

UNIVERSIDADE FEDERAL DE OURO PRETO  
INSTITUTO DE CIÊNCIAS SOCIAIS APLICADAS  
DEPARTAMENTO DE JORNALISMO

Daiane Maciel de Lima Silva

ESPECIALIZANDO O BIG DATA: A PARTICIPAÇÃO DO VOLT DATA LAB EM  
PRODUÇÕES DE JORNALISMO DE DADOS

MARIANA  
2023

Daiane Maciel de Lima Silva

**ESPECIALIZANDO O BIG DATA:**  
A PARTICIPAÇÃO DO VOLT DATA LAB EM PRODUÇÕES DE JORNALISMO DE  
DADOS

Monografia apresentada ao curso de  
Jornalismo da Universidade Federal de Ouro  
Preto, como requisito para a obtenção do título  
de Bacharel em Jornalismo.

Orientadora: Profa. Dra. Natalia Moura  
Pacheco Cortez

MARIANA  
2023

## SISBIN - SISTEMA DE BIBLIOTECAS E INFORMAÇÃO

S586e Silva, Daiane Maciel De Lima.

Especializando o Big Data [manuscrito]: a participação do Volt Data Lab em produções de Jornalismo de Dados. / Daiane Maciel De Lima Silva. - 2023.

71 f.: il.: color.. + Quadro.

Orientadora: Profa. Dra. Natália Moura Pacheco Cortez.

Monografia (Bacharelado). Universidade Federal de Ouro Preto. Instituto de Ciências Sociais Aplicadas. Graduação em Jornalismo .

1. Volt Data Lab. 2. Big data. 3. Jornalismo. 4. Jornalismo científico. I. Cortez, Natália Moura Pacheco. II. Universidade Federal de Ouro Preto. III. Título.

CDU 070:001.5

Bibliotecário(a) Responsável: Essevalter De Sousa - Bibliotecário Coordenador  
CBICSA/SISBIN/UFOP-CRB6a1407



## FOLHA DE APROVAÇÃO

Daiane Maciel de Lima Silva

**Especializando o Big Data: a participação do Volt Data Lab em produções de jornalismo de dados**

Monografia apresentada ao Curso de Jornalismo da Universidade Federal de Ouro Preto como requisito parcial para obtenção do título de bacharel em Jornalismo.

Aprovada em 29 de janeiro de 2024.

### Membros da banca

Prof.(a) Dr(a). Natália Moura Pacheco Cortez - Orientador(a) - Universidade Federal de Ouro Preto  
Prof.(a) Dr(a). Debora Cristina Lopez - Universidade Federal de Ouro Preto  
Prof.(a) Dr(a). Marcelo Freire Pereira de Souza - Universidade Federal de Ouro Preto

Prof.(a) Dr(a). Natália Moura Pacheco Cortez, orientadora do trabalho, aprovou a versão final e autorizou seu depósito na Biblioteca Digital de Trabalhos de Conclusão de Curso da UFOP em 29/01/2024.



Documento assinado eletronicamente por **Natália Moura Pacheco Cortez, PROFESSOR DE MAGISTERIO SUPERIOR**, em 30/01/2024, às 00:28, conforme horário oficial de Brasília, com fundamento no art. 6º, § 1º, do [Decreto nº 8.539, de 8 de outubro de 2015](#).



A autenticidade deste documento pode ser conferida no site [http://sei.ufop.br/sei/controlador\\_externo.php?acao=documento\\_conferir&id\\_orgao\\_acesso\\_externo=0](http://sei.ufop.br/sei/controlador_externo.php?acao=documento_conferir&id_orgao_acesso_externo=0), informando o código verificador **0660102** e o código CRC **5460075F**.

*“Se você torturar os dados por tempo suficiente, eles confessarão qualquer coisa”.*  
(Ronald Coase)

## RESUMO

Esta pesquisa tem o objetivo de analisar contribuições da agência de dados Volt Data Lab em suas parcerias com veículos jornalísticos. Para isso, foi realizado um estudo de caso, comparando cinco projetos da agência com quatro produções vencedoras da edição 2022 do Prêmio Cláudio Weber Abramo de Jornalismo de Dados. O recorte foi estabelecido com o intuito de garantir que as matérias analisadas contassem com um critério de qualidade reconhecido em seu meio, para que, a partir da comparação, fosse possível reconhecer as singularidades das produções do Volt Data Lab. Com isso, buscou-se entender como a especialização da parte do trabalho relacionada a dados resultaria em melhorias e inovações. O foco da pesquisa foram alguns problemas relativos ao *Big Data*, como os vieses de conjunto de dados, investigados por Pasquinelli e Joler (2021), e a opressão algorítmica, termo cunhado por Noble (2018). Além disso, para estabelecer uma base comparativa, este trabalho perpassa o contexto que levou a formação do Jornalismo de Dados, passando pelos estudos de autores como Marcondes Filho (2000), Meyer (2002), Bastos (2005, 2012), Canavilhas (2003, 2006), Manovich (2001), Bradshaw (2011), entre outros. Como resultado da pesquisa, foi possível avaliar que a especialização não é um fator para que o Jornalismo de Dados enfrente as questões do *Big Data*, mas que uma cooperação aprofundada entre a agência e os veículos é necessária para uma conduta equilibrada no que tange à transmissão de vieses.

Palavras-chave: *Big Data*; Jornalismo de Dados; Volt Data Lab

## **ABSTRACT**

This research aims to analyze the contributions of the data agency Volt Data Lab in its partnerships with media outlets. For this purpose, a case study was conducted, comparing five projects of the data agency with four Prêmio Cláudio Weber Abramo de Jornalismo de Dados winner productions. The selection was made to ensure that the analyzed articles met a recognized quality criterion within their field, allowing for the identification of the unique aspects of Volt Data Lab's productions through comparison. The goal was to understand how specialization in the data-related aspects of the work in journalism would lead to improvements and innovations. The research focused on specific issues related to Big Data, such as dataset biases, investigated by Pasquinelli and Joler (2021), and algorithmic oppression, a term coined by Noble (2018). Additionally, to establish a comparative basis, this work delves into the context that led to the formation of Data Journalism, spanning studies by authors such as Marcondes Filho (2000), Meyer (2002), Bastos (2005, 2012), Canavilhas (2003, 2006), Manovich (2001), Bradshaw (2011), among others. As a result of the research, it was possible to evaluate that specialization is not a determining factor for Data Journalism to address Big Data issues, but rather that a deep collaboration between the agency and the media outlets is necessary for a balanced approach regarding the transmission of biases.

Keywords: Big Data; Data Journalism; Volt Data Lab

## LISTA DE ILUSTRAÇÕES

<b>Figura 1</b> - Categoria x Níveis de JD.....	27
<b>Figura 2</b> - Pirâmide invertida do JD.....	28
<b>Figura 3</b> - Frame do vídeo da reportagem ‘4 em cada 10 abortos legais do Brasil são feito fora da cidade onde a mulher mora; pacientes percorreram mais de 1 mil km’.....	34
<b>Figura 4</b> - Mapa da reportagem ‘4 em cada 10 abortos legais do Brasil são feito fora da cidade onde a mulher mora; pacientes percorreram mais de 1 mil km’.....	35
<b>Figura 5</b> - Página ‘Amazônia Minada’.....	36
<b>Figura 6</b> - Mapa da matéria ‘Aquazônia’.....	38
<b>Figura 7</b> - Visão de uma pesquisa no BotPonto.....	39
<b>Figura 8</b> - Recurso visual para a exemplificação de dados na reportagem ‘Quase metade das escolas do Rio viveu momentos de medo com tiroteios nas proximidades’.....	47
<b>Figura 9</b> - Informações referentes à candidatura de Ciro Gomes à presidência em 2022, na página ‘Políticos do Brasil’.....	48
<b>Figura 10</b> - Visão geral do gráfico ‘Gênero e raça na ciência’.....	51
<b>Figura 11</b> - Divisão de faixa-etária entre as mulheres brancas que cursam Ciências Agrárias.....	52
<b>Figura 12</b> - Gráfico da matéria ‘O avanço do mercado das submetralhadoras artesanais (e ilegais) no Brasil’.....	55
<b>Figura 13</b> - Ilustração principal da matéria ‘Tá tudo dominado. Exclusivo: as milícias assumiram o controle do Rio de Janeiro’.....	57
<b>Figura 14</b> - Gráfico com número de denúncias por mês relacionadas a milícias e facções....	58
<b>Figura 15</b> - Gráfico de cidades com maior número de ocorrências nas cidades do RJ.....	59
<b>Figura 16</b> - Visualização de dados com a porcentagem de denúncias das facções e das milícias nas cidades do RJ.....	59
<b>Quadro 1</b> - Lista dos vencedores do Prêmio Cláudio Weber Abramos de Jornalismo de Dados 2022.....	31



## LISTA DE ABREVIATURAS

ANM	Agência Nacional de Mineração
BD	Base de Dados
CNPq	Conselho Nacional de Desenvolvimento Científico e Tecnológico
CPD	Centro de Processamento de Dados
IA	Inteligência Artificial
IIAA	Índice de Impacto das Águas da Amazônia
JD	Jornalismo de Dados
JDBD	Jornalismo Digital em Bases de Dados
JGD	Jornalismo Guiado por Dados
LAI	Lei de Acesso à Informação
PF	Polícia Federal
PSOL	Partido Socialismo e Liberdade
RAC	Reportagem Assistida por Computador
STF	Supremo Tribunal Federal
SUS	Sistema Único de Saúde
TI	Tecnologia da Informação
TSE	Tribunal Superior Eleitoral

## SUMÁRIO

<b>1 INTRODUÇÃO</b> .....	8
<b>2 UM PASSEIO PELAS PRÁTICAS QUE ENVOLVEM DADOS</b> .....	13
2.1 O que são dados?.....	13
2.2 Internet das Coisas, Plataformização e Algoritmos.....	15
2.3 Falta de privacidade e racismo.....	17
<b>3 A CULTURA DE DADOS NO JORNALISMO</b> .....	21
3.1 Primeiros contatos.....	21
3.2 Jornalismo de Dados.....	24
3.3 O papel dos algoritmos.....	28
3.4 O JD na prática.....	30
<b>4 ASPECTOS METODOLÓGICOS</b> .....	42
<b>5 ANÁLISE</b> .....	46
5.1 Descrição dos casos.....	46
5.2 Comparação e resultados.....	60
<b>6 CONSIDERAÇÕES FINAIS</b> .....	63
<b>REFERÊNCIAS</b> .....	65

## 1 INTRODUÇÃO

Desde a década de 1960, quando surgiram os primeiros Centros de Processamento de Dados (CPD) e, principalmente, a partir dos anos 1990, momento da grande popularização da internet, a quantidade de dados armazenados em mídias digitais tem crescido exponencialmente. Para se ter uma dimensão, em 2020, as informações produzidas digitalmente já chegavam a 40 zb (*zettabytes*, equivalente a  $10^{21}$  bytes). A previsão é de que, até 2025, esse número tenha aumentado para, aproximadamente, 175 zb (BARRACHINA e O'DRISCOLL, 2014 apud SANTAELLA e KAUFMAN, 2021, p. 26).

Considerando que o jornalismo é uma profissão baseada na coleta e na divulgação de informações, esta tendência, chamada de *Big Data*, representa um aumento significativo nas fontes disponíveis para a apuração das matérias. Mais do que isso, novas pautas podem ser encontradas entre os dados de empresas e instituições públicas.

Ao mesmo tempo em que o jornalismo tradicional<sup>1</sup> incorporou a datificação às suas rotinas produtivas surgiram veículos destinados apenas a encontrar e reportar, dentro desses sistemas, assuntos de interesse público. Entre as suas denominações estão: Jornalismo de Dados (JD), Jornalismo Digital em Bases de Dados (JDBD) e Jornalismo Guiado por Dados (JGD). Poucas diferenças separam os três tipos, sendo o primeiro definido, “de forma mais abrangente, como o jornalismo que trata grandes quantidades de dados, o segundo aparece muitas vezes como o jornalismo centrado nos dados, por oposição ao terceiro, que utiliza dados para ‘chegar’ a histórias e enriquecê-las” (MARTINHO, 2014, p. 68).

Contudo, a proliferação desse tipo de fonte não significa mais facilidade para o repórter. Isso porque, apesar de numerosos, os dados digitais não vêm em documentos compilados prontos para a leitura. Eles precisam, primeiramente, ser encontrados, entre milhões de outros. Depois, eles ainda têm que ser extraídos, organizados e interpretados. No ritmo humano, dependendo da extensão dos bancos de informações em que a pesquisa é realizada, pode demorar de horas a anos para que a tarefa seja completada.

Tendo isso em vista, a Inteligência Artificial (IA) passou a ser a grande protagonista dos processos tecnológicos, tornando exequível as pesquisas por elementos específicos dentro das bases de dados, além da mecanização de outras atividades. Essa tecnologia compreende a criação de *softwares* que, alimentados (também) com dados, podem reconhecer padrões e realizar previsões, possibilitando o aprendizado de máquina. Como todo programa

---

<sup>1</sup> Aqui, estamos tratando o jornalismo tradicional como aquele que produz notícias de um forma usual, ou seja, por meio unicamente de entrevistas e análises de documentos.

computacional, ela é baseada em algoritmos, que, neste caso<sup>2</sup>, dizem respeito aos passos indicados em linguagem de programação para o seu funcionamento. Assim, programação e Ciências de Dados somam-se ao rol dos requisitos para a atuação de repórteres e, conseqüentemente, apresentam tarefas a mais para serem realizadas no dia a dia das redações.

Sobre isso, é preciso ressaltar que a literatura sobre mídias digitais, independente de usarem ou não as bases de dados, sempre denunciou a sobrecarga de trabalho que um profissional desse meio está propenso a enfrentar. A principal razão para isto é que a mudança nas formas de comunicação levou ao surgimento de novas atividades, como a moderação de comentários e a gerência de redes sociais, que, na ausência de especialistas para executá-las, tornaram-se responsabilidade do jornalista (BASTOS, 2012, p. 286-287). Então, esse profissional “passa a se submeter à lógica imaterial da tecnologia (mais difundida nos meios visuais) e a se adaptar à completa volatilização do ambiente de trabalho” (MARCONDES FILHO, 2000, p. 31).

Desta forma, não é difícil imaginar como o método de apuração baseado em dados pode comprometer ainda mais a rotina já atarefada de quem trabalha com ciberjornalismo. Por outro lado, esse conhecimento pode ser uma solução para as dinâmicas de *deadline* contínuo, que fizeram os repórteres deixarem de lado as reportagens de cunho investigativo e autoral (BASTOS, 2012, p. 286-287). Isto é, apesar de constituir uma tarefa a mais, essa é uma maneira de retomar a tradição de matérias aprofundadas e de apurações minuciosas, perdida com a valorização de um profissional “ágil, mecânico, maquínico” e que seja “acoplável a qualquer parte do sistema de produção” (MARCONDES FILHO, 2000, p. 36).

Além do problema da sobrecarga, existem outras questões inerentes ao *Big Data* e ao uso de algoritmos e IA que precisarão ser reconhecidas e combatidas se o jornalista deseja manter uma abordagem ética. Para começar, será preciso entender que os dados estão sempre imersos em vieses (PASQUINELLI e JOLLER, 2021, p. 1264). Isso significa que as tendências pessoais dos programadores, que, inevitavelmente, vivem em sociedades preconceituosas, são reproduzidas. Seja na escolha de quais dados serão utilizados, na sua interpretação ou na programação das máquinas, todo o processo é permeado pela visão de uma ou mais pessoas. Portanto, nenhum dado é bruto e não deve ser tratado como tal.

Como consequência dos vieses, a discriminação de gênero, raça e classe é amplificada pelos algoritmos (PASQUINELLI e JOLLER, 2021, p. 1275). Noble (2018) chama esse cenário de opressão algorítmica. Ela esclarece que o primeiro passo para formular uma

---

<sup>2</sup> Em outras ocasiões, o termo também está relacionado, popularmente, aos critérios de exibição de conteúdo das plataformas virtuais, com base nas interações (ou nos dados que elas geram) dos usuários.

definição desse conceito é saber que as “fórmulas matemáticas que conduzem decisões automatizadas são feitas por seres humanos [...] as pessoas que tomam essas decisões possuem todos os tipos de valores, muitos dos quais promovem abertamente racismo, sexismo e falsas noções de meritocracia...” (p. 1, tradução nossa)<sup>3</sup>.

Para pensar sobre esses desafios e, principalmente, como enfrentá-los, esta pesquisa investiga delineamentos específicos das práticas de dados em processos jornalísticos, a partir da experiência do Volt Data Lab. O objetivo é identificar se e como a sua contribuição em outros veículos apresenta maneiras inovadoras de lidar com a transmissão de vieses, a opressão algorítmica e outros problemas encontrados na pesquisa. A partir disso, supõe-se que será possível refletir sobre aspectos como a necessidade de tempo para se dedicar à produção — possibilitado pela repartição das tarefas — e de domínio total das Ciências de Dados para superar os dilemas apresentados.

Criado em 2014, o Volt Data Lab, uma “agência de pesquisas, análises e estudos orientados por dados que atua no setor de jornalismo, mídia e comunicação”<sup>4, 5</sup>, teve como ponto de partida o projeto *A Conta dos Passaralhos*, que agrupou em visualizações de dados os números relativos às demissões de jornalistas desde 2012 (e até 2018)<sup>6</sup>. Desde então, a instituição presta serviços para veículos de comunicação que vão desde o aconselhamento sobre o uso de dados até o desenvolvimento de APIs<sup>7</sup> ou a raspagem de dados para a produção de matérias. A agência também possui seus próprios projetos, como o *Atlas da Notícia*, um mapeamento anual de veículos de jornalismo local<sup>8</sup>; o *Núcleo Jornalismo*, uma mídia que produz investigações com base em dados<sup>9</sup>; e o *Science Pulse*, um aplicativo que acompanha as conversas de cientistas na rede social Twitter para saber quais os assuntos mais relevantes dentro do tema<sup>10</sup>.

A investigação dessa mídia será possível através do método de estudo de caso explanatório, que busca responder questões de “como” ou “porquê” sobre um fenômeno e

---

<sup>3</sup> “...mathematical formulations to drive automated decisions are made by human beings. [...] The people who make these decisions hold all types of values, many of which openly promote racism, sexism, and false notions of meritocracy...”

<sup>4</sup> VOLT Data Lab. Disponível em: <<https://voltdata.info/>>. Acesso em 14 de maio de 2023.

<sup>5</sup> Durante a produção da pesquisa, a agência transmitiu as funções de criação de tecnologias para uma ramificação da empresa, chamada Núcleo Tech. O Volt Data Lab passou a atuar apenas com consultoria. Como a essência do trabalho é analisar o que foi feito pela agência anteriormente, não houve a necessidade de inserir essas informações no corpo do texto.

<sup>6</sup> A CONTA dos Passaralhos. Disponível em: <<https://passaralhos.voltdata.info/>>. Acesso em 14 de maio de 2023.

<sup>7</sup> Sigla de *Application Programming Interface*, um padrão de funções realizadas por um programa e que são tornadas disponíveis para outros softwares que possuem objetivos similares. É uma forma de agilizar processos de programação.

<sup>8</sup> ATLAS da Notícia. Disponível em: <<https://www.atlas.jor.br/>>. Acesso em 14 de maio de 2023.

<sup>9</sup> NÚCLEO Jornalismo. Disponível em: <<https://nucleo.jor.br/>>. Acesso em 14 de maio de 2023.

<sup>10</sup> SCIENCE Pulse. Disponível em: <<https://sciencepulse.org/>>. Acesso em 14 de maio de 2023.

descobre as razões do seu sucesso ou da sua falha (CHOPARD e PRZYBYLSKI, 2021, p. 2). A escolha do estudo de caso como uma estratégia de análise apoia-se no objetivo de “estipular um conjunto de elos causais” (YIN, 2001, p. 140). Para além do Volt Data Lab, o nosso caso (que precisa ser um objeto com finitude, de acordo com Stake (2003)) será a sua cooperação com outros veículos, permitindo entender como a mudança nos processos conduzida pela especialização de parte do fluxo de trabalho se liga aos resultados, ou seja, às produções.

Quanto ao recorte, foram escolhidas cinco matérias coproduzidas com o Volt Data Lab. Os critérios para a escolha foram a disponibilidade das páginas onde elas estão publicadas, a variedade de veículos e de tipos de produtos e a quantidade de conteúdo para ser analisado em cada produção. São elas:

- Quase metade das escolas do Rio viveu momentos de medo com tiroteios nas proximidades, reportagem televisiva do *Jornal da Globo*<sup>11</sup>;
- Políticos do Brasil, página de aplicação no site do veículo *Poder 360*<sup>12</sup>;
- Gráfico: Gênero e raça na ciência brasileira, da revista *Gênero e Número*<sup>13</sup>;
- O avanço do mercado das submetralhadoras artesanais (e ilegais) no Brasil, da *BBC News Brasil*<sup>14</sup>;
- Tá tudo dominado. Exclusivo: as milícias assumiram o controle do Rio de Janeiro, reportagem do *The Intercept Brasil*<sup>15</sup>.

Como se trata de produções de diversos veículos, com as suas próprias características, apuração e contribuição, será crucial utilizar a descrição na página do Volt Data Lab sobre qual foi o papel da agência em cada uma delas.

