

UNIVERSIDADE DO VALE DO RIO DOS SINOS — UNISINOS  
UNIDADE ACADÊMICA DE PESQUISA E PÓS-GRADUAÇÃO  
PROGRAMA DE PÓS-GRADUAÇÃO EM COMPUTAÇÃO APLICADA  
NÍVEL MESTRADO

NAIRA KAIESKI

**VIS-SAÚDE - UMA METODOLOGIA PARA VISUALIZAÇÃO E ANÁLISE DE  
DADOS DE SAÚDE PÚBLICA**

São Leopoldo  
2014



Naira Kaieski

**VIS-SAÚDE - UMA METODOLOGIA PARA VISUALIZAÇÃO E ANÁLISE DE  
DADOS DE SAÚDE PÚBLICA**

Dissertação apresentada como requisito parcial  
para a obtenção do título de Mestre pelo  
Programa de Pós-Graduação em Computação  
Aplicada da Universidade do Vale do Rio dos  
Sinos — UNISINOS

Orientador:  
Prof. Dr. Luiz Paulo Luna de Oliveira

Coorientador:  
Prof. Dra. Marta Becker Villamil

São Leopoldo  
2014

DADOS INTERNACIONAIS DE CATALOGAÇÃO NA PUBLICAÇÃO (CIP)

Kaieski, Naira

Vis-Saúde - Uma metodologia para visualização e análise de dados de saúde pública / Naira Kaieski — 2014.

81 f.: il.; 30 cm.

Dissertação (mestrado) — Universidade do Vale do Rio dos Sinos, Programa de Pós-Graduação em Computação Aplicada, São Leopoldo, 2014.

“Orientador: Prof. Dr. Luiz Paulo Luna de Oliveira, Unidade Acadêmica de Pesquisa e Pós-Graduação”.

1. Visualização de informações. Modelagem matemática. Monitoramento da saúde pública. Análise de componentes principais.. I. Título.

CDU 000.004

Bibliotecária responsável: Flávio Nunes — CRB 10/1298

(Esta folha serve somente para guardar o lugar da verdadeira folha de aprovação, que é obtida após a defesa do trabalho. Este item é obrigatório, exceto no caso de TCCs.)



*Ao meu esposo Jacques,  
pela parceria nas noites de branco,  
pela compreensão, carinho e estímulo.*





## **AGRADECIMENTOS**

Ao meu orientador Professor Doutor Luiz Paulo Luna de Oliveira e à Professora Doutora Marta Becker Villamil pelo conhecimento, dedicação e paciência que tiveram durante o desenvolvimento deste trabalho.



## RESUMO

Os sistemas computacionais provêm uma série de ferramentas que possibilitam a análise e a extração de conhecimento em um vasto e complexo conjunto de variáveis. O conjunto de dados que compõem um evento, em uma primeira visão sob a forma numérica pura pode não oferecer qualquer tipo de informação mais relevante que subsidie um aprofundamento do estudo. Contudo o emprego de técnicas de visualização em conjunto com metodologias de análise de dados, possibilitam uma compreensão mais sólida de um evento. Possíveis correlações entre as variáveis podem ser evidenciadas mediante o emprego de metodologias de análise, abrindo novas possibilidades de investigação. Enquanto que o emprego de técnicas de visualização de dados apropriadas ao contexto, momento e conjunto de variáveis podem evidenciar informações importantes sobre o evento que permaneciam à sombra dos números.

O Vis-Saúde desenvolvido neste trabalho utiliza tecnologias *open source* para o desenvolvimento de uma metodologia de análise e visualização de dados em um ambiente *web* onde, se disponíveis, os registros podem ser submetidos à análise e o resultado visualizado em tempo real. Esforços para empregar metodologias de análise e visualização em registros advindos da saúde pública são relevantes a fim de extrair conhecimento do complexo conjunto de dados disponíveis, com o objetivo de orientar a população e os gestores de saúde pública. A metodologia de análise de dados empregada neste trabalho é baseada na análise de componentes principais cujo emprego objetiva encontrar a relação de dependência entre um evento de interesse e outras variáveis relacionadas disponíveis para estudo. A utilização de técnicas de visualização de dados enriquece as possibilidades de compreensão da informação que está sendo estudada. No Vis-Saúde são disponibilizadas duas visualizações georreferenciadas distintas dos registros de saúde, onde ambas são baseadas na região onde o evento ocorreu e apresentadas sobre um mapa geográfico para facilitar a identificação da região de interesse. A primeira contempla a visualização das informações resultantes da análise de dependência entre as variáveis de estudo. A segunda visa facilitar a compreensão da dinâmica da incidência de uma enfermidade com base na apresentação da concentração de ocorrências destas, através de mapas de calor, onde os dados são agrupados em séries temporais e apresentados sob a forma de uma animação mediante a passagem de tempo.

A saúde pública é um segmento da administração governamental muito dinâmica, que está constantemente sujeita a situações endêmicas ou epidêmicas. Tal comportamento demanda um monitoramento constante por parte dos gestores a fim de identificar situações graves para as quais devem ser direcionadas ações e recursos. Para demonstrar o funcionamento da metodologia com dados reais, foram utilizados os registros quanto à incidência de dengue no Brasil. A dengue é uma enfermidade presente em todos os estados da federação e que representa um risco grave à saúde coletiva da população, uma vez que a incidência desta doença tem aumentado nos últimos anos.

**Palavras-chave:** Visualização de informações. Modelagem matemática. Monitoramento da saúde pública. Análise de componentes principais..



## ABSTRACT

Computer systems provide a set of tools that allow the analysis and extraction of knowledge in a large and complex set of variables. The set of data that form an event, in a first analysis under numerical form can not show any important information to support a deep study of the data. Although the use of visualization techniques with data analysis methodologies can allow a best comprehension about an event. Possible correlations between variables can be evidenced by the use of methodologies of analysis that open a set of new possibilities of research. The use of data visualization techniques according to the context, moment and set of variables can show important information about an event that was hidden by numbers.

Vis-Health shown in this paper used open source technologies to develop an analysis and visualization methodology of data in an web environment, where if the data are available, the registers can be processed and the result shown in real time. Efforts to utilize analysis and visualization methodologies in a set of public health data are relevant to extract knowledge of the complex data set available with the goal of guiding the population and public health managers. The methodology analysis used in this paper is based on principal component analysis with the aim to find the dependency relationship between an event of interest and other related variables available for study. The use of data visualization techniques enhances the possibilities of comprehension of the information that is being studied. Vis-Health provided two distinct geo-referenced views of health records, where both are based in the region where the event occurred and displayed on a geographical map for easy identification of the region of interest. The first visualization is showing the result of the dependency analysis between the variables of study. The second one has the aim to make more dynamic the comprehension about the incidence of a disease through the presentation of concentration of events occurred using heat map. In this visualization the data are grouped in temporal series and displayed with an animation with temporal passage.

The public health is a very dynamic segment of government administration that is always threatened with endemic or epidemic situations. This behavior requires a constant monitoring of managers to identify serious situations that can require more attention and resources. With the goal to demonstrate the operation of the methodology with real data, the registers about dengue incidence in Brazil were used. Dengue is a disease present in all Brazil's states and represents a serious risk to the collective health considering that the incidence has increasing in the latest years.

**Keywords:** Information visualization. Mathematical modeling. Monitoring of public health. Principal component analysis..



## LISTA DE FIGURAS

|  |    |
|--|----|
| Figura 1 – Número médio de casos de dengue notificados à OMS nos 30 países/territórios mais endêmicos, 2004-2010 . . . . .   | 23 |
| Figura 2 – Risco global de ocorrência de dengue . . . . .  | 24 |
| Figura 3 – Boletim Epidemiológico . . . . .  | 32 |
| Figura 4 – GHO - Vigilância global da influenza virológica . . . . .   | 33 |
| Figura 5 – Tela inicial do <i>site</i> HealthMap . . . . .   | 34 |
| Figura 6 – Tela inicial do <i>site</i> Flu Near You . . . . .  | 35 |
| Figura 7 – Tela inicial do <i>site</i> BioCaster . . . . .   | 36 |
| Figura 8 – Tela do sistema Google FluTrends . . . . .  | 37 |
| Figura 9 – <i>Site</i> Trulia com as ocorrências de crimes nos EUA . . . . .   | 38 |
| Figura 10 – <i>Site</i> TrackLytics com dados de saúde nos EUA . . . . .   | 39 |
| Figura 11 – Processo de visualização analítica . . . . .   | 45 |
| Figura 12 – <i>Tree map</i> : Estrutura hierárquica representando a organização de continente, país e cidade. . . . .  | 46 |
| Figura 13 – Mapa de calor gerado a partir de micro dados do DNA que representam os valores de expressão dos genes em diversas condições. . . . .                           | 47 |
| Figura 14 – Mapa de calor geográfico: Ocorrência do vírus HIV entre adultos (15-49 anos) em 2011. Regiões estabelecidas pela OMS. . . . .                                  | 48 |
| Figura 15 – Apresentação de pontos em uma matriz . . . . .   | 49 |
| Figura 16 – Apresentação de pontos em uma matriz com o mapa de calor . . . . .   | 49 |
| Figura 17 – Exemplo mapa de calor geográfico com a marcação de pontos de calor . . . . .   | 50 |
| Figura 18 – Arquitetura do serviço Google Fusion Tables . . . . .  | 52 |
| Figura 19 – Diagrama que representa as etapas da metodologia desenvolvida no Vis-Saúde . . . . .   | 53 |
| Figura 20 – (a) Dados de amostra para demonstração do modelo (b) Dados de amostra para demonstração do modelo normalizados . . . . .                                       | 54 |
| Figura 21 – Representação do hiperplano obtido para representação dos dados . . . . .  | 55 |
| Figura 22 – Representação da reta normal ao hiperplano . . . . .   | 55 |
| Figura 23 – Gráfico representando a relação de dependência entre a ocorrência de dengue e fatores climáticos . . . . .   | 56 |
| Figura 24 – Visualização de dados advindos dos relatórios de saúde pública - Concentração da incidência de uma enfermidade e seu comportamento ao longo do tempo . . . . . | 58 |
| Figura 25 – Diagrama do banco de dados relacional para teste da metodologia do Vis-Saúde . . . . .   | 60 |
| Figura 26 – Tela inicial do Vis-Saúde para análise dos dados . . . . .   | 62 |
| Figura 27 – Tela do Vis-Saúde com a apresentação dos dados de dependência entre variáveis . . . . .  | 63 |
| Figura 28 – Representação visual resultante da análise de todos os registros para cada estado brasileiro . . . . .   | 65 |
| Figura 29 – Registros gerados para testes - Rio Grande do Sul (Jan 2013) . . . . .   | 66 |

|   |    |
|---|----|
| Figura 30 – Resultado da análise da ocorrência da dengue em função de todas as demais variáveis de análise - Rio Grande do Sul (Jan 2013) . . . . .                 | 66 |
| Figura 31 – Resultado da análise da ocorrência da dengue em função da temperatura média, precipitação e renda da população - Rio Grande do Sul (Jan 2013) . . . . . | 67 |
| Figura 32 – Exemplo de dados disponíveis no Sinan referentes à incidência de dengue no Brasil no ano de 2006 . . . . .  | 68 |
| Figura 33 – Mapa com a localização das estações meteorológicas das quais podemos consultar dados no BDMEP . . . . .   | 70 |
| Figura 34 – Vis-Saúde - Análise com dados reais . . . . .   | 71 |
| Figura 35 – Vis-Saúde - Resultado detalhado da análise com dados reais . . . . .  | 72 |
| Figura 36 – Vis-Saúde - Resultado da análise para a cidade de São Paulo ao longo dos anos . . . . .   | 73 |
| Figura 37 – Vis-Saúde - Resultado da análise sem as duas variáveis de maior dependência . . . . .   | 74 |
| Figura 38 – Vis-Saúde - Resultado da análise considerando como método de padronização das variáveis a média e o desvio-padrão . . . . .                             | 74 |



## LISTA DE TABELAS

|   |    |
|---|----|
| Tabela 1 – Tabela de comparação entre soluções para visualização inicial dos dados de saúde pública, estes registros apresentados não são contemplados por uma metodologia de análise, assim os dados apresentados estão apenas agrupados geograficamente . . . . . | 35 |
|---|----|



## LISTA DE SIGLAS

|      |   |
|------|---|
| API  | Application Programming Interface (Interface de Programação de Aplicativos)                           |
| CDC  | Center for Disease Control and Prevention of US (Centro para Controle e Prevenção de Doenças dos EUA) |
| EHR  | Eletronic Health Record (Registro Eletrônico de Saúde)  |
| GHO  | Global Health Observatory (Observatório Mundial da Saúde)   |
| GIS  | Geographical Information System (Sistema de Informação Geográfica)                                    |
| GPS  | Global Positioning System (Sistema de Posicionamento Global)  |
| ICT  | Information and Communication Technology (Tecnologia de Informação e Comunicação)                     |
| IP   | Internet Protocol (Protocolo Internet)  |
| ITU  | International Telecommunication Union (União Internacional de Telecomunicações)                       |
| OGP  | Open Government Partnership (Parceria para Governo Aberto)  |
| OMS  | Organização Mundial da Saúde (World Health Organization)  |
| PCA  | Principal Componente Analysis (Análise de Componentes Principais)                                     |
| SGBD | Sistema Gerenciador de Banco de Dados   |
| SMS  | Short Message Service (Serviço de Mensagem Curta)   |
| SUS  | Sistema Único de Saúde  |
| TI   | Tecnologia da Informação  |
| TIC  | Tecnologia da Informação e Comunicação  |
| WHO  | World Health Organization (Organização Mundial da Saúde)  |



## SUMÁRIO

|          |   |           |
|----------|---|-----------|
| <b>1</b> | <b>INTRODUÇÃO</b>   | <b>21</b> |
| 1.1      | Motivação   | 23        |
| 1.2      | Problema  | 24        |
| 1.3      | Objetivos   | 24        |
| 1.4      | Organização do trabalho   | 25        |
| <b>2</b> | <b>REVISÃO BIBLIOGRÁFICA</b>  | <b>27</b> |
| 2.1      | Modelagem matemática  | 27        |
| 2.2      | Monitoramento da saúde pública  | 30        |
| 2.3      | Visualização de densidade em mapas  | 36        |
| <b>3</b> | <b>CONCEITOS BÁSICOS</b>  | <b>41</b> |
| 3.1      | Análise de componentes principais   | 41        |
| 3.1.1    | Cálculo da dependências entre as variáveis consideradas   | 43        |
| 3.2      | Visualização analítica de informações   | 44        |
| 3.3      | Mapa de calor   | 45        |
| 3.4      | Integração entre as técnicas de análise e visualização de dados   | 50        |
| 3.4.1    | Google API  | 50        |
| 3.4.2    | Google Maps API   | 51        |
| 3.4.3    | Google Fusion Tables  | 51        |
| <b>4</b> | <b>DESENVOLVIMENTO DO MODELO</b>  | <b>53</b> |
| 4.1      | Seleção das variáveis de análise  | 54        |
| 4.2      | Processamento dos dados   | 54        |
| 4.3      | Apresentação das informações  | 56        |
| 4.3.1    | Visualização de informações advindas da análise com PCA   | 56        |
| 4.3.2    | Visualização de informações advindas dos relatórios de saúde pública                                    | 57        |
| 4.3.3    | Ferramentas utilizadas no desenvolvimento   | 57        |
| <b>5</b> | <b>RESULTADOS</b>   | <b>61</b> |
| 5.1      | Descrição do funcionamento do ambiente do Vis-Saúde   | 61        |
| 5.2      | Homologação da metodologia de análise e visualização da dependência entre variáveis com dados fictícios | 62        |
| 5.3      | Demonstração com dados reais de saúde pública   | 65        |
| 5.3.1    | Fonte de dados reais para análise   | 66        |
| <b>6</b> | <b>CONCLUSÃO</b>  | <b>75</b> |
|          | <b>REFERÊNCIAS</b>  | <b>77</b> |



## 1 INTRODUÇÃO

A utilização de tecnologias digitais na área da saúde cresce rapidamente neste século como um novo paradigma para as práticas médicas (KIM; CHUNG; KIM, 2009) e cuidado com a saúde coletiva da população. Em um cenário onde a demanda por assistência médica tem se intensificado com o crescimento da população, aumento da longevidade e com o surgimento de novas patologias, vírus e suas mutações, o emprego de soluções digitais pode transformar todo o processo de cuidado com a saúde tornando-o mais eficiente, barato e de melhor qualidade (LIU; PARK; KRIEGER, 2012).

Neste trabalho é apresentado o Vis-Saúde, que consiste em uma metodologia de análise e visualização de informações oriundas de um complexo conjunto de variáveis. A metodologia de análise visa identificar o nível de correlação entre as variáveis estudadas utilizando o método matemático de análise de componentes principais (ACP). Tendo os dados sido analisados é necessário prover um meio de visualização das informações obtidas. As visualizações geradas constituem um componente importante da metodologia desenvolvida, pois a apresentação de um grande conjunto de números puros torna a compreensão da informação um processo complexo e cansativo. Contudo, o emprego de técnicas de visualização de dados, apropriadas ao conjunto de registros sob análise e contextualizadas, pode facilitar a tarefa de abstração de conhecimento uma vez que utilizam o sentido visual humano como um facilitador para perceber a informação e instigar a criatividade para refletir sobre o seu significado.

O Vis-Saúde possibilita duas representações visuais distintas de informações, a primeira é originada dos dados brutos, onde a concentração de eventos é apresentada sob a forma de mapas de calor utilizando séries temporais georreferenciadas. Através desta representação é possível compreender a incidência, a localização e propagação geográfica do evento ao longo do tempo. A segunda representação visual mostra a correlação entre as variáveis de estudo que resulta da análise dos registros utilizando a ACP. Todas as visualizações dos dados são georreferenciadas com base na região de ocorrência do evento e podem ser acessadas através de um ambiente *web* utilizando um navegador independente de plataforma ou dispositivo. Neste ambiente o usuário pode alterar o conjunto de dados sob análise e obter o resultado em tempo real.

Vários ensaios ligados a área de saúde pública buscam relacionar fatores sociais e climáticos à incidência de surtos de doenças, contudo a utilização da análise de componentes principais como proposto no Vis-Saúde, objetivando analisar a correlação entre as variáveis sob estudo, constitui uma nova abordagem com este mesmo intuito. A compreensão da relação de dependência entre as variáveis analisadas frente a um evento de saúde não esgotam as possibilidades de causa do evento mas abrem novas perspectivas de investigação a fim de traçar medidas mais efetivas de controle e prevenção ao surgimento e propagação de doenças.

A saúde pública é um campo multidisciplinar, cujo objetivo é promover a saúde da população, através de esforços comunitários organizados (ASCHENGRAU; SEAGE, 2008). Neste contexto, os serviços de vigilância epidemiológica, que mantêm registros contínuos dos núme-

ros de casos notificados de doenças infecto-contagiosas, constituem uma importante fonte de informações. No Brasil, várias doenças são de notificação compulsória e seus dados de incidência são coletados em intervalos regulares pré-estabelecidos. A monitoração do padrão temporal de incidência permite detectar flutuações sazonais e abruptas alertando os agentes de saúde para eventuais surtos de doenças infecto-contagiosas. O estudo da flutuação de diversas doenças e sua correlação sazonal com outros fatores, permite realizar previsões ou até mesmo estabelecer relações de causa e efeito. Essas informações provenientes da análise de séries temporais podem ser utilizadas sob a perspectiva de direcionar medidas preventivas em relação a ocorrência e propagação das doenças (MASSAD et al., 2004).

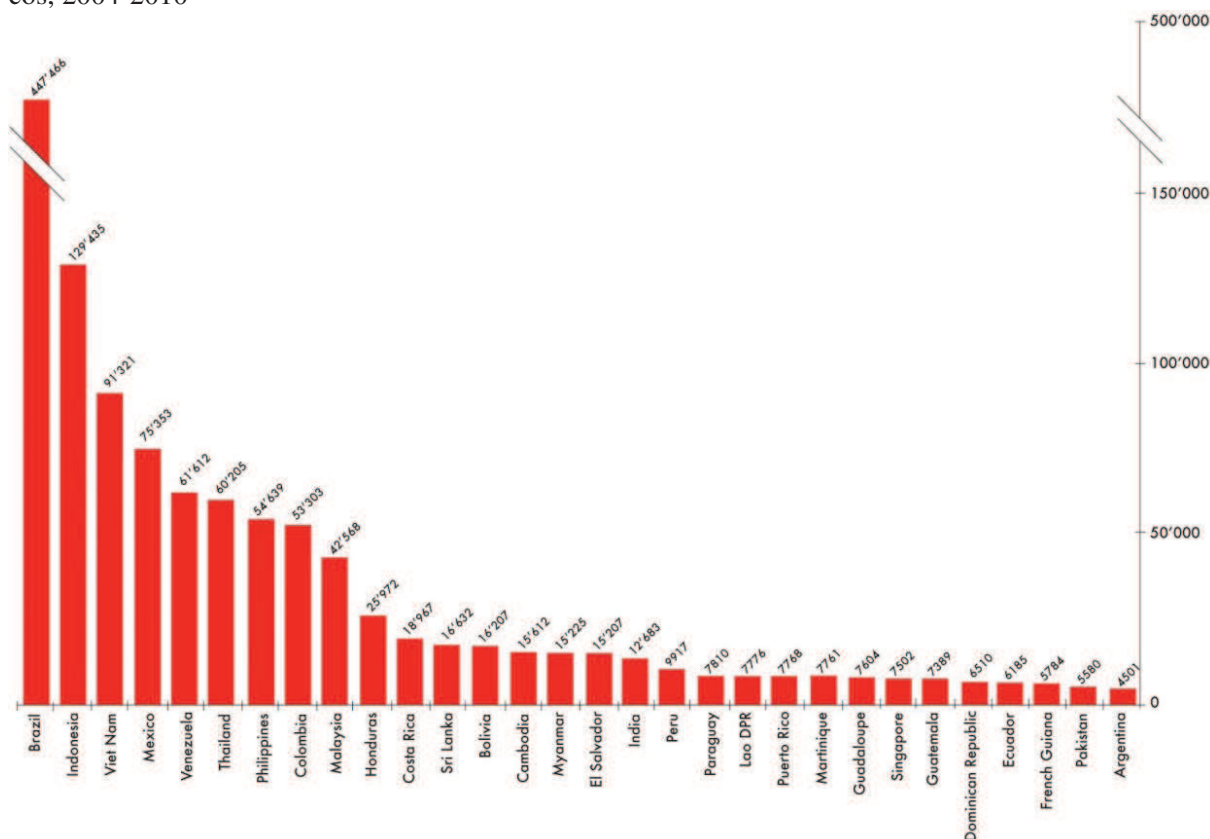
Uma importante fonte de dados, relacionados à saúde pública brasileira, utilizada no Vis-Saúde é o Sinan (Sistema de Informação de Agravos de Notificação) disponibilizado pelo governo. Este sistema disponibiliza informações sobre a ocorrência de diversas enfermidades e está acessível para qualquer cidadão que deseje consultá-las. É uma iniciativa do governo que visa uma maior transparência na administração pública, disponibilizando mais informações à população e incentivando a participação social (Controladoria Geral da União, 2013). Considerando que o objetivo é demonstrar a metodologia de análise e visualização de dados, apenas os registros de casos de dengue serão utilizados na demonstração do Vis-Saúde. A incidência da dengue será avaliada em relação a fatores climáticos e sua ocorrência em relação a análise de séries temporais em determinada região. Embora, historicamente, os casos de dengue tenham uma maior concentração em países do sudeste asiático, o Brasil se tornou um dos países com maior número de casos no mundo (RODRIGUEZ-BARRAQUER et al., 2011). A Figura 1, apresenta o número médio de casos de dengue notificados à Organização Mundial da Saúde (OMS) nos 30 países mais endêmicos entre os anos de 2004 a 2010. Com base nestes dados, o Brasil lidera o ranking negativo de incidência de dengue.

A escolha da dengue para demonstração da metodologia proposta no Vis-Saúde ganha ainda mais relevância quando aspectos globais sobre o risco de sua ocorrência são considerados. Segundo dados da OMS mais de 2,5 bilhões de pessoas, em torno de 40% da população mundial, estão suscetíveis a contrair alguma variante do vírus da dengue (World Health Organization, 2014). A Figura 2, mostra uma representação geográfica mundial das áreas com maior risco de ocorrência de dengue.

A dengue é uma doença causada por quatro sorotipos do vírus transmitidos por artrópodes (mosquito *Aedes Aegypti*) e sua incidência tem aumentado dramaticamente nos últimos 50 anos (KYLE; HARRIS, 2008). O crescimento populacional associado à urbanização desordenada e descontrolada, a redução de recursos para programas de controle e prevenção de doenças infecciosas principalmente em países com um clima tropical acentuam as condições ideais para a proliferação do *Aedes Aegypti*. Esses fatores explicam em grande parte a propagação internacional da dengue, mas é importante analisar e compreender a dinâmica da circulação do vírus em cada espaço geográfico e social (TEIXEIRA et al., 2009).



