

UNIVERSIDADE DE SÃO PAULO

Instituto de Ciências Matemáticas e de Computação

Análise da evolução dos números da pandemia da COVID-19 e das medidas de combate adotadas para a contenção da doença: Uma aplicação em Ciência de Dados

Reinaldo Lepsch Neto

Trabalho de Conclusão de Curso - MBA em Ciência de Dados
(CEMEAI)

UNIVERSIDADE DE SÃO PAULO

Instituto de Ciências Matemáticas e de Computação

Análise da evolução dos números da
pandemia da COVID-19 e das
medidas de combate adotadas para
a contenção da doença: Uma
aplicação em Ciência de Dados

Reinaldo Lepsch Neto

USP - São Carlos

2022

REINALDO LEPSCH NETO

**Análise da evolução dos números da pandemia da COVID-19 e das medidas de
combate adotadas para a contenção da doença: Uma aplicação em Ciência de
Dados**

Trabalho de conclusão de curso apresentado ao Centro de Ciências Matemáticas Aplicadas à Indústria do Instituto de Ciências Matemáticas e de Computação, Universidade de São Paulo, como parte dos requisitos para conclusão do MBA em Ciência de Dados.

Área de concentração: Ciências de Dados

Orientador: Prof. Dr. Wallace Correa de Oliveira Casaca

USP - São Carlos

2022

Ficha catalográfica elaborada pela Biblioteca Prof. Achille Bassi
e Seção Técnica de Informática, ICMC/USP,
com os dados inseridos pelo(a) autor(a)

L611a LEPSCH NETO, REINALDO
Análise da evolução dos números da pandemia da
COVID-19 e das medidas de combate adotadas para a
contenção da doença: Uma aplicação em Ciência de Dados
/ REINALDO LEPSCH NETO; orientador Wallace Correia
de Oliveira Casaca. -- São Carlos, 2022.
76 p.

Trabalho de conclusão de curso (MBA em Ciência
de Dados) -- Instituto de Ciências Matemáticas e de
Computação, Universidade de São Paulo, 2022.

1. Ciência de Dados. 2. Epidemiologia. 3.
Estatística. I. Correia de Oliveira Casaca, Wallace,
orient. II. Titulo.

Bibliotecários responsáveis pela estrutura de catalogação da publicação de acordo com a AACR2:
Gláucia Maria Sáa Cristianini - CRB - 8/4938
Juliana de Souza Moraes - CRB - 8/6176

FOLHA DE AVALIAÇÃO OU APROVAÇÃO

DEDICATÓRIA

Aos valorosos trabalhadores dos serviços de saúde, que tem enfrentado toda sorte de dificuldade, chegando ao sacrifício da própria vida para salvar outras tantas.

Aos profissionais da ciência, que neste inédito período de obscurantismo e celebração da ignorância veem repetir-se padrões de negação próximos àqueles da Idade Média.

Aos (poucos) líderes mundiais, e alguns (mais raros ainda) locais, que decidiram fazer a coisa certa, no momento correto e com isso pouparam inúmeras vidas.

A meu querido amigo **Amilton Silveira**, bem mais que um número perdido entre vítimas anônimas da COVID-19, uma ausência dolorida em minha vida e de sua família.

A meus amados familiares, meus pais, minha amada esposa, sempre paciente a meu lado, e meu filho, meu melhor amigo, sempre próximo mesmo estando tão longe. Não conseguia atravessar tempos tão difíceis sem vocês.

AGRADECIMENTOS

Ao meu orientador, Prof. Dr. Wallace Correa de Oliveira Casaca, por desde o início ter se mostrado um verdadeiro amigo, pelo apoio e esclarecimentos na condução deste projeto, e pela inspiração através do brilhante projeto SP COVID-19 Info Tracker.

Ao Prof. Dr. Miguel A. L. Nicolelis que, embora sem conhecê-lo pessoalmente, também trouxe muita inspiração, pela sua brilhante atuação pública em defesa da ciência brasileira, contra o negacionismo e trabalhando incansavelmente desde o início da pandemia, em especial junto a pesquisadores e gestores públicos da região nordeste. Suas palavras, seja em seu canal no YouTube ou em entrevistas a telejornais, foram decisivas na escolha do tema deste trabalho. Igualmente, ao Dr. Drauzio Varella, felizmente bem conhecido do público brasileiro por sua atuação na mídia televisiva e escrita, sempre transmitindo sua visão humanista e esclarecedora, que me lembro desde garoto no rádio.

Gostaria de adicionar aqui três brilhantes nomes da ciência brasileira, que graças a corajosa atuação da mídia brasileira juntaram-se ao Dr. Nicolelis, marcando presença quase diária nos noticiários e trazendo preciosos esclarecimentos – também com conteúdos muito úteis na Internet: Dr. Atila Iamarino, Dra. Natalia Pasternak e Dra. Luana Araújo. Dedicaram-se, além de nos trazer muito conhecimento, a nos mostrar que **ciência não é difícil – para entendê-la basta amá-la.**

Falando em amar a ciência, concluo acrescentando um nome que me acompanha há mais de quarenta anos e cujas ondas positivas atravessam continuamente o espaço-tempo até mim, desde que sua brilhantíssima mente mergulhou no oceano cósmico, há 25 anos, aqui deixando uma obra magnífica. Obrigado, Dr. CARL SAGAN.

Wheels within wheels in a spiral array

A pattern so grand and complex

Time after time we lose sight of the way

Our causes can't see their effects

(Neil Elwood Peart, 1952-2020)

RESUMO

LEPSCH NETO, R. Análise da evolução dos números da pandemia da COVID-19 e das medidas de combate adotadas para a contenção da doença: Uma aplicação em Ciência de Dados:

2022. 76f. Trabalho de conclusão de curso (MBA em Ciência de Dados) – Instituto de Ciências Matemáticas e de Computação, Universidade de São Paulo, São Carlos, 2022.

Este trabalho apresenta uma análise exploratória dos dados relativos à pandemia da COVID-19, doença infectocontagiosa surgida na China nos últimos meses de 2019 e que adquiriu caráter global em fevereiro de 2020. O ponto central das observações foi o acompanhamento, no tempo, das curvas de novos casos em paralelo com a variação da mobilidade, segundo as categorias e critérios definidos no Relatório de Mobilidade Comunitária do Google. Foi avaliada a correlação linear entre a sequência de casos e de mobilidade, sendo possível concluir que o aumento na mobilidade favoreceu a circulação do vírus. A comparação da mobilidade com os casos foi feita tanto consolidando a mobilidade num único número (média), quanto na análise individual das seis categorias do relatório do Google (residências, trabalho, transporte, parques, comércio e recreação, alimentação e farmácias). O foco principal do estudo foi nas cidades do Estado de São Paulo, complementado com comparações com os números consolidados do Brasil e de dois países escolhidos como *benchmarks* – Reino Unido e Chile. A análise foi estendida com a comparação das curvas de crescimento de casos e outros números no Estado de SP, e por último foram incluídos alguns dados sobre a evolução do processo de vacinação. Como resultado do controle da mobilidade e da aplicação das vacinas, foi possível observar o controle dos casos e óbitos diários da doença. Como conclusão central do estudo, apontamos a importância do acompanhamento contínuo dos dados, através de painéis em tempo real e bases de dados atualizadas recorrentemente, possibilitando desta forma a tomada tempestiva de decisões por gestores públicos comprometidos com a saúde pública.

Palavras-chave: Pandemia. COVID-19. Dados. Modelos. Epidemiologia. Estatística. Ciência de Dados.

ABSTRACT

LEPSCH NETO, R. **Analysis of the COVID-19 pandemic data and the measures of combat against the disease: A Data Science application:**

2022. 76f. Trabalho de conclusão de curso (MBA em Ciência de Dados) – Instituto de Ciências Matemáticas e de Computação, Universidade de São Paulo, São Carlos, 2022.

This work presents an exploratory analysis of data related to the COVID-19 pandemic, an infectious disease that emerged from China in the last months of 2019 which acquired a global status in February 2020. The main goal of this study was to track, over time, the curves of new cases as well as the mobility variation, according to the categories and criteria as properly defined in the Google Community Mobility Report. The linear correlation between the cases and mobility was calculated so that it was found that the increase of mobility favored virus circulation. The comparison of mobility against the cases was performed both consolidating mobility in a single number (average), as well as breaking down into the six categories of the Google report (homes, work, transport, parks, commerce and recreation, food and pharmacies). Our analysis was focused on cities of the State of São Paulo, complemented with comparisons with the consolidated numbers of Brazil and two countries chosen as benchmarks – United Kingdom and Chile. The analysis was complemented with the comparison of case growth curves and other numbers in the State of SP, and finally, some data on the evolution of the vaccination process were also included. As a result of mobility control processes and the application of vaccines, it was possible to control the number of new cases and daily deaths. As a main finding, the study points out the importance of recurrent data monitoring, from real-time panels and data sets, thus enabling timely decision-making by public managers committed to public health.

Keywords: Pandemic. COVID-19. Data. Models. Epidemiology. Statistics. Correlation. Data Science.

LISTA DE FIGURAS

Figura 1 - Painel global da Universidade Johns Hopkins	2
Figura 2 - Mapa da Vacinação Contra Covid-19 no Brasil.....	3
Figura 3 - Comparação de evolução de casos e mobilidade para São Paulo	26
Figura 4 - Comparação de evolução de casos e mobilidade para Botucatu	27
Figura 5 - Comparação de evolução e casos e mobilidade para Barreiros	28
Figura 6 - Comparação de evolução de casos e mobilidade para São Carlos	29
Figura 7 - Comparação de evolução de casos e mobilidade para Paulínia	30
Figura 8 - Comparação de evolução de casos e mobilidade para Campinas	31
Figura 9 - Comparação de evolução de casos e mobilidade para Ribeirão Preto	32
Figura 10 - Comparação de evolução de casos e mobilidade para Franca.....	33
Figura 11 - Comparação de evolução de casos e mobilidade para Itupeva	34
Figura 12 - Comparação de evolução de casos e mobilidade para Itapevi	35
Figura 13 - Comparação de evolução de casos e mobilidade para Itapecerica da Serra.....	36
Figura 14 - Comparação de evolução de casos e mobilidade para São Caetano do Sul	37
Figura 15 - Comparação de evolução de casos e mobilidade para Jales.....	38
Figura 16 - Comparação de evolução de casos e mobilidade para Serra Negra	39
Figura 17 - Comparação de evolução de casos e mobilidade para o estado de São Paulo	40
Figura 18 - Mapa e matriz de correlações para Botucatu.....	46
Figura 19 - Mapa e matriz de correlações para Votorantim.....	47
Figura 20 - Mapa e matriz de correlações para Sorocaba	48
Figura 21 - Mapa e matriz de correlações para Registro	49
Figura 22 - Mapa e matriz de correlações para o estado de São Paulo	50
Figura 23 - Distribuição das categorias principais por município - Estado de São Paulo	51
Figura 24- Distribuição das categorias principais por município - Estado de São Paulo - versão linear	52
Figura 25 - Média geral de variação de mobilidade para janeiro-fevereiro/2021	52
Figura 26 - Ocorrência de casos para abril/2021	53
Figura 27 - Comparação de evolução de casos e mobilidade para o Brasil.....	55
Figura 28 - Comparação de evolução de casos e mobilidade para o Chile.....	56
Figura 29 - Comparação de evolução de casos e mobilidade para o Reino Unido.....	57
Figura 30 - Mapa e matriz de correlações para o Brasil	58
Figura 31 - Mapa e matriz de correlações para o Chile	59

Figura 32 - Mapa e matriz de correlações para o Reino Unido	60
Figura 33 - Casos acumulados por 100 mil habitantes desde fevereiro/2020 - cidades acima 1 MM habitantes.....	61
Figura 34 - Casos acumulados por 100 mil habitantes desde fevereiro/2020 - cidades de 500 mil a 1 MM habitantes	62
Figura 35 - Casos acumulados por 100 mil habitantes desde fevereiro/2020 - cidades até 1500 habitantes.....	62
Figura 36 - Correlação entre óbitos e casos por 100 mil habitantes desde fevereiro/2020.....	63
Figura 37 - Casos acumulados por 100 mil habitantes em 31/12/2020	64
Figura 38 - Mapa e matriz de correlações para vacinação em Americana.....	67
Figura 39 - Mapa e matriz de correlações para vacinação no município de São Paulo.....	68
Figura 40 - Mapa e matriz de correlações para vacinação em Botucatu.....	69

LISTA DE TABELAS

Tabela 1 - Número médio diário de mortos durante a 2 ^a Guerra Mundial por país.....	4
Tabela 2 Print parcial do Google Mobility Report após os devidos acertos manuais prévios.....	24

LISTA DE ABREVIATURAS E SIGLAS

CEMEAI – Centro de Ciências Matemáticas Aplicadas à Indústria, vinculado ao ICMC-USP e que busca através da matemática e suas ferramentas gerar produtos e serviços úteis a parceiros públicos e privados.

CEPID – Centros de Pesquisa, Inovação e Difusão, apoiados pela FAPESP por um período determinado.

COVID-19 - é o nome oficial da doença causada pelo vírus **SARS-CoV-2**. Para facilitar a leitura, iremos nos referir a este último simplesmente como Coronavírus, por ser o único referido neste trabalho – existem outros Coronavírus, que serão explicitamente nomeados de forma distinta quando for o caso

FAPESP – Fundação de Amparo à Pesquisa do Estado de São Paulo. Ligada a Secretaria de Desenvolvimento Econômico do Governo do Estado de São Paulo, é autônoma e possui a missão de avaliar, acompanhar e financiar projetos científicos, além de subvencionar estudos a nível de graduação e pós-graduação que gerem produtos inovadores.

IBGE – Instituto Brasileiro de Geografia e Estatística.

ICMC – Instituto de Ciências Matemáticas e da Computação, pertencente a USP – campus de São Carlos.

SEADE – Fundação Sistema Estadual de Análise de Dados Estatísticos – órgão subordinado à Secretaria de Planejamento e Gestão do Governo do Estado de São Paulo.

SIMI – Sistema de Monitoramento Inteligente – série de indicadores publicados pelo Governo do Estado de São Paulo mostrando a evolução dos números da pandemia no estado em tempo real, incluindo os dados de mobilidade fornecidos (com garantia de anonimidade) pelas operadoras de telefonia celular locais.

UNESP – Universidade Estadual Paulista.

USP – Universidade de São Paulo.

WHO – World Health Organization, ou Organização Mundial da Saúde, mais conhecida pela sua sigla em português OMS.

SUMÁRIO

1 INTRODUÇÃO.....	1
1.1 CONTEXTO DO LEVANTAMENTO	1
1.3 OBJETIVOS DESTE TRABALHO	5
1.4 CORRELAÇÃO E CAUSALIDADE	6
1.5 A PRODUÇÃO CIENTÍFICA EM CIÊNCIA DE DADOS.....	8
2 REVISÃO BIBLIOGRÁFICA	10
3 METODOLOGIA PARA A PESQUISA	14
3.1 PREMISSAS FUNDAMENTAIS.....	14
3.2 BUSCA DO OBJETIVO PRIMORDIAL.....	14
3.3 FONTES DE DADOS	15
3.4 ALGUNS DETALHES DE IMPLEMENTAÇÃO	17
3.5 FERRAMENTAS UTILIZADAS	18
3.6 ESCOPO GEOGRÁFICO.....	20
3.7 ESCOPO NO TEMPO	21
3.8 DISTORÇÕES NAS BASES DE DADOS	22
3.9 MODELOS COMPUTACIONAIS	22
4 RESULTADOS OBTIDOS	23
4.1 ANÁLISE DOS DADOS DE MOBILIDADE DO GOOGLE PARA O ESTADO DE SÃO PAULO (ANÁLISE 001)	24
<i>Evolução no tempo de mobilidade x novos casos.....</i>	<i>26</i>
<i>Correlações por categoria de mobilidade</i>	<i>41</i>
4.2 ANÁLISE COMPARATIVA DOS DADOS DE MOBILIDADE DO GOOGLE PARA OS PAÍSES: BRASIL, CHILE E REINO UNIDO (ANÁLISE 003)	54
4.3 ANÁLISE COMPARATIVA DOS DADOS DO SEADE PARA OS MUNICÍPIOS DO ESTADO DE SÃO PAULO (ANÁLISE 004).....	61
4.4 ANÁLISE COMPARATIVA DOS DADOS DO PROJETO COVID-19 INFO TRACKER SOBRE VACINAÇÃO (ANÁLISE 005).....	65
5 CONCLUSÃO	70
5.1 RESULTADOS E IMPACTOS ESPERADOS – VISÃO PRESCRITIVA	70
5.2 CONSIDERAÇÕES FINAIS	72
REFERÊNCIAS	73
WEBSITES.....	75

1 INTRODUÇÃO

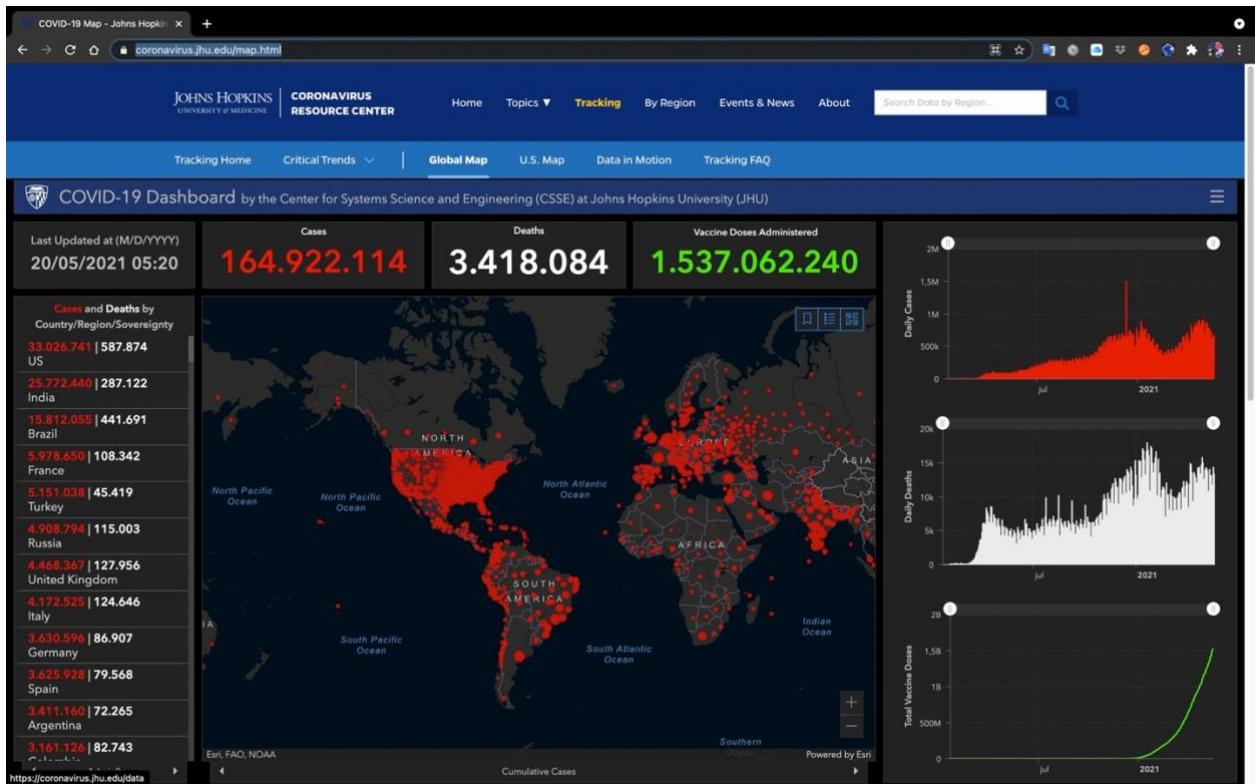
1.1 CONTEXTO E INTRODUÇÃO AO TEMA

Não detalharemos neste trabalho, os primórdios daquilo que está sendo uma das maiores crises sanitárias da História da Humanidade, que é a pandemia da COVID-19, iniciada no final de 2019 e oficialmente reconhecida como tal em fevereiro de 2020 pela Organização Mundial da Saúde (OMS). Portanto, este trabalho endereça uma crise sanitária que ainda está em curso, com desdobramentos bastante graves pela sua grande quantidade de vítimas (entre infectados ainda em tratamento, sobreviventes com sequelas e óbitos) e que se estendem por todos os setores das atividades humanas em todo o planeta, em especial nos campos da saúde, econômico e político. Temos, ainda, observado empresas de todos os portes e tempo de existência iniciando processo de falência ou lutando para se manterem na ativa, além dos altos índices de desemprego. E no âmbito político, governos nacionais, regionais e locais pesadamente afetados, levando-nos à possibilidade de um futuro geopolítico totalmente imprevisível.

Este é, então, um cenário de alta complexidade, que ainda está diariamente se modificando, e suas consequências ainda estão por se desenhar. Porém, ao longo de dois anos de pandemia, um volume considerável de dados foi acumulado, registrando a evolução da doença e, em paralelo, de medidas de combate à mesma, como por exemplo, o isolamento social (desde o início dos casos em âmbito global, e de modo irregular no Brasil) e o processo de vacinação massiva (desde o segundo semestre de 2020, iniciando no Brasil em janeiro/2021).

A rápida evolução da doença, gerando desde o início um volume muito grande de dados, logo exigiu o auxílio dos profissionais de Ciência de Dados. Centros de pesquisa e grandes universidades disponibilizaram seus especialistas e recursos computacionais para tratar e entender tais bases de dados, e o mais importante: criar ferramentas confiáveis e de fácil visualização para ajudar a tomar medidas efetivas no combate à pandemia. Painéis de indicadores, carregados a partir das referidas bases, tornaram-se uma constante em centros médicos, salas de reuniões governamentais e até mesmo no noticiário: a rapidez no desenvolvimento e disponibilização de painéis de dados como, por exemplo, o da Universidade Johns Hopkins (Baltimore, EUA), que foi a primeira ferramenta de visualização de dados em escopo global sobre o fenômeno da COVID-19 (Figura 1).

Figura 1 - Painel global da Universidade Johns Hopkins



Fonte: Centro de Ciências de Sistemas e Engenharia da Universidade Johns Hopkins, Baltimore, EUA. Posição de 20/05/2021, 06:49 Horário de Brasília

Os dados da Johns Hopkins e de outras fontes passaram a ser divulgados publicamente, inclusive por sites especializados em Ciência de Dados como o Kaggle (kaggle.com), passando então a ser estudados por cientistas, engenheiros, analistas e pesquisadores do mundo inteiro, reunindo não só ciências médicas e da saúde, como também computação, matemática, estatística, num esforço acadêmico global para entender e vencer o vírus.

Bem cedo surgiram denúncias de super/subnotificação nas fontes desses mesmos dados, geralmente a partir de serviços públicos de saúde despreparados para lidar com a força e a velocidade de avanço do vírus. A ciência de dados e a medicina ajudaram onde possível, até encontrarem um sério obstáculo: o uso e manipulação política dos números da pandemia por governos de todos os matizes ideológicos em muitos países do mundo, com os objetivos mais diversos. O surgimento de movimentos negacionistas em aliança com o extremismo político colocou todos os profissionais da ciência sob forte pressão. A ciência teve que, muitas vezes, afastar-se do foco de seu complexo trabalho – entender um vírus desconhecido e muito poderoso – para se defender daqueles a quem estava ajudando.

Os conflitos aumentaram quando as primeiras medidas para mitigar o crescimento da pandemia começaram a ser discutidas e propostas. Implicavam em distanciamento social e evitar aglomerações, o que levava diretamente à possibilidade, depois confirmada, do fechamento (ou restrição do horário de funcionamento) de setores particulares, atingindo assim a economia. Aqueles que optaram pelo negacionismo encontravam, enfim, sua justificativa maior – e obstáculo maior para a ciência. Optando simplesmente pela economia, ao invés de uma gestão que priorizasse também a saúde e a vida, alguns governos contribuíram, finalmente, para a perda de controle da pandemia, o que evidentemente não trouxe benefícios econômicos.

O surgimento das primeiras vacinas, no segundo semestre de 2020, não trouxe a esperada segurança e controle da pandemia, já que a vacinação ampla tem se restringido aos países mais ricos. Por exemplo, o negacionismo interferiu na adoção da vacinação no Brasil, iniciada somente em janeiro de 2021 e por iniciativa de um governo local. Segundo o site G1/Globo, até o presente momento (dezembro de 2021), temos aproximadamente 64% da população brasileira imunizada (duas doses aplicadas, ou dose única), destacando, mais uma vez, o esforço de governos e gestores de saúde estaduais e municipais (Figura 2).

Figura 2 - Mapa da Vacinação Contra Covid-19 no Brasil



1.2 A FORÇA DOS NÚMEROS

O forte crescimento dos números na pandemia, em especial os óbitos, tem causado estorricimento, especialmente a partir do ano de 2021, quando o Brasil e o mundo já enfrentavam o problema havia mais de um ano, e meios de controlar o vírus já deveriam ter sido adotados por todos os governos nacionais/locais.

A desigualdade no processo de vacinação, iniciado no mundo ainda em 2020 (no Brasil a partir de janeiro de 2021, mas com extrema lentidão) levou ao surgimento de números assustadores. Podemos exemplificar a comparação do número máximo de óbitos por COVID-19 ocorridos no Brasil num período de 24 horas, no dia 08/04/2021, de 4.249 mortes, com a média diária de mortes por país durante a Segunda Guerra Mundial (1939-1945), para os principais participantes do conflito:

TABELA 1 - Número médio diário de mortos durante a 2^a Guerra Mundial por país

Países	Número de mortos durante a 2a Guerra	inicio participação	fim participação	anos	dias	mortos por dia
URSS	17.000.000	1941	1945	5	1825	9315
Alemanha	5.500.000	1939	1945	7	2555	2152
Polônia	4.000.000	1939	1945	7	2555	1565
China	2.200.000	1939	1945	7	2555	861
Iugoslavia	1.600.000	1939	1945	7	2555	626
Japão	1.500.000	1941	1945	5	1825	821
França	535.000	1939	1945	7	2555	209
Itália	450.000	1939	1945	7	2555	176
Grã-Bretanha	396.000	1939	1945	7	2555	154
EUA	292.000	1941	1945	5	1825	160
TOTAL	33.473.000			1	365	91706

Fonte: Almanaque Abril 1999, apud http://ces.ufpel.edu.br/vestibular/download/2004i/prova_f1_b.pdf

O número de óbitos no Brasil em 08/04/2021 supera, segundo o quadro acima, as mortes diárias de todos os países envolvidos naquele conflito, exceto a URSS. A gravidade da situação se mostra por compararmos os números do pior conflito armado da História (países que sofreram bombardeios e ocupação por exércitos inimigos) com os de uma pandemia ocorrida em tempo de paz, e com processo de vacinação em andamento. Acompanhar como tais números evoluíram até tal ponto extremo e os possíveis caminhos futuros é a motivação principal para esta pesquisa.

1.3 OBJETIVOS

Este trabalho visa estudar a relação entre as medidas de mitigação do contágio (lockdown ou fechamento compulsório de estabelecimentos comerciais não essenciais, incentivo ao uso de máscara e distanciamento social, combate às aglomerações), aliadas à vacinação, utilizando Análise Exploratória de Dados (AED), abordagens estatísticas, técnicas de visualização, e tratamento de dados.

Tratamentos e medicamentos de eficácia questionada pela ciência, bastante difundidos por pessoas “negacionistas” por supostamente não exigirem medidas que afetem a economia, não serão cobertos nesse trabalho, apesar que a utilização de tais recursos tem o seu impacto negativo sobre as medidas comprovadamente eficazes, já que tais expedientes, combinados com o desprezo pela ciência, têm recebido apoio oficial de muitos governos, entre eles, o Governo Federal do Brasil.

Portanto, trata-se de responder às perguntas, aparentemente simples, mas que possuem grande apelo social e sanitário: *até que ponto as medidas de isolamento e distanciamento social e a vacinação contribuíram/têm contribuído para o controle da pandemia da COVID-19?* Para avaliar possíveis respostas nessa direção, faremos uso de estudos de casos com dados de municípios e países em que tais medidas foram implementadas, buscando correlação entre a evolução dos casos e as medidas.

Devemos lembrar e aqui destacar que a pandemia, em âmbito global, ainda se encontrava fora de controle no momento do início desta pesquisa, o que significa números com grande oscilação e cenários que mudam quase que diariamente. Portanto deveremos dar destaque ao momento em que os dados foram coletados e as análises realizadas, procurando

manter os números atualizados até o término deste trabalho. O surgimento das chamadas variantes, oriundas de mutações locais do vírus, mas que rapidamente espalham-se globalmente, tem sido o problema maior, criando situações que se irradiam de múltiplas origens. A resistência das pessoas, de um modo geral e global, a manter as medidas de segurança depois de um período tão longo e sem ver seu fim num futuro próximo, tem sido o principal fator para que tais ondas surjam e se mantenham – podemos dizer que a fadiga à pandemia em si é um problema bastante sério [GOLLOM, 2021].

Serão aqui focados os dados referentes à duas estratégias de combate à pandemia: o isolamento social, medido através de decretação de medidas por governos federal/estaduais/municipais; e a vacinação, iniciada em outros países no final de 2020 e no Brasil em janeiro de 2021. Nossa atenção estará mais dedicada à primeira.

Para não haver dispersão, o foco será a evolução da pandemia contra as medidas adotadas, em âmbito do Estado de São Paulo, que oferece uma massa de dados suficientemente grande e heterogênea. Entretanto, serão também avaliados números em nível consolidado do Brasil e de outros países, sempre com a finalidade de comparação.

1.4 CORRELAÇÃO E CAUSALIDADE

Neste texto, temos uma missão bastante importante, que é guardar a distinção entre CORRELAÇÃO (duas grandezas que variam com um grau mensurável de semelhança) e CAUSALIDADE (uma grandeza é a causa da variação da outra). Este tema é bastante profundo e extenso, e já motivou uma bibliografia bastante rica, da qual destacamos (PEARL, MACKENZIE, 2019). Já foram publicados estudos, com alto grau de seriedade, para mostrar que é possível encontrar no mundo as correlações mais espúrias possíveis, ligando medidas sem absolutamente a menor relação de causa e efeito. Neste presente trabalho, por outro lado, não colocamos em dúvida a relação entre medidas de combate a pandemia e seus efeitos, devidamente comprovados por inúmeros trabalhos já clássicos em epidemiologia. Assim, colocamos ênfase na correlação que se pode medir através de análise exploratória e estatística dos dados.