É importante destacar que todo este trabalho, não apenas o capítulo de análise, faz parte do estudo de caso, já que, considerando os passos sugeridos por Yin (2001), a seleção do que será estudado e a formulação de questões para desenvolver a investigação fazem parte da abordagem metodológica. Porém, duas etapas muito importantes, a “lógica que une os dados às proposições” e os “critérios para interpretar as descobertas”, ocupam a maior parte desta

---

<sup>11</sup> SQUASE metade das escolas do Rio viveu momentos de medo com tiroteios nas proximidades. **Jornal da Globo**. 17/04/2018. Disponível em: <<https://globoplay.globo.com/v/6670902/>>. Acesso em: 11 jun 2023.

<sup>12</sup> POLÍTICOS do Brasil. **Poder 360**. Disponível em: <<https://eleicoes.poder360.com.br/>>. Acesso em: 11 jun 2023.

<sup>13</sup> ASSIS, Carolina. Gráfico: Gênero e raça na ciência brasileira. **Gênero e Número**. 20/06/2018. Disponível em: <<https://www.generonumero.media/reportagens/grafico-genero-e-raca-na-ciencia-brasileira/>>. Acesso em 11 jun. 2023.

<sup>14</sup> BODENMÜLLER, Luiza; SPAGNUOLO, Sérgio. O avanço do mercado das submetralhadoras artesanais (e ilegais) no Brasil. **BBC News Brasil**. 15/04/2019. Disponível em: <<https://www.bbc.com/portuguese/brasil-47858443>>. Acesso em: 11 jun. 2023.

<sup>15</sup> OLLIVEIRA, *et al.* Tá tudo dominado. Exclusivo: as milícias assumiram o controle do Rio de Janeiro. **The Intercept Brasil**. 05/04/2018. Disponível em: <<https://www.intercept.com.br/2018/04/05/milicia-controle-rio-de-janeiro/>>. Acesso em: 11 jun. 2023.

pesquisa, tratando-se do referencial que mostra como funciona o *Big Data*, quais os problemas relacionados a ele, como é o padrão das produções dos veículos de JD que não utilizam o suporte de uma agência e quais são as atuais associações entre jornalismo e algoritmos. A parte teórica ajudará a identificar elementos importantes desses contextos, guiando a análise das matérias que, depois de uma descrição detalhada, serão comparadas com algumas iniciativas de JD que não tiveram envolvimento do Volt Data Lab, permitindo notar a mudança gerada pela contribuição do nosso objeto.

Em relação ao modo como a análise será efetuada, buscaremos observar os tipos de bases de dados utilizadas, qual o papel dos algoritmos nas produções, quais são as estruturas de visualizações de dados presentes e no que a escolha de como representar os dados implica, a presença de interatividade, como os dados são contextualizados e a representação de preconceitos e de estereótipos através dos vieses algorítmicos.

Os capítulos 2 e 3 trazem o referencial teórico que trata dos conceitos de dados, *Big Data*, algoritmos e suas relações. Discute-se, nesta perspectiva, como o jornalismo evoluiu no uso de máquinas até chegar às iniciativas com dados e algoritmos e como podemos defini-las. Essas duas abordagens são necessárias para identificar o que é próprio ou não do Volt Data Lab e inserir a agência em um contexto maior. O quarto capítulo, por sua vez, apresenta de forma mais detalhada a construção do estudo de caso, identificando os autores que serão utilizados para nortear o nosso enfoque, a delimitação das questões que devem ser respondidas posteriormente e a justificativa para cada escolha feita. Por fim, o capítulo 5 traz a análise do nosso objeto de estudo e a comparação com outros casos relevantes.

## 2 UM PASSEIO PELAS PRÁTICAS QUE ENVOLVEM DADOS

### 2.1 O que são dados?

A concepção de dado está popularmente associada a uma ideia de fato, de uma prova incontestável. Como Strasser e Edwards (2017) nos lembram, dizer que possui dados geralmente é suficiente para que um dos lados em um debate possa provar o seu ponto (p. 229). O conceito também está ligado a algo básico, fundamental, como se fosse a pormenorização do conhecimento sobre qualquer coisa. No campo científico, diferentes disciplinas podem nomear desta forma critérios de pesquisa e observação que concernem apenas àquela área específica.

Sobre o seu uso na ciência, Huaylupo (2008) diz que eles sempre foram “o modo através do qual as pessoas e pesquisadores objetivam as suas intuições, percepções, vivências e relações com seus meios sociais e científicos”<sup>16</sup> (p. 128, tradução nossa). A relação de pessoalidade no cerne dos dados científicos desmente a noção popular de que eles são equivalentes aos fatos e nunca mentem. Afinal, eles são construções baseadas no contexto do campo de conhecimento a que são atribuídos e das pessoas pelas quais transitam. Por serem particulares a esses critérios, é difícil construir uma definição unitária, trazendo subjetividade para a sua produção.

Como os dados se caracterizam, então? Alguns atributos parecem ser intrínsecos a eles. A começar, eles são o primeiro momento de transformação da natureza dentro da cadeia de conhecimento. Ou seja, antes de todas as fases da produção científica, o primeiro passo é fazer com que a natureza possa ser entendida por meio de um ou mais dados. Em segundo lugar, eles sempre possuem materialidade. Seja nos papéis que contém números e textos ou nos circuitos eletrônicos que processam os zeros e uns que formam as bases digitais, todo dado pode ser tocado, deslocado e armazenado em um local físico. Por último, como tudo o que é material, eles também são mensuráveis, sendo a forma como serão medidos referente a onde estão armazenados e a partir de qual ponto de vista eles serão analisados (STRASSE e EDWARDS, 2017, p. 329-332).

A noção de dados digitais (que são os que interessam a esta pesquisa) diz respeito à sua representação eletrônica, já que a palavra digital ganhou a conotação de tudo o que é processado por um computador. Apesar disso, muitas informações presentes em bases

---

<sup>16</sup> “...el medio a través del cual las personas e investigadores han objetivado sus intuiciones, percepciones, vivencias y relaciones con sus medios sociales y científicos.”



armazenadas digitalmente foram originalmente gravadas em papel e posteriormente transferidas para um computador (STRASSE e EDWARDS, 2017, p. 332). Contudo, cada vez mais os registros cotidianos, como gastos em lojas e consultas em médicos, passaram a ser armazenados em forma de dados nativos digitais. Além disso, a interação entre humanos e máquinas também se tornou fonte para o abastecimento dessas bases, assim como foram criados serviços cuja gênese é a coleta de dados pessoais.

A agregação desses “pequenos dados” em gigantescas bases digitais recebe o nome de *Big Data* (LUPTON, 2018, p.1). O fenômeno está, portanto, diretamente ligado à característica de mensurabilidade desse elemento, que, neste caso, pode dizer respeito aos *bytes* (conjunto de oito *bits*, menor unidade de informação processada pelos computadores) ou à quantidade de pesquisas feitas no Google, vídeos subidos para o Youtube e outros meios de gerar dados online. Como mostrado por Barrachina e O’Driscoll (2014), as informações do mundo dobram a cada dois anos e, em 2020, já eram aproximadamente 40 zb, enquanto em 2025 será de cerca de 175 zb (apud SANTAELLA e KAUFMAN, 2021, p. 216).

Mais do que um nome para uma grande quantidade de dados, o *Big Data* diz respeito ao fenômeno que leva à sua grande produção e armazenamento, possibilitando a criação de modelos de negócio orientados por esse tipo de “depósito de conhecimento”. Além disso, o processo está tão enraizado em todas as esferas do cotidiano que as lógicas de relacionamento e trabalho estão sendo adaptadas para enfrentar o cenário. Lidar com a gigantesca (e crescente) quantidade de dados disponíveis online não é uma tarefa exequível por humanos. Dar conta do serviço requer o uso de técnicas de IA baseadas no aprendizado de máquina, que, além de automatizarem os fluxos de trabalho, também permitem realizar previsões com base em padrões passados (SANTAELLA e KAUFMAN, 2021, p. 216-217).

Para demonstrar a relevância da IA nos processos que regem o *Big Data*, Pasquinelli e Joler (2020) apresentam o conceito de “nooscópio”, uma analogia com o fato de que em eras de importantes novidades tecnológicas foram inventados objetos que permitiram ver o que o olho nu não poderia (o telescópio e o microscópio). Etimologicamente, *noos* é a palavra grega para conhecimento, enquanto *skopein* significa “para olhar”. Ou seja, a IA alcança o patamar de uma ferramenta para observar o conhecimento, que, embaralhado em centenas de milhares de dados e imerso em configurações muito específicas de linguagem de computador, se torna incompreensível para o olho humano. Logo, a IA é quase tão importante para a captação do conhecimento quanto os dados.

Tendo isso em vista, o alastramento do aprendizado de máquina (assim como das bases de dados) para facilitar o trabalho em praticamente todas as áreas é apenas natural. E o

jornalismo, tendo como base a promoção do conhecimento de um recorte de fatos cotidianos, não teria outro caminho senão a adoção desse “nooscópio” para aprofundar a compreensão de informações e, conseqüentemente, a dimensão das notícias e reportagens. De certa maneira, o JD é o reconhecimento e o acompanhamento das características da produção e assimilação do conhecimento e de tudo que liga-se a ele (como a informação) na atualidade. Resta a este trabalho entender de que forma o *Big Data* é incorporado às práticas jornalísticas.

## 2.2 Internet das Coisas, Plataformização e Algoritmos

Parte de entender como as bases de dados foram geradas e, a partir disso, compreender quais são as fontes do JD, é conceber que a gênese do *Big Data* está na interação via tecnologias virtuais entre atores humanos e não humanos.

Atores humanos contribuem com as grandes bases de dados quando eles participam de atividades como fazer ligações e usar aplicativos em celulares, usando motores de busca online como o Google, comprando produtos ou serviços online ou participando de programas de fidelidade, enviando contribuições para plataformas de mídias sociais, usando *wearables* rastreáveis ou se movendo por espaços que possuem sensores digitais e dispositivos de gravação. Atores não humanos participam da criação de dados digitais através de tecnologias como aparelhos rastreáveis, da ‘agricultura inteligente’, em que plantas e animais são equipados com sensores, da ‘cidade inteligente’, que inclui tecnologias como câmeras CCTV e sistemas de gerenciamento de trânsito baseado em sensores e a ‘casa inteligente’, que envolve utensílios domésticos produzindo e compartilhando dados sobre os hábitos e movimentos dos seus moradores (MICHAEL e LUPTON, 2015, p. 1-2, tradução nossa)<sup>17</sup>.

A interatividade do *Big Data* abrange também as trocas entre humanos e outros humanos e entre objetos e outros objetos. Nesse sentido, a fase atual da internet, conhecida como Internet das Coisas, tem como característica a computação ubíqua, em que os dispositivos digitais estão conectados permanentemente em todos os lugares, fazendo com que os objetos assumam o controle em ações do dia a dia (SANTAELLA, *et al*, 2013, p. 28). Podemos inferir que a Internet das Coisas pode ser tratada tanto como um estímulo quanto como uma consequência da massificação da produção e do armazenamento de dados. Isso porque o conhecimento que compõem as bases digitais acerca dos espaços, dos usuários e das interações entre eles é necessário para que os objetos possam se articular sem tanta

---

<sup>17</sup> “Human actors contribute to big datasets when they engage in activities such as making calls and using apps on mobile phones, using online search engines such as Google, purchasing goods or services online or taking part in customer loyalty programmes, uploading contributions to social media platforms, using wearable self-tracking devices or moving around in spaces that are equipped with digital sensing or recording devices. Nonhuman actors participate in the generation of digital data via technologies such as self-tracking devices, ‘smart agriculture’, in which plants and animals are fitted with sensors, the ‘smart city’, which includes technologies such as CCTV cameras and sensor-based traffic management systems and the ‘smart home’, involving domestic appliances producing and sharing data on the habits and movements of their occupants.”

interferência humana. Ao mesmo tempo, a conexão entre os objetos gera cada vez mais dados, articulando o cenário de *Big Data*.

O mesmo pode ser dito sobre a plataformização, que, segundo Poell, Nieborg e Van Dijck (2020), se trata da “penetração de infraestruturas, processos econômicos e estruturas governamentais de plataformas em diferentes setores econômicos e esferas da vida” (p. 5). Ainda segundo os autores, as plataformas possuem três dimensões institucionais, que seriam: infraestrutura de dados, mercados e governança.

A primeira dimensão diz respeito à transformação de processos em dados e à coleta de metadados comportamentais dos usuários, também chamada de “datificação”. A segunda se refere mais literalmente a uma ideia de plataforma, ligando empresas aos usuários finais. Esse modelo é conhecido como mercado de dois lados ou bilateral e depende das práticas dos usuários finais, já que a competição acirrada faz com que as empresas precisem coletar dados para oferecer um produto que agrade aos consumidores. Quanto à última dimensão, a governança, significa que as plataformas guiam a interação entre os usuários através dos recursos disponibilizados (como os botões de curtir e compartilhar, por exemplo). Essa dimensão é exercida por meio da classificação algorítmica, ou seja, a categorização de quais dados são importantes para serem vistos e quais não são (POELL, NIEBORG e VAN DIJCK, 2020, p. 5-8).

É importante ressaltar que a IA é o que possibilita que as plataformas efetivamente utilizem a grande quantidade de dados à disposição. Santaella *et al* (2013) descreve como as informações produzidas ficariam inutilizáveis sem o recurso de *machine learning*:

A Amazon não pode codificar os gostos do conjunto de seus clientes em um programa de computador, assim como o Facebook desconhece como escrever um programa para identificar as melhores atualizações no Feed de Notícias. A Netflix pode ter cem mil títulos de DVD em estoque, mas se os clientes não souberem como encontrar suas preferências de nada adianta. (SANTAELLA *et al*, 2013, p. 4)

Desta forma, os algoritmos, essência da IA e do aprendizado de máquina, possuem um papel essencial no funcionamento das bases de dados, sendo o que as organiza e as alimenta ao mesmo tempo. Também estão embutidos nos aparelhos eletrônicos e estabelecem a conexão entre máquina/máquina, máquina/indivíduo e indivíduo/indivíduo (SILVA, 2017, p. 33). Assim, os algoritmos determinam o modo de funcionamento das plataformas por meio da classificação dos dados, indicando os assuntos e posts mais acessados, mostrando pessoas e temas que podem ser relevantes aos usuários e apresentando propagandas que dizem respeito aos hábitos de cada pessoa.

Isso tudo mostra como, no cenário atual, os “dados são a primeira fonte de valor e inteligência. Algoritmos são a segunda; eles são as máquinas que computam tal valor e inteligência em um modelo”<sup>18</sup> (PASQUINELLI e JOLLER, 2020, p. 1266, tradução nossa).

### 2.3 Falta de privacidade e racismo

Um dos problemas que surgem com a plataformização e as práticas de se apoiar na coleta de dados como modelo econômico e força motriz de serviços e produtos é que grande parte das informações colhidas pelos algoritmos não são concedidas pelos usuários. Na verdade, são concedidas através dos contratos de usuários, mas, em razão da opacidade do modelo de negócio, isso não fica claro: ou seja, o usuário disponibiliza as informações com seu consentimento, mas não sabe disso. Mais do que isso, as pessoas envolvidas não possuem sequer o poder de escolha sobre como os seus comportamentos serão afetados por essa dinâmica (PANGRAZIO e SEFTON-GREEN, 2020, p. 208).

Os usuários de plataformas podem emitir ou serem usados como dados sem ao menos saber, ficando sem acesso a eles, enquanto as empresas os utilizam para criar produtos e veicular publicidades adaptadas de acordo com o comportamento público. A essa economia baseada na compra e venda de informações sobre usuários, Pasquinelli e Joler (2021) dão nome de “extrativismo do conhecimento”. Os autores atribuem ao *Big Data* as formações de monopólios de IA que, como em qualquer outra economia extrativista, se apoderam de recursos cuja posse é indefinida.

Neste ponto, começam a surgir questionamentos éticos para a atividade do JD, já que, acostumados a pedir referências diretamente da fonte ou retirá-las de documentos acessíveis, os repórteres costumam contar com a permissão de uso. Raros casos, como câmeras escondidas, são, geralmente, apoiados pelo critério de interesse público. O *Código de Ética dos Jornalistas Brasileiros* (2007), por exemplo, informa no artigo 11, inciso III, que o profissional de jornalismo não pode divulgar informações obtidas “com o uso de identidades falsas, câmeras escondidas ou microfones ocultos, salvo em casos de incontestável interesse público e quando esgotadas todas as outras possibilidades de apuração”.

De maneira ligeiramente diferente, Zuboff (2021) aborda a economia voltada à extração de dados colocando no centro das discussões a falta de privacidade da era digital. No contexto do Capitalismo de Vigilância (termo que a autora utiliza para explicar esse cenário),

---

<sup>18</sup> “Data are the first source of value and intelligence. Algorithms are second; they are the machines that compute such value and intelligence into a model.”

como em um livro de distopia, todos os nossos movimentos estão sendo vigiados. Pode parecer uma noção exagerada, mas ao listar os diversos meios através dos quais as grandes empresas de tecnologia, conhecidas como *big techs*, conseguem extrair informações pessoais em um único dia de um único indivíduo, torna-se nítida uma realidade em que o cidadão comum perdeu também o direito à sua privacidade.

O conhecimento sobre comportamentos tornou-se um bem maior do que os eletrônicos e os *softwares*. Muitos dos produtos lançados pelas *big techs* visam criar formas de ter acesso a novos dados.

Se o Google é uma empresa de busca, por que está investindo em dispositivos de *smart home*, dispositivos inteligentes feitos para serem vestidos e carros autodirigidos? Se o Facebook é uma rede social, por que está desenvolvendo drones e realidade aumentada? (ZUBOFF, 2021, p. 163).

Essa situação gera conflitos de cunho democrático e até existencial. De acordo com Véliz (2020), um dos problemas da lógica dessa economia é que a privacidade, assim como o conhecimento, é uma forma de poder, já que ao ter detalhes sobre a vida de uma pessoa é possível antecipar os seus movimentos e influenciá-la (p. 28). Isso faz sentido, considerando que o objetivo das propagandas personalizadas, um dos principais destinos das informações coletadas, é descobrir gostos para, então, prever e influenciar o consumo. Zuboff (2021) é ainda mais subjetiva ao falar que o Capitalismo de Vigilância leva a uma perda do “direito a um tempo futuro”, do “direito a santuários” e do “direito de ser esquecido” (p. 75).

Não é possível que os jornalistas (ou cientistas de dados e técnicos de TI que trabalham no ramo jornalístico) mudem um sistema econômico de alcance global. Porém, o motivo pelo qual as questões relacionadas à privacidade são consideradas importantes nesta pesquisa é para que, através da análise do nosso objeto de estudo em um capítulo posterior, seja possível questionar o papel de quem trabalha nesse contexto em relação ao conflito de interesses na procedência das informações.

Por fim, há também a preocupação sobre como os dados serão usados e como a sociedade os vê. Porque, como foi apresentado, todo dado, criado com base em interações digitais ou não, é uma construção da realidade da forma que ela é percebida por quem o coletou. “Considerar o dado por si próprio como importante ou imprescindível é torná-lo independente, não apenas da realidade de qual faz parte, mas também dos sujeitos que o construíram para usos específicos”<sup>19</sup> (HUAYLUPO, 2008, p. 129, tradução nossa).

No caso da aplicação da IA e do aprendizado de máquina, é imprescindível pensar em quem programou o código e em que contexto essa pessoa está inserida. Como tudo o que é

---

<sup>19</sup> “Considerar al dato por sí solo como importante o imprescindible, es independizarlo, no sólo de la realidad de la cual forma parte, sino también de los sujetos que lo han construido para usos específicos.”

feito por humanos, essas técnicas estão imersas em vieses, assim como erros, limitações, aproximações, falácias e vulnerabilidades (PASQUINELLI e JOLER, 2021, p. 1265). Utilizando a metáfora do nooscópio, é possível pensar no aprendizado de máquina como as lentes que aumentam a probabilidade de reconhecer padrões, porém, para entender o seu impacto, é necessário descobrir em que medida essas lentes também distorcem os resultados através dos vieses (PASQUINELLI e JOLER, 2021, p. 1264).

Essa distorção tendenciosa pode se apresentar em três tipos: os vieses históricos, de conjunto de dados e algorítmicos. O primeiro representa as convenções histórico-sociais. O segundo, como os operadores dos dados transmitem as suas visões de mundo a eles. E o terceiro, a amplificação dos dois outros vieses por meio do aprendizado de máquina e a sua tendência a repetir padrões (PASQUINELLI e JOLER, 2021, p. 1265).

O principal exemplo de como eles atuam na prática são os resultados que até recentemente podiam ser encontrados na página principal e nas sugestões de pesquisa do Google, quando o usuário buscava frases que envolvessem mulheres, negros ou, ainda mais grave, mulheres negras. A especialista em estudos da internet Safiya Noble (2018) percebeu isso ainda em 2010, ao realizar uma pesquisa no motor de busca com o termo “meninas negras” e receber como resultado diversas páginas de sites pornográficos. Mostrando que não se tratava de um caso isolado, seu livro *Algoritmos da Opressão* traz diversos outros experimentos com pesquisas no Google que acabavam demonstrando como os algoritmos desse site reproduzem o racismo e o sexismo da sociedade. A capa da obra na sua edição original, em inglês, é estampada pela pesquisa “Por que mulheres negras são tão...”, a qual os algoritmos da plataforma completam automaticamente com: nervosas, barulhentas, malvadas, atraentes, preguiçosas, irritantes, confiantes, audaciosas e inseguras.

