



UNIVERSIDADE FEDERAL DO RIO DE JANEIRO
ESCOLA DE COMUNICAÇÃO
CENTRO DE FILOSOFIA E CIÊNCIAS HUMANAS
JORNALISMO

**TEM #DDJBR AQUI? Mapeando a presença do
jornalismo de dados no Brasil**

MATHIAS FELIPE DE LIMA SANTOS

RIO DE JANEIRO

2019

UNIVERSIDADE FEDERAL DO RIO DE JANEIRO
ESCOLA DE COMUNICAÇÃO
CENTRO DE FILOSOFIA E CIÊNCIAS HUMANAS
JORNALISMO

TEM #DDJBR AQUI? Mapeando a presença do jornalismo de dados no Brasil

Projeto prático submetido à Banca de Graduação
como requisito para obtenção do diploma de
Comunicação Social / Jornalismo.

MATHIAS FELIPE DE LIMA SANTOS

Orientadora: Profa. Dra. Marialva Barbosa

RIO DE JANEIRO
2019

UNIVERSIDADE FEDERAL DO RIO DE JANEIRO
ESCOLA DE COMUNICAÇÃO

TERMO DE APROVAÇÃO

A Comissão Examinadora, abaixo assinada, avalia o Projeto Prático **TEM #DDJBR AQUI? Mapeando a presença do jornalismo de dados no Brasil**, elaborado por Mathias Felipe de Lima Santos.

Monografia examinada:

Rio de Janeiro, no dia/...../.....

Comissão Examinadora:

Orientadora: Profa. Dra. Marialva Barbosa
Pós-Doutora pelo Centre National des Recherches Scientifiques
Departamento de Expressão e Linguagens – UFRJ

Coorientadora: Profa. Alice Melo
Mestre em Comunicação pelo programa de pós-graduação em Comunicação e Cultura da
Universidade Federal do Rio de Janeiro (UFRJ)

Prof. Dr. Fábio Souza Vasconcellos
Doutor pelo Instituto de Estudos Sociais e Políticos (IESP/UERJ)
Departamento de Jornalismo - UERJ

Prof. Dra. Beatriz Becker
Doutora em Comunicação pela Universidade Federal do Rio de Janeiro (UFRJ)

RIO DE JANEIRO

2019

FICHA CATALOGRÁFICA

DE LIMA SANTOS, Mathias Felipe.

TEM #DDJBR AQUI? Mapeando a presença do jornalismo de dados no Brasil. Rio de Janeiro, 2019.

Projeto prático (Graduação em Comunicação Social/ Jornalismo)
– Universidade Federal do Rio de Janeiro – UFRJ, Escola de
Comunicação – ECO.

Orientadora: Marialva Barbosa; Coorientadora: Alice Melo

DE LIMA SANTOS, Mathias Felipe. **TEM #DDJBR AQUI? Mapeando a presença do jornalismo de dados no Brasil.** Orientadora: Marialva Barbosa. Coorientadora: Alice Melo. Rio de Janeiro: UFRJ/ECO. Projeto prático em Jornalismo.

RESUMO

Este trabalho de conclusão é um produto jornalístico sobre a presença de jornalismo de dados no Brasil. O projeto ‘Tem #DDJBR aqui?’ aborda o tema a partir das organizações que produzem conteúdo baseado em dados. Foi desenvolvido com base em pesquisas e estudos acadêmicos sobre o tema proposto para entender o que engloba a produção jornalística através de dados. Esse trabalho apresenta todo o processo de realização do projeto, desde a coleta, combinação, organização, exploração e visualização de dados para se obter relações que permitam o mapeamento da presença da prática no país, até as dificuldades e questionamentos durante a produção desse projeto. Pretende-se analisar as questões pertinentes ao tema a partir de revisão bibliográfica e ao estudo do processo histórico do jornalismo baseado em dados no país, para entender o quem são as principais organizações produzindo narrativas baseadas em dados no Brasil. O produto resultado desse trabalho é um mapa interativo com as localizações das principais organizações, assim como informações relativas as mesmas.

Dedicatória

Dedico este trabalho a minha irmã,
Maiara Stéfani e aos meus pais,
Jacqueline e Laurilo, pois, sem as
lições e o amor incondicional
deles, nada disso seria possível.

Agradecimentos

A toda a minha família e a Deus, que me acompanham, me apoiam e me incentivam desde sempre, e me apoiaram durante todos os meus estudos. Principalmente a minha irmã, Maiara Stéfani, que sem o apoio e esforço dela não estaria na UFRJ. Aos meus pais, Jacqueline e Laurilo, tenho também que agradecer por me mostrarem quão importante é a educação na vida das pessoas.

Aos amigos que ganhei na faculdade: Pedro Henrique Carvalho, Priscila Firmino, Ana Leticia, Thaís Pedroso, Isabela Izidro, Ix Chel, Mariana Fontes, Fernanda Bastos, Marcella Falcão, Kariny Leal, Mariana Martins, Débora Nascimento, Lavynia Andrade, Kathlen Barbosa, entre outros que provavelmente esqueci de mencionar.

Aos amigos das empresas que passei durante esses anos. Em especial à Alda no Tijuca Tênis Clube, que foi uma professora para meus primeiros passos no jornalismo. Aos parceiros da Olympic Broadcasting Services (OBS), em um evento que nunca poderei esquecer. Aos amigos da Avaaz, em uma missão quase impossível de combater as fake news nas eleições presidenciais de 2018.

Não posso deixar de mencionar dos amigos que chegaram antes e que sempre estiveram presentes ao longo desses anos: Vivian, Leonardo, Shiguelo, Márcio e Wilson. Aos que sugeriram e formaram uma família durante ao período que precisei trancar o curso: Joyce, Michelle, Amalia, Ishaqur, Ana, Ricky, Shahnoza, Maya, Durai Markus, Alena, Alessandro.

A todos os meus colegas, professores e funcionários da Escola de Comunicação da UFRJ. Alguns que marcaram e não poderei deixar de mencionar aqui, como a querida e generosa Ilana Strozenberg, a excelente Andréia Resende, a dedicada Filomena Varejão, ao sábio Leonardo De Marchi, ao inovador Diego Paleontólogo, aos que me ajudaram muito no mundo da fotografia Victa de Carvalho e Dante Gastaldoni e ao profissional mais competente que uma secretaria poderia ter, Henrique Hoehn.

Por fim, e não menos importante, às minhas orientadoras, a excelente, carinhosa e animada Marialva Barbosa, que também foi responsável por melhorar minha redação jornalística, junto com a Alice Melo, uma jornalista perspicaz e tech-savvy que a ECO me deu a honra e o prazer de conhecer.

SUMÁRIO

1. Introdução	1
2. Metodologia: desvendando os caminhos da pesquisa	5
3. Jornalismo de Dados: em torno dos conceitos.....	9
3.1 Jornalismo de dados: um pouco de história.....	9
3.2 RAC, Jornalismo de Dados e Computacional	12
3.3 O desenvolvimento do jornalismo de dados no mundo	13
3.4 Entendendo o jornalismo de dados no Brasil	17
4. Dados sobre dados: Quem trabalha com jornalismo de dados no Brasil?.....	20
4.1 Coleta e limpeza	20
4.2 Organização e exploração de dados	22
4.3 Visualização	23
5. Conclusão	29
6. Referências Bibliográficas	31

1. Introdução

Os dados estão no centro de como nos comunicamos, interagimos e prosperamos em um mundo conectado. Hoje, o mundo é feito de dados e esse é o recurso mais valioso que temos para entender a sociedade. Para saber para onde estamos indo, primeiro precisamos entender onde estamos. As decisões do dia-a-dia precisam de dados e não é fácil entendê-los, por isso, é necessário torná-los visuais e legíveis (GRAY et al., 2012).

Com o surgimento da Internet, as pessoas têm acesso à informação de forma mais fácil, por isso, é necessário entender o que é verdade para que possamos disseminar e criar uma visão crítica do que está acontecendo no mundo. Durante a última eleição presidencial nos EUA, assim como a do Brasil, as redes sociais foram usadas para divulgar informações que poderiam ser consideradas essenciais para decidir a direção tomada pela população. A maior parte do conteúdo compartilhado tinha forma de opinião, mas existiam notícias sobre candidatos que não eram possíveis de saber se eram verdadeiras ou falsas. No entanto, foram disseminadas como verdadeiras. Esse é um problema conhecido como *water cooler effect*, quando as pessoas começam a falar sobre algo sem certeza, são “rumores”, como afirma DIFONZO (2009). “Os rumores proliferam onde quer que as pessoas interajam”¹ (DIFONZO, 2009, tradução nossa).

Porém, os rumores não são algo novo, eles sempre existiram em qualquer sociedade. O que mudou hoje é que as pessoas estão pouco interessadas em saber se algo é de verdade um fato ou um rumor, o que importa para elas é a conveniência em acreditar nele. Dessa forma, “a verossimilhança substituiu a verdade como medida, ‘a arte socialmente recompensada’ passou a ser ‘aquela que faz as coisas parecerem verdadeiras’”, criando uma realidade própria que se torna cada vez mais real para determinados grupos (KAKUTANI, 2018, p. 75). Isso mostra quão relevante é o uso de dados no jornalismo hoje.

O uso de dados nas redações não é algo novo. Já faz quase uma década desde que o jornal britânico *The Guardian* apresentou uma das investigações mais reveladoras no Reino Unido. O jornal publicou 460.000 páginas de relatórios de despesas por membros do Parlamento, pedindo a seus leitores que denunciasses afirmações questionáveis (DE MAEYER et al., 2015; FLEW et al., 2012) e, posteriormente, levou a uma série de demissões do alto escalão da política do país. O jornal britânico disponibilizou

¹ No original: “*Rumors proliferate wherever people interact*”

publicamente dados detalhados que os leitores poderiam pesquisar por itens ou lista de distritos eleitorais do Reino Unido, e enviar comentários ou perguntas para a equipe. Dessa forma, jornalistas do periódico britânico investigaram alegações questionáveis, construindo histórias em torno desses comentários. O fluxo de notícias levou a uma investigação do governo que constatou que muitos membros haviam feito declarações incorretas ou confusas. Um dos mais emblemáticos foi o caso do presidente da Câmara dos Comuns do Parlamento britânico na época, Michael Martin, que teve de responder diversas alegações desde despesas com taxis até subsídio para moradia (SPARROW, 2008). Um benefício chave para *The Guardian* foi que nenhuma outra organização de mídia foi capaz ou estava preparada para replicar o estudo, aumentando a sua reputação para o jornalismo de dados e investigativo (FLEW et al., 2012).

