

Laboratório VISGRAF

Instituto de Matemática Pura e Aplicada

Visualizacao em Tempos de Corona Virus – Parte 2

Julia Giannella and Luiz Velho

Technical Report TR-20-10 Relatório Técnico

June - 2020 - Junho

The contents of this report are the sole responsibility of the authors.
O conteúdo do presente relatório é de única responsabilidade dos autores.

Visualização de dados em tempos de Coronavírus -Parte 2

Julia Giannella e Luiz Velho
Laboratório VISGRAF - IMPA

Resumo:

Atualmente, observamos uma proliferação de visualizações de dados sobre a Covid-19 na mídia o que torna o momento propício para estudo do tema sob a perspectiva de diferentes disciplinas dentre as quais a Matemática e o Design da Informação. Se, por um lado, a abundância de representações dessa natureza sobre a pandemia já constituiria razão legítima para se debruçar sobre o assunto, por outro lado, ela não é a principal motivação da presente discussão. A singularidade do fenômeno epidemiológico que vivemos evidencia novos aspectos a respeito da produção e do uso de visualizações de dados e um deles é sua diversificação para além da contagem e representação visual de eventos relacionados a propagação do vírus. Neste sentido, o artigo discute, através da análise de exemplos, três diferentes abordagens para este tipo de representação esquemática, a saber: visualização de dados hipotéticos, visualizações baseadas em dados secundários e visualização para crítica social e autorreflexão. Em última instância, podemos argumentar que os bons exemplos de visualização ajudam pessoas a entenderem, em diferentes graus e aspectos, as causas e implicações envolvidas no novo coronavírus e incentivam a responsabilidade cívica por meio do autocuidado e da prática do distanciamento social.

1. Introdução

Hoje, ouve-se com bastante frequência a expressão “sem precedentes”. É como se a linguagem que usamos para explicar nosso mundo estivesse desmoronando e os superlativos não pudessem acompanhar a nova realidade trazida pela pandemia do novo coronavírus. Desde março de 2020, nossas vidas foram atingidas por uma avalanche de perguntas difíceis de serem respondidas: “Quando isso vai acabar? Como posso proteger a mim mesmo e aos outros? Até quando as escolas permanecerão fechadas? Vou perder meu emprego?”.

A ironia da situação é que essas dúvidas emergem em um momento de larga e acelerada produção de dados que, no entanto, não parecem ser suficientes para aclarar a situação. A pandemia está gerando grandes quantidades de dados como número de testes realizados, casos confirmados, pacientes recuperados, pessoas que morreram pelo vírus, etc. Ao mesmo tempo, esses dados são muitas vezes incertos, incompletos e difíceis de serem analisados pois estão em contínua revisão e atualização. Diante da complexidade e volatilidade do fenômeno, cientistas, jornalistas e agentes públicos esforçam-se para tornar dados significativos a um público muitas vezes não especializado. Nesse contexto, a visualização de dados assume um papel crucial no processo comunicacional.

Podemos observar uma proliferação de visualizações de dados sobre a Covid-19 na mídia o que torna o momento bastante propício para estudo do tema sob a perspectiva de diferentes disciplinas dentre as quais a Matemática e o Design da Informação. Assim, nasceu o projeto e portal de pesquisa “Coronaviz: visualização em tempos de coronavírus“, uma iniciativa do Visgraf, do Instituto de Matemática Pura e Aplicada, Laboratório com foco em pesquisa e desenvolvimento de novas mídias. O projeto, em andamento, tem como objetivo analisar visualizações de dados produzidas no contexto da pandemia, assim como aclarar conceitos matemáticos importantes para a compreensão do fenômeno. Em última instância, podemos argumentar que os bons exemplos de visualização ajudam pessoas a entenderem, em diferentes graus e aspectos, as causas e implicações envolvidas no novo coronavírus e incentivam a responsabilidade cívica por meio do autocuidado e da prática do distanciamento social.

Se, por um lado, a proliferação de visualizações de dados sobre a atual pandemia já constituiria razão legítima para se debruçar sobre o assunto, por outro lado, ela não é a principal motivação da presente reflexão. A singularidade do fenômeno epidemiológico que vivemos evidencia novos aspectos a respeito da produção de visualizações e um deles é diversificação da visualização de dados para além da contagem de eventos relacionados a propagação da Covid-19, tema deste artigo.

Em um momento inicial, quando a pandemia deflagrou mundialmente em meados de março, mapas, gráficos de barras e gráficos de linhas de casos confirmados, dentre outros tipos de contagens, predominaram na mídia. No entanto, à medida que os efeitos da crise se desdobraram – e continuam a se desdobrar – os meios de comunicação começaram a se atentar para outros tipos de dados que extrapolam a mera contagem de casos confirmados e taxas de mortalidade. Nesta linha de pensamento, o exemplo mais emblemático que podemos citar é o gráfico do “achatamento da curva” que, nos perdoem o duplo sentido, *viralizou* (Figura 1) e ganhou inúmeras adaptações. No início de maio, o resultado da busca por “flatten the curve graphic” no Google Images já exibe incontáveis versões do gráfico. No entanto, a maioria das adaptações são meramente estilísticas, não trazendo nenhuma informação adicional relevante para o conteúdo.

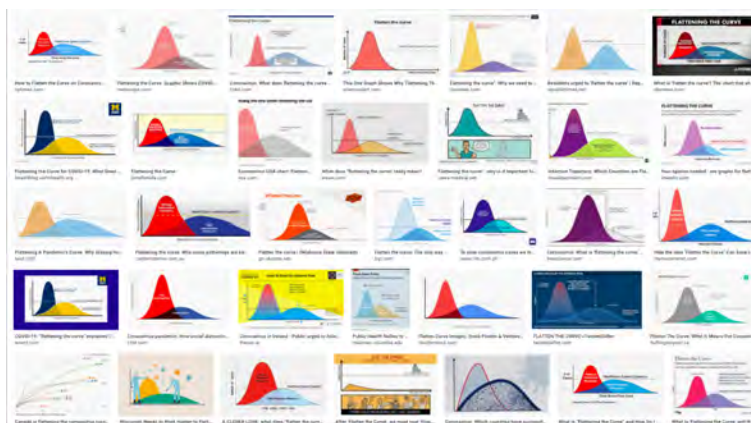


Figura 1 - O resultado da busca por “*flatten the curve graphic*” no *Google Images* exibe incontáveis versões do gráfico.
Fonte: Google/captura de tela.

Mas, para falar mais sobre esse gráfico, precisamos dar um passo atrás. A prática e reflexão de e sobre visualização de dados possui um longo histórico que remonta, primeiramente, a produção empírica de mapas e gráficos por pioneiros como Joseph Priestley, William Playfair e Charles Joseph Minard, e, posteriormente, a sistematização propriamente dita de uma teoria para a linguagem visual-gráfica proposta por estudiosos como Jacques Bertin, Edward Tufte, Johanna Drucker, entre outros. Nos últimos anos, a visualização de dados popularizou-se, tornando-se mais presente em nossas experiência diárias. Nós a vemos em notícias jornalísticas (impressas e televisivas), mídias sociais, livros didáticos, campanhas publicitárias, *dashboards* interativos, estudos científicos, relatórios, etc. sendo usada para diferentes fins: analisar problemas, comunicar informação e tomar decisões baseadas em *evidências*.

