

Trilhas Exploratórias:

Metodologia de visualização de dados

As **Trilhas Exploratórias** se constituem em uma ferramenta interativa e customizada que oferece visualização exploratória, download e compartilhamento de dados de saúde materna-neonatal empregando dados de poluição de 5565 municípios brasileiros ao longo de sete anos (2012-2019).

SUMÁRIO

Objetivo do projeto Ampla Saúde.....	1
Objetivos específicos da ferramenta.....	2
Fontes dos dados.....	2
Informação relevante sobre as relações entre os dados.....	2
Premissas do projeto.....	3
Definições e indicadores.....	5

REALIZAÇÃO



LAB
VIS



FINANCIAMENTO

BILL & MELINDA
GATES foundation



APOIO

PCDAS

OBJETIVO DO PROJETO AMPLIA SAÚDE

A mortalidade materna, neonatal e em crianças menores de cinco anos tem sido reduzida no Brasil ao longo dos últimos anos, predominantemente graças à implementação de políticas públicas na área de saúde, específicas para o eixo materno-infantil, dentro do escopo do Objetivo de Desenvolvimento Sustentável (ODS) 3. A gestação saudável, um nascimento com acolhimento, o aleitamento materno suficiente e imunizações oportunas são reconhecidos fatores que favorecem o crescimento e desenvolvimento da criança. O ambiente em que se desenvolve o feto é essencial para a vida subsequente da criança. Esse ambiente é principalmente definido pelos determinantes sociais e o sistema de saúde (Figura 1), levando em

consideração fatores demográficos e condições socioeconômicas da mãe que a fazem mais ou menos vulnerável; e no acesso a atendimento especializado, ao cuidado pré-natal e à imunização. As condições ambientais consideram apenas as do entorno direto, tais como o acesso à água potável, tratamento de esgoto e coleta de lixo. No entanto, condições climáticas extremas, desastres ambientais e poluição ambiental, podem ter impacto direto sobre a saúde humana e sobre a disponibilidade e o acesso a serviços de saúde. Os determinantes climático-ambientais são cada vez mais frequentes e seu impacto precisa ser medido no contexto da saúde materna-neonatal. Apoiado no princípio de integralidade do SUS, o projeto Amplia Saúde pretendeu ampliar o modelo de análise dos indicadores de morbimortalidade materno-infantil integrando indicadores relacionados à poluição atmosférica.

