

ESTUDOS SOBRE VISUALIZAÇÃO DE DADOS CIENTÍFICOS NO CONTEXTO DA *DATA SCIENCE* E DO *BIG DATA*¹

Email:
adrianacontemporanea@gmail.com
guilhermeataide@gmail.com

Adriana Alves Rodrigues, Guilherme Ataíde Dias

Resumo

Investiga o impacto da visualização de dados científicos e seus desdobramentos no contexto do *Data Science* e da Ciência da Informação por meio de estudos de casos. A pesquisa explora uma abordagem teórico-conceitual a partir da Ciência da Informação e suas relações interdisciplinares - *Big Data*, Ciência de Dados e Visualização de Dados. Considera que o cenário estabelecido e desenvolvido pelo *Data Science* é propício para a geração de novos formatos de visualização de dados de impacto científico e de difusão social através das redes digitais. A metodologia utilizada para a pesquisa é estruturada a partir de uma abordagem quali-quantitativa e se constitui em pesquisa exploratória e de estudos de casos empíricos representativos do fenômeno a partir de observações sistemáticas dos objetos para análise utilizando-se para tal de instrumentos como ficha de observação, visando a captura, tratamento e visualização e construção de modelo de análise. Resultados preliminares indicam que a visualização de dados tratam-se de novos formatos e narrativas inovadoras com aspectos disruptivos.

Palavras-chave: Visualização de Dados. Ciência de dados. Inovação Disruptiva. Mega dados. Ciência da Informação.

Abstract

It investigates the impact of the visualization of scientific data and its unfolding in the context of Data Science and Information Science through case studies. The research explores a theoretical-conceptual approach from the Information Science and its interdisciplinary relationships - Big Data, Data Science and Data Visualization. It considers that the scenario established and developed by Data Science is conducive to the generation of new formats of data visualization of scientific impact and social diffusion through the digital networks. The methodology used for the research is structured from a qualitative-quantitative approach and consists of exploratory research and empirical case studies representative of the phenomenon from systematic

¹ Pesquisa de Doutorado em andamento no Programa de Pós-Graduação em Ciência da Informação na Universidade Federal da Paraíba –PPGCI-UFPB.

observations of the objects for analysis using instruments such as data sheet Observation, aiming at the capture, treatment and visualization and construction of analysis model. Preliminary results indicate that data visualization is about new formats and innovative narratives with disruptive aspects.

Keywords: Data Visualization. Data Science. Disruptive Innovation. Big Data. Information Science.

INTRODUÇÃO

Num mundo cada vez mais baseado na cultura de dados, as tecnologias digitais que permitem coleta, produção e circulação tornam-se cruciais para a exploração e a compreensão dos dados, mais particularmente dos grandes volumes de dados (*Big data*). O emprego das técnicas de visualização de dados (*Dataviz*) nesse campo é uma estratégia emergente e de inovação que tem sido explorado em distintas áreas do conhecimento e recebido cada vez mais atenção em estudos e pesquisas na atualidade (CAIRO, 2012; MANOVICH, 2011; DUR, 2014; JL VALERO, 2014). A visualização de dados trata-se de uma nova caracterização da explosão informacional através da era do *Big Data*. Se a preocupação inicial da Ciência da Informação era, entre outros aspectos, em como coletar, organizar, armazenar e recuperar as informações, na atualização do fenômeno enquanto problema de pesquisa decorre agora em como transformar esses dados em visualização de modo a oferecer uma nova interpretação e significado de dimensão social, informacional e comunicacional.

As investigações em visualização de dados ampliaram seu alcance com as tecnologias digitais da sociedade em rede e da cibercultura (LEMOS, 2002) que possibilitaram que os dados invisíveis, isto é, armazenados em redes digitais, somados aos demais elementos combinatórios (gráficos, mapas, dados científicos, números etc.), fossem transmutados para o ambiente Web e conferissem um maior enriquecimento às estruturas gráficas. Sendo assim, percebemos o vigor do conceito da visualização de dados como central e operacional para a compreensão do fenômeno dos grandes volumes de dados na Ciência da Informação centrado na cultura do *Big Data* (BARLOW, 2013) e do seu impacto tecnológico, cultural e social. Logo, há uma abrangência das apropriações e usos do *Big Data* na economia, comunicação, administração, ciência da informação, medicina, e áreas correlatas que lidam com grandes volumes de dados.

