



UNIVERSIDADE FEDERAL DE MINAS GERAIS
Departamento de Ciência da Computação

TRABALHO PRÁTICO II - INTRODUÇÃO A BANCO DE DADOS

Efeitos Epidemiológicos da Cobertura Vacinal no Brasil: Extração e Análise de
Dados do DataSUS

EDUARDO DE SOUSA
GUILHERME BUENO
GUSTAVO CABRAL
MATHEUS SOARES
RENATO LUCAS

JUNHO DE 2025

ÍNDICE:

1 - Escolha e processo de obtenção dos dados

1.1 - Sobre o tema e contextualização

1.2 - Sobre os dados

2 - Análise exploratória

3 - Estruturação do Banco: Antes e Depois

4 - Análise crítica das fontes de dados

5 - Estudo acerca da correlação entre os dados e seus efeitos práticos

5.1.1 - Visualização dos dados

5.1.2 - Correlações

5.2 - Conclusão

6 - Bibliografia

Link para o repositório do GitHub contendo o código para execução do projeto:

<https://github.com/dudu-soliveira/ibd-tp2>

1 - Escolha e processo de obtenção dos dados

1.1 - Sobre o tema e contextualização

O tema escolhido para o trabalho prático foi a saúde, e mais especificamente o estudo de uma possível correlação entre a cobertura vacinal no Brasil durante a pandemia de covid-19, e taxa de mortalidade da doença no mesmo período. A intenção principal é realizar uma análise crítica acerca dos efeitos práticos de uma boa cobertura vacinal para com a diminuição dos óbitos e transmissão do coronavírus em território brasileiro.

A pandemia causada pelo vírus da covid-19 se iniciou no final do ano de 2019 e chegou ao Brasil no início do ano de 2020. Ocasinou o fechamento de diversos estabelecimentos, e a necessidade de ficar em casa por grande período de tempo para tentar conter a propagação da doença, medida chamada de lockdown. Em relação ao tema do trabalho, a campanha vacinal no Brasil se inicia em 2021, mais precisamente no dia 17 de janeiro, e se estende até os dias atuais. Dessa forma, optamos por escolher os anos de 2021 e 2022 como focos principais das nossas análises, pois além de serem o pico da pandemia no país, também o início da cobertura vacinal e a diminuição gradual da incidência e óbitos no território brasileiro.

1.2 - Sobre as Fontes dos Dados

A busca pela base de dados se baseou na LAI (Lei de Acesso à informação) e buscou sites ligados ao governo que tivessem dados sobre a saúde. Os sites utilizados foram o DataSUS e o do Ministério da Saúde. Seguem os links:

- https://infoms.saude.gov.br/extensions/SEIDIGI_DEMAS_Vacina_C19/SEIDIGI_DEMAS_Vacina_C19.html# - Dados sobre as vacinas.

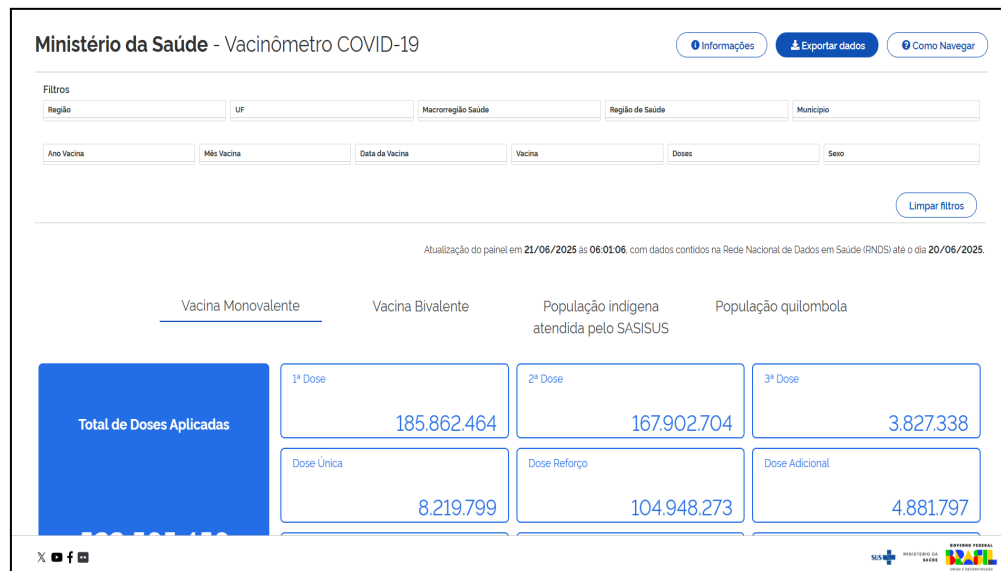


Imagem 1: Painel de vacinação do Ministério da Saúde

- https://infoms.saude.gov.br/extensions/covid-19_html/covid-19_html.html#
Dados sobre os óbitos.



Imagem 2: Painel de estatísticas sobre a COVID-19 do Ministério da Saúde.

- <https://portalsinan.saude.gov.br/calendario-epidemiologico>
Dados sobre as semanas epidemiológicas.
- <https://www.ibge.gov.br/explica/codigos-dos-municipios.php>
Dados sobre os municípios.

2 - Análise Exploratória

Para esta análise, com fins de melhor compreensão do estudo realizado, entende-se, para todos os efeitos, pós-vacina como sendo o período abrangido pelos anos de 2021 a 2025. Assim sendo, em torno desta faixa de tempo se baseará toda a análise efetivada a seguir.

Para tal, dispusemo-nos do cálculo de alguns indicadores não nativos das tabelas do banco de dados exportado, tais como: taxa de mortalidade (por 100 mil habitantes), taxa de incidência (por 100 mil habitantes) e doses aplicadas per capita. Inicialmente, iremos nos ater ao estudo observacional dos indicadores citados em nível municipal.

2.1. - Ranqueamento das taxas de mortalidade, incidência e vacinação per capita

Ao calcularmos a taxa de mortalidade por COVID-19 dos municípios brasileiros no período pós-vacinal (2021-2025), observamos valores significativamente discrepantes: demasiadamente elevados - fenômeno mais visivelmente comum em municípios com população pequena, o que desfavorece uma comparação direta entre as proporções dos municípios -, ou mesmo negativos - anomalia possivelmente gerada por erros de digitação ou de codificação automática quando da migração de dados entre bancos ou sistemas diferentes (por exemplo, de banco hospitalar para o e-SUS ou o SIVEP-Gripe). Dessa forma, a alternativa escolhida foi a visualização ranqueada por agrupamento de municípios em regiões agregadas, seguindo a lógica hierárquica estipulada pelo IBGE. Assim, plotamos as regiões intermediárias - categorização de território hierarquicamente abaixo apenas de unidade federativa, e que engloba regiões imediatas, as quais, por sua vez, englobam municípios sob uma mesma área de influência comum (por questões de saúde, trabalho, compras ou educação) - com o objetivo de amenizar o impacto de possíveis outliers. As figuras 3, 4 e 5, a seguir, apresentam, respectivamente, no âmbito de regiões intermediárias, o ranqueamento referente à aplicação de doses per capita, taxa de mortalidade e incidência da COVID-19.

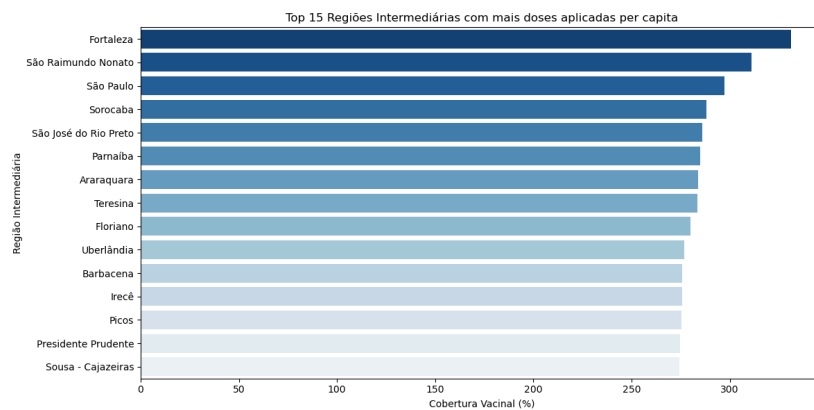


Figura 3: Regiões intermediárias com maior aplicação de doses vacinais per capita.

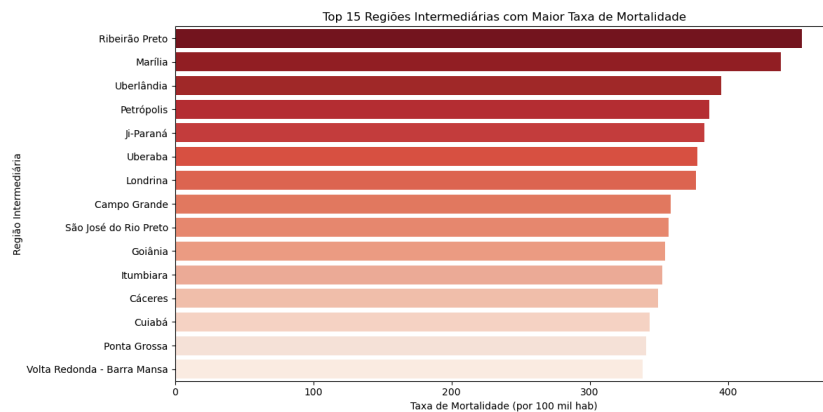


Figura 4: Regiões intermediárias de maior taxa de mortalidade no período pós-vacinal.

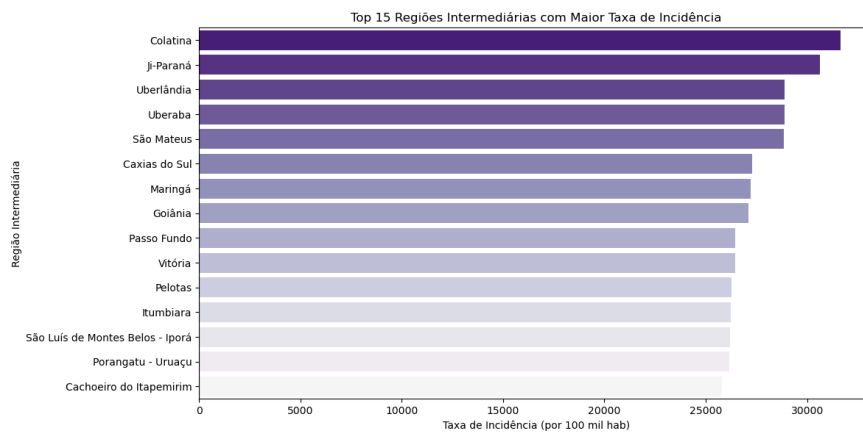


Figura 5: Regiões intermediárias de maior incidência da doença no período pós-vacina.

