

FACULDADE DE ENGENHARIA DA UNIVERSIDADE DO PORTO



FEUP

Visualização de estatísticas pessoais e sociais para um produto Web e Mobile

João Barbosa

Mestrado Integrado em Engenharia Informática e Computação

Orientador: Alexandre Valle de Carvalho (PhD)

Co-orientador: Pedro Melo Campos (MSc)

18 de Junho de 2012

Visualização de estatísticas pessoais e sociais para um produto Web e Mobile

João Barbosa

Mestrado Integrado em Engenharia Informática e Computação

Aprovado em provas públicas pelo Júri:

Presidente: Doutora Cristina Ribeiro (Prof. Auxiliar da FEUP)

Arguente: Doutor José Manuel Torres (Prof. Associado da Universidade Fernando Pessoa)

Orientador: Doutor Alexandre Valle de Carvalho (Prof. Auxiliar da FEUP)

18 de Junho de 2012

Resumo

A visualização de informação é uma área que tem tido uma rápida evolução nos últimos tempos, potenciada pelo desenvolvimento de tecnologias de visualização gráfica, mas também pela constante descoberta de novos algoritmos de visualização que permitem uma melhor perceção e organização de informação. A aplicação destes novos modelos de visualização ainda está um pouco restrita ao domínio da investigação e das ciências da comunicação, sendo ainda pouco utilizada em aplicações de software mais comuns, como as aplicações móveis. É na interseção do domínio da visualização com o domínio das aplicações móveis que surge o *Statz.Me*.

Esta aplicação tem como objetivo não só o registo e partilha de informações pessoais categorizadas, mas também oferecer ao utilizador visualizações mais atrativas sobre essas estatísticas, e que lhe permitam ter uma melhor leitura sobre a sua informação pessoal. É nesta aplicação que foram implementados os modelos de visualização estudados, oferecendo aos utilizadores uma análise mais intuitiva sobre as suas atividades. A integração na aplicação *Statz.Me* de modelos que relacionem múltiplas variáveis, que permitam a interação do utilizador e que ofereçam a possibilidade de uma análise geral imediata, bem como da análise de informação mais detalhada são alguns dos objetivos deste trabalho. A implementação de métodos de visualização que relacionem várias estatísticas, ao mesmo tempo, é o desafio principal deste trabalho, uma vez que estes métodos permitem ao utilizador realizar uma análise sobre as possíveis dependências entre as estatísticas a relacionar, o que o pode ajudar a tirar conclusões importantes sobre as influências de certas atividades no seu comportamento.

Outro dos objetivos do trabalho consiste no desenvolvimento de um motor de visualizações flexível e que tenha potencial para ser utilizado noutras aplicações e noutros ambientes. Desta forma, foi pensada uma arquitetura composta por vários componentes responsáveis pelo tratamento de dados dos diferentes tipos, e que pudesse tirar o maior partido das potencialidades e da flexibilidade da linguagem *JavaScript*, uma vez que esta linguagem pode ser interpretada em aplicações desenvolvidas em ambientes distintos.

Este trabalho analisa alguns dos sistemas de classificação de visualização existentes, e a forma como esta classificação é feita segundo diferentes perspetivas, sublinhando a importância de uma análise prévia não só ao tipo de dados e características temporais, mas também ao domínio de visualização em que os dados se inserem, ao seu tipo de mapeamento e a forma como o utilizador pode interagir com uma visualização. Este estudo permitiu definir uma perspetiva de solução para o desenvolvimento das diferentes visualizações adaptadas a cada tipo de estatística diferente.

Abstract

The visualization of information is an area that has been growing rapidly in the past few years, due to the development of graphic visualization technologies, but also to the constant discovery of new visualization algorithms that give a better perception and organization of information. The application of these new visualization models is still a bit restricted to research areas and communication science, being underused in common software applications, like mobile applications.

It's in the confluence of these fields that Statz.Me emerged. This application has, not only the goal of the track and sharing of categorized personal data, but also the potentiality of offering attractive visualizations about those statistics that give the user a better comprehension about his/her personal data. This application is the working ground for the implementation of the referred models, offering the users an intuitive analysis over their activities. The integration of multi-variate, interactive, overview-oriented, and detailed analysis models are some of the goals of this work. The implementation of visualization methods that simultaneously relate multiple statistics is one of the main challenges of this work, for it can give the user an analysis over possible statistical dependencies, which can help comprehending some influences among certain activities.

Another goal of this work is the development of a flexible and portable visualization engine, that can be used in other applications and in other environments. Considering this, it was thought of an architecture composed by various components that handle the processing of the different data types, and that could get the most of the potential and flexibility of the *JavaScript* language, since that it can be interpreted in applications based on different environments.

The present work focuses on some of the existing visualization classification systems, and the way that this classification is performed under different perspectives, underlying the importance of a prior analysis not only to the data types and the temporal characteristics, but also to the visualization domain in which the data falls, the mapping type, and the way the user is able to interact with the visualization itself. This study allowed a definition of a solution perspective for the development of different and adaptable visualizations models.

Agradecimentos

Gostaria de agradecer a todos os que directa ou indirectamente contribuíram, não só para a realização desta dissertação mas também para todo o meu percurso até aqui.

Um especial obrigado ao meu orientador, Professor Alexandre Carvalho, e ao meu co-orientador, Engenheiro Pedro Campos, por toda a ajuda e tempo dispensados.

Queria também agradecer aos meus pais e à minha irmã por todo o apoio e incentivo que me deram desde o início.

Por fim, gostava de agradecer a todos os amigos da FEUP que me aturaram durante estes cinco anos fantásticos. Foi um enorme prazer todo o tempo que passei com eles.

João Barbosa

*“It took me a long time to get young, and now I consider myself young.
And I’m proud of it. I’m proud that I’m young”*

Bob Dylan

Conteúdo

1	Introdução	1
1.1	Enquadramento	1
1.2	Objetivos	2
1.3	Estrutura da Dissertação	2
2	Visualização de dados	5
2.1	Introdução	5
2.2	Representação fiel e contextualização da informação	7
2.2.1	O problema das visualizações enganosas	7
2.2.2	Contextualização de informação	9
2.3	Sistemas de Classificação de Modelos de Visualização de Informação	10
2.3.1	Taxonomias por tarefas	11
2.3.2	Taxonomias por caracterização de dados/tempo	12
2.3.3	Tabela Periódica dos métodos de visualização	14
2.3.4	Estética de informação	16
2.3.5	Perceção do utilizador vs suposições do designer	18
2.4	Métodos de visualização	19
2.4.1	Métodos de visualização dependentes do tempo	20
2.4.2	Métodos de visualização temporalmente independentes	21
2.5	Conclusões	24
3	Tipos de dados e arquitetura	29
3.1	Tipos de dados do Statz.Me	29
3.2	Arquitetura do motor de visualizações do Statz.Me	30
3.3	Detalhes de implementação	32
3.3.1	Classes relativas a cada tipo de dados do Statz.Me	32
3.3.2	PlotViewController	33
3.3.3	StatzVisualizationEngine	33
3.3.4	NumericStatHandler	33
3.3.5	RatingStatHandler	35
3.3.6	OptionListStatHandler	35
3.3.7	MultiStatHandler	36
3.4	Navegação entre diferentes métodos de visualização	37
3.5	Conclusões	38
4	d3 & Protovis	39
4.1	Bibliotecas de representação gráfica	39
4.2	Conclusões	41

CONTEÚDO

5	Resultados	43
5.1	Visualizações implementadas	43
5.1.1	Zooming	43
5.1.2	Pie Chart	44
5.1.3	Calendar Chart	44
5.1.4	Sankey Chart	45
5.1.5	Moving Bars	45
5.1.6	Timeline Charts - Dot Charts & Dynamic Bubble Charts	47
5.1.7	Burtin's Antibiotics	49
5.1.8	Static Bubble Charts	50
5.1.9	Análise global aos métodos desenvolvidos	51
5.2	Conclusões	51
6	Conclusões e trabalho futuro	55
6.1	Conclusões	55
6.2	Trabalho Futuro	56
6.2.1	Outras Visualizações	56
6.2.2	Novas Funcionalidades	59
	Referências	63

Lista de Figuras

2.1	Definição de visualização de dados de David McCandless.	6
2.2	Esquema de definição de visualização de dados defendido pela comunidade FFunction.	6
2.3	Exemplo de representação falaciosa de informação, proporções diferentes entre comprimento da linha e valor representado.	8
2.4	Exemplo de representação falaciosa de informação, proporções diferentes entre tamanho de cada barril e valor representado.	8
2.5	Representação original da percentagem de população americana pertencente à classe trabalhadora.	9
2.6	Representação alterada, com alteração do eixo vertical.	9
2.7	Representação das percentagens de população americana que pertence / não pertence à classe trabalhadora.	10
2.8	Tabela Periódica dos Métodos de Visualização.	15
2.9	Diagrama de definição de estética de informação.	17
2.10	Spiral Graph.	21
2.11	Theme River.	22
2.12	Calendar View.	22
2.13	Planning Lines.	22
2.14	Perspective Wall.	22
2.15	Gas and Driving.	23
2.16	Candlestick.	23
2.17	Chord Diagram.	23
2.18	Bubble Chart.	24
2.19	Bars of Stuff.	24
2.20	Bullet Charts.	24
3.1	Primeira versão da arquitectura do motor de visualizações do Statz.Me.	31
3.2	Versão atual da arquitetura do motor de visualizações do Statz.Me.	32
3.3	Diagrama do funcionamento do componente Numeric Stat Handler.	34
3.4	Diagrama do funcionamento do componente Rating Stat Handler.	35
3.5	Diagrama do funcionamento do componente Option List Stat Handler.	36
3.6	Diagrama do funcionamento do componente Multi-Stat Handler.	37
4.1	Gráfico de análise de desempenho das tecnologias SVG e Canvas em dois browsers distintos	41
5.1	<i>Screenshot</i> do protótipo do Statz.Me apresentando uma visualização do tipo <i>Zooming</i> para uma estatística numérica.	44

LISTA DE FIGURAS

5.2	<i>Screenshot</i> do protótipo do Statz.Me apresentando uma visualização do tipo <i>Pie Chart</i> para um estatística do tipo "lista de opções".	44
5.3	<i>Screenshot</i> do protótipo do Statz.Me apresentando uma visualização do tipo <i>Calendar Chart</i> para estatísticas que retratam valores numéricos.	45
5.4	<i>Screenshot</i> do protótipo do Statz.Me apresentando uma visualização do tipo <i>Calendar Chart</i> para estatísticas que retratam valores do tipo "lista de opções".	46
5.5	<i>Screenshot</i> do protótipo do Statz.Me apresentando uma visualização do tipo <i>Sankey Chart</i> para estatísticas que retratam valores do tipo "lista de opções".	46
5.6	<i>Screenshot</i> do protótipo do Statz.Me apresentando uma visualização do tipo <i>Moving Bars</i> para uma estatística numérica.	47
5.7	<i>Screenshot</i> do protótipo do Statz.Me apresentando uma visualização do tipo <i>Dot Chart</i> para uma estatística do tipo "lista de opções".	48
5.8	<i>Screenshot</i> do protótipo do Statz.Me apresentando uma visualização do tipo <i>Dynamic Bubble Chart</i> onde são relacionadas duas estatísticas numéricas com uma do tipo "lista de opções".	48
5.9	Versão original do gráfico criado por Will Burtin.	49
5.10	<i>Screenshot</i> do protótipo do Statz.Me apresentando uma visualização do tipo <i>Burtin's Antibiotics</i> onde são relacionadas três estatísticas numéricas com uma do tipo "lista de opções".	50
5.11	<i>Screenshot</i> do protótipo do Statz.Me apresentando uma visualização do tipo <i>Static Bubble Charts</i> onde são relacionadas três estatísticas numéricas com uma do tipo "lista de opções".	51
6.1	Exemplo de utilização da visualização <i>Mapas</i> - cada círculo é representativo de uma ocorrência de um determinado evento, estando a sua cor associada a uma determinada variável.	57
6.2	Exemplo de utilização da visualização <i>Cross Filtering</i> - através da aplicação de filtros nos diferentes gráficos pode ser tiradas conclusões interessantes sobre a informação retratada	58
6.3	Exemplo de utilização da visualização <i>House Hunting</i> - a percepção do utilizador sobre os valores é facilitada, através da utilização de diferentes cores e da análise dos outros gráficos adjacentes.	59

Lista de Tabelas

2.1	Classificação dos métodos de visualização segundo os sistemas analisados	25
2.2	Classificação dos métodos de visualização segundo os sistemas analisados - con- tinuação	26
2.3	Classificação dos métodos de visualização por tipo de dados e características tem- porais	28
4.1	Vantagens das tecnologias Canvas e SVG	40
4.2	Desvantagens das tecnologias Canvas e SVG	40
5.1	Análise das principais características dos métodos criados.	52
6.1	Exemplos de possíveis grupos de estatísticas	61

LISTA DE TABELAS

Abreviaturas e Símbolos

SVG	Scalable Vector Graphics
XML	Extensible Markup Language
API	Application Programming Interface
DOM	Document Object Model
HTML	HyperText Markup Language
XHTML	Extensible Hypertext Markup Language
iOS	iPhone Operative System
FPS	Frames Per Second

Capítulo 1

Introdução

1.1 Enquadramento

A necessidade de conhecer melhor os hábitos regulares que têm maior influência no dia-a-dia é uma característica do ser humano, que tem vindo a ganhar especial importância nos últimos anos. Informações como, por exemplo, a produtividade diária, a nível profissional, o número de horas de sono, a disposição e stress ao longo do tempo, têm sido alvo de especial análise por parte de todos nos dias que correm, com o objetivo focado no aumento da qualidade de vida de cada um. São estes aspetos do comportamento pessoal de cada indivíduo que podem ser respondidos com números, permitindo uma maior facilidade na análise de informação. Relacionado com estas questões, está o conceito de gamification of life que aborda uma visão da vida como um jogo, envolvendo registo de atividades, comparação de resultados, comparação com outros indivíduos e, principalmente, o estabelecimento de objetivos.

Para um registo fiel e prático deste tipo de informações, cujo domínio é bastante disperso, dependendo daquilo que cada indivíduo quiser registar, é necessário algo que seja facilmente acessível e que disponha de um caráter simples e prático. A recente evolução e expansão dos dispositivos móveis constitui então um excelente ponto de partida para aplicações com estes objetivos, dada a sua proximidade com o utilizador. O crescimento da filosofia de partilha de informação subjacente a domínios como o das redes sociais também serviu de ponto de partida para outro dos objetivos deste tipo de aplicações, que consiste na partilha de informações, podendo esta ser na forma de resultados ou atividades semelhantes a realizar. A partilha de informação, aliada a uma filosofia de colaboração, pode ser útil aos utilizadores para atingirem determinados objetivos, uma vez que poderá afetar a sua motivação na realização de determinadas tarefas com eles relacionadas, entrando novamente nesta área o conceito de gamification of life – numa perspetiva de “competição” entre utilizadores.

Para uma correta análise dos dados registados é necessário que o utilizador disponha de métodos de visualização claros, auto-explicativos e que permitam uma interpretação correta da infor-

Introdução

mação. Para a construção destes métodos é necessário ter em conta o tipo de dados a analisar: o número de variáveis e a forma como estas são (ou não) dependentes entre si, bem como a análise das características temporais à qual a informação está ligada. Em alguns casos, a interação do utilizador com esses métodos leva a uma melhor contextualização com a informação e, consequentemente uma melhor análise da mesma. A motivação principal deste trabalho é, então, fornecer aos utilizadores uma visão mais clara sobre cada tipo de estatística, de forma a este poder ter uma melhor ideia das suas atividades, e assim compreender quais são os hábitos diários que influenciam o seu comportamento. Outra das motivações deste trabalho prende-se com o facto de, por vezes, determinadas atividades do dia-a-dia estarem relacionadas ou terem influência sobre outras, o que nem sempre é perceptível a cada indivíduo.

1.2 Objetivos

O objetivo principal desta dissertação prende-se com o desenvolvimento e integração na aplicação *Statz.Me* de métodos de visualização de estatísticas adaptados a cada tipo pretendido pelo utilizador, e que, quando possível, permitam diferentes tipos de análise. São pretendidos métodos eficazes na análise de estatísticas singulares, mas eficazes também na visualização de combinações de estatísticas, suportando também combinações entre tipos de dados diferentes. Este objetivo foi estabelecido com o propósito de dar ao utilizador a possibilidade de descobrir relações entre os valores de diferentes estatísticas através da análise destes métodos de visualização, nomeadamente na descoberta de padrões nos gráficos, posicionamento de objetos representativos de cada registo nos referenciais (quando aplicável), ou análise das cores representativas de cada valor.

1.3 Estrutura da Dissertação

Este documento aborda os aspetos essenciais da implementação, como a arquitetura do motor de visualizações, análise de tecnologias a utilizar e análise dos gráficos implementados. E são abordados também aspetos relacionados com a revisão bibliográfica, como estudo do trabalho relacionado. O documento encontra-se dividido em seis capítulos principais:

No capítulo 2, é analisado o conceito de Visualização de dados, onde são apresentadas algumas definições segundo alguns autores. São abordadas taxonomias que permitem classificar métodos de visualização. Nestas taxonomias é focada a importância da análise dos dados e da forma como estes estão relacionados com o tempo, bem como outras características das visualizações que podem ser usadas como meio de classificação dos respetivos métodos, como as tarefas a executar, o seu domínio de aplicação e o foco nos dados e respetivas técnicas de mapeamento. São ainda referidos alguns dos métodos de visualização a adoptar para posterior integração no *Statz.me*.

No capítulo 3 são analisados os diferentes tipos de dados que são usados para a criação de estatísticas no *Statz.Me* e é descrita a arquitetura do motor de visualizações, analisando todos os

Introdução

componentes fundamentais, o seu funcionamento e a preparação dos dados para posterior representação gráfica, bem como a comunicação que é feita entre os vários componentes e as diferentes camadas.

No capítulo 4, são analisadas as bibliotecas de desenvolvimento de métodos de visualização utilizadas (Protovis e d3) e é justificada a sua utilização em detrimento de outras opções. É também feita uma comparação entre as características das tecnologias SVG (em que as bibliotecas utilizadas são baseadas) e Canvas, (outra tecnologia de desenvolvimento gráfico, que também foi considerada inicialmente).

No capítulo 5, são analisados os resultados do trabalho desenvolvido. É apresentada e justificada a utilização de cada um dos métodos de visualização desenvolvidos, os casos em que cada um é aplicado e os tipos de dados suportado. No fim, é feita uma exposição das suas características principais, analisando vários fatores relevantes no âmbito da visualização de dados. Neste capítulo é ainda feita ainda uma análise dos objetivos cumpridos comparando-os com os que foram de facto implementados.

No capítulo 6, são apresentadas as conclusões finais do trabalho realizado e são ainda expostas algumas ideias que poderiam valorizar a aplicação num eventual desenvolvimento futuro. São apresentados alguns métodos de visualização adicionais que permitem tipos de análise diferentes dos já implementados e que por isso acrescentariam um valor considerável à aplicação no que toca ao domínio da visualização de dados. É referida também a implementação de outras funcionalidades extra que poderiam ser acrescentadas e que se apresentam como casos de utilização interessantes no contexto das aplicações de registo de informação pessoal.

Introdução

Capítulo 2

Visualização de dados

2.1 Introdução

Fornecer métodos adequados de visualização de informação é uma questão fundamental no desenvolvimento de várias aplicações, e muitas vezes uma tarefa complexa. Requer um conhecimento completo do domínio em que a informação se insere, quais os objectivos que se pretendem concretizar com a visualização respetiva. Estes cuidados são essenciais para uma análise correta e eficaz do utilizador final.

