

**Universidade de Trás-os-Montes e Alto Douro**

**Mestrado em Tecnologia Multimédia**

Visualização de dados

# **Padrões Emergentes entre a Crise Habitacional e a População Sem-Abrigo em Portugal (2018-2023)**

*Projeto final interativo*

# Introdução

Este relatório é a terceira e última parte do projeto de Visualização de Dados direcionado para os Padrões Emergentes entre a crise habitacional em Portugal e a evolução da população em situação de sem-abrigo entre 2018 e 2023, cruzando vários dados disponíveis.

No primeiro relatório, a exploração inicial permitiu recolher e preparar os dados, construir visualizações exploratórias e identificar padrões e limitações. No segundo relatório, conseguimos transformar as ideias num protótipo mais estático de “dashboard”, recorrendo ao Tableau Public.

Neste terceiro relatório, apresentamos a materialização final do projeto numa interface web dinâmica e totalmente interativa. Abandonando as restrições das ferramentas de prototipagem. A utilização de código permitiu refinar a estrutura narrativa, adotando um modelo híbrido que guia o utilizador pelo contexto nacional antes de libertar a exploração territorial. Detalha-se aqui o processo de design centrado no utilizador, as soluções técnicas adotadas para garantir a portabilidade da visualização e a validação das escolhas visuais face aos princípios teóricos de percepção e comunicação de dados.

De forma a criar uma visualização final e interativa tivemos recurso a D3.js, focado na análise da crise habitacional em Portugal e a sua correlação com o aumento da população sem-abrigo. O objetivo central foi transformar dados brutos e complexos em informação visionável (Tufte, 1990), permitindo ao utilizador compreender o paradoxo entre a evolução de vários fatores e o aumento da exclusão social.

O painel interativo intitulado de Padrões Emergentes entre a Crise Habitacional e a População Sem-Abrigo está disponível online. O repositório completo do projeto, incluindo código e dados, encontra-se disponível no [GitHub](#) e o projeto final [aqui](#).

## Estrutura Narrativa

De forma a criar uma visualização que equilibre a necessidade de uma mensagem clara com a liberdade de exploração por parte do leitor, a estrutura do projeto fundamentou-se na taxonomia de Segel e Heer (2010). Estes autores identificam um espectro de interatividade entre duas abordagens opostas: Author-Driven e Reader-Driven.

Para equilibrar esses extremos optamos por adotar a estrutura "Martini Glass". Esta estrutura utiliza o formato de um copo de martini como metáfora para o fluxo de interação. Caracteriza-se por iniciar com uma abordagem fortemente guiada pelo autor (Author-Driven), onde se apresenta o contexto nacional da vulnerabilidade social e o paradoxo económico (Gráficos 1.1 e 1.2), assemelhando-se ao pé do copo de martini. Após esta contextualização, a narrativa abre-se para uma exploração livre (Reader-Driven), correspondente ao corpo do copo, onde o utilizador interage com os restantes gráficos interativos (Gráficos 1.3, 2.1, 2.2 e 3) para extrair as suas próprias conclusões focadas no território nacional.

Esta escolha demarca-se das outras estruturas identificadas pelos autores, nomeadamente o 'Interactive Slideshow', que seria demasiado linear e restritivo para a exploração

geográfica, e o 'Drill-Down Story', que, ao oferecer liberdade total desde o início, poderia falhar na comunicação da mensagem central sobre a inflação e a crise estrutural.

## Elementos Narrativos

Em continuidade com a estratégia delineada nas fases anteriores do projeto, implementámos mecanismos de interação desenhados para reforçar a mensagem da narrativa.

Privilegiamos interações simples, mas eficazes, destacando-se a técnica de Linked Highlighting. Esta funcionalidade permite que, ao selecionar um distrito, este seja realçado simultaneamente em todas as visualizações transversais. Tal mecanismo facilita comparações multidimensionais imediatas, essenciais para a percepção de padrões complexos (Few, 2006).

