

PGCOMP - Programa de Pós-Graduação em Ciência da Computação
Universidade Federal da Bahia (UFBA)
Av. Milton Santos, s/n - Ondina
Salvador, BA, Brasil, 40170-110

<https://pgcomp.ufba.br>
pgcomp@ufba.br

A evolução tecnológica em conjunto com a crescente produção massiva de dados demandam criar novos mecanismos para processamento e análise nos mais diferentes domínios de informação. O mesmo ocorre para o domínio de esportes, que possui um grande apelo popular ao redor do mundo e vem conquistando cada vez mais interesse da comunidade científica. Dentre os diferentes gêneros esportivos, o ciclismo vem ganhando atenção especial na comunidade de visualização de informação e Visual Analytics ao longo dos anos. De modo geral, o foco destes trabalhos está na descoberta de padrões para traçar estratégias, exibir informações estatísticas, ou analisar desempenho de grupos de ciclistas. Contudo, há uma lacuna de estudos que foquem em percursos não relacionados a competições profissionais ou ainda que considerem atributos de intensidade em suas análises. Nesse contexto, esse trabalho apresenta o CycleVis, uma abordagem de visualização de dados para ciclismo que permite melhor compreensão do desempenho esportivo de ciclistas e de suas pedaladas. Para tal, utilizaram-se características de intensidade, como velocidade, frequência cardíaca e elevação, além da geolocalização das pedaladas. Para validar a abordagem, foi desenvolvido um caso de uso e o mesmo foi avaliado por meio de estudo com usuários, que resultou em uma avaliação positiva para eficácia e usabilidade. Todos os usuários concordaram que a CycleVis pode auxiliar no desenvolvimento de estratégias para a compreensão da atividade do ciclismo e também ser adaptado para outras modalidades esportivas.

Palavras-chave: Análise esportiva, visualização de dados, análise visual, ciclismo.

CycleVis - Uma abordagem de visualização esportiva para análise de desempenho de ciclistas

Iury Gomes de Oliveira

Dissertação de Mestrado

Universidade Federal da Bahia

Programa de Pós-Graduação em
Ciência da Computação

Setembro | 2023

MSC | 165 | 2023

CycleVis - Uma abordagem de visualização esportiva para análise de desempenho de ciclistas

Iury Gomes de Oliveira

UFBA





Universidade Federal da Bahia
Instituto de Computação

Programa de Pós-Graduação em Ciência da Computação

**CYCLEVIS - UMA ABORDAGEM DE
VISUALIZAÇÃO ESPORTIVA PARA
ANÁLISE DE DESEMPENHO DE CICLISTAS**

Iury Gomes de Oliveira

DISSERTAÇÃO DE MESTRADO

Salvador
14 de Setembro de 2023

IURY GOMES DE OLIVEIRA

**CYCLEVIS - UMA ABORDAGEM DE VISUALIZAÇÃO
ESPORTIVA PARA ANÁLISE DE DESEMPENHO DE CICLISTAS**

Esta Dissertação de Mestrado foi apresentada ao Programa de Pós-Graduação em Ciência da Computação da Universidade Federal da Bahia, como requisito parcial para obtenção do grau de Mestre em Ciência da Computação.

Orientador: Prof. Dr^o Danilo Barbosa Coimbra

Salvador
14 de Setembro de 2023

Ficha catalográfica elaborada pela Biblioteca Universitária de
Ciências e Tecnologias Prof. Omar Catunda, SIBI – UFBA.

O48 Oliveira, Iury Gomes.

CycleVis - Uma abordagem de visualização esportiva
para análise de desempenho de ciclistas. / Iury Gomes de
Oliveira. – Salvador, 2023.

121p.: il.

Orientador: Prof. Dr. Danilo Barbosa Coimbra

Dissertação (Mestrado) – Universidade Federal da
Bahia, Instituto de Computação, 2023.

1. Análise Esportiva. 2. Visualização de Dados. 3.
Ciclismo. I. Coimbra, Danilo Barbosa. II. Universidade
Federal da Bahia. III. Título.

CDU 796:004



“CycleVis - Uma abordagem de visualização esportiva para análise de desempenho de ciclistas”

Iury Gomes de Olivera

Dissertação apresentada ao
Colegiado do Programa de Pós-Graduação em Ciência
da Computação na Universidade Federal da Bahia,
como requisito parcial para obtenção do Título de
Mestre em Ciência da Computação.

Banca Examinadora

Prof. Dr. Danilo Barbosa Coimbra (Orientador)

Prof. Dr. Frederico Araujo Durão - PGCOMP

Prof. Dr. Renato Lima Novais - IFBA

Dedico este trabalho a Deus, sendo a fonte de todo conhecimento e sabedoria. Em Mateus 25:20 está escrito: 'E, chegando o que recebera cinco talentos, apresentou outros cinco talentos, dizendo: Senhor, entregaste-me cinco talentos; eis aqui outros cinco talentos que ganhei', reconheço que cada habilidade e conhecimento que adquiri são talentos que Ele graciosamente me deu. Tive em vista multiplicar estes talentos através do rigoroso trabalho de pesquisa e estudo, com o excelente auxílio de meu orientador. Agora, neste momento decisivo, devolvo estes talentos multiplicados, incorporados em cada página deste trabalho, com o desejo sincero de que sejam usados para a Sua glória e honra. Com gratidão e humildade, entrego este fruto do meu trabalho.

AGRADECIMENTOS

À minha amada esposa, Livia Barbosa Gomes de Oliveira, e à minha mãe, Alzira Costa Gomes, agradeço profundamente pelo amor inabalável, dedicação constante e preocupação incansável. Sem a presença fundamental de vocês na minha vida, a realização deste trabalho não teria sido possível. Vocês representam meu alicerce, minha força motriz.

Para as minhas filhas, Talita Barbosa Gomes de Oliveira e Helena Barbosa Gomes de Oliveira, agradeço pelo amor genuíno e pelos incontáveis momentos de diversão e alegria. Vocês são os alicerces do meu coração e a essência do meu espírito!

Ao Prof. Dr. Danilo Coimbra, minha gratidão vai além do agradecimento pela orientação criteriosa e valiosa. Agradeço também pelo exemplo íntegro e inspirador como educador. Sua confiança, paciência e compreensão foram imprescindíveis nesta jornada. Ao me selecionar como seu orientando, você não apenas transformou a minha carreira, mas também reverberou de maneira positiva na vida da minha família e no meu futuro.

Para os estudantes da turma MATD74, agradeço pela camaradagem e apoio incansável, partilhando os desafios e triunfos ao longo desses dois anos de mestrado. Nossa união nos permitiu superar todos os obstáculos.

Ao IFTO, agradeço pelo suporte robusto e incentivo durante este mestrado, propiciando minha evolução profissional. Espero, ao retomar minhas atividades presenciais, contribuir e agregar valor de maneira significativa, superando as expectativas de todos.

Aos funcionários e professores da UFBA, minha gratidão pelo apoio administrativo e pedagógico. Vocês facilitaram meu caminho e proporcionaram um ambiente propício para o crescimento e aprendizado.

Em muitos momentos da jornada de criação desta dissertação, encontrei refúgio e inspiração nas palavras de Provérbios 13:12, “A esperança adiada desfalece o coração, mas o desejo atendido é árvore de vida”. Cada obstáculo, cada noite tardia de estudo parecia adiar a esperança do término deste trabalho, chegando a desfalecer o coração.

No entanto, as sementes de determinação e resiliência plantadas nesses desafios brotaram, cultivadas pela constante busca por conhecimento e aperfeiçoamento. Agora, o desejo foi finalmente alcançado. Este trabalho, a árvore que brotou dos inúmeros desejos e esperanças, representa mais do que uma simples conquista acadêmica - é a representação tangível da árvore de vida mencionada no versículo.

A chegada desta tão aguardada alegria traz consigo um grande sentimento de realização, objetivos alcançados, missão cumprida. Esta dissertação não é apenas a conclusão de um ciclo, mas também o começo de uma nova fase, onde o conhecimento adquirido e as habilidades aprimoradas serão utilizadas para contribuir com a minha área de atuação.

Assim, concluo esta etapa com muita alegria, sabendo que cada passo, por mais desafiador que tenha sido, valeu a pena. Esta dissertação, portanto, representa o florescimento de uma nova fase que, mesmo quando parecia adiada, jamais foi abandonada.

RESUMO

A evolução tecnológica em conjunto com a crescente produção massiva de dados demandam criar novos mecanismos para processamento e análise nos mais diferentes domínios de informação. O mesmo ocorre para o domínio de esportes, que possui um grande apelo popular ao redor do mundo e vem conquistando cada vez mais interesse da comunidade científica. Dentre os diferentes gêneros esportivos, o ciclismo vem ganhando atenção especial na comunidade de visualização de informação e Visual Analytics ao longo dos anos. De modo geral, o foco destes trabalhos está na descoberta de padrões para traçar estratégias, exibir informações estatísticas, ou analisar desempenho de grupos de ciclistas. Contudo, há uma lacuna de estudos que foquem em percursos não relacionados a competições profissionais ou ainda que considerem atributos de intensidade em suas análises. Nesse contexto, esse trabalho apresenta o CycleVis, uma abordagem de visualização de dados para ciclismo que permite melhor compreensão do desempenho esportivo de ciclistas e de suas pedaladas. Para tal, utilizaram-se características de intensidade, como velocidade, frequência cardíaca e elevação, além da geolocalização das pedaladas. Para validar a abordagem, foi desenvolvido um caso de uso e o mesmo foi avaliado por meio de estudo com usuários, que resultou em uma avaliação positiva para eficácia e usabilidade. Todos os usuários concordaram que a CycleVis pode auxiliar no desenvolvimento de estratégias para a compreensão da atividade do ciclismo e também ser adaptado para outras modalidades esportivas.

Palavras-chave: Análise esportiva, visualização de dados, análise visual, ciclismo.

ABSTRACT

The technological evolution together with the growing mass production of data demand the creation of new mechanisms for processing and analysis in the most different areas of information. The same goes for the field of sports, which has a great popular appeal around the world and has been gaining more and more interest from the scientific community. Among the different sports genres, cycling has been gaining special attention in the community of information visualization and Visual Analytics over the years. In general, the focus of these works is on finding patterns to draw strategies, display statistical information, or analyze performance of groups of cyclists. However, there is a gap of studies that focus on paths not related to professional competitions or that consider attributes of climate and intensity in their analysis. In this context, this work presents CycleVis, a data visualization approach for cycling that allows a better understanding of the sports performance of cyclists. For this, intensity characteristics such as speed, heart rate and elevation were used, in addition to the geolocation of the runs/pedals. To validate the approach, a use case was developed and it was evaluated through a study with users, which resulted in a very positive evaluation for effectiveness and usability. All users agreed that CycleVis can assist in the development of strategies for cycling or be adapted for other sports disciplines.

Keywords: Sports analysis, data visualization, visual analysis, cycling

SUMÁRIO

Capítulo 1—Introdução	1
1.1 Contextualização	1
1.2 Motivação	3
1.3 Problema de Pesquisa	4
1.4 Objetivo geral e Específicos	4
1.5 Questões de Pesquisa	5
1.6 Estrutura do Trabalho	5
Capítulo 2—Revisão Bibliográfica	7
2.1 Considerações Iniciais	7
2.2 Visualização de Dados	7
2.3 Mecanismos de Interação	8
2.3.1 Interações Humanas	9
2.3.2 Símbolos Visuais	9
2.3.3 Cores	11
2.3.4 Operações Visuais	13
2.4 Técnicas de Visualização	14
2.4.1 Visualização de Grafos	15
2.4.2 Visualização de Séries Temporais	18
2.4.3 Visualização Multidimensional	24
2.4.4 Geovisualização	29
2.5 Métodos de Avaliação de Usabilidade	33
2.6 Considerações Finais	35
Capítulo 3—Trabalhos Relacionados	37
3.1 Considerações Iniciais	37
3.2 Visualização de Dados Esportivos	37
3.2.1 Esportes Individuais	42
3.2.2 Esportes em Grupo	45
3.3 Visualização em Ciclismo	50
3.4 Considerações Finais	57

Capítulo 4—CycleVis	59
4.1 Considerações Iniciais	59
4.2 <i>Pipeline</i> de Visualização dos Dados	60
4.3 Arquiteturas do CycleVis	60
4.3.1 Fase de ambiente de desenvolvimento	61
4.3.2 Fase de Ambiente de Produção	62
4.4 Conjunto de dados	64
4.5 Grupos de usuários, Objetivos e Tarefas Analíticas	65
4.5.1 Grupos de Usuário	65
4.5.2 Objetivos	66
4.5.3 Tarefas Analíticas	67
4.6 Disposição Visual	67
4.7 Funcionamento do CycleVis	68
4.7.1 Painéis Visuais para Escolha de Ciclistas e Pedaladas	68
4.7.2 Painéis Visuais para Métricas de Desempenho	72
4.7.3 Painel Visual para Métrica Global de Intensidade	74
4.7.4 Painel Visual Estatístico	76
4.7.5 Painel Visual de Geolocalização	78
4.7.6 Visualizações Coordenadas	78
4.8 Considerações Finais	80
Capítulo 5—Avaliação	83
5.1 Considerações Iniciais	83
5.2 Cenário de Uso	83
5.2.1 Seleção de Ciclistas e Pedaladas	83
5.2.2 Avaliando as Distâncias Percorridas	83
5.2.3 Examinando a Frequência Cardíaca, Velocidade e Elevação	84
5.2.4 Examinando a Intensidade Global	84
5.2.5 Compreendendo os Indicadores Médios	85
5.2.6 Interpretando as Informações Geoespaciais	86
5.2.7 Trabalhando com visualizações coordenadas	86
5.3 Avaliação com Usuários	88
5.3.1 Metodologia e Instruções	88
5.3.2 Demografia	89
5.3.3 Resultados das Tarefas Analíticas	90
5.3.4 Resultados do Questionário de Usabilidade	98
5.3.5 Resultados do Feedback Qualitativo	101
5.4 Considerações Finais	102
Capítulo 6—Conclusões	103
6.1 Contribuições	103
6.2 Limitações	104

SUMÁRIO	xiii
6.3 Trabalhos Futuros	104
Referências Bibliográficas	107
Apêndice A—Apêndice	113

LISTA DE FIGURAS

1.1	Este gráfico enfatiza a evolução da literatura em torno da visualização esportiva (PERIN et al., 2018).	2
1.2	Número médio anual de publicações sobre ciclismo por quinquênios (PUCHER; BUEHLER, 2017).	3
2.1	<i>Pipeline</i> de visualização proposto em Ward, Grinstein e Keim (2010). . .	8
2.2	Atributos visuais que promovem fácil distinção (ALMEIDA; NETO, 2001). . .	10
2.3	Variáveis visuais para visualizações que incluem mapas (SLOCUM et al., 2014).	10
2.4	Atributos visuais que melhor se adequam a determinado tipo de dado (ILINSKY; STEELE, 2011).	11
2.5	a) Modelo de cores no sistema HSL. b) Sensibilidade espectral (MUNZNER, 2014).	12
2.6	Brilho e saturação são ordenados de forma mais natural pelo sistema perceptivo humano, mas a Matiz não é (MUNZNER, 2014).	13
2.7	Exemplo de grafo direcionado com ciclos e pesos nas arestas e vértices. . .	15
2.8	a) Grafo direcionado por força. b) Um grafo enorme de 26.028 vértices e 100.290 arestas sem muita estrutura visível. (MUNZNER, 2014)	16
2.9	a) Visão matricial de um grafo pequeno. b) Visão matricial do grafo em c). Grafo de uma rede maior (MUNZNER, 2014).	17
2.10	a) Um grafo representando a rede de pesquisa do campo de Visualização de dados no concurso InfoVis 2004. b) Os tons em amarelo demonstra a força de colaboração de cada pesquisador, em roxo a força de colaboração entre diferentes pesquisadores de uma mesma comunidade. c) Revela como o autor Ed Chi serve de ponte entre dois subgrupos em uma comunidade de pesquisa. d) Revela padrões de colaboração em comunidades de pesquisa, há comunidades de pesquisa em que existem colaborações entre vários pesquisadores, como também comunidades de pesquisa em que há pouca colaboração entre os pesquisadores (HENRY; FEKETE; MCGUFFIN, 2007).	19
2.11	a) Série temporal da quantidade de passagens aéreas vendidas nos Estados Unidos no período de 1949 a 1960. A linha pontilhada representa uma estimativa de tendência existente nesses dados. b) Variação sazonal existente na série temporal referente à quantidade de passagens aéreas vendidas nos Estados Unidos no período de 1949 a 1960. c) Resíduos extraído da série temporal referente à quantidade de passagens aéreas vendidas em um período de onze anos (MALETZKE, 2009).	20

2.12	Representação linear contra representação cíclica do tempo: diferentes percepções podem ser obtidas (WARD; GRINSTEIN; KEIM, 2010).	21
2.13	a) Exemplo de gráfico de linhas b) Exemplo de Espiral de série temporal (RIBECCA, 2013).	22
2.14	Legibilidade do gráfico de linhas (ORTIGOSSA, 2018).	22
2.15	Exemplo de stream graph (RIBECCA, 2013).	23
2.16	Stream Graph proposto em (HAVRE et al., 2002).	24
2.17	Exemplo de correlações que podem ser inferidas do <i>Scatterplot Matrix</i>	26
2.18	Exemplo de Coordenadas Paralelas (RIBECCA, 2013).	26
2.19	a) Exemplo de Gráfico de Radar (RIBECCA, 2013). b) Gráfico de Radar com muitos polígonos.	27
2.20	Exemplo de Mapa de Calor (RIBECCA, 2013).	28
2.21	a) Exemplo de <i>Stacked Area</i>	28
2.22	Exemplo de <i>Tablelens</i> (GRATZL et al., 2013).	29
2.23	À esquerda um exemplo genérico de mapa coroplético (RIBECCA, 2013). À direita um mapa coroplético do Brasil, que demonstra o interesse sobre o tema Coronavírus (GUSMÃO, 2021).	31
2.24	Exemplo de <i>dotmap</i> que mostra um mapa racial brasileiro.	31
2.25	Exemplo de conexão de mapas (RIBECCA, 2013).	32
2.26	Exemplo de <i>flow map</i> (RIBECCA, 2013).	32
2.27	Exemplo de mapa de bolhas (RIBECCA, 2013).	33
3.1	Técnicas de visualização de dados com alta dimensionalidade. (a) <i>scatter plot</i> (gráfico de dispersão), (b) <i>radar chart</i> (gráfico de radar), (c) <i>parallel coordinate</i> (paralelas coordenadas) e (d) <i>heat map</i> (gráficos de calor).	39
3.2	Exemplos de visualizações de séries temporais. (a) Gráfico de linha (que demonstra uma variação ao longo do tempo a classificação de times de futebol), (b) Histograma (que demonstra a posse de bola de dois times de futebol ao longo do tempo).	39
3.3	Exemplo de visualização hierárquica técnica que pode ser incluída na categoria visualização de rede segundo Du e Yuan (2021). A estrutura hierárquica acima indica os resultados previsíveis no campeonato de basquete.	40
3.4	Visualizações esportivas baseadas em Glifos. (a) Glifo em forma de jogador para descrever eventos com dados de rugby.	41
3.5	Exemplo de visualização de corrida de rua. Cada linha representa um corredor. Na direita a topografia da corrida. Quanto mais brilhante for a cor, maior será o nível de esforço do corredor, para cores mais escuras, menor será o nível de esforço do corredor (OLIVEIRA et al., 2013).	42
3.6	Conforme Lu, Wang e Lin (2014), neste gráfico, os movimentos reais feitos pelos usuários com círculos, os quadrados representam os movimentos estimados pela IA. Na direita o significado dos símbolos, que podem ser combinados para transmitir informações adicionais. Por exemplo, um círculo branco com uma borda preta significa que as peças brancas dão um xeque nas peças pretas, logo as peças pretas devem reagir.	44

3.7	Permite comparar recordes de homens e mulheres para diferentes modalidades (TORRE-ARENAS, 2017).	45
3.8	Visualização em <i>CourtVision</i> . Possibilitar comparar a efetividade dos jogadores em diversos locais das quadras.	47
3.9	Interface principal do <i>SoccerStories</i> . A parte 1 apresenta uma linha do tempo com 10 fases relevantes nas quais o jogador selecionado participou. A parte 2 é um zoom temporal de uma fase selecionada do jogo com a participação do jogador selecionado sobre um <i>layout</i> de campo de futebol. A parte 3 apresenta estatísticas do jogador selecionado, no caso Andres Iniesta. A parte 4 apresenta um resumo em miniatura das 10 fases relevantes do jogador selecionado. A parte 5 apresenta um espaço para anotações textuais.	49
3.10	a) Baseline Bar Display b) Player Map (COX; STASKO, 2006).	50
3.11	<i>Layout</i> principal do TourVis, que consiste em seis visualizações simultâneas: a) Visão Geral da Corrida, b) Visão de uma etapa específica, c) Painel de seleção de times e atletas, d) Ranking da etapa, e) Visualização de detalhes sob demanda e f) Informações específicas de um atleta e sua equipe.	51
3.12	Mapa de perfil proposto por Wood (2015).	53
3.13	Visualização que mostra as posições relativas de 1000 ciclistas em 21 pontos de controle no Audax Londres-Edimburgo-Londres de 2013, com o progresso de um ciclista selecionado em laranja.	54
3.14	Comparativo de Sivtsov (vermelho) com Brajkovic (azul) no prólogo do Tour de France 2012, incluindo seu período de preparação para a corrida, o gráfico de linhas adicionais reflete o progresso do tempo e indica lacunas na medição.	56
3.15	Visualização proposta por Fister, jr e Fister (2016) durante um treinamento para eventos de competição, os valores mostrados nessa figura são tendenciosos, mas o propósito da visualização e demonstrar rapidamente os principais valores obtidos em um possível treinamento.	57
4.1	<i>Pipeline</i> da proposta de visualização de dados.	60
4.2	Arquitetura do CycleVis na fase de ambiente de desenvolvimento.	61
4.3	Arquitetura do CycleVis na fase de ambiente de produção. A Octokit.js ¹ é o SDK (<i>Software Development Kit</i>) disponibilizado pelo Github para integrações com seus repositórios. A Dexie.js ² é a biblioteca utilizada para gerenciar o IndexedDB.	63
4.4	<i>Layout</i> do <i>dashboard</i> visual do CycleVis apresentando painéis visuais justapostos contendo dados de ciclistas e pedaladas, com informações de distância, elevação, frequência cardíaca e geolocalização.	68
4.5	Diagrama de sequência do CycleVis indicando as interações dos principais componentes do software.	69

4.6	a) Checkbox de escolha de ciclistas. b) <i>Slider</i> para definir as distâncias das pedaladas de cada ciclistas a serem comparadas. c) Botão <i>Ver Cycle Lens</i> para acessar a visualização Cycle Lens e selecionar até três pedaladas de cada ciclista. d) Botão Atualizar Painéis para geração do <i>dashboard</i> contendo os demais painéis visuais do CycleVis.	70
4.7	Paleta de Cores utilizada no CycleVis.	70
4.8	Visualização Cycle Lens de dois ciclistas no modo desagrupado e em ordem decrescente. Na parte de cima, o ciclista 4 possui 3 pedaladas selecionadas, enquanto o ciclista 12 possui uma pedalada, sendo essa selecionada e com distância de 67.87 km. Na parte de baixo, a mesma visualização com o modo agrupado ativado	71
4.9	Visualização do Gráfico de Distâncias.	72
4.10	Visualização Métricas de Fluxo. a) Tooltip com valores absolutos b) Régua. c) Seletor de Pedaladas. O manuseio da régua ativa o tooltip, que mostra o valor do fluxo para a distância apontada pela régua	73
4.11	Visualização Intensidade Global. a) Mecanismo de zoom. b) Mecanismo de filtro. A utilização dos mecanismos de filtro e zoom que possibilita visualizar as intensidades entre 14400 e 28800 metros que estão acima da intensidade 0.5	75
4.12	Visualização Gráfico Estatísticos. Parte superior com os indicadores médios dos ciclistas e parte inferior com os atributos médios das pedaladas.	77
4.13	Visualização Mapa de Rastreamento. a) Habilita o zoom no mapa. b) Ativa o <i>fullscreen</i> . c) Ativa a camada de intensidade. d) Ativa a camada de Rotas. e) Ativa a camada de Distância	79
4.14	Coordenação entre Mapa de Rastreamento e Métricas de Fluxo. Ao deslocar as régua no Métricas de Fluxo, ícones indicadores das variáveis de elevação, velocidade e frequência cardíaca são plotadas no Mapa de Rastreamento	80
5.1	Cenário de uso para seleção de ciclistas e pedaladas.	84
5.2	Examinando o comportamento da frequência cardíaca, velocidade e elevação no ponto 39800 metros.	85
5.3	Examinando o nível de intensidade entre o trecho 20000 e 30000 metros.	86
5.4	À esquerda, o gráfico estatístico que possibilita uma compreensão dos indicadores médios dos ciclistas e das pedaladas. À direita o mapa de rastreamento apresentando as informações geoespaciais.	87
5.5	A esquerda, o métricas de fluxo indicando as variações de frequência cardíaca, velocidade e elevação no intervalo entre 20000 e 32000 metros. A direita o mapa de rastreamento que visualiza a correlação entre localização geográfica e as métricas de fluxo	88
5.6	Fluxograma das etapas de avaliação com usuários.	89
5.7	Informações sobre idade, gênero e nível educacional dos participantes.	90
5.8	Informações sobre experiência prévia dos participantes.	91

5.9	Tempo médio em minutos de realização das tarefas dos usuários. Média geral em verde, especialistas em visualização em azul, ciclistas em amarelo.	93
5.10	Tempo individual em minutos de realização das tarefas dos usuários. . . .	94

LISTA DE TABELAS

5.1	Listagem dos objetivos e suas respectivas tarefas.	92
5.2	Resultados da complexidade das questões.	95
5.3	Resultados do questionário de usabilidade.	98
5.4	Tabela com os comentários/feedbacks dos usuários.	101

LISTA DE SIGLAS

ACM CHI	Conferência internacional de Interação Humano-Computador (HCI) organizada pela Association of Computing Machinery
EuroVis	Conferência anual de visualização organizada pelo Eurographics Working Group on Data Visualization
FINA	Federação Internacional de Natação
GPS	Global Positioning System, ou Sistema de Posicionamento Global
GPX	Formato de intercâmbio GPS, um esquema XML designado para transferir dados GPS entre aplicações
HTTP	Hypertext Transfer Protocol
IA	Inteligência Artificial
IEEE Vis	Conferência Anual sobre visualização científica, visualização de informações e análise visual administrada pelo Comitê Técnico de Visualização e Gráficos da IEEE Computer Society
IHC	Interface Humano-Computador
JSON	JavaScript Object Notation
MVC	Model-View-Controller, um padrão arquitetural de sistemas que oferece separação lógica de negócios (Model), a interface do usuário (View) e o controle de fluxo de dados(Controller)
PPA	Sigla utilizada para se referir a pontos por tentativa em (GOLDSBERRY, 2012)
REST	Representational State Transfer
RGB	abreviatura do sistema de cores aditivas formado pelos tons de Vermelho, Verde e Azul
SDK	Software Development Kit
TCX	Formato de intercâmbio GPS, um esquema XML designado para transferir dados GPS entre aplicações
TRID	Transportation Research International Documentation
TRB	Transportation Research Board
UEMs	Do inglês <i>usability evaluation methods</i> (métodos de avaliação de usabilidade)
UML	Do inglês <i>Unified Modeling Language</i> (Linguagem Unificada de Modelagem)

XML É a sigla para "Extensible Markup Language", uma linguagem de marcação com regras para formatar documentos de forma que ele seja facilmente lido tanto por humanos quanto por computadores

LISTA DE ALGORITMOS

1	Lógica do Algoritmo para Calcular os dados no Métricas de Fluxo	74
---	---	----

INTRODUÇÃO

1.1 CONTEXTUALIZAÇÃO

A implementação generalizada de tecnologias de informação e comunicação em variadas esferas da vida humana e da sociedade tem sido uma ferramenta crucial para impulsionar o desenvolvimento econômico, social, cultural e tecnológico. Como resultado desta ampla aplicação, foi possível promover uma coleta e armazenamento em grande escala de dados em diversas áreas da atividade humana (KYSH, 2022). Como exemplo da produção massiva de dados podemos citar as populares redes sociais digitais. Por exemplo, o YouTube possui mensalmente 1.9 bilhões de usuários ativos e em média 300 horas de vídeo são carregadas a cada minuto diariamente. O WhatsApp possui mensalmente 1.5 bilhões de usuários ativos, gerando aproximadamente 60 bilhões de mensagens enviadas diariamente. O Instagram possui mensalmente 1 bilhão de usuários ativos, gerando aproximadamente 95 milhões de postagens diariamente (STOUT, 2022).

Há uma variedade de fontes de dados, tanto estruturados como não estruturados. Dados estruturados incluem bases de dados relacionais, planilhas do Excel, e arquivos em formatos JSON ou XML. Por outro lado, dados não estruturados podem vir de diversas fontes, como e-mail e vídeos, entre outros. Com o maior volume de dados e a maior diversidade de fontes, emerge a preocupação com a veracidade das informações, pois uma análise de dados baseada em informações errôneas ou mal elaboradas gera sérios problemas para o valor da análise de dados, uma das áreas que auxiliam neste processo de entrega de valor é a visualização de dados (ROMANILLOS et al., 2016).

A Visualização de Dados é um campo do conhecimento que utiliza modelos para representar informações em formato visual, buscando ampliar a cognição humana para compreender e assimilar informações. No contexto atual, onde grandes volumes de dados são coletados e acumulados, a relevância desse campo ganha cada vez mais força, pois a compreensão de informação apoiada por interfaces visuais é mais eficiente e ágil do que ler um texto ou uma tabela. Além disso, a representação gráfica pode transpor as barreiras entre diferentes linguagens (MUNZNER, 2014).

Nas últimas duas décadas, a literatura científica na área de Visualização de Dados vem evidenciando um aumento significativo no número de publicações relacionadas a tema

de visualização esportiva nas principais conferências mundiais de Visualização de Dados como a IEEE Vis, EuroVis, e ACM CHI (PERIN et al., 2018). A Figura 1.1 apresenta um gráfico deste estudo com a evolução do número de publicações sobre visualização esportiva, agrupadas por ano e categorizadas entre acadêmicos e não acadêmicos (sites, blogs, infográficos, etc.). Dos 98 trabalhos coletados no estudo, 35 foram considerados não acadêmicos e 63 foram considerados acadêmicos pelo autor.

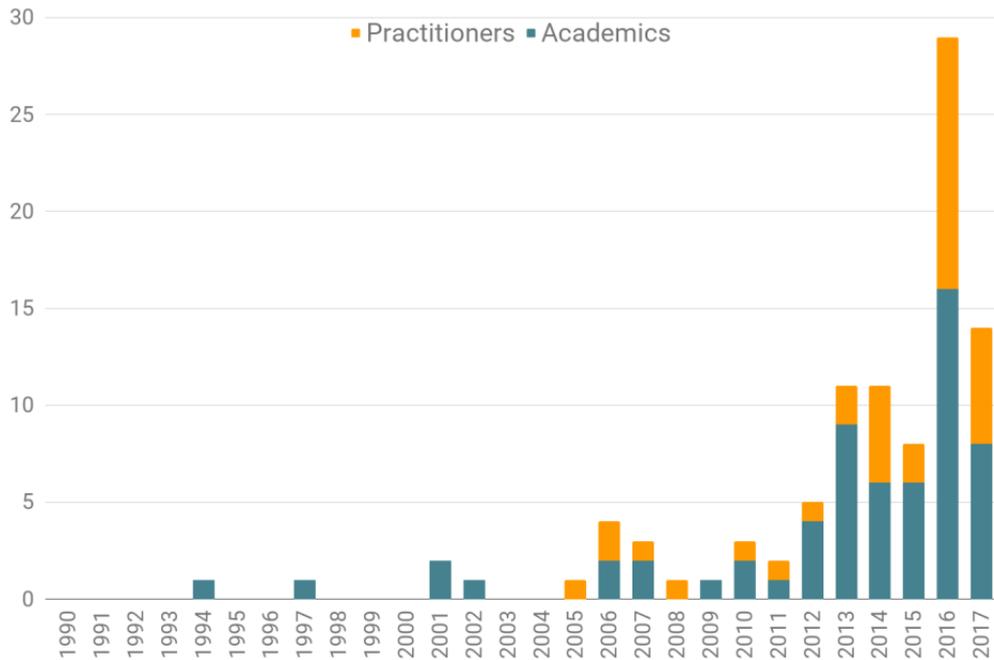


Figura 1.1: Este gráfico enfatiza a evolução da literatura em torno da visualização esportiva (PERIN et al., 2018).

Um esporte que vem atraindo a atenção na área de visualização esportiva é o ciclismo. A Figura 1.2 apresenta o aumento significativo em pesquisas publicadas sobre o ciclismo nas últimas décadas (PUCHER; BUEHLER, 2017). O TRID¹ (Transportation Research International Documentation), uma base de dados que fornece acesso a mais de 1,3 milhões de registros de pesquisa de transportes em todo o mundo, demonstra um aumento de 197 publicações entre 1991 a 1995 para 610 publicações entre 2011 a 2016 relacionadas ao ciclismo. Segundo a Web of Science, o número de artigos revisados por pares aumentou de 22 para 284. O TRB² (Transportation Research Board), um comitê que visa integrar o uso de bicicletas como meio de transporte observando critérios de segurança indica que a média anual de artigos sobre ciclismo saiu de 15 entre 2001 a 2005 para 101 entre 2011 a 2016, evidenciando um crescente interesse em pesquisa acadêmica sobre ciclismo.

¹Disponível em: <https://trid.trb.org/>. Acessado em 15/07/2022

²Disponível em: <https://www.trb.org/PedestriansAndBicyclists/TRBCommittees.aspx>. Acessado em 15/07/2022

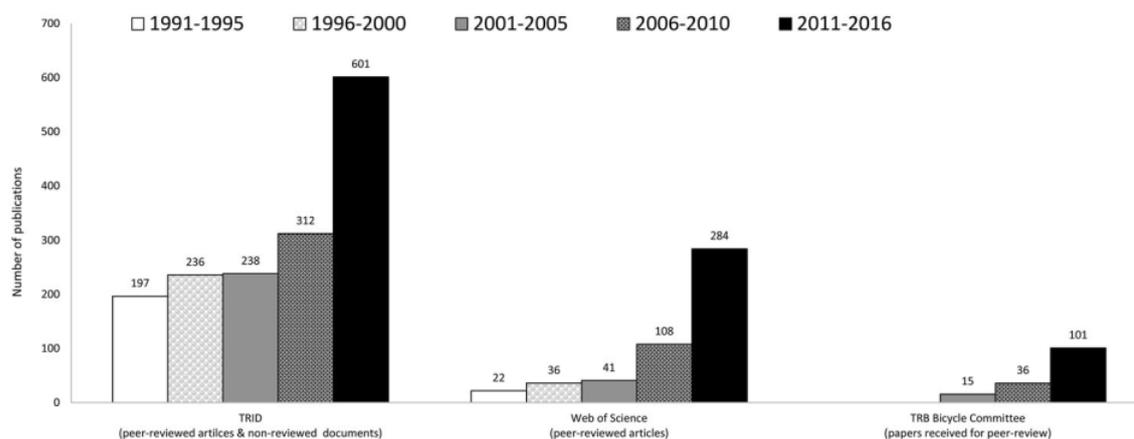


Figura 1.2: Número médio anual de publicações sobre ciclismo por quinquênios (PU-CHER; BUEHLER, 2017).

Com a prevalência crescente do ciclismo, tanto como atividade de lazer quanto como meio de transporte, as cidades têm consistentemente integrado esta prática em suas visões de futuro. Dado o destaque que o ciclismo vem ganhando tanto em debates populares quanto acadêmicos, torna-se crucial entender a experiência real dessa prática, além de realizar e difundir pesquisas de forma impactante e relevante (MAYERS, 2021).

1.2 MOTIVAÇÃO

Neste contexto, faz-se necessário o desenvolvimento de métodos capazes de sumarizar, selecionar e apresentar as informações contidas em dados esportivos, de modo a viabilizar uma melhor análise e exploração visual. No caso particular de ciclismo, há pesquisas em andamento para criar ferramentas de visualização capazes de revelar padrões relevantes deste esporte (DÍAZ; FORT; VÁZQUEZ, 2021; BECK; BURCH; WEISKOPF, 2016; FISTER; JR; FISTER, 2016; WOOD, 2015).

Gráficos de linha ou de barras podem ser amplamente usados na representação de dados esportivos. Embora as metáforas de linha tratem eficazmente certos aspectos dos dados, elas não escalam bem para tratar dados multidimensionais, resultando em problemas como a desordem e o congestionamento visual. Representar dados esportivos multidimensionais eficientemente e compreensivo, permitindo uma análise exploratória por parte dos usuários, está claramente além das capacidades de simples gráficos de linhas e de barras. Portanto, é necessário utilizar métodos adequados como as técnicas de visualização multidimensional (HEER; KONG; AGRAWALA, 2009).

De acordo com Beck, Burch e Weiskopf (2016), abundância de dados esportivos são medidos e armazenados usando sensores e dispositivos eletrônicos modernos. Esses dados, que geralmente variam ao longo do tempo, são registrados e podem ser usados para analisar o desempenho do atleta. Embora seja interessante analisar o desempenho de um único atleta, esta perspectiva às vezes ainda carece de uma referência para definir um

padrão de desempenho e para revelar o potencial de melhoria. McGarry (2009) afirma que o objetivo da análise científica do desempenho esportivo é promover a compreensão de um esporte, visando melhorar os resultados futuros.

É difícil estudar o desempenho de um ciclista com base em um modelo obtido de um único atleta, pois o desempenho do atleta é dependente de inúmeras variáveis, tais como tempo, clima, relevo, treinamento, peso, altura, força, idade, nutrição, etc. Uma maneira melhor é comparar o desempenho no ciclismo tendo como referência vários atletas. Uma comparação como essa poderia dizer o que um atleta faz melhor do que o outro ou poderia revelar diferentes estratégias de pedaladas. Além disso, muitos estudos se concentram predominantemente em percursos relacionados a competições profissionais, negligenciando assim o ciclismo não competitivo ou amador. Logo, encontrar padrões e anomalias na execução das atividades esportivas ajudará a entender melhor os resultados dos atletas, em especial os atletas de ciclismo, sendo interesse tanto dos próprios atletas quanto para outros tipos de profissionais como técnicos e comissão técnica.

1.3 PROBLEMA DE PESQUISA

Os trabalhos na literatura que utilizam visualizações no domínio de ciclismo focam na utilização de dados de competições para atletas profissionais e não consideram variáveis importantes de intensidade como frequência cardíaca e percentual de elevação em simultâneo. Essa proposta preenche essa lacuna ao utilizar dados de corridas não relacionadas a competições, ou seja, relacionadas a atletas amadores, e utiliza variáveis de intensidade para salientar novos padrões de informação acerca do desempenho dos atletas.

Por consequência da quantidade de atributos de intensidade e de informações como distância e localização geoespacial de várias pedaladas e ciclistas, a base de dados inerentemente é multidimensional (ou multivariada). Nesse sentido, técnicas de visualização multidimensional, geoespaciais e temporais são candidatas naturais na análise visual destes dados. Contudo, na literatura não há trabalhos que utilizem em conjunto tais técnicas para o contexto de atividades de ciclismo.

1.4 OBJETIVO GERAL E ESPECÍFICOS

A visualização de informação é um meio de comunicação que representa os dados em formato gráfico, visando simplificar a sua apresentação e interpretação. O desafio de transformar visualizações esportivas em informações compreensivas é um importante tópico de pesquisa e inclui uma ampla variedade de tarefas distintas.

O cerne desta pesquisa reside na criação do CycleVis, uma abordagem analítica especializada na análise visual esportiva de ciclismo. Diferentemente de ferramentas de monitoramento de uso diário, o CycleVis é projetado para a comparação de pedaladas, em contexto de ciclismo amador e também em ciclismo profissional. Esta abordagem visa auxiliar na exploração visual de dados multidimensionais que variam ao longo do tempo utilizando medidas de intensidade como frequência cardíaca e altimetria.

Para concretizar esta pesquisa, foi desenvolvida uma aplicação que não apenas integra técnicas de visualização multidimensional, mas também incorpora mecanismos de

interação avançados. Esses recursos permitem ao usuário reorganizar e interpretar dados de maneira flexível, otimizando o processo de extração de percepções valiosas sobre o desempenho dos ciclistas. Ao abordar desta forma, o CycleVis planeja preencher lacunas na literatura existente, oferecendo uma alternativa para análise de desempenhos dos ciclistas.

As informações resultantes serão exibidas em um layout de janelas justapostas, conhecido como *dashboard*, fundamentado em conceitos e propriedades de percepção visual. Em resumo, esperam-se as seguintes contribuições como objetivos específicos:

- Realizar um levantamento bibliográfico do estado da arte dos trabalhos científicos que abordam a visualização de esportes, prestando atenção especial ao ciclismo;
- Identificar e coletar base de dados que possuam atributos essenciais para o desempenho esportivo de ciclistas, tais como georreferenciamento, informação temporal e intensidade;
- Compilar e estruturar uma base de dados que contenha percurso de atletas de ciclismo;
- Desenvolver uma abordagem para avaliar e validar a abordagem com usuários especialistas de domínio e de computação.

1.5 QUESTÕES DE PESQUISA

Com o problema e os objetivos discutidos, surgiram então as questões de pesquisa que auxiliaram na condução do trabalho e desenvolvimento das respostas para solução do problema. O objetivo dessas questões foi conduzir e auxiliar na validação da solução proposta, dentre as quais destacam-se:

1. Quais atributos são essenciais para compreender o desempenho de um ciclista?
2. Como mapear tais atributos utilizando técnicas de visualização?
3. Como desenvolver uma abordagem que considere todas as informações e visualizações na mesma tela? Considerando tanto características dos ciclistas quanto das pedaladas?
4. Como realizar a avaliação da abordagem desenvolvida?

1.6 ESTRUTURA DO TRABALHO

Para documentar e demonstrar o que é produzido nesta dissertação, este documento está estruturado da seguinte forma:

- No Capítulo 2 descrevem-se os conceitos teóricos que alicerçam o desenvolvimento desta pesquisa, em especial as técnicas de visualização que servem de base no desenvolvimento desta abordagem visual. Além disso, os aspectos de *design* aplicados à visualização de dados. O capítulo prepara o leitor com um esclarecimento sobre os fundamentos teóricos utilizados nesta dissertação;

- No Capítulo 3 apresenta-se uma revisão da literatura sobre alguns dos trabalhos recentes sobre visualização esportiva, particularmente aqueles que tratam de ferramentas de visualização dedicadas à exploração esportiva de ciclismo. As características dessas visualizações foram analisadas, com destaque em suas vantagens;
- No Capítulo 4 é apresentada a abordagem de visualização desenvolvida nesta dissertação, desvendando ao leitor as suas capacidades e as suas funcionalidades. Explica-se sobre as metáforas visuais escolhidas e as possibilidades de extração de conhecimento.
- No Capítulo 5 é detalhado e discutido as avaliações da abordagem, tanto para um cenário de uso quanto para estudo com usuários.
- No Capítulo 6 é descrito a conclusão com as contribuições, limitações e trabalhos futuros.

REVISÃO BIBLIOGRÁFICA

2.1 CONSIDERAÇÕES INICIAIS

Para adentrarmos no desenvolvimento da proposta, é necessário estabelecer uma base conceitual da área que abrange esta proposta. Neste capítulo serão apresentados fundamentos teóricos da área de visualização de dados, as quais são necessárias para alcançar os objetivos e contribuições esperadas desta dissertação.

No tocante à organização do capítulo, a Seção 2.2 introduz o conceito de visualização de dados e seu *pipeline*. Em seguida, a Seção 2.3 trata dos mecanismos de interação, os meios para que as técnicas de visualização apoiem o processo de descoberta de informação pelos usuários. Na Seção 2.4, apresentam-se os conceitos das metáforas visuais escolhidas para essa proposta, em específico abrangendo a visualização de grafos, visualização temporal, visualização multidimensional e geovisualização. Na Seção 2.5, apresenta-se conceitos e métodos de usabilidade. Por fim, as considerações finais estão expostas na Seção 2.6

2.2 VISUALIZAÇÃO DE DADOS

Com os avanços tecnológicos, áreas como a engenharia, estatística, medicina, entre outras, passaram a gerar um alto volume de dados. Dessa forma, para se utilizar as informações dessa abundância de dados, faz-se necessário utilizar técnicas de mapeamento de informação, e um excelente caminho é a representação gráfica dos dados, ou seja, a visualização dos dados. É possível encontrar os termos “multidimensional”, “multivariado”, “multivalorado” para definir conjuntos de dados cujos elementos são compostos por uma coleção de parâmetros ou características. Ao longo desta dissertação, por convenção, será adotado o termo multidimensional.

