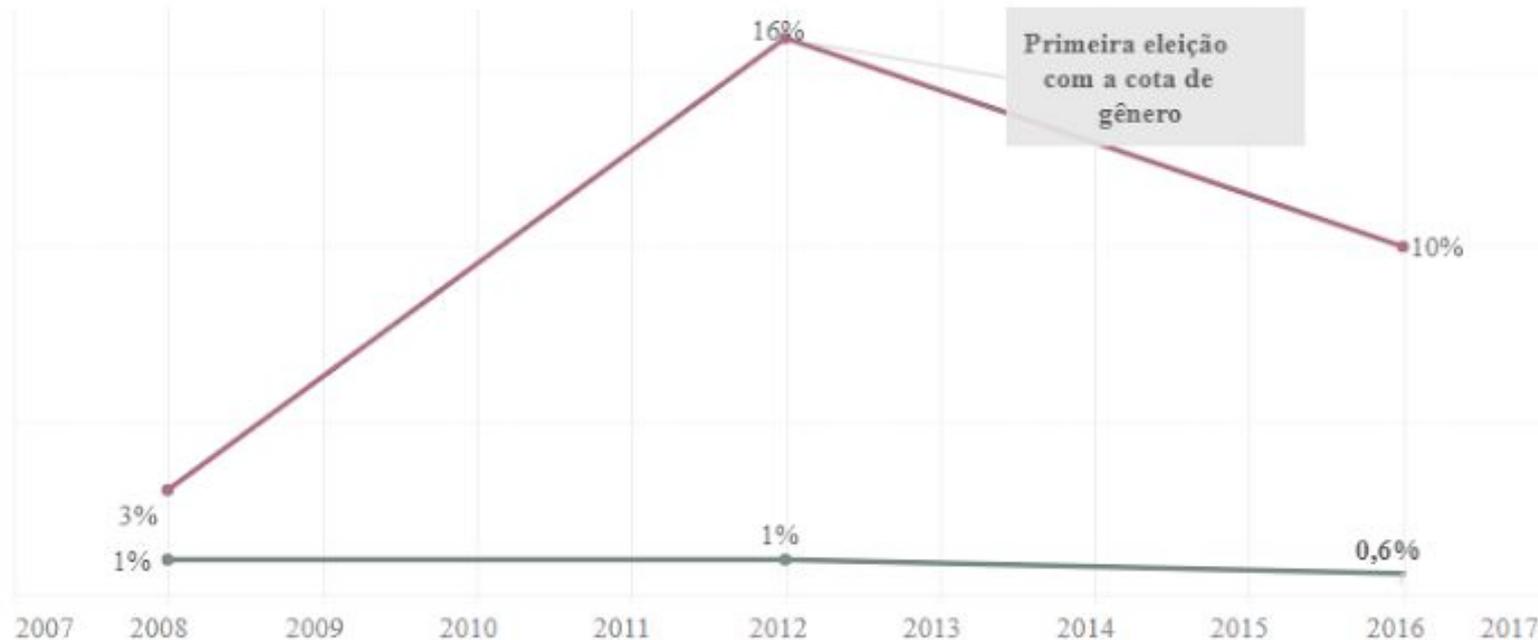
GOVERNO FEDERAL
BRASIL
PAÍS RICO E PAÍS SEM POBREZA

Impacto em políticas públicas a partir de dados abertos

CANDIDATURAS FANTASMAS NAS ELEIÇÕES MUNICIPAIS

Número de candidatas sem votos disparou após adoção da cota de gênero. O mesmo não ocorreu com os homens na disputa.

■ homens
■ mulheres

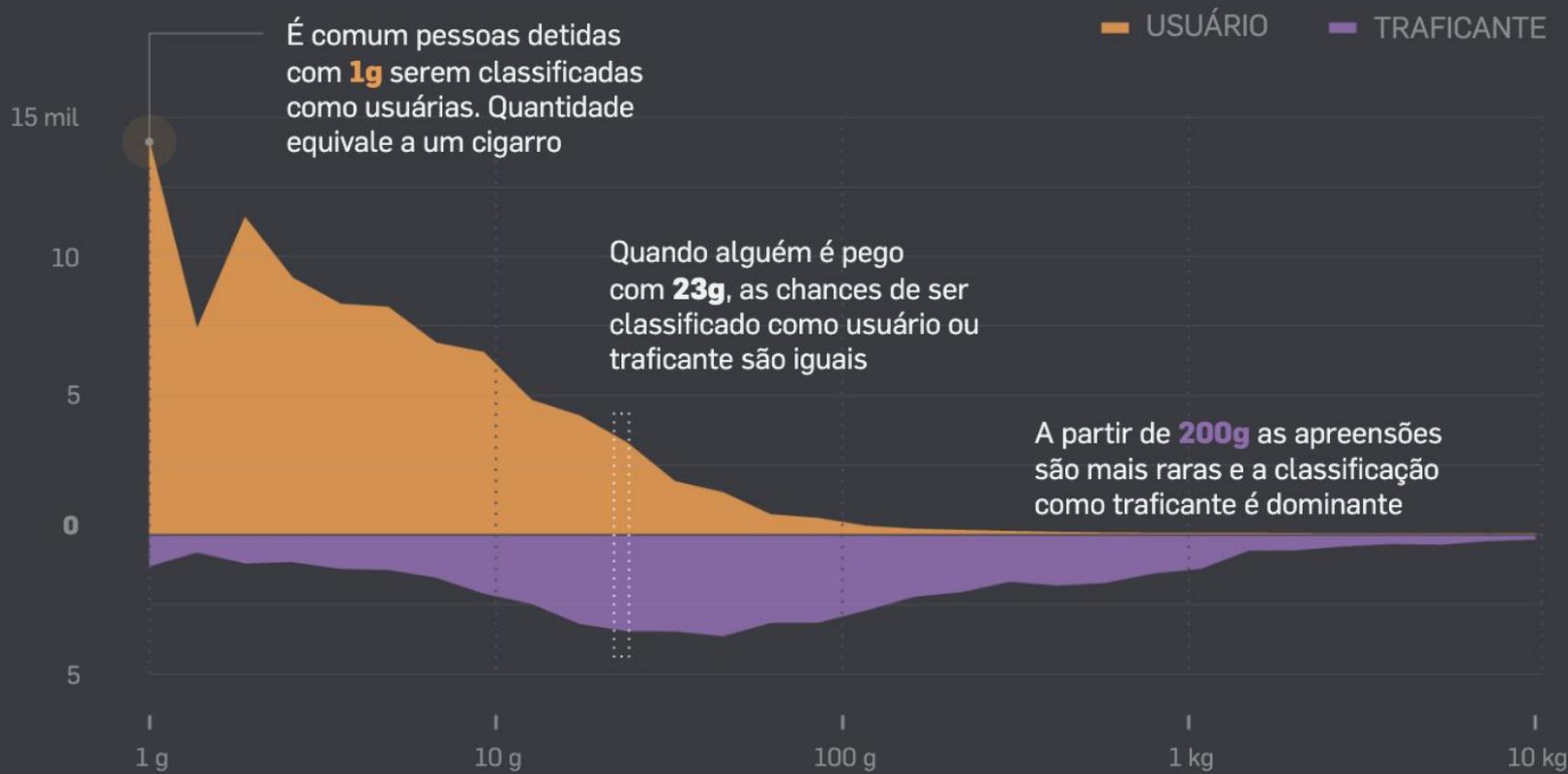


Investigações guiadas por dados



MACONHA

Classificação dos detidos de acordo a quantidade de droga



Dados inéditos para debates públicos

PRESOS PROVISÓRIOS DANOS PERMANENTES

EM 2013 NA CIDADE DO **RIO DE JANEIRO**,
7.734 PESSOAS FORAM PRESAS E MANTIDAS NA PRISÃO,
 EM MÉDIA, POR **101 DIAS** ANTES DO JULGAMENTO.

