

Visualização de Informações e Interação com Prontuários Eletrônicos Semânticos do Paciente

Leandro de Souza Junior^{1,2}, Pedro Ferreira Crocco^{1,2}, Mariangela Dametto¹,
Rodrigo Bonacin¹

l217739@dac.unicamp.br, pcrocco@cti.gov.br, mdametto@cti.gov.br,
rodrigo.bonacin@cti.gov.br

¹ Divisão de Metodologias da Computação - DIMEC, Centro de
Tecnologia da Informação Renato Archer – CTI

² Univesidade Estadual de Campinas - Unicamp

***Abstract.** Due to the large amount of data currently produced by health institutions, data science can significantly contribute to the advancement of research, diagnosis and medical treatments. Information visualization includes useful techniques frequently used in data science projects to support domain experts to make sense of the processed information. Thus, this article presents a study on the application of information visualization techniques using open databases in the field of oncology. Preliminary results with data from the database of Fundação Oncocentro de São Paulo are presented and discussed in this work.*

***Resumo.** Devido à grande quantidade de dados produzidos atualmente pelas instituições de saúde, a ciência de dados pode contribuir significativamente para o avanço das pesquisas, diagnósticos e tratamentos médicos. A visualização de informações inclui técnicas úteis frequentemente usadas em projetos de ciência de dados para ajudar especialistas do domínio a entender as informações processadas. Assim, este artigo apresenta um estudo sobre a aplicação de técnicas de visualização de informações em bancos de dados abertos na área de oncologia. Resultados preliminares com dados do banco de dados da Fundação Oncocentro de São Paulo são apresentados e discutidos neste trabalho.*

1. Introdução

A visualização de dados tem como objetivo facilitar a compreensão de informações complexas de maneira mais fácil e acessível. A forma como abordamos e apresentamos os dados que pretendemos utilizar é um aspecto crucial, pois ajudam os profissionais a tomar decisões acerca dos dados apresentados, entender acerca de quais fatores e comportamentos estão influenciando em determinado fator. Para garantir que a compreensão seja facilitada, é necessário utilizar métodos como tabelas, gráficos e a exploração de tecnologias interativas [Costa 2017]. No contexto médico, a visualização de informações assume uma importância ainda maior, uma vez que engloba a representação visual de dados clínicos, diagnósticos, resultados de testes, imagens

médicas e outras informações relevantes tanto para profissionais da saúde quanto para os próprios pacientes. Em Dowding et al. (2015), por exemplo, é apresentada uma revisão da literatura sobre o uso de *dashboards* de visualização na melhoria do tratamento dos pacientes.

No entanto, é importante destacar que mesmo com a evolução das tecnologias de visualização de informações, o desafio de torna-las eficaz e compreensível continua a ser uma questão presente em diversos cenários. A complexidade em apresentar as informações médicas como uma grande diversidade proveniente de diversas fontes, é um desafio que necessita ser ultrapassado para prover apoio eficiente aos profissionais da saúde. Além disso, as abordagens de visualização nem sempre conseguem lidar adequadamente com o volume e complexidade dos dados gerados por sistemas de saúde. Isso, por sua vez, pode levar a interpretações equivocadas, diagnósticos imprecisos e, em último caso, a decisões clínicas prejudicadas.

O principal objetivo deste trabalho é experimentar alternativas de visualização de informação para apoiar os profissionais do domínio na compreensão de diversos fatores e sua relação com a recidiva de câncer [SBCO 2021]. Isso será alcançado por meio da análise das recidivas, via técnicas de visualização, permitindo ao profissionais de saúde a compreensão das relações entre as intervenções médicas, características sociodemográficas, características dos tumores, tratamentos, entre outros atributos disponível em base de dados pública da Fundação Oncocentro de São Paulo (FOSP), com os desfechos dos pacientes. Ao considerar essa abordagem, espera-se contribuir para uma tomada de decisão mais informada e embasada por evidências, oferecendo *insights* valiosos para profissionais de saúde e pesquisadores no desenvolvimento de estratégias de tratamento mais eficazes.

