

# Visualização de Informações em Bases Abertas Sobre Oncologia

**Leandro de Souza Junior<sup>1,2</sup>, Mariangela Dametto<sup>1</sup>, Rodrigo Bonacin<sup>1</sup>**

1217739@dac.unicamp.br, mdametto@cti.gov.br,  
rodrigo.bonacin@cti.gov.br

**<sup>1</sup> Divisão de Metodologias da Computação - DIMEC, Centro de Tecnologia da Informação Renato Archer – CTI**

**<sup>2</sup> Universidade Estadual de Campinas - Unicamp**

***Abstract.*** *By the analysis of data produced by healthcare institutions, data science can significantly contribute to the advancement of research, public health, diagnosis, and medical treatments. Among the key data science techniques are information visualization techniques, which are useful to support domain experts understand the processed information. In this context, this paper presents a study on the application of information visualization techniques using open databases in the field of oncology. Results with an open dataset from the National Cancer Institute and Oncocentro Foundation of São Paulo are presented and discussed in this work.*

***Resumo.*** *Por meio da análise de dados produzidos pelas instituições de saúde, a ciência de dados pode contribuir significativamente para o avanço da pesquisa, para a saúde pública, para o diagnóstico e para os tratamentos médicos. Entre as principais técnicas de ciência de dados estão as técnicas de visualização de informações, que são úteis para ajudar especialistas do domínio a entender as informações processadas. Neste contexto, este artigo apresenta um estudo sobre a aplicação de técnicas de visualização de informações utilizando bancos de dados abertos na área de oncologia. Resultados com conjunto de dados abertos do Instituto Nacional de câncer e da Fundação Oncocentro de São Paulo são apresentados e discutidos neste trabalho.*

## 1. Introdução

A visualização de dados de saúde é crucial para a compreensão e o tratamento de diversas condições de saúde, especialmente em contextos complexos como o câncer. Com a crescente disponibilidade de grandes volumes de dados, surge a necessidade de ferramentas eficazes para a visualização e interpretação dessas informações [Costa, 2017]. O câncer, com suas múltiplas variáveis e complexidade, representa um desafio significativo para os profissionais de saúde [Clarify, 2021]. Fatores como o abuso de

bebidas alcoólicas são exemplos de fatores que podem influenciar, por exemplo, no diagnóstico, tratamento dos pacientes e recidiva de câncer [Silva, 2018].

A importância da visualização de dados na área da oncologia não pode ser subestimada. Gráficos e *dashboards* bem projetados permitem que médicos e pesquisadores identifiquem padrões e tendências que podem não ser evidentes em tabelas de dados brutos. Este artigo aborda a construção de um *dashboard* interativo utilizando o Power BI<sup>1</sup>, uma ferramenta poderosa para a análise e visualização de dados. O objetivo principal é melhorar a compreensão dos dados sobre diagnóstico, tratamento e estado de pacientes com câncer, correlacionando variáveis como tabagismo e alcoolismo e proporcionando uma análise mais profunda por meio de gráficos e tabelas interativas.

Neste estudo, a ênfase está na integração de diferentes bases de dados para criar um sistema robusto que permita a análise detalhada do estado de pacientes com câncer. O *dashboard* desenvolvido oferece várias visões das informações, incluindo a comparação entre tratamentos e variáveis associadas, e usa gráficos interativos para facilitar a interpretação dos dados. Este artigo examina como o Power BI<sup>1</sup> foi utilizado para construir essa ferramenta, detalha a metodologia empregada e discute as inovações e desafios encontrados ao longo do processo.

## 2. Metodologia e Desenvolvimento do Protótipo

A metodologia deste estudo envolve o desenvolvimento e a implementação do protótipo do *Open Oncology Dashboard* (OODash), um sistema interativo para a visualização de dados oncológicos, com foco específico na análise de fatores ligados ao tabagismo e alcoolismo. O processo foi dividido em várias etapas, cada uma crucial para garantir a eficácia e a utilidade do *dashboard*. A seguir, são descritas cada etapa detalhadamente, incluindo as imagens do *dashboard* que ilustram o trabalho realizado.

