

# What Do Social Media Data Analysts Want? An Analysis from the Perspective of Data Visualization

Caroline Q. Santos

Departamento de Computação – UFVJM  
Diamantina, MG – Brazil  
caroline.queiroz@ufvjm.edu.br

Milene S. Silveira

Escola Politécnica – PUCRS  
Porto Alegre, RS – Brazil  
milene.silveira@pucrs.br

## ABSTRACT

Currently, we observe a growing interest in obtaining information about what circulates on social media. There is interest in extracting and analyzing data obtained from the Web for a number of reasons as users opinions, users behavior, and networks evolution, for instance. Our study seeks to comprehend how data visualization could support social media data analysts in data understanding and decision making. We conducted interviews, and we analyzed the collected data using qualitative research strategies. Our results reveal a context in which analysts believe data visualization can support data analysis and in which they are aware of related tools. Our findings highlight social media analysts' needs in relation to data visualization in order to provide a basis to design of new tools to this audience.

## CCS CONCEPTS

• **Human-centered computing; Information visualization; Visual analytics; Empirical studies in visualization;**

## KEYWORDS

Data analysis; Data visualization; Social media

### ACM Reference Format:

Caroline Q. Santos and Milene S. Silveira. 2018. What Do Social Media Data Analysts Want? An Analysis from the Perspective of Data Visualization. In *17th Brazilian Symposium on Human Factors in Computing Systems (IHC 2018)*, October 22–26, 2018, Belém, Brazil. ACM, New York, NY, USA, 4 pages. <https://doi.org/10.1145/3274192.3274246>

## 1 INTRODUÇÃO

Muitas pessoas, de diferentes áreas do conhecimento - como sociologia, história, comunicação social e administração de empresas, estão extraindo e analisando dados das mídias sociais com diversos objetivos, como observar o comportamento ou entender a opinião da população *online*, por exemplo. O processo de análise consiste em definir quais dados se deseja coletar, como será feita a coleta, como esses dados serão processados e, também, como serão apresentados. Esses passos, ou etapas, são a base para a obtenção de dados, informações e conhecimentos que auxiliam a análise e a tomada de decisão.

Permission to make digital or hard copies of all or part of this work for personal or classroom use is granted without fee provided that copies are not made or distributed for profit or commercial advantage and that copies bear this notice and the full citation on the first page. Copyrights for components of this work owned by others than ACM must be honored. Abstracting with credit is permitted. To copy otherwise, or republish, to post on servers or to redistribute to lists, requires prior specific permission and/or a fee. Request permissions from [permissions@acm.org](mailto:permissions@acm.org).

*IHC 2018, October 22–26, 2018, Belém, Brazil*  
© 2018 Association for Computing Machinery.  
ACM ISBN 978-1-4503-6601-4/18/10...\$15.00  
<https://doi.org/10.1145/3274192.3274246>

O objetivo da visualização é auxiliar o processo de compreensão dos dados, podendo envolver os mais diferentes públicos na sua exploração e análise [3]. Basicamente, é a representação de um grande volume de dados em uma linguagem de mais fácil compreensão, como gráficos e conceitos visuais que indicam o objetivo da visualização [6]. A visualização tem um impacto positivo no processo de tomada de decisão e também auxilia na análise e na comunicação dos dados [6]. Por isso, ferramentas de visualização têm aumentado as possibilidades de analistas de dados explorarem mais facilmente grandes conjuntos de dados, alcançarem novos *insights* e uma melhor compreensão desses dados [2].

Neste estudo, por meio de entrevistas com analistas de dados, buscamos conhecer o trabalho envolvido na análise de dados de mídias sociais e entender melhor este domínio, compreendendo as atividades e necessidades dos analistas. Em um trabalho preliminar sobre esses dados [5], obtivemos “conhecimento para embasar novas questões de pesquisa para o desenvolvimento de visualizações que apoiem a análise de dados de mídias sociais”. Agora, a partir da análise dos dados coletados nas entrevistas, fizemos uma classificação dos perfis dos participantes em *personas*, bem como a associação de cada perfil às necessidades identificadas.

