

# NARRATIVAS DE VISUALIZAÇÕES GUIADAS POR DADOS: *Big Data* e inovação

## *Data-Driven Visualizations Narratives: Big Data and Innovation*

Adriana Alves Rodrigues<sup>1</sup>

(1) Universidade Federal da Paraíba, João Pessoa, [adrianacontemporanea@gmail.com](mailto:adrianacontemporanea@gmail.com)

### Resumo:

Analisa as narrativas de visualização de dados do prêmio *Data Journalism Awards 2018*, bem como seus processos de inovação a partir do cenário de *Big Data*. A metodologia, de caráter exploratório, compreende um Estudo de Caso e discussão teórico-conceituais e empíricas dessas narrativas e suas dimensões no uso de dados a partir de revisão de literatura e aplicação de Ficha de Observação na coleta de dados e análise quali-quantitativa. Os resultados, que fazem parte de uma pesquisa maior em andamento na tese doutoral, demonstram que as narrativas guiadas por dados se constituem em uma forma diferenciada de contar histórias no jornalismo em interface interdisciplinar com a Computação, Design e Ciência da Informação. Conclui-se que a emergência do *Big Data* e da Visualização de Dados aponta para novo formato de narrativa em que os dados exercem função central.

**Palavras-chave:** Visualização de dados; Narrativas; Dados; Big Data

### Abstract:

This paper presents an analysis of data visualization narratives from the 2018 Data Journalism Awards, as well as their innovation processes from the Big Data scenario. The exploratory methodology comprises a case-study and theoretical-conceptual and empirical discussion of these narratives and their dimensions in the use of data from a literature review and the application of an observation sheet in the data collection and qualitative-quantitative analysis. The results, which are part of a larger research project underway in the doctoral thesis, show that data-driven narratives constitute a new way of telling stories in journalism in an interdisciplinary interface with Computing, Design and Information Science. The research finds out that the emergence of Big Data and Data Visualization points to a new narrative format in which the data play a central role.

**Keywords:** Data Visualization; Narratives; Data; Big Data

## 1 Introdução

Em um contexto permeado por grandes volumes de dados, identificados pelo conceito de *Big Data*, as produções de visualizações de dados incorporaram novas linguagens e formatos, tendo como desdobramento a emergência de um novo fenômeno de natureza interdisciplinar. Para Lagoze (2014), o fenômeno do *Big Data* não se restringe a grandes volumes de dados, mas sim, a resultado da mescla de dimensões e características diferenciadas cujos efeitos podem ser revolucionários na produção de conhecimento. Neste sentido, a visualização de dados é uma área que cada vez mais ganha espaço nos debates profissionais e acadêmicos nos mais distintos campos do saber. Para além dos fatores tecnológicos que estão presentes, muitas das

visualizações de dados se pautavam, até então, a apenas “mostrar os dados” (TUFTE, 2001), ou seja, eram produções que se limitavam em explorar aqueles dados por si só. No entanto, a agregação de novas tecnologias e estratégias de estruturação de dados e, principalmente, do *Big Data*, alguns trabalhos vêm apontando como aspecto inovador o fato desses dados conterem elementos de uma narrativa (SEGEL; HEER, 2010; CAIRO, 2011; VIÉGAS, 2013; KOSARA; MACKINLAY, 2013; FIGUEIRAS, 2013) e níveis de complexidade (com os cruzamentos de dados e volume) e deste modo se apresentarem de uma forma mais dinâmica. Viégas (2013) afirma que as visualizações têm a capacidade de captura de um determinado fato da realidade com profundidade através dos aspectos gráficos e

diferentes cruzamentos de dados. Essa exploração do formato de visualização de dados tem ocorrido no campo científico e em organizações jornalísticas como *The New York Times* (EUA), *The Guardian* (Reino Unido), *El País* (Espanha), Estadão (Brasil), entre outros.

