



Universidade do Minho
Escola de Engenharia

Flávia Patricia Alves Pereira

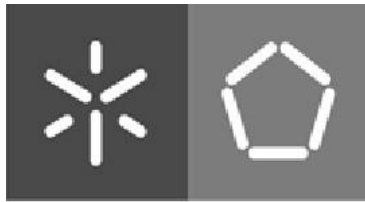
**Big Data e Data Analysis:
Visualização de Informação**

Dissertação de mestrado

Mestrado Integrado em Engenharia e Gestão de
Sistemas de Informação

Trabalho efetuado sob a orientação do
Professor Doutor Jorge Oliveira e Sá

Outubro 2015



Universidade do Minho
Escola de Engenharia

Flávia Patricia Alves Pereira

**Big Data e Data Analysis:
Visualização de Informação**

Dissertação de mestrado

Mestrado Integrado em Engenharia e Gestão de
Sistemas de Informação

Trabalho efetuado sob a orientação do
Professor Doutor Jorge Oliveira e Sá

Outubro 2015

DECLARAÇÃO

Nome: Flávia Patrícia Alves Pereira

Endereço eletrónico: a61562@alunos.uminho.pt

Telefone: 916372620

Cartão do Cidadão: 14181927

Título da dissertação: Big Data e Data Analysis: Visualização de Informação

Orientador: Professor Doutor Jorge Oliveira e Sá

Ano de conclusão: 2015

Designação do Mestrado: Mestrado Integrado em Engenharia e Gestão de Sistemas de Informação.

É AUTORIZADA A REPRODUÇÃO INTEGRAL DESTA DISSERTAÇÃO APENAS PARA EFEITOS DE INVESTIGAÇÃO, MEDIANTE DECLARAÇÃO ESCRITA DO INTERESSADO, QUE A TAL SE COMPROMETE;

Universidade do Minho, __/__/____

Assinatura:

AGRADECIMENTOS

Primeiramente, gostaria de agradecer a todos os colegas e professores que fizeram parte da minha vida enquanto estudante da Universidade do Minho. Porque este momento é o culminar de cinco anos vividos com uma enorme intensidade. Daqui levo na bagagem todas as recordações de pessoas com quem passei muitas horas de trabalho, com quem ultrapassei obstáculos e que sempre se reconheceram como um ombro amigo tanto na vida pessoal como no sucesso académico.

Agradeço aos meus pais e irmão que me propuseram estes anos de estudo pois sem o apoio deles o caminho teria sido mais duro. A educação, a preocupação, o apoio e o carinho por eles demonstrado fizeram com que as dificuldades no meu percurso académico se tornassem desafios. Há minha família agradeço o interesse e disponibilidade que sempre demonstraram em ajudar-me.

Ao Jorginho agradeço toda a fonte de energia e motivação para terminar esta etapa.

Quanto à realização desta dissertação não seria possível sem a orientação do Professor Jorge Oliveira e Sá. Agradeço a oportunidade de trabalhar com ele e explorar um tema tão mediático. A sua disponibilidade e as suas apreciações foram essenciais na evolução do trabalho. O seu profissionalismo e experiência contribuíram positivamente para a minha formação e evolução não só como estudante mas também como pessoa, desenvolvi as minhas capacidades e os meus conhecimentos na área de Engenharia e Sistemas de Informação.

Agradeço a todos os meus amigos que contribuíram para o meu sucesso académico, em especial ao José Fernandes e à Isabel Fidalgo que assistiram à escrita desta dissertação e me auxiliaram na revisão de textos, a eles boa sorte!

A todos muito obrigado!

RESUMO

A revolução da informação está abranger todas as organizações da sociedade moderna, forçando os especialistas na área de Tecnologias de Informação (TI) a transformar os seus processos de aprendizagem para a criação de valor.

As tecnologias produzem e armazenam uma grande quantidade de dados para posteriormente serem produzidas informações. Entender o conjunto heterogéneo de dados e passar a reconhecer dados com significado é o grande objetivo do conceito *Big Data*. A necessidade de compreender e extrair conhecimento a partir do grande conjunto de dados é um processo difícil mas essencial para as organizações que lidam com informação.

Neste contexto é necessário aplicar um processo de análise, limpeza e transformação de dados, denominado de *Data Analysis*. Este processo conduz o utilizador à escolha da técnica mais adequada perante objetivo da sua análise.

A técnica estudada nesta dissertação será a Visualização de Informação (VI). A VI nesta dissertação é estudada com o principal objetivo de se transmitir informação de uma forma clara e efetiva através da utilização de representações gráficas. A mapificação de dados em estruturas visuais (representações gráficas) possibilitam uma vista detalhada sobre o contexto de dados e das suas relações.

Os métodos e técnicas de Visualização de Informação evoluíram nas últimas décadas, em consonância com o avanço tecnológico galopante, daqui surge a necessidade de reformular o Modelo do Processo de Visualização, para facilitar a criação de uma representação visual.

O objetivo principal centra-se na otimização dos métodos de Visualização – produzir uma representação clara e eficiente tem como principal finalidade potencializar a apropriação de dados por meio de representações gráficas. Para este objetivo foi formulado uma Classificação: “Representação Visual: o que pretendo transmitir”, que contempla o estudo de gráficos e das análises que surgem quando se pretende descobrir ou comunicar padrões e tendências nos dados.

A Classificação foi construída como artefacto, com o propósito específico de ajudar o utilizador a decidir qual o gráfico mais adequado para evidenciar um tipo de análise.

Para o estudo desta dissertação optou-se por aplicar como abordagem metodológica o Design Science Research, para a classificação sistemática de conceitos e construção da classificação.

O utilizador como agente crucial durante o Método de Processo de Visualização: deve ser capaz de perceber qual é a análise mais apropriada para os seus dados e qual o tipo de gráfico mais rentável para o seu trabalho.

PALAVRAS-CHAVE: VISUALIZAÇÃO DE INFORMAÇÃO, ANÁLISE DE DADOS, PROPRIEDADES VISUAIS, GRÁFICOS.

ABSTRACT

The revolution of information reaches all organizations of modern society, forcing experts in the field of Information Technology (IT) to transform their learning processes to create more value.

Technology produces and stores a large quantity of data to be able to produce information afterwards. To understand the heterogeneous data and recognize data that matters is the ultimate goal of the concept of Big Data. The need to understand and extract information from a large group of data is a hard but essential process for organizations that deal with information.

In this context comes the need to analyze, clean and transform data. This process is called Data Analysis. The process guides the user to use the most suitable technique depending on the purpose of his analysis.

The technique studied in this thesis will be Information Visualization (IV). The IV in this thesis is studied with the main purpose of transmitting information in a clear and effective way through the use of graphic representations. The mapification of data into visual structures (graphic representations) provide a detailed view of the data context and their relations.

The methods and techniques of IV have evolved in the last decades, in line with the rampant technological progress, hence there is the need to redesign the Model of Visualization Process, to further the making of a visual representation.

The main goal focuses on the optimization of Visualization methods - the main purpose of producing a clear and efficient representation is to enhance the appropriation of data via graphic representations. To attain this purpose was formulated a classification: "Visual Representation: what I intend to transmit", which contemplates the study of graphics and the analysis that arises when one wants to find or communicate patterns and trends in data. The classification was built as an artefact, with the specific purpose of helping the user to decide which graphic is the most suitable for a certain type of analysis. For the study of this thesis it was chosen Design Science Research to apply as methodological approach, for the systematic classification of concepts and the construction of classification. The user, as a key agent during the Visualization Process Method: should be able to acknowledge which is the most appropriate analysis for his data and what type of graphic is the most profitable for his work.

KEYWORDS: INFORMATION VISUALIZATION, DATA ANALYSIS, VISUAL STRUCTURES, GRAPHICS

ÍNDICE

Agradecimentos	iii
Resumo.....	v
Abstract	vii
1. Introdução	15
1.1 Motivação.....	16
1.2 Objetivos	17
1.3 Estrutura do Documento.....	18
1.4 Abordagem metodológica.....	19
1.4.1 Revisão de Literatura.....	19
1.4.2 Design Science Research	20
2. Contexto de Estudo	22
2.1 Big Data	22
2.2 Data Analysis	24
3. Visualização de Informação.....	25
3.1 O Conceito	25
3.2 Perspetiva histórica.....	26
3.3 Dados.....	36
4. Representações visuais	41
4.1 Técnicas de Visualização	41
4.2 Mapeamento Visual	43
4.3 Métodos Analíticos.....	45
5. Gráficos de Informação	46
5.1 Gráficos	46
5.1.1 Gráfico de Linhas	46
5.1.2 Gráfico de Área.....	47
5.1.3 Gráfico de Barras	47
5.1.4 Caixa de Bigodes.....	48
5.1.5 Gráfico de Bolhas.....	48
5.1.7 Histograma	49
5.1.8 Gráfico de Mosaicos	49
5.1.9 Gráfico de Colunas Agrupado.....	49

5.1.10	Pirâmide de População.....	50
5.1.11	Gráfico de Radar.....	50
5.1.12	Gráfico de Barras Radial	51
5.1.13	Gráfico de Dispersão	52
5.1.14	Gráfico de Área Agrupado	52
5.1.15	Gráfico de Barras Agrupado	53
5.1.16	Gráfico de Palmas.....	53
5.2	Diagramas.....	54
5.2.1	Diagrama de Arcos	54
5.2.2	Mapas Mentais.....	54
5.2.3	Diagrama de Cordas	55
5.2.4	Diagrama de Ilustração.....	55
5.2.5	Diagrama de Árvore.....	56
5.2.6	Diagrama de Venn	56
5.2.7	Gráfico da Linha do Tempo	57
5.3	Tabelas	58
5.3.1	Gráfico de Contagem.....	58
5.3.2	Tabela de Tempo	58
5.4	Charts	59
5.4.1	Círculo de Pacotes	59
5.4.2	Gráfico de Donuts.....	59
5.4.3	Matriz de Pontos	60
5.4.4	Gráfico em Caracol	60
5.4.5	Gráfico de Conjuntos Paralelos	61
5.4.6	Gráfico de Imagens.....	61
5.4.7	Gráfico de Tarte.....	62
5.4.8	Gráfico de Área Proporcional	62
5.4.9	Mapas de Árvores.....	62
5.4.10	Nuvem de palavras.....	63
5.5	Mapas.....	64
5.5.1	Mapa de Bolhas.....	64
5.5.2	Mapa de Cores.....	64

5.5.3	Mapa de Fluxo	65
6.	Classificação	66
7.	Conclusão	70
8.	Referências Bibliográficas.....	73

ÍNDICE DE FIGURAS

Figura 1- William Playfair – 1821 (Friendly, 2006)	26
Figura 2 - Dr. John Snow, The Cholera Epidemic in London: 1854.....	27
Figura 3 - Charles Minard “Showing the transportation of comercial goods along the Canal du Center”	28
Figura 4 - Charles Joseph Minard, “Mapa da Campanha Russa de Napoleão”	28
Figura 5 - Tipo de manipulação por Darrel Huff.....	29
Figura 6 - Exemplo de Permutation Matrices Jacques Bertin 1967 adaptado (Bertin Matrices,2014). 30	
Figura 7- Modelo de processamento de Informação, adaptado de Card99	32
Figura 8 - Principios de Gestalt.....	34
Figura 9 – Classificação: "Representação Visual"	68

ÍNDICE DE GRÁFICOS

Gráfico 1- Escala quantitativa vs escala categórica.....	37
Gráfico 2 – Dúas escalas quantitativas.....	38
Gráfico 3 – Áreas de Estudo: Escala Categórica Nominal.....	38
Gráfico 4 - Ranking: Escala Categórica Ordinal	38
Gráfico 5 - Exemplo: Escala Categórica Intervalar.....	39
Gráfico 6 - Line Graph.....	46
Gráfico 7 - Area Graph.....	47
Gráfico 8 - Bar Graph.....	47
Gráfico 9 - Boxplot.....	48
Gráfico 10 - Bubble Graph.....	48
Gráfico 11 - Histogram	49
Gráfico 12 - Marimekko Graph.....	49
Gráfico 13 - Multi Set Bar Graph	50
Gráfico 14 - Population Pyramid	50
Gráfico 15 - Radar Graph.....	51
Gráfico 16 - Radial Bar Graph.....	51
Gráfico 17 - Scatter Plot	52
Gráfico 18 - Stacked Area Graph	52
Gráfico 19 - Stacked Bar Graph	53
Gráfico 20 - Span Chart	53
Gráfico 21 - Arc Diagram	54
Gráfico 22 - BrainStorm.....	54
Gráfico 23 - Chord Diagram.....	55
Gráfico 24 - Illustration Diagram	55
Gráfico 25 - Tree Diagram	56
Gráfico 26 - Venn Diagram	56
Gráfico 27 - Timeline	57
Gráfico 28 - Tally Chart.....	58
Gráfico 29 – Time Table.....	58
Gráfico 30 - Circle Packing.....	59
Gráfico 31 - Donut Chart	59
Gráfico 32 - Dot Matrix Chart.....	60

Gráfico 33 - Nightingale Rose Chart	60
Gráfico 34 - Parallel Sets.....	61
Gráfico 35 - Pictogram Chart	61
Gráfico 36 - Pie Chart	62
Gráfico 37 - Proportional Area Chart	62
Gráfico 38 - Tree Map.....	63
Gráfico 39 - Word Cloud.....	63
Gráfico 40 - Bubble Map	64
Gráfico 41 - Choropleth Map.....	65
Gráfico 42 - Flow Map	65

1. INTRODUÇÃO

A Tecnologia de Informação (TI) é definida como o conjunto de todas as atividades e soluções fornecidas por recursos computacionais que permitem obter o armazenamento, o acesso, a gestão e o uso das informações. As aplicações para a TI são imensas e estão ligadas às mais diversas áreas. A informação é vista como um património que dá valor e sentido às atividades que utilizam ferramentas e sistemas que façam da informação um diferencial. A informação é tão importante que pode determinar a sobrevivência ou a descontinuidade de um negócio por isso é necessário desmembrar o que a informação proveniente das mais diversas fontes pode oferecer a quem a explora.

O avanço tecnológico das comunicações e o crescente desenvolvimento da Internet faz com que os utilizadores e organizações sejam inundadas por uma quantidade enorme de informações, surgindo uma grande dificuldade em perceber o que será útil ou o que não (Bellinger, Castro, & Mills, 2004).

Neste momento encontramos-nos na Era dos Dados que foi intitulada de *Big Data*. Este efeito provocou nas pessoas a necessidade de poder perceber e transmitir estes dados de forma clara e precisa. Na realidade, dados são factos do Mundo, estão sempre corretos e podem ser entendidos como uma descrição deste. Sem perceber, utilizamos dados desde os tempos mais primórdios do Homem para construir o conhecimento do Mundo. A informação constrói-se a partir do processamento, organização e representação de dados num certo contexto fazendo com que o acontecimento se torne útil e compreendido. Depois do trabalho com dados e informação surge o conhecimento, adquirido pela combinação de informações, experiências e *insights* que transferem benefícios aos utilizados e às organizações. (Shneiderman, 1996)

Uma forma de lidar com a interpretação do grande volume dados é o recurso a mecanismos visuais, onde as pessoas conseguem explorar, relacionar, interpretar e apresentar informação útil com rapidez e precisão. Neste sentido surge o tema desta dissertação – Visualização de Informação (VI). As técnicas utilizadas antes desta Era, essencialmente técnicas *data mining* podem não corresponder à rapidez e velocidade de processamento com que a informação aparece. Estas técnicas são realmente úteis na análise de grandes conjuntos de dados para procurar informações relevantes e pertinentes. O que acontece é que as organizações recolhem enormes conjuntos de dados heterogéneos e precisam de acesso rápido e eficiente a esses conjuntos, tomando as decisões mais corretas no menor espaço de tempo devido à competitividade do mercado.

Entende-se por VI o processo de transformar dados em imagens ou representações gráficas com a finalidade de serem interpretadas e/ou apresentadas. As técnicas de visualização são extremamente importantes para a criação de valor que emerge do conceito *Big Data*.

O estudo da VI é pertinente perante a situação atual dentro das TI porque a apresentação de dados em formato gráfico ou em imagens segue o princípio que o cérebro humano tem mais capacidade de interpretar informação exibida na forma visual.

A utilização de ferramentas gráficas durante todo o processo é indispensável. O utilizador tem à sua disposição vários tipos de gráficos que permitem transformar os dados em grandes visualizações, mas no entanto, existem muitos erros quando se pretende escolher a representação gráfica mais adequada consoante a análise e o conjunto de dados que está em causa.

O tipo de dados a explorar, o tipo de análise, a representação gráfica e o resultado obtido identificam várias áreas de estudo como a Visualização de Informação, Visualização Científica, Visualização de Dados e Visualização de Informação Geográfica. A última área destaca-se como sendo a mais antiga, surgindo a partir de mapas, com a característica dos dados representarem uma localização. Nas definições citadas neste trabalho a Visualização de Informação apresenta-se como uma área com domínio da Visualização, com destaque ao tratamento de informação abstrata não tratada na Visualização Científica. Deste modo, não se faz distinção entre as áreas de estudo.

Não existe uma descrição precisa da informação a visualizar como em outras áreas, em que são identificados os tipos de dados que se pretende visualizar. Contudo as afinidades que ligam estas áreas são iguais no processo de conceber representações visuais.

A VI apresenta-se como uma disciplina que ainda não ganhou maturidade suficiente para haver uma caracterização sistemática dos processos que envolvem o seu estudo pois existe e irá sempre existir uma grande variedade de dados abstratos, múltiplas análises, várias representações e utilizadores. Isto implica que a área continue a ser estudada e consequentemente criadas subáreas com focos de estudo específicos.

1.1 Motivação

Um Sistema de Informação é um conjunto de elementos relacionados e desenvolvidos para recolher, processar, armazenar e distribuir informação para facilitar a coordenação, o controle, a análise, a visualização e o processo de decisão de uma organização.

