



Licenciatura em Gestão de Sistemas e Computação

## **Visualização de Informação através de gráficos vetoriais normalizados**

Trabalho final para Laboratório de GSC Aplicado

Elaborado por João Pedro Pereira Batista

Aluno nº 20091233

Orientador: Prof. Doutor José Vasconcelos

Barcarena

Fevereiro 2016

Universidade Atlântica



Licenciatura em Gestão de Sistemas e Computação

## **Visualização de Informação através de gráficos vetoriais normalizados**

Trabalho final para Laboratório de GSC Aplicado

Elaborado por João Pedro Pereira Batista

Aluno nº 20091233

Orientador: Prof. Doutor José Vasconcelos

Barcarena

Fevereiro 2016

Universidade Atlântica

O autor é o único responsável pelas ideias expressas neste relatório

## **Pensamento**

*The purpose of computing is insight, not numbers*

Richard Hemming, Numerical Methods for Scientists and Engineers, 1962

*Whatever can be expressed in numbers, may be expressed by lines*

William Playfair, 1805

*One hundred rumors are not comparable to one look*

Old Chinese inscription

## **Dedicatória**

Este trabalho é dedicado a todos os alunos da Universidade New Atlântica, para que em cada e todos os dias se empenhem arduamente na conquista dos seus objetivos, e para que esses possam resultar em valor acrescido, para si próprios, para a Universidade, e para as suas famílias.

Gostaria de agradecer a todo o corpo docente do curso de Licenciatura em Gestão de Sistemas e Computação, todo o tempo, paciência, e apoio ao longo de três anos, e agora por fim neste trabalho final.

Agradeço aos meus pais que sempre acreditaram que seria capaz, que me ensinaram a ser e a fazer mais e melhor, e que sempre me motivaram a frequentar e concluir com sucesso esta Licenciatura, já em idade adulta.

Gostaria de registrar uma palavra de enorme agradecimento ao meu irmão, sem ele eu não seria o matemático que sou hoje, e que sempre me ajudou em todas as dificuldades cognitivas com que fui lutando ao longo dos três anos de licenciatura, algumas das vezes, bem sei, com prejuízo para a sua própria vida pessoal e profissional. Obrigado!

Esta dedicatória não faria sentido, sem um registo a quem me deu a força necessária para me candidatar à Universidade, e foi necessária muita, à minha querida e linda mulher que sempre me apoiou, muitas foram as noites e as manhãs de sonos trocados e conversas perdidas, não teria sido capaz sozinho.

Este trajeto e trabalho final compensaram por fim!

## Lista de Abreviaturas /Siglas

API	<i>Application Program Interface</i> , ou interface de programação de aplicações, é um conjunto de rotinas, protocolos e ferramentas para a construção de aplicações.
CAE	Código de Atividade Económica
CNAEF	Código Nacional das Áreas de Educação e Formação
CSS	Cascade Style Sheet, é uma linguagem de folha de estilo usado para descrever a apresentação de um documento escrito em uma linguagem de marcação.
CSV	Formato de ficheiro - valores separados por ponto e vírgula, é muitas vezes usado para troca de dados entre aplicações distintas;
D3	Data-Drive Document, ou documento orientado a dados
DGEEC	Direção-Geral de Estatísticas da Educação e Ciência
DOM	Document Object Model, é uma convenção multiplataforma e independente para representar e interagir com objetos em HTML , XHTML, e documentos XML
GeoJSON	Formato para codificação estruturas de dados de informação geográfica
GeoVis	<i>Geographic Visualisation</i> , ou Visualização geográfica
GIF	<i>Graphics Interchange Format</i> , é um formato de imagem bitmap
GIS	<i>Geographic Information System</i> - Sistema de informação geográfica ou sistema de informação geográfica (GIS) é um sistema projetado para capturar, armazenar , manipular, analisar , gerenciar e apresentar todos os tipos de dados espaciais ou geográficas.
HSB	<i>hue, saturation and brightness</i>
HSV	<i>hue, saturation and value</i>
HTML5	HTML5 é uma linguagem utilizada para estruturar e apresentar conteúdos na <i>World Wide Web</i>
INE	Instituto Nacional de Estatística
InfoVis	Information Visualization, Visualização de Informação
JDK	<i>Java Development Toolkit</i>
JPEG	JPEG é uma técnica de compressão com perdas para imagens coloridas . Embora se possa reduzir o tamanho dos ficheiros a cerca de 5 % do seu tamanho normal , algum pormenor é perdido na compressão
JSON	SON ( <i>JavaScript Object Notation</i> ) é um formato de troca de dados simplificado.
NUTS	Nomenclatura das Unidades Territoriais Estatísticas
SciVis	Visualização científica

SDLC	<i>Systems Development Life Cycle</i> , ou ciclo de desenvolvimento de sistemas de informação.
SVG	<i>Scalable Vector Graphics</i>
W3C	<i>World Wide Web Consortium</i>
XML	<i>Extensible Markup Language</i>
XSL	<i>EXtensible Stylesheet Language</i>



## **Resumo**

A Visualização de Informação corresponde à produção de imagens com recurso a computador, interativas, e a representações visuais de dados, nas quais se utilizam várias técnicas de Computação Gráfica e de Interface Homem-Máquina para amplificar a cognição humana sobre a informação. É o estudo de representações visuais, interativas, e que tem como objetivo auxiliar o processo de análise e compreensão de um conjunto de dados, através de representações gráficas suscetíveis de serem manipuladas. Neste trabalho realiza-se um enquadramento teórico para identificar o modelo de referência amplamente reconhecido na visualização de informação, e em seguida procura-se testá-lo, através do desenvolvimento de um protótipo funcional com recurso a representações visuais.

Palavras-chave: Visualização de Informação, Gráficos Vetoriais, Documentos orientados a dados, Ensino Superior

## **Abstract**

Information Visualization corresponds to the production of interactive images using a computer, and visual representations of data, in which various techniques are used of Computer Graphics and Human-Machine Interface in order to amplify human perception of the information. It is the study of interactive visual representations aiming to assist in the process of analysis and understanding of a dataset, using graphical representations which can be manipulated. This paper presents a literature review to identify the widely accepted reference model in the information display and to identify common information visualization techniques. In addition to the literature review, the development of a working prototype using visual representations is also aimed at demonstrating the developed model.

Keywords: Information Visualization, Vectorial Graphics, Data-Driven Documents, Higher Education

## Índice

1. Introdução .....	1
1.1. Problema e motivação .....	1
1.2. Tema .....	2
1.3. Motivos para Visualizar .....	3
1.4. Enquadramento organizacional .....	4
1.5. Objetivos .....	4
1.6. Estrutura do documento .....	5
2. Metodologia .....	6
3. Revisão da Literatura .....	8
3.1. Visualização .....	8
3.1.1. SciVis: <i>Scientific Visualisation</i> .....	8
3.1.2. GeoVis: <i>Geographic Visualisation</i> .....	8
3.1.3. InfoVis: <i>Information Visualisation</i> .....	10
3.2. Princípios gerais da Visualização de Informação .....	11
3.2.1. Perceção visual humana .....	11
3.2.2. Visualizar e interagir .....	11
3.2.3. Focus-plus-Context .....	13
3.2.4. Overview-plus-detail .....	13
3.3. Finalidades da Visualização de Informação .....	14
3.4. Modelo de Referência .....	15
3.4.1. Fase 1 - Transformação dos dados .....	15
3.4.2. Fase 2 - Mapeamento visual .....	16
3.4.3. Fase 3 - Transformações visuais .....	16
3.5. Caracterização de Dados .....	17
3.6. Atributos gráficos .....	18
3.6.1. Cor .....	18
3.6.2. Movimento .....	20
3.6.3. Anotações .....	20
3.7. Representações Visuais .....	21
3.7.1. Biblioteca gráfica – d3.js .....	22
3.7.1.1. Sunburst .....	24

3.7.1.2.	Chord.....	24
3.7.1.3.	<i>Sankey</i> .....	26
3.7.1.4.	<i>Dendogram</i> .....	26
3.7.2.	Biblioteca gráfica – <i>dimple.js</i> .....	27
3.8.	Gráficos vetoriais normalizados.....	31
3.8.1.	Camadas ( <i>layers</i> ).....	32
3.8.2.	Séries cronológicas .....	33
4.	Domínio .....	34
4.1.	Caracterização do ensino superior.....	34
4.1.1.	Alunos Inscritos .....	35
4.1.2.	Alunos Diplomados .....	36
4.2.	Transferência de valor para o mercado de trabalho .....	36
4.2.1.	Setores de atividade económica.....	38
4.2.2.	Volume de negócios por setor de atividade.....	38
4.2.3.	Relação de áreas de educação e formação com setores de atividade.....	40
4.2.4.	Análise por setor económico.....	41
5.	Desenvolvimento .....	42
5.1.	Ciclo de desenvolvimento .....	42
5.2.	Requisitos .....	44
5.3.	Ambiente de desenvolvimento .....	44
5.4.	Protótipo .....	45
5.5.	Métodos clássicos.....	48
6.	Conclusões.....	50
6.1.	Interpretação de resultados.....	50
6.2.	Oportunidades de melhoria futura.....	51

## Índice de Figuras

Gráfico 1 - Processo de transformação de dados em sabedoria.....	2
Gráfico 2 - Uma visualização científica de uma simulação de uma instabilidade Rayleigh - Taylor causada pela mistura de dois fluídos (fonte: <a href="https://goo.gl/xeIXOJ">https://goo.gl/xeIXOJ</a> ) ..	8
Gráfico 3 - Evolução de desemprego desde 2008 (fonte: <a href="http://goo.gl/pdYJH3">http://goo.gl/pdYJH3</a> ) .....	9
Gráfico 4 - Representação gráfica: <i>Treemap</i> (fonte: <a href="http://d3js.com">d3js.com</a> ).....	10
Gráfico 5 - Vista magnificada (fonte: <a href="http://goo.gl/UAydnE">http://goo.gl/UAydnE</a> ).....	13
Gráfico 6 - Exemplo de vista geral, e detalhe (fonte: <a href="http://goo.gl/FVWr56">http://goo.gl/FVWr56</a> ).....	13
Gráfico 7 - Modelo de referência de visualização de informação (Card et Al, 1999)....	15
Gráfico 8 - Gradação de saturação nos modelos HSL e HSV (fonte: wikipedia) .....	19
Gráfico 9 – Representação gráfica: <i>Sunburst</i> ( <a href="http://goo.gl/uguxjh">http://goo.gl/uguxjh</a> ).....	24
Gráfico 10 - Representação gráfica - Sequences subburst (fonte: <a href="http://goo.gl/uguxjh">http://goo.gl/uguxjh</a> ). 24	
Gráfico 11 - Representação gráfica: Chord Diagrams - exemplos (fonte: <a href="http://goo.gl/cnbYD">http://goo.gl/cnbYD</a> ).....	25
Gráfico 12 - Representação gráfica: Chord Diagram (Uber rides, fonte: <a href="http://goo.gl/TMJER">http://goo.gl/TMJER</a> ) .....	25
Gráfico 13 - Representação gráfica: Sankey Diagram.....	26
Gráfico 14 – Representação gráfica (dendograma – protovis, fonte: <a href="http://goo.gl/YkiwKT">http://goo.gl/YkiwKT</a> ) .....	27
Gráfico 15 - Representação gráfica - Radial dendogram (fonte: <a href="http://goo.gl/Z3Wnmd">http://goo.gl/Z3Wnmd</a> ) .....	27
Gráfico 16 - Gráficos de barra - <i>dimple.js</i> (fonte: <a href="http://dimplejs.org">http://dimplejs.org</a> ) .....	28

Gráfico 17 - Gráficos de tarte - <i>dimple.js</i> (fonte: <a href="http://dimplejs.org">http://dimplejs.org</a> ) .....	29
Gráfico 18 - Gráficos de dispersão - <i>dimple.js</i> (fonte: <a href="http://dimplejs.org">http://dimplejs.org</a> ) .....	29
Gráfico 19 - Gráficos de área - <i>dimple.js</i> (fonte: <a href="http://dimplejs.org">http://dimplejs.org</a> ) .....	30
Gráfico 20 - Gráficos de linha - <i>dimple.js</i> (fonte: <a href="http://dimplejs.org">http://dimplejs.org</a> ).....	30
Gráfico 21 - Comparativo entre gráfico raster e vetorial.....	31
Gráfico 22 - Camadas em SVG (fonte: <a href="http://apike.ca/prog_svg_layering.html">http://apike.ca/prog_svg_layering.html</a> ) .....	33
Gráfico 23 - SLDC - Modelo em cascata (fonte: <a href="https://goo.gl/NgIqHJ">https://goo.gl/NgIqHJ</a> ) .....	43
Gráfico 24 - Ambiente de <i>debug</i> - Mozilla Firefox .....	45
Gráfico 25 - Resultado protótipo .....	46
Gráfico 26 - Gráficos de barra e tarte - Excel.....	48
Gráfico 27 - Gráfico em barra - Excel - alternativa .....	48
Gráfico 28 - Gráfico <i>donut</i> - Excel .....	49

## Índice de Tabelas

Tabela 1 - Metodologias adotadas nos objetivos de investigação .....	6
Tabela 2 - Sumário de caracterização de dados.....	17
Tabela 3 - Prioridades dos atributos visuais versus tipo de dado (quantitativo, ordinal, nominal).....	18
Tabela 4 - Classes de representações visuais.....	21
Tabela 5 - Variáveis de análise de alunos inscritos (fonte: DGEEC).....	35

Tabela 6 - Variáveis de análise de alunos diplomados (fonte: DGEEC).....	36
Tabela 7 - Metainformação do indicador: Volume de negócios por setor de atividade .	37
Tabela 8 - Relação de setores de atividade económica e principais setores económicos	39
Tabela 9 - Valor total e médio por empresa /principal setor de atividade económica....	39
Tabela 10 - Tabela de relação de áreas de estudo (CNAEF) com setores de atividade económica .....	40
Tabela 11 - Correspondência entre nº médio de alunos diplomados e volume de negócio, em média.....	41
Tabela 12 - Resultado da agregação de dados, para o ano letivo 2014-15 .....	46
Tabela 13 – Critérios utilizados para escolha de cor de cada dimensão.....	47

## **1. INTRODUÇÃO**

Nos mais diversos domínios em que existe a necessidade de compreender qualquer fenómeno mensurável, existe a necessidade recolher dados e analisar informação estatística sobre o mesmo.