Nesta conjuntura, *Big Data* reivindica uma abordagem inovadora para lidar com os dados científicos. Logo, tratamos a questão por meio do conceito de inovação disruptiva como pertinente para problematizar o horizonte das produções em visualização de dados na perspectiva de novo paradigma que representa o *Big Data*, tendo em vista tratar-se de um contexto ou de uma tecnologia disruptiva. A reflexão que empreendemos visa ampliar o escopo da discussão no âmbito da Ciência da Informação e o aspecto global demarcando e compreendendo o objeto em suas múltiplas dimensões. Posicionamos nosso objeto de estudo como um formato emergente que suscita o estabelecimento de novos padrões de apresentação com investigação focada nas dinâmicas e rupturas que envolvam as mudanças nas linguagens e modos de produção e apropriação das narrativas no tocante ao *Big Data* dentro da Ciência da

Informação. Para esta abordagem convocamos o conceito de Ciência de Dados como instância que escrutina o *Big Data* e reflete sobre o impacto deste na sociedade.

A partir da discussão empreendida, buscamos responder na pesquisa ao seguinte questionamento: Que implicações a Visualização de Dados científicos representa para a produção de narrativas disruptivas quanto à dinâmica do contexto do *Data Science* e do Campo da Ciência da Informação? Partimos do pressuposto de que a Ciência de Dados se constitui em um ferramental teórico e metodológico, ainda em andamento, vital para a análise e construção das visualizações de dados científicos no horizonte visando irromper para inovações disruptivas as estruturas visuais no ambiente digital. Sendo assim, defendemos de que as visualizações de dados científicos no contexto do *Data Science* provocam narrativas complexas disruptivas e geram novos formatos com impacto científico tecnológico e informacional, trazendo novas dinâmicas e formas de compreensão dos dados, tendo em vista que a entrada do *Big data* nas visualizações de dados é um fator que pode reconfigurar tais narrativas. Para tal, analisamos o fenômeno a partir da Ciência de Dados.

2 CIÊNCIA DE DADOS

A Ciência de Dados (*Data Science*) tem sido estudada e considerada como uma área com característica interdisciplinar por parte dos pesquisadores da temática (CONWAY, 2010; STANTON, 2012; ZHU, XIONG, 2015; STREIB, MOUTARI, DHEMER, 2016) ou multidisciplinar (TIERNEY, 2016) em sua origem. Para Porto e Ziviani (2014, p. 2), dentro da conjuntura de grandes volumes de dados, há três linhas de pesquisa que podem ser exploradas com vistas à consolidação da área de Ciência de dados, a saber: 1) Gerência de Dados; 2) Análise de Dados e 3) Análise de Redes Complexas. Conforme alertam: “[...] nesses aspectos fundamentais de análise de dados em larga-escala, há também um grande potencial tecnológico na pesquisa aplicada em ciência de dados com impacto em diferentes áreas do conhecimento e de setores de atuação” (PORTO; ZIVIANI, 2014, p.2). A Ciência de Dados apresenta caráter híbrido com ressignificação dos dados, novas modelagens, metodologias, arquiteturas da informação, bem como a fluidez para novos formatos para o campo da visualização de dados.

A área utiliza métodos e técnicas semelhantes ao da Ciência da Computação, que incluem aquisição dos dados, gestão, armazenamento, segurança, análise e visualização de dados, sendo que de modos diferenciados. Nesta perspectiva, Conway (2010) criou o Diagrama de Venn para especificar as habilidades que concerne à área (Figura 1) abarcando outras disciplinas. No diagrama criado por Conway, a *Ciência de Dados* aparece no centro, em lugar de destaque, indicando a ascensão da área e a correlação com outras capacidades como habilidades hacker, conhecimento de matemática e estatística e expertise substantiva, além de aprendizagem de máquinas. O autor entende que apenas focar na Ciência de Dados não é o suficiente para a compreensão da área em sua totalidade, mas integrar com outras áreas do saber.

