

UNIVERSIDADE ABERTA



**Visualização imersiva de *big data* com
dispositivos móveis**

Tiago Martins Peres

Mestrado em Tecnologias e Sistemas Informáticos Web

2020

UNIVERSIDADE ABERTA



UNIVERSIDADE
AbERTA
www.uab.pt

**Visualização imersiva de *big data* com
dispositivos móveis**

Tiago Martins Peres

Mestrado em Tecnologias e Sistemas Informáticos Web

Dissertação orientada pela Professora Doutora Elizabeth Simão
Carvalho

2020

Resumo

A grande quantidade de dados armazenados atualmente não tem utilidade se não conseguirmos entender as relações e dependências entre os dados e construir modelos significativos. Para identificar novas tendências e padrões mais facilmente faz sentido que os mesmos sejam limpos, processados e apresentados de maneira amigável. Os ambientes imersivos virtuais oferecem grandes oportunidades em diversas áreas, mas existem poucos estudos sobre a sua utilidade como ferramenta de visualização de informação.

Este trabalho assumiu que é possível a visualização de dados do tipo *big data* em ambientes imersivos virtuais móveis de forma satisfatória, permitindo ao utilizador estar dentro da representação visual dos dados. Assim, teve como objetivo propor um modelo visual (designado de *Data-Room*) para verificar a eficácia relativa da visualização de dados do tipo *big data* com a aplicação de técnicas de visualização de dados 3D num ambiente imersivo virtual móvel, comparando com outros ambientes de visualização. O trabalho avaliou o modelo proposto, tendo sido desenvolvido e posteriormente testado para esse efeito o protótipo do *Data-Room*. O modelo foi avaliado em termos de aquisição e apreensão de informação de forma a ser possível identificar as potenciais mais-valias introduzidas. Foram utilizadas seis codificações visuais distintas.

Os resultados indicaram que os ambientes imersivos virtuais móveis são um método válido e útil para a visualização de informação de dados do tipo *big data*. Este ambiente demonstrou ser potencialmente tão ou mais eficaz para visualizar dados desse tipo, comparativamente com outros métodos utilizados na atualidade para o mesmo fim.

Espera-se com este trabalho ter contribuído para um maior conhecimento do valor da utilização de ambientes imersivos virtuais para dispositivos móveis na visualização de dados do tipo *big data* e demonstrar que estes são significativamente mais agradáveis, fáceis de resumir a informação e tão ou mais eficientes quanto a visualização desse tipo de dados em outros ambientes.

Palavras-chave: Big Data, Data Analysis, Immersive Environments, Information Visualization (InfoVis), Smartphones.

Abstract

The vast amount of data stored today is of no use if we cannot understand the relationships and dependencies between the data and build meaningful models. To identify new trends and patterns more easily, it makes sense that they are cleaned, processed and presented in a friendly manner. Virtual immersive environments offer great opportunities in many areas, but there are few studies on their usefulness as an information visualization tool.

This work has assumed that it is possible to view *big data* in mobile virtual immersive environments satisfactorily, allowing users to be inside of the data visual representation. Thus, it aimed to propose a visual model (called *Data-Room*) to verify the relative effectiveness of data visualization of *big data* with the application of 3D data visualization techniques in a mobile virtual immersive environment, comparing with other environments. visualization. The work evaluated the proposed model, having been developed and subsequently tested for this purpose the *Data-Room* prototype. The model was evaluated in terms of acquisition and seizure of information in order to be able to identify the potential added value. Six distinct visual codings were used.

The results indicated that mobile virtual immersive environments are a valid and effective method for visualizing big data. This environment has been shown to be potentially as or more effective for visualizing such data compared to other methods used today for the same goal.

It is hoped that this work has contributed to a greater understanding of the value of using mobile virtual immersive environments to view big data and demonstrate that it is significantly nicer, easier to summarize, and as effective viewing this type of data in other environments.

Keywords: Big Data, Data Analysis, Immersive Environments, Information Visualization (InfoVis), Smartphones.

Quando me lancei no desafio de fazer o Mestrado em Tecnologias e Sistemas de Informação Web, procurei concretizar o desejo de continuar a minha formação académica, seguir a minha área de vocação e seguir um caminho mais direcionado de educação dentro das minhas necessidades.

Informática faz parte da minha vida desde tenra idade, o meu pai é um professor de IT, sangue *geek* corre-me nas veias. Além disso, fui cercado por livros e cultura, cresci com a magia e fantasia dos livros de J. K. Rowling, Harry Potter; por outro lado, tenho uma profunda paixão por livros técnicos que me continuam a ajudar diariamente, inclusive nos dias de hoje. O melhor que aprendi em ambos os mundos foi perceber que tudo é possível, que é tudo uma questão de tempo e muito trabalho, que os objetivos demoram para ser alcançados e que, sem esse trabalho moroso e difícil, são simplesmente impossíveis de alcançar. No momento atual sou um *developer* apaixonado por tecnologias de informação e *design*. Adoro programar e produzir soluções globais que realmente ajudam os outros. Estudei e obtive certificações em 10 universidades diferentes espalhadas pelo mundo, trabalhei em projetos para clientes em 4 continentes diferentes (20 países), fiz mais de 150 cursos online, trabalhei com equipas multi-funcionais e tive de lidar com diferentes fusos horários e vários idiomas. Nos últimos 5 anos, trabalhei em vários projetos e a vários níveis, de projetos iniciantes a sistemas de informação mais complexos e contribuí também para as minhas instituições de caridade favoritas.

O presente trabalho resulta de muitas horas dedicadas a um tema que me apaixonou, mas não o teria conseguido sem o apoio de muitas pessoas à minha volta. Não sendo possível de particularizar todos, não gostaria, no entanto, de deixar de agradecer, especialmente...

.... a Deus por nunca me deixar nem abandonar (Deuteronomio 31:6).
.... à sra. Professora Elizabeth Carvalho pela orientação: reuniões constantes, dúvidas frequentes e todo um esforço e exigência.
.... à minha família (Edmaro Peres, Nídia Peres, Gonçalo Peres, Bárbara Peres, Graciete Peres, Esmeralda Martins, Mónica Bem, ...) por todo o suporte e carinho.
.... ao sr. Professor Leonel Morgado pela coordenação do mestrado, pelo *feedback* e por me ter permitido apresentar este trabalho nas VIII Jornadas InforAberta em Ansião.
.... à secretária do Elisa Antunes pelo esclarecimento de várias dúvidas.
.... aos cientistas e profissionais pela disponibilidade demonstrada e ajuda que me concederam na elaboração deste estudo, sendo de destacar Ciro Donalek (*Virtualitics*) e Amir Ben-Arye (*Graph and Chart*).
.... à rede *Immersive Learning Research Network* (iLRN), nomeadamente Jonathon Richter, Erica Southgate e Heather Dodds pelo apoio e interesse neste trabalho.
.... aos participantes dos inquéritos.

A todos o meu muito obrigado"! ☺

Índice Geral

1. Introdução	1
1.1. Motivações, objetivos e contribuições fundamentais	2
1.2. Metodologia de investigação	5
1.3. Organização da dissertação	6
2. Estado da arte	7
2.1. Fundamentos de big data	9
2.1.1. Dimensões e escalabilidade	10
2.1.2. Integração e processamento	10
2.2. Visualização de informação	13
2.2.1. Técnicas virtuais imersivas	17
2.3. Visualização imersiva em dispositivos móveis	21
2.3.1. Aplicações	21
2.3.2. HMDs	29
2.3.3. Ambientes de desenvolvimento e respetivas bibliotecas	32
2.4. Visualização de big data	38
2.4.1. Visualização de big data em ambientes imersivos	40
2.5. Síntese	42
3. Modelo conceptual	45
3.1. Visão geral	47
3.2. Requisitos do modelo	50
3.3. Mapeamento visual e interação	51
3.4. Síntese	57
4. Implementação e validação do modelo	59
4.1. Implementação	61
4.1.1. Dados utilizados	61
4.1.2. Data-Room	62
4.1.3. Visualizações com Data-Room, Python e R	65
4.2. Validação	76
4.2.1. Teste do modelo	76
4.2.2. Análise de resultados	77
5. Conclusão	87
5.1. Trabalho futuro	88
6. Referências bibliográficas	89
7. Anexos	100

Índice de Tabelas

Tabela 2-1. Comparação das aplicações (realizada dia 16 de novembro de 2019).....	28
Tabela 2-2. Comparação dos HMDs (realizada dia 8 de outubro de 2019).	32
Tabela 2-3. Combinação entre HMD, SDKs de realidade virtual e pacote Unity para soluções em telemóveis.	33
Tabela 2-4. Comparação das bibliotecas para o Unity (realizada dia 25 de abril de 2019).....	34
Tabela 2-5. Comparação das bibliotecas para o Android Studio (realizada dia 25 de abril de 2019).	35
Tabela 2-6. Comparação dos ambientes de desenvolvimento (realizada dia 25 de abril de 2019).	36
Tabela 2-7. Etapas inerentes ao processo de visualização [26, 27].....	38
Tabela 4-1. Características visuais e interativas nos diferentes ambientes.	66
Tabela 4-2. Estatística descritiva do conjunto de dados.	66
Tabela 4-3. Estatística descritiva da amostra usada para a criação das visualizações nas Figuras 4-4, 4-5 e 4-6.	67
Tabela 4-4. Estatística descritiva do conjunto de dados.	68
Tabela 4-5. Estatística descritiva da amostra usada para a criação das visualizações nas Figuras 4-7, 4-8 e 4-9.	68
Tabela 4-6. Estatística descritiva do conjunto de dados usado para a criação das visualizações nas Figuras 4-10, 4-11 e 4-12.	70
Tabela 4-7. Estatística descritiva do conjunto de dados.	71
Tabela 4-8. Estatística descritiva dos dados usados para a criação da visualização na Figura 4-13, 4-14 e 4-15.	71
Tabela 4-9. Estatística descritiva do conjunto de dados usado para a criação das visualizações nas Figuras 4-16, 4-17 e 4-18.	73
Tabela 4-10. Estatística descritiva do conjunto de dados usado para a criação das visualizações nas Figuras 4-19, 4-20 e 4-21.	74
Tabela 4-11. Registo de observações deixadas pelos participantes após utilizar o Data-Room.	79
Tabela 4-12. Reacções médias dos participantes (numa escala Likert de 0 a 5 pontos).	80
Tabela 4-13. Comentários ou sugestões dos participantes.	80
Tabela 4-14. Análise de aquisição.....	81
Tabela 4-15. Análise de retenção.	81
Tabela 4-16. Análise de apreciação das pistas visuais.	83

Índice de Figuras

Figura 2-1. Big data: os 5 Vs. _____	9
Figura 2-2. MapReduce aplicado ao processo de contagem de palavras. _____	11
Figura 2-3. Quadrante mágico para soluções de gestão de dados para análise (Ronthal, Edjlali & Greenwald, 2019). _____	12
Figura 2-4. O processo de visualização de informação (Spence, R., 2007). _____	13
Figura 2-5. Estatísticas oficiais revelam o quanto o coronavírus afetou o mercado de trabalho nos EUA (Bui & Wolfers, 2020). _____	14
Figura 2-6. Ranking de pistas visuais (Cleveland & McGill, 1984). _____	15
Figura 2-7. Os quatro níveis de design de visualizações (Munzner, 2015). _____	16
Figura 2-8. Visualização em CAVE2 (Febretti et al., 2013). _____	17
Figura 2-9. Glass Brain: visualização 3D do cérebro (Glass Brain, sem data). _____	18
Figura 2-10. Rede 3D vista dentro da CAVE (Febretti et al., 2013). _____	18
Figura 2-11. Block Mining Visualizer: visualização de vários conjuntos de dados combinados associados ao processo de mineração subterrânea (Aukstakalnis, 2016). _____	19
Figura 2-12. Gráfico de barras da aplicação 3Data. _____	22
Figura 2-13. SeeSignal. _____	22
Figura 2-14. Visualização de dados geospaciais em plataforma não identificada. _____	23
Figura 2-15. Conceito Climate Data. Visualização de dados climatéricos com o HMD Magic Leap. _____	23
Figura 2-16. A “HyperDesk” do DataView VR. _____	23
Figura 2-17. Vista inicial do Virtualitics. _____	24
Figura 2-18. Carregar grandes quantidades de dados no Virtualitics. _____	24
Figura 2-19. Gráfico de linhas da aplicação Virtualitics. _____	25
Figura 2-20. Screenshot do que se pode fazer com VR Ulysses. _____	25
Figura 2-21. Screenshot do que se pode fazer com a solução. _____	26
Figura 2-22. Nanome. _____	26
Figura 2-23. CalcFlow. _____	27
Figura 2-24. Mapa 3D na aplicação Geometric Space. _____	27
Figura 2-25. Visualização criada com Dali. _____	28
Figura 2-26. Google Cardboard. _____	29
Figura 2-27. Cardboard da marca Splaks (sem data). _____	29
Figura 2-28. Gear VR. _____	30
Figura 2-29. Daydream View. _____	30
Figura 2-30. Bridge by Occipital. _____	30
Figura 2-31. Bridge (Polycarbonate) by Occipital. _____	31
Figura 2-32. Merge VR. _____	31
Figura 2-33. Bibliotecas para criação de gráficos 3D com Unity. _____	34
Figura 2-34. Bibliotecas para criação de gráficos 3D com Android Studio. _____	35
Figura 2-35. Visualização de dados feita na hackathon Unreal Big Data da HammerHeadVR. _____	36
Figura 2-36. Criado com a ferramenta VR-VIZ. _____	37
Figura 2-37. Pipeline de visualização de dados adaptado (Mohan, Lukas & Pangilinan, 2019). _____	39
Figura 2-38. ETL (Microsoft – Documentação, 2019). _____	39
Figura 2-39. Rede criada no VOSViewer com dados provenientes do Dimensions. _____	40

Figura 2-40. Uma demonstração de trabalho de um problema gravitacional de 3 corpos, implantado no MICA com Second Life (2020).	41
Figura 2-41. Um exemplo do interface do utilizador na ferramenta iViz.	41
Figura 2-42. Visualização de tweets num modelo 3D do campus do MIT (Moran et al., 2015).	41
Figura 2-43. Visão geral da taxonomia de visualização imersiva em dispositivos móveis.	44
Figura 3-1. Arquitetura do modelo.	47
Figura 3-2. Pipeline de entrada de dados e resultados.	49
Figura 3-3. Organograma com a hierarquia das funcionalidades esperada.	52
Figura 3-4. Esquício do ambiente a implementar.	54
Figura 3-5. Mockup do ambiente a implementar com recurso ao Adobe Illustrator.	54
Figura 3-6. Mockup do ambiente a implementar com recurso ao Cinema4D.	55
Figura 3-7. Ambiente a implementar visto através da aplicação Zappar após fazer scan do marcador acima - Realidade Aumentada (RA).	56
Figura 4-1. Arquitetura da prova de conceito com escolhas tecnológicas.	63
Figura 4-2. Tipos de amostras.	64
Figura 4-3. Captura de ecrã da solução Data-Room.	65
Figura 4-4. Python Mayavi barchart() (x= anos, y= meses, z= temperatura em °C).	67
Figura 4-5. R latticeExtra (x= anos, y= meses, z= temperatura em °C).	67
Figura 4-6. Data-Room gráfico de barras (x= anos, y= meses, z= temperatura em °C).	68
Figura 4-7. Python Mayavi barchart() (x= anos, y= meses, z= temperatura em °C).	69
Figura 4-8. R latticeExtra (x= anos, y= meses, z= temperatura em °C).	69
Figura 4-9. Data-Room gráfico de barras (x= anos, y= meses, temperatura em °C).	69
Figura 4-10. Python Mayavi barchart() (x= anos, y= meses, z= temperatura em °C).	70
Figura 4-11. R latticeExtra (x= anos, y= meses, z= temperatura em °C).	70
Figura 4-12. Data-Room gráfico de barras (x= anos, y= meses, z= temperatura em °C).	71
Figura 4-13. Python Mayavi points3D() (x= anos, y= dias, z= temperatura em °C).	72
Figura 4-14. R plot3D (x= anos, y= dias, z= temperatura em °C).	72
Figura 4-15. Data-Room gráfico de bolhas (x= anos, y= dias, z= temperatura em °C).	72
Figura 4-16. Python Mayavi points3D() (x= anos, y= dias, z= temperatura em °C).	73
Figura 4-17. R plot3D (x= anos, y= dias, z= temperatura em °C).	74
Figura 4-18. Data-Room gráfico de bolhas (x= anos, y= dias, z= temperatura em °C).	74
Figura 4-19. Python Mayavi points3d() (x= meses, y= dias, z= temperatura em °C).	75
Figura 4-20. R plot3D (x= meses, y= dias, z= temperatura em °C).	75
Figura 4-21. Data-Room gráfico de bolhas (x= meses, y= dias, z= temperatura em °C).	75
Figura 4-22. Ambiente preferido e resposta certa, considerando a técnica gráfico de barras.	82
Figura 4-23. Ambiente preferido e resposta certa, considerando a técnica gráfico de bolhas.	82
Figura 4-24. Ambiente preferido com resposta igual em ambos os inquéritos, considerando a técnica gráfico de barras.	82
Figura 4-25. Ambiente preferido com resposta igual em ambos os inquéritos, considerando a técnica gráfico de bolhas.	83
Figura 4-26. Pistas visuais relevantes nas visualizações.	84

Figura 4-27. Ambientes que permitem uma exploração mais fácil, considerando a técnica gráfico de barras. _____	84
Figura 4-28. Ambientes que permitem uma exploração mais fácil, considerando a técnica gráfico de bolhas. _____	84
Figura 4-29. Ambientes mais apetecíveis de visualizar no ecrã de um telemóvel, considerando a técnica gráfico de barras. _____	85
Figura 4-30. Ambientes mais apetecíveis de visualizar no ecrã de um telemóvel, considerando a técnica gráfico de bolhas. _____	85
Figura 4-31. Pistas visuais e retenção na memória (resposta “Outra” está assinalada com 0%). _____	85

1. Introdução

Este trabalho foi realizado no âmbito do mestrado em Tecnologias e Sistemas Informáticos Web da Universidade Aberta (UAb). O mesmo consiste numa dissertação de mestrado que se enquadra na área da visualização de informação, mais concretamente na avaliação da análise visual de dados e raciocínio, e trata o tema da visualização de dados do tipo *big data* em ambientes imersivos virtuais móveis.

A grande quantidade de dados armazenados atualmente não tem utilidade se não conseguirmos entender as relações e dependências entre os dados e construir modelos significativos. Para identificar novas tendências e padrões mais facilmente, faz sentido que os dados sejam limpos, processados e apresentados de maneira amigável. Os ambientes imersivos virtuais oferecem grandes oportunidades em diversas áreas, mas existem poucos estudos sobre a sua utilidade como ferramenta de visualização de informação.

Como hipótese a ser avaliada, este trabalho assume que é possível a visualização de dados do tipo *big data* em ambientes imersivos virtuais móveis de forma eficaz. Por eficácia definimos a capacidade de adquirir e reter conteúdo, a qualidade da experiência e se é fácil ou não o utilizador concentrar-se. Assim, propomos, como objetivo, um modelo visual para verificar a eficácia relativa da visualização de dados *big data* com a aplicação de técnicas de visualização de dados 3D num ambiente imersivo virtual móvel, comparando com outros ambientes utilizados na atualidade para o mesmo fim. O trabalho irá avaliar o modelo proposto, considerando a reação de participantes, em termos de aquisição e apreensão de informação para determinar mais-valias introduzidas.

Para a implementação deste projeto foi necessário adquirir conhecimento sobre dados do tipo *big data* (dimensões, escalabilidade, integração e processamento), visualização de informação (técnicas visuais imersivas e visualização de *big data*) e sobre o estado atual da visualização de dados em ambientes imersivos virtuais móveis (analisando tanto aplicações já existentes, quanto o hardware necessário e explorando também ambientes de desenvolvimento e bibliotecas que permitem criar este tipo de ambientes).

Espera-se com este trabalho contribuir para um maior conhecimento do valor da utilização de ambientes imersivos virtuais para dispositivos móveis na visualização de dados do tipo *big data* e demonstrar se estes são significativamente mais agradáveis, fáceis de resumir a informação e tão eficazes quanto a visualização desse tipo de dados em outros ambientes de visualização de dados.

O presente capítulo introduz os principais desafios relacionados com a área suprarreferida que motivaram e justificaram a concretização deste projeto, os objetivos desta dissertação e as suas contribuições (secção 1.1.). Posteriormente, é apresentada a metodologia de investigação utilizada (secção 1.2.) e descrita a estrutura do documento (secção 1.3.).

1.1. Motivações, objetivos e contribuições fundamentais

Com o passar do tempo e avanços tecnológico-científicos, as quantidades e complexidade dos dados aumentaram. As nossas vidas também se tornaram traços de dados: transações em supermercados, telefonemas e assim por diante. O conceito de *big data* surgiu neste contexto e trata conjuntos de dados tão volumosos e complexos que os software de processamento de dados tradicionais são inadequados para lidar com eles. Segundo estatísticas (Statista, março de 2018, dezembro de 2018),

- em 2015, 12 *zetabytes* de dados foram criados globalmente e este número está previsto atingir 163 *zetabytes* em 2025;
- em 2018, estima-se que o mercado de software de *big data* valha 14 mil milhões de dólares americanos, enquanto o mercado global valerá 42 mil milhões.

Hal Varian, chefe economista da Google e professor emérito da Universidade da Califórnia, Berkeley, comentou as oportunidades emergentes para profissionais de informática da seguinte forma: “Então, o que está a ficar omnipresente e barato? Dados. E o que é complementar aos dados? Análise. Então, a minha recomendação é fazer muitos cursos sobre como manipular e analisar dados: bases de dados, aprendizagem automática, econometria, estatística, visualização e assim por diante” (Chen, Chiang & Storey, 2012).

Os dados armazenados no presente não têm utilidade, se não conseguirmos entender as relações entre os dados e construir modelos significativos (Aghamirakarimi, Lemire, 2017). Estes mesmos dados estão a ser usados para gerar conhecimento e facilitar a tomada de decisões, trazendo usos e benefícios importantes para as pessoas, comércio, governo e sociedade (Landon-Murray, 2016). Como mencionado por Kammer *et al.* (2018), “o problema dos dados tipo *big data* não consiste apenas em lidar com esse imenso volume de dados; os dados precisam de ser limpos, processados e apresentados de maneira amigável, intuitiva e interativa”. Além disso, de modo a extrair conhecimento dos dados faz sentido “combinar as forças de humanos e computadores para processamento transparente de dados e informação” (Keim *et al.*, 2010). A nossa interpretação desta frase remete para que os dados sejam analisados, estudados ou interpretados utilizando técnicas de visualização, o que permite identificar novas tendências e padrões mais facilmente.

Dietz (2002) observa que um dos sonhos da tecnologia com futuro, mesmo que não seja ainda determinado, é a imersão. Wagner (Dietz, 2002) descreve da seguinte forma “Considerando que o público, aquele representante da vida quotidiana, esquece os confins do auditório, e vive e respira agora apenas na obra de arte que parece a vasta extensão de todo o mundo”. Por outras palavras, de Wagner aos dioramas panorâmicos de Daguerre e à Cratera Roden de James Turrell (Dietz, 2002), os artistas querem obras de arte nas quais o espectador está totalmente imerso. Segundo dados estatísticos, prevê-se que o número de vendas dos *head-mounted-displays* (HMD) cresça mais de 1000% (Statista, janeiro de 2017) em 2020, comparativamente com 2016, o que demonstra que este sonho não é exclusivo do entretenimento.

Quando temos imersão e tecnologia juntos, chegamos rapidamente a ambientes imersivos virtuais. À primeira vista, a métrica de custo-benefício proposta por Chen e Golan (2016) indica que a visualização em ambientes imersivos virtuais pode sofrer das

desvantagens de falta de abstração e alto custo, mas um olhar superficial na história dos ambientes imersivos virtuais e da criatividade neste espaço como um todo sugere que há mais para entender (Chen *et al.*, 2017). Existem, ainda, vários investimentos na área da visualização de dados em ambientes imersivos, como é o caso da instalação imersiva de visualização de 360°, ideia do cientista Arie Kaufman cujo esforço foi financiado em parte por uma doação de 1,4 milhões de dólares americanos da *National Science Foundation* (Williams, 2013). Para uma noção mais clara, é previsto que o mercado de realidade aumentada e virtual totalize os 18,8 mil milhões de dólares americanos em 2020, e espere-se uma expansão drástica nos próximos anos (Statista, novembro de 2019).

Dados estatísticos provenientes de Statista (novembro de 2016) sobre o número de utilizadores de telemóvel em todo o mundo de 2015 a 2020 reforçam a relevância em explorar a eficácia da utilização de ambientes imersivos virtuais móveis ao demonstrar que:

- o número de utilizadores de telemóveis no mundo deve ultrapassar a marca de cinco mil milhões até 2019;
- em 2016, estima-se que 62,9% da população mundial já possuía um telemóvel e prevê-se que a penetração dos telemóveis continue a crescer, chegando a 67% até 2019.

Dentro deste mundo, apontamos para o uso de técnicas 3D, pois o cérebro humano é treinado para sentir e agir em três dimensões, sendo a terceira dimensão percebida como uma vantagem pelo utilizador (Drossis, Margetis & Stephanidis, 2016).

É inegável o interesse num trabalho deste para artistas e curiosos. Este interesse alastra-se a cientistas e investigadores com necessidade de existência de novos métodos para visualização de grandes quantidades de dados, como vemos nos seguintes trabalhos:

1. *An Algorithm to Condense Social Networks and Identify Brokers* (Cavique, Marques & Santos, 2014), que teve como objetivo simplificar o trabalho de um analista de redes sociais que, utilizando grafos para as visualizações num ecrã de computador ou num papel, ao lidar com grandes redes a visualização torna-se incompreensível. Este objetivo foi alcançado com a criação de um novo algoritmo que retorna as informações necessárias para criar uma boa visualização de grandes redes usando um grafo condensado.

2. *Virtual memory palaces: immersion aids recall* (Rodrigues *et al.*, 2017), onde o foco central do artigo é a apresentação de exemplos, de conceções novas para a utilização de realidade virtual, permitindo-nos ter uma noção atempada e privilegiada da reflexão informada que está a ser feita pelas equipas que criaram estas ferramentas (o *iViz* e a da paisagem marciana), para com isso informarmos a nossa própria inspiração na conceção de ferramentas com o mesmo propósito.

Os ambientes imersivos virtuais oferecem grandes oportunidades em diversas áreas, mas existem poucos estudos sobre a sua utilidade como ferramenta de visualização de dados. Posto isto, surge a questão: “serão os ambientes imersivos virtuais móveis mais eficazes para visualizar dados do tipo *big data* comparativamente com os métodos utilizados atualmente?”. Neste trabalho assumimos que sim, ou seja, que o utilizador consegue compreender o que observa, lembra-se do que visualizou, gosta e consegue concentrar-se melhor em ambientes imersivos virtuais móveis.

Este trabalho tem como objetivo abordar os desafios apresentados na secção anterior, em particular, explora a eficácia relativa da visualização de dados tipo *big data* com técnicas 3D em ambientes imersivos virtuais. Mais especificamente foram definidos os seguintes objetivos:

1. Efetuar o levantamento do estado-da-arte da área de visualização de informação, com enfoque nas técnicas 3D, e dos ambientes imersivos, em especial orientados para dispositivos móveis;
2. Propor um modelo visual para verificar a eficácia relativa da visualização de dados *big data* com a aplicação de técnicas de visualização de dados 3D num ambiente imersivo virtual móvel comparando com outros ambientes;
3. Desenvolvimento e implementação de um protótipo com base no modelo visual proposto;
4. Realizar a avaliação do modelo (e respetiva hipótese) com o auxílio do protótipo, realizando testes com um grupo de utilizadores;
5. Propor melhorias e/ou modificação de técnicas visuais utilizadas pelo modelo e/ou eventual criação de alguma nova.

Assim, no tocante a contribuições, apresentaremos um novo modelo visual para visualizar dados *big data* com a aplicação de técnicas de visualização de dados 3D num ambiente imersivo virtual móvel. Com a comparação com outros ambientes, compreendemos a eficácia relativa desta mesma visualização, desenvolvemos um protótipo (ou pré-produto) e melhoramos e / ou modificamos técnicas visuais utilizadas pelo modelo.

1.2. Metodologia de investigação

A presente dissertação será desenvolvida em quatro fases: pesquisa, planeamento de um modelo conceptual, implementação de um protótipo e avaliação/validação do modelo e respetiva hipótese inicial deste trabalho.

Este trabalho categoriza-se como um caso de estudo, em que o caso analisado procura averiguar se é mais eficaz visualizar dados do tipo *big data* com técnicas de visualização de dados 3D em ambientes imersivos virtuais móveis comparativamente com outros ambientes.

O primeiro passo (Pesquisa) consiste na pesquisa bibliográfica, levantamento do estado-da-arte e escrita do trabalho relacionado. Numa segunda fase (Planeamento), focamos no planeamento, onde podemos ver a conceptualização do novo modelo para atingir uma solução melhor. A terceira fase engloba o desenho e implementação do protótipo (ou pré-produto) utilizado para avaliar o modelo conceptual. Por fim, na fase de testes do protótipo, os dados foram recolhidos e analisados.

1.3. Organização da dissertação

O presente documento descreve o trabalho desenvolvido e está organizado em cinco capítulos. Após a presente introdução, no capítulo 2 é feito o levantamento do estado-da-arte. No capítulo 3 é proposto o modelo conceptual de visualização de *big data* em ambientes imersivos virtuais, assumindo dispositivos móveis como ambiente de interação. O capítulo 4 apresenta a implementação e avaliação do modelo, incluindo a realização de testes e análise dos resultados. No capítulo 5 são apresentadas as conclusões e considerações sobre trabalhos a desenvolver futuramente. Posteriormente é possível de aceder às referências bibliográficas, e anexos. Dada a sua dimensão, os anexos foram colocados num documento à parte. Se procuram acesso ao mesmo, enviar email com o assunto “Anexos da Dissertação” para 1701127@estudante.uab.pt ou me@tiagoperes.eu.

2. Estado da arte

Neste capítulo é feita uma reflexão em torno de conceitos considerados importantes para ajudar à compreensão do trabalho desenvolvido. A informação exposta neste capítulo está associada à estrutura estabelecida para o mesmo, na qual constam cinco secções, onde abordaremos, respetivamente, fundamentais de *big data* (secção 2.1.), visualização de informação (secção 2.2.), visualização imersiva em dispositivos móveis (secção 2.3.), visualização de *big data* (secção 2.4.) e, por fim, síntese (secção 2.5.).

2.1. Fundamentos de *big data*

Big data são conjuntos de dados que são tão volumosos e complexos que os software de processamento de dados tradicionais são inadequados para lidar com eles. Em 2001, o analista da Doug Laney (2001) introduziu o conceito dos 3Vs que compõem a dimensionalidade do *big data*: volume (ou a quantidade mensurável de dados), variedade (ou seja, o número de tipos de dados) e velocidade (referente à velocidade de processamento ao lidar com esses dados) (Miller, 2017).

Atualmente consideram-se cinco Vs, ou cinco atributos principais, que definem essa nova tecnologia e que envolvem cinco novos desafios no campo da visualização de informações, mais precisamente como afetarão a visualização de dados (tal como é possível de verificar na Figura 2-1 (Rodrigues *et al.*, 2017)).