ESSAS PRISÕES CUSTARAM APROXIMADAMENTE
R\$ 45 MILHÕES AO CONTRIBUINTE.

COM ESSE DINHEIRO, DARIA PARA
 MANTER POR UM ANO **9,9 MIL ALUNOS DO ENSINO BÁSICO**,
 CONSTRUIR **76 POSTOS DE SAÚDE** OU
 CONSTRUIR **873 CASAS POPULARES**.

QUANDO JULGADOS

4.211 (54,4%) ESTIVERAM PRESOS INDEVIDAMENTE

ABSOLVIDOS : **772**

PROCESSOS ARQUIVADOS : **294**

ACORDOS COM A JUSTIÇA : **814**

PENAS ALTERNATIVAS E REGIME ABERTO : **1.273**

PRISÃO EM REGIME SEMIABERTO : **1.058**

1.437 (18,6%) CONDENADOS AO REGIME FECHADO

1.610 (20,8%) AINDA AGUARDAM JULGAMENTO

AGUARDAM PRESOS : **171**

AGUARDAM EM LIBERDADE : **1.439**

476 (6,2%) OUTROS

SAIBA MAIS ?

Pedro

44 anos, casado, foi preso na Rocinha com 16g de droga. Usuário de drogas, deficiente visual e esquizofrênico ficou **485 dias** preso até ter sua doença reconhecida e seu processo extinto em sua primeira audiência em julho de 2014. Sua prisão custou **R\$ 28 mil** ao contribuinte. E para ele quanto custou?

CASOS EMBLEMÁTICOS

VozData | Telegramas PASO

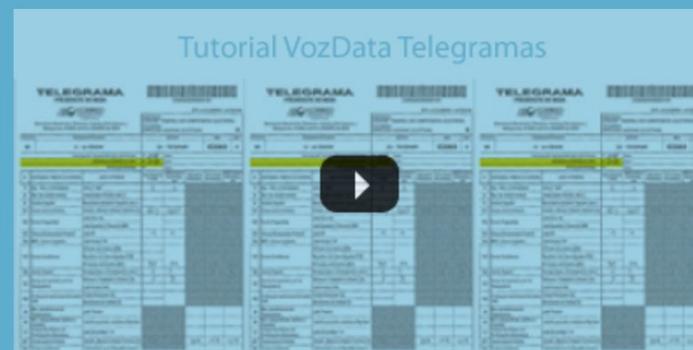
Ayudanos a chequear telegramas de las PASO nacionales y formá parte de esta plataforma colaborativa para detectar inconsistencias. El proceso es rápido y sencillo: elegí una provincia y la herramienta te va a orientar sobre los pasos a seguir.

Elegí la provincia por la que quieres empezar



VideoTutorial

Cómo se usa VozData | Telegramas



Qué buscamos en los telegramas

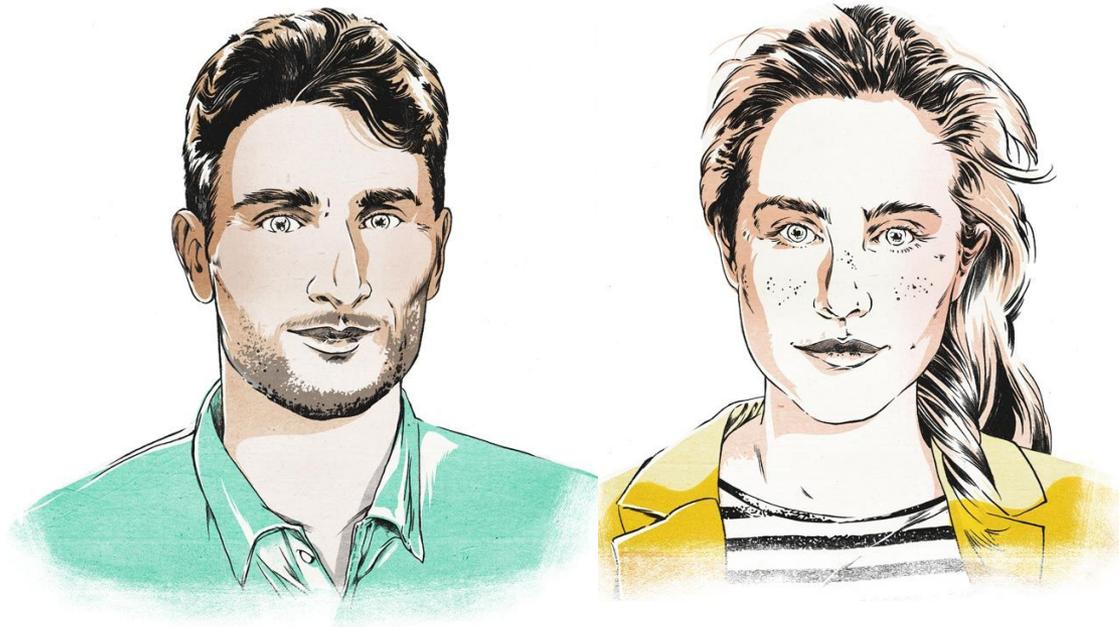


1

Verificar el dato sobre la cantidad de ciudadanos que votaron

Hanna and Ismail –

Racial Discrimination In the Housing Market



Inovação: novas metodologias pra coletar dados inéditos

Publica

Quem são os donos dos ônibus no Rio de Janeiro

Mapeamento dos empresários de ônibus do Rio de Janeiro
Concepção e execução: Adriano Belisário
Checagem: Mariah Queiroz e Luísa Lucciola

