



*O Design de Dashboards para Visualização de Dados sobre  
Malformação Congênita*

**Tatiana Aparecida de Almeida**

Junho / 2021

Dissertação de mestrado em Ciência da Computação

# **O Design de Dashboards para Visualização de Dados sobre Malformação Congênita**

Esse documento corresponde à Dissertação apresentada à Banca Examinadora para Defesa no curso de Mestrado em Ciência da Computação do Centro Universitário Campo Limpo Paulista.

Campo Limpo Paulista, 18 Junho de 2021.

Tatiana Aparecida de Almeida

Prof. Dr. Rodrigo Bonacin (Orientador)

O presente trabalho foi realizado com apoio da Coordenação de Aperfeiçoamento de Pessoal de Nível Superior - Brasil (CAPES) - Código de Financiamento 001.

Ficha catalográfica elaborada pela  
Biblioteca Central da Unifaccamp

A451d

Almeida, Tatiana Aparecida de

*O design de dashboards* para visualização de dados sobre malformação congênita / Tatiana Aparecida de Almeida. Campo Limpo Paulista, SP: Unifaccamp, 2021.

Orientador: Prof<sup>o</sup>. Dr. Rodrigo Bonacin

Dissertação (Programa de Mestrado Profissional em Ciência da Computação) – Centro Universitário Campo Limpo Paulista – Unifaccamp.

1. Visualização de informação em saúde. 2. Informática médica. 3. *Design de dashboards* em saúde. 4. Extração de conhecimento. I. Bonacin, Rodrigo. II. Centro Universitário Campo Limpo Paulista. III. Título.

CDD- 005.75

## **DEDICATÓRIA**

Dedico este trabalho aos meus filhos Daniel Ricardo Teixeira, Maria Clara de Almeida Barbosa, Antonio Bento de Almeida Barbosa e Ana Carolina de Almeida Barbosa, mesmo sem entender estiveram do meu lado, com paciência, carinho e incentivando a nunca desistir. Dedico ao pai Bento Donizete de Almeida(em memória), minha mãe Maria Aparecida de Souza Almeida que sempre incentivou o estudo como forma de sobrevivência. Dedico a minha irmã Viviane Aparecida de Almeida que sempre está disposta a cuidar dos sobrinhos quando necessário. A Minha madrinha Fatima, que posso contar sempre com suas orações. O Tio Antonio, Tia Elaine, Tia Neusa e Tio Val, que sempre me trataram como filha e incetiva o estudo.

## AGRADECIMENTOS

Agradeço a Deus por conduzir meus passos a cada fase da minha vida, pois sem Ele, nada seria possível.

Ao meu orientador, professor Dr. Rodrigo Bonacin, por me acolher e guiar durante o desenvolvimento desta pesquisa.

Aos professores do Programa de Pós-Graduação da FACCAMP pela contribuição durante as disciplinas.

Aos funcionários que com dedicação sempre atendeu muito bem.

Agradeço o Coordenador da Plataforma de Ciência de Dados aplicada á Saúde (PCDaS) da FIOCRUZ, Marcel Pedroso, que orientou sobre a base da dados aplicada.

Agradeço a Professora Daniele Kopp, do curso de Especialização em Estatística e Modelagem Quantitativa da Faculdade Cruzeiro do Sul pelo auxílio com os cálculos estatísticos.

Agradeço a professora da Unicamp Luciana Correira Alvez, Professor Tiago José de Carvalho e Professor Carlos Edurdo Beluzo do IFSP(Instituto Federal de Educação e Tecnologia) Campus Campinas por apresentar a necessidade e dificuldade em visualizar dados demográficos na área da Saúde Pública.

Á todas minhas amigas e amigos de salas o Luiz, Leandro, Jeferson, Cris e Raquel pelo apoio durante ao disciplicas e tardes de estudos.

Agradeço ao Colegiado do Programa de Mestrado em Ciência da Computação da FACCAMP ao coordenador do Curso Professor Osvaldo Luiz de Oliveira, por todo apoio, infraestrutura e pela qualidade do programa de mestrado, essenciais para desenvolvimento acadêmico e profissional.

**Resumo.** A evolução dos sistemas computacionais proporcionou o armazenamento quase ilimitado de informações. Entretanto, para que sejam úteis, elas precisam ser processadas e representadas em soluções de visualização adequadas. A visualização de informação torna-se essencial para que qualquer pessoa possa compreender dados volumosos e complexos, tais como dados relacionados à saúde. Portanto, novos processos de *design* de soluções para visualização de dados devem ser pesquisados a fim de que o vasto volume de informações faça sentido para profissionais de saúde, gestores e pacientes. Informações mais precisas e interpretáveis por métodos de visualização adequados podem, por exemplo, contribuir para mitigar problemas relacionados à malformação congênita. Este trabalho, então, possui como objetivo geral pesquisar e propor um processo de *design* de soluções para visualização de informações sobre malformação congênita. A presente pesquisa utiliza dados de abrangência nacional e propõe soluções de dashboards para apoiar pacientes e profissionais de saúde. Tais soluções são baseadas no estudo da literatura científica da área e são avaliadas por usuários (gestantes) e profissionais de áreas correlatas.

**Palavras-chave:** Visualização de Informação em Saúde, Informática Médica, Design de Dashboards em Saúde, Extração de Conhecimento.

*Abstract.* The evolution of computer based systems provides almost unlimited information storage. However, to be useful, it needs to be processed and represented by suitable visualization solutions. Thus, the visualization of information is necessary to browse and understand voluminous and complex data, such as health data. Therefore, new data visualization design processes must be investigated so that the vast amount of information makes sense for healthcare professionals, healthcare administrators and patients. More accurate and interpretable information through suitable visualization methods can, for instance, contribute to mitigate problems related to congenital malformation. This work aims to investigate and to propose a design process for information visualization solutions of congenital malformation information. This research uses national data and proposes dashboard solutions to support patients and healthcare professionals. Such solutions are based on the study of the scientific literature in the field as well as it is evaluated by users (pregnant women) and experts of related fields.

**Keywords:** Health Information Visualization, Medical Informatics, Health Dashboards Design, Knowledge-Discovery in Databases.

## LISTA DE FIGURAS

Figura 1 – Pirâmide do conhecimento .....	9
Figura 2 – Processo KDD .....	10
Figura 3 – Fluxo para os nascimentos hospitalares .....	11
Figura 4 – Rede de difusão para três assuntos diferentes .....	14
Figura 5 – Demonstração dos dados no formato de agrupamento.....	15
Figura 6 – Visualização de dados informativa.....	16
Figura 7 – Visualização de dados comparativa.....	17
Figura 8 – Visualização de dados de alteração .....	17
Figura 9 – Visualização de dados organizacionais .....	18
Figura 10 – Visualização de dados de relacionamento .....	19
Figura 11 – Exemplo de como se apresentavam os indicadores antigos .....	24
Figura 12 – Indicadores implementados .....	25
Figura 13 – Exemplo UK NHS – <i>Dashboard</i> Operacional Clínico para de Monitorização de Episódios de Urgência .....	26
Figura 14 – Exemplo UK NHS – <i>Dashboard</i> de Otimização da Utilização de Medicamentos .....	27
Figura 15 – Diferentes formas de interação com os sistemas .....	28
Figura 16 – Interação usuário – sistema .....	29
Figura 17 – Aceitabilidade de um sistema.....	29
Figura 18 – Processo do design centrado no usuário (Agnes et al., 2015).....	30
Figura 19 - Análise de conteúdos a serem apresentados na visualização conforme o nível de experiência do usuário (Young et al., 2020).....	32
Figura 20 - Estrutura de usabilidade. Fonte: ISO 9241-11 – ABNT.....	34
Figura 21 - Técnicas para avaliação de usabilidade. Fonte: Adaptado de Cybis.....	34
Figura 22 - Exemplo de Aplicação da SUS.....	35
Figura 23 – Pesquisa por palavra-chave para GIS, epidemiologia molecular e análise de rede social no PubMed.....	38
Figura 24 – Framework para visualização de dados acadêmicos .....	39
Figura 25 – Técnicas de classificação da visualização (adaptado de MacNabb e Laramée, 2019).....	40

Figura 26 – Diferentes visualizações de mapas com informações geográficas .....	41
Figura 27 – Interface para navegação do usuário do aplicativo <i>web HealthMap.org</i> .....	42
Figura 28 – Mapa gerado a partir de casos de coronavírus, em Campinas, SP .....	43
Figura 29 – Arquitetura de um sistema de visualização demográfica .....	44
Figura 30 – Demonstração da integração de visualizações .....	45
Figura 31 – Processo design.....	52
Figura 32 – Solução de mineração adotada.....	53
Figura 33 – Diagrama de Fluxo ETL.....	55
Figura 34 – Na esquerda, a evolução dos nascidos com anomalias; na direita, os nascidos sem anomalia.....	58
Figura 35 – Total de nascimentos X nascidos com anomalia .....	59
Figura 36 – Acumulado de nascidos com anomalia por ano .....	59
Figura 37 – Acumulado de nascidos com anomalia por estado.....	60
Figura 38 – Dashboard com dados gerais .....	61
Figura 39 – Dashboard com análise de dados da mãe .....	63
Figura 40 – Dashboard com filtro das análises das gestações .....	64
Figura 41 – Dashboard com resultados do filtro dos recém-nascidos .....	66
Figura 42 – Mapa de criança recém-nascida com malformação congênita, de 2017, comparada com IDHM (2010) e casos/ 100 mil habitantes .....	67
Figura 43 – Mapas com resultados de crianças recém-nascidas com malformação congênita por categoria e casos/ 100 mil habitantes.....	68
Figura 44 – Distribuição dos nascimentos X Nascimentos com anomalia por região.....	79
Figura 45 - Dashboard com resultados do filtro de comparação (Elaborado pela autora, 2020).. .....	99
Figura 46 – Dashboard com resultados do filtro de Anomalias (Elaborado pela autora, 2020).....	100
Figura 47 – Resultados do perfil profissão que fez avaliação do teste de usabilidade.....	103
Figura 48 – Resultados do perfil sexo que fez avaliação do teste de usabilidade.....	103
Figura 49 – Resultados do perfil estado civil dos participantes na avaliação do teste de usabilidade.....	104
Figura 50 – Resultado do perfil nível de escolaridade na avaliação do teste de usabilidade..	104
Figura 51 – Resultado do perfil: Tem filhos? Quantos? na avaliação do teste de	

usabilidade .....	105
Figura 52 – Resultado do perfil:conhece alguma criança com malformação congênita?.....	105
Figura 53 – Resultados do Teste Usabilidade (Formulario Google 1 a 2 pergunta).....	106
Figura 54 – Resultados do Teste Usabilidade (Formulário Google 3 a 5 pergunta).....	107
Figura 55 – Resultados do Teste Usabilidade (Formulario Google 6 a 8 pergunta).....	108
Figura 56 – Resultados do Teste Usabilidade (Formulario Google 9 a 11 pergunta).....	109
Figura 57 – Resultados do Teste Usabilidade (Formulario Google 12 a 14 pergunta).....	110
Figura 58 – Resultados do Teste Usabilidade (Formulario Google 15 a 17 pergunta).....	111
Figura 59 – Resultados do Teste Usabilidade (Formulario Google 18 a 20 pergunta).....	112
Figura 60 – Resultados do Teste Usabilidade (Formulario Google Sugestão Gerais).....	113

## LISTA DE TABELAS

Tabela 1 – Tabela CID-10.....	12
Tabela 2 – Diagnósticos de anomalias congênitas frequentes em nascidos vivos.....	13
Tabela 3 – Exemplos de informações usadas na visualização.....	19
Tabela 4 – Ferramentas para visualização baseadas em javascript (adaptado de Liu et al., 2017 .....	20
Tabela 5 - Ferramentas para visualização baseadas em outras linguagens (Liu et al., 2017)...	21
Tabela 6 – Tipos de dashboards (Nogueira et al., 2017).....	23
Tabela 7 - Recomendações na construção de dashboards a partir de cada etapa (Adaptado Hartzler et al., 2015).....	33
Tabela 8 - Frequência de pesquisas encontradas na literatura de UDC (Design Centrado no Usuário) voltados para determinado domínio de saúde (Adaptado Ghazali e Ariffin, 2014).	47
Tabela 9 – Trabalhos relacionados recentes com foco no público geral ou que usam UCD...	48
Tabela 10 – Descrição das informações quantitativas por idade da mãe, número de filhos vivos, mortos, semana de gestação, apagar 1, apagar 5 e peso do recém-nascido.....	71
Tabela 11 – Comparação das informações quantitativas das mães por idade, número de filhos vivos, mortos por região do Brasil.....	72
Tabela 12 – Comparação das informações quantitativas dos dados de recém-nascidos e gestação por região do Brasil.....	73
Tabela 13 – Comparação das informações quantitativas das mães, gestação e recém-nascido em relação à presença ou não de malformação.....	74
Tabela 14 – Comparação das informações quantitativas pela presença de malformação por região.....	75
Tabela 15 – Comparação das informações quantitativas pela presença de malformação por região.....	77
Tabela 16 – Análise de correlação entre a cidade da mãe e quantitativos da gestação e do recém- nascido .....	77
Tabela 17 – Análise de correlação entre a idade da mãe e quantitativos da gestação e do recém- nascido .....	78
Tabela 18 – Distribuição dos nascimentos por região .....	79
Tabela 19 – Descrição dos dados categóricos sobre as mães por estado civil e escolaridade	79

Tabela 20 – Descrição das informações quantitativas de dados sobre a gestação (tipo de gravidez, consultas pré-natal, tempo de gestação, parto e local de nascimento) .....	80
Tabela 21 – Descrição das informações quantitativas de dados sobre os recém-nascidos (sexo do bebê, raça/cor, identificação de anomalia).....	81
Tabela 22 – Descrição dos dados categóricos das mães (estado civil e escolaridade) por região de nascimento do recém-nascido .....	82
Tabela 23 – Associação dos dados da gestação (gravidez, consulta pré-natal, tempo de gestação, parto e local de nascimento) por região de nascimento do recém-nascido .....	83
Tabela 24 – Associação dos dados do recém-nascido (sexo, raça/cor, anomalia) com a região de nascimento.....	84
Tabela 25 – Comparação do perfil categórico das mães (estado civil e escolaridade) pela presença de malformação.....	84
Tabela 26 – Comparação do perfil categórico das gestações (gravidez, consultas pré-natal, tempo de gestação, parto, local de nascimento) pela presença de malformação em recém-nascido .....	85
Tabela 27 – Comparação do perfil categórico das recém-nascido (sexo, raça/ cor) pela presença de malformação.....	86
Tabela 28 – Comparação do estado civil das mães pela presença de malformação por região do Brasil.....	87
Tabela 29 – Comparação da escolaridade das mães pela presença de malformação por região do Brasil.....	88
Tabela 30 – Comparação dos dados de gestação com a presença ou não de malformação.....	89
Tabela 31 – Comparação do tempo de gestação e tipo de parto pela presença de malformação por região do Brasil.....	90
Tabela 32 – Comparação do local de nascimento pela presença de malformação por região de nascimento .....	91
Tabela 33 – Comparação dos dados dos recém-nascidos (sexo, raça/ cor) pela presença de malformação .....	92
Tabela 34 – Associação do tipo de anomalia e malformação com o ano de nascimento .....	93
Tabela 35 – Associação do tipo de anomalias por região de nascimento do recém-nascido...	94
Tabela 36 – Associação do tipo de anomalia com o tempo de gestação (premature, tempo normal).....	94

Tabela 37 – Associação de características da gravidez com o tipo de anomalia.....	96
Tabela 38 – Associação da quantidade de consultas por tipo de anomalia .....	97
Tabela 39 – Associação entre o local de nascimento e tipo de anomalia apresentada pelo recém-nascido .....	97
Tabela 40 - Questionário adaptado com base no System Usability Scale (SUS).....	102

## SUMÁRIO

<b>1 INTRODUÇÃO</b> .....	1
1.1 CONTEXTO E MOTIVAÇÃO .....	2
1.2 PROBLEMÁTICA E JUSTIFICATIVA.....	3
1.3 OBJETIVOS, CONTRIBUIÇÕES E MÉTODOS .....	4
1.4 ESTRUTURA DA PROPOSTA.....	5
<b>2 REFERENCIAL TEÓRICO E METODOLÓGICO</b> .....	7
2.1 VISUALIZAÇÃO DE INFORMAÇÃO E MINERAÇÃO DE DADOS.....	7
2.2 ANÁLISE E MINERAÇÃO DE DADOS DE NASCIDOS VIVOS NO BRASIL .....	10
2.3 CONCEITOS E TÉCNICAS DE VISUALIZAÇÃO DE DADOS .....	15
2.4 <i>DASHBOARDS</i> PARA APOIO A DECISÃO EM SAÚDE.....	22
2.5 CONCEITOS DE INTERFACE HUMANO COMPUTADOR ADOTADOS NESTA DISSERTAÇÃO .....	27
<b>3 TRABALHOS RELACIONADOS</b> .....	37
3.1 VISUALIZAÇÃO DE DADOS EM SAÚDE .....	37
3.2 CONSIDERAÇÕES SOBRE OS TRABALHOS RELACIONADOS .....	47
<b>4 PROPOSTA DE DISSERTAÇÃO E RESULTADOS ESPERADOS</b> .....	50
4.1 METODOLOGIA DE PESQUISA.....	50
4.2 PROPOSTA DE SOLUÇÃO.....	52
4.3 PROTOTIPAÇÃO E IMPLIMENTAÇÃO DO SISTEMA .....	57
<b>4.3.1 Implementação das Visualizações Georreferenciadas</b> .....	66
<b>4.3.2 Análise Estatísticas dos Dados do SINASC</b> .....	70
<b>4.3.3 <i>Dashboard</i> Resultantes da Análise Estatísticas</b> .....	92
<b>5 AVALIAÇÃO COM USUÁRIO E ESPECIALISTA</b> .....	101
5.1 MÉTODO DE AVALIAÇÃO E PARTICIPANTES.....	101
5.2 RESULTADOS DA AVALIAÇÃO.. .....	109
5.3 SÍNTESE E DISCUSSÃO.....	113
<b>6.CONCLUSÃO</b> .....	116
6.1.CONTRIBUIÇÕES.....	118
6.2 TRABALHOS FUTUROS.....	122
6.3 PUBLICAÇÕES OBTIDAS E PLANEJADAS.....	123

<b>REFERÊNCIAS BIBLIOGRÁFICAS.....</b>	<b>124</b>
<b>APÊNDICE A – Dicionário de dados do SINASC .....</b>	<b>131</b>
<b>APÊNDICE B – Colunas com a descrição do Tableau.....</b>	<b>140</b>
<b>APÊNDICE C – Mapas com visualização de casos.....</b>	<b>155</b>
<b>APÊNDICE D – Cálculos Usados e Github.....</b>	<b>170</b>

## Lista de abreviaturas e siglas

<b>OMS</b>	Organização Mundial de Saúde
<b>IHC</b>	Interação Humano Computador
<b>SINASC</b>	Sistema de Informações sobre Nascidos Vivos
<b>DNV</b>	Declaração de Nascidos Vivos
<b>CID</b>	Classificação Internacional de Doenças
<b>KDD</b>	Processo de Descoberta de Conhecimento
<b>PCDaS</b>	Plataforma de Ciência de Dados aplicada à Saúde
<b>UDC</b>	<i>Design Centrado no Usuário - User-Centered Design</i>
<b>HCD</b>	<i>Design Centrado em Humanos - Human-Centered Design</i>
<b>Datusus</b>	Departamento de Informática do Sistema Único de Saúde
<b>SUS</b>	<i>System UsabilityScale</i>

## 1 INTRODUÇÃO

A evolução dos sistemas de informação, em conjunto com o aumento da capacidade de processamento e armazenamento, possibilitou a aquisição quase ilimitada de dados das mais diversas fontes. Os dados podem ser utilizados como apoio à tomada de decisão em situações complexas como a gestão da saúde pública, bem como para informar a população. Para isso, segundo Bispo e Cazarini (1998), foram necessárias pesquisas sobre novas ferramentas computacionais que atendessem a essa nova demanda. O conhecimento que pode ser extraído das informações em sistemas computacionais é um grande diferencial competitivo entre as organizações (WERNER, 2014).

Werner (2014) destaca que somente ter as informações armazenadas não apresenta diferencial por si só, sendo necessário encontrar essas informações de maneira ágil, pois, dependendo da volatilidade, ela pode deixar de ter valor em pouco tempo. Para Rodrigues Filho e Gomes (2004), a Tecnologia da Informação passou a ser considerada um elemento indispensável na maior parte das organizações, uma vez que não apenas faz o processamento e organização dos dados, mas também oferece suporte à gestão da informação. Dessa maneira, as empresas vêm buscando, nos últimos anos, alternativas de visualização de dados que lhes permitam uma melhor análise das informações.

Uma das aplicações das técnicas de visualização é a área da saúde, em que podemos realizar uma análise aplicada em doenças, tais como sobre a malformação congênita, tema deste trabalho. A área da saúde oferece desafios para aplicação dos métodos de visualização devido à quantidade de dados disponíveis e também por, de acordo com Carroll *et al.* (2014), existir pouca literatura disponível. Dessa maneira, uma maior quantidade de pesquisas sobre como desenvolver visualizações específicas a este domínio, avaliação da experiência do usuário e técnicas de visualização para dados de saúde heterogêneos são necessários.

A Organização Panamericana da Saúde (OPAS)<sup>1</sup> define malformação congênita como toda anomalia no desenvolvimento do feto e podem ser classificadas como maiores ou menores, podendo muitas vezes levar à morte. Em alguns casos a importância cirúrgica, médica ou estética sobrepõem-se aos fenótipos normais, elas podem ser únicas ou múltiplas e associar-se a malformações maiores.

---

<sup>1</sup> <https://www.paho.org/pt>

Várias técnicas de visualização podem ser usadas para representar dados da saúde pública, incluindo mapas de coropletas, mapas de calor, gráficos de bolhas, gráficos de dispersão e gráficos de barras. Tais técnicas podem usar marcas visuais diferentes (por exemplo, cor, tamanho e forma) para codificar diferentes dimensões da informação. Entretanto, novos designs de soluções para visualização, que explorem em conjunto diferentes técnicas em interfaces para apoio à decisão, devem ser pesquisadas. Neste trabalho é proposto processo de design para:

- Entender como as estruturas de um *dataset* e a variação dos dados podem ser representadas graficamente, funcionando como apoio durante a atividade de transformação dos dados para visualização e interpretação de informações.
- Entender como representações das estruturas de dados podem auxiliar na tomada de decisão no domínio em questão e também como essas representações gráficas podem ser uma ferramenta útil no tratamento e na operação de grande volume de dados complexos em saúde.
- Entender como o uso de técnicas estatísticas e técnicas IHC (Interação Humano Computador) podem ser aplicadas como respostas a perguntas de interesse sobre malformação congênita (anomalia). O processo coloca as gestantes, junto com profissionais de saúde, como parte interessada central em informações sobre malformação congênita em uma visão centrada no usuário.

## 1.1 CONTEXTO E MOTIVAÇÃO

Este trabalho visa contribuir em um contexto complexo para apoio à saúde pública, que acarreta consequências físicas e psicológicas aos envolvidos. Conforme Fontoura (2018), a gestação é uma fase na vida da mulher vivenciada como uma transição, quando, desde o momento da fecundação até o nascimento, a gestante e o recém-nascido passam por experiências psicológicas, fisiológicas e sociais únicas. Isso faz com que a mulher crie dúvidas e expectativas em relação ao novo ser que está por nascer e o papel que deverá assumir. Dessa maneira, a notícia de um feto com malformação congênita pode ser um momento complicado para a gestante (Pelchat, 1992). É importante que a gestante, bem como familiares e profissionais de saúde, tenham acesso a informações sobre o tema de modo a lidar com essa situação.

Segundo Fontoura *et al.* (2014), no Brasil, os índices estatísticos evidenciam que as malformações congênitas constituem a segunda causa de mortalidade neonatal, sendo responsáveis por 22,8% do total de mortes. As regiões Sudeste e Norte apresentam a maior proporção de registros de óbitos por essa causa, perfazendo 35,9% e 24,5%, respectivamente. Em estudo citado por Fontoura *et al.* (2014) sobre prevalência de malformações em São Paulo, concluiu-se que é importante que aconteça o diagnóstico precoce das malformações para a redução da morbimortalidade, principalmente, o neonatal precoce, para a melhora da qualidade de vida e dos índices de sobrevivência.

Mendes *et al.* (2018) citam como principais causas da malformação congênita os transtornos congênitos e perinatais, muitas vezes, associados a agentes infecciosos deletérios à organogênese fetal, tais como os seguintes vírus: rubéola, imunodeficiência humana (HIV), Zika, entre outros. O uso de drogas lícitas e ilícitas, medicações teratogênicas e endocrinopatias maternas também são causas de anomalias. Mendes *et al.* (2018) também mencionam que entre 15% a 25% das malformações congênitas ocorrem devido a alterações genéticas, 8% a 12% são causados por fatores ambientais e 20% a 25% envolvem tanto genes, quanto fatores ambientais (herança multifatorial). Segundo os autores, entre 40% a 60% dos fatores causais ainda possuem origem desconhecida.

As malformações congênitas também contribuem em muito para a mortalidade infantil. De acordo com Bastos e de Macedo (2008), elas afetam uma em cada dez ou vinte crianças que nascem e, não obstante, 50% destas poderiam ser prevenidas por meio de medidas simples. Os autores explicam que apesar de extremamente variáveis em tipos e mecanismo causal, elas surgem de um transtorno no desenvolvimento ontogenético. Dessa maneira, com uso da análise podemos, em um estágio inicial, agir com o objetivo de proteger as gestações. A visualização de dados estatísticos é um elemento essencial para que gestores e profissionais de saúde possam prever tendências e atuar, bem como para informar a população (principalmente gestantes) sobre o tema.

## 1.2 PROBLEMÁTICA E JUSTIFICATIVA

Com a análise dos dados durante o pré-natal, pode-se verificar a probabilidade de haver malformação do recém-nascido até os 28 dias de vida. Apresenta-se, então, a ciência de dados, um campo de estudo com destaque para a capacidade de auxiliar a descoberta de informações

úteis a partir de um grande repositório de dados, como uma ferramenta que pode auxiliar de casos de malformação congênita por meio da análise de diversos fatores. De acordo com o ICICT (2020), a ciência de dados pode ainda ser definida como um conjunto de estratégias, ferramentas e técnicas para coleta, transformação e análise de dados realizadas por equipes multidisciplinares formadas por pesquisadores com conhecimento substantivo do problema em análise.

A grande vantagem da ciência de dados é que ela consegue lidar com grandes volumes de dados complexos e transformá-los dados em uma forma de visualização que possa ser entendida por pessoas que não possuam um conhecimento técnico específico em análise de dados, tais como executivos, gestores, profissionais de saúde e a população em geral. Assim, a visualização de informações é um elemento central para a ciência de dados por permitir que os dados sejam modelados e apresentados em diferentes formas e graus de abstração.

A grande dificuldade encontrada quando se trabalha com visualização de dados é saber qual o tipo de visualização mais adequado para cada caso e usuário. Isso ocorre, principalmente, quando não há um gestor da área disponível para avaliar e dar diretrizes para o design. Dessa maneira, precisa-se de processos de design que produzam resultados adequados na apresentação dos dados, na sua navegabilidade e na forma de visualização. Para tanto, é importante permitir que cada usuário possa manipular os dados conforme suas necessidades, tornando, assim, um problema com maior complexidade. Tais aspectos são agravados em sistemas de saúde que envolvem decisões sobre a vida humana e a gestão de problemas multidisciplinares, tal como lidar com malformações congênitas.

### 1.3 OBJETIVOS, CONTRIBUIÇÕES E MÉTODOS

Apresenta-se como objetivo geral deste estudo o desenvolvimento de um processo de design em uma abordagem centrada no usuário que produza soluções de visualização interativas para apresentação de dados e informações para gestores, profissionais e para a população em geral. Este processo é aplicado em dados estatísticos de saúde referentes à malformação congênita na gravidez para produção de *dashboards* para informar gestantes, profissionais de saúde e gestores. Para tanto, foram analisados métodos de visualização e propostas soluções de *design* para auxiliar na tomada de decisão sobre dados minerados de bases de dados sobre malformações congênitas. Assim, pretende-se responder à seguinte pergunta: *Como realizar o*

*design de soluções de visualização para informar a população e especialistas com o uso de dados minerados sobre bases de malformação congênitas?*

Para atingir esse objetivo e responder à pergunta, apresentam-se os seguintes objetivos específicos e métodos que serão empregados de forma a nortear este trabalho:

- Estudar novas alternativas de *design* para visualização de dados e *dashboards* em saúde;
- Estudar o problema da malformação congênita, de maneira a entender os dados e mecanismos de visualização que podem apoiar na interpretação de como e por que ela ocorre;
- Investigar e propor processo de design centrado no usuário (UCD - *User-Centered Design*), que tem em consideração a população (gestantes), profissionais e gestores;
- Propor um design de *dashboards* para apoio em assuntos relacionados à malformação congênita com base no processo proposto;
- Implementação do protótipo com base no *design* proposto;
- Aplicação da análise em uma base de dados e integração com protótipo de visualização;
- Por fim, investigar e analisar os resultados encontrados com especialistas do domínio e população (gestantes).

Do ponto de vista da computação, este trabalho contribui com processo de UCD que leva em consideração a população e não apenas profissionais. A literatura em *dashboards* em saúde, bem como em outras áreas, focam principalmente especialistas no domínio, e não ambos a população e os especialistas. Isso impõe desafios à área de Interação Humano-Computador (IHC) tais como o uso de linguagem adequada e visualizações de fácil compreensão. Este trabalho constitui contribuição única para área de IHC ao abordar processo de *design* que envolve estudo e a representação de dados provenientes da mineração e análise estatística direcionada a população e especialistas. Mais especificamente, no contexto da saúde, o trabalho se diferencia ao apresentar a aplicação da visualização de dados em uma base de dados que agrega dados de todo o Brasil sobre malformação congênita, em diferentes padrões de qualidade, e também na grande quantidade de atributos contidos na base.

#### 1.4 ESTRUTURA DA DISSERTAÇÃO

De maneira a facilitar a leitura, organizamos o trabalho na forma descrita a seguir.

No capítulo 2 é apresentado o embasamento teórico e metodológico em que esta pesquisa se fundamenta, incluindo métodos de visualização e análise de dados, bem como fundamentos sobre malformação congênita e IHC.

No capítulo 3 são descritos os principais trabalhos relacionados, que visam o uso de técnicas de visualização de informações em saúde.

No capítulo 4 tem-se uma descrição do processo de design proposto, o *design* da proposta de solução e os resultados da instanciação do processo, incluindo conjunto de *dashboards* e a análise estatística dos dados.

O capítulo 5 apresenta a avaliação com usuários e especialistas em áreas relacionadas. São apresentados resultados da avaliação de usabilidade aplicada por meio de conjunto de questões baseadas no modelo de questionário SUS (*System Usability Scale*).

O capítulo 6 apresenta a conclusão e trabalhos futuros.

O Apêndice A detalha o dicionário de dados do SINASC; o Apêndice B detalha as colunas com a descrição do Tableau (2018) e o Apêndice C detalha os mapas com visualização dos casos.

## 2 REFERENCIAL TEÓRICO E METODOLÓGICO

Este capítulo detalha a fundamentação teórica em que o presente trabalho está baseado. O capítulo se inicia com os conceitos básicos sobre a visualização de informação e mineração de dados (seção 2.1) e com um panorama da análise e mineração de dados sobre nascidos vivos no Brasil (seção 2.2). Na sequência, o capítulo detalha conceitos e técnicas de visualização de dados (seção 2.3), a aplicação de *dashboards* para apoio à decisão em saúde (seção 2.4) e, por fim, apresenta os conceitos de IHC utilizados (seção 2.5).

### 2.1 VISUALIZAÇÃO DE INFORMAÇÃO E MINERAÇÃO DE DADOS

Segundo Pino (2016), a visualização de dados, ou *data visualization*, é a representação dos dados em formato gráfico. De acordo com o autor, seu maior objetivo é a promoção de uma comunicação mais clara, simples e eficiente dos resultados obtidos no processo de análise de dados. Conforme Tenbu (2019), uma visão clara do desempenho pode estar no meio das linhas de transação, interações, processos e comportamento dos tipos de dados armazenados. A visualização desses dados permite o reconhecimento em um cenário de alto nível, no qual é possível a observação e a identificação de tendências e padrões, que não seriam facilmente visualizadas “a olho nu”.

A visualização de dados teve origem na década de 1950, quando os primeiros gráficos e imagens foram gerados por computadores, e foi impulsionada, em 1987, com um relatório publicado na National Science Foundation (NSF, 1987) citando a necessidade de criação de novas técnicas de visualização. Com o aumento do poder computacional, foram armazenados dados redes sociais, de elementos microscópios, de dispositivos, entre outros. Dados de diferentes formatos, como imagem, multimídia e texto, foram, então, armazenados e tornaram-se necessárias técnicas avançadas de computação para processar e apresentá-los de forma visual. Nos anos de 1990, conforme Freitas *et al.* (2001), o trabalho com esses dados foi vinculado à área de pesquisa de visualização de dados, como subcampos de visualização da informação, técnicas de modelagem de dados, visualização volumétrica, algoritmos de visualização, entre outros.

Em comparação com outras áreas, esta é relativamente nova e tem ganhado grande destaque nos últimos anos. Isso porque o armazenamento e processamento de uma grande

quantidade de dados (*big data*) se tornou um desafio para os profissionais obterem conhecimento por meio de dados brutos, muitas vezes, não padronizados e incompletos. Pode-se dizer, então, que a ciência de dados, ou *data science*, fundamenta-se na visualização de dados, pois são os profissionais da ciência de dados os responsáveis por lapidar os dados de forma a extrair conhecimento deles. A tomada de decisão formal, segundo Choo (1998), é estruturada por regras e procedimentos que caracterizam papéis, métodos e normas que vão influenciar em como as organizações lidam com os problemas do dia a dia. A extração de conhecimento é um processo importante no contexto de ciência de dados e é utilizado para que pessoas façam sentido sobre o contexto.

Para Rezende (2002), o processo da gestão do conhecimento pode ser dividido em quatro níveis, conforme apresentado na Figura 1, adiante. Podemos observar que, no nível 1, apresenta-se o dado bruto e, conforme ele vai sendo tratado, é possível obter conhecimento a partir dele. Nesse ponto, faz-se importante ressaltar que a qualidade da informação adquirida depende da qualidade dos dados coletados.

O primeiro nível aborda o gerenciamento dos dados, ou seja, a aquisição desses, que podem ser obtidos por meio de sistemas externos, formulários, métricas, entre outros. Um exemplo de aquisição de dados é o Sistema de Informação sobre Nascidos Vivos (SINASC)<sup>2</sup>, cujo objetivo é coletar, produzir, analisar e divulgar dados sobre os nascidos vivos em território nacional. Esse instrumento foi instituído pelo Ministério da Saúde, em 1990, e tem obrigatoriedade de emissão para todos os nascidos vivos, independentemente de como o parto ocorreu.

O nível 2 está ligado à estrutura do dado, ou seja, aos processos relacionados ao tratamento e padronização dos dados. Um exemplo poderia ser a eliminação de registros incompletos, padronização de campos como cidade, CEP, telefone, entre outros. No nível 3 são aplicadas técnicas de tratamento de dados, como, por exemplo, a mineração de dados, o qual espera-se como resultado a interpretação dos dados, resultando na extração de conhecimentos. Por fim, no último nível, apresenta-se o dado bruto transformado em conhecimento, em que também é importante que esteja relacionado a uma pergunta a qual busca-se a resposta, como, por exemplo: Quais as características estão associadas com a malformação ao longo do tempo?

---

<sup>2</sup> Disponível em: <<http://www2.datasus.gov.br/DATASUS/index.php?area=060702>>. Acesso em: 25 mar. 2021.



Figura 1 – Pirâmide do conhecimento (IBC, 2019).

Apesar da quantidade de sistemas computacionais disponíveis, poucos sistemas se preocupam com o uso de técnicas avançadas para melhorar a qualidade dos dados aos profissionais da saúde, como, por exemplo, o uso de técnicas derivadas da mineração de dados.

A mineração de dados, de acordo com Tan *et al.* (2009) e Liu (2012), é o processo de descoberta automática de informações úteis em grandes repositórios de dados e é parte integral do processo de descoberta de conhecimento (KDD – *Knowledge Discovery in Databases*), que têm por objetivo converter dados brutos em informações úteis.

O processo de KDD pode ser observado na Figura 2, adiante, e começa com uma base de dados bruta, que precisa passar por um pré-processamento que terá por objetivo padronizar algumas informações, retirar informações inválidas, nulas ou repetidas. Assim, é necessário compreender o processo e as diferentes necessidades e possibilidades para cada aplicação. As tarefas de aprendizado descritivo, ou não supervisionado, referem-se à identificação de informações relevantes nos dados sem a presença de um elemento externo para guiar o aprendizado. Essas tarefas podem ser divididas em sumarização e associação. A sumarização tem por objetivo encontrar uma descrição simples e compacta dos dados e a associação faz a busca por padrões de associações entre os atributos, de acordo com sua similaridade ou ocorrência.

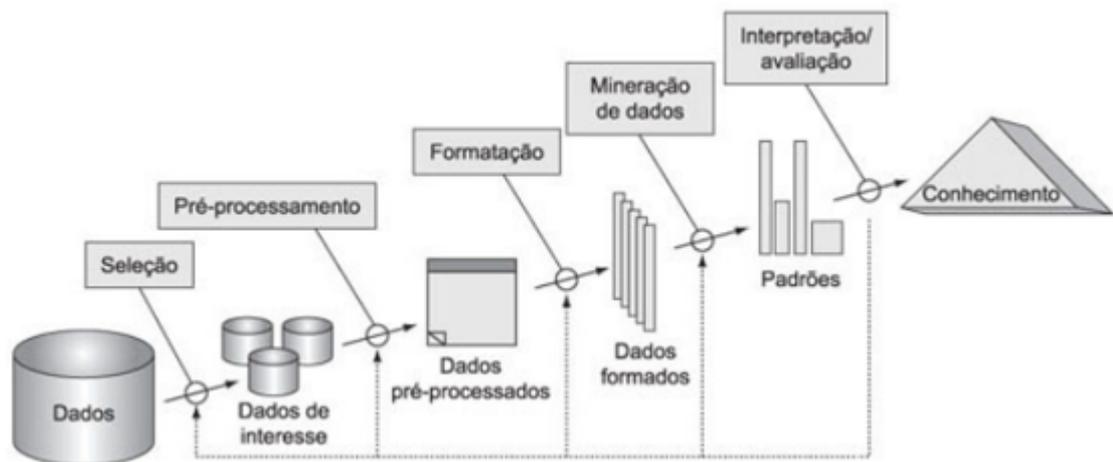


Figura 2 – Processo KDD (TAN *et al.*, 2009).

A diferença e o desafio em se trabalhar com dados ligados à área de saúde é a quantidade de variáveis envolvidas e, conseqüentemente, a visualização disso, pois seu espaço torna-se extremamente grande e complexo. Dessa maneira, são necessários fazer vários testes e simulações de maneira a se chegar em um resultado ideal.

## 2.2 ANÁLISE E MINERAÇÃO DE DADOS DE NASCIDOS VIVOS NO BRASIL

A base de dados SINASC é preenchida de acordo com o fluxograma apresentado na Figura 3. Com base nisso, considera-se como nascido vivo os recém-nascidos que, após expelidos do corpo da mãe, respirem ou apresentem demais sinais de vida, tal como batimento do coração ou pulsação do cordão umbilical. Para os demais (natimortos), em qualquer tipo de gestação, deverá ser preenchida somente a Declaração de Óbito (DO). A declaração apresenta ainda um campo para descrição de malformação congênita e/ ou anomalia cromossômica, bem como campos para inclusão dos códigos referentes a descrição apresentada, de acordo com o Classificação Internacional de Doenças CID 10). Estes defeitos podem apresentar-se isolados ou associados, compondo síndromes de causas genéticas e/ ou ambientais. O registro dessas informações muitas vezes é dificultado pela quantidade de pessoas trabalhando no hospital, bem como por fluxos inadequados, dessa maneira é recomendado pelo SINASC que somente uma pessoa faça o preenchimento do formulário e próximo da data de alta da genitora. Essa informação, posteriormente, é acrescentada no banco de dados da SINASC.

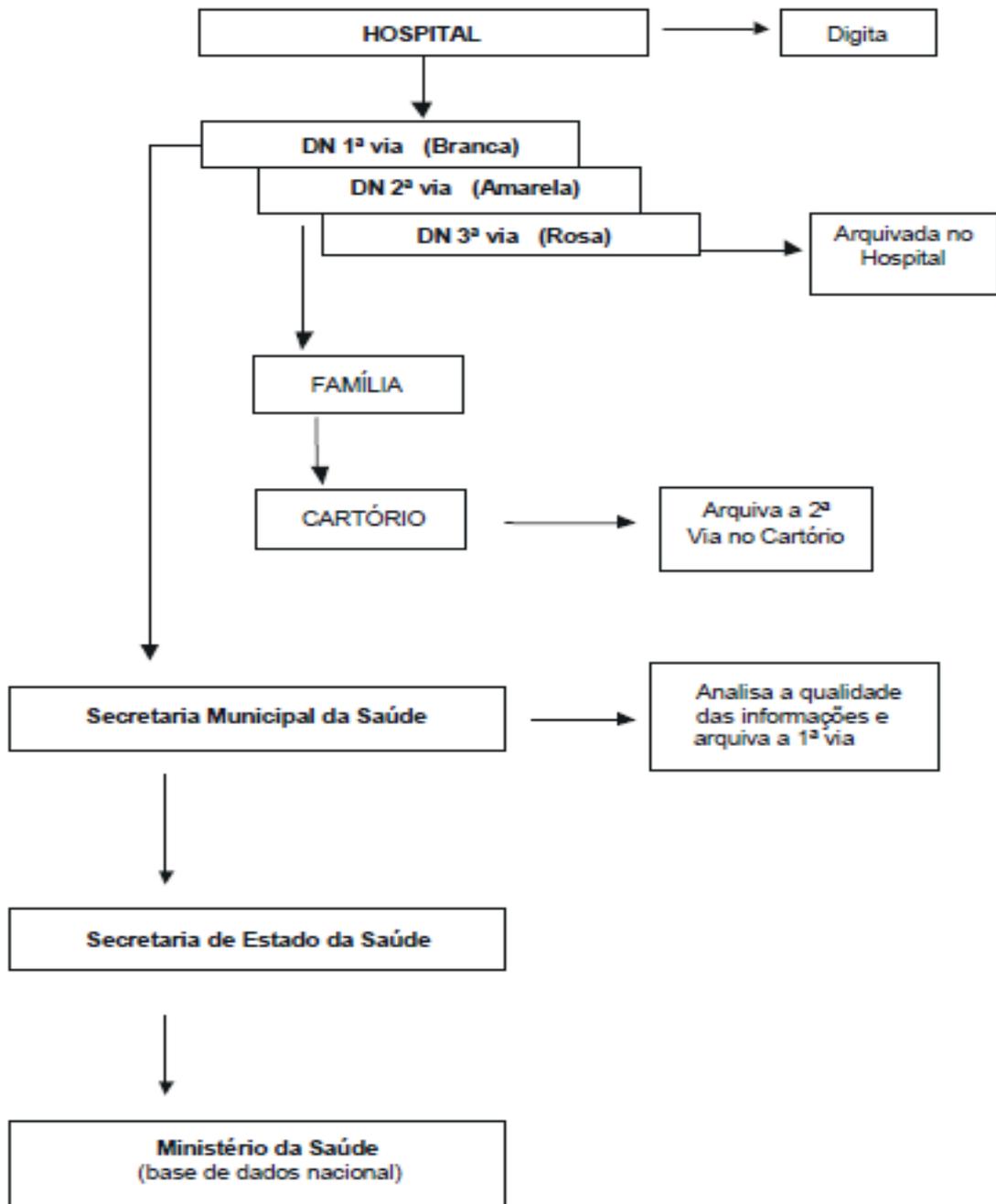


Figura 3 – Fluxo para os nascimentos hospitalares (SINASC, 2017).

Para melhor entendimento sobre malformações congênitas, a Tabela 1, adiante, apresenta o CID-10, que trata a respeito das “Malformações congênitas, deformidades e anomalias cromossômicas” e possui os seguintes agrupamentos, apresentados de acordo com o grau de extensão, frequência e complexidade com que ocorrem.

<b>CID-10</b>	<b>Descrição</b>
Q00-Q07	Malformações congênitas do sistema nervoso
Q10-Q18	Malformações congênitas do olho, do ouvido, da face e do pescoço
Q20-Q28	Malformações congênitas do aparelho circulatório
Q30-Q34	Malformações congênitas do aparelho respiratório
Q35-Q37	Fenda labial e fenda palatina
Q38-Q45	Outras malformações congênitas do aparelho digestivo
Q50-Q56	Malformações congênitas dos órgãos genitais
Q60-Q64	Malformações congênitas do aparelho urinário
Q65-Q79	Malformações e deformidades congênitas do sistema osteomuscular
Q80-Q89	Outras malformações congênitas
Q90-Q99	Anomalias cromossômicas não classificadas em outra parte

Tabela 1 – Tabela CID-10 (Elaborado pela autora, 2020).

Relacionado à malformação, abaixo são apresentadas as principais definições a respeito, sendo que os defeitos podem ocorrer de forma isolada ou em conjunto, de acordo com Brown (1999):

- Malformação – Defeito morfológico de um órgão, parte de um órgão ou região maior do corpo resultante de um processo de desenvolvimento intrinsecamente anormal.
- Disrupção – É um defeito morfológico de um órgão, parte de um órgão, ou uma região maior do corpo, resultante do desarranjo de um processo de desenvolvimento originalmente normal ou de uma interferência sobre ele.
- Deformidade – Forma ou posição anormal de uma parte do corpo causada por forças mecânicas.

- Displasia – Anormalidade da organização das células ao formarem tecidos e seus resultados morfológicos. Em outras palavras, uma displasia é o processo (e consequência) da disistogênese.

As malformações congênitas maiores são as que trazem consequências médica ou estética ao portador, exigem tratamento cirúrgico/ ortopédico e/ ou deixam sequelas funcionais importantes, frequentemente, exibem alta morbimortalidade e algumas são letais. Cerca de 2% dos recém-nascidos vivos apresentam este tipo de defeito, sendo que 1% têm malformações múltiplas, de acordo com de Castro *et al.* (2006). Na Tabela 2 são apresentados alguns dos diagnósticos mais frequentes relacionados a malformações.

Diagnóstico	Nº Casos	Casos por 1000
RN com anomalias associadas	421	10,70
Fosseta pré-auricular	300	7,63
Mama extranumerária	235	5,98
Nevos pigmentados	222	5,65
Ortolani positivo	214	5,44
Apêndice pré-auricular	141	3,59
Hemangiomas	121	3,08
Polidactilia	96	2,44
Artéria umbilical única	89	2,26
Pé torto (diversas formas)	88	2,24
Síndrome de Down	77	1,96
Hipospádias (diversos graus)	66	1,68
Lábio leporino com palato fendido	25	0,63
Anencefalia/espinha bífida/encefalocele	22	0,55
Sindactilia	13	0,33
Microcefalia	10	0,25
Atresia do esôfago	6	0,15
Atresia intestinal	6	0,15
Hidrocefalia	6	0,15
Onfalocele	3	0,08

Tabela 2 – Diagnósticos de anomalias congênitas frequentes em nascidos vivos (baseado em estudo realizado no Hospital dos Servidores Públicos do Estado de São Paulo, com 39.295 recém-nascidos examinados, em 2008) (Elaborado pela autora, 2020).

Conforme Paul e Dredze (2011), a mineração de dados já vem sendo aplicada em vários domínios, sendo que, na área da saúde, podemos citar, por exemplo, o uso para obtenção de dados de redes sociais e, posteriormente, técnicas de agrupamento para avaliar a saúde de uma população e comunidade. Quanto às formas de visualização desses dados, foco desta dissertação, podemos citar, principalmente, a forma de grafos e agrupamento. A forma de grafo nos permite o acompanhamento e comportamento de uma determinada informação, tal como ilustrado na Figura 4. Nela, são apresentados grafos de evolução de três assuntos diferentes, sendo da esquerda para a direita: difusão a respeito do tema de terremotos, tornado e do lançamento de um novo jogo para videogame.

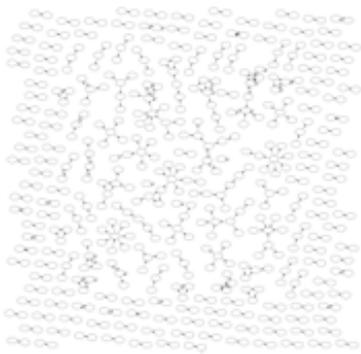


Figure 6: Earthquake information diffusion network.

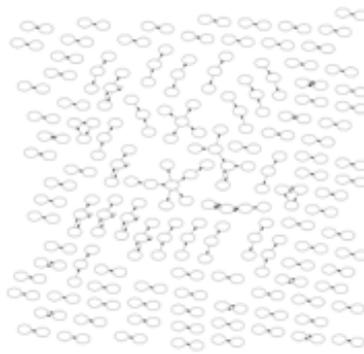


Figure 7: Typhoon information diffusion network.

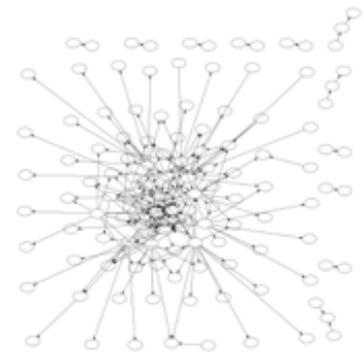


Figure 8: A new Nintendo game information diffusion network.

Figura 4 – Rede de difusão para três assuntos diferentes (SAKAKI *et al.*, 2010).

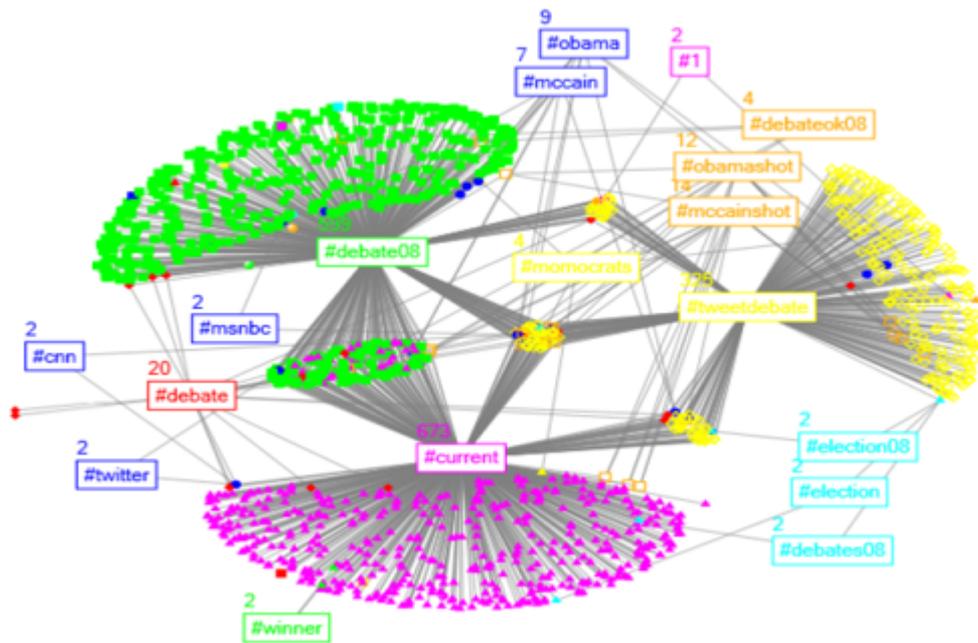


Figura 5 – Demonstração dos dados no formato de agrupamento (SHAMMA *et al.*, 2009).

A visualização em forma de agrupamento pode ser visualizada na Figura 5, acima. Esse tipo de apresentação tem por objetivo a separação dos dados em conjuntos similares e também pode demonstrar a similaridade entre os conjuntos de acordo com sua proximidade.

Conforme as figuras anteriores, podemos observar como um dos grandes desafios, tanto da mineração de dados, quanto da visualização, é a apresentação das informações. Nesse sentido, a visualização na área de saúde é um desafio, pois possui um número muito grande de atributos e fatores relacionados. Por fim, outro fator determinante na forma de visualização dos dados é o que o resultado deve demonstrar. Portanto, podem ser considerados demais gráficos e formas de visualização, como, por exemplo, no caso da malformação congênita, uma forma de visualização seria um mapa demonstrando a quantidade de casos em cada estado ou cidade.

### 2.3 CONCEITOS E TÉCNICAS DE VISUALIZAÇÃO DE DADOS

De acordo com Ouellete (2019), apresentar dados de uma forma fácil de serem compreendidos é uma tarefa complexa para o *designer*. A autora ainda salienta que o objetivo da visualização de dados é facilitar a tomada de decisões de maneira mais efetiva, baseada nos resultados dos gráficos. Ainda conforme a autora, a visualização de dados deve possuir algumas características para que sejam fáceis de usar por todos, sendo elas:

- Facilidade de uso: precisa ser simples para que todos os usuários usem, inclusive os mais iniciantes;
- Recursos: a visualização de dados deve permitir que o usuário navegue pelos dados de forma a interagir com os dados e criar visualizações poderosas;
- Suporte: ser consistente, caso aconteça algum problema;
- Integração: deve haver uma integração das várias fontes de dados, tal como provedores de serviço, *e-mails*, entre outros.

Segundo Nediger (2019), não é possível simplesmente adicionar números em um infográfico e esperar que as pessoas entendam o que se espera dizer com aquilo, de forma que as visualizações de dados devem ser claras e fáceis de entender. Também se deve levar em consideração o tipo de visualização mais adequada para cada caso. As visualizações podem ser separadas em: informativas, comparativas, alterações, organização e relacionamentos.

A visualização informativa pode ser visualizada na Figura 6 e é caracterizada pela apresentação de um único ponto de destaque ou importância nos dados, tal como o uso de um grande número ou pictográfico de forma a demonstrar uma proporção simples. Já na Figura 7, apresentamos alguns exemplos de visualização de dados comparativos, que são utilizados de forma a comparar duas ou mais categorias, ou mostrar a composição de um conjunto. Para isso, pode-se utilizar, por exemplo, uma nuvem de palavras, gráficos em barras, bolhas ou torta.



Figura 6 – Visualização de dados informativa (adaptado de Nedinger, 2019).



Figura 7 – Visualização de dados comparativa (adaptado de Nedinger, 2019).

A visualização de alteração dos dados, apresentada na Figura 8, adiante, é utilizada para demonstrar a evolução/alteração dos dados em um local ao longo do tempo, por meio do uso de linhas do tempo, gráficos em linha, área ou mapa.



Figura 8 – Visualização de dados de alteração (adaptado de Nedinger, 2019).

Os dados também podem ser usados para organizar as informações, sendo a visualização mais conhecida para isso a de lista (Figura 9). Outras visualizações para organização de dados frequentemente adotadas são: diagrama de fluxo para dados mais complexos, diagrama de Venn para agrupamentos de dados, mapa mental para agrupar conexões, pirâmide para demonstrar hierarquia, tabelas, gráficos de barras ordenadas, entre outras.



Figura 9 – Visualização de dados organizacionais (adaptado de Nedinger, 2019).

Por fim, a última forma de visualização destacada por Nedinger (2019) é a de relacionamentos. Essa forma (Figura 10) é utilizada para relevar relacionamentos, correlações e distribuições, por meio do uso de histogramas, gráfico de dispersão e multi-serial.



Figura 10 – Visualização de dados de relacionamento (adaptado de Nedinger, 2019).

Conforme as formas de visualização apresentadas, Nedinger (2019) complementa que, para a definição de qual visualização utiliza, deve-se levar em consideração as que não exijam muito contexto para entendimento, mesmo que para isso seja necessário fazer vários gráficos com diferentes formas de apresentar as informações. Chen *et al.* (2009) fornecem mais alguns exemplos de formas de visualização de dados conforme o tipo de informação que se tem, de acordo com o Tabela 3.

<b>Informação das categorias</b>	<b>Exemplos</b>
Informação sobre o conjunto de dados de entrada	
Geometrias abstratas e características temporais	Esqueletos, características, eventos
Propriedades topológicas	Árvore de contorno para volume de dados, área topológica de vetores, gráfico de rastreamento para dados multivariáveis
Indicadores estatísticos e informações de medidas	Histograma, correção, importância, certeza, entropia, informações mútuas, complexidade e estatísticas locais
Informações sobre os resultados	Histograma de cor, nível e agrupamento
Informações sobre os processos	Padrões de interação, proveniência
Informações sobre as percepções dos usuários	Tempo de resposta, acurácia

Tabela 3 – Exemplos de informações usadas na visualização (adaptado de Chen *et al.*, 2009).