Figura 1 – Número médio de casos de dengue notificados à OMS nos 30 países/territórios mais endêmicos, 2004-2010



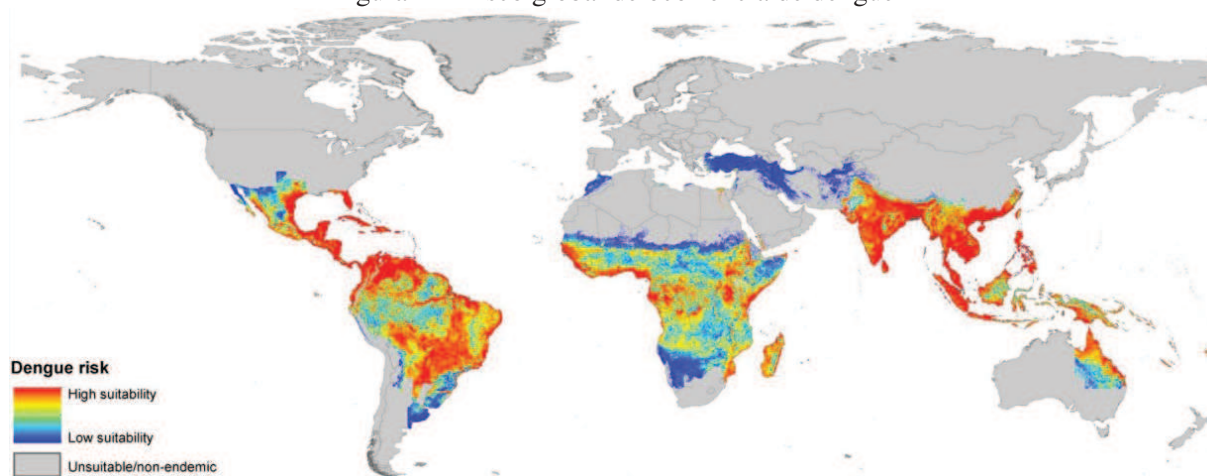
Fonte: (World Health Organization, 2012)

## 1.1 Motivação

O emprego de recursos computacionais na saúde é uma tendência cada vez mais relevante ao passo que as TICs (Tecnologias de Informação e Comunicação) estão se tornando mais ubíquas e pervasivas (WHEELER, 2013). Contudo, não basta apenas o emprego de tecnologia na criação de sensores, meios de armazenamento e transmissão de dados, é necessário o estudo e o desenvolvimento de metodologias de análise do abrangente e complexo conjunto de informações relativas à saúde pública. Monitorar e analisar os dados de saúde pública requer um grande esforço da comunidade de pesquisadores com vistas a encontrar meios de processar, analisar e tornar significativas as informações que podem ser abstraídas dos dados oriundos de saúde pública, diretamente ou da relação destas com outros fatores como os sociais, econômicos e climáticos. Compreender a ocorrência, o comportamento e a propagação de enfermidades ou surtos epidêmicos requer a análise não apenas dos dados de saúde mas também do contexto onde a ocorrência está inserida.

A motivação para a elaboração deste estudo advém da necessidade de explorar meios de analisar e visualizar os dados complexos provenientes da saúde pública e da correlação destes com outros fatores externos como os climáticos e sociais. Não esgotando todas as possibilidades

Figura 2 – Risco global de ocorrência de dengue



Fonte: (BHATT et al., 2013)

mas abrindo novas perspectivas de análise da relação de dependência que pode ser encontrada na ocorrência de determinados eventos de saúde em relação à variáveis externas. Considerando que os dados serão processados e analisados, é relevante considerar métodos de visualização das informações obtidas deste processamento. Assim, a apresentação das informações é tão importante quanto a análise dos dados propriamente, sendo necessário contemplar na visualização dos dados aspectos que objetivam facilitar o entendimento das informações por parte dos gestores de saúde e também da população em geral.

## 1.2 Problema

A ocorrência de um evento de saúde pública como um surto de dengue ou de outra doença infecciosa e/ou contagiosa dificilmente será um evento isolado. Logo, é vital a análise e o entendimento da relação de dependência entre fatores externos (climáticos, econômicos, sociais ou outros) com a incidência de enfermidades na população.

São dois os problemas a serem considerados, um deles é desenvolver e testar uma metodologia de análise da correlação entre as variáveis relacionadas a eventos de saúde pública com outros fatores como os climáticos e sociais. Outro é desenvolver e avaliar técnicas de visualização dos dados sob estudo. São dois os conjuntos de dados a serem visualizados, os resultantes do processamento de análise e os registros sobre a incidência de enfermidades. Para ambos os conjuntos de dados, o emprego de técnicas de visualização podem agregar valor na interpretação dos registros e compreensão de informações.

## 1.3 Objetivos

Desenvolver uma metodologia de análise e visualização do estudo da correlação entre as variáveis associadas a ocorrência de uma enfermidade em uma determinada região. Imple-

mentar técnicas de visualização de dados de saúde pública advindos dos relatórios oficiais do governo brasileiro. Tais dados serão analisado sob a perspectiva de séries temporais quanto a concentração de ocorrências de doenças e seu comportamento ao longo do tempo e do espaço geográfico. Através da metodologia de análise e visualização de informações, proporcionar um meio de estudo mais significativo da relação de dependência entre a incidência de enfermidades e variáveis climáticas.

#### **1.4 Organização do trabalho**

Este trabalho está organizado com a seguinte estrutura: no Capítulo 2 será descrita a revisão bibliográfica com os trabalhos relacionados desenvolvidos; no Capítulo 3 serão abordados os conceitos básicos empregados no desenvolvimento do presente trabalho; o Capítulo 4 apresenta a implementação da metodologia proposta para análise dos dados e da visualizações; o Capítulo 5 apresenta os resultados obtidos com análise de dados fictícios e reais; O Capítulo 6 versa sobre as conclusões do trabalho.



## 2 REVISÃO BIBLIOGRÁFICA

As justificativas de uma metodologia de análise e visualização de dados oriundo dos relatórios de saúde pública foram discutidas na Introdução. As principais necessidades para a existência de uma metodologia como a proposta no Vis-Saúde são registros de incidência de doenças, bases de dados climáticos, recursos de telecomunicação, métodos quantitativos (matemáticos) de análise de dados, ferramentas e métodos de visualização de informações. No que se segue, serão discutidas as possibilidades de viabilização do Vis-Saúde justificando o presente trabalho, ao colocá-lo sob a perspectiva dos trabalhos relacionados a este tema.

### 2.1 Modelagem matemática

O principal diferencial do Vis-Saúde é o emprego da Análise de Componentes Principais para a análise dos dados de saúde pública visando identificar a correlação entre as variáveis de estudo. Considerando especificamente a incidência de dengue para exemplificar o Vis-Saúde, neste Capítulo são apresentados diversos ensaios encontrados na literatura na mesma linha de pesquisa que buscam obter uma explicação através de uma modelagem matemática dos eventos relacionados à saúde pública frente a correlação com outros fatores como climáticos, sociais e econômicos. As abordagens aritméticas empregadas são diversas e os resultados encontrados nem sempre apontam em uma mesma direção. O sucesso na modelagem de um surto epidêmico poderia resultar em sistemas de predição destas situações, contudo os aspectos e fatores que envolvem a saúde são complexos e muitas vezes sazonais, dificultado sua análise de causalidade.

A utilização de modelos matemáticos na medicina e na epidemiologia é quase tão antiga quanto na física. Em 1906, W. H. Hamer estabeleceu o princípio de ação das massas, um importante conceito de epidemiologia matemática, que sugere que o desenvolvimento de uma epidemia depende da taxa de contato entre indivíduos suscetíveis e infecciosos. O princípio de Hamer foi originalmente formulado através de um modelo de tempo discreto, mas, em 1908, Sir Ronald Ross (que descobriu que a malária é transmitida por mosquitos) o generalizou para tempo contínuo, em seus trabalhos sobre a dinâmica da malária. Em 1927, W. O. Kernack e A. G. McKendrick estenderam a teoria com o princípio do limiar, estabelecendo que a introdução de indivíduos infecciosos em uma comunidade não pode levar a um surto epidêmico a menos que a densidade de indivíduos suscetíveis esteja acima de um certo valor crítico (MASSAD et al., 2004).

Em especial nos países com maior incidência de dengue, diversos estudos tem sido desenvolvidos com o intuito de avaliar a ocorrência de dengue em função de fatores climáticos, sociais e econômicos. Dulay, Bautista e Teves (2013) em seu estudo, mostraram a relação entre os fatores climáticos (temperatura máxima, temperatura mínima, precipitação e umidade relativa) com a incidência da dengue e sua variação mais perigosa, a dengue hemorrágica. Este estudo foi realizado com dados da cidade de Iligan, Filipinas, nos anos de 2005 à 2009. A cor-

relação de Pearson foi utilizada para explorar a associação primária entre a incidência dengue e os fatores climáticos enquanto que uma análise de regressão múltipla foi usada para ajustar o modelo estatístico. A análise mostrou que a temperatura máxima, temperatura mínima, precipitação e umidade relativa são preditores significativos da ocorrência de dengue na cidade de Iligan. Sendo a temperatura máxima e temperatura mínima fortemente correlacionadas com o número de casos de dengue, enquanto que a precipitação e umidade relativa são considerados insignificantes. Já a dengue hemorrágica apresentou uma maior correlação com a temperatura mínima e umidade relativa sendo que a temperatura máxima mostrou-se insignificante, neste caso. Diante destes resultados, foi possível concluir no estudo em questão que a temperatura máxima, temperatura mínima e umidade relativa são linearmente relacionados com os casos de dengue na cidade de Iligan. Isso significa que um aumento destas variáveis indicará um aumento nos casos de dengue.

A pesquisa realizada por Machado-Machado (2012) teve como objetivo identificar as áreas propensas à ocorrência de dengue baseado na influencia de variáveis climáticas (19 no total) e socioeconômicas (4 no total) no México. A abordagem desta pesquisa é baseada na modelagem de distribuição de espécies utilizando Maxent, um algoritmo de aprendizado de máquina. A análise de componentes principais foi utilizada neste ensaio de uma forma diferente da proposta no Vis-Saúde, sendo empregada para reduzir o número de variáveis a serem analisadas produzindo novas variáveis independentes. Dois modelos foram gerados, o primeiro utilizava quatro componentes principais (CP) derivadas apenas das variáveis climáticas, enquanto que o segundo usava cinco CP originadas das variáveis climáticas e socioeconômicas. Como resultado da pesquisa, foi possível observar que as variáveis climáticas temperatura e precipitação foram mais relevantes que as socioeconômicas analisadas frente à ocorrência de casos da dengue.

O trabalho de Gomes, Nobre e Cruz (2012) focou em estudar o efeito de fatores sazonais, através da relação entre variáveis climáticas e o risco de dengue na cidade do Rio de Janeiro, com registros de dados entre os anos de 2001 a 2009. Foram utilizados modelos lineares generalizados, com a distribuição Poisson e a binomial negativa. O modelo com melhor ajuste foi o controlado por variáveis indicadoras do ano, que revelou como mais significativas as variáveis de temperatura mínima e precipitação. Neste estudo os autores defasaram os dados climáticos em um mês o que demonstrou a forte relação entre a temperatura mínima, ocorrência de chuvas e a incidência de dengue. Os dados revelaram que o aumento de um grau na temperatura mínima em um mês leva ao aumento de 45% no número de casos de dengue no mês seguinte, enquanto o aumento em 10 milímetros na precipitação leva ao aumento de 6% no número de casos de dengue no mês seguinte.

Valadares, Filho e Peluzio (2013) demonstram em seu ensaio uma tentativa de descrever a incidência da dengue e investigar sua correlação com a densidade vetorial e condições climáticas em dois municípios do Estado do Tocantins. O método matemático empregado para analisar os dados foi a correlação de Pearson sendo aplicado o teste qui-quadrado para investigar a tendência dos coeficientes de incidência de dengue ao longo dos anos. Os resultados

obtidos indicam que não houve correlação significativa entre o coeficiente de incidência da dengue e o índice de infestação predial, a precipitação pluviométrica média e a temperatura média. Logo as condições climáticas não se apresentaram como fatores determinantes na proliferação da dengue, segundo os dados analisados, mas possibilitam as condições ideais para reprodução do vetor.

No estudo de Monteiro et al. (2009) que objetivou descrever os indicadores epidemiológicos e vetoriais da dengue em Teresina-PI, no período de 2002 à 2006, foram analisados dados referentes à ocorrência da dengue, precipitação pluviométrica, temperatura, índice de infestação predial e índice de pendência. Onde o método matemático utilizado para avaliação, foi a correlação de Spearman através da qual foi encontrada correlação positiva entre a incidência de dengue, pluviosidade e temperatura, porém não houve associação entre índice de infestação predial e índice de pendência por estrato. No período analisado, a dengue na cidade de Teresina apresentou maior incidência no primeiro semestre de cada ano, coincidindo com o período de maior índice pluviométrico e de infestação predial. As condições climáticas caracterizadas pela precipitação pluviométrica, temperatura elevada e umidade do ar, em geral mostraram relação positiva com a transmissão da dengue.

A incidência de dengue está diretamente relacionada com o mosquito *Aedes*, que é o seu vetor de transmissão. A natureza sazonal desta enfermidade pode refletir a influência do clima sobre o ciclo de transmissão, onde o aumento da temperatura e da precipitação pode ocasionar o aumento da população de mosquitos (JOHANSSON; DOMINICI; GLASS, 2009). Segundo dados do Centro de Controle e Prevenção de Doenças Americano (CDC) os mosquitos transmissores da dengue são altamente sensíveis às condições ambientais. A temperatura, precipitação e umidade constituem as condições fundamentais para a sobrevivência, reprodução e desenvolvimento do mosquito. Sendo que temperaturas mais elevadas reduzem o tempo necessário para que o vírus se multiplique e propague no mosquito. Se o mosquito torna-se infeccioso mais rápido devido a temperaturas mais elevadas, ele tem mais chance de infectar uma pessoa antes de morrer (CDC, 2012).

Algumas pesquisas, (HARTLEY; DONNELLY; GARNETT, 2002) e (HOPP; FOLEY, 2003), que buscam uma modelagem teórica da dinâmica de transmissão da dengue com base na biologia do mosquito ressaltam a importância da temperatura e da precipitação na determinação dos padrões de transmissão mas sem uma evidência empírica. Em escala global vários ensaios têm destacado algumas características climáticas comuns de áreas onde ocorre a transmissão da dengue (HALES et al., 2002), (PETERSON et al., 2005) e (ROGERS et al., 2006). Enquanto estudos longitudinais têm mostrado consistentemente que a temperatura e a precipitação estão correlacionadas com a transmissão da dengue, mas não demonstraram consistência no que diz respeito às suas funções (HAY et al., 2000), (CAMPBELL et al., 2013), (STEWART-IBARRA; LOWE, 2013), (TUN-LIN; BURKOT; KAY, 2000), (KEATING, 2001), (THAMMAPALO et al., 2005), (LOWEA et al., 2011), (PROMPROU; JAROENSUTASINEE; JAROENSUTASINEE, 2005).



No estudo de Promprou, Jaroensutasinee e Jaroensutasinee (2005), a precipitação mensal acumulada e a temperatura estavam positivamente correlacionadas com o aumento da transmissão de dengue no sul da Tailândia. Já no Golfo da Tailândia, no entanto, o número de dias chuvosos (independentemente da quantidade) e temperatura mínima foram encontrados como positivamente associado com a incidência de dengue. Outro estudo, mais ao norte da Tailândia, descobriu que a temperatura teve um efeito negativo sobre a transmissão da dengue (NAKHA-PAKORN; TRIPATHI, 2005). Stewart-Ibarra e Lowe (2013) concluíram que a precipitação e a temperatura mínima estavam associados positivamente com a incidência de dengue no litoral do Equador. Keating (2001) usando uma análise de regressão linear multivariada para determinar a relação entre a temperatura média mensal defasada e a incidência mensal de dengue em Porto Rico, encontrou indícios de uma autocorrelação positiva. Thammapalo et al. (2005) em sua análise sobre os registros da Tailândia descobriram que um aumento na temperatura está associada a maior incidência de dengue em 9 províncias, enquanto que o aumento da precipitação foi associado a uma diminuição da incidência de dengue em 7 províncias. Sendo que a análise por região mostrou que a incidência de dengue estava negativamente associada com o maior índice de chuvas na região Sul, mas positivamente associado com temperaturas elevadas nas regiões centro e norte.

As conclusões sobre a relevância relativa das diferentes variáveis (climáticas, sociais e econômicas) não são consistentes no sentido de estabelecer um padrão absoluto nos estudos publicados sobre a incidência de dengue. Essa variação pode estar associada a fatores específicos das regiões sob análise como proximidade com zonas tropicais, variações das estações do ano, concentração da população entre outras. Variadas abordagens matemáticas têm sido empregadas na tentativa de modelar eventos na área da saúde pública, contudo a análise da correlação entre as variáveis de interesse como a proposta no Vis-Saúde através da Análise Componentes Principais se apresenta como uma opção inovadora nesta área de pesquisa proporcionando entre outros fatores a possibilidade de processamento em tempo real de um grande conjunto de dados e visualização gráfica dos resultados.

## **2.2 Monitoramento da saúde pública**

Um dos objetivos do Vis-Saúde, além da análise dos dados, é proporcionar formas de apresentação das informações oriundas diretamente dos registros de saúde pública ou da análise matemáticas destes de forma significativa. Dentro desta metodologia os dados são apresentados em dois momentos distintos, onde o primeiro contempla apenas a visualização dos registros "puros" sobre a incidência de enfermidades, compondo um estado inicial de análise e interpretação visual dos registros que são agrupados em regiões de interesse, georreferenciados e apresentados sob a forma de séries temporais. Nesta visualização a concentração da incidência de uma doença, sua propagação e comportamento ao longo do tempo é apresentada através de mapas de calor. Um segundo estágio de visualização contempla o resultado da modelagem matemática



aplicada aos dados onde é representado graficamente a correlação entre as variáveis de análise nas regiões de interesse. Neste contexto, este Capítulo apresenta algumas iniciativas que visam apresentar graficamente os dados de saúde pública sem necessariamente ser aplicado aos dados uma metodologia de análise mais apurada.

O acesso aos dados consolidados de saúde pública auxilia no processo de identificação de endemias, epidemias e pandemias e proporcionam uma forma de analisar seu comportamento e o impacto sobre a população ao longo do tempo. Contudo as informações sobre a saúde pública são complexas e numerosas, logo um cidadão comum pode enfrentar dificuldades em interpretar um conjunto grande de números tabulados. Neste contexto, técnicas de visualização de dados podem ser empregadas para facilitar a apresentação e interpretação destes dados. Em uma forma mais direta, mesmo sem um algoritmo de processamento de dados, o emprego de representações gráficas dos registros de saúde pública podem ser usados com êxito no objetivo de facilitar a interpretação das informações intrínsecas contidas nos relatórios de saúde pública. Agrupar, colorir e georreferenciar os dados, embora simples, agrega valor aos registros apresentados e facilitam sua compreensão. Munidos de informações confiáveis, os gestores de saúde podem direcionar esforços a fim de promover o bem-estar das pessoas, seja com ações de prevenção ou controle de doenças transmissíveis ou crônicas. Nesta seção serão exemplificadas algumas iniciativas que disponibilizam informações sobre a manifestação de doenças através de relatórios ou *sites* disponíveis na internet. O Vis-Saúde, contempla além da modelagem matemática da correlação das variáveis associadas ao evento de saúde também uma visualização inicial dos registros quanto a concentração da incidência de eventos.

O governo brasileiro, através do Ministério da Saúde, elabora e disponibiliza há 43 anos o Boletim Epidemiológico que divulga informações sobre a ação de diversos vírus no Brasil e no mundo. O objetivo deste boletim é retornar aos serviços e profissionais de saúde, os dados informados sobre as doenças depois de consolidados e analisados. Essa ação, de devolver as análises aos fornecedores dos dados, é considerada fundamental para valorizar o ato de notificar e lhe dar sentido, demonstrado sua utilidade para a saúde pública (Secretaria de Vigilância em Saúde, 2012). Por exemplo, o boletim, Secretaria de Vigilância em Saúde (2013), traz dados sobre a circulação do vírus influenza, onde é possível perceber uma maior concentração de atendimentos e óbitos por Síndrome Respiratória Aguda Grave (SRAG) na região sudeste do Brasil, especificamente na cidade de São Paulo. Este dados são compilados mensalmente em um arquivo disponibilizado no *site* do Ministério da Saúde onde as informações são apresentadas em formato textual e gráfico. A Figura 3 mostra um exemplo de como os dados são apresentados no boletim.

A OMS através do Observatório Mundial da Saúde (GHO) disponibiliza uma série de mapas com estatísticas sobre os principais temas relacionados à saúde. No entanto, como estes dados são submetidos pelos governos, existe uma defasagem significativa entre a situação atual da saúde e os dados apresentados nestes mapas. A Figura 4, apresenta um exemplo de um mapa da OMS com os dados de exames com resultado positivo para o vírus influenza no mundo.

Figura 3 – Boletim Epidemiológico



**Influenza 2013: Monitoramento da Semana Epidemiológica 16**

**Informe<sup>2</sup>**

- Na China, até 23 de abril de 2013, 108 casos de influenza A(H7N9) foram confirmados por critério laboratorial, incluindo 22 óbitos. A Organização Mundial de Saúde (OMS) mantém o posicionamento de que não há recomendação de restrições de viagens ou comércio de produtos oriundos da China.

**Contexto Internacional<sup>1</sup>**

Até 13 de abril de 2013, foi observada a seguinte circulação do vírus influenza nas Américas:

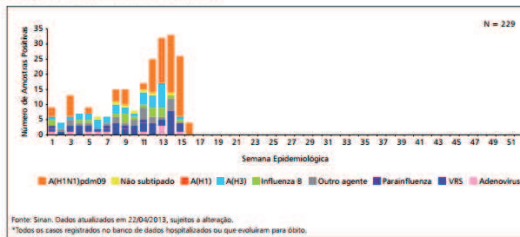


Figura 5 – Casos de SRAG hospitalizados\* segundo vírus identificado e semana epidemiológica do início dos sintomas. Brasil, 2013 (até a SE 16).

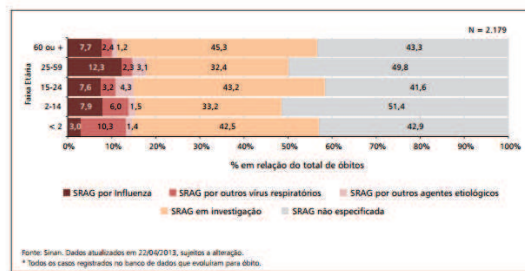
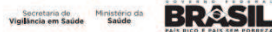


Figura 6 – Casos de SRAG hospitalizados\* por faixa etária e agente etiológico. Brasil, 2013 (até a SE 16).

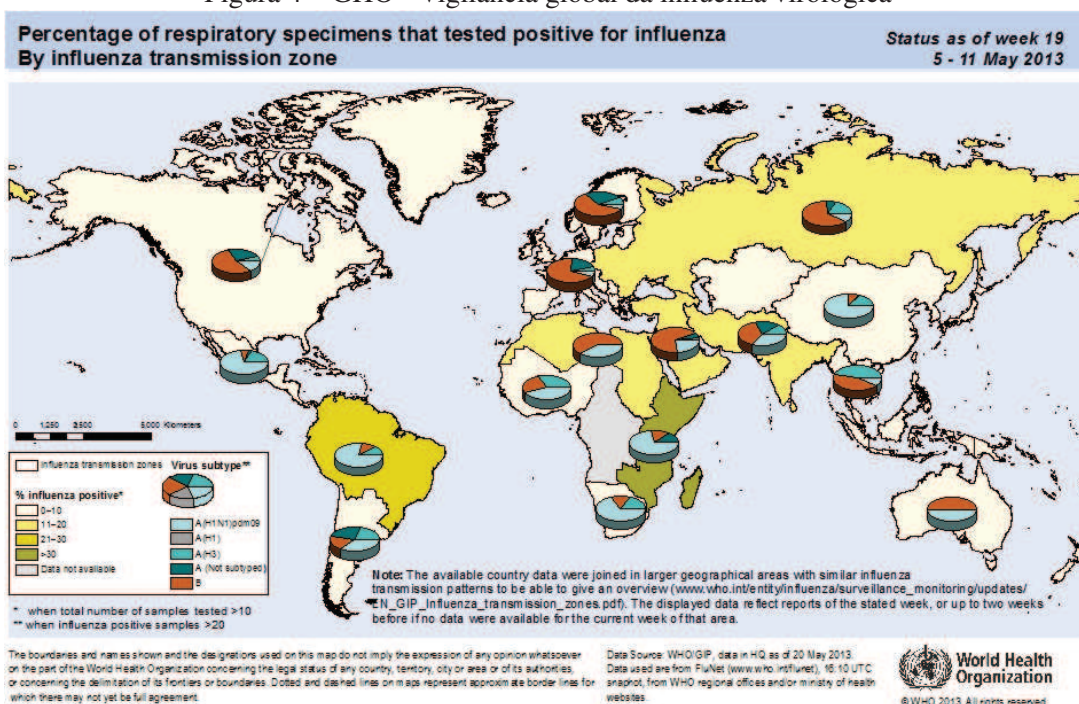


Fonte: (Secretaria de Vigilância em Saúde, 2013)

Outras iniciativas contemplam o desenvolvimento de aplicativos, disponíveis na internet, que possibilitam visualizar a ação de doenças no mundo com o auxílio de mapas geográficos. As informações a serem apresentadas nestes aplicativos são provenientes de distintas fontes. A técnica de visualização de dados utilizando mapas para explicar e prever os padrões de propagação de uma doença e para apoiar o raciocínio epidemiológico está bem estabelecida na pesquisa médica (KOCH, 2011). Utilizando mapas geográficos na medicina é possível explorar a relação entre a incidência observada e causa suspeita.

Considerando aplicativos relacionados ao monitoramento da saúde pública, o HealthMap apresenta sobre um mapa geográfico marcadores de alertas de enfermidades onde a concentração do número de ocorrências de cada doença é representado por uma coloração diferenciada em cada marcador. A Figura 5, apresenta uma tela do aplicativo HealthMap com os marcadores de eventos relacionados à saúde. As informações utilizadas neste sistema são provenientes de diferentes fontes, incluindo agregadores de notícias *online*, relatórios de testemunhas oculares, discussões com especialista e relatórios oficiais. O sistema é atualizado constantemente (24 horas por dia 7 dias por semana) através de um processo automatizado a fim de disseminar informações *online* sobre as doenças emergentes, facilitando a detecção precoce de ameaças

Figura 4 – GHO - Vigilância global da influenza virológica



Fonte: (World Health Organization, 2013)

globais de saúde pública (FREIFELD; BROWNSTEIN, 2013a).

O site "Flu Near You" (FREIFELD; BROWNSTEIN, 2013b), é administrado pelo HealthMap em parceria com a Associação Americana de Saúde Pública e o fundo Skoll contra ameaças globais. Este site apresenta estatísticas quanto a ocorrência de gripe nos Estados Unidos e Canadá de forma semelhante ao HealthMap. Neste sistema o usuário pode visualizar a evolução dos casos de gripe através de uma animação com a passagem de tempo por semana, sendo a concentração de registros demarcada em cada estado. A Figura 6, mostra uma tela do sistema onde é possível visualizar os marcadores sobre um mapa geográfico que indicam a concentração de registros de informações.