### 2.1. Desenvolvimento do OODash para Visualização de Dados Oncológicos

O desenvolvimento do protótipo do OODash envolveu a integração de dados de múltiplas fontes e a criação de uma interface interativa e clara. Inicialmente foram utilizados dados abertos do Instituto Nacional de Câncer (INCA)<sup>2</sup> e da Fundação Oncocentro de São Paulo (FOSP)<sup>3</sup>. Utilizando a plataforma Power BI<sup>1</sup>, a primeira etapa foi a padronização e organização dos dados. Foram combinadas uma base de dados da FOSP, explorada no último período de Iniciação Científica, com a base do INCA na continuidade. A visualização conjunta das bases de dados permitiu relacionar variáveis importantes, como recidivas (SBCO, 2021), tabagismo e alcoolismo, que estavam inicialmente em tabelas diferentes.

A criação do OODash foi orientada para fornecer uma visão abrangente e acessível dos dados oncológicos. As Figuras 1 a 9 apresentam interfaces do *dashboard* proposto. A Figura 1 ilustra a página inicial do OODash, acessível por meio de um ícone de “casinha” localizado no canto superior direito da interface. Esta página serve

<sup>1</sup> <https://www.microsoft.com/pt-br/power-platform/products/power-bi>

<sup>2</sup> <https://www.gov.br/inca/pt-br>

<sup>3</sup> <https://fosp.saude.sp.gov.br/>

como o ponto de partida para a exploração dos dados e exibe informações críticas sobre o perfil demográfico dos pacientes oncológicos, com foco em aspectos relacionados ao alcoolismo e tabagismo. A tela é dividida em seções distintas, onde gráficos e tabelas são utilizados para apresentar dados categorizados por histórico de consumo de álcool, tabagismo, idade e sexo dos pacientes.

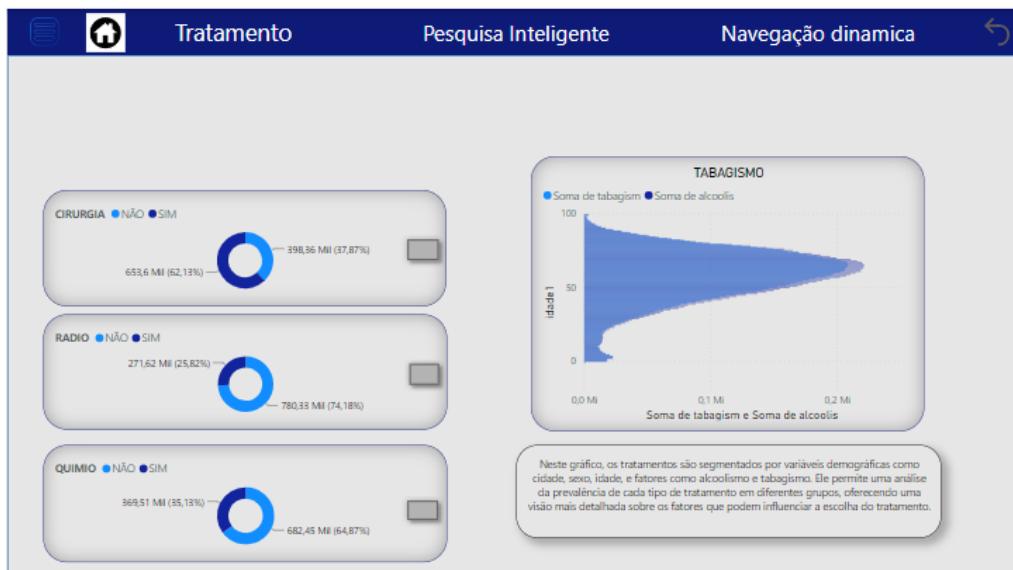
Os dados são apresentados com o objetivo de serem visualmente claros e acessíveis, permitindo que os usuários identifiquem padrões e relações entre os fatores de risco e as características dos pacientes. A inclusão dessas variáveis é essencial, pois o alcoolismo e o tabagismo são conhecidos fatores de risco para diversos tipos de câncer, e sua análise ajuda a entender melhor o perfil de risco dos pacientes. A estrutura da página é projetada para facilitar a navegação e a interpretação dos dados, visando fornecer uma visão geral comprehensível e útil para a tomada de decisões clínicas.