Conforme definido no site Data Hero¹:

“Se correlação não significa causalidade, então o que significa? Uma correlação entre a variável A e a variável B pode significar algumas coisas:

1. *A causa B*
2. *B causa A*
3. *Pura coincidência*
4. *A e B estão conectados pela mesma causa*
5. *Tanto A quanto B contribuem um para o outro em um loop”*

Sabemos que a variação positiva (aumento) na mobilidade aumenta a circulação do vírus, por aumentar o contato interpessoal. E, portanto, na população, vai aumentar o número de casos (partimos da premissa que a grande maioria das pessoas, uma vez em contato com o vírus, vai desenvolver um conjunto mínimo de sintomas que levará a notificação (e pode evoluir para diversos estágios de gravidade, desde um leve mal-estar até o óbito). Ou seja, existe uma relação de CAUSALIDADE entre MOBILIDADE e CASOS, conforme destacam os trabalhos de [KARTAL et al, 2021] e [BERGMAN et al, 2020]. Nossa análise será exploratória e visual, sem a profundidade dessas duas referências.

Portanto, aumento de mobilidade causando aumento de casos é a nossa premissa, recaindo no terreno da epidemiologia. Faremos uma análise comparativa da mobilidade com o número de casos e de que forma variam juntos, seja quantitativamente (via correlação), seja qualitativamente (via gráficos).

Nosso estudo consiste em análises essencialmente exploratórias; olharemos o aspecto dos dados e citaremos, por exemplo, as inter-relações entre as categorias de mobilidade e como influenciam a relação entre mobilidade e casos, sem entrarmos no mérito de como todo esse sistema de inter-relacionamentos funciona, o que nos levaria a estudos além de nosso escopo.

¹ <https://datahero.com/blog/2013/11/21/if-correlation-isnt-causation-what-is-it/> - traduzido pelo autor deste trabalho

1.5 A PRODUÇÃO CIENTÍFICA EM CIÊNCIA DE DADOS

Ciência de dados é um termo de uso relativamente recente, surgida na década passada, para denominar um corpo de conhecimento resultante da evolução, consolidação e amadurecimento de um conjunto de tecnologias e metodologias de tratamento da informação, visando extrair dela valor, para aplicação nas mais diversas áreas do conhecimento humano.

Disciplinas que até alguns anos atrás viviam restritas às salas de aula dos cursos universitários de ciências exatas, como a estatística, a álgebra linear e o cálculo diferencial e integral cresceram e mostraram sua importância e utilidade na produção de ferramentas sofisticadas para o tratamento de dados. Novas linguagens baseadas nas ciências matemáticas e pacotes de funções e bibliotecas avançadas estenderam esse tratamento muito além dos bancos de dados relacionais, passando a abranger textos, som, imagens estáticas, vídeo, e atingindo a fronteira do que – ainda polemicamente – se considera a inteligência artificial.

A Ciência de Dados pode ser considerada uma grande caixa de ferramentas matemáticas e computacionais, que podemos dispor em nossa mesa imaginária para atuar no cenário que pretendemos modelar e analisar, transformando seus números em informações que apoiem decisões. E para entender da melhor forma possível tal cenário, é preciso reunir informações de diferentes fontes e perspectivas visando entender com maior profundidade o contexto da pandemia e suas possibilidades de estudo.

O fenômeno da pandemia de COVID-19 mobilizou, logo de início, a área de epidemiologia em busca de referências na literatura científica para poder entender o novo vírus que surgia. Porém, tais conhecimentos mostraram-se insuficientes, devido a alta letalidade do vírus e rápida disseminação, e isso demandou esforços concentrados dos especialistas da área, o que levou à produção de uma literatura vasta e de rápida vazão, desde o início de 2020, em termos de obras e reuniões científicas.

Desde o início da pandemia, foi possível constatar a importância que a Ciência de Dados possuía no referido contexto, o que inclui seu vasto conjunto de ferramentas e métodos a oferecer aos especialistas da área médica. Com isso, foi ampliado o escopo de aplicação de tais ferramentas nos trabalhos de pesquisa, o que ficou evidente nos trabalhos multidisciplinares publicados, tendo participação crescente de cientistas da computação, matemáticos e estatísticos. Não é exagero postular que essa participação tem sido vital para o avanço dos

trabalhos em epidemiologia, além da construção de “armas de combate” contra o novo Coronavírus, incluindo o uso da Inteligência Artificial e do Big Data para confeccionar e aprimorar vacinas da Covid-19.

De fato, no campo da criação de imunizantes, o empresário brasileiro Felipe Zmoginski, também gerente no grupo chinês Ali Baba, destaca em sua coluna no Portal UOL (2021) a liderança da China na produção de vacinas e insumos contra o Coronavírus, reforçada por suas pesquisas e avanços na área de Inteligência Artificial. Através de *health techs*, empresas de tecnologia totalmente focadas no segmento da saúde, e forte aporte de capital (*venture funds*), o país asiático está surpreendendo a cada dia alargando os limites possíveis da tecnologia. A China colocou todo o seu potencial tecnológico, científico e educacional a favor da luta contra o novo Coronavírus, enquanto outros países ainda politizam essa luta ao invés de buscar convergência nessa batalha, cunhando termos como “vírus chinês”, originado do negacionismo, racismo e xenofobia de algumas frentes ideológicas no Brasil.

Nossa finalidade no presente trabalho é focar em análises de dados simples, a partir de dados e modelos já conhecidos e divulgados na mídia, deixando fora do escopo técnicas mais complexas e visando responder perguntas fundamentais. Por esse motivo, evitamos o acesso a publicações e trabalhos originários exclusivamente da área médica, salvo algumas exceções, e procuramos obras dentro de nosso escopo, na fronteira entre essa área de aplicação e a área computacional.

Portanto, num momento de crise global, não existe oportunidade melhor para aplicar todo esse conhecimento e tecnologias em todo seu potencial. Desde o rigor na coleta dos dados até o cuidado em seu tratamento e organização, a produção de visões diversas que mostram e explicam tudo o que está oculto naquilo que foi coletado, transformado e entendido tem todo o poder de apoiar conclusões e decisões tomadas com base nelas, tudo com o rigor emprestado da metodologia científica e da matemática. Que os números venham, e que deles emergam seus significados.

2 REVISÃO BIBLIOGRÁFICA

Inicialmente, destacamos a importância do Projeto *SP COVID-19 Info Tracker*, um trabalho permanente de coleta, tratamento, e análise de dados no âmbito do Estado de São Paulo, a partir do qual uma série de índices é criada e avaliada diariamente. O projeto é coordenado por pesquisadores da USP, UNESP e CEPID-FAPESP CeMEAI, a saber, Wallace Casaca, Marilaine Colnago, José Alberto Cuminato, Cassio Oishi e Fabio Amaral – e seus dados e resultados encontram-se disponíveis no site <https://www.spcovid.net.br> – que foi de suma importância para a confecção deste trabalho. O intercâmbio de ideias com o Prof. Casaca, nosso orientador, foi fundamental para trazer para este texto e pesquisa o espírito da Ciência de Dados como ferramenta diária de acompanhamento e luta contra a pandemia.

A partir do Projeto *Info Tracker*, selecionamos dois artigos que podem ser considerados centrais neste trabalho, por apresentarem um escopo comum à nossa pesquisa. Os trabalhos de Colnago, Casaca et al. (2021); e Amaral, Casaca et al. (2021) apresentam importantes resultados regionalizados comparando os números de ocorrências da pandemia (infectados, óbitos, taxas hospitalares e de transmissão) em correlação com as políticas locais, como o *lockdown*, que levaram tantas polêmicas quando implementado. Juntamente com os demais dados e resultados oriundos da plataforma *Info Tracker*, podemos considerar tais textos como ponto de partida para nosso trabalho, que procurará enriquecer os com estudos de escopo ampliado, como o Brasil, América Latina, e Europa como comparativos.

Além das obras acima citadas, o trabalho de Nicolelis et al (2021) pode ser considerado outra importante fonte de referência em termos de inspiração e uso da Ciência de Dados. A ativa presença na mídia de seu principal coautor, Miguel Nicolelis², foi um forte motivador na escolha do tema de pesquisa, e a estrutura deste texto traz, de forma explícita, o uso de

² Miguel Nicolelis é um neurocientista, que desde 2020 tem saído de seu principal campo de atuação para se colocar na linha de frente do combate à pandemia no Brasil, e em particular na região Nordeste do Brasil, mobilizando uma grande rede de pesquisadores e profissionais interessados em contribuir na construção do conhecimento acerca da pandemia.

ferramentas de modelagem estatística (regressão linear múltipla, por exemplo), além de visualizações de mapas ricos em detalhes, todas com referências a seus respectivos códigos fontes. Um texto conciso e organizado, que serve como um modelo para nosso trabalho.

Os trabalhos de (NIELSEN, 2021) e (NOVELLA, 2021), citados no tópico de correlação X causalidade, trazem muitos exemplos dessa importante discussão, reforçando o uso da primeira como ferramenta essencial em experimentos. (KARTAL et al, 2021) e (BERGMAN et al, 2020) comprovam nossa premissa básica, que é a relação de causalidade entre a mobilidade e o número de casos, motivo pelo qual não repetiremos aqui tal comprovação. O primeiro tem como foco a Turquia, e o segundo apresenta as visões consolidadas de alguns países.

O texto de Quoc-Viet Pham et al. (2020) traz um panorama da tecnologia de ponta que está sendo utilizada para combater a pandemia, especialmente metodologias que utilizam a Aprendizagem Profunda (do inglês, *Deep Learning*) que, aliada às ferramentas de Ciência de Dados, tem produzido arquiteturas e modelos inteligentes de grande versatilidade. Embora interessante, não empregaremos aqui as suas diversas, por serem muito amplas e ultrapassarem nosso escopo; mas buscaremos aplicar outros modelos igualmente robustos que deixam abertas possibilidades de estudos futuros com a aplicação de tais modelos mais sofisticados, que são as portas de entrada para os trabalhos em Inteligência Artificial. Importante destacar que tal tecnologia tem se aliado ao campo da visão computacional com extremo sucesso, incluindo aí o diagnóstico automatizado a partir da análise de amostras de tecido pulmonar, para citarmos nossa área de aplicação.

Considerando o contexto de dados e modelagem da COVID-19, a obra de [XU et al, 2021] destaca a importância e a forma de compartilhamento de dados sobre a COVID-19, visando assim acelerar na descoberta do conhecimento sobre a pandemia. Já [BERTAGNA, 2021] apresenta um trabalho de porte e complexidade próximos aos que pretendemos desenvolver neste texto. Seu foco foi a evolução de óbitos pela COVID-19 no âmbito da cidade de Campinas – SP. Suas análises se mostraram importantes a fim de que ajustássemos as nossas, mesmo visando explorar outras regiões geográficas. Em termos de modelagem, o trabalho de [CROKIDAKIS, 2020] contempla uma adaptação de ferramentas matemáticas para explorar os dados da pandemia de COVID-19, ainda na fase inicial em que pouco se conhecia sobre o vírus.

Além da análise de dados e modelagem da COVID-19, há ainda outras referências que serviram de base para nossos estudos. Destacamos o trabalho de [BASELLINI et al, 2021], visto que tal obra também emprega em suas análises o Relatório de Mobilidade do Google e apresenta interessantes visualizações de dados. De modo análogo, [BARRIA-SANDOVAL et al, 2021] foi a referência escolhida como *benchmark* para adicionarmos o Chile como o parâmetro comparativo na América Latina. Embora não tenhamos abordado o cenário na Itália, que foi o primeiro país europeu afetado intensamente pela pandemia, foi bastante útil o estudo de [CARTENI et al, 2020] que construiu um modelo de regressão linear múltipla considerando fatores ambientais, mas com uma representação gráfica próxima à nossa.

Também consideramos importante pontuar as fontes para o ferramental computacional prático de nosso trabalho. A bibliografia disponível sobre Ciência de Dados cresce, a cada dia, numa velocidade absolutamente espantosa. Livros de referência em Ciência da Computação existem desde a década de 60, época em que começou a surgir uma “Ciência dos Dados”, a partir das obras seminais sobre SGBDs (Sistemas Gerenciadores de Bancos de Dados), ainda naqueles anos heroicos nos quais “Computação” indicava poucos computadores de grande porte (mainframes) restritos a algumas universidades e empresas na América do Norte, Europa e Japão.

Passando pelas revoluções do computador pessoal, da internet e do telefone celular, foram surgindo, de forma paralela mas sincronizada, duas trilhas: uma de tecnologias relacionadas a dados, novas formas de armazenar e obtê-los, novas linguagens e paradigmas de programação; e uma trilha teórica, destacando modelos matemáticos, ferramentas estatísticas, e arquiteturas que trouxeram à realidade possibilidades jamais imaginadas, como os veículos autônomos, a geração de imagens humanas totalmente artificiais de qualidade realista, tudo baseado em alicerces matemáticos que revalorizaram a base da matemática avançada, como o Cálculo Diferencial e Integral e a Álgebra Linear, conforme já mencionamos.

No setor de programação em Python, uma obra essencial foi o *Python for Data Analysis* [MC KINNEY, 2018], por ter sido editada por um dos criadores da biblioteca *Pandas* (tratamento de datasets). Um texto breve, *How to Use Python and MissForest Algorithm to Impute Missing Data* [RADECIC, 2020] nos apresentou uma técnica simples, mas muito eficiente de preencher valores ausentes em bases de dados, que utilizamos em diversos pontos de nossas análises.

O material didático produzido pelo corpo docente responsável pelo MBA em Ciência de Dados USP, no qual a elaboração deste texto está incluída, foi de suma importância, especialmente os textos conceituais e os exemplos práticos em formato *Jupyter Notebook*, elaborados pelos professores Luís Gustavo Nonato, Francisco Aparecido Rodrigues, Cibele Russo, Moacir Antonelli Ponti e Cristina Dutra, e que foram sendo disponibilizados aos alunos como texto de aula, exercícios resolvidos/propostos e avaliações.

Complementando as fontes acadêmicas, gostaríamos de citar e destacar o importante papel da imprensa na divulgação de informações confiáveis, essenciais diante do fenômeno dos boatos e notícias falsas espalhadas via aplicativos de mensagem e do negacionismo científico por parte de governantes, políticos e formadores de opinião. Assim, os veículos e periódicos abaixo merecem nosso destaque, que citaremos na seção de referências bibliográficas à medida que forem consultados e integrados a este trabalho:

- Brasil: Folha de S. Paulo, Estado de S. Paulo, G1/O Globo, UOL, Revista Veja.
- EUA: Scientific American, The New York Times, The Washington Post.
- Reino Unido: Guardian, The Times, The Economist.
- França: Le Monde
- Espanha: El País

Finalmente, gostaríamos de destacar como uma importante obra, no quesito motivacional para este trabalho, o livro *THE BOOK OF WHY* [PEARL, MACKENZIE, 2019], que trouxe-nos forte interesse pelo estudo da causalidade na ciência, mostrando a evolução da Estatística e de que forma causas e efeitos – e suas correlações – foram medidas no decorrer dos séculos.

3 METODOLOGIA

3.1 PREMISSAS FUNDAMENTAIS

Partimos do princípio de que todas as bases que contabilizam os óbitos pela COVID-19 tomam como referência a definição formal de mortes causadas pela doença, segundo estabelecido em documento da Organização Mundial da Saúde (2020).

Possibilidades de subnotificação ou supernotificação somente serão levadas em consideração onde explicitamente mencionado, especialmente que possa causar distorções nos números apresentados.

3.2 BUSCA DO OBJETIVO PRIMORDIAL

Inicialmente, procuramos mensurar a correlação dos números de casos x medidas de combate, como a redução de mobilidade, e posteriormente a vacinação. Para tal, temos à disposição três grandes coleções de dados.

- Casos da doença (infectados, óbitos, internações)
- Números da variação de mobilidade
- Números da vacinação.

Com relação às medidas restritivas, será importante medir o grau de adoção das mesmas pela população (medidas de variação da circulação de pessoas), pois é um dado relevante e que pode, inclusive, trazer um panorama local/global das atitudes individuais e coletivas em respeito ao bem comum (o fim da crise sanitária).

Buscamos estabelecer padrões, em particular, utilizando técnicas clássicas de Análise Exploratória de Dados. Em todas as análises que exigiram programação, foi utilizada a linguagem Python, na plataforma Jupyter Notebook. Note que não incluímos aqui os códigos,

que serão reunidos em pasta pública específica na plataforma GITHUB e devidamente referenciados a cada resultado publicado.

Análise exploratória e visualização dos dados utilizados podem nos trazer possíveis casos de correlação. Assim, foram utilizadas ferramentas estatísticas (em especial, correlação linear) de forma a entender suas bases e suas estruturas.

3.3 FONTES DE DADOS

A plataforma da Organização Mundial da Saúde (OMS, ou WHO) foi definida como a fonte oficial de definições e nomenclatura sobre a pandemia, além de fonte confiável de dados. Diversos centros médicos e acadêmicos no mundo dedicaram-se a coletar esses dados e aplicar um tratamento rigoroso, tornando-os prontos para serem consumidos por profissionais de diversas áreas, em especial da mídia, e devidamente divulgados de forma segura para a população, vencendo e anulando as fontes de baixa confiabilidade, oriundas de teorias conspiratórias e entidades que buscam espalhar o pânico e o caos, e com isso colher vantagens políticas e financeiras.

O notável trabalho de Xu [XU et al, 2020] gerou um levantamento sobre o grande número de bases de dados que alimentaram um conjunto de milhares de papers (artigos científicos) na área médica. A grande maioria tratando informações de difícil entendimento para profissionais fora da área médica, ou eventualmente análise e processamento de imagens, o que levaria a um foco totalmente fora deste nosso trabalho.

Importante destacar que percorremos diversas bases de dados, mas nem todas tiveram seus números efetivamente utilizados. Muitas foram simplesmente consultadas, sem maiores desenvolvimentos quando não atendiam a perguntas específicas que procuramos responder.

Tomamos como premissas que tais bases receberam o devido tratamento desde a coleta até sua publicação, no sentido de preenchimento de valores ausentes, formatações de tipos de dados específicos – como datas. Bases de dados brasileiras, que apresentem nomes de municípios e de estados com caracteres especiais (acentuação e cedilha) receberam aqui o tratamento adicional de retirada de tais caracteres e uso de maiúsculas (por exemplo, “Consolação” → “CONSOLACAO”), que entendemos que não atrapalharia a análise e

visualização. Reforçemos que estamos tratando bases disponíveis publicamente e bastante utilizadas por veículos de mídia para divulgação de boletins informativos diários, e por instituições e empresas públicas e privadas com finalidades diversas. Qualquer base ligada a empresas ou pessoas que, por sua vez, estejam envolvidas em divulgação de notícias falsas, foi automaticamente descartada.

Com relação às demais fontes de dados acessadas em nosso estudo, todas disponíveis gratuitamente pela internet, podemos listar (os links estão disponíveis nas referências bibliográficas)

- **Projeto SP COVID-19 INFO TRACKER.** Iniciativa de tratamento e apresentação de dados sobre a pandemia, no âmbito do estado de São Paulo, coordenada e realizada por professores e pesquisadores da Universidade de São Paulo (USP) e Universidade Estadual Paulista (UNESP), com o apoio do CEMEAI - Centro de Ciências Matemáticas Aplicadas à Indústria, a qual já foi citada na introdução deste trabalho.
- **Relatórios de Mobilidade da Comunidade – Google (Google Community Mobility Reports).** Foram escolhidos como ponto de partida de nossas análises, devido à sua ampla cobertura geográfica e por abranger observações de datas referência desde o início da pandemia. Basicamente, a mobilidade é registrada através dos sinais identificados distintamente (mas anonimizados) de dispositivos móveis (celulares, tablets, notebooks, aparelhos de GPS) a partir de datas de referência específicas – e a partir destas, registrados em seis categorias a variação percentual da mobilidade: comércio e recreação; mercearias e farmácias; parques; estações de transporte coletivo; locais de trabalho; áreas residenciais.
- **Our World in Data.** Base de dados globais e análises, não restrita aos números da pandemia, e ligada à Universidade de Oxford, Reino Unido.
- **Johns Hopkins University Coronavirus Resource Center.** A primeira e mais famosa base de dados para a pandemia, já citada na parte inicial deste trabalho, segue apresentando um utilíssimo painel, mostrando a distribuição de casos, óbitos, permitindo abertura por países e por dia, e agora inclui também a curva de vacinação. Permite muitas visualizações diferentes além do célebre painel, que nos primórdios da pandemia era praticamente a única fonte confiável citada pela mídia brasileira e global como referência.

- **SEADE – Sistema Estadual de Análise de Dados Estatísticos do Governo do Estado de São Paulo.** Apresenta muitas bases de dados consolidadas para o estado ou detalhadas por município, e possui também um controle próprio de monitoramento da mobilidade, a partir de um acordo com as operadoras de telefonia celular do estado (Oi, Tim, Vivo e Claro), tendo como apoio a ABR (Associação Brasileira de Recursos em Telecomunicações) e o IPT (Instituto de Pesquisas Tecnológicas). Os dados são devidamente anonimizados, visando proteger a privacidade individual.
- **Biblioteca Virtual do Governo do Estado de São Paulo.** Para informações socioeconômicas sobre os municípios paulistas.
- **IBGE** – para consulta sobre estimativas atualizadas das populações dos municípios.
- **Informações sobre lockdown decretados no Estado de São Paulo.** Construímos uma pequena base com uma amostra dos lockdowns decretados no estado de São Paulo pelas prefeituras de alguns municípios, com base no que foi divulgado pela mídia e pelas próprias prefeituras. A finalidade é ter dados adicionais nos gráficos para verificar o respeito aos lockdowns por parte da população.

3.4 ALGUNS DETALHES DE IMPLEMENTAÇÃO

Como há muitos valores ausentes nas colunas, por causa das falhas de medição das variações de mobilidade. Vamos preencher esses vazios utilizando a técnica *Miss Forest* (preenchimento através da utilização do modelo *Random Forest*), conforme definido em [RADECIC, 2020]. O modelo executa em 1 ou 2 minutos para este volume de dados. Em seguida, adicionamos uma coluna com a média das variações das categorias de mobilidade para cada linha (para termos uma medida única).

A partir das colunas de novos casos e da nova média de categorias, acrescentaremos também as respectivas médias móveis, sempre convencionando uma janela de 7 dias. Esta medida auxilia bastante na visualização de grandezas que variam com o tempo, especialmente aquelas que sofrem bastante oscilação nos valores originais. Só temos que, em seguida, truncar a base, excluindo as 6 linhas iniciais que ficarão com lacunas em tais colunas de médias móveis e que podem distorcer os resultados posteriores se forem preenchidas por qualquer técnica.

Para incluir nos gráficos alguma informação sobre lockdown, foi montada uma base de dados auxiliar contendo município, data, início e fim do lockdown, a partir de uma pesquisa no noticiário de 2020 e 2021 (com fontes registradas numa coluna texto caso necessário).

A geração dos gráficos foi feita seguindo uma série de passos num *procedure* em Python:

1. Seleção do município passado por parâmetro
2. Limitação da data final em 30/06/2021
3. Normalização min-max dos valores
4. Desenho da linha com a média móvel geral de mobilidade (calculada a partir da média simples das 6 categorias)
5. Desenho da linha com a média móvel dos casos (deslocada lateralmente segundo parâmetro, com valor padrão = 0)
6. Desenho das linhas de lockdown, se houver (uma linha longa para o início e outra curta para o final).
7. Desenho do gráfico de correlação entre as médias móveis de casos e mobilidade (com cálculo da correlação linear, reta da correlação e escala de cores refletindo o tempo). Aqui é considerada a correlação entre a mobilidade média e os casos, e exibida a categoria principal de mobilidade, calculada previamente. No grupo de gráficos seguintes mostraremos a matriz de correlações aberta por categoria.

3.5 FERRAMENTAS UTILIZADAS

- **Microsoft Excel (Microsoft Corporation)³** – formato de todas as bases de dados aqui tratadas. Alternamos entre os formatos CSV (comma-separated values) e XLSX (formato padrão do Excel) de acordo com a necessidade ou conveniência. Nossas análises foram sempre efetuadas com bases de pequeno porte (em nenhum momento

³ Microsoft Excel ® For Mac, version 16.53 (21091200), Microsoft 365 subscription, © 2021 Microsoft. All rights reserved.

chegando a um milhão de registros), para evitar tempo excessivo de processamento ou problemas de armazenamento, já que por segurança procuramos guardar todos os arquivos do projeto em plataformas de nuvem, com espaço limitado. Mas o fator principal de utilização foi a facilidade em trabalhar com as bases, além de ser uma ferramenta que já utilizamos há bom tempo. Para efeito de tornar as análises mais práticas, procuramos fazer sempre que possível o tratamento prévio dos dados diretamente nas planilhas, e não de forma programática conforme detalhado abaixo (isso será detalhado na seção de resultados).

- **Ferramentas em plataformas abertas:** Jupyter Notebook e Python. A plataforma Jupyter Notebook é um excelente IDE (Integrated Development Environment – Ambiente integrado de desenvolvimento), e permite desenvolver com facilidade código na linguagem de programação adotada neste projeto e no curso de MBA em Ciência de Dados da USP, que é o Python (versão 3.7). Padrão em ciência e engenharia de dados já há alguns anos, o Python apresenta facilidade de aprendizado e mesmo programadores novatos evoluem com muita rapidez. A plataforma Jupyter permite reunir, num único documento (o notebook, ou formato ipynb) código, texto de documentação e ilustrações. Conforme já mencionado, reproduziremos neste documento os resultados e comentários correspondentes das análises de dados, sem incluir os códigos que as geraram – que manteremos e referenciaremos na plataforma GITHUB. Tratamentos prévios dos dados foram feitos no próprio Excel, como mencionamos acima. No código Python, utilizamos diversos pacotes e bibliotecas centrais, como numpy, scikit-learn, e com o tratamento de dados feito pelo pacote pandas; a eles adicionaremos pacotes de análise estatística e visualização.
- **GITHUB** – plataforma global padrão para armazenamento e compartilhamento de desenvolvimento de software, permite que desenvolvedores do mundo inteiro tornem públicos seus trabalhos e os disponibilizem para serem reutilizados ou até mesmo modificados. Aqui manteremos os notebooks Jupyter com nossos códigos para consulta, já que não os incluiremos integralmente neste documento principal.
- **TABLEAU DESKTOP⁴** – aplicativo considerado atualmente o estado da arte em termos de visualização de dados. Para focarmos mais em conclusões específicas sobre

⁴ Tableau Desktop – Professional Edition, 2021.1.5 (20211.21.0819.1914) 64-bit. © 2021 Tableau Software, LLC and its licensors – All rights reserved.

os dados, decidimos delegar o trabalho de visualizações mais complexas/sofisticadas a ferramentas especializadas. O Tableau permite, com facilidade, importar os dados em formato tabular e manipular as visualizações até a obtenção de resultados visualmente úteis e interessantes, incluindo até mesmo mapas geográficos. Permite construir até mesmo painéis com visualizações diferentes (dashboards) e representar os fluxos de dados como stories.

Uma observação sobre as ferramentas de visualização: embora a maioria das visualizações gráficas tenha sido feita diretamente em Python a partir do tratamento e formatação dos dados, em alguns casos delegamos a geração de visualizações mais avançadas para as ferramentas relacionadas acima, que permitiram um ganho de tempo e uma variedade maior de gráficos.

3.6 ESCOPO GEOGRÁFICO

Para evitar dispersão em nosso foco, ou comparações entre cenários por demais diversos, procuramos limitar nosso escopo geográfico principal ao Estado de São Paulo, especialmente por ter o monitoramento rigoroso e atento por parte de iniciativas acadêmicas como o Projeto Info Tracker. Embora exista sempre certa precariedade em dados fornecidos pelo poder público (aqui no caso a Secretaria Estadual da Saúde e o SEADE), é aqui em nosso estado que tais problemas menos comprometem sua confiabilidade.

- Cidades paulistas que adotaram o lockdown em diversos graus (Araraquara, Batatais, outras). Em nossa análise, escolhemos uma amostra delas, segundo seu destaque no noticiário.
- Cidades paulistas com plano de vacinação total para efeitos de pesquisa (Serrana, Botucatu)
- Cidades paulistas que adotaram o mínimo ou nenhuma medida (Mirandópolis, outras).
- Cenário da capital do estado
- Cenário geral do estado de São Paulo
- Cenário geral do Brasil

- Cenário geral de um benchmark na América Latina – Chile. O país foi escolhido por suas características econômicas bastante distintas do restante do continente, as quais, porém, não impediram que fosse atingido com muita força e sofresse forte contração em suas atividades, segundo [BARRIA-SANDOVAL et al, 2021].
- Cenário geral de um benchmark global – Reino Unido

Conforme citado na conclusão do estudo de [BARRIA-SANDOVAL et al, 2021], que foi totalmente focado no Chile, é difícil a comparação da situação entre países distintos, cada um com suas características biopsicossociais próprias. Por essa razão, incluímos os dois países acima citados apenas para termos dois benchmarks internacionais, um mais próximo (América Latina), o outro mais distante (Europa); destacamos que ambos os países tiveram fortes dificuldades e também suas vitórias nessa verdadeira guerra; assumindo nosso foco principal como o estado de São Paulo, a comparação não fica tão imprópria, desde que desconsideremos a forte atuação do governo federal do Brasil no sentido de atrapalhar propositalmente esse combate, visando objetivos político-eleitorais.

3.7 ESCOPO NO TEMPO

Nossas extrações de dados para as análises realizadas sobre a mobilidade no estado de São Paulo foram feitas no intervalo desde abril de 2020 até o meio do ano de 2021, quando passou a haver influência da vacinação, iniciada na capital do estado em janeiro; também devido à vacinação limitamos o estudo para os outros dois países (Chile e Reino Unido) ao final de 2020, devido ao início ágil das campanhas neste último país ainda em dezembro/2020, o que já influenciaria os números da pandemia no primeiro semestre de 2021. Como focamos na comparação entre a mobilidade e os casos, então de fato precisávamos escolher os períodos em que a vacinação ainda era incipiente. Nossas análises sobre a vacinação já selecionaram intervalos mais avançados no ano de 2021.

3.8 DISTORÇÕES NAS BASES DE DADOS

Embora, desde o início da pandemia, fortes distorções tenham sido registradas nos dados – geralmente no sentido de subnotificar os casos e óbitos – não entraremos nesse mérito, pois tais anomalias são, por definição, complexas de serem investigadas e verificadas. Elas podem ocorrer também por alterações nos critérios ou metodologias de coleta e registro de dados. Mas deixemos registrado aqui que o elevado número de óbitos, comprehensivelmente causador de forte comoção na sociedade, especialmente porque poderia ter sido evitado em grande parte, possivelmente foi sensivelmente maior do que o oficialmente registrado – de início pela precariedade nos diagnósticos, e posteriormente por influência política regional ou federal.