Atualmente, ao realizar uma pesquisa desse tipo no Google, tanto em português como em inglês, esses resultados ou sugestões não são mais encontrados. Não é possível saber quando a falha racista foi mudada, mas, ainda em 2018, a etiquetagem automática de fotos pessoais oferecida pelo Google Fotos, outro aplicativo da empresa, definia imagens de pessoas negras como gorilas. A aparente solução que a plataforma encontrou foi, simplesmente, retirar as palavras gorila, macaco e verbetes sobre outros símios das etiquetas, em vez de retrainar os seus algoritmos para que o caso de racismo não se repetisse<sup>20</sup>. Apesar dos inúmeros erros e da falta de uma atitude corretiva, o Google mantém a sua reputação intacta, até porque as suas práticas algorítmicas, utilizadas em prol do capital neoliberal e das

---

<sup>20</sup> SALAS, Javier. Google conserta seu algoritmo “racista” apagando os gorilas. El País, 16 jan. 2018. Disponível em: <[https://brasil.elpais.com/brasil/2018/01/14/tecnologia/1515955554\\_803955.html](https://brasil.elpais.com/brasil/2018/01/14/tecnologia/1515955554_803955.html)>. Acesso em 20 jan. 2023.

elites estadunidenses, levaram a marca a um estado inatacável pela dúvida (NOBLE, 2018, p. 53).

Além disso, casos ainda piores de racismo são propagados por sistemas de IA à medida que a tecnologia avança. A opressão algorítmica se torna também um risco físico quando a maioria das prisões que levam como base programas de IA são de pessoa negras que nem ao menos cometeram crimes<sup>21</sup>.

Tudo isso parte da tentativa de conceber o coletivo como padrões e os comportamentos sociais como distribuições de probabilidade (PASQUINELLI e JOLER, 2021, p. 1275). Não é possível encontrar outro resultado a não ser que se considere a divergência entre os modelos extremamente estatísticos, carregados de uma objetividade engessada, e os seus objetos de análise, humanos e subjetivos. Para nos livrar de uma visão do mundo centrada em dados, Pasquinelli e Joler recomendam “perceber que não é o modelo estatístico que constrói o sujeito, mas o sujeito que estrutura o modelo estatístico”<sup>22</sup> (2021, p. 1277, tradução nossa).

A questão é que a aplicação de padrões para compreender os indivíduos já está em vigor, gerando um sistema de preconceitos e reduzindo pessoas à característica que é relevante para o uso algorítmico em determinado momento. Uma pessoa pode ser vista como a fumante, a obesa, a amante de animais ou qualquer outra característica que seja útil para as empresas, que, por sua vez, podem tratar de modo diferente um grupo de pessoas ao ter conhecimento de seus hábitos por meio da coleta indevida de dados (SILVA, 2017, p. 35). Considerando o objeto desta pesquisa, as práticas de dados no jornalismo, o cuidado, então, deve ser redobrado ao minerar, narrar ou utilizar dados de terceiros, levando em conta a sua relação com a formação da opinião pública e o véu da objetividade jornalística, que muitas vezes esconde do público o fato de que esse também é um trabalho de escolhas e, portanto, de vieses.

---

<sup>21</sup> FIALHO, Ana Carolina Mendes. Reconhecimento Facial: Entenda o impacto da Inteligência Artificial na vida da população negra. Exame, 4 out. 2022. Disponível em: <<https://exame.com/esg/reconhecimento-facial-entenda-o-impacto-da-inteligencia-artificial-na-vida-da-populacao-negra/>>. Acesso em: 20 jan. 2023.

<sup>22</sup> “...realise that it is not the statistical model that constructs the subject, but rather the subject that structures the statistical model.”

## 3 A CULTURA DE DADOS NO JORNALISMO

### 3.1 Primeiros contatos

Para entender como são realizadas as iniciativas de jornalismo automatizado no Brasil, Dalben (2020) entrevistou alguns profissionais envolvidos na produção de robôs que empreendem análises jornalísticas. É o caso do Guilherme Duarte, do veículo Jota, um dos responsáveis pelo robô Rui Barbot.

No artigo de Dalben (2020), o jornalista conta sobre um evento em que demonstrou, em tempo real, como um software de mineração de dados levava apenas 30 minutos para coletar cinco mil ações de inconstitucionalidade do site do Supremo Tribunal Federal (STF). Segundo ele, uma das mulheres da plateia teria praticamente chorado por ter passado um ano realizando a mesma atividade à mão (DALBEN, 2020, p. 499). Esse caso se torna uma anedota interessante para exemplificar o dilema de um repórter contemporâneo: um mar de informações disponíveis, em contraste com a necessidade de conhecimentos técnicos do que até pouco tempo atrás se tratava da função da equipe de TI, além da coexistência com trabalhos maquinicos que podem cumprir melhor o seu próprio ofício.

Para chegar até essa conjuntura, o jornalismo precisou passar por diversas fases, como degraus sendo galgados, evolução que diz mais respeito ao avanço tecnológico do que a uma aproximação da profissão com as tecnologias. Podemos dizer que, desde a invenção da prensa de Gutenberg, que ajudou a popularizar a prática jornalística no século XVII, a dinâmica jornalismo-tecnologia se manteve, especialmente no que tange o auxílio das máquinas em atividades muito difíceis para serem realizadas por humanos. A execução de inúmeras cópias de um texto de forma rápida, propiciada pela prensa móvel, é de certa forma comparável à utilização de linguagem de máquina para minerar milhares de dados em pouco tempo.

No caso dos computadores, a sua inserção no processo de apuração e escrita dos jornais teve início nos anos 1970, década da invenção do microprocessador, que permitiu o avanço dos computadores pessoais (TEMIN, ROMANI e GIL FILHO, 2014, p. 842). O período é equivalente ao começo do que Marcondes Filho (2000) chama de quarta (e última) geração do jornalismo. Para ele, uma das características da fase é “a substituição do agente humano jornalista pelos sistemas de comunicação eletrônica, pelas redes, pelas formas interativas de criação, fornecimento e difusão de informação” (MARCONDES FILHO, 2000, p. 30).



Pode-se dizer que, assim que os computadores invadiram as redações, os dados digitais começaram a fazer parte da rotina de trabalho dos profissionais da mídia. Tanto que, em 1973, o jornalista Philip Meyer (2002) criou o termo Jornalismo de Precisão para caracterizar uma época em que a energia, a dedicação e o talento para escrita já não eram mais as únicas qualidades que um repórter deveria possuir, necessitando também “ser um administrador de bases de dados, um processador de dados e um analista de dados”<sup>23</sup> (p. 1, tradução nossa).

No livro em que escreveu sobre o tema, Meyer (2002) descreveu cinco atitudes que os repórteres deveriam adotar no tratamento dos dados: coletá-los; armazená-los; recuperá-los; analisá-los; reduzi-los; e comunicá-los (p. 7). A obra ainda é pertinente para os jornalistas de hoje, oferecendo sistemas de como referenciar e analisar números, indicando, por exemplo, que sejam feitas comparações com outros índices para que o leitor possua referências. Ela também possui um capítulo sobre como lidar com bases de dados e a importância do computador na produção da matéria. Por essa razão, Meyer também é conhecido como o criador da Reportagem Assistida por Computador ou RAC (GEHRKE e MIELNICZUK, 2017, p. 4), outro conceito que emergiu nos anos 1970 para explicar as novas técnicas nas redações.

Com o tempo, o uso do computador tornou-se, inevitavelmente, mais frequente nas redações, assim como se popularizou no resto da sociedade. E, nos anos 1990, começaram a surgir os veículos jornalísticos digitais. Isso fez com que os jornalistas precisassem aprender a linguagem específica dos meios digitais, abandonando certas práticas dos veículos impressos. Sobre a denominação dessa mídia com práticas próprias da internet, os pesquisadores falam de jornalismo online (ALVES, 2006; CANAVILHAS, 2003, 2006; DEUZE, 1999), jornalismo digital (DEUZE, 1999; SALAVERRÍA, 2019; SCOTT, 2005), ciberjornalismo (BASTOS, 2005, 2012, 2015; CANAVILHAS, 2006) e webjornalismo (CANAVILHAS, 2003, 2006; MIELNICZUK, 2004; QUADROS, 2005).

Muitas vezes usados como sinônimos, os termos tendem a caracterizar qualquer modelo de produção de notícias e reportagens que utilize recursos digitais ou que possua publicação online, sendo o componente tecnológico o seu definidor (BASTOS, 2005; DEUZE, 1999; SALAVERRÍA, 2019). Canavilhas (2006), porém, enxerga os conceitos como fases de desenvolvimento. Para o autor, o jornalismo online é apenas a utilização dos meios digitais para distribuir conteúdos midiáticos tradicionais. Por outro lado,

---

<sup>23</sup> “...to be a database manager, a data processor, and a data analyst.”

na fase a que chamamos webjornalismo/ciberjornalismo, as notícias passam a ser produzidas com recurso a uma linguagem constituída por palavras, sons, vídeos, infografias e hiperligações, tudo combinado de forma a que o utilizador possa escolher o seu próprio percurso de leitura (CANAVILHAS, 2006, p. 114).

De forma geral, independente da denominação escolhida, entende-se que uma mídia tipicamente online deve possuir integração com os recursos que a internet disponibiliza. Para se constituir como um jornalismo proveniente do meio digital, é preciso que as narrativas façam uso de hipertexto, que são conexões com outros conteúdos a partir de hiperligações; de linguagem multimídia, ou seja, incorporar outras mídias como imagens, vídeos, áudios e links; e de linguagem hipermídia, que significa acrescentar a multimídia através do hipertexto (BASTOS, 2005, 2012; CANAVILHAS, 2003, 2006; MIELNICZUK, 2004).

Os pesquisadores da área também apontam as características de personalização, que diz respeito a não-linearidade das narrativas, permitindo uma participação do leitor na construção das informações; a interatividade, que é a possibilidade de comunicação entre veículo e público; e a convergência, que pode se tratar da união tanto das mídias (texto, som, imagem, etc) quanto das formas de comunicação (BASTOS, 2005; DEUZE, 1999; MIELNICZUK, 2004; QUADROS, 2005).

Como consequência, o jornalismo digital alterou a rotina de produção, as habilidades necessárias para atuar na profissão e, inclusive, os modos de ler notícias. A construção de histórias não-lineares e em camadas e o uso de ferramentas interativas exigem que o jornalista tenha conhecimento não apenas sobre o seu público, mas também sobre a máquina que ele opera. Bastos (2005), diz que:

O ciberjornalista tem, conseqüentemente, de tomar decisões sobre qual o formato ou formatos de media que melhor se adaptam a uma determinada estória (multimédia), de considerar opções que permitam ao público responder, interagir ou mesmo personalizar certas estórias (interactividade), e pensar nas maneiras de relacionar a estória com outras estórias, arquivos, e outros recursos através de hiperligações (hipertexto) (BASTOS, 2005, p. 4).

Somado ao fator instantaneidade, já que as notícias ficam à disposição 24 horas por dia e precisam ser constantemente atualizadas, isso faz com que haja uma sobrecarga nos profissionais, que muitas vezes são jovens e mal pagos (BASTOS, 2012, p. 286). Preza-se mais o jornalista que consegue dar conta de um enorme contingente de trabalho em pouco tempo e os textos se tornam mais curtos (MARCONDES FILHO, 2000, p. 35-36, 44).

Além disso, o rigor de verificação foi abandonado, não por um desvio de moral dos novos profissionais, mas porque os deadlines apertados e a grande quantidade de notícias que precisa ser produzida não deixam espaço para uma apuração meticulosa (BASTOS, 2012, p. 287).

### 3.2 Jornalismo de Dados

A RAC, o Jornalismo de Precisão e, posteriormente, a utilização de uma linguagem específica dos meios digitais pavimentaram o caminho até a criação do modelo jornalístico focado em produzir notícias a partir de e sobre dados. A começar pelo JDBD, termo que coloca em evidência o uso de bases de dados (BD). Aqui, elas não são mais uma ferramenta com a qual é preciso aprender a conviver, mas norteiam um tipo específico de jornalismo.

A forma da notícia, os modos para sua classificação interna e externa, assim como a sua atualização, níveis de articulação com o conteúdo inserido numa BD e posterior recuperação vão requerer outro tratamento, conformado a partir das noções de: resolução semântica, metadados, relato imersivo ou narrativa multimídia, e jornalismo participativo (BARBOSA, 2006, p. 8).

O JDBD compreende que as características do jornalismo digital só possuem funcionalidade quando apoiadas pelas bases de dados como cultura e formato (BARBOSA, 2006, p. 8). O conceito parte de noções apresentadas por Manovich (2001) no livro *The Language of the New Media*. Ao inserir as mídias digitais em uma genealogia midiática complexa, o autor considera as bases de dados como correspondentes da era do computador ao que foi a narrativa para a literatura e o cinema: uma forma de expressão cultural privilegiada (MANOVICH, 2001, p. 218). Assim como as narrativas cinematográficas, as bases de dados apresentam um modelo diferente do que seria o mundo. E, neste caso, ele é representado como uma lista de itens não organizados (MANOVICH, 2001, p. 219, 225).

Por toda essa complexidade, o JDBD acaba sendo mais uma concepção acadêmica para estudar possibilidades para o jornalismo no meio digital do que um fazer jornalístico que se concretiza nas redações. De forma mais abrangente existe o JD e o JGD, que nomeiam um modo de trabalho com dados em que o resultado final será um produto tipicamente jornalístico (MARTINHO, 2014, p. 68).

Na introdução do *Manual de Jornalismo de Dados* (GRAY, BOUNEGRU e CHAMBERS, 2012), que vem sendo utilizado como um guia sobre a atuação dos profissionais da área, Paul Bradshaw dá algumas indicações sobre as particularidades do JD:

O que faz o jornalismo de dados diferente do restante do jornalismo? 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. Estas possibilidades aparecem em qualquer estágio do processo, seja usando programas para automatizar o trabalho de combinar informação do governo local, polícia e outras fontes civis, como Adrian Holovaty fez no ChicagoCrime e depois no EveryBlock; seja usando um *software* para achar conexões entre centenas de milhares de documentos, como o The Telegraph fez com o MPs' expenses (GRAY, BOUNEGRU, CHAMBERS, 2012, p. 8).

Para criar um cenário mais nítido, o livro fez a alguns profissionais a seguinte pergunta: “Por que o Jornalismo de Dados é importante?”. Em geral, as respostas destacavam a relevância do conhecimento sobre o processamento de dados, agora que eles são abundantes; a oportunidade de encontrar histórias únicas e exercer a função de *watchdog*; a necessidade de ter acesso aos dados “brutos” para não depender das interpretações oficiais e a vantagem de conseguir dar sentido aos dados e de encontrar informações que de outra maneira não estariam ao alcance do profissional de comunicação (GRAY, BOUNEGRU e CHAMBERS, 2012, p. 13-20).

Pedro Markun, do Transparência Hacker, ressaltou inclusive a questão do aproveitamento do tempo através do uso de máquinas. A sua resposta aos organizadores do livro afirmava que jornalistas não possuem tempo a perder tentando coletar informações de PDFs, sendo a linguagem de programação uma ferramenta valiosa para a aceleração dos processos na redação (GRAY, BOUNEGRU e CHAMBERS, 2012, p. 9).

Essas declarações mostram uma naturalização da presença dos dados nas redações. É como se os profissionais que já os utilizam em suas narrativas estivessem alertando para a inevitabilidade de recorrer a eles. Dentro do cenário de *Big Data*, todos os jornalistas acabarão evocando as técnicas de mineração, análise e visualização.

Nenhuma das respostas presentes no livro previa um especialista em linguagem de programação para minerar os dados necessários entre milhões de outros. Também não se falou em um designer para criar as visualizações que transmitem as informações no JD. Desta forma, não seria completamente errado dizer que a alta sobrecarga dos repórteres, impulsionada pela digitalização dos veículos, aumenta com o JD. Além de saber realizar escolhas em relação à narrativa hipertextual, é preciso conhecer um pouco de programação e de design. Por outro lado, tanto o exemplo utilizado no início do livro quanto os depoimentos do *Manual de Jornalismo de Dados* indicam que a tecnologia, especialmente a IA, é responsável por agilizar demandas e substituir os repórteres em tarefas muito extensas. Então, o JD apresenta a vantagem de fazer com que os curtos deadlines sejam respeitados sem que se exija demais do jornalista.

A segunda versão do manual, lançada em 2021, desvia o seu foco da apresentação do modelo para tratar de reportagens em que os dados foram usados para denunciar ou combater a fome, a miséria e o tratamento indigno dado às minorias e aos imigrantes. É uma resposta ao fato de que:

Dados não oferecem representações neutras e diretas do mundo, visto que se encontram enredados em meio à política e cultura, ao dinheiro e poder. Instituições e infraestruturas que apoiam a produção de dados — de pesquisas a estatísticas, de

ciência climática a redes sociais — passaram a ser questionadas. Logo, cabe perguntar: quais dados, de quem e por quais meios? Dados sobre quais temas e para que finalidade? Quais temas são ricos em dados e quais são pobres? Quem é capaz de se beneficiar destes? Quais tipos de público são agrupados através dos dados, quais competências são suportadas, quais tipos de política pode sancionar e quais os tipos de participação que engendra? (GRAY, BOUNEGRU e CHAMBERS, 2021, p. 5).

A partir disso, percebe-se que o JD já está imbuído da percepção dos vieses como parte inseparável dos dados e da responsabilidade de utilizá-los justamente para pôr sob a luz histórias ignoradas.

Apesar dos manuais possuírem a intenção de responder as questões mais básicas sobre o JD, a partir deles não é possível conceituar o modelo. Eles não dão sinais, por exemplo, para determinar se todo jornalismo que faz uso das técnicas apresentadas em algum momento passa a ser considerado JD, se o conceito só se concretiza caso essas técnicas sejam utilizadas de forma exclusiva ou se o JD é apenas uma ferramenta para desenvolver certas reportagens. Talvez essa seja a intenção. Como os organizadores explicam no começo da segunda versão: “Não presumimos que o ‘jornalismo de dados’ seja um conjunto unificado de práticas. Trata-se, sim, de uma categoria proeminente ligada a uma série de práticas diversas que podem ser estudadas, definidas e experimentadas de maneira empírica” (GRAY, BOUNEGRU e CHAMBERS, 2021, p. 6).

Mesmo que a definição exata de JD não seja indispensável para os objetivos deste trabalho, entender quais são os processos de apuração e construção da notícia é primordial para que se possa inserir o nosso objeto, o Volt Data Lab, dentro desse sistema. Mais do que isso, a partir desse conhecimento, pode-se avaliar como o apoio da agência impacta o resultado das matérias.

Sobre esse assunto, Vasconcellos, Mancini e Bittercourt (2015) propõem quatro dimensões que devem ser englobadas na construção do produto jornalístico com dados: 1) Busca e/ou elaboração própria dos dados (criação da base); 2) Estrutura da base (séries temporais, categorias rankings, tabelas); 3) Visualização dos dados (infografia); 4) Interpretação dos dados (texto). Os autores utilizam essas etapas de construção de matérias para diferenciar o JD de um jornalismo *com* dados. A reportagem que apresenta todas as características se encontra no nível 1 e pode ser considerada JD. No entanto, aquela que não corresponder a nenhuma dessas dimensões fica no nível 5, sendo tratada pelos autores como jornalismo *com* dados (fig 1).

**Figura 1:** Categoria x Níveis de JD

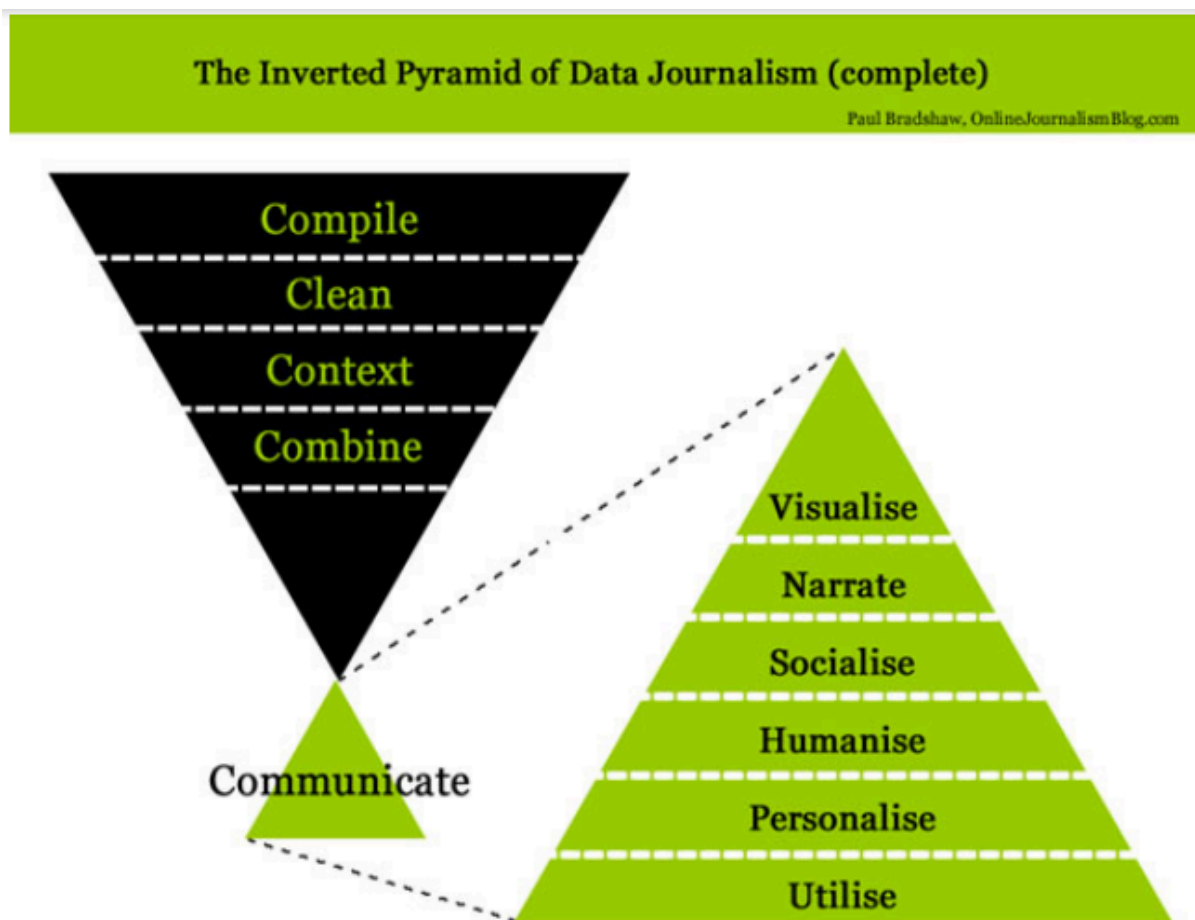
Categorias X Níveis de JD		Busca e/ou Elaboração própria dos dados (criação da base)	Estrutura a base (séries temporais, rankings, tabelas)	Visualização dos dados (infografia)	Interpretação dos dados (texto)
Jornalismo ↑ Com Dados	Nível 1				
	Nível 2				
	Nível 3				
	Nível 4				
	Nível 5				

Fonte: VASCONCELLOS, MANCINI e BITTENCOURT, 2015, p. 17.