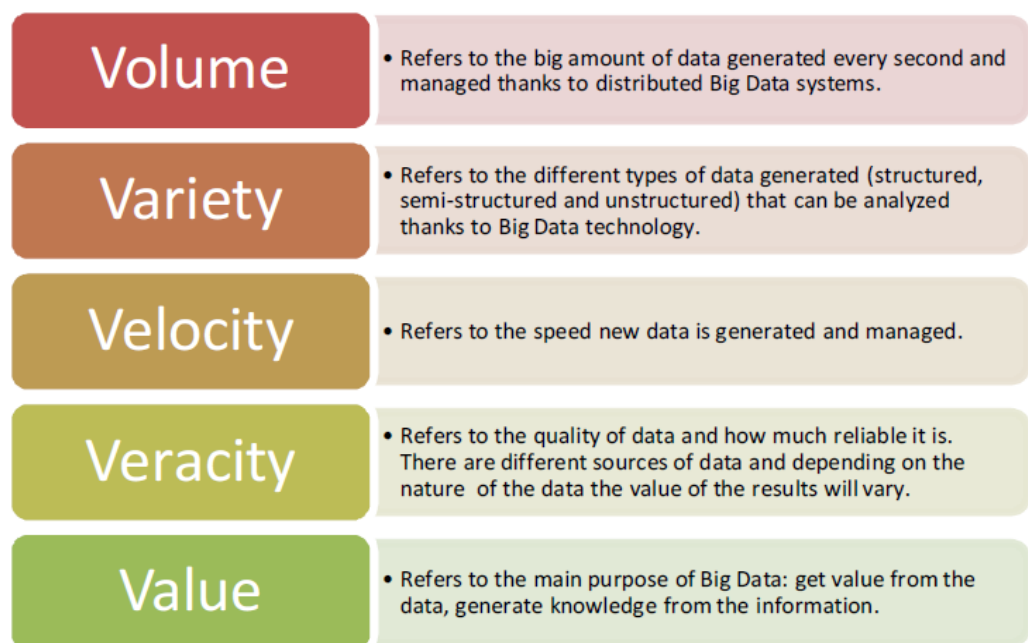


Figura 2-1. *Big data*: os 5 Vs.

Com todas as informações geradas e armazenadas, o grande desafio é como explorar cuidadosamente os dados de uma maneira que permita à organização ou aos investigadores extrair valor para aumentar a eficiência, aumentar os lucros ou descobrir tendências ocultas, recursos importantes ou inter-relações (Aukstakalnis, 2016). O resultado pode ajudar na tomada de decisões empresariais, ajudar organizações de defesa e inteligência a identificar ameaças emergentes, determinar quais os anúncios que provavelmente serão clicados e fornecer informações sobre vários problemas de ciência e engenharia (Cukier, 2010).

A representação visual dos dados é um passo crucial no que toca a *big data*, para poder obter todos os benefícios que esta tecnologia fornece. Entre os principais benefícios das ferramentas de visualização (Wang, Wang & Alexander, 2015), a tomada de decisão melhorada é a principal, mas outros aspetos importantes a considerar são melhores

análises *ad hoc* de dados, melhor colaboração entre utilizadores e aumento do retorno sobre o investimento (ROI) (Rodrigues *et al.*, 2017).

Existem centenas de conjuntos de dados gratuitos disponíveis, prontos para serem usados e analisados por qualquer pessoa. Em Anexos, secção I, Tabela 1, apresentamos uma lista de locais onde podem ser encontrados, tendo a preocupação de obter uma variabilidade de contextos.

2.1.1. Dimensões e escalabilidade

Até então, vimos que *big data* é um termo geral usado para se referir a qualquer coleção de dados tão grande e complexa que excede a capacidade de processamento dos sistemas e técnicas convencionais de gestão de dados. As aplicações de *big data* são infinitas: cada parte do negócio e da sociedade estão a mudar devido ao fato de que agora temos muito mais dados e capacidade de análise.

Big data é comumente caracterizado usando vários V's (Figura 2-1). Os três primeiros são volume, velocidade e variedade: volume refere-se à grande quantidade de dados gerados a cada segundo, minutos, horas e dias no mundo, variedade refere-se às formas cada vez maiores que os dados podem apresentar, como texto, imagens, voz e dados geoespaciais e velocidade refere-se à velocidade na qual os dados estão a ser gerados e ao ritmo no qual os dados movem-se de um ponto para o outro (Altintas, Gupta & Nguen, sem data).

Temos enormes quantidades de dados em diferentes formatos e variando a qualidade que deve ser processada rapidamente. Mais Vs foram apresentados à comunidade de *big data* à medida que foram descobertos novos desafios e formas de definir *big data*. Veracidade e valor são dois desses V's adicionais. A veracidade refere-se ao ruído e anormalidade nos dados, ou seja, às incertezas, veracidade e confiabilidade dos dados, muitas vezes incomensurável. Além disso, há que ter uma estratégia clara e um objetivo com valor que se extrai dos dados do tipo *big data* (Rodrigues *et al.*, 2017).

2.1.2. Integração e processamento

Como vimos na secção 2.1.1. Dimensão e escalabilidade, os dados do tipo *big data* variam em volume e velocidade exigindo o processamento dinâmico e escalável de lotes e fluxos. Para além disso, este tipo de dados possui uma variedade que exige o tratamento de dados em muitos sistemas de dados diferentes e integração de todos em escala. Em comparação com as tecnologias tradicionais de armazenamento, processamento, análise e visualização de dados, a área de *big data* exige tecnologias evolutivas de processamento massivo de dados em plataformas distribuídas e paralelas, como *Apache Hadoop* (*Hadoop* é um *framework* de código *open-source* baseado em Java feita por Doug Cutting com base no *benchmark* do *Google File System* (GFS) (Cho *et al.*, 2014)). *Apache Hadoop* tem como componentes básicos um sistema de ficheiros chamado *Hadoop File System* (HDFS) para armazenar *big data* e *MapReduce* para processamento.

MapReduce (Dean & Ghemawat, 2004), por sua vez, é um modelo de programação e uma implementação para processamento e geração de grandes quantidades de dados com um algoritmo distribuído paralelo num *cluster* (Apache MapReduce, 2019). De um modo geral, um programa *MapReduce* consiste principalmente em duas funções definidas pelo utilizador: *mapper* (que executa a filtragem e classificação) e *reducer* (que executa uma operação de resumo). O *input* de uma tarefa *Hadoop MapReduce* é um conjunto de pares de valores-chave (k, v) e a função *mapper* é chamada para cada um desses pares. A função *mapper* produz zero ou mais pares de valores-chave intermediários (k', v'). De seguida, a estrutura do *Hadoop MapReduce* agrupa esses pares de valores-chave intermediários pela chave intermediária k' e chama a função de *reducer* para cada grupo. Finalmente, a função *reducer* produz zero ou mais resultados agregados (Dittrich & Quiané-Ruiz, 2014).

Para entender num caso prático como o *MapReduce* funciona podemos observar a Figura 2-2 (Yadav, 2017), onde o mesmo é aplicado à contagem de palavras. Primeiro, os dados de entrada são carregados e divididos com base nos pares de valores-chave nos quais o mapeamento é executado. O output desse processo é os pares intermediários de valor-chave (ou seja, List ($K2, V2$)). Posteriormente, essa lista é dada ao redutor e todas as chaves similares são processadas no mesmo (ou seja, $K2, List(V2)$). Finalmente, todas as saídas do *shuffling* são combinadas para formar uma lista final de pares de valores-chave (ou seja, List ($K3, V3$)).

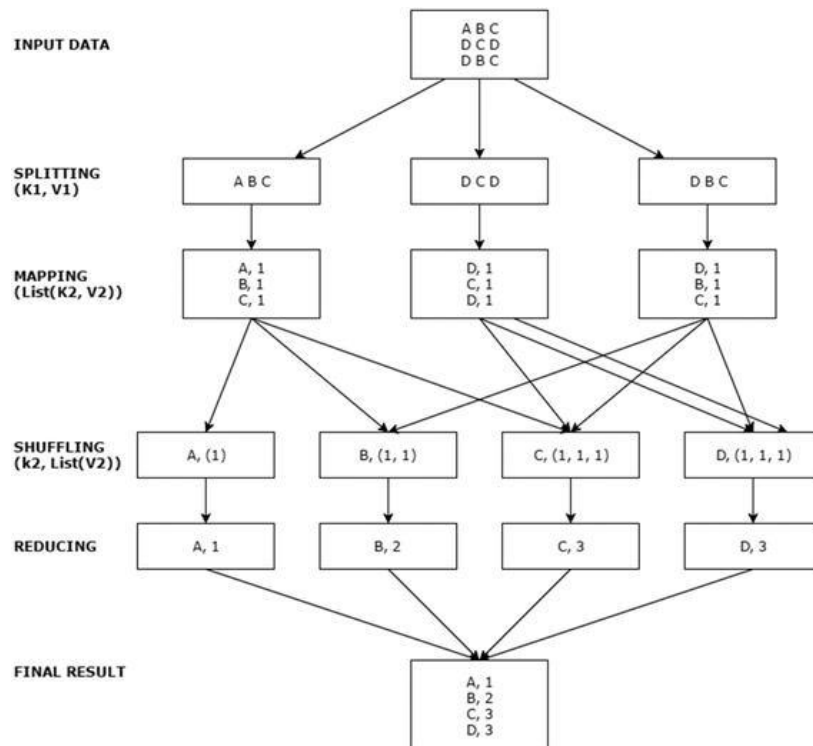


Figura 2-2. MapReduce aplicado ao processo de contagem de palavras.

A estrutura do *Hadoop MapReduce* utiliza um sistema de ficheiros distribuídos para ler e gravar os seus dados. Normalmente, o *Hadoop MapReduce* usa o *Hadoop Distributed File System* (HDFS), que é a contraparte de código aberto do *Google File*

System e, portanto, o desempenho I/O de um trabalho do *Hadoop MapReduce* depende fortemente do HDFS (Dittrich & Quiané-Ruiz, 2014). No HDFS um bloco é geralmente 128 MBs, ou seja, um ficheiro de 1 GB consistiria em 8 blocos e 1 TB em 8000 blocos (Durham, sem data).

Atualmente podemos armazenar os nossos dados nos nossos centros de dados recorrendo ao HDFS, ou Apache HBase, na nuvem ou com uma opção híbrida. Alibaba Cloud E-MapReduce (2020), Amazon EMR (2019) (*Elastic MapReduce*), Cloudera Distributed Hadoop (2020), Google Cloud Dataproc (sem data), MapR (2020), Microsoft Azure HDInsight (2020) e Oracle Big Data Cloud Service (2019) são exemplos de serviços na *cloud* que facilitam a criação e a gestão de *clusters* elásticos que executam o *Hadoop* e outros aplicativos no ecossistema do *Hadoop*. Entre os sistemas de *big data* temos Apache Impala (sem data), o Apache Hive (2014), o Apache Drill (2014) e o Presto (2019).

A Figura 2-3, criada por trabalhadores da Gartner (Ronthal, Edjlali & Greenwald, 2019), tem por base uma análise dos vendedores destas e de outras soluções. Existem ainda soluções que tiram partido de produtos líderes de mercado, não consideradas na Figura 2-3, como é o caso do Trifacta (2020), com versão gratuita, pro e empresa, para limpeza e tratamento de dados. De salientar ainda que a versão gratuita é uma aplicação web que permite importar ficheiros até 100MBs e a solução para empresas suporta plataformas como Cloudera e ambientes de computação na nuvem incluindo Amazon Web Services, Google Cloud e Microsoft Azure.



Figura 2-3. Quadrante mágico para soluções de gestão de dados para análise (Ronthal, Edjlali & Greenwald, 2019).

2.2. Visualização de informação

Visualização de informação designa o estudo da transformação de dados, informações e conhecimento em representações visuais interativas que são interpretadas pelo ser humano (Figura 2-4). Este tipo de representações visuais é muito importante para os utilizadores porque fornece modelos mentais de informação. Assim, temos que visualização da informação é uma área de pesquisa que visa auxiliar os utilizadores a explorar, compreender e analisar dados através de uma exploração visual progressiva e interativa (Shiravi, Shiravi & Ghorbani, A., 2012). De notar que a visualização da informação não é apenas sobre a criação de exibições gráficas de estruturas de informação complexas e latentes; a mesma contribui para uma gama mais ampla de atividades cognitivas, sociais e colaborativas (Chen, 2006).

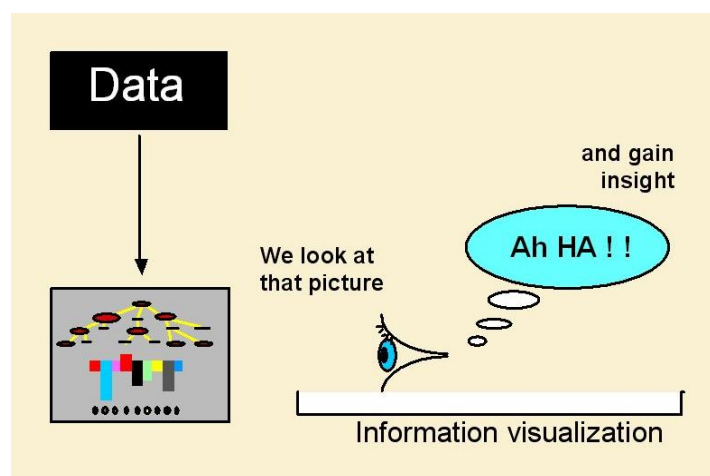


Figura 2-4. O processo de visualização de informação (Spence, R., 2007).

A visualização de informações representa um dos fluxos mais recentes numa tendência de longa data no design moderno de interfaces do utilizador. Camadas crescentes de interface do utilizador estão a ser adicionadas entre o utilizador e o ecrã, mas a interface entre os dois está a tornar-se mais transparente, mais natural e mais intuitiva, como, por exemplo, os já tradicionais "What You See Is What You Get" (WYSIWYG) e "click and drag", ou "fly-through" (Chen, 2006).

Segundo Keim (2002) há muito a ganhar quando os dados são representados numa maneira mais visual. Sviokla (2009) escreve que a visualização de informação confere três benefícios gerais

1. Grandes visualizações são eficientes — por exemplo, permitem que as pessoas vejam grandes quantidades de dados rapidamente;
2. As visualizações podem ajudar analistas ou grupos a obter mais informações sobre a natureza de um problema e descobrir um novo entendimento;
3. Uma ótima visualização pode ajudar a criar uma visão compartilhada de uma situação e alinhar as pessoas às ações necessárias (como vemos na Figura 2-5 que demonstra o quão severamente o coronavírus afetou o mercado de trabalho nos Estados Unidos da América (EUA)).

More Than 3 Million Americans Lost Their Jobs Last Week. See Your State.

Official statistics have revealed how severely coronavirus has hurt the job market. But it may take several months before we know whether this economic disaster will resemble a storm or a long winter.

BY QUOCTRUNG BUI and JUSTIN WOLFERS

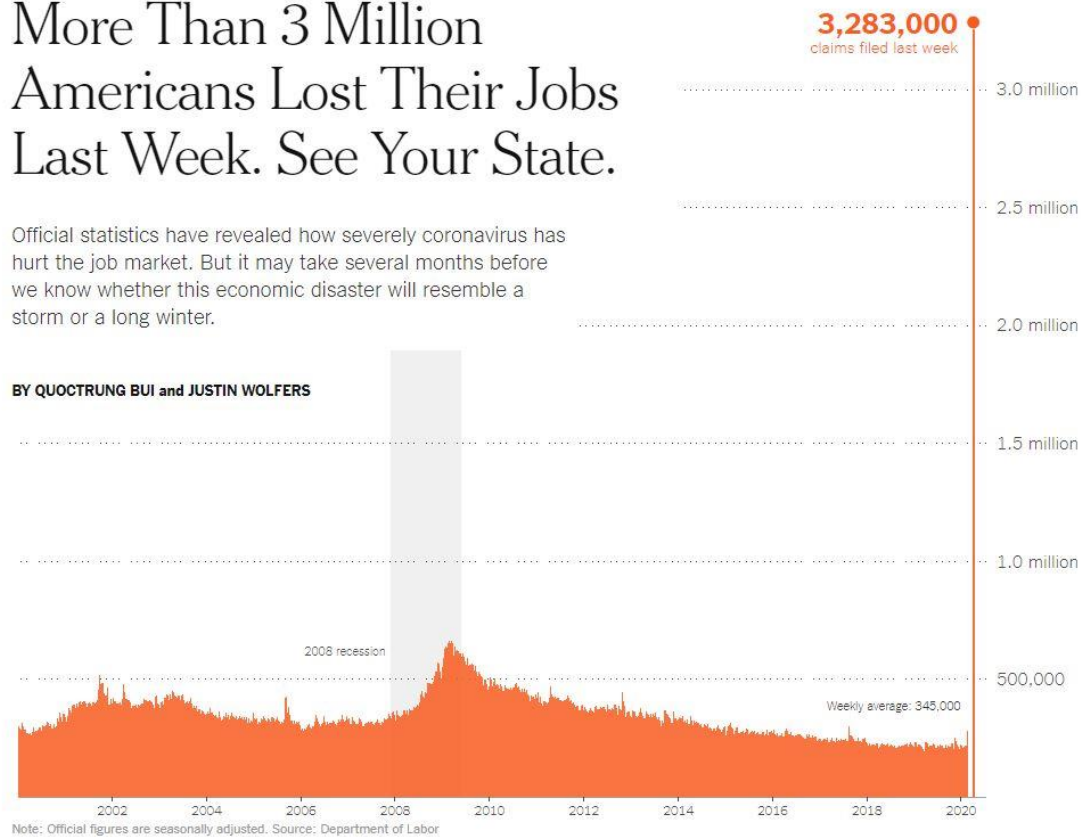


Figura 2-5. Estatísticas oficiais revelam o quanto o coronavírus afetou o mercado de trabalho nos EUA (Bui & Wolfers, 2020).

O cérebro humano é capaz de tarefas extraordinárias no processamento de informação, desde que os dados estejam numa forma que aproveite as capacidades dos mecanismos perceptuais humanos. Destes mecanismos, a visão é o nosso sentido dominante, com cerca de um quarto do nosso cérebro dedicado ao processamento de estímulos visuais e que fornece o canal perceptivo de maior largura de banda em nossos sistemas cognitivos (Reda *et al.*, 2013). É natural que com o crescimento no interesse em *big data*, a visualização de informação (*InfoVis*) seja amplamente utilizada numa variedade de aplicações de análise de dados [5, 6, 7]. Os exemplos incluem análise visual de dados de negócios, dados científicos, histórico de estudantes, dados desportivos, imagens e vídeos, dados de leilão e resultados de pesquisa, entre outros.

Uma melhor compreensão e análise de dados / informação está a tornar-se numa maneira cada vez mais poderosa de crescimento, produtividade e inovação. Para além disso, a visualização de dados é a ponte entre os dados e a intuição humana, ou seja, uma melhor visualização de dados significa exibir os dados de uma maneira que um humano possa encontrar correlações e padrões dentro dos dados facilmente. Por todas estas razões, acredita-se que as técnicas da *InfoVis* são valiosas e, portanto, vale a pena estudar, especialmente as recentes tendências de pesquisa (Liu *et al.*, 2014).

Existem várias técnicas comuns para a representação visual de dados, incluindo tabelas, histogramas, gráficos de bolhas, *dashboards* e *storyboards*. Isso não significa que, no momento de fazer uma visualização, qualquer tipo de representação disponível pode ser escolhido, mas, cada método de visualização tem um objetivo específico e, portanto, é necessário determinar que tipo de informação é desejada para exibir e, de seguida, escolher a representação gráfica mais apropriada. Por exemplo, se quisermos exibir porções numéricas, pode utilizar-se gráficos circulares, enquanto para comparações os gráficos de barras são mais úteis (Rodrigues *et al.*, 2017).

Chen (1999) realça que uma técnica de visualização de informação específica pode ser adaptada para atender a vários requisitos dos utilizadores; por exemplo, as árvores de cone têm sido usadas em vários sistemas com diferentes lógicas, como *LyberWorld* (Hemmje, Kunkel & Willett, 1994), *Cat-a-Cone* (Hearst & Karadi, 1997) e *Hyperbolic 3D* (Munzner, 1998). Uma forma de adaptar é através da utilização de diferentes dimensões, como no estudo de Ware e Franck (1996). Os mesmos estudaram sobre como as pessoas compreendem os gráficos 3D apresentados em 2D e em 3D e notaram que as pessoas tiveram um desempenho significativamente melhor em 3D do que em condições 2D.

Cleveland e McGill estudaram as pistas visuais que as pessoas são capazes de decodificar com mais precisão (Cleveland & McGill, 1984). Por outras palavras, existem pistas visuais mais fáceis de compreender do que outras (Figura 2-6). Os dois classificaram essas pistas na lista a seguir (Simon, 2014)

1. Posição ao longo de uma escala comum (por exemplo, gráfico de dispersão);
2. Posição em escalas idênticas, mas não-alinhadas (por exemplo, múltiplos gráficos de dispersão);
3. Comprimento (por exemplo, gráfico de barras);
4. Ângulo e inclinação (por exemplo, gráfico de pizza);
5. Área (por exemplo, gráfico de bolhas);
6. Volume, densidade e saturação de cor (por exemplo, mapa de calor);
7. Cor (por exemplo, mapa de notícias).

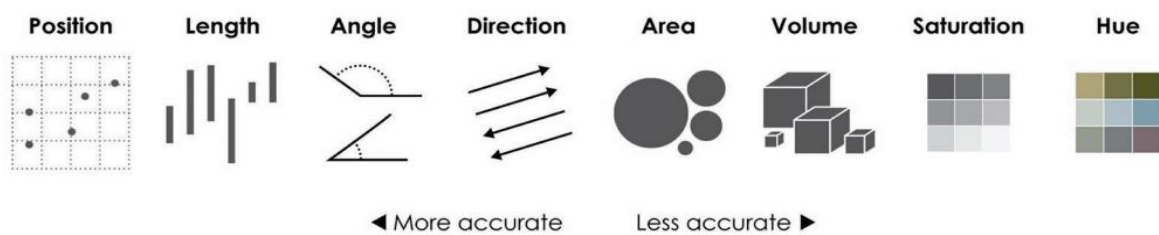


Figura 2-6. Ranking de pistas visuais (Cleveland & McGill, 1984).

O mantra de análise visual diz: "Analise primeiro, mostre o importante, faça *zoom*, filtre e analise ainda mais, detalhes conforme a demanda" (Keim, Mansmann & Thomas, 2010). No entanto, também implica confiar nas ferramentas para analisar os dados e, portanto, para determinar quais informações são valiosas e quais não tem qualquer importância.

Em 2007, Yi, Kang e Stasko (2007) forneceram uma pesquisa abrangente para estudar o papel da interação na *InfoVis*. Eles classificaram técnicas de interação em sete categorias: seleção, exploração, reconfiguração, codificação, abstração / elaboração, filtragem e conexão.

A razão fundamental pela qual o design de visualizações é um empreendimento difícil é que a grande maioria das possibilidades no espaço de design será ineficaz para qualquer contexto de uso específico. Em alguns casos, um design possível tem uma correspondência fraca com as propriedades dos sistemas perceptivos e cognitivos humanos. Noutros casos, o design seria compreensível por um ser humano noutra configuração, mas é uma correspondência ruim com a tarefa pretendida. Apenas um número muito pequeno de possibilidades está no conjunto de escolhas razoáveis, e apenas uma fração ainda menor é uma excelente opção. Escolher aleatoriamente as possibilidades é uma má ideia, porque as possibilidades de encontrar uma solução muito boa são muito baixas (Munzner, 2015).

O elevado número de simulações e *arrays* de sensores, em combinação com o aumento de recursos computacionais, geraram conjuntos de dados grandes, complexos e de alta dimensão utilizados para estudar fenómenos em vários campos de estudo. A visualização desempenha um papel importante na exploração desses conjuntos de dados (Liu *et al.*, 2017). As limitações físicas dos dispositivos de exibição e nossos sistemas visuais impedem a exibição direta e reconhecimento rápido de estruturas com dimensões maiores do que duas ou três. Como guia, é proposta a divisão do problema complexo do design visualizações em quatro níveis em cascata (Figura 2-7), fornecendo uma estrutura de análise que permite abordar diferentes preocupações separadamente.

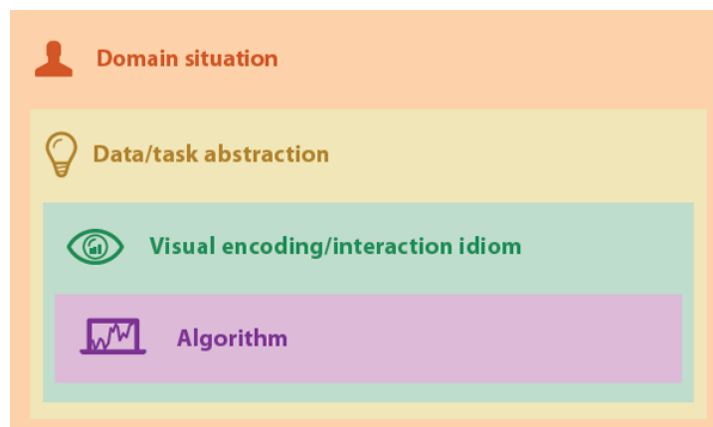


Figura 2-7. Os quatro níveis de design de visualizações (Munzner, 2015).

Existem duas abordagens para o design visual: de cima para baixo ou de baixo para cima. Com o trabalho orientado a problemas, começamos no nível de situação de domínio superior e seguimos o caminho através de decisões de abstração, expressão idiomática e algoritmo. No trabalho orientado por técnica, trabalhamos num dos dois níveis inferiores, design de idioma ou algoritmo, em que o nosso objetivo é inventar novas expressões idiomáticas que melhor suportem abstrações existentes ou novos algoritmos que melhor suportem idiomas existentes (Munzner, 2015).

2.2.1. Técnicas virtuais imersivas

Nos últimos anos, a comunidade científica enfrentou um aumento exponencial na quantidade de dados obtidos através da observação de fenómenos naturais ou gerados por complexas simulações de supercomputadores. Todos esses dados são obtidos em dimensões cada vez maiores, armazenados digitalmente e analisados usando instrumentos de visualização, como o CAVE2 (Febretti *et al.*, 2013) (Figura 2-8).



Figura 2-8. Visualização em CAVE2 (Febretti *et al.*, 2013).

Os dados podem ser estáticos ou podemos estar imersos em processos dinâmicos. A representação tridimensional de dados reais ou modelados é importante para a compreensão dos dados e para a tomada de decisões após esse entendimento, um tópico relevante para vários campos, especialmente no momento do crescimento exponencial de conjuntos de dados (Slater & Sanchez-Vives, 2016). Há uma diferença substancial entre olhar para representações de dados 3D num ecrã e imergir nos dados, navegar por eles, interagir com eles e explorá-los de fora para dentro. Isso, junto com os sistemas de rastreamento corporal, permite uma interação mais natural com os dados (Slater & Sanchez-Vives, 2016), ainda que incorramos em custos mais elevados.

Evidências empíricas sólidas apoiam o benefício da imersão em cenários de exploração e análise de dados (Bowman & McMahan, 2007). A imersão, esta, pode ser física (onde há imersão de órgãos sensoriais em dispositivos físicos) ou imersão psicológica (onde há apenas a sensação de estarmos dentro de um espaço) (Tham *et al.*, 2018).

Ambientes imersivos são os ambientes criados por tecnologias imersivas (Bah *et al.*, 2016). Estes ambientes fornecem um contexto particularmente estimulante para entender as implicações teóricas e práticas de várias questões fundamentais e recursos específicos de visualização de informação (Chen, 2006), mostrando ser uma promessa considerável como uma forma natural de interação humano-computador. As tecnologias de ambientes imersivos virtuais proporcionam uma melhor compreensão de formas e espaços tridimensionais através de fenómenos preceptivos (Schultze & Ramamoorthi, sem data).

Em *The Visual Display of Quantitative Information*, Tufte (1983) escreve que a visualização de dados torna a compreensão humana de grandes conjuntos de dados mais coerente. Ele serve um propósito claro de descrever dados representados em várias formas; por exemplo, como abstrações (gráficos de bolhas, gráficos de barras etc.) e frequentemente como um termo para descrever reconstruções 3D de dados como objetos no espaço 3D (por exemplo, estruturas anatómicas reconstruídas em 3D, como dados cerebrais e pedaços planos de ressonância magnética [MRI] (Figura 2-9) e relações entre genes e doenças (Figura 2-10), ou até mesmo CAVE (Figura 2-11) num ambiente virtual e aumentado).

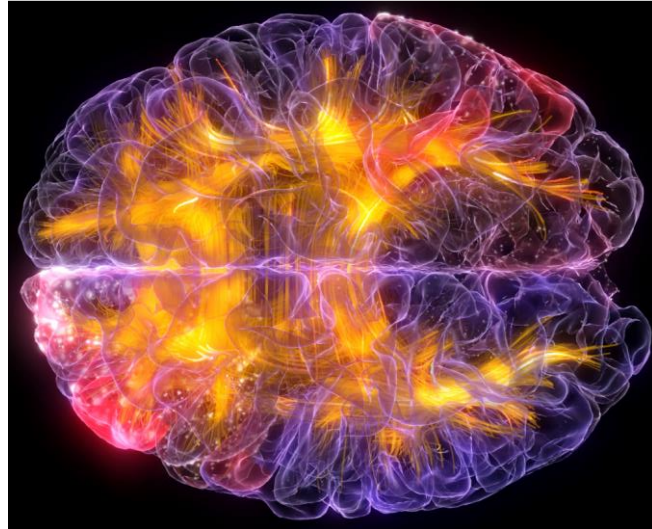


Figura 2-9. *Glass Brain*: visualização 3D do cérebro (Glass Brain, sem data).

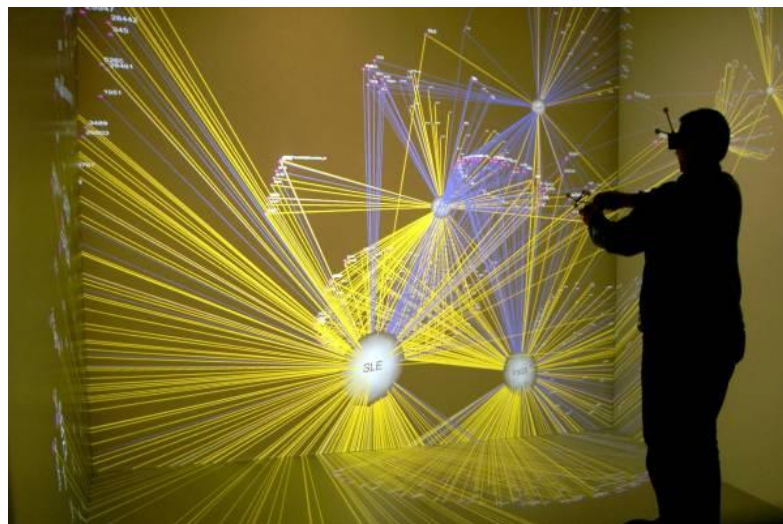


Figura 2-10. Rede 3D vista dentro da CAVE (Febretti *et al.*, 2013).

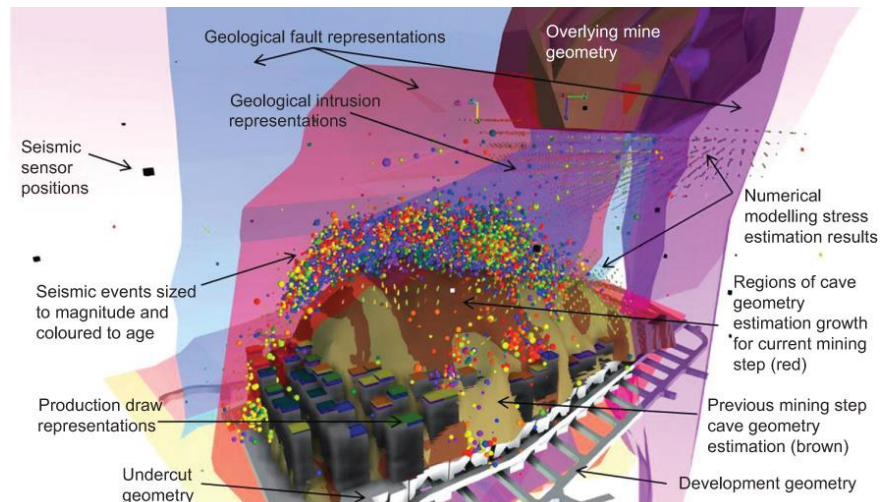


Figura 2-11. *Block Mining Visualizer*: visualização de vários conjuntos de dados combinados associados ao processo de mineração subterrânea (Aukstakalnis, 2016).

Os dados em si são comparativos, relacionais, multivariados e podem permitir que o utilizador investigue questões específicas ou explore os dados em geral para obter um melhor entendimento das suas qualidades. Aqui estão algumas das principais características das visualizações interativas de dados em ambientes imersivos (Mohan, Lukas & Pangilinan, 2019):

- Pode compreender e ordenar dados relacionais, através da integração de descrições para distinguir dados categóricos, sejam dados qualitativos e possam envolver alguns traços estatísticos (foco em dados quantitativos);
- Enfatiza a estética para ajudar os utilizadores a entender os dados através de um bom *design*, não apenas para fins decorativos.