Uma visualização de dados concentra-se em prover aos usuários compreensão das informações expostas, além de permitir que eles interajam com os dados. Telea (2014) explica que a visualização dos dados estimula a cognição e a inteligência humana, aumentando a capacidade de entender e conectar novos padrões. Consequentemente, a visualização de dados usa a percepção humana como primeira maneira de descobrir valor. Keim (2002)

afirma que a visualização é o processo de modificar a informação (muitas vezes textual) em formato visual, facilitando o processo de extração de informação. Assim os usuários conseguem perceber visualmente características que estão oclusas nos dados.

De acordo com Keim e Kriegel (1996), o sistema de percepção humana processa diferentes tipos de dados de forma bastante flexível. Os autores também argumentam que a integração de humanos e computadores no processo de análise visual pode ser muito frutífero para explorar grandes volumes de dados, ao permitir combinar a flexibilidade, criatividade e conhecimento geral dos seres humanos com a vasta capacidade de armazenamento e poder computacional dos dispositivos atuais.

O objetivo de usar a visualização de dados é transformar grandes quantidades de dados brutos, em representações gráficas que buscam aproveitar as aptidões naturais do cérebro humano (TRAINA, 2001). Nesse sentido, o *pipeline* de visualização (Figura 2.1) auxilia a entender a sequência de etapas necessárias para construir metáforas visuais. Inicia-se com os dados brutos que, após transformações (pré-processamento), formam os dados estruturados. Em seguida, os dados são mapeados visualmente em objetos gráficos utilizando variáveis visuais, depois transformados em visões, por meio de técnicas de visualização. Por fim, o usuário interage com as visualizações, interpreta os padrões visuais e extrai as informações presentes nesses padrões.

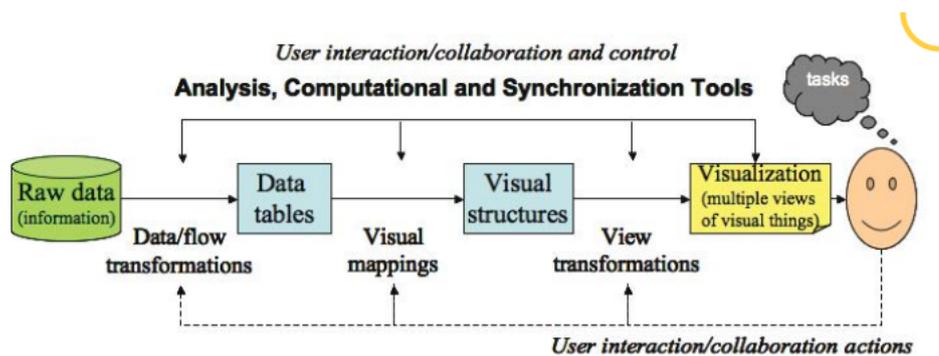


Figura 2.1: *Pipeline* de visualização proposto em Ward, Grinstein e Keim (2010).

2.3 MECANISMOS DE INTERAÇÃO

Ribarsky et al. (1999) afirmam que a visualização de dados visa integrar aspectos de *data mining* (processo de explorar grandes quantidades de dados à procura de padrões consistentes), computação gráfica e IHC (interfaces usuário-computador). Os mecanismos de interação possibilitam meios para que o processo de visualização de dados atenda de forma eficiente e eficaz às necessidades de interatividade dos usuários. Eick (2000) argumenta que a tela de um dispositivo possui um limite de tamanho, o que pode levar a problemas de escalabilidade visual quando a quantidade de informação apresentada é grande.

Os mecanismos de interação permitem a exploração dinâmica de dados, visualização simultânea e conjunta de várias representações gráficas (KEIM; KRIEGL, 1996; EICK,

2000). Isso significa que os dados são vistos de vários ângulos, portanto, são necessárias interações para superar a limitação física das telas, bem como manusear os pontos fortes e fracos das diferentes técnicas de visualização. Keim (2002) relata a importância da interação em visualizações coordenadas, ou seja, quando múltiplas visualizações são colocadas para explorar o mesmo conjunto de dados, de tal forma que ações em uma das visualizações desencadeia ações nas outras. Essa abordagem fornece mais informações se comparado a um sistema visual com um único tipo de visualização.

2.3.1 Interações Humanas

Como apresentada na Figura 2.1, as interações humanas definem as transformações de dados, os mapeamentos visuais, que transformam dados estruturados em representações visuais e as transformações visuais, que criam as visualizações renderizadas ao usuário. A representação visual combina vários elementos para a representação dos dados, tais como posicionamento, cores, forma, tamanho, entre outros. Estes elementos podem ser chamados de atributos visuais.

Almeida e Neto (2001) destacam que a implementação efetiva da interação em alguns dispositivos torna tão fácil explorar os dados que o próprio usuário pode reconhecer novos padrões sem a necessidade de usar algoritmos específicos. Os autores propõem as seguintes funcionalidades:

- Permitir o uso de vários atributos visuais para criar gráficos multidimensionais facilmente interpretáveis;
- Permitir a navegação interativa através da representação visual e o controle interativo sobre as funções;
- Permitir gerenciar interativamente os dados apresentados.

2.3.2 Símbolos Visuais

Card, Shneiderman e Mackinlay (1999) afirmam que a visualização de dados está diretamente ligada à percepção visual humana. Este entendimento depende da adequabilidade da visualização às limitações da percepção humana e tamanho das telas. Outrossim, a definição precisa de símbolos visuais que favorecem para maximizar a eficácia da análise visual não é trivial, a adaptação de símbolos visuais em sistemas visuais costuma ser uma tarefa com vários protótipos até que se chegue ao ideal.

Almeida e Neto (2001) afirmam ser importante observar como as informações são exibidas para os usuários poderem interagir de forma mais eficaz para maximizar a identificação de aspectos relevantes da visualização. Loyola e Meira (2018) relatam que para permitir que os usuários encontrem rapidamente pontos-chave na visualização, estes devem estar representados diferenciadamente dos demais, isso acelera a conscientização do usuário, e levará à identificação de novas tarefas analíticas. A Figura 2.2 mostra como alguns dados são representados por emblemas que facilitam a distinção entre eles. Tais emblemas são chamados de símbolos visuais.

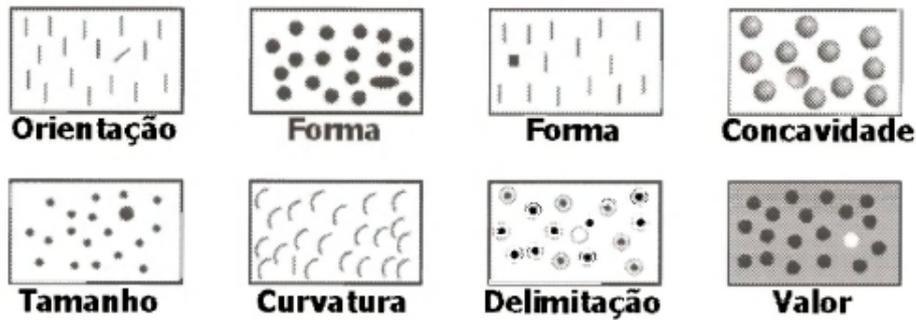


Figura 2.2: Atributos visuais que promovem fácil distinção (ALMEIDA; NETO, 2001).

No que tange a visualização que incluem mapas Ward, Grinstein e Keim (2010) expõem os símbolos visuais que deveriam ser considerados para a escolha de técnicas de visualização (Figura 2.3):

	Size	Shape	Brightness	Color	Orientation	Spacing	Perspective height	Arrangement
Point								
Linear								
Areal								

Figura 2.3: Variáveis visuais para visualizações que incluem mapas (SLOCUM et al., 2014).

Acrescentam-se as observações feitas por Card, Shneiderman e Mackinlay (1999), deve ser dada atenção à natureza dos dados. Certos tipos de dados são mais adequados para determinados elementos. Por exemplo, valores contínuos são melhor mapeados em posições ao longo de um eixo; dados categóricos devem receber emblemas ou formas diferentes, enquanto os dados numéricos podem ser melhor representados por tamanhos diferentes, ou seja, uma vez que se conheça a “forma” dos dados, é possível definir os atributos visuais apropriados.

Em Iliinsky e Steele (2011) encontra-se uma boa maneira de se pensar sobre os atributos visuais, dependendo do tipo de dados que seja necessário codificar (Figura 2.4).

Por exemplo, posicionamento e rótulos podem ser utilizados para codificar qualquer tipo de dado, enquanto forma e ícone são utilizados para dados categóricos. Tamanho, área, ângulo, saturação e brilho são mais adequados para dados ordinais.

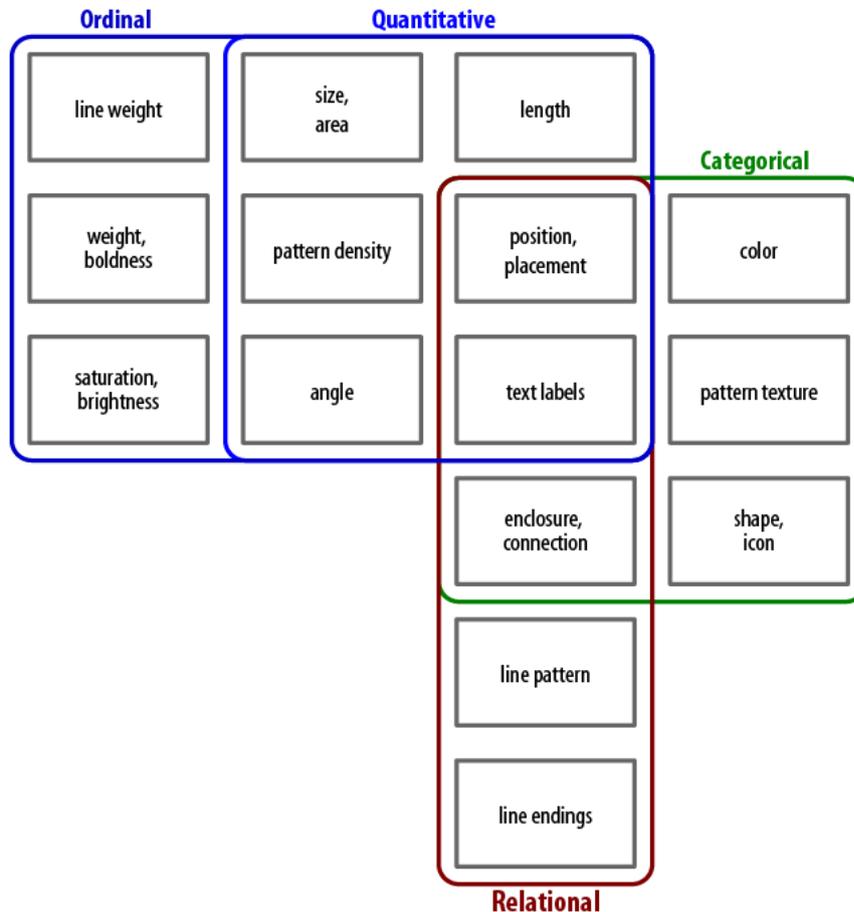


Figura 2.4: Atributos visuais que melhor se adequam a determinado tipo de dado (ILLINSKY; STEELE, 2011).

2.3.3 Cores

O trabalho de Tutida (1998) mostra que o mapeamento de cores é uma parte importante da visualização. É uma maneira fácil de aumentar a interpretação dos dados trazidos visualmente. Segundo a autora, o ponto-chave para o uso de cores na visualização de dados é a escolha adequada de uma combinação que enfatize as características importantes no conjunto de dados, enquanto uma combinação inadequada pode enfatizar os detalhes que não são importantes, criando artefatos visuais irrelevantes evocados pela interação causada entre as cores. O uso excessivo delas em uma mesma página causa fadiga visual, desvia a atenção do usuário do conteúdo e pode tornar o sistema visual ilegível se as cores de fundo e texto não são escolhidas adequadamente.

O espaço de cores que o sistema visual humano pode detectar é o tridimensional, o que significa que pode ser descrito corretamente com três eixos separados. A representação de cores mais comum em computação gráfica é o sistema onde as cores são especificadas como valores triplos de vermelho, verde e azul, o sistema RGB. No entanto, para o contexto de visualizações de dados a renderização no sistema de cor HSL é melhor (MUNZNER, 2014).

Nesse sistema, o eixo de *Hue* (Matiz) descreve a própria cor (sem mistura de branco ou preto), ou seja, o vermelho, azul, roxo etc. O eixo de *saturation* (saturação) representa a quantidade de branco misturado com a cor. Por exemplo, rosa é um vermelho menos saturado. O eixo *lightness* (brilho) é a quantidade de preto misturado com a cor. O modelo HSL pode ser visualizado como um cilindro, onde o brilho é mapeado pela altura, a saturação mapeada pelo raio e as cores (ou matiz) são distribuídas ao redor do círculo de 0 a 360 graus (Figura 2.5a).

Munzner (2014) apresenta a Figura 2.5b mostrando a sensibilidade espectral máxima (eficiência relativa de detecção de luz) do olho humano em condições de luz do dia, que está em uma faixa de comprimento de onda entre 400 e 700 nm. A sensibilidade espectral descreve as diferentes características dos foto pigmentos nos bastonetes e nos cones da retina do olho. Somos mais sensíveis aos comprimentos de onda verde e amarelo do que aos comprimentos de onda vermelho e azul (Figura 2.5b).

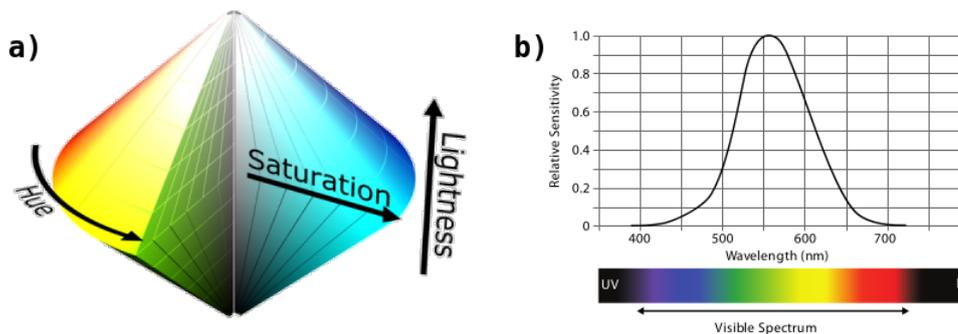


Figura 2.5: a) Modelo de cores no sistema HSL. b) Sensibilidade espectral (MUNZNER, 2014).

Tutida (1998) explica que além da correta correspondência de cores, a escolha de um modelo de cores para a visualização pode ser influenciada por outras questões. Por exemplo, a teoria *Gestalt*, ramo da psicologia, visa compreender e explicar o modo como o sistema cognitivo humano consegue reconhecer os estímulos recebidos, onde a percepção global de uma cena não pode ser explicada simplesmente pela soma de suas partes (WAGEMANS et al., 2012). A percepção cromática humana afeta a interpretação das visualizações, mas é um sistema complexo. Fatores como a distribuição heterogênea das células sensoriais no olho humano, a sensibilidade e alcance do espectro visível ao qual cada um está adaptado, a formação de cor no cérebro humano, exercem influência na percepção visual.

Munzner (2014) explica que a cor pode ser confusa na análise visual porque às vezes é usada para sinalizar tamanho e às vezes usada para sinalizar identidade. A autora aconselha que brilho e saturação devam ser utilizados para diferenciar magnitudes, enquanto a Matiz é usada para identificação. A combinação de brilho e saturação é melhor para apresentar dados ordenados. A Matiz é eficaz para revelar agrupamentos de dados, ideal para dados categóricos. Portanto, uma única cor pode ser usada para mostrar duas ou mais informações ao mesmo tempo (Figura 2.6).



Figura 2.6: Brilho e saturação são ordenados de forma mais natural pelo sistema perceptivo humano, mas a Matiz não é (MUNZNER, 2014).

2.3.4 Operações Visuais

Uma representação visual estática por si só muitas vezes não é suficiente para fornecer as condições necessárias e desejadas para uma melhor compreensão de grandes conjuntos de dados. Por meio de funcionalidades interativas, fornece-se ao usuário meios para que ele opere e explore a visualização. Portanto, novos aspectos do conjunto de dados podem então ser observados na visualização devido à execução dessas operações (FREITAS et al., 2001).

Freitas et al. (2001) também relatam que independentemente dos tipos de operações, sistemas de visualização devem oferecer, também, os meios que permitam limitar a quantidade de informação que os usuários recebem, mas, ao mesmo tempo, informá-los da amplitude do conjunto de dados. Em Shneiderman (1996) existe uma classificação contendo 7 operações ligadas a interação que um sistema de visualização deveria considerar, a saber:

- **overview:** Apresentação de uma visão geral do conjunto de dados. Pode ser adjacente a uma visão mais detalhada e conter um delimitador de observação para controlar esta visão detalhada. O uso consistente dessa estratégia permite que os usuários obtenham zoom em visualizações de médio alcance. A *overview* deve mostrar como os dados estão relacionados, mas depende de fatores limitantes específicos, como o tamanho das telas e do sistema perceptivo humano. A visão geral do conjunto de dados é útil para o usuário, por reduzir o tempo de busca, auxiliar na localização e escolha do próximo passo;

- **zoom:** Exibição pormenorizada de um elemento ou grupo de elementos de interesse. Normalmente, o usuário deseja ver mais detalhes na visualização para superar as limitações da tela. O zoom poderia ser ativado por uma barra de controle, pelo rolamento do mouse, pela indicação de um ponto de localização (foco) seguido de comandos de zoom (fator) consoante a preferência do usuário, entre outras formas. A ideia é dar ao usuário a possibilidade de se aproximar ou se afastar de um ponto de interesse na visualização;
- **filter:** Eliminação de informações desinteressantes para aquele momento. Ao remover itens indesejados, o usuário alcança mais controle do conteúdo da tela e pode então concentrar-se nos interesses de sua análise de forma mais rápida. A filtragem é geralmente combinada com zoom, sendo operada por botões, sliders ou widgets. Ao utilizar filtragem, a atualização rápida da tela é necessária, mesmo ao visualizar dezenas de milhares de itens;
- **details-on-demand:** Apresentação dos detalhes de um elemento ou grupo de elementos após a sua seleção. Ocorre porque os usuários desejam ver detalhes sobre itens específicos enquanto navegam no conjunto de informações. A abordagem usual é simplesmente clicar no elemento desejado e uma janela pop-up aparecerá com mais detalhes;
- **relate:** Observação da relação entre os elementos. Quando um elemento interessante é encontrado, os usuários podem solicitar informações sobre outros elementos com propriedades semelhantes. Desse modo, o sistema visual pode determinar quais elementos na visualização fazem parte de uma mesma relação, ao selecionar um elemento, outros elementos relacionados são destacados;
- **history:** Manter um histórico de execução para suportar operações de desfazer e repetir. A análise de dados é um processo de várias etapas e, por consequência, é de interesse manter o histórico das atualizações, permitindo que o usuário gereencie e compare estados anteriores;
- **extracts:** Exportação dos resultados das operações no sistema visual. Esses resultados podem ser subconjuntos de informações ou parâmetros de consulta. Após a obtenção das informações desejadas, o usuário pode exportá-las e salvar em um arquivo (para facilitar o uso da informação posteriormente), separando os resultados do contexto atual para uma análise mais específica;

2.4 TÉCNICAS DE VISUALIZAÇÃO

Uma das categorizações encontradas na literatura particiona as visualizações segundo o tipo de dado (WARD; GRINSTEIN; KEIM, 2010). Assim, dados temporais proporcionam um conjunto de técnicas similares de visualização, bem como para dados georreferenciados, com relacionamento, multidimensionais, entre outros. A seguir as categorias de técnicas de visualização que abarcam o contexto desse trabalho, a saber: técnicas de visualização com grafos, temporais, multidimensionais e geovisualização.

2.4.1 Visualização de Grafos

A visualização de Grafos (ou redes) mapeia visualmente as características de relacionamento entre seus objetos. Cada objeto é representado por um vértice (ou nó). A ideia principal desta categoria é visualizar grafos utilizando métodos que mapeiam as características de um grafo (ponderado/não-ponderado, direcionado/não-direcionado, cíclico/acíclico, etc.), de forma clara e rápida. A relação entre os vértices é representada por arestas (ligações). Tanto as arestas quanto os vértices podem ter números associados (pesos), então temos um grafo ponderado. Se a aresta tem uma direção (indicada por setas no grafo), temos um grafo direcionado. Alguns grafos possuem ciclos, ou seja, caminhos com comprimento mínimo de três vértices que possuem os mesmos vértices no início e no final (Figura 2.7).

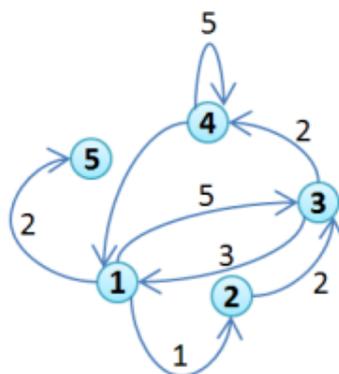


Figura 2.7: Exemplo de grafo direcionado com ciclos e pesos nas arestas e vértices.

Formalmente um grafo é um conjunto de vértices e arestas, representado por $G = (V, E)$, onde $V(G)$ representa um conjunto de vértices e $E(G)$ representam um conjunto de arestas. Para dois vértices $u, v \in V(G)$, uma aresta que os conecta é denotada por $(u, v) \in E(G)$. $|V|$ e $|E|$ indicam, o número de vértices e arestas, respectivamente, em G (CINTRA, 2019). O número de arestas incidentes a um vértice determina o seu grau. Os grafos podem ser usados para interpretar a estrutura de uma rede através da procura de qualquer agrupamento de vértices, quão bem conectados são os vértices ou como o *layout* do diagrama está organizado.

Para visualizar grafos Munzner (2014) e Ward, Grinstein e Keim (2010) explicam sobre a técnica de diagramas de ligação de vértices. Esta técnica procura uma localização para os vértices usando uma abordagem de posicionamento. E a abordagem mais comum para determinar o posicionamento dos vértices utilizando esta técnica é a baseada em força. Esta abordagem alinha os vértices de acordo com uma simulação de forças físicas, onde os vértices se afastam uns dos outros enquanto as arestas atuam como molas.

Os autores concordam que muitos algoritmos com a abordagem baseada em força começam colocando vértices aleatoriamente em uma região espacial e, em seguida, refinam iterativamente suas localizações com base na impulsão e tração dessas forças simuladas

para melhorar gradualmente a visualização. Usam-se símbolos como ícones, círculos ou imagens para representar os vértices e linhas para as arestas. A Figura 2.8a disponível online¹, com livre acesso, simula o uso da metodologia baseada em forças, colocando personagens aparecendo em simultâneo nos capítulos da obra *Os Miseráveis* em posições mais próximas. As cores representam grupos de personagens.

Uma fraqueza desta abordagem é que os **layouts** do grafo são muitas vezes não determinística. O problema com codificações visuais não determinísticas é que a aleatoriedade pode levar a diferentes relações de proximidade a cada vez que o algoritmo é executado, onde as distâncias entre os vértices refletem o valor inicial escolhido aleatoriamente. Outro ponto é a escalabilidade, existe um tempo necessário para posicionar todos os vértices. A abordagem baseada em força cria **layouts** legíveis rapidamente para grafos com dezenas de vértices, mas pode ser difícil entender as relações estruturais do grafo com milhares de nós. Munzner (2014) argumenta que é improvável que esta abordagem produza bons resultados quando o número de arestas for maior que quatro vezes o número de vértices, deixando o grafo emaranhado como se fosse uma bola de cabelo (Figura 2.8b).

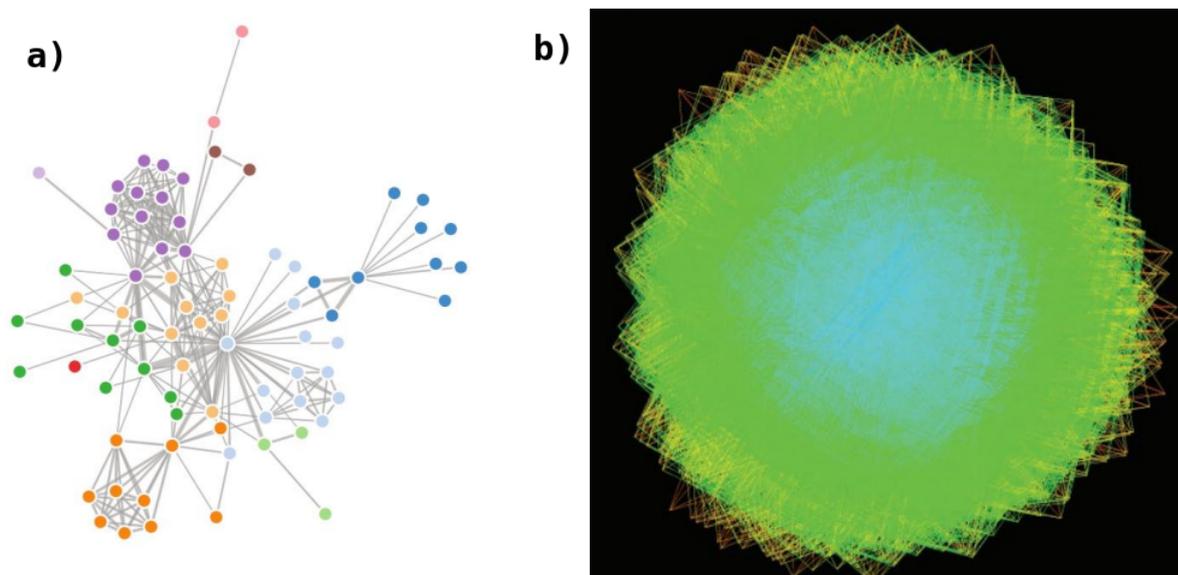


Figura 2.8: a) Grafo direcionado por força. b) Um grafo enorme de 26.028 vértices e 100.290 arestas sem muita estrutura visível. (MUNZNER, 2014)

Em Dwyer, Marriott e Wybrow (2008), há uma proposta de abordagem baseada em restrições que é uma extensão da abordagem baseado em força, usando critérios que limitam a localização dos vértices. As restrições incluem alinhamento horizontal e vertical, vértices não sobrepostos, direção das arestas e proximidade entre grupos de vértices, um software disponível² que implementa essa abordagem pode ser instalado e utilizado. Em Muelder e Ma (2008) existe uma proposta de abordagem multiescala, variante do método

¹Disponível em: <https://bl.ocks.org/mbostock/4062045>. Acessado em 08/06/2022

²Disponível em: <https://github.com/mjwybrow/dunnart>. Acessado em 08/06/2022

baseado em força, que decompõe o grafo em níveis aninhados. Os próximos níveis são subgrafos do original aos quais vários vértices são adicionados dependendo da interação do usuário e assim aumentar os detalhes da visualização. Estas abordagens procuram ser mais rápidas que a abordagem baseada em força. Finalmente, Munzner (2014) argumenta que existem muitas variantes da abordagem de posicionamento baseado em força, e que os algoritmos são geralmente projetados para minimizar as distrações na visualização, como cruzamentos de vértices de borda e sobreposições de vértices.

Outra representação de grafos comum é utilizando uma técnica matricial. Munzner (2014) aborda a técnica de matrizes adjacentes. Esta técnica posta os vértices do grafo ao longo de um eixo vertical e horizontal, e as arestas entre os vértices são indicados colorindo-se a célula da matriz que representa a interseção entre os vértices (Figura 2.9).

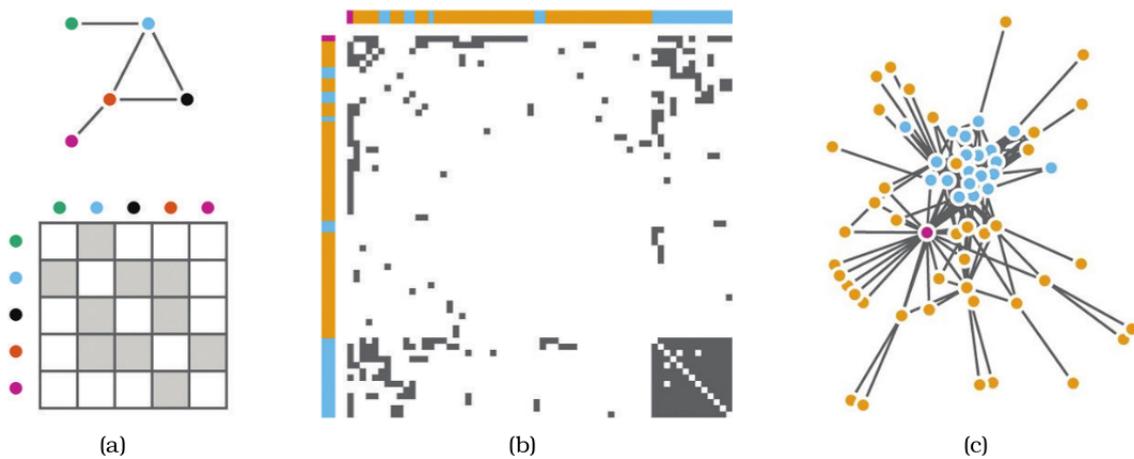


Figura 2.9: a) Visão matricial de um grafo pequeno. b) Visão matricial do grafo em c). Grafo de uma rede maior (MUNZNER, 2014).

A matriz de adjacência permite remover a sobreposição de arestas que ocorre na técnica de ligação de vértices, permite a visualização de grafos com alta densidade. Enquanto as visualizações que utilizam a ligação de vértices suportam grafos com o número de arestas até quatro vezes o número de nós, as visualizações matriciais suportam grafos onde o número de arestas é igual ao número de vértices ao quadrado.

Outro ponto forte das visualizações matriciais é a sua previsibilidade, estabilidade e suporte para reordenamento. As visualizações matriciais podem ser dispostas em um espaço previsível na tela, enquanto as visualizações de ligação de vértices podem exigir uma quantidade de espaço desconhecida, dependendo do tamanho do grafo. Adicionar um novo vértice à visualização matricial causará apenas uma pequena mudança visual, mas em um grafo visualizado pela técnica de ligação de vértices utilizando com posicionamento baseado em força para separar os vértices poderá fazer uma grande diferença, pode ser necessário recalcular o posicionamento de todos os vértices do grafo.

As visualizações matriciais também se destacam por estimar rapidamente o número de vértices em um grafo e oferecer suporte a pesquisas mais rapidamente. Encontrar um

vértice em uma matriz ordenada é mais fácil do que encontrar um vértice que possa ser colocado em qualquer lugar na área bidimensional. Um dos pontos fracos da visualização matricial de grafos é a falta de familiaridade, os usuários interpretam facilmente o diagrama de ligação de vértices para grafos menores sem precisarem de treinamento algum, mas precisam geralmente de algum treinamento para interpretar visualizações de grafos matriciais.

Também estão relatados na literatura visualizações de grafos que combinam as vantagens das técnicas de matriz de adjacências e ligação de vértices. Em Henry, Fekete e McGuffin (2007) é apresentado o *NodeTrix* uma representação híbrida de redes, onde a ligação de vértices é utilizada para mostrar a estrutura global da rede, enquanto os vértices em forma de matriz de adjacências permite detalhar as regiões de uma rede com alta densidade de informações (Figura 2.10).

Segundo o autor, este projeto dispõe de mecanismos de interação que permitem mover, extrair e ampliar os vértices do grafo. Os vértices representam comunidades de pesquisa. Os links entre os vértices indicam quais autores conectam diferentes comunidades de pesquisa. A cor roxa nos vértices do grafo representa a colaboração entre diferentes pesquisadores de uma mesma comunidade. A intensidade das cores representa a força de cada colaboração em uma comunidade. (HENRY; FEKETE; MCGUFFIN, 2007).

2.4.2 Visualização de Séries Temporais

Embora o tempo seja inerentemente contínuo, é possível realizar uma representação computacional discreta do tempo, de modo que as mudanças no tempo correspondem a uma série de intervalos consecutivos de duração, formando uma série temporal. Ortigossa (2018) afirma que as séries temporais são produzidas continuamente e podem ser encontradas em áreas como finanças, saúde, meteorologia, astronomia, etc. O autor apresenta uma definição formal para séries temporais:

$$D = \{(t_1, v_1), (t_2, v_2), \dots, (t_n, v_n)\} \quad (2.1)$$

com

$$v_i = f\langle t_i \rangle \quad (2.2)$$

onde v_i é caracterizado por um único valor, seja qualitativo ou quantitativo, o conjunto é chamado de bidimensional, pois a evolução temporal de seus pares ordenados (t_i, v_i) pode ser representada no plano cartesiano. No entanto, cada instância v_i também pode ser individualmente composta por um subconjunto contendo uma quantidade arbitrária de valores, $v_i = \{v_{i1}, v_{i2}, \dots, v_{ik}\}$. Estes k atributos descrevem as características das variáveis dependentes do conjunto de dados. Neste caso, tal série temporal será denominada de multidimensional, pois suas instâncias são interpretadas como pontos no espaço m -dimensional, em que m é o número total de elementos em (t_i, v_i) .

Maletzke (2009) relata que muitos métodos de análise de séries temporais focam na decomposição de componentes, uma vez que esses componentes são considerados características importantes de uma série temporal, o autor apresenta três características importantes que deveriam ser compreendidas antes de se analisar séries temporais, a saber:

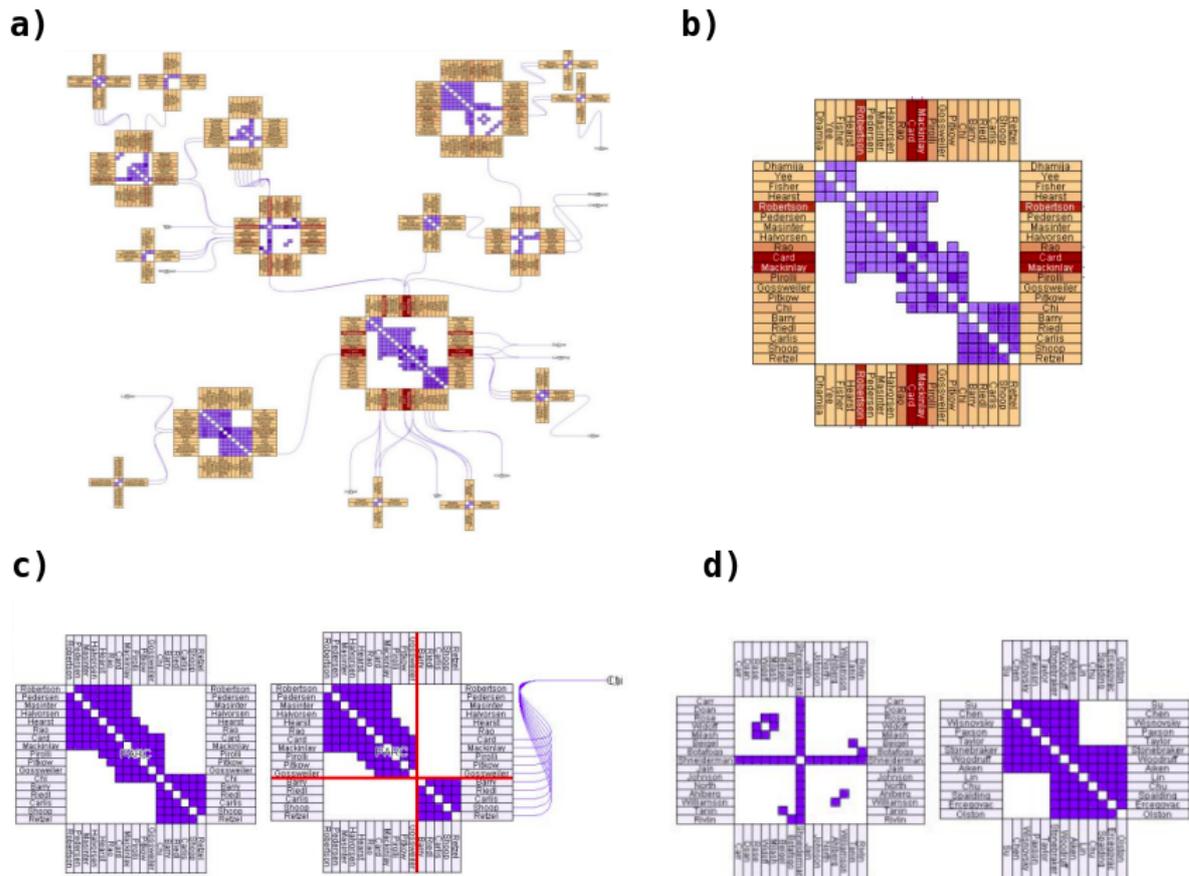


Figura 2.10: a) Um grafo representando a rede de pesquisa do campo de Visualização de dados no concurso InfoVis 2004. b) Os tons em amarelo demonstra a força de colaboração de cada pesquisador, em roxo a força de colaboração entre diferentes pesquisadores de uma mesma comunidade. c) Revela como o autor Ed Chi serve de ponte entre dois subgrupos em uma comunidade de pesquisa. d) Revela padrões de colaboração em comunidades de pesquisa, há comunidades de pesquisa em que existem colaborações entre vários pesquisadores, como também comunidades de pesquisa em que há pouca colaboração entre os pesquisadores (HENRY; FEKETE; MCGUFFIN, 2007).

- **tendência:** Compreende o movimento dominante de uma série temporal, como a série temporal se desenvolve em um longo intervalo de tempo. Possui um comportamento unidirecional, seja crescente ou decrescente. Para ilustrar essa característica, na Figura 2.11a é apresentada uma série temporal demonstrando o número mensal de passagens aéreas vendidas nos Estados Unidos em um período de dez anos, a linha pontilhada representa a tendência estimada por meio de uma regressão linear;
- **sazonalidade:** Refere-se a movimentos similares que ocorrem regularmente em períodos fixos de tempo, como semanas, dias, quinzenas e anos. A Figura 2.11b apresenta um comportamento similar a cada dois anos em uma série temporal que

representa a quantidade de passagens aéreas vendidas nos Estados Unidos entre 1949 a 1960.

- **resíduo:** Os movimentos aleatórios ou irregulares que correspondem a deslocamentos esporádicos de uma série temporal, que não podem ser captados pela tendência ou sazonalidade. A Figura 2.11c apresenta o comportamento residual extraído da série temporal referente à Figura 2.11b

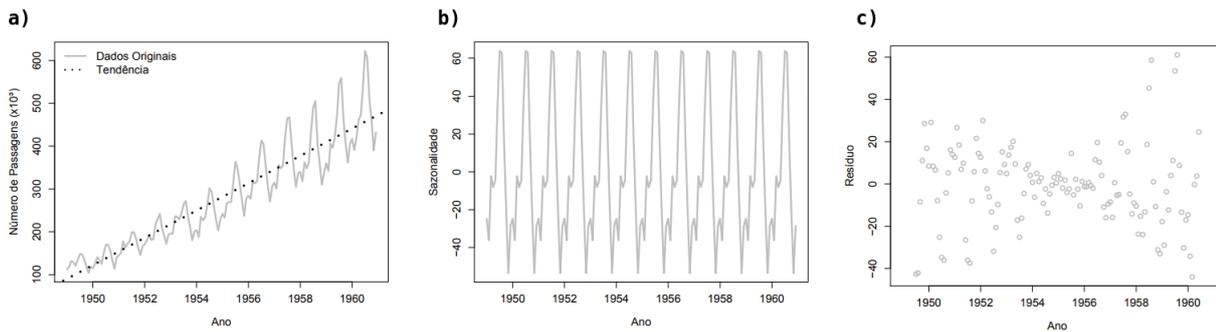


Figura 2.11: a) Série temporal da quantidade de passagens aéreas vendidas nos Estados Unidos no período de 1949 a 1960. A linha pontilhada representa uma estimativa de tendência existente nesses dados. b) Variação sazonal existente na série temporal referente à quantidade de passagens aéreas vendidas nos Estados Unidos no período de 1949 a 1960. c) Resíduos extraído da série temporal referente à quantidade de passagens aéreas vendidas em um período de onze anos (MALETZKE, 2009).

Desse modo, a compreensão das características de uma série temporal provê uma base conceitual para selecionar uma técnica de visualização que seja expressiva, apropriada e eficaz para construir visualizações temporais, identificando características intrínsecas de uma série temporal como a existência de padrões cíclicos nos dados amostrados.

A Figura 2.12 revela a importância da característica temporal para a construção da visualização e conseqüentemente extração de informações. Ela expõe três representações visuais diferentes do mesmo conjunto de dados temporal, que contém o número de casos de gripe que ocorreram no norte da Alemanha durante um período de três anos. Os dados mostram uma forte tendência cíclica. A imagem mais à esquerda a Figura 2.12a usa um gráfico de linha simples para visualizar os dados. Isso facilita visualizar os picos de gripe ao longo do tempo. Mas somente com a visualização em espiral mais à direita na Figura 2.12c é possível detectar um comportamento cíclico dos casos de gripe. No caso da espiral mostrada na figura 2.12b, o ciclo não é visível, e isso se deve ao fato de ter sido definida um padrão cíclico de 24 dias, já a Figura 2.12c está adequadamente parametrizada com um padrão cíclico de 28 dias, desvendando imediatamente a tendência periódica presente nos dados (WARD; GRINSTEIN; KEIM, 2010).

Uma das técnicas mais comuns para visualizar séries temporais é o gráfico de linha, o qual é comumente usado para exibir valores quantitativos em um intervalo contínuo ou

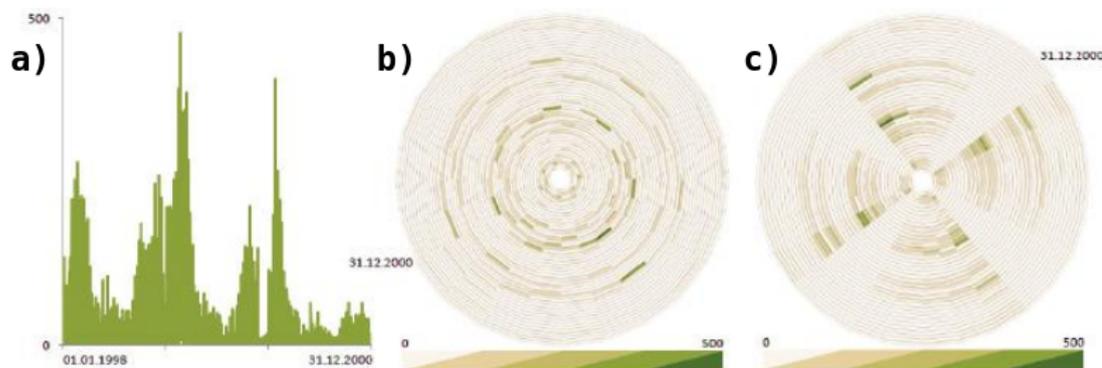


Figura 2.12: Representação linear contra representação cíclica do tempo: diferentes percepções podem ser obtidas (WARD; GRINSTEIN; KEIM, 2010).

período. Possibilita mostrar tendências e analisar como os dados foram modificados ao longo do tempo. Os gráficos de linha são criados plotando-se primeiramente os pontos de dados em uma grade de coordenadas cartesianas, em seguida conecta-se todos esses pontos com uma linha. Usualmente, o eixo Y tem um valor quantitativo, enquanto o eixo X é uma escala de tempo. Os valores negativos podem ser exibidos abaixo do eixo X (Figura 2.13a).

A técnica de espiral de série temporal visualiza dados de tempo usando a espiral de Arquimedes. O gráfico começa no centro de uma espiral e depois avança para fora. Os gráficos em espiral são versáteis e podem usar barras, linhas ou pontos para serem exibidos ao longo do caminho da espiral. Ainda, são ideais para exibir grandes conjuntos de dados e mostrar tendências de longo prazo, facilitando a exibição de padrões periódicos. A cor pode ser personalizada para cada período dividindo a espiral e permitindo a comparação entre cada período. Em Weber, Alexa e Müller (2001) há um bom estudo sobre esta técnica (Figura 2.13b).

Para visualizar séries temporais multidimensionais Keim (2002) sustenta que extrair informações dessas séries é uma tarefa complexa, por serem encontradas nos mais variados formatos e podem ser medidas com diferentes taxas de amostragem em diferentes dimensões, portanto se faz necessário a utilização de metáforas visuais mais robustas.

Ortigossa (2018) argumenta que nesse contexto de séries temporais multidimensionais, é prática comum estudar a evolução de múltiplas séries de uma só vez. A Figura 2.14 realiza uma comparação de legibilidade individual ao se analisar múltiplas séries temporais em um simples gráfico de linhas. Ao aumentar o número de curvas representadas dentro do mesmo espaço, a capacidade de transmitir informações individuais é bastante reduzida.

