Laboratório VISGRAF
Instituto de Matemática Pura e Aplicada

Visualizacao em Tempos de Corona Virus – Parte 1

Julia Giannella and Luiz Velho

Technical Report TR-20-07 Relatório Técnico

May - 2020 - Maio

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

Visualização em Tempos de Coronavírus

Parte 1

Júlia Giannella[1]
Luiz Velho[2]

[1]julia@impa.br
[2]lvelho@impa.br

Capítulo 1

Introdução

1.1 Contexto

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 o 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, nasce o portal "Coronaviz: visualização em tempos de coronavírus", uma iniciativa do Visgraf, do Instituto de Matemática Pura e Aplicada, um laboratório com foco em pesquisa e desenvolvimento de novas mídias. Ao longo deste trabalho, iremos analisar bons e maus exemplos de visualizações de dados produzidos no contexto da pandemia. Em última instância, podemos argumentar que os bons exemplos ajudam pessoas a compreenderem, 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 pande-

mia 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 podem ser situados em dois eixos, elucidados a seguir. Utilizaremos o portal Coronaviz para publicar análises e reflexões cobrindo esses eixos e as diferentes questões que eles abarcam. Eventualmente, tópicos não mencionados neste primeiro momento poderão ser incluídos nas análises do portal.

i) Diversificação da visualização de dados para além da contagem de eventos

O primeiro eixo diz respeito à diversificação da visualização de dados para além da contagem de eventos relacionados à propagação do Covid-19. 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* e ganhou inúmeras adaptações. 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 e Edward Tufte. 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 do Covid-19: a pandemia, no momento, é um evento em processo em um ambiente não controlado.

O fato é que a situação impulsionada pelo Covid-19 é inédita tanto em termos da velocidade e imprevisibilidade como o vírus se propaga, como em termos do rítmo 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* que tem como principal exemplo, no atual contexto, o gráfico do "achatamento 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. Nos aprofundaremos mais sobre esse aspecto nas análises dos gráficos publicados pelo *The Washington Post* (seção 3.1) e *The Economist* (seção 3.2).

Para além do gráfico "achatamento da curva", há uma segunda onda de visualizações de 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. Podemos citar, por exemplo, visualizações que demonstram a mudança dos padrões de gastos em [14], o aumento na taxa de desemprego [25], alterações no comportamento de mobilidade das pessoas [4], a redução da poluição urbana [20] e até mesmo o índice de transparência na divulgação de dados sobre a pandemia [18].

Finalmente, há uma terceira onda de visualizações de dados que emerge quase como uma resposta à ironia mencionada no início do texto. Há uma quantidade sem precedentes de dados sendo produzidos e divulgados 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 que evocam, por um lado, o cuidado com a comunidade e, por outro lado, o auto reflexão.

Por práticas de visualização de dados voltadas para o cuidado da comunidade, queremos enfatizar exemplos que evidenciam o impacto desproporcional do Covid-19 em segmentos mais vulneráveis da sociedade como refugiados [6], profissionais que não podem trabalhar de casa [8], negros [7], pessoas encarceradas [9], etc. Em um perspectiva stricto sensu, são visualizações passíveis de serem questionadas como tais à medida que distanciam-se de certas qualidades como abstração, precisão e objetividade. Apoiando-se em pequenos conjuntos de dados, são visualizações que utilizam recursos figurativos e expressivos para amparar críticas sociais.

Por práticas de visualização de dados voltadas para o auto reflexão, queremos destacar uma mudança mais paradigmática no fluxo de produção de visualizações de dados durante a pandemia. Até o momento, citamos práticas de visualizações de dados claramente produzidas por um meio de comunicação legitimado com objetivo de comunicar informações confiáveis a um público amplo. No entanto, o contexto do Covid-19 evidencia a emergência de visualizações de dados criadas para outro objetivo: auto reflexão e expressão. Nessa linha, iniciativas como "Diario Visual de la cuarentena" [19], "Data Selfie da quarentena" [10] e "Quarantine portrait" [16] utilizam a visualização de dados como uma ferramenta para refletir sobre e representar o cotidiano durante o período de distanciamento social.

ii) Uso responsável da visualização de dados

O segundo eixo diz respeito a responsabilidade cívica envolvida nos processos de aquisição e representação de dados durante a pandemia. A visualização de dados carrega uma aura de confiabilidade e autoridade em seu discurso apoiando a tomada de decisão tanto de gestores da administração pública como dos indivíduos. Quando essa autoridade é exercida de maneira irresponsável, ela pode esgotar a confiança do público, levar à rejeição de medidas de saúde pública e até mesmo incitar pânico.

O uso responsável da visualização de dados possui duas facetas: a primeira diz respeito à proveniência, qualidade e transparência dos dados; a segunda diz respeito aos desafios que conceitos complexos e fenômenos incertos e/ou incompletos impõem à representação visual.