Muitas das grandes infraestruturas de ambientes imersivos virtuais são equipadas com monitores da ordem dos gigapixel. Normalmente, estes monitores têm sido utilizados para visualizar alguns conjuntos de dados muito grandes, como imagens gigapixel, e grandes modelos biomoleculares para simular os comportamentos dinâmicos de milhões de átomos (Chen *et al.*, 2017). A visualização é a ponte principal entre o conteúdo quantitativo dos dados e a intuição humana, e pode-se argumentar que não podemos realmente compreender ou compreender intuitivamente nada (incluindo construções matemáticas), que não podemos visualizar de alguma forma (Schultze & Ramamoorthi, sem data).

Shneiderman (2003) reforça que precisamos de suportar as seguintes três funcionalidades

- Consciencialização rápida da situação através de visões gerais eficazes;
- Número reduzido de ações para realizar tarefas;
- Dar *feedback* significativo às ações do utilizador.

Embora o ambiente imersivo virtual tenha o potencial de aumentar o valor do produto, oferecendo uma experiência de produto virtual imersiva e agradável, o ambiente resultante também pode ser uma distração (Bermúdez *et al.*, 2005). Devemos, assim, colocar restrições sobre o que o utilizador do ambiente imersivo pode fazer, sempre que

faz sentido e quando melhora a usabilidade. Muitas tarefas podem ser feitas muito melhor com restrições sobre elas. Para além disso, dar *feedback* ao utilizador é necessário sempre que este faz uma mudança no ambiente imersivo virtual.

O *feedback* pode ser uma exibição elaborada de informações que mostra ao utilizador o estado de um algoritmo ou um dispositivo que pode medir as coisas. O *feedback* do sistema pode acontecer de diferentes maneiras; no caso dos ambientes imersivos virtuais, pode considerar três sentidos diferentes: o sentido visual, auditivo, e o háptico. É um sinal de um ambiente imersivo virtual de alta qualidade quando é dado *feedback* ao utilizador não apenas por um sentido, mas múltiplos combinados (Schultze & Ramamoorthi, sem data).

Um ponto relevante aqui é a latência, que designa o atraso introduzido pelo computador ou telemóvel a processar um *input* ou atraso através de um cabo, incluindo o cabo de rede. Esse atraso pode ter um impacto severo num ambiente imersivo virtual, devendo haver o mínimo de latência possível ao transmitir uma imagem (este efeito pode também ser neutralizado com uma redução da quantidade de dados exibidos, reduzindo a complexidade dos modelos 3D).

2.3. Visualização imersiva em dispositivos móveis

Atualmente existem vários ambientes que permitem converter informação numa forma gráfica para que seja consumida e interpretada. Em Anexos, secção II, Tabela 2, apresentamos uma lista dos mais utilizados.

Além disso, existem vários projetos com os olhos postos no futuro que investem no auxílio das visualizações imersivas de dados em ambientes imersivos virtuais móveis. Este dispositivo móvel é tipicamente um telemóvel com um HMD (*head mounted display*), ainda que também possam ser HMDs autónomos. Em termos de aplicações, temos o caso das criadas pelas empresas *3Data*, *BadVR*, *DataView VR*, *Virtualitics*, *Virtual Cove*, *Nanome*, *Datavized*, *Deloitte*; no hardware, mais especificamente com os HMDs para usar com o telemóvel, temos, entre outros, o *Cardboard*, o *Gear VR*, o *Daydream*, *Bridge*, *Bridge (Polycarbonate)* e *MergeVR* (ver Anexos, secção III, Tabela 3 para lista com outros HMDs, autónomos ou não). No auxílio de criação de novo software imersivo, com ambientes de desenvolvimento integrado (ADIs), vamos concentrar-nos no *Unity*, *Android Studio*, *Unreal* e bibliotecas que permitem criação de gráficos com dados 3D como *DXR*, *Graph and Chart* e *VR-VIZ* (ver Anexos, secção IV, Tabela 4 para outros possíveis ADIs). O foco nos ADIs acima mencionados deve-se ao reconhecimento dos mesmos (qualidade, documentação, comunidade,...), enquanto que o foco em gráficos com dados 3D, interativos, deve-se aos mesmos demonstrarem comportar-se melhor do que simplesmente 2D como vimos anteriormente. Nesta secção iremos explorar com maior detalhe os softwares e hardwares anteriormente referidos, abordando, por exemplo, bibliotecas até então não referidas no documento. Segue-se a apresentação de algumas aplicações atuais.

2.3.1. Aplicações

3Data

3Data (2019) é uma empresa que dá nome a uma ferramenta que permite visualização de gráficos em realidade virtual, realidade aumentada ou em duas dimensões. A empresa foi inicialmente conhecida por *DatavizVR*, Inc. tendo tornado público, em 2016, a aplicação *DatavizVR* na Steam (2016). Ainda que a *3Data* permita visualizar dados com telemóveis e modificar as técnicas existentes (alterar cores, tamanhos, formas, etc.), de momento não disponibiliza visualizações para ambientes imersivos virtuais, ou seja, todas as visualizações requerem um mundo virtual. A Figura 2-12 é um *screenshot* de um gráfico de barras de uma demonstração da aplicação num *Huawei PRA-LX1*, com versão Android 8.0.0 e giroscópio.

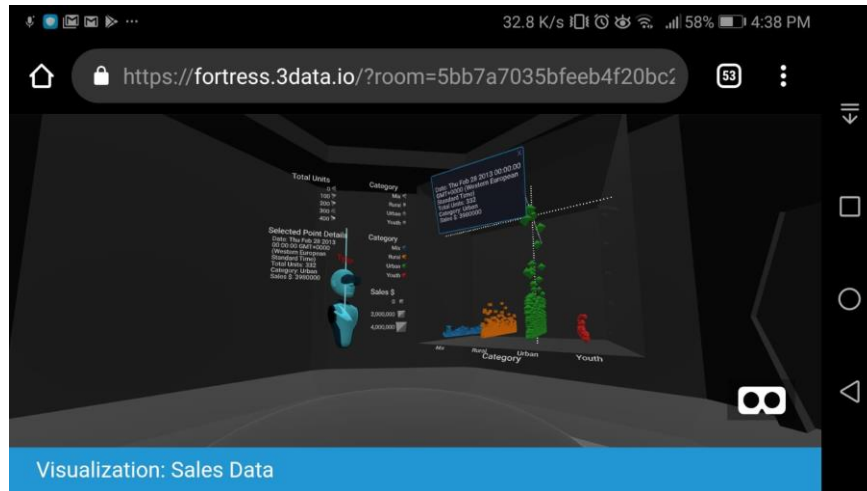


Figura 2-12. Gráfico de barras da aplicação 3Data.

BadVR

BadVR (2019) é uma empresa mais recente que surgiu nesta área, criadores do SeeSignal (2019) que permite visualização em realidade aumentada com HMD da Magic Leap (Figura 2-13). Ainda que não tenhamos conseguido saber mais informações através do sítio da empresa e contacto com a mesma, encontramos dois *screenshots* onde temos uma visualização de dados geospaciais (Figura 2-14) (Ribeca, 2018) (numa plataforma não identificada) e uma prova de conceito (*Climate Data*) que permite visualizar, com Magic Leap, dados climáticos (Figura 2-15) (magicleapdevs, 14 de novembro de 2019).

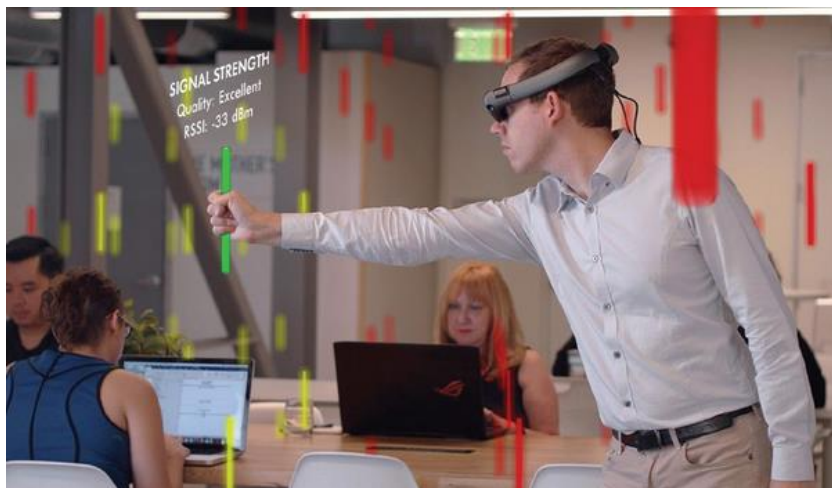


Figura 2-13. SeeSignal.

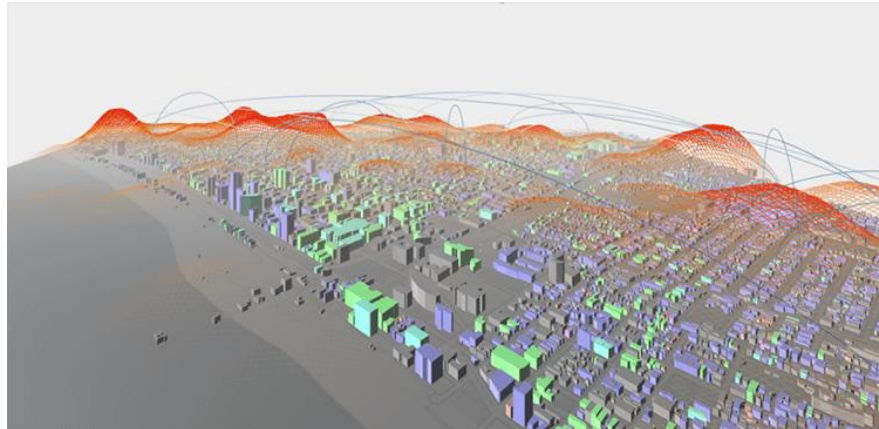


Figura 2-14. Visualização de dados geospaciais em plataforma não identificada.

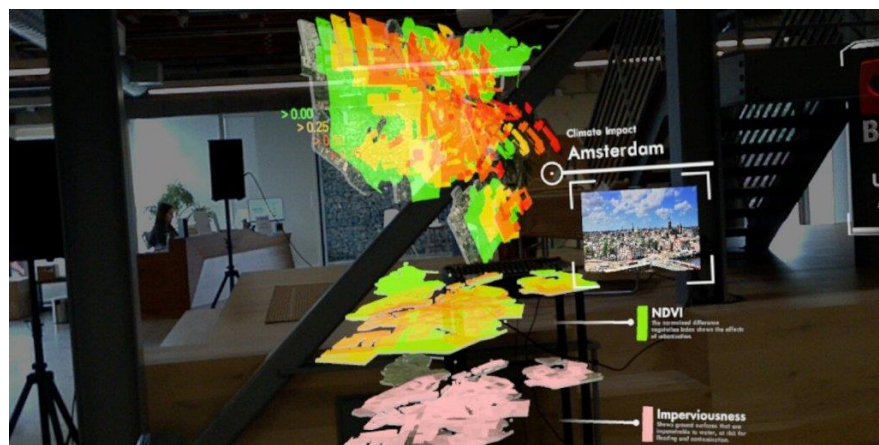


Figura 2-15. Conceito *Climate Data*. Visualização de dados climáticos com o HMD Magic Leap.

DataView VR

DataView VR (sem data) fornece ferramentas para visualizar dados 3D, enquanto facilita a colaboração e apresentação em realidade virtual e aumentada (Figura 2-16).



Figura 2-16. A “HyperDesk” do *DataView VR*.

Virtualitics

Virtualitics (2020) é uma plataforma comercial de visualização imersiva e colaborativa, resultado do desenvolvimento contínuo da ferramenta *iViz* mencionada na secção 2.4. A Figura 2-17 demonstra a vista inicial da aplicação após login, a Figura 2-18 demonstra uma etapa no carregamento de um ficheiro (que ocorreu com sucesso ainda que o computador utilizado não cumprisse com as especificações mínimas) e na Figura 2-19, extraída do sítio da aplicação (Virtualitics – Documentação, 2020), demonstra um gráfico de linhas gerado com essa aplicação.



Figura 2-17. Vista inicial do Virtualitics.

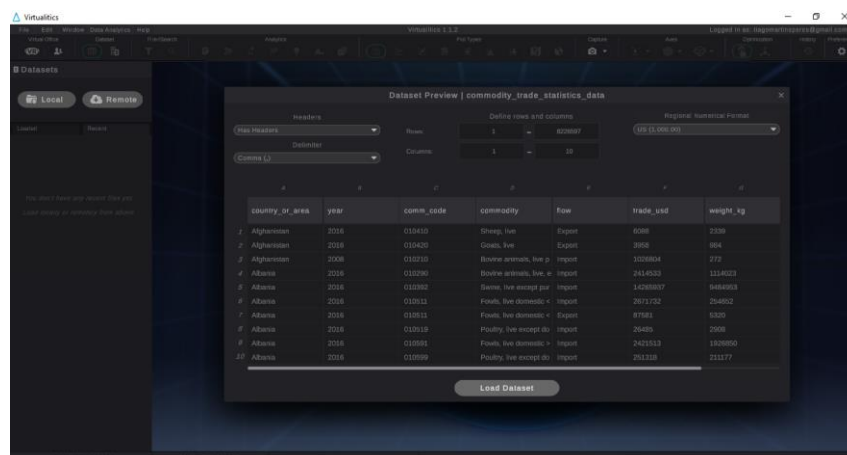


Figura 2-18. Carregar grandes quantidades de dados no Virtualitics.

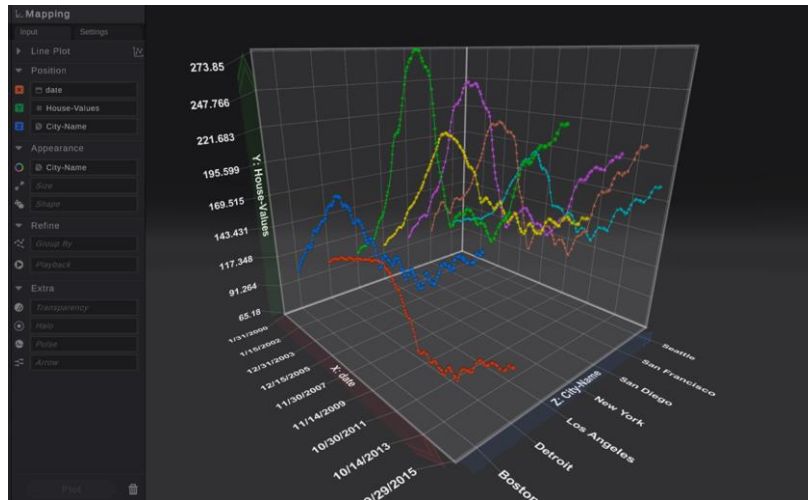


Figura 2-19. Gráfico de linhas da aplicação Virtualitics.

VR Ulysses

VR Ulysses (sem data) é uma empresa que cria soluções para visualização de dados em realidade virtual. Tal como BadVR, não está claro que tipo de aplicações produzem, mas ficamos com a impressão através do Twitter (VR Ulysses – Twitter, 2018) que trabalham apenas com o HMD HTC Vive (Figura 2-20).

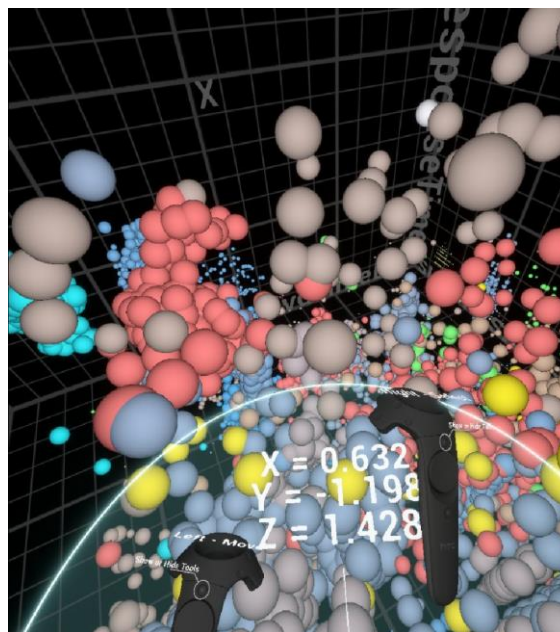


Figura 2-20. Screenshot do que se pode fazer com VR Ulysses.

Virtual Cove

Virtual Cove são os criadores da plataforma *Immersion Analytics* (2019) surgiu através de inspiração num artigo de 2003 de Schneiderman, conforme indicado pelo CEO Bob Levy (2016). Na Figura 2-21 podemos ver o screenshot de algo que podemos fazer com a plataforma.

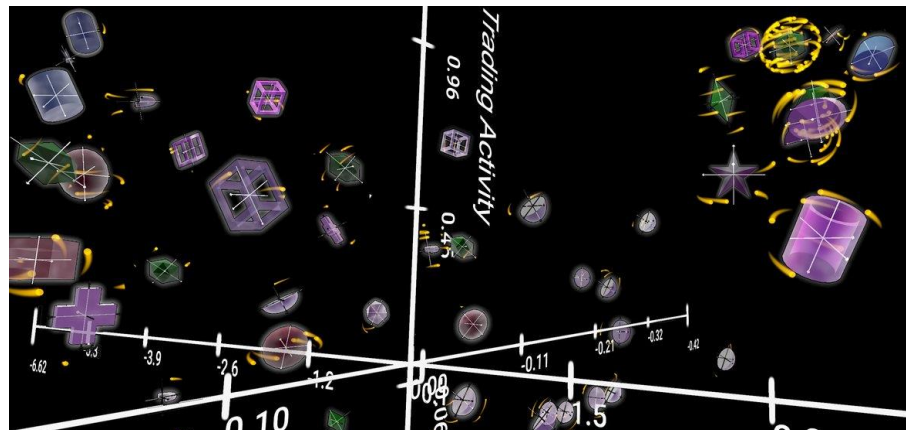


Figura 2-21. Screenshot do que se pode fazer com a solução.

Nanome

Nanome (2020) é uma empresa que cria soluções para visualização de dados em realidade virtual. A mesma é responsável por Nanome (Figura 2-22), que permite visualizar proteínas, moléculas e átomos, e CalcFlow para modelos matemáticos (Figura 2-23).

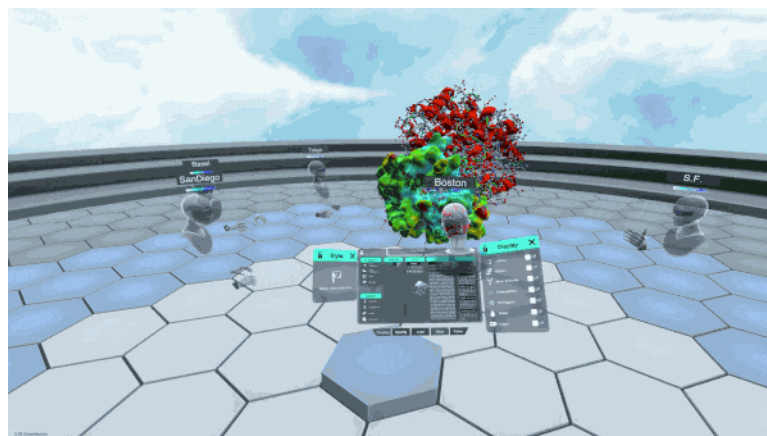


Figura 2-22. Nanome.

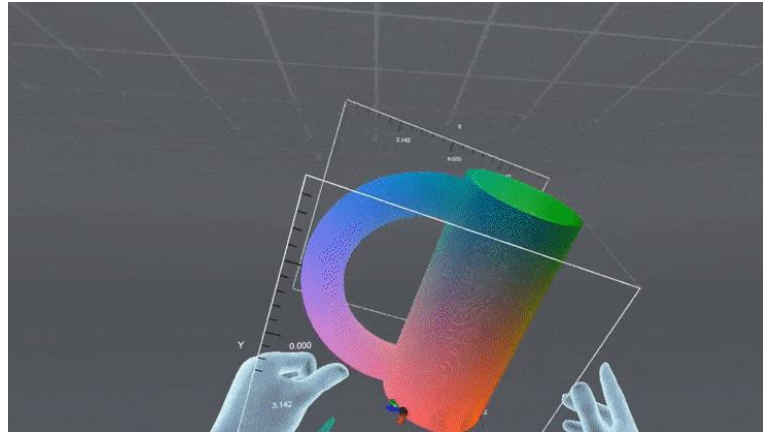


Figura 2-23. CalcFlow.

Datavized

Datavized (2018) é uma empresa que cria soluções para visualização de dados em realidade virtual. A mesma é responsável por Geometric Space (2018), que permite visualizar mapas 3D (Figura 2-24).

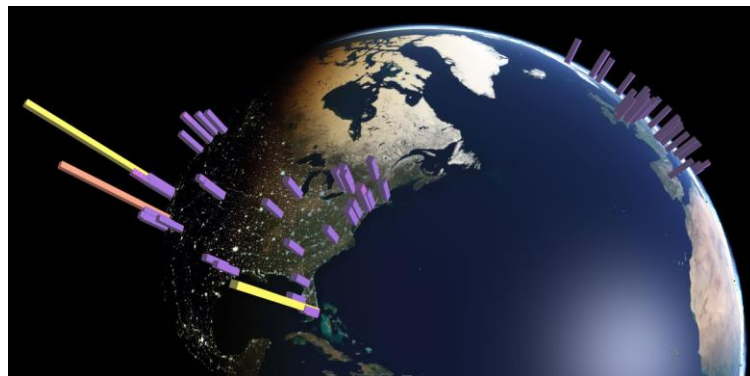


Figura 2-24. Mapa 3D na aplicação Geometric Space.

Deloitte

Deloitte (2020) é uma empresa que presta serviços a nível mundial. Esta empresa criou a plataforma Dali (sem data) que permite criar visualizações de dados em ambientes imersivos (Figura 2-25).



Figura 2-25. Visualização criada com Dali.

Tabela 2-1. Comparação das aplicações (realizada dia 16 de novembro de 2019).

Aplicação	Limitações
3Data	Requer a existência de um mundo.
BadVR	Foco no headset Magic Leap. Não está claro que outros tipos de soluções produzem.
DataView VR	Visualizações apenas para hardware topo de gama, como o Oculus Rift e o HTC Vive.
Virtualitics	Visualizações apenas para hardware topo de gama, como o Oculus Rift e o HTC Vive.
VR Ulysses	Foco no headset HTC Vive. Não está claro que outros tipos de soluções produzem.
Virtual Cove	As limitações em termos de hardware e ambiente não estão claras.
Nanome	Visualizações apenas para hardware topo de gama, como o Oculus Rift e o HTC Vive.
Datavized	Visualizações apenas de mapas 3D.
Deloitte	Visualizações apenas para hardware topo de gama, como o Oculus Rift e o HTC Vive.

2.3.2. HMDs

Cardboard

O *Cardboard* é a plataforma de realidade virtual (RV) mais acessível do mundo, compatível com Android e iOS. O *Cardboard* permite experiências de RV imersivas ao fundir, os dados dos sensores do telemóvel para prever a posição da cabeça do utilizador nos mundos real e virtual. Isso, combinado com o *headset* simples e fácil de usar, torna-o ideal para experiências rápidas de RV. O mesmo suporta dispositivos Android e iOS, tem visualizadores acessíveis que suportam *input* de um botão e interface do utilizador baseada no olhar e ainda disponibiliza SDKs que facilitam a criação de experiências de RV no ambiente de desenvolvimento de sua escolha (*Cardboard*, sem data). No entanto, o uso do *Cardboard* (Figura 2-26) tem como pontos negativos a bateria ter um tempo de vida limitada, os gráficos processam mais lentamente, a precisão dos sensores de movimento é das menores do mercado de HMDs e não permite rastreamento posicional, somente a orientação da cabeça é rastreada (3 graus de liberdade) (Linowes & Schoen, 2016).



Figura 2-26. Google Cardboard.

Dada a simplicidade, várias foram as empresas que conseguiram reproduzir produtos similares com ampla compatibilidade, designs confortáveis e inovadores, de configuração fácil e baixo custo, como é o caso apresentado na Figura 2-27.



Figura 2-27. Cardboard da marca *Splaks* (sem data).

Gear VR

O Gear VR (Figura 2-28) é um *headset* mais robusto do que o *Cardboard*, mas apenas funciona em modelos específicos de *smartphones* Samsung; no entanto, o mesmo tem um sensor de movimento que permite aumentar a precisão do rastreamento do movimento da cabeça e ajuda a reduzir a latência ao atualizar o ecrã (GearVR, sem data).



Figura 2-28. Gear VR.

Daydream View

O *Daydream View* é uma plataforma de RV móvel que oferece RV de alta qualidade e preço acessível na escala do *Android*. O *Daydream View* permite experiências imersivas usando técnicas avançadas para prever a orientação da cabeça do utilizador, aumentar o conforto e reduzir a latência. Em vez de plástico (*Gear VR*) ou cartão (*Cardboard*), o *Daydream* (Figura 2-29) é construído a partir de pano e é empacotado com um controlador de movimento com um *trackpad* e botões, o que o torna excelente para experiências de RV mais longas e mais imersivas (Daydream, sem data). Este tem uma ótica superior em comparação com um *Cardboard*, mas assim como no *Gear VR*, funciona apenas com uma lista muito específica de *smartphones*.



Figura 2-29. Daydream View.

Bridge

O *Bridge* (Figura 2-30) tem uma estrutura de alumínio combinada com um sensor de estrutura occipital (um dispositivo de digitalização 3D móvel) embutido na parte frontal permitindo uma experiência mais imersiva, podendo detectar e criar mapas de profundidade de objetos (Bridge by Occipital, 2020).



Figura 2-30. Bridge by Occipital.

Bridge (Polycarbonate)

O *Bridge (policarbonato)* (Figura 2-31) é uma versão de baixo custo do HMD *Bridge*, em que a estrutura de alumínio é substituída por policarbonato. Este não inclui o sensor de estrutura pois foi projetado para experiências de VR apenas de rotação (3-DoF). É compatível com todos os aplicativos baseados no Cardboard.



Figura 2-31. Bridge (Polycarbonate) by Occipital.

Merge VR

O *Merge VR* (Figura 2-32) permite transformar o *smartphone* iOS ou Android num ambiente imersivo sem necessidade de controlador. Este é feito de espuma macia e leve com lentes ajustáveis, para se adaptar a praticamente qualquer rosto (Marge VR, 2020).



Figura 2-32. Merge VR.

Tabela 2-2. Comparação dos HMDs (realizada dia 8 de outubro de 2019).

HMD	Controlador	Degrees of Freedom (DoF)	Suporta	Material	Peso	Custo
Cardboard da marca Splaks	Não, mas tem um botão no HMD	3	Android de versão 4.1 ou superior. iOS de versão 8.0 ou superior.	Cartão	0,181 kg	13,40 €
Gear VR	Sim	3	Galaxy Note9*, S9, S9+, Note8, S8, S8+, S7, S7 edge, Note5, S6 edge+, S6, S6 edge, A8 Star, A8, A8+. *Galaxy Note9 é compatível apenas com o Gear VR cujo número de modelo é SM-R325NZVC.	Plástico	0,345 kg	115,16 €
Daydream View	Sim	3	<i>Telemóveis Daydream-ready</i> (sem data).	Pano	0,218 kg	53,56 €
Bridge	Sim	6	<i>iPhone 6s, 7, 8, X e XS.</i>	Alumínio	Indeterminado	363,44€
Bridge (Polycarbonate)	Sim	3	<i>iPhone X e XS.</i>	Polycarbonato	Indeterminado	126,56€
Merge VR	Não, mas tem dois botões no HMD	3	<i>Android e iOS de dimensões de 123mm a 158mm.</i>	Espuma	0,340 kg	36,43€

2.3.3. Ambientes de desenvolvimento e respetivas bibliotecas

Unity

A plataforma mais comum para desenvolver ambientes imersivos virtuais é a Unity (2020), um motor de jogo que suporta construção de aplicações numa variedade de plataformas: desde PlayStation a Xbox, de Windows a Mac (e Linux), de Android a iOS (Linowes & Schoen, 2016). Para além disso, existe uma grande comunidade e uma gama de *assets* disponíveis, como modelos 3D e bibliotecas de scripts.

No Unity, os aplicativos são representados como cenas 3D onde é possível adicionar e manipular *GameObjects*, que encapsulam objetos e o seu comportamento. Exemplos de *GameObjects* incluem câmaras, modelos 3D, luzes, efeitos e assim por diante. Os *GameObjects* são organizados em hierarquias pai-filho ou gráficos de cena (Sicat *et al.*, 2018). Estes podem ser guardados como *prefabs* que servem como modelos compartilháveis e reutilizáveis. A Unity possui uma *Asset Store* (2020) online para compartilhar cenas reutilizáveis, *prefabs* e outros ativos. Um script é um código C# ou JavaScript que pode ser anexado a *GameObjects* como componentes e usado para manipular programaticamente *GameObjects* em tempo de execução. É possível editar cenas e *GameObjects* interativamente usando a interface de utilizador do Unity Editor, ou programaticamente usando scripts através da API de *scripting* Unity (2019). A cena pode ser configurada para ser executada em Android, simplesmente especificando o dispositivo de destino nas configurações de implantação do Unity. Em tempo de execução, ou seja, quando a cena é reproduzida, o utilizador pode ver a cena através da exibição no dispositivo e interagir com a mesma (Sicat *et al.*, 2018).

Linowes (2018) criou uma tabela que foi adaptada (Tabela 2-3) para as nossas necessidades onde demonstra a combinação entre os HMD, SDKs de realidade virtual (RV) e pacote Unity que permitem construir soluções a ser utilizadas em *smartphones* Android. Para mais informações, encaminhamos o leitor para a documentação completa do Unity (2019).

Tabela 2-3. Combinação entre HMD, SDKs de realidade virtual e pacote Unity para soluções em telemóveis.

HMD	RV SDK	Pacote Unity
Gear VR	Oculus	Oculus Integration.
Daydream	Daydream	Google VR SDK for Unity and Daydream Elements.
Cardboard	Cardboard	Google VR SDK for Unity.

Podemos encontrar ainda bibliotecas que agilizam a criação de gráficos 3D, destacando-se as apresentadas na Figura 2-33.

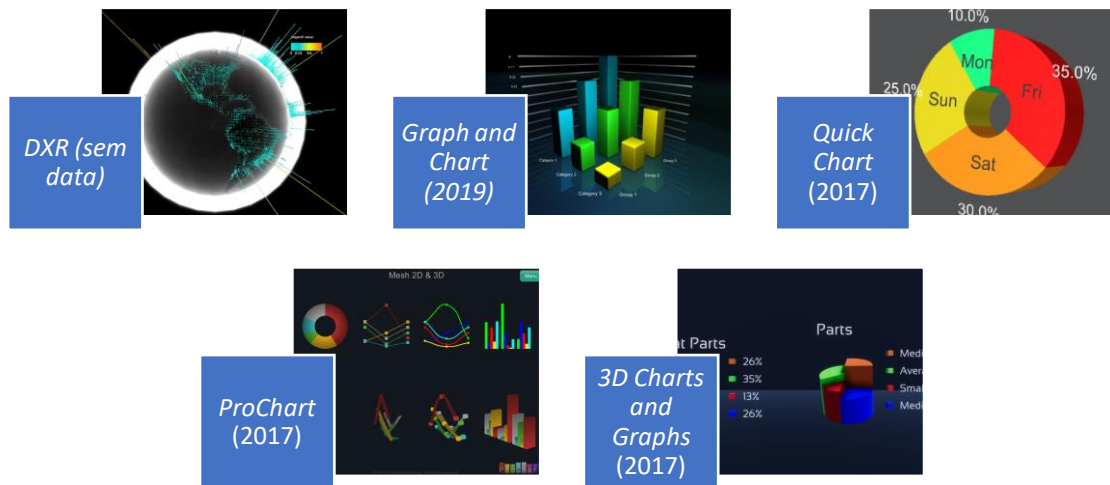


Figura 2-33. Bibliotecas para criação de gráficos 3D com Unity.

