



Manuel António Carvalho Santos

Licenciado em Engenharia Informática

Visualização Cartográfica Assistida num Ambiente SOLAP

Dissertação para obtenção do Grau de Mestre em
Engenharia Informática

Orientador : Doutor João Carlos Gomes Moura Pires,
Professor Auxiliar, Universidade Nova de Lisboa

Júri:

Presidente: Prof. Doutor Nuno Manuel Robalo Correia

Arguente: Prof. Doutora Maribel Yasmina Santos

Vogal: Prof. Doutor João Carlos Gomes Moura Pires



FACULDADE DE
CIÊNCIAS E TECNOLOGIA
UNIVERSIDADE NOVA DE LISBOA

Março, 2013

Visualização Cartográfica Assistida num Ambiente SOLAP

Copyright © Manuel António Carvalho Santos, Faculdade de Ciências e Tecnologia, Universidade Nova de Lisboa

A Faculdade de Ciências e Tecnologia e a Universidade Nova de Lisboa têm o direito, perpétuo e sem limites geográficos, de arquivar e publicar esta dissertação através de exemplares impressos reproduzidos em papel ou de forma digital, ou por qualquer outro meio conhecido ou que venha a ser inventado, e de a divulgar através de repositórios científicos e de admitir a sua cópia e distribuição com objectivos educacionais ou de investigação, não comerciais, desde que seja dado crédito ao autor e editor.

Dedico esta dissertação a todos os que me ajudaram durante a licenciatura e mestrado, especialmente os meus pais.

Agradecimentos

Começo por agradecer ao Professor João Moura Pires por me ter confiado a execução desta dissertação, bem como pela orientação e conselhos que me deu para a realização deste trabalho. Agradeço também ao Ricardo Silva pela ajuda que me deu nas revisões dos documentos bem como pelas secções de *brainstorming* que tivemos para discutir assuntos relacionados com este trabalho. Gostava também de agradecer ao Professor Nuno Correia e ao Nuno Datia pela ajuda na criação do protocolo de validação usado neste trabalho.

Gostaria ainda de agradecer a todos os que me apoiaram durante a execução deste trabalho: André Fidalgo, Bruno Faustino, Filipe Carvalho, Alexandre Gonçalves, Filipe Correia, entre outros.

Por fim um agradecimento especial aos meus pais por todo o apoio que me deram durante a minha licenciatura e mestrado.

Resumo

O conceito *Spatial OLAP* introduz a visualização cartográfica nos sistemas *Business Intelligent/OLAP*. Esta tem grandes potencialidades para melhorar a análise de grandes quantidades de dados georreferenciados e o respetivo processo de tomada de decisão. O potencial que é possível obter da visualização cartográfica depende fortemente do uso apropriado não só das variáveis visuais como também de outros elementos que possam ser utilizados para representar informação no mapa (ex.: gráficos).

Porém, num ambiente *SOLAP*, o analista deveria estar apenas focado em procurar informação útil abstraindo-se de como deve ser construída uma visualização cartográfica adequada. Neste trabalho, é proposto um modelo para assistir o utilizador na construção de visualizações cartográficas.

O modelo proposto considera: (i) a meta-informação existente sobre os modelos multidimensionais; (ii) a interrogação realizada; (iii) os dados resultantes. Através do modelo proposto, pretende-se assistir o utilizador na construção de mapas temáticos apropriados.

Neste trabalho é proposto: (i) um modelo para descrever tipos de visualização; (ii) um modelo para descrever o contexto analítico *SOLAP*; (iii) uma metodologia para a seleção adequada dos tipos de visualização aplicáveis em cada contexto analítico *SOLAP* e sua instanciação.

O modelo de cartografia assistida em ambiente *SOLAP* proposto neste trabalho foi implementado sobre protótipo *SOLAP+*.

Palavras-chave: Cartografia, *SOLAP*, Visualização Assistida, Visualização Cartográfica

Abstract

The concept *Spatial OLAP* brings the cartographic visualization to the *Business Intelligent/OLAP* systems. The cartographic visualization has the potential to improve the analysis of big spatial data sets and the respective decision-making process. The potential obtained from the cartographic visualization is dependent from correct use of the visual variables and other graphics elements used to represent information in the map (ex.: charts).

But, in a *SOLAP* environment, the analyst should be focused in finding useful information abstracting from how to create an appropriate cartographic visualization. In this work, is proposed a model to assist the analyst in the construction of the cartographic visualization.

The proposed model considers: (i) the meta-information in the multidimensional models; (ii) the performed query; (iii) the result data set. By the proposed system, is intended to assist the analyst in the construction of the appropriated thematic maps and the creation of a instance of that visualization.

In this work is proposed: (i) a model to describe visualization types; (ii) a model to describe a *SOLAP* analytic context; (iii) a methodology to select an appropriate applicable visualization type in which *SOLAP* analytic context.

The assisted cartographic model in a *SOLAP* environment, proposed in this work is implemented in the *SOLAP+* prototype.

Keywords: Cartography, SOLAP, Assisted Visualization, Cartographic Visualization

Conteúdo

1	Introdução	1
1.1	Contexto e Motivação	1
1.2	Objetivos	3
1.3	Abordagem e Contribuições	4
1.4	Estrutura do Documento	5
2	Estado da Arte e Trabalho Relacionado	7
2.1	SOLAP	7
2.1.1	SOLAP+	8
2.2	Visualização	11
2.2.1	Visualização de Dados não Espaciais	11
2.2.2	Cartografia num Contexto SOLAP	13
2.2.3	Heurísticas de Visualização de Informação	16
2.3	Sistemas de Visualização Assistida	19
2.4	Conclusão	20
3	Visualização Cartográfica Assistida	21
3.1	Contexto SOLAP	22
3.2	Visualização Cartográfica	25
3.2.1	Variáveis Visuais	26
3.2.2	Tipos de Visualização	27
3.3	Sistema de Visualização Cartográfica Assistida	30
3.3.1	Tipos de Visualização Aplicáveis	30
3.3.2	Fábricas Visuais	33
3.3.3	Mapeamento	34
3.3.4	Preferências	36
4	Arquitetura e Implementação	39
4.1	Arquitetura	39

4.2	Implementação	41
4.2.1	Fábricas Implementadas	43
4.2.2	Limitações Tecnológicas	44
5	Validação	45
5.1	Desempenho e Taxa de Retenção	45
5.2	Protocolo Experimental	46
5.3	Resultados do Protocolo Experimental	48
6	Conclusão e Trabalho Futuro	53
A	Paleta de Tipos de Visualizações	59
B	Questões usadas no Protocolo de Validação	63

Lista de Figuras

1.1	Distribuição da População Portuguesa por Distrito	2
1.2	Distribuição da População Portuguesa por Distrito (com legenda)	3
2.1	Interface do SOLAP+, em [Jor09]	9
2.2	Exemplo de Agrupamento Espacial	10
2.3	Sparklines	11
2.4	Time Spirals	12
2.5	Invasão da Rússia pelas tropas de Napoleão	13
2.6	Exemplo de um Mapa de Símbolos Proporcionais	14
2.7	Relações entre objetos Espaciais	16
2.8	Gráfico contendo "Chartjunk" de Nigel Holmes usado em [BMG ⁺ 10]	18
3.1	Sistema de Visualização Cartográfica Assistida	22
3.2	Modelo Multidimensional de Acidentes Rodoviários nos Estados Unidos	23
3.3	Interrogações SOLAP	23
3.4	Smileys de Várias Cores	26
3.5	Modelo de Tipo de Visualização	28
3.6	Tipo de Visualização: <i>Smileys</i>	28
3.7	Torre de Extração de Petróleo	28
3.8	Tipo de Visualização: Ponto Tamanho	29
3.9	Mapa com Gráficos de Setores	29
3.10	Tipo de Visualização: Gráfico de Setores	29
3.11	Tipo de Visualização: Gráfico de Barras	30
3.12	Exemplo Tipos Aplicáveis: Interrogação	31
3.13	Modelo Para Fábricas Visuais	33
3.14	Tipos de Visualização sobre a Fábrica Ponto	34
3.15	Exemplo de Mapeamento Seletivo	35
3.16	Exemplo de Mapeamento Ordenável	35
3.17	Mapeamento para a Variável Visual <i>Etiquetas</i>	36

4.1	Arquitetura Original do SOLAP+	40
4.2	Arquitetura Atual do Servidor SOLAP+	41
4.3	Exemplo da Sintaxe para Definir Fábricas e Tipos de Visualização	42
4.4	Mapa usando <i>Smileys</i> como Representação Gráfica	43
5.1	Questão usada no Protocolo Experimental	47
5.2	Caminho de Perguntas 1	48
5.3	Caminho de Perguntas 2	48
5.4	Comparação do Tempo de Resposta com e sem VCA	49
5.5	Percentagem de Respostas Certas com e sem VCA	50
5.6	Comparação do Tempo de Resposta Usando Contextos Repetidos	50
5.7	Percentagem de Respostas Certas Usando Contextos Repetidos	51
5.8	Comparação do Tempo de Resposta nas Perguntas Repetidas sem Suporte Visual	51
5.9	Percentagem de Resposta Certas nas Perguntas Repetidas sem Suporte Visual	52

Lista de Tabelas

2.1	Níveis de Organização das Variáveis Visuais	14
2.2	Mapeamento Adequado das Variáveis Visuais para Tipos de Dados	15
3.1	Resultado parcial da Interrogação { <i>Estado</i> , SUM (<i>Mortes</i>), AVG (<i>Pessoas Envolvidas</i>)}	24
3.2	Resultado parcial da Interrogação { <i>Estado</i> , <i>Ano</i> , SUM (<i>Mortes</i>), AVG (<i>Pessoas Envolvidas</i>)}	24
3.3	Resultado parcial da Interrogação { <i>Fábrica</i> , <i>Poluente</i> , SUM (<i>Emissões</i>)}	31
3.4	Fragmento da Paleta de Tipos de Visualização	32
3.5	Variáveis Visuais da Fábrica Ponto	33
3.6	Variáveis Visuais da Fábrica Gráfico de Setores	34
3.7	Transformação da Tabela 3.3	36
3.8	Exemplo do uso de Preferências	37
5.1	Taxa de Retenção	46
5.2	Estatísticas sobre a População	48
5.3	Resultados do Teste Z Para os Tempos de Resposta com e sem <i>VCA</i>	49
5.4	Resultados do Teste Z Usando Contextos Repetidos	50



Introdução

Neste capítulo será apresentada a motivação da realização desta dissertação bem como o contexto em que ela se insere. De seguida serão apresentados os problemas que se pretendem resolver e os objetivos desta dissertação. Por fim serão apresentadas as contribuições que se esperam deste trabalho e será apresentada uma visão global do documento.

1.1 Contexto e Motivação

A adoção generalizada de sistemas de informação para a gestão das operações das organizações, levou à produção de enormes volumes de dados que já são usados desde os anos 80, para analisar o seu desempenho e desse modo servindo de base a muitas decisões, quer estratégicas quer operacionais. O termo *Business Intelligence (BI)*, usado na indústria e posteriormente na academia, engloba um conjunto de atividades, processos, metodologias e instrumentos para acederem, processar e organizar os dados com vista a sua exploração analítica para dar suporte à decisão. Dentro da área de *BI*, uma área que se destacou é conhecida por *OLAP* (cuja expansão do termo é *Online Analytical Processing*) e que se preocupa com a criação de modelos e ferramentas para que os analistas (especialistas do negócio) possam analisar grandes volumes de dados, de modo interativo.

Com a expansão das comunicações móveis e dos serviços *on-line* grande parte dos dados produzidos pelos sistemas operacionais passaram a ser georreferenciados. Os sistemas *OLAP* revelam algumas limitações na análise de dados espaciais. Bédard introduz o conceito *Spatial OLAP (SOLAP)* [RBM01] que resulta da integração de dados espaciais em sistemas *OLAP*. Enquanto que a visualização nos sistemas *OLAP* assenta maioritariamente em gráficos e tabelas nos sistemas *SOLAP*, para além dos recursos já descritos, é possível usar mapas temáticos ou mapas com gráficos.

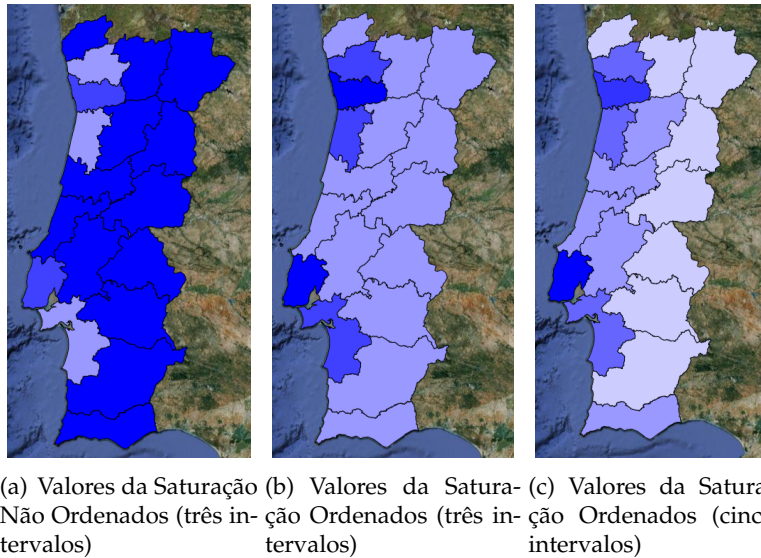


Figura 1.1: Distribuição da População Portuguesa por Distrito

Um mapa temático representa a distribuição de um fenómeno numa área geográfica [KO03]. Ao usar mapas para representar dados ganha-se um poder de análise muito maior [BRP07], mas por outro lado produzir mapas temáticos adequados ao contexto analítico, e que permitam uma rápida interpretação, não é, em geral, uma tarefa fácil. A criação de mapas temáticos é, em geral, um trabalho especializado, normalmente realizado por um cartógrafo. Num contexto analítico, no caso presente *SOLAP*, é essencial que a interação entre o utilizador (em geral um especialista do domínio [Che05]) seja fluida e em cada passo da interação os mapas gerados estejam corretos e adequados.

Os sistemas *SOLAP* são altamente interativos e as análises são feitas de um modo exploratório "*Give me what I say I want, then I can tell you what I really want*" [Inm05]. Deste modo, a criação de mapas deve acompanhar esta fluidez na interação e não estar dependente de um cartógrafo externo.

Para além de problemas relacionados com a correta representação de dados, nas empresas, muitas vezes o resultado de uma análise não é só para ser vista pelo analista mas também por outras pessoas, o que pode levar a problemas de comunicação [DB12].

Na Figura 1.1 está representada a distribuição da população portuguesa por distrito, usando três mapas temáticos diferentes para mostrar essa informação. Nos três mapas os valores a serem representados estão codificados usando a saturação do objeto espacial. No mapa 1.1(a) os intervalos não estão ordenados o que faz com o interior do país tenha um valor de saturação maior, orientando a perceção de que o interior do país tem mais população; no mapa 1.1(b) os intervalos foram ordenados e já se consegue visualizar de forma correta os dados. Apesar do mapa 1.1(b) estar correto segundo as regras cartográficas que serão discutidas na Secção 2.2.2, ainda é possível melhorar a qualidade da visualização. Ao aumentar o número de intervalos de três para cinco, ver mapa 1.1(c), os

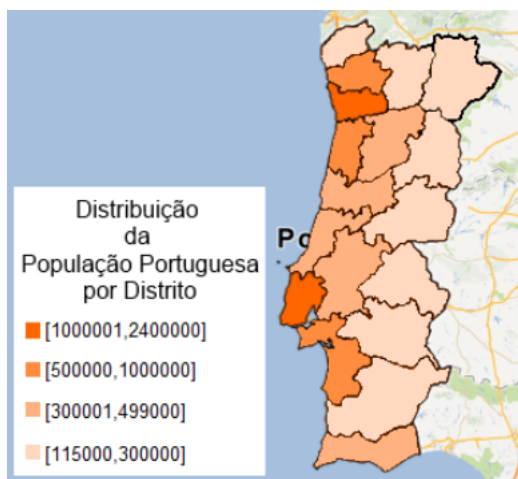


Figura 1.2: Distribuição da População Portuguesa por Distrito (com legenda)

dados visualizados tornam-se ainda mais claros. A melhoria na análise deve-se a capacidade para extrair mais informação no mapa 1.1(c) do que no mapa 1.1(b). Mas o mapa 1.1(c) ainda pode ser melhorado aumentando o contraste entre os dados representados e o mapa base e colocando a respetiva legenda, como mostra o mapa 1.2. A adição da legenda permite ter uma noção dos valores que a saturação está a codificar. Este exemplo mostra que mesmo encontrado um mapa que siga os princípios cartográficos, este pode não ser o mapa mais adequado para representar a informação pretendida.

Os sistemas *SOLAP* devem proporcionar uma forte assistência ao utilizador na escolha de mapas temáticos ou mapas com gráficos que sejam apropriados ao contexto analítico. A não ser em sistemas com fins aplicativos muito específicos ([SP05]), nos atuais sistemas tal assistência é nula ou muito limitada devendo o utilizador escolher o tipo de mapa e a sua aplicação concreta ao contexto analítico.

1.2 Objetivos

Neste trabalho pretendemos endereçar o problema de, no contexto de um sistema *SOLAP* genérico, definir quais as visualizações aplicáveis num dado momento da análise, e garantir que a sua aplicação segue bons princípios de cartografia e da visualização de informação. Naturalmente pretende-se que a abordagem a seguir seja capaz de acomodar outras formas de visualização de informação no mapa além daquelas que forem identificadas durante este trabalho.

O modelo proposto considera: (i) a meta-informação existente sobre os modelos multidimensionais; (ii) a interrogação realizada; (iii) os dados resultantes; (iv) e o histórico de interação com o utilizador. Através do modelo proposto, pretende-se assistir o utilizador na construção de mapas temáticos apropriados.

O primeiro problema a ser resolvido é como representar uma visualização. Para resolver este problema é necessário saber o que é possível representar num mapa e como. Para

além disso também é necessário saber quando é que uma visualização é mais apropriada para um conjunto de dados.

O segundo problema é como guiar o analista na escolha de uma visualização ou mesmo sugerir uma. Para tal é preciso conseguir caracterizar como é que é feita a análise e conseguir avaliar uma visualização para representar um conjunto de dados.

Deste modo, os objetivos deste trabalho passam por: (i) definir o que é uma visualização e quando pode ser usada; (ii) obter informação sobre a análise que está a decorrer; (iii) criar um modelo que consiga guiar o utilizador na escolha de uma visualização para a análise a decorrer ou mesmo sugerir uma para tal efeito.

1.3 Abordagem e Contribuições

A abordagem adotada está dividida em três partes: (i) Caracterizar o que é um contexto analítico *SOLAP*; (ii) Definir o que é uma visualização; (iii) Sugerir e aplicar visualizações para uma análise.

Na caracterização do que é um contexto analítico *SOLAP* é preciso perceber que recursos estão disponíveis durante uma análise que a consigam caracterizar. Estes recursos podem ser o modelo multidimensional, o histórico das ações realizadas, a interrogação realizada, os tipos de dados que estão na clausula *select* e as estatísticas que resultam dos dados. Para além de identificar os recursos é preciso para cada um deles identificar que informação é relevante e para quê.

A segunda tarefa é a modelação das visualizações. Para efetuar a modelação das visualizações é preciso identificar as variáveis visuais que podem ser usadas, como estas se comportam e que propriedades é que existem nos objetos espaciais (pontos, linhas, polígonos). Por fim, é preciso também identificar que tipo de dados devem ser mapeados por uma variável visual

A última tarefa é a criação de um modelo para assistir o utilizador *SOLAP* na construção da visualização cartográfica. Este modelo irá funcionar em duas fases. Na primeira fase o sistema irá filtrar todas as visualizações disponíveis escolhendo apenas as que são aplicáveis naquele contexto analítico. Na segunda fase o sistema vai instanciar as visualizações e no final acaba por sugerir uma ao utilizador. O sistema usará a informação contida no contexto analítico *SOLAP* para a realização das tarefas descritas anteriormente.