Figura 1: Diagrama de Venn desenvolvido por Conway



Fonte: Conway (2010, tradução nossa).

No Diagrama de Venn, aspectos como conhecimentos em estatística e matemática e habilidades *de hacking e aprendizagem de máquinas* aparecem em aspectos cruzados para operacionalizar e gerenciar grandes quantidades de dados em contexto de Ciência de Dados. Deste modo, o Diagrama mostra que é imprescindível habilidades e conhecimentos profundos nessas duas áreas do conhecimento, e, portanto, sem eles, torna-se quase que impossível enveredar por esse caminho. Na zona de perigo, é destinado “para as pessoas que são perfeitamente capazes de extrair dados e estruturação, provavelmente relacionado a um campo que sabem um pouco sobre, mas falta-lhes qualquer compreensão do que esses coeficientes querem dizer” (CONWAY, 2010, *online*, tradução nossa)². O nome *zona de perigo* soa como um alerta para que os profissionais, pesquisadores estejam totalmente aptos para desempenhar suas funções no *Data Science*, tendo, assim, uma visão mais completa dessa nova área em expansão.

Em seguida, o diagrama aponta pesquisas tradicionais ao lado do nome *Data Science*, em um lugar sem pouco destaque. E aqui vemos um equívoco: o de introduzir pesquisas tradicionais quando a emergência da área requer pesquisas inovadoras para o tratamento dos dados, gerenciamento, armazenamento, etc. Áreas emergentes exigem que os procedimentos metodológicos sejam reconfigurados, adaptados e até mesmo criar novas infraestruturas que sejam exploradas as potencialidades do novo fenômeno. Deste modo, há um reposicionamento das pesquisas científicas trazido pelo *Data Science* e de como gerenciar os dados em ambiente digital, estabelecendo pesquisas transversais e explorando novas zonas de interlocução dos resultados e práticas de pesquisa.

Com o intuito de resolver problemas na era do *Big Data* e da Ciência de dados, deve-se incorporar os seguintes aspectos para a garantia de uma gerência adequada: ter infra-estrutura de grandes dados, ciclo de vida e análise de dados, habilidades de gerenciamento de dados e disciplinas comportamentais. A infra-estrutura tenta abarcar tecnologias que operacionalizam os grandes volumes de dados, grandes bancos de dados com memória expandida e computação em nuvem. Neste sentido, o ciclo de vida dos dados inclui todas as etapas de análise de dados, incluindo negócios análise, a compreensão, preparação de dados e integração, construção de

² No original: "know enough to be dangerous," and is the most problematic area of the diagram. In this area people who are perfectly capable of extracting and structuring data, but they lack any understanding of what those coefficients mean".

modelos, avaliação, implementação e monitorização, bem como habilidades de gerenciamento de dados que incluem tradicional modelagem de dados e conhecimento de banco de dados relacional (ZHU; SONG, 2016) que possam transformar as visualizações de dados dentro da Ciência da Informação a partir desse fenômeno que emerge.

3 VISUALIZAÇÃO DE DADOS

A visualização de dados vem se tornando mais frequente, tanto do ponto de vista de abordagem acadêmica, quanto do ponto de vista de alargamento dos usos na mídia impressa e digital, tornando-se comuns como modelos que visam à representação visual de grandes volumes de dados. Na literatura, existem inúmeras terminologias e definições sobre a visualização de dados (VIÉGAS, 2013; CAIRO, 2011; MEIRELES, 2011; RODRIGUES, 2009, MANOVICH, 2012). De uma maneira mais abrangente, a visualização é o resultado de uma tecnologia plural que transforma dados complexos em informação semântica e facilita a interação através de ferramentas para que qualquer usuário complete o processo de modo autônomo. Acrescente-se a essa afirmação o fato de que a visualização se revela em constante evolução e atualização, sendo “uma conjugação de signos de natureza icônica (figurativos) com outros de natureza arbitrária e abstrata (não figurativos: texto, estatísticas, etc)” (CAIRO, 2012, p. 38).