Tabela 2-4. Comparação das bibliotecas para o Unity (realizada dia 25 de abril de 2019).

Bibliotecas	Documentação	Dados	Tipos de Gráficos	Custo	Comentários
DXR	Sim	JSON	Mais de 18	Gratuito	Destinado a prototipagem e projetos de exploração com foco em <i>headsets</i> topo de gama (Microsoft HoloLens, ACER VR, etc). Não é possível de fazer download através do site e o código presente no GitHub gera muitos erros no console que tornam DXR impossível de ser utilizado sem resolver limitações no código.
Graph and Chart	Sim	XML, JSON ou inserção manual	11	62,53 €	Só o “3D Bar” se justifica ser 3D. Percebia-se melhor a informação nos outros gráficos com duas dimensões e cores diferentes. Foco na criação de visualizações com dados de duas dimensões. Excelente suporte.
Quick Chart	Sim	Inserção manual	3	71,47 €	Foco na criação de visualizações com dados de duas dimensões.
ProChart	Sim	Inserção manual	3	26,80 €	Foco na criação de visualizações com dados de duas dimensões.
3D Charts and Graphs	Não	Inserção manual	4	33,05 €	Foco na criação de visualizações com dados de duas dimensões.

Android Studio

O Android Studio é o ambiente oficial de desenvolvimento integrado para Android, desenvolvido com base no software IntelliJ IDEA da JetBrains e projetado especificamente para o desenvolvimento do Android. Este está disponível para download nos sistemas operativos Windows, MacOS e Linux.

O Android Studio integra um número de diferentes ferramentas e processos e o resultado do desenvolvimento de uma aplicação android é um ficheiro .apk que pode ser distribuído a utilizadores pela Google Play Store (Linowes & Schoen, 2016).

Ao utilizar o Cardboard SDK para Android é possível programar em Java usando este ambiente integrado de desenvolvimento (IDE). Segundo Linowes e Schoen (2016), aplicações para Cardboard construídas com Unity são volumosas comparativamente com as desenvolvidas com Android Studio, impactando a utilização da memória, da bateria e requerendo mais tempo a carregar. Para além disso, ao não pagar para uma licença Unity Android, a aplicação começa sempre com o ecrã “Made With Unity”.

Podemos encontrar ainda bibliotecas para o Android Studio que agilizam a criação de gráficos 3D, tal como demonstra a Figura 2-34.



Figura 2-34. Bibliotecas para criação de gráficos 3D com Android Studio.

Tabela 2-5. Comparação das bibliotecas para o Android Studio (realizada dia 25 de abril de 2019).

Bibliotecas	Documentação	Dados	Tipos de Gráficos	Custo	Comentários
Orson Charts For Android	Sim	Inserção manual	7	173,63 €	Foco na criação de visualizações 3D.
TeeChart Java Android	Sim	Inserção manual	50	302,88 €	Foco na utilização em diferentes âmbitos.
NChart3D	Sim	Inserção manual	31	455,58 €	Foco na flexibilidade para visualizar vários dados.

Unreal

O Unreal está entre os motores de jogos mais completos e avançados do mercado. Este foi criado para desenvolver jogos, mas tornou-se viral entre as comunidades de arquitetura, ciência, engenharia e visualização médica. Tal como Unity, o Unreal possui um *marketplace* (2020) para compartilhar cenas reutilizáveis, *prefabs* e outros ativos. No entanto, ainda que tenhamos conhecimento de que já foram criados projetos de visualização de grandes quantidades de dados utilizando o Unreal (Figura 2-35), não encontramos no *marketplace* bibliotecas que agilizam a criação de gráficos 3D. Essa visualização de dados feita na hackathon Unreal Big Data da HammerHeadVR (2018) usa esferas que ficam verdes quando estamos a pairar sobre elas por meio de um controlador RV (Mohan, Lukas & Pangilinan, 2019).

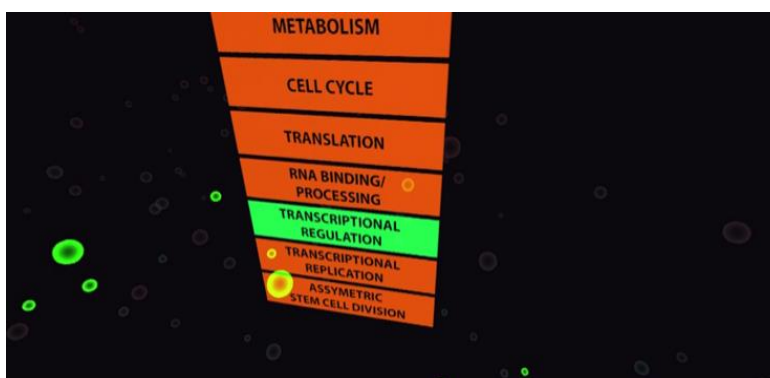


Figura 2-35. Visualização de dados feita na hackathon *Unreal Big Data* da HammerHeadVR.

Tabela 2-6. Comparação dos ambientes de desenvolvimento (realizada dia 25 de abril de 2019).

Ambientes de desenvolvimento	PROS	CONTRAS
Unity3D	Bibliotecas mais baratas. Permite construir soluções para PC, Mac, Linux, iOS, Android, etc.	Bibliotecas maioritariamente focadas na criação de visualizações com dados de duas dimensões.
Android Studio	Bibliotecas com mais funcionalidades.	Bibliotecas mais caras. Permite criar soluções apenas para Android.
Unreal	Permite construir soluções para PC, Mac, Linux, iOS, Android, etc.	Não tem bibliotecas para o fim desejado.

Biblioteca independente (VR-VIZ)

Nesta parte apresentamos uma outra biblioteca não limitada pelos ambientes de desenvolvimento e cujo resultado pode ser visualizado através de um navegador. VR-VIZ (2019), criada por Mustafee Saifee, combina o framework A-Frame (sem data) com as bibliotecas de JavaScript React (2020) (para manipulação de DOM) e D3.js (2019) (para visualizações de dados) para gerar visualizações imersivas (Figura 2-36). De acesso gratuito, VR-VIZ permite criar 16 tipos de gráficos diferentes com dados vindos de

JSON, CSV (com ou sem cabeçalho) e PLY. Ainda que esta biblioteca seja funcional em termos de criação de diferentes tipos de visualizações para dados de três dimensões, a criação de outros objetos 3D (como menus) e ações para o utilizador (como técnicas de locomoção e interações possíveis com o HMD) não são implementadas de forma tão direta como numa plataforma de desenvolvimento de jogos como Unity ou Unreal (que não suportam a mesma).

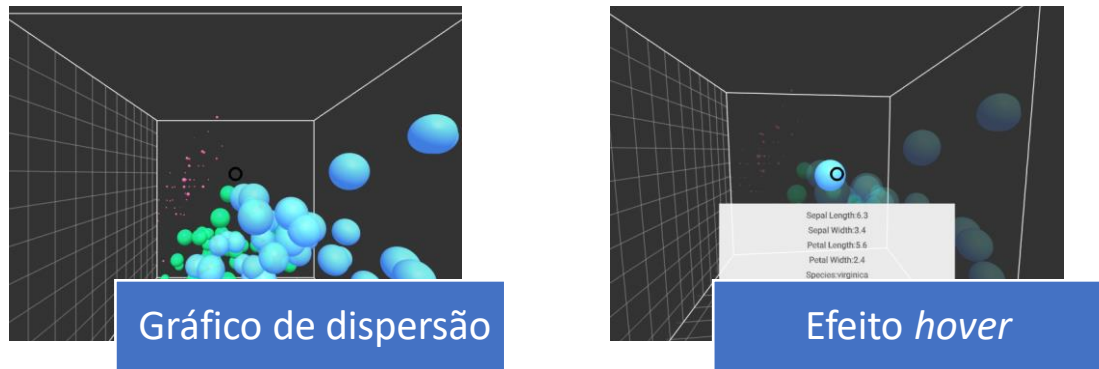


Figura 2-36. Criado com a ferramenta VR-VIZ.

2.4. Visualização de *big data*

Na maioria dos casos, os utilizadores recebem uma quantidade substancial de controlo, portanto, um alto nível de interatividade com as visualizações. Uma grande dificuldade cognitiva em observar uma visualização muito grande é a necessidade de lembrar o que foi observado há pouco. Em geral, é esperado que um ecrã megapixel requeira mais interação, um ecrã gigapixel requeira mais movimento do corpo (Chen *et al.*, 2017). No caso do HMD, como o Cardboard, teremos não tanta interação quanto num ecrã megapixel e não tanto movimento do corpo quanto num ecrã gigapixel.

O tratamento de *big data* requer vários problemas extras, em comparação com os casos gerais de processamento e análise de visualização. De várias fontes que geram dados não estruturados, precisamos de integrar dados considerando diferentes recursos de fontes de dados. Também é necessário ter uma etapa de pré-processamento para manter a qualidade dos dados sem comprometer o caráter inerente dos dados. O resultado da visualização não pode ser produzido se os processos de aquisição, análise e refinamento dos dados não forem executados corretamente (Cho *et al.*, 2014). Para isso, Fry e Maeda (2004) definem sete etapas: aquisição, análise, filtração, mineração, representação, refinamento e interação, conforme mostrado na Tabela 2-7 (Cho *et al.*, 2014).

Tabela 2-7. Etapas inerentes ao processo de visualização [26, 27].

Etapa	Descrição
Aquisição	Obter os dados (arquivo, disco, através da rede).
Análise	Fornecer alguma estrutura para o significado dos dados e ordenar em categorias.
Filtração	Remover tudo, exceto os dados de interesse.
Mineração	Aplicar métodos de estatística ou mineração de dados como forma de discernir padrões ou colocar os dados no contexto matemático.
Representação	Escolher um modelo visual básico, como um gráfico de barras, lista, árvore etc.
Refinamento	Melhorar a representação básica para ser mais clara e mais apelativa visualmente.
Interação	Adicionar métodos para manipular ou controlar quais recursos são visíveis.

De uma forma mais sintética, podemos agrupar certas etapas anteriormente existentes num *pipeline* de criação de visualização de dados com três passos, começando com o pré-processamento de dados, passando pelo passo “Extrair, Transformar e Carregar” (ETL), no qual nós ingerimos dados e os transformamos e convertemos nos formatos apropriados para visualização (Figura 2-37). O trabalho de transformação no ETL (Figura 2-38) ocorre num mecanismo especializado e geralmente envolve o uso de tabelas para manter temporariamente os dados à medida que eles são transformados e

finalmente carregados no destino. Para além disso, geralmente, as três fases do ETL são executadas em paralelo para economizar tempo.

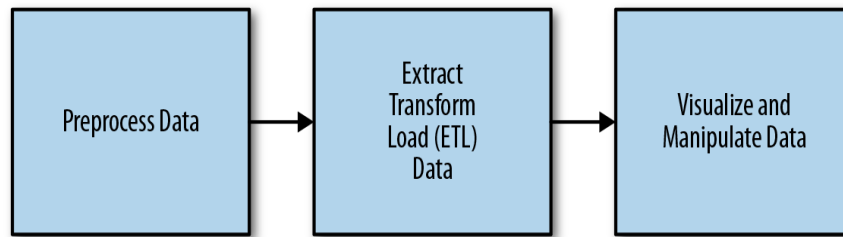


Figura 2-37. Pipeline de visualização de dados adaptado (Mohan, Lukas & Pangilinan, 2019).

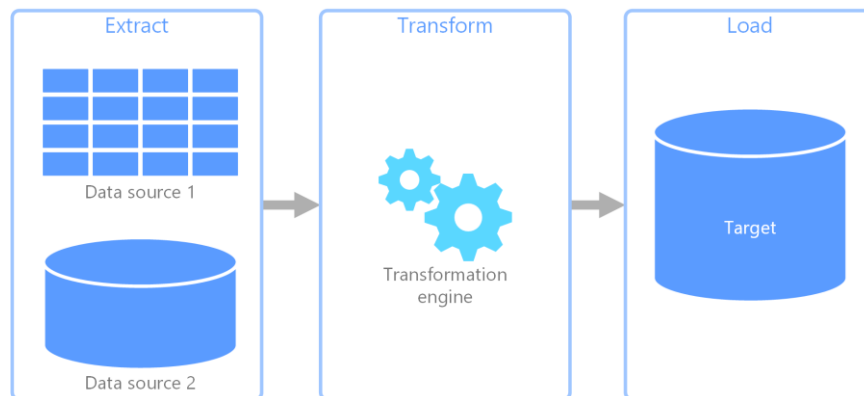


Figura 2-38. ETL (Microsoft – Documentação, 2019).

De notar ainda que a interatividade é crucial para a construção de ferramentas de visualização que lidam com a complexidade. Quando os conjuntos de dados são grandes o suficiente, as limitações de pessoas e exibições impedem apenas mostrar tudo de uma vez; interação onde as ações do utilizador fazem com que a visão mude é o caminho a seguir. Além disso, uma única exibição estática pode mostrar apenas um aspecto de um conjunto de dados. Para algumas combinações de conjuntos de dados e tarefas simples, o utilizador pode precisar apenas de uma única codificação visual. Em contraste, um ecrã que muda de forma interativa suporta muitas consultas possíveis (Munzner, 2015).

2.4.1. Visualização de big data em ambientes imersivos

Nesta parte recorreremos ao Dimensions (onde obtemos dados para mapeamento bibliométrico) e à ferramenta VOSViewer para perceber a relação entre diversos artigos criados entre 2011 e 2019 com termos *Immersive Environments* e *Big Data* (Figura 2-39).

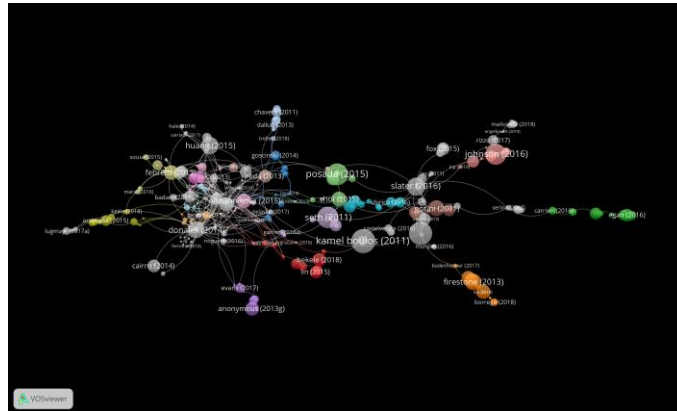


Figura 2-39. Rede criada no VOSViewer com dados provenientes do Dimensions.

O artigo denominado *Immersive and Collaborative Data Visualization Using Virtual Reality Platforms* (Donalek *et al.*, 2014) encontra-se entre os mais citados que se enquadra no tópico tratado neste trabalho. Este artigo insere-se, como os próprios autores indicam, na categoria de um relatório sobre a evolução de trabalho exploratório.

O foco central está na apresentação de exemplos, de conceções novas, para a visualização de dados *big data* em ambientes imersivos, permitindo-nos, ter uma noção do que está a ser feito pelas equipas que criaram as ferramentas. De forma mais concreta, para além da paisagem marciana, neste artigo podemos ler que os autores são pioneiros no desenvolvimento de uma ferramenta de visualização de dados de nome *iViz*, criada através a plataforma Unity 3D™, e que surgiu do progresso do investigador Ciro Donalek e outros após a experiência com o Meta Institute for Computational Astrophysics (MICA) (Donalek *et al.*, 2013) (onde realidade virtual foi utilizada para analisar dados de astrofísica num espaço 3D (Figura 2-40)). Segundo os autores do artigo, já é possível visualizar de forma fácil e rápida 105 a 106 pontos de dados (Figura. 2-41), o que é comparável a outras ferramentas de visualização de última geração.

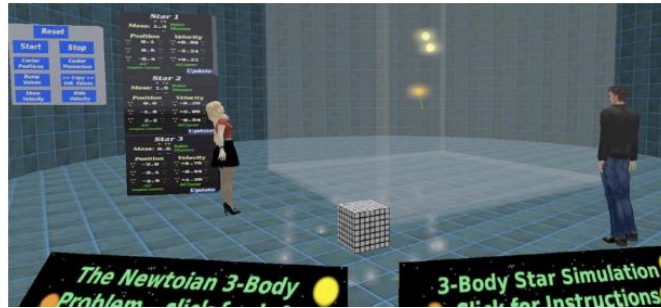


Figura 2-40. Uma demonstração de trabalho de um problema gravitacional de 3 corpos, implantado no MICA com *Second Life* (2020).

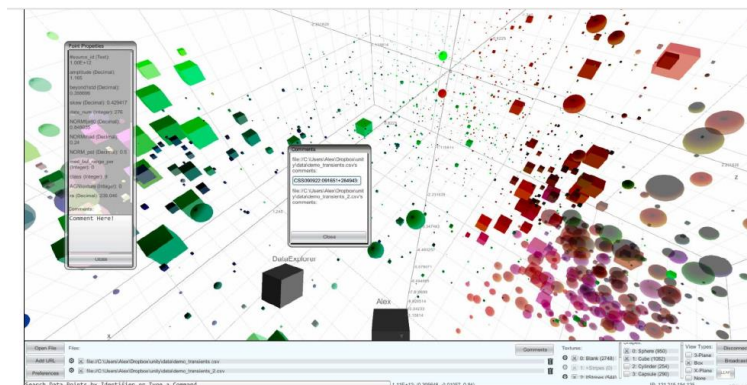


Figura 2-41. Um exemplo do interface do utilizador na ferramenta *iViz*.

Na *iViz* existe um novo recurso para uma exploração de dados visuais colaborativa e multiutilizador. Além do modo padrão de cada utilizador ter seu próprio ponto de vista e navegar pelos dados de forma independente, adicionaram uma função de transmissão onde os utilizadores podem ter uma visão compartilhada associada a um utilizador que então navega pelos dados, o que permite uma exploração de dados colaborativa mais flexível.

Estes projetos, assim como o artigo *Improving Big Data visual analytics with interactive virtual reality* (Moran *et al.*, 2015) em que se focaram na visualização e interação de dados para melhorar a análise visual de *big data* (Figura 2-42), revelam o potencial adicional de como a utilização de ambientes imersivos pode trazer uma experiência visual mais eficaz.

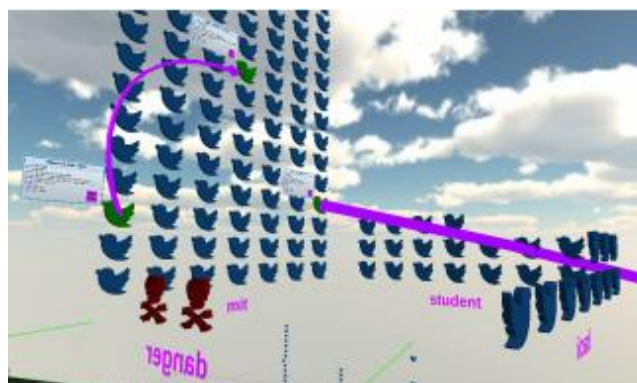


Figura 2-42. Visualização de tweets num modelo 3D do campus do MIT (Moran *et al.*, 2015).

2.5. Síntese

A grande quantidade de dados gerados hoje em dia está a ser usada para criar conhecimento e facilitar na tomada de decisões. No entanto, esta situação conduz a um novo desafio: *como visualizar todos esses dados sem perder informações cruciais a médio / longo prazo?* Para além disso, este processo, que começa a partir de dados para obter informações e, a partir daqui, para alcançar o conhecimento, não se concentra apenas no tratamento dos dados: há que pensar também na maneira correta para que o utilizador entenda o que está a ser exibido e não ficar surpreendido com visualizações complexas e difíceis de entender. Assim, assume-se que em *InfoVis*, as interações do utilizador são tão importantes quanto a apresentação para uma compreensão e análise efetivas da informação.

Tendo por base dados do tipo *big data* e nos 5Vs, a escalabilidade e o dinamismo das representações gráficas podem ser considerados dois dos principais desafios da visualização de dados. No entanto, existem várias funcionalidades que as ferramentas de visualização estão a tentar fornecer agora para representar dados, não apenas para mostrar informação, mas também para obter valor com isso: importar dados de diferentes fontes, combinar a representação visual com rótulos de texto, obter informação e representações de formulário de conhecimento, representar grandes volumes de dados, integrar a mineração de dados, permitir o trabalho colaborativo, facilitar a utilização, atuação e cognição limitada.

Os desafios de *big data* em termos de taxas de dados e volumes são bem conhecidos. Independentemente da escolha (localmente, na nuvem ou híbrida), o armazenamento e processamento de *big data* serão necessariamente distribuídos. Ainda mais interessante do que o crescimento dos volumes de dados é o aumento dramático na complexidade dos dados: conjuntos de dados multidimensionais, muitas vezes, combinam medições numéricas, imagens, espectros, séries temporais, rótulos categóricos, texto, etc. O ponto-chave é que a ciência de *big data* não é sobre dados: é sobre descoberta e compreensão de padrões significativos escondidos nos dados; esta idéia é, de forma subtil, transmitida nas cores Figura 2-1, onde vemos que só ter grandes volumes de dados não traz real valor (vermelho) a não ser que possamos extrair valor dos mesmos (verde).

É inegável a relevância da visualização da informação para extrair conhecimento dos dados. É também importante observar que erros, vieses e outros problemas inesperados geralmente levam a dados que devem ser tratados com cuidado. A falha em descobrir esses problemas geralmente leva a análises erradas e falsas descobertas. Ainda que o tratamento de dados do tipo *big data* implique problemas novos em comparação com os casos gerais de processamento e análise de visualização, existe um *pipeline* (Figura 2-37) pelo qual podemos guiar-nos.

Quando grandes conjuntos de dados devem ser representados, é comum usar representações gráficas em 2D. Em 2D podemos ter várias vistas complementares sob os dados e mostrar as diversas informações que eles contêm. No 3D, a grande mais valia é dispor de uma terceira dimensão para mapear os dados, daí, reduzir o número de vistas que teremos de gerar para mostrar todas a informação que os dados comportam. Com

isto, queremos dizer que podemos mostrar a mesma quantidade de informação num caso e no outro, só que exigirá um número de vistas diferente, por haver uma dimensão a mais disponível. Com isso também se cria a possibilidade de visualização imersiva, que pode auxiliar a perceber melhor a informação que os dados contêm.

Evidências empíricas parecem sugerir que simplesmente aumentar uma interface de 2D para 3D não seria suficiente para aumentar o desempenho da tarefa, a menos que funções adicionais sejam fornecidas para que os utilizadores possam ter controlos maiores de objetos em interfaces 3D. Com a utilização de ambientes imersivos virtuais móveis podemos tirar partido, inclusive, da funcionalidade “fly-through”. Todavia, estes ambientes podem ser desconfortáveis para o utilizador, podendo até comprometer a cognição.

A utilização de ambientes imersivos virtuais mostrou que leva a uma melhor descoberta em domínios que cujas dimensões primárias são espaciais. Os primeiros trabalhos demonstraram que a imersão ajuda os cientistas a investigar mais efetivamente a paleontologia, os tumores cerebrais, a perceção da forma, as estruturas subterrâneas de análise das cavernas, a ressonância magnética, a química orgânica e a física. A exibição de dados mostrou suportar análises multidimensionais altamente abstratas. Muitos investigadores procuram a visualização para apoiar a exploração de grandes conjuntos de dados até porque as visualizações de dados transformam os dados em codificações visuais que ajudam os utilizadores a educar, comunicar, fornecer informações e explorar melhor os dados (Viegas & Wattenberg, 2018).

Empresários, cientistas e curiosos procuram novos métodos para facilitar grandes conjuntos de dados. Existem vários projetos a olhar para o futuro investindo em auxiliar visualizações imersivas de dados. No entanto, as aplicações que existem atualmente no mercado são de empresas com um modelo de negócios B2B (*business to business*, o que logicamente exclui possíveis utilizadores pelo simples facto de não terem um negócio). Para além disso, umas aplicações (como o *3Data*) não permitem visualizar ambientes imersivos virtuais (requer a existência de um mundo), outras (como o *DataView VR* e *Virtualitics*) trabalham apenas com hardware topo de gama, como o Oculus Rift e o HTC Vive. Considerando também as funcionalidades dos ambientes de desenvolvimento atuais, bibliotecas e HMDs, compreendemos que é possível criar soluções para visualizar dados de forma imersiva acessíveis de forma relativamente rápida (Figura 2-43).

Ainda que existam boas perspetivas no que toca à envolvência de visualização de informação e ambientes imersivos, é importante avaliar até que ponto a imersão e a interação com os dados resultam numa mais-valia.

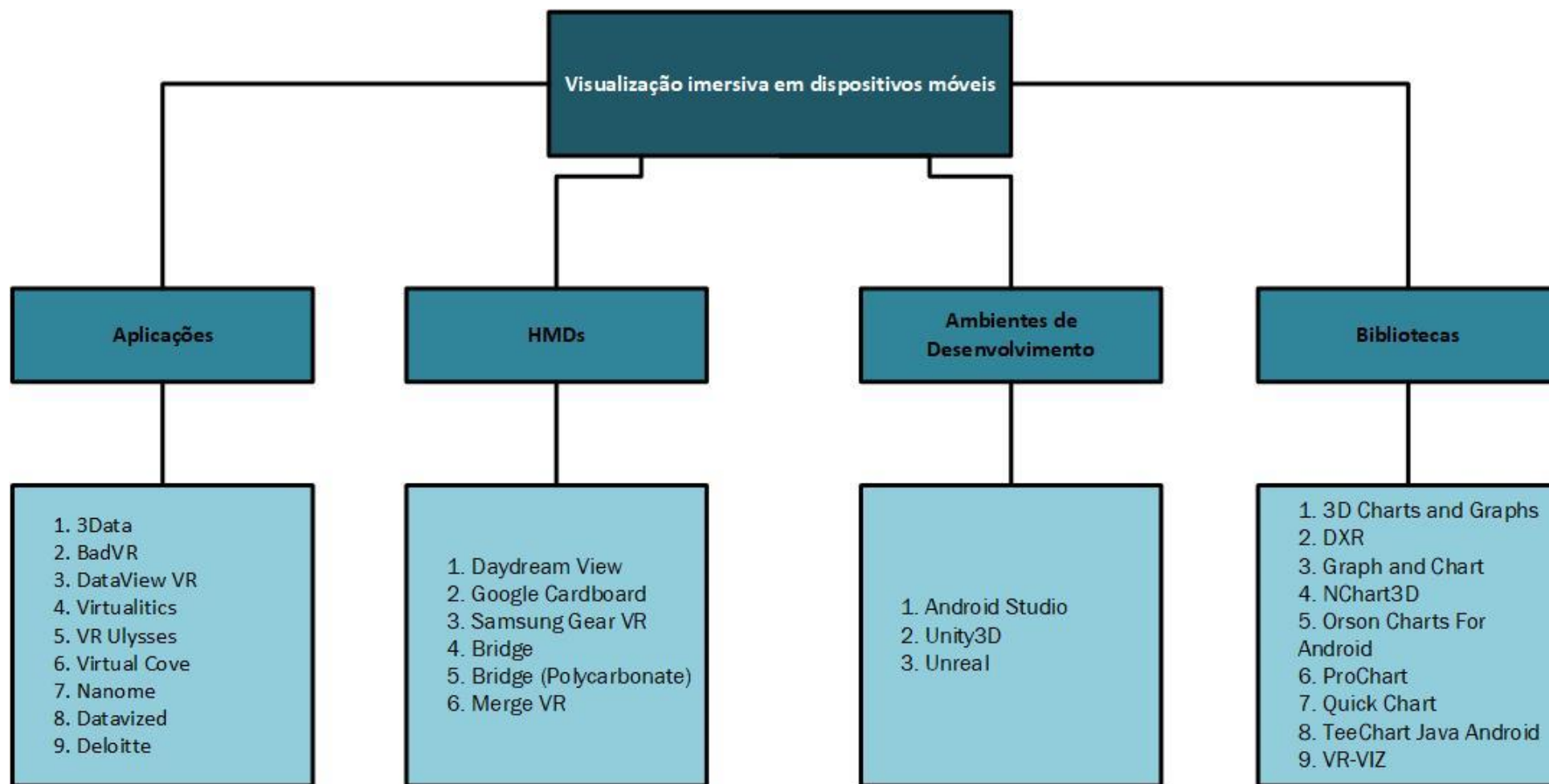


Figura 2-43. Visão geral da taxonomia de visualização imersiva em dispositivos móveis.

3. Modelo conceptual

Este capítulo apresenta o modelo conceptual proposto para esta dissertação (designado de "Modelo-Data-Room"), que permitirá a visualização de *big data* em forma imersiva móvel. Para tal, começaremos pela visão geral (secção 3.1.), seguimos para uma análise dos requisitos do modelo (secção 3.2.) e para o mapeamento visual dos dados (secção 3.3.). O capítulo termina com uma síntese (secção 3.4.), onde será feito um apanhado do que foi exibido com uma devida reflexão.

3.1. Visão geral

Como vimos ainda na secção 2.4., as soluções atuais são projetadas em torno de um *pipeline* de dados, descrevendo como os dados que fluem por meio de uma solução, onde são processados, onde são armazenados e como são consumidos pelo próximo componente no *pipeline*. Para além disso, na secção 2.1., compreendemos que uma arquitetura de *big data* é projetada para lidar com a ingestão, processamento e análise de dados que são muito grandes ou complexos para os sistemas de tradicionais.

Nesta secção apresentamos a visão alto-nível do modelo (Figura 3-1) conjuntamente com um *pipeline* de entrada de dados e resultados.

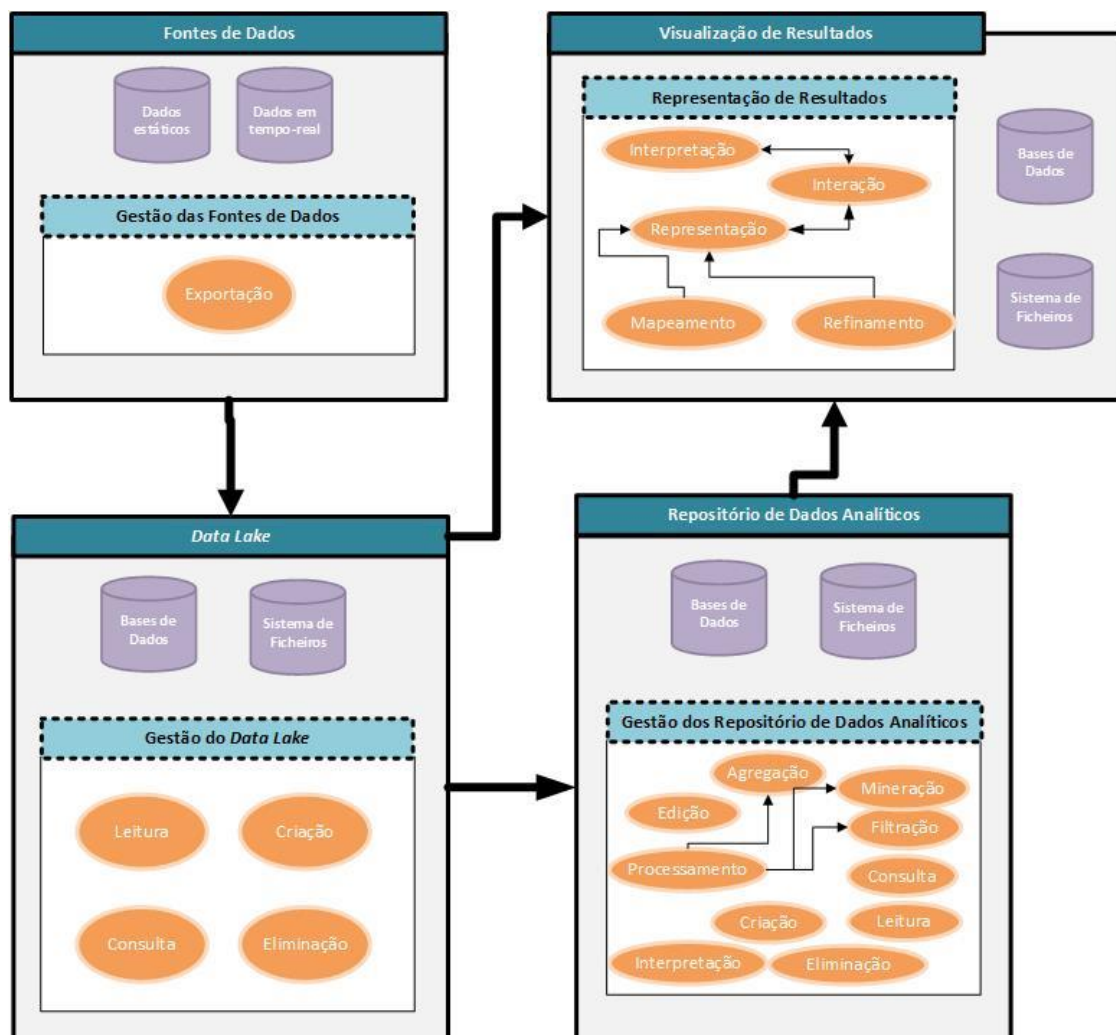


Figura 3-1. Arquitetura do modelo.