Em Havre et al. (2002) é apresentado o *Theme River*, que recebeu o nome de *Stream Graph* em Byron e Wattenberg (2008). Essa visualização é uma variação do gráfico de áreas empilhadas, e visualiza séries temporais multidimensionais empilhando-as uma sobre a outra ao longo dos instantes de tempo. Os fluxos coloridos que fluem dentro do

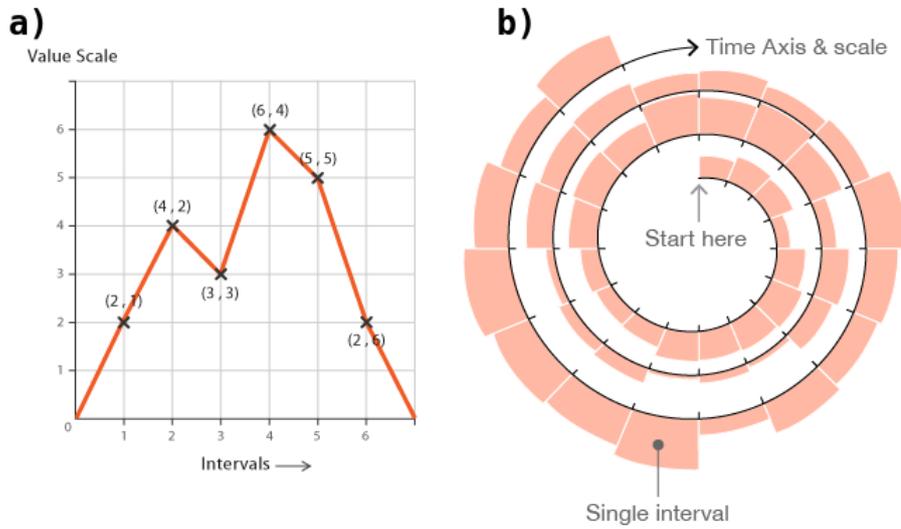


Figura 2.13: a) Exemplo de gráfico de linhas b) Exemplo de Espiral de série temporal (RIBECCA, 2013).

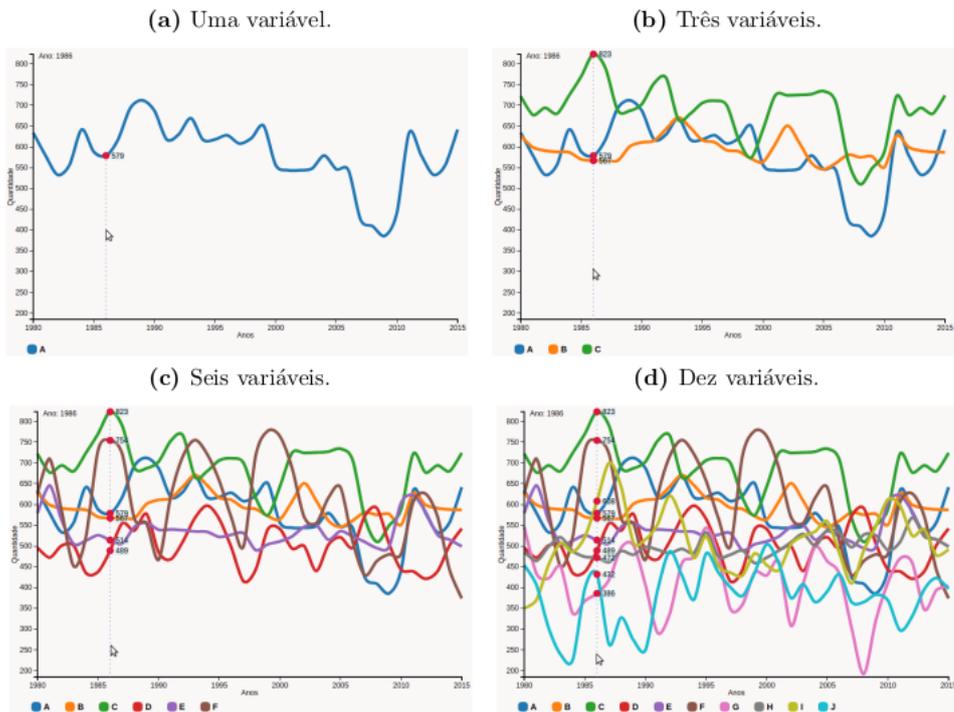


Figura 2.14: Legibilidade do gráfico de linhas (ORTIGOSSA, 2018).

“rio” estreitam ou alargam para indicar diminuições ou aumentos nos dados ao longo do tempo. Os fluxos de certa forma se assemelham a um rio que flui da esquerda para a

direita. Isso torna os *Stream Graphs* esteticamente agradáveis e mais atraentes de se ver. Em um *stream graph*, o tamanho de cada forma de fluxo individual é proporcional aos valores de cada categoria. A desvantagem do *Stream Graph* é não haver nenhum eixo horizontal para usar como referência (Figura 2.15).

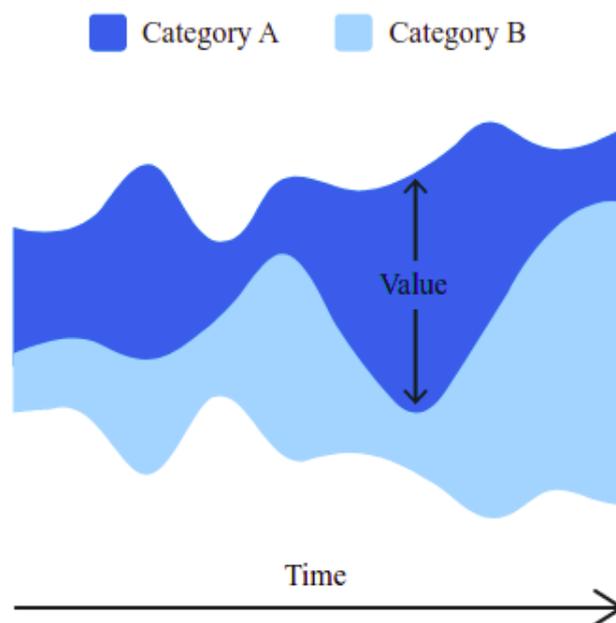


Figura 2.15: Exemplo de stream graph (RIBECCA, 2013).

A Figura 2.16 mostra uma visualização dos discursos de Fidel Castro de 1959 a 1961, destacando com mais frequência os temas abordados por Fidel. Na parte superior da Figura 2.16, há uma sequência cronológica de eventos relativos ao período histórico. Observa-se que os discursos que envolvem o tema armas na cor ciano, são fracos em novembro de 1959, por outro lado, esse tema ganha força em dezembro de 1959. O tema das armas desaparece entre os meses de janeiro e fevereiro de 1960, mas reaparece em março de 1960. Outro ponto que se observa é que no intervalo de tempo da visualização, as falas de Fidel seguem uma linha com temas que envolvem guerrilhas e conflitos.

Finalmente, caso o leitor interessado em conhecer outras técnicas, menciona-se o *Visual Vocabulary*³ e o *DataViz Catalogue*⁴. Em *TimeViz*⁵ existe um repositório interativo para apoiar pesquisadores e profissionais da área de visualização a encontrar visualizações de séries temporais, acompanhadas de leituras, exercícios, *surveys*, *reviews* relacionados.

³Disponível em: <https://ft-interactive.github.io/visual-vocabulary/>. Acessado em 09/06/2022

⁴Disponível em: <https://datavizcatalogue.com/>. Acessado em 09/06/2022

⁵Disponível em: <https://vcg.informatik.uni-rostock.de/ct/timeviz/timeviz.html>. Acessado em 10/06/2022

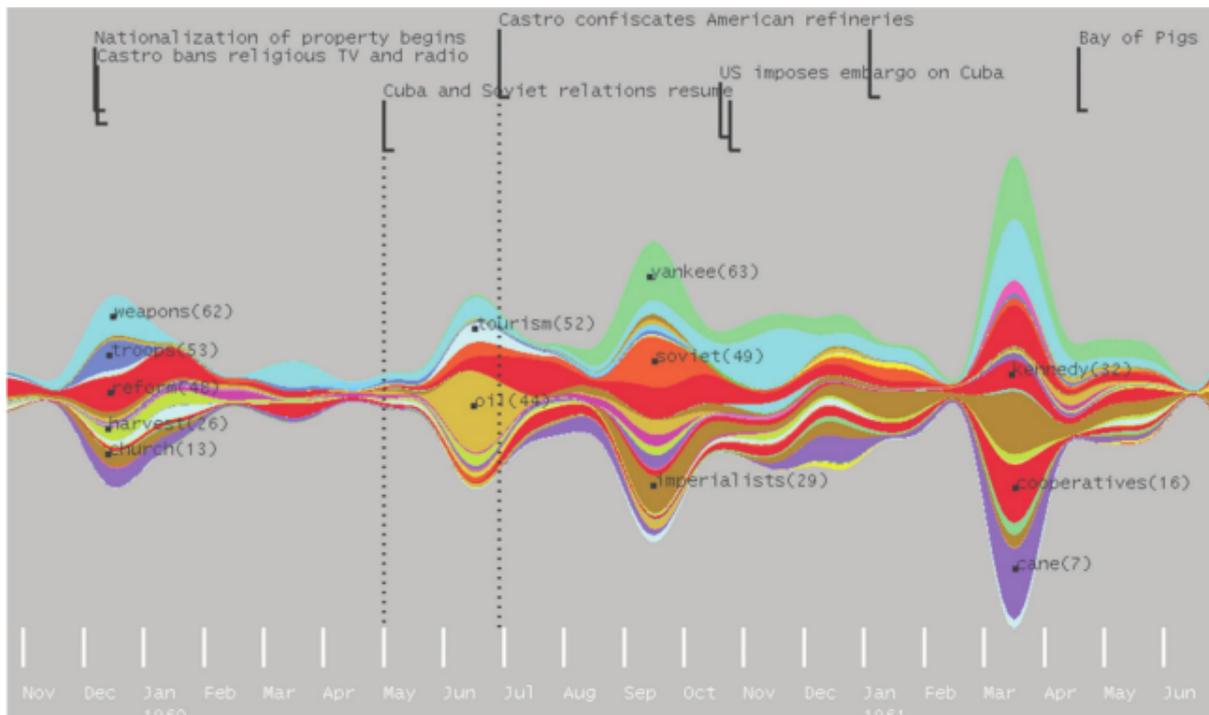


Figura 2.16: Stream Graph proposto em (HAVRE et al., 2002).

2.4.3 Visualização Multidimensional

A visualização de dados multidimensionais é um tema importante no campo da visualização de dados e enfrenta dois problemas: a apresentação de dados multidimensionais no espaço bidimensional ou tridimensional, e as dificuldades do ser humano em compreender esse tipo de dado. Esta dissertação descreve alguns dos principais métodos de visualização de dados multidimensionais seguindo a divisão encontrada em Ward, Grinstein e Keim (2010), a saber:

- ***Point-Based Techniques:*** Em um método baseado em pontos, a visualização exibe valores de um espaço n -dimensional para um espaço visual k -dimensional. Para cada valor, uma representação gráfica, símbolo ou outro objeto estético é desenhado no seu ponto k -dimensional correspondente. O n define as dimensões originais do conjunto de dados;
- ***Line-Based Techniques:*** Nos métodos baseados em linhas, os pontos correspondentes em um determinado valor ou tamanho são conectados por linhas retas, ou curvas. Essas linhas não apenas aprimoram a relação entre valores dos dados, mas também conferem propriedades reconhecíveis aos dados por meio de inclinações, curvaturas, interseções e outros padrões de linha.
- ***Region-Based Techniques:*** Nesta categoria, os polígonos preenchidos são usados para transmitir valores, com base em seu tamanho, forma, cor ou outro atributo.

Várias técnicas baseadas em região foram inicialmente projetadas para dados unidimensionais como gráficos de pizza e barras, sendo posteriormente estendidos para uma versão multidimensional.

- ***Combinations of Techniques:*** Além das técnicas baseadas em pontos, linhas ou regiões, existem técnicas híbridas que combinam características de duas ou mais das técnicas anteriores.

Munzner (2014) apresenta o *Scatterplot Matrix* (Matriz de gráficos de dispersão), uma técnica baseada em pontos, uma variante do *scatterplot* (gráfico de dispersão), onde cada célula da matriz contém um gráfico de dispersão inteiro. Com esta técnica é possível visualizar todas as possíveis correlações das variáveis postas nas linhas e colunas da matriz. A correlação representa como uma variável tende a aumentar ou diminuir, quando a outra variável com a qual se relaciona aumenta ou diminui. O triângulo acima ou abaixo da diagonal da matriz exibe as correlações que se deseja visualizar, a diagonal pode ser omitida para evitar a exibição de uma variável plotada contra si mesma ou utilizada para exibir histogramas ou informações adicionais sobre os dados.

Cada célula do *Scatterplot Matrix* requer espaço suficiente para que se plote um ponto para cada item de forma discernível, portanto a escalabilidade dessa técnica de visualização possui limitações. A Figura 2.17 ilustra um exemplo de *Scatterplot Matrix* com dezesseis gráficos de dispersão aplicados no conjunto de dados Íris. As variáveis, largura da pétala, largura da sépala, comprimento da pétala e comprimento da sépala são mapeadas par a par, gerando uma matriz 4×4 . A cor dos pontos identifica a espécie da flor, azul (setosa), laranja (versicolor) e verde (virgínica).

Infere-se da célula destacada em vermelho que a espécie setosa possui tamanhos de sépalas menores, mas largas de sépalas maiores, a espécie versicolor está no meio das outras duas espécies em termos de comprimento e largura da sépala, e a espécie virgínica possui tamanho de sépalas maiores, mas largas de sépalas menores, essa mesma interpretação pode ser extraída sobre as outras espécies presente na visualização.

A metáfora de coordenadas paralelas, uma técnica baseada em linhas, tornou-se notável após o trabalho de Inselberg e Dimsdale (1990). É usada para plotar dados numéricos multidimensionais, são ideais para comparar muitas variáveis e ver as relações entre elas. Por exemplo, é possível comparar uma série de produtos com os mesmos atributos (comparando especificações de computadores ou carros em diferentes modelos). Nessa visualização cada variável recebe seu próprio eixo e todos os eixos são colocados em paralelo entre si. Cada eixo pode possuir uma escala diferente porque cada variável trabalha com uma unidade de medida diferente, ou todos os eixos podem ser normalizados para manter todas as escalas uniformes. Os valores são plotados como uma série de linhas interconectadas em todos os eixos. Isso significa que cada linha é um conjunto de pontos colocados em cada eixo, que foram todos conectados.

A desvantagem fica por conta da sobreposição das linhas. A melhor maneira de resolver esse problema é usar interatividade como o *Brushing* (destacar uma linha selecionada ou uma coleção de linhas enquanto esmaece todas as outras), isso permitirá foco em seções do gráfico em que se está interessado. Outra desvantagem está associada a descoberta

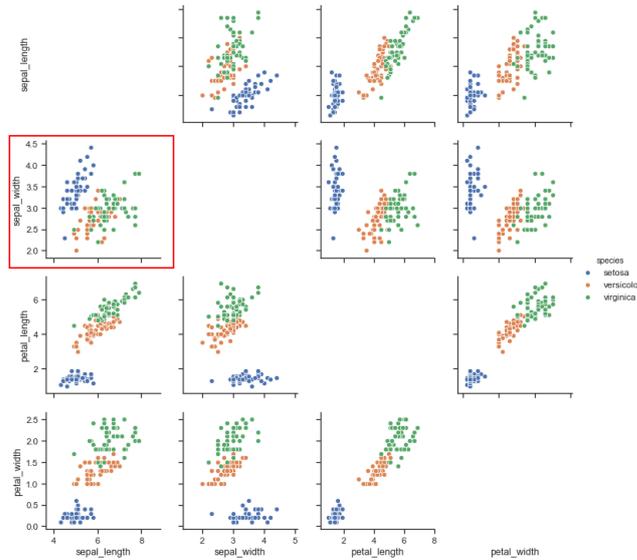


Figura 2.17: Exemplo de correlações que podem ser inferidas do *Scatterplot Matrix*.

de padrões entre eixos que não são vizinhos, mas a interatividade pode aproximar eixos distantes. (Figura 2.18).

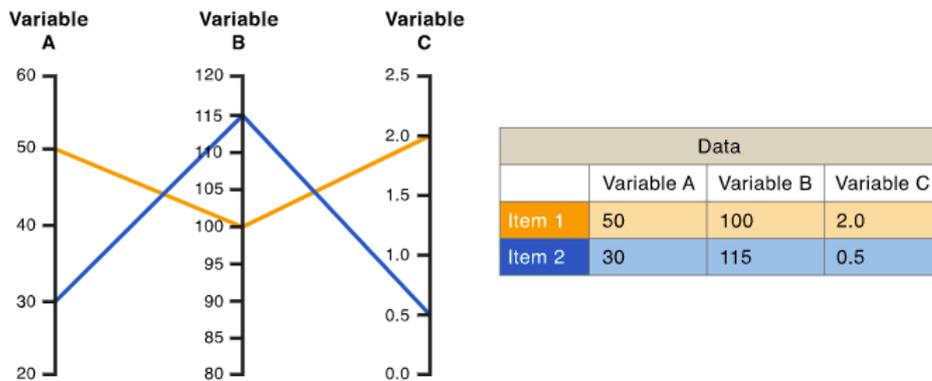


Figura 2.18: Exemplo de Coordenadas Paralelas (RIBECCA, 2013).

O *Radar Chart* (Gráfico de radar), outra técnica baseada em linhas, é uma forma de comparar múltiplas variáveis quantitativas. Isso o torna útil para ver quais variáveis possuem valores semelhantes ou se existem *outliers* entre cada variável. O gráfico de radar permite ver quais variáveis estão pontuando alto ou baixo em um conjunto de dados, tornando-o ideal para exibir o desempenho.

Cada atributo representa um eixo que sai do centro. Todos os eixos são dispostos radialmente com igual distância entre eles e mantendo a escala entre todos os eixos. Cada instância representa linhas que cruzam os eixos nos pontos dos seus valores para aquele atributo, formando um polígono por instância (Figura 2.19a). Vale destacar que

ter múltiplos polígonos em um único gráfico de radar torna a leitura difícil, confusa e muito desordenada. Especialmente se os polígonos forem preenchidos, pois o polígono superior cobre todos os outros polígonos por baixo dele (Figura 2.19b).

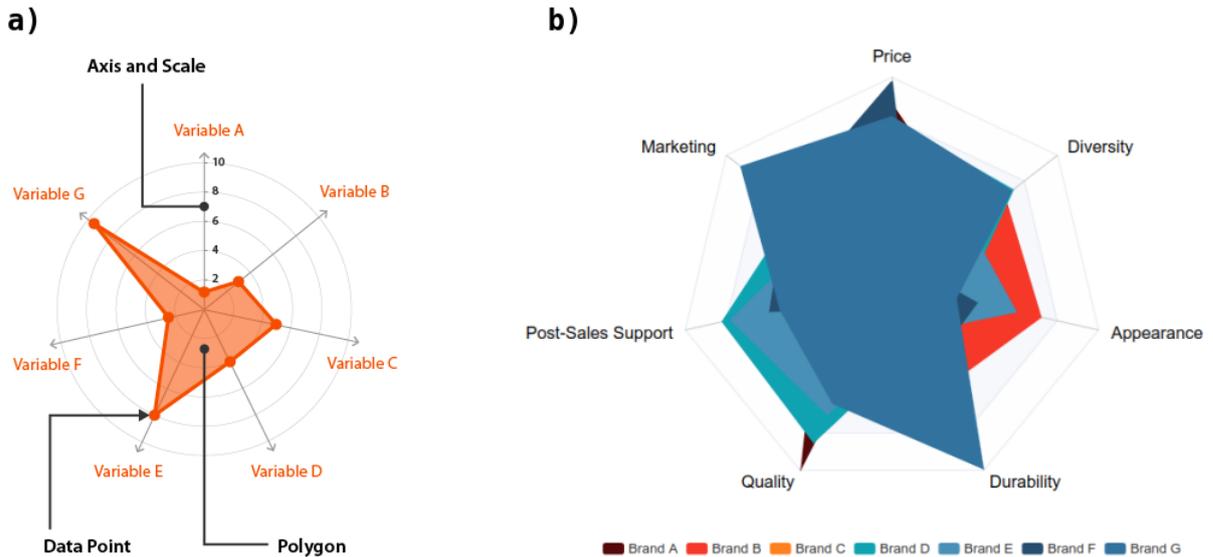


Figura 2.19: a) Exemplo de Gráfico de Radar (RIBECCA, 2013). b) Gráfico de Radar com muitos polígonos.

Ward, Grinstein e Keim (2010) argumenta que dados multidimensionais também podem ser visualizados em tabelas, e os métodos que utilizam essa estrutura variam principalmente nos tipos de interações que suportam. O *Heatmap* (Mapas de Calor) uma técnica baseada em região, visualizam os dados usando diferentes esquemas de cores. Quando aplicados a um formato tabular, os mapas de calor são úteis para o exame cruzado de dados multidimensionais, posicionando variáveis nas linhas e colunas e colorindo as células na tabela. Os mapas de calor são úteis para mostrar a variação entre as variáveis, revelando quaisquer padrões, exibindo se alguma variável é semelhante entre si e para detectar se há correlações entre elas. Os dados contidos em uma célula dependem do relacionamento entre linhas e colunas associadas. Uma legenda pode ser necessária ao lado de um mapa de calor para que ele seja lido corretamente (Figura 2.20).

O *Stacked Area Graph* (Gráfico de área empilhado), outra técnica baseada em região e que se assemelha ao *stream graph*, é uma técnica que divide o mesmo espaço, empilhando uma dimensão sobre a outra. O resultado é um resumo visual dos valores dando uma visão agregada de todas as séries, o gráfico inteiro representa uma visualização multidimensional. Usualmente, são úteis para comparar dados que variam em um intervalo, em situações que a soma ou agregação das séries multidimensionais é tão, ou mais importante quanto a individualização de determinada série (Figura 2.21).

Como uma abordagem baseada em região, o *Table Lens* visa explorar dinamicamente grandes quantidades de dados tabulares. Simplificando, essa visualização é uma maneira de agrupar relacionamentos usando barras. Dependendo da quantidade de espaço de tela

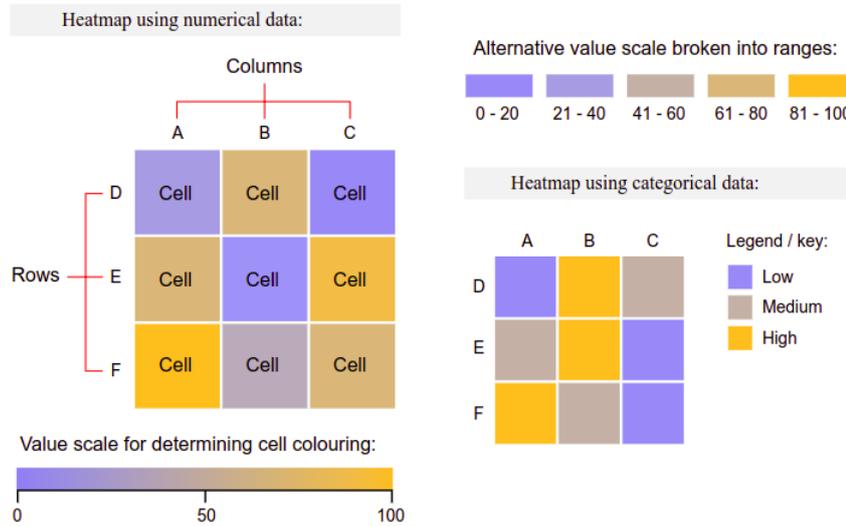


Figura 2.20: Exemplo de Mapa de Calor (RIBECCA, 2013).

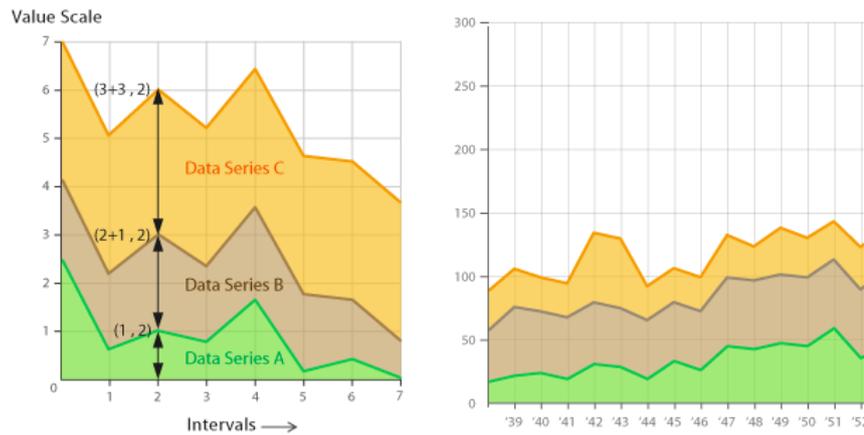


Figura 2.21: a) Exemplo de *Stacked Area*.

alocada pela visualização para o *Table Lens*, os dados podem ser exibidos em colunas e linhas sem barras de rolagem. Essa técnica inclui mecanismos de interação que fornecem recursos de *zoom*, *details-on-demand* e *overview* para exibir visualizações de todo o conjunto de dados enquanto ainda fornece alguns detalhes específicos de algum item. A classificação de colunas ajuda a identificar rapidamente tendências, correlações e discrepâncias. A Figura 2.22⁶ é um exemplo de *Table Lens* aplicado a visualização esportiva, onde um ranking de mais de 1700 itens é exibido como se fossem linhas, com o uso do mouse é possível ver mais detalhes sobre um determinado item específico. A combinação de cores e colunas permite que os usuários definam classificações de dados, executem

⁶Disponível em: <https://lineup.js.org/app/#soccer>. Acessado em 02/07/2022

opções de filtragem e detectem tendências.

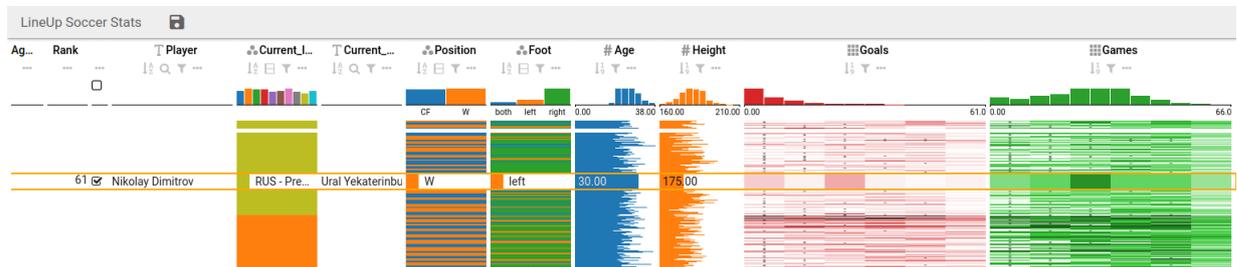


Figura 2.22: Exemplo de *TableLens* (GRATZL et al., 2013).

2.4.4 Geovisualização

Os dados georreferenciados possuem um tipo de característica singular, descrevem dados, com uma localização específica no mundo real. Dados georreferenciados aparecem em muitas aplicações, incluindo pagamentos com cartão de crédito, telefonemas, registros ambientais e dados censitários. Ward, Grinstein e Keim (2010) define o vocábulo “geovisualização”, para se referir a esta categoria de visualização de dados. Um georreferenciamento pode ser vinculado a um par de latitude e longitude, um código postal, um prefixo de número de telefone, um endereço, dados meteorológicos etc, nos quais pode ser importante analisar os padrões, incluindo a localização geográfica. MacEachren e Kraak (2001) definem geovisualização como o uso de representações visuais para tornar visíveis os contextos e problemas espaciais, de modo que a percepção humana esteja envolvida na visualização. Ward, Grinstein e Keim (2010) dispõe principalmente quatro categorias em geovisualização, a saber:

- **Visualização de Dados Espaciais:** A estratégia básica para essa categoria é a utilização de mapas. Os mapas podem ser usados de diferentes maneiras, como fornecer informações específicas sobre locais específicos, fornecer informações gerais sobre padrões espaciais ou para comparar padrões em vários mapas. Nesta categoria a interatividade é importante;
- **Visualização de Ponto de Dados:** Os dados desta categoria são discretos por natureza, mas podem descrever um fenômeno contínuo, como temperatura de locais específicos. Dependendo da natureza dos dados, a geovisualização deve ajustar os símbolos visuais para melhorar a extração de informações.
- **Visualização de Linha de Dados:** A ideia principal é que a geovisualização represente os dados como segmentos de linha entre pares de pontos finais especificados por longitude e latitude. Em geral, os mapeamentos permitem que os parâmetros de largura, tipo, cor e rotulagem sejam utilizados nas linhas. Pontos de interseção também podem ser mapeados e as linhas não precisam ser retas.

- **Visualização de Dados de Área:** Essa categoria representa geralmente os dados utilizando uma perspectiva de área ou região sobre os mapas, ou seja, mapas temáticos.

Os mapas aplicados na geovisualização podem ser subdivididos conforme as propriedades dos dados (qualitativos contra quantitativos, discretos contra contínuos) e segundo as variáveis visuais utilizadas (pontos, linhas, áreas, volumes). De acordo com Ward, Grinstein e Keim (2010) alguns dos mapas resultantes são o symbol maps (mapas de símbolos), dot maps (mapas de pontos), land use maps (mapas de uso do solo), choropleth maps (mapas coropléticos), line diagrams (diagramas de linhas), isoline maps (mapas de isolinhas) e surface maps (mapas de superfície). Frisa-se que essa subdivisão não é impecavelmente ortogonal e que os mesmos dados podem ser visualizados em diferentes tipos de mapa. Em Kraak e Ormeling (2020), há um estudo mais aprofundado sobre os diferentes tipos de mapas.

O *Choropleth Map* (Mapas coropléticos) exhibe áreas geográficas divididas normalmente coloridas, mas podem ser sombreadas ou padronizadas em relação a uma variável de dados. Isso fornece uma maneira de visualizar valores em uma área geográfica que pode apresentar variações ou padrões no local exibido. As variáveis de dados usam progressão de cores para se representar em cada região do mapa. Normalmente, isso pode ser uma mistura de uma cor para outra, uma progressão de tom único, transparente para opaco, claro para escuro ou todo o espectro de cores.

Uma desvantagem de usar cores é que não se pode ler ou comparar com acurácia os valores do mapa. Outra questão é que as regiões maiores aparecem mais enfatizadas do que as menores, de modo que a percepção do observador sobre os valores sombreados é afetada. Um erro comum ao produzir mapas coropléticos é codificar valores de dados brutos (como população) em vez de usar valores normalizados (cálculo populacional por quilômetro quadrado, por exemplo) para produzir um mapa de densidade (Figura 2.23).

O *Dot Map* (Mapa de pontos) é um método de detecção de padrões espaciais, plota-se pontos de tamanho igual em uma área geográfica. Existem dois tipos de mapa de pontos: um para um (um ponto representa uma única contagem ou objeto) e um para muitos (um ponto representa uma unidade específica, por exemplo, 1 ponto = 10 árvores). Os mapas de pontos são ideais para ver como as coisas estão distribuídas em uma região geográfica e podem revelar padrões quando os pontos se agrupam no mapa. Os mapas de pontos são fáceis de entender, sendo melhores para fornecer uma visão geral dos dados, mas não são bons para recuperar valores exatos (Figura 2.24)⁷.

O *Connection Map* (Mapa de Conexão) são desenhados conectando pontos postados em um mapa por linhas retas ou curvas. Os mapas de conexão são ótimos para mostrar conexões e relacionamentos geograficamente, mas também podem ser usados para exibir rotas por meio de uma única cadeia de conexão. Os mapas de conexão também ajudam a revelar padrões espaciais por meio da distribuição de conexões e da concentração de conexões no mapa. Em áreas densas, podem ocorrer problemas de sobreposição de conexões, dificultando a legibilidade da geovisualização (Figura 2.25).

⁷Disponível em: <https://vividmaps.com/brazils-racial-dot-map/>. Acessado em 12/06/2022

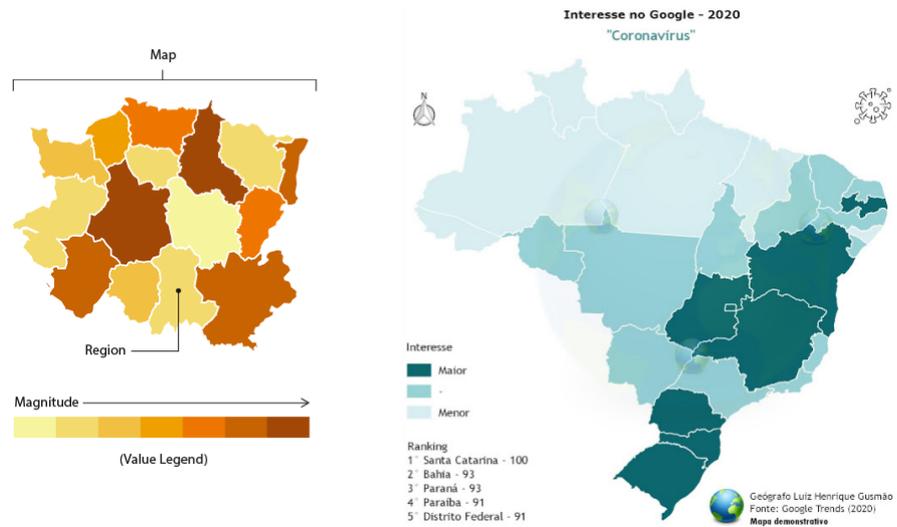


Figura 2.23: À esquerda um exemplo genérico de mapa coroplético (RIBECCA, 2013). À direita um mapa coroplético do Brasil, que demonstra o interesse sobre o tema Coronavírus (GUSMÃO, 2021).



Figura 2.24: Exemplo de *dotmap* que mostra um mapa racial brasileiro.

Os *Flow Maps* (mapas de fluxo) mostram geograficamente o movimento e a quantidade de informações ou objetos de um local para outro. Os Mapas de Fluxo são comumente usados para exibir dados sobre a migração de pessoas, animais e produtos. A magnitude



Figura 2.25: Exemplo de conexão de mapas (RIBECCA, 2013).

dos dados é expressa pela espessura do fluxo. Isso é útil, por exemplo, para visualizar como a migração pode ser distribuída geograficamente. O mapa de fluxo é desenhado a partir de um ponto de origem ramificado a partir de suas “linhas de fluxo”. As setas podem ser usadas para indicar a direção ou se o movimento é de entrada ou saída. O desenho de linhas de fluxo sem setas pode ser usado para representar trajetos percorridos. A técnica de mapa de fluxo é inspirada em um algoritmo de *layout* de grafos que minimiza o cruzamento de vértice de borda. Isso significa que evitar interseções pode ajudar a reduzir a confusão visual no mapa (Figura 2.26).

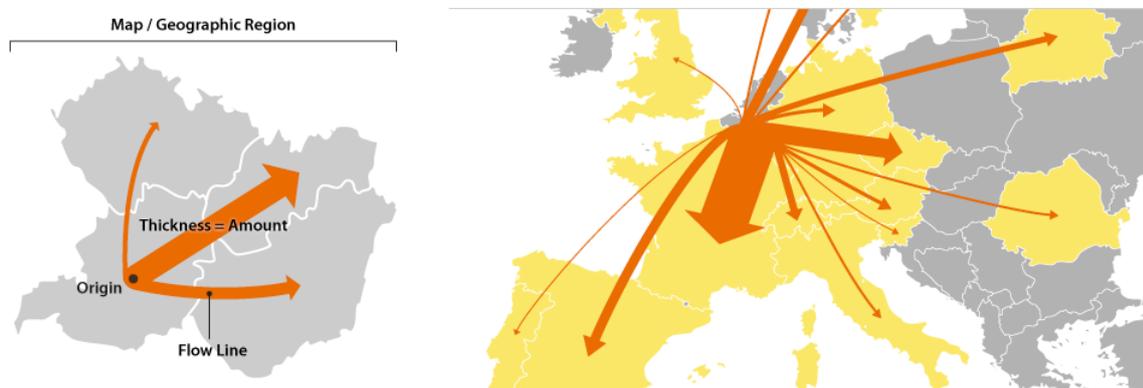


Figura 2.26: Exemplo de *flow map* (RIBECCA, 2013).

Os círculos do *Bubble Map* (mapa de bolhas) são exibidos em uma região geográfica designada com o tamanho do círculo proporcional ao seu valor no conjunto de dados. Os mapas de bolhas são bons para comparar proporções entre áreas geográficas sem os problemas causados pelo tamanho da área regional, algo que pode acontecer no mapa coroplético. No entanto, uma desvantagem do mapa de bolha é que bolhas muito grandes podem se sobrepor a outras bolhas ou regiões no mapa, isso precisa ser considerado na escolha dessa técnica (Figura 2.27).

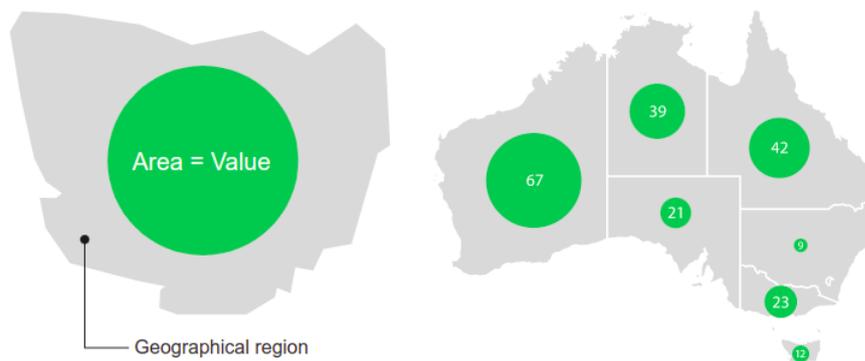


Figura 2.27: Exemplo de mapa de bolhas (RIBECCA, 2013).

2.5 MÉTODOS DE AVALIAÇÃO DE USABILIDADE

Segundo Plaisant (2004), as tarefas designadas durante os estudos com usuários na avaliação da visualização de informações geralmente se concentram em tarefas simples e específicas, que podem ser realizadas em um razoável período e com resultados mensuráveis. Em Forsell e Cooper (2012), discute-se a importância de utilizar questionários para avaliar a usabilidade em visualização de informações. Os autores argumentam que a utilização de questionários é uma grande contribuição para o repertório de métodos e instrumentos usados para avaliação na visualização de informações.

Estudos com usuários podem informar os *designers* sobre a frequência de uso de recursos, orientar *layouts* de tela para acelerar a interação, identificar possíveis fontes de insatisfação e informar a revisão de materiais de ajuda. Eles também podem fornecer evidências de sucesso quando os usuários escolhem uma interface em vez de outra e alocam recursos onde são mais necessários (PLAISANT, 2004).

Nagpal, Mehrotra e Bhatia (2017) afirmam que a ISO 9241-11 é uma norma bastante aceita em IHC, a qual define que usabilidade é: “uma medida na qual um produto pode ser usado por usuários específicos para alcançar objetivos específicos com eficácia, eficiência e satisfação em um contexto específico de uso”. Por outro lado, a ISO 9126-1 bastante utilizada em engenharia de software define que usabilidade é a: “capacidade do produto de software de ser compreendido, aprendido, operado e atraente ao usuário, quando usado sob condições especificadas”.

Portanto, antes de se declarar que um sistema visual está pronto para uso, é importante utilizar algum método para medir seu grau de usabilidade. Assim como testes de funcionalidade são necessários para verificar a robustez da implementação, a avaliação de usabilidade é necessária para verificar se a qualidade de uso de uma visualização possui tempo e esforço necessário para os usuários aprenderem a utilizar o sistema visual com um nível de competência e desempenho aceitáveis.

Fernandez, Insfran e Abrahão (2011) apresentam um estudo detalhado que resume sobre os métodos de avaliação de usabilidades (UEMs) empregados em aplicações Web. Os autores afirmam que a comparação de UEMs para determinar seu desempenho é uma

tarefa complexa, visto que as técnicas de UEMs empregadas não avaliam os mesmos aspectos de usabilidade, como visto anteriormente existem diferentes definições de usabilidade que afetam diretamente como ela é avaliada, por isso houve artigos que utilizaram mais de um tipo de UEMs.

Segundo os autores, 59% dos artigos revisados apresentaram algum tipo de teste com os usuários reais finais. Esse tipo de avaliação coloca um avaliador para observar como o usuário executa as funcionalidades do sistema web. Este é o principal ponto positivo deste tipo de avaliação, observar o usuário final em ação para entender quais são as dificuldades e as facilidades. Isso é geralmente feito no final do desenvolvimento porque é inicialmente feito com um protótipo de alta fidelidade. UEMs que utilizam algum método de inspeção foram utilizados em 43% dos artigos analisados. Esse tipo de avaliação é conduzida sem a participação de usuários reais finais, apenas com especialistas em IHC ou no domínio da aplicação. Os seguintes tipos de métodos de inspeção foram identificados pelos autores, a saber:

- ***Heuristic Evaluation:*** Esse método visa examinar a interface e julgá-la de acordo com princípios reconhecidos pela área de usabilidade. Em Nielsen (1994) há uma proposta de heurísticas para avaliação de usabilidade web com mais de 1811 citações conforme o Google Scholar (examinado em 11/07/2022), que foi atualizada em 2020, disponível online⁸ com livre acesso;
- ***Cognitive Walkthrough:*** Esse método pode ser aplicado no início do desenvolvimento. O profissional de usabilidade simula passo a passo o comportamento do usuário, por isso ele precisa conhecer bem quem é o usuário, às suas características, as habilidades e as deficiências, visto que ele tentará executar uma determinada tarefa como se fosse o usuário;
- ***Perspective-based Inspection:*** Esse método é aplicado quando especialistas realizam uma avaliação baseado em perspectivas de *design* pré-estabelecidas;
- ***Guideline review:*** Semelhante ao primeiro item, com a diferença que este UEM contém diretrizes mais detalhadas em comparação com o *heuristic evaluation*, ocasionando mais tempo de avaliação;

Em 35% dos artigos revisados, algum método de consulta foi utilizado. Esse tipo de avaliação reúne informações subjetivas dos participantes como preferências e os sentimentos após a utilização do artefato avaliado. Os seguintes tipos de métodos de consulta foram identificados pelos autores, a saber:

- ***Questionnaire:*** Trata-se de um conjunto de questões apresentadas aos usuários. Por meio desse método é possível identificar facilmente as preferências, as satisfações e as ansiedades dos usuários.

⁸Disponível em: <https://www.nngroup.com/articles/ten-usability-heuristics/>. Acessado em 08/06/2022

- **Interviews:** Um usuário é entrevistado por um especialista e ambos discutem sobre a atitude do usuário em relação ao artefato avaliado.
- **Focus Group:** Vários usuários participam de uma sessão de discussão sobre suas atitudes em relação ao artefato avaliado.

A modelagem analítica foi utilizada em 21% dos artigos. Esta avaliação apresenta abordagens de engenharia que permitem aos avaliadores medir a usabilidade utilizando modelos matemáticos, a saber:

- **Cognitive Task Analysis:** As tarefas do usuário são modeladas para prever problemas de usabilidade;
- **Task Environment Analysis:** Avaliação da correlação entre os objetivos dos usuários e tarefas executadas pelo usuário na interface;
- **GOMS Analysis:** A usabilidade é medida em termos de Goals (Objetivos), Operators (Operadores), Methods (Métodos) e Selection rules (Regras de seleção) esse método não envolve a participação do usuário final. Ele estipula um tempo possível para um usuário poder decidir, como, por exemplo, ponderar, clicar num botão, encontrar um botão, arrastar o mouse, clicar em um botão, etc. É definido um operador para cada passo que o usuário precisaria fazer para realizar uma ação e a cada operador é atribuído um valor, assim, o tempo de cada ação pode ser calculado por meio da soma dos valores atribuídos a cada operador necessário para executar a ação.

Por fim, em 8% dos artigos utilizaram UEMs que simulam o comportamento do usuário por meio de algum algoritmo.

2.6 CONSIDERAÇÕES FINAIS

Este capítulo apresentou aspectos teóricos importantes para a área de visualização de dados, dentre eles conceitos de *pipeline* de visualização, interação humana, símbolos visuais, cores e operações visuais. A interação humana conduz todo o processo de análise visual e está diretamente relacionada às visualizações geradas para interpretar os dados. A interpretação dos dados é alcançada através do uso adequado de cores, símbolos visuais, métodos de operações visuais e técnicas de visualização para representar adequadamente o conjunto de dados.

Apresentaram-se exemplos de técnicas de visualizações de séries temporais, dados multidimensionais, geovisualização e grafos. Cada grupo auxilia no processo de extração de informação considerando o contexto dos dados que se deseja visualizar. As técnicas de séries temporais facilitam a análise de tendências variantes no tempo, representando adequadamente elementos associados a cada intervalo de tempo. Já as técnicas multidimensionais, por exemplo, podem ser utilizadas para verificar a interdependência entre os dados.

As técnicas de visualização de grafos podem ser utilizadas para a representação de redes. Na literatura as principais técnicas são o diagrama de ligação de vértices, mais intuitivas ao usuário e a matriz de adjacências, mais escaláveis. Por fim, as técnicas de geovisualização são aplicadas quando abordam dados associados a uma localização específica no mapa, por meio de pontos, linhas ou regiões.

Por fim, foi apresentada técnicas de usabilidade para validar e verificar se o que se planejou está sendo realmente feito e, mais importante, analisar se o que está finalizado é utilizável, se é útil, se é acessível a potenciais usuários da visualização.