Meireles (2011, p. 2) conceitua a visualização de dados como “representações de dados que pode assumir diferentes formas, tais como sistemas de notação, mapas, diagramas, explorações de dados interativos, e outras invenções gráficas”. A visualização de dados é o processo de utilização de tecnologias mediadas por computador e digitais para exibir informações quantitativas e qualitativas. Neste sentido, o termo "visualização de dados" é usado aqui para especificar este processo de transformar dados numéricos ou informativos científicos para distinguir a visualização de dados de outras visualizações, tais como a arte não baseada em dados. As visualizações de dados estão cada vez mais se complexificando para narrativas sofisticadas que se utilizam de mapas com dados que permitem interação, conforme defende Fernanda Viégas (2013).

A partir dessas definições, percebemos que a visualização de dados ocupa cada vez mais um papel central na sociedade contemporânea em relação a "explosão de dados" como novo fenômeno da Ciência da Informação. A conjugação entre *Big Data* e visualização de dados delimita uma nova fronteira para o campo da Ciência da Informação, recuperando ou complexificando a pertinência da CI em estudar fenômenos relacionados ao fluxo informacional em sistemas informatizados que lidam com a recuperação e processamento de informações. Partindo desse arcabouço teórico, definimos o conceito de visualização de dados como a técnica de transformar um conjunto complexo de dados em visualizações gráficas de modo a constituir uma representação visível dos dados que estavam “invisíveis” e que passam a ser manipulados por algoritmos em sistemas computacionais para a estruturação de um conteúdo. Operacionalizamos o nosso objeto de estudo a partir da compreensão da visualização de dados por considerar pertinente para o contexto de dados em abundância na ambiência digital.

É pertinente saliente um aspecto dessa definição o fato das visualizações estarem para além de mostrar dados ou números, estatísticas, puros e simples: contam histórias através dos dados. Essa possibilidade ressaltada por Viégas conduz a uma reflexão da importância dos

dados e, que estes dados podem ser aprofundados numa visualização, já que são complexas e tendem a ter uma malha informacional maior que não se restringe a apenas *apresentar*, mas *explorar* e *analisar*. Soma-se a essa afirmação que é preciso contextualizar os dados visando estabelecer sentido, em direção a uma eficácia comunicativa em que “gráficos diferentes contam histórias diferentes” (GRAY; CHAMBERS; BOUNEGRU, 2012, *online*). Como alertam Segel e Heer (2010), às vezes os dados por si só não contam história de maneira mais convincente, mas deve haver uma narrativa que relaciona as consequências reais e causar impacto no interagente.

Segel e Heer (2010) estabelecem dois parâmetros que auxiliam na decodificação dos dados complexos: a) Visualização assistida por informações (*Information-assisted visualization*): É fornecido ao usuário um segundo formato de visualização que normalmente exhibe as informações sobre um conjunto de dados, mas também pode apresentar atributos da visualização do processo, das propriedades dos resultados, ou das características dos comportamentos de percepção do usuário. Com o aumento de tamanho e complexidade de dados, a utilização de informações para ajudar a visualização será inevitável, e se tornará uma necessidade e não uma opção; b) Visualização assistida por conhecimento (*Knowledge-assisted visualization*). Nesta parte, o conhecimento do usuário é um aspecto indispensável, uma vez que este pode atribuir cores e outros elementos dependendo do seu domínio de conhecimento.

Para Murray (2013), a visualização de dados também se configura num campo interdisciplinar, e que na era dos grandes volumes de dados há uma "sobrecarga" que precisa ser decodificada de um modo compreensível, em que o leitor possa entender visualmente a mensagem que se queira passar. "A visualização de dados é um exercício poderoso. A visualização de dados é a maneira mais rápida de comunicá-la aos outros" (MURRAY, 2013, p.1). Por ser uma ferramenta ponderosa, a mesma pode ser usada para distorcer ou mascarar informações verdadeiras. Contudo, ressalta o autor, se utiliza eticamente, o processo de visualização pode auxiliar a enxergar o mundo de uma outra maneira, revelar padrões ocultos e tendências que não estão disponíveis de uma outra forma. Em outras palavras, a visualização de dados tem como especialidade contar histórias. Além disso, "visualização é um processo de mapeamento de informações para imagens"(p.1). Para isso, criam-se regras e metodologias que interpretam dados numa dinâmica muitas vezes complexa.