Este trabalho está organizado da seguinte forma: a próxima seção apresenta a metodologia da pesquisa, seguida das seções de resultados e discussão. Para finalizar, a última seção trata das considerações finais e dos próximos passos da pesquisa.

## 2 METODOLOGIA

Para guiar nosso estudo, formulamos questões de pesquisa com foco nos aspectos sobre organização dos dados para a análise e para a apresentação. A principal questão de pesquisa se refere a compreender “*como os analistas lidam com a visualização no processo de análise de dados de mídias sociais?*”. Além disso, pretendemos responder as seguintes questões: “*Como os analistas de dados de mídias sociais compreendem a importância das visualizações de dados?*” (RQ1) e “*Quais as necessidades dos analistas de dados de mídias sociais em relação à visualização de dados?*” (RQ2).

Neste estudo, apresentamos a evolução de resultados preliminares já discutidos anteriormente [5], fruto de entrevistas semi-estruturadas realizadas com analistas de dados para tentar obter suas percepções sobre visualização de dados e o quanto ela pode contribuir (ou não) no processo de análise. Conseguimos, agora, nos aproximar do contexto de necessidades e práticas específicas da análise de dados, o que nos permitiu elencar novos objetivos, dentre os quais podemos citar: *compreender o quanto a visualização está inserida, de fato, no contexto da análise e obter requisitos para a criação de um modelo de interface de visualização de dados customizável pelo analista.*

**Tabela 1: Perfil dos participantes do estudo.**

Nº	Idade	Formação	Usa ferramentas de visualização?
P1	58	História	Não
P2	42	História	Não
P3	28	Jornalismo	Não
P4	37	Jornalismo	Sim
P5	32	Publicidade	Sim
P6	25	Análise de Sistemas	Sim
P7	28	Administração	Sim
P8	29	Jornalismo	Sim

Além de reforçarmos esses resultados, identificamos e apresentamos as necessidades dos analistas de dados em relação a visualizações. Foram entrevistados oito participantes, que aceitaram participar da entrevista e autorizaram o uso dos dados para a pesquisa (por meio de um termo de consentimento), cujos perfis são descritos na Tabela 1.

As entrevistas foram realizadas com uma amostra de conveniência, por meio de contatos pessoais com profissionais que trabalham com análise de dados de mídias sociais em diferentes Estados (MG, RJ e RS). Elas variaram de 50 a 60 minutos cada, foram gravadas em áudio e depois transcritas e analisadas usando o método de análise de conteúdo [4].

O roteiro da entrevista semi-estruturada consistiu em (mas não estava limitado a) perguntas relacionadas ao conhecimento do participante sobre visualização de dados, conhecimento e uso de ferramentas de visualização, situações em que precisou apresentar dados e que as ferramentas que conhece não atenderam, e o que gostaria que uma ferramenta oferecesse para atender as suas necessidades em relação à apresentação dos dados. Utilizamos o RQDA<sup>1</sup> para categorizar os dados e extrair as informações que serão discutidas.

### 3 RESULTADOS

Em relação ao entendimento de visualização de dados (RQ1), perguntamos aos participantes sobre suas habilidades com ferramentas de visualização e suas opiniões sobre o uso de visualização na análise de dados. Observamos que eles conhecem o conceito e acreditam que o uso de visualizações pode melhorar o processo de análise:

*“Hoje, para além de todo esse esforço, visualizar aumenta mais ainda a possibilidade da gente de discussões e de análises, muito mais que antes. Traz muito mais dados e informações. A informação é o começo pra gente. Hoje algumas visualizações já avançam da informação e chegam na ideia da comunicação. Por isso é fundamental hoje em dia.” (P1).*

*“Eu considero a visualização essencial na análise de dados. Na verdade, na análise de dados em si, não, porque aí o importante é a inteligência analítica que tu tá colocando em cima disso. (...) A visualização é importante no processo de engajamento e compreensão dos dados. Ela aumenta o entendimento daquilo.” (P7).*