2.2 - Visualização comparativa das séries históricas de casos, óbitos e aplicação de vacinas

Para ampla visualização dos agregados nacionais ao longo do período estudado, plotamos gráficos comparativos entre casos e óbitos notificados desde o início dos esforços governamentais de imunização contra a COVID-19 no país. A figura 6 exhibe testes positivos e mortes notificadas após integração dos dados das secretarias municipais de saúde, dispostos por semana epidemiológica.

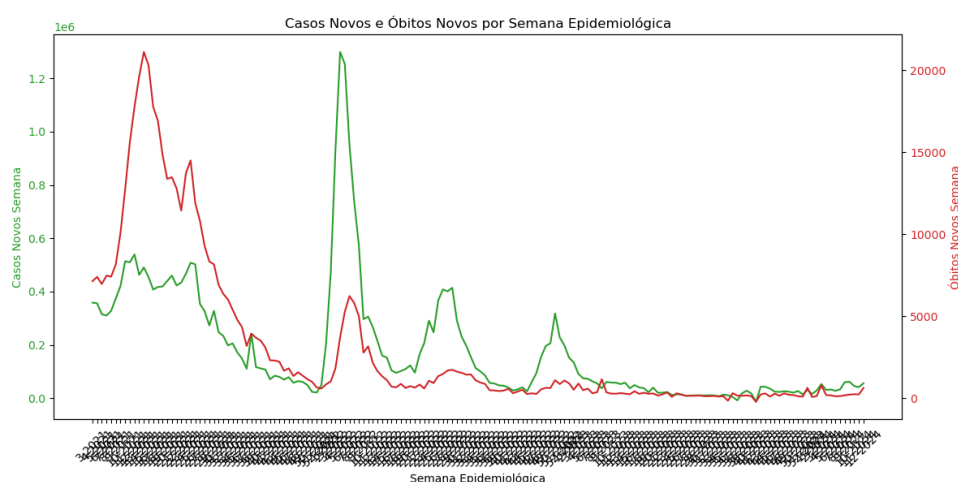


Figura 6: Casos e óbitos por COVID-19 no Brasil no período pós-vacinal.

Visualmente, é nítida a constatação de um descolamento entre as curvas de infecções e óbitos reportados no final do primeiro trimestre de 2021. A hipótese mais factível que justifica esse fenômeno é uma conjunção de fatores: colapso hospitalar em diversas regiões do país^[1], o que levou à indisponibilidade de insumos médicos para tratar pacientes com formas moderadas da doença, contribuindo para o seu agravamento, subnotificação das infecções reais decorrente da ausência de um programa nacional de testagem em massa e de rastreamento de contatos coordenado pelo Governo Federal em parceria com estados e municípios; resultando em poucos testes per capita e indo em contramão do recomendado pela maioria dos cientistas até então, e a introdução de uma nova variante no cenário genômico nacional - a P.1, posteriormente denominada Gama pela Organização Mundial de Saúde (OMS) -, cuja patogenicidade se mostrou significativamente maior em faixas etárias mais jovens^[2].

Paulatinamente, conforme a vacinação progrediu no Brasil, a sensibilidade da curva de óbitos em relação à de casos foi se tornando menor. A esse respeito, é notável uma segunda onda de contaminações pela variante P.1 na virada entre os semestres de 2021, que não é acompanhada proporcionalmente pela curva de mortes, na qual se verificou tão somente um leve repique. A partir de então, até a chegada da variante Ômicron (B.1.1.529) no país, há o primeiro declínio consolidado dos indicadores da doença desde o início da crise sanitária em território brasileiro.

Conforme supracitado, essa tendência é interrompida no início de 2022, com a ascensão das notificações da doença por intermédio da Ômicron e, posteriormente, de suas subvariantes, mantendo ondas de infecção - embora cada vez mais diminutas, em razão da redução dos esforços de testagem em decorrência do encerramento da ESPIN (Emergência de Saúde Pública de Importância Nacional) em maio de 2022^[3]. Apesar disso, os óbitos, parâmetro mais confiável dada sua dificuldade intrínseca de subnotificação, passam a não acompanhar a curva de casos e mantêm uma tendência de declínio a longo prazo, com máximas em surtos continuamente menores que os anteriores.

A seguir, comparamos o agregado nacional de novos casos reportados pelas secretarias municipais de Saúde ao total de imunizantes aplicados por semana epidemiológica.

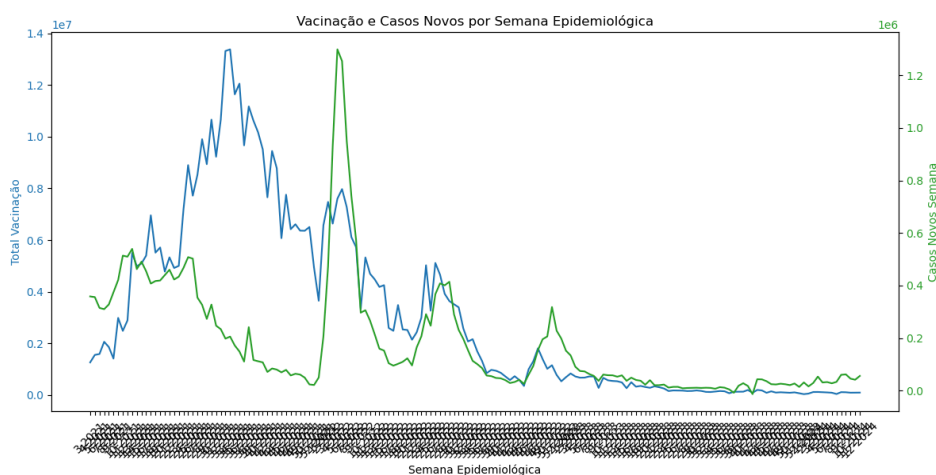


Figura 7: Testes positivos para COVID-19 e total de vacinas aplicadas por semana epidemiológica.

Note que as aplicações semanais das doses dos imunizantes anti-COVID, a despeito de terem se iniciado na 2ª quinzena de janeiro de 2021, somente ultrapassam a

marca simbólica de 1 milhão por dia (7 milhões de doses aplicadas por semana epidemiológica, i. e., $0.7 \cdot 10^7$ no eixo y à esquerda no gráfico) durante a segunda onda da variante P.1, em meados daquele ano. A partir de então, a curva de aplicações rumo a um pico no final do 3º trimestre, concomitantemente a uma queda pronunciável nos novos contágios reportados. Já no ano de 2022, é particularmente notável um acompanhamento da curva de vacinação em relação à de infecções confirmadas, principalmente durante as ondas causadas pelas subvariantes BA.5 - junho a agosto - e BQ.1 - outubro a dezembro. Isto tem a ver com uma maior adesão à imunização durante os períodos de maior transmissão do vírus, somado ao cronograma do Ministério da Saúde que objetivava garantir o esquema vacinal completo a grupos resistentes à vacinação antes e durante picos causados por novas sublinhagens do SARS-CoV-2.

Nos anos seguintes, há uma redução significativa no número de vacinações, mas tal análise exige cautela. Ainda que o estímulo à vacinação tenha se enfraquecido conforme a doença foi se tornando branda à maioria das pessoas, há uma limitação nos horizontes de análise, uma vez que os dados extraídos contemplam apenas a aplicação de doses monovalentes - que tem, como alvo, a proteína de apenas uma versão de vírus, distinta da bivalente, que possui, como alvo, duas variações do novo coronavírus -, as quais, inicialmente, se tornaram obsoletas em 2023, com a introdução da vacina bivalente BA.5 - com alvo na proteína Spike da subvariante Ômicron BA.5 e da cepa ancestral Wuhan-hu-1. Apesar disso, em 2025, ocorre a reintrodução do uso da monovalente, atualizada para combater a subvariante Ômicron JN.1.

Por fim, a última, porém, mais importante, análise visual a ser contemplada, se refere à relação entre óbitos e vacinação per capita, conforme disponível na figura 8.

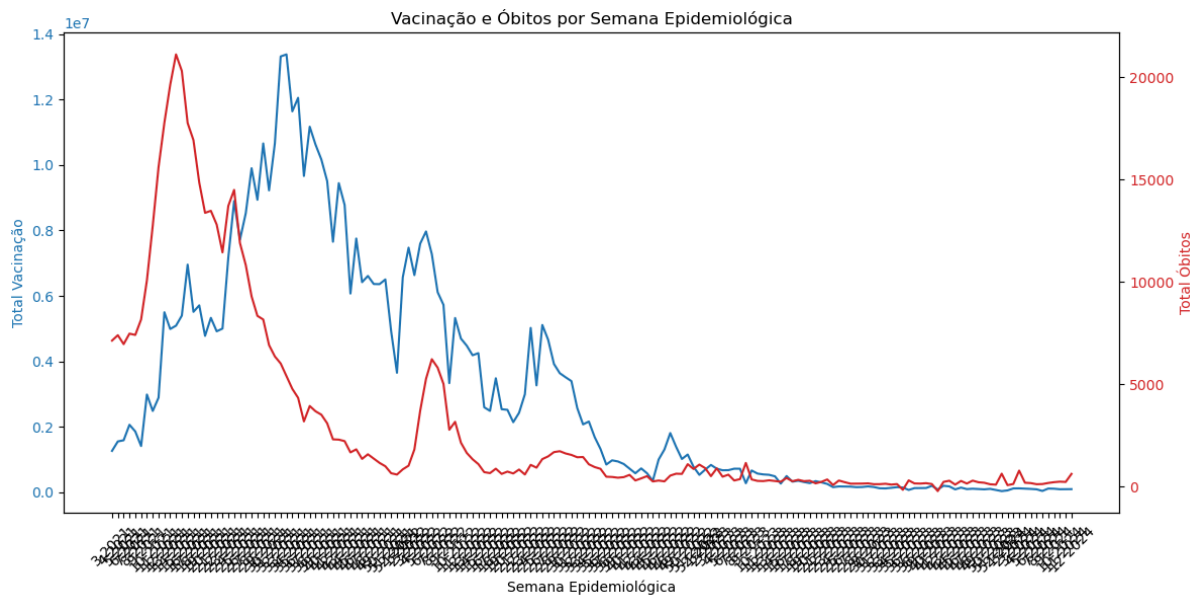


Figura 8: Agregado nacional da aplicação de doses per capita e de novos óbitos por semana epidemiológica.

Observe o quão mais visível é o impacto de médio e longo prazo das vacinas na redução dos óbitos em comparação à curva de casos. Essa realidade atesta acerca da maior eficácia das vacinas em prevenir a evolução dos contágios para suas formas moderada e grave, ao passo que não logra reduzir de forma significativa a transmissão do vírus^[4], principalmente num contexto de subvariantes com alto escape imune, como é o caso das descendentes da Ômicron. Tal análise, por se ater a um caráter puramente visual, não pode ser entendida como conclusiva neste sentido, mas ajuda a fortalecer a ideia supracitada.