TRABALHOS RELACIONADOS

3.1 CONSIDERAÇÕES INICIAIS

A literatura vem investigando e avançando no tema de análise de dados esportivos. Em concomitante, o avanço da tecnologia vem permitindo coletar dados das mais variadas e diferentes atividades esportivas, sejam elas internas (do inglês, *indoor*) ou externas (do inglês, *outdoor*), individuais ou em grupos, presenciais ou remotos (i.e. *online*), amadores ou profissionais, etc. Em especial, a área de visualização de dados aparece com frequência como mecanismo para extrair padrões e, por consequência, informação dos dados capturados.

Neste capítulo será apresentada uma revisão da literatura, abarcando propostas que utilizam visualizações esportivas, assim como os desafios e contribuições para a questão de análise de dados esportivos. O presente capítulo está organizado da seguinte forma. Na Seção 3.2 foram destacados trabalhos que analisam as classificações de visualizações de esportes individuais e coletivos. Na Seção 3.3 trabalhos relacionados a visualização de dados em ciclismo. Por fim, na Seção 3.4 as considerações finais do capítulo.

3.2 VISUALIZAÇÃO DE DADOS ESPORTIVOS

A visualização esportiva provê diferentes abordagens para explorar, entender, e comunicar dados esportivos. Técnicas de visualização de dados e *Visual Analytics* fornecem abordagens que auxiliam essas tarefas. Enquanto a visualização de dados atua com técnicas mais acessíveis e significativas que facilitam a obtenção de percepções (COX; STASKO, 2006), *Visual Analytics* é mais indicada para dados complexos e que permitem uma gama maior de interações para buscar padrões de informação não triviais nos dados.

Aplicações comerciais, como Strava e Garmin, apresentam informações detalhadas de atividades esportivas, incluindo ciclismo. Diversos gráficos de linhas e área apresentam principalmente informações de altimetria, frequência cardíaca, cadência e velocidade. Contudo, esses gráficos são apresentados somente por corrida, sem a possibilidade de comparação visual com outras corridas, seja do mesmo atleta ou com outros atletas. A comparação com o desempenho com outros atletas ocorre somente por tempo mediante

uma listagem, mostrando a sua classificação para um determinado período ou trecho da atividade.

O trabalho de Perin et al. (2018) categoriza as visualizações esportivas em três categorias: *box-score data*, *tracking data* e *meta-data*. A pesquisa contou inicialmente com 281 documentos, entre artigos, infográficos, artigos esportivos, web analytics, etc. 98 artigos foram selecionados para análise, dos quais 63 foram classificados como acadêmicos.

Box-score indica visualizações obtidas a partir de um resumo estatístico dos agentes que participam de um determinado esporte como beisebol, basquete e hóquei. Por exemplo, a pontuação realizada pelos próprios atletas participantes do jogo ou corrida. O vocábulo *tracking data* foi usado para mostrar visualizações esportivas obtidas por dados espaço-temporais coletados por câmeras, sensores, que capturam movimentos de atletas em campo, trajetórias de objetos, etc. Por fim, o vocábulo *meta-data* foi utilizado para designar visualizações esportivas obtidas por informações extras (metadados) de um determinado esporte, como as capacidades do estádio, características físicas dos atletas, informações de patrocinadores etc.

Outro estudo que representa a discussão atual sobre visualização esportiva em formato de *survey* foi realizado por Du e Yuan (2021). Neste trabalho os autores apresentam quatro classificações para a visualização de esportes: visualização de dados com alta dimensionalidade, visualização de séries temporais, visualização de grafos ou rede e visualização baseada em glifos.

A visualização de dados com alta dimensionalidade é usada para se referir a visualizações para dados multidimensionais. Du e Yuan (2021) afirmam que essa categoria é amplamente utilizada para explorar padrões de distribuição na visualização esportiva, porque explora relações e efeitos implícitos em várias dimensões. Incluem métodos baseados em ícones, *pixels*, geometria, hierarquias, gráficos ou uma combinação deles. Exemplos de técnicas de visualização que podem ser utilizadas ou combinadas nesta categoria cita-se o *scatter plot* (gráficos de dispersão), *radar chart* (gráficos de radar), *parallel coordinate* (coordenadas paralelas) e *heat map* (gráficos de calor). (Figura 3.1)

De acordo com Du e Yuan (2021) a visualização de séries temporais descreve perspectivas de visualização derivadas de informações espaço-temporais. São frequentemente integradas à geografia cartográfica. O que possibilita a visualização de dados espaciais e outros modelos de dados que refletem padrões no comportamento de objetos ao longo do tempo. Utilizam-se principalmente gráficos de linha, histogramas ou uma combinação deles. A Figura 3.2a extraída de Perin, Boy e Vernier (2016) apresenta a evolução das equipes no campeonato francês. O gráfico ilustra que quatro equipes são claramente melhores que as demais. Há uma lacuna entre Paris Saint-Germain e Mônaco, e entre Mônaco e Lyon. Os três primeiros se classificam para a Liga dos Campeões. Essa lacuna aparece no início no campeonato e continuou aumentando até o final, revelando uma clara diferença de desempenho entre os times.

A visualização de rede ou grafo foi utilizada por Du e Yuan (2021) para se referir a perspectivas de visualização aplicadas a dados hierárquicos (nós de rede, estruturas topológicas) e inclui metodologias do tipo *node-based graphs* (grafos baseados em nós), *space-filling graphs* (grafos baseados em preenchimento de espaço), *large-scale dense graphs based on edge binding* (grafos densos em grande escala baseados em ligação de bordas)

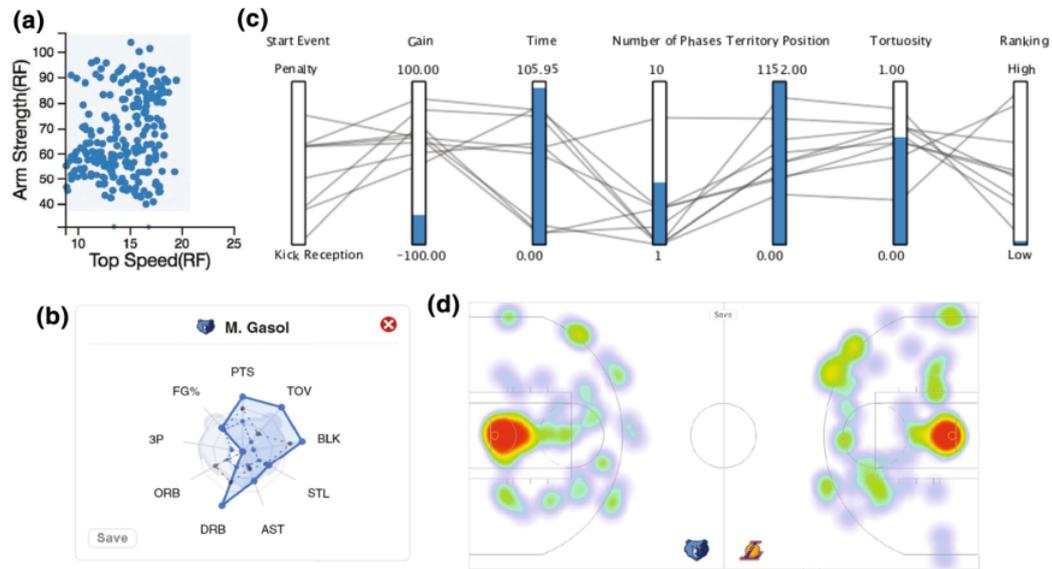


Figura 3.1: Técnicas de visualização de dados com alta dimensionalidade. **(a)** *scatter plot* (gráfico de dispersão), **(b)** *radar chart* (gráfico de radar), **(c)** *parallel coordinate* (paralelas coordenadas) e **(d)** *heat map* (gráficos de calor).

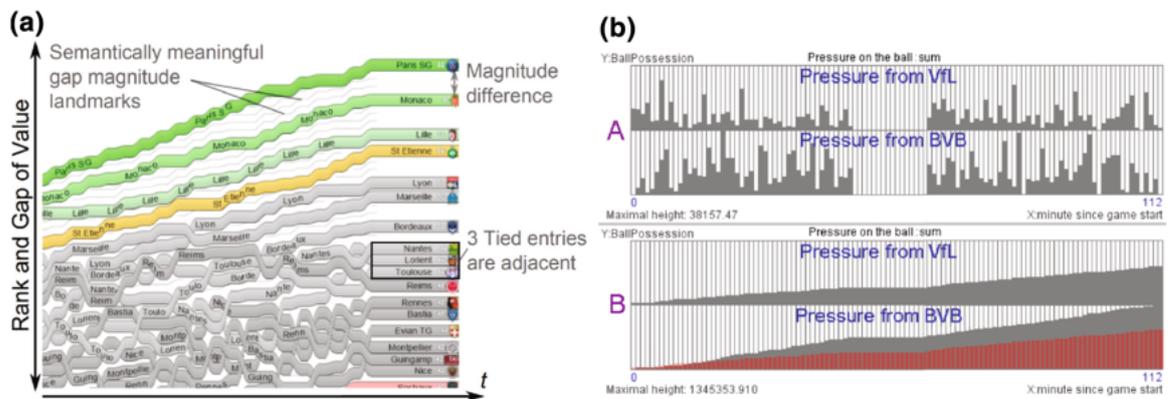


Figura 3.2: Exemplos de visualizações de séries temporais. **(a)** Gráfico de linha (que demonstra uma variação ao longo do tempo a classificação de times de futebol), **(b)** Histograma (que demonstra a posse de bola de dois times de futebol ao longo do tempo).

e *large-scale graphs based on hierarchical clustering* (grafos em grande escala baseados em agrupamento hierárquico). A visualização de rede é menos aplicável às visualizações esportivas, em razão de ser difícil representar o tempo com esse tipo de visualização. No entanto, o *Tree Map* (Mapa de Árvore) e matrizes são exemplos de técnicas de visualização de dados que podem ser colocadas nesta categoria.

A Figura 3.3 apresenta o *Adaptivitree*, extraída de Tan et al. (2007). Um sistema

visual em formato de árvore para dar suporte à visualização de chaves de um torneio. O principal objetivo da exibição de chaves de um torneio é permitir a visualização da redução do número de participantes a um único campeão. Isso geralmente torna uma árvore ou hierarquia uma representação visual útil.

A principal contribuição do *Adaptivitree* é atualizar a estrutura da árvore para que os resultados do torneio possam ser vistos em um formato não textual. O visão inicial do *Adaptivitree* é semelhante à visualização clássica, mas em vez de repetir os nomes dos vencedores em uma nova chave do torneio para indicar vitórias em cada confronto, o *Adaptivitree* move a linha de resultados de um confronto e estende a linha do vencedor, isso significa que o nome de cada competidor aparece apenas uma vez no *Adaptivitree* e o progresso de um competidor no torneio é representado por uma única linha horizontal que se estende de seu nome em direção ao centro. Quando um competidor perde, a sua linha horizontal termina.

As cores mapeiam as previsões do usuário, o verde sem traços, significa uma previsão feita pelo usuário que realmente se materializou no confronto real, o verde tracejado, é uma previsão feita pelo usuário do provável vencedor com chances de acontecer, e o vermelho foi uma previsão que não se concretizou. O *Adaptivitree* não é apenas uma visualização mais compacta e compreensível, mas também permite usar sobreposições que mostram resultados preditivos.

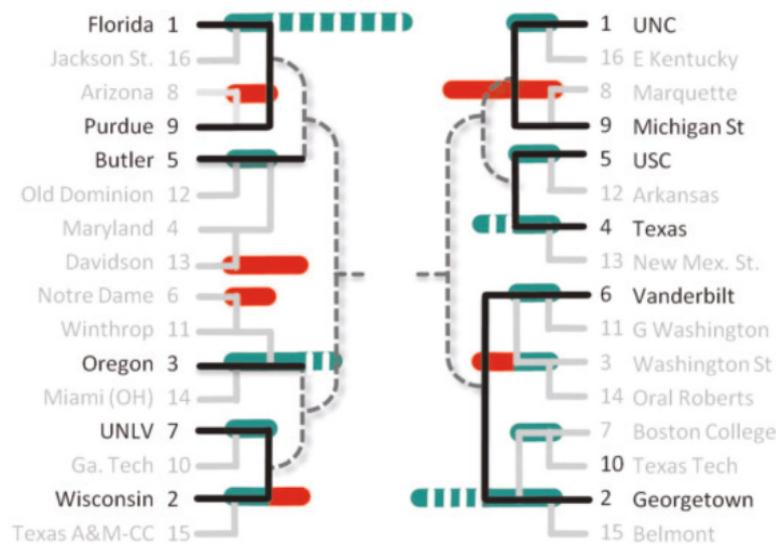


Figura 3.3: Exemplo de visualização hierárquica técnica que pode ser incluída na categoria visualização de rede segundo Du e Yuan (2021). A estrutura hierárquica acima indica os resultados previsíveis no campeonato de basquete.

A visualização baseada em glifos, segundo Du e Yuan (2021), é uma das técnicas de visualização mais utilizadas na visualização esportiva. Ela mapeia os dados usando entidades gráficas e transmite informações por meio de forma, tamanho, cor, posição e

outros canais de codificação. Conseqüentemente, é adequada para exibir conjuntos de dados multidimensionais, mas restrita a uma certa quantidade de variáveis visuais. Normalmente, este tipo de visualização é usada para mostrar o desempenho de um jogador, seu tempo de posse de bola, o número de gols marcados, as estatísticas de um atleta em uma partida específica.

A Figura 3.4 apresenta um exemplo de glifo para rugby que contém uma série de indicações dos locais e quais dados estão sendo coletados. A cor do fundo no quadrado maior é usada para identificar o time, o vermelho para o time da casa, e o azul para o time visitante. Para representar se o time perdeu ou venceu usa-se um círculo colorido colocado entre o glifo principal e a barra de duração, verde para a vitória e laranja para a derrota.

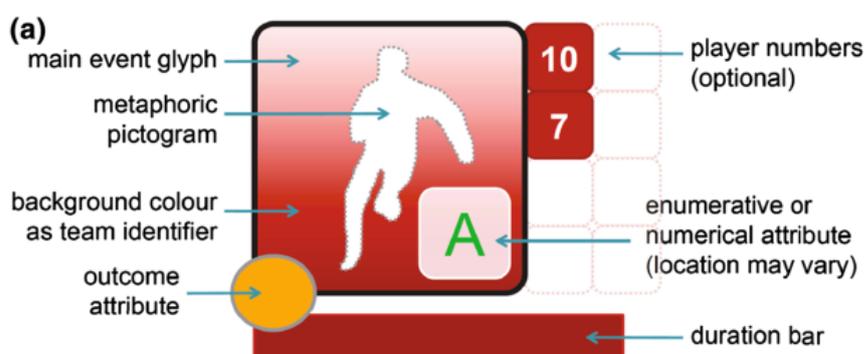


Figura 3.4: Visualizações esportivas baseadas em Glifos. (a) Glifo em forma de jogador para descrever eventos com dados de rugby.

Observa-se que a proposta de categorização de visualizações esportivas em Perin et al. (2018) possui como principal critério como os dados são coletados, além disso, o autor especifica que a categorização da visualização esportiva, seja em *box-score data*, *tracking data* ou *meta-data* não é impecavelmente ortogonal, ou seja, o limite entre esses tipos de dados nem sempre é claro. Por exemplo, os autores afirmam que algumas ligas esportivas profissionais podem usar o tipo *tracking data* para computar *box-score data*, e o tipo *meta-data* geralmente exigem considerações do tipo *box-score* e/ou *tracking data*.

De acordo com Du e Yuan (2021) a visualização esportiva consiste em três tarefas: apresentação, comparação e previsão. Segundo os autores, as técnicas tradicionais de visualização ainda não atendem às necessidades da visualização esportiva. Portanto, resolver os problemas que surgem no processo de análise visual é um desafio e também uma oportunidade. Por isso a proposta de classificação de Du e Yuan (2021) para visualização esportiva considera as diferentes características dos dados entre eles.

Nota-se que ainda não há um consenso sobre as categorizações para as visualizações esportivas. Para facilitar a apresentação dos trabalhos relacionados desta pesquisa, será adotada uma classificação baseada na modalidade do esporte no que tange à quantidade de pessoas, ou seja, se determinado esporte é praticado individualmente ou em grupo

(mais de uma pessoa).

3.2.1 Esportes Individuais

Dentre as atividades esportivas que os atletas, profissionais ou amadores, realizam sozinho sem um parceiro ou equipe durante atividade, aborda-se neste trabalho corridas de rua (OLIVEIRA et al., 2013), natação (TORRE-ARENAS, 2017), xadrez (LU; WANG; LIN, 2014).

Em Oliveira et al. (2013) foi apresentada uma proposta de visualização esportiva que analisa uma determinada corrida, utilizando conjuntos de dados de tamanhos que variam de 60 a 483 corredores. Os dados incluem frequência cardíaca, velocidade e geolocalização. O resultado é uma série de linhas horizontais que começam a partir de um ponto à esquerda e aumentam de comprimento para a direita à medida que os corredores se aproximam do fim da corrida.

A parte mais exigente de uma corrida pode ser identificada onde a cor geral é mais brilhante, e segundo o autor, são geralmente os trechos com mais altitudes, ou o trecho final da prova, onde a competição por uma melhor classificação aumenta. Mudanças de cor significa um aumento ou diminuição na velocidade geral, e ajuda a identificar seções de alta e baixa velocidade geral do percurso (Figura 3.5).

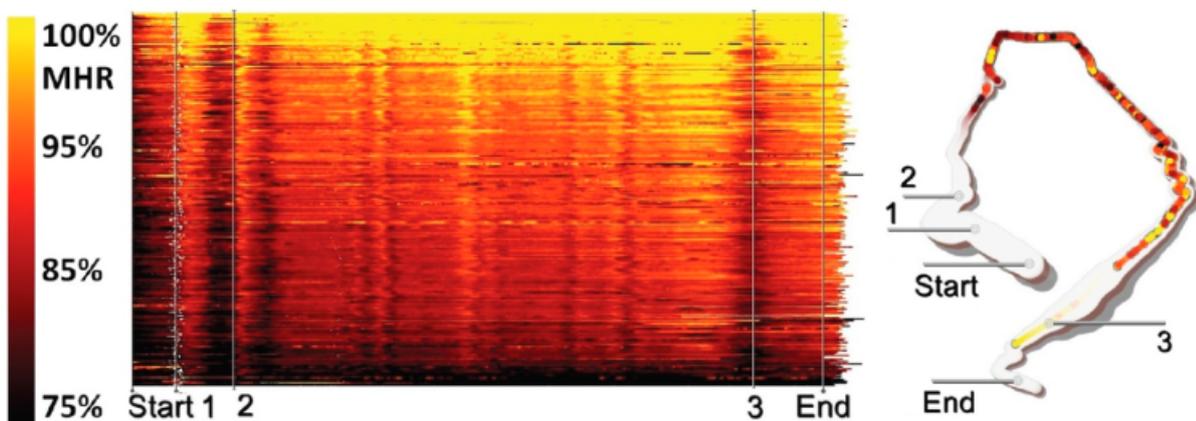


Figura 3.5: Exemplo de visualização de corrida de rua. Cada linha representa um corredor. Na direita a topografia da corrida. Quanto mais brilhante for a cor, maior será o nível de esforço do corredor, para cores mais escuras, menor será o nível de esforço do corredor (OLIVEIRA et al., 2013).

Essa visualização permite comparar o esforço entre diferentes corredores em uma determinada corrida simultaneamente. Os autores se baseiam nos dados de frequência cardíaca para medir o esforço e a intensidade do exercício, mapeando essa informação na visualização. O mapeamento de cores foi projetado para comparar a frequência cardíaca com as respectivas zonas de esforço apropriadas. Oliveira et al. (2013) explicam que os níveis de esforço podem mudar durante a corrida e torna-se importante identificar onde

e por que essas mudanças ocorrem. Essa proposta de visualização permite observar a mudança no nível de esforço ao longo da corrida, observando as mudanças de cor ao longo do eixo horizontal.

Uma das contribuições do trabalho é apresentar dados de múltiplas corridas em simultâneo, sendo cada corrida de atletas diferentes, foi necessário analisar séries temporais multidimensionais gerados por medidores de frequência cardíaca de vários corredores. Segundo Oliveira et al. (2013) às ferramentas de visualização esportiva de frequência cardíaca se concentram em visualizar uma única corrida, muitas vezes falta a capacidade de comparar múltiplas corridas, segundo o autor a análise de dados deste trabalho foi um desafio.

Esportes que preterem atividade mental à física também estão relatados na literatura. Em Lu, Wang e Lin (2014) foi proposta uma visualização esportiva de xadrez para transmitir aos usuários os possíveis desfechos de um jogo de xadrez ao longo das sucessivas jogadas dos participantes. Ao contrário das ferramentas gráficas de xadrez atuais, que se concentram apenas em destacar as peças que estão sob ataque e deixando para o jogador realizar uma investigação mental dos próximos lances assertivos, esta visualização revela os resultados possíveis depois que determinada peça é movida pelo tabuleiro, indicando a vantagem tática que determinado jogador pode ou não ter sobre o oponente.

De acordo com Lu, Wang e Lin (2014) movimentos futuros em jogos de xadrez podem ser previstos usando notações algébricas, no entanto, esta notação não é trivial, sua compreensão é difícil. Na visualização proposta em Lu, Wang e Lin (2014) os usuários podem examinar e visualizar um jogo de xadrez globalmente, visualizando os possíveis resultados, a partir de determinada jogada. A análise de jogo é obtida a partir da utilização de IA para erigir uma *decision tree* (árvore de decisão) a partir do registro das jogadas efetuadas pelos jogadores.

Segundo Lu, Wang e Lin (2014), uma árvore de decisão de xadrez, é essencialmente um grafo, onde cada nó representa uma posição atual das peças do xadrez, possíveis decisões, e as bordas são os possíveis movimentos permitidos e seus resultados. Uma maneira de visualizar a evolução do jogo de xadrez é renderizar completamente toda a *decision tree*. Mas vários nós e bordas podem adicionar complexidade e causar uma confusão visual. Além disso, nós diferentes na árvore podem representar um mesmo estado de peças no tabuleiro, o que pode resultar em percepções erradas Figura 3.6.

A visualização proposta em Lu, Wang e Lin (2014) simplifica a *decision tree* fundindo nós que representam o mesmo estado das peças do jogo no tabuleiro e descartando nós, com pouco efeito sobre o futuro do jogo, melhorando a legibilidade da visualização. As localizações dos nós foram renderizadas usando o sistema *Graph Viz* (Ellson et al. (2002)), as espessuras das bordas são baseadas na qualidade das jogadas. Na visualização, por exemplo, mostra prováveis eventos críticos, como xeque-mates ou *draws*(este é um resultado comum no xadrez quando nenhum lado ganha ou perde, quando nenhum dos lados é forte o suficiente para alcançar uma vitória).

Em Torre-Arenas (2017) foi proposta uma visualização esportiva sobre os recordes mundiais de natação ao longo da história. A autora deste estudo se concentrou em analisar os dados para encontrar padrões, para que posteriormente pudessem ser visualizados por um público mais geral, fazendo-os querer aprender mais sobre natação. Segundo Torre-

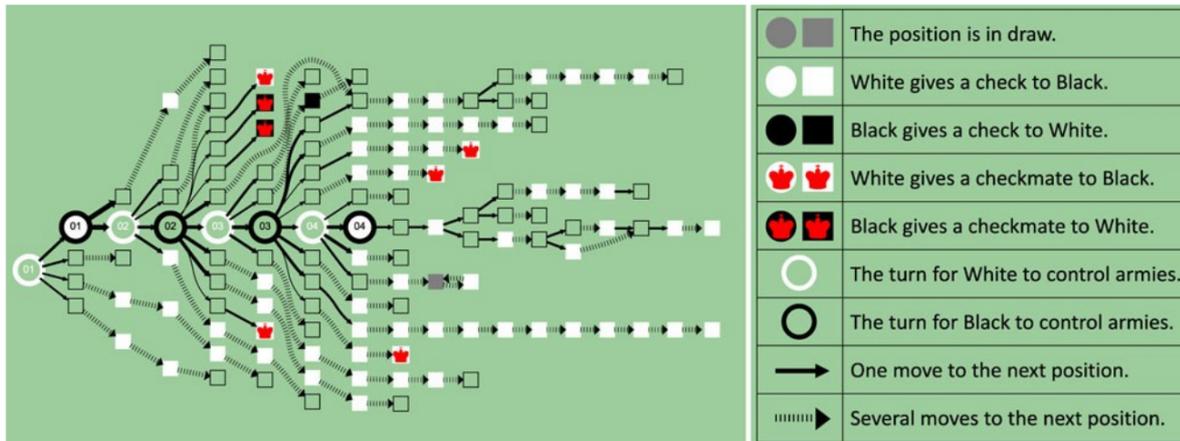


Figura 3.6: Conforme Lu, Wang e Lin (2014), neste gráfico, os movimentos reais feitos pelos usuários com círculos, os quadrados representam os movimentos estimados pela IA. Na direita o significado dos símbolos, que podem ser combinados para transmitir informações adicionais. Por exemplo, um círculo branco com uma borda preta significa que as peças brancas dão um xeque nas peças pretas, logo as peças pretas devem reagir.

Arenas (2017) a Wikipédia foi a fonte de dados para este estudo, pois a FINA (Federação Internacional de Natação) não possui um banco de dados público com todos os recordes, e as informações fornecidas publicamente pela instituição são insuficientes.

De acordo com Torre-Arenas (2017) com a Wikipédia, foi possível coletar um conjunto de dados com 2.359 registros, sendo que cada registro é composto por um nadador que rompeu um recorde. Além disso, foi possível obter informações sobre competições específicas, sexo, horários, datas, nacionalidades, localidades. A velocidade é a variável-chave para este projeto. A autora projetou visualizações, que focam na comparação das velocidades médias de todos os recordes registrados entre 1900 e o início de 2017. Além disso, é possível usar filtros para as modalidades de natação e nacionalidade dos atletas (Figura 3.7).

Após o estudo da proposta de visualização esportiva em Torre-Arenas (2017), uma das primeiras conclusões alcançadas ao examinar os dados foi o quanto os nadadores melhoraram seus desempenhos ao longo do tempo. Por exemplo, em 1902 o nadador Freddy Lane nadou 200 metros livre em 2 minutos e 28.6 segundos, um século mais tarde, Paul Biederman nadou essa mesma distância em 1 minuto e 42 segundos.

Colocado em um contexto mais brasileiro, a visualização em Torre-Arenas (2017), revela que em 1980 o nadador Chris Cavanaugh nadou 50 metros, estilo livre em 23.12 segundos, e em 2009 esse recorde foi superado pelo nadador brasileiro César Cielo com 20.91 segundos. Em 1905 o nadador Zoltán Halmay nadou 100 metros, estilo livre em 1 minuto e 5.8 segundos, em 2009 César Cielo superou esse recorde com 46.91 segundos. Em 1939, Maria Lenk fez 200 metros, estilo peito em 2 minutos e 56 segundos, em 2013 esse recorde foi quebrado por Rikke Moller Pedersen com 2 minutos e 19.11 segundos.

Por fim, muito mais *insights*, podem ser obtidos nesta visualização, como quais foram

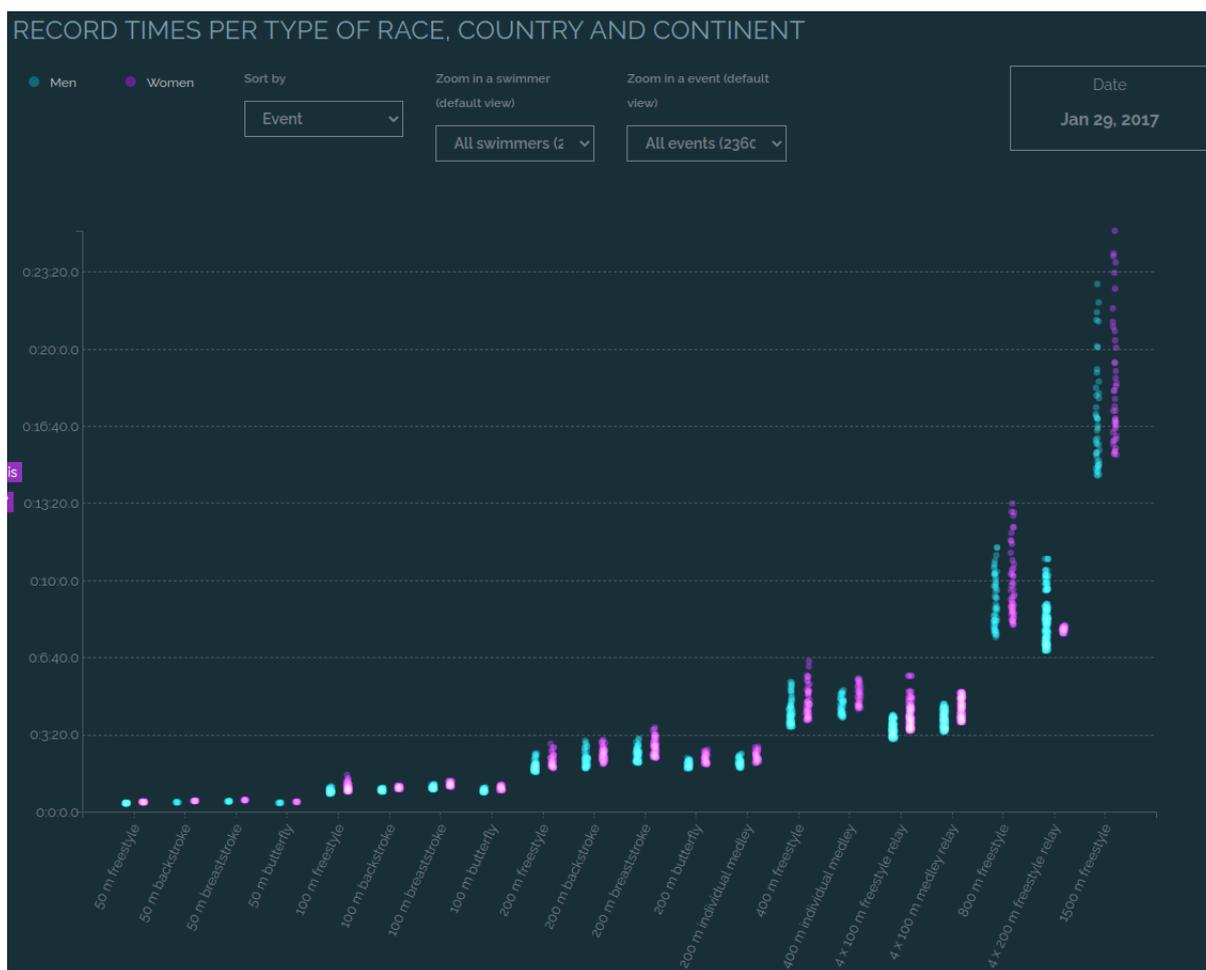


Figura 3.7: Permite comparar recordes de homens e mulheres para diferentes modalidades (TORRE-ARENAS, 2017).

os recordes em todos os estilos de natação entre 1900 e 2017, quais nadadores estabeleceram esses recordes. É possível comparar os recordes entre homens e mulheres, obviamente os recordes masculinos são bem mais baixos se comparados aos femininos, mas a visualização em Torre-Arenas (2017) mostra o quão grande é essa diferença.

3.2.2 Esportes em Grupo

Esportes em grupo podem ser considerados atividades complexas contendo um número significativo de conjuntos de dados abstratos e categorias de participantes, como atletas, treinadores, árbitros e espectadores Page e Moere (2006). Os esportes em equipe são fundamentais na sociedade atual e podem ser vistos na televisão, no rádio, jornais, internet ou smartphones, e ao contrário dos esportes individuais, exigem a colaboração de vários participantes em simultâneo, ou seja, todos os participantes devem ter uma compreensão abrangente, uma visão global dos aspectos do esporte coletivo em questão.

Page e Moere (2006) afirmam que a visualização é aplicada a esportes coletivos de várias maneiras, desde representações científicas a respeito de movimento ou trajetórias, até uma representação mais contextual das regras do jogo e da identidade da equipe. Essa abordagem ampla para a visualização de esportes em grupo decorre da natureza dos dados esportivos que podem ser bastante abstratos. De acordo com Page e Moere (2006), a visualização esportiva de equipe pode ser definida como uma representação visual de dados relacionados ao jogo para auxiliar as pessoas interessadas a entender o esporte. A visualização das características dos esportes coletivos facilita a compreensão do esporte.

Goldsberry (2012) desenvolveu o *CourtVision*, um sistema visual para analisar, quantificar e comunicar os aspectos espaciais do desempenho dos atletas da NBA. Segundo os autores, o basquetebol é um esporte espacial, porque a natureza e a dinâmica de utilização do espaço da quadra (escolha dos locais de arremesso, planejamento defensivo, o espaçamento entre os jogadores) influenciam significativamente o resultado do jogo, ou seja, o desempenho de um time de basquetebol requer raciocínio espacial. O objetivo do *CourtVision* é fornecer uma melhor compreensão da complexa dinâmica espacial dos jogadores e equipes da NBA.

O *CourtVision* através da visualização esportiva revela comportamentos espaciais únicos de jogadores e equipes, que, por sua vez, expõem padrões e anomalias importantes no desempenho dos times que são consideravelmente menos óbvios de serem notados. Por exemplo, o *CourtVision* revela quais posições de arremesso têm as maiores taxas de conversão entre os jogadores, bem como as posições na quadra onde os jogadores são mais bem-sucedidos.

Para visualizar as posições da quadra com as maiores taxas de conversão, Goldsberry (2012) mapeia os dados de arremesso sobre um mapa básico de uma quadra de basquete da NBA subdividida em 1284 células de arremesso com tamanho de 1 ft^2 (aproximadamente $0,09 \text{ m}^2$). Os dados de arremesso foram coletados a partir de um conjunto de dados contendo os jogos da NBA entre 2006 e 2011, esse conjunto de dados inclui o nome do jogador, a localização cartesiana do arremesso e resultado do arremesso para mais de 700.000 arremessos, entre outros dados. De acordo com Goldsberry (2012) as quadras da NBA têm 4.700 ft^2 ($436,64 \text{ m}^2$), e 98% das tentativas de conversão dos arremessos ocorrem em uma área de 1.284 ft^2 ($119,29 \text{ m}^2$) que vai do fundo da quadra até a linha de 3 pontos, esta é a zona de maior pontuação segundo o autor.

Segundo o autor, os jogadores mais eficazes no campeonato eram aqueles capazes de converter arremessos de diferentes lugares da quadra. Para criar uma visualização que revela a posição mais efetiva de um determinado jogador em quadra, Goldsberry (2012) criou uma métrica chamada "*Range*" que é a somatória do número de células de arremesso em que um jogador possui um PPA (pontos por tentativa) de 1. Segundo o autor, uma célula de arremesso possui um PPA de 1 se os pontos convertidos naquela posição foram acima de 1, e possui um PPA de 0 se não houve pontuação daquela célula de arremesso.

Os PPA's variam ao redor da quadra, pois a probabilidade de um arremesso ser convertido não é de 100%. O *Range* é uma métrica que sinaliza as influências espaciais na eficácia do arremesso, é uma contagem das posições da quadra onde o jogador conseguiu converter arremessos. Goldsberry (2012) dividiu o *Range* por 1284 e multiplicou por 100, gerando o *Range%* que indica o percentual de acertos da área total de pontuação.

Segundo o autor, o primeiro lugar foi ocupado por Steve Nash, que revelou um *Range* de 406, ou seja, o jogador é efetivo em 31,6% da área total de pontuação. Ray Allen ficou em segundo lugar com 30,1%, Kobe Bryant em terceiro com 29,8% e Dirk Nowitzki em quarto com 29,0%.

A Figura 3.8 revela os padrões de arremesso de Steve Nash, Ray Allen, Dirk Nowitzki, e Kobe Bryant. Estes quatro jogadores possuem os maiores valores de *Range*, além disso, estas visualizações esportivas revelam diferentes pontos de eficácia entre eles. Por exemplo, Dirk Nowitzki arremessa poucos lances de 3 pontos em comparação com Ray Allen. Além disso, Dirk Nowitzki tem uma boa eficácia nas áreas de médio alcance pelo lado esquerdo da quadra. Ray Allen se destaca nos cantos da quadra onde Steve Nash raramente arremessa. Kobe Bryant demonstrou possuir uma eficácia mais homogênea entre os jogadores comparados.

Os quadrados maiores indicam áreas onde ocorrem mais tentativas, os quadrados menores indicam áreas com menos tentativas. A cor dos quadrados indica uma média de pontos por tentativa para cada posição. As posições mais alaranjadas indicam posições de arremesso com mais conversões, as posições mais azuladas indicam posições de arremesso com menos conversões.

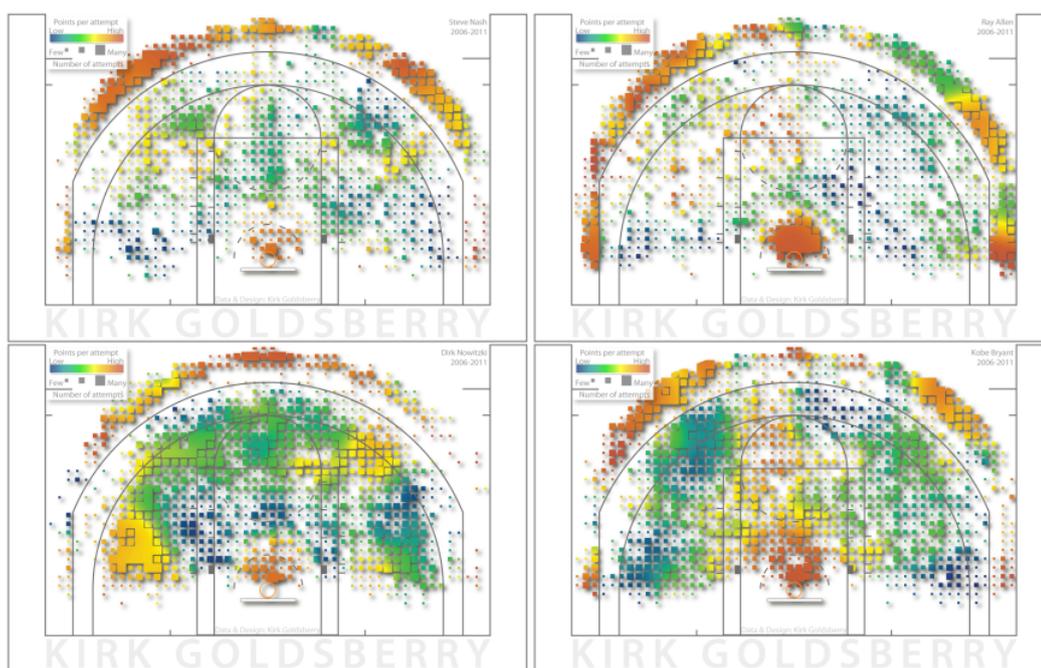


Figura 3.8: Visualização em *CourtVision*. Possibilitar comparar a efetividade dos jogadores em diversos locais das quadras.

Perin, Vuillemot e Fekete (2013) desenvolveram o *SoccerStories* um sistema que combina várias visualizações para permitir que os analistas analisem qualitativamente os dados de futebol explorando as fases do jogo. Em *SoccerStories* as fases do jogo de futebol são descritas como uma sequência de ações de um time até que a bola seja perdida,

ou seja, as fases são separadas pelas transições de posse de bola entre os times.

O sistema funciona da seguinte maneira, um jogo é escolhido e carregado para dentro do sistema. O espaço de trabalho central é um campo de futebol, Perin, Vuillemot e Fekete (2013) constataram que o campo de futebol é o principal objeto de observação e análise dos especialistas, a maioria dos especialistas que colaboraram no estudo, estruturaram seu modelo mental sobre a disposição espacial do time em função dos movimentos dos jogadores ao longo do tempo.

Além disso, uma linha do tempo na parte superior fornece uma visão geral da partida e sinaliza as principais fases do jogador selecionado, expostas em ordem cronológica. Os pontos vermelhos representam chutes ao gol perdidos, pontos verdes chutes ao gol convertidos, pontos azuis chutes ao gol defendidos. Ao se clicar em um ponto, outra fase é exibida. Cada fase selecionada contém outras visualizações menores que mostram mais informações. Há uma barra lateral que exibe as estatísticas do jogador selecionado, a barra lateral também contém uma visão estática do time e usa como *layout* um campo de futebol com a escalação inicial do time.

As conexões entre as pequenas visualizações menores na fase selecionada, codificam um tipo de ação, um passe aéreo é sombreado, um passe pelo campo é tracejado, um jogador que se movimenta com a bola é um traço ondulado. Uma comparação de fases foi adicionado ao lado direito, mostrando todas as fases relevantes do jogador selecionado, provendo uma visão geral de cada fase e permite tanto comparar quanto selecionar outras fases do jogador selecionado. Foi adicionado um espaço para anotações textuais, pois os analistas tendem a escrever textos curtos durante a análise. As pequenas visualizações usando mapas de calor em tons de ciano e marrom representam os escanteios ou cruzamentos feitos entre os jogadores (Figura 3.9).

SoccerStories foi implementado usando *D3.js* e *JQuery*. Segundo os autores, ele funciona em todos os navegadores modernos e consulta um banco de dados para carregar dinamicamente os dados JSON processados do servidor. Essas consultas retornam fases pré-calculadas e estatísticas do jogador, simplificando o processamento das consultas e a transferência de dados resultantes para a aplicação.

Cox e Stasko (2006) desenvolveram o *SportVis*, uma ferramenta para a exploração dinâmica de jogos de beisebol, visando ser flexível o suficiente para apoiar tarefas analíticas relacionadas ao desempenho de times de beisebol durante a temporada. Os autores afirmam que um único jogo de beisebol produz uma grande quantidade de dados. Um jogo de beisebol consiste em nove *innings* (entradas), e cada uma, com múltiplas *plate appearance* (isto é, quando um jogador é creditado cada vez que ele rebate a bola e muda de base). Cada *plate appearance* possui resultados discretos e pode ser descrita por diferentes atributos. É possível ainda separar o *plate appearance* em *individual pitches* (arremessos individuais). Nesse esporte, essa hierarquia permite que esses dados sejam coletados em vários níveis, possibilitando a criação de representações visuais interativas dessas agregações.

Esta versão do *SportVis* se concentrou em analisar o time durante a temporada e analisar um jogador em particular durante a temporada. Duas visualizações foram criadas para auxiliar nesta análise, sendo a primeira chamada de *Baseline Bar Display* (Figura 3.10a) que representa como foram os jogos de um determinado time durante a temporada.

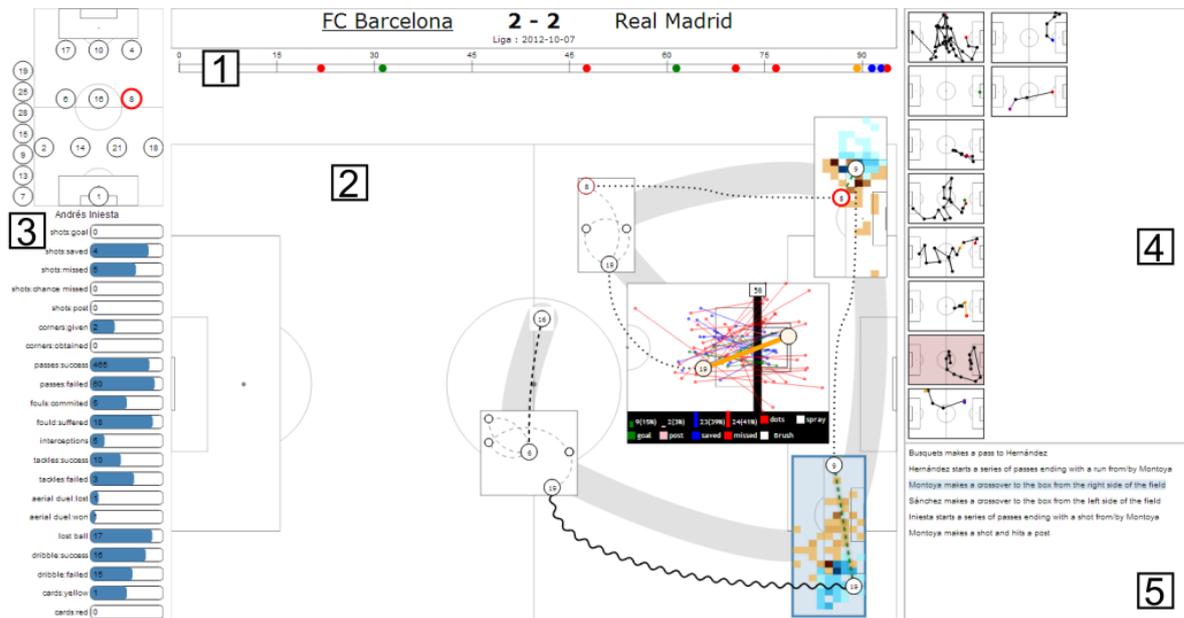


Figura 3.9: Interface principal do *SoccerStories*. A parte 1 apresenta uma linha do tempo com 10 fases relevantes nas quais o jogador selecionado participou. A parte 2 é um zoom temporal de uma fase selecionada do jogo com a participação do jogador selecionado sobre um *layout* de campo de futebol. A parte 3 apresenta estatísticas do jogador selecionado, no caso Andres Iniesta. A parte 4 apresenta um resumo em miniatura das 10 fases relevantes do jogador selecionado. A parte 5 apresenta um espaço para anotações textuais.