Como a Figura 3-1 demonstra, existem quatro locais distintos, nomeadamente Fontes de Dados, *Data Lake*, Repositório de Dados Analíticos e Visualização de Resultados. Como podemos visualizar, a única operação que nos interessa fazer com as Fontes de Dados é extrair os conjuntos de dados, sejam estes estáticos ou dinâmicos, para o *Data Lake*. De notar que consoante o local onde vamos buscar esses dados, poderemos ter, ou não, a possibilidade de especificar com mais ou menos detalhe o que pretendemos

extrair. No entanto, como cada local é diferente, optámos por referir somente extração como ação.

Encontrando-se uma vez no *Data Lake*, os conjuntos de dados podem ser lidos, consultados, eliminados e, inclusive, podem ser criados novos (por exemplo, com dados introduzidos manualmente). Daqui, estes poderão ser enviados para Visualização dos Resultados ou para o Repositório de Dados Analíticos.

Aquando no Repositório de Dados Analíticos, os dados podem ser editados, criados, lidos, eliminados, interpretados ou processados. Dentro do processamento, que pode ser agregação, mineração ou filtração, podemos recorrer a um ou mais computadores para o caso de ser necessário distribuir a computação devido a limites de memória ou processamento encontrados quando usamos só um. Também há espaço para criar novos dados e armazená-los, dados, esses, resultantes de processamento, por exemplo.

Por sua vez, quando em Visualização de Resultados, os dados são primeiramente mapeados para criação da representação visual. Esta pode sofrer refinamentos. Além disso, podemos interagir com a representação para que consigamos interpretar de forma mais eficaz.

Para a criação do *pipeline* (Figura 3-2) usamos o *Date State Reference Model* (DSRM), proposto por Chi (2002), e o *Linked Data Visualization Model* (LDVM), proposto por Brunetti, Auer e García (2012), como inspiração. Este *pipeline* tem início numa extremidade com fonte de dados e resulta na outra extremidade com a vista. Ainda sobre o mesmo, podemos encontrar os seguintes componentes lógicos:

1. *Fonte de dados* — considera-se uma ou mais fontes de dados. As fontes de dados podem ser, por exemplo, bases de dado relacionais ou ficheiros estáticos produzidos por aplicações.
2. *Armazenamento de dados* — recorre-se a um lugar para guardar grandes volumes de arquivos em vários formatos (*Data Lake*). Aqui todos os conjuntos de dados são conhecidos e têm um identificador.
3. *Armazenamento de dados a analisar* — armazenar os dados processados num formato estruturado que pode ser consultado utilizando ferramentas de visualização.
4. *Abstração visual* — informação que pode ser visualizada utilizando uma técnica de visualização.
5. *Vista* — resultado do processo apresentado ao utilizador.

Os dados são propagados pelo *pipeline* de um ponto para outro, aplicando quatro tipos de operadores:

1. *Extração* — recuperação de dados de potencial relevância de uma ou mais fontes e migração para um *data lake*.
2. *Processamento em lotes* — como os conjuntos de dados armazenados são muito grandes, devemos processar um ou mais arquivos de dados usando tarefas em lote de longa execução.
3. *Transformação visual* — condensar os dados num tamanho exibível e criar uma estrutura de dados adequada para visualizações específicas.
4. *Mapeamento visual* — processar as abstrações de visualização para obter uma representação visual.

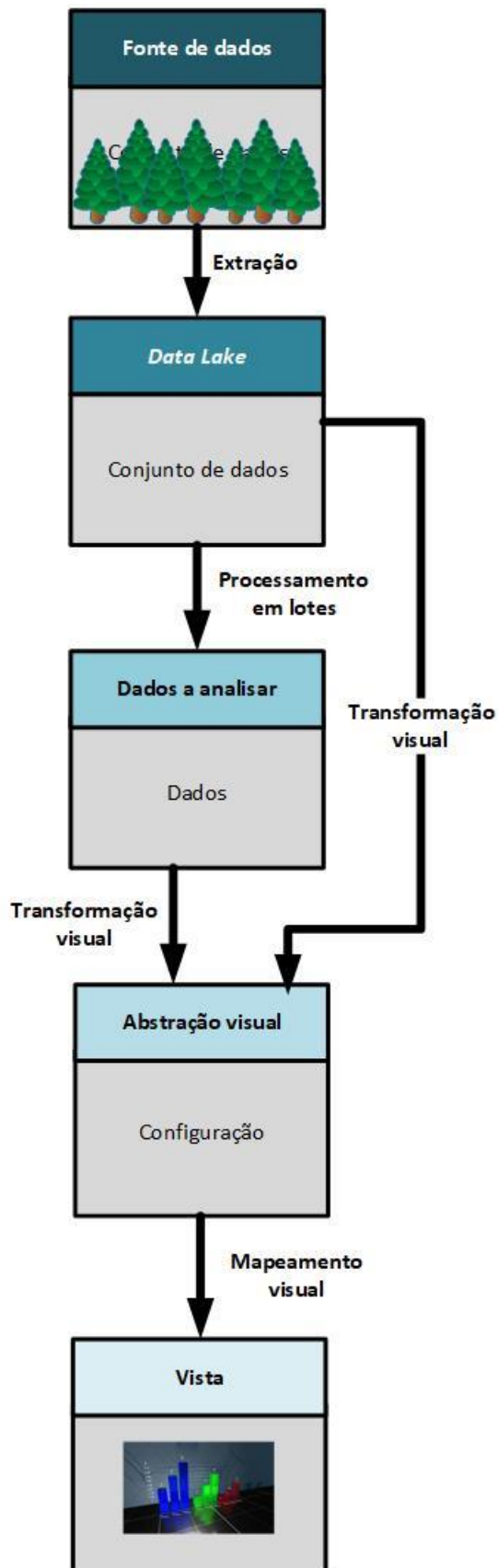


Figura 3-2. Pipeline de entrada de dados e resultados.

3.2. Requisitos do modelo

O modelo que propomos, de nome *Modelo-Data-Room*, permite visualizar *big data* de forma virtual imersiva móvel. Tal modelo pode poupar tempo às pessoas e pode aumentar a possibilidade de estas conseguirem perceber relações entre dados e obter resultados significativos.

Saber com antecedência quais os conjuntos de dados, técnicas de visualização e pistas visuais que as pessoas querem controlar possibilitariam fornecer uma variedade destes e reduzir o tempo de desenvolvimento.

Assume-se que o modelo proposto será composto por uma interface com o utilizador apropriada que estará disponível para visualizar dados de forma imersiva virtual móvel. Para além disso, prevê-se que o modelo terá limitações, nomeadamente um conjunto limitado de conjunto de dados, técnicas de visualização ou pistas visuais.

O modelo proposto inicia numa fonte de dados onde é feita a extração de conjuntos de dados potencialmente relevantes para serem colocados num *data lake*. A limitação da existência, assim como determinadas restrições no acesso a determinados conjuntos de dados, introduz, necessariamente, potenciais riscos para o utilizador. Semelhante raciocínio aplica-se ao mapeamento visual, nomeadamente limitação no número de técnicas de visualização, pistas visuais e, inclusivé, interação.

3.3. Mapeamento visual e interação

O *Modelo-Data-Room* vai permitir aos utilizadores visualizar *big data* de forma imersiva virtual móvel. Um ambiente imersivo virtual móvel pode ser projetado para um único utilizador ou vários utilizadores. O *design* concentrou-se na construção de um ambiente para um único utilizador e em como esse único utilizador pode interagir com o mesmo. Existem funcionalidades específicas para a visualização de informação:

1. Rodar a cabeça para a direita/esquerda, cima/baixo e a câmara mudar em conformidade para que a imersão seja o mais natural possível;
2. Saber para onde está a olhar o utilizador para que se consiga orientar. Para isso, haverá um cursor no centro do ecrã;
3. Ver uma visualização de cada vez para não haver excesso de informação. Apenas uma visualização será disponibilizada de cada vez no ambiente;
4. Selecionar objetos para que possam ser manipulados. O cursor alterará o seu estado quando o utilizador “olha” para objetos específicos e aquando de um clique algo específico acontecer.
5. Refinamento da visualização para que fique mais fácil de compreender. Haverá um menu com um botão que permita trocar de pistas visuais. Ao mudar as pistas visuais, a técnica de visualização e o conjunto de dados continua o mesmo;
6. Trocar a técnica de visualização para encontrar outras correlações e padrões nos dados. Haverá um menu com um botão que permita trocar de técnica de visualização. Ao mudar a técnica de visualização o conjunto de dados continuará o mesmo.
7. Trocar de conjunto de dados para analisar outro tipo de informação. Um botão irá permitir a troca do conjunto de dados. Ao mudar o conjunto de dados a técnica de visualização poderá alterar;
8. Movimentar-se para ver a visualização de formas diferentes. Irá existir um botão que permitirá ao utilizador começar/parar de estar num estado de locomoção.
9. Saber onde se encontra para não se perder no ambiente imersivo. Assim, um mapa será incluído no ambiente imersivo 3D. A localização atualizada do utilizador é vista no mapa.

O organograma presente na Figura 3-3 demonstra de forma alto-nível os requisitos para aceder ao ambiente, nomeadamente *smartphone* Android, com mínimo nível de API 19 (4.4 ‘KitKat’), com giroscópio e acesso à internet para poder fazer download do .apk. Abrindo a aplicação, o mesmo poderá visualizar os dados ou mudar os mesmos. Além disso, poderá realizar outras operações como mover-se, mudar técnica de visualização, mudar pistas visuais, orientar-se e saber para onde está voltado (*Gaze system*). De salientar ainda que apenas iremos considerar duas técnicas de visualização, nomeadamente gráfico de barras e gráfico de bolhas, pela simplicidade, serem adequadas às visualizações que pretendemos, por também serem suficientes para o modelo validar, ou não, a hipótese e que, com um ligeiros ajustes, o gráfico de bolhas pode tornar-se em gráfico de linhas ou num gráfico de dispersão.

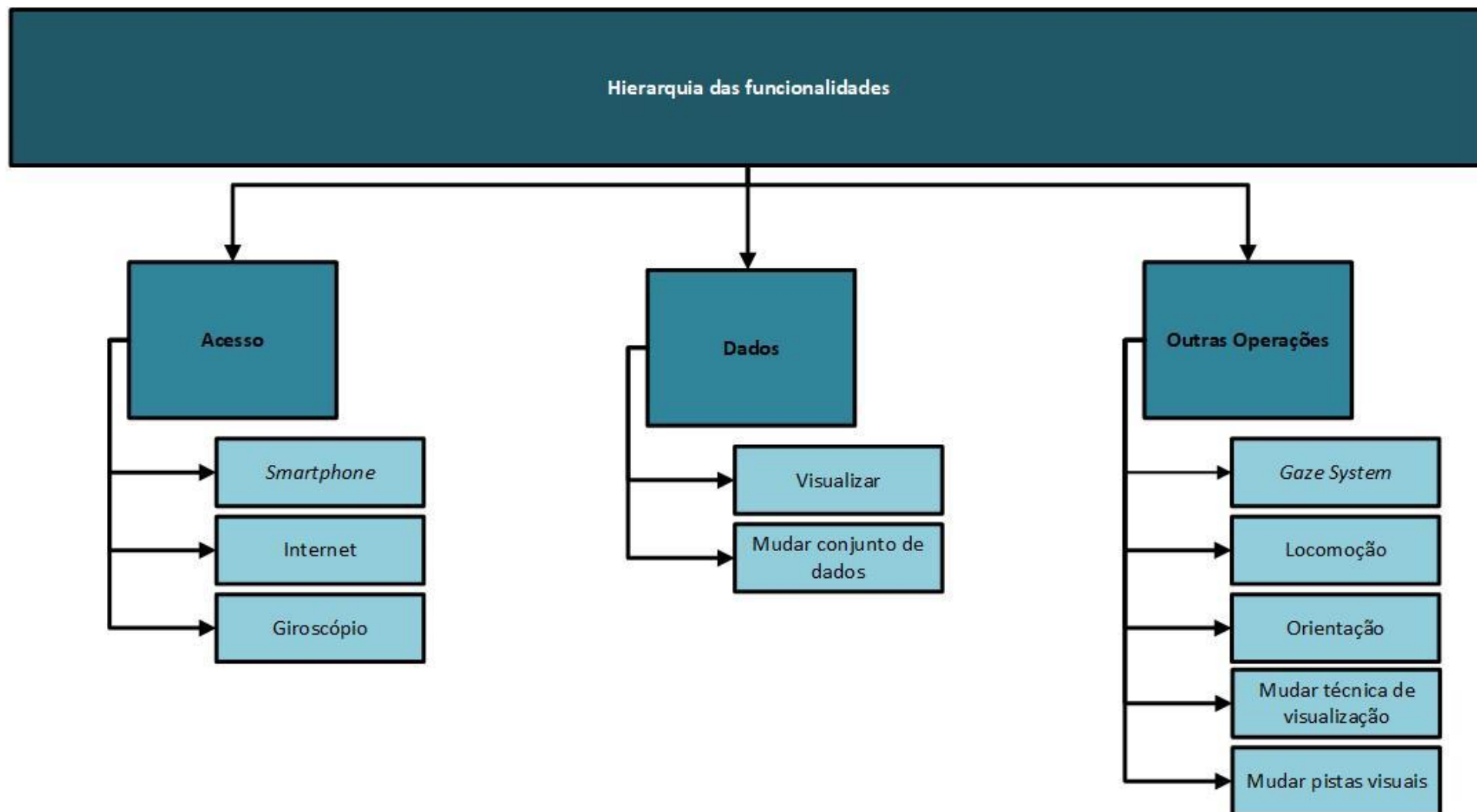


Figura 3-3. Organograma com a hierarquia das funcionalidades esperada.

Considerando as funcionalidades do utilizador acima referidas, criámos um *esquiço* (Figura 3-4) onde é possível visualizar um ambiente com apenas uma técnica de visualização e dois painéis: um contém um mapa e o outro um menu com seis botões. De notar que as ações têm escrito “Action 1”, “Action 2”, ... “Action 6”, porque nesta fase há liberdade no design, uma vez que o importante é salientar que haverá um menu de forma igual ou similar ao demonstrado na Figura 3-4 com botões que permitirão realizar, através de um clique, as funções que pretendemos.

Referente às pistas visuais, iremos ter pistas comuns em ambas as técnicas assim como específicas a cada técnica. Numa perspectiva geral, ambas as técnicas permitirão ao utilizador visualizar de diferentes ângulos e a diferentes distâncias recorrendo a técnicas de locomoção (voar e teletransporte). No tocante a cada técnica e com a classificação proposta por Cleveland e McGill (1984)

- Gráfico de barras
 - Todas as barras começam no 0. As com valores negativos estão voltadas para baixo enquanto que as positivas para cima.
 - Na parte superior da barra iremos incluir o respetivo valor.
 - Ao “olhar” para a barra, a mesma aumenta de tamanho em relação às outras.
- Gráfico de bolhas
 - Diferentes posições ao longo de uma escala comum.
 - Pontos com volume e área diferentes, aumentando quanto mais positivo for o número.

A partir do esquiço, colorimos utilizando Adobe Illustrator (Figura 3-5) simulando através de sombras um efeito 3D. Dado que existem atualmente ferramentas que permitem visualizar modelos 3D de forma aumentada, recorreremos à aplicação Zappar (ver Anexos, secção V, Tabela 5, onde podemos ver uma lista de soluções encontradas para este efeito) e ao Cinema 4D (2020) para criar um componente que permite visualizar o futuro ambiente a implementar em 3D (Figuras 3-6 e 3-7) (para os leitores que possuem Android, podem fazer download em Anexos, secção VI, de uma aplicação criada Vuforia com o mesmo efeito).

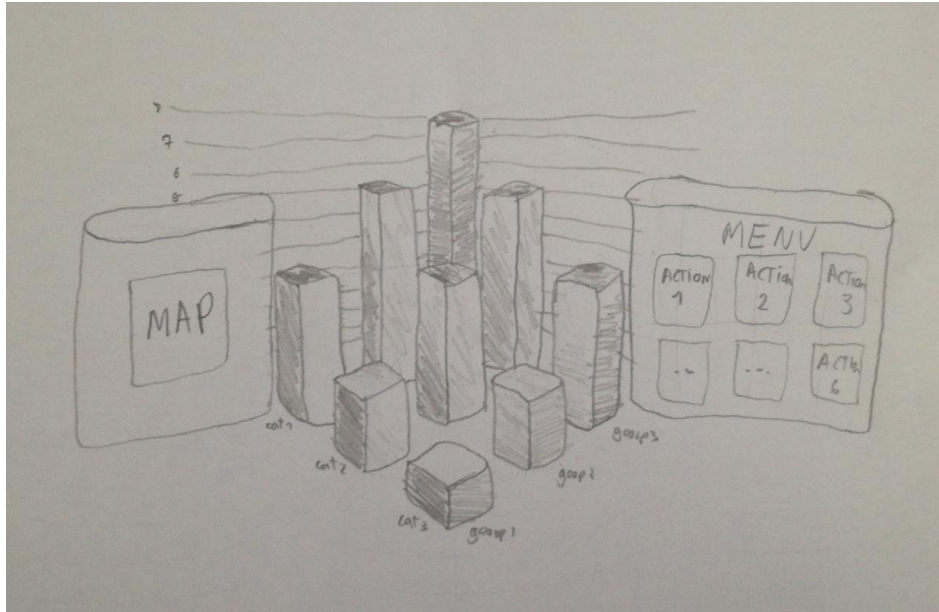


Figura 3-4. Esquiço do ambiente a implementar.

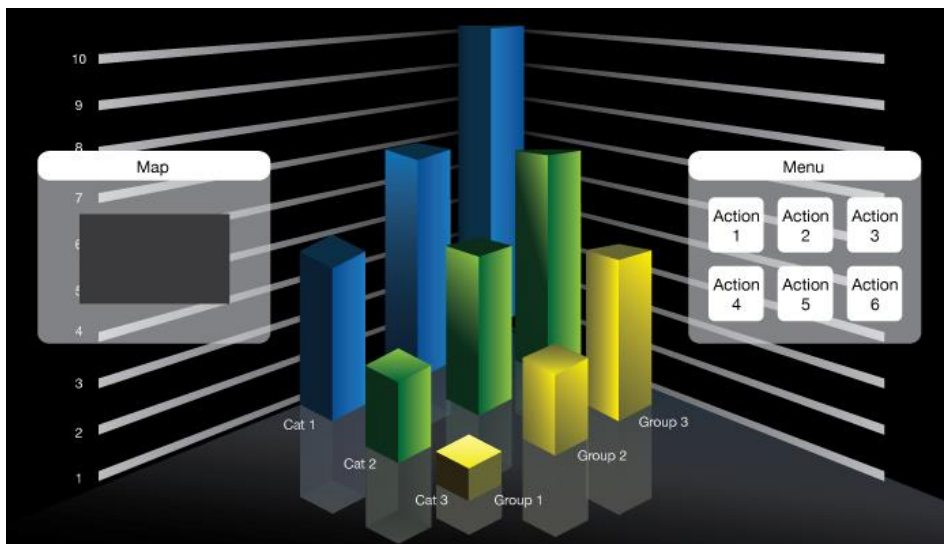


Figura 3-5. Mockup do ambiente a implementar com recurso ao Adobe Illustrator.

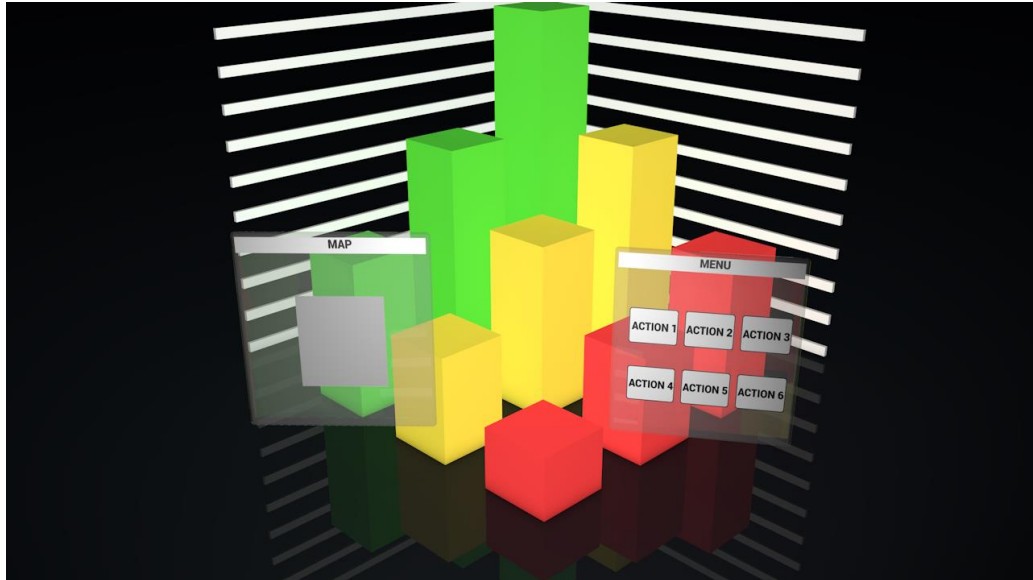


Figura 3-6. *Mockup* do ambiente a implementar com recurso ao Cinema4D.

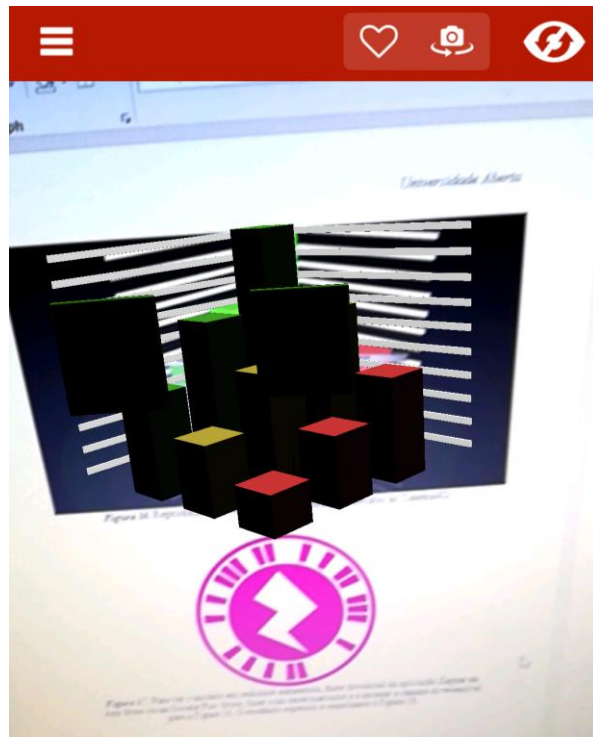


Figura 3-7. Ambiente a implementar visto através da aplicação Zappar após fazer *scan* do marcador acima - Realidade Aumentada (RA).

3.4. Síntese

Para pessoas que querem visualizar *big data* de forma imersiva móvel, o *Modelo-Data-Room* tirará partido de uma aplicação de telemóvel que permite visualizar recorrendo a um ou mais conjuntos de dados, técnicas de visualização e pistas visuais (o organograma da Figura 3-3 ilustra as funcionalidades e a hierarquia existente entre estas). É esperado que seja possível de aceder à aplicação recorrendo a um *smartphone* com acesso a internet. Dentro da aplicação, pretende-se visualizar dados, saber para onde estamos a olhar (*gaze system*), mudar pistas visuais, técnica de visualização e dados, movimentarmo-nos para visualizar os dados de outros ângulos e orientarmo-nos para que saibamos onde nos encontramos no ambiente imersivo virtual.

Dadas as restrições temporais, o *Protótipo-Data-Room* (4.1.2.) será implementado e testado em (pelo menos) um telemóvel Android com versão superior a 4.4 e com giroscópio. Isto deve-se ao facto de ser possível testar o modelo com recurso a apenas a um *smartphone*, o que implica que os participantes do estudo usem todos, em último caso, o mesmo telemóvel (por nós fornecido). Para além disso, para usufruir de três graus de liberdade é necessário que esse telemóvel tenha giroscópio.

O mais importante para validar o modelo é que o utilizador consiga visualizar os dados. Todavia, para que possa haver uma maior variedade de visualizações, será permitido ao utilizador mudar o conjunto de dados, a técnica visual e pistas visuais à distância de um clique (ainda que o utilizador esteja limitado ao que será oferecido no protótipo). De salientar ainda que é esperado que as operações de *gaze system*, de locomoção e de orientação sejam totalmente implementadas, pois permitirão ao utilizador ter uma experiência melhorada no ambiente ao: saber para onde está a olhar, interagir com objetos, deslocar-se no ambiente, de uma ou mais formas diferentes, e saber exatamente onde se encontra em relação aos dados que estão a ser visualizados.

4. Implementação e validação do modelo

Este capítulo descreve os processos de implementação (secção 4.1.) e validação (4.2.) do modelo proposto (descrito no capítulo 3).

4.1. Implementação

Na secção 4.1.1. apresentamos o protótipo *Data-Room* e na 4.1.2. explicitamos onde foram adquiridos os dados e quais serão utilizados para as visualizações. Por fim, na secção 4.1.3. apresentamos as visualizações recorrendo a vários ambientes, nomeadamente, *Data-Room*, Microsoft Excel, Python e R.

4.1.1. *Dados utilizados*

Com o intuito de criarmos visualizações, recorremos à Tabela 1, secção I, em Anexos, de grandes conjuntos de dados abertos ao público e escolhemos um conjunto de dados com resumo dos dados meteorológicos da superfície global (*Global Surface Summary of the Day Weather Data / GSOD*) criado pela Administração Nacional Oceânica e Atmosférica (*National Oceanic and Atmospheric Administration / NOAA*). Foi possível de encontrar informação do mesmo no Kaggle (2019), no entanto os dados foram acedidos utilizando Google BigQuery.

Este conjunto de dados abrange os dados GSOD entre 1929 e o ano atual (atualizados diariamente), colecionados em mais de 9000 estações espalhadas pelo mundo. O resumo global do dia é composto por uma dúzia de médias diárias calculadas a partir dos dados globais das estações horárias. Os elementos climáticos diários incluem diversas dimensões como: temperatura, temperatura do ponto de orvalho, pressão do nível do mar, pressão da estação, visibilidade e velocidade do vento mais a respectiva temperatura máxima e mínima, velocidade máxima sustentada do vento e rajada máxima, quantidade de precipitação, profundidade da neve e indicadores climáticos. Com exceção das estações dos EUA, os períodos de 24 horas são baseados nos horários UTC (*Coordinated Universal Time*).

Reforçamos que estaremos a consumir dados disponibilizados pela NOAA, sendo que o nosso objetivo é efetuar a comunicação no contexto pedagógico, para os leitores desta dissertação. Além disso, como os dados brutos costumam ser muito primitivos para serem usados diretamente para criar visualizações, poderemos necessitar de prepará-los, mais precisamente:

- Lidar com valores ausentes
- Lidar com valores atípicos
- Decidir o nível de granularidade dos dados, por exemplo, as variáveis de tempo devem estar em segundos, minutos ou horas, etc
- Juntar tabelas.

Além de perguntar se os dados fazem sentido lógico, geralmente, é uma boa ideia verificar se os dados fazem sentido. Fazer isso pode ajudar a detectar certos erros nos dados, como dados rotulados incorretamente ou atribuídos a um conjunto incorreto de recursos. Se não forem contabilizados, esses erros podem ter um impacto profundo na análise.

4.1.2. *Data-Room*

Baseados no modelo proposto no capítulo anterior, implementámos o *Protótipo-Data-Room*. Este permite explorar e interagir com um ou mais conjuntos de dados predefinidos de forma imersiva virtual móvel. Diferentes técnicas de visualização (gráfico de barras e gráfico de bolhas), pistas visuais e técnicas de locomoção, orientação e interação são utilizadas.

Este protótipo serve como uma prova de conceito do modelo *Data-Room*. Em termos de meios materiais utilizados para a implementação do protótipo, podem ser divididos em computacionais (Unity 3DTM, com o *asset Graph and Chart*, e serviços da Google Cloud) e físicos (computador, telemóvel Android com versão superior a 4.4 *KitKat*, um cabo USB para ligar o telemóvel ao computador e um Cardboard da marca *Splaks*). A Figura 4-1 demonstra a arquitetura do protótipo com as escolhas tecnológicas, não considerando as mudanças de infraestrutura referentes ao processamento dos dados.

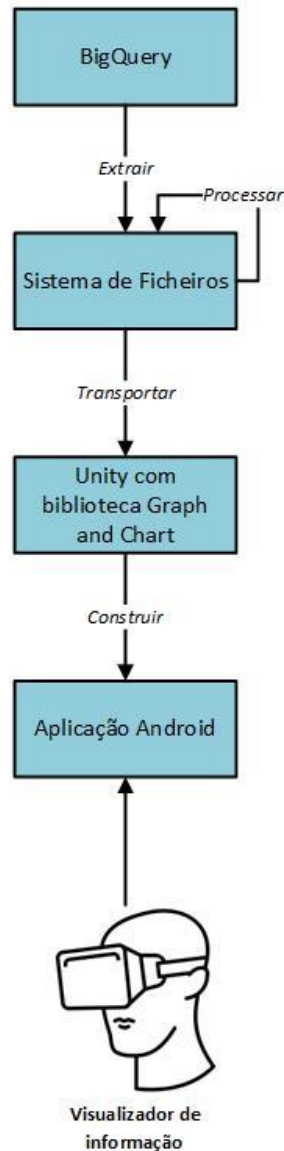


Figura 4-1. Arquitetura da prova de conceito com escolhas tecnológicas.

Optámos por usar BigQuery da Google Cloud como fonte de dados, pois permite que a extração seja feita com a especificidade desejada. O BigQuery foi projetado para lidar com grandes quantidades de dados, distribuindo o trabalho entre várias máquinas utilizando uma técnica chamada *Dremel* (Melnik *et al.*, 2011). Como foi projetado para usar várias máquinas e processamento paralelo, podemos esperar boa escalabilidade e desempenho na especificação dos dados que pretendemos extrair. Além disso, no BigQuery temos não só acesso gratuito ao conjunto de dados público que pretendemos analisar como também existem cursos na Qwiklabs para compreender e utilizar a plataforma. Para perceber como os dados foram extraídos, ver Anexos secção XVI.

Por outro lado, no Google BigQuery podemos apenas exportar ficheiros com um tamanho máximo de 1 GB, ou seja, se o ficheiro que pretendemos exportar tem 9 GBs, teremos de exportar vários ficheiros. Para colmatar esta limitação, exportamos os ficheiros em CSV e juntámo-los localmente com recurso a um script em Python, que os juntou rapidamente (para mais informação sobre esta como é feita a junção, ver Anexos,

secção VII). Além disso, aquando da especificação que é feita através de comandos SQL, não podemos utilizar *ORDER BY* para determinadas *queries*. Como tal, essa ordenação é feita a posteriori no sistema de ficheiros (como os dados foram extraídos do GCS sem cabeçalhos, adicionámos localmente os mesmos e através de um comando Windows podemos ordenar e posteriormente removemos - ver Anexos, secções IX e X).