**Figura 1.** Página Inicial do OODash

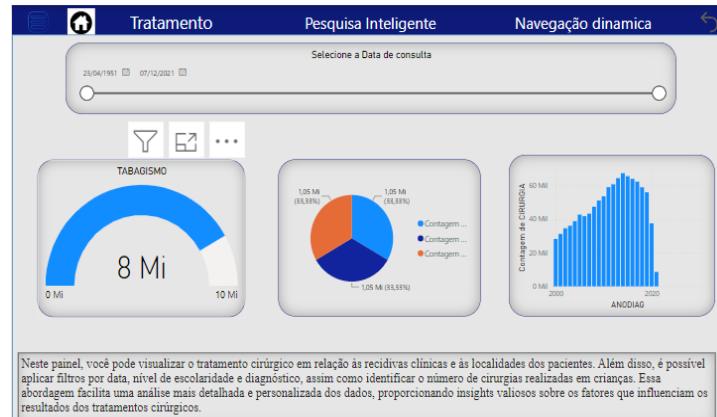
A Figura 2 apresenta uma visão dos tratamentos realizados, com foco nos tipos de tratamentos oncológicos aplicados aos pacientes. Esta imagem inclui três gráficos principais, cada um fornecendo uma perspectiva única sobre os dados de tratamento. O primeiro gráfico é um gráfico de barras, que mostra a quantidade total de tratamentos realizados, segmentados em cirurgia, radioterapia e quimioterapia. Esse gráfico é crucial para entender a distribuição dos diferentes tipos de tratamento e identificar quais são mais ou menos comuns entre os pacientes. Além disso, o gráfico destaca a faixa etária dos pacientes, revelando que a maior concentração de tabagistas e alcoólatras está entre 50 e 60 anos, o que é um dado relevante para a análise do impacto desses fatores de risco sobre o tratamento.

O OODash visa fornecer uma visualização clara e concisa das informações e assim permitir aos profissionais da saúde identificar rapidamente as tendências e padrões no tratamento dos pacientes e avaliar as necessidades de suporte e intervenção. Esse conjunto de gráficos fornece uma visão da distribuição e frequência dos tratamentos, com o objetivo de auxiliar na tomada de decisões informadas e na identificação de áreas que podem necessitar de mais atenção ou investigação adicional.