3.9 MODELOS COMPUTACIONAIS

A análise exploratória será feita basicamente a partir da geração de visualizações em Python ou na ferramenta Tableau®. Os dados coletados, embora venham de fontes já devidamente tratadas para consumo, receberam algum tratamento adicional seguindo uma série de modelos apresentados no curso de MBA em Ciência de dados.

- Limpeza e organização de dados: embora trabalhemos com dados previamente preparados e curados, eles poderão apresentar lacunas, devido a falhas de medição ou problemas de formatação que podem chegar até as ferramentas visuais e comprometer seu funcionamento. Para o preenchimento de lacunas, utilizamos o *MISS FOREST*, baseado num modelo clássico de aprendizagem de máquina, a floresta aleatória, para o preenchimento de valores ausentes [RADECIC, 2020].
- Trabalhamos exclusivamente com dados estruturados, em formato tabular, a partir dos quais poderemos gerar visualizações gráficas que no ajudem a entendê-los.
- O tratamento de dataframes, por considerar massas de dados de pequeno porte e fora de ambiente corporativo, sem necessidade de acesso paralelo, limitou-se às ferramentas do pacote pandas na linguagem Python.
- A normalização max-min foi aplicada para uniformizar as escalas nos gráficos.
- Cálculos envolvendo correlação utilizam a função corr () do pacote numpy.
- A matriz de correlações, uma ferramenta de visualização bastante útil, foi acessada no pacote gráfico seaborn.

4 RESULTADOS OBTIDOS

Nossa discussão terá como foco principal a comparação entre a variação de mobilidade, conforme registrada pelos Relatórios de Mobilidade do Google, e a variação de casos da doença. Em alguns pontos da análise também incluiremos a variação de óbitos. Como complemento, analisaremos a comparação entre casos/óbitos e os números da vacinação, abrangendo os primeiros meses com os seus efeitos.

Apesar de sua importância, não analisaremos os dados da variante ômicron, por fatores de prazo deste trabalho – ela ganhou destaque agora na segunda quinzena de dezembro de 2021. Seus efeitos se estenderão pelos primeiros meses de 2022, e uma análise apurada poderá ser feita mais adiante.

4.1 ANÁLISE DOS DADOS DE MOBILIDADE DO GOOGLE PARA O ESTADO DE SÃO PAULO

Vamos inicialmente analisar a base de mobilidade do Google. Extraímos um relatório do site <https://www.google.com/covid19/mobility/> no dia 14/11/2021, contendo somente registros para o Brasil.

Nesta seção, a proposta é discutir os resultados tomando o trabalho de Bertagna (2021) como linha de base. Nesse sentido, diversos fatores serão citados em nosso levantamento, cuja abrangência geográfica coincide, em parte, com o trabalho acima supracitado.

O tratamento inicial de dados envolve o download e o tratamento do Relatório de Mobilidade do Google (data: 14/11/2021, contemplando os anos de 2020 e 2021) para o Brasil.

TABELA 2 Print parcial do Google Mobility Report após os devidos acertos manuais prévios

Estado	Município	data	comercio_recreacao	alimentacao_farmacia	parkes	estacoes_transporte	localis_trabalho	residencias
00 CONSOLIDADO BRASIL	00 CONSOLIDADO BRASIL	2021-01-01	-45	-2	-2	-37	-71	15
01 CONSOLIDADO BRASIL	00 CONSOLIDADO BRASIL	2021-01-02	-45	-2	-2	-37	-39	11
02 CONSOLIDADO BRASIL	00 CONSOLIDADO BRASIL	2021-01-03	-37	-1	-13	-20	-1	6
03 CONSOLIDADO BRASIL	00 CONSOLIDADO BRASIL	2021-01-04	-23	18	-14	-19	-10	8
04 CONSOLIDADO BRASIL	00 CONSOLIDADO BRASIL	2021-01-05	-22	18	-14	-18	-12	8
05 CONSOLIDADO BRASIL	00 CONSOLIDADO BRASIL	2021-01-06	-21	15	-9	-18	-10	9
06 CONSOLIDADO BRASIL	00 CONSOLIDADO BRASIL	2021-01-07	-20	21	-5	-19	-9	8
07 CONSOLIDADO BRASIL	00 CONSOLIDADO BRASIL	2021-01-08	-18	23	-9	-14	-6	8
08 CONSOLIDADO BRASIL	00 CONSOLIDADO BRASIL	2021-01-09	-20	19	-14	-12	5	6
09 CONSOLIDADO BRASIL	00 CONSOLIDADO BRASIL	2021-01-10	-27	10	-16	-18	6	6
10 CONSOLIDADO BRASIL	00 CONSOLIDADO BRASIL	2021-01-11	-19	17	-20	-16	-5	7
11 CONSOLIDADO BRASIL	00 CONSOLIDADO BRASIL	2021-01-12	-23	12	-21	-21	-10	7
12 CONSOLIDADO BRASIL	00 CONSOLIDADO BRASIL	2021-01-13	-24	6	-20	-23	-10	7
13 CONSOLIDADO BRASIL	00 CONSOLIDADO BRASIL	2021-01-14	-27	5	-20	-24	-13	6
14 CONSOLIDADO BRASIL	00 CONSOLIDADO BRASIL	2021-01-15	-28	4	-24	-24	-12	6
15 CONSOLIDADO BRASIL	00 CONSOLIDADO BRASIL	2021-01-16	-29	4	-25	-23	-4	5
16 CONSOLIDADO BRASIL	00 CONSOLIDADO BRASIL	2021-01-17	-34	-2	-25	-24	-4	5
17 CONSOLIDADO BRASIL	00 CONSOLIDADO BRASIL	2021-01-18	-29	0	-30	-24	-11	7
18 CONSOLIDADO BRASIL	00 CONSOLIDADO BRASIL	2021-01-19	-29	2	-30	-25	-14	7
19 CONSOLIDADO BRASIL	00 CONSOLIDADO BRASIL	2021-01-20	-27	3	-21	-26	-15	8
20 CONSOLIDADO BRASIL	00 CONSOLIDADO BRASIL	2021-01-21	-27	6	-22	-23	-11	7
21 CONSOLIDADO BRASIL	00 CONSOLIDADO BRASIL	2021-01-22	-26	7	-23	-20	-9	7
22 CONSOLIDADO BRASIL	00 CONSOLIDADO BRASIL	2021-01-23	-28	7	-24	-19	-1	6
23 CONSOLIDADO BRASIL	00 CONSOLIDADO BRASIL	2021-01-24	-33	0	-26	-24	-1	6
24 CONSOLIDADO BRASIL	00 CONSOLIDADO BRASIL	2021-01-25	-30	2	-27	-30	-13	8
25 CONSOLIDADO BRASIL	00 CONSOLIDADO BRASIL	2021-01-26	-30	2	-30	-23	-11	7
26 CONSOLIDADO BRASIL	00 CONSOLIDADO BRASIL	2021-01-27	-28	2	-23	-23	-10	8
27 CONSOLIDADO BRASIL	00 CONSOLIDADO BRASIL	2021-01-28	-28	6	-22	-22	-9	7
28 CONSOLIDADO BRASIL	00 CONSOLIDADO BRASIL	2021-01-29	-24	13	-27	-18	-6	6
29 CONSOLIDADO BRASIL	00 CONSOLIDADO BRASIL	2021-01-30	-31	10	-25	-20	-2	6
30 CONSOLIDADO BRASIL	00 CONSOLIDADO BRASIL	2021-01-31	-36	5	-25	-24	-1	5
31 CONSOLIDADO BRASIL	00 CONSOLIDADO BRASIL	2021-02-01	-26	9	-30	-20	-6	6
32 CONSOLIDADO BRASIL	00 CONSOLIDADO BRASIL	2021-02-02	-27	7	-27	-21	-11	6
33 CONSOLIDADO BRASIL	00 CONSOLIDADO BRASIL	2021-02-03	-26	5	-24	-21	-8	6
34 CONSOLIDADO BRASIL	00 CONSOLIDADO BRASIL	2021-02-04	-26	8	-23	-20	-4	6
35 CONSOLIDADO BRASIL	00 CONSOLIDADO BRASIL	2021-02-05	-22	15	-29	-17	-6	5
36 CONSOLIDADO BRASIL	00 CONSOLIDADO BRASIL	2021-02-06	-23	16	-30	-16	0	5
37 CONSOLIDADO BRASIL	00 CONSOLIDADO BRASIL	2021-02-07	-32	6	-32	-24	-1	5
38 CONSOLIDADO BRASIL	00 CONSOLIDADO BRASIL	2021-02-08	-24	10	-33	-17	-5	5
39 CONSOLIDADO BRASIL	00 CONSOLIDADO BRASIL	2021-02-09	-23	11	-32	-17	-8	6
40 CONSOLIDADO BRASIL	00 CONSOLIDADO BRASIL	2021-02-10	-22	9	-26	-17	-6	6
41 CONSOLIDADO BRASIL	00 CONSOLIDADO BRASIL	2021-02-11	-24	11	-28	-18	-7	6
42 CONSOLIDADO BRASIL	00 CONSOLIDADO BRASIL	2021-02-12	-22	15	-31	-16	-4	5

Fonte: <https://www.google.com/covid19/mobility/>

A segunda base de dados aqui utilizada: Leitura do arquivo de casos para o Estado de São Paulo, obtido a partir do repositório de dados do SEADE (Fundação Sistema Estadual de Análise de Dados - Governo de São Paulo), em: <https://github.com/seade-R/dados-covid-sp> . Da mesma forma como feito acima com os dados do Google, a base recebeu uma série de tratamentos e formatações, para ser conectada àquela.

Para gerar os gráficos conforme descrito acima, agrupamos os municípios em amostras segundo critérios de crescimento nos casos (conforme site do projeto Info-Tracker) e densidade populacional, conforme posição de 15/08/2021 na base de dados do SEADE – SP. Dentro de cada amostra escolhemos os municípios mais significativos.

Cada figura consiste em duas visões: uma comparação das curvas de casos e mobilidade no tempo, e a correlação entre mobilidade e casos, onde a mobilidade considerada é o valor médio (média móvel 7) das categorias. A evolução no tempo deve ser observada com atenção – ela assume um perfil distinto para cada cidade – mas é importante examinar o padrão de subidas e quedas de casos novos em comparação com o aumento e diminuição da mobilidade. No segundo gráfico de cada cidade, o de correlação, essa medida é obtida independente do tempo; nele, encontramos padrões e agrupamentos dos mais variados, além de uma correlação positiva entre mobilidade e os casos.

O código fonte que gerou os gráficos pode ser encontrado em: [https://github.com/rlepsch/MBA-USP-TCC-2021⁵](https://github.com/rlepsch/MBA-USP-TCC-2021) no notebook ANALISE001.

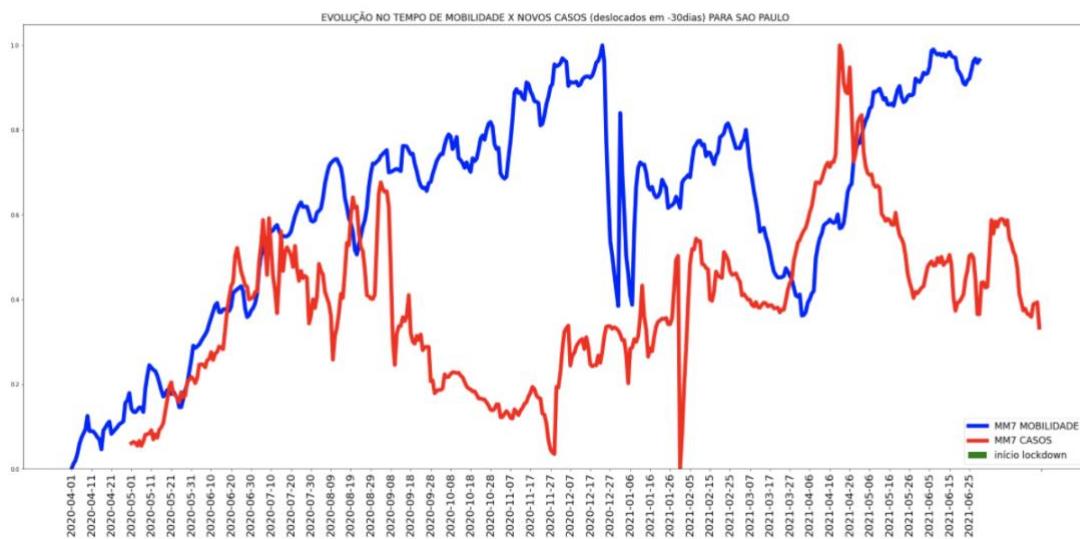
⁵ Os demais arquivos utilizados nesta análise também foram guardados nesse repositório, exceto aqueles com tamanho superior a 25 megabytes, limitação do próprio Github

EVOLUÇÃO NO TEMPO DE MOBILIDADE X NOVOS CASOS

**Capital do Estado de São Paulo, e cidades com alto crescimento no número de casos:
BOTUCATU, BARRETOS, SÃO CARLOS, PAULINIA.**

FIGURA 3 - Comparação de evolução de casos e mobilidade para São Paulo

 ANÁLISE DE MOBILIDADE (RELATÓRIO GOOGLE) X NOVOS CASOS DE INFECÇÃO POR COVID - 19 (BASE SEADE-SP) NO ESTADO DE SÃO PAULO, PARA O MUNICÍPIO DE SAO PAULO - período de 2020-04-01 a 2021-06-30



CORRELAÇÃO LINEAR ENTRE AS MÉDIAS MÓVEIS DE VARIAÇÃO DE MOBILIDADE E DE CASOS NOVOS PARA SAO PAULO
 Período de 2020-04-01 a 2021-06-30 : 0.049196925275355304

>>> A escala de cores do gráfico corresponde ao tempo (valores maiores == datas mais recentes) <<<

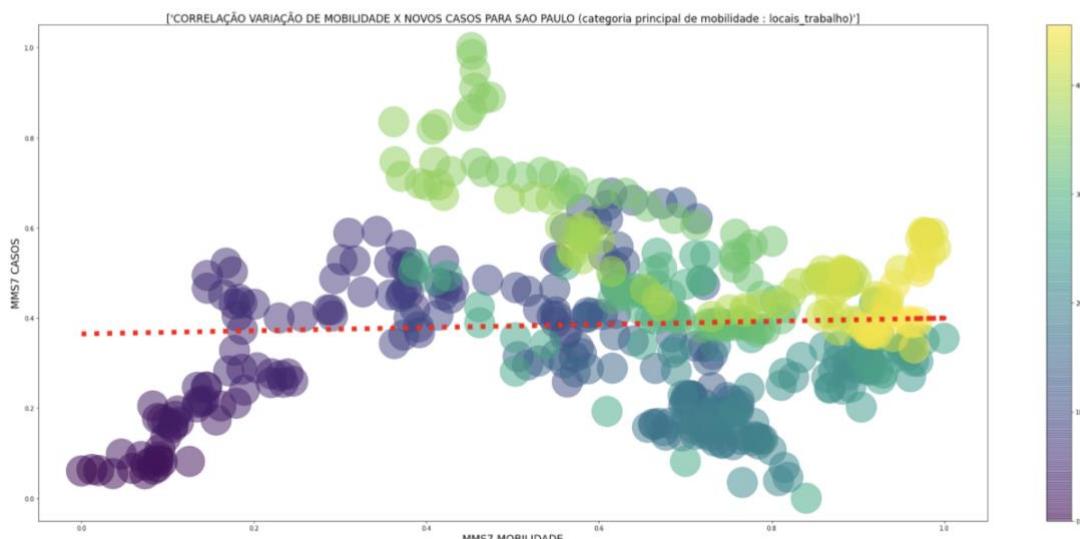
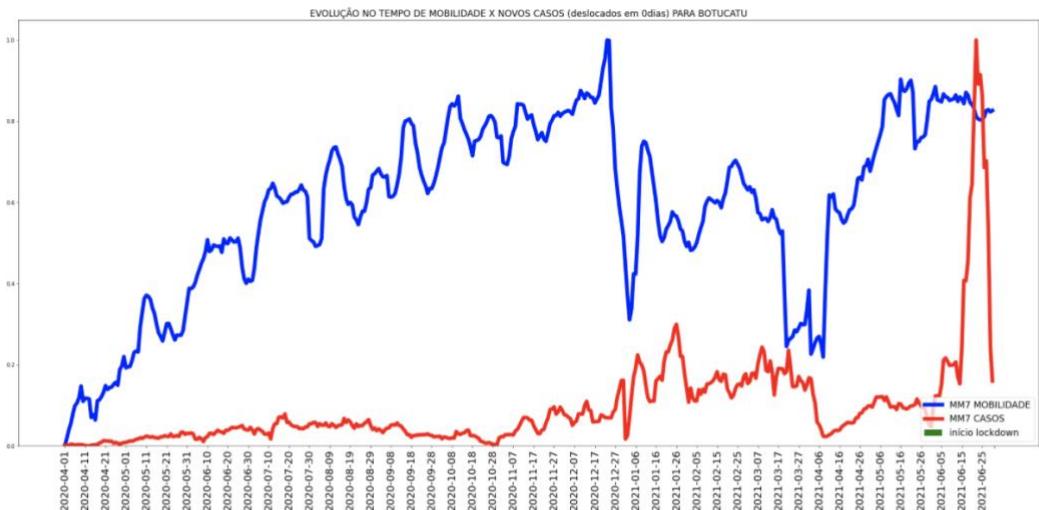


FIGURA 4 - Comparação de evolução de casos e mobilidade para Botucatu

ANÁLISE DE MOBILIDADE (RELATÓRIO GOOGLE) X NOVOS CASOS DE INFECÇÃO POR COVID - 19 (BASE SEADE-SP) NO ESTADO DE SÃO PAULO, PARA O MUNICÍPIO DE BOTUCATU - período de 2020-04-01 a 2021-06-30



CORRELAÇÃO LINEAR ENTRE AS MÉDIAS MÓVEIS DE VARIAÇÃO DE MOBILIDADE E DE CASOS NOVOS PARA BOTUCATU
Período de 2020-04-01 a 2021-06-30 : 0.2521608909268579

>>> A escala de cores do gráfico corresponde ao tempo (valores maiores == datas mais recentes) <<

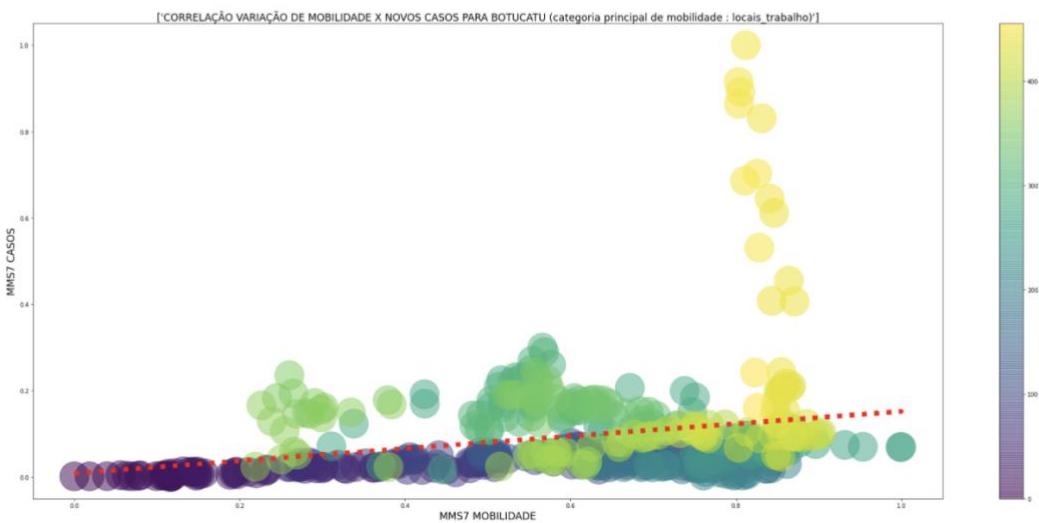
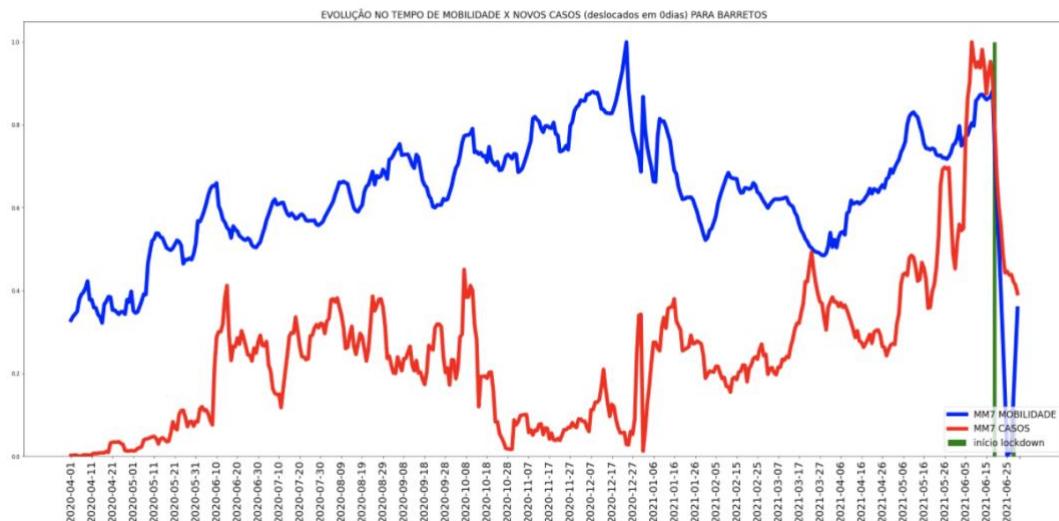


FIGURA 5 - Comparação de evolução e casos e mobilidade para Barretos

 ANÁLISE DE MOBILIDADE (RELATÓRIO GOOGLE) X NOVOS CASOS DE INFECÇÃO POR COVID - 19 (BASE SEADE-SP) NO ESTADO DE SÃO PAULO, PARA O MUNICÍPIO DE BARRETOS - período de 2020-04-01 a 2021-06-30



CORRELAÇÃO LINEAR ENTRE AS MÉDIAS MÓVEIS DE VARIAÇÃO DE MOBILIDADE E DE CASOS NOVOS PARA BARRETOS
 Período de 2020-04-01 a 2021-06-30 : 0.21318611241882446

>>> A escala de cores do gráfico corresponde ao tempo (valores maiores == datas mais recentes) <<<

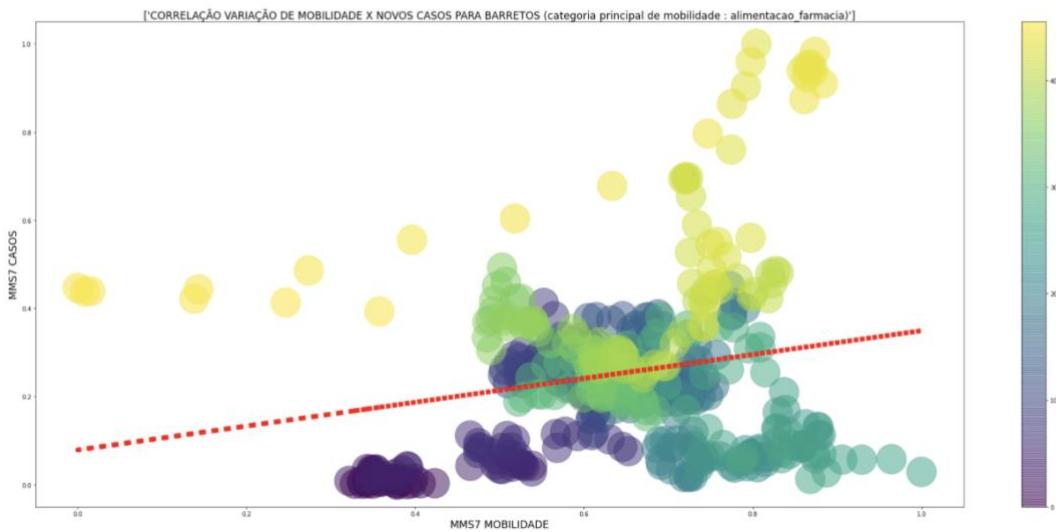


FIGURA 6 - Comparação de evolução de casos e mobilidade para São Carlos

ANÁLISE DE MOBILIDADE (RELATÓRIO GOOGLE) X NOVOS CASOS DE INFECÇÃO POR COVID - 19 (BASE SEADE-SP) NO ESTADO DE SÃO PAULO, PARA O MUNICÍPIO DE SAO CARLOS - periodo de 2020-04-01 a 2021-06-30



CORRELAÇÃO LINEAR ENTRE AS MÉDIAS MÓVEIS DE VARIAÇÃO DE MOBILIDADE E DE CASOS NOVOS PARA SAO CARLOS
Período de 2020-04-01 a 2021-06-30 : 0.208178002209314

>>> A escala de cores do gráfico corresponde ao tempo (valores maiores == datas mais recentes) <<<

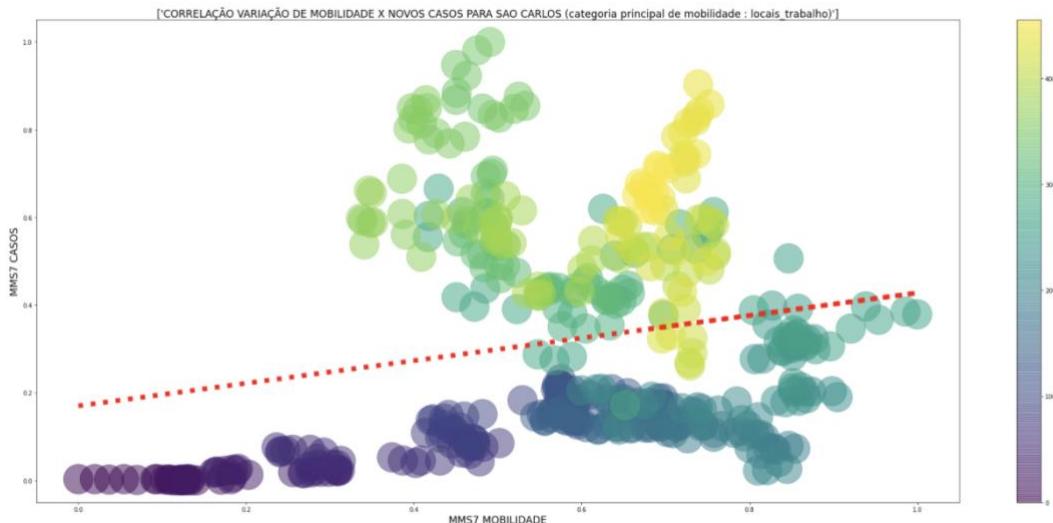
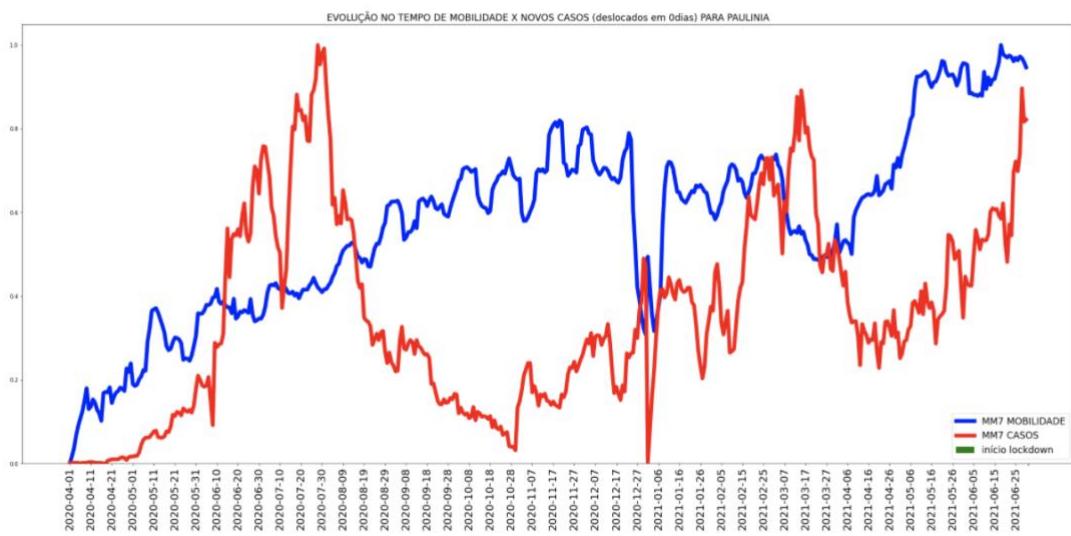


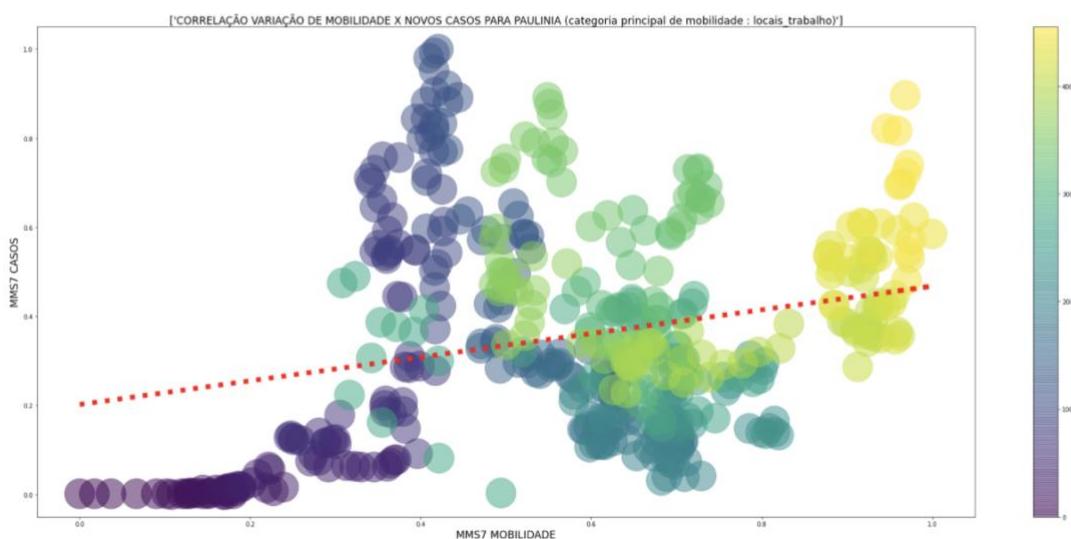
FIGURA 7 - Comparação de evolução de casos e mobilidade para Paulínia