Outro aspecto a ser considerado é quanto a tecnologia utilizada para a visualização de dados. A Tabela 4 apresenta-se um resumo das ferramentas de visualização baseadas em *javascript* destacadas por Liu *et al.* (2017).

<b>Nome do Framework</b>	<b>Formato de entrada dos dados</b>	<b>Gráficos renderizados por:</b>	<b>Tipos de gráficos e mapas</b>	<b>Tipo de licença</b>
D3.js	JSON, CSV e XML	HTML 5 Canvas, SVG e CSS	Poderosa galeria D3 com múltiplos gráficos e maps, incluindo mapa mundial e US.	BSD-3
Chart.js	Javascript API	Apenas HTML 5 Canvas	8 tipos de gráfico, incluindo 23 modelos	Licença MIT
Fusion Charts	JSON, XML	SVG, VML	+ de 90 gráficos e 1.000 tipos de mapas, incluindo todos os continentes e a maior parte dos países	Gratuito para edição básica e avançada versão comercial

Flot Chart	Javascript API	Apenas HTML 5 Canvas	Gráfico de linhas, pontos, área, barras e combinações. Não suporta mapas.	Grátis
ZingChart	Javascript API	HTML 5 Canvas, SVG e VML	Variedade de gráficos na galeria ZingChart. Suporta quase todos os países e áreas.	Gratuito para edição básica e avançada versão comercial

Tabela 4 – Ferramentas para visualização baseadas em *javascript* (adaptado de Liu *et al.*, 2017).

Já o Tabela 5 apresenta ferramentas baseadas em outras linguagens de programação destacadas por Liu *et al.* (2017).

<b>Ferramenta</b>	<b>Formato de entrada dos dados</b>	<b>Linguagem base</b>	<b>Características</b>	<b>Tipo de Licença</b>
Gephi	CSV, arquivos excel	Java, OpenGL	Construída pelo OpenGL, baseado em forçar o layout dos algoritmos em forma de gráficos.	GUN, GPL
Nodebox 3	CSV	Python e Closure	Integra todas as partes funcionais em nós.	GPL
Ggplot 2	R, API	R	Plotagem baseada em camadas e gráficos compostos de camadas.	GUN, GPL e V2
Processamento	Vários formatos de entrada estão disponíveis nessa biblioteca	Java, plugin para python e Javascript	Integrado com o mecanismo OpenGL, possui mais de 100 bibliotecas oferecidas para expansão de uso.	GPL, LGPL

JpGraph	CSV, de bases de dados como SQL	PHP	Pequenos tamanhos de imagens geradas. Imagens anti-spam e 3D são suportadas.	QPL aberto, uso comercial pago
---------	---------------------------------	-----	--	--------------------------------

Tabela 5 – Ferramentas para visualização baseadas em outras linguagens (Liu *et al.*, 2017).

Na proposta de solução desta dissertação, é utilizada uma ferramenta de visualização de dados, o Tableau<sup>3</sup>, que é uma plataforma de análise que permite construir um *dashboard* com desenvolvimento rápido e funcionalidades muito avançadas e, de uma única fonte de dados, um arquivo de extensão HYPER. Esta é uma tecnologia de processamento de dados de grandes volumes que oferece um desempenho de análise rápido considerando *big data*.

#### 2.4 DASHBOARDS PARA APOIO A DECISÃO EM SAÚDE

De acordo com Gomes (2017), um *dashboard* é um painel que apresenta as informações centralizadas juntamente com indicadores e métricas. Geralmente, é um painel utilizado para tomada de decisão e permite que se navegue pelos dados ou seja mudada sua forma de visualização. Segundo o autor, por meio desses dados agrupados disponíveis, é possível se planejar e implementar melhorias nos processos, corrigir falhas e pensar em novas estratégias.

Para Nogueira et al. (2017), um *dashboard* pode ser separado em três tipos, conforme apresentado na Tabela 6. Essa separação é feita levando-se em conta os seguintes fatores: horizonte temporal, periodicidade da informação, destinatários e objetivos. Um exemplo de estudo de caso com resultados sobre *dashboard* é apresentado por Melo e Pelissari (2016). Eles utilizaram a empresa Unicesumar<sup>4</sup> como fonte de análise, sendo essa uma instituição de ensino que atua no mercado de educação. Segundo os autores, a ferramenta utilizada anteriormente para a extração de dados pela empresa encontrava-se descentralizada e apresentava um grau de detalhamento limitado, visto que não possuía uma tecnologia de associação de dados. Dessa maneira, muitos dos relatórios eram simples e feitos em planilhas dinâmicas apenas para atender necessidades pontuais. Como muitas vezes os dados vinham de locais distintos, não havia grande confiança.

<sup>3</sup> Disponível em: <<https://www.tableau.com/>>. Acesso em: 25 mar. 2021.

<sup>4</sup> Disponível em: <<https://www.unicesumar.edu.br/home/>>. Acesso em: 25 mar. 2021.

<i>Dashboards</i> estratégicos	Têm por base a metodologia do <i>Balanced Scorecard</i> de David Norton e Robert Kaplan. Estes, partindo da missão e objetivos estratégicos definidos para uma organização, criam objetivos específicos e mensuráveis, permitindo desta forma a monitorização da sua performance. A definição de um <i>set</i> de Indicadores Chave de Performance (KPIs) é um dos elementos críticos de um <i>dashboard</i> estratégico. Os <i>dashboards</i> estratégicos são utilizados para a monitorização do progresso da organização relativamente aos seus KPIs estratégicos.
<i>Dashboards</i> táticos ou analíticos	Utilizados tradicionalmente em funções de monitorização, os <i>dashboards</i> táticos (também denominados Analíticos) permitem aos seus utilizadores a deteção de tendências internas e de mercado relacionadas com os objetivos definidos, como sejam a análise temporal de vendas ou a evolução de defeitos numa linha de produção ao longo de vários períodos temporais e por comparação com um <i>benchmark</i> definido. Através da análise de comportamentos anteriores é possível definir-se a temporização de lançamento de um novo produto ou a definição de manutenção preventiva ou corretiva.
<i>Dashboards</i> operacionais	Enquanto os <i>dashboards</i> estratégicos e táticos têm características comuns no que diz respeito à condução e orientação para os objetivos, os <i>dashboards</i> operacionais são utilizados para a monitorização <i>online</i> e em tempo real de indicadores pré-definidos. Por comparação direta com o objetivo <i>standard</i> definido e, tendo como finalidade provocar a ação sem grandes necessidades de análise, permitem o alerta e a informação necessária para a correção imediata do desvio.

Tabela 6 – Tipos de *dashboards* (Nogueira *et al.*, 2017).

Apesar do banco de dados ser bem completo, as informações eram dispersas e, muitas vezes, para cruzar os dados, era necessário que as atualizações fossem feitas manualmente. Portanto, existia uma grande dificuldade na emissão dos indicadores mensais e semanais, pois a atualização demandava um grande tempo e esforço, sendo uma tarefa lenta que precisava ser repetida a cada emissão de novo relatório. Na Figura 11 apresenta-se um exemplo de como se apresentavam os indicadores antigos.

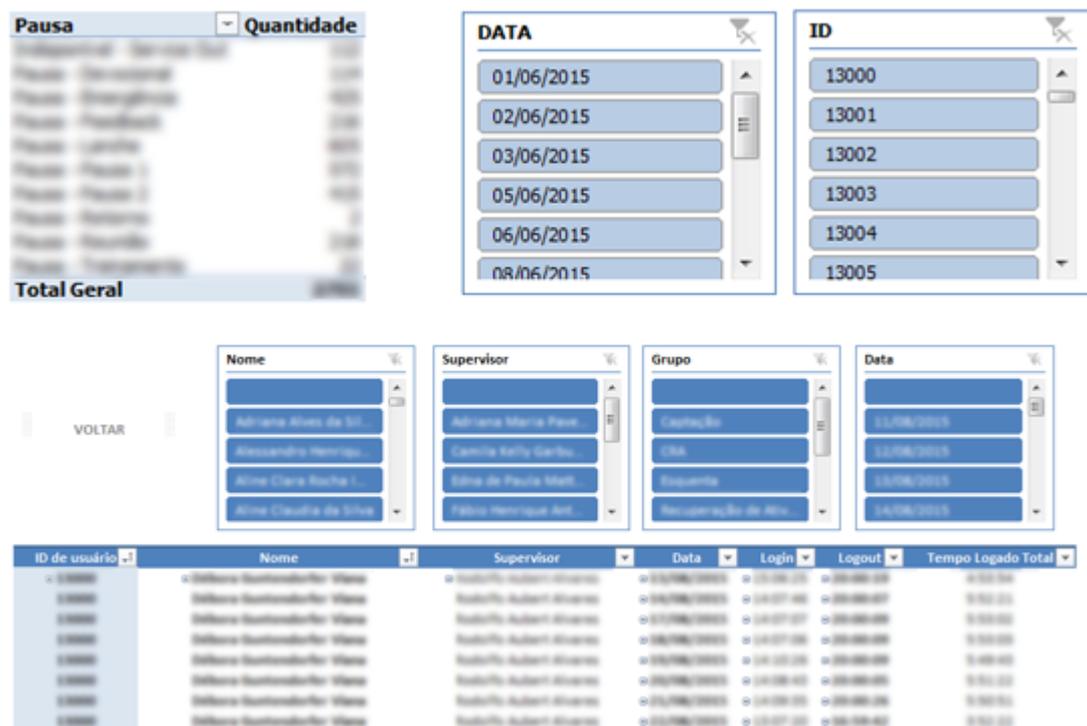


Figura 11 – Exemplo de como se apresentavam os indicadores antigos (Melo e Pelissari, 2016).

Foi necessária a realização de análise dos dados brutos para que se escolhessem os métodos de visualização que seriam utilizados, e somente então foram implementados alguns dos indicadores que a empresa deseja avaliar, tal como apresentado na Figura 12.



Figura 12 – Indicadores implementados (Melo e Pelissari, 2016).

Para Nogueira *et al.* (2017), um *dashboard* voltado para a área da saúde têm como foco indicadores específicos da área, bem como outros indicadores indiretos com impacto na referida área. Os dados são normalmente obtidos por meio de diversos sistemas e a qualidade das informações depende da atenção e cuidado com que foram inseridos no sistema, bem como com o cuidado e detalhamento com que foram inicialmente preenchidos. Em algumas situações, esses dados não são disponibilizados em tempo real.

Um exemplo da complexidade da junção das informações relacionadas à saúde pode ser, por exemplo: informações de farmácias quanto as vendas de remédios, informações de acidentes de trânsito, informações de hospitais e UTI's, informações de bombeiros e atendimentos de primeiros socorros, informações vindas de planos de saúde, informações de

atendimentos particulares (médicos, dentistas, fisioterapeutas, etc.), informações de postos de saúde, informações referentes a vacinas, entre outras. Quanto mais serviços disponibilizam informações em tempo real, ou utilizam sistemas integrados, maior o poder de decisão, porém, maior a dificuldade de apresentados dos dados relacionados (NOGUEIRA *et al.*, 2017). Em alguns casos, é possível prever uma epidemia de acordo com remédios que estão sendo consumidos na farmácia, prever a falta de algum medicamento, monitoramento da eficiência dos serviços, análise, entre outros, mas se fazem necessárias bases de dados integradas e visualizações de dados adequadas.

Nogueira *et al.* (2017) também relatam o elevado custo para manutenção da saúde pública e que o uso dos *dashboards* pode auxiliar na contenção dos custos, no envio da quantidade correta de profissionais para cada posto e no desenvolvimento de novos produtos, tais como vacinas e outros remédios. Nas Figuras 13 e 14 são apresentados exemplos de dois *dashboards* da área da saúde. O primeiro é do United Kingdom National Health Service (UK NHS), que une informações sobre monitoramento e episódios de urgência tal como a quantidade de pacientes e infecções e o tempo para apresentar os dados. O segundo, também do UK NHS, apresenta informações a respeito da utilização de medicamentos, incluindo uma métrica de utilização por estado.

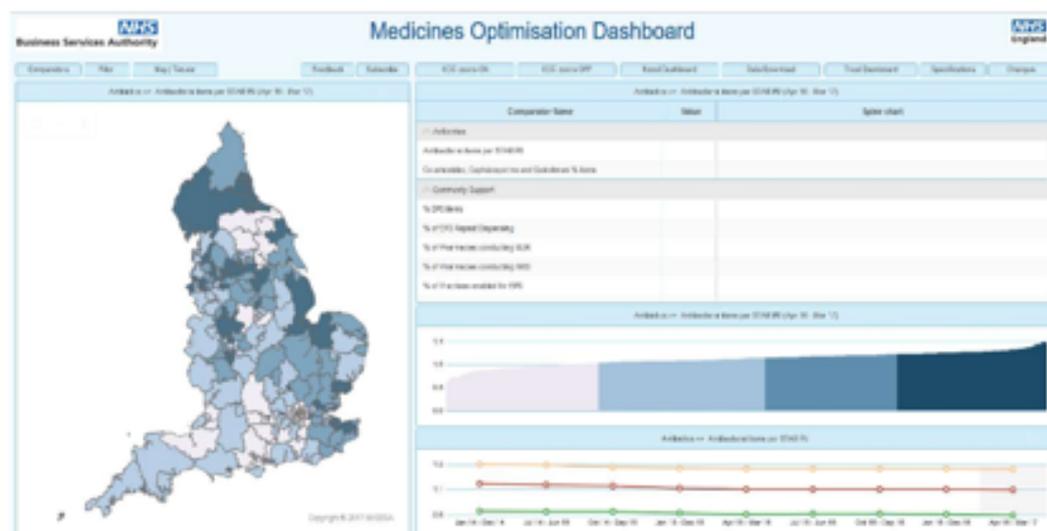


Figura 13 – Exemplo UK NHS – *Dashboard* Operacional Clínico para de Monitorização de Episódios de Urgência (Nogueira *et al.*, 2017).



Figura 14 – Exemplo UK NHS – *Dashboard* de Otimização da Utilização de Medicamentos (Nogueira *et al.*, 2017).

## 2.5 CONCEITOS DE INTERAÇÃO HUMANO-COMPUTADOR ADOTADOS NESTA DISSERTAÇÃO

Segundo Gomes *et al.* (2007), o crescimento do mercado de *software* brasileiro demanda a necessidade constante por inovação e competitividade, bem como qualidade. Com isso, a usabilidade também passou a fazer parte do processo de qualidade, sendo inserido como um diferencial competitivo na criação de valor do produto final. É exigido dos profissionais envolvidos, como engenheiros, *designers* e desenvolvedores, muito mais do que fazer o *software* funcionar, mas algo que agrada o usuário final e seja simples de usar, o que exige uma gama muito maior de conhecimentos, ferramentas e técnicas que precisam ser aplicadas. Em interfaces interativas, tais como *dashboards*, a usabilidade é um elemento central que deve ser levada em consideração no seu *design* e avaliação.

Nielsen (1995) destaca que grande parte dos conhecimentos a que os engenheiros e demais profissionais do desenvolvimento de *software* recebem são relacionados a implementação das tecnologias, não sendo necessariamente centrada no usuário. Ainda para Nielsen (1995), o UCD é essencial para prover a facilidade no aprendizado, melhorar a eficiência de uso, facilitar a memorização, diminuir taxas de erro e aumentar a satisfação.

Segundo Barbosa e Silva (2010), no UCD, abandonamos uma concepção comum na computação em que se pensa nos dados, na lógica e na interface (lado esquerdo da Figura 15)

para o caminho inverso (lado direito da Figura 15). Nesta abordagem comum em IHC, a base começa pela interface, que através da lógica dará acesso aos dados.

O fluxo de mineração de dados e a descoberta de conhecimento são comumente desenhados a partir da análise dos dados para a interpretação dos dados pelos especialistas, tal como o processo apresentado na Figura 2. Esse processo é considerado nesta dissertação (ver capítulo 4), mas também introduzimos uma visão centrada no usuário, percorrendo o caminho do entendimento das visualizações pelos usuários e seu contexto, para o desenvolvimento das interfaces para posterior análise dos dados.

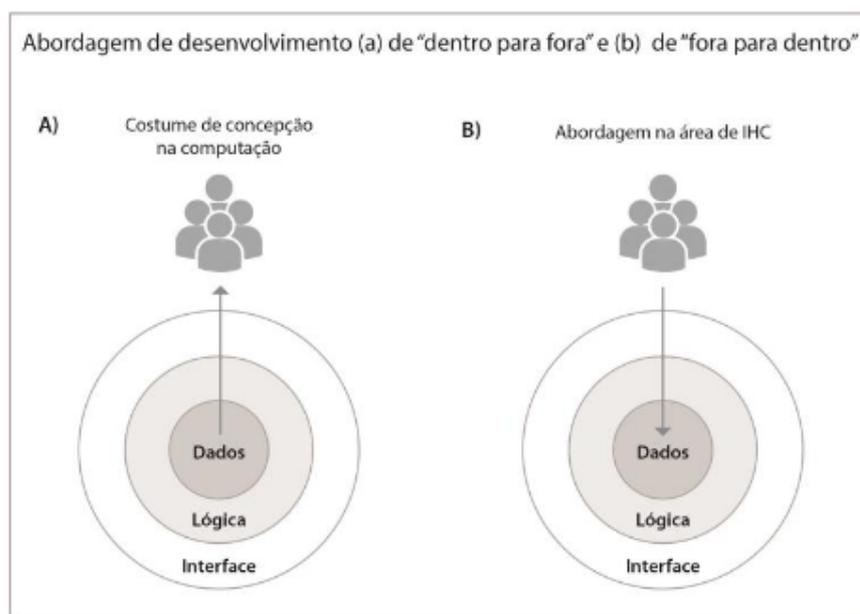


Figura 15 – Diferentes formas de interação com os sistemas (Barbosa e Silva, 2010).

Considera-se, portanto, como parte do método de desenvolvimento de *dashboards*, o projeto, implementação e avaliação das interfaces. Segundo Gomes *et al.* (2007), o projeto está relacionado ao comportamento humano e as técnicas são usadas para criar as interfaces; a implementação trata das técnicas utilizadas para construção das interfaces (princípios de projetos de interfaces, prototipação em papel, prototipação em computadores); e, por fim, a avaliação diz respeito sobre como torna-se avaliado e medida a usabilidade da interface.

De acordo com Santos (2012), dois aspectos centrais a serem considerados são a comunicabilidade e usabilidade em um processo de interpretação e ação (Figura 16). A comunicabilidade permite ao usuário perceber e compreender o que o sistema se propõe a

realizar, ou seja, qual seu objetivo. E a usabilidade permite que o usuário consiga fazer uso do sistema, medindo o grau de satisfação do usuário ao interagir com as interfaces criadas.

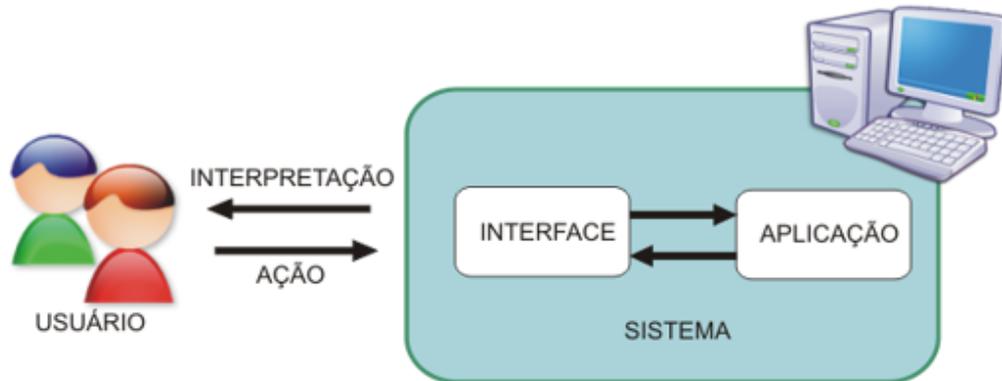


Figura 16 – Interação usuário – sistema (Santos, 2012).

Schneider (2014) considera o termo ergonomia, que trata do estudo e adequação do trabalho as características do usuário, para dar prazer na execução de suas tarefas e preservar sua saúde mental e física. Então, a ergonomia na IHC irá tratar não só os aspectos físicos ao incluir o grau de usabilidade e acessibilidade fornecido por cada sistema. Ao se projetar sistemas deve-se levar em consideração o modelo mental dos usuários, de forma que possa orientar como o sistema deve ser construído, levando-se em consideração a linguagem, padrões e lógica dos usuários. Esses fatores irão determinar a aceitabilidade de um sistema (Figura 17).

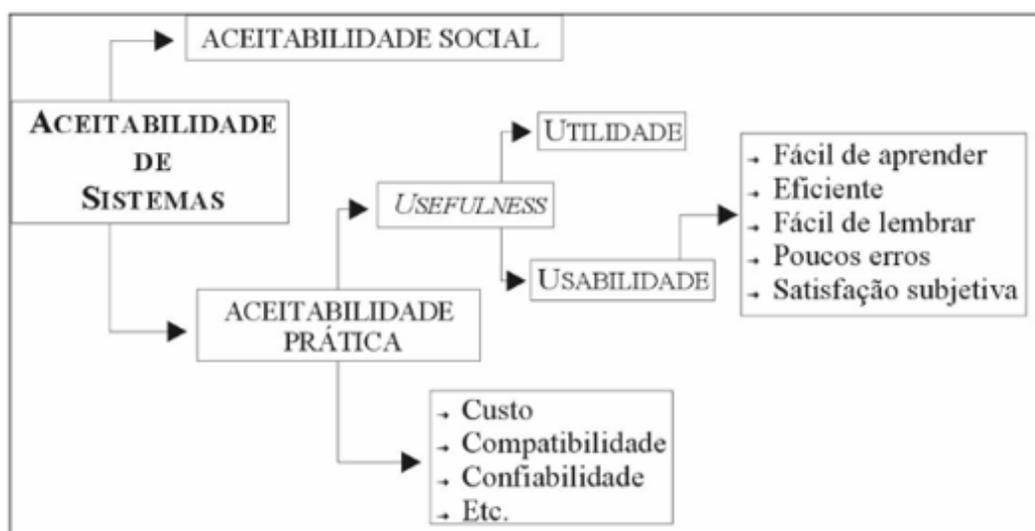


Figura 17 – Aceitabilidade de um sistema (Nielsen, 1995).

Destaca-se que a visualização de dados em *dashboards* está intimamente ligada a IHC. O usuário, ao ler e examinar as informações definidas na tela, pode pensar sobre elas e criar uma resposta sobre o que está vendo, seja alguma forma de agir, seja buscando mais informações dentro do sistema apresentando, recomeçando o ciclo de interação.

A ISO 9241-210 (2010) especifica um *framework* padrão para o UDC ou HCD (*Human-Centered Design*), que estrutura fatores humanos com o objetivo de tornar os sistemas mais fáceis de serem utilizados ao planejar o sistema para os usuários que farão uso do mesmo no futuro. Isso é feito através de fases iterativas que otimizam o design, de acordo com Hartzler *et al.* (2015).

Hartzler *et al.* (2015) apresenta um processo de UCD composto pela sequência de etapas apresentada na Figura 18. Esse processo inicia com entrevistas e pesquisas para que se entenda o contexto de uso, o meio de utilização, o perfil e as características dos usuários e as tarefas que esses desempenham, bem como o ambiente em que o sistema é utilizado. Com base nisso são criados os requisitos de usuário, especificando tudo o que o sistema deve possuir de forma a atender as necessidades dos usuários. Na sequência são criados alguns protótipos de forma a validar se a solução atende os requisitos identificados, e o processo pode ser repetido várias vezes até que todas as necessidades dos usuários tenham sido atendidas.

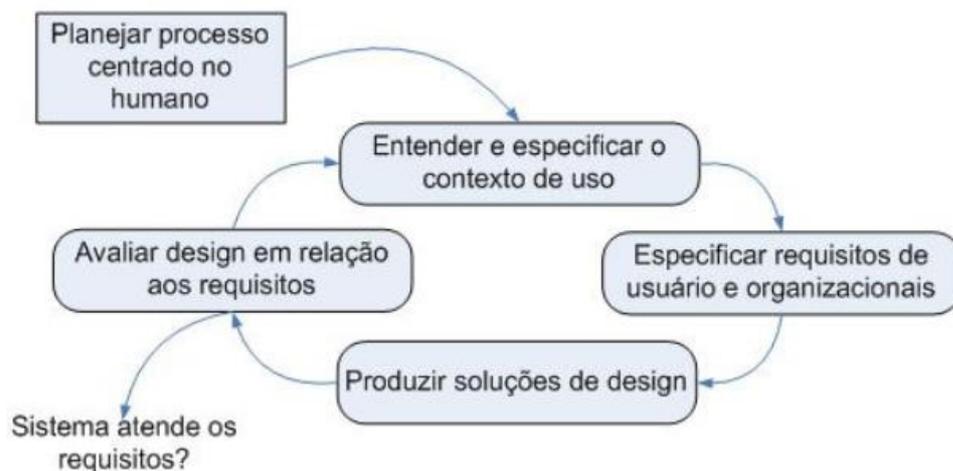


Figura 18 – Processo do design centrado no usuário (Agnes *et al.*, 2015).

De acordo com Ghazali e Ariffin (2014) as abordagens do UCD na área da saúde sugeriram a partir de 1992, seis anos após a introdução do tema por Donald Norman. A pesquisa também foi base para outros estudos, como o *design* centrado no paciente (PCD). De acordo

com Dixit *et al.* (2020), no contexto de visualização em saúde, pode-se iniciar com visualizações básicas, tendo como foco atender a necessidade do usuário ao invés de gastar tempo e energia com visualizações complexas. A solução deve ser aprimorada no decorrer do tempo e com base no retorno dos usuários, sendo que o ritmo de trabalho dos profissionais também precisa ser considerado. Balleggard *et al.* (2008) também menciona a existência de 3 fases comuns no UDC, estando os usuários envolvidos desde as primeiras fases de desenvolvimento, sendo: análise, projeto e implementação. A etapa de análise faz o acompanhamento dos usuários para levantamento dos requisitos, a fase de projeto faz a criação dos protótipos e é na fase de implementação que é feita a avaliação dos projetos propostos.

Johnson *et al.* (2015) destaca que o desenvolvimento de visualizações centradas no usuário em saúde é um processo que coloca as necessidades dos usuários finais como ponto chave das aplicações, tornando primeiramente esse objetivo como prioritário e depois analisar como os dados serão obtidos e agregados. De acordo com Chan (2002), no contexto da saúde, o usuário também precisa ser alfabetizado na utilização do sistema, por meio de uma terminologia correta para o público-alvo. Além disso, deve ser levando em conta os usuários ocasionais e que as métricas apresentadas atendam as necessidades de todos os perfis de usuário. Dessa maneira, diferentes níveis de usuários devem ser considerados (Cotterman e Kumar, 1989), o que é particularmente importante para o contexto desta dissertação, uma vez que ela visa o design de *dashboards* para o público em geral. A Figura 19 destaca que conforme a experiência do usuário diferentes aspectos devem ser priorizados no *design* de *dashboards*.

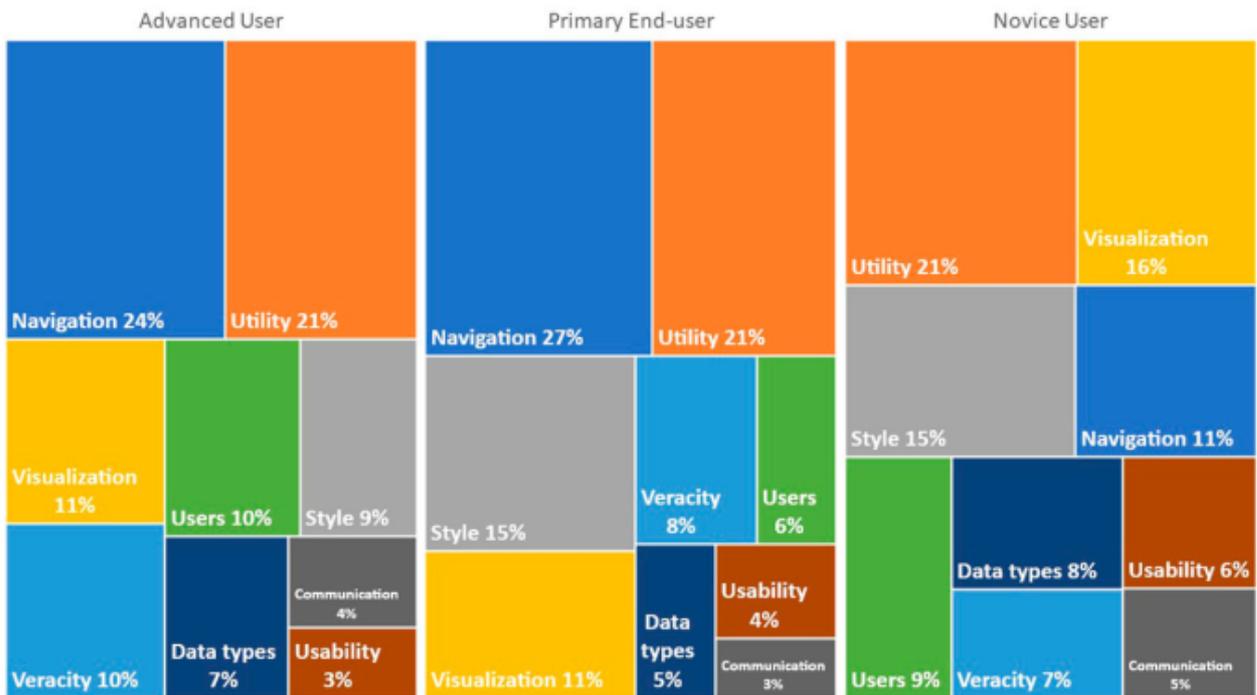


Figura 19 – Análise dos aspectos a serem considerados na visualização conforme o nível de experiência do usuário (Young *et al.*, 2020).

De acordo com Anacleto e Fels (2014) o UDC também pode ser considerado uma filosofia de design que enfatiza a importância das exigências do usuário final, além disso o UCD altamente iterativo e flexível. De acordo com Brennan *et al.* (2010), no contexto de sistemas para saúde, o UCD possibilita a seleção de métodos responsivos e permitem os refinamentos necessários para adequar a ferramenta as particularidades do usuário e do meio onde o sistema estará inserido.

Conforme Hartzler *et al.* (2015), em dashboards para a área de saúde, há o problema de apresentar os resultados em um nível de detalhamento dos dados apropriado, conforme o usuário que estiver visualizando, como por exemplo paciente, provedor, clínica ou instituição. Também devem ser considerados a forma de visualização (por exemplo, cronogramas temporais e periódicos instantâneos) e o nível de interatividade com as informações dentro do contexto em que se encontra o paciente, por exemplo as visualizações rápidas versus as visualizações filtradas para análises mais detalhadas e aprofundadas. Com base nisso os autores apresentam a Tabela 7 com recomendações de como projetar *dashboards* para a área de saúde.

	<b>Funcionalidades de cuidado com os pacientes</b>	<b>Funcionalidades de Melhoria da Qualidade</b>
<b>Chave de suporte</b>	Utilizado para avaliar o progresso do paciente, aconselhar pacientes e compreender a eficácia do tratamento.	Utilizado para aumentar a satisfação do paciente e estabelecer métricas de qualidade assistencial.
<b>Etapa 1 Amostra de wireframes</b>	Fornece visualizações no nível do paciente para monitoramentos individuais. Fornece visualizações em nível de provedor para monitoramento geral de pacientes.	Fornece visualizações em nível de clínica para comparar os dados agregados ao longo do tempo. Fornece visualizações em nível de instituição para comparar os dados agregados em todas as ferramentas.
<b>Etapa 2 Protótipos estáticos</b>	Fornece análises interativas para explorar dinamicamente dados do paciente fora das visitas. Fornece linhas de tempo interativas que criam um gráfico de progresso do paciente durante as visitas	Fornece dados instantâneos e pré-definidos dos pontos finais de monitoramento de cuidados do paciente. Fornece suporte para análises interativas para explorar dinamicamente agregar dados do paciente.
<b>Etapa 3 Protótipos interativos</b>	Fornece dados dinâmicos em nível de paciente e provedor para profunda interação. Fornece relatórios com dados imprimíveis para compartilhar com pacientes individuais.	Fornece visualizações de dados dinâmicas em nível de clínica e instituição para interação. Fornece instantâneos de dados trimestrais de dados agregados do paciente.

Tabela 7 – Recomendações na construção de dashboards a partir de cada etapa (Adaptado Hartzler *et al.*, 2015).

Além do UCD, outro conceito chave adotado nesta dissertação é o conceito de usabilidade. Conforme a ISO 9241-11, usabilidade é definida como medida na qual um produto pode ser usado por usuários específicos para alcançar objetivos específicos com efetividade, eficiência e satisfação num contexto específico de uso. Para tanto, deve ser considerado quem vai usar o sistema, quais os objetivos do usuários, contexto e ambiente de utilização e o que tecnicamente serão necessários. De acordo com a ISO 9241-11, o nível de usabilidade alcançado dependerá das circunstâncias específicas nas quais o produto é utilizado. A Figura 20 apresenta a estrutura de usabilidade segundo a ISO 9241-11.

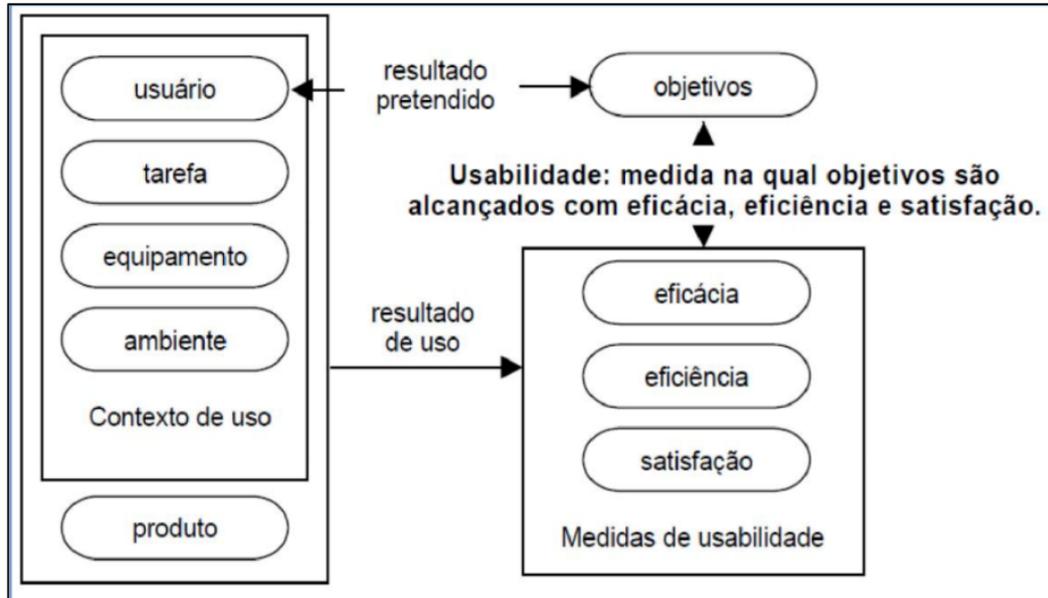


Figura 20 : Estrutura de usabilidade. Fonte: ISO 9241-11 - ABNT, 2011.

Cybis (2003) classifica várias técnicas para avaliar a usabilidade um produto (*Software e Hardware*). Conforme a Figura 21 as técnicas podem ser classificadas entre: prospectivas que buscam a opinião dos usuários; preditivas ou diagnósticas que buscam prever erros de projetos de interfaces sem a participação direta de usuários; e objetivas ou empíricas que observam os usuários ao interagir com o sistema.

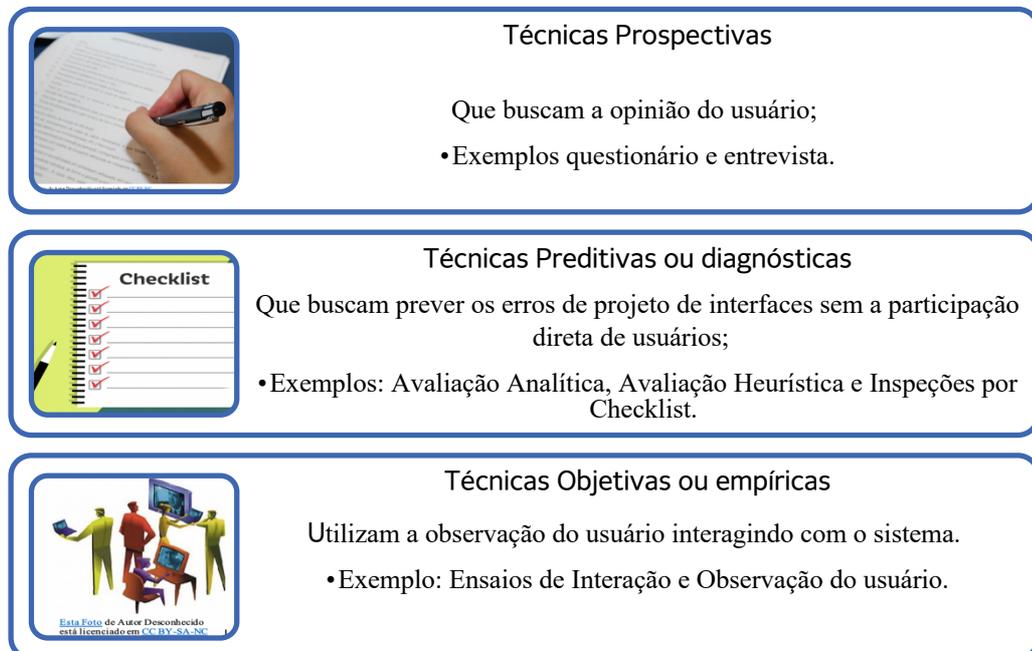
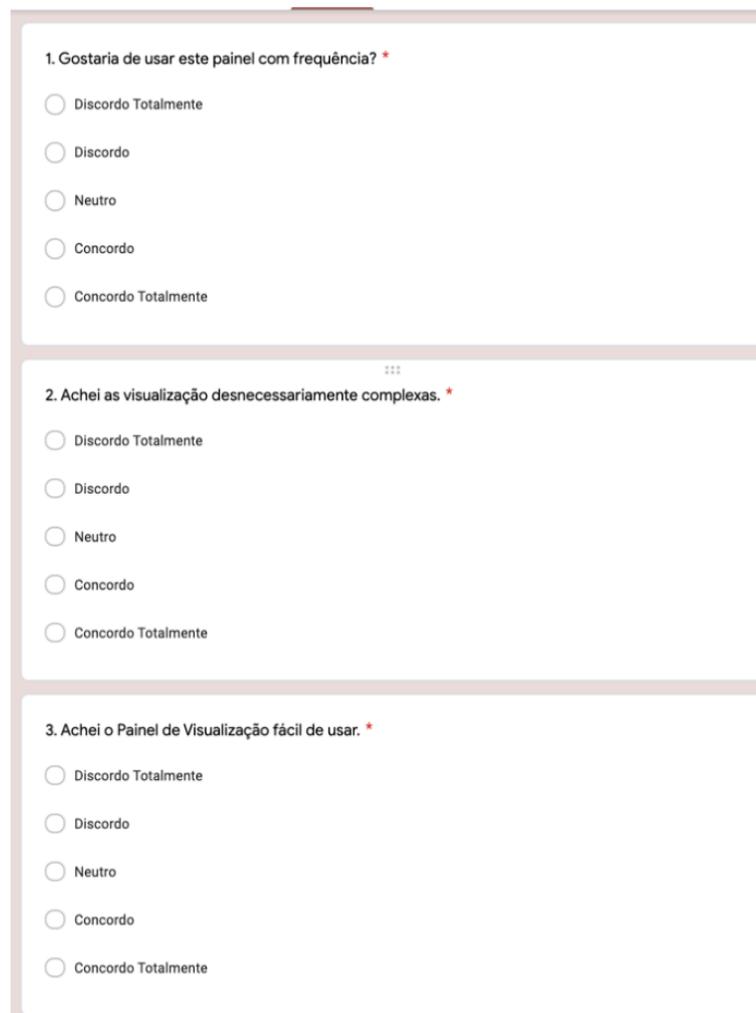


Figura 21 : Técnicas para avaliação de usabilidade. Fonte: Adaptado de Cybis, 2003.

Nesta dissertação foi realizada avaliação com técnicas prospectivas (Capítulo 5), por meio de consulta de questionário baseado no modelo SUS (*System Usability Scale*), cujo referencial é descrito a seguir. O modelo SUS foi desenvolvido por Brooke (1996) para medir o nível de satisfação do usuário em uma escala numérica. Este modelo contém dez questões com 5 opções de resposta que visa medir a usabilidade. O SUS é fácil de administrar e possui boa confiabilidade e referencial que auxilia na interpretação dos dados, bem como não há direitos autorais. O público alvo que responde o questionário SUS assinala sua resposta utilizando a escala *Likert* que varia de “Discordo totalmente” a “Concordo totalmente”, conforme pode ser observado na Figura 22.



The image displays three separate questionnaire items, each within a light-colored rectangular box. Each item consists of a question followed by five radio button options arranged vertically. The options are: 'Discordo Totalmente', 'Discordo', 'Neutro', 'Concordo', and 'Concordo Totalmente'. The first item asks about the frequency of using the dashboard. The second item asks if the visualization is unnecessarily complex. The third item asks if the visualization is easy to use. Each question is marked with a red asterisk to indicate it is required.

1. Gostaria de usar este painel com frequência? \*

Discordo Totalmente

Discordo

Neutro

Concordo

Concordo Totalmente

...

2. Achei as visualização desnecessariamente complexas. \*

Discordo Totalmente

Discordo

Neutro

Concordo

Concordo Totalmente

3. Achei o Painel de Visualização fácil de usar. \*

Discordo Totalmente

Discordo

Neutro

Concordo

Concordo Totalmente

Figura 22 : Exemplo de Aplicação da SUS



### 3 REVISÃO E TRABALHOS RELACIONADOS

Este capítulo apresenta uma revisão exploratória sobre trabalhos e técnicas de visualização de dados em saúde (Seção 3.1), bem como sintetiza e compara características dos trabalhos mais próximos (Seção 3.2) com esta dissertação. Entende-se por trabalhos mais próximos aqueles (da busca exploratória) que propõe *dashboards* voltados para a área de saúde em uma abordagem centrada no usuário ou são voltados ao público geral.

#### 3.1 VISUALIZAÇÃO DE DADOS EM SAÚDE

Conforme citado por Carroll *et al.* (2014), nos últimos 20 anos, houve um aumento da necessidade de informatização referente à visualização analítica na saúde pública, como resultado do aumento de investimentos em sistemas de informação. Assim, esses investimentos geraram novas ferramentas para atividades de saúde pública, incluindo sistemas desenvolvidos em nível federal, estadual e local, bem como pesquisas organizacionais.

Avanços na área eletrônica e interoperabilidade, tecnologia computacional, biotecnologia e outros métodos fizeram com que a área de saúde pública se desenvolvesse e colocassem esses avanços em prática (BOS; BLOBEL, 2007). Khan *et al.* (2010) ressaltam que esses avanços foram particularmente grandes para o apoio a administração de doenças infecciosas em pandemias ou potenciais eventos de bioterrorismo.

Segundo Koch e Denike (2009), uma das primeiras pesquisas a serem transformadas em imagem foi a de John Snow, que criou um mapa de casos de cólera em Londres. Para os autores, essa visualização teve um papel importante no campo da epidemiologia, suporte a comunicação, agregação, análise e uso de dados para testar hipóteses e tomada de decisão.

Conforme Hay *et al.* (2013), as doenças por infecção são um fardo para a população, sejam medidas de forma programática, sejam para outros fins, podendo também ser analisada em termos de distribuição geográfica, fatores de riscos clínicos, dados demográficos ou fontes de exposição. Andre *et al.* (2007) citam ainda que, embora muitos recursos de rotina dos relatórios de saúde pública incluam curvas epidêmicas e mapas, novas formas de visualização, como gráficos de redes sociais e árvores filogenéticas, têm sido cada vez mais usadas para caracterizar surtos de doenças.

Podemos ver, conforme ilustra a Figura 23, adiante, o crescimento das publicações envolvendo GIS, epidemiologia molecular e análise de rede social no PubMed. Sua frequência

em termos de informática biomédica (usabilidade, registro eletrônico de saúde) é mostrado para comparação. AvRuskin *et al.* (2004) citam que esses tipos de dados complexos permitem que profissionais e pesquisadores da saúde pública possam integrar, sintetizar e visualizar informações relativas à vigilância de doenças, sua prevenção e controle. Os autores explicam que o rastreamento da distribuição de doenças com ferramentas de localização geográfica auxilia os profissionais citados a detectar o agrupamento de doenças e analisar sua disseminação nas comunidades e entre territórios, de forma a prever surtos.

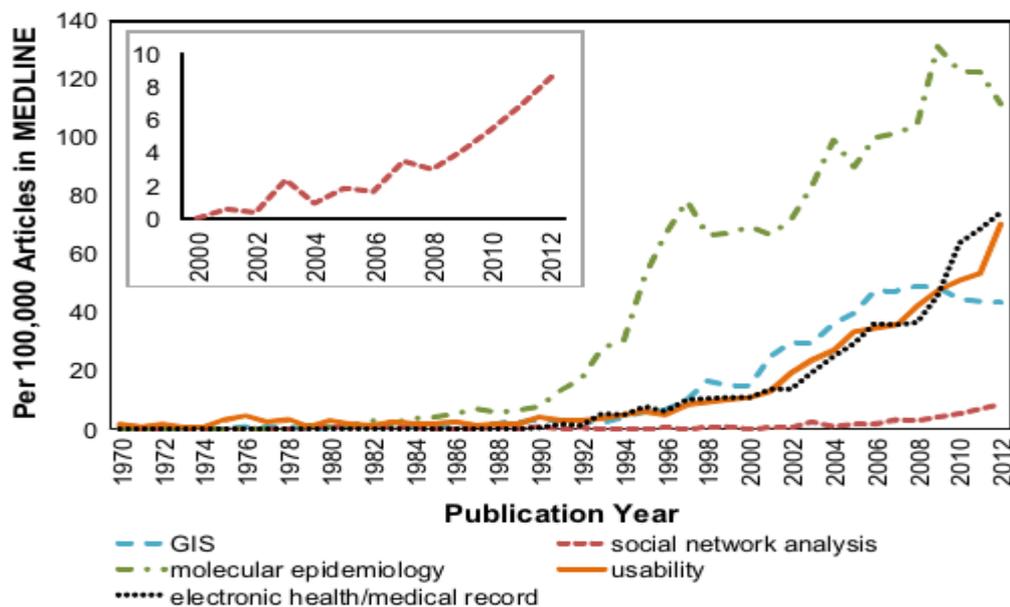


Figura 23 – Pesquisa por palavra-chave para GIS, epidemiologia molecular e análise de rede social no PubMed (Carroll *et al.*, 2014).

Podemos citar a necessidade e possibilidades com a integração de várias fontes de dados tais como: mortalidade, dados clínicos, dados geográficos, relacionamentos, genética do paciente e do patógeno, entre outros. Conforme Carroll *et al.* (2014), essas informações podem ser registradas, armazenadas, acessadas, avaliadas e visualizadas via sistema, em diversos formatos. Folorunso e Ogunseye (2008) destacam que as ferramentas de visualização têm o potencial de melhorar a compreensão desses dados, aumentando os recursos de memória e processamento disponíveis para os usuários, reduzindo a busca por informações, potencializando a detecção de padrões e fornecendo mecanismos para inferência.

Esta seção apresenta, assim, alguns trabalhos relacionados que fizeram uso dessas técnicas de visualização para facilitar o entendimento das informações. Liu *et al.* (2017)

realizaram uma pesquisa a respeito do uso da visualização de dados na área acadêmica, tomando como base as técnicas existentes e os sistemas de análise de volume de dados de diversas fontes. Essa área contém fontes abundantes de dados vindos das mais variadas locais, tal como artigos, livros, patentes e relatórios científicos, e ainda com associação de outros dados, como informações dos autores, citações, figuras, tabelas, entre outros. A solução final proposta pelos autores pode ser visualizada na Figura 24, adiante, com a divisão para a coleta de dados, visualização genérica dos mesmos e exemplos típicos de visualização.

Os autores supracitados abordam de forma detalhada a coleta de dados, testes com ferramentas e implementação da solução. Conforme apresentado, o processo de visualização de dados acadêmicos combina a teoria da visualização científica, visualização de informação e visualização analítica, o que permite aos cientistas terem um melhor entendimento da estrutura e dinamismo da ciência de uma forma visual. A visualização também é amplamente utilizada na área de patentes, conforme apresentado no trabalho de Federico *et al.* (2016). Outras pesquisas relacionadas são a pesquisa de Yang *et al.* (2008), que revisa ferramentas comerciais relacionadas a mineração de dados textuais na visualização por busca de patentes, e a pesquisa de Lup *et al.* (2011), que utiliza a visualização na área de recuperação de patentes.

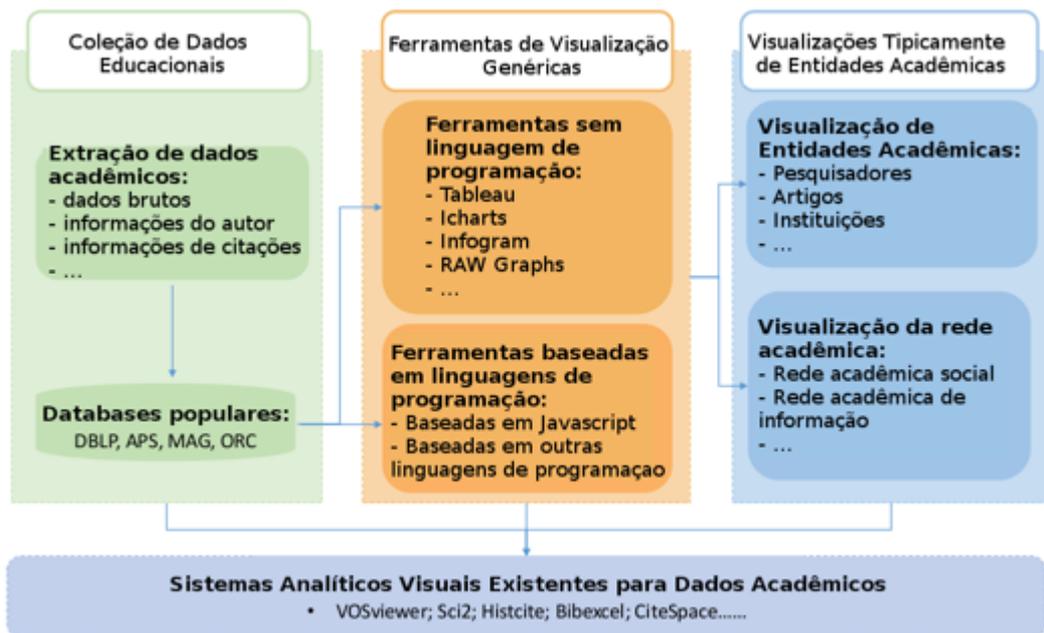


Figura 24 – *Framework* para visualização de dados acadêmicos (adaptado de Liu *et al.*, 2017).

O trabalho de MacNabb e Laramee (2019) apresenta como escrever uma pesquisa na área de visualização de dados, servindo como ponto base para outros trabalhos. Os autores também apresentam alguns exemplos e recomendações de como demonstrar as informações e descrever o processo, como, por exemplo, na Figura 25, em que de acordo com o tipo de dados a ser visualizado a técnica correta a se ser aplicada.

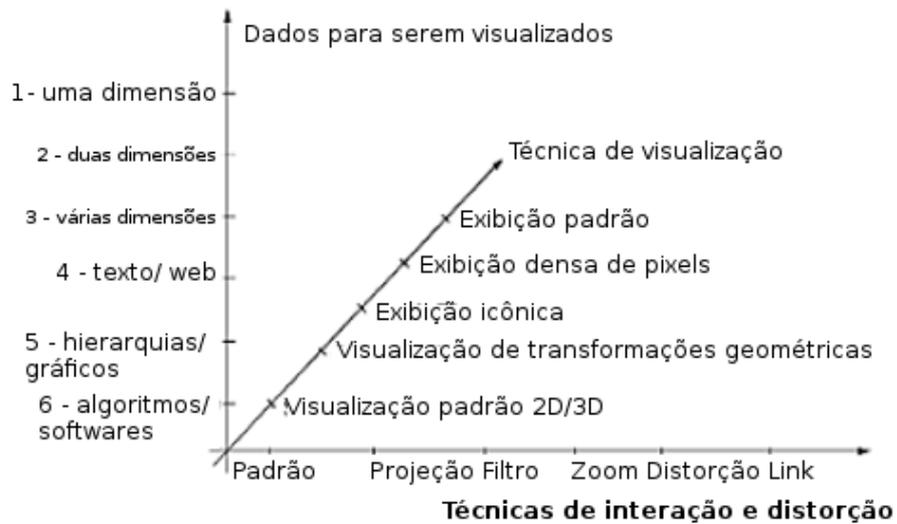


Figura 25 – Técnicas de classificação da visualização (adaptado de MacNabb e Laramee, 2019).

Dentre as técnicas apresentadas, Carroll *et al.* (2014) citam a importância da visualização de dados através de sistemas de informações geográficas, pois essa ferramenta permite que profissionais monitorem e respondam desafios de saúde. Algumas funções dessas ferramentas são: integrar dados de diversas fontes, detecção de grupos e mapeamento de dados.

Dessa forma, a Figura 26 representa casos hipotéticos de doenças infecciosas no estado de Califórnia, no qual cada ponto representa um caso específico de doença. Assim, apresenta-se, na imagem à esquerda, os dados a nível individual, na imagem do meio, os dados agrupados e, por fim, na imagem da direita, os locais com maior incidência de infecção.



Figura 26 – Diferentes visualizações de mapas com informações geográficas (Carroll *et al.*, 2014).

Outro trabalho com visualização em mapas é o de Freifeld *et al.* (2008), com a proposta de criação do aplicativo *web HealthMap.org*, um sistema automatizado para consultar, filtrar, integrar e visualizar relatórios não estruturados sobre surtos de doenças. Para tanto, é feita a integração entre diversas fontes, incluindo mídia de notícias (como *Google News*), contas com curadoria de especialistas (como *ProMED Mail*) e alertas oficiais validados. Por meio do uso de algoritmos de processamento de texto, o sistema classifica os alertas por localização e doença e os sobrepõe em um mapa geográfico interativo, como demonstrado na Figura 27. A precisão dos algoritmos de classificação é medido com base no nível de curadoria humana necessário para corrigir possíveis erros de classificação. Como resultado, o aplicativo representa 87 categorias de doenças e 89 países, sendo que o classificador teve um nível de acerto de 84%.