Partindo da questão de como a visualização de dados se comporta no âmbito da Ciência da Informação, delimitamos nossa situação-problema buscando investigar o seu impacto a partir da interface com o *Big data* e a Ciência de Dados em termos de construção de narrativas disruptivas e seu impacto em termos de representação da informação por meio das tecnologias e estratégias de visualização disponíveis para exploração de mega dados.

4 PROCEDIMENTOS METODOLÓGICOS

Para o percurso metodológico da pesquisa, procedemos com algumas estratégias (quadro 1) visando compreender o fenômeno e o objeto delineado. Para tal empreendemos uma pesquisa exploratória para dimensionamento da visualização de dados científicos e do fenômeno do *Data Science* no campo da Ciência da informação. Ao mesmo tempo, procedemos com uma pesquisa bibliográfica para estabelecimento do estado da arte e recuperação histórico e de definições conceituais sobre visualização de dados científicos e dos conceitos-chaves desta tese como *Big Data*, Ciência de Dados e visualização de dados.

bibliográfica e atenção ao referencial teórico utilizado para compreensão teórico-conceitual do fenômeno.

Após a coleta, submeteremos os resultados à análise e interpretação à luz do referencial teórico, numa abordagem quali-quantitativa que possa demonstrar as diferentes dimensões do fenômeno e a comparação entre os casos em estudo. Por meio destes parâmetros formulamos algumas etapas e estratégias para execução da pesquisa. Primeiro, realizaremos um estudo exploratório para identificar as especificidades, configurando como um mapeamento dos casos de visualização de dados científicos, de modo a delimitar mais adequadamente o objeto e mapear o campo, além de uma ampla revisão de bibliografia para definições teórico-conceituais do objeto e estabelecimento do estado da arte. Esta etapa já encontra-se em andamento desde o início da tese e das orientações e prosseguirá em todo o seu decorrer como processo intrínseco da revisão de literatura e de monitoramento do fenômeno considerando o aspecto de inovação envolvido.

Segundo, o *corpus* da pesquisa está sendo selecionado a partir de casos empíricos representativos e vinculados à perspectiva de inovação e de uso de dados científicos e dados abertos para observação à luz da Ciência de Dados com desdobramento para a Ciência da Informação. A escolha desses casos levará em consideração a representatividade dos mesmos por meio da delimitação do objeto de tese e da perspectiva do panorama com o desenvolvimento sistemático de narrativas com a exploração dos elementos da visualização de dados na *Web*.

A terceira etapa se refere à definição conceitual dos objetos selecionados. Após a revisão de literatura e seleção do *corpus*, é necessário o estabelecimento das categorias de análise e construção dos conceitos. “Toda construção teórica [...], é um sistema cujos eixos são os conceitos e as categorias de análise, unidade de significação que definem a sua forma e o seu conteúdo” (PALACIOS; MACHADO, 2007, p. 207).

Neste sentido, pretendemos remeter a observação e a análise aos objetos empíricos de referência que representem experiências no que se refere à visualização de dados científicos. Devido à natureza mutante do objeto, a pesquisa pretende seguir os rastros dessas experiências para poder delimitar os formatos e linguagens adotadas nas visualizações. A análise e interpretação de dados se darão à luz da análise de conteúdo (BARDIN, 2009) e numa abordagem quali-quantitativa para demonstrar as diferentes dimensões do fenômeno.

5 CONSIDERAÇÕES FINAIS

Neste sentido, pretendemos contribuir para o Campo da Ciência da Informação com pesquisa que complexifique a relação entre visualização de dados científicos e *Big data* a partir da perspectiva da Ciência de Dados. Em decorrência da emergência do fenômeno, compreendemos que estamos diante de um desafio para a cultura de dados na Ciência da Informação e, ao mesmo tempo, contribuição para o estado da arte sobre a temática que permita avançar em termos de compreensão sobre novos formatos de visualização.