No contexto das TI, surge o termo *Big Data* definido como sendo um conjunto de dados extremamente diverso e por este motivo, surge a necessidade de desenvolver e aprimorar técnicas e ferramentas capazes de lidar com o armazenamento, processamento e apresentação dessa informação. A humanidade lida com dados desde os seus primórdios mas os avanços computacionais despoletaram tanto a explosão de informação, como o processo de a tornar perceptível num contexto. Desta forma surgem desafios em analisar informação de forma rápida e eficaz, de maneira a trazer vantagem para quem a analisa.

Num mundo onde voga o conceito de *Big Data* há imensas perguntas por responder e padrões por descobrir, tornando-se uma chave de pesquisa saber como responder a essas perguntas da forma mais correta.

As ferramentas gráficas evoluíram para analisar dados, por conseguinte, a deteção de anomalias nos dados, relacionamentos, padrões ou tendências são altamente evidenciadas através de representações gráficas. Uma vez que existe vantagem na exploração visual dos dados surge uma nova disciplina, o *data visualization* que é entendido um campo recente da estatística descritiva.

Uma representação eficaz sobre os dados ajuda o utilizador a direccionar a sua análise assim como a perceber as evidências escondidas na heterogeneidade de dados. Não só na exploração e análise de dados, as representações gráficas são úteis na apresentação de um negócio, conceito ou ideia.

Um dos problemas mais importantes do estudo da Visualização de Informação é a escolha de uma representação gráfica adequada para o tipo de dados em tratamento. Associar dados a elementos gráficos e a uma localização no espaço pode ser uma tarefa difícil que requer conhecimentos em tipos de dados, análise de dados e representações gráficas. Todas as temáticas serão estudadas ao longo da construção desta Dissertação.

Esta Dissertação surge com a motivação de tornar o Modelo do Processo de Visualização mais fácil, eficaz e adaptado ao cenário de grandes dados. Pretende-se que a dificuldade em escolher uma representação gráfica adequada consoante os relacionamentos que os dados contêm seja desvanecida. Por este motivo será construída uma Classificação que se pretende que seja validada e posta em prática por qualquer utilizador que pretenda construir uma representação visual dos seus dados.

Uma representação visual não só ajuda o utilizador a interpretar e entender um fenómeno, como pode obscurecer o seu significado, impedindo que o utilizador retire informação. Para que a Visualização se torne mais do que uma arte sofisticada, é preciso estudar como os utilizadores abordam e resolvem os seus problemas, percebendo quais são as representações visuais úteis para adicioná-las às técnicas e ferramentas de visualização.

1.2 Objetivos

O principal objetivo desta dissertação é construir uma classificação que contenha representações gráficas. Com este objetivo foram explorados e descritos vários tipos de gráficos, ligando a cada um uma finalidade específica ou os dados que representam.

De modo a melhorar as etapas a percorrer desde a chegada de dados até à sua representação visual, surge o propósito dos temas de estudo: dados, métodos analíticos e representações gráficas.

A pretensão desta dissertação é que um utilizador estude e consulte a classificação desenvolvida sempre que pretender criar uma representação visual, de maneira a eliminar qualquer erro na escolha da representação, na altura de explorar ou apresentar os seus dados.

Além do objetivo de pesquisa principal deste trabalho de investigação, destacam-se outros objetivos:

- Realizar uma análise de literatura relevante sobre **big data, data analysis, visualização de informação, modelos de visualização, técnicas de visualização, dados, propriedades visuais, representações gráficas, gráficos e visualização de dados;**
- Clarificar as temáticas inerentes à Visualização de Informação (VI): *Big Data* e *Data Analysis*, os conceitos que conduzem a aprimorar o estudo da VI;
- Identificar, caracterizar e classificar dados;
- Identificar os elementos relevantes para o mapeamento visual;
- Identificar propriedades visuais e gráficas;
- Descrever e perceber técnicas que explicita a forma como se pode modelar dados em representações gráficas;
- Identificar e descrever um conjunto de representações gráficas;
- Construir uma classificação de representações gráficas ou esquemas visuais;
- Esquematizar a classificação.

1.3 Estrutura do Documento

O presente documento está organizado de acordo com os objetivos propostos na secção anterior. Uma divisão em sete secções organizados de forma a proporcionar ao leitor uma vista concisa e detalhada sobre o problema de investigação.

A presente secção confere ao documento uma visão geral sobre o mesmo. Descrição do motivo implícito no estudo da visualização de informação no contexto das TI e os objetivos principais do trabalho de investigação.

A secção 2 expõe que a dissertação é um documento de investigação científica exploratória com recurso à revisão de literatura e ajuste da abordagem metodológica – *Design Science Research*, para a produção de um artefacto.

A secção 3 surge como uma abordagem inicial para o estudo da Visualização de Informação. O conceito *Big Data* surgiu causando um incomensurável impacto nas organizações, obrigando-as a repensar as suas estratégias. Por sua vez, a análise visual é referida como a implementação de uma análise pertinente por parte das organizações, para lidarem com o conceito inicial.

Segue-se a secção 4 que expõe o conceito principal: Visualização de Informação. A revisão literária apresenta-se com uma sistemática descrição das temáticas envolvidas no processo de visualização e nos seus intervenientes.

A secção 5 descreve um conjunto de técnicas e propriedades visuais que servem de ponte entre a visualização e o resultado de uma representação gráfica valiosa.

A exploração de todos os temas abordados anteriormente esclarecem-se na secção 6, onde são divididos e/ou agrupados vários gráficos, formando famílias de gráficos com características comuns. Os gráficos foram descritos de acordo com a sua utilização e características de conteúdo, que são capazes de suportar quando exibem dados.

A penúltima secção intitulada de Visualização de Dados representa o estudo final da dissertação. Comunicar informação através de uma arte clara e efetiva recorrendo a meios gráficos para alcançar sucesso, é o principal objetivo da VI. A conexão teórica criada entre o propósito de visualizar e o próprio gráfico foi sintetizada num esquema que resulta como parte integrante deste trabalho.

Por fim, na conclusão são discutidos pontos fundamentais para o estudo da Visualização assim como os objetivos que foram alcançados e os que ficaram por alcançar, conduzindo a um incentivo para trabalhos futuros.

1.4 Abordagem metodológica

1.4.1 Revisão de Literatura

A fase de exploração desta dissertação iniciou-se com uma extensiva revisão literária a artigos e trabalhos relacionados com a área de investigação. Foi feita uma seleção de artigos e trabalhos cujos conceitos se enquadram na área de *big data*, *data analytics*, *visual analytics*, *visualização de informação*, *propriedades gráficas*, *tipos de gráficos* e *visualização de dados*. Para a recolha de bibliografia fiável para a comunidade científica foram usadas bases de dados bibliográficas como o *Scopus*, *Web of Science* e *Google Scholar*.

Oriundo do grande volume de dados que precisam ser analisados e interpretados surge o conceito de Visualização, que permite não só explorar dados visualmente como interpretá-los rapidamente, isto porque a perceção humana aumenta quando o sistema do olho tem à sua disposição um conjunto de atributos que formam uma representação. Neste contexto surge o propósito de construir uma classificação, por definição, classificação sistemática de conceitos, que vem a ser utilizada no âmbito da Ciência de Informação como ferramenta de organização intelectual. A organização de informação através deste conceito permite alocar, recuperar e comunicar informações dentro de um sistema de maneira lógica de navegação.

O objetivo incide sobre a construção de uma classificação que facilite o processo de representação gráfica para qualquer utilizador. Se for validada deverá ser colocada em portais institucionais e ou bibliotecas digitais como um novo mecanismo de consulta, ao lado de ferramentas de pesquisa, procurando diminuir os desperdícios de tempo na construção de uma visualização integrada e efetiva. No sentido de proporcionar uma melhor exposição do tema de investigação a exploração de *frameworks e ferramentas analíticas da web* foram indispensáveis para o entendimento da área.

1.4.2 Design Science Research

O método que se pretende usar para uma melhor orientação e execução deste trabalho de dissertação é o *Design Science Research*. Segundo um conjunto de técnicas e perspetivas a metodologia orienta a realização de projetos de pesquisa na área dos Sistemas de Informação.

O Design Science Research consiste principalmente em duas etapas fundamentais: construir e avaliar. Construir é o processo de desenvolver um artefacto para um propósito específico, por outro lado, avaliar consiste na ação de aferir a performance do artefacto na execução daquilo a que ele se propõe e no ambiente para o qual foi planeado para atuar.

Os estudos de Hevner et al (2004) delimitaram um conjunto de sete diretrizes que devem comandar o planeamento e execução de qualquer pesquisa que segue este método. A seguir são descritas as fases de produção do método conciliado com o objetivo principal deste trabalho e dissertação.

Diretriz 1 – Design como Artefacto

A construção de uma classificação para descrever representações consoante métodos analíticos deverá produzir um modelo confiável.

Diretriz 2 – Relevância do Problema

A construção do artefacto tem por motivo melhorar a representação correta dos dados em gráficos.

Diretriz 3 – Avaliação do Design

Propôs-se aferir a qualidade do artefacto produzido utilizando como experiência um conjunto de dados fornecido pelo orientador Jorge Sá. Uma vez que estava a decorrer no Departamento de Sistemas de Informação uma disciplina de Visualização de Informação seria interessante perceber o que os alunos iriam estudar se possível aplicar a classificação.

As propostas de avaliação inicialmente pensadas não foram realizados devido à falta de interesse na área. Contudo, é necessário aplicar mais esforços para o estabelecimento da visualização como uma disciplina que colmata o estudo integrante da área de BI

Diretriz 4 – Contribuições à pesquisa

Espera-se que o artefacto produzido contribua para a finalização de áreas que envolvam Gestão de Sistemas de Informação. Posto que é imprescindível explorar e comunicar resultados de forma clara e precisa, a adoção de representações gráficas torna-se preciosa nesta tarefa.

Diretriz 5 – Rigor da Pesquisa

A adaptativa escolha por uma linguagem clara na explicação de conceitos, temáticas, classificação de dados, gráficos e representações salienta-se devido ao imaturo estado de estudo da visualização de informação no contexto das TI.

Diretriz 6 – Design como Processo de Procura

A produção de um artefacto confiável deverá traduzir-se numa mais-valia, que proporcionará uma base de estudo para a visualização de informação.

Diretriz 7 – Comunicação da pesquisa

Para a comunicação eficaz dos resultados obtidos, espera-se um público-alvo orientado para a exploração de dados, análise e visualização de informação.

2. CONTEXTO DE ESTUDO

2.1 Big Data

O termo popular *Big Data* tem sido utilizado em todo mundo para descrever o crescimento exponencial de dados, desde o início do século vinte e um. O conceito aparece como um conjunto de tecnologias, processos e práticas que permitem às organizações retirar partido da análise do grande volume de dados para a tomada de decisão. A maior fonte de dados *Big Data* são as pessoas, e por isso o termo está presente em todas as áreas que envolvem atividade humana. Como o *Big Data* está associado a várias áreas e sectores a definição do termo tem variado consoante a especificidade de cada área. A primeira definição surge com (Gartner Inc., 2015) que descreve o *Big Data* como sendo uma tecnologia para processar um alto volume de dados, com alta velocidade e de várias naturezas.. Os conhecimentos obtidos devem obedecer ao custo/benefício da procura e uma análise de dados que garante uma maior viabilidade na tomada de decisões (Olshannikova, Ometov, Koucheryavy, & Olsson, 2015).

No *Big Data* há uma série de variáveis que fazem parte da sua composição. Além do volume significativo de dados surge também a sua variedade, uma vez que são coletados de diferentes fontes como *Enterprise Resource Planning* (ERP) ou comentários nas redes sociais. Outra das variáveis é a velocidade porque é necessário realizar análises em tempo real. Além destas começou-se a ter em conta a veracidade e o valor que os dados devem transferir para o negócio. Dados são os recursos naturais da sociedade de informação assim como o petróleo para a sociedade industrial, mas só geram valor se forem tratados e analisados (Olshannikova et al., 2015).

Emerge deste conceito desafios e problemas técnicos que foram classificados em três grupos: dados, processamento e gestão (Akerkar, 2013).

No que diz respeito aos dados muito foi referido como sendo os “5Vs do *Big Data*” (Merritte Stidston, 2014):

- **Volume:** refere-se à grande quantidade de dados gerados por pessoas e máquinas. Esta variável define o tamanho do conjunto de dados que se traduz num problema de armazenamento e análise para a tecnologia de base de dados;
- **Velocidade:** refere-se à rapidez com que dados são gerados e distribuídos, que se traduz num desafio quanto à análise em tempo real de dados;
- **Variedade:** refere-se aos diferentes tipos e formas de dados que derivam de diferentes fontes (dados estruturados e não estruturados);
- **Veracidade** – refere-se à integridade dos dados, que se pode traduzir em falhas de qualidade como incerteza, imprecisão, valores perdidos e distorção de dados;

- **Valor** – refere-se à descoberta de conhecimento que está relacionada com a qualidade da análise do conjunto de dados.

Os desafios de processamento são a recolha de dados, o cruzamento de dados de diversas fontes, a transformação e análise de dados e a representação visual final adequada ao processamento de informação humana – a percepção.

Os desafios de gestão estão relacionados com a segurança na recolha, no processamento e armazenamento de dados, que se traduz em problemas éticos e de privacidade controlados por políticas e regras dentro de uma organização ou noutro contexto que envolva armazenamento (Olshannikova et al., 2015).

Como o Big Data é um conceito ambíguo e a sua natureza ainda não foi definida, surge ainda numa linguagem de negócio os “5Rs do Big Data” (Merritte Stidston, 2014):

- **Relevant** – o volume de dados é tanto que torna-se essencial clarificar que informação é ou não relevante;
- **Real-Time** – acelerar o processo de análise de dados, desde a sua chegada até à sua representação final;
- **Realistic** – o processamento e análise do conjunto pré-definidos de requisitos que apõem todo o negócio. O utilizador deve utilizar a variedade de fontes de dados, assim como a veracidade para oferecer mais valor ao negócio;
- **Reliable** – a qualidade dos dados é fundamental para o apuramento de resultados porque fortes medidas de qualidade relacionam-se com bons resultados. Mais dados não significa mais qualidade, portanto colocar um limite no volume de dados torna-se uma tarefa importante;
- **ROI** – o processamento e a análise eficaz dos dados oferece ao utilizador uma boa tomada de decisão, maximizando o valor do investimento do sistema *Big Data*.

Considerando que o *Big Data* é um conjunto de conceitos teóricos que servem de base ao desenvolvimento de ferramentas para recolher, armazenar e processar dados, surgem desafios na recolha, no armazenamento, na análise, na partilha, na privacidade e na visualização de dados.

A visualização de dados é uma técnica indispensável na aquisição de conhecimento, que está submerso no abundante volume dados. O resultado de uma análise com gráficos e esquemas é utilizada há muito tempo mas o complexo conjunto de dados torna a tarefa mais difícil de ser percebida. O desenvolvimento de uma técnica de visualização para os dados heterogéneos que existe hoje em dia é útil para simplificar e transformar a informação numa representação perceptível ao ser humano (Olshannikova et al., 2015).

2.2 Data Analysis

A gestão e análise de dados têm vindo tomar dimensões significativas dentro da área das TI. Com a oportunidade de incorporar novas tecnologias sobre as práticas de *business intelligence (BI)* e *data analytics* os princípios da gestão de dados mantêm-se:

- Os dados são a fonte da informação e consequentemente de conhecimento e sabedoria, deste modo os dados devem ser apresentados em formato apropriado, conduzindo a um resultado favorável para tomada de decisão;
- Preservar a qualidade dos dados durante todo o processo de recolha e construção de resultado final. Para produzir o valor mais alto é fundamental que os dados se mantêm íntegros;
- O sucesso da utilização das práticas de BI e analytics não depende apenas da forma como os dados são processados. A forma como o resultado final é construído, apresentado e comunicado ao utilizador tem um impacto enorme na exploração de dados.

A transformação de dados em conhecimento é o objetivo da visualização de dados (*dataviz*) que é empregada para alcançar entendimento no conjunto de dados heterogéneo. Uma organização consegue extrair valor quer na visualização de dados, quer na resposta a questões específicas de negócio, quer na descoberta de soluções ou novas perguntas através da necessidade exploratória.

A visualização de dados encaixa na prática de BI por isso é vista como uma prática ou uma disciplina. Mais do que um conjunto de funções numa ferramenta de BI a visualização de dados transmite uma história ou um acontecimento, sendo o seu resultado preciso, atraente, oportuno e relevante.

A prática da visualização de dados apesar de melhorar a tomada de decisão ajuda a aumentar o poder da análise em toda a organização. O poder explicativo e exploratório da visualização de dados permite que o utilizador baseie as suas decisões em recursos visuais. Os meios visuais são mais eficazes do que os dados brutos. As ferramentas oferecem aos utilizadores recursos visuais para expor dados e funcionalidades de interação para uma análise rápida e vantajosa.

Uma solução de BI com o poder de fornecer um conjunto completo de recursos visuais de exploração e apresentação tem um impacto vantajoso, na hora certa com a história certa para uma organização. Se o utilizador deste produto beneficiar de conhecimento acerca de representações gráficas e tipos de dados certamente irá produzir uma análise visual competente (Stodder, 2013).

3. VISUALIZAÇÃO DE INFORMAÇÃO

A palavra visualizar significa tornar algo visual ou visível, já o conceito de visualização é definido como a transformação de conceitos abstratos em imagens reais ou mentalmente visíveis ou ainda, em contexto computacional, visualizar é a conversão de números ou categorias para um formato gráfico que pode ser facilmente compreendido (Spence, 2007).

A visualização deve combinar técnicas de acesso à informação para ajudar a encontrar informação útil. Pode tratar-te de informação da *World Wide Web (WWW)* ou da coleção de documentos organizacionais ou livrarias digitais (Chen, 2010).

3.1 O Conceito

O termo Visualização de Informação (VI) refere-se às representações gráficas interativas de informação, geradas por um computador (Chen, 2010). A VI surge como uma extensão da Interação Humano Computador (IHC). É uma área que integra técnicas de computação gráfica com o objetivo de auxiliar o processo de análise e compreensão de um conjunto de dados através de representações gráficas manipuláveis (Card & Mackinlay, 1997).