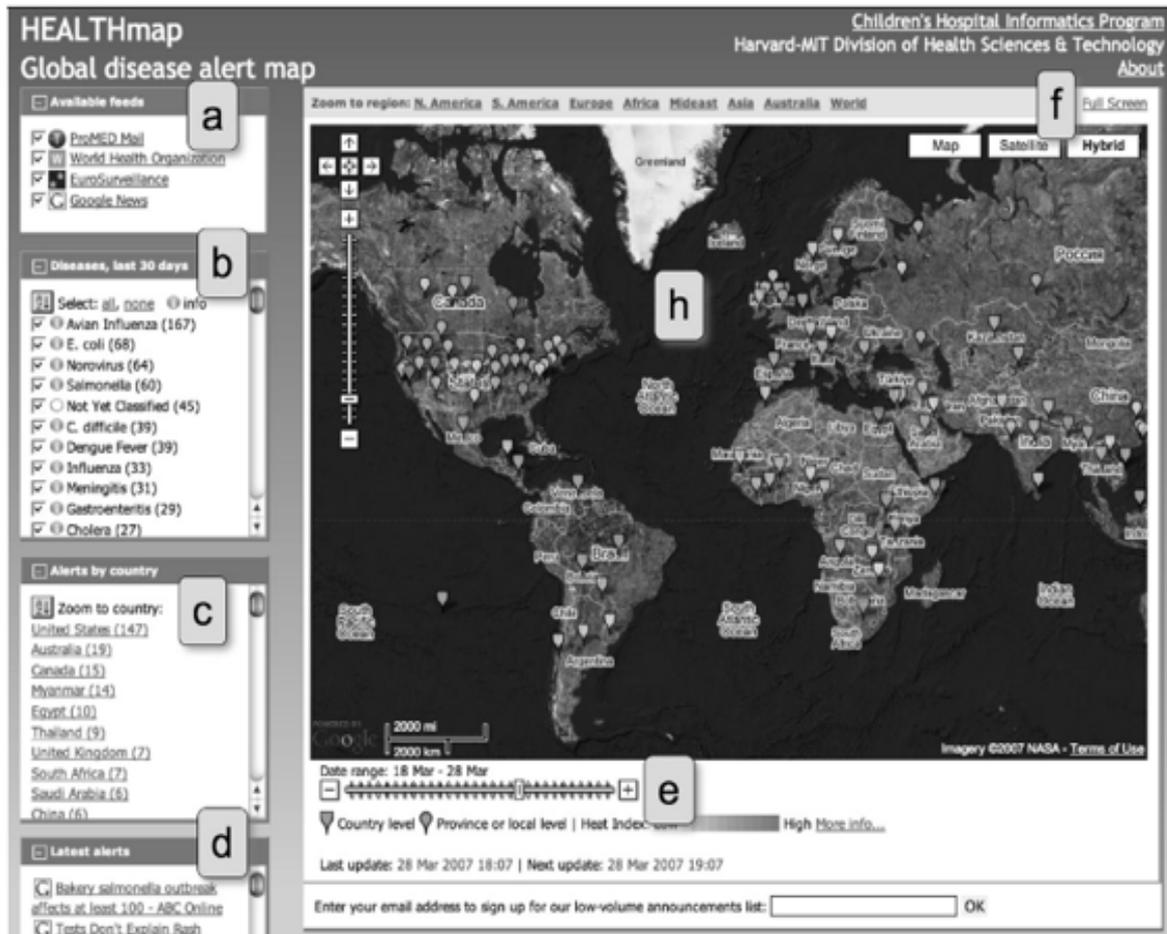


Figura 27 – Interface para navegação do usuário do aplicativo *web* HealthMap.org (Freifeld *et al.*, 2008).

No ano de 2020, com a pandemia causada pelo coronavírus, muitos centros de pesquisa também utilizaram a visualização de dados como forma de manter a população informada sobre a evolução dos casos da doença, como a Universidade de John Hopkings (2020) e a Unicamp (2020). No contexto brasileiro, a Unicamp (2020), criou um mapa interativo de forma a entender como acontece a disseminação dos casos de COVID-19 na região de Campinas, São Paulo, por meio de informações obtidas dos locais onde houve atendimentos relacionados, como o Hospital das Clínicas da Unicamp, Cecom e Caism. Foram, então, catalogados todos os casos de doença desde o dia 3 de março e inseridos no sistema. Dessa forma, o mapa pode ser utilizado para subsidiar ações imediatas no combate a disseminação do vírus. Um dos resultados foi o mapa interativo gerado e apresentado na Figura 28, demonstrando padrões de concentração territorial dos casos confirmados e possíveis tendências de disseminação da doença.

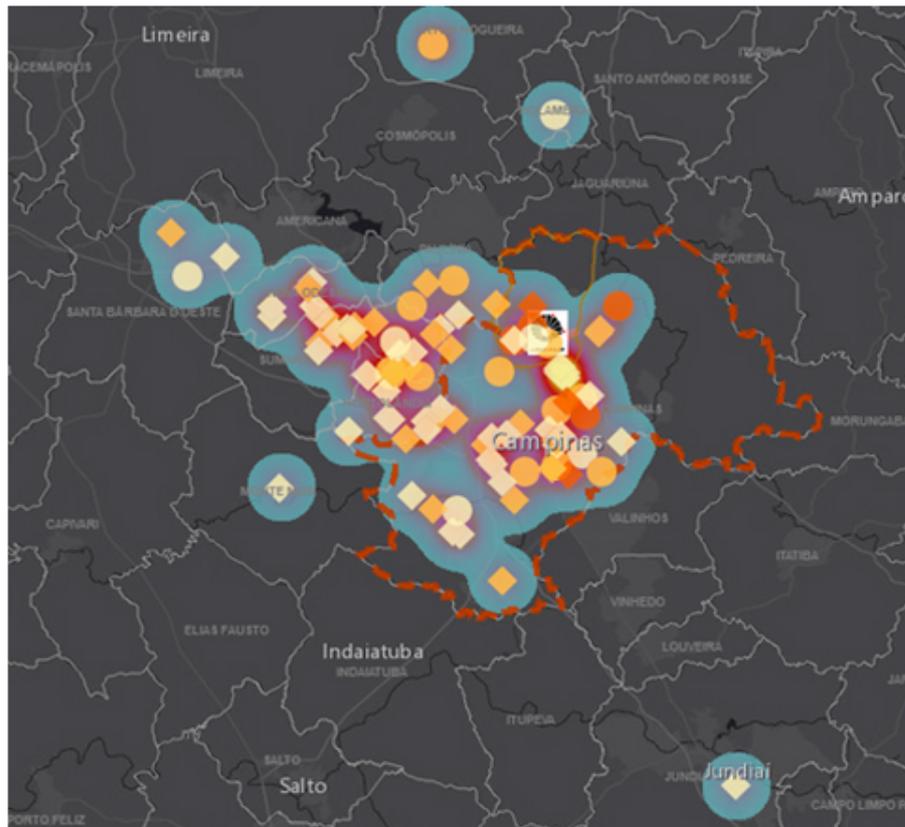


Figura 28 – Mapa gerado a partir de casos de coronavírus, em Campinas, SP (Unicamp, 2020).

Embora a visualização geográfica seja muito prática em questão de visualização das informações, sua criação é um pouco mais complexa, principalmente quando os dados são adquiridos de diversas fontes e precisam ser tratados. Freifeld *et al.* (2008) ilustram (Figura 29) como é a arquitetura da criação de uma aplicação similar a essa. O primeiro passo é a captura das informações utilizadas, que podem vir de diversas fontes, tal como redes sociais, órgãos oficiais do governo, *sites* de notícias, dados de sistemas internos, entre outros. Após isso, a informação precisa passar por um pré-processamento e filtro de forma a colocá-la no padrão com que o sistema irá trabalhar e também para filtrar dados sem relação com a pesquisa. A seguir, os dados devem ser convertidos em um formato de banco de dados, tais como json ou xml, e relacionados com uma determinada localização. Somente a partir desse momento, é criada a interface de visualização e navegação que é disponibilizada para os usuários.

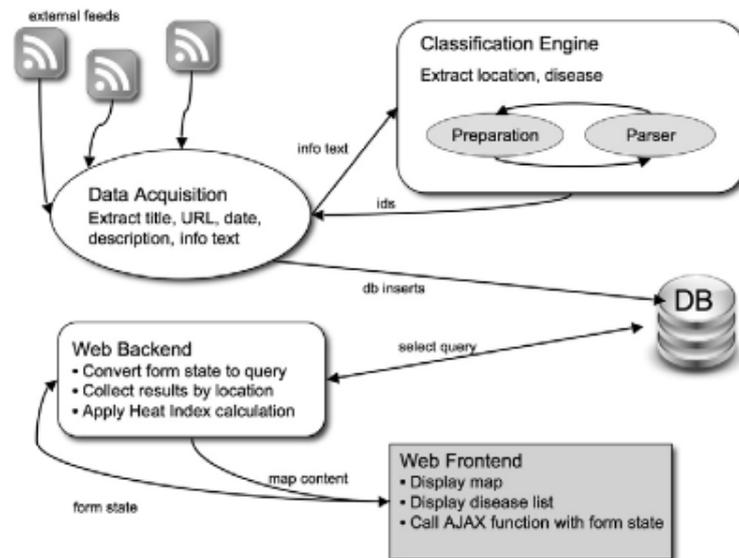


Figura 29 – Arquitetura de um sistema de visualização demográfica (Freifeld *et al.*, 2008).

Khairat *et al.* (2018) apresentam um trabalho amplo de revisão da literatura sobre o uso de painéis de visualização de informações de registros eletrônicos de saúde para profissionais de saúde, sendo que a criação ou uso dessas ferramentas podem facilitar a exibição e recuperação de informações clínicas importantes, auxiliando os resultados e atendimentos aos pacientes. Assim, o foco dos autores foi encontrar artigos que mencionassem essa forma de visualização na área de terapia intensiva, sendo poucas as pesquisas encontradas, restando 17 artigos a serem analisados. A conclusão do trabalho foi que utilizar as soluções de painel de visualização diminuem o tempo gasto na coleta de dados, dificuldade do processo de coleta de dados, carga cognitiva, tempo para conclusão da tarefa, erros e melhoria da consciência da situação, conforme as diretrizes de segurança baseadas em evidências, usabilidade e navegação.

Dowding *et al.* (2015) fizeram uma extensa pesquisa envolvendo diversas visualizações na área médica, em que, de 537 citações, foram filtradas 195 potenciais pesquisas. Dessas, 184 foram descartadas por não estarem acessíveis, não tratarem de *dashboards* ou não conterem dados empíricos, restando, assim, 11 trabalhos. Alguns foram selecionados para serem apresentados a seguir. O estudo de Ahern *et al.* (2012) têm como base pacientes com hipertensão e todos usam painéis eletrônicos para monitorar sua pressão arterial, sendo essas informações utilizadas para sinalizar comunicações com a equipe de cuidados primários. As informações são enviadas via internet e também existe um *dashboard* de luzes para mostrar

pacientes que estão fora de alcance ou precisam de auxílio. Devido a isso, há a necessidade de suporte por parte dos pacientes e a avaliação de qualidade dos dados acaba sendo baixa.

Já o trabalho de McMenamín *et al.* (2011) utiliza como cenário as práticas de primeiros cuidados feito com base em um estudo não comparativo de 15 meses, em que o *dashboard* dos pacientes estava vinculado com o sistema EMR de uma clínica, que utilizava cores como indicadores para marcar as condições de saúde dos pacientes, como, por exemplo, marcar que um paciente é fumante. Cada indicador funciona como uma ferramenta para administrar a saúde do paciente, de forma a verificar outras intervenções relacionadas com o fumo, como o alcoolismo e o câncer de mama. A avaliação da qualidade também foi baixa.

A pesquisa de Zaydfudim *et al.* (2009) teve uma alta avaliação de qualidade, com dados coletados pelas enfermeiras e equipe médica por um período de pré-intervenção, de janeiro de 2005 até junho de 2007. Após isso, houve um período de implementação de 12 meses. A implementação foi composta pela gestão de um painel de controle, apresentado como um protetor de tela com indicadores para cada paciente, sendo os indicadores agrupados em três cores: verde (em conformidade), vermelho (fora de conformidade) e amarelo (alerta). Assim, os gestores recebiam diariamente relatórios sobre esses níveis de conformidade.

Um exemplo de como a integração proposta por Zaydfudim *et al.* (2009) pode facilitar a visualização de dados também é apresentada por Koch *et al.* (2013). Conforme ilustra a Figura 30, a seguir, várias visualizações “tradicionais” são centralizadas de uma maneira integrada, facilitando o acesso a todas as informações em uma única tela.

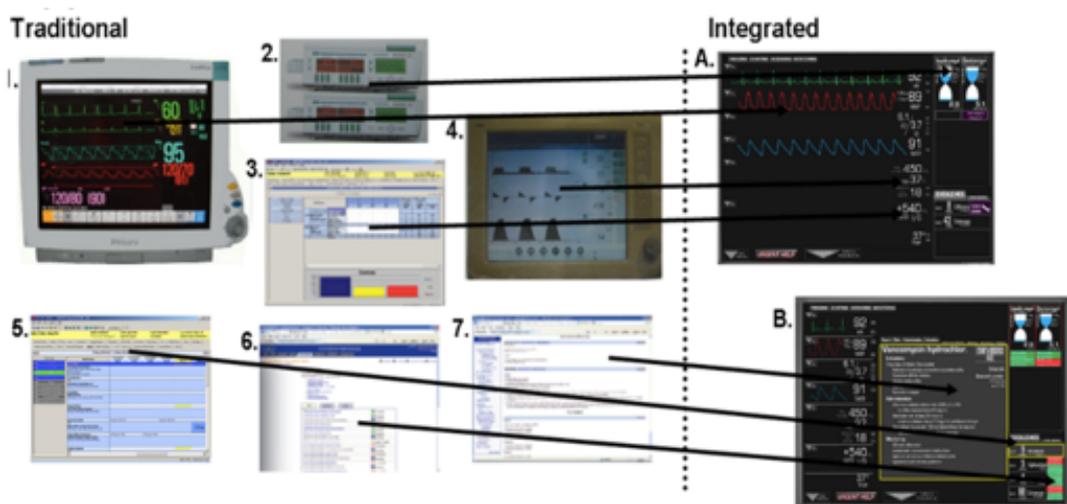


Figura 30 – Demonstração da integração de visualizações (Koch *et al.*, 2013).

O trabalho de Pablato (2009) obteve um resultado bom, tendo como cenário os anestesistas. Os dados foram coletados durante três fases; a primeira, com duração de 20 semanas; a segunda, com duração de 16 semanas; e, por fim, a terceira fase, com duração de 6 semanas. O resultado foi um protetor de tela com painel de controle fornecendo um *feedback* sobre oportunidades para anestesistas na fase de pré-operatório e administração.

Quanto às técnicas de *Machine Learning* utilizadas junto com visualização, Chatzimparmpas *et al.* (2020) apresentam um trabalho com os tópicos mais presentes nas pesquisas, de acordo com cinco categorias. Dessa maneira, apresenta-se que as técnicas mais utilizadas para a visualização de dados é a de agrupamento e análise aplicação.

Já no trabalho de Hartzler *et al.* (2015) o objetivo é a otimização dos painéis de *dashboards* de acordo com as necessidades dos usuários. Para isso foi necessário o engajamento de vários profissionais, como provedores e funcionários que colaboravam com a ferramenta ou tinham interesse na mesma. Durante 15 meses foram desenvolvidas as seguintes atividades: entrevistas com os interessados, construção de um consenso sobre como deveria ser o design dos painéis, design iterativo individual com os usuários de forma a entender as especificações e criar as implementações no *dashboard* e por fim criação de um protótipo a ser implementado na ferramenta existente.

Dixit *et al.* (2020) utilizaram um processo centrado no usuário composto de 5 etapas: obter conhecimento do cenário e domínio, identificar as necessidades do usuário e fontes de dados, projetar e desenvolver as interfaces e refinar iterativamente as mesmas. O cenário utilizado no estudo foi a pandemia causada pelo Covid-19, onde é necessário manter a consistência de informações críticas de saúde tais como os dados da condição do paciente, os tempos de espera e o desempenho da plataforma de telessaúde. Dessa maneira, foram utilizadas visualizações em *dashboards* para dar suporte às operações de telessaúde. O projeto de Dixit *et al.* (2020) também utilizou como base os fundamentos de desenvolvimento de visualizações para telessaúde desenvolvidos por Endsley (1995) e Wright *et al.* (2004). Tal fundamento é composto por 3 etapas: percepção, compreensão e projeção. O nível de percepção seria o mais básico de todos e é onde se percebe o estado atual e é possível fazer o monitoramento dos elementos específicos do ambiente. No nível de compreensão tem-se a capacidade de integrar e sintetizar diferentes elementos de dados, e por fim, no nível de projeção temos a capacidade de prever os estados futuros.

Já Ghazali e Ariffin (2014) faz uma revisão de pesquisas sobre UDC em saúde, compreendendo 60 artigos publicados entre 1992 e Maio de 2014. Conforme o estudo de Ghazali e Ariffin (2014) o maior número de áreas de saúde que fizeram a implementação de UDC foi a diabetes, seguido de monitoramento de idosos. Porém, a maior parte das pesquisas faz a implementação do UDC no sistema geral de saúde, sem ser específico para uma determinada área, conforme ilustrado na Tabela 8.

<b>Domínio</b>	<b>Frequência</b>
Não especificado	16
Diabetes	4
Cuidado de idosos	2
HIV/AIDS, demência, dermatologia, oftalmologia, hipertensão, fisioterapia, doenças crônicas, transplantes de pulmão, entre outros	1

Tabela 8 – Frequência de pesquisas encontradas na literatura de UDC (*Design Centrado no Usuário*) voltados para determinado domínio de saúde (Adaptado Ghazali e Ariffin, 2014).

Por fim, pode ser destacado o trabalho de Beluzo *et al.* (2020) que apresenta um conjunto de ferramentas de análise visuais para *big data* sobre mortalidade neonatal. Esse projeto serviu de inspiração inicial para o desenvolvimento desta dissertação, entretanto possuem objetivos distintos. Em Beluzo *et al.* (2020) foi criado um conjunto de visualizações que auxiliam na compreensão da relação entre variáveis de interesse para facilitar o entendimento das informações públicas relacionadas à mortalidade neonatal, visando a mitigação do problema. Enquanto nesta dissertação o foco está em propor um processo de design de *dashboards*.

### 3.2 CONSIDERAÇÕES SOBRE TRABALHOS RELACIONADOS

Embora as abordagens de UCD e PCD em saúde existam por quase 30 anos, estas não são abordagem dominantes no design de sistemas para área de saúde. Ainda assim, conforme destacado em revisão sistemática realizada por Ghazali e Ariffin (2014), entre os estudos que adotam UCD (e PCD) muitos estão focados em atividades específicas/limitadas baseadas em entrevistas, observações e questionários. Isso resulta na participação limitada do usuário no processo de design.

Os estudos com *dashboards* voltados para a população que fazem uso de UCD são ainda menos frequentes, sendo a maior parte dos estudos são voltados a profissionais de saúde. Apenas um estudo analisado na revisão sistemática apresentada por Dowding *et al.* (2015), sobre *dashboards* para melhoria de cuidados de pacientes, foca a interação com pacientes. Entretanto, os artigos analisados neste capítulo também apontam para necessidade de aprimoramento da interação com *dashboards*, o que destaca a importância e originalidade deste trabalho.

A Tabela 9 apresenta uma análise dos trabalhos relacionados identificados neste capítulo com maior proximidade com o tema desta dissertação, ou seja, que propõe *dashboards* voltados para a área de saúde em uma abordagem centrada no usuário ou são voltados ao público geral. A tabela destaca a base de dados utilizada, o domínio (ou tema específico) de aplicação, o público alvo dos *dashboards* construídos e o processo de design utilizado.

REFERÊNCIA	BASE DE DADOS	DOMÍNIO DE APLICAÇÃO	PÚBLICO ALVO	PROCESSO DE DESIGN
Ahern <i>et al.</i> (2012)	Local/Própria	Hipertensão Arterial	Pacientes com Hipertensão	PCD
Hartzler <i>et al.</i> (2015)	CERTAIN	Cirurgia na coluna	Provedores e funcionários da área da saúde	UCD
Dixit <i>et al.</i> (2020)	-	Covid-19	Provedores e funcionários da área da saúde	UCD
John Hopkings (2020)	Base própria com dados informados por governos	Covid-19	Público geral em escala global	Não identificado
Unicamp (2020)	Base de atendimentos de hospitais ligados à Unicamp	Covid-19	Público geral em escala local	Não identificado
Beluzo <i>et al.</i> (2020)	IBGE, DATASUS, SIM e SINASC	Mortalidade Neonatal	Gestores de Saúde	---
<b><i>Esta Dissertação</i></b>	<b><i>SINASC</i></b>	<b><i>Malformação Congênita</i></b>	<b><i>Público geral (gestantes), profissionais de saúde e gestores</i></b>	<b><i>UCD</i></b>

Tabela 9 – Trabalhos relacionados recentes com foco no público geral ou que usam UCD

Conforme apresenta a Tabela 9, esta dissertação se diferencia das demais por trabalhar com uma base de dados de abrangência nacional (SINASC), seu domínio de aplicação (Malformação Congênita), o público alvo, bem como o processo de design. Isso implica em desafios e contribuições únicas desta pesquisa, tal como o desenvolvimento de um processo de design trabalhar capaz de lidar com bases abrangentes e complexas e produzir soluções de *dashboards* para um público amplo. Para tanto, o processo engloba técnicas de tratamento de grandes volumes de dados, bem como de UCD e avaliação de usabilidade.

## 4 PROPOSTA E RESULTADOS

Neste capítulo, é apresentado o processo de design proposto neste trabalho (seção 4.1), a proposta da solução (seção 4.2) e a prototipação e implementação de um sistema com um conjunto de *dashboards* para análise de dados sobre malformação congênita (seção 4.3).

### 4.1 PROCESSO DE DESIGN

Nesta seção é destacado o processo de design proposto que especifica como as interfaces de visualização de informações e *dashboards* foram concebidas, projetadas, implementadas e avaliadas. Para atingir o objetivo de informar profissionais de saúde, gestores e principalmente a população (com foco em gestantes), o processo proposto é composto de cinco passos, conforme destacado a seguir (Figura 31):

- Passo 1 – *Revisão de soluções tecnológicas e da literatura científica em temas relacionados*. Neste passo, o principal objetivo é substanciar a escolha de métodos de visualização, bem como o desenvolvimento de alternativas de design. No caso da aplicação deste trabalho, a literatura pesquisada focou técnicas e estudos sobre visualização de dados em saúde com foco em soluções tecnológicas que possibilitem o entendimento de conteúdo sobre malformação congênita pela população. Este passo produz um documento com soluções tecnológicas e potenciais soluções de design. Os capítulos 2 e 3 apresentam resultados desta revisão, para o contexto desta dissertação.
- Passo 2 – *Análise e preparação de fontes de dados*. Neste passo, são analisadas as fontes de dados, bem como são preparados os dados necessários para elaboração das soluções de visualização. No caso desta dissertação, utilizamos dados reais e resultados da aplicação de técnicas de mineração de dados em malformação congênita. Para a preparação dos dados são utilizadas ferramentas de Extração, Transformação e Carregamento (ETL - *Extract Transform Load*). A seção 4.2 detalha a execução deste passo, para o contexto desta dissertação. Faz parte deste passo também a análise estatística dos dados, pois a partir dela se pode extrair variáveis de interesse que influenciam nos resultados e visualizações dos dados. A subseção 4.3.2 apresenta análise conduzida no contexto desta dissertação, com o objetivo de identificar e extrair informações (estáticas) que são de relevância a

malformação congênita, que são posteriormente agregadas à apresentação dinâmica e interativa em *dashboards*.

- Passo 3 – *Propor designs de dashboards*. Neste passo, as técnicas visualização selecionadas no passo 1 são integradas em uma solução de *dashboards* com propósitos específicos, tais como a visualização de dados estatísticos sobre a gestações com malformação congênita com propósito de análise de evolução de casos. É proposto neste passo envolver os usuários e outras partes interessadas. Por exemplo, gestantes a procura de informações sobre incidências de malformação congênita, especialistas no domínio (ex: profissionais de saúde e demografia) e gestores. Para tanto, o processo propõe o uso de técnicas de UCD, priorizando aspectos de usabilidade (*cf.*, seção 2.5).
- Passo 4: *Prototipação iterativa do sistema*. Neste passo o design proposto no passo 3 é implementado e aprimorado em um processo iterativo. No caso deste trabalho este processo incluiu a participação de pacientes e especialistas. Aspectos de visualização e usabilidade são avaliados em protótipos funcionais dos *dashboards* propostos. A seção 4.3 apresenta resultados dos passos 3 e 4.
- Passo 5: *Investigação e análise dos resultados*. Neste passo o protótipo final é avaliado por usuários e demais partes interessadas. Uma avaliação com a população (gestantes e interessados no assunto), profissionais de saúde e de demografia em saúde é apresentada no capítulo 5.

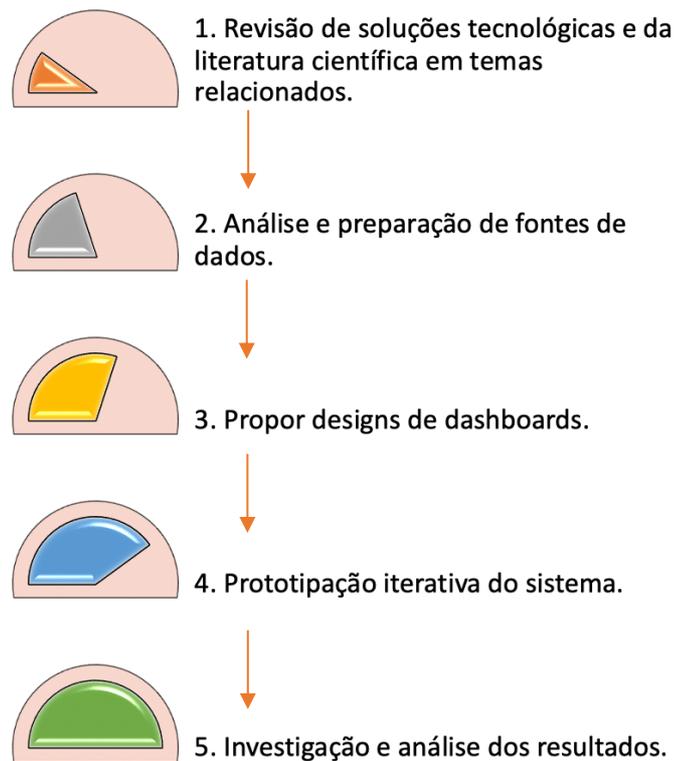


Figura 31 – Processo design (Elaborado pela autora, 2021)

## 4.2 PROPOSTA DE SOLUÇÃO

Nesta seção é apresentado como a solução proposta foi desenvolvida (em uma instância do processo descrito na seção anterior), bem como seus requisitos e concepção inicial. A solução proposta inicia com o uso da base de dados do SINASC, que abrange todo território nacional. Os dados foram reunidos pela Plataforma de Ciência de Dados aplicada à Saúde (PCDaS) relativos a informações sobre nascidos vivos. A base de dados utilizada compreende os anos de 2014 até 2017 e possui mais de 100 atributos.

O primeiro passo realizado com a base de dados consiste na procura de ruídos e padronização de dados. Dessa forma, registros duplicados serão removidos, bem como registros com muitas informações incompletas. Os resultados obtidos e as regras criadas serão enviadas ao *dashboard* e deve permitir a navegação entre eles, bem como o recálculo ao se excluir ou incluir algum atributo. De forma a facilitar a experiência do usuário, utilizamos princípios de usabilidade de forma a tornar o sistema intuitivo para uso, tais como uma interface simples,

campos com nomes que demonstram a que se referem e cores adequadas ao tipo de informação apresentada (destaque, acerto, atenção, entre outros.). Um resumo da mineração dos dados é apresentado na Figura 32.

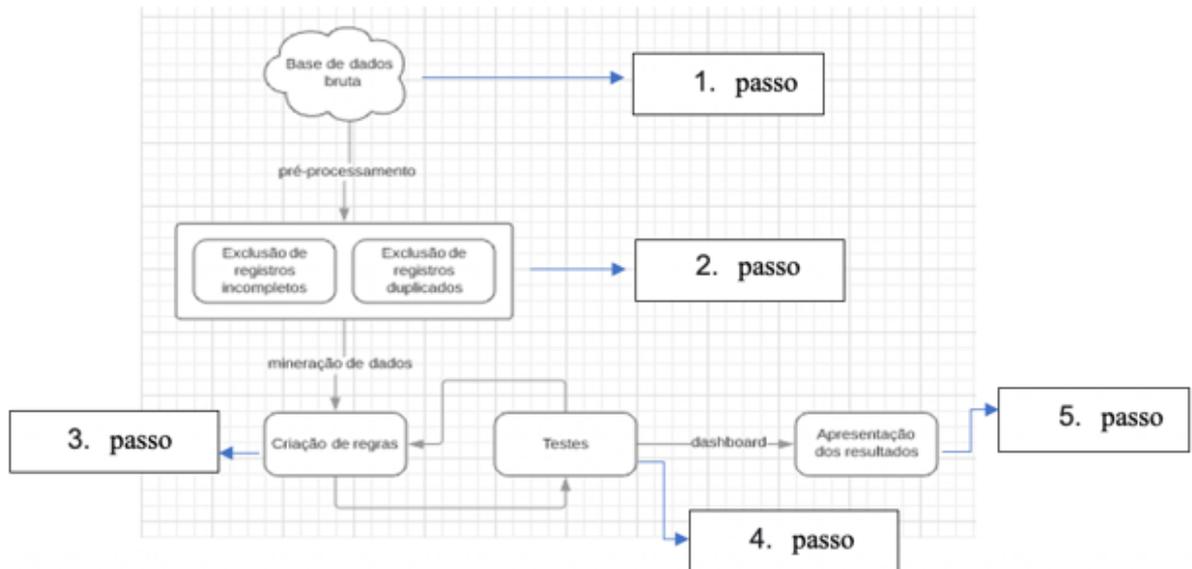


Figura 32 – Solução de mineração adotada (Elaborado pela autora, 2020).

Conforme ilustrado na Figura 32, no primeiro passo, extraímos os dados da PCDaS da base SINASC. No segundo passo, excluimos os dados que não têm relevância para aplicação a malformação congênita, por exemplo, a versão do sistema. No terceiro passo, criamos as regras, separamos os dados da mãe, da gestação e do recém-nascido, separamos pelos estados para entender o que aplicação para estados e Brasil e criamos os cálculos para visualizar as informações de forma clara e precisa. No quarto passo, criamos os testes, avaliamos o *design*, cálculos estáticos e voltamos várias vezes para passo 3 para criar as regras. Finalmente, no passo 5, apresentamos os *dashboards* com os resultados das análises por estados, conforme passos subsequentes no processo de design detalhado na seção anterior.

A solução para criar o processo de visualização de dados, os *dashboard* e a apresentação dos resultados ajudam a entender a motivação final apropriada. É uma etapa crucial para o êxito da visualização de dados, entender e definir o objetivo da visualização. Depois de passar pela mineração de dados, adquirirmos os atributos mais importantes e relevantes. Como forma de garantir que sejam tão completos e precisos quanto possível (que tenha os dados importantes da mãe, gestação e do recém-nascido), devemos filtrar os dados contidos na base, de modo que

o conjunto de final incluía apenas dados de qualidade, completos e necessários para o nosso objetivo, a malformação congênita.

Como diferencial da solução, temos uma análise ampla dos dados da mãe, da gestação e do recém-nascido, com o foco na análise das malformações congênitas (anomalias congênitas) por capítulos, grupos, categoria e subcategoria por estado, e também os dados da mãe, da gestação e do recém-nascido que não teve a malformação congênita. Para entender melhor, criamos mapas com a localização onde se concentram essas crianças. Os dados foram filtrados por estados e por ano da análise. O Apêndice A apresenta a descrição das 119 colunas da base do SINASC. O Apêndice A foi criado utilizando o dicionário da base de dados e conta com os itens que usamos para desenvolver as visualizações desta dissertação.

O processo ETL, realizado nos dados SINASC, foi feito no ambiente da ferramenta Daitaku Data Science Studio (DSS), que centraliza os dados de análises em escala, alimenta análises, aplica modelos de aprendizado de máquina e possui uma conectividade com diferentes formatos de dado para limpeza, conversão, mineração e visualização de dado. Na Figura 33 apresentamos o Diagrama do Fluxo ETL, como visão geral do processo ETL aplicado no SINASC.

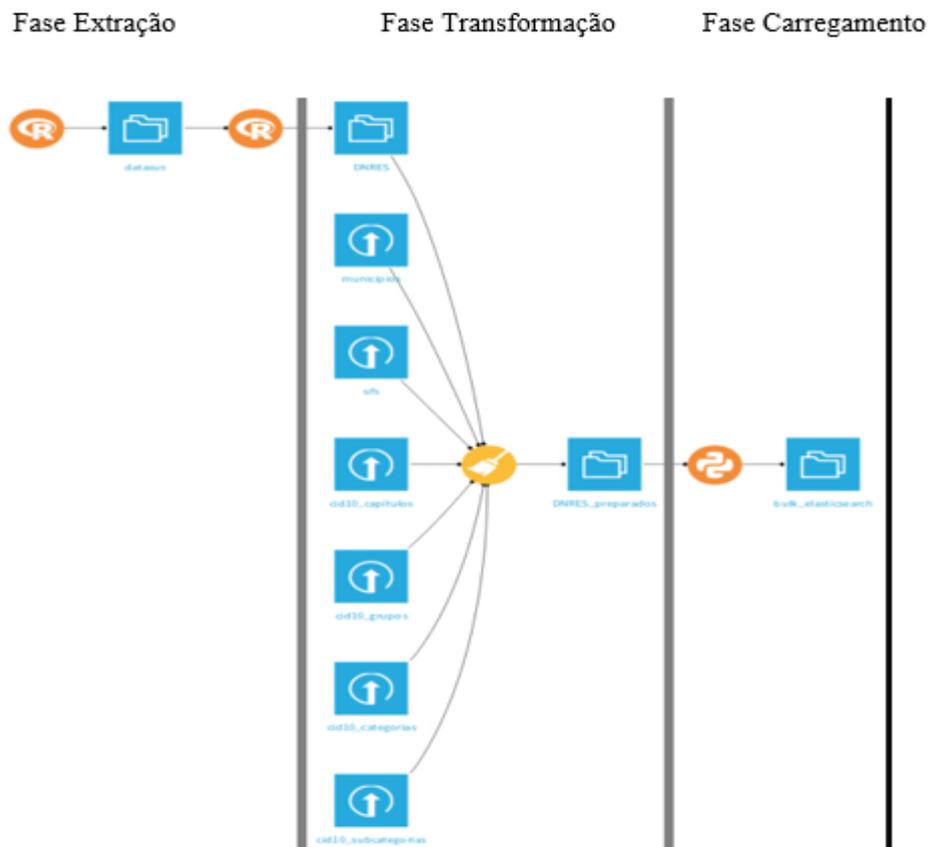


Figura 33 – Diagrama de Fluxo ETL (RAM, 2020).

Na fase Extração, os microdados de cada estado e ano disponível é convertido, lido e unido em único conjunto de dados para tratamento posterior. Na Fase Transformação, são utilizadas as funcionalidades (processos) disponíveis na ferramenta Dataiku Data Science Studio(2020). É neste momento que tratamos os conjuntos de dados, são eliminados os valores inválidos, decodificados e enriquecidos os dados relacionados a município, UF, CID10, data, idade, coordenadas, partos prematuros, conforme o dicionário de dados do SINASC.

Na fase carregamento são indexados os dados já tratados do SINASC na infraestrutura do PCDaS. O resultado de todas as fases é indexado no Elasticsearch<sup>5</sup>, possibilitando o desenvolvimento de *dashboards*.

Conforme destacado anteriormente, na proposta de solução desta dissertação, é utilizada uma ferramenta de visualização de dados, o Tableau, uma plataforma de análise que permite construir um *dashboard* com desenvolvimento rápido e funcionalidades muito avançadas e, de

<sup>5</sup> <https://www.elastic.co/pt/what-is/elasticsearch>

uma única fonte de dados, um arquivo de extensão HYPER, uma tecnologia de processamento de dados de grandes volumes que oferece um desempenho de análise rápido considerando *big data*.

HYPER é resultado de um *workflow* no Tableau Prep (ferramenta de ETL da Tableau) que faz a união das informações, particionadas anualmente no site <https://bigdata-arquivos.icict.fiocruz.br/SINASC/ETLSINASC.zip>, com os dados de 2017, 2016, 2015 e 2014. Foram filtradas apenas as informações da mãe, gestação e recém-nascido das colunas do conjunto de dados do Dicionário de dados do SINASC<sup>6</sup>. Assim, o *dashboard* desenvolvido atualmente permite aos usuários:

- Selecionar o ano a ser analisado (2017, 2016, 2015 e 2014);
- Selecionar o tipo de nascimento com ou sem malformação congênita;
- *Dashboard* de Dados Gerais;
- De valores absolutos de nascimentos Sim, Não e Ignorado;
- *Rank* por cidade;
- Selecionar o número de cidades que é mostrado;
- Classificação por ordem crescente ou decrescente;
- *Dashboard* das análises do recém-nascido;
- Selecionar o ano a ser analisado (2017,2016,2015,2014);
- Selecionar o tipo de nascimento a ser analisado;
- Selecionar o estado no mapa a ser analisado;
- Sexo;
- Raça;
- Idade do pai;
- Peso;
- Tipo de parto;
- Local de nascimento;
- Apgar1<sup>7</sup>;

---

<sup>6</sup>Disponível em

<[https://public.tableau.com/profile/tatiana.almeida#!/vizhome/Workbook\\_v2\\_15972543914630/DadosGerais](https://public.tableau.com/profile/tatiana.almeida#!/vizhome/Workbook_v2_15972543914630/DadosGerais)>. Acesso em: 25 mar. 2021.

<sup>7</sup> O Apgar é um método para resumir rapidamente a saúde do recém-nascido, o Apgar1 se refere à saúde no primeiro minuto e o Apgar5 no quinto minuto

- Apgar5;
- Tipo de anomalia;
- *Dashboard* das análises da mãe;
- Idade da mãe (média, variação anual e *rank* média por cidade);
- Raça;
- Escolaridade;
- Estado civil;
- Quantidade de gestação anteriores (média, variação anual e *rank*: média por cidade);
- Ocupação da mãe;
- Idade da mãe vs Idade do pai;
- Quantidade de filhos vivos (média, variação anual e *rank* média por cidade);
- Quantidade de filhos mortos (média, variação anual, *rank*: média por cidade);
- *Dashboard* dados da gestação;
- Tempo de gestação;
- Tipo de gravidez;
- Tipo de parto;
- Local de nascimento;
- Consulta pré-natais.

Os *dashboards* possuem 100 medidas e colunas calculadas que facilitam a criação de gráficos. Cada medida possui uma lógica e responde de uma maneira segundo o filtro selecionado. No Apêndice B, a lógica implementada em cada medida pode ser visualizada. Os *dashboards* são compostos por um conjunto que contém 81 elementos gráficos que representam os cálculos criados. Tanto os gráficos, quanto as tabelas são sensíveis aos filtros de página.

#### 4.3 PROTOTIPAÇÃO E IMPLEMENTAÇÃO DO SISTEMA

Como resultados preliminares (primeiro ciclo de prototipação), apresentam-se alguns exemplos da visualização de dados que podem ser obtidos, começando com os mais genéricos

e, depois, filtrando para dados mais detalhados. O conjunto de todas as bases de dados representa o total de 65.873.856 nascimentos; desses, 353.898 crianças nasceram com algum tipo de anomalia. A evolução dos nascimentos com anomalia pode ser observada na Figura 34, a seguir, bem como a evolução dos nascidos sem anomalias.

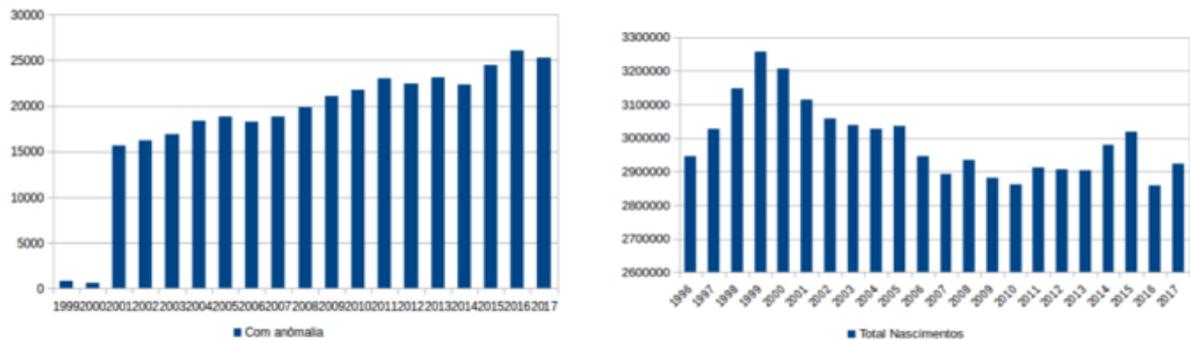


Figura 34 – Na esquerda, a evolução dos nascidos com anomalias; na direita, os nascidos sem anomalias (Elaborado pela autora, 2020).

Percebe-se que, colocando os dados lado a lado, torna-se difícil ter uma ideia da quantidade de nascidos sem anomalia com os nascidos com, devido à proximidade dos gráficos e grande diferença de escala. Como alternativa, pode-se juntar os dois gráficos de forma a facilitar a visualização, conforme exibido na Figura 35. Já, na Figura 36, as bases de dados estão separadas por ano, disponíveis do ano de 1996 até 2014.

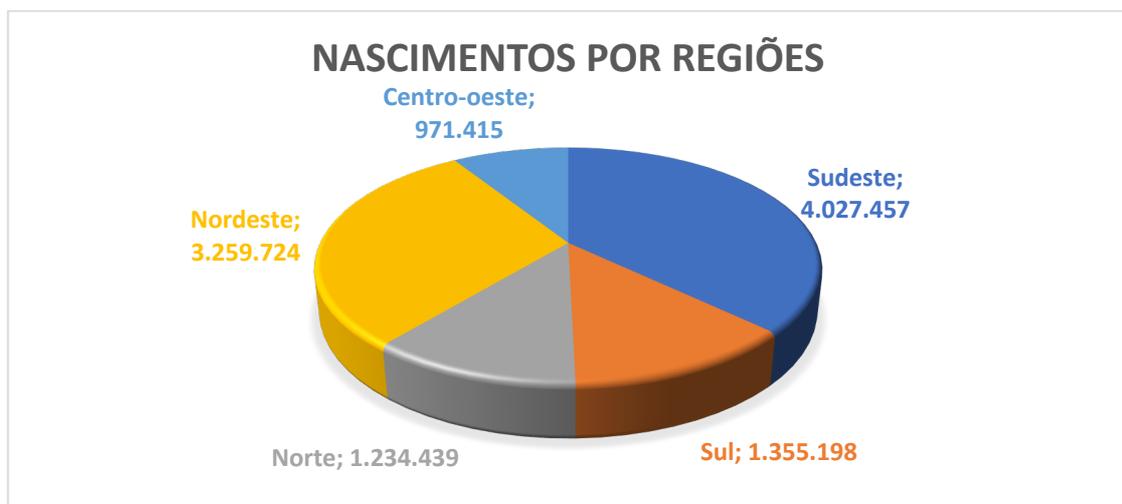


Figura 35 – Total de nascimentos por região (Elaborado pela autora, 2020).

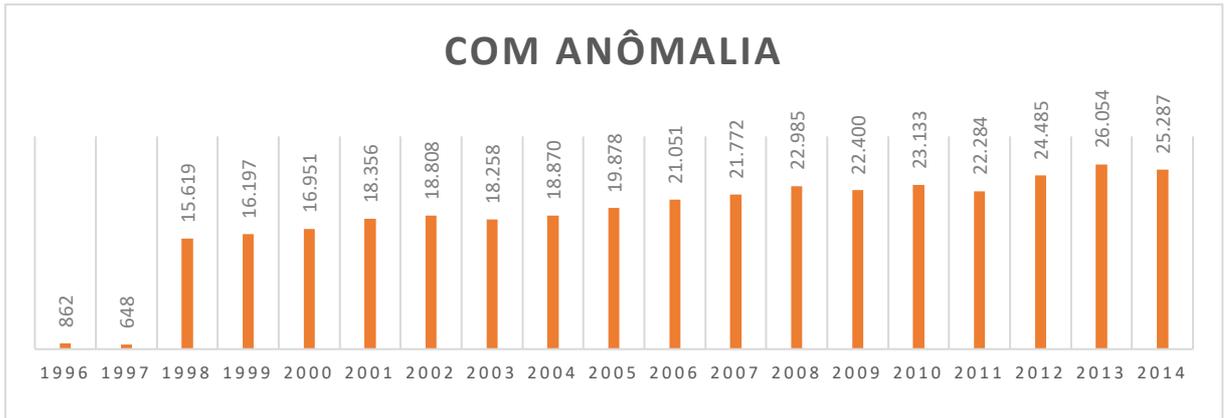


Figura 36 – Acumulado de nascidos com anomalia por ano (Elaborado pela autora, 2020).

Na Figura 37, apresentamos o acumulado por estado de nascidos com anomalias para todo o período. Percebe-se que a maior parte dos casos estão situados em São Paulo. Isto também revela a importância de colocar números relativos à quantidade de nascimento e tamanho da população de cada estado. Tais ponderações foram consideradas na elaboração posterior dos *dashboards* apresentados nos próximos parágrafos.

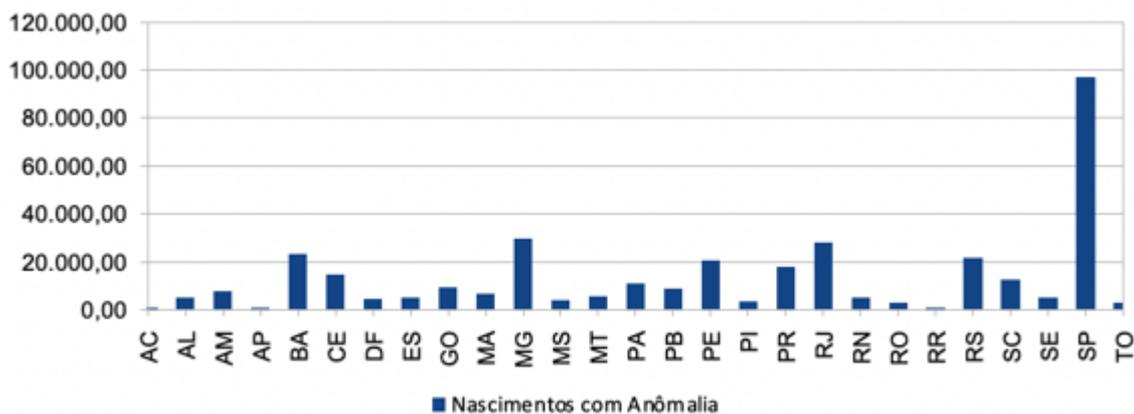


Figura 37 – Acumulado de nascidos com anomalia por estado (Elaborado pela autora, 2020).

O *dashboard* de dados gerais apresentado na Figura 38, adiante, é composto de *cards* verticais com base em princípios de usabilidade. Tais *cards* colocam lado a lado os valores absolutos de nascimentos, sensíveis ao filtro de ano, para cada tipo de nascimento. O objetivo é tornar o sistema simples e intuitivo de se utilizar, atendendo ao princípio que ele deve ser de fácil utilização para qualquer usuário, visto a abrangência de usuários. Assim, pode-se visualizar no item 1 da figura os anos a serem selecionados e, com base nisso, outras

informações são atualizadas e podem ser filtradas, tal como o estado, que é apresentado em forma de mapa.

Assim, as principais informações são apresentadas no topo, tal como o número total de nascimentos, variação anual para nascidos vivos e natimortos. Os dados são apresentados tanto em forma de gráfico, quanto em forma numérica, facilitando sua visualização por quantidade em cada cidade. Na Figura 38, também consta o total nascimentos por valor absoluto de nascimentos, por tipo de nascimento (sensível ao ano), Sim (com malformação), Não (sem malformação) e ignorado. Também inclui a variação anual, comparação do total de nascimentos do ano selecionado no filtro (rosa) comparado ao ano anterior (cinza) aberto no mês. O *rank* por cidade mostra o total de nascimentos das cidades (sensível ao filtro de ano).

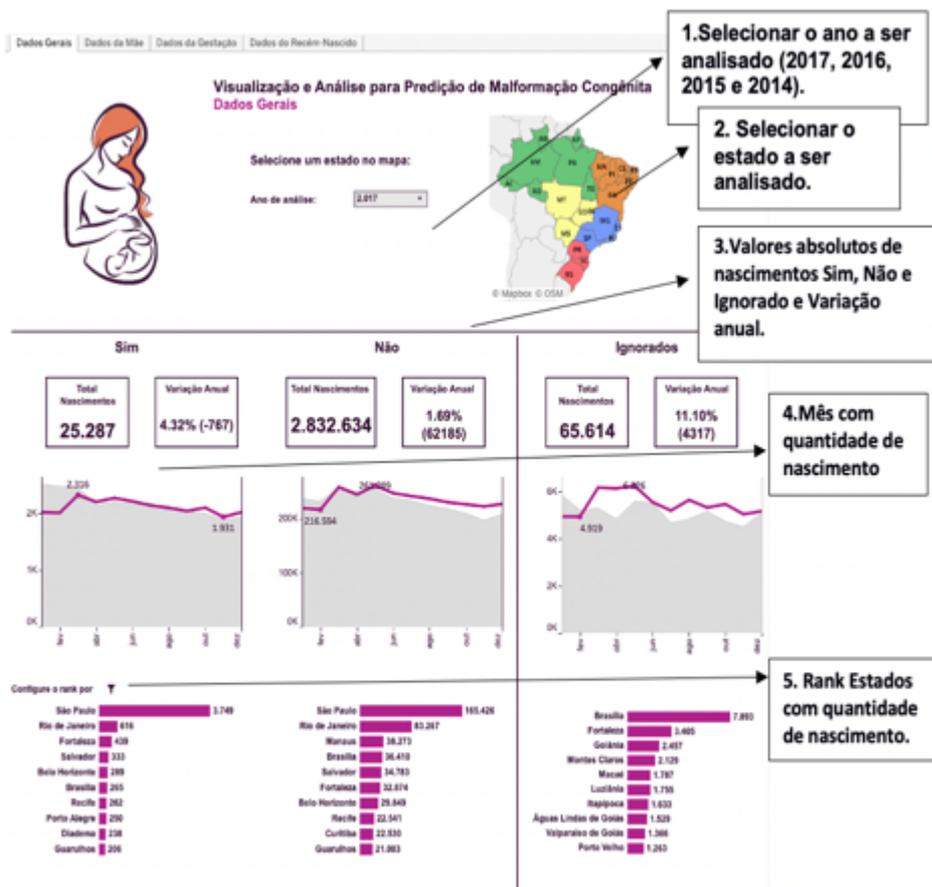
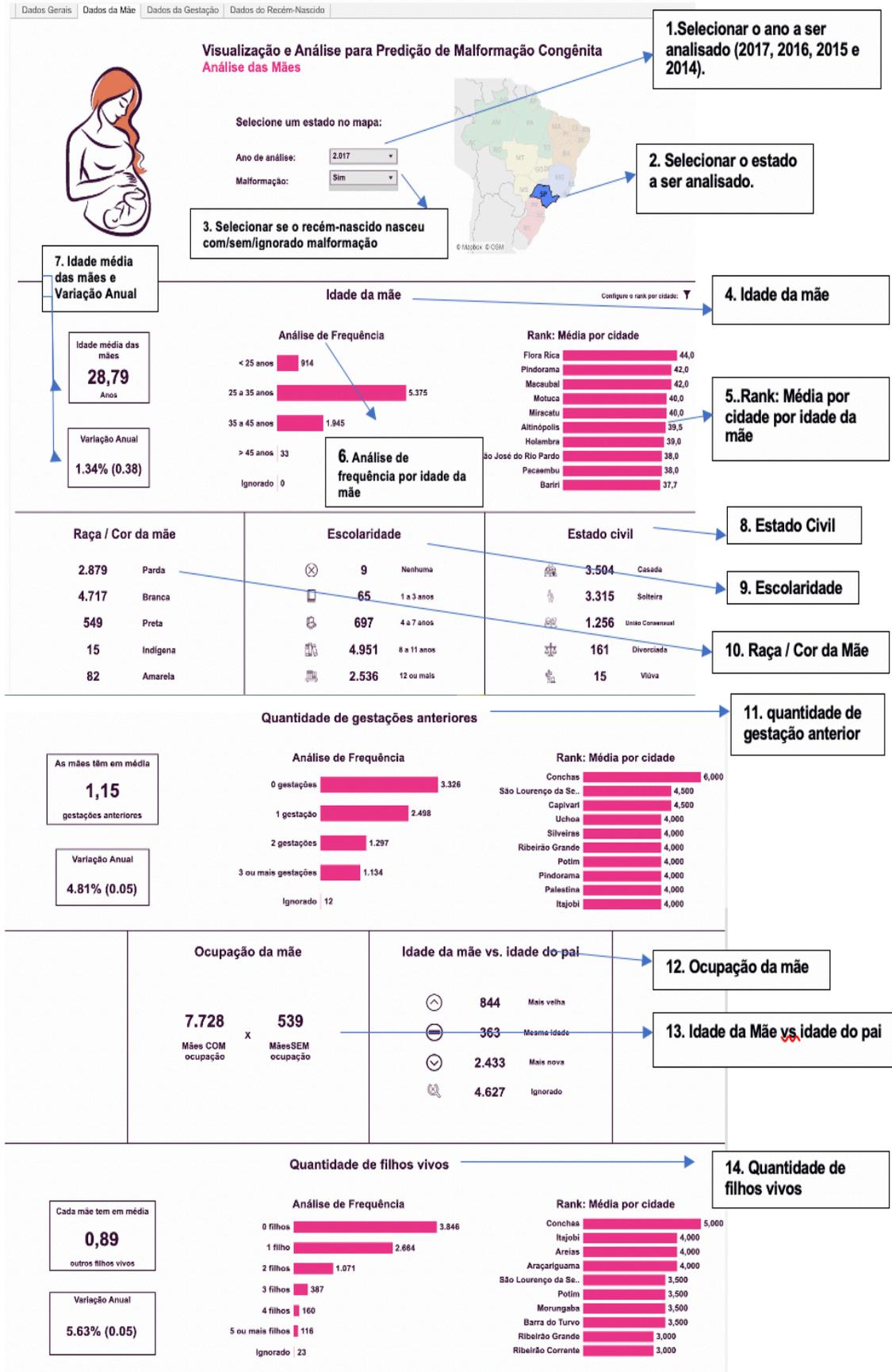


Figura 38 – Dashboard com dados gerais (Elaborado pela autora, 2020).

A análise dos dados da mãe é apresentada na Figura 39, em que se pode filtrar o ano, se a mãe teve criança com malformação e filtrar os dados por estados. O resultado do filtro do ano 2017, em São Paulo, e com recém-nascido com malformação é apresentado na Figura 39.



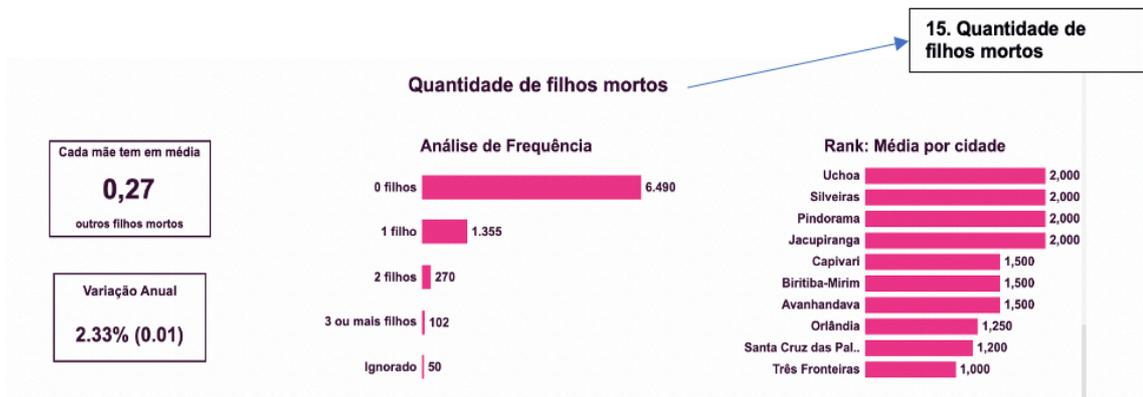


Figura 39 – *Dashboard* com análise de dados da mãe (Elaborado pela autora, 2020).

Um protótipo de *dashboard* com o foco nas análises das gestações é apresentado na Figura 40. Neste *dashboard* pode-se analisar os perfis das gestantes do Brasil, selecionando por estado, por ano (2014, 2015, 2016 e 2016), que teve recém-nascido com malformação ou não, tempo de gestação que iniciou o pré-natal, tipo de parto, local de nascimento, momento da cesária, trabalho de parto induzido, tipo de apresentação do RN, e as médias por cidade.

Já na Figura 41 apresenta *dashboard* para análise dos recém-nascidos. Este *dashboard* também inclui a tabela de anomalias. A tabela tem por objetivo contabilizar o número de nascimentos com malformação congênita (sensível aos filtros de ano e tipo de nascimento) por Capítulo, Grupo, Categoria e Subcategoria. A seta ilustra a funcionalidade de ocultar colunas para aumentar a granularidade da informação. Podemos visualizar por estado selecionado e ano de análise. Com essas interfaces, podemos visualizar por estado as crianças recém-nascidas com malformação congênita em valor absoluto, e visualizar o território nacional também em valor absoluto por capítulo, grupo, categoria e subcategoria.

