

Uma análise da compreensibilidade das visualizações de dados durante a pandemia de COVID-19 no Brasil

Rodrigo Oliveira¹, Cláudia Cappelli², Jonice Oliveira¹

¹Programa de Pós-Graduação em Informática (PPGI) da Universidade Federal do Rio de Janeiro (UFRJ), Rio de Janeiro, RJ, Brasil

²Departamento de Ciência da Computação e Informática da Universidade do Estado do Rio de Janeiro (UERJ), Rio de Janeiro, RJ, Brasil

rodrigo.oliveira@ppgi.ufrj.br, claudia.cappelli@gmail.com,
jonice@dcc.ufrj.br

***Abstract.** We present a pictorial analysis of data visualizations released by Brazilian state governments during the COVID-19 pandemic. During a preliminary evaluation, we identified graphs with the potential to be misinterpreted or misunderstood because they did not follow basic guidelines in the area of data visualization. In this work, we analyze the most recurrent failures, the main types of problematic graphics identified and we call attention to the need for greater care with the use of visualizations.*

***Resumo.** Apresentamos uma análise pictórica de visualizações de dados divulgadas pelos governos estaduais brasileiros durante a pandemia de COVID-19. Durante uma avaliação preliminar identificamos gráficos com potencial de serem mal interpretados ou pouco compreendidos por não seguirem diretrizes básicas da área de visualização de dados. Neste trabalho analisamos as falhas mais recorrentes, os principais tipos de gráficos problemáticos identificados e alertamos à necessidade de um maior cuidado com o uso de visualizações, assim contribuindo que gráficos ineficientes sejam evitados futuramente em outros contextos.*

1. Introdução

O uso de visualizações de dados como um recurso para apoiar o entendimento sobre informações para uma determinada população vem se consolidando. Esse avanço é percebido pela proliferação de representações gráficas no cotidiano através de áreas mais recentes como o jornalismo de dados (Bhaskaran et al., 2022) ou técnicas de visual law (Bolesina et al., 2022). Representações visuais como essas, buscam ajudar as pessoas a realizar tarefas com mais eficiência com o auxílio de imagens geradas pelo computador a partir de um conjunto de dados projetados (Munzner, 2014). No contexto de saúde pública, mas especificamente durante a pandemia da COVID-19, painéis com visualizações surgiram para divulgar os dados da doença em todo mundo por iniciativas públicas ou privadas. Um exemplo é o painel de dados da Organização Mundial de

Saúde (OMS)¹, onde o Brasil está dentre os 10 países mais afetados pela doença e entre os 3 que concentram mais da metade dos casos (EUA, Índia, Brasil). Os estados brasileiros também se valeram de visualizações de dados para comunicar sobre a nova doença, fornecendo boletins recorrentes ou páginas web com informações aos cidadãos. Contudo, no estudo de Oliveira et al. (2021), os autores constataram que tais recursos foram usados em sua maioria de forma incorreta e sem consideração pelos dados, processos e princípios básicos da área de visualização, podendo facilmente sobrecarregar ou dificultar a compreensão dos leitores.

Desde o surgimento da epidemia, especialistas têm alertado para propagação de gráficos sobre a COVID-19 que são mal-projetados² ou manipulados intencional para melhorar a imagem de países atingidos³. É vital aos governos fornecer meios de comunicar a população sobre o avanço da doença em suas localidades como forma de prevenir e controlar a epidemia. O entendimento do público depende da produção de mensagens eficazes para que possam perceber e responder adequadamente à intensidade do desdobramento da pandemia (Malecki et al, 2020). Além disso, os baixos níveis de letramento na população brasileira reforçam essa necessidade de uma comunicação mais eficaz. 63% dos brasileiros não é proficiente ou apenas intermediário para entender informações simples. 3 em cada 10 pessoas têm muita dificuldade para leitura, escrita e reconhecer significados simples de representação gráfica, segundo o último Indicador de Analfabetismo Funcional (2018)⁴. É imprescindível traduzir a mensagem correta de saúde pública para compreensão e acesso de todos.

