

SCC0245

Processamento Analítico de Dados

Prof.^a Dr.^a Cristina Dutra de Aguiar
PAE João Pedro de Carvalho Castro

Trabalho prático 1

Dados sobre saúde e clima no Brasil

Alunos:

- Dalton Hiroshi Sato nUSP 11275172
- Lucas Yuji Matubara nUSP 10734432
- Savio Duarte Fontes nUSP 10737251

Sumário

Descrição do problema	3
Tema	3
Processos de negócios	3
Fatos:	3
Dimensões:	4
Povoamento:	5
Esquemas	7
Esquema Estrela 1	7
Esquema Estrela 2	8
Constelação de fatos	9
Consultas	10
Referências Bibliográficas	10

Descrição do problema

Tema

Com uso de bases de dados provenientes do kaggle e do SUS, nosso tema é sobre a saúde no Brasil, em específico a partir do ano de 2020, com uso de informações sobre relatórios de doenças respiratórias e informações sobre variações do clima e do tempo no Brasil.

Como o mundo está passando por um processo de transição epidemiológica, no qual algumas atividades estão retornando à normalidade, seria de grande interesse conhecer sua evolução e possivelmente detectar algumas características e peculiaridades.

Processos de negócios

O DataWarehouse (DW) possuirá foco em Clima e sua influência nas Doenças Respiratórias, em especial, no Sars-Cov-2.

Por meio dos datasets do Opendatasus (Datusus) e do Kaggle (INMET) os integrantes farão o ETL (Extract, Transform, Load) pelo Pentaho Data Integration, para tratar os dados obtidos.

Essas informações trabalhadas seriam importantes a profissionais de saúde como: médicos, enfermeiros, técnicos em enfermagem, farmacêuticos, o que facilitaria o atendimento dos profissionais disponibilizando uma análise anterior à consulta, também ajudaria no controle das ações governamentais no isolamento social.

Serão feitas buscas quando os profissionais de saúde acharem necessário informações quanto às dimensões trabalhadas para aumentar a produtividade em seus trabalhos e se atualizarem sobre o cenário atual.

Fatos:

- Saúde
 - Atributos:
 - Registro Ocorrência (Aditivo, por meio de SUM, de modo a contabilizar a quantidade de entradas na tabela de fatos)
 - Dimensões:
 - Local
 - Estado Paciente
 - Paciente
 - Data
 - Grupo Sintomas
 - Grupo Doenças
- Vacinação
 - Atributos:
 -
 - Dimensões:
 - Paciente

- Data
 - Vacinas
- Clima:
 - Atributos:
 - Precipitação Total (Aditivo, por meio de SUM para calcular a precipitação total, mediante um filtro por data e/ou região)
 - Temp Máx (Não aditivo)
 - Temp Mín (Não aditivo)
 - Umidade Máx (Não aditivo)
 - Umidade Mín (Não aditivo)
 - Vento Máx (Não aditivo)
 - Dimensões:
 - Local
 - Data

Dimensões:

- Local
 - Cidade - Da admissão do paciente e da medição do clima
 - Estado - Da admissão do paciente e da medição do clima
 - Região - Da admissão do paciente e da medição do clima
 - País - Brasil
 - Hospital- Da admissão do paciente (exclusivo para tabela de saúde)
 - Relacionamento:
 - All > país > região > estado > cidade
- Estado Paciente
 - Estado doente - Se o paciente já foi diagnosticado com alguma doença
 - Estado óbito - Se o paciente foi a óbito, entre data de admissão e data da formação do relatório
- Paciente
 - Nome - Do paciente
 - Idade - Do paciente
 - Sexo - Do paciente
 - Gestante - Se está grávida ou não
 - CPF - Identificação do paciente
 - CEP - CEP do paciente
- Data
 - Data Completa - data em formato DD/MM/AAAA
 - Dia - o dia da data
 - Dia da semana - Segunda, Terça, Quarta etc
 - Semana - a semana em relação ao ano no momento
 - Quinzena - a quinzena em relação ao ano no momento
 - Mês - o mês da data
 - Mês-ano
 - Mês nome
 - Trimestre - o trimestre em relação ao ano no momento
 - Trimestre-ano

