



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

**ESPIÃO ESTATÍSTICO: O JORNALISMO DE DADOS NO
GLOBOESPORTE.COM**

ROBERTO VIEIRA DE SOUSA MALESON

RIO DE JANEIRO

2017

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

**ESPIÃO ESTATÍSTICO: O JORNALISMO DE DADOS NO
GLOBOESPORTE.COM**

Monografia submetida à Banca de Graduação
como requisito para obtenção do diploma de
Comunicação Social/ Jornalismo.

ROBERTO VIEIRA DE SOUSA MALESON

Orientador: Professor Paulo César Castro

RIO DE JANEIRO
2017

UNIVERSIDADE FEDERAL DO RIO DE JANEIRO
ESCOLA DE COMUNICAÇÃO

TERMO DE APROVAÇÃO

A Comissão Examinadora, abaixo assinada, avalia a Monografia **Espião Estatístico: o Jornalismo de Dados no GloboEsporte.com**, elaborada por Roberto Vieira de Sousa Maleson.

Monografia examinada:

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

Comissão Examinadora:

Orientador: Professor Paulo César Castro

Doutor em Comunicação e Cultura pela Escola de Comunicação - UFRJ

Departamento de Expressão e Linguagem - UFRJ

Profa. Cristiane Costa

Doutora em Comunicação e Cultura pela Escola de Comunicação - UFRJ

Departamento de Expressão e Linguagem - UFRJ

Prof. Fernando Ewerton

Doutor em Ciência da Informação pela Escola de Comunicação - UFRJ

Departamento de Expressão e Linguagem - UFRJ

RIO DE JANEIRO

2017

FICHA CATALOGRÁFICA

MALESON, Roberto Vieira de Sousa.

Espião Estatístico: o Jornalismo de dados no GloboEsporte.com.
Rio de Janeiro, 2017.

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

MALESON, Roberto Vieira de Sousa. **Espião Estatístico: o Jornalismo de Dados no GloboEsporte.com**. Orientador: Paulo César Castro. Rio de Janeiro: UFRJ/ECO. Monografia em Jornalismo.

RESUMO

O presente trabalho procura apresentar uma área em crescimento exponencial no jornalismo mundial e brasileiro: o jornalismo de dados. A partir de fundamentação teórica e prática, demonstra-se o significado desta área dentro do meio, quais são suas características e quais são as variáveis envolvidas. É feito um breve percurso histórico para mostrar como o jornalismo de dados surgiu e como se desenvolveu ao longo dos anos, auxiliado pela evolução dos meios tecnológicos e das mudanças nas narrativas jornalísticas. Além disso, procura-se relatar quais são as habilidades e técnicas necessárias para a especialização em jornalismo de dados. Por fim, a pesquisa faz um estudo de caso sobre um produto do GloboEsporte.com, o Espião Estatístico. O objetivo desta parte é analisar como são trabalhados os dados esportivos sobre futebol neste veículo e demonstrar alguns dos conteúdos que são entregues ao público consumidor de estatísticas esportivas pela equipe responsável por gerir o projeto.

SUMÁRIO

1. INTRODUÇÃO

2. JORNALISMO DE DADOS

2.1. DO JORNALISMO DE PRECISÃO AO JORNALISMO DE DADOS

2.1.1. JORNALISMO DE PRECISÃO

2.1.1.1. A TEORIA COLOCADA EM PRÁTICA

2.1.2. REPORTAGEM ASSISTIDA POR COMPUTADOR

2.2. A PIRÂMIDE INVERTIDA DO JORNALISMO DE DADOS

2.3. MOVIMENTO PELA ABERTURA DE DADOS

2.3.1. A LEI DE ACESSO À INFORMAÇÃO NO BRASIL

3. O PERFIL DO JORNALISTA DE DADOS

3.1. HABILIDADES E TÉCNICAS NECESSÁRIAS

3.1.1. COMPILAR

3.1.1.1. BUSCA AVANÇADA DO GOOGLE

3.1.1.2. BUSCA AVANÇADA PELAS REDES SOCIAIS

3.1.1.3. RASPAGEM DE DADOS E NOÇÕES DE PROGRAMAÇÃO

3.1.1.3.1. RASPANDO DADOS COM O GOOGLE SHEETS

3.1.1.3.2. RASPANDO DADOS COM O IFTTT

3.1.1.3.3. RASPANDO DADOS COM O WEBSCRAPER

3.1.1.3.4. POR QUE É BOM SABER PROGRAMAR?

3.1.1.4. CONVERSÃO DE DADOS FECHADOS PARA ABERTOS

3.1.2. LIMPAR

3.1.3. CONTEXTUALIZAR

3.1.4. COMBINAR

3.1.5. COMUNICAR

3.1.6. A MATEMÁTICA NO MANEJO COM NÚMEROS

3.2. O JORNALISMO DE DADOS NO ESPORTE

3.2.1. PANORAMA GERAL NO MUNDO

3.2.2. PANORAMA NO BRASIL

4. ESPIÃO ESTATÍSTICO

4.1. COMO FUNCIONA O PREENCHIMENTO DAS PLANILHAS

4.1.1. OS 13 CABEÇALHOS INICIAIS

4.1.2. ESCALAÇÕES

4.1.3. TRIO DE ARBITRAGEM

4.1.4. OS 13 PASSOS ANTES DO INÍCIO DOS LANCES DO JOGO

4.1.5. OS LANCES DE UMA PARTIDA

4.1.5.1. IMPEDIMENTOS

4.1.5.2. SUBSTITUIÇÕES

4.1.5.3. CARTÕES E PÊNALTIS COMETIDOS

4.1.5.4. FINALIZAÇÕES, FALTAS DIRETAS E PÊNALTIS COBRADOS

4.1.5.5. TEMPO EM CAMPO

4.2. CONTEÚDOS DO ESPIÃO ESTATÍSTICO

4.2.1. ARTILHEIRO DA HORA

4.2.2. JOGADORES DECISIVOS

4.3. JORNALISMO DE DADOS NA TV

4.3.1. SUGESTÃO DE PAUTA

4.3.2 APROVAÇÃO DE PAUTA E PARTICIPAÇÃO NO PROGRAMA

5. CONCLUSÃO

6. REFERÊNCIAS BIBLIOGRÁFICAS

7. APÊNDICE

7.1. FIGURAS

7.2. ENTREVISTA COM SERGIO SPAGNUOLO CONCEDIDA AO AUTOR

1. INTRODUÇÃO

A forma de contar uma história sempre foi importante, ainda mais no jornalismo. Os jornalistas precisam conquistar a atenção do leitor para que ele capte bem a informação transmitida nos conteúdos publicados pela mídia. E percebe-se que, nos últimos anos, as narrativas jornalísticas foram lapidadas em uma técnica que tem ganhado força dentro das redações dos principais veículos: o jornalismo de dados. Para abordar o tema, essa monografia foi organizada em seis capítulos.

O segundo capítulo deste trabalho revela ao leitor justamente como foi esse processo de mudança nas narrativas jornalísticas quando o assunto era trabalhar com qualquer tipo de dado. Os números sempre estiveram presentes nas publicações dos jornais, mas com o avanço da tecnologia e das ferramentas de tratamento, análise e comunicação dos dados, houve uma melhora significativa na apresentação dos mesmos ao público. A tecnologia permitiu análises mais complexas de bases de dados infinitamente maiores e, conseqüentemente, reportagens mais poderosas baseadas em dados. Geralmente, os dados tinham um papel secundário nas narrativas jornalísticas. Mas, esse tipo de jornalismo mudou o paradigma e trouxe protagonismo aos mesmos.

Nesta parte, é apresentado o que é o jornalismo de dados e os cuidados que o repórter deve ter na hora de lidar com números e informações tão precisas quanto as trabalhadas neste meio. Além disso, é feito um breve apanhado histórico de como surgiu a área dentro do jornalismo. A origem disso está na aplicação de técnicas das ciências sociais ao jornalismo pelo americano Phillip Meyer, no final dos anos 1960. Ele foi responsável por quebrar o paradigma da passividade jornalística, principalmente quanto à crença que o jornalista deveria depender sempre das fontes, seja através dos dados oriundos delas ou mesmo das suas declarações. Desse modo, o repórter teria a função apenas de relator dos fatos acontecidos na sociedade. Com o trabalho de Meyer, isso foi deixado de lado e o repórter deixou de ser mero coadjuvante e passou também a atuar para produzir e analisar dados capazes de revelar uma determinada realidade social.

Uma das intenções do trabalho, ainda neste segundo capítulo, é exibir como foi a transição dos anos até que o Jornalismo de Dados ganhasse nome e força, mostrando o que veio antes, desde o Jornalismo de Precisão até a Reportagem Assistida por Computador. Além disso, ao debate é trazido o diagrama apresentado pelo jornalista Paul Bradshaw com as cinco etapas do processo de uma reportagem guiada por dados: compilação, limpeza, contextualização, análise e comunicação. Cada uma delas é explicada com o objetivo de revelar ao leitor o que está por trás das grandes reportagens

do meio. Outro viés abordado neste capítulo trata sobre a importância da abertura dos dados para o jornalismo de dados e que está intimamente ligado às Leis de Acesso à Informação, que são conquistas da militância dos últimos anos em prol da transparência das informações e contra a cultura do sigilo, ainda predominante em muitos países.

O terceiro capítulo tem o objetivo de mostrar qual é o perfil do profissional especialista em dados e quais são as técnicas e habilidades necessárias para o exercício do jornalismo de dados. Nesta parte, é mostrada a importância de saber como utilizar algumas ferramentas de captura e análise de dados e como a programação e a raspagem de dados podem auxiliar a aumentar a produtividade durante uma investigação jornalística. Além disso, ter domínios sobre softwares de planilhas e de banco de dados ou lidar bem com a matemática são habilidades fundamentais para o trabalho no jornalismo de dados. Para a formação destes profissionais, há algumas organizações que oferecem treinamento especializado – presencial ou online, tais como Associação Brasileira de Jornalismo Investigativo (Abraji)¹, a Escola de Dados² e o Centro Knight para o Jornalismo nas Américas³. Estas instituições organizam cursos e workshops direcionados aos jornalistas interessados em aprender ou aprimorar habilidades designadas ao exercício deste tipo de jornalismo. Natália Mazotte (2016) defende que, além da ampliação da oferta de capacitações, é necessário incluir a disciplina Jornalismo de Dados no currículo das faculdades de jornalismo.

Grande parte dos estudantes brasileiros se forma sem o domínio de habilidades do universo digital que hoje em dia são essenciais em todas as etapas da produção jornalística. Para o próximo ano, uma das frentes de trabalho de organizações que se preocupam com o futuro do jornalismo deve ser incentivar essa atualização curricular, integrando novas tecnologias e campos de conhecimento emergentes (MAZOTTE, 2016)⁴.

Ainda nesta parte do trabalho, é revelado o panorama geral das produções de dados na área esportiva. Percebe-se que tanto no contexto internacional quanto no nacional as reportagens guiadas por dados esportivos estão em crescimento, mas ainda são muito baixas se comparadas a outros temas como economia e política, os dois carros-chefes das grandes produções do jornalismo de dados até aqui. Notou-se que faltam equipes de jornalistas de dados dedicados exclusivamente ao esporte. No Brasil,

¹ Disponível em: <http://www.abraji.org.br/>

² Disponível em: <https://escoladedados.org/>

³ Disponível em: <https://knightcenter.utexas.edu/pt-br/>

⁴ Disponível em: <https://jornalismonobrasilem2017.com/as-novas-fronteiras-do-jornalismo-de-dados-no-brasil-cddeede4fb8d>. Acessado em: 11/06/2017.

só identificamos uma equipe desta característica entre os principais veículos de mídia no país: a equipe de dados esportivos do GloboEsporte.com, responsável por gerir o produto Espião Estatístico – objeto de estudo desta monografia.

No capítulo seguinte é feita uma exposição e análise do Espião Estatístico, produto do site GloboEsporte.com. A equipe do Espião trabalha com dados esportivos sobre o futebol, esporte de maior demanda nacional. Neste capítulo final são discutidas a forma de compilação e estruturação dos dados e quais são os tipos de conteúdos possíveis de serem extraídos a partir da base de dados compilada. Por fim, o autor expõe como é o trabalho para expor os resultados deste jornalismo de dados na TV a partir do quadro semanal do Espião Estatístico no programa Troca de Passes⁵, do SporTV⁶.

⁵ Disponível em: <http://sportv.globo.com/site/programas/troca-de-passes/>

⁶ É um canal esportivo por assinatura brasileiro da Globosat.

2. JORNALISMO DE DADOS

O jornalismo de dados é atualmente uma das áreas mais promissoras e com potencial de crescimento e expansão dentro do jornalismo. Ao longo dos últimos anos, as redações passaram a compreender a importância dos dados na produção do conteúdo jornalístico e de como estes podem dar um ângulo diferente ao tradicional e “velho” jornalismo. Porém, isto não quer dizer que os dados só passaram a ser utilizados neste momento atual, em plena segunda década do século XXI, mas sim que o seu melhor uso e aproveitamento vêm desde os anos 1960 até hoje, com crescimento gradual ao longo do tempo. Para Stray (2016), a vida dos dados tem três partes: quantificação, análise e comunicação. “A quantificação é o processo que cria dados. A análise envolve reorganizar os dados ou combiná-los com outras informações para produzir novos conhecimentos. E nada disso é útil sem comunicar o resultado” (STRAY, 2016)⁷.

Se antes o manejo dos dados era simples e pitoresco, com pouca exploração e pouco aprofundamento nas produções midiáticas, hoje eles já assumem papel principal em diversos e relevantes trabalhos do jornalismo. Os jornalistas foram adquirindo, ao longo do tempo, a habilidade de saber como lidar com os dados e tirar deles um conteúdo interessante. Isso é possível graças à interpretação dos dados, conforme descreve Stray (2016):

Toda análise de dados é realmente interpretação de dados, o que requer muito mais do que matemática. Os dados precisam de contexto para significar qualquer coisa: Imagine se alguém lhe deu uma planilha sem colunas nomeadas. Cada conjunto de dados pode ser a fonte de muitas histórias diferentes, e não há nenhuma teoria objetiva que nos diz quais histórias verdadeiras são as melhores. Mas as histórias ainda têm de ser verdade, que é onde o jornalismo de dados se baseia em princípios estatísticos estabelecidos (STRAY, 2016)⁸.

Antes, os dados tinham a função maior de apenas complementar uma reportagem, dando credibilidade ao exposto na matéria a partir da exibição de números que comprovassem o que havia sido escrito em determinada notícia. Hoje, uma reportagem inteira pode ser orquestrada a partir de um conjunto de dados. É comum

⁷ Disponível em: https://www.cjr.org/tow_center_reports/the_curious_journalists_guide_to_data.php. Acessado em: 24/04/2017. Tradução do autor. No original: “Quantification is the process that creates data. Analysis involves rearranging the data or combining it with other information to produce new knowledge. And none of this is useful without communicating the result”.

⁸ Disponível em: https://www.cjr.org/tow_center_reports/the_curious_journalists_guide_to_data.php. Acessado em: 24/04/2017. Tradução do autor. No original: “All data analysis is really data interpretation, which requires much more than math. Data needs context to mean anything at all: Imagine if someone gave you a spreadsheet with no column names. Each data set could be the source of many different stories, and there is no objective theory that tells us which true stories are the best. But the stories still have to be true, which is where data journalism relies on established statistical principles.”

observar jornalistas entrevistando dados⁹ e tirando deles informação de profundo interesse aos cidadãos. A partir de uma base de dados¹⁰ estruturada e organizada pode surgir pautas para diversos tipos de reportagem.

Segundo Bradshaw (apud GRAY et al., 2013), dados e jornalismo são termos problemáticos. O autor avalia que as pessoas pensam os dados apenas como números reunidos numa planilha, mas faz ressalvas:

Há 20 anos, este era praticamente o único tipo de dado com o qual os jornalistas lidavam. Mas nós vivemos num mundo digital agora, um mundo em que quase tudo pode ser (e quase tudo é) descrito com números. A sua carreira, 300 mil documentos confidenciais, todos dentro do seu círculo de amizades; tudo isso pode ser (e é) descrito com apenas dois números: zeros e uns. Fotos, vídeos e áudios são todos descritos com os mesmos dois números: zeros e uns. Assassinatos, doenças, votos, corrupção e mentiras: zeros e uns (apud GRAY et al., 2013: 8).

Ou seja, qualquer coisa que possa ser quantificada e classificada pode vir a ser um dado útil ao trabalho jornalístico. Porém, assim como qualquer fonte, os dados devem ser tratados com ceticismo e, por isso, os jornalistas não podem confiar cegamente neles. Ainda mais por, em sua maioria, se tratarem de números, é necessária uma apuração precisa, uma checagem atenta e uma recheagem ainda mais minuciosa. Um número a mais ou a menos pode fazer muita diferença na história contada em uma reportagem guiada por dados. Além disso, como qualquer ferramenta, os jornalistas devem estar conscientes sobre como os dados podem moldar e restringir as reportagens criadas por meio deles.

Gray et al. (2013) consideram que, um fator diferencial do jornalismo de dados em relação ao restante do jornalismo são as novas possibilidades que se abrem quando se combina o conhecido “faro jornalístico” com a habilidade de contar uma história envolvente com a escala e o alcance absolutos da informação digital agora disponível. O jornalismo de dados aparece como um meio para condensar esse emaranhado de

⁹Entrevistar dados significa descobrir pautas relevantes para uma matéria. Assim como uma entrevista com uma pessoa, que pode passar informações ao jornalista, os dados são capazes de revelar ângulos interessantes para uma pauta jornalística.

¹⁰De acordo com Guimarães, uma base de dados, ou banco de dados, é uma coleção de dados ou informações relacionadas entre si, que representam aspectos do mundo real com significado próprio e que desejamos armazenar para o futuro (GUIMARÃES apud JANSEN, 2011: 8). Sua projeção, construção e estruturação estão submetidas ao uso e conjunto de aplicações que se deseja ter sobre ela e que podem ser simples ou complexas, de tamanhos variados. Atualmente, podem conter dados tanto informações alfanuméricas, quanto imagens, gráficos, vídeos, sons e até outras bases de dados. Bases de dados podem ser simples e pequenas, como a relação de nomes e telefones das pessoas conhecidas por um indivíduo, ou dos bens e valores de uma pessoa física; ou tão grande como o catálogo de todos os livros publicados nos Estados Unidos, ou uma base de dados de todas as fotos recolhidas ao longo dos anos pelo sistema espacial americano (JANSEN, 2011: 8).

informações em uma área confiável de informação. Ele é o instrumento para sublinhar e evidenciar ao leitor uma valiosa informação contida em um terabyte¹¹ de dados, por exemplo. Os jornalistas de dados podem ajudar a reunir, sintetizar e apresentar fontes de informação diversas e muitas vezes difíceis, de modo a fornecer percepções reais sobre questões complexas para a audiência. Ao invés de apenas reciclar press releases e reescrever matérias vistas ou ouvidas anteriormente em outros veículos de comunicação, os jornalistas de dados podem fornecer aos leitores uma perspectiva clara, compreensível, com análises contundentes, dados interativos e gráficos envolventes.

Quanto à questão da apresentação da informação em gráficos interativos e envolventes, Stray (2016) diz que a visualização de dados é muito importante porque se baseia na condição biológica da percepção visual humana, o que facilita a compreensão da informação estatística pelo leitor.

A maioria das pessoas não está acostumada a interpretar dados, e é difícil culpá-los. A visualização de dados pode ser útil porque transfere parte do trabalho cognitivo de compreensão de dados para o poderoso sistema visual humano. Ainda assim, os conceitos fundamentais do trabalho de dados são sutis e às vezes não naturais. As nuances da amostragem, probabilidades, causalidade e assim por diante são estranhas à experiência cotidiana. Mais do que isso, os números não são um meio particularmente empático. Para a maioria das pessoas, mesmo a estatística mais gritante é desconectada da experiência cotidiana. Os jornalistas podem superar isso usando exemplos, metáforas ou histórias para relacionar os números às pessoas. O jornalismo é uma tarefa profundamente humana, não importa os métodos (STRAY, 2016)¹².

2.1. DO JORNALISMO DE PRECISÃO AO JORNALISMO DE DADOS

O jornalismo de dados é uma área que passou por mudanças ao longo dos anos, conforme os avanços da tecnologia e das ferramentas disponíveis para o trabalho jornalístico. Para chegar ao jornalismo de dados conhecido hoje, a área teve suas origens no jornalismo de precisão, doutrinado pelo jornalista americano Phillip Meyer, e

¹¹O terabyte é uma unidade de medida múltipla do byte: corresponde a mil gigabytes ou a um trilhão de bytes.

¹²Disponível em: https://www.cjr.org/tow_center_reports/the_curious_journalists_guide_to_data.php. Acessado em: 24/04/2017. Tradução do autor. No original: “Most people are not used to interpreting data, and it’s hard to blame them. Data visualization can be helpful because it transfers some of the cognitive work of understanding data to the enormously powerful human visual system. Still, the foundational concepts of data work are subtle and at times unnatural. The nuances of sampling, probabilities, causality, and so on are foreign to everyday experience. More than that, numbers are not a particularly empathetic medium. For most people even the most screaming statistic is disconnected from everyday experience. Journalists can overcome this using examples, metaphors, or stories to relate the numbers to people. Journalism is a deeply human task, no matter the methods.”

também pela Reportagem Assistida por Computador (RAC), na qual os computadores foram inseridos ao cotidiano do trabalho dos jornalistas.

2.1.1. JORNALISMO DE PRECISÃO

O jornalismo de dados nasceu, ou teve suas origens, no final dos anos 1960, com o jornalista americano Philip Meyer, quando ele escreveu a obra *Precision journalism: a reporter's introduction to social science methods*. Meyer (1991) diz que, já naquela época, o crescimento exponencial de informação disponível fez com que os jornalistas precisassem se adaptar. Eles passaram a ter a função de organizadores e intérpretes para coletar, filtrar e transmitir os fatos a partir das análises do imenso volume de informações disponíveis. Além disso, era necessário pensar a maneira de colocar estas informações na cabeça dos leitores. Ou seja, como transmitir a informação de forma clara para que a compreensão da mensagem fosse bem sucedida. “Em resumo, um jornalista deve ser um administrador de base de dados, um processador de dados e um analista de dados” (MEYER, 1991)¹³.

O principal a entender é que Meyer introduziu e aplicou métodos de pesquisa das ciências sociais e do comportamento à prática do jornalismo, tais como a investigação qualitativa (análise de casos, grupos focais, análise de discurso, dentre outros) e a investigação quantitativa. Esta última inclui a investigação descritiva (quando se consideram todos os casos, como, por exemplo, em um censo populacional) e o que se chama de investigação inferencial, ou seja, quando se trabalha com um universo de informações, e dada a dificuldade de abordar cada parte dele, se projeta uma mostra representativa. É o caso das pesquisas de opinião ou eleitorais. Assim, Meyer define o jornalismo de precisão como um jornalismo científico.

(Na França, o termo “jornalismo de precisão” tem sido traduzido como “le journalisme scientifique”). Isso significa tratar o jornalismo como se ele fosse uma ciência, adotando método científico, objetividade científica e ideais científicos em todo o processo de comunicação de massa. Se isso soa absurdamente pretensioso, lembre-se que a própria ciência é contida quanto aos seus alcances, suas possibilidades e possui suas próprias sanções contra a pretensão (MEYER, 1991).¹⁴

¹³ Disponível em:

https://www.attorneygeneral.jus.gov.on.ca/inquiries/cornwall/en/hearings/exhibits/Mary_Lynn_Young/pdf/07_Meyer.pdf. Acessado em: 14/04/2017. Tradução do autor. No original: “In short, a journalist has to be a database manager, a data processor, and a data analyst.”

¹⁴ Disponível em:

https://www.attorneygeneral.jus.gov.on.ca/inquiries/cornwall/en/hearings/exhibits/Mary_Lynn_Young/pdf/07_Meyer.pdf. Acessado em: 14/04/2017. Tradução do autor. No original: “(In France, the term

A aplicação de métodos de pesquisa científicos à prática do jornalismo começou a ameaçar o que Phillip Meyer chamou de tradições gêmeas da passividade e da inocência jornalísticas.

A antiga tradição prega que a mídia deve reportar as notícias, não fazer as notícias. O envolvimento da mídia em pesquisas de opinião pública tem sido criticado com base no julgamento de que a mídia não deve fazer as pesquisas, mas esperar passivamente até que outras pessoas as façam para então noticiá-las (MEYER, 1991).¹⁵

Logo, percebe-se que o autor era contra essa antiga tradição. Com o aumento dos meios tecnológicos e aprimoramento das técnicas e ferramentas de compilação e análise de dados, o jornalista passou a ter um papel ativo de buscar essa informação para não mais esperar inocentemente por um conjunto de dados. Até porque a passividade e a inocência, defende Meyer, são fatores que atrapalham o exercício do jornalismo objetivo, uma vez que outras pessoas podem fazer pesquisas tendenciosas já sabendo que o repórter vai apenas noticiá-las, sem ao menos questioná-las.

Para evitar a manipulação, a mídia precisa de mais autoconfiança, e o melhor caminho para a autoconfiança é através do conhecimento. As pesquisas feitas pela mídia proliferaram na década de 1980 precisamente porque os editores não confiavam mais nas pesquisas que os políticos ofereciam e se armaram com suas próprias operações de coleta de dados como forma de autodefesa. Assim, as pesquisas se tornaram não tanto uma forma de produzir notícias quanto uma ferramenta que alavanca o próprio processo de coleta de notícias - desde que, é claro, os jornalistas estejam a cargo de todo o processo, desde a concepção até a montagem da pesquisa, análise e interpretação. O jornalista de precisão não contrata um pesquisador para criar o evento noticioso; o jornalista vira o pesquisador (MEYER, 1991).¹⁶

“precision journalism” has been translated as ‘le journalism scientifique.’) It means treating journalism as if it were a science, adopting scientific method, scientific objectivity, and scientific ideals to the entire process of mass communication. If that sounds absurdly pretentious, remember that science itself is restrained about its achievements and its possibilities and has its own sanctions against pretension.”

¹⁵Disponível em:

https://www.attorneygeneral.jus.gov.on.ca/inquiries/cornwall/en/hearings/exhibits/Mary_Lynn_Young/pdf/07_Meyer.pdf. Acessado em: 14/04/2017. Tradução do autor. No original: “The former tradition holds that media should report news, not make news. Media involvement in public opinion polling has been criticized on the ground that the media should not do polls but should wait passively until other people do them and then report on them.”

¹⁶Disponível em:

https://www.attorneygeneral.jus.gov.on.ca/inquiries/cornwall/en/hearings/exhibits/Mary_Lynn_Young/pdf/07_Meyer.pdf. Acessado em: 14/04/2017. Tradução do autor. No original: “To defend against being manipulated, the media need more self-confidence, and the best route to self-confidence is through knowledge. Media polls proliferated in the 1980s precisely because the editors no longer trusted the polls that politicians tried to give them and armed themselves with their own data collection operations out of self-defense. Thus polling became not so much a way to make news as an enhanced tool of the newsgathering process itself—provided, of course, that journalists are in charge of the entire process, from

E para virar o pesquisador responsável por coletar, armazenar, filtrar e interpretar os números, o jornalista de precisão deve saber o que fazer com estes dados. “Dados brutos, por si, nunca são o suficiente. Para serem úteis, para serem compreendidos, eles devem ser processados, abstraídos, ajustados em algum tipo de estrutura.” (MEYER, 1991).¹⁷ E há quem defenda que os dados, em si só, não são suficientes para produzir uma notícia.