*“A visualização de dados é um pouco a interface entre o trabalho que o jornalista faz e a informação que chega*

*pro usuário, pro leitor, internauta. (...) O analista que não se beneficia das ferramentas de análise de dados está um pouco atrasado.” (P8).*

Apesar disto, alguns consideram as ferramentas disponíveis complexas e difíceis de aprender (*“Eu usei algumas ferramentas de visualização, mas nunca usei o Gephi, por exemplo, porque acho muito difícil. Eu não consegui instalar o Gephi, para você ter ideia.” (P3)*). De uma forma geral, os participantes reconhecem a importância e a necessidade de aprender a manipular dados e gerar visualizações sobre eles. Percebemos que, embora alguns não utilizem ferramentas computacionais (3 participantes), eles sabem quais dados querem coletar e analisar. Assim, as entrevistas permitiram que entendêssemos como as pessoas estão realizando análises de dados de mídias sociais e o que elas consideram importante nesse contexto.

Seis participantes se consideram analistas de dados, P5 não se considera e P8 se considera jornalista de dados, pois, segundo ele, *“eu tenho muito trabalho de análise e não sei se é exatamente o mesmo tipo de análise que um analista de dados faz”*. Os que se consideram analistas de dados não se sentiram completamente seguros para reconhecer esse papel, visto que eles acreditam ser uma denominação que envolve muitas habilidades, como pode ser resumido na fala de P7: *“para ser um analista de dados é preciso ter raciocínio lógico, estatístico, matemático, inteligência analítica, que é o principal, aí as outras habilidades vão depender muito do âmbito em que você trabalha”*. Por outro lado, mesmo não possuindo todas as habilidades que imaginam ser necessárias ao analista, eles se intitularam analistas de dados, pela capacidade de analisar grandes bases de dados.

O participante P4 acredita que a análise feita com base em alguma teoria já consolidada, como a Análise de Redes Sociais, fortalece o papel de analista de dados. Segundo P4, o que ele faz *“passa pela análise de dados, embora isso seja feito com um olhar bastante peculiar, que é um olhar interdisciplinar, um olhar pouco filiado a perspectivas conceituais mais estabelecidas, um olhar para discussões que não exatamente estão dentro de uma perspectiva consolidada dessas análises de dados.”*

Os resultados sugerem que os analistas conhecem ou já ouviram falar de algumas ferramentas de visualização, mas ainda não as dominam. No entanto, eles acreditam que a colaboração é uma maneira de resolver esse problema. Isso pode ser ilustrado pela fala de P1: *“eu não sei como usar essas ferramentas, mas eu as conheço. Assisti a uma oficina sobre o Gephi e achei muito difícil. Eu uso visualizações no meu trabalho, feitas por meus alunos e por meus parceiros, de acordo com minhas demandas. (...) A visualização é importante porque permite compreender a dinâmica, ver que os dados estão em formas diferentes, ao mesmo tempo, em vários lugares.”*

Em geral, os participantes sabem que precisam entender essa realidade relacionada ao volume de dados que é compartilhado nas mídias sociais e o quanto de informação pode ser extraída dele. Segundo P3, *“as pessoas precisam desenvolver a cultura “dateira”, precisam entender que os dados trazem muita informação sobre as pessoas e o mundo.”* Com isso, um ponto importante identificado foi a necessidade de se incentivar e promover a literacia digital das pessoas que querem analisar dados, principalmente nos dias atuais em que há muitas ferramentas para auxiliar os analistas. Alguns

<sup>1</sup>RQDA é um pacote do aplicativo R para análise de dados qualitativos: <http://rqda.r-forge.r-project.org/>

participantes mencionaram conhecer o propósito de uma ou outra ferramenta, mas não saberem utilizá-la, outros já utilizaram várias delas, o que deixa claro que eles sabem da existência de ferramentas para apoio às suas análises.