As contribuições desta tese serão: (i) um modelo para descrever visualizações no contexto *SOLAP*; (ii) um modelo para extrair informação de uma análise *SOLAP*; (iii) um sistema que consiga guiar o utilizador na escolha de uma visualização e apoia-lo na sua aplicação; (iv) um conjunto significativo de visualizações (descritas no modelo proposto) incluído algumas com o fim de demonstrar que o modelo proposto é extensível;

1.4 Estrutura do Documento

Este documento está dividido em seis capítulos: (i) introdução; (ii) estado da arte e trabalho relacionado; (iii) visualização cartográfica assistida; (iv) arquitetura e implementação (v) validação; (vi) conclusão. O Capítulo 2, inicia-se com uma pequena secção em se apresentam sumariamente os conceitos fundamentais do *SOLAP*, e se descrevem alguns sistemas *SOLAP* e em particular o sistema *SOLAP+* que será usado como plataforma de implementação. Neste capítulo também serão discutidos aspetos de visualização de informação em geral e em particular num mapa. Para além de aspetos práticos relativos à visualização serão apresentadas heurísticas usadas para avaliação de visualização. O capítulo termina com a apresentação de alguns sistemas que usam visualização assistida.

No Capítulo 3 será explicada a abordagem tomada para resolver o problema apresentado, dividindo-se em três partes. Na primeira será explicada a obtenção de informação de uma análise. Na segunda parte será apresentado o modelo para criar visualizações. Por fim na última parte será apresentado o modelo proposto para cartografia assistida.

No Capítulo 4 será apresentado como o modelo proposto no capítulo anterior foi implementado bem como as modificações que foram feitas ao sistema *SOLAP+*.

Por fim os últimos dois capítulos serão dedicados a apresentar os resultados obtidos em vários testes realizados, Capítulo 5, e no Capítulo 6 serão discutidas as conclusões obtidas e trabalho futuro.



Estado da Arte e Trabalho Relacionado

Este capítulo inicia-se com uma apresentação de conceitos *SOLAP*. De seguida é feita uma exposição de métodos relativos a visualização de informação (tanto em mapas como em ambientes mais clássicos), e a sua respetiva avaliação. Por fim serão discutidos alguns sistemas que implementam visualização assistida/inteligente/automática.

Na Secção 2.1 serão vistos vários conceitos relacionados com *SOLAP* e com *OLAP*. A apresentação destes conceitos revela-se importante devido ao trabalho realizado nesta dissertação estar enquadrado no contexto *SOLAP*. Após a apresentação de vários conceitos *SOLAP*, na Secção 2.2 são apresentados métodos de visualização usados quer em ambientes clássicos de *Business Intelligence* quer em ambientes *SOLAP*. Na Secção 2.2.3 serão apresentados alguns conjuntos de heurísticas usadas para avaliação de visualizações de informação. Por fim serão analisados vários sistemas de visualização assistida/inteligente/automática, Secção 2.3 sendo discutido as suas vantagens e desvantagens.

2.1 SOLAP

O conceito *SOLAP* foi introduzido por Bédard et al [RBM01] e tem como objetivo entender o conceito *OLAP* para suportar dados geográficos tanto nas dimensões como nas tabelas de facto. Isto introduz dois conceitos [BTMP10]: (i) dimensão espacial; (ii) métrica espacial. Uma dimensão espacial contém pelo menos um atributo geográfico. No caso das métricas espaciais existe trabalho nesta área [Bim10] [MZ07] mas ainda se está longe de chegar a um consenso sobre a sua definição.

Cada dimensão pode indexar os dados a diferentes níveis de granularidade. As hierarquias são definidas utilizando os diferentes níveis da dimensão. Malinowsky e Zimányi definiram vários tipos de hierarquias [MZ05] sendo a mais comum a que representa as dimensões como uma árvore denominada hierarquia espacial simples, por exemplo freguesia, concelho e distrito.

No *SOLAP* existem as operações comuns do *OLAP* tais como: *slices*, *drill-downs* e *roll-ups*. Além destas operações passa a ser possível ter operações sobre os dados geográficos, designadas de *slices* espaciais. Em seguida são apresentados alguns exemplos de *slices* espaciais [BTMP10]: (i) encontrar as cinco lojas mais perto de um local (vizinhos mais próximos); (ii) estar a distância de 1 km (estar a uma distância); (iii) estar dentro de uma certa freguesia (encontrar-se dentro do outro objeto).

Em [RBM01] são propostas diretrizes relativas à visualização de informação em *SOLAP*. Os autores referem a importância da sincronização entre as diferentes formas de visualização de dados (mapa, tabelas e gráficos) e da construção de mapas temáticos adequados. Para além disso, referem que *"a user of a SOLAP tool should be able to modify the graphical semiology according to its specific needs, in order to highlight relevant information. (...) The legend should be modifiable (...) This would facilitate reclassifying for instance, or highlighting categories."* Logo, o potencial que é possível obter da interpretação dos mapas depende, não só da definição correta do mapa, como também da possibilidade de um mapa interativo como por exemplo a capacidade de alterar a legenda, a simbologia do mapa, entre outros.

Baseados nestes princípios foram desenvolvidos vários sistemas. Grande parte destes sistemas são apenas académicos mas nos últimos anos começaram a aparecer sistemas comerciais. Em seguida são mencionados alguns sistemas *SOLAP*: (i) *SOLAP+* [Mat06] [Jor09] [Sil10] [La11]; (ii) *SOVAT* [SP05]; (iii) *Globeolap* [FS10]; (iv) *Gewolap* [BWTM06]; (v) *JMap* [K2 12]; (vi) *Map4Decision* [Int12].

De notar que a Oracle e a Microsoft são dois dos maiores vendedores de suites de *Business Intelligence* e de base de dados mas nenhuma detém sistemas *SOLAP*. Ambas suites de produtos tem suporte para dados geográficos mas nenhum deles tem sistemas *SOLAP*. A Oracle suporta a visualização de dados espaciais através do *Oracle MapViewer* mas este não está diretamente integrado com a suite de *BI*. No caso da Microsoft existe possibilidade de integrar a base de dados com o seu serviço de mapas, mas este não está integrado na sua suite de *BI*.

Dado que o modelo de visualização cartográfica assistida vai ser implementado no sistema *SOLAP+* este vai ser explicado com algum detalhe na Secção 2.1.1.

2.1.1 SOLAP+

O prototipo *SOLAP+* tem sido desenvolvido por uma linha de trabalho orientada pelo Prof. João Moura-Pires ao longo de várias iterações. O objetivo deste projeto é criar uma implementação genérica de *SOLAP*.

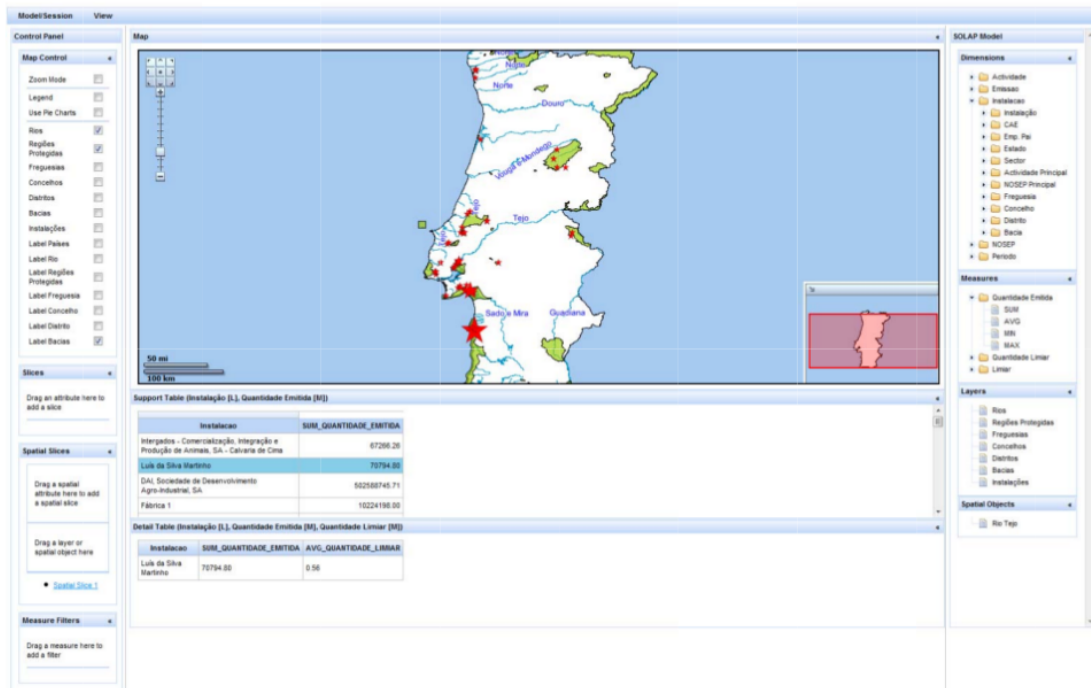


Figura 2.1: Interface do SOLAP+, em [Jor09]

A primeira iteração começou com o trabalho realizado por [Mat06] para definir como seria feita a integração entre dados geográficos e *OLAP*. A primeira iteração do projeto acabou com a implementação de um protótipo em [Jor09]. Este protótipo suportava os métodos de exploração comuns em *OLAP* (*slices, drill-down, roll-up*), estava limitado a atributos espaciais nas dimensões e o mecanismo de visualização não era extensível.

Na Figura 2.1 está representada a interface do *SOLAP+* resultante desta iteração. À direita encontra-se o modelo com a descrição das dimensões, métricas e informação espacial. À esquerda estão os controladores usados para realizar *slices*, filtros e outras operações. No centro está o mapa, a tabela de suporte e a tabela de detalhe.

Uma propriedade fundamental neste sistema é a relação de *1:1* entre a representação no mapa e a tabela de suporte. Esta propriedade é mantida em qualquer situação. Para demonstrar a importância desta propriedade, vamos verificar as consequências da ausência dela em dois casos: (i) múltiplas linhas na tabela de suporte associadas a um objeto gráfico; (ii) múltiplos objetos gráficos associados a uma linha na tabela de suporte. O primeiro caso pode levar a mapas temáticos confusos, pois seria necessário apresentar os dados das diferentes linhas associadas ao objeto espacial no mapa. Além disso, o processo de associação entre os dados presentes na tabela de suporte e o objeto espacial pode ser complicado, caso as linhas pertencentes ao mesmo objeto estejam dispersas pela tabela. Na segunda situação, o utilizador não será capaz de observar de que forma os diferentes objetos espaciais contribuem para a informação guardada na linha da tabela de suporte. Por outro lado, se considerarmos objetos espaciais distantes uns dos outros, o

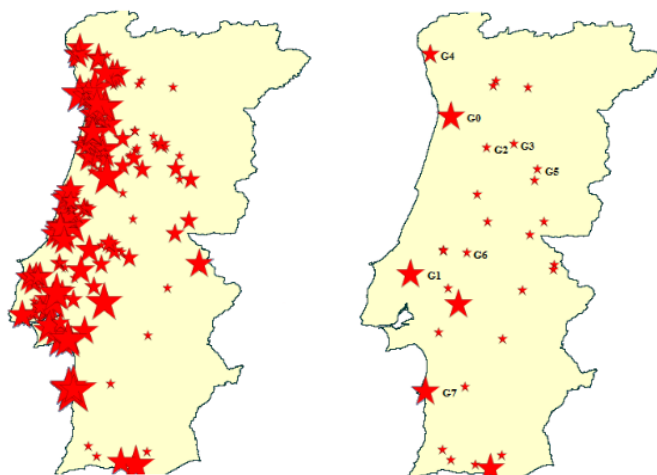


Figura 2.2: Exemplo de Agrupamento Espacial

processo de associação entre a tabela de suporte e o mapa tornar-se-á bastante complicado. A tabela de detalhe serve para realizar *drill-downs* no nível espacial que está a ser mostrado. Por exemplo se na tabela de suporte o nível espacial for distrito e existe um nível concelho abaixo deste na mesma hierarquia, então na tabela de detalhe aparecem os dados agregados por concelho. A relação entre as linhas da tabela de suporte e da detalhe é de $1:N$.

A iteração seguinte desenvolvida por Ricardo Silva em [Sil10], tinha como objetivo resolver dois problemas detetados na primeira iteração: (i) representar relações entre dois objetos espaciais de hierarquias diferentes; (ii) melhorar a visualização de muitos objetos em simultâneo no mapa. As relações entre objetos espaciais vão ser discutidas na Secção 2.2.2. Relativamente ao ponto (ii) foi proposto um modelo de pós-processamento que se baseia na técnica de agrupamento espacial para melhorar a visualização de muitos objetos em simultâneo no mapa, como mostra a Figura 2.2. No mapa da esquerda estão representados os dados relativos às emissões de poluentes por fábrica. À direita está demonstrado o agrupamento espacial aplicado aos dados do primeiro mapa. As estrelas identificadas com "G" representam os agrupamento que foram criados.

Apesar do protótipo já suportar um conjunto interessante de funcionalidades ainda tem algumas lacunas, como não suportar de forma modular novos tipos de visualizações e não existe nenhum tipo de assistência na escolha da visualização. Como foi referido a forma de visualização atual não está extensível, tendo sido definido um conjunto de regras para mapear os valores em: (i) uma ou duas variáveis visuais dos objetos espaciais; (ii) gráficos de linhas, barras e setores. Quando os dados a serem mapeados tinham apenas uma ou duas colunas não espaciais eram usadas as variáveis visuais, caso contrario eram usados os gráficos no mapa.

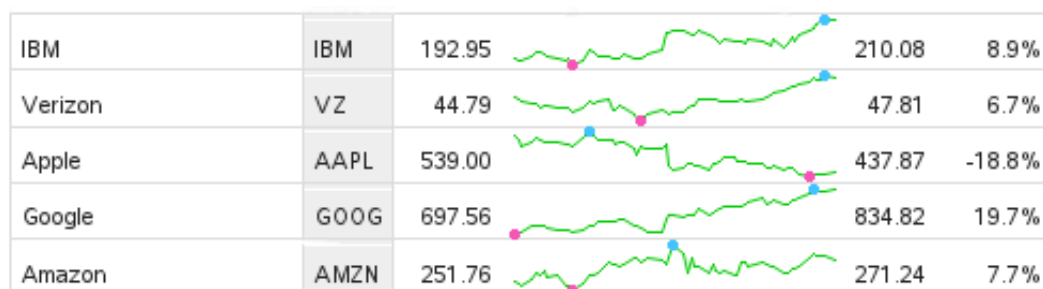


Figura 2.3: Sparklines

2.2 Visualização

Nesta secção serão apresentadas várias formas de visualização de informação. As primeiras técnicas apresentadas são tipicamente utilizadas em ambientes *OLAP* (ver Secção 2.2.1). As técnicas apresentadas serão dedicadas a visualização de dados alfanuméricos e temporais.

Na Secção 2.2.2 serão apresentados tópicos relativos à cartografia nos sistemas *SO-LAP*. A primeira parte desta secção será dedicada a discutir como representar dados associados a objetos espaciais. Na segunda parte serão apresentadas formas de visualização para representar relações entre objetos espaciais.

Por fim, na secção Secção 2.2.3, serão discutidas heurísticas para visualização. Estas foram propostas na literatura utilizando como inspiração as heurísticas para avaliar interfaces.

2.2.1 Visualização de Dados não Espaciais

A visualização de informação em sistemas *OLAP* assenta maioritariamente em tabelas, nomeadamente *tabelas pivô*, e gráficos levando estas visualizações a serem adotadas por um grande número de sistemas. Muitas vezes associados a estes sistemas estão ferramentas para a criação de relatórios ou *dashboards* [RBC09] em que a forma de visualização predileta são os gráficos. Entre os mais usados estão os gráficos de barras, linhas e de sectores [BSBC03] [Few04] [Tuf83].

Para além desta formas de visualização existem outras mais "exóticas" [HBO10], como por exemplo: (i) *Horizon Graphs* [Few08]; (ii) *SPLM (Scatter Plot Matrix)* [Har75]; (iii) *Parallel Coordinates* [Ins85]. Estas tendem a manter a mesma facilidade de uso das anteriores.

Quando se trata da visualização de dados temporais, as visualizações podem dividir-se em duas categorias [EAAB08] : (i) estática; (ii) interativa. Na visualização estática os dados são normalmente representados como gráficos, sendo o mais usado para esta representação o gráfico de linhas [Few04]. Apesar de ser uma representação simples e fácil de interpretar revela alguns problemas associados a mostrar muitas séries diferentes e grandes volumes de dados. Para responder a estes problemas Edward Tufte criou

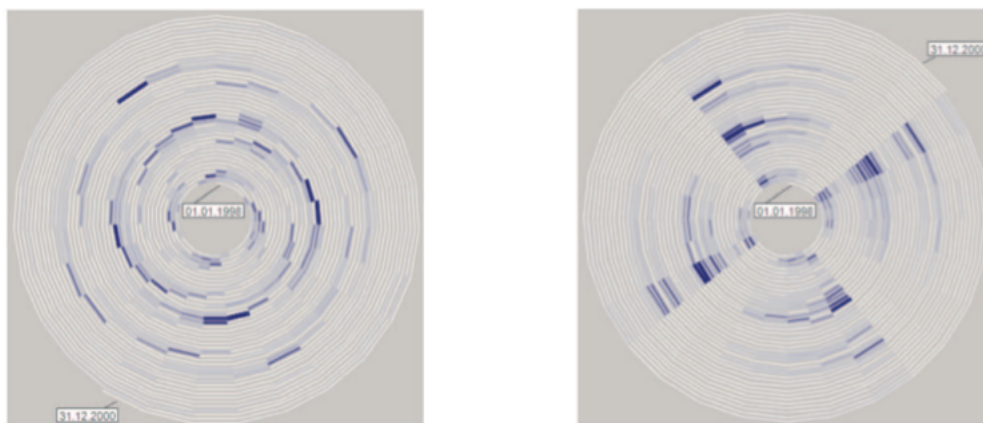


Figura 2.4: Time Spirals

as *sparklines* [Tuf06], estas podem ser definidas como "a font-high, one-word wide time series plot"[TW10]. A título de exemplo, na Figura 2.3 estão representadas *sparklines* que mostram a evolução do valor de algumas ações num período de 60 dias.

Tanto os gráficos de linhas como as *sparklines* tem implícita uma noção de tempo linear. No entanto os dados temporais podem ser modelados de outras formas como foi proposto em [Fra98]. Com estas extensões ganhou-se um maior poder analítico, porque começou-se a ver o tempo de outras formas. Por exemplo ver a semana como um ciclo para comparar dados de cada dia com o período homólogo, podemos tentar encontrar padrões que ocorrem em ciclos. Para representar este tipo de análises foi proposto em [WAM01] o conceito de *Time Spirals*.

Na Figura 2.4 à esquerda está representada uma *Time Spiral* em que o ciclo é de 27 dias, enquanto que a direita estão representados os mesmos dados mas usando um ciclo de 28 dias. Neste caso consegue-se identificar um padrão cíclico nos dados usando como ciclo 28 dias.

As representações apresentadas anteriormente podem ser usadas para contextos não iterativos. Andrienko et al. em [EAAB08] propõe vários mecanismos iterativos para visualizar dados temporais, como uma *timeline* para que consiga mostrar o tempo a várias granularidades. Nesta *timeline* é possível por exemplo ver o tempo ao longo dos anos, depois selecionando um ano a *timeline* passa para meses e assim sucessivamente.

Atualmente no SOLAP+ não existem formas de visualizações dedicadas a representar tempo. Os dados temporais são tratados como dados ordenáveis e em alguns casos são representados como gráficos de linhas [Sil10]. Apesar de o objetivo deste trabalho não ser adicionar visualizações para dados temporais ao protótipo SOLAP+, a apresentação destes tópicos deve-se a possibilidade de o modelo proposto poder suportar tipos de visualização para dados temporais.