O que testemunhamos nos dias de hoje é uma explosão de dados e não uma explosão de informações, dados que nos são continuamente apresentados, e os quais somos continuamente pressionados a observar, processar e a interpretar (Mazza, 2009). Os dados têm diversas origens, e aos quais atribuímos valor e importância variada. Neste cenário, a informação é construída a partir de um fluxo contínuo de dados a que somos sujeitos passivamente ou ativamente.

Torna-se necessária a definição de métodos eficazes que permitam analisar os dados, e a obter destes informação crítica, permitindo assim que sejam auxiliares valiosos em processos de decisão.

### **1.1. Problema e motivação**

A visualização dos resultados numéricos após seu tratamento estatístico não permite uma análise fácil quando o volume de dados é grande ou ainda quando eles apresentam um elevado grau de complexidade (muitas variáveis e dimensões, por exemplo). Nesses casos, mesmo com a utilização de gráficos com tipologias mais clássicas (barras, linhas, etc.), a limitação imposta pelas características desse tipo de gráficos, torna a interpretação difícil, exigindo a construção de vários gráficos complementares e causando um processo cognitivo “forçado” e difícil (Mazza, 2009).

Torna-se necessário então adotar metodologias que permitam incrementar a objetividade e clareza dos dados, bem como acelerar o acesso à informação, face aos métodos clássicos.

## 1.2. Tema

A visualização de informação pode ser uma abordagem diferenciadora neste contexto pois o seu objetivo é transformar um conjunto de dados abstratos em representações gráficas e interativas de forma a auxiliar a análise e compreensão dos dados e na extração de novas informações (Nascimento, 2005).

A Visualização de Informação corresponde à produção de imagens com recurso a computador, interativas, e a representações visuais de dados (Card, Mackinlay and Shneiderman, 1999), nas quais se utilizam várias técnicas de Computação Gráfica e de Interface Homem-Máquina para amplificar a cognição humana sobre a informação (Card, 2007). A Cognição corresponde ao processo de receber, processar, armazenar, e utilizar a informação (Heffner, 2003).

A visualização de informação situa-se entre os dados e a informação, e corresponde a uma atividade cognitiva, que providencia os métodos e ferramentas para organizar e representar dados, e que permite a construção de representações mentais baseada em representações visuais externas (Mazza, 2009).

**Gráfico 1 - Processo de transformação de dados em sabedoria**



O objetivo básico e principal a qualquer visualização é de a partir de um grande volume de dados, extrair o máximo de informação de uma forma rápida, clara, precisa, e eficiente (Ware, 2004). Para isso são utilizadas metáforas visuais adequadas, também denominadas por representações visuais.



O sucesso do desenho de sistemas de visualização de informação depende da correta caracterização da tarefa, do sistema visual humano, das apresentações visuais, e da interação entre todos os anteriores (Ware, 2004), em que se inclui a definição da metáfora visual adequada ao potencial utilizador, à origem e tipo dos dados, ou ao tipo de ênfase desejado ao resultado final.

### 1.3. Motivos para Visualizar

Alguns dos motivos pelos quais os humanos preferem a visualização de dados (D'Orazio, 2012):

1. Colocar as informações no espaço, tornando-as tangíveis e permitindo-nos pensar com olhos, é do agrado dos humanos porque a perceção e cognição do mundo é inerentemente informado pelo espaço;
2. Objetivar informações abstratas em formas, superfícies, volumes e cores;
3. Classificar e Comparar dados, entidades e distribuições;
4. Eles agem como uma memória externa, que nos permite ter em conta um maior número de variáveis e hipóteses, e de se mover facilmente entre raciocínio focado e associações livres;
5. A nossa capacidade de identificar padrões e correlações quando lidamos com números é incrivelmente mais pobre que nossa capacidade de reconhecer e comparar formas;
6. Manipulação direta de objetos em dados dinâmicos. Exemplo: redimensionar uma forma gráfica, como um retângulo, arrastando seus cantos ou bordas com um rato de computador;
7. Iteração contínua: permite incluir um conjunto de hipóteses mais alargadas para a resolução de um problema;

#### **1.4. Enquadramento organizacional**

Em todas as organizações é necessária a análise de dados, sejam de produção, por forma a orientar a sua otimização, sejam de informação do mercado interno ou externo, permitindo assim a estas a inteligência de se posicionarem de forma competitiva num contexto global.

O domínio deste trabalho trata da caracterização do ensino superior nos últimos anos, e na transferência de valor do ensino superior para o mercado de trabalho, em particular no que se refere à adequação da oferta de cursos às necessidades do mercado de trabalho.

Este é um tema sobre o qual se desperta bastante curiosidade, levantando com frequência questões particulares na área da consultoria informática, em gestão de processos, e informação. No âmbito do presente trabalho a área de enfoque particular é a do Ensino Superior

#### **1.5. Objetivos**

A investigação no âmbito do trabalho proposto tem como objetivos:

1. Identificar as técnicas e metáforas visuais mais utilizadas na Visualização de Informação;
2. Compreender a possibilidade de relacionar variáveis de dados com propriedades gráficas que compõem uma representação visual (cor, forma, dimensão, etc.);
3. Compreender o fenómeno de apresentação de dados em camadas (layers) gráficas, através de imagens vetoriais normalizadas;
4. Avaliar a aplicabilidade da representação de séries de dados cronológicas por meio de técnicas de transformação e transição de imagens vetoriais normalizadas;
5. Desenvolvimento de protótipo de geração de representações visuais com recurso à utilização de bibliotecas gráficas;

## **1.6. Estrutura do documento**

Capítulo 1 – Introdução, enquadramento do problema e respetiva motivação. É identificado o tema em questão e o seu enquadramento organizacional, bem como a definição dos objetivos do trabalho;

Capítulo 2 – Apresentação da metodologia utilizada, para atingir cada um dos objetivos identificados;

Capítulo 3 – Revisão de literatura, identificando quais as variantes de visualização, princípios gerais para a visualização de informação, quais as finalidades da visualização, modelo de referência, e quais as características dos dados e como podem estes ser representadas;

Capítulo 4 – Desenvolvimento do domínio com apresentação das variáveis de dados disponíveis bem como o modelo de análise proposto;

Capítulo 5 - Descrição do ambiente de desenvolvimento. Descrição de partes constituintes do protótipo, e modelo de funcionamento. Apresentação de gráficos de acordo com os métodos clássicos sobre as mesmas variáveis de dados em análise;

Capítulo 6 – Conclusões, interpretações de resultados e oportunidades de melhoria;

## 2. METODOLOGIA

Para atingir os objetivos propostos será necessário definir e aplicar diferentes estratégias, através de diferentes metodologias e ferramentas. Estas encontram-se identificadas e descritas em seguida, para cada um dos objetivos. Como objetivo final releva-se o desenvolvimento de um protótipo que procura materializar a revisão de literatura realizada e ainda o domínio em análise.

**Tabela 1 - Metodologias adotadas nos objetivos de investigação**

<b>Objetivo</b>	<b>Metodologia</b>	<b>Ferramentas</b>
Identificar as técnicas e metáforas visuais mais utilizadas na Visualização de Informação	Indutivo	Revisão da Literatura
Compreender a possibilidade de relacionar variáveis de dados com propriedades gráficas que compõem uma representação visual	Indutivo	Revisão da Literatura
Compreender o fenómeno de apresentação de dados em camadas ( <i>layers</i> ) gráficas, através de imagens vetoriais normalizadas	Descritivo	Revisão da Literatura
Avaliar a aplicabilidade da representação de séries de dados cronológicas por meio de técnicas de transformação e transição de imagens vetoriais normalizadas	Correlacional	Revisão da Literatura / Protótipo
Desenvolvimento de protótipo de geração de representações visuais com recurso à utilização de bibliotecas gráficas	Experimental	Desenvolvimento de Protótipo

Identificação das metodologias utilizadas (Barañano, 2008):

- Indutivo “O método indutivo defende que na investigação se deve começar por uma observação para que, no final de um processo, se possa elaborar uma teoria [...] o raciocínio indutivo faz -se do particular para o geral.”
- Descritivo “Este método assenta em estratégias de pesquisa para observar e descrever comportamentos, incluindo a identificação de fatores que possam estar relacionados com um fenómeno em particular.”
- Correlacional “Este método tem como objetivo medir o grau e a direção de uma relação entre duas variáveis que é expressa por um coeficiente de correlação, ou ainda, quando uma teoria científica postula a existência de uma correlação entre duas variáveis que não podem ser analisadas experimentalmente.”
- Experimental “A investigação experimental, sendo um método quantitativo, é objetiva, orientada para o resultado e a para a comprovação, fíável – significando com isso que os dados se apresentam «sólidos» e repetíveis – e generalizável.”

### 3. REVISÃO DA LITERATURA

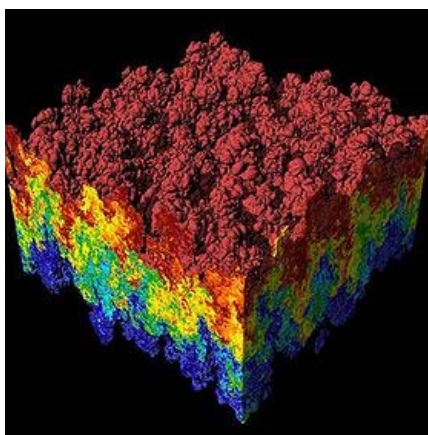
#### 3.1. Visualização

A grande área da Visualização é composta pelas três principais subáreas (Mazza, 2009), descritas nos subcapítulos seguintes.

##### 3.1.1. SciVis: *Scientific Visualisation*

O objetivo desta visualização é ilustrar graficamente dados científicos para permitir aos cientistas compreender, ilustrar, e recolher uma visão dos dados. A representação visual é sugerida por uma forma geométrica que está intrinsecamente associada aos dados (*flows, volumes, surfaces*).

**Gráfico 2 - Uma visualização científica de uma simulação de uma instabilidade Rayleigh - Taylor causada pela mistura de dois fluidos (fonte: <https://goo.gl/xelXOJ>)**

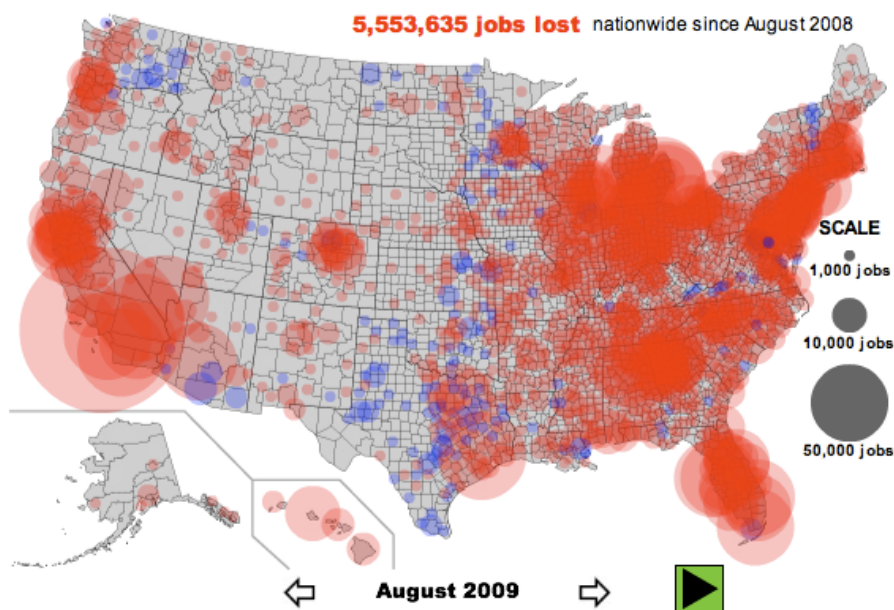


##### 3.1.2. GeoVis: *Geographic Visualisation*

Refere-se a um conjunto de ferramentas e técnicas de análise de dados de suporte geoespacial através da utilização de visualização interativa.

Tal como as subáreas de visualização: Visualização Científica (SciVis) e Visualização da Informação (InfoVis), a geovisualização enfatiza a construção do conhecimento sobre os dados através de representações visuais, para isso a geovisualização comunica informações geoespaciais, através de mapas, de forma que, quando combinado com a compreensão humana, permitem exploração de dados e processos de tomada de decisão.

Gráfico 3 - Evolução de desemprego desde 2008 (fonte: <http://goo.gl/pdYJH3>)



Os mapas estáticos tradicionais têm uma capacidade limitada exploratória enquanto que os sistemas de informação geográfica (GIS) permitem a construção de mapas mais interativos incluindo a capacidade de explorar diferentes camadas do mapa, para ampliar ou reduzir detalhe dos mesmos, e para alterar o aspeto visual do mapa.

Geovisualização representa um conjunto de tecnologias e práticas cartográficas que, em conjunto com a capacidade de processamento dos computadores modernos, permitem efetuar alterações aos mapas em tempo real.

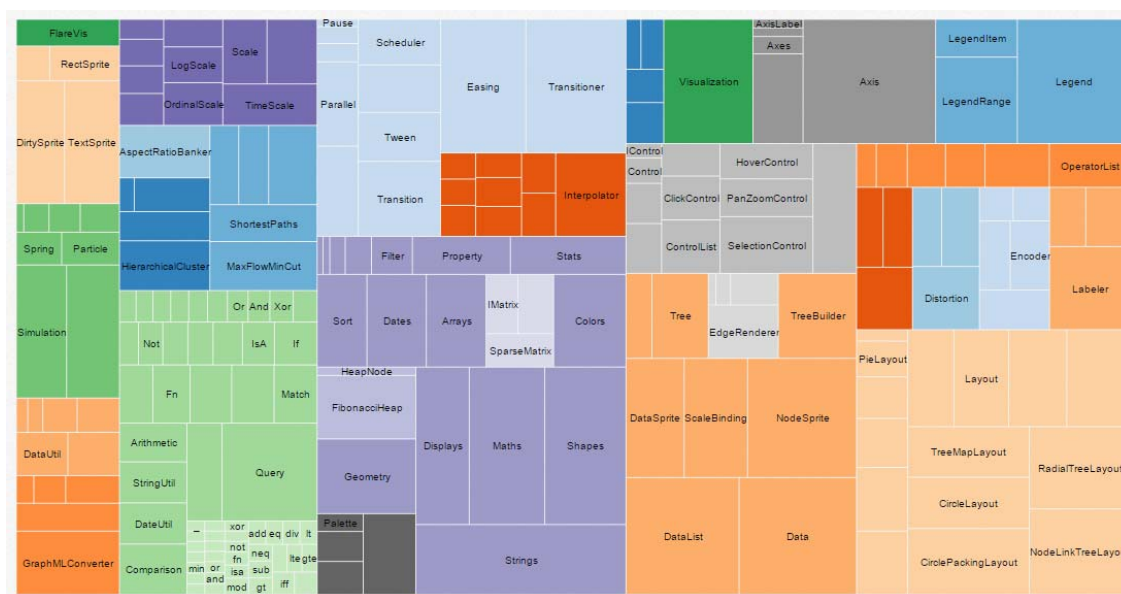
### 3.1.3. InfoVis: Information Visualisation

Visualização de informação é o estudo das representações visuais, interativas, de dados abstratos, para reforçar a cognição humana (Mazza, 2009). Os dados tanto podem ser do tipo numérico como não numérico, tais como texto e informação geográfica. A visualização da informação difere de visualização científica, porque não é definida *a priori* uma representação geométrica, e da geográfica, quando a informação não é representável no espaço.