 ANÁLISE DE MOBILIDADE (RELATÓRIO GOOGLE) X NOVOS CASOS DE INFECÇÃO POR COVID - 19 (BASE SEADE-SP) NO ESTADO DE SÃO PAULO, PARA O MUNICÍPIO DE PAULINIA - período de 2020-04-01 a 2021-06-30



CORRELAÇÃO LINEAR ENTRE AS MÉDIAS MÓVEIS DE VARIAÇÃO DE MOBILIDADE E DE CASOS NOVOS PARA PAULINIA
 Período de 2020-04-01 a 2021-06-30 : 0.2432585360330226

>>> A escala de cores do gráfico corresponde ao tempo (valores maiores == datas mais recentes) <<<



Cidades com baixo crescimento de casos: CAMPINAS, RIBEIRÃO PRETO, FRANCA

FIGURA 8 - Comparação de evolução de casos e mobilidade para Campinas

 ANÁLISE DE MOBILIDADE (RELATÓRIO GOOGLE) X NOVOS CASOS DE INFECÇÃO POR COVID - 19 (BASE SEADE-SP) NO ESTADO DE SÃO PAULO, PARA O MUNICÍPIO DE CAMPINAS - período de 2020-04-01 a 2021-06-30



CORRELAÇÃO LINEAR ENTRE AS MÉDIAS MÓVEIS DE VARIAÇÃO DE MOBILIDADE E DE CASOS NOVOS PARA CAMPINAS
 Período de 2020-04-01 a 2021-06-30 : 0.23740889301539786

>>> A escala de cores do gráfico corresponde ao tempo (valores maiores == datas mais recentes) <<<

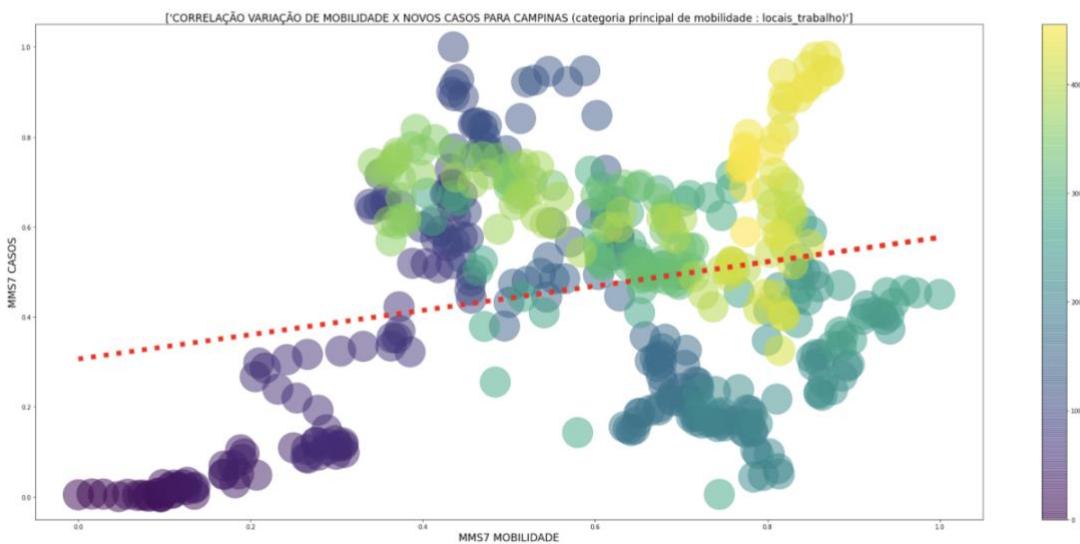
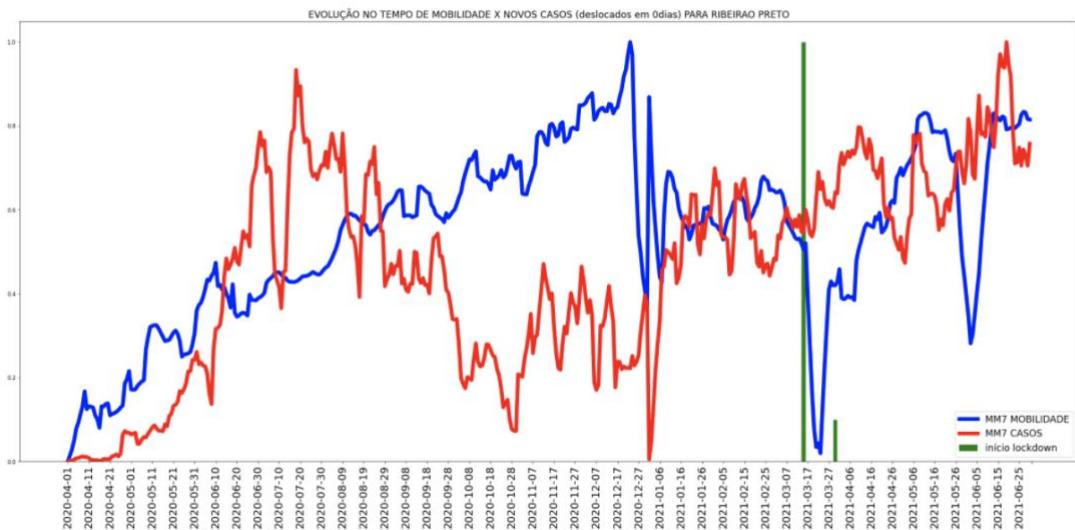


FIGURA 9 - Comparação de evolução de casos e mobilidade para Ribeirão Preto

ANÁLISE DE MOBILIDADE (RELATÓRIO GOOGLE) X NOVOS CASOS DE INFECÇÃO POR COVID - 19 (BASE SEADE-SP) NO ESTADO DE SÃO PAULO, PARA O MUNICÍPIO DE RIBEIRÃO PRETO - período de 2020-04-01 a 2021-06-30



CORRELAÇÃO LINEAR ENTRE AS MÉDIAS MÓVEIS DE VARIAÇÃO DE MOBILIDADE E DE CASOS NOVOS PARA RIBEIRÃO PRETO
Período de 2020-04-01 a 2021-06-30 : 0.2847470032534197

>>> A escala de cores do gráfico corresponde ao tempo (valores maiores == datas mais recentes) <<<

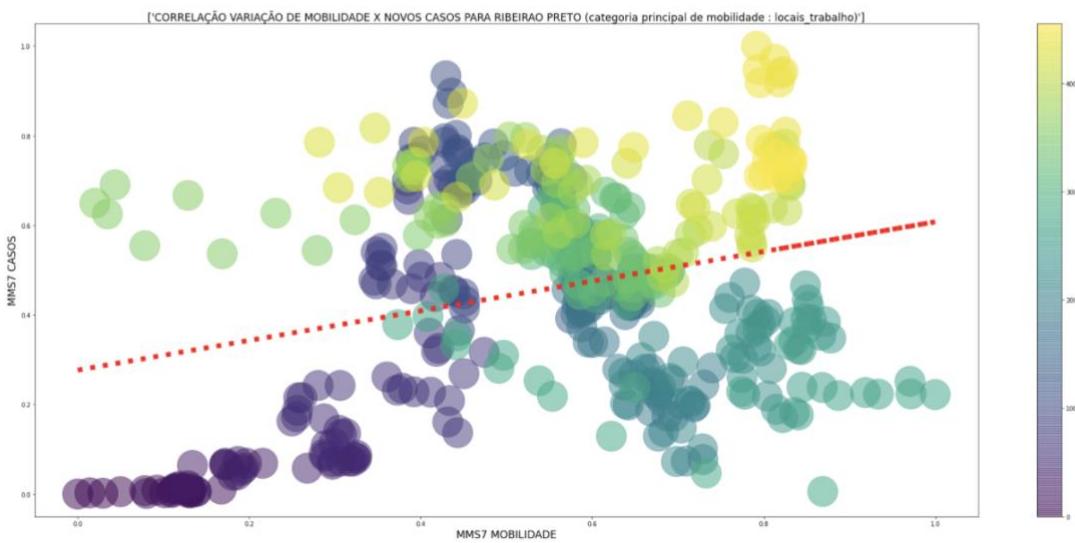


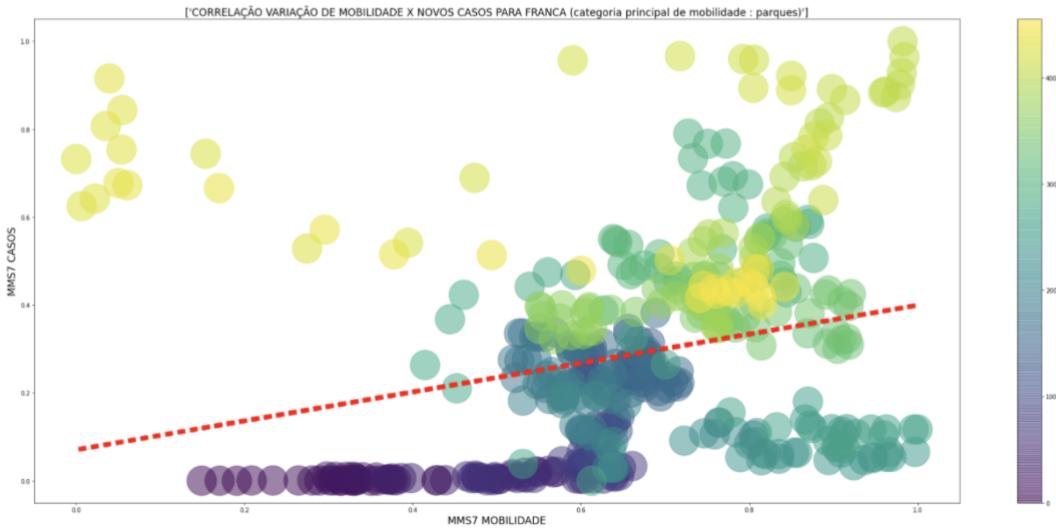
FIGURA 10 - Comparação de evolução de casos e mobilidade para Franca

 ANÁLISE DE MOBILIDADE (RELATÓRIO GOOGLE) X NOVOS CASOS DE INFECÇÃO POR COVID - 19 (BASE SEADE-SP) NO ESTADO DE SÃO PAULO, PARA O MUNICÍPIO DE FRANCA - período de 2020-04-01 a 2021-06-30



CORRELAÇÃO LINEAR ENTRE AS MÉDIAS MÓVEIS DE VARIAÇÃO DE MOBILIDADE E DE CASOS NOVOS PARA FRANCA
 Período de 2020-04-01 a 2021-06-30 : 0.25317363139078175

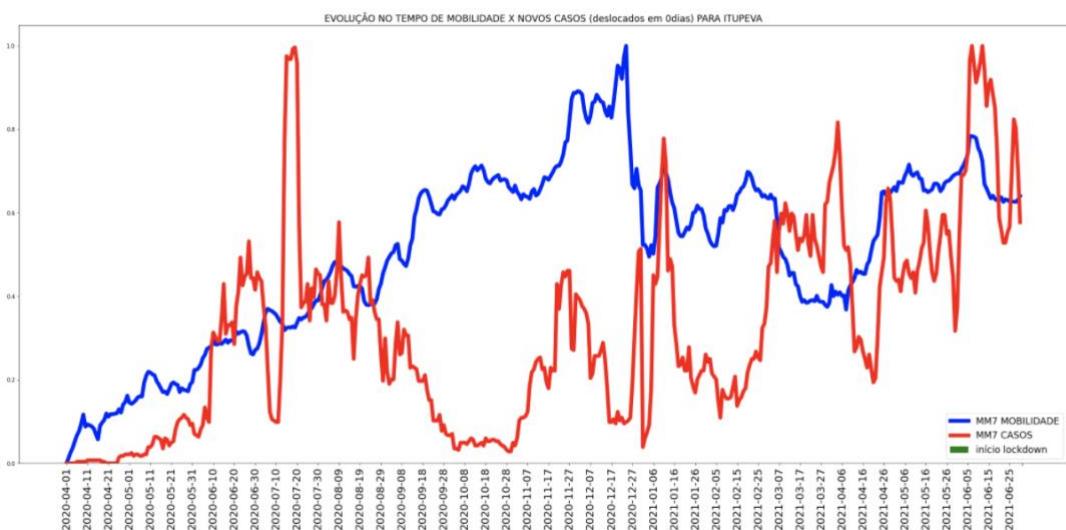
>>> A escala de cores do gráfico corresponde ao tempo (valores maiores == datas mais recentes) <<<



Cidades com alta densidade populacional (acima 9 hab/km²) - ITUPEVA, ITAPEVI, ITAPECERICA DA SERRA

FIGURA 11 - Comparação de evolução de casos e mobilidade para Itupeva

ANÁLISE DE MOBILIDADE (RELATÓRIO GOOGLE) X NOVOS CASOS DE INFECÇÃO POR COVID - 19 (BASE SEADE-SP) NO ESTADO DE SÃO PAULO, PARA O MUNICÍPIO DE ITUPEVA - período de 2020-04-01 a 2021-06-30



CORRELAÇÃO LINEAR ENTRE AS MÉDIAS MÓVEIS DE VARIAÇÃO DE MOBILIDADE E DE CASOS NOVOS PARA ITUPEVA
Período de 2020-04-01 a 2021-06-30 : 0.24150503048065877

>>> A escala de cores do gráfico corresponde ao tempo (valores maiores == datas mais recentes) <<

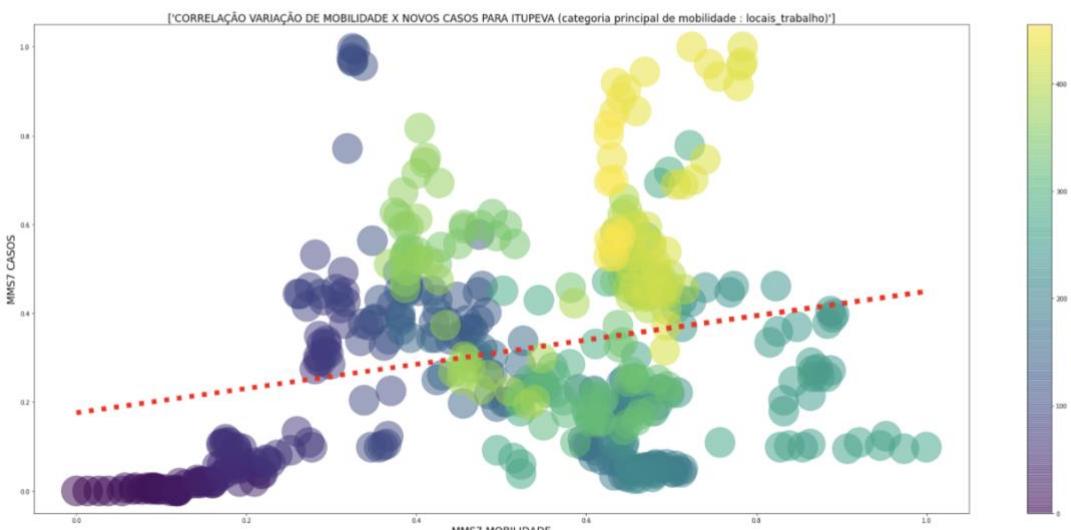
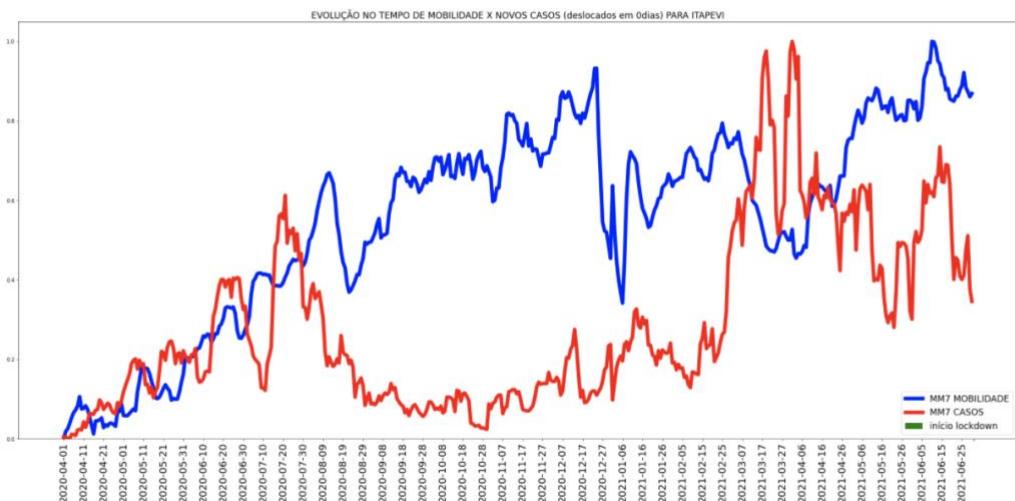


FIGURA 12 - Comparação de evolução de casos e mobilidade para Itapevi

 ANÁLISE DE MOBILIDADE (RELATÓRIO GOOGLE) X NOVOS CASOS DE INFECÇÃO POR COVID - 19 (BASE SEADE-SP) NO ESTADO DE SÃO PAULO, PARA O MUNICÍPIO DE ITAPEVI - período de 2020-04-01 a 2021-06-30



CORRELAÇÃO LINEAR ENTRE AS MÉDIAS MÓVEIS DE VARIAÇÃO DE MOBILIDADE E DE CASOS NOVOS PARA ITAPEVI
 Período de 2020-04-01 a 2021-06-30 : 0.24597257394062408

>>> A escala de cores do gráfico corresponde ao tempo (valores maiores == datas mais recentes) <<<

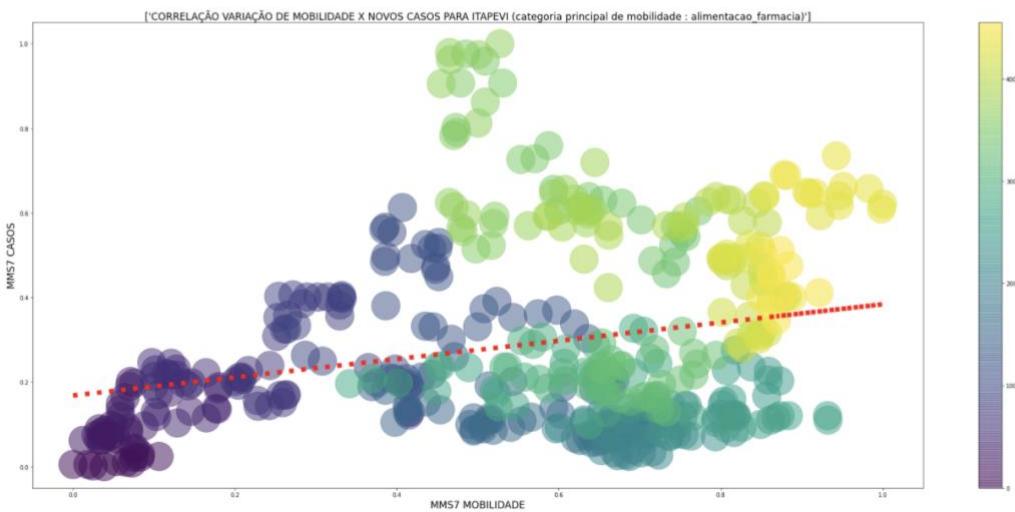
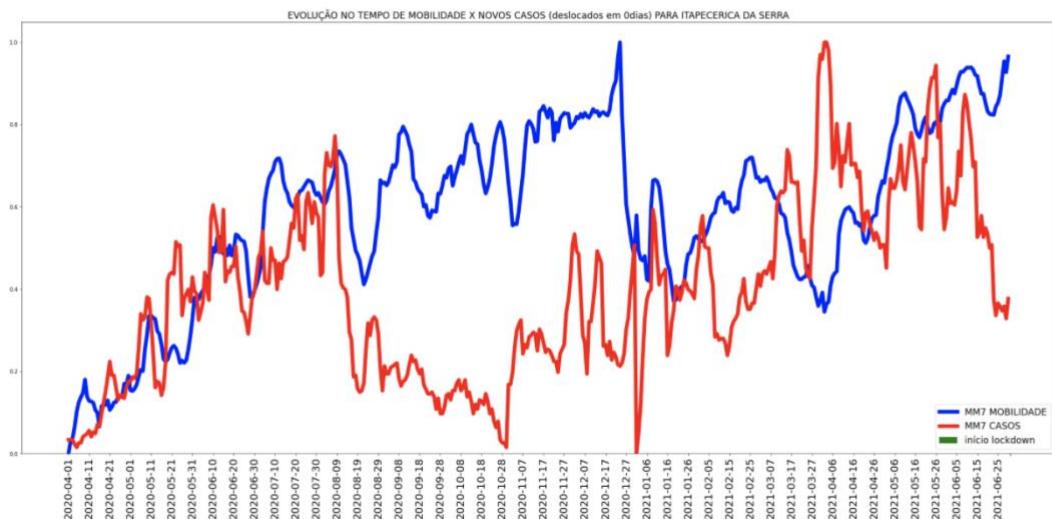


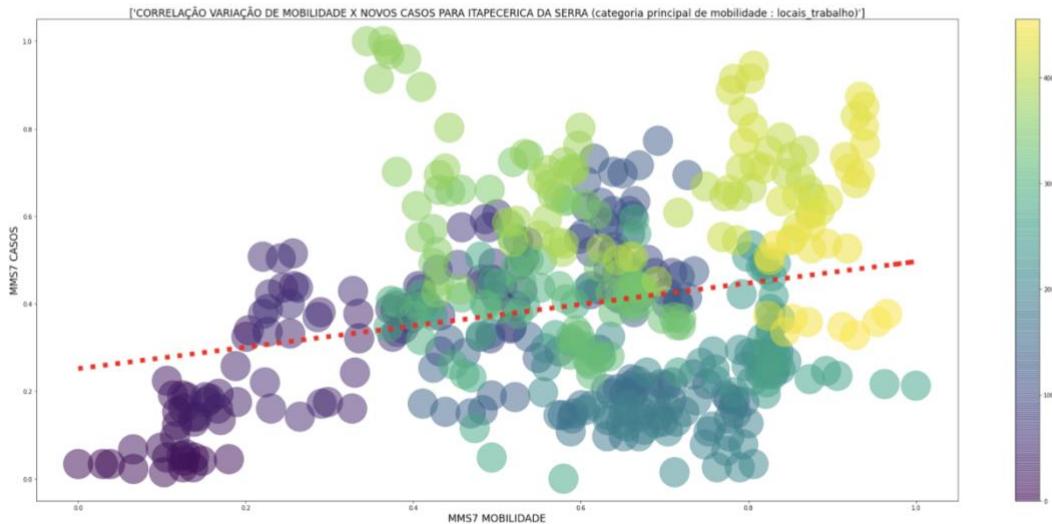
FIGURA 13 - Comparação de evolução de casos e mobilidade para Itapecerica da Serra

 ANÁLISE DE MOBILIDADE (RELATÓRIO GOOGLE) X NOVOS CASOS DE INFECÇÃO POR COVID - 19 (BASE SEADE-SP) NO ESTADO DE SÃO PAULO, PARA O MUNICÍPIO DE ITAPECERICA DA SERRA - período de 2020-04-01 a 2021-06-30



CORRELAÇÃO LINEAR ENTRE AS MÉDIAS MÓVEIS DE VARIAÇÃO DE MOBILIDADE E DE CASOS NOVOS PARA ITAPECERICA DA SERRA
 Período de 2020-04-01 a 2021-06-30 : 0.24418901099341445

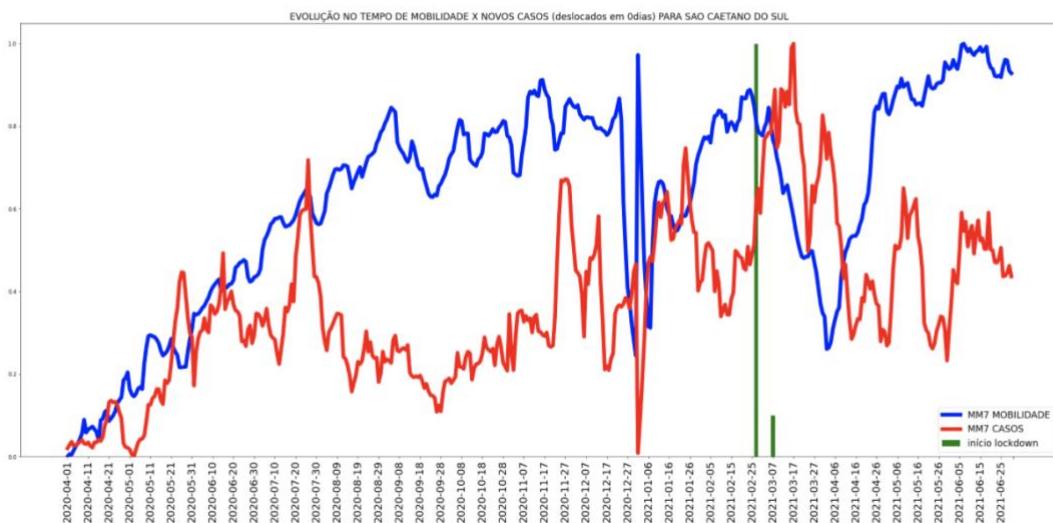
>>> A escala de cores do gráfico corresponde ao tempo (valores maiores == datas mais recentes) <<<



Cidades com baixa densidade populacional (abaixo 9 hab/km²) – SÃO CAETANO DO SUL, JALES, SERRA NEGRA

FIGURA 14 - Comparação de evolução de casos e mobilidade para São Caetano do Sul

 ANÁLISE DE MOBILIDADE (RELATÓRIO GOOGLE) X NOVOS CASOS DE INFECÇÃO POR COVID - 19 (BASE SEADE-SP) NO ESTADO DE SÃO PAULO, PARA O MUNICÍPIO DE SAO CAETANO DO SUL – período de 2020-04-01 a 2021-06-30



CORRELAÇÃO LINEAR ENTRE AS MÉDIAS MÓVEIS DE VARIAÇÃO DE MOBILIDADE E DE CASOS NOVOS PARA SAO CAETANO DO SUL
 Período de 2020-04-01 a 2021-06-30 : 0.3384403359263514

>>> A escala de cores do gráfico corresponde ao tempo (valores maiores == datas mais recentes) <<<

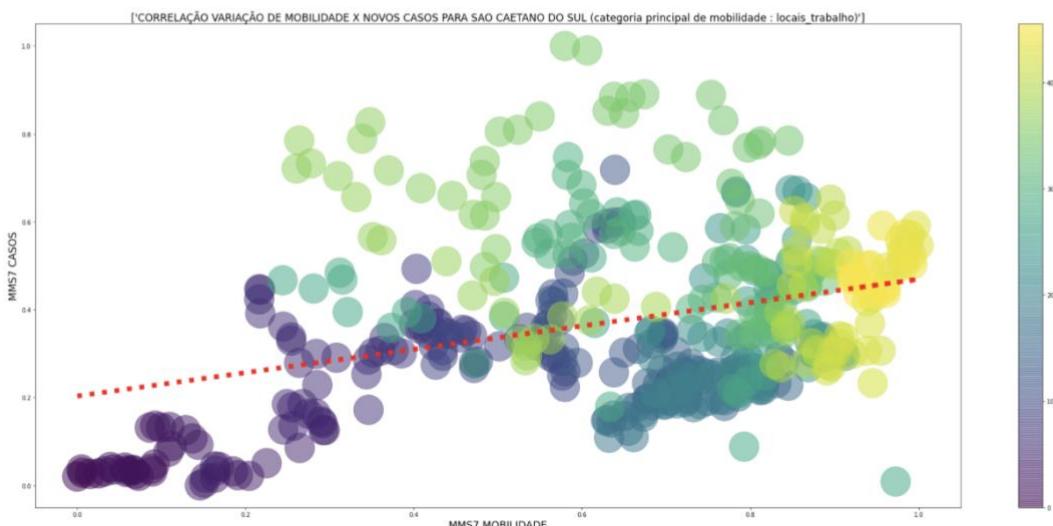
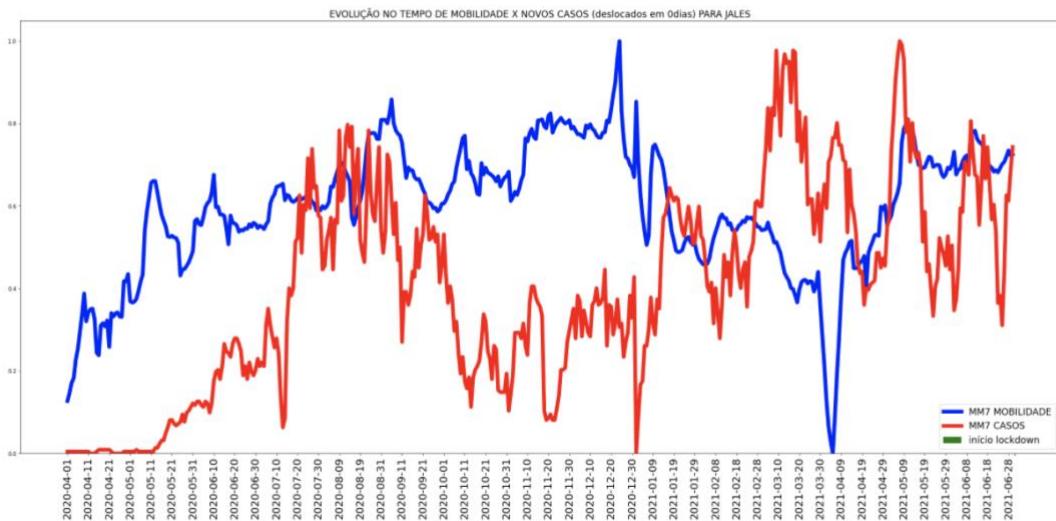


FIGURA 15 - Comparação de evolução de casos e mobilidade para Jales

 ANÁLISE DE MOBILIDADE (RELATÓRIO GOOGLE) X NOVOS CASOS DE INFECÇÃO POR COVID - 19 (BASE SEADE-SP) NO ESTADO DE SÃO PAULO, PARA O MUNICÍPIO DE JALES - período de 2020-04-01 a 2021-06-30



CORRELAÇÃO LINEAR ENTRE AS MÉDIAS MÓVEIS DE VARIAÇÃO DE MOBILIDADE E DE CASOS NOVOS PARA JALES
 Período de 2020-04-01 a 2021-06-30 : 0.10684308838993924

>>> A escala de cores do gráfico corresponde ao tempo (valores maiores == datas mais recentes) <<<