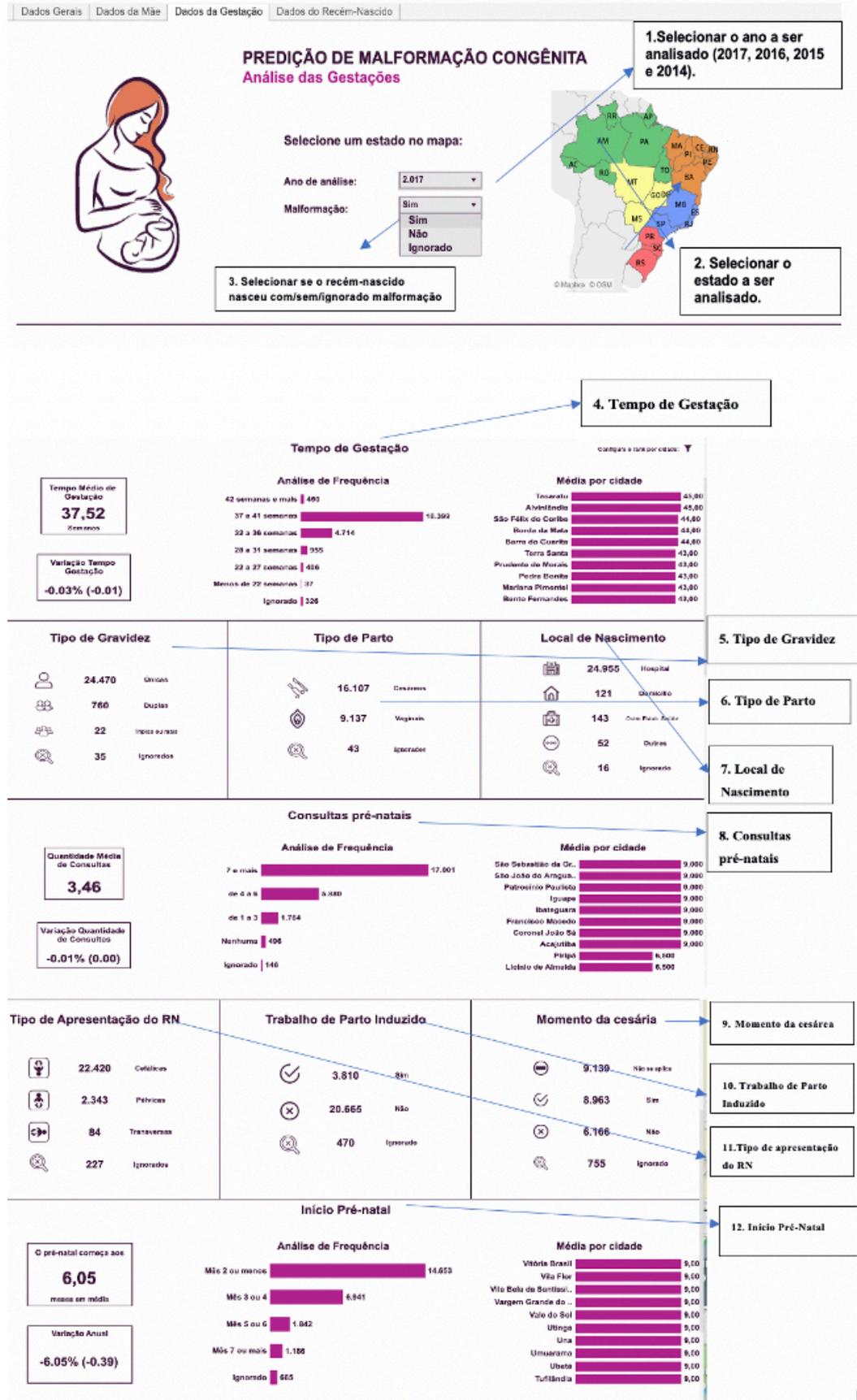
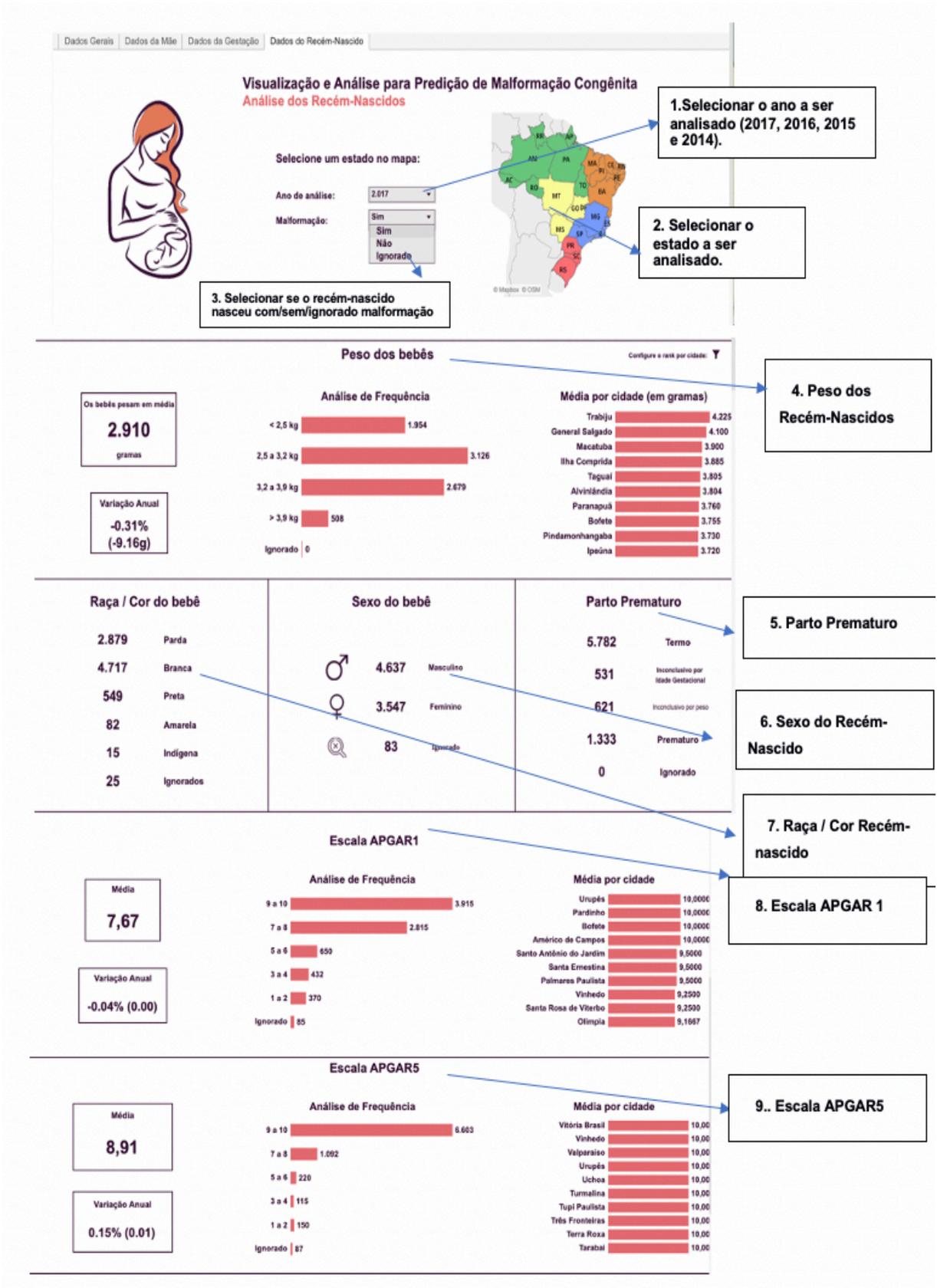


Figura 40 – Dashboard com filtro das análises das gestações.



**Anomalias**

Capítulo	Grupo	Categoria	Subcategoria	Quantidade
II. Neoplasias (tumores)	Neoplasias (tumores) benignas(os)	D18 Hemangioma e linfangioma de qualquer localiz	D18.0 Hemangioma de qualquer localiz	82
XVII. Malformações congênitas e anomalias cromossômicas	Anomalias cromossômicas NCOP	Q80 Síndr de Down	Q80.0 Trissomia 21 não-disjunção meiótica	44
			Q80.1 Trissomia 21 mosaicismo	14
			Q80.2 Trissomia 21 translocação	27
			Q80.9 Síndr de Down NE	785
			Q91.0 Trissomia 18 não-disjunção meiótica	9
		Q91 Síndr de Edwards e síndr de Patau	Q91.1 Trissomia 18 mosaicismo cromossômico	1
			Q91.2 Trissomia 18 translocação	4
			Q91.3 Síndr de Edwards NE	57
			Q91.4 Trissomia 13 não-disjunção meiótica	1

**10. Anomalias**

Figura 41 – Dashboard com resultados do filtro dos recém-nascidos (Elaborado pela autora, 2020).

#### 4.3.1 Implementação das Visualizações Georreferenciadas

De forma a obter um resultado focado em informação geográfica, utilizando a visualização, edição e análise de dados georreferenciados, usamos o QGIS, *software* livre que permite criar mapas através dados geográficos. Para desenvolver o mapa focado com a localização da criança recém-nascida com malformação congênita, classificamos por categoria as malformações (anomalias) e por município (geocódigo), e usamos os dados dos conjuntos de dados do SINASC do ano de 2017.

A quantidade de casos de crianças recém-nascidas com malformação congênita de 2017 foi ponderada pela população dos municípios, segundo a estimativa do IBGE, e padronizada em quantidade de casos por 100 mil habitantes, sendo conectada ao *shapefile* dos municípios brasileiros através do geocódigo. No lado direito da Figura 42 apresenta-se o resultado geral de todas as crianças recém-nascidas com malformação congênita através do mapa geográfico, comparada com os dados do IDHM (2010). Nela, comparamos o IDHM de cada município com

os casos. Comparamos com a quantidade de casos por 100 mil habitantes em relação à quantidade de crianças recém-nascidas com malformação congênita.

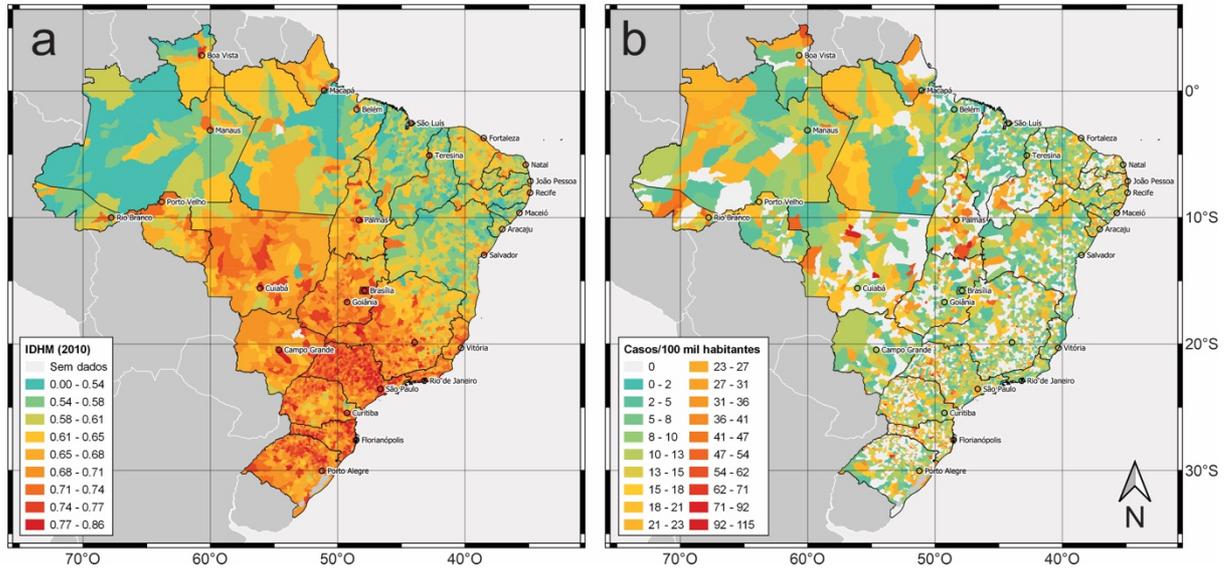


Figura 42 – Mapa de criança recém-nascida com malformação congênita, de 2017, comparada com IDHM (2010) e casos/ 100 mil habitantes (Elaborado pela autora, 2020).

A Figura 43, adiante, apresenta os mapas com resultados de crianças recém-nascida com malformação congênita por casos/100 mil habitantes para diferentes tipos de anomalia. As categorias estão descritas na Tabela CID-10.

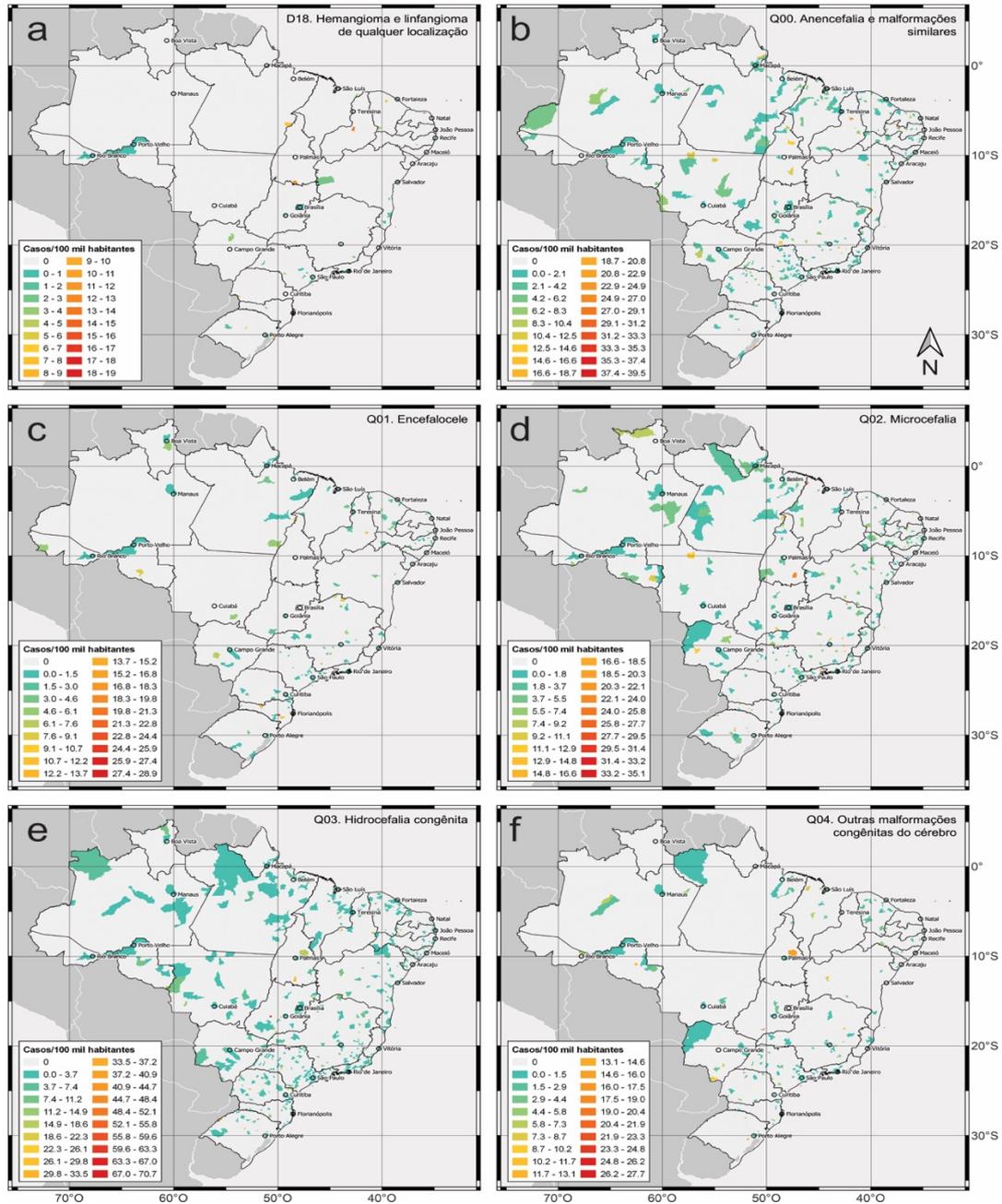


Figura 43 – Mapas com resultados de crianças recém-nascidas com malformação congênita por categoria e casos/ 100 mil habitantes (Elaborado pela autora, 2020).

Os mapas apresentados no Apêndice C mostram diversos tipos de doenças de malformação na infância de modo específico por meio de cores que se encaixam devidamente em cada localidade por estado. As cores são escaladas em diversos tons, começando do azul até chegar ao vermelho, sendo a primeira cor, uma quantidade menor de crianças afetadas por 100 mil habitantes e, a última cor, o maior número de casos no país por 100 mil habitantes.

As regiões com maiores concentrações de crianças com malformação estão no Norte, Nordeste e Sudeste. No Norte, o IDH é menor do que o do país, o que indica que a região possui uma menor qualidade em relação à renda, educação e saúde comparados a maior parte do país. Isso faz com que a população possua menos acesso a um sistema de saúde de qualidade, receba uma educação de qualidade e, conseqüentemente, melhores posições de trabalho. Na maioria dos mapas, a região Norte apresenta grandes proporções coloridas, na maioria, em tons verdes, como hidrocefalia congênita (7,4 e 11,2 crianças por 100 mil habitantes) e fenda palatina 4,9 a 9,9 crianças por 100 mil habitantes).

A região Nordeste é a com o menor IDH, sofre de diversos problemas, como a menor qualidade de vida do Brasil, seca, alta mortalidade infantil (a mais alta do país) e aproximadamente 55% da população não possui saneamento ambiental. Ademais, a região sofre de faltas de políticas públicas eficazes para proporcionar condições adequadas de moradia, saúde, educação, emprego, entre outros fatores, o que facilita para que a população possua altos índices de pobreza e não recebam um tratamento adequado para doenças (Brasil Escola, 2020a). Com falta de pré-natal e bons acompanhamentos na gravidez, a taxa de mortalidade infantil acaba por ser muito alta e, com isso, uma difícil detecção de alguma malformação no feto. Logo, podemos notar que Fortaleza e Recife se destacam em regiões com mais casos de criança com malformação na região, com cor verde, amarela e laranja.

Na Região Sudeste possui IDH menor do que a região Sul, porém, possui alto grau de desenvolvimento tecnológico e industrial, bem como uma rede desenvolvida de atendimento médico-hospitalar (Brasil Escola, 2020b). Provavelmente, devido ao fato de ser a região mais povoada do país e uma região marcada pela desigualdade social, a relação de crianças com malformação também se destaca. Muitas pessoas não possuem acesso a um sistema de saúde particular e não recebem um saneamento adequado em suas residências. As cores que surgem nessa região são, principalmente, a verde e a laranja, por exemplo, a fenda palatina e a fenda labial.

Dessa maneira, é visível como a quantidade de crianças com malformação é maior em regiões em que o IDH possui menores valores, o que leva a entender que a qualidade de vida e um melhor acesso a atendimentos hospitalares favorecem com que a saúde da criança seja mais bem tratada e remediada. Portanto, constatações como essa pode servir para estabelecer políticas voltadas às regiões e municípios com menor IDH.

### 4.3.2 Análise Estatística dos dados do SINASC

Nesta subseção são apresentados os resultados estatísticos referentes a informações do SINASC, dos anos de 2014, 2015, 2016 e 2017. Tais resultados são utilizados na elaboração de *dashboards* apresentados na subseção 4.3.3. Estruturam-se em cálculos estatístico para compreender e buscar resultados significativos na visualização de dados. Estas análises foram calculadas sobre a orientação da Professora Daniele Kopp, da Especialização em Estatística e Modelagem Quantitativa da Faculdade Cruzeiro do Sul, e serão posteriormente convertidas em visualizações para apoiar a interpretação de informações sobre malformações congênitas.

O processo de análise de dados, inicialmente, ocorreu com a obtenção dos bancos de dados do DATASUS, referentes aos nascidos vivos no Brasil por ano. Então, foram selecionadas, de acordo com as normas técnicas, as variáveis adequadas para responder aos objetivos deste estudo. Os bancos de dados passaram pelo processo de limpeza de dados faltantes, categorização de variáveis e transformações de variáveis, caso fosse necessário. Para a análise e a limpeza dos bancos, utilizou-se o *software* STATA versão 12.1<sup>8</sup>. A análise descritiva consistiu em apontar os valores de médias, medianas, desvios padrões, frequências e percentagens. A análise inferencial contou com a aplicação do teste de normalidade de Shapiro Wilk, teste de associação de Qui-quadrado ou Exato de Fisher, de acordo com a quantidade de categorias de cada variável associadas. Também se aplicou o teste *t* para comparação de médias ou alternativo não paramétrico, teste de wilcoxon para medidas independentes, conforme a distribuição de normalidade dos dados.

Aplicou-se uma série de análise descritivas por meio de tabelas e gráficos para que fosse melhorada a capacidade de interpretação dos leitores. Adotou-se ainda o nível de significância de 5% e todos os critérios de pressupostos dos testes foram respeitados. Os cálculos encontram-se na Apêndice D. Os resultados dos cálculos estão em formato de tabelas, apresentadas a seguir, com fácil compreensão e estudo posterior em visualização de dados para aplicação.

A Tabela 10 descreve as informações para o banco completo, com casos com e sem anomalias.

---

<sup>8</sup> <https://www.stata.com/>

**Descrição das informações quantitativas das mães, gestação e recém-nascido.**

	<b>Dados das mães</b>			
	Média	Mediana	Desvio padrão	p-valor
<b>Idade</b>	26,31	26	6,69	<0,001*
<b>Nº de filhos vivos</b>	1,02	1	1,36	<0,001*
<b>Nº de filhos mortos</b>	0,23	0	0,71	<0,001*
	<b>Dados da gestação e recém-nascido</b>			
	Média	Mediana	Desvio padrão	p-valor
<b>Semanas de gestação</b>	38,49	39	2,24	<0,001*
<b>Apgar 1º minuto</b>	8,40	9	2,75	<0,001*
<b>Apgar 5º minuto</b>	9,39	9	2,34	<0,001*
<b>Peso</b>	3187,11	3220	558,28	<0,001*

p-valor referente ao teste de normalidade de shapiro-wilk

\*indica que há significância estatística, ou seja, as variáveis não se adequam a distribuição normal.

Tabela 10 – Descrição das informações quantitativas por idade da mãe, número de filhos vivos, mortos, semana de gestação, apgar 1, apgar 5 e peso do recém-nascido (Elaborado pela autora, 2020).

A média de idade das mães foi de 26,31 anos; quanto à quantidade de filhos, a média geral destes quatro anos que compõem a amostra foi de 1,02 filhos vivos e 0,23 filhos mortos. A média de tempo de gestação foi de 38,5 semanas aproximadamente, o que já configura um tempo adequado para a formação dos recém-nascidos. As medidas de apgar, para um e cinco minutos, apresentaram médias de 8,40 e 9,39, respectivamente. Quanto ao peso ao nascer, a média foi de 3187,11 gramas, também um valor considerado adequado. Quanto à distribuição das variáveis, nenhuma se adequou à normalidade, ou seja, as análises inferenciais consideraram essa informação para a escolha dos testes estatísticos.

A Tabela 11 adiante, apresenta as descrições das informações quantitativas estratificadas por região do país. Apesar de identificarmos valores com pequenas magnitudes de diferença entre as medidas, quando é realizada a estratificação por região do país, o teste

estatístico encontrou um p-valor estatisticamente significativo para a comparação das variáveis idade das mães, quantidade de filhos vivos e mortos, quando essas são comparadas entre as cinco regiões do país. A Tabela 11 apresenta informações quantitativas das mães por regiões do país e a Tabela 12, na sequência, a comparação das informações do recém-nascido e das gestações por regiões do país.

**Comparação dos dados quantitativos das mães por região.**

	<b>Dados das mães</b>			
	Média	Mediana	Desvio padrão	p-valor
<b>Sudeste</b>				
Idade	27,13	27	6,68	<0,0001*
Nº de filhos vivos	0,89	1	1,19	<0,0001*
Nº de filhos mortos	0,22	0	0,63	<0,0001*
<b>Sul</b>				
Idade	27,02	27	6,63	
Nº de filhos vivos	0,87	1	1,14	
Nº de filhos mortos	0,20	0	0,53	
<b>norte</b>				
Idade	24,66	24	6,50	
Nº de filhos vivos	1,44	1	1,68	
Nº de filhos mortos	0,26	0	0,61	
<b>Nordeste</b>				
Idade	25,63	25	6,65	
Nº de filhos vivos	1,11	1	1,51	
Nº de filhos mortos	0,26	0	0,87	
<b>Centro-Oeste</b>				
Idade	26,27	26	6,53	
Nº de filhos vivos	1,03	1	1,33	
Nº de filhos mortos	0,22	0	0,83	

p-valor referente ao teste de comparação múltipla de Kruskal-wallis, alternativo ao teste ANOVA.

\*Indica que há significância estatística (p-valor<0,05), pelo menos um par de grupos difere entre si.

Tabela 11 – Comparação das informações quantitativas das mães por idade, número de filhos vivos e mortos por região do Brasil (Elaborado pela autora, 2020).

**Comparação dos dados quantitativos da gestação e dos recém-nascido por região.**

	Dados do Recém-Nascido e Gestação			
	Média	Mediana	Desvio padrão	p-valor
<b>Sudeste</b>				
Semanas de gestação	38,4	39	2,18	<0,0001*
Apgar 1º minuto	8,52	9	3,73	<0,0001*
Apgar 5º minuto	9,40	9	3,13	<0,0001*
Peso	3157,58	3195	556,54	<0,0001*
<b>Sul</b>				
Semanas de gestação	38,40	39	2,10	
Apgar 1º minuto	8,41	9	1,88	
Apgar 5º minuto	9,42	10	1,59	
Peso	3183,47	3220	553,49	
<b>Norte</b>				
Semanas de gestação	38,61	39	2,34	
Apgar 1º minuto	8,29	9	1,25	
Apgar 5º minuto	9,32	9	1,05	
Peso	3214,73	3240	549,52	
<b>Nordeste</b>				
Semanas de gestação	38,62	39	2,36	
Apgar 1º minuto	8,27	9	2,09	
Apgar 5º minuto	9,29	9	1,86	
Peso	3218,93	3250	567,84	
<b>Centro-Oeste</b>				
Semanas de gestação	38,44	39	2,13	
Apgar 1º minuto	8,43	9	2,10	
Apgar 5º minuto	9,45	10	1,93	
Peso	3172,67	3200	542,44	

p-valor referente ao teste de comparação múltipla de Kruskal-wallis, alternativo ao teste ANOVA.

\*Indica que há significância estatística (p-valor<0,05), pelo menos um par de grupos difere entre si.

Tabela 12 – Comparação das informações quantitativas dos dados de recém-nascidos e gestação por região do Brasil (Elaborado pela autora, 2020).

Todas as variáveis das Tabelas 11 e 12 apresentaram diferenças estatisticamente significativas quando comparadas por região do país onde nasceu o recém-nascido, mesmo que as magnitudes das diferenças não sejam grandes. Destacamos que a região Sudeste foi onde os recém-nascidos apresentaram a menor média de peso (3157,58 gramas), e a Nordeste, a maior média (3218,93 gramas). Todas as médias de peso encontram-se dentro de valores saudáveis.

**Comparação do perfil quantitativo das mães, gestação e recém-nascido pela com e sem malformação.**

	Dados das mães				
	Com malformação		Sem malformação		p-valor
	Média	DP	Média	DP	
<b>Idade</b>	27	7,22	26,31	6,68	<0,001*
<b>Nº de filhos vivos</b>	0,99	1,50	1,02	1,35	<0,001*
<b>Nº de filhos mortos</b>	0,28	0,92	0,28	1,40	<0,001*
	Dados da gestação e recém-nascido				
	Com malformação		Sem malformação		p-valor
	Média	DP	Média	DP	
<b>Semanas de gestação</b>	37,51	3,10	38,50	2,22	<0,001*
<b>Apgar 1º minuto</b>	7,46	4,02	8,41	2,74	<0,001*
<b>Apgar 5º minuto</b>	8,63	3,77	9,40	2,33	<0,001*
<b>Peso</b>	2876,00	762,67	3189,88	554,79	<0,001*

p-valor referente ao teste de comparação de grupos independentes não paramétrico de wilcoxon, \*indica que há diferença estatisticamente significativa(p-valor<0,05).

Fonte: autora.

Tabela 13 – Comparação das informações quantitativas das mães, gestação e recém-nascido em relação à presença ou não de malformação (Elaborado pela autora, 2020).

A comparação das variáveis quantitativas que descrevem as características das mães na Tabela 13, apesar de todas apresentarem diferença estatisticamente significativa, a diferença numérica foi pequena entre os grupos comparados. Já nos dados da gestação e do recém-nascido, apresentaram diferenças maiores quando comparadas pela presença ou não de malformação/ anomalia. Há valores bastante diferentes entre si, como a quantidade de semanas de gestação, em que os recém-nascido com malformação têm cerca de uma semana média a menos que os demais. Os valores de apgar também, em média, são 1 ponto mais altos nos recém-nascido sem malformação e, no peso, há uma diferença de mais de 300 gramas entre as médias dos recém-nascido.

A Tabela 14 apresenta uma comparação do perfil quantitativo das mães por região.

**Comparação do perfil quantitativo das mães com recém-nascido com ou sem malformação, estratificado por região do país.**

	Dados das mães				p-valor
	Com malformação		Sem malformação		
	Média	DP	Média	DP	
<b>Sudeste</b>					
Idade	27,93	7,15	27,14	6,68	<0,001*
Nº de filhos vivos	0,88	1,27	0,89	1,18	<0,001*
Nº de filhos mortos	0,27	0,79	0,22	0,62	<0,001*
<b>Sul</b>					
Idade	27,64	7,28	27,02	6,63	<0,001*
Nº de filhos vivos	0,91	1,21	0,87	1,13	0,3025
Nº de filhos mortos	0,23	0,56	0,20	0,53	<0,001*
<b>Norte</b>					
Idade	25,01	7,06	24,65	6,49	0,0260*
Nº de filhos vivos	1,48	1,79	1,44	1,68	0,5800
Nº de filhos mortos	0,31	0,64	0,26	0,60	<0,001*
<b>Nordeste</b>					
Idade	25,98	7,14	25,62	6,64	<0,001*
Nº de filhos vivos	1,05	1,88	1,11	1,51	<0,001*
Nº de filhos mortos	0,31	1,31	0,26	0,86	<0,001*
<b>Centro-Oeste</b>					
Idade	26,41	6,99	26,26	6,52	0,6866
Nº de filhos vivos	1,08	1,27	1,01	1,21	0,0042*
Nº de filhos mortos	0,26	0,60	0,23	0,64	<0,001*

p-valor referente ao teste de comparação de grupos independentes não paramétrico de wilcoxon, \*indica que há diferença estatisticamente significativa(p-valor<0,05).

Tabela 14 – Comparação das informações quantitativas pela presença de malformação por região (Elaborado pela autora, 2020).

Quando comparadas as informações de idade das mães, quantidade de filhos vivos e mortos, pela presença ou não de malformação, encontrou-se diferença estatisticamente significativa para a idade das mães dentro de todas as regiões do país, exceto para o Centro-Oeste. Já para a quantidade de filhos vivos, não se identificou diferença estatisticamente significativa nas regiões Sul e Norte. As demais apresentaram valores com significância estatística. Quanto à distribuição da quantidade de filhos mortos, encontrou-se diferença estatisticamente significativa na comparação de todas as regiões.

A Tabela 15 apresenta as comparações das variáveis referentes a gestação e aos recém-nascidos por malformação ou não, estratificados também pelas regiões do país.

**Comparação do perfil quantitativo da gestação e dos recém-nascidos com ou sem malformação, estratificado por região do país.**

	Dados da Gestação e dos bebês				
	Com malformação		Sem malformação		p-valor
	Média	DP	Média	DP	
<b>Sudeste</b>					
Semanas de gestação	37,50	2,99	38,41	2,17	<0,001*
Apgar 1° minuto	7,69	4,89	8,53	3,73	<0,001*
Apgar 5° minuto	8,55	4,48	9,47	3,12	<0,001*
Peso	2883,67	750,45	3160,20	553,37	<0,001*
<b>Sul</b>					
Semanas de gestação	37,31	3,00	38,41	2,08	<0,001*
Apgar 1° minuto	7,18	2,70	8,43	1,85	<0,001*
Apgar 5° minuto	8,43	2,41	9,43	1,56	<0,001*
Peso	2852,24	759,23	3186,69	549,93	<0,001*
<b>Norte</b>					
Semanas de gestação	37,64	3,25	38,62	2,33	<0,001*
Apgar 1° minuto	7,22	2,28	8,30	1,24	<0,001*
Apgar 5° minuto	8,41	2,40	9,33	1,03	<0,001*
Peso	2890,69	770,09	3216,56	547,17	<0,001*
<b>Nordeste</b>					
Semanas de gestação	37,62	3,23	38,63	2,34	<0,001*
Apgar 1° minuto	7,32	3,59	8,28	2,07	<0,001*
Apgar 5° minuto	8,49	3,52	9,30	1,84	<0,001*
Peso	2879,21	776,65	3221,85	564,11	<0,001*
<b>Centro-Oeste</b>					
Semanas de gestação	37,39	3,20	38,44	2,11	<0,001*
Apgar 1° minuto	7,24	2,61	8,45	2,13	<0,001*
Apgar 5° minuto	8,45	2,73	9,47	1,97	<0,001*
Peso	2840	775,39	3176,79	538,05	<0,001*

p-valor referente ao teste de comparação de grupos independentes não paramétrico de wilcoxon, \*indica que há diferença estatisticamente significativa(p-valor<0,05).

Tabela 15 – Comparação das informações quantitativas pela presença de malformação por região (Elaborado pela autora, 2020).

Quando se realizou a comparação entre a quantidade de semanas de gestação entre os recém-nascidos com e sem malformação, encontrou-se uma diferença estatisticamente significativas dentro de todas as regiões. A quantidade de semanas foi maior entre os recém-nascidos sem malformação/anomalia. Quando se aplicou o mesmo modelo para os dados de apgar do primeiro minuto, apgar do quinto minuto e peso ao nascer, também se encontrou o mesmo padrão de resultado, em que houve diferença estatisticamente significativa dentro de todas as regiões para essas variáveis.

Dando sequência às análises dos dados quantitativos, foram calculadas regressões para as variáveis de apgar, peso e semanas de gestação em relação a idade da mãe. A Tabela 16 apresenta a análise de correlação entre a idade da mãe e dados quantitativos da gestação e do recém-nascido.

**Análise de correlação entre a idade da mãe e quantitativos da gestação e do recém-nascido.**

		<b>Idade da Mãe</b>
<b>Apgar 1</b>	Coef.	0,04
	p-valor	<0,001*
<b>Apgar 5</b>	Coef.	0,02
	p-valor	<0,001*
<b>Semanas de gestação</b>	Coef.	-0,09
	p-valor	<0,001*
<b>Peso</b>	Coef.	0,04
	p-valor	<0,001*

Coef. Valor do coeficiente de correlação de spearman ( $\rho$ ),

\*apresenta significância estatística (p-valor<0,05).

Tabela 16 – Análise de correlação entre a cidade da mãe e quantitativos da gestação e do recém-nascido (Elaborado pela autora, 2020).

A análise de correlação busca identificar se há associação entre duas variáveis, ou seja, se, a partir do comportamento de uma, podemos esperar determinado comportamento para a outra. Apesar de todas as correlações testadas terem apresentado significância estatística, p-valor<0,05, a força de correlação apresentada pelos coeficientes é muito baixa, não havendo magnitude na influência de uma variável sobre a outra suficientemente forte a ponto de ser relevante. Buscando ir mais a fundo nessas informações, aplicou-se a análise de correlação estratificada pela presença ou não de malformação/anomalia.

Quando as correlações foram realizadas estratificadas por dados de crianças com malformação/anomalia e sem, entre aquelas com malformação/anomalia, não se encontrou correlação significativa entre a idade da mãe e os dois valores de apgar dos recém-nascidos (Tabela 17). Para a quantidade de semanas de gestação e o peso ao nascer, encontrou-se correlação, inversamente proporcional, significativa, mas com muito baixa força de correlação.

**Análise de correlação entre a idade da mãe, quantitativos da gestação e do recém-nascidos.**

		<b>Idade da Mãe</b>
<b>Apgar 1</b>	Coef.	0,04
	p-valor	<0,001*
<b>Apgar 5</b>	Coef.	0,02
	p-valor	<0,001*
<b>Semanas de gestação</b>	Coef.	-0,09
	p-valor	<0,001*
<b>Peso</b>	Coef.	0,04
	p-valor	<0,001*

Coef. Valor do coeficiente de correlação de spearman (rho), \*apresenta significância estatística (p-valor<0,05).

Tabela 17 – Análise de correlação entre a idade da mãe e quantitativos da gestação e do recém-nascido (Elaborado pela autora, 2020).

Entre aqueles sem malformação/anomalia, todas as correlações foram estatisticamente significativas, mas nenhuma com força de correlação significante, todos são classificadas como muito fracas.

Na sequência serão apresentados resultados das análises das variáveis categóricas do banco de dados. De acordo com os critérios estatísticos, as variáveis categóricas foram descritas com a apresentação de valores de frequências e porcentagens e aplicou-se o teste de associação de qui-quadrado para identificar a relação entre variáveis. A Tabela 18 apresenta a distribuição dos nascimentos no período considerado pelo estudo.

**Distribuição dos nascimentos por região**

	Frequência	%
<b>Região</b>		
Sudeste	4.027.457	37,13
Sul	1.355.198	12,49
Norte	1.234.439	11,38
Nordeste	3.259.724	30,05
Centro-oeste	971.415	8,95

Tabela 18 – Distribuição dos nascimentos por região (Elaborado pela autora, 2020).

Para melhor visualização da distribuição dos dados, a Figura 44 ilustra a distribuição da quantidade de nascidos vivos, de 2014 a 2017, estratificados por regiões do país. As regiões que concentraram maiores quantidades de nascimentos foram a Sudeste e a Nordeste.

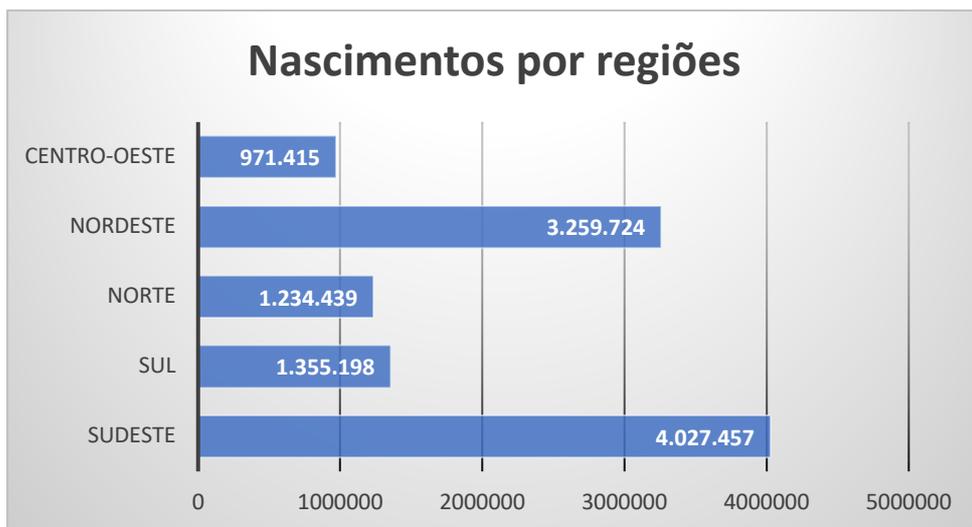


Figura 44 – Distribuição dos nascimentos X Nascimentos com anomalia por região  
(Elaborado pela autora, 2020).

De acordo com a Tabela 19, a maior parte das mães se declararam solteiras (42,41%), seguidas por aquelas que se declararam em uma união consensual (23,65%). Quanto à escolaridade, destaca-se que a grande maioria das mães têm escolaridade entre oito e 11 anos (59,37%), intervalo referente à conclusão do Ensino Fundamental e Médio; em segundo lugar, aquelas que têm entre quatro e sete anos de estudo (18,71%).

**Descrição dos dados categóricos sobre as mães.**

	<b>Frequência</b>	<b>%</b>
<b>Estado civil da mãe</b>		
Solteira	4.546.934	42,41
Casada	3.503.025	32,67
Viúva	19.002	0,18
Separada	116.642	1,09
União consensual	2.535.759	23,65
<b>Escolaridade da mãe</b>		
Nenhuma	61.603	0,58
1 à 3 anos	296.909	2,78
4 à 7 anos	1.997.101	18,71
8 à 11 anos	6.338.141	59,37
12 ou mais	1.982.412	18,57

Tabela 19 – Descrição dos dados categóricos sobre as mães por estado civil e escolaridade  
(Elaborado pela autora, 2020).

A Tabela 20, a seguir, traz os dados sobre a gravidez. Conforme apresenta a Tabela 21, quanto ao nascimento, a maioria dos recém-nascidos nasceram em hospitais (98,41%), com mais de 37 semanas de gestação (88,70%), de gestação única (97,95%); mais da metade dos partos foi cesárea (55,72%) e, para 66,77% dos recém-nascido, as mães foram a pelo menos sete consultas do acompanhamento pré-natal. Destacamos que, mesmo com todos os avanços de tecnologia, conhecimento e investimentos em saúde, 2,28% das mães de recém-nascidos nascidos vivos não participaram de nenhuma consulta pré-natal.

**Descrição dos dados categóricos sobre a gestação.**

	<b>Frequência</b>	<b>%</b>
<b>Gravidez</b>		
Única	10.607.732	97,95
dupla	217.040	2,00
Tripla ou mais	5.346	0,005
<b>Consultas pré-natal</b>		
Nenhuma	245.351	2,28
De 1 à 3	697.580	6,47
De 4 à 6	2.639.992	24,49
7 ou mais	7.198.583	66,77
<b>Tempo de gestação</b>		
Menos de 37 semanas	1.195.963	11,30
37 semanas ou mais	9.390.686	88,70
<b>Parto</b>		
Vaginal	4.798.240	44,28
Cesário	6.037.805	55,72
<b>Local de nascimento</b>		
Hospital	10.667.525	98,41
Outro estabelecimento de saúde	77.844	0,72
Domicílio	76.783	0,71
outros	18.097	0,17

Tabela 20 – Descrição das informações quantitativas de dados sobre a gestação (tipo de gravidez, consultas pré-natal, tempo de gestação, parto e local de nascimento) (Elaborado pela autora, 2020).

Conforme apresenta a Tabela 21, a maior parte dos recém-nascidos são do sexo masculino (51,23%), declarados como raça/cor parda (58,68%) e não apresentaram nenhuma anomalia ou malformação identificada ao nascer. Para melhor compreender as características desses recém-nascidos, as tabelas em sequência apresentam os dados das mães, gestação e recém-nascido estratificados por estado, presença ou não de malformação/ anomalia e região onde nasceram.

<b>Variáveis</b>	<b>Frequência</b>	<b>%</b>
<b>Sexo do bebe</b>		
Masculino	5.556.252	51,23
Feminino	5.289.844	48,77
<b>Raça/cor</b>		
Branca	3.593.151	34,63
Preto	561.949	5,42
Amarelo	40.443	0,39
Pardo	6.088.176	58,68
Indígena	91.301	0,88
<b>Identificação de anomalia</b>		
Sim	88.354	0,84
não	10.484.024	99,16

Tabela 21 – Descrição das informações quantitativas de dados sobre os recém-nascidos (sexo do bebê, raça/cor, identificação de anomalia) (Elaborado pela autora, 2020).

A Tabela 22 traz os dados das mães, associados pela região, tanto a escolaridade, quanto o estado civil das mães, demonstrando o comportamento associado à região identificada pela variável região de nascimento da criança. Na sequência, serão apresentados os dados da gestação e parto também associados com a região de nascimento do recém-nascido. Nela, podemos ver que todas as variáveis referentes à gravidez foram associadas de maneira estatisticamente significativa com a região onde nasceu o recém-nascido. Além das informações sobre as mães e a gestação, também foram associadas com a região de nascimento (Tabela 23) as características dos recém-nascidos, como o sexo e a raça/ cor. Essas informações são apresentadas na Tabela 24. Assim como as demais variáveis descritivas, todas foram estatisticamente significativas quando associadas à região de nascimento dos recém-nascidos.

Associação dos dados categóricos das mães com a região de nascimento do Recém-Nascido.

	Região										p-valor
	Sudeste		Sul		Norte		Nordeste		Centro-oeste		
	Freq.	%	Freq.	%	Freq.	%	Freq.	%	Freq.	%	
<b>Estado civil</b>											<0,001*
Solteira	1.795.106	44,9%	550.039	40,8%	492.318	40,4%	1.321.597	41,3%	387.874	40,5%	
Casada	15.916.89	39,8%	469.244	34,8%	226.770	18,6%	879.477	27,5%	335.845	35,1%	
Viúva	7.673	0,2%	2.494	0,2%	1.347	0,1%	5.936	0,2%	1.552	0,2%	
Separada	65.050	1,6%	17.460	1,3%	5.001	0,4%	17.510	0,5%	11.621	1,2%	
União consensual	536.694	13,4%	309.810	23,0%	493.331	40,5%	975.135	30,5%	220.789	23,1%	
<b>Escolaridade</b>											<0,001*
Nenhuma	7.321	0,2%	2.647	0,2%	18.656	1,5%	29.274	0,9%	3.705	0,4%	
1 à 3 anos	58.712	1,5%	21.330	1,6%	58.648	4,9%	140.974	4,4%	17.245	1,8%	
4 à 7 anos	562.807	14,1%	221.016	16,4%	302.189	25,0%	763.548	24,1%	147.541	15,3%	
8 à 11 anos	2.489.084	62,5%	786.303	58,2%	680.689	56,4%	1.824.479	57,5%	557.586	58,0%	
12 ou mais	865.501	21,7%	319.166	23,6%	147.683	12,2%	414.245	13,1%	235.817	24,5%	

p-valor referente a ao teste de associação de qui-quadrado de associação para variáveis categóricas.

\*indica que há associação estatisticamente significativa (p-valor<0,05).

Tabela 22 – Descrição dos dados categóricos das mães (estado civil e escolaridade) por região de nascimento do recém-nascido (Elaborado pela autora, 2020).

Associação dos dados da gestação com a região de nascimento.

	Sudeste		Sul		Norte		Nordeste		Centro oeste		p-valor
	Freq.	%	Freq.	%	Freq.	%	Freq.	%	Freq.	%	
<b>Gravidez</b>											<0,001*
Única	3.930.854	97,7%	1.323.588	97,7%	1.213.663	98,5%	3.189.210	98,2%	950.417	98,0%	
dupla	90.735	2,3%	30.039	2,2%	18.453	1,5%	58.548	1,8%	19.265	2,0%	
Tripla ou mais	2.492	0,1%	714	0,1%	283	0,0%	1.285	0,0%	572	0,1%	
<b>Consultas pré-natal</b>											<0,001*
Nenhuma	52.503	1,3%	14.801	1,1%	57.893	4,7%	94.828	2,9%	25.326	2,6%	
De 1 à 3	172.997	4,3%	51.672	3,8%	158.028	12,9%	258.440	8,0%	56.443	5,8%	
De 4 à 6	770.794	19,3%	225.654	16,7%	429.843	35,0%	983.593	30,4%	230.038	23,8%	
7 ou mais	3.000.975	75,1%	1.059.595	78,4%	583.472	47,5%	1.900.075	58,7%	654.466	67,7%	
<b>Tempo de gestação</b>											<0,001*
Menos de 37 sem.	443.263	11,1%	145.655	10,8%	142.168	12,1%	360.947	11,5%	103.930	10,9%	
37 semanas ou mais	3.537.555	88,9%	1.200.138	89,2%	1.031.886	87,9%	2.767.721	88,5%	853.386	89,1%	
<b>Parto</b>											<0,001*
Vaginal	1.629.524	40,5%	520.712	38,4%	663.820	53,9%	1.617.707	49,7%	366.477	37,8%	
Cesário	2.394.653	59,5%	833.897	61,6%	568.685	46,1%	1.636.632	50,3%	603.938	62,2%	
<b>Local de nascimento</b>											<0,001*
Hospital	3.999.015	99,3%	1.346.635	99,4%	1.166.174	95,0%	3.194.787	98,0%	960.914	98,9%	
Outro estabelecimento	12.757	0,3%	3.865	0,3%	16.518	1,3%	405.49	1,2%	4.155	0,4%	
Domicílio	11.336	0,3%	3.206	0,2%	39.528	3,2%	18.590	0,6%	4.123	0,4%	
outros	4.227	0,1%	1.455	0,1%	4.846	0,4%	5.498	0,2%	2.071	0,2%	

p-valor referente a ao teste de associação de qui-quadrado de associação para variáveis categóricas. \*indica que há associação estatisticamente significativa (p-valor<0,05).

Tabela 23 – Associação dos dados da gestação (gravidez, consulta pré-natal, tempo de gestação, parto e local de nascimento) por região de nascimento do recém-nascido (Elaborado pela autora, 2020).

**Associação dos dados dos Recém-nascido com a região de nascimento**

	Sudeste		Sul		Norte		Nordeste		Centro Oeste		p-valor
	Freq.	%	Freq.	%	Freq.	%	Freq.	%	Freq.	%	
<b>Sexo</b>											0,009*
Masculino	2.060.498	51,2%	695.121	51,3%	633.102	51,3%	1.670.658	51,3%	496.873	51,2%	
Feminino	1.966.306	48,8%	659.940	48,7%	601.157	48,7%	1.588.259	48,7%	474.182	48,8%	
<b>Raça/cor</b>											<0,001*
Branca	1.780.732	45,4%	1.083.073	81,0%	99.100	8,2%	379.451	12,5%	250.795	28,7%	
Preto	290.058	7,4%	53.369	4,0%	25.901	2,1%	159.129	5,2%	33.492	3,8%	
Amarelo	20.043	0,5%	2.937	0,2%	2.507	0,2%	9.430	0,3%	5.526	0,6%	
Pardo	1.823.981	46,5%	192.861	14,4%	1.033.357	85,4%	2.469.415	81,4%	568.562	65,1%	
Indígena	6.197	0,2%	5.105	0,4%	49.844	4,1%	14.886	,5%	15.269	1,7%	
<b>anomalia</b>											<0,001*
Sim	38.752	1,0%	10.740	0,8%	6.616	0,5%	26.042	0,8%	6.204	0,7%	
não	3.912.246	99,0%	1.335.203	99,2%	1.205.687	99,5%	3.131.489	99,2%	899.399	99,3%	

p-valor referente a ao teste de associação de qui-quadrado de associação para variáveis categóricas.

\*indica que há associação estatisticamente significativa (p-valor<0,05).

Tabela 24 – Associação dos dados do recém-nascido (sexo, raça/cor, anomalia) com a região de nascimento (Elaborado pela autora, 2020).

**Comparação do perfil categórico das mães pela presença ou não de Malformação**

	<b>Anomalias / malformação</b>				p-valor
	Com Malformação		Sem Malformação		
	Freq.	%	Freq.	%	
<b>Estado civil da mãe</b>					<0,001*
Solteira	37.319	42,6%	4.394.126	42,4%	
Casada	28.878	33,0%	3.389.125	32,7%	
Viúva	181	0,2%	18.310	0,2%	
Separada	1.072	1,2%	112.889	1,1%	
União consensual	20.065	22,9%	2.458.663	23,7%	
<b>Escolaridade da mãe</b>					<0,001*
Nenhuma	508	0,6%	59.366	0,6%	
1 à 3 anos	2.409	2,8%	285.277	2,8%	
4 à 7 anos	15.990	18,3%	1.924.099	18,6%	
8 à 11 anos	51.430	58,8%	6.138.302	59,4%	
12 ou mais	17.145	19,6%	1.922.925	18,6%	

p-valor referente a ao teste de associação de qui-quadrado de associação para variáveis categóricas.

\*indica que há associação estatisticamente significativa (p-valor<0,05).

Tabela 25 – Comparação do perfil categórico das mães (estado civil e escolaridade) (Elaborado pela autora, 2020).

Conforme apresenta a Tabela 25, quando testada a associação entre a escolaridade e estado civil da vida e a presença de malformação/anomalia, encontrou-se um p-valor estatisticamente significativo nas testagens, mesmo que haja diferenças muito pequenas entre os valores encontrados.

Na Tabela 26, são apresentadas as informações da gestação associadas à presença ou não de malformação/anomalia. Houve associação estatística em todas as comparações por presença de anomalia ou não e destaca-se que a magnitude das diferenças é muito pequena, quando olhamos as proporções observadas. Muitas vezes, diferem apenas um ponto percentual ou menos.

**Comparação do perfil categórico da gestação pela presença de malformação**

	<b>Anomalias/ malformação</b>				p-valor
	<b>Com malformação</b>		<b>Sem malformação</b>		
	Freq.	%	Freq.	%	
<b>Gravidez</b>					<0,001*
Única	85.558	97,0%	10.257.270	97,9%	
dupla	2.531	2,9%	209.758	2,0%	
Tripla ou mais	71	0,1%	5.171	0,1%	
<b>Consultas pré-natal</b>					<0,001*
Nenhuma	2.127	2,4%	226.889	2,2%	
De 1 à 3	6.772	7,7%	671.600	6,4%	
De 4 à 6	22.378	25,5%	2.546.322	24,4%	
7 ou mais	56.345	64,3%	6.976.164	66,9%	
<b>Tempo de gestação</b>					<0,001*
Menos de 37 semanas	20.953	24,1%	1.144.732	11,2%	
37 semanas ou mais	65.929	75,9%	9.106.979	88,8%	
<b>Parto</b>					<0,001*
Vaginal	32.814	37,23%	4.637.627	44,28%	
Cesário	55.336	62,77%	5.835.966	55,72%	
<b>Local de nascimento</b>					<0,001*
Hospital	87.108	98,7%	10.311.741	98,4%	
Outro estabelecimento	561	0,6%	74.642	0,7%	
Domicílio	415	0,5%	72.996	0,7%	
outros	204	0,2%	17.145	0,2%	

p-valor referente a ao teste de associação de qui-quadrado de associação para variáveis categóricas.

\*indica que há associação estatisticamente significativa (p-valor<0,05).

Tabela 26 – Comparação do perfil categórico das gestações (gravidez, consultas pré-natal, tempo de gestação, parto, local de nascimento) (Elaborado pela autora, 2020).

Os dados dos recém-nascidos, sexo e raça/cor da pele, também foram associados de maneira estatisticamente significativa com a presença ou não de anomalias/malformação

(Tabela 27). Para buscar desenhar um panorama melhor do perfil relacionado às malformação/anomalias, as tabelas em sequência associam as variáveis descritivas das mães, da gestação e dos recém-nascidos com a presença ou não de anomalias/malformação, estratificadas por região de nascimento.

**Comparação do perfil categórico dos recém-nascidos pela presença de malformação**

	<b>Anomalias/ Malformação</b>				p-valor
	Com malformação		Sem malformação		
	Freq.	%	Freq.	%	
<b>Sexo</b>					<0,001*
Masculino	49.462	56,80%	5.366.009	51,20%	
Feminino	37.572	43,20%	5.118.013	48,80%	
<b>Raça/cor</b>					<0,001*
Branca	31.545	37,30%	3.516.443	34,90%	
Preto	5.593	6,60%	544.594	5,40%	
Amarelo	457	0,50%	38.690	0,40%	
Pardo	46.341	54,80%	5.894.596	58,50%	
Indígena	628	0,70%	89.163	0,90%	

p-valor referente a ao teste de associação de qui-quadrado de associação para variáveis categóricas.

Tabela 27 – Comparação do perfil categórico das recém-nascido (sexo, raça/ cor) pela presença de malformação (Elaborado pela autora, 2020).

Quanto às associações do estado civil das mães com a presença ou não de malformação, apenas a região Norte não apresentou associação estatisticamente significativa do estado civil da mãe com anomalias. A Tabela 28 apresenta tais estatísticas.