Sobre a **RQ2**, que busca levantar quais as necessidades dos analistas de dados de mídias sociais em relação à visualização de dados, conseguimos identificar algumas necessidades de forma bem direta. O participante P4, por exemplo, relatou que precisa combinar o uso de várias ferramentas para tentar obter o resultado desejado para suas análises, porém ainda não conseguiu isso. Ele utiliza bastante o Gephi e tem feito alguns trabalhos no Tableau com o propósito de criar visualizações que se aproximem de algo que mostre as redes se formando e se transformando ao longo do tempo. Esse foi outro ponto importante que surgiu nas entrevistas: a necessidade de observar a mudança de contexto, relacionamentos e foco ao longo do tempo, como ilustrado pelos discursos:

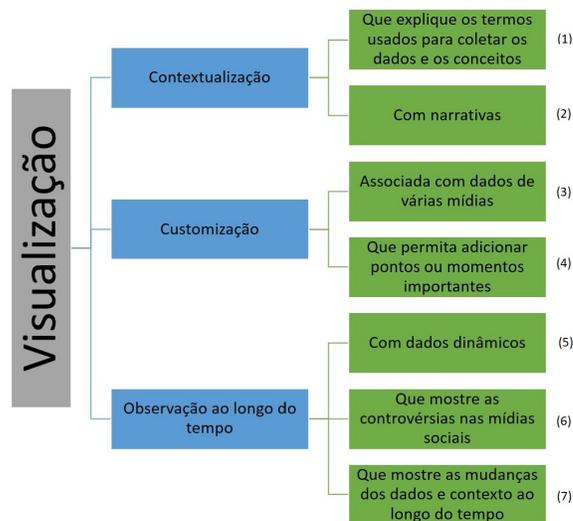
*“Eu adoraria poder ver as controvérsias nas mídias sociais, especialmente ao longo do tempo. Se eu pudesse ver, retroagir em intensidade como as questões foram aparecendo ao longo do tempo e onde. Por exemplo: eu preciso distinguir aborto, enquanto aborto opção da mulher, de aborto política pública. Quando é o primeiro tem uma carga de moralismo muito grande, e vai ser respondido ali”* (P1).

*“No meu negócio, para eu melhorá-lo, eu preciso analisar historicamente os picos, por exemplo, e outras informações complementares (dos sites, por exemplo)”* (P5).

*“Eu acho que um recurso de versão resolveria a questão do tempo, que aí você teria a versão 1, foi a que você fez a coleta do dia x ao dia y, e naquele dia ali o movimento estava só no início; aí você vê aquela versão e depois você tem a versão 2, que você continua a coleta até um mês depois, por exemplo, e aí você tem a versão 3, que aquilo já está há 6 meses acontecendo, e aí você vê a mudança, então, esse controle de versão facilitaria”* (P8).

Alguns participantes apontaram a importância das visualizações e da infografia na análise de dados em uma tentativa de relacionar isso ao que uma visualização de dados precisa conseguir informar aos seus leitores (*“Vi uma visualização sobre a crise da Volkswagen. Foi um resumo, mas ajudou as pessoas a entenderem o que aconteceu e não simplesmente se assustarem com a manchete das notícias: ‘O presidente da Volkswagen renuncia por causa de (...)’ Então, a visualização de dados nos permite entender a complexidade de certas coisas”*. (P1)). Outros já mencionaram diretamente sobre associar narrativas com visualizações (*“Agregar narrativas nas visualizações tem uma grande importância porque as pessoas querem receber a informação mastigada mas, se eu só jogar gráficos com números, não vai ter graça nenhuma. Então o contexto todo cria o ambiente e o entendimento da pessoa pra ela entender aquilo. É fundamental.”* (P3)).

Os participantes também estão preocupados com a customização. Alguns relataram que muitas vezes precisavam usar *plug-ins*, baixar bibliotecas, importar pacotes ou escrever *scripts* para organizar



**Figura 1: Necessidades dos analistas para ferramentas de visualização.**

as visualizações como eles queriam, e se isso fosse oferecido na ferramenta, seria muito melhor.