A visualização de dados pode também ser vista como a transformação de dados de uma forma abstrata para um conjunto de informação organizada e devidamente contextualizada, ou num sentido mais metafórico, uma forma de contar uma história com os dados. [Yau11]. Segundo Friedman [Fri08], o objetivo principal da visualização de dados é comunicar a informação de forma clara e efetiva através de meios gráficos.

Para serem transmitidas as ideias de uma forma exata, o autor responsável pela visualização deve ter em conta aspetos funcionais (“dados em estado bruto”) e estéticos, de modo a ser conseguida uma perceção mais intuitiva de conjuntos de dados complexos.

É comum afirmar que a visualização de dados está intimamente relacionada com a visualização de informação e a análise de gráficos estatísticos. No entanto é uma área que aborda um domínio ainda mais abrangente, em termos de interdisciplinaridade, estando também relacionada com as áreas de interação humano-computador, extração de informação e *design*.

Conceitos que estão tipicamente relacionados com este assunto [FFu10]: a transmissão de informação, simplicidade, perceptibilidade, relevância e conhecimento, uma vez que são necessários por parte do autor cuidados com a forma, interesse, integridade e função dos dados a analisar, de forma a encontrar um equilíbrio entre tipos de visualização agradáveis para o utilizador final, mas ao mesmo tempo com conteúdo válido e objetivo. Esta é uma das principais dificuldades no desenvolvimento de métodos de visualização de dados atualmente. Muitos autores caem frequentemente na tentação de apresentarem gráficos ou outros métodos de visualização que procuram

Visualização de dados

chamar a atenção do utilizador final, mas que apresentam conteúdo redundante ou com pouco interesse dentro do contexto em que os dados se inserem.

Quando questionado sobre o conceito de visualização de dados, David McCandless [McC09] respondeu a esta questão com uma figura onde combinava os conceitos de interesse, função, forma e integridade. Esta definição de McCandless pode ser vista na figura 2.1.

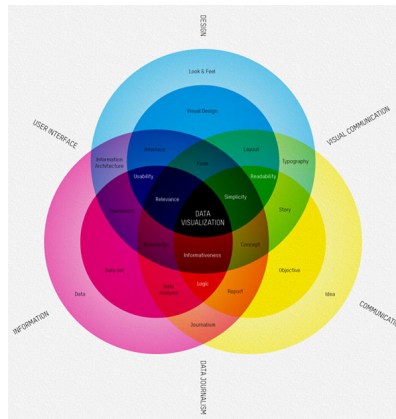


Figura 2.1: Definição de visualização de dados de David McCandless.

Mais recentemente, a comunidade especializada em interfaces e visualização de dados FFunction [FFu10] tentou simplificar a definição deste do conceito de visualização de dados, propondo uma mistura de apenas três conceitos: informação, design e comunicabilidade. A mistura de conceitos pode ser observada na figura 2.2.

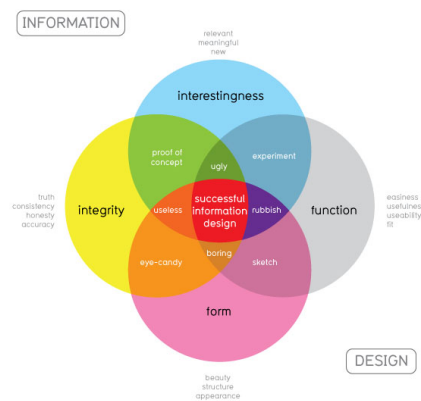


Figura 2.2: Esquema de definição de visualização de dados defendido pela comunidade FFunction.

Na figura 2.2, a visualização de dados surge do cruzamento de diferentes áreas como a comunicação, design e informação. Engloba três conceitos fundamentais:

- Look and feel ou aspeto, envolve o ponto de vista estético da visualização de dados.
- Os dados, são a informação propriamente dita que se pretende analisar.

- A ideia, retrata o objetivo da visualização. O propósito para o qual esta foi criada. Pode ser vista como a mensagem que se pretende passar para o utilizador.

Para uma abordagem correta no desenvolvimento de métodos de visualização, autores como Shneiderman *et al* [Shn96], Aigner *et al* [AMM⁺07b] [AMM⁺07a] e Lengler *et al* [LE07] abordaram diferentes taxonomias e modelos com diferentes perspetivas e que permitem uma melhor organização e definição de um vasto leque de técnicas de visualização que podem ser adaptados em conformidade com o tipo de dados, objetivos de análise ou perceções dos utilizadores alvo.

Numa abordagem de alto nível baseada nas perceções feitas através da análise de cada representação, é razoável afirmar que muitos métodos de visualização têm objetivos distintos, no que toca por exemplo à descoberta de padrões nos dados, perceção e interpretação de questões sociais/culturais [LV07], e realização de tarefas de análise dos dados, como por exemplo investigação de detalhes e extração de informação para posterior utilização [Shn96].

Existem já comunidades *online* que abordam continuamente novos métodos de visualização de informação, apresentando e analisando novas técnicas, novas estéticas, e divulgando aplicações criadas por terceiros que constituem um trabalho interessante no âmbito da visualização de dados, e em certos casos, também no âmbito das aplicações de registo de informação pessoal como é o caso do Statz.Me. [DV111], [IA111], [JE112].

Estas comunidades apresentam ainda várias abordagens interessantes sobre problemas comuns do domínio da visualização de informação. Um dos problemas mais abordados envolve a contextualização em que surgem os dados analisados e a forma como são representados visualmente, tendo em conta a perceção e as características cognitivas comuns dos utilizadores. Neste capítulo são abordadas estas questões demonstrando porque constituem por vezes uma dificuldade na criação de métodos de visualização e na forma como é demonstrada a informação contida nos dados em análise. São abordadas também modelos de classificação de técnicas de visualização que foram desenvolvidos abordando diferentes perspetivas. Posteriormente serão analisados também os métodos de visualização estudados.

2.2 Representação fiel e contextualização da informação

2.2.1 O problema das visualizações enganosas

Ao longo dos tempos têm surgido em variados domínios, representações de dados que distorcem a informação nelas contida levando muitas vezes o leitor a ter uma perceção errada da informação. Para Edward Tufte não há razões para crer que as representações visuais de informação não são vulneráveis a mentiras [Tuf86]. No foco da questão envolvendo representações de dados enganosas, está o facto de que os métodos de visualização são considerados como um dos meios principais para mostrar informação óbvia a alguém sem o mínimo conhecimento no domínio dos dados em questão, sendo por isso uma preocupação constante por parte dos autores.

Edward Tufte dá exemplos de gráficos falaciosos na perceção que passam para os leitores:

Visualização de dados

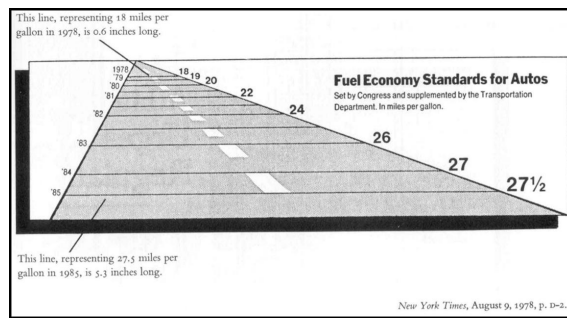


Figura 2.3: Exemplo de representação falaciosa de informação, proporções diferentes entre comprimento da linha e valor representado.

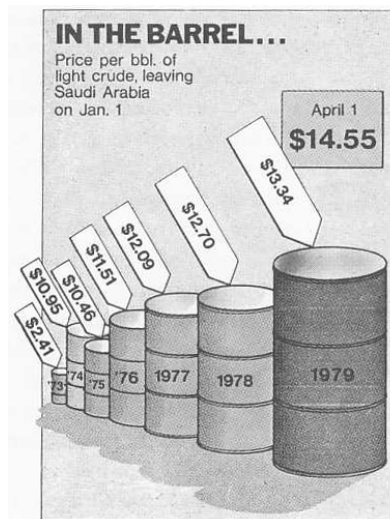


Figura 2.4: Exemplo de representação falaciosa de informação, proporções diferentes entre tamanho de cada barril e valor representado.

Nas figuras 2.3 e 2.4 podemos verificar exemplos de visualização que falham na sua função de darem uma percepção correta da informação ao utilizador uma vez que as figuras utilizadas em cada representação apresentam proporções erradas para os valores que representam. Na figura 2.3 a proporção entre o primeiro valor retratado (18) e o respetivo comprimento da linha que o representa, não é a mesma que é usada no caso da ultima linha para o valor a ela correspondente.

Este tipo de erro ocorre com maior frequência quando são utilizadas representações gráficas recorrendo a figuras do quotidiano. Embora a utilização destas apresente a vantagem de oferecer uma perspetiva mais intuitiva do objeto que é alvo de representação, e oferecer também uma visão mais atrativa dos valores retratados, tornam mais difícil para o leitor a percepção correta das diferenças entre os seus valores. Diferenças pouco acentuadas nos valores de cada elemento são praticamente impercetíveis ao olho do leitor, levando por vezes os autores a caírem na tentação de criarem representações incorretas.

2.2.2 Contextualização de informação

O domínio da contextualização da informação representada é também um ponto fundamental na criação de métodos de visualização.

Drew Skau apresenta uma análise interessante da importância deste fator [Dre12]. O autor defende que o argumento de que se os valores dos eixos estiverem corretos, então a visualização é correta, não é válido, uma vez que toda a informação relacionada com o gráfico deve ser visual e devidamente representada. Ou seja, é fundamental dar ao leitor a representação completa e íntegra dos dados que estão a ser tratados. As figuras 2.5 e 2.6 ilustram exemplos onde tal pode ser demonstrado.

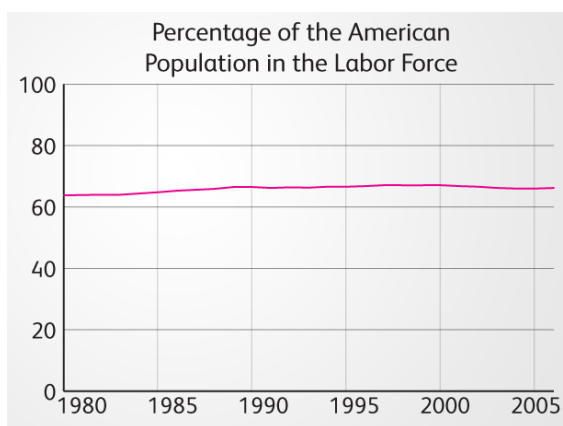


Figura 2.5: Representação original da percentagem de população americana pertencente à classe trabalhadora.

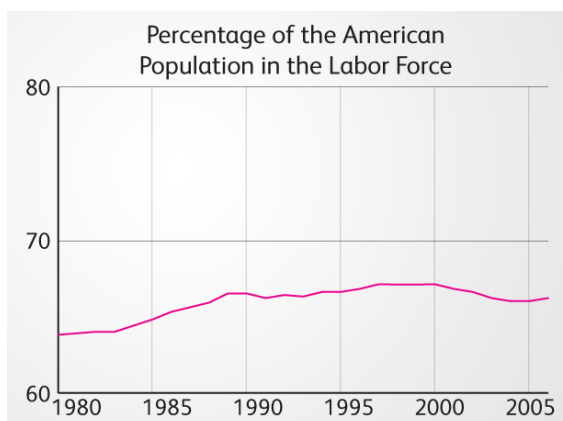


Figura 2.6: Representação alterada, com alteração do eixo vertical.

Nestes exemplos é mostrada a percentagem da classe trabalhadora da população americana. Fazendo um ajustamento do eixo vertical, tornam-se mais perceptíveis as variações dos valores ao longo do tempo, no entanto, se for tido em conta o facto de que a linha não é o único elemento

representativo informação, e que as áreas delimitadas por esta representam respetivamente a percentagem de população que pertence e não à classe trabalhadora, este ajuste vertical no gráfico acaba por destruir a perceção desta relação entre os valores, perdendo-se uma perceção que é importante na análise da informação.

A figura 2.7 mostra uma representação onde são ilustradas de forma mais destacada a referida relação entre os valores.

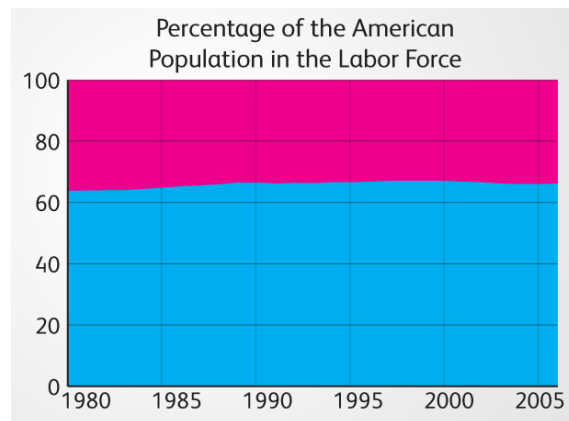


Figura 2.7: Representação das percentagens de população americana que pertence / não pertence à classe trabalhadora.

Este tipo de contextualização dos valores ajuda por vezes o leitor a ter uma perceção mais adequada dos dados, não sendo importante apenas em representações deste tipo, mas também na análise de outros tipos de dados, como por exemplo, dados quantitativos. Neste tipo de análises, é importante que os eixos comecem em zero, para que os valores da quantidade real representada possam ser visualizados corretamente.

O autor defende que esta contextualização da informação deve ser aplicada sempre que possível, uma vez que oferece uma visão mais clara e intuitiva dos dados tratados.

2.3 Sistemas de Classificação de Modelos de Visualização de Informação

A exploração de informação é uma tarefa que deve dispor de meios que permitam uma experiência agradável e intuitiva para o utilizador, que leve este a conseguir uma análise exata dos dados a visualizar e que consiga evitar o erro bastante comum de provocar uma “superabundância” de informação por vezes inútil e/ou redundante [Tuf86]. Evitar que isto aconteça não é uma tarefa fácil uma vez que quanto maior for a complexidade dos dados a analisar maior é a tendência de a visualização se tornar confusa ou pouco detalhada.

A visualização de informação é também uma área científica que tem vindo a crescer nos últimos anos, potenciada pelo desenvolvimento de novas tecnologias que permitem métodos de visualização mais atrativos e interativos para o utilizador. No entanto é uma área com domínio

ainda pouco estruturado e que envolve disciplinas bastantes díspares como a de interação humano-computador, *design* gráfico, gestão de informação, estatística e extração de conhecimento. Desta forma, havendo variados domínios paralelos, o avanço numa determinada área pode não ser acompanhado pelas restantes, pelo que é necessário, na definição de um método de visualização, uma integração racional das diferentes áreas, de forma a uma melhor estruturação do método propriamente dito. É também necessário ter em conta os requisitos indiretamente definidos pelo tipo de dados a analisar para ser possível o desenvolvimento de métodos que permitam a criação e partilha de conhecimento.

Existem variadas diretrizes no que toca a abordagem de métodos de visualização. No entanto, um princípio básico desenvolvido por Ben Shneiderman [Shn96], defende o seguinte: “Primeiro, um plano geral, *zoom* e filtragem a seguir; detalhes a pedido do utilizador”. O mesmo autor defende também a implementação de taxonomias orientadas às tarefas pretendidas pelo utilizador na exploração dos dados.

Uma fórmula geral utilizada por Ralph Lengler [LE07] na definição de métodos de visualização é: “Um método de visualização é uma representação gráfica, sistemática, baseada em regras que retracta informação de uma forma que leva à aquisição de perceções e a uma compreensão elaborada, ou através da comunicação de experiências”.

Existem já modelos teóricos que organizam o tipo de visualização de acordo com diferentes perspetivas. A *Taxonomia por Tarefas* abordada por Shneiderman *et al.* [Shn96] que organiza os diferentes métodos de visualização segundo as tarefas que estes permitem executar sobre os dados. A taxonomia por tipo de dados de Aigner *et al.* [AMM⁺07b], Bertone *et al.* [ABM⁺07] e Muller *et al.* [MS03] que analisa o tipo dos dados em questão bem como as características das variáveis temporais a que estes estão associados. A *Tabela Periódica dos Métodos de Visualização* de Lengler *et al.* [LE07] organiza os diferentes métodos tendo em conta o seu contexto, propósito, e tipo de representação. O modelo de Lau *et al.* [LV07] aborda uma integração entre uma perspetiva mais artística de visualização e uma perspetiva mais funcional, baseada nos dados propriamente ditos, criando um modelo de informação estética. Outro modelo, de Tory *et al.* [TM04], aborda uma classificação de métodos baseada em fatores de interdisciplinaridade como, por exemplo, a relação entre as expectativas do utilizador e as suposições do *designer* da visualização. As secções seguintes descrevem cada um destes modelos teóricos em pormenor.

2.3.1 Taxonomias por tarefas

Este sistema de classificação recorre à definição das tarefas que são executadas sobre os dados [Shn96]. Shneiderman identificou sete tarefas principais:

- **Visão Geral:** tem como objetivo proporcionar ao utilizador a perceção geral dos dados, dando-lhe também a opção de efetuar uma análise em detalhe. Uma das técnicas utilizadas recorre ao efeito de *fish-eye*, em que o utilizador navega ao longo do gráfico e na área seleccionada consegue obter uma visão mais detalhada sobre os dados. Semelhante a análise de um mapa com a utilização de uma lupa.

- **Zoom:** possibilidade de uma análise mais pormenorizada dos dados já que é comum os utilizadores terem interesse sobre uma determinada porção da coleção de dados. Efeitos de transição suaves de *zoom-in* e *zoom-out* ajudam na sensação de posição e contexto.
- **Filtragem:** tem como objetivo uma análise restringida apenas aos dados relevantes para o utilizador. Uma das chaves deste tipo de ação é o recurso a *queries* dinâmicas aplicadas aos dados.
- **Detalhes a pedido:** permite uma análise em detalhe de um determinado grupo de dados escolhido pelo utilizador. Uma forma de aplicar este tipo de tarefa poderia consistir na seleção de numa determinada área do gráfico que se pretende analisar com maior detalhe.
- **Relacionamento:** são criados meios que permitam, ao utilizador analisar determinados dados que possam estar relacionados. Como por exemplo ver os atores que contracenam em determinado filme. Este tipo de funcionalidades pode ser feito com a utilização de *queries* aos dados seleccionados, de forma a serem pesquisados no *dataset* atributos que possuam informação relacionada.
- **Histórico:** uma vez que o processo de visualização dos dados requer vários passos, é útil manter um registo de todas as ações efetuadas e dessa forma, permitir operações como retroceder, repetir e o refinamento das mesmas ações.
- **Extração:** permite que determinados dados/resultados sejam extraídos para um ficheiro num formato que permita a sua reutilização noutros domínios, como apresentações, impressões em papel, etc.

2.3.2 Taxonomias por caracterização de dados/tempo

Aigner *et al.* [AMM⁺07b], Bertone *et al.* [ABM⁺07] e Muller *et al.* [MS03] apresentaram sistemas de classificação de técnicas de visualização de informação orientadas à natureza temporal dos dados. No trabalho destes autores são analisadas as diferentes, características temporais dos dados, de forma a serem construídos métodos de visualização compatíveis com essas mesmas características.

2.3.2.1 Caracterização do tempo

Quando se tratam dados orientados ao tempo é estritamente necessário, para a adoção de métodos de visualização eficazes e auto-explicativos uma análise às diferentes características temporais adjacentes ao problema em questão [LAB⁺11]. Este pode ser analisado tendo em conta os seguintes aspetos:

- **Foco: tempo por intervalo vs tempo por pontos:** No que toca ao foco da variável do tempo, esta pode ser definida baseada em pontos, em que cada valor corresponde a apenas um momento no tempo, em que mais nada se sabe sobre o intervalo entre dois pontos. Por

exemplo: valores de abertura e fecho da bolsa). Um foco orientado a intervalos implica uma abordagem contínua do eixo temporal, isto é, implica a existência de durações como dias, meses etc.