O site BioCaster (BIOCASTER, 2013), oportuniza a visualização, através de marcadores em um mapa geográfico mundial, de eventos, sejam eles terremotos, tufões, eventos nucleares, químicos ou biológicos que afetam humanos, animais ou plantas. Este sistema não é especializado em eventos relacionados à saúde pública embora estes também possam ser visualizados. O BioCaster apresenta uma cobertura mundial e possibilita a visualização dos eventos em séries temporais com animação em função da passagem de tempo. Algumas das fontes de informações deste sistema são: ProMED-mail, Google News, a Organização Mundial de Saúde, MedISys, Baidu News, SOSO.com e a rede de integração de dados Wildlife. A Figura 7, apresenta a tela inicial do site BioCaster com alguns marcadores de eventos.

O Google por sua vez, desenvolveu uma ferramenta experimental chamada Google Trends que possibilita investigar a atividade de enfermidades como gripe e dengue com base nos termos de pesquisa utilizado pelos usuários do serviço de busca do Google em determinadas regiões

Figura 5 – Tela inicial do *site* HealthMap

Fonte: Site HealthMap (<http://www.healthmap.org>)

do mundo. Na Figura 8, é apresentada um tela da ferramenta Google Flu Trends. Os dados são exibidos em um mapa geográfico com a delimitação dos países. A atividade da gripe é classificada em 5 níveis, onde uma cor diferente é atribuída a cada nível, quando o usuário do sistema clica sobre um país, uma nova tela é apresentada com dados específicos desta região e gráficos que indicam a atividade estimada da gripe em cada mês no país selecionado (GOOGLE, 2013).

Baseado nas iniciativas de visualização de incidentes e ocorrências de eventos de saúde apresentadas neste Capítulo foi elaborada a Tabela 1, que destaca as principais diferenças entre as iniciativas citadas e o Vis-Saúde no que tange a apresentação inicial dos dados sem uma análise matemática dos mesmos. A forma de visualização dos dados proposta no Vis-Saúde emprega uma técnica de representação de concentração de ocorrências através de mapas de calor, que utiliza uma animação baseada em séries temporais georreferenciadas considerando a região onde o evento ocorreu. A concentração da incidência de uma enfermidade será apresentada sobre um ponto geográfico da cidade ou estado onde o paciente reside, sem caracterizar a localização da residência da mesma. Os dados a serem apresentados na visualização inicial proposta são oriundos dos relatórios consolidados de saúde pública. Ao usuário do sistema será permi-



Figura 6 – Tela inicial do *site* Flu Near You

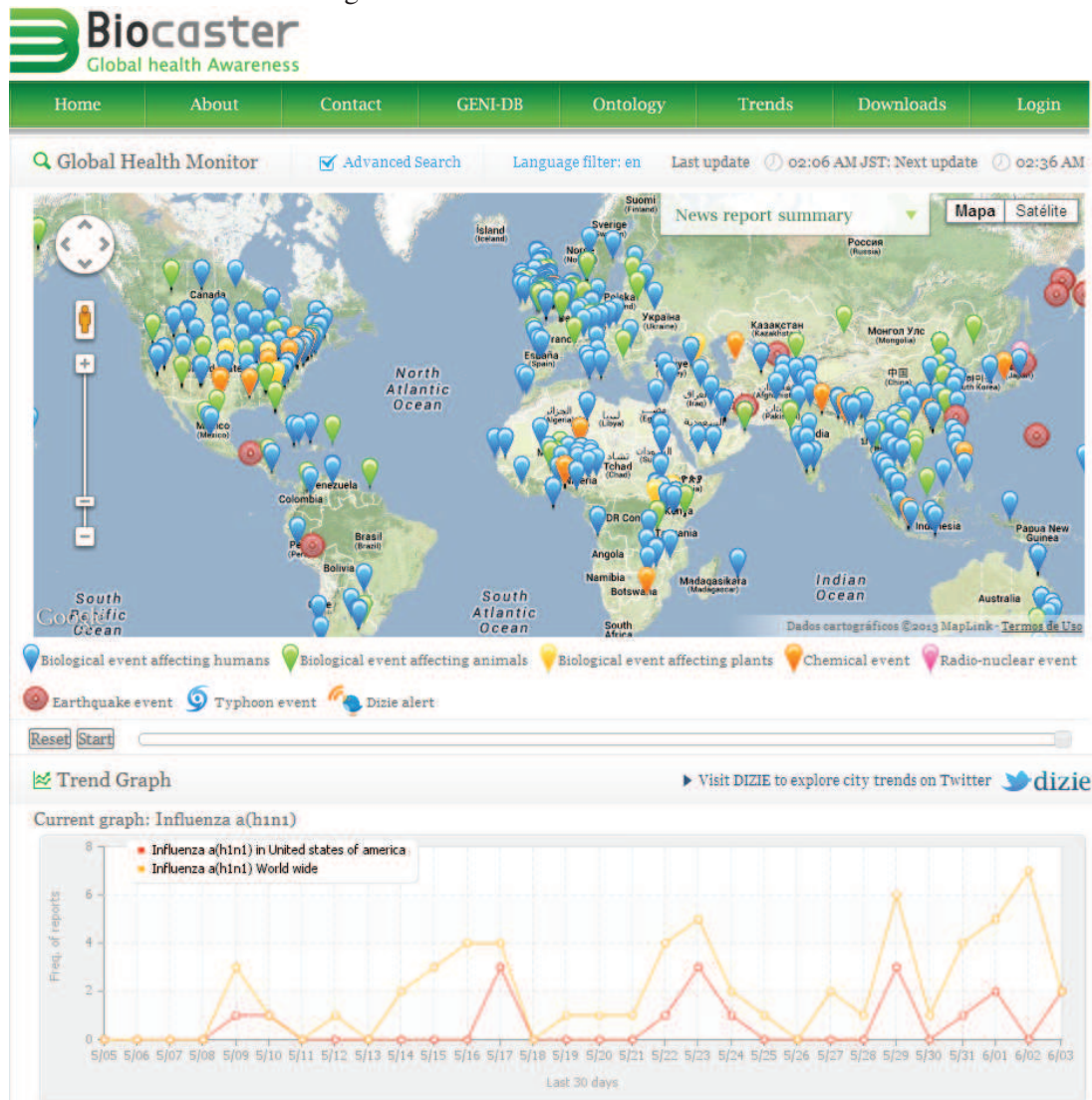
Fonte: *Site* Flu Near You (<https://flunearyou.org>)

tida a personalização da visualização (ou animação) com a seleção do período de tempo a ser visualizado. Como as informações quanto ao número de ocorrências de uma enfermidade não caracterizam um indivíduo em particular e não implicam em nenhuma informação privada que possa causar constrangimento, esta visualização pode ser publicada para uso amplo e irrestrito da população.

Tabela 1 – Tabela de comparação entre soluções para visualização inicial dos dados de saúde pública, estes registros apresentados não são contemplados por uma metodologia de análise, assim os dados apresentados estão apenas agrupados geograficamente

| Solução       | Apresentação dos dados   | Fonte de dados                                   | Série histórica | Possui animação |
|---------------|--|--|-----------------|-----------------|
| Vis-Saúde     | Mapa de calor (com marcadores em gradiente em cada ponto) sobre um mapa geográfico | Relatório de dados consolidados de saúde pública | Sim             | Sim             |
| HealthMap     | Marcadores sobre um mapa geográfico  | Notícias publicadas na internet                  | Não             | Não             |
| Fly Near You  | Marcadores sobre um mapa geográfico  | Notícias publicadas na internet                  | Sim             | Não             |
| BioCaster     | Marcadores sobre um mapa geográfico  | Notícias publicadas na internet                  | Sim             | Sim             |
| Google Trends | Mapa de calor geográfico (delimitação de áreas geográficas inteiras) e gráficos    | Termos de pesquisa no Google                     | Sim             | Não             |

Nesta seção foram apresentados alguns sistemas que visam apresentar graficamente em mapas geográficos dados quanto à incidência de eventos, relacionados exclusivamente à saúde

Figura 7 – Tela inicial do *site* BioCaster

Fonte: *Site* BioCaster (<http://born.nii.ac.jp>)

pública ou não. A fonte de dados utilizados por alguns destes sistemas nem sempre podem ser considerados tão confiável quanto os dados utilizados pelo Vis-Saúde que são relatórios governamentais. A visualização do registros de eventos de saúde pública organizados por região podem ser utilizados como uma fonte importante de informações para orientar investigações mais profundas e o destino de ações emergenciais de tratamento do evento.

### 2.3 Visualização de densidade em mapas

A representação gráfica inicial dos dados de saúde pública apresentada no Vis-Saúde tem por objetivo caracterizar a localização e gravidade da incidência das enfermidades e seu comportamento ao longo do tempo, sem o emprego de uma metodologia de análise matemática dos dados. Esta apresentação básica, embora simples, permite que a pessoa interessada tenha

Figura 8 – Tela do sistema Google FluTrends



Fonte: Site Google FluTrends (<http://www.google.org/flu-trends>)

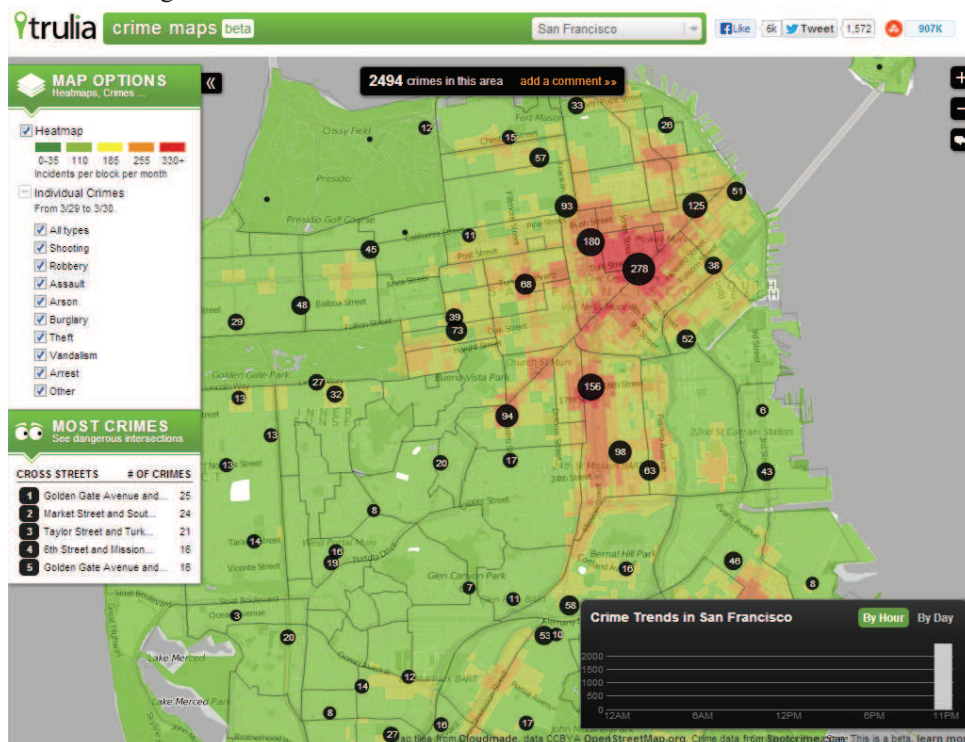
uma visão do panorama atual e histórico dos dados que podem ser submetidos a uma posterior análise a fim de estudar a correlação das variáveis de interesse ao fenômeno sob análise.

Uma das formas de visualização de dados que tem sido amplamente explorada é a baseada em sistemas de informação geográficas que permitem ao usuário do sistema compreender as informações que lhe são apresentadas com base na localização de sua ocorrência. Esta técnica de visualização foi exemplificada em parte na Seção 2.2 deste trabalho. É possível explorar a apresentação dos dados de forma a representar a densidade de eventos em determinada região, para este propósito é comum a utilização de mapas de calor sobre mapas geográficos. Os mapas de calor são formas de representar graficamente a concentração de quaisquer eventos, e consiste em uma camada de visualização com cores que é aplicada sobre o objeto no qual ocorre o evento. Como objeto, sobre o qual as informações serão representada podem ser utilizados mapas geográficos, *sites*, representações de objetos ou corpos, locais de prática de esporte, superfícies entre outras opções. A concentração de eventos é representada por cores, sendo que as mais frias como azul e verde representam uma menor concentração de eventos e cores mais quentes como tons vermelho significam uma maior densidade de eventos. Diversos sistemas têm sido desenvolvidos para apresentar diferentes concentrações de eventos utilizando como forma de representação de dados os mapas de calor. Esta técnica também é empregada no Vis-Saúde para a representação gráfica inicial dos registros de saúde pública no que se refere a concentração da incidência de casos de dengue em determinada região.

Como exemplo de aplicações que utilizam mapa de calor temos *sites* como o Trulia que apresenta sobre um mapa geográfico um mapa de calor com a concentração de ocorrência de crimes registrados nos departamentos de polícia dos Estados Unidos. A Figura 9, apresenta um exemplo da visualização de crimes proposta pelo Trulia. Já o WikiCrimes, apresenta a densi-

dade de crimes e permite a participação do usuário no registro de ocorrências (WIKICRIMES, 2013). Mapas de calor tem sido utilizados também no esporte para analisar a movimentação dos desportistas, em *sites* para avaliar as áreas mais acessadas e o comportamento do usuário no *site* (SESSIONCAM, 2013) e em projetos que envolvem informação à população como transportes (BLOCH, 2010), saúde (CDC, 2013), clima, investimentos, governança entre outros assuntos.

Figura 9 – Site Trulia com as ocorrências de crimes nos EUA



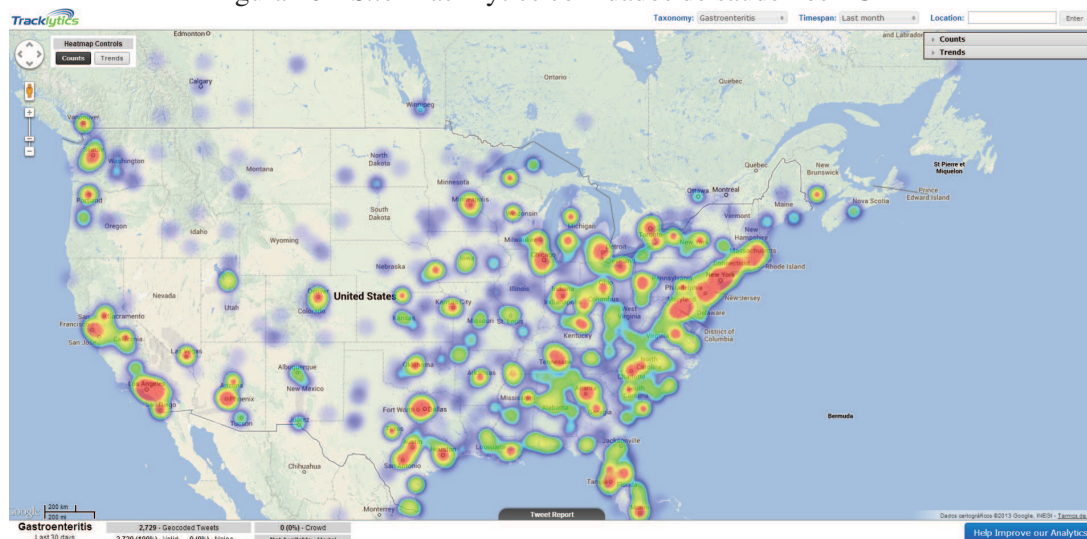
Fonte: Site Trulia Crimes (<http://www.trulia.com/crime>)

Na área da saúde, a representação da densidade de dados através de mapas de calor pode ser utilizada a fim de estudar a ocorrência de enfermidades e o modo de propagação das mesmas junto à população. Algumas iniciativas foram citadas anteriormente como o *site* "Flu Near You" e os mapas da OMS. O *site* TrackLytics, é outro exemplo, que utiliza mapas de calor para apresentar a ocorrência de determinadas enfermidades nos EUA. A Figura 10, mostra a concentração de casos de gastroenterite no mês de abril de 2013.

Considerando que o emprego de técnicas de visualização de dados constitui uma ferramenta importante no processo de construção do entendimento e uso significativo dos dados de saúde pública, o Vis-Saúde utiliza este recurso em várias etapas da metodologia desenvolvida. Em um estágio inicial da visualização dos dados de saúde pública foi utilizado mapas de calor para apresentar sobre um mapa geográfico a concentração de ocorrências de enfermidades em uma determinada região. Um dos diferenciais desta visualização é a fonte de dados (relatórios oficiais do governo quanto a incidência de doenças de notificação compulsória). A apresentação dos dados em tempo real e a possibilidade de utilizar uma animação no mapa de calor com a passagem de tempo e apresentação dos dados em séries temporais constitui outra contribuição que



Figura 10 – Site TrackLytics com dados de saúde nos EUA



Fonte: Site TrackLytics (<http://hhs.tracklytics.com>)

visa facilitar o entendimento dos dados pelo seu espectador. Um dos atrativos da apresentação dos dados em séries temporais com passagem do tempo é o estudo do comportamento sazonal e da propagação das patologias que afetam a população. A apresentação e compreensão destes dados, possibilita que a população em geral e os agentes de saúde tomem ações a fim de minimizar os danos à saúde pública pois é necessário considerar que a visualização apresentada é de casos confirmados de enfermidades.

A etapa mais importante do Vis-Saúde contempla submeter os registros de saúde pública a uma análise mediante a aplicação de uma metodologia matemática do estudo da correlação entre as variáveis de interesse ao evento sob análise. Nesta fase é possível estudar o evento sob a perspectiva de variáveis correlacionadas à ele. Como resultado desta etapa são geradas informações que são apresentadas graficamente e georreferenciadas em regiões de incidência do evento. Nesta visualização, da correlação entre as variáveis, as informações são apresentadas em gráficos com formato de pizza sobre um mapa geográfico, onde cada porção do gráfico representa a correlação de uma variável de interesse com as demais submetidas ao estudo. A análise matemática das variáveis relacionadas ao evento de saúde pública juntamente com a técnica de visualização de dados proposta, constitui a metodologia de análise e visualização dos dados do Vis-Saúde. Neste Capítulo foram apresentadas algumas iniciativas que aproximam a ideia de representação visual de informações desenvolvida no Vis-Saúde e a técnica de modelagem matemática dos registros empregada.



### 3 CONCEITOS BÁSICOS

Neste capítulo serão explorados os conceitos fundamentais que foram utilizados no desenvolvimento do Vis-Saúde. A primeira seção aborda o método matemático de análise de componentes principais (PCA) que foi empregado no Vis-Saúde para efetuar análise da correlação entre as variáveis do evento sob estudo. As demais seções versam sobre a visualização de informações relatando a relevância do emprego de técnicas de apresentação dos dados com o objetivo de prover informações significativas a partir dos registros de dados. Será justificada a escolha da técnica de visualização da concentração de ocorrências de eventos através de mapas de calor bem como o emprego de mapas geográficos para georreferenciar as informações a serem apresentadas. As tecnologias empregadas para viabilizar a visualização dos dados georreferenciados em um ambiente *web* também serão apresentadas neste capítulo.

#### 3.1 Análise de componentes principais

A análise de componentes principais é uma das técnicas estatísticas mais antigas e conhecidas para a análise multivariada de dados. A ideia central da PCA é reduzir a dimensionalidade do conjunto de dados analisados no qual existe um grande número de variáveis correlacionadas, mantendo o máximo possível da variação dos dados originais (JOLLIFFE, 2002). Esta redução da dimensão dos dados é obtida através da transformação em um novo conjunto de variáveis, os componentes principais (PCs), que são em igual número as variáveis originais, porém, não são correlacionadas e resultam da combinação linear das variáveis originais. Estas novas variáveis são ordenadas de tal forma que as primeiras retêm grande parte da variação presente em todas as variáveis originais. Assim ao utilizar componentes principais, um grande conjunto de dados complexos pode ser representado por um número relativamente pequeno de novas variáveis. Esta simplificação aliada a técnicas de visualização de dados, possibilita avaliar visualmente semelhanças e diferenças entre os dados e determinar se pode haver um agrupamento de informações (RINGNÉR, 2008).

Consideremos um conjunto de dados envolvendo  $V$  variáveis, isto é, o conjunto é constituído por  $V$  séries temporais. Este conjunto também terá  $V$  componentes principais, que são combinações lineares das variáveis originais. Onde as combinações lineares são dadas por manipulações algébricas da matriz de covariância das  $V$  variáveis. O método de análise por PCs encontra o mesmo número de novas variáveis (componentes principais), no entanto, apenas algumas delas geralmente explicam a maior parte da variabilidade do processo. Assim, é possível diminuir o número de variáveis de análise, sem ter perda significativa de informações (JACKSON, 1991). Dado um conjunto de dados com  $V$  variáveis e  $N$  observações, para efeito de análise, é possível organizá-los na seguinte matriz:

$$W = \begin{bmatrix} w_1^1 & w_2^1 & \dots & w_V^1 \\ w_1^2 & w_2^2 & \dots & w_V^2 \\ \vdots & \vdots & \ddots & \vdots \\ w_1^N & w_2^N & \dots & w_V^N \end{bmatrix}, \quad (3.1)$$

onde, o subscrito indica a variável e o sobrescrito indica a observação. Assim, temos nas colunas da matriz as variáveis de análise e nas linhas as observações. Em geral, consideramos as variáveis de análise como vetores-coluna  $N$ -dimensionais, isto é:

$$\mathbf{w}_1 = [w_1^1 w_1^2 \dots w_1^N]' \quad (3.2)$$

$$\mathbf{w}_2 = [w_2^1 w_2^2 \dots w_2^N]' \quad (3.3)$$

$$\vdots \quad (3.4)$$

$$\mathbf{w}_V = [w_V^1 w_V^2 \dots w_V^N]', \quad (3.5)$$

onde os apostrofes indicam transposição de matrizes/vetores.

O método de obtenção das componentes principais, considerando este processo com  $V$  variáveis, consiste em encontrar a matriz de covariância de  $W$ , dada por:

$$S = \begin{bmatrix} s_1^2 & s_{12} & \dots & s_{1V} \\ s_{12} & s_2^2 & \dots & s_{2V} \\ \vdots & \vdots & \ddots & \vdots \\ s_{V1} & s_{V2} & \dots & s_V^2 \end{bmatrix}, \quad (3.6)$$

onde, como de costume,  $s_k^2$  é a variância da variável  $k$  e  $s_{ij}$  é a covariância entre as variáveis  $i$  e  $j$ . Note que  $s_k^2$  e  $s_{ij}$  são, essencialmente, produtos internos entre as variáveis de análise. De fato, considerando que as variáveis de análise estejam centralizadas, isto é, cada observação de cada variável foi diminuída de sua média, podemos definir  $s_k^2$  e  $s_{ij}$  como:

$$s_k^2 = \frac{\langle \mathbf{w}_k, \mathbf{w}_k \rangle}{N-1} \quad s_{ij} = \frac{\langle \mathbf{w}_i, \mathbf{w}_j \rangle}{N-1}, \quad (3.7)$$

onde  $\langle \cdot, \cdot \rangle$  representa o produto interno usual. A análise por CPs é baseada na álgebra linear, com efeito, repare que  $S$  é uma matriz simétrica, e portanto diagonalizável. Mais precisamente, existe uma matriz ortonormal  $U$  tal que:

$$U' S U = L, \quad (3.8)$$

onde  $U'$  denota a transposta da matriz  $U$ , e  $L$  é a matriz diagonal cujos elementos da diagonal principal são os autovalores  $l_1 > l_2 > \dots > l_V$  de  $S$ . Mais ainda, as colunas  $\mathbf{u}_1, \mathbf{u}_2, \dots, \mathbf{u}_V$  de  $U$  são os respectivos autovetores normalizados (com norma 1) de  $S$  respectivamente associados

aqueles autovalores, naquela ordem. Tais autovetores são mutualmente ortogonais entre si. Desta forma, as matrizes  $U'$  e  $U$  são matrizes de mudança de base entre as variáveis antigas (originais), em direções não necessariamente ortogonais, e as variáveis novas, ortogonais entre si. Maiores informações sobre PCA podem ser encontradas em (JACKSON, 1991).

### 3.1.1 Cálculo da dependências entre as variáveis consideradas

Uma vez encontradas as componentes principais,  $\mathbf{u}_1, \mathbf{u}_2, \dots, \mathbf{u}_V$ , tomamos a última ( $V$ -ésima) para compor a equação do hiperplano de  $\mathbb{R}^V$  que melhor aproxima os  $N$  pontos  $V$ -dimensionais dados pelas linhas de  $W$ . De fato,  $\mathbf{u}_V$  é ortogonal a esse plano. Portanto, o plano procurado é aquele dos pontos  $\mathbf{x} = (x_1, x_2, \dots, x_V) \in \mathbb{R}^V$  tais que

$$\mathbf{u}_V \cdot \mathbf{x} = \mathbf{u}_V \cdot \mathbf{w}_0, \quad (3.9)$$

onde  $\mathbf{w}_0$  é a média daqueles  $N$  pontos. Denotando  $\mathbf{n} = \mathbf{u}_V$ , esta equação pode ser colocada na forma

$$\sum_{v=1}^V n_v x_v = d, \quad (3.10)$$

onde  $d = \mathbf{u}_V \cdot \mathbf{w}_0$  e  $x_v, v = 1, 2, \dots, V$  variáveis.

Se queremos explicitar a correlação,  $\alpha$ , de uma variável  $k$  em função de uma das  $V$  variáveis restantes escrevemos

$$\alpha_v(k) = \frac{\frac{n_v}{n_k}}{\sum_{\substack{v=1 \\ v \neq k}}^V \frac{n_v}{n_k}} = \frac{n_v}{\sum_{\substack{v=1 \\ v \neq k}}^V n_v} \quad (3.11)$$

onde  $k = 1, 2, \dots, V$  variáveis.