Para prevenir sobrecarga no computador ou afetar significativamente o desempenho dadas as limitações de memória (inclusivé na ordenação acima referida), inspirámo-nos na ferramenta Trifacta Wrangler e usámos as amostras para criar as visualizações. Existem vários tipos de amostras (Figura 4-2) e, ainda que tenham vários problemas associados (podendo, inclusivé, não representar corretamente os dados), na próxima secção vamos prosseguir com a utilização de amostras aleatórias, que são geradas rapidamente, para criar as visualizações (para saber como foram criadas as amostras, ver Anexos, secção VIII). Em maior detalhe:

- *Amostras aleatórias* — seleção aleatória de um subconjunto de linhas no conjunto de dados.
- *Amostras baseadas em filtros* — encontrar valores específicos em uma ou mais colunas. Para o conjunto de valores correspondente, uma amostra aleatória é gerada.
- *Amostras baseadas em anomalias* — encontrar dados incompatíveis ou ausentes ou ambos em uma ou mais colunas. Aqui podemos especificar uma ou mais colunas e se a anomalia trata incompatibilidade, ausência de ou ambos.
- *Amostras estratificadas* — encontrar todos os valores exclusivos em uma coluna e criar uma amostra que contenha os valores exclusivos, até o limite de tamanho da amostra. A distribuição dos valores da coluna na amostra reflete a distribuição dos valores da coluna no conjunto de dados. Os valores amostrados são classificados por frequência, em relação à coluna especificada.
- *Amostras baseadas em grupos* — encontrar linhas contíguas no conjunto de dados que correspondem a uma seleção aleatória dos valores exclusivos em uma coluna. Todas as linhas correspondentes aos valores exclusivos selecionados aparecem na amostra, até o tamanho máximo da amostra.

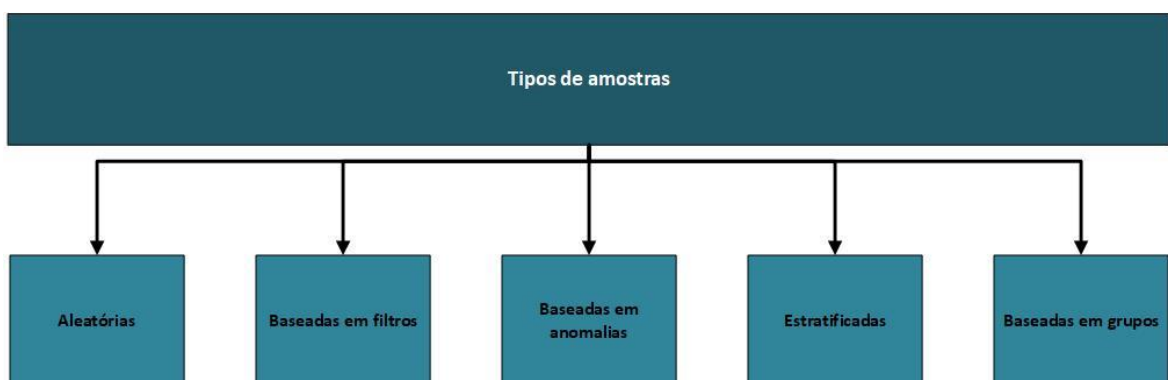


Figura 4-2. Tipos de amostras.

Como os dados que estamos a utilizar estão em formato CSV, recorremos a um CSVReader (ver Anexos, secção XI) que foi adaptado, por conveniência, para que permita introduzir dados que não tenham cabeçalhos. Para criar a representação visual dos dados recorremos à biblioteca *Graph and Chart*. Esta biblioteca inclui, até ao momento, apenas uma técnica de visualização de dados que podemos definir como 3D –

gráfico de barras. No entanto, ao modificar a categoria e profundidade conseguimos exibir informações 3D discretas num gráfico de bolhas (para mais informação sobre esta modificação, ver Anexos, secções XI e XII). Assim, passamos de uma a duas técnicas de visualização em 3D, nomeadamente: gráfico de barras e gráfico de bolhas. Na Figura 4-3, podemos observar o ambiente *Data-Room* com todas as funcionalidades mencionadas no capítulo anterior. Uma esfera verde que aponta a direção da cabeça do utilizador e indica que podemos interagir com o botão para onde está direcionada, dois painéis (mini-mapa e menu) com um dos botões ativos e um gráfico de barras 3D com dados aleatórios.

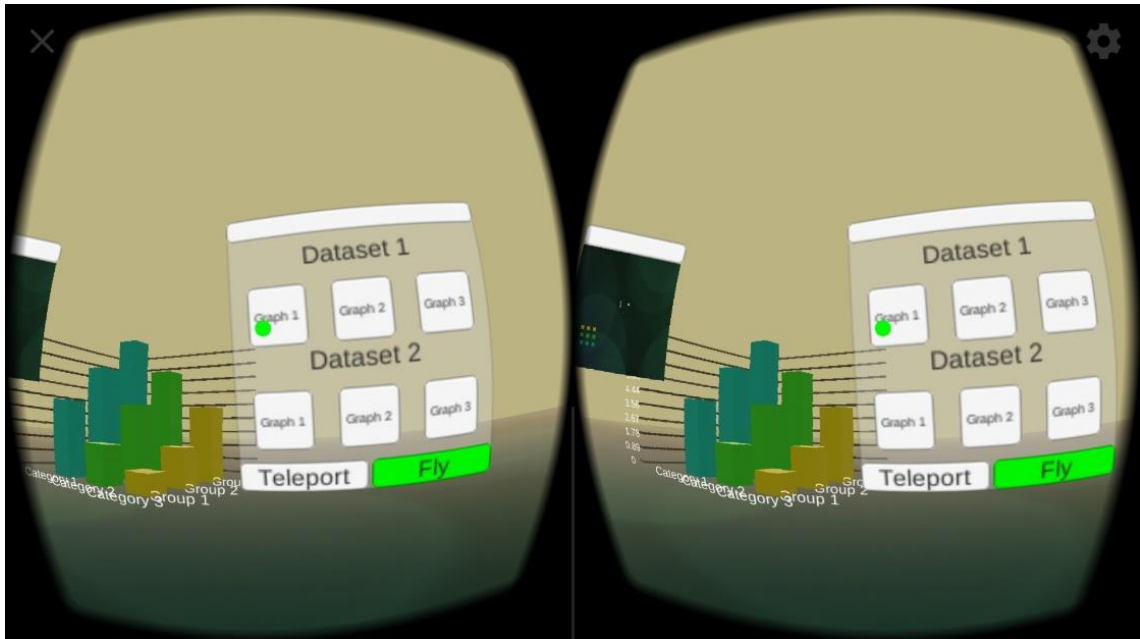


Figura 4-3. Captura de ecrã da solução *Data-Room*.

No tocante a considerações de implementação, o protótipo de visualização de informação foi desenvolvido para um telemóvel Android (Huawei PRA-LX1) com versão superior a 4.4 com giroscópio. O protótipo é possível de ser adquirido no seguinte sítio: https://github.com/tiago-peres/mestrado/blob/master/app_final.apk (acedido dia 11 de maio de 2020).

4.1.3. Visualizações com *Data-Room*, *Python* e *R*

Uma vez que estamos a efetuar uma comunicação no contexto pedagógico, ou seja, as codificações visuais serão utilizadas para testar o modelo e apresentar os resultados no presente documento, optámos por ter em consideração a forma lógica como disponibilizámos as codificações visuais. Para isso, considerámos uma abordagem *Top-Bottom*, onde, de acordo com os dados disponibilizados, iremos analisar numa perspectiva mais alto nível o panorama geral e, afunilar a uma perspectiva mais baixo nível, focando numa cidade e nos últimos anos.

Na Tabela 4-1 registamos as diferenças visuais presentes nas representações criadas por nós, num Asus Laptop Intel Core™ i7-4510U CPU @ 2.00 GHz com 8 GBs de RAM e sistema operativo Windows 10, nos diferentes ambientes. O processo inerente à criação

das visualizações abaixo expostas pode ser visualizado nos Anexos, secções XVII, XVIII e XIX. (optámos também por incluir as secções XIV e XV para explicar como os ambientes em R e Python foram instalados e configurados).

Tabela 4-1. Características visuais e interativas nos diferentes ambientes.

Ambiente	N. de pontos	Interativo	Imersão física	Cor	Eixos e legendas
Python Mayavi barchart()	10 ⁵	No computador sim	Não	Escala de cor	Fácil de configurar
R latticeExtra	10 ⁵	Não	Não	Escala de cor	Fácil de configurar
Data-Room gráfico de barras	10 ⁴	Sim tanto no computador quanto no <i>smartphone</i>	Sim	Constante e invariável	Fácil de configurar
Python Mayavi points3d()	10 ⁴ ~10 ⁵	No computador sim	Não	Escala de cor	Fácil de configurar
R plot3D	10 ⁴	Não	Não	Escala de cor	Fácil de configurar
Data-Room gráfico de bolhas	10 ³	Sim tanto no computador quanto no <i>smartphone</i>	Sim	Constante e invariável	Fácil de configurar dois dos eixos

1.ª Codificação Visual

Com esta visualização foi analisada a relação dos anos e / ou meses com as temperaturas, desde que existem registos, em mais de 9000 estações espalhadas pelo mundo. Na Tabela 4-2 podemos ver a estatística descritiva (ver Anexos, secção XX para perceber como a mesma foi calculada) do conjunto de dados extraído do BigQuery que serviu como base para a criação das visualizações. Como podemos ver, os dados de temperatura mínima e máxima fazem sentido pois estão entre os recordes mundialmente registados. Essa situação ocorre nas demais codificações também.

A partir deste conjunto de dados foi gerada uma amostra aleatória com um número de linhas correspondente à capacidade do ambiente capaz de lidar com o menor número de pontos (Tabela 4-3). Nós tivemos o cuidado de utilizar a mesma amostra para os três ambientes de forma a não criar bias na interpretação dos resultados em função de cada ambiente ter limites de processamento diferentes.

Tabela 4-2. Estatística descritiva do conjunto de dados.

Variável	Número de pontos	Média	Desvio Padrão	Mínimo	Máximo
ano	154038545	-	-	1929	2019
mês		-	-	janeiro	dezembro
temperatura (°C)		11,64	13,30	-83,61	43,33

Tabela 4-3. Estatística descritiva da amostra usada para a criação das visualizações nas Figuras 4-4, 4-5 e 4-6.

Variável	Número de pontos	Média	Desvio Padrão	Mínimo	Máximo
ano	999	-	-	1940	2019
mês		-	-	janeiro	dezembro
temperatura (°C)		10,86	13,56	-62,39	35,94

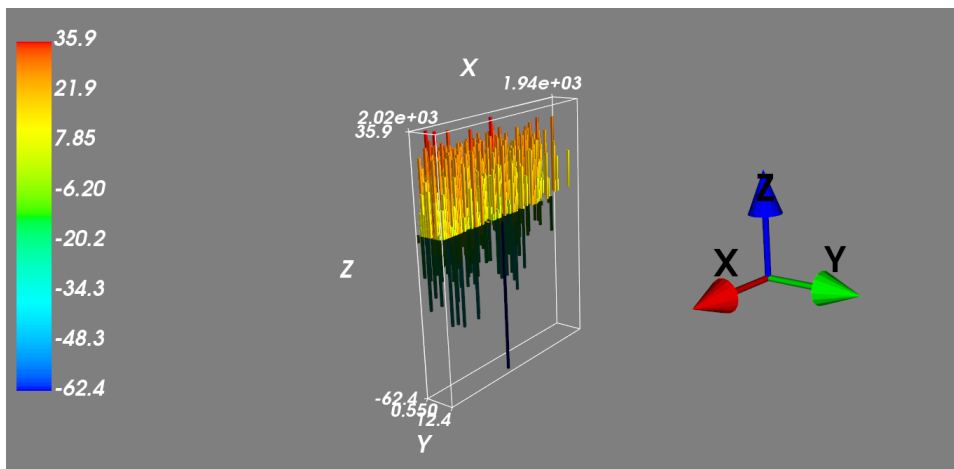


Figura 4-4. Python Mayavi barchart() (x= anos, y= meses, z= temperatura em °C).

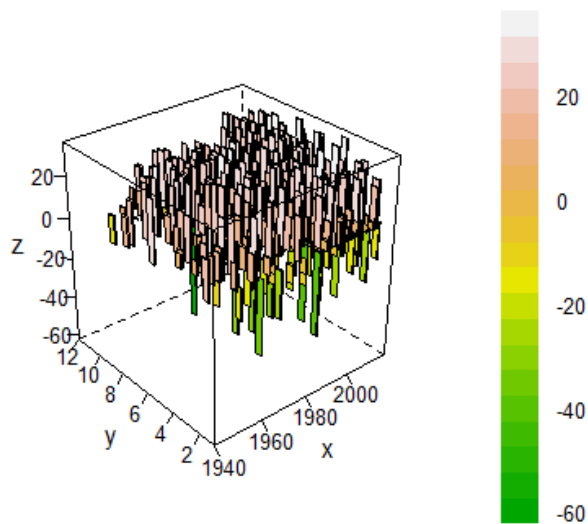


Figura 4-5. R latticeExtra (x= anos, y= meses, z= temperatura em °C).

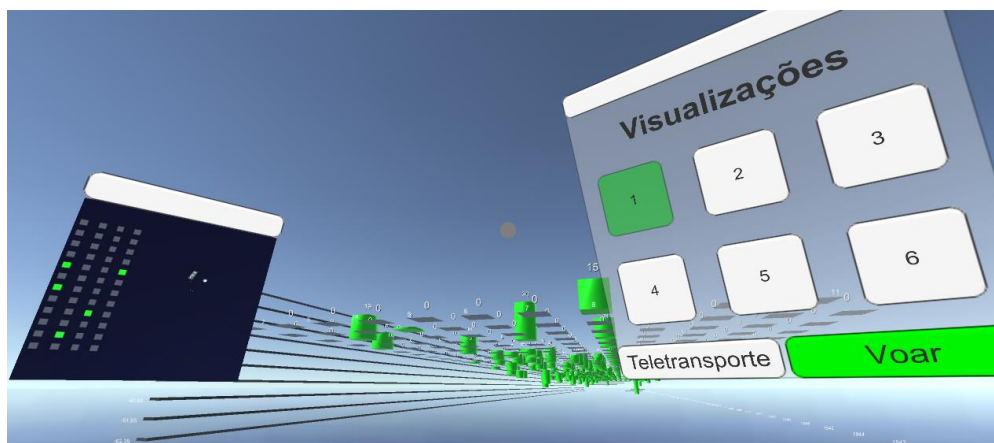


Figura 4-6. Data-Room gráfico de barras (x= anos, y= meses, z= temperatura em °C).

2.ª Codificação Visual

Com esta visualização pretendemos analisar a relação dos anos e / ou meses com as temperaturas, recorrendo a dados dos últimos 10 anos, em mais de 9000 estações espalhadas pelo mundo. Na Tabela 4-4 podemos ver a estatística descritiva do conjunto de dados extraído do BigQuery e na Tabela 4-5 podemos ver a estatística descritiva da amostra criada a partir desses mesmos dados que serviu para a criação das visualizações.

Tabela 4-4. Estatística descritiva do conjunto de dados.

Variável	Número de pontos	Média	Desvio Padrão	Mínimo	Máximo
ano	43636699	-	-	2009	2019
mês		-	-	janeiro	dezembro
temperatura (°C)		12,74	12,86	-81,50	43,33

Tabela 4-5. Estatística descritiva da amostra usada para a criação das visualizações nas Figuras 4-7, 4-8 e 4-9.

Variável	Número de pontos	Média	Desvio Padrão	Mínimo	Máximo
ano	9999	-	-	2009	2019
mês		-	-	janeiro	dezembro
temperatura (°C)		12,81	12,76	-67,17	41,61

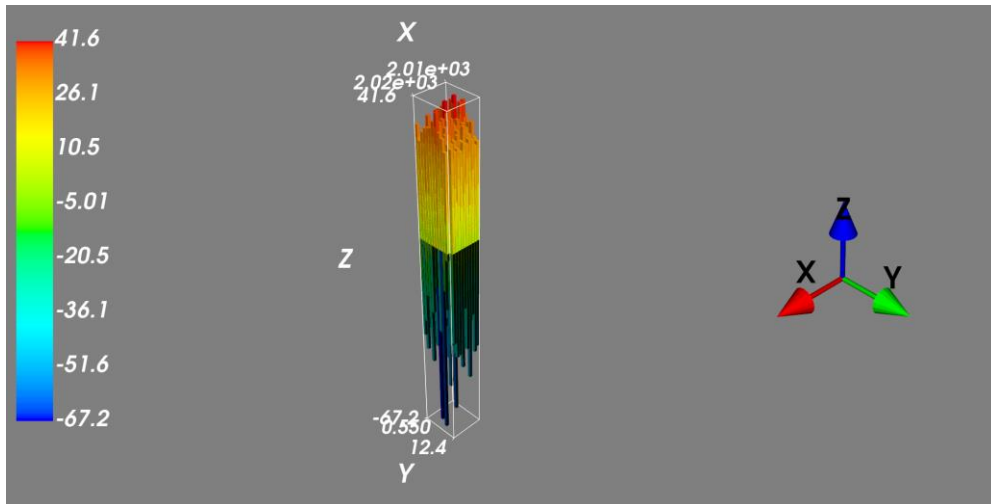


Figura 4-7. Python Mayavi barchart() (x= anos, y= meses, z= temperatura em °C).

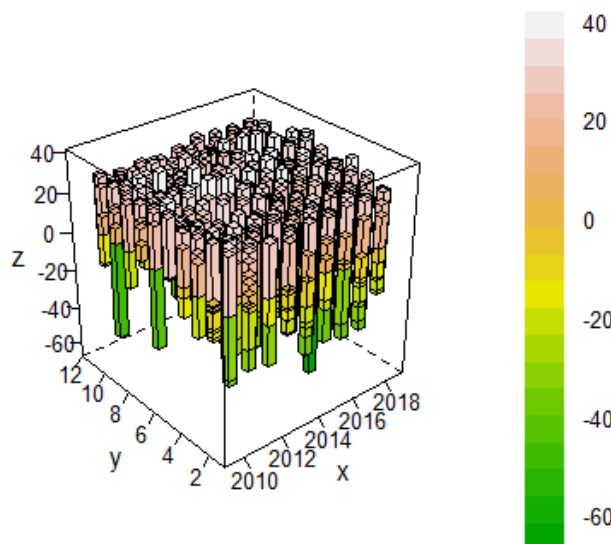


Figura 4-8. R latticeExtra (x= anos, y= meses, z= temperatura em °C).

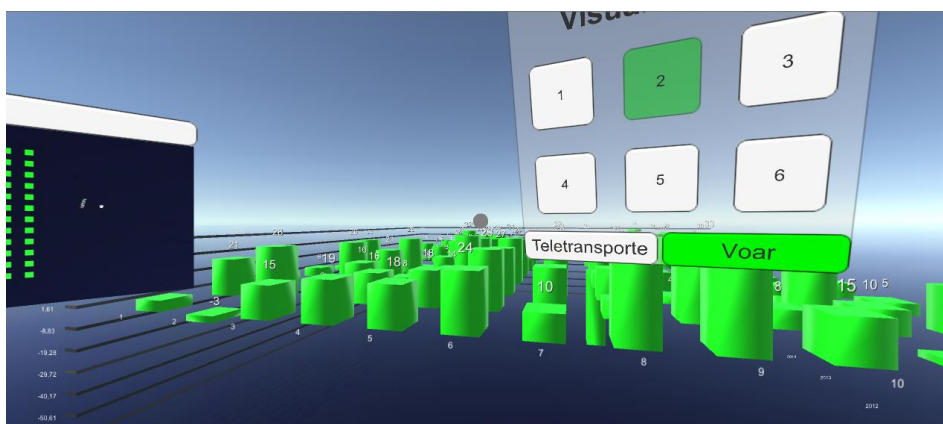


Figura 4-9. Data-Room gráfico de barras (x= anos, y= meses, temperatura em °C).

3.ª Codificação Visual

Com esta visualização pretendemos analisar a relação dos anos e / ou meses com as temperaturas, recorrendo a dados dos últimos 10 anos no Cabo Carvoeiro, Peniche. Na Tabela 4-6 podemos ver a estatística descritiva do conjunto de dados extraídos do BigQuery. Neste caso não foi necessário gerar uma amostra.

Tabela 4-6. Estatística descritiva do conjunto de dados usado para a criação das visualizações nas Figuras 4-10, 4-11 e 4-12.

Variável	Número de pontos	Média	Desvio Padrão	Mínimo	Máximo
ano	638	-	-	2009	2010
mês		-	-	janeiro	dezembro
temperatura (°C)		15,99	3,21	3,89	23,61

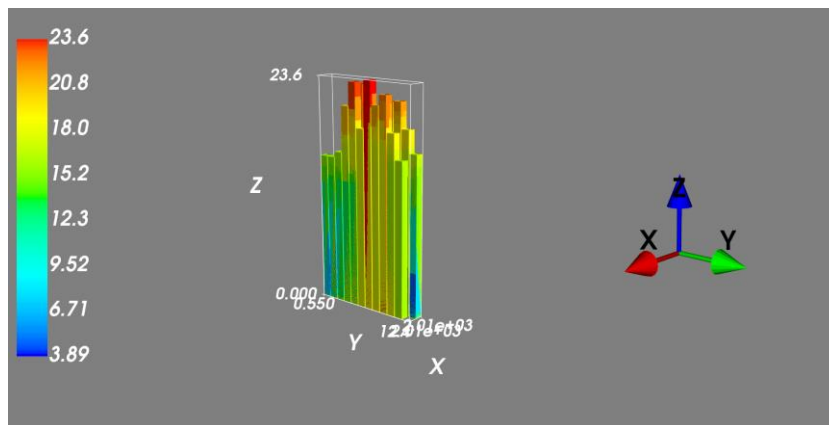


Figura 4-10. Python Mayavi barchart() (x= anos, y= meses, z= temperatura em °C).

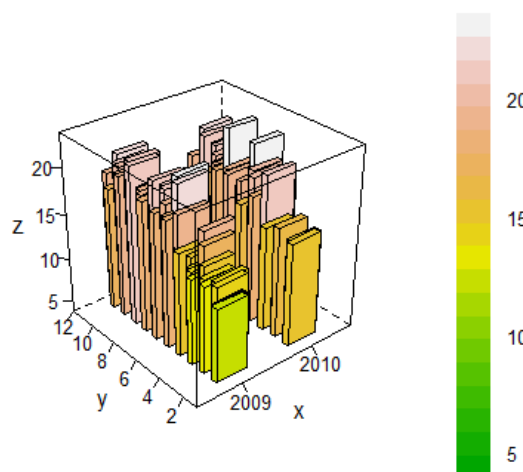


Figura 4-11. R latticeExtra (x= anos, y= meses, z= temperatura em °C).

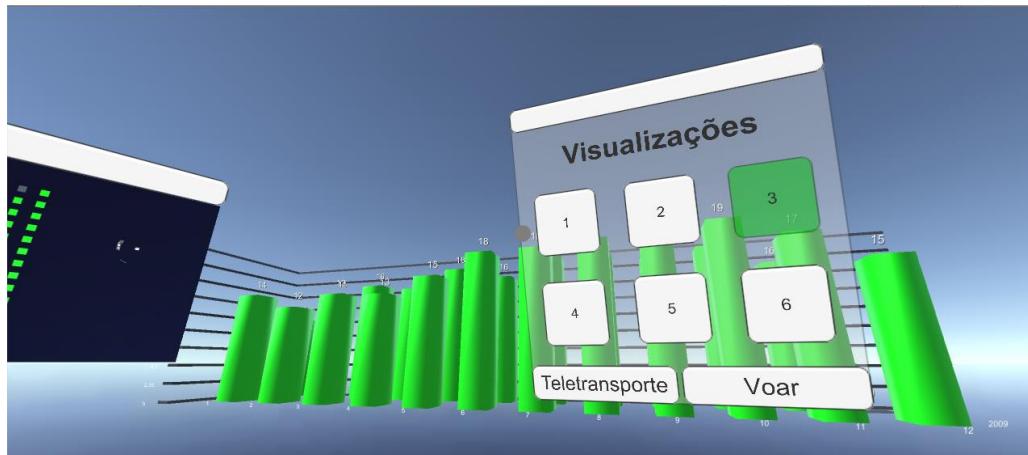


Figura 4-12. Data-Room gráfico de barras (x= anos, y= meses, z= temperatura em °C).

4.ª Codificação Visual

Com esta visualização pretendemos analisar a relação dos anos e / ou dias com as temperaturas no Cabo Carvoeiro, desde que existem registos. Na Tabela 4-7 podemos ver a estatística descritiva do conjunto de dados extraído do BigQuery e na Tabela 4-8 podemos ver a estatística descritiva da amostra criada a partir desses mesmos dados que serviu para a criação das visualizações.

Tabela 4-7. Estatística descritiva do conjunto de dados.

Variável	Número de pontos	Média	Desvio Padrão	Mínimo	Máximo
ano	9170	-	-	1973	2010
dia		-	-	1	31
temperatura (°C)		24,33	2,92	3,89	24,33

Tabela 4-8. Estatística descritiva dos dados usados para a criação da visualização na Figura 4-13, 4-14 e 4-15.

Variável	Número de pontos	Média	Desvio Padrão	Mínimo	Máximo
ano	999	-	-	1973	2010
dia		-	-	1	31
temperatura (°C)		15,39	2,92	3,89	22,33

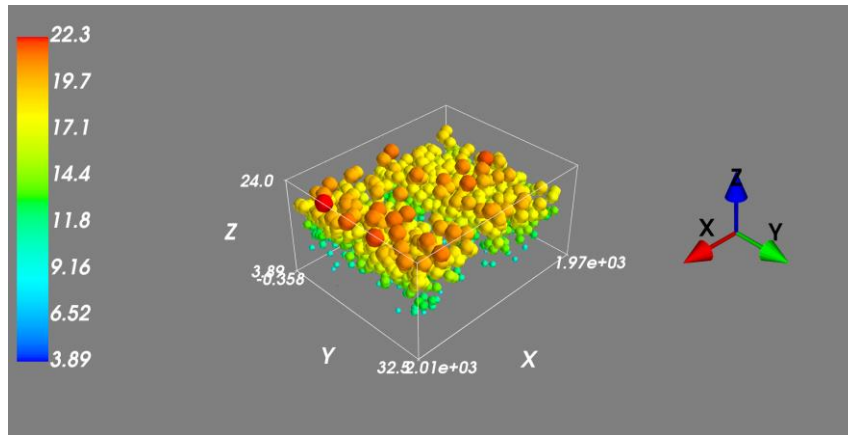


Figura 4-13. Python Mayavi points3D() (x= anos, y= dias, z= temperatura em °C).

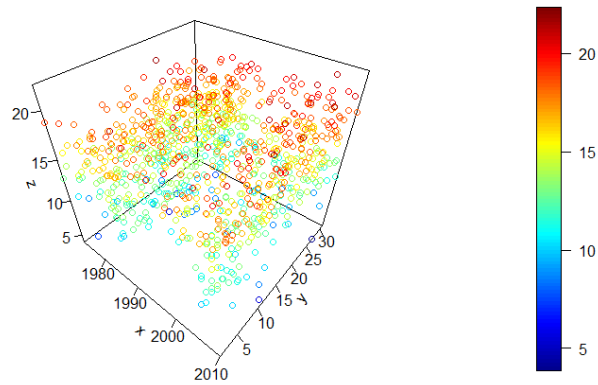


Figura 4-14. R plot3D (x= anos, y= dias, z= temperatura em °C).

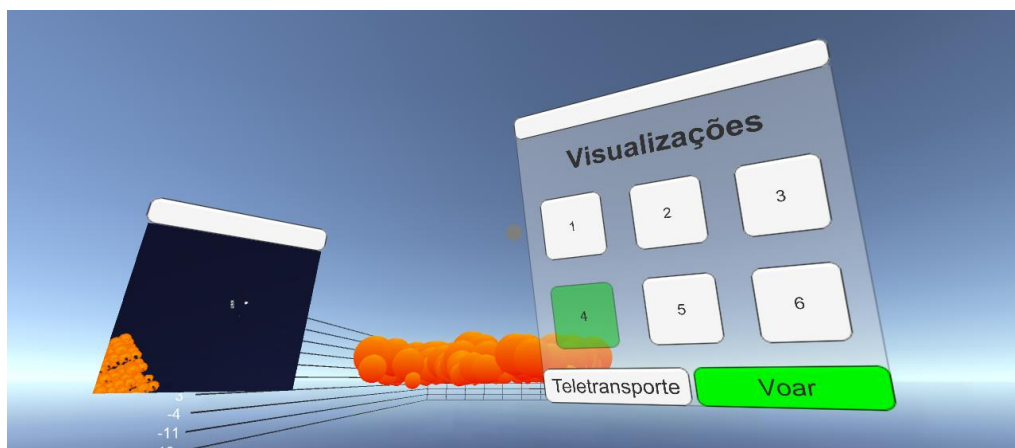


Figura 4-15. Data-Room gráfico de bolhas (x= anos, y= dias, z= temperatura em °C).

5.ª Codificação Visual

Com esta visualização pretendemos analisar a relação dos anos e / ou dias com as temperaturas, recorrendo a dados dos últimos 10 anos no Cabo Carvoeiro, Peniche. Na Tabela 4-9 podemos ver a estatística descritiva do conjunto de dados extraído do BigQuery e , a partir desses mesmos dados, criámos as visualizações.

Tabela 4-9. Estatística descritiva do conjunto de dados usado para a criação das visualizações nas Figuras 4-16, 4-17 e 4-18.

Variável	Número de pontos	Média	Desvio Padrão	Mínimo	Máximo
ano	638	-	-	2009	2010
dia		-	-	1	31
temperatura (°C)		15,99	3,21	3,89	23,61

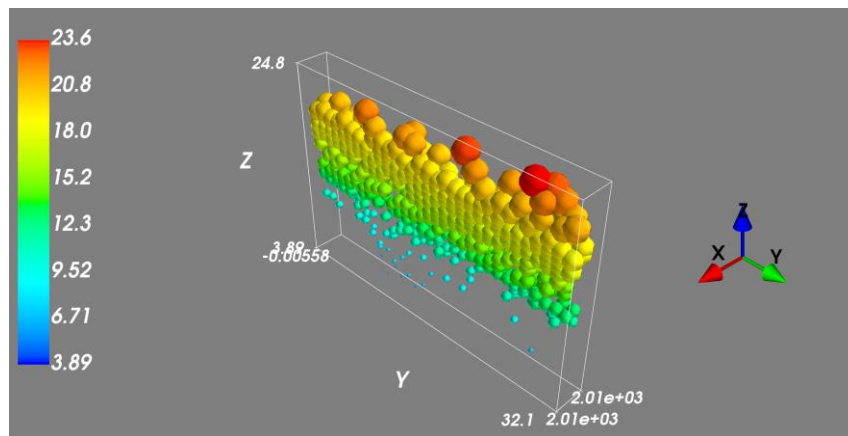


Figura 4-16. Python Mayavi points3D() (x= anos, y= dias, z= temperatura em °C).

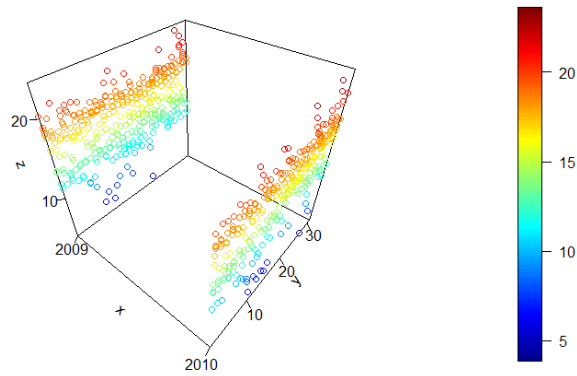


Figura 4-17. R plot3D (x= anos, y= dias, z= temperatura em °C).