O restante deste artigo está organizado da seguinte maneira: a seção 2 apresenta a metodologia adotada neste trabalho, a seção 3 apresenta o design de protótipo sobre ferramenta de visualização de informações para recidiva de câncer, por fim, a seção 4 discute os resultados, conclui o trabalho e apresenta os trabalhos futuros.

2. Metodologia

Neste trabalho as técnicas empregadas na visualização de informação foram segmentadas em duas etapas distintas: (1) organização de dados e (2) construção de visualizações. A primeira etapa envolve a organização dos dados fornecidos pela FOSP em uma tabela unificada. Inicialmente, foram verificadas possíveis faltas de padronização e ausência de dados. Isso demandou uma organização preliminar, com o objetivo de facilitar a construção dos gráficos em etapas posteriores.

A Tabela 1 apresenta exemplos de dados extraídos da base da FOSP. Esse pré-processamento é essencial para aplicativos de visualização criar representações confiáveis bem como para interpretação correta do significado de cada gráfico gerado. Por exemplo, ao lidar com o atributo "cidade", é fundamental esclarecer que se trata precisamente de uma localidade urbana, evitando quaisquer ambiguidades, bem como estabelecer que cidades estão relacionadas a um código IBGE específico, conforme ilustra a Tabela 1.

Tabela 1. Exemplos de dados pré-processados extraídos da base da FOSP

Column1	ESCOLARI	IDADE	SEXO	UFNASC	UFRESID	IBGE	CIDADE	CATEATEND	DTCONSULT	CLINICA
0	9	113	2	SP	SP	3550308	SÃO PAULO	2	quinta-feira, 16 de março de 2017	
1	9	109	2	BA	SP	3550308	SÃO PAULO	1	terça-feira, 9 de abril de 2019	
2	9	107	2	SP	SP	3550308	SÃO PAULO	9	segunda-feira, 15 de agosto de 2011	
3	2	106	2	MG	SP	3536604	PAULO DE FARIA	2	quarta-feira, 29 de julho de 2020	
4	2	106	2	MG	SP	3536604	PAULO DE FARIA	2	quarta-feira, 29 de julho de 2020	
5	2	106	1	SP	SP	3540804	POTIRENDABA	2	quarta-feira, 11 de março de 2020	
6	9	105	1	SI	SP	3518800	GUARULHOS	1	terça-feira, 1 de dezembro de 2020	
7	9	105	1	SI	SP	3518800	GUARULHOS	1	terça-feira, 1 de dezembro de 2020	
8	1	105	2	BA	SP	3529401	MAUA	2	quinta-feira, 4 de fevereiro de 2016	
9	9	105	1	SP	SP	3541406	PRESIDENTE PRUDENTE	2	terça-feira, 19 de janeiro de 2016	
10	9	105	2	PI	SP	3550308	SÃO PAULO	2	quarta-feira, 6 de julho de 2011	
11	1	105	2	SP	SP	3524303	JABOTICABAL	9	segunda-feira, 26 de março de 2007	
12	1	105	2	SP	SP	3546603	SANTA FE DO SUL	9	terça-feira, 13 de setembro de 2005	
13	2	104	2	SP	SP	3525904	JUNDIAI	2	segunda-feira, 14 de dezembro de 2020	
14	9	104	2	SP	SP	3550308	SÃO PAULO	2	quinta-feira, 10 de agosto de 2017	
15	1	104	1	CE	SP	3508108	BURITAMA	2	quarta-feira, 23 de julho de 2014	
16	2	104	1	MG	MG	3127107	FRUTAL	2	terça-feira, 13 de março de 2012	
17	3	104	1	SP	SP	3512605	CORONEL MACEDO	2	sexta-feira, 5 de novembro de 2010	
18	1	103	2	MS	GO	5213103	MINEIROS	2	quinta-feira, 17 de setembro de 2020	
19	9	103	2	BA	SP	3550308	SÃO PAULO	2	quarta-feira, 11 de julho de 2018	
20	2	103	1	SP	SP	3512605	CORONEL MACEDO	2	terça-feira, 8 de novembro de 2016	
21	1	103	1	BA	SP	3515806	FLORA RICA	2	segunda-feira, 4 de dezembro de 2017	