- **Escala: ordinal vs discreto vs contínuo:** Uma abordagem temporal discreta descreve o tempo como ocorrências isoladas e independentes, instâncias sem duração e são apenas representados na forma de pares <ponto-temporal, valor>. Por outro lado, uma abordagem por intervalos implica a utilização de um eixo temporal explicitamente escalado (por exemplo por dias, meses anos etc.) e em que cada evento possui uma determinada duração. Uma abordagem ordinal ocorre quando são aplicados termos como “X ocorre antes de Y”, ou seja, quando existe uma sequência de eventos passíveis de serem relacionados temporalmente.
- **Estrutura: cíclica vs linear:** Uma representação linear assume um ponto inicial e um eixo temporal disposto desde o passado até ao presente (e futuro em caso de sistemas que implementem sistemas de predição). Este tipo de abordagem permite uma análise contínua dos dados, fornecendo ao utilizador uma perspetiva histórica de toda a informação [SC00]. Por sua vez uma estrutura cíclica tem como objetivo uma representação mais intuitiva de eventos que ocorram periodicamente, como por exemplo as estações do ano. A ordem dos acontecimentos nesta abordagem não tem significado relevante uma vez que a ordem dos acontecimentos é ambígua. Por exemplo, o verão vem depois da primavera, mas a primavera também vem depois do verão. Uma adaptação do intervalo temporal pode levar a interpretações da eventos periódicos de uma forma mais direta, uma vez que os itens destacados na visualização serão dispostos na mesma fatia do gráfico circular [WAM01], [CK98].
- **Ponto de vista: ordenado vs ramificado vs perspetivas múltiplas:** Os eixos temporais ordenados consideram eventos que se seguem um após o outro, por outras palavras, eixos ordenados permitem afirmar que o elemento X_i sucede o elemento X_{i-1} e precede o elemento X_{i+1} . Os eixos temporais ramificados podem ser usados em cenários em que determinadas sequências de ações podem ser previstas, criando uma visualização com múltiplas alternativas, dependendo de decisões futuras. Esta abordagem permite situações em que um determinado elemento A_x pode ser seguido pelos elementos A_{x+1} ou por A_{x+1} . Esta abordagem suporta ainda processos de tomada de decisões. Finalmente, os eixos temporais de perspetivas múltiplas permitem observar mais do que um elemento para cada instância temporal, dando assim a possibilidade de representação de eventos paralelos. (Por exemplo relação entre duas variáveis).
- **Granularidade: nula vs única vs múltipla:** A granularidade representa as frações em que pode ser dividido o tempo (minutos, horas, dias etc.). Pode ser nula, em que não está definida, ou por outras palavras, em que não há um padrão definido para a periodicidade de cada nova instância do elemento a analisar. A representação da granularidade pode também

ser única, quando os elementos são representados com (apenas) uma periodicidade definida, ou múltipla em que são mostrados os valores adaptados a cada periodicidade diferente.

2.3.2.2 Modelação dos dados

Depois de analisadas as características do domínio temporal, é também necessária uma análise das características dos dados a visualizar, de modo a ser efetuada uma modelação apropriada para os dados em questão. As características dos dados são analisadas de seguida.

- **Número de variáveis:** cada elemento a analisar pode ser representado com recurso a mais do que uma variável (por exemplo: na meteorologia, em que são analisadas as variáveis: temperatura, humidade, nível de precipitação etc.), ou apenas por uma única variável representativa do elemento que se pretende analisar (por exemplo: velocidade de um automóvel). No caso de existirem múltiplas variáveis estas podem estar correlacionadas, influenciando os seus valores. (por exemplo: o nível de precipitação tem influência sobre o valor da humidade).
- **Enquadramento: espacial vs abstrato:** Os dados a visualizar podem possuir um contexto espacial, isto é, dependerem de uma determinado contexto sobre o qual ocorre a ação, ou possuir um contexto abstrato, isto é, não estão associados a nenhum contexto em específico.
- **Tipos de dados: eventos vs estados:** Os tipos de dados podem ser classificados como estados, quando o valor da variável em observação é feito em passos discretos ao longo do tempo (por exemplo: semáforo). São definidos como eventos quando ocorre algum acontecimento que leva à alteração dos valores dos dados em análise (por exemplo: eventos políticos que alteram valor dos mercados bolsistas).

Esta taxonomia é utilizada na *framework* conceptual abordada por Aigner *et al.* [ABM⁺07], de forma a auxiliar a criação e classificação de técnicas de visualização mais adaptadas às características não só dos dados propriamente ditos, mas também da dimensão temporal. O autor defende que a realização de uma análise prévia baseada numa taxonomia pode trazer melhoramentos consideráveis em sistemas de simulação que lidam com dados relacionados com o tempo.

2.3.3 Tabela Periódica dos métodos de visualização

Um pouco à semelhança das taxonomias aplicadas aos dados e à modelação da variável temporal, os próprios métodos de visualização necessitam de alguma análise e categorização tendo em conta características como o seu domínio de aplicação, complexidade e nível de detalhe. Ralph Lengler apresenta um sistema para a categorização de métodos de visualização, organizando-os por diferentes categorias, dimensões e pela estratégia de visualização aplicada, de forma análoga à organização dos diferentes elementos químicos na tabela periódica [LE07].

Esta representação, ilustrada na figura 2.8 permite ainda uma análise relativamente à complexidade de visualização e a área de aplicação de cada método uma vez que a tabela é organizada

A PERIODIC TABLE OF VISUALIZATION METHODS

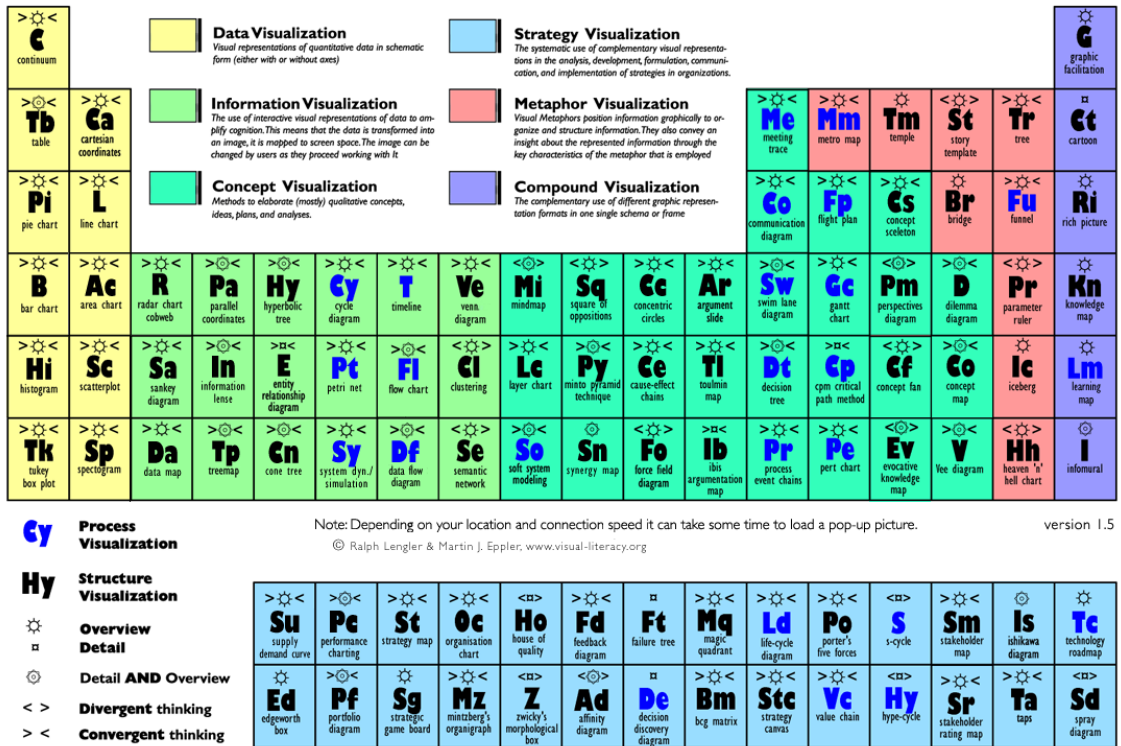


Figura 2.8: Tabela Periódica dos Métodos de Visualização.

segundo estes parâmetros da mesma forma que a tabela periódica está organizada em grupos e períodos. Nesta abordagem a complexidade de visualização corresponde aos períodos e a área de aplicação aos grupos. Os seis “grupos” principais pelos quais são divididos os diferentes métodos são:

- **Visualização de dados:** inclui formatos quantitativos como gráficos de barras, circulares ou de linhas. Sendo por isso a área de aplicação mais comum.
- **Visualização de informação:** engloba redes semânticas ou mapas de árvore, é normalmente associado à utilização de representações interativas de forma a permitir uma maior transmissão de conhecimento.
- **Visualização de conceitos:** utilizado como mapa conceptual. Informação é apresentada de forma esquemática, com ligações representando relações entre os diferentes conceitos (normalmente representados com círculos ou caixas).
- **Visualização metafórica:** possui a característica de familiarizar o utilizador com o conteúdo dos dados em análise. Requer uma diferente estruturação de informação. Esta própria representação de métodos de visualização recorrendo a uma tabela periódica pode ser considerada uma visualização metafórica.

- **Visualização de estratégia:** é uma abordagem que permite uma melhor análise e desenvolvimento de estratégias em organizações. É útil na tomada de decisões e nas previsões de resultados.
- **Visualização composta:** pode ser vista como uma mistura das abordagens anteriores. Pode envolver mapas conceituais com a utilização de gráficos circulares, ou tipos de visualização metafóricas de uma perspectiva interativa.

Além destes grupos, existem ainda outras dimensões que é conveniente ter em consideração no agrupamento dos métodos:

- **Dependência de tarefas:** dependendo das tarefas pretendidas a visualização pode realçar certos aspetos dos dados. Este domínio pode ser dividido em três tipos principais: focado num plano geral, focado nos detalhes, ou misto (segundo a abordagem de Ben Shneiderman: “plano geral em primeiro lugar, *zoom* e filtragem de seguida, detalhes a pedido”)[Shn96].
- **Processo cognitivo:** é definido pelo tipo de reflexão que o método provoca no utilizador, pode ser definido em dois tipos:
 - Convergente, em que é provocada uma redução de complexidade na análise dos dados.
 - Divergente, processo oposto, em que é aumentada a complexidade da análise e em que o utilizador pode obter perceções mais personalizadas relativamente a uma determinada questão.
- **Informação representada:** pode ser definida em dois tipos:
 - Estrutural: hierarquias ou redes
 - Processo: cíclica, contínua ou sequencial

Outra característica desta organização dos diferentes métodos de visualização é uma percepção de que pode não existir apenas um método apropriado para um caso em específico. Pelo contrário, a combinação de diferentes técnicas pode levar a um realce dos resultados pretendidos.

Com esta analogia Lengler *et al.* consegue de uma forma intuitiva mostrar que os vários métodos podem ser combinados de forma a realçarem os resultados, bem como a indicação das principais características de cada um, de forma a que os investigadores destes métodos possam aplica-los de forma mais adequada às suas necessidades.

2.3.4 Estética de informação

Outro modelo de categorização de métodos de visualização é o apresentado por Lau *et al.* [LV07], orientado à estética de informação. A representação estética pode ser vista como a interseção entre três pontos fundamentais da visualização de informação: os dados, a estética, e a interação. Estas características podem ser analisadas de forma mais clara na figura 2.9:

Visualização de dados

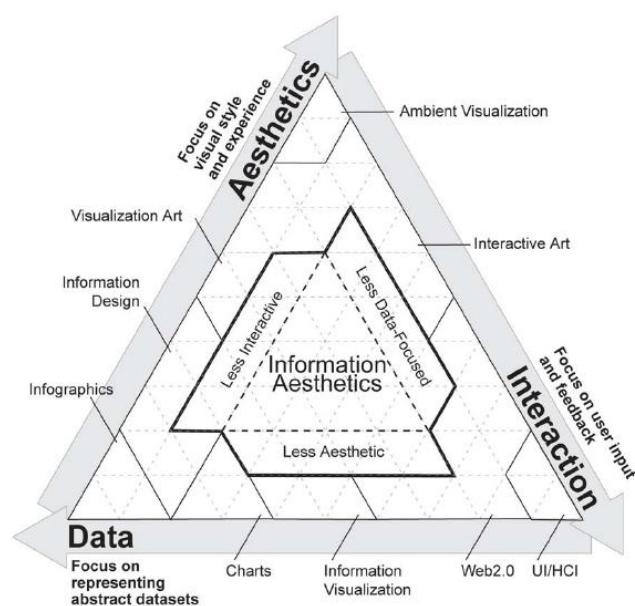


Figura 2.9: Diagrama de definição de estética de informação.

Na figura 2.9 pode ser visto que a estética de informação consiste, então, na aplicação de métodos de representação de dados abstratos (visualização de informação), na utilização de mecanismos que permitam uma interface interativa (arte interativa) e na adoção de meios de representação atrativos para o utilizador (visualização artística).

Este modelo tem por objetivo transmitir uma ideia mais exata da integração entre as áreas de visualização artística e de visualização de informação. Envolve fatores de foco nos dados, essenciais na visualização informativa, e de foco nas técnicas de mapeamento, relevantes para o domínio da visualização artística. O foco nos dados é determinado através da análise sobre em que aspetos a visualização permitirá a aquisição de conhecimento, e é baseada em observações mais objetivas e funcionais. Por sua vez, o foco nas técnicas de mapeamento concretiza-se através de observações feitas com o objetivo de analisar quais os métodos de representação usados para mapear os dados em informação visual, ou por outras palavras, o processo de conversão dos valores dos dados em representações visuais.

- **Foco nos dados:** O foco nos dados pode ser feito de duas formas distintas. De forma intrínseca com o objetivo de facilitar a compreensão dos dados e permitir um mapeamento visual que permita uma boa análise cognitiva. Este foco nos dados de forma intrínseca pode ser visto como ferramenta de auxílio para a execução de tarefas e disseminação de informação por parte dos utilizadores. Permitindo a descoberta de certos padrões nos dados. A focagem dos dados de forma extrínseca facilita uma exteriorização de significado que está relacionado com os dados analisados. Esta abordagem tem como objetivos principais a apreciação, interpretação e reflexão pessoal. Métodos de visualização mais orientados a aspetos artísticos são habitualmente sinónimos de focagem extrínseca do significado dos

dados, e permitem uma percepção de mais alto nível nomeadamente em âmbitos sociais ou culturais.

- **Técnicas de mapeamento:** No que toca a técnicas de mapeamento, estas podem ser divididas em dois tipos. Diretas, que são geralmente conduzidas por standards de cognição visual como por exemplo as regras de Gestalt, e de psicologia da percepção. A utilização de técnicas de mapeamento sistemáticas é dominante em tipos de visualização orientadas a representações diretas. Outra característica das técnicas de mapeamento direto é que estas são reversíveis, ou seja o utilizador pode inferir os valores dos dados a partir da representação visual destes. Por outro lado as técnicas de mapeamento interpretativas envolvem influências estilísticas e visões mais subjetivas. O *design* da visualização pode derivar de influências multidisciplinares. Este mapeamento subjetivo não permite a possibilidade de os dados serem inferidos na análise da representação, ou seja os utilizadores tem maior dificuldade em compreender os valores dos dados a partir da representação.

Este modelo proposto por Andrea Lau *et al.* demonstra que as técnicas de mapeamento e de foco nos dados estão correlacionados, isto é, a escolha de uma determinada técnica de mapeamento geralmente determina de imediato qual o tipo de foco de dados resultante (e vice-versa). Técnicas de visualização baseadas em mapeamentos diretos geralmente envolvem um foco intrínseco, enquanto mapeamentos interpretativos levam a um foco extrínseco dos dados. Uma análise mais aprofundada mostra que estes fatores podem ser identificados como os campos principais da visualização informativa e visualização artística, havendo no entanto um espectro de técnicas de visualizações mais alargado entre estes dois tipos, estando a definição de representação estética situada entre eles.

2.3.5 Percepção do utilizador vs suposições do designer

O modelo proposto por Tory *et al.* [TM04] aborda os conceitos relacionados com as suposições e percepções dos dados, por parte dos *designers*, e as expectativas e interpretação dos mesmos, por parte dos utilizadores. Abordando a perspectiva do *designer* surge o conceito de modelo de *design* que engloba as suas suposições sobre os dados e respetivos algoritmos para o seu desenho. Da perspectiva do utilizador surge o conceito de modelo de utilizador, que engloba ideias pré-concebidas sobre a visualização dos dados propriamente ditos e a sua interpretação.

Uma vez que o mesmo utilizador pode ter percepções diferentes do mesmo modelo em alturas diferentes, para uma boa definição de um modelo de utilizador são necessárias suposições sobre os dados, desenvolvendo hipóteses sobre estes e procurando evidências nos mesmos que suportem estas hipóteses. Estes modelos podem ser refinados com recurso a interações sobre os dados através de outros métodos de visualização. Esta interação permite uma ideia mais exata dos dados, e conseqüentemente um melhor suporte ao desenvolvimento de hipóteses.

Por seu lado, a classificação de modelos de *design* é feita segundo uma taxonomia que se baseia na classificação do tipo de dados como discretos ou contínuos, e de outros atributos adjacentes ao

algoritmo de representação. Esses atributos podem ser características espaciais (que funcionam como uma contextualização dos dados), características do tempo, cor e transparência.

As características espaciais podem ser dadas, isto é, definidas por determinadas restrições dos dados, ou escolhidas pelo *designer*, adaptando-as conforme as suas suposições dos dados e das tarefas a que estes vão estar associados.

Dentro das classificações de dados discretos, há ainda a considerar os que detêm caráter estrutural e os que representam valores. Por seu lado os tipos de dados contínuos podem ser escalares, vetoriais ou com múltiplas variáveis.

Os autores abordaram ainda que tarefas de visualização são permitidas por este modelo:

- Modelos estruturais discretos, que permitem uma análise das relações de conexão (por exemplo: “relações pai/filho”);
- Modelos baseados em valores discretos, que permitem uma análise de padrões, como identificação de clusters, outliers etc.;
- Modelos discretos, que permitem aos utilizadores um estudo dos detalhes de cada item bem como a filtragem de dados;
- Modelos contínuos, que dão ao utilizador a possibilidade de estudar tendências numéricas como o aumento ou diminuição do valor de determinadas variáveis.

Em síntese, o modelo proposto por Tory *et al.* permite uma análise aos métodos de visualização seguindo uma comparação entre as perceções esperadas pelo utilizador e as suposições do *designer*. Os modelos que seguem as perceções do *designer* são denominados modelos de *design* e são classificados tendo em consideração restrições dos meios de representação mas, principalmente, através da classificação do tipo de dados entre contínuos e discretos, e para cada um deste tipo de dados são organizados os métodos/técnicas de visualização mais apropriadas.

Existem ainda outras taxonomias que são usadas como sistemas de classificação de métodos de visualização. Uma taxonomia abordada por [Chi00] analisa o processo de criação de uma visualização em quatro partes fundamentais: obtenção de valores, abstração analítica, abstração de visualização e por último a fase de visualização. Para cada transição destas fases está associado processo de transformação de dados (entre as duas primeiras fases) e um processo de mapeamento visual (entre a segunda e a terceira fase, e entre as duas últimas fases). O autor define também tarefas de análise a aplicar a cada uma das fases, como por exemplo a normalização de valores (transformação de dados, entre as duas primeiras fases) e mapeamento dos valores recorrendo a diferentes técnicas (última fase).