Em uma investigação com um conjunto de visualizações, Segel e Heer (2010) definiram alguns aspectos que podem subsidiar na composição de narrativa com dados, como por exemplo, apresentação de *slides* em movimento e histórias com detalhamento, isto é, com profundidade informativa, cujo foco maior, foi equilibrar o dados com os aspectos interativos. Neste contexto de narrativas e propriedades de dados em sua construção, a investigação de Nora Paul (2012) estabelece tipologias para as narrativas digitais, a saber: 1) Mídia: suporte da narrativa que combina recursos digitais como texto, fotos, gráficos, animação); 2) Ação: refere-se a dois aspectos distintos das narrativas digitais: o movimento do próprio conteúdo e a ação requerida pelo usuário para acessar o conteúdo; 3) Relacionamento: está ligado ao relacionamento entre o usuário do conteúdo e o próprio conteúdo. 4) Contexto: Capacidade de ofertar dados extras, remetendo a outros materiais, a exemplo dos *links*. e 5) Comunicação: Capacidade de conectar com outros tipos de mídias.

A pesquisa partiu da questão: Como as narrativas guiadas por dados endereçam formatos inovadores para a visualização de dados? Portanto, pretendemos observar e caracterizar esses elementos agregados às visualizações de dados e *Big data* visando compreender a perspectiva no contexto aqui explorado.

## 2 Objetivos

Esta investigação tem como objetivo compreender o processo de utilização dos

elementos narrativos que estão presentes nas visualizações de dados, realizando um estudo exploratório, a partir das categorias de análise: Elementos gráficos, interação, Dados estruturados, Proveniência dos dados, Tecnologias usadas e Inovação.

## 3 Procedimentos Metodológicos

Para atender o objetivo deste estudo, procedemos com um Estudo de Caso e nos baseamos na literatura específica além do aspecto empírico da observação e análise dos recursos narrativos nas visualizações de dados vencedoras do *Data Journalism Awards 2018*<sup>1</sup>, evento que acontece anualmente desde 2012, e esse ano foi em Lisboa, Portugal, e premia os melhores trabalhos com dados em jornais, revistas e agências de notícias mundiais<sup>2</sup>. Em 2018, foram 630 trabalhos submetidos ao prêmio de 58 países. Integralizando um total de 86 finalistas, com 13 trabalhos vencedores. Partimos para nossa pesquisa dessa amostra de 13 trabalhos e delimitamos um *corpus* de 8 visualizações que atendiam ao recorte ou escopo de jornalismo de dados. Descartamos do nosso estudo as produções que não se adequam à nossa amostra, como *startup*, *app*, *sites* e portfólios. O procedimento envolveu revisão de literatura, monitoramento e aplicação de Ficha de Observação<sup>3</sup> submetidas para essas visualizações no período de 17 a 26 de setembro de 2018. Consideramos as seguintes categorias de observação, quais sejam: **Elementos gráficos de visualização de dados** para averiguar a presença de ilustrações, vídeos, gráficos estáticos, gráficos dinâmicos, áudios, textos curtos, fotos estáticas, fotos em movimentos, imagens em 3D, mapas

<sup>1</sup> Disponível em: <https://goo.gl/TeKwkw> Acesso em: 14 set 2018.

<sup>2</sup> A competição é organizada pelo *Global Editors Network* e tem apoio do *Google News Initiative* e *Knight Foundation*.

<sup>3</sup> Disponível em <https://goo.gl/CBFr1w> acesso em 3 out.2018

dinâmicos ou estáticos, dentre outros, de modo que facilitem a narrativa; **Interação:** Aspecto fundamental nas visualizações, Cairo (2008) define três aspectos: Instrução (nível básico, quando a narrativa é linear e apresenta botões de avançar e retroceder); Manipulação (Poder mudar os objetos na visualização); e Exploração (Tipo de interação por imersão, ex. imagens de 360, com navegação hipertextual); **Dados Estruturados (ou semânticos):** A partir dos estudos de Cairo (2011), essa categoria se refere em apresentar os dados de modo compreensível e com sentido a esse conjunto de dados, se está estruturado de modo compreensível a partir dos formatos (verticalizados, horizontalizados). “Também criamos informação estruturada quando desenhamos gráficos que codificam os números, o que de outra maneira seria impossível extrair algo útil” (CAIRO, 2011, p. 31, tradução nossa)<sup>4</sup>; **Proveniência dos dados:** Identificar a origem dos dados que foram explorados na visualização, se foram dados públicos, privados, de empresas, agências de notícias, etc. A ideia é identificar como esses dados estão sendo explorados; **Tecnologias usadas:** Identificar quais foram as tecnologias e linguagens mais utilizadas para a execução das produções de visualização de dados: E **inovação** como construto dessas narrativas em base de dados com estratégias de uso de novos recursos.

