

Covid Timeline Paper

1st Cleomar Felipe Rabelo Antoszczyszyn

*Escola de Matemática Aplicada
FGV - RJ*

Rio de Janeiro, Brasil
cleomarfelipe@gmail.com

2nd Gabriel Pastori

*Escola de Matemática Aplicada
FGV - RJ*

Rio de Janeiro, Brasil
gabrielpastorifigueira@gmail.com

3rd Gustavo Ramalho

*Escola de Matemática Aplicada
FGV - RJ*

Rio de Janeiro, Brasil
gustavoramalho384@gmail.com

Abstract—Este documento é um artigo sobre a criação de uma página de visualização de dados na web. A página traz informações sobre o COVID-19 e possui iterações que permitem ao usuário navegar de 2020 até 2023.

I. INTRODUÇÃO

Desde o início de 2020, uma pandemia varreu o mundo, trazendo incertezas, inseguranças e transformando a vida de bilhões de pessoas. Agora, em 2023, à medida que refletimos sobre o passado, testemunhamos uma jornada repleta de desafios e superações, embora ainda enfrentemos as consequências devastadoras do COVID-19. É dentro desse contexto que criamos uma página com dados relacionados a essa doença, com o objetivo de oferecer uma compreensão mais clara e acessível por meio da abstração de um grande volume de dados em visualizações interativas e fluidas.

Essa nova página criada teve como objetivo dar continuidade ao projeto da A3, onde foi possível ver a evolução dos casos de COVID pelo mundo. Nesta etapa, o grupo teve como objetivo principal ampliar a visualização e criar uma página dinâmica em formato de storytelling, de modo que cativasse o leitor e seja possível compreender mais profundamente o contexto da doença em diferentes épocas.

Os dados que utilizamos para termos informações sobre quantidade de óbitos e novos casos foram retirados da *OWID* e podem ser encontrados *aqui*. Para a extração de notícias, pegamos dados da *Uol*, *G1* e da *ONU*.

II. OBJETIVOS

A página "COVID Timeline" foi criada com o objetivo de fornecer uma visão histórica da pandemia de COVID-19, desde o seu início em 2020 até o presente ano de 2023. Ela busca lembrar ao leitor como foi esse evento por meio de notícias e outros dados relacionados. Nesse sentido, buscamos atingir nossos resultados por meio de 3 seções:

A. Covid Evolution

Esta seção utiliza números e estatísticas para ajudar a entender a magnitude da pandemia, identificar tendências e tomar decisões informadas. Ela permite que o usuário acesse dados brutos sobre a evolução da pandemia ao longo do tempo, incluindo informações sobre novos casos, mortes e a média móvel em todos os países do mundo.

B. News Timeline

Esta seção destaca a importância de acompanhar a progressão das notícias relacionadas ao COVID-19. Ela permite que o usuário veja notícias de cada dia da pandemia, oferecendo uma perspectiva jornalística de cada momento enfrentado.

C. WordCloud

Esta seção oferece uma rápida visualização das informações mais relevantes de um determinado período de tempo. Ela cria uma nuvem de palavras-chave ou termos mais mencionados, oferecendo insights valiosos sobre os principais tópicos, preocupações e debates que estavam em evidência naquele momento específico.

III. DESCRIÇÃO DO SISTEMA

A estrutura da página passa inicialmente por uma breve descrição da conjuntura que estamos trabalhando e posteriormente temos três visualizações, já comentadas anteriormente: *Covid Evolution*, *News Timeline* e *WordCloud*. Vamos passar por cada uma delas e comentar detalhes das criações:

A. Desenvolvimento - Covid Evolution

Nessa primeira visualização, basicamente temos o trabalho da A3 implementado em nosso website, agora com uma ideia que agrega no storytelling e com melhorias significativas. Pegamos os dados referenciados na introdução e, com o auxílio de uma box no canto superior esquerdo, conseguimos ver a evolução dos países que mais tiveram mortes, casos e média móvel por época. A intensidade da cor em cada país pelo mapa mostra como o país se comporta em relação ao dado analisado. Colocamos um botão para as datas serem passadas automaticamente, além de também ser possível ajustar o *step* de acordo com o que desejar. Por fim, quando clicamos em um país, conseguimos ver toda a série temporal da variável selecionada para aquele país. Vale ressaltar a implementação de melhorias na visualização do ranking numérico no canto superior esquerdo da página, que passa agora a ser representado por um gráfico de barras, além da capacidade de escolha de uma data específica através de um input calendário.