A segunda foi chamada de *Player Map* (Ver Figura 3.10b) que fornece uma visão geral rápida para identificar grupos de jogadores ineficazes durante o jogo.

De acordo com Cox e Stasko (2006) o *design* da *Baseline Bar Display* foi inspirado no gráfico meteorológico anual do *New York Times*. O gráfico do *Times* codifica múltiplas variáveis relacionando a temperatura ao longo de um eixo em uma série temporal, Cox e Stasko (2006) planejaram atingir uma densidade de informação semelhante na visualização *Baseline Bar Display*. O *design* do *Player Map* foi inspirado no *Smart Money Map of the Market* (Uma das primeiras visualizações na web do mercado de ações).

Na *Player Map* cada retângulo representa um jogador, e o tamanho do retângulo representa o número de vezes em que o jogador foi selecionado para o jogo ofensivo. A cor (codificada para exibir um espectro de verde a vermelho) representa valores altos para baixos. Cada retângulo codifica estatísticas específicas que o usuário pode observar e selecionar. De acordo com Cox e Stasko (2006), embora a média de rebatidas seja a medida tradicional de um rebatedor, outras estatísticas podem ser relacionados ao seu *run scored* (é a corrida no qual o jogador avança pela primeira, segunda, e terceira base, retornando com segurança à base inicial, tocando as bases nessa ordem).

Esta visualização provê uma maneira mais eficiente de alocar jogadores ao longo da temporada. Um grande retângulo vermelho e fortemente saturado sinaliza que um joga-

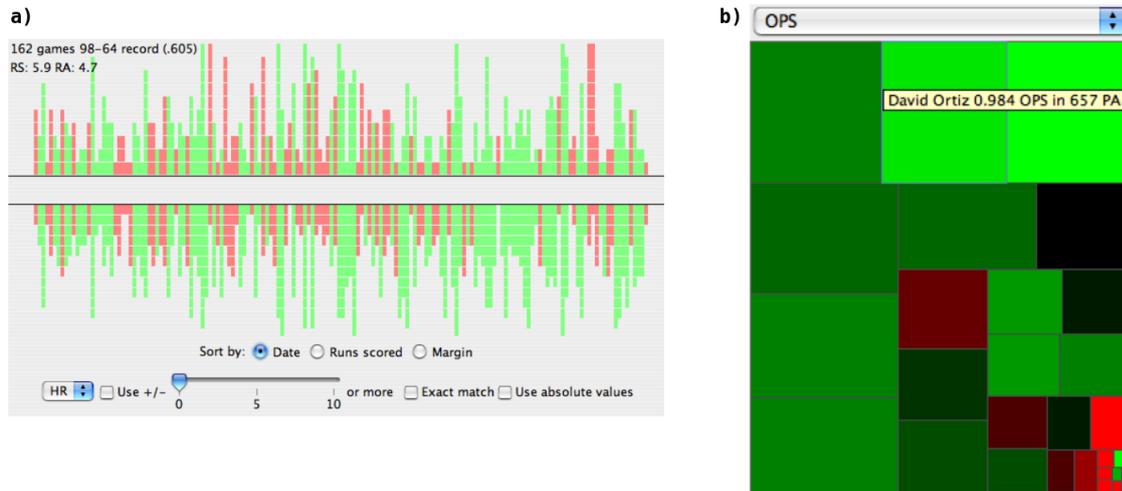


Figura 3.10: a) Baseline Bar Display b) Player Map (COX; STASKO, 2006).

dor ofensivo ruim está tendo muitas chances ou um jogador com habilidades defensivas recebendo oportunidades ofensivas. Um pequeno retângulo verde pode ser um jogador ofensivo mais eficiente, e candidato a receber mais oportunidades ofensivas, e que provavelmente pode não estar sendo escalado ou que pode estar sob tratamento médico.

3.3 VISUALIZAÇÃO EM CICLISMO

Díaz, Fort e Vázquez (2021) desenvolveram o *TourVis*. O foco deste aplicativo é explorar o progresso de todos os ciclistas em uma única etapa ou em todas as etapas de uma competição de ciclismo de estrada. A temporada profissional de ciclismo de estrada é composta por uma série de eventos. Há também corridas de um dia, como os campeonatos de *Milão-San Remo*, *Tour de Flanders*, *Paris-Roubaix*, *Liège-Bastogne-Liège*. Outras são corridas de várias etapas que vão desde uma semana como *Paris Nice* ou *Tour de Romandie* a três semanas de duração como o *Tour de France*, ou *La Vuelta*.

Díaz, Fort e Vázquez (2021) explicam que em corridas de várias etapas, vence o ciclista que acumular menos tempo nas etapas. Existem também dois tipos de etapas: corridas regulares e corridas cronometradas. Nas etapas regulares, todos os ciclistas largam ao mesmo tempo. Nas provas cronometradas, cada ciclista inicia após dois minutos do ciclista anterior, em ordem inversa em relação à classificação geral individual, sendo que o líder atual é o último a começar na etapa. Isso dá ao líder a vantagem de conhecer as parciais dos ciclistas anteriores. O *TourVis* fornece uma visualização que permite ao usuário ver o que aconteceu durante uma etapa, visualizando a história completa de cada etapa, como os ataques e fugas feitas por ciclistas, a mudança de líder durante uma ou mais etapas concluídas, a distância em relação ao líder, o comportamento dos diferentes pelotões formados durante uma determinada etapa, etc.

De acordo com Díaz, Fort e Vázquez (2021) em provas de ciclismo de estrada, geralmente os 4 dados a seguir são importantes: localização do ciclista (rastreamento conti-

nuamente por GPS), as classificações atuais dos ciclistas na prova, as informações dos ciclistas e equipes durante a prova, por fim as informações geográficas do local da corrida (o nível de inclinação, por exemplo, é muito relevante em provas de ciclismo de estrada). Neste sentido, informações desta natureza devem ser incluídas em uma ferramenta de visualização esportiva de ciclismo.

O *TourVis* facilita a exploração de provas de ciclismo de estrada em várias etapas. Esta aplicação permite que os usuários visualizem eventos importantes que ocorreram ao longo da prova, e não apenas a classificação final. Embora inovador, o trabalho apresenta limitações específicas, pois se restringe apenas a competições profissionais e estruturadas, contendo etapas e grupos de equipes. Adicionalmente, a falta de métricas de intensidade ou medidas de frequência cardíaca no *TourVis* apresenta uma limitação substancial na análise de desempenho individual dos ciclistas.

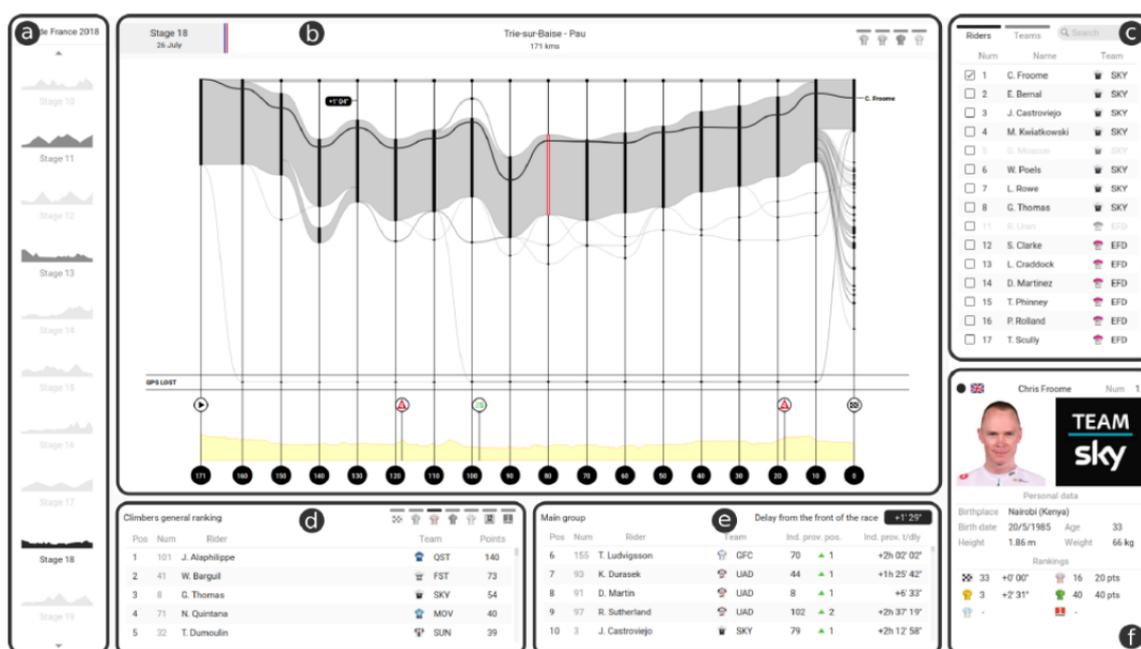


Figura 3.11: *Layout* principal do TourVis, que consiste em seis visualizações simultâneas: a) Visão Geral da Corrida, b) Visão de uma etapa específica, c) Painel de seleção de times e atletas, d) Ranking da etapa, e) Visualização de detalhes sob demanda e f) Informações específicas de um atleta e sua equipe.

Ainda sobre a visualização esportiva no ciclismo de estrada, duas visualizações merecem destaque no trabalho desenvolvido em Wood (2015). A primeira foi construída como alternativa para apoiar a história pessoal do ciclista, a segunda possibilita um comparativo entre os ciclistas. As visualizações também auxiliam na elaboração da estratégia de corrida e demonstração de desempenho de ciclistas em eventos de longa duração. Segundo Wood (2015), às corridas de longa distância geralmente envolvem muitas horas e muitos quilômetros para serem percorridos por ciclistas, tais eventos requerem dos atletas

uma significativa resistência, neste sentido os ciclistas pedalam sozinhos ou em pequenos grupos por muito tempo, ou até mesmo dias e, por isso, é comum os ciclistas escreverem relatos sobre suas próprias experiências após uma etapa para compartilhá-las com outros atletas.

Em regra, estes relatos são escritos em forma de diário, muitas vezes apoiados por fotografias. No entanto, com a melhor disponibilidade das tecnologias de rastreamento atuais se tornando acessíveis, há mais oportunidades para criar visualizações pessoais sobre a participação em atividades esportivas. Wood (2015) usou eventos de ciclismo de longa duração para demonstrar que visualizar o ciclismo com mapas de perfil e gráficos de coordenadas paralelas, bem como outros dados específicos do atleta, pode melhorar a escrita das narrativas pessoais sobre a prova de ciclismo.

Segundo Wood (2015), três critérios são importantes e devem ser considerados na estratégia para percorrer longas distâncias. Em primeiro lugar, a inclinação do terreno, esta impõem ao ciclista um excesso de esforço. Em segundo lugar, o vento pode ser um fator decisivo para o ciclista, que, dependendo da direção, ajuda ou atrapalha o progresso do atleta. Por último, os pontos de controle, onde os ciclistas têm oportunidade de repousar e comer.

Segundo Wood (2015), o mapa de perfil proposto originalmente por Huffman (2009) oferece uma visão alternativa que permite visualizar simultaneamente a elevação do terreno, alguns aspectos geográficos do percurso e adicionar narrativas pessoais. Esta visualização minimalista mapeia a rota como uma linha preta, e sua espessura varia proporcionalmente conforme o grau de elevação apresentado em cada trecho ao longo de todo o percurso. Essa proposta oferece a possibilidade de mapear outras variáveis, como a direção e velocidade do vento, e as cores podem ser utilizadas para representar outras variáveis, porém Wood (2015) afirma que não haviam dados suficientemente detalhados para expor a experiência do ciclista sobre o vento, pois este varia com a localização, altitude e clima. No entanto, ao manter um mapeamento geográfico da rota, é possível estimar a influência da direção do vento ao longo do percurso. O mapa de perfil proposto em Wood (2015) pode ser construído automaticamente após um evento com o auxílio de um DEM (Modelo de Elevação Digital, banco de dados topográfico com as elevações da superfície e elementos geográficos existentes sobre ela), ou fornecida por equipamentos de GPS.

A Figura 3.12 mostra um mapa de perfil gerado automaticamente do Audax Mille-Cymru de 1.400 km. Este *design* minimalista permite acrescentar narrativas de texto ao espaço com relativa facilidade, e, ao mesmo tempo, enfatizando a altitude do terreno, ou seja, demonstra as mudanças na dificuldade do terreno. Esta visualização pode proporcionar aos ciclistas bons *insights* para elaborarem melhor uma estratégia de corrida.

A segunda técnica de visualização proposta em Wood (2015) que merece destaque proposta é visualizar a posição relativa de um ciclista em relação a outros ciclistas. É muito interessante para os ciclistas entenderem o quão bem eles se saem em comparação com outros atletas no mesmo evento. Os ciclistas em eventos de longa duração podem estar separados por muitos quilômetros, horas, ou até mesmo dias. A visualização de posições relativas pode dar aos ciclistas a oportunidade de comparar o desempenho e experiências pessoais com outros atletas.

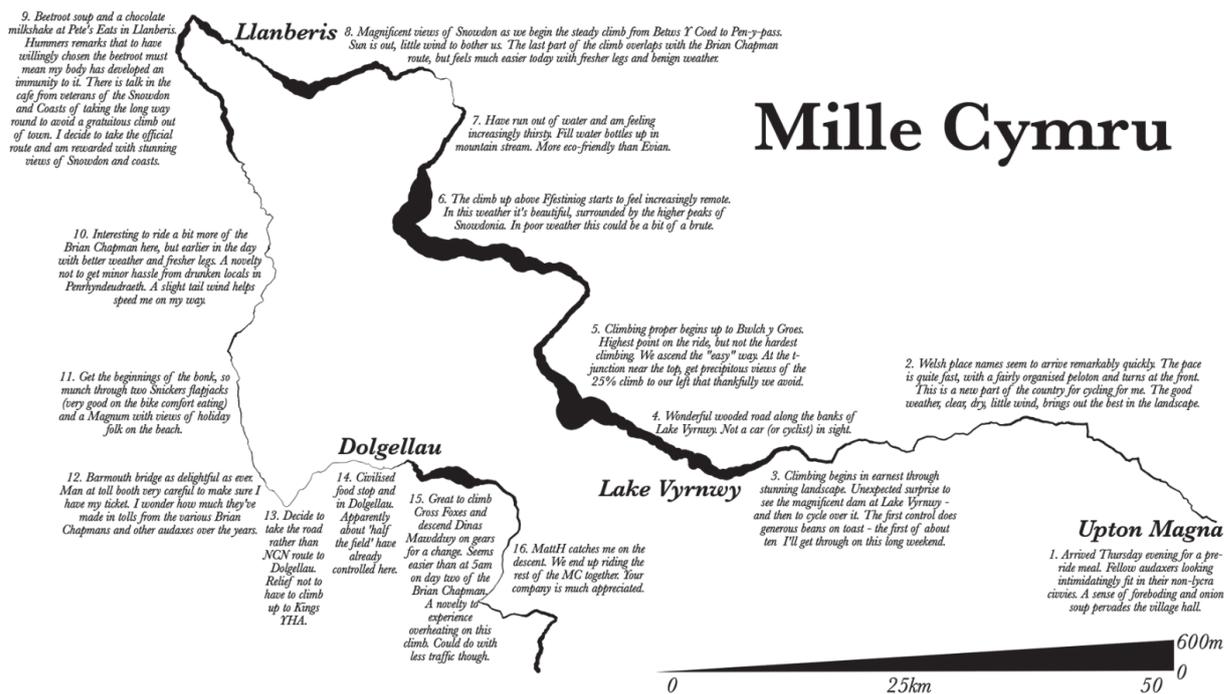


Figura 3.12: Mapa de perfil proposto por Wood (2015).

A visualização da posição relativa em eventos ciclísticos cronometrados é mais comumente representada como uma tabela classificada de tempos de conclusão, com tempos parciais medidos em pontos de verificação. Um evento Audax longo (1.000 km ou mais) geralmente pode incluir de 10 a 20 desses pontos de verificação, representando uma oportunidade de mostrar como a posição se altera no decorrer de um evento. A visualização da posição relativa foi proposta inicialmente por Batty (2006). A proposta de Wood (2015) usando a técnica de coordenadas paralelas é visualizar o progresso de todos os ciclistas comparativamente ao invés de mostrar o progresso de um único ciclista, ver Figura 3.13.

Esta visualização foi criada usando *Javascript* e *Processing*, para permitir uma interação baseada na Web. Os pontos mais escuros criados pela intersecção de muitas linhas demonstram o lugar geográfico onde houve mais mudanças da posição entre os ciclistas, segundo Wood (2015) os ciclistas podem ter optado por dormir enquanto outros continuaram. Além disso, Wood (2015) explica que os ciclistas mais rápidos no início do evento tendem a manter sua posição relativa ao longo do percurso, ou seja, a linha do ciclista apresenta menos ondulação no gráfico. Da mesma forma, aqueles que tiveram o tempo mais lento no início do evento tenderam a permanecer assim durante a maioria do evento. Segundo Wood (2015) a maioria das mudanças na posição relativa dos ciclistas ocorre entre *Market Rasen* e *Thirsk*.

Em Beck, Burch e Weiskopf (2016), foi proposta uma visualização de como as séries temporais podem ser visualmente comparadas usando uma representação matricial aplicada ao ciclismo. Esse conceito de exibir comparação de pontos de dados com células

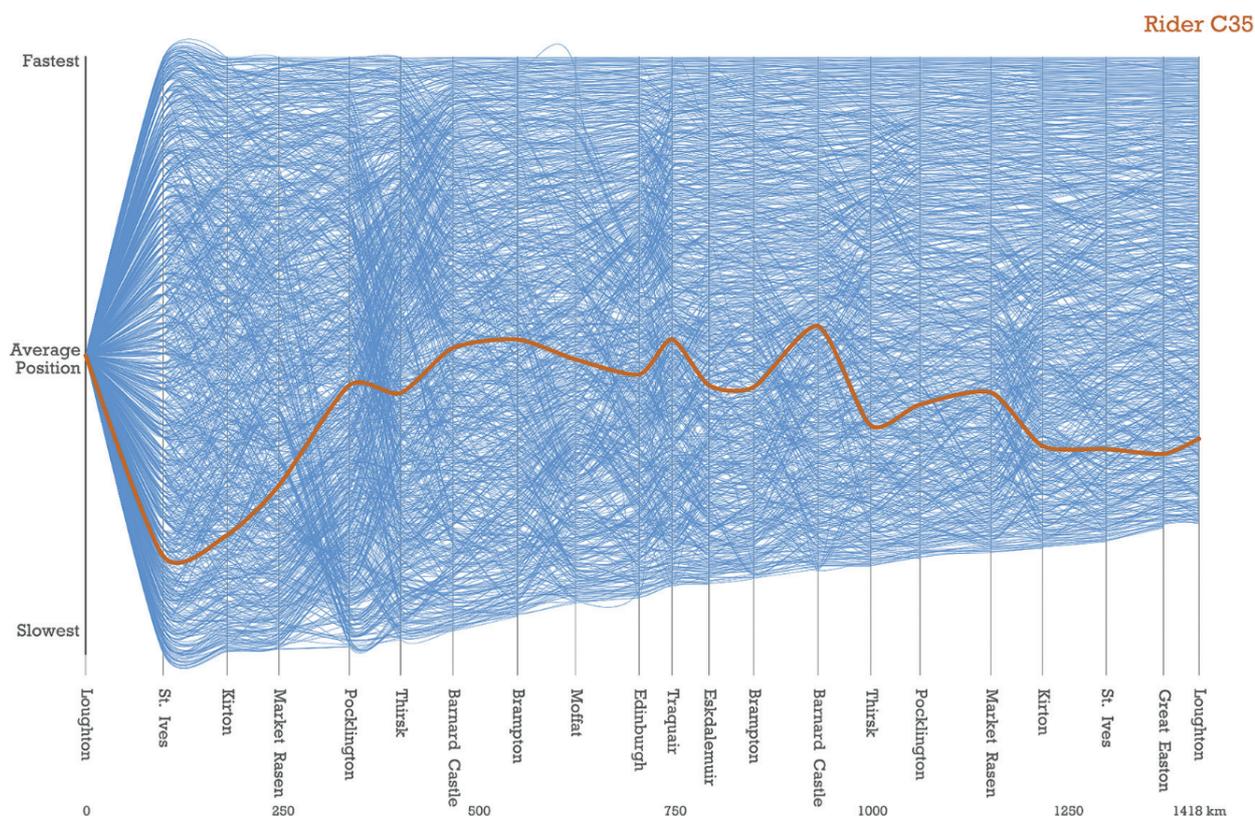


Figura 3.13: Visualização que mostra as posições relativas de 1000 ciclistas em 21 pontos de controle no Audax Londres-Edimburgo-Londres de 2013, com o progresso de um ciclista selecionado em laranja.

matriciais codificadas por cores pode ser poderosa porque permite a comparação de diferentes intervalos de dados em um único gráfico. Neste estudo de caso, Beck, Burch e Weiskopf (2016) investigaram e analisaram os dados de dois ciclistas durante o prólogo do *Tour de France* de 2012.

O prólogo é uma prova de curta duração onde todos os ciclistas são enviados na pista separadamente, um após o outro, além disso, é proibido o *slipstreaming* (técnica utilizada, especialmente pelos ciclistas que se aproveitam de um fluxo de ar menos intenso, em torno de outros ciclistas em movimento, num esforço para reduzir o próprio arrasto), ou seja, é proibido pedalar em um pelotão, pois um ciclista pode economizar bastante energia em relação ao ciclista da frente, em virtude das condições aerodinâmicas favoráveis em um pelotão. O prólogo geralmente ocorre em áreas urbanas, marcando o início dos principais campeonatos de ciclismo de estrada. No caso do *Tour de France* em 2012, segundo Beck, Burch e Weiskopf (2016) o prólogo, foi realizado em Liège, Bélgica, em uma pista de 6,4 km sem nenhuma subida relevante. Entre o prólogo e a última etapa, são disputadas várias outras, incluindo etapas de montanha, cronometragem individual e por equipe.

Beck, Burch e Weiskopf (2016) compara o desempenho de Kanstantsin Sivtsov, um ciclista bielorrusso da equipe Sky Procycling, com o ciclista Janez Brajkovic, um membro

esloveniano da equipe Astana Pro. Sivtsov terminou em 63^o lugar de 198 ciclistas, com 25 segundos de atraso em relação ao vencedor e Brajkovic ficou em 32^o lugar com 22 segundos de atraso em relação ao vencedor, Fabian Cancellara um especialista suíço em provas de tempo da equipe Radioshack-Nissan, venceu a etapa em 7 minutos e 13 segundos. Para tornar a exibição da matriz mais eficiente, já que havia períodos sem dados registrados, foi adicionado um gráfico linear que mede a progressão ao longo do tempo, ver Figura 3.14.

Ambas as séries temporais começam no canto superior esquerdo. A série temporal vermelha de Sivtsov é orientada horizontalmente e termina à direita, a série temporal azul de Brajkovic é orientada verticalmente e termina na parte inferior da visualização, as matrizes no centro fornecem uma comparação visual das séries temporais de desempenho dos ciclistas, e codificam todas as comparações de ponto de dados $n \times m$ possíveis. Segundo Beck, Burch e Weiskopf (2016), as partes B1 e B2 da Figura 3.14 comparam o desempenho dos ciclistas na primeira fase de preparação para o prólogo, nota-se que os gráficos possuem bastante variação, provavelmente os ciclistas precisam desacelerar para passar por curvas ou obstáculos. Já na parte B5 e B6 que representa as intersecções de B1 e B2, o vermelho é dominante em B5, revelando que Sivtvo teve melhor desempenho nas três primeiras voltas, enquanto Brajkovic teve melhor desempenho na última volta representada por B6.

Segundo Beck, Burch e Weiskopf (2016) a parte C1 e C2 representam a segunda fase de preparação em um ambiente controlado, no caso, representa os dados dos ciclistas em uma bicicleta de estrada estacionária, semelhante às bicicletas de academia, mas com adaptações para simular uma bicicleta de ciclismo de estrada. Em C1 observa-se que Brajkovic treinou múltiplos intervalos de intensidade, mas Sivtsov apresenta um padrão de treinamento diferente em C2, aumentando gradativamente a intensidade de seu treinamento na primeira parte, o retângulo C3 que representa a interseção de C1 com C2 apresenta mais padrões diagonais em vermelho. Enfim, este estudo de caso ilustra a aplicabilidade e utilidade da visualização esportiva, a visualização revelou estratégias distintas de preparação para a prova dos dois ciclistas comparados, esta visualização foi aplicada ao ciclismo mas também pode ser adequada para analisar o treinamento de outros tipos de esportes de resistência.

Em Fister, jr e Fister (2016) foi proposta uma visualização do treinamento de ciclistas, utilizando-se o software *ImageMagick* (software de edição de imagens em modo texto e de código aberto) e a linguagem de programação *Ruby*, a visualização visa ser simples e intuitiva ao mesmo tempo. É possível deduzir a distância percorrida, a duração, a posição do vento, a velocidade média, a altitude do terreno, a frequência cardíaca do atleta, a potência da pedalada e o sentimento do atleta. Os dados do atleta são obtidos e interpretados a partir de um arquivo de dados em formato .GPX ou .TCX e com base no esforço do atleta, são construídos valores específicos, a partir dos quais é possível obter informações significativas para cada categoria da atividade de treinamento realizada (Figura 3.15).

Segundo Fister, jr e Fister (2016), cada vez mais ciclistas estão usando rastreadores GPS para mensurar, monitorar e controlar os dados obtidos durante seus treinamentos, bem como realizar mais análises e planejamentos. Todos esses dados, que se tornaram

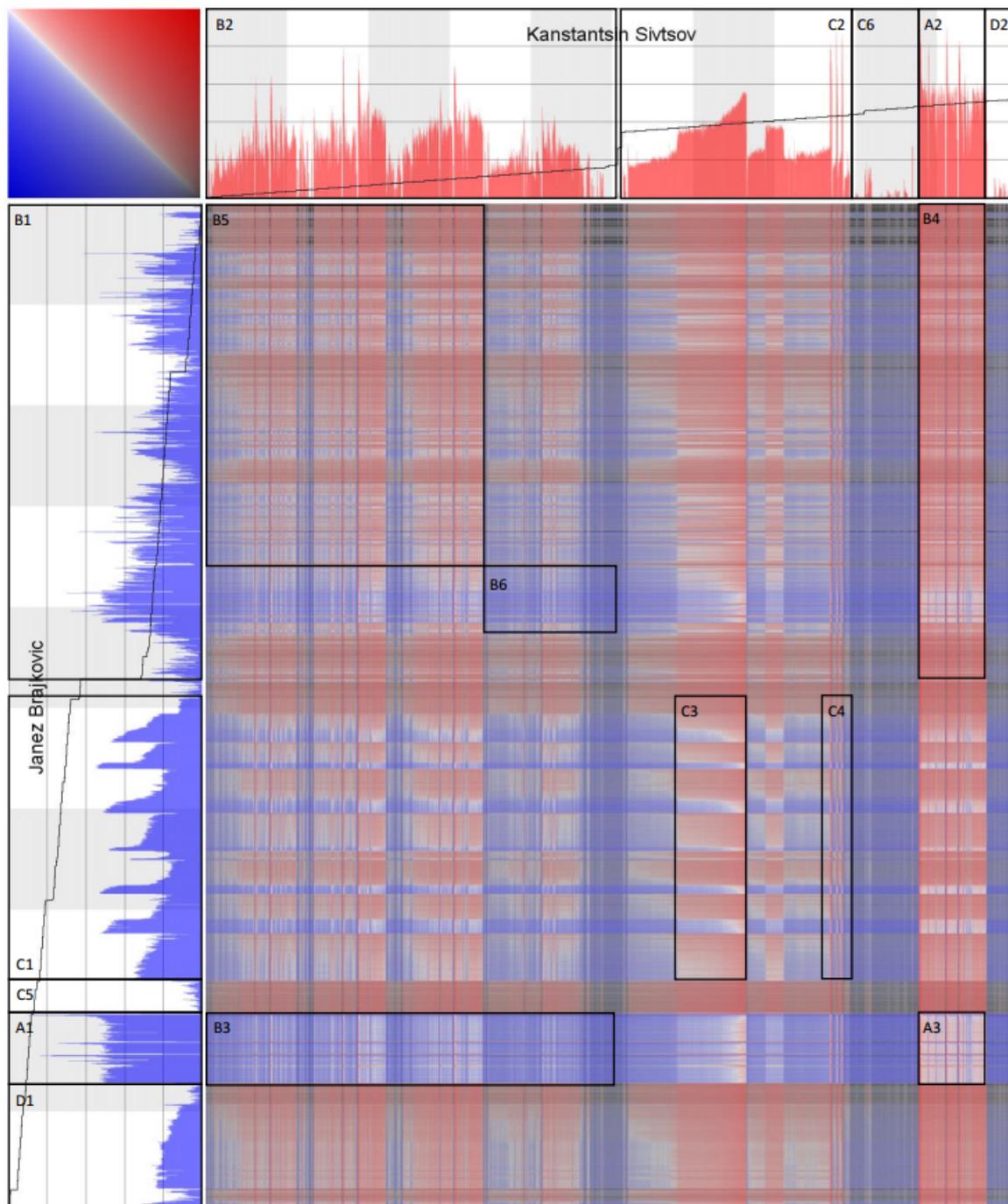


Figura 3.14: Comparativo de Sivtsov (vermelho) com Brajkovic (azul) no prólogo do Tour de France 2012, incluindo seu período de preparação para a corrida, o gráfico de linhas adicionais reflete o progresso do tempo e indica lacunas na medição.

mais acessíveis aos ciclistas, levam a grandes e complexas quantidades de dados, por exemplo, diferentes gráficos, coeficientes, valores médios, previsão de fadiga, avaliação

de desempenho, entre outros, podem ser extraídos dos conjuntos de dados. Esses dados devem ser interpretados corretamente para se tornarem informações úteis para um ciclista.

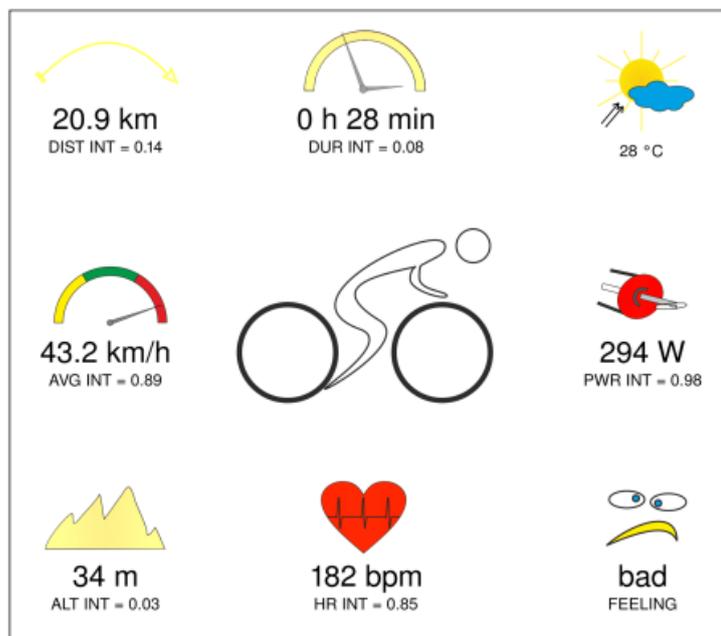


Figura 3.15: Visualização proposta por Fister, jr e Fister (2016) durante um treinamento para eventos de competição, os valores mostrados nessa figura são tendenciosos, mas o propósito da visualização e demonstrar rapidamente os principais valores obtidos em um possível treinamento.

Para uma leitura correta da Figura 3.15, é necessário considerar tanto a cor dos elementos quanto o valor ao lado deles. A curva representando a distância atualmente em amarelo vivo significa que o atleta aplicou pouco esforço neste item. O número abaixo representa a distância total pedalada e serve como informação de apoio na visualização. O velocímetro representa a velocidade e está dividido em três seções, para mostrar a velocidade de passeio, velocidade de treinamento e velocidade de competição. Um valor quantitativo é adicionado também para evitar uma leitura errônea da inclinação do velocímetro. O mesmo se aplica à duração, altitude, frequência cardíaca e do medidor de potência. As cores são escolhidas conforme o desempenho do atleta, elas variam entre amarelo brilhante, como baixo esforço, e vermelho-escuro, como alto esforço.

3.4 CONSIDERAÇÕES FINAIS

Este capítulo destaca alguns dos trabalhos do estado da arte em visualização esportiva, evidenciando que as visualizações esportivas tendem a ser promissoras com múltiplas estratégias de visualização.

Os trabalhos relacionados tem o propósito de trazer luz ao estado da arte e os desafios do tema da pesquisa, bem como servir de inspiração para a solução proposta. Nesse

sentido, foram apresentados trabalhos de visualização envolvendo qualquer esporte, seja em grupo ou individual. Posteriormente, foram discutidos trabalhos de visualização para o ciclismo, tema desse projeto de pesquisa.

Dentre os trabalhos de ciclismo, não há comparações de corridas de diferentes ciclistas em diferentes circuitos. Ademais, há também uma escassez de trabalhos que utilizam técnicas de visualização multidimensional considerando atributos de intensidade como frequência cardíaca e altimetria. Nesse sentido, o próximo capítulo apresenta uma abordagem para contornar essas lacunas.

CYCLEVIS**4.1 CONSIDERAÇÕES INICIAIS**

Conforme mencionado no capítulo anterior, as técnicas tradicionais de visualização têm limitações quando aplicadas aos esportes. Essas facilitam apenas a inspeção de uma ou de algumas variáveis e pode reduzir a compreensão de uma perspectiva esportiva específica. Para Qin (2019) um sistema de visualização de informações eficaz deve ser projetado com o objetivo primordial de minimizar a carga cognitiva associada à análise de dados, com recursos e funções que tornam os dados mais compreensíveis para a mente humana e orientam o analista a descobrir informações não evidentes e potencialmente importantes.

Nesse contexto, emerge o desafio de criar novas ferramentas de visualização esportiva que sejam versáteis o suficiente para suportar tarefas interativas, ao mesmo tempo, em que promovam a exploração que resulte em ganhos analíticos sobre a dinâmica dos dados. O sistema de análise visual desenvolvido e apresentado neste capítulo visa contornar uma lacuna de comparação visual de desempenho no ciclismo ao apresentar o desempenho de diferentes ciclistas em diferentes circuitos, utilizando técnicas de visualização multidimensional e medidas de intensidade.

A Seção 4.2 descreve o *pipeline* de visualização dos dados deste trabalho. Na Seção 4.3 apresenta-se as duas fases do desenvolvimento da arquitetura desta proposta, a fase de ambiente de desenvolvimento e a fase de ambiente de produção. Na Seção 4.4 aborda-se o conjunto de dados e as estratégias de pré-processamento adotadas para “limpar” e normalizar seus valores. Na Seção 4.5, especificam-se os grupos de usuários e objetivos de projeto sobre os quais o CycleVis focou seu desenvolvimento e as tarefas analíticas relacionadas à compreensão desses dados esportivos. A Seção 4.6 exibe a disposição visual do projeto. A Seção 4.7 explica o funcionamento do CycleVis e seus painéis visuais que permitem analisar e conduzir uma exploração de dados sob demanda, guiando o usuário na descoberta de informações de seu interesse. Por fim, na Seção 4.8 estão as considerações finais deste capítulo.

4.2 PIPELINE DE VISUALIZAÇÃO DOS DADOS

A Figura 4.1 apresenta o *pipeline* de visualização do CycleVis influenciada pelo modelo de Visual Analytics proposto por Ward, Grinstein e Keim (2010). Os dados a serem visualizados, estão no Github¹, estruturados para garantir rápido acesso, facilitando sua visualização. A seleção de dados envolve identificar o subconjunto de dados que será potencialmente visualizado, isso ocorre sob controle do usuário por meio das ferramentas de seleção do CycleVis.

O mapeamento de dados para visualização é o cerne do *pipeline* de visualização e realiza o mapeamento dos valores dos dados para os painéis visuais do CycleVis. Esse mapeamento envolve o processamento dos dados antes do mapeamento, por exemplo, cálculos matemáticos, ordenação de dados, normalização. A renderização ou geração da visualização específica dos objetos de visualização consiste no uso das técnicas de visualização, incluindo todas as informações suplementares para facilitar a interpretação dos dados, neste ponto acontece a extração de informações pelo usuário.

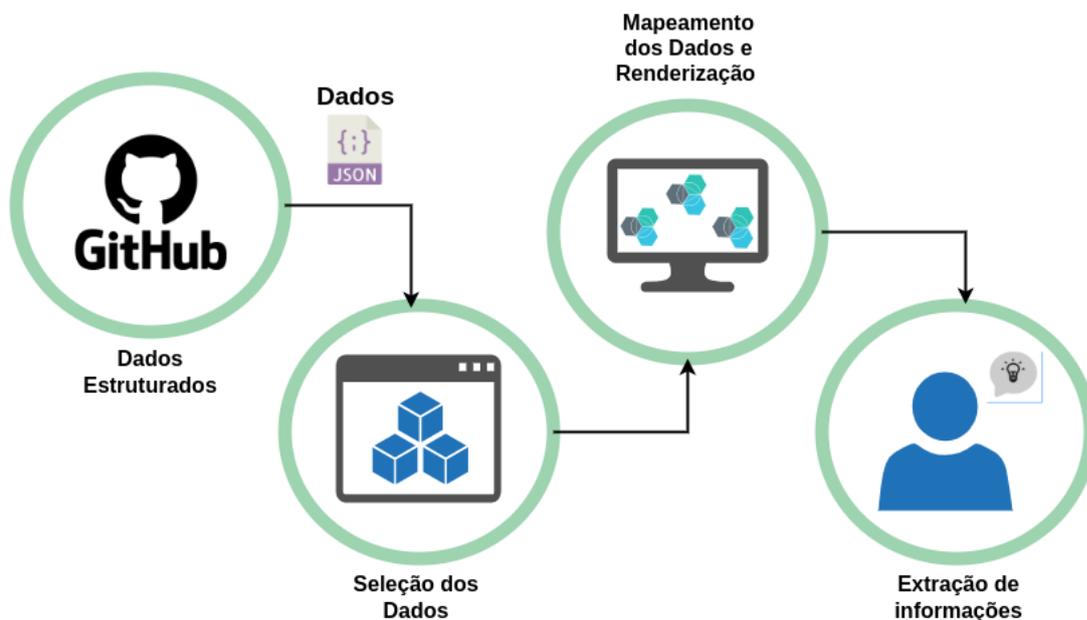


Figura 4.1: *Pipeline* da proposta de visualização de dados.

4.3 ARQUITETURAS DO CYCLEVIS

A arquitetura do sistema atravessou duas fases, a fase de ambiente de desenvolvimento e ambiente de produção. A primeira consistiu na implementação de um *design* de arquitetura Model-View-Controller (MVC) para possibilitar a construção mais rápida da aplicação, a fim de investir mais tempo na construção das visualizações. Já a segunda fase, o *front-end* (parte da aplicação com a qual o usuário irá interagir diretamente) foi

¹Disponível em: <https://github.com/>. Acessado em 27/06/2023

adaptado para funcionar similar a uma aplicação em formato SPA (*Single Page Application*), adaptando-se às interfaces de API disponibilizadas pelo Github, o foco foi era obter escalabilidade e renderização das visualizações sem incorrer em custos financeiros adicionais de hospedagem.

4.3.1 Fase de ambiente de desenvolvimento

Durante esta fase, a arquitetura foi construída utilizando um ambiente local (localhost), empregando a linguagem de programação PHP² (*Hypertext Preprocessor*). A arquitetura MVC (*Model-View-Controller*), é uma abordagem de projeto de software que separa a aplicação em três componentes interconectados, a lógica de negócios (*Model*), a interface do usuário (*View*) e o controle de fluxo de dados (*Controller*).

Foi utilizado a linguagem PHP por ser voltada para web e oferecer suporte à maioria dos servidores web atuais como Apache e Nginx. O SGBD utilizado foi o MySQL por ser de código aberto. Para virtualizar a infraestrutura foi escolhido o Docker³, o qual também é de código aberto e poderia facilitar em eventual instalação e implantação do sistema em servidores VPS (Servidor Virtual Privado), minimizando conflitos entre versões, bibliotecas e sistemas operacionais (Figura 4.2).

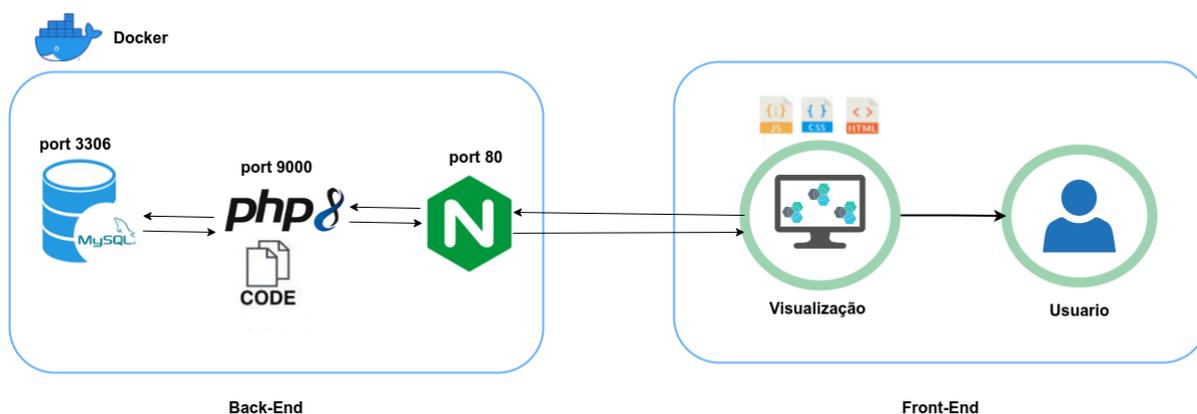


Figura 4.2: Arquitetura do CycleVis na fase de ambiente de desenvolvimento.

O fluxo de dados nesta arquitetura segue geralmente a seguinte ordem:

1. Uma solicitação do cliente chega ao servidor Nginx.
2. Se a solicitação for para um recurso estático, o Nginx servirá o arquivo diretamente.
3. Se a solicitação for para um script PHP, o Nginx passa a solicitação para o PHP-FPM.
4. O PHP-FPM processa o script, que pode envolver a comunicação com o banco de dados MySQL para buscar ou atualizar dados.

²Disponível em: <https://www.php.net/>. Acessado em 14/06/2022

³Disponível em: <https://www.docker.com/>. Acessado em 14/06/2022

5. Uma vez processado o script, o PHP-FPM retorna o conteúdo gerado ao Nginx.
6. O Nginx, por fim, retorna a resposta ao cliente.

Essa arquitetura oferece uma opção de rápido desenvolvimento e sem grande complexidade para servir aplicações web. Ao usar o Docker para separar cada componente em seu próprio contêiner, permite-se uma gestão e atualização mais fácil de cada serviço. Ao mesmo tempo, o uso de PHP-FPM e Nginx permite o processamento mais acelerado dos scripts PHP e por consequência, mais velocidade na entrega de conteúdo dinâmico e estático ao usuário.

O código-fonte PHP utilizado na fase de ambiente de desenvolvimento, para auxiliar no desenvolvimento do *front-end* da aplicação, está disponível *online*⁴, com livre acesso. No *front-end*, a visualização é renderizada utilizando principalmente HTML5, CSS3 e JavaScript através das bibliotecas Bootstrap, D3.js, JQuery, Turf.js, Leaflet.js e Apache Echarts.

4.3.2 Fase de Ambiente de Produção

Após o desenvolvimento da aplicação e com a base de dados consolidada, o projeto se deparou com desafios relacionados a escalabilidade, armazenamento de dados e à hospedagem sem custos financeiros. Para superar esses obstáculos, o Github foi adotado como solução para funcionar como *back-end* da aplicação. A utilização do Github como *back-end* apresentou duas vantagens principais.

Primeiramente, o serviço é gratuito, reduzindo os custos associados ao armazenamento de dados e código da aplicação. Em segundo lugar, o Github disponibiliza uma API robusta, permitindo a integração fácil e eficaz com o *front-end* da aplicação, possibilitando que o *front-end* funcionasse similar a um SPA. Esta API facilita a comunicação entre os dados e o *front-end*, tornando a troca de informações mais fluida e eficiente. Ademais, utilizar o Github garantiu que todas as questões técnicas referentes à hospedagem e configuração de servidores não diminuíssem o tempo valioso destinado ao desenvolvimento das interfaces de visualização. Essa vantagem foi essencial para otimizar o foco principal deste estudo.

As aplicações do tipo SPA são um tipo de aplicação web que carrega uma única página HTML e, em seguida, atualiza dinamicamente essa página conforme o usuário interage com a aplicação. Ao contrário das aplicações web tradicionais que recarregam a página inteira quando um usuário solicita uma nova página ou conteúdo, as SPA's usam JavaScript e APIs para atualizar seletivamente partes da página, melhorando a agilidade e desempenho do sistema, além de melhorar também a experiência do usuário.