Apesar da área de visualização de dados estar bem consolidada com diversos aplicativos ou websites com a função de criar gráficos como o Microsoft Excel⁵, Tableau⁶, Infogram⁷ ou Microsoft Power Bi⁸. As boas práticas sobre gráficos não chegam ao conhecimento de todos e esses sistemas não seguem esses padrões na hora de criar seus gráficos (Su, 2008), (Wnuk; Dębski, 2016). A autora Cole Knaflic cita em seu livro *Storytelling com dados*: “Ninguém quer fazer um gráfico ruim. Mas acontece. Repetidamente. Em todas as empresas e por todo tipo de pessoas” (Knaflic, 2015). Segundo o Índice de Transparência da Covid-19 da Open Knowledge Brasil⁹, 90% dos estados avaliados no começo da pandemia tinham um nível “Baixo” de clareza das informações divulgadas.

Este artigo se propõe a apresentar uma análise pictórica das visualizações de monitoramento do começo da epidemia do novo coronavírus nos estados brasileiros durante o período de junho e julho de 2020 a partir de um conjunto de princípios de visualização relatados na pesquisa preliminar de Oliveira et al. (2021). Buscamos contribuir com uma avaliação detalhada dos principais gráficos em desacordo a essas diretrizes que se tornaram potenciais ferramentas de incompreensão e desinformação

¹ <https://covid19.who.int/>

² <https://www.esri.com/arcgis-blog/products/product/mapping/mapping-coronavirus-responsibly/>

³ <https://towardsdatascience.com/stopping-covid-19-with-misleading-graphs-6812a61a57c9>

⁴ <https://ipm.org.br/inaf>

⁵ <https://www.microsoft.com/pt-br/microsoft-365/excel>

⁶ <https://www.tableau.com/pt-br>

⁷ <https://infogram.com/>

⁸ <https://powerbi.microsoft.com/pt-br/>

⁹ <https://transparenciacovid19.ok.org.br/v1/>

para a população. Assim como, classificar as violações aos principais cuidados no design de visualizações de forma a evitar que outras experiências similares produzam o mesmo resultado, principalmente por não especialistas de órgãos governamentais envolvidos no monitoramento epidemiológico.

Este trabalho apresenta na próxima seção uma breve descrição da metodologia utilizada. A seção 3 reporta os resultados da avaliação e finalizamos com as conclusões e trabalhos futuros na seção 4.

2. As Visualizações de COVID-19 no Estados do Brasil

A metodologia utilizada para a definição das diretrizes assim como a seleção das visualizações deste estudo se seguem no trabalho de Oliveira et al. (2021). Os gráficos analisados foram encontrados a partir da busca nos sites e portais dos governos estaduais e federal, foram selecionados os endereços das páginas que continham painéis de dados com visualizações que acompanham a doença (23 estados). Na ausência deste, foram consultados os boletins epidemiológicos diários (4 estados), que também apresentavam gráficos da epidemia durante o período de 15 de junho a 06 de julho de 2020. Tabelas de dados, contagem simples de indicadores, gráficos sobre prestação de contas públicas ou isolamento social não foram considerados válidos. Cada estado teve o total de gráficos válidos contabilizados e submetidos a avaliação com base no conjunto de diretrizes definido. As diretrizes incluídas no conjunto que forma a régua de avaliação foram encontradas em uma revisão de trabalhos da área de visualização de dados e totalizaram cerca de 50 diretrizes, por exemplo: não usar gráficos em 3D, adicionar sempre um título ao gráfico, adicionar cores legíveis a pessoas com daltonismo. Há práticas mais contextuais, como: retirar informações irrelevantes do gráfico, ou destacar itens com cores culturalmente aceitáveis, por exemplo, vermelho para perigo ou verde para positivo. Essas recomendações se restringem aos tipos de gráficos mais usados como: barras, pizzas, linhas e mapas, não considerando a interatividade em sua análise. Não foi realizado um julgamento de valor sobre as diretrizes, assume-se que todas têm igual significância. A ausência de qualquer uma delas, se considera uma falha, afetando a compreensão do gráfico em algum aspecto ou deixando de torná-lo mais agradável ou mais fácil para a leitura possível. A Tabela 1 expõe um resumo das informações coletadas, assim como o link para cada portal de origem desses gráficos, os tipos de formato, os softwares onde foi possível identificar que as visualizações foram desenvolvidas e divulgadas, a quantidade de gráficos encontrada e o total de falhas nesses gráficos para cada estado.