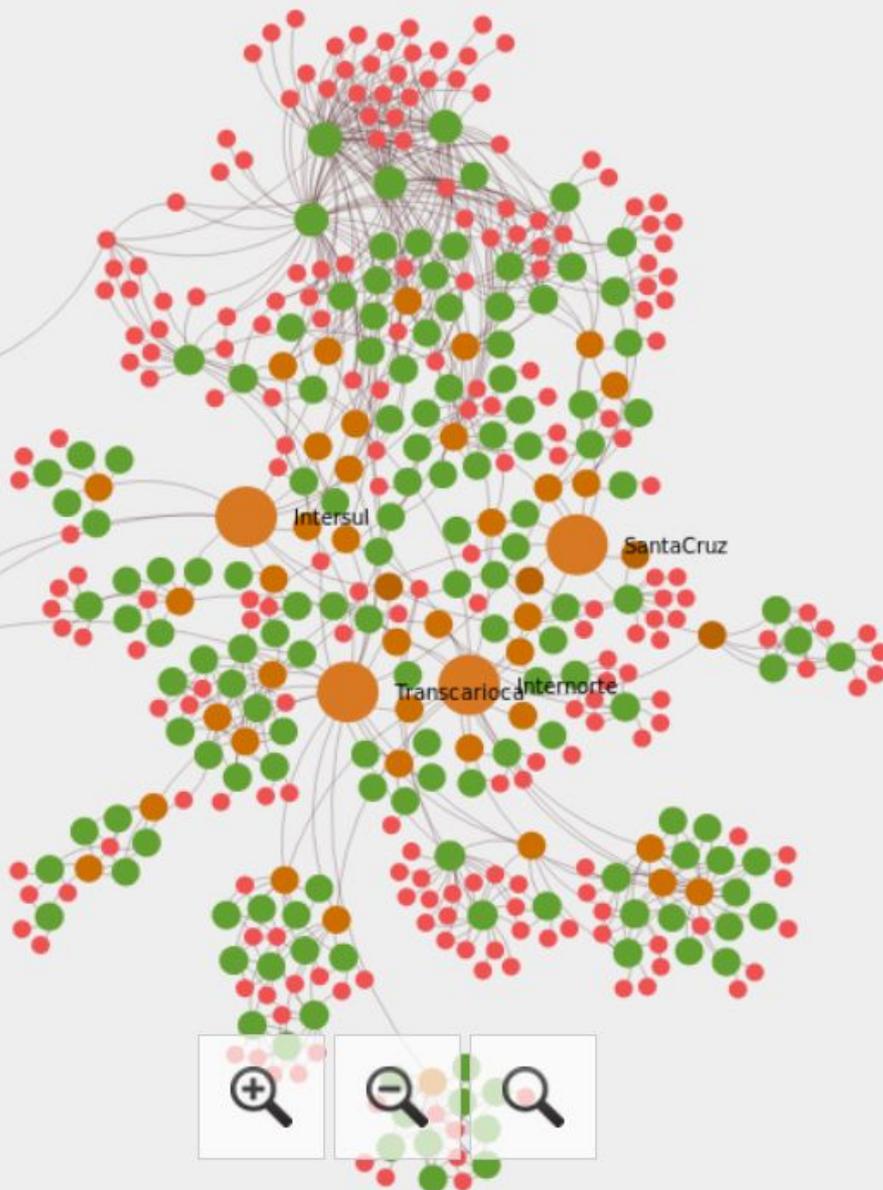
i Mais sobre esta visualização

Legend:

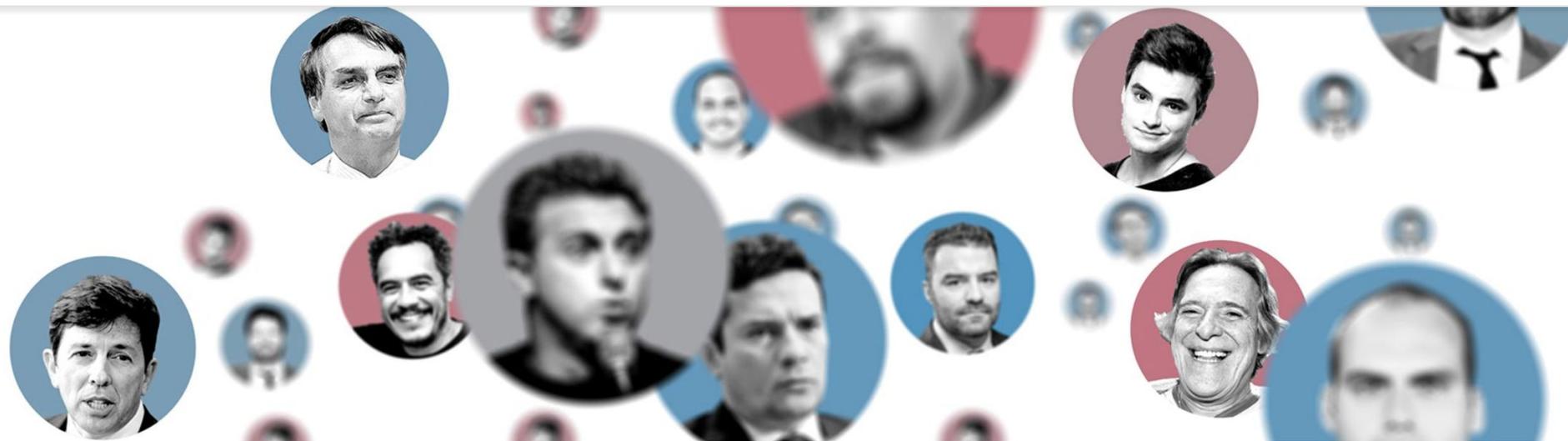
- Pessoas e empresas
- Relações de Propriedade
- Marrom: empresas de ônibus; Verde: Pessoas Físicas; Rosa: Outras empresas

Busca:

Selecione um tipo:



Inovação: novas metodologias pra coletar dados inéditos



GPS Ideológico

Análise do debate político no Twitter

Inovação: algoritmos de ciência política no jornalismo

O que revela uma análise das emoções dos candidatos durante o debate

Usando uma inteligência artificial desenvolvida pela Microsoft, detectamos as expressões faciais dos presidentiáveis no debate do último domingo, 9



Inovação: reconhecimento facial e IA

How Y'all, Youse and You Guys Talk

What does the way you speak say about where you're from?
Answer all the questions below to see your personal dialect map.

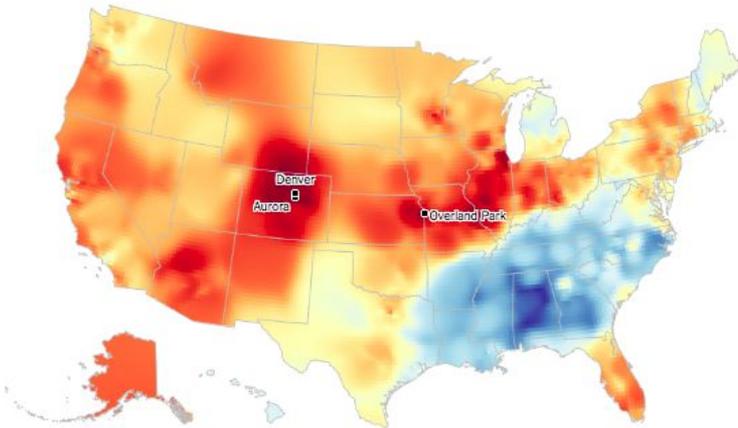
Your Map

See the pattern of your dialect in the map below. Three of the most similar cities are shown.