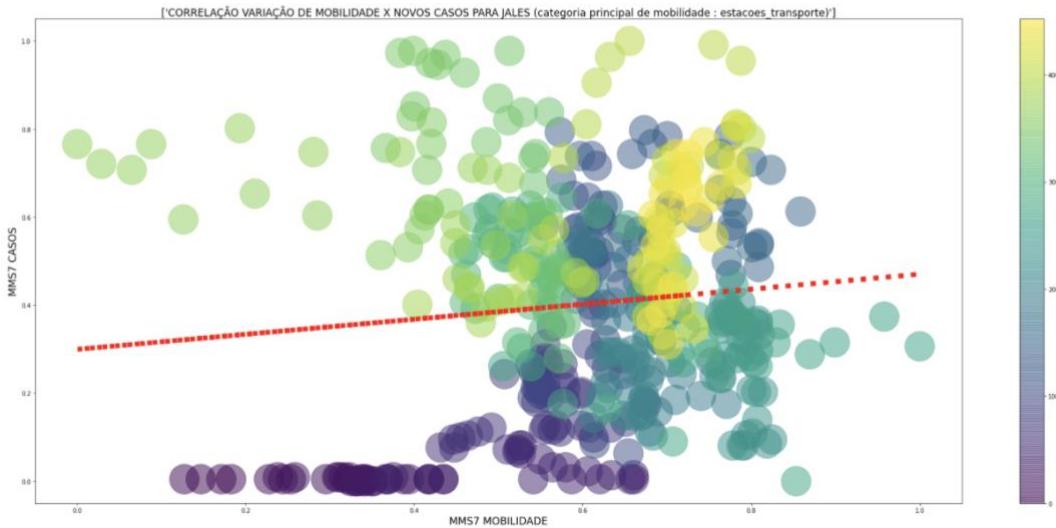
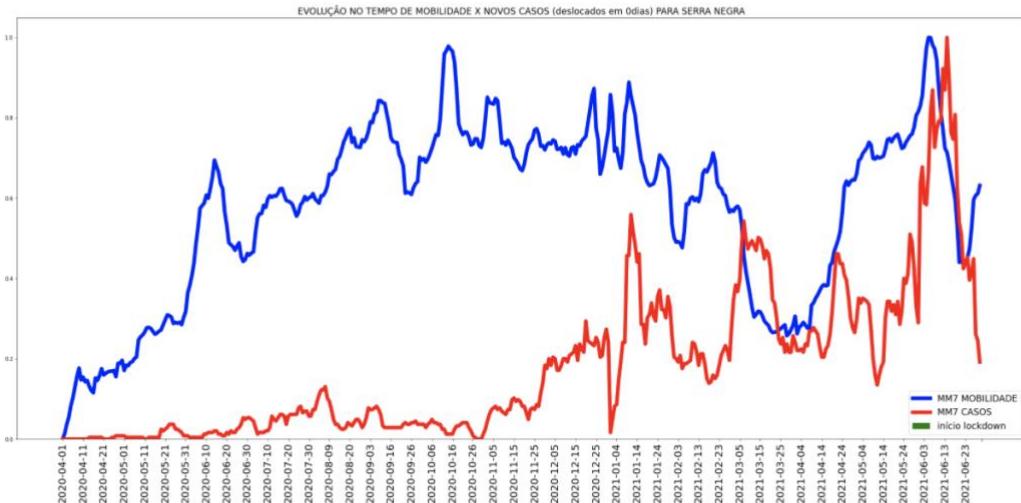


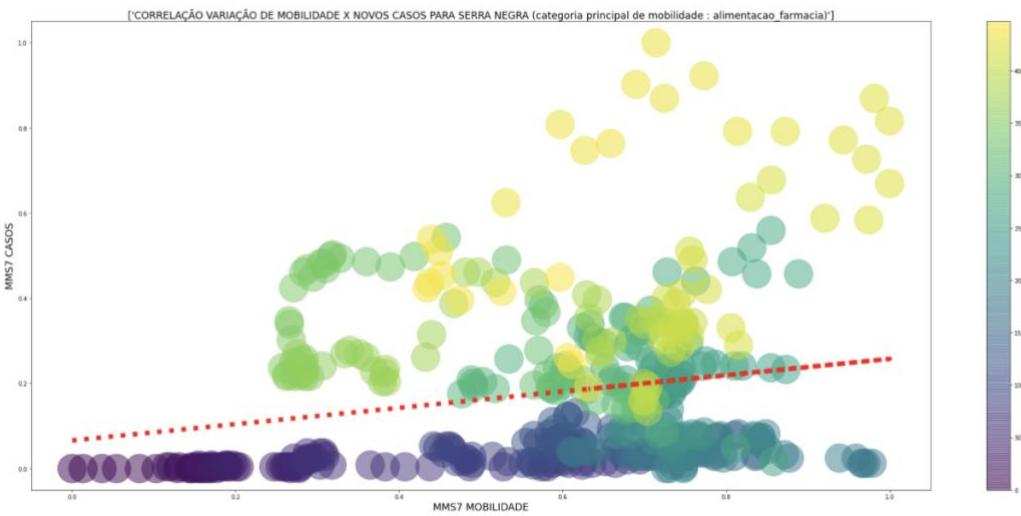
FIGURA 16 - Comparação de evolução de casos e mobilidade para Serra Negra

 ANÁLISE DE MOBILIDADE (RELATÓRIO GOOGLE) X NOVOS CASOS DE INFECÇÃO POR COVID - 19 (BASE SEADE-SP) NO ESTADO DE SÃO PAULO, PARA O MUNICÍPIO DE SERRA NEGRA - período de 2020-04-01 a 2021-06-30



CORRELAÇÃO LINEAR ENTRE AS MÉDIAS MÓVEIS DE VARIAÇÃO DE MOBILIDADE E DE CASOS NOVOS PARA SERRA NEGRA
 Período de 2020-04-01 a 2021-06-30 : 0.21358725355570715

>>> A escala de cores do gráfico corresponde ao tempo (valores maiores == datas mais recentes) <<<



CONSOLIDADO PARA O ESTADO DE SÃO PAULO:

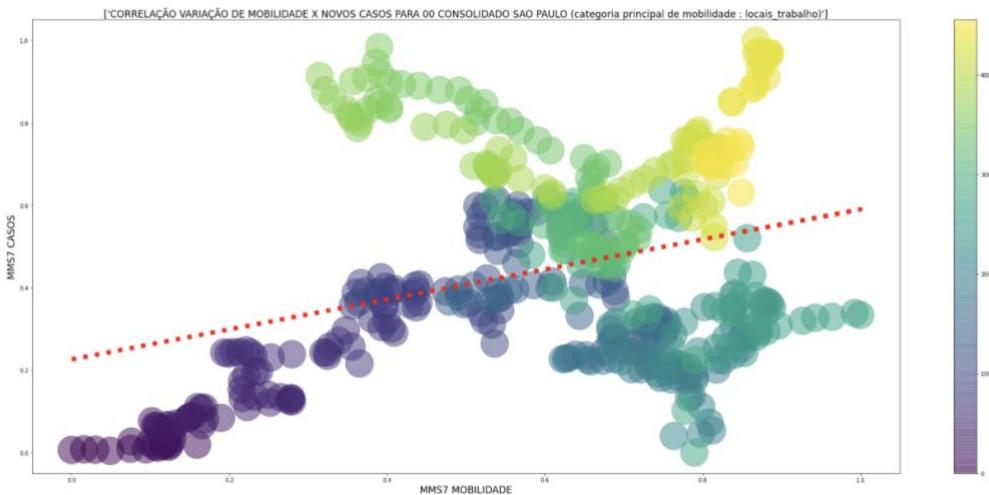
FIGURA 17 - Comparação de evolução de casos e mobilidade para o estado de São Paulo

 ANÁLISE DE MOBILIDADE (RELATÓRIO GOOGLE) X NOVOS CASOS DE INFECÇÃO POR COVID - 19 (BASE SEADE-SP) NO ESTADO DE SÃO PAULO, PARA O MUNICÍPIO DE 00 CONSOLIDADO SAO PAULO - período de 2020-04-01 a 2021-06-30



CORRELAÇÃO LINEAR ENTRE AS MÉDIAS MÓVEIS DE VARIAÇÃO DE MOBILIDADE E DE CASOS NOVOS PARA 00 CONSOLIDADO SAO PAULO
 Período de 2020-04-01 a 2021-06-30 : 0.3314182213805825

>>> A escala de cores do gráfico corresponde ao tempo (valores maiores == datas mais recentes) <<<



Nos gráficos acima apresentados, pudemos notar diversos padrões de evolução no decorrer do tempo, como por exemplo a sincronia de crescimento das curvas para o estado de São Paulo (consolidado – Figura 17) em 2020, seguida de um descolamento de ambas mais para o final do ano. Ainda não havia vacinas, mas diversas categorias de mobilidade, por exemplo, trabalho, tiveram peso, com empresas colocando seus quadros completos de funcionários em home-office (trabalhando em casa), enquanto outras categorias de mobilidade cresceram muito, como as idas ao comércio de rua, shoppings e parques. Esse padrão também foi observado em cidades interioranas de grande porte (Campinas, figura 8; Ribeirão Preto, figura 9). e cidades da Grande São Paulo (Itapevi, figura 12; Itapecerica da Serra, figura 13; S. Caetano do Sul, figura 14).

Uma observação bastante interessante, que é coberta pelo conhecimento em epidemiologia, é que sempre que há uma diminuição na mobilidade leva um tempo até isso se refletir na curva de casos, e vice-versa. Esta informação ataca diretamente o resultado de medidas públicas contra a desinformação.

Como mostram as barras verdes indicadoras de lockdown oficial em algumas cidades, tal procedimento decretado pelas prefeituras municipais em geral foi seguido pela população – quedas abruptas de mobilidade geral podem ser vistas junto a tais barras (exemplo, Ribeirão Preto, Figura 9). Evitamos colocar nos gráficos de evolução no tempo as categorias de mobilidade detalhadas e os números da pandemia detalhados em óbitos, casos, internações, porque isso tornaria os mesmos por demais carregados e de difícil visualização. A capital paulista (Figura 3) seguiu um perfil bastante parecido com o estado. Já em Botucatu (Figura 4) vimos a curva de casos bastante controlada em relação a mobilidade (lembrando que outros fatores como uso de máscaras e distanciamento individual em locais públicos, aqui não incluídos por serem de mais difícil mensuração, tem um peso importante). Perfil semelhante em Franca (Figura 10), e Serra Negra (Figura 16) onde vemos o perfil da curva de casos acompanhar a de mobilidade, mas com algum deslocamento – de fato, esse padrão pode ser visto em quase todos os gráficos de tempo, variando de acordo com a cidade e a época.

Os gráficos de evolução no tempo são essenciais para a tomada de decisões de restrição de mobilidade, sempre considerando fatores locais e o perfil da mobilidade. Mais adiante analisamos a movimentação de pessoas aberta em suas subcategorias, o que pode direcionar medidas restritivas aos setores de maior correlação positiva (por exemplo fechando parques mas preservando o comércio).

Os gráficos da parte de baixo de cada figura, que mostram a correlação da mobilidade geral (média móvel da média das seis categorias) com o número de casos (também média móvel) mostram sempre o valor positivo, mas com correlação linear baixa. Como veremos adiante, as categorias de mobilidade interagem entre si (exemplo, empresas vazias decorrem de residências cheias, e, portanto, haverá mais casos de transmissão doméstica que nas empresas, se considerarmos, claro, os dois extremos, com gente demais em casa e ninguém no trabalho. Se o excesso de pessoas em casa significar mais gente no comércio, então já veremos outro tipo de interação. O objetivo aqui foi fazer o paralelo entre a mobilidade geral, consolidada, e as ocorrências.

CORRELAÇÕES POR CATEGORIA DE MOBILIDADE

A partir das análises de correlação acima, vamos agora gerar um novo tipo de gráfico, a matriz de nuvens de pontos, mostrando qualitativamente como a correlação se distribui, nas diversas categorias de variação de mobilidade. A maior de todas as correlações fica ainda abaixo de 0.5, portanto vemos que os valores são muito pequenos -- lembremos que estamos olhando apenas correlações positivas, onde as categorias de variação de mobilidade e os novos casos crescem juntos.

É importante pontuar que as categorias de mobilidade não variam de forma isolada, elas ocorrem todas juntas num mesmo contexto geográfico e, portanto, cada uma pode ser influenciada por várias outras. Não está em nosso escopo analisar e medir tais influências, mas a matriz de correlações e as nuvens de pontos podem exibir um quadro qualitativo onde veremos onde essas influências são mais fortes.

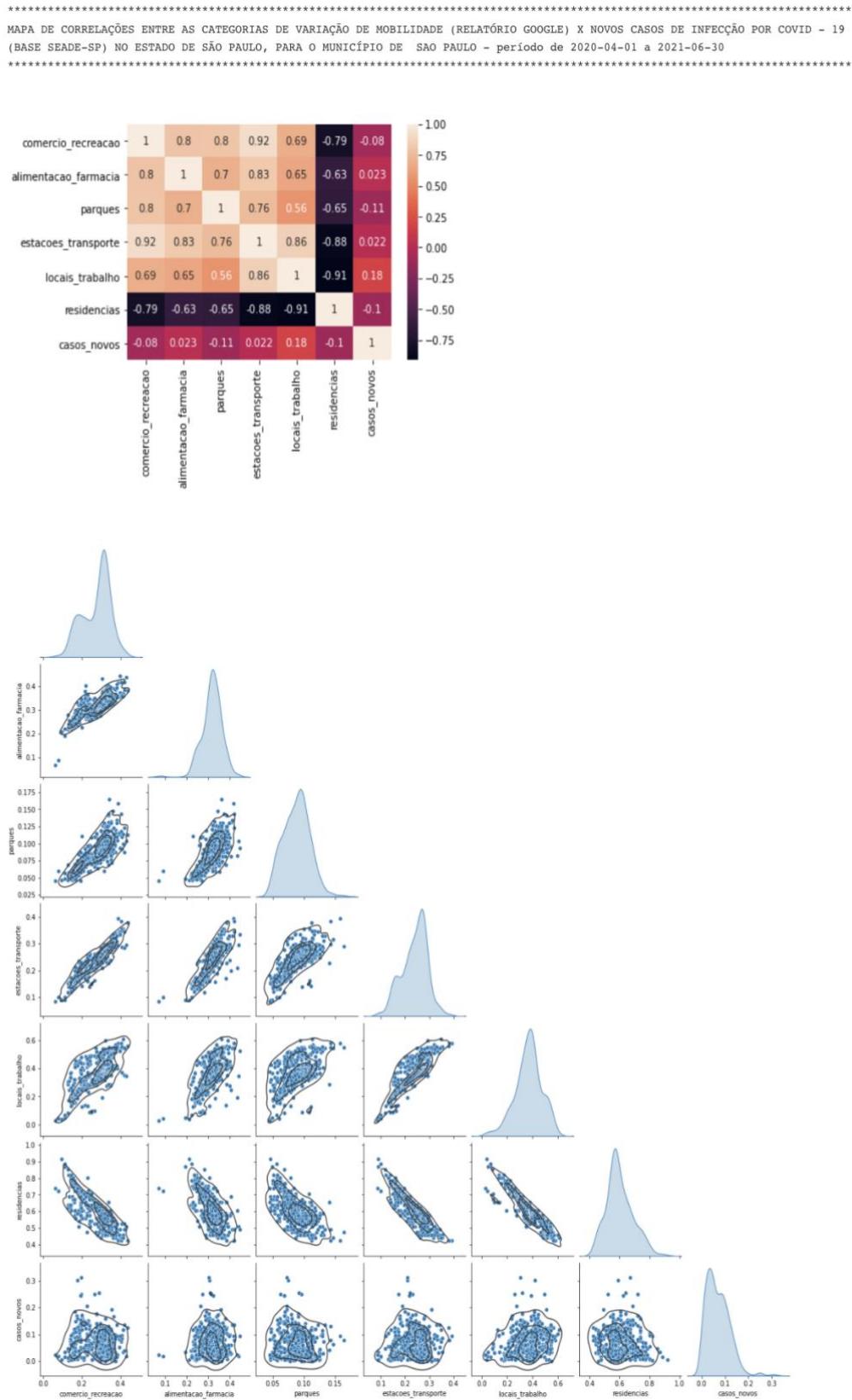
A seguir, vamos fazer uma importante verificação: as correlações entre as 6 categorias de variação de mobilidade, e especialmente a correlação entre cada categoria e o número de novos casos. A ideia é verificar qual categoria de variação de mobilidade mais influencia no surgimento de novos casos. Inicialmente vamos padronizar as colunas de estudo (as 6 categorias mais os casos novos).

No início, colocamos o quadro completo com os valores das correlações em cada célula e seguindo uma escala de cores (heat map), para estabelecer o estudo quantitativo.

Para o estudo qualitativo, temos em seguida a matriz de correlações, onde para cada par (i,j) vemos a correlação da coluna i com a coluna j , e na diagonal - (i,i) - vemos o histograma da coluna i . As colunas de interesse continuam sendo as 6 categorias de mobilidade mais a coluna de casos novos (foi definido na lista 'colunas' anteriormente).

Aqui utilizamos a matriz de correlações da Seaborn, tomando apenas as células abaixo da diagonal, pois aquelas acima são equivalentes. Também aproveitamos para incluir os contornos de densidade, que envolvem agrupamentos e áreas mais densas na plotagem, ajudando a indicar possíveis padrões.

FIGURA 18 - Mapa e matriz de correlações para o município de São Paulo



Podemos notar fortes correlações positivas e negativas entre as categorias de mobilidade na matriz acima, como por exemplo estacoes_transporte x locais_trabalho (positiva) e residencias x locais_trabalho (negativa). Mas o ponto maior de atenção é a última coluna/linha (categorias de mobilidade X casos): o tipo de correlação é menos óbvio. Todos apresentam um formato semelhante, mas as conclusões ficam mais complexas: as categorias se somam e se subtraem, tornando as correlações entre cada categoria e os casos resultantes de tais interações.

Por exemplo: a categoria RESIDENCIA (penúltima linha do quadro) possui correlações negativas com as outras 5 categorias, como seria natural supor. Mas a correlação RESIDENCIA x CASOS parece menos visível, levemente positiva. Pelo calculado anteriormente, a categoria de maior correlação positiva com casos para a cidade de SÃO PAULO é a de LOCAIS DE TRABALHO, que qualitativamente difere pouco: visualmente, para poucos casos, vemos mais variação em empresas que em residências - possivelmente resultado de medidas tomadas nos locais de trabalho para evitar o contágio, o que é de mais difícil efetivação dentro de uma residência.

Portanto, qualitativamente, vemos acima a tendência para a correlação entre as categorias de mobilidade e os novos casos apresentar pequenos valores devido à interação entre as categorias, e visualmente apresentarem agrupamentos mais circulares, não oblíquos.

FIGURA 18 - Mapa e matriz de correlações para Botucatu

 MAPA DE CORRELACOES ENTRE AS CATEGORIAS DE VARIAÇÃO DE MOBILIDADE (RELATÓRIO GOOGLE) X NOVOS CASOS DE INFECÇÃO POR COVID - (BASE SEADE-SP) NO ESTADO DE SÃO PAULO, PARA O MUNICÍPIO DE BOTUCATU - período de 2020-04-01 a 2021-06-30

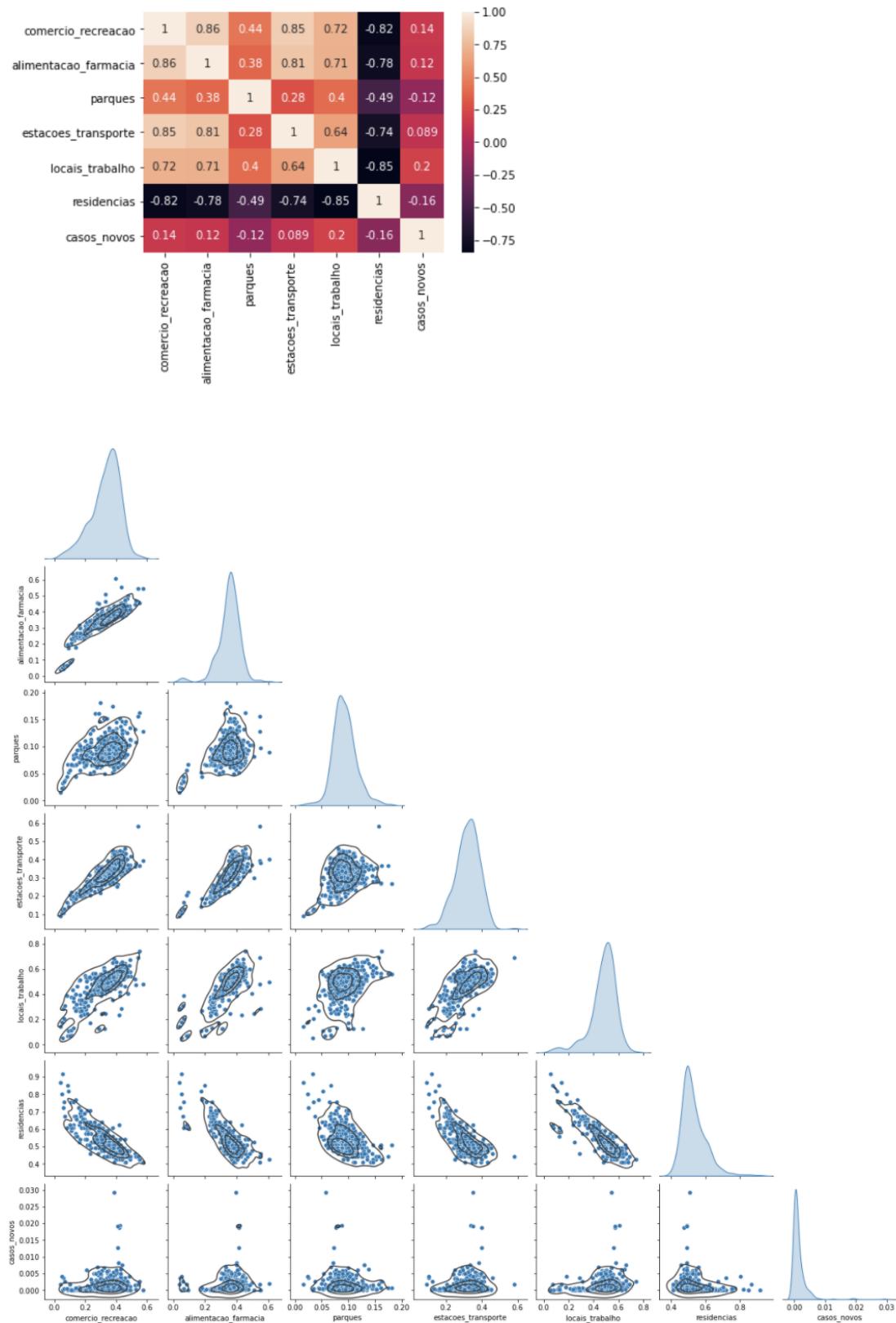


FIGURA 19 - Mapa e matriz de correlações para Votorantim

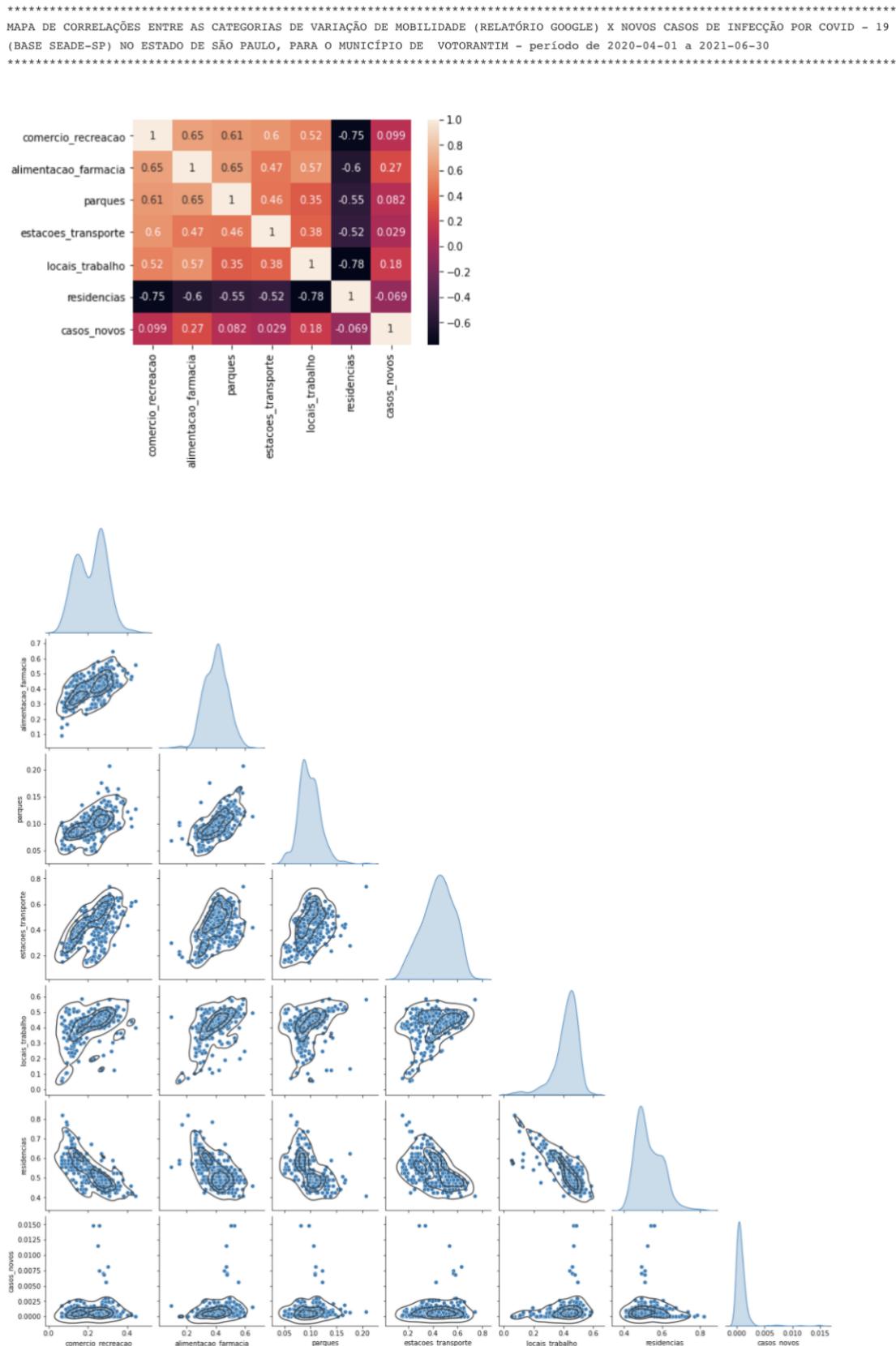


FIGURA 20 - Mapa e matriz de correlações para Sorocaba

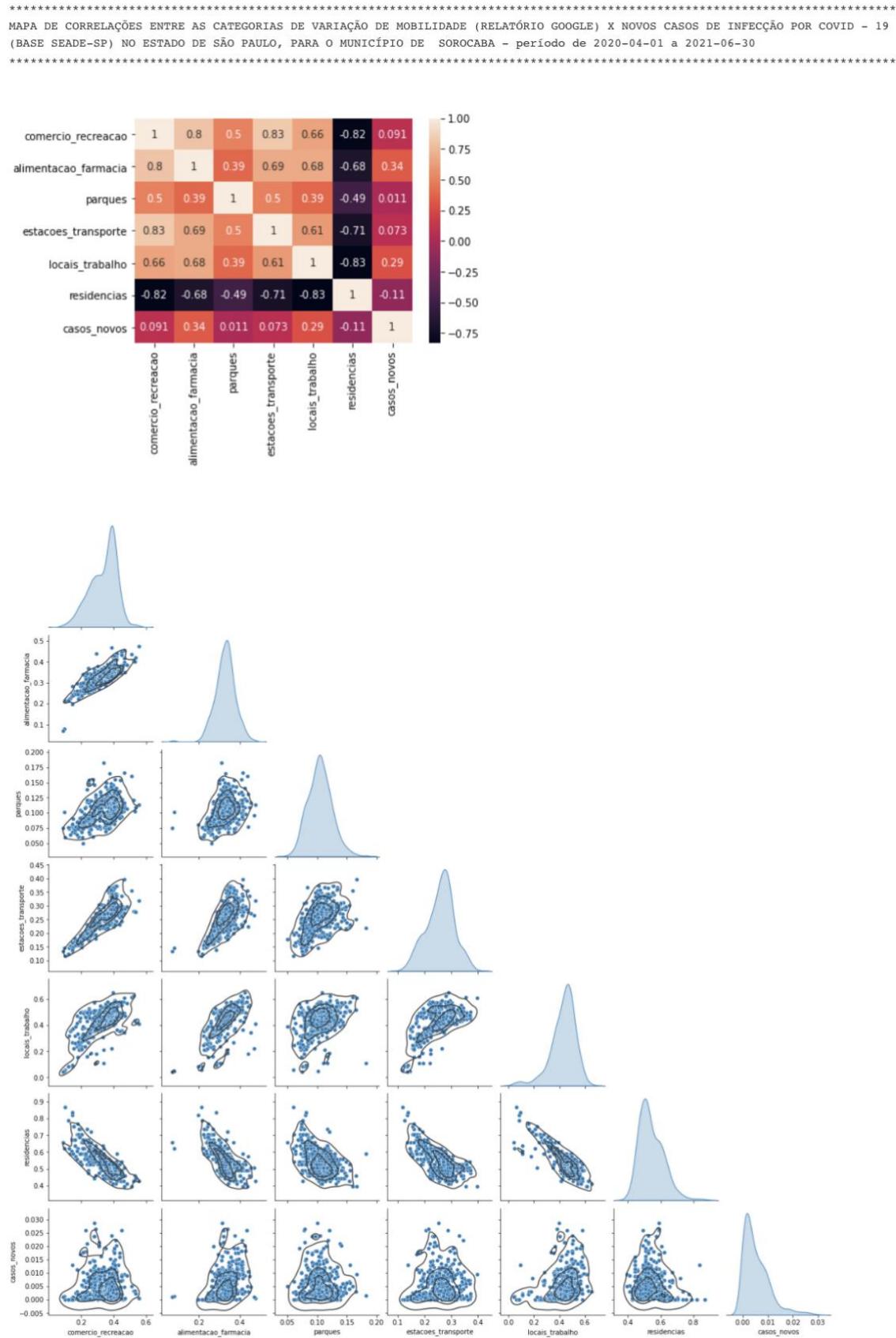
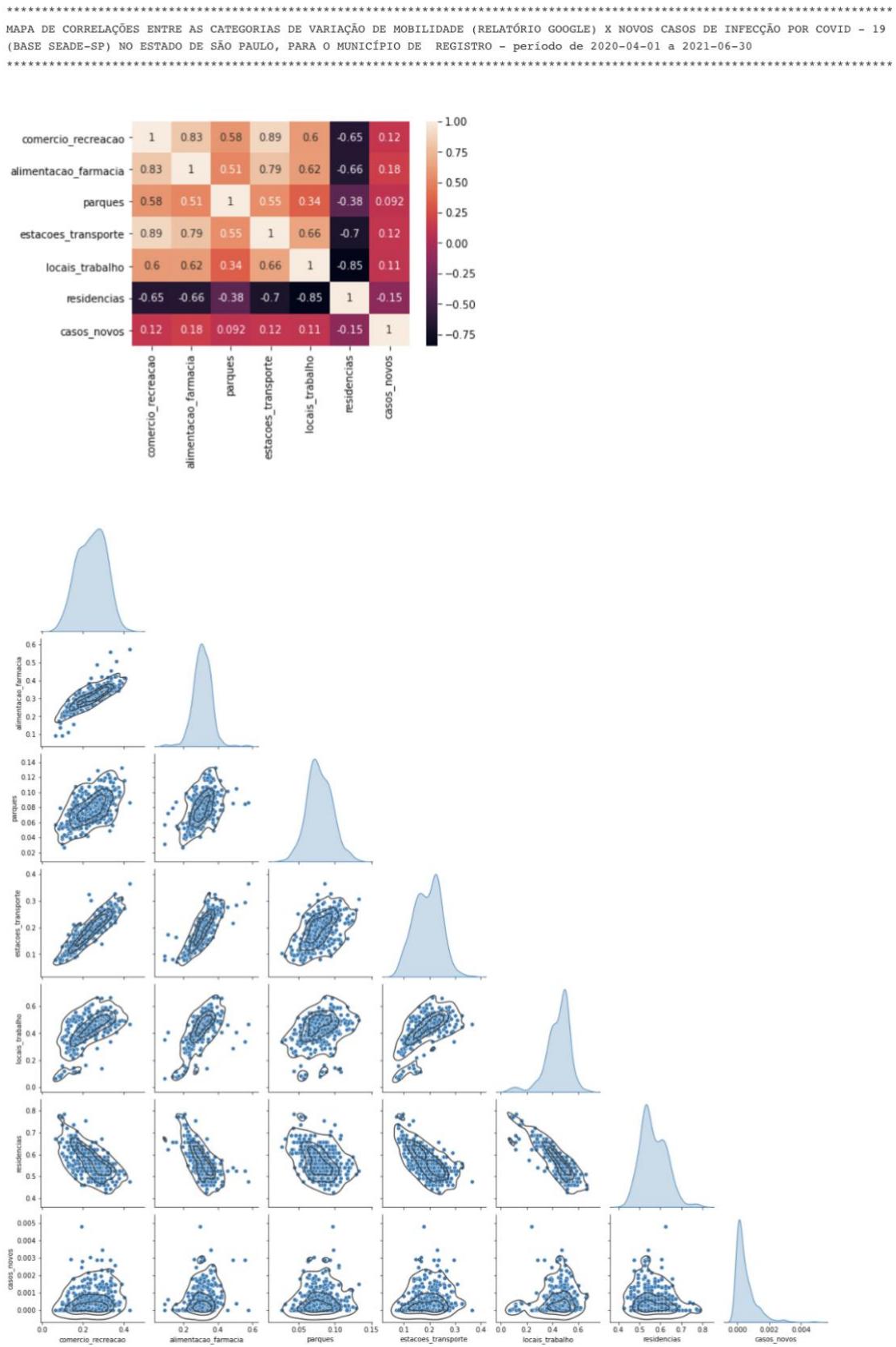