Situações de crise como uma pandemia exigem uma comunicação abundante, consistente e com fontes confiáveis. E diferentes agentes assumem a mediação desse processo. Em primeiro lugar, organizações de saúde, em suas diferentes instâncias, precisam ser abertas e transparentes. Precisam esclarecer a metodologia utilizada na coleta de dados e devem informar o que sabem e, principalmente, o que desconhecem ou não têm certeza a respeito dos dados que divulgam. Em segundo lugar, os repositórios online que agregam e disponibilizam conjuntos de dados sobre o Covid-19 na internet precisam ter competência para fazê-lo. As universidades Johns Hopkins (Estados Unidos) e Imperial College (Inglaterra) são exemplos de instituições acadêmicas que criaram repositórios online confiáveis e facilmente acessíveis à comunidade [2, 1]. Em terceiro lugar, os meios de comunicação — dos grandes jornais aos veículos independentes — precisam ser criteriosos e sensatos ao selecionar e estruturar dados em visualizações e dashboards. Visualizações são recursos poderosos para comunicação de informações, mas também podem enganar, e desinformar. Sobre essa questão, a especialista Amanda Makulec publicou uma orientação para criar visualizações de dados responsáveis diante de uma crise de saúde pública [15].

Finalmente, a pandemia aponta alguns desafios para a representação visual de dados: como explicar e descrever o crescimento exponencial de forma clara e acessível? Como comparar a propagação do vírus em diferentes países de modo correto? Como explicar visualmente a incerteza em números como contagem de casos quando nem sempre os testes realizados são suficientes para torná-los indicadores confiáveis de casos reais? Ao invés de reduzir a complexidade do fenômeno, os bons exemplos de visualizações criados até o momento valem-se de convenções visuais específicas para o contexto, bem como fazem uso recorrente de anotações e ressalvas.

Capítulo 2

A Matemática da Epidemia

2.1 Formulação do Problema

Quais são os conceitos matemáticos necessários para entender a evolução de uma epidemia? Para poder identificá-los precisamos primeiro entender o fenômeno sanitário do processo epidêmico, i.e., como ele acontece no mundo real. Uma epidemia pode ser definida como a rápida disseminação de uma doença contagiosa para um grande número de pessoas de uma dada população num curto espaço de tempo. Dessa definição, destacamos "rápida disseminação ... em um curto espaço de tempo". A partir daí, podemos identificar dois conceitos matemáticos importantes relacionados com o processo epidêmico: crescimento exponencial ; e sistema dinâmico. O primeiro está relacionado com o aumento rápido de uma quantidade, enquanto o segundo diz respeito à evolução temporal.

Baseada nesses conceitos, a matemática proporciona instrumentos que potencialmente podem contribuir de maneira fundamental no estudo, contrôle e enfrentamento das epidemias. Esses instrumentos são, em última instância, modelos matemáticos que possibilitam entender e fazer predições sobre esse fenômeno através de simulações.

Várias áreas da Matemática estão envolvidas no desenvolvimento de modelos efetivos sobre as epidemias. Dentre elas, se destacam as áreas de Equações Diferenciais, Sistemas Dinâmicos, Probabilidade, Estatística, além da Ciência de Dados e Computação. Isto porque, o processo epidêmico se dá no mundo real e, como consequência, se torna necessário desenvolver modelos matemáticos abstratos que reflitam fidedignamente os aspectos relevantes do fenômeno concreto em questão. Nesse sentido, é importante enfatizar que como, em geral, não se tem acesso a todas as informações pertinentes do processo, faz-se uma amostragem (usando técnicas estatísticas) para obter um conjunto de dados que necessariamente contém algum grau de incerteza (mensurado probabilisticamente). Esses dados são incorporados em modelos da dinâmica do sistema (usando equações diferenciais) que permitem fazer simulações (por métodos computacionais).

O estudo descrito acima faz parte da Epidemiologia Matemática, um campo multidisciplinar que congrega especialistas das Ciências Médicas, Sociais e Matemáticas. Nesse contexto, estão reunidas a pesquisa e a prática. Nele, os modelos

desenvolvidos tem um duplo papel: por um lado, deseja-se entender o fenômeno conceitualmente de maneira teórica — com isso valida-se o modelo que é usado para testar hipóteses; por outro lado, o objetivo real é poder interferir no processo de modo a controlar a epidemia — o que se faz com base em predições obtidas por simulações para subsidiar a tomada de decisões. No primeiro caso, trabalha-se sobre situações ideais usando dados hipotéticos. No segundo caso, utiliza-se uma série de amostras do processo em andamento que permitem fazer análises e extrapolações.

2.2 O Processo Epidêmico e a Exponencial

Depois dessas preliminares, voltamos ao entendimento do processo epidêmico, focando nos dois aspectos destacados: a disseminação e seu efeito ao longo do tempo. A disseminação de uma doença infecciosa se dá pelo contágio, através do qual um indivíduo infectado transmite a doença a outros indivíduos não-infectados da população. Quando a incidência da doença na população é pequena, temos uma situação endêmica e depois que o número de casos da doença cresce e ultrapassa um limite superior, temos uma situação epidêmica. Ou seja, a taxa de incidência define a gravidade do efeito da doença numa determinada comunidade e se, por acaso a disseminação se espalhar por vários países temos uma pandemia.

Qual é a relação do processo epidêmico com o crescimento exponencial? Em matemática, a exponencial é uma função da forma

$$f(t) = a * b^t$$

cujo gráfico pode ser visto na Figura 2.1 (onde $a = 1$ e $b = e = 2.71828\ldots$).