2.4 Métodos de visualização

Depois do estudo dos sistemas de classificação dos vários métodos de visualização existentes, são apresentados de seguida modelos de visualização que foram estudados, a sua relação com o

tempo e respetiva classificação segundo os cinco sistemas analisados anteriormente. Neste capítulo não são considerados os métodos mais comuns como gráficos de barras, linhas, ou gráficos de áreas uma vez que os seus propósitos de visualização são já senso comum, e por isso não é considerada necessária uma análise aprofundada a estes métodos. Apesar de não serem abordados nesta secção, estes métodos são, no entanto, considerados na implementação dos métodos de visualização no *Statz.Me* uma vez que, apesar de básicos, são métodos bastante eficazes no domínio da visualização de dados.

2.4.1 Métodos de visualização dependentes do tempo

Dentro dos métodos de visualização dependentes do tempo temos:

- **Spiral Graph:** método ilustrado na figura 2.10 e utilizado na análise de eventos cíclicos. Com a adaptação do intervalo de tempo, permite uma melhor percepção da periodicidade dos eventos [AMM⁺07b] [MS03].
- **Theme River:** método ilustrado na figura 2.11 e que analisa várias variáveis na mesma visualização. O valor de cada variável é ilustrado como o caudal de um rio. Permite uma boa análise da evolução dos dados ao longo do tempo. [AMM⁺07b] [MS03]
- **Calendar View:** ilustrado na figura 2.12, este método atribui uma determinada cor a cada dia de um calendário. Permite uma boa interpretação dos valores dos dados durante o ano, ou do período que se pretenda analisar [MS03].
- **Planning Lines:** ilustrado na figura 2.13 esta visualização semelhante a diagramas de Gantt, é frequentemente utilizada para planeamentos e eventos sequenciais temporalmente ordenados. Permite uma boa percepção da ordem dos eventos e da respectiva duração [AMM⁺07b].
- **Perspective Wall:** ilustrado na figura 2.14, é um método que possui um conceito semelhante ao de *PlanningLines* mas com a possibilidade de navegação temporal. Método recorre a uma distorção da visualização dos dados passados e futuros de forma a dar ao utilizador a ideia de perspectiva [SC00].
- **Gas and Driving:** ilustrado na figura 2.15, é baseado na aplicação de um método de visualização que relaciona o preço do combustível com a média da distância percorrida em milhas pela população dos Estados Unidos da América. O recuo da linha do gráfico ilustra a diminuição da distância percorrida, pelo que uma leitura da esquerda para a direita nem sempre implica uma ordem cronológica [GD112].
- **Candlestick:** ilustrado na figura 2.16, é utilizado para ilustrar os valores iniciais e finais de uma determinada variável dentro de um período de tempo. A cor de cada barra é utilizada para a comparação com os valores finais registados no período de tempo anterior. Pode ser utilizado como ferramenta de análise para a bolsa de valores. [C1212].

2.4.2 Métodos de visualização temporalmente independentes

- **Chord diagram:** permite o relacionamento entre várias entidades na mesma visualização. Na figura 2.17 pode ser visto um exemplo da utilização deste diagrama, onde são analisados os fluxos monetários de doadores privados para os diferentes partidos políticos alemães [GPD12].
- **Bubble Charts:** ilustrado na figura 2.18, é um método que pode relacionar entre duas a quatro variáveis distintas recorrendo a coordenadas no referencial e a diferentes áreas e cores. Permite uma rápida interpretação de todas as variáveis em análise [BC112].
- **Bars of stuff:** ilustrado na figura 2.19, é uma visualização metafórica que recorre à utilização de figuras que definem de forma mais intuitiva os dados em análise. Os valores dos dados são definidos com o tamanho de cada figura. [BS112].
- **Bullet Charts:** ilustrado na figura 2.20, é uma visualização que permite a categorização de valores (exemplo: mau, suficiente, bom) e a comparação com o último valor registado (marca triangular da figura) [Bul12].

Estes métodos podem ser classificados segundo as taxonomias abordadas na secção 2.3 como pode ser visto nas tabelas 2.1, 2.2 e 2.3. Cada método está classificado à luz das características fundamentais das taxonomias estudadas, sendo apresentadas para cada um as características que são alvo da classificação segundo os sistemas estudados, como as tarefas de interação permitidas, estudadas por Shneiderman *et al.*, área de aplicação (grupo da tabela periódica) definidas por Lenglet *et al.*, características do foco de dados e do mapeamento abordadas por Lau *et al.*, e características de dados que são alvos das perspetivas do *designer*, características estudadas por Tory *et al.*.

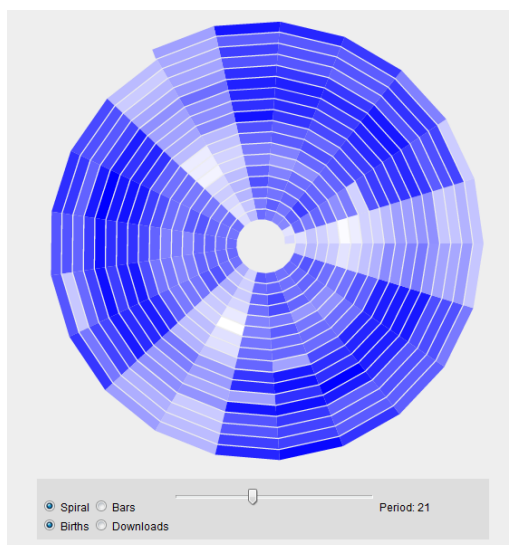


Figura 2.10: Spiral Graph.

Visualização de dados

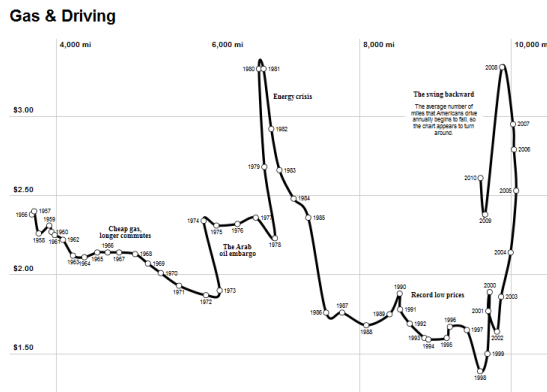


Figura 2.15: Gas and Driving.

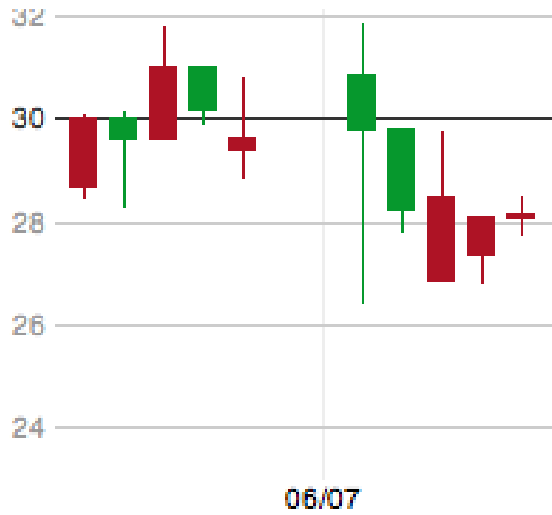


Figura 2.16: Candlestick.

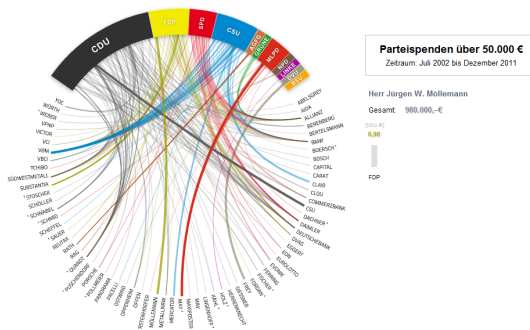


Figura 2.17: Chord Diagram.

Visualização de dados

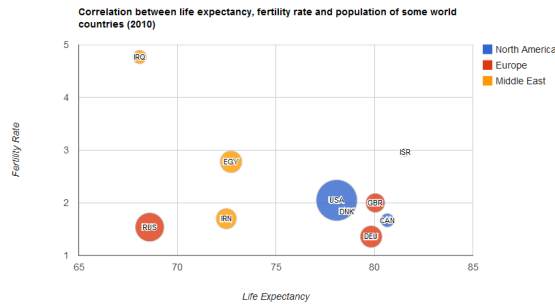


Figura 2.18: Bubble Chart.



Figura 2.19: Bars of Stuff.

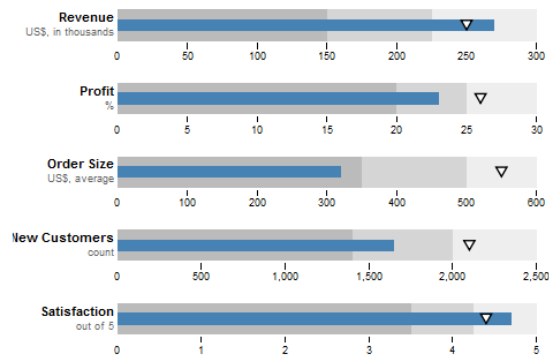


Figura 2.20: Bullet Charts.

2.5 Conclusões

Neste capítulo foram abordadas definições básicas do conceito de visualização de dados, tendo sido identificados os componentes principais que constituem o domínio desta área como, por exemplo, o aspeto da visualização, a integridade dos dados representados e o propósito ou a mensagem que se pretende transmitir ao utilizador com uma determinada visualização de dados.

Na secção 2.2 foi feita uma análise a um dos problemas principais na criação de métodos de visualização, que consiste no desenvolvimento de gráficos com potenciais leituras incorrectas, e analisadas formas de como estes erros surgem nos variados tipos de visualização: muitas vezes por

Visualização de dados

Tabela 2.1: Classificação dos métodos de visualização segundo os sistemas analisados

	Taxonomia por tarefa	Grupo - Tabela Periódica	Estética de informação (foco de dados /mapeamento)	Perspectiva designer
Spiral Graph	Visão Geral	Visualização de informação	Intrínseco (descoberta de padrões) / Direto (standards de cognição visual)	Dados discretos (valores), aplicada mesma cor com diferentes tonalidades conforme valor dos dados
Theme River	Visão Geral	Visualização de informação	Intrínseco (perceção de tendências) / Direto (standards de cognição visual)	Dados contínuos, aplicadas diferentes cores aos dados, limitação de espaço para dados situados no meio do gráfico
Calendar View	Visão Geral	Visualização de informação	Intrínseco (descoberta de padrões) /Direto (aplicáveis standards de cognição visual)	Dados discretos (valores), aplicadas diferentes cores aos diferentes valores dos dados
Planning Lines	Visão Geral, Zoom	Visualização conceptual	Intrínseco (descoberta de padrões) / Direto (aplicáveis standards de cognição visual)	Dados contínuos, comprimento das barras define duração de eventos.
Perspective Wall	Visão Geral, Navegação temporal, Zoom, Detalhes a pedido	Visualização conceptual	Intrínseco (descoberta de padrões) / Direto (aplicáveis standards de cognição visual)	Dados contínuos, comprimento de barras define duração de eventos, distorção de imagem para efeito de perspectiva
Gas and Driving	Visão Geral	Visualização de dados	Extrínseco(perceção da relação das diferentes variáveis ao longo do tempo) / Interpretativo (permite interpretação de informação e sua associação a eventos que possam ter influenciado os valores)	Dados contínuos

tentação dos autores em darem uma perceção mais evidente das diferenças de valores representadas. Foi também analisada a importância da contextualização dos dados no processo de criação de cada novo método de visualização, e demonstrados cuidados a ter neste aspeto como, por exemplo, em casos de análises de dados quantitativos, a representação dos eixos de cada gráfico que deve ser iniciada em zero, de forma a ter-se uma ideia mais fiel da quantidade real.

Na secção 2.3 foi feita uma análise a várias taxonomias que são utilizadas para a classificação de variados métodos de visualização abordando diferentes perspetivas.

- **Taxonomias por tarefas** - em que os métodos são classificados com base no tipo de tarefas que permitem efetuar sobre os dados em análise. Como por exemplo, obter uma visão geral da informação retratada, operações de *zooming*, navegação temporal para uma perspetiva mais detalhada ou operações de obtenção de detalhes a pedido do utilizador.
- **Taxonomias por caracterização de dados/tempo** - onde a classificação dos métodos de visualização é feita com base nas características dos dados e nas respetivas características temporais destes. Quanto aos dados as características envolvidas são o número de variáveis; ponto de referência – pode ter um contexto espacial, ou abstrato – e os tipos de dados, que podem ser eventos ou estados. No que toca às características temporais estas são divididas em:

- **Escala** - valores registados podem ter um contexto temporal contínuo ou discreto;

Visualização de dados

Tabela 2.2: Classificação dos métodos de visualização segundo os sistemas analisados - continuação

	Taxonomia por tarefa	Grupo - Tabela Periódica	Estética de informação (foco de dados /mapeamento)	Perspectiva designer
Candlestick	Detalhes a pedido	Visualização de dados	Intrínseco (identificação de tendências) / Direto (dados podem ser inferidos a partir da visualização)	Dados discretos (valores), desenho da visualização recorre a uma linha que define os valores máximos e mínimos registados, e uma caixa que define valores de abertura e encerramento.
Chord Diagram	Filtragem, Detalhes a pedido	Visualização de informação	Intrínseco (apenas possui ligações diretas entre dados) / Direto (aplicável cognição visual nas ligações entre dados)	Dados discretos (estruturais), ligação entre os diferentes dados representada com cores e áreas diferentes
Bubble Charts	Visão Geral, Detalhes a pedido	Visualização de dados	Extrínseco (permite interpretação e transmite significado) / Direto (aplicável standards de cognição visual - perceção de áreas)	Dados discretos (valores), dados diferentes representados com cores diferentes, valores representados recorrendo a áreas e posições no referencial
Bars of Stuff	Visão Geral	Visualização metafórica	Intrínseco (valores dos dados representados pelo tamanho das figuras) / Direto (aplicáveis standards de cognição visual - comparação dos tamanhos das figuras)	Dados representados com recurso a figuras
Bullet Charts	Visão Geral	Visualização de dados	Intrínseco (valores definidos com comprimento de barras) / Direto (aplicáveis standards de cognição visual - comparação dos tamanhos das barras)	Aplicadas diferentes cores às respectivas características qualitativas dos valores.

– **Foco temporal** - define se os dados são registados com base em pontos ou intervalos de tempo;

- **Estrutura** - Dados podem ser analisados linearmente, ou de forma cíclica que pode levar a diferentes conclusões;
 - **Pontos de vista** - Dados podem ser analisados de forma ordenada, ou tendo em conta diferentes perspectivas, o que no último caso envolve sequências de eventos;
 - **Granularidade** - envolve a periodicidade do registo de cada ocorrência. Pode ser nula - no caso de não haver uma periodicidade definida; múltipla – no caso de a análise temporal puder ser feita com diferentes intervalos de tempo; e única – em que o intervalo de tempo é sempre o mesmo para toda a visualização.
- **Classificação com base na Tabela Periódica dos métodos de visualização** - nesta taxonomia a classificação é feita recorrendo essencialmente a um agrupamento dos métodos pelo tipo de visualização a que pertencem, em que os vários tipos podem ser: visualização de informação, visualização de dados, visualização de conceitos, visualização metafórica, visualização composta, visualização estratégica.
 - **Classificação com base na estética de informação** - esta taxonomia baseia-se essencialmente na relação entre três pontos fundamentais no âmbito da visualização: os dados, a estética e a interação. É uma taxonomia que tem como objetos de estudo o foco nos dados e as técnicas de mapeamento utilizadas, e estabelece uma relação entre estas duas variáveis. Esta taxonomia demonstra que as técnicas de mapeamento e de foco nos dados estão correlacionados. As técnicas de visualização baseadas em mapeamentos diretos (conduzidas por standards de cognição visual) geralmente envolvem um foco de dados intrínseco (facilita a compreensão dos dados e permite boa análise cognitiva), enquanto mapeamentos interpretativos (envolvem influências estilísticas e visões mais subjectivas) levam a um foco extrínseco dos dados (exteriorização de significado que está relacionado com os e permite a apreciação, interpretação e reflexão pessoal sobre a informação).
 - **Classificação com base nas percepções de utilizador vs percepções do *designer*** - aborda os conceitos relacionados com as suposições e percepções dos dados, por parte dos *designers*, e as expectativas e interpretação dos mesmos, por parte dos utilizadores. A perspectiva do *designer* engloba as suas suposições sobre os dados e respetivos algoritmos para o seu desenho. Por outro lado, da perspectiva do utilizador surge o conceito de modelo de utilizador, que engloba ideias pré-concebidas sobre a visualização, o contexto dos dados propriamente ditos, e a sua interpretação.

Por último, foram apresentados métodos de visualização que permitem diferentes tipos de análise de dados. Na sequência, foi apresentada a classificação desses métodos com base nas taxonomias estudadas. Alguns foram tidos como ponto de partida para outros métodos, aplicados no Statz.Me.

Tabela 2.3: Classificação dos métodos de visualização por tipo de dados e características temporais

	Características temporais					Características de dados		
	Escala	Foco tem- poral	Estrutura	Pontos-de- vista	Granularidade	Nº de variáveis	Referência	Tipos de da- dos
Spiral Graph	Discreto	Pontos	Cíclico	Ordenado	Única	1	Abstrato	Estados
Theme River	Contínuo	Intervalos	Linear	Ordenado	Nula	Múltiplas variáveis	Abstrato	Eventos
Calendar View	Discreto	Pontos	Linear	Ordenado	Única	1	Abstrato	Estados
Planning Lines	Contínuo	Intervalos	Linear	Múltiplas perspetivas	Nula	Múltiplas variáveis	Abstrato	Eventos
Perspective Wall	Contínuo	Intervalos	Linear	Múltiplas perspetivas	Nula	Múltiplas variáveis	Abstrato	Eventos
Candlestick	Discreto	Pontos	Linear	Ordenado	Nula	4	Espacial	Estados
Gas and Driving	Contínuo	Intervalos	Linear	Ordenado	Única	2	Espacial	Eventos
Chord Diagram	Não aplicável					Múltiplas variáveis		
Bubble Charts	Não aplicável					Até 5 variáveis		
Bars of stuff	Não aplicável					1		
Bullet Charts	Não aplicável, apenas comparável com ocorrência anterior					1		

Capítulo 3

Tipos de dados e arquitetura

Neste capítulo são analisados os tipos de dados suportados pela aplicação para a criação das estatísticas, bem como a arquitetura do motor de visualizações do Statz.Me, o funcionamento de cada componente e a sua comunicação com os dados da aplicação.