O termo jornalismo de dados se tornou popular desde a última década com a chegada do *Datablog* do jornal britânico *The Guardian*. A prática dados é vista como um novo método de investigar e contar histórias de maneira a informar e engajar o público em maior escala. Além disso, os jornalistas de dados têm papel importante para ajudar a diminuir as barreiras à compreensão e à imersão em dados e aumentar a alfabetização em dados de seu público em larga escala.

Apesar de ser um projeto de nicho em 2009, neste estágio de sua evolução, o jornalismo de dados, tanto como profissão quanto como disciplina acadêmica, ganhou força significativa para crescer e se tornar uma disciplina em forte amadurecimento. O jornalismo de dados se consolidou como a arte de coletar dados de várias fontes, verificá-los e processá-los para produzir histórias com base no que eles dizem (DE MAEYER et al., 2015). Um exemplo evidente é o lançamento do *Panama Papers* há dois anos, que reuniu nada menos que 370 jornalistas em 100 organizações de mídia de 80 países diferentes, que revelaram os negócios secretos da economia *offshore* (BURGESS, 2016; CARSON & FARHALL, 2018). Esse trabalho contou com a colaboração da audiência e atores do jornalismo de dados, como jornalistas, designers e profissionais de tecnologia da informação (TI) que desempenham papel crucial ao divulgar essas histórias de maneira mais envolvente.

A história do uso de dados no jornalismo começou no século XIX quando a enfermeira Florence Nightingale usou um gráfico de pizza para representar as mortes da guerra da Criméia. A difusão de tecnologias digitais nos meios de comunicação ao longo do século XX alterou a prática e incorporou novos aspectos ao trabalho de jornalismo. Dessa forma, surgiram novas formas de informação, como resultado do aumento da digitalização

(BRADSHAW, 2014). O jornalista Philip Meyer estabeleceria um marco no campo de RAC – Reportagem com Auxílio de Computador (do inglês, *CAR – Computer-Assisted Reporting*). Foi em 1967, o ano dos protestos dos direitos civis em Detroit, umas das manifestações mais violentas da história americana, com 43 mortos e mais de 7.000 presos. Meyer com a ajuda de um *mainframe* fez levantamentos estatísticos para mostrar que as pessoas que frequentavam a faculdade eram igualmente suscetíveis a promoverem revoltas, como as que haviam chegado ao ensino médio (GRAY et al., 2012).

A prática de jornalismo de dados utiliza da digitalização para lidar com grande quantidade de dados, que jamais poderiam ter sido imaginados, dando nova ênfase ao uso de ferramentas tecnológicas para o processo de produção de notícias. Jornalismo de dados serve como um lembrete de quão importante a análise de *big data* (grandes massas de dados) pode ajudar no processo de trabalho de notícias (HAMMOND, 2017; LEWIS & WESTLUND, 2015). De tão relevante que essa categoria se tornou relevante para o jornalismo, passou a ser estudada no ambiente acadêmico. No entanto, as pesquisas acadêmicas focadas em estudar a prática são frequentemente centradas em casos de países que adotaram precocemente, também conhecido como *early-adopter*, como os Estados Unidos e o Reino Unido. Os debates acadêmicos sobre jornalismo de dados têm sido estimulados por vários estudiosos nos últimos anos, particularmente com relação às normas, práticas e adoção do jornalismo de dados em nível nacional. Os estudos têm como foco os Estados Unidos (PARASIE & DAGIRAL, 2013; YOUNG & HERMIDA, 2015), Reino Unido (BORGES-REY, 2016; HANNAFORD, 2015), Canadá (HERMIDA & YOUNG, 2017), Bélgica (DE MAEYER et al., 2015), Noruega (KARLSEN & STAVELIN, 2014), Suécia (APPELGREN & NYGREN, 2014) e Espanha (TERUEL RODRÍGUEZ & BLANCO CASTILLA, 2016).

Entender os fenômenos do jornalismo de dados e o surgimento de novas mídias² é um dos maiores desafios nos estudos contemporâneos do jornalismo. Como a prática amadureceu desde então, os pesquisadores usaram de diferentes métodos para alcançar seus objetivos de pesquisa, como entrevistas qualitativas ou questionários para identificar e compreender quem são os principais atores envolvidos nas principais redações e como a

² O surgimento de estudos sobre a nova mídia evidenciam o reconhecimento do papel cultural chave desempenhado pelos computadores e das redes na sociedade global e na mídia em geral. Autores consideram nova mídia os métodos e as práticas sociais de comunicação, representação e expressão que se desenvolveram usando o computador digital, multimídia, a rede e os modos que essas máquinas utilizam para transformar o trabalho em outras mídias: de livros a filmes, de telefones para a televisão. Como referência, as redes sociais são consideradas nova mídia (LISTER et al., 2009).

prática está sendo implementada (YOUNG et al., 2018). Porém, os métodos tradicionais não são mais a única forma de compreender quem são os principais envolvidos na atividade do jornalismo de dados do Brasil. Eles não são capazes de dar a dimensão desse campo em expansão, hoje é necessário que ele seja estudado e representado em outros formatos e formas.

Portanto, este trabalho deliberadamente pretende criar um produto jornalístico que mapeia e dá visibilidade ao jornalismo de dados no Brasil. Dessa forma, esse projeto sai da área geográfica mencionada anteriormente de grande estudo acadêmico, em vez disso, foca neste caso ainda pouco pesquisado do Brasil (RODRIGUES, 2014; TRÄSEL, 2013, 2018; VASCONCELLOS et al., 2015; VENTURA, 2018). Na indústria de notícias brasileiras menos inovadora e com menos recursos (VENTURA, 2018), até que ponto se pode afirmar que o jornalismo de dados realmente chegou nas organizações de mídia? Quão concentrado é o jornalismo de dados nas grandes capitais brasileiras? Essas são duas questões chaves que nortearão o trabalho. O objetivo é fazer um cadastro nacional de jornalistas de dados e divulgar as instituições que aplicam a prática no país através de um mapa interativo que permite ao leitor buscar por organizações que produzam conteúdos jornalísticos baseado em dados.

Através de uma metodologia que combinará busca por força bruta, questionários e uso de busca em redes sociais (usando hashtags populares na área, como #jornalismodedados e a que nomeia a esse trabalho, #ddjbr, referente a *data-driven journalism Brazil*), os dados serão coletados para construir o mapa do jornalismo de dados no Brasil. Esse trabalho apresenta todo o processo de realização do projeto, desde a coleta, organização e exploração de dados para se obter relações que permitam o mapeamento da presença da prática no país, até as dificuldades e questionamentos durante a produção desse projeto. Conclui-se com uma agenda para possíveis desdobramentos do projeto.

Dessa forma, a reportagem baseada em dados digitais é relevante devido ao impacto que está tendo sobre o papel tradicional e o poder do jornalismo. Investigar como a prática vem sendo aplicada e onde vem sendo no Brasil, nos ajuda a entender o nível de maturidade e quem são os principais atores trabalhando com a prática no país. Outro fator relevante desse trabalho, é a falta de pesquisas acadêmicas focadas no país, com dimensões continentais e líder na América Latina, o Brasil ainda não chamou a atenção de pesquisadores para entender como a produção de reportagem baseada em dados vem sendo aplicada no país.

2. Metodologia: desvendando os caminhos da pesquisa

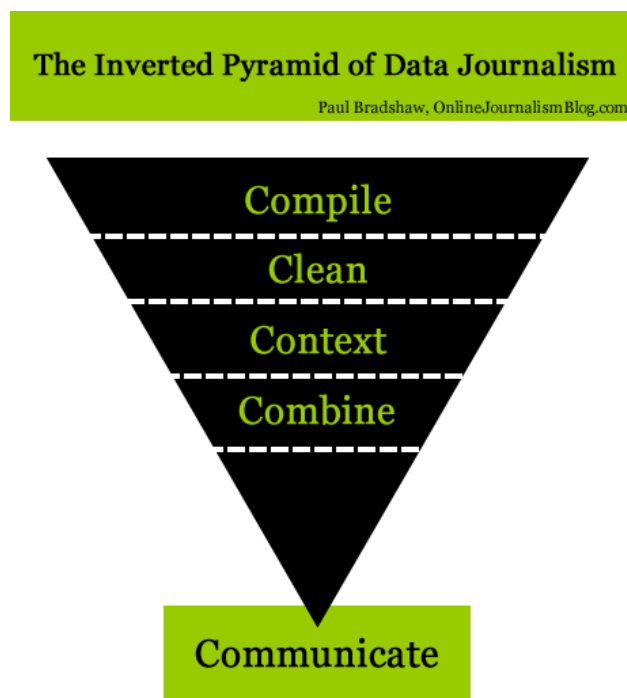
Optamos por esmiuçar os caminhos da pesquisa a partir da descrição da metodologia adotada neste capítulo do trabalho, por entender que na própria escolha metodológica estão contidos os pressupostos teóricos que adotamos, bem como os percursos e as ferramentas utilizadas para a construção da proposta.

Assim, esse capítulo tem como objetivo central apresentar as etapas de construção do trabalho, ao mesmo tempo em que descreve o caminho de produção de uma história baseada em dados.

O trabalho utilizou, portanto, de diferentes práticas para a produção do produto jornalístico resultante, baseando na pirâmide invertida desenvolvida por Bradshaw (2007), conforme mostra figura 1, para descrever o pipeline de produção de uma história baseado em dados. Segundo o autor existem duas formas de como começar uma reportagem de dados: ou você tem uma pergunta que precisa de dados para ser respondida ou possui um conjunto de dados que precisa ser questionado. Seja qual for, a primeira fase é a compilação de dados. O mesmo defende que esse é o estágio mais importante, não apenas porque todo o resto depende dessa parte, mas porque é provavelmente o estágio que será retornado mais vezes em cada um dos estágios subsequentes.

Para criar uma base de dados com os meios que produzem histórias baseadas em dados foi utilizado um questionário disponibilizado em grupos de WhatsApp de dados abertos, de Facebook de jornalismo de dados e dados abertos, além de coletar dados do Twitter com as hashtags #ddjbr (referente a *data-driven journalism Brazil*) e #jornalismodedados. Agregando os dados de todas essas fontes foi permitido gerar um banco de dados com cinquenta e dois ($n=52$) organizações que produzem histórias baseadas em dados, descrito com mais detalhes no capítulo 4. Manualmente foram adicionados os endereços corretos e as informações de descrição de cada meio.

Figura 1 – Pirâmide Invertida



Fonte: (BRADSHAW, 2007).