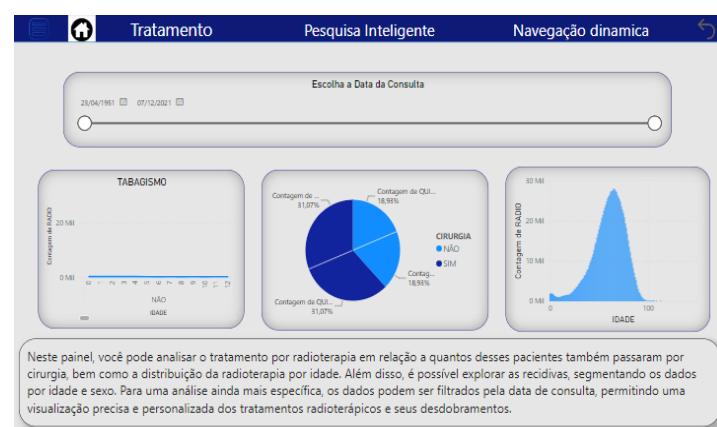


**Figura 2.** Interface com visão sobre dos tratamentos realizados

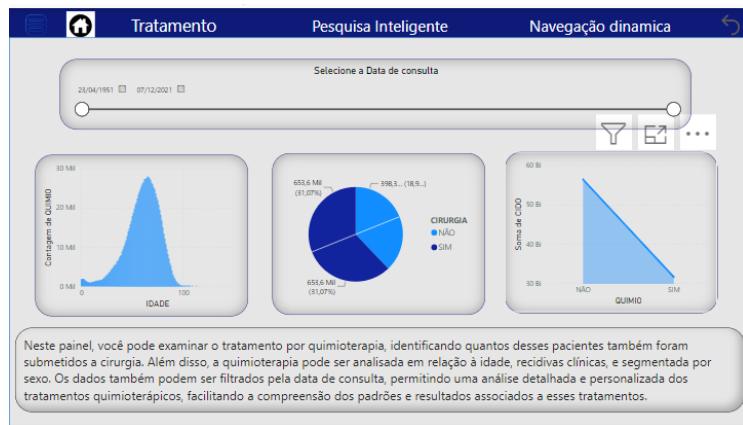
As Figura 3, 4 e 5 do OODash oferecem uma visão dos tratamentos oncológicos realizados, com foco na cirurgia, radioterapia e quimioterapia, respectivamente. Cada uma dessas figuras fornecem uma perspectiva sobre esses tratamentos, além de como eles se relacionam com o histórico de tabagismo e alcoolismo dos pacientes.



**Figura 3.** Visão sobre tratamentos oncológicos realizados com foco na cirurgia



**Figura 4.** Visão sobre tratamentos oncológicos realizados com foco na radioterapia



**Figura 5.** Visão sobre tratamentos oncológicos realizados com foco na quimioterapia

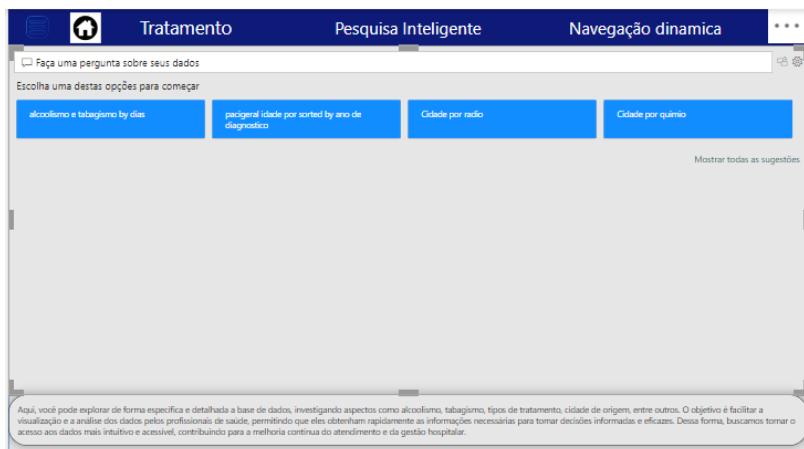
A Figura 3 oferece um botão dentro do gráfico que fornece acesso a dados adicionais sobre os tratamentos cirúrgicos. Esse oferece uma ferramenta que permite aos usuários explorar informações relacionadas às cirurgias realizadas. Ao interagir com essa funcionalidade, o usuário é direcionado a uma nova camada de dados, que inclui aspectos importantes como a data da consulta, o sexo e a idade dos pacientes, além de fatores relacionados ao tabagismo e alcoolismo. Esse nível de detalhamento possibilita uma análise dos impactos das variáveis individuais na necessidade e eficácia dos procedimentos cirúrgicos. O gráfico também apresenta uma comparação entre cirurgia, radioterapia e quimioterapia, evidenciando que a cirurgia é o tratamento mais frequente, enquanto a radioterapia é menos comum. Esta comparação visual visa apoiar o entendimento da prevalência relativa dos tratamentos e como diferentes fatores de risco podem influenciar a escolha e a frequência dos procedimentos cirúrgicos.

As visualizações revelam que a radioterapia (Figura 4) é menos comum em comparação com a cirurgia (Figura 3), sugerindo um papel complementar na maioria dos casos. Os dados indicam que a radioterapia é frequentemente usada para tratar cânceres como o de mama e próstata, onde a radiação pode ser eficaz na erradicação de células cancerosas residuais após a cirurgia. A visualização também aponta que a radioterapia pode ser aplicada em casos onde a cirurgia não é viável, seja por razões de saúde do paciente ou pela localização do tumor. Além disso, o gráfico de barras na Figura 4 mostra que a radioterapia é mais comum em pacientes mais jovens e em casos onde o câncer é detectado em estágios avançados.