As representações visuais são expostas ao utilizador como uma tecnologia que combina pontos fortes de análise e apresentação de dados, onde o ser humano e o computador cooperam com as respetivas capacidades para obter os resultados mais eficazes.

Transformar dados não espaciais em representações gráficas intuitivas é o processo principal da área. É fundamental perceber:

- As representações gráficas que as ferramentas oferecem aos utilizadores;
- O tipo de dados que são passíveis de análise;
- Os dados que as representações albergam;
- A finalidade na utilização de uma representação.

O objetivo da VI é produzir uma representação visual interativa de dados abstratos com reforço da cognição humana, proporcionando ao utilizador a aquisição de conhecimento inerente à estrutura heterogénea de dado. Os dados heterogéneos ou abstratos não apresentam uma estrutura espacial intrínseca, o que dificulta o mapeamento de qualquer geometria, mas apoia-se em meios proporcionados pela VI para a sua representação visual (Chen, 2010).

Os modelos gráficos ou representações visuais variam de acordo com a natureza dos dados, que podem representar conceitos abstratos, valores ou relacionamentos do mundo físico associados a um espaço geográfico ou temporal. Qualquer que seja o caso o utilizador é responsável por associar o tipo de dados na mais adequada representação gráfica.

A capacidade de visualizar na totalidade um conjunto de dados mapeado numa representação é essencial para descobrir relações interessantes entre os dados.

A VI tem evoluído como a abordagem que transforma conjuntos complexos de dados em informações úteis para o utilizador através de um meio rápido e apelativo. O método básico do utilizador é gerar representações visuais interativas de informação que exploram capacidades percetuais do sistema visual humano e os recursos interativos da solução de problema cognitivos (Ware, 2004).

A aplicação entre o domínio da natureza de dados abstrata e o domínio das representações gráficas designa-se por mapeamento.

3.2 Perspetiva histórica

Desde há várias gerações que o processo de VI é colocado em prática em várias áreas, por isso é necessário estudar a sua evolução e impacto para perceber como as representações visuais mudaram ao longo do tempo. Elas são uma parte fundamental na aprendizagem e compreensão humana.

Existem dois tipos de representações visuais, ambas importantes mas com requisitos diferentes quando se constroem visualizações. O primeiro tipo é nomeado de Apresentação que utiliza dados para comunicar e persuadir uma audiência através da comunicação visual. O segundo tipo é nomeado de Visualização e refere-se à ideia de utilizar uma forma visual para adquirir conhecimento. Este termo implica que pessoas estejam envolvidas na sua prática para responder a questões.

Um trabalho importante que marcou a história da Visualização foi de William Playfair, um engenheiro considerado o pai da apresentação estatística. Ele foi pioneiro na técnica de representar propriedades visuais tais como barras, linhas e áreas para representar dados visualmente (ver figura 1). Uma forma simples e eficaz de descrever e organizar dados quantitativos. No entanto, se o conjunto de dados tem várias dimensões a representação pode tornar-se confusa (Friendly, 2006).

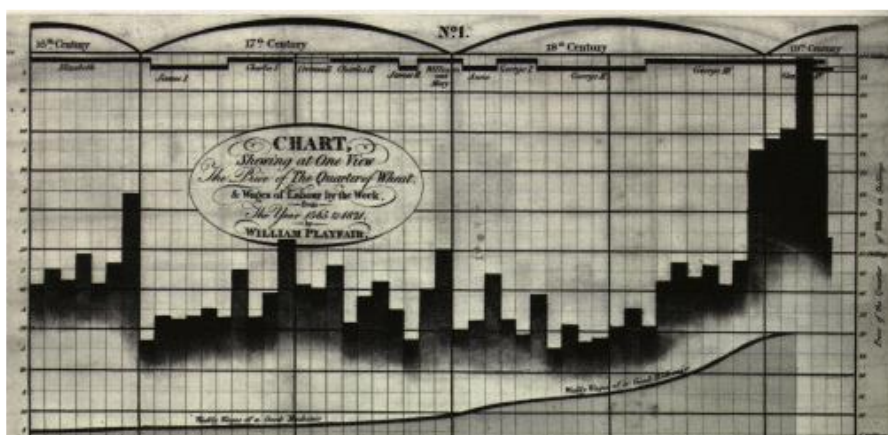


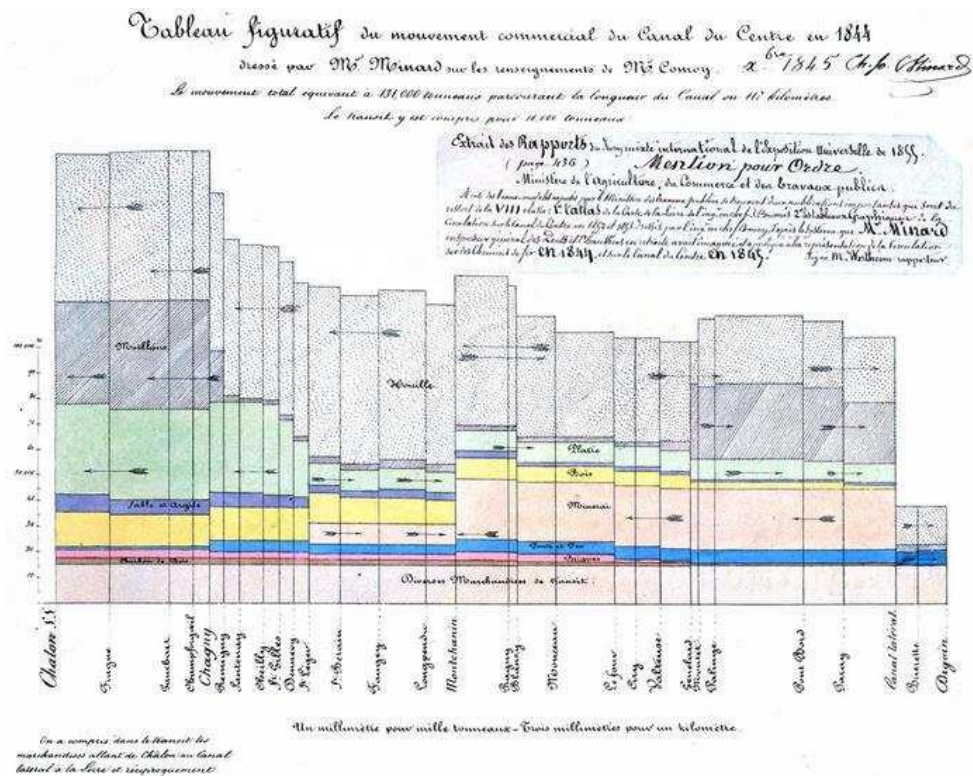
Figura 1- William Playfair – 1821 (Friendly, 2006)

Em Outubro de 1831, o surto da epidemia da Cólera abalou a Grã-Bretanha, matando mais de 52.000 pessoas. A causa da doença era desconhecida até 1855, quando Dr. John Snow produziu o famoso Mapa de Bolhas (ver figura 2). Num mapa da cidade John Snow representou as mortes das pessoas doentes através de pontos vermelhos. Depois de observar o Mapa, constata-se que a maioria dos casos insidiam sobre uma bomba de água na *Broad Street*. A mapificação das mortes provocadas pela doença na cidade ajudou a descobrir a fonte de contaminação (Mackinlay & Winslow, 2015).



Figura 2 - Dr. John Snow, *The Cholera Epidemic in London: 1854*

Na mesma época a utilização de gráficos começou a ser reconhecida na área da Economia. A utilização deste método na área foi reconhecido pelos trabalhos de Charles Joseph Minard. Um dos seus trabalhos exhibe o transporte comercial de mercadoria na França (ver figura 3). A largura de cada barra vertical exhibe a distância ao longo do percurso e as diferentes alturas de cada barra exibem a quantidade de mercadoria transportada, para mostrar a área proporcional ao custo do transporte (Friendly, 2006).



Outro trabalho famoso de Charles Minard é a apresentação da viagem histórica de Napoleão até Moscou (ver figura 4). O gráfico exibe uma barra castanha sombreada que representa o tamanho do exército e a espessura da barra declina progressivamente à medida que o exército caminha até Moscou. Na parte inferior do gráfico há uma linha que representa a temperatura exterior, o que influenciou a quantidade de tropas durante a viagem (Mackinlay & Winslow, 2015).

Mais tarde o reconhecimento de gráficos estatísticos foram entendidos como uma forma de alterar dados. O americano Darrel Huff em 1955 publicou o livro - *How to Lie with Statistics*, um exemplo de manipulação sobre dados (ver figura 5). O gráfico da esquerda representa a forma inicial dos dados, enquanto o gráfico da direita foi manipulado para um nível de detalhe mais elevado, onde se pode observar com mais detalhe a subida dos gastos. Este livro é importante para a utilização de técnicas de VI porque mostra como as representações visuais podem ser manipuladas para contar histórias diferentes (Mackinlay & Winslow, 2015).

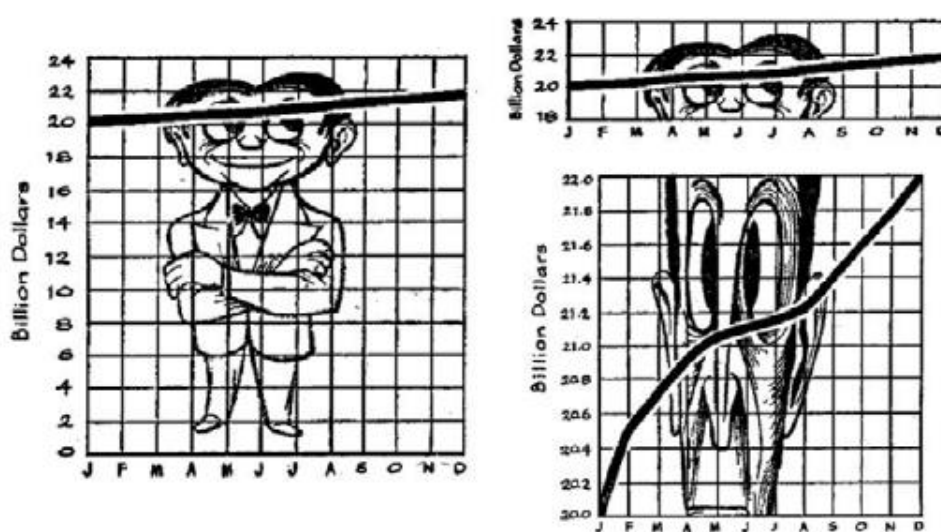


Figura 5 - Tipo de manipulação por Darrel Huff

Jacques Bertin introduziu um método de apresentação e estratégias de análise para múltiplos dados. O contributo da publicação do livro - *Semiology of Graphics*, foi importante na medida em que identifica três marcas visuais: linhas, pontos e áreas, e onde cada marca contém determinadas propriedades. Sendo elas: posição, cor, tamanho, forma, orientação e textura. A sua técnica designada de *Permutation Matrices* (ver figura 6) explica como manipular linhas e colunas numa tabela, com o objetivo de encontrar padrões nos dados desorganizados e mostrar a correlação entre os valores. Neste exemplo os valores de dados estão distribuídos pelos doze meses do ano durante dois anos (eixo horizontal), onde existe uma relação entre as várias categorias (eixo vertical), que representam característica da população que frequenta um hotel. O uso das cores representa os meses com mais ou menos afluência de uma determinada característica. Por exemplo: percebe-se que o preço (*prix*) é mais alto nos meses de Janeiro, Fevereiro, Março, Abril, Agosto, Outubro e Novembro, repetindo a prática para os dois anos representados.

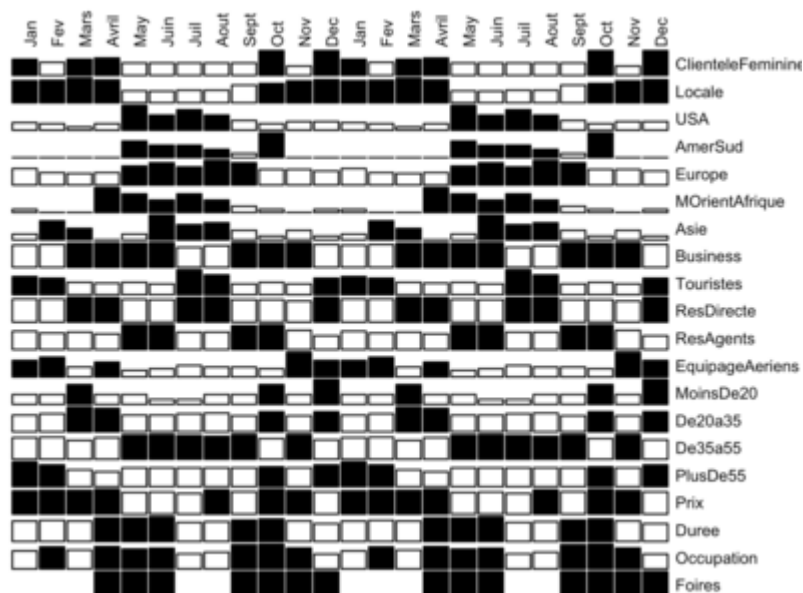


Figura 6 - Exemplo de Permutation Matrices Jacques Bertin 1967 adaptado (Bertin Matrices,2014)

O processamento informático dos dados estatísticos começou em 1957, com a criação da primeira linguagem de alto nível para a computação – *Fortran* – surgiu a possibilidade de construir novas formas gráficas em programas de computador.

Durante o avanço tecnológico outros temas começaram a surgir, abrindo portas para os primeiros exemplares de sistemas interativos para gráficos estatísticos 2D e 3D. A evolução dos métodos gráficos e técnicas de visualização estavam dependentes dos avanços tecnológicos, da recolha de dados e da teoria estatística.

Desde 1970 a 1980 estavam em causa gráficos estatísticos para dados quantitativos multidimensionais projetados para o analista observar a relação entre dimensões. Neste período os gráficos estatísticos tornaram-se a corrente principal, com a publicação de livros que incluíam capítulos dedicados a diagramas e gráficos, frequência absoluta e acumulada, métodos para encontrar a mediana e quartis, escolha de escalas e linhas para a representação visual de diferentes proporções, series temporais.

Em 1915 o trabalho de John W. Tukey, *“The Future of Data Analysis”* apelava à análise de dados, como um ramo da estatística e da matemática. Ele inventou novos gráficos com a publicação do livro *“Exploratory Data Analysis”* – gráficos de relação caule-folha, *boxplots*, matrizes, e muitos mais que entraram no vocabulário de implementação estatística.

Até ao século XXI novos gráficos históricos foram criados utilizando dados disponíveis, métodos, tecnologia e percepção humana. A aplicação da VI em várias áreas surge de objetivos concretos e

práticos: a necessidade de perceber comportamentos e relações por meio de representações gráficas eficientes (Friendly, 2006).

A aplicação de métodos modernos para a visualização de informação depende da própria história da visualização, onde vários autores contribuíram para a explosão da área entre a comunidade científica.

Modelo do Processo de Visualização

Os modelos do processo de visualização são uma referência para o utilizador, apresentando um conjunto de tarefas que deverá facilitar a criação de uma representação visual. Primeiro os dados brutos são recolhidos de sistemas operacionais e transformados em tabelas de dados, através do processo de transformação de dados; em seguida as tabelas de dados devem ser modificadas para estruturas visuais, através de métodos de mapeamento visual, com o objetivo de atingir a visualização efetiva. No processo de visualização identificam-se três fases básicas:

- 1) **Pré-processamento:** a transformação de dados compreende a modificação de dados brutos em relações lógicas, mais estruturadas para a compreensão humana;
- 2) **Mapeamento visual:** associação entre os dados e as representações gráficas. As estruturas visuais são constituídas: substrato espacial, marcas e propriedades gráficas;
- 3) **Representação:** é criada a imagem da representação visual.

A proposta para formalizar este modelo foi feita por (Card & Mackinlay, 1999), (ver figura 5) onde identificaram componentes importantes no processo de transformar os dados numa representação visual perante a utilização de uma técnica.

O modelo de referência sugerido por (Card & Mackinlay, 1999) foi adaptado ao contexto de estudo desta Dissertação, continuando a ser descrito como um mapeamento de dados brutos para uma representação visual, em que o utilizador tem influência sobre cada etapa do processo. No entanto é pertinente a inclusão de mais dois estágios no modelo, tendo em conta o contexto *big data*: **filtragem**, que seleciona o conjunto de dados que se pretende visualizar e a **interação**, onde o utilizador se relaciona com todo o processo de visualização.

A **filtragem** não é propriamente uma fase, uma vez que pode ser aplicada em qualquer momento do processo de VI. Um primeiro filtro deve ser utilizado após a recolha de dados, que devido ao imenso volume o utilizador deve criar funções com grau de interesse, escolher um conjunto de dados e restringir as variáveis que pretende explorar. Outros filtros podem ser incluídos na mapificação, se o conjunto de dados continuar a ser demasiado vasto. O número de representações não será suficiente para representar todas as variáveis, e estas terão de ser eliminadas. Ou ainda, se a escolha da representação se relacionar com uma escala de representação pequena, alguns elementos não vão ser representados.

Na fase de mapeamento visual a associação de dados a elementos gráficos, com o objetivo de criar informação, define-se pelo critério de escolha de qual a representação a utilizar na visualização. Os critérios estabelecidos podem ser condicionados:

- Pelas características das saídas de gráficos, por exemplo: a escala em que é feita a representação, que envolve conhecimento em escalas gráficas e tipos de dados;
- Pela escolha da análise que se pretende produzir nos dados através de relacionamentos, por exemplo: comparação ou distribuição.