3.1 Tipos de dados do Statz.Me

De modo a englobar um vasto número de possibilidades para a criação de estatísticas são considerados vários tipos de dados. Assim, foram considerados sete tipos diferentes:

- **Numeric:** tipo de dados representativo de quantidades/dimensões, sendo por isso representado com valores decimais. Na sua representação gráfica, todas as atualizações que tenham a mesma data são agrupadas no mesmo registo, com o intuito de ser representado o valor total verificado em cada dia. Exemplos: "distância percorrida", "combustível abastecido", "peso", "altura"
- **Rating:** tipo de dados utilizado em estatísticas em que os valores em análise são classificações / categorizações. Os valores possíveis são números inteiros entre uma e cinco estrelas. Exemplos: "disposição diária", "qualidade do almoço"
- **Counter:** tipo de dados utilizado em estatísticas relativas a valores numéricos que possuem um carácter incremental, ou seja, funcionam como um contador dos valores registados. Por este motivo, os valores tratados são números inteiros. Exemplos: "copos de vinho", "cafés tomados", "cigarros fumados"
- **Duration:** tipo de dados utilizado utilizado para representar um intervalo de tempo. É calculado recorrendo a um cronómetro que é controlado pelo utilizador. Para a respetiva representação gráfica os tempos registados são convertidos em minutos e tratados posteriormente como valores numéricos. Exemplos: "tempo de viagem até casa"

- **Date:** tipo de dados utilizado para representar datas de ocorrências de eventos associados à estatística em análise. Exemplos: “dias em que cheguei atrasado”
- **Option List:** tipo de dados utilizado para a representação de registos nominais de uma lista de opções diferentes. As diferentes opções são adicionadas no processo de criação da própria estatística. Exemplos: “ementa de almoço”, “bares visitados”
- **Free Text:** tipo de dados utilizado em estatísticas com propósito semelhante a blocos de notas em que são guardados registos textuais. Exemplos: “anedotas guardadas”, “frases motivadoras”

No que diz respeito ao motor de visualizações, os tipos de dados *Numeric*, *Counter* e *Duration* são processados da forma análoga, uma vez que tratam valores que podem ser contabilizados e cuja representação gráfica é semelhante. A cada novo registo relativamente a uma variável, é guardado, juntamente com o valor registado, a data de ocorrência dessa atualização. Essas datas são posteriormente utilizadas para a representação gráfica dos dados de modo a permitirem uma análise da evolução temporal da informação.

3.2 Arquitetura do motor de visualizações do Statz.Me

Considerando que os métodos de visualização desenvolvidos foram baseados nas linguagens *JavaScript* e *HTML*, tornou-se necessária a utilização de uma *WebView* (um componente de interface disponível em variadas plataformas móveis como, por exemplo, o *iOS*), uma vez que se trata de um componente com capacidade para interpretar código nas referidas linguagens, de forma a permitir a renderização dos gráficos. Esta *WebView* é o componente principal de uma classe que foi criada na camada *iOS*, denominada *PlotViewController*, que é responsável pela interface dos dados relativos a cada estatística com a respetiva visualização, bem como com a apresentação de cada gráfico no ecrã. A estrutura da criação das visualizações de estatísticas do Statz.Me passou por duas fases distintas, descritas de seguida.

Inicialmente o tratamento de dados e adaptação do seu formato a cada um dos diferentes gráficos foi realizado na camada *iOS* (em código *Objective-C*) nas classes correspondentes ao respetivo tipo de dados. Depois desse tratamento os dados eram enviados para a classe *PlotViewController*, responsável pela escolha do gráfico a renderizar. Esta arquitetura está representada na figura 3.1.

Nesta primeira versão da arquitetura de visualização estatística do Statz.Me a escolha do gráfico a apresentar era feita também na camada *iOS*. Cada classe de dados enviava, juntamente com o pacote de dados, uma lista de identificadores (*id's*) de cada gráfico para a classe *PlotViewController*. Esta classe executava, com base no identificador colocado no topo da lista, o gráfico correspondente.

Uma vez que, nesta arquitetura, o processamento dos dados e escolha das visualizações a apresentar é feita apenas na camada *iOS*, e como tal baseada em código *Objective-C*, verificou-se que a reutilização das visualizações era bastante limitada, pelo que foi necessária a procura de uma solução alternativa que permitisse uma maior flexibilidade para uma posterior reutilização.

Tipos de dados e arquitetura

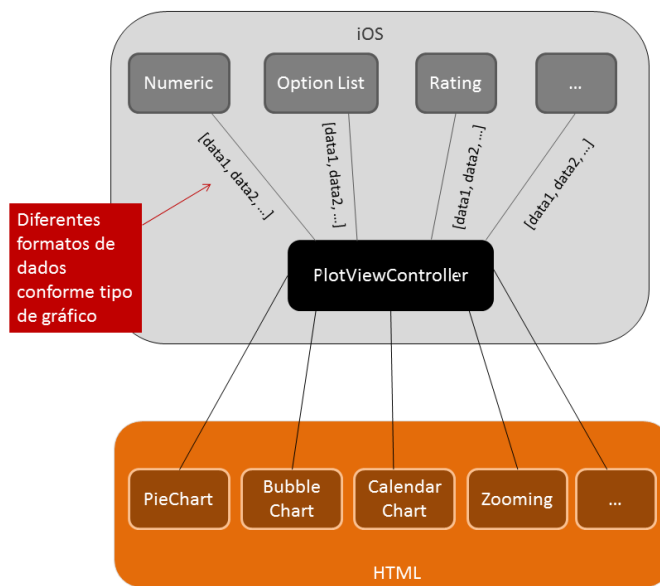


Figura 3.1: Primeira versão da arquitetura do motor de visualizações do Statz.Me.

Com o objetivo de permitir a exportação das visualizações desenvolvidas para um eventual desenvolvimento da aplicação noutros sistemas operativos, foram aplicadas alterações à arquitetura adotada anteriormente, de modo a centralizar o processamento dos dados, adaptação do seu formato e escolha da visualização, na camada *JavaScript*, ao contrário do que era realizado na primeira abordagem. Para o efeito, foi criado um componente central responsável pela gestão da renderização de todos os tipos de visualizações – *StatzVisualizationEngine* – e outros componentes responsáveis por cada tipo de dados e o respetivo processamento – *NumericStatHandler*, (responsável por dados do tipo *Numeric*, *Counter* e *Duration*), *RatingStatHandler*, *OptionListStatHandler* e *MultiStatHandler* (responsável pela gestão da visualização de múltiplas estatísticas em simultâneo). Esta nova estrutura é representada na seguinte figura 3.2.

Dentro da camada *JavaScript*, depois de processados os dados no respetivo componente, é devolvido ao componente principal (*StatzVisualizationEngine*) uma lista de identificadores dos gráficos que podem ser renderizados tendo em conta o tipo de dados em análise. Em seguida é apresentado o gráfico cujo identificador está no topo da lista recebida, e é reenviado para a classe *PlotViewController* (camada *iOS*) a lista dos identificadores. Através de um botão presente na interface desta classe, podem ser mostrados os restantes gráficos que suportam o tipo de dados em causa. Este processo é feito com base na seleção de um novo *id* da lista recebida, e enviado novamente para o componente *StatzVisualizationEngine* que será responsável pela renderização do novo gráfico.

Os vários componentes da arquitetura final são analisados de seguida.

Tipos de dados e arquitetura

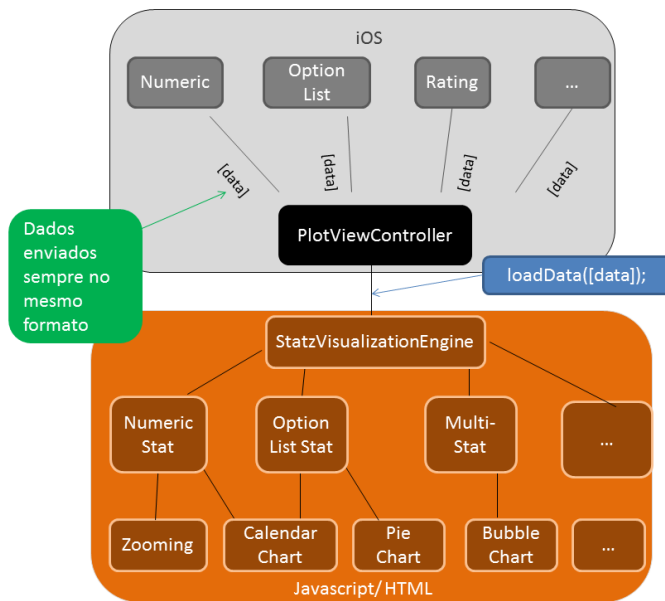


Figura 3.2: Versão atual da arquitetura do motor de visualizações do Statz.Me.

3.3 Detalhes de implementação

Nesta secção são descritos com maior detalhe os vários componentes do motor de visualizações, ilustrando a forma como é feita a escolha dos gráficos conforme o resultado do processamento dos dados das estatísticas que são efetuados por cada componente, e a forma como é feita a comunicação entre estes.

3.3.1 Classes relativas a cada tipo de dados do Statz.Me

Estas classes são responsáveis pela gestão das estatísticas cujos dados são do respetivo tipo retratado. É nestas classes que são adicionados novos registos a cada estatística, bem como a manutenção do histórico de valores. Foi a partir destas classes que se iniciou a criação dos métodos de visualização da aplicação. Inicialmente foi criada uma ligação de cada uma destas classes ao componente *PlotViewController* responsável pela manutenção da *WebView* onde são renderizados os gráficos. No momento em que esta ligação é invocada a classe em execução inicia o processamento de todos os dados registados e envia-os para o componente *PlotViewController* em formato de *string* (texto) representando um objeto *JavaScript*. (exemplo `{“date”: <data>, “value”: <valor do registo>}`, para o caso de valores numéricos, e `{“date”: <data>, “object”: <valor do registo>}`, para valores textuais. No caso de estatísticas do tipo *Date* o formato dos dados é apenas: `{“date”: <data>}`).

3.3.2 PlotViewController

Esta classe, que tem como função principal o carregamento para a respetiva *WebView* do componente *StatzVisualizationEngine* (ficheiro *HTML*) e a inclusão nesse mesmo ficheiro de código *JavaScript*, utiliza o pacote de dados recebido e invoca a execução da função *loadData(<pacote de dados>)* do componente *StatzVisualizationEngine*, que além de apresentar o respetivo gráfico, devolve a lista de identificadores dos gráficos disponíveis para o tipo de dados em questão. Esta lista é utilizada para permitir a visualização de vários gráficos relativos a uma mesma estatística. É nesta classe que está definida a interface que permite a visualização de vários métodos de visualização para uma mesma estatística, através de um botão que foi criado para o efeito. É também nesta classe que é feita a ligação para a interface da escolha de estatísticas a relacionar.

3.3.3 StatzVisualizationEngine

Ficheiro *HTML* onde vão ser renderizados os diferentes gráficos (através da inclusão de código *JavaScript*) e que funciona como componente central do motor de visualizações. É responsável pela comunicação com a camada *iOS* (com a classe *PlotViewController*) recebendo desta os dados das estatísticas que se pretendem representar. Depois de recebidos os dados, são analisadas de seguida as suas propriedades e, conforme o tipo de dados, é invocado o respetivo componente que, no final da sua execução, retorna uma lista de identificadores dos gráficos que é possível representar. A invocação destes componentes é iniciada depois de executada a função *loadData(<pacote de dados>)* e de serem verificados os tipos de dados recebidos. De seguida, é renderizado o gráfico cujo identificador está no topo da lista recebida e reenviada a lista de identificadores para a camada *iOS*, uma vez que é nesta camada que está feita a interface que permite a funcionalidade de análise de diferentes gráficos. Esta funcionalidade tem como base o reenvio de um novo *id* da classe *PlotViewController* para este componente que irá executar novamente, com o novo *id* como argumento, a função responsável pela renderização dos gráficos.

3.3.4 NumericStatHandler

Componente responsável pela gestão de dados de estatísticas do tipo numérico. Tem como função o agrupamento de registos por dia, isto é, todos os registos recebidos que tenham a mesma data são agrupados num só registo, com a data em questão e com a soma dos seus valores. Outra função deste componente é o cálculo do grau de frequência de atualização da estatística recebida, que é utilizado como condição na ordenação dos *id's* a enviar para o componente principal. A figura 3.3 demonstra o funcionamento deste componente.

Depois de calculada a frequência de atualização da estatística, se o seu valor for inferior a 25%, as visualizações a implementar ficam limitadas ao gráfico *Zooming* e *MovingBars*. Esta condição foi feita com base na ideia de que graus inferiores a este valor iriam apresentar no gráfico *Calendar Charts* um número de elevado de “dias em branco” (quadrículas por preencher, dado que este gráfico é feito com o preenchimento com diferentes cores de quadrículas, semelhante a um calendário) o que poderia ser um entrave a uma análise mais generalizada por parte do

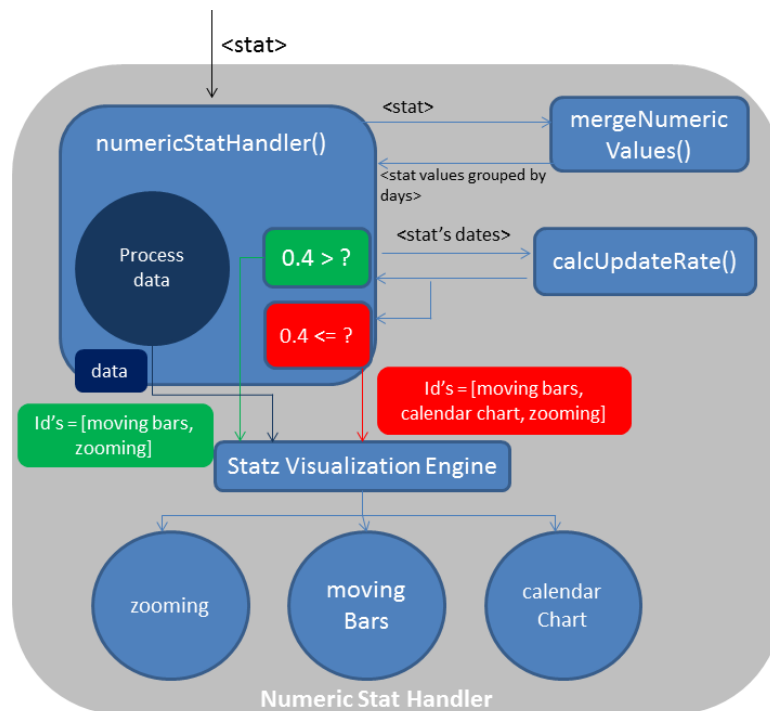


Figura 3.3: Diagrama do funcionamento do componente Numeric Stat Handler.

utilizador. O valor de 25% foi decidido devido a representar uma taxa de atualização relativamente baixa em comparação com aquilo que se pensa ser o grau de atualização comum para estatísticas em aplicações deste género. Uma vez que se trata de uma aplicação cujo contexto principal é a análise de informação pessoal, é aceitável considerar que grande parte das estatísticas a analisar sejam baseadas em dados com um cariz mais prático, e que por isso, à partida, terão um grau de atualização quase diário.

Outro aspeto que é influenciado pelo grau de atualização da estatística tem a ver com a intenção de dar ao utilizador uma análise mais intuitiva do gráfico de forma a facilitar a leitura dos dados. Desta forma, pretende-se que o gráfico que permite uma melhor leitura esteja no topo da lista, para ser o primeiro a ser apresentado. Dado que para este tipo de dados existem gráficos que permitem tipos de análise diferentes, um dos fatores que diferencia estes tipos de análise é precisamente o grau de frequência de atualização, uma vez é possível haver estatísticas cujas perspetivas de visualização sejam diferentes: algumas cujo objetivo da sua análise é apenas a ilustração da evolução dos valores a médio/longo prazo e que, por isso, o seu grau de atualização é menor; e outras cuja funcionalidade é o registo de valores para cada período de tempo (geralmente dias), e cujo grau de atualização tem tendência a ser mais elevado.

Gráficos possíveis: *Zooming, Calendar Chart, Moving Bars*

3.3.5 RatingStatHandler

Componente responsável pelo tratamento de estatísticas do tipo *Rating*. Permite análises temporais e estáticas dos dados. À semelhança do que acontece no tratamento de estatísticas numéricas, é calculado também aqui o grau de atualização de cada estatística e ordenada a lista de *id's* em conformidade.

O funcionamento deste componente pode ser analisado na figura 3.4:

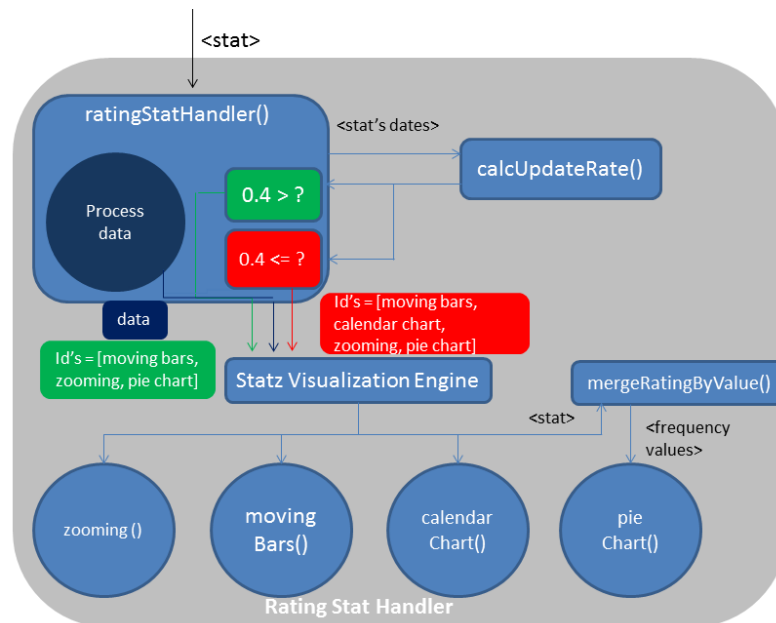


Figura 3.4: Diagrama do funcionamento do componente Rating Stat Handler.

Depois de enviados os *id's* para o componente principal, e no caso do *id* a processar ser o de *Pie Chart*, é executado um processamento de preparação dos dados para este tipo de gráfico separado do componente *RatingStatHandler*. Esta separação é feita uma vez que facilita o processo de preparação dos dados quando ocorre uma atualização do gráfico a renderizar por parte do utilizador, dado que neste processo os dados das estatísticas já não estão sujeitos a um processamento de dados executado nos respetivos componentes, uma vez que é apenas necessária (no máximo) uma adaptação dos dados em cada atualização. O processo de navegação entre visualizações é analisado mais á frente.

Gráficos possíveis: *Pie Chart, Zooming, Calendar Chart, Moving Bars*

3.3.6 OptionListStatHandler

Trata da gestão de estatísticas do tipo “lista de opções”. Pode apresentar até quatro visualizações diferentes que permitem dois tipos de análises: estática e temporal. A análise estática apenas permite ter uma ideia da frequência com que ocorreram as diferentes opções, enquanto que, na análise temporal também é possível analisar o registo das ocorrências ao longo do tempo. Para

este último caso existe ainda outra fator a ter em conta para os gráficos a apresentar, que envolve a ocorrência de mais do que um registo no mesmo dia. Se tal for verificado a visualização de *Calendar Charts* não será considerada, uma vez que apenas permite o registo de uma atualização por dia. Este fator pode ser exemplificado na figura 3.5:

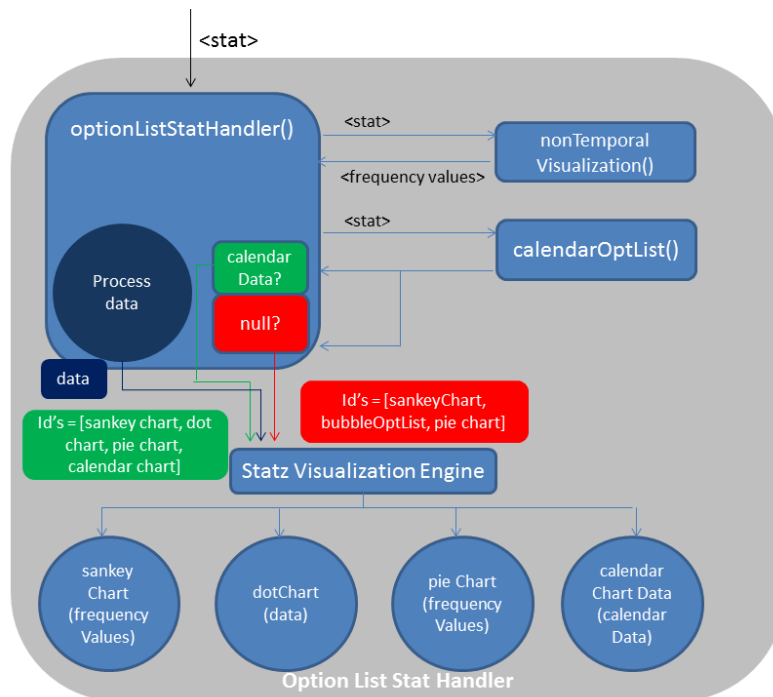


Figura 3.5: Diagrama do funcionamento do componente Option List Stat Handler.