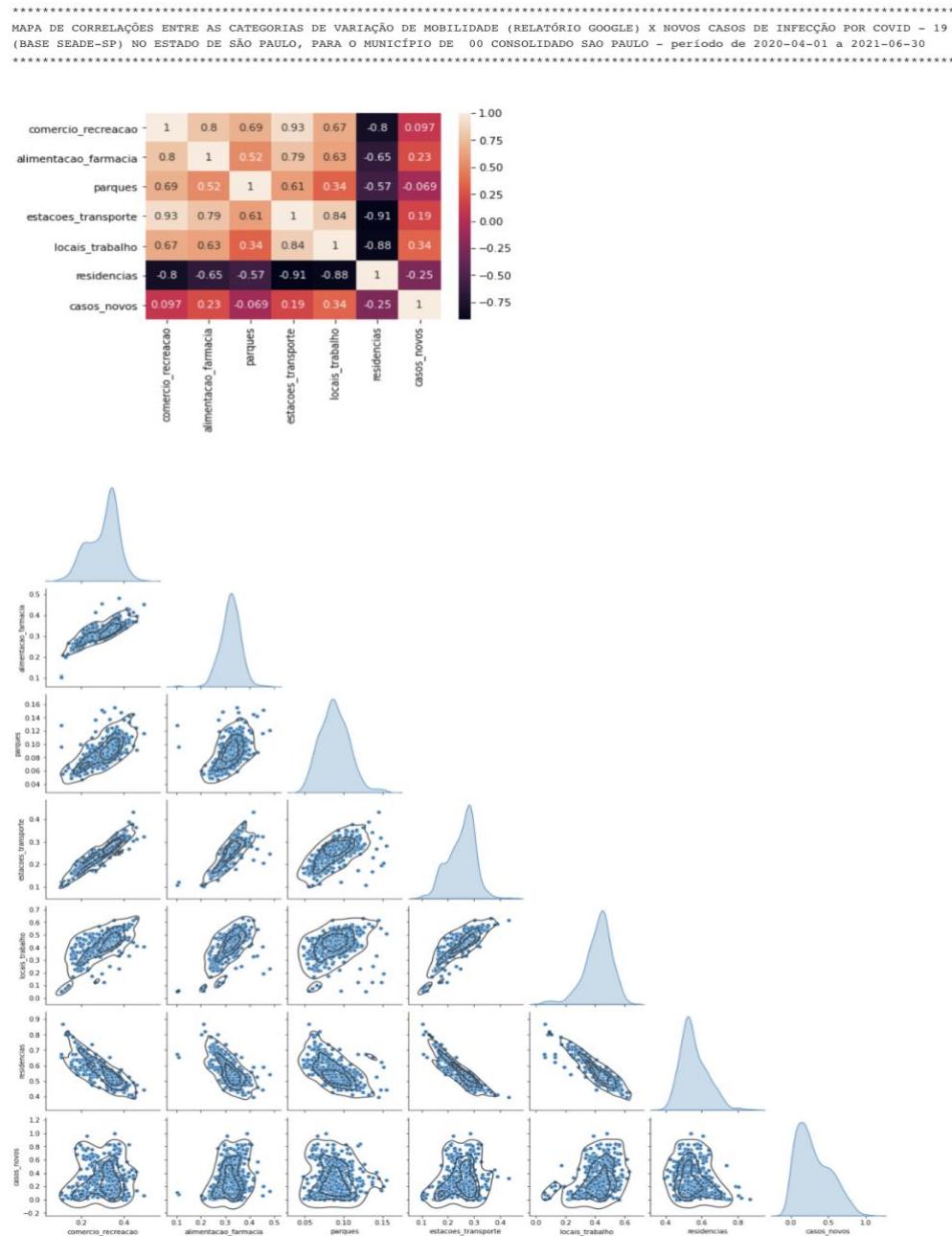
FIGURA 21 - Mapa e matriz de correlações para Registro



Visualmente, podemos observar também que cidades de maior porte tendem a ter agrupamentos mais circulares e homogêneos para a correlação entre categorias e casos, enquanto cidades menores apresentam distribuições mais oblíquas para os pontos - indicando possivelmente menor interação entre as categorias, levando ao predomínio de algumas e, portanto, tendências mais claras.

Visão consolidada para o estado de São Paulo:

FIGURA 22 - Mapa e matriz de correlações para o estado de São Paulo



As duas visualizações a seguir (Figuras 23 e 24) foram geradas na ferramenta TABLEAU ®, a partir da base de dados (todas as datas de 15/02/2020 a 30/06/2021 para todos os municípios paulistas), e são equivalentes.

Para cada município do estado foi calculada a categoria de variação principal, ou seja, aquela que possui maior correlação positiva com o número de casos no município dentro do intervalo de datas considerado. Cada categoria principal de variação de mobilidade é representada por uma cor, e no caso da primeira visualização (em círculos), cada ponto é uma linha da tabela (todas as linhas de um município possuem o mesmo valor para essa categoria principal). Na visualização circular, o cursor foi colocado bem em cima do ponto que mostra, para o município de Cruzeiro, na data de 17/02/2021, a categoria principal para toda a base de dados (locais_trabalho) e qual sua correlação com o número de casos (0.0885). A visualização abaixo contém exatamente a mesma informação em formato de barras horizontais.

FIGURA 23 - Distribuição das categorias principais por município - Estado de São Paulo

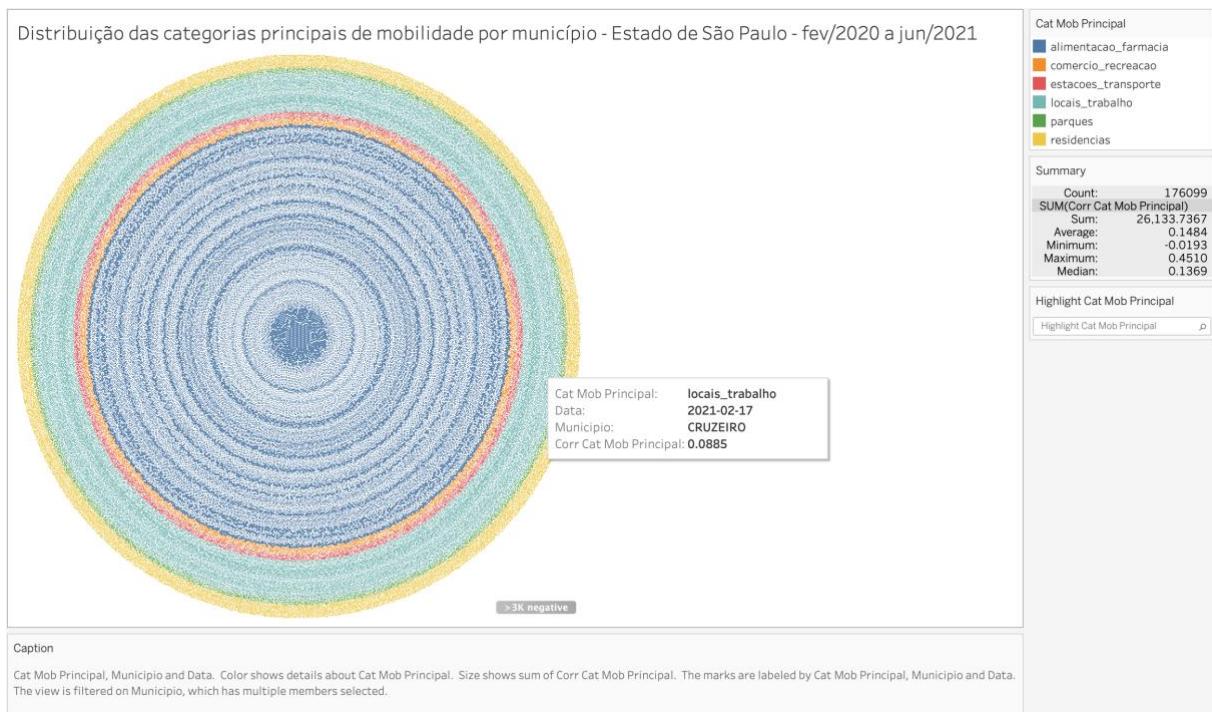
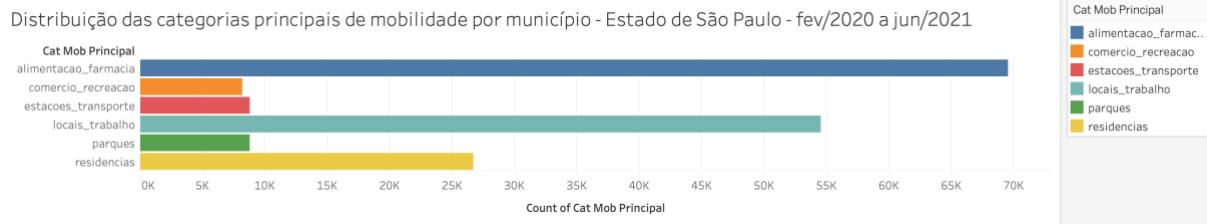


FIGURA 24- Distribuição das categorias principais por município - Estado de São Paulo - versão linear

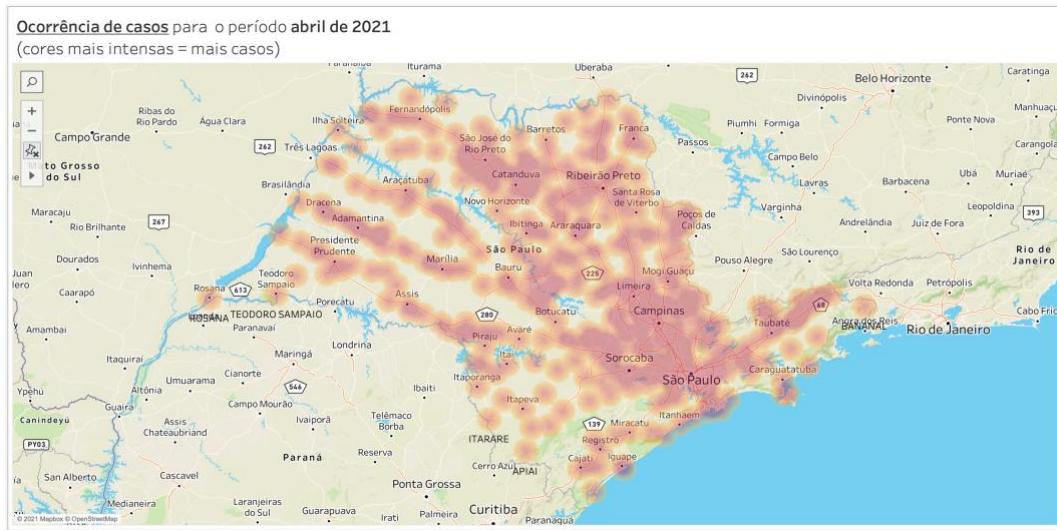


Abaixo, os mapas gerados também na ferramenta Tableau®, comparando a mobilidade em janeiro e fevereiro de 2021 e o número de casos em abril de 2021. A distribuição da mobilidade e dos casos ao longo das mais importantes rodovias do estado, como podemos ver nos mapas, confirma o ponto levantado por Miguel Nicolelis (NICOLELIS et al, 2021) do importante papel das redes de estradas brasileiras como vias através das quais a COVID-19 espalhou-se dos grandes centros para o interior - e depois, via "efeito bumerangue", houve o fluxo de doentes em sentido contrário, em busca de tratamento adequado em cidades de maior porte.

FIGURA 25 - Média geral de variação de mobilidade para janeiro-fevereiro/2021



FIGURA 26 - Ocorrência de casos para abril/2021



4.2 ANÁLISE COMPARATIVA DOS DADOS DE MOBILIDADE DO GOOGLE PARA OS PAÍSES: BRASIL, CHILE E REINO UNIDO

Da mesma data que extraímos os dados da base Google para o estado de São Paulo, o fizemos para BRASIL, CHILE e REINO UNIDO. Os tratamentos prévios realizados na base, tanto diretamente no Excel como no Jupyter Notebook, foram os mesmos da base anterior.

A escolha do Reino Unido teve como referência o trabalho de [BASELLINI et al, 2019], que também utilizou os dados do Google Mobility Reports e, embora tenha abrangido um escopo geográfico um pouco mais restrito (apenas Inglaterra e País de Gales) mostrou um cenário bastante detalhado das medidas pelo governo britânico, no sentido de controlar o avanço da doença, em especial os índices de mortalidade da mesma.

Para esta massa de dados, mais adiante, limitaremos a análise ao ano de 2020, já que os procedimentos de vacinação no Reino Unido começaram ainda em dezembro daquele ano e já no primeiro semestre de 2021 o seu efeito começou a influenciar a curva de casos. Temos agora, então, uma base em formato “excel” contendo o período fevereiro a dezembro/2020 para a variação das categorias de mobilidade para Brasil, Chile e Grã-Bretanha.

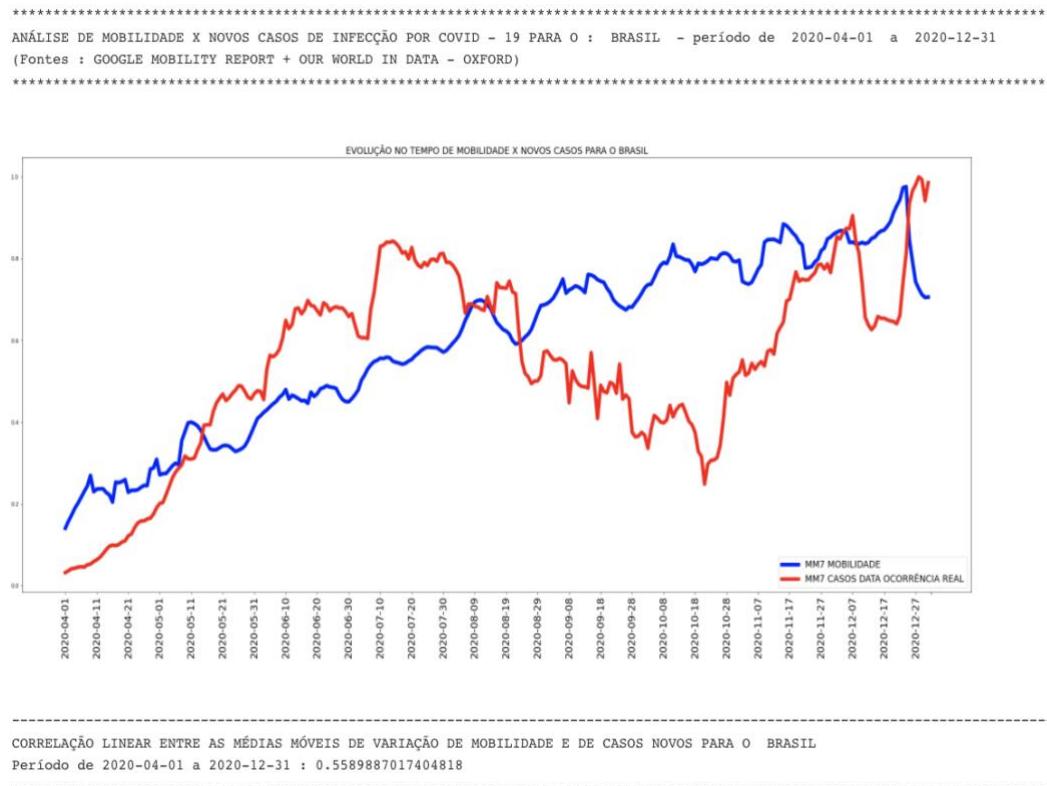
Fonte de dados para casos - Brasil, Chile e Grã-Bretanha - extraído da base do site OUR WORLD IN DATA (OWID - Oxford) em 09/09/2021.

Foi realizada a inclusão de uma coluna da data aproximada de contaminação (utilizando o tempo médio entre contaminação e notificação de 15 dias para o Brasil, 10 dias para o Chile e 3 dias para Reino Unido). Em seguida foi feito o merge das bases de mobilidade e de casos acima, amarrando por país e por data - na base de casos, deve-se utilizar a data da contaminação (aproximada). Fixamos a data inicial em 01/04/2020 para seguir o padrão da análise realizada para o estado de São Paulo na seção anterior. Chama a atenção a diferença de perfil entre cada gráfico, pois aqui vemos países diferentes, com políticas bem distintas de reação à pandemia.

Destacamos o fato que, embora para o caso do Reino Unido e Chile a correlação entre a mobilidade global (resultante da média das categorias) e o número de casos seja negativa, a abertura em categorias, que nos permite comparar cada categoria com os casos, tem várias delas positivas, como poderemos ver nos heatmaps.

O código fonte que gerou os gráficos pode ser encontrado em:
<https://github.com/rlepsch/MBA-USP-TCC-2021> no notebook ANALISE003.

Figura 27 - Comparação de evolução de casos e mobilidade para o Brasil



>>> A escala de cores do gráfico corresponde ao tempo (valores maiores == datas mais recentes) <<<

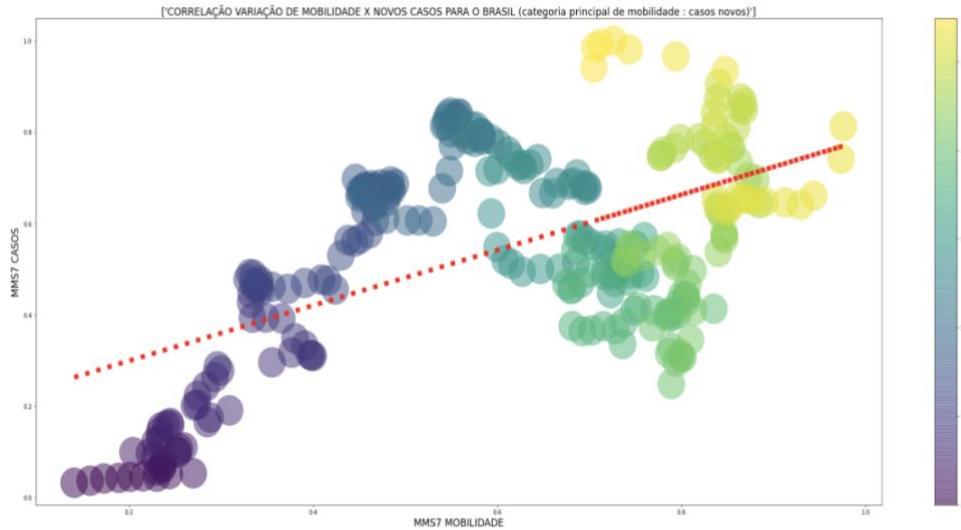
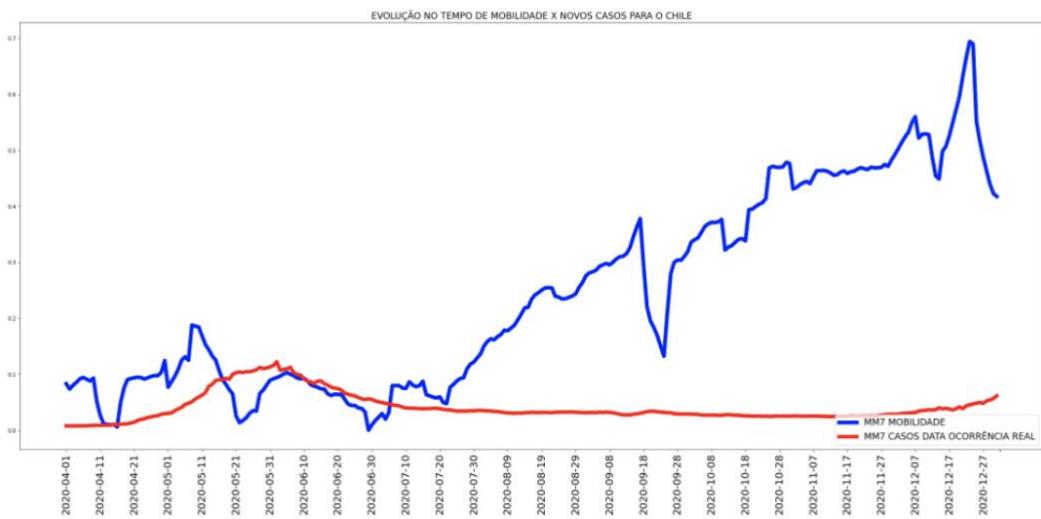


FIGURA 28 - Comparação de evolução de casos e mobilidade para o Chile

 ANÁLISE DE MOBILIDADE X NOVOS CASOS DE INFECÇÃO POR COVID - 19 PARA O : CHILE - periodo de 2020-04-01 a 2020-12-31
 (Fontes : GOOGLE MOBILITY REPORT + OUR WORLD IN DATA - OXFORD)



CORRELAÇÃO LINEAR ENTRE AS MÉDIAS MÓVEIS DE VARIAÇÃO DE MOBILIDADE E DE CASOS NOVOS PARA O CHILE
 Período de 2020-04-01 a 2020-12-31 : -0.3592111496275798

>>> A escala de cores do gráfico corresponde ao tempo (valores maiores == datas mais recentes) <<<

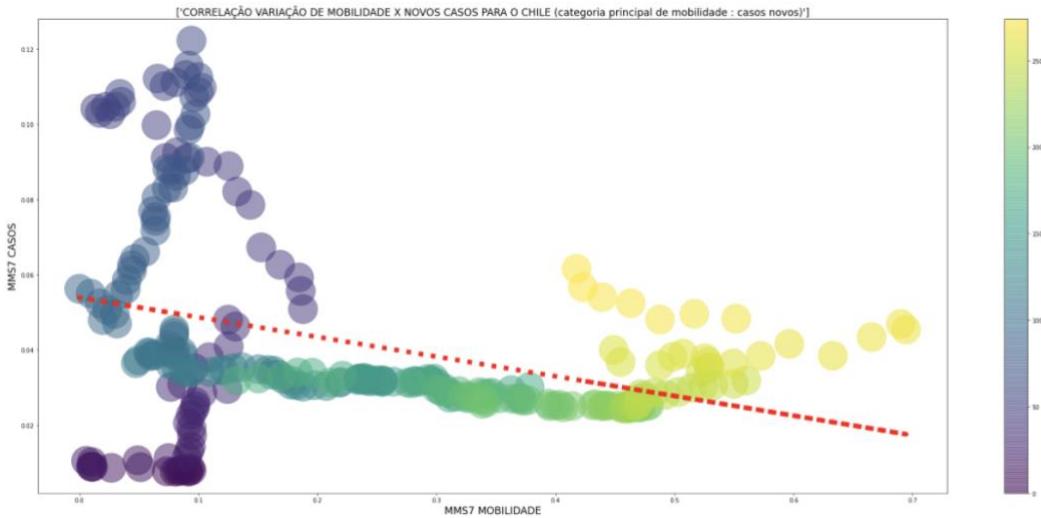
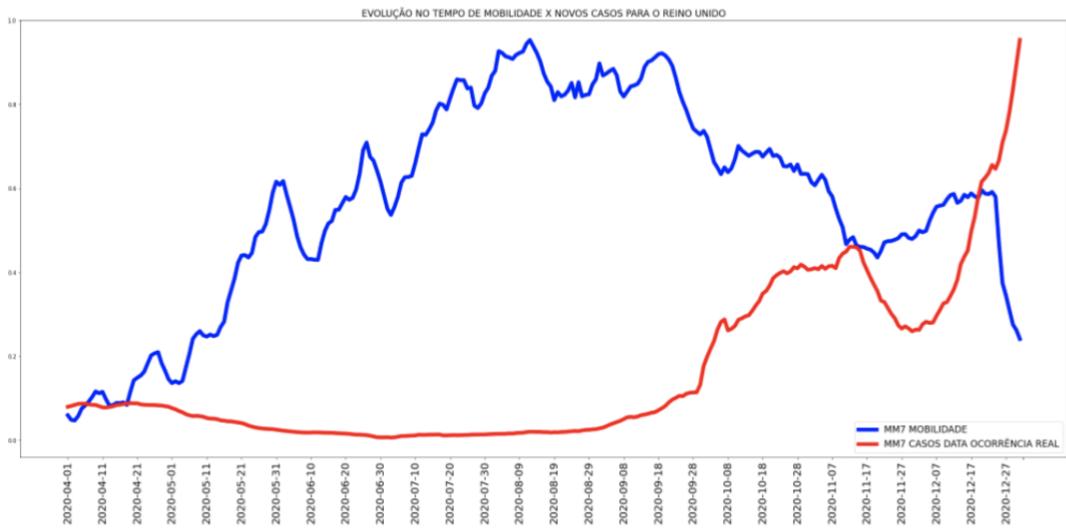


FIGURA 29 - Comparação de evolução de casos e mobilidade para o Reino Unido

 ANÁLISE DE MOBILIDADE X NOVOS CASOS DE INFECÇÃO POR COVID - 19 PARA O : REINO UNIDO - período de 2020-04-01 a 2020-12-31
 (Fontes : GOOGLE MOBILITY REPORT + OUR WORLD IN DATA - OXFORD)



CORRELAÇÃO LINEAR ENTRE AS MÉDIAS MÓVEIS DE VARIAÇÃO DE MOBILIDADE E DE CASOS NOVOS PARA O REINO UNIDO
 Período de 2020-04-01 a 2020-12-31 : -0.1550953903056572

>>> A escala de cores do gráfico corresponde ao tempo (valores maiores == datas mais recentes) <<<

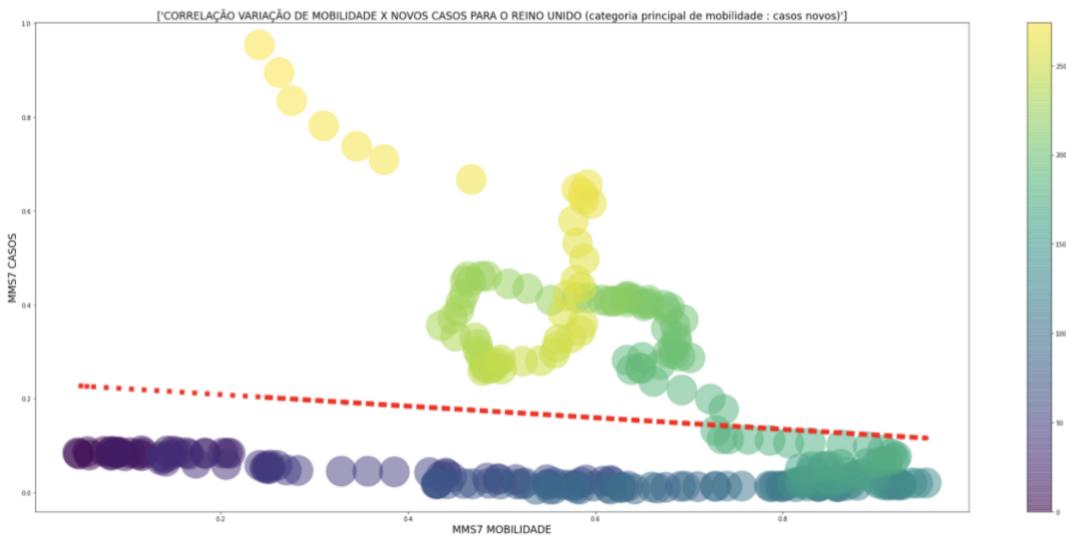


FIGURA 30 - Mapa e matriz de correlações para o Brasil

 MAPA DE CORRELACÕES ENTRE AS CATEGORIAS DE VARIAÇÃO DE MOBILIDADE (RELATÓRIO GOOGLE) X NOVOS CASOS DE INFECÇÃO POR COVID - 19
 (BASE OWID - OXFORD) NO BRASIL - período de 2020-04-01 a 2020-12-31