Least similar Most similar

Show least similar

SHARE YOUR MAP:   



TECHNOLOGY

The New York Times' Most Popular Story of 2013 Was Not an Article

It was that dialect quiz.

4.8k

94



ROBINSON MEYER | JAN 17, 2014

The Year's Most Visited All Content - Multiplatform

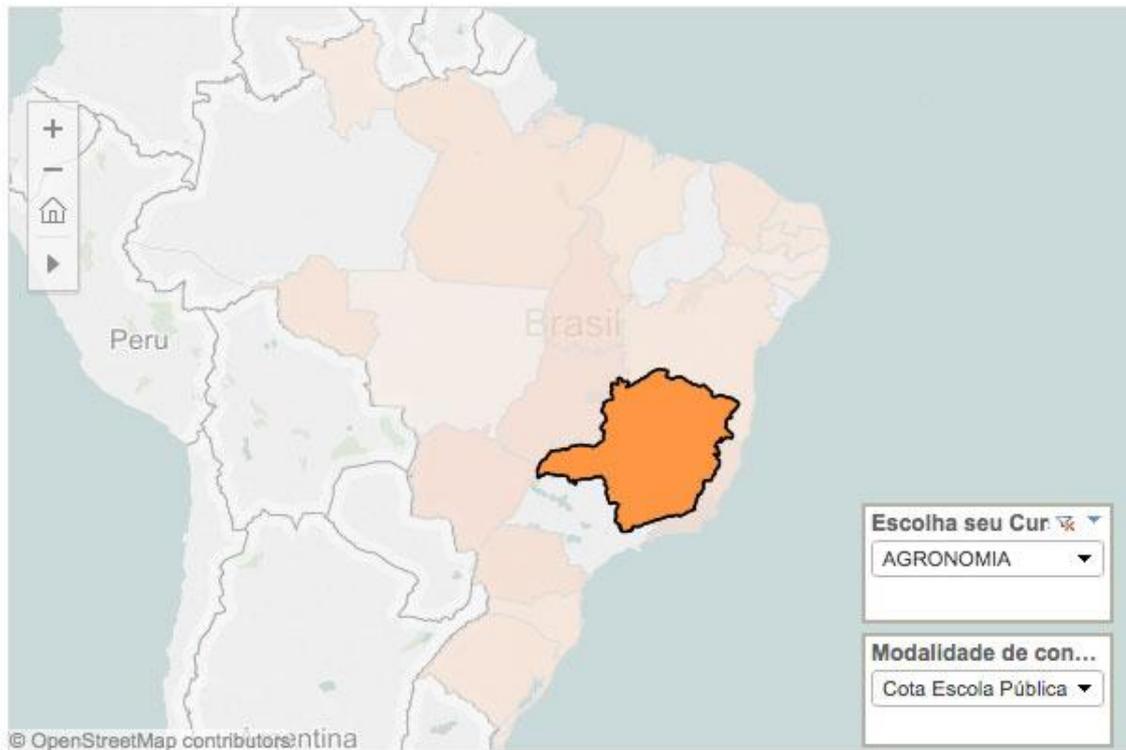
2013

2013 Rank

1 NYTIMES.COM, M.NYTIMES.COM + iOS

HOW Y'ALL, YOUSE AND YOU GUYS TALK, by Josh Katz & Wilson Andrews (December 21).
What does the way you speak say about where you're from? Answer the questions to see your personal dialect map.

Mapa de sotaques do NYT: página mais visitada do veículo em 2013



Entenda as cores: Quanto mais escuro, maior a nota de corte mínima, quanto mais claro, menor a nota de corte mínima

Ranking das Universidades

Sigla

UFU (CAMPUS UMUARAMA - INTEGRAL)	673.32
IFSEMG (CAMPUS BARBACENA - INTEGRAL)	662.56
UFLA (LAVRAS - UNIDADE SEDE - INTEGRAL)	662.2
UFV (UNIVERSIDADE FEDERAL DE VIÇOSA - INTEGRAL)	659.88
UFMG (CAMPUS REGIONAL DE MONTES CLAROS - MATUTINO)	653.84
UFV (CAMPUS DE FLORESTAL - UNIVERSIDADE FEDERAL DE VIÇOSA - INTEGRAL)	650.34
UFU (CAMPUS MONTE CARMELO - INTEGRAL)	646.25
UFVJM (CAMPUS PRESIDENTE JUSCELINO KUBITSCHEK DE OLIVEIRA - CAMPUS II - INTEGRAL)	645.38

Navegador das notas do SISU: novas audiências

Scale states equally

Scale states by population

 Where your friends live

Connect to Facebook to see the rights of states where your friends live. Your information will not be saved. 