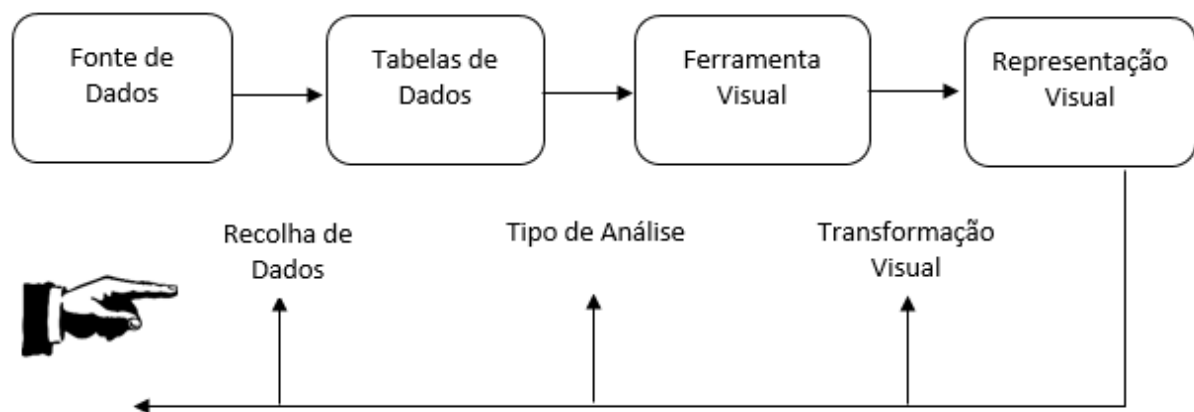


Figura 7- Modelo de processamento de Informação, adaptado de Card99

A **interação** não pode ser considerada uma fase mas um mecanismo de vantagem que o utilizador pode usar durante o processo de visualização:

- Aceder a um novo conjunto de dados para um novo pré-processamento, caso o conjunto de dados processados não for o total;
- Modificar os critérios de mapeamento visual quer na escolha de outras representações quer na escolha de novas variáveis;
- Alterar os critérios de filtragem aplicado ao conjunto de dados.

As setas indicam o fluxo de transformações que vão desde os dados não tratados até ao utilizador e vice-versa. Os dados brutos podem estar em diferentes formatos e são transformados em tabelas de dados. As tabelas de dados são projetadas em estruturas visuais, representações gráficas definidas por uma estrutura espacial, símbolos e propriedades gráficas. Por fim, estas estruturas são visualizadas de acordo com mecanismos de interação escolhidos pelo utilizador e resultado deve ser igual a uma **visualização efetiva**.

A expressividade, a efetividade, a consistência e a ordem são características que devem ser conseguidas na representação de dados, tornando a visualização mais clara e eficiente. Apenas informações relevantes devem ser tratadas pelo que uma representação deve ser de rápida interpretação. As

caraterísticas da imagem devem corresponder às caraterísticas dos dados e reunirem a informação mais importante com o método mais eficaz (Bertin, 1983).

Percepção Humana no Processo de VI

O meio ambiente físico e social estão envolvidos na recolha de dados, o primeiro refere-se como sendo a fonte dos dados, o segundo mais complexo determina a maneira como esses dados são armazenados e explorados.

As capacidades básicas do olho estão relacionadas com o problema da representação de uma visualização ideal. A visão humana contém milhões de fotorreceptores que são capaz de processar em paralelo o reconhecimento rápido de padrões (Ware, 2004). A largura de banda da visão como modo de comunicação leva à transferência de dados armazenados digitalmente na mente humana. Um dos benefícios mais importantes é a capacidade humana de raciocinar visualmente sobre os dados e extrair maior nível de conhecimento, além da transferência de dados simples (Card & Mackinlay, 1991).

A partir de uma representação visual facilmente se identifica uma ideia no conjunto de dados, deduzida por meio de conhecimento do padrão visual. Ideias que a partir de uma estrutura desorganizada de dados não são facilmente reconhecidas.

A VI é uma técnica efetiva porque funciona como um equilíbrio entre a percepção e a cognição. A visão corresponde à percepção visual do Homem, que ocorre desde o olho até ao córtex cerebral localizado na parte de trás do cérebro. O pensamento corresponde à cognição, o ato de adquirir conhecimento ocorre em primeiro pelo córtex cerebral e depois na parte frontal do cérebro. A visão tem um processo mais rápido e eficiente do que o pensamento.

A Visualização requer a utilização de um pensamento consciente para todo o trabalho, ao tirar partido dos olhos para uma maior utilização da percepção visual.

O primeiro contributo para a ciência da percepção foi desenvolvido pela *Gestalt School of Psychology* e tinha como objetivo descobrir como o Ser Humano percebia um padrão, uma forma ou como organizava tudo o que via. Desta forma, os estudiosos descobriram que organizamos tudo o que vemos de forma a fazer sentido, chamando a este acontecimento de comportamento visual.

A pesquisa resultou em princípios que auxiliam o utilizador no processo da visualização, através da criação dos seguintes grupos:






Proximidade	Objetos estão próximos uns dos outros que podem ser percebidos como grupos	
Similaridade	Objetos com atributos similares podem ser percebidos como grupos	
Caixa	Objetos que aparecem limitados por uma zona em comum podem ser percebidos como grupos	
Continuidade	Objetos aparecem alinhados ou como uma continuação do anterior podem ser percebidos como grupo	
Conexão	Objetos que aparecem interligados podem ser percebidos como um grupo	

Figura 8 - Principios de Gestalt

Estes princípios levam a que o utilizador seja capaz de obter novos *insights* sobre a perceção visual e a cognição, que são fatores importantes para a Interação Humano-Computador (IHC).

A capacidade do ser humano de processar informação em formato visual é muito mais rápida do que a de processar informação verbal. Neste sentido duas técnicas surgiram:

- *Preattentive visual processing*
- Mecanismos e limitações na atenção e memória

O processamento visual ocorre automaticamente no cérebro antes da cognição consciente e ocorre ao longo de várias etapas tratadas por neurónios especializados que se sintonizam para detetar atributos contidos na informação visual. Esses atributos podem ter diferentes comprimentos, tamanhos, cores, intensidade das cores, ângulos, texturas e formas.

O reconhecimento de manter vários itens em simultâneo na consciência leva a que a memória e a atenção sejam meios poderosos no armazenamento de informação. Um método é colocar diferentes visualizações de informação na frente dos olhos ao mesmo tempo, o que aumenta a capacidade de explorar dados multidimensionais e fazer comparações. Boas técnicas de visualização de informação quando usadas corretamente estendem o pensamento para novos domínios de análise analítica (Shneiderman, 1996).

Para alguns estudiosos, a visualização deve ser entendida como uma linguagem e não como uma ciência. Baseiam-se no facto da visualização ter como essência a construção de representações feitas de símbolos, e como podem transmitir significado. Neste sentido é essencial o estudo da Semiótica dos Gráficos. O estudo de símbolos e como eles transmitem significado chama-se Semiótica, esta disciplina foi originária nos Estado Unidos e desenvolvida na Europa pelo francês Ferdinand.

A obra de Bertin (1983) classifica marcas gráficas e como permitem expressar dados.

O modelo de processar informação na percepção visual humana apresenta-se em três fases:

- 1) **Processamento** – A informação visual é processada rapidamente por biliões de neurónios que trabalham em paralelo extraindo caraterísticas do campo visual como a cor, a textura, padrões e movimentos;
- 2) **Padrões** – Os processos ativos dividem o campo visual em regiões de padrões simples como os mesmos contornos de regiões, as mesmas cores, regiões com os mesmos padrões de textura e movimento. A fase do conhecimento padrão do processamento visual é flexível e influenciada pela quantidade de informação disponível a partir da primeira fase;
- 3) **Sequência visual de alto nível** – O nível mais alto de percepção são os objetos mantidos na memória de trabalho.

O processo de mapeamento visual fornece um ponto de partida para um projeto que envolva visualização, mas são necessários métodos mais avançados, porque a complexidade dos dados aumenta. A identificação de estruturas dentro da informação ajuda a orientar o processo de *design*, proporcionam organização de alto nível para um conjunto de dados e orientações para a conceção adequada de visualizações. Estas estruturas são importantes para modelos mentais dos utilizadores de informação, e devem ser utilizados no mapeamento visual.

3.3 Dados

A visualização de dados define-se como um meio para a exploração visual ou interativa de dados, de uma qualquer dimensão e a sua correspondente representação gráfica.

Se o objetivo da pesquisa na área de visualização é transformar dados em uma representação visual perceptiva e eficiente é essencial identificar a natureza dos dados, e como estes se dispõem num gráfico. Sabendo que a visualização de dados se baseia na capacidade humana de interpretação visual, e que através dessa interpretação se assimilam relacionamentos e padrões na procura de novos conhecimentos, é um interesse de estudo nesta área descobrir as relações que se podem traduzir em conhecimento. Os dados apresentam características com mais de duas dimensões, o que permite a representação geométrica desses mesmos. No entanto, existem dados com dimensões desconhecidas que não podem ser representados da mesma forma e nesse caso trata-se de informação, na forma abstrata, que se definem por dados categóricos.

As representações gráficas permitem identificar fenómenos e tendências não visíveis com outras técnicas, graças ao cruzamento de variáveis nos dados. As visualizações são criadas através do cruzamento de dados quantitativos e categóricos que justificam a escolha do tipo de representação gráfica a utilizar.

Para descrever métodos de visualização tendo em conta a qualidade dos atributos e o que são capazes de transmitir, é útil considerar dados quantitativa e informação categórica.

Vários autores desenvolveram pesquisas na área de VI distinguem naturezas de dados e um caminho útil para considerar a qualidade dos dados é a taxonomia escalas definida pelo estatístico (Stevens, 1946). De acordo com este há três escalas de medida nominal, ordinal e intervalar (Ware, 2004). Entenda-se por escala a representação dos tipos de dados categóricos numa representação gráfica, geralmente eixos vertical e horizontal.

A informação quantitativa por si só não representa apenas números, mas também dados que dizem o que os números medem.

Por esta razão não há vários tipos de dados mas sim duas variáveis em questão – os números – dados quantitativos e – os fenómenos que os números medem – dados categóricos.

Dados quantitativos são apenas números, em percentagem, inteiros ou decimais, são valores que pode representar ou ter uma ligação com uma categoria (Tufte, 2001).

Agrupar informação em categorias reduz grande quantidade do conhecimento armazenado. As categorias são a chave para organizar o conhecimento espacial, isto é, são conhecidas pelo seu relacionamento com o Mundo. Podemos categorizar relações e acontecimentos atribuindo-lhes grupos de especificações. As propriedades são consideradas características ou atributos de uma categoria (Few, 2005).

Por exemplo, dizer aleatoriamente a alguém - 943, aparentemente este número não quer dizer nada, mas se referenciar este valor como sendo os quilómetros da Costa Portuguesa já acrescentei conhecimento à informação que parecia confusa.

Um gráfico de duas-dimensões é representado por diferentes escalas: escala quantitativa, no eixo vertical e a escala categórica, no eixo horizontal.

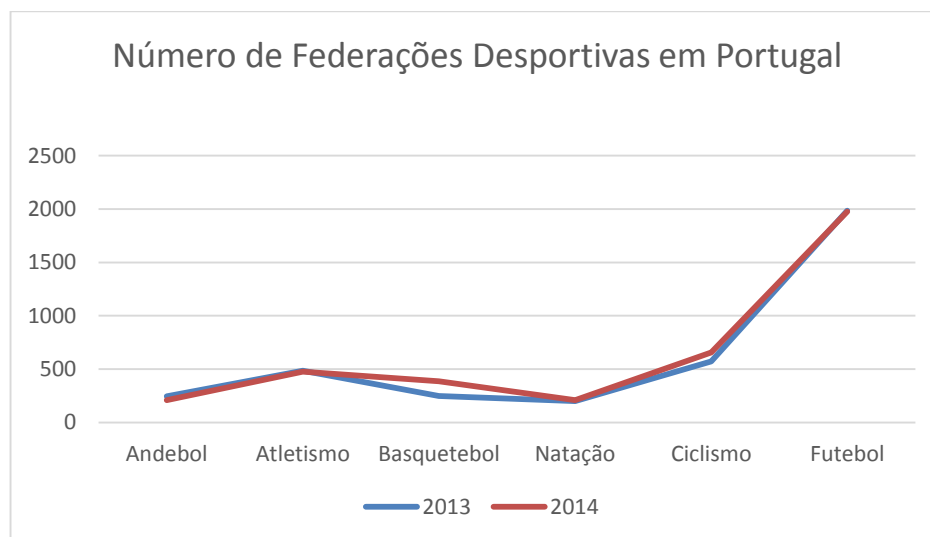


Gráfico 1- Escala quantitativa vs escala categórica

O gráfico 1 exhibe como as duas escalas se relacionam, e desta forma qual a melhor disposição para o mostrar. A escala categórica é medida pelos valores da escala quantitativa. É a escala categórica que identifica o objeto de análise, no entanto só é possível analisar esta informação se existir informação numérica, dada pela escala quantitativa.

O eixo vertical é representado pela quantidade de federações desportivas em Portugal, e o eixo horizontal representa as atividades desportivas (Andebol, Futebol...). As linhas contemplam duas referências temporais diferentes, o ano de 2013 (linha azul) e o ano 2014 (linha vermelha).

Desta maneira, o gráfico de linhas é a representação visual mais adequada aquando a comparação de várias variáveis.

Os gráficos de duas-dimensões exibem duas escalas diferentes, com exceção dos gráficos de dispersão, em que as suas exibições são definidas apenas por escalas quantitativas (ver gráfico 2).

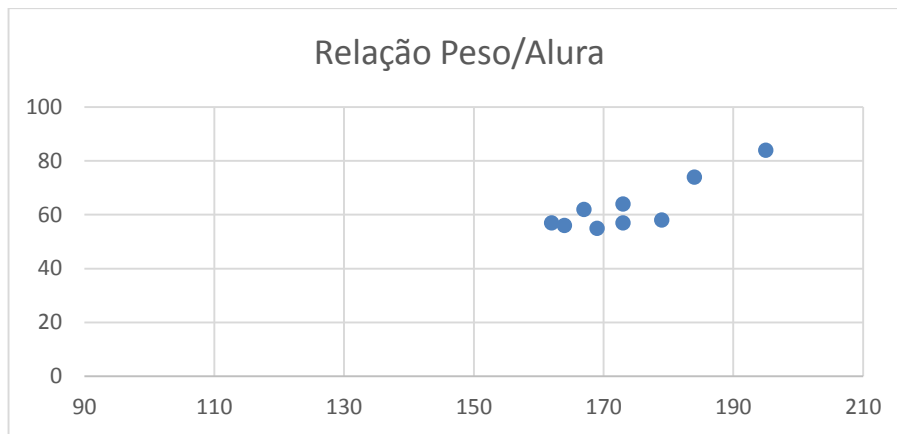


Gráfico 2 – Duas escalas quantitativas

As escalas categóricas quando são definidas nos gráficos, assumem três tipos (Stevens, 1946):

- Nominal: Elementos discretos e sem ordem, que mesmo pertencentes à mesma categoria não têm nenhuma relação direta entre si (ver gráfico 3).



Gráfico 3 – Áreas de Estudo: Escala Categórica Nominal

- Ordinal: Elementos que têm uma ordem intrínseca, e tal como as escalas nominais não se relacionam uns com os outros (ver gráfico 4).

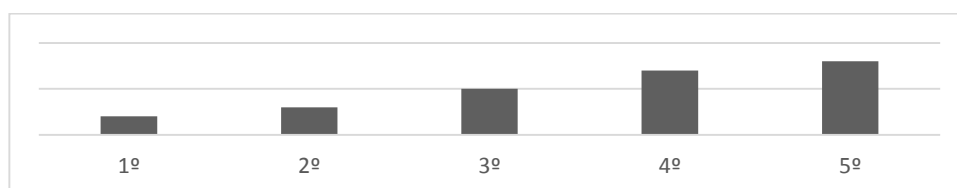


Gráfico 4 - Ranking: Escala Categórica Ordinal

- Intervalar: Elementos que têm uma ordem intrínseca e representam valores quantitativos. Esta escala começa como escala quantitativa que se converte numa escala categórica que se divide em intervalos (ver gráfico 5).

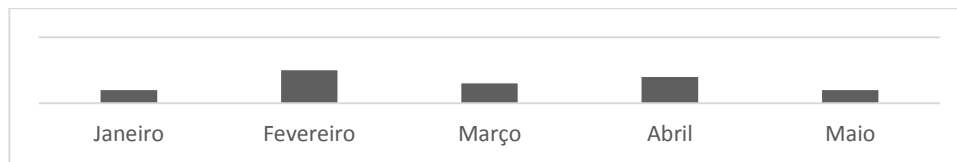


Gráfico 5 - Exemplo: Escala Categórica Intervalar

Neste caso, o tempo é dividido em períodos com intervalos iguais. Como os meses, os dias ou as horas são elementos que correspondem a medidas temporais também estão definidas para as escalas quantitativas.

4. REPRESENTAÇÕES VISUAIS

A VI pretende a exibição gráfica de informação sem significado, para análise e posterior apresentação. Os dados contêm histórias importantes e a área da VI oferece recursos poderosos para a descoberta dessas histórias assim como a sua apresentação a uma audiência.

As representações visuais são um poderoso recurso, que auxiliam o mapeamento entre os dados e um conjunto de propriedades gráficas com características específicas. Através de variáveis visuais, substrato espacial e marcas visuais é possível criar uma imagem exploratória e representativa.

Converter dados em informações com conhecimento em tempo útil, é uma mais-valia para o mercado atual. Analisar dados para favorecer o negócio, melhorar a experiência dos clientes, corrigir erros nos processos operacionais, prever situações indesejadas ou comunicar uma ideia são objetivos na representação gráfica de dados.

Um dos desafios da área de VI é facilitar o processo de criação visual e aquisição de conhecimento, através de representações adequadas e satisfatórias que explorem os dados no seu todo (Spence, 2007).

4.1 Técnicas de Visualização

A análise de dados, também conhecida como *data mining* ou a descoberta de conhecimento permite extrair informações preciosas a partir de dados desorganizados por meio de algoritmos (Keim, Andrienko, Fekete, & Carsten, 2008).

O processo de limpar, transformar e modelar dados para descobrir informação útil surge agora na análise visual. A exploração e análises visuais que resultam da utilização de técnicas de visualização oferecem conhecimento suficiente para substituir a utilização de outras técnicas de *data mining* (Keim et al., 2008).

No processo de VI, deve ter-se em conta a escolha das melhores técnicas e aplicá-las corretamente. Porque a utilização inadequada dessas técnicas geram resultados incorretos ou inconclusivos. De forma a evitar esses erros é necessário conhecer a técnica de visualização que irá conduzir a uma representação gráfica adequada para os dados.