A implementação técnica, através de código, permitiu refinar a interação e simplificar a análise, unificando as escalas de preços de "compra" e "renda" num único gráfico 2.1. A alternância entre estas variáveis recorre a transições suaves (Bostock et al., 2011), o que assegura a consistência visual e permite ao utilizador alterar os dados sem perder o contexto cognitivo (Ware, 2012).

Por fim, uma evolução crucial nesta última fase foi a integração de textos narrativos dinâmicos junto de cada gráfico. Estes funcionam como uma legenda narrativa que transcende a mera descrição dos dados (Knaflic, 2015), oferecendo também a sua interpretação explicativa.

## Criação automática ou ao uso de LLMs

O desenvolvimento deste projeto enfrentou um desafio técnico significativo: a elevada complexidade da biblioteca D3.js e a curva de aprendizagem acentuada da sua sintaxe, contrastando com a proficiência de programação limitada da equipa (Bostock et al, 2011).

Para mitigar este desnível e viabilizar a implementação de uma visualização interativa avançada, adotamos uma metodologia de desenvolvimento assistido por Inteligência Artificial, recorrendo aos modelos Gemini (Google), ChatGPT (OpenAI) e Claude Sonet 4.5 (Anthropic's). A utilização destas ferramentas não foi passiva, mas sim integrada num fluxo de trabalho iterativo.

À medida que se delineavam os requisitos funcionais e a componente visual, os LLMs atuaram como técnicos, convertendo essas intenções de design num código funcional. Isto permitiu-nos ultrapassar a barreira de entrada técnica descrita por Bostock et al. (2011) sobre a gramática complexa do D3.

A geração automática de código não produziu resultados definitivos nas primeiras iterações. Este processo revelou-se incremental, exigindo o desenvolvimento de competências de leitura e análise estrutural para a identificação de erros. Foi necessário iterar os pedidos

(*prompts*) aos modelos de linguagem para obter correções precisas e, subsequentemente, consolidar os diversos fragmentos de código num ficheiro único e coeso.

Ao delegar a escrita da sintaxe à Inteligência Artificial, conseguimos concentrar os nossos esforços cognitivos na validação do design. Isto garantiu que as escolhas visuais fossem fiéis aos princípios estudados, em vez de serem limitadas pelo que sabíamos ou não programar.

Os Modelos de Linguagem (LLMs) funcionaram como um acelerador técnico indispensável, permitindo que um grupo com limitações de programação produzisse um artefacto final de complexidade elevada, mantendo o controlo autoral sobre a narrativa e a integridade dos dados.

## **Limpeza, transformação e adaptação inteligente dos dados com apoio de IA**

Ao longo do desenvolvimento do projeto, a Inteligência Artificial foi utilizada como ferramenta central de apoio, atuando não apenas na limpeza dos dados, mas sobretudo na sua transformação e adaptação dinâmica às necessidades analíticas do trabalho. Dando continuidade à metodologia aplicada nos trabalhos anteriores, o tratamento dos dados não foi todo manual, havendo automatizações através de scripts gerados com o auxílio de IA, que permitiram harmonizar fontes distintas e calcular novas variáveis, num processo integralmente documentado, incluindo ajustamentos manuais pontuais sempre que foram detectadas incoerências.