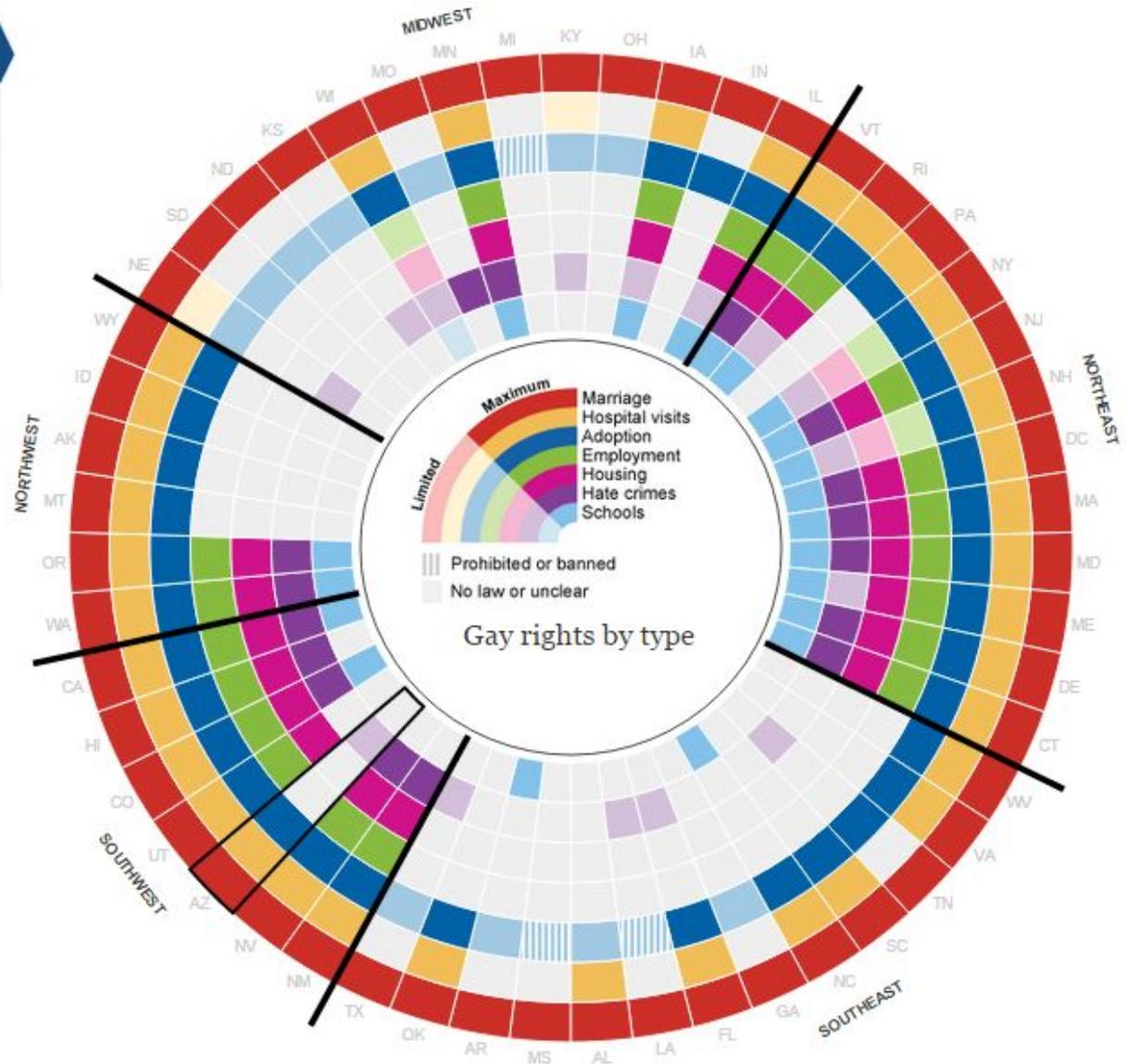
 Share your state on Facebook

You are not connected to Facebook.

Arizona

2010 population: 6,392,017

-  **Marriage:** Allows same-sex marriage.
-  **Hospital visitation:** Allows visitation by same-sex partners or spouses.
-  **Adoption:** Allows adoption by a single person and joint adoption by same-sex couples.
-  **Employment:** No protection against discrimination or the law is unclear.
-  **Housing:** No protection against discrimination or the law is unclear.
-  **Hate crimes:** Law addresses hate crimes related to sexual orientation.
-  **Schools:** No protection or the law is unclear.



Conteúdo perene: produção que mantém audiência



OjoPúblico

Sección OjoPúblico

¡Suscríbete!

HALLAZGOS

OPINIÓN

REACCIONES

¿QUÉ ES CUENTAS JURADAS?



Cuentas juradas de Lima Metropolitana

Mostrar cuentas juradas

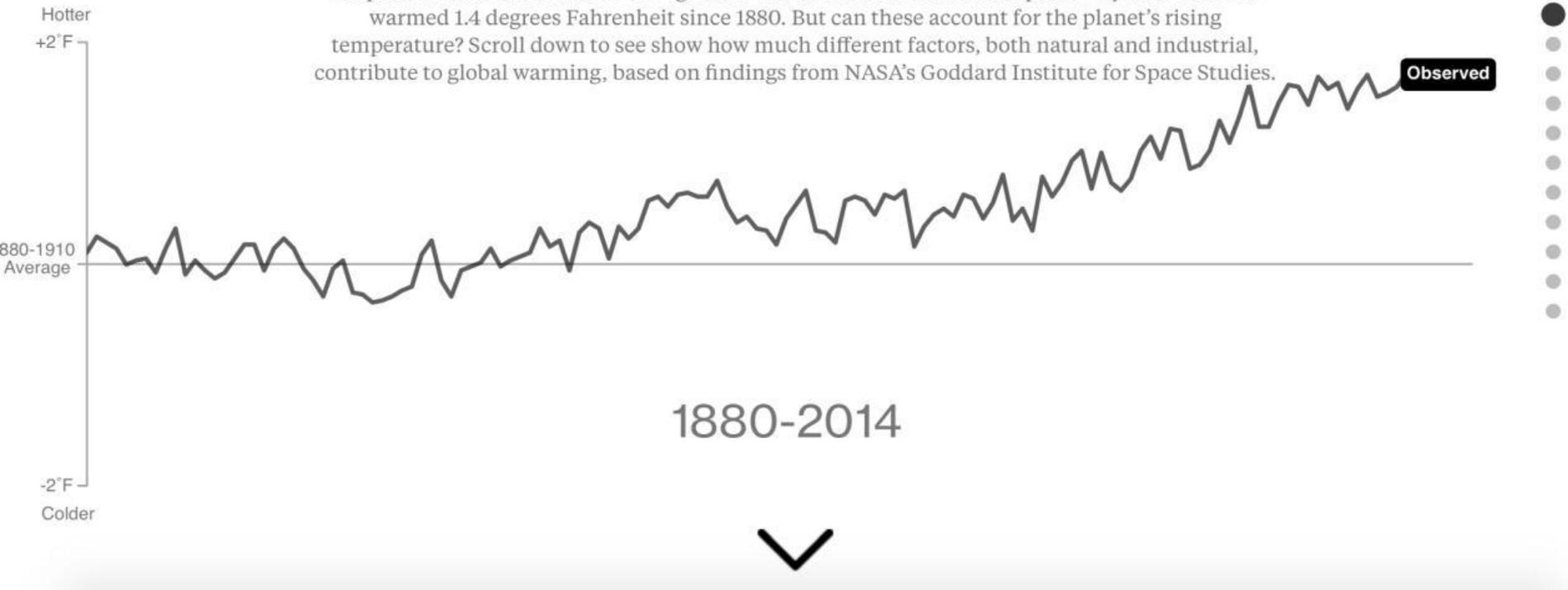


Pequenos podem ser grandes

What's Really Warming the World?

By Eric Roston  and Blacki Migliozi  | June 24, 2015

Skeptics of manmade climate change offer various natural causes to explain why the Earth has warmed 1.4 degrees Fahrenheit since 1880. But can these account for the planet's rising temperature? Scroll down to see how much different factors, both natural and industrial, contribute to global warming, based on findings from NASA's Goddard Institute for Space Studies.



Destrinchando a complexidade

The Best and Worst Places to Grow Up: How Your Area Compares

Children who grow up in some places go on to earn much more than they would if they grew up elsewhere. MAY 4, 2015 |

RELATED ARTICLE

GENDER

INCOME PERCENTILE

Switch to another county ...