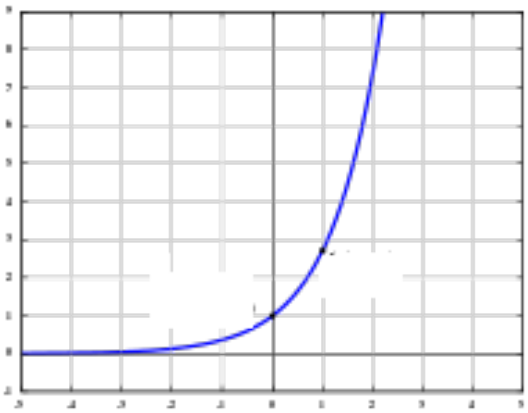

Figura 2.1: Função exponencial

O crescimento exponencial é a forma como uma quantidade aumenta com o tempo segundo $f(t)$, ou seja o parâmetro t da função é o tempo e o seu valor a quantidade de interesse. Na equação $a * b^t$, a constante a é o valor inicial no tempo $t = 0$, e a constante b é o "fator de crescimento". Uma característica da exponencial é que sua derivada (taxa de crescimento) é diretamente proporcional

ao valor da função. Podemos ver isso, no caso de uma discretização de t em intervalos de tempo unitários (i.e., $t = 1, 2, \ldots$). Temos então

$$f(t+1) = a * b^{t+1} = a * b^t * b = f(t) * b$$

A inversa da função exponencial $f = e^t$ é o *logaritmo*, denotado por log, devido a esse fato, a exponencial é também conhecida como *anti-logarítmo*. A função logarítmica desempenha papel importante em cálculos envolvendo a exponencial.

A relação da exponencial com a modelagem de epidemias é devido ao fato que, segundo estudos epidemiológicos, o primeiro período de um surto epidêmico segue um crescimento exponencial. Para ilustrar esse fato, consideramos um caso hipotético no qual, temos uma única pessoa infectada (i.e., $a = 1$) e cada pessoa infectada transmite a doença para duas outras pessoas (i.e., $b = 2$). Isso resulta na formula para essa epidemia $f(t) = 1 * 2^t$

Com essa fórmula, podemos calcular o valor de $f(t)$ para $t = 1, \ldots, 14$, obtendo assim o número de pessoas infectadas em cada intervalo. Vemos que, com um fator de crescimento de 2, teremos mais de 1600 casos após 14 dias — como pode ser verificado na Tabela 2.1 e no gráfico da Figura 2.2 (baseado em [12]).

Tempo	Infectados
0	1
1	2
2	4
3	8
4	16
5	32
6	64
7	128
8	256
9	512
10	1024
11	2048
12	4096
13	8192
14	16384

Tabela 2.1: Evolução da epidemia

Ocorre que, depois dessa fase inicial a evolução da propagação da epidemia deixa de seguir o modelo exponencial pois a população tem um tamanho finito e todos seus individuos são infectados de modo que o crescimento termina. Além disso, deve-se considerar que os individuos após contrair a doença eventualmente serão curados e deixarão de ser transmissores dela.

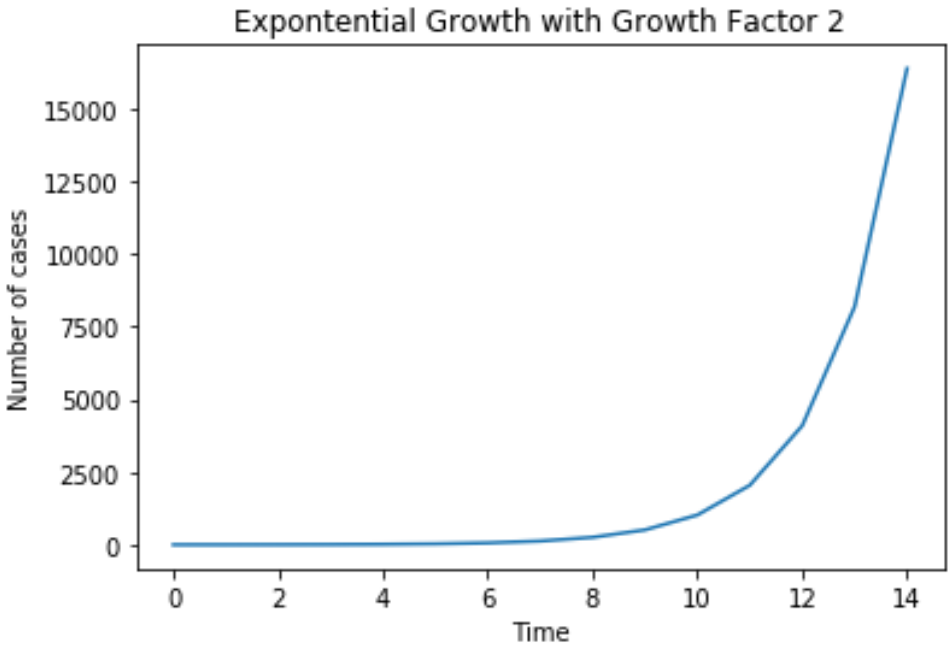

Figura 2.2: Gráfico do crescimento exponencial com os dados da Tabela 2.1

2.3 A Modelagem pela Função Logística

Vimos anteriormente que a fase inicial de uma epidemia segue um crescimento dado pela função exponencial.

Na realidade, para se obter um modelo matemático da epidemia na totalidade, englobando suas várias fases, temos que considerar que após o período inicial de crescimento, numa segunda fase deve ocorrer um decrescimento até se chegar num patamar máximo.

A função *Logística* tem as características descritas acima e estudos epidemiológicos concluíram que processo completo da epidemia pode ser modelado pelo crescimento logístico.