O menor autovalor, ou seja  $l_V$  da matriz  $L$ , possui uma considerável relevância para a análise da correlação entre as variáveis. Este autovalor representa as informações que não puderam ser explicadas pela correlação entre as  $V$  variáveis. Para mensurar a informação não explicada podemos utilizar a equação

$$NE = \frac{l_V}{\sum_{v=1}^V l_v} \quad (3.12)$$

Por consequência, a explicação obtida com a análise da correlação entre as variáveis é dado por

$$E = 1 - NE \quad (3.13)$$

Logo, considerando o total de explicação obtida um função da análise da correlação entre as  $V$  variáveis, se queremos mensurar a proporção de explicação obtida pela correlação de uma variável  $k$  em função de uma das demais  $V$  variáveis de análise, usamos a equação

$$\rho_v(k) = \frac{n_v}{\sum_{\substack{v=1 \\ v \neq k}}^V n_v} * E \quad (3.14)$$

A metodologia desenvolvida no Vis-Saúde visa proporcionar justamente a análise e visualização da correlação entre variáveis relacionadas a um evento de saúde pública. Nesta visualização será considerado o percentual de explicação obtida com as variáveis sob análise e o percentual de informação não explicada com as variáveis analisados. Sendo a explicação total subdividida proporcionalmente entre as variáveis consideradas na análise no momento da geração do gráfico. Estas informações obtidas a partir da análise com a PCA serão visualizadas na forma de gráficos georreferenciados, sendo que para cada área de interesse da análise um gráfico da correlação entre as variáveis estudadas será apresentado no mapa.

### 3.2 Visualização analítica de informações

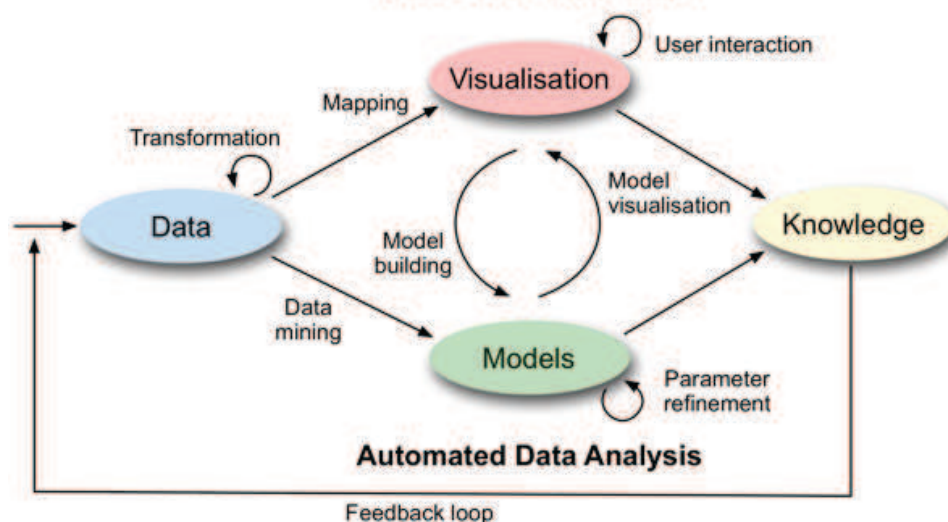
Segundo Lu et al. (2013), visualização e visualização analítica de informações são áreas de estudo muito próximas, ambas objetivam o desenvolvimento de técnicas visuais para revelar informações significativas dos dados em aplicações do mundo real. A visualização, como área de pesquisa tem sua origem na computação gráfica e tornou-se uma área de pesquisa popular ao longo dos anos. Já a visualização analítica é relativamente jovem, com foco em proporcionar um raciocínio analítico através de interfaces visuais interativas.

A visualização de informações, de modo geral, está relacionada com a geração de representações visuais cognitivamente úteis dos dados analisados. Esta representação pode ser através de diagramas, gráficos, mapas ou outros recursos visuais. Segundo Arabnia (1999), a "visualização sintetiza e exibe imagens derivadas de modelos dinâmicos utilizados para analisar os fenômenos do mundo real". O objetivo é empregar o sentido visual humano como um meio para perceber a informação e estimular a criatividade para pensar sobre o seu significado.

Sistemas utilizados para visualização analítica de um volumoso e complexo conjunto de dados, integram algoritmos computacionais de descoberta de conhecimento com técnicas de visualização interativa de informações (SCHRECK; KEIM, 2013). A descoberta de conhecimento utiliza métodos avançados de análise e mineração de dados a fim de encontrar padrões em grandes conjuntos de dados. De acordo com Keim et al. (2010), a visualização analítica é a ciência do raciocínio analítico apoiado por interfaces visuais interativas. Esta técnica permite aos tomadores de decisão combinar a flexibilidade humana, a criatividade e o conhecimento da causa com as capacidades de processamento e armazenamento dos computadores, a fim de proporcionar a compreensão de problemas complexos. A Figura 11, mostra uma visão abstrata dos diferentes estágios (ovais) e suas transições (setas) do processo de visualização analítica.

Contextualizando a estrutura organizacional da visualização analítica no Vis-Saúde, a fonte de dados será proveniente dos relatórios de registros já consolidados de saúde pública. A visualização dos dados será georreferenciada e representada através de gráficos e mapas de calor. A modelagem e análise dos dados será obtida empregando a análise estatística multivariada de componentes principais. O conhecimento extraído dos dados integrará as informações geradas pelo Vis-Saúde com a experiência dos gestores de saúde em interpretar os dados e selecionar as

Figura 11 – Processo de visualização analítica  
**Visual Data Exploration**



Fonte: (KEIM et al., 2010)

variáveis a serem submetidas aos processos de análise e visualização no Vis-Saúde.

### 3.3 Mapa de calor

No Vis-Saúde foi implementada uma visualização dos registros quanto a ocorrência de eventos de saúde pública sem passar por qualquer método matemático de análise de dados. Esta visualização objetiva apenas apresentar um panorama da incidência de eventos segundo o período determinado para visualização. A técnica selecionada para apresentação destes registros foi o mapa de calor sobre um mapa geográfico. O mapa de calor irá promover o entendimento da concentração dos eventos, e se utilizado em conjunto com o recurso de agrupamento em séries temporais irá prover também informações sobre o comportamento do evento ao longo do tempo segundo sua distribuição geográfica. O mapa geográfico sobre o qual o mapa de calor será apresentado é utilizado como base para localização das ocorrências dos eventos e permite o estudo espacial do evento.

Um mapa de calor ou *heat map*, pode ser considerado qualquer forma de representação de dados através de cores em uma imagem de duas dimensões (GREENBERG; XU; KUMAR, 2013). Geralmente, esta técnica é empregada para a visualização de dados com várias dimensões (WALKER; FIELDSEND; EVERSON, 2012) e indicam a concentração de eventos ou o valor de um dado. Diversas áreas de pesquisa como mobilidade urbana, esportes, comportamental, social, política e saúde tem utilizado mapas de calor para apresentar seus dados. Nesta visualização, são utilizadas representações gráficas dos dados onde os valores de uma matriz são representados como cores (PEREZ-LLAMAS; LOPEZ-BIGAS, 2011), que indicam os valores numa escala de máxima, cores quentes como tons de vermelho, a mínima, cores frias como tons de azul ou verde (WALKER; EVERSON; FIELDSEND, 2013).

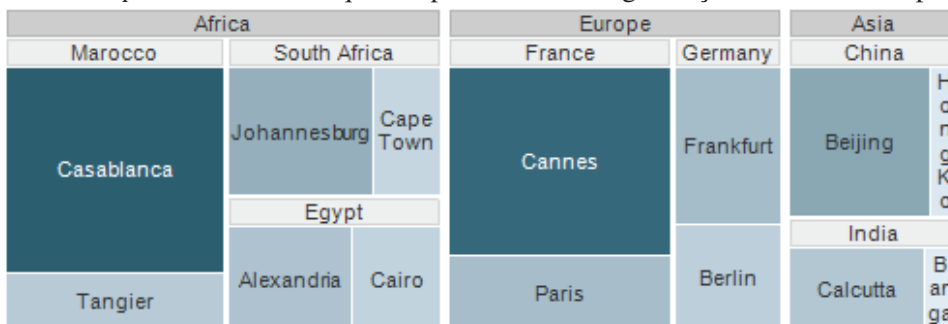


Existem diferentes tipos de mapas de calor que utilizam técnicas de visualização de dados distintas. A maioria dos *heat maps* utilizam técnicas de mapeamento para representar os dados ao invés de tabelas ou gráficos, proporcionando uma visualização única dos dados que não é possível obter em tabelas e gráficos. Nos próximos parágrafos serão descritos alguns tipos de mapas de calor.

- Mapas de árvore (*tree maps*): esta técnica geralmente é empregada na representação de um grande conjunto de dados complexos, como por exemplo, na análise do mercado acionário, gestão de riscos, análise de portfólio de projetos, análise de participação de mercado e gerenciamento de rede. Utilizando a visualização através de mapas de árvore é possível apresentar detalhes de um conjunto de dados e ao mesmo tempo um resumo geral visual, de modo que as tendências, anomalias, distribuições e relacionamento podem ser vistos em um contexto e em vários níveis de uma só vez. O desenvolvimento dos mapas de árvore, data do ano de 1990, e é atribuído a Ben Shneiderman da Universidade de Maryland (SHAH; MITCHELL; SHOOK, 2005).

A técnica de *tree maps*, explora maneiras de mostrar os dados hierárquicos em um *layout* com espaço restrito. A área disponível é dividida em retângulos alternando o orientação entre horizontal e vertical para representar a hierarquia dos dados (SHNEIDERMAN, 1992). As dimensões dos retângulos e as cores indicam detalhes quantitativos dos dados representados. Os retângulos são posicionados do canto superior esquerdo da visualização para o canto inferior direito, com o maior retângulo posicionado no canto superior esquerdo e o menor retângulo no canto inferior direito. Os níveis de hierarquia nos *tree maps* são visualizados como retângulos dentro de outros retângulos. Cada conjunto de retângulos no mesmo nível de hierarquia representam uma coluna ou uma expressão da tabela de dados. Cada retângulo individual no nível da hierarquia representa uma categoria na coluna de dados.

Figura 12 – *Tree map*: Estrutura hierárquica representando a organização de continente, país e cidade.



Fonte: <http://stn.spotfire.com/stn/Configure/VisualizationTypes.aspx>

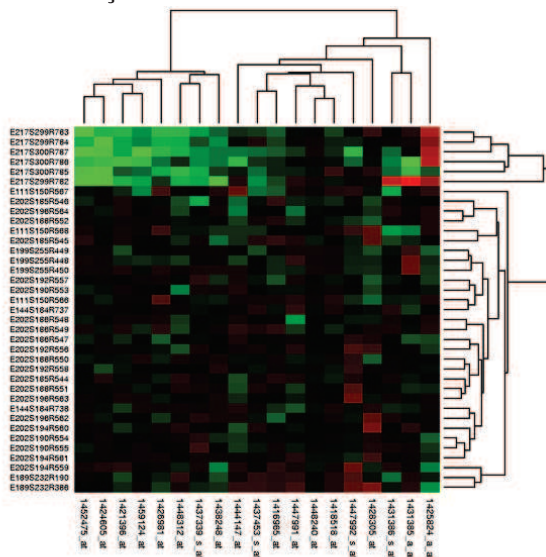
Exemplificando uma estrutura hierárquica de continente, país e cidade: um retângulo, que representa um continente poderá conter vários outros retângulos, um para cada país do continente. Cada retângulo que representa um país, poderá conter outros retângulos



que representam as cidades do país. Esta estrutura hierárquica, pode ser visualizada na Figura 12, onde a coloração dos retângulos foi atribuída com base no número de vendas de produtos em cada cidade. Os retângulos com cores mais escuras indicam um maior número de vendas. Diversos algoritmos foram desenvolvidos para determinar o tamanho e a ordem dos retângulos em um mapa de árvore, a seleção da implementação a ser aplicada depende dos dados a serem apresentados.

- Mapa de matriz de micro dados (*microarray maps*): é outra técnica de apresentação de dados que constitui um tipo especial de mapa de calor. Comumente, esta técnica é utilizada em biologia molecular para representar dados dos genes. Esta técnica utiliza uma grade de cores interligada a um dendrograma (um diagrama de árvore) para agrupar hierarquicamente os genes. As linhas da matriz representam os genes e cada coluna da linha representa uma amostra diferente. A cor de cada célula da matriz é definida com base no nível de expressão do gene na referida amostra, um exemplo de um mapa de calor com matriz de micro dados pode ser visualizado na Figura 13. Esta visualização auxilia os pesquisadores da biologia a encontrar grupos de genes que possuem um comportamento semelhante durante os experimentos (Lab Escape, 2013).

Figura 13 – Mapa de calor gerado a partir de micro dados do DNA que representam os valores de expressão dos genes em diversas condições.

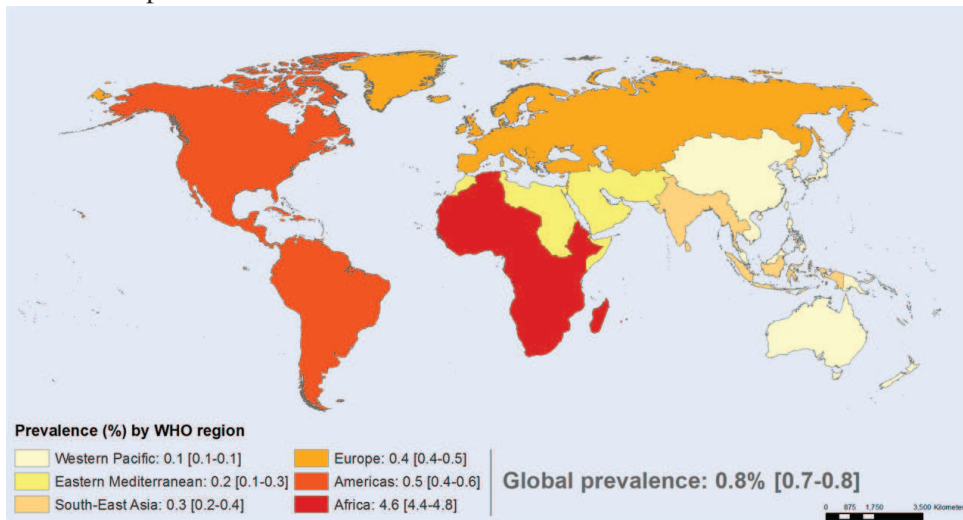


Fonte: <http://www.labescape.com/info/articles/what-is-a-heat-map.html>

- Mapas geográficos: a representação de mapas de calor em mapas geográficos, é uma das formas de representação de concentração de dados mais tradicional. Esta técnica é usada para visualizar os dados com base na sua localização geográfica de ocorrência. Esta forma de apresentação dos dados geralmente é utilizada para destacar tendências, anomalias e distribuições através de uma área geográfica. A coloração do mapa de calor

geográfico pode ser definida através de diversos algoritmos diferentes, no entanto sempre irão representar os dados a serem visualizados. A coloração pode ser aplicada a uma região geográfica inteira (continente, país, estado, cidade, etc) ou em pontos geográficos específicos que devido a sua proximidade e concentração podem ou não serem agrupados por um algoritmo de geração da cor. A Figura 14, apresenta um exemplo de mapa de calor onde a coloração, que representa a concentração de dados, é definida em toda uma região geográfica.

Figura 14 – Mapa de calor geográfico: Ocorrência do vírus HIV entre adultos (15-49 anos) em 2011. Regiões estabelecidas pela OMS.



Fonte: OMS - [http://gamapserver.who.int/mapLibrary/Files/Maps/HIV\\_adult\\_prevalence\\_2011.png](http://gamapserver.who.int/mapLibrary/Files/Maps/HIV_adult_prevalence_2011.png)

Na presente proposta será apresentado outro tipo de mapa de calor, baseado na definição de gradientes em cada um dos marcadores a serem inseridos na camada de apresentação dos dados. Neste tipo de mapa de calor, a cada marcador  $i$  de localização  $P_i$  no mapa, associamos uma função

$$C_i(\mathbf{x}) = I_i f(\mathbf{x} - P_i), \quad (3.15)$$

onde  $\mathbf{x}$  é uma posição genérica no mapa,  $f$  é uma função de  $x$  radial em relação ao centro  $P_i$ , isto é, dependente somente da distância  $|\mathbf{x} - P_i|$ , e  $I_i$  uma intensidade associada ao marcador. Por exemplo, podem-se ter  $f$  como do tipo Gaussiana, ou mesmo uma função afim em relação a  $|\mathbf{x} - P_i|$ . A função  $f$  adotada pode ou não ser de suporte compacto.

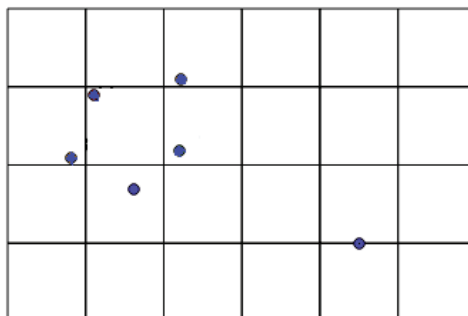
Quando, em um mapa, existe um conjunto de marcadores localizados em  $P_1, P_2, \dots, P_M$ , tem-se em cada ponto arbitrário  $\mathbf{x}$  do mapa,

$$C_R(\mathbf{x}) = \sum_{m=1}^M C_i(\mathbf{x}). \quad (3.16)$$

O Resultado  $C_R$  é então transformado em cor de acordo com algum esquema, a ser escolhido pelo usuário.

A Figura 15, apresenta um espaço matricial onde alguns marcadores foram inseridos para exemplificar a composição da intensidade das cores no mapa de calor com marcadores em gradiente.

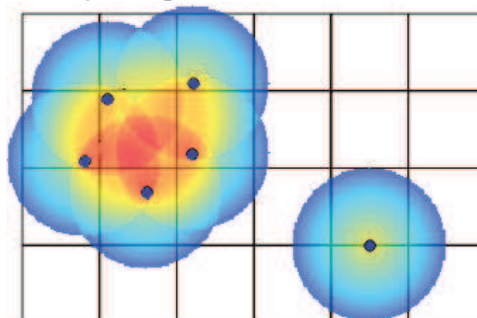
Figura 15 – Apresentação de pontos em uma matriz



Fonte: <http://gis.stackexchange.com/questions/55424/heatmap-algorithm-to-visualise-point-diversity>

A intensidade das cores depende do número total de marcadores inseridos no mapa de calor. Assim na medida que novos marcadores vão sendo adicionados, o mapa de calor é recalculado com novas intensidades de cores e interpolação dos marcadores próximos. Na Figura 16, foi adicionado à matriz apresentada anteriormente, o mapa de calor equivalente aos marcadores inseridos. Nesta imagem é possível visualizar um único marcador no mapa de calor e interação/interpolação de cores quando são inseridos vários marcadores próximos.

Figura 16 – Apresentação de pontos em uma matriz com o mapa de calor

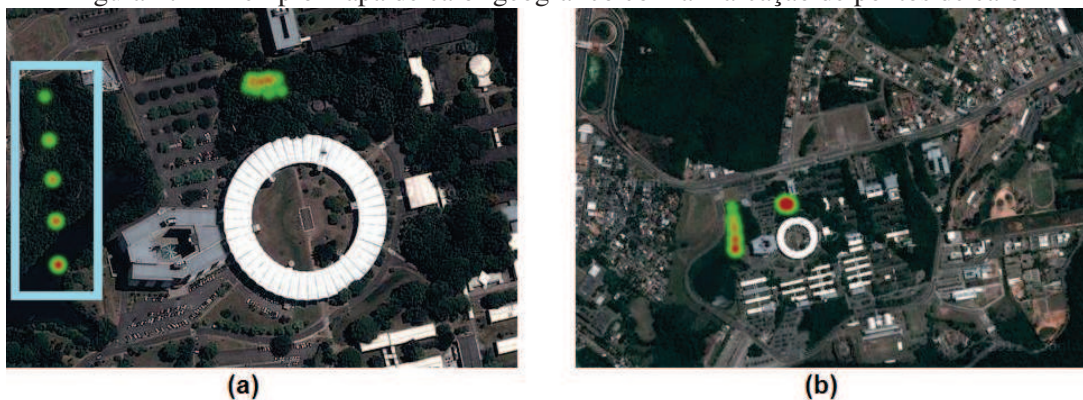


Fonte: <http://gis.stackexchange.com/questions/55424/heatmap-algorithm-to-visualise-point-diversity>

Na Figura 17, são apresentadas duas imagens do mesmo mapa de calor porém com níveis de aproximação (*zoom*) diferentes. Quando a proximidade da visualização é alterada, o mapa de calor é recalculado, pois a distância entre os marcadores é visualmente diferente e estes deverão ser agrupados ou separados. Na Figura 17, na imagem (a), foram inseridos diversos marcadores a fim de exemplificar visualmente a coloração e o agrupamento dos mesmos. O retângulo azul destaca 5 grupos de marcadores, com diferentes concentrações, no mapa de calor. Inicialmente há um único marcador, em seguida são inseridos 2, 3, 4

e 5 marcadores, respectivamente, sobre um mesmo ponto a fim de apresentar a alteração na coloração de acordo com a concentração de pontos no mapa de calor. Na Figura 17(a), foi inserido também um grupo de 15 marcadores próximos que geraram uma espécie de "mancha" de ocorrência no mapa de calor, no entanto como os pontos não estavam muito próximos a coloração do grupo indica uma mínima concentração. A Figura 17(b), é o mesmo mapa de calor da Figura 17(a), porém com uma visão mais distante (menor *zoom*), logo os pontos estão mais próximos visualmente e foram agrupados na visualização.

Figura 17 – Exemplo mapa de calor geográfico com a marcação de pontos de calor



### 3.4 Integração entre as técnicas de análise e visualização de dados

A análise da correlação entre as variáveis de análise empregada no Vis-Saúde é obtida através da análise de componentes principais. O emprego da PCA puramente resulta em números estatísticos que são representados visualmente através de gráficos apresentados sobre um mapa geográfico com as delimitações dos espaços geográficos. Para viabilizar a visualização do resultado obtido com a PCA em um ambiente *web* com a possibilidade de apresentar os dados em tempo real é necessário o emprego de tecnologias que permitam a dinâmica necessária ao ambiente. Com este propósito foram utilizadas algumas das tecnologias disponíveis na interface de programação de aplicativos (API) do Google que serão apresentadas nas sub-seções abaixo:

#### 3.4.1 Google API

Com o desenvolvimento da *web 2.0*, combinar dados de mais de uma fonte em uma única aplicação ou *site* para criar um novo serviço (*mashup*) tem se tornado uma tendência. Os serviços disponíveis através da API (*Application Programming Interface* - Interface de Programação de Aplicativos) do Google são um exemplo típico de recursos que podem ser utilizados em aplicações do tipo (LIN; GAO; XU, 2009). O Google disponibiliza diversos serviços e ferramentas através de suas APIs, que consistem em especificações usadas por componentes de *software* para comunicar uns com os outros (W3SCHOOLS, 2014). Uma API é constituída por um conjunto de estruturas de dados, classes de objetos ou funções que podem ser usadas por

um programador utilizando uma linguagem de programação como JavaScript, PHP ou outra linguagem para fazer uso dos recursos da API.

Considerando aspectos de integração e funcionalidades necessárias, foi utilizado no Vis-Saúde a API do Google Maps para exibir o mapa de referência dos dados a serem analisados e visualizados. A delimitação dos estados brasileiros foi adicionada com o recurso de Fusion Tables do Google, onde são armazenadas as informações que compõe o polígono de cada estado. Já os gráficos de visualização da dependência entre as variáveis analisadas foram inseridas através do API Google Visualization.

### 3.4.2 Google Maps API

O Google Maps revolucionou o serviço de aplicações que utilizam mapas *online* na internet com um novo tipo de interação cliente/servidor baseado em AJAX (*Asynchronous JavaScript and XML*), capaz de manter uma comunicação contínua entre o cliente e o servidor para fazer a apresentação imediata de informações no mapa (UDELL, 2008). O JavaScript é a linguagem de programação nativa do Google Maps para manipulação de funções da API, onde o programador pode personalizar os mapas usando etiquetas XHTML (*Extensible HTML*) para definir a formatação do mapa através de diretivas CSS (*Cascading Style Sheet*).

A API do Google Maps é um serviço gratuito que permite incorporar o Google Maps em páginas da *web* ou aplicativos móveis. Esta API normalmente é usada em aplicativos não comerciais, de teste e desenvolvimento. O serviço que utiliza a API do Google Maps deve ser acessado de forma gratuita pelos usuários finais. Esta API permite sobrepor seus próprios dados em um mapa personalizado do Google. É possível criar aplicativos para *web* ou dispositivos móveis com a plataforma de mapeamento do Google, incluindo imagens de satélite, visão de ruas (*Street View*), relevo, rotas de carro, mapas estilizados, informações demográficas, análises e um grande banco de dados de lugares. A versão 3 da API estendeu o suporte também para navegadores de dispositivos móveis com suporte a Javascript tornando o Google Maps JavaScript API uma das API's mais utilizadas em aplicações que usam mapas *online* (HU; DAI, 2013). As visualizações de dados apresentadas no Vis-Saúde são baseadas em API do Google Maps que proporcionam a integração de mapas geográficos, camadas de mapas de calor e gráficos as aplicações *web*.

### 3.4.3 Google Fusion Tables

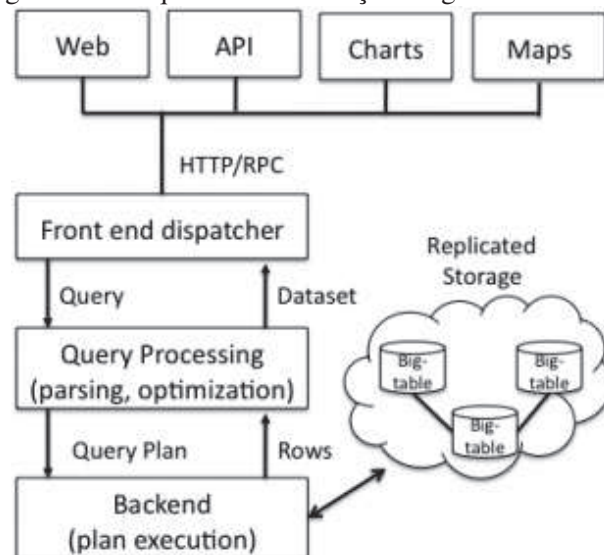
O serviço Google Fusion Tables permite o gerenciamento e integração de grandes estruturas de dados baseado no conceito de computação em nuvem. Através deste serviço os usuários podem transferir seus arquivos de dados para a nuvem, efetuar consultas e visualizá-los imediatamente após sua transferência. Os meios de visualização de dados disponíveis são variados incluindo mapas, linhas do tempo ou gráficos. Uma vez que os dados foram transferidos para



o serviço de Fusion Tables eles podem ser consultados diretamente ou podem ser utilizados para gerar novas visualizações do conjunto de dados com algum filtro ou método de agregação específico (GONZALEZ et al., 2010).