#### 4 Resultados

A partir dos dados coletados<sup>5</sup> da observação e do procedimento da análise dos mesmos, foi possível identificar que a utilização de elementos gráficos nas visualizações fazendo composição com os

dados (exemplo de mapas, infografias, fotos e interações) para dar vazão a camada de dados e de construção de visualizações atrativas que possam dar sentido a dados complexos e cruzamento de dados. Nos 8 casos analisados (Apêndice A), os mapas estáticos predominam com 87,5% como lugar para receber dados, com exceção do *PostMedia*, que não trouxe nem mapas estáticos e nem dinâmicos. Em seguida, o uso de vídeos enquanto recurso das narrativas com 75% de uso. Em compensação, tivemos 4 gráficos interativos em detrimento de 3 gráficos estáticos. Gráficos guiados por dados são 2 (*Infotimes* e *RunRun*), em outros momentos os dados vêm incorporados em outros elementos.

No aspecto interação, como recurso que permite maior participação do usuário no manejo desses dados e de toda a narrativa, identificamos produções mais avançadas e com uso mais intenso de características de interação com o predomínio do tipo de interação por Exploração, total de 100% dos casos analisados (8 de 8), além de 2 casos de Manipulação (*PostMedia* e *Infotimes*, que também concomitantemente apresentaram o grau de exploração, totalizando 25% para esse item). Nenhum caso foi registrado do nível mais básico, o de Instrução. Essa possibilidade de interação é um aspecto pertinente para o formato de visualização de dados porque quanto maior o grau de interação, mas aumenta a possibilidade de cruzamento de dados em nível de personalização pelo usuário. Logo, constitui-se de um recurso fundamental na construção atual dessas visualizações.

Em relação aos dados estruturados (Apêndice C), as narrativas de visualização de dados têm se sofisticado e utilizado novas possibilidades de apresentação. Identificamos a tendência de verticalização, totalizando 87,5% da amostra (7 casos)

<sup>4</sup> No original: “También creamos información estructurada cuando dibujamos gráficos que codifican los números, lo que de otra manera sería imposible extraer algo útil”.

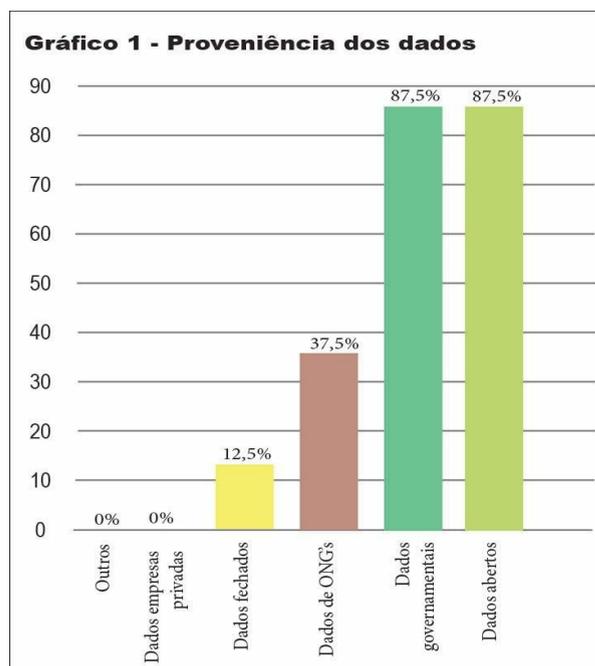
<sup>5</sup> Disponível em <https://goo.gl/HoVvAu> acesso em 3 out. 2018.

contra 2 casos (25%) horizontais. O modelo vertical se tornou tendência a partir do *SnowFall* do *The New York Times*, que a partir de 2012<sup>6</sup> iniciou o fenômeno da verticalização como formato para reportagens com dados permitindo aprofundamento para uso de recursos diversos como mapas, vídeos, mapas guiados por dados e outros recursos multimídia que dialogam diretamente com os dados em grande volume. A partir de então, os demais jornais mundiais começaram a aderir a esse formato, cuja “abundância do texto verbal sinaliza um resgate da qualidade, apuração e contextualização” (LONGHI; WINQUES, 2015, p. 8).