A formula da função logística é dada pela seguinte expressão

$$f(t) = \frac{c}{1 + a * e^{-bt}}$$

onde $f(t)$ é o número de casos no tempo t, a constante $b > 0$ e a capcidade máxima de f é dada pelo valor limite c.

Além disso, temos que o *valor inicial* de f (i.e., o número de casos no começo do surto) é dado por $\frac{c}{(1+a)}$ e a *taxa máxima de crescimento* ocorre quando $t = \frac{\ln(a)}{b}$ e $f(t) = \frac{c}{2}$.

Para exemplificar o uso da função logística como modêlo da evolução de uma epidemia, podemos supor um caso hipotético onde o número máximo de pessoas doentes é 1000 (i.e., tamanho da população) e o processo começa com apenas uma pessoa doente, que pode infectar duas outras pessoas.

Dessa forma, colocamos na formula os seguintes valores: $c = 1000$, $a = 999$ e $b = 2$. O gráfico da função logística correspondente a simulação de um periodo de

10 dias (i.e., $t = 0, \ldots, 10$), pode ser visto na Figura 2.3, Note que a curva cresce de forma acelerada até perto de $t = 3$, onde começa a desacelerar para atingir o patamar de 1000.

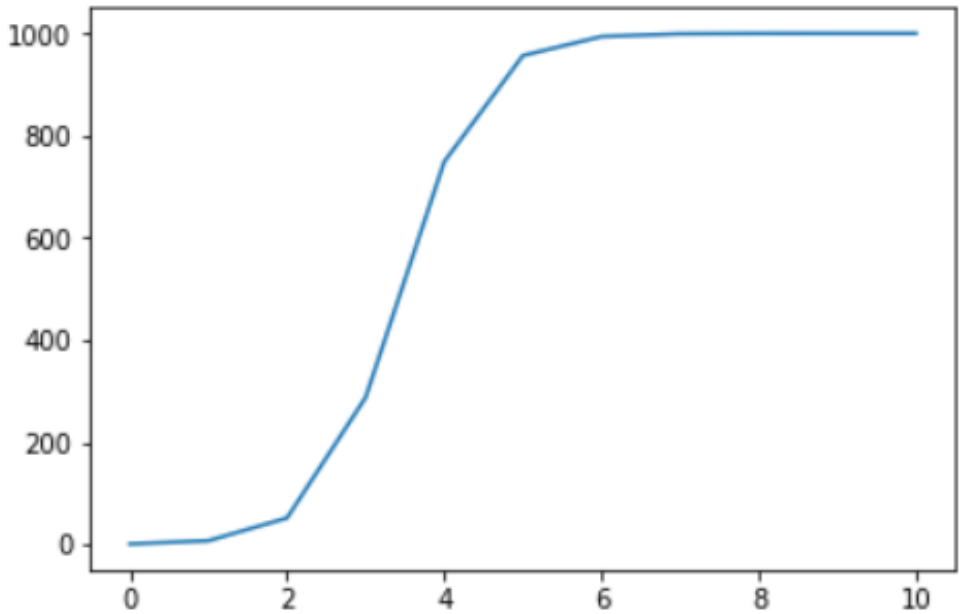

Figura 2.3: Função logística (de [13])

2.4 A Dinâmica da Epidemia e o Modelo SIR

Nesse ponto, para continuarmos nosso endendimento da epidemia, devemos estudar o processo como um sistema dinâmico. A classe de modelos epidemiológicos para calcular o número teórico de pessoas infectadas por uma doença contagiosa ao longo do tempo em uma população fechada, é conhecido com modelo SIR. Ele tem esse nome porque consiste de um conjunto de equações diferenciais acopladas relacionando o número de indivíduos da população de acordo com o seu estado da doença. Assim, temos as pessoas *Susceptíveis*, $S(t)$, as pessoas *Infectadas*, $I(t)$ e as pessoas *Removidas*, $R(t)$ (i.e., que foram curadas ou morreram, deixando de influir no processo epidêmico).

O mais simples dessa classe de modelos é o o chamado modelo de [11]. Esse modelo assume que o tamanho da população é fixo, o período da incubação do agente infectado é instantâneo e a duração da infectividade é mesma da doença. Alem disso, ele supõe atuação igual sobre toda a população.

O modelo consiste de um sistema de três equações diferenciais ordinárias não-lineares,

$$\frac{dS}{dt} = -\beta S(t)I(t)$$
$$\frac{dI}{dt} = \beta S(t)I(t) - \gamma I(t)$$
$$\frac{dR}{dt} = \gamma I(t)$$

onde β é a taxa de infecção e γ é a taxa de recuperação.

A evolução desse sistema de equações ao longo do tempo é governada pelo valor do *limiar epidemiológico*

$$R_0 = \frac{\beta S(t)}{\gamma}$$

Quando $R_0 > 1$ cada pessoa doente irá infectar mais de uma pessoa e a epidemia crescerá exponencialmente. Por outro lado, se $R_0 < 1$ cada pessoa doente infectará menos que uma pessoa, causando uma diminuição da epidemia.

Para simular computacionalmente o processo, recorremos à integração numérica das equações, usando o método de Euler.

$$s_n = s_{n-1} - \beta s_{n-1} i_{n-1} \Delta t$$
$$i_n = i_{n-1} + (\beta s_{n-1} i_{n-1} - \gamma i_{n-1})\Delta t$$
$$r_n = r_{n-1} + \gamma i_{n-1} \Delta t$$

Baseado nos dados reais provenientes de amostras coletadas sobre um determinado período, podemos estimar os valores dos parâmetros β e γ e realizar simulações.