Com os dados obtidos, a próxima fase é a limpeza. Ter confiança nas histórias escondidas atrás dos dados, significa poder confiar na qualidade dos mesmos, e isso é possível através de um processo de limpeza cuidadoso. Para Bradshaw (2007), essa etapa tem dois objetivos básicos: remover o erro humano e converter os dados em um formato consistente com outros dados que estão sendo utilizados. Nesse produto jornalístico, essa etapa foi fundamental para padronização de dados e limpar erros humanos que foram adicionados através do formulário. A ferramenta *open source* OpenRefine foi usada para auxiliar no processo de limpeza. Essa plataforma é utilizada para ajudar na limpeza através da execução de quatro tarefas essenciais: remoção de registros duplicados, separação de valores contidos no mesmo campo, análise da distribuição de valores ao longo de um conjunto de dados e agrupamento de diferentes representações do mesmo tipo.

A próxima etapa na pirâmide é a contextualização. Como qualquer em qualquer fonte jornalística, os dados nem sempre podem ser confiáveis. Eles podem vir com suas próprias histórias, preconceitos e objetivos. Então Bradshaw (2007) pede para questionar seus dados: quem coletou, quando e com que propósito? Como foi reunido? O que exatamente eles querem dizer? Para o autor também é preciso entender jargões, códigos,

categorias ou terminologias especializadas. Nesse caso, como os dados foram coletados com a ajuda de terceiros, foi preciso fazer a checagem de cada link para ver a conformidade com o que foi definido como jornalismo de dados no questionário. Seguindo a definição de Vasconcellos et al. (2015), o jornalismo de dados é a prática jornalística que exige a coleta, organização e exploração de dados para se obter relações que podem ser significativas entre os dados. Com base nisso, foram checadas todas as histórias e portais das organizações para verificar o alinhamento com o jornalismo de dados. Após cuidadosa análise, foram considerados para o mapeamento organizações que não são apenas de mídia, mas que produzem conteúdo jornalístico feito por atores com formação na área, como é o caso da Sociedade Brasileira de Pediatria e do Conselho Federal de Medicina.

A penúltima fase dessa pirâmide invertida é a combinação. O autor, Bradshaw (2007), acredita que boas reportagens podem ser encontradas em um único conjunto de dados, mas geralmente é necessário combinar duas ou mais bases de dados. Assim como no jornalismo, é melhor uma história de múltiplas fontes do que uma única fonte. A combinação também pode vir através de outras formas, com o uso de dados e visualizações nas histórias. Conforme mencionado antes, esse projeto seguiu os preceitos de Bradshaw (2017) e utilizou de dois canais diferentes para mapeamento de dados para a geração de uma base de dados. Os dados foram coletados através de um questionário e através do Twitter. Essas duas fontes permitiram gerar uma tabela com as organizações que produzem jornalismo de dados no Brasil.

A última fase do jornalismo de dados é a ‘comunicação’. Considera a mais relevante, pois visualizar os resultados – em geral, através de um mapa, gráfico ou infográfico, ou ainda em uma animação ou produto interativo – é o que diferencia a prática do uso de dados pelo jornalismo (BRADSHAW, 2007). Como a monografia, através do produto jornalístico produzido (utilizando como ferramental os mecanismos de produção do jornalismo de dados) tem o objetivo de responder onde tem jornalismo de dados no Brasil, foi utilizado um mapa interativo que permite ao usuário selecionar cada ponto no mapa e obter mais informação sobre cada organização, assim como acessar um *link* para uma história baseada em dados. Para ajudar o trabalho de jornalistas, *startups* estão desenvolvendo soluções prontas que eliminam a necessidade do jornalista de programar e oferecem formas de visualização dos dados. Essas ferramentas são conhecidas como *out-of-the-box* e ajudam na produção de reportagens principalmente para pequenas e médias organizações de mídia que

não tem equipes suficientes. Nesse caso, a ferramenta *out-of-the-box* Carto³ foi escolhida para apresentação dos dados através de um mapa utilizando em associação com programação para web.

³ <https://carto.com>

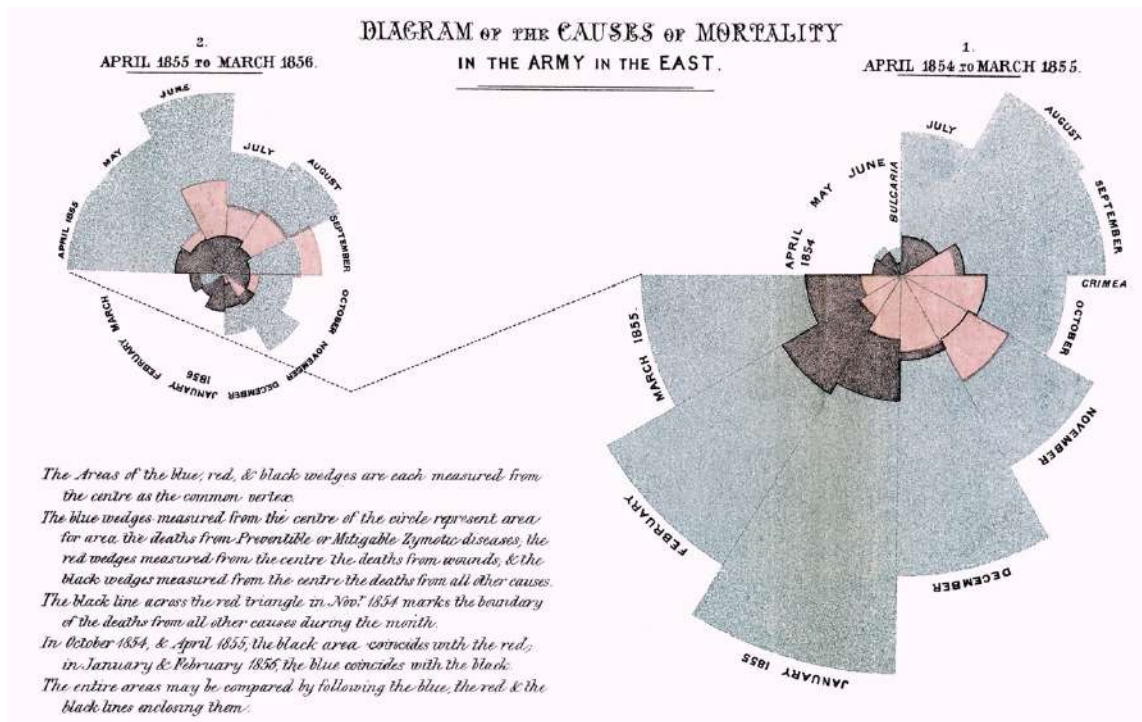
3. Jornalismo de Dados: em torno dos conceitos

Esse capítulo conceituará o jornalismo de dados, explicando seu processo histórico de utilização de conhecimentos estatísticos para a produção de notícias até a nomenclatura conhecida hoje como jornalismo de dados. Também apresentará um panorama do que vem sendo estudado sobre jornalismo de dados e como a prática vem moldando redações e organizações em todo o mundo. Por fim, serão abordados aspectos de como a prática vem sendo utilizada como recurso para atrair audiência e caminho para enfrentar a crise no modelo de negócio do jornalismo atual.

3.1 Jornalismo de dados: um pouco de história

Apesar da terminologia ‘jornalismo de dados’ ter se popularizado nos últimos anos com o *Datablog* do jornal britânico *The Guardian*, o uso de dados e conhecimentos estatísticos para a produção de notícias é bem anterior. A história do jornalismo de dados começa para os estudiosos em 1850 com Florence Nightingale, uma enfermeira e a primeira mulher a fazer parte da *Royal Statistical Society*. Ela teve um grande impacto, pois foi pioneira na utilização de métodos de representação visual de informações estatísticas para representar as mortes dos soldados britânicos na Guerra da Criméia, conforme mostra a figura 2 (ROGERS, 2010).

Figura 2 – Diagrama para mostrar as causas da morte na guerra por Florence Nightingale



Fonte: (ROGERS, 2010).

Seu trabalho mostrou claramente que as mortes foram ocasionadas por infecção e que as doenças contraídas pelos soldados ocorreram em hospitais, algo que rotineiramente era subestimado na época. Com esses dados, Nightingale convenceu o parlamento britânico a criar normas de higiene que puderam salvar muitas vidas (SCHACHT, 2017).

Desde então, existem poucos relatos de histórias que aplicaram métodos baseado em dados para desenvolver reportagens. Apenas com a chegada dos primeiros computadores ao mercado que foi possível coletar e analisar dados em larga escala. Em 1952, a equipe da rede americana de radiodifusão CBS foi uma das primeiras a usar esses métodos para prever resultados eleitorais. A equipe alimentava os resultados e pesquisas de antigas eleições presidenciais utilizando um UNIVAC (sigla para *Universal Automatic Computer*, em inglês) da Remington Rand, um dos primeiros computadores comerciais (GRAY et al., 2012).

Quando o UNIVAC declarou que as probabilidades eram de 100:1 a favor do general Dwight D. Eisenhower, a equipe da CBS supôs que fosse um *bug* no computador e não transmitiram os resultados. Porém a máquina estava certa, Eisenhower venceu, e o

computador estimou o resultado da eleição com precisão de até um ponto percentual - um avanço para a indústria de computadores, bem como sua aplicação ao jornalismo.

Nos primeiros anos, até mesmo décadas da tecnologia da informação, ter acesso a uma dessas máquinas não era nada fácil e barato. A maior parte do trabalho ainda era feita à mão, mesmo que fosse tecnologicamente possível para um computador fazer a tarefa. Realizar uma pesquisa eleitoral em 1962 ainda significava ir de porta em porta, coleccionar questionários e classificá-los manualmente por partido, candidato, idade, sexo - muitas horas de trabalho para algo que, hoje, levaria apenas alguns cliques no computador. O ano de 1952 foi um marco, pois foi a primeira vez que foi utilizado o RAC – Reportagem com Auxílio de Computador (do inglês, *CAR – Computer-Assisted Reporting*) para prever o resultado da eleição presidencial (GRAY et al., 2012).

Mais tarde, no ano de 1967, o então jornalista Philip Meyer estabeleceria outro marco no campo. Esse período foi marcado pelos protestos dos direitos civis em Detroit, numa das manifestações mais violentas da história americana, com 43 mortos e mais de 7.000 presos, como já ressaltamos anteriormente. Meyer com a ajuda de um *mainframe* fez levantamentos estatísticos para mostrar que as pessoas que frequentavam a faculdade eram igualmente suscetíveis a promoverem revoltas, assim como as que haviam chegado ao ensino médio (GRAY et al., 2012; SCHACHT, 2017).