⁴Disponível em: <https://github.com/iurygdeoliveira/mestrado>

⁴Disponível em: <https://getbootstrap.com/docs/5.0/getting-started/introduction/> Acessado em 26/06/2023

⁴Disponível em: <https://d3js.org/> Acessado em 26/06/2023

⁴Disponível em: <https://jquery.com/> Acessado em 26/06/2023

⁴Disponível em: <https://turfjs.org/> Acessado em 26/06/2023

⁴Disponível em: <https://leafletjs.com/> Acessado em 26/06/2023

⁴Disponível em: <https://echarts.apache.org/> Acessado em 26/06/2023

No sistema em questão, os dados são devolvidos pelo Github, entregando objetos JSON⁵ ao *front-end* do CycleVis. Estes pacotes contêm uma variedade de informações relativas às pedaladas selecionadas, incluindo pontos de latitude e longitude, valores de distância, velocidade, frequência cardíaca, elevação e tempo gasto entre cada par de pontos de latitude e longitude. Além disso, os pacotes também contêm o nome do país e da localidade correspondente, a duração das pedaladas, o total de trackpoints, e as médias de velocidade, frequência cardíaca, elevação e temperatura.

Esses dados são então armazenados na memória em um formato projetado especificamente para atender os requisitos de entrada das bibliotecas javascript utilizadas na análise visual. Simultaneamente, esses dados são armazenados em cache no IndexedDB⁶. Esta estratégia tem o objetivo de acelerar a criação das visualizações caso os mesmos dados sejam solicitados novamente, ao permitir recuperar as informações do IndexedDB, evita-se a realização de novas requisições ao Github, que têm um custo de processamento mais elevado.

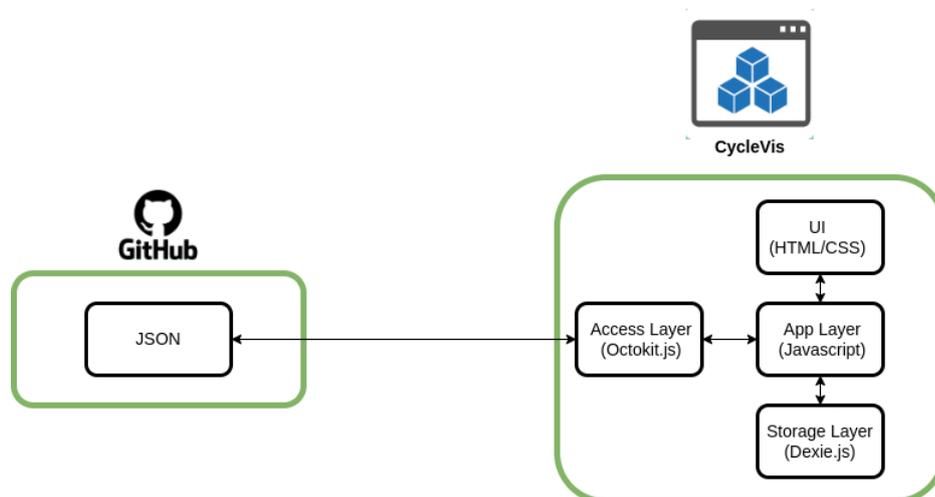


Figura 4.3: Arquitetura do CycleVis na fase de ambiente de produção. A Octokit.js⁷ é o SDK (*Software Development Kit*) disponibilizado pelo Github para integrações com seus repositórios. A Dexie.js⁸ é a biblioteca utilizada para gerenciar o IndexedDB.

Nessa fase, a comunicação entre Github e *Front-end* ocorre via arquitetura REST (FIELDING, 2000) com trocas de objetos JSON. A arquitetura REST usa protocolos e métodos HTTP (GET, POST, PUT e DELETE) para enviar e receber recursos. Isso permite, por exemplo, que dois sistemas se comuniquem em qualquer lugar do mundo sem estado, sessão ou cache. O JSON é um formato de dados compacto e independente

⁵Disponível em: <https://www.json.org/json-en.html>. Acessado em 14/06/2022

⁶Disponível em: https://developer.mozilla.org/en-US/docs/Web/API/IndexedDB_API Acessado em 26/06/2023

⁷Disponível em: <https://github.com/octokit/octokit.js/> Acessado em 26/06/2023

⁸Disponível em: <https://dexie.org/> Acessado em 26/06/2023

de linguagem de programação, que utiliza texto legível por humanos no formato atributo-valor.

O fluxo de dados nesta arquitetura segue geralmente a seguinte ordem:

1. **Repositório:** A origem dos dados é o repositório CycleVis_Dataset⁹. Objetos JSON são obtidos a partir deste repositório para atender as requisições do usuário.
2. **Camada de Acesso (Access Layer):** Esta camada atua como o ponto de entrada para os dados no Cyclevis. É implementada usando o Octokit.js, um SDK fornecido pelo GitHub que simplifica a interação com a API do GitHub. A camada de acesso recebe os objetos JSON e os passa para a próxima camada no sistema.
3. **Camada de Aplicação (App Layer):** Esta camada é implementada em JavaScript. Ela recebe os dados da Camada de Acesso e, dependendo do tipo de solicitação feita pelo cliente, a Camada de Aplicação pode se comunicar com a Camada de Armazenamento (Storage Layer) para armazenar os dados no IndexedDB ou iniciar o processo de renderização das visualizações. A Camada de Aplicação monta os painéis visuais para o usuário, onde são renderizados em um formato legível para o usuário.
4. **Camada de Armazenamento (Storage Layer):** Esta camada é responsável pelo armazenamento persistente dos dados no *front-end* da aplicação. Utiliza a tecnologia IndexedDB, um banco de dados baseado em índice que permite armazenar dados estruturados no lado do cliente. A Camada de Armazenamento recebe e armazena os dados passados para ela pela Camada de Aplicação.
5. **Renderização da UI:** Se a Camada de Aplicação determinar que os dados recebidos devem ser exibidos ao usuário, ela inicia o processo de renderização da interface do usuário. Esta é implementada usando HTML e CSS, tecnologias padrão para a criação e estilização de interfaces web.

A aplicação CycleVis, está disponível *online*¹⁰, com livre acesso.

4.4 CONJUNTO DE DADOS

O conjunto de dados que permeia a proposta do trabalho está relacionado ao esporte de ciclismo, contendo informações de ciclistas e pedaladas. Foi um enorme desafio encontrar um conjunto de dados contendo informações de diferentes ciclistas e diversas pedaladas para cada ciclista, e também contendo informações de desempenho como frequência cardíaca. O conjunto de dados foi obtido no material de Rauter e Fister (2015)¹¹, o qual apresenta informações mais detalhadas de como iniciar o pré-processamento dos dados

⁹Disponível em: https://github.com/iurygdeoliveira/cyclevis_dataset

¹⁰Disponível em: <https://iurygdeoliveira.github.io/cyclevis.github.io/>

¹¹Disponível em: <https://academictorrents.com/browse.php?search=sport&c6=1>. Acessado em 12/02/2022

como, por exemplo, quais bibliotecas utilizar para iniciar o *parse* (extração das informações) dos dados, visto que os dados das pedaladas estão armazenados em arquivos GPX ou TCX, as quais são formatos de intercâmbio GPS, esquemas XML para transferir dados GPS entre aplicações.

Cada pedalada é representada por um arquivo estruturado, nos formatos TCX ou GPX. Identificou-se uma inconsistência na padronização dos dados, mesmo em arquivos de formato idêntico, provavelmente devido à utilização de dispositivos GPS variados na coleta de dados. Para resolver esta questão, implementou-se uma rotina utilizando uma expressão regular que visou uniformizar a estrutura dos dados. Adicionalmente, várias técnicas de pré-processamento foram empregadas, incluindo a eliminação de dados imprecisos, padronização, remoção de entradas duplicadas, tratamento de dados omissos, mitigação de outliers, discretização de atributos e composição de atributos.

Do total de 18185 arquivos, foi possível extrair 7079 pedaladas, contendo os atributos de frequência cardíaca, latitude, longitude, velocidade, distância e temperatura. O atributo de elevação também foi adicionado à aplicação, porém foi obtido pela API do Google por meio da geolocalização (lat/long).

4.5 GRUPOS DE USUÁRIOS, OBJETIVOS E TAREFAS ANALÍTICAS

Nesta seção, especificamos os grupos de usuários, os objetivos da abordagem (OB1-OB7) sobre as quais o CycleVis focou seu desenvolvimento em relação aos cenários de extração de informação. São especificadas também tarefas analíticas (T1-T7) que um usuário deve conseguir realizar para obter informações ao realizar a análise visual exploratória.

4.5.1 Grupos de Usuário

No campo do ciclismo, a literatura existente se concentra predominantemente em auxiliar, especialistas em ciclismo e atletas profissionais a compreender e comparar o desempenho de suas pedaladas. No entanto, neste estudo, ampliamos nosso escopo para incluir grupos de usuários adicionais: atletas amadores e especialistas em visualização de dados, que podem contribuir significativamente para o avanço das ferramentas de análise de desempenho visual. À medida que essas ferramentas ganham adoção em diversos níveis de prática esportiva, observa-se que atletas amadores, apesar de possuírem um conhecimento limitado de análise de desempenho, estão interessados em utilizá-las para auxiliar em seu treinamento e na tomada de decisões.

No entanto, a confiança desses atletas nas decisões baseadas em tais ferramentas pode ser comprometida pela falta de conhecimento de como essas ferramentas analisam e interpretam seus dados de desempenho. Por outro lado, especialistas em visualização de dados possuem um conhecimento substancial sobre a interpretação e análise visual de dados, mas podem não estar familiarizados com as necessidades específicas dos atletas amadores. Portanto, este trabalho também auxilia a apresentar uma solução para os atletas amadores com o apoio de especialistas em visualização de dados.

4.5.2 Objetivos

Os objetivos (OB) foram definidos com inspiração em ideias encontradas nos trabalhos relacionados na Seção 3.1. A título de exemplo, o trabalho de Oliveira et al. (2013) identifica regiões de intensidade no percurso, possibilitando verificar como o contexto geográfico influencia na intensidade. As limitações observadas em *TourVis*, como relatado por Díaz, Fort e Vázquez (2021), propiciaram percepções valiosas para os objetivos do CycleVis. Inicialmente, destacou-se a importância de uma coleta de dados mais granular, um fator que motivou o CycleVis a adotar técnicas de coleta mais sofisticadas. A incapacidade do *TourVis* em capturar dados automaticamente em tempo real serviu de estímulo para desenvolver um sistema que suportasse atualizações dinâmicas e amostragens ajustáveis, possibilitando visualizações adaptáveis as pedaladas escolhidas, com funcionalidades de zoom.

O trabalho de Fister, jr e Fister (2016) também considerou a potência, obtida por aparelhos, e para reduzir o nível de ruídos. Em relação aos ciclistas, o estudo de Beck, Burch e Weiskopf (2016) desenvolveu uma matriz de comparação individual entre ciclistas, uma abordagem que possibilita o estudo de estratégias eficientes de treinamento no ciclismo, por isso consideramos relevante para nosso trabalho. Por fim, o trabalho de Wood (2015) foi inspirador, ao desenvolver uma visualização para comparar vários ciclistas ao longo do mesmo percurso realizado, um aspecto que consideramos crucial para nosso estudo. Todos esses trabalhos nos auxiliaram para definir os objetivos de projeto deste estudo.

- **OB1:** Selecionar e visualizar ciclistas e pedaladas. Espera-se que os usuários possam filtrar ciclistas e pedaladas utilizando determinados critérios, com o intuito de melhor compreender o desempenho dos ciclistas nas pedaladas.
- **OB2:** Identificar os diferentes padrões visuais para apresentar as distâncias nas pedaladas. Espera-se que o usuário possa compreender e comparar efetivamente as distâncias das pedaladas, identificando tendências, padrões e anomalias.
- **OB3:** Permitir a descoberta de padrões visuais para a compreensão do desempenho das pedaladas e ciclistas. Espera-se que o usuário compreenda como métricas de intensidade como frequência, elevação e velocidade influenciam no desempenho das pedaladas e ciclistas.
- **OB4:** Explorar a ideia e os gráficos da intensidade global como métrica que contempla todas as métricas de intensidade. Espera-se que o usuário identifique padrões no esforço global das pedaladas selecionadas.
- **OB5:** Descobrir se os usuários conseguem interpretar e comparar os indicadores médios de cada ciclista e das pedaladas selecionadas. Espera-se que essa comparação possa ajudar a identificar áreas de força ou fraqueza, padrões de desempenho, ou mudanças no desempenho ao longo do tempo.
- **OB6:** Avaliar a identificação de padrões georreferenciados nas pedaladas selecionadas. Espera-se que o usuário entenda melhor a geolocalização das pedaladas,

podendo servir de apoio para a descoberta de informações para métricas de intensidade.

- **OB7:** Descobrir diferentes informações por intermédio das coordenações entre as visualizações. Desse modo, possibilita compreender melhor os padrões de informações escondidos.

4.5.3 Tarefas Analíticas

Para cumprir os objetivos de projeto, foram determinadas um conjunto de tarefas analíticas a serem realizadas no CycleVis. Cada uma das tarefas a seguir estão relacionadas com o objetivo de projeto de mesmo número.

- T1: Medir a habilidade do usuário em filtrar e selecionar pedaladas de um ciclista específico.
- T2: Conferir a capacidade do usuário identificar os padrões visuais que mapeiam a distância das pedaladas.
- T3: Rastrear a perspectiva de comparação nas métricas de desempenho considerando múltiplas pedaladas e somente uma pedalada.
- T4: Verificar se o usuário consegue interpretar, compreender e comparar as intensidades globais das pedaladas.
- T5: Inspeccionar a perspectiva dos indicadores médios no painel visual Gráficos Estatísticos.
- T6: Mensurar a competência do usuário em explorar a localização geográfica das pedaladas selecionadas.
- T7: Apurar a habilidade do usuário em interagir com visualizações coordenadas.

4.6 DISPOSIÇÃO VISUAL

A disposição (*layout/design*) da interface visual foi projetada para uma experiência de navegação em monitores com resolução de Full HD (1920 x 1080). Contudo, teste exploratórios e de navegação responderam bem a resoluções a partir de 1760 x 768.

Optou-se por desenvolver uma aplicação Web contendo grupos de visualizações justapostas, distribuídas espacialmente para permitir ao usuário analisar e explorar visualmente o conjunto de dados, com suas respectivas instâncias (ciclistas e pedaladas). A Figura 4.4 apresenta o *dashboard* visual da abordagem, contendo todos os painéis visuais justapostos, com suas respectivas visualizações e filtros. Detalhes sobre os seus funcionamentos serão discutidos posteriormente.

Como mencionado, a composição dos painéis visuais segue a configuração classificada por Javed, Ghani e Elmquist (2012) como justaposição. A justaposição, como um método de apresentação de *layout*, possui vários benefícios, um dos quais é a capacidade facilitada

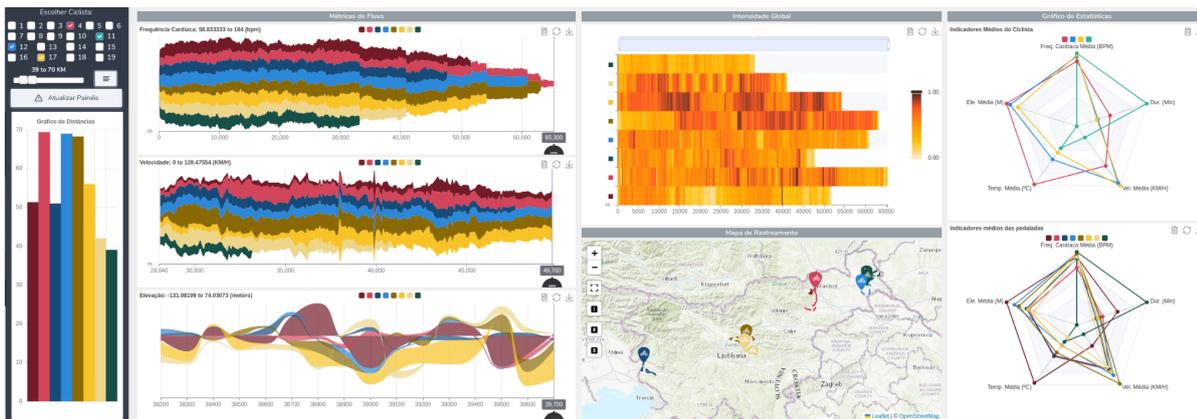


Figura 4.4: *Layout* do *dashboard* visual do CycleVis apresentando painéis visuais justapostos contendo dados de ciclistas e pedaladas, com informações de distância, elevação, frequência cardíaca e geolocalização.

de comparar informações diretamente. Quando visualizações correlatas são posicionadas em proximidade imediata, os usuários conseguem comparar as informações exibidas com mais eficácia, uma vez que não precisa usar rolagem ou mudar de tela, dificultando a capacidade cognitiva. Adicionalmente, essa abordagem se beneficia da predominância de monitores de computador com formatos retangulares, nos quais a dimensão horizontal é frequentemente mais extensa. Essa característica permite que se faça um ajuste para acomodar as visualizações para aproveitar o máximo de espaço disponível, otimizando, assim, a exibição dos resultados renderizados (SADANA; STASKO, 2016).

4.7 FUNCIONAMENTO DO CYCLEVIS

A Figura 4.5 apresenta uma versão do diagrama de sequência inspirado na UML e visa representar graficamente os componentes principais da abordagem com suas respectivas interações do usuário, dispostos em uma sequência temporal. Em linhas gerais, este diagrama demonstra como o controle de fluxo é realizado durante a utilização do CycleVis, a ser apresentado com mais detalhes a seguir.

4.7.1 Painéis Visuais para Escolha de Ciclistas e Pedaladas

A primeira e segunda etapa da Figura 4.5 é apresentada na Figura 4.6. O primeiro passo consiste na escolha de até cinco ciclistas, limitação estabelecida para evitar sobrecarga visual na renderização dos painéis visuais (Figura 4.6a). Durante a seleção dos ciclistas, o sistema mantém na memória do navegador (RAM) apenas os dados atualizados das distâncias de todas as pedaladas percorridas pelos ciclistas selecionados. A segunda etapa é realizada no local indicado na Figura 4.6b, e envolve a seleção do intervalo de distância percorrida considerando todas as pedaladas dos ciclistas selecionados, quanto maior o intervalo maior a quantidade de pedaladas a serem selecionadas na visualização Cycle Lens.

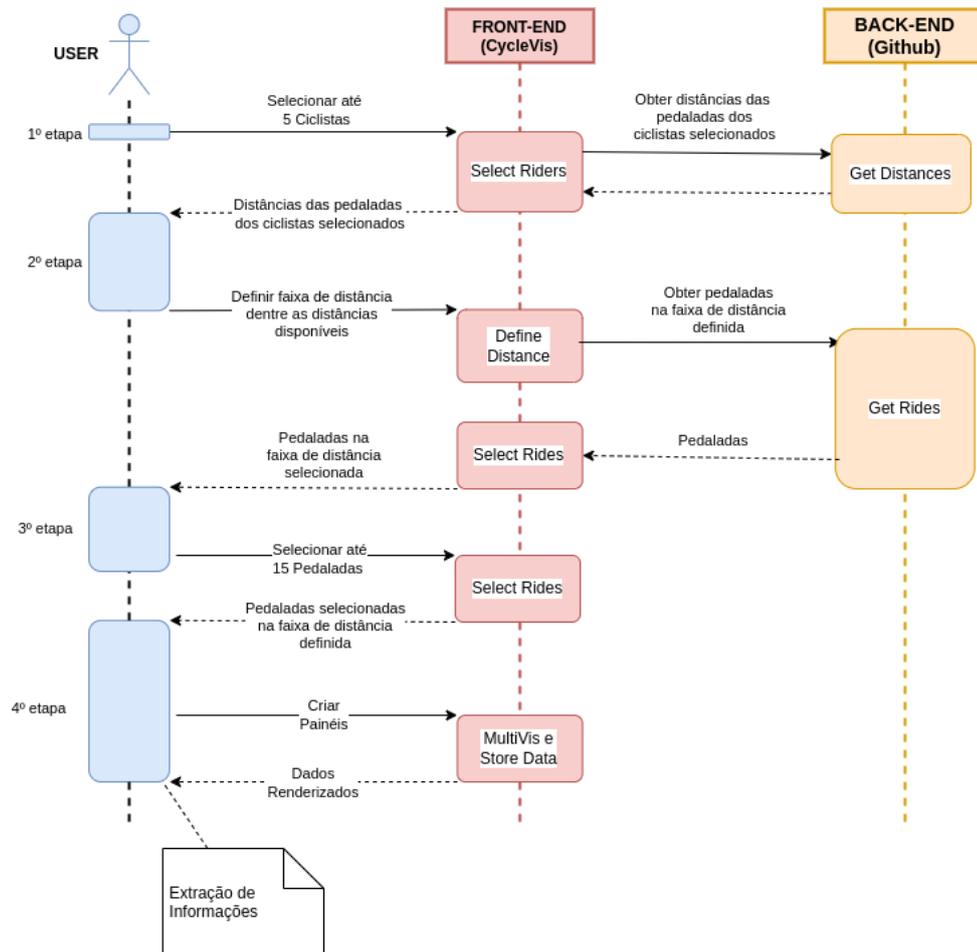


Figura 4.5: Diagrama de sequência do CycleVis indicando as interações dos principais componentes do software.

Cinco cores com matiz distintas foram escolhidas para representar cada um dos ciclistas, ilustrada na paleta de cores da Figura 4.7 (obtidas em Colors¹²). As linhas representam a saturação para cada ciclista, enquanto as colunas alteram brilho dessa cor, representando as pedaladas. Optou-se por essa paleta em razão do alto contraste da cor branca de fundo com os objetos do primeiro plano, aprimorando a legibilidade das informações Ware (2019). Importante salientar que na Figura 4.6a, a cor definida na primeira etapa irá definir os tons de cores para cada ciclista e suas pedaladas em todas as outras visualizações presentes na abordagem, facilitando a legibilidade.

A terceira etapa é acionada pelo botão “Ver Cycle Lens” conforme indicado na Figura 4.6c. Todas as pedaladas dos ciclistas selecionados são apresentadas na visualização *Cycle Lens* (Figura 4.8), que sumariza todas as pedaladas no formato de linhas e cada ciclista representa uma coluna. O tamanho da linha representa a distância da pedalada.

Essa visualização tem o modo compactado que diminui a espessura da linha e pos-

¹²Disponível em: <https://colors.co/> Acessado em 14/06/2022



Figura 4.6: a) Checkbox de escolha de ciclistas. b) *Slider* para definir as distâncias das pedaladas de cada ciclistas a serem comparadas. c) Botão *Ver Cycle Lens* para acessar a visualização *Cycle Lens* e selecionar até três pedaladas de cada ciclista. d) Botão *Atualizar Painéis* para geração do *dashboard* contendo os demais painéis visuais do *CycleVis*.



Figura 4.7: Paleta de Cores utilizada no *CycleVis*.

sibilita a exibição de um volume maior de pedaladas simultaneamente, ou o modo sem compactação que apresenta linhas mais grossas, permitindo mais detalhes sobre a distância. Recursos de interatividade, como: ao selecionar uma linha, que representa uma pedalada (seu comprimento de distância), ela é destacada e aumenta de tamanho e muda para uma cor mais escura. O valor nominal da distância é apresentado quando o usuário realiza o mouse-over em cima da linha. Outra interação possível é a ordenação das pedaladas por maior e menor distância. Essas ações representam a terceira etapa da Figura 4.5 e constam no Objetivo de Projeto 1.

Após selecionar as pedaladas na visualização *Cycle Lens*, é necessário clicar no botão “Atualizar Painéis” (Figura 4.6d), para dar início a quarta etapa. Nesta etapa os painéis visuais serão criados e renderizados em um *dashboard* para o usuário. De modo detalhado, os dados das pedaladas selecionadas são ordenados, primeiramente por ciclistas e depois por pedaladas, antes de serem fornecidos para as bibliotecas de visualização. Os dados são armazenados em vetores, dispostos de acordo com o ciclista e a pedalada, e

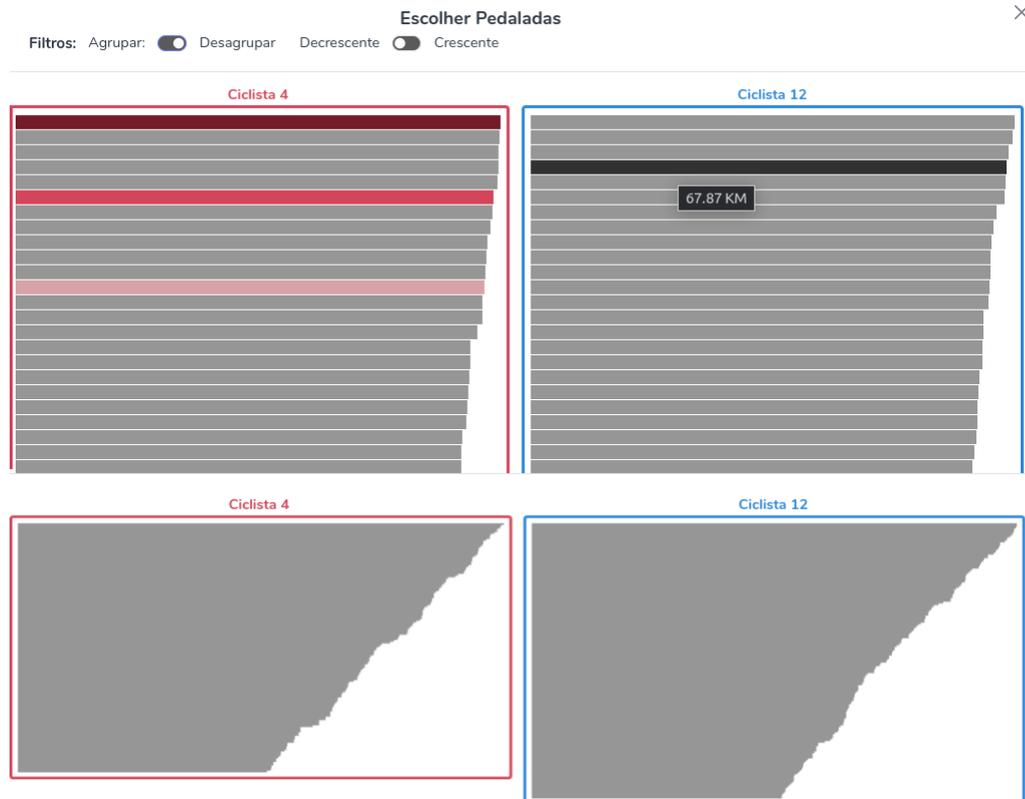


Figura 4.8: Visualização Cycle Lens de dois ciclistas no modo desagrupado e em ordem decrescente. Na parte de cima, o ciclista 4 possui 3 pedaladas selecionadas, enquanto o ciclista 12 possui uma pedalada, sendo essa selecionada e com distância de 67.87 km. Na parte de baixo, a mesma visualização com o modo agrupado ativado

compõem uma parte significativa da estratégia adotada para otimizar a velocidade de renderização e tempo de resposta de interação no CycleVis. Paralelamente, esses dados são armazenados em cache no IndexedDB, um banco de dados do navegador. Assim, a latência é reduzida na busca dos mesmos dados no Github, e por consequência, otimizar ainda mais a renderização das visualizações.

As próximas visualizações fazem parte do *dashboard* visual do CycleVis (Figura 4.4). O gráfico de barras vertical foi escolhido para representar a distância das pedaladas selecionadas (Figura 4.9). O eixo Y mapeia a distância em quilômetros, enquanto o eixo X mapeia as pedaladas. As cores remetem ao ciclista e as pedaladas escolhidas anteriormente. Como interação, ao passar o mouse em uma barra, é exibida a distância (em quilômetros). Essa visualização está entre uma das visualizações que pode auxiliar a obter o Objetivo de Projeto 2.

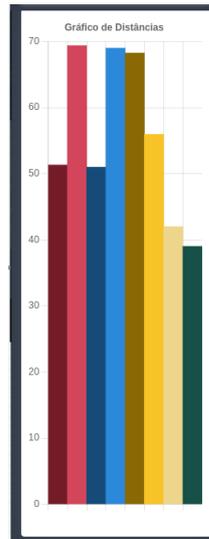


Figura 4.9: Visualização do Gráfico de Distâncias.

4.7.2 Painéis Visuais para Métricas de Desempenho

Para apresentar os dados relacionados ao desempenho, como os atributos de frequência cardíaca, velocidade e elevação, foram utilizadas visualizações baseadas no gráfico de área empilhado, mapeando a distância no eixo x e no eixo y o respectivo valor do atributo de desempenho (Figura 4.10). Para frequência cardíaca e velocidade foi utilizado o *Stream Graph* com sua representação em formato de ondas, a qual permite comparar facilmente o desempenho do atributo em diferentes pedaladas ao longo do percurso. Para apresentar a elevação foi utilizada um gráfico de área empilhado com formato suavizado das áreas em formato de ondas, por possibilitar verificar os declives (valores negativos) e aclives (valores positivos) ao longo do percurso, ou seja, o comportamento da elevação. Esse grupo de visualizações foi denominada de Métricas de Fluxo e estão localizadas no centro da interface (Figura 4.4f). Ao passar o cursor do mouse sobre os fluxos ocorre um *highlight* nos fluxos/linhas. Além disso, uma régua móvel ativa um *tooltip*, revelando informações detalhadas das variáveis (Figura 4.10A). Isso amplia a exploração do usuário uma vez que fornece detalhes sobre demanda, permitindo a descoberta de valores ponto-a-ponto.

Usando a régua móvel (Figura 4.10B) e mecanismos de zoom, é possível comparar simultaneamente as três métricas de desempenho para o mesmo intervalo de distância em todos as três visualizações, aprimorando a capacidade de comparação e identificação de padrões de desempenho para ciclistas e pedaladas distintas. É possível também selecionar pedalada(s) na legenda de cor (Figura 4.10C). Vale ressaltar que para cada fluxo/área, o desempenho de cada pedalada representa o tamanho da área entre as curvas dos seus respectivos fluxos/áreas. Espera-se que esse painel auxilie no Objetivo de Projeto 3.

O Algoritmo 1 apresenta a lógica para o painel de Métricas de Fluxo. Devido à natureza inerentemente ruidosa dos dados de GPS, a média móvel aritmética foi empregada como um método de suavização para diminuir os ruídos e melhorar a confiabilidade dos

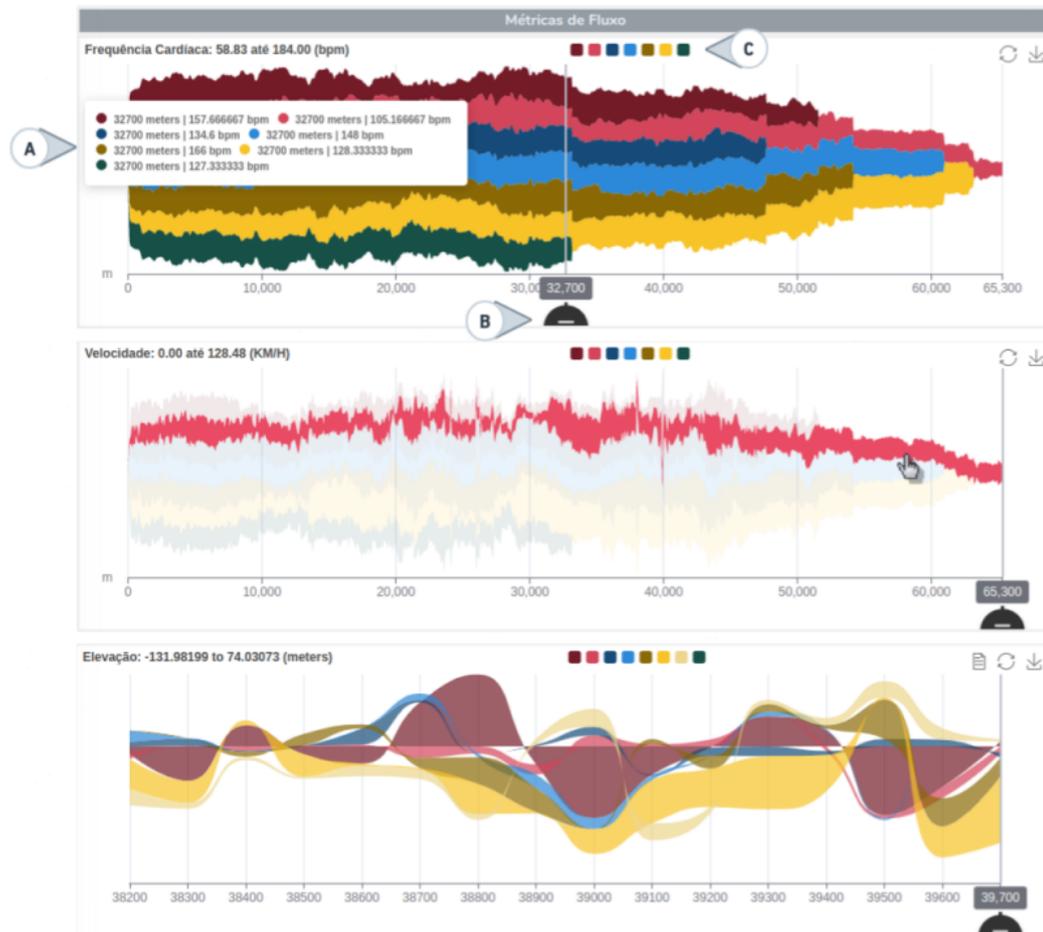


Figura 4.10: Visualização Métricas de Fluxo. a) Tooltip com valores absolutos b) Régua. c) Seletor de Pedaladas. O manuseio da régua ativa o tooltip, que mostra o valor do fluxo para a distância apontada pela régua

dados de frequência cardíaca, elevação e velocidade ao longo das pedaladas. Especificamente, os dados foram processados em segmentos de 100 metros, calculando-se a média dos valores dentro de cada segmento.

A linha 1 declara uma variável global “pedal_selecionados”, o qual é um vetor contendo as pedaladas selecionadas. A linha 2 demarca o início do procedimento “atualizarPedaladas”, cuja função é processar e atualizar as informações relativas às pedaladas. A linha 3 define uma variável local “segmentos”, destinada a armazenar segmentos de 100 metros de pedaladas. A linha 4 inicia uma estrutura de repetição (*foreach*), que percorrerá todos os itens no vetor “pedal_selecionados”. Na linha 5, uma estrutura condicional verifica se o elemento atual do vetor “pedal_selecionados” é nulo através da função “ChecarNulo”. Se a condição for verdadeira, o código prossegue. Na linha 6, a função “CriarSegmento” é chamada para cada item não nulo em “pedal_selecionados”, resultando em um novo segmento atribuído à variável “segmentos”. As linhas 7 a 9 chamam

Algoritmo 1: Lógica do Algoritmo para Calcular os dados no Métricas de Fluxo

```

1: global var pedal_selecionadas[tam] /* Pedaladas Seleccionadas */
2: procedure ATUALIZARPEDALADAS
   /* Segmentos de 100 metros */
3:   local var segmentos = {FreCard, elevacao, velocidade}
4:   foreach item in pedal_selecionados
   /* Calcula os dados das pedaladas apenas uma vez */
5:     if(ChecarNulo(pedal_selecionadas[item.id]) == true)
6:       segmentos ← CriarSegmento(item)
7:       segmentos.FreCard ← CalcDados(pedal_selecionados[item.id], 'freqCard')
8:       segmentos.elevacao ← CalcDados(pedal_selecionados[item.id], 'elevacao')
9:       segmentos.velocidade ← CalcDados(pedal_selecionados[item.id], 'velocidade')
   /* Armazena em Cache e atualiza em memória */
10:    pedal_selecionados ← armazenarPedalada(item, segmentos)
11:   endif
12: endfor
13: end procedure

14: procedure CALCDADOS(pedalada, atributo)
15:   local var trecho100
16:   local var dadosCalculados[tam]
17:   foreach item in pedalada
18:     trecho100 ← Media100(item[atributo])
19:     dadosCalculados ← empilha(trecho100)
20:   endfor
21:   return dadosCalculados
22: end procedure

```

a função “CalcDados” para cada item em “pedal_selecionados”, calculando e atribuindo os dados de frequência cardíaca, elevação e velocidade para o respectivo segmento.

Na linha 10, a função “armazenarPedalada” é chamada para cada item e seu respectivo segmento, atualizando o vetor “pedal_selecionados” com as novas informações. O procedimento “CalcDados” itera sobre cada pedalada, e calcula a média aritmética de um trecho de 100 metros para o atributo passado como referência na chamada da função, então é empilhado o resultado no vetor “dadosCalculados”.

4.7.3 Painel Visual para Métrica Global de Intensidade

O objetivo de apresentar visualmente os dados dos ciclistas e suas pedaladas é compreender melhor o(s) seu(s) desempenho(s), seja analisando cada atributo em separado (localmente) ou mesmo analisando um conjunto deles (modo global). Nesse sentido, foi desenvolvida uma métrica para quantificar o último caso, ou seja, uma métrica para analisar o desempenho global do ciclista. Para tal, foram utilizados os atributos de desempenho de velocidade, elevação e frequência cardíaca.



Figura 4.11: Visualização Intensidade Global. a) Mecanismo de zoom. b) Mecanismo de filtro. A utilização dos mecanismos de filtro e zoom que possibilita visualizar as intensidades entre 14400 e 28800 metros que estão acima da intensidade 0.5

A Figura 4.11 apresenta a visualização com o cálculo da intensidade global, baseado em *Heatmap*. Cada linha representa uma pedalada e cada barra vertical dentro da linha representa um segmento de aproximadamente 100 metros. A intensidade global de cada barra é computada como uma média de cada trecho de 100 metros de cada um dos atributos de velocidade, elevação e frequência cardíaca (Fórmula 4.1), seguido de um processo de normalização Min-Max usando a Fórmula 4.2. A cor mapeia o nível de intensidade global de cada segmento. No caso da figura, as cores brancas e amarelo

claro representam níveis de intensidade muito baixo, enquanto marrom e preto níveis de intensidade muito alto.

$$IG = (Frequencia100 + elevacao100 + velocidade100)/3.0 \quad (4.1)$$

$$IGn_j = \frac{IGi_j - \min(IG)_j}{\max(IG)_j - \min(IG)_j} \quad (4.2)$$

Na Fórmula 4.2, IGn_j representa o valor de intensidade global normalizado da pedalada j , IGi_j é o valor do segmento i da pedalada j , $\min(IG)_j$ é o valor mínimo de intensidade global da pedalada j , e $\max(IG)_j$ é o valor máximo da intensidade global da pedalada j .

Para aprimorar a experiência do usuário, foram adicionados mecanismos de interação na visualização. O recurso de zoom (Figura 4.11(A)) permite aos usuários aproximar os segmentos de intensidade das pedaladas, proporcionando uma visão mais detalhada dos dados. Isso pode ser particularmente útil para usuários que desejam analisar de perto um determinado intervalo e a variação da intensidade ao longo de suas pedaladas. A legenda de cor é interativa e permite selecionar o valor de intensidade na legenda de cor (Figura 4.11(B)), facilitando a identificação de trechos de maior ou menor intensidade no mapa de rastreamento, proporcionando aos usuários ajustar sua análise de desempenho livremente. Com base nessas considerações, espera-se que o Objetivo 4 seja alcançado com a implementação desta visualização.

4.7.4 Painel Visual Estatístico

A Figura 4.12 compila os dados quantitativos em duas metáforas visuais baseadas na visualização *Radar Chart*. Esta técnica permite a comparação de múltiplos atributos simultaneamente por meio da comparação de seus polígonos resultantes. Os atributos utilizados foram: frequência cardíaca (batimentos por minuto - BPM), elevação (metros - M), velocidade (quilômetros por hora - KM/H), temperatura (graus Celsius - C) e duração (minutos). Na parte de “Indicadores Médios do Ciclista” cada instância/polígono representa um ciclista e os vértices dos polígonos são obtidos pelo cálculo da média aritmética de todas as pedaladas selecionadas para aquele ciclista. Desse modo, é possível compreender na Figura 4.12 que a maior duração das pedaladas e menor temperatura média são do ciclista de cor verde, com 428,17 minutos de média, enquanto a maior temperatura média é a do ciclista vinho, com 19,6 graus Celsius. Essa análise visual multivariada entre diferentes ciclistas permite identificar padrões ou discrepâncias que podem ser úteis para estudos futuros ou para entender e otimizar o desempenho dos ciclistas. Como mecanismos de interação adicionou-se um *mouse-over* que apresenta uma dica com os valores médios dos ciclistas e habilita as pedaladas deste ciclista na parte de “Indicadores Médios das Pedaladas”.

Nesta visualização cada instância/polígono representa uma pedalada. Para identificar qual pedalada pertence a um ciclista, basta verificar as linhas com mesma saturação de

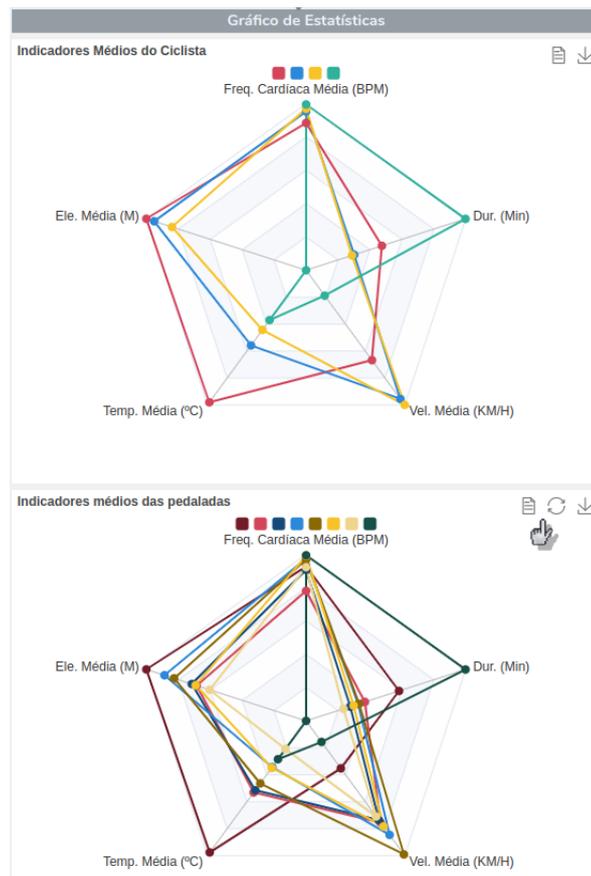


Figura 4.12: Visualização Gráfico Estatísticos. Parte superior com os indicadores médios dos ciclistas e parte inferior com os atributos médios das pedaladas.

cor. Do mesmo modo, é possível ler na Figura 4.12 que a pedalada de maior velocidade é do ciclista amarelo e a pedalada de menor elevação é a do ciclista verde. Como mecanismos de interação adicionou-se um mouse-over que apresenta uma dica com os valores das pedaladas. Estas visualizações podem ajudar a identificar pontos positivos e a melhorar no treinamento dos atletas ou mesmo em competições, permitindo aprimorar o desempenho geral.

Esta visualização permite uma comparação direta entre diferentes categorias. Isso é particularmente útil ao tentar identificar em quais variáveis os ciclistas estão se destacando. Em adição à comparação direta, também serve para analisar o desempenho individual de um ciclista em todas as variáveis simultaneamente. Esta análise pode ajudar a identificar áreas de força e fraqueza e sugerir onde o treinamento pode ser focado para melhorar o desempenho geral. Existindo correlações entre variáveis, por exemplo, uma elevada frequência cardíaca correlacionando-se com uma alta velocidade, essas tendências podem ser facilmente identificadas nesta visualização. Com base nessas considerações, espera-se que Objetivo 5 seja alcançado com a implementação desta visualização.

4.7.5 Painel Visual de Geolocalização

O Mapa de Rastreamento da Figura 4.13 contém múltiplas camadas que permite uma análise geoespacial multidimensional. Cada camada oferece uma perspectiva única sobre o comportamento dos ciclistas e suas pedaladas. Ao comparar diferentes camadas, podemos entender como os vários fatores se inter-relacionam e influenciam o desempenho dos ciclistas.

Neste estudo, definiu-se uma camada de rotas que pode fornecer informações valiosas sobre quais rotas são mais usadas ou preferidas pelos ciclistas. Isso pode servir para planejar ou otimizar rotas, por exemplo, para organizações que construam ciclovias ou para ciclistas que buscam rotas mais eficientes ou desafiadoras. Outra camada foi a de intensidade que pode oferecer percepções sobre onde os ciclistas tendem a pedalar com maior intensidade, o que pode ser um indicativo de áreas de treinamento específicas ou de terrenos mais desafiadores. Por exemplo, observa-se que o ciclista amarelo na Figura 4.13 pedala em maior intensidade em relação ao ciclista vermelho. Isso poderia, por exemplo, informar estratégias de treinamento ou planos de desenvolvimento de infraestrutura do local da pedalada. A camada de distância pode ajudar a entender o alcance geral dos ciclistas e os locais que eles tendem a alcançar ao pedalar. Essas informações podem ser úteis para planejar eventos ou serviços de ciclismo.