Lembre-se também que um banco de dados sozinho não é uma história. Em vez disso, é um campo de informação que precisa ser colhido cuidadosamente com discernimento e cautela. Ele precisa ser comparado e aumentado com observações e entrevistas (HOUSTON, 2015: 4).¹⁸

O jornalista tem, então, essa missão de transformar esse conjunto de informação em notícia, em algo palpável de ser percebido pelo público. Segundo Meyer (1991), há duas fases que devem ser consideradas e solucionadas: a fase da entrada, em que os dados são coletados, organizados e analisados, e a fase da saída, na qual os dados são preparados para serem comunicados ao leitor. O autor elabora seis passos sobre o que fazer com os dados:

1. Coletá-los. Quer você tenha ou não tentado imitar os cientistas em seus métodos de coleta de dados, você pode lucrar sabendo alguns de seus truques. Sempre vale a pena lembrar, o que me disse o professor H. Douglas Price em Harvard, na primavera de 1967, que “os dados não vêm da cegueira.”
2. Armazená-los. Os jornalistas dos velhos tempos armazenavam dados em pedaços de papel em suas mesas, nos cantos do escritório ou, se fossem muito organizados, em arquivos de recortes. Computadores fazem isso melhor.
3. Recuperá-los. As ferramentas do jornalismo de precisão podem ajudar a recuperar dados que você mesmo coletou e guardou, dados que alguém armazenou pensando em um usuário como você ou dados que alguém guardou por motivos completamente alheios ao seu interesse, talvez sem ter a mínima ideia de que um jornalista ou usuário público poderia algum dia estar interessado neles.
4. Analisá-los. A análise jornalística muitas vezes consiste em meramente examinar para encontrar e listar os desvios interessantes. Mas pode também envolver pesquisas por causação implícita, por

conception, through research design, to analysis and interpretation. The precision journalist does not hire a pollster to create a news event; the journalist becomes the pollster.”

¹⁷Disponível em:

https://www.attorneygeneral.jus.gov.on.ca/inquiries/cornwall/en/hearings/exhibits/Mary_Lynn_Young/pdf/07_Meyer.pdf. Acessado em: 14/04/2017. Tradução do autor. No original: “Raw data alone can never be enough. To be useful, to be understood, data have to be processed, abstracted, fit into some kind of structure.”

¹⁸Tradução do autor. No original: “Also remember that a database alone is not a story. Instead, it is a field of information that needs to be harvested carefully with insight and caution. It needs to be compared with and augmented with observations and interviews.”

padrões que possam sugerir que fenômenos diferentes variam juntos por motivos interessantes, ou até para avaliar a efetividade de políticas públicas.

5. Reduzi-los. A redução de dados se tornou uma habilidade tão importante no jornalismo quanto a coleta de dados. Uma boa matéria jornalística é definida pelo que deixa de fora, quase tanto quanto pelo que ela inclui.

6. Comunicá-los. Uma reportagem não lida ou não entendida é uma reportagem perdida. Você pode filosofar que, como o som de uma árvore caindo em uma floresta, ela não chega a existir (MEYER, 1991).¹⁹

2.1.1.1. A TEORIA COLOCADA EM PRÁTICA

Philip Meyer viu em um movimento de convulsão social a oportunidade para testar sua teoria do jornalismo de precisão. Quando começou uma série de protestos raciais nos anos 1960, primeiro na área de Watts, em Los Angeles, depois em Newark e Detroit, surgiram algumas teorias populares para explicar as causas destes levantes sociais. Uma delas, famosa entre os editorialistas, dizia que os manifestantes tendiam a ser as pessoas frustradas e sem esperança, de baixa renda econômica, e que se rebelavam porque não tinham outros meios de avanço ou de expressão. Phillip Meyer conseguiu refutar essa teoria ao fazer uma pesquisa do tipo *survey* (com tabulação de entrevistas padronizadas).

Se você puder identificar os rebeldes e compará-los com os não-rebeldes, você vai descobrir, se a teoria for verdadeira, que os menos educados são mais propensos a participar de rebeliões. O *Detroit Free Press* fez uma pesquisa desse tipo, em 1967, e descobriu que jovens universitários tinham tanta chance de participar dos protestos quanto aqueles que nem chegaram a terminar o segundo grau. A teoria, então, foi reprovada pelos dados (MEYER, 1991)²⁰.

¹⁹Disponível em:

https://www.attorneygeneral.jus.gov.on.ca/inquiries/cornwall/en/hearings/exhibits/Mary_Lynn_Young/pdf/07_Meyer.pdf. Acessado em: 14/04/2017. Tradução do autor. No original: “1. Collect it. Whether or not you ever try to emulate scientists in their data collection methods, you can profit from knowing some of their tricks. It is always worth remembering, as Professor H. Douglas Price told me at Harvard in the spring of 1967, that ‘data do not come from the stork.’ 2. Store it. Old-time journalists store data on stacks of paper on their desks, incorners of their offices, and, if they are really well organized, in clip-files. Computers are better. 3. Retrieve it. The tools of precision journalism can help you retrieve data that you collected and stored yourself, data that someone else stored with a user like you in mind, or data that someone else stored for reasons completely unrelated to your interest, perhaps with no earthly idea that a journalist or public user would ever be interested in it. 4. Analyze it. Journalistic analysis often consists of merely sorting to find and list the interesting deviances. But it can also involve searches for implied causation, for patterns that suggest that different phenomena vary together for interesting reasons. 5. Reduce it. Data reduction has become as important a skill in journalism as data collection. A good news story is defined by what it leaves out as well as what it includes. 6. Communicate it. A report unread or not understood is a report wasted. You can make a philosophical case that, like the sound of a tree falling in the forest, it does not exist at all.”

²⁰Disponível em:

https://www.attorneygeneral.jus.gov.on.ca/inquiries/cornwall/en/hearings/exhibits/Mary_Lynn_Young/pdf/07_Meyer.pdf. Acessado em: 14/04/2017. Tradução do autor. No original: “If you can identify the

Outra teoria popular da época era que a causa básica das rebeliões era a dificuldade que os negros vindos do sul dos Estados Unidos tinham em ser assimilados na cultura do norte do país. Forçados pela repressão do sul e pelos efeitos da escravidão a assumir um papel passivo, teriam encontrado uma válvula de escape para a agressão reprimida ao deixarem suas regiões de origem. Esta teoria também foi facilmente operacionalizada por Meyer (1991). “Se verdadeira, as rebeliões deveriam ser um comportamento mais frequente para imigrantes do sul do que para os negros criados no norte. A pesquisa do *Free Press* produziu um resultado diferente” (MEYER, 1991).²¹

Com isso, pôde-se perceber como Philip Meyer conseguiu aplicar a metodologia dos estudos das ciências sociais para negar teorias populares e principalmente em um veículo de comunicação, o que foi impactante para a época. A pesquisa feita para essa reportagem derrubou as duas teorias até então aceitas sobre os atos de vandalismo na cidade de Detroit – a quase mil quilômetros de distância de Nova York. Ao contrário do que se pensava as depredações não partiam predominantemente de pessoas com baixo nível de instrução e de negros oriundos do Sul.

Pela cobertura, intitulada *The People Beyond 12th Street*, Meyer e a equipe do *Detroit Free Press* ganharam o Prêmio Pulitzer²² de 1968, na categoria *Local General or Spot News*. No mesmo ano, Meyer foi intitulado “a computer reporter”, pelo periódico *Newsweek*, e definiu aquilo que chamaria de *Precision Journalism*, uma metodologia para a coleta de informações utilizando o método científico e as técnicas de pesquisa das ciências sociais aliados ao uso do computador. (JANSEN, 2011: 13).

O crucial é entender que esta reportagem trouxe grandes avanços na forma como o jornalismo era pensado. Meyer conseguiu destituir duas teorias populares da época por meio de compilação, cruzamento e análise de dados. Ele conseguiu compreender que aplicar métodos científicos no trabalho jornalístico poderia ser uma ferramenta

rioters and compare them with the non-rioters, you will find, if the theory is true, that the less-educated are more likely to be riot participants. The Detroit Free Press did perform such a survey in 1967, and found that people who had attended college were just as likely to participate in the riot as those who had failed to finish high school. The theory was not supported by the data.”

²¹Disponível em:

https://www.attorneygeneral.jus.gov.on.ca/inquiries/cornwall/en/hearings/exhibits/Mary_Lynn_Young/pdf/07_Meyer.pdf. Acessado em: 14/04/2017. Tradução do autor. No original: “If true, rioting should be a more frequent behavior for immigrants from the South than for those blacks raised in the North. The Free Press survey produced a different result.”

²²Criado em 1917, o Prêmio Pulitzer é uma premiação norte-americana voltada para a produção jornalística impressa e online, e para a produção literária e musical. Seu nome é uma homenagem a seu criador, o editor de jornais Joseph Pulitzer.

poderosa para a prática da imprensa. Não bastava mais esperar passivamente. Este trabalho foi o princípio do que se conhece, atualmente, como jornalismo de dados.

2.1.2. REPORTAGEM ASSISTIDA POR COMPUTADOR

A Reportagem Assistida por Computador (RAC), ou Reportagem com Auxílio do Computador, é uma variante do Jornalismo de Precisão. Garrison (1998) a define como a aplicação do uso de computadores para reunir informação na apresentação do noticiário. Basicamente, pode-se afirmar que a RAC é toda e qualquer apuração jornalística que faça uso da informática para encontrar as informações desejadas, ou para descobrir informação inesperada em bases de dados.

Houston (2015) pondera que a reportagem assistida por computador não se refere a jornalistas sentados em frente ao teclado digitando histórias ou navegando na internet. “Refere-se ao download de bases de dados e análise de dados que possam fornecer contexto e profundidade para histórias diárias” (HOUSTON, 2015: 6).²³

As ferramentas da reportagem assistida por computador não substituirão a imaginação de um bom jornalista, a capacidade de conduzir entrevistas reveladoras ou o talento para desenvolver fontes. Mas um jornalista que saiba como usar computadores no dia-a-dia e com o trabalho de longo prazo irá reunir e analisar informações mais rapidamente, e desenvolver e fornecer uma compreensão mais profunda. O jornalista estará melhor preparado para entrevistas e poderá escrever com mais autoridade. Esse jornalista também vai ver histórias potenciais que nunca teria ocorrido a ele ou ela. (HOUSTON, 2015: 6-7).²⁴

Percebe-se, portanto, que as técnicas e ferramentas da RAC vieram para organizar, aperfeiçoar e elevar o trabalho jornalístico a um patamar antes inimaginável. Prova disso são as premiações às produções jornalísticas feitas com auxílio das ferramentas disponíveis no computador. Graças às técnicas, o jornalismo pôde produzir reportagens com números extensos de dados em um tempo consideravelmente menor.

Grandes histórias investigativas estão sendo produzidas com computadores. Alguns artigos divertidos para ler também estão sendo gerados. Nos últimos anos, os projetos assistidos por computador

²³Tradução do autor. No original: “It refers to downloading databases and doing data analysis that can provide context and depth to daily stories.”

²⁴Tradução do autor. No original: “The tools of computer-assisted reporting won’t replace a good journalist’s imagination, ability to conduct revealing interviews, or talent to develop sources. But a journalist who knows how to use computers in day-to-day and long-term work will gather and analyze information more quickly, and develop and deliver a deeper understanding. The journalist will be better prepared for interviews and be able to write with more authority. That journalist also will see potential stories that would never occurred to him or her.”

ganharam Prêmios Pulitzer e outros reconhecimentos nacionais por excelência em jornalismo (GARRISON: 1998, 14).²⁵

É crucial mencionar que a RAC não veio para substituir às velhas boas práticas jornalísticas, como checar os fatos, entrevistar o outro lado ou procurar mais de uma fonte de informação. As práticas jornalísticas com o computador, pelo contrário, só se tornaram ainda mais essenciais. Por exemplo, o jornalista não pode confiar cegamente em uma base de dados. Assim como acontece quando entrevista uma fonte humana, ele deve desconfiar e checar as informações. No caso da base de dados, é importante cruzá-la com outros dados e organizá-la de uma forma que a interpretação seja correta.

Garrison (1998) sugere que a Reportagem Assistida por Computador se refere ao uso da ferramenta em dois níveis para o enriquecimento de uma reportagem:

Primeiro, a RAC inclui o uso de computadores para procurar informações e recuperá-las de outros computadores e de seus bancos de dados. Isso é definido como pesquisa online. Segundo, o termo inclui o uso de computadores para analisar bancos de dados originais e bases de dados de outras fontes para obter informações para as notícias. Isso às vezes é chamado de jornalismo de banco de dados. (GARRISON, 1998: 11).²⁶

Houston (2015) complementa o sugerido por Garrison (1998) com três habilidades que surgiram com a RAC: recursos online, planilhas e gerenciadores de banco de dados. Ele as explica da seguinte maneira:

Recursos online estão disponíveis para os jornalistas através de uma variedade de maneiras. Os recursos on-line incluem e-mail, grupos de discussão, mídias sociais e bancos de dados ativos e arquivados, onde os registros são armazenados. Com recursos online você pode procurar registros judiciais, recuperar registros de campanha ou censo nacional de países, e encontrar milhares de outras bases de dados.

Software de planilha como o Microsoft Excel é bom para analisar números. Você deve pensar em usar uma planilha sempre que você estiver olhando para salários, orçamentos, dados do censo, preços ou relatórios estatísticos. Uma planilha lhe permite filtrar e somar rapidamente colunas de números, compará-los, classificá-los e colocar seus resultados em gráficos. Enquanto uma planilha pode permitir que você faça muito mais, estes são usos rotineiros para reportagens básicas assistidas por computador.

Um gerenciador de banco de dados, como o Microsoft Access, é bom para pesquisar, resumir e relacionar arquivos diferentes conhecidos

²⁵Tradução do autor. No original: “Major investigative news stories are being produced with computers. So are ordinary, day-to-day stories. Some fun-to-read feature articles are being generated as well. In recent years, computer-assisted projects have won Pulitzer Prizes and other national recognition for excellence in journalism.”

²⁶Tradução do autor. No original: “First, CAR includes use of computers to search for information and retrieve it from other computers and their databases. This is referred to as online research. Second, the term includes use of computers to analyze original databases and databases from other sources for information for news stories. This is sometimes referred to as database journalism.”

como tabelas. Um gerente de banco de dados pode agrupar tipos de informações semelhantes e vincular arquivos diferentes por meio de palavras-chave ou números de identificação. (HOUSTON, 2015: 10-11).²⁷

Com o passar do tempo e a evolução dos meios tecnológicos, outras ferramentas se juntaram às três citadas acima, como novos softwares estatísticos, as localizações geográficas precisas com o GPS (sigla em inglês para sistema de posicionamento global) e análises de mídias sociais.

Assim como no *data journalism*, as reportagens com auxílio do computador cruzam dados em planilhas eletrônicas, fazem buscas avançadas na internet e em banco de dados públicos, fazem raspagem de dados, usam as redes sociais – como Facebook, Twitter e Instagram – como fonte de coleta de informações e, por fim, elaboram a visualização dos dados, seja em gráficos, mapas ou qualquer outro tipo de apresentação.

2.2. A PIRÂMIDE INVERTIDA DO JORNALISMO DE DADOS

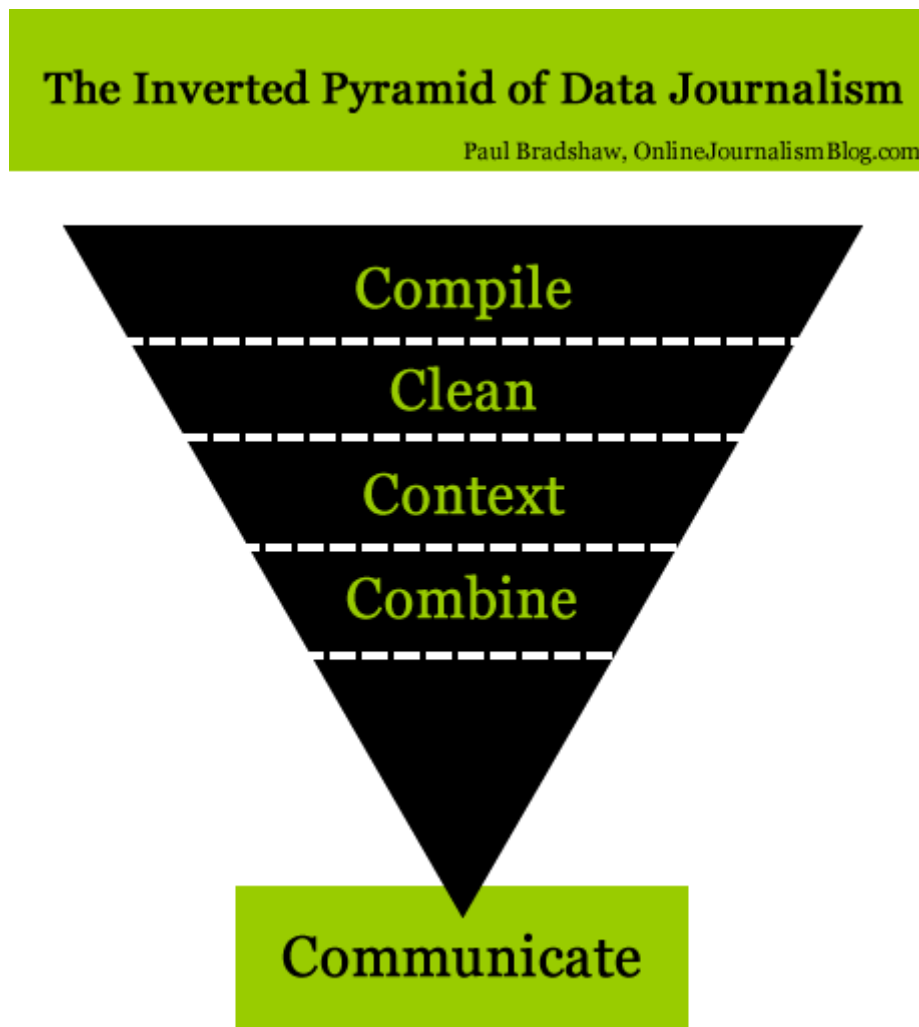
O jornalista inglês Paul Bradshaw formulou, em 2011, o que ele chamou de a pirâmide invertida do jornalismo de dados. Para ele, é invertida porque, no jornalismo de dados, a reportagem começa com um conjunto grande de informação e, conforme os dados vão sendo depurados, ela diminui, ganha um enfoque e chega ao ponto em que os dados podem ser comunicados. Esta pirâmide norteia as etapas envolvidas no processo de elaboração das reportagens guiadas por dados. O diagrama pensado pelo jornalista fica claro nesta imagem:

²⁷Tradução do autor. No original: “Online resources are available to journalists through a variety of ways. Online resources include email, discussion groups, social media, and active and archived databases, where records are stored. With online resources you can look up court records, retrieve campaign records or national census of countries, and find thousands of other databases.

Spreadsheet software such as Microsoft Excel is good for analyzing numbers. You should think about using a spreadsheet whenever you are looking at salaries, budgets, census data, prices, or statistical reports. A spreadsheet allows you to quickly filter and sum columns of numbers, compare them, sort them, and put your results into charts. While a spreadsheet can allow you to do much more, these are routine uses for basic computer-assisted reporting.

A database manager such as Microsoft Access is good for searching, summarizing, and relating different files known as tables. A database manager can group similar kinds of information and link different files through keywords or identification numbers.”

Figura 1



Em seu blog “Online journalism blog”, Paul Bradshaw, elabora a seguinte explicação para cada uma das etapas:

- 1) Compilar – o jornalismo de dados começa de uma dessas duas maneiras: ou o jornalista tem uma pergunta que requer o uso de dados ou tem uma base de dados que precisa ser interrogada. De qualquer forma, a compilação da informação – ou seja, a aglutinação de um conjunto de dados – é o que se define como um ato de jornalismo de dados e pode ocorrer de várias maneiras: a) uma organização, pública ou privada, envia a um ou vários jornalistas, um conjunto de dados para que eles trabalhem em cima; b) a partir de técnicas de busca avançada podem ser obtidas bases de dados valiosas. A busca avançada do Google é uma importante ferramenta para encontrar informações “escondidas” na enorme quantidade de informações disponíveis na internet; c) a raspagem de dados também permite extrair bases de dados de sites que não disponibilizam o conjunto de informação por meio

de ferramentas como Web Scraper, OutWit Hub e Scraperwiki; d) da conversão de documentos em algo que possa ser analisado. A chamada transformação de arquivos fechados, como arquivos pdf, em arquivos abertos, como planilhas de Excel. Isso pode ser feito com ferramentas como DocumentCloud, Zamzar e Tabula; e) a partir uma coleta de dados feita pelo próprio jornalista por meio de pesquisas, formulários online, dentre outros. Esta fase da compilação é a mais importante, - não só porque tudo depende disso, mas porque é provavelmente a mais revisitada no processo. Em cada uma das fases subsequentes - limpeza, contextualização, cruzamento e comunicação - pode ser preciso compilar mais informações. Por isso, é necessário muita atenção nesta parte, pois a precisão do trabalho tem grande responsabilidade nesta fase da apuração.

- 2) Limpar – basicamente aqui, o importante é clarear todo o conjunto de dados estruturado e torná-lo apto a uma análise confiável. Nesta parte, é necessário remover o erro humano e converter o dado em um formato consistente com outra base de dados utilizada. Por exemplo, as bases de dados, muitas vezes, incluem alguns ou todos os erros a seguir: entradas duplicadas, entradas vazias, utilização de valores padrão para economizar tempo, formatação incorreta (por exemplo, palavras em vez de números), entradas corrompidas ou entradas com código HTML²⁸, vários nomes para a mesma coisa (por exemplo, I.B.G.E., IBGE e Instituto Brasileiro de Geografia e Estatística), além da falta de dados. Há formas simples de limpar os dados no Excel ou no Google Sheets²⁹, como, por exemplo, com o uso do “localizar e substituir”, classificando para encontrar entradas estranhamente altas ou baixas ou vazias, e usando filtros para que só apareçam as entradas duplicadas. Para limpezas maiores e mais elaboradas, uma excelente ferramenta é o Google Refine. Uma função interessante desta ferramenta é “*common transforms*”, que pode, por exemplo, converter todas as entradas para letras minúsculas, maiúsculas ou apenas a inicial maiúscula. Pode remover HTML, remover espaços antes e após as entradas (que podem não ser vistas pelo usuário, mas que os computadores vão ver como uma diferença para os mesmos dados que não tenham espaço), remover espaços duplos, juntar e dividir células, e formatá-las de forma consistente.
- 3) Contextualizar – como toda reportagem, é necessário contextualizar os dados para que eles não fiquem difusos nem confusos na cabeça do leitor.
- 4) Combinar – nesta etapa, é realizado o cruzamento de dados que permitem ao jornalista uma análise interessante da base ou das bases de dados. Uma

²⁸ HTML é a linguagem base da internet, utilizada na construção de websites. HTML é a sigla de *HyperTextMarkupLanguage*, expressão inglesa que significa "Linguagem de Marcação de Hipertexto".

²⁹É a ferramenta de planilha do Google. Disponível em: sheets.google.com

combinação trivial seria cruzar duas ou mais bases de dados com uma coluna em comum. Por exemplo, o nome de um jogador de futebol, um time ou uma localização. Isso usualmente significa assegurar que o ponto em comum esteja formatado com o mesmo nome em cada uma das bases de dados.

- 5) Comunicar – esta seria a parte “legal” do trabalho do jornalismo de dados, pois é neste momento em que a criatividade é necessária na hora de escolher a melhor e mais eficiente forma de visualizar os dados, seja em um mapa, um gráfico ou uma animação. As visualizações exercem múltiplos papéis na hora da publicação porque elas podem tanto ilustrar um ponto levantado no texto de forma mais atraente como excluir do texto dados técnicos desnecessários. Além disso, vale destacar que uma informação visual tem um impacto imediato e bem mais profundo nos espectadores do que um texto corrido sem imagens, ou seja, quando o jornalista apresenta uma visualização com informações interativas a um usuário, estas informações serão bem mais efetivas e serão mais bem processadas pelo leitor.³⁰

2.3. MOVIMENTO PELA ABERTURA DE DADOS

A cultura de dados abertos apesar de estar crescendo, ainda precisa ser mais difundida porque a cultura do sigilo das informações ainda está presente em muitos países, principalmente em regimes de governantes mais conservadores. Para o jornalismo, a militância pela abertura dos dados é crucial. Afinal, os números são a matéria prima do trabalho dos jornalistas de dados. Para se fazer esse tipo de jornalismo muitos caprichos podem ser deixados de lado, mas a matéria prima não pode faltar. Com ela, os jornalistas podem produzir matérias esclarecedoras ao cidadão sobre, por exemplo, como determinada parte da administração pública de uma cidade, estado, país ou organização está sendo gerida. Uma cultura de dados abertos, principalmente em organizações governamentais, que gerem informações públicas, é essencial para o exercício da transparência.

No relatório da terceira Conferência Internacional de Dados Abertos, realizada em Ottawa, no Canadá, em maio de 2015, foi exposto o que são os dados abertos e quais os benefícios da difusão dos mesmos na sociedade:

³⁰Disponível em: <https://onlinejournalismblog.com/2011/07/07/the-inverted-pyramid-of-data-journalism/>. Acessado em: 17/04/2017.

Os dados abertos são dados disponibilizados para que qualquer pessoa possa acessar, usar e compartilhar. Em vez de aumentar a assimetria da informação - onde algumas organizações e instituições mantêm os dados públicos fechados - com dados abertos, o terreno é estabelecido para todos os cidadãos se envolverem na formulação da política. Com os dados abertos, as organizações e os indivíduos podem dedicar tempo à solução colaborativa de problemas, ao invés de negociar o acesso a um conjunto de dados. Com dados abertos, inovação, conhecimento e novos modelos de negócios são possíveis, gerando impactos de grande escala. (CONFERÊNCIA INTERNACIONAL DE DADOS ABERTOS, 2015: 12).³¹

Por isso, percebe-se a importância da cultura de dados abertos. Uma sociedade com as informações públicas disponíveis de forma ampla tende a ter uma atuação maior sobre as medidas tomadas pelo poder público, uma vez que tem conhecimento sobre as suas próprias necessidades. O exercício da cidadania passa a ser mais efetivo. Além disso, o trabalho do jornalista que têm o papel de divulgar a informação correta ao público é facilitado. É mais fácil e ágil trabalhar com algo que já está disponível do que ter que solicitar uma informação, que, por princípio, já deveria estar pública e acessível.

A militância pela abertura de dados passou, então, a ser maior a partir do momento que foram percebidas mudanças positivas em torno dos dados abertos. Ou seja, jornalistas, cientistas e interessados passaram a pressionar governos a liberarem bases de dados públicas e, principalmente, a disponibilizá-las de forma integral (sem cortes ou censura) e aberta. A Associação Brasileira de Jornalismo Investigativo (Abraji), por exemplo, foi essencial nesta luta pela liberação dos dados.

Em prol da transparência, as demandas pela disponibilização dos dados governamentais ganharam força e a pressão no corpo político fez com que leis de acesso à informação fossem elaboradas em diversos países. A Acess Info Europe³² anunciou, em 18 de setembro de 2014, que o Paraguai era o centésimo país a promulgar uma lei de acesso à informação no mundo. A organização notou ainda que repetidas pesquisas de monitoramento mostram que países com leis de acesso à informação geralmente têm melhores níveis de resposta a pedidos de informação, mas a maioria dos países ainda

³¹Disponível em: <http://1a9vrva76sx19qvtvg1ddvt6f.wpengine.netdna-cdn.com/wp-content/uploads/2015/11/opendatacon-report-en-web.pdf>. Acessado em: 17/04/2017. Tradução do autor. No original: “Open data is data that is made available for anyone to access, use, and share. Instead of increasing information asymmetry—where a few organizations and institutions keep public data locked up—with open data the groundwork is laid for all citizens to get involved in shaping policy. With open data, organizations and individuals can spend time on collaborative problem solving, rather than negotiating access to datasets. With open data, innovation, insights, and new business models are made possible, generating impacts that can scale.”

³²É uma organização de direitos humanos dedicada à promoção e proteção do direito de acesso à informação. Ela possui uma série de projetos destinados a alavancar o direito à informação, a fim de aumentar a participação e a prestação de contas, defender os direitos humanos e promover a democracia.

têm problemas com implementações inconsistentes, uso exagerado de exceções e altos níveis de silêncio administrativo.