É conveniente explorar características dos dados que influenciam as técnicas de visualização para estabelecer diretrizes.

As diretrizes escolhidas são técnicas de visualização *standard*, que descrevem cinco classes de técnicas (Keim, 2002):

- Gráficos padrão 1D, 2D e 3D;
- Técnicas iconográficas;

- Técnicas geométricas;
- Técnicas orientadas a *pixel*;
- Técnicas baseadas em gráficos ou hierarquias.

Os **gráficos padrão 1D, 2D e 3D** são mais utilizados na estatística para alcançar uma estimativa, uma certeza, uma hipótese, a distribuição ou frequência de um atributo no conjunto de dados. Normalmente representam um a três atributos e são utilizados para análise de dados quantitativos. Todos os gráficos desta categoria estão preparados para representar dados quantitativos. Esta classe fornece gráficos para fazer comparações, classificações e determina a ligação entre atributos. Gráficos estatísticos são usados na análise de dados para descobrir padrões, estruturas nos dados e identificar *outliers*.

Nas **técnicas iconográficas** os atributos são mapeados de acordo com propriedades que variam de acordo com o valor dos atributos. Gráficos de ícones representam cada dado individual o que permite perceber o comportamento de padrões dos dados.

Nas **técnicas geométricas**, atributos multidimensionais são mapeados num plano de duas-dimensões, que permite a visualização de todos os atributos. Os gráficos representados por estas técnicas são adotados para representar dados qualitativos e quantitativos. Gráficos que permitem uma vista pormenorizada ou global sobre todos os dados e que verificam a correlação entre atributos, padrões, comportamentos e regras são uma opção eficaz para grandes conjuntos de dados.

Nas **técnicas orientadas a *pixel*** cada atributo é mapeado num *pixel* de cor e colocado no ecrã dividido em janelas que correspondem a um atributo. Neste caso podem ser analisada a correspondência de relações entre os atributos.

Técnicas baseadas em **gráficos e hierarquias** têm uma estrutura de relação entre os dados. Podem ser representadas por mapas de árvore, gráficos de mosaicos ou outros. Os gráficos utilizados para esta técnica são ideais para exibir quando existe uma ou mais relações entre os dados.

Keim, classifica outros parâmetros importantes na adoção de técnicas de visualização: tipos de dados, tarefas do utilizador, dimensionalidade e posição dos atributos no gráfico. Estes parâmetros influenciam a escolha da técnica de visualização em função da característica dos dados.

O tipo de dados é um fator determinante na escolha de uma técnica de visualização. A classificação dos dados está de acordo com a dimensionalidade (capacidade de representar atributos), e com a natureza dos dados, (numéricos ou categóricos) (Shneiderman, 1996).

As tarefas do utilizador são outro fator determinante na escolha de uma técnica. O que corresponde às atividades que o utilizador pode executar de acordo com o objetivo da representação gráfica. Na literatura de (Keim, 2002), (Shneiderman, 1996) e (Pillat, Valiati, & Freitas, 2005) refere-se um conjunto das atividades de interação:

- Zoom – Vista total ou detalhada sobre todo o conjunto de dados;
- Correspondência entre atributos – Grau de relações entre variáveis pode revelar padrões, comportamentos ou tendências;
- Identificação de Outliers – Conjuntos de dados com comportamentos irregulares quando comparados com o resto dos dados;
- Identificação de características importantes;
- Identificação de conjuntos de atributos com comportamentos semelhantes.

4.2 Mapeamento Visual

Os gráficos têm dois usos distintos, o primeiro é utilizado como um meio para comunicar um tipo de informação, e o segundo é utilizado como processamento gráfico, em que cada utilizador pode manipular objetos gráficos, de forma a aumentar conhecimento do que está a visualizar (Bertin, 1983). Este processamento inicia-se com dados destrutturados que existem na mais variada forma e precisam ser codificados na representação visual.

Existem duas formas fundamentais de dados - os valores as estruturas de dados que foram apelidadas de entidades e relacionamentos. Entidades são os objetos de dados que se pretende avaliar e os relacionamentos são o propósito da visualização. Estruturas e padrões que se pretendem descobrir (Steele & Iliinsky, 2010):

1. Entidades: são geralmente o objeto de interesse, dados recolhidos de fontes aleatórias;
2. Relacionamentos: baseiam nas estruturas que relacionam os atributos e criam um conhecimento. Neste trabalho optou-se por substituir a designação de relacionamentos para métodos analíticos, que permitem a inferência de conhecimento. Os métodos analíticos são uma abordagem da VI por isso serão descritos com mais pormenor na secção: Visualização de Dados.

Substrato espacial: Um aspeto fundamental acerca da VI é a utilização do espaço. A posição espacial é uma boa forma de codificar dados (Card & Mackinlay, 1999). Existem quatro tipos elementares de eixos:

U = Unstructured Axis (sem eixos)

N = Nominal Axis (um País dividido em Distritos)

O = Ordinal Axis (ordenar os Distritos)

Q = Quantitative Axis (Os Países têm métricas)

Os eixos são um alicerce base da construção de uma representação gráfica e seguindo esta afirmação passa a ser uma característica espacial dos gráficos três sistemas de coordenadas:

Sistema de Coordenadas Cartesiano – Chama-se de espaço cartesiano ao esquema onde se especificam pontos num espaço com dimensões (D). Um sistema de 2D (eixos x,y) permite localizar um ponto acima e abaixo, além de esquerda e direita, mas a inclusão de mais uma reta (eixo z) permite localizar o objeto para a frente ou para trás - sistemas 3D;

Sistemas de Coordenadas Polares – Chamam-se coordenadas bidimensionais quando um ponto num plano é definido pela sua distância em relação à origem e ao ângulo do raio;

Sistema de Coordenadas Geográfica – O sistema de mapeamento da Terra, expressa qualquer posição no planeta. As coordenadas geográficas assentam em várias linhas imaginárias, horizontais e verticais traçadas sobre o globo terrestre.

O globo é dividido em latitude, que vai desde de 0 a 90 graus (Norte ou Sul) e longitude, que vai desde 0 a 180 graus (Leste ou Oeste).

Variáveis visuais: O significado da representação gráfica é expressa através de um conjunto de variáveis:

Tamanho e Valor – são variáveis de imagem, porque a constroem;

Orientação, Cor e Textura – são variáveis de separação, porque separam e diferenciam os elementos da imagem.

Marcas: As variáveis visuais são implementadas através das marcas que definem o objeto visual que ocorre no espaço. Há quatro tipos elementares de marcas:

P = Points

L = Lines

A = Areas

V = Volumes

Princípios da Visualização

A visualização é um método de manipular dados com a finalidade de os ver em novas formas.

Existem cinco princípios para criar representações visuais (Tufte, 2001):

1. Aplicar **comparações** nas visualizações – usar diferentes tamanhos e cores para realçar comparações visuais;
2. Mostrar **causalidade** – reforçar qual a relação que existe entre algo que acontece e existe e mostrar o acontecimento que causa isso;
3. Mostrar dados com **múltiplas variáveis** – usar dados com duas dimensões ou mais, acrescentar mais informação;

4. Integrar **todos os elementos visuais** – usar textos, imagens, números e formas adequadas para visualização em vez de reduzir a representação ao seu título;
5. Criar **conteúdo** de *design* – qualidade de dados, relevância na apresentação e integridade.

4.3 Métodos Analíticos

As representações gráficas contemplam uma diversidade de funcionalidades que se expressam em diferentes tipos de dados, tarefas, cenários e propósitos.

De acordo com Few (2005), identificam-se oito tipos de métodos analíticos que permitem classificar representações gráficas para que seja mais intuitivo comunicar uma mensagem visual e eficaz a partir de conjuntos de dados. Os métodos a reter são:

1. **Séries de tempo:** a observação de variáveis capturadas ao longo de um período de tempo permitem a verificação de uma tendência;
2. **Classificação:** observação ordenada de variáveis categóricas medidas em função do seu desempenho;
3. **Parte de um todo:** as subdivisões de variáveis categóricas pode ser observada quando apresentada como uma proporção em relação ao total do conjunto;
4. **Desvio:** a observação de um conjunto e variáveis comparado com um conjunto de variáveis de referência;
5. **Distribuição:** a observação dos valores de uma variável específica, num determinado intervalo permite perceber o seu comportamento;
6. **Correlação:** a observação da distribuição de duas variáveis permite determinar o movimento/comportamento de uma variável em função da outra;
7. **Comparação:** comparar variáveis categóricas sem qualquer ordem que influencie o comportamento da variável;
8. **Geográficos:** observação de uma variável relacionada com um espaço físico.

Os métodos surgem na medida em que explorar dados na forma visual permite a descoberta de novas coisas e novos comportamentos. Aplicar princípios da estatística juntamente com características dos dados e dos seus relacionamentos irá facilitar a forma como os dados podem ser organizados para a descoberta de novas informações.

5. GRÁFICOS DE INFORMAÇÃO

Os gráficos surgem com o propósito de explorar, analisar e comunicar dados. Efetuar uma análise através de representações visuais aumenta capacidade de entendimento e absorção de informações. O avanço das tecnologias e ferramentas gráficas, impulsionaram a diversidade de modelos gráficos que o utilizador pode adotar como recurso. Através da exploração de propriedades gráficas, surgem vários tipos de representações visuais. Desde tabelas a esquemas de diagramas, de gráficos não convencionais designados de *charts* a simuladores e medidores de desempenho. As propriedades gráficas devem ser aplicadas nas mais variadas formas, desde que revelem certezas no conjunto de dados.

Como tal, existem vários gráficos e relacionar o propósito da visualização com as características gráficas é ponto de partida para a descrição dos trinta e seis gráficos que surgem a seguir (Schreiner, 1997). Deste variado conjunto foram classificadas cinco principais famílias de representações: gráficos, diagramas, tabelas, charts e mapas.

As representações descritas foram escolhidos e definidas com base nas características da sua utilização, sistema de coordenadas, marcas gráficas e preparação para a exibição de dados quantitativos e qualitativos.

5.1 Gráficos

A classe de representações designadas de gráficos funcionam quando se pretende analisar ou comunicar uma mensagem que está contida na forma dos dados.

5.1.1 Gráfico de Linhas

Os gráficos de linhas são utilizados para exibir dados quantitativos ao longo de um período de tempo (contínuo ou intervalar). Ideais para revelar tendências e relações quando agrupados com outras linhas (nova serie de dados).

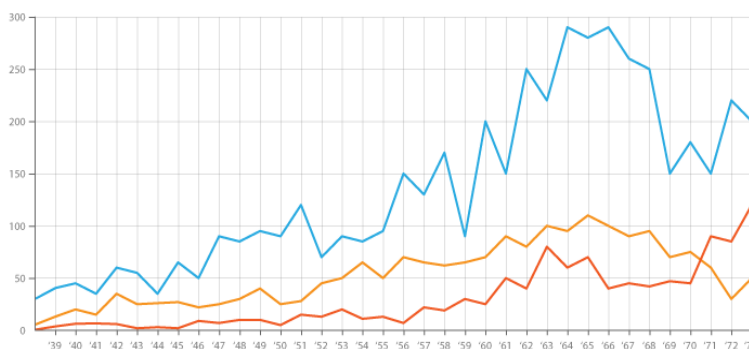


Gráfico 6 - Line Graph

5.1.2 Gráfico de Área

Os gráficos de área são equivalentes aos gráficos de linhas mas com a área por baixo das linhas preenchidas por uma cor ou textura. Os dados são exibidos através do sistema de coordenadas cartesianas. São utilizados para acompanhar o desenvolvimento de dados quantitativos ao longo de um período, úteis para exibir tendências e relações, uma vez que transmite valores específicos.

As variações deste gráfico podem ser: gráfico de Áreas Agrupadas (o eixo de partida é o zero) e gráfico de Área Empilhadas (cada nova série de dados parte da série antiga).

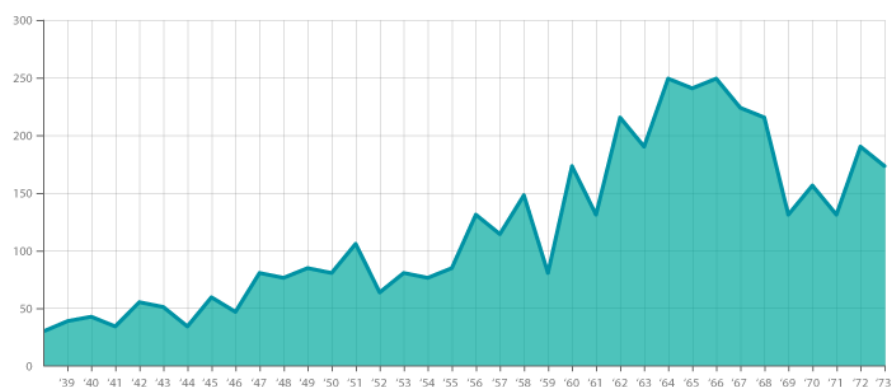


Gráfico 7 - Area Graph

5.1.3 Gráfico de Barras

Estes gráficos podem contemplar o uso de barras horizontais ou barras verticais (conhecido como gráfico de colunas). São utilizados para exibir comparações numéricas e discretas entre categorias. Enquanto um eixo representa valores quantitativos (normalmente eixo vertical), o outro representa dados com categorias específicas (normalmente eixo horizontal).

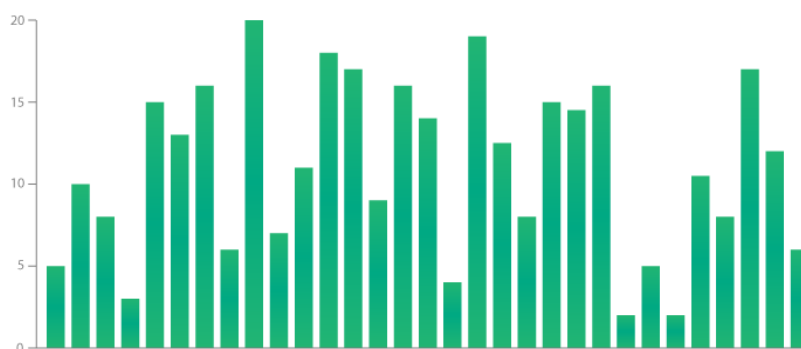


Gráfico 8 - Bar Graph

5.1.4 Caixa de Bigodes

É uma representação ideal para exibir grupos de dados quantitativos, através dos quartis. As linhas do gráfico indica os quartis inferior e superior e os *outliers* são apresentados por pontos individuais fora das caixas. O seu uso é relevante para a estatística descritiva.

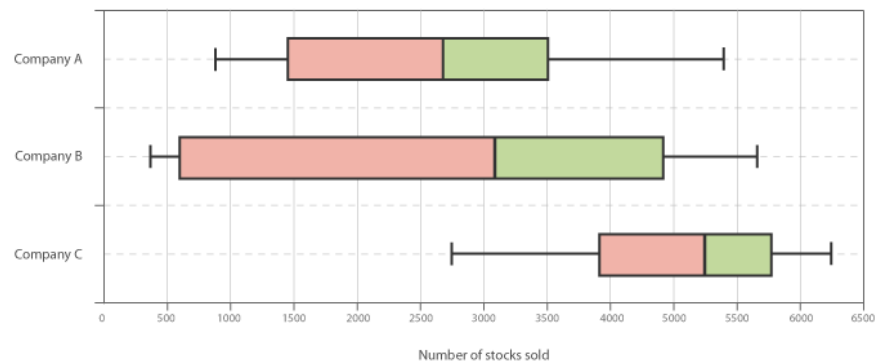


Gráfico 9 - Boxplot

5.1.5 Gráfico de Bolhas

O gráfico de bolhas representa múltiplas variáveis. É uma combinação de um gráfico de dispersão com um gráfico de área proporcional. As variáveis são exibidas através do sistema de coordenadas cartesianas.

Nesta representação quando há excesso de informação (apresentadas por bolhas) dificulta a sua leitura por parte do utilizador, o que pode ser contornado por ferramentas de interatividade.

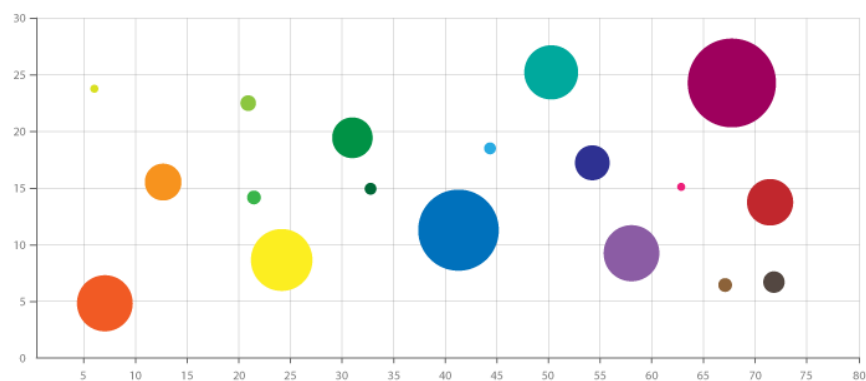


Gráfico 10 - Bubble Graph

5.1.7 Histograma

É uma representação ideal para exibir como os dados estão distribuídos num período de tempo (contínuo ou intervalar). Cada barra do gráfico é igual à quantidade de dados e são utilizadas para estima se concentram os valores de dados.

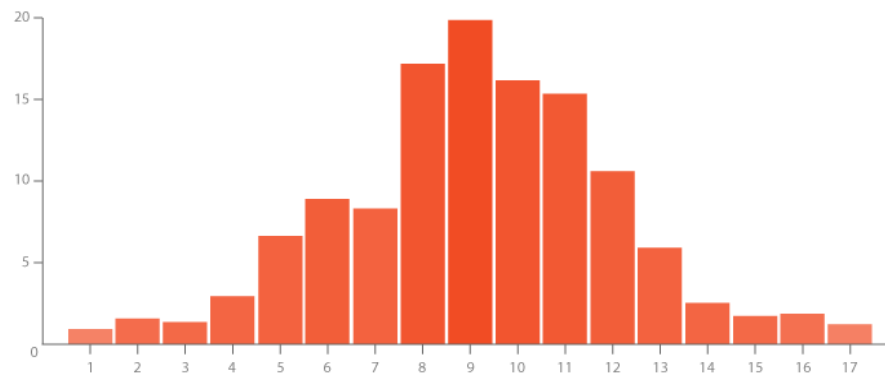


Gráfico 11 - Histogram

5.1.8 Gráfico de Mosaicos

Esta representação exibe informação categórica através de várias variáveis. Ambos os eixos contêm um conjunto de variáveis, determinadas pela largura e altura de cada segmento. Ideal para revelar relações entre as categorias e subcategorias. Utiliza o sistema de coordenadas cartesiano.