Embora o uso de computadores no jornalismo remonte à década de 1950 (COX, 2000), o padrinho *de fato* do RAC é Philip Meyer, que delineou uma nova forma de produção de conteúdo, conhecido como jornalismo de precisão, que rendeu um livro de mesmo nome (MEYER, 2002). O jornalismo de precisão foi modelado a partir das ciências sociais, usando métodos empíricos (particularmente pesquisas e análise de conteúdo) e análise estatística para obter respostas mais definitivas às questões jornalísticas (ibid.). Desde 1960, jornalistas, principalmente do campo investigativo, buscam monitorar o Estado através das análises de base de dados de registros públicos usando métodos científicos. Também conhecido como “jornalismo de serviço público”, os defensores dessas técnicas assistidas por computador têm procurado revelar tendências, desconstruir o conhecimento popular sobre certos tópicos e revelar injustiças perpetradas por autoridades públicas e empresas privadas (GRAY et al., 2012; SCHACHT, 2017).

A reportagem de Bill Dedman sobre “*The Color of Money*” (A cor do dinheiro, tradução nossa) nos anos 80 revelou viés racial sistêmico nas políticas de empréstimo das principais instituições financeiras. No artigo “*What Went Wrong*” (O que deu errado,

tradução nossa), Steve Doig procurou analisar os padrões de danos causado pelo furacão Andrew no início dos anos 1990 para entender o efeito de políticas e práticas de desenvolvimento urbano erradas. As reportagens baseados em dados ajudaram a implementar políticas públicas, prestando um serviço valioso, e renderam prêmios famosos aos jornalistas (CODDINGTON, 2015; GRAY et al., 2012; SCHACHT, 2017).

Apenas no final dos anos 80 e início dos anos 90 que o jornalismo de precisão, desde a reformulação como RAC, começou a fazer incursões significativas nas redações, liderado por várias histórias vencedoras do Prêmio Pulitzer, que se tornaram um forma de reconhecimento e validação profissional da prática, evoluindo para o que é conhecido hoje como jornalismo de dados (CODDINGTON, 2015; SCHACHT, 2017). Embora o jornalismo de dados tenha sido usado informalmente por profissionais de RAC durante décadas, o primeiro uso registrado do termo por uma grande organização de mídia é do *The Guardian*, que lançou seu *Datablog* em março de 2009. Embora a paternidade do termo seja contestada, é amplamente utilizado desde que os documentos da Guerra do Afeganistão do *Wikileaks* vazaram em julho de 2010. A cobertura do *The Guardian* dos registros de guerra tirou proveito das ferramentas gratuitas de visualização de dados, as soluções *out-of-the-box*, como o *Google Fusion Tables*, outro aspecto comum do jornalismo de dados para produzir conteúdo relevante para sua audiência (CODDINGTON, 2015; GRAY et al., 2012; SCHACHT, 2017).

3.2 RAC, Jornalismo de Dados e Computacional

Apesar de RAC e jornalismo de dados estarem relacionados, autores discordam que a prática possam ser consideradas a mesma. Coddington (2015) defende no seu artigo “*Clarifying journalism’s quantitative turn*” que o jornalismo quantitativo pode ser dividido em três tipos principais: RAC – Reportagem com Auxílio de Computadores, Jornalismo de Dados e Jornalismo Computacional. Para ele, o importante é diferenciar esses três tipos, uma vez que há uma dificuldade para muitos pesquisadores de distingui-los e todos eles têm uma importância para ampliar a tarefa do jornalista para alcançar pessoas de forma a permitir que elas tenham acesso à democracia e ao pensamento crítico.

A Reportagem com Auxílio de Computador (RAC) foi popularizada, como mencionado na seção anterior, durante os anos cinquenta, quando o uso de pesquisa, análise de conteúdo e estatística eram utilizados para validar questões jornalísticas, mas sua base

está no jornalismo de precisão criado por Meyer. Até o final dos anos 80 e início dos anos 90, a técnica de jornalismo de precisão, desde a reformulação como RAC, começou a abrir caminho nas redações e tornou-se muito ligado ao jornalismo investigativo, sendo muitas vezes visto como ferramenta para auxiliar em projetos de jornalismo de longo prazo e de assuntos públicos (CODDINGTON, 2015).

O Jornalismo de Dados ou Jornalismo Guiado por Dados teve um grande *boom* na última década, em que a comunicação digital permitiu a produção de conteúdo visual de forma fácil. A principal diferença é que os valores de *design* de visualização e jornalismo estão no centro do jornalismo de dados (GORDON, 2013 apud CODDINGTON, 2015). O termo jornalismo de dados entrou em voga, pois foi o escolhido pela indústria de notícias para o jornalismo baseado na análise de dados e na apresentação de tal análise ao fim dos anos 2000 (CODDINGTON, 2015).

O último tipo defendido por Coddington (2015) no seu artigo é o jornalismo computacional, que é o uso de um conjunto de ferramentas para um “jornalismo orientado tecnologicamente e centrado em uma aplicação de computação e pensamento computacional”⁴ (p. 332, tradução nossa). Esse tipo usa da combinação de algoritmos, dados e conhecimento das ciências sociais para complementar a função social do jornalismo. Um exemplo de utilização desse tipo de jornalismo que reúne informações ou as apresenta, por exemplo, são os jornalistas robôs, também conhecido como *bots*. O jornalismo computacional é construído sob dois pilares: abstração e automação. Abstração é a capacidade de decompor informações ou problemas além de seu contexto material imediato, ou seja, é um processo cognitivo. A computação, então, é simplesmente a automação de informações abstratas e processos. Esses processos de automação geralmente tomam a forma de algoritmos, que ocasionalmente são considerados um terceiro elemento do pensamento computacional (CODDINGTON, 2015).

3.3 O desenvolvimento do jornalismo de dados no mundo

Os debates acadêmicos sobre jornalismo de dados foram implementados por vários pesquisadores ao longo dos últimos anos, em geral com respeito às normas, práticas e

⁴ No original: “[...] technologically oriented journalism centered on the application of computing and computational thinking [...]”

culturas de jornalismo de dados a nível nacional. Dentre esses estudos incluem-se os dos Estados Unidos (FINK & ANDERSON, 2015; PARASIE & DAGIRAL, 2013; ROYAL & BLASINGAME, 2015; YOUNG & HERMIDA, 2015), Reino Unido (BORGES-REY, 2016; HANNAFORD, 2015), Canada (HERMIDA & YOUNG, 2017; TABARY et al., 2016), Bélgica (DE MAEYER et al., 2015), Noruega (KARLSEN & STAVELIN, 2014), Suécia (APPELGREN & NYGREN, 2014) e Espanha (FERRERAS RODRÍGUEZ, 2016). Já no Brasil, a prática ainda é pouco estudada tendo apenas contribuições de poucos autores com enfoque em tipologias ou análise de conteúdos de alguns meios, como o Nexo (ARAÚJO, 2018; MANCINI & VASCONCELLOS, 2016; RODRIGUES, 2014; TRÄSEL, 2013, 2014, 2018; VASCONCELLOS et al., 2015; VENTURA, 2018).

Como a prática tem amadurecido desde então, pesquisadores tem usados diferentes métodos para responder a problemática estudada, tais como entrevistas qualitativas ou levantamentos para identificar e entender como as partes interessadas estão se envolvendo nos principais veículos e como estão sendo implementadas (HERMIDA & YOUNG, 2017).

A pesquisa etnográfica de Royal e Blasingame (2015) teve enfoque em como a equipe de interativos do jornal americano *The New York Times* utilizava dados como parte dos projetos de reportagem baseada em dados. Os autores inferiram que as partes envolvidas os viam como jornalistas afetados pelos valores do mundo digital de colaboração, inovação e criatividade. Em outro trabalho nos Estados Unidos, os autores Parasie e Dagiral (2013) mostram que em Chicago existe a presença dos jornalistas-programadores, também conhecido como *journo-coders*, que surgiram com o desenvolvimento do jornalismo de dados e a possibilidade de utilização de tecnologia para interação do usuário em sites de notícias. Já o estudo de Young e Hermida (2015) mudou de enfoque de uma região ou país e concentrou em um veículo de mídia tradicional dos EUA, o *Los Angeles Times*. Suas descobertas refletiram o que Parasie e Dagiral (2013) haviam percebido em Chicago, um movimento de crescente da demanda por jornalistas-programadores pelo veículo. Eles também sugeriram que trazer programadores para a redação ajuda moldar os ideais jornalísticos, onde a tecnologia ajuda a dar formas e contornos ao jornalismo, assim como o processo inverso também ocorre (c.f. HERMIDA & YOUNG, 2017).

Ainda na América do Norte, os autores Tabary, Provost e Trottier (2016) usaram diferentes abordagens para analisar as reportagens de dados publicadas por uma empresa de mídia da região de Quebec no Canadá. Eles sugeriram que as reportagens são produzidas

por atores especializados e apontaram que habilidades computacionais e estatísticas não eram tão desenvolvidas nos veículos da região.

Na Europa, os estudos mostraram diferentes alinhamentos em relação à tecnologia e à identidade do jornalismo. No Reino Unido, muito menos tem sido mencionado sobre os *jouno-coders* e sobre sua existência nas organizações de mídia, embora alguns estudos de casos da primeira edição do Manual de Jornalismo de Dados (*Data Journalism Handbook*, em inglês) mostrem a sua importância (GRAY et al., 2012; HERMIDA & YOUNG, 2017).

Nos estudos conduzidos por Hannaford (2015), os dois maiores veículos de mídia do Reino Unido, *BBC* e *Financial Times*, foram analisados e os resultados mostraram a presença de equipes híbridas trabalhando na produção de dados, com programadores, jornalistas e designers trabalhando juntos para produzir conteúdo multimídia e interativo (HANNAFORD, 2015; HERMIDA & YOUNG, 2017). Essa abordagem foi amplamente adotada por outras empresas de mídia no mundo, como na Austrália e Brasil (MIELNICZUK; TRÄSEL, 2014; TRÄSEL, 2017).

Outro autor que obteve resultados semelhantes em sua pesquisa foi Borges-Rey(2016), que conduziu 24 entrevistas semi-estruturadas com profissionais de mídias, tais como *The Guardian*, *BBC*, *Financial Times*, *Channel 4*, *Trinity Mirror*, *The Times*, *CNN*, *Thomson Reuters*, *Telegraph* e outros veículos de menor porte. As descobertas do autor se baseiam também na análise de produção de conteúdo de jornalístico e descobriram que o uso de interativos e formas envolventes para aumentar a experiência do usuário informacional que “oferecem conteúdo de várias camadas, multiplataforma, gamificado e dinâmico vinculado a base de dados”⁵ (BORGES-REY, 2016, p. 161, tradução nossa).

Na Europa continental, Karlsen e Stavelin(2014) também conduziram entrevistas nas seis maiores redação da Noruega e encontraram que “as percepções de continuidade superaram a mudança entre os entrevistados com as habilidades técnicas (...) subordinadas à tradição ininterrupta do jornalismo” (c.f. HERMIDA & YOUNG, 2017, p. 161, tradução nossa), mostrando a necessidade de projetos de longo prazo para a produção de jornalismo de dados (KARLSEN & STAVELIN, 2014). Ainda nos países nórdicos, os pesquisadores Appelgren & Nygren (2014) apontaram um crescente interesse em jornalismo de dados na Suécia pelos atores envolvidos, e indicaram que a prática ainda está em sua infância no país.