Evidência é a qualidade daquilo que é evidente, do que não dá margem à dúvida. Constitui-se como uma das principais prerrogativas para o emprego de metodologias e técnicas baseadas em dados. Nesse sentido, visualizações de dados baseadas no tempo tradicionalmente são empregadas em dois tipos de situação:

- Evidenciação de eventos concretos e finalizados, isto é, conjuntos de dados com início, meio e fim conhecidos;
- Evidenciação de eventos concretos e em processo, isto é, conjunto de dados atualizados em tempo real em um ambiente controlado.

Nenhuma das situações, contudo, descreve o fenômeno Covid-19: a pandemia, no momento, é um evento em processo em um ambiente não controlado.

O fato é que a situação impulsionada pela Covid-19 é inédita tanto em termos da velocidade e imprevisibilidade como o vírus se propaga, como em termos do ritmo acelerado em que notícias e análises são publicadas e divulgadas na mídia. A simultaneidade entre fenômeno e análise do fenômeno faz com que se consolide uma nova abordagem para a visualização de dados. Chamaremos essa nova abordagem de *visualização de dados hipotéticos e monitorados* que tem como principal exemplo, no atual contexto, o gráfico do “achamento da curva”. Nessa abordagem, a visualização de dados desloca-se de sua função primordial que é evidenciar eventos concretos para demonstrar conceitos e modelos matemáticos a partir de dados obtidos em simulações.

2. Visualização de dados hipotéticos

A matéria do The Washington Post publicada em 14 de março e assinada por Harry Stevens tem como objetivo principal explicar a um público leigo diferentes cenários de propagação da Covid-19. Para tal, faz uso de recursos gráficos e narrativos. Primeiramente, a matéria introduz o conceito de curva exponencial através de um gráfico (Figura 2) que mostra a

evolução da doença nos Estados Unidos entre 22 de janeiro (data do primeiro caso confirmado no país) até 13 de março. A curva exponencial se comporta como uma função exponencial. O gráfico é representado em escala linear e é eficaz na evidenciação da curva acentuada que retrata a rápida propagação do vírus na população.



Figura 2 – A forma do gráfico é equivalente a uma função exponencial. Fonte: The Washington Post.

O uso de ícone no meio da frase (Figura 3) enfatiza a forma típica de uma curva exponencial e reforça o argumento de que, caso o número de casos continue a dobrar a cada três dias, haverá cerca de cem milhões de casos nos Estados Unidos até maio. A matéria, contudo, não oferece uma visão do gráfico em escala logarítmica.


Essa chamada **curva exponencial**  preocupa os especialistas. Se o número de casos continuar dobrando a cada três dias, haverá cerca de cem milhões de casos nos Estados Unidos até maio.

Figura 3 – O uso de ícone no meio do texto reforça a forma típica de uma função exponencial. Fonte: The Washington Post.

Na sequência, a matéria se esforça para esclarecer porque a simulação da propagação de epidemias em cenários de “distanciamento social”, segundo profissionais de saúde pública, é a melhor abordagem para desacelerar o contágio e, conseqüentemente, minimizar o impacto da doença. Para elucidar tal argumento, a matéria introduz noções básicas de sistemas dinâmicos em situações epidemiológicas a partir da constituição de um cenário de propagação de uma doença fictícia, chamada de *simulitis*. A invenção de uma doença para ilustrar o caso ao invés de basear-se na própria Covid-19 não é claramente justificada. No entanto, pode ser compreendida como uma estratégia narrativa para simplificar a explicação, eliminando parâmetros não essenciais para uma compreensão inicial.

A explicação da propagação da doença *simulitis* é feita em dois níveis: o primeiro para aclarar as categorias de pessoas envolvidas em um cenário epidemiológico e o segundo para simular a disseminação em uma pequena população.

Em um primeiro nível de explanação, sem mencionar que está explicando o modelo *SIR*, o texto apresenta a divisão de uma dada população em três categorias: pessoa saudável (equivalente à categoria *S*, de Suscetível), pessoa doente (equivalente à categoria *I*, de Infectada) e pessoa recuperada (equivalente à categoria *R*, de Recuperada). Para esclarecer a dinâmica de contágio, a matéria apresenta duas animações consecutivas. A primeira animação reforça algo talvez já bem compreendido no senso comum que é a disseminação da doença viral a partir do contato entre uma pessoa doente e uma pessoa saudável. Quando uma bolinha marrom encosta em uma bolinha azul, a última adquire a cor da primeira, uma metáfora visual do contágio (Figura 4). A segunda animação demonstra que uma pessoa doente se transforma, provavelmente, em uma pessoa recuperada (Figura 5).



Figura 4 - Quadro da animação explicando a dinâmica de contágio. Fonte: The Washington Post.



Figura 5 - Quadro da animação explicando a dinâmica de contágio e recuperação. Fonte: The Washington Post.

O texto, ao mesmo tempo que descreve a animação serve de legenda visual para entender as cores utilizadas (Figura 6).

Vamos chamar nossa falsa de doença de 'simulitis'. Ela se espalha ainda mais facilmente do que a covid-19: sempre que uma **• pessoa saudável** entra em contato com uma **• pessoa doente**, ela também fica doente.

Na vida real, é claro, as pessoas acabam se recuperando. Uma **• pessoa recuperada** não pode transmitir simulitis a uma pessoa saudável nem adoecer novamente depois de entrar em contato com uma pessoa doente.

Figura 6 - O texto é codificado visualmente para funcionar como legenda das animações. Fonte: The Washington Post.

Em um segundo nível de explanação, a matéria simula o modelo *SIR* em uma cidade também fictícia com população de 200 pessoas diante de quatro cenários: *a* - circulação livre para todos; *b* - quarentena forçada (ou tentativa de quarentena); *c* - distanciamento moderado e; *d* - distanciamento extensivo. Cada cenário é comunicado ao leitor através de uma animação composta por três elementos visual-gráficos: 1) contagem; 2) gráfico de mudança no tempo e; 3) uma simulação de circulação e contágio entre indivíduos da população. A Figura 7 ilustra

os três elementos para um determinado momento de contágio no cenário *a*, isto é, sem nenhuma ação de distanciamento social e a Figura 8 ilustra um determinado momento do cenário *b*, isto é, de quarentena forçada tendo como referência a prática imposta pelo governo chinês na província de Hubei, o marco zero da Covid-19. A animação deste último cenário inclui uma barreira que separa pessoas doentes de pessoas saudáveis, mas à medida que o tempo passa a barreira se rompe, demonstrando à impraticabilidade de quarentenas forçadas sem uma campanha de conscientização da população.

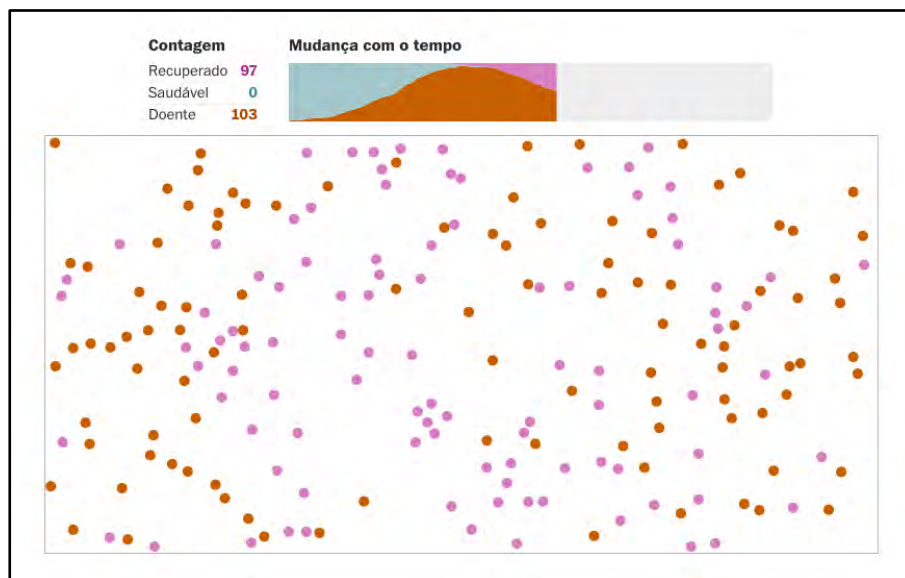


Figura 7 - Quadro da simulação da propagação do *similitis* no cenário *a*, circulação livre para todos. Fonte: The Washington Post.