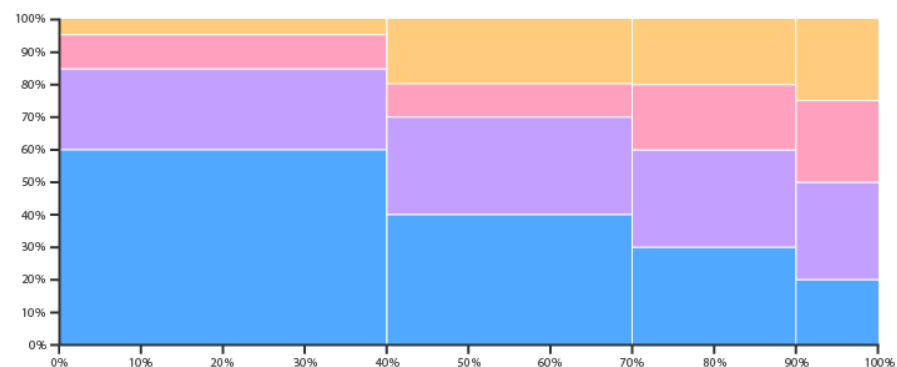


Gráfico 12 - Marimekko Graph

5.1.9 Gráfico de Colunas Agrupado

Esta representação exibe dois ou mais conjuntos de dados. São dispostos lado a lado e agrupados em categorias, todas dispostas no mesmo eixo (normalmente eixo horizontal). Assim como o gráfico de barras, cada comprimento de barra exibe

comparações numéricas das categorias. Cada série de dados é representada por uma cor de maneira a distingui-las.

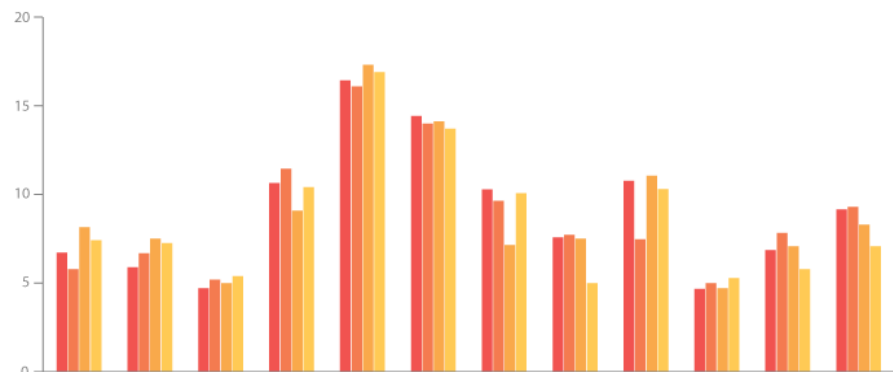


Gráfico 13 - Multi Set Bar Graph

5.1.10 Pirâmide de População

Esta representação é equiparável a um histograma, que melhor se adequa às características da uma população. Exibe a distribuição de dados populacionais nos dois eixos. O eixo horizontal normalmente representa números da população, enquanto o eixo vertical exibe as características.

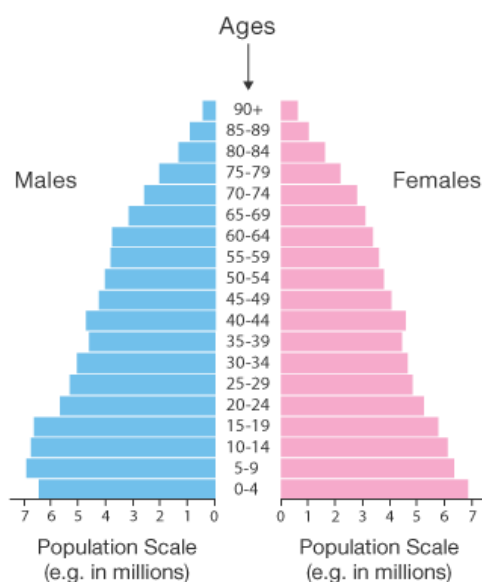


Gráfico 14 - Population Pyramid

5.1.11 Gráfico de Radar

O gráfico de radar é utilizado para comparar dados quantitativos com múltiplas variáveis. É possível ver quais as variáveis com valores semelhantes ou discrepantes.

São ideais para exibir as variáveis que marcam um alto ou baixo dentro de um conjunto de dados, tornando-se o principal gráfico quando se trata da representação de um desempenho. Podem ser comparadas diferentes categorias de desempenhos.

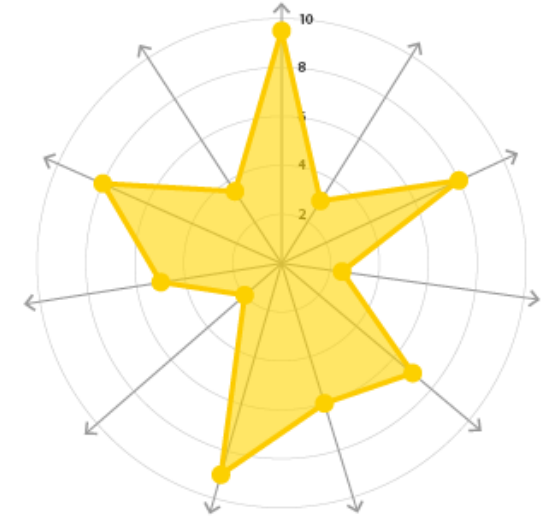


Gráfico 15 - Radar Graph

5.1.12 Gráfico de Barras Radial

São uma representação equiparável ao gráfico de barras mas disposto num sistema de coordenadas polares, em vez de cartesiano. A desvantagem deste é que os comprimentos das barras podem ser mal interpretados uma vez que cada barra de fora fica ligeiramente maior do que a de dentro. Não existe proporção entre o tamanho da barra e o valor que ela representa.



Gráfico 16 - Radial Bar Graph

5.1.13 Gráfico de Dispersão

Os gráficos de dispersão exibem um conjunto de pontos através do sistema de coordenadas cartesianas, para revelar valores entre duas variáveis. Atribuem uma variável a cada eixo, o que permite detetar se existe uma ou mais relações entre as duas variáveis. Na análise visual surgem os típicos padrões dos gráficos de dispersão: valores positivos (aumentam em conjunto), negativos (valores aumentam enquanto outros diminuem), nulos (não há correlação), lineares e exponenciais.

A sua utilização é ideal para avaliar o impacto que uma variável tem na outra.

Se o utilizador pretender pode ainda acrescentar uma linha ou curva para complementar a análise, conhecida como linha de tendência.

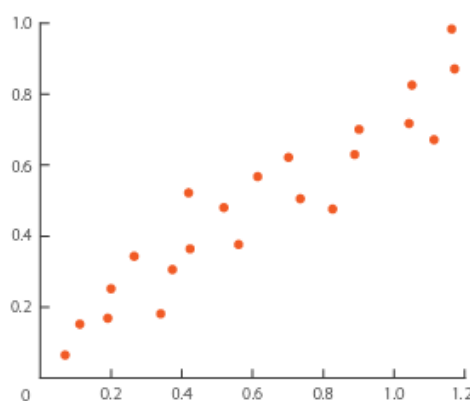


Gráfico 17 - Scatter Plot

5.1.14 Gráfico de Área Agrupado

São representados como os gráficos de área, exceto se o utilizador pretender exibir múltiplos conjuntos de dados. Cada conjunto começa no ponto que acaba a série de dados anterior. Utiliza a área para comunicar números inteiros.

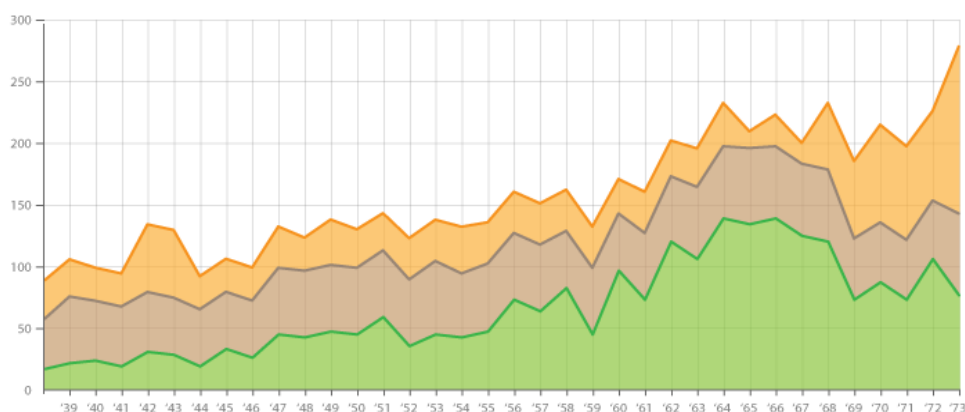


Gráfico 18 - Stacked Area Graph

5.1.15 Gráfico de Barras Agrupado

Estes gráficos representam numa barra, segmentos de outras barras que correspondem a vários conjuntos de dados. São utilizados para demonstrar como uma categoria maior é dividida em categorias menores e qual a relação de cada parte sobre o todo. Existem dois tipos deste gráfico:

Gráficos de barras simples: Os valores são postos após o segmento anterior. O valor total da barra é igual à soma de todos os segmentos. Ideal para comparar montantes totais através de cada barra.

Gráfico 100% agrupado: Exibe a percentagem de uma parte de um grupo total. É representado pela percentagem que cada valor tem para o montante total de cada grupo.

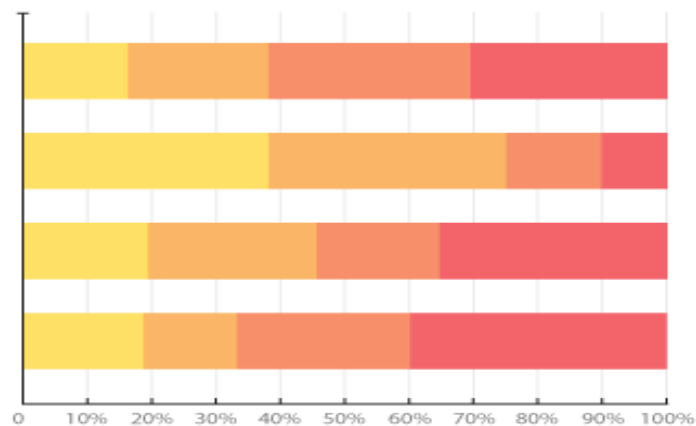


Gráfico 19 - Stacked Bar Graph

5.1.16 Gráfico de Palmas

Os gráficos de palmas são utilizados para exibir intervalos de conjuntos de dados. A sua desvantagem é concentrar os dados em valores extremos, não realçando a informação de valores que se encontram no meio.

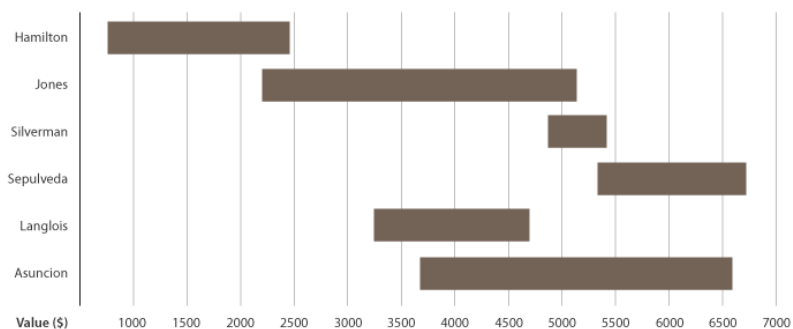


Gráfico 20 - Span Chart

5.2 Diagramas

A classe de representações formada pelos diagramas funcionam quando se pretende comunicar uma mensagem que está contida nas relações internas de um conjunto de dados ou entre diferentes conjuntos.

5.2.1 Diagrama de Arcos

Este diagrama é uma alternativa para quando se pretende representar gráficos 2D. Os nós são posicionados ao longo do eixo 1D (eixo horizontal) e os arcos são utilizados para mostrar as ligações entre os nós. A largura das linhas dos arcos podem ser utilizadas para representar a frequência desde a fonte até ao nó-alvo. São ideais para encontrar ocorrências nos dados.

A desvantagem desta representação é que eles não mostram as estruturas e as conexões entre os nós tão bem como gráficos 2D e muitas ligações desordenadas podem fazer com que o diagrama seja difícil de ler.

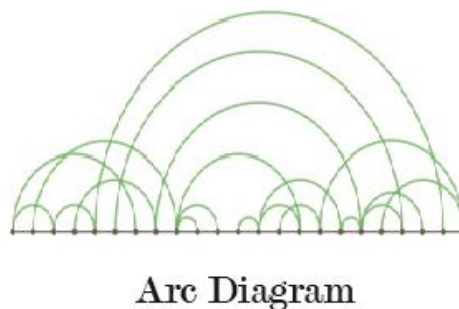


Gráfico 21 - Arc Diagram

5.2.2 Mapas Mentais

Estes diagramas são utilizados como mapas, para associar palavras, imagens, ideias ou conceitos. Geralmente utilizados como ferramenta nas primeiras etapas de um projeto, para gerar ou clarificar ideias, fazer associações, organizar informação e visualizar essa estrutura.

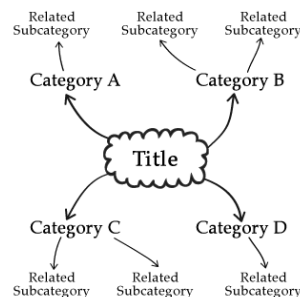


Gráfico 22 - BrainStorm

5.2.3 Diagrama de Cordas

O diagrama de cordas exibe as relações internas entre as entidades. As ligações entre entidades mostram que elas partilham algo em comum. Ideais para comparar semelhanças entre o mesmo conjunto de dados ou entre conjuntos diferentes.

Os nós estão dispostos em volta de um círculo, onde as relações entre os pontos ligadas são arcos ou curvas. Os valores são atributos entre cada ligação que é representada em proporção com o tamanho de cada arco. O uso da cor pode ser utilizada para distinguir duas ou mais séries de dados.

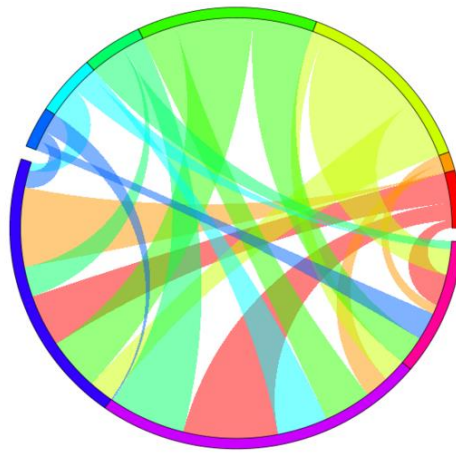


Gráfico 23 - Chord Diagram

5.2.4 Diagrama de Ilustração

Estes diagramas utilizam-se para explicar conceitos ou métodos, descrever objetos ou lugares e mostrar como coisas se movem ou mudam. São representados por imagens acompanhadas com legendas.

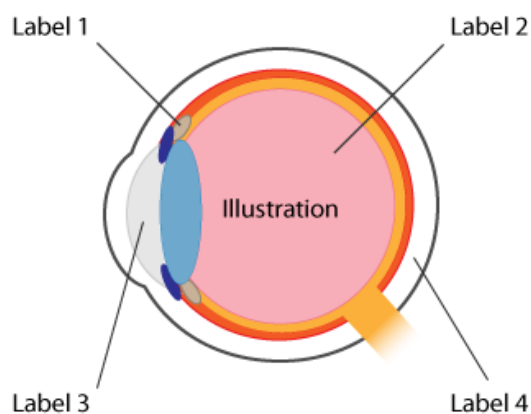


Gráfico 24 - Illustration Diagram

5.2.5 Diagrama de Árvore

O diagrama de árvore representa hierarquias nas suas estruturas. Geralmente a estrutura destes diagramas consistem em elementos com um nó-raiz e um membro que não tem pai superior. Podem existir nós que estão ligados em conjunto em conexões de linha, como ramos, que representam as relações entre os membros. Por fim os nós de folha são membro sem nós-filho.

A sua utilização recai essencialmente na exibição de famílias com relações descendentes, na classificação científica, na ciência da computação, na matemática e em empresas e organizações para fins de gestão.

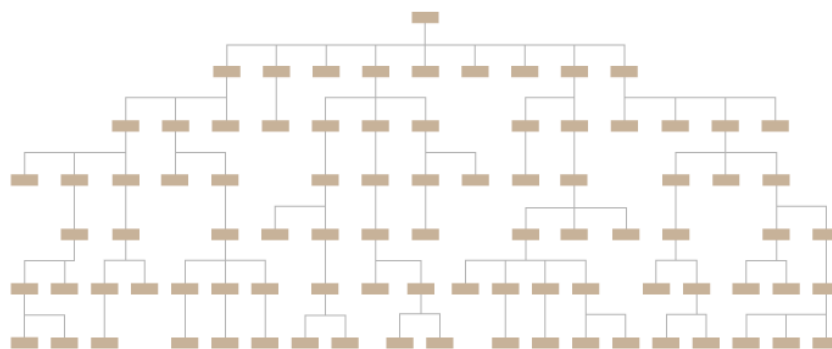


Gráfico 25 - Tree Diagram

5.2.6 Diagrama de Venn

Este diagrama exibe visualmente todas as relações lógicas possíveis entre uma coleção de um conjunto, normalmente representado em círculos. Cada conjunto é uma coleção de dados que tem algo em comum. Quando múltiplos círculos se sobrepõem acontece a interseção, o local onde os dados que têm qualidades em comum se misturam.