All kids

Boys

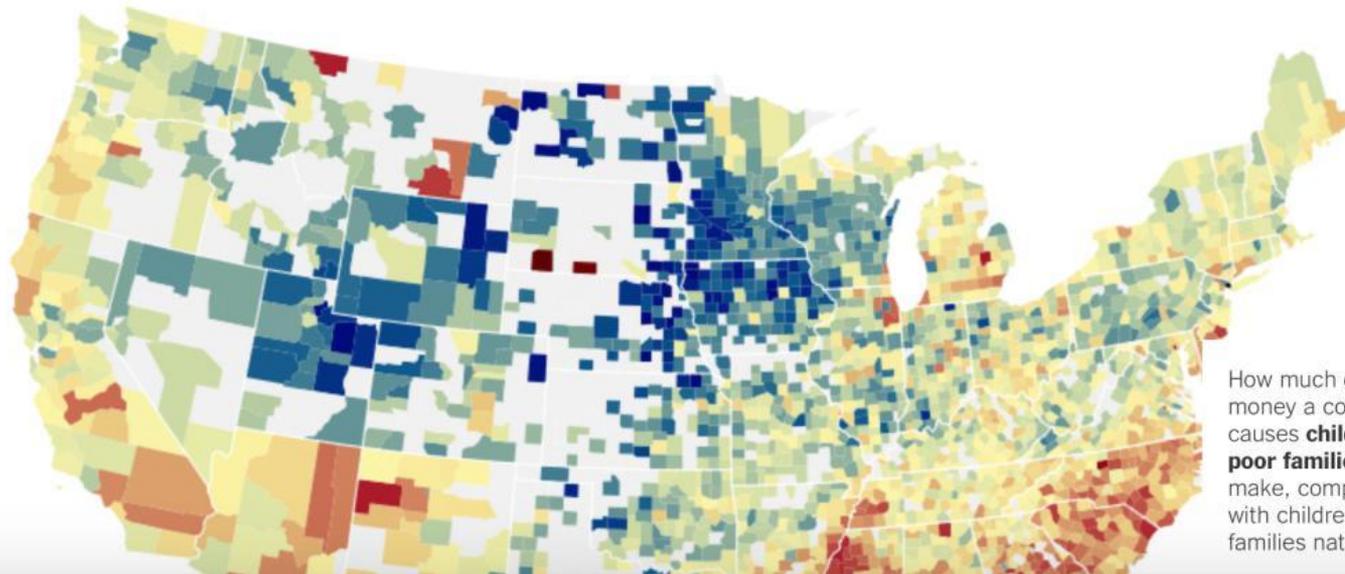
Girls

25th

50th

75th

99th



How much extra money a county causes **children in poor families** to make, compared with children in poor families nationwide.

Influenciando o debate público com evidências

Tell your own story with data from ProPublica.

Our premium data products are designed to deliver insights, with data, analysis, and practical documentation you won't find anywhere else.

Police Use of Force Reports (New Jersey)



DATA SET

PREMIUM

This research-ready data set includes data on more than 70,000 police use-of-force incidents across all 468 New Jersey municipal police departments and the New Jersey State Police between 2012 and 2016.

 Provided in collaboration with NJ Advance Media

[Learn more →](#)

Dollars for Docs Data



DATA SET

PREMIUM

Ready-to-use data on more than \$9 billion dollars paid by pharmaceutical companies to medical providers and health care institutions. *Updates annually*

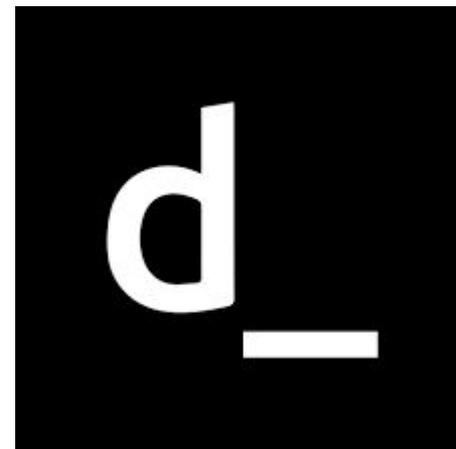
[Learn more →](#)

ProPublica Data Store: novos modelos de negócio

THE // INTERCEPT



INFOAMAZONIA 



 **FiveThirtyEight**

Novos veículos



ESCOLA
DE
DADOS

Habilidades de outras áreas

- Ciência da Computação
- Design
- Estatística
- Análise de Dados
- Sociologia e Antropologia
- Engenharia
- ...!

Data Pipeline



Fluxo de trabalho com dados:

cada uma das etapas requer uma série de habilidades diferentes

Como começar

Pergunta que precisa de bases de dados para ser respondida;

Ligue o radar: onde estão os números? Busque fontes de dados disponíveis (e não disponíveis)

Trate os dados como outra fonte qualquer: é confiável? O que essa fonte é capaz de me dizer?

Defina seu objeto com questões, e não afirmações

Limpeza

Garante a qualidade dos dados - integridade e previsibilidade;

Duas formas: retirar erros humanos e converter os dados a um formato consistente com outros já utilizados;

Contextualizar

Base de dados nem sempre é confiável e já vem com objetivos e preconceitos.

Devemos perguntar: quem a coletou, quando, como (metodologia) e por quais razões.

Tudo isso provavelmente levará a uma nova etapa de compilação.

Combinar

Boas histórias vêm de mais de uma base de dados;

Quanto mais fontes, melhor!

Comunicar

Visualização de dados;

Novas narrativas para deixar que os dados também contem a história: mapas, gráficos, infografias, animações, etc;

Precisa programar?

Depende.

Ferramentas com interface gráfica

Workbench;

Tableau;

Metabase (nível intermediário);

Linguagens mais comuns

R e Python;

Javascript (D3);

SQL para entrevistar bases de dados;

Habilidades essenciais

Jornalismo é a primeira delas

"Eu sempre digo que tem três coisas que você precisa saber pra trabalhar no meu time. Uma é jornalismo, então você precisa ser um bom jornalista, ter o que chamamos de senso editorial, entender o que é a história, qual a sua intenção, como contar os fatos. Também precisa ter algum conhecimento de design, capacidade de pelo menos seguir um template. E, por fim, código. E, de longe, o mais fácil desses três para ensinar alguém é como programar." (Scott Klein, Editor da ProPublica - do livro *Nerd Journalism*)