Enfim, percebemos que há um grande leque de necessidades dos analistas em relação às ferramentas de visualização e isso pode significar um espectro que vai desde o desconhecimento sobre o uso da ferramenta ou de suas funcionalidades até o seu domínio e conhecimento do que a ferramenta atende (ou não) em termos de apresentação dos dados.

## 4 DISCUSSÃO

A partir dos dados analisados nas entrevistas, conseguimos construir a nossa compreensão de o quanto a visualização de dados está presente nas atividades de trabalho dos participantes e conhecer a opinião deles sobre a importância do uso de visualizações para apresentação de seus dados. Para os jornalistas, percebemos que a combinação de técnicas de visualização podem ser úteis para a obtenção e a divulgação de informações, podendo ser um recurso valioso em suas atividades de trabalho.

Para os analistas de dados, percebemos que a visualização é um conceito não apenas conhecido mas também necessário no processo de análise de dados de mídias sociais. No entanto, para eles, há ainda dificuldades relatadas em aprender a utilizar as ferramentas de visualização e há também demandas que eles acreditam que não sejam atendidas pelas ferramentas existentes.

Com isso, conseguimos: *i*) identificar as necessidades/demandas dos analistas para ferramentas de visualização; e *ii*) criar *personas* para representar os diferentes perfis de participantes do estudo. Esses pontos foram fundamentais para a compreensão sobre o que querem os analistas de dados. Em relação às demandas/requisitos para ferramentas de visualização, nossas descobertas sugerem três pontos a serem considerados: (1) contextualização, (2) customização e (3) observação ao longo do tempo, conforme mostrado na Figura 1.

Tabela 2: Personas

Foto	Descrição
	<p>Nome: João Chaves, Idade: 50</p> <p>João tem formação nas áreas de humanas e de ciências sociais. Trabalha em universidade e suas pesquisas envolvem analisar fatos, acontecimentos, situações, contextos históricos, sociais e econômicos a partir de conteúdos de mídias sociais online, assim como observar e analisar as mudanças (e o que as impactam) ao longo do tempo nesses aspectos. João não utiliza software para coletar os dados e apoiar a análise, mas ele procura parceiros para fazer isso. Ele gostaria de poder apresentar suas demandas ao parceiro que irá criar/executar as ferramentas de software necessárias para que este lhe forneça os resultados esperados.</p>

Tabela 3: Personas

Foto	Descrição
	<p>Nome: Maria Santos, Idade: 28</p> <p>Maria é da área de comunicação social e suas atividades de trabalho envolvem promover o marketing digital da empresa, analisar as métricas das redes e mídias sociais do perfil da empresa, criar relatórios a partir dessas análises, observar o comportamento dos usuários online, e planejar estratégias de melhoria das métricas. Maria não utiliza software para coletar os dados e, para apoiar a análise, ela utiliza os recursos que os próprios sites oferecem, além do Google Analytics. Ela gostaria de ter tempo para aprender a usar ferramentas de visualização, como o Tableau e o Gephi, porque acredita que poderiam apoiar melhor a criação de gráficos sobre os dados que darão suporte às decisões a serem tomadas pela empresa.</p>

Tendo em vista estas necessidades e as informações obtidas a partir do perfil dos participantes, criamos 3 *personas* [1] para representar grupos de usuários segundo seu padrão de comportamento nas atividades de análise, uso de tecnologias, preferências, etc. A *persona* 01 (Tabela 2) foi pensada para representar os analistas que têm pouca ou nenhuma habilidade com ferramentas para coleta e apresentação de dados. As suas demandas são referentes aos itens 1, 3, 5, 6 e 7 da Figura 1.

A *persona* 02 (Tabela 3) foi pensada para representar os analistas com alguma habilidade com ferramentas para coleta e apresentação de grande volume de dados: aqueles que conhecem algumas ferramentas, mas não aprenderam a utilizá-las ou tiveram dificuldades. As suas demandas são referentes aos itens 2, 3 e 4 da Figura 1.