2.3 - Análises de correlação

Para analisar como os indicadores de interesse do estudo (mortalidade, incidência e vacinação per capita) se comportaram entre os municípios, aplicou-se dispersão aos dados, de dois em dois, com foco na relação com a cobertura vacinal considerando doses totais. Uma observação importante para a análise é que, dada a elevada cifra de municípios, optou-se, novamente, por agrupar as estatísticas por regiões intermediárias, de modo a facilitar as comparações e que a verificação de uma correlação seja mais tangível.

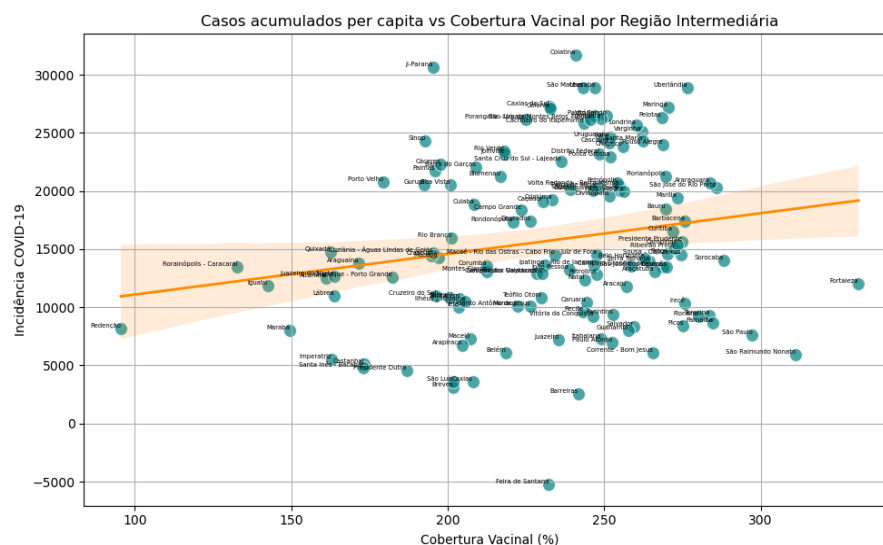


Figura 9: Regressão linear baseada na incidência da COVID-19 e na cobertura vacinal.

Conforme esperávamos, a correlação entre a incidência da doença em termos de infecções e a cobertura vacinal - contabilizando-se doses totais aplicadas, i.e., esquema inicial e doses de reforço - foi baixa, sendo praticamente desprezível com $R^2=0.0343$.

A surpresa está na correlação traçada entre mortalidade e cobertura vacinal, a qual se esperava ser forte e negativa: se mostrou positiva e mais forte do que a de incidência, mas ainda fraca, com $R^2=0.1113$. Esse resultado, no entanto, deve ser entendido com cautela: é sugerível que municípios com predominância de faixas etárias de maior risco de agravamento para a COVID-19 também possam maior adesão vacinal, enquanto regiões populadas por grupos de menor risco tendem a se eximir da imunização, por menor perspectiva de risco.

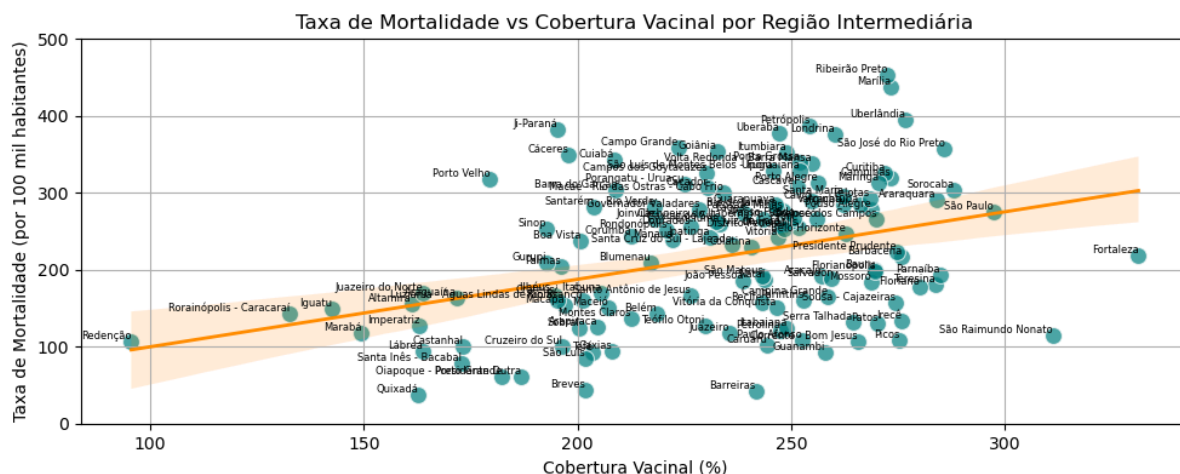


Figura 10: Regressão linear baseada na mortalidade por COVID-19 e na cobertura vacinal.

Abaixo, agrupamos os dados por estado e recalculamos as taxas de mortalidade e de incidência do SARS-CoV-2. O resultado obtido pode ser visualizado na figura 11:

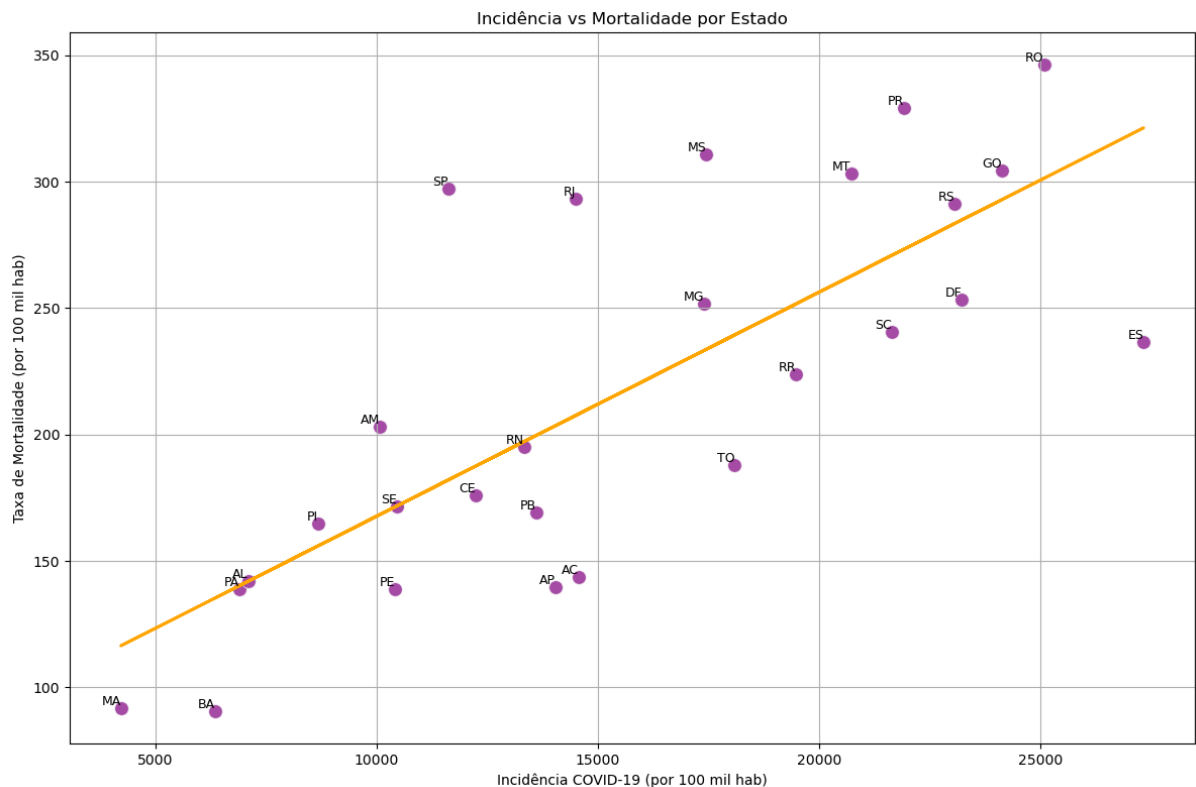


Figura 11: Correlação entre incidência e mortalidade estaduais pela COVID-19.

Com um $R^2=0.5828$, indicando correlação moderada a forte, o gráfico evidencia um aspecto bastante intuitivo: uma maior incidência da doença propicia uma maior taxa de mortalidade. É curioso, no entanto, observar que os estados de São Paulo e do Rio de Janeiro, dois entre os três estados mais populosos do Brasil, se destacam com uma elevada taxa de mortalidade, a despeito de uma incidência relativamente menor se comparada a de outras unidades federativas. Essa aparente maior letalidade da doença nesses estados aparenta não ter relação com versões específicas do vírus que circularam localmente, dado que, em nível nacional, as linhagens virais circulantes em território brasileiro tiveram uma abrangência aproximadamente homogênea^[5], embora não necessariamente em termos temporais. Assim sendo, a causa mais provável está relacionada a uma menor testagem per capita, i.e., proporcional à população, embora ainda elevada em números absolutos. Essa hipótese parece fazer sentido tendo em conta que a maior parte dos exames para detecção do novo

coronavírus foram obtidos por compra intermediada pela pasta nacional da Saúde, e não diretamente pelas secretarias estaduais. Desse modo, caberia ao Ministério da Saúde definir a distribuição dos testes ao sistema hospitalar capilarizado do Brasil por uma ordem de prioridade interna. Como o custo de aquisição desse aparato em tempos de alta demanda tende a ser naturalmente elevada, há uma tendência de que estados populosos recebam uma carga elevada, mas não rigidamente proporcional à sua população.

Por outro lado, o estado do Espírito Santo se destaca por uma baixa mortalidade, a despeito de uma elevada incidência da doença. Novamente, isso não está necessariamente relacionado a um menor impacto direto do vírus na população local por fatores genéticos, mas representa um caso de sucesso de uma unidade federativa que se beneficiou de uma taxa elevada de testagem - em relação a seus pares - e da implementação de uma política de rastreamento de contatos, com posterior estímulo à vacinação em massa. O ES é um dos poucos estados brasileiros que, ainda em 2025, reportam dados diários referentes à COVID-19^[6], o que evidencia o legado de maior compromisso do governo local no combate à doença, refletido nos principais indicadores.

2.4 - Mapas de incidência e mortalidade por COVID-19

Com o intuito de se visualizar o impacto em termos de incidência e mortalidade da COVID-19 a nível municipal, fez-se uso do recurso shapefile, disponibilizado pelo IBGE, para uma junção via código dos municípios, possibilitando associar cada município à sua forma geometricamente delimitada por fronteiras no mundo real. Um detalhe importante nos mapas das figuras 12 e 13 a seguir é que os municípios classificados como outliers tanto para incidência quanto para mortalidade foram desconsiderados, resultando em algumas poucas regiões em branco.