Este comportamento evidencia a capacidade da IA de ajustar a representação da informação em função do contexto, contribuindo para uma solução tecnicamente eficiente. A existência de dados organizados de formas distintas não constitui um erro, mas sim um requisito técnico fundamental: diferentes visualizações implicam diferentes estruturas de dados.

```
//visu 1.3 small multiples
const inequalityData = [
  { distrito: "Viana do Castelo", values: [
      {year:2018,value:7.9}, {year:2019,value:7.8}, {year:2020,value:7.7},
      {year:2021,value:7.7}, {year:2022,value:7.7}, {year:2023,value:7.8}
    ],
  },
  { distrito: "Braga", values: [
      {year:2018,value:8.3}, {year:2019,value:8.1}, {year:2020,value:8.5},
      {year:2021,value:8.4}, {year:2022,value:8.6}, {year:2023,value:8.6}
    ],
  },
  { distrito: "Porto", values: [
      {year:2018,value:12.6}, {year:2019,value:12.1}, {year:2020,value:12.8},
      {year:2021,value:12.8}, {year:2022,value:13.0}, {year:2023,value:13.1}
    ],
  },
  { distrito: "Vila Real", values: [
      {year:2018,value:9.9}, {year:2019,value:9.2}, {year:2020,value:9.5},
      {year:2021,value:8.9}, {year:2022,value:9.0}, {year:2023,value:9.2}
    ],
  },
  { distrito: "Bragança", values: [
      {year:2018,value:8.8}, {year:2019,value:9.2}, {year:2020,value:9.2},
      {year:2021,value:9.3}, {year:2022,value:9.0}, {year:2023,value:9.1}
    ],
  }
];
```

**Figura 1:** Organização de dados feita por AI para Small multiples. Originalmente escrito em CSV.

**Figura 2:** Organização de dados feita por AI para Dumbbell charts, Scatter plots e o “mapa” coroplético. Originalmente escrito em CVS

```
//visu 1.1
const data1_1 = [
  {year: 2018, idosos: 2273940, pobreza: 2222100, desemprego: 713090},
  {year: 2019, idosos: 2327150, pobreza: 2173000, desemprego: 735166},
  {year: 2020, idosos: 2385368, pobreza: 2056200, desemprego: 737324},
  {year: 2021, idosos: 2436949, pobreza: 2311500, desemprego: 645278},
  {year: 2022, idosos: 2486274, pobreza: 2083600, desemprego: 732821},
  {year: 2023, idosos: 2537740, pobreza: 2103600, desemprego: 708738}
];
```

**Figura 3:** Organização de dados feita por AI para gráfico de linhas sem análise geográfica. Originalmente escrita em CSV

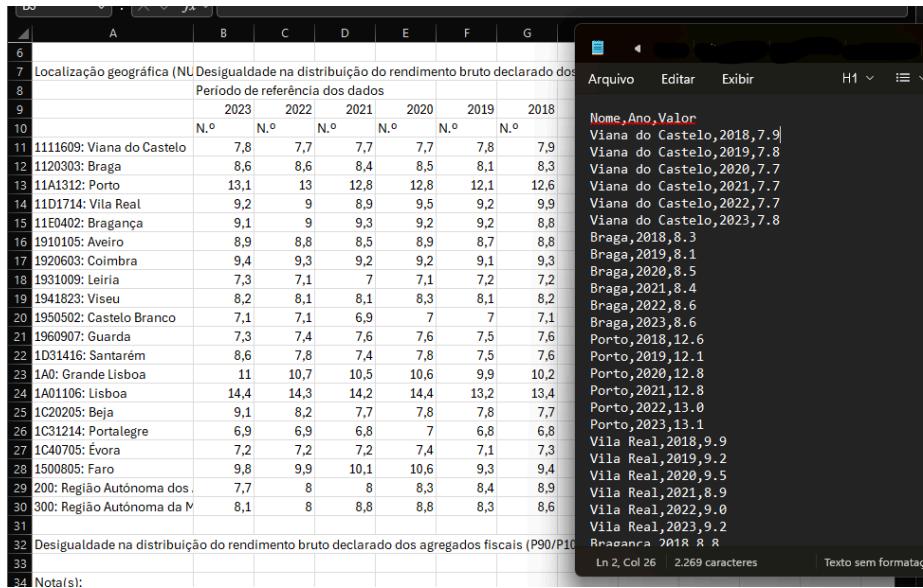
Durante este processo, a organização dos dados sofreu alterações face ao plano inicial. Estas mudanças resultaram de adaptações automáticas propostas pela IA, que identificou a necessidade de reestruturar os dados para satisfazer os requisitos técnicos específicos de cada visualização em D3.js. Deste modo, os dados foram reorganizados em diferentes formatos consoante a tarefa analítica (Figuras 1, 2 e 3):