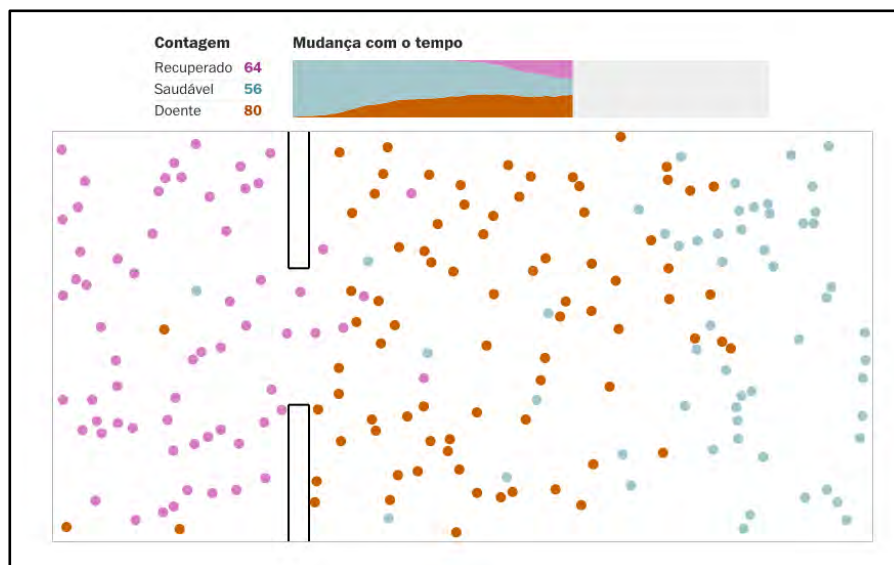


Figura 8 - Quadro da simulação da propagação do *similitis* no cenário *b*, quarentena forçada. Fonte: The Washington Post.

Após apresentar as animações geradas para cada cenário, a matéria reforça que elas são baseadas em dados hipotéticos e que, a cada nova simulação, os dados serão renderizados de modo diferente. Contudo, mesmo com dados diferentes a cada novo *play* da animação, há

uma forma padrão nos gráficos que é mantida. A Figura 9 sintetiza a forma padrão dos gráficos gerados para cada simulação.



Figura 9 - A forma padrão dos gráficos simulados para cada um dos cenários de propagação de uma epidemia. Fonte: The Washington Post.

Esses gráficos são, mais especificamente falando, gráficos de área que comparam mudanças no tempo mostrando a proporção do total que cada categoria ocupa em determinado momento. Este tipo de visualização gera *insights* sobre tendências gerais e valores relativos. Por exemplo, nota-se que no:

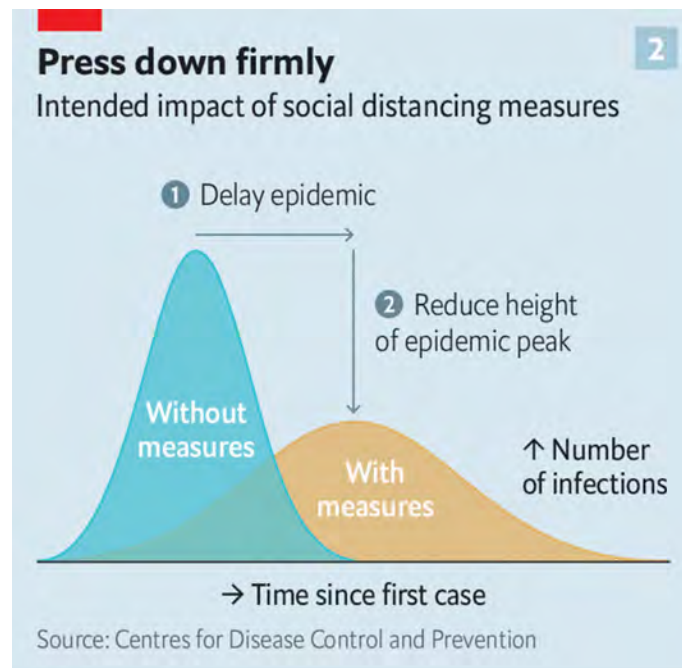
- cenário *a* - circulação livre para todos: aproximadamente toda população é infectada e a epidemia é erradicada em menor período de tempo;
- cenário *b* - tentativa de quarentena: aproximadamente toda população é infectada, mas a epidemia acontece em duas ondas o que leva mais tempo para erradicar totalmente a doença;
- cenário *c* - distanciamento moderado: uma parcela menor é infectada em um período maior de tempo, o que leva mais tempo para erradicar totalmente a doença;
- cenário *d* - distanciamento extensivo: uma parcela ainda menor é infectada e a epidemia é erradicada em um período de tempo semelhante ao do cenário *a*.

Sem se aprofundar na análise dos gráficos e nas implicações de cada um dos cenários, o jornalista sustenta que "o distanciamento social moderado geralmente supera a tentativa de quarentena, e o amplo distanciamento social geralmente é o que funciona melhor".

A matéria do The Washington Post simplifica o complexo e dinâmico fenômeno de propagação de uma epidemia. Em condições concretas, outros fatores precisam ser considerados, inclusive a taxa de mortalidade de uma doença que, no caso da Covid-19, é real. Com exceção do primeiro gráfico (Figura 1) que ilustra o conceito de curva exponencial a partir de dados de um evento real e completo, as visualizações apresentadas na matéria são produzidas a partir de dados hipotéticos a partir de um modelo genérico. A riqueza da abordagem da matéria não é prever a evolução da Covid-19 para determinado país ou população, mas sim visualizar diferentes cenários de propagação de uma epidemia. Esses cenários, por sua vez, são essenciais à compreensão do conceito "achatamento da curva" que, embora não mencionado nesta matéria, foi amplamente disseminado por órgãos de saúde pública e meios de comunicação.

O jornal The Economist foi o primeiro grande veículo de comunicação a divulgar amplamente o gráfico que ficou popularmente conhecido como o modelo conceitual do "achatamento da curva" (do inglês, *flatten the curve*). Publicada em 29 de fevereiro, a matéria traz um gráfico (Figura 10) criado pelo jornalista visual Rosamund Pearce baseado em um

gráfico previamente criado pelo órgão americano *Centers for Disease Control and Prevention* (CDC), em 2017 (Figura 11).



The Economist

Figura 10 – O primeiro gráfico do “achatamento da curva” a ganhar visibilidade na mídia. Fonte: The Economist.