Em 1994, apenas 15 países tinham leis de acesso à informação promulgadas. Desde então, as demandas por dados abertos e governo aberto, além do trabalho dos Defensores da Liberdade de Informação ajudaram a promover esta cultura da abertura dos dados. Desde 2006, grandes grupos internacionais de direitos humanos, incluindo a *UN Human Rights Committee* (Comitê de Direitos Humanos das Nações Unidas), *European Court of Human Rights* (Corte Europeia de Direitos Humanos) e o *Inter-American Court of Human Rights* (Corte Interamericana de Direitos Humanos) confirmaram que o acesso à informação é um direito humano fundamental ligado à liberdade de expressão.

Segundo o relatório Índice de Dados Abertos para o Brasil, produzido pela Diretoria de Análise de Políticas Públicas da Fundação Getulio Vargas (FGV/DAPP)³³ em parceria com a Open Knowledge Brasil³⁴, lançado no dia 27 de abril de 2017, o Brasil alcançou uma pontuação de 64%. Isso coloca o país em oitavo lugar no ranking mundial de dados abertos. Entre os países da América Latina, o Brasil é o primeiro da lista. A conclusão do documento é positiva, mas algumas ressalvas são importantes.

O Brasil teve um bom desempenho no índice quando comparado aos demais países. No entanto, a pontuação de 65% indica que ainda há muito espaço para aprimoramentos. Apenas 40% das dimensões do índice receberam a nota total e três delas nem mesmo existem publicamente. A maior parte dos problemas identificados está mais relacionada à usabilidade (completude, atualização e formato) do que a processo (meios de acesso e licença de uso dos dados).

O problema mais recorrente é a dificuldade de se trabalhar os dados e, em segundo lugar, a não disponibilização da base de dados completa para download. Isso indica que os órgãos responsáveis pelo provimento das informações talvez não tenham a compreensão de que a facilitação do processo de edição e manipulação dos dados é também sua responsabilidade e parte importante da transparência.

Tais resultados podem auxiliar na abertura de dados no âmbito nacional, direcionando a atenção para dimensões prioritárias e problemas mais comuns de acordo com a definição de dados abertos. Em outras palavras, este documento traz algumas diretrizes para uma política de dados abertos do país.

³³A Diretoria de Análise de Políticas Públicas da Fundação Getulio Vargas (FGV/DAPP) é um centro de pesquisa social aplicada voltado à inovação para políticas públicas. O centro produz análise de ponta com uso intensivo de redes sociais e conhecimento interdisciplinar. Além disso, a FGV/DAPP tem como missão aprimorar a gestão pública brasileira e qualificar o debate público na sociedade em rede, por meio da transparência e do diálogo entre o Estado e a cidadania.

³⁴A Open Knowledge Brasil (OKBr) é uma organização da sociedade civil sem fins lucrativos e apartidária, fundada em 2013. A OKBr utiliza e desenvolve ferramentas cívicas, faz análises de políticas públicas, trabalha com jornalismo de dados e promove o conhecimento livre para tornar a relação entre governo e sociedade mais transparente e para que haja uma participação política mais efetiva e aberta.

Essa iniciativa pode ter uma série de benefícios, tanto internamente quanto externamente à administração pública. Internamente, porque o próprio processo de abertura de dados cria um mecanismo de cobrança interna para resultados das diferentes unidades, o que contribui para uma maior eficiência da gestão pública. Além disso, a abertura de dados permite que os cidadãos possam contribuir de diversas formas, seja com análises, controle social, criação de aplicativos, replicação da informação, entre outras, o que acaba por tornar a gestão mais eficiente ao congrega esforços por meio da participação da sociedade. Por outro lado, há também benefícios externos: a administração passa a ser vista como mais *accountable* aos seus cidadãos. Além dos ganhos de imagem, a possibilidade de se tornar referência internacional em dados abertos contribui positivamente para o *advocacy* da questão da transparência em outros países. (ÍNDICE DE DADOS ABERTOS PARA O BRASIL, 2017: 21).³⁵

2.3.1. A LEI DE ACESSO À INFORMAÇÃO NO BRASIL

Conforme Medeiros (2014), o acesso à informação é indispensável para a promoção de uma maior transparência e conseqüentemente menos corrupção nos órgãos públicos. Quanto mais transparente é uma empresa, maior a tendência de não haver corrupção entre seus membros e deles com os demais. “A transparência no acesso às informações auxilia a promover a confiança da população no governo e viabiliza sua prestação de contas” (MEDEIROS, 2014: 55).

Segundo Stiglitz (apud MEDEIROS, 2014: 59), nas sociedades onde existe pouca transparência nos atos da administração pública são comuns práticas paternalistas, clientelistas, corrupções e outras formas de utilização dos bens públicos para atingir interesses particulares.

Por esta razão, esforços têm sido empregados com mais frequência na tentativa de promoção de uma maior transparência das ações governamentais. Isso ocorre porque a melhora do acesso à informação pública e a criação de regras que permitem a disseminação das informações produzidas pelo governo reduzem o escopo dos abusos que podem ser cometidos (MEDEIROS apud STIGLITZ, 2014: 59).

Com isso claro, um grande passo para a democracia brasileira foi tomado a partir da promulgação da Lei de Acesso à Informação (LAI) – Lei 12.527 –, em novembro de 2011. Com a regulamentação da LAI em maio de 2012, os Poderes Executivo, Legislativo e Judiciário da União, dos Estados e dos municípios, as autarquias, fundações públicas, empresas públicas, sociedades de economia mista e outros órgãos controlados pela União, Estados ou municípios e as instituições sem fins lucrativos que

³⁵Disponível em: <http://dapp.fgv.br/transparencia-orientamentaria/indices/IndexeDadosAbertosBrasil2017.pdf>. Acessado em: 07/05/2017.

recebam dinheiro público passaram a ter determinadas obrigadoriedades que antes não eram previstas por leis.

Um exemplo disso é que os órgãos públicos passaram a ter que disponibilizar todos os registros de repasses ou transferências de recursos financeiros e despesas do órgão, além de informações de licitações realizadas e em andamento (editais, resultados e contratos, notas de empenho). Outra obrigadoriedade que vale destaque é a necessidade de divulgação de dados sobre programas, ações, projetos e obras, com indicação da unidade responsável, principais metas e indicadores de resultado e impacto. Isso tudo foi um ganho para a sociedade que antes vivia às margens do que o jornalismo conseguia “arrancar” destas organizações e torná-las públicas por meio de reportagens. Neste momento, qualquer pessoa ou empresa pode fazer pedidos de informação a órgãos públicos por qualquer meio válido, tais como: telefone, carta, e-mail, fax, pessoalmente ou por formulário em sites oficiais. Ao realizar a solicitação, o órgão público é inviabilizado, segundo a lei, de questionar os motivos pelos quais se está pedindo a informação. A lei ainda determina um prazo de 20 dias para resposta do órgão público, que pode ser estendido por no máximo mais dez dias, caso este necessite e justifique de acordo com as regras vigentes. Caso o órgão não cumpra as determinações, cabe ao cidadão denunciar aos órgãos superiores competentes.

Para os jornalistas, categoria que passou a fazer uso da LAI, a norma fez com que a apuração da informação fosse facilitada. Ela passou a representar a conquista de uma nova ferramenta de trabalho. Dutra (2015) relatou que a lei virou uma fonte de informações oficiais alternativa às assessorias de imprensa dos órgãos públicos.

Os profissionais identificaram o seguinte movimento nas assessorias: quando se trata de assunto mais delicado, os próprios assessores de imprensa orientam os jornalistas a solicitarem a informação via Lei de Acesso. Alguns profissionais relatam isso como um aspecto positivo, visto que em determinados casos não se trata de ineficiência do assessor, mas sim um reflexo da postura assumida por seus superiores hierárquicos naquele órgão. Por outro lado, alguns acreditam que essa atitude muitas vezes é banalizada pelas assessorias, que indicam a Lei como uma alternativa para se esquivarem de disponibilizar as informações solicitadas. Os jornalistas também destacaram como uma consequência positiva da lei a redução da dependência das assessorias e imprensa, e a própria mudança na postura de algumas delas no sentido de não questionar a motivação, ou a finalidade da informação solicitada (DUTRA, 2015: 9)³⁶.

³⁶Disponível em: http://www.abraji.org.br/seminario/PDF/2/luma_poletti_dutra.pdf. Acessado em: 23/04/2017.

Para Medeiros (2014), a lei 12.527 é um avanço na efetivação do direito ao acesso à informação no país. Ela pontuou alguns temas importantes da LAI:

Dentre os aspectos relevantes, pode-se citar a estipulação de prazos razoáveis para a obtenção de informações junto aos órgãos públicos, o uso das tecnologias da informação e de comunicação como facilitador do acesso às informações públicas. Além da obrigatoriedade dos órgãos públicos promoverem a realização de audiências ou consultas públicas, incentivando a participação popular ou a outras formas de divulgação. (MEDEIROS, 2014: 70-71).

Por fim, Medeiros (2014) fez uma ressalva sobre a lei ao contar sobre as dificuldades em relação a como ela impactará as rotinas dos órgãos públicos e seus servidores, devido ao possível aumento na demanda de solicitação de informações pelos cidadãos aos órgãos públicos. Isso porque muitos desses órgãos não têm estrutura suficiente para suportar a alta demanda que a lei sugere. “Outro entrave a ser superado refere-se a não estipulação de um parâmetro na cobrança de taxas, o que pode levar a cobranças abusivas e, conseqüentemente, a diminuição de pedidos de informação” (MEDEIROS, 2014: 71).

3. O PERFIL DO JORNALISTA DE DADOS

Além da observação das velhas práticas³⁷ para se tornar um bom repórter, os interessados em se especializar no jornalismo de dados precisam adquirir algumas habilidades específicas, tais como ter domínio de línguas estrangeiras – principalmente inglês e espanhol – e saber manejar algumas das diversas ferramentas disponíveis na internet, que ajudam a cumprir as etapas dos trabalhos jornalísticos guiados por dados.

Além disso, é importante declarar que o jornalista de dados precisa de uma formação contínua e de conhecimento sobre assuntos pelos quais os jornalistas tendem a querer evitar, como a matemática e a programação. A formação contínua se faz necessária porque as tecnologias evoluem diariamente e novas ferramentas surgem para atender a demandas de todo o processo das reportagens guiadas por dados, desde a obtenção até a visualização da história contada por dados.

Sandra Crucianelli (2013) acrescenta ainda que é preciso conhecer bem o funcionamento da administração pública, saber interpretar as leis e decretos, uma vez que os dados públicos são importantes fontes para o trabalho do jornalista de dados. “Se um jornalista não entende como funcionam as instituições de seu país, dificilmente poderá tomar o caminho com êxito” (CRUCIANELLI, 2013: 89)³⁸.

3.1. HABILIDADES E TÉCNICAS NECESSÁRIAS

Uma base de dados é a fonte primordial do trabalho do jornalista de dados. Mas, o caminho para obtê-los nem sempre é o mais acessível. Assim como disse Sophie Warnes (2016), trabalhar com os dados sem precisar organizar ou limpar³⁹ seria basicamente um caso de muita sorte, pois é necessário um trabalho prévio na obtenção e estruturação dos mesmos.

Normalmente, você precisa encontrar os dados online em algum lugar e raspá-los para que você possa usá-los. Então você precisa limpá-los. Em alguns casos, os dados são separados e não estão em um formato

³⁷Velhas práticas aqui referem-se à capacidade do jornalista de saber apurar e checar bem as histórias, de saber escrever, de ter um excelente domínio da língua na qual escreve (para jornalistas de veículos brasileiros seria o português). Além de atuar eticamente durante todo o processo de elaboração de uma reportagem.

³⁸ Disponível em: <https://knightcenter.utexas.edu/books/HDPP.pdf>. Acessado em: 14/05/2017. Tradução do autor. No original: “Si un periodista no comprende como funcionan las instituciones de su país, dificilmente podrá emprender el camino con éxito”.

³⁹Limpar para o trabalho do jornalista de dados refere-se à etapa de organizar os dados de uma forma que seja possível analisar o conteúdo da base de dados claramente e corretamente. Dados que não estão bem estruturados impossibilitam uma análise propícia e podem ser prejudiciais se usados de forma incorreta na produção do conteúdo jornalístico.

digital, por exemplo, recibos. Isso pode significar digitar e organizar todos os recibos físicos (WARNES, 2016)⁴⁰.

Logo, a primeira dificuldade encontrada é a obtenção dos dados necessários para uma reportagem. A pirâmide invertida do jornalismo de dados (apresentada no Capítulo 2 deste trabalho) divide o trabalho do jornalista de dados em cinco etapas: compilar, limpar, contextualizar, combinar e comunicar. Por isso, o especialista em dados deve conhecer cada uma delas para que a mensagem possa ser transmitida ao leitor da melhor maneira possível. Aquele que não tem conhecimentos sobre alguma das etapas pode ter o fluxo do trabalho prejudicado e, conseqüentemente, uma falha na comunicação da reportagem. Ou seja, é importante que o repórter tenha as técnicas necessárias em cada uma das etapas e saiba usar as ferramentas disponíveis na rede para construir com eficácia uma reportagem guiada por dados.

3.1.1. COMPILAR

A primeira etapa de compilar os dados é, muitas vezes, a mais árdua no processo inteiro e também primordial para a definição do caminho da reportagem. O jornalista que não encontrar os dados almejados inicialmente ou só encontrá-los parcialmente terá que rever a forma de contar a história, uma vez que a base estrutural principal da matéria não estará completa. Por isso a importância da fase de compilação dos dados.

É importante que o especialista em dados tenha domínio sobre algumas habilidades necessárias para encontrar e capturar dados, tais como: busca avançada do Google e de rede sociais – como Twitter, Facebook e Instagram, raspagem de dados e conversão de dados fechados⁴¹ em abertos.

3.1.1.1. BUSCA AVANÇADA DO GOOGLE

A busca avançada do Google auxilia a se obter com maior precisão e rapidez a informação que o cidadão almeja. O guia de ferramentas do Google para jornalistas,

⁴⁰Disponível em: <https://medium.com/@sophiewarnes/what-is-data-journalism-what-does-a-data-journalist-do-bd914bed815b>. Acessado em: 14/05/2017. Tradução do autor. No original: “Usually, you need to find the data online somewhere and scrape it so you can use it. Then you need to clean it. In some cases, the data is separate and not in a digital format, for instance, receipts. This might mean typing up and organising all physical receipts.”

⁴¹Dados fechados são aqueles nos quais o trabalho sobre eles é dificultado, uma vez que eles não podem ser reorganizados. Por exemplo, arquivos em PDF ou Jpeg são considerados fechados porque não é possível manejar os dados de acordo com as necessidades do jornalista. Para conseguir trabalhar com as informações da forma como deseja, o jornalista deve, então, converter, por exemplo, este arquivo para um formato aberto, como o CSV ou XML: formatos para trabalhar em programas de planilhas e de banco de dados.

elaborada por Mauricio Marín (2010) apresenta algumas dicas de como fazer as pesquisas no navegador. Primeiro, começar a busca pelos termos principais. Depois, delimitar a pesquisa com mais termos caso os resultados anteriores não forem satisfatórios. Escrever os termos buscados de forma lógica, por exemplo, é melhor escrever “Farmácia Tamoio Rio de Janeiro” do que “Tamoio Rio de Janeiro Farmácia”. Além disso, Marín (2010) dá uma dica interessante sobre realizar as buscas quando logado em uma conta do Google ou do Gmail (pertencente ao Google).

[...] uma recomendação: abra uma conta Google ou Gmail (ou faça login com ela quando você iniciar a pesquisa, se você já tiver uma conta). Pesquisar e navegar com uma conta do Google não só irá permitir que você personalize melhor as pesquisas, mas também armazená-las usando o recurso Histórico da web, e usar uma variedade de serviços do Google, desde Google Docs, Calendar e Google Reader até o YouTube e Blogger, entre outros (MARÍN, 2010: 14)⁴².

Há ainda a gramática do Google, que permite alcançar os resultados com uma maior precisão. Neste trabalho, serão delimitadas algumas das principais ferramentas desta gramática que auxilia na busca de termos específicos da pesquisa desejada:

1) Exclusão de palavras (-): em determinadas ocasiões, os resultados encontrados na busca não se referem ao desejado inicialmente ou se confundem com outros termos. Por exemplo, se for feita uma pesquisa por “Roberto Carlos”, podem ser exibidas informações tanto do cantor brasileiro quanto do ex-lateral esquerdo da seleção brasileira. Para delimitar informações apenas do ex-jogador, basta inserir na pesquisa “-cantor”, que todos os resultados com cantor serão excluídos automaticamente.

2) Busca de frases específicas (“frase”): a busca por termos entre aspas facilita a pesquisa porque o resultado exibirá apenas os termos entre as aspas, de acordo com a maneira como foram digitadas.

3) Pesquisa em sites específicos (site:): quando uma busca é feita e já se sabe, previamente, de onde se quer encontrar o resultado, basta digitar o site desejado que o Google exibirá os conteúdos publicados por aquele site específico.

4) Busca por domínios da internet (site:edu, site:com, site:com.br): a pesquisa por domínios ajuda no momento da delimitação de que tipos de resultados se deseja

⁴² Tradução do autor. No original: “una recomendación: abra una cuenta de Google o Gmail (o ingresar com ella al iniciar una sesión de búsquedas, si ya tiene una). Hacer búsquedas y navegar con una cuenta de Google no solo le permitirá personalizar mejor las búsquedas, sino también almacenarlas mediante la función Historial Web, y utilizar una gran variedad de servicios de Google, desde Google Docs, Calendar y Google Reader hasta YouTube y Blogger, entre otros”.

encontrar. Se a pessoa quer apenas resultados de sites governamentais, basta delimitar por “site:org”, por exemplo.

5) Pesquisa por tipo de documento (filetype:pdf): escolher o formato do conteúdo buscado é importante pois acelera o desempenho, por exemplo, do jornalista de dados, que já consegue encontrar o arquivo no formato desejável para se trabalhar.

Além destas ferramentas de busca, o usuário pode pesquisar diretamente pela aba da pesquisa avançada do Google⁴³. Nesta página, o usuário pode pesquisar conforme as delimitações descritas acima e mais outras, tais como descreve Marín (2010: 18)⁴⁴: “[...] o idioma desejado nos resultados e na origem dos sites web dos resultados (o país onde estejam os servidores que os alojam), características que podem ser úteis para investigações específicas”.

Sandra Crucianelli (2013) publicou um artigo no portal Scidev.Net⁴⁵ relatando os benefícios da busca avançada do Google no trabalho no jornalístico, que valoriza as ferramentas apresentadas acima:

A busca avançada do Google permite restringir os resultados por extensão de domínio, ajudando-o a procurar dados acadêmicos ou do governo, por formatos de arquivo, como arquivos em Excel, em que é mais provável encontrar tabelas de números e estatísticas. Tabelas e gráficos são geralmente salvos como imagem, por isso a pesquisa de dados também deve incluir Flickr e Google Imagens (CRUCIANELLI, 2013)⁴⁶.

3.1.1.2. BUSCA AVANÇADA PELAS REDES SOCIAIS

Tanto o Twitter⁴⁷, quanto o Facebook⁴⁸ e o Instagram⁴⁹ podem ser fontes de dados para artigos jornalísticos. Dentre eles, o Twitter pode ser considerado o mais relevante para o trabalho de um jornalista porque é uma rede social que funciona como se fosse um “tempo real” dos assuntos principais e relevantes do momento. É ali que os cidadãos vão falar sobre problemas no tráfego, política, criminalidade, comentar sobre o

⁴³Link para pesquisar pela Busca Avançada do Google: https://www.google.com.br/advanced_search.

⁴⁴ Tradução do autor. No original: “[...] el idioma deseado para los resultados y el origen de los sitios web de los resultados (el país donde estén los servidores que los alojan), características que pueden ser útiles para investigaciones específicas.”

⁴⁵ Disponível em: <http://www.scidev.net>

⁴⁶ Disponível em: <http://www.scidev.net/america-latina/periodismo/guia-practica/periodismo-de-datos-c-mo-hacer-historias-con-n-meros.html>. Acessado em: 21/05/2017. Tradução do autor. No original: “La búsqueda avanzada de Google le permite delimitar los resultados por la extensión del dominio, ayudándolo a buscar datos académicos o del gobierno y el formato del archivo, como archivos en Excel, em los cuales es más probable encontrar tablas de cifras y estadísticas. Los cuadros y gráficos generalmente se cargan como imagen, por lo que su búsqueda de datos también debe incluir Flickr y Google Images”.

⁴⁷ Disponível em: www.twitter.com

⁴⁸ Disponível em: www.facebook.com

⁴⁹ Disponível em: www.instagram.com

jogo de futebol que está acontecendo, a notícia do dia ou a declaração que alguma autoridade pública acabou de fazer. “Essas características fazem da plataforma uma ferramenta poderosa de monitoramento. O Twitter é uma boca por onde as redações emitem conteúdos, mas é também um ouvido apurado da sociedade” (STAMILLO & CIPRIANI, 2015: 3).

Graças a essas características, o Twitter é uma boa fonte para medir a reação das pessoas sobre determinado acontecimento da sociedade. A medição costuma ser feita a partir da quantidade de tweets⁵⁰ que citaram tal acontecimento. Quanto mais tweets com o tema, mais claro fica que o assunto viralizou na rede. Por exemplo, o jornal Nexo⁵¹ publicou, em maio de 2017, a reação das pessoas sobre a divulgação de dois áudios envolvendo presidentes da República: Dilma e Temer. A primeira gravação virou notícia no dia 16 de março de 2016 e a segunda no dia 17 de março. Com isso, o jornal foi ao Twitter para medir, nestes dias, como o público reagiu às notícias. A matéria⁵² foi reveladora e mostrou como a rede social pode ser uma fonte alternativa interessante para analisar os impactos de um fato da sociedade entre as pessoas.

O Twitter pode ser ainda um excelente buscador de fontes para matérias e reportagens. Isso porque a busca avançada da rede⁵³ permite, por exemplo, encontrar pessoas que “tweetaram”⁵⁴ sobre determinado assunto perto do local onde aconteceu o fato. Por exemplo, se ocorre um deslizamento de terra no morro do Bumba, em Niterói, região metropolitana do Rio de Janeiro, às 22h de um domingo, e uma pessoa publica no Twitter um post sobre o assunto após o desastre, ela vira um potencial personagem para a matéria de um jornal. Tanto o tweet pode ser exibido na reportagem quanto a pessoa pode ser localizada para testemunhar sobre o fato.

Para encontrar quem está “tweetando” sobre determinado acontecimento, basta ir até a busca avançada do Twitter e delimitar como deseja realizar a busca. Pode-se pesquisar, por exemplo, por palavras específicas, por idioma, por hashtags ou até por geolocalização. O último é importante porque consegue delimitar os tweets para um raio de distância próximo de onde ocorreu determinado acontecimento. Isso enriquece a pesquisa e faz com que a busca seja ainda mais precisa.

⁵⁰Tweet é o nome dado pelo Twitter a todas as publicações feitas nesta rede social. A postagem é curta com textos de até 140 caracteres.

⁵¹ Disponível em: <https://www.nexojornal.com.br>

⁵² ALMEIDA, Rodolfo; OSTETTI, Vitória. Como o Twitter reagiu às gravações de Temer e Dilma. Nexo, 2017. Disponível em: <https://www.nexojornal.com.br/grafico/2017/05/19/Como-o-Twitter-reagiu-%C3%A0s-grava%C3%A7%C3%B5es-de-Temer-e-Dilma>. Acessado em: 21/05/2017.

⁵³Link para pesquisar pela busca avançada do Twitter: <https://twitter.com/search-advanced>.

⁵⁴ Vem do verbo tweetar, neologismo criado na rede para indicar a ação de publicar um tweet.

O GloboEsporte.com publicou uma matéria⁵⁵ sobre a cronologia dos fatos após o acidente envolvendo o avião da Chapecoense⁵⁶, que deixou 71 mortos em novembro de 2016. A reportagem capturou tweets colombianos relatando como os fatos eram apurados e divulgados na rede. A partir disso, o site organizou os posts na rede e mostrou a ordem dos fatos divulgados pelas autoridades colombianas.

Enquanto o Twitter tem essa característica de relatar em pequenas publicações os fatos que acontecem na sociedade, o Facebook e o Instagram são redes sociais que divulgam aspectos mais pessoais dos indivíduos. As características são diferentes, mas ainda assim são importantes ferramentas para auxiliar o trabalho do jornalista.

O Facebook é mais heterogêneo. Nele, podem ser buscadas publicações pessoais de indivíduos célebres da sociedade, grupos sobre determinado interesse em comum ou até eventos sociais para reuniões ou manifestações. Isso tudo faz com a rede seja uma excelente fonte de trabalho dos jornalistas.

Por exemplo, caso um repórter queira personagens para uma matéria sobre uma greve em uma universidade federal. Ele pode ir até grupos do Facebook da universidade para conversar com alguns estudantes e professores. Pode buscar páginas do Facebook que estejam organizando ou envolvidas com o movimento. Além de contactar, pelo processo de busca da rede social, possíveis fontes para a matéria.

O Instagram, rede social destinada à publicação de fotos e vídeos, pode servir para encontrar publicações de pessoas famosas, como atletas, atores e políticos, para coletar anúncios de eventos importantes de algumas organizações ou para ser estruturada uma base de dados com fotos publicadas sobre determinado tema.

Por exemplo, o Instagram funciona muito bem para medir uma manifestação. Isso porque muitas pessoas postam na rede fotos ou vídeos no protesto. Essas publicações podem ser localizadas por meio de hashtags incluídas nas postagens ou por geolocalização – quando as publicações inserem a localização da foto ou vídeo. Isso tudo pode ser coletado e pode servir de indicador da participação das pessoas no ato.

⁵⁵ GLOBOESPORTE.COM. Veja a cronologia da tragédia a partir da visão colombiana. GloboEsporte.com, 2016. Disponível em: <http://www.playbuzz.com/globoesportecom/veja-a-cronologia-da-trag-dia-a-partir-da-vis-o-colombiana>. Acessado em: 21/05/2017.

⁵⁶No dia 29 de novembro, um acidente de avião chocou o Brasil. Jogadores, jornalistas e tripulação foram vítimas de um acidente aéreo, que matou 71 deles. Os atletas da Chapecoense estavam viajando rumo à Medellín para disputar a final da Copa Sul-Americana contra o Atlético Nacional, time da cidade.

Uma ferramenta simples que pode auxiliar esta coleta é o IFTTT (*IfThisThanThat*)⁵⁷. O site tem uma série de serviços que programam atividades de forma automática por meio de um robô virtual. A lógica da ferramenta é a seguinte: se algo acontecer, então faça isso para mim. As fotos postadas no Instagram durante uma manifestação poderiam ser facilmente programadas para serem capturadas e inseridas em determinado local. Neste caso, poderíamos criar uma receita⁵⁸ com a seguinte ordem: se alguém publicar uma foto no Instagram perto de onde ocorre a manifestação, então capture a foto e o link e cole em uma linha de uma planilha do Excel. Para informar quais fotos devem ser capturadas, pode-se programar, pelo próprio IFTTT, o raio de quilômetro desejado, os termos que podem aparecer na publicação e as hashtags usadas no ato. Tudo isso junto pode ser coletado pela ferramenta e faz com que as fotos se tornem uma base de dados a ser analisada posteriormente pelo jornalista para elaboração da matéria.

3.1.1.3. RASPAGEM DE DADOS E NOÇÕES DE PROGRAMAÇÃO

Saber como raspar dados⁵⁹ e ter conhecimentos de programação são importantes, principalmente, no processo de compilação dos dados de uma reportagem.