⁵ No original: “[...] interactive and engaging forms of informational user experience that offer multi-layered, multiplatform, gamified, database-linked dynamic content [...]”

Na Bélgica, De Maeyer et al. (2015) encontraram uma tendência que os *early adopters* estavam produzindo histórias baseadas em dados “simples” sem adoção pelas organizações de mídia tradicionais, criando um “sentimento geral de resignação com grandes projetos” contra as restrições das rotinas de produção de notícias e a “falta de apoio institucional ou estrutural” dos meios midiáticos (DE MAEYER et al., 2015, p. 444, tradução nossa). Os autores destacaram a necessidade de promover jornalismo de dados com o apoio estrutural nas redações, bem como a necessidade de ser valorizado pelo público. Para os autores, avaliar se e como o jornalismo de dados tem sido útil para o desenvolvimento de práticas no jornalismo ajuda na sua promoção e adoção pelas redações (DE MAEYER et al., 2015).

O jornalismo de dados representa, para muitos autores, a democratização de recursos, ferramentas, técnicas e métodos que antes eram restritos a poucos, como especialistas, governos, entre outros. Os profissionais que trabalham com a prática têm papel essencial em atenuar as barreiras para a compreensão e imersão nos dados e aumentar a alfabetização de dados em larga escala da audiência (GRAY et al., 2012). Entretanto, organizações de mídia são atores fundamentais nesse processo, uma vez que os jornalistas estão acostumados a um mundo onde eles são responsáveis pela produção de notícias baseadas em entrevistas ou fatos. Criar ferramentas e práticas que permitam preparar sua força de trabalho para lidar e entender os valores do jornalismo de dados é essencial. Para estar pronto para esse novo ambiente, veículos de mídia precisam olhar para todo o ecossistema midiático e entender os papéis de cada ator na produção de notícias baseadas em dados e a importância do jornalismo de dados na produção de conteúdo de qualidade. Os veículos tradicionais estão acostumados a uma hierarquia estrita, que tem um impacto negativo na inovação e produção de novas formas e formatos noticiosos, gerando um efeito direto no jornalismo de dados. “Recursos, vontade da liderança e possibilidade para implantar [reportagens baseadas em dados] são fatores determinantes para que as habilidades de jornalistas de dados e computacionais exerçam suas atividades” (HERMIDA & YOUNG, 2017, p.171, tradução nossa).

Uma das limitações para não comprometer a integridade jornalística é a simplificação excessiva de reportagens devido à limitação de tempo ou recursos (USKALI & KUUTTI, 2015). Essas decisões podem levar a uma falta de confiança no jornalismo e diminuição da qualidade dos artigos produzidos pelas organizações que não estão preocupados com este ponto. Dessa forma, é essencial que os *stakeholders* envolvidos

entendam a importância de uma reportagem de jornalismo de dados e o *pipeline* de produção para que a equipe possa ser reconhecível por sua produção de conteúdo de qualidade e isento (BRADSHAW, 2007, 2013).

O jornalismo de dados foi capaz de criar uma ‘ponte’ entre tecnólogos e jornalistas nas redações, expandindo seus horizontes nas práticas e nos processos e evoluindo para além das ilhas de conhecimento, permitindo explorar novos tópicos em um nível mais profundo (CAIRO, 2016; HERMIDA & YOUNG, 2017). Portanto, promover o jornalismo de dados é essencial para a criação de um único nome e imagem para a audiência e para as redações, principalmente através de uma consciência sobre a importância de estabelecer um significado e uma presença de qualidade nos mercados de notícias para atrair e reter clientes leais.

O jornal americano *The New York Times* atingiu recentemente um marco com seu modelo de negócios que tem como principal objetivo criar um maior compromisso com a audiência, levando a um aumento maciço na receita de assinaturas (PEISER, 2018). O britânico *The Guardian*, que possui um modelo de negócios baseado em doação dos leitores, anunciou no último ano que “a companhia estava no caminho para o *break-even* (ponto de equilíbrio)” (SWENEY, 2018). O que esses dois veículos tradicionais têm em comum? Eles fazem um massivo investimento em jornalismo de dados e em outras formas de jornalismo digital, criando conteúdo de qualidade e envolvente para o público, e assim, fortalecendo a organização contra o debate público em torno da confiança no jornalismo.

Entender como o público vê e como as organizações de mídia alcançaram um nível de maturidade no desenvolvimento de jornalismo de dados é essencial para mostrar a sua importância e não utilizar como desculpa a falta de dinheiro e tempo para não investir nesse campo. Dessa forma, há um fortalecimento da capacidade competitiva dos veículos para sobreviver na era digital (HERMIDA & YOUNG, 2017; OBERMAYER & OBERMAIER, 2016).

3.4 Entendendo o jornalismo de dados no Brasil

Enquanto no mundo jornalismo de dados está em expansão e consolidação, a prática no Brasil ainda está em desenvolvimento. Um dos pioneiros no trabalho de jornalismo de dados e RAC foi Cláudio Weber Abramo, que morreu em 2018, aos 72 anos de idade.

Abramo era bacharel em matemática pela USP e mestre em filosofia da ciência pela Unicamp (ASSIS, 2018; ROCHA, 2018; VENTURA, 2018).

O jornalista foi editor de Economia da Folha de São Paulo (1987) e secretário-executivo da Gazeta Mercantil (1987-88), além de ter passado por outras publicações como Valor Econômico, Abril e IstoÉ. Pioneiro do jornalismo de dados no Brasil com projetos como os ‘Às Claras’ e ‘Excelências’, Abramo era defensor da integridade das instituições públicas, da transparência, do acesso a dados públicos e uma referência no combate à corrupção no Brasil.

Como pai da prática decidiu fundar em 2000, junto a um grupo de indivíduos e entidades não-governamentais comprometidos com o combate à corrupção, a Transparência Brasil, organização que se dedica ao acesso à informação e ao combate à corrupção. Entre os vários projetos desenvolvidos pela organização está o ‘Excelências’, uma base de dados online com informações de processos e do desempenho de parlamentares, ganhador do Prêmio Esso de Melhor Contribuição à Imprensa em 2006. Outro projeto de destaque de Abramo foi o ‘Às Claras’, uma plataforma que organizava e disponibilizava gastos das eleições (ASSIS, 2018; ROCHA, 2018). O matemático-jornalista também foi um dos mais importantes articuladores para a elaboração da Lei de Acesso à Informação (LAI), aprovada em 2011, que assegurava o direito de acesso aos dados produzidas por órgãos e entidades da União, Estados e Municípios (ibid).

O trabalho de Abramo abriu portas para outras organizações começarem a trabalhar com dados abertos e jornalismo de dados. Em 2012, o Estado de São Paulo criou o Estadão Dados, a primeira equipe voltada para a produção de reportagens baseadas em dados em uma redação. Considerado hoje um dos casos mais bem-sucedidos, a equipe já criou reportagens baseadas em dados do Twitter, como “O que 15 mil tweets revelam sobre seu candidato” para as eleições presidenciais de 2018 (ASSIS, 2018; ROCHA, 2018).

Outra equipe que vem se destacando é a G1 Dados, do portal G1.com pertencente ao grupo Globo. Com base na cidade de São Paulo, eles foram um dos primeiros brasileiros premiados no maior prêmio de jornalismo de dados global, o *Data Journalism Awards*. Liderados pelo jornalista de dados Thiago Reis do maior grupo de mídia brasileiro, o projeto foi premiado na categoria escolha pública de 2018. O “Monitor da Violência” prometia investigar os assassinatos no Brasil ao longo de uma semana. O objetivo era claro: entender o que está impulsionando a violência, humanizar os dados e apelar ao governo que tome uma ação. O grosso da investigação ocorreu entre 21 e 27 de agosto de 2017 e envolveu

230 jornalistas de 55 redações afiliadas da emissora em todo o país. O projeto foi tão grande que reuniu não só jornalistas do grupo, como também uma universidade e uma ONG para fazer uma investigação a nível nacional sobre as mortes pela violência. Entre os resultados assustadores obtidos, eles descobriram que houve um assassinato a cada oito minutos (BOUCHART, 2017).

Um ano antes da equipe do G1, a ABRAJI (Associação Brasileira de Jornalistas Investigativos) ganhou o mesmo prêmio na edição de 2017, porém na categoria de pequenas redações jornalísticas. Competindo com 99 trabalhos de 30 países, o projeto CTRL+X mapeava as tentativas de censura a veículos de mídia no Brasil. O portal apresentava na época do prêmio cerca de 2500 processos judiciais de políticos contra empresas de mídia. Segundo Tiago Mali, coordenador do projeto, a ferramenta foi útil para acompanhar a liberdade de expressão no país, principalmente durante o período eleitoral de 2016 (NAHSER, 2018).

Foi tão importante o trabalho de Abramo para o jornalismo de dados, que abriu a portas para esses reconhecidos projetos. Hoje, o jornalista-matemático dá nome ao recém-criado prêmio para reconhecer a excelência na produção de trabalhos jornalísticos guiados por dados no Brasil, o prêmio Cláudio Weber Abramo de Jornalismo de Dados⁶, que terá sua primeira edição ao final de 2019. Com esse reconhecimento, espera-se que o Brasil consiga se desenvolver ainda mais na produção de jornalismo de dados e atinja a maturidade alcançada por outros países.

⁶ <https://jornalismodedados.org>, Acesso em: 1º de maio de 2019.

4. Dados sobre dados: Quem trabalha com jornalismo de dados no Brasil?

Quando comecei a estudar sobre o assunto, a ideia inicial era fazer um estudo sobre quem estava trabalhando com o jornalismo de dados no Brasil. Em conversa com minhas orientadoras, percebemos a necessidade de usar o jornalismo de dados para mostrar sua presença no país e tentar entender, dessa forma, quem são os principais atores que estão desenvolvendo essa prática no país. A primeira tarefa e talvez mais desafiadora era como mapear a maior parte dessas organizações. No mundo, o jornalismo de dados é conhecido por ser uma prática jornalística que tem muitos valores do mundo *hacker*, como colaboração e tecnofilia (TRÄSEL, 2018). Acreditando nesses valores, utilizamos um questionário que foi enviado para grupos de pessoas que trabalham com dados no Brasil. Esses grupos que tinham como foco tópico de dados abertos quanto jornalismo de dados foram de grande ajuda no mapeamento inicial das organizações. Com o objetivo de obter o maior número possível de empresas que trabalham com jornalismo baseado em dados, recorremos ao Twitter para encontrar outros atores que não chegaram até o questionário.