O objetivo da segunda etapa do processo é aproveitar as características que foram previamente organizadas na tabela anterior de modo a gerar boas visualizações. Essas características são utilizadas para a construção de gráficos através dos métodos de visualização selecionados. Com a estruturação adequada dos dados, torna-se viável criar uma grande variedade de elementos visuais, incluindo gráficos, tabelas, diagramas, mapas, infográficos e painéis. Esses elementos visuais são escolhidos para estar em sintonia com as informações contidas no banco de dados e de modo que façam sentido ao público escolhido (especialistas no domínio).

Os recursos visuais selecionados devem ser uma representação visual eficaz, destacando de forma clara e impactante o efeito das recidivas em diferentes grupos de indivíduos. Isso é particularmente relevante ao considerarmos os diversos tipos de câncer. O objetivo é fornecer uma visualização que seja facilmente compreensível e que permita uma análise intuitiva do impacto das recidivas em diferentes contextos. Essa abordagem visa proporcionar uma compreensão mais profunda da situação, permitindo uma análise mais informada. Por exemplo, esses gráficos podem facilitar a identificação de tendências, padrões ou correlações entre os tipos de câncer e as recorrências. Dessa maneira, a tomada de decisão pode ser enriquecida, baseada nas percepções visualizadas e interpretadas nos gráficos.

3. Ferramenta de Visualização de Informações sobre Recidiva de Câncer

Nesta seção primeiramente é apresentada a tecnologia e design da ferramenta proposta para visualização de informações sobre recidiva de câncer (subseção 3.1). Na sequência, são apresentadas as principais funcionalidades dessa ferramenta (subseção 3.2).

3.1. Tecnologia e principais decisões de design

Após analisar os dados, realizar o pré-processamento e compreender o significado de cada dado, a plataforma Power BI¹ foi escolhida para o desenvolvimento das visualizações. Esta ferramenta foi escolhida devido a sua capacidade de criar um design apropriado que seja esteticamente agradável e, ao mesmo tempo, facilite a compreensão para quem observa. Embora possa apresentar algumas limitações para visualização de dados clínicos, esta é uma ferramenta amplamente adotada em negócios, que pode construir visualizações customizadas para o propósito deste trabalho e é suficiente para a construção de protótipos dentro do escopo desta iniciação científica.

Inicialmente, foram realizadas reuniões com os envolvidos do projeto, incluindo especialista no domínio para a definição e design de diferentes tipos de gráficos com diferentes especificações apresentadas em um conjunto de abas.

A Figura 1 apresenta tela principal do protótipo com abas na lateral para seleção de diferentes tipos de visualizações, conforme definido em design em conjunto com especialista no domínio.