Alternando entre as diferentes camadas, é possível observar as correlações entre as variáveis. A título de exemplo, um aumento na intensidade pode estar ligado a uma rota específica ou a uma distância específica. Identificar essas correlações pode abrir caminho para estudos mais aprofundados. Um mapa de rastreamento é uma ferramenta visualmente atraente e fácil de interpretar. Ele pode ser uma forma eficaz de comunicar resultados e conclusões para uma variedade de públicos, incluindo ciclistas, analistas de desempenho, planejadores urbanos e pesquisadores. Com base nessas considerações, espera-se que o Objetivo 6 seja alcançado com a implementação desta visualização.

4.7.6 Visualizações Coordenadas

Por fim, foi desenvolvido o recurso de coordenação entre as visualizações Métricas de Fluxo e Intensidade Global, e entre as visualizações Intensidade Global e Mapa de Rastreamento, no qual a cada interação de uma determinada visualização, atualiza-se a outra(s), atuando como filtro dos dados.

Intensidade Global -> Métricas de Fluxo. As manipulações efetuadas na funcionalidade de *datazoom* na Visualização da Intensidade Global promovem a correspondente aproximação por distância na Visualização das Métricas de Fluxo. Tal interação permite que o usuário acompanhe a interação de uma determinada distância em Intensidade Global, sendo refletida nas variáveis de elevação, velocidade e frequência cardíaca. Essa relação tem o intuito de proporcionar uma melhor compreensão de como essas variáveis impactam coletivamente o desempenho do ciclista.

Métricas de Fluxo -> Mapa de Rastreamento. Ao manipular as réguas na visualização Métricas de Fluxo, ícones que representam elevação, velocidade e frequência cardíaca são posicionados no Mapa de Rastreamento. Cada ícone associa-se a um segmento de 100 metros no mapa, permitindo ao usuário uma correlação direta entre essas variáveis e as

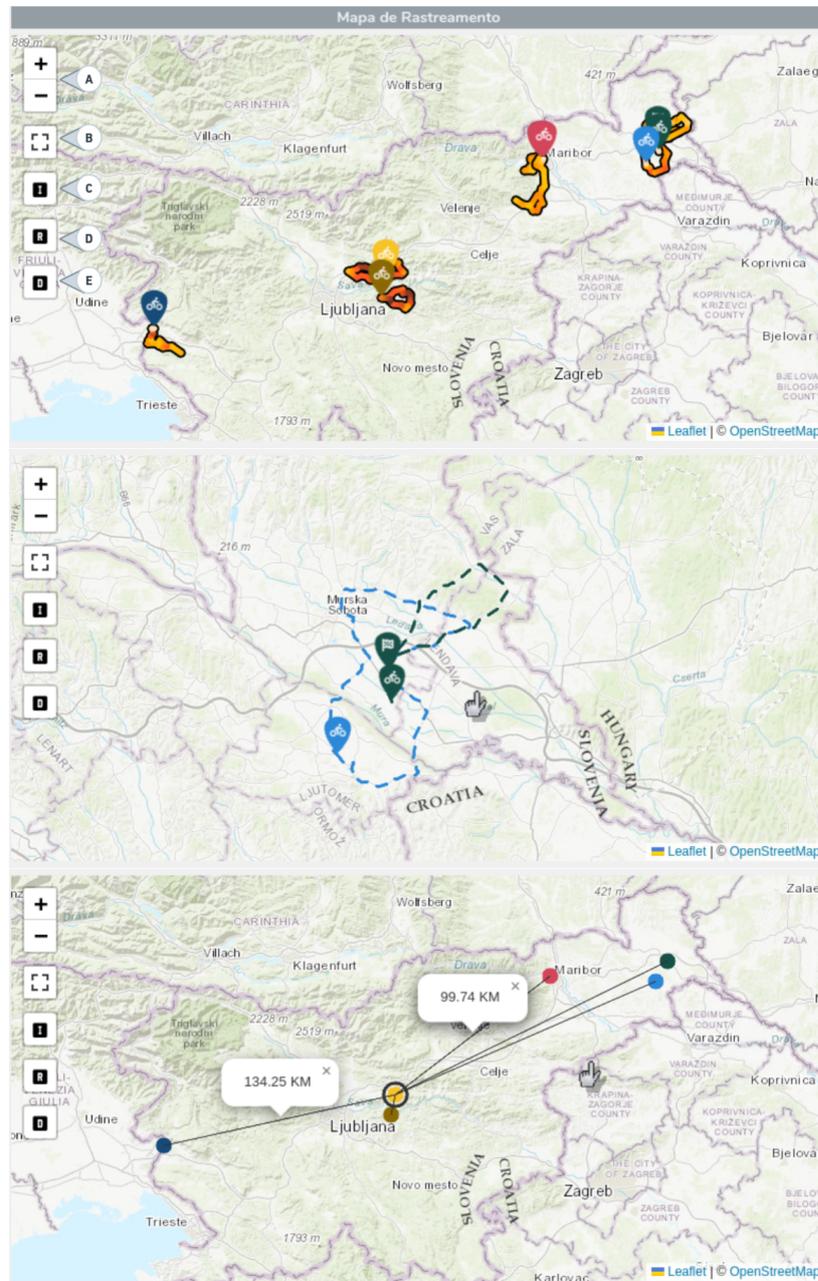


Figura 4.13: Visualização Mapa de Rastreamento. a) Habilita o zoom no mapa. b) Ativa o *fullscreen*. c) Ativa a camada de intensidade. d) Ativa a camada de Rotas. e) Ativa a camada de Distância

condições geográficas específicas do percurso. Esta funcionalidade oferece uma análise granular por trechos, enriquecendo a compreensão das variáveis em contexto geográfico.

A coordenação entre diferentes visualizações permitirá uma análise mais eficaz dos dados ao longo do tempo. Os usuários poderão rastrear como as variáveis de desempenho

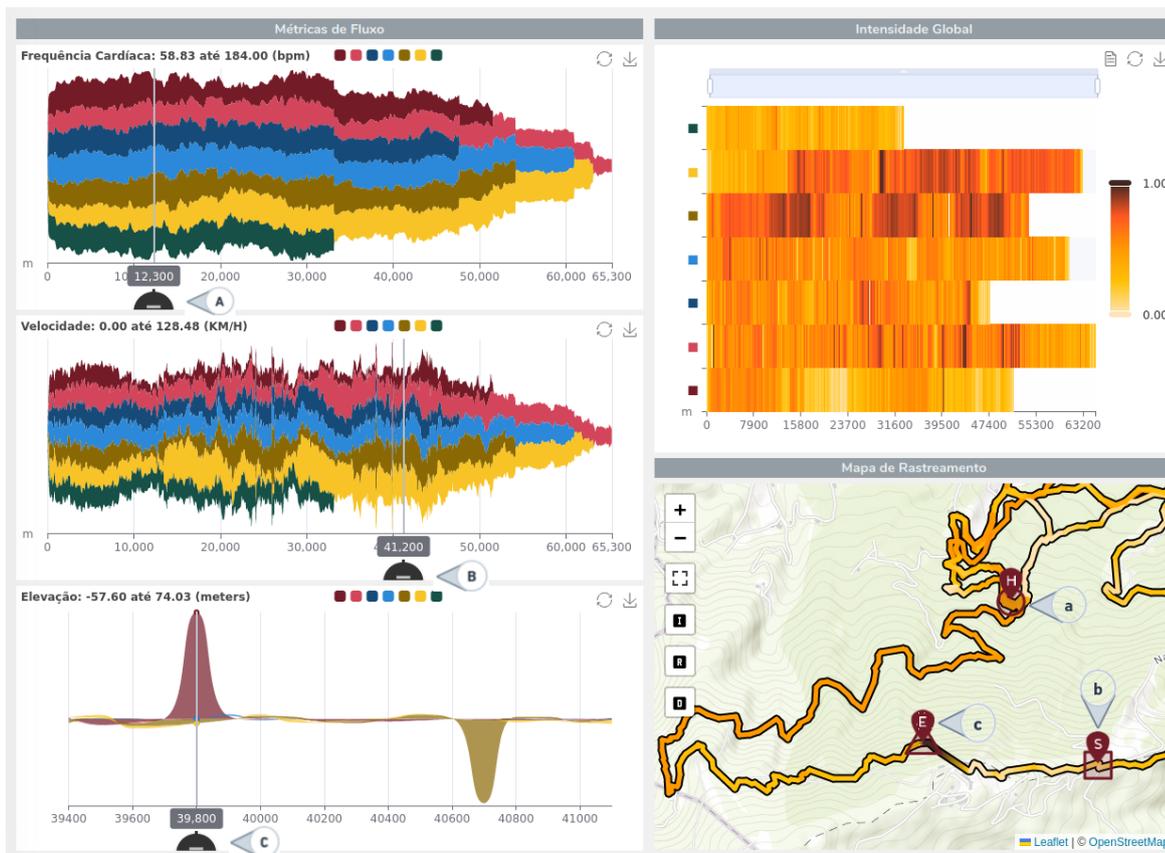


Figura 4.14: Coordenação entre Mapa de Rastreamento e Métricas de Fluxo. Ao deslocar as régulas no Métricas de Fluxo, ícones indicadores das variáveis de elevação, velocidade e frequência cardíaca são plotadas no Mapa de Rastreamento

evoluem durante todo o trajeto de pedalada e como as mudanças em uma variável (por exemplo, velocidade) podem impactar outras variáveis (como a intensidade global e a frequência cardíaca).

Em conclusão, a utilização de visualizações coordenadas por meio de ferramentas de interação fornecerá aos usuários um método eficaz e intuitivo para explorar os dados, possibilitando a identificação de padrões e tendências complexas que seriam desafiadoras de se perceber por meio de uma única visualização. Isso melhora a compreensão geral do desempenho do ciclista e permite uma análise mais precisa das variáveis que afetam esse desempenho. Com base nessas considerações, espera-se que o Objetivo 7 será efetivamente alcançado com a implementação destas coordenações.

4.8 CONSIDERAÇÕES FINAIS

Este capítulo apresentou a proposta de dissertação relacionado a visualização esportiva aplicada ao esporte de ciclismo. Visando propor uma nova abordagem de exploração visual com técnicas de visualização multidimensionais, séries temporais, grafos e geo-

visualização, para correlacionar atributos de frequência cardíaca, velocidade, elevação, temperatura, duração, intensidade e do percurso do terreno.

Para cumprir os objetivos de projeto, foram apresentadas técnicas que envolve a descrição das visualizações propostas. Como contribuição deste trabalho, foi desenvolvida uma nova abordagem de exploração visual multidimensional para exibir o contexto geral dos dados do ciclista e permitir que o usuário desenvolva sua análise por meio da identificação e seleção das informações que lhe sejam relevantes, utilizando ferramentas de interação com a interface, em vez de apenas consumir informações estatísticas e previamente apuradas. Isso permitirá uma análise de desempenho mais apurada em esportes de ciclismo ou outros esportes de resistência como o atletismo.

AVALIAÇÃO**5.1 CONSIDERAÇÕES INICIAIS**

O processo de avaliação da pesquisa desenvolvida é uma etapa crucial na metodologia científica. Os métodos de avaliação utilizados foram o de realização de tarefas analíticas, questionário de usabilidade, e feedback qualitativo apresentados na Seção 2.5. Este capítulo apresenta a avaliação dos resultados obtidos provenientes deste trabalho de pesquisa. Para tal, será apresentado na Seção 5.2 um cenário de uso hipotético onde praticantes de ciclismo desejam analisar e comparar suas pedaladas. Na Seção 5.3 foi realizado um estudo com usuários de modo que eles pudessem avaliar a eficiência e usabilidade da ferramenta. Por fim, a Seção 5.4 apresenta as considerações finais do capítulo.

5.2 CENÁRIO DE USO**5.2.1 Seleção de Ciclistas e Pedaladas**

Seguindo o início do fluxo de trabalho apresentado na Figura 4.5, deve-se filtrar e selecionar pedaladas dos ciclistas específicos. Para isso, a Figura 5.1 apresenta um cenário com as etapas de seleção de ciclistas e suas respectivas pedaladas. Inicialmente um usuário seleciona os ciclistas de 4, 12, 17 e 11, estabelecendo assim um critério de filtragem específico. Em seguida, é definido um intervalo de distância entre 39 e 70 quilômetros, o que permite focar em pedaladas dentro desse intervalo de distância. Depois disso, é necessário clicar no botão “Ver Cycle Lens” então selecionam-se pedaladas específicas para cada ciclista, com base na distância percorrida. Por fim, ao fechar a janela do *Table Lens* e clicar em “Atualizar Painéis”, os usuários podem ver os dados renderizados nas visualizações com base nas pedaladas selecionadas. Esse novo painel com as visualizações justapostas permite ao usuário analisar o desempenho do ciclista de uma perspectiva mais ampla, considerando as pedaladas selecionadas.

5.2.2 Avaliando as Distâncias Percorridas

No *dashboard* que apresenta as visualizações das pedaladas dos ciclistas selecionados, o usuário pode identificar as pedaladas no gráfico de distâncias de maior ou menor valor.

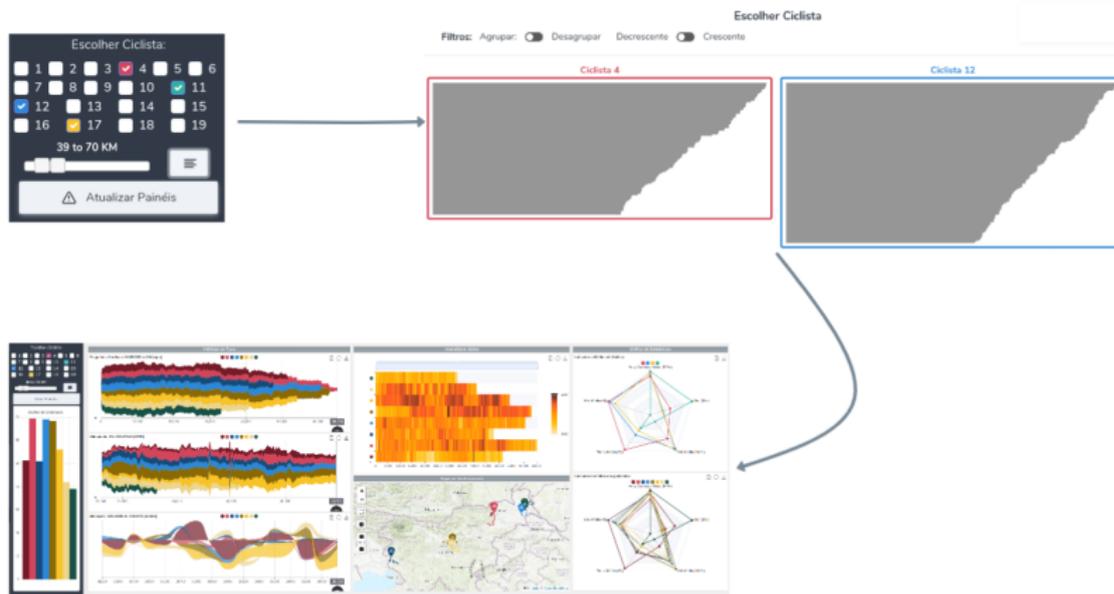


Figura 5.1: Cenário de uso para seleção de ciclistas e pedaladas.

Por exemplo, é possível identificar que a pedalada verde corresponde ao ciclista 11 e possui a menor distância. Essa identificação pode ser feita preferencialmente pelo Gráfico de Distâncias (Figura 4.9), ou pelo eixo horizontal do Métricas de Fluxo (Figura 5.2) ou da Intensidade Global (Figura 5.3).

5.2.3 Examinando a Frequência Cardíaca, Velocidade e Elevação

Analisando o painel visual de Métricas de Fluxo (Figura 5.2) o usuário identifica que a velocidade global/média da pedalada de cor vermelho-escuro é menor do que as outras pedaladas. Em seguida, eles selecionam apenas a pedalada de cor vermelha escura em cada gráfico que mostra a elevação, velocidade e frequência cardíaca.

Depois disso, ajusta-se a régua nos gráficos de elevação, velocidade e frequência cardíaca para a distância de 39800 metros. Isso permite verificar com mais detalhes como essas métricas variam em um ponto específico do percurso, sendo possível identificar que quando a elevação sobe em 39.800 metros, a velocidade até chegar em 39.800 metros diminui, evidenciando a correlação da elevação e velocidade. Essa sequência de interações pode demonstrar que os usuários estão extraíndo informações significativas sobre o desempenho do ciclista ao longo de um percurso específico, entendendo como as métricas de desempenho variam ao longo do percurso e o que essas variações podem indicar sobre o desempenho do ciclista.

5.2.4 Examinando a Intensidade Global

Na próxima etapa, o usuário analisa o painel visual de Intensidade Global (Figura 5.3) identificando que o ciclista 17 (cor amarela) tem pedaladas com maiores intensidades

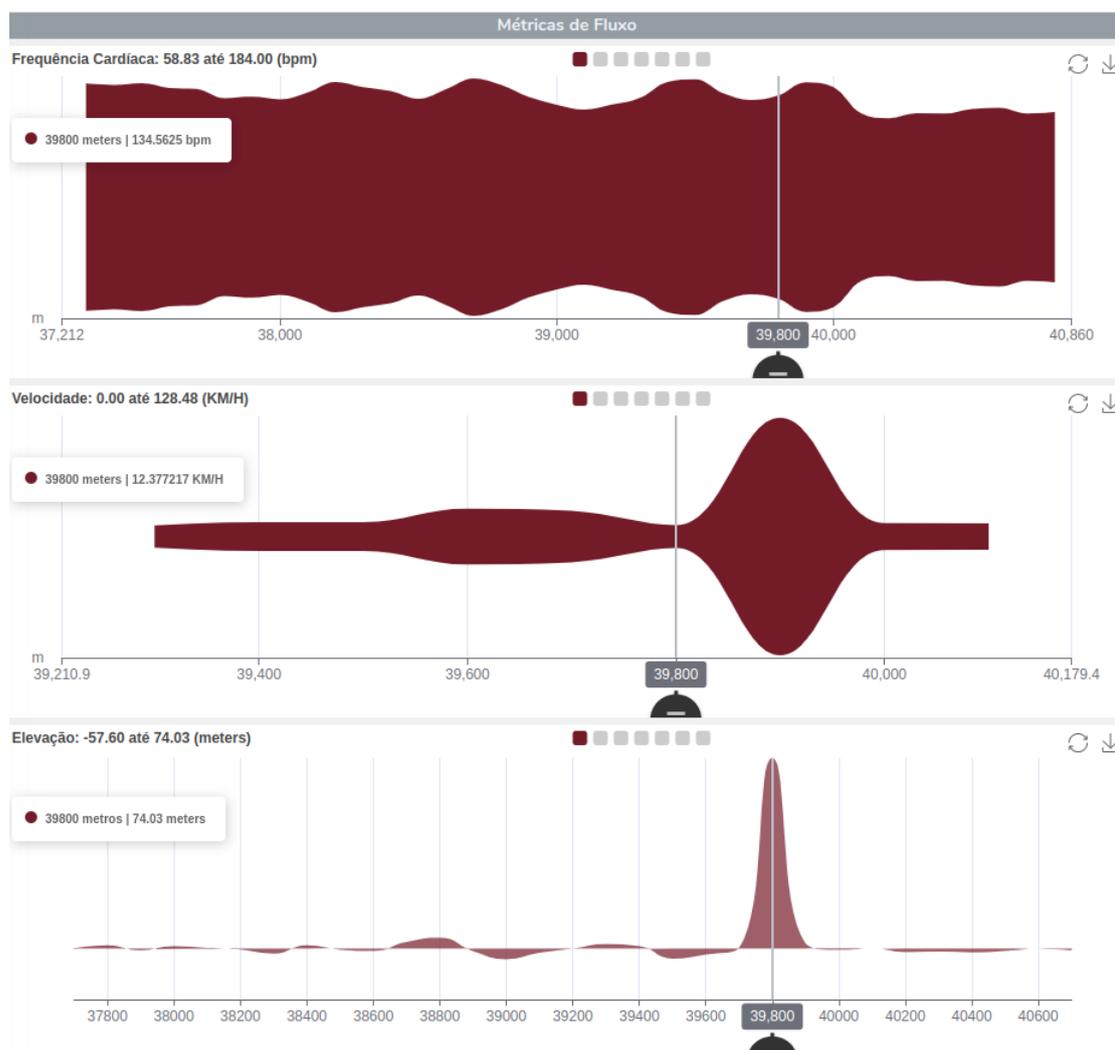


Figura 5.2: Examinando o comportamento da frequência cardíaca, velocidade e elevação no ponto 39800 metros.

globais do que os outros ciclistas. Em seguida, selecionam um intervalo específico no datazoom, o que permite que eles se concentrem em um segmento específico do percurso e analisem a intensidade do esforço dos ciclistas nesse segmento. Eles identificam que o ciclista 17 possui pelo menos uma pedalada que consegue chegar no nível máximo de intensidade global. Ao realizar esta tarefa é analisado a intensidade do esforço do ciclista nas pedaladas realizadas, identificando padrões na intensidade nesse esforço, o que pode ajudar a entender melhor o desempenho do ciclista.

5.2.5 Compreendendo os Indicadores Médios

Quando analisado os Gráficos Estatísticos (Figura 5.4), o usuário identifica que a maior elevação média e maior temperatura é do ciclista 4 (cor vermelha). Em seguida, mo-

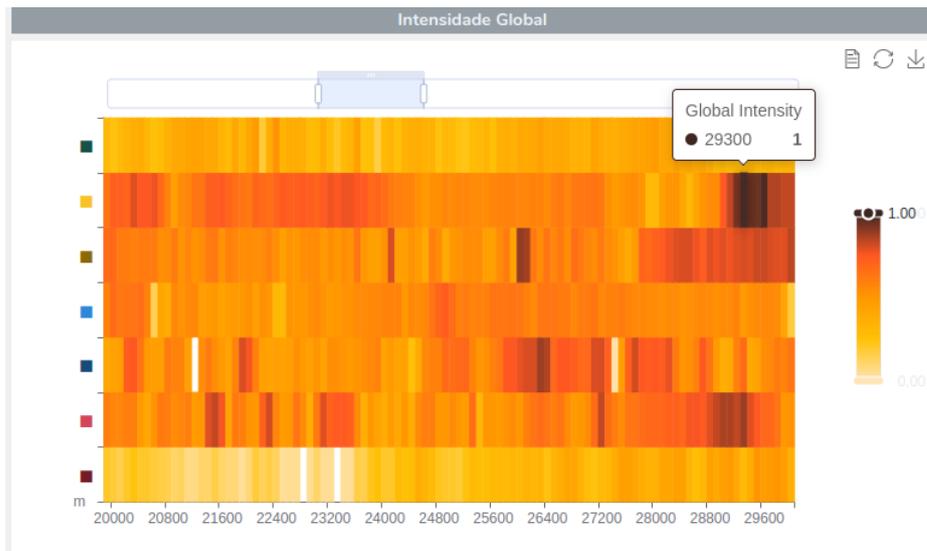


Figura 5.3: Examinando o nível de intensidade entre o trecho 20000 e 30000 metros.

vendo o *mouse* sobre o polígono azul, eles identificam que o ciclista 12 (cor azul) possui pedalada com temperatura média abaixo de 9 graus Celsius. Ao realizar esta tarefa eles estão interpretando e comparando os indicadores médios de cada ciclista e das pedaladas selecionadas, identificando diferenças e semelhanças. Isso permite que eles identifiquem áreas e atributos de força ou fraqueza com os respectivos padrões de desempenho.

5.2.6 Interpretando as Informações Geoespaciais

Ao utilizar o painel visual de Mapa de Rastreamento (Figura 5.4), seleciona-se a camada de distância. O usuário identifica que, em relação à pedalada azul-claro, a pedalada vermelha clara está mais próxima à vermelha escura. Em seguida, analisa a camada de rotas e identifica que nem todas as pedaladas estão no formato de circuito, ou seja, o ponto de início e final não é necessariamente o mesmo lugar. Por fim, é analisada a camada de intensidade e identifica que a pedalada amarelo-escuro tem alta intensidade na região de Spodnji Polsnik. Desse modo, o usuário está explorando o mapa e identificando a localização das pedaladas selecionadas, além de compreender melhor o desempenho do ciclista ao mapear a intensidade global no trajeto das pedaladas.

5.2.7 Trabalhando com visualizações coordenadas

A coordenação entre visualizações atua como um filtro de dados para outras visualizações, mecanismo também conhecido como visualizações vinculadas. Além dos exemplos discutidos antes, podemos detalhar as seguintes interações de coordenação:

Métricas de Fluxo x Intensidade Global: o usuário utiliza o *datazoom* no painel visual da Intensidade Global (Figura 5.5), para selecionar o intervalo de distância entre 20000 metros e 32000 metros aproximadamente. Isso possibilita focar em um segmento especí-

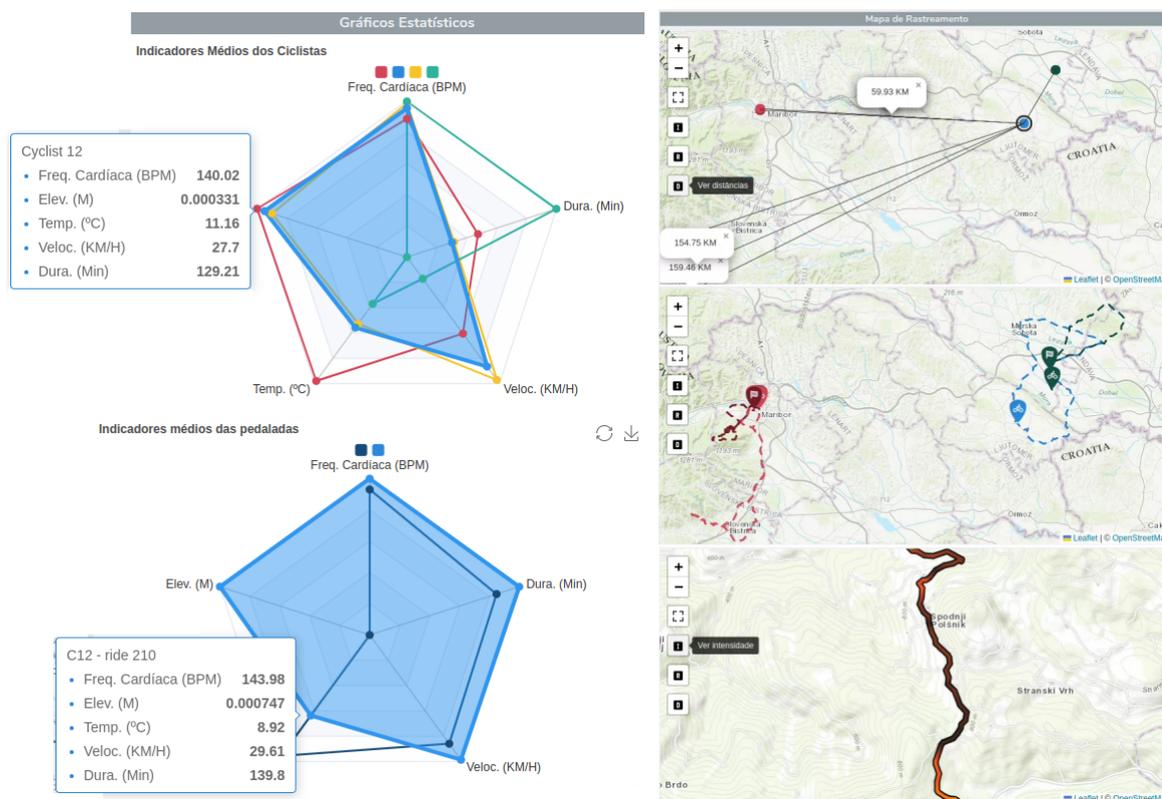


Figura 5.4: À esquerda, o gráfico estatístico que possibilita uma compreensão dos indicadores médios dos ciclistas e das pedaladas. À direita o mapa de rastreamento apresentando as informações geoespaciais.

fico das pedaladas e entender como as variáveis de desempenho variam nesse intervalo, automaticamente o painel Métricas de Fluxo ajusta-se no intervalo definido. Em seguida, é analisado o painel visual de Intensidade Global e o painel visual de Métricas de Fluxo concomitantemente, permitindo identificar a ocorrência de intensidade global alta na pedalada amarelo-claro e aumento da velocidade na mesma pedalada para esse intervalo de distância.

Intensidade Global x Mapa de Rastreamento: posteriormente, observa-se no Mapa de Rastreamento como a localização geográfica influenciou especificamente as variáveis de frequência cardíaca, velocidade e elevação. Ao realizar estas interações utiliza-se a coordenação entre diferentes visualizações a fim de aprimorar a compreensão das relações entre as variáveis que influenciam o desempenho do ciclista. O objetivo é compreender como a intensidade global e a velocidade se relacionam em um segmento específico das pedaladas, o que pode auxiliar na melhor interpretação do desempenho do ciclista nesse segmento.

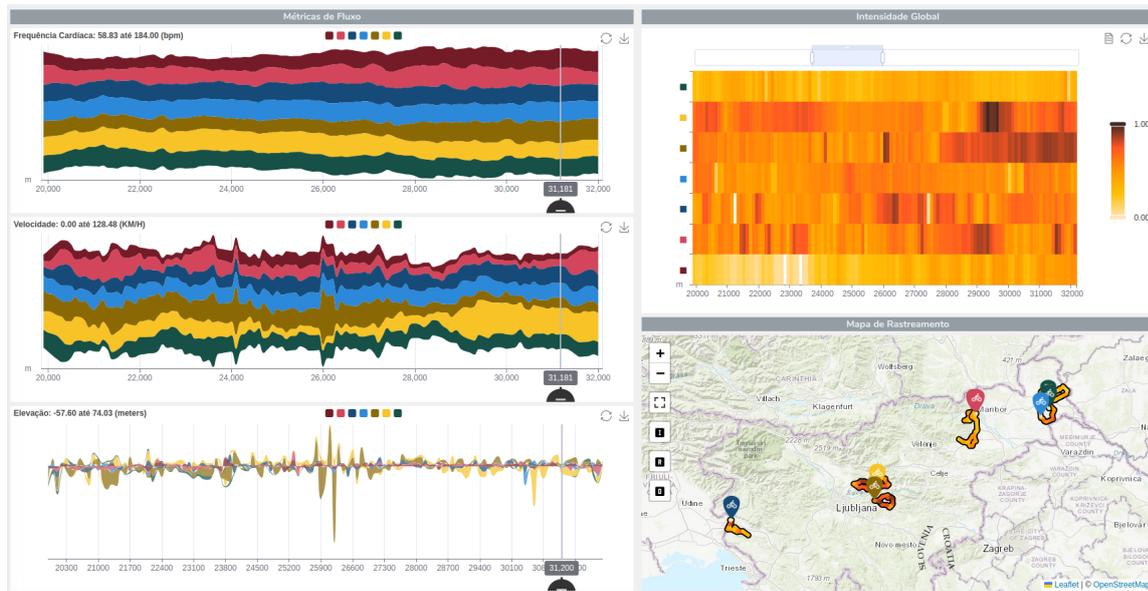


Figura 5.5: A esquerda, o métricas de fluxo indicando as variações de frequência cardíaca, velocidade e elevação no intervalo entre 20000 e 32000 metros. A direita o mapa de rastreamento que visualiza a correlação entre localização geográfica e as métricas de fluxo

5.3 AVALIAÇÃO COM USUÁRIOS

Algumas avaliações com usuários ocorreram presencialmente e outras remotamente, utilizando vídeo chamadas *online* via plataforma *Google Meet*¹. As vantagens abarcam a economia de tempo e de custo de deslocamento para realização do procedimento, possibilitando o recrutamento de uma quantidade maior de usuários, uma vez que não é necessária a presença física.

5.3.1 Metodologia e Instruções

A Figura 5.6 ilustra o processo de avaliação, dividido em cinco etapas distintas. A etapa inicial envolveu a introdução do contexto da ferramenta, além da explicação de seus objetivos e a condução de uma sessão de treinamento que durou aproximadamente de 10 a 15 minutos. Esta etapa possuía o objetivo de familiarizar os usuários com o projeto e a ferramenta e prepará-los para as etapas subsequentes. A segunda etapa envolveu a aplicação de um questionário demográfico, cujo propósito era mapear o perfil dos usuários. A terceira etapa consistiu na realização das sete tarefas analíticas, ou seja, a responder às questões relacionadas às tarefas da Seção 4.5. Para responder cada uma das tarefas, o usuário interagia com a ferramenta e respondia à tarefa seguida de uma pergunta relacionada a complexidade daquela tarefa (escala Likert de 5 níveis, indo de

¹Google Meet - Videochamadas, reuniões e conferências *online* em tempo real. Disponível em: <https://meet.google.com>. Acessado em: 20/07/2023

complexidade muito fácil ao muito difícil). A quarta etapa envolveu a aplicação de um questionário de usabilidade, cujo objetivo era avaliar quantitativamente a usabilidade do usuário com a ferramenta. A quinta e última etapa foi a coleta de feedback qualitativo dos usuários, que visou proporcionar uma oportunidade para os usuários expressarem suas opiniões e experiências abertamente, possibilitando a obtenção de percepções extras que podem ser usadas para melhorar o CycleVis.

O documento original contendo as tarefas e instruções conforme mostrado aos participantes está disponível no Apêndice 1.



Figura 5.6: Fluxograma das etapas de avaliação com usuários.

5.3.2 Demografia

A Figura 5.7 compila dados relativos à distribuição etária, grau de formação educacional e gênero dos participantes que participaram da pesquisa em questão. Compuseram o conjunto de participantes, doze indivíduos, sendo oito homens e quatro mulheres. A faixa etária dos participantes situa-se no intervalo de 21 a 50 anos. Com maior quantidade na faixa de 36 a 40 anos. Quanto à formação, observa-se que dois dos participantes possuem ensino médio completo, enquanto o restante possui formação ao nível superior. Dentre os participantes que possuem superior completo, há 4 pessoas com especialização, 2 com mestrado e 1 com doutorado. Importante destacar que não houve, entre os participantes, familiaridade prévia com o conjunto de dados utilizado no estudo.

A Figura 5.8 apresenta informações sobre a familiaridade com as técnicas computacionais, familiaridade em relação à prática de ciclismo e a frequência com esse esporte. No tocante a experiência com visualização de dados, três deles não possuíam nenhuma experiência anterior, dois apresentaram um grau de familiaridade considerado baixo, outros três demonstraram um nível médio de experiência e quatro dos participantes tinham um alto nível de experiência com visualização de dados.

Em relação à experiência dos participantes com ciclismo, três deles não possuíam experiência anterior com a prática, enquanto nove deles confirmaram ter algum grau de experiência nessa atividade. Dentre os que praticavam, dois deles praticam essa atividade diariamente, dois praticam de 2 a 3 vezes por semana, apenas um realiza uma vez por semana e quatro participantes pedalam algumas vezes por mês. Foi registrado que nenhum dos participantes pratica ciclismo de 1 a 5 vezes por semestre. Por fim, no que se refere à distância percorrida pelos participantes durante suas práticas de ciclismo, não foi registrada nenhuma distância inferior a 5km ou superior a 50km. No entanto, dois participantes costumam pedalar entre 5km e 15km, quatro costumam pedalar entre 15km

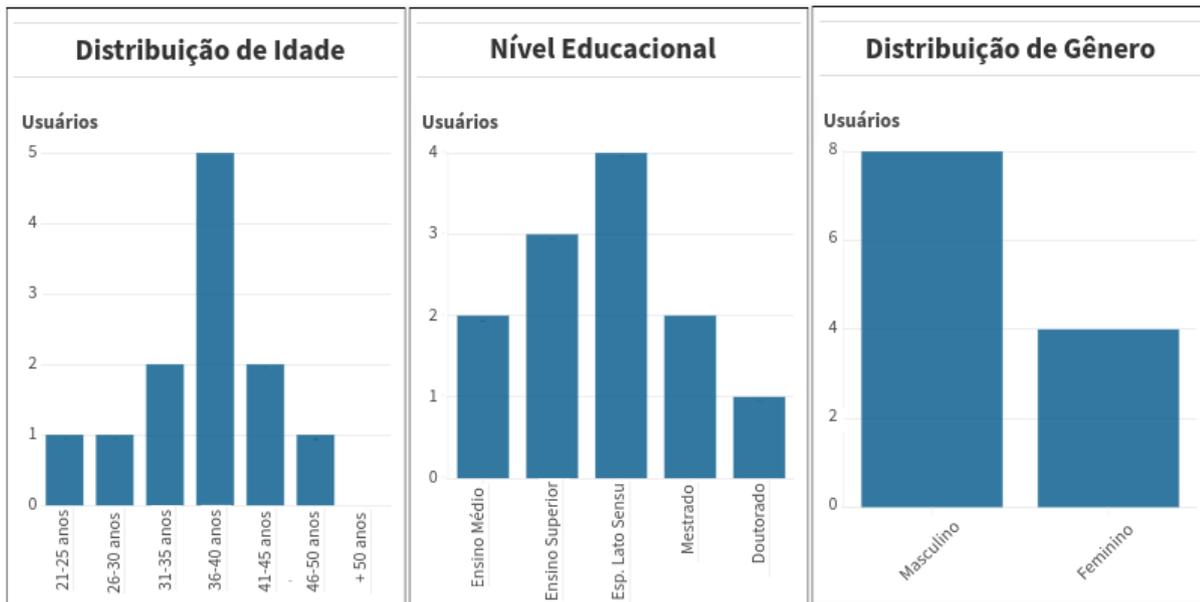


Figura 5.7: Informações sobre idade, gênero e nível educacional dos participantes.

e 30km e três costumam pedalar entre 30km e 50km.

5.3.3 Resultados das Tarefas Analíticas

A Tabela 5.1 apresenta lado-a-lado os objetivos e suas respectivas tarefas analíticas definidos na Seção 4.5. Cada tarefa possui uma pergunta com 4 alternativas, sendo somente uma alternativa correta. Além disso, cada questão possui uma opção para selecionar um dentro cinco níveis de complexidade (muito fácil, fácil, moderado, difícil e muito difícil). Todas as questões relacionadas às tarefas estão relatadas no Apêndice 1.

A Figura 5.9 apresenta o tempo médio de conclusão da avaliação por todos os usuários (cor verde). É possível notar também a média de cada grupo em separado, os grupos de ciclistas (cor amarelo) e visualizadores (especialistas em visualização - cor azul). Dentre as principais informações obtidas, destacam-se:

- **Variação entre Grupos de Usuários:** Nota-se que, em geral, os ciclistas tendem a levar mais tempo para completar as tarefas em comparação com os especialistas em visualização, ao serem mais rápidos somente em 3 das 11 tarefas. Isso pode indicar que os ciclistas demoram por não terem familiaridade com os gráficos e interações durante o processo de exploração visual dos dados.
- **Tarefas mais demoradas:** A tarefa T 03 apresenta o maior tempo médio de realização para ambos os grupos, com uma média geral de 8 minutos e 18 segundos. Isso indica que a tarefa é mais complexa visto que é necessário interagir com 3 gráficos, além de um nível de precisão maior na interação. As alternativas alocadas como respostas podem exigir um nível maior de atenção por parte do usuário também.

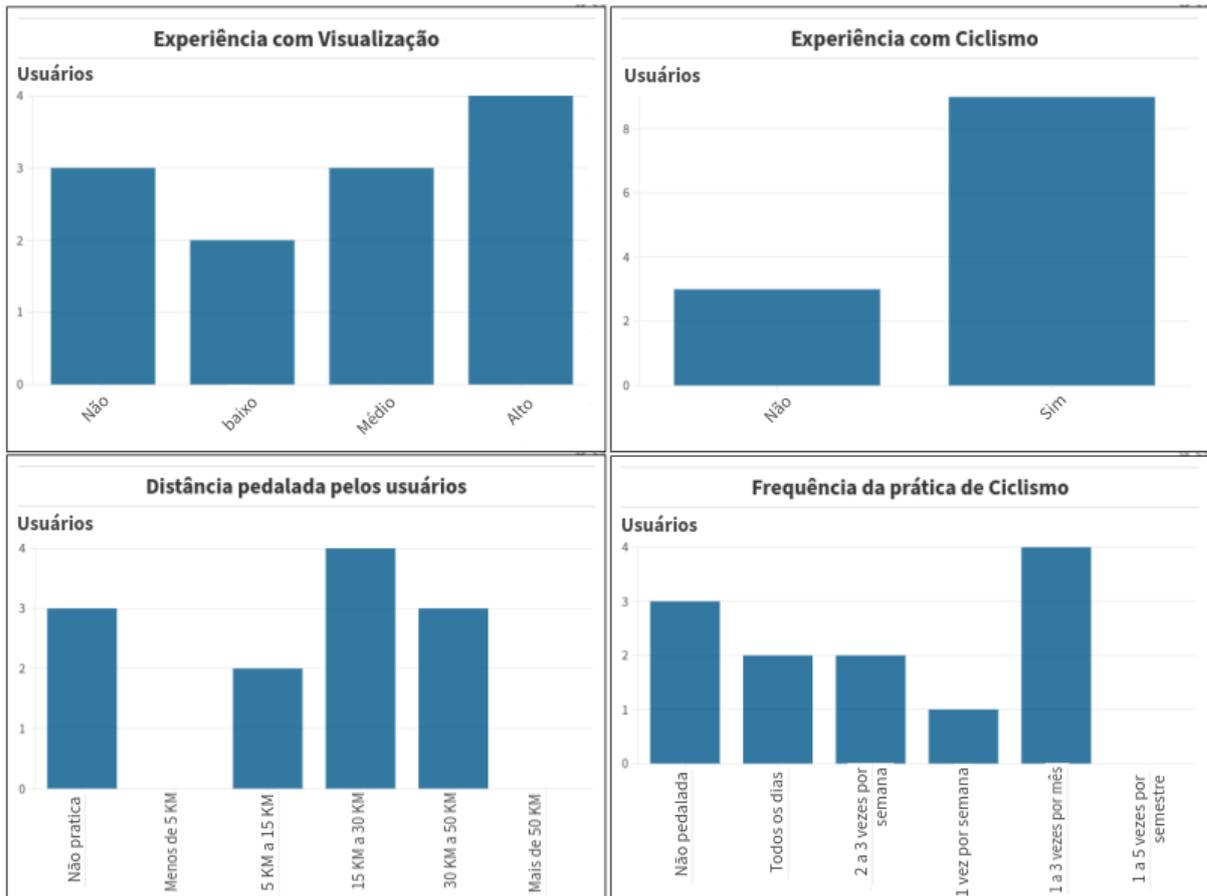


Figura 5.8: Informações sobre experiência prévia dos participantes.

- **Tarefas mais rápidas:** A tarefa T 05.1 obteve a média geral mais rápida, enquanto a T 02 foi a mais rápida dentre os visualizadores e a T 04.1 dentre os ciclistas. Isso denota que o gráfico estatístico foi de fácil interpretação para ambos grupos. Analisando em separado, o gráfico de barras foi mais intuitivo para os visualizadores e o de intensidade global para os ciclistas, sendo uma possibilidade para esse último ser mais intuitivo por mapear a informação de intensidade na cor.
- **Consistência entre Grupos:** Algumas tarefas, como T 05.1 e T 06.1, mostram tempos médios iguais e de curta duração entre visualizadores e ciclistas, o que pode indicar que as visualizações associadas são bem projetadas para serem universalmente compreensível para todos os usuários.
- **Maiores diferenças:** As tarefa iniciais T 01, T 02 e T 03 possuem uma diferença significativa entre diferente grupos de usuários, com maior tempo para os ciclistas. Uma possibilidade é o tempo de adequação para o grupo entender como funciona as interações de seleção e filtro na aplicação, ou seja, ocorreu uma duração de tempo maior para ciclistas se familiarizarem com o processo de exploração visual

Objetivo	Tarefa
O1: Selecionar e visualização de ciclistas e pedaladas	T1: Selecione os ciclistas 4, 12, 17 e 11. Em seguida, estabeleça o intervalo entre 39 e 70 quilômetros (KM). Depois, clique no botão “Ver Table Lens” para ver detalhes sobre os ciclistas selecionados
O2: Identificar os diferentes padrões visuais para apresentar as distâncias nas pedaladas	T2: Identifique as pedaladas de maior ou menor distância e responda com a afirmação verdadeira
O3: Permitir a descoberta de padrões visuais para a compreensão do desempenho das pedaladas e ciclistas	T3: No painel visual métricas de fluxo selecione apenas a pedalada de cor vermelha escura em cada gráfico que mostra a elevação, velocidade e frequência cardíaca. Em seguida, use a régua e ajuste nos gráficos de elevação, velocidade e frequência cardíaca para a distância de 39800 metros. Qual a afirmação verdadeira?
O4: Explorar a ideia e os gráficos da intensidade global como métrica que contempla todas as métricas de intensidade	T4.1: No painel visual de Intensidade Global, assinale a alternativa correta
	T4.2: No painel visual de Intensidade Global, selecione o intervalo de 20.000 metros e 30.000 metros no Datazoom e assinale a alternativa correta
O5: Descobrir se os usuários conseguem interpretar e comparar os indicadores médios de cada ciclista e das pedaladas selecionadas	T5.1: Analise o painel visual de Gráficos Estatísticos, em específico os Indicadores Médios dos Ciclistas, e responda à alternativa correta
	T5.2: Analise o painel visual de Gráficos Estatísticos, em específico os indicadores médios das pedaladas, e responda à alternativa correta
O6: Avaliar a identificação de padrões georreferenciados nas pedaladas selecionadas	T6.1: Analise o painel visual de Mapa de Rastreamento, selecione a camada de distância e responda com a afirmativa correta
	T6.2: Analise o painel visual de Mapa de Rastreamento, selecione a camada de rotas e responda com a afirmativa incorreta
	T6.3: Analise o painel visual de Mapa de Rastreamento, selecione a camada de intensidade e responda com a afirmativa correta
O7: Descobrir diferentes informações por intermédio das coordenações entre as visualizações	T7.1: Selecione no DataZoom no painel visual da Intensidade Global o intervalo de distância entre 20000 e 32000 aproximadamente. Analise o painel visual de Intensidade Global e o painel visual de Métricas de Fluxo e assinale a alternativa correta

Tabela 5.1: Listagem dos objetivos e suas respectivas tarefas.