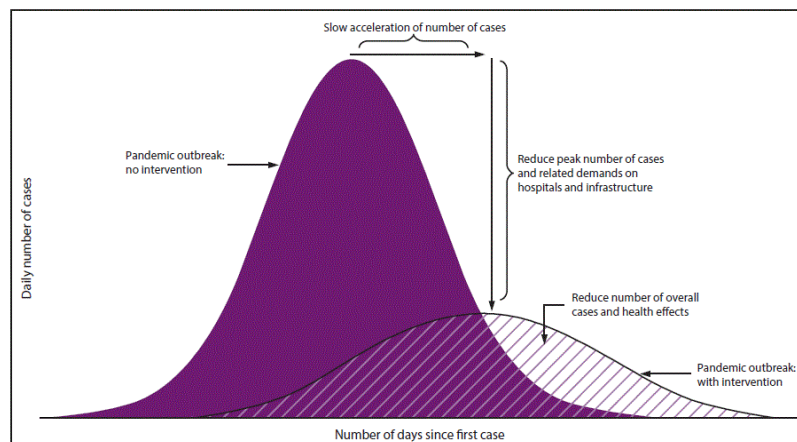


Figura 11 – O gráfico original do “achatamento da curva” publicado em relatório de pesquisa. Fonte: CDC.

A Figura 10 é uma representação simplificada e esquemática para a quantidade de pessoas infectadas ao longo do tempo. Não há dados concretos sendo mapeados neste gráfico, tanto que não há unidades de medição representadas. O gráfico é uma abstração de curvas obtidas em simulações de propagação de uma dada epidemia, no caso a Covid-19. Como mencionamos na análise da matéria do The Washington Post, essas simulações são criadas a partir do modelo *SIR* levando em consideração diferentes cenários de propagação. Embora os resultados das simulações variem e dependam de diferentes parâmetros, há tendências-chave esperadas para a dinâmica de contágio. O que a Figura 10 mostra é a sobreposição de duas tendências de curva para pessoas infectadas. A primeira delas, em azul e legendada como

without measures, tem um pico acentuado, indicando o surto de coronavírus em um curto período de tempo. Essa curva pode ser associada à curva de pessoas infectadas em um cenário sem medidas de contenção da doença (equivalente ao cenário de circulação livre para todos na matéria do The Washington Post). A outra curva, em laranja e legendada como *with measures*, é mais suave (ou achatada), indicando uma taxa mais gradual de infecção por um longo período de tempo. Essa curva pode ser associada à curva de pessoas infectadas em um cenário com medidas de contenção da doença (equivalente ao cenário de distanciamento social na matéria do The Washington Post). O desenho das curvas é modelado por uma variável chamada taxa reprodutiva (R) que diz respeito ao número de casos que cada novo caso de infecção dará origem.

A matéria do The Economist ressalta que a curva mais suave resulta em menos pessoas infectadas simultaneamente, o que reduz as chances de um eventual colapso do sistema de saúde resultando em menos mortes. No entanto, essa informação é trazida somente no texto. Após sua primeira publicação na grande mídia, o gráfico do “achatamento da curva” foi adaptado incontáveis vezes, dentro e fora do contexto jornalístico. *Vox* (Figura 3), *New York Times* (Figura 4) e *Nexo* (Figura 5) foram alguns dos veículos que adaptaram o gráfico do “achatamento da curva” tornando mais claro para um público leigo o que realmente está em jogo entre uma curva e outra: a capacidade dos sistemas de saúde – e aqui podemos considerar a disponibilidade de profissionais, número de leitos hospitalares, respiradores, etc. – de prestar serviço àqueles que precisam. Para destacar essa diferença, as novas versões do gráfico do “achatamento da curva” trazem uma linha pontilhada representando o limite da capacidade do sistema de saúde. Assim como os outros elementos gráficos presentes no gráfico, a linha não é mapeada em termos de valores concretos. Não se conhece ao certo o número de casos da Covid-19 que irá sobrecarregar o sistema de saúde, mas a posição da linha tangente à curva achatada reitera visualmente o argumento de que existe somente um cenário no qual a epidemia pode ser enfrentada sem grandes riscos à população: o cenário de medidas de contenção através de distanciamento social.

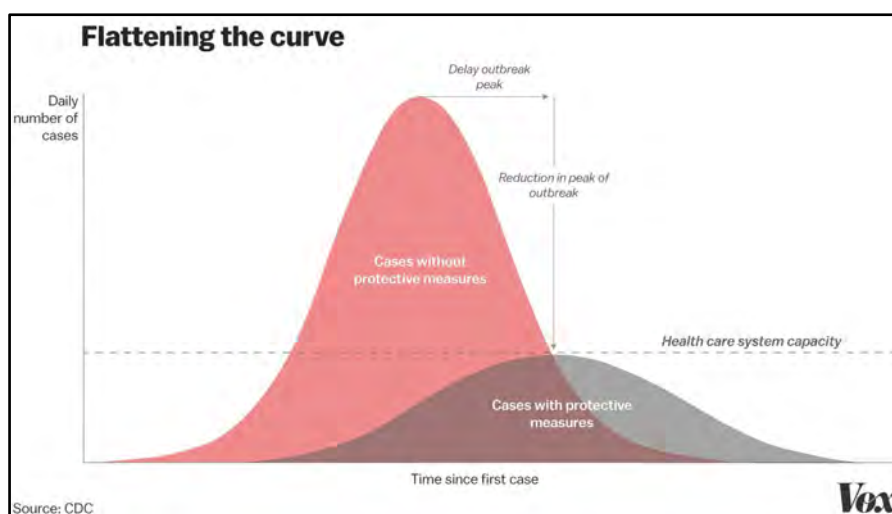


Figura 12 - Gráfico de “achatamento da curva” pelo veículo Vox. Fonte: Vox.

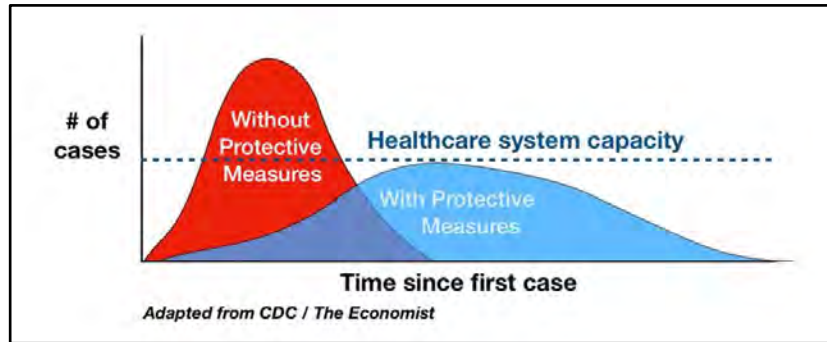


Figura 13 - Gráfico de “achato da curva” pelo veículo New York Times. Fonte: NYT.

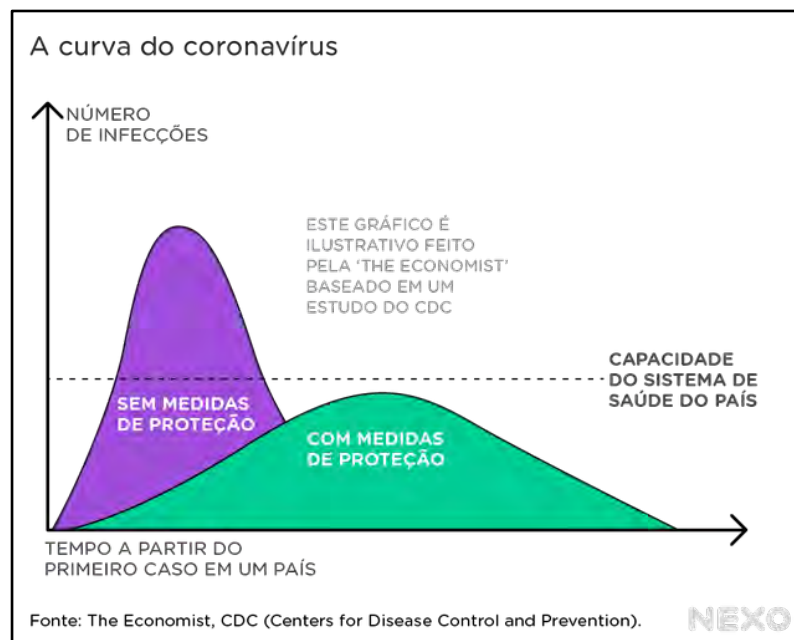


Figura 14 - Gráfico de “achato da curva” pelo veículo Nexo. Fonte: Nexo.