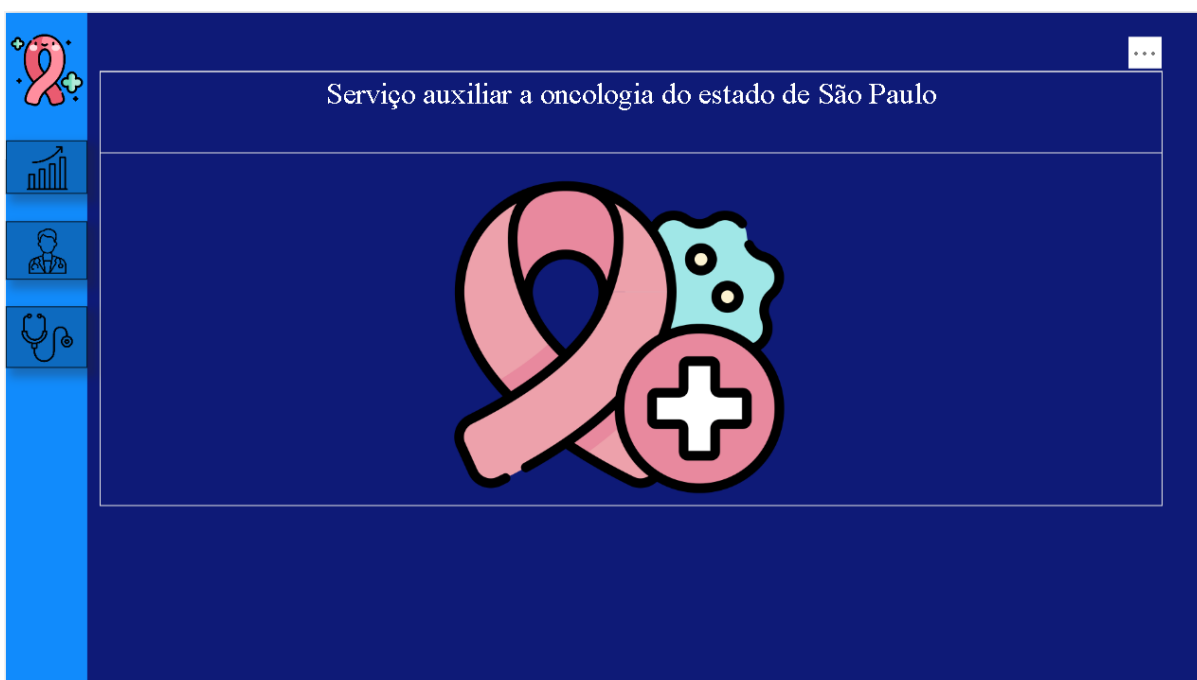


Figura 1. Tela principal com abas para seleção de diferentes visualizações

Ao dividirmos o conteúdo em abas (lado esquerdo da Figura 1), criamos uma estrutura que facilita a exploração aprofundada de temas específicos, mantendo a clareza e organização, bem como um design minimalista. Sendo assim, a primeira aba permite uma análise detalhada das recorrências, oferecendo a oportunidade de visualizar relações entre a localização geográfica, características demográficas dos pacientes e datas de consulta, com a possível ocorrência de recidivas. O objetivo foi facilitar a compreensão e a identificação de padrões e *insights* relevantes por parte de especialistas no domínio.

¹¹ <https://learn.microsoft.com/en-us/power-bi/>

Essa abordagem foi adotada pois outras alternativas poderiam restringir a quantidade de informações que poderíamos abordar e limitar a exploração de fatores cruciais relacionados às recidivas. Ao optar pelas abas, ampliamos consideravelmente a diversidade de tópicos e variáveis que podem ser considerados e, por conseguinte, enriquecemos a nossa compreensão sobre o fenômeno das recidivas de câncer.

Conforme ilustra a Figura 2 dentro de cada aba os usuários podem selecionar e examinar os diversos tipos de tratamento, como quimioterapia, radioterapia e cirurgia. Isso permite que eles realizem análises comparativas entre os tratamentos, que compreendam as recidivas associadas a cada um deles e identifiquem padrões que possam ser relevantes para suas necessidades de pesquisa (ou prática clínica).