A raspagem de dados é uma técnica em que um programa de computador extrai informações de uma interface feita para a leitura humana. No jornalismo, essa técnica vem sendo aplicada principalmente para extrair informações de páginas da Web e de documentos que não são organizados de forma estruturada, como PDFs. É, sem dúvidas, uma das ferramentas mais importantes que um jornalista pode ter em investigações jornalísticas que envolvem a Web e sistemas digitais [...] Diversas informações estão espalhadas em sites da Web e em documentos PDF fornecidos por agências do governo, empresas e organizações. Não só o jornalismo investigativo, mas o de serviço, o esportivo e o cultural também se beneficiam das técnicas de raspagem. A transformação dessas informações que estão espalhadas por aí num formato estruturado pode permitir descortinar possíveis irregularidades, correlações, levantar questões importantes e abrir caminhos para a produção de serviços e ferramentas úteis, educativas ou de entretenimento (PIRES, 2015)⁶⁰.

⁵⁷Disponível em: <https://ifttt.com/>. IFTTT é uma sigla em inglês que significa *If This, Than That* ou seja, “Se isso, então aquilo”.

⁵⁸ Receita é a sequência de ações que o usuário vai concatenar usando o IFTTT. As receitas são as programações dadas aos robôs para que eles atuem da forma como o usuário deseja.

⁵⁹Raspar de dados é uma maneira utilizada para extrair dados escondidos em documentos como páginas da web e PDFs e torná-los usáveis, possíveis de serem analisados.

⁶⁰ Disponível em: <https://escoladedados.org/manual/cursos/raspagem/raspagem-jornalismo-de-dados/>. Acessado em: 21/05/2017.

Percebe-se, portanto, que a técnica é fundamental para um jornalista de dados. O conhecimento dela faz com que reportagens trabalhosas sejam simplificadas e a produtividade multiplicada.

A raspagem de dados também pode ajudar jornalistas a monitorarem a prestação de contas de agências de governos que, apesar de publicarem informações na Web, ainda não o fazem de forma estruturada. É o caso da Secretaria da Segurança Pública do estado da Bahia, que divulga informações atualizadas sobre os homicídios no estado. As tabelas estão publicadas no site da secretaria, mas servem apenas para consulta humana. Em vez de tabular esses dados manualmente, o que potencialmente poderia levar a erros na contagem, um programador pode facilmente desenvolver um raspador que extrai as informações para uma planilha, de modo que os dados poderiam ser facilmente analisados e contextualizados (PIRES, 2015)⁶¹.

A primeira premissa para fazer uma raspagem de dados é observar se a fonte de onde pretende-se coletar as informações não está disponível da maneira como se deseja. Por isso, a saída está na raspagem. Um jornalista não-programador pode fazer a raspagem de três formas: com o Google Sheets⁶², com o IFTTT (descrito no subcapítulo anterior) e com outras ferramentas disponíveis na internet, como o Web Scraper⁶³.

3.1.1.3.1. RASPANDO DADOS COM O GOOGLE SHEETS

A raspagem de dados no Google Sheets é feita por meio de uma fórmula que importa a informação de uma página na internet para as células da planilha eletrônica. É necessário conexão com internet para que a importação seja feita. A fórmula para que o procedimento seja executado é a seguinte: “=import(“URL”; “table”; N)”. Como pode-se observar, esta fórmula apresenta três parâmetros⁶⁴. Os dois primeiros precisam estar entre aspas. O primeiro refere-se ao endereço da página na internet (URL). O segundo é “table” (tabela) ou “list” (lista), variando de acordo com a tag do HTML (TABLE ou LIST) com a qual as informações estão organizadas na página que será raspada. O terceiro é um número, começando do 1, que identifica a posição da tabela ou da lista na página que o usuário queira raspar; se a página tem mais de uma tabela, por exemplo, o número representará a ordem em que elas aparecem. Esta fórmula deve ser digitada na primeira célula da planilha do Google Sheets, a A1. Após digitados os parâmetros da

⁶¹Disponível em: <https://escoladedados.org/manual/cursos/raspagem/raspagem-jornalismo-de-dados/>. Acessado em: 21/05/2017.

⁶² É o programa de planilhas do Google. Disponível em: sheets.google.com.

⁶³ É uma extensão do navegador Chrome, que faz raspagens buscando elementos do HTML da página.

⁶⁴Na ciência da computação, os parâmetros se usam para mandar valores à função, com os quais ela trabalhará para realizar as ações. São os valores de entrada que recebem uma função. Neste caso, os parâmetros são os valores antes do ponto e vírgula (;).

fórmula, a importação é feita automaticamente e os dados da página serão carregados na planilha do Google Sheets.

3.1.1.3.2. RASPANDO DADOS COM O IFTTT

Assim como descrito no subcapítulo “Busca avançada pelas redes sociais”, o IFTTT é uma ferramenta que faz da raspagem de dados a forma para compilar a informação desejada. No subcapítulo citado foi descrito o processo de captura de fotos postada por usuários que foram a algum protesto. O IFTTT cria um robô virtual que tem uma tarefa automatizada: ele coleta dados de um site e cola as informações em uma linha de uma planilha do Google Sheets, por exemplo. Essa é uma técnica que acelera o processo de captação da informação.

É o caso, por exemplo, do jornalista que queira saber quantas pessoas postaram fotos ou vídeos no Instagram com a “#ForçaChape”⁶⁵ em manifesto em prol da Chapecoense; ele poderia utilizar o IFTTT para fazer essa contagem. A tarefa que deveria ser criada é a seguinte: a cada postagem no Instagram com a hashtag #ForçaChape, cole a foto e o link da postagem em uma linha do Google Sheets. Com isso, o jornalista obterá de forma ágil a quantidade de manifestos de solidariedade nos dias que se seguiram à tragédia com o avião da Chapecoense, em novembro de 2016.

3.1.1.3.3. RASPANDO DADOS COM O WEB SCRAPER

O Web Scraper é uma extensão do navegador Chrome que faz a captura de dados de forma automatizada, devendo ser configurada da maneira que se deseja a realização da raspagem. Para fazer a raspagem é necessária a instalação⁶⁶ da ferramenta no navegador. Cumprida essa etapa, um ícone de teia de aranha aparecerá ao lado da barra de endereços do navegador. Com isso, basta clicar no ícone para iniciar a programação da raspagem de dados desejada.

Para iniciar a raspagem, é preciso ir até a página da internet em que estão as informações a serem coletadas. Nela, o usuário deve clicar com o botão direito do mouse e selecionar “Inspecionar Elemento” para abrir a janela do Webinspector⁶⁷ do Chrome. No menu superior do Webinspector, passa a aparecer uma opção chamada

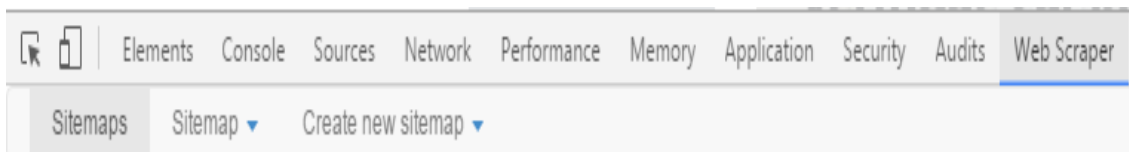
⁶⁵A hashtag ForçaChape foi a mais utilizada nos dias após o acidente com o avião da Chapecoense. A hashtag foi adotada como manifesto de solidariedade à tragédia de novembro de 2016.

⁶⁶O link para instalação da extensão do Chrome é este: <https://chrome.google.com/webstore/detail/web-scraper/jnhgnonknehpejjnehehlkklplmbmhn>.

⁶⁷É uma janela que se abre no navegador, onde se pode acessar, dentre outras coisas, o código HTML da página. Esta é uma parte familiar aos programadores e onde os jornalistas começam a ficar intimidados.

Web Scraper. O usuário deve, então, clicar na opção para que a programação seja iniciada. A imagem abaixo mostra a opção do Web Scraper no Webinspector:

Figura 2



Conforme pode-se observar na imagem, há três opções no menu do Web Scraper. O “Sitemaps” lista os “robôs” de raspagem que o usuário vai criar. O robô de raspagem é a rotina de captura de informações. Cada site raspado tem uma rotina diferente, então cada site terá um robô diferente. Na aba “Sitemap” se configura o robô. E, por fim, no “Create new sitemap” são criados os robôs.

Para começar, basta clicar em “Create new sitemap” e nomear o robô. Em seguida, deve-se colocar o endereço da página onde o robô fará a captura dos dados. Para exemplificar, serão raspados dados de uma tabela do Portal da Transparência⁶⁸ com informações sobre os gastos do governo brasileiro com entidades sem fins lucrativos em 2017. Os dados a serem raspados estão expostos na imagem abaixo:

Figura 3

CNPJ	Razão social [Nome Fantasia]	Total no Ano (R\$) *
33.641.663/0001-44	FUNDAÇÃO GETULIO VARGAS	25.045.069,47
02.485.183/0001-08	INSTITUICAO EDUCACIONAL MATOGROSSENSE- IEMAT	23.871.013,67
08.469.280/0001-93	FUNDAÇÃO NORTE RIO GRANDENSE DE PESQUISA E CULTURA [FUNPEC]	21.477.656,96
63.083.869/0001-67	INSTITUTO EDUCACIONAL DO ESTADO DE SÃO PAULO - IESP	21.422.378,65
83.476.911/0001-17	FUNDAÇÃO DE AMPARO A PESQUISA E EXTENSÃO UNIVERSITÁRIA	21.005.894,52
00.331.801/0001-30	UNIAO BRASILENSE DE EDUCACAO E CULTURA [UNIAO BRASILENSE DE EDUCACAO E CULTURA - UBEC]	19.644.837,92
03.365.403/0001-22	ASSOCIACAO IGREJA ADVENTISTA MISSIONARIA - AIAMISS [AIAMISS]	16.465.003,69
34.174.896/0001-47	INSTITUTO DE PROFESSORES PUBLICOS E PARTICULARES	16.454.268,96
17.878.554/0001-99	FUNDAÇÃO DE ENSINO E TECNOLOGIA DE ALFENAS	15.799.117,49
18.720.938/0001-41	FUNDAÇÃO DE DESENVOLVIMENTO DA PESQUISA	15.520.387,57
05.572.870/0001-59	FUNDAÇÃO DE AMPARO E DESENVOLVIMENTO DA PESQUISA [FADESP]	15.073.649,71
28.638.393/0001-82	ASSOCIACAO SALGADO DE OLIVEIRA DE EDUCACAO E CULTURA	13.739.232,80
37.174.034/0001-02	INSTITUTO EURO AMERICANO DE EDUCACAO CIENCIA TECNOLOGIA [UNEURO]	12.990.715,43
33.423.575/0001-76	CONFEDERACAO NACIONAL DO COMERCIO DE BENS, SERVICOS E TURISMO - CNC	12.862.150,95
75.517.151/0001-10	ASSOCIACAO PARANAENSE DE ENSINO E CULTURA [UNIVERSIDADE PARANAENSE - UNIPAR]	12.800.155,38

O nosso robô vai se chamar “raspagem_sem_fins_lucrativos”. O endereço da página a ser colado será a URL da página. Um detalhe importante é que, neste caso, são

⁶⁸ É um canal pelo qual o cidadão pode acompanhar a utilização dos recursos federais arrecadados com impostos no fornecimento de serviços públicos à população, além de se informar sobre outros assuntos relacionados à Administração Pública Federal. Disponível em: <http://www.portaldatransparencia.gov.br>

202 páginas a serem raspadas e é preciso informar o robô que ele deve percorrer todas as 202 páginas e não somente a primeira. É necessário informar ao Web Scraper qual é o endereço da primeira página e a progressão que vai da primeira até à última. Para isso, na hora de colar o endereço de onde o robô iniciará a raspagem, deve-se relatar a progressão de páginas. Neste exemplo, a raspagem inicia na página 1 e vai até a 202. A imagem abaixo detalha o procedimento a ser escrito para que todas as páginas sejam devidamente raspadas:

Figura 4

Selecione o(a) "Razão social" para obter o detalhamento do valor Caso queira outra classificação, clique no título da coluna correspondente

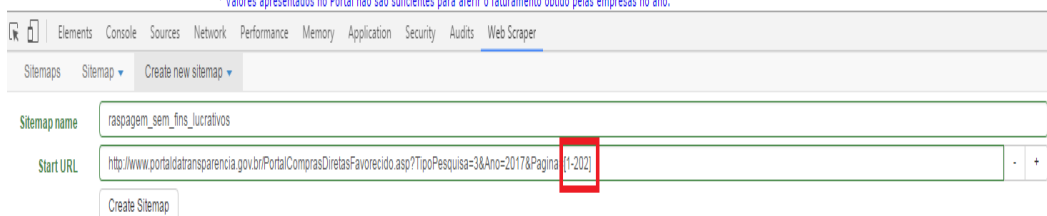
CNPJ	Razão social [Nome Fantasia]	Total no Ano (R\$) *
33.641.663/0001-44	FUNDAÇÃO GETULIO VARGAS	25.045.069,47
02.485.183/0001-08	INSTITUICAO EDUCACIONAL MATOGROSSENSE-IEEMAT	23.871.013,67
08.469.200/0001-93	FUNDAÇÃO NORTE RIO GRANDENSE DE PESQUISA E CULTURA [FUNPEC]	21.477.656,96
63.083.869/0001-67	INSTITUTO EDUCACIONAL DO ESTADO DE SAO PAULO - IESP	21.422.378,65
83.476.911/0001-17	FUNDAÇÃO DE AMPARO A PESQUISA E EXTENSÃO UNIVERSITÁRIA	21.005.894,52
00.331.801/0001-30	UNIAO BRASILENSE DE EDUCACAO E CULTURA [UNIAO BRASILENSE DE EDUCACAO E CULTURA - UBEC]	19.844.837,92
03.365.403/0001-22	ASSOCIACAO IGREJA ADVENTISTA MISSIONARIA - AIAMIS [AIAMISS]	16.465.003,69
34.174.896/0001-47	INSTITUTO DE PROFESSORES PUBLICOS E PARTICULARES	16.454.268,96
17.878.554/0001-99	FUNDAÇÃO DE ENSINO E TECNOLOGIA DE ALFENAS	15.799.117,49
18.720.938/0001-41	FUNDAÇÃO DE DESENVOLVIMENTO DA PESQUISA	15.520.387,57
05.572.870/0001-59	FUNDAÇÃO DE AMPARO E DESENVOLVIMENTO DA PESQUISA [FADESP]	15.073.649,71
28.638.393/0001-82	ASSOCIACAO SALGADO DE OLIVEIRA DE EDUCACAO E CULTURA	13.739.232,80
37.174.034/0001-02	INSTITUTO EURO AMERICANO DE EDUCACAO CIENCIA TECNOLOGIA [UNIEURO]	12.990.715,43
33.423.575/0001-76	CONFEDERACAO NACIONAL DO COMERCIO DE BENS, SERVICOS E TURISMO - CNC	12.862.150,95
75.517.151/0001-10	ASSOCIACAO PARANAENSE DE ENSINO E CULTURA [UNIVERSIDADE PARANAENSE - UNIPAR]	12.800.155,38

Pesquisar:

Página 2 / 202

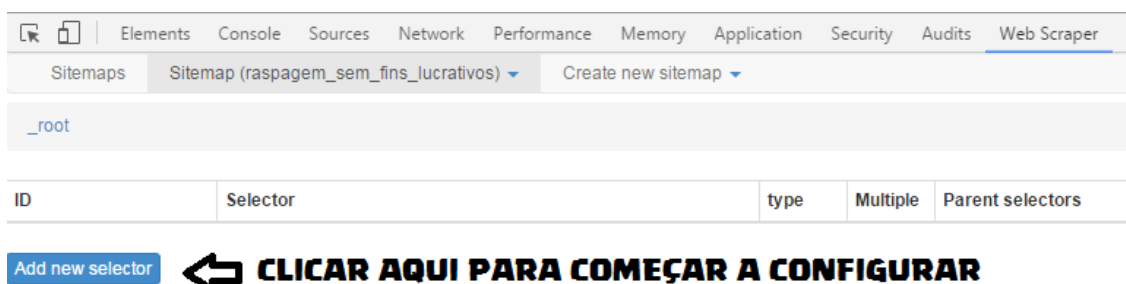
« Primeira | < Anterior | Próxima > | Última » | Página:

* Valores apresentados no Portal não são suficientes para aferir o faturamento obtido pelas empresas no ano.



Feito isso, basta clicar no botão “Create Sitemap” para criar o robô. Por enquanto ele segue sem nenhuma atribuição. No momento é apenas um robô de raspagem “vazio”, que não tem nenhuma função ou tarefa. A próxima etapa é designar uma tarefa a ele. Para isso, o usuário deve clicar no botão azul “Add new selector” para poder começar a adicionar os elementos da raspagem.

Figura 5



Após clicar neste botão (conforme exibido na imagem acima), uma nova janela aparecerá com opções de campos a serem preenchidos. No primeiro chamado “*Id*”, o usuário deve digitar “*tabela*”, uma vez que o robô vai raspar os dados em formato de tabela. O segundo “*Type*” é um campo com diversas opções a serem selecionadas. Ali, o usuário escolhe o tipo de raspagem a ser feita. Neste caso, será “*Table*” porque o robô terá a tarefa de raspar a tabela com os gastos diretos do governo com entidades sem fins lucrativos em 2017.

O terceiro campo “*Selector*” é onde se informa ao robô o local do código HTML da página que ele deve coletar as informações. Para isso, é preciso clicar em “*Select*” e depois no topo da tabela a ser raspada. Após a ação, a tabela será destacada, automaticamente, em vermelho. Conforme pode-se observar na imagem abaixo:

Figura 6

GASTOS DIRETOS POR FAVORECIDO (ENTIDADES SEM FINS LUCRATIVOS)

Total destinado pelo Governo Federal em âmbito nacional em 2017 - Aplicações Diretas: R\$ 526.909.703.038,11
 Total destinado pelo Governo Federal em âmbito nacional em 2017 - Entidades sem fins lucrativos: R\$ 2.687.200.406,98

CLICAR AQUI PARA MOSTRAR AO ROBÔ QUAL É A TABELA

CNPJ	Razão social [Nome Fantasia]	Total no Ano (R\$) *
33.641.663/0001-44	FUNDAÇÃO GETULIO VARGAS	25.045.069,47
02.485.183/0001-08	INSTITUICAO EDUCACIONAL MATOGROSSENSE-IEMAT	23.871.013,67
08.469.280/0001-93	FUNDAÇÃO NORTE RIO GRANDENSE DE PESQUISA E CULTURA [FUNREC]	21.477.656,96
63.083.869/0001-67	INSTITUTO EDUCACIONAL DO ESTADO DE SAO PAULO - IESP	21.422.378,65
83.476.911/0001-17	FUNDAÇÃO DE AMPARO A PESQUISA E EXTENSÃO UNIVERSITÁRIA	21.005.894,52
00.331.801/0001-30	UNIAO BRASILENSE DE EDUCACAO E CULTURA [UNIAO BRASILENSE DE EDUCACAO E CULTURA - UBEC]	19.844.837,92
03.365.403/0001-22	ASSOCIACAO IGREJA ADVENTISTA MISSIONARIA - AIAMISS	16.465.003,69
34.174.896/0001-47	INSTITUTO DE PROFESSORES PUBLICOS E PARTICULARES	16.454.268,96
17.878.554/0001-99	FUNDAÇÃO DE ENSINO E TECNOLOGIA DE ALFENAS	15.799.117,49
18.720.938/0001-41	FUNDAÇÃO DE DESENVOLVIMENTO DA PESQUISA	15.520.387,57
05.572.870/0001-59	FUNDAÇÃO DE AMPARO E DESENVOLVIMENTO DA PESQUISA [FADESP]	15.073.649,71
28.638.393/0001-82	ASSOCIACAO SALGADO DE OLIVEIRA DE EDUCACAO E CULTURA	13.739.232,80
37.174.034/0001-02	INSTITUTO EURO AMERICANO DE EDUCACAO CIENCIA TECNOLOGIA [IUNEURO]	12.990.715,43
33.423.675/0001-76	CONFEDERACAO NACIONAL DO COMERCIO DE BENS, SERVICOS E TURISMO - CNC	12.862.150,95
75.517.151/0001-10	ASSOCIACAO PARANAENSE DE ENSINO E CULTURA [UNIVERSIDADE PARANAENSE - UNIPAR]	12.800.155,38

Pesquisar:

Página 2/202

APÓS A IDENTIFICAÇÃO CORRETA DA TABELA, CLICAR NO BOTÃO AZUL

div#listagem table S P C Done selecting!

Elements Console Sources Network Performance Memory Application Security Audits Web Scraper

Sitemaps Sitemap (raspagem_sem_fins_lucrativos) Create new sitemap

Id

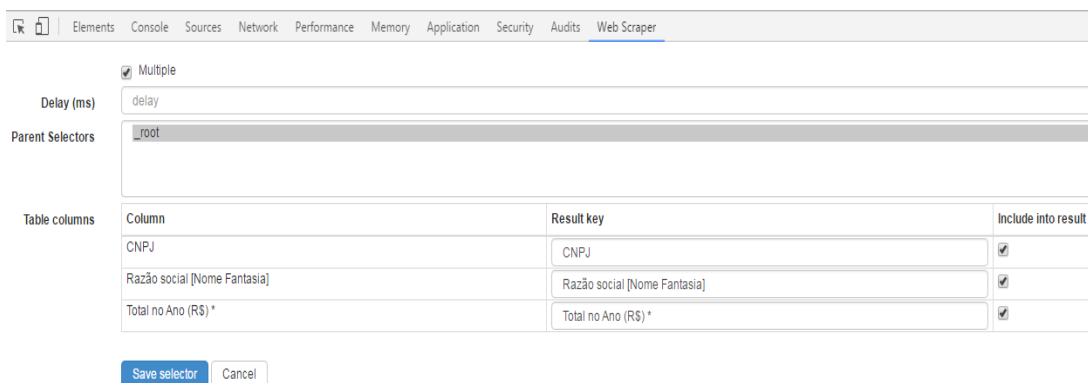
Type

Selector

Após a correta identificação da tabela, o usuário deve clicar no botão azul “*Doneselecting!*” para que a referência da tabela na página seja salva às configurações do robô “*raspagem_sem_fins_lucrativos*”. Como são muitos campos a serem raspados, deve-se selecionar o quadradinho “*Multiple*” para que o robô pegue todas as informações selecionadas da tabela.

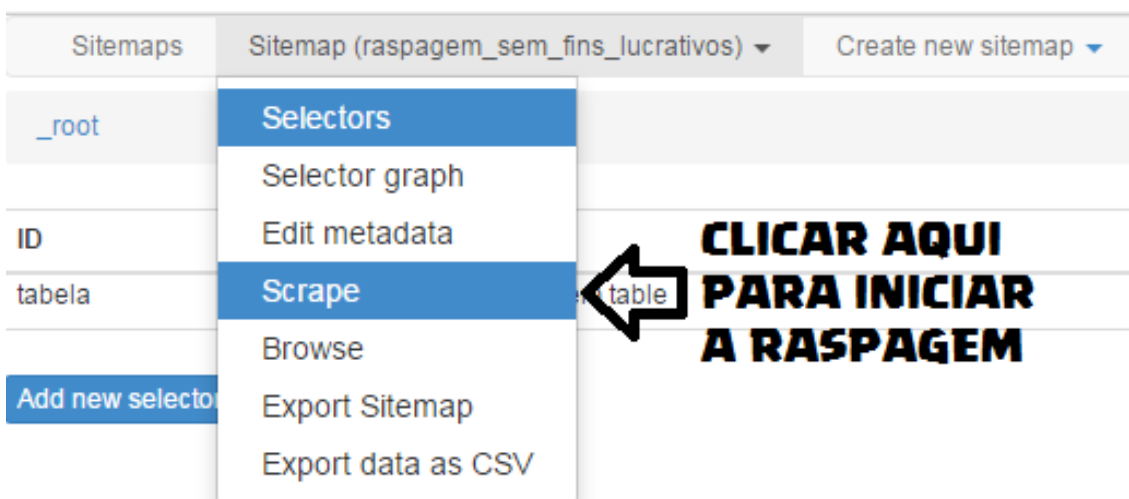
Cumprida esta etapa, já passa a ser exibido na parte mais inferior da janela do Web Scraper quais serão as colunas da tabela. Após a verificação das informações das colunas, o usuário precisa clicar no botão azul “*Save selector*” para finalizar a configuração das ações do “*raspagem_sem_fins_lucrativos*”. O detalhamento final desta parte está detalhado na imagem abaixo:

Figura 7



Feito isso, o robô já está configurado e pronto para iniciar a raspagem. Para fazê-la, basta clicar na aba “*Sitemap(nomedorobô)*”, neste caso, em “*Sitemap(raspagem_sem_fins_lucrativos)*”. Quando o usuário clicar na aba, algumas opções serão expostas e ele deve, então, selecionar a opção “*Scrape*” para iniciar os trabalhos do robô. A imagem a seguir detalha isto:

Figura 8

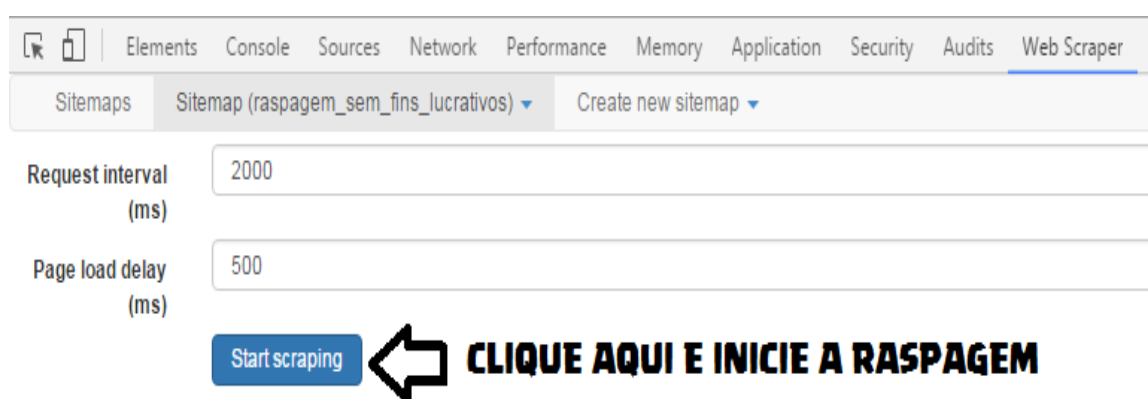


Antes do início de fato da raspagem, o usuário precisa configurar o tempo que o robô deve esperar para mudar de página em microssegundos. O padrão são 2000 microssegundos, equivalente a dois segundos. A delimitação do tempo de raspagem é

importante porque se for feita de maneira muito acelerada, pode prejudicar o funcionamento do servidor da página, que pode ficar sobrecarregada.

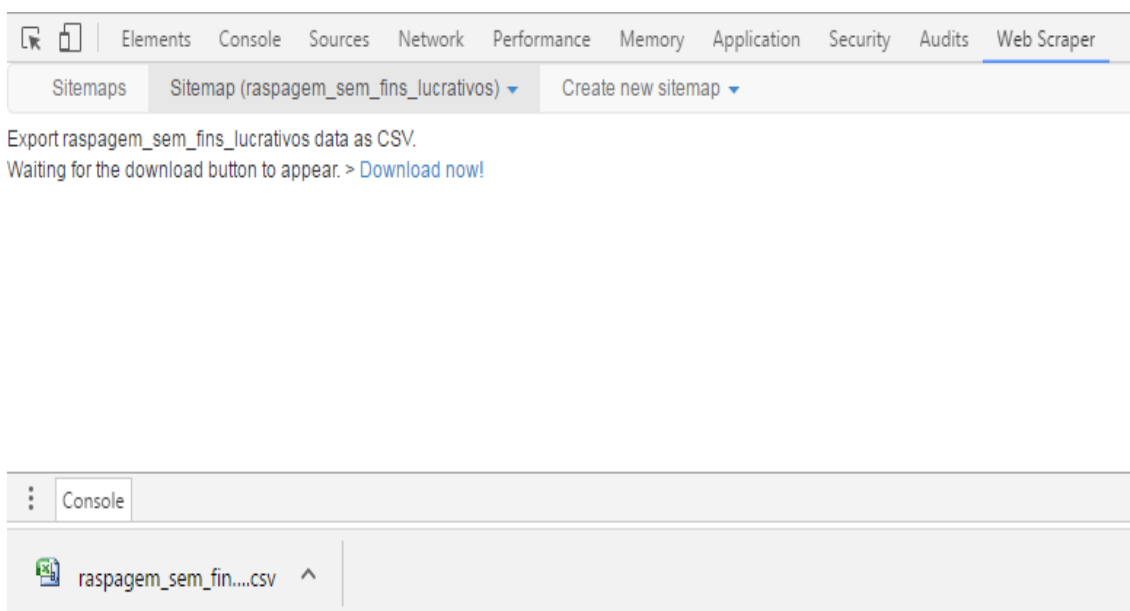
Além disso, é preciso configurar quanto tempo o robô deve esperar antes de raspar os dados da página. Ou seja, ele tem que esperar a página carregar para, então, coletar os dados ali expostos. Isso vai depender da conexão de internet do usuário com a conexão do servidor da página raspada. O padrão são 500 microssegundos, o que vale meio segundo. Se for preciso, o usuário deve aumentar esse tempo de espera do carregamento de informações da página.