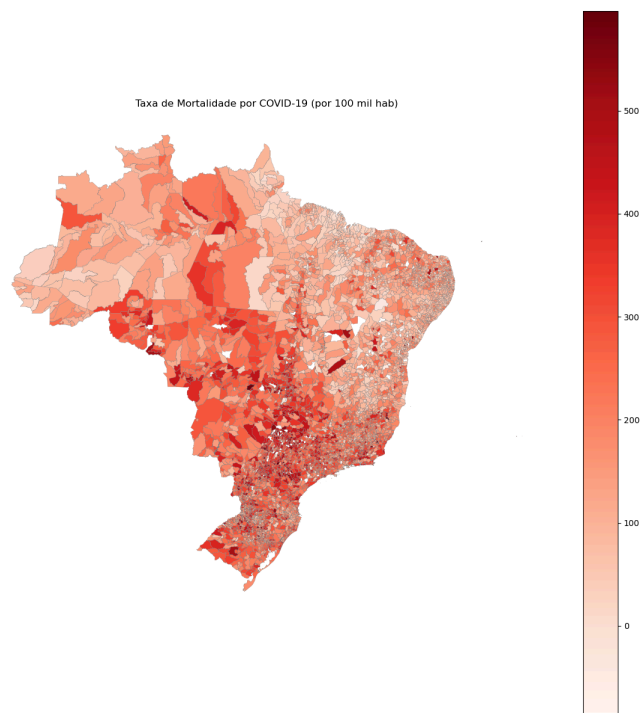


Figura 12: Taxa de mortalidade municipal para COVID-19 no período pós-vacinal.

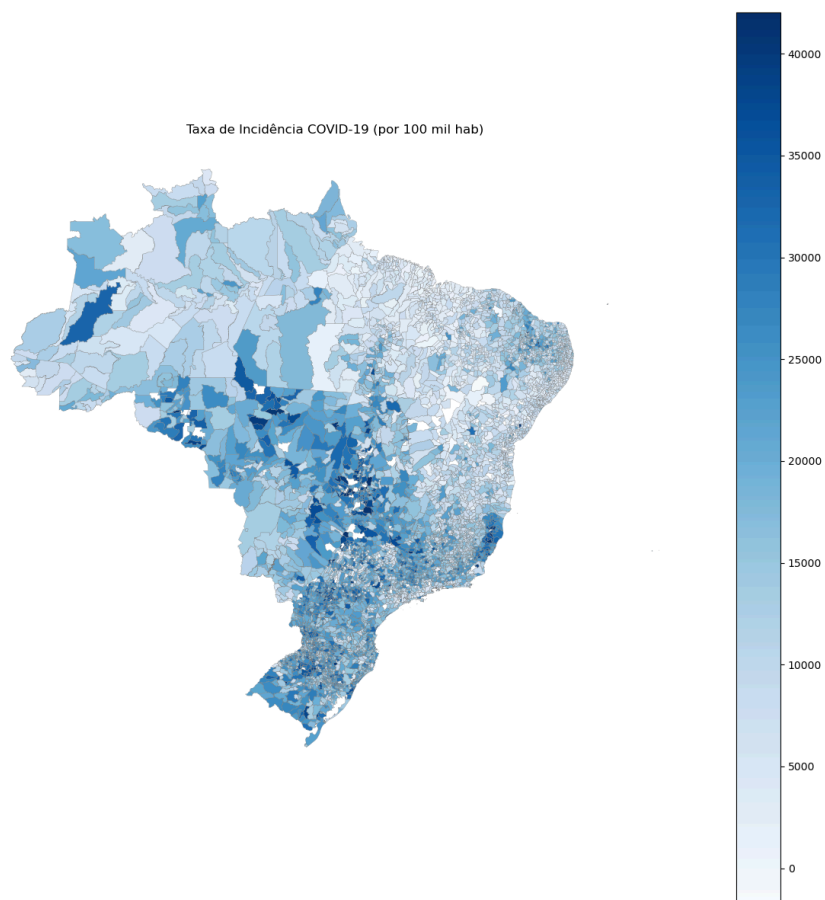


Figura 13: Taxa de incidência municipal da COVID-19 no período pós-vacinal.

Em ambas as figuras, é nítido observar uma aparente maior impacto do novo coronavírus nos municípios da região Centro-oeste.

3 - Estruturação do Banco: Antes e Depois

Os dados foram coletados de sites diferentes e não tinham, inicialmente, uma relação direta entre eles. Isto é, algumas tabelas não tinham chaves primárias e não tinham chaves que fizessem a conexão entre tabelas diferentes. Por exemplo, a tabela de “Vacinas” estava organizada em município e data da aplicação, logo, não estava diretamente relacionada com nenhuma outra tabela, tal que as outras estavam organizadas por semanas epidemiológicas. Logo, foi necessário um tratamento dos dados para condicioná-los às suas relações.

Conforme indicados nos diagrama e tabela 1, segue-se o esquema conceitual do banco de dados e o dicionário de dados correspondente referentes aos dados **não tratados**, como na coleta.

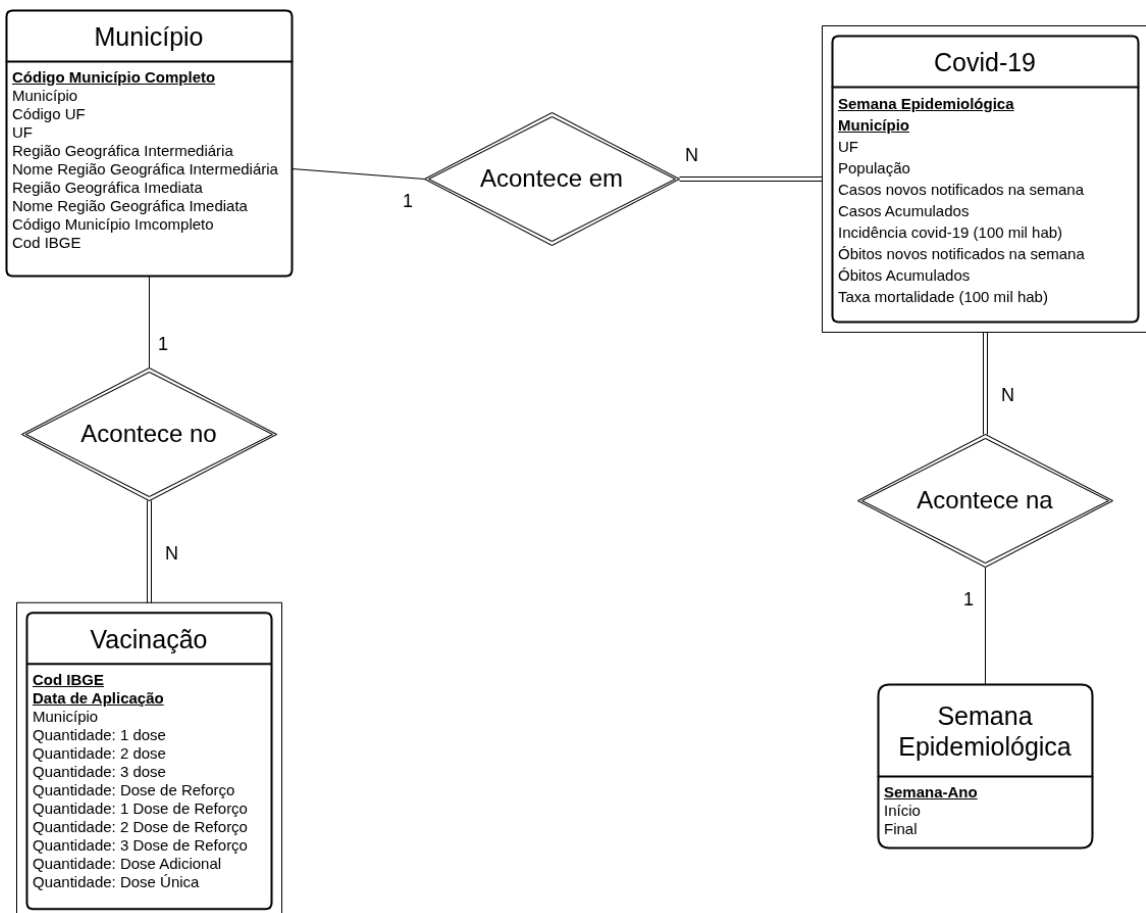


Diagrama 1: Esquema conceitual do banco em estágio inicial (desnormalizado).

Relação	Atributo	Tipo/Largura	Unico	Null	Valores Permitidos	Restrições e Comportamento se Chave Estrangeira
Município	Código Município Completo	INT	S	N		PK
Município	Município	VARCHAR(100)	N	N		
Município	Código UF	VARCHAR(2)	N	N		
Município	UF	VARCHAR(20)	N	N		
Município	Região Geográfica Intermediária	INT	N	N		
Município	Nome Região Geográfica Intermediária	VARCHAR(100)	N	N		
Município	Região Geográfica Imediata	INT	N	N		
Município	Nome Região Geográfica Imediata	VARCHAR(100)	N	N		
Município	Código Município Incompleto	INT	N	N		
Município	Cod IBGE	INT	N	N		
Vacinação	Cod IBGE	INT	N	N		PK, FK references Município "Código Município Completo"
Vacinação	Data de Aplicação	DATE	N	N		PK, FK references "Semana Epidemiológica" Semana-Ano
Vacinação	Município	VARCHAR(100)	N	N		
Vacinação	Quantidade: 1 dose	INT	N	N		
Vacinação	Quantidade: 2 dose	INT	N	N		
Vacinação	Quantidade: 3 dose	INT	N	N		
Vacinação	Quantidade: Dose de Reforço	INT	N	N		
Vacinação	Quantidade: 1 Dose de Reforço	INT	N	N		
Vacinação	Quantidade: 2 Dose de Reforço	INT	N	N		
Vacinação	Quantidade: 3 Dose de Reforço	INT	N	N		
Vacinação	Quantidade: Dose Adicional	INT	N	N		
Vacinação	Quantidade: Dose Única	INT	N	N		
Covid-19	Semana Epidemiológica	DATE	N	N		PK, FK references "Semana Epidemiológica" Semana-Ano
Covid-19	Município	VARCHAR(100)	N	N		PK, FK references Municípios "Código Município Completo"
Covid-19	UF	VARCHAR(20)	N	N		
Covid-19	População	INT	N	N		
Covid-19	Casos novos notificados na semana	INT	N	N		
Covid-19	Casos Acumulados	INT	N	N		
Covid-19	Incidência covid-19 (100 mil hab)	FLOAT	N	S		
Covid-19	Óbitos novos notificados na semana	INT	N	N		
Covid-19	Óbitos Acumulados	INT	N	N		
Covid-19	Taxa mortalidade (100 mil hab)	FLOAT	N	S		
Semana Epidemiológica	Semana-Ano	DATE	S	N		PK
Semana Epidemiológica	Início	DATE	S	N		
Semana Epidemiológica	Final	DATE	S	N		

Tabela 1: Dicionário de dados referente às tabelas desnormalizadas.