Além das narrativas verticais, os resultados apontaram para 75% (6 dos 8 casos analisados) das produções baseadas em mapas guiados por dados, cujos dados ficaram visíveis, seguidos de 25% do formato horizontalizados. Essas construções podem apontar tendência no jornalismo e na Ciência da Informação como formatos de apresentação. Esse tipo de formato também é denominado de narrativa *longform*, “um nível mais aprofundado de relato, que vai além do padrão cotidiano da produção e narrativas atraentes, frequentemente com elementos multimídia, que realçam o artigo” (LONGHI; WINQUES, 2015, p. 3).

As narrativas de visualização guiadas por dados se caracterizam, na origem das fontes, pelo uso de dados abertos (gráfico 1). Dos 8 casos analisados de visualização de dados, a maioria (7, ou seja, 87,5%) utilizam dados abertos. Igualmente o uso de dados governamentais (87,5%) e 3 casos de uso de dados de ONGs e similares (37,5%).<sup>7</sup> Dados fechados apenas no caso da agência *Reuters* (12,7%), que concomitantemente

também utiliza dados abertos e governamentais nas suas produções. Portanto, nesse quesito os dados abertos são predominantes nos projetos de visualização de dados apresentados demonstrando avanço nesta política. Buscamos, nesse caso, compreender o avanço do uso de dados abertos e governamentais e sua abundância para construção de narrativas ou projetos.



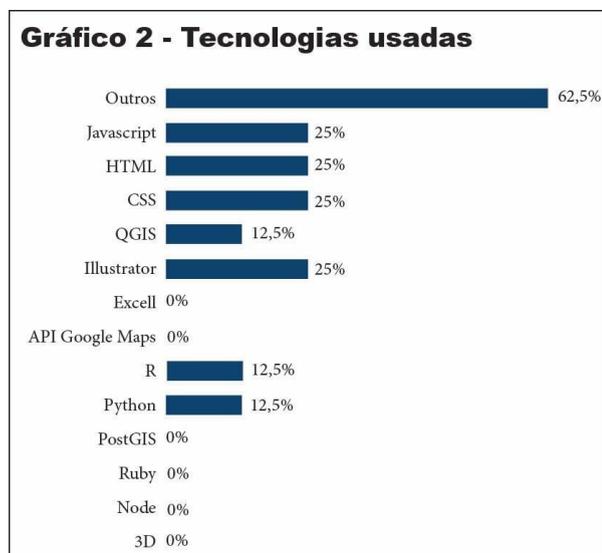
Fonte: Dados da Pesquisa, 2018.

Sobre as tecnologias utilizadas (gráfico 2), esse item é mais complexo de analisar tendo em vista que as tecnologias e linguagens dependem do tipo de projeto e da complexidade do mesmo e, ao mesmo tempo, do tamanho e da especialidade da equipe envolvida em tais projetos. O predomínio na nossa observação foi de tecnologias fora do escopo do nosso levantamento prévio e enquadrando-se em "outros". Dos 8 casos, 5 utilizaram tecnologias distintas e novas (65,5%). Há uma variedade de tecnologias e linguagens observada na análise, a exemplo da linguagem *R* e *Python*

<sup>6</sup> Disponível em: <https://goo.gl/zUufx2> Acesso em: 18 set 2018.

<sup>7</sup> Neste levantamento, os resultados somam mais de 100% porque era possível selecionar mais de um elemento por caso

com 12,5% do total. Outros como JavaScript, CSS, por exemplo, totalizaram 25%.



Fonte: Dados da Pesquisa, 2018.