É abrangente o conjunto de visualizações disponíveis pelo serviço Fusion Tables sendo que a escolha de uma depende do tipo dos dados a serem apresentados. Por exemplo, a apresentação dos dados através de um mapa só será possível se em alguma coluna dos dados houver informações de localização, como coordenadas de latitude e longitude. É possível também integrar dados de diversas origens através de junção de consultas em diferentes tabelas de dados de usuários distintos. Os usuários deste serviço, podem manter seus dados privados, públicos ou compartilhados com outros usuários. A visualização dos dados é obtida por meio da API de visualização do Google que renderiza no navegador do usuário, usando Javascript ou Flash, a visualização dos dados obtidos através de uma interface específica de consulta. A Figura 18, apresenta a arquitetura do serviço Google Fusion Tables.

Figura 18 – Arquitetura do serviço Google Fusion Tables



Fonte: (GONZALEZ et al., 2010)

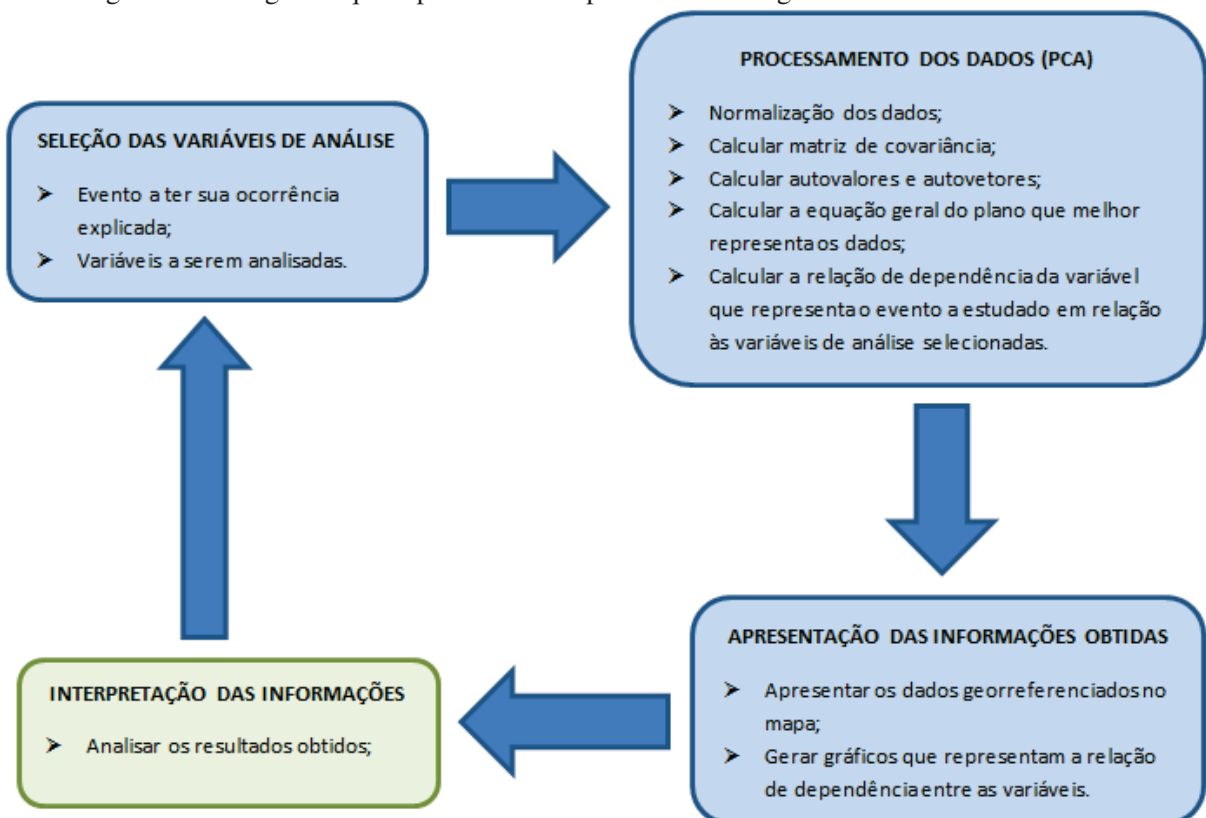
Um componente muito popular do serviço de Fusion Tables é a representação de um grande conjunto de dados geográficos. É possível transferir tabelas com diretivas de pontos, linhas, polígonos ou dados de localização geográfica (latitude e longitude) que podem ser renderizados como mapas. Esta renderização dos dados pode ser feita no servidor da aplicação e ao cliente é enviado um conjunto de pequenas imagens que contém o mapa já renderizado. Este recurso do serviço Google Fusion Tables foi utilizado no Vis-Saúde para apresentar no mapa geográfico a delimitação da área geográfica dos estados brasileiros.

#### 4 DESENVOLVIMENTO DO MODELO

Neste capítulo serão apresentadas informações relativas as etapas da metodologia de análise e visualização desenvolvida no Vis-Saúde sob os aspectos de aplicação e desenvolvimento do algoritmo de análise dos dados e dos métodos de visualização das informações.

Um dos escopos do trabalho é propor uma metodologia de análise e representação da relação de dependência entre as variáveis selecionadas para análise, com o intuito de explicar a ocorrência de um determinado evento em função destas. Logo, o estágio inicial do processamento dos dados requer a definição das variáveis de análise e do evento a ser estudado. Em uma segunda etapa é necessário processar os dados utilizando a análise de componentes principais a fim de se obter o estudo da relação de dependência entre as variáveis propostas e o evento sob análise. Finalizada a análise, a próxima etapa consiste em visualizar as informações geradas, onde serão utilizados recursos da API do Google para apresentar os dados sob a forma de gráficos georreferenciados sobre um mapa. Apresentados os dados resultantes da análise, cabe ao usuário do sistema interpretá-los, e decidir entre tomar ações ou continuar a análise dos dados alterando as variáveis sob análise. A Figura 19, apresenta um diagrama que visa ilustrar as etapas da metodologia de análise do Vis-Saúde de forma global.

Figura 19 – Diagrama que representa as etapas da metodologia desenvolvida no Vis-Saúde



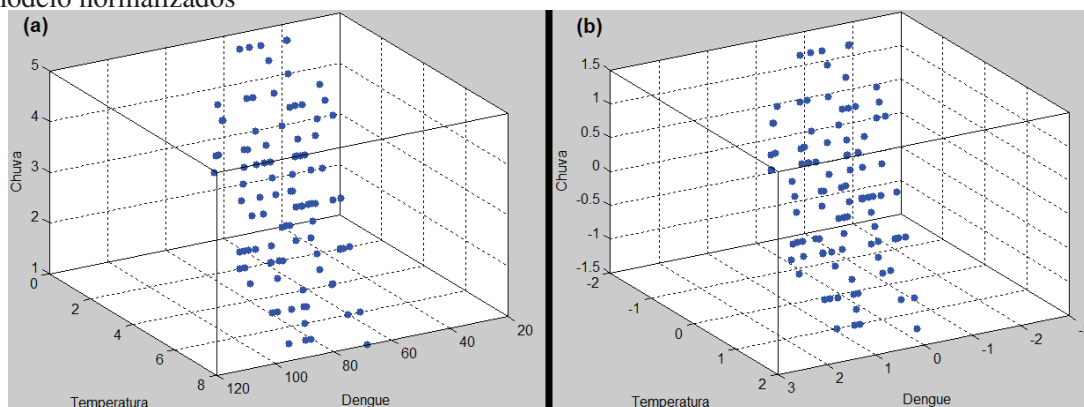
#### 4.1 Seleção das variáveis de análise

A primeira etapa da metodologia de análise e visualização do Vis-Saúde é a definição das variáveis de análise. Deve ser definida uma variável que representa o evento a ser estudado frente as demais variáveis de análise. A metodologia irá prover informações que definem a relação de dependência entre a variável que representa o evento e as demais variáveis de análise definidas. Para exemplificar a aplicação da metodologia foram definidas de forma arbitrária três variáveis: dengue (número de casos no mês), temperatura média do mês (em graus Celsius) e pluviosidade (milímetros de chuva acumulada no mês). O evento a ser estudado será a incidência de casos de dengue em função da temperatura média e da pluviosidade.

#### 4.2 Processamento dos dados

Uma vez definido o evento a ser estudado frente ao conjunto de variáveis, é necessário aplicar o algoritmo de processamento de dados. A Figura 20a, apresenta o conjunto de dados para teste gerado para exemplificar o funcionamento da análise proposta no Vis-Saúde. Cada ponto no espaço tridimensional representa um registro que é composto por número de casos de dengue, temperatura média e pluviosidade acumulada. Estes registros são submetidos a um pré-processamento onde os dados são normalizados. O processo de normalização pode considerar apenas a média de cada variável, neste caso focando a correlação dos dados propriamente. Ou considerar a média e o desvio padrão de cada variável, neste caso, com ênfase na correlação da variação dos dados. O método de normalização empregado para demonstração inicial considera a média e o desvio-padrão de cada variável. A representação dos dados de análise após o processo de normalização pode ser visto na Figura 20b.

Figura 20 – (a) Dados de amostra para demonstração do modelo (b) Dados de amostra para demonstração do modelo normalizados

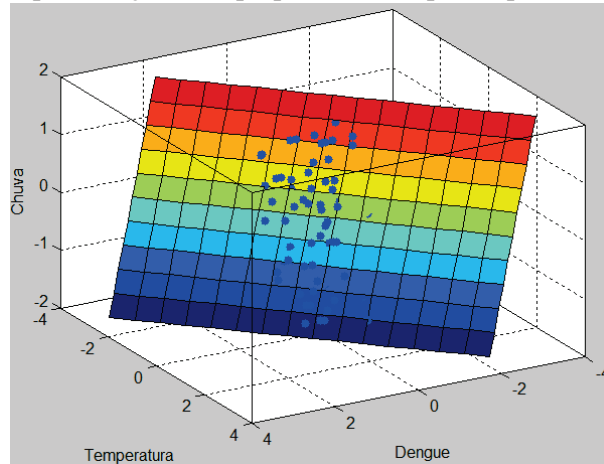


Uma vez executado o processo de normalização dos dados a análise de componentes principais é aplicada. Sendo necessário calcular a matriz de covariância dos dados e os seus autovalores e autovetores conforme descrito na seção 3.1. Através dos valores obtidos dos cálculos do



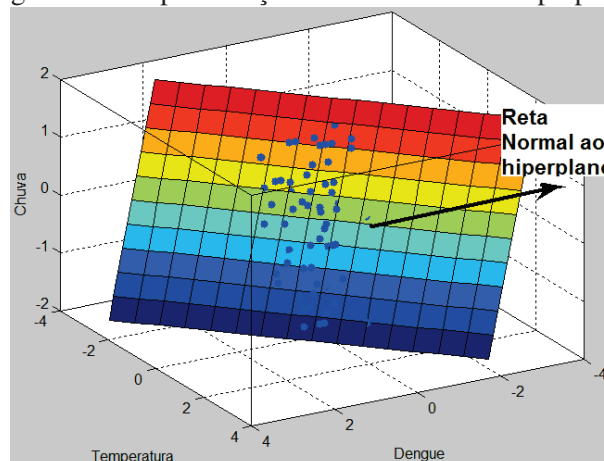
autovalores é possível obter a equação geral do plano que melhor representa os dados sob análise conforme apresentado na equação 3.10. A Figura 21, a traz uma representação dos dados com o plano encontrado como mais adequado para representação dos mesmos.

Figura 21 – Representação do hiperplano obtido para representação dos dados



Uma característica importante para o modelo de cálculo da dependência entre as variáveis são as propriedades matemáticas do menor autovalor resultante da matriz de covariância. Este autovalor representa a menor variância dos dados analisados e seu respectivo autovetor representa o vetor da reta normal ao plano que melhor representa os dados sob análise. Uma representação desta reta normal ao hiperplano pode ser vista na Figura 22. Este autovalor também pode ser utilizado para calcular a quantidade de informações que não puderam ser auferidas mediante as variáveis analisadas conforme equação 3.12 apresentada anteriormente.

Figura 22 – Representação da reta normal ao hiperplano



Para calcular a relação de dependência de uma variável em função de outra é necessário aplicar a equação 3.11, que irá resultar em um coeficiente que posteriormente será convertido para percentual. Na visualização de dados proposta no Vis-Saúde foi considerado ainda o percentual de informações que não pode ser explicado com as variáveis de análise, salientando

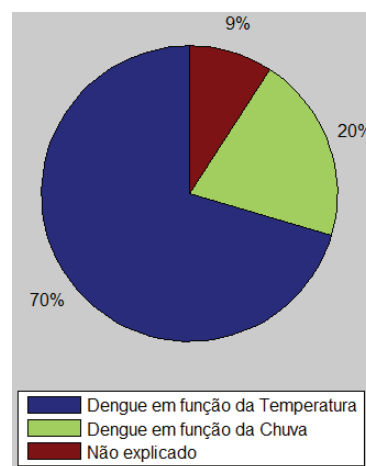
uma possível lacuna nas variáveis analisadas para explicar o evento e abrindo possibilidades de ampliar a análise a um conjunto maior de variáveis. Esta informação é obtida mediante o cálculo de representatividade do menor autovalor encontrado no cálculo da PCA em relação aos demais autovalores conforme equação 3.12.

### 4.3 Apresentação das informações

#### 4.3.1 Visualização de informações advindas da análise com PCA

Sendo o conjunto de variáveis processado utilizando a metodologia descrita na seção anterior, estes devem ser apresentados segundo uma técnica de visualização apropriada. Para fins de análise, os dados são agrupados em regiões de interesse, e para cada área geográfica de estudo, um gráfico com a apresentação das informações resultantes da análise é exibido. Considerando todas as formas possíveis de apresentações gráficas de dados e a natureza das informações a serem exibidas, optou-se por utilizar uma representação através de gráficos em formato de pizza. Esta representação é usualmente empregada para a apresentação de dados estatísticos e são adequados para serem exibidos sobre o mapa geográfico possibilitando uma interpretação intuitiva das informações. Neste gráfico, para cada variável de análise é associada uma cor (temperatura média e pluviosidade) e existe ainda o percentual de variância dos dados que não pode ser explicado com a análise das variáveis selecionadas. Esta informação é tão relevante quanto o percentual de explicação obtida, pois instiga a exploração da análise dos dados frente a novas variáveis. Na Figura 23, é possível verificar, através de um exemplo, o gráfico gerado a partir da análise de um conjunto de variáveis visando explicar a ocorrência da dengue em uma região.

Figura 23 – Gráfico representando a relação de dependência entre a ocorrência de dengue e fatores climáticos



Neste exemplo, 9% de informações permanecem sem explicação mediante a análise das variáveis temperatura média e pluviosidade, logo foi obtido 91% de explicação do evento com a análise frente as variáveis citadas. Sendo destes 91% de explicação do evento, 70% está

associado com a variável temperatura e 20% está associado com a variável pluviosidade.

#### 4.3.2 Visualização de informações advindas dos relatórios de saúde pública

A segunda visualização proposta no Vis-Saúde apresenta os dados organizados em séries temporais sobre a ocorrência de determinado evento de saúde em uma região. Os dados são apresentados sobre um mapa geográfico onde a representação da ocorrência de uma enfermidade em uma região é exibida sobre o aspecto de concentração das mesmas através de mapas de calor. Para cada ocorrência de um enfermidade é inserido um marcador no mapa, a concentração de marcadores no mapa de calor gera representações diferenciadas baseadas em cores, onde tons de vermelho representam uma maior concentração de eventos e uma área de maior calor. Esta representação de dados foi desenvolvida adicionando a ela um recurso de animação que possibilita uma análise da série temporal e sua evolução ao longo do tempo e do espaço geográfico. Na medida que o tempo passa a apresentação de dados é alterada sendo possível visualizar a distribuição e propagação geográfica de uma enfermidade ao longo do tempo.

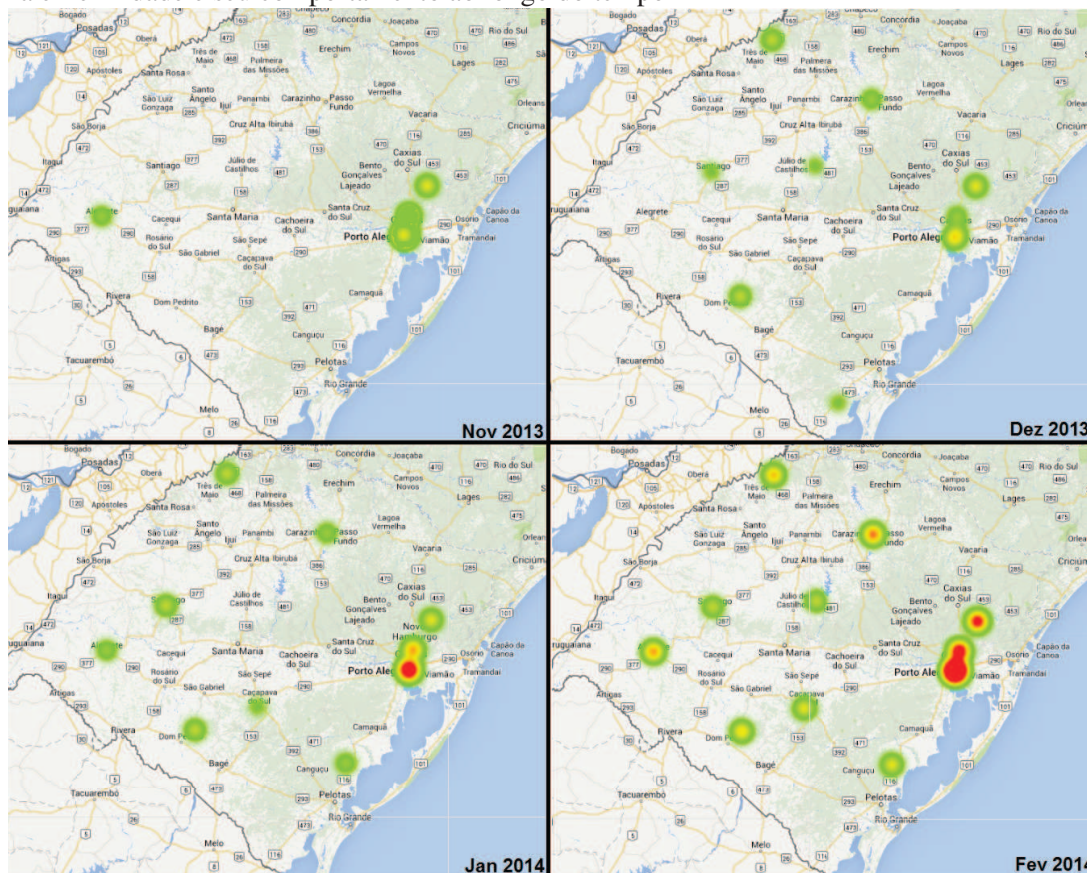
Esta visualização apresenta os dados dos relatórios de saúde sem a adição de nenhum algoritmo de pré-processamento ou análise. Contudo, o emprego de técnicas de visualização que proporcionam uma forma distinta de apresentação de dados constitui uma ferramenta importante para a extração de conhecimento sobre o dados apreciados. A visualização da incidência de enfermidades baseada na localização geográfica e comportamento ao longo do tempo possibilita benefícios tanto para a população quanto aos gestores de saúde pública. A população pode visualizar os dados e obter conhecimento sobre a situação de saúde pública de uma determinada região de seu interesse buscando medidas de prevenção pessoal ou mesmo evitando a visitação de algum local onde um surto endêmico de alguma enfermidade esteja ocorrendo. Já os gestores podem embasar suas decisões estratégicas de prevenção e combate à surtos de doenças em dados dos próprios órgãos públicos que uma vez visualizados de forma mais amigável podem ser facilmente compreendidos. A Figura 24, apresenta uma visualização baseada na concentração da ocorrência de uma enfermidade e sua evolução longo do tempo.

#### 4.3.3 Ferramentas utilizadas no desenvolvimento

No desenvolvimento do ambiente de protótipo para teste da metodologia proposta no Vis-Saúde foram utilizadas diferentes ferramentas computacionais, sendo preferencialmente empregadas tecnologias *web* e *open source* que permitem o acesso ao sistema a partir de qualquer dispositivo que disponha de uma conexão com a rede mundial de computadores e um navegador (*browser*) para acesso às páginas *web*.

Para a geração dos dados de prova de conceito da metodologia do Vis-Saúde e análise inicial da correlação entre as variáveis foi utilizado o MATLAB. Este aplicativo é composto por uma linguagem de alto nível e um ambiente interativo que dispõe de ferramentas e funções

Figura 24 – Visualização de dados advindos dos relatórios de saúde pública - Concentração da incidência de uma enfermidade e seu comportamento ao longo do tempo



matemáticas que podem ser utilizadas para analisar conjuntos de dados, gerar visualizações de informações, desenvolver algoritmos e criar modelos e aplicações voltados para soluções matemáticas (MATLAB, 2014). Como o MATLAB é uma ferramenta proprietária não passível de ser integrada ao ambiente *web* proposto no Vis-Saúde, o uso desta ferramenta foi restrita apenas para a fase inicial de testes da metodologia desenvolvida no Vis-Saúde.

O Vis-Saúde disponibiliza a visualização das informações, advindas da análise, com base em um sistema georeferenciado através de uma plataforma *web* independente de arquitetura e acessível através de *softwares* navegadores (*browser*). As visualizações geradas pelo Vis-Saúde são publicadas na internet através do servidor *web* Apache.

O paradigma de programação utilizado na codificação do protótipo para teste da metodologia do Vis-Saúde foi o orientado a objetos (OO). Este paradigma consiste em uma abordagem de desenvolvimento onde a estrutura do software é baseado em objetos interagindo entre si para realizar uma tarefa (CLARK, 2013). A programação orientada a objetos possui grande capacidade de escalabilidade e reutilização de código, podendo ser utilizada tanto em rotinas simples como em sistemas complexos (BUDD, 2002).

A linguagem de programação utilizada no desenvolvimento do protótipo para testes foi o *Hypertext Preprocessor* (PHP). A escolha pela linguagem PHP foi baseada na integração da mesma com o SGBD escolhido e pela existência de bibliotecas de análise de dados estatísticos

que foram vitais no desenvolvimento da análise de dados implementada no Vis-Saúde. A análise da correlação entre as variáveis foi desenvolvida utilizando a PCA que por sua vez depende de uma série de cálculos matemáticos para sua implementação. Dentre as rotinas matemáticas necessárias está o cálculo de autovalores da matriz de dados, para o cálculo dos autovalores foram utilizadas rotinas implementadas pela biblioteca LAPACK. Esta biblioteca foi escrita na linguagem de programação Fortran 90 e fornece rotinas para cálculos de sistemas de equações lineares, autovalores e autovetores entre outros problemas matemáticos complexos (NETLIB, 2014)

Visando contemplar o maior número possível de usuários e compatibilizar o sistema com diversos dispositivos existentes, sistemas operacionais e navegadores, a interface visual do aplicativo foi desenvolvida utilizando a linguagem *HyperText Markup Language* (HTML) e o *framework* Bootstrap. O Bootstrap foi construído por integrantes da equipe do Twitter e utiliza *Cascading Style Sheets* (CSS) responsivo que permite uma melhor visualização do *site* desenvolvido em diversos dispositivos, navegadores e distintos tamanhos e resoluções de tela (BOOTSTRAP, 2014).

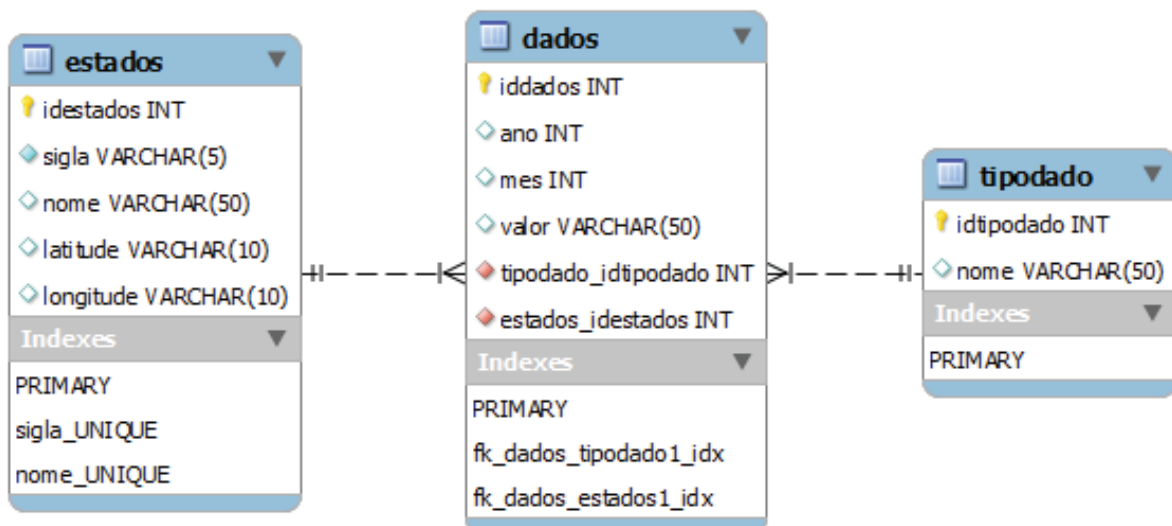
Os dados a serem analisados pelo Vis-Saúde são armazenados em um banco de dados MySQL, este sistema gerenciador de banco de dados (SGBD) figura entre os quatro SGBD relacionais mais utilizados no mundo (DB-ENGINES, 2014). O banco de dados utilizado no Vis-Saúde possui uma estrutura simplificada com apenas 3 tabelas, sendo elas:

- estados: nesta tabela são armazenadas informações (nome, sigla, localização geográfica da capital e um identificador para referência) sobre os estados da federação brasileira.
- tipodado: contém informações (nome e um identificador para referência) sobre as variáveis que serão utilizadas na análise.
- dados: esta é a tabela com maior número de registros, nela são armazenados os dados a serem submetidos a análise. Contém todos os registros para todas as variáveis com referência do tipo do dado ao qual o registro se refere e estado ao qual está vinculado além do ano, mês e valor da variável propriamente.

