

Faculdade de Ciências e Tecnologia
Universidade de Coimbra

VISUALIZAÇÃO DA CORRELAÇÃO DE PADRÕES AMBIENTAIS NA CHINA

Mestrado em Design e Multimédia

João Miguel Rocha Lameira

Orientação

Penousal Machado

Evgheni Polisciuc

23 de Janeiro de 2017

Resumo

A visualização de informação é uma disciplina e área de investigação relativamente recente, que ganhou maior ênfase nas últimas décadas com a evolução tecnológica. Através desta nova ferramenta de exploração, interpretação e apresentação de dados, tornou-se possível a implementação de novas formas de representação gráfica, que abrange várias áreas do conhecimento desde as ciências até às artes.

Esta dissertação foca-se na criação de métodos de visualização georreferenciados que representam graficamente os dados da poluição, do vento e do uso do solo na China, com o objetivo de construir narrativas visuais que permitam explorar a informação. Assim, pretende-se aplicar técnicas de visualização de informação capazes de sintetizar e interpretar os dados. Estes dados, de carácter geográfico, permitem-nos tanto representar os níveis de poluição, velocidade e direção do vento, em diferentes pontos da China, como analisar o seu dinamismo ao longo do tempo.

Para visualizar estes conjuntos de informação, foram utilizadas duas abordagens distintas. A poluição e o uso do solo são representados sobre o mapa por *isolines*, enquanto o vento é representado por partículas. Desta forma, pretende-se tanto representar os dados como sobrepor os padrões por camadas com o intuito dos correlacionar. Posteriormente, também é proposto a implementação de uma interface gráfica que permita ao utilizador explorar e analisar os dados.

Abstract

The visualization of information is a discipline and field of research relatively recent, that gained more emphasis in the last decades with the technological evolution. Through this new tool of exploration, interpretation and presentation of data, it became possible to implement new forms of graphic representation, which covers various areas of knowledge, from sciences to arts. This dissertation focuses on the creation of georeferenced visualization methods, which graphically represent the pollution, wind and land use data in China, in order to construct visual narratives that allow the exploration of information.

Thus, it is intended to apply information visualization techniques capable of synthesizing and interpreting the data. These geospatial data allow us to represent pollution levels, wind speeds and direction in different parts of China, and to analyze their dynamism over time. To view this data, two different approaches were used. The pollution and land use with representations on the map, through isolines, and the wind with particle representation. This way, it is intended both to represent the data and to superimpose them from several layers in order to identify patterns and correlate them. In addition, the implementation of an interface is proposed for a better exploration and analysis of the data.

Palavras-chave

Visualização de informação, visualização de dados, correlação, poluição, uso do solo, vento, padrões, mapas temáticos.

Agradecimentos

Aos meus pais

Ao meu irmão

À minha namorada

Ao meus orientadores e amigos

Um especial agradecimento

“The purpose of information visualization is to amplify cognitive performance, not just to create interesting pictures. Information visualizations should do for the mind what automobiles do for the feet.” Card, 2008.

Índice

Resumo
Abstract
Agradecimentos

1	<u>Introdução</u>	15
1.1	Motivação	16
1.2	Âmbito	17
1.3	Objetivos	18
1.4	Estrutura da dissertação	19
2	<u>Estado de arte</u>	21
2.1	Qualidade do ar China	22
2.2	Teoria da visualização	26
2.2.1	Semiologia gráfica	27
	Análise da informação	
	Propriedades do sistema gráfico	
2.2.2	Princípios da excelência gráfica	44
2.2.3	Princípio da integridade gráfica	45
2.2.4	Interatividade	79
	Níveis de interatividade	
	Técnicas de interação	
2.3	Mapas temáticos	54
2.3.1	Mapa de Pontos	57
2.3.2	Mapa de Símbolos	60
2.3.3	Mapa Coroplético	64
2.3.4	Mapas isométricos e isolinhas	68
2.3.5	Mapas de fluxo	74
2.3.6	Cartogramas	76

2.4	Casos de estudo	
2.4.1	Wind map	80
2.4.2	The world air quality index	85
3	<u>Trabalhos relacionados</u>	91
3.1	Projeto UKKO	92
3.2	Iso-edges for the Geovisualization of Consumptions	97
4	<u>Desenvolvimento da aplicação</u>	103
4.1	Visualização da pouição	104
4.2	Visualização do vento	127
4.3	Visualização da correlação	137
5	Conclusão	147
6	Bibliografia	159
7	Anexos	163

1 Introdução

A poluição é um tema que tem ganho relevância com o passar dos anos e, a partir da Revolução Industrial, tornou-se sinónimo de grandes problemas para a humanidade, em proporção global. A forte presença das indústrias, o aumento crescente das áreas urbanas e da sua população, elevaram os índices de poluentes do ar numa escala alarmante, o que impulsionou um maior interesse no controlo dos gases concentrados.

Por ser o maior país da Ásia Oriental e o mais populoso do mundo, a China, é também um dos maiores emissores de gases poluentes. Nela, os dados sobre os gases concentrados são recolhidos e armazenados diariamente criando grandes volumes de informação. Esses dados, que abrangem o índice de Ozono, Dióxido de Nitrogénio, Dióxido de Enxofre, Monóxido de Carbono e de Partículas Suspensas, presentes em várias cidades do país, originam-se em diversas situações, naturais ou antropogénicas.

Para representar esta quantidade elevada de dados abstratos, ou seja, que não são físicos, a perceção visual e a cognição humana podem ajudar a compreender as informações. Segundo Jacques Bertin, se os dados forem transmitidos graficamente, ao invés de grandes tabelas, a comparação torna-se fácil e pode ser instantânea. Esta vantagem é conseguida porque o sistema de perceção visual do ser humano deteta e reconhece padrões organizados hierarquicamente.^[1]

A visualização de informação, por ser um conjunto de tecnologias que usam a computação gráfica para ampliar a cognição humana, permite acelerar a compreensão e a ação num volume de informação cada vez maior. Esta maneira de visualizar a informação consegue-se por meio dum método de interpretação de dados representados graficamente, que tem o propósito de analisar e comunicar a informação.^[2]

^[1] Bertin, J. (1967 | 1983) *Semiology of graphics: Diagrams, networks, maps*, Madison, WI, University of Wisconsin Press, pp 2-3.

^[2] Card, S. K., Moran, T. P., & Newell, A. (2008). *The psychology of human-computer interaction*. Boca Raton, Fla.: CRC Press. pp 542.

Motivação

A visualização geográfica sempre desempenhou um papel importante na história da humanidade, muito antes das representações suportadas pelo computador se tornarem populares. Na cartografia, a arte e ciência de *map-making*, desenvolveu-se diversas técnicas de representação de dados geográficos, o que contribuiu para uma grande base de conhecimento cartográfico, sobre o qual a visualização suportada pelo computador pode ser construída.

Contudo, no contexto atual, devido à quantidade e complexidade da informação, e à sua constante alteração no tempo, estas técnicas de visualização não se adequam para representar os dados de uma forma dinâmica. Sendo assim, esta é uma área na qual faz sentido investigar novas metodologias e modelos visuais personalizados que sejam mais eficientes na exploração e apresentação dessas informações.

Visto que nos encontramos num momento em que os índices de poluição em todo o mundo são elevados, e que há, uma grande quantidade de dados armazenados, surge o interesse em criar uma aplicação que permita representar os gases presentes no ar. Desta forma, tornar a informação mais acessível ao utilizador permitindo interagir, através de uma interface gráfica, extrair novas informações e, conseqüentemente, conhecimento a partir dos padrões visualizados.

Neste sentido, o projeto fundamenta-se no desenvolvimento de uma aplicação que permita, não só visualizar os dados da poluição como a sua correlação com o uso do solo e o movimento do vento, possibilitando a comparação de informação de forma interativa.

Âmbito

Realizado no âmbito da dissertação de mestrado do curso de Design e Multimédia da Universidade de Coimbra, o projeto desenvolveu-se no laboratório de *Computational Design and Visualization (CDV LAB)*, pertencente ao *Cognitive and Media Systems Group (CMS)* do *Centre of Informatics and Systems of the University of Coimbra (CISUC)*.

O projeto insere-se na área da representação gráfica, mais especificamente no campo de mapeamento de dados em elementos visuais, particularmente nos mapas temáticos. Neste projeto, é proposto como objeto de estudo a análise e visualização cartográfica de elementos envolvendo a problemática ambiental da poluição do ar, mais especificamente, no território da República Popular da China.

O seu domínio de atuação abrange o campo da visualização de informação, onde foi realizada uma pesquisa do estado da arte, exploradas e analisadas diversas técnicas em função do seu sucesso em comunicar. Posteriormente foram tiradas ilações que ajudaram a definir melhor as condições necessárias na implementação da aplicação. A outra área sobre a qual se incidirá esta dissertação é meio ambiente, em particular a poluição do ar, um fenómeno espacial e temporal. Os dados devido à sua natureza dinâmica e multidimensional foi realizada uma pesquisa e exploração de técnicas de interatividade com a finalidade de criar uma visualização nova que permita a interpretação dos dados através de elementos visuais representativos.

Objetivos

A visualização de informação é uma área ampla, que abrange diversas técnicas de computação gráfica, geralmente interativas, que proporciona a compreensão e análise de um conjunto de dados por meio de representações gráficas. Neste sentido, torna-se necessário definir os objetivos a serem alcançados, a fim de realizar o projeto de uma forma sólida e com êxito.

Os objetivos deste projeto incluem:

- Investigação do estado da arte na área da visualização de informação, mais especificamente no campo da cartografia e mapas temáticos.
- Analisar e compreender visualizações desenvolvidas com técnicas similares para ajudar na escolha e desenvolvimento de uma representação gráfica coerente aos dados em estudo.
- Analisar os dados e desenvolver protótipos adequados à sua estrutura.
- Desenvolver modelos gráficos que representem a poluição do ar, o uso do solo e o vento na China.
- Criar um programa capaz de mapear dados em elementos visuais de uma forma dinâmica.
- Juntar os artefactos visuais desenvolvidos para que seja possível sobrepor os dados.
- Construir uma interface que permita ao utilizador interagir com os dados no tempo e no espaço.

Estrutura da dissertação

O **Capítulo 2**, inserido no estado da arte, salienta a teoria da visualização de informação, a qual abrange a semiologia, os princípios da excelência e integridade gráfica, interatividade e os mapas temáticos. Por fim, o capítulo descreve trabalhos produzidos na área da visualização de dados sobre o mesmo tema em questão, sendo o primeiro sobre o vento e o último sobre o tema da poluição. Ambos os trabalhos abordam práticas que unem dados específicos para cada área de um mapa.

O **Capítulo 3** enfatiza o estudo e a análise sobre outros trabalhos desenvolvidos com técnicas de visualização de dados similares. Os exemplos descritos, neste capítulo, expõem o uso de representações gráficas que utilizam como base dados georreferenciados para sintetizar e simplificar uma vasta e complexa quantidade de informação. Estes exemplos, embora analisem diferentes tipos de informações, assumem grande importância pela forma com que os dados são representados em mapas e as informações são transformadas em visualizações dinâmicas e acessíveis.

O **Capítulo 4** especifica a modo em que se desenvolve a aplicação e são especificadas as primeiras abordagens os dados e as ferramentas utilizadas para o desenvolvimento do projeto. Nele, são descritos os problemas encontrados, ao longo do desenvolvimento, o modo como foram solucionados e como, por consequência, direcionaram a utilização de novos métodos de representações e até chegar a representação final. Ademais, o capítulo descreve o *layout*, as interações e as animações que complementam a estrutura final do projeto.

O **Capítulo 5** descreve uma conclusão sobre o projeto, contributos futuros, e limitações.

2 Estado de arte

O objetivo deste capítulo é aprofundar o campo da visualização de informação, em como representar os dados e comunicar a informação de forma clara e universalmente compreensível. Assim, está dividido em três secções: teoria, mapas temáticos e casos de estudo.

A primeira aborda os princípios teóricos definidos por Jacques Bertin na área da representação gráfica, análise de informação e propriedades do sistema gráfico, bem como, os estudos feitos por Edward Tufte que definem boas e más práticas na construção duma visualização de informação.

Na segunda secção, são analisadas e exemplificadas técnicas de representação de dados geográficos, usadas para representar diferentes tipos de informações. Por último, são apresentadas visualizações desenvolvidas com dados iguais aos nossos para nos ajudar a evitar problemas e entender como são construídas.

2.1 Qualidade do ar na China

Como poluição, compreende-se a presença de determinadas substâncias, variáveis quanto a duração e concentração, que são nocivos ou impróprios ao bem-estar humano e ambiental. A poluição atmosférica, em particular, refere-se à contaminação do ar por meio de gases ou partículas, naturais ou antropogénicas, procedentes de diversos tipos de fontes, fixas ou não.

Entre as fontes naturais de poluição, pode-se citar os gases de emissões vulcânicas, incêndios florestais, descargas elétricas, ações bacterianas, decomposição de plantas e animais, erosão do solo, pólen, entre outros. As fontes antropogénicas de poluentes, entretanto, são as mais preocupantes e crescentes entre todas as fontes de poluentes atmosféricos. Estas, geradas pela atividade humana, são, em sua grande maioria, produzidas pelas indústrias, veículos, usinas termoelétricas, incineradores de lixo, etc.

O problema dessa poluição pode ser retratado através de três componentes essenciais: as fontes que emitem, a atmosfera e os receptores. Quando lançados na atmosfera pelas fontes de emissão tornam-se livres de circulação e interação que, com o auxílio da energia solar e da presença de vapores de água dispersam-se em grandes dimensões. Uma vez lançada em ambientes de locomoção propícios, a poluição presente no ar torna-se pouco seletiva, atingindo a todos, sem distinção, que a ela for exposto.^[3]

Para manter o controlo entre as emissões de poluentes e os riscos à saúde humana e ambiental, foram formuladas considerações, definidas legalmente, na tentativa de administrar os danos emitidos ao planeta. Estas considerações são conhecidas como Padrões de Qualidade do Ar.

^[3] Calijuri, Maria do Carmo; Gasparini, Davi. (2012) Engenharia ambiental. Conceitos, Tecnologia e Gestão. Editora Campus.

Os Padrões de Qualidade do Ar são definidos através de leis, que determinam um limite máximo para a concentração de poluentes atmosféricos num determinado local. Por serem baseados em estudos científicos, proporcionam níveis de segurança adequados, como meio de garantir a proteção de saúde e bem-estar da população.

Os poluentes atmosféricos, são substâncias classificadas como partículas suspensas, *Particulate Matter* (PM). Sob um ponto de vista global, são constituídas por todo tipo de material sólido ou líquido, poeiras ou fumo, que se mantém suspenso na atmosfera, devido ao seu diâmetro reduzido. Em geral, são classificados pelo diâmetro aerodinâmico médio das partículas, com relação entre o seu tamanho e a possível penetração no trato respiratório.

As partículas inaláveis grossas classificadas como, PM 2.5 e PM 10, são aquelas que possuem diâmetro médio, medidos em micrómetro, menores que 2.5 μm e do que 10 μm , respetivamente.

Outros poluentes atmosféricos a serem levados em consideração são: Ozono (O_3), Monóxido de Carbono (CO), Dióxido de Enxofre (SO_2) e Dióxido de Nitrogénio (NO_2). O Monóxido de Carbono, definido por ser um gás além de inodoro, incolor, é emitido por processos de queima de combustíveis. Por esses motivos, são encontrados com grande presença nas áreas urbanas através de fontes móveis de poluição.

A China, assume o posto de um dos maiores emissores de gases poluentes do planeta, com índices de poluição do ar alarmante. Diversas cidades do país, com Beijing, Shanghai, Guangzhou e Xi'an, possuem níveis de partículas tóxicas elevadas, mais do que os máximos determinados pelos Padrões de Qualidade do Ar e permitidos pela Organização Mundial de Saúde.

Por ser o país com um dos maiores consumos de carvão no mundo, com uma produção de energia fornecida 60% pelo mesmo,^[4] a China sofre com uma das maiores taxas de poluição, por partículas suspensas de PM 2.5, do planeta. O nevoeiro de fumo e partículas nocivas a saúde conhecido como *Smog* assola cidades como Beijing, Shanghai e Hong Kong, onde a saúde humana assume problemas, negativamente, avassaladores. Isso significa elevados riscos para a saúde desde problemas respiratórios, à probabilidade de cancro e morte prematura em números, consideravelmente, altos.

^[4] Greenpeace. Reduce Air Pollution. Acedido em 31 de agosto de 2016. Disponível em www.greenpeace.org/eastasia/campaigns/air-pollution/.

Tendo em conta que a poluição do ar, pode ser deslocada de acordo com a velocidade e intensidade do vento, analisaremos, brevemente, a medição do mesmo e equipamentos utilizados para realizar tais atividades.

O vento é um acontecimento meteorológico, formado pelo movimento do ar na atmosfera. Por ser um fenómeno de origem natural, são diversos os fatores que podem influenciar a sua formação, como o movimento de rotação e translação, realizados pelo planeta terra, a pressão atmosférica, a humidade do ar e radiação solar. Esses fatores são também os responsáveis pela velocidade que o vento assume, podendo ser fenómenos leves e suaves como catastróficos e destruidores, como por exemplo o caso dos furações.

De modo geral, a velocidade do vento é medida em metros por segundo (m/s), podendo variar, dependendo da sua velocidade, para outras unidades, como nós (equivalente a 1,852 km/h) ou para quilómetros por hora (km/h). Em suma, a presença do vento é de extrema importância para o ser humano, tanto por possibilitar o seu uso na produção de energia, como por facilitar a dispersão, mesmo que mínima, de poluentes acumulados em áreas populosas, que por possuírem interferência de construções, que formando de barreiras, tende a se intensificar em determinados locais.

Neste sentido, o projeto procura obter dados de natureza geográfica, relacionados com ambos os assuntos referidos, tanto a presença da poluição no território chinês, quanto a relação que esta desenvolve com a presença do vento e sua consequente dispersão de poluentes, e representa-los por meio de uma visualização de informação. Para isto, o trabalho aqui desenvolvido, propõe analisar a correlação existente entre a poluição e vento e a concentração de poluentes em determinadas áreas geográficas.

O projeto desenvolve-se no intuito de analisar grandes quantidades de informações sobre poluentes, e interpreta-los de maneira acessível. Nele, é proposto uma visualização de dados que analisa o dinamismo da poluição, paralelo ao uso do solo, apresentados no mapa da China através de um mapa isométrico (*contour map*) que representa os níveis de poluição e as áreas afetadas, bem como o uso do solo, enquanto os ventos são representados através de um mapa de fluxo (partículas). Ambos são apresentados no mapa de forma interativa e dinâmica, o que possibilita ao utilizador uma compreensão mais simples e rápidas sobre os dados disponíveis.

2.2 Teoria

A compreensão geral da teoria pode orientar o processo de desenvolvimento de uma visualização de informação de forma eficaz e evitando erros. A construção de uma visualização geralmente depende das decisões do designer. No entanto, existem uma série de estudos bem documentados que ajudam na tomada de decisões. Um gráfico é composto por componentes, que por sua vez são constituídos por elementos que podem ser representados de várias formas e variar consoante o seu valor.

Desta forma, a teoria de Jaques Bertin é útil para analisar as informações e escolher os elementos gráficos para a representar. Tufte foca-se mais em aspetos funcionais da representação e no uso eficiente do design para explorar dados complexos. Assegurando que o observador consegue ler a informação, através da uniformidade dos gráficos. Por fim, são abordadas as técnicas de interação, mais comuns, que evoluíram junto com os avanços computacionais e tornam os gráficos ricos ao permitir a apresentação de um número maior de informação, bem como, a exploração dos mesmo.

2.2.1 Semiologia gráfica

^[5] Bertin, J. (1967 | 1983) *Semiology of graphics: Diagrams, networks, maps*, Madison, WI, University of Wisconsin Press.

Para criar uma representação visual que permita comunicar os dados de forma eficiente e para que possa ser entendida a informação, iremos estudar a abordagem prática proposta por Jacques Bertin, um teórico e cartógrafo francês, que ficou conhecido pelo seu livro “*Semiologie Graphique*” editado em 1967.^[5] A primeira e mais abrangente tentativa para fornecer uma base teórica para a visualização de informação. Antes disso, as visualizações eram criadas ao gosto de cada um.

No seu estudo ele organizou os elementos visuais de acordo com as características dos dados, mas não incluiu o movimento, a variação de distância e a iluminação. Por essa razão, apenas considera as representações visuais que são próprias para imprimir num papel branco, visíveis num relance, lidas à distância dum livro, debaixo duma iluminação constante, utilizando meios gráficos prontamente disponíveis.

Para comunicarmos a informação com sucesso, no processo de transformação de dados em representações visuais, precisamos de separar o conteúdo, informação a ser transmitida, da forma e propriedades do sistema gráfico. Bertin não discute o conteúdo da informação, o que interessa numa boa representação gráfica é a qualidade e a eficiência de comunicar. Só assim nos permite avaliar completamente o conteúdo da informação.

Análise da informação

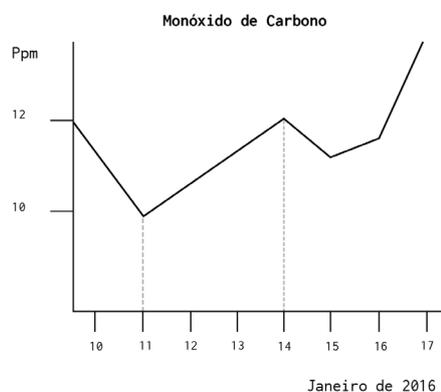
O pensamento apenas pode ser expresso dentro de um sistema de sinais. O mimetismo é uma forma natural de codificação; a linguagem verbal é um código de sinais auditivos (que deve ser aprendido para comunicar com os outros); a linguagem e a representação gráfica também. O armazenamento de memória em discos, fitas ou em computadores necessita de codificações apropriadas.

A representação gráfica é a transcrição, num sistema de sinais gráficos, da informação. Sendo assim, a representação gráfica pode ser abordada pela semiologia, a ciência que lida com todos os sistemas de sinais.

Qualquer transcrição leva necessariamente à separação do *content* (o conteúdo) do *container* (a forma). O ‘*content*’, são elementos do pensamento que podem permanecer constantes, independentemente do sistema de sinais em que são traduzidos, devem ser separados do ‘*container*’, os meios disponíveis em um dado sistema e as leis que governam o seu uso. Esses elementos são constantes, qualquer que seja o pensamento a ser transcrito.

Sendo assim, se estamos a estudar os meios, propriedades e limites do sistema gráfico, é necessário primeiro separar rigorosamente o *content* (a informação a ser transmitida), do *container* (as propriedades do sistema gráfico). O que importa é a eficácia e a qualidade da transcrição gráfica. Aliás é, a característica de uma boa transcrição gráfica. Que nos permite avaliar completamente a qualidade do conteúdo da informação.^[6]

Sabendo que cada sistema de sinais tem as suas propriedades, o seu estilo e a sua estética, que constante pode ser isolada num pensamento, ao longo de várias transcrições? O pensamento é a relação entre vários conceitos que foram reconhecidos e isolados num dado momento de um conjunto de imagináveis conceitos.



^[6] Bertin, J. (1967 | 1983) *Semiology of graphics: Diagrams, networks, maps*, Madison, WI, University of Wisconsin Press, pp. 4-5.

Figura 1 - Gráfico com as médias de Monóxido de Carbono na China.

Por exemplo, no gráfico da figura 1, no dia 11 de janeiro de 2016, o nível médio Monóxido de Carbono é de 10 ppm e no dia 14 de janeiro de 2016 o nível médio de Monóxido de Carbono é 12 ppm. Qualquer que seja a forma assumida pela frase, o conteúdo envolverá sempre alguns elementos pertinentes:

- O conceito ‘quantidade de ppm’, ou a variação do número de partes por milhão;
- O conceito ‘tempo’, ou a variação da data;
- Um elemento - Monóxido de Carbono - do conceito ‘diferentes médias de Monóxido de Carbono na China’. Este elemento é por definição, a invariante, uma vez que constitui a relação entre número de ppm e a data.

Na representação gráfica, o conteúdo traduzível do pensamento será chamado de informação. É constituída essencialmente por uma ou várias correspondências pertinentes entre um conjunto finito de conceitos variáveis e um invariante.

Sempre que falarmos de informação a ser transmitida, a noção central comum a todas as correspondências pertinentes será chamada, por Bertin, de **invariante**. Os conceitos variáveis envolvidos serão chamados de **componentes**. Assim, o exemplo da figura 1, pode-se dizer que tem uma invariante – uma dada média de Monóxido de Carbono - que relaciona dois componentes, a variação do número de ppm (partes por milhão), e a variação no tempo.

Seja qual for o sistema de sinais utilizado, pelo menos dois componentes serão precisos para representar a informação. No sistema gráfico, duas componentes visuais são normalmente utilizadas: as duas dimensões do plano. De forma a facilitar a explicação, Bertin, chamou aos componentes de um sistema gráfico de sinais, **variáveis visuais**, e as duas variáveis fornecidas pelo plano de **dimensões planares**. A informação será, portanto, formada por correspondências pertinentes entre os componentes e a sua representação gráfica por correspondências entre as variáveis.^[7]

^[7] Bertin, J. (1967 | 1983) *Semiology of graphics: Diagrams, networks, maps*, Madison, WI, University of Wisconsin Press, pp.16-27.

A determinação do número dos componentes é o primeiro nível na análise da informação. Componentes e variáveis são por definição, divisíveis. As diferentes partes identificáveis de um componente ou de uma variável são chamadas por Bertin de **elementos** ou **categorias**.^[8] Por exemplo, as categorias ‘bovino’, ‘ovino’ e ‘caprino’ pertencem à componente ‘animais domésticos’. A complexidade de uma figura está ligada ao número de categorias em cada componente.

^[8] Bertin, J. (1967 | 1983) *Semiology of graphics: Diagrams, networks, maps*, Madison, WI, University of Wisconsin Press, pp.28-32.

O termo **comprimento**^[9] é usado para descrever o número de elementos ou categorias presentes numa determinada componente ou variável. Esta é a segunda etapa da análise de informação. Desta forma, pode-se dizer que o componente “sexo” tem um comprimento igual a dois, “feminino” e “masculino”. Num componente quantitativo não devemos confundir o “comprimento”, o número de passos úteis, com o intervalo da série, a razão entre o maior e o menor número na série estatística.

^[9] Bertin, J. (1967 | 1983) *Semiology of graphics: Diagrams, networks, maps*, Madison, WI, University of Wisconsin Press, pp.33.

Os gráficos mostram as relações estabelecidas entre os elementos ou componentes. Estas relações definem três níveis de organização, e cada componente ou variável visual localiza-se num, destes níveis:

O nível **qualitativo** (ou nível nominal) inclui todos os conceitos numa simples diferenciação (comercio, produtos, religião, cores e etc). Normalmente envolve duas abordagens percentuais: isto é similar aquilo, permite combinar todos os componentes num único grupo (associação). Isto é diferente daquilo, pertence a outro grupo (diferenciação).

O nível **ordenado** envolve todos os conceitos que permitem uma classificação dos elementos numa forma universal, por exemplo, uma ordem temporal: 10 anos-25 anos-60 anos; uma ordem sensorial: frio-morno-quente, preto-cinza-branco e uma ordem de discriminação intelectual ou moral: fraco-bom-excelente, pequeno-médio-grande.

O nível é **quantitativo** (intervalo de níveis) quando existe unidades contáveis, por exemplo: isto é um quarto daquilo. Estes níveis são sobrepostos, o que é quantitativo é da mesma forma ordenado e qualitativo. O que é ordenado também qualitativo. O que é qualitativo não é quantitativo nem ordenado, mas é arbitrariamente ordenável.

^[10] Bertin, J. (1967 | 1983) *Semiology of graphics: Diagrams, networks, maps*, Madison, WI, University of Wisconsin Press, pp.34-39.

Os níveis de organização^[10] são a terceira etapa da análise de informação, os quais formam um domínio universal do significado, de analogias fundamentais, que podem ser comprovadas pela representação gráfica. Todas as formas de significado que não estão nas relações acima, são na verdade exteriores ao gráfico e servem apenas para ligar o sistema gráfico aos conceitos do mundo exterior. Elas devem contar com uma explicação codificada em outro sistema (legenda) ou uma analogia figurativa de uma forma ou cor (símbolos), os quais são baseados em hábitos adquiridos ou convenções aprendidas e não podem ser universais. Cada variável visual tem as suas propriedades particulares de comprimento e nível. É importante que cada componente transmitida por uma variável visual tenha pelo menos uma correspondente comprimento e nível. Os gráficos preocupam-se com a representação destes três níveis de organização. No entanto, as relações de similaridade e ordem, com base na métrica, são as que constitui a fundação para o processo de informação e análise.

Propriedades do sistema gráfico

Depois de analisarmos o conteúdo da informação, é necessário representa-la num sistema gráfico. Tal como, a escrita é uma codificação da linguagem, a informação a ser transmitida tem que ser codificada por marcas.

A **localização** tipicamente significa a posição da marca (símbolo) em relação a um sistema de coordenadas espaciais projetadas. A localização é considerada uma variável visual “indispensável”, e

tem primazia visual sobre as outras. A localização é usada principalmente para representar a componente espacial da informação. No entanto, pode ser usada para representar informações do atributo. Além disso, os mapas de *isolines* manipulam a localização, interpolando os valores dos atributos para produzir uma superfície espacial, em seguida, desenha as *isolines* através das localizações dos valores com atributos iguais dentro da superfície.

Cada marca visível tem uma localização, que pode variar através da correspondência entre as duas dimensões do plano. Fixada num dado ponto do plano pode ser desenhada de diferentes formas, e variar as suas propriedades visuais^[11] de:

Tamanho: descreve a quantidade de espaço ocupado pela marca no plano (sistema gráfico). O tamanho é a primeira variável visual a ser manipulada nos mapas de símbolos graduados. O tamanho também é manipulado em mapas de fluxo que escala a espessura das linhas de fluxo para um dado valor.

Forma: descreve o contorno do veículo de sinalização. A forma da variável visual é essencial para o desenho das marcas (símbolos) qualitativos frequentemente usados no mapeamento por referência. As formas podem variar e ser, abstratas, como os círculos, quadrados ou triângulos e icônicas, imitando diretamente o referente representado pelo símbolo.

Orientação: descreve a direção ou rotação da marca no plano. A orientação normalmente é relativa à *neatline* do plano (explicitamente incluída ou inferida pelo espaço negativo), mas em alguns casos pode ser relativa ao sistema de coordenada espaciais. A orientação pode ser manipulada pelos mapas de fluxo para representar a direção do fluxo.

^[11] Bertin, J. (1967 | 1983) *Semiology of graphics: Diagrams, networks, maps*, Madison, WI, University of Wisconsin Press, pp.42-43.

Cor: descreve o comprimento da onda dominante do símbolo (marca) na proporção visível do espectro eletromagnético, por exemplo azul, verde e vermelho. É particularmente relevante para o mapeamento de coropléticos, bem como outras formas de mapas de referência e mapas temáticos que usam cores para designar categorias ou classes. Um esquema de cores espectral manipula a tonalidade da cor enquanto controla os outros componentes de cor.

Valor: descreve a quantidade relativa de energia emitida ou refletida pela marca. A variação no valor da cor resulta na percepção de sombreamento, ou áreas de luz relativa (alta emissão) e escuras (baixa emissão). A manipulação do valor da cor é importante para os mapas coropléticos que representam informações ordinais ou numéricas. Em esquema de cores sequencial ajusta o valor da cor em uma direção, às vezes cruzando duas ou três cores diferentes para melhor a discriminação.

Textura: a textura descreve a grossura do padrão de preenchimento dentro da marca (símbolo). É comum em mapas coropléticos manipular a textura, usando técnicas de *halftone* ou *dithering* a fim de imitar sombreamento. Caivano^[12] descreveu a textura como uma dimensão visual de ordem superior com três componentes constituintes: a direção das unidades de textura (relacionada à orientação da variável visual), o tamanho das unidades de textura (relacionadas ao tamanho da variável visual) e a densidade das unidades de textura (aproximando-se do efeito perceptivo do sombreamento associado ao valor da variável visual cor).

^[12] Caivaio, J. L. (1990). Visual texture as a semiotic system. *Semiotica*, 80(3/4), 239-252) in MacEachren, Alan M. (1995) *How maps work: representation, visualization, and design*. Guilford Press.

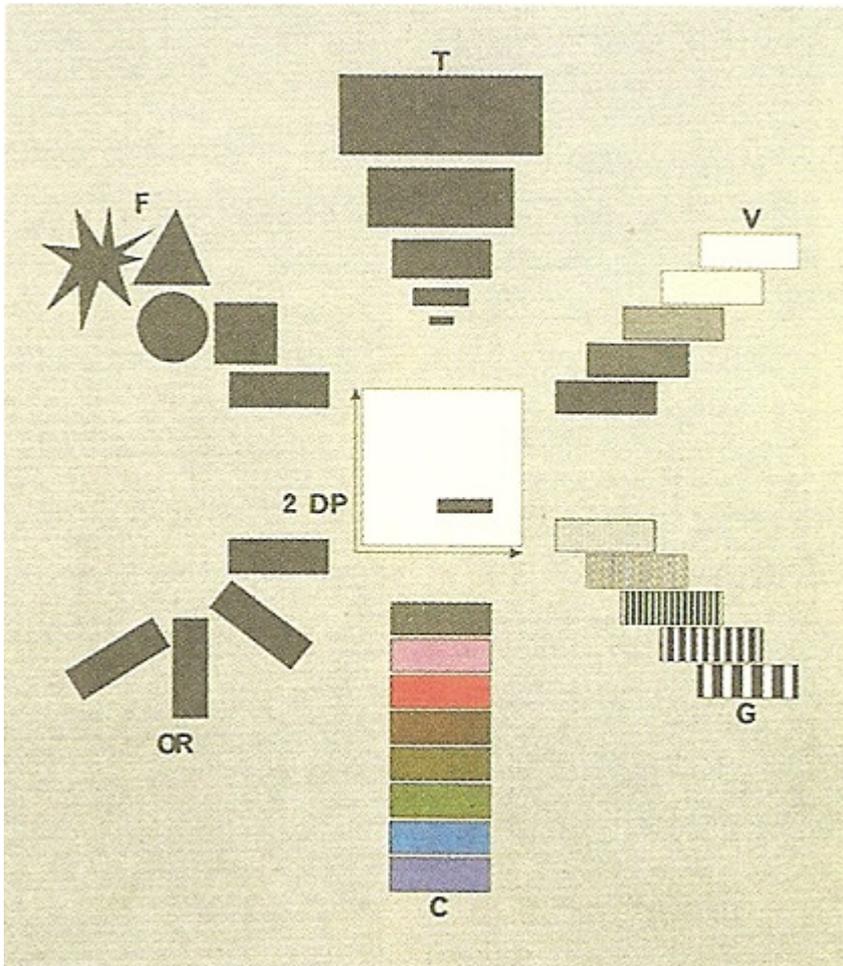


Figura 2 - Variáveis visuais introduzidas por Jacques Bertin.

O conjunto de variáveis visuais apresentadas por Bertin foram estendidas pelo Geógrafo e Professor da Universidade do Ohio, Alan MacEachren. No seu livro publicado em 1995, *How maps work*, ele identificou cinco novas variáveis visuais.^[13]

Saturação da cor: descreve o pico espectral da marca através do espectro visível. É a terceira variável, das três variáveis visuais, associada à percepção de cor. As cores saturadas emitem ou refletem a energia numa banda concentrada do espectro visível.

^[13] MacEachren, Alan M. (1995) *How maps work: representation, visualization, and design*. Guilford Press, pp. 272-276.

Do ponto de vista de design, o valor da cor pode ser concetualizado como a quantidade de preto num símbolo enquanto a saturação é a quantidade de cinza. A saturação de core também pode ser referida com “croma” e “intensidade” na teoria da cor.

Disposição: descreve o *layout* das marcas gráficas que constituem o plano. A disposição das variáveis visuais pode variar: regular, isto é, as marcas gráficas estão perfeitamente alinhadas numa estrutura com a forma de grelha; irregular, as marcas gráficas são colocadas aleatoriamente ou se fundem em agrupamentos.

Nitidez: descreve a nitidez do limite do símbolo. Também referido como *depth-of-field* e *fuzziness* em visualização de informação. MacEachren descobriu que a nitidez era uma variável visual eficaz para representar a incerteza no contexto da simbolização pontual.

Resolução: descreve a precisão espacial em que um símbolo é exibido no plano. A resolução como uma variável visual alavanca diferentes níveis de abstração para codificar a informação. Numa representação *raster*, a resolução refere-se ao tamanho da grelha. Numa representação vetorial, a resolução refere-se à quantidade de detalhes (nós).

Transparência: descreve a quantidade de mistura gráfica entre um símbolo e o fundo ou outros símbolos subjacentes. Originalmente, MacEachren, referiu-se à transparência como “névoa” para sugerir uma barreira parcialmente opaca que impacta a clareza dos símbolos subjacentes (o canal alfa indica a transparência em computação gráfica).

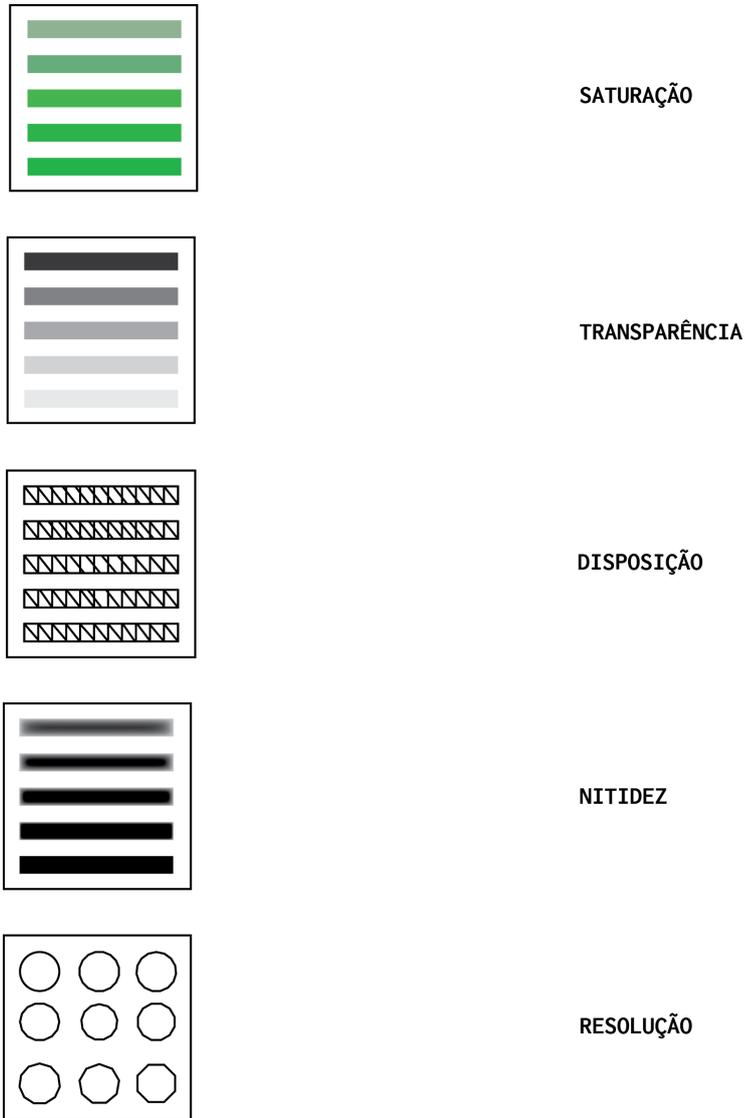


Figura 3 - Variáveis visuais introduzidas por Alan MacEachren.

Sendo assim, o Designer tem ao seu dispor doze variáveis visuais, duas do plano e dez de retina, que são os componentes de um sistema gráfico. Estas variáveis visuais têm diferentes propriedades e diferentes capacidades para descrever um dado tipo de informação, tal como, os componentes cada variável é caracterizada pelo seu nível de organização e pelo seu comprimento.