**Comparação do estado civil das mães pela presença de malformação, estratificadas por regiões do país.**

Região		Anomalias/ malformação					
		Com malformação			Sem malformação		
		Estado civil	Freq.	%	Freq.	Freq.	%
Sudeste	Estado civil					0,045*	
	Solteiro	17.062	44,3%	1.744.166	44,9%		
	Casada	15.410	40,0%	1.546.073	39,8%		
	Viúva	74	0,2%	7.425	0,2%		
	Separada	664	1,7%	63.124	1,6%		
	União consensual	5.322	13,8%	522.237	13,4%		
Sul	Solteiro	4.476	41,9%	541.321	40,7%	0,036*	
	Casada	3.573	33,4%	462.608	34,8%		
	Viúva	23	0,2%	2.454	0,2%		
	Separada	147	1,4%	17.207	1,3%		
	União consensual	2.466	23,1%	305.711	23,0%		
Norte	Solteiro	2.669	41,0%	483.244	40,6%	0,108	
	Casada	1.256	19,3%	221.078	18,6%		
	Viúva	11	0,2%	1.310	0,1%		
	Separada	30	0,5%	4.875	0,4%		
	União consensual	2.537	39,0%	480.346	40,3%		
Nordeste	Solteiro	10.424	40,6%	1.274.332	41,3%	<0,001*	
	Casada	6.708	26,1%	845.864	27,4%		
	Viúva	58	0,2%	5.704	0,2%		
	Separada	155	0,6%	16.887	0,5%		
	União consensual	8.342	32,5%	939.994	30,5%		
Centro Oeste	Solteiro	2.688	44,0%	35.1063	39,6%	<0,001*	
	Casada	1.931	31,6%	313.502	35,3%		
	Viúva	15	0,2%	1.417	0,2%		
	Separada	76	1,2%	10.796	1,2%		
	União consensual	1.398	22,9%	210.375	23,7%		

p-valor referente a ao teste de associação de qui-quadrado de associação para variáveis categóricas.

\*indica que há associação estatisticamente significativa (p-valor<0,05).

Tabela 28 – Comparação do estado civil das mães pela presença de malformação por região do Brasil (Elaborado pela autora, 2020).

Todas as associações de escolaridade (Tabela 29) da mãe em relação à presença de anomalias/malformação, em todas as regiões mensuradas, foram estatisticamente significativas. Na sequência, serão discriminadas as associações em relação aos dados da gestação.

Comparação da escolaridade das mães pela presença de malformação, estratificadas por regiões do país.

Região	Escolaridade	Anomalias/ malformação				p-valor
		Com malformação		Sem malformação		
		Freq.	%	Freq.	%	
Sudeste	Nenhuma	90	0,2%	7.061	0,2%	<0,001*
	1 à 3 anos	598	1,6%	56.606	1,5%	
	4 à 7 anos	5.166	13,4%	544.053	14,1%	
	8 à 11 anos	23.032	59,9%	2.419.731	62,5%	
	12 ou mais	9.595	24,9%	842.894	21,8%	
Sul	Nenhuma	37	0,3%	2.585	0,2%	<0,001*
	1 à 3 anos	231	2,2%	20.843	1,6%	
	4 à 7 anos	1.949	18,2%	217.071	16,3%	
	8 à 11 anos	6.126	57,2%	775.453	58,3%	
	12 ou mais	2.365	22,1%	314.709	23,7%	
Norte	Nenhuma	127	2,0%	18.090	1,5%	0,023*
	1 à 3 anos	343	5,3%	56.974	4,8%	
	4 à 7 anos	1.634	25,1%	294.615	25,0%	
	8 à 11 anos	3.634	55,8%	666.598	56,5%	
	12 ou mais	769	11,8%	143.911	12,2%	
Nordeste	Nenhuma	221	0,9%	28.086	0,9%	0,002*
	1 à 3 anos	1.102	4,3%	135.234	4,4%	
	4 à 7 anos	6.202	24,2%	735.523	24,1%	
	8 à 11 anos	14.975	58,4%	1.760.106	57,6%	
	12 ou mais	3.147	12,3%	398.840	13,0%	
Centro Oeste	Nenhuma	33	0,5%	3.544	0,4%	<0,001*
	1 à 3 anos	135	2,2%	15.620	1,8%	
	4 à 7 anos	1.039	16,9%	132.837	14,9%	
	8 à 11 anos	3.663	59,7%	516.414	58,0%	
	12 ou mais	1.269	20,7%	222.571	25,0%	

p-valor referente a ao teste de associação de qui-quadrado de associação para variáveis categóricas.

\*indica que há associação estatisticamente significativa (p-valor<0,05).

Tabela 29 – Comparação da escolaridade das mães pela presença de malformação por região do Brasil (Elaborado pela autora, 2020).

Tanto a variável gravidez, quanto as consultas pré-natal foram associadas de maneira significativa com a presença de anomalias/malformação em todas as regiões do país (Tabela 30).

Associação dos dados de gestação e consultas pré-natais com a presença se má formações, estratificadas por região de nascimento dos recém-nascido

<b>Anomalias/ Malformação</b>						
		<b>Com</b>		<b>Sem</b>		
		<b>malformação</b>		<b>Malformação</b>		
<b>Região</b>	<b>Gravidez</b>	<b>Freq.</b>	<b>%</b>	<b>Freq.</b>	<b>%</b>	<b>p-valor</b>
<b>Sudeste</b>	Única	37.502	96,9%	3.818.693	97,7%	<0,001*
	Dupla	1.167	3,0%	88.131	2,3%	
	Tripla ou mais	41	0,1%	2.434	0,1%	
<b>Sul</b>	Única	10.400	96,9%	1.304.227	97,7%	<0,001*
	Dupla	320	3,0%	29.481	2,2%	
	Tripla ou mais	10	0,1%	694	0,1%	
<b>Norte</b>	Única	6.438	97,8%	1.185.595	98,5%	<0,001*
	Dupla	143	2,2%	18.007	1,5%	
	Tripla ou mais	1	<0,01%	279	<0,01%	
<b>Nordeste</b>	Única	25.216	97,1%	3.068.773	98,2%	<0,001*
	Dupla	732	2,8%	56.176	1,8%	
	Tripla ou mais	14	0,1%	1.222	<0,01%	
<b>Centro oeste</b>	Única	6.002	97,2%	879.982	97,9%	<0,001*
	Dupla	169	2,7%	17.963	2,0%	
	Tripla ou mais	5	0,1%	542	0,1%	
<b>Consultas pré-natal</b>						
<b>Sudeste</b>	Nenhuma	548	1,4%	50.228	1,3%	<0,001*
	1 à 3	2.164	5,6%	167.405	4,3%	
	4 à 6	7.699	20,0%	748.018	19,3%	
	7 ou mais	28.030	72,9%	2.917.866	75,1%	
<b>Sul</b>	Nenhuma	187	1,7%	14.007	1,1%	<0,001*
	1 à 3	574	5,4%	50.695	3,8%	
	4 à 6	2.171	20,3%	221.962	16,7%	
	7 ou mais	7.772	72,6%	1.045.399	78,5%	
<b>Norte</b>	Nenhuma	365	5,5%	56.220	4,7%	<0,001*
	1 à 3	916	13,9%	154.348	12,9%	
	4 à 6	2.401	36,5%	419.724	35,0%	
	7 ou mais	2.895	44,0%	570.491	47,5%	
<b>Nordeste</b>	Nenhuma	810	3,1%	84.402	2,7%	<0,001*
	1 à 3	2.589	10,1%	248.169	8,0%	
	4 à 6	8.424	32,7%	944.574	30,4%	
	7 ou mais	13.938	54,1%	1.832.511	58,9%	
<b>Centro oeste</b>	Nenhuma	217	3,5%	22.032	2,5%	<0,001*
	1 à 3	529	8,6%	50.983	5,7%	
	4 à 6	1.683	27,4%	212.044	23,7%	
	7 ou mais	3.710	60,4%	609.897	68,1%	

p-valor referente a ao teste de associação de qui-quadrado de associação para variáveis categóricas. \*indica que há associação estatisticamente significativa (p-valor<0,05).

Tabela 30 – Comparação dos dados de gestação com a presença ou não de malformação (Elaborado pela autora, 2020).

O tempo de gestação e o tipo de parto mostraram-se associados de maneira estatisticamente significativa com a presença de malformação/anomalias, em todas as regiões mensuradas, indicando a relação entre as variáveis (Tabela 31).

**Associação dos dados de tempo de gestação e tipo de parto com a presença ou não de malformações, estratificadas por região de nascimento dos recém-nascido**

		Anomalias/ Malformação				p-valor
		Com Malformação		Sem Malformação		
		Freq.	%	Freq.	%	
<b>Região</b>	<b>Tempo de gestação</b>					<0,001*
Sudeste	prematureo	9.031	23,50%	425.786	11,00%	
	tempo normal	29.445	76,50%	3.445.782	89,00%	
Sul	prematureo	2.775	26,00%	141.485	10,70%	<0,001*
	tempo normal	7.884	74,00%	1.184.750	89,30%	
Norte	prematureo	1.551	24,40%	137.900	12,00%	<0,001*
	tempo normal	4.794	75,60%	1.009.982	88,00%	
Nordeste	prematureo	6.100	24,10%	344.238	11,40%	<0,001*
	tempo normal	19.226	75,90%	2.674.308	88,60%	
Centro-Oeste	prematureo	1.496	24,60%	95.323	10,70%	<0,001*
	Tempo normal	4.580	75,40%	792.157	89,30%	
	<b>Tipo de parto</b>					
Sudeste	vaginal	13.557	35,00%	1.586.404	40,60%	<0,001*
	cesário	25.147	65,00%	2.322.962	59,40%	
Sul	vaginal	3.188	29,70%	514.053	38,50%	<0,001*
	cesário	7.545	70,30%	820.619	61,50%	
Norte	vaginal	2.967	45,10%	648.438	53,90%	<0,001*
	cesário	3.611	54,90%	555.507	46,10%	
Nordeste	vaginal	11.156	43,00%	1.559.562	49,90%	<0,001*
	cesário	14.803	57,00%	1.567.488	50,10%	
Centro-Oeste	vaginal	1.946	31,50%	329.170	36,60%	<0,001*
	cesário	4.230	68,50%	569.390	63,40%	

p-valor referente a ao teste de associação de qui-quadrado de associação para variáveis categóricas.

\*indica que há associação estatisticamente significativa (p-valor<0,05).

Tabela 31 – Comparação do tempo de gestação e tipo de parto pela presença de malformação por região do Brasil (Elaborado pela autora, 2020).

Assim como ocorreu com ademais variáveis, as características do local de parto também foram associadas com a presença de malformação/anomalias, em todas as regiões, mesmo que as magnitudes das diferenças não tenham sido significativas (Tabela 32).

**Associação dos dados de local de nascimento com a presença ou não de malformações, estratificadas por região de nascimento dos recém-nascido.**

Região	Local de nascimento	Anomalias/ malformação				
		Com malformação		Sem Malformação		p-valor
		Freq.	%	Freq.	%	
Sudeste	Hospital	38.386	99,1%	3.885.005	99,3%	<0,001*
	Outro estabelecimento de saúde	193	0,5%	12.294	0,3%	
	Domicilio	103	0,3%	10.829	0,3%	
	Outros	63	0,2%	4.020	0,1%	
Sul	Hospital	10.647	99,1%	1.326.935	99,4%	0,003*
	Outro estabelecimento de saúde	48	0,4%	3.786	0,3%	
	Domicilio	26	0,2%	3.046	0,2%	
	Outros	18	0,2%	1.405	0,1%	
Norte	Hospital	6.309	96,0%	1.139.673	95,1%	<0,001*
	Outro estabelecimento de saúde	70	1,1%	16.167	1,3%	
	Domicilio	159	2,4%	38.193	3,2%	
	Outros	32	0,5%	4.642	0,4%	
Nordeste	Hospital	25.655	98,6%	3.070.654	98,1%	<0,001*
	Outro estabelecimento de saúde	202	0,8%	38.432	1,2%	
	Domicilio	98	0,4%	17.030	0,5%	
	Outros	76	0,3%	5.123	0,2%	
Centro-oeste	Hospital	6.111	98,5%	889.474	98,9%	0,001*
	Outro estabelecimento de saúde	48	0,8%	3.963	0,4%	
	Domicilio	29	0,5%	3.898	0,4%	
	Outros	15	0,2%	1.955	0,2%	

p-valor referente a ao teste de associação de qui-quadrado de associação para variáveis categóricas.

\*Indica que há associação estatisticamente significativa (p-valor<0,05).

1

Tabela 32 – Comparação do local de nascimento pela presença de malformação por região de nascimento (Elaborado pela autora, 2020).

Tanto a raça/cor, quanto o sexo dos recém-nascidos foram associados de maneira estatisticamente significativa com a presença de anomalias/malformação detectadas ao nascer(Tabela 33).

**Associação dos dados dos recém-nascidos com a presença ou não de Malformações, estratificadas por região de nascimento dos recém-nascido.**

		Anomalia / Malformação				
		Com Malformação		Sem Malformação		
		Freq.	%	Freq.	%	p-valor
Região	Sexo					
Sudeste	Masculino	21.744	56,80%	1.999.536	51,10%	<0,001*
	Feminino	16.531	43,20%	1.912.710	48,90%	
Sul	Masculino	6.109	57,50%	684.321	51,30%	<0,001*
	Feminino	4.512	42,50%	650.882	48,70%	
Norte	Masculino	3.635	55,90%	618.194	51,30%	<0,001*
	Feminino	2.863	44,10%	587.491	48,70%	
Nordeste	Masculino	14.539	56,90%	1.604.255	51,20%	<0,001*
	Feminino	11.011	43,10%	1.527.234	48,80%	
Centro-Oeste	Masculino	3.435	56,40%	459.703	51,10%	<0,001*
	Feminino	2.655	43,60%	439.696	48,90%	
Região	Raça/cor	Freq.	%	Freq.	%	p-valor
Sudeste	Branca	18.196	47,90%	1.738.319	45,50%	<0,001*
	Preta	2.997	7,90%	281.404	7,40%	
	Amarela	326	0,90%	19.447	0,50%	
	Parda	16.415	43,20%	1.775.982	46,50%	
	Indígena	66	0,20%	6.053	0,20%	
Sul	Branca	8.470	79,80%	1.068.096	81,00%	<0,001*
	Preta	585	5,50%	52.309	4,00%	
	Amarela	26	0,20%	2.892	0,20%	
	Parda	1.497	14,10%	190.279	14,40%	
	Indígena	37	0,30%	5.020	0,40%	
Norte	Branca	573	8,80%	96.989	8,20%	<0,001*
	Preta	163	2,50%	25.289	2,10%	
	Amarela	9	0,10%	2.446	0,20%	
	Parda	5.425	83,60%	1.011.358	85,40%	
	Indígena	319	4,90%	48.622	4,10%	
Nordeste	Branca	2.855	11,90%	371.388	12,60%	<0,001*
	Preta	1.588	6,60%	154.887	5,30%	
	Amarela	69	0,30%	9.157	0,30%	
	Parda	19.425	80,90%	2.393.019	81,30%	
	Indígena	82	0,30%	14.594	0,50%	
Centro-Oeste	Branca	1.451	26,70%	241.651	29,60%	<0,001*
	Preta	260	4,80%	30.705	3,80%	
	Amarela	27	0,50%	4.748	0,60%	
	Parda	3.579	65,80%	523.958	64,20%	
	Indígena	124	2,30%	14.874	1,80%	

p-valor referente a ao teste de associação de qui-quadrado de associação para variáveis categóricas.

\*indica que há associação estatisticamente significativa (p-valor<0,05).

Tabela 33 – Comparação dos dados dos recém-nascidos (sexo, raça/ cor) pela presença de malformação (Elaborado pela autora, 2020).

Outra análise realizada foi referente ao tipo de anomalia. Para tanto, foram usadas as definições de classificação da Estatística Internacional de Doenças e Problemas Relacionados com a Saúde, criado pela Organização Mundial da Saúde (OMS), a partir dos dados do banco, que considera quatro anos. Identificou-se 88.848.232 nascidos vivos com indicação de anomalia/malformação. Com estes, observou-se o código de identificação da CID-10 para classificar as anomalias de acordo com a nomenclatura adequada. Excluíram-se dados de 286 crianças pelo código indicar Hemangioma, que é um tipo de tumor benigno. Os demais dados que foram considerados adequados estão apresentados nos parágrafos seguintes.

Encontrou-se associação estatisticamente significativa entre o tipo de anomalia e o ano de nascimento do recém-nascido. Identificamos que, para a maioria das variáveis, há uma tendência de aumento na quantidade de observações com o avanço dos anos (Tabela 34).

#### Associação do tipo de anomalia/ malformação com o ano de nascimento.

Tipo de anomalia	2014		2015		2016		2017		p-valor
	N	%	N	%	N	%	N	%	
sistema nervoso	1.755	7,90%	2.359	16,10%	3.685	14,20%	2020	8,00%	<0,001*
face, pescoço, olhos e ouvido	964	4,30%	530	3,60%	1.138	4,40%	1.169	4,70%	
aparelho circulatório	1.483	6,70%	489	3,30%	1.694	6,50%	1.957	7,80%	
aparelho respiratório	148	0,70%	83	0,60%	133	0,50%	162	0,60%	
fenda no paláto ou lábios	1.118	5,00%	688	4,70%	1.217	4,70%	1.179	4,70%	
aparelho digestivo	736	3,30%	475	3,20%	983	3,80%	1.066	4,20%	
orgãos genitais	1.561	7,00%	929	6,30%	1.595	6,10%	1.778	7,10%	
aparelho urinário	327	1,50%	157	1,10%	345	1,30%	391	1,60%	
osteomuscular	8.447	38,10%	5.301	36,20%	8.605	33,20%	8.685	34,50%	
outras malformações	1.078	4,90%	694	4,70%	1.061	4,10%	1.065	4,20%	
cromossomicas	915	4,10%	597	4,10%	989	3,80%	1.054	4,20%	
Múltiplas anomalias	3.667	16,50%	2.334	15,90%	4.491	17,30%	4.613	18,30%	

p-valor referente a ao teste de associação de qui-quadrado de associação para variáveis categóricas. \*indica que há associação estatisticamente significativa (p-valor<0,05). n.a: não foi possível aplicar o teste de qui-quadrado por não atender aos pressupostos do mesmo, apresentou casa zerada.

Tabela 34 – Associação do tipo de anomalia e malformação com o ano de nascimento

(Elaborado pela autora, 2020).

Também se encontrou associação estatística entre a região de nascimento e o tipo de anomalia malformação. Destacamos que as regiões Sudeste e Nordeste apresentam normalmente maiores frequências que as demais (Tabela 35).

**Associação do tipo de anomalias por região de nascimento da criança.**

Tipo de anomalia	Região de nascimento										p-valor
	Sudeste		Sul		Norte		Nordeste		Centro-oeste		
	N	%	N	%	N	%	N	%	N	%	
sistema nervoso	3.365	8,70%	776	7,30%	909	13,90%	4.086	15,70%	683	11,10%	<0,001*
face, pescoço, olhos e ouvido	1.803	4,70%	422	3,90%	348	5,30%	1.001	3,90%	227	3,70%	
aparelho circulatório	3.742	9,70%	822	7,70%	142	2,20%	705	2,70%	212	3,40%	
aparelho respiratório	268	0,70%	75	0,70%	30	0,50%	121	0,50%	32	0,50%	
fenda no paláto ou lábios	1.659	4,30%	677	6,30%	488	7,50%	1.036	4,00%	342	5,50%	
aparelho digestivo	1.582	4,10%	395	3,70%	302	4,60%	757	2,90%	224	3,60%	
orgãos genitais	2.434	6,30%	736	6,90%	257	3,90%	2.053	7,90%	383	6,20%	
aparelho urinário	652	1,70%	189	1,80%	47	0,70%	262	1,00%	70	1,10%	
osteomuscular	12.703	32,90%	3.563	33,30%	2.520	38,50%	9.827	37,90%	2.425	39,30%	
outras máformações	1.505	3,90%	487	4,60%	225	3,40%	1.331	5,10%	350	5,70%	
alterações cromossômicas	1.570	4,10%	622	5,80%	231	3,50%	918	3,50%	214	3,50%	
mais de uma identificação de anomalia	7.272	18,90%	1.922	18,00%	1.045	16,00%	3.856	14,90%	1.010	16,40%	

Tabela 35 – Associação do tipo de anomalias por região de nascimento do recém-nascido (Elaborado pela autora, 2020).

Conforme apresenta a Tabela 36, encontrou-se associação significativa entre a gravidez e o tempo de gestação. Destacamos que, nas crianças nascidas com menos de 37 semanas de gestação, apresentaram-se maiores porcentagens de situações em que mais de um código de anomalia (26,1%) foi identificada em comparação às nascidas no tempo adequado (14,4%).

Tipo de anomalia	Tempo de gestação				p-valor
	prematureo		tempo normal		
	N	%	N	%	
sistema nervoso	2.533	12,10%	7.077	10,80%	<0,001*
face, pescoço, olhos e ouvido	553	2,60%	3.185	4,90%	
aparelho circulatório	1.237	5,90%	4.346	6,60%	
aparelho respiratório	164	0,80%	349	0,50%	
fenda no paláto ou lábios	680	3,30%	3.454	5,30%	
aparelho digestivo	701	3,40%	2.495	3,80%	
orgãos genitais	1.121	5,40%	4.654	7,10%	
aparelho urinário	407	1,90%	805	1,20%	
osteomuscular	5.752	27,50%	24.732	37,70%	
outras máformações	1.375	6,60%	2.442	3,70%	
alterações cromossômicas	918	4,40%	2.580	3,90%	
mais de uma identificação de anomalia	5.449	26,10%	9.431	14,40%	

Tabela 36 – Associação do tipo de anomalia com o tempo de gestação (prematureo, temponormal) (Elaborado pela autora, 2020).

Não foi possível aplicar o teste de associação entre o tipo de anomalia e as características da gravidez (Tabela 37), pois havia uma quebra do pressuposto do teste, apresentando uma casa com frequência zero.

Associação de características da gravidez com o tipo de anomalia							
Tipo de anomalia	GRAVIDEZ						p-valor
	única		dupla		3 ou mais		
	N	%	N	%	N	%	
sistema nervoso	9.555	11,20%	230	9,10%	7	9,90%	n.a
face, pescoço, olhos e ouvido	3.705	4,40%	90	3,60%	1	1,40%	
aparelho circulatório	5.372	6,30%	235	9,30%	10	14,10%	
aparelho respiratório	502	0,60%	23	0,90%	1	1,40%	
fenda no paláto ou lábios	4.083	4,80%	103	4,10%	3	4,20%	
aparelho digestivo	3.165	3,70%	85	3,40%	2	2,80%	
orgãos genitais	5.635	6,60%	205	8,10%	7	9,90%	
aparelho urinário	1.179	1,40%	37	1,50%	3	4,20%	
osteomuscular	30.174	35,40%	773	30,60%	15	21,10%	
outras másformações	3.667	4,30%	217	8,60%	6	8,50%	
alterações cromossomicas	3.499	4,10%	47	1,90%	0	0,00%	
mais de uma identificação de anomalia	14.583	17,10%	481	19,00%	16	22,50%	

Tabela 37 – Associação de características da gravidez com o tipo de anomalia (Elaborado pela autora, 2020).

Conforme apresenta a Tabela 38, encontrou-se associação estatisticamente significativa entre a quantidade de consultas pré-natal e o tipo de anomalia malformação dos recém-nascidos, mas destacamos que as diferenças nas porcentagens observadas não apresentaram grande magnitude. Um dos padrões mais claros foi que, para as malformações do sistema nervoso, a quantidade observada diminui em relação à quantidade de consultas realizadas pelas mães.

Tipo de anomalia	CONSULTAS								p-valor
	nenhuma		1 à 3		4 à 6		7 ou mais		
	N	%	N	%	N	%	N	%	
sistema nervoso	378	17,90%	1.020	15,10%	2.836	12,70%	5.475	9,80%	<0,001*
face, pescoço, olhos e ouvido	75	3,50%	261	3,90%	896	4,00%	2.542	4,50%	
aparelho circulatório	77	3,60%	282	4,20%	1.049	4,70%	4.177	7,50%	
aparelho respiratório	11	0,50%	44	0,70%	126	0,60%	341	0,60%	
fenda no paláto ou lábios	102	4,80%	257	3,80%	1.055	4,70%	2.763	4,90%	
aparelho digestivo	64	3,00%	226	3,40%	733	3,30%	2.217	4,00%	
orgãos genitais	126	6,00%	426	6,30%	1.455	6,50%	3.810	6,80%	
aparelho urinário	18	0,90%	80	1,20%	266	1,20%	843	1,50%	
osteomuscular	693	32,80%	2.311	34,30%	8.080	36,30%	19.741	35,20%	
outras máformações	148	7,00%	377	5,60%	1.080	4,80%	2.252	4,00%	
alterações cromossomicas	74	3,50%	198	2,90%	815	3,70%	2.444	4,40%	
mais de uma identificação de anomalia	347	16,40%	1.254	18,60%	3.888	17,50%	9.451	16,90%	

p-valor referente a ao teste de associação de qui-quadrado de associação para variáveis categóricas.

\*indica que há associação estatisticamente significativa (p-valor<0,05). n.a: não foi possível aplicar o teste de qui-quadrado por não atender aos pressupostos do mesmo, apresentou casa zerada.

Tabela 38 – Associação da quantidade de consultas por tipo de anomalia (Elaborado pela autora, 2020).

Conforme apresenta a Tabela 39, encontrou-se associação com significância estatística na comparação dos tipos de anomalias do recém-nascidos. Destacamos que as diferenças nas percentagens observadas não foram grandes, mas, conforme o tipo de anomalia apresentada pelo recém-nascido, há uma variação nos locais de nascimento. Supõe-se que pode haver um problema estrutural nessa variável, pois algumas malformações podem ficar despercebidas, caso o nascimento não ocorra com o acompanhamento adequado, sendo essa uma situação que talvez precise de estudos mais aprofundados.

LOCAL DE NASCIMENTO									
Tipo de anomalia	Hospital		outro estabelecimento de saúde		Domicilio		outros		p-valor
	N	%	N	%	N	%	N	%	
sistema nervoso	9.693	11,20%	52	9,30%	44	11,10%	26	12,90%	<0,001*
face, pescoço, olhos e ouvido	3.740	4,30%	27	4,80%	27	6,80%	4	2,00%	
aparelho circulatório	5.588	6,40%	24	4,30%	6	1,50%	5	2,50%	
aparelho respiratório	522	0,60%	0	0,00%	3	0,80%	1	0,50%	
fenda no paláto ou lábios	4.130	4,80%	29	5,20%	26	6,60%	15	7,40%	
aparelho digestivo	3.202	3,70%	33	5,90%	17	4,30%	6	3,00%	
orgãos genitais	5.793	6,70%	36	6,50%	21	5,30%	12	5,90%	
aparelho urinário	1.208	1,40%	8	1,40%	2	0,50%	2	1,00%	
osteomuscular	30.549	35,20%	237	42,50%	163	41,30%	73	36,10%	
outras máformações	3.848	4,40%	18	3,20%	15	3,80%	11	5,40%	
alterações cromossômicas	3.488	4,00%	30	5,40%	25	6,30%	12	5,90%	
mais de uma identificação de anomalia	14.956	17,20%	63	11,30%	46	11,60%	35	17,30%	

p-valor referente a ao teste de associação de qui-quadrado de associação para variáveis categóricas.

\*indica que há associação estatisticamente significativa (p-valor<0,05). n.a: não foi possível aplicar o teste de qui-quadrado por não atender aos pressupostos do mesmo, apresentou casa zerada.

Tabela 39 – Associação entre o local de nascimento e tipo de anomalia apresentada pelo recém-nascido (Elaborado pela autora, 2020).

### 4.3.3 Dashboards Resultantes da Análise Estatística

Com base os resultados obtidos com cálculos estatísticos apresentados na última subseção, foram desenvolvidos *dashboards* comparativo e um específico para anomalias. Os cálculos foram essenciais para identificar as variáveis de maior influência e incorpora-las aos *dashboards*. Enquanto os cálculos estatísticos representam uma visão mais abrangente e estática (dos dados em momento) dos dados, os *dashboards* proveem cálculos dinâmicos, bem como uma interação dinâmica com os dados e visualizações focadas nas principais variáveis identificadas durante a análise estatística.

Assim, o *dashboard* comparativo desenvolvido permite aos usuários:

- Selecionar um estado no mapa.
- Selecionar o ano a ser analisado, todos (2014, 2015, 2016 e 2017).
- Comparar informações com malformação, sem malformação com estado civil da mãe.
- Comparar informações com malformação, sem malformação com escolaridade da mãe.
- Comparar informações com malformação, sem malformação com tipo da gravidez.
- Comparar informações com malformação, sem malformação com consultas pré-natal.

- Comparar informações com malformação, sem malformação com tempo da gestação.
- Comparar informações com malformação, sem malformação com tipo do parto.
- Comparar informações com malformação, sem malformação com o local de nascimento.
- Comparar informações com malformação, sem malformação com o sexo do recém-nascido.
- Comparar informações com malformação, sem malformação com a raça/ cor do recém-nascido.

A Figura 45 apresenta resultados nos últimos 4 anos (2014, 2015, 2016 e 2017), os quais incluem tanto gestações com malformação como sem malformação. O estado civil mais frequente é o solteiro, o tempo de escolaridade da mãe 8 a 11 anos, o tipo de gravidez única, as consultas pré-natal foi de 7 ou mais consultas, tempo da gestação é de 37 semanas ou mais, tipo de parto é cesáreo, o local de nascimento é hospital, o sexo do recém-nascido é o masculino e a raça/cor do recém-nascido é a parda.

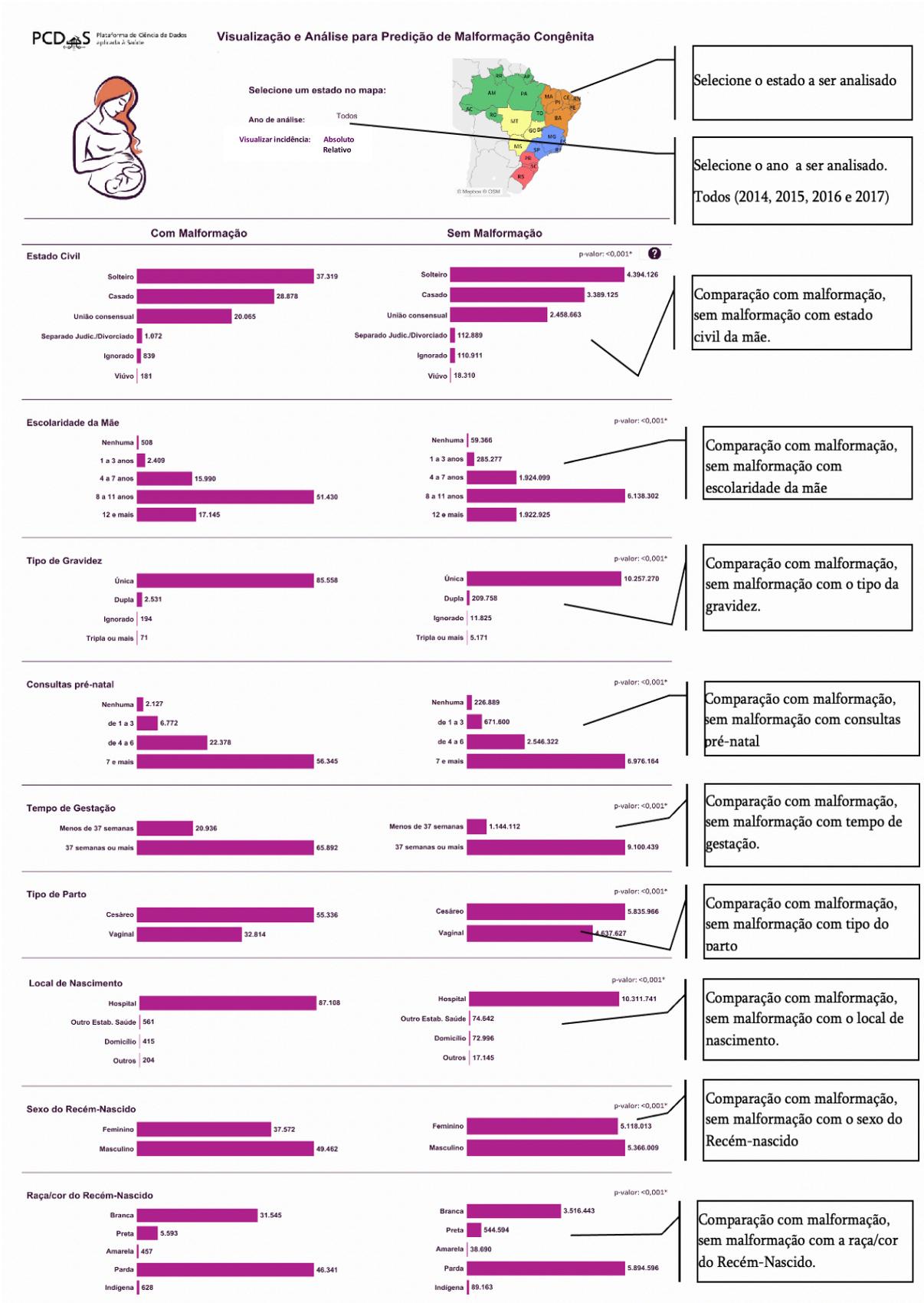


Figura 45 – Dashboard com resultados do filtro de comparação (Elaborado pela autora, 2020).

Na Figura 46 os *dashboard* com foco em Anomalias permite aos usuários:

- Selecionar o ano a ser analisado (base 2014, 2015, 2016 e 2017).
- Selecionar o estado a ser analisado.
- Filtrar anomalias por Capítulo, Grupo, Categoria, Subcategoria e Quantidade casos.
- Filtrar por Ano, Região, Tempo de Gestação, Tipo de Gestação, Consultas e Local de Nascimento.
- Analisar incidência da anomalia, por Região, Estado e cidade e quantidade de anomalia.

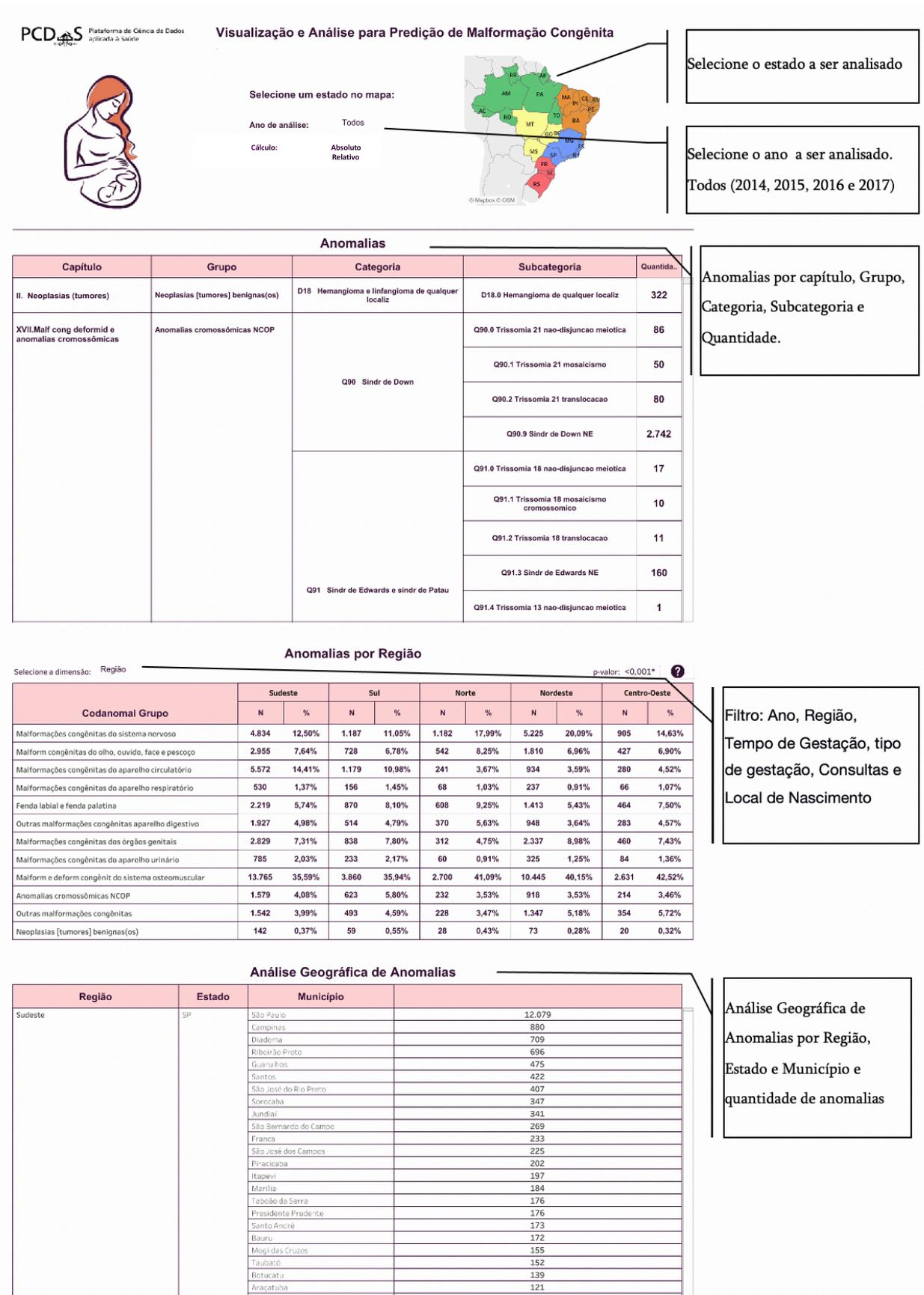


Figura 46 – Dashboard com resultados do filtro de Anomalias (Elaborado pela autora, 2020).

## 5. AVALIAÇÃO COM USUÁRIOS E ESPECIALISTAS

Este capítulo apresenta a avaliação de usabilidade dos *dashboards* apresentados no capítulo anterior. O público alvo inclui todos os interessados neste objeto de estudo incluindo mães, agentes de saúde e da área da demografia. A seção 5.1 apresenta o método de avaliação, bem como o perfil detalhado dos participantes; a seção 5.2 destaca os resultados obtidos do ponto de vista quantitativo e qualitativo; por fim, a seção 5.3 discute os resultados.

### 5.1 MÉTODO DE AVALIAÇÃO E PARTICIPANTES

A coleta de dados para avaliação da usabilidade se realizou através do questionário SUS aplicado por meio de um formulário Google para mães, estudantes e pesquisadores. Os participantes foram recrutados por meio de convite a grupos de interesse nas redes sociais Facebook e Instagram. Também participou especialista na área de Demografia e Saúde, convidada por e-mail.

Conforme destacado na seção 2.5, o SUS é utilizado em pesquisa onde é definido como método de averiguação de nível de usabilidade de um sistema. É um questionário quantitativo de usabilidade que consiste em 10 questões respondidas que avalia em uma escala 1 a 5 tipo *Likert*, onde o 1 significa “discordo completamente” e 5 “concordo completamente”. A Figura 40 apresenta 20 perguntas do formulário aplicado nesta avaliação, além destas perguntas o formulário também contou com 10 perguntas adicionais sobre características dos *dashboards* e visualizações georeferenciadas. Ao final foi feita uma pergunta aberta sobre sugestões de melhorias. A tabela 40 apresenta as questões realizadas.

1. Gostaria de usar este painel com frequência?
2. Achei as visualização desnecessariamente complexas.
3. Achei o Painel de Visualização fácil de usar.
4. Achei que seria necessário o apoio de um técnico para poder usar e concluir as análises.
5. As funções destas visualização estavam bem integradas
6. Achei as visualização dos dados muito inconsistente
7. Imagino que a maioria das pessoas aprenderiam visualizar este gráficos rapidamente.
8. Achei as visualização muito complicado de usar
9. Eu me senti muito confiante com o visualização.
10. Eu preciso aprender visualizar gráfica antes de continuar usando este painel.
11. Eu me senti confortável com esta visualização.
12. Foi fácil encontrar a informação que eu precisava
13. Eu gostei de usar a interface do Painel.
14. A Interface do gráficos é agradável
15. A organização de informações na tela do painel é clara
16. A figura do mapa: os estados brasileiros que apresentam as análises teve dificuldade na visualização no mapa ?
17. A figura do mapa: Consegue visualizar os dados de cada estados nas Análise Gerais?
18. A figura do mapa: Consegue visualizar os dados de cada estados nas análise da mãe?
19. A figura do mapa: Consegue visualizar os dados de cada estados nas análise da gestação?
20. A figura do mapa: Consegue visualizar os dados de cada estados nas análise da recém-nascido?
Sugestão Gerais:

Tabela 40: Questionário adaptado com base no System Usability Scale (SUS).

O estudo foi realizado durante janeiro de 2020. Ao todo 29 pessoas com diferentes perfis participaram neste estudo, a saber: 4 especialistas em demografia (2 professores e 2 pesquisadores), 6 profissionais da área da saúde (enfermeiras, dentista e veterinário); 2 cientistas da computação, e 16 pessoas de outras especialidades e profissões sobre o papel do paciente (por exemplo, 3 professores, 2 operadores de *call center*, 2 menores alunos de pós-graduação, etc.). A Figura 47 apresenta a distribuição dos participantes de acordo com a profissão<sup>9</sup>.

<sup>9</sup> Alguns profissionais de saúde estão identificados como servidor público ou professor

**Profissão:**

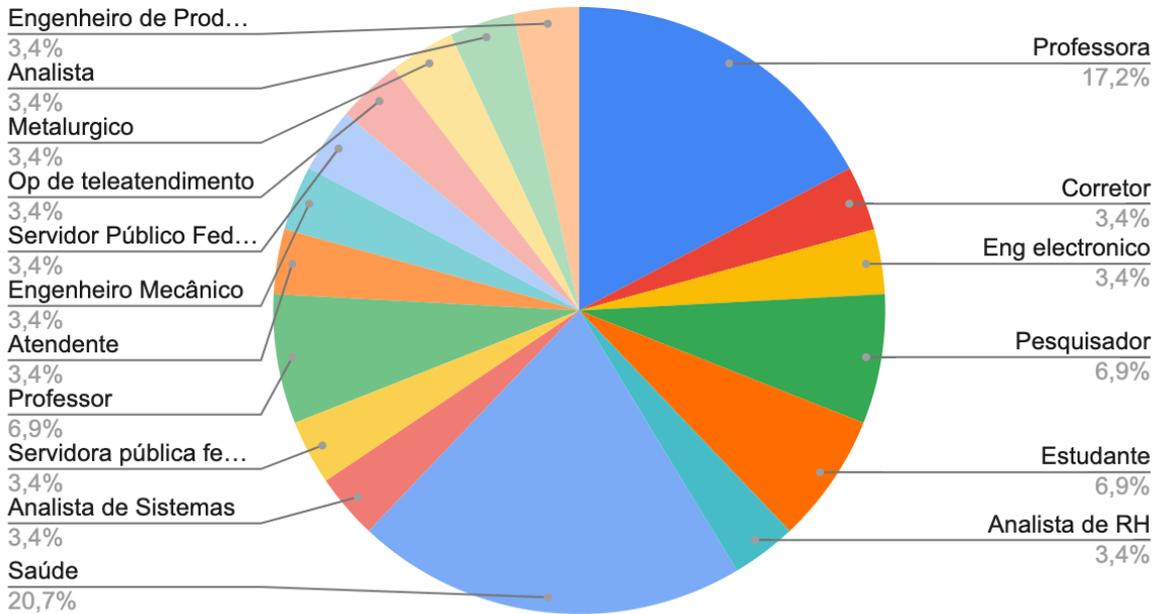


Figura 47 – Resultados do perfil profissão que fez avaliação do teste de usabilidade.

A Figura 48 apresenta a distribuição por sexo, sendo 19 participantes são mulheres e 10 homens.

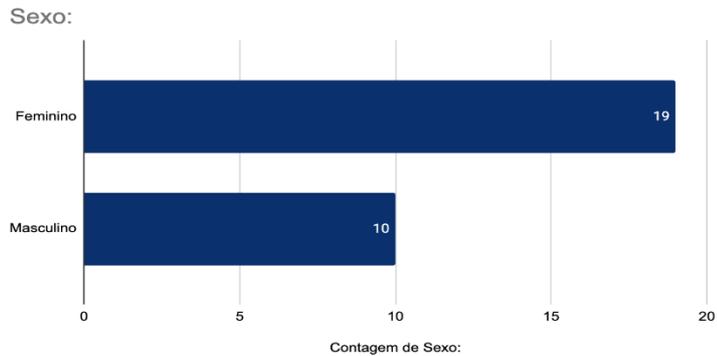


Figura 48 –Perfil por sexo dos participantes

A Figura 49 apresenta a distribuição por estado civil, sendo 13 solteiros, 12 casados e 4 divorciados. Os 12 participantes conhecem uma criança com (qualquer) congênita malformação.

Estado Civil

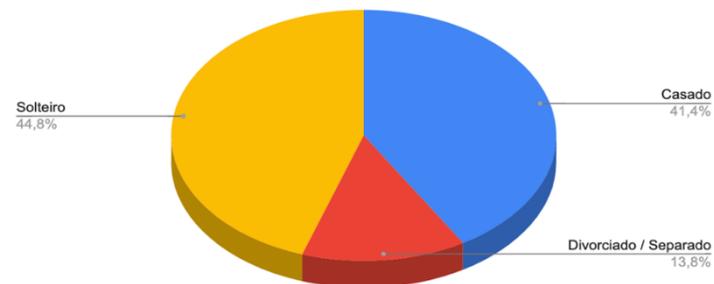


Figura 49 – Perfil por estado civil dos participantes

A Figura 50 apresenta o perfil por nível de escolaridade dos participantes da avaliação.

Escolaridade

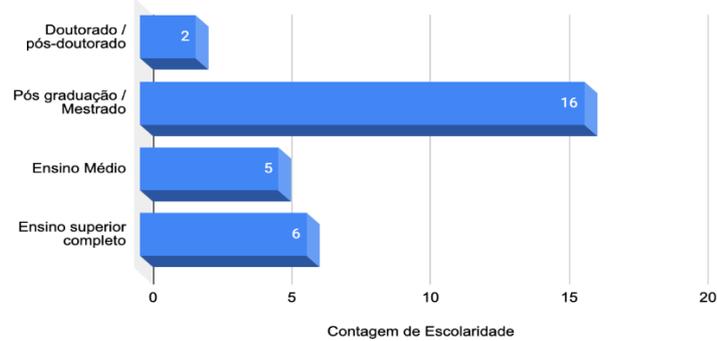


Figura 50 – Perfil nível de escolaridade dos participantes

Na Figura 51, constata-se que 15 participantes possuem nenhum filho (com o maior índice), sete tem 1 filho, 5 tem dois filhos e 2 tem três filhos.

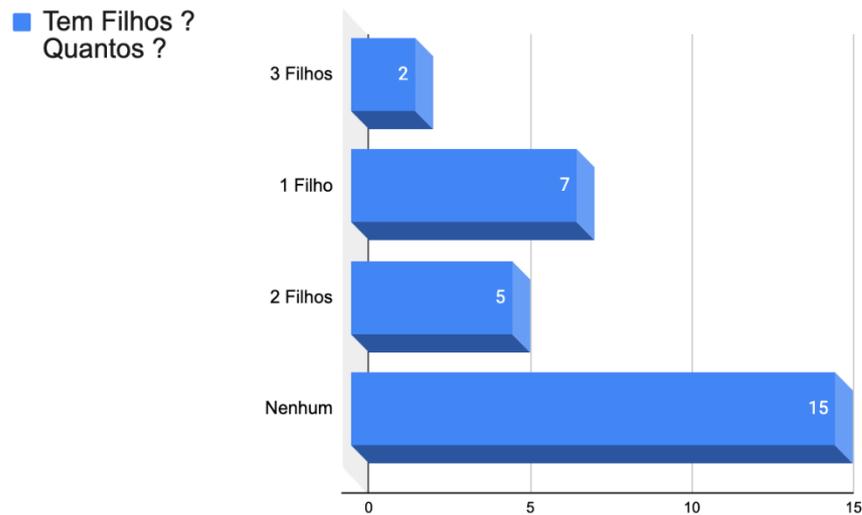


Figura 51 – Perfil dos participantes quanto ao número de filhos

A Figura 52 mostra que 16 pessoas não conheciam nenhuma criança com malformação congênita, 11 pessoas disseram que sim e 2 pessoas ficaram em dúvida.

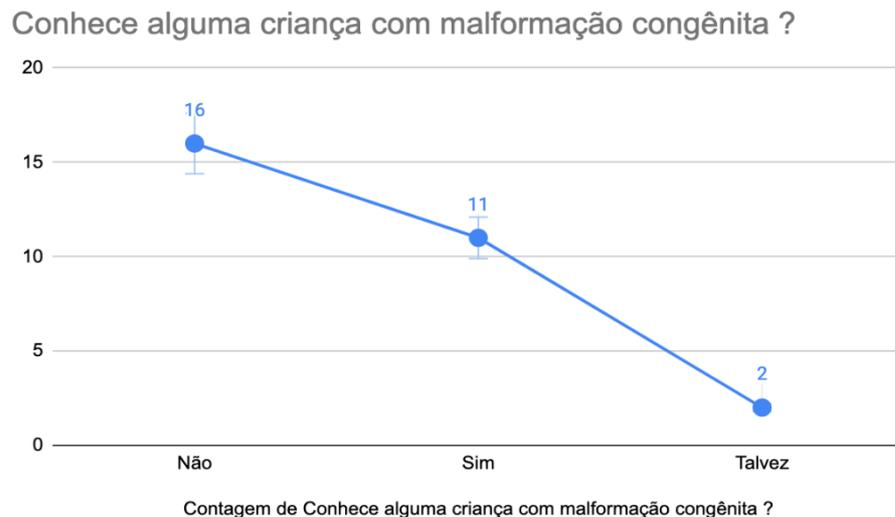


Figura 52 – Resultado do perfil: conhece alguma criança com malformação congênita? .

Os seguintes passos foram realizados para execução desta avaliação:

1. Os participantes foram recrutados em redes sociais (conforme supracitado).
2. Antes de iniciar a avaliação tiveram que concordar com um termo de consentimento para a avaliação online.
3. Na sequência, os participantes interagiram com *dashboard*, sem limitações de tempo, ou seja, o tempo que eles sentissem necessário explorar a interface para conhece-la.

4. Após explorar a interface, foram direcionados a formulário de avaliação no Google forms disponível em  
([https://docs.google.com/forms/d/e/1FAIpQLSfagYDUUnw3DoIunp5NDyJBn3Ur3RwfInVSOs8FBVjpa38lJ5w/viewform?usp=sf\\_link](https://docs.google.com/forms/d/e/1FAIpQLSfagYDUUnw3DoIunp5NDyJBn3Ur3RwfInVSOs8FBVjpa38lJ5w/viewform?usp=sf_link))
5. Por fim, foi realizada a síntese dos resultados quantitativos e qualitativos descritos a seguir.

## 5.2 RESULTADOS DA AVALIAÇÃO

Esta seção apresenta os resultados da avaliação, incluindo questões de múltipla escolha e questão em aberto. Os resultados da Figura 53 apresenta opiniões positivas referente as duas primeiras perguntas da avaliação realizada. A questão 1 tem a maior porcentagem em 51,7% em concordo e 10,3% em concordo totalmente em relação a vontade de utilizar com frequência. Já a questão 2 na questão 2, 44,8% assinalou discordo e 13,8% assinalou discordo totalmente em relação às visualizações serem desnecessariamente complexas.

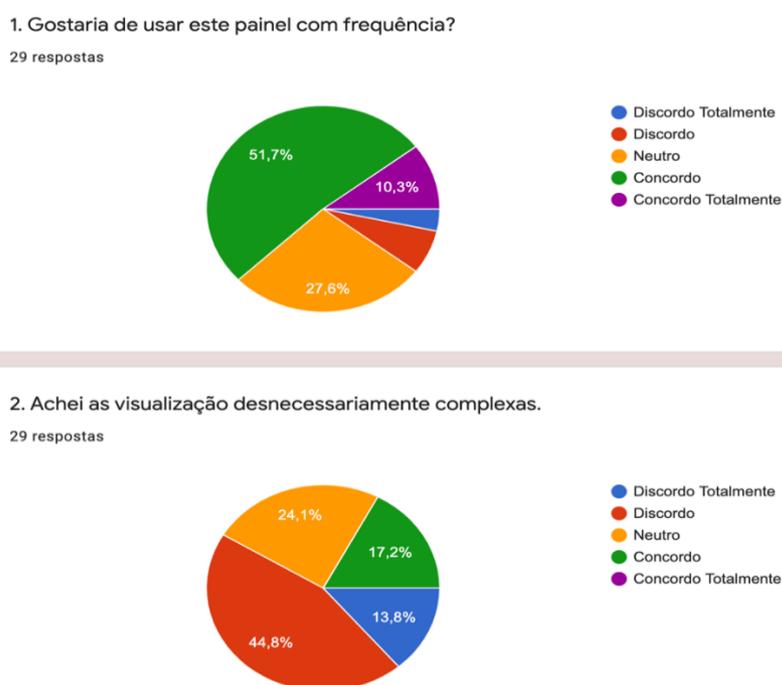


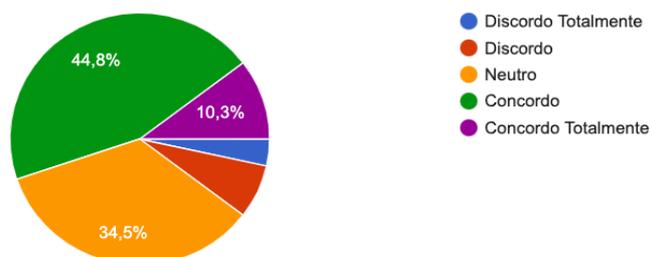
Figura 53 – Resultados da avaliação de Usabilidade (Formulário Google 1 a 2 pergunta).

A Figura 54 destaca que na questão 3, 44,8% responderam que concordam e 10,3% concordam totalmente que a visualização é fácil de usar. Já na questão 4, 44,8% responderam

que discordam e 3,4 % que discordam totalmente que seria necessário apoio de um técnico. Na questão 5, 65%, dos entrevistados concluíram que as visualizações estavam bem integradas.

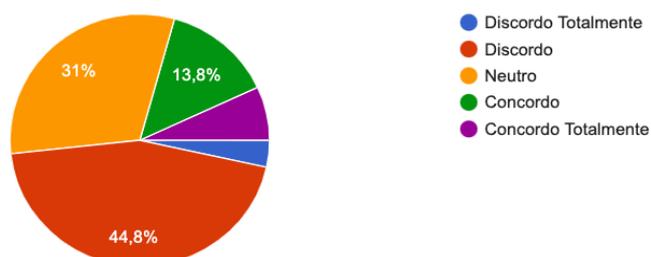
3. Achei o Painel de Visualização fácil de usar.

29 respostas



4. Achei que seria necessário o apoio de um técnico para poder usar e concluir as análises.

29 respostas



5. As funções destas visualização estavam bem integradas

29 respostas

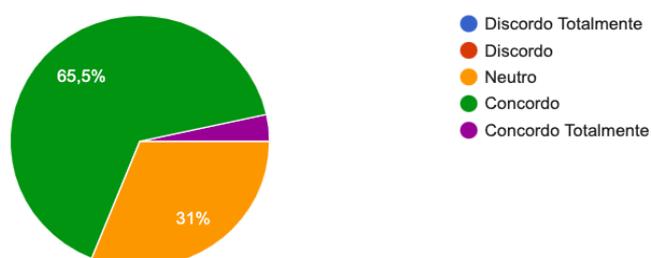


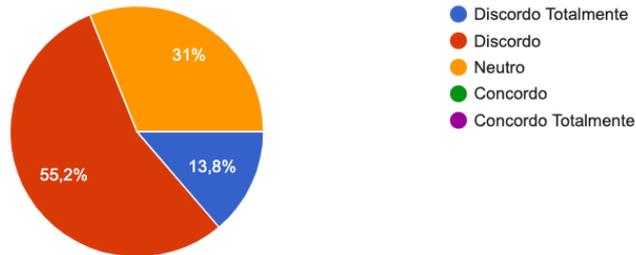
Figura 54 – Resultados do Teste Usabilidade (Formulário Google 3 a 5 pergunta).

A Figura 55 apresenta que na questão 6, 55,2% quando discordam que achou a visualização de dados muito inconsistente. Na questão 7, 41,4% concordam e 3,4 % concordam

totalmente que as pessoas aprenderiam visualizar rapidamente. Na questão 8, 58,6% discordaram que a visualizações são muito difíceis de usar.