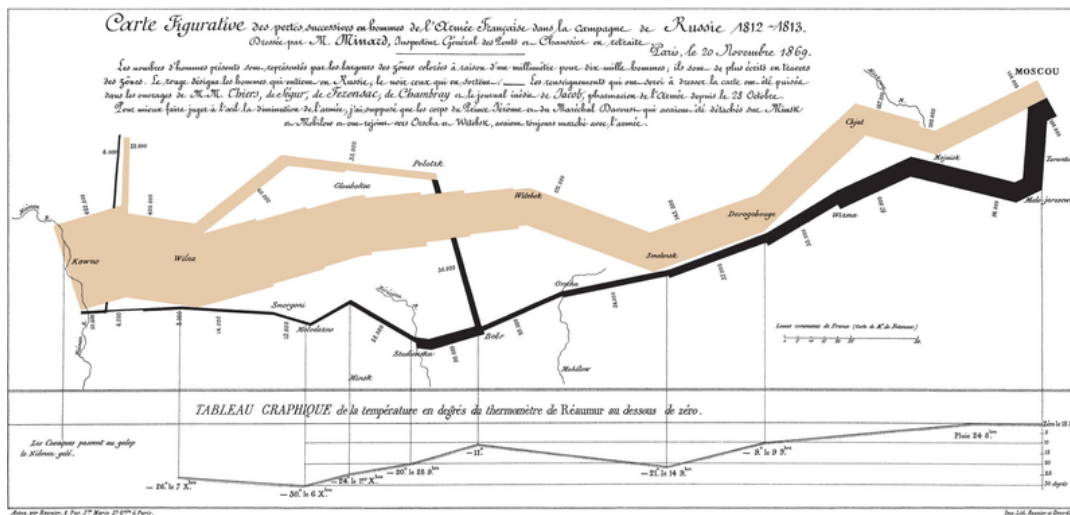


Figura 2.5: Invasão da Rússia pelas tropas de Napoleão

2.2.2 Cartografia num Contexto SOLAP

O uso de mapas para representar dados, como acontece no SOLAP, é algo usado na cartografia há dezenas de anos. O mapa da Figura 2.5 foi criado por Charles Minard em 1869, que representa a tentativa de invasão da Rússia por Napoleão. No mapa estão presentes várias variáveis como o número de tropas, as rotas que foram tomadas por estes, os rios que atravessaram e o calendário de eventos. De notar que é considerado um dos melhores mapas estatísticos alguma vez criados [Tuf83]. Apesar da qualidade do mapa apresentado, as técnicas usadas na execução deste não se aplicam facilmente a um sistema SOLAP. O mapa apresentado representa um fenómeno específico, durante um período específico e numa geografia particular. Dificilmente este mapa estatístico é generalizável a outros fenómenos e noutros espaços. No contexto SOLAP é essencial conseguir abordar a geração dos mapas temáticos de forma genérica o suficiente para que a aplicabilidade do SOLAP seja tão vasta como a do OLAP.

Os dados resultantes das interrogações realizadas numa aplicação SOLAP podem ser representados no mapa como objetos espaciais, usando as suas variáveis visuais, ou outros elementos tais como gráficos ou *chorems* [DCDFL⁺11]. Uma variável visual deve expressar a lógica do significado dos dados mapeados. De modo a conceber mapas para que possam ser "vistos" e não "lidos" é necessário utilizar metodologias que cumpram as regras cartográficas [KO03].

Bertin [Ber67] identificou as seguintes variáveis visuais: tamanho, saturação, cor, forma e textura. Estas variáveis visuais são aplicáveis consoante o objeto espacial em causa (ponto, linha ou polígono). Cada variável visual traduz um ou mais significados, como mostra a Tabela 2.1. Por exemplo, Bertin considera que a variável *tamanho* tende a traduzir uma ordenação ou quantidade, mas o mesmo não se verifica para a *cor*, que está associada a significados seletivos (ex.: género). A lista de variáveis visuais apresentada em cima tem vindo a ser estendida com a introdução de novas variáveis como a

Níveis de Organização		
Seletivo	Ordenado	Quantitativo
Forma Cor	Saturação Tamanho	Tamanho
Textura		

Tabela 2.1: Níveis de Organização das Variáveis Visuais

orientação [KO03].

Na cartografia clássica a denominação dos tipos de mapas está associada aos objetos representados e quais as variáveis visuais utilizadas [Fre92]. Por exemplo *proportional symbol maps* são mapas que representam pontos em que os valores associados usam o seu tamanho ou a saturação, ver Figura 2.6. Outros tipos de mapas incluem *choropleth maps* (polígonos variando a cor ou/e saturação), *symbol maps* (ponto variando a forma) entre outros.

Os atributos presentes numa interrogação (à exceção de atributos espaciais) podem ser categorizados em diferentes classes [ASW11]. A principal distinção é: dados quantitativos e qualitativos. Quanto aos primeiros, os dados podem ser escalas de razão, escalas de intervalos (por exemplo a temperatura) ou valores numéricos. Um bom exemplo deste tipo de dados são as métricas numéricas presentes nas tabelas de facto. Ao contrário dos dados quantitativos, os qualitativos apenas podem tomar valores pertencentes a um conjunto finito de valores. Este tipo de dados pode ainda ser subdividido em duas categorias: (i) nominal onde não existe uma ordem implícita (ex.: género); (ii) ordinal em que existe uma ordem implícita (ex.: escala de Mercalli).



Figura 2.6: Exemplo de um Mapa de Símbolos Proporcionalis

Objetos Espaciais	Tipo de Dados	Tamanho	Saturação	Cor	Forma	Textura
Pontos	Nominal	Não	Não	Sim	Sim	Sim
	Ordinal	Sim	Sim	Sim(*)	Não	Não
	Quantitativo	Sim	Sim	Sim(*)	Não	Não
Linhas	Nominal	Não	Não	Sim	Sim	Sim
	Ordinal	Sim	Sim	Sim(*)	Não	Não
	Quantitativo	Sim	Sim	Sim(*)	Não	Não
Polígonos	Nominal	Não	Não	Sim	Não	Sim
	Ordinal	Não	Sim	Sim(*)	Não	Não
	Quantitativo	Não	Sim	Sim(*)	Não	Não

Tabela 2.2: Mapeamento Adequado das Variáveis Visuais para Tipos de Dados

Com base nas características perceptivas das variáveis visuais e no tipo de dados, em [Pas04] foi proposto em que condições uma variável visual era apropriada, ou não, para mapear os respectivos dados. Na Tabela 2.2 é possível identificar para cada objeto espacial quais as variáveis visuais que este pode utilizar, e para um tipo de dados. Quando um valor desta tabela é dado como adequado mas esteja marcado com um * a sua utilização deve ser de forma cuidada (ex.: gradiente de cores). Adicionalmente, Bertin afirma também que apenas se devem utilizar até três variáveis visuais em simultâneo [Ber67].

Ao representar informação numérica num mapa temático, esta é em geral agrupada por classes. Considere que se pretende apresentar a população portuguesa por distrito, utilizando a saturação. É necessário criar um conjunto de classes de valores para agrupar de modo a atribuir um valor de saturação a cada intervalo. A escolha quer do número de classes quer da forma como estas são definidas [KO03] tem um papel importante na construção da visualização cartográfica. A definição do número de classes não deve ser totalmente livre e está dependente não só do tipo do objeto espacial (ponto, linha e polígono) como também das variáveis visuais que se pretende utilizar [KO03]. Por exemplo, se for utilizada a cor (em polígonos) para mapear um atributo do tipo nominal, oito classes é o máximo recomendado para que seja possível identificar todas elas de forma quase imediata [KO03].

A forma como são definidas as classes não tem influência na capacidade do utilizador ter ou não uma percepção imediata do mapa. Influencia sim, a interpretação que o utilizador tem dos dados. Para o tipo de dados quantitativos ou ordinais as classes podem ser definidas com intervalos iguais, através de quebras naturais, entre outras formas [KO03]. Já para dados nominais deverá existir uma classe para cada valor distinto do atributo que está a ser mapeado.

Até então foram discutidas formas de representação de informação quando apenas está envolvido um objeto espacial. No entanto, em [Sil10] foi proposto casos de interação onde estão dois objetos espaciais (multi-geometria). No contexto deste documento entende-se por multi-geometria quando numa dada análise existem dois atributos espaciais de hierarquias diferentes. Como foi referido na Secção 2.1.1 este problema foi identificado na primeira iteração do prototipo *SOLAP+*. Para resolver este problema foi

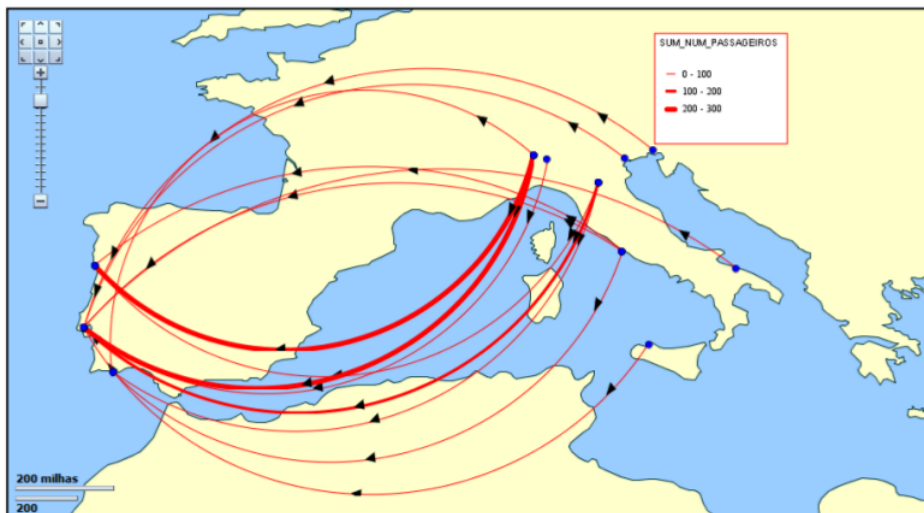


Figura 2.7: Relações entre objetos Espaciais

proposto em [Sil10] uma abordagem que usa arcos para definir as relações entre os objetos como está representado na Figura 2.7.

Cada relação é representada por um arco, que contém as variáveis visuais presentes nas linhas, e pelos respectivos objetos espaciais envolvidos. Portanto para cada relação serão usados três objetos, o arco e os objetos envolvidos. Ao tratar um arco da mesma maneira que uma linha é possível associar gráficos ou outras representações ao arco.

Para além de poder usar o arco para visualizar dados é também possível usar os objetos envolvidos na relação para representar dados. Atualmente existem seis possíveis relações [Sil10], estas são identificadas pelos objetos envolvidos: *ponto-ponto*; *ponto-linha*; *ponto-polígono*; *linha-linha*; *linha-polígono*; *polígono-polígono*.

2.2.3 Heurísticas de Visualização de Informação

As heurísticas têm sido bastante usadas para a avaliação da qualidade e usabilidade de interfaces [NM94]. Com base nesta ideia têm sido propostos conjuntos de heurísticas para avaliar métodos de visualização de informação. Para a definição dos conjuntos de heurísticas foram realizados alguns estudos [FJ10] [ZSN⁺06]. Estes estudos usam como base conjuntos de recomendações/heurísticas de vários autores e tentam criar um conjunto usando as que melhor conseguem explicar os vários problemas analisados.

Nos dois estudos acima referenciados existem três conjuntos de heurísticas que são mencionados em ambos: (i) Zuk e Carpendale "*Selection of perceptual and cognitive heuristics*"[ZC06]; (ii) Shneiderman "*Visual Information-Seeking Mantra*"[Shn96]; (iii) Amar e Stasko "*Knowledge and task-based framework*"[AS04]. Em seguida serão apresentados estes três conjuntos de heurísticas.

O conjunto de heurísticas definido em (i) foi criado com base no trabalho de Bertin, Tufte e Ware. Cada um destes autores deu contributos na sua área. Bertin com os níveis de organização das variáveis visuais [Ber67], vistas anteriormente na Secção 2.2.2;

Tufte com o contributo para o desenho de gráficos estatísticos [Tuf83]; e Ware com o trabalho desenvolvido no livro *Information Visualization: Perception for Design* [War04]. Este conjunto é composto por treze heurísticas, que fornecem recomendações para o uso correto de cores nas visualizações, melhorar a percepção da informação obtida e garantir a qualidade dessa informação.

No que respeita ao uso da cor as heurísticas indicam que esta não deve mostrar dados de natureza ordenável (*Don't expect a reading order from color*), tal como foi mostrado na Figura 2.1. Para além disto também deve-se ter em conta o daltonismo para a escolha de cores (*Consider people with color blindness*). Por fim no que respeita às cores ainda é chamada a atenção para dois pormenores: (i) a percepção da cor pode variar consoante o tamanho dos objetos (*Color perception varies with size of colored item*); (ii) quando existem várias *layers* temáticas sobrepostas no mapa deve-se ter em atenção as cores de cada uma de modo a que estas não sejam confundidas (*Local contrast affects color & gray perception*). Para além das *layers* temáticas as de referência também podem influenciar a facilidade da visualização.

No que respeita à qualidade da informação mostrada, isto é sem induzir erros, são indicadas algumas heurísticas de como se devem usar as variáveis visuais (ver Figura 2.1): estas devem ter domínio suficiente para mostrar todos os valores de modo a não omitir informação (*Ensure visual variable has sufficient length*); para além do tamanho do domínio também se deve ter em conta os valores a mapear para cada variável (*Quantitative assessment requires position or size variation*), tal como foi discutido na Secção 2.2.2. Outras recomendações neste domínio referem-se ao uso de texto para ajudar a identificar informação sempre que seja necessário (*Integrate text wherever relevant*) e deve-se preservar a dimensionalidade dos dados para a visualização (*Preserve data to graphic dimensionality*). Isto significa que a visualização não deve "distorcer" os dados, por exemplo quando se representa dados num gráfico de barras o tamanho das barras deve ser proporcional ao valor dos dados.

Por fim neste conjunto de heurísticas existem várias recomendações de como melhorar a qualidade percetiva da visualização de dados. Nestas encontram-se referências às leis de *Gestalt* [Kof55] usadas na teoria da percepção, à forma como deve ser apresentada a informação (*Provide multiple levels of detail*) e (*Preattentive benefits increase with field of view*). Neste dois casos as recomendações são para as visualizações consigam apresentar a informação a diferentes níveis e as anomalias são mais fáceis de encontrar com o aumento do campo de visualização. Por fim as últimas heurísticas dizem respeito ao desenho, defendendo que deve-se usar o menor espaço possível para criar a visualização e que se devem eliminar-se todos os elementos gráficos que não contribuam para a visualização, "chartjunk", (*Put the most data in the least space*) e (*Remove the extraneous ink*).

Estes elementos externos decorativos, como mostrado na Figura 2.8, são considerados por alguns autores [Tuf83] como "chartjunk". Recentemente têm aparecido trabalhos

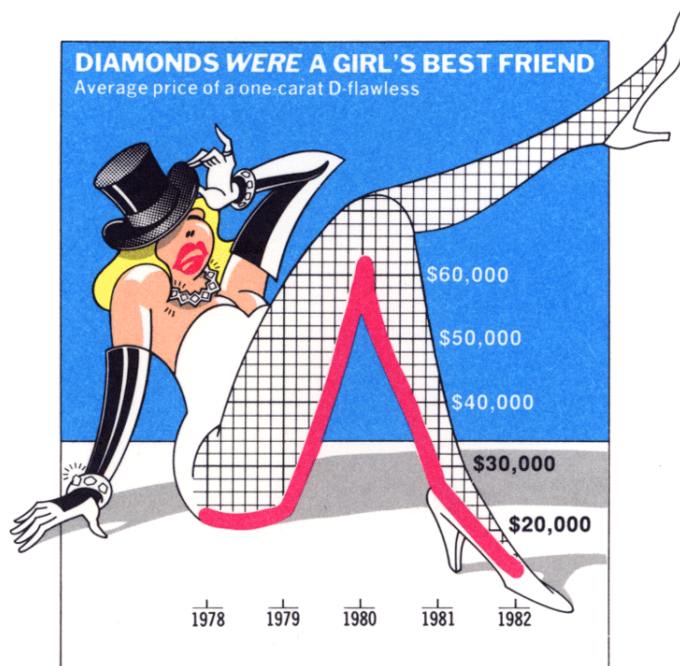


Figura 2.8: Gráfico contendo "Chartjunk" de Nigel Holmes usado em [BMG⁺10]

[BMG⁺10] que defendem a utilização destes elementos gráficos para melhorar certos aspectos como a memória. Nesta linha de trabalho foram desenvolvidas um conjunto de recomendações [HAS11] de modo a melhorar a memorização das visualizações.

As heurísticas referentes à *Visual Information-Seeking Mantra* tem como objetivo melhorar a forma como se analisam os dados, enquanto que o conjunto de heurísticas anteriores está mais relacionado com a percepção de informação. Este conjunto contém um total de seis heurísticas: (i) *Overview first*; (ii) *Zoom and filter*; (iii) *Details on demand*; (iv) *Relate*; (v) *Extract*; (vi) *History*. Como foi referido, este conjunto de heurísticas foi criado a partir da forma como se analisam os dados. A primeira parte da análise é conseguir obter uma visão geral do problema que está a ser analisado (i), depois de obter a visão geral usa-se o *zoom* e *filter* para restringir o domínio da procura (ii). Após restringir o domínio deve ser possível obter mais informação sobre os dados atuais (iii). Os utilizadores também devem ter facilidade de voltar atrás na análise (iv). Por fim o utilizador deve ter a capacidade de conseguir relacionar os dados ao longo da análise (v) e ser capaz de extrair informação destes (vi).

Por fim no conjunto "*Knowledge and task-based framework*", as heurísticas presentes têm como objetivo garantir propriedades que facilitam a interpretação e o processo cognitivo do analista. Este conjunto também é constituído por seis heurísticas:

Expose uncertainty Esta heurística defende que sempre que haja alguma incerteza nos dados esta deve poder ser visível, de modo a que o analista tenha conhecimento da

sua existência.

Concretize relationships Esta heurística é válida se for possível relacionar a informação visualizada de modo a conseguir inferir relações .

Determination of Domain Parameters Nesta heurística é defendido que a visualização deve conseguir mostrar os dados no contexto do seu domínio., isto é, para cada conjunto de valores deve-se ter em conta as suas unidades (quando existem) e qual é a relação entre os dados a serem mostrados e o seu domínio.

Multivariate Explanation Esta heurística defende que sempre que melhorem a capacidade de análise devem ser usados mecanismos para transformação de dados como por exemplo passar os valores para escalas logarítmicas.

Formulate cause & effect Nesta heurística é defendido que deve ser possível identificar anomalias nos dados, para se conseguir detetar a causa deles.

Confirm Hypotheses Por fim nesta heurística é defendido que se deve ajudar o analista quando este tenta confirmar a sua hipótese usando para tal outras medidas de comparação como por exemplo a distribuição normal.

Este conjuntos poderão vir a ter dois usos na neste trabalho: (i) estas heurísticas estarão presentes no modelo proposto de forma a que a construção da visualização cartográfica seja o mais correta possível (ver Secção 3); (ii) serem usadas para a avaliar a qualidade das visualizações criadas pelo sistema (ver Secção 5).

2.3 Sistemas de Visualização Assistida

Uma vez que o objetivo deste trabalho passa por construir um sistema de visualização cartográfica assistida, nesta secção serão discutidos alguns sistemas que implementam meios de assistir o utilizador na escolha de visualizações.

O sistema SOLAP+, apresentado na Secção 2.1.1, inclui um mecanismo para escolha de visualizações [Sil10]. A seleção da visualização a ser aplicada é obtida através de uma árvore de decisão. A árvore recebe como parâmetros o tipo de objeto espacial a ser representado, o número de métricas, o número de colunas numéricas e semânticas. O resultado é a visualização a ser representada que pode usar as variáveis visuais apresentadas na Tabela 2.2, ou gráficos como o gráfico de setores ou de barras.

Outro sistema, denominado *Show Me* também usa um método similar para a escolha de visualização [MHS07]. Neste sistema também é usada uma árvore de decisão para escolher qual a visualização que se pode usar. Este sistema foi implementado na aplicação *Tableau*¹, sendo o conjunto de visualizações suportadas tanto por mapas como outras mais tradicionais, gráficos ou tabelas.

¹www.tableausoftware.com

Por fim, o último sistema que vai se apresentado é o *Common GIS* descrito em [AAV03]. Neste sistema também é utilizada uma árvore de decisão para obter as visualizações a serem usadas. Neste caso todas as representações visuais são feitas usando mapas temáticos. Ao contrário dos dois sistemas anteriores a escolha da visualização cabe ao utilizador sendo que o sistema só indica quais as que são validas.

De notar que estes sistemas apresentam duas falhas que o sistema proposto neste trabalho quer colmatar. A primeira está na criação da visualização, enquanto que os sistemas apresentados têm só em conta a estrutura dos dados para a escolha da visualização, isto tende a não chegar. Em muitos casos existem, para um tipo de visualização, diversas maneiras de a criar. Estes sistemas tendem a ignorar isto usando para cada visualização um conjunto de parâmetros por defeito.