DataVis: *Data Visualisation* = InfoVis + GeoVis.

No trabalho proposto será explorada a área de visualização de informação – InfoVis.

Gráfico 4 - Representação gráfica: *Treemap* (fonte: d3js.com)



Na visualização de dados e de informação, *treemapping* é um método muito utilizado para representar dados que se relacionam hierarquicamente usando retângulos de diferentes dimensões (Carvalho and Fernandes, 2009). Cada ramo da árvore é definido por um retângulo, preenchido com retângulos menores lado a lado, os quais representam, por sua vez, sub-ramos. Cada retângulo tem uma área proporcional a uma variável especificada nos dados. Muitas vezes, os ramos são coloridos para mostrar valores possíveis de uma variável de dados. Como as dimensões de cor e tamanho são



correlacionados com a estrutura da árvore (ou do treemap) podem-se, muitas vezes, ver padrões que seriam difíceis de detetar em outras formas de visualização. Uma outra vantagem do *treemapping* é que, por construção, eles fazem uso eficiente do espaço, e como resultado podem exibir de forma legível uma grande quantidade de dados.

## **3.2. Princípios gerais da Visualização de Informação**

### **3.2.1. Perceção visual humana**

A visualização de informação permite a apresentação de dados em formas gráficas de modo a que o utilizador possa utilizar a sua perceção visual para melhor analisar e compreender as informações (Card, Mackinlay and Shneiderman, 1999). Os humanos têm excelentes capacidades perceptivas:

1. Para observar, reconhecer e recordar imagens rapidamente.
2. Para detetar rapidamente e automaticamente padrões e mudanças no tamanho, cor, forma, movimento ou textura.

A visualização de informação visa apresentar informações visualmente, no essencial para diminuir o esforço do trabalho cognitivo para o sistema de perceção visual humana.

### **3.2.2. Visualizar e interagir**

O suporte à interação com o utilizador é tão importante como a representação visual subjacente, através de funções como: *overview*, *zoom* e filtro (Shneiderman, 1996). Uma representação visual estática, por si só, não é suficiente para disponibilizar as condições ideais necessárias à compreensão de grandes conjuntos de dados. Por norma são disponibilizadas funções através das quais o utilizador pode explorá-los. A aplicação destas funções poderá, ocasionalmente, provocar alterações nas representações visuais.

As funções disponíveis ao utilizador para exploração dos dados podem ser classificadas em três níveis (Freitas et al., 2001):

1. Nível um – básico

Navegação por deslocação horizontal /vertical, quando em representação plana, ou por rotação dos dados quando em representação 3D.

2. Nível dois – secundário

Funções de seleção de dados de interesse que podem provocar um reposicionamento do conjunto de dados para uma melhor visualização, ou ainda a realização de *zoom* (semântico) que permite aumentar o detalhe de uma parte do conjunto de dados, e supressão do detalhe do restante conjunto. Funções: poda (*prunning*), agrupamento (*clustering*), ou de expansão.

Nestas operações podem ocorrer três situações:

- a. A representação é alterada para mostrar apenas a região de interesse selecionada;
- b. A região de interesse ocupa o campo de visão principal e o restante conjunto de dados é mantido em área à parte (*Focus-plus-Context*);
- c. A região de interesse e a visão geral são exibidas em simultâneo (*Overview-plus-Detail*);

3. Nível três – superior

Funções de filtro que permitem representar visualmente apenas partes do conjunto de dados. Podem ser simples filtros de dados. Este nível pode-se considerar fora do âmbito da visualização e contido no nível aplicacional.

As transições, quando utilizadas, devem ser animadas suavemente, durante cerca de 1 seg. - objeto constante elimina a necessidade de reassimilação da imagem (Robertson, Jock and Card, 1991).

### 3.2.3. Focus-plus-Context

Esta função de seleção de dados foca-se nas áreas de interesse, permitindo o detalhe máximo de informação disponível da área selecionada, enquanto se mantêm o contexto de informação adjacente, com significativamente menos detalhe. É exemplo disso, a vista magnificada, ou de “olho de peixe”: que corresponde a uma distorção geométrica que resulta no zoom de uma determinada área de interesse (Furnas, 1981).

Gráfico 5 - Vista magnificada (fonte: <http://goo.gl/UAydne>)



### 3.2.4. Overview-plus-detail

Esta função de seleção de dados corresponde ao modelo de apresentação da informação em dois ecrãs, separados, mas sincronizados, uma que corresponde a uma vista geral e abrangente e uma outra que corresponde à vista detalhada.

Gráfico 6 - Exemplo de vista geral, e detalhe (fonte: <http://goo.gl/FVWr56>)



### **3.3. Finalidades da Visualização de Informação**

A Visualização de Informação tem como finalidade potencializar a apropriação de informação pelo utilizador, por meio de recursos gráficos. A visualização de informação é uma área de aplicação de técnicas de computação gráfica interativas, que pretendem auxiliar a análise e a compreensão de um conjunto de dados (Freitas et al., 2001).

A utilização dos princípios de visualização de informação, através da construção e mapeamento de representações visuais permite (Mazza, 2009):

1. Entender sistemas complexos e dados;
2. Tomar decisões;
3. Procurar informação;
4. Comunicar (ideias, conceitos, ou propriedades dos dados).

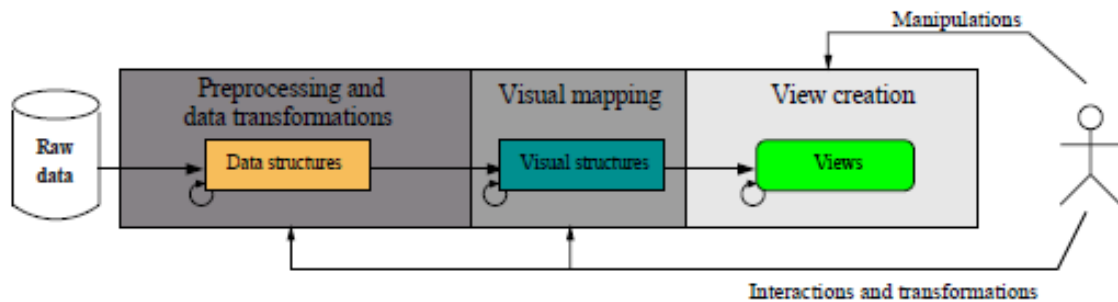
Uma representação visual tem uma de duas utilizações (Mazza, 2009):

1. Análise exploratória de dados
  - a. Identificar propriedades dos dados
  - b. Relações entre os dados
  - c. Identificar regularidades ou padrões
  - d. Identificar possíveis tendências
2. Análise confirmatória
  - a. Relações entre séries de dados
  - b. Confirmar ou inferir hipóteses

### 3.4. Modelo de Referência

O processo de computação que converte a informação numa visualização é conhecido como modelo de referência para a visualização, e é composto por três fases:

Gráfico 7 - Modelo de referência de visualização de informação (Card et Al, 1999)



#### 3.4.1. Fase 1 - Transformação dos dados

Nesta os dados em bruto são transformados, se necessário ou aplicável, através de funções de grupo, podem ser efetuadas conversões ou outras operações matemáticas, que se identificam necessárias para a obtenção dos dados a serem mapeados visualmente.

Nesta fase, e no contexto da transformação dos dados, poderá ser aplicável a computação de estatística sobre os dados em bruto. A Estatística é a ciência que recolhe, organiza, sumariza, e apresenta dados (estatística descritiva) ou analisa e realiza inferências sobre os dados (estatística inferencial) (Jaisingh, 2000).

O resultado corresponde ao armazenamento dos dados em estruturas de dados pré-definidas como suporte à representação de visual pretendida.

### **3.4.2. Fase 2 - Mapeamento visual**

Está relacionado com a escolha da representação visual para mapear os dados armazenados na tabela.

É nesta fase de mapeamento visual que são aplicadas as técnicas de visualização, que procuram representar graficamente dados de um determinado domínio, de modo a que a representação visual gerada explore a capacidade de perceção do homem e este, a partir das relações espaciais exibidas, interprete e compreenda as informações apresentadas e deduza novos conhecimentos (Freitas et al., 2001)

### **3.4.3. Fase 3 - Transformações visuais**

Consistem nas possíveis interações que o utilizador realiza sobre a visualização.

Na fase inicial de definição de sistemas de visualização de informação, deve-se considerar a melhor forma de mapear informações através de uma representação gráfica que facilite a sua interpretação pelos utilizadores, bem como possibilitar diferentes formas de manipulação do conjunto de dados, tanto geométrica, sejam rotações ou de zoom da representação gráfica, bem como analiticamente, com a redução ou expansão do conjunto de dados exibido de acordo com critérios determinados pelo utilizador.

Uma das principais considerações a ser feita no processo de construção de visualizações é o de determinar qual a técnica que deve ser utilizada, em função do tipo de informação que se pretende analisar e explorar. Alguns autores propõem classificações de técnicas visuais por forma a auxiliar a identificação de quando utilizar que técnica. Shneiderman (1996), classifica as técnicas de visualização por tipo de dados/tarefas, em função da dimensão do domínio (1D, 2D, 3D, ou n-D), dirigidas á visualização de hierarquias e de relacionamentos (grafos), suporte de tarefas como a obtenção de uma visão geral, obtenção de visão detalhada, *zooming*, filtragem, ou identificação de relacionamentos.

### 3.5. Caracterização de Dados

Os dados são valores, ou medidas, de variáveis que descrevem a ocorrência de um determinado evento, fenómeno, ou processo que é objeto de estudo ou de análise (Freitas et al., 2001).

Desta forma, os dados correspondem a atributos que podem ser caracterizados de acordo com diferentes critérios. A identificação dessas características irá influenciar na escolha de uma técnica de visualização (Ware, 2004).

**Tabela 2 - Sumário de caracterização de dados**

<b>Critério</b>	<b>Tipo</b>	<b>Exemplo</b>
Classe de informação, ou Tipo de Informação	Qualitativa Nominal	Género
	Qualitativa Ordinal	Habilitações Literárias
	Quantitativa	Temperatura
	Relacionamento	Link para um documento
Tipo de valores, ou Tipo de dados primitivo	Alfanumérico	Género
	Numérico (inteiro, real)	Temperatura
	Simbólico	Link para documento
Natureza do domínio	Discreto	Marcas de automóvel
	Contínuo	Superfície de terreno
	Contínuo-discretizado	Anos (tempo discretizado)
Dimensão do domínio	1D	Fenómeno que ocorre no tempo
	2D	Superfície de terreno
	3D	Volume de dados médicos
	n-D	Dados de uma população

Os dados podem conter informação numérica que discrimina quantidades ou uma ordem sequencial, ou ainda algum tipo de “rótulo” (label) que identifica uma classe ou conjunto. Por outro lado, cada dado pode ser considerado como sendo uma posição espacial num espaço de dimensão  $n$  (ponto), ou possuir um valor numérico (escalar). Os escalares também podem constituir conjuntos discretos de funções contínuas com 1 (linear), 2 (imagem), 3 (volumétrica) ou mais dimensões. Num vetor, cada elemento é considerado uma reta com um determinado comprimento, num espaço de dimensão  $n$ .

### 3.6. Atributos gráficos

De uma forma geral, algumas prioridades visuais podem ser observadas consoante o tipo de dado (Carvalho and Fernandes, 2009). A Tabela 3 apresenta algumas das situações existentes, estando a prioridade ordenada de forma crescente (1ª linha apresenta a maior, enquanto que a última a menor).

Tabela 3 - Prioridades dos atributos visuais versus tipo de dado (quantitativo, ordinal, nominal)

<b>Quantitativo</b>	<b>Ordinal</b>	<b>Nominal</b>
Posição	Posição	Posição
Comprimento	Densidade	Matiz
Ângulo	Saturação	Densidade
Inclinação	Matiz	Saturação
Área	Comprimento	Forma
Densidade	Ângulo	Comprimento
Saturação	Inclinação	Ângulo
Matiz	Área	Inclinação
Forma	Área	Área

#### 3.6.1. Cor

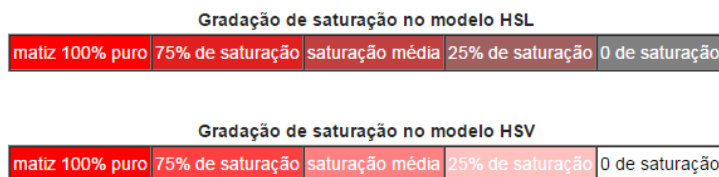
Não existe nenhuma teoria completa e é descrita em função da gama de cores (Kunni and Shinagawa, 1991). A sua dimensão em termos de perceção é caracterizada pelo seu matiz, saturação e intensidade/brilho (Healey, 1997).

O contexto visual onde as técnicas são inseridas, também é importante (Tufte, 1994). As cores devem ser utilizadas segundo os seus significados mais padronizados (de acordo com os hábitos do utilizador), o mesmo ocorre com a utilização de símbolos (*glyphs*) (Tullis, 1990).

Em colorimetria e também no modelo de cor HSV (ou HSB), saturação ou grau de pureza da cor é um parâmetro que especifica a qualidade de um matiz de cor pelo grau de mesclagem com a cor branca. Outras modalidades: matiz, brilho, e canal alfa.