Especificamos apenas as linguagens e tecnologias mencionadas no resumo executivo (apresentação) dos projetos vencedores sem exploração interna dos projetos para tentar identificar pelo código fonte ou outro meio tendo em vista que isso exigiria um trabalho de "bastidores" dos projetos de difícil caracterização visando a precisão. Logo, partimos para o próprio resumo em que cada projeto indicava essas tecnologias utilizadas, de modo a termos um recorte mais coerente. Essa exploração nos fornece um panorama e aponta para o uso crescente de tecnologias específicas de visualização de dados como o *Tableau Public*<sup>8</sup>. Na nossa observação, exploramos o conceito de inovação, partimos da observação de cada caso e dos próprios apontamentos no resumo executivo dos projetos para poder apontar os avanços em termos de inovação nessas narrativas baseadas em visualização guiada por dados. Nesta direção, identificou-se que todas, em algum aspecto, inclusive pela natureza do

<sup>8</sup> Software livre que permite criar visualizações de dados na Web. Disponível em: <https://public.tableau.com/pt-br/s/> Acesso em: 1 out 2018.

prêmio, apresentam alguns elementos de inovação e de avanço quanto ao estado das produções do momento, apresentando inovações disruptivas para os formatos de visualização de dados. De um modo geral, podemos apontar alguns aspectos inovadores como: 1) Dados Geolocalizados por satélite e atualizáveis em tempo real; 2) Mapas guiados por dados como bússola para as informações; 3) Imagens em 360 grau conferindo imersão na narrativa e 4) *Machine Learning* e Inteligência Artificial como potencializadores dessas narrativas.

## 5 Considerações Finais

As narrativas em ambientes digitais criaram novos paradigmas em relação às narrativas tradicionais, tendo em vista o potencial da ambiência digital para a construção das mesmas. Nesta direção, esse estudo tentou compreender os processos de construção das narrativas com dados nestas visualizações e demarcar as estratégias utilizadas como forma de compreensão dos dados. Observamos alguns aspectos que servem para reflexão: 1) Formato *Longform* segue como uma prática que vem sendo cada vez mais explorado nas narrativas com dados; 2) Mapas guiados por dados como estrutura principal de apresentação dos dados, o que confere um aspecto de inovação dos dados; 3) Narrativas com dados em profundidade, o que quebra com o paradigma de somente mostrar, mas contextualizar, dar sentido àqueles dados visualmente.

Os textos curtos e blocados ao lado das imagens foram bem recorrentes, assim como o uso dos *hiperlinks*. 4) Dados abertos cada vez mais sendo explorados nestas produções, o que pode transparecer uma certa ética ao creditar nas narrativas, muito embora não se sabe se todos os dados brutos foram trabalhos massivamente. Vale destacar a imagem em satélite como

elemento diferencial produzindo um ambiente de imersão naquela estrutura de dados, ainda que só a *Reuters* tenha utilizado o recurso timidamente. As narrativas seguem as estratégias de tendências de mercado que estão postas, como formatos *Longform* e imagens de satélite de modo imersivo, para que se sinta dentro da narrativa, muito embora apenas a *Agência Reuters* utilizou, mas segue sendo um indício promissor desse recurso. A opção pela interação por exploração, cuja navegação se deu mais pela força do *link*, condiz com a proposta da narrativa, de contextualizar os dados, contar a história dos acontecimentos, o que pode parecer pouco dinâmico à primeira vista. Em nossa análise, alguns veículos analisados poderiam explorar mais imagens em 360 graus, por exemplo, como ferramenta imersiva das narrativas, o que invaria no manejo com os dados.

É válido ressaltar que não há modelos fechados para realizar análises e estudos, mas as existências de vários ângulos de abordagens em diversas áreas, tendo em vista que “a narrativa é capaz de corresponder a uma cronologia cronológica e não cronológica, extraindo uma sucessão de eventos” (BOASE, 2013, p. 5)<sup>9</sup>. Assim sendo, a pesquisa avança no sentido de demarcação de como esses elementos narrativos estão sendo explorados e seus potenciais na era do *Big Data*, de modo em o processo analítico-visual seja um dos expoentes, e ainda que resguardando as limitações do estudo. Para trabalhos futuros, busca-se realizar um estudo de recepção a fim de averiguar como os usuários recebem e compreendem esses dados a partir dessas visualizações, bem como a exploração mais

detalhada das tecnologias utilizadas pelos profissionais.