É importante ressaltar que a diferenciação entre infografia e texto, colocando o texto como uma linguagem verbal imprescindível para narrar os dados, está equivocada. Isso porque a infografia é, por si só, uma narrativa completa que pode ou não possuir elementos verbais, tratando-se de um tipo textual. Além disso, esse recurso também é uma maneira de transmitir informações complexas que não seriam compreendidas de outra maneira (LIMA, 2015, p. 111). De qualquer modo, essa categorização oferece um panorama do que tem sido considerado importante pela literatura do tema, desde a busca pelas informações à escolha de como elas serão narradas.

Bradshaw (2011), por sua vez, propõe em seu blog outros processos como constituintes do JD. Reestruturando a velha pirâmide invertida para se adequar à apuração com dados, ele hierarquizou as seguintes ações: compilar, limpar, contextualizar, combinar e comunicar. Em um diagrama separado, ele dividiu “comunicar” em uma pirâmide cuja base corresponde a visualizar, seguido por narrar, socializar, humanizar, personalizar e, na base, utilizar (fig 2). Menos limitadora e muito completa (no sentido que prevê as pequenas ações envolvidas na mineração e narração dos dados), essa estrutura deixa à escolha do jornalista o modo pelo qual ele comunicará as informações encontradas. E, mostrando uma articulação com o início do uso dos computadores na redação, a sua esquematização se parece muito com as ideias de Meyer (2002), quando ele sugere coletar, armazenar, recuperar, analisar, reduzir e comunicar.

Figura 2: Pirâmide invertida do JD



Fonte: BRADSHAW, 2011

### 3.3 O papel dos algoritmos

Na lógica da era do computador, o mundo não é reduzido apenas às estruturas de dados, como também aos algoritmos (MANOVICH, 2001, p. 223). São eles que permitem que as informações sejam mineradas e, como apresentado no início do capítulo, colocam nas mãos dos jornalistas cinco mil ações de inconstitucionalidade do STF. Não existe interpretação de dados sem os algoritmos.

Esses intermediários de *software* operam por trás das câmeras e influenciam uma ampla gama de atividades; a seleção de notícias online via motores de busca e agregadores de notícias, o consumo de música e vídeos de entretenimento via sistemas de recomendação, a escolha de serviços e produtos em lojas online, a exibição de mensagens de status nas redes sociais e os robôs de investimento na bolsa de valores são os exemplos mais proeminentes dessa tendência onipresente (SAURWEIN, JUST e LATZER, 2015, p. 35, tradução nossa)<sup>24</sup>.

<sup>24</sup> These software intermediaries operate behind the scenes and influence a wide range of activities; the selection of online news via search engines and news aggregators, the consumption of music and video entertainment via recommender systems, the choice of services and products in online shops, the display of status messages on social online networks and algorithmic trading in stock exchange markets are the most prominent examples of this omnipresent trend.

O JDBD pode ter considerado as bases de dados como centro cultural do jornalismo digital, contudo, são os algoritmos que possibilitam que as narrativas funcionem e que dão perspectivas criativas para a invenção de novos produtos. Mesmo que o termo tenha sido pouco abordado na pesquisa sobre JD, é de algoritmos que falamos quando nos referimos a usar as bases para encontrar novas histórias, exercer a função de *watchdog* e poupar tempo. Os algoritmos revolucionaram o ofício informativo por mecanizar o trabalho intelectual e cognitivo, assim como outras ferramentas mecanizaram o trabalho físico nos séculos XIX e XX (DIAKOPOULOS, 2019, p. 17). É o caso da comparação da prensa móvel com o *software* de mineração de dados.

Ao usar algoritmos como o centro da apuração, um leque de possibilidades é aberto e o significado de produto jornalístico diversifica-se, sem precisar que a coleta de dados tenha como fim uma produção textual (considerando o sentido amplo de texto como o que pode ser lido). Diakopoulos (2019) escreveu, em *Automating the News: how algorithms are rewriting the media*, um capítulo para cada recurso que os jornalistas podem desenvolver usando algoritmos. São eles: mineração jornalística de dados, produção de conteúdo automatizado, *newsbots* (sobre o uso de robôs para gerar conteúdos), jornalheiros digitais (sobre a distribuição de notícias) e relatório de responsabilidade algorítmica (sobre o uso de engenharias reversas para conter os erros causados pelo uso de algoritmos por instituições como a polícia).

Em 2020, Lemos e Domingo organizaram o dossiê *Journalism and Algorithms* para a revista *Brazilian Journalism Research*. Assim como o livro de Diakopoulos (2019), as utilizações dos algoritmos citadas nessas pesquisas oscilavam entre uma ampla gama de assuntos, como automatização de notícias e experiência de usuário (DIERICKX, p. 432-457), novos valores-notícia (SANTOS, p. 458-475), robôs jornalísticos no Twitter (DALBEN, p. 476-501), plataformização (JURNO e ANDREA, p. 502-525), *agenda-setting* guiada por algoritmos (OLIVEIRA e COSTA, p.550-575) e a inserção de notícias em assistentes pessoais (PASE, et al., p. 576-603). Essa relação origina debates labirínticos já que ela diz respeito a criar informação sobre a forma que somos vistos pelo processo de datificação (LEMOS e DOMINGO, 2020, p. 405), que nunca será única ou limitada. O conteúdo das bases de dados, gerado sobre “nós”, pode possuir tantas interpretações quanto o número de pessoas que tentarem analisá-lo e a quantidade de ferramentas que forem usadas com esse fim.

Como vimos no capítulo anterior, podemos ver os algoritmos de duas formas. Na primeira, em seu conceito puro, como um instrumento da Ciência da Computação, base para o



funcionamento de *softwares*. A segunda forma está ligada às práticas de recomendações de acordo com o perfil individual dos usuários das plataformas. Esta definição de algoritmos estende-se para todos os aspectos da vida, fora e dentro das redes. Mas, neste capítulo, trataremos apenas da primeira definição.

Assim, em sua concepção inicial, os algoritmos são como etapas pré-estabelecidas para a resolução de um problema ou para atingir um resultado. São como os passos de uma receita de bolo (DIAKOPOULOS, 2019, p. 16, 33). Envolvidos na receita estão também os utensílios e as ferramentas, que podem se tratar da linguagem de máquina e da IA. Em uma definição ainda mais complexa, Lemos e Domingo (2020) dizem que algoritmos são, ao mesmo tempo, canal, código, mensagem e visão de mundo (p. 405). Esses objetos midiáticos correspondem, de maneira geral, ao modo através do qual o JD buscará informações, produzirá o seu conteúdo, veiculará as suas narrativas e alcançará o seu público.

Por ser necessária abstração para conceber como as etapas de uma “receita” levam a um resultado tão impressionante quanto cinco mil ações coletadas em 30 minutos, é preciso ter pensamento computacional. Isto é, prever os caminhos pelos quais os algoritmos realizariam a tarefa e os possíveis empecilhos e suas soluções. “Ver um problema específico e reconhecer que isso é a instância de um problema mais geral permite que pessoas com pensamento computacional reconheçam oportunidades para colocar computadores para resolver o problema geral de maior escala” (DIAKOPOULOS, 2019, p. 32, tradução nossa)<sup>25</sup>.

O pensamento computacional será um aspecto importante para entender as técnicas do Volt Data Lab, tendo em consideração que se trata de uma agência focada no uso da tecnologia. Assim, podemos analisar, a partir do nosso objeto, as práticas jornalísticas que, nesse contexto, alternam abordagens conceituais e metodológicas da computação e quantificação, através da idealização e implementação de pensamento e habilidades computacionais e matemáticas (LEWIS, 2014, p. 323).

### **3.4 O JD na prática**

Visualizar os exemplos de matérias que se denominam JD é a melhor forma de compreender o manejo de dados e de algoritmos, assim como a extensão do uso da pirâmide invertida de Bradshaw (2011), e ter uma base comparativa para a análise do nosso objeto. Para isso, abordaremos quatro reportagens vencedoras do *Prêmio Cláudio Weber Abramo de*

---

<sup>25</sup> Seeing a specific problem and recognizing that it is an instance of a more general problem allows computational thinkers to recognize opportunities for applying computers to solve the larger-scale general problem.

*Jornalismo de Dados* de 2022<sup>26</sup>. O concurso é o pioneiro desta categoria no Brasil e, apesar de nenhum critério de júri poder ser tido como o padrão em um assunto, utilizá-lo como recorte estabelece uma base de qualidade mínima e permite o contato com veículos diferentes.

Aliás, as próprias categorias da premiação revelam o que a instituição considera importante em termos de JD. São elas:

- Dados abertos – trabalhos que utilizam bases de dados públicas ou a Lei de Acesso à Informação (LAI);
- Inovação – reportagens com experimentação;
- Investigação – uso de dados para descobrir tópicos de interesse público;
- Visualização – além do viés jornalístico, é considerada a produção estética na forma de organizar as informações coletadas.

Esses quatro termos — dados abertos, inovação, investigação e visualização — coincidem com os comentários apresentados pelos profissionais entrevistados para o *Manual de Jornalismo de Dados*, especialmente no que diz respeito a encontrar mais informações e exercer a função de *watchdog*. Ademais, mostra o peso da visualização, ainda que esse recurso figure no topo e não na base da pirâmide de Bradshaw (2011). Abaixo, estão os principais detalhes sobre as quatro matérias premiadas pelo *Prêmio Cláudio Weber Abramo*, na edição de 2022, que é a mais recente no momento desta pesquisa (quadro 1). Uma delas, inclusive, foi produzida pelo *Núcleo Jornalismo*, veículo do Volt Data Lab.

**Quadro 1:** Lista dos vencedores do Prêmio Cláudio Weber Abramos de Jornalismo de Dados 2022

<b>Título</b>	<b>Categoria</b>	<b>Veículo</b>	<b>Tipo</b>	<b>Resumo</b>
4 em cada 10 abortos legais do Brasil são feitos fora da cidade onde a mulher mora; pacientes percorreram mais de 1 mil	Investigação	<i>GI</i>	Reportagem multimídia	Matéria com texto, vídeos e recursos de infografia sobre as dificuldades enfrentadas por mulheres que buscam o aborto concedido a

<sup>26</sup> PRÊMIO Cláudio Abramo de Jornalismo de Dados. Disponível em: <<https://premio.jornalismodedados.org/edicoes/2022/>>. Acesso em: 17 jun. 2023.

km <sup>27</sup>				termos previsto em lei nos hospitais públicos brasileiros.
Amazônia Minada <sup>28</sup>	Dados Abertos	<i>InfoAmazonia</i>	<i>Dashboard</i>	Mapa com informações sobre regiões da Amazônia em que há um processo de requerimento de mineração aberto.
Aquazônia <sup>29</sup>	Visualização	<i>Ambiental Media</i>	Reportagem multimídia	Reportagem com visualizações criada para apresentar o <i>Índice de Impacto nas Águas da Amazônia</i> (IIAA), que estuda os impactos nas águas dos rios da Amazônia.
BotPonto <sup>30</sup>	Inovação	<i>Núcleo Jornalismo</i>	Robô	Ferramenta de monitoramento de <i>fake news</i> no Youtube.

**Fonte:** Elaboração própria com base em informações presentes na página do Prêmio Cláudio Weber Abramoz e em cada uma das matérias.

A primeira produção desta lista corresponde a uma reportagem multimídia, com base em dados solicitados pelo *GI* ao Sistema Único de Saúde (SUS) através da LAI, sobre

<sup>27</sup> FARIAS, Victor; FIGUEIREDO, Patrícia. 4 em cada 10 abortos legais do Brasil são feito fora da cidade onde a mulher mora; pacientes percorreram mais de 1 mil km. **GI**. 09/06/2022. Disponível em: <<https://g1.globo.com/sp/sao-paulo/noticia/2022/06/09/4-em-cada-10-abortos-legais-no-brasil-sao-feitos-fora-da-cidade-onde-a-mulher-mora-pacientes-percorreram-mais-de-1-mil-km.ghtml>>. Acesso em: 17 jun. 2023.

<sup>28</sup> AMAZÔNIA Minada. **InfoAmazonia**. Disponível em: <<https://minada.infoamazonia.org/>>. Acesso em: 16 jun. 2023.

<sup>29</sup> RIBEIRO, Ronaldo; KLEIN, Leticia; DAMASIO, Kevin. Aquazônia. **Ambiental Media**. Disponível em: <<https://aquazonia.ambiental.media/>>. Acesso em: 17 jun. 2023.

<sup>30</sup> BOTPONTO. Núcleo Jornalismo. Disponível em: <<https://nucleo.jor.br/botponto/>>. Acesso em: 17 jun. 2023.

pacientes que fizeram aborto por se encaixarem nas circunstâncias previstas em lei (estupro, risco à vida da gestante e feto com malformação do cérebro). Como gancho, os repórteres usaram a descoberta de que a maioria das mulheres era moradora de uma cidade diferente àquela onde o procedimento foi realizado. Chama a atenção, portanto, a possibilidade de dar uma finalidade aos conhecimentos subutilizados pelos órgãos públicos e, como citado anteriormente, a importância do JD para que jornalistas não dependam das informações divulgadas de maneira oficial (ainda que partam de fontes oficiais).

Outras fontes de dados utilizadas foram uma pesquisa do Datafolha e o levantamento, cujo nome não foi citado, realizado pela doutoranda em Saúde Coletiva na Universidade Federal de Santa Catarina (UFSC) Marina Jacobs. No geral, não foi feito uso de bases que compilam conhecimento sem o consentimento dos seus correspondentes, sendo todas de interesse público e/ou produzida a partir de métodos científicos.

Partindo dessa visão científica, uma das principais diferenças dessa produção e de outras peças jornalísticas tradicionais é a presença de um intertítulo de metodologia, onde foi explicado como a equipe teve acesso às informações junto ao Ministério da Saúde. “No pedido, foi solicitada a quantidade de mulheres que fizeram aspiração manual intrauterina (AMIU) ou curetagem — os dois procedimentos usados no Brasil para abortar — com o código da paciente O04, que identifica o aborto por razões médicas e legais” (FARIAS e FIGUEREIDO, 2022). O espaço também foi importante para justificar certas decisões: “A reportagem optou por solicitar o CID<sup>31</sup> e os procedimentos, em vez de somente o CID, de modo a garantir que não há pacientes duplicadas no sistema” (FARIAS e FIGUEIREDO, 2022).

Quanto à construção narrativa, a matéria do G1 apresenta com mais destaque um texto redigido de acordo com as características tradicionais de uma reportagem (contextualizando e aprofundando o tema do gancho a partir de documentos ou entrevistas). Ainda no texto, podemos perceber os desdobramentos da pirâmide “comunicar” de Bradshaw (2011), especialmente no que tange aos passos personalizar, humanizar e socializar. Essas etapas são observadas nos esforços dos repórteres para trazer casos pessoais que mostram os dados na prática, apresentar o contexto das personagens que vai além da sua relação com os números e mostrar porque o problema apresentado é, de fato, um problema.

Correspondendo ao aspecto multimidiático do jornalismo digital e do JD, o levantamento rendeu também:

- fotos ilustrativas;

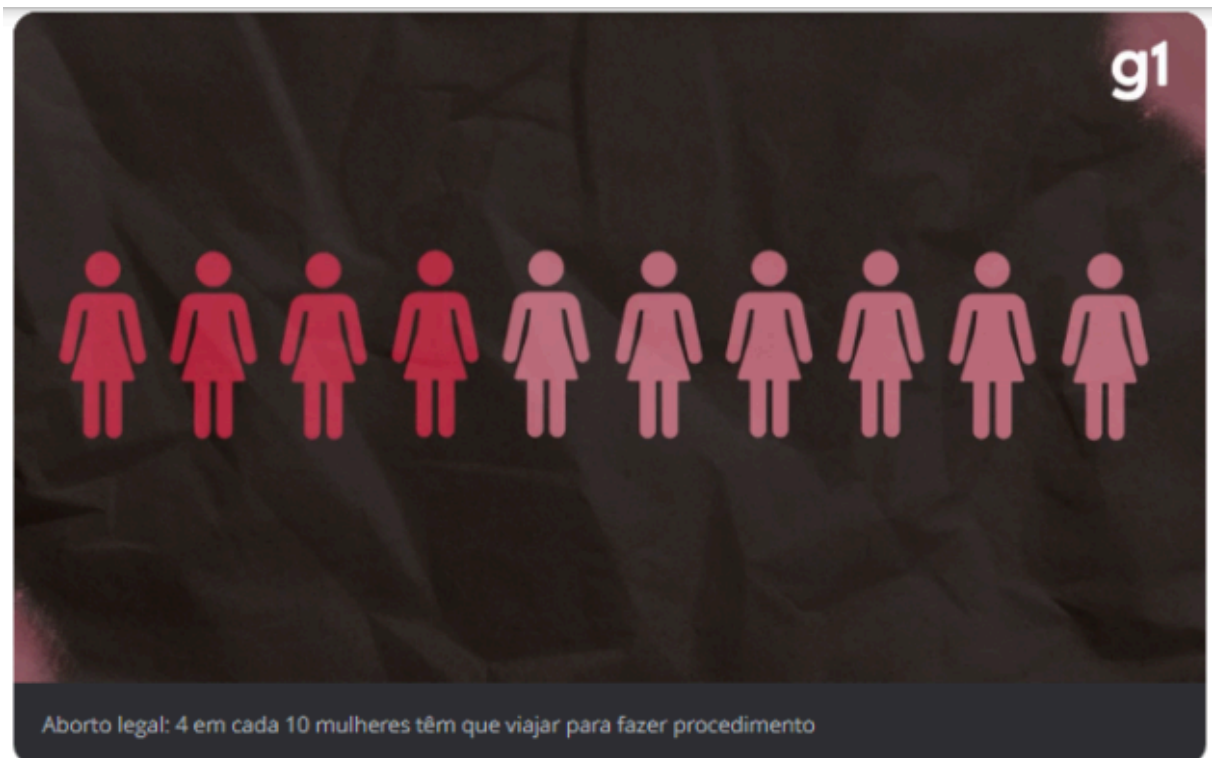
---

<sup>31</sup> Classificação Estatística Internacional de Doenças e Problemas Relacionados com a Saúde.

- um vídeo que apenas apresenta dados sobre aborto legal no Brasil;
- um vídeo que aborda a história das personagens;
- um mapa mostrando os principais deslocamentos das mulheres que precisaram sair de suas cidades para realizar um aborto legal;
- um infográfico com o caminho realizado de ônibus e carro por quem sai de Santa Maria das Barreiras para realizar o procedimento em Belém, além dos custos da viagem;
- infográfico em movimento, mostrando a concentração de abortos nas cidades brasileiras;
- gráfico de barras com o número de abortos nas sete cidades que mais realizam o procedimento.

A imagem a seguir corresponde a um frame do vídeo informativo, em um momento em que se ilustra a informação de que quatro entre dez mulheres que buscam terminar a gestação de forma legal precisam viajar para isso (fig 3).

**Figura 3:** *Frame do vídeo da reportagem ‘4 em cada 10 abortos legais do Brasil são feito fora da cidade onde a mulher mora; pacientes percorreram mais de 1 mil km’*



**Fonte:** FARIAS e FIGUEIREDO, 2022

A partir disso, percebe-se que as escolhas são elementos importantes no JD, pois é preciso decidir o que fazer com os dados coletados, como torná-los mais acessíveis ao entendimento geral dentro das suas especificidades, entre outros. O mapa interativo (fig 4), por exemplo, mostrou-se uma ferramenta adequada para comunicar as centenas de trajetos entre duas cidades que já foram percorridas por mulheres em busca de dar um fim à gestação de modo legítimo. O fato de que ele foi separado por quadros que exibem os detalhes mais importantes, como as cidades que mais receberam as mulheres grávidas, ajuda o leitor a se guiar através de tantas localidades destacadas.

