

Como os

DADOS

Aconteceram

*da Era da Razão à
Era dos Algoritmos*

AMOSTRA

CHRIS WIGGINS
MATTHEW L. JONES

Como os
DADOS
Aconteceram

*da Era da Razão à
Era dos Algoritmos*



ALTA BOOKS
GRUPO EDITORIAL
Rio de Janeiro, 2024

Como os Dados Aconteceram

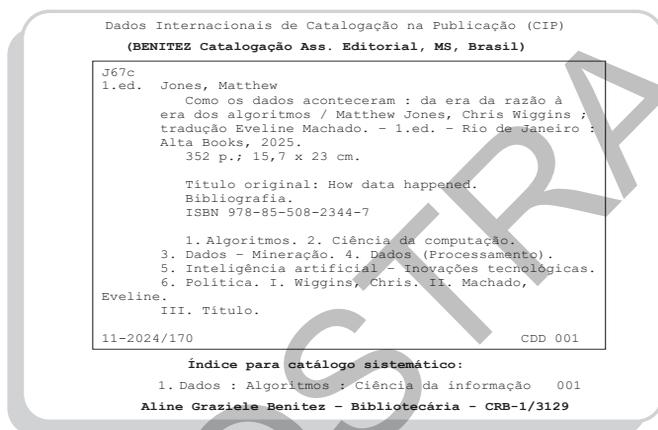
Copyright © 2025 Starlin Alta Editora e Consultoria Eireli.

Copyright © 2023 Matthew Jones and Chris Wiggins.

ISBN: 978-85-508-2344-7

Translated from original How Data Happened. Copyright © 2023 by Matthew Jones and Chris Wiggins. This translation is published and sold by Andrew Nurnberg Associates Limited. ISBN 978-1-324-00673-2, the owner of all rights to publish and sell the same. PORTUGUESE language edition published by Starlin Alta Editora e Consultoria Eireli, Copyright © 2025 by Starlin Alta Editora e Consultoria Eireli.

Impresso no Brasil – 1ª Edição, 2025 – Edição revisada conforme o Acordo Ortográfico da Língua Portuguesa de 2009.



Todos os direitos estão reservados e protegidos por Lei. Nenhuma parte deste livro, sem autorização prévia por escrito da editora, poderá ser reproduzida ou transmitida. A violação dos Direitos Autorais é crime estabelecido na Lei nº 9.610/98 e com punição de acordo com o artigo 184 do Código Penal.

O conteúdo desta obra fora formulado exclusivamente pelo(s) autor(es).

Marcas Registradas: Todos os termos mencionados e reconhecidos como Marca Registrada e/ou Comercial são de responsabilidade de seus proprietários. A editora informa não estar associada a nenhum produto e/ou fornecedor apresentado no livro.

Material de apoio e erratas: Se parte integrante da obra e/ou por real necessidade, no site da editora o leitor encontrará os materiais de apoio (download), errata e/ou quaisquer outros conteúdos aplicáveis à obra. Acesse o site www.altabooks.com.br e procure pelo título do livro desejado para ter acesso ao conteúdo.

Suporte Técnico: A obra é comercializada na forma em que está, sem direito a suporte técnico ou orientação pessoal/exclusiva ao leitor.

A editora não se responsabiliza pela manutenção, atualização e idioma dos sites, programas, materiais complementares ou similares referidos pelos autores nesta obra.

Produção Editorial: Grupo Editorial Alta Books

Diretor Editorial: Anderson Vieira

Editor da Obra: J. A. Ruggeri

Vendas Governamentais: Cristiane Mutús

Gerência Comercial: Claudio Lima

Gerência Marketing: Viviane Paiva

Produtora Editorial: Isabella Gibara

Tradução: Eveline Cardoso

Capitulos: Carol Suiter

Revisão: Thais Cotts

Diagramação: Alice Sampaio

Capa: Lorrainh Cândido


ALTA BOOKS
GRUPO EDITORIAL

Rua Viúva Cláudio, 291 – Bairro Industrial do Jacaré

CEP: 20.970-031 – Rio de Janeiro (RJ)