Como exemplo, vamos usar a epidemia pelo virus da gripe de Hong Kong na cidade de Nova Iorque no final dos anos 1960. Estimando um período médio de infecção de três dias resulta em $\gamma = \frac{1}{3}$. Supondo que cada pessoa infectada irá contaminar outra pessoa a cada dois dias, temos $\beta = \frac{1}{2}$. Temos como valores iniciais da simulação $S_0 = 7.900.000$ (população de Nova Iorque em 1968-1969), $I(0) = 10$ (nível de inicial de infecção por 10 pessoas) e $R(0) = 0$.

De posse desses dados e escolha de valores dos parâmetros, calculamos as curvas de solução do modelo para um periodo de 140 dias que pode ser vista no gráfico da figura 2.4, onde as quantidades estão normalizadas (i.e., $s(t) = \frac{S(t)}{N}$, $i(t) = \frac{I(t)}{N}$, $r(t) = \frac{R(t)}{N}$ e $N = S_0$.

Duas observações importantes sobre esse gráfico. Primeiro, a curva $r(t)$ segue a função *Logistica* de crescimento que, ao contrario da função exponencial, modela a evolução da infecção em todas suas fases [13]. Segundo, apesar do nível de $i(t)$ parecer baixo, a sua influêcia é grande devido ao fator exponencial. Note que o ponto de máximo dessa curva corresponde ao ponto de inflexão de $r(t)$ onde a derivada (taxa de crescimento) muda de sinal.

Embora o modelo de Kermack-McKendrick descreva a essência da dinâmica de uma epidemia, ele inclui apenas os elementos principais do processo. Modelos

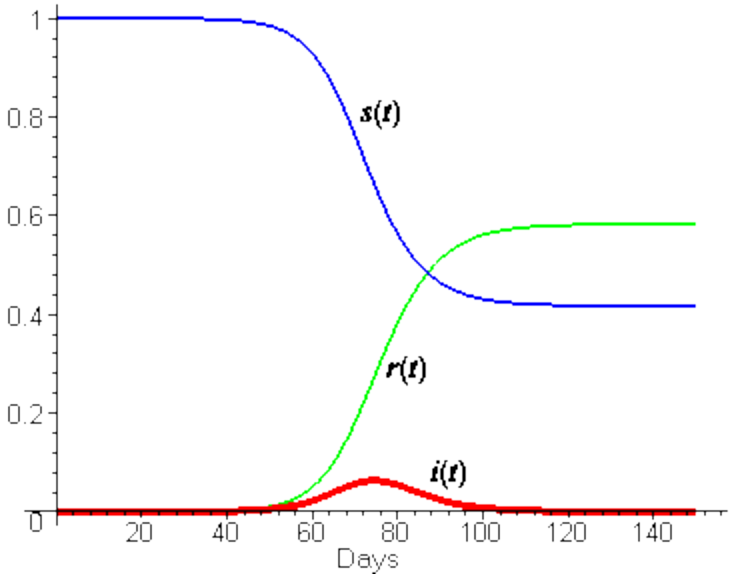

Figura 2.4: Epidemia da gripe de Hong Kong em Nova Iorque (de [23]).

mais complexos incluem outras categorias de individuos na população, como o modelo SIRD (em inglês: *Susceptible-Infected-Recovered-Deceased*), que divide as pessoas removidas em *Recuperadas* e *Mortas*. Com isso, outros parâmetros são incluídos, nesse caso as taxas de recuperação e mortalidade.

Além disso uma análise completa deve considerar outros aspectos epidemiológicos e sociais que são necessários para modelar de maneira fidedigna o fenômeno.

Capítulo 3

Análise

3.1 Visualização de dados hipotéticos

A matéria do *The Washington Post* publicada em 14 de março e assinada por Harry Stevens [24] tem como principal objetivo explicar a um público leigo diferentes cenários de propagação do 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 3.1) 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. O uso de ícone no meio da frase (Figura 3.2) 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.

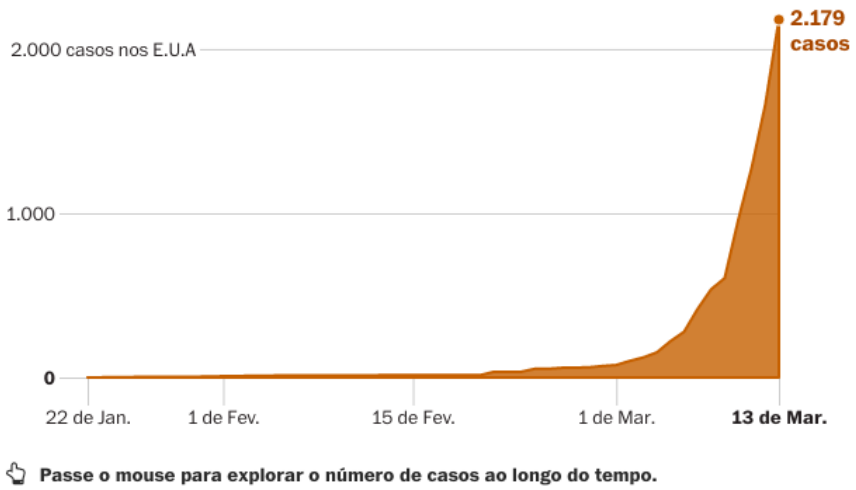

Figura 3.1: A forma do gráfico é equivalente a uma função exponencial.