Classes de representação (implantação)

O plano é o centro de todas as representações gráficas, é homogêneo e tem duas dimensões. No plano uma marca pode ser representada por um ponto, uma linha, ou uma área, que constituem as três figuras elementares do plano geométrico, a que Bertin chamou de **implantação** ou **classe de representação**.^[14] Para distinguir as classes de representação, Bertin definiu que o tamanho da variável de retina e o seu uso varia consoante a classe de representação envolvida; a representação de quantidades varia segundo o ponto, a linha ou a área utilizada; as diferenças na classe de representação são seletivas e numa imagem o mesmo conceito não pode ser representado por diferentes implantações.

^[14] Bertin, J. (1967 | 1983) *Semiology of graphics: Diagrams, networks, maps*, Madison, WI, University of Wisconsin Press, pp. 44.

Um **ponto** representa uma localização no espaço que não tem teoricamente tamanho ou área, apenas coordenadas e consequentemente pode variar de posição. A sua representação gráfica, no entanto, carece de uma marca que tenha tamanho suficiente para torna-lo visível. Numa representação cartográfica este fenómeno tem o nome de ponto de representação.

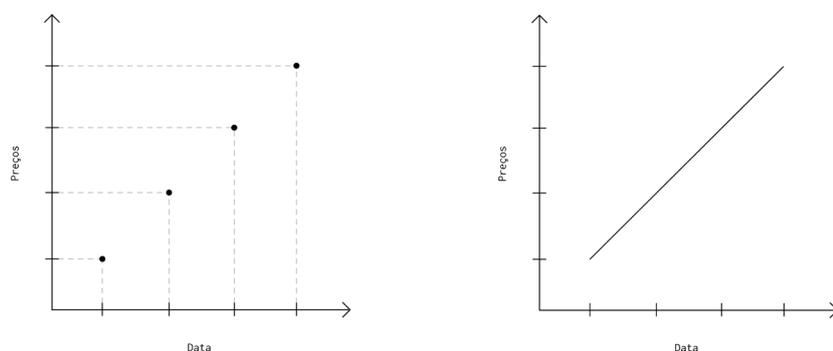
Uma **linha** é basicamente uma conexão entre dois pontos, que tem comprimento e posição, mas não tem teoricamente área. Pode ser descrita por um conjunto infinito de pontos e definir a fronteira entre duas áreas. Assim uma linha pode variar de posição e a marca que a torna visível variar em tamanho, textura, cor, valor, orientação e forma.

Uma **área** é a última classe de representação, que pode ser formada por um ponto e uma linha. É algo que existe no espaço do plano que tem tamanho teoricamente medível, e pode variar de posição, mas a marca que a representa não pode variar de tamanho, forma ou orientação sem causar uma variação do seu significado. Contudo pode variar de valor, textura e cor.

O **plano** é contínuo, limitado apenas pelas fronteiras da percepção e pelos limites da diferenciação gráfica. Desde que a sua continuidade não seja afetada por quebras, a informação não sofre quaisquer lacunas.

Segundo Bertin, a utilização das dimensões planares, ou **imposição**,^[15] depende da natureza das correspondências expressas no plano, ou seja, a forma como o designer as utiliza define quatro grupos de representação: diagramas, redes, mapas ou símbolos.

Um grupo de representação gráfica é considerado um **diagrama**, quando a correspondência no plano pode ser estabelecida entre todos os elementos de um componente e entre todos os elementos de outro componente.

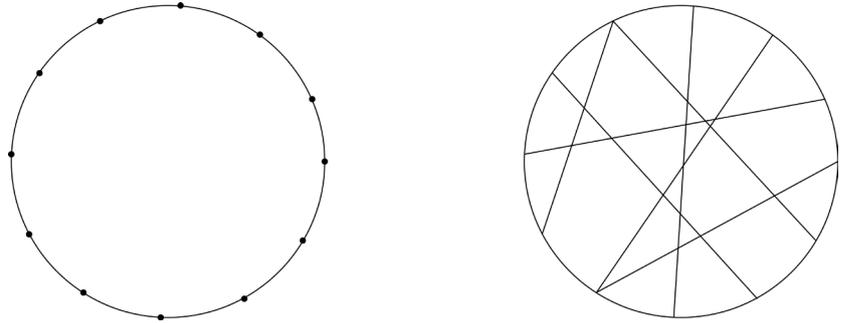


^[15] Bertin, J. (1967 | 1983) *Semiology of graphics: Diagrams, networks, maps*, Madison, WI, University of Wisconsin Press, pp. 50-51.

Figura 4 - Exemplo da construção de um diagrama de linhas.

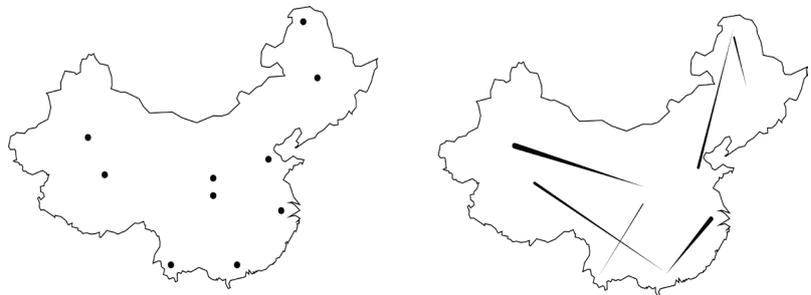
Por exemplo, as tendências do stock X no mercado de Paris ao longo do tempo tem duas componentes: tempo e preço, a correspondência de ambas no plano resultam num gráfico de linhas a que chamamos de diagrama. O processo de construção de um diagrama passa pela definição da representação dos componentes e pelo registo das correspondências. Quando a correspondência no plano pode ser estabelecida entre todos os elementos do mesmo componente, a este segundo grupo de representação gráfica dá-se o nome de redes.

Figura 5 - Exemplo da construção de um gráfico radial.



Por exemplo, numa conversação entre indivíduos sentados à volta da mesma mesa, qualquer indivíduo é capaz de comunicar com outro qualquer. Pode ser representado, como na figura 5, por um gráfico radial onde cada ponto é um indivíduo sentado à mesa e as conexões são as conversas registadas entre eles. No final o registo das correspondências observadas constitui a informação, que pode ser simplificada, ordenando os elementos.

Figura 6 - Exemplo da construção de um mapa.



Uma construção gráfica é considerada um mapa quando as correspondências no plano podem ser estabelecidas entre todos os elementos da mesma componente e organizadas numa ordem geográfica, também conhecido como mapa geográfico. Por exemplo, o inventário geográfico das estradas, é constituído por um conjunto de correspondências estabelecidas entre elementos geográficos, geralmente um conjunto de cidades/províncias, organizadas em ordem geográfica. Desde que a rede geográfica

não possa ser reordenada arbitrariamente, a imagem pode ser simplificada pela eliminação de determinadas correspondências. O processo de construção de um mapa é simplificar tudo, reproduzindo em ordem geográfica e registrando as correspondências. Isto resolve qualquer problema na escolha das duas dimensões planares.

Uma construção gráfica é considerada um símbolo quando a correspondência não é estabelecida no plano, mas entre um único elemento e o leitor, ou seja, é exterior ao gráfico. É um problema que envolve estudos semióticos de simbolismo, geralmente baseados em analogias de forma ou cor. Tal como nos sinais de estrada ou de ferrovia, os códigos convencionais são utilizados na tipografia, agricultura, geologia ou na indústria, por exemplo, sinais de segurança ou símbolos militares. Também conhecidos como ícones.

No geral, qualquer construção gráfica, seja qual for o grupo a que pertença irá ser sempre denominada de representação ou gráfico. Nos diagramas e redes, a imposição é variada. O plano pode ser utilizado de diferentes maneiras. Os componentes podem ser inscritos de acordo com uma disposição, dispersos por todo o plano ou de acordo com uma construção **retilínea**, **circular**, **ortogonal** ou **polar**.

Figura 7 - Tipos e grupos de imposição por Jacques Bertin

IMPOSITION		TYPES OF IMPOSITION				
		ARRANGEMENT	RECTILINEAR	CIRCULAR	ORTHOGONAL	POLAR
GROUPS OF IMPOSITION	DIAGRAMS		 	 	 	 
	NETWORKS	 	 	 	 	 
	MAPS	 GEO 				
	SYMBOLS					

Variáveis visuais e níveis de representação

Dentre de todas as variáveis visuais, o plano é o único que fornece variáveis que processam todas as propriedades perpetuais. As duas dimensões do plano têm o maior nível de organização e pode assim representar qualquer componente da informação.

O uso das dez variáveis de retina, quer seja para representar uma terceira componente ou substituir as duas dimensões do plano, tem o nome de **elevações**, que podem ser combinadas com todos os tipos de imposição a fim de formar um tipo de construção.

Assim qualquer variável de retina pode ser usada para representar qualquer componente, mas nem todas as variáveis são adequadas para todos os componentes, é a definição dos níveis de organização que ajuda a resolver este problema.

Enquanto o plano é seletivo, associativo, ordenado e quantitativo, as variáveis de retina apenas tem algumas destas propriedades, os seus níveis de organização diferem. Assumindo particular importância na cartografia, devido as duas dimensões do plano serem utilizadas pela base geográfica, sempre que apareça um segundo componente na informação, tem que se utilizar variáveis de retina. As propriedades percentuais de uma variável determinam o seu **nível de organização**.^[16]

^[16] Bertin, J. (1967 | 1983) *Semiology of graphics: Diagrams, networks, maps*, Madison, WI, University of Wisconsin Press, pp. 64-69.

A percepção é **seletiva**, quando permite ao olho isolar todos os elementos de uma dada categoria, e compreender a imagem formada pela categoria. Tal percepção pode ser imediata, neste caso, a variável é seletiva e cada categoria forma uma família. Por outro lado, a percepção pode necessitar de analisar sinal por sinal, nestes casos a variável não é seletiva. Para todas as implementações: ponto, linha e área, a forma e a orientação não são seletivos quando representadas pela área.

A percepção **associativa** é útil quando nos permite agrupar imediatamente todas as categorias de maneira diferente, e quando procuramos uniformizar a variação. Os testes feitos por Bertin mostram que o tamanho e o valor são dissociativos, enquanto as outras variáveis são associativas.

A percepção **ordenada** é útil quando comparamos várias ordens, tal comparação é imediata no caso de a variável ser ordenada. Quando uma variável é ordenada, não é necessário consultar a legenda para ser capaz de ordenar as categorias, torna-se óbvio que isto vem antes daquilo e depois do outro. Assim a forma, a orientação e as cores não são ordenadas, cada pessoa pode estabelecer qualquer ordem, porque nenhuma se afirma imediatamente. Reciprocamente a textura, o tamanho e o valor impõem uma ordem universal e imediatamente perceptível.

A percepção **quantitativa** é útil quando tentamos definir numericamente a relação entre dois sinais e quando tentamos agrupar homogeneamente os sinais. Quando a percepção é quantitativa, a relação numérica entre dois sinais é imediata e não precisa de recorrer a legenda, o leitor vê logo que isto é o dobro daquilo ou metade daquele. Apenas a variação de tamanho pode ser utilizada para representar valores quantitativos.

Figura 8 - Níveis de organização das variáveis de retina, Jacques Bertin. e MacEachren.

	QUANTITATIVO	ORDENADO	SELETIVO	ASSOCIATIVO	DISSOCIATIVO
LOCALIZAÇÃO	■	■	■	■	□
TAMANHO	■	■	■	■	■
NITIDEZ	□	■	□	■	■
RESOLUÇÃO	□	■	□	■	□
TRANSPARÊNCIA	□	■	■	■	■
VALOR	■	■	□	■	■
SATURAÇÃO	■	■	□	■	■
COR	■	■	■	■	■
TEXTURA	■	■	■	■	■
ORIENTAÇÃO	■	■	■	■	□
DISPOSIÇÃO	□	□	■	■	□
FORMA	□	□	■	□	□

2.2.2 Excelência gráfica

Tal como, Jacques Bertin, Edward Tufte^[17] fez um estudo sobre as representações visuais de dados, mas mais focado nos aspectos funcionais do gráfico, descrevendo boas e más práticas na construção de uma visualização. Assim, a excelência em gráficos estatísticos consiste em comunicar ideias complexas com **clareza**, **precisão** e de forma **eficaz**.

A disposição do gráfico necessita de mostrar os dados; induzir o leitor ao pensamento na substância ao invés da metodologia, no design gráfico e na tecnologia da produção do gráfico; evitar distrações no que o dado tem a dizer; mostrar vários números em um espaço pequeno; organizar grandes quantidades de dados em coerência; encorajar o olho a comparar as diferentes partes da informação; revelar os diferentes níveis de detalhe a partir de uma visão ampla para uma mais fina; servir ao propósito de descrever, explorar, tabular e decorar, estar integrado com a estatística e a descrição verbal das informações

“Graphical excellence is that which gives to the viewer the greatest number of ideas in the shortest time with the least ink in the smallest space.”^[18]

A excelência gráfica é uma apresentação bem desenhada de informações interessantes, consiste na substância, estatística e design. É quase sempre múltipla narrativa e requer verdades na representação da informação. Os gráficos revelam informações. O princípio sustenta-se em vários aspectos. Elementos adicionais devem ser incluídos, como as grelhas, números, para poder criar um contexto ao redor da informação.

^[17]Tufte, E. R. (2006) Beautiful evidence, 1st ed. Cheshire, Connecticut: Graphics Press.

^[18] Tufte, E. R. (1983). The visual display of quantitative information. Cheshire, Conn. (Box 430, Cheshire 06410): Graphics Press.

O objetivo é minimizar a quantidade de elementos visuais que possam distrair o leitor da observação, estudo e compreensão da informação importante apresentada no gráfico. Apagar a informação desnecessária, minimizar a redundância do gráfico e maximizar o uso dos elementos cruciais, são maneiras de tornar a visualização mais limpa. Isso focará a atenção na essência do gráfico e a sua mensagem principal.

2.2.3 Integridade gráfica

Segundo Edward Tufte há gráficos que distorcem a informação e que dificultam a percepção da verdade. Os gráficos não são diferentes das palavras, também podem ser manipulados e enganar.

No século XX os gráficos eram associados como uma forma de distorção da verdade, utilizados para enganar o observador ingênuo. No centro da preocupação com a representação gráfica estava a suposição de que os dados gráficos eram principalmente dispositivos que mostravam o óbvio aos mais desatentos. Esta presunção levou a dois caminhos graficamente estéreis nos anos de 1930 a 1970.

O primeiro caminho, o gráfico deve estar “vivo”, “comunicar dinamicamente”, ter muita informação extra e ser exagerado. O segundo caminho, a tarefa principal da análise gráfica era detectar e denunciar erros.

Mais tarde, Jonh Tukey fez os gráficos estatísticos serem respeitados, pondo um fim à visão que os gráficos eram apenas para a decoração de números. Desta forma, Tukey mudou o pensamento sobre o design, usando-o eficientemente para explorar dados complexos.

Os gráficos eram utilizados como instrumentos de raciocínio sobre informação quantitativa. Contudo, os gráficos falsos continuam a existir. Os erros serão sempre confrontados sempre que uma mentira seja detetada. A excelência gráfica começa por dizer a verdade sobre os dados. Mas há vários gráficos que falham em dizer a verdade. Um gráfico não é distorcido se as representações visuais dos dados forem consistentes com a representação numérica.

Diferentes pessoas veem as mesmas áreas de maneiras diferentes, a percepção muda com a experiência e a experiência é dependente de contexto. De qualquer modo, dadas as dificuldades de percepção, o melhor que podemos esperar é alguma uniformidade nos gráficos e assegurar que os observadores têm uma boa chance de obter os números certos.

Tufte definiu **seis princípios**^[19] que conduzem em direção a esses objetivos e conseqüentemente aumentam a integridade gráfica:

^[19] Tufte, E. R. (2006). Beautiful evidence, 1st ed. Cheshire, Connecticut: Graphics Press.

Primeiro, a representação dos números, como dimensão física da superfície do gráfico deve ser diretamente proporcional à quantidade numérica representada.

Segundo, deve ser usada uma legenda clara, detalhada e profunda para debater a distorção gráfica e a ambigüidade, escrever explicações sobre os dados no próprio gráfico e legendar eventos importantes. Um exemplo clássico a evitar é adicionar a terceira dimensão onde não é necessário. A violação do primeiro princípio constitui um gráfico deturpado, medido pelo fator de mentira segundo Tufte.

O fator de mentira é uma conta, que divide o tamanho do efeito no gráfico pelo tamanho do efeito nos dados, se for igual a um, o gráfico pode fazer um trabalho preciso na representação dos números subentendidos, correspondendo à verdade. Se o fator de mentira do gráfico for maior ou menor que um, então existe distorção na informação, nesse caso “mentira”.

Cada parte do gráfico gera expectativas visuais sobre a outra parte e na economia da percepção gráfica, estas expectativas geralmente determinam o que os olhos veem. A confusão entre variação de design e variação de dados numa superfície de um gráfico leva a ambiguidade e erros, para os olhos mudanças no design representam mudanças nos dados.

Tufte define assim o **terceiro** princípio da integridade gráfica: mostrar a variação dos dados, não variação do design.

O **quarto** princípio, numa visualização que envolva dinheiro em séries temporais, as unidades deflacionadas e padronizadas de medição monetária são quase sempre melhores do que unidades nominais. Existem consideráveis ambiguidades de como as pessoas percebem as duas dimensões da superfície e depois a convertem numa dimensão numérica. O uso de duas ou três dimensões variáveis para mostrar dados numa dimensão é uma técnica fraca e ineficiente, apenas capaz de lidar com um pequeno conjunto de dados, geralmente com erros de design e de ambiguidade na percepção.

Estes desenhos causam tantos problemas, segundo Tufte enumerando o **quinto** princípio da integridade gráfica, devem evitar que a dimensão da variável representada não exceda o número da dimensão nos dados. No geral isso implica não utilizar a terceira dimensão para representar algo que apenas precisa de duas dimensões.

Tufte define o último princípio da integridade gráfica, dizendo que os gráficos não devem cotar dados fora do contexto, para serem verdadeiros e reveladores.

2.2.4 Interatividade gráfica

Uma das grandes vantagens da evolução tecnológica, para além, do processamento de grandes quantidades de dados é a interatividade. A interatividade consiste num conjunto funcionalidades oferecidas ao utilizador pela aplicação, que através de uma interface gráfica permite controlar os dados e mudar a visualização. A aplicação recebe uma entrada do utilizador e responde um resultado de saída, fundamentalmente muda a informações exibidas pela visualização permitindo ao utilizador perceber o que acontece quando muda as variáveis.

Níveis de interação

Existem três níveis de interação, o **primeiro nível**, também conhecido como **momento psicológico**, tem uma resposta imediata de cerca de 0.1' (um décimo de segundo). Estímulos com um intervalo de um décimo de segundo fundem-se numa única percepção: duas imagens similares vistas com um intervalo de um décimo de segundo fundem-se numa percepção de movimento.

O mesmo acontece numa representação gráfica se um botão trocar de cor ou posição num intervalo de tempo de um décimo de segundo, ao ser tocado no ecrã, o toque irá parecer ter causado a mudança. Se os dados mudar com um intervalo de um décimo de segundo ao mover um *slider*, irá parecer que o *slider* muda diretamente os dados. Uma ação e a sua percepção ocorrem dentro desse intervalo de tempo apresentando uma relação de causa-efeito.

O **segundo nível** de interação tem um tempo de resposta de um segundo, é o tempo de uma **conversa improvisada**. Eventos que acontecem em menos de um segundo, acontecem muito rápido para o utilizador responder a menos que esteja preparado. Um automóvel requer menos de um segundo para travar num evento inesperado. Um diálogo humano tem um tempo de conversa constante. Se um orador parar de falar por mais de um segundo, o ouvinte sente-se obrigado a perguntar se esta alguém, para ter a certeza que a conversa não acabou. Este nível de interação pode ser explorado por interações suaves com o utilizador: animações que ocorrem em um segundo podem transmitir a informação sem abrandar o utilizador.

O **terceiro nível** de interação tem um tempo de resposta cíclico entre cinco e trinta segundos, conhecido também como **unidade de tarefa**. Este é o ritmo normal de ciclos de interações elementares em um sistema interativo, é o ritmo de uma capacidade cognitiva em que o utilizador pode executar um procedimento. Por exemplo, o tempo de uma interação de rotina com um editor de texto, isto deve incluir o tempo para seleccionar o texto no ecrã e modifica-lo.

O tempo é importante na interação, mas o efeito funciona de forma diferente consoante a escala de tempo da interação. O tempo de resposta de um computador precisa de ser ajustado a essa constante de tempo humano. A interação pode mesmo ser retardada em alguns casos: simplesmente mostrar o estado inicial e o final sem uma transição animada, pode fazer com que o utilizador interprete mal quais são os objetos transformados. O tempo significado. A interação usa mecanismos de programação para impor a animação em outras constantes de tempo.^[20]

^[20] Card, S., Mackinlay, J. and Shneiderman, B. (1999) Readings in information visualization – using vision to think. San Francisco: Academic. pp 231.

Técnicas de interação

Existem muitas técnicas de interação que podem ser implementadas na criação de uma visualização. São essencialmente uma forma de selecionar um subconjunto de objetos na Tabela de Dados. Isto permite que elas sejam utilizadas para localizar dados, revelar padrões, ou para selecionar os argumentos de outras transformações. Sendo assim, o utilizador tem várias maneiras de interagir com a informação, tais como, *data transformation*, *visual mapping* e *view transformations*.^[21]

^[21] Card, S., Mackinlay, J. and Shneiderman, B. (1999) Readings in information visualization – using vision to think. San Francisco: Academic. pp 233-234.

Data transformations é um grupo composto por seis tipos de interações, que modificam a visualização, através da seleção de casos e variáveis que são representados.

dynamic queries: é utilizada para selecionar e filtrar dados num intervalo de variáveis da Tabela de dados. Permitindo ao utilizador obter informações mais específicas, através do controle de uma interface gráfica como um botão ou um *slider*.

direct walk: resulta de um conjunto de ligações de caso para caso.

^[22] Um *Web Browser*, onde o utilizador navega de página em página é um bom exemplo de *direct walk*. Através de um conjunto de *clicks* na visualização o utilizador pode pesquisar a informação e modifica-la.

^[22] Card, S.K., Pirolli, P. and Mackinlay, J.D. (1994) The Cost of Knowledge Characteristic Function: Display Evaluation for Direct-Walk Dynamic Information Visualizations in Card, S., Mackinlay, J. and Shneiderman, B. (1999) Readings in information visualization – using vision to think. San Francisco: Academic. pp 233-234.

details-on-demand: expande um conjunto pequeno de objetos para revelar mais variáveis escondidas. Isso permite a economia de espaço no layout e o mapeamento mais elementos;

attribute walk: o utilizador seleciona um caso e depois pesquisa por outros casos com atributos similares, é também utilizada para comparar dados escolhidos. Esta tecnologia foi desenvolvida num sistema chamado *Rabbit* para pesquisar base de dados.

brushing: é a seleção dos mesmos dados em múltiplas visualizações para comparar os resultados.^[23] O utilizador destaca um caso de uma visualização e outras visualizações mostram as informações sobre o mesmo caso. Um *slider* pode ser visto como uma forma elementar de *brushing*. O *slider* representa todos os objetos ordenados por uma variável.

direct manipulation: pode ser utilizada para modificar transformações, permitindo a parametrização de vários elementos.

O *visual mapping* of elements não tem um número finito de aplicações. Podem ser implementados vários tipos de transformações de várias formas possíveis e é constituído por duas técnicas:

dataflow diagram: é uma técnica comum usada por sistemas de visualização comerciais para mapear dados em formas visuais. A ideia é usar uma representação explícita como um diagrama *node-link* para representar o mapeamento;

pivot table: permite ao utilizador rapidamente manipular os dados apresentados e os mapear facilmente para linhas e colunas de um gráfico.

Finalmente, *view transformations*, um grupo de cinco interações: *direct selection*: refere-se a um conjunto de métodos que evoluíram para selecionar e destacar um objeto ou grupos de objetos, melhorando a aparência da estrutura visual, muitas das vezes identificam um conjunto de objetos que serão argumentos de uma ação;

^[23] Cleveland, W. S. and McGill, M. E. (1988) Dynamic Graphics for statistic, Wadsworth and Brooks/Cole, Pacific Grove, California in Card, S., Mackinlay, J. and Shneiderman, B. (1999) Readings in information visualization – using vision to think. San Francisco: Academic. pp 233-234.

camera movement: é uma mudança da posição do observador especialmente utilizada em três dimensões, a visualização de informação pode por vezes beneficiar permitindo ao utilizador observa-la de outro ângulo e revelando partes desconhecidas, sem obstruções;

magic lenses: seleciona objetos consoante a sua posição x, y e depois aplica outra técnica de seleção, por exemplo, *dynamic queries*.^[24] Além disso pode-se aplicar transformações de visualização e de dados aos itens selecionados;

overview + detail: utiliza dois ou mais níveis de visualização ligados.^[25] Uma das visualizações dá uma visão geral de todos os objetos ou daqueles com maior importância. A outra visualização é detalhada, sobrepõe a visualização geral e mostra mais informação dos objetos selecionados.

zooming: envolve a redução do número de objetos que são visíveis, mas possivelmente aumenta o número de variáveis por objeto que são exibidos.

^[24] Bier, E. A., Stone, M. C., Ken Pier, Buxton, W. and DeRose, T. D. (1993) Toolglass and Magic Lenses: The See - Through Interface, Proc. Siggraph 93', ACM Press, New York, pp 73-80 in Card, S., Mackinlay, J. and Shneiderman, B. (1999) Readings in information visualization - using vision to think. San Francisco: Academic. pp 233-234.

^[25] Plaisant C., Carr D., Shneiderman (1995) Image Browser Taxonomy and Guidelines for Designers, IEEE software, vol. 28, #3, pp 21-32 in Card, S., Mackinlay, J. and Shneiderman, B. (1999) Readings in information visualization - using vision to think. San Francisco: Academic. pp 233-234.

Figura 9 - Tabela com as técnicas de interação, Card et al 1999

Interaction techniques		
Modifies Data Transformations	Modifies Visual Mappings	Modifies View Transformations
Dynamic queries	Dataflow	Direct selection
Direct walk	Pivot tables	Camera movement
Details-on-demand		Magic lens
Attribute walk		Overview + detail
Brushing		Zoom
Direct manipulation		

2.3 Mapas temáticos

As visualizações geográficas sempre desempenharam um papel importante na história humana. Muito antes das visualizações computacionais se tornarem populares. As primeiras visualizações geográficas datam de antes da idade da pedra, como por exemplo, as pinturas rupestres.

Desde a cartografia, a arte e ciência da elaboração de mapas, evolui continuamente até aos dias de hoje. Um dos exemplos conhecidos da visualização estática além dos mapas geográficos são os mapas temáticos, que mostram o padrão espacial de um tema, como por exemplo características climatéricas ou densidade populacional.

“A thematic map is concerned with portraying the overall form of a given geographical distribution. It is the structural relationship of each part to the whole that is important. Such a map is a kind of graphic essay dealing with the spatial variations and interrelationships of some geographical distribution.” [26]

[26] Arthur H. Robinson in Meirelles, I. (2013) Design for Information: An Introduction to the Histories, Theories, and Best Practices Behind Effective Information Visualizations. Rockport Publishers, Beverly, MA.

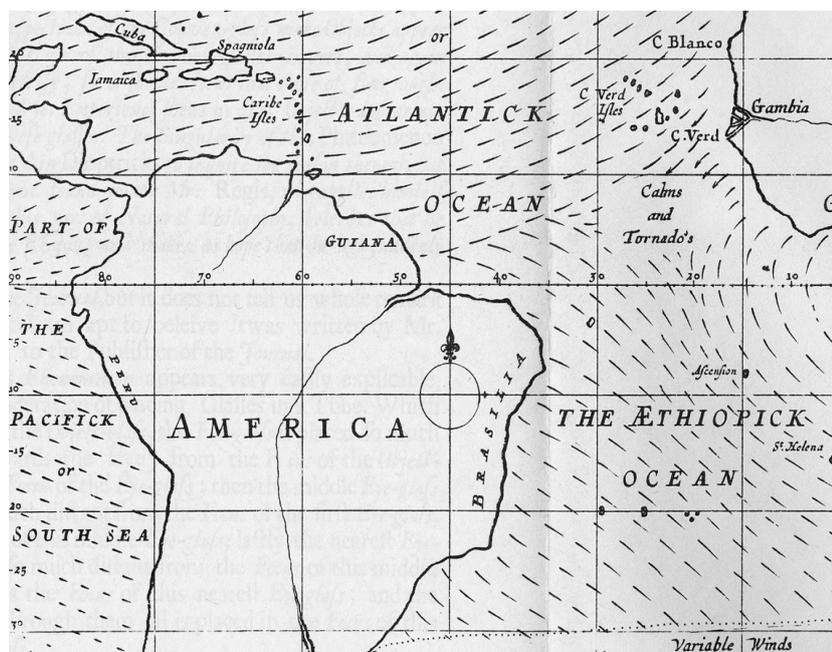


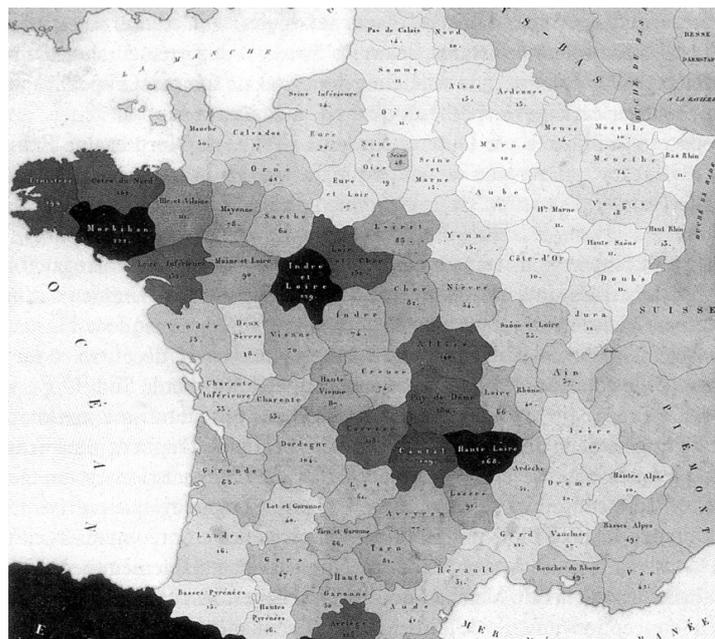
Figura 10 - Primeiro mapa meteorológico, criado por Edmond Halley em 1686.

Os mapas temáticos são representações de dados quantitativos e qualitativos, que têm como base um mapa. Este é determinado pela posição fixa dos dados que são definidos pela geometria, de tal forma que a relações espaciais são representadas utilizando um sistema de referência de localização, por exemplo latitude e a longitude. Por outras palavras, os mapas temáticos mostram um tema que pode ser um número de fenómenos, tais como, social, político, económico ou cultural, com o propósito de revelar padrões e frequências na geografia que ocorrem.

Os mapas temáticos remontam à segunda metade do século XVII e tiveram grandes avanços no século XIX quando foram criados a maioria dos métodos gráficos.^[27] Inicialmente os mapas temáticos representavam fenómenos naturais, um dos primeiros, foi criado pelo geógrafo britânico Edmond Halley, em 1686, que interpretava os ventos e moções observáveis nos mares entre os trópicos representados num mapa geográfico. Como podemos observar na imagem da figura 10, o vento é representado pela direção e pelo comprimento das linhas.

^[27] Meirelles, I. (2013) Design for Information: An Introduction to the Histories, Theories, and Best Practices Behind Effective Information Visualizations. Rockport Publishers, Beverly, MA, pp 116-117.

Figura 11 - “Carte figurative de l’instruction populaire de la France”, Pierre Charles Dupin, 1829. Isabel Meirelles, 2013.



Mais tarde, cerca de um século depois, começou-se a interpretar os fenômenos sociais. O primeiro mapa estatístico moderno aparece pelas mãos do francês Pierre Charles Dupin em 1826, ver imagem da figura 11. O mapa, também conhecido como o primeiro mapa coroplético, descreve em tons de cinza a intensidade e a distribuição do analfabetismo em França.

Durante o iluminismo, à medida que os dados se acumularam, a partir de observações e medições ambientais, a atenção com as informações, mudou do lugar para o espaço. Sendo assim, o foco que até então era analítico e se preocupava com o posicionamento dos atributos passou a ser holístico a se preocupar com a extensão e variação espacial. Assim nasce a ideia de distribuição. Este salto, do lugar para o espaço, levou a criação de representações de distribuições que passaram a ser denominadas como mapas temáticos.

Durante a época dos Egípcios, Gregos e Romanos a contagem da população era registada, com finalidades administrativas, como por exemplo, a tributação. Segundo Meireles, foi a partir dos Romanos que nasceram os censos, uma palavra derivada do latim *census*, que significa “estimativa”. Entretanto, o armazenamento sistemático de dados sociais, apenas começou no final do século XVIII com o primeiro censo populacional registado na Suécia, seguido por outros países ao longo do tempo.

Apesar dos avanços gráficos alcançados por William Playfair no século XVIII, foi durante meados do século XIX que os métodos gráficos, para análise de dados, se desenvolveram sem precedentes, em diversos aspetos. Desta forma, o reconhecimento da importância numérica de informação, para o planeamento geral da população, se desenvolveu em diversos países.

2.3.1 Mapas de Pontos

Os mapas de pontos têm como principal objetivo mostrar a distribuição espacial de fenômenos utilizando o ponto com variável visual. Os mapas podem representar dois conjuntos de dados discretos: fenômenos discretos com informação da localização geográfica, como o mapeamento médico ou fenômenos com variação regular, como a maioria dos mapas que representam os dados de censos, no qual os símbolos são distribuídos dentro de uma área geográfica correspondente a fim de retratar densidades.

São uma maneira simples, mas eficaz, de mostrar diferenças de densidade em distribuições geográficas. Os mapas de pontos foram populares durante muitos anos, por serem de fácil e rápida compreensão. De uma forma intuitiva, revelam agrupamentos, por exemplo.

Existem dois tipos básicos de mapas de pontos, um-para-um, onde um ponto corresponde a um objeto no mapa, ou um-para-muitos, em que um ponto pode representar vários objetos, por exemplo, dez mil pessoas. Há, no entanto, uma tendência de subestimar o número de pontos e as diferenças de densidades entre as áreas. Como tal, é importante que o valor unitário seja um número redondo e claramente indicado. Isso ajuda a fornecer as legendas que ilustram as diferentes densidades no mapa.

A atribuição de unidades de valor e unidades de tamanho afeta em grande parte o mapa e a sua percepção. Por exemplo, se o tamanho do ponto é muito pequeno, e cada unidade equivale a grandes números, o mapa pode ser percebido de forma errada. Da mesma forma, se o tamanho do ponto é muito grande e cada unidade equivale a números pequenos, o mapa pode dar uma impressão imperfeita de alta densidade.

Este problema torna-se maior quando existe um conjunto de dados com grandes variações de densidade, para o qual alguns cartógrafos usaram uma combinação de métodos de pontos graduados e distribuídos. O posicionamento do ponto não é trivial e pode também afetar a percepção da informação. Para posicionar manualmente os pontos é recomendado agrupar as características de acordo com o centro de gravidade dentro da unidade estatística, bem como relacionar com outras informações geográficas, como a topografia e a localização das cidades.



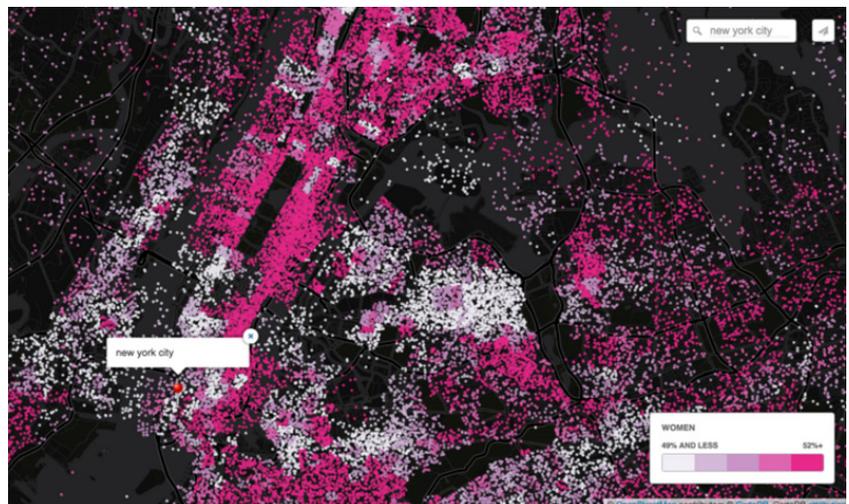
Figura 12 - Mapeamento do surto de cólera em Londres. Criado por John Snow, 1854.

Um dos primeiros mapas de pontos utilizados com a finalidade de traçar padrões de doenças foi criado pelo anestesista britânico Dr. John Snow. Este mapa descreve a epidemia de cólera de 1854 no sul de Londres, considerada uma das piores de sempre, matando

cerca de 500 pessoas em dez dias num raio de 250 metros. Snow em um esforço para defender a sua teoria de a cólera era uma doença transmitida pela água, e não uma doença transmitida pelo ar, como se acreditava. Mapeou as mortes individuais recorrendo ao relatório semanal de mortalidade do Escritório de Registo Geral de Londres e representou-as no mapa por pontos.

As bombas de água, onze no total, são representadas por cruces no mapa, aos dados originais acrescentou conhecimento local, como as informações fornecidas pelo Reverendo Whitehead, que também tinha mapeado o surto. O mapa de Dr. Snow não pôs fim à epidemia de cólera, nem tão pouco convenceu as autoridades sanitárias da teoria da aquosa. A discussão em torna da natureza da cólera só acabou em 1883, quando a bacteriologista Robert Koch identificou o *Vibrio Cholera* como um agente aquoso. O mapa, no entanto, ajudou a avançar com a investigação, revelando o padrão da doença no contexto espacial.

Figura 13 - “A Dot Map of All the U.S. Ladies”, criado por CartoDB’s Andrew Hill. Screenshot da cidade de Nova York.



[28] A Dot Map of All the U.S. Ladies [online]. Acedido em 3 de junho de 2016. Disponível em https://team.carto.com/u/andrew/viz/d28c3cb2-9954-11e5-b837-0e674067d321/embed_map.

O mapa de pontos, *A Dot Map of all the U.S. ladies* [28] criado por Andrew Hill em 2016, permite visualizar onde vivem as mulheres americanas e comparar os números com o resto da população.

Para criar esta representação, foram utilizados dados do censo de 2014, no qual cada ponto no mapa corresponde a uma quantidade de 1000 mulheres. A cor atribuída a cada ponto é calculada entre a proporção de mulheres e a população total de cada localização. Assim, os pontos mais claros (brancos) indicam os locais em que as mulheres representam menos de 49 por cento da população total, e os pontos mais cor de rosa representam a taxa de mulheres superior a 52 por cento.

Existem vários espaços a preto no mapa sem qualquer ponto, muitos deles em cidades densamente povoadas, o que não quer dizer que não existam mulheres nessas áreas. O seu número é reduzido faz com que não sejam representadas pelo algoritmo utilizado.

Se olharmos para o mapa dos Estados Unidos como um todo, duas tendências se destacam. Uma delas é que as cidades têm grupos mais densos de pessoas no geral. Mas esse padrão também pode ser explicado, em parte, pela relação que as mulheres têm, especialmente as mais novas, com as cidades.

As cidades oferecem empregos, redes sociais e amenidades, o que torna mais fácil para as mulheres terceirizar tarefas que até então tem sido tradicionalmente solicitada a fazer. Em segundo lugar, o mapa de pontos mostra que as cidades da Costa Leste, como Nova York, geralmente têm conjuntos mais densos de pontos rosa do que os da Costa Oeste, como Seattle. Há provavelmente várias razões para este desequilíbrio geográfico, que corresponde a uma tendência já identificada. ^[29]

^[29] Citylab. A Dot Map of All the U.S. Ladies. Acedido em 3 de junho de 2016. Disponível em <http://www.citylab.com/work/2016/03/dot-map-american-ladies-census-cartodb/473376/>.