**Figura 4:** Mapa da reportagem ‘4 em cada 10 abortos legais do Brasil são feito fora da cidade onde a mulher mora; pacientes percorreram mais de 1 mil km



Fonte: Ministério da Saúde

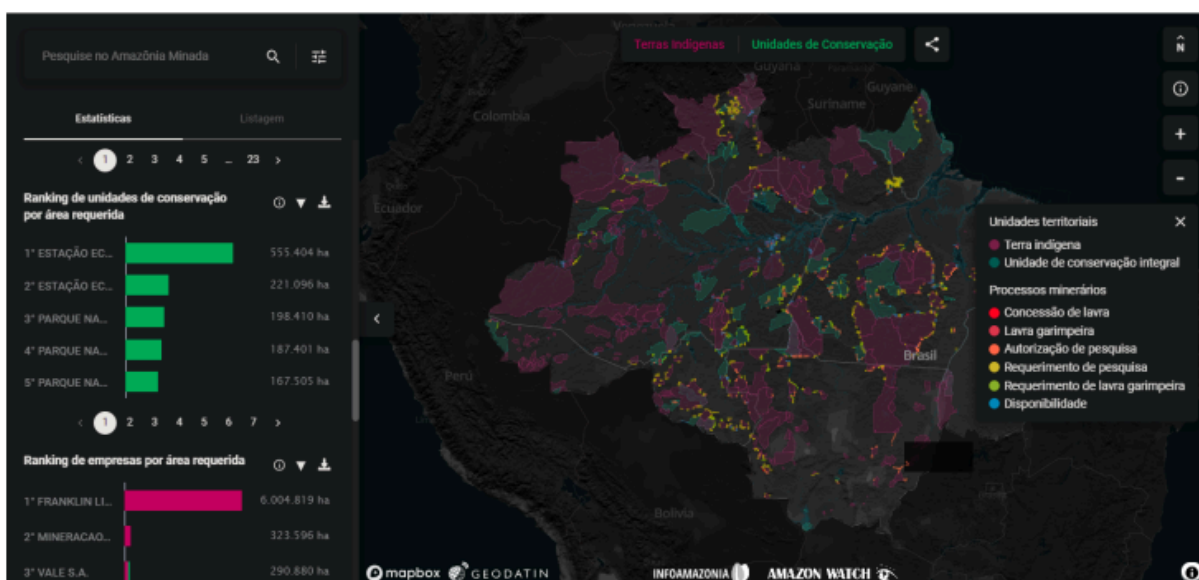
Fonte: FARIAS e FIGUEIREDO, 2022

Nota-se também a abundância de informações que são conquistadas sobre um assunto por meio das bases de dados. Mesmo que o texto estivesse completo, contendo os principais índices e entrevistas com personagens e especialistas, ainda foi possível inserir novos conhecimentos nos vídeos e infográficos.

Não há indicações claras do papel dos algoritmos nesse trabalho, mas podemos inferir que eles foram importantes para a classificação e agrupamento das estatísticas, além de estarem presentes no mapa interativo. Também pode-se notar que, ainda que a matéria possua um viés de conjunto de dados muito específico, em que os jornalistas focaram em apresentar que existe uma grande negação de atendimento para a interrupção da gravidez na maioria das cidades, ele está a favor de uma minoria histórico-social.

O Amazônia Minada, segunda produção indicada na tabela, é um *dashboard* (painel de visualização de métricas) que mostra requerimentos de mineração em terras com demarcação indígena ou em unidades de conservação na Amazônia. A visualização consiste em um mapa marcando a região onde cada um dos processos foi realizado (fig 5). Passando o cursor sobre uma localidade onde há mineração, é aberto um quadro com o nome do solicitante, ano, fase e número do processo, área declarada, minério, entre outros detalhes. No lado esquerdo da página, um menu destaca as principais estatísticas. Além disso, o usuário pode filtrar as informações que deseja que apareçam no mapa ou fazer buscas específicas. São tantas as funções que, assim como a matéria anterior contava com um capítulo de metodologia, essa produção traz um vídeo explicativo de como utilizar os recursos.

Figura 5: Página ‘Amazônia Minada’



Fonte: Amazônia Minada

Os dados, coletados pela Agência Nacional de Mineração (ANM), são atualizados diariamente. E, assim como as fontes da reportagem sobre os abortos legais, a base usada é de interesse público e não prejudica a privacidade de um grupo de pessoas. O fato de que o

*dashboard* é atualizado automaticamente quando há mudança nos processos em andamento ou quando novos requerimentos são feitos permite inferir que os algoritmos que “puxam” e catalogam as informações são essenciais em um cenário de rapidez no meio digital e contribuem para o fator de atualidade, importante para jornalismo.

Esse painel também torna mais claro o conceito de nooscópio aplicado no jornalismo, já que podemos pensar em cada uma das cores no mapa como um conhecimento que antes estava disperso nos catálogos da ANM e que agora podem ser enxergados, ou, usando o termo do JD, visualizados, com forma e cor.

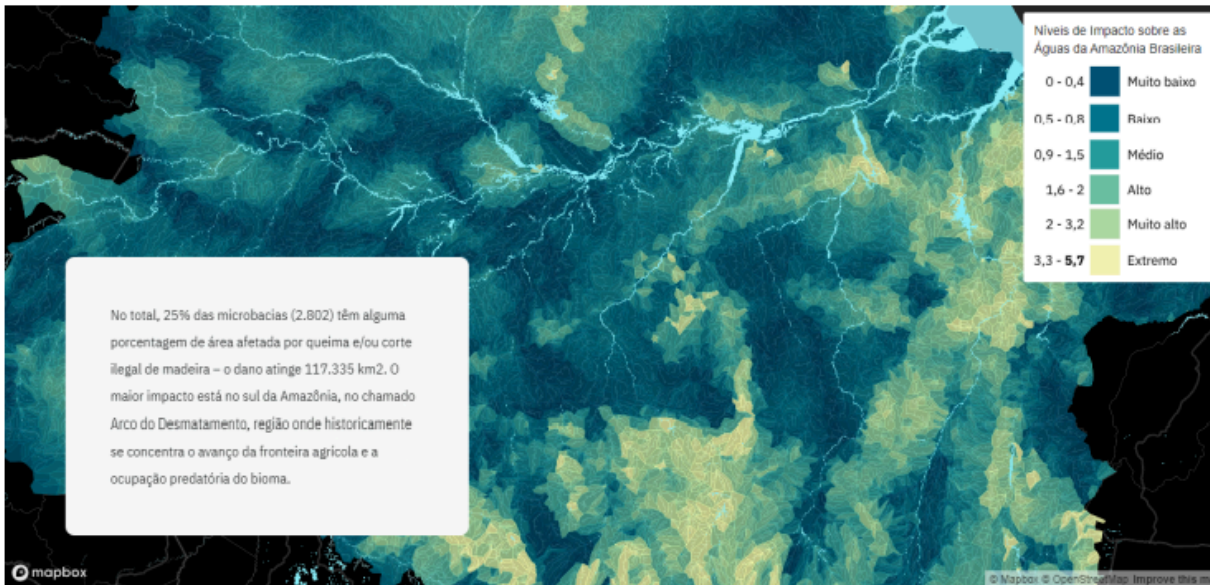
Nessa produção, não é possível identificar alguns dos fatores citados por Bradshaw (2011), principalmente no que diz respeito a comunicar. São tantos detalhes que é difícil entender o que eles significam em sua totalidade. Não há um recurso que personalize ou humanize os dados, eles são apenas apresentados, e a falta de uma estratégia para guiar o leitor leva a uma ausência de narrativa.

Isso parece dificultar o reconhecimento de como os dados foram percebidos durante a construção da peça. Provavelmente, existe algum grau de viés de conjunto de dados, que pode estar presente em escolhas simples dos programadores e jornalistas envolvidos, como a decisão por fazer um mapa e não um outro tipo de visualização ou inserir o tamanho das terras requeridas em hectares e não pela porcentagem do território. Contudo, não é possível analisar apenas com a observação do *dashboard*.

A vencedora da categoria visualização, a reportagem Aquazônia, por outro lado, apresenta uma narrativa contínua que ordena texto, mapas, infografias e fotos. Apesar de estruturada no sentido tradicional das páginas virtuais, de cima para baixo, as visualizações são exibidas uma de cada vez, na ordem desejada pelos produtores, ocupando todo o espaço disponível da tela. A contemplação de cada elemento precisa ser feita individualmente, no ritmo ditado pela matéria. Inclusive, ao lado da parte textual, é apresentado uma guia, mostrando em qual intertítulo o leitor se encontra.



**Figura 6:** Mapa da matéria ‘Aquazônia’



**Fonte:** RIBEIRO, KLEIN e DAMASIO, 2023

Os dados utilizados na produção são de autoria da própria Ambiental Media, que faz conteúdo jornalísticos a partir de pesquisas científicas, com a parceria do Instituto Serrapilheira. O estudo, chamado IIAA, foi a base para a reportagem e as visualizações. O grande foco no levantamento científico pode ser visto, mais uma vez, nas explicações sobre como os dados foram selecionados, o que geralmente não faz parte de uma narrativa jornalística tradicional. O trecho a seguir exemplifica isso:

Durante o processo de desenvolvimento do índice, muitas vezes foi preciso flexibilizar as medidas dos impactos, seja pela ausência de dados seja pelo fato de que cada microbacia é um complexo ecológico individual, pleno de particularidades ambientais ou legais. Por exemplo: o impacto causado por barragens de hidrelétricas ou desmatamento é mais claro. (RIBEIRO, KLEIN e DAMASIO, 2023)

Esse tipo de depoimento sobre como os dados foram coletados, escolhidos ou aplicados é importante para desenvolver no público o sentido de que as informações partem de algum ponto de vista.

Outro ponto relevante dessa reportagem é que, assim como a matéria do *GI*, esta também segue todos os passos contidos na pirâmide comunicar do Bradshaw (2011), preocupando-se inclusive em personalizar, humanizar e socializar os números, índices e estatísticas. A produção apresenta especialistas e tenta mostrar a dimensão humana dos danos do impacto ambiental nos rios da Amazônia, tratando da insegurança alimentar das famílias ribeirinhas que dependem dos peixes. Essa parece ser uma das principais diferenças entre uma organização que produz estudos com intenções jornalísticas ou meramente científicas, já que a humanização e a inserção em um contexto maior não são necessárias para esta última.

Por último, o BotPonto, criação do *Núcleo Jornalismo*, é uma “ferramenta aberta para facilitar a checagem de informação em vídeos no Youtube durante as eleições 2022”<sup>32</sup>. Como a descrição indica, este foi um projeto com um objetivo temporário, que era mostrar quantos e quais vídeos contendo desinformação foram postados no período eleitoral. Mesmo assim, como ele é abastecido por dados do Youtube, cuja organização é feita por algoritmos, o BotPonto continua funcionando passado o momento para o qual foi concebido e ainda que os vídeos sobre o tema tenham diminuído consideravelmente.

Nele, o usuário pode buscar datas, termos, vídeos e canais específicos. Os resultados mostrarão o nome e o link do conteúdo no Youtube, canais com a desinformação, minutagem em que a *fake news* aparece, termos mais mencionados, entre outros. Abaixo, está um exemplo de um trecho dos resultados do robô, com o filtro “último 30 dias” aplicado (fig 7).

**Figura 7:** Visão de uma pesquisa no BotPonto

The screenshot shows the BotPonto search interface. At the top, there are filters for 'Filtro de data' (set to 'Últimos 30 Dias'), 'Título do Vídeo', 'Termos', and 'Canal'. Below the filters, a message reads: 'Veja todos os vídeos em que o BotPonto encontrou algum termo potencialmente desinformativo. Clique no título do vídeo para encontrar o minuto exato em que esses termos são ditos, com link direto para o ponto a ser checado. Use os filtros acima para limitar as datas de postagem dos vídeos, os termos e os canais. Os vídeos estão ranqueados por maior número de visualizações.' Below this is a table titled 'Todos os vídeos que mencionam um termo monitorado' with columns for 'Canal', 'Visualizações', and 'Título (Clique para encontrar os termos na minutagem)'. The table lists four results:

Canal	Visualizações	Título (Clique para encontrar os termos na minutagem)
Folha Política	34.391	CPI do MST faz 'batida' de surpresa em acampamento de invasores de terra e descobr
Jovem Pan News	8.548	Pesquisador analisa qual o real impacto da MP dos carros populares
Jovem Pan News	2.075	Pedido de exploração de petróleo vai ser reanalisado pelo Ibama
Jovem Pan News	1.860	Ibovespa tem queda em dia de poucos negócios; Luís Artur Nogueira analisa

**Fonte:** Núcleo Jornalismo

Mais uma vez, um texto de instruções é apresentado para ajudar o usuário na sua trajetória e as fontes usadas (vídeos abertos ao público) não possuem impedimentos éticos. Neste caso, assim como no *dashboard* Amazônia Minada, não são apresentados contextos, humanização ou personalizações.

<sup>32</sup> BOTPONTO. Núcleo Jornalismo. Disponível em: <<https://nucleo.jor.br/botponto/>>. Acesso em: 17 jun. 2023.

Quanto aos vieses, eles se apresentam sob a forma dos termos que foram considerados desinformativos e aqueles que não foram ensinados para os algoritmos. O perigo dessa catalogação, do que é desinformativo ou não, está no contexto, já que os vídeos no Youtube podem citar expressões como exército, golpista ou ladrão (enquadrados pelo robô para a coleta de dados) sem transmitir uma *fake news*, assim como outros termos podem constar em notícias falsas. De qualquer forma, não há nenhuma explicação sobre como eles lidaram essa questão, se houve uma avaliação dos canais ou se foi usada alguma IA capaz de analisar discursos, mostrando a importância da explicação sobre a metodologia no JD.

Isso nos leva a tentar entender o caráter jornalístico do robô, que apenas compila dados e os exhibe. Se considerarmos a importância de comunicar equívocos e falácias para que as pessoas não acreditem nelas, é possível enxergar a relação dessa produção com o jornalismo. Porém, sem saber qual foi o critério para a seleção dos termos, é difícil entender qual seria o papel de um profissional da comunicação nesse empreendimento, já que poderia ser executado por um programador sozinho. Além disso, o robô parece levar ao extremo a função de automatizar o jornalismo de checagem de fatos, atualizando as listas de fake news diária e rapidamente. Por outro lado, não existe uma apuração para, literalmente, checar os fatos, falhando na qualidade.

Em resumo, considerando os ganhadores do *Prêmio Cláudio Weber*, podemos classificar o JD como as etapas e o resultado da transformação de uma base de dados com relevância pública em um sistema midiático de fácil acesso e entendimento para o leitor. Nem sempre a comunicação das informações precisa estar contextualizada, podendo partir apenas da apresentação de um produto baseado em algoritmos e extraído de bases de dados, que, por si só, apresentam o seu próprio universo de desdobramentos. E, como indicado na literatura sobre uso de algoritmos no jornalismo, esse é um expediente para criar robôs, programas, visualizações e quaisquer outras ferramentas que possam atender ao interesse público. Afinal, os algoritmos são canal, código, mensagem e visão de mundo (LEMOS e DOMINGO, 2020, p. 205)

Ademais, o JD é o exemplo máximo da mecanização do jornalismo, já que, em muitos casos, as produções são atualizadas automaticamente, pois foram programadas para isso. Outras características importantes a serem destacadas é que as matérias possuem uma abordagem típica das ciências, citando a metodologia utilizada. Isso acaba sendo importante para amenizar a transmissão de vieses e até manter o público mais consciente sobre as escolhas que estão envolvidas na articulação dos dados (assim como na mídia como um todo). Sobre as fontes escolhidas, percebe-se que são utilizadas bases de dados de órgãos públicos,

pesquisas conduzidas por conta própria ou informações abertas da internet. No recorte em questão, não houve o cruzamento de bases, sendo escolhida geralmente uma única fonte para nortear a apuração. Já em relação à apresentação do conteúdo, a maioria é multimidiático, configuração comum no meio digital. Apesar de algumas matérias não apresentarem texto verbal, nenhuma era completamente isenta de visualizações de dados, o que parece ser um recurso importante para transmitir a dimensão da pesquisa para os leitores. O JD, além de estar imerso nas mesmas configurações do jornalismo digital, amplia a possibilidade de participação do leitor, que pode selecionar dados para serem visualizados nos mapas e gráficos, e a disponibilidade das notícias, atualizadas de forma mecânica.

## 4 ASPECTOS METODOLÓGICOS

Tendo em vista a proposta de investigação, adotou-se como método o estudo de caso. Essa escolha se justifica porque esse tipo de pesquisa trata-se de “uma investigação empírica que investiga um fenômeno contemporâneo dentro de seu contexto da vida real, especialmente quando os limites entre o fenômeno e o contexto não estão definidos” (YIN, 2001, p. 32). Além disso, essa estratégia é útil para entender novidades conceituais (VENTURA, 2007, p. 386). De certa forma, uma agência de dados jornalísticos pode ser considerada assim.

De acordo com Stake (2003), o estudo de caso deve ser tanto o processo de investigação quanto o seu resultado (p. 136). Ele também explica que ele não se trata exatamente de uma escolha metodológica, mas da decisão sobre o que será estudado, que é um caso (STAKE, 2003, p. 134). Ele explica que para ser considerado dessa forma, um objeto de estudo precisa, inerentemente, possuir finitude ou um limite (p. 135). Ou seja, é preciso saber onde ele começa e onde acaba.

Desta forma, nesta pesquisa são destacadas como um conjunto as matérias criadas para veículos jornalísticos pelo Volt Data Lab, no contexto do recorte estabelecido. Assim, será possível alcançar essa finitude, já que o trabalho da agência será limitado a um número específico de colaborações.

Além disso, Ventura (2007) recomenda buscar um caso que seja típico (representando um tipo ideal), extremo ou atípico (p. 385). A escolha de uma agência que trabalha apenas com dados é um caso atípico dentro do tema jornalismo e, justamente por isso, será uma referência individual para pensar em hipóteses sobre o todo.

Rackin e Becker (1992, apud DUARTE e BARROS, 2011) dizem que, no âmbito das Ciências Sociais,

não é suficiente observar um fenômeno social, um evento histórico ou destacar certos comportamentos com o objetivo de declará-los ‘casos’. Se desejarmos falar sobre um ‘caso’, precisamos dos meios de interpretá-lo ou contextualizá-lo em uma realidade. Um caso compõe sua uniformidade não das ferramentas teóricas usadas para analisá-lo, mas do modo como ele toma forma; nomeado como um fato social ou histórico que combina toda a sorte de elementos dentro de um conjunto de papéis sociais, uma instituição, um movimento social, ou a lógica de ação de uma comunidade (p. 218).

Sobre isso, o contexto de trabalho e os problemas na articulação da lógica algorítmica do JD apresentados nos capítulos anteriores podem configurar o cenário em que o caso se insere.

No âmbito dessa metodologia/escolha do que será estudado, algumas definições podem ser feitas para ajudar a entender o objetivo da pesquisa. Entre elas, a noção de estudo de caso explanatório é a que mais se encaixa nos nossos propósitos. Em primeiro lugar, porque esse é o modelo recomendado para as questões de “como” ou “por que” (YIN, 2001, p. 25; CHOPARD e PRZYBYLSKI, 2021, p. 2). Mas, para além disso, porque “‘explicar’ um fenômeno significa estipular um conjunto de elos causais em relação a ele” (YIN, 2001, p. 140).

Outros pesquisadores podem possuir uma classificação diferente desse estudo, entre eles, podemos ressaltar Stake (2003). Dentro da divisão feita por esse autor, a ideia de estudo de caso instrumental é a que mais se adequa a este trabalho. Ele é caracterizado pelo uso de um caso particular para idealizar uma generalização. O objeto não seria o interesse primário da pesquisa, mais focada no contexto geral que o rodeia, mas uma forma de alcançá-lo (STAKE, 2003, p. 137). O estudo de caso instrumental liga-se a este trabalho pois o nosso objetivo principal está mais ligado a entender possibilidades para lidar com os problemas que a adoção de dados pode gerar do que a traçar um panorama do trabalho do Volt Data Lab.

Tendo estabelecido do que se tratam as nossas estratégias de análise e porque elas serão aplicadas, esmiuçaremos como a investigação ocorreu. Para isso, foram usadas as etapas propostas por Yin (2001), que são: a elaboração de questões; as proposições de estudo; a definição do caso; a lógica que une os dados às proposições e os critérios para interpretar as descobertas. Outros autores também fazem considerações relevantes sobre como deve ser a execução do estudo de caso (todos de forma muito similar, sendo as maiores diferenças em relação ao nome dos processos), porém a nossa escolha baseou-se no fato de que esse pesquisador traz etapas mais detalhadas do que outros.

Na sua perspectiva, o primeiro passo é a elaboração de questões (YIN, 2001, p. 42). Aqui, é considerada também a visão de Stake (2003), que define que as perguntas não devem gerar respostas simples (como “Quais tipos de projetos o Volt Data Lab desenvolve para os veículos que o contratam/fazem parceria com a agência?). A ideia é que sejam problemas temáticos e mais longos (STAKE, 2003, p. 142). Quatro questões guiaram este trabalho:

- De quais maneiras os veículos que utilizam os serviços da agência conseguem produzir produções inovadoras?

- Como as matérias produzidas em conjunto com a agência se caracterizam em relação aos problemas do *Big Data*?
- Como especializar a parte da matéria que corresponde ao trabalho com dados e algoritmos faz com que o resultado final seja diferente?
- Como o domínio na área de dados da Volt Data Lab influencia o resultado de suas participações no âmbito jornalístico (ex.: as matérias usam bancos de dados muito extensos ou recursos tecnológicos complexos)?