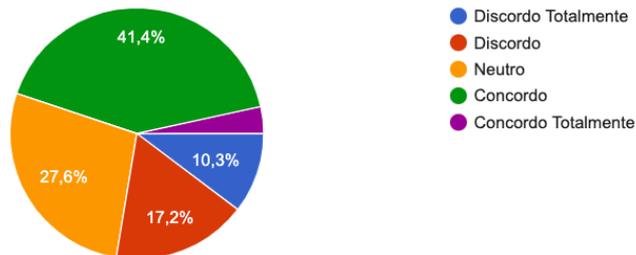
6. Achei as visualização dos dados muito inconsistente

29 respostas



7. Imagino que a maioria das pessoas aprenderiam visualizar este gráficos rapidamente.

29 respostas



8. Achei as visualização muito complicado de usar

29 respostas

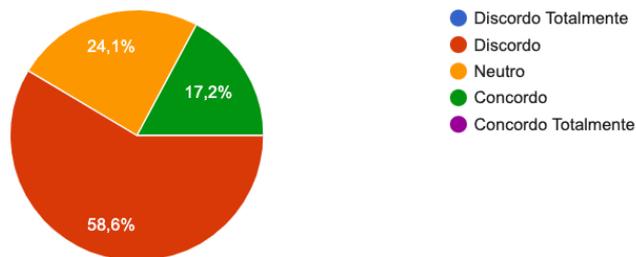


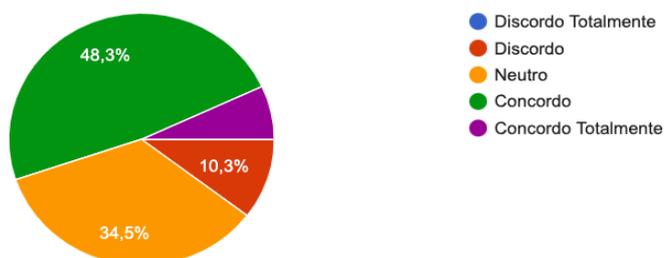
Figura 55 – Resultados do Teste Usabilidade (Formulário Google 6 a 8 pergunta).

A Figura 56 apresenta que na questão 9, o maior índice foi 48,3% onde o publico concorda que sente muito confiança em usar as visualizações, na questão 10 o maior índice foi de 48,3% onde o publico discorda em precisar aprender visualização gráfica para depois

usar o painel e na questão 11 o maior índice foi 55,2% onde o público se sente confortável em usar as visualizações.

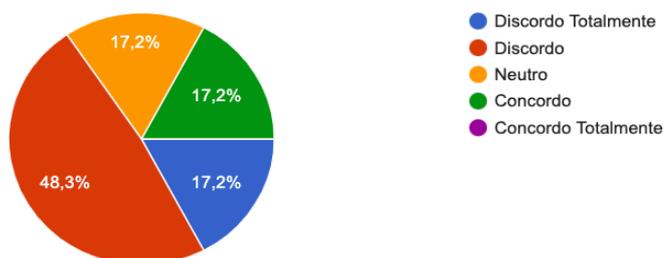
9. Eu me senti muito confiante com o visualização.

29 respostas



10. Eu preciso aprender visualizar gráfica antes de continuar usando este painel.

29 respostas



11. Eu me senti confortável com esta visualização.

29 respostas

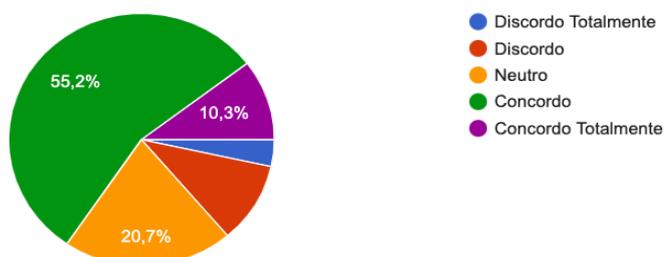


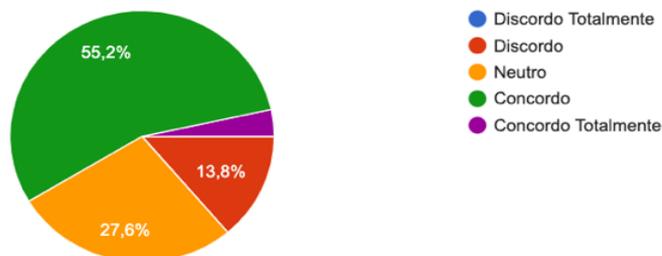
Figura 56 – Resultados do Teste Usabilidade (Formulário Google 9 a 11 pergunta).

A Figura 57 apresenta que na questão 12, 55,2% concordam e 3,4% concordam totalmente que foi fácil encontrar a informação necessária. Já na questão 13, 51,7% concordam

e 3,4% concordam totalmente que gostaria de utilizar o painel. Na questão 14, 65,5% concordam e 6,9% concordam totalmente que as interfaces dos gráficos são agradáveis.

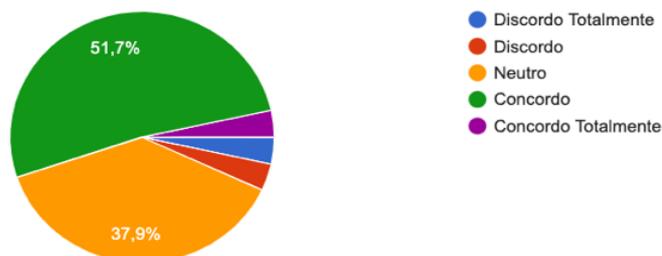
12. Foi fácil encontrar a informação que eu precisava

29 respostas



13. Eu gostei de usar a interface do Painel.

29 respostas



14. A Interface do gráficos é agradável

29 respostas

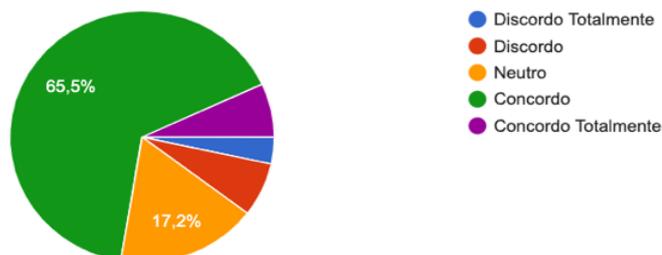


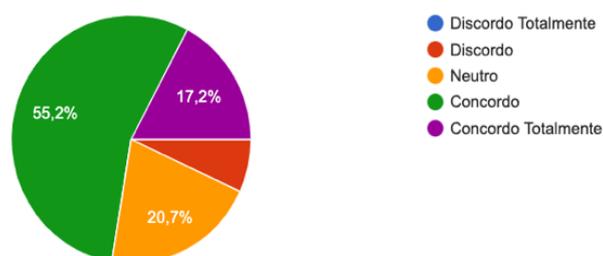
Figura 57 – Resultados do Teste Usabilidade (Formulário Google 12 a 14 pergunta).

A Figura 58 apresenta que na questão 15, 55,2% concordam e 17,2% concordam totalmente que a organização do painel está clara. Já na questão 16, 44,8% discordaram que

tiveram dificuldades com os mapas; enquanto que na questão 17, 69% concordam e 3,4% concordam totalmente que conseguem visualizar os estados.

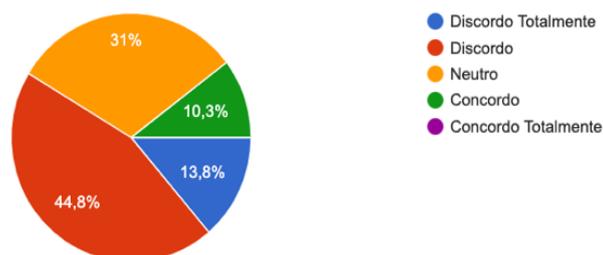
15. A organização de informações na tela do painel é clara

29 respostas



16. A figura do mapa: os estados brasileiros que apresentam as análises teve dificuldade na visualização no mapa ?

29 respostas



17. A figura do mapa: Consegue visualizar os dados de cada estados nas Análise Gerais?

29 respostas

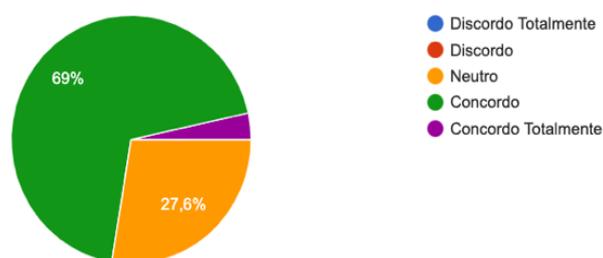
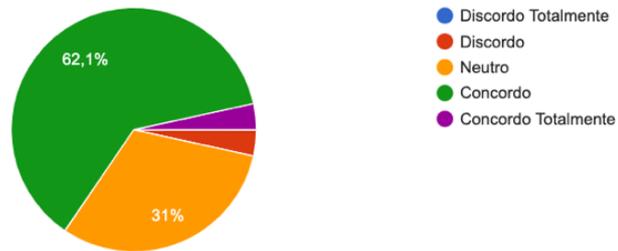


Figura 58 – Resultados do Teste Usabilidade (Formulário Google 15 a 17 pergunta).

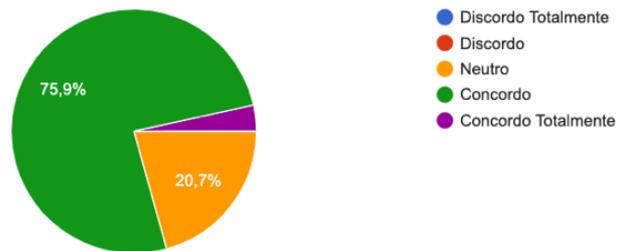
A Figura 59 apresenta que na questão 18, 62,1% concordam e 3,4% concordam totalmente que conseguem visualizar os dados de cada estado na análise sobre a mãe. Já na questão 19, 75,9% concordam e 3,4% concordam totalmente que conseguem visualizar os

dados de cada estado na análise sobre a gestação; entanto, na questão 20, 75,9% concordam e 6,9% concordam totalmente que conseguem visualizar os dados de cada estado na análise sobre o recém-nascido.

18. A figura do mapa: Consegue visualizar os dados de cada estados nas análise da mãe?  
29 respostas



19. A figura do mapa: Consegue visualizar os dados de cada estados nas análise da gestação?  
29 respostas



20. A figura do mapa: Consegue visualizar os dados de cada estados nas análise da recém-nascido?  
29 respostas

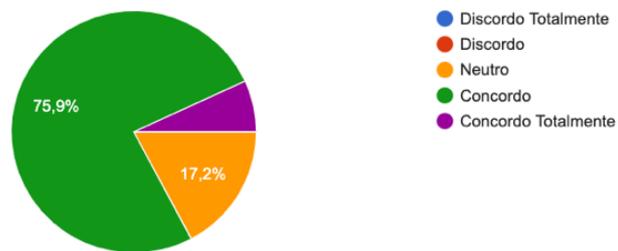


Figura 59 – Resultados do Teste Usabilidade (Formulário Google 18 a 20 pergunta).

A Figura 60 apresenta resposta a questão aberta sobre sugestões gerais. Dos 29 participantes, 8 sugeriram melhorias. Em geral, as sugestões indicam a necessidade de melhorar

as visualizações e fornecer novos filtros. Os participantes também expressaram a necessidade de acessibilidade e soluções mais minimalistas (sintéticas). O uso de cores e vocabulários adaptados foram propostos como alternativas para melhorar a interface. Um participante também apontou dificuldades na visualização das informações dos painéis em dispositivos móveis.

Sugestão Gerais:

8 respostas

Acredito que por eu visualizar no celular, não é a melhor opção para analisar. Seria melhor visualizar os dados impressos no papel.

Eu tenho sugestões de visualizações melhores para os olhos, mas não sei se no tableau é possível. Também não sei se eu sou parâmetro pq conheço de estatística etc. Acho que vale a pena mandar para gestor, por exemplo.

Bem apresentado. Como sugestão, utilizar cores mais distintas para facilitar a visualização, o que seria interessante,, principalmente, quando necessário abrir várias abas para cruzar dados (por mais que tenha gostado muito da paleta de cores utilizada). Um datasus apresentado dessa maneira seria maravilhoso! Parabéns pelo trabalho

acho que seria bom por um filtro de busca por cidade, ou permitir escolher uma cidade x para saber as informações respectivas

Pode-se pensar em mostrar proporção quando mostra quantidade. Por exemplo, SP e RJ podem sempre ter maior quantidade mas proporcionalmente pode não representar tanto em relação a outras cidades, pode usar isso para poder analisar algum surto em algumas regiões.

Filtrar o público alvo. É preciso um vocabulário mais direto. Perguntas mais Talhadas aonde a resposta seja sim não talvez nunca. Na prática a Dona Maria q nunca fez faculdade precisa ler e entender. Essa é a clareza q sugiro, pois é um assunto que abrange classe baixa, média e alta. Bjs Tati. Att.

Muita informação num mesmo slide. Acredito que deveria ser mais sintético

Permitir ao usuário colocar um esquema de cores diferente, pensando em daltônicos ou pessoas que tenham dificuldade com cores quentes. Também sugiro a possibilidade de aumentar o tamanho da fonte (procurei e não encontrei). As informações estão ótimas e muito bem organizadas, pensaria apenas em adicionar os requisitos de acessibilidade web no dashboard.

Figura 60 – Resultados do Teste Usabilidade (Formulario Google: Sugestões Gerais)

### 5.3 SÍNTESE E DISCUSSÃO

Os resultados apontam positivamente para a qualidade das visualizações e *dashboards* construídos com o método proposto. De maneira geral, foram obtidas boas pontuações em relação aos critérios de usabilidade, bem como os participantes se mostraram propensos a utilizar os *dashboards* e visualizações georeferenciadas.

Em síntese, a maioria dos participantes afirmam que usariam os painéis, que os painéis são fáceis de usar e que estão bem integrados. Também concordam que aprenderiam rapidamente, declaram sentir confiança, estarem confortável, encontrarem informações facilmente e consideraram a organização da informação clara. A maioria também gostou de usar os *dashboards* e consideraram sua interface agradável. Já em relação às interfaces em mapas, a maioria considera fáceis de interpretar nas quatro perspectivas (geral, mãe, gravidez e recém-nascido).

Com relação às perguntas negativas, a maioria discorda que as visualizações são desnecessariamente complexas, que o apoio de um especialista é necessário, que as visualizações são inconsistentes e que as visualizações são complicadas de usar. A maioria também discorda de dificuldades na interpretação dos valores sobre os estados nas interfaces baseadas em mapas.

Embora seja difícil estabelecer uma relação de causa efeito, a abordagem de utilizada no processo de design, ou seja, o UCD, pode ter contribuído para esses resultados ao prover uma visão das necessidades dos usuários e o contexto onde o sistema está inserido. Para tanto, foi necessário o acompanhamento das atividades desempenhadas pelos profissionais, entrevistas com os interessados e criação de protótipos. Isso torna o processo menos ágil, sendo necessário mais esforços para envolvimento dos usuários, entendimento das suas necessidades e coleta de *feedback*.

Portanto, pode-se citar como limitante o tempo de implementação do sistema, visto que dependendo do tamanho da equipe, dos perfis dos usuários e dos diferentes níveis de visualização de dados pode ser necessário fazer uma grande quantidade de protótipos, entrevistas, questionários e análises. Sendo assim, pode ser mais rápida a utilização de outros métodos, como por exemplo testes de usabilidade de usuário somente na entrega e a avaliação heurística restrita aos especialistas. Como destacado por Dixit *et al.* (2020), em algumas situações específicas, a abordagem UCD não é possível de ser utilizada quando os dados precisam ser visualizados no mesmo dia (ou seja, existe uma demanda de visualização urgente). Para tanto é sugerida uma abordagem mais leve, sendo então melhor ter uma visualização básica das informações ao invés de nenhuma. O balanceamento entre recursos necessários para execução de um processo e a qualidade dos resultados obtidos é um tema que necessita pesquisas em longo prazo, principalmente quando esses processos envolvem atividades com usuários.

Outro aspecto a ser destacado, é o balanceamento entre a exposição de dados mais precisos (e complexos) e visualizações voltadas a informar a população. Os *dashboards* construídos se destinam a um público amplo, que possui diferentes interesses e formações. Para tanto, foi abordada a estratégia de incluir visualizações alternativas, onde o usuário pode ter uma visão geral sobre os dados, bem como obter mais dados mais técnicos tais como a distribuição por anomalia. A visualização de dados absolutos versus relativos/incidência (de acordo com regiões e demais variáveis) também teve que ser conciliado. A exemplo de outros *dashboards* recentemente construídos para informar o público (e.g., John Hopkings (2020) e Unicamp (2020)), optou-se por apresentar com destaque inicial aos números absolutos, mas também possibilitar a visualização informações relativas, incidência e informações técnicas destinadas para usuários mais avançados.

É importante destacar que também foram enfrentados desafios no momento de aquisição de dados para análise, bem como dificuldades observadas na construção dos sistemas de visualização. Tal constatação está em linha com Dixit *et al.* (2020) que mencionam a complexidade envolvida no caso de crescimento dos sistemas, colocando o armazenamento, junção e processamento dos dados como um desafio. Como outros obstáculos, em linha com Hartzler *et al.* (2015), destacamos que as ferramentas não estão integradas, e com isso dificultam a obtenção dos dados, causam problemas de segurança que podem ocorrer no compartilhamento dos dados e geram a dificuldades de apresentar diversos tipos de visualizações para vários perfis de usuário. Algumas limitações são decorrentes das tecnologias empregadas, tais como o tableau. O uso de várias ferramentas com diferentes tecnologias e funcionalidades na elaboração dos *dashboards* também deve ser considerado.

Uso em logo prazo também deve ser analisado. Técnicas de avaliação em campo, no qual seja possível criar estatísticas consistentes com um número maior de usuários, bem como possibilitar a análise qualitativa do uso na prática podem contribuir com novos aprimoramentos aos *dashboards* construídos. Os resultados práticos como identificação de tendências e descoberta de conhecimento em longo prazo também devem ser analisados.

## 6 CONCLUSÃO

O *design* de soluções de visualização de grandes massas de dados direcionadas a um grande volume de usuários vem sendo um desafio da IHC. Embora a literatura na área de IHC tem apresentado diversos processos baseados em UCD há mais de 30 anos, ainda existem lacunas em aberto no que diz respeito a considerar as necessidades de aplicações que englobam amplo volume e variedades de usuários. Tais lacunas são ainda mais evidentes quando se considera novas tecnologias e desafios relacionados a visualização de grandes bases de dados complexos. Soluções de *dashboards* são empregadas para possibilitar a visualização integrada e prover meios para extrair conhecimento dessas bases, bem como apoiar a decisão. Os processos de *design* baseados em UCD para *dashboards* são espaços, e ainda mais raros ao considerar soluções voltadas à saúde. Sendo essas normalmente voltadas aos profissionais de saúde ou a um perfil específico de paciente.

No âmbito da saúde pública, os sistemas computacionais são utilizados, por exemplo, na previsão de epidemias, na identificação de epicentros de doenças, assim como podem ser empregados na análise e previsão de malformações congênitas. O *design* adequado de soluções de visualização é essencial para que a grande quantidade de informações armazenadas em sistemas computacionais seja corretamente interpretada por profissionais de saúde, gestores e pacientes.

Várias técnicas podem ser usadas para visualizar informações de saúde, por exemplo, mapas de calor e gráficos (bolhas, barras, dispersão, entre outros). Porém, essas formas de visualização em geral apresentam as informações de forma estática, não permitindo uma navegação dinâmica, onde os dados podem ser filtrados de acordo com as necessidades dos usuários.

Esta dissertação propõe o uso de UCD em um processo de design de *dashboards* para área de saúde. Um conjunto de *dashboards* sobre malformação congênita é proposto com uso de técnicas de visualização de informação, de mineração e tratamento de dados, bem como e princípios de UCD e representação de informação georreferenciada. Visou-se construir *dashboards* intuitivos e flexíveis, com diversos filtros e visões alternativas sobre a base de dados SINASC, de abrangência nacional. Nosso objetivo é permitir que usuários com diferentes níveis de conhecimento de informática e do domínio possam usar as informações.

Os *dashboards* construídos foram avaliados por 29 pessoas com diferentes perfis. Os usuários experimentaram o uso dos *dashboards* e responderam questões relacionadas a usabilidade baseadas no SUS. Os resultados positivos e apontam para a viabilidade da proposta, bem como possíveis aprimoramentos.

A seção 6.1 apresenta as principais contribuições desta proposta, enquanto a seção 6.2 apresenta os futuros e, por fim, a questão 6.3 lista as publicações relacionadas a esta dissertação até a presente data.

## 6.1 CONTRIBUIÇÕES

Esta dissertação apresenta contribuições para as áreas de IHC e Informática Médica, mais especificamente com proposta de processo de *design* baseado em UCD para visualização de informações de grandes volumes de dados, mas também com soluções de *dashboards* para visualização de malformação congênita.

Com o objetivo de responder à questão de pesquisa que norteia este trabalho “*Como realizar o design de soluções de visualização para informar a população e especialistas com o uso de dados minerados sobre bases de malformação congênitas?*”, foi realizado proposta de processo de *design*, elaboração de uma solução com este processo e avaliação dos resultados com base no sistema desenvolvido. Assim, a seguir são destacadas as contribuições obtidas nesta dissertação:

- *Revisão de técnicas e estado da arte.* Esta dissertação apresenta (capítulos 2 e 3) revisão sobre conceitos e técnicas de mineração e design de *dashboards*. Tal revisão abrange soluções de *dashboards* para saúde que utilizam UCD.
- *Proposta de processo de design.* Esta dissertação propõe e experimenta processo de design para *dashboards* destinados a sintetizar grandes volumes de dados e a serem utilizados por público amplo de usuários. Tal processo utiliza UCD em conjunto com prototipação, técnicas de ETL (para manipulação de grandes massas de dados) e análise estatística dos dados (visando selecionar variáveis e informações relevantes ao domínio). Uma versão inicial deste processo de *design* é apresentada em artigo publicado em conferência internacional (Almeida, Rosa e Bonacin, 2021).
- *Instanciação do processo no contexto de malformação congênita.* Esta dissertação apresenta como o processo de *design* foi instanciado no contexto de visualização de

dados de malformação congênita. As etapas da instanciação do processo de design, incluindo execução do ETL, o design dos *dashboards* e a análise estatística são detalhados nesta dissertação.

- *Desenvolvimento de conjunto de dashboards para malformação congênita.* Esta dissertação também apresenta um conjunto de *dashboards* e visualizações georeferenciadas construídas como resultado da instanciação do processo. Versões iniciais (que não incluem *dashboards* apresentados da subseção 4.3.3) foram publicados inicialmente em workshop local (Almeida e Bonacin, 2020) e posteriormente, versão revisada, em conferência internacional (Almeida, Rosa e Bonacin, 2021).
- *Avaliação de dashboards para malformação congênita.* Esta dissertação apresenta resultados de avaliação de usabilidade dos *dashboards* construídos com por 29 pessoas de perfil diversos. Uma avaliação preliminar com um número menor de usuários foi apresentada em conferência internacional (Almeida, Rosa e Bonacin, 2021).

## 6.2 TRABALHOS FUTUROS

Com o objetivo de expandir a solução de *dashboards* apresentada, pretende-se desenvolver mecanismos para visualização de resultados da aplicação de técnicas de aprendizado de máquina, de modo a prever tendências e identificar fatores contribuintes. Objetiva-se também expandir as bases de dados para outras fontes além da SINASC, incluindo o estudo de integração de dados e normalização.

Do ponto de vista de aprimoramento do processo, como próximos passos se propõe investigar o uso de técnicas de *design* participativo e *design thinking*, de modo a gerar ideias e soluções para *dashboards* flexíveis e adaptados ao público alvo.

Também se planeja a avaliação em longo prazo por profissionais de saúde, gestores e pacientes, e com banco de dados constantemente atualizado (*online*).

## 6.3 PUBLICAÇÕES OBTIDAS E PLANEJADAS

Como resultado desta pesquisa foram elaborados 2 artigos, um aceito, publicado e premiado em evento local e outro aceito e enviado para publicação em evento internacional. Também é planejada (em elaboração) publicação para ser submetida para revista internacional com

resultados ainda não publicados tais como: revisão do processo de design, detalhamento da instanciação do processo, *dashboards* com base em estatísticas e avaliação de usabilidade (completa) com usuários.

Segue abaixo lista de publicações:

- Workshop de Computação da UNIFACCAMP (WCF) 2020, evento promovido pelo Centro Universitário Campo Limpo Paulista (Unifaccamp).  
Titulo: Estudos sobre Métodos de Visualização para Predição de Malformações Congênitas  
Premiado como melhor artigo do WCF 2020.
- Conferência HCI International 2021 (Conference Management System).  
Titulo: *Designing Data Visualization Dashboards to Support the Prediction of Congenital Anomalies*  
Envida para publicação no Springer LNCS (Lecture Notes in Computer Science).

## REFERÊNCIAS BIBLIOGRÁFICAS

AGNES, J.; CHAVES, R.; KIPPER, L.E. Análise dos formulários de satisfação dos clientes de uma empresa do setor tecnológico a luz dos conceitos da usabilidade. **Tecno-Lógica**. 19. 51. 10.17058/tecnolog.v19i1.5670. 2015.

AHERN, D. K. *et al.* E-health blood pressure control program. **J. Med. Pract. Manage**, v. 28, n. 2, p. 91-100, 2012.

ALMEIDA, T. A.; ROSA, F.F.; BONACIN, R. **Designing Data Visualization Dashboards to Support the Prediction of Congenital Anomalies**. 23RD INTERNATIONAL CONFERENCE ON HUMAN-COMPUTER INTERACTION, 2021. 20 p. In Press.

ALMEIDA, T. A. ; BONACIN, R. . **Estudos sobre Métodos de Visualização para Predição de Malformações Congênitas**. Anais do XVI Workshop de Computação da UNIFACCAMP, 2020. p. 73-78.

ANACLETO, J. and S. Fels, Lessons from ICT design of a healthcare worker-centered system for a chronic mental care hospital, in **Proc. CHI, ACM**, Apr 26 – May 01 2014, Toronto, ON, Canada, pp. 1687-1692. 2014.

ANDRE, M. *et al.* Transmission network analysis to complement routine tuberculosis contact investigations. **Am J Pub Health**, v. 97, n. 3, p. 470-477, 2007.

AVRUSKIN, G. A. *et al.* Visualization and exploratory analysis of epidemiologic data using a novel space time information system. **Int J Health Geogr**, v. 3, n. 1, p. 26, 2004.

BALLEGAARD, S. A. T. R., Hansen, M. Kyng, Healthcare in everyday life - Designing healthcare services for daily life, in Proc. 2008 CHI, ACM, April 5-10, Florence, Italy, pp. 1807-1816. 2008.

BARBOSA, S.; SILVA, B.. **Interação humano-computador**. Rio de Janeiro: Campus, 2010.

BASTOS, M. S.; DE MACEDO, R. M. G.. Prevenção de malformações congênitas. **IV Mostra Interna de Trabalhos de Iniciação Científica do Cesumar**, Maringá, 2008.

BELUZO, C. E., Pimentel, L. R., Alves, L. C., & de Carvalho, T. J. (2020). **Big Data Visualization Methods Applied in the Context of Neonatal Mortality**. *Anais do Computer on the Beach*, 11(1), 592-595.

BISPO, C. A. F.; CAZARINI, E. W.. A evolução do processo decisório. *In: ENCONTRO NACIONAL DE ENGENHARIA DE PRODUÇÃO*, 1998, Niterói. **Anais...** Niterói: UFF/TEP, 1998.

BOS, L.; BLOBEL, B. Semantic interoperability between clinical and public health information systems for improving public health services. **Med Care Compunetics Four**, n. 4, p. 256, 2007.

BRASIL ESCOLA. **Região Nordeste**. 2020a Disponível em: <<https://brasilecola.uol.com.br/brasil/regiao-nordeste.htm>>. Acesso em: 09 nov 2020.

BRASIL ESCOLA. **Região Sudeste**. 2020b Disponível em: <<https://brasilecola.uol.com.br/brasil/regiao-sudeste.htm>>. Acesso em: 09 nov. 2020.

BRENNAN PF, D. S, Casper, G. Project HealthDesign: Rethinking the power and potential of personal health records. **J Biomed Inform**. 43(5):S3-5. PMID:20937482. 2010.

BROOKE, J. SUS - A quick and dirty usability scale. **Usability Eval Ind [Internet]**. 1996;189(194):4–7. Available from: <http://hell.meiert.org/core/pdf/sus.pdf> acessado 02 Dezembro de 2020.

BROWN, T. A. **Genética, um enfoque molecular**. 3. ed. Rio de Janeiro: Guanabara Koogan. 1999.

CARROLL, L. N.; AU, A. P.; DETWILER, L. T.; FU, T.; PAINTER, I.S.; ABERNETHY, N. F. Visualization and analytics tools for infectious disease epidemiology: A systematic review. **Journal of Biomedical Informatics**, n. 51, p. 287-298, 2014.

CHAN W. Increasing the success of physician order entry through human factors engineering. **J Healthc Inf Manag**; 16 (1): 71–9. 2002.

CHEN, M.; EBERT, D.; HAGEN, H.; LARAMEE, R.; LIERE, R.; KWAN-LIU, M.; RIBARSKY, W.; SCHEUERMANN, G.; SILVER, D. Data, Information, and Knowledge in Visualization. **IEEE Computer Graphics and Applications**, v. 29, n. 1, 2009.

CHOO, C. W. **The knowing organizations**: how organizations use information to construct meaning, create knowledge, and make decisions. Nova Iorque: Oxford University, 1998.

CYBIS, W. A. **Engenharia de Usabilidade**: uma abordagem ergonômica. 2003. Apostila para o curso de Pós-Graduação em Engenharia de Produção - Laboratório de Utilizabilidade de Informática - Universidade de Santa Catarina: Florianópolis

COTTERMAN, W. W. Kumar, K. User Cube: A Taxonomy of End-Users. **Communications of the ACM** 32: 11, 1313–1320. 1989.

DATAIKU. **Conecte-se a todas as suas fontes de dados**. Disponível em: <<https://www.dataiku.com/product/features/connectivity/>>. Acesso em: 12 ago. 2020.

DE CASTRO, M. L. S. *et al.* Frequência das malformações múltiplas em recém-nascidos na Cidade de Pelotas, Rio Grande do Sul, Brasil, e fatores sócio demográficos associados. **Cad. Saúde Pública**, Rio de Janeiro, v. 22, n. 5, p. 1009-1015, 2006.

DIXIT, Ram A., Hurst, S. A, Katharine T., Boxley, C., Lysen-Hendershot, K., Bennet, S., S. Booker, E. R., R. M. Rapid development of visualization dashboards to enhance situation awareness of COVID-19 telehealth initiatives at a multihospital healthcare system. **Journal of the American Medical Informatics Association**, 27(9), 1456–1461. 2020.

DOCUMENTAÇÃO completa do processo de ETL dos dados do SINASC. Disponível em: <<https://bigdata-metadados.icict.fiocruz.br/dataset/sistema-de-informacoes-de-nascidos-vivos-sinasc/resource/9d664e65-2dbd-44cf-8f4a-65499affa27c>>. Acesso em: 12 ago. 2020.

DOWDING, D.; RANDELL, R.; GARDNER, P.; FITZPATRICK, G.; DYKES, P.; FAVELA, J.; HAMER, S.; WHITEWOOD-MOORES, Z.; HARDIKER, N.; BORYCKI, E.; CURRIE, L. C.. Dashboards for improving patient care: review of the literature. **International Journal of Medical Informatics**, n. 84, 2015.

EDUCA MAIS BRASIL. **Região Norte**. É a maior região do Brasil. Disponível em: <<https://www.educamaisbrasil.com.br/enem/geografia/regiao-norte>>. Acesso em: 09 dez. 2020.

ENDSLEY, MR. Measurement of situation awareness in dynamic systems. **Hum Factors**; 37 (1): 65–84. 1995.

ESCOLA E EDUCAÇÃO. **População da Região Sudeste do Brasil** – Habitantes, IDH, PIB, Estados. Disponível em: <<https://escolaeducacao.com.br/populacao-da-regiao-sudeste-do-brasil-habitantes-idh-pib-estados/>>. Acesso em: 09 dez. 2020.

ESTADÃO. **Região Sul**. Disponível em: <<https://economia.estadao.com.br/noticias/geral,regiao-sul-e-socialmente-a-mais-desenvolvida-imp,557299https://economia.estadao.com.br/noticias/geral,regiao-sul-e-socialmente-a-mais-desenvolvida-imp-,557299#:~:text=O%20Sul%20det%C3%A9m%20o%20maior,m%C3%A9dia%20de%20vida%20entre%20outros>>. Acesso em: 10 nov. 2020.

FEDERICO, P.; HEIMERL, H.; KOCH, S.. **A Survey on Visual Approaches for Analyzing Scientific Literature and Patents**. 2016. Disponível em: <[https://publik.tuwien.ac.at/files/PubDat\\_251231.pdf](https://publik.tuwien.ac.at/files/PubDat_251231.pdf)>. Acesso em: 19 maio 2020.

FOLORUNSO, O.; OGUNSEYE, O. Challenges in the adoption of visualization system: a survey. **Kybernetes**, v. 37, n. 9-10, p. 1530-1541, 2008.

FONTOURA, F. C. *et al.* Ansiedade de mães de recém-nascidos com malformações congênitas nos períodos pré e pós-natal. **Rev. Latino-Am. Enfermagem**, Ribeirão Preto, v. 26, e3080, 2018. Disponível em: <[http://www.scielo.br/scielo.php?script=sci\\_arttext&pid=S0104-11692018000100380&lng=en&nrm=iso](http://www.scielo.br/scielo.php?script=sci_arttext&pid=S0104-11692018000100380&lng=en&nrm=iso)>. Acesso em: 10 abr. 2020.

FREIFELD, C. C. *et al.* HealthMap: global infectious disease monitoring through automated classification and visualization of Internet media reports. **J Am Med Inf Assoc**, v. 15, n. 2, p. 150-157, 2008.

FREITAS, C. M. D. S.; CHUBACHI, O. M.; LUZZARDI, P. R. G.; CAVA, R. A. Introdução à Visualização de Informações. **RITA**, v. VII, n. 2, 2001.

GOMES, A. S.; MELO, C. A.; ARCOVERDE, D. F.; SABIÁ, G. A. **O Ensino de Interação Humano-Computador com Ênfase na Formação de Competências Relacionadas à Inovação Tecnológica: Design da Interação**. Recife: UFPE. 2007.

GOMES, P. C. T.. **O que é um dashboards?** O guia completo e definitivo! 2017. Disponível em: <<https://www.opservices.com.br/o-que-e-um-dashboard/>>. Acesso em: 09 abr. 2020.

GHAZALI, M. A., Nurul, A. Mat. User Centered Design Practices in Healthcare: A Systematic Review. 3rd **International Conference on User Science and Engineering (i-USER)**. 2014.

HAY, S. I. *et al.* Global mapping of infectious disease. **Philos Trans Roy Soc B BiolSci**, v. 368, n. 1614, 2013.

HARTZLER, A. L.; Chaudhuri, S.; Fey, B. C.; Flum, D. R.; Lavalley, D.. Integrating Patient-Reported Outcomes into Spine Surgical Care through Visual Dashboards: Lessons Learned from Human-Centered Design. **eGEMs (Generating Evidence & Methods to improve patient outcomes)**: Vol. 3: Iss. 2, Article 2. 2015.

ICICT. **Ciência de dados aplicada à saúde**. Disponível em: <<https://www.icict.fiocruz.br/ci%C3%Aancia-de-dados-aplicada-%C3%A0-sa%C3%Bade>>. Acesso em: 10 abr. 2020.

ISO 9241-210:2010. Ergonomics of human-system interaction -- **Part 210: Human-centred design for interactive systems**. Disponível em <[http://www.iso.org/iso/catalogue\\_detail.htm?csnumber=5207](http://www.iso.org/iso/catalogue_detail.htm?csnumber=5207)>. Acessado em 15 de abril de 2021.

JOHNSON, C. M., Zhang J. A user-centered framework for redesigning health care interfaces. *J Biomed Inform.* 38(1):75-87. 2005.

JOHN'S, H.. **Coronavirus Resource Center**. Disponível em: <<https://coronavirus.jhu.edu/us-map>>. Acesso em: 09 jun. 2020.

JOLINS, J.; ANCUKIEWICZ, M.; DELONG, E.; PRYOR, D.; MUHLBAIER, L.; MARK, D. Discordance of databases designed for claims payment versus clinical information systems: implications for outcomes research. **Ann. Intern. Med.**, n. 119, p. 844-850, 1993.

KHAIRAT, S. S.; DUKKIPATI, A.; LAURIA, H. A.; BICE, T.; TRAVERS, D.; CARSON, Sh. S. The Impact of Visualization Dashboards on Quality of Care and Clinician Satisfaction: Integrative Literature Review. **JMIR Hum Factors**, v. 5, n. 2, e22, 2018.

KHAN, A. S. *et al.* The next public health revolution: public health information fusion and social networks. **Am J Pub Health**, v. 100, n. 7, p. 1237-1242, 2010.

KOCH, S. H.; WEIR, C.; WESTENSKOW, D.; GONDAN, M.; AGUTTER, J.; HAAR, M.; *et al.* Evaluation of the effect of information integration in displays for ICU nurses on situation awareness and task completion time: A prospective randomized controlled study. **Int J Med Inform**, v. 82, n. 8, p. 665-675, 2013.

KOCH, T.; DENIKE, K. Crediting his critics' concerns: Remaking John Snow's map of Broad Street cholera, 1854. **Soc Sci Med**, v. 69, n. 8, p. 1246-1251, 2009.

LIU, H. G. Mining social media: A brief introduction. **Inform**s, 2012.

LIU, J.; TANG, T.; WANG, W, Xu, Bo; KONG, Xiangjie & XIA, Feng. A Survey of Scholarly Data Visualization. **IEEE Access**, 2017.

MANISH, J. G., em **Mobile Sensors and Context-Aware Computing**, 2017. Disponível: < <https://www.sciencedirect.com/topics/computer-science/usability> > Acesso em : 08 abr.2020

MCMENAMIN, J.; NICHOLSON, R.; LEECH, K. Patient Dashboard: the use of a colour-coded computerised clinical reminder in Whanganui regional general practices. **J. Primary Health Care**, v. 3, n. 4, p. 307-310, 2011.

MCNABB, L.; LARAMEE, R. S. How to Write a Visualization Survey Paper: A Starting Point. **Eurographics**, 2019.

MELO, L. R.; PELISSARI, W. R.. Estudo de caso sobre Business Intelligence como instrumento de apoio à tomada de decisão. **XI Ciclo de Estudos da Faculdade Cidade Verde**, 2016.

MENDES, I. C.; JESUINO, R. S. A.; PINHEIRO, D. S.; REBELO, A. C. S.. Anomalias congênitas e suas principais causas evitáveis: uma revisão. **Revista Médica de Minas Gerais**, n. 28, e-1977, 2018. Disponível em: <<http://www.rmmg.org/exportar-pdf/2329/e1977.pdf>>. Acesso em: 07 abr. 2020.

MUNDO EDUCAÇÃO. IDH do Brasil. Disponível em: <<https://mundoeducacao.uol.com.br/geografia/idh-brasil.htm#:~:text=O%20IDH%20do%20Brasil%20%C3%A9,humano%3B%20e%20baixo%20desenvolvimento%20humano>>. Acesso em: 09 nov. 2020.

NEDIGER, M.. **How to choose the best types of charts for your data**. 2019. Disponível em: <<https://venngage.com/blog/how-to-choose-the-best-charts-for-your-infographic/>>. Acesso em: 21 maio 2020.

NIELSEN, J. Ten Usability Heuristics. 1993. Disponível em: <https://www.nngroup.com/articles/ten-usability-heuristics/> acessado 10 abril de 2021

NIELSEN, J. Usability 101: **Introduction to usability** 2003. Disponível em: <https://www.nngroup.com/articles/usability-101-introduction-to-usability/> acessado 10 abril 2021

NIELSEN, J. **Usability Engineering**. Burlington: Morgan Kaufmann, 1993.

NOGUEIRA, P.; MARTINS, J.; RITA, F.; FATELA, L.. Dashboard da saúde: passado, presente e futuro. Uma perspectiva da evolução em Portugal. **Séries informação e análise**, ano 1, n. 2, 2017.

NSF. **National Bureau of Standards Research Reports**. 1987. Disponível em: <<https://digital.library.unt.edu/ark:/67531/metadc502656/>>. Acesso em: 17 maio 2020.

OUELLETE, C.. **5 Best Survey Data Visualization Tools**. 2019. Disponível em: <<https://optinmonster.com/best-survey-data-visualization-tools/>>. Acesso em: 21 maio 2020.

PABLATE, J. **The Effect of Electronic Feedback on Anesthesia Providers' Timely Preoperative Antibiotic Administration**. University of North Florida, 2009.

PAUL, M. J.; DREDZE, M.. You Are What You Tweet: Analyzing Twitter for Public Health. **Proceedings of the Fifth International AAAI Conference on Weblogs and Social Media**, 2011.

PELCHAT, D. Processus d'adaptation des parents d'un enfant atteint d'une déficience et élaboration d'un programme d'intervention précoce à leur intention. **Revue Canadienne de Santé Mentale Communautaire**, n. 1, p. 63-80, 1992.

PINO, I. **Desafios e tendências da visualização de dados**. DataViz Basics: 1 de 4. 2016. Disponível em: <<https://blog.dp6.com.br/desafios-e-tend%C3%Aancias-da-visualiza%C3%A7%C3%A3o-de-dados-dataviz-basics-1-de-4-8208cf5fbc6>>. Acesso em: 08 abr. 2020.

PNUD BRASIL. **Desenvolvimento Humano e IDH**. Disponível em: <<https://www.br.undp.org/content/brazil/pt/home/idh0.html>>. Acesso em: 09 nov. 2020.

PORTAL IBC. Disponível em: <<https://www.ibccoaching.com.br/porta/entenda-como-funciona-piramide-do-conhecimento-e-seus-quatro-componentes/>>. Acesso em: 06 ago. 2019.

RAM. **Fluxo ETL**. Disponível em: <[https://raw.githubusercontent.com/bigdata-icict/ETL-Dataiku-DSS/master/SINASC/ETL\\_divided.png](https://raw.githubusercontent.com/bigdata-icict/ETL-Dataiku-DSS/master/SINASC/ETL_divided.png)>. Acesso em: 07 ago. 2020.

REZENDE, Y. A falsa utilidade da biblioteca de empresa. **O Estado de São Paulo**, Caderno Empresas, p. 2, São Paulo, 25 mar. 1992.

RODRIGUES FILHO, J.; GOMES, N. P.. Tecnologia da Informação no Governo Federal. **RAP**, 2004.

SAKAKI, T.; OKAZAKI, M.; MATSUO, Y.. Earthquake Shakes Twitter Users: Real-Time Event Detection By Social Sensors. **Proceedings of the 19th international conference on Word Wide Web**, 2010.

SANTOS, A. P.. A Importância da Interação Humano-Computador. 2012. Disponível em: <<http://tiqx.blogspot.com/2012/02/compreenda-importancia-da-interacao.html>>. Acesso em: 05 dez. 2020.

SAS. **First Things First**. Highmark makes healthcare-fraud prevention top priority with SAS. 2006.

SCHEIDER, H. N. **Ergonomia das interfaces humano-computador como princípio de qualidade em Ead**. 2014. Disponível em: <<http://www.uece.br/endipe2014/ebooks/livro4/10.%20DID%C3%81TICA,%20TECNOLOGIAS%20E%20INOVA%C3%87%C3%95ES%20PEDAG%C3%93GICAS%20NA%20ESCOLA%20B%C3%81SICA.pdf>>. Acesso em: 06 dez. 2020.

SHAMMA, D. A.; KENNEDY, L.; CHURCHILL, E.. Tweet the Debates: Understanding Community Annotation of Uncollected Sources. **WSM'09 Proceedings of the first SIGMM workshop on social media**, 2009.

STAGGERS, A. G, N. A systematic review of the designs of clinical technology: Findings and recommendations for future research. **ANS Adv Nurs Sci**. 32(3):252-79. 2009.

TABLEAU. **A Tableau apresenta o Hyper, uma nova tecnologia de processamento de dados que oferece um desempenho de análise sem precedentes**. 10 jan. 2018. Disponível em: <<https://www.tableau.com/pt-br/about/press-releases/2018/tableau-launches-hyper-new-data-engine-technology-delivering-unprecedented>>. Acesso em: 12 ago. 2020.

TAN, P.; STEINBACH, M.; KUMAR, V. **Introduction to data mining – mineração de dados**. Rio de Janeiro: Ciência Moderna Ltda., 2009.

TENBU. **Vantagens e benefícios de visualizações de dados de qualidade**. 2019. Disponível em: <<https://tenbu.com.br/data-visualization-analytics/>>. Acesso em: 09 abr. 2020.

TENÓRIO, J. M. et al. **Desenvolvimento e Avaliação de um Protocolo Eletrônico para Atendimento e Monitoramento do Paciente com Doença Celíaca**. 2011. Disponível em:<<http://www.sbis.org.br/cbis11/arquivos/693.pdf>>. acessado 10 abril 2021

TOMAR, D.; AGARWAL, S.. A survey on Data Mining approaches for Healthcare. **International Journal of Bio-Science and Bio-Technology**, v. 5, n. 5, p. 241-266, 2013.

UNICAMP. **Unicamp cria mapa interativo para acompanhar casos de coronavírus dos seus hospitais**. 2020. Disponível em:

<<https://www.unicamp.br/unicamp/noticias/2020/05/04/unicamp-cria-mapa-interativo-para-acompanhar-casos-de-coronavirus-dos-seus>>. Acesso em: 20 maio 2020.

WERNER, W.; WERNER, I. A. Gestão do conhecimento: Ferramentas tecnológicas e portais do Conhecimento para empresas desenvolvedoras de tecnologias de médio e pequeno portes. **Revista Terra Viva**, 2014.

WRIGHT, MC, TAEKMAN, J.M, E. MR. Objective measures of situation awareness in a simulated medical environment. **Qual Saf Health Care** **2004**; 13 (suppl\_1): i65–71. 2004.

YANG, Y.; AKERS, L.; KLOSE, T.; YANG, C. B. Text mining and visualization tools - Impressions of emerging capabilities. **WorldPatent Inform.**, v. 30, n. 4, p. 280-293, 2008.

YOUNG, G. W. Kitchin, R, Naji, J.. Building City Dashboards for Different Types of Users. **Journal of Urban Technology**. 2020.

ZAYDFUDIM, V. *et al.* Implementation of a real-time compliance dashboard to help reduce SICU ventilator-associated pneumonia with the ventilator bundle. **Arch. Surg.**, p. 656-662, 2009.

## APÊNDICE A – Dicionário de dados do SINASC

### Campo com X ; parâmetros na construção da abordagem ( design)

Id	type	description
NUMERODN		Número do DN, sequencial por informante da UF e por ano
CODINST	texto	Código de instalação da geração dos registros.
ORIGEM	bigint	
NUMERODV	bigint	Número do Dígito Verificador
PREFIXODN	bigint	Prefixo da DN
CODESTAB	texto	Código de estabelecimento
CODMUNNASC ✘	bigint	Município de ocorrência, codificação idêntica a CODMUNRES, conforme tabela TABMUN
LOCNASC ✘	bigint	Local de ocorrência do nascimento, conforme tabela: 9: Ignorado 1: Hospital 2: Outro Estab Saúde 3: Domicílio 4: Outros
def_loc_nasc ✘	texto	Local de nascimento (Nominal, com as seguintes características: Hospital; outros produtos de saúde; Domicílio; Via pública; Outros; Ignorado)
IDADEMAE ✘	bigint	Idade da mãe em anos
ESTCIVMAE ✘	bigint	Estado civil, conforme tabela: 1: Solteira 2: Casada 3: Viúva 4: Separado judicialmente / Divorciado 5: União consensual (anteriores) 9: Ignorado
def_est_civil ✘	texto	Estado civil (Situação conjugal: Solteiro; Casado; Viúvo; Separado judicialmente / divorciado; União estável; Ignorado)

ESCMAE 	bigint	Escolaridade, anos de estudo concluídos: 1: Nenhuma 2: 1 a 3 anos 3: 4 a 7 anos 4: 8 a 11 anos 5: 12 e mais 9: Ignorado
def_escol_mae 	texto	Escolaridade da mãe (Nenhuma; de [poiu1 a 3 anos; de 4 a 7 anos; 8 a 11 anos; 12 anos e mais; Ignorado)
CODOCUPMAE 	texto	Ocupação, conforme Classificação Brasileira de Ocupações (CBO-2002)
QTDFILVIVO 	texto	Número de filhos vivos
QTDFILMORT 	texto	Número de filhos mortos
CODMUNRES 	bigint	Município de residência da mãe, em código idêntico ao CODMUNNASC, conforme tabela TABMUN
GESTACAO 	bigint	Semanas de gestação, conforme tabela: 9: Ignorado 1: Menos de 22 semanas 2: 22 a 27 semanas 3: 28 a 31 semanas 4: 32 a 36 semanas 5: 37 a 41 semanas 6: 42 semanas e mais
def_gestacao 	texto	Semana de gestação (Nominal, com as seguintes variações: menos de 22 semanas; 22 a 27 semanas; 28 a 31 semanas; 32 a 36 semanas; 37 a 41 semanas; 42 semanas ou mais)
GRAVIDEZ 	bigint	Tipo de gravidez, conforme tabela: 9: Ignorado 1: Única 2: Dupla 3: Tripla e mais
def_gravidez 	texto	Tipo de gravidez (Nominal, com as seguintes características: Única; Dupla; Tripla e mais e ignorada)
PARTO 	bigint	Tipo de parto, conforme tabela: 9: Ignorado 1: Vaginal 2: Cesáreo
def_parto 	texto	Tipo de parto (Nominal, com as seguintes características: Vaginal; Cesáreo; Ignorado)

CONSULTAS 	bigint	Número de consultas pré-natal: 1: Nenhuma 2: de 1 a 3 3: de 4 a 6 4: 7 e mais 9: Ignorado
def_consultas 	texto	Número de consultas durante o pré-natal (Nenhuma; de 1 a 3; de 4 a 6; 7 e mais; Ignorado)
DTNASC 	texto	Data do nascimento, no formato ddmmaaaa
data_nasc 	encontro	Data de nascimento
ano_nasc 	bigint	Ano do nascimento
dia_semana_nasc	texto	Dia da semana em que ocorreu o nascimento (dom; seg; ter; qua; qui; sex; sáb)
HORANASC	texto	Horário de nascimento
SEXO 	bigint	Sexo, conforme tabela: 0: Ignorado, não informado 1: Masculino 2: Feminino
def_sex0 	texto	Sexo (Nominal, com as seguintes listas: Masculino; Feminino; Ignorado)
APGAR1 	texto	Apgar no primeiro minuto 00 a 10
APGAR5 	texto	Apgar no quinto minuto 00 a 10
RACACOR 	bigint	Raça / Cor: 1: Branca 2: Preta 3: Amarela 4: Parda 5: Indígena
def_raca_cor 	texto	Raça / cor (Nominal, com as seguintes características: Branca; Preta; Amarela; Parda; Indígena)
PESO 	texto	Peso ao nascer, em gramas
IDANOMAL 	bigint	Anomalia congênita: 9: Ignorado 1: Sim 2: Não
def_anomalia 	texto	Anomalia congênita (Ignorado; Sim; Não)
DTCADASTRO	texto	Dados do cadastro do DN no sistema
CODANOMAL	texto	Código de malformação congênita ou anomalia cromossômica, de acordo com o CID-10

NUMEROLOTE	bigint	Número do lote
VERSAOSIST	texto	Versão do sistema
DTRECEBIM	texto	Dados de recebimento no nível central, dados do último upgrade de registro.
DIFDATA	texto	Diferença entre dados de óbito e dados do recebimento original do DO ([DTNASC] - [DTRECORIG])
DTRECORIGA	texto	Dados do 1o recebimento do lote, dados pelo Sisnet.
NATURALMAE	bigint	Se uma mãe para estrangeira, constar o código do país de nascimento.
CODMUNNATU	bigint	Código do município de naturalidade da mãe
CODUFNATU	bigint	Código da UF de naturalidade da mãe
ESMAE2010	bigint	Escolaridade 2010. Valores: 0 - Sem escolaridade; 1 - Fundamental I (1a a 4a série); 2 - Fundamental II (5a a 8a série); 3 - Médio (antigo 2o Grau); 4 - Superior incompleto; 5 - Superior completo; 9 - Ignorado.
SERIESC <sup>10</sup> MAE	bigint	Série escolar da mãe. Valores de 1 a 8.
DTNASCMAE	texto	Data de nascimento da mãe
RACACORMAE	bigint	Raça / cor da mãe
QTDGESTANT	texto	Número de gestações anteriores
QTDPARTNOR	texto	Número de partos vaginais
QTDPARTCES	texto	Número de partos cesáreos
IDADEPAI	bigint	Idade do pai
DTULTMENST	texto	Dados da última menstruação (DUM): dd mm aaaa
SEMAGESTAC	bigint	Número de semanas de gestação.