2.3.2 Mapas de símbolos graduados

Os mapas de símbolos graduados são mapas temáticos, que utilizam a variável visual tamanho, para representar a magnitude dos dados discretos, por exemplo, contagem de pessoas. O tamanho é proporcional às quantidades representadas e não depende da área

em que se encontram inscrito. Esta característica ajuda a evitar problemas de confusão entre áreas geográficas e os valores dos dados, como acontece nos mapas coropléticos. Na construção de mapas de símbolos graduados o principal objetivo é que o leitor consiga estimar o valor de um símbolo. Existem duas variáveis que devem ser consideradas ao representá-lo: a forma do símbolo e a escala. A forma do símbolo pode variar, sendo a mais comum o círculo, apesar de vermos também barras retangulares e triângulos, esta facilita a tarefa cognitiva.^[30]

^[30] Meirelles, I. (2013) *Design for Information: An Introduction to the Histories, Theories, and Best Practices Behind Effective Information Visualizations*. Rockport Publishers, Beverly, MA, pp 138

A escolha do método de escala é talvez o maior desafio nos mapas de símbolos proporcionais, bem como nos coropléticos. Há duas técnicas para escalar o tamanho dos símbolos: classificado, quando o tamanho varia gradualmente, e não classificado, quando os tamanhos seguem um sistema proporcional. Em sistemas não classificados, o número de categorias é igual ao número dos valores dos dados. Se houver cinco valores, haverá cinco codificações. Para representar um número grande de valores, a estratégia mais comum é utilizar percentagens.

Duas questões devem ser consideradas em sistemas classificados porque influenciam a maneira como os dados são representados e percebidos: o número de classes e o método para dividir os dados. Um número diferente de classes influencia os padrões revelados na visualização, para tal deve-se fazer experiências com número de grupos antes de tomar uma decisão. É recomendado primeiro analisar os dados para compreender certas características, como a distribuição.

Por exemplo, usar métodos de quartil, para dividir as quantidades em grupos de números iguais. Mas se a intenção for representar números distorcidos, esta não é a melhor opção porque os valores idênticos serão divididos em grupos diferentes e diferentes valores serão agrupados. Os valores da cor e da saturação podem ser usados para mapear outros atributos dos dados, no entanto, a percepção é dificultada em símbolos pequenos.^[31]

^[31] Meirelles, I. (2013) *Design for Information: An Introduction to the Histories, Theories, and Best Practices Behind Effective Information Visualizations*. Rockport Publishers, Beverly, MA, pp 139.

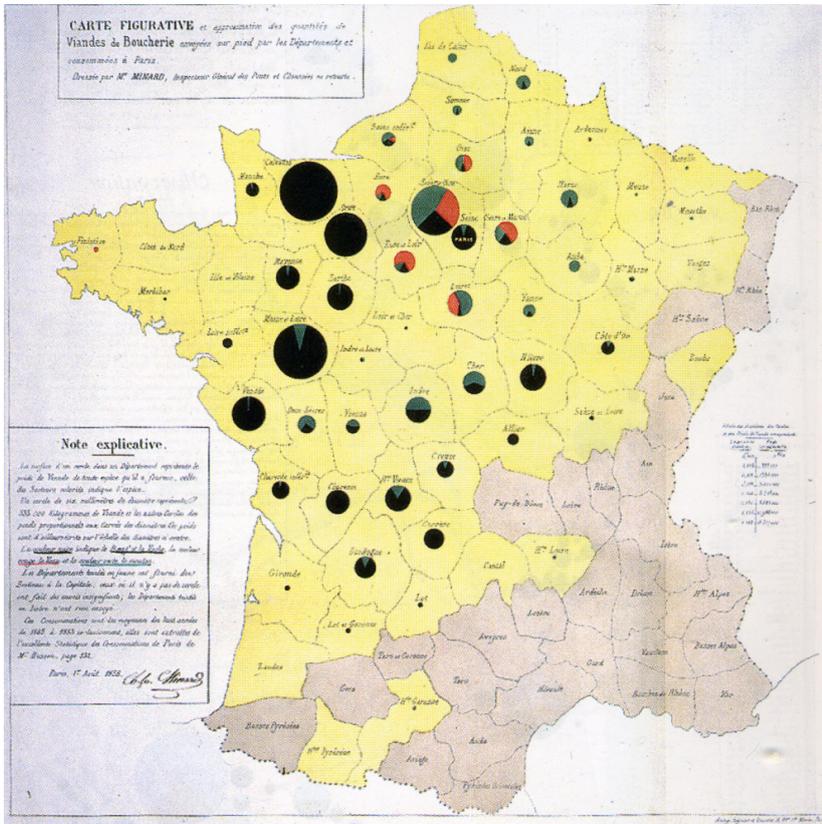
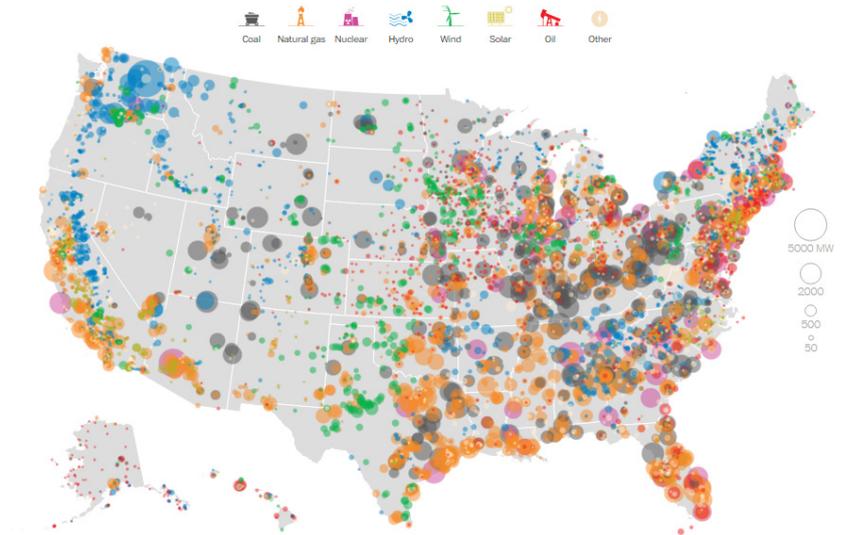


Figura 14 - “Carte figurative et approximative des quantités de viandes de boucherie envoyées sur pied par les départements et consommées à Paris”, Charles Minard, 1858.

O mapa da figura 14, criado pelo engenheiro francês Charles Minard, em 1858, foi o primeiro a usar *pie chart* graduados, para descrever mais do que uma variável no mapa.^[32] Cada círculo é dividido e dimensionado para representar as quantidades proporcionais de carne fornecida por cada distrito de França ao mercado de Paris entre 1845 e 1853. A área do círculo indica o peso de carne fornecida e a codificação da cor representa o tipo de carne, se for verde é carne de carneiro/ovelhas, se for vermelho bezerros e preto carnes vaca. Os distritos que fornecem carne aparecem pintados a amarelo como cor de fundo e os distritos que não fornecem carne sem cor (bege claro).

^[32] Meirelles, I. (2013) Design for Information: An Introduction to the Histories, Theories, and Best Practices Behind Effective Information Visualizations. Rockport Publishers, Beverly, MA, pp 138

Figura 15 - United States electricity map, criado pelo jornal The Washington Post, 2015.



^[33] The Washington Post. Mapping how the United States generates its electricity [online]. Acedido em 10 de junho de 2016. Disponível em <https://www.washingtonpost.com/graphics/national/power-plants/>.

O mapa de símbolos graduados da figura 15, criado pelo jornal The Washington Post,^[33] em 2015, representa cada estação de energia elétrica nos Estados Unidos. A cor dos círculos denota o tipo de estação que é, e o tamanho descreve a capacidade de energia, que pode de ser de 50 megawatts, 500 megawatts, 2000 megawatts ou 5000 megawatts.

As usinas movidas a carvão são representadas pela cor cinza escuro (preto). Este combustível é a fonte de eletricidade mais comum, a par do gás natural, no país. Nele existem cerca de 511 usinas que geram 34 por cento da eletricidade total num ano. Nos anos 80 era a principal fonte de energia da nação, mas agora cria apenas um terço da energia total.

O gás natural é representado pela cor laranja. Existem 1.740 usinas movidas a gás natural nos Estados Unidos que geram cerca de 30 por cento da energia elétrica do país num ano. A cor-de-rosa são representadas as usinas movidas a energia nuclear. Existem 99 reatores em 63 centrais elétricas nucleares nos Estados Unidos que geram 20 por cento da energia elétrica do país num ano.

A azul são representadas as usinas hidroelétricas, existem cerca de 1436 em todo o país que geram 7 por cento da energia elétrica num ano e geram aproximadamente 48 por cento da energia em cinco estados.

A energia elétrica gerada pelas usinas movidas pelo vento nos Estados Unidos e representada pela cor verde. Existem cerca de 853 em todo o país que representam aproximadamente 5 por cento da energia produzida num ano e estão localizadas em zonas onde o vento é mais forte e os espaços são abertos, como Dakota. A Amarelo escuro são representadas as usinas elétricas movidas pela energia solar. Nos Estados Unidos há cerca de 772 que produzem 1 por cento da energia total do país num ano. Estão localizadas principalmente no sudoeste do país, onde o sol é mais forte, como a Califórnia.

Por fim, as usinas movidas pelo petróleo, são representadas pela cor vermelha. Existem aproximadamente 1098 centrais que produzem cerca de 1 por cento da energia elétrica do país num ano e mostram que o petróleo não é mais uma fonte popular de geração de eletricidade.

2.3.3 Mapas coropléticos

Os mapas coropléticos são talvez a técnica mais popular para representar dados estatísticos usando a área.^[34] Tipicamente exibem dados que foram agregados por unidades administrativas ou valores que foram normalizados, por exemplo, densidades e médias.

Um dos problemas dos mapas coropléticos é o tamanho da área de codificação que influencia a percepção da quantidade que está a ser representada. Para evitar confundir a área geográfica com os valores de dados, é essencial normalizar os dados, em vez, de utilizar dados absolutos. As densidades e médias devem ser

[34] Pine, J. C. (2009). Natural hazards analysis: reducing the impact of disasters. Boca Raton, FL.: Auerbach Publications. pp 102.

calculadas antes da codificação. As variáveis visuais utilizadas nos mapas coropléticos, para codificar dados quantitativos, são o valor da cor, saturação da cor e a textura, ou a combinação de ambas. O tom da cor é muitas vezes usado para diferenciar dados categóricos no caso de mapas multivariados.

O valor da cor e da saturação são variáveis ordenadas, enquanto a tonalidade da cor não é. Esta é a razão pela qual o valor da cor é geralmente utilizado em mapas coropléticos, representando uma gama de dados classificados. Os elementos a considerar na conceção de um mapa coroplético são o tamanho, a forma da área, o número de classes, e o método que é utilizado para classificar os dados.

Devido ao facto de a codificação visual ser uniformemente distribuída dentro das regiões dos mapas coropléticos, a impressão que dá é que os fenómenos representados também são uniformemente distribuídos, o que não acontece na maioria das vezes. A impressão geral dos fenómenos será mais significativa se as áreas estatísticas tiverem uma forma semelhante e tamanho reduzido. Sempre que possível é recomendado evitar o uso de áreas com grande variação de tamanho e forma.

Tal como, no caso dos mapas de símbolos graduados, o número de classes e a forma como os dados são divididos em grupos, influenciam os padrões resultantes. A distribuição dos dados fornece informações significativas para o número de classes. A classificação dos dados, influência, em boa parte, quais as características dos dados que são enfatizadas e suprimidas. Se por um lado ter um elevado número de classes fornece resultados mais detalhados, por outro lado, existe limites para o número de classes que somos capazes de distinguir. No geral, é garantido conseguir-se diferenciar um máximo de cinco a oito classes, porque este intervalo encaixa numa zona cognitivamente eficiente. ^[35]

^[35] Meirelles, I. (2013) Design for Information: An Introduction to the Histories, Theories, and Best Practices Behind Effective Information Visualizations. Rockport Publishers, Beverly, MA, pp 142.

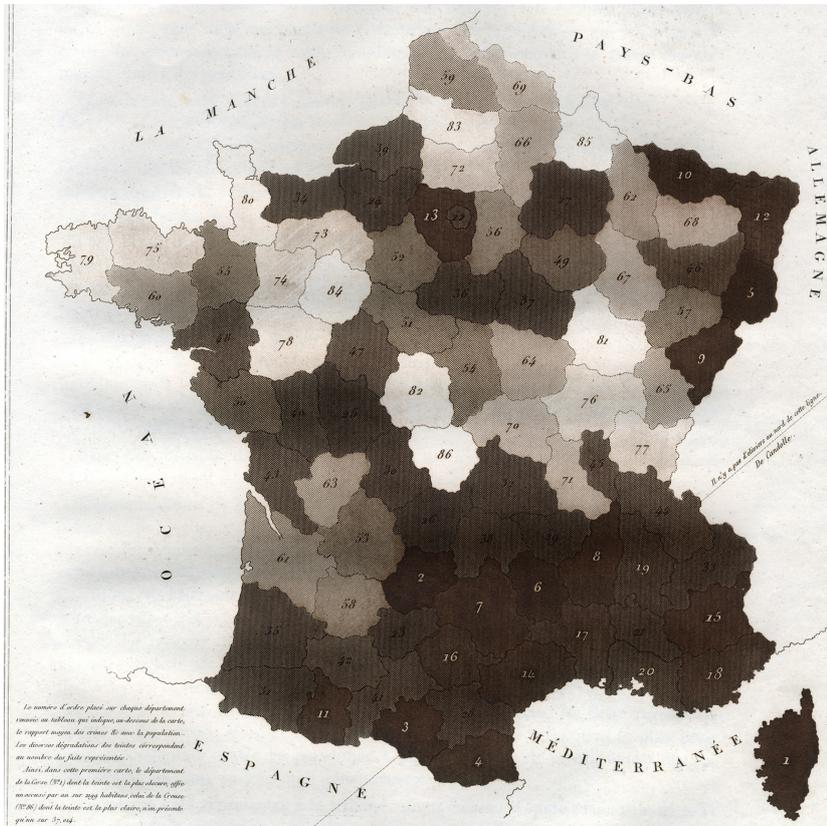


Figura 16 - “Crimes contre les personnes”, criado por André Michel Guerry em 1833.

O mapa coroplético da figura 16, criado pelo advogado e estatístico André Michel Guerry, em 1833, foi o primeiro a mapear a criminalidade em França, entre 1825 e 1830, por distrito. Para tal, ele utilizou as estatísticas sobre os crimes publicados pelo departamento judiciário francês. Ao estudar as tabelas com os dados ficou impressionado com a regularidade com que, ano após ano, cometiam-se crimes graves no país. Um dos temas que mais o fascinou foi a incidência numérica dos crimes em relação aos níveis de educação.

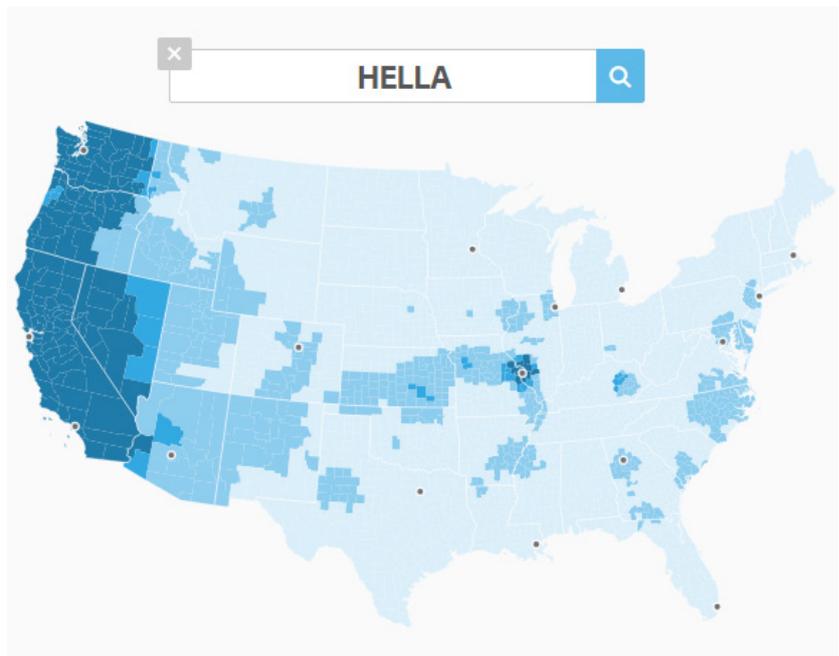
Sendo assim, avaliou o nível de educação por referência à alfabetização, usando dados sobre o número de alfabetizados e analfabetos entre os recrutas para o exercito francês. Esses números mostraram que, contrariamente a todas as suposições e declarações que foram feitas, as regiões com mais altos níveis de educação apresentaram a maior incidência de crimes.^[36]

^[36] Friendly, M. (1802-1866) The Life and Works of André-Michel Guerry, York University, Toronto, Canadá.

^[33] Meirelles, I. (2013) *Design for Information: An Introduction to the Histories, Theories, and Best Practices Behind Effective Information Visualizations*. Rockport Publishers, Beverly, MA, pp 143.

Guerry utilizou diferentes tons de cor, cerca de sete, para representar os diferentes níveis de criminalidade. O castanho escuro indica os distritos com uma taxa alta de criminalidade e o branco os que tem uma taxa baixa. Os distritos são numerados em função da taxa de criminalidade, do maior para o mais pequeno, e o exato número de crimes é dado na tabela debaixo do mapa.^[33]

Figura 17 - “The great American word mapper”, criado por Nikhil Sonnad, 2014.



^[37] Quartz. The great American word mapper. Acedido em 17 julho de 2016. Disponível em https://qz.com/862325/the-great-american-word-mapper/#int/words=dinner_supper&smoothing=3.

Na imagem da figura 17 é apresentado um conjunto de mapas coropléticos, criado pelo jornalista e programador Nikhil Sonnad, da *Quartz*.^[37] Eles permitem visualizar as 100.000 palavras mais utilizadas no *Twitter* nos Estados Unidos, categorizando-as por assunto, e em seguida, por região. Para realizar este projeto Sonnad utilizou dados extraídos de bilhões de *tweets* coletados pelo geógrafo Diansheng Guo em 2014. Jack Grieve, linguista forense da Universidade Aston no Reino Unido, juntamente com Andrea Nini da Universidade de Manchester, identificou o top

100.000 de palavras usadas no *Twitter* e quantas vezes elas foram utilizadas em cada estado nos Estados Unidos, continental, com base em dados de localização. A gama de cor é a variável visual que identifica a intensidade com que as palavras são utilizadas, num total de quatro classes, e a forma o estado em questão. Sendo assim, o azul escuro representa as palavras mais comuns e o azul claro (turquesa) as palavras menos comum. O utilizador tem a possibilidade de pesquisar e visualizar as palavras por cidades, alimentos, equipas desportivas, palavras de calão e outras referências culturais. Ao mesmo tempo que compara diferentes padrões nos mapas, tem a possibilidade de os partilhar.

2.3.4 Mapas isométricos e isopletos

Isarithms ou *isarithmetic lines*, normalmente chamadas de *isolines*, é um termo genérico para uma linha que junta todos os pontos que têm o mesmo valor acima ou abaixo de algum *datum* ou ponto de referência. Existem vários tipos de *isarithmetic lines*, a maioria tem nomes específicos, embora todas sejam baseadas no mesmo conceito.^[38] Algumas das linhas frequentemente encontradas estão descritas na tabela da figura 18.

^[38] Tyner, J. A. (2010) *Principles of Map Design*. The Guilford Press, New York, NY, pp 170.

Isarítmos normalmente utilizados	
Isobars	Linhas de igual pressão barométrica
Isobaths	Linhas de igual profundidade de água
Isochrones	Linhas de igual tempo mínimo de viagem
Isogones	Linhas de igual magnetismo terrestre (declinação)
Isohyets	Linhas de igual precipitação
Isohypsies	Linhas de igual elevação do solo (contour lines)
Isotherms	Linhas de igual temperatura

Figura 18 – Tabela com os tipos de linhas isaritmicas. Tyner (2010) *Principles of Map Design*.

Se passarmos um plano horizontal por uma superfície estatística, a linha de intersecção do plano e da superfície quando projetada para baixo é uma *isarithmic line*. Podemos ver que cada ponto ao longo da linha de intersecção tem a mesma distância do *datum*. Se dois planos paralelos intersectarem a superfície, a distância entre os planos é chamada de ***isarithmic interval***. Ver imagem da figura 19.

Existem duas categorias principais de *isarithms*, as **linhas isométricas** e as **linhas isopletas**, embora os termos genéricos *isarithm* ou *isoline* sejam mais utilizados. Estas categorias são baseadas na natureza dos valores de *z* e na forma como os dados são recolhidos.

Se os dados forem recolhidos em pontos reais, como estações meteorológicas, ou pontos de controlo de elevação, para os quais os valores realmente existem (pontos verdadeiros), o tipo de linha é isométrica. Os valores de *z* podem ser absolutos, tal como a temperatura em um específico ponto, derivados, como a temperatura média nesse.

Se os pontos para os quais os dados obtidos são assumidos ou arbitrários (pontos conceituais), como o centro de distribuição ou o centro de uma área de enumeração como um município, os valores referem-se a uma área e não a um ponto, e os valores de *z* são derivados, isto é, proporções, médias e percentagens. Este tipo de linha é isopleta (*isopleth*) ou isoplética (*isoplethic*). Isopletas (*isopleths*) são frequentemente utilizados para representar densidades populacionais e dados similares.^[39]

^[39] Tyner, J. A. (2010) Principles of Map Design. The Guilford Press, New York, NY, pp 171.

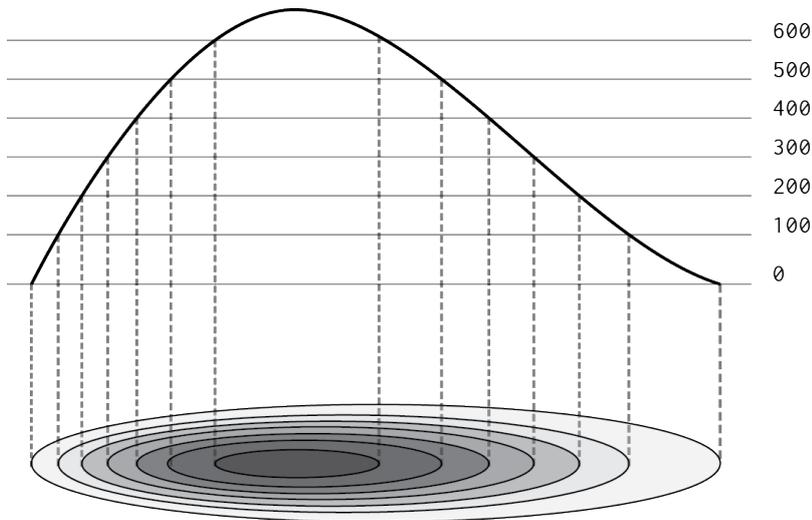


Figura 19 - Formação de linhas isométricas.

Apesar dos métodos de construção destes dois tipos de *isarithmetic lines* sejam semelhantes, o tipo de informação que eles retratam e a que o leitor pode extrair dos mapas são diferentes. Uma linha isométrica permite que os valores sejam determinados interpolando ou estimando os valores entre as linhas.

Interpolação é a estimativa de valores pela comparação das distâncias a todos os outros pontos de controle vizinhos. Quando se lê um mapa isométrico, o leitor interpola ou “lê entre as linhas” para estimar o valor de qualquer ponto. Se o intervalo *isarithmetic* for par, a taxa de mudança é encontrada a partir do espaçamento dos *isarithmetic*. As linhas eventualmente espaçadas representam uma taxa uniforme, as linhas estreitamente espaçadas representam um gradiente íngreme, e as linhas amplamente espaçadas representam um gradiente suave.

As linhas isopletas não permitem a estimativa de valores em qualquer ponto porque os pontos de controle representam valores de áreas, não pontos reais. Assim, permitem ao leitor observar padrões e gradientes gerais. Da mesma forma que a taxa de mudança pode ser determinada a partir de um mapa isométrico, áreas de mudança rápida ou gradual podem ser encontradas em mapas isopleto.

Antes de começar a recolher dados, um pesquisador seleciona os pontos de controle para os quais os dados são recolhidos. Nos mapas isométricos, esses pontos podem já existir, como estações meteorológicas, mas para os mapas isopleto os pontos são conceituais e devem ser escolhidos. Há muitas opções, e a escolha do ponto de controle faz a diferença na aparência do mapa. Em alguns tipos de estudos, especialmente estudos biogeográficos, os pontos podem ser simplesmente espaçados regularmente. Neste caso, deve-se evitar um padrão retangular porque resulta em situações ambíguas. Triangular ou hexagonal funcionam melhor.

Para mapas isométricos o intervalo *isarithmetic*, idealmente, é sempre o mesmo. Isso torna o mapa mais fácil de interpretar uma vez que o espaçamento das linhas diz ao leitor a natureza da superfície. Se um for escolhido um intervalo desigual, o espaçamento entre as linhas pode ser enganador. Para mapas isopleto, os dados geralmente apresentam diferenças acentuadas e amplas variações, dificultando a escolha de um intervalo par. Neste caso, o intervalo irregular deve ser claramente mostrado na legenda. As técnicas para escolher uma categoria para os mapas coropleto podem ser usadas nos mapas isopleto.

A largura das linhas nos mapas *isarithmetic* são uniformes. Assim, a principal consideração de design é a sombra entre as linhas e o esquema utilizado para sombreamento. Se outras informações forem mostradas no mapa, como nos mapas topográficos, as linhas são simplesmente numeradas e não há nenhuma outra variável. No entanto, se os dados representados pelas linhas são o assunto principal do mapa, o sombreamento chamado de *layer tints* é adicionado para tornar o mapa mais fácil de ler.

As linhas que mostram elevações são normalmente chamadas de *contour lines* e os tons de *hypsometric tints*. As opções para camadas de tons são: a luminosidade (valor do tom), *double-ended*, sequências espectrais.



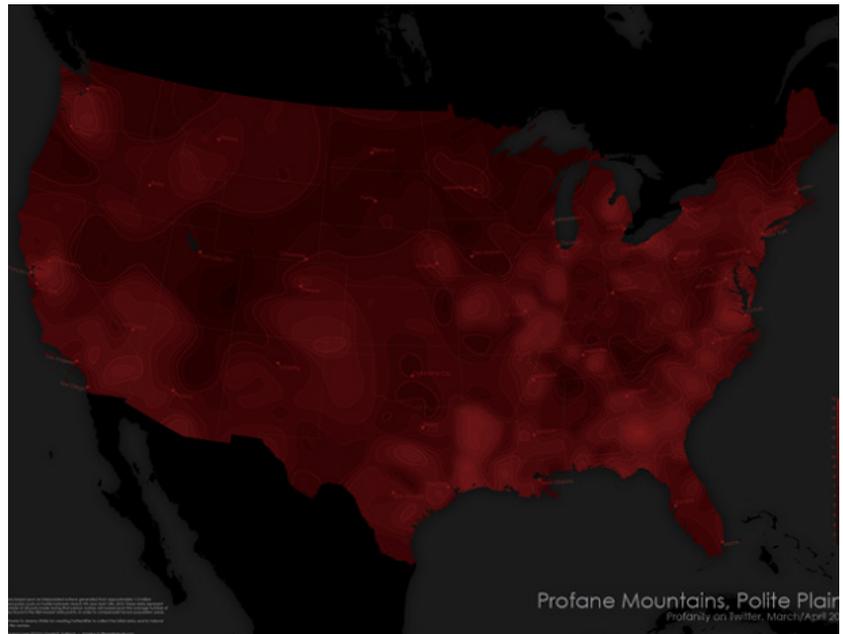
Figura 20 - Mapa de isolinhas que representa campos magnéticos, publicado em 1701 pelo inglês Edmond Halley.

O mapa de linhas magnéticas, criado por Edmond Halley em 1700, e publicado em 1701 é considerado o primeiro mapa a fazer uso de linhas de igual valor, *isolines*, para codificar dados. A declinação ou variação magnética é a diferença angular entre o norte magnético e o norte geográfico para qualquer ponto na superfície da Terra. Se o ângulo for maior que o norte geográfico, a variação é leste, se o ângulo for inferior ao norte geográfico, a variação é ocidental e se ambos estiverem na mesma direção, não há variação.

Na época, os navegadores marítimos podiam calcular a latitude onde quer que estivessem observando o Sol, desde que fosse visível para eles. O interesse de Halley pela declinação

magnética, prendia-se em resolver o problema em calcular a longitude. Os dados utilizados resultaram de observações que o mesmo coletou durante duas viagens marítimas. A maioria dos mapas de *isoline* que encontramos, hoje em dia, representam fenômenos naturais como, temperaturas ou precipitação.

Figura 21 - “Profane Mountains, Polite Plains”, criado por Daniel Huffman, 2010.



O mapa da figura 21, criado por Daniel Huffman, em 2010, representa os locais onde as pessoas são mais propensas a jurar. Este mapa mostra a frequência com que algumas profanidades aparecem em um *tweet* com base na localização do mesmo. As *isolines* são baseadas em uma superfície gerada a partir de aproximadamente 1.5 milhões de *tweets* públicos no *Twitter* entre 9 de março e 12 de abril de 2010.

Estes dados representam apenas uma amostra de todos os *tweets* feitos durante esse período. As *isolines* baseiam-se no número médio de profanidades encontradas nos 500 pontos mais próximos, a fim de compensar as áreas de baixa população. Quanto mais clara é a cor maior é o número de profanidades.

2.3.5 Mapas de fluxos

Os mapas ou redes de fluxos retratam fenómenos lineares que envolvem a conexão de lugares. Mostram o movimento de bens, pessoas e migrações a partir de um ponto de origem, passando por locais intermédios e, finalmente, o destino.^[40]

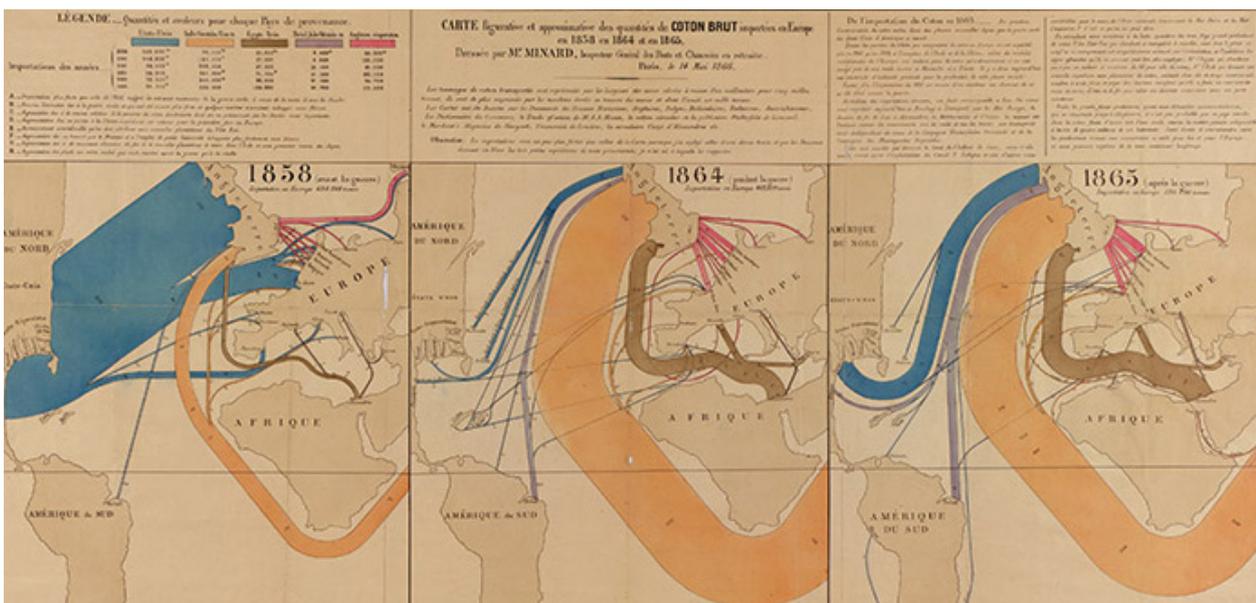
A maioria dos mapas codifica os dados através de atributos visuais como espessura da linha, ou o peso, geralmente para representar o volume de fluxo. Às vezes, são utilizadas linhas pontilhadas, tracejadas, sólidas e duplas para mostrar relações ordinais entre os fluxos. As linhas podem ser coloridas para representar diferentes cargas ou suportes.

Como muitos mapas de fluxo são de pequena escala, as linhas de fluxo não representam rotas reais e devem ser usadas apenas como uma indicação geral de conectividade. Mapas de fluxo em grande escala, como mapas de trânsito de passageiros, representam melhores aproximações de trajetos reais de mapas de viagem. A migração e o movimento de bens comerciais são commumente representados por mapas de fluxo.^[41]

^[40] Johnson, J. M (2003) Geographic Information. Greenwood Press, Westport, CT, pp 48.

^[41] Meirelles, I. (2013) Design for Information: An Introduction to the Histories, Theories, and Best Practices Behind Effective Information Visualizations. Rockport Publishers, Beverly, MA, pp 152-153.

Figura 22 - “Map of European Cotton Imports during the American Civil War”, criado por Minard, 1866.



O mapa de fluxo da figura 22, criado pelo engenheiro francês Charles Minard, mostra o impacto da guerra civil americana, entre 1861 e 1865, sobre os fluxos de algodão Europeu. Os fluxos antes, durante e depois da guerra são retratados com faixas coloridas. A largura das faixas representa a quantidade de algodão transportado. Antes da guerra civil a maior parte dos países europeus exportava a matéria-prima exclusivamente do sul dos Estados Unidos da América, representado pela faixa azul. Com a guerra civil houve um bloqueio das exportações e os padrões de comércio mudaram.

A estratégia era simples, as fábricas de têxteis britânicas e francesas dependiam do algodão do sul, e se esse algodão fosse cortado por causa da guerra, forçaria as potências europeias a intervir no conflito para salvar as suas indústrias. No entanto, a estratégia falhou, apesar da quase eliminação do algodão do sul do mercado internacional durante a guerra. O mapa de Minard mostra porquê. A europa começou a exportar o algodão da Ásia e da África, representado no mapa pelas linhas amarelas e castanhas.

Existem vários aplicativos que permitem controlar o trajeto de atividades físicas, gravando a velocidade, o tempo, a elevação, e a localização no telemóvel. Mas quando olhamos para os cronogramas as atividades aparecem individualmente.

^[42] FlowingData.Where People Run in Major Cities. Acedido em 8 agosto de 2016. Disponível em https://qz.com/862325/the-great-american-word-mapper/#int/words=dinner_supper&smoothing=3.

A visualização^[42] criada por Nathan Yau, em 2014, não só permite visualizar as atividades em conjunto, como proporciona aos coletores dos dados planejar os exercícios futuros de uma forma melhor. Ao mesmo tempo permite aos Urbanistas, certificar-se que os cidadãos têm ciclovias adaptadas e caminhos para correr.



Figura 23 - "Where People Run in Major Cities" (Dallas), criado por Nathan Yau, 2014.

Os dados utilizados para desenvolver esta visualização são fornecidos pela aplicação *RunKeeper*, em várias cidades no mundo. Podemos observar que as pessoas preferem fazer exercício físico longe dos carros e das grandes artérias, deslocando-se para perto da água ou mesmo em parques. Isto pode ser visualizado na imagem da figura 23, onde as linhas são mais grossas existe uma maior afluência de pessoas.

2.3.6 Cartogramas de áreas e de distâncias

Os cartogramas são mapas abstratos, que combinam informação estatística com localização geografia. Enquanto os mapas físicos ou topográficos apenas mostram áreas relativas e distâncias, estes distorcem as áreas e as distâncias para transmitir dados. As áreas e as distâncias são aumentadas e reduzidas para refletir o valor do atributo que está a ser descrito no mapa. Por exemplo, a forma de um estado pode ser proporcionada ao número de habitantes.^[43]

^[43] Johnson, J. M. (2003). *Geographic information: how to find it, how to use it*. Westport, CT: Greenwood Press. pp 48.

Os **cartogramas de áreas** foram concebidos com a finalidade de revelar padrões geográficos e espaciais. A forma e a localização são mantidas ao máximo. É importante salientar que existem dois tipos de cartogramas de área, os contíguos e os não contíguos. Nos cartogramas de área não contíguos, os objetos não têm de manter a conectividade com os adjacentes. Ao libertar os objetos adjacentes, eles podem crescer ou diminuir de tamanho, mantendo a forma. Ao contrário dos cartogramas de área contíguos, em que a forma é sacrificada para preservar a topologia

Os **cartogramas de áreas** foram concebidos com a finalidade de revelar padrões geográficos e espaciais. A forma e a localização são mantidas ao máximo. É importante salientar que existem dois tipos de cartogramas de área, os contíguos e os não contíguos. Nos cartogramas de área não contíguos, os objetos não têm de manter a conectividade com os adjacentes. Ao libertar os objetos adjacentes, eles podem crescer ou diminuir de tamanho, mantendo a forma. Ao contrário dos cartogramas de área contíguos, em que a forma é sacrificada para preservar a topologia.

Um **cartograma de distâncias** tipicamente mostra a relação entre o tempo e distância a partir de um ponto numa rede. Assim, os cartogramas de distância mostram a distribuição espacial dos atributos num mapa. Isto permite ao leitor visualizar facilmente conceitos, tais como, o curso do tempo e comparar distâncias umas com as outras.^[44]

^[44] Meirelles, I. (2013) Design for Information: An Introduction to the Histories, Theories, and Best Practices Behind Effective Information Visualizations. Rockport Publishers, Beverly, MA, pp 156-157.

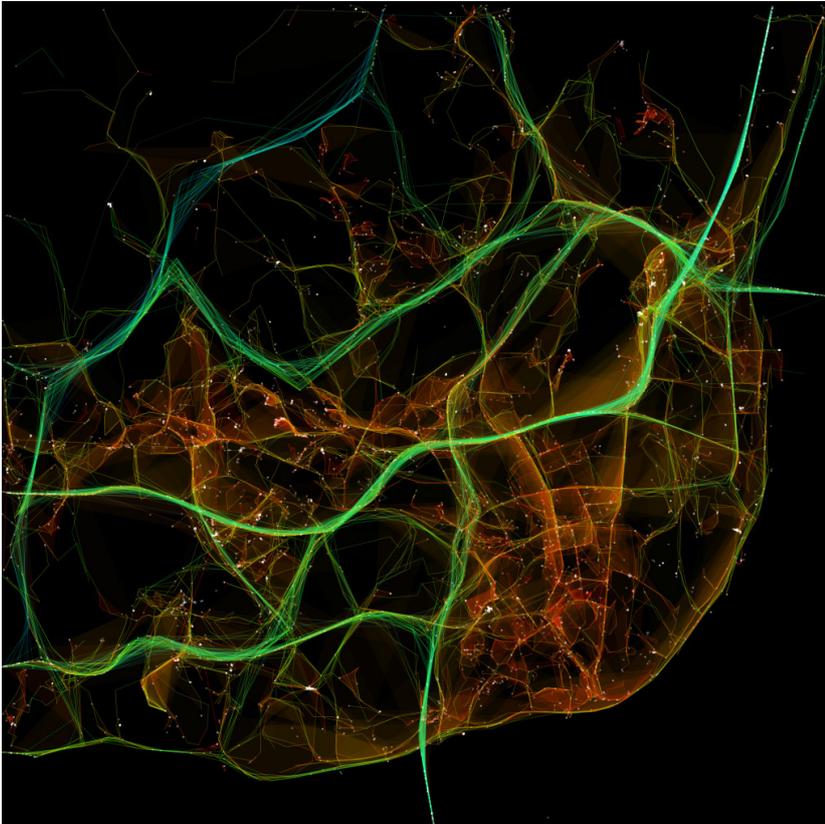


Figura 24 - Pedro Cruz, Penousal Machado, and João Bicker (Universidade de Coimbra com MIT CityMotion), Portugal: "Traffic in Lisbon," 2010.