B. Desenvolvimento - News Evolution

Essa visualização foi com certeza a que mais demandou tempo do grupo. Um dos grandes desafios foi a extração de dados, visto que portais de notícias, em geral, procuram dificultar a raspagem automática de dados, por razões óbvias. Apesar disso, utilizando-se de duas bibliotecas de Python - BeautifulSoup e Selenium - e da biblioteca JQuery, do javascript, conseguimos extrair os dados de 3 websites - Uol, UN e G1 - e exibimos as notícias referentes a cada dia da pandemia. No canto superior da tela, colocamos uma box que permite a seleção da fonte da notícia, ou seja, pegar apenas da fonte que desejar. Além disso, há uma opção de "Ver mais notícias" em que abre um modal em que é exibido as outras notícias do dia para a fonte selecionada.

C. Desenvolvimento - WordCloud

Também com python, o WordCloud mostra as palavras que mais são vistas nessas notícias a cada mês. Para melhor fluidez da visualização, tiramos algumas palavras que costumavam aparecer e que não necessariamente tinha a ver com o contexto, mas são comuns para construções textuais (como conectivos). Outras palavras que desconsideramos são as que fazem referência ao nome da doença, como COVID-19 ou CORONAVIRUS. Optamos por fazer isso porque a adição delas não agrega no objetivo da WordCloud, que é apresentar as informações em destaque relacionadas ao COVID em cada época. Se considerássemos essas palavras, elas apareceriam em todos os meses, uma vez que todas as notícias da pandemia citam o nome dela.

IV. TRABALHOS RELACIONADOS E MOTIVAÇÕES

Ao final do trabalho A3, ainda era uma dúvida se iríamos partir para uma nova base de dados e buscar outros tipos de visualização ou seguiríamos na mesma linha que estávamos nesse trabalho. Entretanto, duas questões importantes nos incentivaram a optar pela 2ª opção:

A primeira questão foi que, ao consultar os exemplos de projetos finais, havia um website com dados sobre covid (que pode ser encontrada clicando *aqui*). Apesar da ideia ser diferente, percebemos que era possível criar um storytelling em um website dinâmico que cativasse o leitor. O segundo motivo foi uma crítica que tivemos no antigo trabalho: "O que nossa visualização, que faz referência a dados sobre o Covid pelo mundo, diferencia de uma visualização que seja possível encontrar facilmente no google?". Por esse motivo, achamos que seria um desafio interessante a ser seguido. Por questões de originalidade, resolvemos relacionar notícias de datas específicas com o período da doença que estávamos enfrentando. Agora o leitor tem uma perspectiva mais jornalística e compreende melhor a situação da época que desejar. Obviamente essa é apenas uma parte das ideias implementadas, porém foi o principal motivo para darmos continuidade a essa ideia.

V. RESULTADOS

A principal intenção do website é "contar uma história". Portanto, o nosso objetivo era conseguir relacionar as visualizações e compreender informações por trás delas. O grupo considera que o objetivo foi cumprido, visto que isso acontece. Por exemplo, quando estava analisando a 1ª visualização, notei que diversos países tinham um grande pico de aumento de média móvel dos casos no início de 2022. Como exemplo, o Brasil:

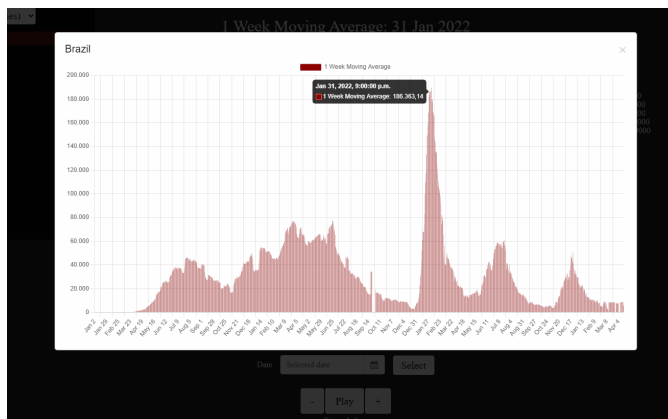


Fig. 1. Aumento - média movel

Com apenas o auxílio da visualização de notícias, busquei procurar qual era o motivo desse pico nessa época:



Fig. 2. Surgimento da Ômicron

Como pode ser visto, entendemos que o motivo era o surgimento da variante Ômicron e sua disseminação pelo mundo. Se buscássemos pelo WordCloud, também encontraríamos o nome da variante nos títulos de notícias:

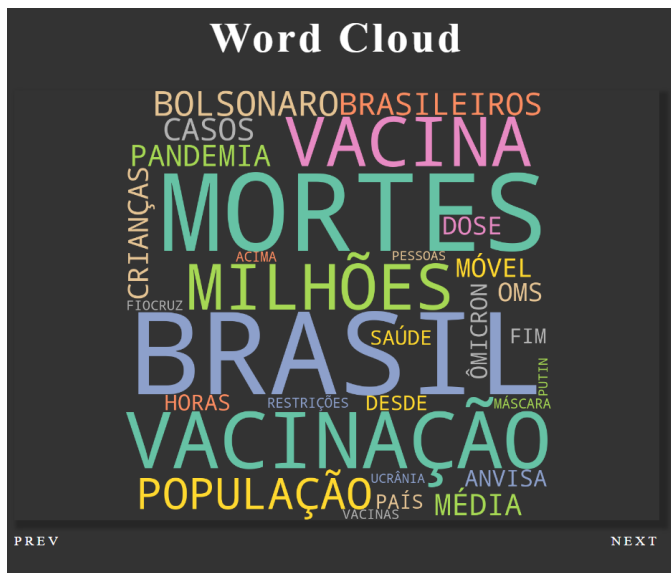


Fig. 3. Ômicron - WordCloud

Esse é apenas um exemplo da efetividade do nosso website. Para melhor entendimento e imersão no trabalho, é necessário ver diretamente no *GitHub*. O usuário tem a chance de obter diferentes insights e compreender os mais diversos contextos da doença pelo mundo.

VI. DISCUSSÃO

Agora, a página consegue permitir que o usuário, além de ter informações a dados estatísticos, possa ter uma visão mais clara da época em que está procurando. Por exemplo, quando queremos saber qual era o contexto e em que ponto necessariamente estávamos da pandemia. Conseguimos saber quando surgiram novas variantes, começo da vacinação, medidas políticas e outras informações. Tudo isso acessando uma visualização única, sem necessidade de pesquisas além ou qualquer outra coisa. Caso queira algo mais versátil e rápido, a nuvem de palavras nos fornece isso: conseguimos identificar quando assuntos estavam em alta, como a compra de vacinas pelo governador João Dória, a criação da coronavac e vacina da Pfizer, surgimento da ômicron... Por fim, o usuário consegue fazer essas relações entre as três visualizações, sendo mais simples de compreender os motivos por trás da variação dos dados brutos da 1ª visualização, como foi mostrado na seção de Resultados.

VII. CRÍTICAS DO MVP

Após a apresentação do MVP, algumas críticas foram levantadas como sugestão de melhoria para o projeto. Selecionamos algumas que achamos válidas, considerando o tempo que tínhamos para implementação e também a relevância da crítica. Os aperfeiçoamentos foram realizados até chegar na página final, entre eles estão:

A. Separação da Covid Evolution e News Timeline

Um ponto importante que foi questionado é a quantidade de informações que estávamos transmitindo. A tela ficava poluída e dificultava a visualização, portanto, acatamos a recomendação e separamos em duas visualizações.

B. Criação de uma WordCloud e design da página

Outra crítica que acatamos foi a melhora do design da página: agora temos um storytelling com maior fluidez e uma dinâmica interessante para observar as informações. Nesse mesmo comentário, foi sugerido a criação de uma nuvem de palavras, que, como visto nesse paper, também foi implementado.

C. Sidebar de notícias

Agora, é possível ver todas as notícias do dia de cada site com a ajuda de uma Sidebar na visualização *News Timeline*. Não nos prendemos necessariamente a uma notícia que pode não ter a informação que o usuário procura.

D. Seleção de data

Uma funcionalidade simples que não tínhamos era a seleção de datas específicas, pois anteriormente era necessário ir pelo scroll até chegar na data que queríamos. Agora, há um campo nas duas primeiras visualizações que permitem isso.

E. Evolução em gráficos

Por fim, substituímos o ranking com os dados na *Covid Evolution*, que antes estavam apenas com números, por um gráfico de barras que torna a visualização mais efetiva.

VIII. TRABALHOS FUTUROS

Obviamente, nem todas as críticas foram implementadas. Selecionamos as que mais agregariam em nossa página, além de também considerarmos as sugestões que mais pareciam frequentes entre as enviadas. Entretanto, críticas válidas que não foram implementadas nesse momento poderão ser válidas para trabalhos futuros que possam melhorar ainda mais nossas visualizações. Uma melhora que queríamos muito implementar é a seleção de notícias por país, onde possivelmente seria um grande diferencial para a página. Quando colocássemos o mouse por cima de algum país, veríamos as novidades dele no dia, permitindo uma maior interação do mapa com as notícias. Entretanto, é uma implementação que leva muito mais tempo para ser colocada do que o prazo atual permitia, portanto, ainda não há no website. Além disso, gostaríamos de facilitar ainda mais a facilidade que o leitor veja correlações entre as 3 visualizações, sem precisar necessariamente ir em uma e voltar na outra.

REFERENCES

- [1] COVID-19 Measures Final Project. Disponível em: <https://6859-sp21.github.io/final-project-covid-19-measures/>
- [2] OWID COVID-19 Data. Disponível em: <https://github.com/owid/covid-19-data/tree/master/public/data>
- [3] UOL Notícias. Disponível em: <https://noticias.uol.com.br/>
- [4] G1 - O portal de notícias da Globo. Disponível em: <https://g1.globo.com>
- [5] United Nations News. Disponível em: <https://news.un.org/>