3. Visualizações baseadas em dados secundários

Para além do gráfico “achato da curva”, há uma segunda onda de visualizações do dados que reflete os desdobramentos implicados na prática de distanciamento social. São visualizações baseadas em dados secundários, isto é, indiretamente relacionados ao avanço da pandemia. Seleccionamos algumas visualizações para comentar e as dividimos em três eixos: impacto sócio-ambiental, impacto socioeconómico e hábitos de consumo e de busca na internet.

3.1 Impacto sócio-ambiental

O The New York Times publicou no dia 17 de março imagens de satélites que mostram os efeitos do novo coronavírus na Itália e na China. O referido satélite (no caso, o Sentinel-5P) detecta traços de atividade humana - emissões de tubos de escape de carros e caminhões, combustível fóssil queimado em usinas de energia e outras atividades industriais - que podem ser sintetizados em imagens, mais especificamente visualizações científicas. Quando comparadas com imagens do mesmo período há um ano atrás, as visualizações deste ano

evidenciam uma redução significativa das concentrações de dióxido de nitrogênio na Itália (Figura 15) e na China (os primeiros países a adotarem medidas restritivas visando a limitação de movimento de pessoas).

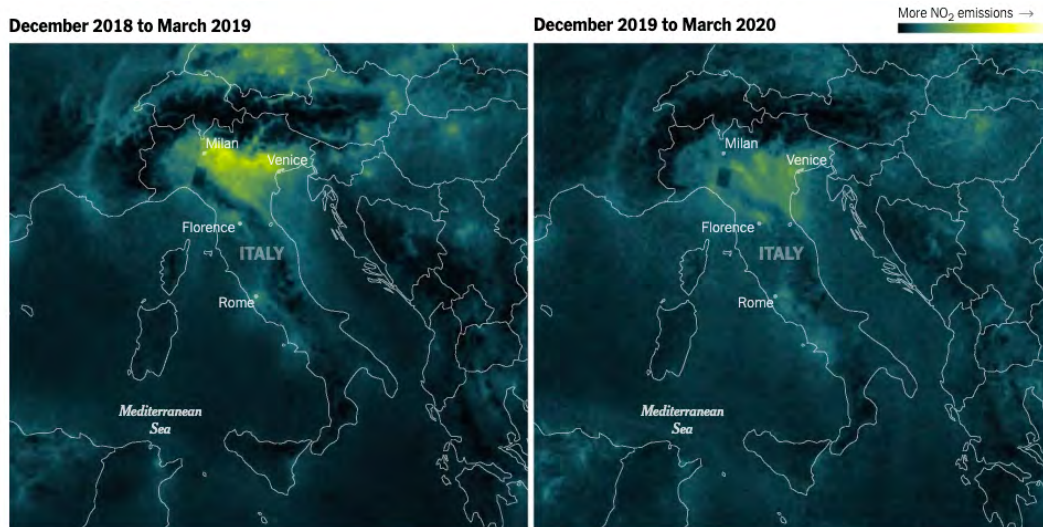


Figura 15: cada quadro na animação representa um composto de 20 dias de concentrações de dióxido de nitrogênio, de meados de dezembro a meados de março. Fonte: dados de satélite do Sentinel-5P processados pelo Descartes Labs.

O MIT Media Lab buscou uma abordagem semelhante ao publicar um estudo sobre como a prática de distanciamento social está alterando a mobilidade das pessoas na cidade de Nova York. No entanto, a natureza e origem dos dados é completamente diferente de imagens de satélite. Os dados de mobilidade utilizados na visualização foram fornecidos pela Cuebiq, uma empresa de inteligência e medição de localização. A empresa fornece registros anônimos de localizações GPS de usuários que optaram por compartilhar seus dados anonimamente nos EUA, de 26 de fevereiro de 2020 a 25 de março de 2020.

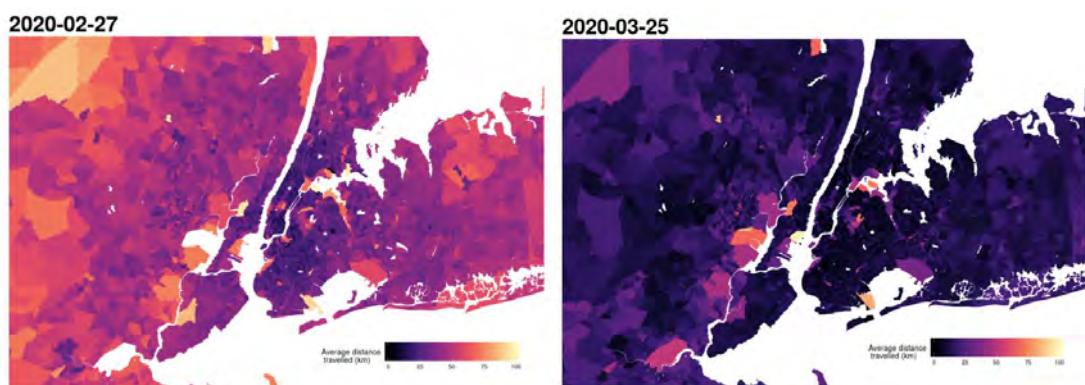


Figura 16: cada quadro na animação representa uma média de distância percorrida por dia por usuários de GPS anônimos na cidade de Nova York. Fonte: MIT Media Lab.

3.2 Impacto sócio-econômico

O jornal The New York Times, em sua versão impressa, publicou nos dias 27 de março e 9 de maio de 2020, duas capas visualmente impactantes (Figura 17) sobre a onda de desempregos que atinge os Estados Unidos nas últimas semanas. As duas capas utilizam uma estratégia

similar: recorrem a gráficos em escala linear para destacar o pico desproporcional de desempregos nas últimas semanas.



Figura 17: Capas do The New York Times de 27 de março e 9 de maio utilizam gráficos em escala linear para enfatizar o aumento de desemprego nos Estados Unidos devido à pandemia. Fonte: NYT.

Em ambas capas, os gráficos foram utilizados para apontar uma correlação entre o aumento de desemprego nos Estados Unidos e a evolução da pandemia. E ao ocupar a coluna da direita exclusivamente com dados dos gráficos, as capas reforçam visualmente a intensidade da crise: a verticalidade da coluna intensifica a agudeza ascendente ou descendente da última linha do gráfico.

O gráfico da capa de 27 março mostra o histórico de solicitações semanais por seguro desemprego no país desde os anos 2000 até a última semana de março de 2020. Os dados são do Departamento do Trabalho dos Estados Unidos. A maioria das linhas variam poucos centímetros de altura umas das outras. O maior índice de desemprego semanal anterior até então é de cerca de 600.000 durante a recessão de 2008-09 (essa informação aparece anotada no gráfico como um elemento comparativo). Mas essa linha é superada radicalmente pela última linha que representa o vertiginoso aumento de reivindicações por seguro desemprego na semana de 20-27 de março: a linha ultrapassa outras manchetes alcançando o topo página. Os quase 3,3 milhões de pedidos por seguro desemprego neste período é cerca de 5 vezes maior que o segundo maior índice.