- **Entidade-Linha (Wide Format):** Para visualizações como *scatter plots*, mapas coropléticos e *dumbbell charts*, os dados foram organizados de modo a que cada linha correspondesse a um distrito, agregando todos os indicadores relevantes para permitir comparações pontuais entre anos (ex: 2018 vs 2023).
  - **Séries Temporais/Small Multiples (Long Format):** Outras visualizações exigiram estruturas orientadas para a evolução temporal, onde a granularidade dos dados foi ajustada para alimentar múltiplos gráficos simultâneos.

|   |    |   |   |
|---|----|---|---|
|   | Aa | Quadro extraído em 02 de Dezembro de 2025 (19:14:04)  | ↓ |
| 1 |    | <a href="http://www.ine.pt">http://www.ine.pt</a>   |   |
| 2 |    | Localização geográfica (NUTS - 2024);Desigualdade na distribuição do rendimento bruto declarado dos agregados fiscais (P90/P10 - N.º) |   |
| 3 |    | ;Período de referência dos dados;;;;;   |   |
| 4 |    | ;2023;2022;2021;2020;2019;2018;   |   |
| 5 |    | ;N.º ;N.º ;N.º ;N.º ;N.º ;N.º ;   |   |
| 6 |    | 1111609: Viana do Castelo;7   |   |
| 7 |    | 1120303: Braga;8  |   |
| 8 |    | 11A1312: Porto;13   |   |

**Figura 4:** Leitura do ficheiro em excel “Desigualdade na distribuição do rendimento bruto declarado

dos agregados fiscais 2018-2023” por Rawgraphs.



The image shows a side-by-side comparison of data processing. On the left, a Microsoft Excel spreadsheet titled "Localização geográfica (NU) Desigualdade na distribuição do rendimento bruto declarado dos agregados fiscais 2018-2023" is displayed. It contains a table with columns for years 2018 through 2023 and various regions. On the right, a screenshot of the RAWGraphs interface shows the same data converted into a CSV file, with the header "Nome,Ano,Valor" and rows of data such as "Viana do Castelo,2018,7.9". A status bar at the bottom of the RAWGraphs window indicates "Linha 2, Coluna 26 | 2.269 caracteres" and "Texto sem formatação".

**Figura 5:** Exemplo de processo de harmonização e limpeza do ficheiro excel “Desigualdade na distribuição do rendimento bruto declarado dos agregados fiscais 2018-2023” (esquerda) para CSV(direita) manualmente.

O tratamento dos dados seguiu uma abordagem evolutiva e híbrida, iniciando-se na segunda fase do projeto com uma limpeza preliminar manual, essencial para converter ficheiros Excel em CSV e filtrar colunas incompatíveis com a ferramenta RAWGraphs (Figuras 4 e 5). Posteriormente, estes dados harmonizados foram submetidos a um *pipeline* de processamento assistido por IA, que permitiu corrigir inconsistências profundas, normalizar nomenclaturas e calcular variáveis derivadas (como taxas por 10.000 habitantes), encontrando-se todo o código documental destas transformações disponível no repositório [GitHub](#).

Outro exemplo prático do refinamento manual e estratégico foi a decisão de excluir as Regiões Autónomas (Açores e Madeira) do Gráfico 1.3. Embora existissem dados fiscais para estas regiões, a sua ausência nos datasets de habitação e sem-abrigo (fontes utilizadas nos restantes gráficos) criaria lacunas na análise comparativa transversal. A deteção desta discrepancia levou a um ajuste nos dados para garantir a coerência visual e analítica de todo o dashboard.

## Erros, Limitações e Problemas Identificados

O desenvolvimento deste projeto enfrentou constrangimentos estruturais fundamentais para a contextualização dos resultados. Como referenciado nas fases anteriores, a recolha de dados foi circunscrita ao intervalo 2018–2023. Esta restrição deveu-se à inexistência de datasets completos e transversais para todos os indicadores (sem-abrigo e imobiliário) fora

deste período. A extensão da análise temporal resultaria em lacunas graves (*missing values*), pelo que se optou por manter este intervalo fechado para garantir a integridade e a comparabilidade estatística.