A Figura 5 oferece uma visão sobre a quimioterapia, destacando sua frequência e eficácia em diversos tipos de câncer. A quimioterapia é um tratamento sistêmico crucial para combater cânceres que se disseminaram além do local original, utilizando medicamentos para destruir células cancerosas em todo o corpo. A visualização ilustra a aplicação da quimioterapia em diferentes tipos de câncer, mostrando como a frequência e a eficácia variam conforme a doença. Além disso, a interface compara a eficácia da quimioterapia com outros tratamentos, como cirurgia e radioterapia, fornecendo uma perspectiva sobre seu papel no tratamento oncológico.

A Figura 6 mostra um filtro de pesquisa que permite ao usuário configurar perguntas específicas e obter resultados em forma de gráficos ou tabelas. Esta funcionalidade aumenta a flexibilidade do OODash, permitindo consultas

personalizadas. A pesquisa iterativa é um componente fundamental na criação e aprimoramento de visualizações de dados em ferramentas como o Power BI®.



**Figura 6.** Filtro de pesquisa para perguntas específicas

Este processo iterativo (Figura 6) envolve a criação inicial de representações visuais, como gráficos, tabelas e painéis, com base em requisitos e objetivos específicos. Esses primeiros designs visuais são desenvolvidos para responder a perguntas-chave e apresentar dados de maneira clara e informativa. Nesse contexto, o processo iterativo envolve a criação e refinamento contínuo das visualizações com base nas interações e nas necessidades do usuário final. O aplicativo é projetado para se adaptar e evoluir conforme o uso real, permitindo ao próprio usuário ajustar e personalizar a visualização dos dados para melhor atender às suas necessidades específicas.

A interação direta do usuário com a ferramenta iterativa (Figura 6) permite uma abordagem mais personalizada para a análise de dados. Por exemplo, o usuário pode ajustar visualizações para destacar informações específicas, como a frequência de tratamentos em diferentes categorias de câncer, ou detalhar a eficácia dos tratamentos com base em novas variáveis de interesse. Ao utilizar essa abordagem, o OODash não apenas responde às necessidades atuais do usuário, mas também se prepara para futuros ajustes e melhorias. Isso visa assegurar que as visualizações de dados permaneçam relevantes e eficazes, apoiando a análise contínua e a tomada de decisões estratégicas com base nas informações mais atualizadas e pertinentes.

O OODash também faz uso do ArcGIS Maps<sup>4</sup>, uma ferramenta de mapeamento geográfico que permite uma representação visual interativa dos dados em um contexto espacial. Essa tecnologia permite a criação de mapas interativos que ilustram a frequência e ocorrência de eventos de interesse, tais como das recidivas. Conforme ilustra a Figura 7, com a utilização do ArcGIS Maps é possível mapear as áreas com maiores índices de recidiva após diferentes tipos de tratamento, incluindo a radioterapia. Também é possível, por exemplo, observar como a presença de fatores de risco, como tabagismo e alcoolismo, pode influenciar a taxa de recidiva em diferentes localidades.

<sup>4</sup><https://www.esri.com/pt-br/arcgis/geospatial-platform/overview#:~:text=O%20que%20%C3%A9%20ArcGIS?,vantagem%20na%20tomada%20de%20decis%C3%B5es>.

Além disso, a análise georreferenciada por revelar padrões que podem não ser evidentes em gráficos tradicionais, oferecendo apoio para futuras intervenções e pesquisas direcionadas.



**Figura 7.** Filtro de pesquisa para perguntas específicas

### 3. Discussão: limitações e aprimoramentos

A discussão aborda as implicações e possíveis melhorias no *dashboard* OODash, além de propor futuras atualizações e aprimoramentos.

Uma das principais propostas para melhorar o OODash é a implementação de atualizações em tempo real. Atualmente, o OODash utiliza dados estáticos que, embora úteis, podem rapidamente se tornar desatualizados. Integrar uma funcionalidade que permita a atualização automática dos dados proporcionará uma visualização mais precisa e relevante. Essa mudança garantirá que as informações exibidas estejam sempre alinhadas com as últimas atualizações dos conjuntos de dados disponíveis.