O projeto "*Traffic in Lisbon*", é uma série de animações da evolução do tráfego em Lisboa, durante um período de 24 horas.^[45] Nele são mapeados 1.534 veículos durante o mês de outubro de 2009 e exploradas diferentes metáforas visuais da cidade como um organismo com problemas circulatórios. Na imagem da figura 24, os caminhos são codificados por cores de acordo com a velocidade dos veículos: verde e ciano para os veículos mais rápidos; amarelo e vermelho para os mais lentos. Este é um bom exemplo de um cartograma de distâncias em que as estradas são dilatadas ou comprimidos consoante o fluxo de carros.

^[45] Visualization et al. Traffic in Lisbon and the visualization of trajectories. Acedido em 20 agosto de 2016. Disponível em <http://pmcruz.com/information-visualization/traffic-in-lisbon-condensed-in-one-day>.

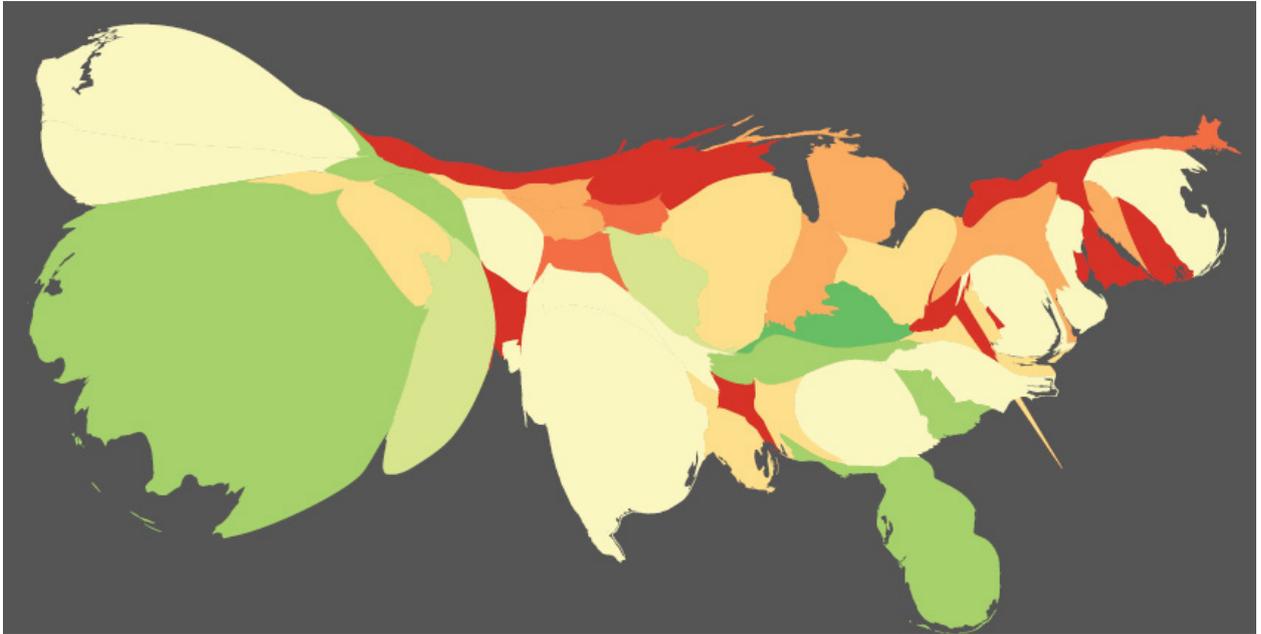


Figura 25 - “Pulse of the Nation”, criado por Alan Mislove, Sune Lehmann, Yong-Yeol Ahn, Jukka-Pekka Onnela, and J. Niels Rosenquist, 2011.

O projeto “*Pulse of the Nation*”, é um cartograma de área, que examina o humor dos Estados Unidos inferindo mais de 300 milhões de *tweets* coletados entre o mês de setembro de 2006 e agosto de 2009. A imagem da figura 25, pertence a um conjunto de cartogramas, criados em 2011 por uma equipa interdisciplinar da Universidade Northeastern e Harvard, que preservam a forma geográfica o máximo possível. Os tamanhos das aéreas dos estados são dimensionados de acordo com número de *tweets* em cada região. O resultado é um mapa que uniformiza as densidades. A cor codifica o nível de humor, por meio de uma escala, que varia entre o vermelho (infeliz), o amarelo (neutro) e o verde (feliz). A partir desta representação é possível observar tendências tanto como variações de humor, em horas específicas do dia, como variações geográficas entre a costa oeste e leste.

2.4 Casos de estudo

Nesta secção, com o propósito de evitar possíveis problemas práticos, são analisadas outras representações e animações gráficas relacionadas com a poluição e o vento. São estudadas as técnicas utilizadas para representar os dados, para que assim possa ser escolhida a melhor forma de visualizar e de interagir com a informação, com o intuito de facilitar a apresentação e organização dos dados, bem como, a comodidade de pesquisas por parte do utilizador.

2.4.1 Wind map

Desenvolvido pelos codiretores do grupo de pesquisa de visualização “*Big Picture*” da *Google*, Fernanda Viégas e Martin Watterberg em 2012, o mapa animado “*Wind Map*”^[46] foi criado com o intuito de levar dados horários do *National Digital Forecast Database* para apresentar fluxos de vento em todo território dos Estados Unidos. As informações, reproduzidas no mapa com dados em tempo real, formam uma visualização de informação que reproduz o movimento do vento de forma simples e inteligível. Nele, as linhas brancas representadas, indicam a intensidade das forças das rajadas dos ventos, com variações diárias de representação visual.

^[46] Wind map [online]. Acedido em 22 agosto de 2016. Disponível em <http://hint.fm/wind/>.

wind map

August 22, 2016

1:35 pm EST
(time of forecast download)

top speed: 31.6 mph
average: 8.6 mph

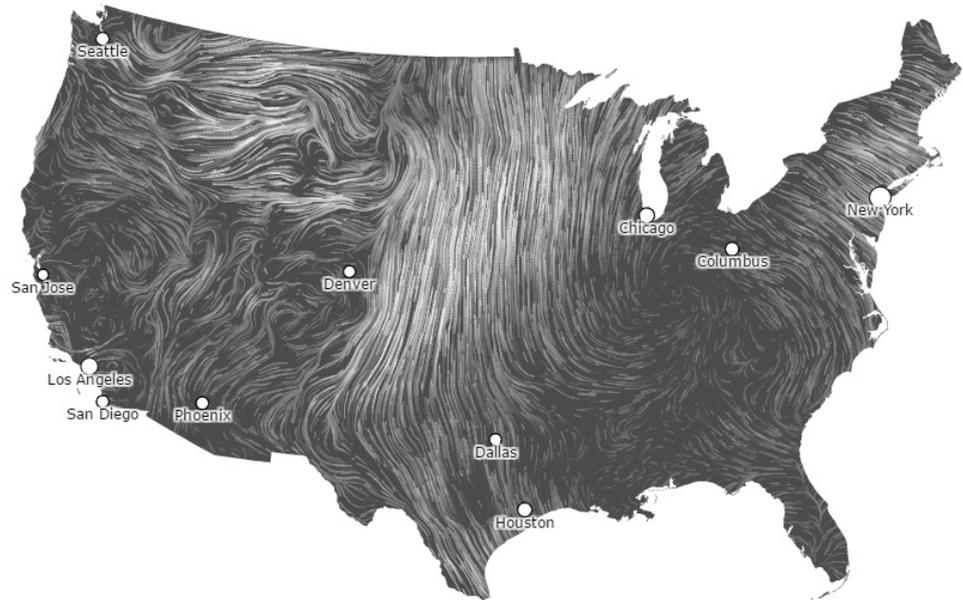


Figura 26 - Screenshot da visualização Wind Map criada pela Fernanda Viégas e Martin Wattenberg, 2012.

Estas representações visuais, apresentadas no mapa, podem ser utilizadas para demonstrar casos esporádicos de fenômenos naturais (furacões, tornados, etc) como o caso do furacão Sandy, registrado no dia 30 de outubro de 2012, e o furacão Isaac, registrado nos dias 28 e 29 de agosto de 2012. Nestes casos, consegue-se perceber, claramente, a concentração do vento em um ponto específico do mapa, ou seja, durante o período em que há mudanças meteorológicas intensas, como os casos citados, o mapa apresenta grande intensidade de vento em uma área específica do país, enquanto nas outras áreas, não atingidas, a concentração do vento diminui consideravelmente. A partir da visualização dos dados, representados no mapa, é possível perceber que a direção do vento, no país, move-se majoritariamente do Sul para o norte, com variações entre 10 e 30 mph, havendo algumas exceções pontuais nas extremidades do mapa onde o vento desenvolve-se com menor velocidade, atingindo de 1 a 5 mph.

Os investigadores do projeto, afirmam que o mapa do vento também é utilizado por diversos usuários para as mais diversas funções, desde observadores de pássaros que seguem os padrões de migração de acordo com o vento há ciclistas que planeiam viagens com informações retiradas do mapa e teóricos que utilizam essas informações para rastrear produtos químicos que se propagam no ar.

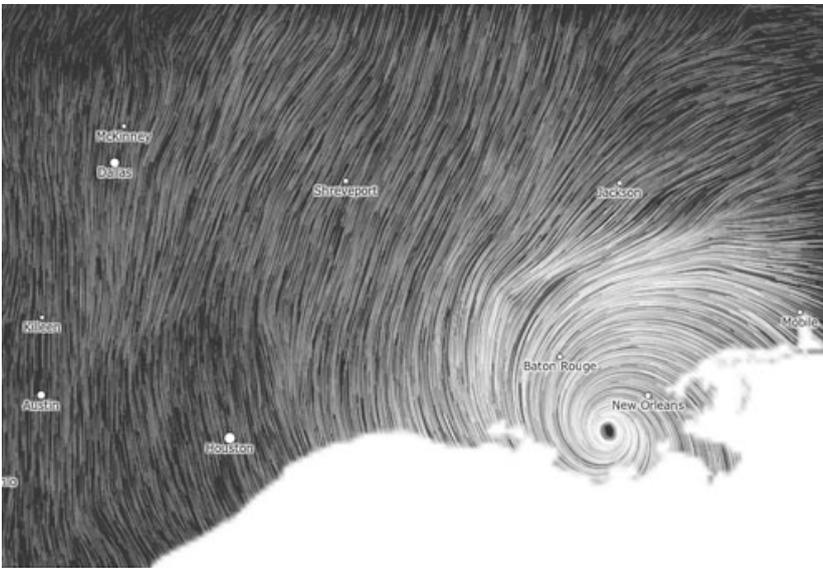


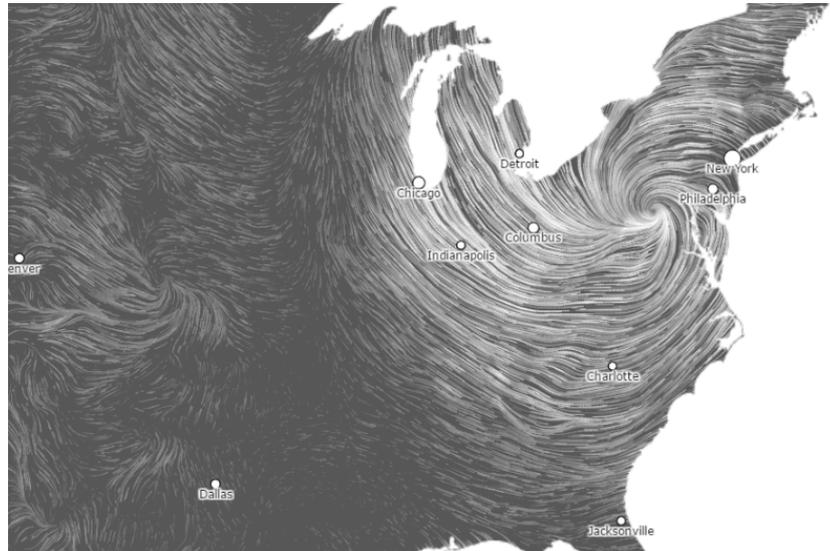
Figura 27 - Screenshot do furacão Isaac.

“As an artwork that reflects the real-world, its emotional meaning changes from day to day. On calm days, it can be a soothing meditation on the environment; during hurricanes, it can become ominous and frightening.” [47]

[47] Viégas, F; Wattenberg, M. (2012). Wind Map. Disponível em <http://hint.fm/projects/wind/>. Acedido em 22 de agosto de 2016.

O projeto, atualizado de hora em hora, utiliza um padrão de fácil distinção, devido à espessura e orientação das linhas que revelam a geografia escondida. O mapa consegue incorporar, simultaneamente, requisitos de visualização científica, com clareza e precisão, sem perder a atração visual, o impacto artístico e as qualidades metafóricas da arte.

Figura 28 – Screenshot do furacão Sandy.



[48] Wright, F. L. (1949). *Genius and the Mobocracy*, London in Kevin Nute: *Frank Lloyd Wright and Japan: The Role of Traditional Japanese Art and architecture in the work of Frank Lloyd Wright*, Routledge, London and New York, 2000, pp 80.

“Form follows function—that has been misunderstood. Form and function should be one, joined in a spiritual union.” [48]

Assim como referência Wright, arquiteto americano do século XIX, o projeto de Viégas e Watternberg introduzem em seu trabalho a preocupação com a forma e, em paralelo, a função, de informar o utilizador sobre os padrões, a força e as direções do vento. O mapa construído coloca cara a cara o sentido da estética sobre substância, aspetos do design que devem ser combinados e reunidos de maneira harmónica, sem comprometer nem um nem o outro, em uma união equilibrada, esteticamente convidativa e funcionalmente eficaz.

O projeto caracteriza-se como um exemplo de representação gráfica capaz de transmitir grandes quantidades de informação de uma forma rápida e intuitiva. Segundo Viégas e Wattenber^[49] existiam outras informações que poderiam ser implementados na sua função, como a temperatura, a precipitação ou a poluição, entretanto alegam que a intenção ao criar o mapa era focar-se no vento e na representação artística que resultaria desses dados.

A representação gráfica do “*Wind Map*”, ao usarem uma sucessão de linhas finas e sombreadas, trazem à tona uma associação artística às linhas utilizadas no período renascentista,^[50] embora neste mapa de fluxos, estejam representadas em constante movimentação. Por ser considerado uma obra artística com grande qualidade de representação, o mapa entrou em exposição no *Museum of Modern Art* (MoMA), em Nove Iorque, entre 2 de março de 2013 e permaneceu até o dia 31 de janeiro de 2014.

^[49] Viégas F., Wattenberg, M. (2013). Visualizing Natural and Cultural Phenomena. Disponível em <https://www.youtube.com/watch?v=gORdy0YEsGU>. Acedido em 27 de agosto de 2016.

^[50] Popeson, P. (2013). MoMA PS1 Blog. Collection & Exhibitions. Department Preparator, Department of Architecture and Design. Disponível em https://www.moma.org/explore/inside_out/2013/02/28/watching-the-wind-viegas-and-wattenbergs-wind-map/. Acedido em 25 de agosto de 2016.



Figura 29 - Exposição do Wind map no MoMA.

2.4.2 The world air quality index

O projeto social *The World Air Quality Index* (AQI), baseado em Pequim, na China, e iniciado em 2007, promove a publicação de mapas, através de uma fonte unificada de informações sobre a poluição, em todo o mundo, em tempo real. Os mapas fornecem aos habitantes dados atualizados sobre os níveis de partículas e gases nocivos no ar, provenientes de diversas fontes poluentes, desde ao fumo libertado pelos automóveis à queima de combustíveis fósseis, madeira, moléculas biológicas e outros materiais tóxicos presentes na atmosfera.

As representações gráficas do mapa são construídas a partir de informações disponibilizadas pelas *Environmental Protection Agencies* (EPAs) que exibem a localização dos sensores de qualidade do ar utilizando como base um mapa do *Google*. As informações, representadas no mapa, são recolhidas a partir de 9.000 estações em 600 grandes cidades de 70 países. São conhecidas mais 20.000 estações de monitoramento da qualidade do ar ao redor do mundo, das quais 9.000 são publicadas no *World Air Quality Index*. Essas publicações, que assumem um alto nível de consistência, publicam apenas informações de estações que apresentem partículas suspensas (PM2.5, PM10).

Ademais, a fim de garantir um alto nível de precisão, para os valores do Índice de Qualidade do Ar (AQI), diversas máquinas de processamento são utilizadas. As informações extraídas, por serem verificadas em tempo real, possibilitam detetar possíveis estações de monitoramento defeituosas, de modo a removê-las, caso seja necessário, da representação gráfica no mapa. Em sua representação gráfica, as categorias dos níveis de qualidade do ar, apresentam os dados através de diferentes cores e numerações, determinadas consoante a intensidade de poluentes encontrados em cada região analisada. Os dados utilizados na construção desta visualização são recolhidos constantemente através de equipamentos profissionais, em diversas estações ao redor do mundo.

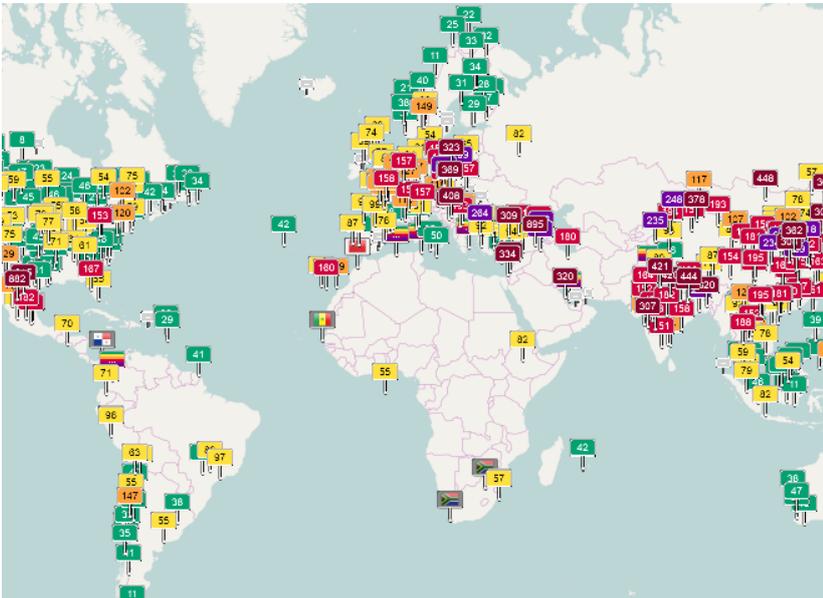


Figura 30 – Screenshot do mapa geral, The World Air Quality Index.

Durante a recolha de informações, são obtidos um conjunto de dados sobre os mais variados gases poluentes, como partículas suspensas PM2.5, PM10, Ozono, Dióxido de Nitrogénio, Dióxido de Enxofre e Monóxido de Carbono. Os cálculos para a obtenção dos valores de qualidade do ar são obtidos através de três inputs; o poluente, a unidade e a concentração. Os resultados desses cálculos, gerados a partir de um conversor específico (*AQI Calculator: Concentration to AQI*),^[51] mostram o Índice de Qualidade do Ar de cada área. A partir do valor resultante o Índice de Qualidade do Ar é classificado entre diferentes grupos conforme a sua periculosidade em relação aos habitantes.

^[51] AQI Calculator: Concentration to AQI. Disponível em <http://aqicn.org/faq/2015-03-20/a-comparison-of-worldwide-air-quality-scales-part-1/>. Acedido em 10 de agosto de 2016.

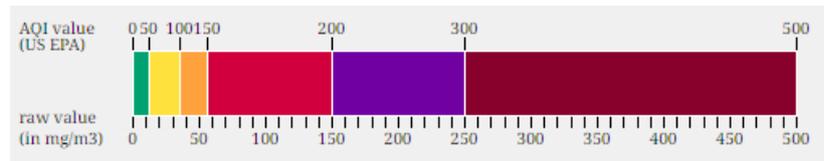
Figura 31 – Screenshot da legenda dos níveis de perigo para a saúde consoante o do índice de qualidade do ar.

AQI	Air Pollution Level	Health Implications
0 - 50	Good	Air quality is considered satisfactory, and air pollution poses little or no risk
51 -100	Moderate	Air quality is acceptable; however, for some pollutants there may be a moderate health concern for a very small number of people who are unusually sensitive to air pollution.
101-150	Unhealthy for Sensitive Groups	Members of sensitive groups may experience health effects. The general public is not likely to be affected.
151-200	Unhealthy	Everyone may begin to experience health effects; members of sensitive groups may experience more serious health effects
201-300	Very Unhealthy	Health warnings of emergency conditions. The entire population is more likely to be affected.
300+	Hazardous	Health alert: everyone may experience more serious health effects

^[52] World Air Quality Index Project. A comparison of worldwide Air Quality Scales - Part 1. Publicado em 20 de março de 2015. Disponível em <http://aqicn.org/faq/2015-03-20/a-comparison-of-worldwide-air-quality-scales-part-1/>. Acesso em 10 de agosto de 2016.

Segundo o projeto *World Air Quality Index*, existem diversos tipos de escalas para a classificação dos níveis de perigo do ar, ao redor do mundo, o que deixa a leitura do Índice de Qualidade do Ar mais complexo. Na China, assim como vários outros países, a escala utilizada é própria e conhecida como HJ 633-2012,^[52] no entanto, no projeto em questão, a escala utilizada baseiam-se no padrão das EPAs dos Estados Unidos. Devido à grande variedade de escalas encontradas, a *World Health Organisation* (WHO) está disponibilizar diretrizes para a análise da qualidade do ar com base em revisões de evidências científicas acumuladas.

Figura 32 - Screenshot da escala padrão Environmental Protection Agencies USA, utilizada no projeto.



O mapa visual do Índice de Qualidade do Ar é em tempo real, no qual representa a poluição atmosférica no mundo, alguns dados em especial foram escolhidos para resumir e simplificar a transmissão das informações sobre cada local indicado no mapa. Em um primeiro momento, em cada país ou cidade (a depender do tamanho do local e da quantidade de informações disponíveis) é apresentado placas que sinalizam o Índice de Qualidade do ar e com a cor referente ao nível de periculosidade de cada local específico do mundo. Em um segundo momento, ao escolher o local em que o utilizador deseja obter mais informações, é apresentado uma janela na qual mostra as informações sobre a qualidade do ar de maneira mais detalhada. Nela, são apresentados gráficos referentes ao Índice de Qualidade do Ar de todos os poluentes encontrados no local, tais como Dióxido de Nitrogênio (NO₂), Dióxido de Enxofre (SO₂), Ozono, e as suas respectivas variações durante determinados períodos na semana.

Em um terceiro e último momento, essas informações sobre os poluentes podem ser mais detalhadas e apresentar dados como o poluente primário do local (em alguns casos), as taxas mínimas e máximas de cada poluente, durante determinado período de tempo, o valor calculado do Índice de Qualidade do Ar de cada poluente presente, assim como informações adicionais referentes ao tempo, como a temperatura, pressão atmosférica, humidade, vento, etc, do local em análise.

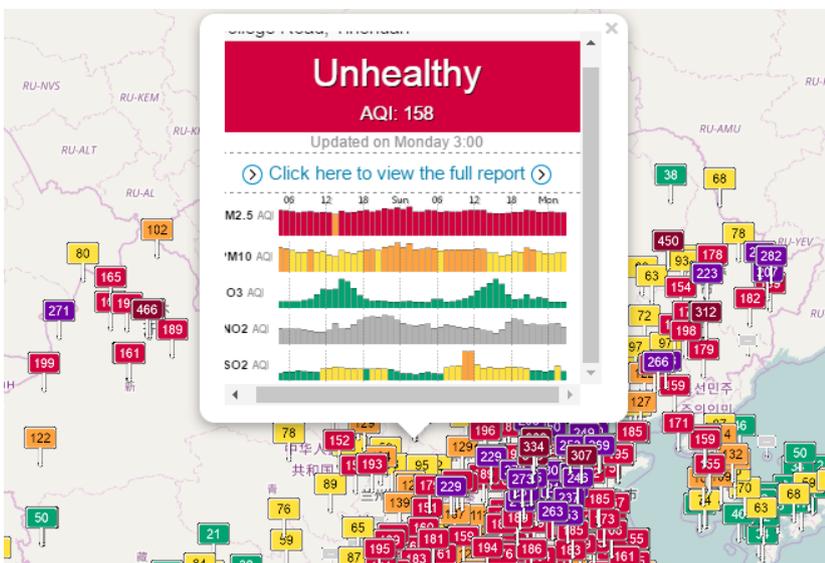


Figura 33 - Screenshot mouse click (2º momento).

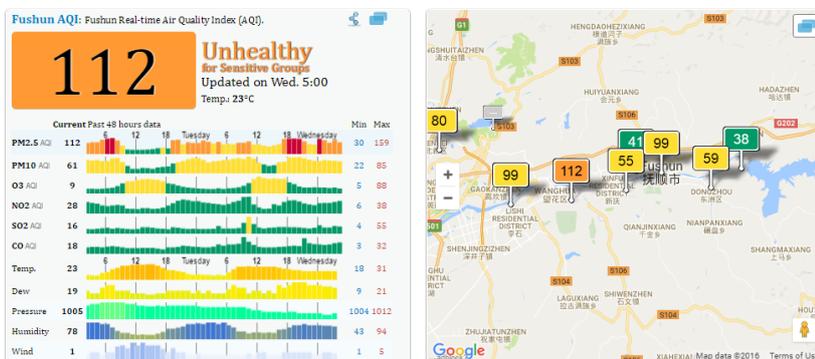
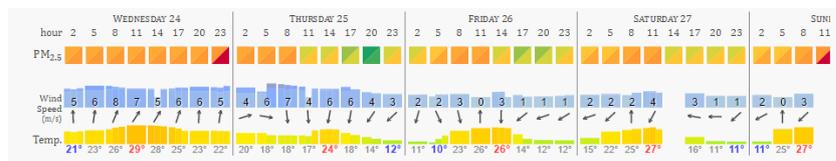


Figura 34 - Screenshot da vista completa da informação (3º momento).

Outras informações adicionais, encontradas no projeto *World Air Quality Index*, são as previsões para a qualidade do ar (*Air Quality Forecast*) para os dias seguintes ao dia em análise. Nesta representação gráfica, são apresentadas as datas das previsões, juntamente com a hora específica do dia, em que as previsões são feitas, o índice do poluente PM2.5 presente no local, referenciado com a numeração e a cor referente a periculosidade, a velocidade do vento (em metros por segundo) das horas analisadas, junto com sua direção, e por último a temperatura.

Todas essas informações descritas podem ser visualizadas num mapa dinâmico de previsões (*Forecast Map*), onde os poluentes são representados com cores específicas de cada nível de poluição e movimentam-se no mapa de maneira orgânica, como uma mancha de cores, consoante a mudança de intensidade dos poluentes ao longo das horas e dos dias em análise. Para além disto, a animação no mapa pode ser filtrada quando o utilizador escolhe qual poluente analisar, ou seja, caso o usuário tenha interesse em analisar um poluente específico, o mapa apresenta a opção de filtro para o poluente escolhido e o representa no mapa de modo isolado, o que simplifica e facilita a leitura da informação.

Figura 35 - Screenshot das informações adicionais.



3 Trabalhos relacionados

Com o objetivo de compreender os diferentes meios de visualizar grandes quantidades de informações, estes capítulos propõem analisar diferentes técnicas desenvolvidas em outros projetos que utilizam a visualização de dados como ferramenta escolhida para sintetizar, interpretar e apresentar as informações de modo intuitivo e dinâmico.

Neles, as informações são representadas através de dados geográficos, que permitem representar os dados em pontos específicos dos mapas e analisar as suas mudanças ou evoluções ao longo do tempo, que pode ser tanto dias, quanto períodos mais longos, como meses ou anos.

Os projetos analisados, são implementados com uma interface que permite a representação das informações de maneira resumida, além de permitir ao utilizador a possibilidade de interação e exploração dos dados de forma simples e acessível.

3.1 Projecto Ukko

O Projeto *Ukko* consiste em um novo serviço climático inovador que visualiza previsões sazonais do vento para apoio a decisões voltadas ao setor de energia. Nomeado em homenagem ao Deus finlandês do céu, do tempo e do trovão, o Projeto *Ukko* cria visualizações e uso de interface para previsões probabilísticas avançadas. Desenvolvido pela IC3 com o apoio de *Minero España*, como parte do *Resilience Prototype* no projeto EUROPIAS, o projeto abre novos caminhos no esforço para melhoria a resiliência da sociedade à variabilidade e mudança climática.

O projeto utiliza a visualização de informação para comunicar previsões climáticas que auxiliam a compreensão de condições futuras do vento e facilitam o uso de energia eólica limpa. Neste caso, os dados projetados não são voltados para os próximos dias, mas sim para meses e estações.

Durante o período de dois anos, a empresa de arte e design *FutureEverything* trabalhou em conjunto com cientistas climáticos do *Barcelona Supercomputing Center* (BSC) e *Met Office*, juntamente com o principal projeto europeu de apoio à adoção de serviços de previsão climática, o *EUPORIAS*, para o desenvolvimento deste projeto. Considerado o primeiro do seu tipo, o Projecto *Ukko*, criado pelo designer alemão, e especialista em visualização de informações, Moritz Stefaner, com o apoio de Dominikus Baur e Stefanie Posavec, que trabalhou junto com artistas, designers e cientistas climáticos para desenvolver novas possibilidades de representações a serviço do clima e demonstrar o potencial de previsões climáticas avançadas com apelo visual, interatividade e acessibilidade.

Com as diversas variações climáticas, ao longo do tempo, a sociedade desenvolveu novas fontes de energias renováveis. Entretanto, por serem sujeitas a constantes variáveis, tornaram-se



Figura 36 - Screenshot da visualização geral do Projeto Ukko.

menos confiáveis do que as fontes de energia convencionais. Por este motivo, é de essencial importância manter constante o monitoramento de previsões do vento para que a gestão da rede de energia alcance um equilíbrio e estabilidade entre a demanda e a produção. Neste sentido, o projeto foi desenvolvido e direcionado ao setor de energia eólica, o qual procura oferecer apoio aos operadores e comerciantes desta energia na tomada de decisões quanto ao planeamento, manutenção e novos investimentos de parques eólicos. Neste setor, a compreensão das condições futuras do vento é considerada crucial, por este motivo a interface criada procura detetar e informar com simplicidade padrões e movimentações do vento ao redor do mundo. Através da visualização criada o utilizador consegue explorar futuras fontes de energia eólica e localizar regiões suscetíveis a altas e baixas velocidades de vento e flutuações substanciais.

Segundo o laboratório de cultura digital *Future Everything*, a ferramenta também apresenta finalidade de auxiliar pessoas e indústrias vulneráveis a alterações provocadas pelo clima e pelo

tempo, proporcionando soluções e planeamento à condições futuras com meses ou até anos de antecedência. Segundo os mesmos, o ambiente natural como um todo, bem como a saúde humana, infraestrutura, e acesso a água potável estão todos a mercê das condições climáticas. Neste sentido, o projeto, além de apresentar funções informativas a respeito do clima e do tempo, posiciona-se como meio de gestão inteligente e cuidadoso dos recursos naturais. Garantindo resiliência a eventos climáticos extremos e a prevenção aos riscos ligados as mudanças climáticas, economia e segurança alimentar para todo o mundo.^[53]

^[53] Future Everything. Project Ukko. Disponível em: <http://futureeverything.org/projects/ukko/>. Acedido em: 17 de outubro de 2017

A visualização gráfica dos dados e o design de interface, na aplicação, permite ao utilizador uma navegação interativa para maior e melhor exploração dos dados. No mapa, onde as informações do vento são sobrepostas, o utilizador observa diversas linhas, com diferentes espessuras e diferente variações de cores. Cada característica, representada em cada linha, possui uma função de informação específica.



Figura 37 - Screenshot da legenda do projeto Ukko

As variações de espessuras das linhas representam a força do vento prevista, as quais variam entre cinco intensidades de vento diferentes. Dentre elas, a mais grossa representa uma força de vento de intensidade elevada e varia até a mais fina, que representa uma força de vento de baixa intensidade. A opacidade das linhas representa a habilidade do vento, que varia entre quatro

tonalidades, das quais as linhas mais brilhantes e visuais representam que a ferramenta tem sido muito útil, na região assinalada, no passado, e as mais escuras representam o oposto. Nesta representação, não é utilizada nenhuma linha caso nenhuma previsão útil possa ser feita. Por último, e considerado o mais usual no uso de informações sobre a velocidade do vento, está codificado na cor e rotação da linha. As linhas amarelas com inclinação para cima indicam um aumento esperado no vento na região assinalada, enquanto as linhas azuis com inclinação para baixo indicam ventos mais baixos.

Na representação gráfica, também é possível visualizar a quantidade de energia eólica instalada, conseqüentemente o utilizador pode visualizar pontos no mapa onde as previsões de mudanças na força do vento são substanciais e onde são as localizações dos parques eólicos no mundo. Ao selecionar uma região, o utilizador também consegue visualizar informações mais detalhadas, como previsões passadas e futuras e a velocidade média sazonal do vento em metros por segundo da região em análise.



Com alternativa a visualização panfônica no mapa, os dados da previsão do vento, utilizados no Projeto *Ukko*. São representados em uma visualização adicional que abrange 51 conjuntos de dados. Esta visualização, denominada como *Ambient Installation*, exibe mais de 100.000 dados usando uma projeção de mapa personalizado (equidistante azimutal) representados em todo o mundo. Essa visualização animada percorre planeta, roda e amplia em algumas áreas específicas, e permite ao utilizador uma outra visualização, diferenciada dos dados eólicos no mundo.

Figura 38- Screenshot da visualização num ponto específico no mapa

3.2 Iso-edges for the Geovisualization of Consumptions

O projeto de visualização de dados *Iso-edges for the Geovisualization of Consumptions*, desenvolvido pelo Departamento de Engenharia Informática, CISUC, da Universidade de Coimbra, em parceria com a empresa de retalho portuguesa SONAE, promove a visualização de grandes quantidades de dados, sobre a evolução dos consumos de 729 hipermercados e supermercados, da empresa referida, por um período de dois anos.

Esta visualização propõe analisar e compreender uma vasta quantidade de informações, obtidas ao longo desses dois anos, para que as mesmas possam ser sintetizadas e reproduzidas graficamente de maneira compreensiva e objetiva. A representação dos dados, realizada geograficamente, permite recompilar as informações de modo a facilitar a compreensão sobre as mudanças nos padrões de consumo, que ocorrem nas cadeias de hipermercados e supermercados da empresa, tanto nas variações sazonais, assim como em épocas festivas do ano, quanto na gestão de stocks e decisões administrativas de abertura ou renovações de novas, possíveis, instalações.

Para a realização deste projeto, os realizadores tiveram acesso a um conjunto de dados que abrange informações de todas as transações realizadas nos supermercados e hipermercados no período de maio de 2012 a abril de 2014, bem como informações dos códigos postais de todos os clientes das cadeias da empresa, o que possibilitou uma estimativa quanto a posição geográfica dos clientes e das principais edificações de Portugal.



Figura 39 - Visualização do consumo em dezembro 2012 (à esquerda) e janeiro 2013 (à direita).

Segundo o artigo,^[54] escrito pelos realizadores do projeto, em uma primeira abordagem foram utilizadas curvas de nível, ou *isolines*, que marcaram as áreas com valores de consumo similar e permitiu delimitar e enfatizar as áreas de alto e baixo consumo e representar as suas mudanças geográficas. Como havia pouca densidade de *isolines* a legibilidade da visualização dos dados tornou-se prejudicada, o que resultou na utilização de uma técnica complementar, designada como *isoedges*, definida como *isolines* abertas, que auxiliou no aumento da densidade e, conseqüentemente, na melhoria da visualização. Neste projeto, os dados de consumo, representados através do uso das *isolines*, foram representados no mapa de Portugal juntamente com a projeção dos principais edifícios e a localidade dos clientes da cadeia de hipermercados e supermercados da empresa. Com isso, os realizadores do projeto, conseguiram criar uma visualização com as

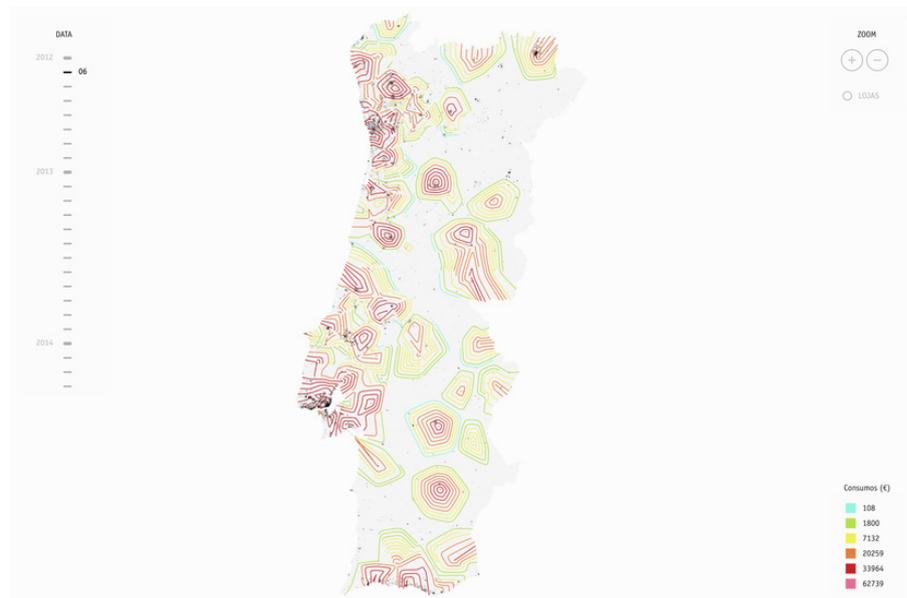
^[54] Maças, C., Polisciuc, E., Machado, P., et al. (2016). Iso-edges for the Geovisualization of Consumptions. Departamento de Engenharia. CISUC, Universidade de Coimbra, Coimbra, Portugal.

principais áreas de conglomerados de edifícios e residências. Estas, representadas no mapa como pequenos círculos pretos, de tamanhos iguais para não sobrecarregarem a visualização, fornecem aos utilizadores uma compreensão visual simplificada da informação. Visto que o projeto tencionava analisar a intensidade de consumo de diferentes meses do ano, os dados representados foram calculados a partir do consumo médio de cada mês. Cada linha, utilizada para representar as informações no mapa, equivale a um valor de consumo e separam as áreas com alto e baixo consumo. A técnica utilizada para gerar as *isolines* é proveniente de uma triangulação entre os centros geométricos de cada distrito português, que utiliza o algoritmo da triangulação de *Delaunay*.

^[55] Maças, C., Polisciuc, E., Machado, P. et al. (2016). Iso-edges for the Geovisualization of Consumptions. Departamento de Engenharia. CISUC, Universidade de Coimbra, Coimbra, Portugal.

^[55] Para conseguir uma distância similar entre as linhas e favorecer uma melhor legibilidade da informação, foi definido uma distância mínima entre as linhas e representada apenas as linhas que respeitassem essa distância. Visto que as linhas resultantes eram demasiado afastadas, utilizou-se a técnica de *isoedges* para complementar e enfatizar a informação, principalmente em zonas em que os distritos são mais próximos uns dos outros.

Figura 40 - Screenshot da interface gráfica Iso-edges for the Geovisualization of Consumptions



A interface, criada para o projeto, propõe uma interação intuitiva e simplificada na qual permite que o utilizador consiga explorar as informações, representadas no mapa, com maior acessibilidade. Nela, o utilizador consegue ampliar a visualização através de um *zoom*, o que proporciona um nível mais detalhado de informação, escolher entre os diferentes meses, dos quais deseja obter as informações, apenas ao clicar no mês pretendido e visualizar a localização das lojas nas quais ocorreram o consumo, ao clicar em “lojas” no canto superior direito. Ademais, uma legenda com as cores utilizadas na visualização, que identifica os valores de consumo, é posicionada junto ao mapa para facilitar a compreensão das informações representadas.