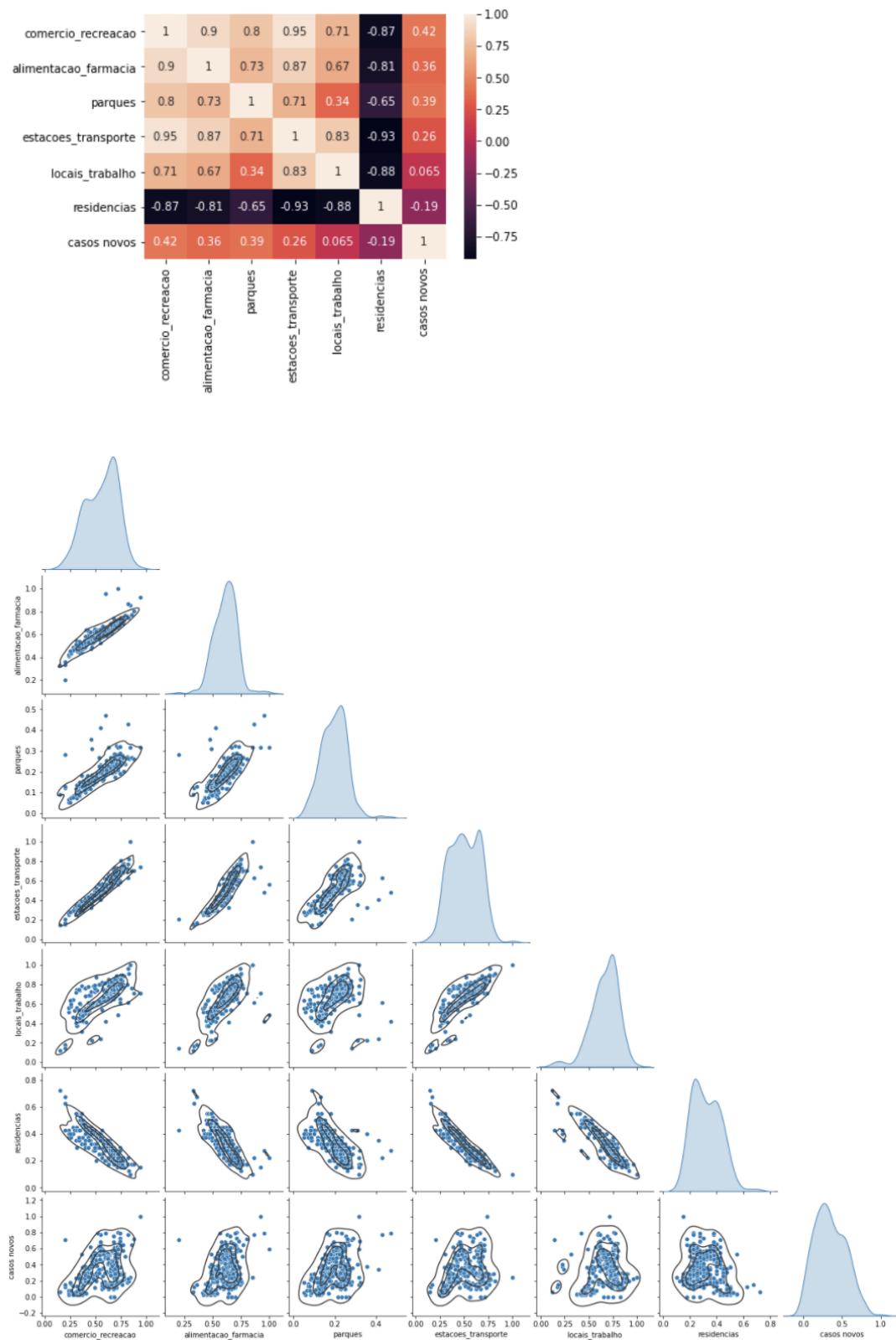


FIGURA 31 - Mapa e matriz de correlações para o Chile

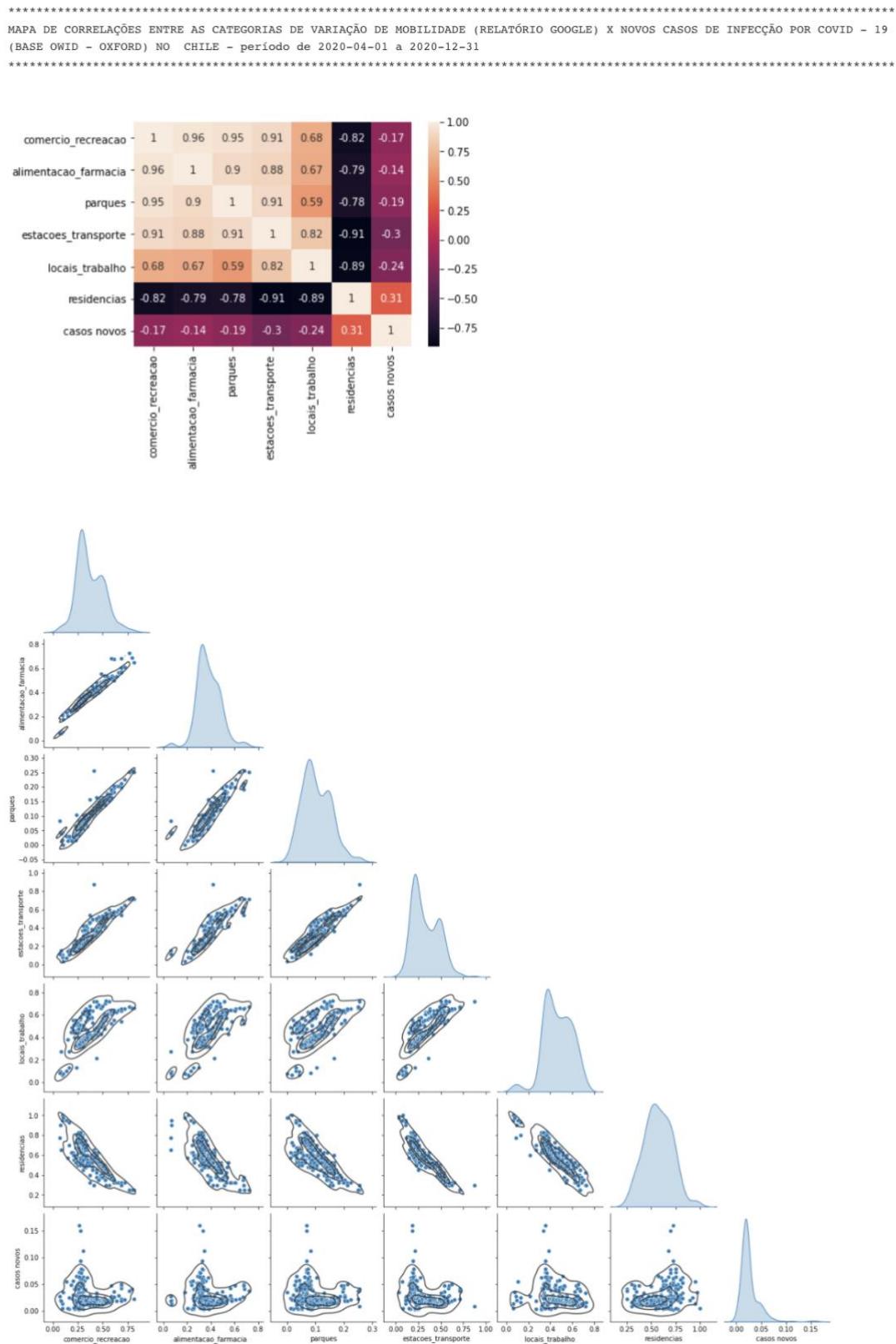
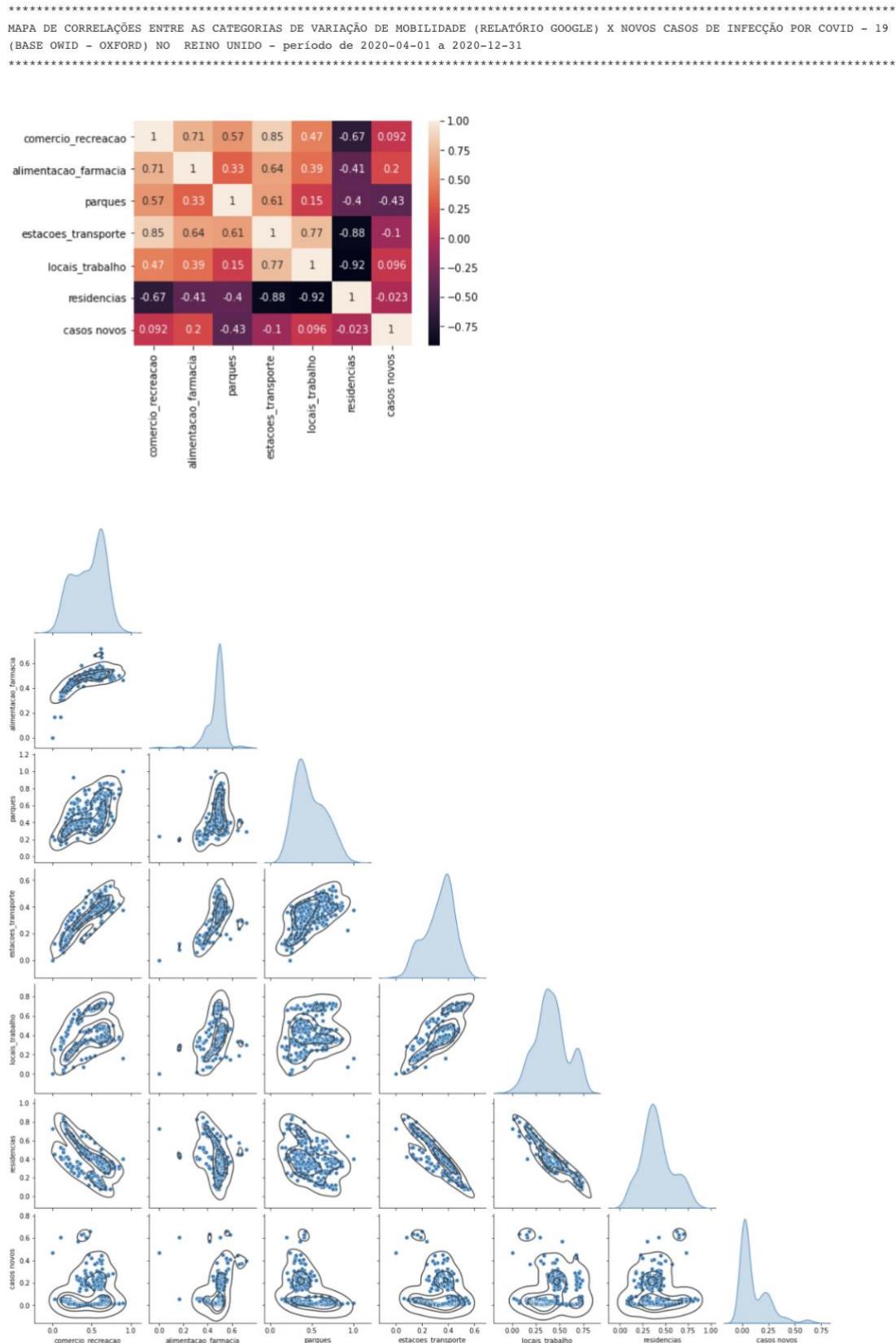


FIGURA 32 - Mapa e matriz de correlações para o Reino Unido



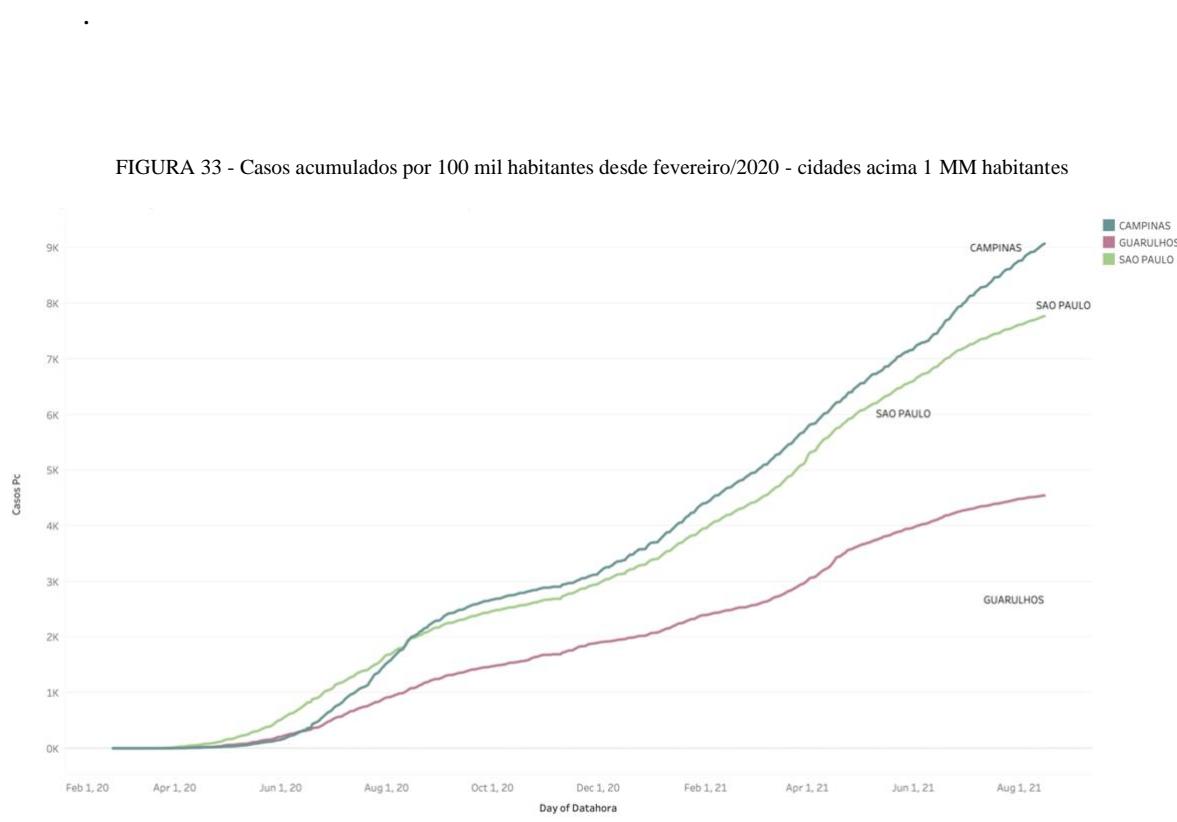
4.3 ANÁLISE COMPARATIVA DOS DADOS DO SEADE PARA OS MUNICÍPIOS DO ESTADO DE SÃO PAULO

Aqui faremos o tratamento prévio da base de casos e óbitos do SEADE, apenas para municípios do estado de São Paulo e acessado no dia 15/08/2021. Os resultados serão gravados no mesmo arquivo para visualização através da ferramenta Tableau, na qual produzimos todos os gráficos deste item. A ideia aqui neste notebook é apenas fazer o tratamento de caracteres acentuados, padronizando para sem acentuação e em maiúsculas.

Fonte dos dados:

https://raw.githubusercontent.com/seade-R/dados-covid-sp/master/data/dados_covid_sp.csv

Abaixo, as visualizações obtidas no Tableau® a partir da base gravada, algumas curvas de evolução comparada de casos, para cidades de portes semelhantes. Os arquivos do Tableau podem ser encontrados em: [https://github.com/rlepsch/MBA-USP-TCC-2021⁶](https://github.com/rlepsch/MBA-USP-TCC-2021).



⁶ Para este item não há código em Python

FIGURA 34 - Casos acumulados por 100 mil habitantes desde fevereiro/2020 - cidades de 500 mil a 1 MM habitantes

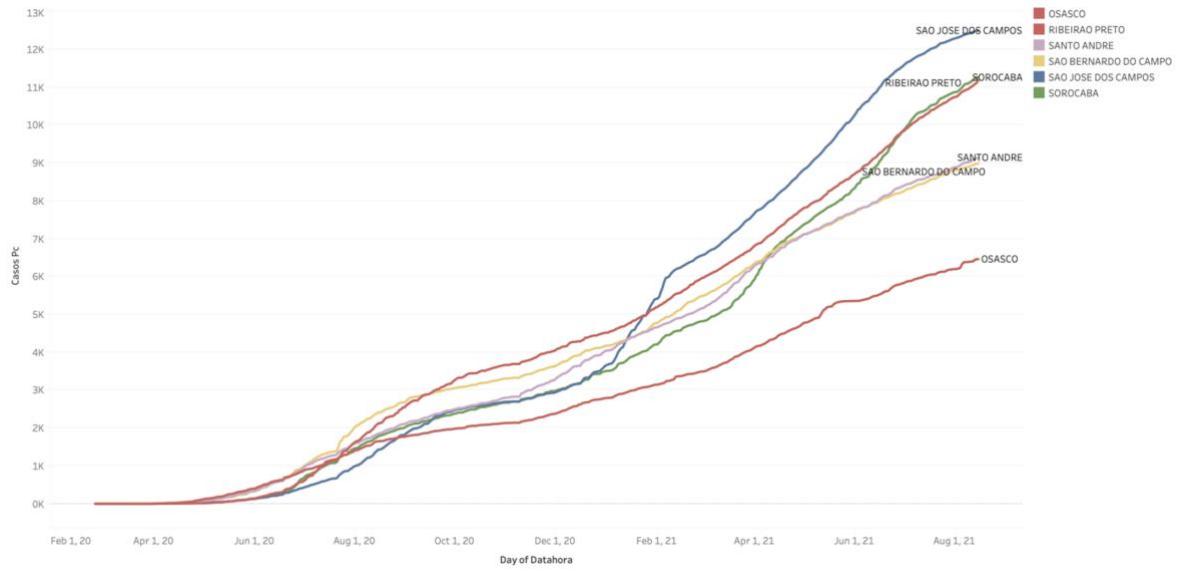
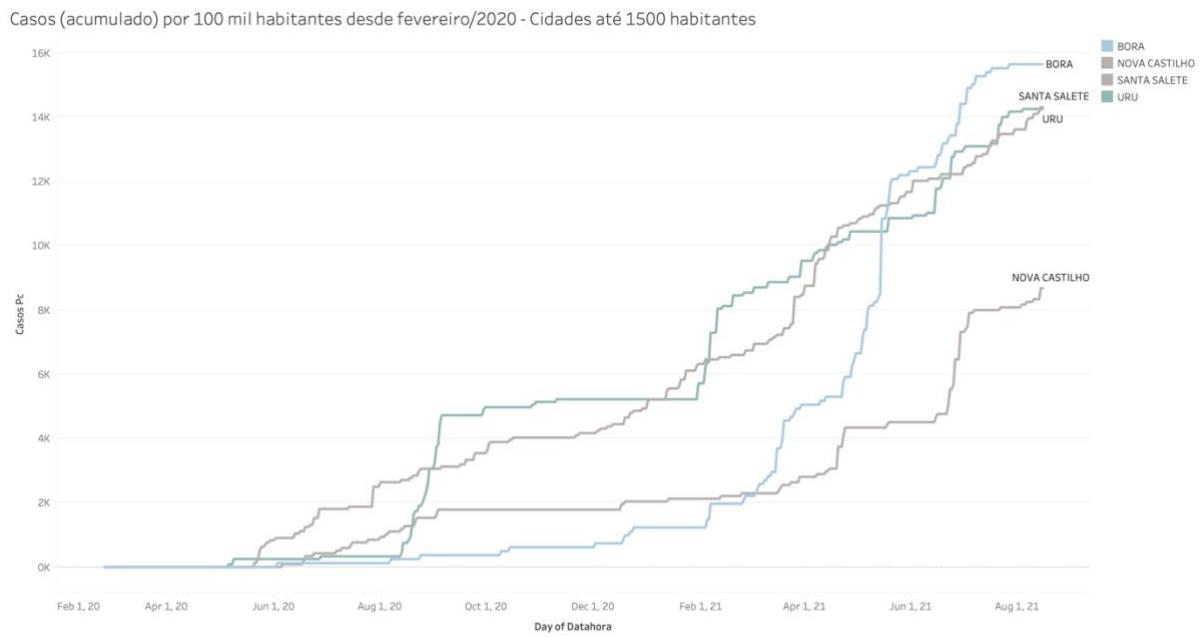
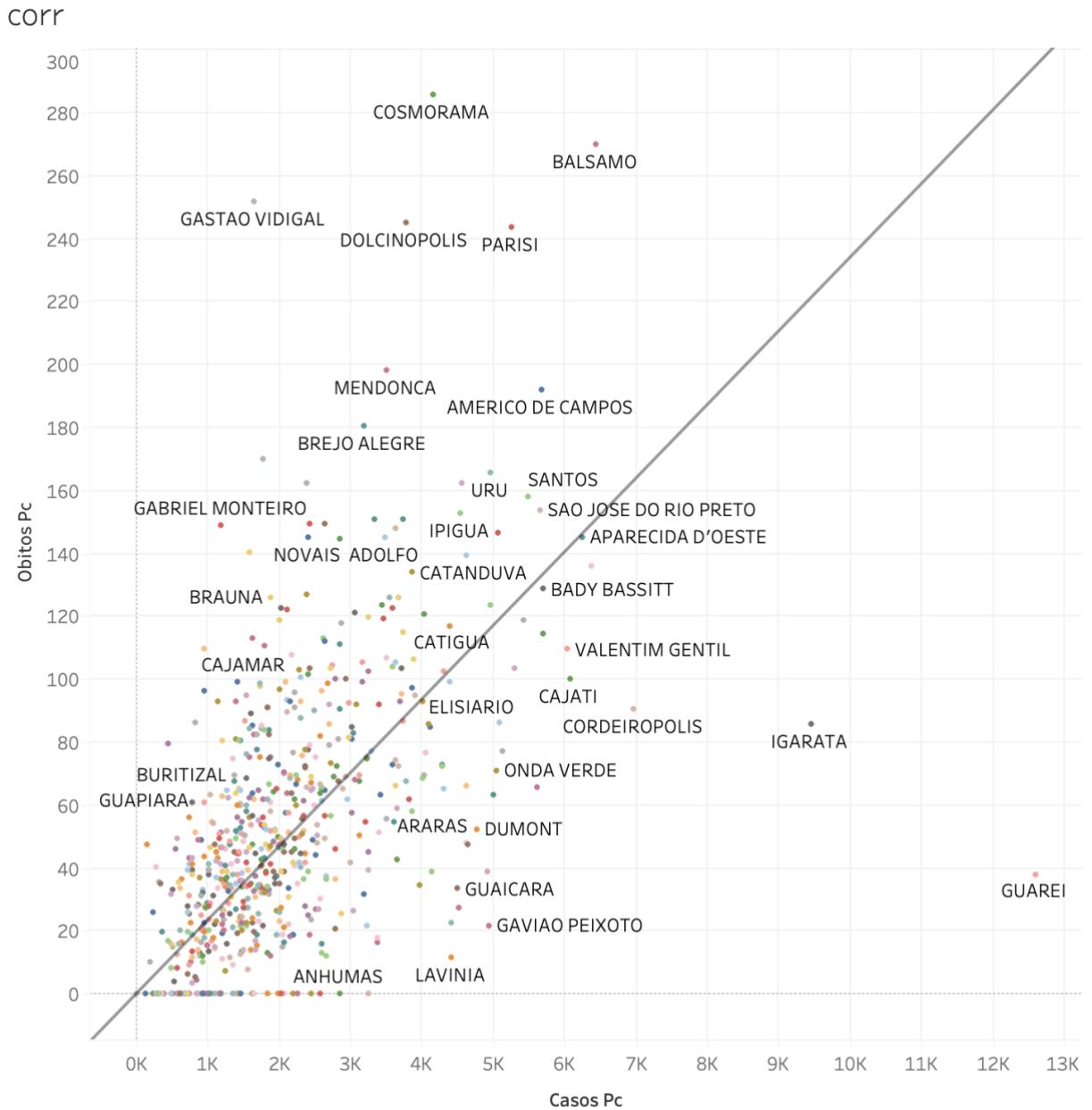


FIGURA 35 - Casos acumulados por 100 mil habitantes desde fevereiro/2020 - cidades até 1500 habitantes



O gráfico de correlação da Figura 36 (entre casos e óbitos por cada 100 mil habitantes) destaca o panorama dos municípios do Estado de São Paulo, com os números acumulados desde o início da pandemia até o dia 24/10/2021, mostrando a divisão entre municípios com mais óbitos (acima da diagonal) e aqueles com menos.

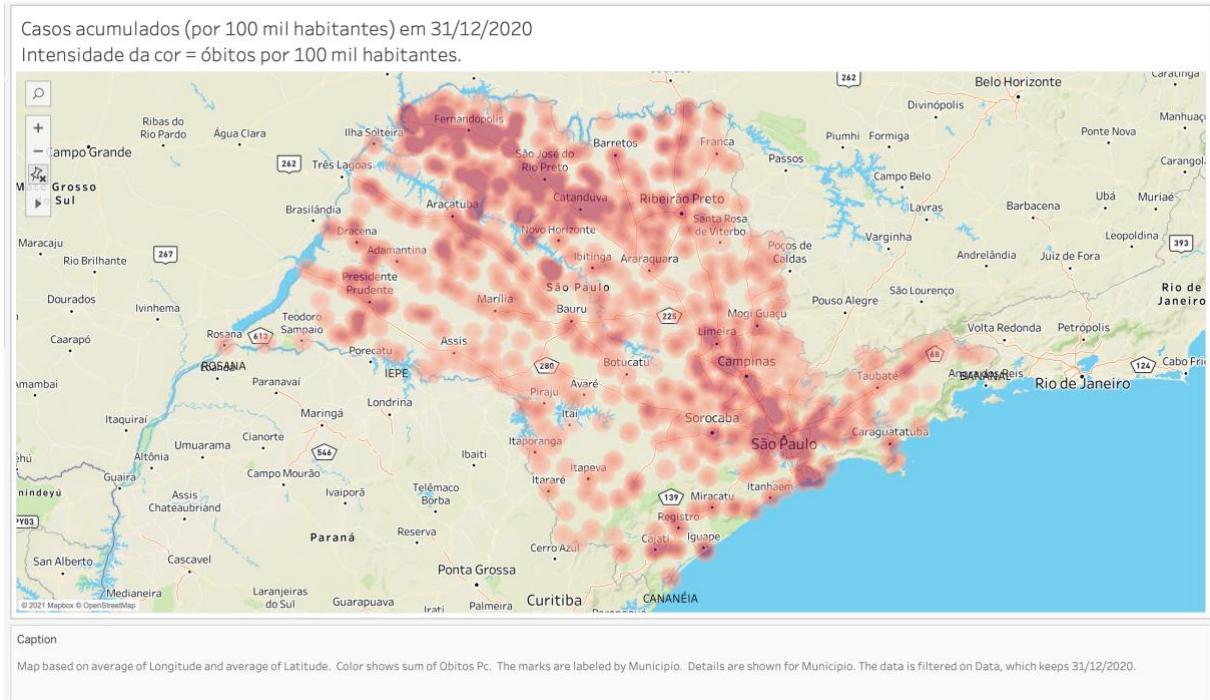
FIGURA 36 - Correlação entre óbitos e casos por 100 mil habitantes desde fevereiro/2020



Sum of Casos Pc vs. sum of Obitos Pc. Color shows details about Municipio. Size shows sum of Casos Pc. The marks are labeled by Municipio. The data is filtered on Datahora, which ranges from 24/10/2020 to 24/10/2020. The view is filtered on Municipio, which keeps 646 of 646 members.

Finalmente o mapa abaixo, gerado no Tableau® mostra a distribuição geográfica dos casos acumulados no estado de São Paulo, fixando a data do final de 2020. Como esperado, acompanha a distribuição populacional dos municípios.

FIGURA 37 - Casos acumulados por 100 mil habitantes em 31/12/2020



4.4 ANÁLISE COMPARATIVA DOS DADOS DO PROJETO COVID-19 INFO TRACKER SOBRE VACINAÇÃO

O foco deste trabalho desde o início foi acompanhar a relação entre as categorias de mobilidade e a evolução regional da pandemia. A vacinação iniciou-se no estado (e no país) por volta de um ano após o aparecimento dos primeiros casos, no final de janeiro de 2021, por iniciativa do governo estadual e por isso seus efeitos só começariam a surgir no segundo semestre (este trabalho teve sua elaboração iniciada antes de julho/2021).

Porém, a evolução da campanha de vacinação – que, como não podia deixar de ser, sofreu oposição do governo federal e dependeu das gestões estaduais e municipais para seguir adiante – produziu muitas massas de dados interessantes que não poderiam deixar de ser analisadas.

Os números de óbitos a nível nacional, que se aproximaram da triste cifra de 5000 em 24 horas, no início do mês de abril/2021, mostraram uma forte evolução positiva a partir do segundo semestre, à medida que a população, apesar do negacionismo e das interferências políticas e mentiras divulgadas pelo presidente da república, aderiu massivamente à campanha de vacinação. Atualmente, estamos vendo (dezembro/2021) números diários de óbitos, no país, na casa dos 3 dígitos; novas ondas e variantes poderosas, como o ômicron, podem alterar esses números, mas as vacinas já se mostraram eficientes no caso da Delta e outras cepas.

Neste item, concentraremos nossa análise nos números curados e apresentados pelo projeto COVID-19 Info Tracker. Não realizamos distinção entre as marcas/origens das vacinas; consideramos a data efetiva de imunização como 15 dias após a aplicação da segunda dose ou dose única.

Para esta análise, foi criada somente a matriz de correlações, onde comparamos:

- Taxa de isolamento (medida pelo SIMI – Sistema de Monitoramento Inteligente do governo de SP)
- Novos casos
- Novos óbitos
- Novos imunizados

O código fonte que gerou os gráficos pode ser encontrado em: <https://github.com/rlepsch/MBA-USP-TCC-2021> no notebook ANALISE005.