TPMETESTIM	bigint	Método utilizado. Valores: 1 - Exame físico; 2 - Outro método; 9 - Ignorado.
CONSPRENAT 	texto	Número de consultas pré-natal
MESPRENAT 	texto	Mês de gestação em que começou o pré-natal
TPAPRESENT 	bigint	Tipo de apresentação do RN. Valores: 1 - Cefálico; 2 - Pélvica ou podálica; 3 - Transversa; 9 - Ignorado.
STTRABPART 	bigint	Trabalho de parto induzido? Valores: 1 - Sim; 2 - Não; 3 - Não se aplica; 9 - Ignorado.
STCESPARTO 	bigint	Cesárea ocorreu antes do trabalho de parto? Valores: 1 - Sim; 2 - Não; 3 - Não se aplica; 9 - Ignorado.
TPNASCASSI 	bigint	Nascimento foi assistido por? Valores: 1 - Médico; 2 - Enfermeira / obstetrix; 3 - Parteira; 4 - Outros; 9 - Ignorado.
TPFUNCRESP 	bigint	Tipo de função responsável pelo preenchimento. Valores: 1 - Médico; 2 - Enfermeiro; 3 - Parteira; 4 - Funcionário do cartório; 5 - Outros.
TPDOCRESP	bigint	Tipo de documento do responsável. Valores: 1-CNES; 2-CRM; 3- COREN; 4 - RG; 5 - CPF.
DTDECLARAC	texto	Dados da declaração: dd mm aaaa
ESMAEAGR1 	texto	Escolaridade 2010 agregada. Valores: 00 - Sem Escolaridade; 01 - Fundamental I Incompleto; 02 - Fundamental I Completo; 03 - Fundamental II Incompleto; 04 - Fundamental II Completo; 05 - Ensino Médio Incompleto; 06 - Ensino Médio Completo; 07 - Incompleto Superior; 08 - Superior Completo; 09 - Ignorado; 10 - Fundamental I Incompleto ou Inespecífico; 11 - Fundamental II Incompleto ou

		Inespecífico; 12 - Ensino Médio Incompleto ou Inespecífico.
STDNEPIDEM	bigint	Status de DO Epidemiológica. Valores: 1 - SIM; 0 - NÃO.
STDNNOVA	bigint	Status de DO Nova. Valores: 1 - SIM; 0 - NÃO.
CODPAISRES	bigint	Código do país de residência
TPROBSON	texto	Código do Grupo de Robson, criado pelo sistema
PARIDADE	bigint	
KOTELCHUCK	bigint	
nasc_MUNNOME	texto	Nome do município de nascimento
nasc_MUNNOMEX	texto	Nome do município de nascimento em maiúsculas e sem acentos
nasc_AMAZONIA	texto	Indica (S / N), se o município de nascimento faz parte da Amazônia Legal (conforme IBGE)
nasc_FRONTEIRA	texto	Indica (S / N) se o município de nascimento fizer parte da faixa de fronteira (conforme IBGE)
nasc_CAPITAL	texto	Indica (S / N) no município de nascimento e capital da UF
nasc_MSAUDCOI	bigint	Código de Saúde Macrorregional do Município de Nascimento
nasc_RSAUDCOI	bigint	Código de Saúde Regional do Município de Nascimento
nasc_CSAUDCOD	bigint	Código de Saúde Microrregional do Município de Nascimento
nasc_LATITUDE	flutuador	Latitude do município de nascimento
nasc_LONGITUDE	flutuador	Longitude do município de nascimento
nasc_ALTITUDE	bigint	Altitude do município de nascimento
nasc_AREA	flutuador	Área do município de nascimento

nasc_codigo_adotado	bigint	Armazém ou código atribuído ao município de nascimento atualmente, tratando os casos em que os códigos utilizados foram usados para o mesmo município ao longo do tempo
res_MUNNOME	texto	Nome do município de residência
res_MUNNOMEX	texto	Nome do município de residência em maiúsculas e sem acentos
res_AMAZONIA	texto	Indica (S / N) se o município de residência faz parte da Amazônia Legal (conforme IBGE)
res_FRONTEIRA	texto	Indica (S / N), se o município de residência fizer parte da faixa de fronteira (conforme IBGE)
res_CAPITAL	texto	Indica (S / N) no município de residência e capital da UF
res_MSAUDCOD	bigint	Código de Saúde Macrorregional do Município de Residência
res_RSAUDCOD	bigint	Código Regional de Saúde que é o Município de Residência
res_CSAUDCOD	bigint	Código de Saúde Microrregional do Município de Residência
res_LATITUDE	flutuador	Latitude do município de residência
res_LONGITUDE	flutuador	Longitude do município de residência
res_ALTITUDE	bigint	Altitude do município de residência
res_AREA	flutuador	Área do município de residência
res_codigo_adotado	bigint	Armazém ou código atribuído ao município de residência atualmente, tratando os casos em que os códigos utilizados foram usados para o mesmo município ao longo do tempo
nasc_SIGLA_UF	texto	Sigla da unidade de federação de nascimento
nasc_CODIGO_UF	bigint	Código da UF de nascimento

nasc_NOME_UF	texto	Nome da unidade de federação de nascimento
res_SIGLA_UF	texto	Sigla da unidade de federação de residência
res_CODIGO_UF	bigint	Código da UF de residência
res_NOME_UF	texto	Nome da unidade de federação de residência
nasc_REGIAO	texto	Nome da região da unidade de federação de nascimento
res_REGIAO	texto	Nome da região da unidade de federação de residência
codanomal_capitulo ✘	texto	Capítulo CID-10 de malformação congênita ou anomalia cromossômica
codanomal_grupo ✘	texto	Grupo CID-10 de malformação congênita ou anomalia cromossômica
codanomal_categoria ✘	texto	Categoria CID-10 de malformação congênita ou anomalia cromossômica
codanomal_subcategoria ✘	texto	Subcategoria CID-10 de malformação congênita ou anomalia cromossômica
nasc_coordenadas	texto	Coordenadas do município de nascimento
res_coordenadas	texto	Coordenadas do município de residência
parto_prematuro ✘	bigint	Indica a prematuridade do nascimento. 0: não há indícios de prematuridade; 1: há indício de prematuridade dado pela idade gestacional (GESTACAO <= 4); 2: há índice de prematuridade dado pelo peso ao nascer (PESO <2500); 3: a idade gestacional e o peso ao nascer mostram prematuridade
def_parto_prematuro ✘	texto	Indica a prematuridade do nascimento. Termo: não há indícios de prematuridade; Inconclusivo-IG: há indício de prematuridade devido à idade gestacional (GESTACAO <= 4); Peso inconclusivo: há indício de prematuridade devido ao peso ao nascer (PESO <2500); Prematuro: a

		idade gestacional e o peso ao nascer causam prematuridade
--	--	---

## APÊNDICE B – Colunas com a descrição do Tableau

Página do Relatório	View	Medida	Descrição
Dados Gerais	total_nascimentos_sim	conta_nascimentos	Contagem do campo Idanomal com objetivo de contar o número de nascimentos. Há um filtro nessa view para trazer somente os casos com anomalia.
Dados Gerais	total_nascimentos_ nao	Conta_nascimentos	Contagem do campo Idanomal com objetivo de contar o número de casos. Há um filtro nessa view para trazer somente os casos sem anomalia.
Dados Gerais	total_nascimentos_ ignorado	Conta_nascimentos	Contagem do campo Idanomal com objetivo de contar o número de casos. Há um filtro nessa view para trazer somente os casos ignorados.
Dados Gerais	Todas views que envolvem crescimento anual	tx_variacao_[...]	Cálculo de variação anual padrão: Subtrair 1 da contagem de nascimentos no atual dividido pela contagem de nascimentos no ano anterior (porcentagem).
Dados Gerais	Todas views que envolvem crescimento anual	variacao_[...]	Subtração da contagem de nascimentos no atual pela contagem de nascimentos no ano anterior (valor absoluto).
Dados Gerais	variacao_anual_sim	Nasc_variacao_str	O campo nasc_variacao_str concatena os campos tx_variacao_nasc e variacao_nasc com alguns elementos de texto pra que seja possível trazer o crescimento em relação ao ano anterior.
Dados Gerais	variacao_anual_ nao	Nasc_variacao_str	O campo nasc_variacao_str concatena os campos tx_variacao_nasc e variacao_nasc com alguns elementos de texto pra que seja possível trazer o crescimento em relação ao ano anterior.
Dados Gerais	variacao_anual_ ignorado	Nasc_variacao_str	O campo nasc_variacao_str concatena os campos tx_variacao_nasc e variacao_nasc com alguns elementos de texto pra que seja possível trazer o crescimento em relação ao ano anterior.
Dados Gerais	crescimento_anual_sim	conta_nascimentos_Y-1	Contagem do campo Idanomal com objetivo de contar o número de nascimentos no ano anterior. Há um filtro nessa view para trazer somente os casos com anomalia.
Dados Gerais	Crescimento_anual_sim	Conta_nascimentos	Contagem do campo Idanomal com objetivo de contar o número de nascimentos. Há um filtro nessa view para trazer somente os casos com anomalia.

Dados Gerais	nascimentos_cidade_sim	Conta_nascimentos	Contagem do campo Idanomal com objetivo de contar o número de nascimentos. Há um filtro nessa view para trazer somente os casos com anomalia.
Análise das Mães	idade_mae_media	Media_idade_mae	Média de Idademae por Idanomal, Ano Nasc e res SIGLA UF.
Análise das Mães	variacao_idade_mae	tmp_media_idade_mae_string	O campo tmp_media_idade_mae_string concatena os campos tx_variacao_idade_mae e variacao_media_idade_mae com alguns elementos de texto pra que seja possível trazer o crescimento em relação ao ano anterior.
Análise das Mães	frequencia_idade_mae	Conta_casos	Conta nascimentos em que o ano e Idanomal são iguais aos valores selecionados nos parâmetros.
Análise das Mães	idade_mae_por_cidade	media_idade_mae_cidade	Média de Idademae por Idanomal, Ano Nasc, res SIGLA UF e res MUNNOMEX
Análise das Mães	Raça / Cor Mãe	Conta_casos	Conta nascimentos com filtro para diferentes valores de cc_classifica_etnia_mae
Análise das Mães	Raça / Cor Mãe	cc_classifica_etnia_mae	Se Racacormae = 1, Branca; Se Racacormae = 2, Preta; Se Racacormae = 3, Amarela; Se Racacormae = 4, Parda; Se Racacormae = 5, Indígena;
Análise das Mães	Escolaridade	Conta_casos	Conta nascimentos com filtro para diferentes valores de Def Escol Mae
Análise das Mães	Estado Civil	Conta_casos	Conta nascimentos com filtro para diferentes valores de Def Est Civil
Análise das Mães	Ocupação da Mãe	cc_classifica_ocupacao_mae	Se Codocupmae é nulo, então "sem ocupação"; se Codocupmae é diferente de nulo, então "com ocupação".
Análise das Mães	Ocupação da Mãe	Conta_casos	Conta nascimentos com filtro para diferentes valores de cc_classifica_ocupacao_mae

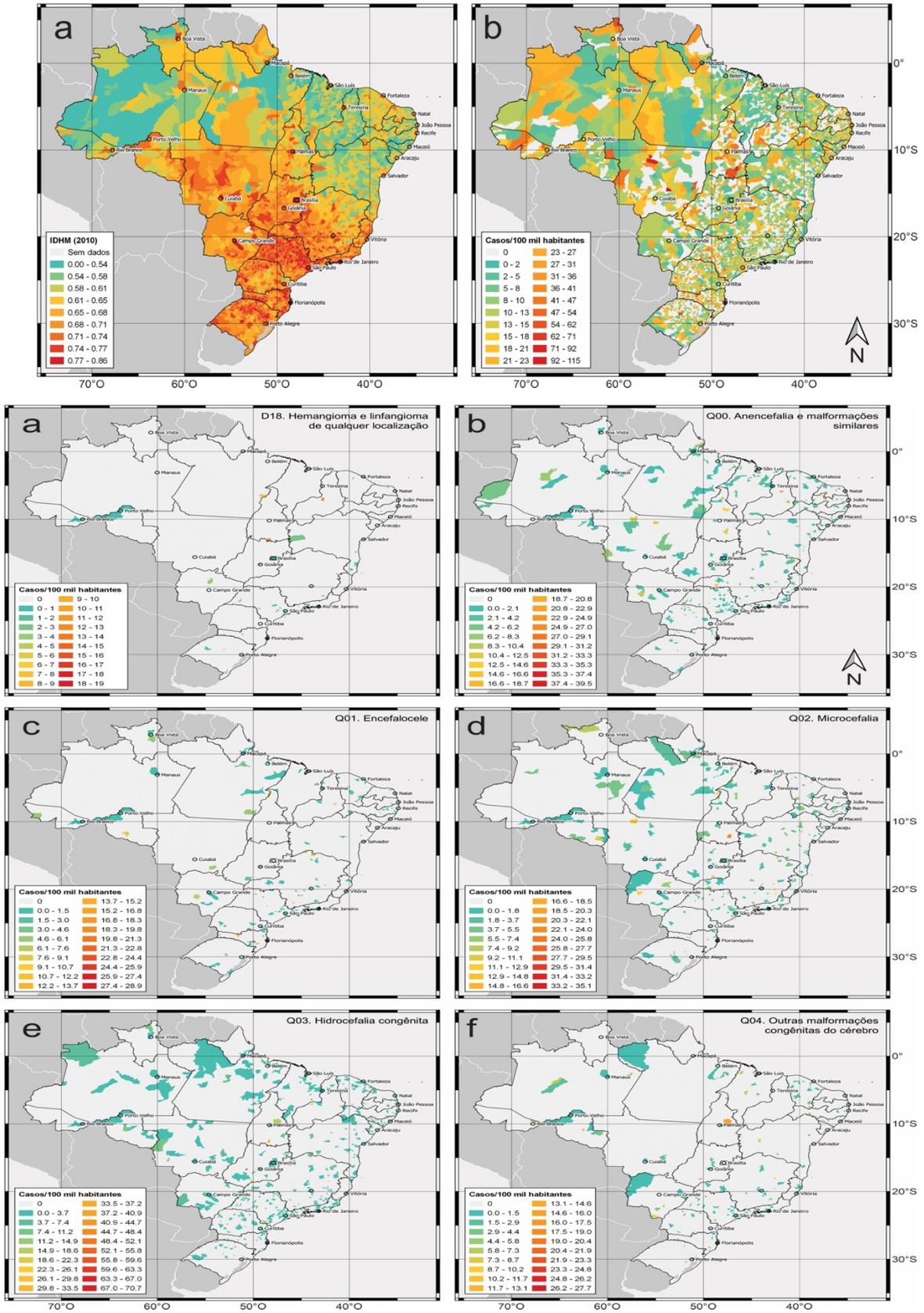
Análise das Mães	Idade da mãe vs. idade do pai	cc_classifica_idademaes_vs_pai	Se Idadepai > Idademaes, 'Mais nova'; Se Idadepai = Idademaes, 'Mesma idade'; Se Idadepai < Idademaes, 'Mais velha'; Se Se Idadepai = nulo ou Idademaes = nulo, 'Ignorado';
Análise das Mães	Idade da mãe vs. idade do pai	Conta_casos	Conta nascimentos com filtro para diferentes valores de cc_classifica_idade_mae_vs_pai
Análise das Mães	qtd_gestacoes_mae	media_qtd_gestações	Média de Qtdgestant por Idanomal, Ano Nasc, res SIGLA UF e res MUNNOMEX
Análise das Mães	variacao_qtd_gestacoes_mae	tmp_media_qtd_gestacoes_string	O campo tmp_media_qtd_gestacoes_string concatena os campos tx_variacao_qtd_gestacoes e variacao_qtd_gestacoes com alguns elementos de texto pra que seja possível trazer o crescimento em relação ao ano anterior.
Análise das Mães	frequencia_qtd_gestacoes	cc_qtd_gestacoes	Se Qtdgestant = 0, '0 gestações'; Se Qtdgestant = 1, '1 gestações'; Se Qtdgestant = 2, '2 gestações'; Se Qtdgestant >= 3, '3 ou mais gestações'
Análise das Mães	quantidade_gestacoes_mae_municipio	media_qtd_gestacoes_cidade	Média de Qtdgestant por Idanomal, Ano Nasc, res SIGLA UF e res MUNNOMEX
Análise das Mães	qtd_filhos_vivos	media_filhos_vivos	Média de Qtdfilvivo por Idanomal, Ano Nasc, res SIGLA UF
Análise das Mães	variacao_qtd_filhos_vivos	tmp_media_filhos_vivos_string	O campo tmp_media_filhos_vivos_string concatena os campos tx_variacao_filhos_vivos e variacao_filhos_vivos com alguns elementos de texto pra que seja possível trazer o crescimento em relação ao ano anterior.
Análise das Mães	frequencia_qtd_filhos_vivos	classifica_qtd_filhos_vivos	Se Qtdfilvivo = 0, '0 filhos'; Se Qtdfilvivo = 1, '1 filhos'; Se Qtdfilvivo = 2, '2 filhos'; Se Qtdfilvivo = 3, '3 filhos'; Se Qtdfilvivo = 4, '4 filhos'; Se Qtdfilvivo >= 5, '5 ou mais filhos';

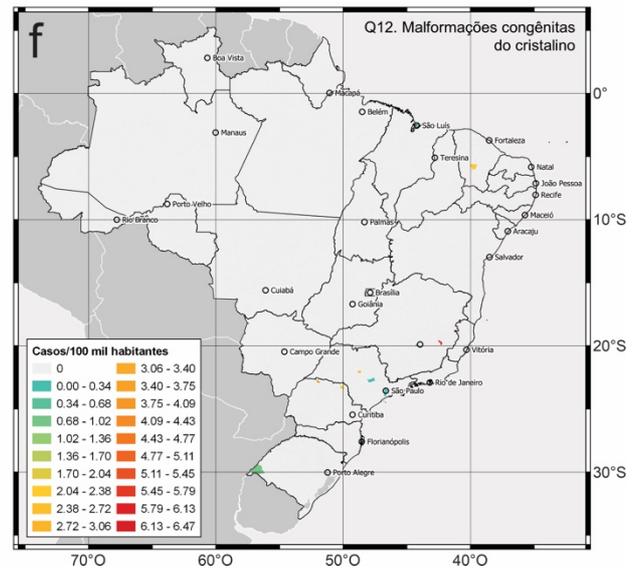
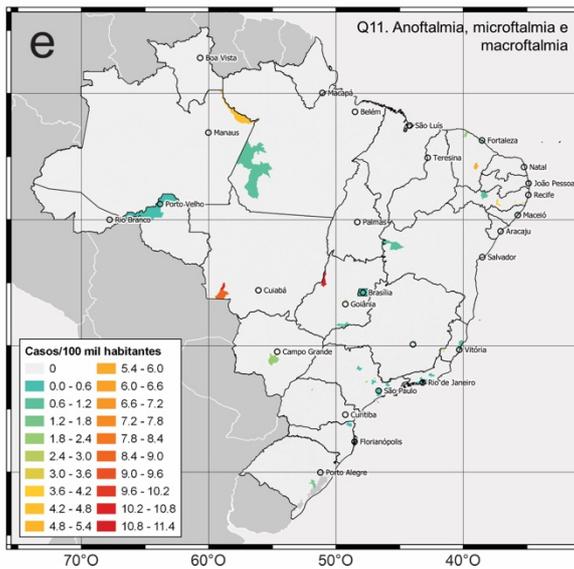
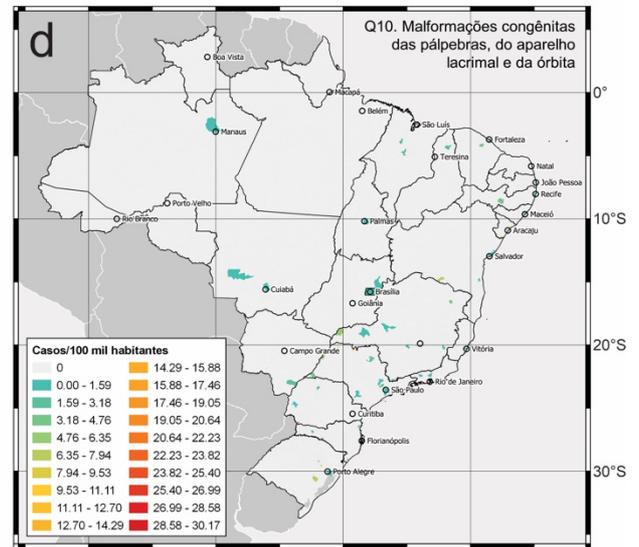
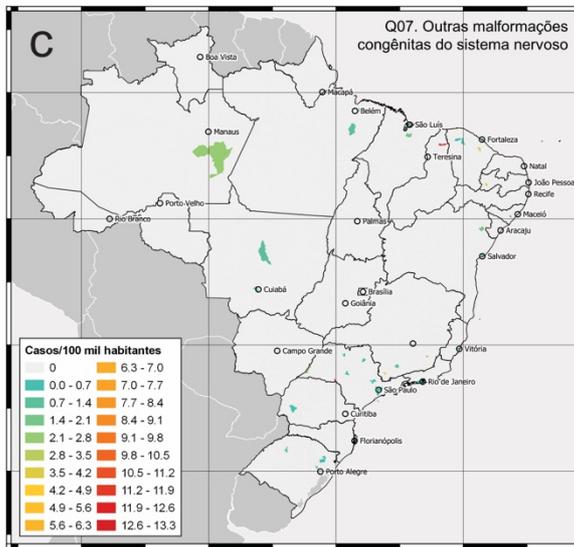
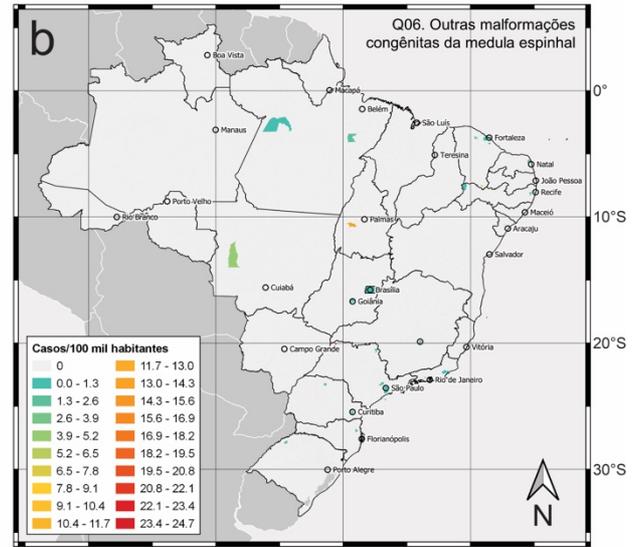
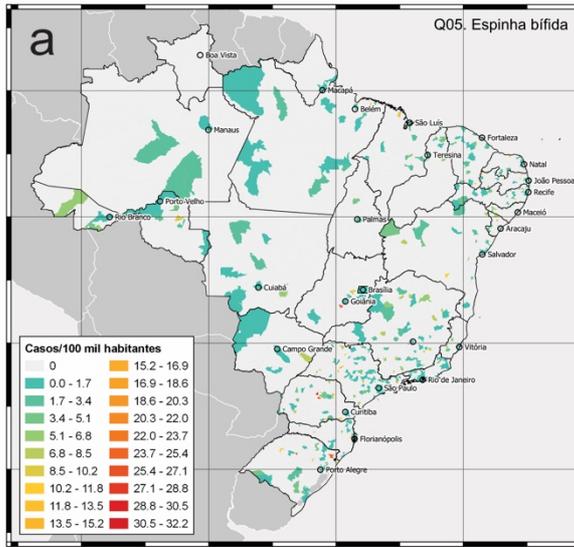
Análise das Gestações	Tipo de Gravidez	Conta_casos	Conta nascimentos com filtro para diferentes valores de Def Gravidez
Análise das Gestações	Tipo de Parto	Conta_casos	Conta nascimentos com filtro para diferentes valores de Def Parto
Análise das Gestações	Local Nascimento	Conta_casos	Conta nascimentos com filtro para diferentes valores de Def Loc Nasc
Análise das Gestações	Tipo de Apresentação do RN	cc_classifica_tipo_apresentação	Se Tpapresent = 1, 'Cefálico'; Se Tpapresent = 2, 'Pélvica'; Se Tpapresent = 3, 'Transversa'; Se Tpapresent = 9, 'Ignorado';
Análise das Gestações	Tipo de Apresentação do RN	Conta_casos	Conta nascimentos com filtro para diferentes valores de cc_classifica_tipo_apresentação
Análise das Gestações	Trabalho de Parto Induzido	Conta_casos	Conta nascimentos com filtro para diferentes valores de cc_classifica_tipo_apresentação
Análise das Gestações	Momento da cesária	cc_classifica_parto_cesarea	Se Stcesparto = 1, 'Sim'; Se Stcesparto = 2, 'Não'; Se Stcesparto = 3, 'Não se aplica'; Se Stcesparto = 9, 'Ignorado';
Análise das Gestações	Momento da cesária	Conta_casos	Conta nascimentos com filtro para diferentes valores de cc_classifica_parto_cesarea
Análise das Gestações	mes_medio_pre_natal	mes_medio_primeira_consulta	Média de Mesprenat por res SIGLA UF;
Análise das Gestações	variacao_mes_primeira_consulta	variacao_mes_medio_primeira_consulta_string	O campo variacao_mes_medio_primeira_consulta_string concatena os campos tx_variacao_mes_medio_primeira_consulta e variacao_mes_medio_primeira_consulta com alguns elementos de texto pra que seja possível trazer o crescimento em relação ao ano anterior.
Análise das Gestações	frequencia_meses_prenatal	cc_meses_prenatal	Se Mesprenat >= 0 e Mesprenat <= 2, 'Mês 2 ou menos'; Se Mesprenat >= 3 e Mesprenat <= 4, 'Mês 3 ou 4'; Se Mesprenat >= 5 e Mesprenat <= 6, 'Mês 5 ou 6'; Se Mesprenat >= 7, 'Mês 7 ou mais'; Se Mesprenat = nulo, 'Ignorado';

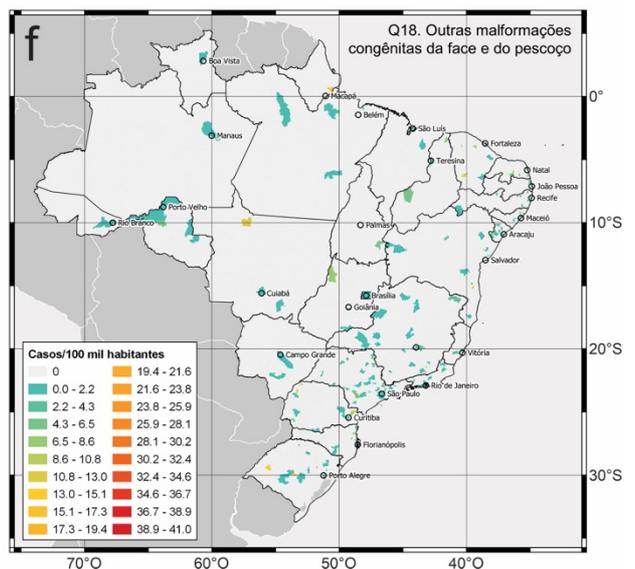
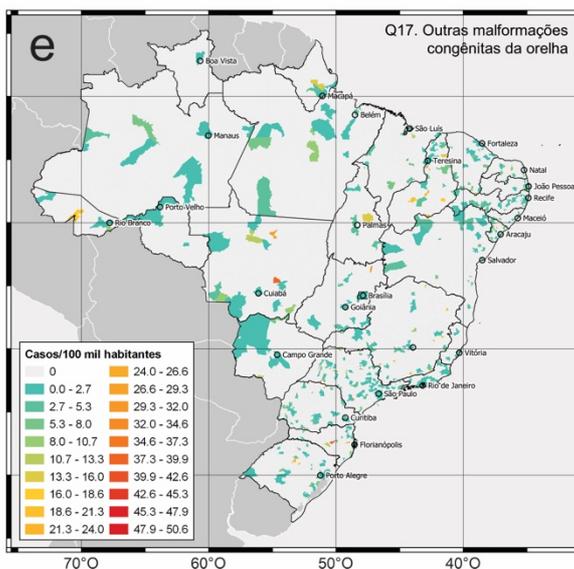
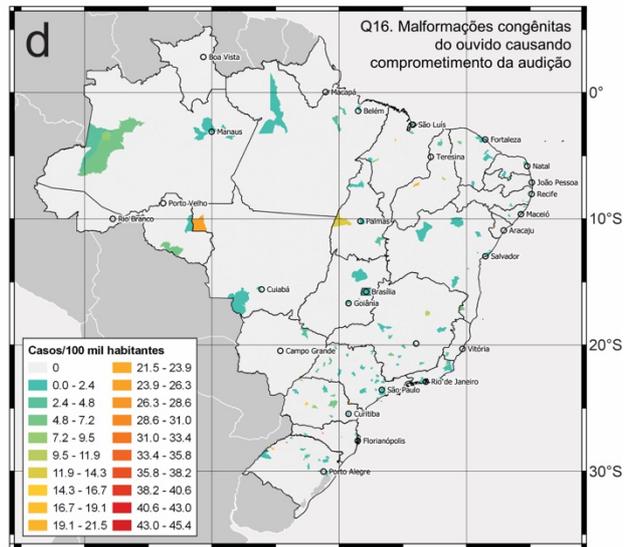
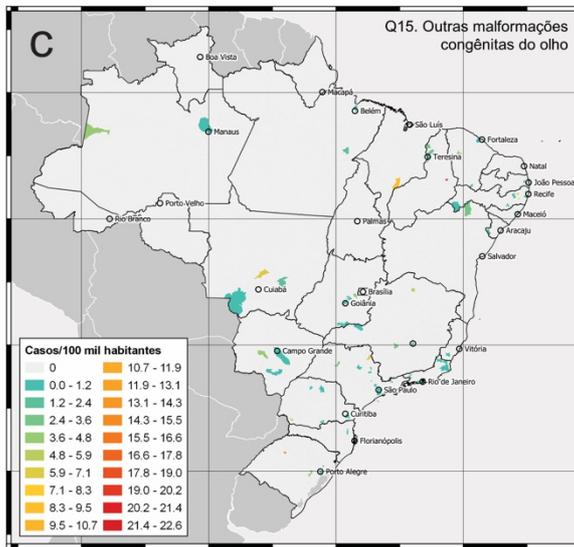
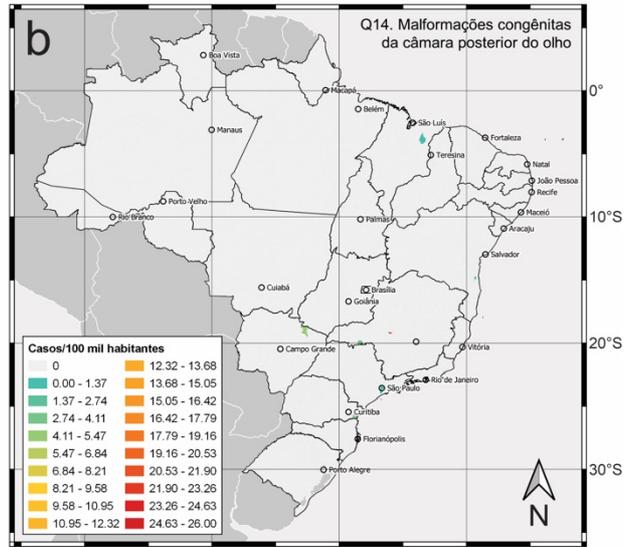
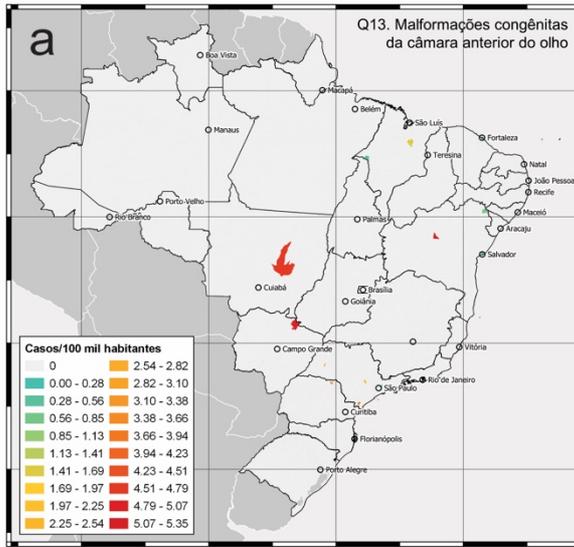
Análise das Gestações	frequencia_meses_prenatal	Conta_casos	Conta nascimentos com filtro para diferentes valores de cc_meses_prenatal
Análise das Gestações	mes_prenatal_por_cidade	mes_medio_primeiro_prenatal_cidade	Média de Consprenat por Idanomal, Ano Nasc, res SIGLA UF e res MUNNOMEX
Análise dos Recém-Nascidos	peso_medio_bebes	media_peso_bebe	Média de Peso por Idanomal, Ano Nasc, res SIGLA UF
Análise dos Recém-Nascidos	variacao_peso_bebes	tmp_peso_bebe_str	O campo tmp_peso_bebe_str concatena os campos tx_variacao_peso_bebe e variacao_peso_bebe com alguns elementos de texto pra que seja possível trazer o crescimento em relação ao ano anterior.
Análise dos Recém-Nascidos	frequencia_peso_bebes	cc_classifica_peso	Se Peso < 2500, '< 2,5 kg'; Se Peso >= 2500 e Peso < 3200, '2,5 a 3,2 kg'; Se Peso >= 3200 e Peso < 3900, '3,2 a 3,9 kg'; Se Peso >= 3900, '> 3,9kg'; Se Peso = nulo, 'Ignorado';
Análise dos Recém-Nascidos	peso_bebes_por_cidade	media_peso_bebe_cidade	Média de Peso por Idanomal, Ano Nasc, res SIGLA UF e res MUNNOMEX
Análise dos Recém-Nascidos	Raça / Cor do Recém-nascido	Conta_casos	Conta nascimentos com filtro para diferentes valores de Def Raca Cor
Análise dos Recém-Nascidos	Sexo do Recém-nascido	Conta_casos	Conta nascimentos com filtro para diferentes valores de Def Sexo
Análise dos Recém-Nascidos	Parto Prematuro	Conta_casos	Conta nascimentos com filtro para diferentes valores de Def Parto Prematuro
Análise dos Recém-Nascidos	media_apgar1	media_apgar1	Média de Apgar1 por Idanomal, Ano Nasc, res SIGLA UF
Análise dos Recém-Nascidos	variacao_media_apgar1	tmp_apgar1_str	O campo tmp_apgar1_str concatena os campos tx_variacao_apgar1 e variacao_apgar1 com alguns elementos de texto pra que seja possível trazer o crescimento em relação ao ano anterior.

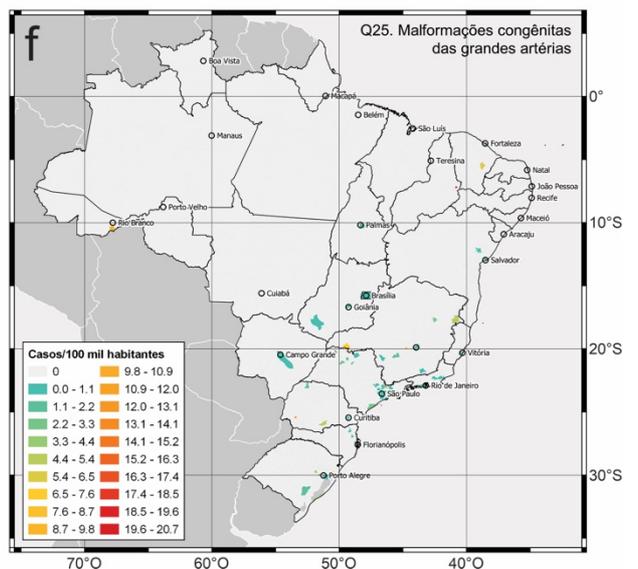
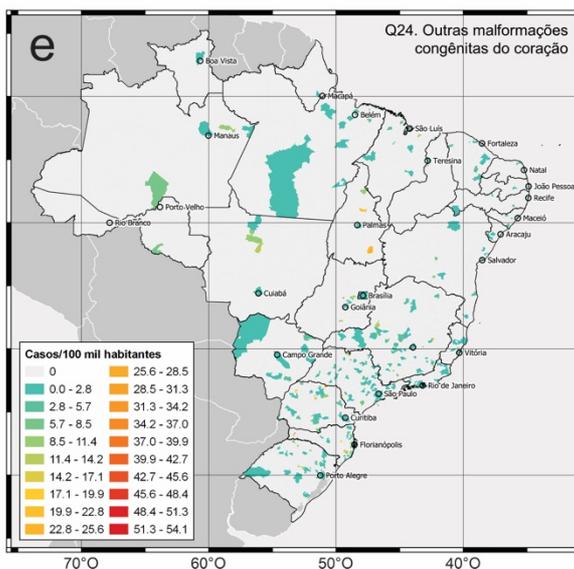
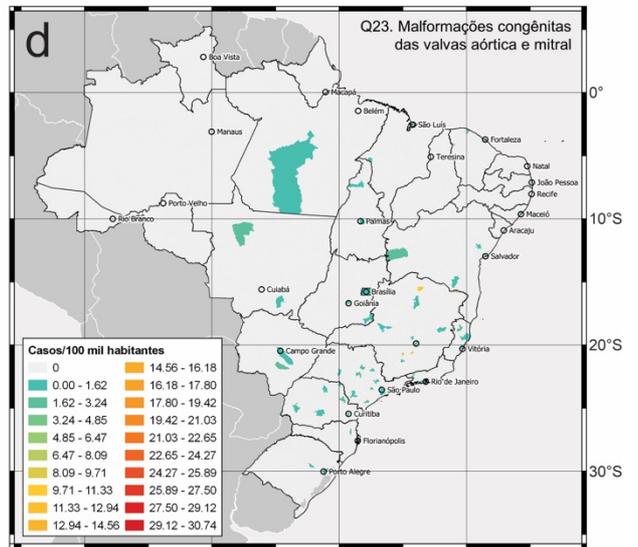
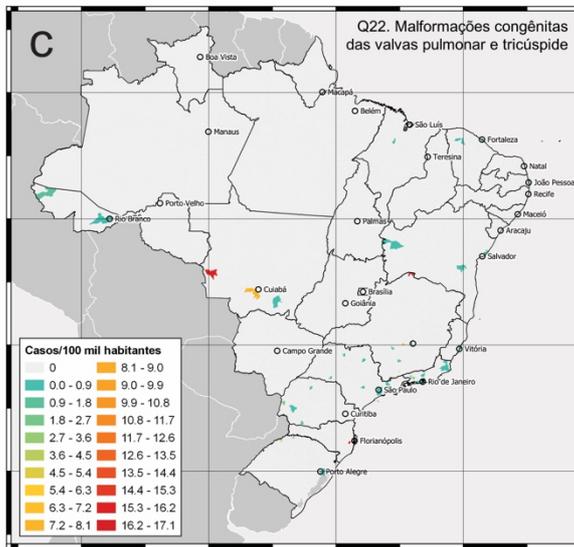
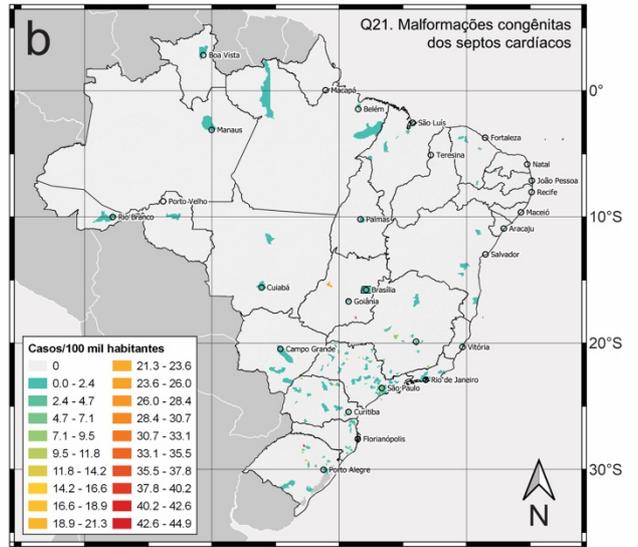
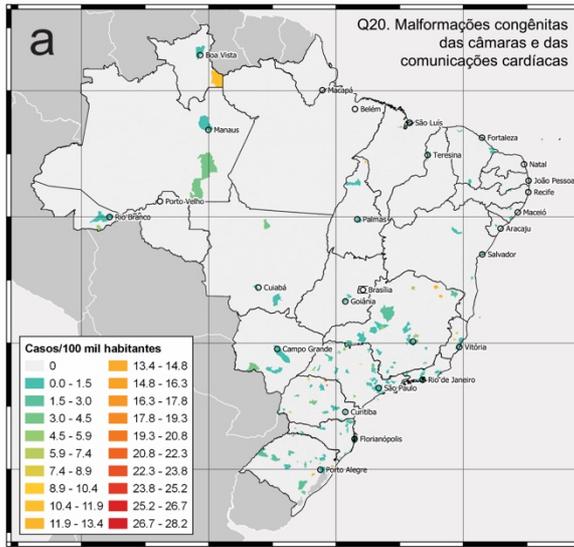
Análise dos Recém-Nascidos	frequencia_apgar1	classifica_apgar1	Se Apgar1 >= 1 e Apgar1 <= 2 , '1 a 2'; Se Apgar1 >= 3 e Apgar1 <= 4 , '3 a 4'; Se Apgar1 >= 5 e Apgar1 <= 6 , '5 a 6'; Se Apgar1 >= 7 e Apgar1 <= 8 , '7 a 8'; Se Apgar1 >= 9 e Apgar1 <= 10 , '9 a 10'; Senão, 'Ignorado';
Análise dos Recém-Nascidos	media_apgar1_por_cidade	media_apgar1_cidade	Média de Apgar1 por Idanomal, Ano Nasc, res SIGLA UF e res MUNNOMEX
Análise dos Recém-Nascidos	media_apgar5	media_apgar5	Média de Apgar5 por Idanomal, Ano Nasc, res SIGLA UF
Análise dos Recém-Nascidos	variacao_media_apgar5	tmp_apgar5_str	O campo tmp_apgar5_str concatena os campos tx_variacao_apgar5 e variacao_apgar5 com alguns elementos de texto pra que seja possível trazer o crescimento em relação ao ano anterior.
Análise dos Recém-Nascidos	frequencia_apgar5	classifica_apgar5	Se Apgar5 >= 1 e Apgar5 <= 2 , '1 a 2'; Se Apgar5 >= 3 e Apgar5 <= 4 , '3 a 4'; Se Apgar5 >= 5 e Apgar5 <= 6 , '5 a 6'; Se Apgar5 >= 7 e Apgar5 <= 8 , '7 a 8'; Se Apgar5 >= 9 e Apgar5 <= 10 , '9 a 10'; Senão, 'Ignorado';
Análise dos Recém-Nascidos	media_apgar5_por_cidade	media_apgar5_cidade	Média de Apgar5 por Idanomal, Ano Nasc, res SIGLA UF e res MUNNOMEX

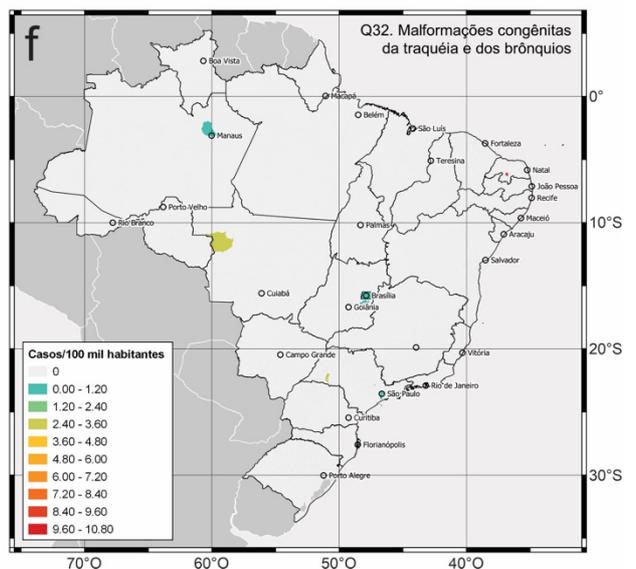
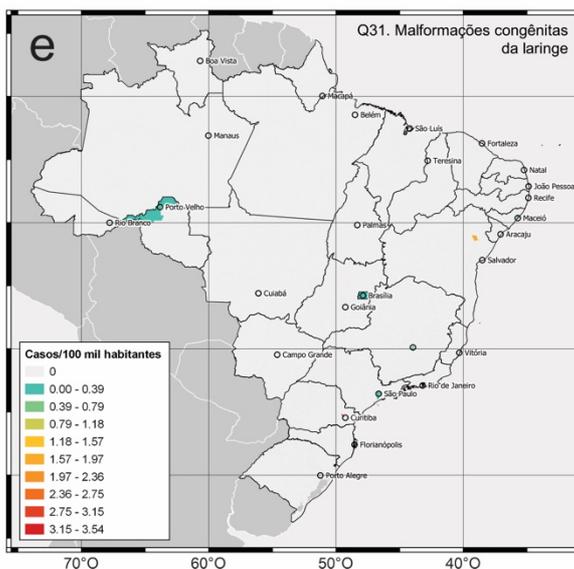
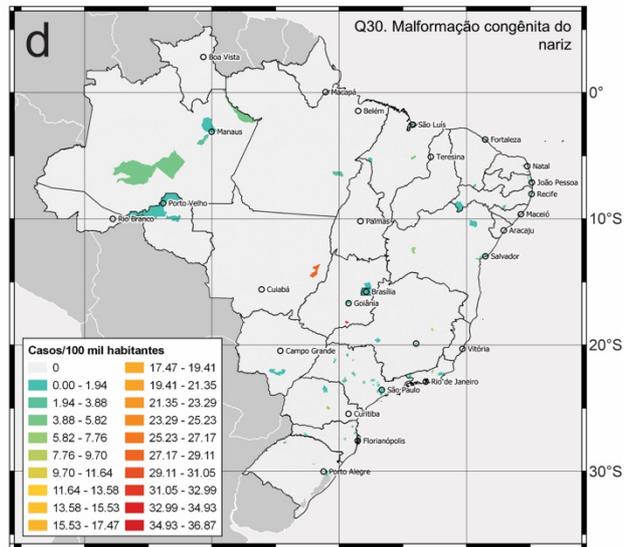
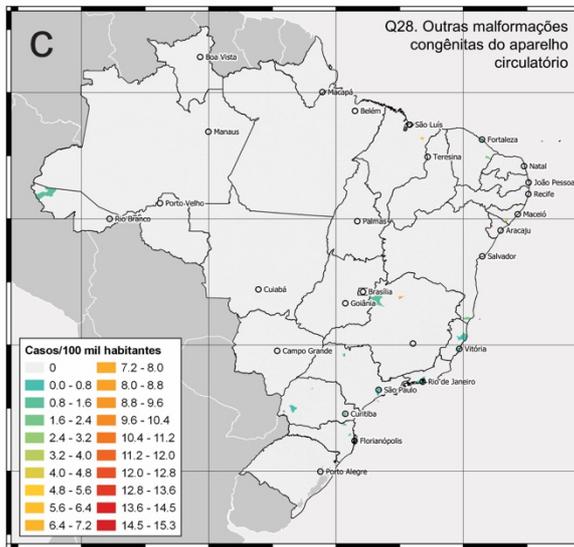
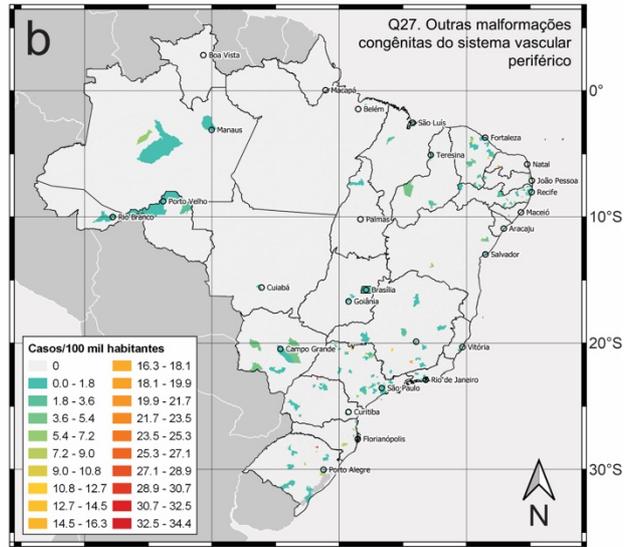
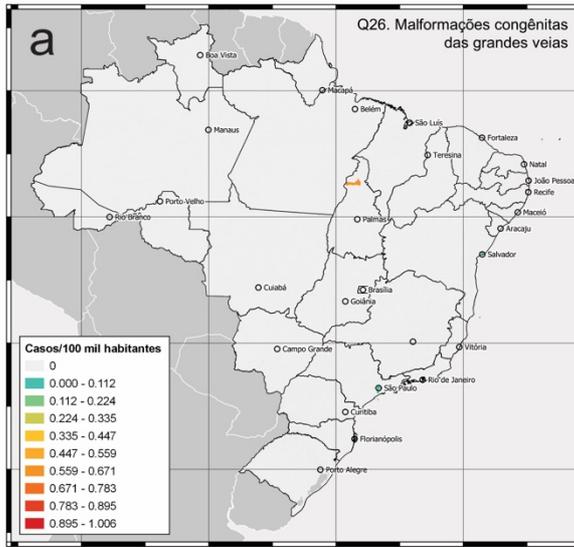
APÊNDICE C – Mapas com visualização de casos

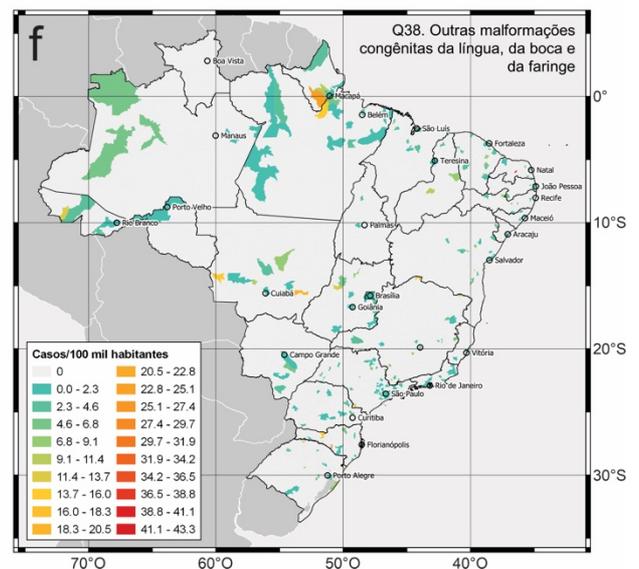
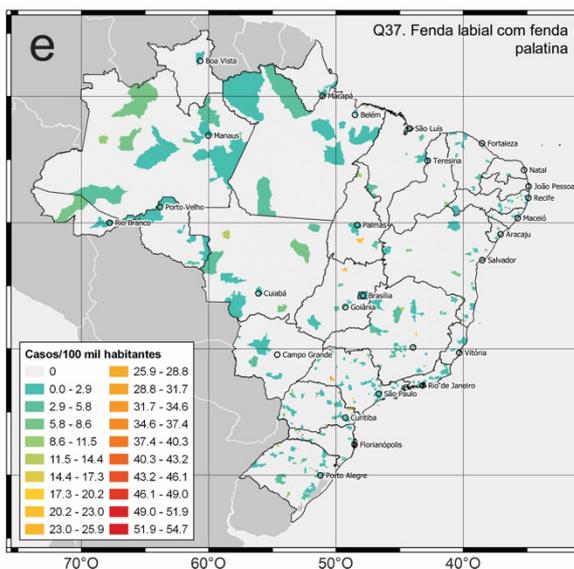
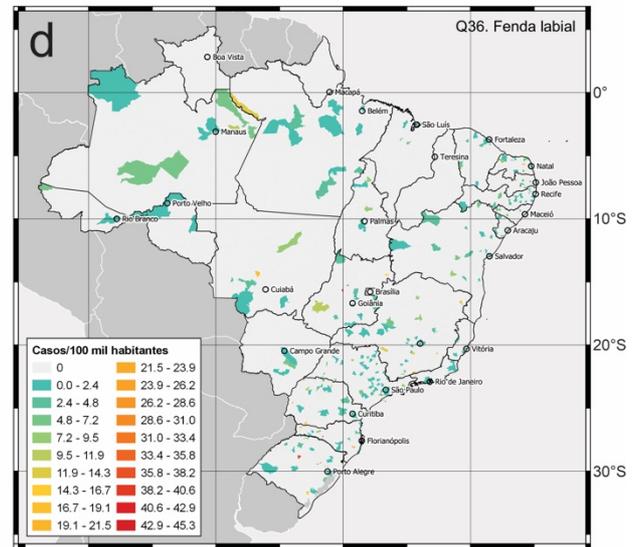
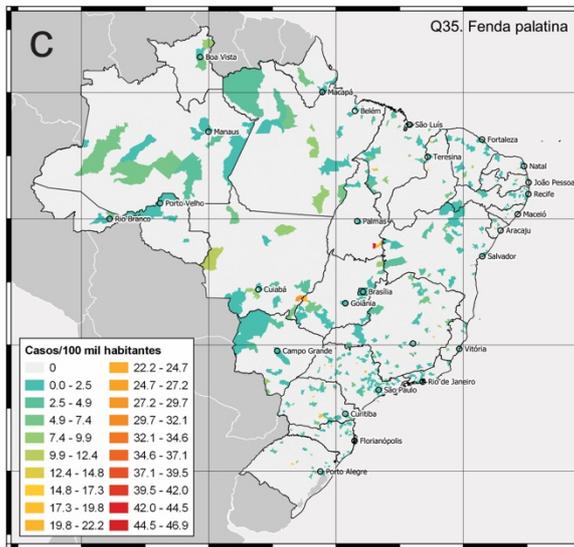
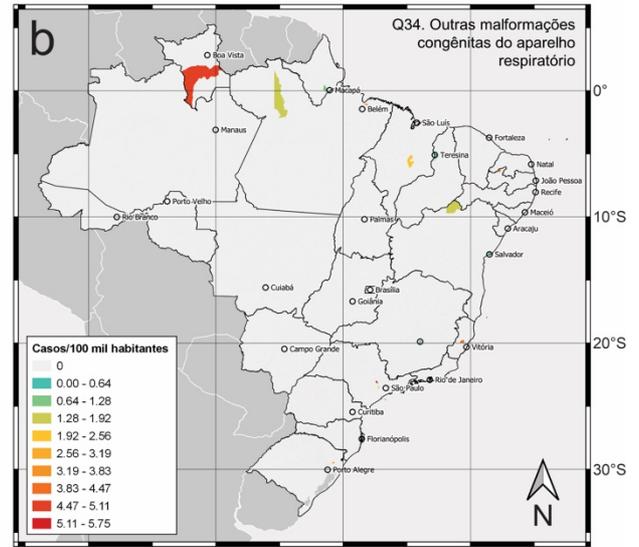
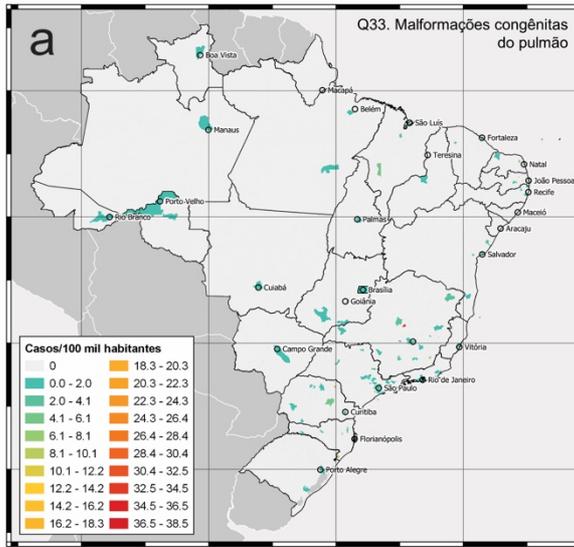


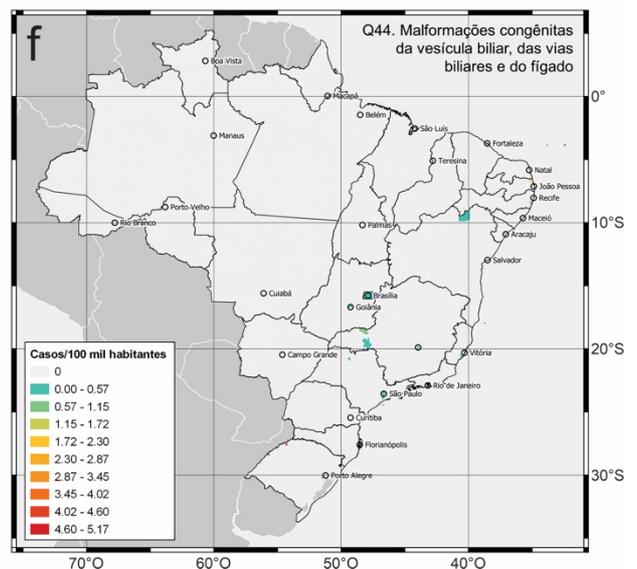
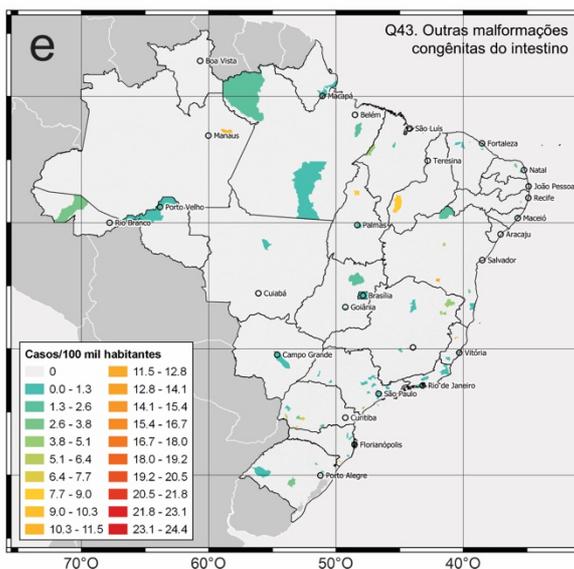
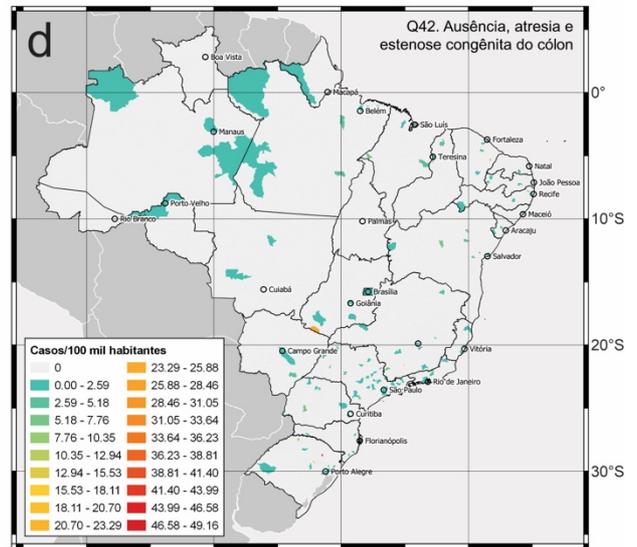
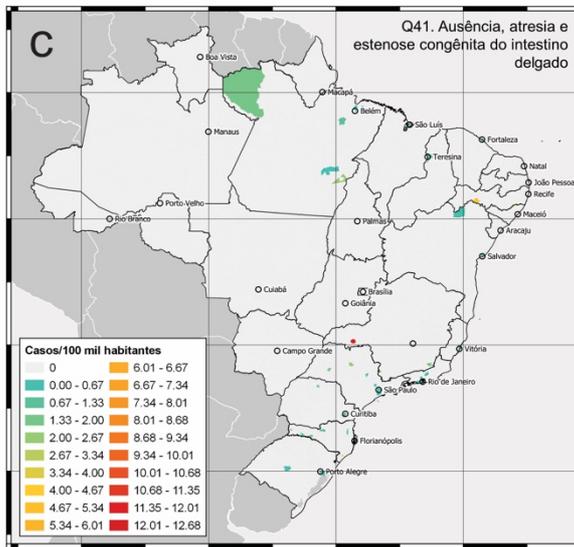
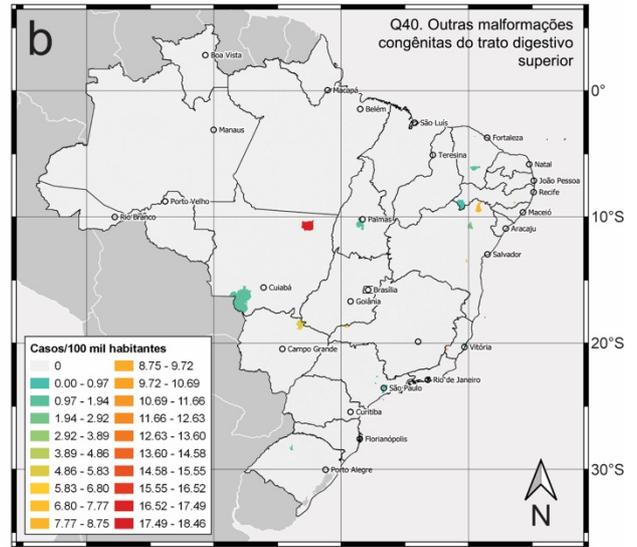
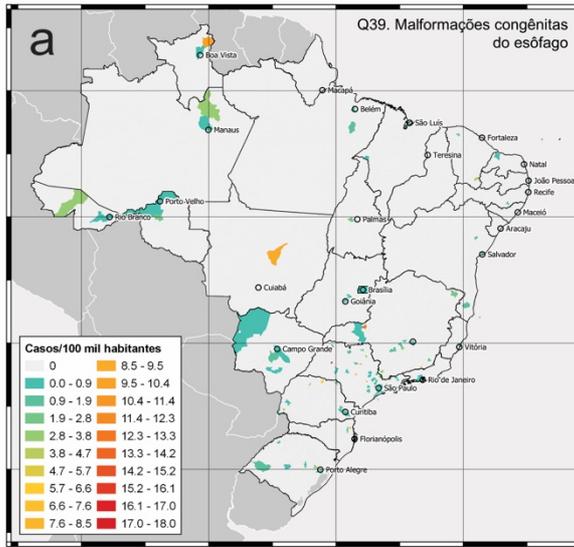