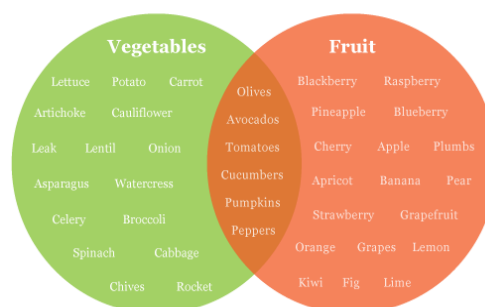


Gráfico 26 - Venn Diagram

5.2.7 Gráfico da Linha do Tempo

Esta representação exibe uma lista de eventos em ordem cronológica. A sua principal função é comunicar informação relacionada ao longo do tempo, para análise ou apresentação visual de acontecimentos.

Se for construído em escala permite observar eventos distribuídos ao longo do tempo, coisas que estão a acontecer ou vão acontecer. Utilizado para perceber como dados categóricos podem sofrer alterações ao longo do tempo.

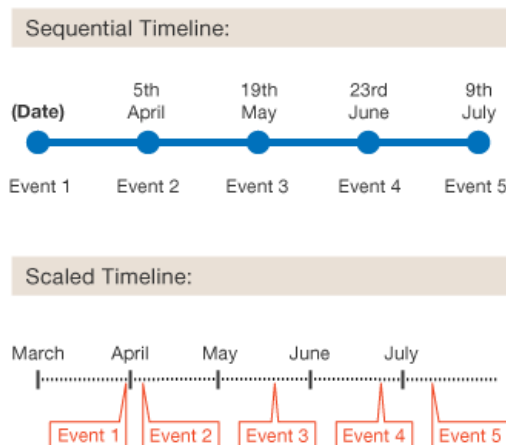


Gráfico 27 - Timeline

5.3 Tabelas

A classe de representações designada de tabelas, funcionam quando se pretende analisar ou comunicar uma informação precisa e fácil de referenciar.

5.3.1 Gráfico de Contagem

Esta representação exibe a frequência e a distribuição dos dados, utilizando o sistema numérico de marcas. Categorias, valores, ou intervalos são colocados num eixo ou coluna (normalmente eixo vertical ou primeira coluna da esquerda), e cada vez que um valor aparece, uma marca de registo acontece. Quando todos os dados estão representados, exibe-se o total de ocorrências. No final o resultado assemelha-se a um histograma.

Title	Tally	Totals
Categories/ Values or Intervals		38
		20
		37
		35
		33
		15
		30
		20
		34
		25

Gráfico 28 - Tally Chart

5.3.2 Tabela de Tempo

Estas tabelas são utilizadas como ferramenta de gestão para referenciar eventos agendados, tarefas ou ações.

Organizar dados numa tabela com uma ordem cronológica ou em ordem alfabética permite aos utilizadores uma consulta rápida. Por este motivo são utilizadas para exibir datas e horários de partidas e chegadas de transportes públicos.

▼ Times						
Category A	08:35	08:48	09:03	09:15	09:50	10:23
Category B	09:27	10:32	11:11	12:00	14:10	15:59
Category C	07:13	11:03	13:47	16:10	18:00	22:22

	Mon	Tues	Wed	Thru	Fri	Sat	Sun
07:00	Event A		Event B		Event C	Event D	Event E
08:00		Event F		Event G	Event H		
09:00			Event I				Event J
10:00	Event K			Event L	Event M	Event N	
11:00			Event O		Event P		

Gráfico 29 – Time Table

5.4 Charts

A classe de representações designadas de charts são utilizadas como referência a um gráfico de informação. Chart é uma representação que proporciona um meio visual para consolidar e mostrar dados, com o objetivo de analisar, planejar, controlar e comunicar.

5.4.1 Círculo de Pacotes

São uma variante de mapas de árvores que utiliza círculos em vez de retângulos. Cada círculo representa uma hierarquia: cada ramo de árvore é representado como um círculo e os seus sub-ramos são representados por sub-círculos.

A área de cada círculo é utilizada para representar valor adicionais, como por exemplo a quantidade. A cor também pode ser utilizada para exibir diferentes categorias ou variáveis.

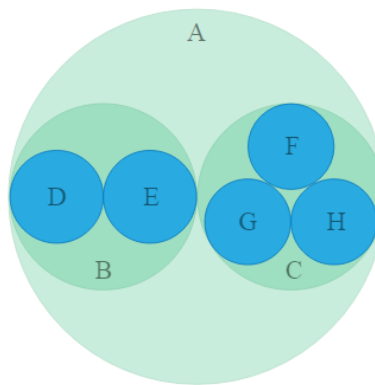


Gráfico 30 - Circle Packing

5.4.2 Gráfico de Donuts

Este gráfico apresenta-se como um gráfico de tarte mas com a área do centro cortada, ocupando menos espaço que o outro. A vantagem é que proporciona ao utilizador uma leitura mais fácil na distinção do tamanho dos elementos por meio de um arco.

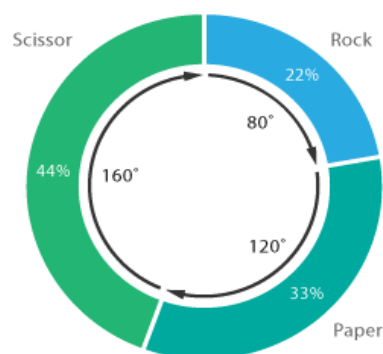


Gráfico 31 - Donut Chart

5.4.3 Matriz de Pontos

Esta representação exibe dados através de unidades de pontos coloridos, onde cada um representa uma categoria em particular. São úteis para alcançar uma visão geral da proporção que cada categoria assume num conjunto de dados. Pontos com a mesma cor são exibidos em conjunto.

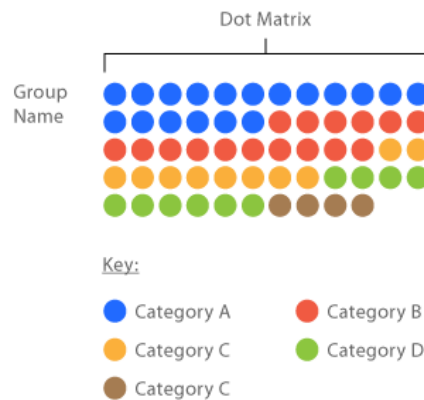


Gráfico 32 - Dot Matrix Chart

5.4.4 Gráfico em Caracol

Esta representação foi utilizada pela primeira vez pelo médico Florence Nightingale para comunicar as mortes que se podiam ter evitado na Guerra da Crimeia.

Os dados são exibidos através de um sistema de coordenadas polares. Cada categoria ou intervalo de dados é dividido num segmento igual, onde cada segmento se alarga a partir do centro em proporção ao valor que representa. Cada anel parte do centro da grade polar e serve como escala para traçar o tamanho do segmento que representa o valor mais elevado. A área representa o valor da variável.

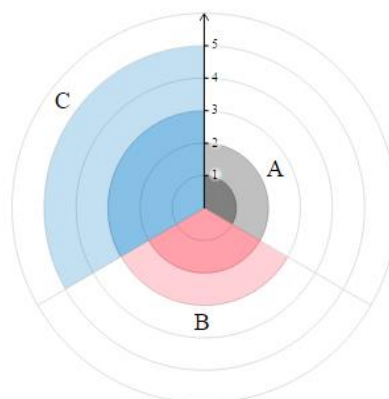


Gráfico 33 - Nightingale Rose Chart

5.4.5 Gráfico de Conjuntos Paralelos

Estes gráficos exibem um caminho de fluxos e proporções. O caminho do fluxo é representado por conjuntos de linhas, onde cada conjunto corresponde a uma dimensão e os seus valores/categorias são representados em cada divisão do conjunto de linhas. A largura da linha é o trajeto determinado pelo valor proporcional do total da categoria. Cada caminho de fluxo deve ser colorido para comparar a distribuição entre as diferentes categorias.

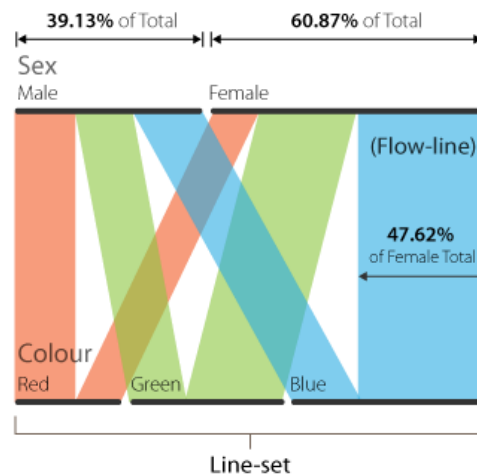


Gráfico 34 - Parallel Sets

5.4.6 Gráfico de Imagens

Estes gráficos utilizam ícones para revelar uma visão mais profunda na exibição de pequenos conjuntos de dados discretos. Normalmente um ícone representa dados unitários ou com muitas unidades. Exemplo: dados populacionais representam-se por ícones de humanos.

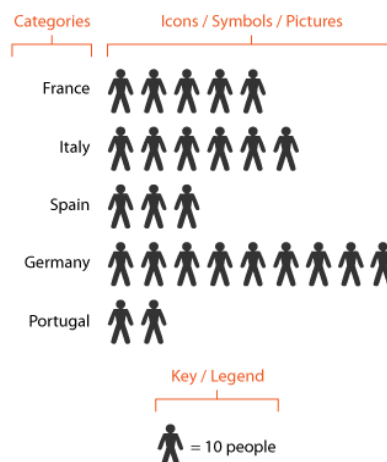


Gráfico 35 - Pictogram Chart

5.4.7 Gráfico de Tarte

Estes gráficos são utilizados em apresentações porque ajudam a comunicar proporções e percentagens entre categorias, dividem um círculo em segmentos proporcionais. Cada fatia representa uma proporção e o círculo completo representa a soma total de dados.

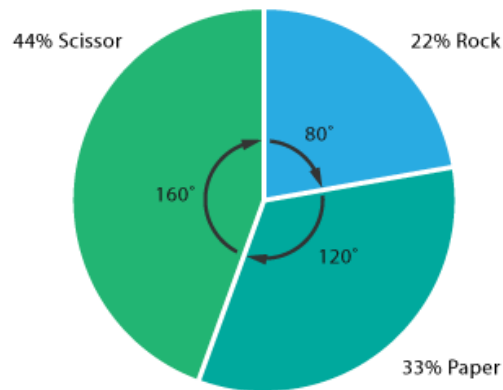


Gráfico 36 - Pie Chart

5.4.8 Gráfico de Área Proporcional

Estes gráficos são ideais para comparar valores e exibir proporções (tamanho e quantidade). Dão uma visão geral e rápida do tamanho de dados, sem o uso de escalas. Uma desvantagem é dificuldade em estimar valores através da área, o que significa que são mais utilizados para a comunicação do que para análise.

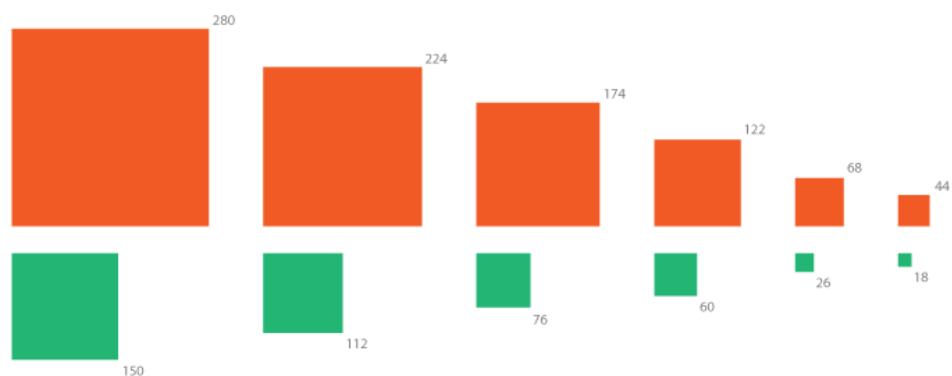


Gráfico 37 - Proportional Area Chart

5.4.9 Mapas de Árvores

Shneiderman (1996) desenvolveu originalmente mapas de árvores como forma de visualizar um vasto arquivo de diretórios num computador, sem ocupar muito espaço

Estes mapas são uma alternativa aos diagramas de árvores. Para visualizar estruturas de hierarquias enquanto se conhece as quantidades de cada categoria, através do tamanho da área. Cada categoria é associada à área de um retângulo com subcategorias em retângulos sobrepostos ao principal. Quando a quantidade é atribuída a uma categoria, o tamanho de área é exibido em proporção a essa quantidade e às outras quantidades da mesma categoria.



5.4.10 Nuvem de palavras

São geralmente utilizadas em *website's* ou *blogs*. Usam-se para comparar discursos.



5.5 Mapas

A classe de representações designadas de mapas funcionam bem quando se pretende comunicar dados com uma localização ou relação espacial.

5.5.1 Mapa de Bolhas

Estes mapas representam vários círculos, ao longo de uma região geográfica, com a área do círculo proporcional ao conjunto de dados. São ideais para comparar proporções em regiões geográficas.

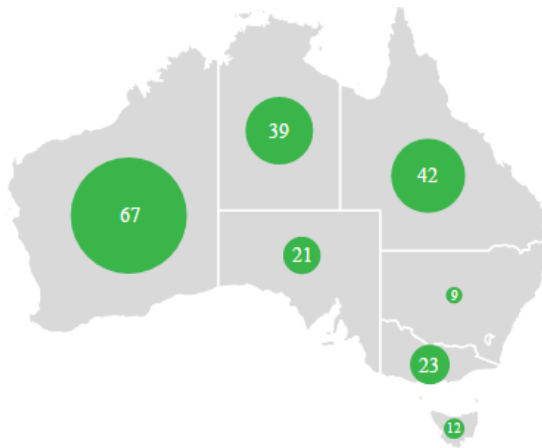


Gráfico 40 - Bubble Map

5.5.2 Mapa de Cores

Esta representação exibe áreas geográficas divididas em regiões, que mantêm uma relação com uma variável de dados. Através das cores, de áreas sombreadas ou padronizadas. A visualização de valores por uma área geográfica revela variações ou padrões em todo o local exibido.

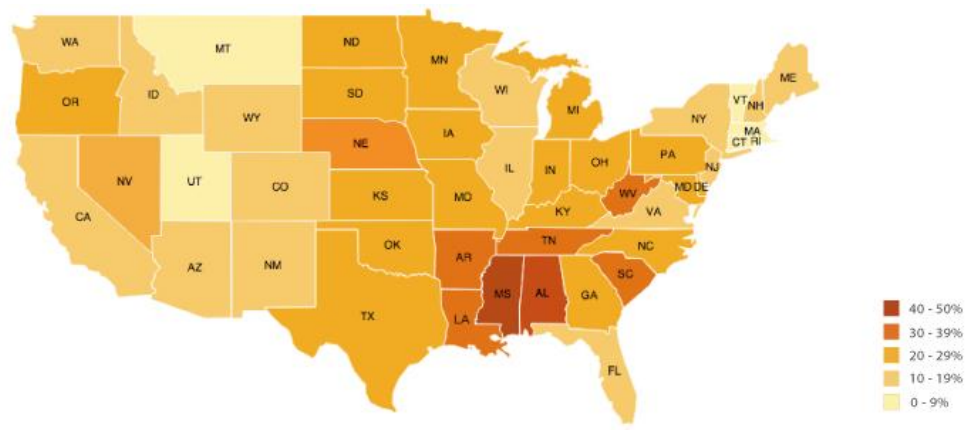


Gráfico 41 - Choropleth Map

5.5.3 Mapa de Fluxo

Os mapas de fluxo geográfico acompanham o movimento da informação ou objetos desde uma localização para outra. Normalmente utilizados para exibir dados de população migratória, animais ou produtos. A quantidade é representada no tamanho de uma linha ou seta (que indica o movimento). Estas representações permitem perceber dados de distribuição geográfica.

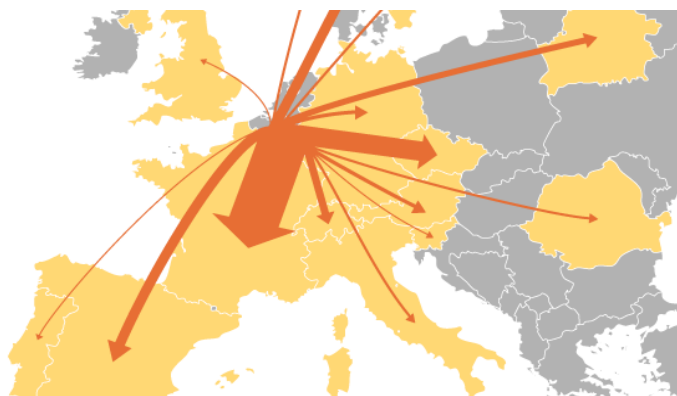


Gráfico 42 - Flow Map

6. CLASSIFICAÇÃO

As principais preocupações para a construção desta classificação foram: como evidenciar as características dos dados numa representação visual, e ainda tornar a escolha do tipo de gráfico mais rápida, por parte de quem utiliza as técnicas de VI.

Primeiramente, é importante que o processo de Visualização de inicie com um propósito e que o(s) conjunto(s) de dados assumam as suas posições na representação visual.

Se existe um propósito para produzir uma Visualização é conveniente pensar nas etapas a percorrer para atingir esse propósito - desde qual a forma mais eficiente para evidenciar as características dos dados até ao tipo de gráfico mais adequado para visualizar essas características.

Vários *frameworks* (*"Types of visualization insights"*, *"Key visualization types"*, *"Visualization framework designed to ease the selection and design of data visualizations"*) têm sido desenvolvidos com o objetivo de facilitar o processo de converter dados em imagens - um complexo processo que envolve a cognição e percepção humana, *data mining*, algoritmos de visualização e IHC. Assim, cinco classes de representações visuais foram definidas, servindo como base para o mapeamento visual:

Tabelas – Linhas ou colunas dispostas por uma grelha geométrica que contêm dados ou imagens.