FIGURA 38 - Mapa e matriz de correlações para vacinação em Americana

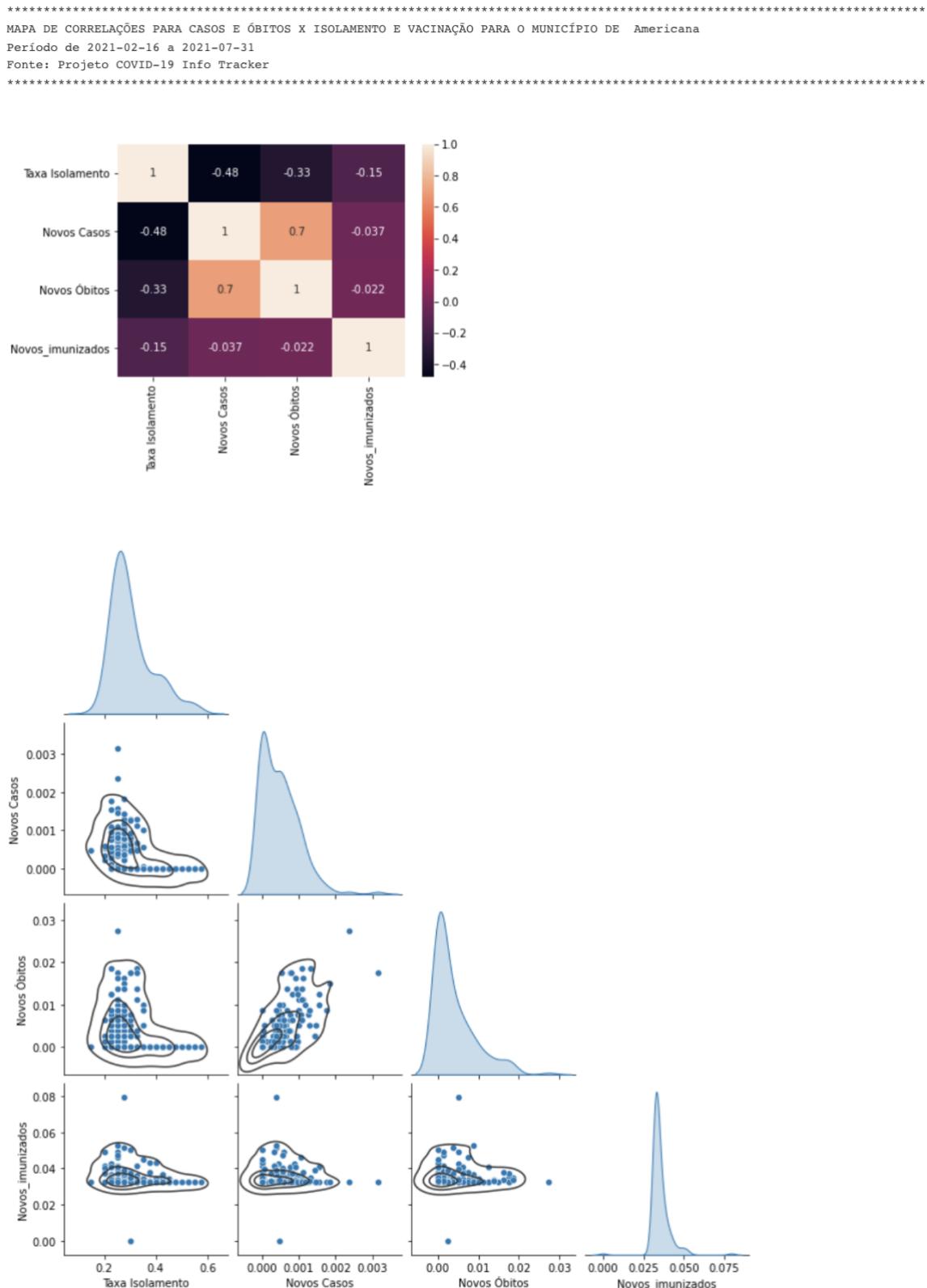


FIGURA 39 - Mapa e matriz de correlações para vacinação no município de São Paulo

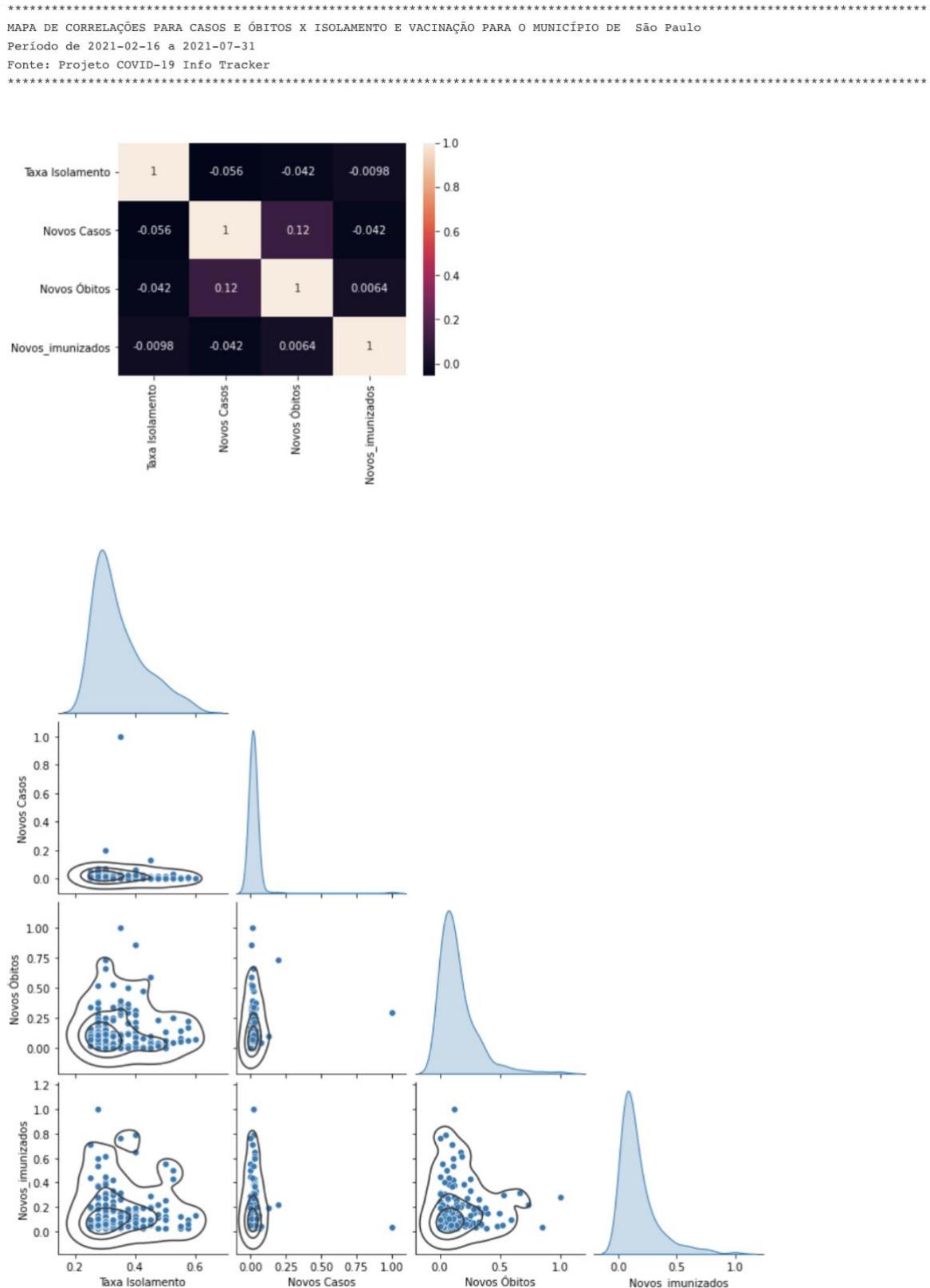
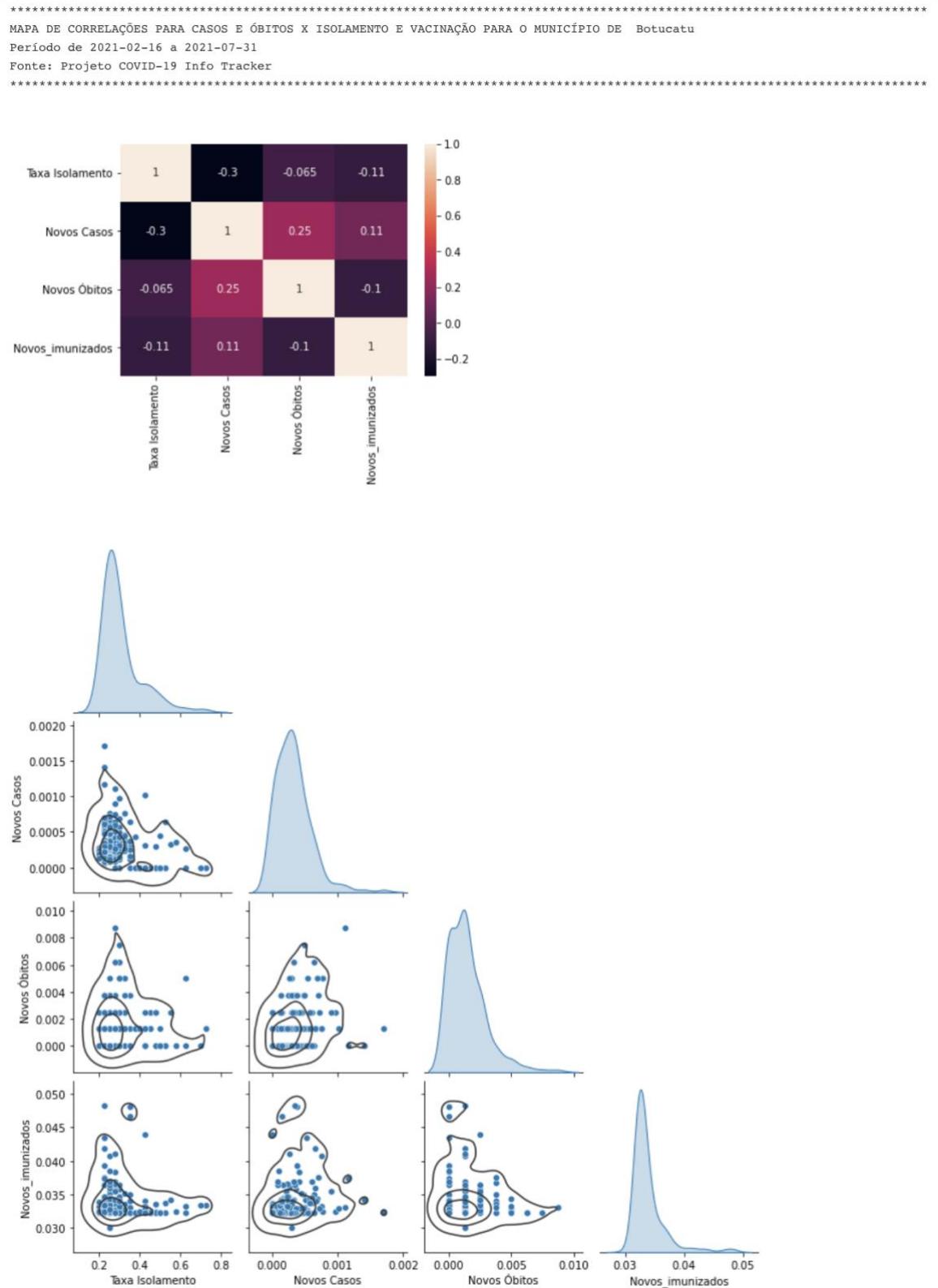


FIGURA 40 - Mapa e matriz de correlações para vacinação em Botucatu



5 CONCLUSÃO

5.1 RESULTADOS E IMPACTOS ESPERADOS – VISÃO PRESCRITIVA

Voltamos a reforçar aqui que os estudos apresentados foram realizados entre os meses de abril e dezembro de 2021, trabalhando com dados produzidos desde fevereiro de 2020, quando já estava caracterizada e estabelecida a pandemia da COVID-19 no mundo, e que foram também gerados durante a pesquisa. Daí a importância em pontuar o momento da extração e preparação dos dados.

Medidas a serem tomadas e políticas a serem seguidas, motivadas pelo acompanhamento da evolução dos números da pandemia, têm se tornado realidade em diversos países e estados/municípios brasileiros. A vacinação tem tido o impacto maior, mostrando-se ainda, até o presente momento, extremamente desigual em sua distribuição global. Como efeitos positivos da imunização na América do Norte e Europa, eventos com presença do público voltaram a acontecer, sempre de forma cuidadosamente supervisionada, no segundo semestre de 2021, embora o cenário possa mudar nos próximos meses (início de 2022) devido à poderosa variante ômicron, neste momento (dezembro/2021) ainda intensamente estudada, e que pode levar a pandemia ao seu terceiro ano. Bom lembrar que as variantes mais antigas, já bem conhecidas, ainda causam ondas de casos e óbitos em locais onde a vacinação é incipiente por questões de falta de recursos ou, pior, de movimentos negacionistas. Nesta véspera de início do inverno no hemisfério norte, estamos vendo por lá até mesmo o surgimento de epidemias paralelas, como a gripe comum, em outros tempos tão facilmente controlada.

Com relação à mobilidade populacional, e voltando o nosso foco para a situação do Brasil e de nosso estado, estamos vendo um fenômeno preocupante: os brasileiros de um modo geral, independente de terem adotado ou não o isolamento social ou tomado a vacina, “decretaram” o fim da pandemia – fato que está longe de acontecer – e ocupam livremente shopping centers, restaurantes, praias, ruas de comércio popular. Como pontuado em nosso estudo, o segundo semestre de 2021 não teve os números da mobilidade analisados, devido a atuação dos primeiros efeitos da vacinação. Sabemos que os números de casos e óbitos tiveram queda sensível (como visto em nossa análise sobre a vacinação, e como podemos ver no noticiário); por outro lado, países europeus já enfrentavam novas ondas de variantes anteriores do vírus, devidas a reabertura geral do comércio, antes mesmo do surgimento da preocupante

variante ômicron, foco de alertas da OMS e de especialistas em todo o mundo e no Brasil. Governos oscilam, indecisos quanto às medidas a tomar; o Carnaval 2022 segue suspenso, embora iniciativas populares possam vir a acontecer, dentro do desprezo geral da população pelos alertas de especialistas. Se plotarmos as curvas de mobilidade atuais, certamente encontraremos picos equivalentes a épocas de absoluta normalidade.

Centros de pesquisa epidemiológica do mundo todo – incluindo aqui o Instituto de Ciências Biomédicas da USP – seguem trabalhando dia e noite para isolar e entender a estrutura e forma de ação das variantes do vírus, o que tem se mostrado um formidável desafio, especialmente num país como o nosso, que tem sofrido nos últimos dois anos um lamentável desmonte da educação, da ciência e da tecnologia. Neste aspecto as universidades públicas paulistas têm sido um núcleo, não somente de excelência científica, mas também de resistência contra esse desmonte, juntamente com entidades de apoio como a FAPESP e o CEMEAI. E vemos também a chama científica de nossos profissionais brilhar em centros no exterior, onde, devidamente valorizados, produzem resultados ainda mais brilhantes – a equipe que identificou a variante ômicron, no Centro para Respostas a Epidemias e Inovação da África do Sul, é chefiada pelo brasileiro Túlio de Oliveira, diretor do centro.

Os números aqui mostrados evidenciam que, sim, a mobilidade está fortemente correlacionada com os casos da doença, e como essa correlação está distribuída. Como apontamos em nossas considerações iniciais, a correlação deve ser utilizada como um apontador, um sinal amarelo, de onde devemos olhar e concentrar medidas, sempre monitorando em tempo real os efeitos. O encontro entre a volta precoce da população às suas atividades normais e a nova variante ômicron – que nada indica que será a última – poderá voltar a elevar os números diários de casos. As vacinas, não custa lembrar, foram desenvolvidas numa janela de tempo extremamente reduzida, com pesquisadores correndo contra o relógio, condições muito difíceis de testes e a forte oposição de movimentos negacionistas; e, portanto, ainda não se conhece totalmente seu grau de eficácia contra um vírus que, convenhamos, tem se mostrado um mecanismo formidável em termos de adaptação e mutabilidade.

E contra tal mecanismo, aproveitemos essa oportunidade para desenvolver, além das ferramentas médicas e epidemiológicas, o ferramental matemático e computacional que as apoia. Desde visualizações de dados como aquelas que aqui apresentamos, e que deveriam abastecer painéis atualizados em tempo real em salas de gestão pública; até ferramentas baseadas em arquiteturas de redes neurais profundas, capazes de analisar imagens médicas com

alta velocidade e precisão, ou produzir agentes que busquem e analisem padrões em grandes massas de dados, indicando pontos de atenção.

5.2 CONSIDERAÇÕES FINAIS

Procuramos, neste trabalho de pesquisa através de dados, extrair dos mesmos um panorama mostrando, com critérios objetivos, as relações entre a pandemia da COVID-19 e as medidas de combate a mesma, buscando substituir a polêmica da efetividade ou não de tais medidas por mensuração, análise visual e exploratória, estatística e modelos. Tal fenômeno global – segundo alguns, o primeiro efetivamente com essa escala (Fernando Pires, Fiocruz) tem abastecido uma quantidade formidável de bases de dados por todo o mundo, motivando desde estudiosos amadores – hobbistas dos números e da estatística – até acadêmicos em pesquisas de pós-doutorado e cientistas de dados ou da computação em projetos de monitoramento como o SP COVID-19 Info Tracker. Iniciativas fundamentais, de alto valor e que na maioria dos casos aconteceram sem apoio de governos locais ou federal, ou às vezes mesmo contando com dificuldade de acesso aos dados por fatores políticos. E, em tempo, tendo que ainda por cima enfrentar o terrorismo cibernético, na forma do ataque de hackers que tirou do ar bases de dados em dezembro de 2021 e inutilizou aplicativos de comprovação de vacinação e até mesmo a análise dos dados. Tivéssemos estendido a nossa coleta até o final de 2021/início de 2022, e o trabalho de análise teria sido seriamente prejudicado.

Esta é uma pequena contribuição à Ciência, salvadora maior de nossas vidas, e tão infelizmente atacada nestes tempos de negacionismo, teorias da conspiração, disseminação de notícias falsas e extremismo político e religioso, onde a opinião de um especialista altamente gabaritado em epidemiologia, com décadas de estudo e muitos trabalhos publicados, chega a valer menos do que a de arautos do obscurantismo, espalhada e compartilhada sem o menor critério pelas redes sociais.

Entre a ignorância fanática e o conhecimento científico, fiquemos com o último. Somente ele nos permitirá chegar ao futuro.

REFERÊNCIAS

- AMARAL, Fábio; CASACA, Wallace; OISHI, Cássio M.; CUMINATO, José A. **Simulating Immunization Campaigns and Vaccine Protection Against COVID-19 Pandemic in Brazil.** IEEE Access, Digital Object Identifier 10.1109/ACCESS.2021.3112036, versão atual 17/09/2021.
- AMARAL, Fábio; CASACA, Wallace; OISHI, Cássio M.; CUMINATO, José A. **Towards Providing Effective Data-Driven Responses to Predict the Covid-19 in São Paulo and Brazil.** Sensors, 21(2), 540, 2021.
- BARRIA-SANDOVAL, Claudia; FERREIRA, Guillermo; BENZ-PARRA, Katherine; LOPEZ-FLORES, Pablo. **Prediction of confirmed cases of and deaths caused by COVID-19 in Chile through time series techniques: A comparative study.** PLOS One Open Access, Peer-reviewed research article, 29/04/2021. <https://journals.plos.org/plosone/article?id=10.1371/journal.pone.0245414> - acessado em 29/06/2021.
- BASELLINI, Ugofillippo; ALBUREZ-GUTIERREZ, Diego; DEL FAVA, Emanuele; PERROTTA, Daniela; BONETTI, Marco; CAMARDA, Carlo G.; ZAGHENI, Emilio. **Linking excess mortality to mobility data during the first wave of COVID-19 in England and Wales.** Science Direct – SSM – Population Health, 2021. <https://www.sciencedirect.com/science/article/pii/S2352827321000744?via%3Dihub> – acessado em 29/06/2021.
- BERGMAN, Nittai K.; FISHMAN, Ram. **Correlations of Mobility and Covid-19 Transmission in Global Data.** <https://doi.org/10.1101/2020.05.06.20093039>. 02/06/2020, medRXiv – The preprint server for health sciences. Acessado em 06/01/2022.
- BERTAGNA, Julia. **Estudo do Excesso de Mortes nas cidades de São Paulo - SP e Campinas - SP durante a pandemia de Covid-19.** UNICAMP – IMECC – Dept. Matemática Aplicada, 2021.
- CARTENI, Armando; DI FRANCESCO, Luigi; MARTINO, Maria. **How mobility habits influenced the spread of the COVID-19 pandemic: Results from the Italian case study.** Science Direct - Science of the Total Environment - 741 (2020) 140489. <https://doi.org/10.1016/j.scitotenv.2020.140489>. 0048-9697/© 2020 Elsevier B.V. All rights reserved.
- COLNAGO, Marilaine; CASACA, Wallace; OISHI, Cássio M.; AMARAL, Fábio V.; CUMINATO, José A. **Lockdown contra a Covid-19 funciona ou não? Um estudo de caso com o município de Araraquara.** Proceeding Series of the Brazilian Society of Computational and Applied Mathematics. 2021.
- CROKIDAKIS, Nuno. **Data analysis and modeling of the evolution of COVID-19 in Brazil.** Instituto de Física, Universidade Federal Fluminense, Niterói – RJ, Brasil, 30/03/2020.

https://www.researchgate.net/publication/340270886_Data_analysis_and_modeling_of_the_evolution_of_COVID-19_in_Brazil - acessado em 29/06/2021.

GOLLOM, Mark. **Yes, pandemic fatigue is an issue, experts say. Will Omicron make it worse?** CBC News, Toronto, 18/12/2021. <https://www.cbc.ca/news/health/pandemic-fatigue-omicron-covid-19-1.6290026>, acessado em 07/01/2022.

KARTAL, Mustafa Tevfik; DEPREN, Özer; DEPREN KILIÇ, Serpil. **The relationship between mobility and COVID-19 pandemic: Daily evidence from an emerging country by causality analysis.** Transportation Research Interdisciplinary Perspectives, Volume 10, June 2021, 100366. <https://doi.org/10.1016/j.trip.2021.100366>. Copyright © 2022 Elsevier B.V. or its licensors or contributors. Acessado em 06/01/2022

MC KINNEY, Wes. **Python for Data Analysis - Data Wrangling with Pandas, NumPy, and IPython** (Second Edition). O'Reilly, 2018.

NICOLELIS, Miguel A. L; RAIMUNDO, Rafael L. G.; PEIXOTO, Pedro S.; ANDREAZZI, Cecilia S. **The impact of super-spreader cities, highways, and intensive care availability in the early stages of the COVID-19 epidemic in Brazil.** Scientific Reports – Nature Portfolio. Publicado online em 21/06/2021. <https://www.nature.com/articles/s41598-021-92263-3> - acessado em 23/06/2021.

NIELSEN, Michael. **If correlation doesn't imply causation, then what does?** DDI – Data Driven Intelligence, 23/01/2012. <https://michaelnielsen.org/ddi/if-correlation-doesnt-imply-causation-then-what-does/>. Acessado em 22/09/2021.

NOVELLA, Steven. **Evidence in Medicine: Correlation and Causation.** Science-Based Medicine, 18/11/2009. <https://sciencebasedmedicine.org/evidence-in-medicine-correlation-and-causation/>. Acessado em 22/09/2021.

PEARL, Judea; MACKENZIE, Dana. **The Book of Why – The New Science of Cause and Effect.** Penguin, 2019.

QUOC-VIET PHAM, DINH C. NGUYEN, THIEN HUYNH-THE, WON-JOO HWANG, PUBUDU N. PATHIRANA. **Artificial Intelligence (AI) and Big Data for Coronavirus (COVID-19) Pandemic: A Survey on the State-of-the-Arts.** IEEE Access, July 2020 PP(99):1-1

RADECIC, Dario. **How to Use Python and MissForest Algorithm to Impute Missing Data - Step-by-step guide on using Random Forests to handle missing data.** <https://towardsdatascience.com/how-to-use-python-and-missforest-algorithm-to-impute-missing-data-ed45eb47cb9a> - Towards Data Science, acessado em 05/11/2020.

WORLD HEALTH ORGANIZATION. **DIRETRIZES INTERNACIONAIS PARA A CERTIFICAÇÃO E CLASSIFICAÇÃO (CODIFICAÇÃO) DA COVID-19 COMO CAUSA DE MORTE – Baseadas na CID – Classificação Estatística Internacional de Doenças.** Abril/2020.

https://www.who.int/classifications/icd/Guidelines_Cause_of_Death_COVID-19-20200420-PT_Apr_24.pdf?ua=1#:~:text=COVID%2D19%20deve%20ser%20registada,ou%20contribu%C3%ADdo%20para%20a%20morte. – acessado em 26/06/2021.

XU Zuo; YONG Chen; OHNO-MACHADO, Lucila; HUA Xu. **How do we share data in COVID-19 research? A systematic review of COVID-19 datasets in PubMed Central Articles.** *Briefings in Bioinformatics*, Volume 22, Issue 2, March 2021, Pages 800–811, Oxford Academic. <https://doi.org/10.1093/bib/bbaa331> - acessado em 30/06/2021.

WEBSITES

<https://www.spcovid.net.br/> - Plataforma COVID-19 Info Tracker. Acessado em 24/06/2021.

<https://www.saopaulo.sp.gov.br/planosp/simi/dados-abertos/>. Acessado em 04/07/2021.

Google LLC "Google COVID-19 Community Mobility Reports".

<https://www.google.com/covid19/mobility/> Acessado em: 30/05/2021.

Our World in Data (University of Oxford) - <https://ourworldindata.org/explorers/coronavirus-data-explorer?zoomToSelection=true&time=2020-03-01..latest&facet=none&pickerSort=asc&pickerMetric=location&Metric=Confirmed+cases&Interval=7-day+rolling+average&Relative+to+Population=true&Align+outbreaks=false&country=BRA~CHL~GBR>. Acessado em: 09/09/2021.

<https://coronavirus.jhu.edu/map.html> - COVID-19 Dashboard by the Center for Systems Science and Engineering (CSSE), University Johns Hopkins, Baltimore, USA. Acessado em 20/05/2021.

http://ces.ufpel.edu.br/vestibular/download/2004i/prova_f1_b.pdf, Tabela de mortos por país na Segunda Guerra Mundial, publicada originalmente no Almanaque Abril Edição 1999.

<https://especiais.g1.globo.com/bemestar/vacina/2021/mapa-brasil-vacina-covid/> - Mapa da vacinação contra a COVID-19 no Brasil – Site G1/Globo. Acessado em 21/05/2021.

<https://www.dw.com/pt-br/em-novo-recorde-brasil-registra-4249-mortes-por-covid-19-em-24-horas/a-57139302#:~:text=Coronav%C3%ADrus-,Em%20novo%20recorde%2C%20Brasil%20registra%204.249%20mortes%20por%20covid%2D19,mil%20novos%20casos%20da%20doen%C3%A7a.> - Em novo recorde, Brasil registra 4.249 mortes por covid-19 em 24 horas. Acessado em 30/05/2021.

<http://www.revistahcsm.coc.fiocruz.br/reflexoes-insones-em-noites-pandemicas/> - Fernando A. Pires Alves, REFLEXÕES INSONES EM NOITES PANDÊMICAS – Blog da revista História Ciência Saúde – Fiocruz/Manguinhos. Acessado em 16/06/2021.

https://support.google.com/covid19-mobility/answer/9824897?hl=pt-BR&ref_topic=9822927#zippy=%2Ccomo-a-data-do-n%C3%A9mero-principal-est%C3%A1-relacionada-%C3%A0-data-do-relat%C3%B3rio – Página de apoio e documentação do Relatório de Mobilidade Global do Google. Acessado em 17/06/2021

[https://www.who.int/emergencies/diseases/novel-coronavirus-2019/technical-guidance/naming-the-coronavirus-disease-\(covid-2019\)-and-the-virus-that-causes-it](https://www.who.int/emergencies/diseases/novel-coronavirus-2019/technical-guidance/naming-the-coronavirus-disease-(covid-2019)-and-the-virus-that-causes-it). Página da Organização Mundial da Saúde. Acessado em 25/06/2021.

<https://www.cdc.gov/>. Site Centers for Disease Control and Prevention (EUA). Acessado em 27/06/2021.

<https://www.ons.gov.uk/peoplepopulationandcommunity/healthandsocialcare/conditionsanddiseases>. Site do Office for National Statistics (Reino Unido). Acessado em 27/06/2021.

<https://olhardigital.com.br/2021/02/05/videos/covid-19-nova-inteligencia-artificial-pode-projetar-vacinas-em-segundos/> . Acessado em 29/06/2021.

<https://www.uol.com.br/tilt/colunas/felipe-zmoginski/2021/01/21/uso-de-ia-e-big-data-explicam-ascensao-da-china-na-pesquisa-de-vacinas.htm>. Acessado em 29/06/2021.

<https://portal.fiocruz.br/pergunta/qual-e-o-tempo-de-incubacao-do-novo-coronavirus>. Acessado em 08/07/2021.

<https://brasil.elpais.com/brasil/2020-07-24/sem-alarde-sao-paulo-muda-divulgacao-de-obitos-da-covid-19-e-especialistas-criticam.html>. Acessado em 03/08/2021.

<https://www.statology.org/how-to-read-a-correlation-matrix/> - Acessado em 15/09/2021

<https://datahero.com/blog/2013/11/21/if-correlation-isnt-causation-what-is-it/> - Acessado em 15/09/2021

<https://en.wikipedia.org/wiki/Correlation> - Acessado em 15/09/2021

https://en.wikipedia.org/wiki/Correlation_does_not_imply_causation - Acessado em 15/09/2021

<https://g1.globo.com/bemestar/vacina/noticia/2020/12/07/reino-unido-anuncia-que-vacinacao-contra-covid-19-comeca-nesta-terca-8.ghtml> - Acessado em 18/11/2021

<http://www.bibliotecavirtual.sp.gov.br/temas/sao-paulo/sao-paulo-populacao-dos-municípios-paulistas.php>, acessado em 21/07/2021

<https://jornal.usp.br/ciencias/pesquisadores-da-usp-isolam-variante-omicron-do-sars-cov-2/>, acessado em 17/12/2021

<https://exame.com/ciencia/brasileiro-ajuda-rastrear-nova-variante-da-covid-na-africa-do-sul/>, acessado em 18/12/2021