- Semestre - o semestre em relação ao ano no momento
- Semestre-ano
- Ano - o ano da data
- Relacionamento:
 - ALL > ano > semestre > trimestre > mês > quinzena > semana > dia
- Sintomas
 - Chave sintoma - chave do sintoma
 - Sintoma - Nome do sintoma
- Ponte Grupo Sintoma
 - Grupo sintoma - Grupo de sintomas apresentados pelo paciente
 - Sintoma - Chave do sintoma
 - Fator de ponderação - Auxilia na construção de somatórios
- Doença
 - Chave doença - chave da doença
 - Doença - Nome da doença
- Ponte Grupo Doença
 - Grupo doença - Grupo de doenças apresentados pelo paciente
 - Doença - Chave da doença
 - Fator de ponderação - Auxilia na construção de somatórios
- Vacinas
 - Nome_empresa - Nome da empresa que produz a vacina
 - Dose - Tipo da dose tomada (1ª, 2ª, reforço)

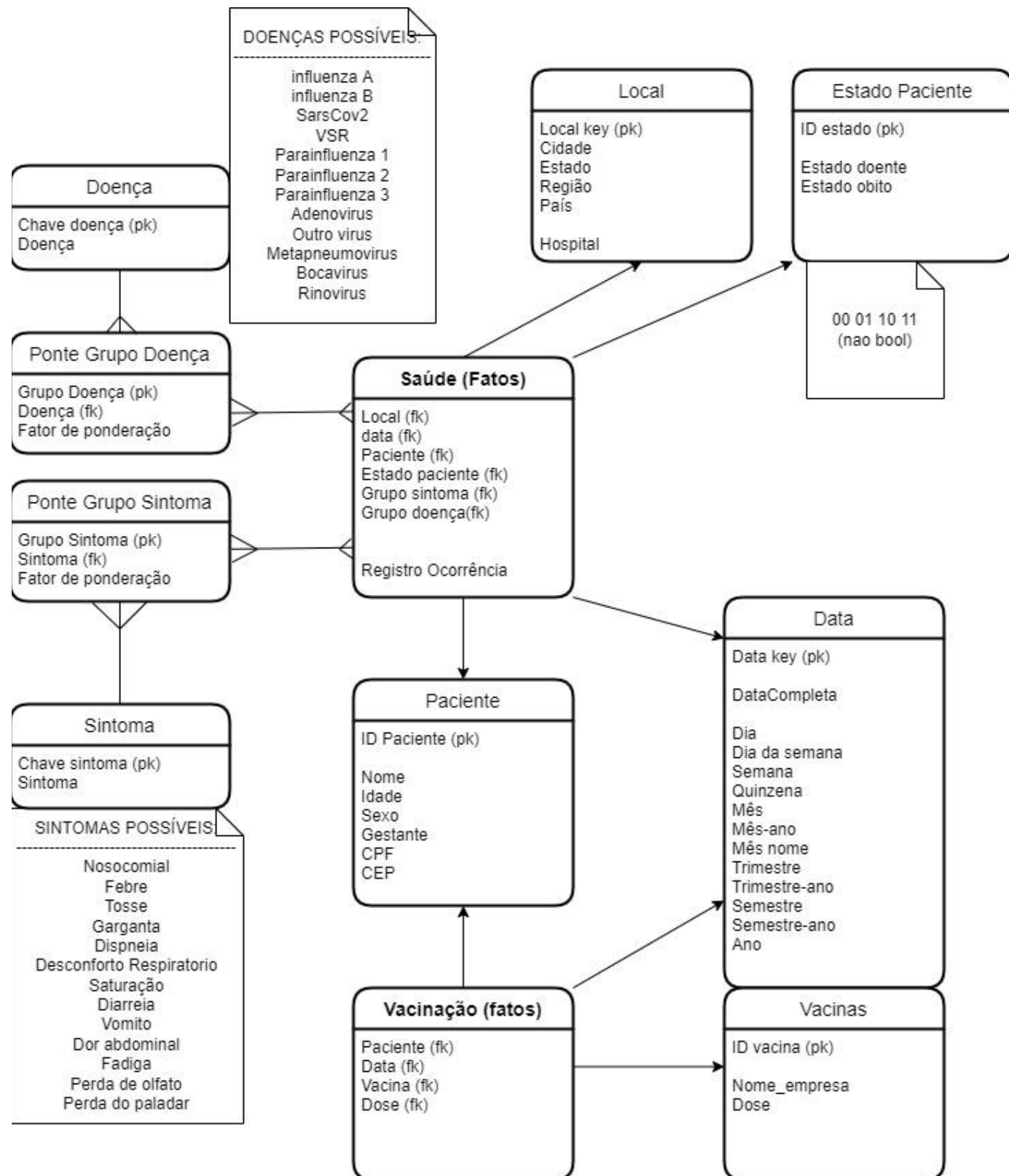
Povoamento:

- Dimensão Data será povoada por meio de programação
- Dimensão Local será povoada por meio da integração das nossas duas bases de dados, o DataSus e o Kaggle, pegando os atributos referentes à Dimensão Local
- Dimensão Paciente será povoada utilizando os dados referentes aos atributos da Dimensão Paciente que estão na base de dados DataSus
- Dimensão Estado Paciente será povoada utilizando os dados da base do dataSus referentes se ele está doente e se ele veio a óbito
- Dimensão Vacina será povoada utilizando os nomes das vacinas que se encontram na base de dados do DataSus
- Dimensão Grupo Doenças será povoada utilizando os dados da base do dataSus de forma que todos as doenças sejam colocadas na tabela de doenças e os grupos de doenças que os pacientes têm serão colocados na tabela Grupo Doenças
- Dimensão Grupo Sintomas será povoada utilizando os dados da base do dataSus de forma que todos os sintomas sejam colocadas na tabela de sintomas e os grupos de sintomas que os pacientes têm serão colocados na tabela Grupo Sintomas
- Fato Registro Ocorrência recebe um inteiro de valor 1 para fazermos operações aditivas em cima da tabela de fatos
- Fato Precipitação total será povoado utilizando a soma diária das precipitações que estão na base de dados do Kaggle
- Temp Máx será povoado utilizando a temperatura máxima diária que está na base de dados do Kaggle