Outra proposta de aprimoramento é a criação de uma rede de perfil para usuários. Esta funcionalidade permitirá que os usuários se cadastrem e acessem informações personalizadas com base em suas necessidades e interesses. Com perfis individuais, os usuários poderão salvar configurações de visualização, preferências e consultas frequentes, tornando a ferramenta mais adaptável às suas necessidades específicas. A integração de redes sociais e opções de login pode ampliar o acesso e a personalização do OODash.

Também é proposto expandir o OODash com mais abas e visuais interativas com o objetivo de proporcionar uma análise mais detalhada e abrangente dos dados. A adição de novas abas pode incluir seções específicas para diferentes tipos de câncer, variáveis adicionais e análises comparativas. Os visuais interativos, como gráficos dinâmicos e tabelas rotativas, permitirão uma exploração mais aprofundada dos dados.

O aprimoramento com novos em gráficos e tabelas também é crucial para melhorar a interpretação dos dados. Gráficos interativos que permitem zoom e filtragem dinâmica podem oferecer uma análise mais profunda das informações. Tabelas com funcionalidades de classificação e filtragem também podem aumentar a eficiência na visualização e interpretação dos dados. Esses aprimoramentos visam tornar o dashboard mais intuitivo e funcional para os usuários.

Outra possibilidade é a integração com projetos correlacionados. Ao unir diferentes projetos que se complementam, será possível criar um ambiente propício para ampliar a riqueza das visualizações disponíveis. Um exemplo de integração seria com o trabalho apresentado em Crocco e outros (2023), que aborda clusterização de casos de recidiva de câncer de mama, e será possível aprimorar substancialmente a maneira como os dados são visualizados e interpretados.

#### 4. Conclusão

A criação e desenvolvimento do OODash visou dar apoio significativo para a análise e compreensão dos fatores associados ao câncer, tais como alcoolismo e tabagismo, bem como analisar a recidiva. Este projeto ilustrou como a integração e visualização eficaz dos dados podem proporcionar apoio a geração de *insights* sobre relações entre variáveis no tratamento e estudo do câncer.

O OODash permite uma análise dos dados sobre tratamentos e recidivas, destacando a importância da relação entre variáveis como tabagismo e alcoolismo. A combinação das bases da FOSP e INCA, juntamente com a padronização das variáveis, facilitou a visualização de tendências e padrões que podem influenciar o desenvolvimento, estado do paciente e recidiva de câncer. As funcionalidades interativas, como os gráficos dinâmicos e filtros de pesquisa tendem a aumentar a capacidade dos usuários de explorar e interpretar os dados de maneira mais eficaz.

#### Referências

- Clarify (2021). “Qual a importância da visualização de dados?”. URL: <https://clarify.com.br/blog/qual-a-importancia-da-visualizacao-de-dados/> (último acesso 09 de setembro de 2024).
- Costa, F. G. D. (2017). “Visualização de dados e sua importância na era do Big Data”. Trabalho de Conclusão de Curso de Bacharelado em Estatística, Universidade Federal do Rio Grande do Sul. URL: <https://lume.ufrgs.br/handle/10183/169109> (último acesso 09 de setembro de 2024).
- Crocco, P.F., Souza Jr, L., Dametto, M., Bonacin R. (2023) “Clusterização de Dados Abertos em Oncologia Usando Técnicas de Aprendizado de Máquina: um estudo preliminar sobre recidiva de câncer de mama”. Jornada de Iniciação Científica do CTI Renato Archer, XXV JICC – 2023. URL: <https://www.gov.br/cti/pt-br/publicacoes/producao-cientifica/jicc/xxv-jicc-2023> (último acesso 09 de setembro de 2024).
- SBCO, Sociedade Brasileira de Cirurgia Oncológica (2021). "Recidivas do câncer: entenda como acontece". URL: <https://sbc.org.br/recidivas-do-cancer-entenda-como-acontece/> (último acesso 09 de setembro de 2024).

Silva, J.A.G.S (2018). “Capa: É possível falar em cura?”. Revista Rede Câncer, 40.  
URL: <https://www.inca.gov.br/publicacoes/revistas/rede-cancer-no-40> (último acesso 09 de setembro de 2024).