Apesar da robustez das visualizações alcançadas, é fundamental reconhecer as "zonas de sombra" da análise. O projeto descreve com eficácia *onde* o problema é grave, mas os dados demográficos e imobiliários disponíveis são insuficientes para explicar cabalmente os determinantes de certos *outliers*, como o caso de Beja. Adicionalmente, a janela temporal limitada impede a distinção clara entre tendências estruturais (longo prazo) e picos conjunturais (pós-pandemia ou crise inflacionária), não sendo possível afirmar inequivocamente se a situação de 2023 representa um "novo normal" ou um pico isolado.

Reconhece-se, ainda, que a priorização da eficácia funcional e da lógica de dados levou a compromissos na interface, nomeadamente na responsividade visual. Em ecrãs de menor dimensão, a densidade de informação e a redução de elementos textuais podem prejudicar a legibilidade imediata, exigindo do utilizador um esforço de interpretação acrescido.

## Síntese Analítica e Conclusões

Nesta fase final, a integração de um gráfico de síntese (Gráfico 3) revelou-se crucial para uma leitura imediata do panorama nacional, permitindo desconstruir percepções baseadas apenas em números absolutos. A análise cruzada entre Pressão Imobiliária e Exclusão Social destacou o distrito de Beja como um caso crítico: apesar da baixa densidade populacional e menor pressão imobiliária, apresenta uma intensidade de sem-abrigo desproporcionalmente elevada. Em contrapartida, Lisboa, embora lidere nos números absolutos, segue uma tendência mais expectável face à sua demografia.

Esta síntese visual permitiu concluir que a crise tem naturezas distintas: no litoral, correlaciona-se fortemente com o mercado de habitação; no interior, operam fatores de exclusão social sistémica independentes do valor das rendas. A principal aprendizagem extraída reside, portanto, na identificação desta dualidade no fenómeno, dialogando diretamente com a literatura sobre exclusão social e demografia ao demonstrar que a exclusão habitacional opera segundo lógicas territoriais distintas.

O desenvolvimento deste projeto representou uma curva de aprendizagem significativa, dada a experiência limitada do grupo em programação web. A estratégia de iniciar o trabalho com prototipagem em ferramentas visuais como o Tableau, foi fundamental para estruturar o raciocínio lógico antes da transição para o ambiente de código em D3.js.

Apesar das dificuldades técnicas inerentes a esta biblioteca, os resultados foram satisfatórios. A utilização de Inteligência Artificial funcionou como um acelerador indispensável, auxiliando na superação de barreiras de sintaxe e na implementação de lógicas complexas (como os *Small Multiples*), tornando o processo de desenvolvimento mais acessível. Contudo, a implementação do código gerado pela IA não foi isenta de desafios, exigindo um esforço significativo de depuração para corrigir inconsistências em

escalas, tipografias e cores, garantindo a coerência final do artefacto.

Em suma, este projeto demonstra como a visualização de dados interativa pode transformar tabelas estáticas em narrativas visuais poderosas, capazes de expor crises invisíveis e informar a tomada de decisão social.

## Referências

- Bostock, M., Ogievetsky, V., & Heer, J. (2011). D3: Data-Driven Documents. *IEEE Transactions on Visualization and Computer Graphics*, 17(12), 2301–2309.
- Few, S. (2006). *Information dashboard design: The effective visual communication of data*. O'Reilly.
- Knafllic, C. N. (2015). *Storytelling with data: A data visualization guide for business professionals*. Wiley.
- Segel, E., & Heer, J. (2010). Narrative Visualization: Telling Stories with Data. *IEEE Transactions on Visualization and Computer Graphics*, 16(6), 1139–1148.
- Tufte, E. R. (1990). *Envisioning information*. Graphics Press.
- Ware, C. (2012). *Information visualization: Perception for design* (3.<sup>a</sup> ed.). Morgan Kaufmann.