**Gráfico 8 - Gradação de saturação nos modelos HSL e HSV (fonte: wikipedia)**



A estes três componentes, poderá ainda ser adicionado um quarto – Textura. Podem variar de forma independente ou não. Em termos gerais tem-se que:

Componente	Utilização efetiva
Matiz	Dados do tipo ordinal e nominal (escala de cores) (Keller and Keller, 1993)
Saturação	Dados do tipo ordinal
Brilho	Dados do tipo ordinal e quantitativo (Marcus and Dam, 1991)
Textura	Dados do tipo nominal (Cuccu and Moltedo, 1993)

Cuidados na sua utilização:

- Matiz
  - Azul (cores “frias”) – para valores afastados do ponto de observação, pequenos, negativos, ou associados com temperaturas baixas;
  - Vermelho (cores “quentes”) – para valores próximos, altos, positivos, associados a temperaturas altas ou algo de perigoso;
  - O matiz pode apresentar uma aparência diferente de acordo com a cor de fundo utilizada;
  - Descontinuidade nas escalas de cor;
- Saturação
  - Quando se deseja interpretar o brilho e a saturação independentemente.

- Brilho
  - Os objetos brilhantes num fundo escuro parecem ser maiores do que os objetos escuros num fundo claro.
  - A utilização de brilho gradiente fornece uma melhor interpretação da distância ou profundidade.
  - Alterações de brilho ou a utilização de um brilho total não são apreendidos linearmente.
  - O contraste do brilho influencia a perceção do brilho.
- Textura
  - Não sobrepor texturas.
  - As texturas podem dar origem a outras impressões, exemplo: densidade.

### **3.6.2. Movimento**

Devem ser mostrados numa velocidade mínima de 10 quadros/segundo. Piscar duas ou mais imagens para realçar diferenças ou similaridades.

### **3.6.3. Anotações**

A anotação é outro detalhe que também auxilia a interpretação de atributos visuais: Rótulos, títulos, legendas.

- Escala de cor e brilho, ou significado de texturas
- Escalas de distâncias que relacionam as coordenadas do mundo real com as existentes no ecrã.
- Símbolos de orientação (por ex. norte/sul, leste/oeste).

- Animações com indicação temporais: duração, momento, etc.

### 3.7. Representações Visuais

Representações visuais ou gráficas correspondem às "figuras" ou "imagens" utilizadas para representar o conjunto (ou subconjunto) de dados em análise (Mazza, 2009).

Além dos gráficos tradicionalmente utilizados para apresentação de dados como os gráficos de pontos, de linha, de barras, de tarte, e histogramas de frequência, que permitem observar relações entre atributos, uma série de representações gráficas mais ou menos complexas podem ser utilizadas para codificar através de elementos visuais (cores ou símbolos geométricos) tanto valores como relações entre elementos de dados.

**Tabela 4 - Classes de representações visuais**

Classe	Tipo	Utilização
Gráficos 2D, 3D	Pontos Circulares Linhas Barras Superfícies (para 3D)	Representação da distribuição dos elementos no espaço/domínio, representação da dependência ou correlação entre atributos
Ícones Glifos Objetos geométricos	Elementos geométricos 2D ou 3D diversos	Representação de entidades num contexto, representação de grupos de atributos de diversos tipos
Mapas	De pseudo-cores De Linhas De Superfícies De ícones, símbolos diversos	Representação de campos escalares ou categorias Representação de linhas de contorno de regiões, idem no espaço 3D Representação de grupos de atributos (categóricos. Escalares, etc.)
Diagramas	Nodos e Arestas	Representação de relacionamentos diversos: É-um É-parte-de, Comunicação, Sequência. Referência.

Quando a partir de um universo de dados, precisamos de ilustrar conceitos, ideias e propriedades intrínsecas a esses dados, o uso de representação visual oferece-nos uma ferramenta de comunicação válida. O desafio reside na definição da representação visual que permita efetivamente alcançar o seu objetivo.

Para Tufte, um dos principais estudiosos contemporâneos da disciplina, a excelência em gráficos estatísticos consiste em ideias complexas comunicadas com clareza, precisão e eficiência (Tufte, 1983).

### 3.7.1. Biblioteca gráfica – d3.js

D3.js, ou apenas D3 para *Data-Driven Documents*, é uma biblioteca *JavaScript* para produzir visualizações de dados, dinâmicas, interativas em ambiente de navegação de Internet. Utiliza gráficos vetoriais (SVG) para visualização gráfica das representações visuais em uso, bem como HTML5, e CSS para apresentação destes gráficos.

Esta biblioteca - D3.js permite um elevado controle sobre o resultado visual final, ao nível da interação e experiência do utilizador. A sua versão inicial 1.0 data de 18 de fevereiro de 2011, e a sua última versão 3.5.12, estável, data de 17 de dezembro de 2015, o que significa para um projeto open-source, que mantém grande atividade.

Sítio da Internet: <http://d3js.org>

Esta biblioteca está disponível ao abrigo da licença BSD que permite que o software distribuído sob a licença seja incorporado em produtos proprietários. Todo o trabalho desenvolvido com esta biblioteca pode ser incorporado em produtos com licença proprietária.

A biblioteca JavaScript D3.js usa funções JavaScript pré-construídas para selecionar elementos, criar objetos SVG, associar-lhes estilo (CSS), ou adicionar transições, efeitos dinâmicos ou textos de ajuda (*hints*). Grandes conjuntos de dados podem ser facilmente ligados a objetos SVG e a uma ou mais camadas, usando funções D3.js simples para gerar gráficos. Os dados podem estar em vários formatos, mais comumente JSON, valores separados por vírgula ( CSV ) ou GeoJSON, mas se necessário podem ser programados para ler outros formatos de dados.

A interface de programação (API) da biblioteca gráfica - D3.js contém largas centenas de funções, que podem ser agrupadas nos seguintes contextos lógicos:

1. Seleção (*Selections*)
2. Transições (*Transitions*)
3. Listas (*Arrays*)
4. Matemática /Cálculo (*Math*)
5. Cor (*Color*)
6. Escala (*Scales*)
7. Gráficos vetoriais (*SVG*)
8. Tempo (*Time*)
9. Apresentação (*Layouts*)
10. Geografia (*Geography*)
11. Geometria (*Geometry*)
12. Comportamento (*Behaviors*)

Critérios a ter em conta na escolha de uma representação visual:

1. Caracterização dos dados: numéricos, não-numéricos, etc;
2. Complexidade dos dados: número de variáveis;
3. Utilização do espaço disponível, representatividade dos dados;
4. Representação dos dados num eixo temporal;
5. Transições visuais necessárias, para representar alterações de dados, em função do tempo ou da inclusão /exclusão de uma ou mais variáveis de dados;
6. Funções de seleção e filtro a disponibilizar para criar foco sobre a totalidade dos dados disponíveis;
7. Interatividade do utilizador com a representação visual;

De entre a totalidade das representações visuais disponíveis na versão 3.5.2 da biblioteca gráfica javascript – D3.js, identifique e descreva algumas das que considero serem as mais pertinentes considerando os critérios identificados anteriormente.

### 3.7.1.1. Sunburst

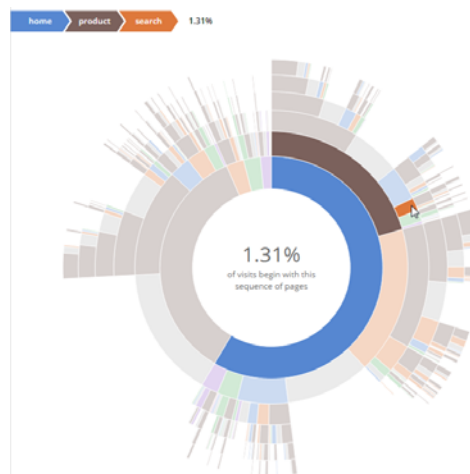
Um *sunburst* é semelhante ao *treemap*, só que usa uma disposição radial. O nó raiz da árvore está no centro, com as folhas na circunferência. A área (ou ângulo, dependendo da aplicação) de cada arco corresponde ao seu valor.

A representação visual deve ser cuidadosamente desenhada ou inventada (abstract structures), contudo não é suficiente, é também necessário providenciar ao utilizador interação através de funções de navegação e manipulação.

Gráfico 9 – Representação gráfica: *Sunburst*  
(<http://goo.gl/uguxjh>)



Gráfico 10 - Representação gráfica - Sequences suburst (fonte: <http://goo.gl/uguxjh>)



### 3.7.1.2. Chord

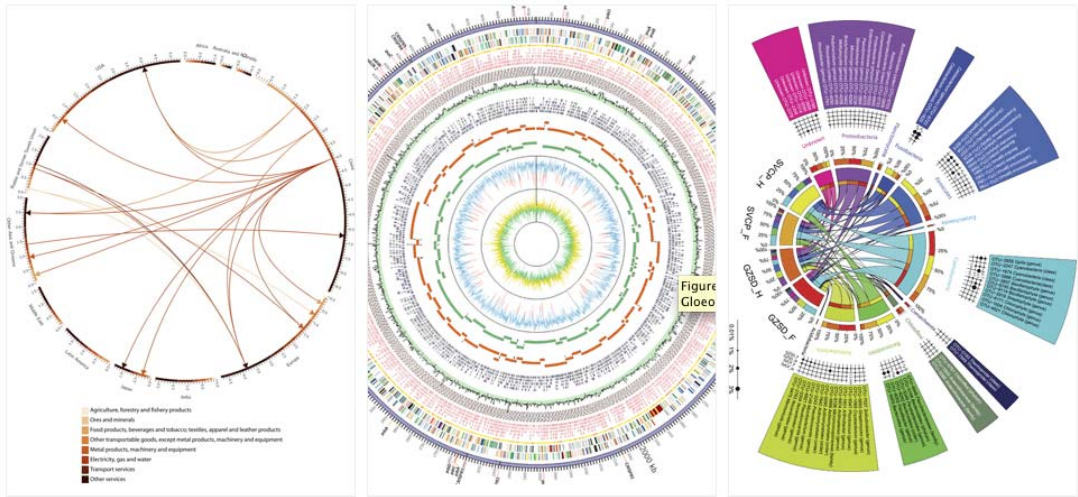
Estes diagramas representam as relações entre um grupo de entidades, e permite visualizar os dados num layout circular, ideal para explorar as relações entre objetos ou posições. Este exemplo também demonstra interatividade simples permitindo o filtro da

Visualização de Informação através de gráficos vetoriais normalizados

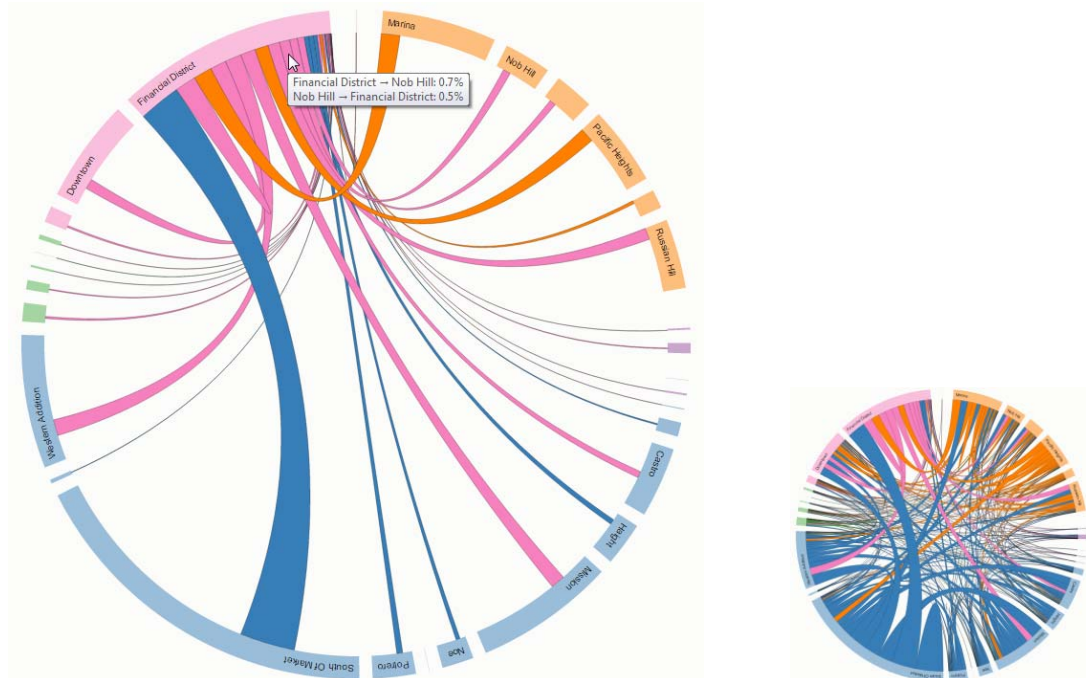
Licenciatura em Gestão de Sistemas e Computação

informação visível através de *mouseover*. Permite a visualização interativa de dados que são interdependentes.

**Gráfico 11 - Representação gráfica: Chord Diagrams - exemplos (fonte: <http://goo.gl/cnbYD>)**



**Gráfico 12 - Representação gráfica: Chord Diagram (Uber rides, fonte: <http://goo.gl/TMJER>)**

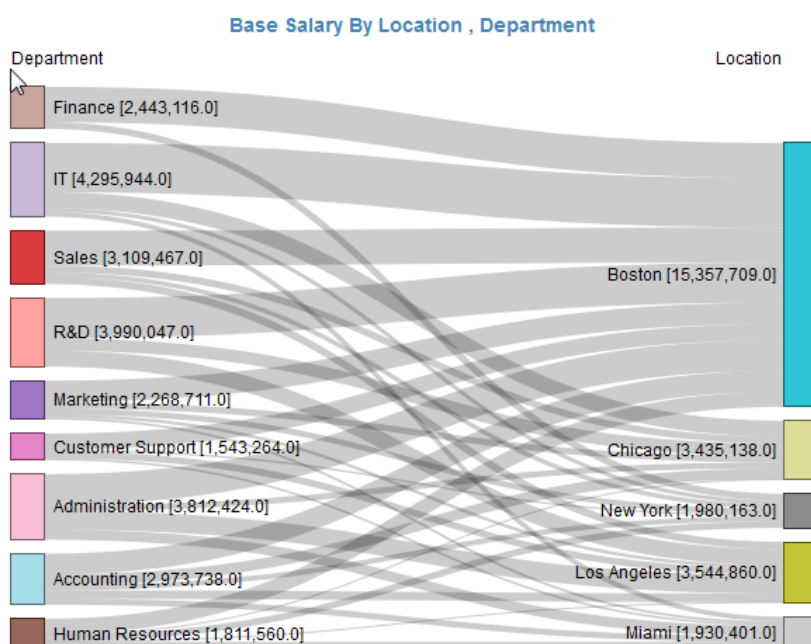


À direita, em baixo, apresenta-se a representação total de trajetos.