Com o desenvolvimento deste projeto, pode-se compreender o nível de consumo e as suas diferenças dentre os municípios do país, bem como analisar as áreas em que os hipermercados ou supermercados da SONAE não se fazem presentes. Foi possível, também, verificar áreas promissoras para a abertura de novas lojas, visto que apresentam um baixo consumo, embora tenha uma grande população. Através da visualização desenvolvida, surgiu uma nova fonte de apoio a tomada de decisão administrativas, por parte da empresa, bem como nova ferramenta simplificada e acessível para compreensão e síntese de grandes e complexas informações sobre o consumo dos 729 hipermercados e supermercados da empresa, em Portugal.

3.3 Conclusão

Em suma, depois de analisar os casos de estudo é de realçar a importância da escolha das variáveis visuais e da técnica de representação em função dos dados a utilizar, para se conseguir transmitir a informação de uma forma clara, rápida e intuitiva.

Nenhuma técnica de representação é perfeita, conseqüentemente novas formas de representação iram surgir para melhorar métodos antigos e desatualizados, acompanhada pela evolução tecnológica que desempenha um papel importante no desenvolvimento de novas formas de representação e extração de informação.

As técnicas e gráficos descritos neste capítulo serviram com fonte de inspiração no desenvolvimento de visualizações durante a pesquisa. Com este capítulo pretende-se analisar erros na representação e na comunicação de outras visualizações para não os repetir. O objetivo é melhorar a representação gráfica dos dados da poluição e do vento, para encontrar padrões na informação que nos levem à descoberta de novo conhecimento.

4 Desenvolvimento da aplicação

Neste capítulo será discutido todo o processo de visualização dos dados do vento e da poluição. Começando com a descrição dos dados e das ferramentas utilizadas. Posteriormente, são apresentadas as primeiras abordagens para visualizar os dados, descritas as dificuldades encontradas no seu desenvolvimento e como as conclusões tiradas influenciaram as representações seguintes. Os nossos modelos de representação serão então discutidos em detalhe, terminando com a apresentação das técnicas de interação utilizadas para explorar os dados.

Sendo assim, este capítulo será dividido em três partes: a primeira incidirá sobre análise e representação os dados da poluição; a segunda parte sobre os dados do vento; e por fim a junção das duas representações numa.

4.1 Visualização da poluição

Dados e Ferramentas

Os dados utilizados para representar a poluição são fornecidos pelo Centro de Análise Geográfica da Universidade de Harvard através da hiperligação (<http://aqi.cga.harvard.edu/china/>). Os ficheiros são disponibilizados desde o dia 1 e janeiro de 2015, três vezes por dia (1h, 9h, 17h), em formato *Tape Archive* (.tar).

Em primeiro lugar, foi escolhido um *sample* de dados de uma semana (do dia 10 de janeiro de 2016 até 16 de janeiro de 2016). Em seguida, foram baixados os arquivos com os dados para uma pasta de trabalho e descomprimidos para formato *Comma-separated values* (.csv).

Posteriormente, foi analisada a informação necessária para representar os dados. O conjunto de dados é composto pelo nome das estações (inglês e chinês) onde são feitas as medições, pelo id da localização das estações, pela localização geográfica (latitude e longitude) das estações, pelos diferentes gases presentes no ar como o Dióxido de enxofre pphm (partes por 100 milhões), o Dióxido de nitrogénio pphm (partes por 100 milhões), o Ozono pphm (partes por 100 milhões), o Monóxido de carbono ppm partes por milhão), as Partículas Suspensas 2.5 µg/m³ (microgramas por metro cúbico), as Partículas Suspensas 10 µg/m³ (microgramas por metro cúbico), a temperatura (Celsius), o ponto de condensação da água(celsius), a pressão (milibares), a humidade (gramas por metro cúbico), a velocidade do vento (quilómetros por hora) e a data de recolha da informação.

Como referido anteriormente, devido ao elevado número de dados, foi utilizada uma amostra de semana e extraídas apenas as informações necessárias na representação da poluição como

a localização, os diferentes gases, a data e a hora. A informação relativa à velocidade e à direção do vento que podiam ser utilizadas na representação do vento, secção 4.2, não foi usada porque os dados são inconsistentes, ou seja, faltam-lhes valores. As restantes informações porque não influenciavam na representação da poluição.

Para isso foi utilizado o *Processing*, uma linguagem de programação derivada do Java, como ferramenta de apoio no desenvolvimento da visualização, o que nos permitiu carregar os ficheiros com os dados, filtrar e tratar a informação.

Primeiras abordagens

Numa primeira, abordagem aos dados da poluição, foram calculados os gases poluentes mais elevados em cada localização, ou seja, foram percorridos todos os valores referentes aos seis gases (Monóxido de Carbono, Dióxido de Enxofre, Dióxido de Nitrogénio, Ozono, Partículas Suspensas 2.5 e Partículas Suspensas 10) e representados os que tinham a maior taxa. A ideia era descobrir qual era o maior poluente em cada área geográfica.

Desta forma, a técnica utilizada para representar os dados da poluição foi um mapa de pontos, devido à sua natureza geográfica. Em primeiro lugar, foi necessário representar o mapa da China e mapear a longitude e a latitude dos dados para posições (x e y) no ecrã do computador, utilizando a biblioteca *Unfolding Map*.

Cada localização é representada por um círculo no mapa e a cor foi utilizada como variável visual, onde o verde representava o Dióxido de Enxofre, o amarelo as Partículas Suspensas 2.5, o roxo o Ozono, o vermelho o Monóxido de Carbono, o laranja o Dióxido de Nitrogénio e o vermelho escuro as Partículas Suspensas 10.

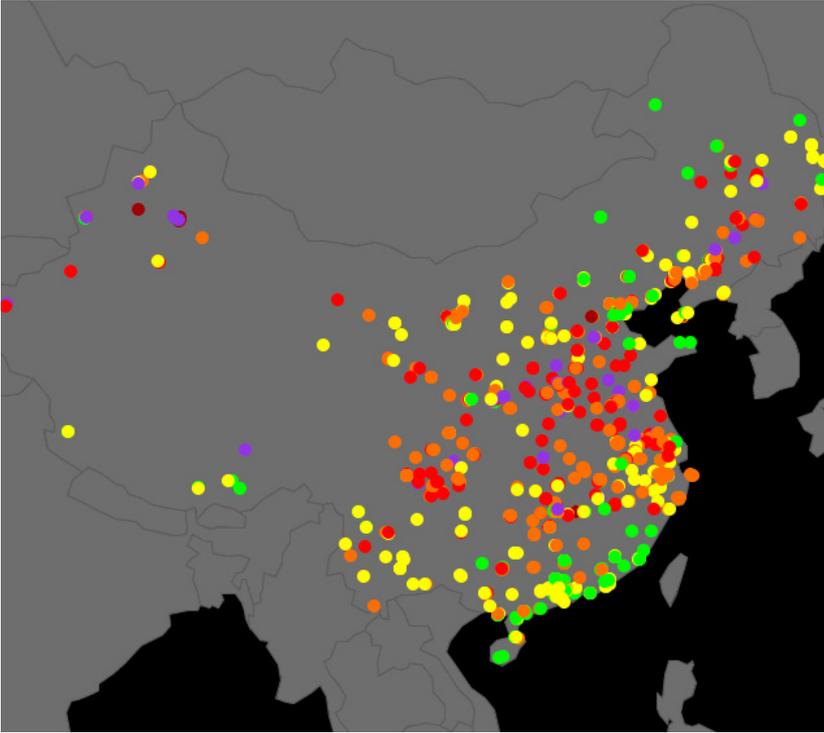


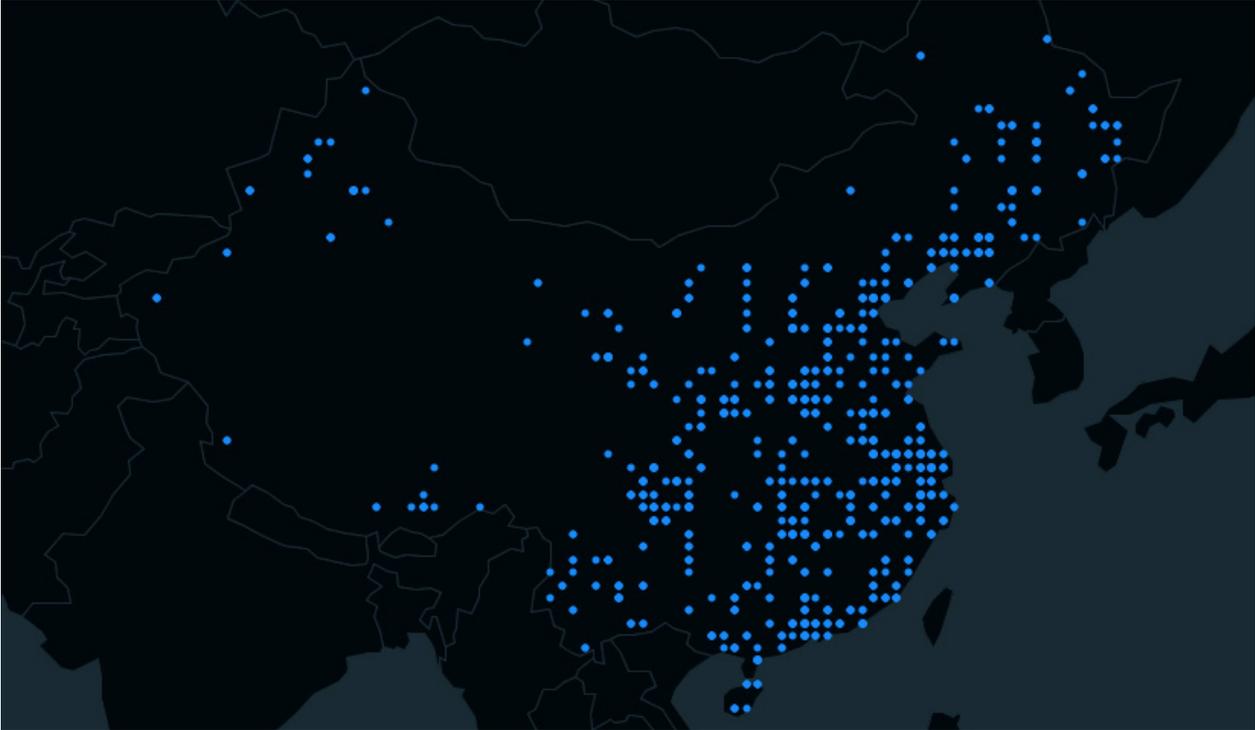
Figura 41 - Screenshot da visualização dos gases mais elevados.



Figura 42 - Screenshot da grelha aplicada ao mapa da China.

Posteriormente à representação dos gases poluentes, reparou-se que haviam dados sobrepostos no mapa, o que dificultava a leitura da informação pelo utilizador e conseqüentemente escondia os dados. Para resolver este problema e reduzir o ruído visual resolveu-se aplicar uma grelha quadrangular de (5 pixeis por 5 pixeis). A grelha devia garantir que os pontos ficavam equidistantes uns dos outros e assim, organizar a representação, tornando-a mais clara. Por outro lado, a grelha acrescentou um erro gráfico à visualização fez com que se perdesse a localização verdadeira dos dados. Isto é, foi calculada a distância entre a posição dos dados e as várias posições na grelha e escolhida a mais próxima para trocar.

Cada ponto no mapa, da imagem da figura 42, representa o centro de uma área geográfica criada pela grelha. A paleta de cores também foi trocada, inicialmente pensamos em aplicar a cor preta aos oceanos e o branco aos países, mas achamos que ficava melhor uma paleta com tons da mesma cor (Azul). Assim sendo, foi equacionada a opção de durante o dia utilizar tons mais claros e a noite tons mais escuros para indicar a mudança de horas.



Como podemos ver na imagem da figura 43, depois de se ter aplicado a grelha, o *layout* dos pontos ficou mais organizado e a informação é transmitida de uma forma clara e limpa. Contudo, existem dados que ficaram com a mesma posição, ou seja, sobrepostos.

Figura 43 - Screenshot da grelha aplicada aos dados.

Para resolver este problema foi necessário fazer uma contagem dos dados que tinham a mesma posição e ver qual tinha o índice de poluição mais elevado de todos. Esse valor foi mapeado no raio do círculo e posteriormente atribuído à área geográfica em que está inserido na grelha.

A cor que até então era a variável visual utilizada para representar o gás poluente, foi substituído pelo tamanho do círculo e representado o nível de poluição. Desta forma, os círculos com o raio maior representam os gases poluentes com os níveis mais elevados e os menores os níveis mais baixos.

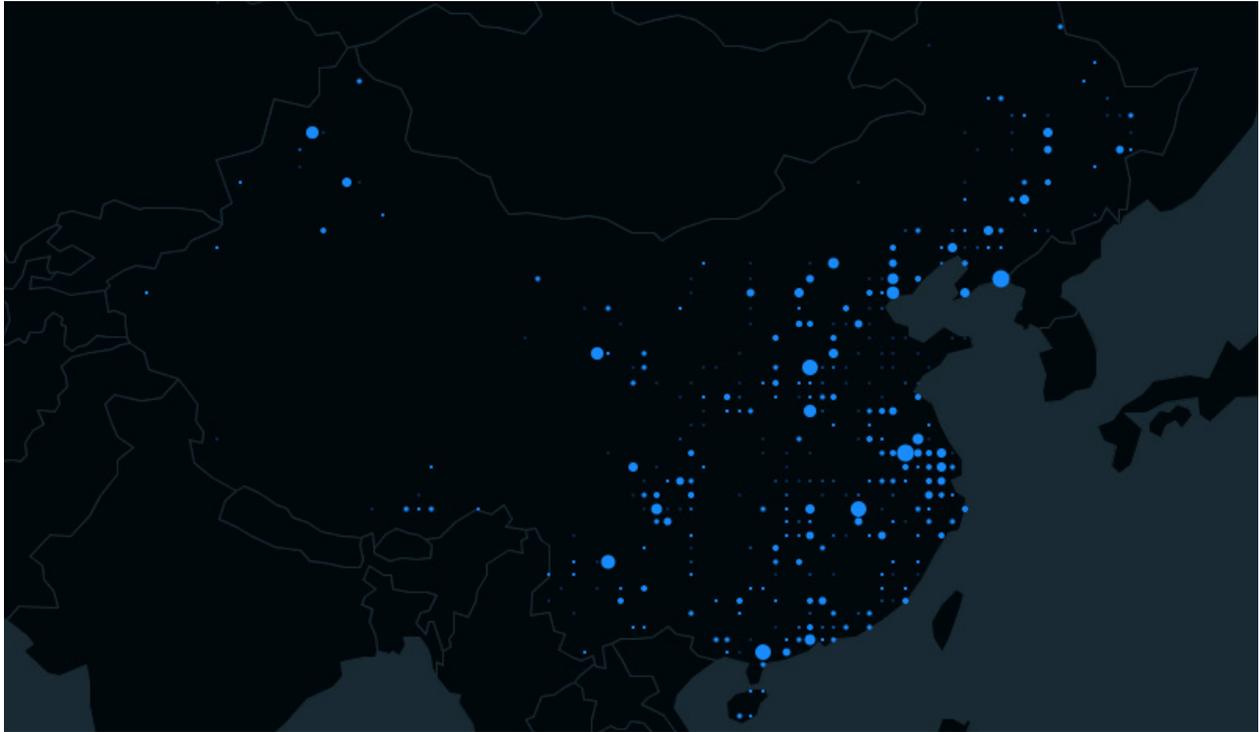


Figura 44 - Screenshot à visualização dos dados da poluição mapeados pelo variável tamanho.

Em suma, esta primeira abordagem aos dados da poluição tem falhas em vários aspetos fundamentais. Em primeiro lugar, não se consegue ter a perceção do tipo de gás está a ser visualizado, só permite ver o nível de poluição que dado pelo tamanho do raio do círculo. Em segundo lugar, existem mais dados que foram omitidos por ter a mesma posição geográfica, ou seja, só é representado o que tem o valor mais elevado. Em terceiro lugar, devido ao uso da grelha, a posição dos dados foi trocada podendo levar o utilizador a tirar conclusões erradas.

Assim, foi procurada outra forma mais eficiente de representar os dados da poluição que permita representar os diferentes gases independentemente do nível de poluição e da sua localização geográfica.

Abordagem final

Esta última fase de representação da poluição resulta da combinação de várias técnicas e tem como finalidade a criação de um mapa isométrico, composto por linhas *isohypses* ou *contour lines*, linhas de igual altura do solo.

Para tal, num primeiro momento foi utilizada a formula matemática conhecida como *distance decay*. Um termo geográfico que descreve o efeito da distância em interações espaciais, mais precisamente entre duas localizações separadas. É um conceito que liga a geografia e a ocorrência ou frequência de um padrão de atividade. A *distance decay* pode ser expressa como uma função exponencial ou potencial que diminui à medida que a distância aumenta. Como é ilustrado na imagem da figura 45.

$$P_i = \sum_{j=1}^n \frac{M_j}{d_{ij}^c}$$

M_j = valor da poluição j ; d_{ij} = distância entre pixel (i) e localização dado (j) ; P_i = número potencial i (cor mapeada no pixel); c = expoente da função potencial (distribuição)

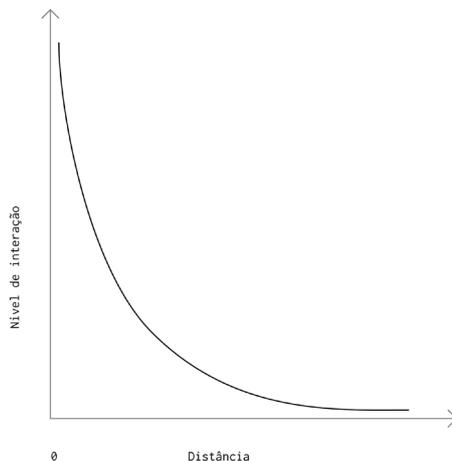


Figura 45 - Imagem ilustrativa do algoritmo do *distance decay*.

Assim, são percorridos todos os pixels dentro do espaço do mapa e para cada um são calculadas as distâncias a todos os pontos de poluição. Sempre que é calculada uma distância nova também é feito o somatório da divisão do valor da poluição pela distância.

No final de calcular todas as distâncias o valor do somatório é mapeado na cor do pixel, mas para isso é necessário, antes de tudo, calcular o valor máximo do somatório. As cores dos pixels são mapeadas numa escala de cinza onde o preto representa as maiores distâncias (níveis baixos de poluição) e o branco as menores distâncias (níveis altos de poluição) aos pontos. Todos os pontos de poluição têm influência, no espaço do mapa, sobre a cor mapeada nos pixels.

FUNÇÃO DISTANCE DECAY: PSEUDOCÓDIGO

1 **Percorre** todas as posições x no ecrã

2 **Percorre** todas as posições y no ecrã

2.1 **Se** estiver dentro do mapa então

2.2 **Percorre** todos dos dados da poluição

2.2.1 Calcula a **distância** entre a posição pixel e o ponto de poluição

2.2.2 Faz o **somatório** da divisão do valor da poluição pela distância

Fim do ciclo da poluição

Fim do se está dentro do mapa

Fim do ciclo dos pixels

Desta forma, são criados mapas de altura (*heights maps*), imagens bidimensionais utilizadas para armazenar valores relativos à elevação de uma superfície, neste caso os valores da poluição. Quanto maior for o valor da poluição mais alta é a “montanha” criada pela fórmula, que utiliza uma distribuição *gaussiana* (normal). A abertura da distribuição pode ser manipulada pelo valor c da função, o que torna a “montanha” mais ou menos íngreme, ou seja, faz com que os pixels branco se concentrem em torno do ponto ou se distribuam.

Esta técnica é a primeira etapa na criação de mapas isométricos, como tal, foram geradas imagens para cada tipo de gás (Monóxido de Carbono, Ozono, Dióxido de Nitrogénio, Dióxido de Enxofre, Partículas Suspensas 2.5, Partículas Suspensas 10), três vezes por dia (1h, 9h, 17h) durante uma semana (10 de janeiro de 2016 a 16 de janeiro de 2016) e guardadas em pastas diferentes. Foi necessário criar um programa unicamente para o efeito, visto que, se as imagens fossem geradas em tempo real sobrecarregava a aplicação e demoraria a iniciar.

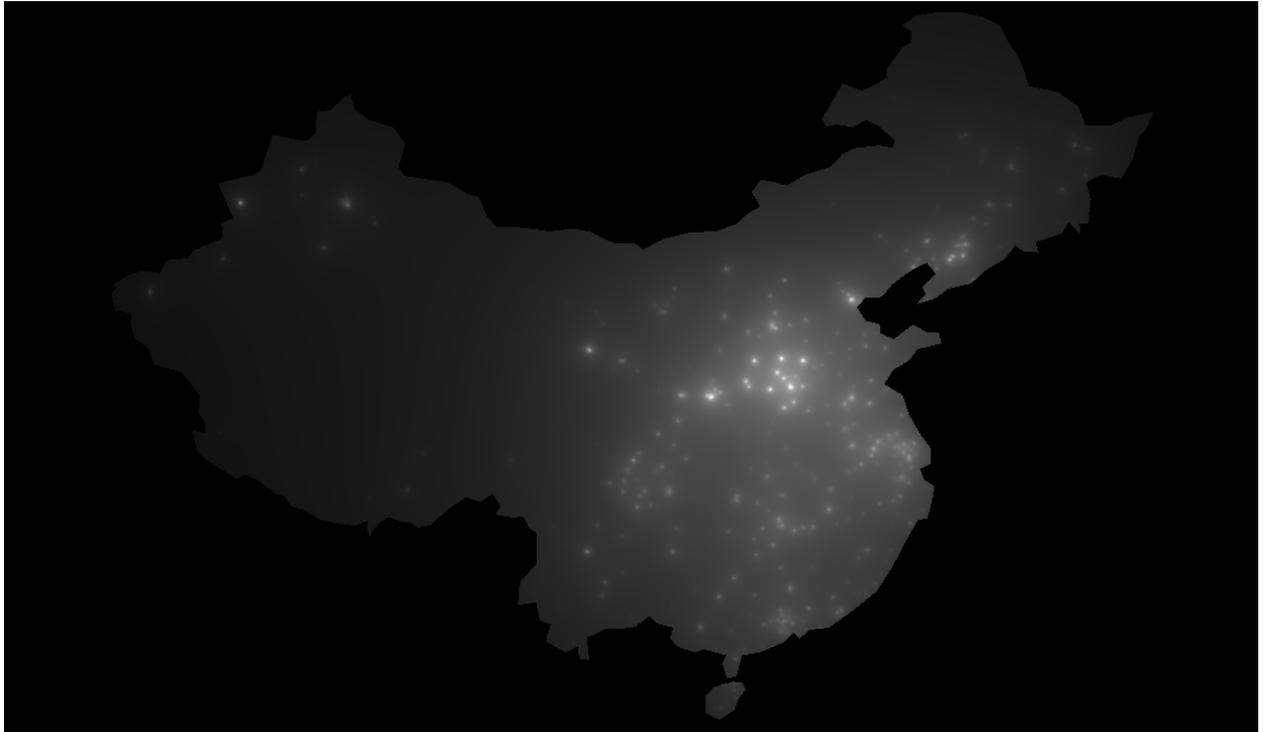


Figura 46 - Height map gerado com os dados do CO (Monóxido de Carbono).

Como podemos visualizar, na figura da imagem 46, as manchas brancas no mapa representam os picos de poluição onde a concentração de gases poluentes é maior. Apesar das imagens serem estáticas e não termos a variável tempo, associada, conseguimos com alguma facilidade identificar áreas de elevada poluição. É evidente que na parte oriental da China existe uma maior concentração de gases independentemente do seu tipo.

O próximo passo na criação do mapa isométrico é a geração das *contour lines* a partir das imagens produzidas anteriormente (*heights maps*). Para isso foi utilizada a biblioteca *BlobDetection*,^[56] desenvolvida especificamente para o *Processing*. Esta biblioteca destina-se a encontrar “*blobs*” em uma imagem, isto é, áreas cujo brilho está acima ou abaixo de um específico valor. O objetivo da deteção de *blobs* é identificar e marcar essas regiões.

^[56] BlobDetection [online], acedido em 7 de Novembro de 2016, disponível em <http://www.v3ga.net/processing/BlobDetection>

O algoritmo utilizado, por detrás da biblioteca, para a deteção das *blobs* é baseado no *threshold* (limiar), ou seja, no brilho que queremos que seja detetado. Assim sendo, foi atribuído um *threshold* máximo e um mínimo, que posteriormente foi dividido por níveis ou número de *contours*, neste caso foram utilizados 35 níveis. Por exemplo, se tivermos uma “montanha” ou uma superfície estatística e passarmos por ela vários planos paralelos com a mesma distância, cada plano representa um *threshold* diferente e a projeção dos cortes obtidos as áreas de igual valor que queremos detetar. Como ilustra a imagem da figura 47.

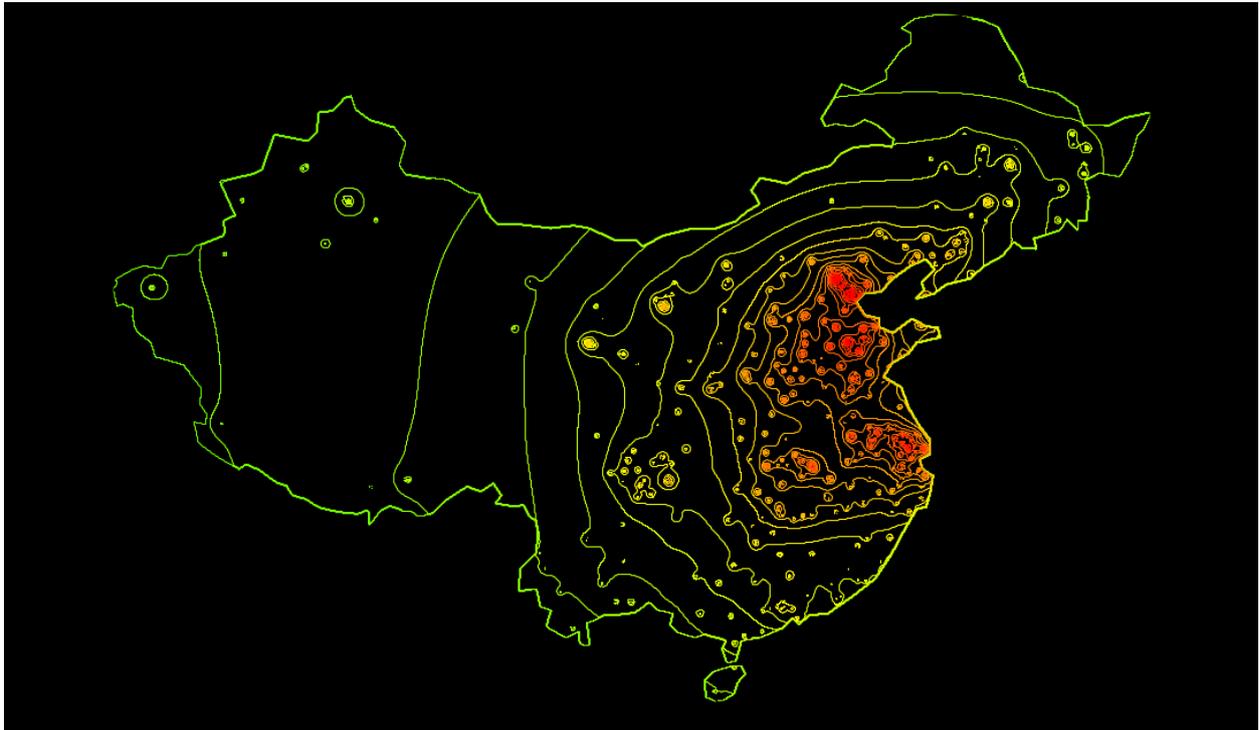


Figura 47 - Contour map gerado a partir da detecção das blobs dos dados da poluição (2D)

Como pode ser visualizado na imagem da figura 47, os níveis de poluição são representados através de linhas que utilizam a cor como variável visual. Assim, o verde corresponde aos níveis de baixa concentração e o vermelho aos níveis de elevada poluição. Na imagem tridimensional da figura 48 as cores foram invertidas para testar a percepção dos valores, chegando-se à conclusão que o vermelho é mais intuitivo, para o utilizador, na representação de valores altos do que baixos.

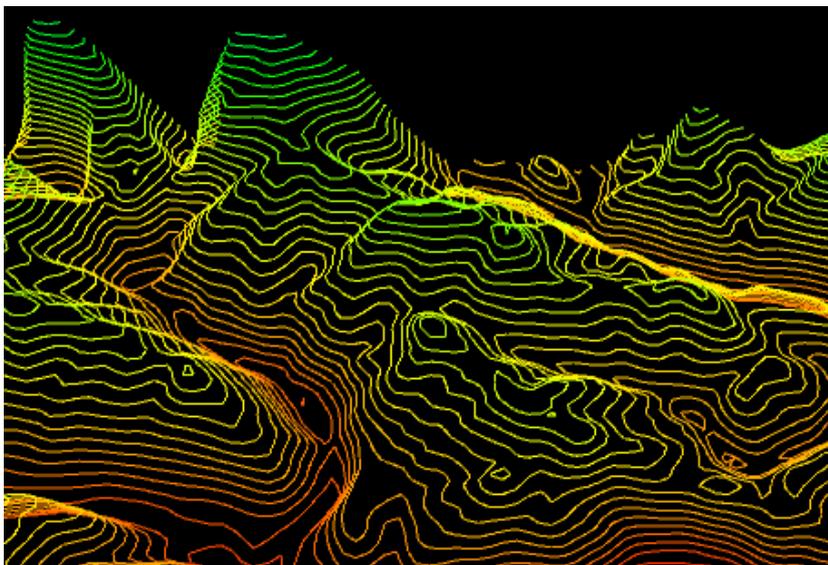


Figura 48 - Contour map gerado a partir da deteção das blobs dos dados da poluição (3D).

Para conseguir ter uma visão tridimensional do *contour map*, gerado, foi utilizada a biblioteca *Peasycam v202*,^[57] desenvolvida pelo engenheiro Jonathan Feinberg para o *Processing* e que fornece uma câmara *dead-simple mouse-driven*. Isto é, permite ao utilizador rodar a câmara em torno do objeto que está a ser visualizado e fazer *zoom in* ou *zoom out* utilizando o rato.

^[57] Peasycam v202 [online], acedido em 21 de Novembro de 2016, disponível em <http://mrfeinberg.com/peasycam>.

A terceira dimensão quando utilizada na visualização de informação dá uma visão errônea dos valores representados, isto por causa da perspectiva. Por exemplo, se compararmos as alturas das “montanhas” da figura 48 os valores podem parecer iguais, mas isso não acontece na realidade. Sendo assim, foi utilizada uma visão bidimensional em detrimento da terceira dimensão.

Depois de identificar as áreas com os mesmos níveis de poluição, delimitadas por linhas (*isolines*), foram geradas superfícies para facilitar a leitura dos dados e a identificação de padrões. Isto é, os pontos que compõem as linhas foram ordenados e extraídos para ficheiros .CSV pelo tipo de gás, dia e hora.

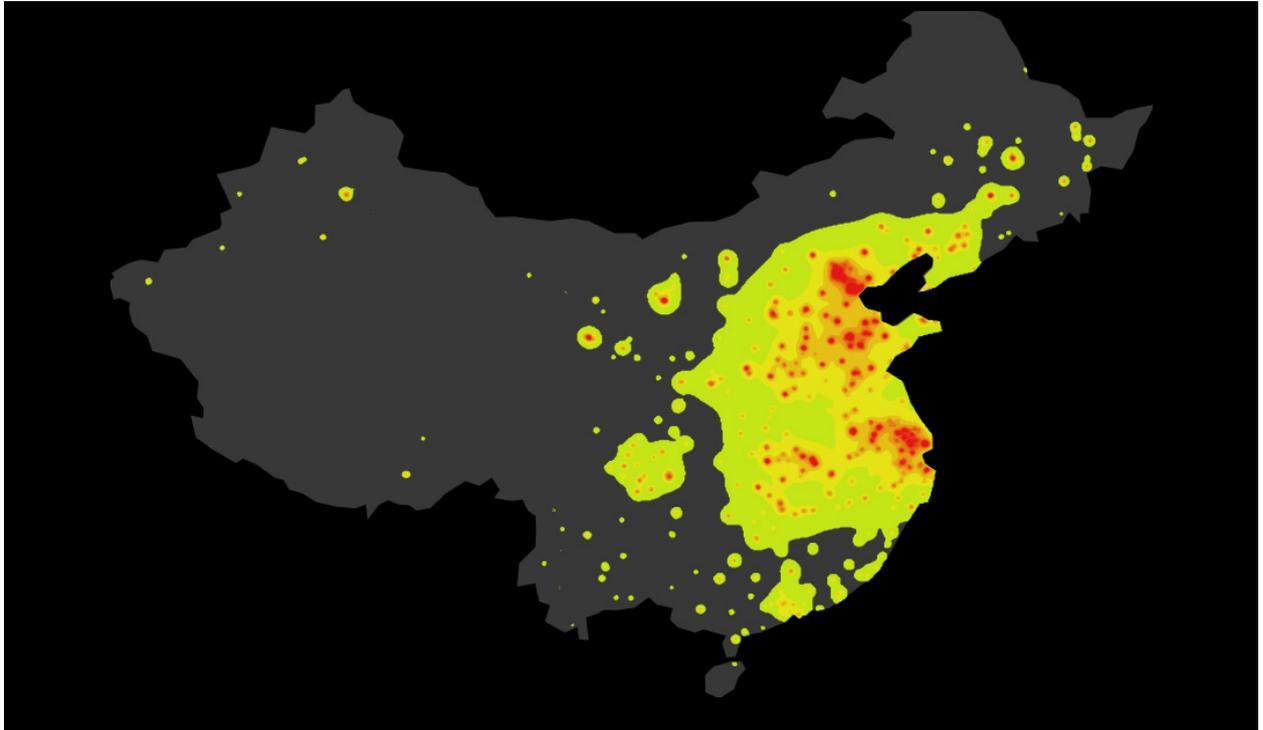


Figura 49 - Superfícies criadas a partir das isolines ou *contour map*.

Desta forma, houve a necessidade de criar um novo programa para evitar que a aplicação ficasse lenta. No qual, posteriormente foram carregados os dados dos ficheiros e desenhadas as *shapes* (formas) da poluição sobre o mapa.

As cores utilizadas para mapear o nível de poluição no *contour map* foram também usadas para mapear os valores nas superfícies geradas. Esta paleta de cores resultaria se apenas quiséssemos representar um tipo de gás de cada vez. Como é ilustrado pela imagem da figura 49.

^[58] Meirelles, I. (2013) Design for Information: An Introduction to the Histories, Theories, and Best Practices Behind Effective Information Visualizations. Rockport Publishers, Beverly, MA, pp 142

Devido ao facto de existirem seis gases diferentes para representar, decidiu-se atribuir uma cor a cada tipo de poluente para simplificar o *layout* e utilizar a transparência (50%) como variável visual. Visto que, a percepção humana consegue distinguir, com alguma facilidade, até um máximo de oito cores.^[58]

Assim, numa primeira fase foi atribuída a cor azul ao Dióxido de Enxofre, o lilás ao Ozono, o verde ao Dióxido de Nitrogénio, o laranja as Partículas Suspensas 2.5, o amarelo as Partículas Suspensas 10 e o vermelho ao Monóxido de Carbono.

Ao gerar as várias imagens para cada dia da semana foi identificado um problema com a biblioteca *BlobDetection*. Haviam pontos (vértices) que não eram computados, ou seja, estava em falta o que fazia com que as *shapes* (superfícies) não fechassem completamente criando “cortes” (falhas) e sobreposições que não existiam. Como é ilustrado pela imagem da figura 51.

^[59] Diewald_CV_kit [online], acedido em 23 de novembro de 2016, disponível em https://github.com/diwi/diewald_CV_kit.

Este problema fez com que fosse necessário procurar uma nova biblioteca que deteta-se “blobs”. Depois de vários testes foi escolhida a biblioteca *diewald_CV_kit*,^[59] criada por Thomas Diewald em 2011, especificamente para o *Processing* e que contém ferramentas utilizadas no campo da visão computacional. Além disso, foi essencial criar um programa novo que deteta-se as “blobs” e guarda-se os vértices das superfícies em ficheiros .CSV.

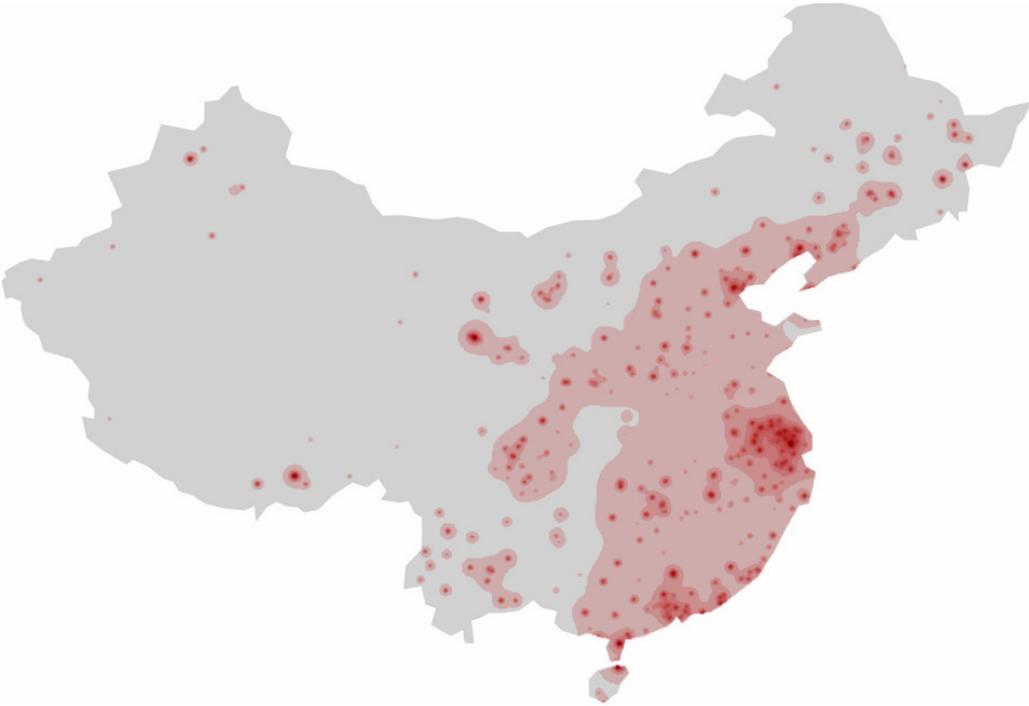


Figura 50 - Representação dos dados do Monóxido de Carbono.