## 6 Referências

BOASE, C. Digital **Storytelling for Reflection and Engagement**: a study of the uses and potential of digital storytelling. Centre for Active Learning & Department of Education, University of Gloucestershire. (2013) Disponível em: <https://goo.gl/MNJg9H> Acesso em 3 set 2018.

CAIRO, A. **El arte funcional: infografía y visualización de la información**. Madrid: Alamut, 2011. (2008) Infografía 2.0: visualización interactiva de información en prensa. Alamut, Madrid.

FIGUEIRAS, A. A typology for data visualization on the web. In Information Visualisation (IV), 2013 **17th International Conference**, pages 351–358, July 2013. Disponível em: <https://goo.gl/TNhtEm> Acesso em: 18 mai 2018.

KOSARA, R; MACKLINLAY, J. Storytelling: The next step for visualization. **Computer**, 46(5):44–50, 2013.

LAGOZE, C. Big Data, data integrity, and the fracturing of the zone control. **Big Data & Society**. jul./dec. 2014. Disponível em: <https://goo.gl/TcsW1r> Acesso em: 10 abr. 2018.

LONGHI, R.; WINQUES, K. O lugar do longform no jornalismo online: qualidade versus quantidade e algumas considerações sobre o consumo. Brasília. 24º Compós2015, - **Anais..** Brasília: Universidade de Brasília, 2015. v. 1. p. 1-19.

PAUL, N. Elementos das narrativas digitais. In P. Ferrari (Org). **Hipertexto e hipermídia**: as novas ferramentas da Comunicação digital. São Paulo: Contexto, 2012.

SEGEL, E; HEER, J. Narrative visualization: Telling stories with data. **Visualization and Computer Graphics**, IEEE Transactions on, 16(6):1139–1148, 2010. Disponível em: <https://goo.gl/TmNgFu> Acesso em: 3 set 2018.

TUFTE, E. **The visual display of quantitative information**. Cheshire, Connecticut: Graphic Press, 2001.

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: 4 jul 2017.

<sup>9</sup> No original: “Narrative is able to combine both a chronological and non-chronological dimension into a meaningful whole by extracting a configuration from a followable succession of events”.

## 7 Apêndice A – Elementos gráficos nas visualizações

TIPO	Agência Reuters	La Nación	PostMedia	Infotimes	Caixin Media	RunRun	BuzzFeed	G1	TOTAL
Gráficos interativos	X		X	X	X				4 (50%)
Gráficos guiados por dados				X		X			2 (25%)
Gráficos estáticos	X				X	X			3 (37,5)
Mapas interativos	X			X	X	X			4 (50%)
Mapas estáticos	X	X		X	X	X	X	X	7 (87,5%)
Foto interativa	X								1 (12,5%)
Foto estática	X			X		X	X		4 (50%)
Imagem em 360	X								1 (12,5%)
Ilustrações	X		X	X	X		X		5 (62,5%)
Video		X		X	X	X	X	X	6 (75%)
Imagem de satélite	X								1 (12,5%)
Infográficos estáticos								X	1 (12,5%)
Infográficos interativos									0 (0%)
Outros			X						1 (12,5%)

Fonte: Dados da Pesquisa, 2018.

## Apêndice B – Interação

TIPO	Agência Reuters	La Nación	PostMedia	Infotimes	Caixin Media	RunRun	BuzzFeed	G1	TOTAL
Instrução (linear e aparenta botões de avançar e retroceder)									0 (0%)
Manipulação (quando é possível mudar os objetos dentro da visualização)			X	X					2 (25%)
Exploração (tipo de interação com navegação hipertextual, explorando dados e/ou por imersão)	X	X	X	X	X	X	X	X	8 (100%)

Fonte: Dados da Pesquisa, 2018.

## Apêndice C – Dados estruturados

TIPO	Agência Reuters	La Nación	PostMedia	Infotimes	Caixin Media	RunRun	BuzzFeed	G1	TOTAL
Verticalizados	X	X	X	X	X		X	X	7 (87,5%)
Horizontalizados						X	X		2 (25%)
Mapas guiados por dados	X	X		X	X	X		X	6 (75%)
Outro									0 (0%)

Fonte: Dados da Pesquisa, 2018.