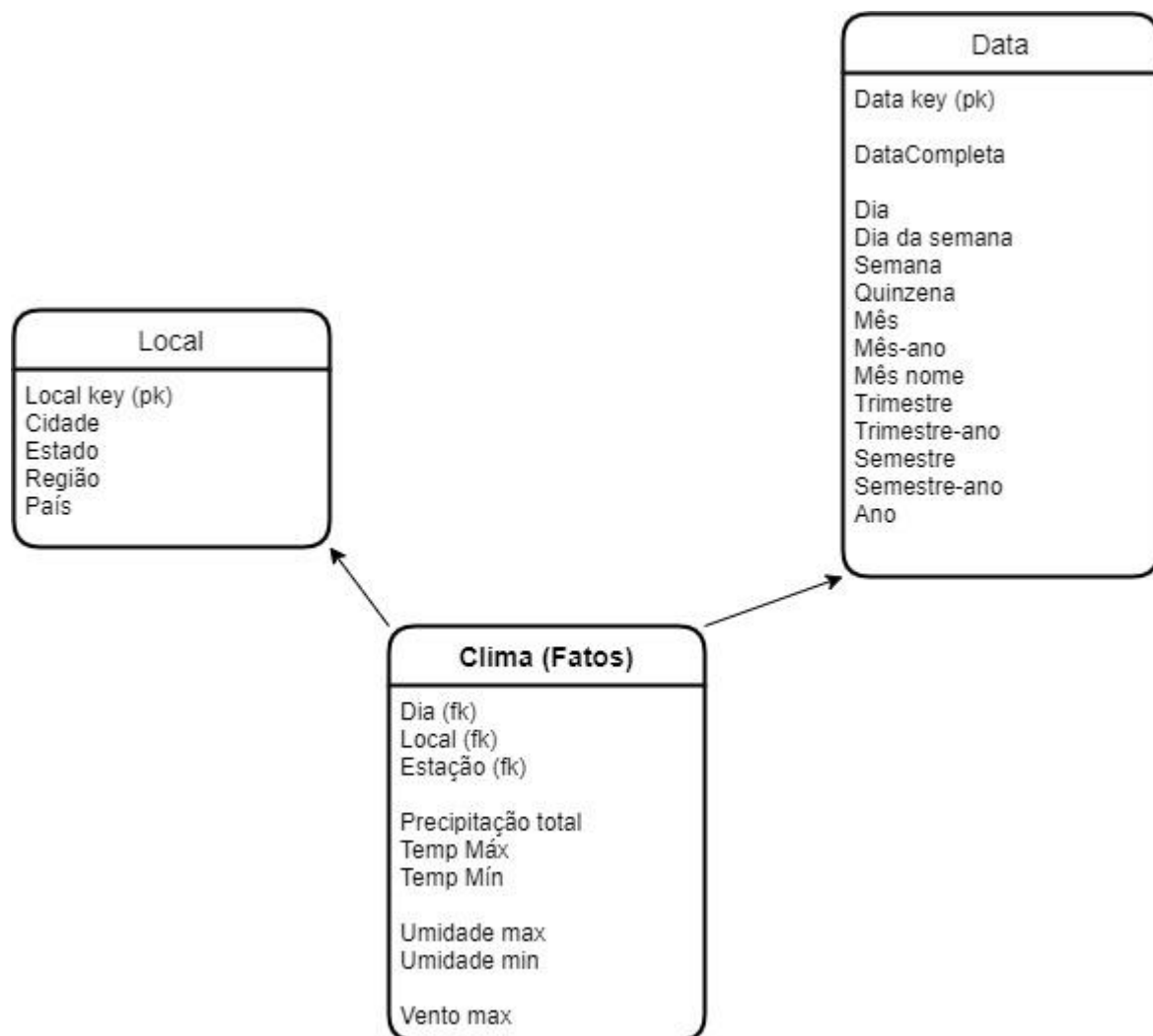
- Temp Mín será povoado utilizando a temperatura mínima diária que está na base de dados do Kaggle
- Umidade Máx será povoado utilizando a umidade máxima diária que está na base de dados do Kaggle
- Umidade Mín será povoado utilizando a umidade mínima diária que está na base de dados do Kaggle
- Vento Máx será povoado utilizando a medição máxima diária do vento que está na base de dados do Kaggle
- Caso seja observado que os dados extraídos não sejam suficientes para uma análise, e que as consultas não apresentem uma quantidade de dados suficientes, dados sintéticos para os esquemas de Vacinação e de Saúde podem ser inseridos.