Em segundo lugar, o esquema conceitual e o dicionário de dados referentes aos dados pós tratamento.

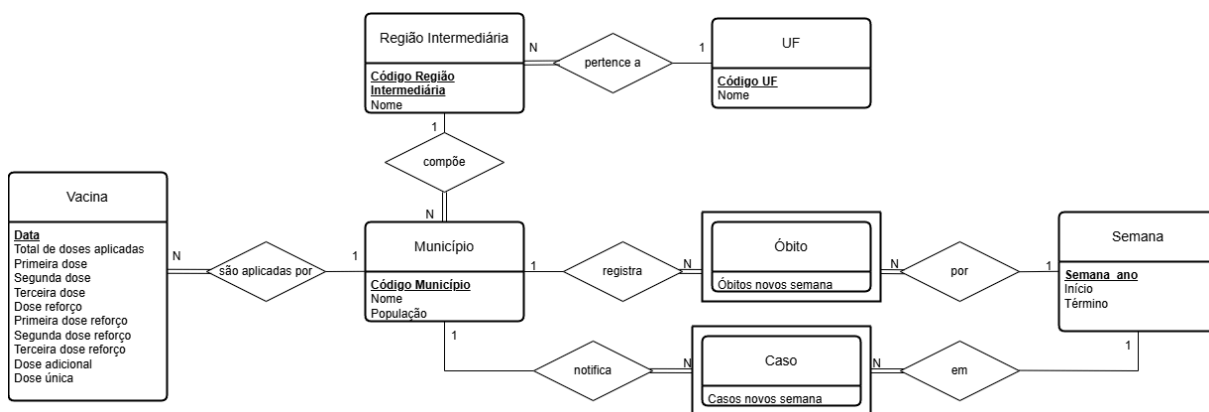


Diagrama 2: Esquema conceitual do banco normalizado.

Relação	Atributo	Tipo/Largura	Nulo?	Único?	Valores permitidos	Restrições adicionais e comportamento se chave estrangeira
UF	Código UF	CHAR(2)	N	S		PK
UF	Nome	CHAR(2)	N	N		
Região Intermediária	Código Região Intermediária	CHAR(4)	N	S		PK
Região Intermediária	Nome	VARCHAR(255)	N	N		
Município	Código Município	CHAR(6)	N	S		PK
Município	Nome	VARCHAR(255)	N	N		
Município	Código UF	CHAR(2)	N	N		FK references UF(Código UF)
Município	Código Região Intermediária	CHAR(4)	N	N		FK references Região Intermediária(Código Região Intermediária)
Município	Código Região Imediata	CHAR(6)	N	N		FK references Município(Código Município)
Município	População	INT	N	N		
Óbito	Código IBGE	CHAR(6)	N	N		FK references Município(Código Município)
Óbito	Semana epidemiológica	VARCHAR(7)	N	N		FK references Semana(Semana_ano)
Óbito	Óbitos novos semanas	INT	N	N		
Caso	Código IBGE	CHAR(6)	N	N		FK references Município(Código Município)
Caso	Semana epidemiológica	VARCHAR(7)	N	N		FK references Semana(Semana_ano)
Caso	Casos novos semanas	INT	N	N		
Vacina	Código IBGE	CHAR(6)	N	N		PK(1); FK references Município(Código Município)
Vacina	Data	DATE	N	N		PK(2);
Vacina	Total de doses aplicadas	INT	N	N		
Vacina	Primeira dose	INT	N	N		
Vacina	Segunda dose	INT	N	N		
Vacina	Terceira dose	INT	N	N		
Vacina	Dose reforço	INT	N	N		
Vacina	Primeira dose reforço	INT	N	N		
Vacina	Segunda dose reforço	INT	N	N		
Vacina	Terceira dose reforço	INT	N	N		
Vacina	Dose adicional	INT	N	N		
Vacina	Dose única	INT	N	N		
Semana	Semana_ano	VARCHAR(7)	N	S		PK
Semana	Início	DATE	N	S		
Semana	Término	DATE	N	S		

Tabela 2: Dicionário de dados referente ao sistema já normalizado.

4 - Análise crítica das fontes de dados

Durante o processo de obtenção dos dados relacionados à saúde, o grupo enfrentou algumas dificuldades para a manipulação correta deles. Apesar dos dados estarem em um site organizado e se apresentarem, em grande maioria, completos e fáceis de serem interpretados: ocorreram alguns problemas que tivemos que contornar e tivemos que realizar operações para o aprimoramento dos dados. Seguem alguns dos desafios enfrentados:

- Subnotificação
- Coleta manual dos dados
- Realização de códigos
- Despadronização do Banco
- Incompatibilidade de período de tempo

A subnotificação é um importante fator a ser considerado principalmente em análises de epidemias como foi a de COVID-19. Muitas prefeituras, principalmente de cidades menores em relação ao tamanho da população, não enviaram dados completos de casos, óbitos e vacinação e assim foi necessário realizar uma análise criteriosa e tentar considerar as cidades com mais de 100.000 habitantes. Além disso,

existia por parte da população um grande estigma relacionado à doença e por isso muitas pessoas que tiveram contato com contaminados e posteriormente tiveram os mesmos sintomas não realizaram o teste, na época muitos diziam “não pega covid quem não faz teste”.

Em relação ao processo de coleta dos dados, eles não se encontraram em um estado imediato de importação do site e análise pronta. Foi necessário uma varredura, verificar os períodos de tempo dos dados de vacinação e óbitos, selecionar os filtros de anos manualmente, e assim importá-los de forma compatível e que fosse possível os comparar.

Ademais, existiam algumas inconsistências no banco referentes ao tratamento dos atributos nulos. Na tabela de Óbitos, por exemplo, existia uma coluna em que os atributos de valor nulo eram sinalizados por “-” no lugar do elemento, e outra coluna em que os atributos nulos eram sinalizados por uma string especificando o porque daquele elemento estar nulo, sendo que havia somente um motivo. Ocorrências como essa atrapalham o tratamento dos dados. Seria mais produtivo que os autores dos dados criassem um documento especificando a causa dessas exceções e as representassem de forma padrão no banco.

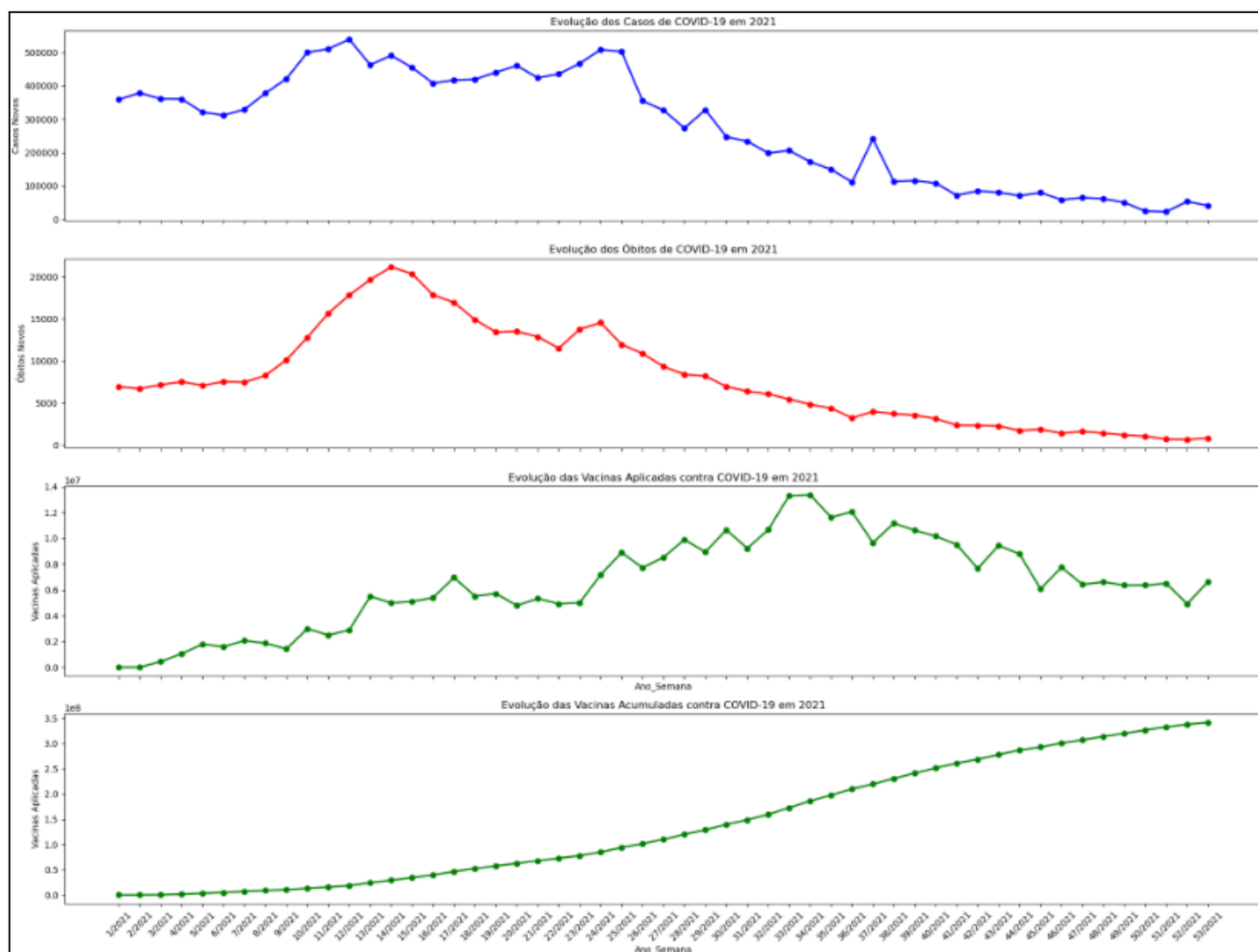
Outro contratempo que enfrentamos está relacionado a incompatibilidade do período de tempo entre as tabelas de vacinação, de óbitos e de casos com a tabela de localidade de COVID-19.