Em seguida, após a formulação de questões, vêm as proposições de estudo, que são como hipóteses para guiar o pesquisador sobre a forma como ele deveria começar a pesquisa, afinando ainda mais o recorte (YIN, 2001, p. 42). No nosso caso, não teremos uma proposição fechada para cada questão apresentada. Como ponto de partida para a busca de resposta e um contorno para a pesquisa, foi usado o pressuposto de que a divisão do trabalho jornalístico com outra equipe que possui um grande conhecimento sobre dados é um diferencial para as redações. Assim, toda a investigação parte da observação desses aspectos dentro do escopo das questões propostas.

O terceiro passo trata da definição do caso. Como foi explicado, o foco aqui são os trabalhos com dados realizados pelo Volt Data Lab para outros veículos jornalísticos. Porém, o número de amostras ainda seria muito extenso se todos esses projetos fossem considerados, totalizando 11. Eliminando as páginas fora do ar ou as que tinham acesso restrito para usuários, chegamos a sete produções. Para diminuir ainda mais a quantidade de casos, foi excluída uma matéria da Revista Piauí que continha pouco conteúdo para ser analisado. E visando a prevalência de uma variedade em relação à estrutura das reportagens e aos veículos, eliminamos também uma reportagem da Gênero e Número, pois já havia outra amostra desse mesmo veículo. A opção descartada foi uma produção majoritariamente textual, que se repetia em outros casos, mantendo na análise um infográfico do mesmo veículo. No fim, sobraram os seguintes “sub-casos” para compor o estudo:

- Quase metade das escolas do Rio viveu momentos de medo com tiroteios nas proximidades, produção audiovisual do *Jornal da Globo*;
- Políticos do Brasil, página de aplicação no site do veículo *Poder 360*;
- Gráfico: Gênero e raça na ciência brasileira, da revista *Gênero e Número*;
- O avanço do mercado das submetralhadoras artesanais (e ilegais) no Brasil, da *BBC News Brasil*;

- Tá tudo dominado. Exclusivo: as milícias assumiram o controle do Rio de Janeiro, reportagem do *The Intercept Brasil*.

As duas últimas etapas, que Yin (2001) chama de “lógica que une os dados às proposições” e “critérios para interpretar as descobertas”, não são muito abordadas pela literatura do estudo de caso (p. 48). Entende-se, então, que o modo de realização delas deve estar atrelada aos valores individuais de cada pesquisa.

Para os nossos fins, a lógica que une os dados às proposições foi a revisão bibliográfica (já apresentada no primeiro e no segundo capítulo) sobre três temas considerados pertinentes às proposições: *Big Data* e os seus problemas; JD; e uso de algoritmos no jornalismo. Já o critério para interpretar as descobertas foi o método comparativo, contrastando os dados coletados com o estado da arte sobre o assunto.

É lançando mão de um tipo de raciocínio comparativo que podemos descobrir regularidades, perceber deslocamentos e transformações, construir modelos e tipologias, identificando continuidades e descontinuidades, semelhanças e diferenças, e explicitando as determinações mais gerais que regem os fenômenos sociais (SCHNEIDER e SCHIMITT, 1998, p. 49).

A comparação foi feita tanto entre a parte teórica e as nossas amostras quanto entre os exemplos de JD, apresentados no fim do capítulo anterior, e o nosso caso. Sobre esta última comparação, é pertinente citar que o *Botponto*, iniciativa da própria equipe do Volt Data Lab, será ignorado já que poderia fazer parte das amostras de análise. Quanto ao resto, espera-se que possam fornecer noções de aproximações e distanciamentos em relação às cooperações com a agência.



## 5 ANÁLISE

### 5.1 Descrição dos casos

A primeira produção a ser descrita é uma reportagem televisiva que foi ao ar no *Jornal da Globo* (e depois no *Bom Dia Rio* e no *GI*, do conglomerado de mídia Grupo Globo), chamada “Quase metade das escolas do Rio viveu momentos de medo com tiroteios nas proximidades”, disponível agora no site de *streaming* da emissora. A sua pauta trata da quantidade de tiroteios no entorno das escolas da cidade do Rio de Janeiro e as suas consequências para os alunos, que ficam constantemente sem aulas.

O papel do Volt Data Lab foi realizar o levantamento dos dados, comparando os endereços de escolas e creches da rede pública de ensino com as localizações dos tiroteios. Isto é, a agência foi contratada para executar uma atividade que requer acessar bases de dados diferentes; minerar os dados desejados entre os demais; comparar as informações; e desenvolver maneiras simples de demonstrar a dimensão dos números (por meio de porcentagens, por exemplo). Através de um esforço interpretativo, podemos dizer que o nosso objeto de estudo foi responsável por suprir uma necessidade para a qual o veículo talvez não possuía especialistas.

É importante citar que a tarefa foi praticável por mérito da existência do Fogo Cruzado<sup>33</sup>, uma plataforma colaborativa que reúne ocorrências de tiroteios pela capital fluminense, formando uma base de dados própria. Esse projeto também foi realizado com a assistência do Volt Data Lab, fato que mostra como as empresas de comunicação e outras instituições poderiam se articular para apoiar empreendimentos de organização de dados com o potencial de desencadear inúmeras matérias, páginas e *softwares*, servindo de fontes para o JD.

Além disso, a utilização do Fogo Cruzado para entender o panorama dos tiroteios próximos às escolas cariocas segue o padrão do JD visto no terceiro capítulo desta pesquisa, ao recorrer a bases de dados que não desrespeitam princípios de privacidade, reunindo apenas fatos ocorridos no Rio de Janeiro. Como foi visto, em geral, as informações desse tipo de jornalismo partem da compilação própria ou de algum órgão público sobre o funcionamento das cidades e instituições.

Apesar de ser uma produção televisiva, a característica de visualização de dados também pode ser verificada na matéria do *Jornal da Globo*. Os trechos que apresentam os

---

<sup>33</sup> FOGO Cruzado. Disponível em: <<https://api.fogocruzado.org.br/>>. Acesso em: 18 jul. 2023.

dados levantados pela agência o fazem por meio de representações gráficas para que o público possa visualizar e acompanhar os números citados, reforçando a relevância de recursos didáticos para a elucidação dos dados (fig 8).

**Figura 8:** Recurso visual para a exemplificação de dados na reportagem ‘Quase metade das escolas do Rio viveu momentos de medo com tiroteios nas proximidades’



**Fonte:** Jornal da Globo, 2018

Sobre os impactos da terceirização do levantamento dos números e índices sobre tiroteios próximos a escolas através de uma organização especializada, não é possível dizer que esse recurso gerou mudanças significativas no produto final, pelo menos não do ponto de vista de inovação do gênero de reportagem televisiva, que tem o “propósito de transmitir as informações por meio de entrevistas, imagens que revelem o ocorrido” (XAVIER e RODRIGUES, 2013, p. 2). No geral, a matéria do *Jornal da Globo* segue o padrão já estabelecido para esse tipo de narrativa jornalística, especialmente por ser obra de um programa tradicional. Além dos resultados possibilitados pela pesquisa nas bases de dados, a reportagem apresentou diversos personagens, para contextualizar e humanizar a pauta, e foram ouvidos especialistas que pudessem aprofundar o assunto. Porém, é preciso ressaltar que, neste contexto, o Volt Data Lab permitiu que a pauta se tornasse mais assertiva ao mensurar, com base em uma comparação inédita entre as duas bases de dados, o que antes só era perceptível para quem enfrenta o problema em questão.

Com uma abordagem diferente, a página Políticos do Brasil, do veículo digital *Poder 360*, traz uma aplicação (produzida pelo Volt Data Lab) que mostra o histórico de candidatos

às eleições políticas brasileiras de 1998 a 2022. Tendo como fonte as bases de dados do Tribunal Superior Eleitoral (TSE), é possível verificar informações de candidatos a presidente; vice-presidente; governador; vice-governador; senador; deputado federal, estadual e distrital; 1º e 2º suplente de senador; prefeito; vice-prefeito e governador. Para que o usuário faça um recorte dos elementos oferecidos, esses dados devem ser filtrados por cargo, ano, unidade federativa, município, partido ou nome do candidato. Ao clicar em um dos resultados apresentados, tem-se acesso — além do nome, da data de nascimento, da sigla política, da profissão e de outras informações pessoais — ao plano de governo, certidão criminal, quantidade de votos em cada eleição e ficha completa da declaração de bens do candidato selecionado.

A imagem abaixo mostra uma captura de tela da Políticos do Brasil exibindo uma parte dos dados relacionados a Ciro Gomes, candidato à Presidência da República em 1998, 2002, 2018 e 2022 e ao cargo de deputado federal pelo Ceará em 2006. No momento em que a imagem foi salva, a aplicação mostrava a ficha da sua candidatura de 2022 (fig 9).

**Figura 9:** Informações referentes à candidatura de Ciro Gomes à presidência em 2022, na página ‘Políticos do Brasil’

The screenshot shows the profile of Ciro Gomes for the 2022 presidential election. At the top, there are navigation tabs for the years 2022, 2018, 2006, 2002, and 1998. The profile section includes a photo of Ciro Gomes and the following personal data table:

Nome completo	CIRO FERREIRA GOMES
Data de nascimento	06/11/1957
Município de nascimento	PINDAMONHANGABA / SP
Nacionalidade	Brasileira
Gênero	Masculino
Estado civil	Divorciado
Grau de instrução	Ensino superior completo
Ocupação principal	ADVOGADO

To the right of the profile is a table titled 'Candidatura' with the following details:

Cargo disputado	Presidente
Vice	ANA PAULA MATOS
Cidade/UF da candidatura	Brasil
Situação da candidatura	DEFERIDO
Nome na urna	CIRO GOMES
Nº na urna	12
Nome do partido	PARTIDO DEMOCRATICO TRABALHISTA
Sigla/Nº do partido	PDT/12
Coligação	PARTIDO ISOLADO (PDT)
Plano de Governo	Baixar
Certidão Criminal	Baixar

**Fonte:** Poder 360

Nesta produção, o Volt Data Lab foi responsável por todo o desenvolvimento da aplicação, desde a mineração dos dados à montagem de uma arquitetura para a navegação. Isso indica que, ao fazer divisões de tarefas dentro do JD, assim como no jornalismo digital como um todo, o repórter não precisaria ocupar o seu tempo com o que originalmente não é

de sua alçada e dedicar-se a manter a qualidade de apuração independente da velocidade das publicações nas mídias digitais. Isso é importante pois uma iniciativa que não seja necessariamente baseada na apuração e na elaboração de narrativas é capaz de atender a princípios de divulgação de fatos de interesse público, além de serem úteis para os próprios profissionais de jornalismo. A página Políticos do Brasil, por exemplo, poderia tornar-se também um facilitador do repórter em outras matérias sobre eleições, já que este teria uma ferramenta rápida para apurar o tema.

Neste contexto, é plausível considerar essa produção como uma organização dos dados do TSE de forma que fiquem “legíveis” para o público, cumprindo o objetivo inerente dos jornalistas de divulgar fatos. Também pode-se dizer que essa é uma forma de tornar visíveis os conhecimentos sobre os candidatos políticos e os seus interesses com o passar dos anos e das candidaturas, antes “perdidos” nas bases de dados. Tendo isto em vista, entre as produções analisadas, esta é a que mais representa a ideia do nooscópio ou da possibilidade de ampliar a visão do conhecimento a partir de algoritmos, dados digitais e outros recursos tecnológicos.

A sistematização dos dados com filtragem através de botões e de uma hierarquia de informações dentro de outras também é responsável pela característica de interatividade desse projeto, sendo que o usuário decide o que deve buscar e o que fazer com os resultados. É, portanto, similar ao Bot Ponto e ao Amazônia Minada, constituídos como uma disposição de conhecimentos para que o leitor consulte e explore. Com base nisso, conclui-se que as aplicações, como um produto do JD, caracterizam-se por oferecer a ordenação e a transformação de bases de dados em um formato navegável, sem guiar o caminho de quem as utiliza. Mas, então, de certo modo, o mar de informações continua sendo um mar de informações, sem um laço que as una ou as aparta de seu contexto, tendo o seu objetivo atrelado ao próprio objetivo do usuário.

Seria possível, por exemplo, criar recortes para estabelecer melhor a utilidade e a articulação dos dados, cumprindo não apenas o papel de facilitar o seu acesso e leitura, como também de oferecer ferramentas interpretativas para eles. Os elementos da página Políticos do Brasil possuem potencial para serem divididos, com uma aplicação só para tratar das declarações de bens, outra para os planos de governo e assim por diante. Essa divergência entre as possibilidades e o que de fato foi construído pode se tratar simplesmente de uma questão entre a expectativa condicionada pelas narrativas jornalísticas homogêneas e os novos interesses desencadeados pelo uso das bases de dados. Outra maneira de perceber a situação é questionando a ênfase nos dados em detrimento de uma abordagem em que eles são os

motores para uma história maior. Ademais, no caso descrito, seria necessário conceber também a participação do *Poder 360* na articulação do projeto e qual foi o pedido feito para o Volt Data Lab. Mas, como a estratégia deste estudo de caso parte da comparação com outras iniciativas do JD, não será viável comprovar as suposições, relacionadas ao modo de trabalho e ao entendimento dos profissionais envolvidos.

Por outro lado, esse assunto facilita a compreensão sobre os vieses apresentados na aplicação. Tratando-se somente da reunião de informações sobre cada candidato, os pontos de vista pessoais dos responsáveis pela iniciativa não são observáveis no produto final. De qualquer maneira, é possível verificar reproduções de concepções sócio-históricas brasileiras na própria compilação feita anteriormente pelo TSE. O fato de que foi considerado importante citar o gênero e o estado civil dessas pessoas, por exemplo, diz respeito a valores sociais (além de atender a um propósito de criação de estatísticas para o entendimento da composição política nacional).

Em linhas gerais, essa produção apresenta vieses indissociáveis da sociedade brasileira, pela forma que a própria base de dados foi construída, e mostra um certo esforço para não somar as tendências dos produtores a eles. Entretanto, justamente por este último enfoque, a *Políticos do Brasil* deixa de trazer reflexões sobre os dados, arriscando reforçar a ideia de inexistência de vieses ou a sua ausência nesse conteúdo e, como foi apresentado no segundo capítulo, isso acaba apartando as informações da realidade em que ela está inserida. Transformar as bases de dados em nada mais do que um sumário consultável também leva ao risco de mecanizar demais o trabalho no JD, focando somente em compilar e exibir informações, sendo que, como foi visto anteriormente, os dados são uma ferramenta para encontrar novas histórias e abordar temas já explorados com mais profundidade.

Essa tendência de comunicar os dados se atendo, particularmente, a produção de sentidos explorada pela própria configuração das bases de dados também foi observada no gráfico “Gênero e raça na ciência brasileira”, da revista *Gênero e Número*. A publicação trata de um pequeno texto introdutório (de dois parágrafos), contextualizando o assunto, seguido por um gráfico interativo. A parte verbal limita-se a uma explicação superficial sobre as complicações a respeito da presença de mulheres e pessoas negras trabalhando em pesquisas científicas, sem apresentar as causas ou as consequências da questão. Em síntese, o texto esclarece que “um dos dados que atestam a exclusão das pessoas negras da ciência e da pesquisa brasileiras é o fato de que, nos últimos cinco anos, a proporção de bolsistas do CNPq que se identificam como pretos e pardos não chega a 30% – isso em um país com 54% da população negra” (ASSIS, 2018). A introdução também cita a fonte dos dados, retirados de

currículos registrados na plataforma Lattes, do Conselho Nacional de Desenvolvimento Científico e Tecnológico (CNPq), de 2013 a 2017.

Já o gráfico, que concerne ao trabalho realizado pelo Volt Data Lab, traz o número de bolsas recebidas por homens e mulheres em oito campos: Ciências Agrárias; Ciências Biológicas; Ciências da Saúde; Ciências Exatas e da Terra; Ciências Humanas; Ciências Sociais Aplicadas; Engenharias e “outros” (fig 10). Ao passar o cursor sobre qualquer botão, aparece o número exato de auxílios dados aos pesquisadores daquela categoria. Clicar em uma das áreas científicas apresentadas, por sua vez, exibe a divisão de homens e mulheres por raça. Já ao abrir o índice de uma raça específica (como por exemplo, branca, parda ou preta), a divisão se expande para o número de bolsas por faixa etária para o gênero e a raça escolhidos. Também é possível pegar um atalho e selecionar, na visualização inicial, um gênero em uma das áreas da ciência para acessar todos os dados sobre raça e faixa etária de uma só vez (fig 11). É permitido, dentro da formatação do gráfico, retroceder em todas as compilações feitas para selecionar novos recortes. Além disso, existe um menu para filtrar o tipo de bolsa que será observada, como Doutorado, Mestrado, Iniciação Científica, Iniciação Científica/PIBIC e Produtividade em Pesquisa.

**Figura 10:** Visão geral do gráfico ‘Gênero e raça na ciência’



**Fonte:** ASSIS, 2018

**Figura 11:** Divisão de faixa-etária entre as mulheres brancas que cursam Ciências Agrárias



**Fonte:** ASSIS, 2018

Novamente, não há nessa produção uma contextualização ou humanização dos números apresentados, tratando da compilação de muitas informações. Assim, existe um risco de fazer os dados parecerem soltos, o que pode partir de um cuidado para não reforçar os estereótipos, mas que acaba consolidando a ideia da existência de dados brutos, como se eles não estivessem naturalmente ligados a um contexto que é, necessariamente, humano. Isso mostra como, de certo modo, as entrevistas utilizadas no jornalismo tradicional são capazes de levar o trabalho com dados a uma dimensão mais humanizada, por darem conta desse problema.

Para além, o formato de gráfico com tantas hierarquizações na informação faz com que o leitor precise clicar em diversos botões, indo e voltando, até que seja possível compreender um pouco das questões de gênero e raça e como elas se relacionam com a idade, interpretando-as à sua maneira. Essas questões apontam a importância da cooperação entre o veículo e a agência, ao recorrer a esta última para executar atribuições especializadas. Ou seja, nas produções em que o Volt Data Lab for usado para organizar os dados e fazer uso dos conhecimentos sobre API e algoritmos, a equipe de repórteres, editores e outros profissionais devem propor aprofundamentos para a tecnologia que será desenvolvida e pensar nos seus limites em relação à exposição de fatos.

O gráfico permite comprovar, por meio de dados, que o preconceito de gênero e raça se estende à vida acadêmica e que mesmo a idade interfere nas oportunidades oferecidas para certos grupos sociais, enquanto o tipo de formação não influencia tantas mudanças nesse contexto. No caso do gênero, também é demonstrado que certas áreas da ciência são dominadas por homens e outras por mulheres. Mas isso é tudo o que pode ser extraído da produção. Não é possível descobrir, por exemplo, se o motivo pelo qual a população negra recebe menos bolsas parte do fato de que, na época em que esse gráfico foi desenvolvido, em 2018, ela estava em menor número nas universidades<sup>34</sup>, ou se parte de ações de racismo deliberada de professores. Mesmo que não fosse viável dar uma resposta que compreendesse todos os casos, através de entrevistas, os repórteres teriam a opinião de sociólogos que estudam o tema e conheceriam a experiência de alguns personagens, o que seria mais elucidativo para essa produção.

O mesmo pode ser dito a respeito da divisão por área das ciências, que, sem uma explicação, gera a impressão de que mulheres recebem mais bolsas de um modo geral, quando há um motivo, explicado pelos próprios padrões de gênero, para homens serem a maioria nas Ciências Exatas, mas não em áreas como Ciências da Saúde ou Linguística, Letras e Artes.

Tratar um tema sério desta maneira acaba recaindo no equívoco de conceber os coletivos como padrões, já que leva os leitores a verem grupos de pessoas brancas e não brancas, faixas etárias e cursos universitários sem contemplar o fato de que isso faz parte de um problema social maior e não se trata apenas de uma série de padrões observáveis. Ainda que os vieses de conjuntos de dados dessa produção baseiem-se em mostrar opressões que existem na sociedade, não trazer as causas para isso é ignorar as circunstâncias que formaram essa situação.

Em relação aos vieses históricos (que estão de certa forma ligados aos de conjunto de dados do CNPq), podemos ressaltar que a ideia de dividir a população em dois gêneros, atrelados ao sexo biológico (e excluindo os intersexuais), é uma construção e está longe de ser universalmente aceita. Considerando a pauta e até a linha editorial da revista, essa divisão faz sentido, permitindo a contemplação de como essa mesma bipartição manteve as mulheres em desvantagem. Porém, em outras produções, talvez seja importante pensar em estratégias para diminuir a propagação desse aspecto contido nas bases de dados.

---

<sup>34</sup> MORENO, Ana Carolina. Taxa de jovens negros no ensino superior avança, mas ainda é metade da taxa dos brancos. G1. 06/11/2019. Disponível em: <<https://g1.globo.com/educacao/noticia/2019/11/06/taxa-de-jovens-negros-no-ensino-superior-avanca-mas-ainda-e-metade-da-taxa-dos-brancos.ghtml>>. Acesso em 28 jul. 2023.



Seguindo para a matéria “O avanço do mercado das submetralhadoras artesanais (e ilegais) no Brasil”, da *BBC News Brasil*, antes de realizar esta descrição, foi necessário entender o significado de “análise de dados” dentro do *Big Data*, visto que, segundo o Volt Data Lab, essa foi a sua função na produção. Também chamada de *Analytics*, a análise de dados é feita, inicialmente, pelo tratamento dos dados a partir de técnicas estatísticas que os separa e reúne em conjuntos.

Adicionalmente, para executar a tarefa também pode-se fazer uso de técnicas para categorização, limpeza e transformação dos dados, utilizando, inclusive, a visão da proveniência (fontes de origem) dos dados para auxiliar no processo de categorização. Ao final desta fase é possível chegar à definição e preparação de modelos (fase de *data preparation* e *model planning*) que serão úteis na construção do grande conjunto de dados, chamado de lago de dados (*data lake*). A carga de dados (denominada fase de *ingest*) ocorrerá em seguida e será realizada para povoar o lago de dados. No lago estarão reunidos todos os dados que serão alvo de análise. (RIBEIRO, 2014, p. 100).