### 3.7.1.3. Sankey

Este é um tipo de diagrama de fluxo, em que a largura das ligações representa a proporção da quantidade de fluxo. Estes diagramas colocam ênfase visual sobre as transferências de fluxo entre a origem e o destino, estes são úteis na localização de contribuições dominantes para um fluxo global.

Gráfico 13 - Representação gráfica: Sankey Diagram

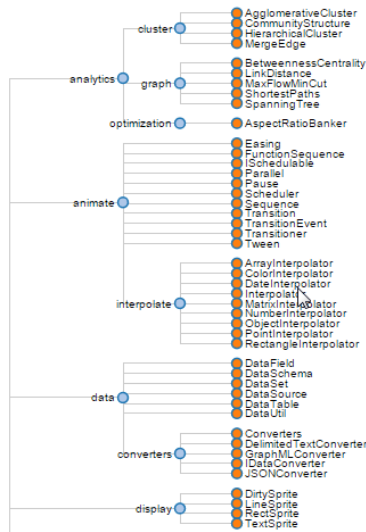


### 3.7.1.4. Dendogram

Um dendrograma (ou layout de cluster) é um diagrama *node-link* que coloca os nós na mesma profundidade. Neste exemplo que se apresenta em seguida, as classes (nós laranja) estão alinhados na margem direita, com os diretórios (azul nós internos) para a esquerda. Tal como acontece com outros esquemas de árvores, dendrogramas também pode ser orientada radialmente.



**Gráfico 14 – Representação gráfica (dendograma – protovis, fonte: <http://goo.gl/YkiwKT>)**



**Gráfico 15 - Representação gráfica - Radial dendrogram (fonte: <http://goo.gl/Z3Wnmd>)**



### 3.7.2. Biblioteca gráfica – dimple.js

Esta é uma biblioteca gráfica – *javacript* de construção de gráficos denominados por “clássicos”, ou seja: barras, tarte, linhas, área, dispersão, ou bolha.

Sítio da internet: <http://dimplejs.org>

**Gráfico 16 - Gráficos de barra - dimple.js (fonte: <http://dimplejs.org>)**

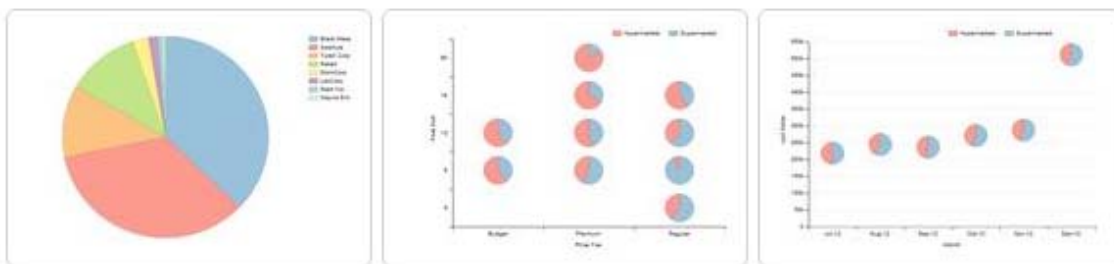


Um gráfico de barras é um gráfico que usa barras horizontal ou vertical para mostrar as comparações entre as categorias . Um eixo do gráfico mostra as categorias específicas a serem comparadas , e o outro eixo representa um valor discreto.

Os gráficos de barras podem também ser usados para comparações de dados mais complexas com gráficos de barras agrupadas e gráficos de barras empilhados. Num gráfico de barras agrupadas, para cada grupo de cada categoria existem duas ou mais barras. Estas barras são utilizadas para representar um determinado agrupamento codificados por cores. Por exemplo, uma empresa com duas lojas pode representar num gráfico de barras agrupadas diferentes barras coloridas para representar cada loja: o eixo horizontal (x) poderá representar os meses do ano e o eixo vertical (y) irá mostrar a receita.

Em alternativa, um gráfico de barras empilhado pode ser usado, em que os grupos se fazem representar uns em cima dos outros. A altura da barra resultante mostra o resultado combinado dos grupos. No entanto, os gráficos de barras empilhadas não são adequados para conjuntos de dados em que alguns grupos têm valores negativos. Em tais casos, os gráficos de barras agrupadas são preferíveis.

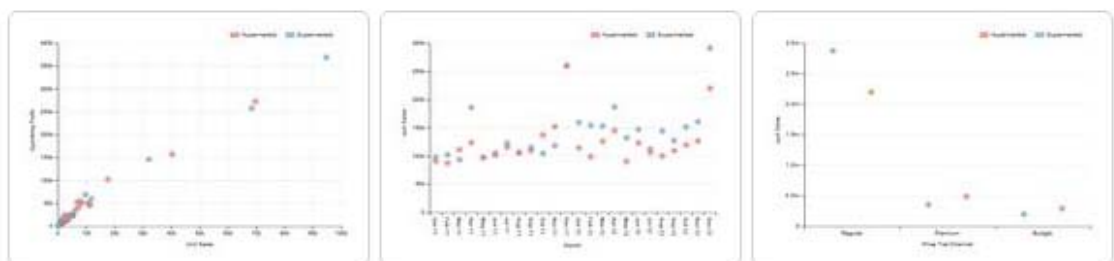
**Gráfico 17 - Gráficos de tarte - dimple.js (fonte: <http://dimplejs.org>)**



Um gráfico de tarte é um gráfico estatístico circular, que é dividido em fatias para ilustrar proporção numérica, no qual o comprimento do arco de cada fatia é proporcional à quantidade que representa.

Os gráficos de tarte são amplamente utilizados no mundo empresarial, contudo verifica-se nos últimos tempos uma elevada crítica à sua utilização que indica ser difícil comparar diferentes seções de um dado gráfico de tarte, ou ainda a comparabilidade de dados entre diferentes gráficos de tarte. Os gráficos de tarte podem ser substituídos na maioria dos casos, por gráfico de barras, ou gráficos de pontos.

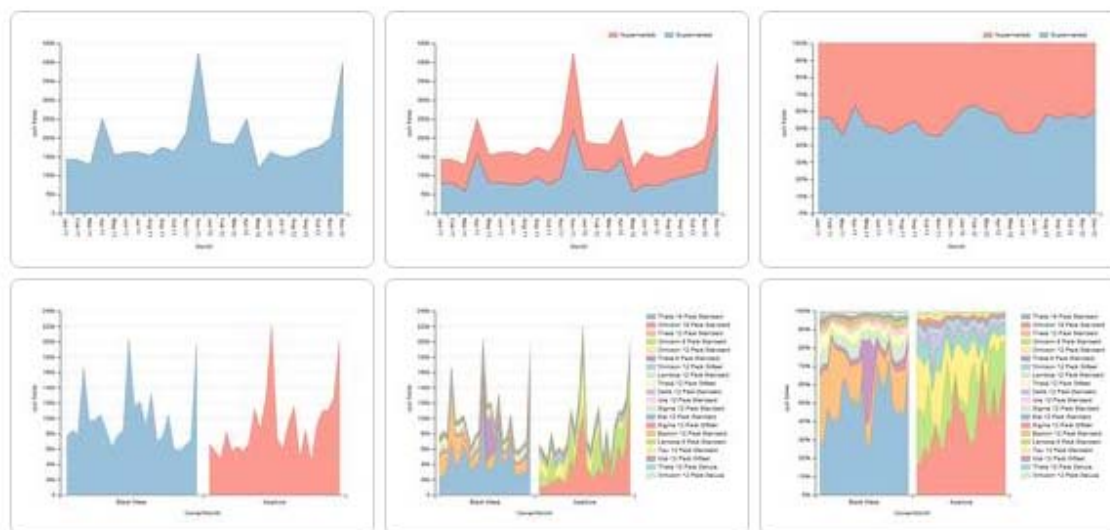
**Gráfico 18 - Gráficos de dispersão - dimple.js (fonte: <http://dimplejs.org>)**



Os gráficos de dispersão XY, como se pode entender diretamente do nome, trabalham com eixos específicos e são marcados quando os dados numéricos de ambos os eixos se cruzam. Este gráfico é específico para a representação de várias séries de dados. É importante sublinhar que o gráfico dispersão (XY) precisa de pares de valores para trabalhar. Se não houver valores em par, é impossível determinar um ponto.

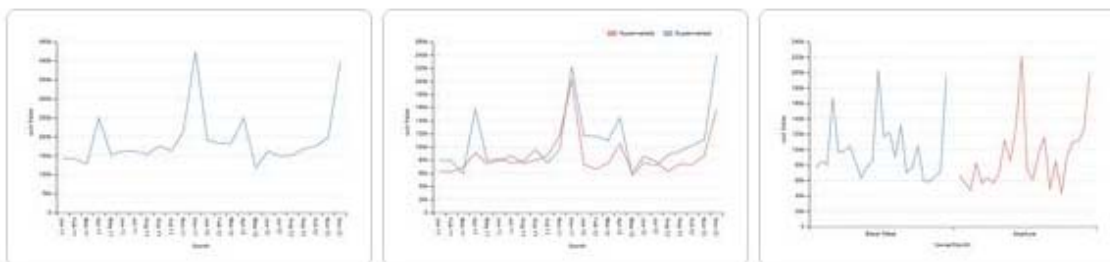
É interessante usá-lo em situações em que é necessária comparação de dados ao longo de um eixo temporal.

**Gráfico 19 - Gráficos de área - dimple.js (fonte: <http://dimplejs.org>)**



Os gráficos em área são excelentes para destacar mudanças e oscilações de uma categoria de acordo com o tempo, por exemplo. Assim como todas as outras opções de gráfico, este gráfico também conta com diversos subtipos em duas ou três dimensões, empilhados e 100% empilhados. Contudo, se apenas for utilizada uma série de dados, poderá ser mais adequado utilizar um gráfico de linhas para que todas as informações apareçam sem interferência de outras categorias.

**Gráfico 20 - Gráficos de linha - dimple.js (fonte: <http://dimplejs.org>)**



Os gráficos de linha são aqueles que aprendemos mais cedo a construir e que são visíveis em muitos locais e contextos, porque são excelentes para representar sequências de dados numa escala de tempo dividida em períodos iguais. Normalmente, no eixo horizontal temos a divisão do tempo em dias, meses ou qualquer unidade de tempo (nos casos em que se procura representar a evolução de uma determinada variável ao longo do tempo); e no eixo vertical os valores.

Os rótulos são bem evidentes neste tipo de gráfico quando a quantidade deles é inferior a dez, caso contrário, talvez seja preferível adotar gráficos de dispersão, onde os pontos poderão ser vistos sem interferências.

### 3.8. Gráficos vetoriais normalizados

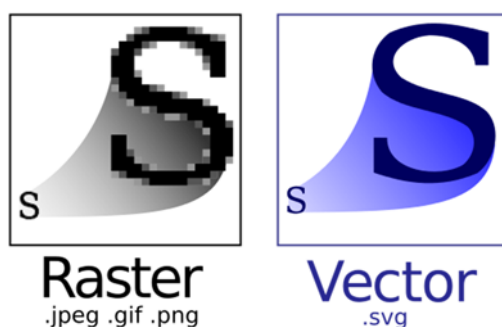
Bitmap é um “Mapa de Bits” e também é chamado de “Raster” ou “Imagem Rasterizada”. Estas imagens são produzidas através da caracterização de cada pixel que compõe a imagem. O tratamento de imagens deste tipo requer ferramentas especializadas, geralmente utilizadas em fotografia, pois envolvem cálculos muito complexos, como interpolação, álgebra matricial, etc.

*Scalable Vector Graphics* (SVG) é um formato de imagem vetorial baseado em XML para gráficos bidimensionais com suporte para interatividade e animação. A especificação SVG é um padrão aberto desenvolvido pelo *World Wide Web Consortium* (W3C) desde 1999.

As imagens SVG e os seus comportamentos são definidos em linguagem XML, o que significa que podem ser pesquisadas, geradas, e comprimidas. Sendo um ficheiro XML, imagens SVG podem ser criadas e editadas com qualquer editor de texto, mas são mais frequentemente criadas com recurso a software de desenho.

Todos os principais *browsers* web tem pelo menos algum grau de suporte SVG e pode renderizar a imagem diretamente.

**Gráfico 21 - Comparativo entre gráfico raster e vetorial**



Em resumo, uma imagem vetorial – SVG é:

1. Usada para definir gráficos vetoriais para a web
2. Define os gráficos em formato XML
3. Gráficos SVG não perde qualidade se forem ampliada ou redimensionada
4. Cada elemento e cada atributo em arquivos SVG pode ser animado
5. É uma recomendação W3C
6. Integra com outros padrões W3C , como o DOM e XSL

As vantagens do uso do formato SVG em detrimento de formatos raster (como JPEG e GIF) são:

1. Imagens SVG podem ser criadas e editadas com qualquer editor de texto
2. Imagens SVG pode ser pesquisadas, geradas por scripting, e comprimidas
3. Imagens SVG são escaláveis
4. Imagens SVG podem ser impressas com alta qualidade em qualquer resolução
5. Imagens SVG são *zoomable* (i.e. imagem pode ser ampliada sem degradação )
6. O formato SVG é um padrão aberto
7. Um ficheiro SVG é composto por 100% de linguagem XML

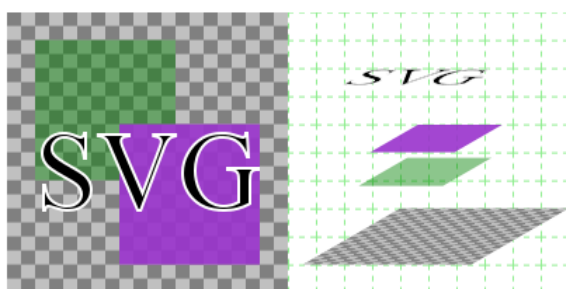
### **3.8.1. Camadas (layers)**

Uma única representação gráfica pode ser composta por várias componentes gráficas vetoriais, estas por sua vez podem ser agrupadas constituindo camadas gráficas que podem ser reordenadas, ou cuja visibilidade pode ser controlada pelo utilizador.