Tels.: (21) 3278-8069 / 3278-8419

www.altabooks.com.br – altabooks@altabooks.com.br

Ouvidoria: ouvidoria@altabooks.com.br



AMOSTRA

Para nossas famílias, que tornaram possível este livro

AMOSTRA

SUMÁRIO

PRÓLOGO	IX
---------------	----

PARTE I

01. Os Envolvidos	3
02. Física Social e l'homme moyen	17
03. A Estatística do Divergente	31
04. Dados, Inteligência e Política	49
05. O Batismo Matemático dos Dados	69

PARTE II

06. Dados em Tempos de Guerra	89
07. Inteligência sem Dados	107
08. Volume, Variedade e Velocidade	125
09. Máquinas, Aprendizado	155
10. A Ciência dos Dados	175

PARTE III

11. A Batalha pela Ética dos Dados	207
12. Persuasão, Publicidade e Capital de Risco	229
13. Soluções além do Solucionismo	253

AGRADECIMENTOS	275
----------------------	-----

NOTAS	279
-------------	-----

ÍNDICE	331
--------------	-----

AMOSTRA

PRÓLOGO

EM UMA MANHÃ de abril em 2018, conforme a luz do sol da primavera entrava pela janela leste da sala do seminário no Schermerhorn Hall da Universidade Columbia, eu (Wiggins) fui para o quadro-negro explicar a coisificação quantitativa, processo mágico pelo qual uma correspondência numérica da observação empírica se torna uma coisa. Armado com a história de Adolphe Quetelet, que visava revelar o homem ideal usando os dados que ele obteve sobre as medições físicas dos soldados escoceses, tracei no quadro a imortal “curva normal”. Conhecida pelos matemáticos como curva gaussiana e contestada como a famosa “curva em sino” dos testes de QI, a curva normal significa para os cientistas naturais que os dados revelaram algo real, mesmo sendo algo transcendental. Virei para os alunos, na esperança de ver em seus olhos que eles compartilhavam do meu entusiasmo. Um me encarou e levantando as mãos para o céu, perguntou: “Podemos falar sobre o Facebook agora?”

Naquela manhã, jornais e feeds de notícias digitais igualmente abriram fogo contra Washington, acabando com toda dissimulação. O CEO irreverente de uma empresa de tecnologia que mudaria a cultura no Vale do Silício estava sendo chamado ao Senado dos Estados Unidos. Em nome de todos os cidadãos, senadores buscavam entender como os dados pessoais de milhões de pessoas, inclusive dos nossos alunos, foram comprometidos, o *New York Times* explicava, usados para fins escusos que violavam as normas sobre privacidade e nosso processo

político.¹ No fim do depoimento no Congresso naquela semana, os alunos reconheceram o tamanho da lacuna entre como os representantes eleitos entendiam sua realidade mediada digitalmente e seu conhecimento pessoal a partir do crescimento com os algoritmos.

A história dos dados está repleta de disputas: disputas para definir o que é verdade, disputas para usar dados e aumentar o poder de alguém, às vezes, disputas para usar algoritmos e dados para iluminar a escuridão e empoderar os indefesos. Este livro nasceu do nosso ensino de centenas de alunos curiosos, junto de nossas próprias experiências, como um historiador da ciência e um cientista de dados em exercício, como cidadãos tentando entender como chegamos nessa realidade mediada por algoritmos, como podemos escolher viver de modo diferente. Como todos os usuários, desenvolvedores e temas da tecnologia, tentamos entender para onde vai tudo isso, além de como modelaremos coletivamente o futuro. Tentamos contar uma história com ideias e tecnologias, mas também uma história com verdade e poder.

Deixando de lado o giz, concordamos que Quetelet teria sua vez. Mas primeiro, precisaríamos explicar como um astrônomo belga desconhecido se encaixa na história dos dados: como os dados e os meios para analisá-los passariam de uma preocupação do Estado para as universidades, as Forças Armadas e o setor privado.

Usamos os “dados” aqui como um atalho para a vastidão dos sistemas de tomada de decisão algorítmica baseada em dados nos cercando por quase todos os lados. Exploramos como os dados foram criados e escolhidos, além de como novas técnicas matemáticas e de computação que se desenvolveram para lidar com os dados servem para modelar pessoas, ideias, a sociedade, operações militares e economias. Junto com os dados vem o poder, inclusive o poder de modelar o que é percebido como verdade. Embora a tecnologia e a matemática sejam centrais, a história dos dados diz respeito a um jogo instável entre Estados, corporações e pessoas.