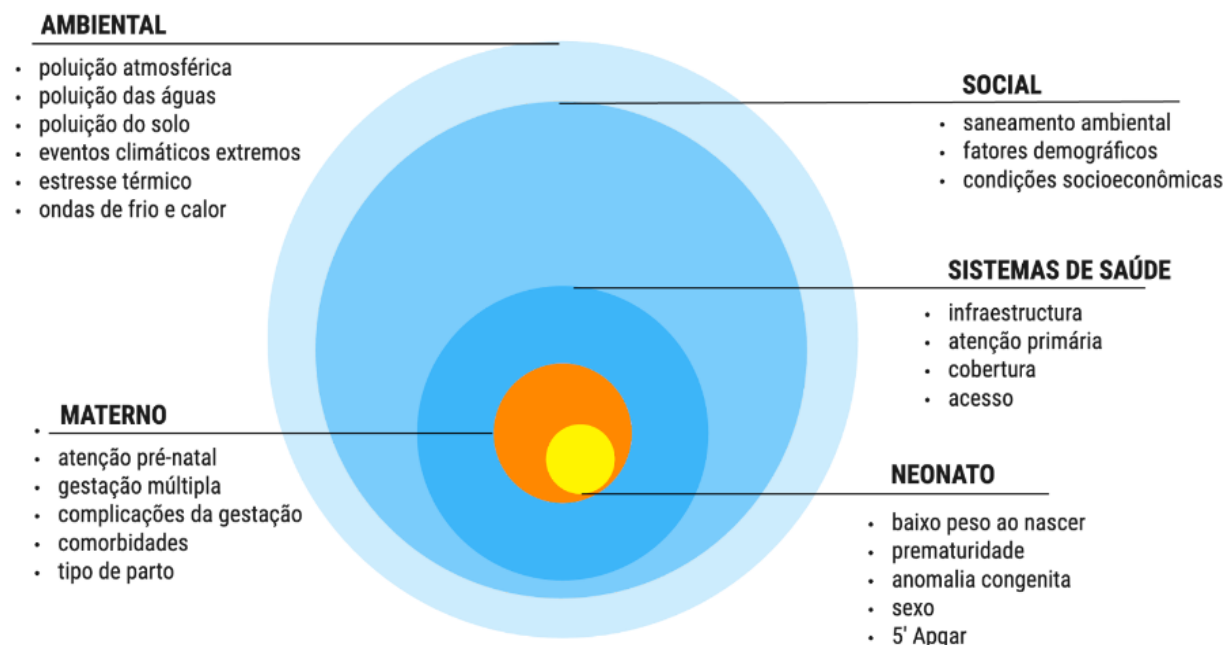


Figura 1. Marco conceitual de análise de morbidade e mortalidade neonatal

OBJETIVOS ESPECÍFICOS DA FERRAMENTA

- Oferecer uma **abordagem inovadora de análise visual** de dados de saúde materna-infantil do período perinatal, tendo em vista fatores ambientais (poluição atmosférica)
- Criar uma **ferramenta interativa que permita explorar** essa relação

FONTE DOS DADOS

A ferramenta de visualização emprega dados disponibilizados abertamente pelos seguintes órgãos governamentais:

- DATASUS, Departamento de Informática do Sistema Único de Saúde do Brasil - Ministério da Saúde
- Programa Queimadas do INPE, Instituto Nacional de Pesquisas Espaciais
- IBGE, Instituto Brasileiro de Geografia e Estatística

A vigilância do óbito infantil e fetal é obrigatória nos serviços de saúde (públicos e privados) que integram o Sistema Único de Saúde (SUS) desde 2010, por portaria do Ministério da Saúde. Assim, especificamente, do DATASUS, foram utilizados dados dos sistemas SIM (Sistema de Informação sobre Mortalidade) e SINASC (Sistema de Informação sobre Nascidos Vivos) a partir de 2012. Além dos dados de mortalidade, também foram empregadas as bases Declaração de Óbitos Fetais (DOFET-SIM), que foram extraídos, tratados e enriquecidos pelo PCDaS (Plataforma de Ciência de Dados aplicada à Saúde – ICICT, Instituto de Comunicação e Informação Científica e Tecnológica em Saúde da Fundação Oswaldo Cruz), utilizando uma metodologia própria de ETL (*Extract, Transform, Load*).

Do INPE, foi utilizado o SISAM, Sistema de Informações Ambientais Integrado à Saúde, desenvolvido através de uma parceria do INPE com a OPAS/OMS (Organização Pan-Americana da Saúde/Organização Mundial da Saúde) e o FUNDEP (Fundação de Desenvolvimento da Pesquisa). Foram utilizados dados de poluentes atmosféricos, material particulado fino de 2,5 micras (MP2,5) de

2012 até 2019, último ano disponibilizado na base. Outros dados, tais como concentração de ozônio, monóxido de carbono, e velocidade e direção do vento, embora disponíveis, não se mostraram consistentes e, portanto, foram abandonados após os primeiros testes. O download dos dados do SISAM foi realizado com o auxílio de um coletor de dados automático (*web scraping crawler*). A base oferece até quatro medições diárias de MP2,5 embora tenham sido encontradas variações nos diferentes municípios. Deste modo, optou-se por uma agregação semanal para os dados de poluição.

Do IBGE foram utilizados o IDH (Índice de Desenvolvimento Humano) para o ano de 2010, último censo realizado até o momento, e a geolocalização dos municípios.