Através da revisão da literatura, não foi possível perceber qual o nível de adoção desta técnica – adoção de camadas gráficas para representação de dados – mas poderá

vir a ser considerada bastante útil e ajustada para a componente de interação com o utilizador, nomeadamente para a função de filtro e seleção dos dados a serem visualizados, na medida em que a cada uma destas camadas gráficas poderá ser associada uma variável dos dados em representação, aplicando assim a fase 3 – transformações do modelo de referência (Card, Mackinlay and Shneiderman, 1999).

Gráfico 22 - Camadas em SVG (fonte: [http://apike.ca/prog\\_svg\\_layering.html](http://apike.ca/prog_svg_layering.html))



### 3.8.2. Séries cronológicas

Em estatística, uma série cronológica é um conjunto de observações feitas em períodos sucessivos de tempo, durante um certo intervalo. A análise de uma série cronológica permite perceber a evolução do fenómeno em análise, e identificar tendências. A tendência é um movimento suave e consistente ao longo de um período grande de tempo (Fernandes, 1999).

No domínio geral da representação gráfica dos dados, uma série cronológica é representável através de gráficos em que um dos eixos representa o tempo. A visualização de informação, propõe que a visualização destas séries de dados se realize através de animações gráficas que representem as variações que ocorreram nas variáveis em análise, e às quais denomina de transições.

As transições, quando utilizadas, devem ser animadas suavemente, durante cerca de 1 seg. - objeto constante elimina a necessidade de reassimilação da imagem (Robertson, Jock and Card, 1991).

## **4. DOMÍNIO**

Ao iniciar a pesquisa para recolha dos dados de educação, tanto no sitio da internet do INE ou do PORDATA, verificou-se que os dados em bruto não estão disponíveis, e os dados que são oferecidos, em formato tabela ou em gráfico, correspondem já a um indicador representativo de apenas uma variável (ex. Tipo de ensino, Nível de Formação, Área de Formação e Educação, etc.), mas em nenhuma destas origens foi possível obter a relação de todas as variáveis.

Verificou-se ainda que não está disponível nenhuma relação entre o sistema de classificação de áreas de formação e educação – CNAEF, pelo qual se organizam e distribuem os cursos de ensino superior, e o sistema de classificação de setores de atividade económica – CAE. Neste sentido, no capítulo de transferência de valor do ensino superior para o mercado de trabalho, é proposta no presente trabalho uma relação destas áreas de formação com os principais setores de atividade económica: primário, secundário, e terciário.

### **4.1. Caracterização do ensino superior**

Através do sitio da internet da DGEEC foi possível obter os dados em bruto dos alunos inscritos de todas as instituições de ensino superior, desde o ano letivo 2004/05 até ao ano letivo 2014/15, e para os alunos diplomados desde o ano letivo 2007/08 até ao ano letivo 2013/14. Estes dados necessitam contudo de alguma transformação dado que as variáveis recolhidas nem sempre foram as mesmas em todos os anos letivos.

A ordenação do território nacional NUTS II e NUTS III de 2013 foi ignorada para permitir a comparabilidade entre todos os anos letivos disponíveis, tendo sido adotada a classificação de 2002.



#### 4.1.1. Alunos Inscritos

As variáveis disponíveis para a caracterização dos alunos inscritos, tendo por base a totalidade de variáveis disponíveis para os dados de 2014/15 são as seguintes:

**Tabela 5 - Variáveis de análise de alunos inscritos (fonte: DGEEC)**

Variável	Classe	Tipo	Observações
Código Estabelecimento	Qualitativo Nominal	Numérico	
Estabelecimento	Qualitativo Nominal	Alfanumérico	
Código Unidade Orgânica	Qualitativo Nominal	Numérico	
Unidade Orgânica	Qualitativo Nominal	Alfanumérico	
NUTS II - 2002	Qualitativo Nominal	Alfanumérico	
NUTS III - 2002	Qualitativo Nominal	Alfanumérico	
Distrito	Qualitativo Nominal	Alfanumérico	
Concelho	Qualitativo Nominal	Alfanumérico	
Subsistema de Ensino	Qualitativo Nominal	Alfanumérico	
Tipo de Ensino	Qualitativo Nominal	Alfanumérico	
Código Curso	Qualitativo Nominal	Numérico	
Curso	Qualitativo Nominal	Alfanumérico	
Nível de Formação	Qualitativo Ordinal	Alfanumérico	
CNAEF - Código Grande Grupo	Qualitativo Ordinal	Numérico	
CNAEF - Grande Grupo	Qualitativo Ordinal	Alfanumérico	
CNAEF - Código Área de Estudo	Qualitativo Ordinal	Numérico	
CNAEF - Área de Estudo	Qualitativo Ordinal	Alfanumérico	
CNAEF - Código Área de Educação e Formação	Qualitativo Ordinal	Numérico	
CNAEF - Área de Educação e Formação	Qualitativo Ordinal	Alfanumérico	
Sexo	Qualitativo Nominal	Alfanumérico	Disponível a partir do ano letivo 2007/08 (inclusive)
Nacionalidade	Qualitativo Nominal	Alfanumérico	Disponível a partir do ano letivo 2008/09 (inclusive)
Estudante-Trabalhador	Qualitativo Nominal	Alfanumérico	
Primeira Vez	Qualitativo Nominal	Alfanumérico	
Tempo Parcial	Qualitativo Nominal	Alfanumérico	
Bolseiro	Qualitativo Nominal	Alfanumérico	

Considerando que todas as variáveis apenas estão disponíveis a partir do ano letivo 2008/09, será analisada uma série de 7 anos letivos, de 2008/09 a 2014/15.

#### 4.1.2. Alunos Diplomados

As variáveis disponíveis para a caracterização dos alunos diplomados são as seguintes:

**Tabela 6 - Variáveis de análise de alunos diplomados (fonte: DGEEC)**

Variável	Classe	Tipo	Observações
Código Estabelecimento	Qualitativo Nominal	Numérico	
Estabelecimento	Qualitativo Nominal	Alfanumérico	
Código Unidade Orgânica	Qualitativo Nominal	Numérico	
Unidade Orgânica	Qualitativo Nominal	Alfanumérico	
NUTSII 2002	Qualitativo Nominal	Alfanumérico	
NUTSIII 2002	Qualitativo Nominal	Alfanumérico	
Distrito	Qualitativo Nominal	Alfanumérico	
Concelho	Qualitativo Nominal	Alfanumérico	
Subsistema de Ensino	Qualitativo Nominal	Alfanumérico	
Tipo de Ensino	Qualitativo Nominal	Alfanumérico	
Código Curso	Qualitativo Nominal	Numérico	
Curso	Qualitativo Nominal	Alfanumérico	
Nível de Formação	Qualitativo Ordinal	Alfanumérico	
CNAEF - Código Grande Grupo	Qualitativo Ordinal	Numérico	
CNAEF - Grande grupo	Qualitativo Ordinal	Alfanumérico	
CNAEF - Código Área de Estudo	Qualitativo Ordinal	Numérico	
CNAEF - Área de Estudo	Qualitativo Ordinal	Alfanumérico	
CNAEF - Código Área de Educação e Formação	Qualitativo Ordinal	Numérico	
CNAEF - Área de Educação e Formação	Qualitativo Ordinal	Alfanumérico	
Género	Qualitativo Nominal	Alfanumérico	
Nacionalidade	Qualitativo Nominal	Alfanumérico	

Considerando que todas as variáveis estão disponíveis em todos os anos letivos disponíveis, será analisada uma série de 7 anos letivos, de 2007/08 a 2013/14.

#### 4.2. Transferência de valor para o mercado de trabalho

A quantificação deste indicador será procurada através da relação entre áreas de formação e os principais setores de atividade económica. Os dados de alunos inscritos e

diplomados do ensino superior identificados e caracterizados no capítulo - 4.1 Caracterização do ensino superior.

Como indicador de performance das empresas foi utilizado o volume médio de negócios das empresas: total e por setor de atividade económica (PORDATA, 2016).

**Tabela 7 - Metainformação do indicador: Volume de negócios por setor de atividade**

<b>Série</b>	Setores de atividade económica
<b>Unidade de Medida</b>	Euro
<b>Tipo Valor</b>	Média
<b>Escala</b>	N.º
<b>Fórmula</b>	Volume de negócios por setor de atividade económica no ano civil / Empresas por setor de atividade económica no ano civil
<b>Fórmula matemática</b>	$VNAE(0,t) / EAE(0,t)$ $VNAE(0,t)$ = Volume de negócios por setor de atividade económica entre os momentos 0 e t; $EAE(0,t)$ = Empresas por setor de atividade económica entre os momentos 0 e t.
<b>Conceitos</b>	
<b>Atividade Económica</b>	<p>Uma atividade resulta de uma combinação de recursos, tais como equipamentos, mão de obra, técnicas de fabricação, redes de informação e produtos, que conduzem à criação de bens ou de serviços determinados.</p> <p>Qualquer atividade se caracteriza por uma entrada de produtos (bens ou serviços), um processo de produção e uma saída de produtos.</p> <p>É frequente agrupar as atividades económicas em três grandes setores: 1. Primário, incluindo agricultura, floresta, caça, pesca e extração mineral; 2. Secundário, incluindo indústria transformadora e construção; e 3. Terciário, incluindo os serviços, tais como comércio, transportes, administração pública, educação ou saúde.</p>
<b>Volume de Negócios</b>	O volume de negócios é o montante obtido por uma empresa com a venda de bens e a prestação de serviços, excluindo impostos.

#### **4.2.1. Setores de atividade económica**

A economia de um país pode ser dividida em setores (primário, secundário e terciário) de acordo com os produtos produzidos, modos de produção e recursos utilizados.

##### **Setor Primário**

O setor primário está relacionado com a exploração de recursos da natureza, como por exemplo: agricultura, exploração de minas, pesca, pecuária, extração vegetal e caça. É o setor primário que fornece a matéria-prima para a indústria de transformação. Este setor da economia é muito vulnerável, pois depende muito dos fenómenos da natureza como, por exemplo, do clima.

##### **Setor Secundário**

É o setor da economia que transforma as matérias-primas (produzidas pelo setor primário) em produtos industrializados (roupas, máquinas, automóveis, alimentos industrializados, eletrónicos, casas, etc).

##### **Setor Terciário**

É o setor económico relacionado aos serviços. Como atividades económicas deste setor económicos, podemos identificar como exemplos: comércio, educação, saúde, telecomunicações, serviços de informática, seguros, transporte, serviços de limpeza, serviços de alimentação, turismo, serviços bancários e administrativos, transportes, etc.

#### **4.2.2. Volume de negócios por setor de atividade**

Para obter o valor médio referente ao volume de negócios das empresas por setor de atividade económica, foi necessário efetuar alguma transformação aos dados originais, recolhidos do PORDATA e já descritos através da metainformação do indicador.

Os dados originais do indicador estão agrupados por setores de atividade económica e foi necessário converter estes para os principais setores económicos.

**Tabela 8 - Relação de setores de atividade económica e principais setores económicos**

<b>Setor de atividade económica</b>	<b>Setor económico</b>
Agricultura, produção animal, caça, silvicultura e pesca	Primário
Indústrias extrativas	Primário
Indústrias transformadoras	Secundário
Eletricidade, gás e água	Secundário
Construção	Secundário
Comércio por grosso e a retalho (...)	Terciário
Transporte e armazenagem	Terciário
Alojamento, restauração e similares	Terciário
Atividades financeiras e de seguros	Terciário
Atividades imobiliárias	Terciário
Educação	Terciário
Atividades de saúde humana e apoio social	Terciário

Tendo por base a associação identificada anteriormente, foi possível obter um valor médio, em Euros, de qual o volume de negócios das empresas por principal setor económico.

**Tabela 9 - Valor total e médio por empresa /principal setor de atividade económica**

<b>ANO</b>	<b>total primário</b>	<b>valor médio primário</b>	<b>total secundário</b>	<b>valor médio secundário</b>	<b>total terciário</b>	<b>valor médio terciário</b>
<b>2004</b>	775.798,90	387.899,45	8.410.292,00	2.803.430,67	2.422.151,60	346.021,66
<b>2005</b>	853.647,10	426.823,55	8.765.105,20	2.921.701,73	2.647.420,20	378.202,89
<b>2006</b>	953.797,00	476.898,50	9.639.663,80	3.213.221,27	2.970.062,40	424.294,63
<b>2007</b>	1.032.008,00	516.004,00	11.538.339,40	3.846.113,13	3.321.171,90	474.453,13
<b>2008</b>	960.734,60	480.367,30	12.788.039,50	4.262.679,83	3.673.765,70	524.823,67
<b>2009</b>	889.206,90	444.603,45	10.533.134,20	3.511.044,73	3.388.330,70	484.047,24
<b>2010</b>	967.374,30	483.687,15	12.809.352,20	4.269.784,07	3.588.296,20	512.613,74
<b>2011</b>	1.014.345,60	507.172,80	13.232.688,30	4.410.896,10	3.489.943,90	498.563,41
<b>2012</b>	997.117,40	498.558,70	12.943.280,50	4.314.426,83	3.388.306,40	484.043,77
<b>2013</b>	906.208,80	453.104,40	12.952.390,90	4.317.463,63	3.272.609,60	467.515,66

### 4.2.3. Relação de áreas de educação e formação com setores de atividade

De modo a permitir de alguma forma a correlação cursos de ensino superior e volume de negócios médio das empresas em Portugal, foi necessário relacionar os sistemas de classificação CNAEF e CAE, como já indicado anteriormente. Na tabela apresentada é proposta uma associação de setor de atividade económica com cada uma das áreas de educação e formação que classificam cursos de ensino superior.