Já o gráfico da capa de 9 de maio exibe a variação mensal de empregos desde 1946 até abril de 2020. Os dados também são do Departamento do Trabalho dos Estados Unidos. Quando há aumento de empregos entre meses subsequentes, a variação é positiva e é representada por uma linha ascendente em cor cinza. Quando há diminuição de desempregos, a variação é negativa e é representada por uma linha descendente em cor vermelha. Assim, a última linha – que representa o mês de abril – indica uma variação negativa de menos 20,5 milhões de empregos em relação ao mês anterior.

Ao criar visualizações, designers da informação devem eleger a melhor maneira de apresentar os dados e a decisão por um tipo ou outro de escala é uma dessas escolhas. Em uma escala linear, como a utilizada em ambos gráficos do NYT, o valor entre os intervalos no eixo y é constante, o que pode acarretar em picos acentuados. Esse tipo de escala tem prós e contras, mas, no presente contexto, a desproporcionalidade dos picos em relação ao restante dos gráficos corrobora o tom das matérias que é destacar a gravidade da crise.

É interessante destacar que dados não são unidades de informação neutras e que questioná-los é, na maioria das vezes, uma atitude crítica saudável. A matéria da Vox de 9 de abril expõe algumas possíveis ponderações sobre os dados utilizados para tratar do atual problema de desemprego. No caso do primeiro gráfico, é possível argumentar que o aumento de pedidos de seguro desemprego nos EUA aumentou drasticamente nas últimas semanas não só por causa de possíveis demissões, mas também porque uma lei recentemente promulgada no país expandiu a elegibilidade de beneficiários. Ao mesmo tempo, há uma provável subnotificação de pedidos de seguro desemprego pois os sistemas estaduais responsáveis por receber as reivindicações (tanto telefônico como online) não suportam a alta demanda de solicitações. Isso leva a crer que o real número de solicitações excede aquele registrado pelos órgãos oficiais.

Sobre o gráfico do dia 9 de maio, podemos ponderar sobre o fato da variação mensal de desemprego ser medida em números absolutos. Se fosse baseada na taxa de desemprego, será que a forma do gráfico teria a mesma força visual? Esclarecendo para os menos familiarizados com terminologias estatísticas, taxa é a razão entre duas grandezas. Então, taxa de desemprego pode ser compreendida como o número de trabalhadores desempregados dividido pela força de trabalho total. O próprio NYT pondera sobre tal aspecto ao trazer à capa um gráfico secundário – bem menor que o gráfico principal (Figura 18) – apresentando a taxa de desemprego no mesmo período. A taxa de desemprego em abril de 2020 é a maior registrada na história, mas mostrá-la visualmente é menos impactante do que apresentar a variação em números absolutos.

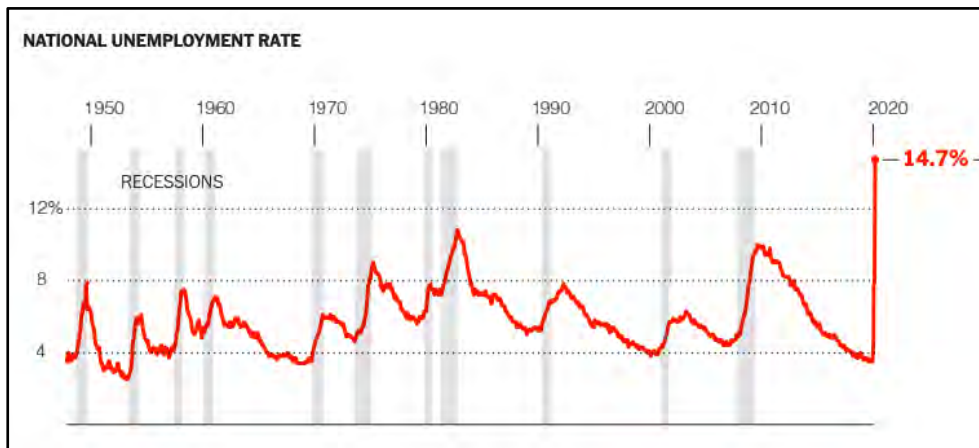


Figura 18 - Taxa mensal de desemprego nos Estados Unidos desde 1946 até abril de 2020. Fonte: NYT.

3.3 Hábitos de consumo e de busca na internet

A matéria do The New York Times de 11 de abril traz uma análise sobre como o vírus alterou profundamente a vida da população norte-americana, transformando, inclusive, a maneira como cidadãos gastam seu dinheiro. A primeira visualização (Figura 19) é uma série histórica que ilustra a variação percentual em gastos por categoria entre 8 de janeiro e 1 de abril. É nítida a queda de gastos em todas as categorias, com exceção de groceries – que inclui gastos em supermercados, lojas de conveniência, entre outros. Além disso, a acentuada curva de crescimento em groceries na primeira quinzena de março coincide com o momento de supermercados lotados no país, quando à população norte-americana, incitada pelo pânico, esvaziou prateleiras em inúmeros estabelecimentos.

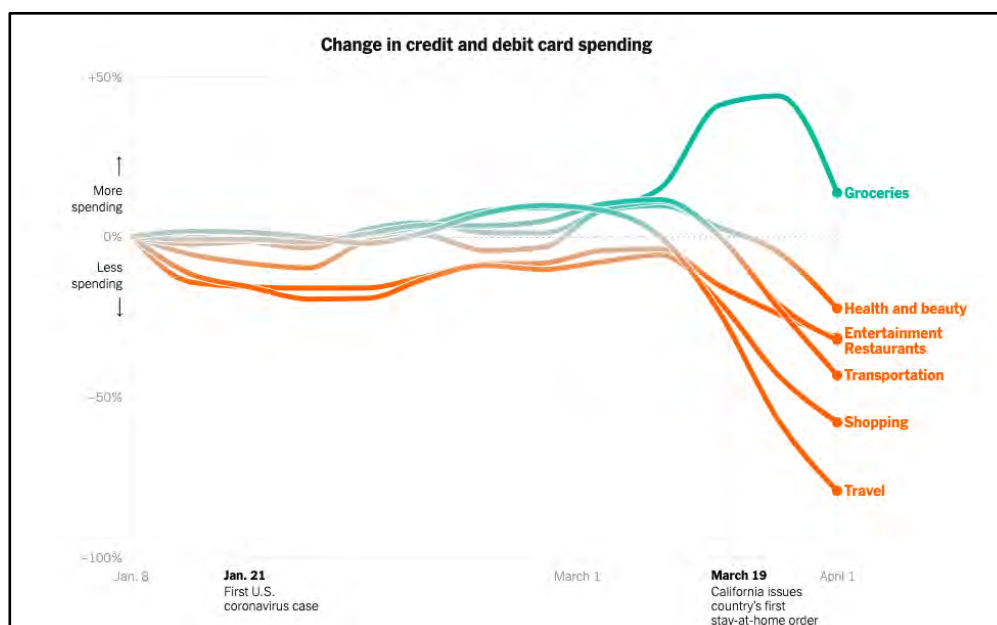


Figura 19 - Variação percentual em gastos (por categoria) dos norte-americanos desde o início do ano. Cada linha é uma média das duas semanas anteriores, o que suaviza as anomalias semanais. Fonte: NYT.