Agregando as duas bases de dados, foi obtido um total de cinquenta e dois ($n=52$) organizações que produzem histórias baseadas em dados e com links para exemplos desses trabalhos. O trabalho estava só por começar, como Bradshaw (2007) estabeleceu. A compilação de dados era apenas a etapa inicial, seria preciso ainda limpar para eliminar informações duplicadas ou faltantes. Depois dessa fase, era preciso contextualizar os dados, checar se eles eram confiáveis e combinar com informações de outras fontes para então apresentá-los de forma visual. A visualização é uma das etapas mais relevante a ser feita, pois mostrar os resultados é o que diferencia a prática do uso de dados pelo jornalismo. Outro ponto considerado nesse trabalho foi deixar a experiência do usuário da melhor forma possível, criando uma visualização clara, explícita e acionável. Nos itens seguintes, cada etapa será descrita detalhadamente até a obtenção do produto jornalístico.

4.1 Coleta e limpeza

Conforme foi mencionado antes no capítulo de Metodologia, o trabalho utilizou de diferentes práticas para a produção do produto jornalístico resultante, baseando na pirâmide invertida desenvolvida por Bradshaw (2007), conforme mostrado figura 1. Essa pirâmide descreve o *pipeline* de produção de uma história baseado em dados. Para o objetivo de obter um mapa de onde há produção de matérias baseadas em dados, foi criando um questionário

que foi enviado para grupos de WhatsApp de dados abertos, de Facebook de jornalismo de dados e dados abertos.

O questionário foi dirigido a dois públicos alvos: pessoas que trabalham com jornalismo de dados e pessoas que se interessam sobre o assunto. Conforme o perfil do informante, as perguntas eram diferentes. Para separar os dois grupos, as questões iniciais focavam em entender se a pessoa trabalhava em alguma organização com jornalismo de dados ou ela pessoalmente trabalhava com a prática. Aos informantes que trabalham com jornalismo de dados foram, então, direcionadas questões que perguntavam o nome, cidade, estado, CEP da organização, assim como *link* para uma história baseada em dados e contatos do responsável pela área de jornalismo de dados. Já aqueles que responderam que não trabalham em organizações de mídia foram pedidos os dados similares, porém com enfoque de cadastro de organizações de mídia. Ao todo foram coletados vinte ($n=19$) entradas com dados de organizações.

Em paralelo, além desse questionário, foram coletados de dados do Twitter com as hashtags #ddjbr e #jornalismodedados para complementar a base de dados de organizações que trabalham com jornalismo de dados. Foram coletados dados do Twitter durante o início de 2019, de março até meados de abril, através da plataforma Workbench⁷. Apesar da comunidade de jornalismo de dados brasileira produzir pouco conteúdo no Twitter comparado a mundial, através das hashtags #ddjbr e #jornalismodedados foi possível encontrar outros veículos que não haviam sido mapeados pelos informantes, assim como um perfil @ddj_br, que produzia newsletter até o fim de 2018 sobre reportagens baseadas em dados. Esse perfil foi essencial para mapear exemplos de histórias das grandes organizações.

Foram agregados os dados de todas essas fontes, que, então, foram limpados para eliminar informações duplicadas ou faltantes. Após essa etapa, foi possível gerar um banco de dados com cinquenta e dois ($n=52$) nomes e locais das organizações de mídia. Das entidades mapeadas é possível ver que grande parte delas são organizações de mídia independentes ou nativas digitais (51,92%), desde as mais populares como UOL, Nexo, G1, Poder360, Intercept até as menores e mais desconhecidas da grande população como Colaborados, Datalab, Filtro Fact-Checking, Volt Data Lab, entre outras. Logo atrás estão as mídias tradicionais que representam 36,5% do total, a lista inclui empresas tradicionais brasileiras, como TV Globo, Globonews, Estado de S. Paulo, Folha de S. Paulo, assim como

⁷ Workbench - <http://workbenchdata.com/>

oriundas de fora do país, como BBC e El País. A menor parte é composta de organizações que não empresas de mídia, porém contam com jornalistas de em seu corpo de funcionários produzindo história baseada em dados, além de um blog de um jornalista de dados.

Tabela 1 – Número de Organizações por cidade

<i>Tipo de Organização</i>	<i>Organização de mídia</i>
Mídia Tradicional	19 (36,5%)
Nativa Digital	27 (51,92%)
Organização não-midiática	4 (7,69%)
Outros	2 (4%)
Total	52

Fonte: Elaborada pelo autor com base nos dados coletados.

4.2 Organização e exploração de dados

Os dados coletados foram organizados em uma planilha do Google Drive para serem facilmente atualizados e estarem disponíveis em qualquer lugar, já que ficam salvos na nuvem. Na planilha foi utilizada a programação através de *scripts* e fórmulas para atualização de alguns campos da tabela, que permitiu obter de forma automatizada partes dos dados.

Figura 3 – *Script* usado para gerar longitude e latitude.

```
function GEOCODE_LAT(address) {  
  if (address.map) {  
    return address.map(GEOCODE_GOOGLE)  
  } else {  
    var r = Maps.newGeocoder().geocode(address)  
    for (var i = 0; i < r.results.length; i++) {  
      var res = r.results[i]  
      return res.geometry.location.lat  
    }  
  }  
}  
  
function GEOCODE_LONG(address) {  
  if (address.map) {  
    return address.map(GEOCODE_GOOGLE)  
  } else {
```

```
var r = Maps.newGeocoder().geocode(address)
for (var i = 0; i < r.results.length; i++) {
  var res = r.results[i]
  return res.geometry.location.lng
}
}
```

Fonte: Elaborada pelo autor.

Com os dados obtidos pelos informantes e através do Twitter, pode-se verificar que existiam informações que não estavam completas, devido a falta de conhecimento sobre esse dado por parte do informante. Por isso, as organizações que possuíam apenas as cidades, já que entre os dados pedidos no formulário estavam ‘cidade’ e ‘CEP’ da empresa (caso o informante conhecesse), e não possuíam endereço disponíveis tanto no Google Maps, quanto em seus próprios websites, foram colocadas em pontos centrais da cidade onde se localizam. As demais, foram adicionados manualmente os endereços corretos buscando no Google Maps e adicionados os *links* em uma planilha junto as informações de descrição de cada meio. Através de um *script* foi possível coletar a latitude e longitude de cada um dos endereços disponíveis e colocar com precisão onde se localiza cada meio, conforme figura 3. Foi utilizado os códigos de geolocalização da API do Google Maps para coletar a latitude e longitude exatas.

Os dados de informação das empresas, assim como os logos, foram coletados manualmente. As imagens com os logos foram armazenadas no Dropbox e foram adicionados os *links* de cada imagem na base de dados. Já os *links* de exemplo de história baseados em dados e do portal das organizações foram inseridos pelos informantes. Após todas essas etapas, foi obtida uma base de dados com cinquenta e dois ($n=52$) organizações que trabalham com jornalismo de dados.

4.3 Visualização

Com os dados coletados, compilados e limpos, foi necessário voltar a pergunta proposta para entender a melhor forma de visualizá-los. O que esse projeto propõe é responder onde há jornalismo de dados no Brasil. Tendo essa proposta em mente, a melhor

forma para fazer isso seria a utilização de uma ferramenta para mapeamento. Com o objetivo de ajudar o trabalho de jornalistas, *startups* estão desenvolvendo ferramentas prontas que eliminam a necessidade de programação do usuário e oferecem elegantes formas de visualização dos dados. Essas soluções são conhecidas como *out-of-the-box* e ajudam na produção de reportagens principalmente para pequenas e médias organizações de mídia que não possuem força de trabalho suficiente, como já enfatizamos. Nesse caso, a ferramenta *out-of-the-box* Carto⁸ foi escolhida para apresentação dos dados, pois permite a visualização e análise de dados geográficos, além de um mapa utilizando em associação com programação para web. CARTO, antes chamada de CartoDB, é uma plataforma de computação em nuvem de Software como Serviço (*SaaS*) que fornece ferramentas de sistema de informação geográfica (SIG ou *GIS - Geographic Information System*, do acrônimo em inglês) e ferramentas de mapeamento na web para exibição em um navegador.

Apesar de ser uma ferramenta simples, o Carto possui interface para programação onde é possível alterar *layout, design*, assim como a leitura dos dados com código SQL. Para o produto jornalístico foi preciso utilizar codificação CSS e HTML para ajustar detalhes da visualização, como o número de cores a serem exibidas. A primeira etapa foi a escolha do tipo de mapa de base que seria utilizado para exibição dos dados. Foi escolhido o *Voyage* com base nos princípios de clareza e evitar uma poluição visual com rodovias e informações hidrográficas que não serão úteis nesse caso (CAIRO, 2016). O mapa base apresenta as divisões de estados e países, facilitando a localização visual dos dados apresentados, conforme mostra a figura 4.

Tabela 2 – Número de Organizações por cidade

Cidades	Total de organizações por cidade.
Brasília	3
Campo Grande	1
Curitiba	2
Florianópolis	2
Fortaleza	1
Porto Alegre	1
Recife	2
Rio Claro	1
Rio de Janeiro	13

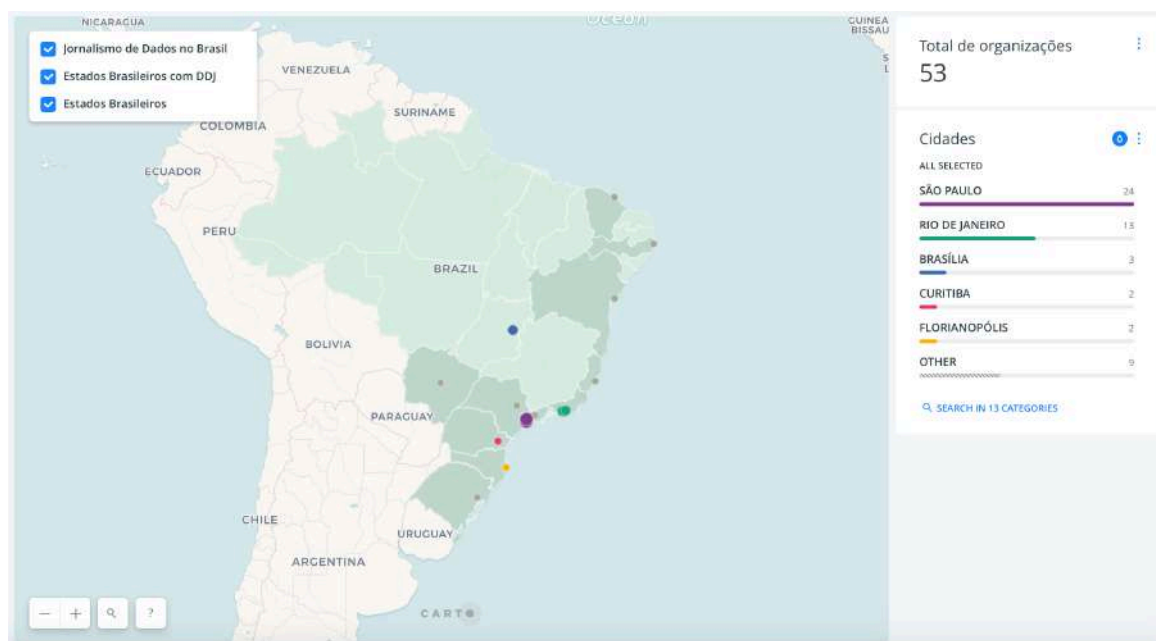
⁸ Disponível: <https://carto.com>. Acesso no período de novembro de 2018 a maio de 2019. Como se trata do uso constante da plataforma que foi utilizada durante toda realização do projeto foram realizados múltiplos acessos compreendendo o período em referência.