A segunda falha está relacionada com a extensibilidade do sistema. Num estudo recente [ZSB⁺12] uma das maiores limitações encontradas foi a falta de extensibilidade dos métodos de visualização. Em todos os sistemas apresentados não é possível estender as formas de visualização. O modelo proposto neste trabalho também tenta resolver este problema.

2.4 Conclusão

Neste capítulo primeiro foram apresentados conceitos relativos aos sistemas *SOLAP* sendo que foi dado mais destaque ao *SOLAP+* uma vez que é nele que este trabalho vai ser realizado.

Após a apresentação de conceitos relacionados com o *SOLAP* e do sistema *SOLAP+*, foram discutidas técnicas para a visualização de informação com ou sem componente espacial.

Ainda nesta secção foram apresentados três conjuntos de heurísticas, cada um criado para um objetivo diferente: (i) qualidade da visualização; (ii) interatividade com o sistema; (iii) facilitar a interpretação dos dados. O principal uso destas heurísticas, para o modelo de visualização cartográfica assistida, é garantir a qualidade das visualizações sugeridas. Muitas partes do modelo foram criadas tendo em mente algumas destas heurísticas.

Por fim na última secção foram apresentados alguns sistemas que usam visualização de dados assistida. Nesta secção foram comparadas as funcionalidades destes com o sistema proposto neste trabalho.

3

Visualização Cartográfica Assistida

A visualização cartográfica é indispensável num ambiente *SOLAP*, uma vez que os resultados das interrogações realizadas são apresentados principalmente através de mapas. Com o objetivo de garantir um papel efetivo destes, é proposto um sistema para assistir o utilizador na escolha e construção das visualizações cartográficas adequadas. Este capítulo começa por dar uma visão global sobre a abordagem seguida. São descritos em pormenor os principais parâmetros do sistema proposto nas secções 3.1 e 3.2. Por fim, são apresentados na Secção 3.3 as componentes responsáveis por assistir o utilizador na construção de visualizações cartográficas.

Num ambiente *SOLAP* existem diversos fatores que influenciam a representação e apresentação adequada dos resultados. Aqueles que foram identificados como relevantes para o sistema proposto são: (i) a meta-informação existente sobre os modelos multidimensionais (ii) a interrogação realizada; (iii) os dados resultantes da interrogação. Todas estas fontes de informação daqui em diante passarão a ser designadas de *Contexto SOLAP*.

Para além do *Contexto SOLAP*, o sistema proposto, Figura 3.1, recebe também como parâmetro de entrada uma paleta de tipos de visualização. O conceito tipo de visualização e todas as definições que o suportam serão apresentados na Secção 3.2.

Com base no *Contexto SOLAP* e na paleta de tipos de visualização a tarefa de construção de visualizações cartográficas é realizada em duas fases: (i) identificação dos tipos aplicáveis; (ii) aplicação de visualização. A primeira fase identifica quais os tipos de visualizações cartográficas (presentes na paleta de tipos de visualizações cartográficas) que são adequados para um dado *Contexto SOLAP*. Posteriormente, na segunda fase, é aplicada um tipo de visualização aos dados presentes no *Contexto SOLAP* usando algumas das recomendações discutidas na Secção 2.2.3.

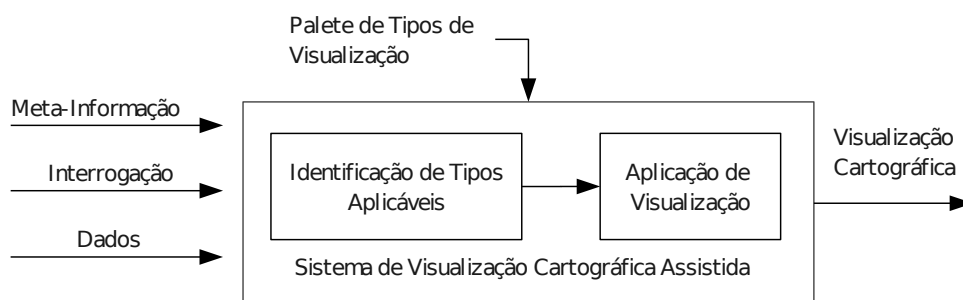


Figura 3.1: Sistema de Visualização Cartográfica Assistida

3.1 Contexto SOLAP

Um dos parâmetros de entrada do sistema de visualização cartográfica assistida é denominado de *Contexto SOLAP*. Este é composto por três elementos: (i) a meta informação associada ao modelo multidimensional; (ii) a interrogação realizada; (iii) os dados obtidos.

A meta-informação sobre o modelo multidimensional inclui informação sobre as dimensões que existem. Uma dimensão é composta por um conjunto de níveis e um conjunto de hierarquias. Uma hierarquia tem uma sequência de níveis. Um nível tem um conjunto de atributos associados e os respectivos tipos de dados. Nesta dissertação, os tipos de dados considerados podem ser espaciais (pontos, linhas ou áreas) ou semânticos (numérico, alfanumérico ou temporais). Um nível diz-se espacial se tiver atributos de tipo espacial. Uma dimensão diz-se espacial se tiver níveis espaciais. Adicionalmente, o modelo contém também informação acerca das métricas numéricas presentes nas tabelas de factos e os operadores de agregação aplicáveis sobre as métricas. De notar que nesta dissertação não se consideram dados espaciais nas tabelas de facto, este assunto tem vindo a ser discutido em vários trabalhos [Bim10] [MZ07] mas não existe consenso em como deve ser tratado.

Na Figura 3.2 está representado um modelo multidimensional, muito simplificado, para um *Data Warehouse* sobre acidentes rodoviários nos Estados Unidos. Neste modelo estão representadas duas dimensões (*Tempo*, *Local*) ambas com uma hierarquia. A dimensão *Tempo* é composta pelos níveis dia, mês e ano. A dimensão *Local* contém dois níveis, local e estado, ambos espaciais. Por fim neste modelo também está representada uma tabela de facto com as métricas *Número de Mortes* e *Número de Pessoas Envolvidas*. Este modelo vai ser usado mais a frente para apresentar alguns exemplos ilustrativos.

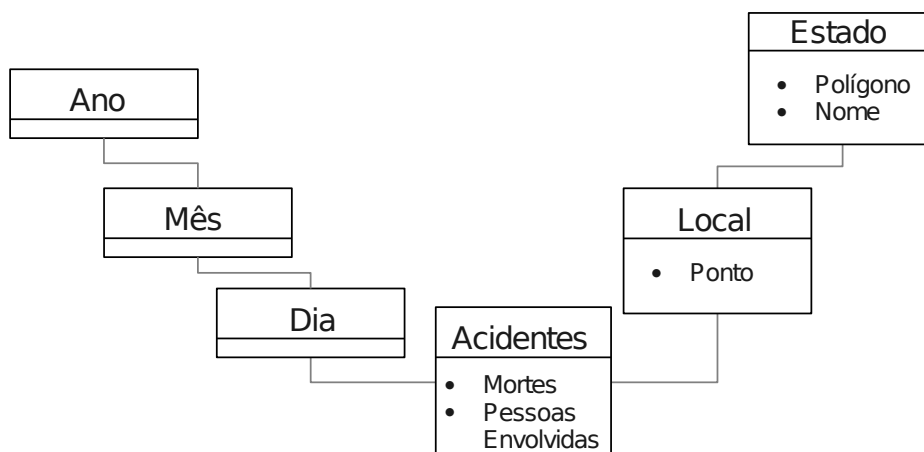


Figura 3.2: Modelo Multidimensional de Acidentes Rodoviários nos Estados Unidos

As interrogações *SOLAP* são realizadas de modo interativo indicando os elementos do modelo multidimensional a serem usados e que operações executar sobre eles. A interrogação realizada irá dar origem a uma procura na base de dados que retornará os dados atualmente a serem analisados. Uma interrogação *SOLAP* indica quais os níveis, atributos e métricas alvo da análise do utilizador. Para além destes elementos, contém a descrição sobre os operadores de agregação das métricas e potenciais filtros sobre os atributos e/ou métricas. Uma vez que estamos num ambiente *SOLAP* onde a visualização é feita através de mapas, nesta dissertação considera-se também que numa interrogação está incluída informação sobre o nível *zoom* e a *Minimum Bounding Box* atual (área de interesse).

Usando o modelo apresentado na Figura 3.2 considere as interrogações da Figura 3.3. Na primeira interrogação seleciona-se o nível espacial *Estado* e as duas métricas da tabela de facto. Numa usa-se o operador de agregação soma e noutra a média. Ao adicionar o nível *Ano* do nível *Tempo* obtemos a segunda interrogação.

$$\{\textit{Estado}, \text{SUM}(\textit{Mortes}), \text{AVG}(\textit{Pessoas Envolvidas})\}$$

$$\{\textit{Estado}, \textit{Ano}, \text{SUM}(\textit{Mortes}), \text{AVG}(\textit{Pessoas Envolvidas})\}$$

Figura 3.3: Interrogações *SOLAP*

O último parâmetro de entrada relativamente ao *Contexto SOLAP* é constituído pelos dados resultantes da interrogação *SOLAP* bem como estatísticas sobre estes. Cada coluna tem associado um tipo de dados, este pode ser obtido cruzando a informação da cláusula *select* na interrogação e os respetivos elementos do modelo multidimensional, sendo que os tipos de dados podem ser: espaciais, reais, nominais e tempo. Quando uma coluna tem como tipo de dados um valor real as estatísticas associados são distribuição de valores, média, mediana, desvio padrão e quantis. Caso o tipo de dados seja nominal então as estatísticas associadas são o número de valores diferentes, a distribuição de valores e a

moda.

Ao executar as interrogações da Figura 3.3 obtemos como parte do resultado as Tabelas 3.1 e 3.2. Para efeitos de representação do resultado da interrogação sob a forma de uma tabela assume-se que existe um atributo textual para cada nível espacial. Esse atributo pode desempenhar o papel de identificador ou legenda de cada objeto espacial no mapa. Nestas interrogações só está a ser usado um nível espacial mas pode acontecer interrogações em que se usem dois níveis espaciais, caso isso aconteça usamos a metodologia proposta em [Sil10]. Quando isto acontece significa um de três casos: (i) dois níveis de dimensões diferentes; (ii) dois níveis da mesma dimensão de hierarquias diferentes; (iii) dois níveis da mesma dimensão e da mesma hierarquia. No primeiro caso será criado um arco para representar o objeto como foi proposto em [Sil10]. Para o segundo usa-se a intersecção dos dois objetos. Por fim no último caso usa-se o nível mais baixo na hierarquia. Concluindo para efeitos de representação será usado apenas um valor espacial para cada linha de resultado.

<i>Estado</i>	SUM(Mortes)	AVG(Pessoas Envolvidas)
Alabama	9097	2.41
Arizona	7829	2.95
Arkansas	5694	2.42
California	33873	2.77
Colorado	5212	2.64
Connecticut	2607	2.33

Tabela 3.1: Resultado parcial da Interrogação {*Estado*, **SUM(Mortes)**, **AVG(Pessoas Envolvidas)**}

<i>Estado</i>	<i>Ano</i>	SUM(Mortes)	AVG(Pessoas Envolvidas)
Alabama	2009	841	2.28
Alabama	2010	858	2.28
Arizona	2009	755	2.77
Arizona	2010	651	2.56
Arkansas	2009	592	2.38
Arkansas	2010	562	2.25
California	2009	3075	2.70
California	2010	2679	2.36
Colorado	2009	465	2.58
Colorado	2010	448	2.33
Connecticut	2009	223	2.16
Connecticut	2010	319	2.02

Tabela 3.2: Resultado parcial da Interrogação {*Estado*, *Ano*, **SUM(Mortes)**, **AVG(Pessoas Envolvidas)**}

Para a discussão da representação no mapa dos dados obtidos numa interrogação

SOLAP, importa discutir o relacionamento entre o nível espacial que participa na interrogação e os outros níveis (da mesma dimensão ou de outras dimensões) bem como o relacionamento do nível espacial com as métricas envolvidas na interrogação.

Pela natureza do modelo multidimensional qualquer nível de uma dimensão tem, em geral, um relacionamento $1:N$ para cada uma das métricas e conseqüentemente o relacionamento entre o nível espacial e qualquer uma das métricas, envolvidas no interrogação, é $1:N$. Quanto ao relacionamento de um nível espacial (E) com um nível semântico (A) é necessário distinguir os seguintes casos: (i) E e A são de dimensões distintas e nesse caso, pela própria definição do modelo multidimensional existe, em geral, uma relação $1:N$ entre E e A; (ii) E e A são da mesma dimensão e E é de nível igual ou inferior a A então $1:1$; (iii) E é de nível superior a A então a relação será de $1:N$; (iv) por fim caso E seja incomparável a A então a relação é de $1:N$.

O mesmo relacionamento observado no modelo multidimensional pode também ser observado nos dados obtidos de uma interrogação *SOLAP*. Nos dados obtidos de uma interrogação, a relação entre o nível espacial (coluna espacial) e qualquer outra coluna é de $1:1$ se para cada um dos valores diferentes da coluna espacial existir apenas um valor diferente na referida coluna. Se para algum valor espacial existir mais do que um valor diferente na referida coluna então a relação é de $1:N$.

Quando no modelo multidimensional a relação observada é de $1:1$ então a relação observada nos dados é necessariamente de $1:1$. Quando a relação segundo o modelo multidimensional é de $1:N$ então a observada nos dados pode ser $1:N$ ou $1:1$. Sendo que as relações segundo o modelo podem induzir algumas relações "menos corretas", para efeitos de visualização assistida será tida em conta apenas as relações observadas nos dados.

Ao analisar os resultados das interrogações realizadas, Figura 3.3, conseguimos identificar a relação entre a coluna espacial (no caso dos exemplos a coluna *Estado*) e a outras colunas. Para a primeira interrogação, Tabela 3.1, a relação observada das duas colunas não espaciais com a espacial é de $1:1$. Para a segunda interrogação as relações observadas nos dados, Tabela 3.2, são de $1:N$ para todas as colunas.

3.2 Visualização Cartográfica

No *SOLAP* a visualização de dados é feita preferencialmente sobre mapas. Para tal é preciso mapear os dados obtidos numa interrogação *SOLAP* para um mapa (visualização cartográfica). Nesta dissertação o conceito de visualização cartográfica é entendido como um conjunto de representações gráficas (composta por variáveis visuais), sendo que estas estão associadas a um objeto espacial (ponto, linha, polígono, ou composto resultante da relação entre objetos espaciais), no local definido pelas suas coordenadas.

Para obter o efeito acima descrito temos de particionar os dados, obtidos numa interrogação, por cada objeto espacial. Para cada partição é criada uma representação gráfica

que usa como parâmetros os valores das colunas não espaciais associadas ao objeto espacial.

Usando os dados apresentados na Tabela 3.1, para este exemplo cada objeto espacial está presente apenas numa única linha. Deste modo a partição será feita linha a linha. Por exemplo para o estado do Alabama ficam associados os valores: $\{\text{SUM}(\text{Mortes}) : 9097, \text{AVG}(\text{Pessoas Envolvidas}) : 2.41\}$. Neste caso o nível espacial tem uma relação de 1:1 com cada uma das outras duas colunas, que neste caso são métricas.

Para a segunda interrogação cujos dados são apresentados na Tabela 3.2 existe mais do que uma linha por objeto espacial. Neste exemplo os dados obtidos associados ao Alabama são os seguintes: $\{\text{Ano} : [2009, 2010], \text{SUM}(\text{Mortes}) : [841, 858], \text{AVG}(\text{Pessoas Envolvidas}) : [2.28, 2.28]\}$. Neste caso a relação existente entre a coluna espacial e a outras colunas é de 1:N.

A forma como se mapeiam os dados obtidos nas interrogações SOLAP para as respectivas representações gráficas é usando tipos de visualização. Estes vão ser descritos em pormenor na Secção 3.2.2. Mas antes de explicar os tipos de visualização é necessário explicar como foram tratados vários problemas relacionados com a forma como se criam representações gráficas.

3.2.1 Variáveis Visuais

Nesta dissertação, o significado atribuído ao termo *variável visual* difere da interpretação clássica dada às variáveis de Bertin [Ber67]. Enquanto que na definição Bertin interpreta-se que uma variável visual é um conceito como cor ou saturação, neste trabalho uma variável visual é uma propriedade visual de uma característica gráfica associada a uma representação gráfica. Na definição proposta podemos usar a mesma propriedade visual em varias partes da representação. Usando como base os *smileys* da Figura 3.4 com a definição de variável visual proposta podemos identificar como variáveis visuais por exemplo a cor do *smiley* e a cor do olhos do *smiley*.



Figura 3.4: Smileys de Várias Cores

Para além da repetição de conceitos mostradas acima, a definição proposta consegue identificar variáveis visuais que não foram identificadas por Bertin, ver Tabela 2.1. Por exemplo considerando um gráfico de setores, com a definição proposta podemos identificar duas variáveis visuais: (i) as etiquetas do gráfico; (ii) os valores referentes a cada sector.

Para perceber a diferença entre as duas noções de variáveis visuais, com a definição de Bertin quando falamos da cor estamos a referir a cor de toda a representação gráfica ou se estamos a falar de uma parte da representação então não podemos usar a cor como variável visual noutra parte da representação. Daqui em diante nesta dissertação as variáveis visuais definidas por Bertin passam a serem denominadas variáveis visuais clássicas.

Bertin além de identificar várias variáveis visuais também as classifica consoante o seu significado percetivo, ver Tabela 2.1. São identificados três tipos de variáveis visuais: (i) seletivas; (ii) ordenáveis; (iii) quantitativas. Ao estender o significado de variável visual de Bertin também foi sentida a necessidade de estender os tipos de variáveis visuais.

Nesta dissertação é proposta uma nova classificação para as variáveis visuais, com base na classificação de Bertin sendo composta por quatro classes: (i) seletivo; (ii) ordenável; (iii) conjunto seletivo; (iv) conjunto ordenável. A classe seletiva e ordenável tem os mesmos significados dados por Bertin. No caso dos conjuntos (seletivo e ordenável) classificam variáveis visuais capazes de traduzir diversos significados seletivos ou ordenáveis numa só variável visual. Por exemplo a cor do *smiley* é uma variável visual seletiva, já no caso do tamanho do *smiley* esse será ordenável. Já o caso de um gráfico de setores ambas as variáveis visuais serão conjuntos, as etiquetas serão um conjunto seletivo enquanto que os valores serão um conjunto ordenável.

Nas classes propostas nota-se a ausência da classe quantitativa proposta por Bertin. Esta classe foi excluída de modo a simplificar o modelo sendo que a informação perdida não é significativa. A classe quantitativa é uma sub-classe da ordenável logo qualquer variável visual quantitativa também será ordenável. Para além de ser uma sub-classe de outra na generalidade o seu uso tende a ser complicado. O objetivo desta classe é transmitir razões, isto é se $A = 2 * B$ então essa relação tem de ser mantida na representação. Muitas vezes estas diferenças não se conseguem interpretar sendo que a interpretação mais comum é $A > B$.

Com a discussão das propriedades e comportamento das variáveis visuais em seguida serão apresentados os tipos. Na definição proposta para os tipos de visualização as variáveis visuais serão as em cima proposta.

3.2.2 Tipos de Visualização

O segundo parâmetro do sistema de visualização cartográfica assistida consiste numa paleta de tipos de visualização. Um tipo de visualização τ corresponde fundamentalmente a uma representação gráfica composta por variáveis visuais associadas a um ou mais tipos de dados aceitáveis para controlar essas variáveis visuais. A notação para definir um tipo de visualização está apresentada na Figura 3.5.

$$\tau = (\text{Nome}, \text{Objeto Espacial}, \text{Fábrica}, [(\text{Nome Variável Visual}, \text{Classe da Variável Visual}, \{\text{Tipo}_1, \dots, \text{Tipo}_n\}, [\text{Cardinalidade}]), \dots])$$

Figura 3.5: Modelo de Tipo de Visualização

Para definir um tipo de visualização é necessário especificar: (i) o nome do tipo de visualização e (ii) o tipo do objeto espacial (exigido para que se possa instanciá-lo) (iii) a fábrica onde está suportado o tipo de visualização (iv) e, as variáveis visuais pretendidas para mapear os dados. Para cada variável visual definida é especificado: (i) a classe da variável visual; (ii) o tipo de dados aceitáveis; (iii) e a cardinalidade desta.