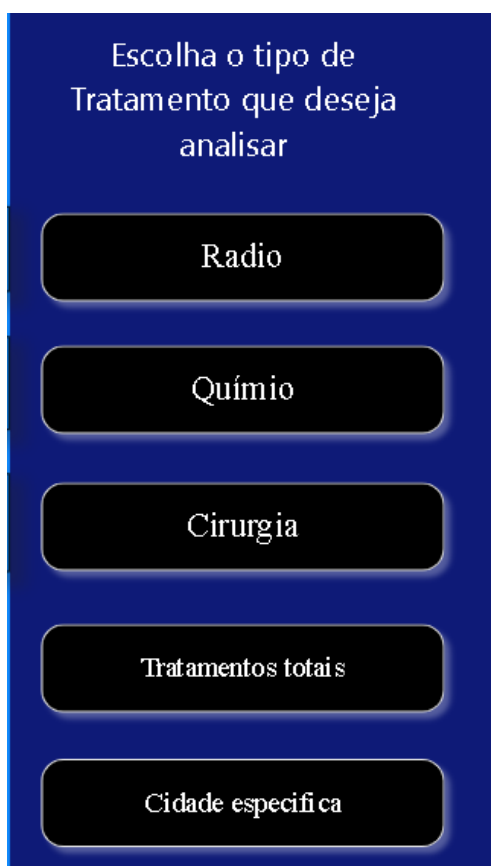


Figura 2. Parâmetros de seleção para uma aba específica

3.2. Funcionalidades e utilização da ferramenta

Após a especificação e estruturação geral do design a ser adotado, direcionamos nossa atenção para o próximo passo: a utilização eficiente dos dados. Nesse contexto, consideramos cuidadosamente que tipos de gráficos, tabelas ou mapas seriam os mais adequados para facilitar a compreensão dos profissionais de saúde. Dessa forma, optamos por empregar gráficos de setores, gráficos de linhas e gráficos de colunas, a fim de detalhar os tratamentos que apresentaram recidivas e destacar o número total de pacientes submetidos a diferentes tipos de tratamentos, tais como quimioterapia, radioterapia e cirurgia.

Conforme ilustra o exemplo da Figura 3, os gráficos de setores permitem uma visão proporcional das recorrências em relação aos diferentes tratamentos, enquanto os gráficos de linhas e de colunas fornecem uma visualização temporal e comparativa das recidivas. Dessa forma, viabilizamos uma análise mais eficaz das tendências ao longo do tempo e das variações entre os diferentes métodos de tratamento.

A variável 'RECENHUM' neste estudo sobre recorrência de câncer é fundamental para identificar e categorizar os pacientes que não apresentaram reincidência da doença. Essa variável, representando a ausência de recorrência do câncer em determinados pacientes, desempenha um papel crucial na análise dos dados e na compreensão dos resultados obtidos a partir do estudo.



Figura 3. Exemplo de visualização de recidivas por tratamento (Radioterapia)

Por meio do exemplo exibido na Figura 3, um especialista no domínio pode, por exemplo, visualizar que dentre as 271,6 mil pessoas submetidas a tratamentos de radioterapia 37 mil delas apresentaram recidiva, ligeiramente acima da média. O gráfico exibido no exemplo também permite o especialista identificar em quais localidades houve um maior índice de recidivas em relação ao tratamento por radioterapia.

Além das abordagens supracitadas, também foram explorados outros métodos para apoiar visualmente aqueles que estão fazendo a observação dos dados. Um exemplo é a utilização de um recurso do Power BI chamado ArcGIS Maps². Esse

recurso é uma ferramenta de mapeamento geográfico que permite uma representação visual interativa dos dados em um contexto espacial.

Conforme ilustra exemplo na Figura 3, ao incorporar o ArcGIS Maps ao protótipo foi possível criar mapas interativos que fornecem *insights* adicionais sobre a distribuição geográfica das recidivas de câncer. Esse recurso nos permitiu mapear as áreas onde as recorrências foram mais prevalentes após diferentes tipos de tratamentos. Por meio dessa abordagem, é possível apoiar a identificação de regiões com maiores índices de recidiva, como também estabelecer conexões geográficas entre os resultados dos tratamentos e as condições locais.