Um diagrama que representa a organização do banco de dados para teste da metodologia do Vis-Saúde é apresentado na Figura 25.

Todo o processamento matemático necessário para os cálculos que compõem a metodologia do Vis-Saúde são executados no servidor do protótipo, no lado do cliente apenas a visualização das informações é renderizada pelo navegador.

Figura 25 – Diagrama do banco de dados relacional para teste da metodologia do Vis-Saúde





## 5 RESULTADOS

Nos Capítulos anteriores, a metodologia do Vis-Saúde foi descrita detalhadamente. Neste Capítulo serão apresentados os resultados obtidos com bases em testes executados com dados fictícios e dados reais oriundos dos relatórios sobre a incidência de dengue correlacionados com dados climáticos. Os testes com dados fictícios são relevantes tendo em vista que são dados gerados artificialmente e assim facilitam a comprovação da metodologia de análise da relação de dependência entre as variáveis estudadas. Testes com dados reais muitas vezes agregam dificuldades na obtenção de informações confiáveis, contudo, compõe outra importante forma de demonstração da aplicação da metodologia desenvolvida.

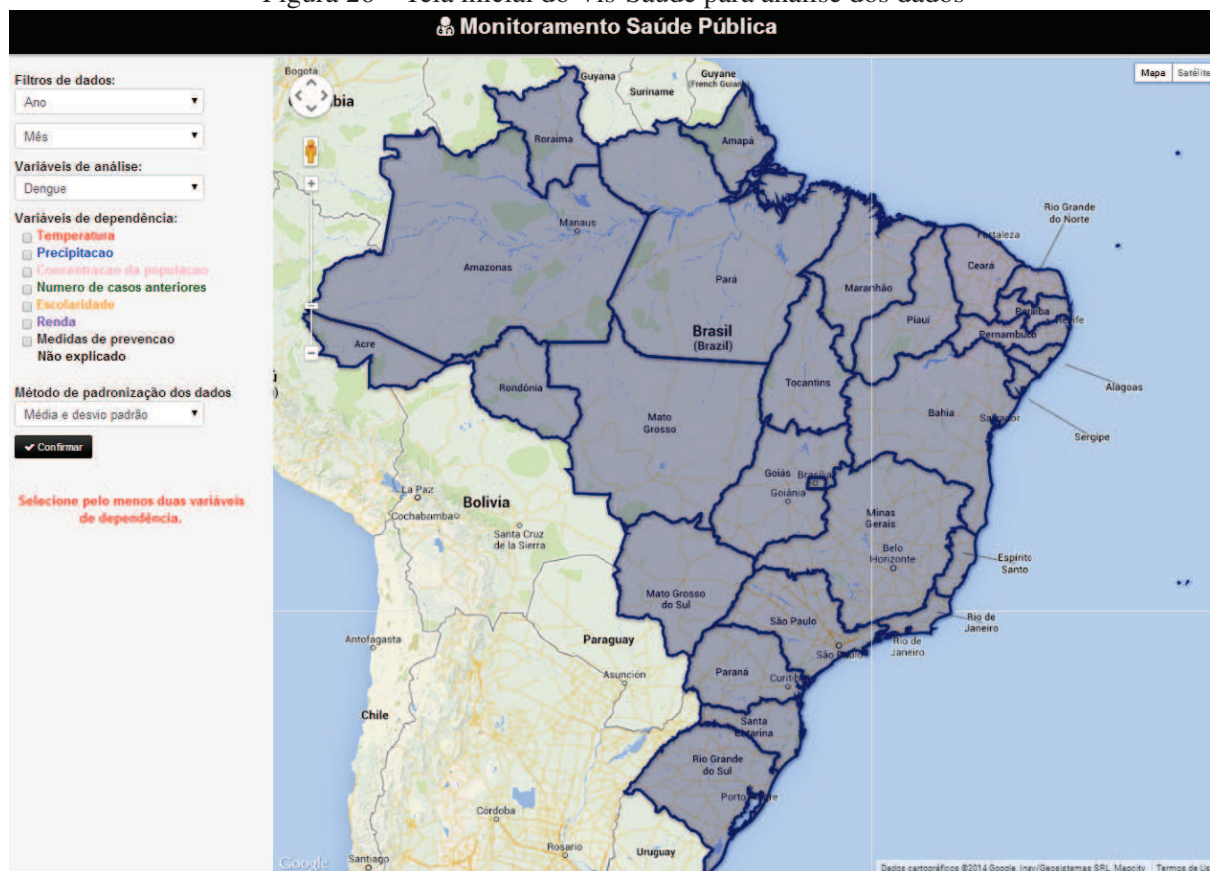
### 5.1 Descrição do funcionamento do ambiente do Vis-Saúde

O Vis-Saúde contempla uma metodologia de análise e visualização de dados consolidados de saúde pública onde um evento de saúde, como por exemplo a incidência de dengue, é analisada frente às variáveis possivelmente correlacionadas como as climáticas. O objetivo da análise é identificar e estudar a relação de dependência entre as variáveis consideradas a fim de buscar explicações sobre o evento foco do estudo.

Também é contemplado pelo sistema a implementação de técnicas de visualização dos registros de saúde organizados em séries temporais quanto à concentração de incidência de dengue em uma região. Todos os dados apresentados pelo Vis-Saúde são georreferenciados e apresentados sobre um mapa. O usuário do sistema deve definir o período de dados a ser analisado, selecionar uma variável de análise (que corresponde ao evento a ser investigado) e as demais variáveis de dependência. É possível também selecionar o método de padronização das variáveis a ser utilizado na etapa de pré-processamento dos dados. A análise utilizando os dados padronizados apenas pela média permite o estudo da correlação entre os dados propriamente. Enquanto que a análise utilizando como método de padronização a média e o desvio-padrão irão fornecer uma análise da correlação da variação dos dados. A Figura 26, apresenta a tela inicial do Vis-Saúde para a análise de dependência entre as variáveis.

O processamento da relação de dependência entre as variáveis pode ser calculado em tempo real em um ambiente *web* passível de ser desenvolvido com soluções *open source*, como foi descrito no ambiente de protótipo desenvolvido para teste da metodologia. Logo, um utilizador pode alterar as variáveis de análise segundo julgar necessário para facilitar a compreensão das informações e obter a visualização dos novos resultados em poucos instantes. A apresentação da relação de dependência de um evento em relação às variáveis de análise é expressa através de gráficos no formato de pizza que são apresentados sobre o mapa, um para cada estado ou região de interesse. A Figura 27, apresenta uma ilustração da tela de protótipo do Vis-Saúde com a apresentação das informações geradas pela análise de dependência entre as variáveis. Cada variável é sempre representada pela mesma cor no gráfico a fim de facilitar o entendimento

Figura 26 – Tela inicial do Vis-Saúde para análise dos dados



do observador. Quando um utilizador clica sobre um gráfico resultante da análise apresentado no mapa, este é visualizado em detalhes onde é possível observar o sentido da correlação (representado por um sinal positivo ou negativo) entre as variáveis de análise e o evento foco do estudo. Na seções seguintes serão detalhados os gráficos apresentados nesta visualização.

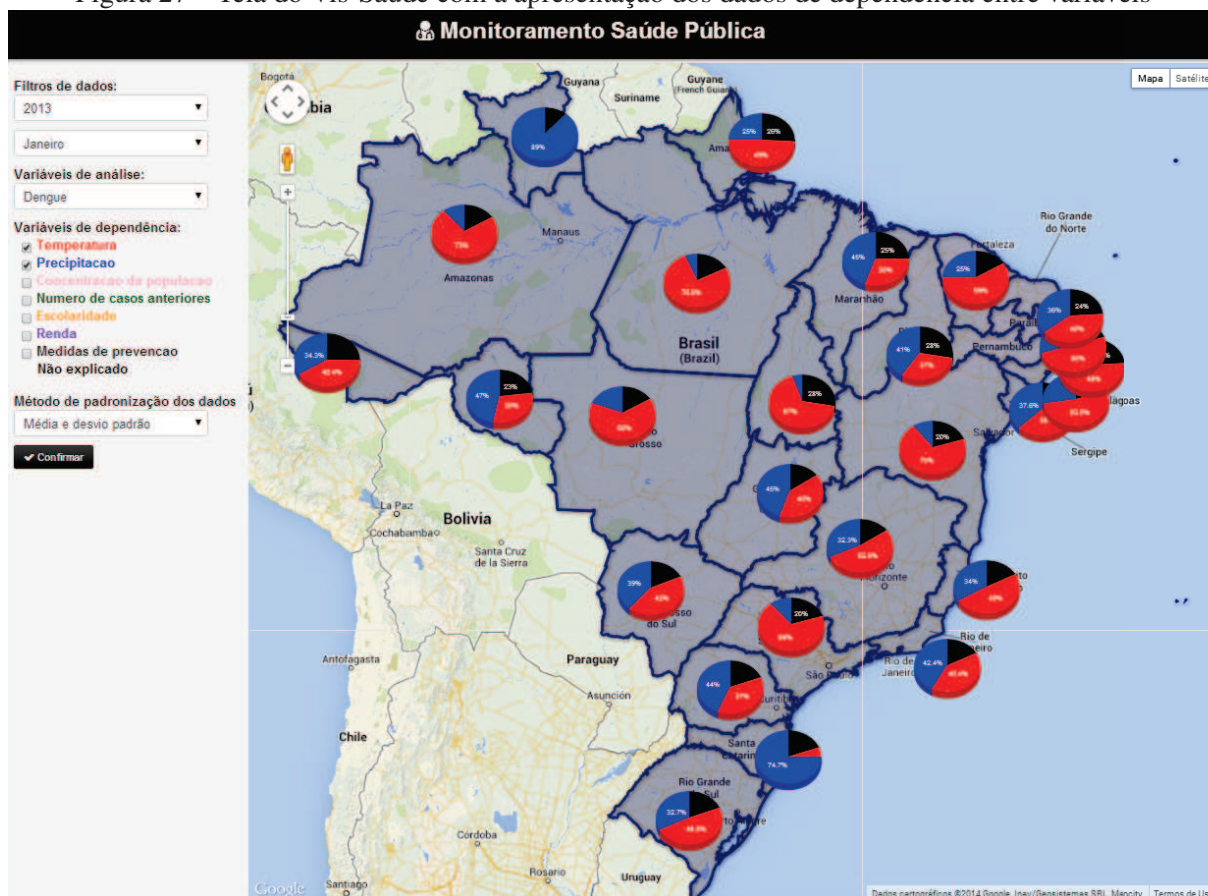
## 5.2 Homologação da metodologia de análise e visualização da dependência entre variáveis com dados fictícios

Para verificar o funcionamento da metodologia de análise proposta do Vis-Saúde testes iniciais foram realizados com dados fictícios. Tais dados foram gerados artificialmente para facilitar a verificação dos resultados obtidos e simulam a incidência mensal de dengue nos estados do Brasil. Logo, os registros gerados para testar a metodologia de análise, simulam a incidência mensal de dengue em função de variáveis sociais, climáticas e econômicas nos 26 estados do Brasil. As variáveis consideradas na geração do dados foram:

- Número de casos de dengue registrados na região
- Temperatura média da região (em graus Celsius)
- Volume de chuva acumulada (milímetros acumulados)



Figura 27 – Tela do Vis-Saúde com a apresentação dos dados de dependência entre variáveis



- Concentração de pessoas
- Número de casos de dengue registrados no período anterior de análise
- Nível de escolaridade da população (índice entre 0 e 10)
- Ações de prevenção à incidência de dengue (índice entre 0 e 10)
- Renda média da população (número de salários mínimos)

Para cada uma das 26 regiões de interesse consideradas, foram gerados dados mensais compostos por 50 amostras para cada uma das variáveis citadas, sendo considerado o período de anos de 2010 a 2013. No total, 499.200 registros de dados para as variáveis consideradas foram inseridos no banco de dados para testes. Tais dados foram gerados utilizando o aplicativo MATLAB e depois importados para o banco de dados do Vis-Saúde através de uma rotina própria implementada para importar o arquivo de dados gerado pelo MATLAB.

Na geração dos dados, foi utilizada uma função de geração de números inteiros pseudo-aleatórios distribuídos uniformemente em um intervalo pré-estabelecido para cada uma das variáveis sociais, climáticas e econômicas consideradas na análise. Já a incidência de dengue resulta do produto entre um coeficiente atribuído arbitrariamente para cada uma das variáveis

de análise com o próprio valor aleatório gerado para cada uma das variáveis. A equação utilizada para a geração de dados foi

$$\begin{aligned} \text{Dengue} = & (2 * \text{temperatura}) + (3 * \text{chuva}) + (2 * \text{populacao}) + (2 * \text{renda}) \\ & + (4 * \text{casosanteriores}) + (3 * \text{escolaridade}) + (4 * \text{prevencao}) + 5 + \text{RUIDO} \end{aligned} \quad (5.1)$$

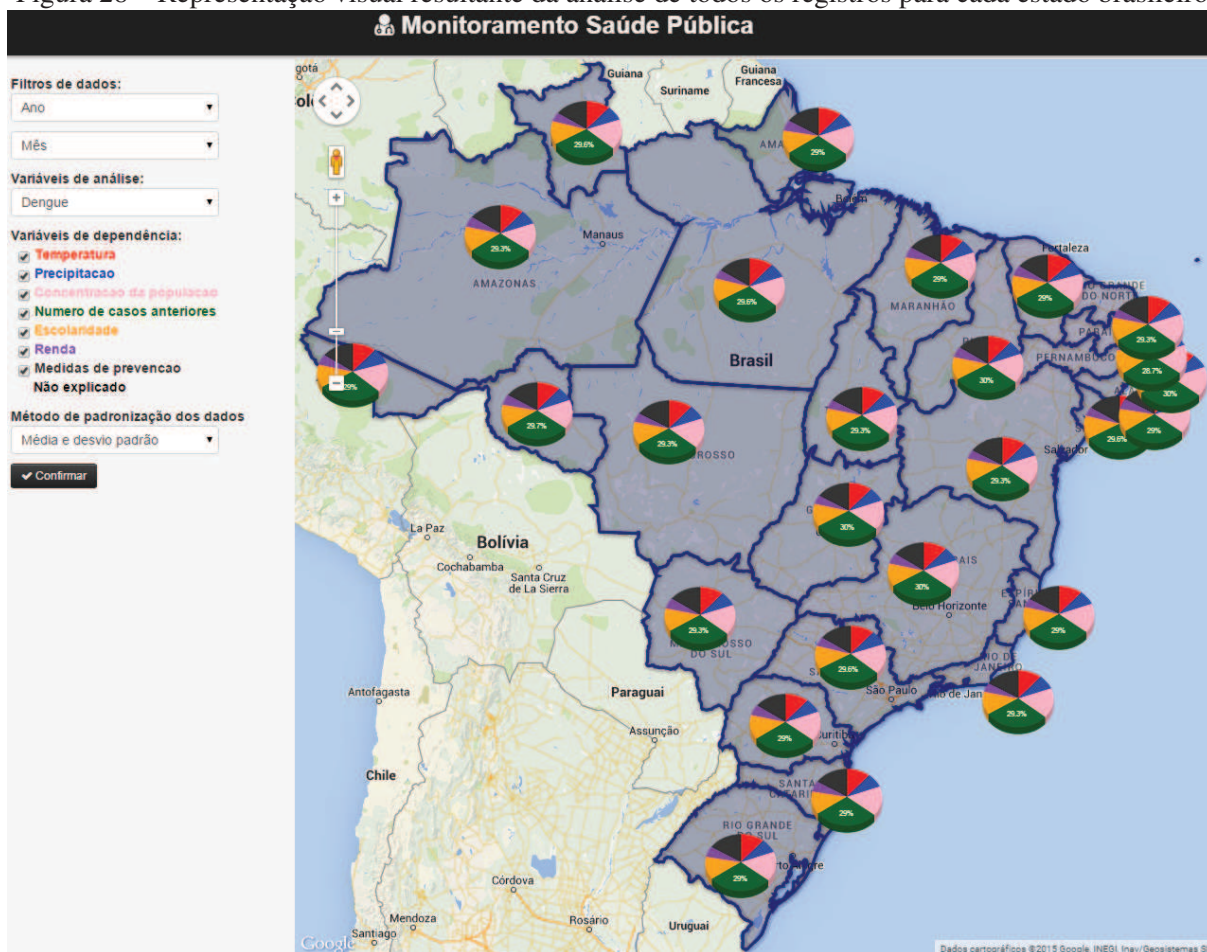
O ruído adicionado na equação de geração dos dados para testes representa uma flutuação dos mesmos, visando aproximar estes de uma situação real. Contudo, não é foco deste teste simular fielmente a incidência de dengue ou outra variável de análise citada, estes dados visam apenas gerar um banco de dados populado com registros para demonstração da metodologia de análise e visualização do Vis-Saúde.

Na análise, os dados são agrupados por estado segundo os critério de seleção estabelecidos pelo usuário que interage definindo as variáveis. Para cada estado, um gráfico com o resultado da metodologia é apresentado. Para cada região de interesse, o sistema faz uma busca dos registros segundo o período selecionado pelo usuário. Uma vez os dados selecionados e agrupados por regiões de interesse, a metodologia de análise dos dados utilizando a análise de componentes principais é aplicada para calcular a relação de dependência entre o evento e as variáveis selecionadas. Em seguida é exibido sobre um mapa, geográfico, base do sistema de visualização um gráfico no formato de pizza que corresponde ao resultado obtido com a análise dos dados. Neste gráfico, as informações são apresentadas na forma de fatias. A divisão inicial do gráfico é em duas partes, sendo que uma destas é reservada para apresentar o percentual de informação que não pôde ser explicada com a análise das variáveis. Outra parte, que representa o percentual de informação que pode ser explicada com a análise das variáveis relacionadas com o evento, é sub-dividida em fatias entre as variáveis de análise de dependência. Cada variável recebe uma fatia de tamanho proporcional ao seu percentual de relevância para a obtenção da explicação, considerando o percentual de explicação total que foi obtido. A Figura 28 apresenta uma representação gráfica resultante da análise de todos os registros de dados para cada estado brasileiro.

A fim de observar detalhadamente o resultado obtido com a análise e visualização dos dados, uma restrição ao conjunto de registros foi aplicada como exemplo. O conjunto de dados sob análise compreendem apenas o estado do Rio Grande do Sul e o mês de Janeiro de 2013. A Figura 29, mostra a representação dos dados para cada uma das 50 amostras geradas para as sete variáveis (casos de dengue, temperatura média, pluviosidade, concentração da população, número de casos no período anterior, escolaridade, ações de prevenção e renda) de análise propostas para testes. Na geração de dados foi atribuído um coeficiente de multiplicação para cada uma das variáveis, estes fatores não estão sendo exibidos nos gráficos dos dados apresentados.

Considerando a análise da dengue em função de todas as demais variáveis de teste temos um percentual de explicação dos dados que ultrapassa o limiar de 99,9%, isso porque o ruído inserido na geração de dados foi muito baixo. Logo, para o período e região definidos (Rio Grande do Sul, Janeiro de 2013) o gráfico de explicação da ocorrência de dengue em função

Figura 28 – Representação visual resultante da análise de todos os registros para cada estado brasileiro



das demais variáveis de análise pode ser observado na Figura 30. Neste caso a maior correlação da incidência da dengue é com as variáveis número de casos anteriores na região, concentração da população e medidas de prevenção. É possível observar uma correlação positiva (sinal ao lado de cada variável) entre a incidência de dengue e todas as demais variáveis de análise. Este resultado era esperado devido a formulação utilizada na geração dos dados para teste da metodologia.

Para o mesmo período e região, porém analisando a incidência de casos de dengue apenas em função das variáveis de análise temperatura, precipitação e renda é possível observar que um percentual maior de informação não pôde ser explicado em função das variáveis de análise selecionadas. Este comportamento era esperado devido as características de geração dos dados artificiais. A Figura 31, apresenta o gráfico resultante desta análise com um conjunto menor de variáveis de análise da dependência.

### 5.3 Demonstração com dados reais de saúde pública

Uma relevante etapa do desenvolvimento da metodologia proposta no Vis-Saúde é a demonstração da metodologia com dados reais advindos dos relatórios consolidados de saúde

Figura 29 – Registros gerados para testes - Rio Grande do Sul (Jan 2013)

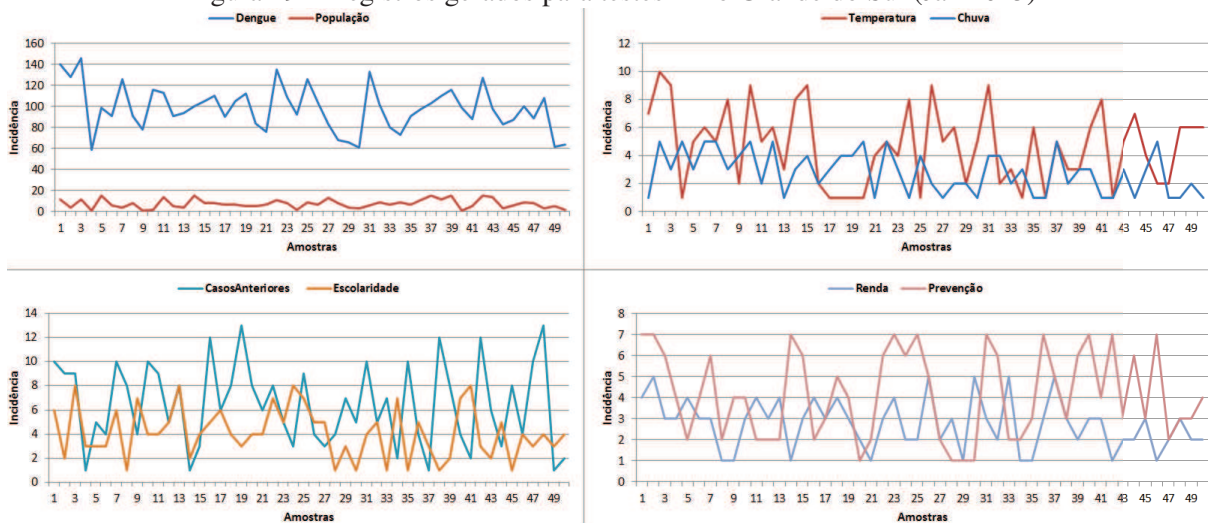
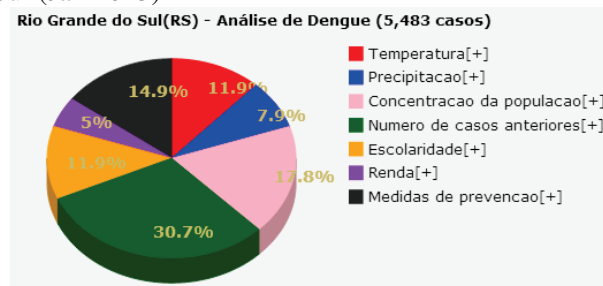


Figura 30 – Resultado da análise da ocorrência da dengue em função de todas as demais variáveis de análise - Rio Grande do Sul (Jan 2013)



pública e de estações meteorológicas espalhas pelo território brasileiro. Considerando o avanço tecnológico nos meios de transmissão de dados, armazenamento e processamento, mesmo os municípios mais distantes podem e devem registrar as informações sobre a incidência de enfermidades de notificação compulsória ou de interesse regional. Conforme será descrito nas próximas seções os dados disponibilizados pelo governo constituem uma importante fonte de informações a ser explorada pelos pesquisadores, contudo o acesso e a ontologia dos dados ainda gera algumas dificuldades na exploração das mesmas nos meios acadêmicos.

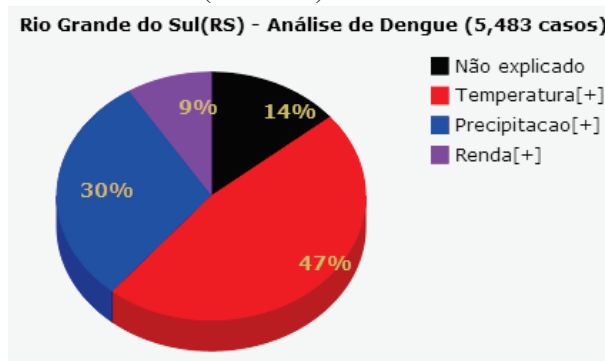
### 5.3.1 Fonte de dados reais para análise

#### 5.3.1.1 Dados de saúde pública

Os dados utilizados no Vis-Saúde referentes à ocorrências reais de dengue são advindos do Sinan (Sistema de Informação de Agravos de Notificação). Este sistema foi implantado no Brasil de forma gradual a partir de 1993 e tem por objetivo coletar, transmitir e disseminar dados gerados rotineiramente pelo Sistema de Vigilância Epidemiológica das três esferas de governo, por intermédio de uma rede informatizada, para apoiar o processo de investigação e



Figura 31 – Resultado da análise da ocorrência da dengue em função da temperatura média, precipitação e renda da população - Rio Grande do Sul (Jan 2013)



dar subsídios à análise das informações de vigilância epidemiológica das doenças de notificação compulsória (Ministério da Saúde, 2007). O Sinan é alimentado, principalmente, pela notificação e investigação de casos de doenças e agravos que constam da lista nacional de doenças de notificação compulsória (Portaria GM/MS Nº 104, de 25 de janeiro de 2011), mas é facultado a estados e municípios incluir outros problemas de saúde importantes em sua região. Em 1998, o uso do Sinan foi regulamentado, tornando obrigatória a inserção regular de registros na base de dados nacional pelos municípios, estados e Distrito Federal. Como forma de instigar o registro de informações no Sinan os municípios que não adicionarem informações ao banco de dados do Sinan, por dois meses consecutivos, têm suspensos os recursos do Piso de Assistência Básica (PAB).