5 - Estudo acerca da correlação entre os dados e seus efeitos práticos

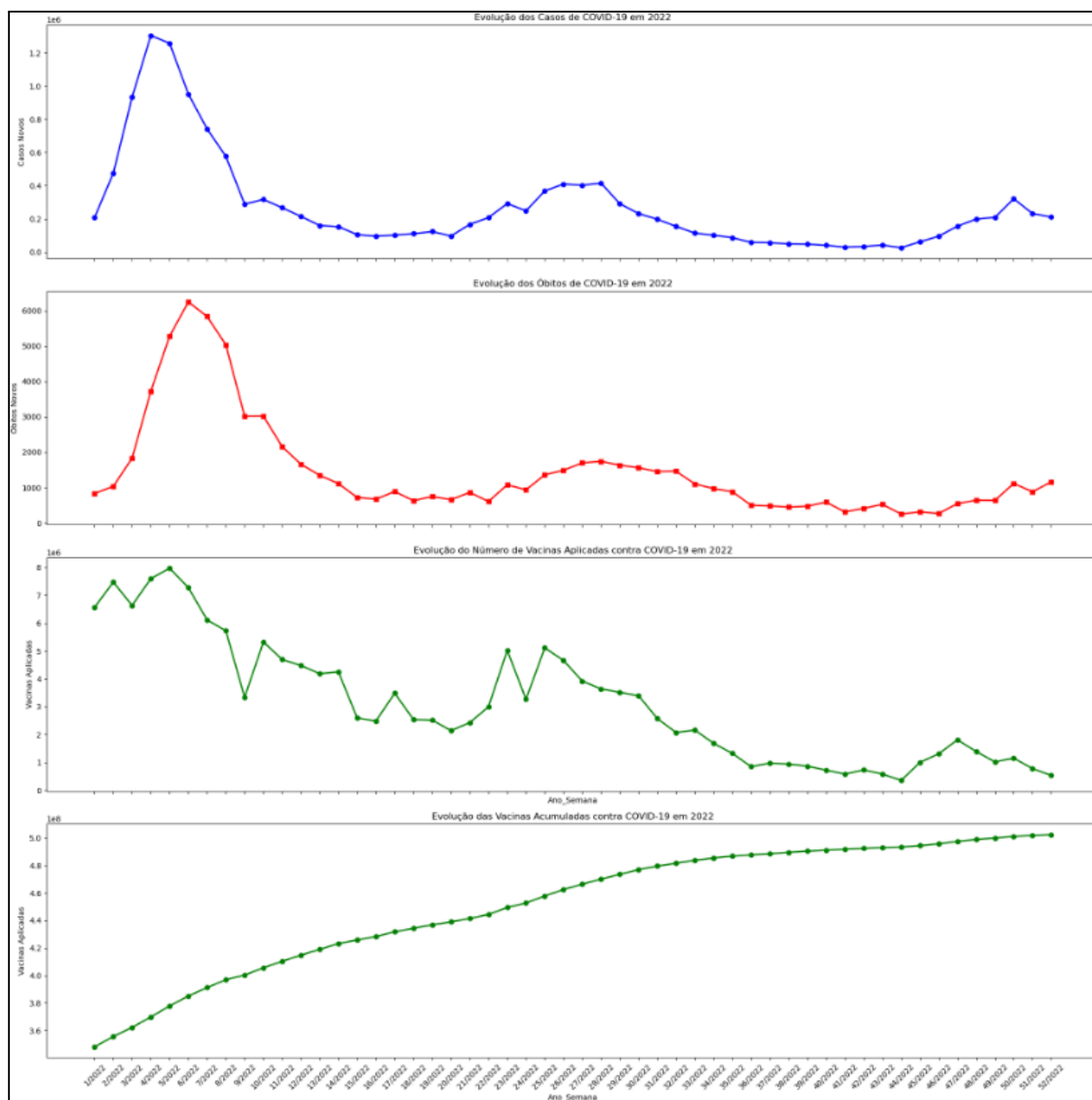
5.1.1 - Visualização dos dados

Abaixo estão os gráficos relacionados ao ano de 2021 a 2025 de dados sobre covid-19, que são respectivamente: casos, óbitos, vacinas aplicadas e número de vacinas acumulado.

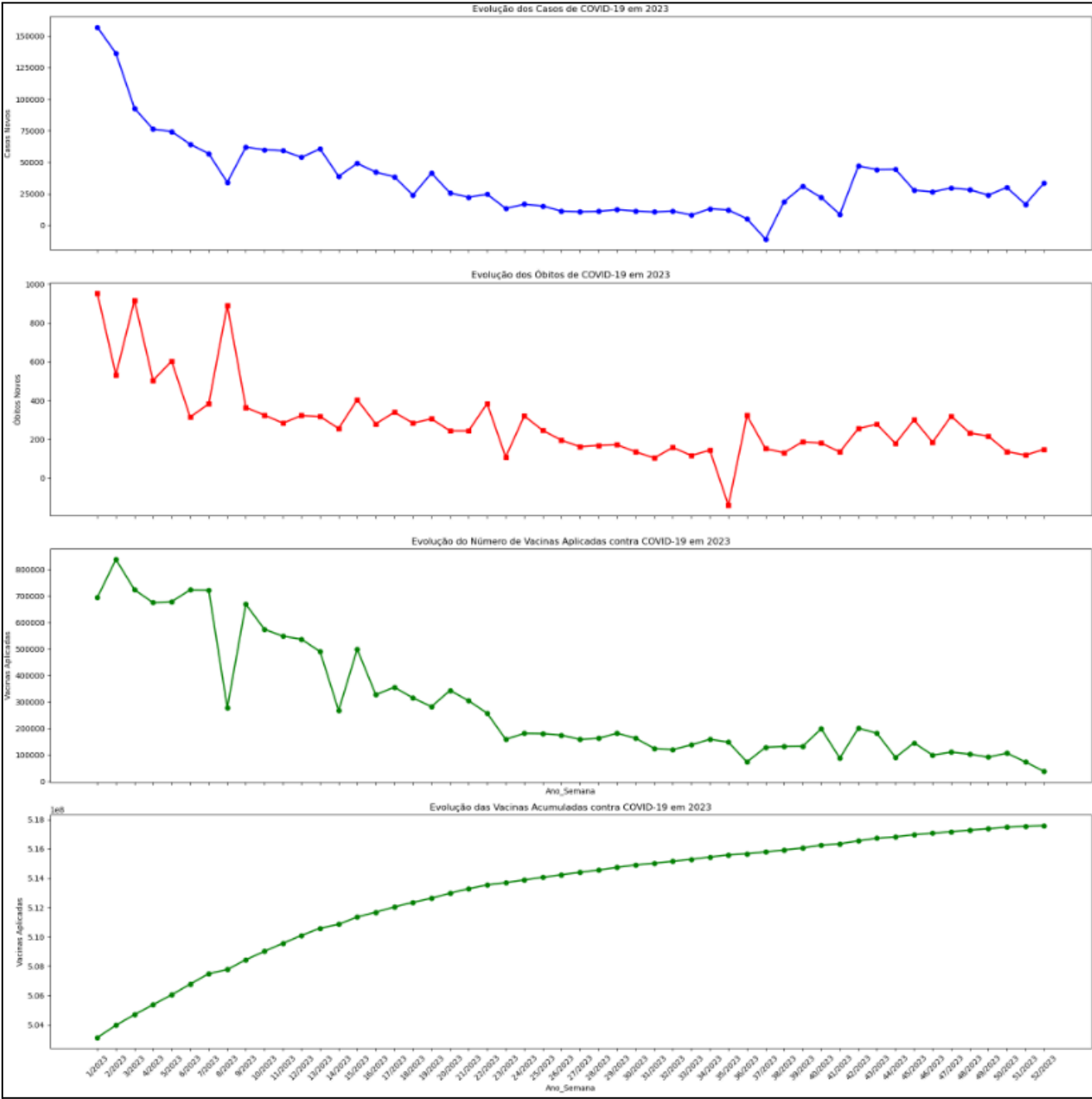
2021:



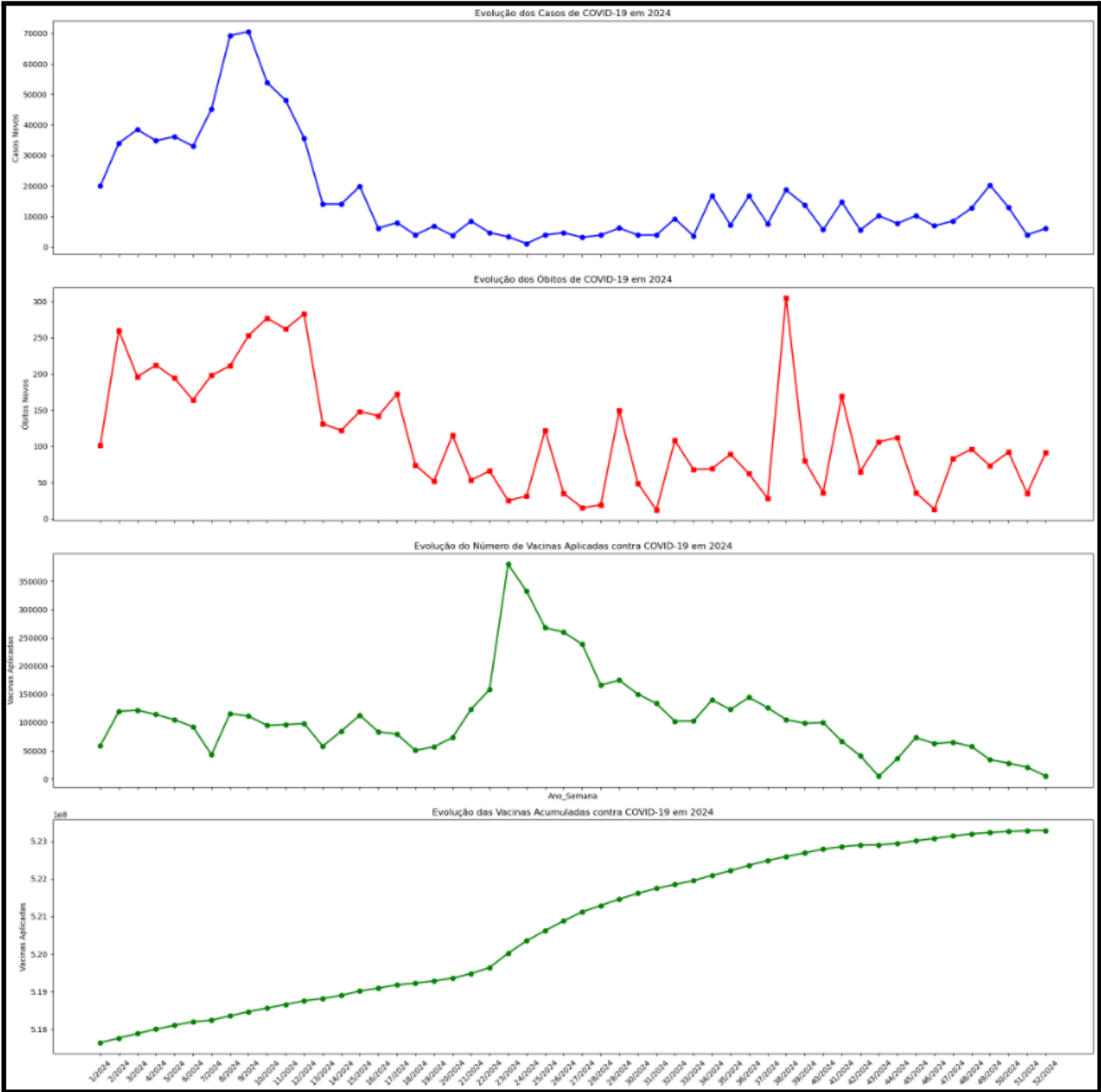
2022:



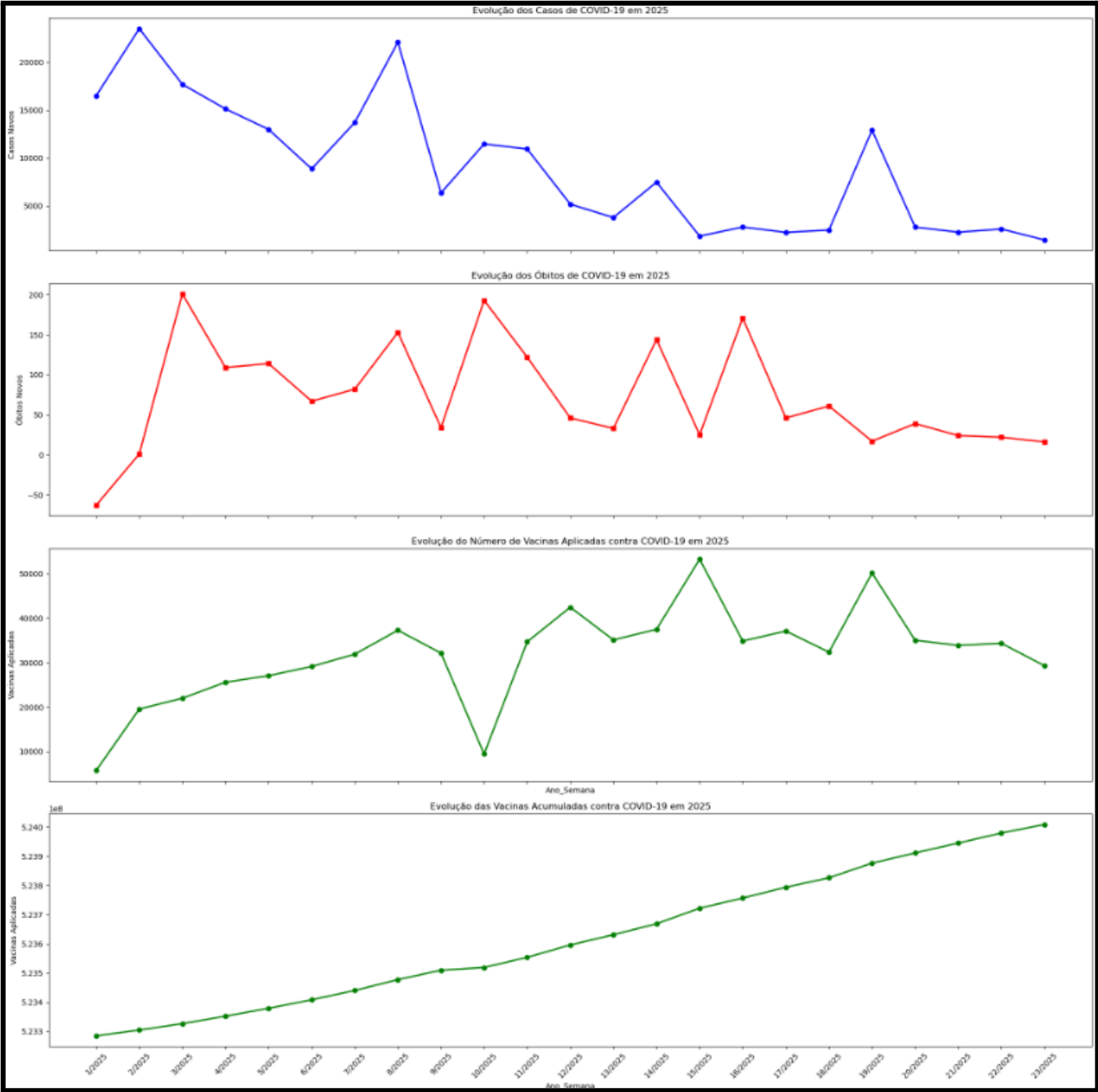
2023:



2024:



2025:



5.1.2 - Correlações:

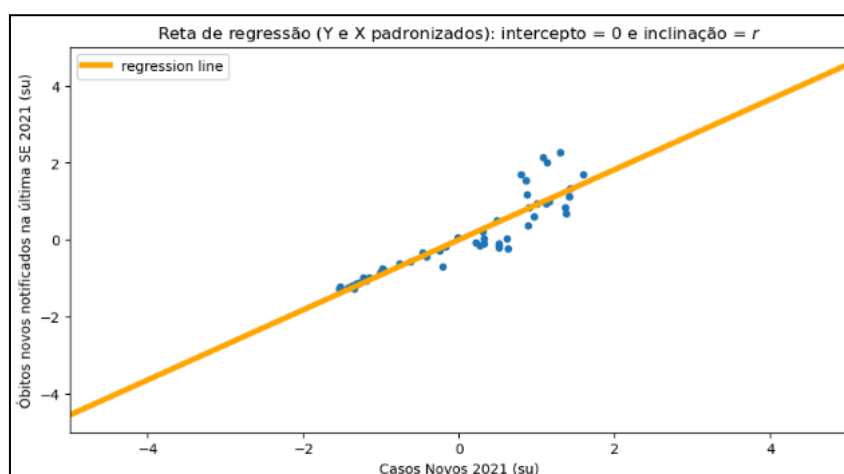
Correlações entre óbitos e casos da COVID - 19

2021	0.9117263091608317
2022	0.8082457778443201
2023	0.7400372305754842
2024	0.7695781310676145
2025	0.22625786427644795

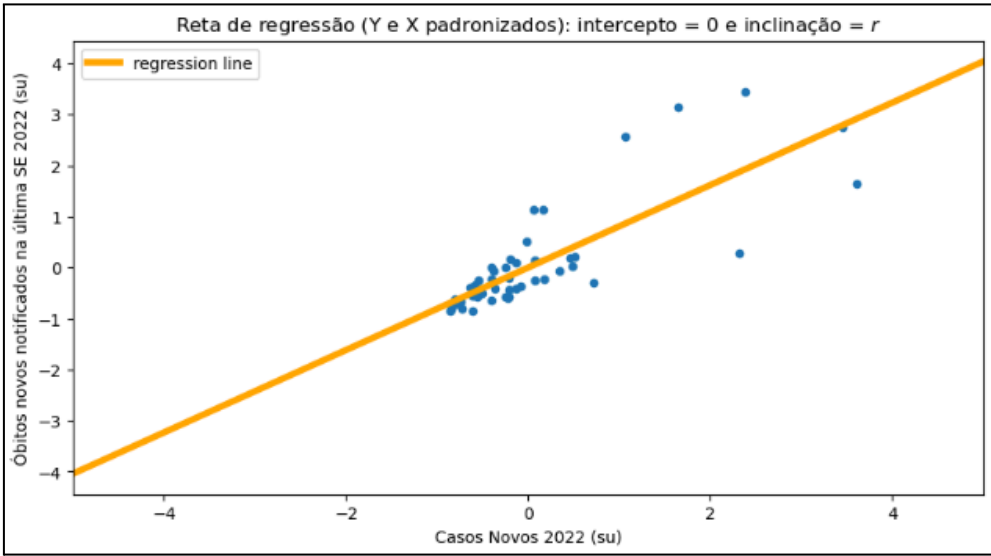
Podemos ver que os dados possuem um índice de correlação de Pearson consideravelmente alto, mas eles diminuem ao decorrer do tempo, ou seja, mesmo com uma alta nos números de contaminados, o número de óbitos não os acompanhava. Isso tem a ver com os efeitos da vacinação, avanço das técnicas de tratamento e um maior conhecimento acerca da doença. Isso pode ser visto também nos gráficos da seção anterior, visto que o maior número de casos em uma semana de 2021 foi de aproximadamente 500 mil pessoas e 20 mil óbitos, uma média de 1 óbito a cada 25 contaminados. Já em 2022 a semana com maior número de casos teve 1 milhão e 200 mil ocorrências e 6 mil óbitos, uma média de 1 óbito a cada 200 casos.

Seguem-se os gráficos de correlação com a reta de regressão, o gráfico a seguir representa o ano de 2021, com reta de regressão com inclinação de aproximadamente 0.911.

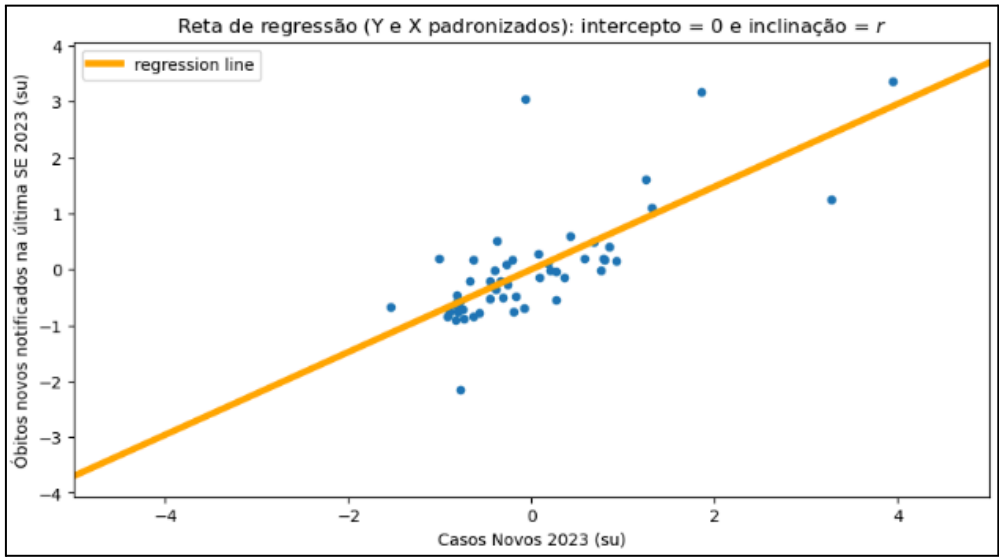
2021:



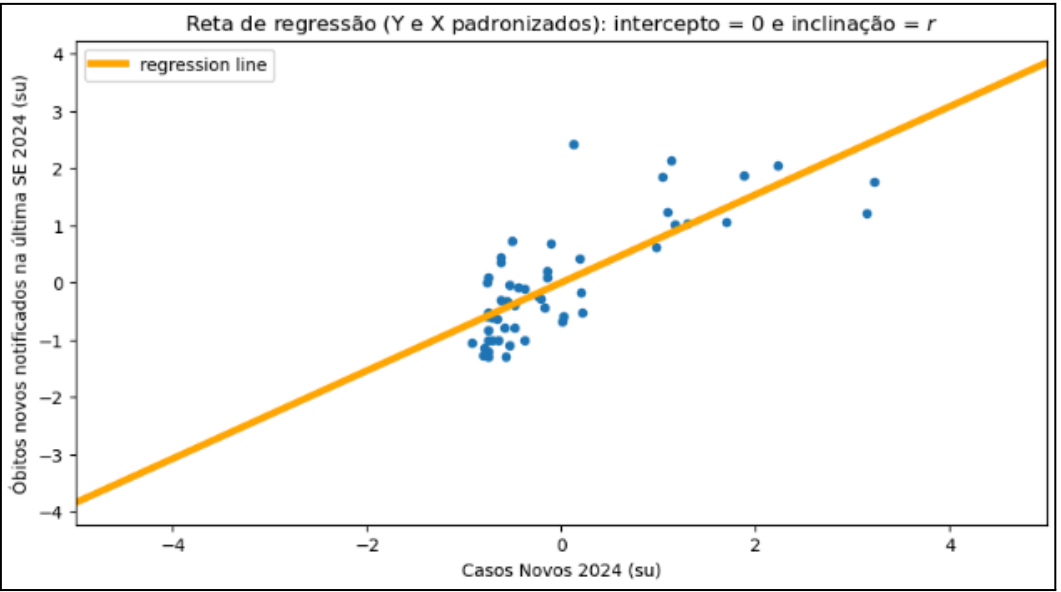
2022:



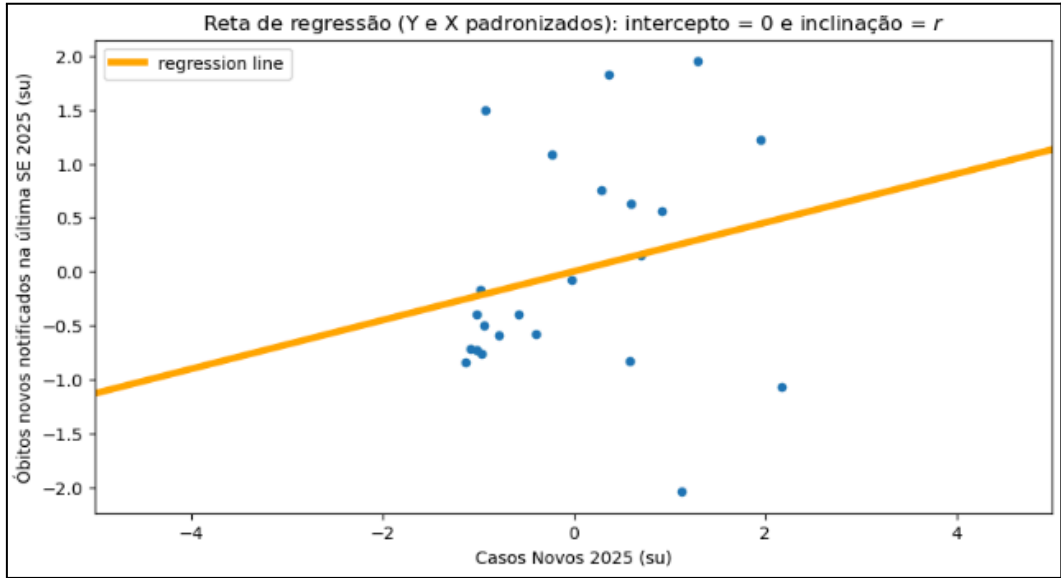
2023



2024:



2025:

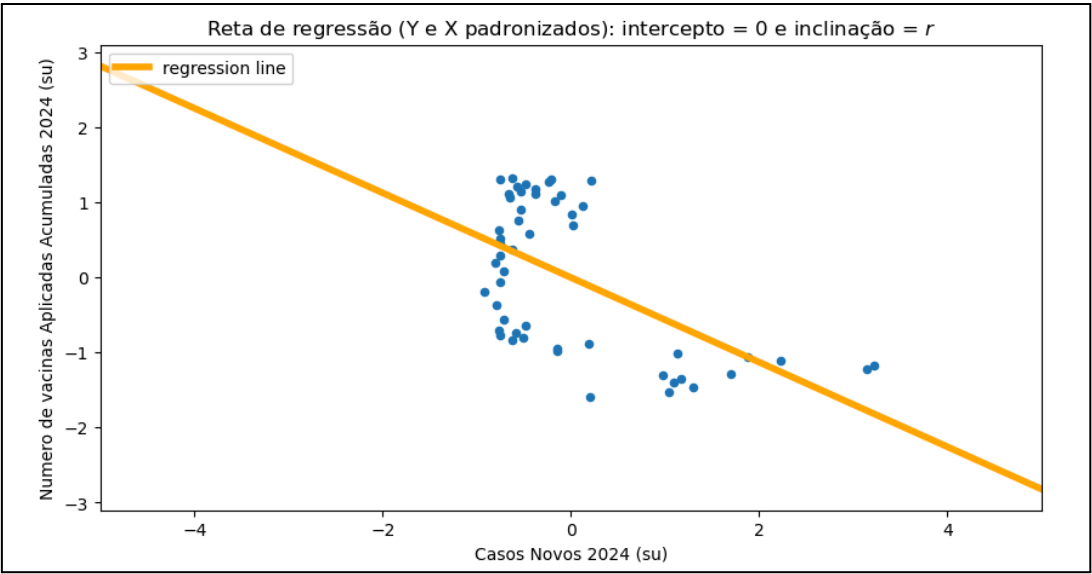


Correlações entre casos da COVID - 19 e vacinação acumulada>

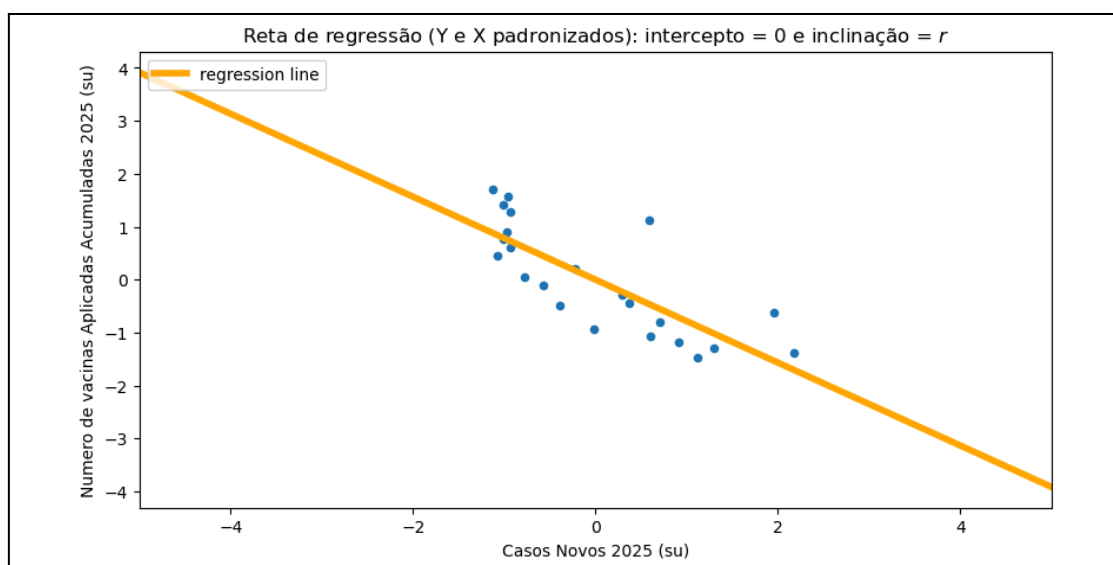
2021	-0.9143571233924207
2022	-0.6394070335032497
2023	-0.8002284746236002
2024	-0.5638451690373697
2025	-0.783022358736976

Podemos ver que os dados possuem um índice de correlação de Pearson consideravelmente baixo, todos eles abaixo de -0.5, o que indica que quanto mais os valores de vacinação acumulada aumentam, o número de casos tende a diminuir. Isso segue um conceito de imunidade coletiva ou imunidade de rebanho, que diz que a partir de um número de pessoas de uma comunidade forem vacinadas contra determinada doença, o vírus causador dessa enfermidade tende a ter maior dificuldade de se proliferar entre as pessoas, pois uma grande parte delas está imune.

2024:



2025:



Correlações entre óbitos de COVID - 19 e vacinação acumulada:

2021	-0.785417271089891
2022	-0.5993424411815771
2023	-0.7717786361768516
2024	-0.5432328391695033
2025	-0.2520329696108848

A partir desses índices de correlação de Pearson podemos concluir que o número de óbitos e de vacinação acumulada são negativos para todos os anos, ou seja, quando um aumenta o outro tende a diminuir. Esses números já eram esperados, pois como óbitos e casos são correlacionados positivamente e casos e vacinação acumulada tem correlação negativa, então de maneira transitiva podemos confirmar a hipótese de que óbitos e vacinação acumulada também tem correlação negativa.

5.2 - Conclusão

Em termos gerais, o trabalho de coletar os dados de forma correta e manipulá-los para serem mais fáceis de serem interpretados ajudou a comprovar um ponto de extrema importância para a saúde global: o impacto da vacinação no controle seja de doenças epidêmicas ou não, ou seja, seu impacto para a diminuição do número de óbitos e também de casos também.

6 - Bibliografia

- [1] **Observatório Covid-19 aponta maior colapso sanitário e hospitalar da história do Brasil | Portal Fiocruz.** Disponível em:
<<https://fiocruz.br/noticia/2021/03/observatorio-covid-19-aponta-maior-colapso-sanitario-e-hospitalar-da-historia-do>>. Acesso em: 25 jun. 2025.
- [2] MAGENTA, M. **Covid: Variante acelera intubação de jovens e SP orienta procurar ajuda no 10 dia de sintomas.** Disponível em:
<<https://noticias.uol.com.br/ultimas-noticias/bbc/2021/04/19/covid-variante-variante-sao-paulo-intubacao-jovens-coronavirus.htm>>. Acesso em: 25 jun. 2025.
- [3] **Ministério da Saúde declara fim da Emergência em Saúde Pública de Importância Nacional pela Covid-19.** Disponível em:
<<https://www.gov.br/saude/pt-br/assuntos/noticias/2022/abril/ministerio-da-saude-declara-fim-da-emergencia-em-saude-publica-de-importancia-nacional-pela-covid-19>>.
- [4] NUÑO DOMÍNGUEZ. **Por que a vacina contra a covid-19 não impede a transmissão do vírus.** Disponível em:
<<https://brasil.elpais.com/ciencia/2021-11-18/por-que-a-vacina-contr-a-covid-19-nao-impede-a-transmissao-do-virus.html>>. Acesso em: 25 jun. 2025.
- [5] **Dashboard-pt.** Disponível em: <<https://www.genomahcov.fiocruz.br/dashboard-pt/>>.
- [6] **COVID-19 - Painei COVID-19 - Estado do Espírito Santo.** Disponível em:
<<https://coronavirus.es.gov.br/painel-covid-19-es>>.