Depois de recebidos os dados da estatística, estes são processados pela função *nonTemporalVisualization()* responsável pela preparação de dados para uma análise estática, ou seja, a preparação dos dados para os gráficos *Sankey* e *Pie Charts*. De seguida é então verificada se existe alguma ocorrência de mais do que um registo por dia através da função *calendarOptList* que verifica se é ou não possível a renderização de um *Calendar Chart*. Se a renderização deste gráfico for possível, então a função devolve um pacote com os dados já preparados para a sua construção e prontos para serem enviados para o componente *StatzVisualizationEngine*. No caso contrário, a função devolve o valor *null*, e o *id* de *Calendar Chart* não é enviado. Os *id's* das visualizações a renderizar são de seguida enviados, juntamente com os respetivos dados preparados para cada gráfico, para *StatzVisualizationEngine*.

Gráficos possíveis: *Pie Chart*, *Sankey Chart*, *Calendar Chart* e *Dot Chart*

3.3.7 MultiStatHandler

Este componente é responsável pela gestão de dados de múltiplas estatísticas. Tem como funções principais, a verificação das combinações de tipos de dados que foram recebidos, o seu agrupamento (tendo em conta a combinação de dados recebida e a respetiva data), ordenação e

reenvio desses dados já preparados, juntamente com a lista de identificadores dos gráficos que suportam os dados em questão. O funcionamento deste componente é demonstrado na figura 3.6.

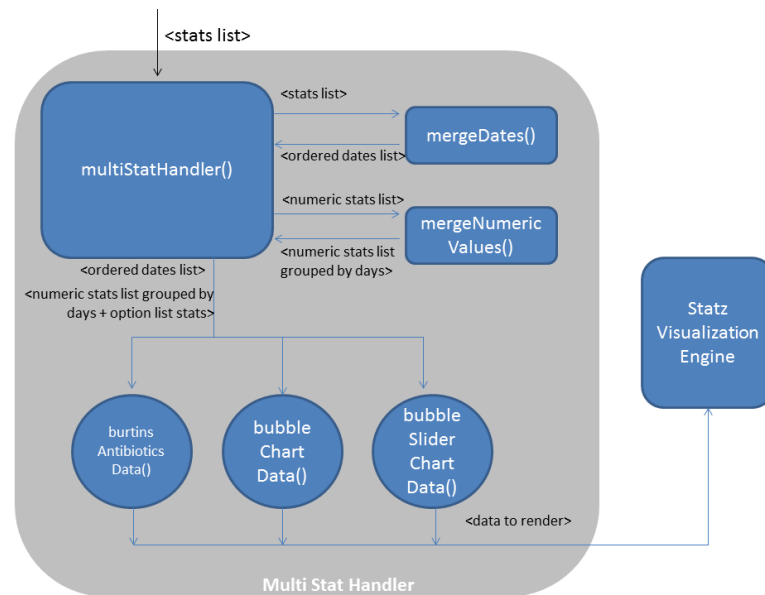


Figura 3.6: Diagrama do funcionamento do componente Multi-Stat Handler.

Depois de recebida a lista de estatísticas, é invocada a função *mergeDates()*, que tem como função a ordenação de todas as datas associadas aos registos recebidos. Esta função devolve uma lista de datas ordenadas, que é usada, em seguida, como auxiliar na adaptação dos dados a cada gráfico, pelas respetivas funções. No caso de terem sido recebidas estatísticas numéricas, é ainda executada a função *mergeNumericValues()* que tem como função o agrupamento de valores numéricos por dia. Depois de processadas as estatísticas numéricas, estas são novamente inseridas na lista e enviadas, juntamente com a lista de datas, para as funções responsáveis pela adaptação dos dados a cada gráfico. Cada uma destas funções reenvia os resultados do seu processamento para o componente *StatzVisualizationEngine*.

Gráficos possíveis: *Burtin's Antibiotics*, *Bubble Slider Chart*, *Bubble Chart*

3.4 Navegação entre diferentes métodos de visualização

Um dos objetivos do trabalho realizado é dar ao utilizador a possibilidade de escolher entre vários métodos de visualização para uma mesma estatística. Um pouco à semelhança do que acontece com o processamento inicial de cada estatística (ou grupo de estatísticas) o processo de renderização é realizado com base na avaliação de um identificador próprio para cada gráfico, que é processado pelo componente *StatzVisualizationEngine*. Na altura da primeira representação gráfica de cada estatística, são enviados para o componente *PlotViewController* (camada *iOS*) os identificadores de cada gráfico que é possível renderizar. Quando é pressionado o botão de navegação entre as visualizações, o componente *PlotViewController* envia para o motor de visualizações

o *id* seguinte na lista juntamente com os dados da estatística no seu formato original. Com base no *id* recebido é invocada, de seguida, a função responsável pela adaptação dos dados originais ao gráfico para o qual o identificador aponta.

3.5 Conclusões

Neste capítulo foram demonstrados os tipos de dados da aplicação, analisando as suas principais características e a sua funcionalidade, tanto a nível das estatísticas que podem suportar como do propósito de utilização que está subjacente ao tipo de dados representado. No processamento das visualizações, os tipos de dados *Numeric*, *Counter* e *Duration* são fundidos num tipo único, uma vez que o seu tratamento para as respetivas visualizações é análogo.

Foi também analisada neste capítulo toda a arquitetura por detrás do motor de visualizações, bem como analisados todos os componentes que dele fazem parte, foram ainda descritos os detalhes mais relevantes da implementação de cada componente.

Depois de analisadas as duas fases pelas quais passou a arquitetura do motor de visualizações, pode-se concluir que a segunda versão é mais vantajosa, uma vez que permite uma maior flexibilidade e portabilidade para um eventual desenvolvimento do Statz.Me noutras plataformas, uma vez que o processamento dos dados das estatísticas é baseado em código *JavaScript* ao contrário da primeira versão em que este processamento era feito em código *Objective-C*. Este fator permite que o motor seja reutilizável noutros ambientes graças à capacidade destes em interpretar código *JavaScript*.

Neste capítulo ainda foi descrito todo o processo de escolha dos gráficos a renderizar, nomeadamente no que toca ao processamento dos dados das estatísticas, análise de condições restritivas, quando aplicável (por exemplo o cálculo do grau de atualização em estatísticas numéricas que pode limitar a utilização da visualização *zooming*), e envio dos identificadores dos gráficos para o componente principal e para a camada *iOS*.

Capítulo 4

d3 & Protovis

Neste capítulo é justificada a utilização das bibliotecas de visualização gráfica *d3* e *Protovis* ao invés do recurso a outras bibliotecas de representação gráfica existentes e baseadas na linguagem *JavaScript*. São enumeradas quais as bibliotecas estudadas, e é feita também uma comparação entre as tecnologias de representação SVG e *Canvas*, uma vez que são as tecnologias em que as bibliotecas analisadas se baseiam.

4.1 Bibliotecas de representação gráfica

As tecnologias estudadas e seleccionadas para suporte ao trabalho de renderização dos gráficos foram as bibliotecas *d3* e *Protovis*, baseadas na linguagem *JavaScript* e com recurso à utilização de SVG – uma família de especificações baseadas em XML que permite a representação de gráficos bidimensionais, estáticos ou dinâmicos. No início do trabalho foi também estudada a possibilidade da utilização de outras bibliotecas baseadas em SVG (*JFreeCharts* [JFC12], *Flot Examples* [FE112], *jQuerySparklines* [JQS12] e *JavaScript InfoVis Toolkit* [JIT12] mas cuja variedade era um pouco limitada. Foi ainda analisada a possibilidade do desenvolvimento dos métodos de visualização com recurso a bibliotecas baseadas em *Canvas/HTML 5* como, por exemplo, as bibliotecas *RGraph* [RGr] e *SenchaTouchCharts* [Sen]. No entanto, mais uma vez, a variedade de opções das bibliotecas *d3* e *Protovis*, aliada às comparações entre as vantagens e desvantagens das duas tecnologias tendo em conta o tipo de renderizações que se pretende, acabou por pesar na decisão final a favor da opção de SVG, sendo então escolhidas as bibliotecas *d3* e *Protovis*. As vantagens principais da utilização de SVG prendem-se principalmente com o facto de derivar de elementos do DOM (*Document Object Model*), uma convenção multi-plataforma e independe de linguagem para a representação e interação de objetos declarados em documentos *HTML*, *XHTML* e *XML*, e que está especificada na sua própria API. Outras vantagens e desvantagens destas duas tecnologias são apresentadas nas tabelas 4.1 e 4.2.

Tabela 4.1: Vantagens das tecnologias Canvas e SVG

Canvas	SVG
Alta performance em renderizações bidimensionais.	Independente da resolução em que é aplicado — o que torna o SVG melhor para interfaces multi-plataforma como, por exemplo, a renderização em browsers de smartphones e browsers comuns uma vez que permite escalamento automático para todas as resoluções de ecrã.
A Performance é constante uma vez que tudo na renderização são pixels. A performance baixa apenas quando a resolução da imagem aumenta.	O SVG apresenta um bom suporte para animações recorrendo a JavaScript, ou à respetiva sintaxe de cada elemento do DOM.
A imagem resultante pode ser guardada com format .png ou .jpg	Existe um controlo total sobre cada elemento utilizado a API do DOM para SVG em JavaScript.
Mais indicado para renderização de gráficos <i>bitmap</i> (como, por exemplo, em jogos). As operações de edição de imagem requerem manipulação ao nível do pixel.	Um elemento de SVG é um ficheiro com formato XML, o que leva a uma melhor acessibilidade por parte de cada <i>browser</i> .

Tabela 4.2: Desvantagens das tecnologias Canvas e SVG

Canvas	SVG
Não são criados elementos do DOM para qualquer tipo de renderização. É tudo baseado em pixels.	Renderização mais lenta à medida que a complexidade do documento aumenta.
Não existe uma API de suporte à animação. Estas são feitas recorrendo a timers e a outros eventos que são responsáveis por atualizar o Canvas quando necessário.	Pouco apropriada para aplicações como jogos, ou cuja complexidade seja relativamente elevada.
Pouco eficaz na renderização de texto.	
Pouco apropriado para utilização em aplicações com interface de utilizador, uma vez que estas necessitam de ser dinâmicas e interativas, e Canvas necessita de renderização manual de cada elemento.	

Quanto às respetivas performances das duas tecnologias, estas podem ser analisadas na figura 4.1 [Svg12]

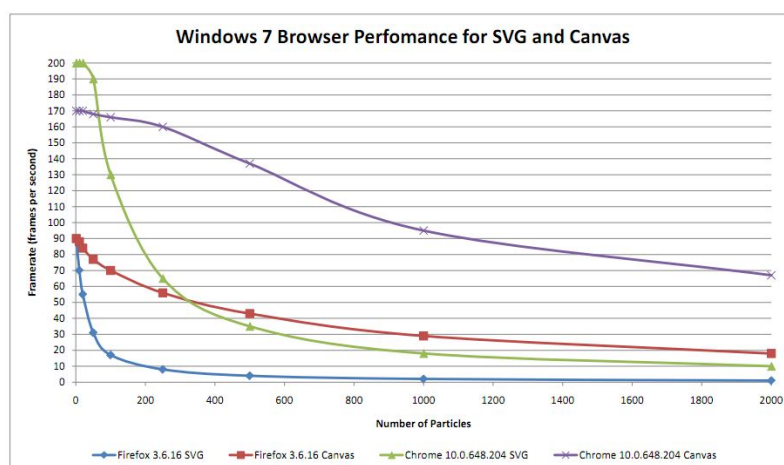


Figura 4.1: Gráfico de análise de desempenho das tecnologias SVG e Canvas em dois browsers distintos

Testado em ambiente *Windows*, foi calculada a performance baseada no nível de FPS (*frames per second*) obtidas com SVG e *Canvas* em diferentes *browsers*. Como se pode analisar no gráfico, para renderizações de poucas partículas a utilização de SVG obtém melhores resultados do que *Canvas*, com a utilização do *Google Chrome*. A partir das 100 partículas, a performance de SVG baixa consideravelmente, devido ao facto de o processamento de elementos do DOM ser mais lento à medida que a sua complexidade aumenta.

No que toca às duas bibliotecas utilizadas, pode-se afirmar que ambas estão bem documentadas e possuem variados exemplos de aplicações, que recorrem às suas capacidades, dado serem uma alternativa bastante viável e bastante utilizada no desenvolvimento de métodos de visualização fundamentalmente ao nível da *Web*. O seu funcionamento no *iPhone* também é bastante eficaz, bem como a interface das interações implementadas nos diferentes gráficos.

4.2 Conclusões

Neste capítulo foi justificada a utilização das bibliotecas *d3* e *Protovis* tendo sido justificada a sua utilização em detrimento de outras bibliotecas de visualizações de dados baseadas tanto em SVG como em *Canvas*. O fator principal para esta decisão acabou por ser a grande variedade de gráficos destas duas bibliotecas, mas o facto de estas recorrerem a SVG teve também um papel importante na decisão final, uma vez que o bom suporte para animações (recorrendo a *JavaScript*) aliado à possibilidade de controlo total sobre cada elemento utilizando a API e sintaxe de elementos do DOM foram argumentos mais fortes face às vantagens apresentadas pela utilização de *Canvas*. Neste capítulo também é feita uma comparação entre as vantagens entre estas duas tecnologias, de onde se pode concluir que para representações gráficas pouco complexas,

SVG apresenta melhor performance do que *Canvas*, no entanto, à medida que a complexidade das representações aumenta *Canvas* acaba por conseguir um melhor desempenho. Uma vez que os métodos de visualização desenvolvidos para o *Statz.Me* não têm uma complexidade elevada, a escolha de bibliotecas baseadas em SVG acabou por garantir também um melhor desempenho.

Capítulo 5

Resultados

Nesta secção são apresentadas as visualizações implementadas no Statz.Me, descritas as suas principais características e analisados os motivos principais que levaram à sua implementação para os respetivos tipos de dados.

5.1 Visualizações implementadas

5.1.1 Zooming

Este método de representação visual é aplicado na visualização de estatísticas do tipo numérico (*numeric, duration, counter, rating*). Permite uma perspetiva geral dos dados ao longo do tempo, uma boa perceção de mudanças repentinas de valores, e uma visão detalhada de informação recorrendo a uma interação de *zooming* e de navegação temporal ao longo do eixo horizontal. É especialmente indicada para estatísticas cujo grau de atualização é relativamente baixo, uma vez que nestes casos é mais relevante ter uma perceção geral da evolução dos valores do que a informação sobre o dia em que esses foram alterados. Por exemplo, numa estatística sobre a evolução do peso, é mais interessante uma perceção geral imediata da sua evolução, do que propriamente o dia exato em que os valores de peso foram registados. No entanto esta informação pode também ser obtida com o recurso à interação que permite navegabilidade e *zooming*. Esta representação gráfica é utilizada no Statz.Me como visualização principal (a primeira a ser representada para cada estatística numérica) para estatísticas que possuam um grau de atualização inferior a 25%. Isto deve-se ao facto de este gráfico ser mais indicado para a ilustração da evolução de valores numéricos a médio/longo prazo, ao contrário da representação da relação “tempo - valor correspondente”, que está mais inerente a outros tipos de visualização. Este valor de 25% foi decidido devido a representar uma taxa de atualização relativamente baixa em comparação com aquilo que se pensa ser o grau de atualização comum para estatísticas em aplicações deste género.

Este grau de atualização é calculado com base no número de dias em que ocorreram registos da estatística a dividir pelo total de dias entre a primeira e a última atualização.

Resultados

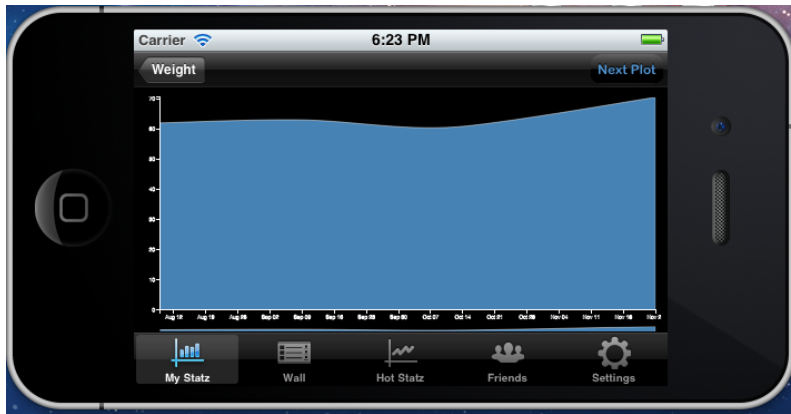


Figura 5.1: *Screenshot* do protótipo do Statz.Me apresentando uma visualização do tipo *Zooming* para uma estatística numérica.

5.1.2 Pie Chart

Esta representação gráfica é aplicada em estatísticas cujos dados de base são “lista de opções” ou *ratings*, em que o conjunto de valores possíveis de registo está pré-definido. Permite uma percepção geral da distribuição da frequência dos valores registados através da área de cada “fatia” do gráfico circular representada. Não permite uma análise da evolução temporal dos valores, pelo que, uma estatística representada deste modo deve ser complementada com outros métodos de representação que permitam uma análise da evolução temporal.

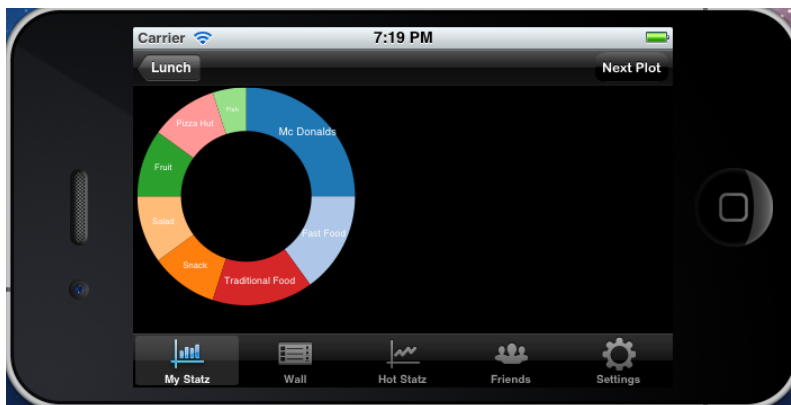


Figura 5.2: *Screenshot* do protótipo do Statz.Me apresentando uma visualização do tipo *Pie Chart* para um estatística do tipo "lista de opções".

5.1.3 Calendar Chart

Esta representação gráfica é utilizada tanto para estatísticas numéricas como para estatísticas cujos dados de base são “lista de opções”. Neste último caso, este gráfico só é utilizado se houver apenas uma atualização por dia da respetiva estatística.

Resultados

O *Calendar Charts* permite ao utilizador ter uma boa perspetiva geral da informação e pode facilitar a descoberta de padrões nos dados, ao longo do tempo, através de análise visual. O funcionamento deste gráfico baseia-se no preenchimento de cada quadrícula representativa de um dia com a cor correspondente a um determinado valor. Para uma melhor interpretação torna-se adequado complementar com uma legenda, para ilustrar os valores correspondentes a cada cor. A representação de valores numéricos recorre a uma interpolação de cores. No exemplo da Figura 5.3 a escala de cores varia entre verde e vermelho. Estas cores foram escolhidas por serem comumente consideradas como valores positivos/negativos. Sendo o verde correspondente a valores mais altos, e o vermelho a valores mais baixos. Para o caso de valores do tipo “lista de opções”, cada opção tem atribuída uma cor aleatória.