Salvador	1
São José dos Campos	1
São Paulo	23
Vitória	1
Total de organizações	52

Fonte: Elaborada pelo autor.

Foram utilizadas treze (13) cores para representar as cidades onde existe pelo menos uma organização atuando com jornalismo de dados. O tamanho da circunferência é maior conforme a maior presença de organizações dentro da mesma cidade, São Paulo e Rio de Janeiro lideram, respectivamente, como mostra a tabela 2. Para visualizar essas informações em forma numérica foi decidido colar um *widget* lateral que apresenta o total de organizações e o total por cada cidade. Conforme o usuário aproxima o mapa de uma cidade e as demais não aparecem, tanto o número total de organizações quanto as cidades atualizam para apresentar apenas os dados correspondentes das cidades que estão enquadradas no momento.

Figura 4 – Mapa base *Voyage* com os pontos onde há presença de jornalismo de dados.

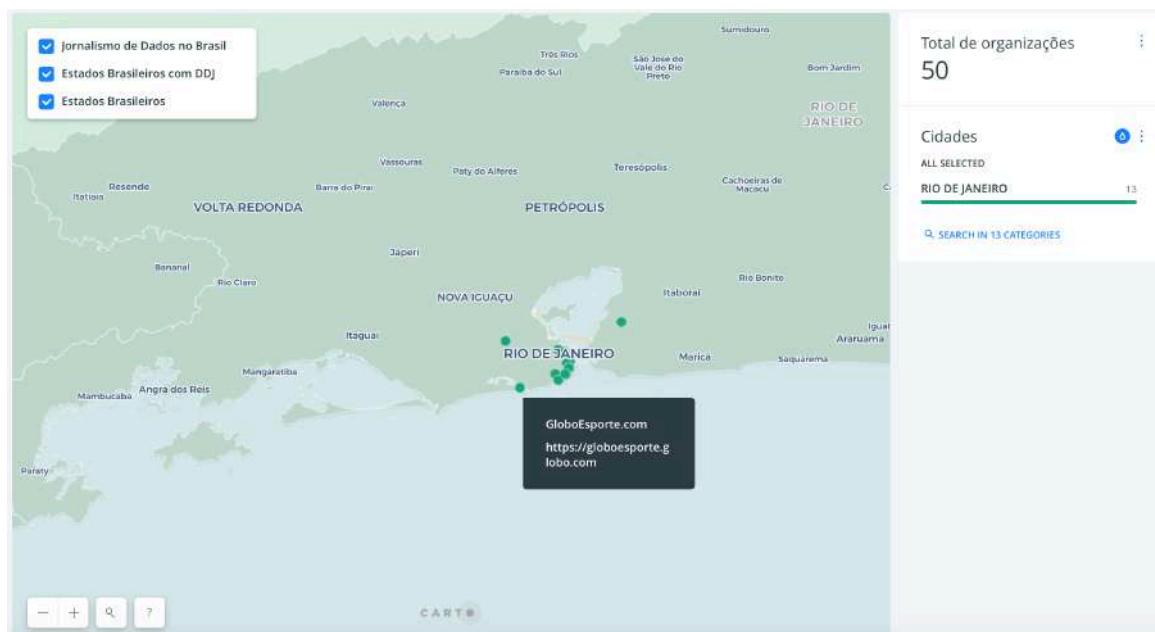


Fonte: Captura realizada pelo autor.

Para melhorar a experiência do usuário e permitir que ele encontre mais fácil as organizações, foi criado uma janela de fundo preto que aparece toda vez que o usuário parar o *mouse* sobre um ponto. Foi decidido que nessa janela deveria ter o mínimo de informação

de forma a não poluir visualmente, para o usuário decidir se deveria clicar ou não no ponto, por isso cada ponto quando passa o *mouse* aparece o nome da organização e o *site*, conforme mostra figura 5. Caso o usuário deseje obter mais informações será necessário clicar no ponto.

Figura 5 – Quando o usuário passa o *mouse* sobre um ponto, uma janela aparece.



Fonte: Captura realizada pelo autor.

Ao clicar em qualquer ponto, o usuário encontrará um janela com fundo branco com mais informações, conforme figura 6. Nessa janela *pop-up* aparece quatro informações principais: logo da organização, nome, texto descritivo e *link*. Os logos das organizações aparecem no topo da janela, com uma pequena caixa com o nome no canto inferior esquerdo. Abaixo encontra-se uma descrição da organização e sobre a área que trabalha com jornalismo de dados na entidade, quando disponível. Nesse texto, *hyperlinks* para histórias baseadas em dados e para organização aparecem sobre o texto. Por último aparece novamente o *link* da organização, caso o usuário não tenha percebido a existência de *links* na descrição.

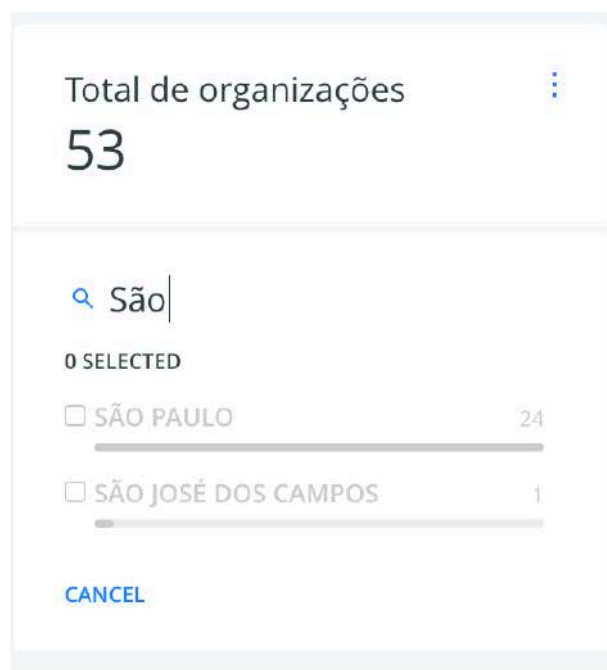
Figura 6 – Janela com as mais informações aparece ao clicar em um ponto.



Fonte: Captura realizada pelo autor.

Conforme mencionado anteriormente, foi adicionado um *widget* lateral que permite ao usuário buscar uma determinada cidade. Ao digitar o nome inicial de uma cidade aparece uma sugestão de todas as cidades que estão presentes na base de dados e iniciam com o mesmo nome, facilitando a busca do usuário, ver figura 7.

Figura 7 – Busca por cidade



Fonte: Captura realizada pelo autor.

Esse mapa foi produzido de forma que possa ser facilmente embutido em um *website*, permitindo que assim outras pessoas ou organizações possam adicionar em seus portais disseminando o jornalismo de dados para mais brasileiros.

Esse produto jornalístico, “Tem #DDJBR aqui? Mapa do jornalismo de dados no Brasil”, pode ser acessado através do link abaixo: <http://bit.ly/MapaDDJBr>

5. Conclusão

O projeto não só coloca em prática o jornalismo de dados, mas permite fazer análise críticas sobre a relevância dessa disciplina no país. A presença massiva de organizações nas duas maiores capitais, São Paulo (23 empresas) e Rio de Janeiro (13), mostram que o jornalismo de dados ainda não está consolidado no país. A limitação dessa prática nas capitais pode ter reflexo no papel fundamental do jornalismo na atualidade. No nosso entendimento o papel da imprensa como cão de guarda⁹ para sociedade pode ser influenciado por essa falta de desenvolvimento do jornalismo de dados.

O país tem avanços significativos na política de dados abertos com a Lei da Transparência. Essa lei brasileira, sancionada em 2009, obriga a União, os estados e os municípios a divulgar seus gastos na Internet em tempo real. As investigações jornalísticas tradicionais estão longe de abraçar dados como fonte de informação, por isso o jornalismo de dados é fundamental para a atuação do jornalismo.

O presente projeto não aborda as razões que levam as organizações de mídia brasileira a não adotarem a prática. Mas pesquisas no mundo mostraram que a limitação de tempo ou recursos influenciam adoção do jornalismo de dados. A crise da indústria jornalística afeta a decisão por reportagens que demandam tempo, pois cada recurso investido (funcionário) necessita ser bem aproveitado. O erro está exatamente nessa concepção, jornais tradicionais que investiram em novas formas de conteúdo para engajar a audiência e conseguir criar novos modelos de negócios sustentáveis, como *The Guardian* e *The New York Times*, são exemplos de sucesso.

Além da concentração de jornalismo de dados nas cidades do Rio e de São Paulo, o mapa mostra que veículos de imprensa locais que trabalham com jornalismo de dados ainda são poucos ou não tem presença destacada na rede mundial de computadores. Isso reflete que menos pessoas serão expostas e alfabetizadas no uso de dados.

Esse projeto também mostrou que não são apenas organizações de mídias que estão produzindo conteúdo jornalístico com dados. Com a crise, jornalistas estão sendo contratados para trabalhar em outras organizações e passam a fazer conteúdo para os

⁹ Sobre jornalismo como “cão de guarda” cf. (WAISBORD, 2000)

canais dessas empresas. Esses profissionais aproveitam do jornalismo de dados para criarem histórias que permitam envolver suas audiências. As organizações Sociedade Brasileira de Pediatria, Conselho Federal de Medicina e APCEF/SP são apenas algumas das organizações em que se pode identificar a presença de jornalistas trabalhando com dados.

Outro ponto a se destacar é a grande presença de mídias nativas digitais na produção de conteúdo baseado em dados. Apesar da prática existir em mídias tradicionais, a presença dos nativos digitais é mais massiva. Além disso, existem organizações de mídia que se apresentam como focadas apenas em dados, como Volt Lab, Gênero e Número, Fiquem Sabendo, entre outras.