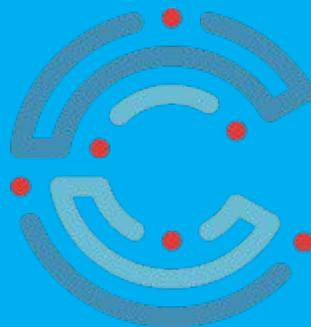
AFRô
PYTHON



coursea



meetup





ESCOLA
DE
DADOS

PERGUNTAS CRUCIAIS PROS DADOS

Dados



Informação



Apresentação



Conhecimento



Quantificações da realidade

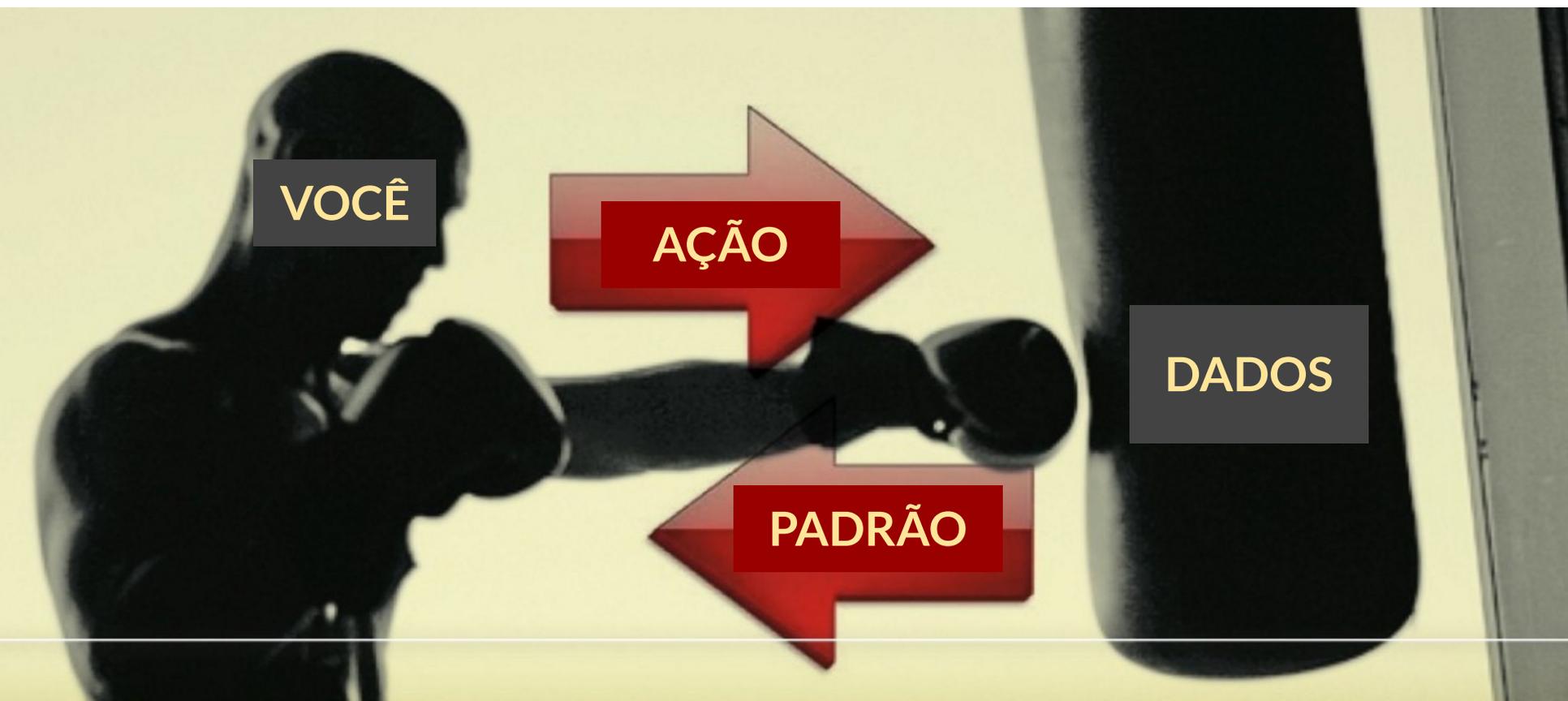
5 estrelas?

————— RATE YOUR RIDE —————



Representações objetivas?

O que quantificar já é uma escolha subjetiva



Desempregados?



Perguntas cruciais

- De onde a base vem? A metodologia de sua coleta está acessível?
- O que ela está, de fato, medindo?
- O que as escolhas adotadas sobre o que é medido nos diz?
- Há algum viés nessas escolhas?
- Os "a priori" dessas escolhas condizem com o que já conhecemos sobre aquele tema?
- Como os dados foram coletados?
- Como foi calculada a amostra?
- Qual a margem de erro dos dados coletados?

Ligar o radar!



ESCOLA
DE
DADOS

BUSCA AVANÇADA

O que o Google faz?

- **calculadora aritmética** (experimente colocar $9*2$ na caixa de busca. Depois experimente algo mais complexo: $\sqrt{\cos(x)}*\cos(300x)+\sqrt{\text{abs}(x)-0.7}*(4-x*x)^{0.01}$, $\sqrt{6-x^2}$, $-\sqrt{6-x^2}$ from -4.5 to 4.5)
- **conversor de moedas** (experimente colocar 100 brl em euro na caixa de busca)
- **conversor de medidas** (experimente colocar 100kg em libras na caixa de busca)
- **relógio mundial** (experimente colocar horário em Bogotá na caixa de busca)

O que o Google faz?

- **listagem de obras de autores e cantores** (experimente colocar a expressão livros de Machado de Assis na caixa de busca)
- **sistema de respostas diretas** (experimente colocar moeda da Jordania na caixa de busca)
- **jogos estúpidos da sua infância** (experimente colocar a expressão “atari breakout” na caixa de busca do Google Imagens)

Operadores de busca

- **Digite termos entre aspas**
 - "acidentes de trânsito"
- **Afunile sua pesquisa**
 - sinal de menos para excluir
 - sinal pode ser repetido
- **Amplie a busca com um coringa**
 - asterisco no meio de uma frase