O tipo do objeto espacial é usado para identificar qual o objeto a ser desenhado na representação. A fábrica indica qual a interface que é exposta ao tipo de visualização. Esta interface é constituída por um conjunto de variáveis visuais e que gera uma representação gráfica (imagem), usada na construção de mapas. As fábricas serão discutidas em pormenor na Secção 3.3.2.

O outro elemento do tipo de visualização é uma lista de variáveis visuais. Para cada variável visual desta lista é indicado o seu nome (este nome deverá estar declarado na respetiva fábrica), a classe percetiva proposta na Secção anterior, os tipos de dados aceitáveis (reais, nominais ou tempo) e a cardinalidade (será explicado mais a frente).

$$\tau = (\text{Smiley}, \text{Ponto}, \text{Smiley}, [(\text{Cor}, \text{Seletivo}, \{\text{Nominal}\}, [1]), (\text{Cor dos Olhos}, \text{Seletivo}, \{\text{Nominal}\}, [1])])$$

Figura 3.6: Tipo de Visualização: *Smileys*

Por exemplo se quisermos criar um tipo de visualização que use como representação gráfica os *smileys*, expressão da Figura 3.6, indicamos que o objeto espacial será um ponto e que a fábrica será a fábrica de *smileys*, considerando que a representação gráfica é a da Figura 3.4. Neste tipo de visualização estamos a utilizar duas variáveis visuais, a cor do *smiley* e a cor dos olhos. Para ambas a classe é seletiva, o tipo de dados é nominal e tem cardinalidade 1.

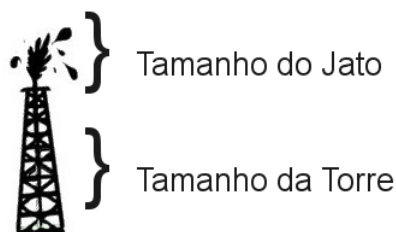


Figura 3.7: Torre de Extração de Petróleo

Por exemplo na Figura 3.7, a torre de extração é vista como uma representação gráfica. Com base na definição de tipo de visualização proposta é possível definir um tipo de

visualização em que tenha uma variável visual que varia o tamanho da torre, e uma outra que varia o tamanho do jato no topo. Neste tipo de visualização, Figura 3.8, as diferentes variáveis visuais estão a utilizar a mesma característica gráfica (ex: tamanho).

$$\tau = (\text{Torre de Extração}, \text{Ponto}, \text{Torre}, [(\text{Tamanho da Torre}, \text{Ordenável}, \{\text{Real}\}, [1]), (\text{Tamanho do Jato}, \text{Ordenável}, \{\text{Real}\}, [1])])$$

Figura 3.8: Tipo de Visualização: Ponto Tamanho

No mapa da Figura 3.9, a informação é representada usando gráficos de setores (que corresponde ao tipo de visualização da Figura 3.10), nos quais são usadas duas variáveis visuais: (i) o valor associado a cada setor; (ii) e as respetivas cores, ambas conjuntos.

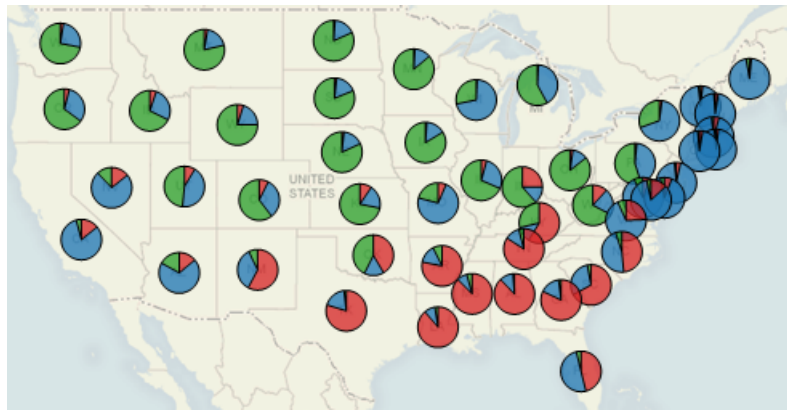


Figura 3.9: Mapa com Gráficos de Setores

No tipo de visualização mostrado na Figura 3.10 referente ao mapa da Figura 3.9 as duas variáveis visuais são conjuntos. Isto acontece porque num gráfico de setores conseguimos representar vários valores em simultâneo. No caso das etiquetas do gráfico serão mapeados para a variável visual cores, enquanto que os valores respetivos são a variável visual valores. Uma vez que os valores das etiquetas são textuais a classe será um conjunto seletivo, no caso dos valores estes são reais logo o conjunto é ordenável.

$$\tau = (\text{Gráfico}, \text{Polígono}, \text{Gráfico de Setores}, [(\text{Cores}, \text{Conjunto Seletivo}, \{\text{Nominal}\}, [1]), (\text{Valores}, \text{Conjunto Ordenável}, \{\text{Real}\}, [1])])$$

Figura 3.10: Tipo de Visualização: Gráfico de Setores

Por fim, as restrições sobre cardinalidade dos dados indica o número de valores possíveis de mapear numa dada variável visual, isto é séries. Se o valor da cardinalidade for fixo este é um número (neste caso não existem séries), caso contrário é representado por um intervalo na forma $n0..n1$. Este intervalo é fechado nas duas extremidades. No tipo de visualização 3.10 ambas as cardinalidades têm o valor 1. Se em vez de um gráfico de setores fosse usado como representação gráfica um gráfico de barras (tipo de visualização 3.11) então para a cardinalidade da variável visual que representa os valores de yy

o valor passava a ser representado pelo intervalo de um a cinco. Este intervalo deve-se a possibilidade de o gráfico de barras desta representação gráfica permitir entre uma e cinco séries para os valores de yy .

$$\tau = (\text{Gráfico}, \text{Ponto}, \text{Gráfico de Barras}, [(Eixo\ xx, \text{Conjunto Seletivo}, \{Nominal\}, [1]), \\ (Eixo\ yy, \text{Conjunto Ordenável}, \{Real\}, [1..5])])$$

Figura 3.11: Tipo de Visualização: Gráfico de Barras

A extensibilidade alcançada através da definição de tipo de visualização proposta é bastante importante, uma vez que se pode utilizar a mesma característica visual sobre mais do que uma componente da representação gráfica, os conjuntos permitem a integração de gráficos e outras representações complexas e por fim a própria existência do tipo de visualização é um mecanismo de abstração. No Anexo A está representada parte da paleta de tipos de visualização (a paleta contém mais de quarenta tipos) criada no âmbito desta dissertação, de notar que essa lista não é de forma alguma exaustiva.

O modelo proposto para descrever os tipos de visualização permite saber em que contextos *SOLAP* podem ser usados e como eles se comportam, mas não tem indicações de como instanciar os tipos de visualização para criar a representação gráfica. Para conseguir descrever como desenhar as representações gráficas são propostas as fábricas.

3.3 Sistema de Visualização Cartográfica Assistida

O sistema de visualização cartográfica assistida opera com base num *Contexto SOLAP* e num conjunto disponível de tipos de visualizações (os quais se assume terem sido especificados por perito em cartografia). Por um lado, o *Contexto SOLAP* contém informação sobre as análises realizadas por utilizadores (e respetivos dados), e por outro, os tipos de visualização contém informação sobre as representações gráficas e as respetivas variáveis visuais que podem ser utilizadas para mapear informação.

Deste modo, desde a análise realizada por um utilizador até ao retorno de uma visualização cartográfica, o modelo realiza duas tarefas. Em primeiro lugar é necessário identificar os tipos de visualizações disponíveis que sejam aplicáveis ao *Contexto SOLAP* em questão. Uma vez realizada essa tarefa aplica-se um tipo de visualização identificado anteriormente. As secções subsequentes detalham estas duas tarefas.

3.3.1 Tipos de Visualização Aplicáveis

De modo a identificar os tipos de visualização aplicáveis, do conjunto disponível, é necessário que se verifiquem algumas condições. Primeiro é necessário verificar se o tipo do objeto espacial declarado no tipo de visualização corresponde ao atributo espacial contido no *Contexto SOLAP*. Se tal não se verificar o tipo de visualização não é aplicável.

Posteriormente avalia-se se é possível associar cada variável visual do tipo de visualização a uma coluna não espacial do conjunto de dados (criar associações), sem se repetir

nenhuma variável visual a não ser quando a cardinalidade for diferente de um. Quando as variáveis tem cardinalidades maiores de um, estas podem ser um número válido para o intervalo declarado. Por exemplo para um cardinalidade [1..3] a variável visual pode ser repetida uma, duas ou três vezes. É possível que numa associação o número de variáveis visuais seja menor ou igual ao número de colunas a serem representadas.

Por fim temos que verificar se uma associação declarada é válida, para tal verificamos se o tipo de dados da coluna nos dados está nos tipos de dados aceitáveis para este tipo de visualização. Caso esteja então por fim verificamos se as classes de variável visual conseguem ser respeitadas pelos dados. Para conseguir tal efeito comparamos a classe da variável visual com o relação das colunas com a espacial. Quando uma variável visual é da classe *seletivo* ou *ordenável* então a relação verificada nas colunas tem de ser de 1:1. Caso a classe seja um dos conjuntos a relação tem de ser de 1:N. Se esta propriedade não se verificar então esta associação não é válida.

Para um tipo de visualização e um conjunto de dados pode haver várias associações possíveis, se existir uma ou mais válidas então o tipo de visualização passa a ser declarado como aplicável. Caso contrário o tipo não é aplicável e é descartado.

Considere como parâmetros de entrada do sistema proposto a paleta descrita no Anexo A e o *Contexto SOLAP* contendo a interrogação da Figura 3.12. Nesta interrogação as fábricas são representadas por um ponto. Nos dados obtidos, ver Tabela 3.3, pode-se verificar que as duas colunas não espaciais tem um relacionamento de 1:N com a coluna espacial. Por fim o tipo de dados da coluna *Poluente* tem como tipo de dados nominal e a soma das emissões é um real.

{*Fábrica*, *Poluente*, **SUM**(*Emissões*)}

Figura 3.12: Exemplo Tipos Aplicáveis: Interrogação

<i>Fábrica</i>	<i>Poluente</i>	SUM (<i>Emissões</i>)
Agro-Pec. das Barreiras, S.A.	PM10	39163.69
Agro-Pec. do Alto da P. - S. Bartolomeu dos Galegos	N2O	0
Agro-Pec. do Alto da P. - Alto da Palhoça	HFC	0
Agro-Pec. Valinho, S.A.	As	73582.8
Agro-Pec. Valinho, S.A.	Cd	21844.12
Alfa Sul - Alumínios do Sul, Lda	CO2	280386.2242

Tabela 3.3: Resultado parcial da Interrogação {*Fábrica*, *Poluente*, **SUM**(*Emissões*)}

	Nome	Fábrica	Variáveis Visuais			
			<i>a</i>	<i>b</i>	<i>c</i>	<i>d</i>
A1	Cor	Ponto	Cor	[1]	Seletiva	Nominal
A2	Tamanho	Ponto	Tamanho	[1]	Ordenável	Real
A3	Saturação	Ponto	Saturação	[1]	Ordenável	Real
A4	Cor e Tamanho	Ponto	Cor Tamanho	[1] [1]	Seletiva Ordenável	Nominal Real
A5	Cor e Saturação	Ponto	Cor Saturação	[1] [1]	Seletiva Ordenável	Nominal Real
A6	Tamanho e Saturação	Ponto	Saturação Tamanho	[1] [1]	Ordenável Ordenável	Real Real
A7	Gráfico de Setores	Gráfico de Setores	Etiquetas Valores	[1] [1]	Conj. Seletivo Conj. Ordenável	Nominal Real
A8	Gráfico de Barras	Gráfico de Barras	Etiquetas Valores	[1] [1..5]	Conj. Seletivo Conj. Ordenável	Nominal Real
A9	Gráfico de Linhas	Gráfico de Linhas	Etiquetas Valores	[1] [1..5]	Conj. Ordenável Conj. Ordenável	Real Real

Tabela 3.4: Fragmento da Paleta de Tipos de Visualização

^aVariável Visual^bCardinalidade^cClasse^dTipo de Dados

Usando o processo descrito anteriormente o primeiro passo é filtrar os tipos de visualização na paleta consoante o tipo de objeto espacial contido no *Contexto SOLAP*. Neste caso vamos aceitar apenas tipos de visualização cujo objeto espacial seja um ponto. Os tipos de visualização que resultam nos que têm identificadores entre A.1 e A.9. Estes tipos de visualização estão representados na Tabela 3.4.

O segundo passo é criar as associações entre as variáveis visuais e as colunas. Ao verificar os tipos de dados descartamos os tipos A.6 e A.9 uma vez que nas duas colunas tem como tipo de dados reais enquanto que no dados uma coluna é nominal e outra é real. Por fim verificamos se as classes das variáveis visuais estão de acordo com relação entre as colunas não espaciais e a espacial. Uma vez que as duas colunas a serem representadas detêm uma relação de 1:N com a espacial isto obriga a que as variáveis visuais sejam da classe *Conjunto Seletivo* ou *Conjunto Ordenável*. Os únicos tipos de visualização que conseguem respeitar esta condição são o A.7 e A.8 onde as variáveis visuais são conjuntos. Estes dois tipos de visualização são declarados como aplicáveis sendo os outros descartados.

Com a explicação da identificação dos tipos de visualização aplicáveis o próximo passo é a aplicação de um tipo aplicável. Para conseguir isso é necessário criar um mapeamento usando as fábricas de visualização descritas na próxima secção. Por fim pode acontecer casos onde o número de tipos de visualização aplicáveis seja diferente de um. Se não existirem tipos aplicáveis então não existe forma de representar os dados do atual *Contexto SOLAP*. Caso haja mais do que um tipo aplicável, exemplo apresentado em cima, é preciso criar uma forma de desempate, esta será explicada na Secção 3.3.4.

3.3.2 Fábricas Visuais

Uma fábrica representa uma interface para um serviço capaz de criar representações gráficas. Para definir uma fábrica visual, expressão da Figura 3.13, temos de indicar o nome da fábrica, a localização para o serviço (*URI*) e uma lista de variáveis visuais. Para cada variável visual é indicado o seu nome, o domínio de valores possíveis e o valor por defeito. Os valores por defeito são indicados apenas se for necessário esta decisão está dependente da fábrica e da representação gráfica.

$$\vartheta = (\text{Nome}, \text{URI}, [(\text{Nome da Variável Visual}, \text{Valor por Defeito}, \{\text{Valor}_1, \dots, \text{Valor}_n\}), \dots])$$

Figura 3.13: Modelo Para Fábricas Visuais

O domínio de uma variável visual é uma lista finita de valores aceitáveis para o uso da respetiva variável visual. Por exemplo a variável visual cor cujo domínio é uma lista de cores, pode ter como domínio: *Azul*, *Vermelho*, *Amarelo* e *Verde*. Mas se os valores do domínio forem um intervalo numérico também devem ser especificados, por exemplo a variável visual saturação (usando como modelo de cor *HSV*): *20*, *40*, *60*, *80* e *100*. De notar caso os valores do domínio sejam ordenáveis então devem ser declarados de forma ordenada. O valor por defeito de variável visual deve ser um dos valores contidos nessa lista.

A separação entre as fábricas visuais e os tipos de visualização permite que sobre uma fábrica existam vários tipos de visualização. Outra vantagem da separação de conceitos é que permite modificar as fábricas sem ter que modificar os respetivos tipos de visualização, desde que a interface seja mantida.

Nesta dissertação foram criadas sete fábricas visuais: (i) Ponto; (ii) Linha; (iii) Polígono; (iv) Gráfico de Linhas; (v) Gráfico de Setores; (vi) Gráfico de Barras; (vii) *Smileys*. Na Tabela 3.5 estão apresentadas as variáveis visuais contidas na fábrica *Ponto*. Nas variáveis visuais *Cor*, *Saturação* e *Tamanho* os domínios são descritos em valores reais. Isto acontece porque a cor e saturação são definidas em função do *HSV*. Uma vez que esta três variáveis visuais são reais, sujeitas a ordenação, logo a ordem de definição do domínio é importante de manter. A variável visual *Forma* tem como domínio dois valores textuais que representam os dois tipos de forma do ponto.

Cor	0	{0, 60, 120, 180, 240, 300}
Saturação	100	{20, 40, 60, 80, 100}
Tamanho	3	{1, 2, 3, 4, 5}
Forma	Círculo	{Círculo, Estrela}

Tabela 3.5: Variáveis Visuais da Fábrica Ponto

Na Tabela 3.6 estão representadas as variáveis visuais da fábrica gráfico de setores. Para um gráfico de setores foram consideradas duas variáveis visuais, as etiquetas e os os valores a mostrar. No caso das etiquetas estas têm associadas cores, como tal no domínio

desta variável visual está declarado as cores aceites pela fábrica. Já no caso dos valores não existe domínio porque os valores a serem enviados para a fábrica serão os dados obtidos no *Contexto SOLAP*.

Etiquetas	{Vermelho, Azul, Verde, Amarelo, Laranja, Magenta}
Valores	{}

Tabela 3.6: Variáveis Visuais da Fábrica Gráfico de Setores

Usando como base a fábrica visual de pontos, Tabela 3.5, podemos ter vários tipos de visualização, sendo que alguns usam apenas um sub-conjunto das variáveis visuais disponíveis na fábrica. Na Figura 3.14 estão representados dois tipos de visualização sobre pontos. No tipo τ_1 são usadas duas variáveis visuais da fábrica respetiva, enquanto que no tipo τ_2 é apenas usada uma. A possibilidade de escolher apenas algumas das variáveis visuais de uma fábrica para um tipo de visualização permite um aproveitamento muito maior da capacidade da representação gráfica usada.

Apesar de as outras variáveis visuais não serem usadas quando for criado o mapeamento, ver Secção 3.3.3, será necessário associar valores a elas. Por exemplo para o mapeamento do τ_1 também é necessário associar as variáveis visuais *Tamanho* e *Saturação* com os respetivos valores por defeito.

$$\begin{aligned}\tau_1 &= (Forma\ e\ Cor, Ponto, Ponto, [(Forma, Seletivo, Nominal, [1], \\ &\quad (Cor, Seletivo, Nominal, [1])]) \\ \tau_2 &= (Tamanho, Ponto, Ponto, [(Tamanho, Seletivo, Nominal, [1])])\end{aligned}$$

Figura 3.14: Tipos de Visualização sobre a Fábrica Ponto

Para cada variável visual de um tipo de visualização podemos aumentar a sua informação com a informação na fábrica respetiva. Deste modo podemos associar os domínios e os valores por defeito a cada variável visual. Este aumento de informação irá ser necessário no próximo passo, que é criar um mapeamento usando um tipo de visualização e um *Contexto SOLAP*.

3.3.3 Mapeamento

Um mapeamento corresponde a mapear os valores presentes em cada coluna dos dados do *Contexto SOLAP* para uma respetiva variável visual de um tipo aplicável. No final é criado um pedido à fábrica correspondente do tipo de visualização mapeado, contendo o mapeamento criado e a uma forma reduzida dos dados.

O mapeamento entre uma coluna e uma variável visual pode ser de um cinco tipos. Quatro destes cinco tipos são referentes as classes de variáveis visuais apresentadas na Secção 3.2.1 (Seletiva, Ordenável, Conjunto Seletivo e Conjunto Ordenável). A quinta está relacionada com os valores por defeito. Para os valores por defeito o mapeamento dessa variável visual consiste em associar-lhe o valor por defeito.

Para um mapeamento de uma variável visual cuja classe seja seletiva é necessário atribuir cada valor diferente na coluna de dados a mapear um valor diferente do domínio. Caso o número de valores diferentes na coluna a mapear seja superior ao tamanho do domínio então o mapeamento não é válido. Se um mapeamento for inválido então o tipo de visualização não é possível de ser aplicado.

$$\begin{aligned} \text{Dados da Coluna} &: [A_1, A_2, A_3, A_4] \\ \text{Domínio da Variável Visual} &: [B_1, B_2, B_3, B_4] \\ \text{Mapeamento Resultante} &: \{A_1 : B_1, A_2 : B_2, A_3 : B_3, A_4 : B_4\} \end{aligned}$$

Figura 3.15: Exemplo de Mapeamento Seletivo

Na Figura 3.15 está exemplificado um mapeamento seletivo para cada valor único da coluna a mapear é associado um valor do domínio. No final resulta um lista de pares em que a chave é referente aos dados da coluna e os valores pertence ao domínio da variável visual.