Esquemas

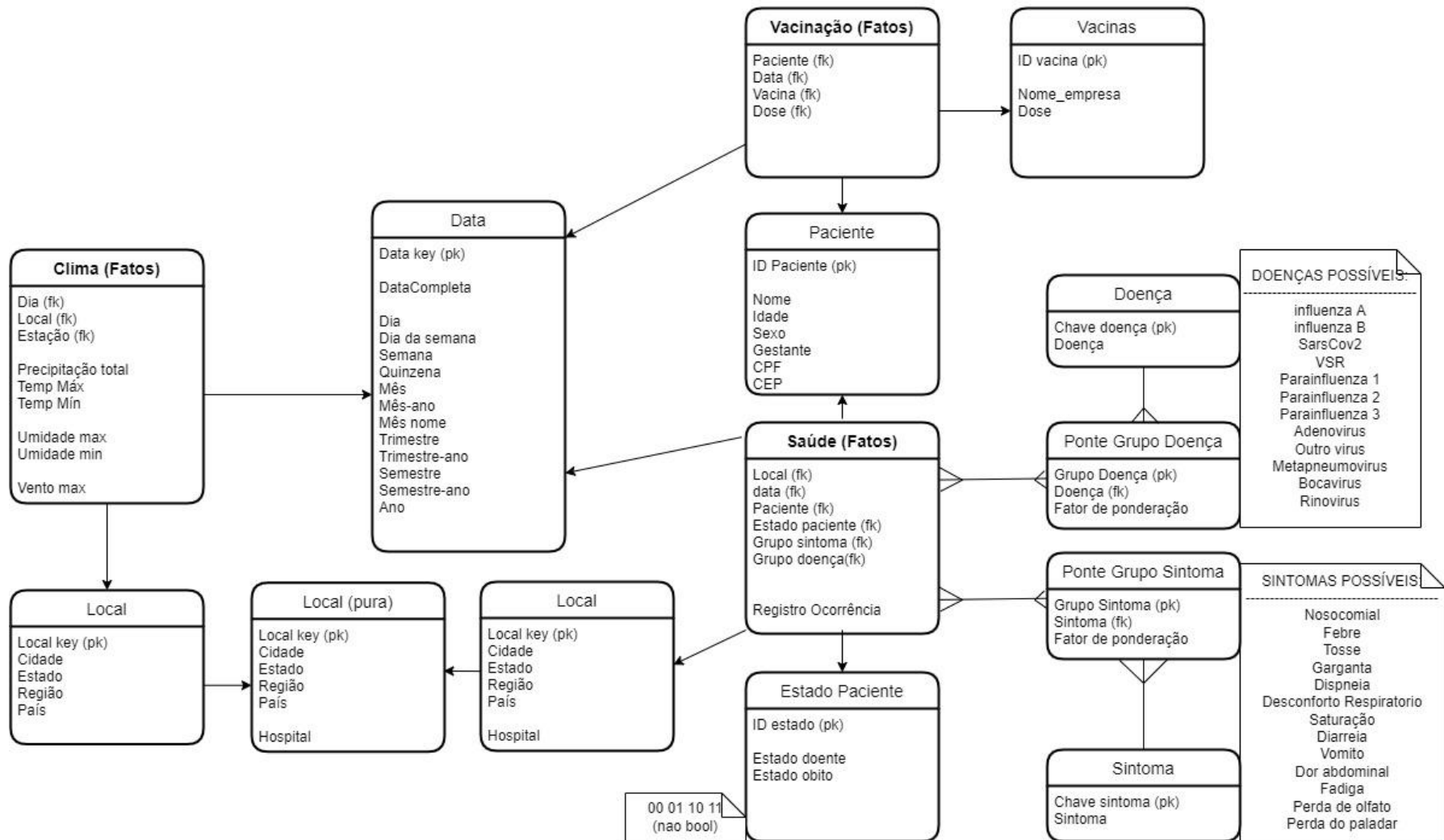
Esquema Estrela 1



Esquema Estrela 2



Constelação de fatos



A constelação de fatos foi projetada de modo que nossos 3 esquemas estrelas (Saúde, Vacinação e Clima) pudessem se interligar. No projeto, Clima se relaciona com Saúde de acordo com a data e com o local, de modo que seja possível fazer uma associação de admissões em hospitais de acordo com a condição climática em certas regiões (ar seco coopera com aparição de alguns sintomas e agravantes de algumas doenças respiratórias).

Ao mesmo tempo, Saúde se relaciona com Vacinação, de modo que seja possível verificar a ocorrência de certas doenças de acordo com o grau de vacinação, em específico, o SarsCov-2.

Consultas

- **Slice and Dice:** Verificar os contágios mensais que aconteceram na cidade de São Paulo no ano de 2020
- **Drill-Across:** Verificar os contágios semanais nas semanas em que a precipitação total foi menor que 100 mm no mesmo local.
- **Roll-Up:** Verificar precipitação total por mês por região.
- **Drill-Down:** Verificar por dia quantos foram os vacinados em um determinado semestre.
- **Pivot:** Verificar número de contágios por trimestre por estado, trocando as perspectivas de tempo e local.

Referências Bibliográficas

Datasus. "SRAG 2021 e 2022 - Banco de Dados de Síndrome Respiratória Aguda Grave - incluindo dados da COVID-19 - OPENDATASUS." *OPENDATASUS*, 2022, <https://opendatasus.saude.gov.br/dataset/srag-2021-e-2022>. Accessed 18 May 2022.

INMET. "Climate Weather Surface of Brazil - Hourly." *Kaggle*, outubro 2021, <https://www.kaggle.com/datasets/PROPPG-PPG/hourly-weather-surface-brazil-south-east-region>. Accessed 18 May 2022.