Figura 9



Após a configuração descrita acima, o usuário deve clicar no botão azul “*Start scraping*” (conforme exposto na imagem acima) para, enfim, raspar os dados informados ao robô da página. O tempo de execução da tarefa designada vai variar de acordo com a quantidade de informações a serem capturadas pelo robô. Neste caso, eram 202 páginas. O “raspagem_sem_fins_lucrativos” levou 11 minutos e 32 segundos para raspar todos os dados das páginas. Por fim, para obter os dados raspados, o usuário pode exportar os dados para um arquivo CSV⁶⁹. Após selecionar a opção “*Export data as CSV*” na aba “Sitemap(raspagem_sem_fins_lucrativos)”, uma opção de “Download now!” aparecerá. Basta clicar ali para obter, por fim, os dados raspados.

⁶⁹ CSV (Comma-Separated Values) é formato simples de armazenamento, que agrupa as informações de arquivos de texto em planilhas. Os dados são separados por um delimitador, que usa a vírgula e a quebra de linha para separar os valores. O formato também usa as aspas em campos nos quais são usados os caracteres reservados (vírgula e quebra de linha). Cada linha em um texto CSV representa uma linha em uma planilha. Cada célula é geralmente separada por vírgula ou um outro caractere, como tabulador.

Figura 10

Com as informações colhidas, percebem-se os benefícios da raspagem de dados. Se o usuário optasse por fazer a coleta das informações manualmente, ele certamente levaria algumas horas para finalizar o trabalho. Um jornalista, com habilidades de raspagem de dados, pode resolver essa tarefa em alguns minutos. Neste caso exemplificado neste subcapítulo, configurar o robô e esperar até que ele finalizasse a tarefa levou cerca de 20 minutos, tempo mínimo se comparado às horas de trabalho caso o repórter optasse por fazer a coleta manualmente. Logo, a produtividade aumenta consideravelmente quando o jornalista detém a técnica da raspagem. Por isso, é importante para o jornalista de dados saber como programar um robô para fazer este tipo de trabalho mecânico.

3.1.1.3.4. POR QUE É BOM SABER PROGRAMAR?

Com os benefícios da raspagem descritos, entende-se que a iniciação do jornalista na programação passa pelas técnicas de raspagem. Mas ele pode ir além. Se saber raspar dados já facilita o trabalho, saber programar coloca o polivalente profissional com amplas possibilidades de melhora no trabalho e manipulação de dados cotidianamente. Scott Klein (2014) dá dicas de como começar a aprender a programar. Para o autor, o mais importante, quando começar a aprender a programar, é buscar um projeto no qual o repórter já esteja comprometido a finalizar.

Escolha algo que irá te decepcionar ou o seu patrão, caso você não o termine. Há duas razões para isso. Primeiramente, você aprenderá

melhor se você atribuir o conhecimento novo ao velho. Observar um código por meio da lente de um problema que você sabe como resolver é uma forma inestimável para compreendê-lo e memorizá-lo. Em segundo lugar, isso vai te dar o impulso para escalar as partes mais íngremes da curva de aprendizagem. Você precisa ter mais medo de perder o seu prazo do que aprender programação (KLEIN, 2014)⁷⁰.

Portanto, apesar de ser um universo diferente do que as pessoas estão acostumadas, o processo de imersão no mundo da programação não precisa ser traumático. O importante é observar os benefícios que o conhecimento adquirido trará para a própria produção dentro do trabalho jornalístico diário. Scott Klein (2014) ainda dá mais outros conselhos aos aspirantes a programadores. Ele diz que o próximo passo é escolher uma linguagem para aprender. No universo da ciência da computação, há algumas linguagens diferentes que têm diferentes funcionalidades. O jornalista deve procurar saber qual se adequa melhor a sua rotina de trabalho. JavaScript, Python, Ruby, R, e C++ são algumas das linguagens de programação mais conhecidas e usadas atualmente, em plena segunda década do século 21. Cabe ao interessado, descobrir qual pode ser mais útil a ele.

Vá para uma livraria e folheie o primeiro capítulo de um livro sobre Python e um livro sobre Ruby ou navegue na internet para introduções às linguagens. Existem outras linguagens, mas estas duas são excelentes. Escolha alguma linguagem que você se interesse. [...] Se você não pode decidir basta jogar uma moeda. [...] E elas são como linguagens românticas: se você é bom em uma você poderá aprender a outra rapidamente (KLEIN, 2014)⁷¹.

Para um jornalista acostumado com texto, observar um código pode soar aterrorizante a princípio, mas pode ser a solução e a alternativa que faltava para finalizar reportagens intermináveis e acelerar todo o fluxo do trabalho. O jornalista-programador Sérgio Spagnuolo (2017) compartilhou, em entrevista a este trabalho, como foi a sua experiência de imersão no mundo da programação. Spagnuolo é fundador do Volt Data

⁷⁰ Disponível em: <https://www.propublica.org/nerds/item/how-to-start-learning-how-to-program>.

Acessado em: 24/05/2017. Tradução do autor. No original: “Pick something that will disappoint you or your employer if you don't finish it. There are two reasons for this. First, you'll learn best if you attach new knowledge to old. Seeing code through the lens of a problem you know how to solve is an invaluable way to understand it and remember it. Second, it will give you the momentum to scale the steeper parts of the learning curve. You need to be more afraid of missing your deadline than you are of programming”.

⁷¹ Disponível em: <https://www.propublica.org/nerds/item/how-to-start-learning-how-to-program>.

Acessado em: 24/05/2017. Tradução do autor. No original: “Go to a bookstore and flip through the first chapter of both a Python book and a Ruby book or browse the web for introductions to the languages. There are other languages but these two are excellent. Pick whichever language delights you more. [...] If you can't decide, simply flip a coin. [...] And they're like Romance languages: If you're good at one you'll be able to learn the other pretty quickly”.

Lab⁷², uma agência de jornalismo de dados que produz – em uma ponta – pesquisas, investigações, levantamentos e reportagens para redações e projetos de mídia, e – em outra – inteligência para comunicação de ONGs, terceiro setor e até algumas empresas. O jornalista contou que aprendeu “na unha” a programar, ou seja, correu atrás sozinho para conseguir adquirir as habilidades da programação.

Buscava tutoriais online, via exemplos em funcionamento e buscava replicar até funcionar para o que precisava. Eventualmente, a repetição leva ao conhecimento, e foi o que aconteceu. Sou jornalista primeiro, não programador, e ainda tenho muito a aprender. Mas aprender com exemplos é a melhor coisa. Sites como Stock Overflow e blogs com tutoriais específicos são uma benção para quem quer aprender a programar. Cursos e workshops são excelentes para começar, mas a prática é a melhor forma de fazer, aprender e desenvolver habilidades de jornalismo de dados (SPAGNUOLO, 2017)⁷³.

Segundo Sérgio Spagnuolo, a programação permite que o jornalista consiga analisar, visualizar e obter uma quantidade de dados muito maior do que se fosse “à mão”, em termos de volume. “No Volt, as linguagens que mais usamos são R (para análise), SQL (para query de dados), JavaScript (para visualização). Ainda estamos desenvolvendo melhor nossas capacidades em Python para extração de dados.” (SPAGNUOLO, 2017).

3.1.1.4. CONVERSÃO DE DADOS FECHADOS PARA ABERTOS

Os dados fechados são um obstáculo no trabalho dos jornalistas de dados. Mas, são apenas um obstáculo, não um muro inteiro capaz de impedir totalmente o manejo da informação ali inserida. Há algumas ferramentas disponíveis na internet que permitem a conversão dos dados em formato fechado para o formato aberto. Considera-se um formato fechado arquivos que não permitem a manejo da informação ali contida, tais como pdf ou jpeg. Já os formatos abertos são aqueles nos quais os jornalistas podem trabalhar com a informação da maneira como desejam.

É comum órgãos governamentais oferecerem documentos em formato pdf aos jornalistas. Normalmente, as informações vêm bonitas e organizadas a olho nu. Mas, quando trabalhadas em um computador, elas são pouco úteis, já que não podem ser trabalhadas. Por exemplo, se a instituição responsável pelo censo demográfico de um país divulga as informações em tabelas organizadas e “bonitas” (com cores, marcações,

⁷² Disponível em: <http://www.voltdata.info/>

⁷³ Entrevista concedida ao autor.

imagens etc) no formato pdf para que os jornalistas as noticiem, seria uma missão complicada de ser cumprida. Isso porque os jornalistas não conseguiriam visualizar de maneira correta os principais ganchos do conteúdo apresentado. É importante que os dados possam ser cruzados entre si para que ângulos diferentes surjam para serem noticiados. A instituição pode apresentar um ponto de vista no conteúdo divulgado, mas os próprios dados podem fornecer outros pontos e, muitas vezes, ainda mais esclarecedores. E, por isso, nestes casos, os jornalistas devem converter o pdf para .xls (formato do Excel⁷⁴) ou csv (formato acessível para programas de planilhas ou de banco de dados) para trabalhar com os dados.

Nesta etapa de conversão de formatos, há alguns sites e ferramentas que auxiliam os jornalistas na tarefa. Normalmente, para fazer a conversão de formato só é preciso informar um email para receber o arquivo convertido. O site Zamzar⁷⁵ é um dos mais usados por jornalistas por permitir a conversão de diferentes formatos de forma rápida e gratuita. No site, pode-se transformar arquivos PDF em Word, por exemplo. Há outras ferramentas e sites disponíveis na internet que podem auxiliar o repórter que necessita converter o formato de um arquivo para a execução de uma reportagem. Outro site interessante é o Free-OCR⁷⁶ que executa a tarefa de reconhecimento ótico de caracteres. Isso é importante, por exemplo, em documentos escaneados ou fotografados que são enviados aos jornalistas em formato jpeg ou png. Caso o repórter queira usar um trecho do documento, ele teria de escrever toda a parte que ele deseja usar, o que acarreta perda de tempo. Caso o formato viesse em um documento do Microsoft Word, o jornalista poderia manejar a informação de forma muito mais rápida. Com o Free-OCR, as informações contidas na foto são lidas e transformadas em formato de texto passível de edição, o que facilita a vida do jornalista, que, caso contrário, teria que transcrever as partes do documento que fossem usadas.

Estes são dois sites úteis no trabalho da conversão de formatos de arquivos. No entanto, há uma infinidade de outros websites para esta finalidade disponíveis na rede. Os interessados devem sempre fazer pesquisas nos buscadores da internet para descobrir que sites ou ferramentas conseguem executar a conversão que o usuário almeja obter.

⁷⁴ É o programa de planilhas eletrônicas da Microsoft.

⁷⁵ Disponível em: www.zamzar.com

⁷⁶ Disponível em: <http://www.free-ocr.com>

3.1.2. LIMPAR

Após compilar e estruturar toda a base de dados necessária para a reportagem guiada por dados, há um trabalho muito importante a ser feito em seguida: a limpeza dos dados. Ela garante que não haja nenhuma incongruência nos dados e que todas as pequenas falhas da base sejam excluídas ou consertadas. Por exemplo, uma base que contém três nomes diferentes para um mesmo órgão está incorreta e pode prejudicar a análise posterior. Se IBGE está escrito em três grafias diferentes – IBGE, I.B.G.E. e Instituto Brasileiro de Geografia e Estatística –, consequentemente, a base de dados apresenta um problema. É necessário um padrão para que os resultados sejam harmonizados. Caso a limpeza desses nomes não seja feita, a reportagem corre o risco de apresentar erros por conta de falha na análise dos dados. Paul Bradshaw (2010) aponta alguns cuidados a serem tomados nesta etapa.

[...] procure por nomes diferentes para a mesma coisa, erros de ortografia e pontuação, campos mal formatados (por exemplo, as datas que são formatadas como texto), dados inseridos incorretamente e informações que estão faltando por completo (BRADSHAW, 2010)⁷⁷.

Sandra Crucianelli (2013) complementa dizendo que não importa qual seja o procedência dos dados, é provável que sempre haja necessidade de uma limpeza para que eles sejam úteis e utilizáveis.

Qualquer que seja a procedência dos seus dados, é provável que precise de uma 'limpeza' para que sejam úteis. Isto pode implicar muitas coisas, desde a reorganização e eliminação de dados que não necessite, até o uso de ferramentas como OpenRefine (anteriormente Google Refine) para torná-los mais consistentes.[...] É possível ordenar os dados de menor para maior ou por localização (CRUCIANELLI, 2013)⁷⁸.

E para ajudar nesta árdua tarefa de limpar bases de dados, há algumas ferramentas que os jornalistas de dados devem ter conhecimento para que as usem quando necessário. A primeira é o OpenRefine⁷⁹ (também conhecido como Google

⁷⁷ Disponível em: <https://www.theguardian.com/news/datablog/2010/oct/01/data-journalism-how-to-guide>. Acessado em: 24/05/2017. Tradução do autor. No original: “[...] look out for different names for the same thing, spelling and punctuation errors, poorly formatted fields (e.g. dates that are formatted as text), incorrectly entered data and information that is missing entirely”.

⁷⁸ Disponível em: <http://www.scidev.net/america-latina/periodismo/guia-practica/periodismo-de-datos-c-mo-hacer-historias-con-n-meros.html>. Acessado em: 21/05/2017. Tradução do autor. No original: “Cualquiera sea la procedencia de sus datos, es probable que necesiten una ‘limpieza’ para que le Sean útiles. Esto puede implicar muchas cosas, desde la reorganización y eliminación de datos que no necesite, hasta usar herramientas como OpenRefine (anteriormente Google Refine) para hacerlos más consistentes [...]. Es posible ordenar los datos de menor a mayor o por ubicación. Podría buscar promedios o juntar o comparar dos conjuntos de datos”.

⁷⁹ Disponível em: openrefine.org

Refine): um aplicativo, que deve ser baixado para o desktop da máquina do usuário, que facilita o tratamento de dados. Há diversos tutoriais na internet sobre as funcionalidades do OpenRefine; então, os interessados devem procurar pelos tutoriais que mostram passo a passo como executar diversas tarefas por meio da ferramenta. Entre outras funções, o Refine busca semelhanças entre os dados e sugere agrupamentos, facilitando a padronização de dados mal formatados. Ou seja, aquele exemplo do IBGE, citado no primeiro parágrafo deste subcapítulo, seria facilmente identificado e rapidamente corrigido por meio de um recurso do programa para homogeneizar a base de dados. O programa oferece ainda inúmeras possibilidades de encadeamento nas buscas, auxilia no manuseio de planilhas que possam conter erros humanos e/ou de sistemas e permite refazer as ações em caso de erro.

Outra ferramenta interessante para auxiliar o jornalista na limpeza da base de dados é o Data Wrangler⁸⁰: um aplicativo gratuito que funciona no próprio navegador e foi desenvolvido por uma equipe de pesquisadores da Universidade Stanford. A interface é intuitiva, o que facilita a limpeza dos dados pela ferramenta. Assim como o OpenRefine, há diversos tutoriais na internet que exibem as funcionalidades do programa. Um exemplo é que, com apenas um clique, o usuário consegue dividir uma coluna com nomes e sobrenomes em duas diferentes. O Data Wrangler funciona reconhecendo padrões na tabela. Com o reconhecimento dos padrões, as transformações são feitas automaticamente. No entanto, o principal inconveniente do software é que ele só trabalha com bases de dados de pequeno porte, até 1000 linhas e 40 colunas.

Estes são dois programas interessantes disponíveis na internet quando o assunto é limpeza de base de dados. No entanto, não são os únicos. Apesar de o Excel e o Google Sheets serem voltados para o processamento de planilhas, também é possível fazer limpezas rápidas usando algumas funções, tais como exclusão de linhas ou colunas desnecessárias ou formatação automática de células da planilha. O intuito deste trabalho é demonstrar quais são os principais softwares e os mais utilizados, além de sublinhar a importância da limpeza da base de dados no contexto de execução de uma reportagem guiada por dados. E também informar que, para isso, há alguns programas que auxiliam o repórter nesta empreitada.

⁸⁰ Disponível em: <http://vis.stanford.edu/wrangler/>

3.1.3. CONTEXTUALIZAR

Nesta etapa, o jornalista tem de ser, primeiramente, um bom contador de histórias. Apesar de a fonte das reportagens guiadas por dados serem os próprios dados, eles também podem contar e revelar excelentes histórias assim como as fontes humanas. Então, o repórter tem que saber com que base de dados está lidando para, assim, tirar os melhores ângulos possíveis dali. Além disso, o repórter deve identificar a metodologia e os critérios usados na formulação de uma base de dados para que a transmissão do relato seja a melhor possível. Sublinhar a fonte de onde vem a informação também é importante para transmitir confiabilidade nos dados informados na reportagem.

É importante também que o jornalista de dados tenha uma escrita apurada e clara. Como a fonte primordial da história são os dados, é necessária muita coesão e objetividade na transmissão da informação para que o leitor não se confunda com os números apresentados. A mistura de muitos dados diferentes pode confundir o leitor e atrapalhar a compreensão da mensagem passada pela reportagem. Portanto, é fundamental que o repórter seja cuidadoso na escolha de como vai transmitir a mensagem da reportagem para o público para que os dados não fiquem descontextualizados e a informação não seja bem captada pelos leitores.

Por exemplo, se o indivíduo tem em mãos uma base com dados da Unicef⁸¹ sobre a fome entre as crianças do mundo, o jornalista que for reportar a história deve tomar cuidado na contextualização destas informações. Se esta base de dados diz que 100 milhões de crianças africanas são subnutridas, o repórter não pode apenas noticiar o número sem antes entender o contexto dos dados ali expostos. Caso as nações do norte do continente sejam as mais afetadas – com 90% das crianças subnutridas ali –, cabe ao jornalista informar o leitor sobre isso e não generalizar um número para a África toda. Para saber se 90 milhões é um número realmente alarmante, é preciso contextualizar os dados. Ou seja, verificar a quantidade de crianças nestas nações do norte do continente. Hipoteticamente, este número representa quase que a totalidade de crianças dos países afetados. Logo, o jornalista já pode destacar que quase todas as crianças das nações do norte da África são subnutridas.

⁸¹ O Fundo das Nações Unidas para a Infância (Unicef) é um órgão das Nações Unidas (ONU) que tem como objetivo promover a defesa dos direitos das crianças, ajudar a dar resposta às suas necessidades e contribuir para o seu desenvolvimento.

3.1.4. COMBINAR

A etapa de combinar é crucial porque é nela que o repórter vai formular as perguntas que vão guiar a reportagem. É preciso então questionar os dados, entrevistá-los, de fato. Assim como uma boa entrevista tem a possibilidade de revelar uma ótima matéria, uma entrevista cuidadosa e minuciosa sobre a base de dados é crucial para conseguir produzir uma reportagem guiada por dados de excelência.

Dê aos dados tratamento de 'fonte': pergunte-lhes como se fosse o seu público. E faça muitas perguntas: a resposta pode não ser a primeira que vem a sua cabeça. Por exemplo, uma planilha de retrações de uma revista poderia sugerir um aumento na detecção de fraudes, mas terão que ser feitas perguntas se existem outras interpretações. Pense cuidadosamente sobre os seus resultados, soam plausíveis? É melhor verificar e reavaliar os cálculos. Não arruíne a sua reputação por conta de um erro básico. Você pode consolidar suas descobertas ou formular novas perguntas com uma análise estatística simples. Por exemplo, você poderia localizar mais tempestades catastróficas em seu país a cada ano durante 20 anos. Mas é um resultado significativo ou poderia ser a probabilidade de uma variação natural? Ferramentas como o R-Project e RStudio pode ajudá-lo a julgá-lo. Você também pode querer verificar suas descobertas com peritos ou outros jornalistas de dados experientes, especialmente quando você está começando (CRUCIANELLI, 2013)⁸².

Portanto, a análise tem que ser minuciosa. Como Sandra Crucianelli (2013) disse acima, um erro pode arruinar a reputação de um repórter. É necessária atenção extrema porque um número expressado de forma incorreta pode ser motivo de dor de cabeça ao repórter que revelou, imprudentemente, tal dado. O jornalista de dados deve ter como mantra a checagem e a re Checagem dos dados; não é possível confiar cegamente em uma base de dados que não foi inspecionada previamente pelo repórter.

Comece verificando a matemática dos representantes, relatórios oficiais e orçamentos. Procure por números implausíveis e figuras que possam ter sido distorcidas para serem mais bem vistas. Não assuma que a pessoa que preparou a documentação tinha boas habilidades matemáticas. E não suponha que a pessoa esteja acima da

⁸² Disponível em: <http://www.scidev.net/america-latina/periodismo/guia-practica/periodismo-de-datos-c-mo-hacer-historias-con-n-meros.html>. Acessado em: 21/05/2017. Tradução do autor. No original: “Dele a los datos tratamiento de ‘fuente’: fórmúlese preguntas como si lo hiciera a su audiencia. Y haga montones de preguntas: la respuesta podría no ser la primera que se le viene a la cabeza. Por ejemplo, una hoja de cálculo de las retractaciones de una revista podría sugerir un aumento em la detección de fraudes, pero tendrá que preguntarse si caben otras interpretaciones. Piense con sumo cuidado sobre sus resultados, ¿suenan plausibles? Es mejor verificar y volver a comprobarlos cálculos. No arruíne su reputación por un error básico. Puede consolidar sus conclusiones o plantearse nuevas preguntas con un simple análisis estadístico. Por ejemplo, podría ubicar más tormentas catastróficas em su país cada año durante 20 años. Pero ¿es un resultado significativo o podría ser la probabilidad de una variación natural? Herramientas como R-Project y RStudio pueden ayudarle a juzgarlo. También podría querer comprobar sus conclusiones con expertos u otros experimentados periodistas de datos, especialmente cuando está comenzando”.

manipulação dos números para tornar seu caso mais forte. Em outras palavras, entreviste os números com o mesmo cuidado que você entrevista as pessoas (WICKHAM, 2003: 18)⁸³.

No entanto, apesar de a tarefa de minimizar os erros ser complicada, há algumas ferramentas disponíveis que auxiliam o repórter a conseguir checar as informações para que tudo esteja correto. Este trabalho indicará duas delas, mas vale ressaltar que há muitas outras espalhadas pela internet. Os interessados devem usar os fóruns de discussão, debates de congressos do tema e os próprios navegadores da rede como fonte para descobrir e avaliar programas de análise estatística.

Um deles é o Excel, que é um excelente programa para fazer análise de dados e, por isso, um fiel escudeiro do repórter. Com ele, dados repetidos ou incorretos podem ser identificados rapidamente e, assim, corrigidos. Uma função que auxilia a análise e combinação de dados é o recurso “Filtro” do programa. Em uma base de dados extensa, o filtro pode ajudar o repórter a fazer combinações mais facilmente e, conseqüentemente, chegar a novos ângulos para os dados. O jornalista deve, então, sempre buscar correlações entre eles para agregar valor aos resultados apresentados.

Além deste recurso, há um outro muito importante chamado “Tabela dinâmica”. A Tabela dinâmica é um recurso disponível no Excel que permite uma análise dos dados de forma mais rápida. Com poucos cliques, o recurso permite uma exploração e análise dos dados de forma simples e precisa. Para criá-la, é preciso que a base de dados esteja organizada em formato tabular e sem nenhuma linha ou coluna vazia. Além disso, os tipos de dados devem ser os mesmos, portanto o usuário não pode misturar datas e textos em uma mesma coluna, por exemplo. Aos interessados, existem muitos tutoriais disponíveis na internet que mostram o passo a passo de criação e utilização da Tabela dinâmica do Excel.

Outro programa interessante para análise estatística de dados é o software Statwing⁸⁴. A interface da ferramenta é simples e permite que o usuário construa e visualize relações entre diferentes variáveis. Para isso, é necessário carregar o arquivo

⁸³ Disponível em: <https://books.google.com.br/books?hl=pt-BR&lr=&id=RYtYmMD2ReAC&oi=fnd&pg=PA10&dq=Math+for+journalists&ots=kAXGE-S4-2&sig=zVVv3mwx5FEJkirilO6jYUQd2s4#v=onepage&q&f=false>. Acessado em: 27/05/2017. Tradução do autor. No original: “Start by checking the math of speakers, official reports and budgets. Look for implausible numbers and figures that may have been tweaked to look better. Don’t assume the person who prepared the documentation had good math skills. And don’t assume that person is above manipulating the numbers to make his or her case stronger. In other words, interview the numbers with the same care that you interview people”.

⁸⁴ Disponível em: <https://www.statwing.com/>

com os dados, marcar as variáveis que se deseja explorar e verificar suas respectivas relações no formato de gráfico escolhido. Com o software, o cruzamento de informações é facilitado e, por meio dele, descobre-se se há relação estatística significativa entre duas variáveis. São vários recursos disponíveis e o próprio site do software apresenta vídeos tutoriais com exemplificações dos recursos do programa.

3.1.5. COMUNICAR

“O jornalismo de dados lida com um grande volume de dados. Sem uma visualização adequada para acompanhar a análise desses dados, seria impossível visualizar de maneira simples as variáveis apontadas em um estudo” (CRUCIANELLI, 2013: 102)⁸⁵. Conforme resumiu Sandra Crucianelli (2013), saber como comunicar um dado é fundamental para a compreensão da mensagem pelo leitor. O jornalista de dados deve, portanto, ter domínio sobre as técnicas e ferramentas de visualização para que a captação e o entendimento dos dados sejam facilitados conforme a apresentação dos mesmos. Além da boa apresentação dos dados, o repórter tem de fazer com que o público entenda o que eles significam.

Assim como um líde bem escrito que captura a atenção do leitor, uma visualização de dados deve ser atrativa para seduzir o público para ler a reportagem. No entanto, as reportagens guiadas por dados são peças jornalísticas e, por isso, não devem perder sua essência, que é “vender” ao leitor o que mais de importante há em uma base de dados.

Minha última recomendação é se concentrar no que mais importa, o mais relevante. Muitos designers visuais se concentram em gráficos triviais em vez de se concentrar em coisas mais importantes, como a desigualdade, saúde, acesso à água potável, etc. Existem algumas visualizações que mostram gráficos muito bem feitos, porém triviais, sem muita importância. (CAIRO, 2013)⁸⁶.

Logo, a visualização de dados não é um meio para enfeitar a reportagem e sim uma ferramenta pela qual o jornalista deve se apropriar para facilitar a compreensão dos dados informados na matéria. A visualização escolhida deve ter qualidade informativa

⁸⁵ Disponível em: <https://knightcenter.utexas.edu/books/HDPP.pdf>. Acessado em: 27/05/2017. Tradução do autor. No original: “El periodismo de datos maneja un volumen grande de datos. Sin una adecuada visualización que acompañe el análisis de esos datos, sería imposible visualizar de manera sencilla las variables que se ponen en estudio”.