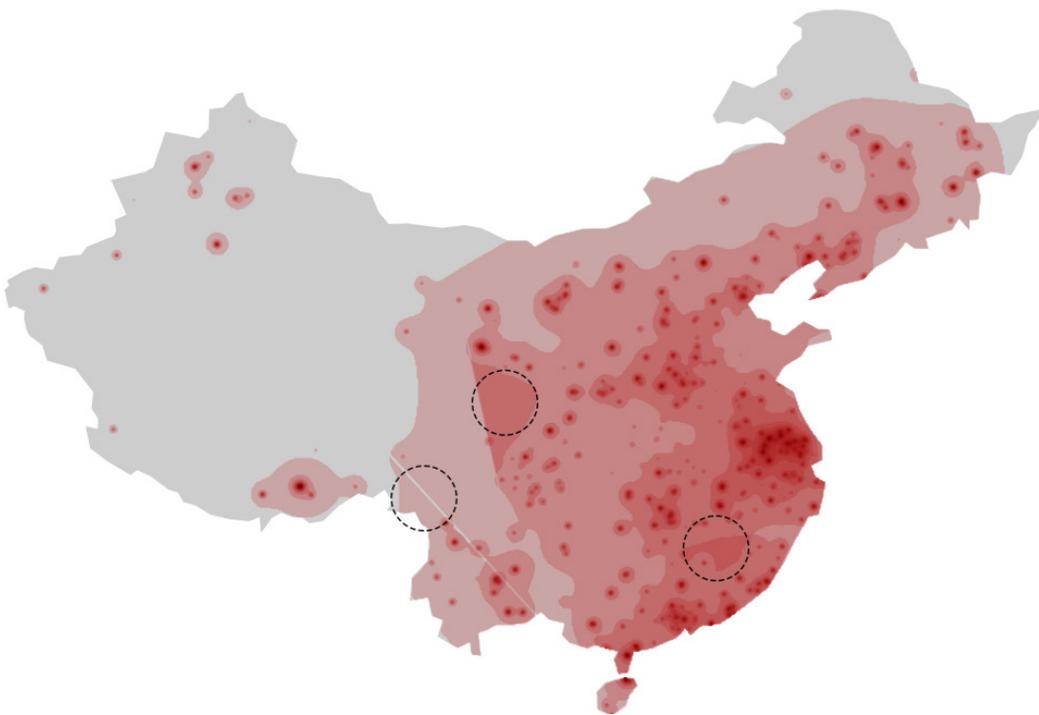
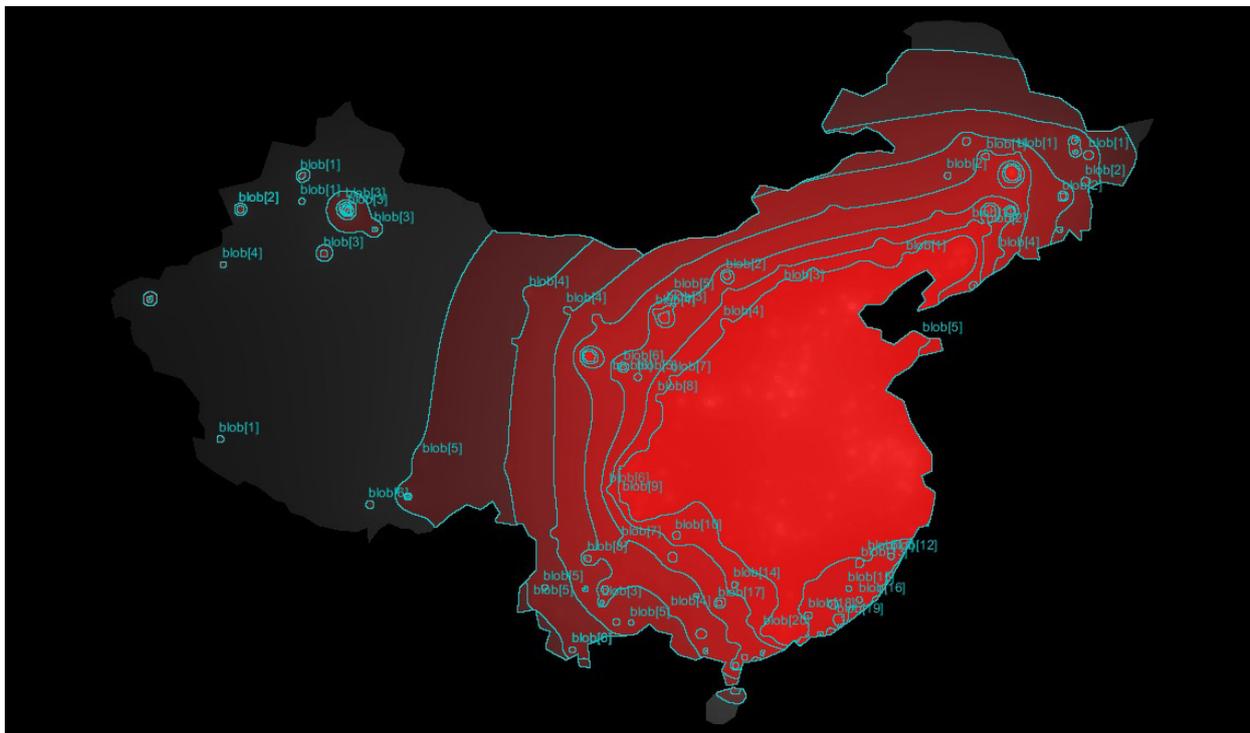


Figura 51 - Defeitos da biblioteca BlobDetection.



Nesta fase foi aproveitado para fazer algumas alterações na captação das “blobs”. O número de níveis foi reduzido, passou de 35 para 20. Tal como, o número mínimo de vértices que passou a ser de 20, ou seja, só são detetadas e guardadas as superfícies com um número de vértices superior a 20.

Figura 52 - Detecção das “blobs” com recurso à biblioteca diwald_CV_kit.

Por fim, depois de captadas as “blobs” para todos os dias, foi criado o primeiro artefacto visual que permite visualizar os diferentes gases presentes no ar e escolher que tipo poluente. A seleção do gás é feita através do menu “Air”, do lado direito, que contém os nomes dos poluentes. Ao iniciar a aplicação é carregada automaticamente a imagem em formato .PNG referente ao Dióxido de Nitrogénio e se clicarmos no nome de outro poluente a imagem troca, diminuindo a opacidade da primeira e aumentando a da segunda.

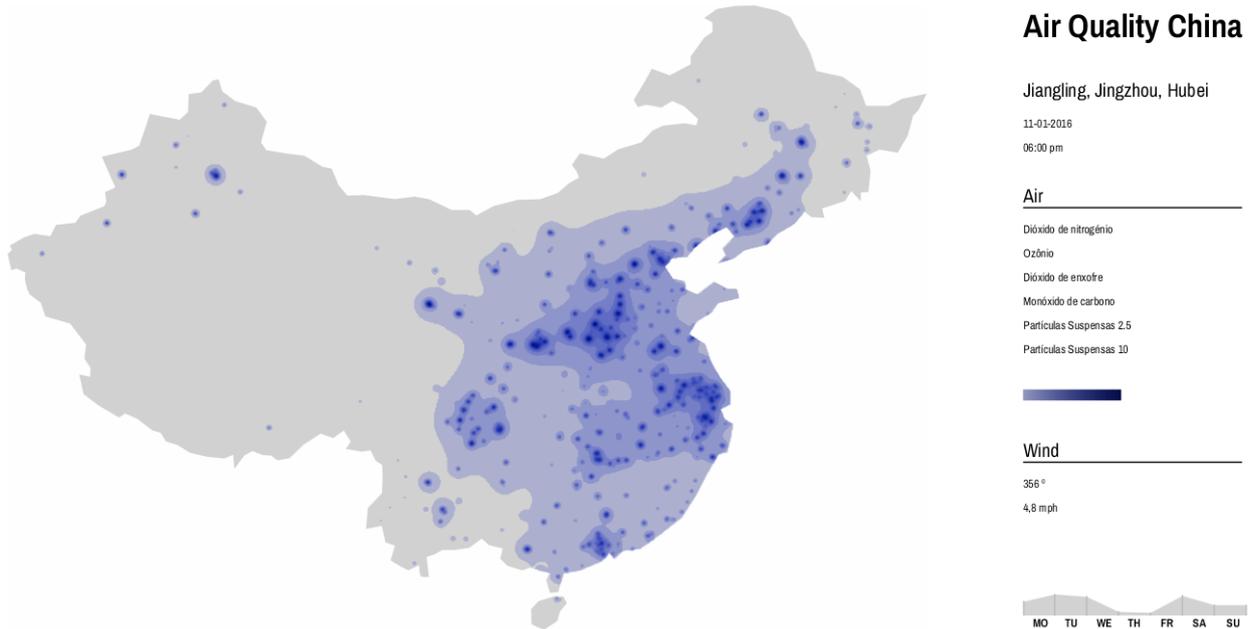


Figura 53 - Screenshot do artefacto visual depois dos melhoramentos gráficos

A legenda da imagem com as cores também muda ao clicar no nome. Nesta primeira fase foram utilizadas imagens, ao invés de representar as formas por ponto, numa tentativa de aumentar o processamento da aplicação e consequentemente a velocidade. Na secção 4.2 é abordado com mais pormenor o *layout* bem com as restantes funcionalidades da aplicação.

A componente tempo ainda não está associada a este artefacto visual, razão pela qual só permite visualizar apenas uma hora do dia. Para incorporar a componente tempo na visualização e se conseguir ver o dinamismo da poluição foi feito um *morphing* entre as *shapes* de hora para hora e consequentemente de dia para dia.

O *morphing* é uma técnica que transforma uma *shape* A em outra *shape* B com uma transição suave. Para fazer a transformação das *shapes*, os pontos que as constituem são interpolados utilizando a função *lerp()* do *Processing*, criada para descrever movimento ao longo de um vetor e que recebe como valores de entrada o ponto A, ponto B e quantidade a interpolar entre os pontos (um valor

entre 0 e 1). Mas para isso acontecer é necessário que ambas as *shapes* tenham o mesmo número de ponto, o que não acontece na maioria das vezes.

Sendo assim, em primeiro lugar, foi necessário criar um algoritmo capaz de escolher entre todas as *shapes* a melhor para fazer o *morphing*, ou seja, a *shape* com o menor número de pontos diferentes e que estivesse mais próxima da de origem. Para isso, foram calculados os centroides de todas as *shapes* em dois momentos diferentes, por exemplo à 1h e às 9 h.

De seguida, são percorridos os centroides da primeira hora (1h) e procurados centroides correspondente (próximos) na segunda hora (9h), a uma distância máxima (normalmente o raio da circunferência em que se insere a forma). Como é ilustrado pela imagem da figura 54, em que a *shape* do primeiro momento (1h) encontra duas correspondências no segundo momento (9h). Para escolher uma das duas formas o algoritmo procura o centro mais próximo e a forma que tenha o menor número pontos diferentes.

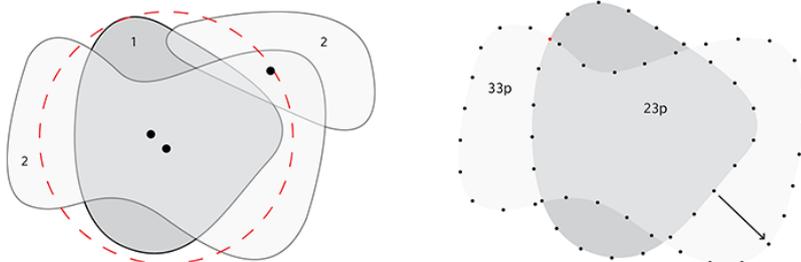


Figura 54 - Imagem ilustrativa da procura da shape ideal para o *morphing* (à esquerda), contagem dos pontos e adição de novos (direita).

Depois de encontrar a *shape* ideal (a mais parecida) para fazer o *morphing* é necessário contar número de vértices de cada uma e identificar a que tem o menor número. Visto que, elas têm de ter o mesmo número de pontos para fazer a transformação, são adicionados vértices equidistantes à forma mais pequena até ficar com o mesmo número.

Posteriormente, são procuradas interseções entre as duas *shapes* que queremos transformar. Caso exista algum ponto coincidente, o conjunto de vértices de ambas as *shapes* são reordenados a partir desse ponto (como é ilustrado na imagem da direita da figura 54 pelo um ponto a vermelho). Se não existirem pontos coincidentes é traçada uma reta do centro da primeira *shape* e reordenados os pontos a partir do ponto em que ela intersectar a segunda *shape*.

Os pontos das *shapes* são reordenados para evitar que existam deformações ou rotações das formas em torno de si mesma. Sempre que o *morphing* é feito entre *shapes* de níveis diferentes de poluição a cor dela também muda a mesma velocidade da transformação. O *morphing* das *shapes* (superfícies estatísticas) de uma hora para outra, por exemplo da 1h para as 9h, permite-nos “descobrir” informação nova, ou seja, estimar os níveis de poluição e as áreas afetadas neste intervalo de tempo que não são fornecidos pelos dados.

Com o intuito de comparar os diferentes padrões gerados pelos dados da poluição e conseguir encontrar algum tipo de correlação entre ambos, foi criada uma nova aplicação que permite sobrepor as representações e ver o seu dinamismo ao longo do tempo. Esta visualização representa os dados por *layers* (camadas), mas não sobrepõe níveis de poluição inferiores aos superiores, ou seja, cada nível de poluição é apresentado por ordem crescente independentemente do tipo de gás que é escolhido em primeiro.

Ao clicarmos no nome do poluente, no menu “*Air*”, ao contrário do que acontecia anteriormente, em que um padrão desaparecia e aparecia outro, desta vez ficam ambos visíveis e é o utilizador que define o tipo de poluente quer ver e quais quer sobrepor. Além disso, é desenhado um quadrado, ao lado do nome do gás poluente, com a cor da *layer* que está a ser representada, dando a indicação que está ativa. Se quisermos desativar uma *layer* basta clicar com botão direito do rato sobre o nome e tanto o quadrado como a *layer* desaparecem. Para auxiliar a visualização foi criado

um gráfico de linhas que mapeia as médias da poluição durante a semana. Cada linha é desenhada com a cor correspondente à do gás poluente e sempre que desativamos a *layer* ela também desaparece. Este gráfico permite comparar os diferentes gases e encontrar correlações entre ambos, bem como, detetar picos de poluição e analisar os comportamentos dos gases poluentes ao longo do tempo. Sendo assim, o gráfico está dividido por horas e dias, e ao clicarmos na tecla “UP” do teclado é incrementada uma hora, a cada 8 horas é carregado um novo *morphing*. Para o utilizador se situar no tempo foi criado um ponteiro, no gráfico, que marca a hora e acrescentada a data completa por baixo.

As cores escolhidas para representar os dados da poluição foram baseadas no *Color Brewer 2.0*,^[60] uma ferramenta que aconselha cores para cartografia consoante a natureza dos dados que são utilizados (quantitativos, qualitativos) e o número de classes. A transparência da cor, aproximadamente 50%, é a variável visual que mapeia o valor da poluição para cada tipo de poluente. Como podemos visualizar na imagem da figura 55 as cores dos picos de poluição sobressaem independentemente do número de *layers* ativas. Por exemplo, a cor verde sobressai em algumas zonas do mapa (Xangai e Beijing) à cor vermelha porque os níveis de poluição do Dióxido de Nitrogénio são mais elevados do que os de Monóxido de Carbono. Quando os níveis de poluição são parecidos as cores misturam-se, neste caso, formando castanho.

^[60] Color Brewer 2.0 [online], acedido em 27 de novembro de 2016, disponível <http://colorbrewer2.org>

Na imagem da figura 56 podemos visualizar que os níveis de Dióxido de Enxofre são mais elevados no norte da China, enquanto os níveis de Monóxido de Carbono são mais elevados no centro e sul. Se repararmos no gráfico com médias, podemos ver que os desenhos das linhas são muitos parecidos, chegando até a ser coincidentes em alguns momentos da semana, o que indica que existe correlação entre os gases.

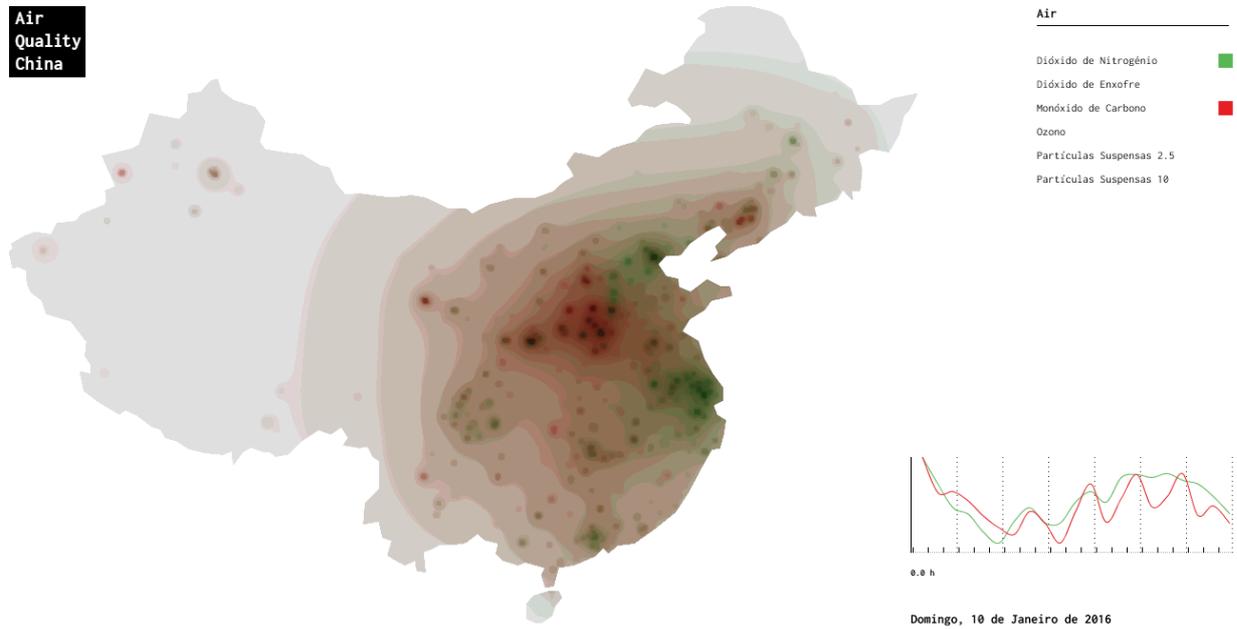


Figura 55 - Screenshot da visualização do Dióxido de Nitrogénio e do Monóxido de Carbono (sobrepostos).

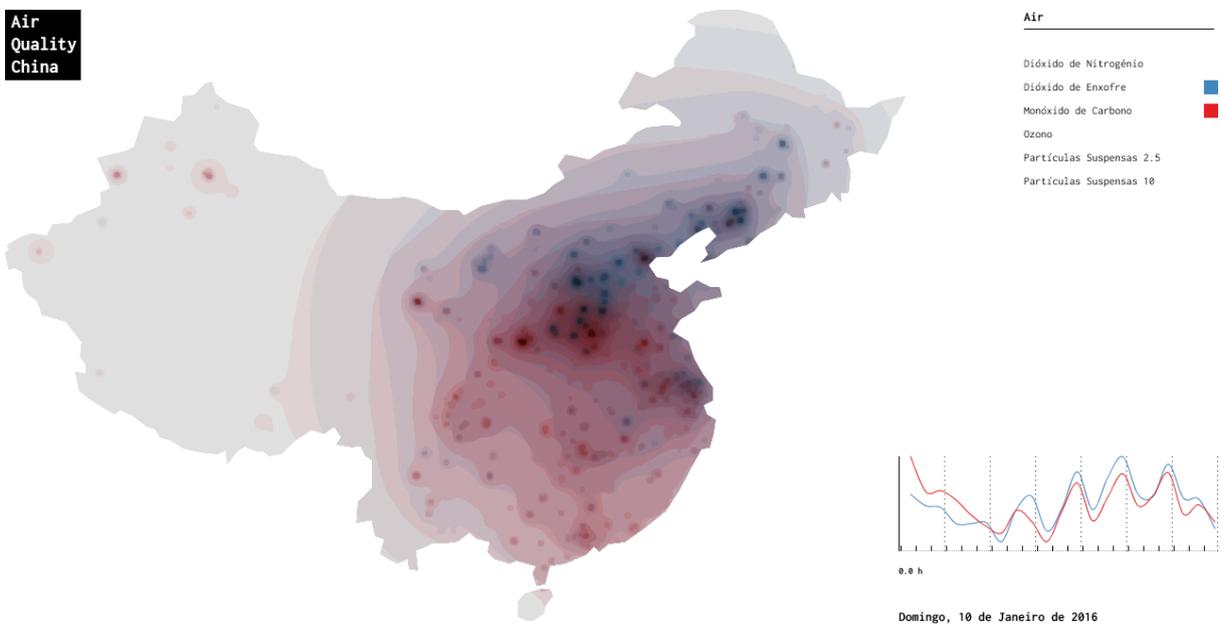
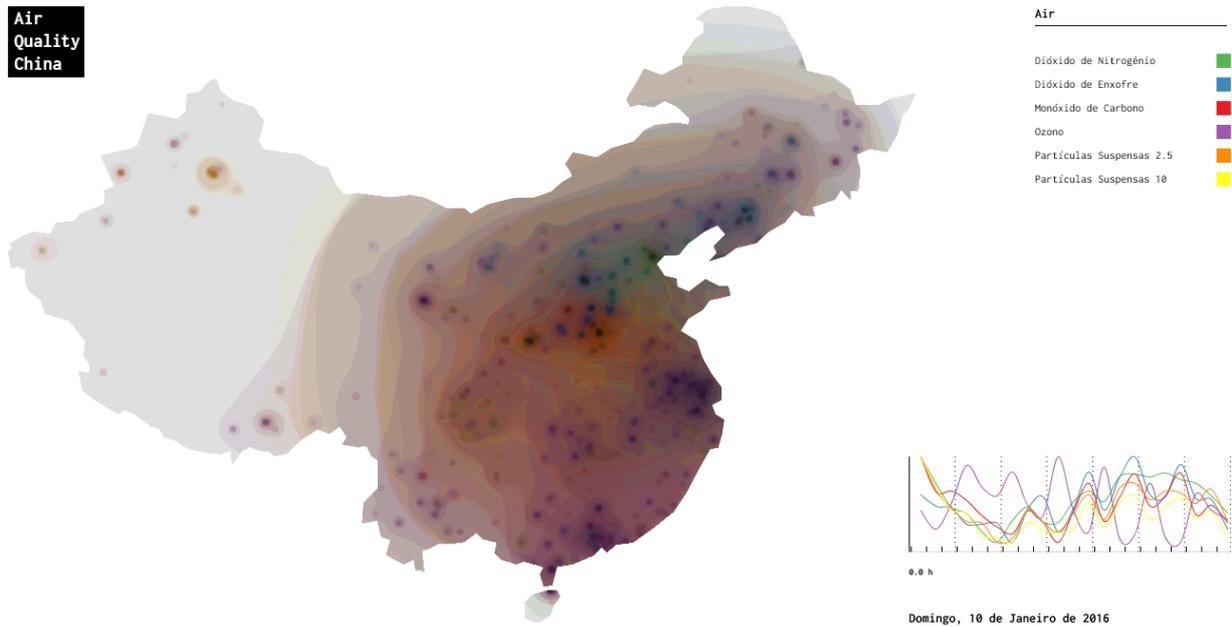


Figura 56 - Screenshot da visualização do Dióxido de Enxofre e do Monóxido de Carbono.



Mesmo com a visualização de todos os gases ativos, consegue-se identificar diferentes picos de poluição, como o Monóxido de Carbono (vermelho) e Partículas Suspensas 2.5 (laranja) no centro do país mais precisamente na região de Shaanxi, o Dióxido de Nitrogénio (verde) no noroeste do país perto da cidade de Beijing e o Ozono um pouco por toda a China com mais intensidade nas cidades de Xangai e Hong-Kong. No gráfico, à exceção do Ozono, os desenhos das linhas são muito parecidos, indiciando a existência de correlação entre os restantes gases.

Figura 57 - Screenshot da visualização com todos os gases ativos (sobrepostos).

4.2 Visualização do vento

Dados e Ferramentas

Previamente à recolha dos dados do vento, foi necessário fazer uma pesquisa de bases de dados que disponibilizassem informações meteorológicas na China. Sendo assim, foram testadas várias fontes como a *Weather Unlocked API*, *OpenWeatherMap API* e *Weather Underground API*. Mas devido a diversos problemas como o número reduzido de pedidos permitidos e em alguns casos a falta de informação (inconsistências nos dados), levaram a que fosse escolhida a API, *The Dark Sky Forecast*. Uma fonte consistente e menos limitadora, que apenas não fornece dados da região Chinesa conhecida como Tibete.

Devido ao elevado número de pedidos feitos ao iniciar a aplicação, o que a tornava lenta e caso não houvesse ligação à Internet a impossibilitaria de funcionar, houve a necessidade de criar um programa, unicamente, para pré-armazenar os dados em ficheiros de texto (.txt).

Para tal, foram feitos vários pedidos à API em formato JSON. Inicialmente, de 3 em 3 horas, durante uma semana para visualizar se haviam desvios nos padrões gerados, ou seja, se os padrões gerados variavam muito nesse curto espaço de tempo. Por fim, os pedidos foram realizados em função do *sample* de dados de uma semana (10 de janeiro de 2016 a 17 de janeiro de 2016) de poluição, ou seja, 3 vezes por dia (1h,9h,17h). Posteriormente os dados recolhidos foram tratados e armazenadas em vários ficheiros de texto por dia e hora.

Contudo, estes pedidos também foram influenciados pelo uso de uma grelha que divide o país em várias áreas geográficas. Assim, os pedidos foram feitos com base na localização do centro (posição x e y) de cada quadrado. Os dados recolhidos contêm informações relacionadas com a localização (latitude e longitude), velocidade e direção do vento.

Por fim, foram recolhidos dados com as fronteiras (limites) da China, em formato CSV. O que permitiu representar o mapa do país e localizar os dados sobre o mesmo. Anteriormente, tinham sido feitos testes com a biblioteca *unfolding maps*, mas devido a incompatibilidades com a nova versão do *Processing* não funcionava. O *Processing* é uma linguagem de programação, baseada no Java, que fornece todas as ferramentas necessárias para trabalhar com elementos gráficos, como desenhar uma linha ou criar uma animação.

Primeiras abordagens

Depois de recolher e armazenar os dados, em vários ficheiros separados, para aumentar a velocidade de processamento e consequentemente a velocidade da aplicação, foram realizados os primeiros testes visuais com os dados. Em primeiro lugar, devido à natureza dos dados geográficos, foi necessário utilizar a forma do país, China, para representar as informações. Ao mesmo tempo foi mapeada a latitude e a longitude para medidas do ecrã (x e y).

Como já referido, anteriormente, foi utilizada uma grelha (invisível) que divide o país em pequenas áreas (quadrados) e ajuda a retratar o movimento do vento. A grelha mapeia as localizações no mapa, sem ela seria difícil gerar padrões com os dados. Assim, os dados referentes ao vento (velocidade, direção e localização geográfica) foram carregados para cada célula da grelha (quadrado) ao iniciar a aplicação.

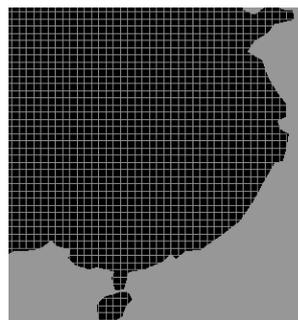


Figura 58 - Screenshot do mapeamento da grelha quadrangular, China.

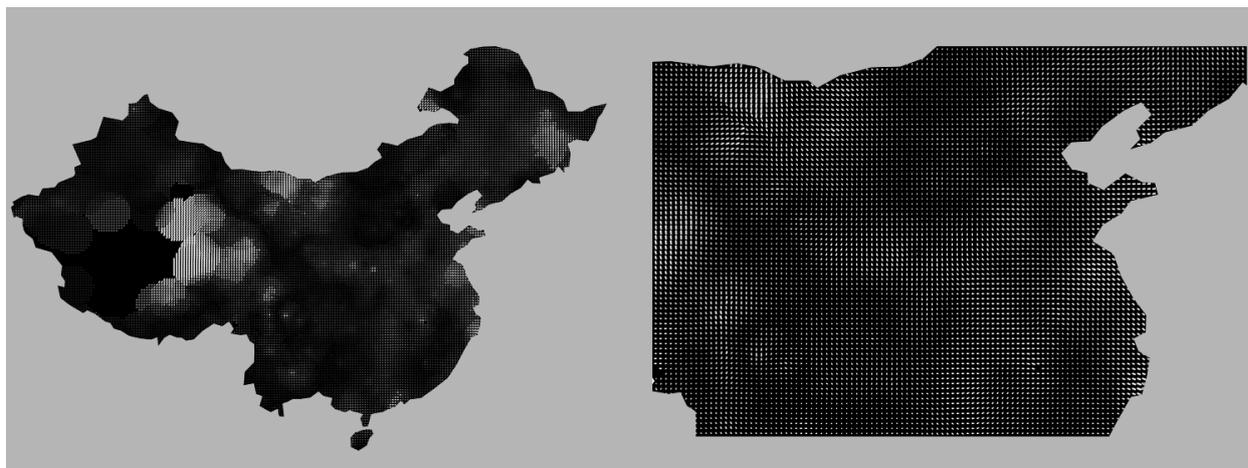
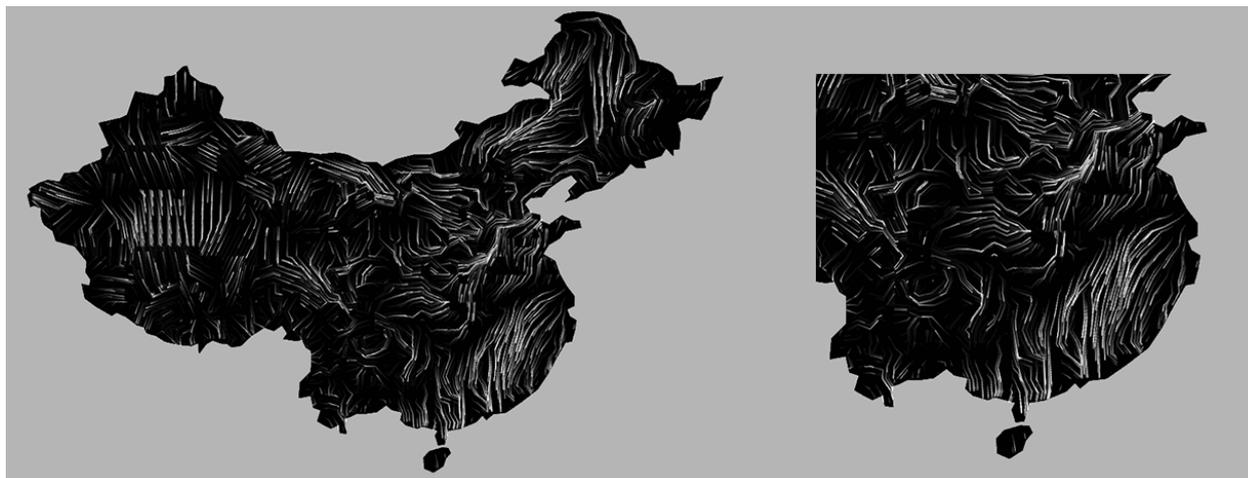


Figura 59 - Screenshot da primeira representação dos dados do vento.

Numa primeira tentativa de representar os dados do vento, foram mapeados os valores carregados na grelha para linhas. Assim, à semelhança do trabalho desenvolvido por Snow e mais recentemente pelo Projecto *Ukko*, foi criado um mapa de “linhas”.

Em cada quadrado da grelha é desenhada uma linha, a orientação da linha representa a direção do vento e a espessura a velocidade. Como podemos ver na imagem da figura 59, os padrões gerados permitem-nos identificar zonas de maior e menor intensidade de vento, no entanto, que direção do vento não é tanto perceptível. Desta forma, numa segunda fase decidimos animar as linhas e criar um mapa de fluxo.

Inicialmente foi utilizada uma grelha de 2 pixels por 2 pixels, mas a aplicação ficava lenta devido ao elevado número de células e de partículas que tinham de ser percorridas pelo programa. Por isso, foram feitos testes com uma grelha de 10 pixels por 10 pixels, mas o nível de detalhe dos padrões gerados era baixo e o movimento do vento era pouco perceptível. Sendo assim, foi utilizada uma grelha de 5 pixels por 5 pixels, que permite ter uma perceção coerente do dinamismo do vento e ao mesmo tempo um processamento menos pesado.



Esta foi a primeira representação animada do vento criada com os dados. Retratando o movimento do vento com partículas brancas que se movem dentro do espaço do mapa, mais precisamente os valores da velocidade e da direção do vento em cada ponto do mapa (localização) num determinado intervalo de tempo.

Figura 60 - Screenshot da representação animada do vento por partículas.

A direção do vento é representada no mapa pelo rastro deixado por cada partícula. Desta forma, o fundo (*background*) não é limpo, ou invés disso, é atualizada a cor do mapa com uma transparência de aproximadamente 85%, o que permite que sejam visualizados os padrões desenhados pelas partículas no fundo.

A direção norte é fornecida nos dados pelo valor 0 (graus), só que no *Processing* esse valor representa a direção este. Portanto, a todas as direções foi necessário retirar 90 graus ou, então, somar 270 graus. Isto porque os ângulos no *Processing* são contados ao contrário, ou seja, no sentido horário. Só desta forma é possível representar corretamente a direção do vento.

A percepção da velocidade do vento é dada pelo tempo de vida das partículas, mas também pela espessura da linha. As partículas são criadas em cada quadrado da grelha e quanto maior for a

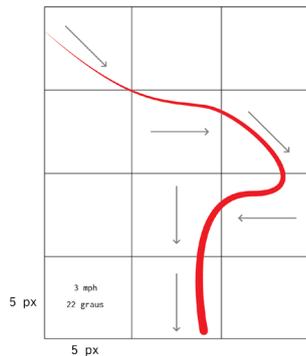


Figura 61 - Imagem ilustrativa do movimento das partículas.

velocidade menor é o tempo de vida, isto é, a cada *frame* a velocidade é descontada ao tempo de vida. Assim, quando uma partícula “morre”, ela volta a “nascer” no quadrado de origem. Este ciclo repete-se criando padrões com mais ou menos densidade conforme a velocidade do vento na área geográfica em questão. O mesmo acontece quando uma partícula ultrapassa dos limites do país. Para que as partículas circulem dentro dos limites do mapa, o algoritmo criado percorre todas as células da grelha e compara a sua localização com a localização das partículas, se ambas forem iguais, ou seja, se a partícula estiver dentro da área do quadrado as suas variáveis (velocidade e direção) são atualizadas com os valores pré-carregados na célula. Como é ilustrado na imagem da figura 61.

Posteriormente à representação do movimento do vento, foi criado um gráfico, auxiliar, que mostra, de forma mais detalhada, a velocidade do vento ao longo do tempo, o que facilita e ajuda na identificação de picos de velocidade num determinado ponto do mapa. O algoritmo reage ao *click* do rato, no espaço do mapa, guardando a posição x e y selecionada. De seguida, carrega e percorre os ficheiros de texto com os dados da semana e procura as velocidades do vento, para cada hora (de 3 em 3 horas), referentes à localização guardada. Ao mesmo tempo, calcula a velocidade máxima durante a semana, necessária para representar corretamente os dados. Por fim, as informações encontradas são mapeadas sobre eixos verticais, cada um com 100 pixels e de igual espaçamentos entre eles.

Figura 62 - Screenshot do gráfico auxiliar que representa as velocidades do vento para um dado ponto do mapa durante uma semana.



Como pode ser visualizado na imagem da figura 62, a forma pintada de preto representa a hora do dia que estamos a ver os dados, neste caso 06:00 da manhã. Para ajudar a ler o gráfico foi acrescentada uma legenda com os valores da velocidade e da direção do vento, auxiliada pelo nome da cidade que é obtido com um pedido à API da *google maps* através das coordenadas (latitude e longitude) do ponto selecionado.

Além disso, também foram acrescentados à legenda os dias das semanas, a hora e a data completa. Informações, essas, extraídas dos ficheiros de texto carregados. A tipografia escolhida para esta fase foi a *Raleway*, uma fonte *free*, da *google fonts*.

Assim, o utilizador pode explorar os dados do vento através da seleção (ao clicar no mapa) e se pretender trocar a hora ou o dia da visualização pode fazê-lo usando a tecla “UP” ou “DOWN” do teclado do computador que incrementa 3 horas e provoca um novo carregamento dos dados. Sempre que os dados são carregados as partículas desaparecem e voltam a ser desenhadas de novo em cada célula da grelha.

O dia da semana que estamos a visualizar os dados é rasurado. Como se pode ver na imagem da figura 63, em que *wednesday* (*we*) aparece com uma linha sobre o mesmo. O incremento também é aplicado à hora e à data, neste caso, para a data trocar são necessários sete incrementos, porque os dados são carregados de 3 em 3 horas.



Figura 63 - Screenshot da legenda complementar à visualização e ao gráfico.



Figura 64 - Screenshot da primeira aplicação criada com os dados do vento (3 horas do dia 27 de fevereiro de 2016).

Abordagem final

Nesta fase do desenvolvimento da visualização do vento foram feitos melhoramentos gráficos na representação dos dados. Em primeiro lugar, foram identificados alguns erros evidentes. A legenda do lado direito é muito grande ocupando aproximadamente 1/3 do ecrã. É dado muito destaque a informações, menos importantes, como a data, as horas e os dias da semana. Estes tipos de dados precisam de ser hierarquizados e organizados.

O nome da visualização está em falta, quem não estiver por dentro do projeto não consegue facilmente identificar o país em questão apenas pela forma, nem que tipo de dados estão a ser representados. A tipografia utilizada não é a melhor opção para apresentar informações em ecrãs. A cor do fundo é escura, é necessário repensar os aspetos gráficos do artefacto.

O gráfico auxiliar, por baixo do mapa, é muito grande ocupando também cerca de 1/3 do espaço do ecrã. Não é intuitivo que tipo de informação que está a ser visualizada. A legenda do gráfico não pode estar separada do mesmo, isso faz com que o utilizador perca a atenção e tenha de desviar o olhar para encontrar a informação.

O resultado deste desperdício de espaço é que o artefacto visual (principal) ocupa menos de metade do ecrã. É necessário aumentar o tamanho do mapa e reduzir o da lenda e do gráfico auxiliar, para que os padrões gerados pelas partículas, retratando o movimento do vento, sejam visualizados com detalhe e precisão.



Figura 65 – Screenshot da abordagem final aos dados do vento.

O tamanho da janela de visualização foi alterado, a largura passou de 900 pixels para 1400 pixels e a altura de 600 pixels para 700 pixels o que, conseqüentemente, aumentou o tamanho do mapa. Além disso, foi libertado mais espaço para o mapa ao reduzir o tamanho do gráfico auxiliar e trocando-o de lugar.

As informações (legenda) e os botões foram organizados por blocos, mais precisamente três blocos de informação. O primeiro contém com o nome da visualização, “*Air Quality China*”, o nome da cidade e do estado, a data completa e a hora. Os tamanhos do tipo de letra também foram reduzidos e a tipografia trocada pela *ArchivoNarrow*.

No segundo bloco de informação, “*Air*”, foram colocados os botões com os nomes dos gases poluentes, um tema abordado na secção a seguir. Por último o bloco de informação, “*Wind*”, é composto pela velocidade do vento em milhas por hora, pela direção do vento em graus e pelo gráfico auxiliar no qual foram colocados

os dias da semana, por baixo de cada área referente. A cor fundo da visualização que era cinzento foi mudado para branco o que, juntamente com os ajustes gráficos que foram feitos, a tornam mais limpa e intuitiva para o utilizador.

O sistema de partículas utilizado também sofreu alterações. Os rastros que até então eram desenhados no fundo (*background*), passaram a ser guardadas num *array* com um tamanho igual 16, ou seja, as últimas 16 posições percorridas por cada partícula são guardadas.