Esse caso reforça a suposição de que os veículos recorrem ao Volt Data Lab para suprir uma necessidade para a qual não há especialistas nas redações, principalmente porque a análise de dados é uma função que, segundo Ribeiro (2014), requer conhecimentos matemáticos e estatísticos. Ademais, “a análise de dados, especialmente no ambiente de *Big Data*, conta com vários desafios, como redundância, inconsistência, ruído, heterogeneidade, discretização e categorização de dados, o alto custo com infraestrutura e a necessidade de métodos de análises específicas para trabalhar com grandes volumes de dados” (ROSA, 2018, p. 31). Posto isto, verifica-se que, por mais que os jornalistas de dados tenham acesso às bases, encontrar as informações e torná-las aproveitáveis requer, além de conhecimentos específicos, a autoridade para enfrentar contratempos inerentes ao *Analytics*.

Sobre essa atividade, é essencial considerar que não se trata de um único modo de lidar com informações digitais, mas de um conceito que abrange mineração de dados, análise de redes sociais, *web mining*, *machine learning* e métodos de otimização<sup>35</sup>, cada qual com seus usos, técnicas e ferramentas (ROSA, 2018, p. 46). Na reportagem da *BBC News Brasil* sobre apreensões de metralhadoras produzidas artesanalmente (a confecção de armas de fogo em casa é ilegal no país), os dados foram obtidos por meio da LAI e, mesmo que o órgão consultado não tenha sido divulgado, fica claro que eles fazem parte dos registros da Polícia Federal (PF) — a reportagem usou dados de 2013 a 2018, período em que a instituição cuidava desse assunto. Logo, o trabalho da agência foi fazer a mineração dos dados de

---

<sup>35</sup> Também podem ser chamados de algoritmos de otimização e servem para resolver problemas quantificáveis de alta complexidade. Eles podem atuar, por exemplo, comparando várias opções até achar a mais próxima de estar correta.

ocorrências da PF, separando o grupo das que eram relativas a apreensões de armas e, dentro dessa classificação, quais diziam respeito a metralhadoras, especialmente as artesanais.

Assim, foi possível quantificar os estados que fazem a discriminação do tipo de produção das armas apreendidas, os estados onde mais há apreensão de armas artesanais, a porcentagem de metralhadoras apreendidas que possuem fabricação artesanal e dizer se houve e de quanto foi o aumento desse tipo de apreensão em todo o território nacional. Ou seja, assim como outras produções descritas neste capítulo, o Volt Data Lab foi responsável por oferecer um panorama completo de um assunto a partir de bases de dados. Neste caso, por se tratar de uma reportagem, as informações são distribuídas no decorrer do texto e apenas uma visualização, reproduzida abaixo, foi utilizada.

**Figura 12:** Gráfico da matéria ‘O avanço do mercado das submetralhadoras artesanais (e ilegais) no Brasil



**Fonte:** BODENMÜLLER e SPAGNUOLO, 2019

É importante reforçar o que foi dito no terceiro capítulo, a respeito do texto verbal não ser imprescindível para transmitir as informações do JD e que as visualizações não precisam

ser completadas por esse recurso. Entretanto, nas produções descritas anteriormente, a aplicação e o gráfico deixavam, pela maneira que foram construídos, mais perguntas do que respostas. Neste caso, porém, através do conforto de um modelo narrativo mais tradicional, os números receberam profundidade e nenhuma lacuna foi deixada.

A reportagem explica, por exemplo, o que é uma submetralhadora e porque ela é um problema. “A adaptação ou transformação de armas em submetralhadoras aumenta a capacidade de destruição: com um aperto no gatilho, é possível ter uma rajada com 30 ou mais disparos” (BODENMÜLLER e SPAGNUOLO, 2019). Também é esclarecido o motivo para que esse tipo de arma tenha se tornado popular (o seu preço é inferior), quais as razões para o crescimento do uso desse tipo de arma (vídeos divulgados na internet que ensinam a produzir submetralhadoras e a crise econômica a partir de 2014) e quais as possíveis soluções para o problema, com base em especialistas.

Assim como muitos exemplos observados no referencial teórico, a reportagem esclarece a sua abordagem dos dados ao leitor, como um objetivo similar à explicação metodológica em uma pesquisa científica. Isso pode ser encontrado no seguinte trecho: “Analisar os dados sobre apreensões de armas artesanais é um desafio, uma vez que há um problema crônico de classificação dos órgãos de segurança. Marcas, calibres e tipos de armas são comumente trocados, dificultando a interpretação” (BODENMÜLLER e SPAGNUOLO, 2019).

Por último, tratando de uma das preocupações desta pesquisa, os vieses, não foi possível verificá-los apenas com as informações disponíveis. Saber quais dados conquistados por meio da LAI foram deixados de fora da apuração seria uma forma mais profícua de entender as tendências tanto da PF, ao formar as bases, quanto dos profissionais do Volt Data Lab e da *BBC News Brasil*, ao escolher o que utilizar.

O mesmo não acontece com a reportagem do Intercept Brasil, “Tá tudo dominado. Exclusivo: as milícias assumiram o controle do Rio de Janeiro”. A matéria, que utilizou uma análise de palavras-chave do banco de dados do Disque Denúncia para calcular o aumento da formação de milícias no estado do Rio de Janeiro e a sua proporção em relação às facções, tinha um claro propósito de contar uma história de dominação em que, ao contrário do senso popular, a milícia é tão responsável quanto às facções, senão mais, pelos crimes nas comunidades fluminenses. Com vieses claramente a favor da linha editorial, busca-se, então, trabalhar o imaginário popular a respeito desse tipo de grupo criminoso para construir o cenário que levou ao assassinato da vereadora Marielle Franco, do PSOL (Partido Socialismo

e Liberdade), em 2018. A matéria constitui a parte sete da série “O caso Marielle”, que aborda outras nuances importantes para desvendar o crime.

**Figura 13:** Ilustração principal da matéria ‘Tá tudo dominado. Exclusivo: as milícias assumiram o controle do Rio de Janeiro’



**Fonte:** OLLIVEIRA, *et al.*, 2018

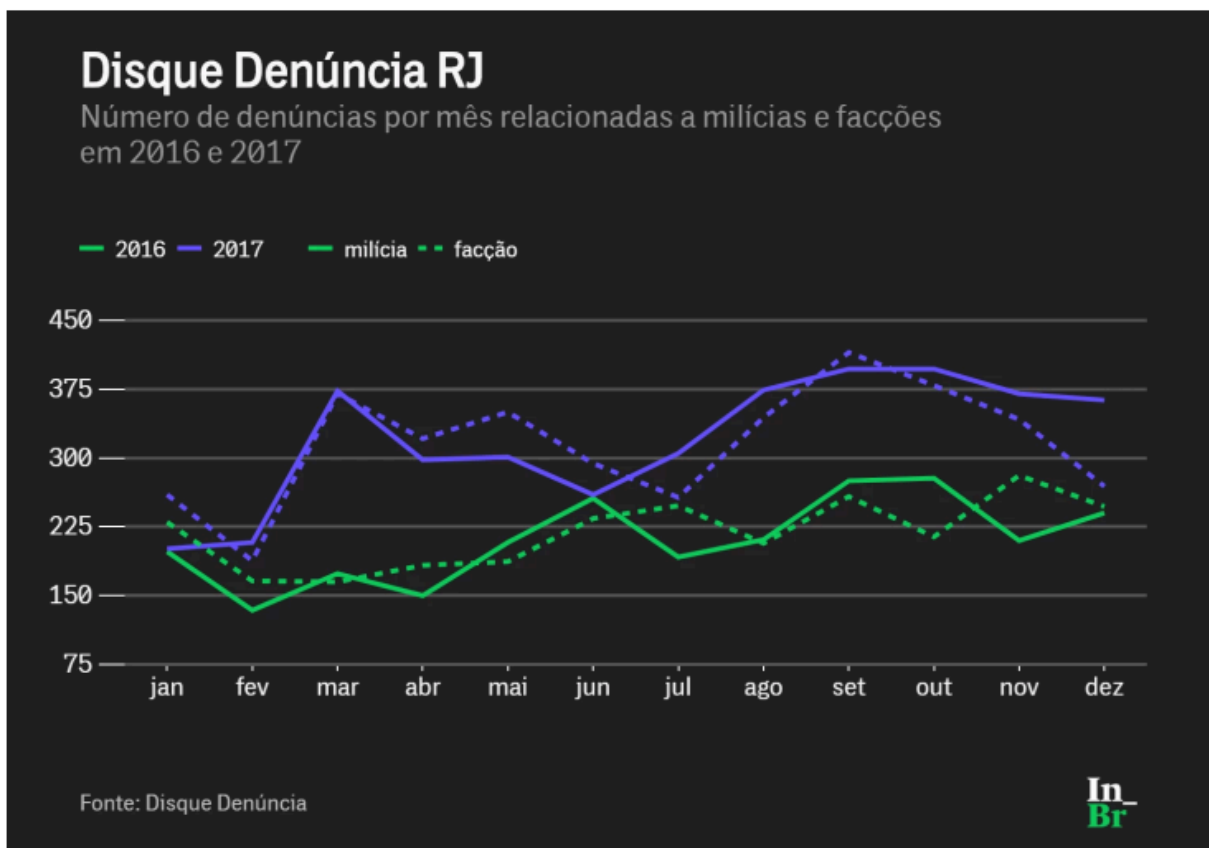
Se a designação da produção como parte da série não fosse suficiente, alguns trechos da reportagem fazem a sua parte em indicar o seu objetivo. “Enquanto milicianos voltaram a frequentar os palácios do poder, Freixo, ameaçado de morte, segue até hoje com escolta policial. E Marielle foi morta, ao que tudo indica, por esse mesmo poder” (OLLIVEIRA, *et al.*, 2018). A cobrança por respostas em relação à morte da vereadora e a denúncia da autoridade da milícia em regiões inteiras, perturbando a ordem local, são orientações legítimas para matérias jornalísticas, cumprindo os seus encargos sociais. Porém, saber que existe sim uma tendência a nortear a análise de dados, o que o veículo deixa claro, sem intenções de esconder, é importante para que o leitor entenda que, possivelmente, existem mais dados além dos que foram apresentados, constituindo um recorte da situação (o que, no fim, todo produto jornalístico é).

Segundo a descrição que a própria agência deu para esse projeto em seu site, “o Volt Data Lab foi responsável por organizar, analisar e visualizar esses dados a partir de interativos e gráficos estáticos, em um esforço coletivo com a equipe do veículo”<sup>36</sup>. A partir dessa explicação é difícil saber onde começa e onde acaba a participação do Volt Data Lab. Mas, em

<sup>36</sup> Disponível em: <<https://voltdata.info/projetos/milicias/>>. Acesso em 31 jul. 2023.

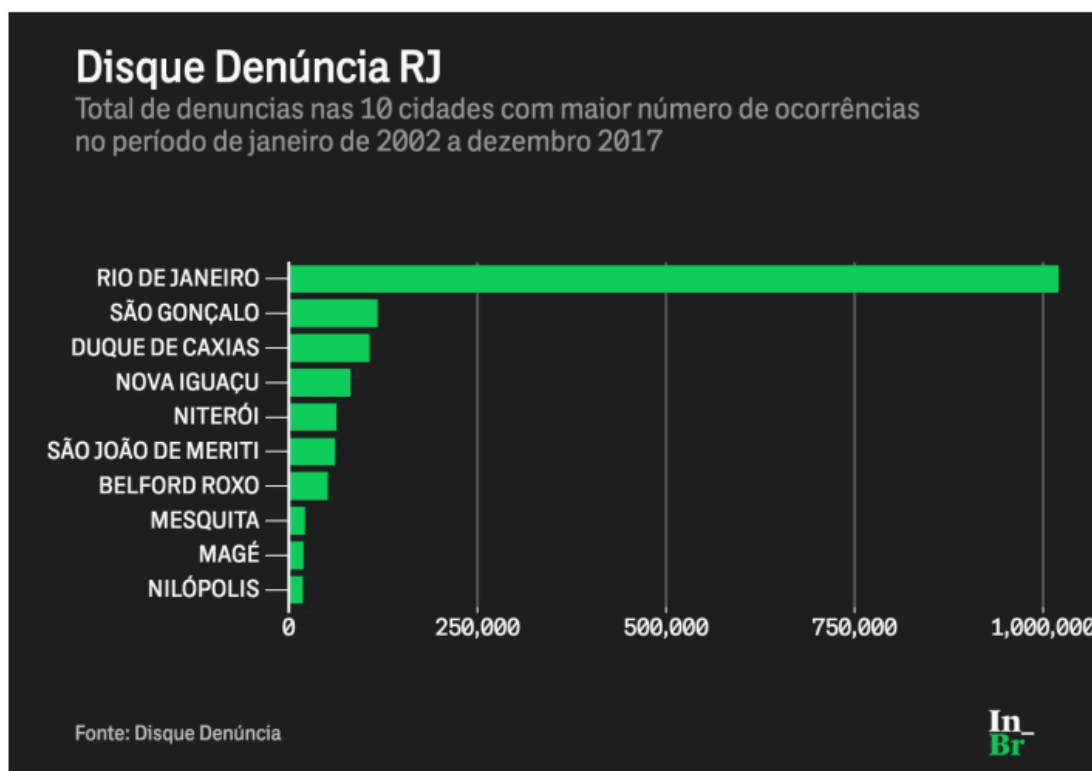
conjunto com os repórteres e editores do Intercept Brasil, os dados foram usados para determinar a porcentagem das ligações de denúncias anônimas que versavam sobre a milícia, de quais cidades partiram a maioria delas, comparar a quantidade de denúncias a respeito das facções e da milícia, os tipos de crimes denunciados e em quais períodos as denúncias, tanto sobre facções quanto sobre a milícia, foram maiores. Ou seja, apenas com dados foi possível contar uma história completa, mostrando o cenário da milícia no Rio de Janeiro. Isso porque diferentes visualizações de dados foram usadas, todas com objetivos específicos a respeito da informação que deveriam transmitir e com uma articulação para completarem a narrativa (fig 14, 15 e 16).

**Figura 14:** Gráfico com número de denúncias por mês relacionadas a milícias e facções



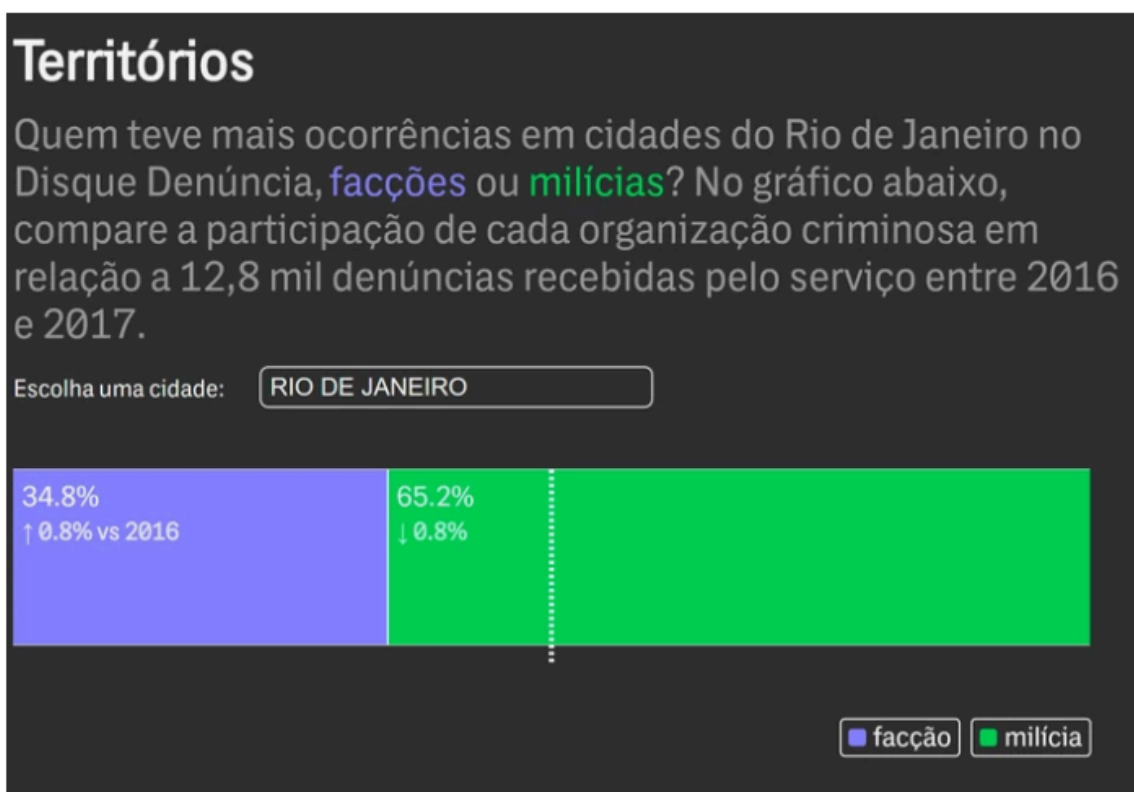
Fonte: OLLIVEIRA, et al., 2018

Figura 15: Gráfico de cidades com maior número de ocorrências na cidade do RJ



Fonte: OLLIVEIRA, et al., 2018

Figura 16: Visualização de dados com a porcentagem de denúncias das facções e das milícias no RJ



Fonte: OLLIVEIRA, et al., 2018

Fazendo uma comparação com a caracterização do JDBD, presente no terceiro capítulo, essa abordagem mostra que é possível entender as bases de dados como um centro cultural, mas que isso funciona melhor quando há um esforço de interpretação e de disposição em uma ordem narrativa, separando as nuances mais importantes, em vez de entendê-las como uma “lista de itens não organizados”.

Indo além de apresentar os dados de forma estruturada e finalizada, a matéria do Intercept Brasil não deixa de humanizar os índices, a partir da apresentação de fontes e histórias das regiões assoladas pelas milícias. Aliás, o próprio propósito de tentar entender a morte de Marielle Franco, um caso com um fim trágico envolvendo a atuação de milicianos, é uma ferramenta para que o público personifique a história que foi mostrada nas visualizações.

## 5.2 Comparação e resultados

Independente da participação do Volt Data Lab, os casos escolhidos para a análise seguem o padrão estabelecido anteriormente como características do JD: transformam uma base de dados de interesse público, cujo acesso geralmente é obtido através da LAI, em um sistema navegável e acessível, geralmente multimidiático e interativo. Os seus formatos variam, podendo se apresentar como uma reportagem textual ou audiovisual ou gráficos e páginas de consulta. Exceto pela reportagem que foi ao ar no Jornal da Globo, os projetos efetuados pelo Volt Data Lab também se concentravam em retirar toda a história de uma única base de dados.

Não é surpreendente, considerando que essas iniciativas continuam sendo parte do modelo do JD, que elas contenham em si os seus fundamentos. Isso não significa que a agência gere um impacto zero nos seus empreendimentos. Analisando os casos, podemos perceber, em primeiro lugar, que essas produções possuem uma tendência maior de mecanizar as informações e apresentar as descobertas em dados de forma desconectada de seus contextos. Isto acontece porque, provavelmente, a especialização das tarefas nem sempre é considerada pelo veículo que a contratou como parte dos processos da redação e não há integração entre as duas equipes (sendo o caso do *Intercept Brasil* uma exceção positiva).

Se por um lado os profissionais de dados diminuem a sobrecarga dos repórteres, por outro, existe um risco de que essa parte do trabalho — a análise, mineração, cruzamento, entre outros — seja vista como completa por si só, quando, aludindo às pirâmides de Bradshaw (2011), deveria ser processos independentes com suas próprias demandas e que se ligam ao ato de comunicar.

Essas produções mais desgarradas das formas narrativas tradicionais, como a página Políticos do Brasil e o gráfico “Gênero e Número na Ciência Brasileira”, poderiam significar, de certa maneira, uma forma tímida de inovação. Contudo, esse é um conceito amplo, que, sem uma definição, pode ser interpretado como qualquer coisa entre o desenvolvimento de um formato ou gênero e a aplicação de técnicas diferentes para resolver problemas cotidianos das redações. Por esta razão, é importante esclarecer que,

enquanto inventar é um termo associado à geração de uma ideia ou produto totalmente novo – incluindo desde o desenvolvimento de pequenos artefatos até invenções mais marcantes, como a pólvora, a escrita, a lâmpada, os meios de transporte e de comunicação, entre outros – podemos entender inovação como a introdução de diferenciais ou melhorias em tais descobertas, reconfigurando-as. [...] A inovação no jornalismo pode ser compreendida sob certos aspectos, seja nas formas narrativas, nos processos produtivos, nos dispositivos tecnológicos e/ou interativos ou mesmo nas etapas de circulação e consumo das informações. (MARTINS, 2018, p. 36 e 37)

Além dessa significação, Martins (2018) sugere, com base em Barbosa (2014) e em Meyer (2010), que, no jornalismo, a inovação está atrelada ao dinamismo, que é a chave dos produtos jornalísticos contemporâneos (MARTINS, 2018, p. 38). Nesse contexto, o jornalismo é inovador quando o seu “conteúdo é continuamente renovado, interrelacionado, contextualizado, trabalhado segundo distintos formatos multimídia” (BARBOSA, 2014, p. 1 apud MARTINS, 2018, p. 38).