Este gráfico tem, no entanto, uma limitação considerável, que consiste na impossibilidade de serem representados diferentes valores para um mesmo dia. Para estatísticas onde seja necessário, este gráfico é substituído por uma adaptação de um *Bubble Chart* onde, para um mesmo dia, pode ser visto mais do que um registo e a hora em que ocorreu.

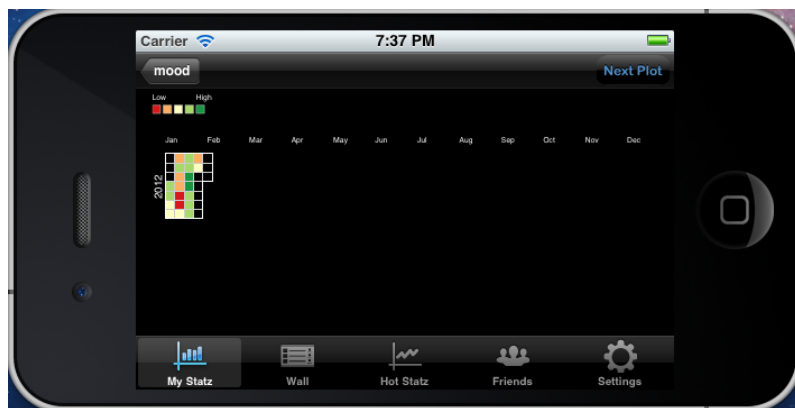


Figura 5.3: Screenshot do protótipo do Statz.Me apresentando uma visualização do tipo *Calendar Chart* para estatísticas que retratam valores numéricos.

5.1.4 Sankey Chart

Esta representação gráfica tem um propósito idêntico às representações *Pie Charts*: é utilizada na análise da frequência de cada valor em estatísticas baseadas em “lista de opções”, não sendo possível uma análise temporal da informação. No entanto, ao contrário do que acontece nos *Pie Charts*, este gráfico indica os valores exatos de cada frequência, no final de cada barra. Tal como os *Pie Charts* a vantagem principal deste gráfico é a perspetiva geral dos dados, devido à percepção que é dada pelos tamanhos de cada barra.

5.1.5 Moving Bars

Esta representação gráfica é utilizada na apresentação de estatísticas numéricas suportada em gráficos de barras, com a possibilidade de análise de evolução temporal. Um gráfico produzido

Resultados



Figura 5.4: *Screenshot* do protótipo do Statz.Me apresentando uma visualização do tipo *Calendar Chart* para estatísticas que retratam valores do tipo "lista de opções".

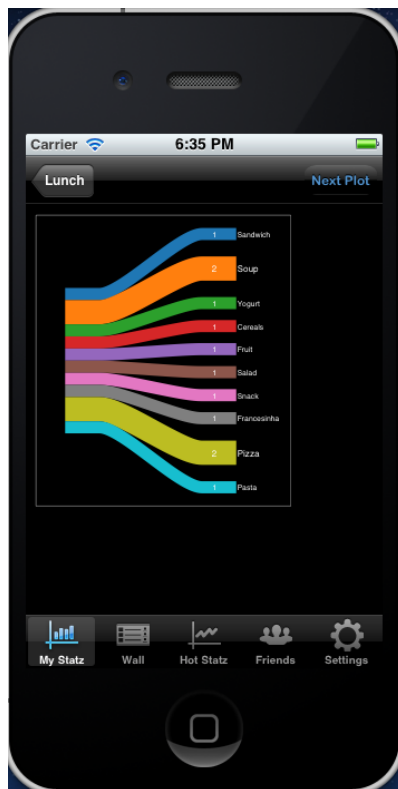


Figura 5.5: *Screenshot* do protótipo do Statz.Me apresentando uma visualização do tipo *Sankey Chart* para estatísticas que retratam valores do tipo "lista de opções".

recorrendo a esta representação é dividido em partes de sete *slots*, podendo cada *slot* ser ou não preenchida com uma barra representativa dos valores de determinado dia (sete *slots* correspondendo a intervalos de sete dias). Os *slots* vazios representam intervalos de tempo superiores a um dia entre cada atualização. Com esta opção pretendeu-se dar ao utilizador uma ideia de interrupção entre as atualizações. Desta forma, as atualizações em dias consecutivos são representadas com

Resultados

barras sem espaços entre elas, e as atualizações que ocorram com mais do que um dia de intervalo são representadas com um espaço de intervalo. Através de uma interação de *slide* o utilizador pode navegar ao longo do gráfico para efetuar uma análise dos dados ao longo do tempo.

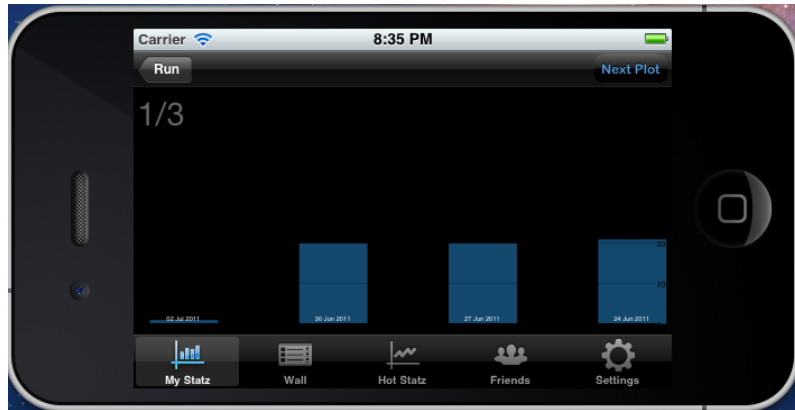


Figura 5.6: *Screenshot* do protótipo do Statz.Me apresentando uma visualização do tipo *Moving Bars* para uma estatística numérica.

5.1.6 Timeline Charts - Dot Charts & Dynamic Bubble Charts

Esta representação gráfica é utilizada para vários tipos de visualizações, suportando a visualização de estatísticas múltiplas, e ainda a visualização de estatísticas singulares para variáveis do tipo “lista de opções”.

Relativamente à visualização de múltiplas estatísticas, este gráfico suporta dois tipos de relações: entre várias estatísticas numéricas; e entre duas estatísticas numéricas e uma do tipo “lista de opções”, ambas com possibilidade de análise de dados ao longo do tempo.

No caso de relações entre valores apenas numéricos, cada bolha representa uma determinada atualização, em que a cor define qual a estatística que representa, e o tamanho da bolha o seu respetivo valor. As posições horizontal e vertical de cada bolha definem respetivamente o dia e hora em que foi efetuado o respetivo registo.

Por outro lado, se as estatísticas a analisar forem do tipo numérico e de “lista de opções” a representação é feita da seguinte forma: a cor de cada bolha representa o valor da variável da lista de opções, enquanto o tamanho da bolha e a sua posição vertical definem os valores das variáveis numéricas. Para estes dois casos, na eventualidade de existirem dias sem registos de alguma das estatísticas a analisar, a sua representação no gráfico é feita com valores pré-definidos: *undefined*, no caso de serem valores do tipo “lista de opções” e cuja cor da bolha é definida a preto. No caso de valores numéricos, para a variável correspondente ao tamanho das bolhas, o registo deixa de ser identificado por uma bolha e passa a ser representado por outra figura geométrica, e a variável que é identificada pela posição vertical é representada com uma altura igual a zero, ou seja situada sobre o eixo horizontal.

Resultados

No caso de a visualização tratar apenas uma estatística do tipo “lista de opções”, é usada a alternativa *Dot Charts*, semelhante à anterior mas com um tamanho pré-definido para cada bolha uma vez que este fator não é representativo de nenhuma variável nesta visualização. Cada ponto representa um registo diferente, em que o valor do tipo “lista de opções” é definido pela sua cor, e a sua posição vertical define a hora a que o mesmo registo ocorreu. Esta funcionalidade foi acrescentada devido à necessidade de representar mais do que um registo por dia para estatísticas deste tipo.

Em ambos os casos é possível, através de interação, uma análise temporal dos dados através de navegação ao longo do gráfico. É ainda possível uma leitura de informação mais detalhada, recorrendo a etiquetas que são mostradas através de toque em cada ponto.

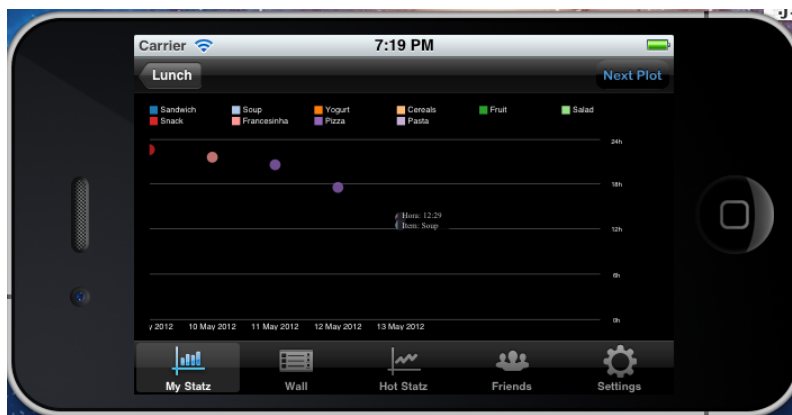


Figura 5.7: Screenshot do protótipo do Statz.Me apresentando uma visualização do tipo *Dot Chart* para uma estatística do tipo "lista de opções".

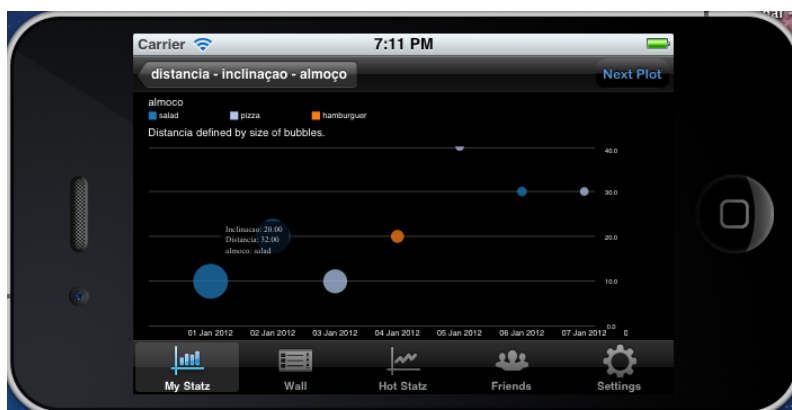


Figura 5.8: Screenshot do protótipo do Statz.Me apresentando uma visualização do tipo *Dynamic Bubble Chart* onde são relacionadas duas estatísticas numéricas com uma do tipo "lista de opções".

Resultados

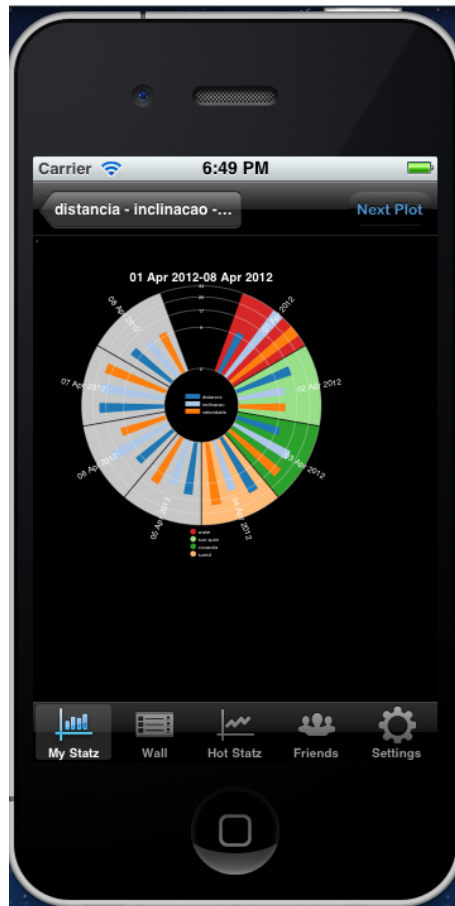


Figura 5.10: *Screenshot* do protótipo do Statz.Me apresentando uma visualização do tipo *Burtin's Antibiotics* onde são relacionadas três estatísticas numéricas com uma do tipo "lista de opções".

5.1.8 Static Bubble Charts

Esta representação gráfica está vocacionada para a representação de múltiplas estatísticas com suporte de relações até três variáveis numéricas e uma variável do tipo “lista de opções”. Permite ainda uma análise de informação ao longo do tempo, ainda que de forma menos intuitiva do que as visualizações analisadas anteriormente, uma vez que, neste gráfico os dados não estão dispostos horizontalmente por ordem cronológica, devido ao facto da posição horizontal de cada registo ser definida pelo valor de uma das variáveis numéricas, e não pela data em que ocorreram como acontece nas outras visualizações analisadas. É a par da visualização *Burtin's Antibiotics* a representação com maior suporte de variáveis diferentes, tendo no entanto a vantagem de permitir uma leitura geral dos dados mais intuitiva uma vez que podem ser descobertas mais facilmente relações de proporcionalidade, ou proporcionalidade inversa os valores representados através das suas posições no gráfico.



Figura 5.11: Screenshot do protótipo do Statz.Me apresentando uma visualização do tipo *Static Bubble Charts* onde são relacionadas três estatísticas numéricas com uma do tipo "lista de opções".

5.1.9 Análise global aos métodos desenvolvidos

As visualizações implementadas podem ser analisadas de forma mais global na tabela 5.1, onde são relacionadas algumas características fundamentais com o tipo de visualização analisado.

5.2 Conclusões

Neste capítulo foram analisados os métodos de visualização implementados no Statz.Me, bem como as características mais importantes de cada um e os tipos de estatísticas em que são aplicados. É apresentada também para cada gráfico a justificação da sua utilização e as perspectivas de análise que permite.

Para estatísticas cujos valores sejam numéricos os métodos utilizados são: *zooming*, *calendar charts* e *moving bars*. A visualização *zooming*, é caracterizada principalmente pela sua interação que permite uma navegação temporal e visão mais detalhada dos dados. *Calendar Charts* são caracterizados por permitirem descoberta de padrões nos dados ao longo do tempo graças ao mapeamento dos valores num calendário, com diferentes cores conforme o intervalo de valores para o respetivo registo.

Estatísticas do tipo lista de opções são mapeadas com os métodos: *Pie Chart*, *Calendar Charts* (adaptados para a representação de valores deste tipo), *Sankey Chart*, *Dot Chart*. *Pie Charts*, mostram a percentagem das ocorrências de cada valor da lista de opções num gráfico circular, dando ao utilizador uma perceção geral de quais os valores registados com maior frequência. O funcionamento dos *Calendar Charts* já foi explicado anteriormente, sendo aplicado da mesma forma que nas estatísticas para valores numéricos. *Sankey Charts*, funcionam de forma análoga aos *Pie Charts*, apresentado a vantagem de mostrarem os valores exatos da frequência de ocorrência dos valores.

Para o caso de múltiplas estatísticas são utilizados os métodos: *Static Bubble Charts*, *Dynamic Bubble Charts* e *Burtin's Antibiotics*. *Static Bubble Charts* permitem uma representação de quatro

Resultados

Tabela 5.1: Análise das principais características dos métodos criados.

	Estatísticas singulares	Múltiplas estatísticas	Valores Numéricos	Valores “lista de opções”	Análise temporal	Percepção de padrões / relações entre valores	Interação
<i>Zooming</i>	Sim	Não	Sim	Não	Sim	Percepção da altura do gráfico em relação a determinados períodos de tempo	<i>Zooming</i> e navegação temporal
<i>Calendar Chart</i>	Sim	Sim (estatísticas tipo “date” e outro)	Sim	Sim	Sim	Percepção de alguns padrões de cores no gráfico em determinadas alturas	Não aplicável
<i>Pie Chart</i>	Sim	Não	Valores de rating apenas	Sim	Não	Não	Não aplicável
<i>Sankey Chart</i>	Sim	Não	Não	Sim	Não	Não	Não aplicável
<i>Moving Bars</i>	Sim	Não	Sim	Não	Sim	Não	<i>Slide</i> para navegação entre blocos de tempo
<i>Dot Chart</i>	Sim	Não	Não	Sim	Sim	Percepção da repetição de alguns valores em determinadas horas, com base na posição dos pontos	<i>Slide</i> para navegação temporal
<i>Dynamic Bubble Charts</i>	Não	Sim	Sim	Sim	Sim	Percepção de quantidade dada pela ilustração dos valores através do tamanho das bolhas	<i>Slide</i> para navegação temporal.
<i>Burtin’s Antibiotics</i>	Não	Sim	Sim	Sim	Sim	Percepção de quantidade dada pela comparação de tamanhos das barras, e a sua relação com os valores do tipo “lista de opções” que são definidos pela cor de fundo	<i>Slide</i> para navegação entre blocos de tempo.
<i>Static Bubble Chart</i>	Não	Sim	Sim	Sim	Sim	Relação entre posição das bolhas no espaço com o seu tamanho	<i>Slide</i> para navegação entre blocos de tempo

Resultados

variáveis diferentes (três numéricas, e uma do tipo “lista de opções”), através do posicionamento de bolhas num referencial cujos eixos definem os valores de duas das variáveis numéricas. A terceira variável numérica é definida pelo tamanho de cada bolha, a cor desta define a variável do tipo “lista de opções”. Esta visualização permite uma análise ao longo do tempo, ainda que de forma menos intuitiva do que as restantes, uma vez que a informação temporal é dada apenas por etiquetas com a respetiva data que são posicionadas junto de cada bolha, e não através do posicionamento das bolhas da esquerda para a direita como no caso das visualizações restantes em que os dados podem ser vistos com uma linha temporal. *Dynamic Bubble Charts* permitem a representação de várias combinações de dados: três variáveis diferentes (duas numéricas e uma do tipo “lista de opções” ou duas variáveis do tipo “lista de opções” e uma do tipo numérico) e um qualquer número de variáveis numéricas. A visualização *Burtin's Antibiotics* relaciona até três variáveis numéricas e uma do tipo lista de opções. As variáveis numéricas são representadas com barras, e a variável do tipo “lista de opções” com a cor de fundo de cada fatia (correspondente a um dia) do gráfico.

Resultados

Capítulo 6

Conclusões e trabalho futuro

6.1 Conclusões

A visualização de dados é uma área que tem vindo a crescer a um ritmo acelerado nos últimos anos, impulsionada pelo desenvolvimento das várias disciplinas que lhe estão associadas, pelo avanço tecnológico, pela ubiquidade das comunicações e pelo surgimento e vulgarização de dispositivos móveis. Têm surgido com elevada frequência métodos inovadores nos vários domínios da visualização, como a visualização de dados, a visualização conceptual ou a visualização de informação. Alguns dos aspetos inovadores destes métodos estão relacionados com a utilização de novos meios interativos (que permitem uma melhor exploração de informação); novas influências estilísticas (mais focadas na interpretação de informação e reflexão pessoal), e com o desenvolvimento de algoritmos de visualização que permitem a análise de múltiplas variáveis em simultâneo.

Apesar deste desenvolvimento, os métodos de visualização de informação mais recentes não são ainda aplicados frequentemente no domínio das aplicações móveis, em geral, e em aplicações de registo de informação pessoal em particular. Considera-se que a inversão deste facto pode tornar as aplicações móveis mais atrativas para o utilizador, além de lhe proporcionarem uma visão melhorada sobre os dados registados.