No mapeamento ordenável os valores da coluna a serem mapeados são divididos em intervalos sendo associado a cada intervalo um valor do domínio. Para tal é preciso identificar duas coisas o algoritmo para calcular os intervalos e número de intervalos.

Neste trabalho apenas foi tratado o problema da escolha do algoritmo, não sendo abordado o segundo problema. Foram escolhidos três algoritmos para calcular os intervalos: (i) intervalos uniformes; (ii) normal; (iii) quebras naturais. Nos intervalos uniformes os intervalos são calculados usando como degrau o subtração do valor mínimo ao máximo dividindo pelo número de intervalos. Os intervalos calculados usando como base a normal usam como degrau desvio padrão dos dados, sendo este somado e subtraído a média dos dados. Para este algoritmo os intervalos são sempre adicionados em números pares. Por fim as quebras naturais foram apresentadas na Secção 2.2.2. No que diz respeito ao número de intervalos foi estipulado que este poderá variar entre três e oito. Atualmente a escolha dos intervalos é feita de forma manual pelo utilizador.

Para a escolha do algoritmo de criação de intervalos foi proposto, que se o número de valores diferentes for menor que sete então usa-se os intervalos uniformes com o número de valores diferentes como parâmetro. Caso contrário criamos distribuições (normal, exponencial, uniforme) usando como base os dados a representar. Depois comparamos cada distribuição com o conjunto de dados usando um *F Test*. Se a distribuição normal for a mais próxima então o algoritmo usado será os intervalos normais. Se for a distribuição uniforme a mais então será a os intervalos uniformes, por fim se for a exponencial será as quebras naturais.

$$\begin{aligned} \text{Dados da Coluna} &: [n_1, n_2, n_3, n_4, n_5, n_6] \\ \text{Domínio da Variável Visual} &: [B_1, B_2, B_3, B_4] \\ \text{Mapeamento Resultante} &: \{[n_1 - n_3[: B_1, [n_3 - n_5[: B_2, [n_5 - n_6] : B_3\} \end{aligned}$$

Figura 3.16: Exemplo de Mapeamento Ordenável

Na Figura 3.16 está demonstrado um mapeamento ordenável, usando os valores da coluna são criados três intervalos $[n_1 - n_3]$, $[n_3 - n_5]$, $[n_5 - n_6]$ sendo associado a cada um valor do domínio.

Para o mapeamento dos conjuntos (ordenável e seletivo) caso exista domínio é preciso criar mapeamentos. Estes mapeamentos serão iguais ao seletivo ou ordenável dependendo se o conjunto a ser representado é um conjunto seletivo ou um conjunto ordenável.

Usando para exemplo do mapeamento os dados da Tabela 3.3 e o tipo de visualização A.9 da Tabela 3.4, a informação contida na fábrica respectiva está mostrada na Figura 3.6.

Para este tipo de visualização vai ser necessário criar um mapeamento do tipo conjunto seletivo para a variável visual *Etiquetas* uma vez que esta tem um domínio declarado. Já para a variável visual valores não é necessário uma vez que esta não tem domínio declarado. Na Figura 3.17 está demonstrado o mapeamento resultante para a respectiva variável visual.

$$\{PM10 : Vermelho, N2O : Azul, HFC : Verde, \\ As : Amarelo, Cd : Laranja, CO2 : Magenta\}$$

Figura 3.17: Mapeamento para a Variável Visual *Etiquetas*

Num pedido a uma fábrica para além do mapeamento também é necessário adicionar os dados. Os dados são enviados de forma agrupada por objeto espacial, isto quando existe mais do que uma linha com o mesmo objeto espacial os valores nas colunas não espaciais são agrupados em lista. Na Tabela 3.7 está a transformação resultante dos dados usados no exemplo anterior.

<i>Fábrica</i>	<i>Poluente</i>	SUM(Emissões)
Agro-Pec. das Barreiras, S.A.	[PM10]	[39163.69]
Agro-Pec. do Alto da P. - S. Bartolomeu dos Galegos	[N2O]	[0]
Agro-Pec. do Alto da P. - Alto da Palhoça	[HFC]	[0]
Agro-Pec. Valinho, S.A.	[As, Cd]	[73582.8, 21844.12]
Alfa Sul - Alumínios do Sul, Lda	[CO2]	[280386.2242]

Tabela 3.7: Transformação da Tabela 3.3

Por fim a resposta a um pedido vem sobre a forma de um conjunto de representações gráficas, uma para cada objeto espacial, e a legenda para a respectiva visualização.

3.3.4 Preferências

Pode acontecer que na identificação dos tipos aplicáveis resulte mais do que um tipo aplicável, caso isso aconteça é preciso criar um fator de desempate. Para tal efeito é proposta nesta dissertação um modelo de desempate baseado em preferências.

Uma preferência é um valor entre zero e um associado às variáveis visuais de um tipo

de visualização. Cada variável visual tem associado os parâmetros relativos ao mapeamento, se for seletivo o número de valores diferentes a mapear, ordenável o número de intervalos. Por exemplo para o tipo de visualização *Forma e Cor*, Figura 3.14, podemos ter como preferências a Tabela 3.8.

Forma	Cor	Preferência
1	3	0.7
1	5	0.9
2	5	0.9

Tabela 3.8: Exemplo do uso de Preferências

Cada tipo de visualização contém um conjunto de preferências deste modo é possível comparar entre vários tipos de visualização. Caso não haja valor de preferência para as variáveis visuais de um tipo de visualização então esse valor é considerado zero. Se no final ainda existir mais do que um tipo de visualização aplicável então é escolhido um de forma aleatória.

4

Arquitetura e Implementação

Neste capítulo vai ser explicado como foi implementado o sistema de cartografia assistida sobre o protótipo *SOLAP+*, apresentado na Secção 2.1.1, e que modificações foi preciso fazer a este para poder ser implementado. De notar que o protótipo *SOLAP+* foi grande parte refeito, tendo inclusive sido necessário modificar a sua arquitetura para facilmente permitir novas extensões, nomeadamente o sistema de visualização cartográfica assistida proposto nesta dissertação.

4.1 Arquitetura

O *SOLAP+* tem como objetivo ser uma implementação de *SOLAP* genérica, independente dos dados, interativa e que implemente várias funções comuns em *SOLAP*. No *SOLAP+* a visualização dos dados é feita usando mapas temáticos, tabelas ou gráficos.

A arquitetura resultante da implementação do protótipo desenvolvido em [Sil10] é a apresentada na Figura 4.1. O cliente é a componente que expõe a interface gráfica ao utilizador tanto para apresentar resultados, como para permitir que o utilizador explicito o que pretende. Esta componente realiza um pedido ao servidor *SOLAP+*, o servidor vai criar a interrogação *SQL* correspondente ao pedido, usando a meta-informação disponível. Após a geração da interrogação esta é executada na base de dados sendo os resultados processados para serem apresentados ao utilizador. Para além dos dados a componente cliente também recebe um pedido *SQL* com a interrogação espacial. Esta interrogação espacial é depois redirecionada pelo cliente para o servidor *Oracle MapViewer*. A resposta deste servidor é o mapa com a informação a ser mostrada. A indicação do tipo de mapa é passada na interrogação sendo este de uma lista definida a priori no

servidor. Apesar da lista poder ser adaptada as formas de visualização disponíveis estão limitadas às implementadas no próprio servidor. A arquitetura apresentada levanta vários problemas nomeadamente: (i) estar dependente da base de dados *Oracle* uma vez que o servidor *MapView* assim o obriga; (ii) esta estranha interação em que o servidor tem que interagir com o *MapView* de forma indireta através da base de dados e através do cliente pelo envio da interrogação que este tem que realizar posteriormente sobre o *MapView*; (iii) a forte dependência de toda a solução relativamente ao *MapView* para a geração dos mapas temáticos.

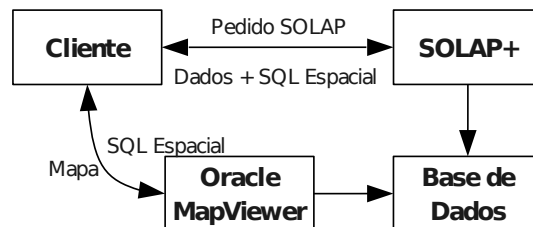


Figura 4.1: Arquitetura Original do SOLAP+

Deste modo a arquitetura adotada no *SOLAP+* levanta vários problemas para a implementação de um modelo de cartografia assistida nesta arquitetura. O primeiro está relacionado com a falta de capacidade de definir formas de visualização para além das presentes no *MapView*. A segunda está relacionada com as instruções para a construção do mapa temático, na visualização cartográfica assistida existe a necessidade de gerar estas instruções de forma dinâmica. No *MapView* estas instruções têm de estar declaradas num ficheiro de configuração.

Para tentar resolver os problemas apresentados, principalmente os relacionados com a falta de extensibilidade das formas de visualização, foi criada uma nova arquitetura para o servidor *SOLAP+*, apresentada na Figura 4.2. Nesta arquitetura um cliente faz um pedido ao servidor *SOLAP+*, este gera a interrogação *SQL* correspondente e executa. O resultado pode passar por fases de pós-processamento sendo depois enviado para uma módulo de visualização. Neste módulo é produzida a visualização que será enviada como resposta ao cliente.

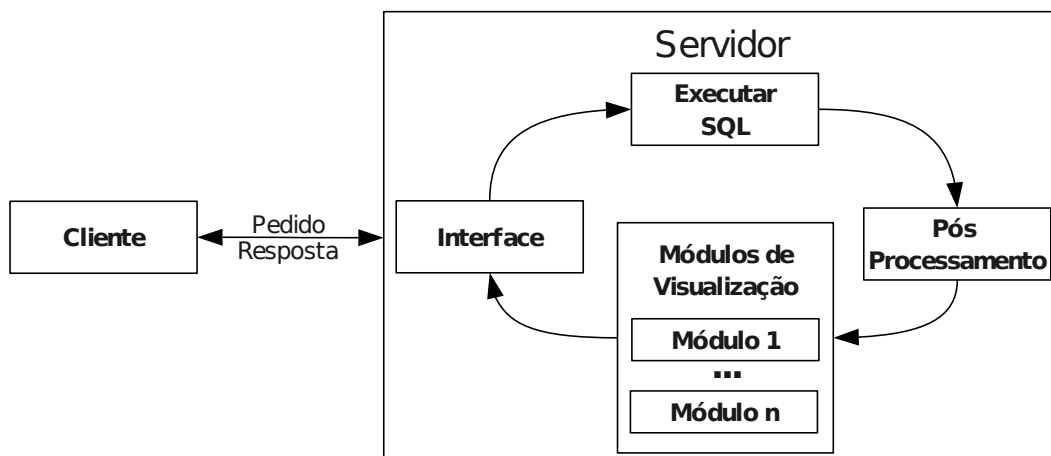


Figura 4.2: Arquitetura Atual do Servidor SOLAP+

Uma das preocupações desta nova arquitetura foi a modularidade, de modo a permitir adicionar novas funcionalidades ao servidor de forma simples e extensível. A primeira forma de modularidade inserida foi a independência da base de dados, uma vez que ao retirarmos o *MapView* podemos utilizar diferentes base de dados sem limitações. A segunda está relacionada com a possibilidade de criar unidades de pós-processamento. Nestas unidades podem realizar-se várias ações sobre os dados, como *data-mining*. Por fim e a mais relevante para este trabalho é a possibilidade de criar módulos de visualização.

Na arquitetura criada os módulos de visualização permitem que seja possível com que os clientes peçam os resultados em diferentes formas de visualização. Por exemplo é possível ter um módulo em que as visualizações sejam baseadas em gráficos e ter um módulo em que as visualizações produzidas são mapas temáticos. Para implementar a visualização cartográfica assistida foi criado um módulo para tal efeito, este módulo vai ser explicado na próxima Secção.

Nesta nova arquitetura foi criado um novo protocolo de comunicação com o servidor. Neste protocolo o pedido é criado indicando quais as métricas, níveis e slices a realizar. Para além disto também é passada informação sobre as unidades de pós-processamento a usar. Na resposta vem um *geojson* com a informação sobre os objetos espaciais a representar, um arquivo contendo a visualização (representações gráficas e legenda). Para além disto também pode vir um *rowset* com os respetivos dados. Tanto para o pedido como para a resposta o formato usada é *json*.

4.2 Implementação

O modelo de cartografia assistida proposto na Secção 3 foi implementado no protótipo criando um módulo de visualização para o efeito. Primeiro vai ser explicado como foi implementada a parte relativa aos tipos de visualização, sendo depois explicado a implementação de todo o processo de seleção e aplicação destes.

O modelo que representa os tipos de visualização e as fábricas foram implementados usando como base os modelos descrito na Secção 3.2.2. Foram criadas duas *XML Schemas* para representar cada um dos modelos. Na Figura 4.3 estão apresentados dois pequenos exemplos de como são os *XML* usados para declarar os tipos de visualização e as fábricas respetivamente.

```
1 <visualization spatialObject="Point" name="Point_Color" factory="point">
2   <variable name="color" cardinality="1" visualClass="Selective">
3     <type>Text</type>
4   </variable>
5 </visualization>
6
7 <factory factoryUrl="@intern" name="point">
8   <variable name="color">
9     <domainValue name="000" />
10    (...)
11    <domainValue name="330" />
12  </variable>
13 </factory>
```

Figura 4.3: Exemplo da Sintaxe para Definir Fábricas e Tipos de Visualização

Na criação de tipos de visualização, linhas 1 a 5, declaramos o objeto espacial, a fábrica e o respetivo nome. Para cada variável visual do tipo de visualização têm como atributos o nome, a cardinalidade e a classe visual. Na cardinalidade os valores podem ser num formato numérico ou no formato $n0..n1$ sendo $n0$ e $n1$ números. Para adicionar tipos de dados a uma variável visual criamos elementos filhos com os respetivos tipos de dados.

Na criação de fábricas declaramos a *URL* em que podemos encontrar o serviço que cria a representação gráfica e o seu nome. Apesar de no modelo a localização ser uma *URI* na implementação usa-se uma *URL* porque os serviços para as fábricas vão estar sobre a forma de serviços *REST*. As fábricas contêm ainda a possibilidade de definir elementos, com as mesmas propriedades que foram declaradas no modelo.

Depois de definir os tipo de visualização e as fábricas, o próximo passo é implementar o modelo de cartografia assistida. Na implementação deste foram usados todos os passos descritos na Secção 3.3.

Na seleção de tipos aplicáveis começa-se por carregar a paleta de tipos de visualização e criar estatísticas sobre os dados obtidos. Depois são realizadas as fases descritas na Secção 3.3.1. Na aplicação junta-se a informação da fábricas às respetivas variáveis visuais dos tipos aplicáveis. De seguida, são criados os mapeamentos das variáveis visuais para cada tipo aplicável logo criando as instanciações para cada tipo de visualização. Sendo no final escolhida uma usando as preferências declaradas sobre os tipos de visualização.

A resposta a uma fábrica vem sobre a forma de um arquivo comprimido contendo as representação gráficas para cada objeto espacial e a legenda. Na implementação as

representações gráficas e a legenda obtidas de uma fábrica devem estar sobre a forma de um *bitmap*. De seguida vão ser explicadas em mais detalhe algumas das fábricas implementadas neste trabalho.

4.2.1 Fábricas Implementadas

Neste trabalho foram implementadas sete fábricas: três delas produzem os objetos espaciais básicos (ponto, linha e polígono); outras três foram criadas para produzir gráficos (setores, linhas, polígonos) como representações gráficas; por fim a última fábrica foi desenvolvida por Marta Lidon no âmbito do Programa de Introdução à Investigação Científica em Engenharia Informática. A representação gráfica usada nesta fábrica é *smiley*.

As fábricas sobre os objetos espaciais têm como variáveis visuais comuns a cor e a saturação, ambas representadas usando o modelo de cores *HSV*. Para a fábrica ponto também contém as variáveis visuais tamanho e forma. Na fábrica linha também se usa como variável visual o tamanho e a textura. A variável visual textura indica se a linha é representada usando um traço sólido ou tracejada.

As fábricas onde as representações gráficas são gráficos foram construídas usando a biblioteca *matplotlib*¹. Para o gráfico de setores foram consideradas duas variáveis visuais, a etiquetas e os valores. Para os gráficos de linhas e barras, as variáveis visuais são as etiquetas e os valores. Para a variável visual valor podem existir até cinco séries.

Por fim a fábrica que usa como representação os *smileys* tem como variáveis visuais a cor do olhos, a cor do *smiley*, o tamanho deste e o sorriso (triste, contente, indiferente). Na Figura 4.4 está demonstrado um mapa onde são usados *smileys* como representação gráfica, neste mapa estão a ser usado como variável visual o tamanho do *smiley* e o sorriso.

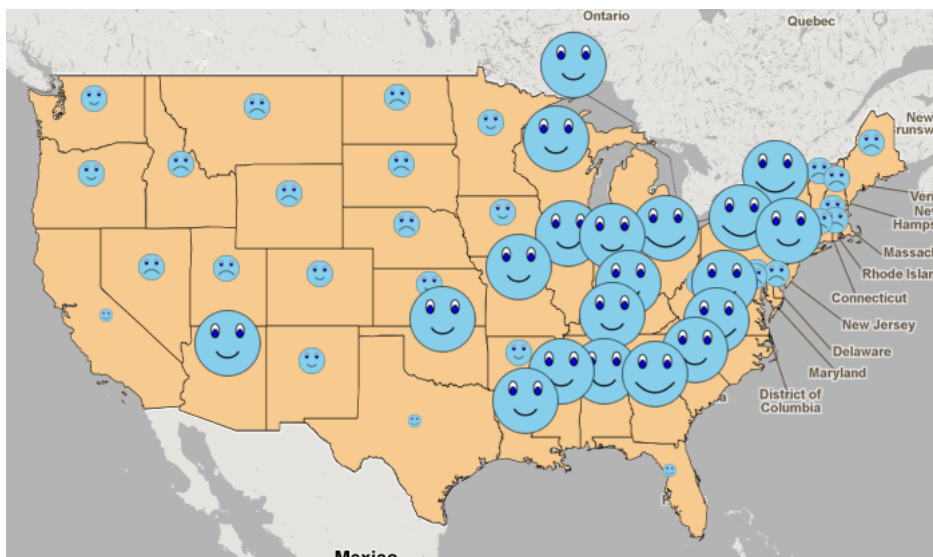


Figura 4.4: Mapa usando *Smileys* como Representação Gráfica

¹<http://matplotlib.org/>

4.2.2 Limitações Tecnológicas

Durante a implementação da *Visualização Cartográfica Assistida* foram encontradas algumas limitações nas tecnologias usadas, nomeadamente na *API* para os mapas e na biblioteca de criação de gráficos.

Um dos problemas encontrados foi a criação de imagens contendo gráficos (setores, linhas e barras). Das bibliotecas disponíveis em *Java* para a criação de gráficos nenhuma se adaptava ao pretendido. Deste modo foi escolhido usar a biblioteca *matplotlib* escrita em *Python* para a criação de gráficos. Apesar de ser uma limitação ao plano inicial mostrou-se a extensibilidade do modelo proposto.

A limitação tecnológica problemática está relacionada com a *API* de mapas, neste caso o *Google Maps*. No modelo proposto para cada objeto espacial é criada uma imagem com a sua representação. Na implementação para as fábricas ponto, linha e polígono isto não acontece. Para cada uma destas fábricas as representações são criadas usando as primitivas nativas da *API*. Para além disto sempre que se associa uma representação a uma linha ou polígono, usa-se a *API* para desenhar esse objeto com um conjunto de valores por defeito.

Usar a *API* para desenhar linhas e polígonos deve-se a esta não suportar o uso de figuras para representar estes objetos espaciais, sendo apenas possível em pontos. Logo quando usamos a fábrica linha ou polígono usa-se a *API*, caso seja outra fábrica associa-se um ponto ao objeto espacial e coloca-se nesse ponto a representação gráfica respetiva.

5

Validação

Neste capítulo vão ser apresentados os testes realizados no âmbito desta dissertação bem como os resultados obtidos. Foram realizados dois conjuntos de testes, o primeiro relacionado com o desempenho da solução implementada bem como a sua taxa de retenção, isto será descrito na Secção 5.1. Para o segundo conjunto de testes foi criado um protocolo em que são usados questionários respondidos pelos utilizadores. Isto será apresentado na Secção 5.2.