Ao longo de todo o projeto "Amplia Saúde", foram utilizados apenas dados secundários agregados e sem possibilidade de identificação individual, inclusive sem campo comum que pudesse relacionar os eventos nas bases SINASC e do SIM.

INFORMAÇÃO RELEVANTE SOBRE AS RELAÇÕES ENTRE OS DADOS

A ferramenta de dados desenvolvida no âmbito do projeto Amplia Saúde compreende um conjunto de visualizações exploratórias que proporcionam a combinação de um enorme volume de dados do período perinatal com atributos socioeconômicos, geográficos e de incidência da poluição atmosférica. As visualizações exploratórias podem colaborar com a identificação de padrões nos dados, a obtenção de insights e a formulação ou testagem de hipóteses.

Por reunir e combinar dados heterogêneos e em grande quantidade (*big data*) obtidos a partir de bases não vinculadas, estas visualizações não são indicadas para buscar relações diretas de causa e efeito entre poluição atmosférica e desfechos da saúde dos recém nascidos, demandando estudos posteriores que fundamentem as conexões identificadas.

PREMISSAS DO PROJETO

Por tratar de grandes volumes de dados coletados ao longo de vários anos, em diferentes formatos a partir de diferentes localidades, foi necessário fazer um levantamento inicial das bases de dados disponíveis e verificar a sua qualidade, além das possibilidades de agregação e interação, criação de indicadores ou utilização de indicadores existentes. Nesta etapa inicial, foi criado um diagrama na plataforma Miro (Figura 2), onde dados das fontes relacionadas aos temas saúde materna-neonatal e poluição ambiental foram listados. Posteriormente, algumas das bases selecionadas, acabaram não sendo empregadas, como foi o caso do CNES (Cadastro Nacional de Estabelecimentos de Saúde) e do SIH (Sistema de Informações Hospitalares e Ambulatoriais).

A Ferramenta oferece informação de uma forma exploratória, sob demanda e detalhada. É uma camada dirigida àqueles que buscam realizar consultas específicas, formular ou testar hipóteses e tomar decisões orientadas por dados. Portanto, foi necessário criar uma arquitetura customizada para a ferramenta (Figura 3). Para gerar as visualizações interativas, adotamos uma arquitetura onde os dados são armazenados e recuperados a partir de um servidor em nuvem, mas as visualizações são desenhadas por programas que rodam no navegador do usuário. Para contornar problemas de desempenho típicos de sistemas de visualização de grandes volumes de dados, foi introduzida na arquitetura do sistema uma camada adicional (rotulada como “middleware” na Figura 3) que permite comprimir os dados enviados pela rede desde o servidor até o cliente e permitir um acesso mais rápido aos dados.