Um dos pontos essenciais deste trabalho foi o estudo das taxonomias e de outros fatores essenciais na criação de tipos de visualização, e conhecer, estudar e classificar métodos de visualização de informação. O levantamento de informação realizado conduziu à criação de conhecimento que foi fundamental na seleção, criação e aplicação de métodos de visualização, que foram incorporados na aplicação *Statz.Me*.

Antes de ser iniciada a fase de implementação foram ainda estudadas as tecnologias com capacidade de suportar a implementação das visualizações, tendo sido feita uma comparação entre elas e apresentadas as conclusões que permitiram a decisão sobre qual das opções a utilizar para o desenvolvimento dos métodos de visualização.

Conclusões e trabalho futuro

Relativamente ao conhecimento adquirido inicialmente no domínio da visualização, fatores essenciais como a visão geral dos dados, visão mais detalhada, perspetivas de análise distintas, descoberta de padrões e representação de correlações foram tidos em consideração nos métodos desenvolvidos para o Statz.Me, e cuja implementação cobre estes pontos.

Comparando os objetivos propostos inicialmente com o que foi de facto atingido no final do trabalho, pode concluir-se que os resultados são satisfatórios na medida em que:

- Foram implementados métodos de visualização para todos os tipos de dados da aplicação.
- Foi feita a implementação de mais do que um método de visualização para cada estatística singular, sendo assim possível a navegação entre vários métodos de visualização para a mesma estatística.
- Foram desenvolvidos métodos de visualização com recurso a interações, de forma a que o utilizador possa efetuar uma análise mais intuitiva dos dados.
- Foi desenvolvido mais do que um método que permite a análise de múltiplas estatísticas simultaneamente.

Existem, no entanto, novas perspetivas de funcionalidades e tipos de visualização que podem ser acrescentados à aplicação, e que vão ser analisados na secção seguinte.

6.2 Trabalho Futuro

Uma vez terminadas as fases de desenvolvimento e de implementação dos métodos de visualização foram desenvolvidas ideias no sentido de valorizar tanto o Statz.Me, como as visualizações já implementadas. Algumas dessas ideias baseiam-se em outros métodos de representação gráfica a implementar com diferentes tipos de análise e funcionalidade. Outras porém, envolvem uma nova abordagem da apresentação das estatísticas, de modo a serem agrupadas com base num contexto escolhido pelo utilizador. Estas novas ideias são apresentadas de seguida.

6.2.1 Outras Visualizações

Um dos pontos principais para um eventual desenvolvimento futuro do Statz.Me é a implementação de outros tipos de métodos de visualização que permitam diferentes tipos de análises e que, aliados à implementação de outras funcionalidades que foram pensadas, acrescentam um valor considerável à aplicação. Os métodos de visualização pensados são descritos nesta secção.

6.2.1.1 Mapas

Este tipo de representação gráfica oferece um contexto espacial às estatísticas. Pode ser utilizada para representar duas estatísticas do tipo “lista de opções” que seriam mapeadas pela cor de cada bolha e pela respetiva etiqueta textual. Esta visualização surgiu da ideia de uma outra funcionalidade que foi pensada para uma eventual implementação, no futuro, e que consiste em

Conclusões e trabalho futuro

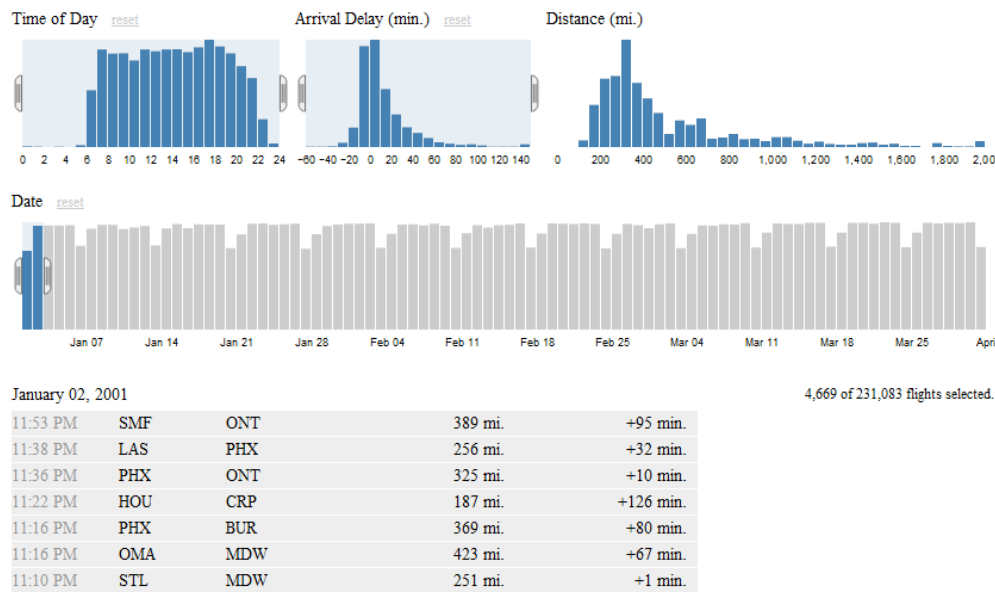


Figura 6.2: Exemplo de utilização da visualização *Cross Filtering* - através da aplicação de filtros nos diferentes gráficos pode ser tiradas conclusões interessantes sobre a informação retratada

Considera-se que esta representação gráfica é bastante útil na aplicação, uma vez que permitirá a ilustração entre várias estatísticas, mostrar a sua relação e, graças à filtragem eficaz, produzir uma visão mais detalhada sobre os respetivos valores de cada estatística, o que pode levar a novas conclusões sobre a influência entre os seus valores.

6.2.1.3 House Hunting

Visualização que recorre à utilização de diferentes cores para o mapeamento de valores numéricos tendo em conta dia e a hora em que foram registados. Esta visualização recorre ainda a mais três gráficos, que são usados para uma filtragem e organização de dados. No exemplo da figura 6.3, a matriz colorida ilustra a pesquisa de habitações por dia e hora no estado do Dakota do Sul nos EUA. O gráfico circular ilustra a partir de que dispositivos foram feitas essas pesquisas (computadores ou dispositivos móveis), e o gráfico de barras ilustra os mesmos valores numéricos da matriz colorida mas apenas para o dia selecionado na matriz (neste exemplo, Th (quinta-feira), 7AM).

Considera-se que esta representação gráfica valoriza a extração de conhecimento a partir da aplicação Staz.Me, uma vez que oferece uma perceção imediata do comportamento geral dos valores registados tendo em conta o dia e a hora de registo. Permite, também, através dos gráficos secundários, a realização de outras análises, importantes para a estatística.

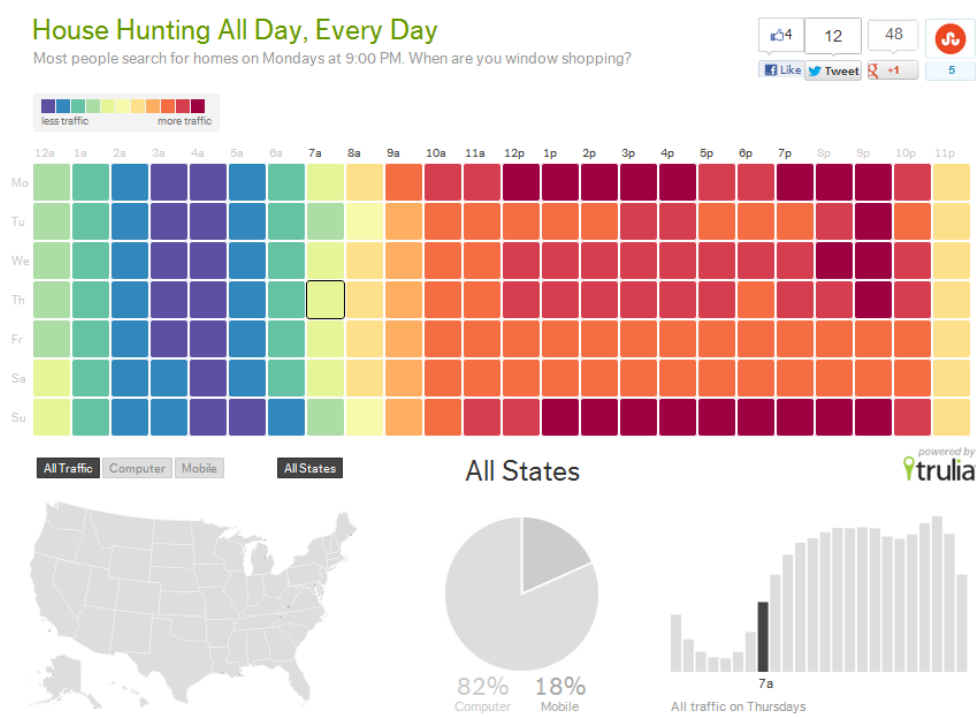


Figura 6.3: Exemplo de utilização da visualização *House Hunting* - a percepção do utilizador sobre os valores é facilitada, através da utilização de diferentes cores e da análise dos outros gráficos adjacentes.

6.2.2 Novas Funcionalidades

6.2.2.1 Contexto espacial

Como foi dito na secção 6.2.1.1, outra das ideias a desenvolver num eventual trabalho futuro consiste na implementação da funcionalidade de registo de variáveis de localização no processo de atualização de cada estatística.

A implementação desta funcionalidade traz grandes vantagens à aplicação uma vez que acrescenta mais uma variável (local) que pode ser alvo de uma análise enriquecedora do comportamento do utilizador, uma vez que este comportamento pode ser influenciado pelo local onde o utilizador está situado.

Por exemplo, é aceitável a pertinência de uma comparação entre o número de copos de cerveja ingeridos, para alturas do dia semelhantes, entre um local com um elevado número de bares na proximidade e mais propício ao lazer/entretenimento e outro com um número de bares inferior. Isto é, as características do local podem influenciar as ações do utilizador. Este tipo de informação será, posteriormente, representado com base no método de visualização *Mapas* apresentado na secção anterior, onde a informação é ilustrada num mapa, nos locais em que os registos ocorreram.

6.2.2.2 Grupos de estatísticas

A ideia da criação de grupos de estatísticas surgiu na altura da implementação de *Static* e *Dynamic Bubble Charts*, em que foi pensado que essas visualizações seriam mais apropriadas à representação de estatísticas dentro de um contexto comum. Ou seja, uma estatística composta por várias estatísticas relacionadas entre si. Por exemplo, considerando um grupo de estatísticas denominado “Corrida”, podem ser consideradas as estatísticas “distância”, “inclinação” e “tipo de percurso”, onde as estatísticas “distância” e “inclinação” são numéricas e, por isso, representadas através da posição vertical e tamanho das bolhas, enquanto que a estatística “tipo de percurso” será do tipo “lista de opções” e, portanto, será representada através da cor de cada bolha.

Neste contexto, os grupos de estatísticas não estarão limitados a um número fixo de estatísticas relacionadas, nem estarão limitados a um único tipo de visualização. Podem ser criados de forma livre pelo utilizador, estando depois, no entanto, as suas visualizações limitadas ao número máximo de variáveis suportado para cada visualização. De seguida são apresentados outros exemplos de grupos de estatísticas que podem ser constituídos e passar a integrar a aplicação Staz.Me:

Estes grupos de estatísticas constituem, também, uma motivação adicional para a implementação dos métodos de visualização adicionais, que foram apresentados neste capítulo, uma vez que estes são apropriados para agrupamento, pois abordam várias variáveis em simultâneo.

Tabela 6.1: Exemplos de possíveis grupos de estatísticas

Grupo	Estatísticas constituintes
Futebol	Nº golos (numérico), nº assistências (numérico), tempo jogado (duração), nº faltas (numérico), posição jogada (lista de opções)
Viagem	Nº km's (numérico), tempo (duração), velocidade média (numérico), partida/chegada (texto livre)
Tempo livre	Tempo leitura (duração/numérico), tempo a ouvir música (duração/numérico), tempo a ver cinema (duração/numérico)
Álbuns de música	Banda (texto livre), género (lista de opções), classificação (rating)
Filmes	Realizador (texto livre), género (lista de opções), classificação (rating)
Restaurantes	Ementa escolhida (texto livre), satisfação (rating), preço (numérico)
Bares	Ambiente (rating/lista de opções), localização (coordenadas registadas / texto livre), qualidade música (rating)
Locais de interesse	Tipo (lista de opções), classificação (rating), localização (coordenadas registadas / texto livre)

Conclusões e trabalho futuro

Referências

- [ABM⁺07] Wolfgang Aigner, Alessio Bertone, Silvia Miksch, Christian Tominski e Heidrun Schumann. Towards a conceptual framework for visual analytics of time and time-oriented data. In *2007 Winter Simulation Conference*, pages 721–729. IEEE, December 2007.
- [AMM⁺07a] Wolfgang Aigner, S. Miksch, W. Muller, H. Schumann e C. Tominski. Visualizing time-oriented data—A systematic view. *Computers & Graphics*, 31(3):401–409, 2007.
- [AMM⁺07b] Wolfgang Aigner, Silvia Miksch, Wolfgang Müller, Heidrun Schumann e Christian Tominski. Visual methods for analyzing time-oriented data. *IEEE transactions on visualization and computer graphics*, 14(1):47–60, 2007.
- [BC112] Visualization: Bubble charts - google charts tools, 2012. Disponível em <http://code.google.com/intl/pt-PT/apis/chart/interactive/docs/gallery/bubblechart.html>, acessado a última vez em 05 de janeiro de 2012.
- [BS112] Bars of stuff - sample, 2012. Disponível em <http://visapi-gadgets.googlecode.com/svn/trunk/barsofstuff/doc.html>, acessado a última vez em 05 de janeiro de 2012.
- [Bul12] Protovis - bullet charts, 2012. Disponível em <http://mbostock.github.com/protovis/ex/bullet.html>, acessado a última vez em 05 de fevereiro de 2012.
- [Bur] Burtin antibiotics illustrations | forth go. Disponível em <http://www.forthgo.com/blog/2009/01/11/burtin-antibiotic-illustrations/>.
- [C1212] Candlesticks - australian securities exchange, 2012. Disponível em <http://www.asx.com.au/research/candlestick.htm>, acessado a última vez em 25 de janeiro de 2012.
- [Chi00] Ed H. Chi. A taxonomy of visualization techniques using the data state reference model. In *Proceedings of the IEEE Symposium on Information Vizualization 2000*, INFOVIS '00, pages 69–, Washington, DC, USA, 2000. IEEE Computer Society.
- [CK98] J.V. Carlis e J.A. Konstan. Interactive visualization of serial periodic data. In *Proceedings of the 11th annual ACM symposium on User interface software and technology*, pages 29–38. ACM, 1998.
- [Dre12] The importance of being earnest, or showing context in data visualization, 2012. Disponível em <http://blog.visual.ly/>

REFERÊNCIAS

- [context-in-data-visualization/](#), acessado a última vez em 13 de junho de 2012.
- [DV111] Datavisualization.ch, 2011. Disponível em <http://datavisualization.ch/>, acessado a última vez em 23 de novembro de 2011.
- [FE112] Flot examples, 2012. Disponível em <http://people.iola.dk/olau/flot/examples/>, acessado a última vez em 07 de fevereiro de 2012.
- [FFu10] FFunction. What is data visualization?, 2010. Disponível em http://infosthetics.com/archives/2010/11/what_is_data_visualization.html, acessado a última vez em 20 de dezembro de 2011.
- [Fri08] Vitaly Friedman. Data visualization and infographics, 2008. Disponível em <http://www.smashingmagazine.com/2008/01/14/monday-inspiration-data-visualization-and-infographics/>, acessado a última vez em 19 de dezembro de 2011.
- [GD112] Gas & driving - new york times, 2012. Disponível em <http://www.nytimes.com/imagepages/2010/05/02/business/02metrics.html>, acessado a última vez em 05 de fevereiro de 2012.
- [GPD12] visualcomplexity.com | german party donations, 2012. Disponível em http://www.visualcomplexity.com/vc/project_details.cfm?id=741&index=741&domain=, acessado a última vez em 05 de fevereiro de 2012.
- [IA111] Information aesthetics, 2011. Disponível em <http://infosthetics.com/>, acessado a última vez em 23 de novembro de 2011.
- [JE112] Jess3, 2012. Disponível em <http://jess3.com/>, acessado a última vez em 11 de dezembro de 2011.
- [JFC12] Jfreechart, 2012. Disponível em <http://www.jfree.org/jfreechart/>, acessado a última vez em 07 de fevereiro de 2012.
- [JIT12] Javascript infovis toolkit, 2012. Disponível em <http://thejit.org/>, acessado a última vez em 07 de fevereiro de 2012.
- [JQS12] jquery sparklines, 2012. Disponível em <http://omnipotent.net/jquery.sparkline/>, acessado a última vez em 07 de fevereiro de 2012.
- [LAB⁺11] T Lammarsch, W Aigner, A Bertone, S Miksch, A Rind, S Miksch e G. Santucci. Towards a Concept how the Structure of Time can Support the Visual Analytics Process. In *Proc. Int. Workshop Visual Analytics (EuroVA 2011) in conjunction with EuroVis*, pages 9–12, 2011.
- [LE07] Ralph Lengler e M.J. Eppler. Towards a periodic table of visualization methods for management. *Proceedings of Graphics and Visualization in Engineering (GVE 2007)*, 2007.
- [LV07] Andrea Lau e Andrew Vande Moere. Towards a Model of Information Aesthetics in Information Visualization. In *2007 11th International Conference Information Visualization (IV '07)*, pages 87–92. IEEE, July 2007.

REFERÊNCIAS

- [McC09] David McCandless. Interesting, easy, beautiful, true?, 2009. Disponível em <http://www.informationisbeautiful.net/2009/interesting-easy-beautiful-true/>, acessado a última vez em 19 de dezembro de 2011.
- [MS03] W. Muller e H. Schumann. Visualization methods for time-dependent data - an overview. In *Proceedings of the 2003 International Conference on Machine Learning and Cybernetics (IEEE Cat. No.03EX693)*, volume 1, pages 737–745. IEEE, 2003.
- [RGr] Rgraph: Html5/javascript charts for your website. Disponível em <http://www.rgraph.net/examples/index.html>.
- [SC00] S.F. Silva e Tiziana Catarci. Visualization of linear time-oriented data: a survey. In *Proceedings of the First International Conference on Web Information Systems Engineering*, volume 1, pages 310–319. IEEE Comput. Soc, 2000.
- [Sen] Html5 charts for mobile devices. Disponível em <http://www.sencha.com/products/touch/charts>.
- [Shn96] Ben Shneiderman. *The eyes have it: a task by data type taxonomy for information visualizations*. IEEE Comput. Soc. Press, 1996.
- [Svlg12] Svg vs canvas performance, 2012. Disponível em <http://joeloughton.com/blog/web-applications/svg-vs-canvas-performance/>, acessado a última vez em 11 de junho de 2012.
- [TM04] M. Tory e T. Moller. Rethinking visualization: A high-level taxonomy. In *Information Visualization, 2004. INFOVIS 2004. IEEE Symposium on*, pages 151–158. IEEE, 2004.
- [Tuf86] Edward R. Tufte. *The visual display of quantitative information*. Graphics Press, Cheshire, CT, USA, 1986.
- [WAM01] Marc Weber, M. Alexa e Wolfgang Müller. Visualizing time-series on spirals. In *proceedings of the IEEE Symposium on Information Visualization*, page 7, 2001.
- [Yau11] Nathan Yau. *Visualize This: The FlowingData Guide to Design, Visualization, and Statistics*. Wiley Publishing Inc., Indianapolis, IN, USA, 2011.