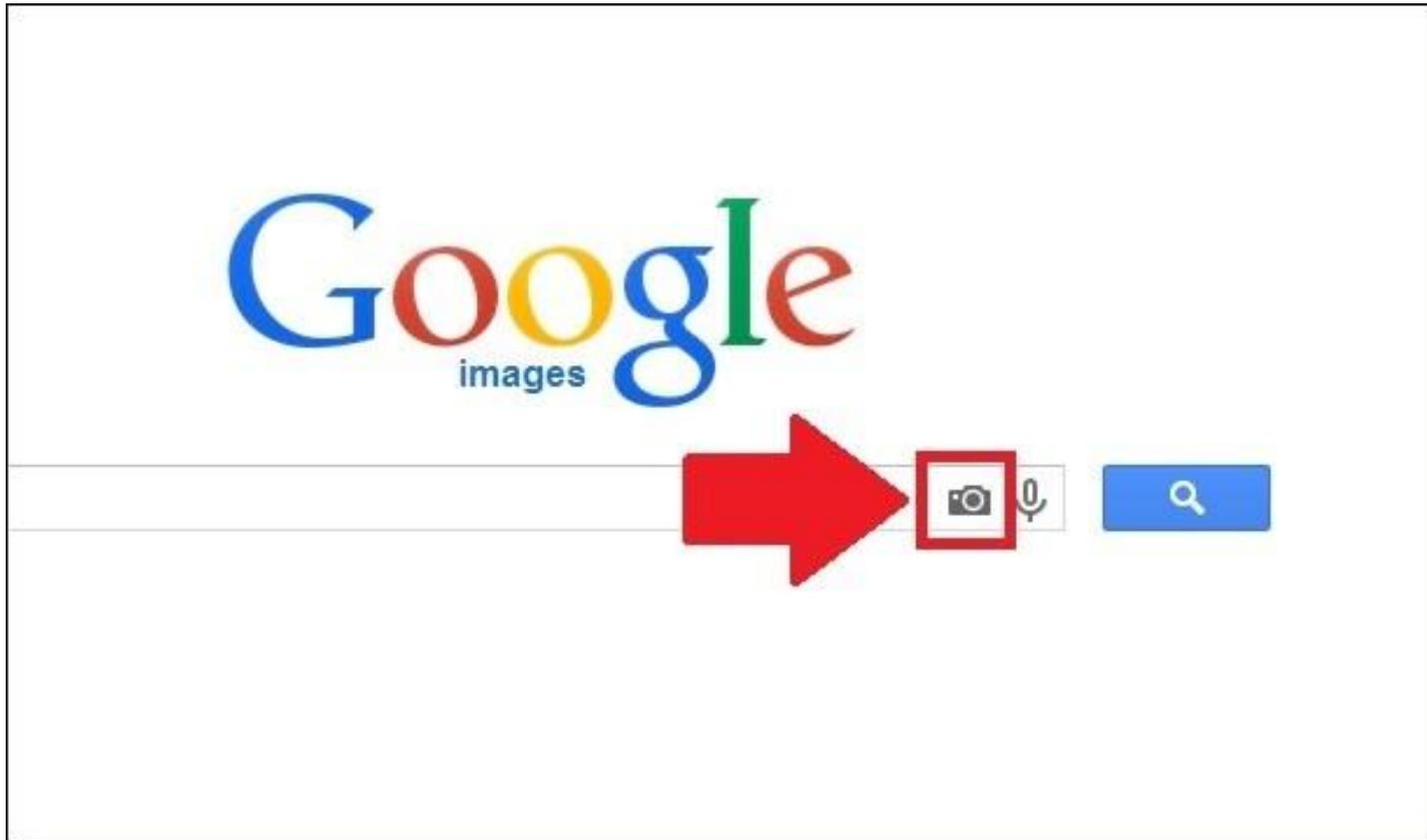
Operadores de busca

- **Procurar dentro de um site**
 - site:gov.br
 - site:data.gov.br
- **Procurar em sites registrados em um país**
 - site:br retorna sites registrados no NIC.br
 - site:br não retorna sites brasileiros de fora

Operadores de busca

- **filetype:xls**
 - retorna planilhas do Excel
- **filetype:doc**
 - retorna documentos do Word
- **filetype:pdf**
 - retorna arquivos em PDF

Busca reversa de imagens



Busca reversa de imagens



Pesquisa por imagem ×

Pesquise no Google com uma imagem em vez de texto. Tente arrastar uma imagem aqui.

Colar o URL da imagem **Envie uma imagem** 

Escolher arquivo Nenhum arquivo selecionado

Busca reversa de imagens

Yandex

Bonus track!



Products · Accounts

Reverse Image Search

Search by image and find where that image appears online

 Upload or enter Image URL 

[How to use TinEye](#)





ESCOLA
DE
DADOS

LAI

Lei de Acesso à Informação

Transparência ativa

Em seu artigo 3º, a LAI estabelece diretrizes para sua aplicação, entre elas a de que as **informações de interesse coletivo ou geral deverão ser divulgadas de ofício pelos órgãos públicos, espontânea e proativamente**, independentemente de solicitações. Institui-se, neste artigo, o conceito de Transparência Ativa, que deriva do princípio da publicidade das atividades administrativas.

Estas devem representar o interesse público, o que justifica a excepcionalidade do sigilo. Portanto, o caminho inicial para quem quer obter dados públicos é verificar se estão disponíveis no site do órgão sobre o qual se busca a informação.

Transparência passiva

Qualquer cidadão pode fazer pedidos de acesso à informação pública para os órgãos submetidos à LAI. A “Transparência Passiva” se dá **quando algum órgão ou ente é demandado pela sociedade a prestar informações** que sejam de interesse geral ou coletivo, desde que não sejam resguardadas por sigilo (veja no quadro abaixo as situações que permitem o sigilo).

O que é informação pública?

- informação produzida ou acumulada por órgãos e entidades públicas;
- informação produzida ou mantida por pessoa física ou privada decorrente de um vínculo com órgãos e entidades públicas;
- informação sobre atividades de órgãos e entidades, inclusive relativa à sua política, organização e serviços;
- informações pertinentes ao patrimônio público, utilização de recursos públicos, licitação e contratos administrativos;
- informações sobre políticas públicas, inspeções, auditorias, prestações e tomadas de contas.

TRANSPARÊNCIA ATIVA

Órgão ou Entidade



Internet



TRANSPARÊNCIA PASSIVA



Pedido



Serviço de
Informações
ao Cidadão



Área responsável
pela informação



Levantamento
em arquivos
ou sistemas

Resposta



Digitalização
e copiagem



Informação sigilosa

- Três níveis:
 - Ultrassecreto - 25 anos
 - Secreto - 15 anos
 - Reservado - 5 anos
- Renovável apenas uma vez. **Prazo máximo: 50 anos**
- Informação é automaticamente tornada pública após o prazo.

Informação sigilosa

- Pôr em risco a defesa e a soberania nacionais e planos ou operações estratégicos das Forças Armadas.
- Prejudicar a condução de negociações ou as relações internacionais do país, se as informações foram fornecidas em caráter sigiloso por outros estados e organismos internacionais.
- Pôr em risco a vida, a segurança ou a saúde de terceiros.
- Oferecer elevado risco à estabilidade financeira, econômica ou monetária do país.
- Prejudicar ou causar risco a projetos de pesquisa e desenvolvimento científico ou tecnológico.
- Pôr em risco a segurança de instituições ou 'altas' autoridades nacionais ou estrangeiras e seus familiares.
- Comprometer atividades de inteligência, bem como de investigação ou fiscalização em andamento.
- Estar sob sigilo de justiça ou ser segredo industrial
- Estar relacionadas à intimidade, vida privada, honra e imagem de uma pessoa.

O que pedir?

Como pedir?

Pontos importantes

- **Objetividade e precisão:** Utilize linguagem clara e seja específico em relação à informação que procura.
- **Privacidade:** Deixe claro no seu pedido que, caso as informações que você procura de alguma forma violem o princípio de privacidade, solicite a edição das informações de modo a não violar o princípio.
- **Simplicidade:** Pedidos que exigem trabalho extra de consolidação ou análise são rejeitados
- **Metadados:** Buscar sempre os dados que descrevem os dados. Quais campos existem em determinada base? Metadado também da Lei de Acesso: Quais pedidos foram feitos usando a Lei de Acesso à Informação e quais foram deferidos? Para saber quais bases já foram abertas.

Dicas preciosas

- **Nada de pedidos genéricos:** se o pedido for genérico demais, ele pode ser recusado. Qual é o local? Qual é o período? Qual item ou itens você procura? Ex: Últimos 5 anos de profissionais do programa Mais Médicos em São Paulo
- **Quem detém a informação?** Faça uma pesquisa inicial para saber se a matéria que você pesquisa é de responsabilidade de União, Estados ou Municípios e qual órgão seria competente por aquela informação.
- **Embase o seu pedido:** mostrar que você sabe não só o que está pedindo, mas que também conhece precedentes que embasam o requerimento. Decisões de instâncias superiores, formato aberto "nos termos do art. 8º, §3º, III da Lei Federal 12.527/11 e art. 24, V da Lei Federal 12.965/14", "anonimizando as informações consideradas sigilosas na base".