Já a segunda visualização (Figura 20) é um gráfico de bolhas que mostra a variação de gastos por setor entre a última semana de março de 2019 e o mesmo período em 2020. A área das

bolhas são dimensionadas pelas vendas do setor e as cores representam se as variações são negativas (em laranja) ou positivas (em verde). É fácil notar que os setores aéreo e cultural são fortemente impactados enquanto setores de jogos e compras online registram forte crescimento.

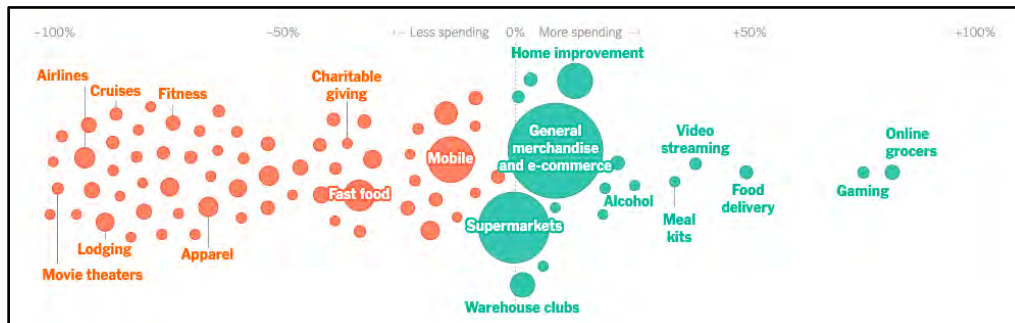


Figura 20 - Variação de gastos por setor entre a última semana de março de 2019 e o mesmo período em 2020. Fonte: NYT.

O texto da matéria elucida as fontes dos dados: *“todos os gráficos deste artigo são baseados em uma análise de dados do Earnest Research, do New York Times, que rastreia e analisa as compras com cartão de crédito e débito de quase seis milhões de pessoas nos Estados Unidos. Embora os dados não incluam transações em dinheiro e, portanto, não reflitam todas as vendas, eles fornecem um forte instantâneo do impacto do vírus na economia.”*

No dia 11 de maio, o Nexo publicou uma série de visualizações sobre o aumento de buscas no Google por certos termos durante quarentena. Segundo a matéria, o *Google Trends* (ferramenta do Google que mostra os mais populares termos buscados em um passado recente) chegou a registrar 142% mais buscas por *pijamas* no começo de maio em relação à média do ano. *Aspirador* e *quebra-cabeça* também tiveram um aumento de mais de 50% no interesse.

Há dois tipos de visualização de dados na matéria. O primeiro tipo (Figura 21) exibe quatro faixas de variação (-50% ou menos; 50% a 0%; 0% a 50%; e mais de 50%) de buscas em relação à média do ano (janeiro a maio de 2020). Os termos são divididos em categorias: computador/escritório, cozinha, uso pessoal, diversão, exercícios físicos e casa. O recorte abaixo apresenta a variação para os termos na categoria computador/escritório.

O segundo tipo de visualização é um detalhamento da variação de busca por termos mais buscados. Essas visualizações contam com elementos ilustrativos e evidenciam a variação dia-a-dia. Em comparação com a visualização anterior, é possível perceber agora o quanto mais um termo é mais buscado que outro.

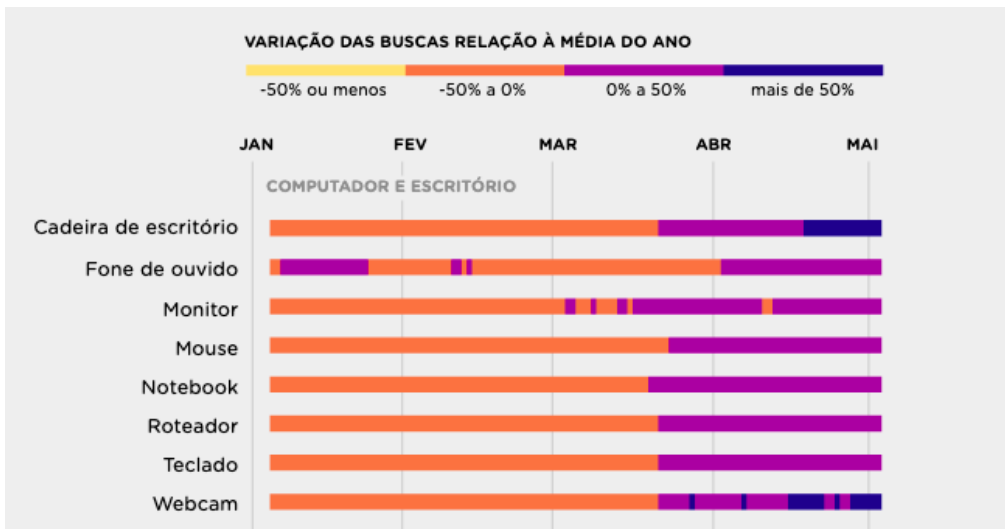


Figura 21 - Variação de buscas por termos no Google em relação à média do ano. Fonte: Nexø.

O segundo tipo de visualização (Figura 22) é um detalhamento da variação de busca por termos mais buscados. Essas visualizações contam com elementos ilustrativos e evidenciam a variação dia-a-dia. Em comparação com a visualização anterior, é possível perceber agora o quão mais um termo é mais buscado que outro.



Figura 22 - Variação de buscas pelos termos pijamas, cadeira de escritório e quebra-cabeça no Google em relação à média do ano. Fonte: Nexø.

4. Crítica social e autorreflexão

Há uma terceira onda de visualizações de dados alavancada no contexto da pandemia que extrapola a questão da visualização de dados hipotéticos através de simulações e o uso de dados secundários para evidenciar os desdobramentos da crise. Estamos vivenciando uma crise sanitária que produz enormes quantidades de dados que, por sua vez, são processados e divulgados quase que em tempo real. Contudo, esses dados, mesmo quando visualizados, não são capazes de dar conta de todas as dúvidas e amenizar nossas ansiedades. Diante, da incerteza, incompletude e complexidade do fenômeno, surgem práticas de visualização de dados guiadas por outras motivações. Podemos citar, por um lado, exemplos que promovem a crítica social e o cuidado com a comunidade e, por outro lado, exemplos que propiciam autorreflexão.

4.1 Os efeitos desiguais da pandemia

Por práticas de visualização de dados voltadas para crítica social e cuidado da comunidade, queremos enfatizar exemplos que evidenciam o impacto desproporcional da Covid-19 em segmentos mais vulneráveis da sociedade. A contribuição da jornalista e editora de dados do The Guardian, Mona Chalabi, é significativa nesta direção. Talvez por ser descendente de iraquianos e ter crescido em Londres, a jornalista desenvolveu um estilo e abordagem únicos para comunicar dados: suas visualizações (ou, se preferirem, ilustrações infográficas), desenhadas a mão expõem a imprecisão dos dados, colocam o dedo em feridas sociais e desafiam o senso comum.

No contexto do novo coronavírus, Mona desenvolveu uma série temática autoral com cerca de 20 visualizações sobre o assunto. A série, em andamento, foi divulgada em sua conta pessoal no Instagram e algumas visualizações foram publicadas em sua página. A seguir, uma seleção de três visualizações de dados que evidenciam a situação de maior vulnerabilidade de refugiados, negros e encarcerados diante da pandemia.