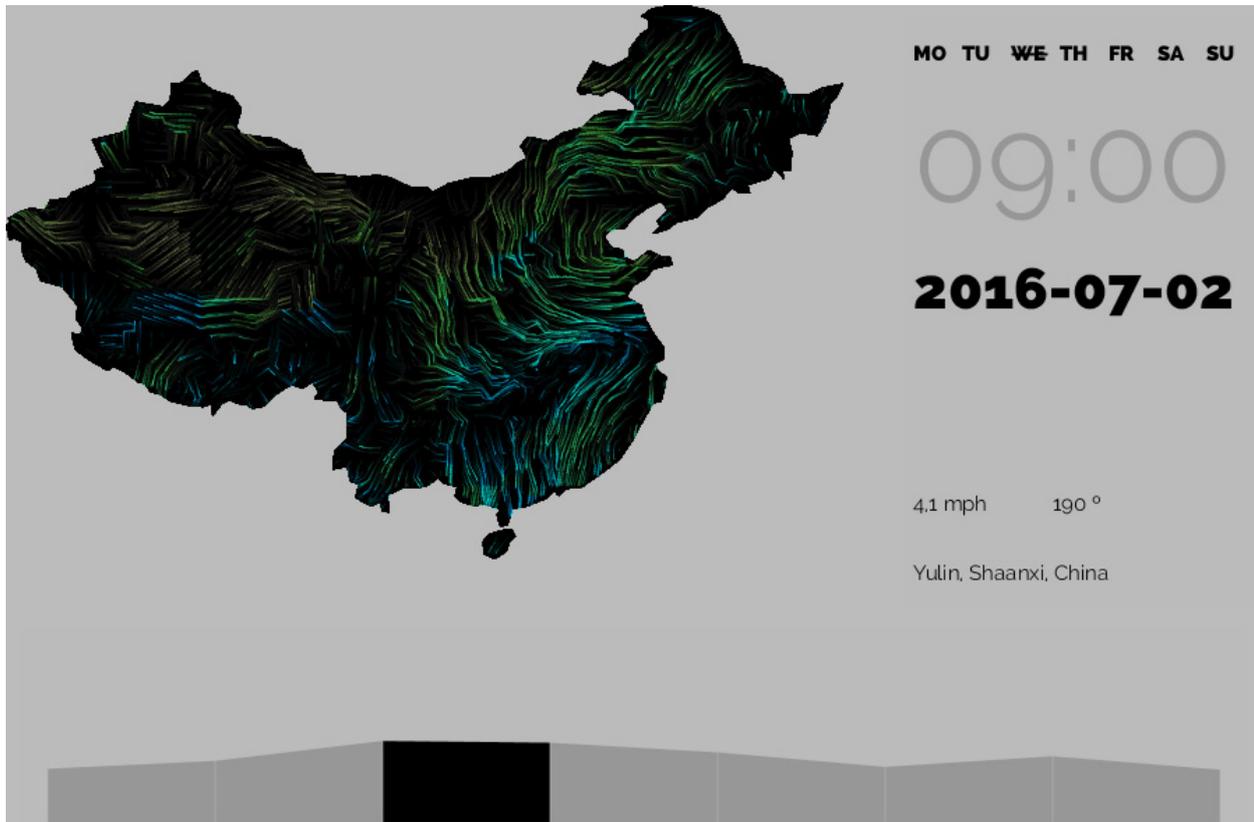
Posteriormente são desenhados os rastros por linhas que conectam os pontos com uma transparência que é mapeada a partir do tempo de vida de cada partícula. Isto porque para representar a poluição, tema abordado na secção a seguir, é necessário limpar o fundo devido às animações utilizadas. Caso contrário não era possível sobrepor este tipo de informações. Esta técnica resolve um problema, mas acarreta outro, o processamento da aplicação fica mais pesado o que faz com que as partículas circulem mais devagar.

4.3 Sobreposição das visualizações

Nesta secção, pretende-se mostrar os testes feitos para cruzar a informações fornecidas pelos dados da poluição e do vento. Esta fase completa as secções abordadas anteriormente, 4.1 e 4.2, o principal objetivo é encontrar algum tipo de correlação entre ambos e tirar conclusões que possam levar à descoberta de conhecimento novo. Desta forma, serão apresentadas as primeiras abordagens feitas e as dificuldades encontradas que nos levaram à construção de novas formas de representação até chegar a um resultado final

Primeiras abordagens

Inicialmente os dados da poluição foram mapeados diretamente na grelha juntamente com os dados do vento, como já foi aprofundado na secção 4.2. A cor das partículas foi utilizada como variável visual para a representar (poluição), ou seja, o nível de poluição era indicado pela cor das partículas consoante a área geográfica em que se encontravam. Como podemos ver na imagem da figura 66 o azul representava os níveis baixos de poluição e o verde os mais altos. A ideia era conseguir visualizar algum tipo de correlação entre ambos, por exemplo se o vento de alguma forma transportava a poluição para outras localizações ou se a velocidade do vento influenciava os níveis de poluição.



Ao fim de alguns testes chegou-se à conclusão que esta técnica não era a melhor opção para representar os dados da poluição, devido ao facto de em algumas zonas do mapa não existir vento, ou o mesmo ser reduzido, o que fazia com que não conseguimos ver a poluição porque esta dependia da representação do vento. Foi então que se pensou noutra forma de representar os dados da poluição, que permitisse comparar com a visualização do vento e correlaciona-los.

Assim sendo, desenvolveu-se um mapa isométrico ou *contour map* que representa as áreas afetadas e os níveis de poluição, como foi aprofundado com mais detalhe na secção 4.1.

Figura 66 - Screenshot da primeira abordagem à correlação dos dados.

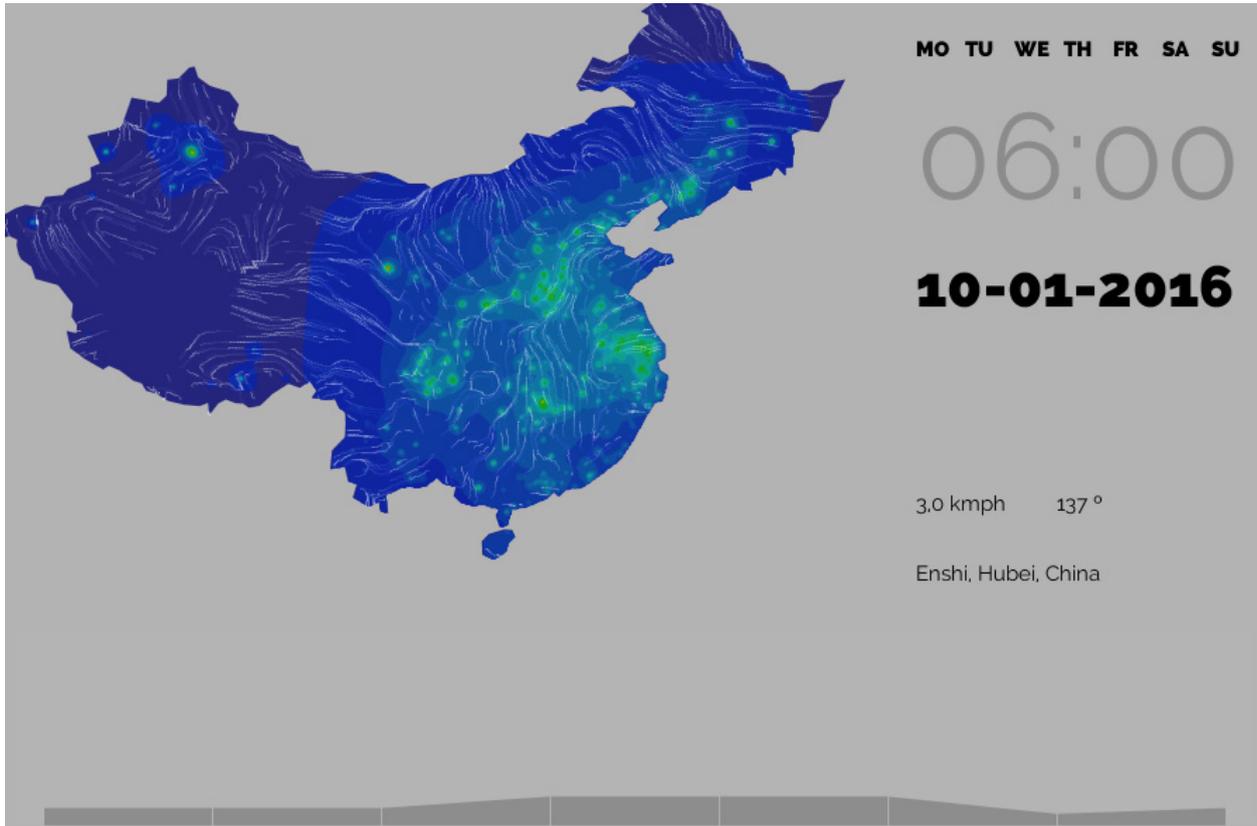
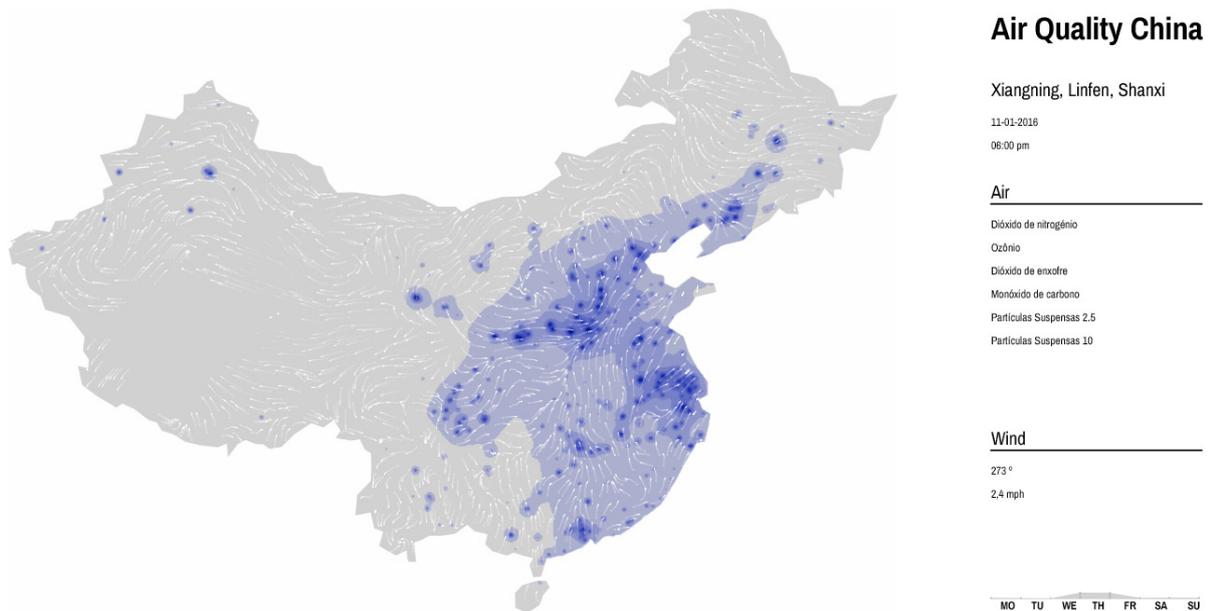


Figura 67 - Screenshot da segunda abordagem à correlação dos dados.

Nesta segunda abordagem, os níveis de poluição foram mapeados numa escala de verde e azul, onde verde representava os níveis de poluição mais elevados e o azul os mais baixos. O rastro das partículas passou a ser guardado, ao invés, de ser desenhado no fundo do mapa. Isto porque, o *background* tinha de ser limpo a cada *frame* para que os padrões da poluição não fossem desenhados uns em cima dos outros. Desta forma, os dados podiam ser sobrepostos (cruzados), ao mesmo tempo, sem inconsistências. Entretanto, a aplicação precisava de melhoramentos tanto ao nível de cores, organização da informação, como técnicos (otimização). O menu do lado direito era muito grande e o mapa tinha um tamanho reduzido. Além disso, não era perceptível a correlação entre os dados. A visualização da poluição era estática, por esse motivo não se conseguia ver se o vento influenciava no seu movimento para outras zonas geográficas.



Na imagem, da figura 68, podemos ver os melhoramentos gráficos que foram feitos no artefacto visual. Tanto as cores das *shapes*, atribuída uma diferente a cada gás poluente, como a cor do fundo foram mudados. A informação (outputs) foi organizada e hierarquizada por menus. O utilizador pode interagir com a aplicação e trocar a visualização dos gases, como as velocidades no gráfico do vento. Apesar da janela ter sido aumentada, o espaço está melhor distribuído e é dado mais ênfase ao mapa. Entretanto, devido ao facto da visualização da poluição ser estática, ainda não podemos visualizar a sua correlação com o vento. Para tal, foi criado o *morphing* entre as *shapes* que permite visualizar o seu dinamismo ao longo do tempo e mapear dados que eram desconhecidos.

Figura 68 - Screenshot da terceira abordagem à correlação dos dados.

Abordagem final

Nesta última fase decidimos adicionar os dados do uso do solo à visualização, não só para a completar, mas também correlacionar com os dados da poluição, ou seja, para visualizar se de alguma forma influenciam na geração de gases poluentes.

Os dados do *land use* (uso do solo) utilizados para o desenvolvimento da aplicação foram recolhidos em formato *Shapefile* no site *GeoFabrik*^[61] que trabalha diretamente com dados *free* do projeto *OpenStreetMap*.^[62] Os dados são disponibilizados por país e atualizados diariamente. Depois de fazer o *download* dos arquivos para uma pasta de trabalho, foi necessário utilizar um *software GIS (Geographic Information System)*, neste caso o *Quantum GIS*, para processar a informação e converter o formato.

Das vinte classes do uso do solo, disponíveis, decidimos só utilizar apenas cinco (industrial, cultivo, comercial e *retail*) porque achamos que as restantes não têm relevância para o trabalho. Devido ao facto dos dados serem representados por áreas, como podemos ver na imagem da figura 69, foi necessário utilizar o *plugin, real centroid*, para encontrar o centro de cada *shape* e posteriormente guarda-los em formato JSON num ficheiro de texto.

^[61] GeoFabrik [online], acedido em 20 de janeiro de 2017, disponível em <http://download.geofabrik.de/asia/china.html>

^[62] O projeto OpenStreetMap foi fundado no Reino Unido em 2004 e tem como objetivo criar um conjunto de dados geográficos disponíveis em todo o mundo. O foco é principalmente nas infraestruturas de transporte (ruas, caminhos, caminhos-de-ferro, rios), mas também recolhe uma multiplicidade de pontos de interesse, edifícios, características naturais, informação sobre o uso do solo, bem como costas e fronteiras administrativas.



Os ficheiros de texto criados com as informações dos dados do *land use* foram posteriormente carregados para um programa *Processing*. Em primeiro lugar foi calculada a densidade de pontos, ou seja, foram percorridas todas localizações do *land use* e para cada uma foi calculado o número de vizinhos num raio de 100 pixels, como é ilustrado na imagem da figura 70. Assim, utilizando a mesma técnica para representar os dados da poluição foi gerado um *height map*, em escala de cinzas, para cada classe diferente do uso do solo. Como os dados não tinham valores associados à área, o valor da densidade, calculado, foi mapeado no M (massa) da fórmula do *distance decay*.

Em seguida, foi aplicada a deteção de *blobs* as imagens geradas com as diferentes categorias do uso do solo e criados *contour maps*. Dos quais foram extraídas as superfícies com diferentes concentrações, por exemplo de indústrias. Como mostra a imagem da figura 71.

Figura 69 - Screenshot do ficheiro Shapefile com os dados do land use no programa QGIS.

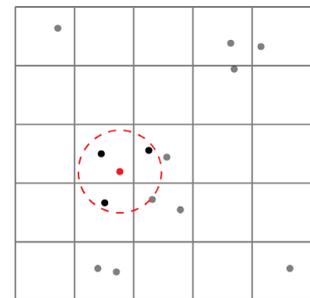


Figura 70 - Ilustração do cálculo de densidades para cada ponto do land use.



Figura 71 - Imagem ilustrativa do cálculo de densidades de pontos.

Em suma, depois de representar os dados do *land use*, as visualizações foram juntas num único artefacto visual. À semelhança dos dados da poluição, as representações do uso do solo são sobrepostas por camadas no mapa. Isto, permite correlacionar os padrões com os dados da poluição e tirar conclusões específicas. Desta forma, foram adicionados botões à aplicação que permitem ativar as diferentes categorias. As superfícies estatísticas são, assim, representadas pela cor preta e a variável visual utilizada é a transparência (50%), que permite identificar as zonas de maior e menor concentração do uso do solo. Como podemos ver na imagem da figura 72 as zonas mais escuras (pretas) é onde existe maior concentração de áreas.

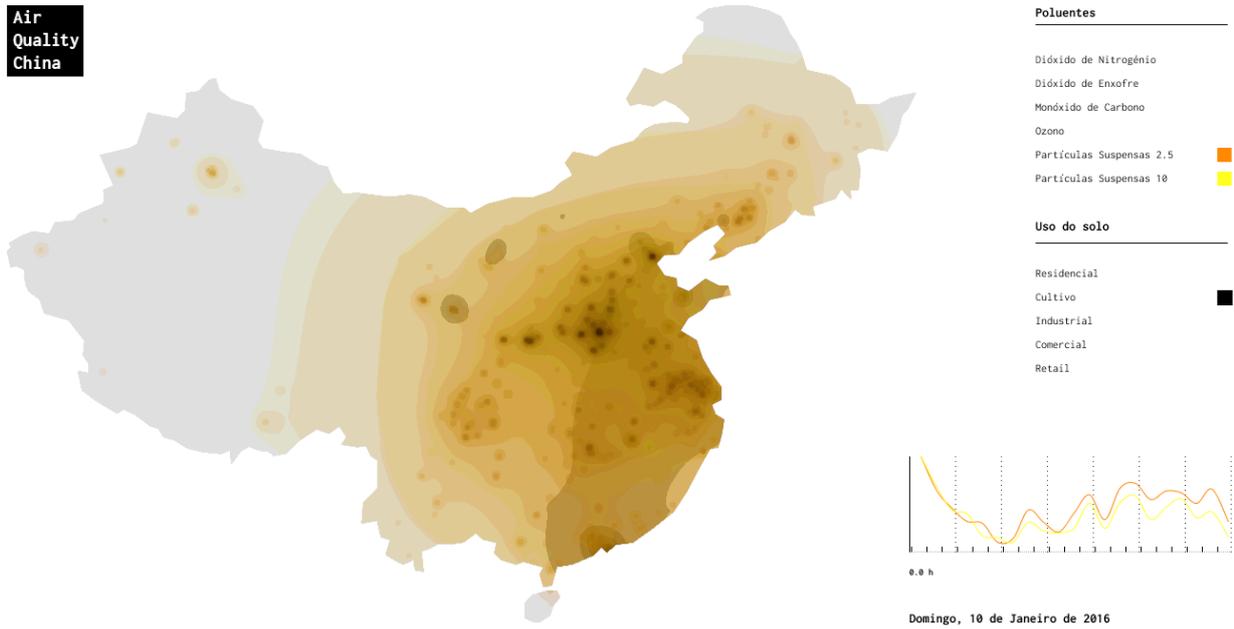


Figura 72 - Sobreposição dos dados do land use.

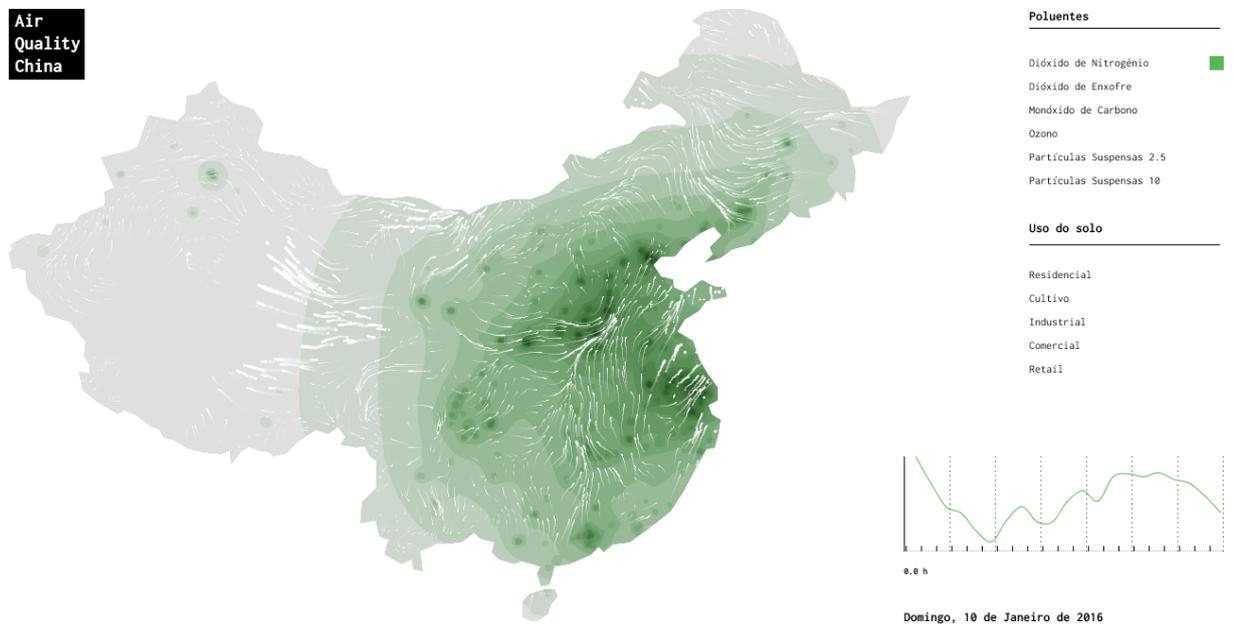


Figura 73 - Sobreposição com os dados do vento.

5 Conclusão

Esta dissertação, começa por apresentar vários conceitos sobre a representação de dados, ou seja, visualização de informação. Com o intuito de desenvolver uma visualização que fosse eficaz, foi desenvolvido o estado da arte, que abrange a teoria e as técnicas de visualização existentes para representar os dados utilizados. Segundo Bertin, para transmitir-se a informação de forma eficiente, o designer, tem de ter em conta o conteúdo da informação (*content*) e as propriedades do sistema gráfico (*container*). Sempre que falarmos de informação, a noção central comum a todas as correspondências tem o nome de invariante e conceitos variáveis envolvidos, os quais serão chamados de componentes. As diferentes partes identificáveis de um componente ou de uma variável são chamadas, por Bertin, de elementos ou categorias.

As relações estabelecidas entre os componentes ou elementos, definem quatro níveis de organização: associativo, seletivo, ordenado ou quantitativo. Para representar a informação, o designer tem à sua disposição um sistema gráfico, e a codificação da informação e feita por marcas. Todas as marcas podem variar de posição no ecrã ou no papel, mas só algumas de tamanho, textura, valor, cor, orientação e forma.

Desta forma, existem doze variáveis visuais que podem ser aplicadas ao ponto, à linha ou à área, que constituem as três figuras elementares do plano geométrico e a que Bertin deu o nome de implantação ou classes de representação. O ponto representa uma localização e como tal pode variar de todas as formas. A linha é definida por um conjunto de pontos e pode variar enquanto marca visual de tamanho, forma, cor, valor e orientação. A área é formada pelas duas classes de representação anteriores, ponto e a linha, podendo variar de valor, textura, cor.

A forma como são expressas as correspondências no plano, permitem-nos definir quatro grupos de representação: diagramas, redes, mapas ou símbolos. Existem variáveis visuais que são mais adequadas para representar a informação do que outras, a percepção da mesma define o seu nível de organização. Bertin definiu cinco níveis de organização: associativo, dissociativo, seletivo, ordenado e quantitativo.

Tal como, Bertin, Edward Tufte fez um estudo sobre as representações visuais de dados, embora mais focado nos aspetos funcionais do gráfico, descrevendo boas e más práticas na construção de uma visualização. Assim, a excelência em gráficos estatísticos consiste em comunicar ideias complexas com clareza, precisão e de forma eficaz. Um gráfico não é distorcido se as representações visuais dos dados forem consistentes com a representação numérica. Diferentes pessoas veem as mesmas áreas de maneiras diferentes, a percepção muda com a experiência e a experiência é dependente de contexto.

De qualquer modo, dadas as dificuldades de percepção, o melhor que podemos esperar é alguma uniformidade nos gráficos, e assegurar que os observadores têm uma boa chance de obter os números certos. Tufte, definiu seis princípios que conduzem em direção a esses objetivos e conseqüentemente aumentam a integridade gráfica: primeiro, a representação dos números de ser diretamente proporcional as quantidades numéricas representadas; segundo, deve ser utilizada uma legenda para completar a representação gráfica e debater a ambigüidade; terceiro, deve ser mostrada a variação dos dados não a variação do design; quarto, se existirem eventos importantes nos dados devem ser escritos no próprio gráfico; quinto, em séries de tempo que envolvam dinheiro, as unidades deflacionadas e padronizadas são quase sempre melhores o que as unidades nominais e sexto, o número de variáveis não deve exceder a dimensão nos dados.

Uma das grandes vantagens da evolução tecnológica para além do processamento de grandes quantidades de dados é a interatividade. O tempo é importante na interação, mas o efeito funciona de forma diferente consoante o tempo de resposta. Assim sendo, são definem três níveis de interatividade: o primeiro nível, também conhecido como momento psicológico, tem um tempo de resposta um décimo de segundo, estímulos com este intervalo de tempo fundem-se numa perceção de movimento; o segundo nível de interação, tem um tempo de resposta de um segundo que é o tempo de uma conversa improvisada; o terceiro nível de interação, tem um tempo cíclico de resposta de cinco a trinta segundos, conhecido como unidade de tarefa, é o ritmo de uma capacidade cognitiva em que o utilizador faz uma tarefa.

Para interagir com um gráfico, o utilizador tem diferentes possibilidades. Existem três grupos de técnicas de interação com os dados, *data transformation* é um grupo formado por seis interações: *dynamic queries* permitem ao utilizador selecionar e filtrar dados da base de dados, *direct walk* é navegação de página para página, *details-on-demand* expande um conjunto pequeno de objetos, *attribute walk* permite cruzar dados, *brushing* é a seleção dos mesmos dados em múltiplas visualizações para comparar os resultados e *direct manipulation* permite a parametrização de vários elementos; *visual mapping* é um grupo composto por duas interações: *dataflow diagram* usa uma representação explícita como um diagrama *node-link* para representar o mapeamento, *pivot table* permite ao utilizador rapidamente manipular os dados e mapeá-los em linhas ou colunas; *view transformations* um grupo de cinco interações: *direct selection* é seleção e destaque de um objeto ou conjunto de objetos, *camera movement* é a mudança de posição do observador, *magic lenses* seleciona objetos consoante a sua posição, *overview + detail* é a sobreposição de visualizações e *zooming*, envolve a redução do número de objetos que são visíveis.

A segunda parte do estado da arte foca-se nas técnicas de visualização, mais precisamente nos mapas temáticos, tais como mapas de pontos, mapas de fluxo, mapas de símbolos graduados, cartogramas, mapas isométricos e isopleto e mapas coropléticos. Os mapas de pontos têm como principal objetivo mostrar a distribuição espacial de fenómenos utilizando o ponto com variável visual. Assim, podem representar dois conjuntos de dados discretos: fenómenos discretos com informação da localização geográfica, como o mapeamento médico ou fenómenos com variação regular, como a maioria dos mapas que representam os dados de censos, no qual os pontos são distribuídos dentro de uma área geográfica correspondente, a fim, de retratar densidades.

Um dos primeiros mapas de pontos utilizados com a finalidade de traçar padrões de doenças foi criado por John Snow. Nele, Snow mapeou as mortes e representou-as por pontos, com a finalidade de comprovar a sua teoria. O segundo exemplo apresentado é mais recente e representa densidades demográficas, mas especificamente onde as mulheres americanas vivem e a sua proporção comparada com a proporção global, num específico ponto. A cor é utilizada, neste mapa, para além da posição (ponto) para representar diferentes tipos de densidades. Os mapas de símbolos graduados, utilizam a variável tamanho para representar a magnitude dos dados discretos.

Na construção de mapas de símbolos graduados o principal objetivo é que o leitor consiga estimar o valor de um símbolo. Para isso, existem duas variáveis que devem ser consideradas para representar um símbolo: a forma e a escala. No primeiro exemplo apresentado são utilizados, por Minard, pela primeira vez *pie chart*, que descrevem mais que uma variável no mapa. Cada círculo é dividido para representar diferentes tipos de carne e o tamanho representa o peso. O segundo exemplo, mais recente, foi criado pelo jornal *The Washington Post* e representa cada estação de energia elétrica nos Estados Unidos. A variável cor é utilizada para indicar os diferentes tipos de usinas e o tamanho a capacidade de produção.

Os mapas coropléticos representam os dados utilizando a área e a cor, como variável visual. Normalmente os dados são agrupados por áreas administrativas e normalizados os valores. Assim, Guerry utilizou sete tons de cor diferentes no seu mapa para representar os níveis de criminalidade, por distrito em França. No segundo exemplo criado por Sonnad, são utilizados dados do *Twitter* nos Estados Unidos e categorizados por assunto, em seguida, por região. A cor é atribuída consoante a intensidade de utilização de uma palavra naquela região.

Os mapas isopleto, tal como, os mapas isométricos, são construídos por linhas que resultam da intersecção de um plano com a superfície estatística. Os mapas isopleto não permitem a estimativa de valores em qualquer ponto, porque os pontos de controle representam valores de áreas e não pontos reais, como é exemplificado pelo mapa de Daniel Huffman. Enquanto, os mapas isométricos os dados são recolhidos em pontos verdadeiros, por exemplo estações meteorológicas, para os quais os valores realmente existem, permitindo que os valores sejam determinados interpolando ou estimando valores entre linhas, tal como, o mapa criado por Edmond Halley.

Os mapas de fluxo retratam fenómenos lineares que envolvem a conexão de lugares e mostram o movimento de pessoas, bens, etc. A codificação é feita pela espessura da linha ou peso, e geralmente representa o volume de fluxo. Minard no seu gráfico utiliza espessura da linha para representar o fluxo de algodão da Europa, enquanto Yau mapeia o fluxo de pessoas a fazer exercício pelo recurso ao peso das linhas.

Por fim, os cartogramas que são mapas que destroem áreas ou distâncias para transmitir dados. Desta forma, dividem-se em dois grupos: os cartogramas de área, existem dois tipos os que preservam a forma, não contíguos, e os que preservam a topologia os contíguos; os cartogramas de distâncias, normalmente representam a relação entre o tempo e a distância. O projeto "*Traffic on Lisbon*" é um bom exemplo de cartogramas de distâncias em que

as estradas são dilatadas ou comprimidas consoante o fluxo de carros (trânsito). O “*Pulse of the Nation*”, criado por Mislove é um cartograma de áreas que representa os dados sobre o humor na América. O resultado é um mapa que uniformiza densidades.

Para terminar o nosso estudo na área da visualização de informação foram analisadas diferentes visualizações desenvolvidas, que utilizam dados parecidos com os nosso. O principal objetivo é entender como eles são representados, ou seja, que técnicas e variáveis são utilizadas. Sendo assim, primeiramente apresentamos o trabalho desenvolvido pela Fernanda Viégas e Martin Wattenberg, “*Wind Map*”, um mapa de fluxos que representa os dados do vento, nos Estados Unidos, no tempo e no espaço. Os dados utilizados neste projeto são atualizados diariamente o que permite ter uma visualização em tempo real, a direção e a intensidade do vento são indicadas por partículas brancas que se movimentam no espaço do mapa. Este projeto, consegue incorporar simultaneamente requisitos de uma visualização científica, sem perder a atração visual, o impacto artístico.

O segundo trabalho apresentado, “*The World Air Quality Index*”, utiliza os dados de poluição em todo o mundo e em tempo real. Os dados são representados depois de calcular o índice de qualidade do ar e a cor é mapeada consoante o nível de perigo para a saúde atribuído ao índice. Num segundo momento, ao interagir com o mapa, o utilizador pode ver os dados em específicos sobre ponto. A informação é expandida e abre uma janela com os gases e partículas presentes no ar.

No terceiro capítulo, desta dissertação, são abordados trabalhos relacionados com o nosso e que sobrepõe várias camadas de dados. O Projeto *Ukko* é o exemplo disso, e permite visualizar a intensidade do vento e ao mesmo tempo as centrais eólicas construída em todo o mundo no mapa. O objetivo deste trabalho é auxiliar as empresas de energia eólica na tomada de decisões, tais como,

prever locais novos para construir uma central ou centrais possivelmente menos produtivas. Ao mesmo tempo, como é uma aplicação online, qualquer utilizador pode a utilizar para prever o vento na sua área geográfica. Isto é, os dados utilizados no desenvolvimento desta ferramenta são referentes aos meses futuros.

O projeto *Iso-edges for the geovisualization of consumption* é uma visualização que utiliza *isolines* para mapear dados de consumo geograficamente. Cada linha utilizada para representar a informação no mapa, equivale a um valor e separam áreas de elevado e baixo consumo. Como a representação dos dados é feita no espaço e no tempo permite ao utilizador identificar simetrias, por exemplo entre o interior do país e o litoral.

Assim, iniciamos o projeto de visualização do vento com um conjunto de dados que continham direção, velocidade e localização. Decidimos que a visualização devia usar uma linguagem visual simples, clara e ser interativa para permitir ao utilizador explorar os dados. A ferramenta utilizada no desenvolvimento do trabalho foi o *Processing*. A nossa primeira tentativa, para representar os dados do vento baseou-se em linhas (partículas) que se movimentavam no espaço do mapa. Mas para isso foi necessário aplicar uma grelha quadrangular ao mapa e posteriormente aos dados para que as partículas não saíssem dos limites do país.

Sendo assim, foi necessário voltar a fase de recolha de dados e fazer pedidos à API com base no centro de cada quadrado da grelha e carrega-los para a mesma. Inicialmente o rastro das partículas que indicava a direção e a velocidade do vento era desenhado no fundo do mapa e foi adicionado um gráfico auxiliar para ajudar a ver-la ao longo da semana.

Numa segunda abordagem aos dados do vento, devido à sobreposição com a visualização da poluição, houve a necessidade de guardar o rastro das partículas e desenhá-lo posteriormente, visto que o *background* tinha de ser limpo a cada *frame*, o que nos trouxe problemas de otimização da aplicação e fez com que ela ficasse lenta. Além disso, também foram feitos melhoramentos gráficos, a informação foi organizada e hierarquizada por menus, ao mapa foi dado mais destaque e as cores mudadas o que tornou a visualização mais limpa e clara.

Na segunda parte do desenvolvimento da nossa aplicação, foram representados os dados da poluição na China, um conjunto de informações relativas aos índices de Dióxido de Enxofre, Monóxido de Carbono, Dióxido de Nitrogénio, Ozono, Partículas suspensas 2.5 e Partículas suspensas 10, com referente localização geográfica.

A nossa primeira tentativa de representar os dados baseou-se num mapa de pontos em que cada ponto representava o gás mais poluente em cada lugar, pela cor. Devido à sobreposição de pontos no ecrã, e ao ruído visual, foi necessário aplicar a mesma grelha que foi aplicada ao vento. Isto, fez com que os pontos trocassem de lugar, provocando inconsistências na visualização ao mesmo tempo que organizou a informação. Existiam pontos sobrepostos, isto é, com a mesma localização. Para colmatar esse problema foram percorridos todos os pontos com a mesma localização e desenhado apenas o que tinha o valor mais alto. O nível de poluição foi mapeado no raio do círculo.

Ao fim de alguns testes achamos que esta não era a melhor técnica para representar os dados da poluição, em primeiro lugar porque seria difícil de sobrepor com os dados do vento e em segundo lugar porque só eram representados os gases mais poluentes em cada localização, omitindo assim informações.

Assim, numa segunda abordagem aos dados da poluição e à semelhança do projecto *Iso-edges for the geovisualization of consumption*, decidimos criar um *contour map* ou mapa isométrico a partir dos dados da poluição. Para tal, num primeiro momento, foi utilizada a formula matemática, *distance decay*, que nos permitiu gerar imagens bidimensionais (*height maps*) numa escala de cinza com os dados, onde as zonas mais claras representavam os picos de poluição e as zonas mais escuras os locais com menores índices. Posteriormente foi utilizada a biblioteca *BlobDetection*, que se baseia no *threshold*, para detetar as diferentes áreas com iluminações diferentes. Desta forma, foi gerado um *contour map* para cada tipo de gás poluente e atribuída uma cor diferente a cada um.

As áreas com diferentes níveis de poluição que eram delimitadas pelas linhas de contorno foram pintadas com a respectiva cor e a transparência utilizada como variável visual dos níveis de poluição. Esta técnica, permitiu-nos visualizar as áreas afetadas pela poluição, bem como, detetar picos de poluição e simetrias, por exemplo entre ocidente e o oriente do país. Posteriormente, foi feito o *morphing* entre as *shapes*, áreas de diferentes níveis de poluição, de uma hora para outra hora com o objetivo de representar o dinamismo dos gases e consequentemente mapear dados que não eram fornecidas. Em seguida os diferentes padrões gerados pelos dados foram sobrepostos, por camadas, e adicionados botões que permitiam ligar ou desligar as *layers* (camadas). Ao mesmo tempo foi construído um gráfico de linhas com as médias de poluição de cada gás e as linhas desenhadas com as cores do gás representado. Isto permitiu não só visualizar o comportamento de cada gás ao longo do tempo com também detetar pico de poluição e correlações entre os gases.

Na terceira e última parte do desenvolvimento da aplicação foram feitas várias tentativas de juntar as duas visualizações, ou seja, sobrepor os dados do vento aos dados da poluição.

Inicialmente foram mapeados os dados da poluição na grelha, tal como, dos dados do vento e o nível de poluição representado pela cor das partículas. Entretanto, ao fim de alguns testes chegamos à conclusão que quando não existe vento não podíamos ver a poluição, mesmo que ela existisse. Desta forma, numa segunda abordagem à sobreposição dos dados e com uma representação diferente da poluição, utilizando um mapa isométrico, foi necessário representar os padrões gerados pelo vento de outra maneira, isto é, guardar o rastro das partículas e depois desenhá-lo, visto que o *background* tinha de ser limpo. Apesar deste melhoramento na representação do vento e conseguir sobrepor as duas visualizações ainda não era perceptível existir correlação entre os dados. A visualização do vento era dinâmica, mas a da poluição era estática.

Nesta fase, ainda não tinha sido aplicado o *morphing* às shapes, por esse motivo não conseguíamos ver se o vento deslocava a poluição para outras localizações. Assim, numa última abordagem para sobrepor as visualizações foram adicionados dados do uso do solo (*land use*) para complementá-la e correlacionar com a poluição.

Trabalhos futuros

Devido ao facto das partículas sobrecarregarem o processamento da aplicação, como trabalho futuro pretendemos aplicar o algoritmo *quadtree* que divide o ecrã em pequenos quadrados e representa apenas o que estamos a ver, aumentando assim a velocidade da aplicação. Isto, possibilitaria também a implementação do *zoom*, visto que, as imagens são rasterizadas não podemos fazer *scale* porque ficam despixelizada, o que nos obriga a gerar imagens maiores para posteriormente as aumentar. Além do que já foi referido, com a implementação do *quadtree* o movimento das partículas ficaria mais fluído o que permita uma visualização melhor da correlação do vento com a poluição.

6 Bibliografia

Yau, N. (2013). *Data points: visualization that means something*. Indianapolis, IN: John Wiley & Sons, Inc.

Tufte, E. R. (2006). *Beautiful evidence*. Cheshire, CT: Graphics Press.

Meirelles, I. (2013). *Design for information: an introduction to the histories, theories, and best practices behind effective information visualizations*. Beverly, MA: Rockport .

Fry, B. (2008). *Visualizing data*. Beijing: O'Reilly Media, Inc.

Bertin, J. (1983). *Semiology of graphics*. Madison, WI: University of Wisconsin Press.

Tufte, E. R. (1983). *The visual display of quantitative information*. Cheshire, Conn. (Box 430, Cheshire 06410): Graphics Press.

Johnson, J. M. (2003). *Geographic information: how to find it, how to use it*. Westport, CT: Greenwood Press.

Tyner, J. A. (2010). *Principles of map design*. New York: Guilford Press.

Pine, J. C. (2009). *Natural hazards analysis: reducing the impact of disasters*. Boca Raton, FL.: Auerbach Publications.

Card, S. K., Mackinlay, J. D., & Shneiderman, B. (1999). *Readings in information visualization: using vision to think*. San Francisco, CA: Morgan Kaufmann

Bertin, J. (1983). *Semiology of graphics*. Madison, WI: University of Wisconsin Press.

MacEachren, A. M. (1995). *How maps work: representation, visualization, and design*. New York: Guilford Press.

Card, S. K., Moran, T. P., & Newell, A. (2008). *The psychology of human-computer interaction*. Boca Raton, Fla.: CRC Press.

Lupton, E., & Phillips, J. C. (2008). *Graphic design: the new basics*. New York: Princeton Architectural Press.

Tufte, E. R. (1997). *Visual explanations: images and quantities, evidence and narrative*. Cheshire, CT: Graphics Press.

Cleveland, W. S. (1994) *The Elements of Graphing Data, Revised*. New Jersey: Hobart Press, 1994.

Maças, C., Polisciuc, E., Machado, P. et al. (2016). *Iso-edges for the Geovisualization of Consumptions*. Departamento de Engenharia. CISUC, Universidade de Coimbra, Coimbra, Portugal.