⁸⁶ Disponível em: <http://manual.periodismodedatos.org/alberto-cairo.php>. Acessado em: 28/05/2017. Tradução do autor. No original: “Mi última recomendación es enfocarse en aquello que más importe, lo más relevante. Muchos diseñadores visuales se enfocan en gráficas triviales en lugar de enfocarse en cosas de más importancia, como inequidad, salud, acceso al agua potable, etc. Existen algunas visualizaciones que muestran gráficos muy bien hechos pero triviales, sin mucha importancia”.

própria, pois isto auxilia a compreensão da mensagem transmitida na reportagem. “O primeiro objetivo de um infográfico não é ser bonito apenas por causa do apelo visual, mas, acima de tudo, ser compreensível primeiro, e bonito depois disso; ou ser bonito graças à sua funcionalidade requintada” (CAIRO, 2013)⁸⁷.

Cabe, além disso, explicar a diferença entre um infográfico e uma visualização de dados. O primeiro tem enfoque na narrativa construída a partir dos dados, oferecendo, assim, um caminho específico para o público entender o que foi recortado como importante pelo repórter. Já em uma visualização de dados, geralmente não há um componente narrativo, mas um componente exploratório, no qual o público pode navegar pelos dados quantitativos apresentados. A ênfase está na apresentação de dados complexos para torná-los mais facilmente exploráveis para que o público extraia suas próprias histórias deles.

A visualização procura revelar evidências, permitir análises pelo público, que, com isso, consegue extrair informação e significado por trás de uma base de dados complexa. Por isso, as decisões ao criar uma visualização devem passar pelos princípios de design e da psicologia cognitiva porque são fundamentais na percepção da mensagem pelo leitor. O jornalista de dados deve ter conhecimento de quatro princípios do design para o momento da elaboração de uma visualização ou infografia: simplicidade (usar apenas a informação necessária para a reportagem), unidade (harmonia da composição como um todo, cada parte deve trabalhar junta como um tema visual completo), contraste (o destaque que desperta interesse na apresentação) e hierarquia (ordem na qual o olho humano percebe o que vê. Essa ordem é criada pelo contraste visual, entre outras formas, num campo de percepção). Portanto, é necessária uma definição dos objetivos da visualização, além de se pensar qual tipo de informação poderá ser extraída pelo público. “O primeiro e principal objetivo de qualquer gráfico ou visualização é ser uma ferramenta para seus olhos e cérebro para perceber o que está além do seu alcance natural” (CAIRO, 2013: 9-10)⁸⁸.

Com isso claro, o jornalista deve, então, identificar qual é o melhor tipo de visualização para alcançar os objetivos pretendidos com a apresentação dos dados. Cada conjunto de informação tem uma visualização ideal de ser mostrado ao público. Por

⁸⁷Tradução do autor. No original: “The first goal of an infographic is not to be beautiful just for the sake of eye appeal, but, above all, to be understandable first, and beautiful after that; or to be beautiful thanks to its exquisite functionality.”

⁸⁸Tradução do autor. No original: “The first and main goal of any graphic and visualization is to be a tool for your eyes and brain to perceive what lies beyond their natural reach”.

exemplo, os gráficos de linha são adequados para evoluções ao longo do tempo, enquanto gráficos de barra são ideais para comparar dados por categorias. É importante, portanto, que o jornalista de dados busque o melhor tipo de visualização para o dado que ele quer apresentar ao leitor.

Para visualizações com dados geolocalizados – como, por exemplo, dados sobre quais os locais com maiores índices de criminalidade em uma cidade –, dois ótimos softwares para elaborar mapas interativos e funcionais são o Google FusionTables⁸⁹ e o Tableau Public⁹⁰. O primeiro tem uma grande vantagem pela sua capacidade de processar planilhas com grandes volumes de dados e pela ótima documentação disponível. Com o programa, é possível transformar as informações em mapas (colunas com localizações são automaticamente interpretadas), cronologias ou rankings simples. Além disso, o usuário pode geocodificar endereços sem grandes dificuldades e exportar os dados em KML (formato de arquivo usado para exibir dados geográficos). Para se aprofundar nos recursos do software, os interessados encontram muitos vídeos tutoriais que apresentam as funcionalidades e os tipos de visualizações possíveis de serem feitas

Já o Tableau Public é um programa de visualizações interativas de dados que não requer habilidades de programação, mas exige do usuário um bom entendimento de organização de bases de dados e formatos de gráficos. Ele é ideal para visualizar conjuntos de dados multidimensionais. É necessária a instalação do software no desktop para utilização de suas ferramentas. Há uma vasta gama de recursos de edição e é possível unir tabelas e gráficos para construir visualizações mais complexas. O site oferece tutoriais e treinamentos para quem quiser explorar as funcionalidades do programa. Além destes dois softwares indicados neste trabalho, pode-se recomendar a utilização de mais outros três programas para criar histórias interativas baseadas em mapas: CartoDB⁹¹, Odyssey.js⁹², StoryMap JS⁹³.

Para criar visualizações menos elaboradas, porém eficazes, as opções de sites e programas que auxiliam o jornalista de dados são variadas. Neste trabalho, serão indicados a título de exemplificação quatro sites de visualização de dados: Infogr.am⁹⁴, Piktochart⁹⁵, Plot.ly⁹⁶, Datawrapper⁹⁷ e Visual.ly⁹⁸. Em comum entre estas ferramentas

⁸⁹ Disponível em: <https://fusiontables.google.com>

⁹⁰ Disponível em: <https://public.tableau.com/s/>

⁹¹ Disponível em: <https://carto.com/>

⁹² Disponível em: <https://cartodb.github.io/odyssey.js/>

⁹³ Disponível em: <https://storymap.knilab.com/>

⁹⁴ Disponível em: <https://infogr.am/>

⁹⁵ Disponível em: <https://piktochart.com/>

está o fato de que suas interfaces intuitivas permitem que o usuário consiga criar visualizações funcionais em um curto espaço de tempo, caso ele já tenha uma base de dados estruturada quando iniciar a projeção da visualização.

O Infogr.am é um site que oferece, gratuitamente, uma série de layouts e gráficos para construir visualizações interativas. É possível inserir os dados manualmente em uma planilha do próprio programa ou fazer um upload de uma base de dados já estruturada. Com isso, o usuário pode, em seguida, já escolher – entre as opções disponíveis – qual tipo de gráfico utilizará para representar os dados da planilha. Simples e prático, o site oferece a possibilidade de publicar e divulgar o projeto criado ou incorporá-lo ao próprio local de publicação da reportagem.

Assim como o Infogr.am, o Piktochart permite criar infográficos interessantes de maneira ágil. A versão gratuita oferece alguns layouts personalizáveis e os usuários podem retirar e adicionar elementos "arrastando e soltando". Terminado o trabalho de edição, é possível publicar o infográfico como HTML (ou seja, em uma página própria no navegador), compartilhá-lo em redes sociais ou incorporá-lo ao site do usuário. Aos interessados, há a possibilidade de adquirir a versão paga, que oferece mais de 100 modelos e outros recursos adicionais.

O Plot.ly é um software flexível que permite a criação de diferentes tipos de gráficos, apresentações e painéis. Com ele, é possível fazer análise de dados usando programação – com JavaScript, Python, R, por exemplo – ou fazendo upload de uma planilha do Excel. Há uma livreria de visualizações disponível no site, que oferece ao usuário a possibilidade de ver e experimentar visualizações criadas no site. Isso auxilia o jornalista que esteja em dúvida sobre qual recurso usar e de como representar os dados que tem a sua disposição.

Outro software para construção de gráficos é o Datawrapper, que foi criado por uma instituição alemã que promove treinamentos para jornalistas. Para iniciar, o jornalista deve colar a base de dados na primeira tela do programa e seguir as instruções para construir e personalizar a visualização. Existem alguns tipos de gráficos disponíveis e suas variáveis, tais como de linha, barra, pizza e tabela. Uma desvantagem da ferramenta é que não é possível combinar vários gráficos em uma única visualização.

⁹⁶ Disponível em: <https://plot.ly/>

⁹⁷ Disponível em: <https://www.datawrapper.de/>

⁹⁸ Disponível em: <https://visual.ly/product/infographic-design>

Por fim, o Visual.ly é uma poderosa ferramenta, que oferece uma vasta gama de gráficos para diferentes tipos de necessidade, basta que o usuário escolha a que julgar mais apropriada conforme seus respectivos objetivos. O site oferece um conjunto de ferramentas simples para a construção de visualizações de dados interessantes, bem como uma plataforma para compartilhar suas criações. Um detalhe importante do programa é o seu caráter de mídia social, uma vez que o site permite a interação entre os usuários, que curtem e comentam os gráficos criados no software. É um ótimo local, portanto, para debater ideias e discutir quais os melhores tipos de visualizações para determinados dados disponíveis.

Em suma, percebe-se que o jornalista de dados está munido de uma variada rede de ferramentas de visualizações que lhe permite construir visualizações interativas, mesmo sem o domínio da programação. Como as mudanças são constantes, novas ferramentas surgem a cada ano e novos recursos aparecem para suprir carências de visualizações dos anos anteriores. Por isso, a formação contínua e a atualização constante sobre o que há de novo no mercado é fundamental. Sem isto, o jornalista pode ficar desatualizado em relação aos programas e softwares mais usados na elaboração de reportagens guiadas por dados. Cabe ao repórter, então, pesquisar e sempre se atentar às ferramentas e programas utilizados nas grandes redações e entre outros jornalistas de dados do mundo. Cursos online, fóruns de debate de jornalismo de dados e comunidades e páginas do tema em redes sociais são importantes meios para isso.

3.1.6. A MATEMÁTICA NO MANEJO COM NÚMEROS

Para ser jornalista de dados é preciso lidar com a matemática. Muitas pessoas escolhem a profissão de jornalista para fugir da matemática, mas não há escapatória para os interessados em seguir carreira de dados. Conforme descreveu Sandra Crucianelli (2013), é importante ter conhecimentos de matemática básica e estatística, tanto descritiva quanto inferencial.

Um jornalista “anumérico” não terá a capacidade de enfrentar o processo conhecido como “a entrevista aos números.” Isto nos interessa especialmente quando se trata de análises de orçamentos públicos, impostos e outras questões como criminalidade e poluição ambiental (CRUCIANELLI, 2013)⁹⁹.

⁹⁹ Disponível em: <https://knightcenter.utexas.edu/books/HDPP.pdf>. Acessado em: 29/05/2017. Tradução do autor. No original: “Un periodista ‘anumérico’ no tendrá la capacidad para enfrentar el proceso conocido como ‘la entrevista a los números’. Esto nos interesa especialmente cuando se trata de análisis de presupuestos públicos, impuestos y otros temas como criminalidad y contaminación ambiental”.

Kathleen Wickham (2003) corrobora o exposto por Sandra Crucianelli (2013) ao dizer que todos os jornalistas precisam de habilidades básicas de matemática. Isso porque ela está presente no dia-a-dia do repórter.

Os orçamentos governamentais, os relatórios de criminalidade e as pesquisas estatísticas formam a espinha dorsal do jornalismo diário. Descobrir a mudança de porcentagem, calcular as equações básicas de tempo/distância e entender as taxas imobiliárias fazem parte do trabalho diário de um jornalista (WICKHAM, 2003: 12)¹⁰⁰.

Portanto, o repórter de dados, que tem o número como fonte primordial, não precisa ser um matemático para lidar com números, mas é importante ter alguma familiaridade e disposição com a disciplina para, a partir dela, conduzir investigações em dados e, assim, produzir grandes reportagens guiadas por dados.

É primordial que um jornalista consiga trabalhar com números acima do que sua formação universitária ensina. Analisar dados é uma forma muito eficaz de produzir conhecimento e gerar boas pautas. Trabalhar com conceitos básicos de estatística é fundamental nessa área, assim como conseguir interpretar gráficos complexos e extrair informações de bancos de dados (SPAGNUOLO, 2017).

“As estatísticas são ferramentas matemáticas para interpretar o significado dos dados e para prever tendências futuras com base nesses resultados” (WICKHAM, 2003: 56)¹⁰¹. Por isso, é importante que o jornalista tenha habilidades com a estatística para conseguir interpretar o real significado de um conjunto de dados e, a partir dele, inferir conclusões elucidativas a cerca de algum tema.

Ter um conhecimento básico de estatística e o papel desempenhado pela manipulação dos números é um elemento importante na caixa de ferramentas de um jornalista. Jornalistas são frequentemente solicitados a avaliar pesquisas e estudos; a menos que eles saibam como os números foram utilizados eles não conseguem informar precisamente acerca dos resultados (WICKHAM, 2003: 39)¹⁰².

¹⁰⁰ Disponível em: <https://books.google.com.br/books?hl=pt-BR&lr=&id=RYtYmMD2ReAC&oi=fnd&pg=PA10&dq=Math+for+journalists&ots=kAXGE-S4-2&sig=zVVv3mwx5FEJkiriO6jYUQd2s4#v=onepage&q&f=false>. Acessado em: 29/05/2017. Tradução do autor. No original: “Government budgets, crime reports and research statistics form the backbone of daily journalism. Figuring out percentage change, calculating basic time/distance equations and understanding real estate tax assessments are all part of a journalist’s daily job”.

¹⁰¹ Disponível em: <https://books.google.com.br/books?hl=pt-BR&lr=&id=RYtYmMD2ReAC&oi=fnd&pg=PA10&dq=Math+for+journalists&ots=kAXGE-S4-2&sig=zVVv3mwx5FEJkiriO6jYUQd2s4#v=onepage&q&f=false>. Acessado em: 29/05/2017. Tradução do autor. No original: “Statistics are mathematical tools for interpreting what data mean and for forecasting future trends based on those findings”.

¹⁰² Disponível em: <https://books.google.com.br/books?hl=pt-BR&lr=&id=RYtYmMD2ReAC&oi=fnd&pg=PA10&dq=Math+for+journalists&ots=kAXGE-S4-2&sig=zVVv3mwx5FEJkiriO6jYUQd2s4#v=onepage&q&f=false>. Acessado em: 29/05/2017. Tradução do autor. No original: “Having a basic understanding of statistics and the role played by the manipulation of numbers is an important element in a journalist’s toolbox. Journalists are frequently asked to evaluate

3.2. O JORNALISMO DE DADOS NO ESPORTE

Área em constante transformação e crescimento, o jornalismo de dados está cada vez mais produzindo conteúdos esportivos. É notável que os principais clubes e seleções dos principais esportes do mundo – como futebol, basquete, baseball, tênis, dentre outros – têm investido na formação de equipes de analistas de dados¹⁰³. Principalmente, para avaliar a performance dos atletas para buscar melhorias no rendimento de cada um. Com isso, é importante ressaltar que o dado é uma peça valiosa dentro do mundo esportivo. O jornalista de dados esportivo tem, basicamente, a missão de transformar informações complexas e números complicados de partidas, campeonatos esportivos ou séries históricas de confrontos em algo maleável e mastigado para o leitor. Logo, percebe-se que o jornalismo de dados tem um campo fértil pela frente para explorar e produzir grandes reportagens guiadas por dados.

No entanto, enquanto a política e a economia ganharam força dentro do contexto do jornalismo de dados, o esporte ainda está a caminho de crescer como força expressiva no meio dos números. Até o momento da publicação deste trabalho, não há uma pesquisa oficial que descreva como está a distribuição das publicações jornalísticas de dados por temas. Mas, sabe-se que os principais conteúdos das reportagens guiadas por dados vêm das editorias de política, economia e cidade (polícia, saúde, transporte, educação etc.). O esporte é uma área que está em crescimento exponencial, com cada vez mais publicações notáveis de dados em torno de temas esportivos.

3.2.1. PANORAMA GERAL NO MUNDO

Os principais veículos de mídia no mundo já têm, em suas respectivas redações, equipes de jornalismo de dados. O que já representa um avanço para uma área tão nova se comparada às demais do meio. Entretanto, quando o assunto é sobre equipes especializadas em jornalismo de dados no esporte, o eixo é mais reduzido. O que há, na verdade, são grupos de jornalistas de dados que publicam conteúdo sobre diversos temas, sendo um deles o esporte.

surveys and studies; unless they know how the numbers were used they cannot report accurately on the results”.

¹⁰³ Um exemplo clássico foi o modelo aplicado ao time de basebol americano Oakland Athletics pelo gerente geral Billy Beane, que baseava sua busca de transformar a equipe em competitiva através de uma sofisticada análise estatística dos jogadores. A história foi parar no cinema como o nome Moneyball: o homem que mudou o jogo (2011).

Uma das maiores referências em jornalismo de dados no mundo é o site FiveThirtyEight – ou 538¹⁰⁴ –, cujas reportagens inovadoras trazem diferentes tipos de visualizações para variados conjuntos de dados. Dividido em departamentos, a editoria de esporte¹⁰⁵ do FiveThirtyEight tem papel relevante dentro do veículo, com reportagens especiais e conteúdos analíticos interessantes, todos sempre baseados em dados. Neste caso, esportivos. Os esportes mais populares nos Estados Unidos são os que têm mais destaque na publicação, tais como Futebol Americano, Basquete e Baseball.

Para fugir do tradicional, o 538 publicou, em maio de 2016, uma reportagem especial sobre a história por trás da luta Sumô¹⁰⁶, atividade pouco praticada no mundo, com mais adeptos concentrados em países asiáticos, como a Mongólia, onde o esporte é bem forte. Além de uma grande história narrada por meio de uma envolvente escrita, a reportagem tem destaque pelos dados apresentados – característica fundamental presente nas matérias publicadas no veículo. São diferentes tipos de gráficos, que trazem valiosas informações sobre a luta, como, por exemplo, a importância da altura e do peso no Sumô. Esta matéria prova como o esporte – independentemente da modalidade – pode ser fonte para diferentes tipos de reportagens guiadas por dados.

O Financial Times¹⁰⁷, um dos veículos midiáticos mais influenciadores sobre economia e político do mundo, dedica um blog para reportagens esportivas baseadas em dados. Produzido semanalmente pelo jornalista John Burn-Murdoch, o blog “*Baseline*”¹⁰⁸ traz diferentes visualizações de dados usando números para contar histórias, testar hipóteses e acabar com estereótipos. O foco das análises estatísticas não se limita apenas ao futebol e cobre outros esportes como tênis, atletismo, futebol americano, baseball e Fórmula 1, por exemplo.

No entanto, as reportagens esportivas guiadas por dados no Financial Times não se limitam apenas ao blog citado acima. Um exemplo fora do blog é o especial produzido em setembro de 2016 sobre a história do tênis feminino¹⁰⁹. Com cinco

¹⁰⁴ Disponível em: <https://fivethirtyeight.com>

¹⁰⁵ Link para o site da editoria de esportes do 538: <https://fivethirtyeight.com/sports/>

¹⁰⁶ MORRIS, Benjamin. The Sumo Matchup Centuries In The Making. FiveThirtyEight, 2016. Disponível em: <https://fivethirtyeight.com/features/the-sumo-matchup-centuries-in-the-making/>. Acessado em: 31/05/2017.

¹⁰⁷ Disponível em: <http://www.ft.com/>

¹⁰⁸ Disponível em: <http://blogs.ft.com/ftdata/category/baseline/>

¹⁰⁹ BURN-MURDOCH, John. A visual history of women’s tennis. Financial Times, 2016. Disponível em: <https://ig.ft.com/sites/visual-history-of-womens-tennis/>. Acessado em: 31/05/2017.

visualizações estáticas de dados, o autor apresenta, visualmente, um conjunto enorme de informações sobre a história da modalidade.

Outro grande nome do jornalismo de dados é o tradicional jornal americano The New York Times, que tem um grande departamento especializado em dados dentro de sua redação. Com jornalistas de dados, designers e programadores, o veículo produz reportagens inovadoras e impressionantes, como um projeto revolucionário publicado em 2012, chamado SnowFall¹¹⁰. O projeto traz uma maneira nova de apresentar informações, misturando texto, imagens, áudios, vídeos e animações interativas. Todos estes elementos podem também estar presentes em uma reportagem guiada por dados.

Apesar de o departamento de dados do New York Times ter enfoque em materiais sobre temas como política e economia, o esporte também ganha espaço nas produções da editoria. Um exemplo de reportagem guiada por dados foi uma publicada antes da Copa do Mundo de Futebol de 2014¹¹¹, que analisava as chances de cada país em ser sorteado para grupos mais fáceis ou difíceis na primeira fase da competição. Outro material do New York Times que merece menção neste trabalho é o projeto¹¹² publicado após as Olimpíadas do Rio de Janeiro de 2016, que analisou, historicamente, quais países dominaram os Jogos Olímpicos desde a primeira edição do torneio, em 1896, em Atenas, na Grécia.

Percebe-se, portanto, que há um esforço contínuo das grandes redações em trazer o jornalismo de dados para o esporte. Mas, ainda é notável que as grandes produções do meio dão prioridades a outras áreas. As equipes formadas por jornalistas de dados nos principais veículos de mídia do mundo produzem conteúdos em diversos temas, e o esporte tem sido uma editoria que tem ganhado um olhar atento dos jornalistas de dados. Todavia, observa-se que ainda há falta de equipes especializadas em esportes e em dados e, por isso, as produções na área são menores quando comparadas a outras.

¹¹⁰ BRANCH, John. Snow Fall: The Avalanche at Tunnel Creek. New York Times, 2012. Disponível em: <http://www.nytimes.com/projects/2012/snow-fall/#/?part=tunnel-creek>. Acessado em: 31/05/2017.

¹¹¹ AISCH, Gregor; QUEALY, Kevin. A fairer world cup draw. New York Times, 2014. Disponível em: https://www.nytimes.com/interactive/2014/06/03/upshot/world-cup-draw-simulation.html?_r=0. Acessado em: 31/05/2017.

¹¹² AISCH, Gregor; BUCHANAN, Larry. A Visual History of Which Countries Have Dominated the Summer Olympics. New York Times, 2016. Disponível em: <https://www.nytimes.com/interactive/2016/08/08/sports/olympics/history-olympic-dominance-charts.html>. Acessado em: 31/05/2017.

3.2.2. PANORAMA NO BRASIL

No Brasil, o cenário não é diferente. É ainda mais escasso o trabalho do jornalismo de dados esportivo. Isso também por conta de o jornalismo de dados ainda não ser tão desenvolvido e estar tão presente nas redações brasileiras quanto o é nas redações dos principais veículos de mídia estrangeiros. Nas grandes redações, o Estadão Dados é uma das maiores referências na área e produz grandes peças jornalísticas de dados sobre temas como política, educação e cidades, por exemplo. Além disso, há equipes com trabalhos no jornalismo de dados em veículos como G1, O Globo, Folha, TV Globo, Editora Abril, Jornal Correio, Zero Hora, Nexo, dentre outros.

No entanto, percebe-se que não há equipes de jornalistas de dados dedicados exclusivamente ao esporte, tal como se nota em outros grandes veículos fora do país. Em território brasileiro, as reportagens guiadas por dados no esporte são produtos pontuais e especiais. Por exemplo, durante as Olimpíadas do Rio de 2016, o jornal Zero Hora produziu um especial sobre as bolas olímpicas¹¹³. Basicamente, o jornal compilou algumas variáveis (peso, tamanho, preço e marca olímpica – tudo isso era um dado da base de dados elaborada para esta reportagem) sobre as 11 bolas olímpicas usadas em algumas modalidades do torneio. A matéria apostou no visual para transmitir as informações. Apesar de não conter gráficos ou tabelas, que são típicos de reportagens guiadas por dados, trata-se de um material de jornalismo de dados, uma vez que cada variável explicada era um dado do conjunto de informações transmitido.

¹¹³ CONSTANTE, Ohana; MENEZES, Paula. As bolas olímpicas. Zero Hora, 2016. Disponível em: <http://especiais.zh.clicrbs.com.br/especiais/bolas-olimpicas/>. Acessado em: 31/05/2017.

4. ESPIÃO ESTATÍSTICO

O Espião Estatístico é um produto do GloboEsporte.com, que, basicamente, gera dados sobre futebol para conteúdos do site e/ou da TV Globo e SporTV. Para isso, há uma equipe de jornalista de dados esportivos dedicada à compilação estruturada de dados, contextualização, análise e comunicação da informação. Portanto, é uma iniciativa pioneira no país por ser a única equipe especializada em dados esportivos. No entanto, os dados se resumem a apenas um esporte: o futebol.

Os dados compilados são fruto de vários aspectos que envolvem uma partida, desde o local, data e horário de um jogo até um impedimento, finalização ou um cartão vermelho nos acréscimos do segundo tempo (todos os componentes que são compilados e analisados pelo Espião serão descritos no subcapítulo 4.2).

A ferramenta de trabalho utilizada para todos os dados relacionados a este produto é a planilha Excel¹¹⁴. Há uma planilha-modelo que serve de referência para o preenchimento de suas células¹¹⁵ – variando de acordo com a partida que for feita. É necessário um padrão para tudo que é realizado para o Espião Estatístico, de modo que, na hora da análise, as subjetividades sejam minoradas e as estatísticas se tornem consistentes. Por exemplo, na hora de escrever qual foi o motivo de um cartão amarelo ou vermelho recebido por determinado jogador é preciso que haja um padrão para que um mesmo motivo não seja escrito de duas maneiras diferentes. Se um jogador deixa a perna por trás do adversário e o derruba, o motivo correto a se escrever é “Calçou”. Caso não seja seguido este princípio e se o repórter escreve da forma como quer, não é possível formular estatísticas confiáveis. Por isso, toda planilha é feita por um dos membros da editoria de Dados do GloboEsporte.com e, em seguida, revisada por outro integrante (os revisores são os que estão há mais tempo na equipe e que já sabem bem como funciona o fluxo da planilha e as particularidades de seu preenchimento).

Como a demanda é alta e a editoria não tão grande – em março de 2017 a equipe contava com 11 pessoas, sendo que oito faziam e/ou revisavam planilhas –, nem todas as planilhas são completas. Ou seja, há uma escolha entre campeonatos mais ou menos importantes para se decidir quais devem ter uma planilha completa e quais são “resolvidos” apenas com uma simples. A diferença entre ambas é a seguinte: em uma

¹¹⁴Os dados são organizados no programa da Microsoft em linhas e colunas, sendo as primeiras referidas com números [1, 2, 3...] e as segundas com letras [A, B, C... AA, AB...].

¹¹⁵Célula é a unidade de uma planilha formada pela interseção da linha com a coluna do Excel. Ela é identificada pela coluna e linha correspondente. Por exemplo, célula A1: pertence a primeira coluna e a primeira linha da planilha.

planilha simples entram apenas os lances cruciais de uma partida, isto é, gol, cartão amarelo, cartão vermelho e substituição da equipe grande¹¹⁶. Portanto, ficam de fora as finalizações e os impedimentos. Os nomes dos bandeiras (árbitros assistentes) não são colocados nas planilhas simples, uma vez que os impedimentos não são registrados.

A competição de maior relevância no futebol brasileiro e que, conseqüentemente, tem maior audiência no GloboEsporte.com é o Campeonato Brasileiro. Portanto, o Espião Estatístico foca bastante neste torneio. Logo, a este são destinadas as planilhas completas, com todos os lances possíveis de uma partida registrados nas linhas do Excel. Além do Brasileirão, ganham também planilhas completas as partidas das Eliminatórias Sul-Americanas para a Copa do Mundo.

Os campeonatos com planilha simples são os seguintes: Estaduais, Copa do Brasil, Copa do Nordeste, Primeira Liga, Sul-Americana, Taça Libertadores e Brasileirão Série B. Destas competições, apenas os jogos dos times grandes são “planilhados”¹¹⁷. A partida deve ter ao menos um clube da Série A para que seja feita uma planilha simples deste jogo. Por exemplo, na Copa do Nordeste, Sport (Série A em 2017) x Juazeirense necessita de ser “planilhado”. Já Juazeirense x Sampaio Corrêa não. Caso o Sampaio Corrêa avance para a Série A do Campeonato Brasileiro em 2018, por exemplo, todos os seus jogos, inclusive no Campeonato Maranhense, passariam a ser “planilhados” em 2018.

A única exceção para a regra de apenas os 20 da Série A “ganharem” planilha em todos os jogos oficiais do ano vai para algum dos 12 maiores times¹¹⁸ do Brasil, que, porventura, venha a cair para a Série B do Brasileirão. Em 2017, o Internacional é o caso. Isto quer dizer que todos os jogos em competições oficiais do Internacional neste ano serão “planilhados”, mesmo que com apenas planilhas simples.