Tabela 1. Visualizações de dados por estado brasileiro. Fonte: <oculto para revisão>.

Estado	Link	Formato	Qtd Gráficos	Total de Falhas	Estado	Link	Formato	Qtd Gráficos	Total de Falhas
Acre	[Link]	Painel	15	8	Paraíba	[Link]	Painel	13	46
	[Link]	Painel	8	8	Paraná	[Link]	Boletim	21	33
Alagoas	[Link]	Painel	17	10	Pernambuco	[Link]	Painel	24	33
Amapá	[Link]	Painel	7	2	Piauí	[Link]	Painel	5	20

Amazonas	[Link]	Painel	21	5	Rio de Janeiro	[Link]	Painel	19	89
Bahia	[Link]	Boletim	9	5	Rio Grande do Norte	[Link]	Boletim	25	44
Ceará	[Link]	Painel	9	3	Rio Grande do Sul	[Link]	Painel	9	56
Distrito Federal	[Link]	Painel	6	6	Rondônia	[Link]	Painel	20	65
Espírito Santo	[Link]	Painel	10	3	Roraima	[Link]	Painel	4	25
Goiás	[Link]	Painel	17	3	Santa Catarina	[Link]	Painel	23	91
Maranhão	[Link]	Painel	4	3	São Paulo	[Link]	Painel	19	32
Mato Grosso	[Link]	Boletim	4	1	Sergipe	[Link]	Painel	16	56
Mato Grosso do Sul	[Link]	Painel	5	3	Tocantins	[Link]	Painel	13	23
Minas Gerais	[Link]	Painel	3	1					
Pará	[Link]	Painel	21	6					

3. Resultados da Avaliação das Visualizações

Todos os estados apresentaram ao menos dois gráficos com problemas referente às diretrizes selecionadas neste trabalho. Um resumo deste resultado é apresentado na Figura 2 em forma gráfica. Na imagem demonstramos que o gráfico de barras foi o tipo de gráfico mais utilizado. Além disso, identificamos que o software mais utilizado para a criação dos gráficos foi o Power Bi (6 estados), seguido de vários estados onde não foi possível identificar o sistema utilizado (6 estados). A biblioteca de código aberto Chart Js¹⁰ seguiu em terceiro lugar com 4 estados utilizando suas funções, seguido do sistema Qlik sense¹¹ (2 estados) e ArcGIS¹² (1 estado) respectivamente.

A diretriz mais desrespeitada entre os gráficos analisados foi a ausência de uma legenda no gráfico (28 ocorrências), seja para todos os dados ou alguns deles, prejudicando o usuário de identificar sobre como o gráfico apresenta suas variáveis. Em seguida, a identificação de cada linha poderia ser melhorada com um rótulo direto, tornando a legenda desnecessária em um gráfico de linhas (18 ocorrências). Ainda tivemos a escala de cores com tons de uma mesma cor muito próximos tornando difícil a comparação entre eles com um total de 18 ocorrências. Muitas visualizações também apresentaram algum nível de desordem visual com elementos sobrepostos ou desnecessários (15 ocorrências) e eixos sem rótulos também atrapalhando a identificação dos dados utilizados.