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 3.2: O uso de ícone no meio do texto reforça a forma típica de uma função exponencial.

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, consequentemente, 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 no próprio 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 3.3). A segunda animação demonstra que uma pessoa doente se transforma, provavelmente, em uma pessoa recuperada (Figura 3.4).

Figura 3.3: A animação explica a dinâmica de contágio.

Figura 3.4: A animação explica a dinâmica de contágio e recuperação.

O texto da matéria, ao mesmo tempo que descreve a animação serve de legenda visual para entender as cores utilizadas (Figura 3.5).

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 3.5: O texto é codificado visualmente para funcionar como legenda das animações.

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.

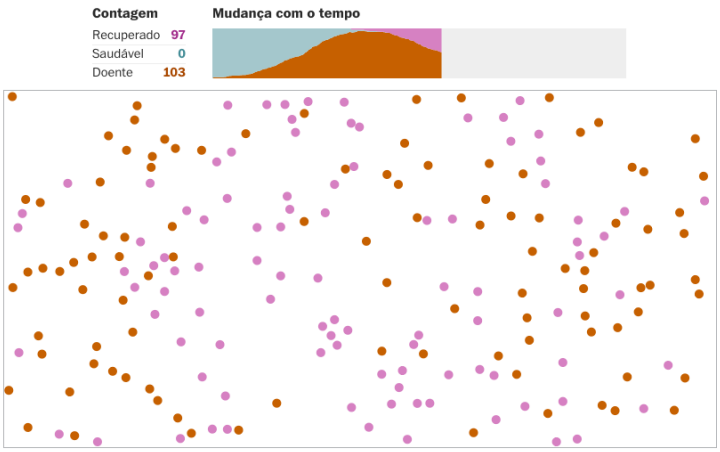

Figura 3.6: Simulação da propagação do *simulitis* no cenário a – circulação livre para todos.

A Figura 3.6 ilustra os três elementos para um determinado momento de contágio no cenário a, isto é, sem nenhuma ação de distanciamento social. A Figura 3.7 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

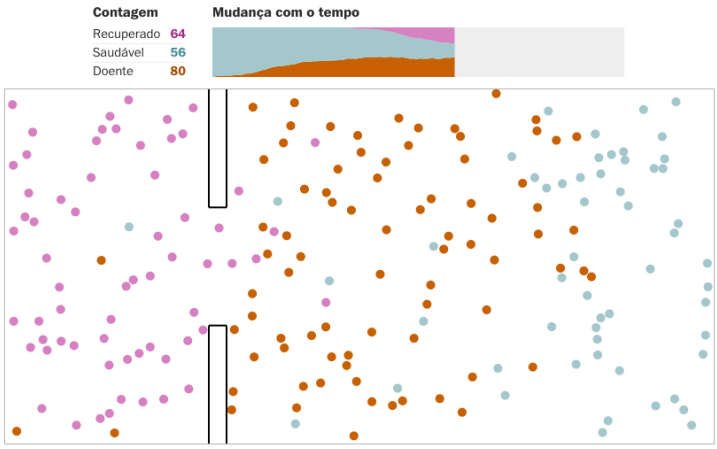

Figura 3.7: Simulação da propagação do *simulitis* no cenário b – quarentena forçada.

marco zero da Covid-19. A animação desse 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.

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 3.8 sintetiza a forma padrão dos gráficos gerados para cada simulação.

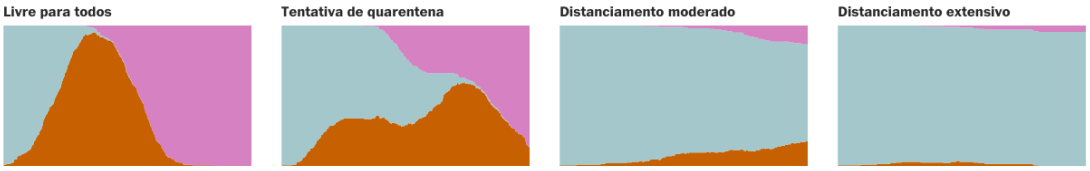

Figura 3.8: A forma padrão dos gráficos simulados para cada um dos cenários de propagação de uma epidemia.

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 e 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 autor da matéria 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 do Covid-19, é real. Com exceção do gráfico 3.1 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 do 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. Para falar sobre tal conceito, analisaremos, a seguir, o gráfico publicado pelo *The Economist* que foi rapidamente repercutido por diversos outros meios veículos, sendo inclusive adaptado para outros formatos.

3.2 Achatando a curva

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" [3]. Publicada em 29 de fevereiro, a matéria traz um gráfico (Figura 3.9) criado pelo jornalista visual Rosamund Pearce baseado em um gráfico previamente criado pelo órgão americano Centers for Disease Control and Prevention (CDC), em 2017 [21] (Figura 3.10) .