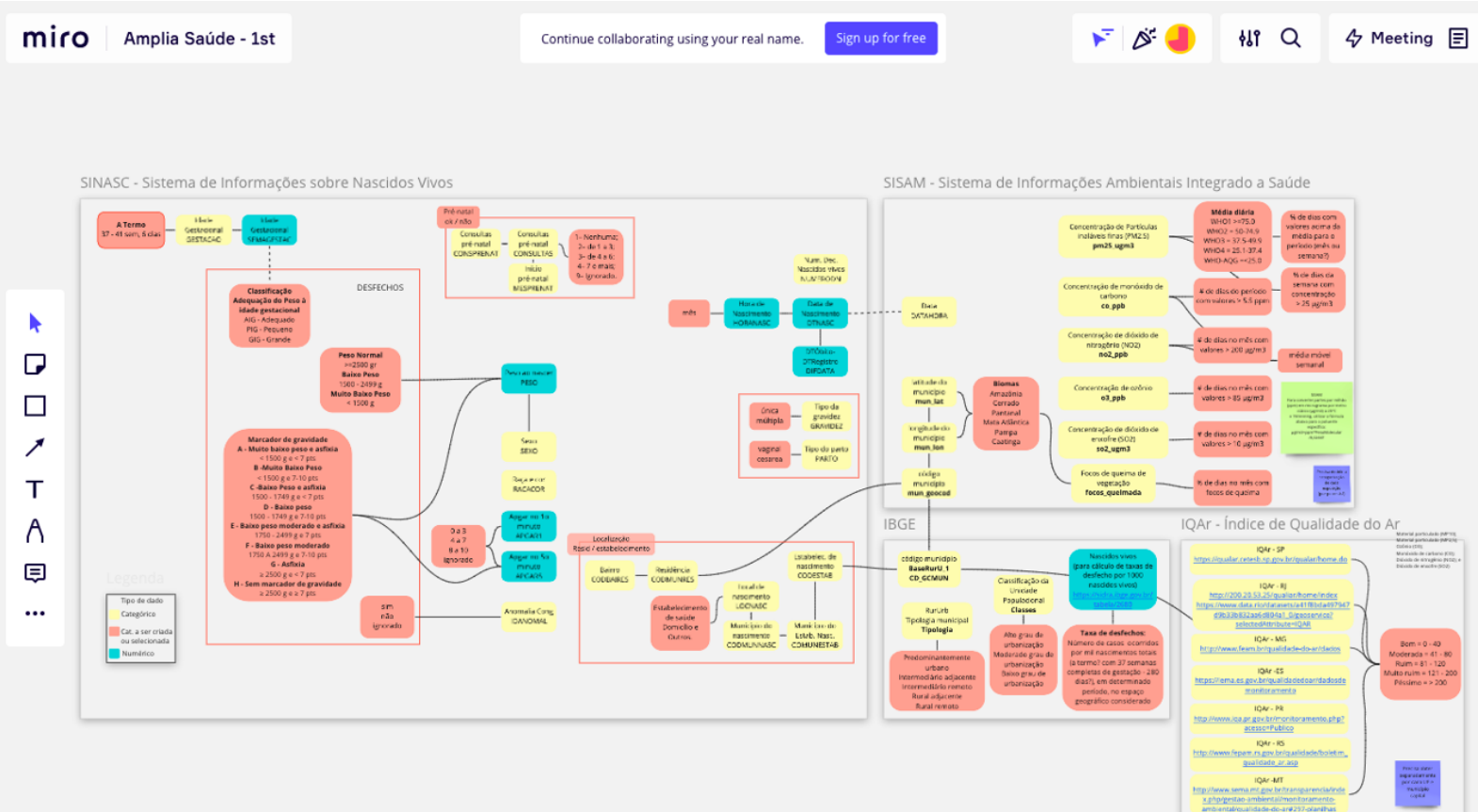


Figura 2. Diagrama relacional das bases de dados

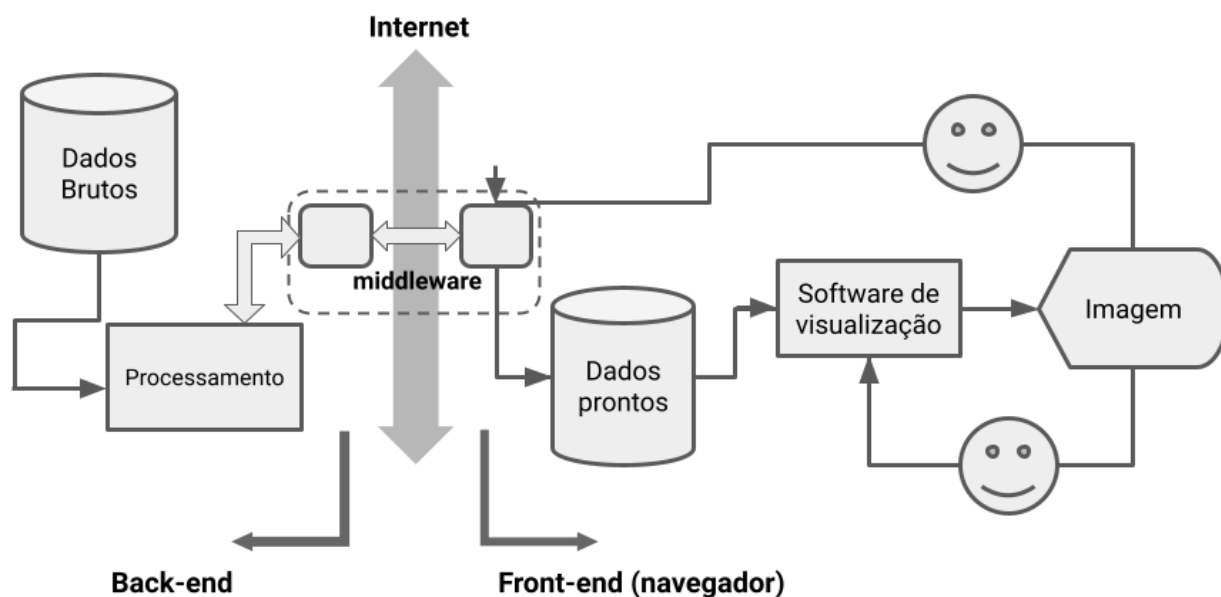


Figura 3. Arquitetura do sistema de dados/ferramenta

Na produção das visualizações, diversas decisões de design foram tomadas à luz do grande volume de dados que pode ser exibido de diferentes maneiras em função das escolhas realizadas pelo usuário. Por exemplo, limitamos em quatro os municípios que podem ser plotados simultaneamente na visualização Trilhas. Adicionalmente, os gráficos de poluentes de municípios com baixo índice de poluição atmosférica podem não mostrar a variação existente quando comparados com municípios mais poluídos, pois o eixo foi fixado para abarcar todos os valores existentes.

Tendo em vista os diversos desafios observados ao longo do processo de estruturação dos dados e criação das visualizações, foram determinadas cinco premissas de projeto, que serão detalhadas a seguir.

AGREGAÇÃO DOS DADOS E APROXIMAÇÃO DE RELAÇÃO TEMPORAL

Dados de saúde materna-neonatal e dados de poluição possuem granularidades temporais diferentes. Enquanto os primeiros são normalmente organizados e analisados por mês ou ano, os segundos são medidos várias vezes ao dia. De forma a possibilitar a análise visual simultânea dos dois conjuntos de dados (saúde e poluição atmosférica), foi necessário criar um único modo

de agregação. Assim, todos os dados foram agregados por semana epidemiológica e por ano, tendo em vista o seu emprego, respectivamente, nos grupos de visualizações das trilhas exploratórias e no mapa / gráfico de dispersão.

SELEÇÃO DE NASCIDOS EM GESTAÇÕES ÚNICAS

Para permitir uma base de comparação entre os períodos das gestações, e também pela morbimortalidade fetal mostrar-se aumentada nas gestações com dois ou mais fetos no interior do útero, optou-se por eliminar os nascidos de gestação múltipla da base de dados do projeto.

ÍNDICES DE POLUIÇÃO E SAÚDE HUMANA

Para indicar visualmente os índices de poluição inadequados para a saúde humana foram utilizados os valores de referência determinados pelo CONAMA (Conselho Nacional do Meio Ambiente) sobre qualidade do ar em medições diárias de material particulado 25 μ (MP2.5), que por sua vez foram baseados nas recomendações da OMS (Organização Mundial da Saúde - World Health Organization. *Air Quality Guidelines - Update 2021*. Copenhagen, Denmark: WHO Regional Office for Europe) sobre limites diários (15 μ g/m³) e anuais (5 μ g/m³) não prejudiciais à saúde humana.