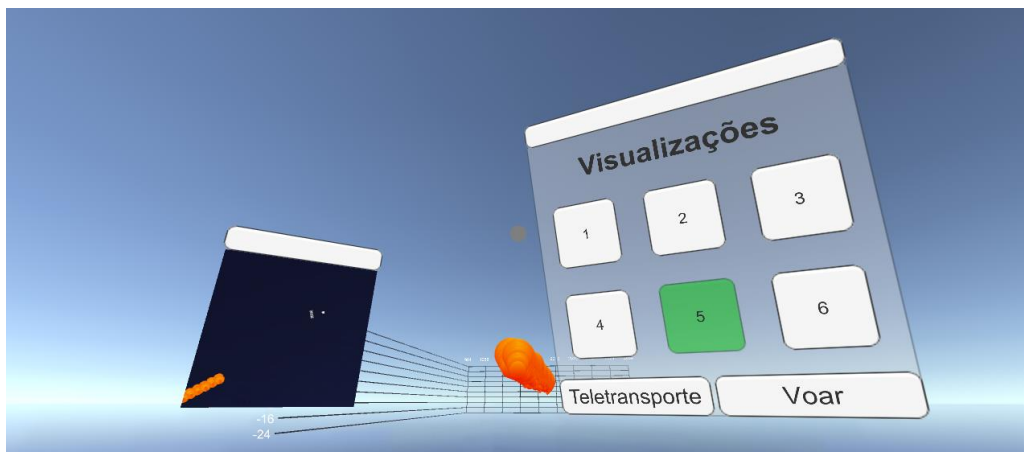


Figura 4-18. Data-Room gráfico de bolhas (x= anos, y= dias, z= temperatura em °C).

6.ª Codificação Visual

Com esta visualização pretendemos analisar a relação dos meses e / ou dias com as temperaturas, recorrendo a dados dos últimos 10 anos no Cabo Carvoeiro, Peniche. Na Tabela 4-10 podemos ver a estatística descritiva do conjunto de dados extraídos do BigQuery. Neste caso não foi necessário gerar uma amostra.

Tabela 4-10. Estatística descritiva do conjunto de dados usado para a criação das visualizações nas Figuras 4-19, 4-20 e 4-21.

Variável	Número de pontos	Média	Desvio Padrão	Mínimo	Máximo
mês	638	-	-	janeiro	dezembro
dia		-	-	1	31
temperatura (°C)		15,99	3,21	3,89	23,61

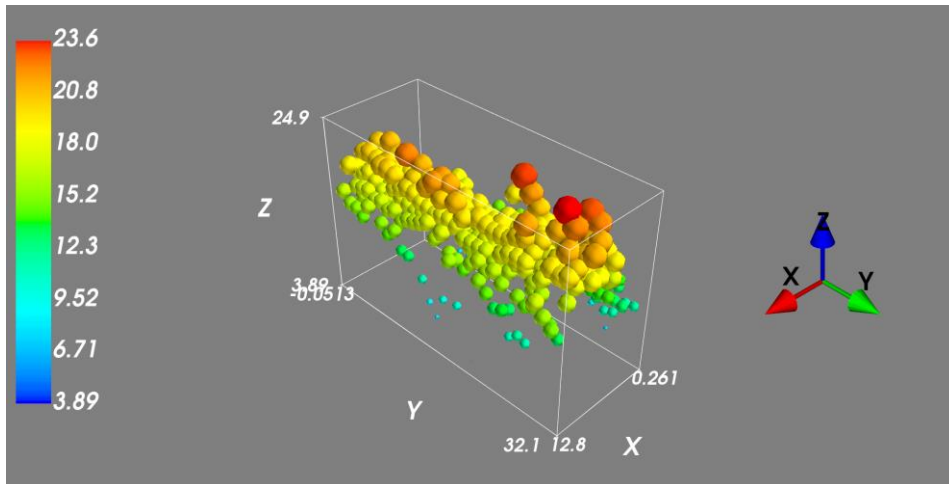


Figura 4-19. Python Mayavi points3d() (x= meses, y= dias, z= temperatura em °C).

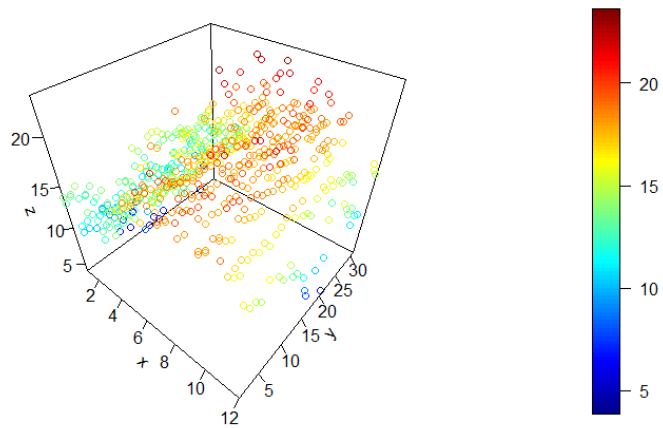


Figura 4-20. R plot3D (x= meses, y= dias, z= temperatura em °C).

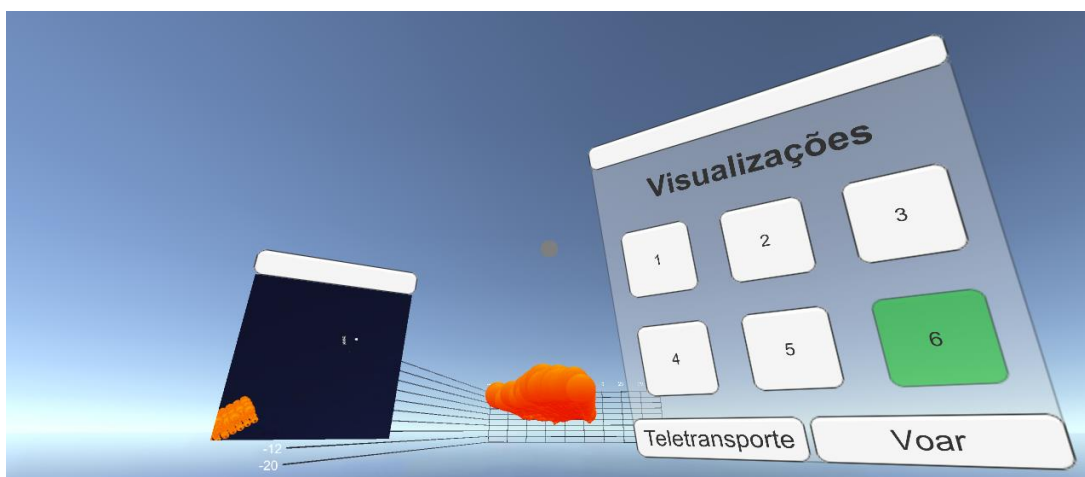


Figura 4-21. Data-Room gráfico de bolhas (x= meses, y= dias, z= temperatura em °C).

4.2. Validação

Na secção 4.2.1. mostramos como o teste do modelo foi realizado e em 4.2.2. expomos e analisamos os resultados obtidos.

4.2.1. Teste do modelo

Para testar o modelo recrutámos um total de oito participantes de idades compreendidas entre os 15 e os 30 anos. Foram produzidas experiências com ambientes diferentes, nomeadamente com o protótipo Data-Room, Python e R.

Este estudo consistiu em duas fases, nomeadamente fase de visualização (*in loco*) e fase de apreensão (realizada cinco dias após a fase inicial para que haja distanciamento mas que não seja demasiado grande ao ponto de correr risco dos participantes não se lembrarem do sucedido). Os resultados foram avaliados através de dois inquéritos: o primeiro (inquérito de visualização) realizado aquando das visualizações; o segundo (inquérito de apreensão) foi enviado por *email* aos participantes sete dias após a experiência de visualização.

Os dois inquéritos, que podem ser lidos com maior detalhe nos Anexos secção XXI, contiveram as seguintes secções:

A. Inquérito de visualização: correio electrónico dos participantes; dados demográficos; experiência com a visualização de dados; experiência em ambientes imersivos virtuais móveis; codificações visuais, reação / empenho do participante (escala de Likert de cinco pontos); descrição do sucedido; perguntas sobre a satisfação da visualização e concentração; aquisição de informação - seis perguntas sobre a visualização possíveis de obter com a experiência de visualização; memorabilidade das questões da experiência; valor percebido da visualização - questões baseadas em opiniões.

B. Inquérito de apreensão: Retenção da informação - as mesmas seis perguntas do inquérito de visualização.

De um modo mais ilustrativo, na fase inicial de visualização, cada participante acedeu com o seu telemóvel a um sítio *web* específico para preencher o inquérito de visualização por conta própria; para que este inquérito fosse realizado, disponibilizámos um Cardboard igual ao da Figura 2-27, um telemóvel Huawei PRA-LX1 com a aplicação Data-Room instalada, e, com o *feedback* do participante, colocámos tudo à feição para que o utilizador pudesse focar-se na resolução do inquérito. Este inquérito teve duração máxima de 90 minutos para todos os participantes e foram registadas observações.

A fase dois decorreu cinco dias após a experiência de visualização. Os participantes receberam o inquérito de apreensão de informação online, que continha as mesmas seis perguntas usadas para a obtenção de informação no primeiro inquérito. Este inquérito foi concluído pelos participantes em casa no seu próprio horário. Os participantes tiveram dois dias, no total, para completar o inquérito, para que o mesmo não ocorra durante um período de tempo indefinido.

4.2.2. Análise de resultados

Os resultados dos inquéritos podem ser vistos em Anexos, secção XXII. Das oito pessoas que fizeram parte do estudo a maioria identifica-se como sendo do sexo masculino (62,5%, ou seja, cinco são homens), têm idades compreendidas entre os 16 e os 26 anos, são de nacionalidade portuguesa, com grau de literacia variado (sendo que dois dos participantes são estudantes a tempo inteiro) e de diferentes áreas do conhecimento. (ciências e tecnologias, informática, artes e línguas e humanidades). Dos 8 participantes, 75% tem experiência a visualizar informação e apenas 25% tem experiência com ambientes imersivos virtuais móveis, realidade virtual ou realidade aumentada.

1.^a Codificação Visual

Com esta visualização pretendíamos saber qual a maior temperatura registada. Como vemos na Tabela 4-3, o valor correto é 35 C°. No entanto, o Data-Room demonstrou 34 C° como o valor correto, dada uma limitação da visualização que, quando há dois ou mais dados com o mesmo ano e mês, apenas o último a ser encontrado será renderizado. Um total de 50% das pessoas indicou que a codificação mais fácil para conseguir saber qual é o maior valor é a do *Data-Room gráfico de barras*. No entanto, apenas duas das oito pessoas escolheram 35 C°, enquanto 5 escolheram 34 C° (62,5%). As primeiras, segundo o *feedback* que nos foi dado durante o inquérito, assumiram que a escala na codificação criada com *Python Mayavi barchart()* apresentava o valor mínimo e o máximo, no entanto, os outros participantes só conseguiram ter completa certeza quando viram com o *Data-Room gráfico de barras*. Também houve quem escolhesse “nenhum desses” por não ter encontrado 34 C° no *Data-Room gráfico de barras* e não confiar que a escala no *Python Mayavi barchart()* apresentava o mínimo e o máximo.

3.^a Codificação Visual

Esta visualização teve como objetivo saber qual o menor valor de temperatura registado. Como vemos na Tabela 4-5, o valor correto é -67 C°. No entanto, o Data-Room mostrou -61 C° como o valor correto, dada uma limitação da visualização já referida na Codificação 1. Um total de 75% das pessoas indicou que a codificação mais fácil para conseguir saber qual é o maior valor é a do *Data-Room gráfico de barras*. No entanto, 7 (88%) das 8 pessoas escolheram -61 C°, que é possível de ser observado no Data-Room.

3.^a Codificação Visual

A visualização visou saber qual o último ano desde que há registos. Como vemos na Tabela 4-6, o valor correto é 2010. A maioria, isto é, 7 dos participantes (88%) indicaram que a codificação mais fácil para conseguir saber o último ano em que temos registos é a do *R latticeExtra*. Não houve dúvidas que esse ano é 2010. No geral a codificação 3 criada com Python foi ignorada para este propósito.

4.^a Codificação Visual

Com esta visualização desejávamos saber o intervalo de variação de temperatura. Como vemos na Tabela 4-8, o valor correto seria entre 3 C° e 25 C°. Metade

das pessoas indicou que a codificação mais fácil para conseguir saber o intervalo de variação de temperatura é a do *Python Mayavi points3d()*. No entanto, apenas duas das oito pessoas (25%) escolheram 35 C°, enquanto 5 escolheram 34 C° (62,5%). Neste caso percebemos que o *Data-Room gráfico de bolhas* seria melhor se tivesse uma espécie de régua para que conseguissem balizar os valores.

5.^a Codificação Visual

Nesta visualização quisemos saber qual a maior temperatura. Como vemos na Tabela 4-9, o valor correto é 23 C°. 75% das pessoas indicaram que a codificação mais fácil para conseguir saber qual é o maior valor é a *Data-Room gráfico de barras*. No entanto, apenas quatro das oito pessoas escolheram 23 C° (50%). As primeiras, segundo o *feedback* que nos foi dado durante o inquérito, assumiram que a escala na codificação criada com o *Python Mayavi barchart()* apresentava o valor mínimo e o máximo, no entanto, os outros participantes só conseguiram ter completa certeza quando viram com o *Data-Room gráfico de barras*. Também houve quem escolhesse “nenhum desses” por não ter encontrado 34 C° no *Data-Room gráfico de barras* e não confiar que no *Python Mayavi bubblechart()* a escala de cor apresentava valores entre o mínimo e o máximo. No geral a codificação 1 criada em R foi ignorada para este propósito porque o ângulo escolhido não era o mais indicado.

6.^a Codificação Visual

A sexta visualização pretendeu saber qual o mês com a temperatura mais alta (julho). Cinco participantes indicaram que a codificação mais fácil era a do *R plot3d()*, sendo que apenas três destes participantes acertaram que era o mês de julho.

A partir da codificação 4, os participantes sentiram-se quase como forçados a assumir que a escala visível em *Python Mayavi* compreendia mesmo os valores mínimos e máximos, pois caso contrário teriam muita dificuldade de arranjar respostas para as perguntas, ou seja, aquela desconfiança inicial, ainda que presente, já não existia em grande quantidade como no início. Além de que com essa assunção, responder à questão tornar-se-ia mais fácil. Os participantes tinham tempo limitado e queriam finalizar o inquérito de forma mais rápida. A tabela 4-11 lista as observações relatadas pelos participantes em relação ao *Data-Room*.

Tabela 4-11. Registo de observações deixadas pelos participantes após utilizar o Data-Room.

Participante	Observações
1	<ul style="list-style-type: none"> • O ecrã do telemóvel desligou. • Demorou a clicar no botão “VOAR”. • Ia caíndo por estar perto de uma cadeira.
2	<ul style="list-style-type: none"> • Demorou a clicar no botão “VOAR”. • Enjoou a partir da terceira visualização imersiva.
3	<ul style="list-style-type: none"> • Doeu à cara. • Ficou zonha passados 10 minutos de imersão. • Não percebia que, no Data-Room gráfico de barras, as barras iam em direções diferentes.
4	<ul style="list-style-type: none"> • Bastante entusiasmado. • Um bocado de <i>lag</i>. • Marca do HMD no nariz. • “Como se fecha o menu?”
5	<ul style="list-style-type: none"> • Mexeu-se muito. • Gostou muito de estar no ambiente imersivo. • “Estou debaixo de água” • “Estou num pêndulo” • “Estou às cambalhotas no espaço” • “Como posso colocar o menu no lixo?” • Data-Room gráfico de bolhas fez lembrar a série Wipeout. • “Quero uma régua para o Data-Room gráfico de bolhas”. • “Onde estão os meses? Só vejo números” • Marca do HMD no nariz. • “Sou um(a) despachado(a), faço tudo a correr, mas aqui sou forçado(a) a ser mais lento(a) e eu gosto disso”.
6	<ul style="list-style-type: none"> • Cabeça e nariz pequenos, foi necessário ajustar o HMD. • Não gostou do menu. • “Como sei que estou a voar?” • Clicou três vezes no botão 3 do menu enquanto voava.
7	<ul style="list-style-type: none"> • Falta de vista mas sabe como procurar os dados. • <i>Lag</i>. • Bolhas muito juntas. • Teve uma atualização do telemóvel enquanto respondia ao inquérito.
8	<ul style="list-style-type: none"> • Primeiro comentário interessante do participante: “Será que influencia se as pessoas têm vertigens ou claustrofobia?” • Segundo comentário interessante do participante: “Estou bem, mas pode ser da fome.”

Quando pedimos aos participantes para avaliar os ambientes, obtivemos as médias apresentadas na Tabela 4-12.

Tabela 4-12. Reações médias dos participantes (numa escala Likert de 0 a 5 pontos).

Ambiente	Empenho	Visualização	Satisfação	Concentração
Python Mayavi barchart()	4,375	4,000	3,875	3,750
R latticeExtra	4,125	3,625	3,375	3,125
Data-Room gráfico de barras	4,500	3,875	4,125	3,250
Python Mayavi poits3d()	4,250	3,750	3,625	3,750
R plot3D	3,750	3,250	3,25	2,875
Data-Room gráfico de bolhas	3,750	3,250	3,625	2,750

Sete dos oito participantes (87,5%) referiram que gostariam de ter o Data-Room no seu telemóvel. Além disso, quatro participantes (50%) deixaram os comentários disponíveis na Tabela 4-13.

Tabela 4-13. Comentários ou sugestões dos participantes.

Comentários ou sugestões
Gostaria de ter (o Data-Room no telemóvel), mas mais fluído
Gostei bastante da experiência e como visualizei estes gráficos, tirando os gráficos das bolhas que foi-me difícil compreender os vários valores, todos os outros gráficos usando o Google box (Cardboard) me pareceram bastantes úteis em vários aspetos com rápido acesso aos dados e sendo uma maneira mais interessante de o fazer.
Muito interessante para leitura de dados, de certeza que para um emprego em que seja super necessário esta ferramenta seja muito útil.
O Data-Room poderia ter o valor nas bolhas para facilitar a visualização em alguns casos.

As tabelas 4-14 e 4-15 ilustram os resultados, respectivamente, a aquisição e retenção da informação visualizada. As colunas indicam o número da codificação/pergunta e as linhas o participante. Para cada codificação existem 3 colunas; na primeira podemos ver qual o ambiente selecionado pelo participante como ideal para responder à questão, na segunda vemos a resposta à questão no primeiro inquérito (inquérito de visualização) e na terceira vemos a resposta à questão após sete dias (no inquérito de apreensão). As células pintadas a verde significam as respostas totalmente coerentes entre si. As de cor vermelha são as respostas que apresentaram incoerência entre elas e as demais. Para representar os ambientes, utilizou-se apenas as iniciais (P – Python, D – Data-Room e R) e o traço (-), para indicar que a resposta indica que nenhuma das opções de resposta apresentadas tinha o valor correto.

Nas tabelas vemos três casos em que o ambiente escolhido pelo participante foi o Python ou o R, mas onde o valor obtido e utilizado como resposta teve origem na visualização no Data-Room. Isto significa que podem ter sido escolhidos os ambientes de forma errada. Através destas tabelas criámos os gráficos 4-22, 4-23, 4-24, 4-25, 4-26 e 4-27, que ilustram de forma resumida os resultados obtidos. Estas tabelas serão úteis para medir:

- aquisição de informação - ambiente preferido e respostas corretas.
- retenção de informação - o ambiente definido como ideal e resposta igual em ambos os inquéritos.

Tabela 4-14. Análise de aquisição.

Participante	1			2		3		4		5		6						
1	P	35	43	D	-61	-61	R	2010	2010	P	3-25	4-25	P	23	23	R	agosto	agosto
2	P	-	-	P	-67	-	R	2010	2009	D	4-25	-	P	30	-	R	julho	julho
3	P	35	43	D	-61	-61	D	2010	2013	D	5-20	4-25	R	23	23	R	maio	agosto
4	D	34	-	P	-61	-	R	2010	2009	P	3-25	5-20	P	23	23	P	agosto	maio
5	R	34	34	D	-61	-	R	2010	2010	R	3-25	3-25	P	30	30	R	julho	julho
6	D	34	35	D	-61	-67	R	2010	2010	P	3-25	4-25	P	23	23	P	agosto	julho
7	D	34	34	D	-61	-67	R	2010	2009	P	4-25	-	P	-	30	R	julho	agosto
8	D	34	34	D	-61	-61	R	2010	2010	D	-	-	D	30	30	D	agosto	agosto

Tabela 4-15. Análise de retenção.

Participante	1			2		3		4		5		6						
1	P	35	43	D	-61	-61	R	2010	2010	P	3-25	4-25	P	23	23	R	agosto	agosto
2	P	-	-	P	-67	-	R	2010	2009	D	4-25	-	P	30	-	R	julho	julho
3	P	35	43	D	-61	-61	D	2010	2013	D	5-20	4-25	R	23	23	R	maio	agosto
4	D	34	-	P	-61	-	R	2010	2009	P	3-25	5-20	P	23	23	P	agosto	maio
5	R	34	34	D	-61	-	R	2010	2010	R	3-25	3-25	P	30	30	R	julho	julho
6	D	34	35	D	-61	-67	R	2010	2010	P	3-25	4-25	P	23	23	P	agosto	julho
7	D	34	34	D	-61	-67	R	2010	2009	P	4-25	-	P	-	30	R	julho	agosto
8	D	34	34	D	-61	-61	R	2010	2010	D	-	-	D	30	30	D	agosto	agosto

Na Tabela 4-14 e 4-15 vemos um total de três casos assinalados a vermelho, significando que o ambiente escolhido como ideal pelo participante foi o Python ou R, mas onde o valor foi obtido e utilizado como resposta veio através do Data-Room. Isto indica que os participantes podem ter escolhido o ambiente de forma errada. Através destas tabelas e considerando as técnicas de visualização utilizadas (gráfico de barras e gráfico de bolhas), criámos as Figuras 4-22, 4-23, 4-24 e 4-25. Na Figura 4-22, o gráfico A demonstra as preferências, sem considerar a possível escolha errada do ambiente, enquanto que o gráfico B considera que todos os casos assinalados a vermelho escolheram de forma errada o ambiente (feita com base na visualização disponível no Data-Room, porém o ambiente indicado como ideal foi outro). Na Figura 4-23, o Data-Room gráfico de bolhas aparece assinalado com 0%. Na Figura 4-25, o gráfico A demonstra as preferências, sem considerar a possível escolha errada do ambiente, enquanto que o

gráfico B considera que todos os casos assinalados a vermelho escolheram de forma errada o ambiente (feita com base na visualização disponível no Data-Room, porém o ambiente indicado foi outro).

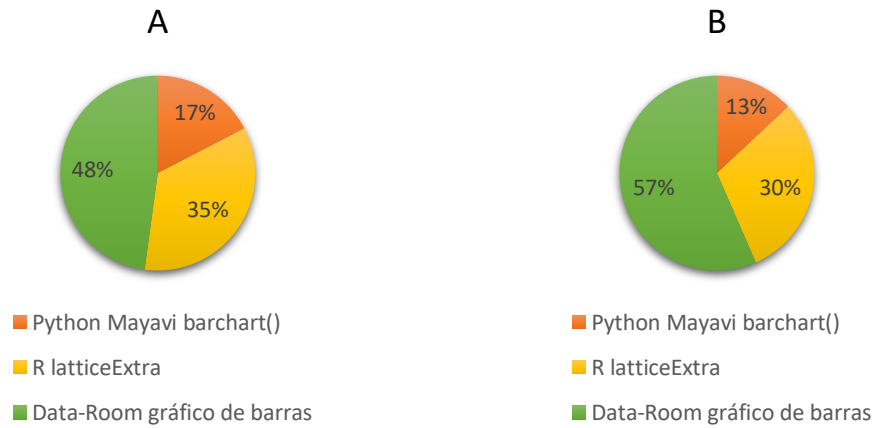


Figura 4-22. Ambiente preferido e resposta certa, considerando a técnica gráfico de barras.

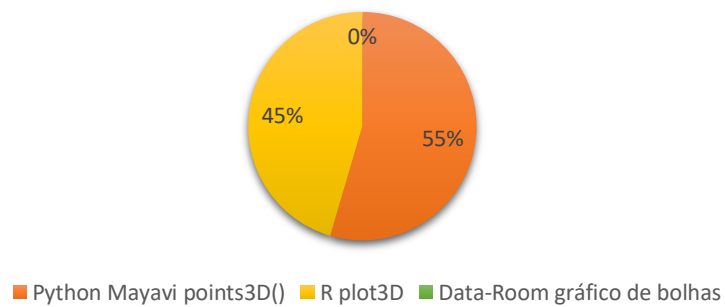


Figura 4-23. Ambiente preferido e resposta certa, considerando a técnica gráfico de bolhas.

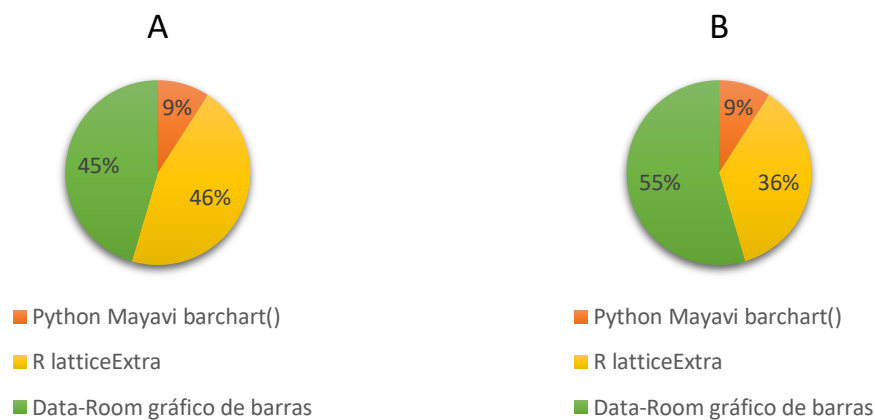


Figura 4-24. Ambiente preferido com resposta igual em ambos os inquéritos, considerando a técnica gráfico de barras.

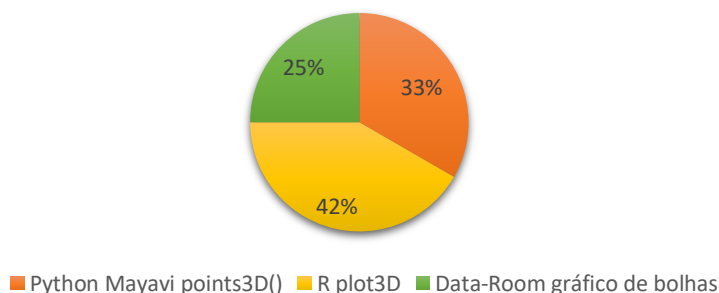


Figura 4-25. Ambiente preferido com resposta igual em ambos os inquéritos, considerando a técnica gráfico de bolhas.

Com base nas perguntas existentes nos inquéritos, foram elaborados os gráficos ilustrados nas figuras 4-26 a 4-31. Através dessas figuras conseguimos perceber que:

- Os eixos é uma das pistas mais apreciadas (Tabela 4-16, Figura 4-26) e que, a capacidade de fazer Zoom foi uma das pistas menos apreciadas. No tocante à retenção, todavia, a cor foi a pista que melhor se reteve na memória (Figura 4-31);

Tabela 4-16. Análise de apreciação das pistas visuais.

Codificação	Pistas					
	Eixos	Cores	Visão Panorâmica	Imersão	Zoom	Quantidade de dados
1	2°	3°	2°	3°	2°	1°
2	2°	3°	4°	3°	2°	1°
3	3°	1°	3°	1°	2°	2°
4	1°	2°	2°	3°	4°	4°
5	3°	1°	2°	3°	3°	4°
6	1°	3°	2°	2°	4°	3°
Média	2°	2,2°	2,5°	2,5°	2,8°	2,5°

- Para a visualização de gráficos de barras, o Data-Room é o ambiente que permite uma exploração mais fácil com 58% dos votos (Figura 4-27) e o mais apetecível de ser visualizado no ecrã de um telemóvel com 54% dos votos (Figura 4-29);
- Para a visualização de gráficos de bolhas, o Python é o ambiente que permite uma exploração mais fácil com 50% dos votos (Figura 4-28) e o mais apetecível de ser visualizado no ecrã de um telemóvel com 50% dos votos (Figura 4-30).

Pergunta: O que consideram relevante nas visualizações?

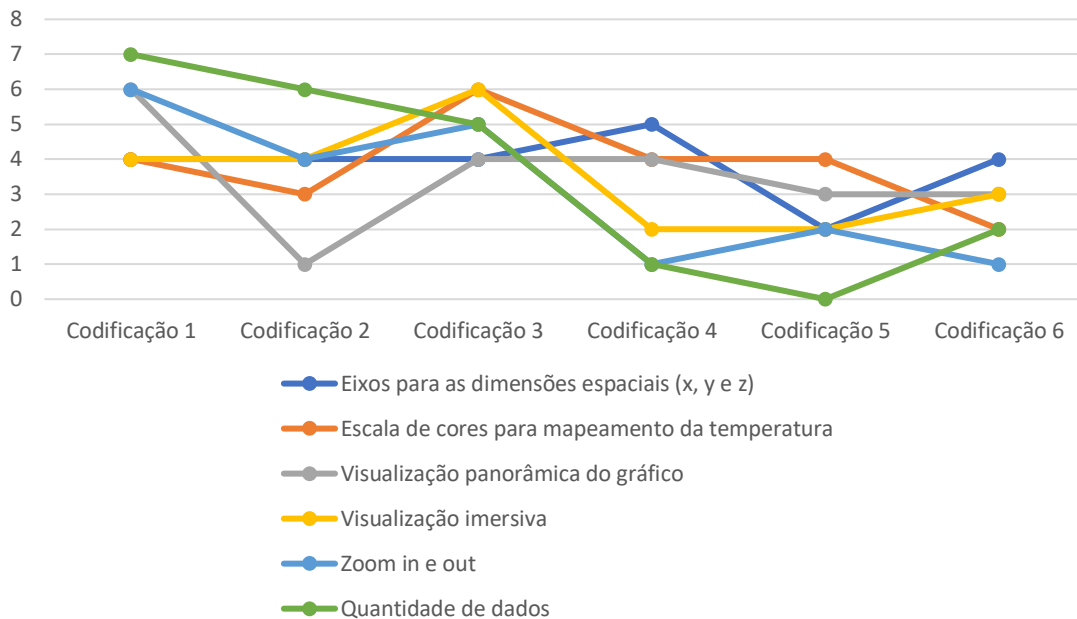


Figura 4-26. Pistas visuais relevantes nas visualizações.

Pergunta: Qual dos ambientes permite uma exploração mais fácil?

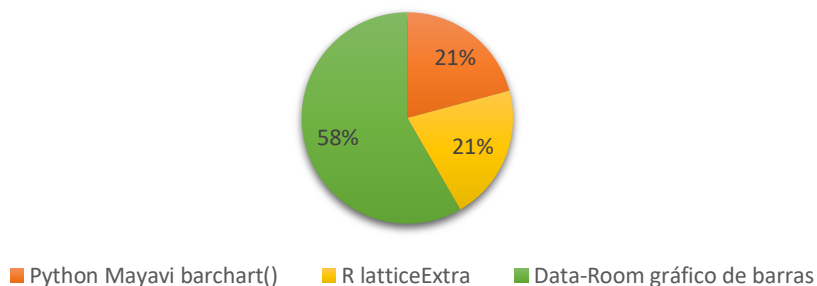


Figura 4-27. Ambientes que permitem uma exploração mais fácil, considerando a técnica gráfico de barras.

Pergunta: Qual dos ambientes permite uma exploração mais fácil?

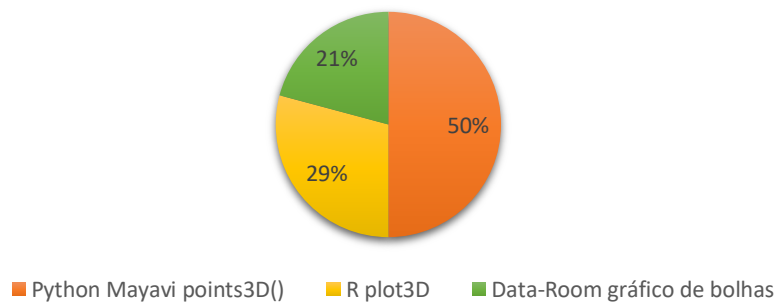


Figura 4-28. Ambientes que permitem uma exploração mais fácil, considerando a técnica gráfico de bolhas.

Pergunta: Qual ambiente é mais apetecível ser visualizado no ecrã de um telemóvel?