A *persona* 03 (Tabela 4) foi pensada para representar os analistas com habilidades em ferramentas para coleta e apresentação de dados: aqueles que conhecem algumas ferramentas, utilizam-nas e buscam o aprendizado de outras. As suas demandas são referentes aos itens 2, 3, 4, 5 e 7 da Figura 1.

## 5 CONSIDERAÇÕES FINAIS

Este estudo investigou o conhecimento e utilização de ferramentas de visualização de dados por analistas de dados de mídias sociais e quais são as suas necessidades em relação a isso. Com as entrevistas, nos aproximamos do universo da análise de dados de mídias sociais e compreendemos como os analistas realizam esse processo.

Percebemos que há dificuldades em representar graficamente os dados da forma que desejam, mas, na prática, eles fazem uso de várias ferramentas para tentar obter o resultado desejado. Eles sentem falta de ferramentas de fácil uso que os permitam construir

Tabela 4: Personas

Foto	Descrição
	<p>Nome: José Pereira, Idade: 31</p> <p>José tem formação na área de comunicação social/ jornalismo. Trabalha com pesquisas que envolvem investigar e analisar fatos, acontecimentos, situações, contextos históricos, sociais e econômicos a partir de conteúdos de mídias e redes sociais online, assim como observar e analisar as mudanças (e o que as impactam) ao longo do tempo nesses aspectos. José utiliza software para coletar e fazer limpeza dos dados e para apoiar a análise, e também conhece e utiliza ferramentas de visualização de dados. Ele gostaria que essas ferramentas tivessem recursos para apresentar a mudança de contexto ao longo do tempo mas, que não fosse uma espécie de “fotografia” de cada momento.</p>

as visualizações sem a necessidade de se utilizar várias ferramentas para isso. Percebemos que existem diferentes perfis de analistas de dados de mídias sociais, os quais foram categorizados por meio de *personas*.

Em relação à nossa principal questão de pesquisa, nossos resultados sugeriram que a relação entre analistas de dados e visualização está em evolução. Eles conhecem o conceito e acreditam na importância do seu uso, mas alguns ainda consideram as ferramentas complexas e difíceis de aprender; quando eles precisam, pedem um parceiro para criar as visualizações.

Para apoiar a reflexão dos *designers* sobre como ajudar a resolver os problemas dos analistas, identificamos requisitos para uma possível ferramenta de visualização de dados. Como trabalhos futuros, pretendemos realizar uma comparação entre nossos resultados e os existentes na literatura, projetar uma ferramenta que contemple as necessidades apontadas pelos analistas e criar um protótipo a fim de ser possível realizar uma prova de conceito e verificar se essas necessidades foram atendidas.

## REFERÊNCIAS

- [1] Kathy Baxter, Catherine Courage, and Kelly Caine. 2015. *Understanding your users: a practical guide to user research methods*. Morgan Kaufmann.
- [2] Michael Brooks. 2015. *Human Centered Tools for Analyzing Online Social Data*. Ph.D. Dissertation. University of Washington.
- [3] Jeffrey Heer, Michael Bostock, and Vadim Ogievetsky. 2010. A Tour Through the Visualization Zoo. *Queue* 8, 5, Article 20 (May 2010), 20:20–20:30 pages.
- [4] Jonathan Lazar, Jijuan Heidi Feng, and Harry Hochheiser. 2010. *Research methods in human-computer interaction*. John Wiley & Sons.
- [5] Caroline Queiroz Santos, Milene Selbach Silveira, and Isabel Harb Manssour. 2016. Visualization and Social Media Data Analysis: Preliminary Studies About Data Analysts' Perception. In *Proceedings of the 15th Brazilian Symposium on Human Factors in Computer Systems (IHC '16)*. ACM, New York, NY, USA, Article 35, 4 pages.
- [6] Matthew O Ward, Georges Grinstein, and Daniel Keim. 2015. *Interactive data visualization: foundations, techniques, and applications*. CRC Press.