Seguindo esta visão, o JD como um todo é um modelo de inovação, que com base nos algoritmos cria produtos que se atualizam sozinhos e sempre utilizam mais de um tipo de mídia. Contudo, devemos pensar no que é inovação dentro do JD, já que os sinais de renovação jornalística são simplesmente as características desse modelo.

Em termos de “introdução de diferenciais ou melhorias”, o Volt Data Lab não deve ser considerado um originador de inovação. As produções analisadas, independente da sua alta qualidade, ou não apresentam diferenciais, seguindo à risca o modelo do JD, ou se distanciam de melhorias, por uma má articulação entre veículos e agências. Mesmo com a aplicação e o gráfico interativo, as produções menos tradicionais entre as analisadas, não podem ser consideradas como inovadoras, já que exemplos como o Amazônia Minada são feitos regularmente no JD.

Ou seja, apesar do domínio do Volt Data Lab no âmbito de dados, a sua influência no produto final é mínima, se considerarmos que outras iniciativas também utilizam bases extensas para alcançar resultados que atendem a diversos tipos de necessidades jornalísticas, como fornecer a pauta, completar a apuração ou estruturar páginas de consulta de informações de interesse público. Isso também se aplica ao uso das tecnologias disponíveis,



que também foi bem diverso e inesperado nas produções que não tinham a participação de uma instituição especializada para repartir o trabalho.

Quanto aos problemas relativos à coleta indevida de dados e à perpetuação de preconceitos, a participação do Volt Data Lab nas matérias examinadas também não traz aprimoramento. Mas isso acontece porque a própria prática do JD no Brasil se apoia na LAI para a confecção das reportagens e, em alguns casos, faz a sua própria compilação de informações. Além disso, os propósitos estão sempre (pelo menos na totalidade dos casos estudados) servindo a elucidação das visões discriminatórias. Por outro lado, a abordagem da página *Políticos do Brasil* e do gráfico “Gênero e Raça na Ciência Brasileira” expõe que recorrer à agência sem uma pauta aprofundada (que não precisa ser necessariamente textual, pode ser um robô ou uma aplicação, portanto que não considere apenas a exibição dos dados) é ignorar a construção social em volta das bases e dos algoritmos entregar uma ilusão de que os dados existem sem um contexto.

É importante salientar que mesmo que os resultados encontrados mostrem pouco impacto do nosso objeto de estudo no JD, isso não é um aspecto negativo. A partir das conclusões alcançadas podemos perceber, primeiramente, que os jornalistas do JD estão bem preparados para as funções que assumem e que a especialização de algumas atividades pode ser importante. Apenas não em seu aspecto técnico, que não teve tanto peso na pesquisa.

Além disso, foi possível observar cenários em que o JD tange para um trabalho mecanizado, o que pode ser explorado em investigações futuras, para que se possa evitar. Por fim, a pesquisa deixou ainda mais claro o peso dos vieses nesse tipo de produção, sendo necessário um balanço entre não divulgar concepções sociais negativas e não levar ao equívoco de que elas não existem ou não estão presentes nos dados. Assim, a despeito dos resultados contrários às hipóteses, este trabalho ainda foi capaz de produzir considerações para o tema trabalhado.

## 6 CONSIDERAÇÕES FINAIS

Na era do *Big Data*, os dados e os algoritmos tornaram-se moedas de alto valor. Eles ampliam o comércio de bens e serviços, por meio das publicidades personalizadas, tem grande influência no cenário político internacional, são ferramentas para as tomadas de decisões em grandes empresas e também na vida de milhões de pessoas que, conectadas constantemente em plataformas, são levadas a consumir, assistir, ler, ouvir, usar, entre outras ações indicadas por algoritmos com base em suas preferências.

Se as empresas e os consumidores foram tomados por esse sistema de geração massiva de conhecimentos com base na relação entre humanos e tecnologias, é impossível que as instituições voltadas para a comunicação, incluindo o jornalismo, não compartilhem dessas práticas. Como o estabelecimento de uma cultura formada por dados digitais não pode acontecer completamente isenta de complicações, nesta pesquisa, focamos no que a visão de um mundo através de padrões, de forma aparentemente matemática e equivocadamente objetiva, pode gerar, especialmente para o modelo jornalístico que se baseia no *Big Data*.

Para dar ênfase ao aspecto do conhecimento técnico necessário para atuar nesse meio, a proposta foi analisar cinco produções do JD co-elaboradas com a agência de dados Volt Data Lab para descobrir maneiras de lidar com o *Big Data* no campo midiático. Já os objetivos secundários eram identificar características do JD no Brasil e investigar se a especialização de tarefas nesse modelo jornalístico levava a melhorias, inovações e a uma diminuição da sobrecarga entre profissionais da área.

Fazendo um levantamento bibliográfico e o estudo dos casos, foi possível alcançar quase todos os objetivos, chegando ao resultado principal de que as práticas do Volt Data Lab não destoam daquelas do JD, sendo focadas na análise de uma base de dados de notório interesse público, de articulação própria ou acessada por meio da LAI, e na organização dos achados em visualizações e *softwares* que podem ou não ser completados por uma reportagem escrita ou audiovisual. Para além, verificou-se que a especialização das tarefas, independentemente do grande conhecimento dos profissionais, não foi suficiente para provocar melhorias. Dessa forma, conclui-se que, para que o jornalista possa se inserir no contexto de datificação sem reproduzir hábitos negativos que estereotipam, padronizam e desumanizam as pessoas, o conhecimento técnico não é o único caminho. Escolher um formato narrativo que una as informações do seu contexto real possui um impacto maior nessa conjuntura. O único objetivo que não foi propriamente atingido foi o que diz respeito à sobrecarga de trabalho, por não ser viável de analisar com a metodologia escolhida, abrindo

margem para estudos baseados em pesquisa de campo e entrevistas. Outro possível desdobramento é investigar quais são os gêneros e formatos jornalísticos desenvolvidos no JD, entendendo o que é esperado deles e quais são os seus processos de produção.

Já as hipóteses, que consideravam que a pesquisa encontraria uma relação positiva entre especialização e resolução dos problemas inerentes ao *Big Data*, não foram comprovadas. Isso mostra que, em futuras pesquisas, a preocupação com a técnica pode dar espaço a outras questões, como a proposta de delineamentos éticos para novos formatos jornalísticos, que é o caso das aplicações, dos robôs e dos *dashboards*.

## REFERÊNCIAS

A CONTA dos Passaralhos. Disponível em: <<https://passaralhos.voltdata.info/>>. Acesso em 14 de maio de 2023.

ALVES, Rosental Calmon. Jornalismo digital: Dez anos de web... e a revolução continua. **Comunicação e Sociedade**, Minho, v. 9-10, p. 93-102, 2006.

AMAZÔNIA Minada. **InfoAmazonia**. Disponível em: <<https://minada.infoamazonia.org/>>. Acesso em: 16 jun. 2023.

ASSIS, Carolina. Gráfico: Gênero e raça na ciência brasileira. **Gênero e Número**. Disponível em: <<https://www.generonumero.media/reportagens/grafico-genero-e-raca-na-ciencia-brasileira/>>. Acesso em 16 jun. 2023.

ATLAS da Notícia. Disponível em: <<https://www.atlas.jor.br>>. Acesso em 14 de maio de 2023.

BARBOSA, Suzana. O que é Jornalismo Digital em Bases de Dados? *In*: Encontro Anual da COMPOS, 15, 2006, Bauru.

BASTOS, Helder. Ciberjornalismo e narrativa hipermídia. **Prisma.com**, Porto, n. 1, p. 3-15, 2005.

BASTOS, Helder. A diluição do jornalismo no ciberjornalismo. **Estudos em jornalismo e mídia**, Porto, v. 9, n. 2, p. 284-298, ago. 2012.

BASTOS, Helder. Ciberjornalismo, jornalismo e democracia. **Media & Jornalismo**, Lisboa, n. especial, p. 93-105, 2015.

BODENMÜLLER, Luiza; SPAGNUOLO, Sérgio. O avanço do mercado das submetralhadoras artesanais (e ilegais) no Brasil. **BBC News Brasil**. Disponível em: <<https://www.bbc.com/portuguese/brasil-47858443>>. Acesso em: 11 jun. 2023.

BOTPONTO. **Núcleo Jornalismo**. Disponível em: <<https://nucleo.jor.br/botponto/>>. Acesso em: 17 jun. 2023.

BRADSHAW, Paul. The Inverted Pyramid of Data Journalism. Online Journalism Blog, 2011. Disponível em: <<https://onlinejournalismblog.com/2011/07/07/the-inverted-pyramid-of-data-journalism/>>. Acesso em 29 maio 2022.

CANAVILHAS, João Manuel Messias. Webjornalismo: considerações gerais sobre jornalismo na Web. **Informação e Comunicação Online - Projecto Akademia**, Covilhã, v. 1, p. 63-73, 2003.

CANAVILHAS, João Manuel Messias. Do jornalismo online ao webjornalismo: formação para a mudança. **Comunicação e Sociedade**, Minho, v. 9-10, p. 113-119, 2006.

CHOPARD, Kay; PRZYBYLSKI, Roger. **Methods brief**: Case studies. Washington, Justice Research and Statistics Association, nov. 2021.

DALBEN, Silvia. Robôs no jornalismo brasileiro: Três estudos de caso. **Brazilian Journalism Research**, Brasília, v. 16, n. 3, p. 476-501, dez. 2020.

DEUZE, Mark. Journalism and the Web: An analysis of skills and standards in an online environment. **International Communication Gazette**, v. 61, n. 5, p. 373-390, 2009.

DIAKOPOULOS, Nicholas. **Automating the news**: how algorithms are rewriting the media. Massachusetts: Harvard University Press, 2019.

DIERICKX, Laurence. The social construction of news automation and the user experience. **Brazilian Journalism Research**, Brasília, v. 16, n. 3, p. 432-457, dez. 2020.

DUARTE, Jorge; BARROS, Antonio. **Métodos e técnicas de pesquisa em comunicação**. São Paulo: Atlas, 2011.

FARIAS, Victor; FIGUEIREDO, Patrícia. 4 em cada 10 abortos legais do Brasil são feito fora da cidade onde a mulher mora; pacientes percorreram mais de 1 mil km. **G1**. 09/06/2022. Disponível em: <https://g1.globo.com/sp/sao-paulo/noticia/2022/06/09/4-em-cada-10-abortos-legais-no-brasil-sao-feitos-fora-da-cidade-onde-a-mulher-mora-pacientes-percorreram-mais-de-1-mil-km.ghtml>. Acesso em: 17 jun. 2023.

FEDERAÇÃO NACIONAL DOS JORNALISTAS. Código de Ética dos Jornalistas Brasileiros, 2007.

FIALHO, Ana Carolina Mendes. Reconhecimento Facial: Entenda o impacto da Inteligência Artificial na vida da população negra. **Exame**, 4 out. 2022. Disponível em: <https://exame.com/esg/reconhecimento-facial-entenda-o-impacto-da-inteligencia-artificialna-vida-da-populacao-negra/>. Acesso em: 20 jan. 2023.

FOGO Cruzado. Disponível em: <https://api.fogocruzado.org.br/>. Acesso em: 18 jul. 2023.

GEHRKE, Marília; MIELNICZUK, Luciana. Philip Meyer, the outsider who created Precision Journalism. **Intexto**, Porto Alegre, n. 39, p. 4-13, maio/ago. 2017.

GRAY, Jonathan; BOUNEGRU, Liliana; CHAMBERS, Lucy. **Manual de jornalismo de dados**. Open Knowledge Foundation, 2012.

GRAY, Jonathan; BOUNEGRU, Liliana; CHAMBERS, Lucy. **Manual de jornalismo de dados 2**. Open Knowledge Foundation, 2021.

HUAYLUPO, Juan. La relatividad y significación de los datos. **Cinta moebio**, Santiago, v. 32, p. 127-152, 2008.

JURNO, Amanda Chevtchouk; D'ANDRÉA, Carlos. Entre parcerias, infraestruturas e produtos: Facebook Journalism Project e a plataforma do jornalismo. **Brazilian Journalism Research**, Brasília, v. 16, n. 3, p. 502-525, dez. 2020.

LEMOS, André. DOMINGO, David. Journalism and Algorithms. **Brazilian Journalism Research**, Brasília, v. 16, n. 3, p. 404-409, dez. 2020.

LEWIS, Seth C. Journalism in an era of big data. **Digital Journalism**, v. 3, n. 3, p. 321-330, 2014.

LIMA, Ricardo Oliveira da Cunha. O que é infografia jornalística? **InfoDesign - Revista Brasileira de Design da Informação**, São Paulo, v. 12, n. 1, p. 111-127, 2015.

LUPTON, Deborah. How does data come to matter? Living and becoming with personal data. **Big Data and Society**, v. 5, n. 2, p. 1-11, 2018.

MANOVICH, Lev. **The language of the new media**. Massachusetts: The MIT Press, 2001.

MARCONDES FILHO, Ciro. **Comunicação e jornalismo: A saga dos cães perdidos**. São Paulo: Hacker, 2000.

MARTINHO, Ana Pinto. Jornalismo de dados: caracterização e fluxos de trabalho. **Exedra**, Coimbra, n. 9, p. 65-73, dez. 2014.

MARTINS, Elaide. Modos e sentidos da inovação no jornalismo. **Comunicação & Inovação**, v. 19, n. 39, p. 35-49, jan-abril 2018.

MEYER, Philip. **Precision journalism: A reporter's introduction to social science methods**. Lanham: Rowman & Littlefield Publishers, 2002. 4ª ed.

MICHAEL, Mike; LUPTON, Deborah. Toward a manifesto for the public understanding of big data. **Public Understanding of Science**, v. 25, n. 1, p. 104-116, 2016.

MIELNICZUK, Luciana. Webjornalismo de Terceira Geração: continuidades e rupturas no jornalismo desenvolvido para a Web. *In: Encontro dos Núcleos de Pesquisa da Intercom*, 23, 2004, Porto Alegre.

MORENO, Ana Carolina. Taxa de jovens negros no ensino superior avança, mas ainda é metade da taxa dos brancos. G1. 06/11/2019. Disponível em: <<https://g1.globo.com/educacao/noticia/2019/11/06/taxa-de-jovens-negros-no-ensino-superior-avanca-mas-ainda-e-metade-da-taxa-dos-brancos.ghtml>>. Acesso em 28 jul. 2023.

NOBLE, Safiya Umoji. **Algorithms of oppression: How search engines reinforce racism**. New York: New York University Press, 2018.

NÚCLEO Jornalismo. Disponível em: <<https://nucleo.jor.br/>>. Acesso em 14 de maio de 2023.

OLIVEIRA, Daniela Borges de; COSTA, Belarmino Cesar Guimarães da. Pauta guiada por algoritmo: conteúdo e forma no Estadão Infográficos. **Brazilian Journalism Research**, Brasília, v. 16, n. 3, p. 550-575, dez. 2020.

OLLIVEIRA et al. Tá tudo dominado. Exclusivo: as milícias assumiram o controle do Rio de Janeiro. *The Intercept Brasil*. 05/04/2018. Disponível em:

<<https://www.intercept.com.br/2018/04/05/milicia-controle-rio-de-janeiro/>>. Acesso em: 11 jun. 2023.

PANGRAZIO, Luci; SEFTON-GREEN, Julian. The social utility of ‘data literacy’. **Learning, Media and Technology**, v. 45, n. 2, p. 208-220, 2020.

PASE, André Fagundes, et al. Quem controla a voz? O uso jornalístico e o domínio informacional em transatores vocais. **Brazilian Journalism Research**, Brasília, v. 16, n. 3, p. 576-603, dez. 2020.

PASQUINELLI, Matteo; JOLER, Vladan. The nooscope manifested. **AI & Society**, Brighton, n. 36, p. 1263–1280, dez. 2021.

POELL, Thomas; NIEBORG, David; VAN DIJCK, José. Plataformização. **Fronteiras – estudos midiáticos**, São Paulo, v. 22, n. 1, p. 2-10, jan./abr. 2020.

POLÍTICOS do Brasil. Poder 360. Disponível em: <<https://eleicoes.poder360.com.br/>>. Acesso em: 11 jun. 2023.

PRÊMIO Cláudio Abramo de Jornalismo de Dados. Disponível em: <<https://premio.jornalismodedados.org/edicoes/2022/>>. Acesso em: 17 jun. 2023.

QUADROS, Cláudia Irene de. A participação do público no webjornalismo. *In: Encontro de Pesquisadores em Jornalismo*, 3, 2005, Florianópolis.

QUASE metade das escolas do Rio viveu momentos de medo com tiroteios nas proximidades. **Jornal da Globo**. 17/04/2018. Disponível em: <<https://globoplay.globo.com/v/6670902/>>. Acesso em: 11 jun. 2023.

RIBEIRO, Cláudio José Silva. Big Data: os novos desafios para o profissional da informação. **Informação & Tecnologia (ITEC)**, João Pessoa/Marília, v. 1, n. 1, p. 96-105, jan./jun. 2014.

RIBEIRO, Ronaldo; KLEIN, Letícia; DAMASIO, Kevin. Aquazônia. **Ambiental Media**. Disponível em: <<https://aquazonia.ambiental.media/>>. Acesso em: 17 jun. 2023.

ROSA, Caroline Silvério. **Estudo sobre as técnicas e métodos de análise de dados no contexto de Big Data**. 2018. 60 p. Engenharia Eletrônica e de Telecomunicações — Universidade Federal de Uberlândia (UFU), Patos de Minas, 2018.

SALAS, Javier. Google conserta seu algoritmo “racista” apagando os gorilas. **El País**, 16 jan. 2018. Disponível em: <[https://brasil.elpais.com/brasil/2018/01/14/tecnologia/1515955554\\_803955.html](https://brasil.elpais.com/brasil/2018/01/14/tecnologia/1515955554_803955.html)>. Acesso em 20 jan. 2023.

SALAVERRÍA, Ramón. Periodismo Digital: 25 años de investigación. **El profesional de la información**, Barcelona, v. 28, n. 1, 2019.

SANTAELLA, Lucia, et al. Desvelando a internet das coisas. **Revista GEMInIS**, São Carlos, v. 1, n. 2, p. 19-32, 2013.

SANTAELLA, Lucia; KAUFMAN, Dora. Os dados estão nos engolindo? **Civitas: Revista de Ciências Sociais**, São Paulo, v. 21, n. 2, p. 214-223, ago. 2021.

SANTOS, Márcio Carneiro dos. A operação jornalística guiada por dados: reconfigurando a percepção de valor da notícia com algoritmos, inteligência artificial e personalização estendida. **Brazilian Journalism Research**, Brasília, v. 16, n. 3, p. 458-475, dez. 2020.

SARWEIN, Florian; JUST, Natascha; LATZER, Michael. Governance of algorithms: options and limitations. **Info**, v. 17, p. 35-49, 2015.

SCIENCE Pulse. Disponível em: <<https://sciencepulse.org/>>. Acesso em 14 de maio de 2023.

SCOTT, Ben. A contemporary history of digital journalism. **Television & New Media**, v. 6, n. 1, p. 89-126, feb. 2005.

SCHNEIDER, Sergio; SCHIMITT, Cláudia Job. O uso do método comparativo nas Ciências Sociais. **Cadernos de Sociologia**, Porto Alegre, v. 9, p. 49-87, 1998.

SILVA, Sivaldo Pereira. Algoritmos, comunicação digital e democracia: Dimensões culturais e implicações políticas nos processos de Big Data. *In*: MEHL, João Paulo; SILVA, Sivaldo Pereira. **Cultura digital, internet e apropriações políticas**: experiências, desafios e horizontes. Rio de Janeiro: Folio Digital: Letra e Imagem, 2017.

STAKE, Robert E. Case Studies. *In*: DENZIN, Norman D.; LINCOLN, Yvonna S. **Strategies of Qualitative Inquiry**. Thousand Oaks: Sage, 2003, p. 134-164.

STRASSER, Bruno J.; EDWARDS, Paul N. Big data is the answer... But what's the question? **Osiris**, Chicago, v. 32, nº 1, p. 328-345, 2017.

TEMIN, Roberto; ROMANI, Elizabeth; GIL FILHO, Vicente. Os primeiros caracteres digitais: Um estudo a partir dos computadores pessoais. p. 841-850. *In*: Congresso Brasileiro de Pesquisa e Desenvolvimento em Design, 11, **Anais**, São Paulo: Blucher, 2014, p. 841-850.

VASCONCELLOS, Fábio; MANCINI, Leonardo; BITTENCOURT, Carolina. Cinco categorias de Jornalismo de Dados ou uma proposta para problematizar o Jornalismo a partir de dados no Brasil. *In*: Seminário de Pesquisa em Jornalismo Investigativo, 2, 2015, São Paulo.

VÉLIZ, Carissa. **Privacidade é poder**. São Paulo: Contra Corrente, 2021.

VENTURA, Magda Maria. O estudo de caso como modalidade de pesquisa. **Rev SOCERJ.**, v. 20, n. 5, p. 383-386, 2007.

VOLT Data Lab. Disponível em: <<https://voldata.info/>>. Acesso em 14 de maio de 2023.

XAVIER, Aline; RODRIGUES, Liliana. Técnicas e práticas para elaborar reportagens telejornalísticas. *In*: Congresso de Ciências da Comunicação na Região Norte, 12, 2013, Manaus.

YIN, Robert K. **Estudo de caso**: planejamento e métodos. Porto Alegre: Bookman, 2001.



ZUBOFF, Shoshana. **A era do capitalismo de vigilância**: a luta por um futuro humano na nova fronteira do poder. Rio de Janeiro: Intrínseca, 2021.