Vale ressaltar, que esse projeto possui limitações que não permite analisar o nível de maturidade de cada mídia na produção de jornalismo de dados. Além dessas limitações já mencionadas, esse projeto não apresenta um *widget* para busca por macrorregiões, como bairros e ruas. Infelizmente nem todas as organizações puderam ser contatadas ou possuíam informações que garantiam a precisão do endereço, por isso uma busca por macrorregião não seria tão fidedigna. Outro ponto, que foi testado com a audiência e não foi possível melhorar devido as limitações da plataforma Carto e de seu uso em aplicações móveis, foi as janelas de informações sobre as organizações. Torná-las mais colorida ou atrativa para o usuário aumentaria o espaço requisitado da tela, dificultando a utilização do mapa por um usuário de celular.

Com base em todos os pontos mencionados acima, pode-se ver que o jornalismo de dados tem grande potencial de exploração no país e que precisa de atores espalhados pelas diversas capitais. Só com uma presença massiva no país, o jornalismo de dados pode ser apreciado e, assim criar de um único nome e imagem para a audiência e para as redações brasileiras. Assim, criaremos uma consciência sobre a importância de estabelecer um significado do jornalismo de dados como produto de qualidade para atrair e reter clientes leais.

6. Referências Bibliográficas

APPELGREN, E.; NYGREN, G. Data Journalism in Sweden. **Digital Journalism**, v. 2, n. 3, p. 394–405, 3 jul. 2014.

ARAÚJO, L. V. A web e o jornalismo de dados: mapeamento de conceitos chave. **Dispositiva**, v. 5, n. 1, p. 144–163, 2018.

ASSIS, C. Morre Cláudio Weber Abramo, defensor da transparência e pioneiro do jornalismo de dados no Brasil. Disponível em: <<https://knightcenter.utexas.edu/pt-br/blog/00-20034-morre-claudio-weber-abramo-defensor-da-transparencia-e-pioneiro-do-jornalismo-de-dados>>. Acesso em: 4 de maio de 2019.

BORGES-REY, E. Unravelling Data Journalism. **Journalism Practice**, v. 10, n. 7, p. 833–843, 2 out. 2016.

BOUCHART, M. Holding the powerful accountable, using data. Disponível em: <<https://medium.com/data-journalism-awards/holding-the-powerful-accountable-using-data-f0bd5ebb24e3>>. Acesso em: 7 maio. 2019.

BRADSHAW, P. The inverted pyramid of data journalism. p. 1–7, 2007.

BRADSHAW, P. **Stories and Streams - Teaching collaborative journalism with peer to peer learning**. [s.l.: s.n.].

BRADSHAW, P. Data Journalism. In: **Ethics for Digital Journalists**. [s.l.] Routledge, 2014. p. 214–232.

BURGESS, M. Panama Papers: How the 11.5 million documents were analysed | WIRED UK. Disponível em: <<https://www.wired.co.uk/article/panama-papers-data-leak-how-analysed-amount>>. Acesso em: 19 mar. 2019.

CAIRO, A. **The Functional Art**. 1. ed. Berkeley: New Riders Publishing, 2016.

CARSON, A.; FARHALL, K. Understanding Collaborative Investigative Journalism in a “Post-Truth” Age. **Journalism Studies**, v. 19, n. 13, p. 1899–1911, 2018.

CODDINGTON, M. Clarifying Journalism's Quantitative Turn. **Digital Journalism**, v. 3, n. 3, p. 331–348, 4 maio 2015.

COX, M. The Development of Computer-Assisted Reporting. **Newspaper Division, Association for Education in Journalism and Mass Communication, Southeast Colloquium**, v. 2030, n. 305, p. 22, 2000.

DE MAEYER, J. et al. Waiting for Data Journalism. **Digital Journalism**, v. 3, n. 3, p. 432–446, 4 maio 2015.

DIFONZO, N. **The Watercooler Effect: An Indispensable Guide to Understanding and Harnessing the Power of Rumors**. [s.l.] Penguin Publishing Group, 2009.

FERRERAS RODRÍGUEZ, E. M. El Periodismo de Datos en España. **Estudios sobre el Mensaje Periodístico**, v. 22, n. 1, p. 255–272, 28 jun. 2016.

FINK, K.; ANDERSON, C. W. Data Journalism in the United States. **Journalism Studies**, v. 16, n. 4, p. 467–481, 4 jul. 2015.

FLEW, T. et al. The promise of computational journalism. **Journalism Practice**, v. 6, n. 2, p. 157–171, abr. 2012.

GRAY, J.; BOUNEGRU, L.; CHAMBERS, L. **The Data Journalism Handbook**. 1st. ed. Sebastopol: O'Reilly Media, 2012.

HAMMOND, P. From computer-assisted to data-driven: Journalism and Big Data. **Journalism**, v. 18, n. 4, p. 408–424, 2017.

HANNAFORD, L. Computational journalism in the UK newsroom. **Journalism Education**, v. 4, n. 1, p. 6–21, 2015.

HERMIDA, A.; YOUNG, M. L. Finding the Data Unicorn. **Digital Journalism**, v. 5, n. 2, p. 159–176, 7 fev. 2017.

KAKUTANI, M. **A morte da verdade: Notas sobre a mentira na era Trump**. [s.l.] Editora Intrínseca, 2018.

KARLSEN, J.; STAVELIN, E. Computational Journalism in Norwegian Newsrooms. **Journalism Practice**, v. 8, n. 1, p. 34–48, 2 jan. 2014.

LEWIS, S. C.; WESTLUND, O. Big Data and Journalism. **Digital Journalism**, v. 3, n. 3, p. 447–466, 4 maio 2015.

LISTER, M. et al. **New Media: a Critical Introduction**. [s.l.: s.n.].

MANCINI, L.; VASCONCELLOS, F. Jornalismo de Dados: conceito e categorias. **Revista Fronteiras - estudos midiáticos**, v. 18, n. 1, p. 69–82, 2016.

MEYER, P. **Precision Journalism : a Reporter’s Introduction to Social Science Methods**. [s.l.] Rowman & Littlefield Publishers, 2002.

MIELNICZUK, L.; TRÄSEL, M. Data-driven journalism as professional innovation and its challenges for education. **Contemporanea Comunicação e Cultura**, v. 15, n. 2, p. 609–629, 2017.

NAHSER, F. Behind the scenes of a murder investigation that mobilised 230 journalists all over Brazil. Disponível em: <<https://medium.com/data-journalism-awards/behind-the-scenes-of-a-murder-investigation-that-mobilised-230-journalists-all-over-brazil-e5136e2ea1ea>>. Acesso em: 7 maio. 2019.

OBERMAYER, B.; OBERMAIER, F. **The Panama papers : breaking the story of how the rich & powerful hide their money**. 1. ed. London: Oneworld Publications, 2016.

PARASIE, S.; DAGIRAL, E. Data-driven journalism and the public good: “Computer-assisted-reporters” and “programmer-journalists” in Chicago. **New Media & Society**, v. 15, n. 6, p. 853–871, 18 set. 2013.

PEISER, J. New York Times Co. Reports Revenue Growth as Digital Subscriptions Rise - The New York Times. Disponível em: <<https://www.nytimes.com/2018/05/03/business/media/new-york-times-earnings.html>>. Acesso em: 12 abr. 2019.

ROCHA, C. Como Claudio Weber Abramo trabalhou pela transparência. **Nexo**, v. 2018, p. 19–22, 2018.

RODRIGUES, Kelly de Conti. Jornalismo de dados na web: estudo da produção de sentido na infografia do blog do Estadão Dados e do La Nación Data Blog. *In: 10º Interprogramas de Mestrado em Comunicação da Faculdade Cásper Líbero*. [s.l.]: Faculdade Cásper Líbero, 2014.

ROGERS, S. Florence Nightingale, datajournalist: information has always been beautiful. **The Guardian**, 13 ago. 2010.

ROYAL, C.; BLASINGAME, D. Data journalism: An explication. **International Symposium in Online Journalism**, v. 5, n. 1, p. 24–46, 2015.

SCHACHT, K. The history of data driven journalism –. Disponível em: <<https://datentaeter.de/the-history-of-data-driven-journalism/>>. Acesso em: 1 maio. 2019.

SPARROW, A. Michael Martin: what are the allegations? **The Guardian**, 2008.

SWENEY, M. Guardian on track to break even as company halves its losses | Media | The Guardian. Disponível em: <<https://www.theguardian.com/media/2018/apr/25/guardian-on-track-to-break-even-as-company-halves-its-losses>>. Acesso em: 12 abr. 2019.

TABARY, C.; PROVOST, A. M.; TROTTIER, A. Data journalism's actors, practices and skills: A case study from Quebec. **Journalism**, v. 17, n. 1, p. 66–84, 2016.

TERUEL RODRÍGUEZ, L.; BLANCO CASTILLA, E. El periodismo de datos en los medios de referencia españoles. **Trípodos**, v. 38, p. 13–30, 2016.

TRÄSEL, M. JORNALISMO GUIADO POR DADOS: relações da cultura hacker com a cultura jornalística. **Anais do XXII Encontro Anual da Compós, Universidade Federal da Bahia**, p. 1–16, 2013.

TRÄSEL, M. Jornalismo guiado por dados: aproximações entre a identidade jornalística e a cultura hacker. **Estudos em Jornalismo e Mídia**, v. 11, n. 1, p. 291, 2014.

TRÄSEL, M. Hacks and hackers: the ethos and beliefs of a group of Data-Driven Journalism professionals in Brazil. **Repórteres e hackers: o ethos e as crenças de um**

grupo de profissionais de Jornalismo Guiado por Dados no Brasil, v. 25, n. 1, p. 1–14, 2018.

USKALI, T. I.; KUUTTI, H. Models and Streams of Data Journalism. **The Journal of Media Innovations**, v. 2, n. 1, p. 77, 2015.

VASCONCELLOS, F. et al. The five Data Journalism categories, or how to understand Data Journalism in Brazil. p. 1–24, 2015.

VENTURA, M. P. Jornalismo de Dados como diferencial: o caso do Nexo. **Pauta Geral - Estudos em Jornalismo**, v. 5, n. 2, p. 240–254, 2018.

WAISBORD, S. **Watchdog Journalism in South America**. 1. ed. [s.l.] Columbia University Press, 2000.

YOUNG, M. L.; HERMIDA, A. From Mr. and Mrs. Outlier To Central Tendencies. **Digital Journalism**, v. 3, n. 3, p. 381–397, 2015.

YOUNG, M. L.; HERMIDA, A.; FULDA, J. What Makes for Great Data Journalism?: A content analysis of data journalism awards finalists 2012–2015. **Journalism Practice**, v. 12, n. 1, p. 115–135, 2018.