A Figura 3.9 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 o Covid-9. 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, existe tendências-chave esperadas para a dinâmica de contágio. O que a Figura 3.9 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

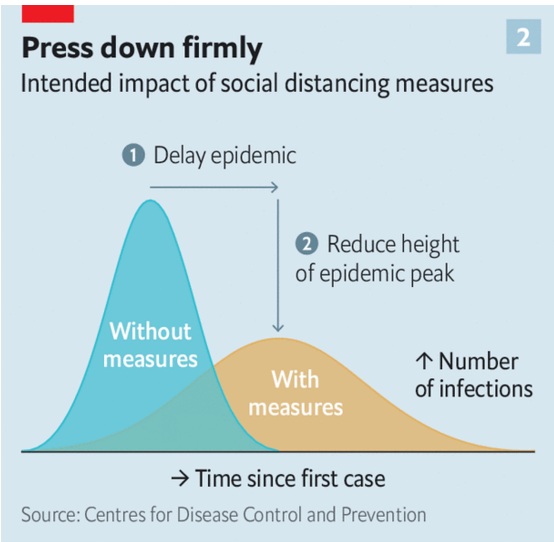

Figura 3.9: O primeiro gráfico do "achatamento da curva" a ganhar visibilidade na mídia. Fonte: *The Economist*.

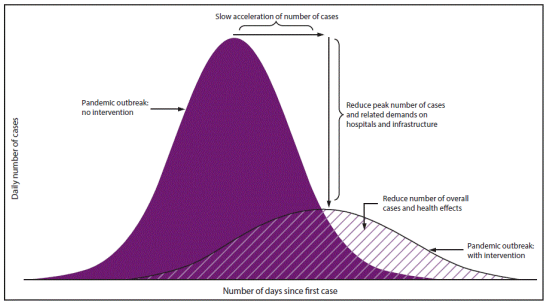

Figura 3.10: O gráfico original do "achatamento da curva" publicado em relatório de pesquisa. Fonte: CDC.

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 infeção dará origem. Em última análise, a taxa reprodutiva pode ser associada ao limiar epidemiológico que é comentado na seção 2.

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 in-

formação 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 3.11) [5], *The New York Times* (Figura 3.12) [22] e *Nexo* (Figura 3.13) [17] foram alguns dos veículos que adaptaram o gráfico 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 sabe, ao certo, o número de casos de 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.

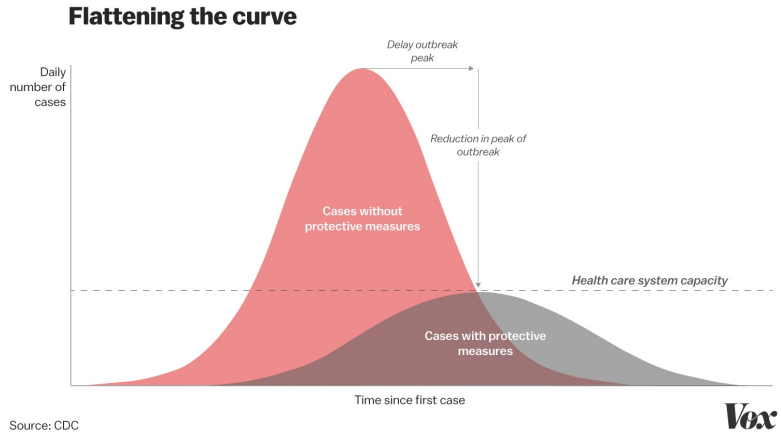

Figura 3.11: Gráfico de "achatamento da curva" pelo veículo *Vox*.

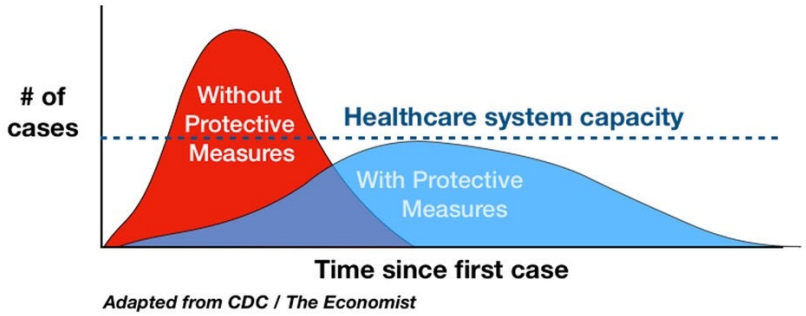

Figura 3.12: Gráfico de "achatamento da curva" pelo veículo *The New York Times*.

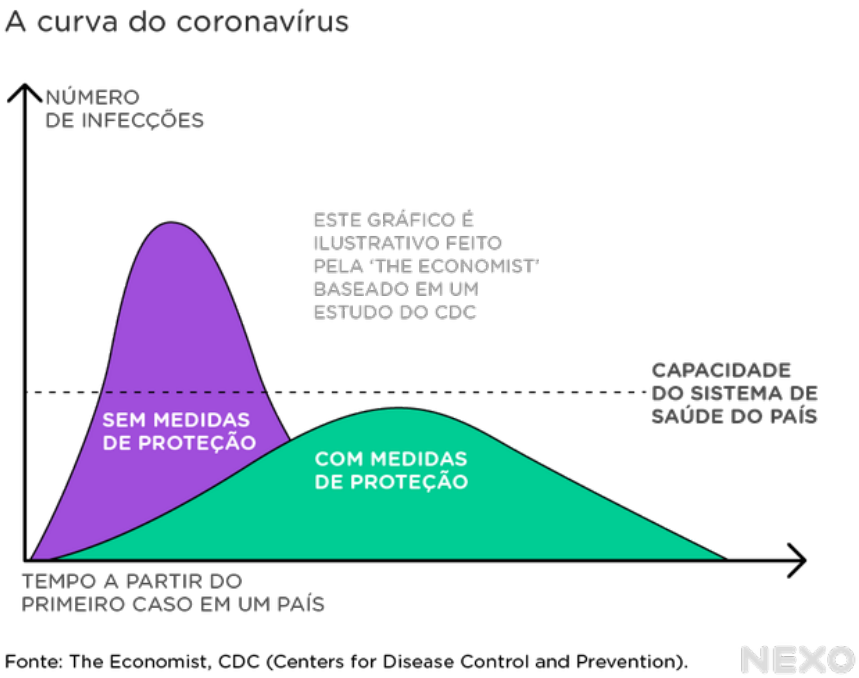

Figura 3.13: Gráfico de "achatamento da curva" pelo veículo *Nexo*.