¹⁰ <https://www.chartjs.org/>

¹¹ <https://www.qlik.com/pt-br/products/qlik-sense>

¹² <https://www.arcgis.com/index.html>

Destaques em resumo dos resultados da análise das visualizações de dados sobre a COVID-19 no Brasil

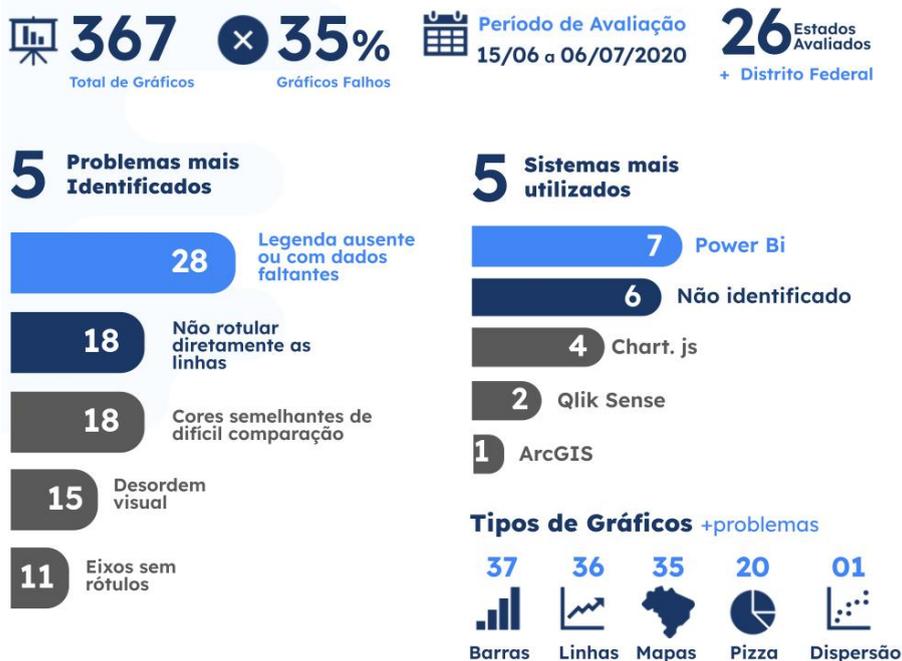


Figura 1. Gráficos de resumo dos resultados da avaliação. Fonte: Autor.