**Tabela 10 - Tabela de relação de áreas de estudo (CNAEF) com setores de atividade económica**

Área de estudo	Sector de Atividade
01 Programas de base	Terciário
08 Alfabetização	Terciário
09 Desenvolvimento pessoal	Terciário
14 Formação de professores/formadores e ciências da educação	Terciário
21 Artes	Terciário
22 Humanidades	Terciário
31 Ciências sociais e do comportamento	Terciário
32 Informação e jornalismo	Terciário
34 Ciências empresariais	Terciário
38 Direito	Terciário
42 Ciências da Vida	Terciário
44 Ciências físicas	Terciário
46 Matemática e estatística	Terciário
48 Informática	Terciário
52 Engenharia e técnicas afins	Secundário
54 Indústrias transformadoras	Secundário
58 Arquitetura e construção (1)	Terciário
62 Agricultura, silvicultura e pescas	Primário
64 Ciências veterinárias (2)	Primário
72 Saúde (3)	Terciário
76 Serviços sociais	Terciário
81 Serviços pessoais	Terciário
84 Serviços de transporte	Terciário
85 Proteção do ambiente	Terciário
86 Serviços de segurança	Terciário
99 Desconhecido ou não especificado	Terciário

- (1) Com exceção à área de educação e formação 582 - Construção civil e engenharia civil – Setor de atividade: Secundário
- (2) Com exceção à área de educação e formação 640 - Ciências veterinárias – Setor de atividade: Terciário
- (3) Com exceção à área de educação e formação 727 - Ciências farmacêuticas – Setor de atividade: Secundário

Depois de analisados e transformados os dados em bruto dos alunos diplomados, foi possível obter o número de alunos médio por setor económico, por ano letivo.

#### 4.2.4. Análise por setor económico

Mediante uma análise diferenciada por setor económico, é possível comparar os resultados de volume de negócio, em média por empresa, com o número de diplomados, em média por curso. Os alunos diplomados de um determinado ano letivo, ex. 2007/08 concluem o curso no decorrer do ano 2008, e por isso, e para o propósito do presente trabalho, se convencionou que a transferência de valor se regista no ano seguinte ao do diploma.

Exemplo: Aluno diplomado em 2007/08, resultados no volume de negócio em 2009.

**Tabela 11 - Correspondência entre nº médio de alunos diplomados e volume de negócio, em média**

Ano letivo	Primário		Secundário		Terciário	
	Alunos	Volume Negócios	Alunos	Volume Negócios	Alunos	Volume Negócios
2007/08	37	444.603,45	75	3.511.044,73	47	484.047,24
2008/09	22	483.687,15	64	4.269.784,07	46	512.613,74
2009/10	21	507.172,80	57	4.410.896,10	45	498.563,41
2010/11	19	498.558,70	56	4.314.426,83	44	484.043,77
2011/12	17	453.104,40	55	4.317.463,63	44	467.515,66

## 5. DESENVOLVIMENTO

Para o desenvolvimento do protótipo funcional que pretende integrar bibliotecas gráficas existentes, foram utilizadas linguagens de programação direcionadas para a *world wide web* (www) maximizando assim a utilização multiplataforma, nomeadamente o computador pessoal, o *tablet* ou *smartphone*, bem como permitir a utilização de todas as funcionalidades do ambiente de construção, análise, exploração e também a consulta dos resultados das representações visuais obtidas.

Para a componente de dados e sua transformação foi utilizado um modelo de dados orientado a objetos, através de XML, que é simultaneamente a linguagem de construção de imagens vetoriais normalizadas – SVG – *Scalable vector graphics* (<http://www.w3.org/TR/SVG/>). Desta forma procura-se uma associação direta entre atributos gráficos de uma imagem e os elementos de dados da origem selecionada.

Para a representação visual das origens de dados, foram utilizados os modelos de visualização propostos pelas bibliotecas de Javascript d3.js e dimple.js, que permitem a construção de gráficos vetoriais (SVG).

A customização de elementos gráficos, no que se refere a cor, textura, linha, e outros efeitos visuais, será implementada através de CSS (cascade style-sheet).

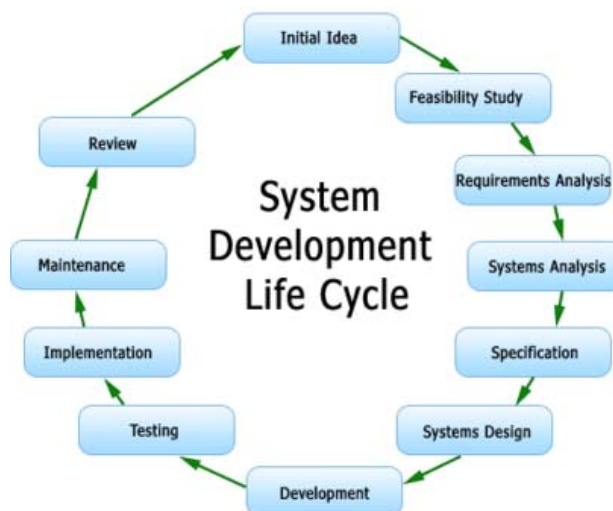
Pretende-se adotar para a construção da *framework* a linguagem Java, considerando os recursos existentes que se pretendem integrar para a solução global.

### 5.1. Ciclo de desenvolvimento

Um processo de desenvolvimento de software é um conjunto de atividades, ordenadas, com a finalidade de obter um produto de *software*. O Ciclo de Vida de Desenvolvimento de Sistemas (SDLC) é um modelo conceptual utilizado em gestão de projetos, que descreve as etapas envolvidas num projeto de desenvolvimento de sistema de informação.



Gráfico 23 - SLDC - Modelo em cascata (fonte: <https://goo.gl/NgIqHJ>)



O modelo em cascata, mais clássico, é um modelo de desenvolvimento de software sequencial no qual o desenvolvimento é visto como um fluir constante para frente (como uma cascata) através das fases de análise de requisitos, projeto, implementação, testes (validação), integração, e manutenção de software.

O modelo em cascata move-se para a próxima fase somente quando a fase anterior está completa e perfeita, não se verificando saltos entre as fases, tanto para a frente como para trás, ou sobreposição de fases.

Nos dias de hoje, tem sido cada vez mais crescente a adoção de metodologias ágeis ou magras (*lean*) de modo a permitir disponibilizar tão cedo quanto possível uma versão do produto, para que possa ser avaliado, testado, e continuamente melhorado, desta forma entende-se que o custo de cada fase e do projeto na sua totalidade poderá ser reduzido.

No contexto deste trabalho, foi adotada uma metodologia ágil, em que cada versão é determinada por um conjunto de requisitos inicialmente identificados, e que serão revistos no fim do desenvolvimento de cada versão. A lista de requisitos de cada versão é uma sublista da registo total de requisitos para o produto a desenvolver.

## 5.2. Requisitos

Lista de requisitos para o desenvolvimento do protótipo de visualização de informação através de gráficos vetoriais:

REQ 1 – Permitir construir gráficos vetoriais representativos de um conjunto de dados definido *à priori*;

REQ 2 – Permitir a visualização de “n” dimensões de dados simultaneamente, permitindo o acesso mais rápido à totalidade da informação;

REQ 3 – Permitir a interação do utilizador com as representações visuais obtidas, permitindo assim construir mais do que uma vista da informação, através de funções de apresentação/ocultação de dimensões de dados, ou de filtros sobre os dados;

REQ 4 – Disponibilizar mecanismos de animação /transição das representações visuais de modo a permitir a visualização de uma série de dados cronológica;

## 5.3. Ambiente de desenvolvimento

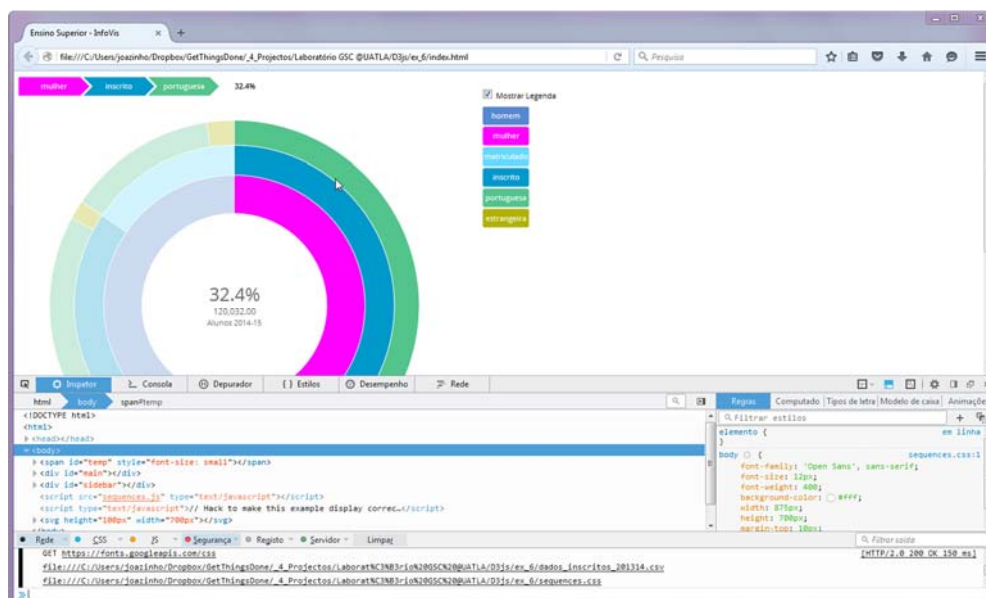
O protótipo em desenvolvimento tem como resultado final uma página *web*, acessível através de qualquer *browser* de internet (ex. Mozilla Firefox, Internet Explorer, ou Google Chrome).

As linguagens de programação utilizadas foram as seguintes: HTML para construção de páginas *web* e disposição dos conteúdos no ecrã, em conjunto com o CSS para definição de estilos e ainda para o posicionamento dos objetos. A biblioteca gráfica para construção do mapeamento visual e vistas disponíveis foi a *d3.js (javascript)*. Os dados foram recolhidos em formato – XLS (folha de cálculo) e foram transformados em CSV para associação ao mapeamento visual escolhido.

Para repositório dos ficheiros de código-fonte do projeto foi utilizado o GitHub (repository): <https://github.com/jppbatista>.

A totalidade do código-fonte produzida foi escrita sem necessidade de um interface de desenvolvimento, bastando um editor de texto e as ferramentas de *debug* disponíveis no *browser* de internet.

Gráfico 24 - Ambiente de *debug* - Mozilla Firefox



## 5.4. Protótipo

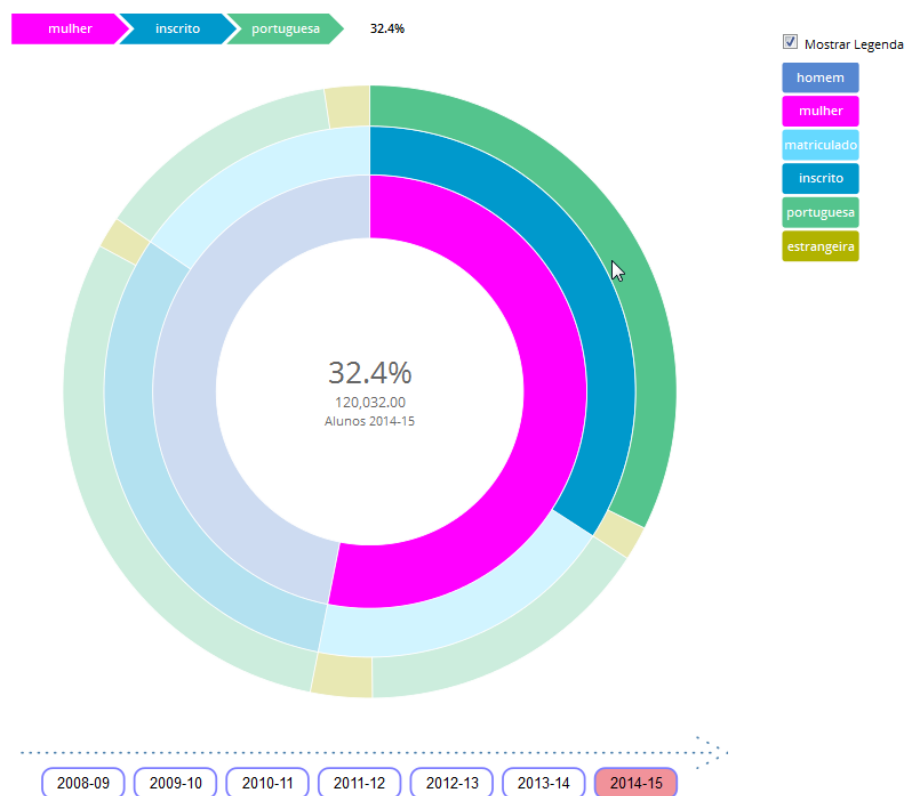
O protótipo é composto por uma representação visual denominada por *sunburst partition* que permite representar dados que sejam hierarquicamente relacionáveis, como são os casos das variáveis escolhidas: sexo, tipo de inscrição (matrícula – primeira inscrição no par estabelecimento/curso, ou inscrição), e nacionalidade (portuguesa, ou estrangeira). Cada variável é representada por um dos arcos da visualização, pela seguinte ordem, do interior para o exterior da representação visual: 1) sexo, 2) tipo de inscrição, e 3) nacionalidade.

As variáveis foram escolhidas de entre as disponíveis para a caracterização dos alunos inscritos, e foi necessário transformar os dados recolhidos originalmente, de modo a obter um número total de alunos por cada ocorrência das variáveis: sexo, tipo de inscrição, e nacionalidade.

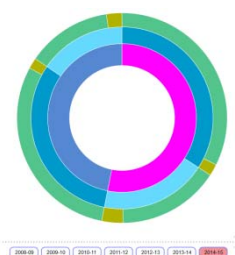
**Tabela 12 - Resultado da agregação de dados, para o ano letivo 2014-15**

Chave (sexo-tipoinscrição-nacionalidade)	Nº de alunos
homem-inscrito-portuguesa	105673
homem-inscrito-estrangeira	6088
homem-matriculado-portuguesa	46955
homem-matriculado-estrangeira	9218
mulher-inscrito-portuguesa	114757
mulher-inscrito-estrangeira	6461
mulher-matriculado-portuguesa	57046
mulher-matriculado-estrangeira	12252

**Gráfico 25 - Resultado protótipo**



À direita, é apresentada a visualização inicial da representação visual, na qual não está selecionada nenhuma sequência em consulta, e não está também visível a legenda, que é de apresentação opcional.









São partes constituintes do protótipo, as seguintes:

- Topo: sequência em consulta;
- Lateral direita: Legenda, é possível mostrar/ocultar;
- Centro: representação visual, e ao centro desta surge uma legenda com a percentagem da sequência em consulta, bem como o valor absoluto de alunos, e ainda informação do ano letivo em consulta;
- Rodapé: linha temporal, com os anos letivos disponíveis para consulta. O ano letivo em consulta encontra-se destacado através de uma cor de fundo.

Na representação desenvolvida, é utilizada a saturação da cor HSV para destacar qual a sequência escolhida e em consulta. No acesso inicial todos os arcos são apresentados sem qualquer saturação.