5.1 Desempenho e Taxa de Retenção

Nesta secção serão apresentados resultados relativos a dois testes executados: (i) desempenho; (ii) taxa de retenção. O desempenho da solução implementada é importante uma vez que os sistemas *SOLAP* tendem a ser bastante interativos logo uma solução que degrade muito o desempenho não é desejável. A taxa de retenção é calculada dividindo o número de tipos de aplicáveis pelo número de tipos de visualização na paleta dado como parâmetro para a visualização cartográfica assistida.

Todos os testes foram realizados usando uma paleta contendo quarenta e um tipos de tipos de visualização, sobre cinco *contextos SOLAP*. Três destes em que os objetos espaciais são pontos e dois onde são polígonos. Para os contextos dos pontos os elementos a representar vão ser: (i) uma métrica numérica; (ii) uma métrica numérica e valor textual da mesma hierarquia que o nível espacial; (iii) duas métricas numéricas. Nos contextos onde os níveis espaciais são polígonos vão conter: (iv) uma métrica numérica; (v) uma métrica numérica e valor textual de uma hierarquia diferente do nível espacial. Nestes testes foram usados um conjunto de dados composto por registo de emissões poluentes em Portugal. Os pontos representam as fábricas onde foram registadas as emissões e os

polígonos as freguesias onde estão localizadas as fábricas.

Para cada contexto em análise foram realizadas cinco corridas, sendo apresentados os tempos obtidos e a respetiva média para cada contexto. Na Tabela 5.1 estão apresentados os resultados relativos á taxa de retenção, sendo apresentados o número total de tipos aplicáveis em quarenta e um presentes na paleta e a respetiva taxa de retenção.

	Pontos			Polígonos	
	<i>Contexto (i)</i>	<i>Contexto (ii)</i>	<i>Contexto (iii)</i>	<i>Contexto (iv)</i>	<i>Contexto (v)</i>
Tipos Aplicáveis	2	4	2	1	2
Taxa de Retenção	4.88 %	9.76 %	4.88 %	2.44 %	4.88 %

Tabela 5.1: Taxa de Retenção

Relativamente ao desempenho os tempos observados a executar a visualização cartográfica assistida não representam sequer 1% to tempo despendido. As maiores frações do tempo estão concentradas no acesso a base de dados e na geração do mapa, logo a introdução da visualização cartográfica assistida não representa problemas de desempenho.

5.2 Protocolo Experimental

Nesta Secção será apresentado o protocolo experimental com recurso a utilizadores, e os resultados obtidos deste. Para a realização deste protocolo foi criado um questionário onde são apresentadas análises ao utilizador e são pedidas conclusões sobre essas análises. Este protocolo tem como objetivo testar a visualização cartográfica assistida em três métricas: *(i)* velocidade de aquisição de informação; *(ii)* retenção de informação; *(iii)* exatidão da informação obtida.

A velocidade de aquisição da informação pode ser calculada usando como base o tempo despendido para responder a cada grupo de perguntas. A exatidão de informação pode ser calculada usando como base o número de resposta certas às perguntas. Por fim a retenção de informação pode ser calculada fazendo repetir perguntas anteriores com menos suporte visual, isto é sem o mapa.

Para o protocolo foi criado um questionário onde para cada questão é apresentada uma descrição do problema, um ou duas perguntas e os respetivos mapas, bem como as respostas possíveis. Na Figura 5.1 está representada uma das questões usadas durante o questionário. Nesta questão são apresentados dois mapas, com as respetivas perguntas. O uso de um questionário estatístico em vez do protótipo deve-se à necessidade de focar os utilizadores para a análise, isto é a qualidade da visualizações. Se fosse usado o protótipo havia o risco de os utilizadores começarem a avaliar a interface deste em vez das visualizações cartográficas.

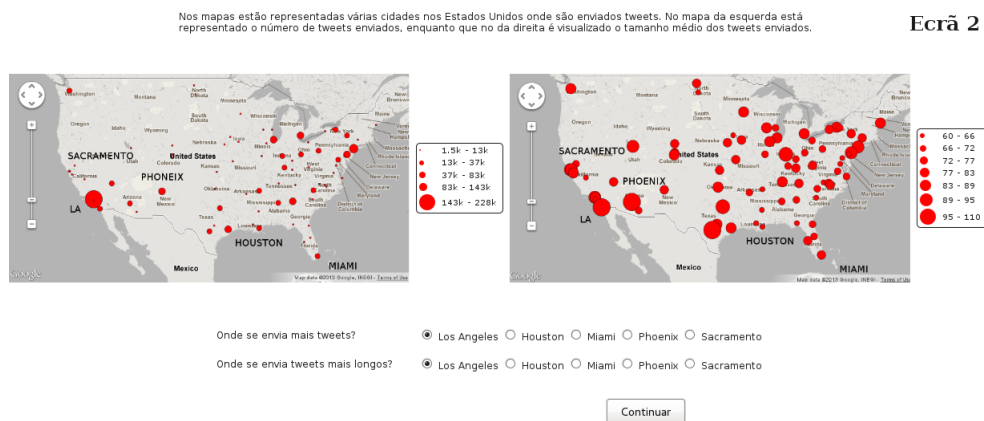


Figura 5.1: Questão usada no Protocolo Experimental

No protocolo as questões vão ser de um de quatro tipos: (i) visualização cartográfica assistida (*VCA*) vs *SOLAP+* (utilizando o mecanismo não extensível presente no protótipo *SOLAP+* implementando em [Sil10]); (ii) questões repetidas do tipo (i) sem suporte visual, isto é sem mapa; (iii) questões do tipo (i) mas onde são usados dois mapas e duas perguntas sobre um contexto; (iv) por fim questões onde os mapas foram gerados usando a visualização cartográfica assistida sendo estas geradas aos pares para cada contexto.

As questões do tipo (i) vão servir para comparar a *VCA* com o *SOLAP+* tanto na velocidade de aquisição de informação como na qualidade da informação obtida. As questões do tipo (ii) irão servir para testar retenção de informação tanto das visualizações com a *VCA* e do *SOLAP+*. No caso das questões do tipo (iii) servem para o mesmo efeito do que as do tipo (i) mas são contextos mais desafiantes para os utilizadores. Por fim as perguntas do tipo (iv) servem para comparar duas visualizações, geradas pela *VCA*, para o mesmo contexto, mas fazendo apenas pequenas alterações como por exemplo variar o número de intervalos.

O protocolo começa por apresentar três questões do tipo (i) ou (iii), seguidas de uma do tipo (ii), este padrão volta a ser repetido mais uma vez. Depois serão apresentadas três questões do tipo (iv), seguidas de duas do tipo (i). As últimas 3 perguntas pertencem ao tipo (iv). Na Figuras 5.2 e 5.3 estão representados os dois caminhos de questões possíveis. As questões são representadas usando um número, para além do número podem estar anotadas com: *VCA* análise em que os resultados são apresentados usando a visualização cartográfica assistida; *SOLAP+* os resultados são apresentados usando as formas de visualização da implementação anterior do *SOLAP+*; *M* indica que são perguntas repetidas sem suporte visual. Cada utilizador irá responder apenas a um caminho sendo estes alternados entre utilizadores.

1(SOLAP+) → 2(VCA) → 3(SOLAP+) → 1M → 4(VCA) → 5(SOLAP+) → 6(VCA) → 4M → 9 → 11 → 13 → 7(SOLAP+) → 8(VCA) → 10 → 12 → 14

Figura 5.2: Caminho de Perguntas 1

1(VCA) → 2(SOLAP+) → 3(VCA) → 1M → 4(SOLAP+) → 5(VCA) → 6(SOLAP+) → 4M → 10 → 12 → 14 → 7(VCA) → 8(SOLAP+) → 9 → 11 → 13

Figura 5.3: Caminho de Perguntas 2

As análises foram criadas sobre três conjuntos de dados: (i) dados sobre emissões poluentes em Portugal, é um conjunto de dados de pequena dimensão onde grande parte dos objetos espaciais são polígonos; (ii) dados provenientes do *tweeter*, conjunto de dados com grande dimensão mas poucos objetos espaciais diferentes; (iii) acidentes nos Estados Unidos durante oito anos, grande diversidade de objetos espaciais. O Anexo B contém as perguntas efetuadas neste protocolo. Os *screenshots* de todas as questões realizadas no protocolo experimental estão disponíveis em <https://dl.dropboxusercontent.com/u/1188536/protocolo-experimental.zip>.

De seguida, irão ser apresentados os resultados obtidos na execução do protocolo apresentado anteriormente. De início serão apresentadas estatísticas sobre a população, seguidos dos resultados referentes às três categorias de questões.

5.3 Resultados do Protocolo Experimental

A população que respondeu a este questionário é composta por 30 indivíduos com uma média de idades de 24.9 anos. Sendo 80% dos indivíduos do sexo masculino e 20% do sexo feminino. Para averiguar se os indivíduos estavam familiarizados com alguns conceitos usados em análise de dados foram feitas as perguntas que estão apresentadas na Tabela 5.2. Para cada pergunta são apresentados os resultados obtidos.

	Nunca	Algumas Vezes	Frequentemente
Com que frequência realiza análise de dados ?	40 %	40 %	20 %
Com que frequência usa mapas (Google Maps, Bing Maps, etc.) ?	0 %	27 %	73 %
Com que frequência usa mapas temáticos ?	23 %	67 %	10 %

Tabela 5.2: Estatísticas sobre a População

Os primeiros resultados a serem apresentados são relativos á comparação entre análises com e sem a visualização cartográfica assistida. No gráfico da Figura 5.4 são apresentados os tempos médios de resposta e os respetivos desvios padrão, para cada questão. Os tempos estão separados entre as análises com *VCA* (Visualização Cartográfica Assistida) e sem *VCA*. Como mostra a Figura 5.4 não existem diferenças significativas entre os tempos registados nas questões com e sem *VCA* com exceção da primeira questão. Isto é confirmado quando recorrendo ao teste Z como mostra a Tabela 5.3. Foi testada a hipótese de os tempos usando *VCA* serem menores a uma taxa de confiança de 95 %, isto é o valor *P* ser menor ou igual de 0.05.

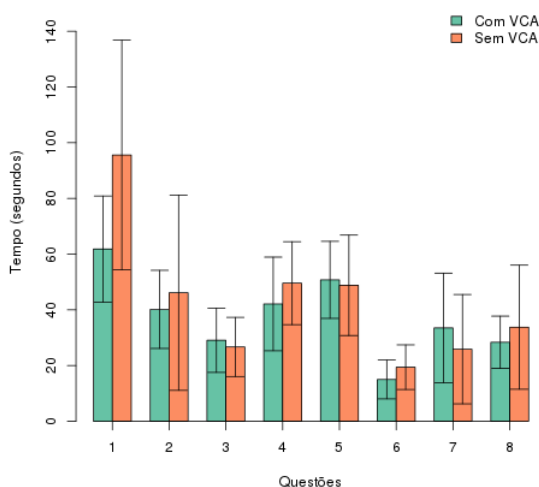


Figura 5.4: Comparação do Tempo de Resposta com e sem *VCA*

Questão	1	2	3	4	5	6	7	8
Valor <i>P</i>	0.0007	0.2538	0.8090	0.0260	0.6630	0.0164	0.9340	0.1732

Tabela 5.3: Resultados do Teste Z Para os Tempos de Resposta com e sem *VCA*

Como mostra a Figura 5.5 relativamente a percentagem de respostas certas não existe grande diferença entre os teste com e sem *VCA*, com exceção da questão 1 onde a percentagem de respostas certas com *VCA* é bastante superior. Isto mostra que o modelo de visualização cartográfica assistida não tem perdas significativas, sendo por vezes melhor quando comparado com a implementação anterior do protótipo *SOLAP+*. Na questão 5 ambos os sistemas usaram como visualização uma representação baseada em gráfico de setores, quando a análise pedia para comparar valores absolutos o que torna esta forma de visualização pouco adequada para este caso, apesar de ser considerada válida por ambos os sistemas. Isto prova que visualizações inadequadas podem prejudicar as análises executadas.

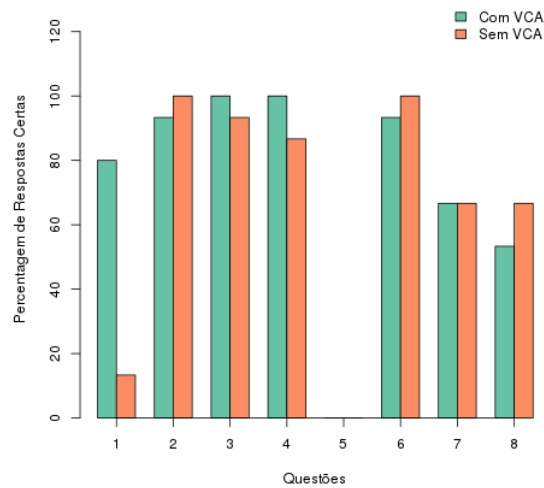


Figura 5.5: Percentagem de Respostas Certas com e sem VCA

Quando comparados os tempos de resposta usando os contextos repetidos em análises apenas, gráfico da Figura 5.6, com VCA notamos que estes tendem a ser menores da segunda vez que o contexto aparece. Os valores relativos ao teste Z, Tabela 5.4 teste para 95 % de confiança isto é o valor P ser menor ou igual a 0.05, comprovam que os tempos de resposta são significativamente melhores.

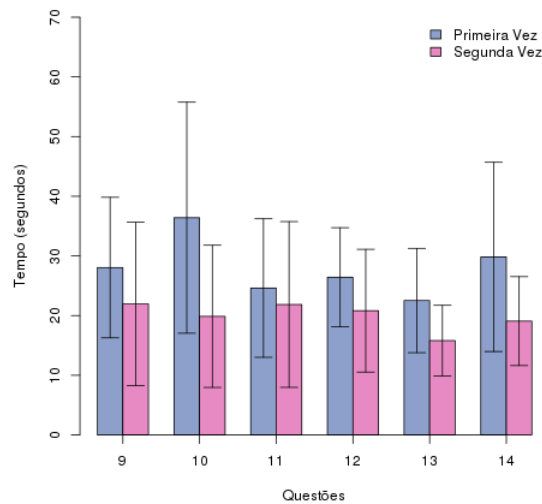


Figura 5.6: Comparação do Tempo de Resposta Usando Contextos Repetidos

Questão	9	10	11	12	13	14
Valor P	0.0227	0.0004	0.1790	0.0044	0.0014	0.0042

Tabela 5.4: Resultados do Teste Z Usando Contextos Repetidos

No caso das respostas certas usando contextos repetidos, Figura 5.7, podemos verificar que as questões como segunda vez tem uma maior percentagem de resposta certas

ou são similares ao da primeira vez. Olhando para os valores tanto do tempo como das percentagens de resposta certas podemos verificar que conseguimos melhores tempos a quando da segunda vez com percentagens de respostas certas similares ou melhores. Com esta relação em mente podemos afirmar que a *VCA* consegue algumas melhorias na retenção de informação por parte dos utilizadores.

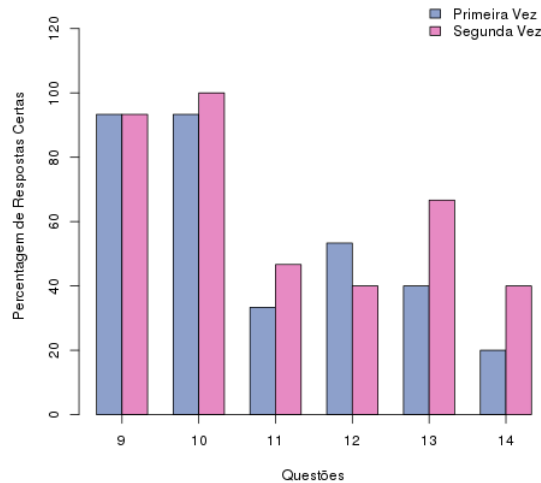


Figura 5.7: Percentagem de Respostas Certas Usando Contextos Repetidos

Na Figura 5.8 estão representados a média e desvio padrão dos tempos de resposta nas perguntas repetidas sem suporte bem como nas respetivas perguntas com o suporte visual. Nas perguntas repetidas existe uma redução bastante grande no tempo de resposta tanto como e sem *VCA* quando comparando com as perguntas originais. Mas se a comparação for feita só entre as perguntas com e sem *VCA* não existe variações significativas.

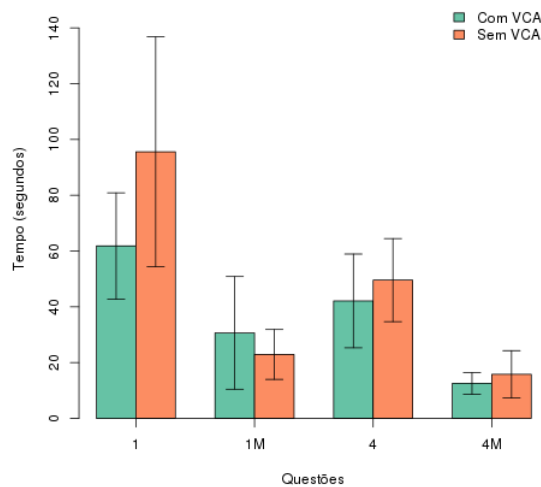


Figura 5.8: Comparação do Tempo de Resposta nas Perguntas Repetidas sem Suporte Visual

Quando examinando a percentagem de respostas, gráfico da Figura 5.9, podemos verificar que a percentagem de respostas certas com VCA é mais elevada sendo os valores na pergunta retidas similares á pergunta original. Apesar de existir uma grande redução do tempo de resposta nas perguntas com memória este parece não afetar a percentagem de resposta certas. Tal como já tinha sido nas perguntas com contextos repetidos, a VCA consegue melhorar a retenção de informação por parte dos utilizadores, melhorando os tempos de resposta mantendo percentagens similares de respostas certas.

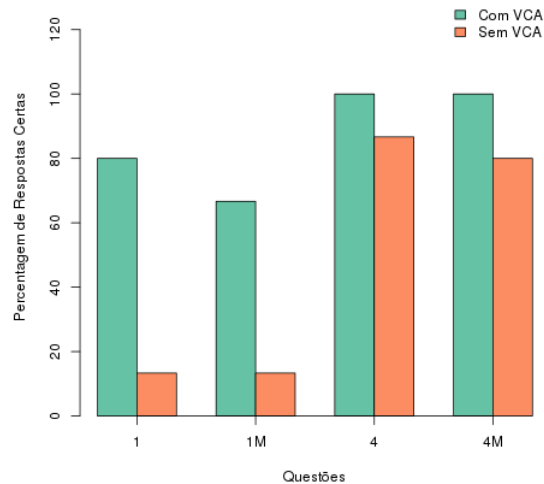


Figura 5.9: Percentagem de Resposta Certas nas Perguntas Repetidas sem Suporte Visual

Com os resultados obtidos podemos concluir que a visualização cartográfica assistida comporta-se de forma semelhante na qualidade da informação obtida, sendo que em alguns casos consegue ser melhor do que o SOLAP+ onde a implementação está *hardcoded*. Do ponto de vista de retenção de informação a visualização cartográfica assistida consegue resultados melhores do que a implementação de referencia no SOLAP+. Por fim com os resultados obtidos nas perguntas repetidas para o mesmo contexto podemos concluir que é mais importante a escolha do tipo de visualização aplicável do que a afinação deste (por exemplo número de intervalos).



Conclusão e Trabalho Futuro

O objetivo inicialmente definido neste trabalho consistia em propor e implementar uma solução para o problema de, no contexto de um sistema *SOLAP* genérico, determinar quais as visualizações, baseadas em mapas, que são aplicáveis num dado momento da análise, e garantir que a sua aplicação segue os bons princípios de cartografia e da visualização de informação. Um requisito importante definido desde cedo é a capacidade de a solução adotada vir a acomodar outras formas de visualização de informação no mapa além daquelas que forem identificadas durante este trabalho.

A abordagem seguida assenta nas propostas dos modelos seguintes: (i) um modelo para descrever o contexto analítico, que inclui os modelos multidimensionais *SOLAP*, a interrogação realizada, e os dados obtidos. O enriquecimento com estatísticas sobre os dados e das relações entre as colunas que participam nos resultados são aspetos importantes para a determinação das formas de visualização mais apropriadas a cada contexto; (ii) um modelo abstrato para descrever tipos de visualização (mapas temáticos), de maneira a que seja possível determinar a sua aplicabilidade num contexto analítico descrito pelo modelo anterior. No quadro deste modelo foi proposto uma extensão às variáveis visuais definidas por Bertin; (iii) por fim foi também proposto um modelo de fábrica de visualizações capaz de enquadrar a implementação de visualizações, que suportam os tipos de visualização anteriormente definidos.