E por isso, naquela manhã, falamos não só sobre dados, mas sobre os riscos de um mundo mediado por eles.

Contexto

A ideia de uma aula sobre como os dados aconteceram nasceu em novembro de 2015, em uma conversa rápida durante o jantar com alguns alunos da graduação na Columbia com uma combinação de formações nas áreas de engenharia e humanas. Na época, suspeitávamos que os alunos estivessem muito interessados na história sobre data science e pensávamos que nossas perspectivas combinadas e complementares forneceriam uma visão útil, com material novo igualmente para engenheiros e não tecnólogos. Quando demos a aula pela primeira vez em janeiro de 2017, percebemos rápido que os alunos não só estavam interessados em como chegamos lá, como também buscavam uma estrutura analítica e viável para entender a ética e a política dos dados.²

Com “política” não queremos dizer uma definição limitada como “votar”, mas “relacionada à dinâmica de poder”. Nosso objetivo é fornecer uma estrutura para entender o papel persistente dos dados em reorganizar o poder: poder corporativo, poder estatal e poder do povo. Nosso arco histórico fornece uma vantagem crítica que aponta para uma compreensão compartilhada do presente, assim como as armas e as ferramentas à nossa disposição para modelar o futuro.

Este Livro

Toda história deve começar em algum lugar e achamos que um ponto de partida útil é o final do século XVIII, na época em que a palavra “estatística” apareceu pela primeira vez no idioma inglês. Nossa história fica entre o difícil trabalho de coletar dados, inclusive criar uma infraestrutura para coletá-los e torná-los públicos, e o desenvolvimento de novas técnicas matemáticas e de computação para estudá-los — novos modos de entender e fazer afirmações sobre esses dados, usando tais afirmações para tomar decisões, muitas vezes mudando profundamente as vidas, para melhor ou pior. Em cada capítulo consideramos uma transição intelectual. Debates sobre como uma nova capacidade técnica e científica foi desenvolvida; quem apoiou, fez avanços ou financiou essa capacidade ou transição; como a transição foi contestada; e como essa nova capacidade reorganizou o poder, mudando quem podia fazer o quê, a partir

de quê e para quem.³ Focamos não apenas as reorganizações do poder militar ou financeiro, mas em geral as transições com uma valência ética e política: em que os dados afetam os direitos, reorganizam os danos, apoiam ou impedem a justiça.

Como os Dados Aconteceram começa com dados a serviço do estadismo, antes de voltar o uso dos dados para a melhoria da sociedade e o batismo matemático dos dados com a criação de um novo campo acadêmico chamado “estatística matemática”. A segunda parte abre com a aplicação militar dos dados na Segunda Guerra Mundial para quebrar o código, coincidindo com o nascimento da computação digital. Seguimos desde o Bletchley Park na Inglaterra ao Bell Labs nos EUA, e até as aplicações comerciais e de engenharia de dados como resultado da Segunda Guerra. Fazendo a transição do poder corporativo para as reações no poder do Estado e no “poder do povo”, exploramos o impacto do registro digital e pessoal mantendo nossa compreensão de privacidade, em particular o desejo público de privacidade como uma defesa contra o excessivo poder do Estado nos anos 1970. Traçamos o primeiro nascimento e morte do campo da “inteligência artificial” e seu renascimento, saindo das cinzas na forma de “aprendizado de máquina” com base em repositórios cada vez maiores de dados sobre cidadãos, consumidores e adversários militares do mesmo modo.

A última seção do livro conecta esse passado ao nosso presente e futuro. Examinamos como dados e poder saíram de uma preocupação do Estado e passaram a ser uma preocupação corporativa, vendo organizações financeiras e modelos de negócio que permitiram que uma corporação dominasse setores inteiros rapidamente com a ajuda de tecnologias capacitadas por dados. Um debate duvidoso sobre ética estruturou muitas das possíveis soluções para o poder corporativo; traçamos a história da ética aplicada na pesquisa e como ela impactou o modo como os algoritmos capacitados por dados são implantados como produtos, modelando nossa realidade pessoal e política.