REORGANIZAÇÃO DOS DADOS POR DATA DE CONCEPÇÃO

O principal desafio para visualizar e analisar os possíveis efeitos da poluição sobre as gestações e a saúde materna-neonatal encontra-se na limitação imposta pelo emprego de dados secundários. Ou seja, não há como acompanhar individualmente as gestações e nascimentos. Em qualquer momento, há um grande número de mulheres grávidas e em diferentes períodos de gestação, que terão duração variável. Um mesmo período de pico de poluição atmosférica pode encontrar mulheres no início da gravidez enquanto outras estão no meio ou no final. Mais ainda, sabe-se que o período de maior suscetibilidade para o feto ocorre entre a 22ª e a 38ª semanas de gestação. No entanto, ao se analisar esta exposição sobre um conjunto de gestações, não é possível observar de forma diferenciada os diversos períodos. A partir destas considerações, decidimos utilizar uma estimativa da data de concepção. Este cálculo foi feito subtraindo-se a duração da gestação em semanas da data de nascimento. Na visualização, empregamos o termo “janela de suscetibilidade” para demarcar este período entre a 22ª e a 38ª semana.

DADOS EM DESTAQUE

As bases de dados do SIM e do SINASC possuem mais de cem atributos, quando levamos em conta os atributos originais e os acrescentados posteriormente como forma de enriquecimento. No entanto, o peso ao nascer trata-se de uma variável da mais extrema importância, uma vez que essa informação é essencial para o cálculo de vários indicadores, incluindo a adequação à idade gestacional. Tendo estabelecido a relevância do atributo peso, foram apenas incluídos casos com a variável peso ao nascer corretamente preenchida. O campo de número de semanas de gestação foi incluído na base do SINASC em 2009 mas os dados mostraram-se consistentes em todos os municípios a partir de 2012. Esta data determinou o início do período compreendido pelas visualizações.

DEFINIÇÕES E INDICADORES

A ferramenta compreende dois conjuntos de visualizações: **Mapa** e **Trilhas**.

MAPA

A visualização Mapa oferece diversas alternativas para comparação entre municípios, filtragem e seleção de até quatro municípios, tendo em vista o aprofundamento da análise a ser realizada posteriormente com a visualização Trilhas. No mapa propriamente dito é possível observar o tamanho da população de cada município e o valor médio anual de uma variável principal escolhida entre 6 possíveis, facilitando a seleção com base no posicionamento geográfico. Esta mesma página também contém um gráfico de dispersão para análise de até três variáveis de todos os municípios disponíveis, obedecendo às filtrações desejadas. Há ainda um gráfico de barras mostrando a variação da variável principal ao longo dos anos para até quatro municípios.

INDICADORES-CHAVE:

- Poluição
- Poluição média por ano
- Baixo Peso
- Mortalidade Neonatal
- Mortalidade Perinatal
- Mortalidade Infantil

TRILHAS

Nas Trilhas, é possível proceder ao aprofundamento da investigação sobre os municípios selecionados no Mapa, mas também, opcionalmente, modificar esta escolha. As Trilhas compreendem três conjuntos investigativos: Nascidos Vivos, Características do Nascimento e Mortalidade Perinatal e Infantil. Todos eles possuem um gráfico de linha com a incidência do poluente atmosférico em paralelo com outros gráficos relacionados especificamente aos dados da saúde neonatal abordados. Os gráficos de linha da saúde neonatal irão apresentar variações de acordo com os atributos escolhidos no menu principal ou no menu de comparação

NASCIDOS VIVOS

INDICADORES-CHAVE:

- Poluente
- Peso ao Nascer
- Adequação à idade gestacional
- Total de nascidos vivos

ATRIBUTOS:

- **Peso ao Nascer:** Normal, Baixo peso, Muito baixo peso, Ignorado
- **Idade Gestacional:** Pós-termo, A termo tardio, A termo completo, A termo precoce, Pré-termo moderado, Muito pré-termo, Pré-termo extremo, Ignorado
- **Adequação à idade gestacional:** GIG (grande para a idade gestacional), AIG (adequado para a idade gestacional), PIG (pequeno para a idade gestacional), Ignorado.
- **Sexo:** Feminino, Masculino, Indeterminado
- **Anomalia Congênita:** Não, Sim
- **Raça / Cor da Mãe:** Branca, Preta, Parda, Outros, Ignorado
- **Escolaridade da Mãe:** Superior, Ensino médio, Fundamental completo, Fundamental incompleto, Sem, Ignorado
- **Estado Civil da Mãe:** Solteira, Casada ou união estável, Viúva, Separada ou divorciada, Ignorado
- **Idade da Mãe:** 10-14, 15-19, 20-29, 30-39, 40-49, 50+, Ignorado
- **Consultas de Pré-Natal:** 7+, 4 a 6, 1 a 3, 0, Ignorado
- **Tipo de parto:** Vaginal, Cesáreo, Indeterminado
- **Trabalho de Parto Induzido:** Sim, Não
- **Adequação Pré-Natal:** Não fez, Inadequado, Intermediário, Adequado, Mais que adequado, Ignorado

CARACTERÍSTICAS DO NASCIMENTO

INDICADORES-CHAVE:

- Poluente
- Peso ao Nascer
- Grupos de Robson
- Total de nascidos vivos

ATRIBUTOS:

- **Peso ao Nascer:** Normal, Baixo peso, Muito baixo peso, Ignorado
- **Idade Gestacional:** Pós-termo, A termo tardio, A termo completo, A termo precoce, Pré-termo moderado, Muito pré-termo, Pré-termo extremo, Ignorado

- **Adequação à idade gestacional:** GIG (grande para a idade gestacional), AIG (adequado para a idade gestacional), PIG (pequeno para a idade gestacional), Ignorado
- **Sexo:** Feminino, Masculino, Indeterminado
- **Anomalia Congênita:** Não, Sim
- **Raça / Cor da Mãe:** Branca, Preta, Parda, Outros, Ignorado
- **Escolaridade da Mãe:** Superior, Ensino médio, Fundamental completo, Fundamental incompleto, Sem, Ignorado
- **Estado Civil da Mãe:** Solteira, Casada ou união estável, Viúva, Separada ou divorciada, Ignorado
- **Idade da Mãe:** 10-14, 15-19, 20-29, 30-39, 40-49, 50+, Ignorado
- **Apgar 5o. Minuto:** Boa vitalidade, Asfíxia moderada, Asfíxia grave, Ignorado
- **Tipo de parto:** Vaginal, Cesáreo, Indeterminado
- **Trabalho de Parto Induzido:** Sim, Não
- **Adequação Pré-Natal:** Não fez, Inadequado, Intermediário, Adequado, Mais que adequado, Ignorado

MORTALIDADE PERINATAL E INFANTIL

INDICADORES-CHAVE:

- Poluente
- Óbitos Fetais
- Óbitos em relação ao parto
- Óbitos totais

ATRIBUTOS:

- **Peso ao Nascer:** Normal, Baixo peso, Muito baixo peso, Ignorado
- **Sexo:** Feminino, Masculino, Indeterminado
- **Raça / Cor da Mãe:** Branca, Preta, Parda, Outros, Ignorado
- **Escolaridade da Mãe:** Superior, Ensino médio, Fundamental completo, Fundamental incompleto, Sem, Ignorado
- **Idade da Mãe:** 10-14, 15-19, 20-29, 30-39, 40-49, 50+, Ignorado
- **Tipo de parto:** Vaginal, Cesáreo, Indeterminado.

Como citar este documento:

KOSMINSKY, D.; ESPERANÇA, C.; ILLARRAMENDI, X.; GIANNELLA, J. R.; MAURO, R.; GOMES JUNIOR, S. C.; BARCELLOS, L. **Trilhas Exploratórias:** Metodologia de visualização de dados. 2023. Disponível em:
<<http://ampliasaude.org/pt/metodologia/>>