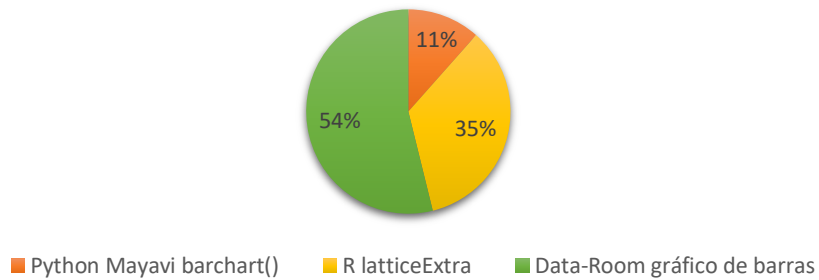


Figura 4-29. Ambientes mais apetecíveis de visualizar no ecrã de um telemóvel, considerando a técnica gráfico de barras.

Pergunta: Qual ambiente é mais apetecível ser visualizado no ecrã de um telemóvel?

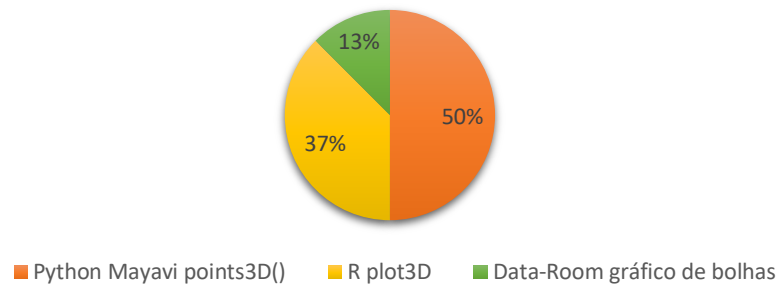


Figura 4-30. Ambientes mais apetecíveis de visualizar no ecrã de um telemóvel, considerando a técnica gráfico de bolhas.

Pergunta: Qual das pistas visuais retém melhor na memória?

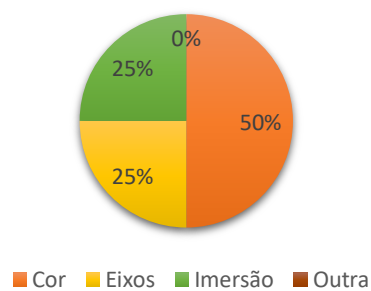


Figura 4-31. Pistas visuais e retenção na memória (resposta “Outra” está assinalada com 0%).

5. Conclusão

Os resultados obtidos nesta dissertação indicam que os ambientes imersivos virtuais móveis são um método válido e eficaz para a visualização de informação. Aliás, os mesmos demonstram que os ambientes imersivos virtuais móveis são potencialmente tão ou mais eficazes para visualizar dados do tipo *big data* comparativamente com outros métodos utilizados atualmente, sendo que, como referido na secção 1.1., definimos eficácia como:

- A capacidade de adquirir e reter conteúdo;
- A qualidade da experiência;
- Se é fácil ou não o utilizador concentrar-se.

Independentemente dos participantes terem, ou não, experiência com ambientes imersivos móveis, realidade virtual ou realidade aumentada, os valores obtidos demonstram que das vezes que foi obtida uma resposta correta às questões colocadas, 48-57% das vezes o Data-Room foi escolhido como método ideal para obter a resposta a tal resposta, enquanto que o segundo melhor método (Python) tinha 35-30%. Por outras palavras, Data-Room é o melhor ambiente para adquirir conteúdo.

Além disso, os resultados demonstram que das vezes em que as respostas foram iguais em ambos os inquéritos (inquéritos, estes, realizados com uma diferença de sete dias entre eles), 45-55% o Data-Room foi usado como método ideal para obter a resposta, enquanto o segundo melhor método (Python) tinha 36-46%. Mesmo não estando a resposta correta, a informação extraída pelos mesmos nesse ambiente aguentou-se na memória mais tempo do que através dos outros métodos. Com isto, vemos que o Data-Room é o ambiente mais capaz para reter informação adquirida.

A nossa investigação demonstrou que as pessoas estão muito satisfeitas com este tipo de ambientes imersivos virtuais móveis, demonstrando empenho e satisfação (Tabela 4-12). No tocante à concentração, a Tabela 4-12 considera que o Python o melhor ambiente e o Data-Room e R variando de posição consoante a técnica utilizada. Dentro do ambiente Data-Room, existe uma clara preferência para o gráfico de barras em relação ao gráfico de bolhas. Enquanto o *Data-Room gráfico de barras* foi votado como ambiente que permite uma exploração mais fácil com 50% dos votos e mais apetecível de ser visualizado no num ecrã de telemóvel (54% dos votos), o *Data-Room gráfico de bolhas* ficou áquem do esperado. Através dos inquéritos e das observações registadas (Tabela 4-11, Figura 4-26, e) compreendemos que o Data-Room gráfico de bolhas, ainda que tenha *lag* (tal como o Data-Room gráfico de barras), não foi tão eficaz, comparativamente com Python ou R, porque um dos eixos não tinha legenda (e a presença de eixos bem definidos foi a pista mais valorizada (Tabela 4-16)) e as bolhas estavam muito próximas.

Os resultados indicam que este tipo de ambientes pode ser mais eficaz para visualizar dados de forma móvel, em termos de aprender e reter informação, entusiasmo e concentração, do que visualizações criadas com Python ou R. Concluímos que os ambientes imersivos virtuais móveis podem providenciar experiências de visualização de dados melhores do que outras técnicas, ao ponto das quererem ter tais soluções nos seus telemóveis (87,5% dos partipantes).

5.1. Trabalho futuro

Mostrar os dados do tipo *big data* em 3D dinamicamente foi um desafio. Além disso, a adição de interatividade para a manipulação dos dados e respectiva análise adicionou também um outro nível de dificuldade. Tendo isso em conta, a criação de um artefacto digital melhorado com menos *lag*, capaz de abarcar mais dados, com mais e melhoradas técnicas de visualização, pistas visuais, um maior leque de acções possíveis e, inclusive, realizar testes de utilização, nomeadamente de funcionalidades e conforto, é um potencial trabalho futuro.

Os resultados obtidos tiveram como base as perguntas colocadas aos participantes. Seria interessante alargar o leque de questões, incluir perguntas que permitissem explorar a eficácia relativa de ambientes imersivos virtuais móveis com um grupo maior de participantes. Idealmente, abordar uma associação com uma empresa, como por exemplo no mundo financeiro, para que seja mais perceptível de que forma este tipo de ambientes afeta os funcionários nos seus locais de trabalho.

Explorar os resultados além da concentração, satisfação, aquisição e retenção de conhecimento também pode ser interessante. Ainda que todos os participantes tiveram o mesmo objetivo para alcançar dentro de cada codificação visual, a técnica de procura empregada pelos mesmos nem sempre foi a mesma, o tempo de resolução também foi diferente, uns precisaram de ajustar o HMD mais do que outros, etc.

6. Referências bibliográficas

- 3D Charts and Graphs, 2019. Disponível online em: <https://assetstore.unity.com/packages/3d/props/tools/3d-charts-and-graphs-114377> [acedido dia 10 de janeiro de 2020].
- 3Data, 2019. Disponível online em: <https://3data.io/> [acedido a 10 de janeiro de 2020].
- A-frame, sem data. Disponível online em: <https://aframe.io/> [acedido dia 25 de abril de 2019].
- Aghamirkarimi, D., Lemire, D., 2017. *Discovering the Smart Forests with Virtual Reality*. Disponível online em: <https://dl.acm.org/citation.cfm?id=3102254.3102287> [acedido a 2 de novembro de 2018].
- Alibaba Cloud E-MapReduce, 2020. Disponível online em: <https://www.alibabacloud.com/products/emapreduce> [acedido a 10 de janeiro de 2020].
- Altintas, I., Gupta, A. & Nguyen, M., sem data. *Big Data Specialization*. Disponível online em: <https://www.coursera.org/specializations/big-data> [acedido dia 30 de dezembro de 2018].
- Amazon EMR, 2019. Disponível online em: <https://aws.amazon.com/emr/> [acedido a 10 de janeiro de 2020].
- Apache Drill, 2014. Disponível online em: <https://drill.apache.org/> [acedido dia 19 de agosto de 2019].
- Apache Hive, 2014. Disponível online em: <https://hive.apache.org/> [acedido dia 19 de agosto de 2019].
- Apache Impala, sem data. Disponível online em: <https://impala.apache.org/> [acedido dia 19 de agosto de 2019].
- Apache MapReduce, 2019. Disponível online em: https://hadoop.apache.org/docs/r1.2.1/mapred_tutorial.html [acedido a 10 de janeiro de 2020].
- Aukstakalnis, S., 2016. *Practical Augmented Reality: A Guide to the Technologies, Applications, and Human Factors for AR and VR*. Disponível online em: <https://learning.oreilly.com/library/view/practical-augmented-reality/9780134094328/> [acedido dia 23 de abril de 2019].
- Azure HDInsight, 2020. Disponível online em: <https://azure.microsoft.com/en-us/services/hdinsight/> [acedido a 10 de janeiro de 2020].

- BadVR, 2019. Disponível online em: <https://badvr.com/> [acedido dia 16 de novembro de 2019].
- Bah *et al.*, 2016. *Immersive Analytics: Exploring Future Interaction and Visualization Technologies for Data Analytics*. Disponível online em: <https://dl.acm.org/citation.cfm?id=2996365> [acedido dia 16 de novembro de 2019].
- Bermúdez *et al.*, 2005. *Between Art, Science and Technology: Data Representation Architecture*. Disponível online em: <https://www.mitpressjournals.org/doi/abs/10.1162/0024094054762061> [acedido a 2 de novembro de 2018].
- Bowman, D, McMahan, R., 2007. *Virtual Reality: How Much Immersion Is Enough*. Disponível online em: <https://ieeexplore.ieee.org/document/4287241/> [acedido dia 16 de novembro de 2019].
- Bridge by Occipital, 2020. Disponível online em: <https://bridge.occipital.com> [acedido dia 19 de agosto de 2019].
- Brunetti, J., Auer, S., García, R., 2012. *The Linked Data Visualization Model*. Disponível online em: https://www.researchgate.net/publication/268485197_The_Linked_Data_Visualization_Model [acedido dia 6 de maio de 2019].
- Bui, Q., Wolfers, J. (2020). More Than 3 Million Americans Lost Their Jobs Last Week. See Your State. Disponível online em: <https://www.nytimes.com/interactive/2020/03/26/upshot/coronavirus-millions-unemployment-claims.html> [acedido dia 11 de maio de 2020].
- Cardboard, sem data. Disponível online em: <https://developers.google.com/vr/discover/cardboard> [acedido a 10 de janeiro de 2020].
- Cardboard da marca Splaks, sem data. Disponível online em: <https://www.amazon.com/Cardboard-Splaks-Virtual-Reality-Forehead/dp/B01LWSSQTY> [acedido dia 25 de abril de 2019].
- Cavique, L., Marques, N., Santos, J., 2014. *An Algorithm to Condense Social Networks and Identify Brokers*. Disponível online em: https://link.springer.com/chapter/10.1007%2F978-3-319-12027-0_27 [acedido a 28 de novembro de 2017].
- CDH, 2020. Disponível online em: <https://www.cloudera.com/products/open-source/apache-hadoop/key-cdh-components.html> [acedido a 10 de janeiro de 2020].

- Chen, C., 1999. *Information Visualization and Virtual Environments*. Disponível online em: <https://www.springer.com/gp/book/9781447136224> [acedido dia 5 de maio de 2019].
- Chen, C., 2006. *Information Visualization: Beyond the Horizon*. Disponível online em: <https://www.springer.com/gp/book/9781852337896> [acedido dia 5 de maio de 2019].
- Chen, H., Chiang, R., Storey, V., 2012. *Business Intelligence and Analytics: From Big Data to Big Impact*. Disponível online em: <https://www.jstor.org/stable/41703503> [acedido a 2 de novembro de 2018].
- Chen, M., Golan, A., 2016. *What may visualization processes optimize?*. Disponível online em: <https://ieeexplore.ieee.org/document/7368928/> [acedido a 15 de outubro de 2018].
- Chen *et al.*, 2017. *Cost-benefit Analysis of Visualization in Virtual Environments*. Disponível online em: <https://arxiv.org/abs/1802.09012> [acedido a 22 de outubro de 2018].
- Chi, E., 2002. *Data State Reference Model*. Disponível online em: https://link.springer.com/chapter/10.1007/978-94-017-0573-8_2 [acedido dia 6 de maio de 2019].
- Cho *et al.*, 2014. *Big Data Analysis with Interactive Visualization using R packages*. Disponível online em: <https://dl.acm.org/citation.cfm?id=2644168> [acedido a 2 de novembro de 2018].
- Cinema 4D, 2020. Disponível online em: <https://www.maxon.net/en-us/try/> [acedido a 11 de maio de 2020].
- Cleveland, W., McGill, R., 1984. *Graphical Perception: Theory, Experimentation, and Application to the Development of Graphical Methods*. Disponível online em: <https://www.jstor.org/stable/2288400> [acedido dia 24 de abril de 2019].
- Cukier, K., 2010. *Data, Data Everywhere: A Special Report on Managing Information*. Disponível online em: <https://www.economist.com/special-report/2010/02/27/data-data-everywhere> [acedido dia 22 de abril de 2019].
- D3.js, 2019. Disponível online em: <https://d3js.org/> [acedido a 10 de janeiro de 2020].
- Dali, sem data. Disponível online em: <http://www.dali-vr.com/> [acedido dia 16 de novembro de 2019].
- DataView VR, sem data. Disponível online em: <https://www.dataviewvr.com/> [acedido dia 3 de maio de 2019].

- Datavized, 2018. Disponível online em: <https://datavized.com/> [acedido dia 16 de novembro de 2019].
- DatavizVR, 2016. Disponível online em: https://store.steampowered.com/app/551960/DatavizVR_Demo/ [acedido dia 16 de novembro de 2019].
- Daydream, sem data. Disponível online em: <https://developers.google.com/vr/discover/daydream> [acedido a 10 de janeiro de 2020].
- Dean, J., Ghemawat, S., 2004. *MapReduce: Simplified Data Processing on Large Clusters*. Disponível online em: <https://dl.acm.org/doi/10.1145/1327452.1327492> [acedido dia 23 de maio de 2019].
- Deloitte, 2020. Disponível online em: <https://www2.deloitte.com/global/en.html> [acedido a 10 de janeiro de 2020].
- Dietz, S., 2002. *Ten dreams of technology*. Disponível online em: <https://www.jstor.org/stable/1577257> [acedido a 7 de outubro de 2018].
- Dittrich, J, Quiané-Ruiz, J., 2014. *Efficient Big Data Processing in Hadoop MapReduce*. Disponível online em: <https://dl.acm.org/citation.cfm?id=2367562&dl=ACM&coll=DL> (acedido dia 11 de dezembro de 2018);
- Donalek *et al.*, 2013. *The MICA Experiment: Astrophysics in Virtual Worlds*. Disponível online em: <https://arxiv.org/abs/1301.6808> [acedido dia 16 de novembro de 2019].
- Donalek *et al.*, 2014. *Immersive and Collaborative Data Visualization Using Virtual Reality Platforms*. Disponível online em: <http://ieeexplore.ieee.org/abstract/document/7004282/> [acedido a 28 de novembro de 2017].
- Drossis, G., Margetis, G., Stephanidis, C., 2016. *Towards Big Data Interactive Visualization in Ambient Intelligence Environments*. Disponível online em: https://link.springer.com/chapter/10.1007/978-3-319-39862-4_6 [acedido a 2 de novembro de 2018].
- Durham, G., sem data. *Foundations for Big Data Analysis with SQL*. Disponível online em: <https://www.coursera.org/learn/foundations-big-data-analysis-sql> [acedido dia 19 de agosto de 2019].
- DXR, sem data. Disponível online em: <https://sites.google.com/view/dxr-vis> [acedido a 11 de novembro de 2018].

- Febretti *et al.*, 2013. *CAVE2: A Hybrid Reality Environment for Immersive Simulation and Information Analysis*. Disponível online em: https://www.researchgate.net/publication/258813622_CAVE2_A_Hybrid_Reality_Environment_for_Immersive_Simulation_and_Information_Analysis [acedido dia 16 de novembro de 2019].
- Fry, B., Maeda, J., 2004. *Computational information design*. Disponível online em: <https://dl.acm.org/citation.cfm?id=1037443> [acedido a 5 de novembro de 2018].
- Gear VR, sem data. Disponível online em: <https://www.samsung.com/global/galaxy/gear-vr> [acedido a 10 de janeiro de 2020].
- Geometric, 2018. Disponível online em: <http://geometric.space/> [acedido dia 16 de novembro de 2019].
- Glass Brain, sem data. Disponível online em: <https://neuroscape.ucsf.edu/technology/#glass-brain> [acedido dia 25 de abril de 2019].
- Google Cloud Dataproc, sem data. Disponível online em: <https://cloud.google.com/dataproc/> [acedido a 10 de janeiro de 2020].
- Graph and Chart, 2019. Disponível online em: <https://assetstore.unity.com/packages/tools/gui/graph-and-chart-78488> [acedido dia 10 de janeiro de 2020].
- HammerHeadVR, 2018. Disponível online em: <https://www.hammerheadvr.com/> [acedido dia 4 de abril de 2019].
- Hearst, M., Karadi, C., 1997. *Cat-a-Cone: an interactive interface for specifying searches and viewing retrieval results using a large category hierarchy*. Disponível online em: <https://dl.acm.org/citation.cfm?id=258582> [acedido dia 5 de maio de 2019].
- Hemmje, M., Kunkel, C., Willett, A., 1994. *LyberWorld – a visualization user interface supporting fulltext retrieval*. Disponível online em: <https://dl.acm.org/citation.cfm?id=188563> [acedido dia 5 de maio de 2019].
- Immersion Analytics, 2019. Disponível online em: <https://www.immersionanalytics.com/> [acedido dia 16 de novembro de 2019].
- Kaggle NOAA GSOD, 2019. Disponível online em: <https://www.kaggle.com/noaa/g sod> [acedido dia 16 de novembro de 2019].
- Kammer *et al.*, 2018. *VisBIA 2018: Workshop on Visual Interfaces for Big Data Environments in Industrial Applications*. Disponível online em: <https://dl.acm.org/citation.cfm?id=3206603> [acedido a 2 de novembro de 2018].

- Keim, D., 2002. *Information Visualization and Visual Data Mining*. Disponível online em: <https://dl.acm.org/citation.cfm?id=614508> [acedido dia 16 de novembro de 2019].
- Keim *et al.*, 2010. *Mastering the Information Age - Solving Problems with Visual Analytics*. Disponível online em: https://www.researchgate.net/publication/277007765_Mastering_The_Information_Age_-_Solving_Problems_with_Visual_Analytics [acedido a 22 de outubro de 2018].
- Keim, D., Mansmann, F., Thomas, J., 2010. *Visual analytics: How much visualization and how much analytics*. Disponível online em: <https://dl.acm.org/doi/10.1145/1809400.1809403> [acedido a 28 de novembro de 2017].
- Landon-Murray, M., 2016. *Big Data and Intelligence: Applications, Human Capital, and Education*. Disponível online em: <https://scholarcommons.usf.edu/jss/vol9/iss2/6/> [acedido a 22 de outubro de 2018].
- Laney, D., 2001. *3D Data Management: Controlling Data Volume, Velocity, and Variety*. Disponível online em: <http://blogs.gartner.com/doug-laney/files/2012/01/ad949-3D-Data-Management-Controlling-Data-Velocity-and-Variety.pdf> [acedido dia 22 de abril de 2019].
- Levy, B., 2016. *How does virtual reality contribute to data visualization?*. Disponível online em: <https://www.quora.com/How-does-virtual-reality-contribute-to-data-visualization/answer/Bob-Levy-4> [acedido dia 16 de novembro de 2019].
- Linowes, J., Schoen, M., 2016. *Cardboard VR Projects for Android*. Disponível online em: <https://www.packtpub.com/eu/application-development/cardboard-vr-projects-android> [acedido dia 10 de janeiro de 2020].
- Linowes, J., 2018. *Unity Virtual Reality Projects – Second Edition*. Disponível online em: <https://www.packtpub.com/eu/game-development/unity-virtual-reality-projects-second-edition> [acedido dia 10 de janeiro de 2020].
- Liu *et al.*, 2014. *A survey on information visualization: recent advances and challenges*. Disponível online em: <https://link.springer.com/article/10.1007%2Fs00371-013-0892-3> [acedido a 28 de novembro de 2017].
- Liu *et al.*, 2017. *Visualizing High-Dimensional Data: Advances in the Past Decade*. Disponível online em: <http://ieeexplore.ieee.org/document/7784854/> [acedido a 28 de novembro de 2017].
- magicleapdevs (14 de novembro de 2019), “Climate Data is a new concept on Magic Leap World. We believe devs & #spatialcomputing can raise global awareness,

- improve understanding of complex data & systems, and make a positive impact on sustainable development goals. Learn more [@UNDP @BadVR_Inc](https://magi.ca/climatetw@UNDP)". Disponível online em: <https://twitter.com/magicleapdevs/status/1195040416028073985?s=20> [acedido dia 16 de novembro de 2019].
- MapR, 2020. Disponível online em: <https://mapr.com/> [acedido a 10 de janeiro de 2020].
- Melnik *et al.*, 2011. *Dremel: Interactive Analysis of Web-Scale Datasets*. Disponível online em: <https://research.google/pubs/pub36632/> [acedido dia 16 de outubro de 2019].
- Merge VR, 2020. Disponível online em: <https://mergevr.com/headset#> [acedido dia 19 de agosto de 2019].
- Microsoft – Documentação, 2019. *Extract, transform, and load (ETL)*. Disponível online em: <https://docs.microsoft.com/en-us/azure/architecture/data-guide/relational-data/etl> [acedido dia 25 de abril de 2019].
- Miller, J., 2017. *Big Data Visualization*. Disponível online em: <https://www.packtpub.com/big-data-and-business-intelligence/big-data-visualization> [acedido dia 22 de abril de 2019].
- Mohan, V., Lukas, S., Pangilinan, E., 2019. *Creating Augmented and Virtual Realities*. Disponível online em: <https://learning.oreilly.com/library/view/creating-augmented-and/9781492044185/> [acedido dia 4 de abril de 2019].
- Moran *et al.*, 2015. *Improving Big Data visual analytics with interactive virtual reality*. Disponível online em: <https://ieeexplore.ieee.org/document/7322473> [acedido dia 16 de novembro de 2019].
- Munzner, T., 1998. *Exploring large graphs in 3D hyperbolic space*. Disponível online em: <https://ieeexplore.ieee.org/abstract/document/689657> [acedido dia 5 de maio de 2019].
- Munzner, T., 2015. *Visualization Analysis and Design*. Disponível online em: <https://www.cs.ubc.ca/~tmm/vadbook/> [acedido dia 10 de abril de 2019].
- Nanome, 2020. Disponível online em: <https://nanome.ai/> [acedido dia 16 de novembro de 2019].
- NChart3D, 2020. Disponível online em: <https://nchart3d.com/nchart/> [acedido a 10 de janeiro de 2020].
- Oracle Big Data Cloud Service, 2019. Disponível online em: <https://www.oracle.com/big-data/big-data-cloud-service/> [acedido a 10 de janeiro de 2020].

- Orson Charts For Android, 2017. Disponível online em: <http://www.object-refinery.com/orsoncharts/android/index.html> [acedido dia 10 de janeiro de 2020].
- Presto, 2019. Disponível online em: <http://prestodb.github.io/> [acedido dia 19 de agosto de 2019].
- ProChart, 2017. Disponível online em: <https://assetstore.unity.com/packages/tools/gui/prochart-46203> [acedido dia 10 de janeiro de 2020].
- Quick Chart, 2017. Disponível online em: <https://assetstore.unity.com/packages/tools/modeling/quick-chart-20406> [acedido dia 10 de janeiro de 2020].
- React, 2020. Disponível online em: <https://reactjs.org/> [acedido a 10 de janeiro de 2020].
- Reda *et al.*, 2013. *Visualizing Large, Heterogeneous Data in Hybrid-Reality Environments*. Disponível online em: <https://ieeexplore.ieee.org/abstract/document/6497028> [acedido dia 23 de abril de 2019].
- Ribeca, S., 2018. *Who's Offering Data Visualization VR?*. Disponível online em: <https://datavizcatalogue.com/blog/whos-offering-data-visualization-vr/> [acedido dia 16 de novembro de 2019].
- Rodrigues *et al.*, 2017. *Massive Data Visualization Analysis Analysis of current visualization techniques and main challenges for the future*. Disponível online em: <http://ieeexplore.ieee.org/document/7975704/> [acedido a 28 de novembro de 2017].
- Ronthal, A., Edjlali, R., Greenwald, R., 2019. *Magic Quadrant for Data Management Solutions for Analytics*. Disponível online em: <https://www.gartner.com/doc/reprints?id=1-3U1LC65&ct=170222&st=sb> [acedido dia 20 de maio de 2019].
- Shneiderman, B., 2003. *Why Not Make Interfaces Better than 3D Reality?*. Disponível online em: https://www.researchgate.net/publication/3209133_Why_Not_Make_Interfaces_Better_than_3D_Reality_2004 [acedido dia 5 de maio de 2019].
- Schultze, J., Ramamoorthi, R., sem data. *Virtual Reality (VR) App Development*. Disponível online em: <https://www.edx.org/professional-certificate/virtual-reality-vr-app-development> [acedido dia 30 de dezembro de 2018].
- Second Life, 2020. Disponível online em: <https://secondlife.com/> [acedido dia 16 de novembro de 2019].

- SeeSignal, 2019. Disponível online em: <https://www.seesignal.com/> [acedido dia 16 de novembro de 2019].
- Shiravi, H., Shiravi, A., Ghorbani, A., 2012. *A survey of visualization systems for network security*. Disponível online em: <http://ieeexplore.ieee.org/document/6007132/> [acedido a 28 de novembro de 2017].
- Sicat *et al.*, 2018. *DXR: A Toolkit for Building Immersive Data Visualizations*. Disponível online em: <https://ieeexplore.ieee.org/document/8440858> [acedido a 7 de novembro de 2018].
- Simon, P., 2014. *The Visual Organization: Data Visualization, Big Data, and the Quest for Better Decisions*. Disponível online em: <https://www.wiley.com/en-cw/The+Visual+Organization%3A+Data+Visualization%2C+Big+Data%2C+and+the+Quest+for+Better+Decisions-p-9781118858349> [acedido dia 24 de abril de 2019].
- Slater, M., Sanchez-Vives, M., 2016. *Enhancing Our Lives with Immersive Virtual Reality*. Disponível online em: <https://www.frontiersin.org/articles/10.3389/frobt.2016.00074/full> [acedido dia 16 de novembro de 2019].
- Spence, R., 2007. *Information Visualization: Design for Interaction (2nd Edition)*. Disponível online em: <https://dl.acm.org/citation.cfm?id=1196684> [acedido a 7 de novembro de 2018].
- Statista, novembro de 2016. *Number of mobile phone users worldwide from 2015 to 2020 (in billions)*. Disponível online em: <https://www.statista.com/statistics/274774/forecast-of-mobile-phone-users-worldwide/> [acedido a 22 de outubro de 2018].
- Statista, janeiro de 2017. *Sales of virtual reality head-mounted displays worldwide in 2016 and 2020 (in million units)*. Disponível online em: <https://www.statista.com/statistics/697159/head-mounted-display-unit-sales-worldwide/> [acedido a 22 de outubro de 2018].
- Statista, março 2018. *Big data revenue worldwide from 2016 to 2027, by major segment (in billion U.S. dollars)*. Disponível online em: <https://www.statista.com/statistics/301566/big-data-factory-revenue-by-type/> [acedido a 15 de outubro de 2018].
- Statista, dezembro de 2018. *Volume of data/information created worldwide from 2005 to 2025 (in zetabytes)*. Disponível online em: <https://www.statista.com/statistics/871513/worldwide-data-created/> [acedido a 22 de outubro de 2018].
- Statista, novembro de 2019. *Forecast augmented (AR) and virtual reality (VR) market size worldwide from 2016 to 2023 (in billion U.S. dollars)*. Disponível online

- em: <https://www.statista.com/statistics/591181/global-augmented-virtual-reality-market-size/> [acedido a 10 de janeiro de 2020].
- Sviokla, J., 2009. *Swimming in Data? Three Benefits of Visualization*. Disponível online em: <https://hbr.org/2009/12/swimming-in-data-three-benefit> [acedido dia 24 de abril de 2019].
- TeeChart Java Android, sem data. Disponível online em: https://www.steema.com/product/java_android [acedido a 10 de janeiro de 2020].
- Telemóveis Daydream-ready, sem data. Disponível online em: <https://vr.google.com/daydream/smartphonevr/phones/> [acedido dia 25 de abril de 2019].
- Tham *et al.*, 2018. *Understanding Virtual Reality: Presence, Embodiment, and Professional Practice. reality*. Disponível online em: <https://ieeexplore.ieee.org/document/8316900> [acedido dia 16 de novembro de 2019].
- Trifacta, 2020. Disponível online em: <https://www.trifacta.com/> [acedido a 10 de janeiro de 2020].
- Tufte, E., 1983. *The Visual Display of Quantitative Information*. Disponível online em: https://www.edwardtufte.com/tufte/books_vdqi [acedido dia 4 de abril de 2019].
- Unity, 2020. Disponível online em: <https://unity3d.com/> [acedido a 10 de janeiro de 2020].
- Unity – API de *scripting*, 2019. Disponível online em: <https://docs.unity3d.com/ScriptReference/> [acedido a 7 de novembro de 2018].
- Unity - Asset Store, 2020. Disponível online em: <https://assetstore.unity.com/> [acedido a 10 de janeiro de 2020].
- Unity - Documentação, 2019. Disponível online em: <https://docs.unity3d.com/Manual/index.html> [acedido a 7 de novembro de 2018].
- Unreal Engine Marketplace, 2020. Disponível online em: <https://www.unrealengine.com/marketplace/store> [acedido a 10 de janeiro de 2020].
- Viegas, V., Wattenberg, M., 2018. *Visualization for Machine Learning*. Disponível online em: <https://www.facebook.com/nipsfoundation/videos/203530960558001/> [acedido dia 4 de abril de 2019].

- Virtualitics, 2020. Disponível online em: <https://www.virtualitics.com/> [acedido a 10 de janeiro de 2020].
- Virtualitics – Documentação, 2020. Disponível online em: <https://docs.virtualitics.com/plot-types/line-plot/> [acedido a 10 de janeiro de 2020].
- VR Ulysses, sem data. Disponível online em: <https://www.vrulysses.com/> [acedido dia 16 de novembro de 2019].
- VR Ulyses – Twitter, 2018. Disponível online em: https://twitter.com/Ulysses_VR_Data [acedido dia 16 de novembro de 2019].
- VR-VIZ, 2019. Disponível online em: <https://github.com/mustafasaifee42/VR-Viz> [acedido dia 25 de abril de 2019].
- Wang, L., Wang, G., Alexander, C., 2015. *Big data and visualization: Methods, challenges and technology progress*. Disponível online em: <http://pubs.sciepub.com/dt/1/1/7/> [acedido a 28 de novembro de 2017].
- Ware, C., Franck, G., 1996. *Evaluating stereo and motion cues for visualizing information nets in three dimensions*. Disponível online em: <https://dl.acm.org/citation.cfm?id=234975> [acedido dia 5 de maio de 2019].
- Williams, S., 2013. *Immersive visualization*. Disponível online em: <http://www.pnas.org/content/110/12/4438> [acedido a 15 de outubro de 2018].
- Yadav, V., 2017. *Processing Big Data with Azure HDInsight: Building Real-World Big Data Systems on Azure HDInsight Using the Hadoop Ecosystem*. Disponível online em: <https://www.springer.com/de/book/9781484228685> [acedido dia 23 de maio de 2019].
- Yi, J., Kang, Y., Stasko, J., 2007. *Toward a Deeper Understanding of the Role of Interaction in Information Visualization*. Disponível online em: <http://ieeexplore.ieee.org/document/4376144/> [acedido a 28 de novembro de 2017].

7. Anexos

Dada a sua dimensão, os anexos foram colocados num documento à parte. Se procuram acesso ao mesmo, enviar email com o assunto “Anexos da Dissertação” para 1701127@estudante.uab.pt ou me@tiagoperes.eu.