Quanto a coleta de dados pelo Sinan, as mesmas podem ser operacionalizadas nas unidades de saúde, seguindo a orientação de descentralização do Sistema Único de Saúde (SUS). Os dados são coletados a partir da Ficha Individual de Notificação (FIN) que é preenchida pelas unidades assistenciais para cada paciente quando da suspeita da ocorrência de problema de saúde de notificação compulsória ou de interesse nacional, estadual ou municipal. Esse instrumento deve ser encaminhado aos serviços responsáveis pela informação e/ou vigilância epidemiológica das Secretarias Municipais, que devem repassar semanalmente os arquivos em meio magnético para as Secretarias Estaduais de Saúde (SES). A comunicação das SES com a Secretaria de Vigilância em Saúde (SVS) deverá ocorrer quinzenalmente, de acordo com o cronograma definido pela SVS no início de cada ano. Caso não ocorra nenhuma suspeita de doença, as unidades de saúde precisam preencher o formulário de notificação negativa, que tem os mesmos prazos de entrega. Além da Ficha Individual de Notificação (FIN), e da Notificação Negativa, o Sistema ainda disponibiliza a Ficha Individual de Investigação (FII), que é um roteiro de investigação, que possibilita a identificação da fonte de infecção, os mecanismos de transmissão da doença e a confirmação ou descarte da suspeita. Ainda são utilizados para a coleta de dados a planilha de surtos e os boletins de acompanhamento de casos de Hanseníase e Tuberculose (Ministério da Saúde, 2007).

O Sinan nacional, disponível no endereço <http://dtr2004.saude.gov.br/sinanweb>, apresenta

os dados na forma de tabelas ou arquivos csv (*Comma-Separated Values* - Valores Separados por Vírgula). Para serem utilizados no Vis-Saúde os dados disponibilizados pelo Sinan foram exportados no formato csv e importados para o banco de dados do Vis-Saúde mediante uma rotina própria desenvolvida. A Figura 32, apresenta um exemplo de dados tabulados gerados pelo Sinan e que foram utilizados pelos Vis-Saúde após a devida importação.

Figura 32 – Exemplo de dados disponíveis no Sinan referentes à incidência de dengue no Brasil no ano de 2006

| Ministério da Saúde  |               |               |               |               |               |               |               |              |              |              |              |               |                | DATASUS                                   |
|--|---------------|---------------|---------------|---------------|---------------|---------------|---------------|--------------|--------------|--------------|--------------|---------------|----------------|---|
|  |               |               |               |               |               |               |               |              |              |              |              |               |                | Tecnologia da Informação a serviço do SUS |
| DENGUE - Notificações registradas no Sistema de Informação de Agravos de Notificação - Sinan |               |               |               |               |               |               |               |              |              |              |              |               |                |   |
| Notificações por Mês 1º Sintoma(s) segundo UF Notificação                                    |               |               |               |               |               |               |               |              |              |              |              |               |                |   |
| Período: 2006  |               |               |               |               |               |               |               |              |              |              |              |               |                |   |
| UF Notificação   | Jan           | Fev           | Mar           | Abr           | Mai           | Jun           | Jul           | Ago          | Set          | Out          | Nov          | Dez           | Total          |   |
| <b>TOTAL</b>   | <b>20.453</b> | <b>27.347</b> | <b>51.457</b> | <b>62.041</b> | <b>39.600</b> | <b>18.493</b> | <b>11.547</b> | <b>7.487</b> | <b>4.215</b> | <b>4.559</b> | <b>8.377</b> | <b>10.797</b> | <b>266.373</b> |   |
| Rondonia   | 1.955         | 1.427         | 408           | 124           | 62            | 32            | 17            | 15           | 13           | 49           | 84           | 250           | 4.436          |   |
| Acre   | 80            | 38            | 27            | 19            | 8             | 10            | 5             | 2            | 5            | 7            | 30           | 39            | 270            |   |
| Amazonas   | 74            | 81            | 126           | 75            | 62            | 38            | 36            | 38           | 31           | 28           | 28           | 28            | 645            |   |
| Roraima  | 26            | 8             | 3             | 15            | 77            | 86            | 118           | 132          | 107          | 125          | 141          | 135           | 973            |   |
| Para   | 1.806         | 1.125         | 1.051         | 708           | 576           | 306           | 138           | 140          | 142          | 142          | 227          | 381           | 6.742          |   |
| Amapa  | 395           | 356           | 286           | 123           | 104           | 61            | 47            | 57           | 59           | 104          | 98           | 121           | 1.811          |   |
| Tocantins  | 504           | 668           | 1.080         | 1.010         | 999           | 265           | 91            | 46           | 140          | 107          | 367          | 622           | 5.899          |   |
| Maranhao   | 351           | 538           | 478           | 539           | 706           | 300           | 303           | 393          | 266          | 381          | 322          | 448           | 5.025          |   |
| Piaui  | 235           | 264           | 425           | 445           | 753           | 1.261         | 540           | 284          | 134          | 108          | 74           | 236           | 4.759          |   |
| Ceara  | 961           | 1.106         | 1.826         | 2.839         | 5.755         | 5.804         | 4.405         | 2.652        | 1.077        | 760          | 901          | 475           | 28.561         |   |
| Rio Grande do Norte  | 193           | 260           | 403           | 698           | 1.159         | 2.010         | 1.827         | 886          | 408          | 196          | 204          | 136           | 8.380          |   |
| Paraíba  | 229           | 201           | 356           | 438           | 313           | 313           | 317           | 275          | 188          | 186          | 144          | 128           | 3.088          |   |
| Pernambuco   | 412           | 369           | 827           | 1.209         | 1.940         | 1.391         | 942           | 648          | 333          | 219          | 236          | 247           | 8.773          |   |
| Alagoas  | 193           | 202           | 148           | 139           | 267           | 320           | 488           | 499          | 274          | 250          | 216          | 166           | 3.162          |   |
| Sergipe  | 50            | 57            | 93            | 178           | 246           | 125           | 86            | 54           | 51           | 72           | 87           | 45            | 1.144          |   |
| Bahia  | 586           | 731           | 1.263         | 1.377         | 1.028         | 390           | 226           | 154          | 138          | 220          | 371          | 327           | 6.811          |   |
| Minas Gerais   | 2.706         | 4.523         | 7.285         | 7.719         | 4.738         | 896           | 186           | 71           | 77           | 236          | 637          | 905           | 29.979         |   |
| Espirito Santo   | 441           | 1.081         | 1.762         | 2.398         | 2.561         | 845           | 366           | 199          | 91           | 104          | 136          | 118           | 10.102         |   |
| Rio de Janeiro   | 1.980         | 3.886         | 7.787         | 6.863         | 3.150         | 976           | 392           | 267          | 150          | 198          | 287          | 745           | 26.681         |   |
| Sao Paulo  | 1.460         | 4.080         | 16.199        | 25.171        | 10.266        | 1.750         | 514           | 281          | 174          | 250          | 446          | 1.041         | 61.632         |   |
| Parana   | 42            | 59            | 336           | 548           | 131           | 31            | 10            | 4            | 7            | 10           | 24           | 43            | 1.245          |   |
| Santa Catarina   | 5             | 9             | 13            | 12            | 9             | 5             | 1             | 2            | -            | 1            | -            | 2             | 59             |   |
| Rio Grande do Sul  | 5             | 16            | 11            | 9             | 12            | 6             | -             | 1            | -            | 1            | 6            | 2             | 69             |   |
| Mato Grosso do Sul   | 160           | 548           | 2.707         | 3.680         | 1.361         | 337           | 99            | 44           | 48           | 104          | 763          | 1.602         | 11.453         |   |
| Mato Grosso  | 1.001         | 1.045         | 1.490         | 908           | 603           | 224           | 145           | 183          | 116          | 354          | 1.885        | 1.719         | 9.673          |   |
| Goias  | 4.519         | 4.617         | 4.982         | 4.694         | 2.627         | 670           | 234           | 151          | 172          | 335          | 635          | 794           | 24.430         |   |
| Distrito Federal   | 84            | 52            | 85            | 103           | 87            | 41            | 14            | 9            | 14           | 12           | 28           | 42            | 571            |   |

Fonte: Sinan

Para a análise de dados proposta no Vis-Saúde foram utilizados apenas os registros de dengue, agrupados por mês, para os anos de 2003 até 2012. O período selecionado contempla os anos em que os registros estavam completos e confiáveis. Os registros de 2013 e 2014 estavam disponíveis no Sinan mas foram removidos sob alegação de que estavam inconsistentes. Nesta análise inicial estão contempladas apenas 7 capitais de estados brasileiros, entre eles, alguns com maior incidência histórica de dengue no Brasil. As capitais selecionadas para a análise foram: Porto Alegre, Florianópolis, São Paulo, Rio de Janeiro, Belo Horizonte, Fortaleza e Salvador. Para as capitais consideradas na análise, foram contabilizados apenas os casos de dengue registrados para moradores das respectivas cidades. O critério utilizado para seleção das capitais foi a incidência histórica de dengue e também a presença de estações climáticas das quais serão consultadas as variáveis climáticas adicionadas na análise. É importante citar que as capitais a terem seus dados analisados com o Vis-Saúde abrangem regiões climáticas e sociais muito distintas e estas variações podem influenciar os resultados obtidos na análise.

### 5.3.1.2 Dados sobre variáveis climáticas

A metodologia proposta e apresentada no Vis-Saúde proporciona a análise da correlação da ocorrência de um evento em função de variáveis de análise pré-estabelecidas. O emprego desta metodologia de análise na área de saúde pública possibilita a análise da correlação de eventos de saúde com fatores externos a fim de verificar a relação de dependência destes e assim abrir novas possibilidades de investigação das causas dos eventos analisados. A prova de conceito do Vis-Saúde com dados reais visa analisar a incidência de casos de dengue frente a fatores climáticos. As variáveis climáticas utilizadas são advindas do Banco de Dados Meteorológicos para Ensino e Pesquisa (BDMEP) disponibilizado pelo Instituto Nacional de Meteorologia (INMET), são elas:

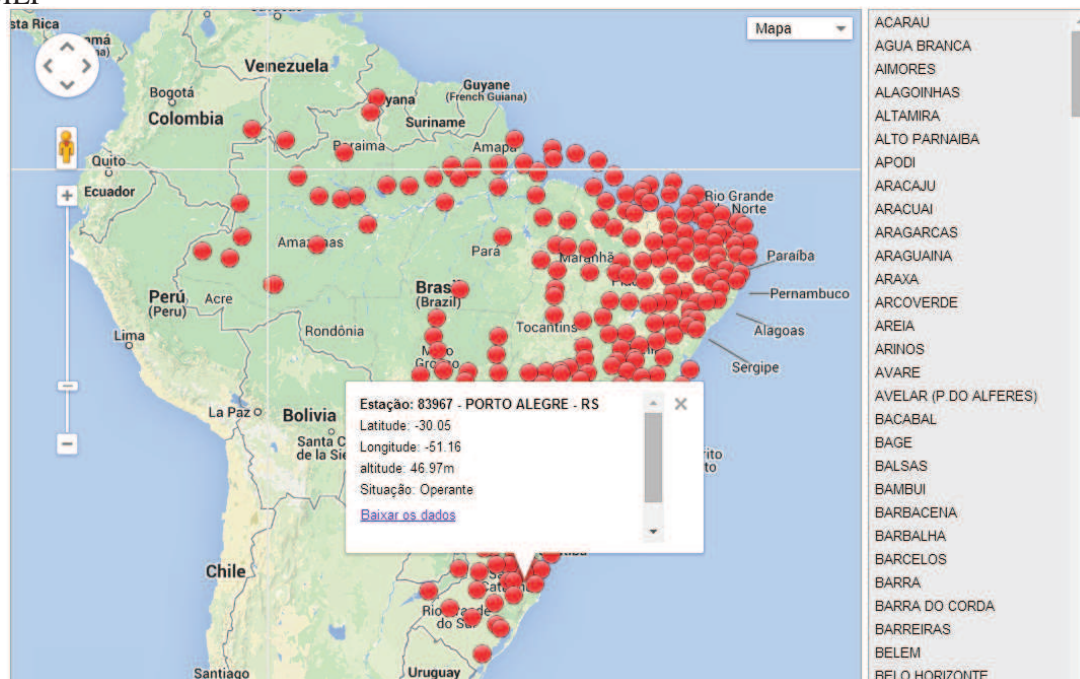
- Número de dias com precipitação
- Pluviosidade (volume de chuva acumulada em milímetros no mês)
- Temperatura máxima média (em graus Celsius)
- Temperatura mínima média (em graus Celsius)
- Umidade relativa do ar média (em percentual)

O BDMEP é um banco de dados para apoiar as atividades de ensino e pesquisa e outras aplicações em meteorologia, hidrologia, recursos hídricos, saúde pública e meio ambiente. Esta base de dados abriga dados meteorológicos diários em formato digital, referentes a séries históricas da rede de estação do INMET (291 estações meteorológicas convencionais) num total de cerca de 3 milhões de informações, referentes às medições diárias, de acordo com as normas técnicas internacionais da Organização Meteorológica Mundial. No BDMEP estão acessíveis os dados diários a partir de 1961 das estações para as quais se disponha, em formato digital, de pelo menos 80% dos dados que foram registrados naquele período (INMET, 2014).

As variáveis atmosféricas disponibilizadas para consulta no BDMEP são: precipitação ocorrida nas últimas 24 horas, temperatura do bulbo seco, temperatura do bulbo úmido, temperatura máxima, temperatura mínima, umidade relativa do ar, pressão atmosférica ao nível da estação, insolação, direção e velocidade do vento. A Figura 33, apresenta a localização das estações meteorológicas das quais o BDMEP disponibiliza informações. Nesta base de dados também é possível consultar dados referentes a série histórica com os dados consolidados por mês, estes foram os dados empregados para popular a base de dados para testes do Vis-Saúde.

As variáveis climáticas utilizadas na análise do Vis-Saúde são correspondentes as área geográficas delimitadas como de interesse para os dados de saúde pública (as 7 capitais citadas anteriormente) no mesmo intervalo de tempo definido para estas, ou seja, dados mensais consolidados.

Figura 33 – Mapa com a localização das estações meteorológicas das quais podemos consultar dados no BDMEP



Fonte: Site BDMEP

### 5.3.1.3 Análise com dados reais

A metodologia de análise e visualização da correlação entre variáveis apresentada no Vis-Saúde foi empregada na análise de dados de saúde pública reais obtidos do Sinan e da base de dados climáticos BDMEP. O objetivo desta análise é identificar se os resultados obtidos com o emprego da metodologia desenvolvida no Vis-Saúde são coerentes com as conclusões que foram auferidas pelas pesquisas citadas na seção 2.1 de trabalhos relacionados. As variáveis utilizadas na análise com dados reais foram as listadas abaixo:

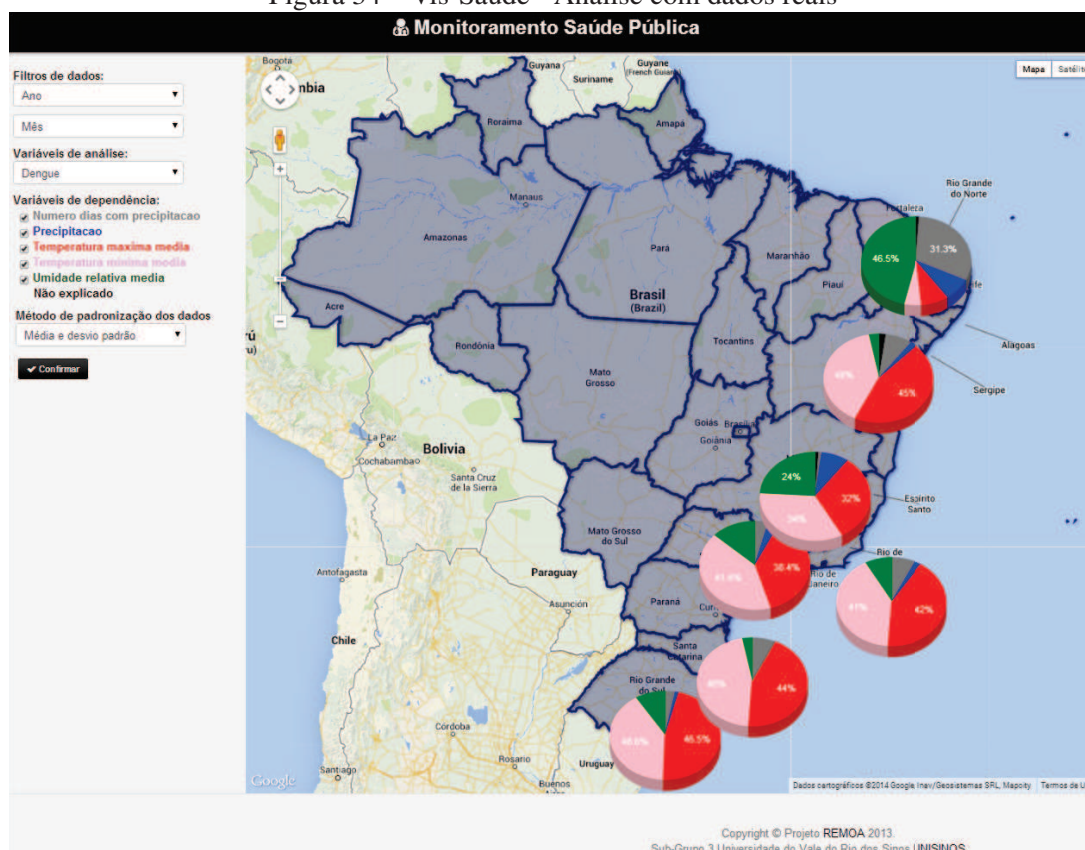
- Saúde Pública: número de casos de dengue (consolidados por mês)
- Climáticas: número de dias com precipitação, precipitação, temperatura máxima média, temperatura mínima média e umidade relativa do ar média (consolidados por mês)

As variáveis citadas acima foram obtidas consolidadas por mês para o período de 2003 até 2012, totalizando 10 anos, para 7 capitais de estados brasileiros: Porto Alegre, Florianópolis, São Paulo, Rio de Janeiro, Belo Horizonte, Fortaleza e Salvador. Logo, a base de dados com registros reais para teste da metodologia somou um total de 5040 registros entre variáveis climáticas e referentes a ocorrência de dengue.

As possibilidades de investigação da relação de dependência entre a ocorrência de dengue frente as variáveis climáticas passa pela definição do período de análise dos dados, se este será mensal, anual ou compreenderá todos os registros de dados. A Figura 34, apresenta o resultado obtido analisando todos os registros de dados sem delimitação de mês ou ano.

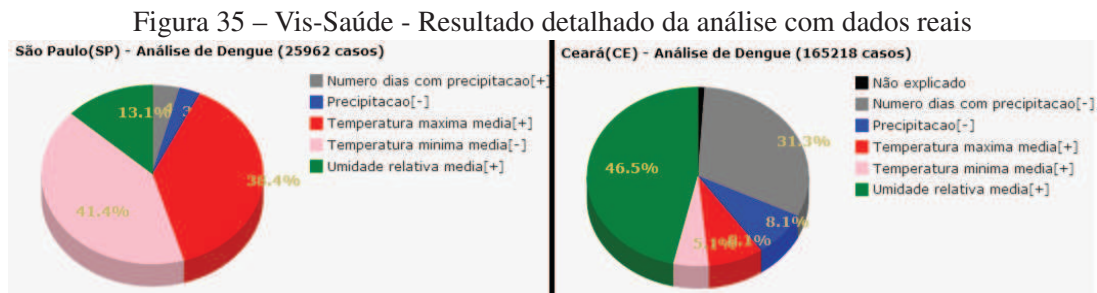


Figura 34 – Vis-Saúde - Análise com dados reais



O resultado apresentado indica uma relação de dependência consistente entre a incidência de dengue e a temperatura máxima média (dependência positiva), temperatura mínima média (dependência negativa) e a umidade relativa média do ar (dependência positiva) em 6 das 7 capitais analisadas. A única cidade que apresentou um resultado mais adverso foi justamente a localizada em uma região mais norte do Brasil, Fortaleza. Nesta capital a análise indica uma relação de dependência mais significativa com o número de dias com precipitação (dependência negativa) e a umidade relativa média do ar (dependência positiva). A variação nos resultados obtidos com a análise proposta no Vis-Saúde pode ser atribuída às condições mais áridas que são encontradas na região nordeste do país onde a cidade está localizada. Os pesquisadores Promprou, Jaroensutasinee e Jaroensutasinee (2005), em seu ensaio citado na seção 2.1 encontraram discrepâncias semelhantes ao avaliar a incidência da dengue em regiões distintas da Tailândia. A pesquisa de citetextoTHAMMPALO2005 corrobora para os resultados obtidos com a análise do Vis-Saúde, pois concluiu que o aumento na temperatura está associada a maior incidência de dengue. Ainda na pesquisa citada a análise por região mostrou que a incidência de dengue estava negativamente associada com o maior índice de chuvas na região Sul, mas positivamente associado com temperaturas elevadas nas regiões centro e norte. Assim é necessário considerar que as particularidades climáticas de cada região acabam por influenciar de forma determinando nos resultados ainda mais se considerado um país tropical de privilegiada extensão territorial como o Brasil. A Figura 35, mostra em detalhes os gráficos gerados com os resultados obtidos

com a análise de dependência entre as variáveis para as cidade de São Paulo e Fortaleza.



Tomando a cidade de São Paulo, como exemplo, e alterando o ano de análise dos dados, as variáveis com maior relação de dependência não se alteram de forma muito significativa permanecendo as de maior coeficiente como sendo a temperatura máxima média, a temperatura mínima média e a umidade relativa média do ar. É possível perceber uma alteração considerável no nível de relação da dengue com as variáveis climáticas precipitação e número de dias com precipitação ao longo dos anos, a Figura 36, evidencia esta variação.

Outra análise relevante que pode ser considerada para a demonstração da metodologia é remover da análise as duas variáveis com maior relação de dependência encontrada, ou seja, remover da análise as variáveis temperatura máxima média e temperatura mínima média. Neste caso, a análise indica um percentual maior de informação que não pôde ser explicada mediante a análise da incidência de dengue em função das variáveis número de dias com precipitação, precipitação acumulada e umidade relativa do ar média. A Figura 37 mostra o resultado desta análise onde é possível verificar o aumento da fatia que indica a quantidade de informação que não pôde ser explicada com as variáveis de análise.

Em todas as análises apresentadas acima, o método de padronização dos dados utilizado considerou a média e o desvio-padrão de cada variável. A análise utilizando apenas a média dos dados enfatizou de forma ainda mais considerável a correlação da incidência da dengue com a temperatura máxima média, temperatura mínima média e umidade relativa do ar média.