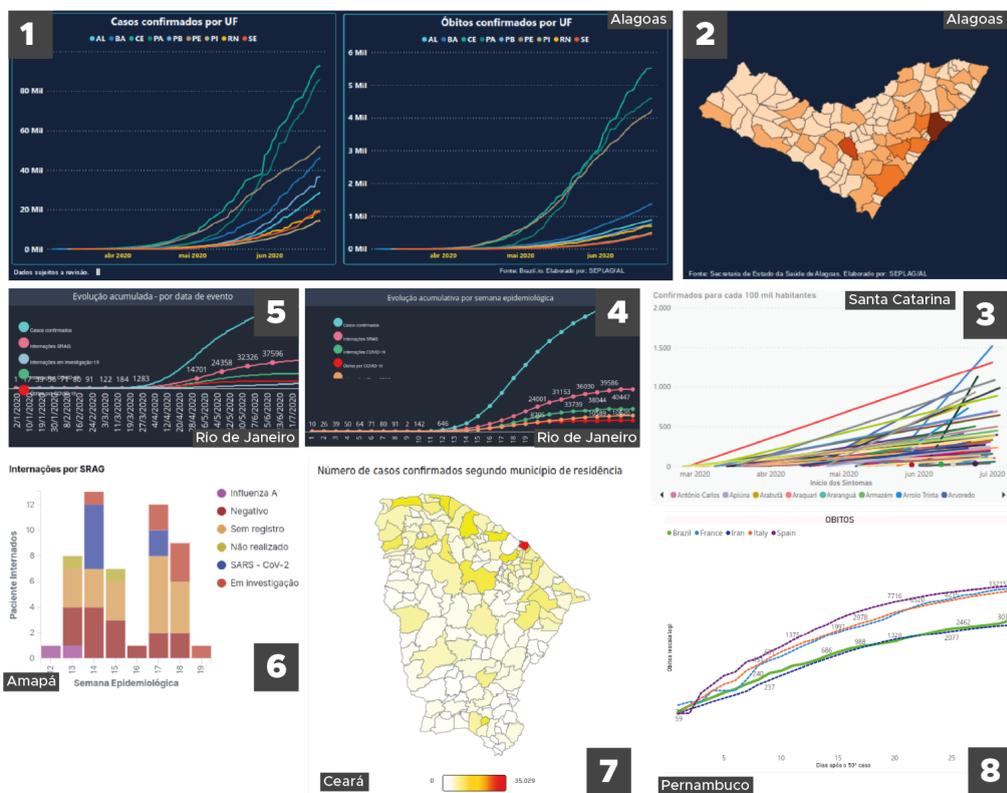


Figura 2. Gráficos avaliados com maior número de problemas. Fonte: 1) Gráficos de barras de Alagoas 2) Mapa de Alagoas 3) Gráfico de Linhas de Santa Catarina 4) Gráfico de Linhas do Rio de Janeiro 5) Gráfico de Linhas do Rio de Janeiro 6) Gráfico de Barras Empilhadas do Amapá 7) Mapa de Ceará 8) Gráfico de Linhas do Pernambuco.

A Figura 2 apresenta alguns exemplos dos gráficos coletados durante a avaliação. Foram selecionadas as visualizações que apresentaram mais problemas durante a pesquisa. Sendo os gráficos de linhas considerados os mais graves, por exemplo nos itens 1, 3, 4, 5 e 8 todos possuem várias cores similares que dificultam a identificação das informações. É recomendado no máximo 5 tons de cores sem prejudicar a leitura (Wilke, 2019). Além desses gráficos não apresentarem uma rotulagem direta nas linhas para evitar que os leitores precisem percorrer o olhar na legenda buscando identificar qual cor corresponde a cada linha fazendo uma sobrecarga mental ao usuário para leitura desse gráfico. Os itens 2 e 7 apresentam mapas também com problemas nas cores escolhidas, onde a escala é de difícil leitura tornando muitas cores semelhantes nas áreas pintadas, além da falta de legendas. Por fim, no gráfico 6 temos barras empilhadas com a ordem da legenda, não acompanhando a ordem das barras e também cores das categorias semelhantes, sem contar que nenhum dos gráficos se atenta ao uso de escala de cores que atendam a população daltônica.

4. Conclusões

Este trabalho se debruçou a detalhar os fatores mais recorrentes que afetam a compreensibilidade das visualizações públicas brasileiras selecionadas durante a pandemia de COVID-19. Como trabalho futuro é necessário um estudo com outros tópicos que podem afetar o entendimento de um gráfico, como a interatividade e usabilidade. Todavia, o mérito deste artigo é apontar caminhos que evitem que problemas como os elucidados nesta pesquisa não ocorram novamente em gráficos públicos de outros contextos. Além do alerta aos estados brasileiros da necessidade de mais esforços, para divulgar as informações com recursos mais eficientes.

Referências

- Bolesina, Iuri, and Jeverson Lima Lemes. "Visual law: um conceito emergente do encontro entre direito e design." *Revista Thesis Juris* 11.1 (2022): 155-171.
- Bhaskaran, Harikrishnan, Geeta Kashyap, and Harsh Mishra. "Teaching Data Journalism: A Systematic Review." *Journalism Practice* (2022): 1-22.
- Oliveira, R., Cappelli, C., & Oliveira, J. (2021, June). An indicator of inefficient visualizations: The challenge of transparency during the covid-19 pandemic in brazil. In XVII Brazilian Symposium on Information Systems (pp. 1-11).
- Malecki, K., Keating, J. A., & Safdar, N. Crisis Communication and Public Perception of COVID-19 Risk in the Era of Social Media. *Clinical Infectious Diseases*, 2020.
- Knaflic, Cole Nussbaumer. *Storytelling with data: A data visualization guide for business professionals*. John Wiley & Sons, 2015.
- Su, Y. S. (2008). It's easy to produce chartjunk using Microsoft® Excel 2007 but hard to make good graphs. *Computational Statistics & Data Analysis*, 52(10), 4594-4601.
- Wnuk, A., & Dębski, K. J. (2016). Should we trust graphs' default settings in the R packages?. *Communications in Biometry & Crop Science*, 11(2).
- Wilke, C. O. (2019). *Fundamentals of data visualization: a primer on making informative and compelling figures*. O'Reilly Media.