Figura 4. Exemplo de visualização sobre recidiva de câncer utilizando o ArcGIS Maps

4. Discussão, Conclusão e Trabalhos Futuros

A ferramenta Power BI se mostrou útil para a construção dos protótipos e trouxe benefícios para nossa abordagem de visualização de dados ao incorporar visualizações disponíveis. De maneira geral, em uma análise preliminar ela apresentou boa usabilidade, e potencial para possibilitar que profissionais de saúde e pesquisadores tenham *insights* derivados das informações médicas. Entretanto, é importante destacar que também encontramos alguns desafios e limitações no uso dessa ferramenta.

Uma das possíveis desvantagens levantadas do Power BI é a complexidade no uso de uma ampla variedade de serviços adicionais e um conjunto diversificado de ferramentas relacionáveis. Essa complexidade pode dificultar a navegação e o entendimento completo de todas as funcionalidades oferecidas. Além disso, a curva de aprendizado para explorar plenamente essas ferramentas pode ser íngreme, especialmente para aqueles que estão menos familiarizados com a plataforma.

Outro ponto negativo que enfrentamos está relacionado à manipulação de informações provenientes do banco de dados. Foram encontradas limitações ao tentar renomear campos ou colunas de maneira flexível, o que pode afetar a precisão e a compreensão das informações apresentadas. Essa restrição pode impactar a capacidade de personalizar a visualização para atender às necessidades específicas da análise.

Apesar dessas limitações, os benefícios da usabilidade, interatividade e poder de visualização oferecidos pelo Power BI superaram os desafios encontrados. Para lidar com esses obstáculos, procuramos soluções alternativas e trabalhamos para otimizar a utilização da ferramenta.

Como próximos passos desta iniciação científica pretende-se ampliar a compreensão Power BI e criação de visualizações adicionais. Este é um passo crucial para a evolução do projeto, principalmente porque permitirá refinar a abordagem e direcionar o foco para casos mais específicos e detalhados.

Outro aprimoramento importante é a integração da visualização com resultados da aplicação de técnicas de ciência de dados e inteligência artificial a base da FOSP. Por exemplo, é possível elaborar recursos para visualização de clusters de pacientes a partir de técnicas de aprendizado de máquina como as apresentadas em Crocco et al. (2022).

Por fim, destaca-se como trabalho futuro o registro e a disponibilização da ferramenta na Internet, bem como a análise de seu uso por profissionais de saúde.

5. Agradecimentos

Este trabalho foi financiado em parte pelo Conselho Nacional de Desenvolvimento Científico e Tecnológico (CNPq) processo nº 405352/2021-2 e pelo programa PIBIC do CTI Renato Archer. Agradecemos a Fundação Oncocentro de São Paulo por disponibilizar publicamente os dados utilizados neste projeto.

6. Referência

Costa, F. G. D. (2017). Visualização de dados e sua importância na era do Big Data. Trabalho de conclusão de graduação. Repositório Universidade Federal do Rio Grande do Sul - UFRGS, <http://hdl.handle.net/10183/169109> acessado em 21 de agosto de 2023.

Crocco, P.F., Maeda, A.E., Ruppert, G.C.S, Dametto, M., Bonacin. R. Clusterização de Dados Abertos em Oncologia Usando Técnicas de Aprendizado de Máquina: um estudo preliminar sobre recidiva de câncer de próstata. XXIV Jornada de Iniciação Científica do Centro de Tecnologia da Informação Renato Archer - JICC'2022 PIBIC/CNPq/CTI - Outubro de 2022 – Campinas – São Paulo. Disponível em: <https://www.gov.br/cti/pt-br/publicacoes/producao-cientifica/jicc/xxiv-jicc-2022> Último acesso em 20 de agosto de 2023.

Dowding D., Randell R., Gardner P., et al. (2015) Dashboards for improving patient care: review of the literature. *Int J Med Inform.* 84(2):87-100. doi:10.1016/j.ijmedinf.2014.10.001

SBCO (2021) Recidivas do câncer: entenda como acontece, Sociedade Brasileira de Cirurgia Oncológica. Disponível em <https://sbco.org.br/recidivas-do-cancer-entenda-como-acontece/>