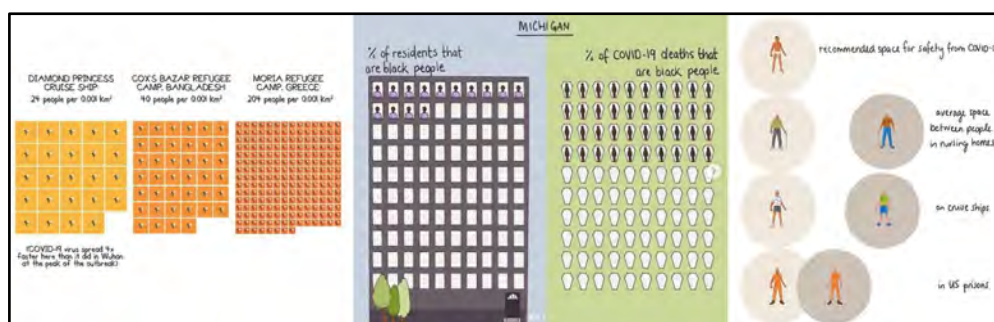


Figura 23 - Ilustrações infográficas por Mona Chalabi. Fonte

No contexto brasileiro, o economista Gabriel Vaz de Melo desenvolveu uma série de visualizações de dados bastante pertinente sobre os deslocamentos populacionais por serviços de saúde diante da escassez de leitos e equipamentos em diversas regiões do país. A partir de dados da pesquisa REGIC (Regiões de Influência das Cidades) de 2018 divulgada pelo IBGE (Instituto Brasileiro de Geografia e Estatística), Gabriel criou um mapa animado (Figura 24) que evidencia o fluxo de deslocamentos por acesso a serviços de saúde de alta complexidade. Em suas palavras: "... enquanto nas regiões Sul e Sudeste os deslocamentos para os serviços

de alta complexidade são, em média, de 100 km, no Norte e Centro-Oeste essa média sobe para 276 e 256 km, respectivamente."

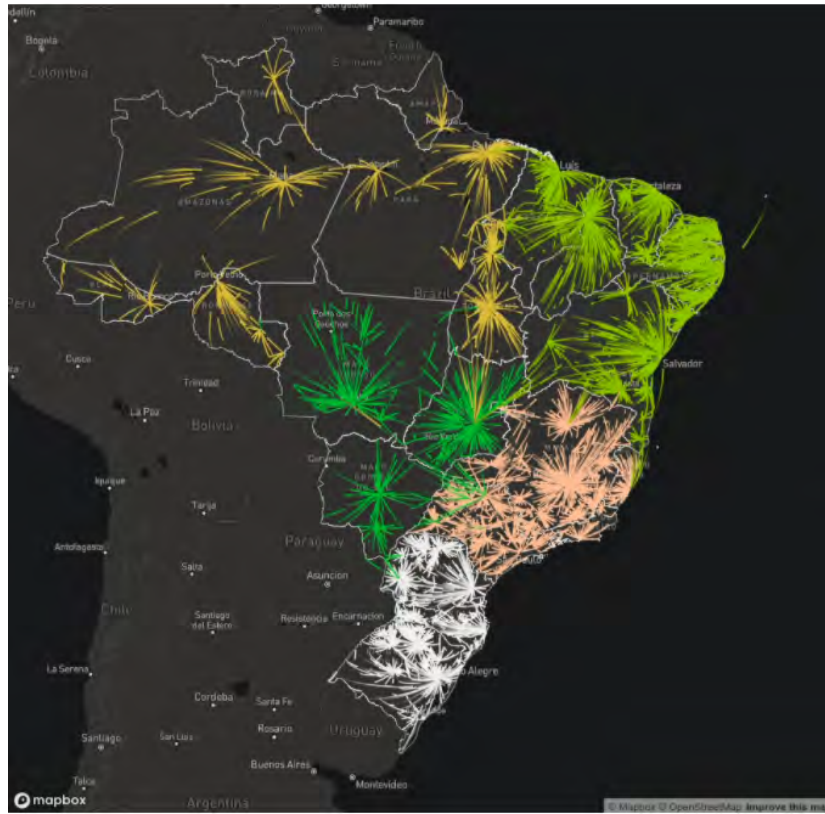


Figura 24 - Principais deslocamentos por serviços de saúde de alta complexidade segundo a REGIC 2018. Fonte: IBGE, 2020. Elaboração: Gabriel Vaz de Melo.

A partir de outras base de dados, Gabriel também criou uma visualização com dois mapas justapostos (Figura 25) em que é possível relacionar quais municípios possuem leitos de UTI no SUS e quais municípios possuem casos confirmados de Covid-19. A diferença é gritante aos olhos.

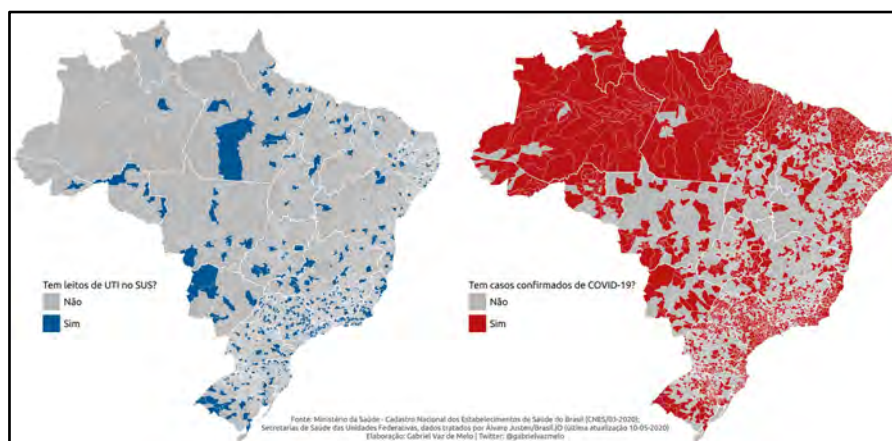


Figura 25 - Municípios que possuem leitos de UTI no SUS (à esquerda) e municípios que possuem casos confirmados de Covid-19. Elaboração: Gabriel Vaz de Melo.

4.2 Visualização como forma de autorreflexão e expressão

Por práticas de visualização de dados voltadas para o autorreflexão e expressão, queremos destacar uma mudança mais paradigmática no fluxo de produção de visualizações de dados durante a pandemia. Até o momento, citei práticas de visualizações de dados produzidas por meios de comunicação (oficiais ou alternativos) com objetivo de informar, com clareza e confiabilidade, um público amplo. Para além dessa finalidade, o contexto do Covid-19 evidencia a emergência de visualizações de dados criadas para outros objetivos como a autorreflexão e a expressão. Nessa linha, iniciativas como *Diario Visual de la cuarentena*, *Data Selfie* da quarentena e *Quarantine portrait* utilizam a visualização de dados como uma ferramenta para refletir sobre e representar o cotidiano durante o período de distanciamento social.

Data Selfie da quarentena é uma oficina virtual idealizada e ministrada por mim, Julia Giannella, para os estudantes do curso de Comunicação Visual Design da UFRJ. Inspirada no projeto *Dear Data* (Giorgia Lupi e Stefanie Posavec, 2015), a oficina tem como objetivo estimular que o participante reflita sobre o período da quarentena a partir da coleta, representação e comunicação de dados pessoais que constituem o pano de fundo de suas atividades diárias durante o isolamento. A seguir, algumas das *data selfies* criadas pelos estudantes.



Figura 26 - "Uma semana de saúde" por Marina Ramos.



Figura 27 - "Uma semana de interações" por Ana Ferreira.