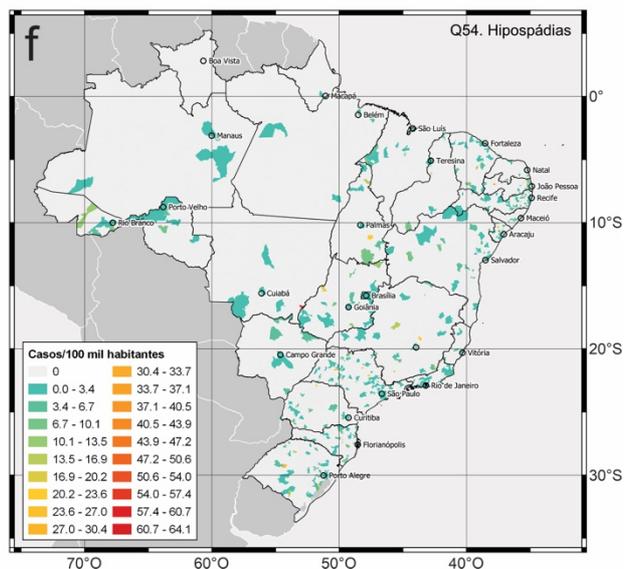
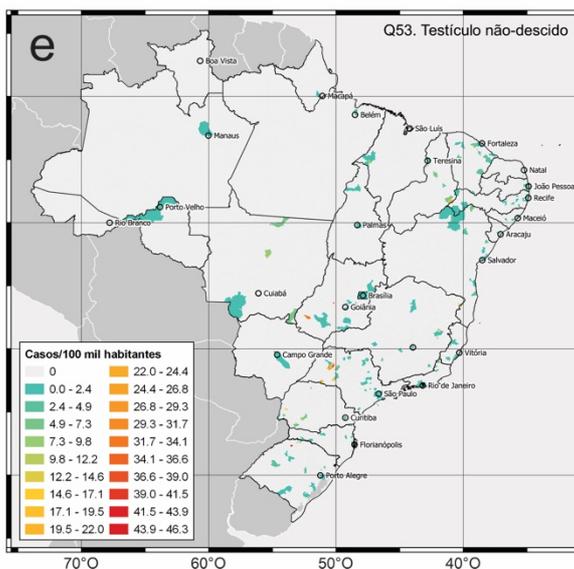
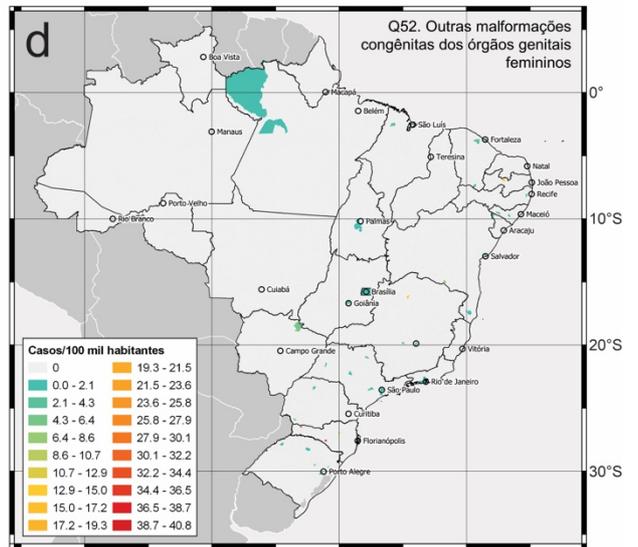
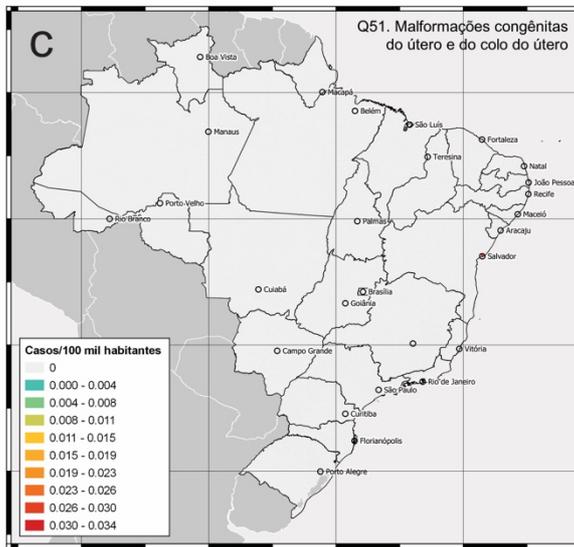
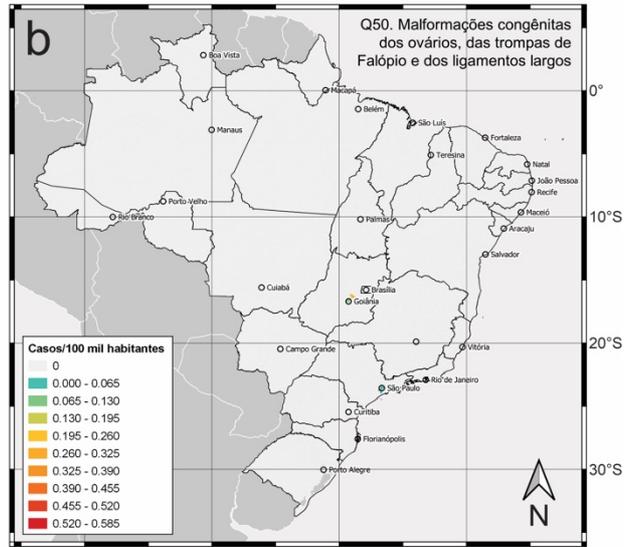
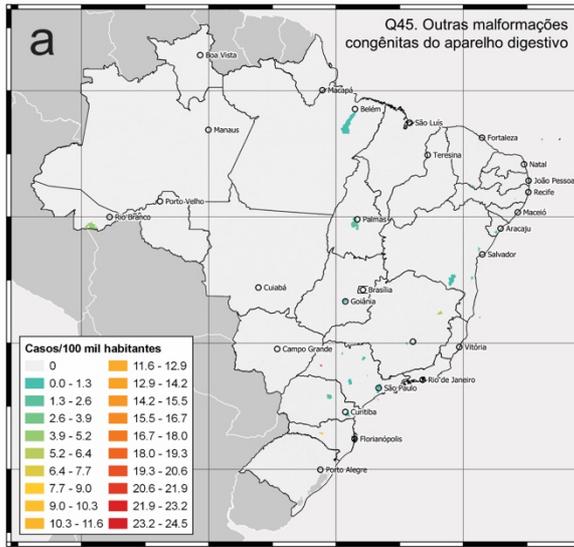


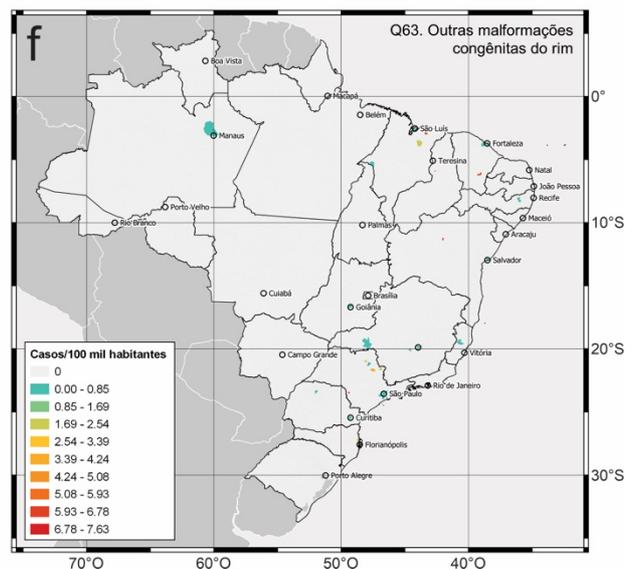
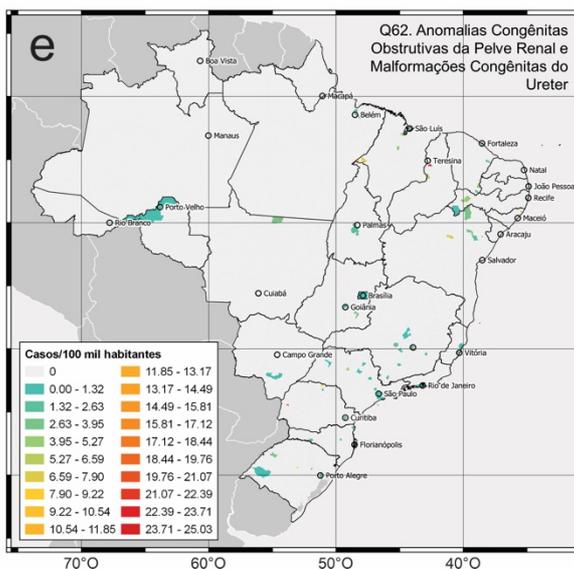
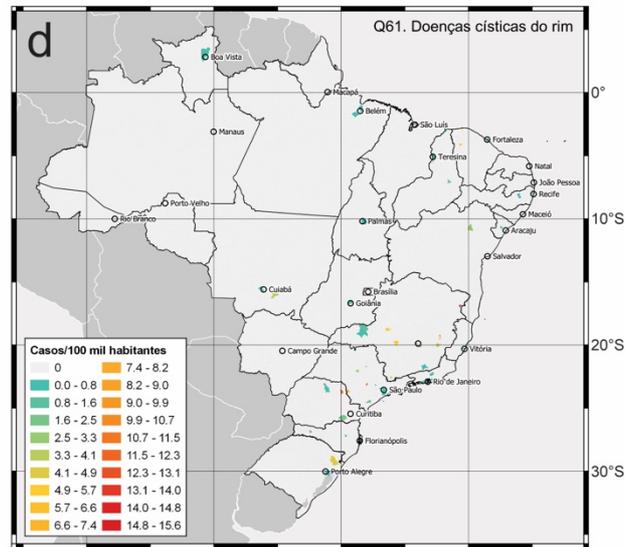
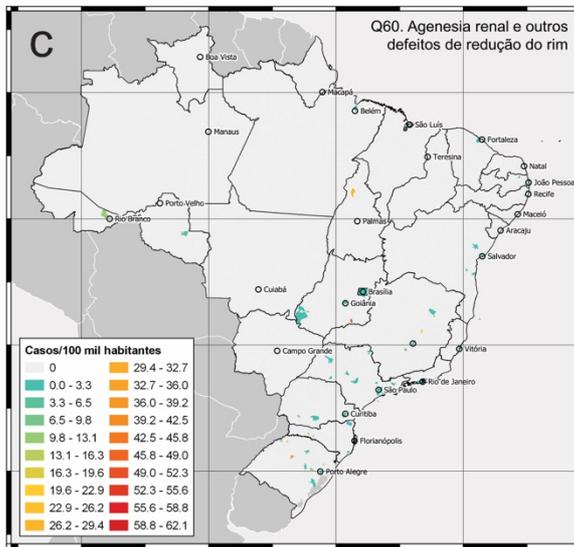
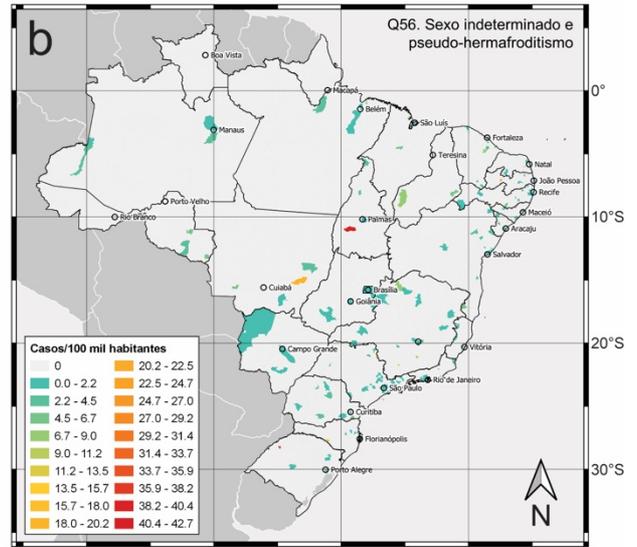
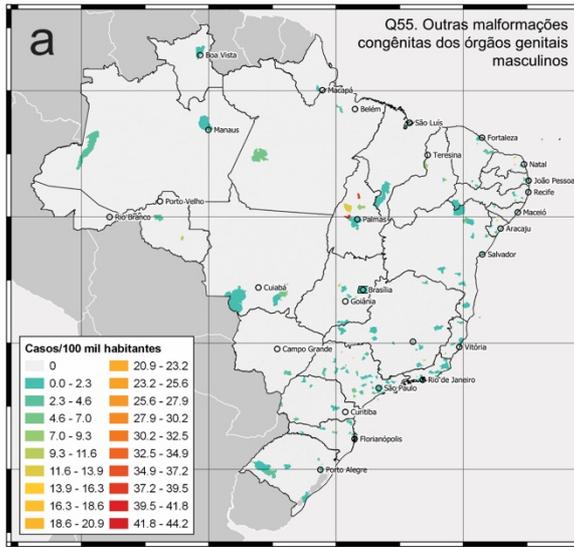


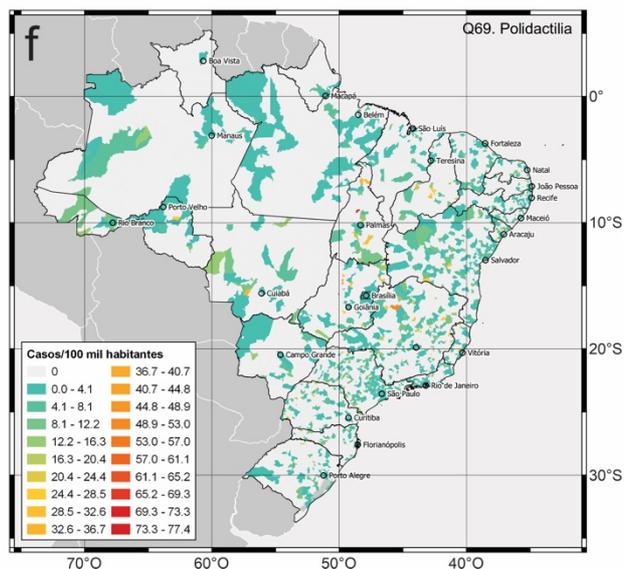
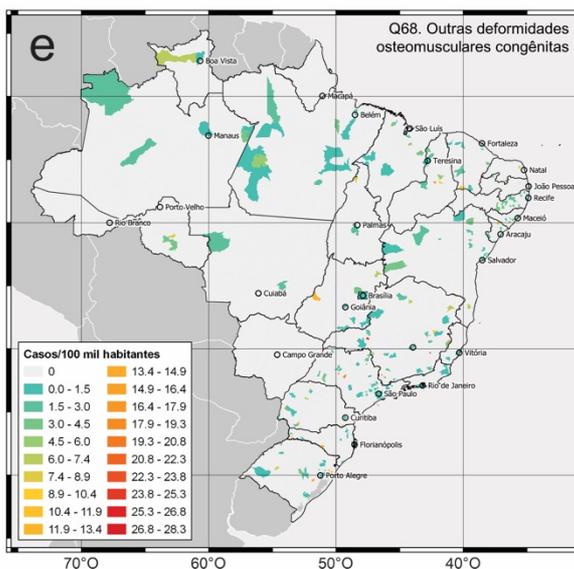
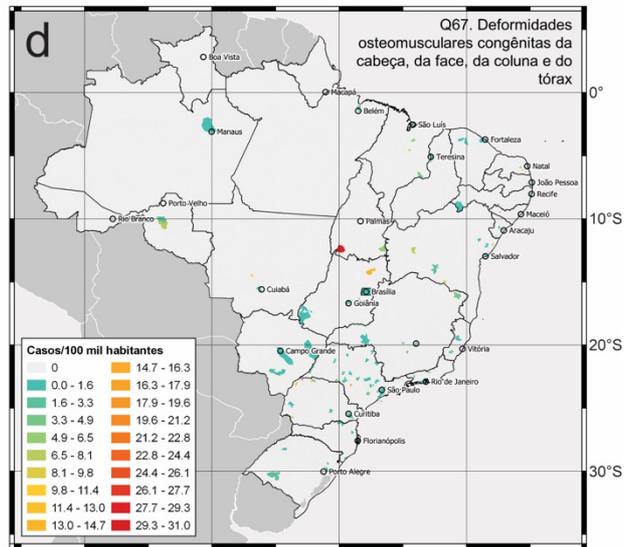
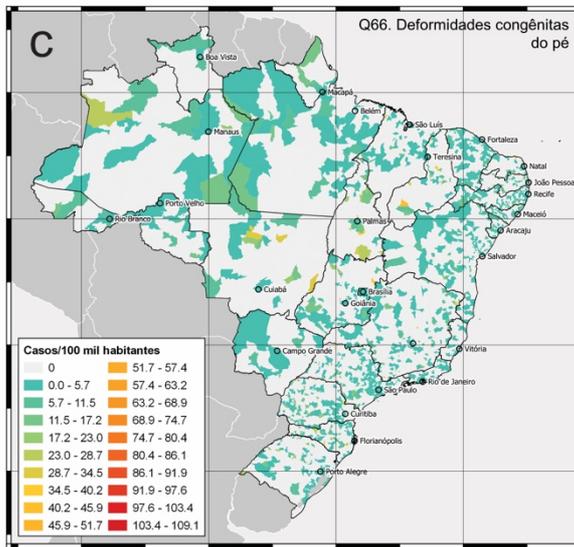
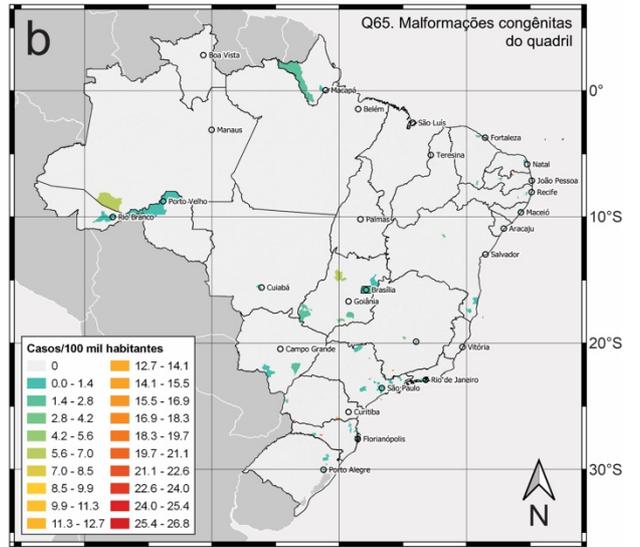
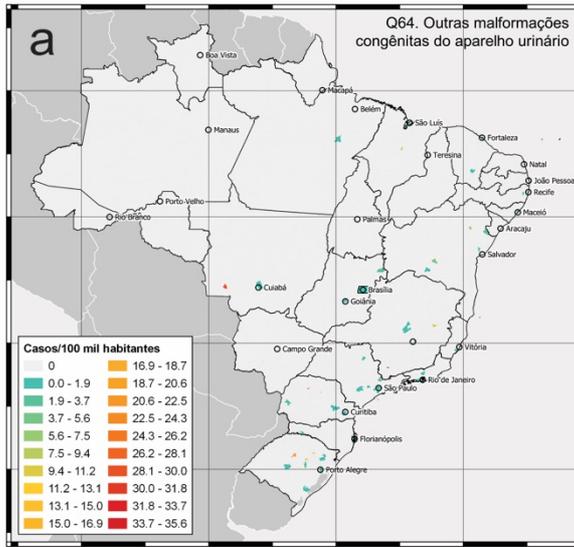






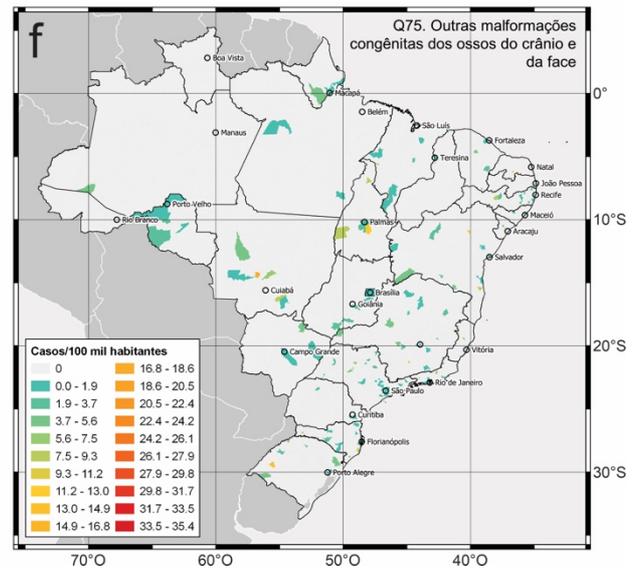
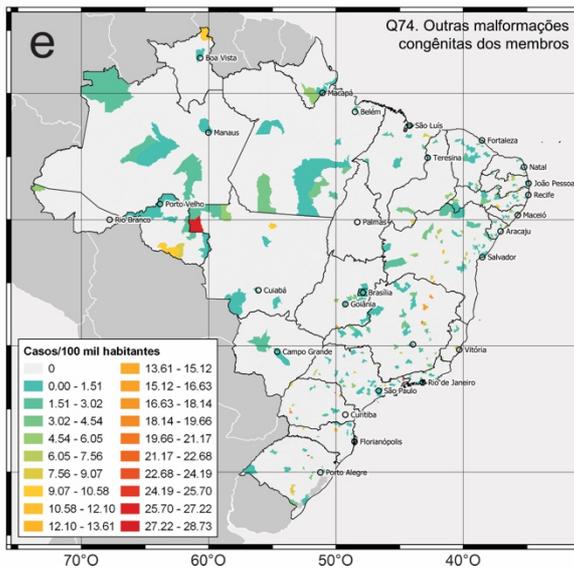
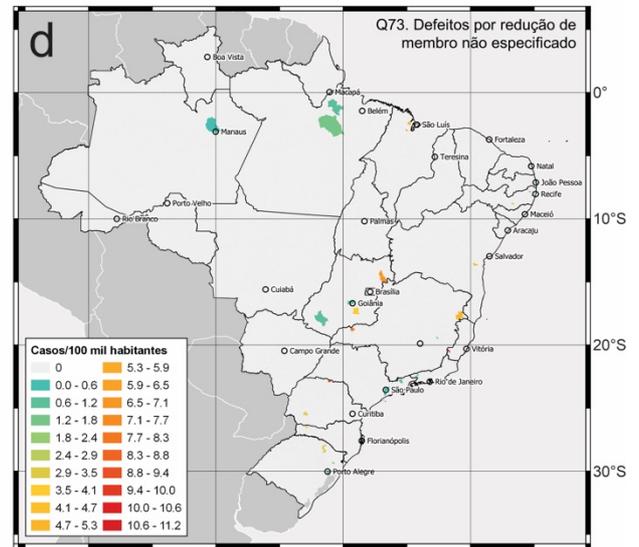
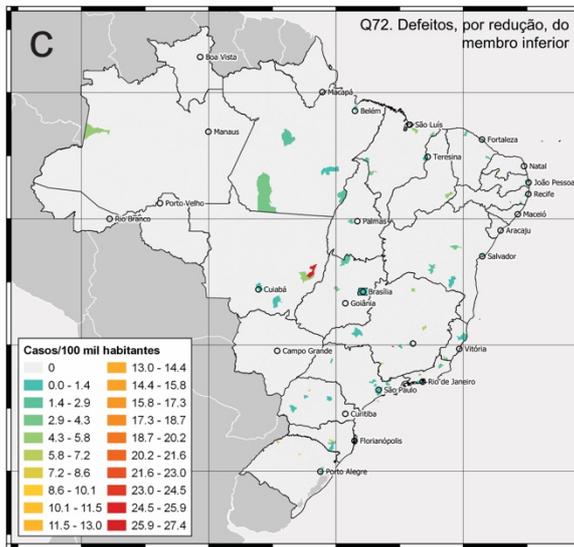
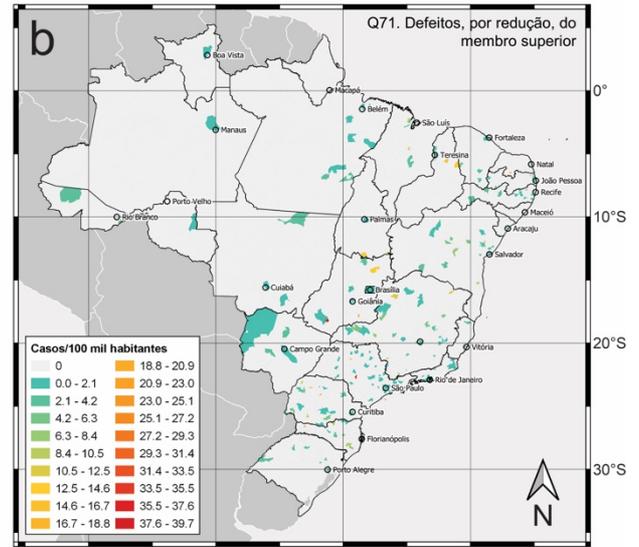
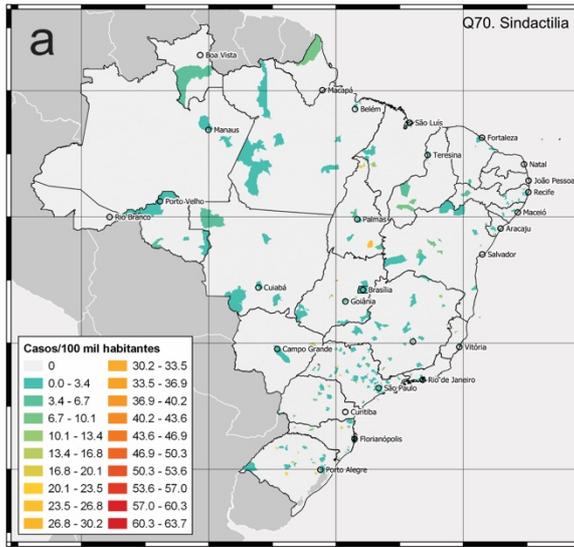


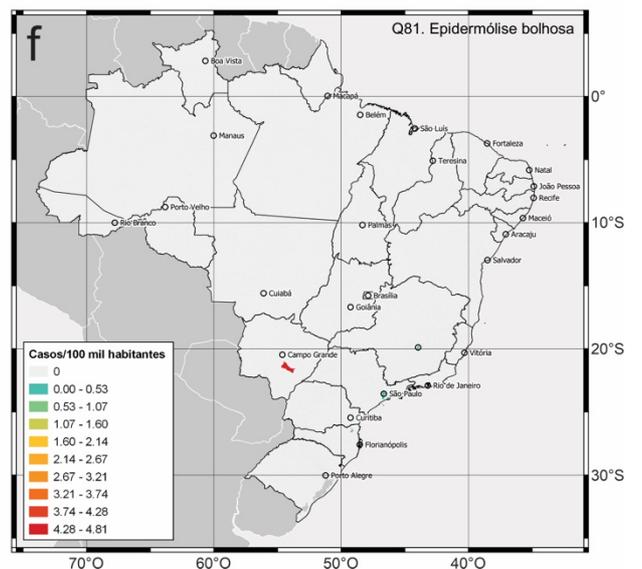
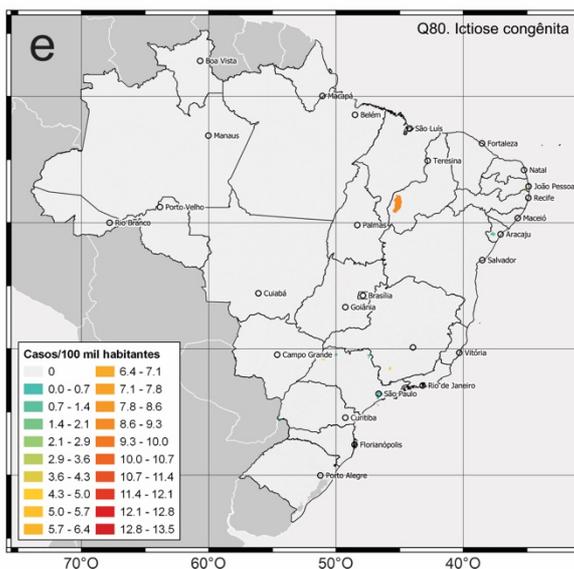
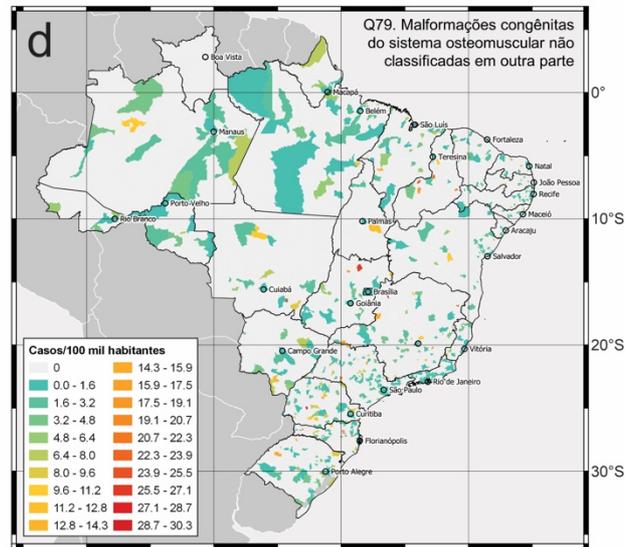
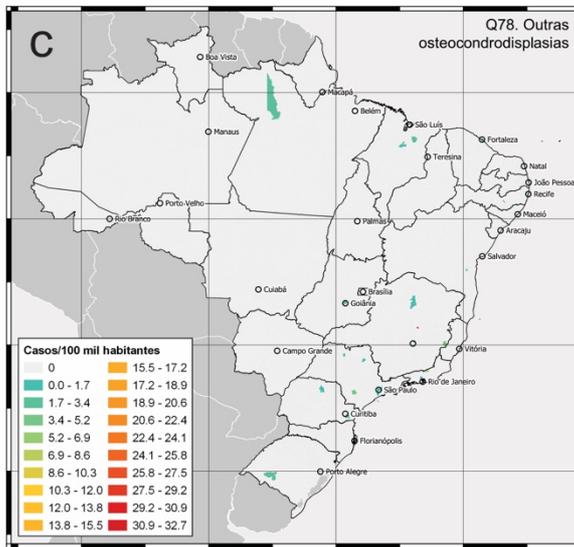
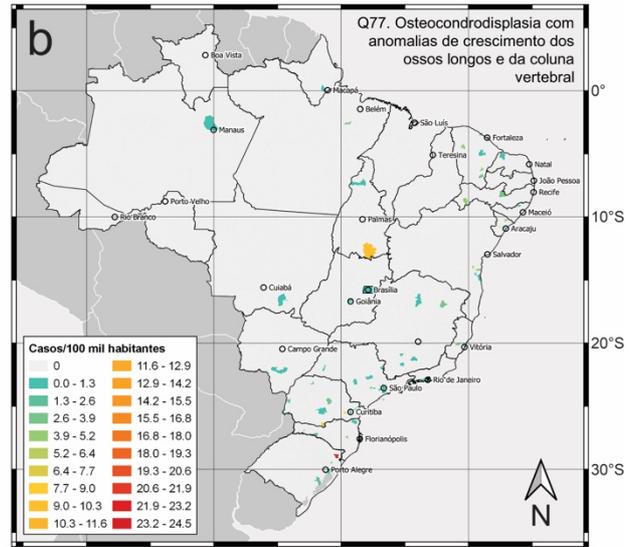
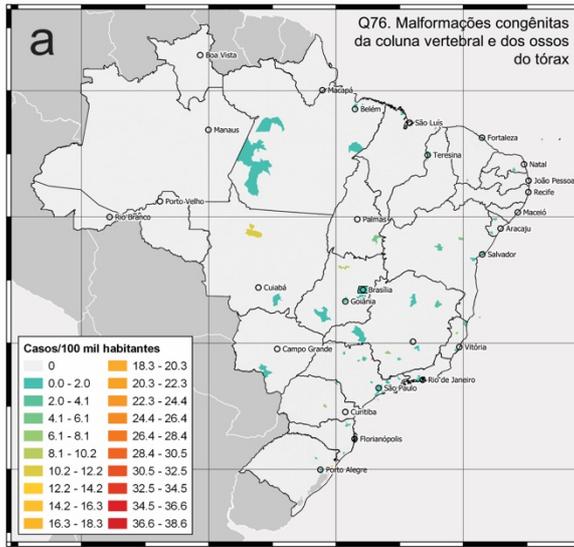


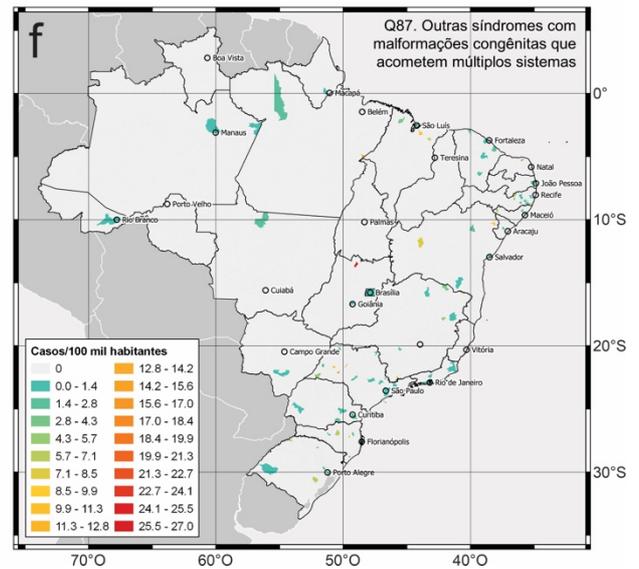
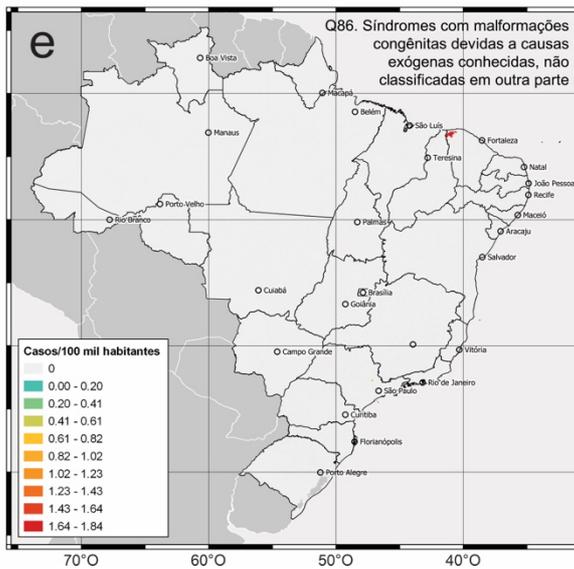
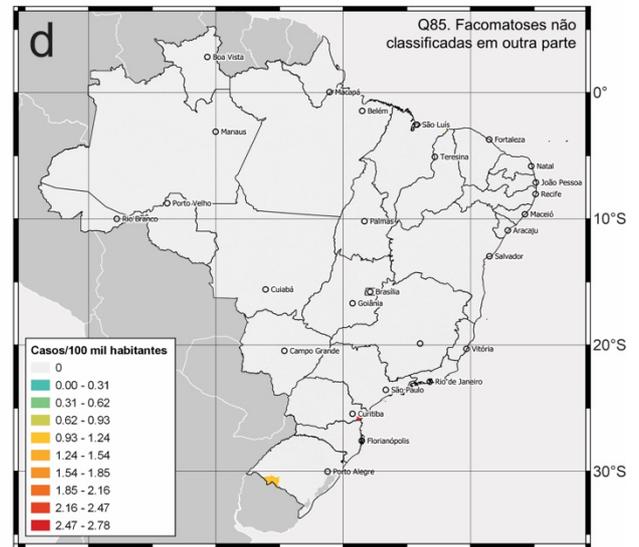
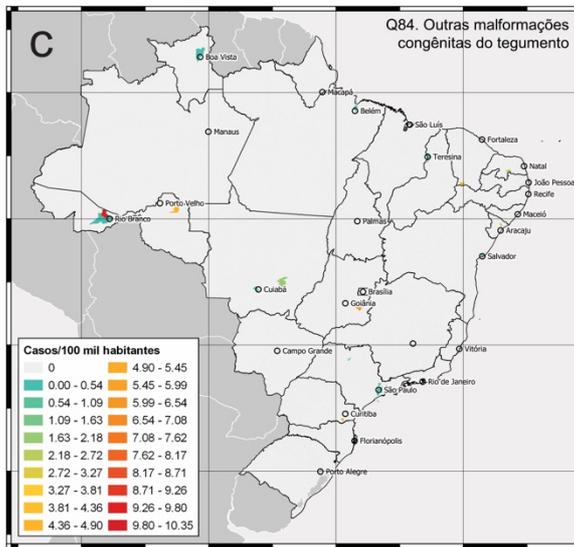
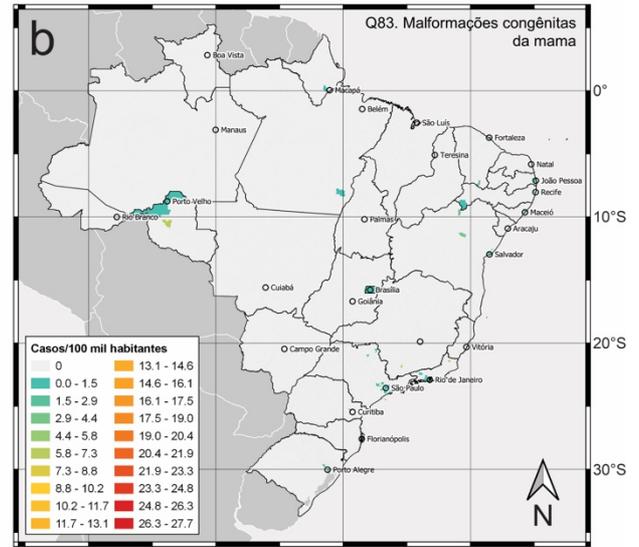
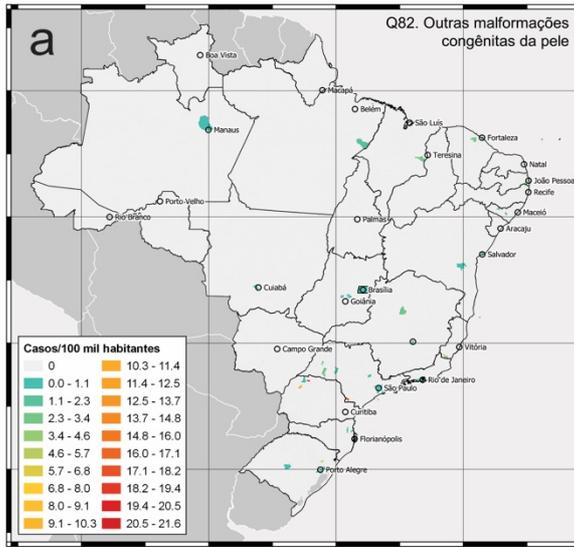


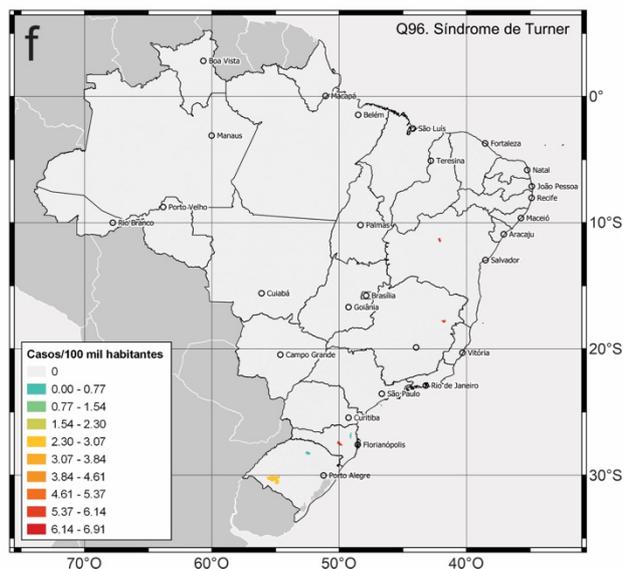
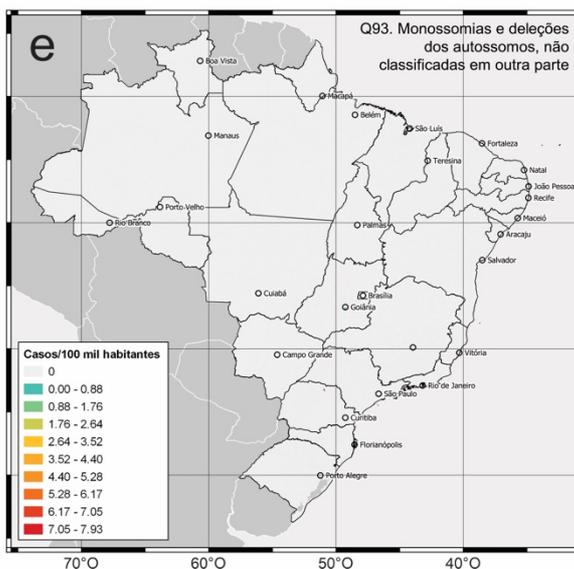
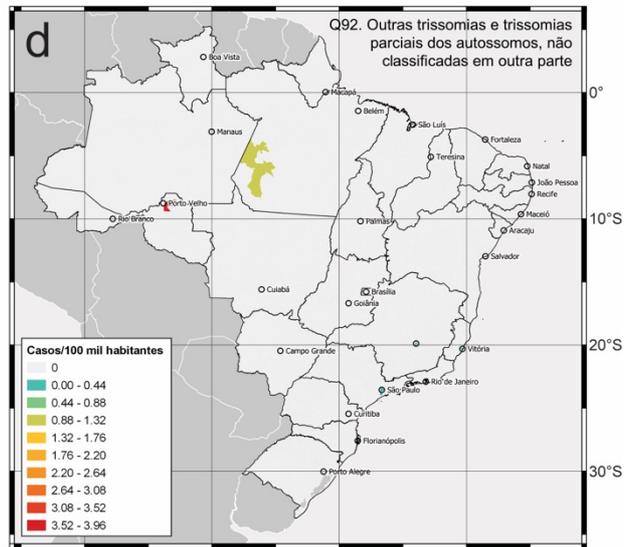
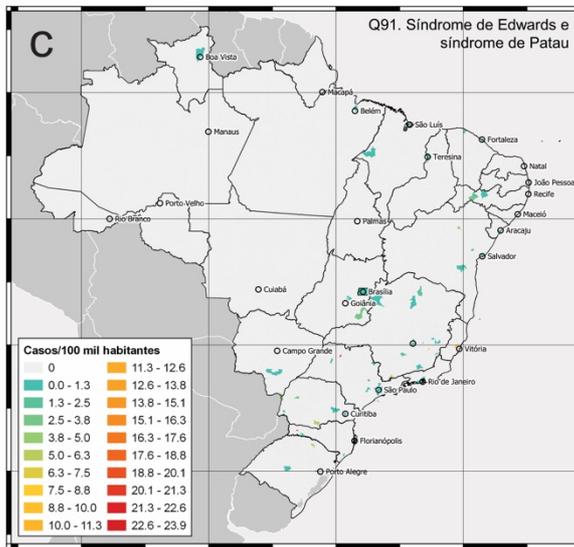
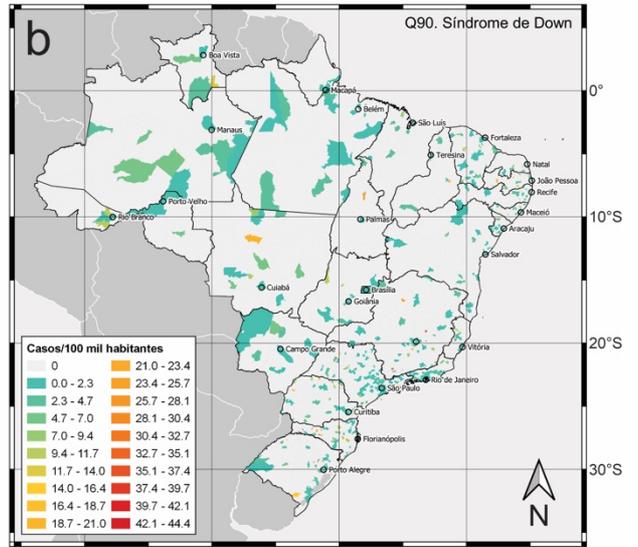
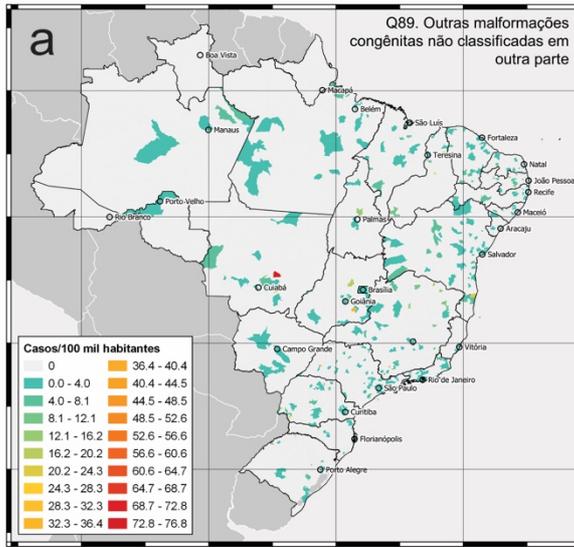
70°O 60°O 50°O 40°O

70°O 60°O 50°O 40°O



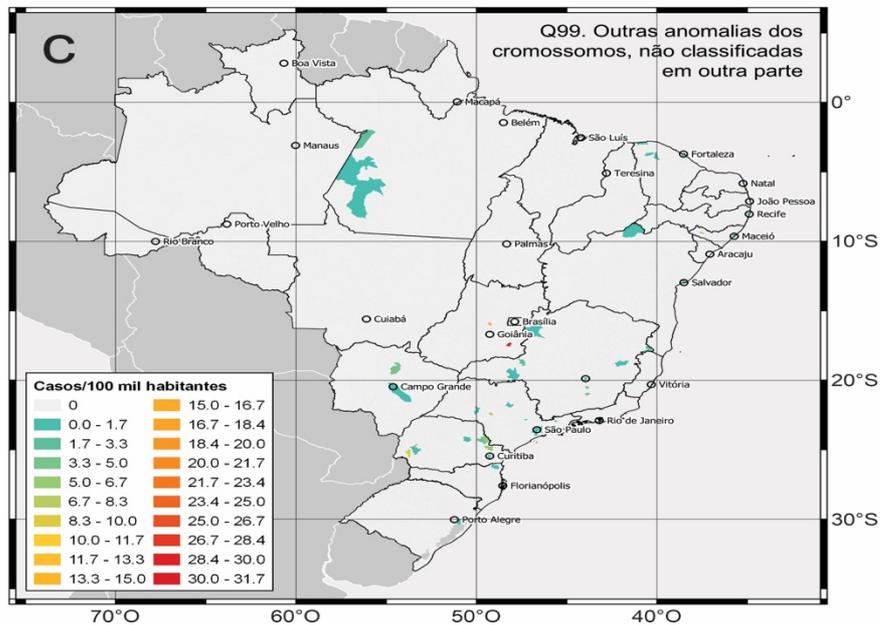
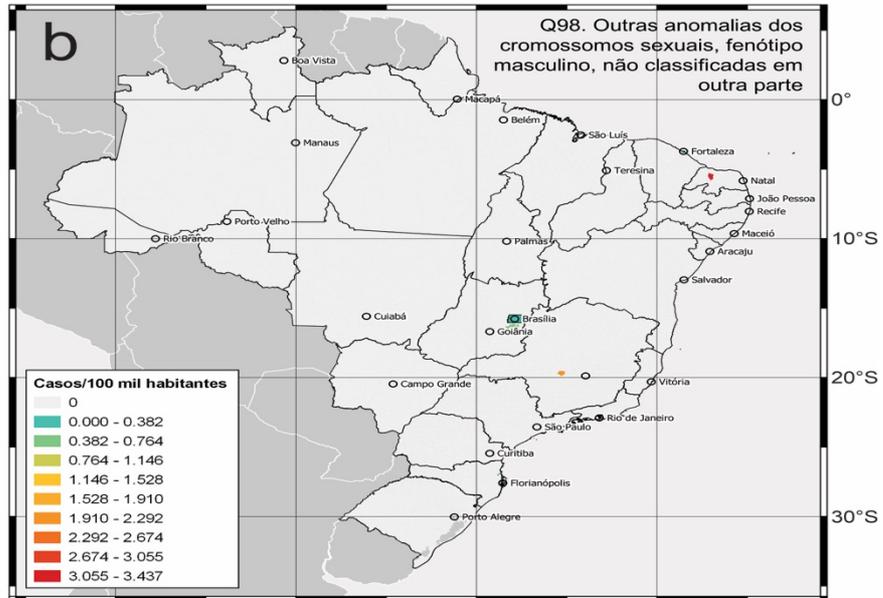
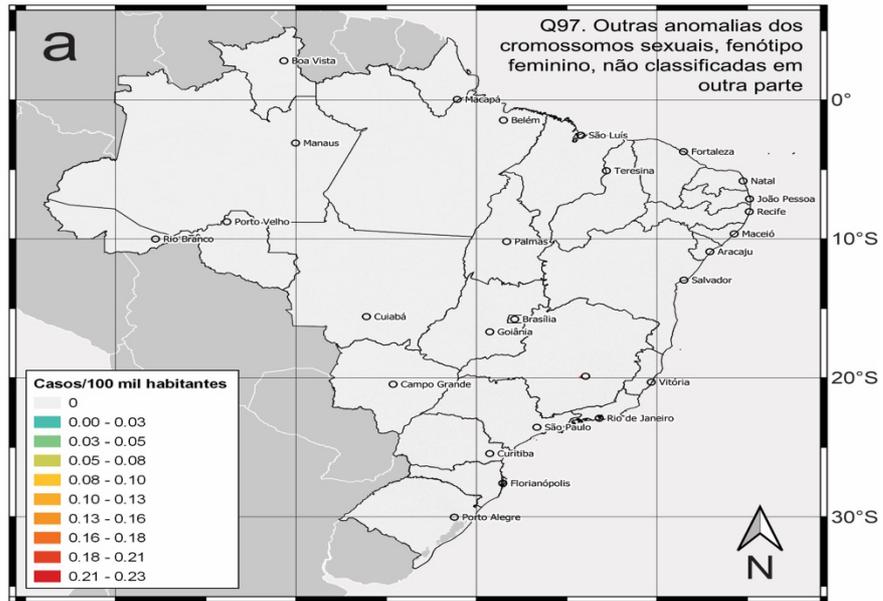






70°O 60°O 50°O 40°O

70°O 60°O 50°O 40°O



## APÊNDICE D – Cálculos Usados e Github<sup>11</sup>

Teste de normalidade de shapiro-wilk, usado para identificar se uma amostra pertence a uma população que se adequa a distribuição normal.

- 1) Ordena-se a amostra.
- 2) Calcular a soma de quadrados dos desvios:

$$S^2 = \sum_{i=1}^n (y_i - \bar{y})^2 = \sum_{i=1}^n y_i^2 - \frac{\left(\sum_{i=1}^n y_i\right)^2}{n}$$

- 3) Calcular b:

Em caso de n par n=2K:

$$b = \sum_{i=1}^k a_{n-i+1} (y_{n-i+1} - y_i)$$

Em caso de n impar n=2K+1:

$$b = a_n (y_n - y_1) + \dots + a_{k+2} (y_{k+2} - y_k)$$

- 4) Calcula a estatística:

$$W = \frac{b^2}{S^2}.$$

- 5) Tomada de decisão se o W calculado for maior que o W tabelado, aceita-se a hipótese de normalidade, se W calculado for menor que o W tabelado rejeita-se a hipótese de normalidade.

Teste de comparação múltipla de Kruskal-wallis é um teste não paramétrico usado para comparar 3 ou mais grupos, testando a hipótese que todos os grupos possuem funções de distribuições iguais. É um teste alternativo não paramétrico para o teste F utilizado no ANOVA.

- 1) Ordenamos todas as N observações das k amostras em ordem crescente, sendo  $r_{ij}$  como sendo o posto de  $X_{ij}$ :

$$R_i = \sum_{j=1}^{n_i} r_{ij} \quad \text{e} \quad R_i = \frac{R_i}{n_i}, \quad i = 1, \dots, k.$$

- 2) Cálculo da estatística H:

---

<sup>11</sup> <https://github.com/orgs/congenita/people/Tatianaalmeidacc>

$$H = \frac{\frac{12}{N(N+1)} \sum_{i=1}^k n_i \left( R_i - \frac{N+1}{2} \right)^2}{1 - \frac{\sum_{j=1}^g t_j^2 - t_j}{N^2 - N}} = \frac{\left( \frac{12}{N(N+1)} \sum_{i=1}^k \frac{R_i^2}{n_i} \right) - 3(N+1)}{1 - \frac{\sum_{j=1}^g t_j^2 - t_j}{N^2 - N}}$$

- 3) Encontramos os valores críticos para o nível de significância determinado.

Utilizamos estatística Qui-quadrado.

- 4) Se a estatística calculada H for maior que o valor da estatística qui-quadrado tabelada rejeitamos a hipótese de que as amostras provêm de populações igualmente distribuídas (ou seja que as amostras são iguais).

Teste de comparação de grupos independentes não paramétrico de wilcoxon é usado para comparar se duas amostras independentes são igualmente distribuídas.

- 1) Ordena-se os valores das duas amostras.
- 2) Associa-se os postos dos elementos da amostra 1 e 2.
- 3) Calcula-se os valores de:

$$U_m = S_m - \frac{1}{2}m(m+1)$$

$$U_n = S_n - \frac{1}{2}n(n+1).$$

Como  $S_m + S_n$  é a soma das duas amostras:

$$S_m + S_n = \frac{1}{2}(m+n)(m+n+1)$$

Os valores de  $U_m$  e  $U_n$  são relacionados:

$$U_m = mn - U_n$$

- 4) Calcula-se  $U_m$  e  $U_n$  e a estatística W é dada por  $U_n$ .
- 5) Calcula-se os valores críticos da estatística W, compara-se o valor da estatística W calculada e do valor tabelado, se o valor calculado for maior que o valor tabelado, rejeita-se a hipótese nula.

Qui-quadrado de associação: é usado para identificar se há associação entre a variável da linha e da coluna para tabelas de duas variáveis.

- 1) Organiza-se a tabela.
- 2) Aplica-se o cálculo para estatística qui-quadrado do teste:

$$Q_{obs}^2 = \sum_{i=1}^r \sum_{j=1}^c \frac{(O_{ij} - E_{ij})^2}{E_{ij}}$$

3) Calcula-se a estatística tabelada com  $(r-1)X(c-1)$  graus de liberdade.

Compara-se o valor calculado com o valor tabelado se o valor calculado for maior que o valor crítico tabelado, rejeita-se a hipótese nula.