Figura 36 – Vis-Saúde - Resultado da análise para a cidade de São Paulo ao longo dos anos

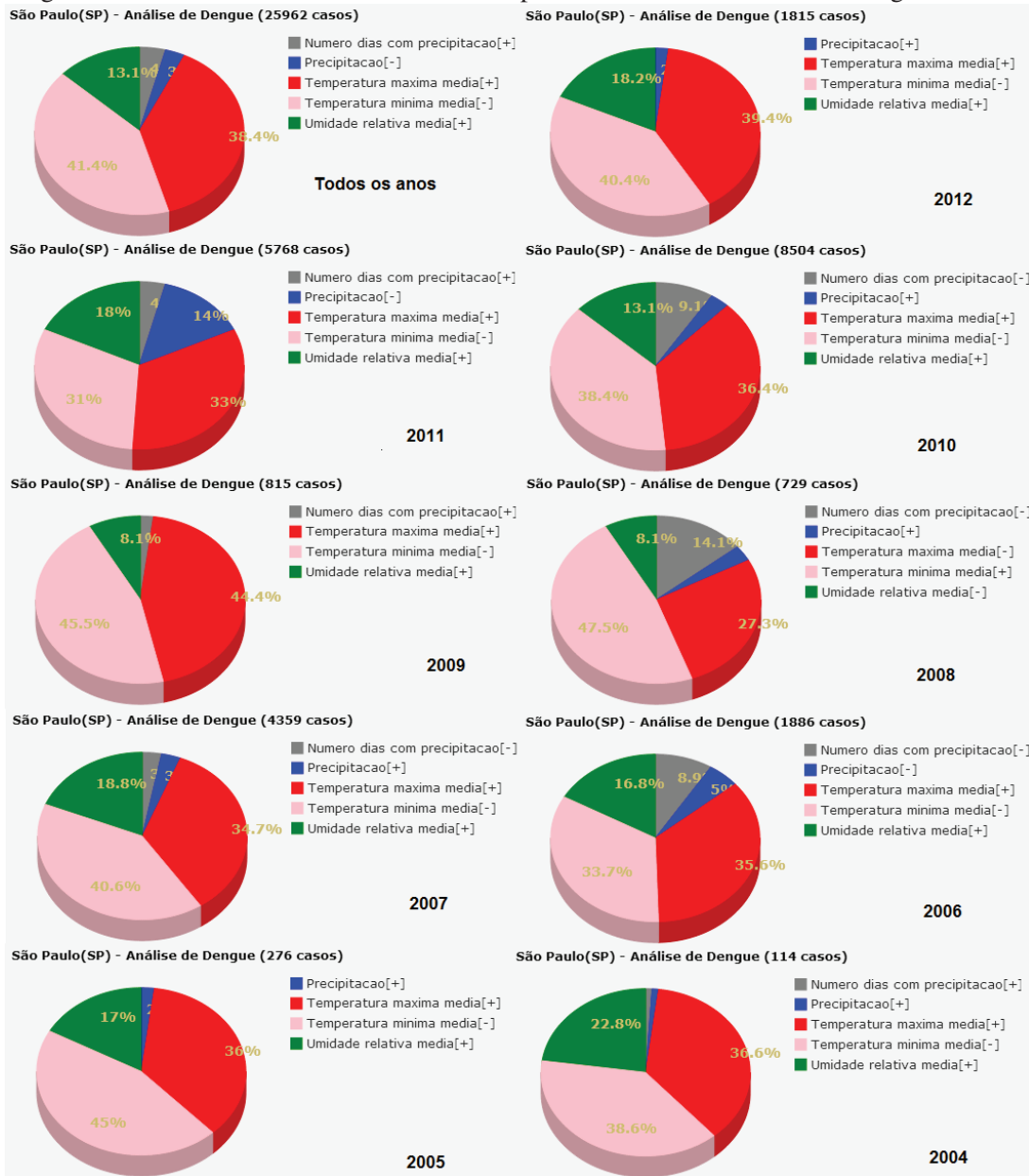


Figura 37 – Vis-Saúde - Resultado da análise sem as duas variáveis de maior dependência

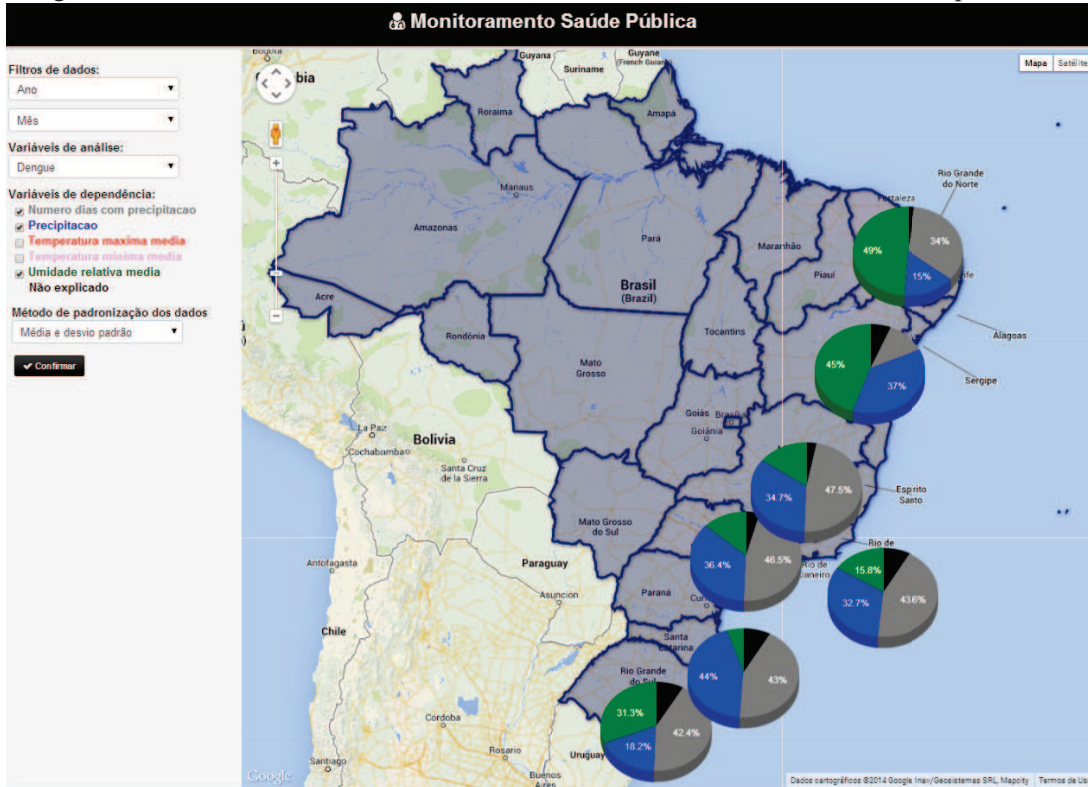
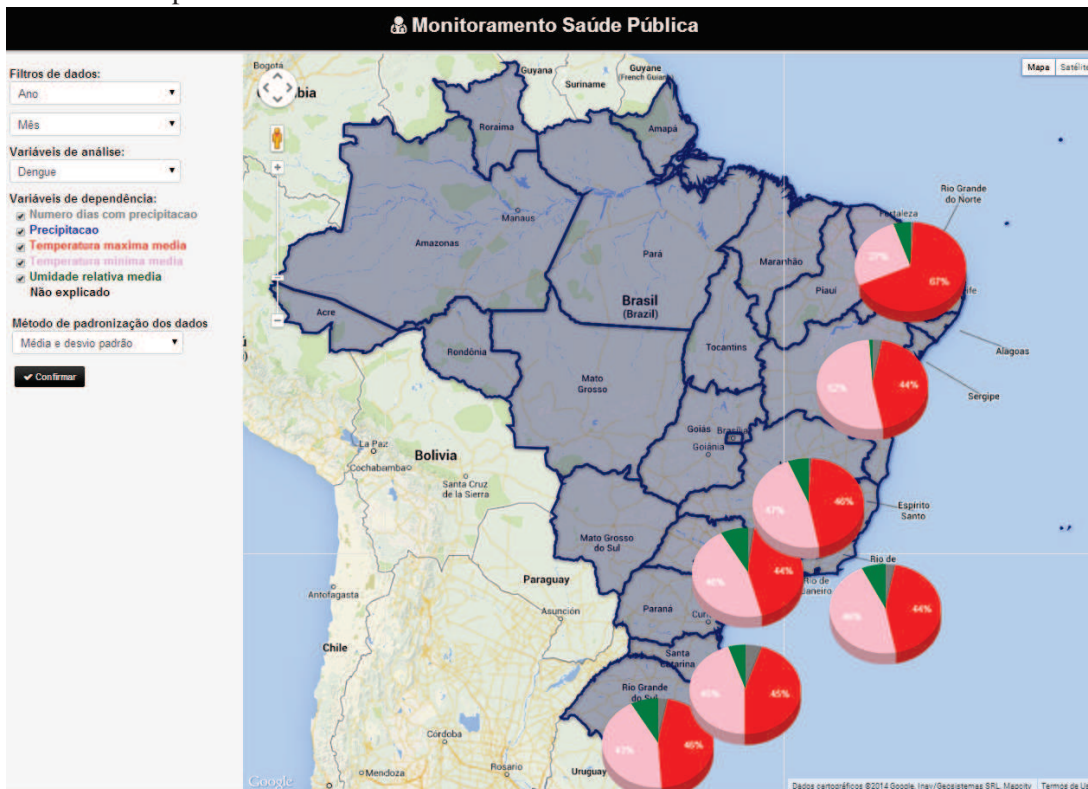


Figura 38 – Vis-Saúde - Resultado da análise considerando como método de padronização das variáveis a média e o desvio-padrão





## 6 CONCLUSÃO

Neste trabalho foi apresentada uma metodologia de análise e visualização de relação de dependência entre variáveis de interesse à um evento. A metodologia desenvolvida é genérica podendo ser aplicada a qualquer conjunto de dados, contudo para ilustrar de forma mais clara a aplicação da metodologia esta foi utilizada em registros de saúde pública, na análise da incidência de dengue em função de variáveis climáticas. O objetivo da análise com dados de saúde não foi estudar propriamente a incidência de dengue mas avaliar a metodologia desenvolvida.

O ambiente de protótipo para testes da metodologia do Vis-Saúde utiliza apenas ferramentas *open source* e é disponibilizada em um ambiente *web* acessível através do navegador de preferência do utilizador, independente de plataforma. No ambiente de protótipo do Vis-Saúde o usuário pode alterar o conjunto de variáveis a serem analisadas e obter os novos resultados da análise em tempo real, visto que todo o processamento ocorre em tempo de execução no servidor.

A metodologia proposta no Vis-Saúde foi testada com dois conjuntos de dados distintos. Os resultados obtidos com a análise dos dados fictícios demonstrou a acurácia da metodologia desenvolvida, que ao ser utilizada em registros de saúde pública frente a variáveis climáticas apresentou resultados consoantes com as demais pesquisas na área que apontam uma correlação significativa entre a incidência da dengue e a temperatura máxima, temperatura mínima e umidade relativa do ar conforme citados nos trabalhos elencados na seção 2.1. Contudo diversas pesquisas têm sido realizadas com intuito de avaliar a correlação da incidência de dengue com variáveis climáticas, sendo que, a maioria dos estudos abrange uma região geograficamente restrita. A análise da incidência da dengue no Brasil em função de variáveis climáticas não pode ser considerada uma tarefa trivial. O país abriga características peculiares devido a sua vasta extensão territorial e por possuir diversos micro-climas distintos em suas regiões. No norte temos uma situação de aridez com temperatura média anual elevada enquanto que na região sul o clima apresenta as quatro estações de forma bem definida. Apesar destes fatos a análise dos dados reais de dengue mostrou uma relação de dependência consistente entre a incidência da dengue e a temperatura máxima média, temperatura mínima média, umidade relativa do ar média e número de dias com precipitação.

Além da análise dos dados, a metodologia contempla uma visualização da concentração de ocorrências de um determinado evento baseado em séries temporais. O emprego de uma técnica de visualização aplicada a dados brutos foi capaz de trazer à tona informações sobre a dinâmica do evento sob análise que antes ficavam à sombra dos números brutos tabulados. Com base nesta visualização é possível identificar uma área epidêmica, bem como sua evolução ao longo do tempo e do espaço geográfico. Até o momento um artigo com ênfase neste tema foi publicado no Workshop de Computação Aplicada em Governo Eletrônico - 2014.

A principal contribuição do desenvolvimento da metodologia de análise e visualização de dados do Vis-Saúde é a utilização da análise de componentes principais para identificar a cor-

relação entre as variáveis associadas a um evento de interesse. Sendo que esta correlação ainda pode ser identificada como positiva ou negativa trazendo ainda mais significância à análise dos dados. As técnicas de visualização de dados propostas no Vis-Saúde se mostraram apropriadas ao objeto de análise proposta para demonstração da aplicação da metodologia. Considerando que a metodologia do Vis-Saúde pode ser aplicado à conjuntos de dados relacionados a outros eventos, talvez seja necessário uma adequação base sobre a qual os gráficos devem ser apresentados. Contudo a metodologia de visualização da relação de dependência entre as variáveis utilizando gráficos em formato de pizza mostrou-se satisfatória a nível informacional.

A metodologia de análise de dados mostrou-se robusta e capaz de identificar a relação de dependência entre as variáveis de análise, não esgotando as possibilidades de explicação do evento mas abrindo novas perspectivas de investigação do evento sob estudo. As técnicas de visualização de dados empregadas mostraram-se adequadas ao propósito de facilitar a abstração e a compreensão das informações apresentadas. Sendo que a visualização da concentração da ocorrência de enfermidades surgiu como uma forma inicial de análise puramente visual dos registros de saúde mas que resulta em ganhos consideráveis para estudar a dinâmica da propagação temporal e espacial de uma enfermidade bem como sua ocorrência sazonal ou endêmica em uma determinada região.

## REFERÊNCIAS

- ARABNIA, H. Reading in information visualization: using vision to think. **IEEE MultiMedia**, [S.l.], v. 6, n. 4, p. 93, 1999.
- ASCHENGRAU, A.; SEAGE, G. R. **Essentials of Epidemiology in Public Health**. 2. ed. [S.l.]: Jones and Barlett, 2008. 516 p.
- BHATT, S.; GETHING, P. W.; BRADY, O. J.; MESSINA, J. P.; FARLOW, A. W.; MOYES, C. L.; DRAKE, J. M.; BROWNSTEIN, J. S.; HOEN, A. G.; SANKOH, O.; MYERS, M. F.; GEORGE, D. B.; JAENISCH, T.; WINT, G. R. W.; SIMMONS, C. P.; SCOTT, T. W.; FARRAR, J. J.; HAY, S. I. The global distribution and burden of dengue. **Nature**, [S.l.], v. 496, p. 504–507, Mar. 2013.
- BIOCASTER. **BioCaster Global health Awareness**. Disponível em: <<http://born.nii.ac.jp>>. Acesso em: junho 2013.
- BLOCH, M. **Tracking Taxi Flow Across the City**. Disponível em: <<http://www.nytimes.com/interactive/2010/04/02/nyregion/taxi-map.html>>. Acesso em: junho 2013.
- BOOTSTRAP. **Introducing Bootstrap**. Disponível em: <<http://twitter.github.io/bootstrap/index.html>>. Acesso em: fev. 2014.
- BUDD, T. **An Introduction to Object-Oriented Programming**. 3. ed. [S.l.]: Addison Wesley Longman, 2002.
- CAMPBELL, K. M.; LIN, C. D.; IAMSIRITHAWORN, S.; SCOTT, T. W. The Complex Relationship between Weather and Dengue Virus Transmission in Thailand. **The American Journal of Tropical Medicine and Hygiene**, [S.l.], v. 89, n. 6, p. 1066–1080, 2013.
- CDC. **Dengue and Climate**. Disponível em: <<http://www.cdc.gov/Dengue/entomologyEcology/climate.html>>. Acesso em: maio 2014.
- CDC. **County Level Estimates of Diagnosed Diabetes — U.S. Maps**. Disponível em: <[http://apps.nccd.cdc.gov/DDT\\_STRS2/NationalDiabetesPrevalenceEstimates.aspx](http://apps.nccd.cdc.gov/DDT_STRS2/NationalDiabetesPrevalenceEstimates.aspx)>. Acesso em: junho 2013.
- CLARK, D. **Beginning C# Object-Oriented Programming**. 2. ed. [S.l.]: Apress Berkely, 2013. 384 p.
- Controladoria Geral da União. **Parceria para Governo Aberto**. Disponível em: <<http://www.cgu.gov.br/PrevencaodaCorrupcao/CompromissosInternacionais/GovernoAberto>>. Acesso em: março 2014.
- DB-ENGINES. **DB-Engines Ranking of Relational DBMS**. Disponível em: <<http://db-engines.com/en/ranking/relational+dbms>>. Acesso em: mar. 2014.
- DULAY, A. V. S.; BAUTISTA, J. R.; TEVES, F. G. Climate Change and Incidence of Dengue Fever (DF) and Dengue Hemorrhagic Fever (DHF) in Iligan City, Lanao del Norte, Philippines. **International Research Journal of Biological Sciences**, [S.l.], v. 2, n. 7, p. 37–41, 2013.

FREIFELD, C.; BROWNSTEIN, J. **About HealthMap**. Disponível em: <<http://www.healthmap.org/about>>. Acesso em: junho 2013.

FREIFELD, C.; BROWNSTEIN, J. **Flu Near You - Do you have it in you?** Disponível em: <<https://flunearyou.org>>. Acesso em: junho 2013.

GOMES, A. F.; NOBRE, A. A.; CRUZ, O. G. Análise temporal da relação entre dengue e variáveis meteorológicas na cidade do Rio de Janeiro, Brasil, no período de 2001 a 2009. **Cadernos de Saúde Pública**, Rio de Janeiro, v. 28, n. 11, p. 2189–2197, Nov. 2012.

GONZALEZ, H.; HALEVY, A.; JENSEN, C. S.; LANGEN, A.; MADHAVAN, J.; SHAPLEY, R.; SHEN, W. Google Fusion Tables: data management, integration and collaboration in the cloud. **ACM Symposium on Cloud Computing**, Indianapolis, June 2010.

GOOGLE. **Google Tendências da Gripe**. Disponível em: <<http://www.google.org/flutrends>>. Acesso em: junho 2013.

GREENBERG, I.; XU, D.; KUMAR, D. **Processing: creative coding and generative art in processing 2**. 2. ed. [S.l.]: Springer, 2013. 445 p.

HALES, S.; WET, N. de; MAINDONALD, J.; WOODWARD, A. Potential effect of population and climate changes on global distribution of dengue fever: an empirical model. **Lancet**, [S.l.], n. 360, p. 830–834, 2002.

HARTLEY, L.; DONNELLY, C.; GARNETT, G. The seasonal pattern of dengue in endemic areas: mathematical models of mechanisms. **Trans R Soc Trop Med Hyg**, [S.l.], v. 96, n. 4, p. 387–397, 2002.

HAY, S. I.; MYERS, M. F.; BURKE, D. S.; VAUGHN, D. W.; ENDY, T.; ANANDA, N.; SHANKS, G. D.; SNOW, R. W.; ROGERS, D. J. Etiology of interepidemic periods of mosquito-borne disease. **Proceedings of the National Academy of Sciences of the United States of America**, [S.l.], v. 97, n. 16, p. 181–220, 2000.

HOPP, M. J.; FOLEY, J. A. Worldwide fluctuations in dengue fever cases related to climate variability. **Climate Research**, [S.l.], v. 25, n. 1, p. 85–94, 2003.

HU, S.; DAI, T. Online Map Application Development Using Google Maps API, SQL Database, and ASP.NET. **International Journal of Information and Communication Technology Research**, [S.l.], Mar. 2013.

INMET. **BDMEP - Banco de Dados Meteorológicos para Ensino e Pesquisa**. Disponível em: <<http://www.inmet.gov.br/portal/index.php?r=bdmep/bdmep>>. Acesso em: abr. 2014.

JACKSON, J. E. **A User's Guide to Principal Component Analysis**. 1. ed. [S.l.]: John Wiley & Sons, 1991. 592 p.

JOHANSSON, M. A.; DOMINICI, F.; GLASS, G. E. Local and Global Effects of Climate on Dengue Transmission in Puerto Rico. **PLoS Neglected Tropical Diseases**, [S.l.], v. 3, n. 2, 2009.

JOLLIFFE, I. T. **Principal component analysis**. 2. ed. [S.l.]: Springer series in statistics, 2002. 487 p.



KEATING, J. An investigation into the cyclical incidence of dengue fever. **Social Science and Medicine**, [S.l.], v. 55, n. 9, p. 1691, 2001.

KEIM, D.; KOHLHAMMER, J.; ELLIS, G.; MANSMANN, F. **Mastering The Information Age - Solving Problems with Visual Analytics**. 1. ed. [S.l.]: Eurographics Association, 2010. 168 p.

KIM, H.-C.; CHUNG, G.-S.; KIM, T.-W. A Framework for Health Management Services in Nanofiber Technique-based Wellness Wear Systems. **Healthcom**, Sydney, p. 70–73, 2009.

KOCH, T. **Disease Maps: epidemics on the ground**. 1. ed. [S.l.]: University of Chicago Press, 2011. 344 p.

KYLE, J. L.; HARRIS, E. Global Spread and Persistence of Dengue. **Annual Review of Microbiology**, [S.l.], v. 62, p. 71–92, Oct. 2008.

Lab Escape. **Heat Map Learning Center**. Disponível em:  
<<http://www.labescape.com/info/articles/what-is-a-heat-map.html>>. Acesso em: junho 2013.

LIN, S.; GAO, Z.; XU, K. Web 2.0 traffic measurement: analysis on online map applications. **Proceedings of the 18th International Workshop on Network and Operating Systems Support for Digital Audio and Video**, New York, NY, USA, p. 7–12, 2009.

LIU, W.; PARK, E.; KRIEGER, U. eHealth Interconnection Infrastructure Challenges and Solutions Overview. **Healthcom**, Beijing, p. 255–260, Oct. 2012.

LOWEA, R.; BAILEYA, T. C.; STEPHENSONA, D. B.; GRAHAMB, R. J.; COELHO, C. A.; CARVALHOD, M. S.; BARCELLOSD, C. Spatio-temporal modelling of climate-sensitive disease risk: towards an early warning system for dengue in brazil. **Computers & Geosciences**, [S.l.], v. 37, n. 3, p. 371–381, 2011.

LU, A.; EBERT, D.; GAO, J.; ZHANG, S.; JOSHI, A. Special Issue on Visualization and Visual Analytics. **Tsinghua Science and Technology**, [S.l.], v. 18, n. 4, p. 109–110, 2013.

MACHADO-MACHADO, E. A. Empirical mapping of suitability to dengue fever in Mexico using species distribution modeling. **Applied Geography**, [S.l.], v. 33, p. 82–93, Apr. 2012.

MASSAD, E.; RENÉ X, d. M.; SILVEIRA, P. S. P.; ORTEGA, N. R. S. **Métodos quantitativos em medicina**. 1. ed. Barueri, SP: Manole, 2004. 570 p.

MATLAB. **MATLAB The Language of Technical Computing**. Disponível em:  
<<http://www.mathworks.com/products/matlab>>. Acesso em: mar. 2014.

Ministério da Saúde. **Sistema de Informação de Agravos de Notificação: normas e rotinas**. 2. ed. Brasília: Editora do Ministério da Saúde, 2007. 68 p.

MONTEIRO, E. S. C.; COELHO, M. E.; CUNHA, I. S. da; AMPARO SALMITO CAVALCANTE, M. do; AMORIM CARVALHO, F. A. de. Aspectos epidemiológicos e vetoriais da dengue na cidade de Teresina, Piauí – Brasil, 2002 a 2006. **Epidemiologia e Serviços de Saúde**, Brasília, v. 18, n. 4, p. 365–374, 2009.

NAKHAPAKORN, K.; TRIPATHI, N. K. An information value based analysis of physical and climatic factors affecting dengue fever and dengue haemorrhagic fever incidence. **International Journal of Health Geographics**, [S.l.], v. 4, p. 31, 2005.

NETLIB. **LAPACK — Linear Algebra PACKage**. Disponível em:

<<http://www.netlib.org/lapack>>. Acesso em: abr. 2014.

PEREZ-LLAMAS, C.; LOPEZ-BIGAS, N. Gitoools: analysis and visualisation of genomic data using interactive heat-maps. **PLoS ONE**, [S.l.], 2011.

PETERSONA, A. T.; MARTÍNEZ-CAMPOSB, C.; NAKAZAWAC, Y.; MARTÍNEZ-MEYER, E. Time-specific ecological niche modeling predicts spatial dynamics of vector insects and human dengue cases. **Transactions of the Royal Society of Tropical Medicine and Hygiene**, [S.l.], n. 99, p. 647–655, 2005.

PROMPROU, S.; JAROENSUTASINEE, M.; JAROENSUTASINEE, K. Climatic Factors Affecting Dengue Haemorrhagic Fever Incidence in Southern Thailand. **WHO Regional Office for South-East Asia**, [S.l.], v. 29, p. 41–48, 2005.

RINGNÉR, M. What is principal component analysis? **Nature Biotechnology**, [S.l.], v. 26, n. 3, p. 303–304, Mar. 2008.

RODRIGUEZ-BARRAQUER, I.; CORDEIRO, M. T.; BRAGA, C.; SOUZA, W. V. de; MARQUES, E. T.; CUMMINGS, D. A. T. From Re-Emergence to Hyperendemicity: the natural history of the dengue epidemic in brazil. **PLoS Neglected Tropical Diseases**, [S.l.], Jan. 2011.

ROGERS, D.; WILSON, A.; HAY, S.; GRAHAM, A. The Global Distribution of Yellow Fever and Dengue. **Advances in Parasitology**, [S.l.], v. 62, p. 181–220, 2006.

SCHRECK, T.; KEIM, D. Visual Analysis of Social Media Data. **Computer**, [S.l.], v. 46, n. 5, p. 68–75, 2013.

Secretaria de Vigilância em Saúde. Especial Tuberculose. **Boletim Epidemiológico**, [S.l.], v. 43, 2012.

Secretaria de Vigilância em Saúde. Influenza 2013: monitoramento da semana epidemiológica 16. **Boletim Epidemiológico**, [S.l.], v. 44, n. 4, 2013.

SESSIONCAM. **Behavioural Website Heatmap**. Disponível em:

<<http://www.sessioncam.com/features-of-sessioncam/website-heatmap>>. Acesso em: junho 2013.

SHAH, S. L.; MITCHELL, W.; SHOOK, D. Tree mapping technology as a means of visualising control performance problems. **Computing & Control Engineering Journal**, [S.l.], v. 16, n. 4, p. 35–39, 2005.

SHNEIDERMANN, B. Tree visualization with tree-maps: 2-d space-filling approach. **ACM Transactions on Graphics**, [S.l.], v. 11, n. 1, p. 92–99, Jan. 1992.

STEWART-IBARRA, A. M.; LOWE, R. Climate and Non-Climate Drivers of Dengue Epidemics in Southern Coastal Ecuador. **The American Journal of Tropical Medicine and Hygiene**, [S.l.], v. 88, n. 5, p. 971–981, 2013.

TEIXEIRA, M. G.; CONCEIÇÃO N. COSTA, M. da; BARRETO, F.; BARRETO, M. L. Dengue: twenty-five years since reemergence in brazil. **Caderno Saúde Pública**, Rio de Janeiro, v. 25, 2009.

THAMMAPALO, S.; CHONGSUWIWATWONG, V.; MCNEIL, D.; GEATER, A. The climatic factors influencing the occurrence of dengue hemorrhagic fever in Thailand. **Southeast Asian Journal Tropical Medicine and Public Health**, [S.l.], v. 36, n. 1, p. 191–196, 2005.

TUN-LIN, W.; BURKOT, T. R.; KAY, B. H. Effects of temperature and larval diet on development rates and survival of the dengue vector *Aedes aegypti* in north Queensland, Australia. **Medical and Veterinary Entomology**, [S.l.], v. 14, n. 1, p. 31–37, 2000.

UDELL, S. **Beginning Google Maps Mashups with Mapplets, KML, and GeoRSS**: from novice to professional. 1. ed. Berkely, CA, USA: Apress Berkely, 2008. 416 p.

VALADARES, A. F.; FILHO, J. R. C.; PELUZIO, J. M. Impacto da dengue em duas principais cidades do Estado do Tocantins: infestação e fator ambiental (2000 a 2010). **Epidemiologia e Serviços de Saúde**, Brasília, v. 22, n. 1, p. 59–66, Mar. 2013.

W3SCHOOLS. **Google Maps API Tutorial**. Disponível em: <<http://www.w3schools.com/googleapi/default.asp>>. Acesso em: abril 2014.

WALKER, D.; FIELDSEND, J.; EVERSON, R. Visualising many-objective populations. **International Conference on Genetic and Evolutionary Computation Conference Companion**, [S.l.], p. 451–458, July 2012.

WALKER, D. J.; EVERSON, R. M.; FIELDSEND, J. E. Visualizing Mutually Nondominating Solution Sets in Many-Objective Optimization. **IEEE Transactions on Evolutionary Computation**, [S.l.], v. 17, n. 2, p. 165–184, 2013.

WHEELER, B. Point-of-care Technologies. **IEEE Pulse**, [S.l.], v. 4, n. 3, p. 6, 2013.

WIKICRIMES. **WiKiCrimes**. Disponível em: <<http://www.wikicrimes.org/>>. Acesso em: junho 2013.

World Health Organization. Global strategy for dengue prevention and control 2012-2020. , [S.l.], 2012.

World Health Organization. **Global Health Observatory - Map gallery**. Disponível em: <[http://www.who.int/gho/map\\_gallery/en](http://www.who.int/gho/map_gallery/en)>. Acesso em: junho 2013.

World Health Organization. Dengue and severe dengue. **Fact sheet**, [S.l.], v. 177, Mar. 2014.