4.1. COMO FUNCIONA O PREENCHIMENTO DAS PLANILHAS?

A etapa de preenchimento das células das planilhas do Espião Estatístico é fundamental para que todo o trabalho da equipe que o faz seja preciso. Um fator que

¹¹⁶Os times considerados grandes são os 20 da Série A do Campeonato Brasileiro e mais algum outro que tenha caído para a Série B do mesmo campeonato. Em 2017, trata-se do Internacional. Em 2016 foi o caso do Vasco da Gama.

¹¹⁷ Expressão interna do GloboEsporte.com para se referir aos jogos que necessitem de planilhas.

¹¹⁸ No jornalismo esportivo brasileiros, 12 clubes são considerados os maiores do país: Flamengo, Vasco, Fluminense, Botafogo, Corinthians, Palmeiras, Santos, São Paulo, Grêmio, Internacional, Cruzeiro e Atlético-MG. Ou seja, as quatro maiores forças do Rio de Janeiro e São Paulo e as duas maiores do Rio Grande do Sul e de Minas Gerais.

auxilia a compilação dos dados do Espião é a forma como estes são estruturados no Excel. De acordo com o criador do Espião Estatístico, Valmir Storti, o produto já recebeu elogios de analistas e especialistas matemáticos justamente pela estruturação criada para armazenar os dados. Isso é importante porque já elimina uma etapa do processo de construção de uma reportagem guiada por dados, que é a limpeza. Como os dados já são capturados de maneira organizada, não é necessário limpar a base de dados para, então, conseguir analisar e extrair informações dela.

É importante entender que há uma metodologia por trás do lançamento das informações na planilha de um jogo. Como esclarecido acima, o Espião dedica uma planilha para cada jogo analisado. No entanto, o “carro-chefe” do Espião é a planilha-mãe¹¹⁹. As análises para os conteúdos publicados pela equipe vêm das informações extraídas deste local e, como este é um produto da Globo, a disponibilização e o acesso aos conteúdos do Espião Estatístico se restringe aos membros da equipe.

Para que a planilha-mãe esteja completa, ou seja, com todos os jogos oficiais dos times da elite do futebol “planilhados”, é preciso fazer a planilha de cada jogo. Então, é necessário partir do micro (planilha individual de uma partida) para só assim chegar ao macro (planilha-mãe com todos os jogos).

A compilação inicial dos dados de uma partida deve ser feita a partir de uma cópia da planilha-modelo¹²⁰ com o jogo que será “planilhado”. O preenchimento das planilhas do Espião Estatístico é feito, normalmente, após o fim das partidas quando as súmulas dos jogos já foram divulgadas. A divulgação das súmulas é importante para que as informações contidas na planilha do Espião estejam de acordo com o que há de oficial nos jogos “planilhados” do campeonato. Além disso, vale mencionar que a forma de acompanhamento dos lances de uma partida é por meio de gravações feitas a partir das transmissões dos jogos. Caso algum não tenha sido televisionado, a única forma de acompanhar é através do tempo real da partida, da súmula e também por meio de lances divulgados em sites da internet, como o YouTube¹²¹.

¹¹⁹A planilha-mãe é a planilha do Excel onde são coladas todas as planilhas individuais dos jogos planilhadas durante o ano.

¹²⁰ Há uma planilha-modelo para o preenchimento de todos os jogos que ganham planilhas durante o ano. Este modelo serve como base inicial para que o repórter possa completar a planilha de acordo com as informações do jogo na estruturação de dados usada pelo Espião.

¹²¹YouTube é um site de compartilhamento de vídeos enviados pelos usuários através da internet. Disponível em: www.youtube.com.br

4.1.1. OS 13 CABEÇALHOS INICIAIS

Ao iniciar o preenchimento de uma planilha do Espião, há 13 campos no cabeçalho que devem ser completados antes mesmo do início dos lances de uma partida. São eles: competição¹²²; rodada; momento¹²³; placar¹²⁴; placar¹²⁵; diferença¹²⁶; jogo; casa ou fora; tempo¹²⁷; minuto¹²⁸; time; evento; quem¹²⁹. Estas variáveis dão nome às colunas onde os dados serão preenchidos. Por exemplo, se a coluna da competição for “Libertadores”, todas as células abaixo serão preenchidas com o nome “Libertadores”.

O lançamento destas informações individuais da partida se faz necessário para que a planilha ganhe uma identidade. Cada um dos 13 itens descritos acima são variáveis padronizadas presentes em todas as planilhas individuais de cada jogo “planilhado” pelo Espião. Com isso, caso haja necessidade de uma pesquisa futura sobre dados de alguma partida, estes preenchimentos iniciais auxiliam na rápida identificação de qual jogo se trata. Por meio de filtros do Excel isto é possível. Por exemplo, caso o repórter deseje saber qual time aplicou a maior goleada no Campeonato Brasileiro, esta informação pode ser encontrada na planilha-mãe através de três filtros destes 13 cabeçalhos.

O primeiro filtro é sobre a variável competição para listar apenas as planilhas do Campeonato Brasileiro. O segundo é na coluna rodada para que sejam selecionadas as rodadas da competição que se deseja o recorte. O último é na variável evento – nesta coluna estão presentes diversos registros da planilha (serão explicados ao longo deste capítulo) – no qual o jornalista deve filtrar por “Mandante” e “Visitante”, de forma que apareça uma lista com todos os jogos daquele campeonato naquelas rodadas de todos os times que jogam dentro ou fora de casa. Com isto, o repórter deve olhar para a coluna jogo onde aparece o resultado das partidas e analisar, um a um, qual foi o mais elástico deles. Portanto, o trabalho é facilitado graças à maneira como os dados estão estruturados na planilha-mãe.

¹²² Competição é o campo que se refere ao torneio pelo qual a partida é válida.

¹²³ Momento é o instante do lance que mostra se o time estava vencendo, empatando ou perdendo.

¹²⁴ Placar do mandante: quantidade de gols marcados.

¹²⁵ Placar do visitante: quantidade de gols marcados.

¹²⁶ Diferença numérica da subtração do placar do mandante pelo placar do visitante.

¹²⁷ Tempo mostra se o lance foi na primeira ou na segunda etapa do jogo.

¹²⁸ Refere-se ao exato instante em que o lance ocorreu.

¹²⁹ Quem é o time Mandante e quem é o time Visitante.

4.1.2. ESCALAÇÕES

A próxima etapa no preenchimento da planilha são as escalações de cada um dos times. Deve-se preencher conforme o seguinte padrão: os 11 titulares, os jogadores que permaneceram no banco, os reservas que entraram na partida – neste caso, no máximo três, e, por fim, o técnico da equipe. É importante preencher o nome de todos os jogadores com a grafia correta para que não haja jogadores duplicados no momento posterior de análise dos dados. Conforme exposto no Capítulo 2 deste trabalho, é necessário um cuidado extremo com os dados, já que um erro na apuração pode ser o motivo de uma matéria publicada com números distorcidos. Stray (2016) complementa a ideia de que é necessário saber com que dados o jornalista está lidando e principalmente deixar claro que, caso não o saiba, reconheça isso.

Nem todos os dados vêm com erros de medição anexados. Às vezes você tem que ler a cópia fina para encontrar para fora, ou chamar alguém e pedir. Mas se você não sabe e não pode razoavelmente adivinhar as fontes e as magnitudes do possível erro, então você realmente não sabe o que os dados significam. [...] Assim como um cientista, um jornalista não pode inventar dados, ignorar evidências ou tolerar falácias lógicas. É igualmente importante saber quando você não sabe, quando você não pode responder à pergunta a partir de dados disponíveis (STRAY, 2016)¹³⁰.

Por isso, um simples detalhe como a grafia do nome do jogador na planilha pode custar caro mais à frente na hora da análise, da combinação e da comunicação dos dados. É importante tomar cuidado sempre e checar tudo com muita atenção. Conforme disse o criador do Espião, Valmir Storti (2017), a atenção na compilação do dado deve ser redobrada porque um erro pode comprometer todo o trabalho.

Quando você coleta dados tem o fator humano que você pode errar a coleta do dado. Já aconteceu de a gente publicar matéria com informação errada. Isso compromete a credibilidade. Credibilidade você constrói, conquista e perde. Às vezes, um erro compromete todo o seu trabalho. Tem que se tomar muito cuidado com todo o processo de coleta, tabulação e extração de conteúdo porque você tem uma responsabilidade muito grande (STORTI, 2017)¹³¹.

Além do nome de cada um dos atletas relacionados para a partida, é preciso colocar a abreviação da posição em que cada jogador iniciou a partida. Esta abreviação

¹³⁰ Disponível em: https://www.cjr.org/tow_center_reports/the_curious_journalists_guide_to_data.php. Acessado em: 24/04/2017. Tradução do autor. No original: “Not all data comes with measurement errors attached. Sometimes you have to read the fine print to find out, or call someone and ask. But if you do not know and cannot reasonably guess the sources and magnitudes of possible error, then you don’t really know what the data means. [...] Just like a scientist, a journalist can’t make up data, ignore evidence, or condone logical fallacies. It’s equally important to know when you don’t know, when you can’t answer the question from available data.”

¹³¹Entrevista concedida ao autor.

é incluída na célula ao lado do nome de cada atleta. No Espião, há um padrão que todos os repórteres devem seguir ao preencher esta informação:

Posição	Sigla
goleiro	gol
lateral direito	lad
zagueiro direito	zad
zagueiro esquerdo	zae
lateral esquerdo	lae
volante	vol
meio-campo	mec
atacante	ata
técnico	tec

4.1.3. TRIO DE ARBITRAGEM

Outra parte fundamental na planilha é o preenchimento do trio de arbitragem do jogo. Nesta etapa, é necessária uma atenção em dois aspectos: escrever corretamente o nome do árbitro e de cada um dos dois bandeiras, e apontar o nome dos times que têm seus lances acompanhados por cada assistente no primeiro tempo (células imediatamente à esquerda dos nomes dos auxiliares 1 e 2). O time que ataca, no primeiro tempo, para o lado onde o assistente 1 está atuando deve ser informado na planilha. Assim como o time que ataca, no primeiro tempo, para o lado onde o bandeira 2 está atuando também deve ser informado. Na imensa maioria dos estádios, a TV mostra o auxiliar 1 à direita do vídeo. Ele está sempre à direita do quarto árbitro.

Apontar quem é o bandeira responsável por assinalar os impedimentos da equipe no primeiro tempo é essencial para que os lances de impedimentos registrados durante o jogo sejam encaminhados para o assistente correto que os assinalou. No segundo tempo, as equipes trocam de lado de ataque e, automaticamente, o árbitro 1 que assinalou os impedimentos de uma equipe no primeiro tempo, passa a marcar os impedimentos do time adversário no segundo tempo.

4.1.4. OS 13 PASSOS ANTES DO INÍCIO DOS LANCES DO JOGO

Antes do início do preenchimento do que aconteceu na partida, ou seja, todos os lances compilados em uma planilha completa, ainda são necessários os preenchimentos de mais 13 campos do documento do jogo: classificação (do time visitante e do mandante antes do início da rodada), local (nome do estádio em que aconteceu a partida), clima (seco ou chuva), dia (segunda-feira, terça-feira etc), data (preenche-se no formato de ano, mês e dia; por exemplo: 2017 03 27), horário (21h45, por exemplo),

posse de bola (do time mandante e do visitante) e faltas cometidas (do mandante e do visitante). Além destes treze campos, há dois espaços para o preenchimento do público e da renda do jogo. Mas, isto é inserido apenas pelo revisor responsável e, no caso das Eliminatórias, não há esse preenchimento.

Novamente, é crucial seguir o padrão no lançamento destas informações na planilha para que a planilha-mãe contenha um conjunto homogêneo de dados. Isso auxilia a etapa de combinação (análise) das informações para fundamentação de pautas e reportagens. Por exemplo, com o lançamento dos horários das partidas na planilha, é possível fazer uma lista, ao fim do Campeonato Brasileiro, para saber quais foram os horários que mais tiveram partidas ao longo da competição.

4.1.5. OS LANCES DE UMA PARTIDA

Como já foi citado anteriormente, são vários os lances de um jogo que devem ser registrados na planilha durante os 90 minutos, indo desde um impedimento até uma finalização, cartão ou substituição. Um detalhe importante é que os lances são sempre preenchidos tomando o exato momento do início da jogada, mais precisamente o primeiro segundo em que ela tem início. Ou seja, se uma finalização começou com um passe na defesa, assistência de um jogador na intermediária ou chute na pequena área, deve-se iniciar a contagem da jogada a partir do momento (ou seja, a partir do segundo) que a bola saiu do pé do atleta que deu o passe inicial. A exceção para esta regra são os gols. Para este caso específico, o que vale é o exato momento em que a bola ultrapassou a linha do gol adversário.

4.1.5.1. IMPEDIMENTOS

Na planilha, o impedimento é marcado a partir do instante em que a bola sai do pé de quem deu o passe ao jogador impedido. O jornalista deve informar o nome do atleta flagrado na banheira¹³² e a região do campo que ele estava no momento do passe. A equipe do Espião adota um padrão para registro das partes do gramado em que tal lance acontece. A arte a seguir detalha isso melhor:

¹³²Termo cunhado pelo futebol para designar a posição de impedimento do jogador.

Figura 11



Fonte: Espião Estatístico – GloboEsporte.com

O campinho com as regiões do campo ajuda a homogeneizar os registros dos lances. Isto possibilita uma melhor estruturação da informação de forma mais consistente. Com isso, um lance escrito em uma linha da planilha pode ser facilmente imaginado pela pessoa que o observa sem assistir ao mesmo.

Além destas informações, é necessário fazer uma avaliação da infração assinalada pelo bandeira. Primeiro, é dito se ele acertou ou não. O jornalista tem três opções: “correto”, “errou” ou “duvidoso”. Depois, há ainda uma avaliação subjetiva da dificuldade de marcação do impedimento: “fácil” ou “difícil”. Caso não seja possível identificar a infração, o repórter precisa informar nesta linha da planilha, no campo dos comentários¹³³, que a TV ou gravação não mostra o exato momento do impedimento.

Este tipo de registro na planilha é um dos principais destaques do Espião, uma vez que este é um dos poucos produtos, senão o único, que consegue estruturar uma base de dados confiável para avaliar o desempenho dos árbitros assistentes. E mais. O

¹³³Há uma coluna designada a comentar os lances da planilha, serve tanto para impedimento ou finalização quanto para um cartão ou substituição.

Espião nasceu quando Valmir Storti (2017) percebeu que não havia, no cenário do jornalismo esportivo brasileiro, alguma equipe que medisse a eficiência das marcações de impedimento no Brasileirão.

[...] ninguém analisa o desempenho dos bandeirinhas. Quem é o melhor bandeirinha do Brasil? Impedimento é um negócio que dá tanta polêmica. Eu vou ver todos os jogos para saber isso.[...]Para saber quem era o melhor bandeirinha do Brasil eu criei o Índice Bandeira Branca, que mostrava os bandeiras que conseguiam um aproveitamento de acerto em pelo menos 90% dos impedimentos que marcavam. São muitos impedimentos marcados, são mais de mil por campeonato (STORTI, 2017).

Em 2009, ano que Valmir Storti (2017) teve esta ideia, ele conseguiu assistir 209 dos 380 jogos do Campeonato Brasileiro. A ideia de analisar o desempenho dos árbitros auxiliares foi o ponto de partida para estruturação de tudo que a planilha do Espião Estatístico consegue trazer.

[...] Aí, eu pensei: se eu vou ter que ver todos os jogos, eu não vou anotar só os impedimentos. Conforme eu ia fazendo os jogos, eu ia tendo ideia de pauta. E a cada pergunta que eu não conseguia extrair os dados que eu queria, eu mudava a estruturação dos dados na planilha para conseguir responder àquela pergunta (STORTI, 2017).

Logo, percebe-se que a planilha expandiu, ao longo do tempo, os registros dos lances de uma partida. Se antes, a ideia era ver apenas os impedimentos, isto evoluiu para outros, conforme este trabalho apresentará em seguida.

4.1.5.2. SUBSTITUIÇÕES

Por jogo, cada time pode fazer, no máximo, três substituições¹³⁴. Na planilha, há, então, três marcações possíveis: substituição 1, substituição 2 e substituição 3. Primeiro, o jornalista deve informar quem saiu do campo e a posição dele na partida. Ou seja, ata, mec, lad ou gol, por exemplo. Em seguida, o repórter também precisa informar, na mesma linha, quem entrou no lugar dele e em qual posição.

Além disso, foi padronizado que o momento da substituição é o instante em que os jogadores (substituído e substituto) se cumprimentam na beira do gramado ou no segundo em que o substituto pisa em campo pela primeira vez. Isso é importante para a definição do tempo em campo dos titulares e reservas de uma partida. O tempo em campo serve como parâmetro para diversos produtos do Espião Estatístico. Por

¹³⁴Substituição é uma ação realizada durante uma partida de futebol, em que um jogador que iniciou o jogo no time titular dá lugar a um outro que estava no banco de reservas até aquele momento.

exemplo, a partir do cruzamento de dados, há um produto chamado “Artilheiro da Hora”, que será explicado no subcapítulo 4.2.1.

4.1.5.3. CARTÕES E PÊNALTIS COMETIDOS

Há quatro padrões de cartões na planilha: “amarelo”, “amarelo anulado”, “vermelho” e “vermelho direto”. O amarelo anulado refere-se ao primeiro amarelo tomado por um jogador que foi expulso por ter levado dois cartões amarelos na partida. É uma forma de diferenciação dos cartões. Já a diferença entre vermelho e vermelho direto é que o último refere-se ao jogador que é expulso de campo sem receber o segundo amarelo. Normalmente, por falta grave ou ofensa ao juiz ou torcida.

O preenchimento entre os cartões amarelos e os pênaltis cometidos é o mesmo. Quanto ao instante do lance, deve ser o exato momento da falta e não quando o árbitro mostra o cartão ao atleta. Para os cartões vermelhos, o preenchimento é no momento em que o árbitro adverte o atleta com o cartão. A diferença deve-se ao tempo em campo do atleta. Os jogadores expulsos têm o tempo em campo contabilizado até o instante em que o cartão vermelho é mostrado. Portanto, o mais importante é registrar o momento da exibição do cartão do que do instante da infração cometida.

Para registrar os lances na planilha, o jornalista deve dizer qual o tipo de cartão, quem foi o atleta advertido e em que região do campo ele estava. Além disso, é necessário informar o nome do jogador que puxou o cartão, ou seja, quem motivou o cartão recebido pelo adversário. Por fim, no campo dos comentários, coloca-se o motivo do cartão ou pênalti cometido. Há uma lista padronizada de motivos que deve ser seguida por todos que fazem as planilhas. São 19 opções:

Figura 12

Motivo
Agarrou
Atitude antidesportiva (discussão, excesso de faltas, falta técnica etc)
Braçada (tapa)
Cabeçada
Calçou
Carrinho (tesoura)
Cera(substituição, atendimento médico, cobrança de lateral etc)
Comemoração de gol

Cotovelada
Empurrou
Joelhada
Mão
Motivo não identificado
Pisão
Pontapé
Reclamação
Simulação
Solada
Trombada

Fonte: Espião Estatístico – GloboEsporte.com

A padronização dos motivos de cartão ou pênalti se fez importante para descobrir quais eram os tipos de faltas mais comuns que resultaram em cartões ou pênaltis. Isso possibilita a elaboração de um ranking com os motivos de cartões mais comuns em uma competição ou, por exemplo, quantos cartões por reclamação determinado time tomou durante a temporada.

4.1.5.4. FINALIZAÇÕES, FALTAS DIRETAS E PÊNALTIS COBRADOS

As faltas diretas e os pênaltis são dois tipos de finalizações de bola parada. Como o início e o fim da jogada se concentram em um lance – a batida do jogador em direção à meta adversária –, este tipo de ataque tem início e fim no mesmo momento. Portanto, na planilha, preenche-se apenas qual é o tipo de jogada, se um pênalti ou uma falta direta, o nome do jogador incumbido da cobrança, a forma como ele a realizou (se de pé direito, esquerdo, calcanhar etc) e a região do campo.

Diferentemente das cobranças de bola parada, as finalizações com bola rolando são preenchidas desde o início da jogada, ou com últimos três toques na bola, ou quatro em caso de arremate aéreo. Há diferentes tipos de registros para o início de uma jogada, pode-se marcar “passe”, “contra-ataque”, “roubada de bola”, “lateral” não aéreo e “escanteio” quando curto e não alçado na área. O preenchimento dos campos de um lance de finalização sempre segue o seguinte padrão: o que foi (passe, lateral, finalização; quem (nome do jogador); como (com que perna ele fez isso); e onde (qual região do campo). Este é o padrão de preenchimento para cada etapa de uma

finalização. Entenda etapa como quem participa da jogada. Por exemplo, se o jogador X cobra um lateral com a mão na ponta direita, o companheiro W cruza na área a partir do bico direito para a assistência de cabeça do atleta Y para a finalização de voleio do atacante G, todos estes quatro lances devem ser registrados em uma linha da planilha.

Há dois tipos diferentes de finalizações com a bola rolando: aéreas e rasteiras. Nas aéreas, há seis tipos diferentes de marcações: “bola levantada”, “cruzamento”, “falta levantada”, “lançamento”, “lateral” (– só quando longo e para dentro da área adversária, quando for uma jogada de ataque) e “escanteio”. O último tem uma especificidade quanto ao preenchimento. Diferentemente dos demais que seguem o mesmo padrão de indicar o como (qual pé) e onde (região do campo), no escanteio o como vira em que lado do ataque: se direito é “ld” ou se é esquerdo “le”. Já o onde se refere ao local onde a bola batida chega dentro da área: “primeiro pau”, “segundo pau”, “entrada da pequena área”, “marca do pênalti” ou “outro lado da área”.

Nas rasteiras, há algumas possibilidades: “assistência”, “tabela¹³⁵”, “falta dois toques¹³⁶”, “jogada individual”, “rebatida¹³⁷”, “rebatida do goleiro” ou “rebatida de finalização”. As duas primeiras são marcações em que o jogador que a fez será credenciado como “garçom” do lance. Isso porque o fim da jogada é uma finalização e eles foram os responsáveis pelo último toque na bola antes de o companheiro chutar. Todas as duas devem ser preenchidas conforme o padrão já descrito anteriormente, com nome do jogador, com qual perna e em que região do campo. Já as rebatidas de finalização nada mais são que finalizações. No entanto, o motivo para serem “rebatida de finalização” e não “finalização” é para indicar que houve um outro chute após esse primeiro. Toda vez que for observado uma rebatida de finalização, já se sabe que, em seguida, o lance conterà uma nova.

Uma peculiaridade importante das finalizações, faltas diretas e pênaltis é o resultado das mesmas. Ou seja, para onde a bola foi. O jornalista deve marcar: “gol”, “trave”, “travessão”, “fora”, “goleiro defende”, “goleiro espalma”, “goleiro salva”, “por cima”, “alta por cima”, “defesa corta”, “defesa salva”, “bloqueado” e “barreira”. Se a bola tomou o rumo da meta, deve-se descrever os quadrantes do gol, que é dividido em nove partes. Por exemplo, se a bola vai no ângulo direito do goleiro, o quadrante a se

¹³⁵É um tipo de lance em que dois jogadores trocam passes entre eles.

¹³⁶A falta dois toques indica que um jogador tocou na bola antes do outro finalizar ou levantar na área. Ou então, o jogador que realizou a cobrança optou pelo passe curto. Mas, como o passo originou-se de uma falta, marca-se, neste caso, falta dois toques.

¹³⁷É uma ação executada pelo time adversário que tenta recuperar a bola, mas apenas a rebate para alguma parte do gramado.

apontar é 1. São nove quadrantes: de um a três para as partes superiores da meta, indo da esquerda para a direita; de quatro a seis para as partes a meia altura do gol, indo da esquerda para a direita; e de sete a nove para as partes inferiores das traves defendidas pelo goleiro, indo da esquerda para a direita.

Isso é importante, por exemplo, para que, em cobranças de pênalti, o Espião Estatístico tenha um histórico de onde o jogador costuma bater as penalidades. Além de saber onde ele bate, pode-se revelar quantas vezes ele fez o gol, quantas foram para fora, quantas pararam na trave ou em quantas o goleiro fez a defesa.

Ainda neste tópico, o campo de comentários é importante para algumas hashtags, famoso jogo da velha (#), convencionadas pelo Espião. A # mais usada em finalizações é a #deprimeira para indicar que o finalizador bateu de primeira na bola. Além dela, há algumas outras hastags como: “#defesa, #drible, #erro (da arbitragem), #frango, #golperdido, #golaço e #mico. Todas elas auxiliam no momento posterior de buscas e análises dos lances compilados no Espião Estatístico. Por exemplo, se surgirem demandas para descobrir quantos frangos de goleiro aconteceram em um campeonato, esta informação pode ser mais facilmente encontrada ao se fazer um filtro na planilha buscando as #frango.

Por fim, quanto à minutagem destes três tipos de lance devem seguir o padrão do início em que a bola sair do pé do jogador que iniciou a jogada. A única exceção a esta regra são os gols porque a contagem do minuto deve ser a do instante em que a bola cruzar a linha da meta defendida pelo goleiro adversário.

4.1.5.5. TEMPO EM CAMPO

A última etapa do preenchimento de uma planilha de um jogo é colocar o tempo em campo de todos os jogadores da partida. Para isso, basta acrescentar duas linhas durante a marcação dos lances: “fim do primeiro tempo” (quando o juiz apitar o fim da etapa inicial) e “fim de jogo” (quando o árbitro assinalar o final da partida. É crucial que a hora, minuto e segundo do “fim do primeiro tempo” e “fim de jogo” sejam exatamente as mesmas dos instantes em que o árbitro soar o apito ou erguer os braços. Isso ajuda na precisão dos tempos em campo de atleta. Às vezes, dois jogadores estiveram em campo durante um campeonato pela mesma quantidade de minutos, mas um ficou mais segundos em campo que o outro, então o primeiro jogou mais tempo que o segundo.

Na planilha-modelo, há formulas pré-programadas que quando as linhas de fim de etapas são preenchidas, os tempos em campo passam a aparecer automaticamente em

uma célula designada. Depois, basta colar o tempo calculado na coluna designada ao tempo em campo dos atletas na planilha. Com isso, o tempo de permanência de um jogador naquela partida já fica registrado. Porém, ainda é necessário regular o tempo dos substituídos e dos substitutos, uma vez que eles não estiveram em campo durante toda a partida. Novamente por meio de fórmulas, basta copiar o tempo exato das substituições e colar nos lugares indicado, que, automaticamente, a planilha constará com o tempo correto de todos os titulares e reservas.

Este armazenamento de dado é crucial para diversas análises ou conteúdos entregues pelo Espião, como, por exemplo: saber quem é o jogador que leva um cartão no período mais curto de tempo de jogo. Ou seja, com o tempo em campo dos atletas, basta somar o tempo de todos os jogos deste atleta e dividir pela quantidade de cartões levados por ele nesta competição. O resultado revela quanto tempo determinado jogador leva para ser “premiado” pelo juiz.

4.2. CONTEÚDOS DO ESPIÃO ESTATÍSTICO

Conforme indicado no subcapítulo 2.2 deste trabalho, o fluxo de trabalho de um jornalista de dados segue cinco passos: compilar, limpar, contextualizar, combinar e, por fim, comunicar. O subcapítulo 4.1 mostrou que o preenchimento organizado da planilha engloba dois destes cinco passos: compilar e limpar. Isso porque como os dados produzidos no Espião já são compilados de maneira organizada e estruturada, eles já passam a ser passíveis de análise assim que a planilha estiver feita.

As etapas seguintes – contextualizar, combinar e comunicar – são as necessárias para que todas as estatísticas do Espião nasçam e possam ser compartilhadas com os leitores do GloboEsporte.com. Para isso, a próxima fase é combinar, ou melhor, cruzar os dados e interpretá-los.