A investigação caminha, dentro do seu plano de trabalho, para o aprofundamento da coleta de novos dados que permitam a caracterização das visualizações de dados em contexto de Ciência de Dados e da Ciência da Informação e seus desdobramentos para os estudos sobre

a cultura dos dados e contribuição para o estado da arte da temática a partir dos objetivos propostos.

REFERÊNCIAS

BARDIN, L. **Análise de Conteúdo**. Lisboa, Portugal; Edições 70, LDA, 2009.

BARLOW, M. **The culture of Big Data**. Sebastopol-CA: O'Reilly Media, 2013.

BORGMAN, C. **Big data, little data, no data**. Scholarship in the networked world. Cambridge-London: The MIT Press, 2015. (Kindle version).

CAIRO, A. **El arte funcional: infografía y visualización de información**. Madrid: Alamut, 2012.

CONWAY, D. (2003). **The data science venn diagram**. Disponível em: <http://migre.me/vII5x>
Acesso em: 22 out 2016.

DUR, B. U. **Data Visualization and Infographics in Visual Communication Design Education at the Age of Information**. TOBB University of Economics and Technology, Faculty of Fine Arts, Design & Architecture, Department of Visual Communication Design. 2014.

GRAY, J. BOUNEGRU, L. CHAMBERS, L. (Ed.). **The Data Journalism Handbook**. How Journalists Can Use Data to Improve the News. Sebastopol: O'Reilly Media, 2012.

LEMOS, A. **Cibercultura, tecnologia e vida social na cultura contemporânea**. Porto Alegre: Sulina, 2002.

MANOVICH, L. **Trending: The Promises and the Challenges of Big Social Data**. Disponível em: <http://goo.gl/IqIlgGF> Acesso em: 3 jun. 2015.

MEIRELES, I. **Visualizing data: new pedagogical challenges**. Spinillo, Farias & Padovani (Eds). Selected Readings of the 4th Information Design International Conference. São Paulo: SBDI | **Brazilian Society of Information Design**, 2010.

MURRAY, S. **Interactive Data Visualization for the Web**. Sebastopol: O'Reilly Media, 2013.

PALACIOS, M; MACHADO, E. **Modelos de jornalismo digital**. Salvador: Calhandra/Edições GJol, 2003.

PORTO, F.; ZIVIANI, A. **Ciência de Dados**. 2014. Disponível em <https://goo.gl/OomBCW>. Acesso em 24 dez. 2016.

RODRIGUES, A. A. **Infografia em base de dados no jornalismo digital**. (dissertação de mestrado) - FACOM, UFBA, Salvador, 2009. Disponível: <https://goo.gl/s7XICG> Acesso em: 20 maio 2014.

SEGEL, E; HEER, J. Narrative visualization: Telling stories with data. *Visualization and Computer Graphics*, **IEEE Transactions on**, 16(6):1139–1148, 2010

STANTON, J. **An Introduction data science**. Syracuse University's School of Information Studies, 2012.

STREIB, F. E; MOUTARI, S; DHEMER, M. **The Process of Data in the Emergent Feature of Data Science**. *Frontiers in Genetics*. Volume 7 February 2016.

TIERNEY, B. **Data Science is multidisciplinary**. 2016. Disponível em: <http://migre.me/vlsaS>
Acesso em: 25 out 2016

VIÉGAS, F. **Designer explica como a visualização de dados pode ser atraente**. Rio de Janeiro, TV Globo, 20 jun. 2013. Entrevista a Globo Universidade. Disponível em: <http://goo.gl/c35Ztn> Acesso em: 4 jul. 2016.

ZHU, Y. Y; XIONG, Y. (2011) **Dataology and Data Science: Up to Now**. Retrieved from the World Wide Web November 16, 2014: <https://goo.gl/UcL6Ca> Acesso em: 22 out. 2016.