Por fim, debatemos sobre o futuro. Por mais difícil que seja fazer previsões, um modo incisivo de organizar nossa compreensão do futuro é descrever as disputas presentes entre os poderes, além das áreas nas quais essas disputas serão decididas. Fechamos o livro vendo o que consideramos ser as disputas

presentes mais importantes entre os poderes corporativo, do Estado e do povo, junto com a possibilidade de novas formas de solidariedade. As soluções dessas disputas modelarão nosso futuro coletivo, tendendo mais para a justiça — ou talvez não.

Nosso objetivo aqui é um entendimento prático da história. Não fugiremos de nossos papéis como cidadãos, tecnólogos e indivíduos; somos usuários desses produtos e, como mencionado lá nos anos 1970, já que estamos em uma economia de publicidade, também somos o produto.

Trouxemos para o livro duas perspectivas complementares, cada uma com limites e tendências. Wiggins desenvolve métodos de aprendizado de máquina para entender a biologia e a saúde há mais de 20 anos como professor em Columbia e, desde 2013, desenvolve e implanta métodos e produtos de aprendizado de máquina como cientista de dados chefe no *New York Times*. No outro lado das “duas culturas” de C. P. Snow, Jones é historiador de ciências, tendo rastreado como os modos matemáticos de pensar e o debate se tornaram um jeito oficial e essencial de estudar a natureza e a política a partir da “Revolução Científica” do século XVII em diante. Em particular, ao examinar como muitos usos dos dados aumentam a disparidade, nos baseamos bastante na escrita esclarecedora de muitos acadêmicos e ativistas que expuseram esses processos. Muitos, se não a maioria, dos críticos mais incisivos e brilhantes vêm essencialmente de origens e experiências muito diferentes das nossas — dois acadêmicos brancos do sexo masculino. Nosso trabalho se baseia e celebra a obra e a visão deles por toda parte. Indicaremos uma excelente literatura sobre o impacto global dos algoritmos e das tecnologias capacitadas por dados, e histórias dos dados na organização de nossas sociedades, economias e instituições educacionais. Nosso material mais contemporâneo foca basicamente os EUA. Fornecemos notas finais para refletir não só onde descobrir mais sobre os tópicos dados em aula e escritos sobre nossas publicações acadêmicas, mas também muitos trabalhos importantes, inclusive a literatura acadêmica, que encorajamos que os leitores vejam para aprofundar o conhecimento.

Procuramos dar uma visão clara das tensões históricas e atuais entre os poderes corporativo, do Estado e do povo, focando o papel dos dados ao estabelecer a verdade e modelar as disputas entre eles. Esperamos mostrar como chegamos lá coletivamente, para representar as pequenas coincidências, as escolhas subjetivas de

design e as decepções que se calcificaram no que apenas *parece* como as coisas “devem ser”. Entender essas transições e contingências revelará como problemas parecidos foram resolvidos no passado. E, por sua vez, nos ajudará a retratar como poderemos romper e redefinir os ossos do sistema que, por vezes, empodera o indefeso — embora com frequência tenha fortalecido o empoderado.

Mostrando como resultados aparentemente imutáveis dependeram de escolhas no passado, podemos ver como escolher um futuro diferente de maneira coletiva.

AMOSTRA

01.

Os Envolvidos

A tecnologia não é boa nem má, e também não é neutra.

— *Primeira lei da tecnologia de Kranzberg, 1986*

Ministro um curso na Universidade de Michigan chamado “a Internet é um fiasco” e não tenho que explicar a ninguém o que significa... Aturamos isso por muito tempo; não parecemos saber nada diferente.

— *Lisa Nakamura, 2019*

EM DEZEMBRO DE 2014 no Palais des Congrès de Montreal, a cientista de computação Hanna Wallach defendia uma revolução diante um público de tecnólogos, advogados e ativistas. Falando para os principais cientistas de computação que trabalham com “aprendizado de máquina”, ela propôs que seu próprio campo precisava desesperadamente questionar como os algoritmos que desenvolviam e as tecnologias que os algoritmos forneciam, desafiavam nossos valores de “equidade, responsabilidade e transparência”. Embora filósofos, sociólogos