Este é um estágio rico, uma vez que muitos conteúdos podem ser produzidos pelo Espião a partir do cruzamento de dados. Com poucos filtros na planilha pode-se descobrir, por exemplo, quantos gols de falta aconteceram em um determinado campeonato. O leque de opções é variado. Mas, para não se estender muito, esta monografia vai dissertar sobre dois conteúdos que são produzidos, periodicamente, ao final de cada edição do Campeonato Brasileiro. São eles: Artilheiro da Hora e Jogadores decisivos.

4.2.1. ARTILHEIRO DA HORA

O Artilheiro da Hora é um conteúdo exclusivo do Espião Estatístico que avalia quem são os goleadores que precisam de menos minutos em campo para fazer um gol. Isso pode ser medido em uma competição apenas, como o Ranking deste quesito no Campeonato Brasileiro de 2016¹³⁸. Outra saída é fazer uma classificação geral de todas as competições que têm a participação dos times da “elite” do futebol – 20 clubes da Série A do Brasileirão –, conforme o publicado no início de 2017 pelo setorista do São Paulo do GloboEsporte.com, Marcelo Hazan, na matéria “Gilberto e Pratto engrenam com boa média de gols por minutos jogados”.¹³⁹

Para que este tipo de objeto seja produzido são necessários alguns cruzamentos de dados. Antes de tudo, o autor precisa definir se quer um ranking por competição ou um contando todas. O Campeonato Brasileiro de 2016 servirá como exemplificação da elaboração deste tipo de conteúdo.

Primeiramente, deve-se checar se a planilha-mãe¹⁴⁰ está atualizada com todos os jogos feitos e revisados já colados nela. Isto é necessário para que a análise não seja feita sem a presença de alguma partida do campeonato, é importante que todos os jogos estejam presentes na planilha-mãe. Após esta etapa, o primeiro filtro a ser feito é na coluna da competição para escolher apenas jogos do Brasileirão. Em seguida, o repórter deve filtrar nos resultados para que apenas os gols sejam exibidos. Com isso, já são revelados todos os lances com gols deste torneio.

Para prosseguir é preciso definir antes o critério a ser adotado para a publicação deste material. Normalmente, para que algum jogador entre no ranking ele deve ter feito pelo menos 40% da quantidade de gols do artilheiro do campeonato. Por exemplo, se o artilheiro da competição tem 20 gols, somente os atletas que já fizeram cinco ou mais podem concorrer ao “prêmio” de artilheiro da hora. Isso porque, senão, um jogador que esteve em campo em apenas uma partida e fez dois gols levaria vantagem sobre os demais que representaram as respectivas camisas de seus clubes por mais jogos.

¹³⁸AIRES, Eduardo; MENEZES, Marcio. Sassá faz gols em metade do tempo do artilheiro Fred; veja o ranking. GloboEsporte.com, 2016. Disponível em: <http://globoesporte.globo.com/futebol/brasileirao-serie-a/noticia/2016/10/sassa-faz-gols-em-metade-do-tempo-do-artilheiro-fred-veja-o-ranking.html>. Acessado em: 30/03/2017

¹³⁹HAZAN, Marcelo. Gilberto e Pratto engrenam com boa média de gols por minutos jogados. GloboEsporte.com, 2017. Disponível em: <http://globoesporte.globo.com/futebol/times/sao-paulo/noticia/2017/02/artilheiro-da-hora-gilberto-e-pratto-poem-sao-paulo-forte-na-disputa-veja.html>. Acessado em: 30/03/2017

¹⁴⁰A planilha-mãe é onde são coladas todas as planilhas individuais feitas para cada jogo dos times da “elite” do futebol. Isso permite que a análise possa ser completa, uma vez que todas as partidas do ano destes clubes se encontram em um mesmo arquivo de Excel.

Com isso, apenas os goleadores com certa quantidade de gols concorrerão ao quesito. Agora, uma primeira lista de artilheiros já foi elaborada. Basta saber o tempo em campo de cada um, cruzar com a quantidade de gols e descobrir quanto tempo eles levam para marcar na competição. O último filtro é desmarcado e outro será feito.

Para averiguar o tempo em campo de cada atleta é preciso filtrar, na planilha, por “titular” e “reserva”. Logo, serão exibidas as informações de todos os jogadores que entraram em campo em todas as rodadas do Brasileirão. A próxima fase é fazer o filtro na variável quem e selecionar o nome de algum jogador. Neste caso, para o ranking de artilheiros da hora, deve-se selecionar nome a nome da lista prévia de goleadores para revelar o tempo em campo de cada. Por exemplo, seleciona-se Ricardo Oliveira, do Santos. A planilha passa a exibir todos os jogos em que ele entrou em campo, seja como titular ou reserva. Cada linha é um jogo diferente. Agora, basta somar os minutos em campo em todas as partidas. Portanto, a soma total revela o tempo no gramado de Ricardo Oliveira atuando pelo Santos no Campeonato Brasileiro de 2016. Depois é necessário fazer o mesmo com todos os artilheiros selecionados previamente.

Com a definição do tempo em campo de todos os goleadores, basta dividir isto pela quantidade de gols de cada um. O resultado será quanto tempo cada atleta necessita para balançar as redes. Em seguida, os resultados são classificados do menor para o maior e, assim, são definidos os “Artilheiros da Hora”.

4.2.2. JOGADORES DECISIVOS

Neste ranking, o objetivo é mostrar quem são os atletas que costumam resolver o problema quando a situação está delicada. O jogador decisivo é aquele faz um gol ou dá uma assistência quando sua equipe está empatando ou perdendo por um gol de diferença. Quanto mais ele participa dos gols do time nestas condições, mais decisivo ele é. No Campeonato Brasileiro do ano passado, Marinho, que ajudou o Vitória a escapar do rebaixamento, foi o melhor jogador neste quesito.¹⁴¹

Para elaborar esta lista é preciso montar, primeiramente, dois rankings: um de goleadores e outra de garçons¹⁴². Os filtros necessários são os seguintes. Primeiro, na coluna dos resultados, seleciona-se “gol” para que todos os gols sejam exibidos. Depois,

¹⁴¹SOUSA, Eduardo; MARÇAL, Guilherme; MALESON, Roberto; STORTI, Valmir. Duplas dinâmicas, goleiros decisivos, artilheiros: as estatísticas do Brasileirão. GloboEsporte.com, 2016. Disponível em: <http://globoesporte.globo.com/futebol/brasileirao-serie-a/noticia/2016/12/duplas-dinamicas-goleiros-decisivos-artilheiros-estatisticas-do-brasileirao.html>. Acessado em: 03/04/2017.

¹⁴²Na terminologia do mundo do futebol, os garçons são os jogadores que dão assistências para finalização dos companheiros de time.

é preciso colocar o campeonato que se deseja analisar, neste caso o Brasileirão. Com isso, todos os gols desta competição passam a estar listados na planilha.

A próxima etapa é filtrar a variável momento em “d” ou “e”. Isto significa que, assim, somente serão exibidos os gols feitos nos momentos em que o time perdia ou empatava a partida. No entanto, os tentos realizados quando o time já perdia por dois ou mais gols de diferença não interessam para este tipo de ranking, uma vez que não são decisivos naquele instante da partida. Por isso, na coluna diferença, tudo que for maior que “-1”, “0” e “1” deve ser desmarcado. Esta coluna exibe as diferenças dos placares entre mandante e visitante: por exemplo, se o mandante ganhar de 3 a 1 o jogo, após abrir 3 a 0 de vantagem e levar um no final da partida. O único gol decisivo seria o primeiro que inaugurou o placar. O segundo e o terceiro não foram decisivos e, por isso, no filtro, devem ser desmarcados para que apenas os decisivos apareçam.

Para o conteúdo dos jogadores decisivos do Espião, o mais importante não é o dado geral, mas o individual de cada atleta. Para chegar nesta fase é preciso elaborar uma lista dos artilheiros e assistentes decisivos para depois somar e montar um ranking final. Isso porque este tipo de atleta tanto pode dar assistências como pode marcar gols nestas situações. A lista dos jogadores decisivos é formada por aqueles que têm a maior quantidade de gols e assistências nestes critérios explicados anteriormente. Esta etapa trabalhosa do cruzamento de dados faz parte da fase “combinar” dos cinco passos da pirâmide invertida do jornalismo de dados¹⁴³. Para prosseguir os métodos de Paul Bradshaw, as etapas seguintes são “contextualizar” e “comunicar”. Estas são feitas no momento da escrita da reportagem. É necessário simplicidade, coesão e boa familiaridade com os dados para que o conteúdo possa ser transmitido da forma mais clara possível ao leitor.

A questão do “comunicar” no Jornalismo de Dados está muito atrelada à visualização de dados de um artigo jornalístico. Há diversas formas atrativas de visualização dentro do universo dos dados no jornalismo. Mas, os conteúdos gerados pelo Espião Estatístico esbarram na falta de mão de obra para conseguir produzir especiais mais visuais e interativos, características importantes deste tipo de jornalismo. Muitas vezes, estes números do Espião são transmitidos com uma arte estática dos

¹⁴³BRADSHAW, PAUL. The inverted pyramid of data journalism. Disponível em: <https://onlinejournalismblog.com/2011/07/07/the-inverted-pyramid-of-data-journalism/>. Acessado em: 17/04/2017.

dados. A imagem pode até resolver o problema, mas ainda não é o ideal, visto os inúmeros trabalhos interativos e visuais neste meio.

4.3. JORNALISMO DE DADOS NA TV

Em fevereiro de 2017, o Espião Estatístico ganhou um quadro semanal no programa Troca de Passes, do SporTV. Os números compilados e analisados pela equipe do Espião passaram a ganhar ainda mais público com a apresentação todas terças e sextas-feiras na TV. A ideia é compartilhar com o telespectador os números interessantes e atraentes tanto de algum jogador específico como sobre o desempenho da equipe em determinado ano ou competição.

Na edição de sexta-feira, dia 07 de abril de 2017, o Espião apresentou os números do meio-campista Camilo¹⁴⁴, que foi assunto daquele dia, quando o técnico do Botafogo foi questionado sobre a mudança de posição¹⁴⁵ e a consequente queda no desempenho no ano do referido atleta. Os dados apresentados no Troca de Passes revelam que, realmente, Camilo teve uma queda brusca no rendimento em comparação com o ano anterior. A imagem abaixo mostra como esses dados são apresentados:

Figura 13



¹⁴⁴Disponível em: <http://sportv.globo.com/troca-de-passes/videos/v/espiao-estatistico-analisa-numeros-de-camilo-e-aproveitamento-dos-clubes-da-serie-a/5786330/>. Acessado em: 08/04/2017

¹⁴⁵ Disponível em: <http://globoesporte.globo.com/futebol/times/botafogo/noticia/insatisfeito-com-barracao-no-botafogo-camilo-cobra-jair-e-vai-embora-sem-treinar.ghtml>. Acessado em: 08/04/2017

Em 2016, o camisa 10 do alvinegro carioca participou diretamente de 28% dos gols do time, com seis gols e seis assistências, enquanto em abril de 2017, ele tinha uma participação nos tentos do time de apenas 4%, com uma única assistência a gol até aquele momento da temporada. Logo, percebe-se que os números corroboram com a insatisfação de Camilo e são boa fonte de debate entre os comentaristas do programa.

4.3.1. SUGESTÃO DE PAUTA

Para cada participação no Troca de Passes, a equipe do Espião Estatístico sugere pautas aos coordenadores do programa. Estas podem ser sobre algum jogador específico, sobre algum time em geral ou sobre algum assunto quente do dia ou semana. Primeiramente, estas pautas são sondadas no grupo do WhatsApp entre a equipe do Espião e coordenadores do programa do SporTV. E depois, enviada com os números consolidados por email.

As pautas enviadas sempre são pensadas em torno de dados que gerem debate entre os comentaristas do programa no dia da participação do Espião no SporTV. É importante que os números sejam claros e suficientes para que possam ser compreendidos, primeiramente, pela equipe de coordenadores do Troca de Passes e, por fim, pelo telespectadores.

4.3.2. APROVAÇÃO DA PAUTA E PARTICIPAÇÃO NO PROGRAMA

A aprovação do que entrará na programação do Troca de Passes depende tanto do interesse e atração dos dados quanto também do espaço disponível no programa. Em dias com a programação apertada, o quadro do Espião pode vir a ser mais enxuto e com menos temas. Usualmente, a participação apresenta dados sobre um ou dois temas diferentes com duração média de cerca de dois a minutos cada. E quando a programação não está carregada de temas, a participação do produto do GloboEsporte.com pode vir a ser mais ampla, com um tempo médio de participação bem maior.

O importante a perceber é que o jornalismo de dados não precisa e não deve ficar restrito apenas ao meio online e pode se expandir para outras plataformas. A prova disso é o Espião Estatístico no Troca de Passes. Há um público interessado em saber acerca dos números do futebol brasileiro e que gosta do debate gerado em torno deles.

5. CONCLUSÃO

Com origens no fim da década de 1960, o jornalismo de dados se remodelou ao longo dos anos e, principalmente, se adaptou aos novos meios para conseguir entregar uma variedade maior de conteúdos. Se antes, este jornalismo se restringia ao papel, ou seja, a gráficos e tabelas estáticos publicados nos jornais e revistas, atualmente, o jornalismo de dados está presente em todas as plataformas, inclusive a televisiva, conforme exposto no subcapítulo 4.3 deste trabalho.

Ficou claro que, para se tornar um especialista em dados, é preciso muita dedicação e disposição para aprender a usar as novas ferramentas e adquirir habilidades que sempre foram negadas entre os jornalistas, como a programação e a matemática. Um jornalista de dados lida com estas áreas cotidianamente e precisa ter familiaridade com elas para ter uma maior produtividade e para conseguir fazer análises profundas de bases de dados e reportagens de grande interesse público.

E se especializar na área, seja com treinamentos presenciais ou virtuais, é importante também para estar mais bem preparado para suprir as necessidades das redações. Segundo Natália Mazotte (2016), a alta demanda por treinamentos com o enfoque no jornalismo de dados tem justificativa.

Os profissionais capacitados para lidar com dados têm sido mais valorizados entre editores e líderes de empresas de mídia. Em uma pesquisa recente divulgada pelo Reuters Institute, 76% de 130 editores e CEOs de iniciativas digitais de diferentes países afirmaram ser extremamente importante melhorar o uso de dados nas redações (MAZOTTE, 2016)¹⁴⁶.

Apesar de a demanda ser maior por profissionais desta categoria no mercado, ainda são poucos os jornalistas que detêm conhecimentos acerca da área. Não há dúvidas que há um crescimento quanto a isso, mas ainda é pequeno, principalmente quando comparado a países onde o jornalismo de dados é mais forte. No Brasil, são poucas as redações com equipes especializadas em dados.

O Espião Estatístico, objeto de estudo desta monografia, mostrou pioneirismo quando o assunto é trabalhar com dados esportivos no meio do futebol brasileiro. O trabalho consistente há pelo menos cinco anos fez com que a equipe produzisse conteúdos jamais feitos no país, como o Artilheiro da Hora, por exemplo. Mas, apesar disso, o Espião tem limitações como, por exemplo, a forma de armazenamento das informações, que é feito na planilha do Excel, mas, sem ela, o produto não existiria.

¹⁴⁶ Disponível em: <https://jornalismomonobrasilem2017.com/as-novas-fronteiras-do-jornalismo-de-dados-no-brasil-cddeede4fb8d>. Acessado em: 11/06/2017.

O Espião hoje é o que ele pode ser. A cada ano, a equipe vai sendo montada de acordo com as possibilidades. O Espião sempre foi aquilo que as pessoas que estão fazendo ele conseguem fazer. [...] Qualquer produto ou serviço sempre pode ser melhorado, seja sendo feito por pessoas mais competentes, mais criativas, com ferramentas mais apropriadas. A gente faz no Excel, mas ele não é a ferramenta para isso. A gente sabe que o Excel não foi feito pra isso, mas a gente usa e consegue tirar valor dele. [...] Às vezes os seus méritos são suas fraquezas. Uma fraqueza do produto é que ele é feito por pessoas, mas sem elas ele não existiria. Uma das fraquezas é que ele é feito em Excel, mas sem ele, ele não existiria. (STORTI, 2017).

Conforme expôs o criador do Espião acima, o produto trabalha de acordo com as próprias possibilidades. Apesar de o Excel não ser a ferramenta mais adequada, o Espião Estatístico cumpre com excelência as duas primeiras etapas da pirâmide invertida de Paul Bradshaw: compilação e limpeza. Isso porque a estruturação dos dados pensada para serem registradas de acordo com os lances de uma partida de futebol é única. O Espião conseguiu homogeneizar atos de um jogo e fazer deles dados para análises estatísticas.

A etapa de combinação ou cruzamento de dados também é feita no Excel, mas, conforme foi exposto no subcapítulo 3.1.4, há outras ferramentas que poderiam agilizar este processo. Isto não impede a entrega dos conteúdos pela equipe do Espião, que é feita regularmente e com eficácia. Mas é importante expor que usar o Excel nesta fase não é tão vantajoso, uma vez que não é uma ferramenta tão eficiente na análise de um banco de dados. No entanto, a equipe do Espião reconhece isto e já trabalha para buscar soluções que acelerem o trabalho e aumentem a produtividade. Uma parceria com matemáticos da Escola de Matemática Aplicada da Fundação Getúlio Vargas promete ser um meio para solucionar este “problema”.

Por fim, as etapas de contextualização e comunicação dos dados têm sido bem resolvidas pelos integrantes da equipe do Espião Estatístico na hora da apresentação dos conteúdos para o site. A visualização dos dados ainda depende muito do auxílio das equipes de Infografia e Desenvolvimento do GloboEsporte.com. Esta dependência impede que alguns conteúdos ganhem o formato pensado inicialmente. Por exemplo, fazer com que um conteúdo seja interativo depende da dedicação dos profissionais da equipe de programação do site e também da equipe de profissionais da infografia para a produção do layout do conteúdo. Com isso, caso estas duas equipes estejam muito atarefadas, o produto pode sofrer mudanças para que ele se torne menos complexo e, conseqüentemente, mais plausível de ser produzido pelas equipes citadas.

Portanto, nota-se que é necessário um esforço conjunto para que haja celeridade nas reportagens guiadas por dados do Espião. Talvez uma maneira para isso seja buscar sites e ferramentas que “resolvam o problema” das visualizações propostas pela equipe de dados do GloboEsporte.com. Enquanto isso, o Espião trabalha com o que tem para continuar produzindo os conteúdos de qualidade estatística valorável para o mundo do jornalismo esportivo brasileiro. Pode-se dizer que a equipe faz o que propõe o filósofo contemporâneo brasileiro Mario Sergio Cortella¹⁴⁷: “Faça o teu melhor, na condição que você tem, enquanto você não tem condições melhores, para fazer melhor ainda!¹⁴⁸”.

¹⁴⁷ Mario Sergio Cortella é um filósofo, escritor, educador e professor universitário brasileiro

¹⁴⁸ Link do vídeo do pensamento do filósofo: <https://www.youtube.com/watch?v=BsEH-DbtDB4>

6. REFERÊNCIAS BIBLIOGRÁFICAS

BRADSHAW, Paul. **How to be a data journalist**. The Guardian, 2010. Disponível em: <https://www.theguardian.com/news/datablog/2010/oct/01/data-journalism-how-to-guide>. Acessado em: 24/05/2017.

BRADSHAW, Paul. **The inverted pyramid of data journalism**. Online Journalism Blog, 2011. Disponível em: <<http://onlinejournalismblog.com/2011/07/07/the-inverted-pyramid-of-data-journalism>>. Acessado em: 17/04/2017.

CAIRO, Alberto. **Infografías y visualizaciones que simplifican em exceso o confunden datos**. In: Manual de Periodismo de Dados Iberoamericano. 2013. Disponível em: <http://manual.periodismodedatos.org/alberto-cairo.php>. Acessado em: 28/05/2017.

CAIRO, Alberto. **The Functional Art: An introduction to information graphics and visualization**. Berkeley: New Riders, 2013.

CONFERÊNCIA INTERNACIONAL DE DADOS ABERTOS. **Enabling the Data Revolution – An International Data Roadmap**. Relatório. Ottawa, 2015.

CRUCIANELLI, Sandra. Herramientas digitales para periodistas. Centro Knight para el Periodismo en las Américas de la Universidad de Texas, 2013. Disponível em: <https://knightcenter.utexas.edu/books/HDPP.pdf>. Acessado em: 14/05/2017.

CRUCIANELLI, Sandra. Periodismo de datos: cómo hacer historias con números. Scidev.Net, 2013. Disponível em: <http://www.scidev.net/america-latina/periodismo/guia-practica/periodismo-de-datos-c-mo-hacer-historias-con-numeros.html>. Acessado em: 21/05/2017.

DUTRA, Luma Poletti. **Lei de Acesso à Informação e jornalismo: usos e desafios**. In: II Seminário de Pesquisa em Jornalismo Investigativo, 2., 2015, São Paulo. Anais... São Paulo: ABRAJI, 2015. P.1-14. Disponível em: <http://www.abraji.org.br/seminario/PDF/2/luma_poletti_dutra.pdf>. Acessado em: 23/04/2017.

GARRISON, Bruce. **Computer-Assisted Reporting**. 2. ed. Londres: LEA Publishers, 1998.

GRAY, Jonathan; CHAMBERS, Lucy; BOUNEGRU, Liliana (orgs.). **Manual de jornalismo de dados: como os jornalistas podem usar dados para melhorar suas reportagens**. São Paulo: Abraji/EJC, 2013. Disponível em: <<http://datajournalismhandbook.org/pt/>>. Acessado em: 23/04/2017

HOUSTON, Brant. **Data Journalism: What Computer-Assisted Reporting Is and Why Journalists Use It**. In: HOUSTON, Brant. Computer-assisted reporting: a practical guide. 4. ed. Nova York: Routledge, 2015. Cap. 1. p. 3-16.

HOWARD, Alexander. **The art and science of data-driven journalism**: when journalists combine new technology with narrative skills, they can deliver context, clarity and a better understanding of the world around us. Nova York: TOW Center for Digital Journalism, 2014.

Disponível em: <<http://towcenter.org/wp-content/uploads/2014/05/Tow-Center-Data-Driven-Journalism.pdf>>. Acessado em 07/05/2017.

ÍNDICE DE DADOS ABERTOS PARA O BRASIL. Relatório. Rio de Janeiro, 2017. Disponível em: <<http://dapp.fgv.br/transparencia-orcamentaria/indices/IndiceDadosAbertosBrasil2017.pdf>>. Acessado em: 07/05/2017.

JANSEN, Thiago Pinheiro. **Data-driven journalism: o uso e a visualização de bases de dados na produção jornalística**. Rio de Janeiro, 2011.

MARÍN, Mauricio Jaramillo. **Guía de herramientas Google para periodistas**. Relatório. Google, 2010.

MEYER, Phillip. **The new precision journalism**. Bloomington: Indiana University Press, 1991. Disponível em: <https://www.attorneygeneral.jus.gov.on.ca/inquiries/cornwall/en/hearings/exhibits/Mary_Lynn_Young/pdf/07_Meyer.pdf>. Acessado em: 14/04/2017.

MEDEIROS, Simone Assis et al. **Lei de acesso à informação: em busca da transparência e do combate à corrupção**. Informação & Informação, Londrina, v. 19, n. 1, p.55-75, jan. 2014. Quadrimestral. Universidade Estadual de Londrina. <http://dx.doi.org/10.5433/1981-8920>. Disponível em: <<http://www.uel.br/revistas/uel/index.php/informacao/article/view/13520/14207>>. Acessado em: 17/04/2017.

PIRES, Marco Túlio. **Raspagem & Jornalismo de dados**. Escola de Dados, 2015. Disponível em: <https://escoladedados.org/manual/cursos/raspagem/raspagem-jornalismo-de-dados/>. Acessado em: 21/05/2017.

STAMILLO, Leonardo & CIPRIANI, Luis. **Programa de Treinamento do Twitter**. Apostila. Twitter, 2015.

STRAY, Jonathan. **The curious journalist's guide to data**. Columbia Journalism Review, 2016. Disponível em: <http://www.cjr.org/tow_center_reports/the_curious_journalists_guide_to_data.php>. Acessado em: 24/04/2017.

WARNES, Sophie. **What is data journalism? What does a data journalist do?** Medium. 2016. Disponível em: <https://medium.com/@sophiewarnes/what-is-data-journalism-what-does-a-data-journalist-do-bd914bed815b>. Acessado em: 14/05/2017.

WICKHAM, Kathleen Woodruff. **Math tools for journalists**. Marion Street Press: Oak Park, 2003. Disponível em: <<https://books.google.com.br/books?hl=pt-BR&lr=&id=RyYmMD2ReAC&oi=fnd&pg=PA10&dq=Math+for+journalists&ots=kAXGE-S4-2&sig=zVVv3mwx5FEJkirilO6jYUQd2s4#v=onepage&q&f=false>>. Acessado em: 13/05/2017.

7. APÊNDICE

7.1. FIGURAS

Figura 1: Página 15

Figura 2: Página 32

Figura 3: Página 32

Figura 4: Página 33

Figura 5: Página 33

Figura 6: Página 34

Figura 7: Página 35

Figura 8: Página 35

Figura 9: Página 36

Figura 10: Página 37

Figura 11: Página 63

Figura 12: Página 65

Figura 13: Página 73

7.2. ENTREVISTA COM SERGIO SPAGNUOLO CONCEDIDA AO AUTOR

ROBERTO MALESON: Primeiro, queria que você contasse um pouco sobre o Volt Data Lab e quais são os tipos de produto que vocês entregam?

SERGIO SPAGNUOLO: O Volt é uma agência de jornalismo de dados que produz, em uma ponta, pesquisas, investigações, levantamentos e reportagens para redações e projetos de mídia, e, em outro, inteligência para comunicação de ONGs, terceiro setor e até algumas empresas.

RM: Nesse contexto, quais as habilidades e técnicas você considera mais importantes para um que um jornalista consiga entregar esses tipos de produto?

SS: É primordial que um jornalista consiga trabalhar com números acima do que sua formação universitária ensina. Analisar dados é uma forma muito eficaz de produzir conhecimento e gerar boas pautas. Trabalhar com conceitos básicos de estatística é fundamental nessa área, assim como conseguir interpretar gráficos complexos e extrair informações de bancos de dados. Outra coisa, talvez a mais importante, é entender a

estrutura de dados: qual a melhor forma de organizar aquilo que você quer ver ou transmitir.

RM: Qual a importância da programação nisso tudo? E quais são as linguagens que você mais usa no seu trabalho?

SS: A programação permite que você analise, visualize e obtenha muito mais dados do que é possível fazer à mão, em termos de volume. No Volt, as linguagens que mais usamos são R (para análise), SQL (para query de dados), javascript (para visualização). Ainda estamos desenvolvendo melhor nossas capacidades em Python para extração de dados.

RM: Gostaria que você contasse como foi o seu processo de aprendizagem a programar. Como foi isso pra você? Fez cursos, leu livros, quais foram as etapas cumpridas?

SS: Eu aprendi na unha. Buscava tutoriais online, via exemplos em funcionamento e buscava replicar até funcionar para o que precisava. Eventualmente, a repetição leva ao conhecimento, e foi o que aconteceu. Sou jornalista primeiro, não programador, e ainda tenho muito a aprender. Mas aprender com exemplos é a melhor coisa. Sites como Stock Overflow e blogs com tutoriais específicos são uma benção para quem quer aprender a programar.

RM: Por fim, queria você desse algumas dicas a um aspirante a jornalista de dados e dissesse o porquê é importante saber programar ou saber as ferramentas que te ajudem nessa etapa.

SS: Minha dica é produzir. Vá atrás, produza gráficos, levantamentos, reportagens, mesmo que nenhum veículo vá publicar e que pouca gente vá ler, e mesmo que você tenha que fazer tudo sozinho. Cursos e workshops são excelentes para começar, mas a prática é a melhor forma de fazer, aprender e desenvolver habilidades de jornalismo de dados.