São estes os ingredientes principais do sistema proposto que, perante um contexto *SOLAP* descrito de acordo com o primeiro modelo, e tendo em conta os tipos de visualização disponíveis, seleciona os que são aplicáveis, escolhendo um e apoiando a sua aplicação segundo os bons princípios de cartografia e da visualização de informação.

Esta proposta foi implementada sobre uma implementação genérica de *SOLAP* e desenvolvida no departamento, o sistema *SOLAP+*. No entanto, tal como foi explicado

na Secção 4, foi necessária uma profunda revisão da arquitetura e implementação do mesmo. Com este processo ganhou-se em modularidade e extensibilidade do sistema e independência do *SGBD* onde se encontra implementado o modelo multidimensional.

A modularidade proposta que se ganhou permitiu implementar o sistema proposto nesta dissertação como um módulo de visualização. A validação do sistema proposto apenas seria viável com a ilustração de um número de visualizações significativas. Foram implementadas seis fábricas de visualização que dão suporte a cerca de quarenta tipos de visualização. De maneira a ilustrar que o sistema proposto pode ser acomodar outras formas de visualização de informação no mapa além daquelas que foram identificadas neste trabalho, foi implementada uma fábrica que usa como representação gráfica *smileys*.

Para realizar a validação do sistema proposto foi criado um questionário onde são pedidas várias análises aos utilizadores. Com os resultados obtidos nestas análises pode-se concluir que o sistema implementado consegue resultados semelhantes aos do método de visualização não extensível implementado anteriormente no *SOLAP+*, tanto na qualidade da informação obtida como no tempo necessário para a adquiri-la. Os resultados obtidos para a retenção de informação mostram uma melhoria da visualização cartográfica assistida face ao *SOLAP+*.

Como trabalho futuro existem vários tópicos que podem ser abordados. O primeiro seria a criação de mais paletes de tipos de visualização e respetivas fábricas, de modo a ter uma maior diversidade de representações. Para além das visualizações podiam ser criadas heurísticas (usando como base as recomendações apresentadas na Secção 2.2.3) para melhorar a escolha das instanciações de visualizações.

Para além destas funcionalidades algo que seria interessante para estender o modelo proposto seria visualizações interativas. Atualmente as visualizações geradas são estáticas sem elementos dinâmicos. Com a introdução de interatividade iria ser possível utilizar novas formas de análise. Para além da interatividade o modelo proposto poderia ser estendido para conseguir adaptar os conjuntos de dados aos visualizações.

Por fim a ultima proposta para trabalho futuro seria adicionar ao *Contexto SOLAP* o histórico da interação dos utilizadores e conseguir reconhecer a intenção analítica de cada análise de modo a conseguir melhorar a escolha de visualizações no modelo.

Bibliografia

- [AAV03] Gennady Andrienko, Natalia Andrienko, e Hans Voss. GIS for everyone: the CommonGIS project and beyond. *Maps and the Internet*, pág. 131–146, 2003.
- [AS04] R. Amar e J. Stasko. A knowledge task-based framework for design and evaluation of information visualizations. In *Information Visualization, 2004. INFOVIS 2004. IEEE Symposium on*, pág. 143–150, 2004.
- [ASW11] David R. Anderson, Dennis J. Sweeney, e Thomas A. Williams. *Statistics for Business and Economics*. Cengage Learning, Março 2011.
- [Ber67] J. Bertin. Sémiologie graphique. les diagrammes, les réseaux, les cartes. *The Geographical Journal*, 135:144–146, 1967.
- [Bim10] S. Bimonte. On modeling and analysis of multidimensional geographic databases. *Data Warehousing Design and Advanced Engineering Applications: Methods for Complex Construction*, pág. 96, 2010.
- [BMG⁺10] Scott Bateman, Regan L. Mandryk, Carl Gutwin, Aaron Genest, David McDine, e Christopher Brooks. Useful junk?: the effects of visual embellishment on comprehension and memorability of charts. In *Proceedings of the 28th international conference on Human factors in computing systems*, pág. 2573–2582, 2010.
- [BRP07] Y. Bédard, S. Rivest, e M. J. Proulx. Spatial on-line analytical processing (SOLAP): concepts, architectures, and solutions from a geomatics engineering perspective. *Data Warehouses and OLAP: Concepts, Architectures, and Solutions*, pág. 298, 2007.
- [BSBC03] S. Bigwood, M. Spore, e B. Baumgartner-Cohen. *Presenting numbers, tables, and charts*. Oxford University Press, 2003.

- [BTMP10] S. Bimonte, A. Tchounikine, M. Miquel, e F. Pinet. When spatial analysis meets olap: Multidimensional model and operators. *International Journal of Data Warehousing and Mining (IJDWM)*, 6(4):33–60, 2010.
- [BWTM06] S. Bimonte, P. Wehrle, A. Tchounikine, e M. Miquel. Gewolap : A web based spatial olap proposal. In *On the Move to Meaningful Internet Systems 2006: OTM 2006 Workshops*, pág. 1596–1605, 2006.
- [Che05] C. Chen. Top 10 unsolved information visualization problems. *Computer Graphics and Applications, IEEE*, 25(4):12–16, 2005.
- [DB12] Z. Dobesova e J. Brus. Intelligent systems in cartography. *Intelligent Systems*, pág. 257, 2012.
- [DCDFL⁺11] D. De Chiara, V. Del Fatto, R. Laurini, M. Sebillio, e G. Vitiello. A chorembased approach for visually analyzing spatial data. *Journal of Visual Languages & Computing*, 2011.
- [EAAB08] R. Edsall, G. Andrienko, N. Andrienko, e B. Buttenfield. Interactive maps for exploring spatial data. *ASPRS Manual of GIS*, 2008.
- [Few04] S. Few. *Show me the numbers: Designing Tables and Graphs to Enlighten*. Number ISBN 0970601999. Analytics Press, 2004.
- [Few08] S. Few. Time on the horizon. *Visual Business Intelligence Newsletter*, pág. 1–7, 2008.
- [FJ10] C. Forsell e J. Johansson. An heuristic set for evaluation in information visualization. In *Proceedings of the International Conference on Advanced Visual Interfaces*, pág. 199–206, 2010.
- [Fra98] A. U Frank. Different types of "times" in GIS. *Spatial and temporal reasoning in geographic information systems*, pág. 40, 1998.
- [Fre92] U. Freitag. *Kartographische konzepte*. Berliner Geowissenschaftliche Abhandlungen, Reie C, Band 13, 1992.
- [FS10] V. R.T Ferraz e M. T.P Santos. Globeolap : Improving the geospatial realism in multidimensional analysis environment. In *12th International Conference on Enterprise Information Systems*, pág. 99–107, 2010.
- [Har75] J. A. Hartigan. Printer graphics for clustering. *Journal of Statistical Computation and Simulation*, 4(3):187–213, 1975.
- [HAS11] Jessica Hullman, Eytan Adar, e Priti Shah. Benefitting InfoVis with visual difficulties. *Visualization and Computer Graphics, IEEE Transactions on*, 17(12):2213–2222, 2011.

- [HBO10] Jeffrey Heer, Michael Bostock, e Vadim Ogievetsky. A tour through the visualization zoo. *Communications of the ACM*, 53(6):59, Junho 2010.
- [Inm05] W. H. Inmon. *Building the data warehouse*. John Wiley & Sons, 2005.
- [Ins85] A. Inselberg. The plane with parallel coordinates. *The Visual Computer*, 1(2):69–91, 1985.
- [Int12] Intelli3. Map4Decision. http://www.intelli3.com/en/map4decision_en, 2012.
- [Jor09] Ruben Jorge. *SOLAP+ : extending the interaction model*. Master thesis, Universidade Nova de Lisboa - Faculdade Ciências e Tecnologia, Monte de Caparica, 2009.
- [K2 12] K2 geospatial. JMap. <http://www.k2geospatial.com/gis/jmap>, 2012.
- [KO03] M. J. Kraak e F. Ormeling. *Cartography : visualization of geospatial data*. Pearson Education, 2003.
- [Kof55] K. Koffka. *Principles of Gestalt psychology*. New York, 1955.
- [La11] Tiago Dias Silva Leão. *Caracterização espacial utilizando indução orientada aos atributos no SOLAP+*. Master thesis, Universidade Nova de Lisboa - Faculdade Ciências e Tecnologia, Monte de Caparica, 2011.
- [Mat06] Rosa Matias. *Integração de Informação Geográfica em Sistemas {OLAP}*. Master thesis, Universidade Nova de Lisboa - Faculdade Ciências e Tecnologia, Monte de Caparica, 2006.
- [MHS07] Jock D. Mackinlay, Pat Hanrahan, e Chris Stolte. Show me: Automatic presentation for visual analysis. *Visualization and Computer Graphics, IEEE Transactions on*, 13(6):1137–1144, 2007.
- [MZ05] E. Malinowski e E. Zimányi. Spatial hierarchies and topological relationships in the spatial MultiDimER model. *Database: Enterprise, Skills and Innovation*, pág. 181–194, 2005.
- [MZ07] E. Malinowski e E. Zimányi. Logical representation of a conceptual model for spatial data warehouses. *Geoinformatica*, 11(4):431–457, 2007.
- [NM94] J. Nielsen e R. L Mack. *Usability inspection methods*. 1994.
- [Pas04] J. Pastor. *Conception d'une légende interactive et forable pour le SOLAP*. Tese de Doutoramento, Université Laval, 2004.
- [RBC09] Nils Rasmussen, Manish Bansal, e Claire Y. Chen. *Business Dashboards: A Visual Catalog for Design and Deployment*. John Wiley & Sons, Março 2009.

- [RBM01] S. Rivest, Y. Bedard, e P. Marchand. Toward better support for spatial decision making: Defining the characteristics of spatial on-line analytical processing (SOLAP). *Geomatica-Ottawa*, 55(4):539–555, 2001.
- [Shn96] B. Shneiderman. The eyes have it: A task by data type taxonomy for information visualizations. In *Visual Languages, 1996. Proceedings., IEEE Symposium on*, pág. 336–343, 1996.
- [Sil10] Ricardo Filipe Silva. *SOLAP+*. Master thesis, Universidade Nova de Lisboa - Faculdade Ciências e Tecnologia, Monte de Caparica, 2010.
- [SP05] M. Scotch e B. Parmanto. SOVAT : Spatial OLAP visualization and analysis tool. In *System Sciences, 2005. HICSS'05. Proceedings of the 38th Annual Hawaii International Conference on*, pág. 142b–142b, 2005.
- [Tuf83] E. R Tufte. *The visual display of quantitative information*, volume 7. Graphics press Cheshire, CT, 1983.
- [Tuf06] E. R Tufte. *Beautiful evidence*, volume 23. Graphics Press Cheshire, CT, 2006.
- [TW10] V. Tirronen e M. Weber. Sparkline histograms for comparing evolutionary optimization methods. In *Proceedings of International Conference on Evolutionary Computation*, 2010.
- [WAM01] M. Weber, M. Alexa, e W. Müller. Visualizing time-series on spirals. In *proceedings of the IEEE Symposium on Information Visualization*, pág. 7, 2001.
- [War04] C. Ware. *Information visualization: perception for design*, volume 22. Morgan Kaufmann, 2004.
- [ZC06] T. Zuk e S. Carpendale. Theoretical analysis of uncertainty visualizations. In *Proceedings of SPIE*, volume 6060, pág. 66–79, 2006.
- [ZSB⁺12] Leishi Zhang, Andreas Stoffel, Michael Behrisch, Sebastian Mittelstadt, Tobias Schreck, René Pompl, Stefan Weber, Holger Last, e Daniel Keim. Visual analytics for the big data era - a comparative review of state-of-the-art commercial systems. In *Visual Analytics Science and Technology (VAST), 2012 IEEE Conference on*, pág. 173–182, 2012.
- [ZSN⁺06] T. Zuk, L. Schlesier, P. Neumann, M. S Hancock, e S. Carpendale. Heuristics for information visualization evaluation. In *Proceedings of the 2006 AVI workshop on BEyond time and errors: novel evaluation methods for information visualization*, pág. 1–6, 2006.



Palete de Tipos de Visualizações

Neste anexo está descrita parte da paleta de tipos de visualização criada no âmbito deste trabalho, serão apresentados dezoito tipos de visualização de um total de quarenta e três. De notar que a lista a ser apresentada não é de modo algum exaustiva, dado que só serão apresentados tipos de visualização para algumas das variáveis visuais clássicas, e tipos de visualização que usam gráficos (gráfico de barras, gráficos de setores e gráficos de linhas).

Para cada tipo de visualização irá ser apresentado o seu nome, a fábrica a que se refere as variáveis visuais, o objeto espacial a que este se aplica e quais as variáveis visuais presentes. Para tornar mais compacta a representação das variáveis visuais estas serão representadas usando a seguinte forma:

{Nome, [Cardinalidade], Classe Visual, Tipos de Dados Aceitáveis}

A.1 Ponto Cor

Objeto Espacial : Ponto, Fábrica : Ponto

1. *{Cor, [1], Seletiva, Nominal}*

A.2 Ponto Tamanho

Objeto Espacial : Ponto, Fábrica : Ponto

1. *{Tamanho, [1], Ordenável, Real}*

A.3 Ponto Saturação

Objeto Espacial : Ponto, Fábrica : Ponto

1. {Saturação, [1], Ordenável, Real}

A.4 Ponto Cor e Tamanho

Objeto Espacial : Ponto, Fábrica : Ponto

1. {Cor, [1], Seletiva, Nominal}
2. {Tamanho, [1], Ordenável, Real}

A.5 Ponto Cor e Saturação

Objeto Espacial : Ponto, Fábrica : Ponto

1. {Cor, [1], Seletiva, Nominal}
2. {Saturação, [1], Ordenável, Real}

A.6 Ponto Tamanho e Saturação

Objeto Espacial : Ponto, Fábrica : Ponto

1. {Saturação, [1], Ordenável, Real}
2. {Tamanho, [1], Ordenável, Real}

A.7 Ponto Gráfico de Setores

Objeto Espacial : Ponto, Fábrica : Gráfico de Setores

1. {Etiquetas, [1], Conjunto Seletivo, Nominal}
2. {Valores, [1], Conjunto Ordenável, Real}

A.8 Ponto Gráfico de Barras

Objeto Espacial : Ponto, Fábrica : Gráfico de Barras

1. {Etiquetas, [1], Conjunto Seletivo, Nominal}
2. {Valores, [1..5], Conjunto Ordenável, Real}

A.9 Ponto Gráfico de Linhas

Objeto Espacial : Ponto, Fábrica : Gráfico de Linhas

1. {Etiquetas, [1], Conjunto Ordenável, Real}
2. {Valores, [1..5], Conjunto Ordenável, Real}

A.10 Polígono Cor

Objeto Espacial : Polígono, Fábrica : Polígono

1. {Cor, [1], Seletiva, Nominal}

A.11 Polígono Saturação

Objeto Espacial : Polígono, Fábrica : Polígono

1. {Saturação, [1], Ordenável, Real}

A.12 Polígono Cor e Saturação

Objeto Espacial : Polígono, Fábrica : Polígono

1. {Cor, [1], Seletiva, Nominal}
2. {Saturação, [1], Ordenável, Real}

A.13 Polígono Gráfico de Setores

Objeto Espacial : Polígono, Fábrica : Gráfico Setores

1. {Etiquetas, [1], Conjunto Seletivo, Nominal}
2. {Valores, [1], Conjunto Ordenável, Real}

A.14 Polígono Gráfico de Barras

Objeto Espacial : Polígono, Fábrica : Gráfico Barras

1. {Etiquetas, [1], Conjunto Seletivo, Nominal}
2. {Valores, [1..5], Conjunto Ordenável, Real}

A.15 Linha Cor

Objeto Espacial : Linha, Fábrica : Linha

1. {Cor, [1], Seletiva, Nominal}

A.16 Linha Saturação

Objeto Espacial : Linha, Fábrica : Linha

1. {Saturação, [1], Ordenável, Real}

A.17 Linha Textura

Objeto Espacial : Linha, Fábrica : Linha

1. {Textura, [1], Seletiva, Nominal}

A.18 Linha Cor e Saturação

Objeto Espacial : Linha, Fábrica : Linha

1. {Cor, [1], Seletiva, Nominal}
2. {Saturação, [1], Ordenável, Real}



Questões usadas no Protocolo de Validação

Neste anexo estão declaradas as questões usadas no protocolo de validação. Para cada questão é apresentado o texto relativo a descrição do contexto analítico, as perguntas realizadas e a indicação das respetivas respostas certas. As respostas corretas aparecem marcadas com um asterisco.

B.1 Questão 1

Nos mapas estão representadas várias cidades nos Estados Unidos onde são enviados tweets. No mapa da esquerda está representado o número de tweets enviados, enquanto que no da direita é visualizado o tamanho médio dos tweets enviados.

- Onde se envia mais tweets?
 - Los Angeles (*); Houston; Miami; Phoenix; Sacramento
- Onde se envia tweets mais longos?
 - Los Angeles; Houston; Miami; Phoenix (*); Sacramento

B.2 Questão 2

No mapa estão representadas várias fábricas em Portugal. Para cada fábrica é mostrado em que bacia hidrográfica ela está situada.

- Quais as Bacias com mais instalações?

– Tejo (*); Algarve; Douro; Norte; Sado e Mira

- Quais as Bacias com menos instalações?

– Tejo; Algarve (*); Douro; Norte; Sado e Mira

B.3 Questão 3

No mapa estão representados o número de acidentes de viação para cada estado, nos Estados Unidos.

- Onde existem mais acidentes?

– Ohio; California (*); New York; New Mexico; Florida

B.4 Questão 4

Nos mapas estão representado para cada estado o número de pessoas mortas em acidentes (mapa da esquerda) e o número de pessoas embriagadas envolvidas em acidentes (mapa da direita).

- Onde morreram mais pessoas?

– Nevada; Florida; Washington; California (*); New Hampshire

- Onde existem mais acidentes envolvendo pessoas embriagadas?

– Nevada; Florida; Washington; California (*); New Hampshire

B.5 Questão 5

No mapa está representado o número de tweets enviados em cada estado por língua em são escritos.

- Onde existem mais tweets em português?

– New York; Texas; California (*); Maryland; Ohio

B.6 Questão 6

No mapa está representada a quantidade de emissões poluentes em cada distrito.

- Onde existe mais poluição?

– Porto; Setúbal (*); Lisboa; Santarém; Braga

B.7 Questão 7

No mapa está representada a quantidade de emissões poluentes em cada distrito.

- Qual o distrito onde não existe dados referentes as emissões?
 - Vila Real; Guarda; Beja; Évora; Bragança (*)

B.8 Questão 8

No mapa está representado o número de tweets enviados por fonte em cada estado.

- Qual as duas formas mais comuns de enviar tweets na costa oeste?
 - Web e Twitterfeed (*); Twitter for iPhone e Twitter for Android; Web e Twitter for iPhone; Twitter for Android e Twitterfeed; Web e Twitter for Android

B.9 Questão 9

No mapa estão representados vários locais de acidentes no Estados Unidos, para cada acidente é mostrado o número de vítimas mortais.

- Onde morrem mais pessoas em acidentes?
 - Costa Oeste; Costa Este (*); Centro

B.10 Questão 10

No mapa estão representados vários locais de acidentes no Estados Unidos, para cada acidente é mostrado o número de vítimas mortais.

- Onde morrem mais pessoas em acidentes?
 - Costa Oeste; Costa Este (*); Centro

B.11 Questão 11

No mapa estão representadas várias fábricas em Portugal. Para cada fábrica é mostrado o valor das suas emissões poluentes.

- Qual a zona com mais emissões?
 - Norte; Centro; Vale do Tejo(*); Sul

B.12 Questão 12

No mapa estão representadas várias fábricas em Portugal. Para cada fábrica é mostrado o valor das suas emissões poluentes.

- Qual a zona com menos emissões?
 - Norte; Centro; Vale do Tejo; Sul(*)

B.13 Questão 13

No mapa está representado o número de tweets enviados em cada estado.

- Qual a zona com menos tweets?
 - Norte (*); Centro; Este; Sul; Oeste

B.14 Questão 14

No mapa está representado o número de tweets enviados em cada estado.

- Qual a zona com mais tweets?
 - Norte; Centro; Este; Sul(*); Oeste