Bibliografia

[1] Imperial College COVID-19 Response Team. Technical report. http://www.imperial.ac.uk/medicine/departments/school-public-health/infectious-disease-epidemiology/mrc-global-infectious-disease-analysis/covid-19/.

[2] Novel Coronavirus (COVID-19) Cases, provided by JHU CSSE. https://github.com/CSSEGISandData/COVID-19.

[3] The world gets ready - Covid-19 is now in 50 countries, and things will get worse | Briefing | The Economist. *The Economist*, February 2020. https://www.economist.com/briefing/2020/02/29/covid-19-is-now-in-50-countries-and-things-will-get-worse.

[4] Michiel Bakker, Alex Berke, Matt Groh, Alex Pentland, and Esteban Moro. Social Distancing in New York City. *MIT Media Lab*, March 2020. http://curveflattening.media.mit.edu/posts/social-distancing-new-york-city/.

[5] Eliza Barclay and Dylan Scott. How canceled events and self-quarantines save lives, in one chart. *Vox*, March 2020. https://www.vox.com/2020/3/10/21171481/coronavirus-us-cases-quarantine-cancellation.

[6] Mona Chalabi. COVID-19 and refugees, April 2020. https://www.instagram.com/p/B-ce_gSlgSJ/.

[7] Mona Chalabi. Racial disparities in COVID-19, April 2020. https://www.instagram.com/p/B-z2OVVlBKj/.

[8] Mona Chalabi. Who can work from home?, March 2020. https://www.instagram.com/p/B-PqSFBlFyT/.

[9] Lola Ferreira and Marília Ferrari. Como está a visitação nos presídios do Brasil em tempos de isolamento? *Gênero e Número*, April 2020. http://www.generonumero.media/visitas-presidios-isolamento/.

[10] Julia R. Giannella. Data selfie da quarentena, 2020. http://www.juliagiannella.com.br/dataselfiequarentena/.

[11] W. O. Kermack and A. G McKendrick. A contribution to the mathematical theory of epidemics. *Proc. Roy. Soc. Lond.*, A(115):700–721, 1927.

[12] Joos Korstanje. *Modeling Exponential Growth*, 2020. https://towardsdatascience.com/modeling-exponential-growth-49a2b6f22e1f.

[13] Joos Korstanje. *Modeling Logistic Growth*, 2020. https://towardsdatascience.com/modeling-exponential-growth-49a2b6f22e1f.

[14] Lauren Leatherby and David Gelles. How the Virus Transformed the Way Americans Spend Their Money. *The New York Times*, April 2020. https://www.nytimes.com/interactive/2020/04/11/business/economy/coronavirus-us-economy-spending.html.

[15] Amanda Makulec. Ten Considerations Before You Create Another Chart About COVID-19, March 2020. https://medium.com/nightingale/ten-considerations-before-you-create-another-chart-about-covid-19-27d3bd691be8.

[16] Manlio Massimetti and Angela Testa. Quarantine Portrait, 2020. https://quarantineportrait.art/.

[17] Matthew McQueen. Como desacelerar a pandemia de coronavírus. *Nexo Jornal*, March 2020. https://www.nexojornal.com.br/externo/2020/03/13/Como-desacelerar-a-pandemia-de-coronavirus.

[18] OKBR. Transparência Covid-19, 2020. https://transparenciacovid19.ok.org.br/.

[19] Zachary Petit. One Design Firm's Quarantine, in Data, May 2020. https://www.printmag.com/information-design/designing-quarantine-data/.

[20] Nadja Popovich. Watch the Footprint of Coronavirus Spread Across Countries - The New York Times. *The New York Times*, March 2020. https://www.nytimes.com/interactive/2020/climate/coronavirus-pollution.html.

[21] Noreen Qualls. Community Mitigation Guidelines to Prevent Pandemic Influenza — United States, 2017. *MMWR. Recommendations and Reports*, 66, 2017. https://www.facebook.com/CDCMMWR.

[22] Siobhan Roberts. Flattening the Coronavirus Curve. *The New York Times*, March 2020. https://www.nytimes.com/article/flatten-curve-coronavirus.html.

[23] David Smith and Lang Moore. *The SIR Model for Spread of Disease.* Mathematical Association of America, 2020. https://www.maa.org/press/periodicals/loci/joma/the-sir-model-for-spread-of-disease.

[24] Harry Stevens. Essas simulações mostram como achatar a curva de crescimento do coronavírus. *Washington Post*, March 2020. https://www.washingtonpost.com/graphics/2020/health/corona-simulation-portuguese/.

[25] Matthew Yglesias. US rocked by 6.6 million more initial unemployment claims last week. *Vox*, April 2020. https://www.vox.com/2020/4/9/21213449/unemployment-initial-claims-us-april-4.