As cores escolhidas para cada arco /variável seguiram os critérios descritos na tabela 13.

**Tabela 13 – Critérios utilizados para escolha de cor de cada dimensão**

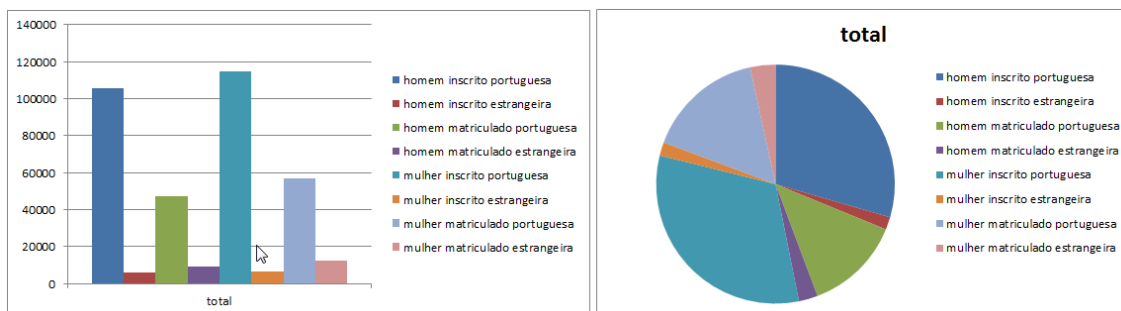
Variável	Valor	Cor	Descrição
Sexo	Homem		
	Mulher		
Tipo de Inscrição	Matrícula		O azul claro identifica o início do percurso académico, por oposição ao azul escuro que corresponde a uma inscrição subsequente.
	Inscrição		
Nacionalidade	Portuguesa		O verde está presente na bandeira nacional e facilmente se associa a Portugal. O vermelho foi excluído por ser uma cor muito forte que iria desviar a atenção para o resto.
	Estrangeira		O amarelo escolhido, como o verde, é uma cor clara, mas quente, em oposição ao verde que é uma cor fria.

Por último, é possível alternar manualmente a consulta pelos anos letivos disponíveis na linha temporal, no rodapé da imagem. Em alternativa, é possível animar as transições por entre todos os anos letivos disponíveis.

### 5.5. Métodos clássicos

Neste capítulo procura-se construir a representação gráfica das mesmas variáveis de dados, apresentadas na descrição do protótipo, de acordo com os métodos mais utilizados, que nos dias de hoje corresponde aos gráficos do tipo barras, ou tarte.

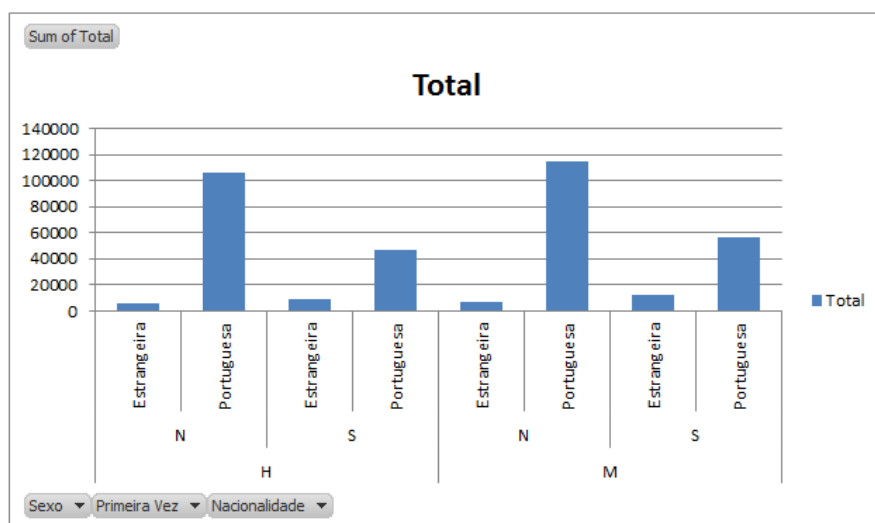
**Gráfico 26 - Gráficos de barra e tarte - Excel**



No gráfico 26, no gráfico em barra, são apresentadas tantas séries de dados (entradas na legenda) quantas as conjugações possíveis das três variáveis em estudo: sexo, tipo de inscrição, e nacionalidade, e nas categorias (eixo do X) é apresentado o valor total.

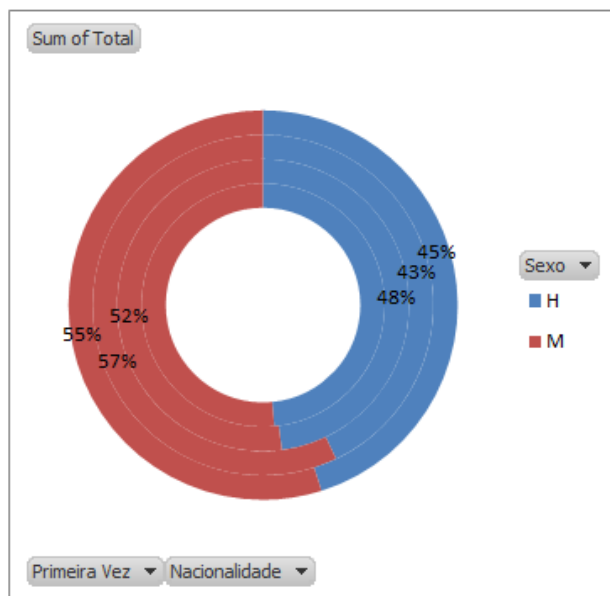
No gráfico 27, é apresentado um gráfico de barras que corresponde à variante de a série conter o valor total, e nas categorias (eixo do X) contém as conjugações possíveis das três variáveis em estudo.

**Gráfico 27 - Gráfico em barra - Excel - alternativa**



O gráfico 28 apresentado em seguida corresponde a uma aproximação à representação visual utilizada pelo protótipo e que se encontra disponível no Excel. Este contém quatro arcos que corresponde à multiplicação de valores possíveis das variáveis primeira vez (sim, não), com a nacionalidade (portuguesa, estrangeira).

**Gráfico 28 - Gráfico *donut* - Excel**



## 6. CONCLUSÕES

Este trabalho não pressupõe uma avaliação empírica do modelo de representação visual proposto, em oposição a outros modelos, como é o caso dos gráficos apresentados no capítulo 5.5 Métodos Clássicos. Uma avaliação empírica permitiria obter *feedback* sobre a funcionalidade, eficiência, eficácia e utilidade do protótipo desenvolvido. Assim, neste capítulo de conclusões procura-se avaliar o nível de satisfação dos objetivos propostos e identificar oportunidades de melhoria futuras.

### 6.1. Interpretação de resultados

Foi possível pela adoção da biblioteca gráfica d3.js implementar o modelo de referência proposto por Card. et. Al (1999), tendo sido necessário planear e executar todas as etapas desde a recolha de dados, processamento e armazenamento, associação dos dados a uma representação visual, e por último a interação do utilizador.

Da revisão da literatura realizada, é possível concluir que a biblioteca gráfica (d3.js) coloca o foco na história que os dados procuram transmitir, e procura extrair o máximo de valor possível dos dados, ao invés de apenas se focar na simples estatística dos dados. Isso é possível através de mapeamentos visuais e interação com o utilizador.

O elemento gráfico – cor – é um aspeto essencial no sucesso da mensagem que se procura transmitir, a experiência e predefinição do seu significado permite uma compreensão imediata da informação que os dados procuram transmitir. Este sucesso depende por isso dos seus utilizadores, e da sua compreensão para com a representação visual proposta, como um todo, onde se incluem naturalmente os atributos gráficos.

Na visualização de informação um aspeto que se destaca é o da capacidade de explorar os dados através das valências de interação do utilizador com os gráficos, permitindo-lhe explorar os dados e descobrir as informações que lhe estão subjacentes. Este aspeto é diferenciador ao de abordagens mais clássicas em que a descoberta de informação é contemplativa de uma imagem estática.



## 6.2. Oportunidades de melhoria futura

Estes tipos de visualização de dados e informação surgem nos dias de hoje quase sempre associados a trabalhos científicos, onde os dados que se procuram representar são complexos e onde as pessoas que produzem e consomem esta informação têm uma apetência natural para a compreensão de diferentes representações de dados através de novas representações visuais, e possuem os conhecimentos técnicos necessários para o seu desenvolvimento.

São oportunidades de melhoria as seguintes:

1. Disponibilizar uma *framework* de visualização de informação, que permita ao utilizador, sem conhecimentos técnicos, construir uma representação visual, de acordo com o modelo de referência proposto por Card. et. al. (1999), sem necessidade de programação, para isso teria de ser necessário ao utilizador:
  - a. Selecionar os dados;
  - b. Selecionar a representação visual, entre uma das disponíveis das bibliotecas *d3.js* e *dimple.js*;
  - c. Mapear os dados, e as suas variáveis, com atributos gráficos;
2. Avaliar a função de *drill-down* gráfico, ou seja, permitir consultar o detalhe de um gráfico através da consulta de outro(s) gráfico(s);
3. Através da *framework* proposta, permitir gravar sequências de visualização gráfica com conjugação de caixas de texto (*hints*) de modo a acompanhar o utilizador na leitura pretendida que este faça dos gráficos apresentados;
4. Explorar outros modelos visuais como são o caso: *Chord*, e *Sankey*, que permitem identificar fluxos, ou contributos, entre variáveis dos dados, como por exemplo identificar a transferência de valor do ensino superior para o mercado de trabalho, de acordo com a proposta metodológica descrita no presente trabalho;

## **Bibliografia**

- Andrews, K. (2012) *Information Visualisation - Course Notes*, 09 May, [Online].
- Barañano, A.M. (2008) *Métodos e Técnicas de Investigação em Gestão*, Edições Sílabo.
- Bederson, and Shneiderman, B. (2003) *The Craft of Information Visualization: Readings and Reflections*, Morgan Kaufmann.
- Benin, J. (1981) *Graphics and Graphic Information Processing*, New York: De Gruyter.
- Card, S. (2007) *The Human-Computer Interaction Handbook: Fundamentals, Evolving Technologies, and Emerging Applications*, A. Sears and J.A. Jacko edition, Lawrence Erlbaum Assoc Inc.
- Card, Mackinlay and Shneiderman (1999) *Readings in Information Visualization: Using Vision to Think*, Morgan Kaufmann Publishers.
- Carvalho, E. and Fernandes, A. (2009) *Visualização de informação*, Guimarães: Centro de Computação Gráfica.
- Chen, C. (2005) 'Top 10 Unsolved Information Visualization Problems', *IEEE Computer Graphics and Applications*, pp. 25(4):12-16.
- Chen, C. (2006) *Information Visualization: Beyond the Horizon*, Springer.
- Cuccu, F. and Moltedo, L. (1993) 'Texture Mapping for Scientific Visualization Environments', *IEEE - CG&A*, vol. 17, pp. 131-136.
- D'Orazio, F. (2012) *10 reasons why we visualize data*, 26 Apr, [Online], Available: <http://www.slideshare.net/Facegroup/10-reasons-why-we-visualise-data> [07 Nov 2015].
- Fernandes, E.M.d.G.P. (1999) *Estatística*, Braga: Universidade do Minho.

Freitas, Chubachi, Luzzardi and Cava (2001) 'Introdução à Visualização de Informações', *RITA – Revista de Informática Teórica e Aplicada, Instituto de Informática UFRGS, Porto Alegre, RS*, vol. VIII, n.2, pp. 143-158.

Freitas and Wagner (1995) 'Ferramentas de suporte às tarefas da análise exploratória visual', *Revista de Informática Teórica e Aplicada*, vol. 2(1), pp. 5-36.

Furnas, G.W. (1981) *The Fisheye View: A New Look at Structured Files*, Technical Memorandum, Available: <http://www.si.umich.edu/>.

Healey, C.G. (1997) "*On the Use of Percetual Cues and Data Mining for Effective Visualization of Scientific Datasets*", EECS Department, University of California at Berkeley.

Heffner (2003) *Psychology Dictionary*, [Online], Available: <http://allpsych.com/dictionary/> [maio 2012].

Heilmeyer and Tesler (1992) *The 3D File.Sjvstem Navigator*, Mountain XIieW, Silicon Graphics.

Jaisingh, L. (2000) *Statistics for the Utterly Confused*, The McGraw-Hill.

Keller, P. and Keller, M. (1993) *Visual Cues*, IEEE Society Press.

Kunni, T.L. and Shinagawa, Y. (1991) *Visualization Modeling: Making Visualization a Creative Discipline*, New Trends in Animation and Visualization, Eds.

Marcus, A. and Dam, A.V. (1991) 'User Interface Developments for the Nineties', *IEEE CG&A*, setembro, pp. 49-56.

Mazza, R. (2009) *Introduction to Information Visualization*, Springer-Verlag London Limited.

Nascimento (2005) 'Visualização de Informações – uma abordagem prática', *XXV Congresso da Sociedade Brasileira de Computação*, vol. A Universidade da Computação: um agente de inovação e conhecimento, pp. 1262 – 1312.

PORDATA (2016) *Volume médio de negócios das empresas: total e por setor de atividade económica - Portugal*, [Online], Available: <http://www.pordata.pt/Portugal/Volume+m%C3%A9dio+de+neg%C3%B3cios+das+empresas+total+e+por+setor+de+atividade+econ%C3%B3mica-2921> [11 Jan 2016].

Robertson, G.G., Jock, D.M. and Card, S.K. (1991) 'Cone Trees: Animated 3D Visualizations', *Proc. SIGCHI Conference on Human Factors in Computing*, pp. 189-194.

Senay, H. and Hignatius, E. (1996) *Rules and Principles of Scientific Data Visualization*, The George Washington University.

Shneiderman, B. (1996) 'The eyes have it: a task by data type taxonomy for information visualizations', *Proceedings of IEEE Symposium on Visual Languages*, pp. 336-343.

Tufte, E.R. (1983) *The Visual Display of Quantitative Information*, Graphics Press, Cheshire.

Tufte, E. (1994) *Principles of Information Display for Visualization Practitioners*, NASA Ames Research Center.

Tullis, T.S. (1990) *Screen Design*, Handbook of Human Computer Interaction.

Ware, C. (2000) *Information Visualization: Perception for design*, San Francisco, CA: Morgan Kaufmann.

Ware, C. (2004) *Information Visualization: Perception for Design - Interactive Technologies*, Ed. Morgan Kaufman.