Greenpeace. *Reduce Air Pollution*. Acedido em 31 de agosto de 2016. Disponível em www.greenpeace.org/eastasia/campaigns/air-pollution/.

Calijuri, M. C., Gasparini, D. (2012) *Engenharia ambiental. Conceitos, Tecnologia e Gestão*. Editora Campus.

Friendly, M. (1802-1866) *The Life and Works of André-Michel Guerry*, York University, Toronto, Canadá.

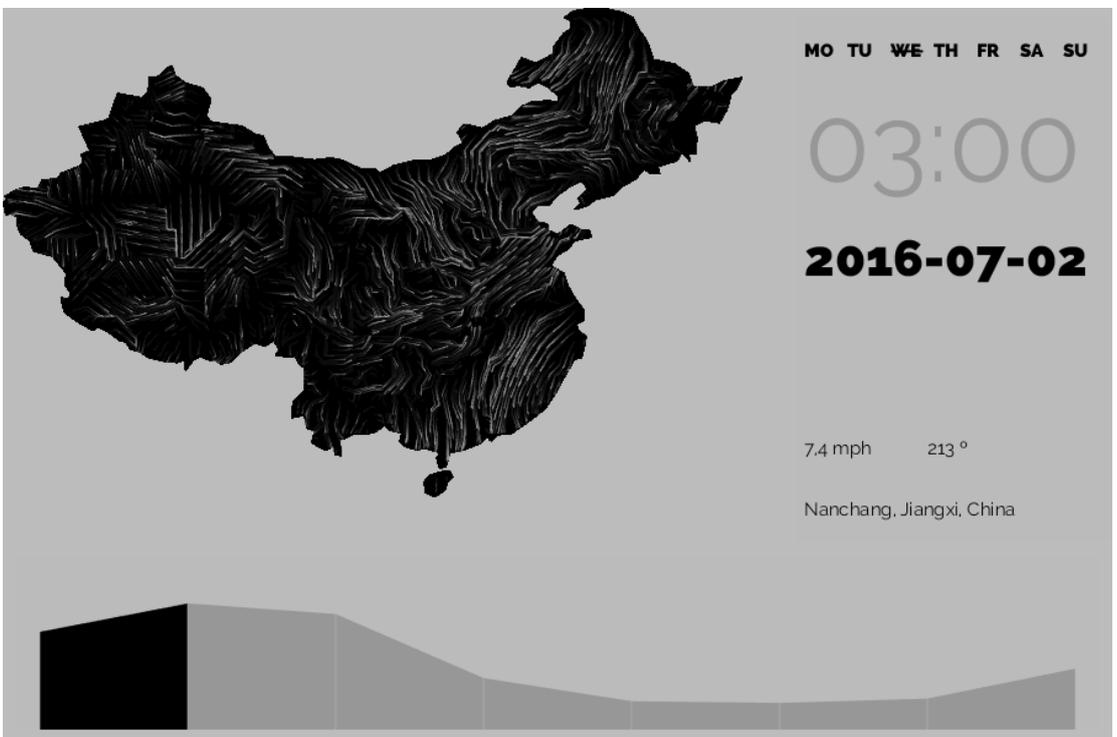
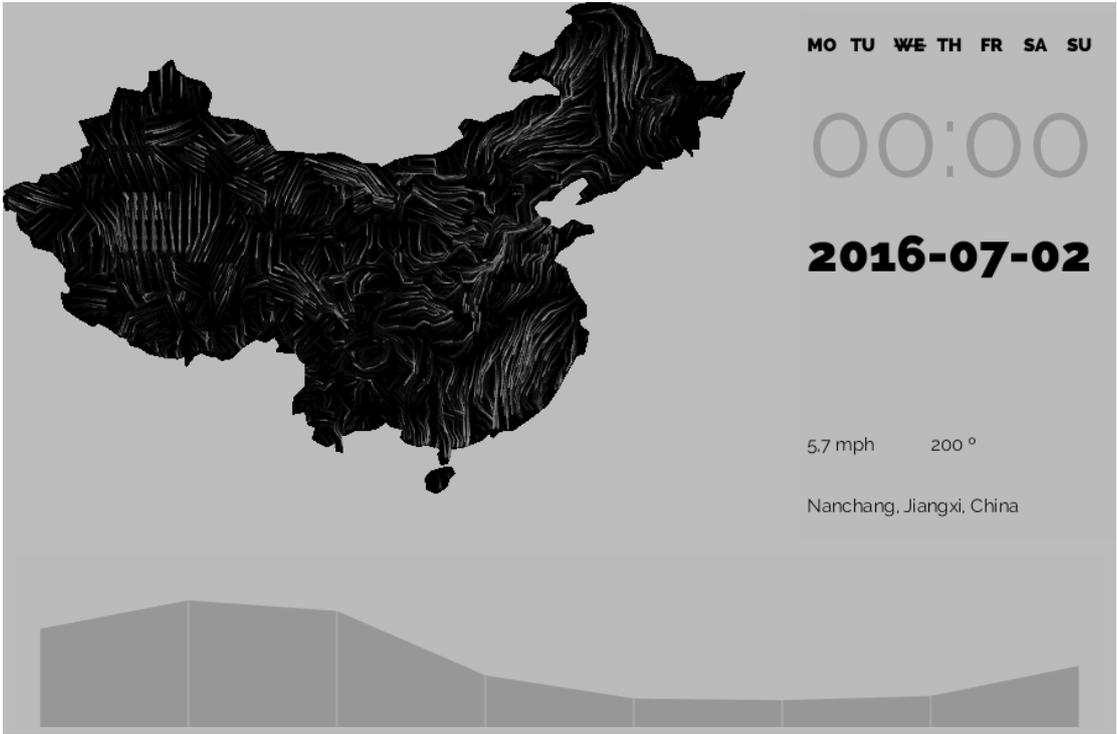
Kamolov, R. (2013) *Visualizing Patterns of Flow and Usage in Wireless Networks*, University of Coimbra, Coimbra, Portugal.

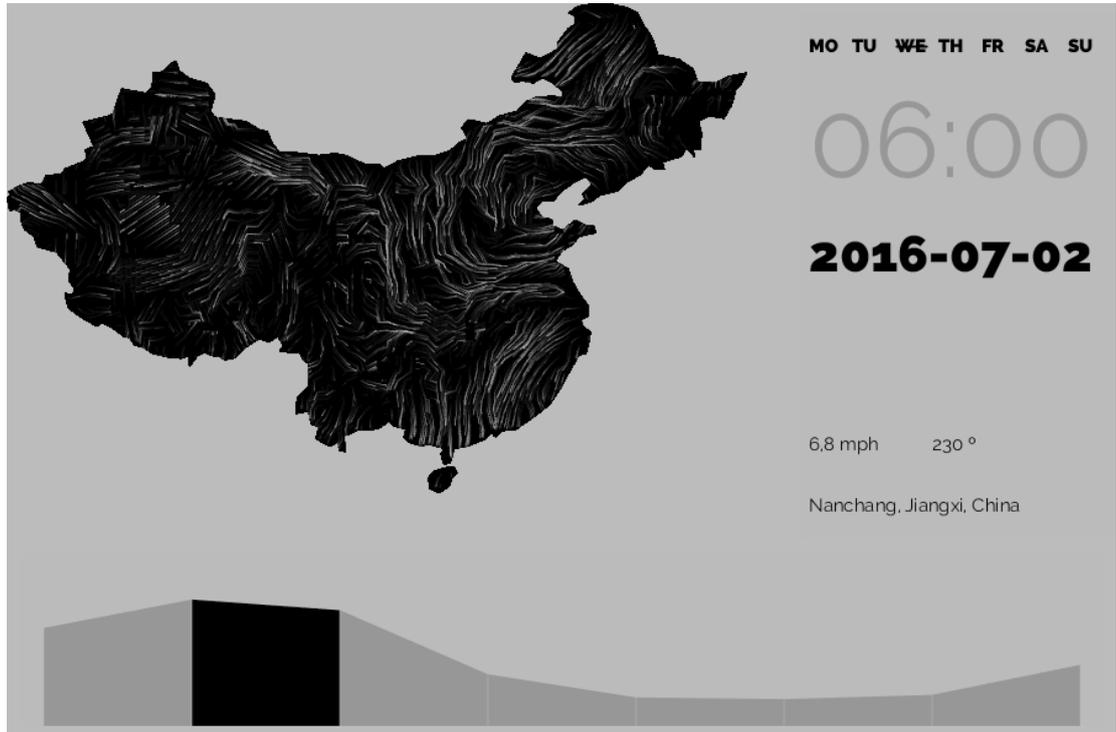
Polisciuc, E. (2013) “*Visual Tools for the Study of Urban Mobility, Design Strategies for Visualization and Analysis, of Row and Semantic Data to Understand Urban Mobility*”, University of Coimbra, Coimbra, Portugal.

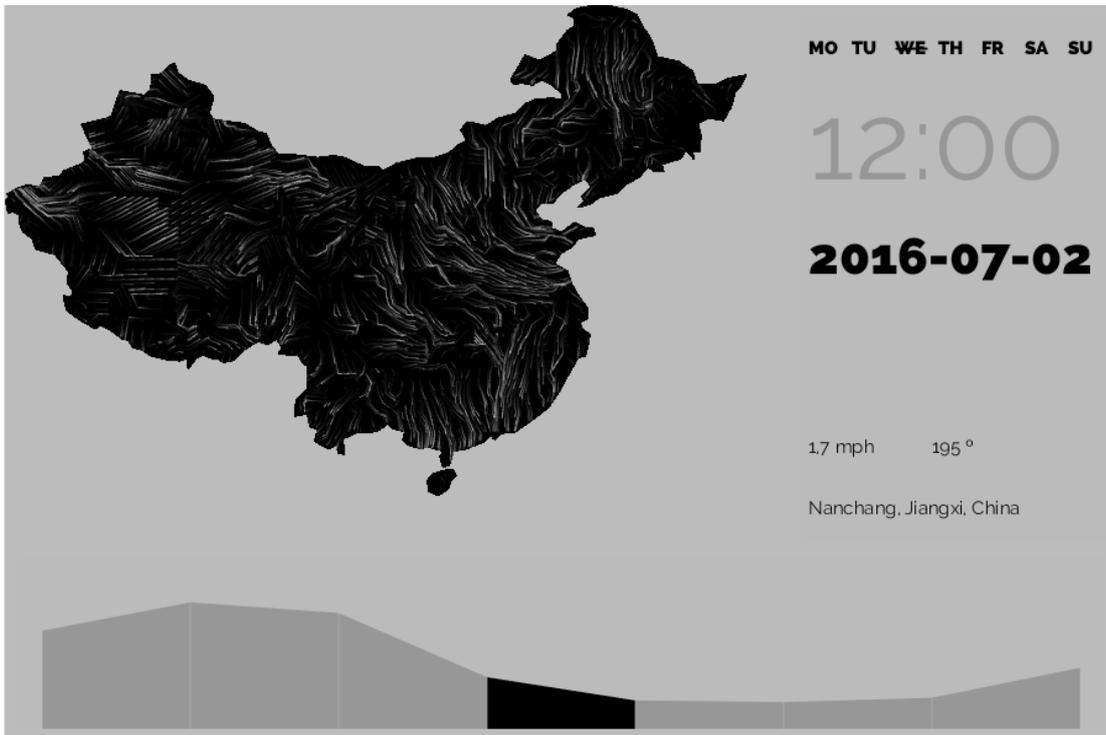
7 Anexos

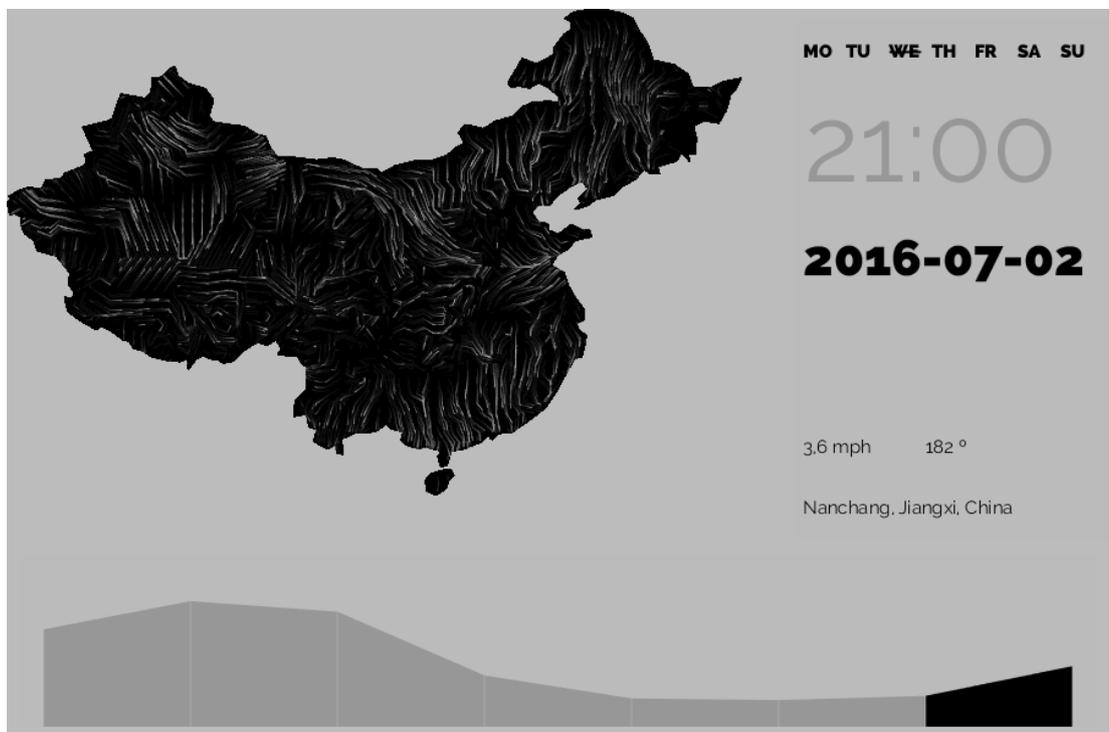
7.1 Visualização do vento

Abordagem Inicial





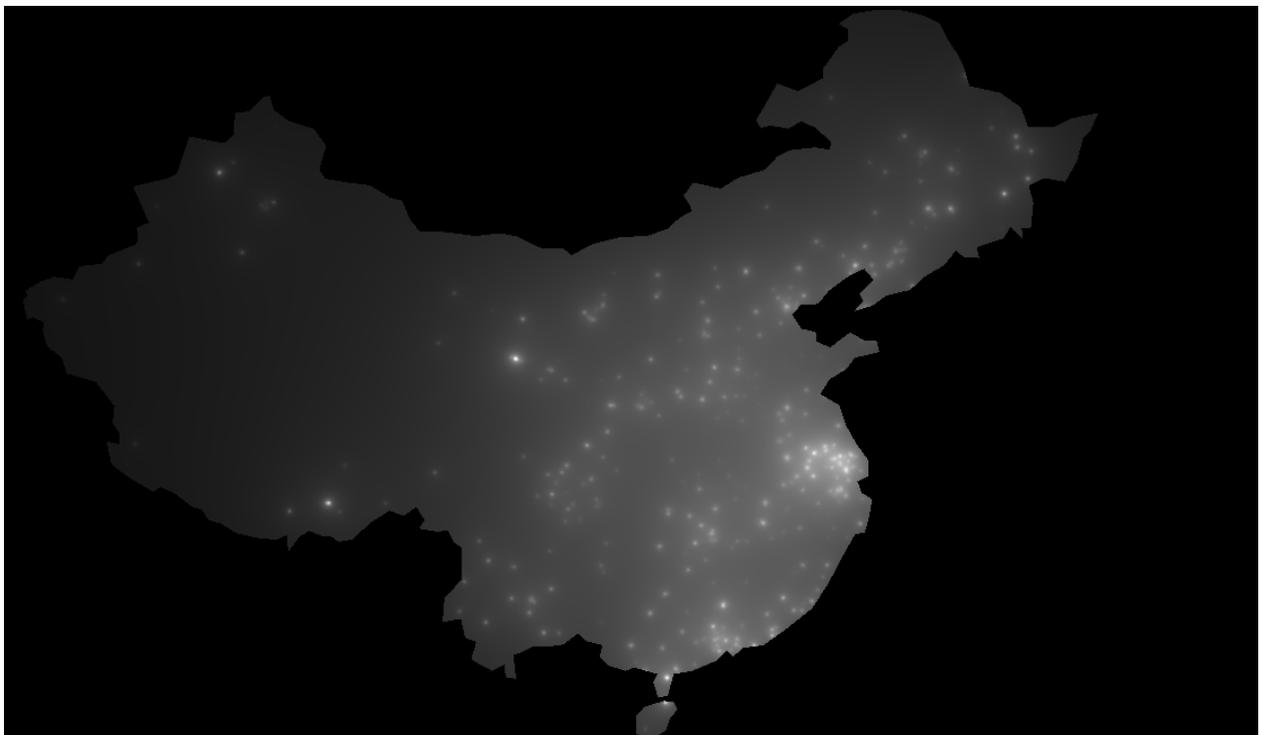




7.2 Visualização da poluição

Height maps

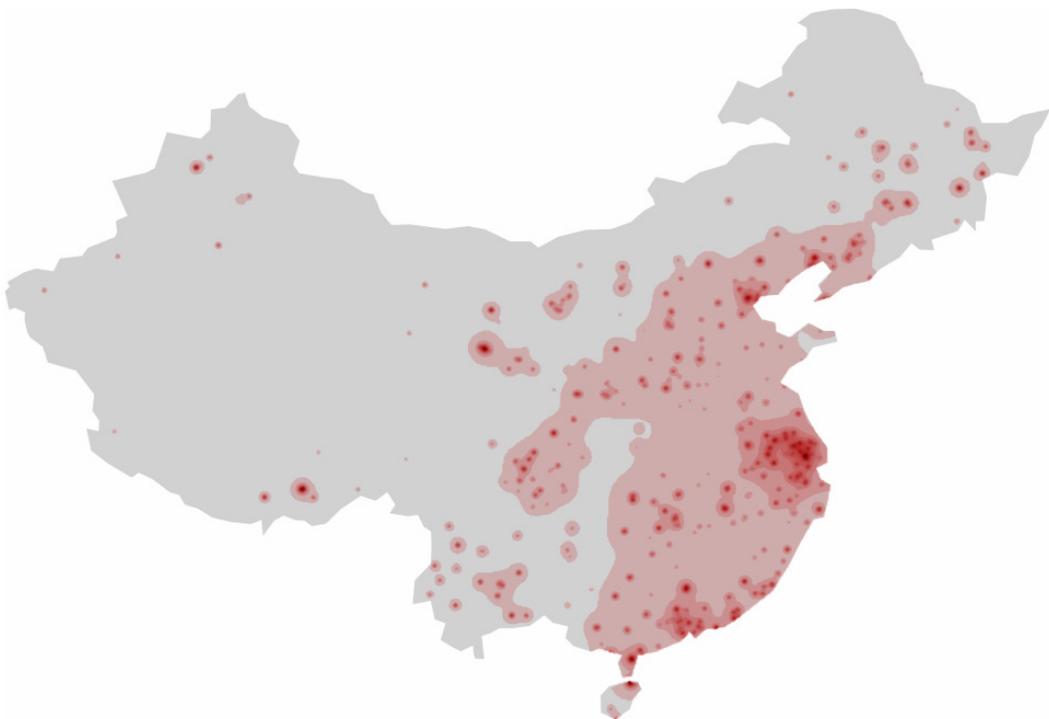
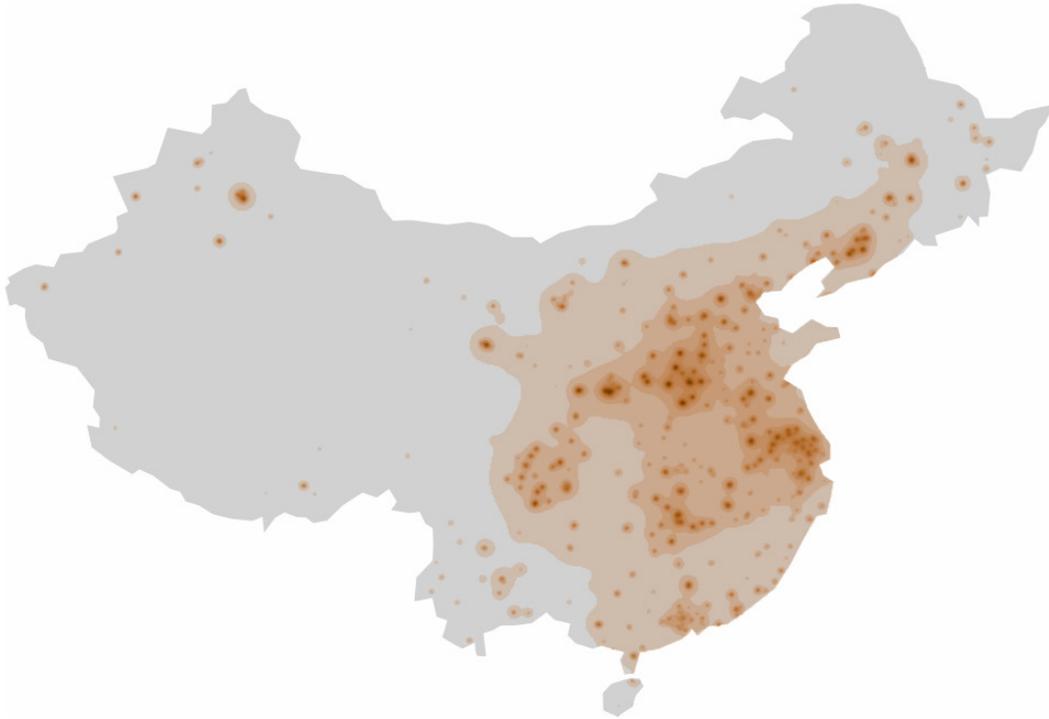


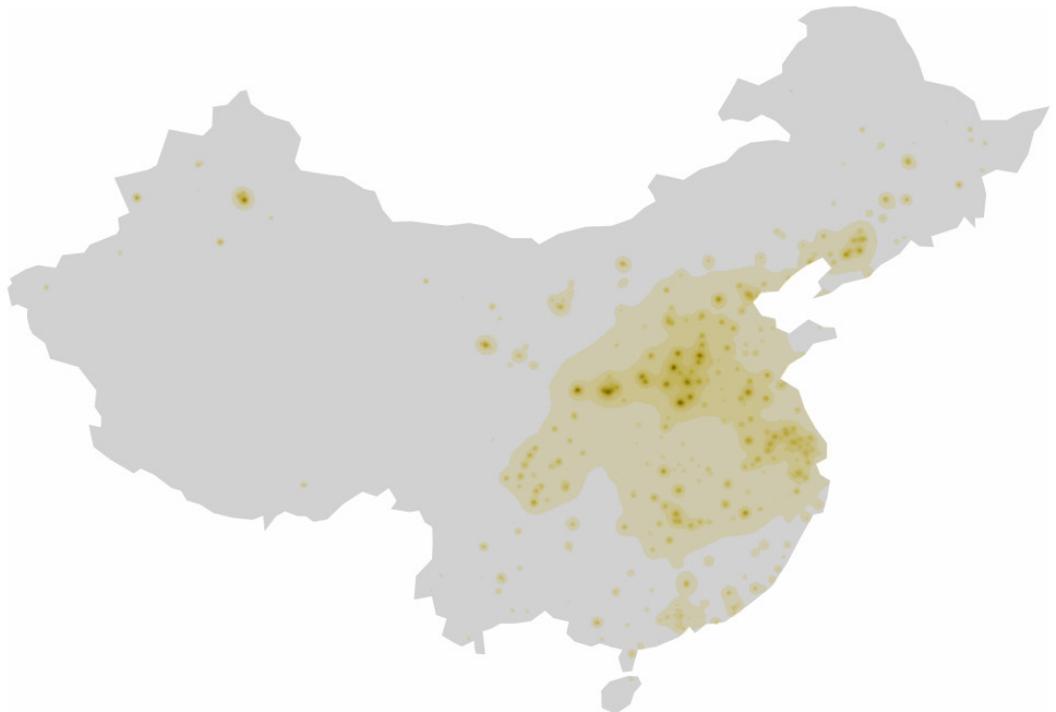


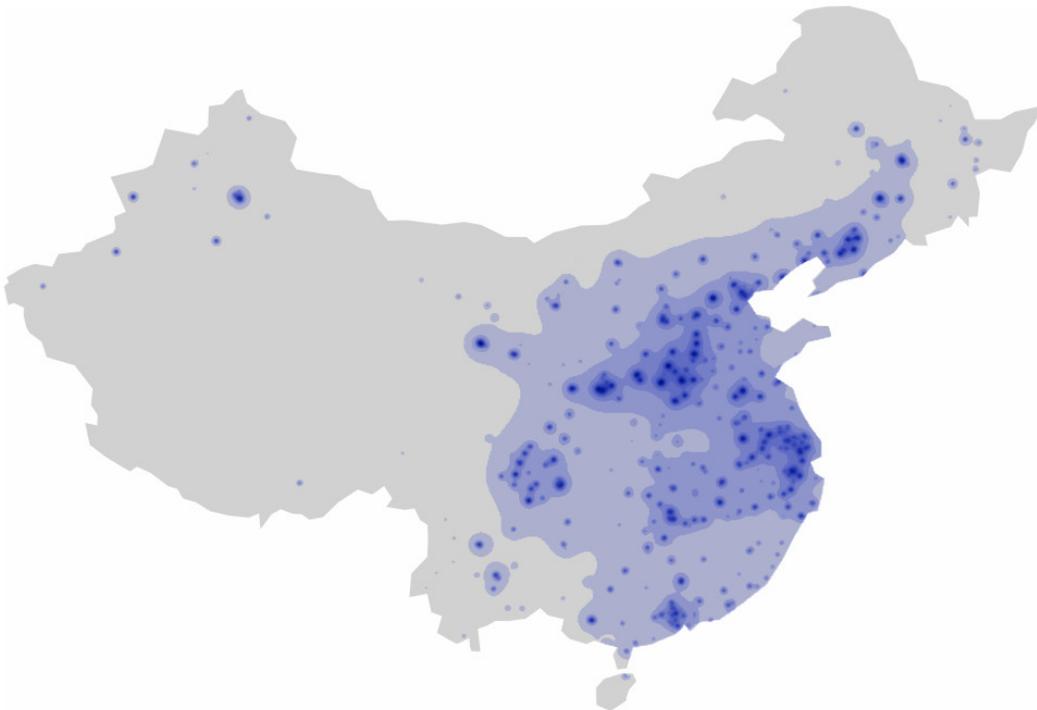
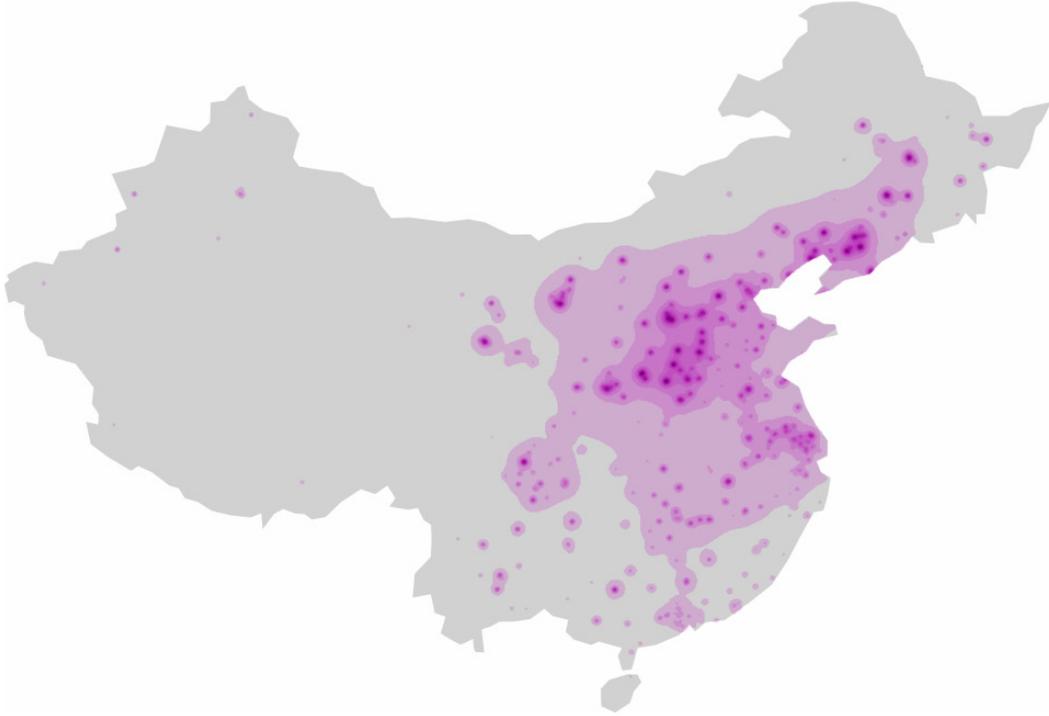


7.3 Visualização da poluição

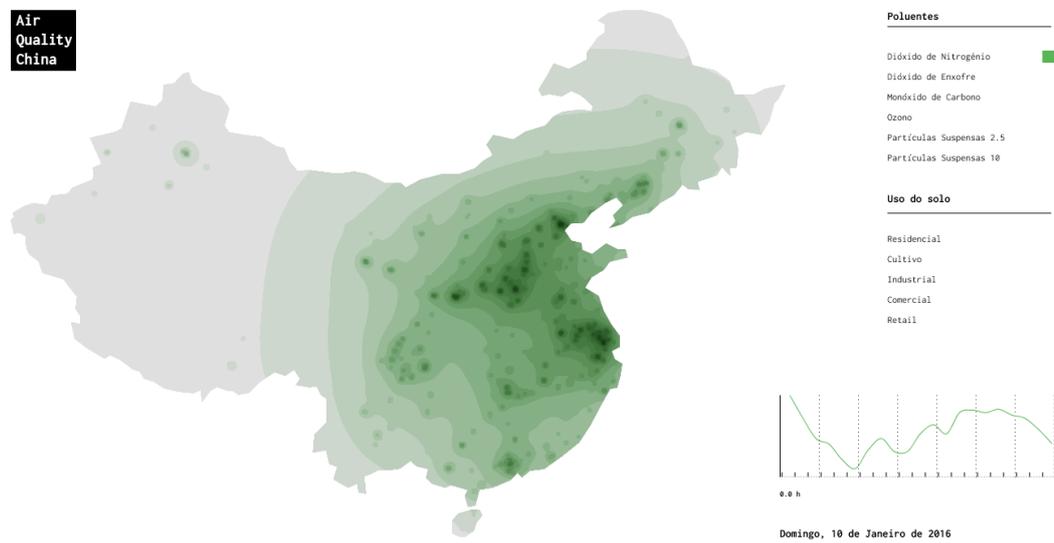
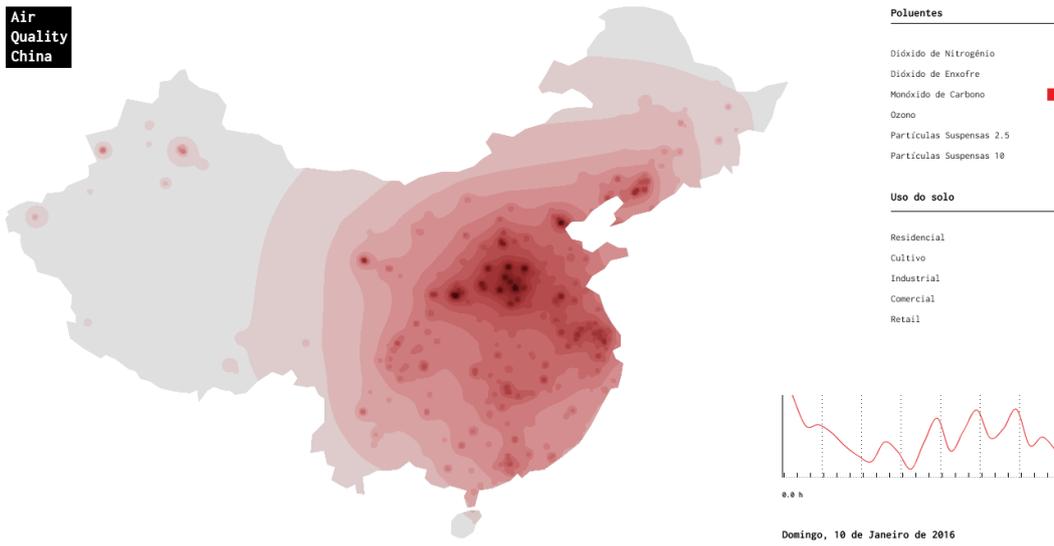
Mapas isométricos



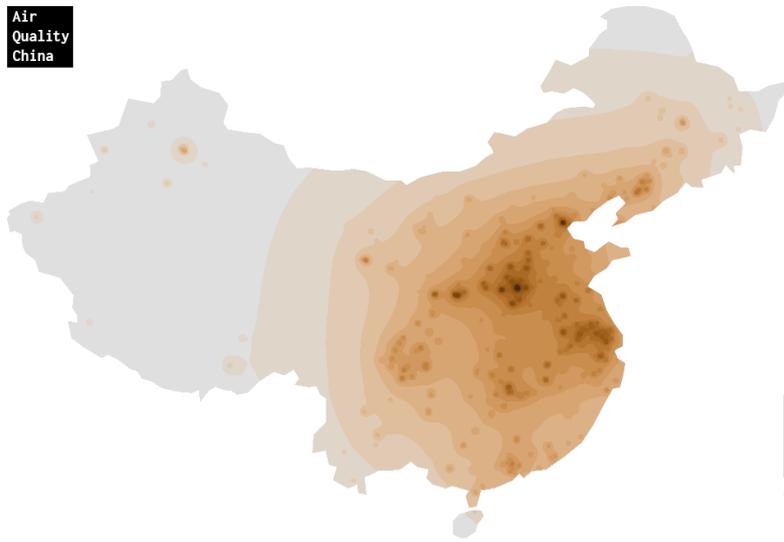




7.4 Aplicação final



Air Quality China

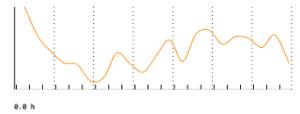


Pouentes

- Dióxido de Nitrogénio
- Dióxido de Enxofre
- Monóxido de Carbono
- Ozono
- Partículas Suspensas 2.5
- Partículas Suspensas 10

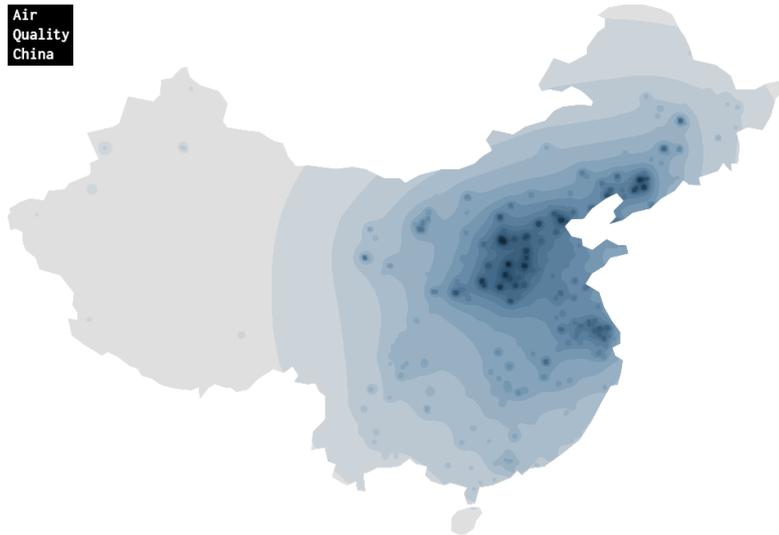
Uso do solo

- Residencial
- Cultivo
- Industrial
- Comercial
- Retail



Domingo, 10 de Janeiro de 2016

Air Quality China

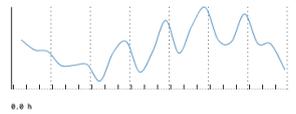


Pouentes

- Dióxido de Nitrogénio
- Dióxido de Enxofre
- Monóxido de Carbono
- Ozono
- Partículas Suspensas 2.5
- Partículas Suspensas 10

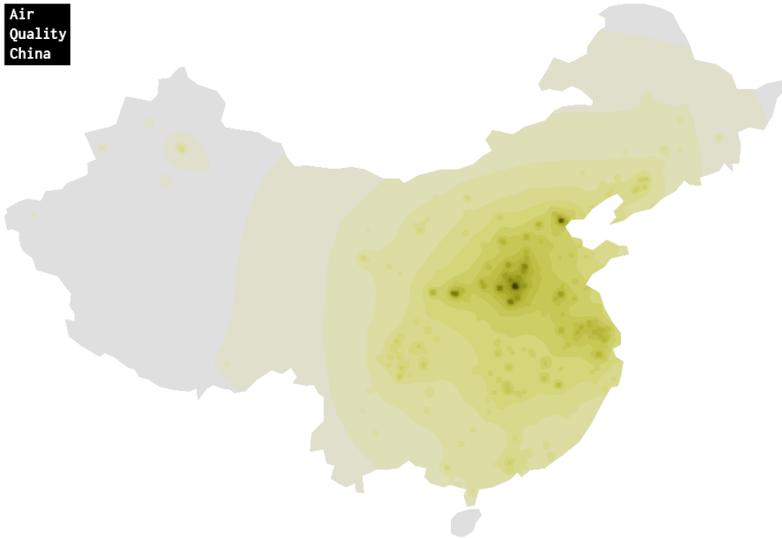
Uso do solo

- Residencial
- Cultivo
- Industrial
- Comercial
- Retail



Domingo, 10 de Janeiro de 2016

Air Quality China

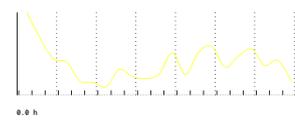


Poluentes

- Dióxido de Nitrogênio
- Dióxido de Enxofre
- Monóxido de Carbono
- Ozono
- Partículas Suspensas 2.5
- Partículas Suspensas 10

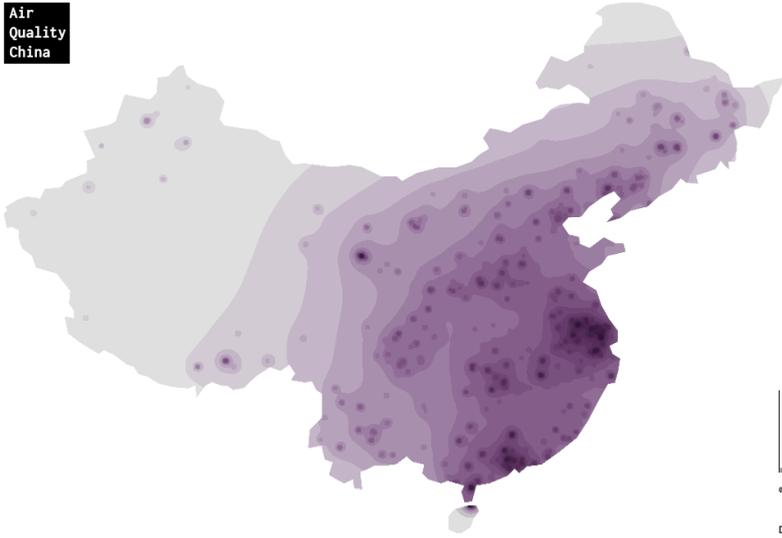
Uso do solo

- Residencial
- Cultivo
- Industrial
- Comercial
- Retail



Domingo, 10 de Janeiro de 2016

Air Quality China

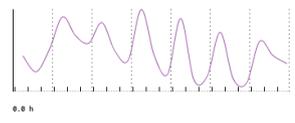


Poluentes

- Dióxido de Nitrogênio
- Dióxido de Enxofre
- Monóxido de Carbono
- Ozono
- Partículas Suspensas 2.5
- Partículas Suspensas 10

Uso do solo

- Residencial
- Cultivo
- Industrial
- Comercial
- Retail



Domingo, 10 de Janeiro de 2016