Gráficos – Traçar variáveis quantitativas e qualitativas como se fossem coordenadas, num sistema de referencia bem definido, em eixo vertical ou horizontal.

Charts – Traçar dados quantitativos e qualitativos sem o uso de um sistema de referência.

Mapas – Exibe dados associados a um espaço físico.

Diagramas – Utiliza nós para representar registos de dados e ligações entre nós, que representam relações de sequências dentro do conjunto de dados.

As etapas do processo de visualização são um conjunto de decisões que o utilizador deve tomar antes do resultado final. Perceber qual o tipo de dados que o conjunto contém e defini-los: dados quantitativos ou categórico; conhecer os tipos de gráficos e seleccioná-los; saber quais as análises que pode adotar para evidenciar as características dos dados; dominar a utilização de propriedades gráficas e construir uma história, transporta o utilizador para o domínio de representar dados visualmente.

Antes de iniciar o processo de visualização o utilizador deve questionar-se cientificamente, caso contrário não surge o propósito da visualização. Responder a questões temporais como "quando", espaciais como "onde", tópicos como "o quê" ou relacionais como "com quem", permite que o utilizador direcione e justifique as suas escolhas na construção de uma visualização.

Perante o estudo das classes de representações gráficas apresentadas no capítulo anterior e de acordo com os métodos analíticos descritos na secção 5.4, apresentam-se de seguida doze tipos de análises *standard*, que permitem agrupar as representações gráficas:

- Comparação - Análise que revela semelhanças ou diferenças entre os valores de dados;
- Proporção - Análise que utiliza o tamanho para exibir semelhanças e diferenças entre os valores de dados num todo;
- Relações - A relação entre atributos pode ser encontrada comparando atributos do eixo vertical com o horizontal;
- Hierarquia - Análise que exibe uma classificação ordenada de dados em estruturas;
- Conceitos - Análise que ajuda a simplificar ideias e conceitos;
- Localização - Análise que exibe dados ao longo de uma região geográfica;
- Parte de um todo - Análise que exibe uma parte ou partes de uma variável num total;
- Distribuição - Análise que exibe a frequência de dados e como eles são espalhados ou agrupados ao longo de um intervalo;
- Movimento - Análise que ajuda a exibir movimentos e mudanças nos dados;
- Padrões - Análise que exibe formas e padrões para revelar significado aos dados;
- Ao longo do tempo - Análise que exibe dados durante um período de tempo;
- Análise de Texto - Análise que exibe palavras de um texto visualmente.

Na medida em que a VI deve ser utilizada num contexto de exploração e comunicação de dados de forma organizada em que se pretende revelar novas informações ou prever novos comportamentos surgem os métodos analíticos descritos.

O principal motivo de representar dados visualmente é evidenciar e revelar as suas relações, as suas características, diferenças e semelhanças. Poder compara-los ou distribui-los ao longo de uma variável e ainda organizar os dados em hierarquia é possível e existem gráficos ideais para revelar estes comportamentos.

Os doze métodos analíticos foram definidos aquando a descrição dos gráficos no capítulo anterior. Doze métodos básicos e simples que se aproximam à escolha de uma representação gráfica ideal para realçar certos tipos de análises.

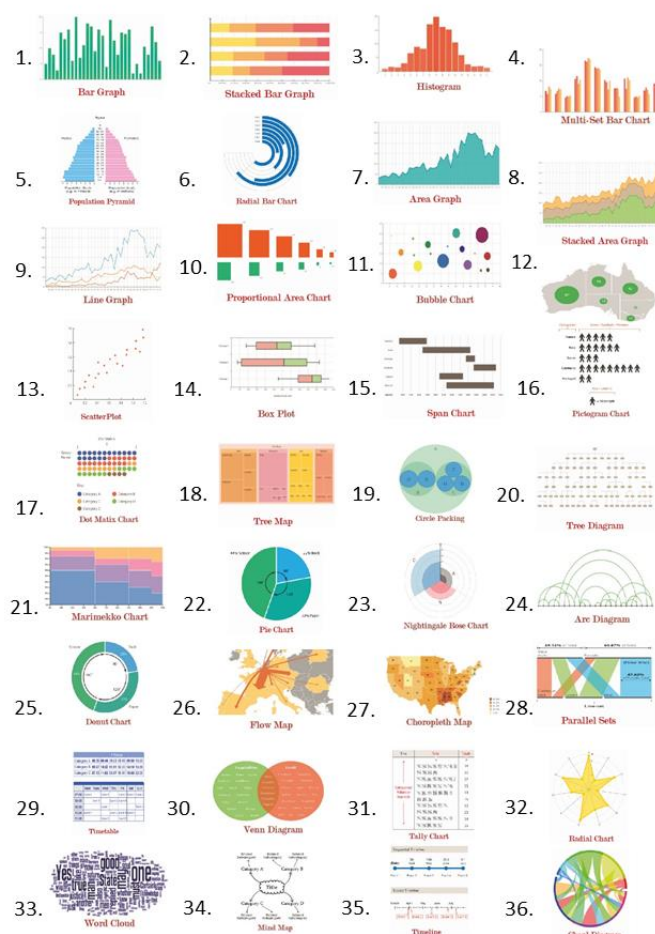
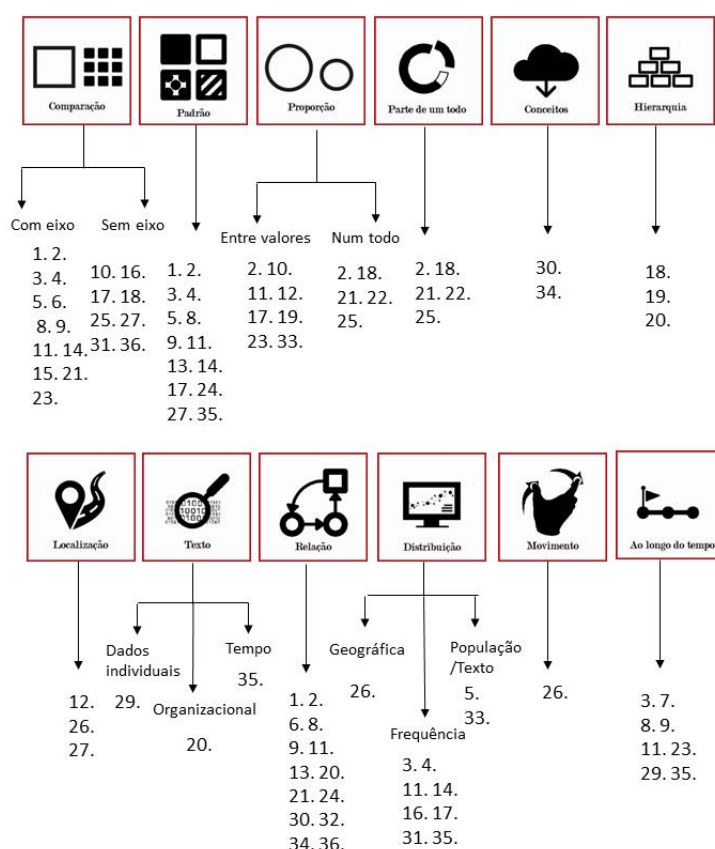
Com o objetivo de aproximar dois grandes grupos da VI: a análise de dados e representações gráficas foi construída a Classificação - “Representações Visuais – o que pretendo transmitir?” (ver figura 9). O evidência a relação entre os métodos analíticos e os vários tipos de gráficos.

Na Classificação, o grupo à esquerda da imagem representa os doze métodos analíticos, que de forma básica possibilitam escolher representações gráficas. Isto é, a escolha adequada de uma representação gráfica está intimamente ligada à análise ou comportamento que o utilizador pretende revelar na sua visualização.

Exibir uma comparação, um padrão, uma proporção, uma parte de um todo, um conceito, uma hierarquia, uma localização, um texto, uma relação, uma distribuição, um movimento ou um evento ao longo do tempo são critérios escolhidos para se iniciar o processo de visualização e todos os de métodos analíticos são capaz de atribuir significado ao conjunto diverso de dados, na medida em que são adequados às suas reais características e necessidades de exploração.

Representação Visual: o que pretendo transmitir?

Métodos Analíticos



Legenda: Tipos de Gráfico

Figura 9 – Classificação: "Representação Visual"

Durante o processo de Visualização após a fase em que o utilizador percebe qual o propósito da análise e o método analítico que melhor se relaciona com as principais características dos seus dados, é necessário perceber qual a utilização gráfica mais eficiente.

O grupo à direita da imagem representa um conjunto de trinta e seis diferentes tipos de gráficos, com a respetiva legenda [1 a 36]. Um total de trinta e seis gráficos que contemplam desde os três sistemas

de coordenadas: cartesianas, polares e geográficas, às mais variadas marcas gráficas: linhas pontos e áreas, com diversas variáveis visuais.

Através do esquema é visível que será mais fácil alcançar uma comparação correta de um conjunto de dados através da utilização de um *bar graph* do que um *pie chart*, isto é possível devido às características específicas deste dois tipos de gráfico.

A representação final destes dois grandes grupos da VI permite perceber a proximidade com que um gráfico deve ser escolhido consoante a análise pretendida, porque desta forma são reveladas com mais pormenor as variáveis de dados e os seus comportamentos. Desta maneira o processo de visualização será finalizado com sucesso.

Posto isto, associado ao conjunto de métodos analíticos e representações gráficas surge a oportunidade de contextualizar a exploração ou representação através e um conjunto de acontecimentos que desencadeiam o entendimento da representação visual final. A utilização das propriedades gráficas ao longo de um trabalho de visualização irá evidenciar comportamentos, padrões e tendências nos dados que após serem estudado e contextualizados devem ser comunicados através de uma narrativa visual (Segel & Heer, 2010).

7. CONCLUSÃO

Este trabalho de dissertação tem como objetivo principal uma classificação detalhada de gráficos que explique os seus tipos, as suas categorias e as suas características, realçando qual a utilização mais adequada para cada gráfico. Desta forma, além da descrição detalhada dos gráficos procedeu-se à construção de um esquema apelidado de Representações Visuais, para facilitar a compreensão da tarefa entre a análise/relação pretendida e a escolha do gráfico ideal nesse contexto.

Para tal, o projeto iniciou-se com a concretização do enquadramento teórico do tema através do levantamento dos primeiros gráficos, que se traduz num marco importantíssimo para o estudo da área. Posteriormente, realizou-se um levantamento de questões que impulsionam o estudo da Visualização de Informação em contexto de Tecnologias de Informação e aferiu-se que o *Big Data* é a locomotiva dos dados que produzem informação. De seguida, procedeu-se ao levantamento de toda a bibliografia relevante para a contextualização da Visualização de Informação de forma a perceber a vantagem em visualizar e transformar dados abstratos em visualizações. Assim foi realizado um levantamento de todas as temáticas envolvidas no tema central, no sentido de melhor dar a conhecer os conceitos tratados. Desta forma aferiu-se que o utilizador é um agente crucial neste processo e é sobre ele que recaem todas as decisões que objetivam auxiliar a análise e compreensão de um conjunto diverso de dados.

Por conseguinte, uma pequena alteração no modelo de visualização seria indispensável devido ao grande e variado conjunto de dados, neste sentido a implementação de técnicas interativas de filtragem durante todo o processo de visualização é indispensável para uma seleção coerente de dados. Procedeu-se à síntese de características essenciais ao mapeamento gráfico, indispensáveis à adaptativa natureza dos dados. Foi aferido que a escolha de propriedades gráficas irá influenciar todo o aspeto e conteúdo da representação assim como as conexões e contextos entre os factos dos dados. Desta maneira, os gráficos foram descritos tendo a julgar pelo seu conteúdo e pela capacidade em transmitir um tipo de informação ou relação. A construção de um artefacto fiável para a comunidade científica baseou-se na metodologia *Design Science Research*, embora algumas das suas diretrizes não tenham sido executadas devido à falta de aplicação da área no contexto das TI, contudo espera-se aumentar o estudo e desenvolvimento da área, uma vez que esta enriquece toda a informação explorada e transmitida por um Sistema de Informação.

Resumidamente, a Visualização de Informação impulsiona a capacidade humana para absorver informação contida num conjunto de dados. A combinação do *big data* com ferramentas analíticas e com a visualização de dados confere uma vantagem única e preciosa para as organizações, uma vez que estas lidam com um cenário de dados não estruturado e de fontes externas à organização. A necessidade de análises imediatas ao voluptuoso e rápido conjunto de dados, a capacidade de explorar

dados e conseguir transforma-los em informação útil devem ser simplificados pela agilidade em apresentar os dados em representações gráficas. O método de visualização deve ser adotado pelos vários utilizadores e pelos diversos departamentos de uma organização, uma vez que permite uma interação quase real com os dados.

A continuação do estudo da visualização que exprima novas características de relações nos dados e representações gráficas será pertinente para a consolidação da área. Uma sugestão de trabalhos futuros seria a inclusão de casos de estudo com acesso a fontes de informação para a exploração, representação e apresentação sistemática de resultados.

8. REFERÊNCIAS BIBLIOGRÁFICAS

- Akerkar, R. (2013). *Big Data Computing. Chapman and Hall/CRC*. Retrieved from <http://books.google.com/books?hl=en&lr=&id=mXr6AQAAQBAJ&pgis=1>
- Bellinger, G., Castro, D., & Mills, A. (2004). Data, information, knowledge, and wisdom. URL: <Http://www.Systems-....> Retrieved from <http://geoffreyanderson.net/capstone/export/37/trunk/research/ackoffDiscussion.pdf>
- Bertin, J. (1983). *Semiology of Graphics. University of Wisconsin Press*. <http://doi.org/10.1007/s00371-013-0892-3>
- Bertin Matrices. (2014). Retrieved October 19, 2015, from <http://www.statlab.uni-heidelberg.de/projects/bertin/>
- Card, & Mackinlay. (1991). The information visualizer, an information workspace. In *Proceedings of the SIGCHI conference on Human factors in computing systems Reaching through technology - CHI '91* (pp. 181–186). New York, New York, USA: ACM Press. <http://doi.org/10.1145/108844.108874>
- Card, & Mackinlay. (1999). *Information visualization: using vision to think*.
- Card, S. K., & Mackinlay, J. (1997). The Structure of the Information Visualization Design Space. *Xerox PARC*, 92–99. <http://doi.org/10.1109/INFVIS.1997.636792>
- Chen, C. (2010). Information visualization. *John Wi Ley & Sons, Inc*, 2, 387–403. <http://doi.org/10.1002/wics.89>
- Few, S. (2005). Quantitative vs. Categorical Data: A Difference Worth Knowing, (April), 1–5.
- Friendly, M. (2006). A Brief History of Data Visualization. *Handbook of Computational Statistics: Data Visualization*, 1–46. http://doi.org/10.1007/978-3-540-33037-0_2
- Gartner Inc. (2015). Technology Research | Gartner Inc. Retrieved October 20, 2015, from <http://www.gartner.com/technology/home.jsp>
- Keim, D. (2002). Information visualization and visual data mining. *IEEE Transactions on Visualization and Computer Graphics*, 8(1), 1–8. <http://doi.org/10.1109/2945.981847>
- Keim, D., Andrienko, G., Fekete, J., & Carsten, G. (2008). Visual Analytics : Definition , Process , and Challenges, 154–175.
- Mackinlay, J. D., & Winslow, K. (2015). *Designing Great Visualizations*.

- Merritte Stidston. (2014). Business Leaders Need R's not V's: The 5 R's of Big Data | MapR. Retrieved October 19, 2015, from <https://www.mapr.com/blog/business-leaders-need-r%E2%80%99s-not-v%E2%80%99s-5-r%E2%80%99s-big-data>
- Olshannikova, E., Ometov, A., Koucheryavy, Y., & Olsson, T. (2015). Visualizing Big Data with augmented and virtual reality: challenges and research agenda. *Journal of Big Data*, 2(1), 22. <http://doi.org/10.1186/s40537-015-0031-2>
- Pillat, R. M., Valiati, E. R. A., & Freitas, C. M. D. S. (2005). Experimental study on evaluation of multidimensional information visualization techniques. *CLHC '05: Proceedings of the 2005 Latin American Conference on Human-Computer Interaction*, 20–30. <http://doi.org/10.1145/1111360.1111363>
- Schreiner, D. E. (1997). *Information graphics: A comprehensive illustrated reference. Technical Communication* (Vol. 44). [http://doi.org/10.1002/\(SICI\)1097-4571\(19980401\)49:4<383::AID-ASI11>3.0.CO;2-V](http://doi.org/10.1002/(SICI)1097-4571(19980401)49:4<383::AID-ASI11>3.0.CO;2-V)
- Segel, E., & Heer, J. (2010). Narrative visualization: telling stories with data. *IEEE Transactions on Visualization and Computer Graphics*, 16(6), 1139–48. <http://doi.org/10.1109/TVCG.2010.179>
- Shneiderman, B. (1996). The eyes have it: a task by data type taxonomy for information\nvisualizations. *Proceedings 1996 IEEE Symposium on Visual Languages*, 336–343. <http://doi.org/10.1109/VL.1996.545307>
- Spence, R. (2007). *Information Visualization: Design for Interaction. Proceedings of CHI 2005 Conference on Human Factors in Computing Systems* (Vol. 7). <http://doi.org/10.1075/idj.17.3.15ehr>
- Steele, J., & Iliinsky, N. (2010). *Beautiful Visualization. Beautiful Visualization*. Retrieved from <http://www.amazon.fr/Beautiful-Visualization-Julie-Steele/dp/1449379869>
- Stevens, S. S. (1946). On the theory of scales of measurement. *Science (New York, N.Y.)*, 103(2684), 677–680. <http://doi.org/10.1126/science.103.2684.677>
- Stodder, D. (2013). Data Visualization and Discovery for Better Business Decisions. *TDWI Best Practices Report, Third Quarter*, 1(July), 36.
- Tufte, E. R. (2001). The Visual Display of Quantitative Information. *Technometrics*. <http://doi.org/10.1198/tech.2002.s78>
- Ware, C. (2004). *Information visualization: perception for design. Information Visualization*. <http://doi.org/10.1016/B978-0-12-381464-7.00018-1>