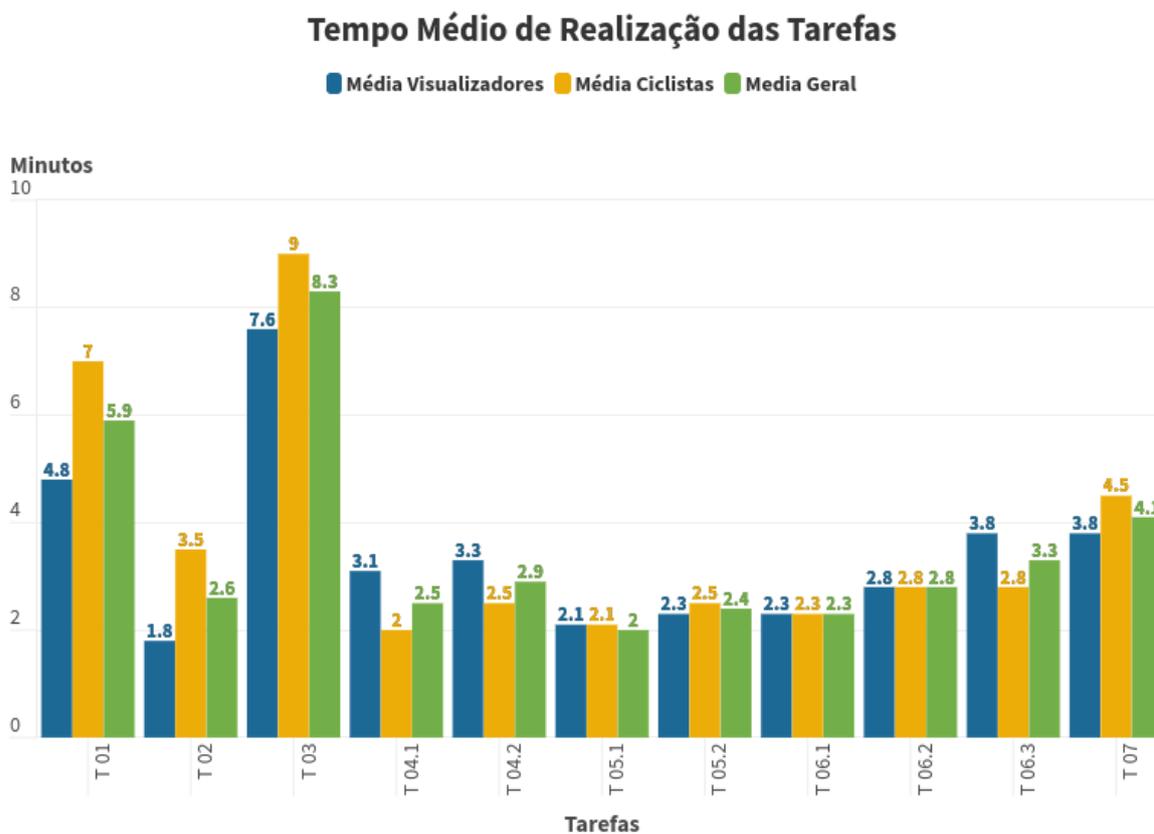


Figura 5.9: Tempo médio em minutos de realização das tarefas dos usuários. Média geral em verde, especialistas em visualização em azul, ciclistas em amarelo.

interativa.

Em relação aos tempos individuais de cada participante, a Figura 5.10 apresenta a duração em minutos para a realização de cada tarefa para cada usuário (sendo C_i o usuário i do grupo de ciclista, e V_i o usuário i do grupo de visualizadores). É possível observar que o usuário que levou mais tempo para realizar todas as tarefas foi o visualizador V5, ao passo que o de menor duração foi o visualizador V4. Além das tarefas T 01 e T 03 que mais tomou tempo da maioria dos ciclistas, a tarefa T 07 tomou um tempo maior para os usuários C6 e V5. Vale ressaltar que o único usuário que não demorou nas tarefas T 01 e T 03 foi o V4.

Os dados referentes a complexidade das tarefas estão presentes na Tabela 5.2. Adotou-se o intervalo de confiança médio com 95% de confiança (ICM 95%). Essa medida fornece uma estimativa do intervalo dentro do qual a verdadeira média da população provavelmente se encontra. Quanto mais estreito o intervalo, mais precisas são as estimativas. A margem de erro (ME) é uma medida da incerteza em torno da estimativa da média. Quanto menor a margem de erro, mais confiável é a estimativa. O desvio padrão (DP) é uma medida da variabilidade na estimativa da média. Quanto menor o erro padrão, mais

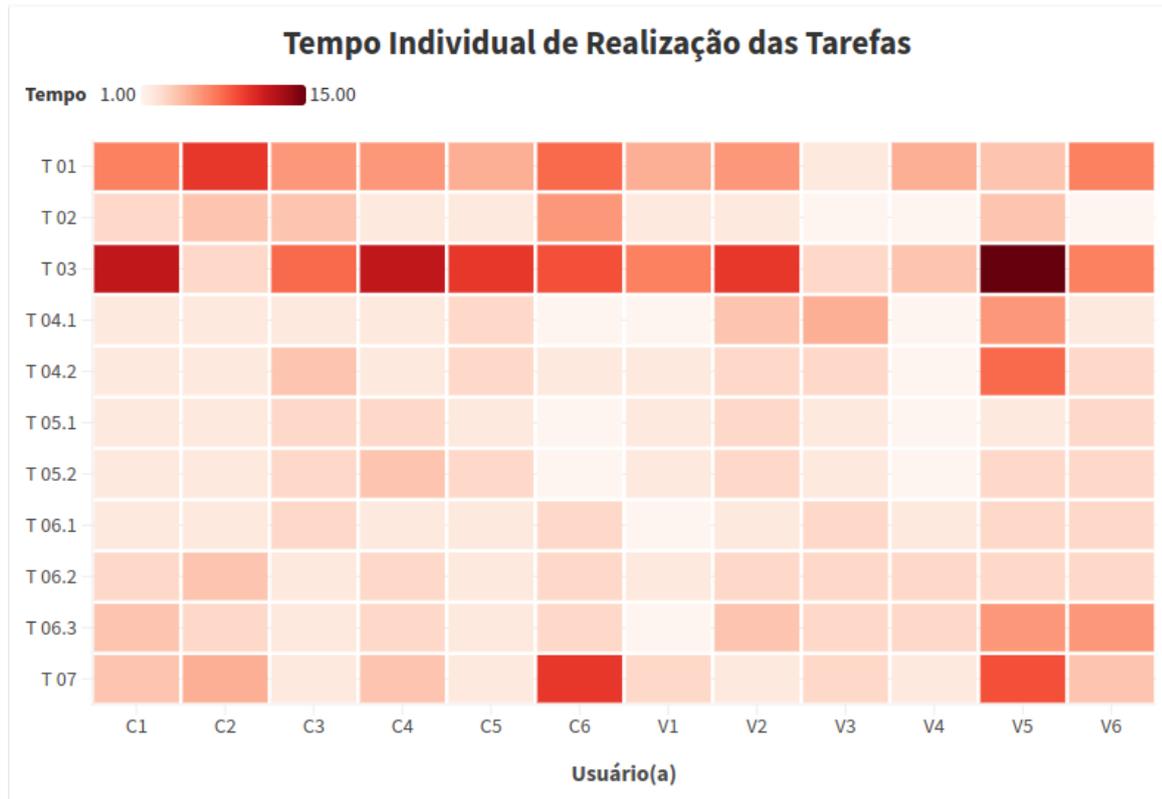


Figura 5.10: Tempo individual em minutos de realização das tarefas dos usuários.

confiável é a estimativa.

Assim, a Tabela 5.2 agrupa a complexidade das tarefas definidas na Seção 4.5, com níveis de complexidade variando de 1 a 5 para cada tarefa, sendo 1 muito baixa e 5 muito alta. As linhas de C1 até C6 representa o grupo de ciclistas, as linhas de V1 até V6 representa o grupo de visualizadores. Cada coluna representa os resultados de cada tarefa. MU representa a média do usuário em todas as tarefas. MC representa a média do grupo de ciclistas. MV representa a média do grupo de visualizadores. MG representa a média geral. PA representa o percentual de acertos médio na tarefa. PE representa o percentual de erros na tarefa.

Tarefa 1: A média de complexidade (MC) para os ciclistas foi maior do que os visualizadores, com médias de 2,5 e 2,3 respectivamente, denotando um nível entre fácil e intermediário de complexidade. Ambos grupos tiveram 1 usuário que classificou a complexidade dessa tarefa como alta (4). Houve sugestões de usuários sobre a possibilidade de exibir os valores de distâncias no *Table Lens* sem a necessidade de interação manual através do *mouse*, isso pode tornar mais fácil para os usuários selecionar pedaladas específicas reduzindo a complexidade percebida da tarefa. Além disso, o percentual de acerto (PA) para os usuários na tarefa T1 foi de 100%, indicando que estes conseguiram completar a tarefa T1 com sucesso, e, portanto, o percentual de erro (PE) foi de 0%. O grupo

Usuários	T1	T2	T3	T4.1	T4.2	T5.1	T5.2	T6.1	T6.2	T6.3	T7	MU
C1	1	1	3	1	1	1	2	2	1	1	1	1,3
C2	4	3	3	1	2	1	2	1	1	1	3	1,8
C3	2	1	4	2	2	2	2	5	3	3	5	2,5
C4	3	3	4	2	4	4	5	4	2	4	3	3,1
C5	2	1	3	1	1	1	3	3	1	2	2	1,6
C6	3	1	5	1	1	3	1	3	2	1	4	2,8
V1	1	1	3	1	3	1	2	2	1	1	2	1,5
V2	2	1	4	2	1	2	3	4	3	5	3	2,5
V3	1	1	4	3	3	2	2	4	2	3	3	2,3
V4	3	1	2	1	1	1	1	3	2	2	2	1,5
V5	3	1	5	3	3	2	4	4	3	4	4	3
V6	4	1	3	1	1	1	1	2	1	3	3	1,7
ICM	1,73	0,84	3,01	1,08	1,23	1,14	1,55	2,34	1,62	1,62	2,23	1,73
95%	3,11	1,83	4,16	2,09	2,61	2,36	3,12	3,82	2,36	3,38	3,61	2,54
DP	1,08	0,78	0,9	0,79	1,08	0,97	0,36	1,16	0,83	1,38	1,08	0,64
ME	±0,7	±0,5	±0,5	±0,5	±0,7	±0,6	±0,8	±0,7	±0,5	±0,9	±0,7	±0,4
	T1	T2	T3	T4.1	T4.2	T5.1	T5.2	T6.1	T6.2	T6.3	T7	
MC	2,5	1,6	3,6	1,3	1,8	2	2,5	3	1,6	2	3	2,2
MV	2,3	1	3,5	1,8	2	1,5	2,1	3,1	2	3	2,8	2,2
MG	2,4	1,3	3,55	1,5	1,9	1,7	2,3	3,05	1,8	2,5	2,9	2,2
PA	100%	100%	75%	92%	92%	100%	92%	100%	75%	92%	92%	91%
PE	0%	0%	25%	8%	8%	0%	8%	0%	25%	8%	8%	9%

Escala de Complexidade:

1 2 3 4 5
 +Fácil +Complexo

Tabela 5.2: Resultados da complexidade das questões.

de visualizadores completou a tarefa 2 minutos e 12 segundos mais rápido do que o grupo de ciclistas, o que pode indicar que a experiência ou familiaridade com a visualização de dados pode afetar um papel na eficiência do uso da ferramenta.

Tarefa 2: Todos os visualizadores indicaram essa tarefa com pouca complexidade. Para os ciclistas, dois usuários indicaram complexidade média. A média geral de complexidade da Tarefa 2 foi a menor dentre todas as tarefas realizadas, com 1,3 de média, colocando a complexidade dessa tarefa mais próxima do nível de dificuldade muito fácil. Isso sugere que o gráfico de distâncias foi eficaz em comunicar informações sobre as distâncias das pedaladas. O percentual de acerto (PA) para os usuários foi novamente de 100%, indicando que todos também conseguiram concluir a tarefa com sucesso. Um comportamento interessante observado foi que 75% dos usuários utilizaram o Gráfico de Distâncias para compreender as distâncias das pedaladas, enquanto um quarto dos usuários também utilizaram as Métricas de fluxo (25%), indicando que diferentes visualizações podem ser úteis para diferentes usuários. Os visualizadores completaram a tarefa mais rapidamente do que os ciclistas, com um tempo médio de conclusão de 1 minuto e 48 segundos, enquanto os ciclistas levaram em média 3 minutos e 30 segundos para completar a tarefa.

Tarefa 3: Esta tarefa foi considerada a mais complexa de todas para todos os usuários se considerarmos a maior média geral de 3,55. Foi a única que dois usuários de grupos distintos classificaram como muito complexa (nível 5) e 2 usuários de grupos distintos classificaram como complexa. Somente um visualizador classificou como fácil, com o restante atribuindo nível médio de complexidade. O tempo médio geral de conclusão da tarefa foi o maior de todas, com 8 minutos e 18 segundos. Esse fator acaba refletindo na complexidade da tarefa e vice-versa. Tal resultado pode ser decorrente da necessidade de interpretar várias métricas de desempenho ao longo do circuito. O percentual de acerto (PA) foi de 75%, indicando que a maioria dos usuários conseguiu completar a tarefa com sucesso, embora um quarto deles tenha encontrado alguma dificuldade, conforme indicado pelo percentual de erro (PE) de 25%.

Tarefa 4: Esta tarefa foi dividida em duas subtarefas (T 4.1 e T 4.2). Para ambas, a média de complexidade foi do nível muito fácil para fácil, tendo uma média menor para a T 4.1. A tarefa T 4.2 obteve uma avaliação de nível de complexidade alta de um usuário de ciclismo. O grupo de visualizadores levou mais tempo para completar ambas as atividades em comparação com o grupo de ciclistas. Por um lado, isso pode sugerir que os ciclistas, possuem mais familiaridade com a identificação do desempenho do ciclista por meio de *Heatmap*. Embora a maioria dos usuários tenha conseguido interpretar corretamente a intensidade do esforço do ciclista, uma pequena porcentagem de usuários (8%) interpretou incorretamente as informações.

Tarefa 5: Esta tarefa foi dividida em duas subtarefas (T 05.1 e T 05.2). A Tarefa T 5.1 foi considerada mais fácil que a Tarefa 5.2, tanto para a média geral quanto analisando cada grupo individualmente. A Tarefa T 5.2 foi considerada muito complexa por um ciclista. O tempo médio para conclusão foi muito próximo entre os grupos, com 12 segundos mais rápido para os visualizadores. Os usuários conseguiram interpretar corretamente os indicadores médios de cada ciclista e das pedaladas selecionadas. Isso é evidenciado pelo fato de que 100% dos usuários identificaram corretamente que a maior elevação média e maior temperatura é do ciclista 4, e 92% dos usuários identificaram corretamente que o ciclista

12 possui pedalada com temperatura média abaixo de 9 graus Celsius. Isso sugere que os usuários encontraram a interface do painel visual de Gráficos Estatísticos intuitiva e fácil de usar. No entanto, uma pequena porcentagem de usuários (8%) errou na subtarefa T5.2, sugerindo que pode haver algum aspecto da tarefa ou do treinamento que não estava claro para esses usuários. Isso pode ser uma área para investigação e melhoria futura.

Tarefa 6: Esta tarefa foi dividida em três subtarefas (T 6.1, T 6.2 e T 6.3). A Tarefa 6.1 foi a segunda mais complexa entre todas as 11 analisadas, com média geral de 3,05 (nível intermediário de complexidade). Metade dos visualizadores classificaram como de nível complexo, enquanto 1 ciclista classificou como muito complexo e outro como complexo. Apesar do nível de complexidade mais alto que as demais, todos acertaram a resposta. A Tarefa T 6.2 teve a menor complexidade média de 1,8 (nível fácil). Contudo, teve a menor taxa de acerto, com 75%. Uma possibilidade é a curta duração de tempo para responder à tarefa. A Tarefa 6.3 foi observada como a mais complexa para os visualizadores, com nível intermediário, ao passo que nível fácil para ciclistas. A taxa de acerto foi de 92%, indicando que, mesmo mais complexo que a anterior, teve uma taxa de acerto maior.

Quanto a análise de rotas das pedaladas, a maioria dos usuários (75%) identificou corretamente a afirmativa incorreta, sugerindo que eles entenderam que nem todas as pedaladas ocorrem no formato de circuito. Essa é uma conclusão valiosa que confirma a eficácia da ferramenta em comunicar a natureza das rotas de pedalada aos usuários. Quanto a interpretação da intensidade das pedaladas, a grande maioria dos usuários (92%) conseguiu identificar corretamente que a pedalada amarelo-escuro tem alta intensidade na região de Spodnji Poljsnik. Isso demonstra que os usuários podem interpretar com sucesso os dados de intensidade exibidos pela ferramenta. Vale ressaltar que os ciclistas, mesmo possuindo menor conhecimento abrangente em interpretação visual, foram melhores e 12,68% mais rápidos que os visualizadores na realização desta tarefa. Assim, sugere-se que as tarefas que necessitam de interpretação de informações espaciais podem ser menos desafiadoras para ciclistas, pois isso se correlaciona diretamente com a experiência prática que possuem na modalidade esportiva. Portanto, a experiência pessoal e o conhecimento prévio de cada grupo podem influenciar diretamente na percepção de complexidade das tarefas e na forma como cada um interage e interpreta as informações apresentadas na ferramenta de análise visual. Além disso, o tempo relativamente curto necessário para concluir as atividades indica que a visualização Mapa de Rastreamento pode ser usada eficientemente para análise espacial de dados de ciclismo.

Tarefa 7: Esta tarefa possui um nível muito semelhante de complexidade para todos os grupos, muito próximo da complexidade média. Destaca-se um usuário de ciclismo relatando um nível de complexidade muito alto e uma taxa de acerto relativamente alta, com percentual de erro (PE) de 8%. Isso indica que, embora a atividade fosse desafiadora, a maioria dos usuários sentiu que conseguia lidar com ela. Os visualizadores cumpriram essa tarefa, na média, 24 segundos mais rápido. Esses dados corroboram que utilizar coordenadas em visualizações não é uma tarefa complexa e pode ser interessante em projetos futuros envolvendo esportistas.

Por fim, as médias aritméticas para todas as tarefas considerando *i*) todos os usuários,

ii) o grupo de ciclistas e o *iii)* o grupo de visualizadores foram as mesmas, 2,2 indicando que, de modo geral, as tarefas tiveram uma média de complexidade considerada fácil. Individualmente, somente os usuários C4 e V5 obtiveram a média da complexidade de nível intermediário. A média de Porcentagem de Acerto (PA) ficou acima de 90%, com 91%, enquanto a média da Porcentagem de Erro (PE) ficou em 9%.

5.3.4 Resultados do Questionário de Usabilidade

Os resultados do questionário de usabilidade estão apresentados na Tabela 5.3 e também seguiu o padrão da escala Likert (sendo 1 para “Discordo Totalmente” e 5 para “Concordo Totalmente”) para as respostas.

Res	Q1	Q2	Q3	Q4	Q5	Q6	Q7	Q8	Q9	Q10	Q11
1	0/12	0/12	0/12	0/12	2/12	0/12	0/12	0/12	0/12	0/12	0/12
	0%	0%	0%	0%	17%	0%	0%	0%	0%	0%	0%
2	0/12	0/12	3/12	1/12	0/12	0/12	0/12	0/12	0/12	0/12	0/12
	0%	0%	25%	8%	0%	0%	0%	0%	0%	0%	0%
3	5/12	2/12	1/12	2/12	1/12	0/12	0/12	1/12	2/12	0/12	0/12
	42%	17%	8%	17%	8%	0%	0%	8%	17%	0%	0%
4	3/12	1/12	2/12	3/12	3/12	7/12	3/12	3/12	1/12	2/12	3/12
	25%	8%	17%	25%	25%	58%	25%	25%	8%	17%	25%
5	4/12	9/12	6/12	6/12	6/12	5/12	9/12	8/12	9/12	10/12	9/12
	33%	75%	50%	50%	50%	42%	75%	69%	75%	83%	75%
	100%	100%	100%	100%	100%	100%	100%	100%	100%	100%	100%
NC 95%	3,92	4,58	3,92	4,17	3,92	4,42	4,75	4,58	4,58	4,58	4,75
	+	+	+	+	+	+	+	+	+	+	+
	0,57	0,5	0,83	0,65	0,96	0,33	0,29	0,42	0,5	0,25	0,29

Tabela 5.3: Resultados do questionário de usabilidade.

A seguir, segue uma discussão mais detalhada acerca dos resultados para o questionário de usabilidade.

Questão 1: Você concorda que foi fácil selecionar pedaladas e ciclistas na ferramenta?

A média das respostas foi de 3,92. Isso sugere que os usuários concordaram como relativamente fácil selecionar pedaladas e ciclistas na ferramenta, com 33,3% concordando totalmente e 25% concordando (totalizando 58,3% em concordância). No entanto, 41,7% dos usuários tiveram uma opinião neutra, o que pode sugerir que algumas melhorias podem tornar a tarefa de seleção mais fácil, esse resultado reforça os resultados obtidos na Tarefa 01, que avaliou os aspectos de seleção de pedaladas por meio do *Table Lens*.

Questão 2: Você concorda que foi útil compreender as variáveis de esforço, como elevação e frequência cardíaca e velocidade?

A média das respostas foi de 4,58. Portanto, a maioria dos usuários concordou que compreender as variáveis de esforço, como elevação e frequência cardíaca e velocidade, foi útil. Isto indica que a ferramenta satisfaz em comunicar estas variáveis, as quais são críticas para o desempenho do ciclismo.

Questão 3: Você concorda que foi útil compreender a localização das pedaladas entre diferentes ciclistas?

A média das respostas foi de 3,92. Isso sugere que a ferramenta satisfaz sobre a utilidade de compreender a localização das pedaladas entre diferentes ciclistas, 66,7% dos usuários concordaram (totalizando as categorias 4 e 5). Já alguns (33,3% nas categorias 2 e 3) não concordam que a ferramenta satisfaz. Esse resultado corrobora os resultados da Tarefa 06. Essa tarefa contém a maior quantidade de usuários que discorda de uma determinada questão. Uma possibilidade é a dificuldade nas interações no mapa para localizar as pedaladas.

Questão 4: Você concorda que foi fácil compreender a intensidade das pedais no mapa?

A média das respostas foi de 4,17 reforçando o resultado da Tarefa 04. Quando questionados se era fácil compreender a intensidade das pedaladas no mapa, a maioria dos usuários (75%) concordou ou concordou totalmente. No entanto, 25% dos usuários (categorias 2 e 3) sentiram que a ferramenta poderia ser melhorada neste aspecto.

Questão 5: Você concorda que a ferramenta serviu para comparar a média das variáveis de desempenho entre os ciclistas?

A média das respostas foi de 3,92, ou seja, obteve-se um resultado relativamente positivo. Enquanto 75% dos usuários concordaram ou concordaram totalmente que a ferramenta satisfaz a afirmação, 17% discordaram totalmente, sugerindo que a ferramenta pode precisar de aprimoramentos para tornar essa comparação mais eficaz.

Questão 6: Você concorda que a ferramenta serviu para comparar a média das variáveis de desempenho das pedaladas de um ciclista?

A média das respostas foi de 4,42. Todos os usuários (100%) concordaram ou concordaram totalmente que a ferramenta serviu para comparar a média das variáveis de desempenho das pedaladas de um ciclista. Isto indica que a ferramenta é bem-sucedida nesta tarefa específica.

Questão 7: Você concorda que a ferramenta serviu para comparar a intensidade global das pedaladas?

A média das respostas foi de 4,75. A comparação da intensidade global das pedaladas foi altamente valorizada pelos usuários, com 100% de concordância (categorias 4 e 5). Isto sugere que esta funcionalidade é uma força chave da ferramenta.

Questão 8: Você concorda que foi útil utilizar mais de um painel visual para encontrar a informação desejada?

A média das respostas foi de 4,58. A maioria dos usuários (91,7%) concordou ou concordou totalmente que foi útil utilizar mais de um painel visual para encontrar a informação desejada. Isto sugere que a visualização ligada multifacetada foi bem aceita para a grande maioria dos usuários.

Questão 9: Você acredita que a ferramenta proposta poderia ser utilizada em competições profissionais de ciclismo?

A média das respostas foi de 4,58. A maioria dos usuários (83,3%) concordam totalmente que a ferramenta proposta poderia ser utilizada em competições profissionais de ciclismo. Isto indica que os usuários concordam com a aplicabilidade prática da ferramenta.

Questão 10: Você concorda que a ferramenta pode auxiliar no desenvolvimento de estratégias para a atividade do ciclismo?

A média das respostas foi de 4,58. Todos os usuários (100%) concordaram ou concordaram totalmente que a ferramenta pode auxiliar no desenvolvimento de estratégias para a atividade do ciclismo. Essa questão obteve a maior quantidade de usuários que concordam totalmente. Isso sugere que a ferramenta é vista como uma adição valiosa para a estratégia e planejamento do treinamento de ciclismo.

Questão 11: Você concorda que a ferramenta proposta poderia ser adaptada para análise de outras modalidades esportivas?

A média das respostas foi de 4,75. Todos os usuários (100%) concorda que a ferramenta proposta poderia ser adaptada para análise de outras modalidades esportivas. Isto é uma indicação de que os usuários veem potencial na aplicação desta ferramenta além do ciclismo, abrindo novas possibilidades para futuras adaptações e melhorias.

De modo geral, os usuários concordaram ou concordaram totalmente com os objetivos das funcionalidades da ferramenta. Somente para as questões Q1, Q3 e Q5 obtiveram médias gerais menores, sendo importante observar melhorias nas respectivas funcionalidades. As áreas para potencial melhoria incluem a simplificação da seleção de pedaladas e ciclistas, além de tornar a comparação entre as médias de variáveis de desempenho mais intuitiva. A adaptação da ferramenta para outras modalidades esportivas também é vista como uma oportunidade valiosa.

5.3.5 Resultados do Feedback Qualitativo

Os comentários/feedbacks dos usuários, conforme apresentado na Tabela 5.4, oferece uma visão mais crítica sobre a percepção dos usuários enquanto utilizavam o CycleVis. Os comentários são na maioria positivos, mas também apontam para áreas que requerem melhorias e ajustes.

Usuários	Comentários
C1	Uma ferramenta moderna, de alta precisão, bem útil para o mundo do ciclismo para o competidor quanto para o amador.
C2	Muito útil a ferramenta e gostaria de vê-la sendo usada no futuro.
C3	A ferramenta é útil, porém os painéis todos juntos dificultam a visão.
C4	No gráfico de intensidades as cores poderiam ser diferentes ao invés de um degrade. No gráfico de elevação mesmo sendo selecionado vários poderia ser apresentado apenas um por vez e como já contém a opção de selecionar as pedaladas o mesmo seria utilizado para demonstrar as disparidades entre os competidores.
C5	A ferramenta pode ser útil e aplicada a realidade, pode melhorar nas legendas para que qualquer pessoa possa compreender melhor cada uma das suas funcionalidades/utilidade.
C6	Não houve comentário.
V1	A ferramenta é de fácil utilização e foi muito simples entender também as métricas utilizadas para a visualização. O Mapa de Rastreamento acredito que foi o mais complexo, mas devido a região onde as pedaladas foram realizadas e a visualização na tela de todas as pedaladas juntas.
V2	Informar ao usuário para utilizar o zoom para analisar as atividades da Métricas de fluxo (Tarefa 3). Colocar uma legenda na intensidade global para identificar as pedaladas dos ciclistas. Aumentar o tamanho dos gráficos estatísticos para tela pequena. Colocar legenda para identificar os ciclistas no Mapa de rastreamento.
V3	Oclusão do rótulo de distâncias em alguns momentos.
V4	A ideia é muito boa. Acredito que algumas alterações podem dar uma maior personalização da interface, como por exemplo, a possibilidade de exibir os valores no Table Lens sem a ter que passear com o mouse. Também é interessante por a legenda de cores quando maximiza o mapa para não ter que ficar minimizando para saber qual é a cor de cada ciclista.
V5	Adicionar mouseover para identificar os ciclistas.
V6	Muito útil a ferramenta, apresenta as informações necessárias de forma ágil simples. Talvez terá dificuldades para manusear no celular. Na tela Table Lens exibir o rótulo do eixo. Gostaria de ser usuário da ferramenta.

Tabela 5.4: Tabela com os comentários/feedbacks dos usuários.

No que tange aos pontos positivos, os comentários/feedbacks indicaram:

- **Facilidade de Uso:** Usuários como V1 e V6 destacaram a facilidade de uso da ferramenta e a simplicidade na compreensão das métricas apresentadas.

- **Utilidade e Precisão:** Comentários de C1, C2 e V6 enfatizam a utilidade e a precisão da ferramenta, tanto para competidores quanto para amadores no ciclismo.
- **Aplicabilidade Prática:** Usuários como C2, C5 e V6 veem a ferramenta como algo que pode ser efetivamente aplicado no mundo real, indicando sua relevância prática.

Quanto as melhorias que podem ser realizadas, os feedbacks indicaram:

- **Complexidade Visual:** C3 e V1 apontam para a complexidade visual quando múltiplos painéis ou métricas são exibidos simultaneamente.
- **Navegação e Interface:** V2, V4 e V5 sugerem melhorias na interface do usuário, como a adição de funcionalidades de zoom e *mouseover*, para facilitar a navegação e a interpretação dos dados.
- **Legendas e Rótulos:** C4, C5 e V2 pedem por legendas e rótulos mais claros para auxiliar na interpretação dos gráficos e mapas.
- **Visibilidade e Tamanho:** V2, V3 e V6 mencionam que o tamanho dos gráficos e a visibilidade dos rótulos podem ser um problema, especialmente em dispositivos com telas menores.

Portanto, embora a ferramenta CycleVis seja, em geral bem recebida, existem áreas onde melhorias podem ser implementadas. Estas percepções indicam direções para o desenvolvimento contínuo da ferramenta em pesquisas futuras.

5.4 CONSIDERAÇÕES FINAIS

A avaliação foi composta por um cenário de uso e um estudo com usuários. O cenário de uso demonstrou como um usuário pode extrair informações de dados ciclísticos, na prática. O estudo realizado com usuários envolveu usuários com diferentes níveis de experiência e formação, variando de ensino médio a doutorado, e incluiu uma série de tarefas analíticas projetadas para avaliar diferentes aspectos da ferramenta. Os resultados do estudo sugerem que a ferramenta é geralmente eficaz e intuitiva, embora possa haver espaço para melhorias em termos de usabilidade. Por exemplo, os usuários levaram um tempo relativamente longo para completar uma determinada tarefa, além disso, uma pequena porcentagem de usuários interpretou incorretamente algumas das informações nesta tarefa, o que sugere que a ferramenta pode precisar de melhorias para ser mais eficiente, principalmente nas Métricas de Fluxo.

Considerando as diferenças entre visualizadores e ciclistas, uma abordagem de *design* mais personalizada pode ser benéfica. Para tarefas mais complexas, recursos de treinamento ou documentação adicional podem ser úteis. As análises que incluíram as tarefas, complexidade e usabilidade fornecem uma avaliação quantitativa cruciais para futuras iterações do *design* da ferramenta e melhorias contínuas.

CONCLUSÕES

Este capítulo conclui o presente trabalho apresentando uma visão geral das contribuições. Além disso, são respondidas às questões de pesquisa e pontuadas as limitações e direcionamentos para trabalhos futuros.

6.1 CONTRIBUIÇÕES

O CycleVis é uma abordagem de visualização projetada para explorar visualmente os dados de ciclismo, facilitando e orientando o usuário na etapa de extração de informações não evidentes e potencialmente importantes. Essa dissertação apresentou nove painéis interativos que apresenta visualizações multidimensionais e coordenadas que auxiliam na análise de diferentes tipos de métricas utilizadas na avaliação das práticas do ciclismo.

Durante o desenvolvimento deste trabalho, diversas questões guiaram o processo de pesquisa. A primeira questão de pesquisa está relacionada aos atributos essenciais para compreender o desempenho do ciclista. O trabalho atual não considera apenas os atributos de desempenho comumente relatados na literatura, como velocidade e informação temporal, mas adiciona frequência cardíaca, elevação, localização geoespacial e indicadores de estatística relacionando a média de desempenho de ciclistas e pedaladas. A segunda questão de pesquisa aborda a maneira de mapear os atributos utilizando as técnicas de visualização. Como a característica dos dados são multidimensionais, adotaram-se técnicas de visualização com tais características. Gráficos de área para os atributos de intensidade mapeando múltiplas dimensões como as pedaladas. Indicadores médios para ciclistas e suas pedaladas por meio de gráficos de radar. *Heatmaps* para apresentar um novo atributo global de intensidade e, por fim, mapas contendo todas as pedaladas. Para as visualizações contendo informação temporal, adotou-se a ideia do plano cartesiano com o eixo X para mapear o tempo. Para a terceira questão de pesquisa adotou-se o *layout* mais utilizado na literatura científica e também nos setores comerciais, o de apresentar as visualizações em uma janela com painéis visuais contendo visualizações justapostas, comumente chamada de *dashboards*. Estes painéis mapeiam tanto informações de pedaladas quanto de ciclistas, a depender das interações realizadas pelo usuário. Por fim,

para responder a quarta questão da pesquisa, em relação à avaliação da abordagem, foi desenvolvido um cenário de uso e o mesmo foi aplicado em um estudo com usuários para validar a eficiência, complexidade e usabilidade da ferramenta. Questionários guiaram doze usuários com diferentes perfis durante o uso e aplicabilidade da ferramenta. Outra contribuição foi a aquisição do registro de software de número BR512023002189-0

6.2 LIMITAÇÕES

A primeira limitação da abordagem foi delimitar uma quantidade máxima de ciclistas e pedaladas e, portanto, não tornar uma abordagem escalável nesse sentido. Essa decisão foi consequência das limitações das próprias técnicas de visualização utilizadas, pois inserir mais observações de ciclistas e pedaladas poderia ocasionar uma sobrecarga visual e dificultar o processo de extração de informação.

Em relação às limitações técnicas, a técnica de visualização Cycle Lens requer aprimoramentos para facilitar a identificação das pedaladas. À visualização Métricas de Fluxo, mostrou-se de complexidade média a alta, sugerindo um foco maior no treinamento para usuários com pouca experiência e familiaridade com *Stream Graphs*. A Visualização Intensidade Global e Mapa de Rastreamento poderiam incluir melhorias relacionadas a inclusão de legendas para melhor interpretação dos dados. Outra limitação é a adaptação para telas menores, as visualizações não são adequadas para todos os formatos de tela.

6.3 TRABALHOS FUTUROS

Na literatura, não há trabalhos que relacionem pedaladas em locais distintos para diferentes perfis de ciclistas, tornando essa uma contribuição inédita. Ao passo que utilizar outros atributos podem contribuir para uma análise mais completa, como, por exemplo, atributos climáticos como temperatura, umidade relativa e/ou direção e velocidade do vento. Atributos relacionados à fisiologia do ciclista pode auxiliar a encontrar determinados padrões de desempenho para pedaladas distintas. Peso, altura, frequência cardíaca máxima, VO₂, entre outros.

Os resultados advindos da avaliação com usuários e respectivas tarefas elencando suas complexidades, em conjunto com o questionário de usabilidade, contém informações valiosas para a condução de atividades futuras. Com base no questionário de usabilidade, a maioria dos usuários acredita que a ferramenta tem potencial para ser utilizada em competições profissionais de ciclismo, o que indica haver um interesse em explorar como a CycleVis pode ser adaptada e otimizada para atender às necessidades específicas desse contexto. Isso pode incluir a incorporação de métricas adicionais relevantes para o ciclismo profissional, aprimoramentos na visualização de dados para facilitar a interpretação rápida durante as competições e a integração com outras ferramentas de análise de desempenho utilizadas no ciclismo profissional.

Além disso, todos os usuários concordaram que a CycleVis pode auxiliar no desenvolvimento de estratégias para a atividade do ciclismo. Isso sugere que futuras pesquisas podem se concentrar em como a ferramenta pode ser aprimorada para apoiar o planejamento estratégico de treinamento e competição. Em especial, como a intensidade do

exercício pode ter um impacto significativo na saúde e bem-estar do ciclista, esta capacidade de análise apoia a tomada de decisões, seja no planejamento de treinamentos ou na avaliação do desempenho. Os usuários podem ajustar, inclusive, sua ingestão alimentar conforme a demanda energética, maximizando os benefícios do exercício para a saúde e o condicionamento físico.

A abordagem do *TourVis* em visualização de dados, interatividade e adaptabilidade pode servir de inspiração para o desenvolvimento contínuo do *CycleVis*, especialmente na incorporação de visualizações mais dinâmicas e interativas e na adaptação para diferentes usuários. Finalmente, todos os usuários concordaram que a *CycleVis* poderia ser adaptada para a análise de outras modalidades esportivas. Isso indica haver um potencial significativo para expandir a aplicação da ferramenta além do ciclismo, como corridas e demais esportes que precisam estudar a resistência de atletas de modo mais enfático.

REFERÊNCIAS BIBLIOGRÁFICAS

- ALMEIDA, M. O.; NETO, M. M. Uso de interfaces abundantes em informação para mineração visual de dados. *Relatórios técnicos: RT-NUPERC-2001-5/p*). Bahia: Universidade Salvador, 2001.
- BATTY, M. Rank clocks. *Nature*, v. 444, p. 592–6, 12 2006.
- BECK, F.; BURCH, M.; WEISKOPF, D. A matrix-based visual comparison of time series sports data. In: HULLIN, M.; STAMMINGER, M.; WEINKAUF, T. (Ed.). *Vision, Modeling and Visualization*. [S.l.]: The Eurographics Association, 2016. ISBN 978-3-03868-025-3.
- BYRON, L.; WATTENBERG, M. Stacked graphs – geometry & aesthetics. *IEEE Transactions on Visualization and Computer Graphics*, v. 14, 2008.
- CARD, M.; SHNEIDERMAN, B.; MACKINLAY, J. *Readings in information visualization: using vision to think*. [S.l.]: Morgan Kaufmann, 1999.
- CINTRA, D. S. *Visualização exploratória de grafos bipartidos com o paradigma multinível*. Tese (Doutorado) — Universidade de São Paulo, 2019.
- COX, A.; STASKO, J. Sportsvis: Discovering meaning in sports statistics through information visualization. In: *Compendium of Symposium on Information Visualization*. [S.l.: s.n.], 2006. p. 114–115.
- DU, M.; YUAN, X. A survey of competitive sports data visualization and visual analysis. *Journal of Visualization*, Springer, v. 24, n. 1, p. 47–67, 2021.
- DWYER, T.; MARRIOTT, K.; WYBROW, M. Dunnart: A constraint-based network diagram authoring tool. In: SPRINGER. *International Symposium on Graph Drawing*. [S.l.], 2008. p. 420–431.
- DÍAZ, J.; FORT, M.; VÁZQUEZ, P. Tourvis: Narrative visualization of multi-stage bicycle races. *Computer Graphics Forum*, v. 40, n. 3, p. 531–542, 2021. Disponível em: <<https://onlinelibrary.wiley.com/doi/abs/10.1111/cgf.14327>>.
- EICK, S. Visual discovery and analysis. *IEEE Transactions on Visualization and Computer Graphics*, v. 6, n. 1, p. 44–58, 2000.
- ELLSON, J. et al. Graphviz— open source graph drawing tools. In: MUTZEL, P.; JÜNGER, M.; LEIPERT, S. (Ed.). *Graph Drawing*. Berlin, Heidelberg: Springer Berlin Heidelberg, 2002. p. 483–484. ISBN 978-3-540-45848-7.

- FERNANDEZ, A.; INSFRAN, E.; ABRAHÃO, S. Usability evaluation methods for the web: A systematic mapping study. *Information and Software Technology*, v. 53, n. 8, p. 789–817, 2011. ISSN 0950-5849. Advances in functional size measurement and effort estimation - Extended best papers. Disponível em: <<https://www.sciencedirect.com/science/article/pii/S0950584911000607>>.
- FIELDING, R. T. Rest: architectural styles and the design of network-based software architectures. *Doctoral dissertation, University of California*, 2000.
- FISTER, D.; JR, I. F.; FISTER, I. Visualization of cycling training. In: . [S.l.: s.n.], 2016.
- FORSELL, C.; COOPER, M. Questionnaires for evaluation in information visualization. In: *Proceedings of the 2012 BELIV Workshop: Beyond Time and Errors - Novel Evaluation Methods for Visualization*. New York, NY, USA: Association for Computing Machinery, 2012. (BELIV '12). ISBN 9781450317917. Disponível em: <<https://doi.org/10.1145/2442576.2442592>>.
- FREITAS, C. M. D. S. et al. Introdução à visualização de informações. *Revista de informática teórica e aplicada. Porto Alegre. Vol. 8, n. 2 (out. 2001), p. 143-158*, 2001.
- GOLDSBERRY, K. Courtvision : New visual and spatial analytics for the nba. In: . [S.l.: s.n.], 2012.
- GRATZL, S. et al. Lineup: Visual analysis of multi-attribute rankings. *IEEE Transactions on Visualization and Computer Graphics (InfoVis '13)*, v. 19, n. 12, p. 2277–2286, 2013.
- GUSMÃO, L. H. *Cartografia Temática de Inventário*. Facilitageo, 2021. Disponível em: <<https://facilitageo.com/2021/08/01/cartografia-tematica-de-inventario/>>.
- HAVRE, S. et al. Themeriver: Visualizing thematic changes in large document collections. *IEEE transactions on visualization and computer graphics*, IEEE, v. 8, n. 1, p. 9–20, 2002.
- HEER, J.; KONG, N.; AGRAWALA, M. Sizing the horizon: the effects of chart size and layering on the graphical perception of time series visualizations. In: *Proceedings of the SIGCHI conference on human factors in computing systems*. [S.l.: s.n.], 2009. p. 1303–1312.
- HENRY, N.; FEKETE, J.-D.; MCGUFFIN, M. J. Nodetrix: a hybrid visualization of social networks. *IEEE Transactions on Visualization and Computer Graphics*, v. 13, n. 6, p. 1302–1309, 2007.
- HUFFMAN, D. A technique for encoding elevation changes along a route. *Cartographic Perspectives*, n. 63, p. 83–86, Jun. 2009. Disponível em: <<https://cartographicperspectives.org/index.php/journal/article/view/cp63-huffman>>.

- ILIINSKY, N.; STEELE, J. *Designing data visualizations: Representing informational Relationships*. [S.l.]: "O'Reilly Media, Inc.", 2011.
- INSELBERG, A.; DIMSDALE, B. Parallel coordinates: a tool for visualizing multi-dimensional geometry. In: *Proceedings of the First IEEE Conference on Visualization: Visualization '90*. [S.l.: s.n.], 1990. p. 361–378.
- JAVED, W.; GHANI, S.; ELMQVIST, N. Polyzoom: Multiscale and multifocus exploration in 2d visual spaces. In: . [S.l.: s.n.], 2012. v. 2012.
- KEIM, D. Information visualization and visual data mining. *IEEE Transactions on Visualization and Computer Graphics*, v. 8, n. 1, p. 1–8, 2002.
- KEIM, D. A.; KRIEGEL, H.-P. Visualization techniques for mining large databases: A comparison. *IEEE Transactions on knowledge and data engineering*, IEEE, v. 8, n. 6, p. 923–938, 1996.
- KRAAK, M.-J.; ORMELING, F. *Cartography: visualization of geospatial data*. [S.l.]: CRC Press, 2020.
- KYSH, L. Information technologies in the agricultural sector of ukraine. *Scientific Journal of Polonia University*, 2022.
- LOYOLA, G. F. M.; MEIRA, M. T. *Proposta de aplicação de visualização de dados focada em planejamento urbano acessível*. Dissertação (B.S. thesis) — Universidade Tecnológica Federal do Paraná, 2018.
- LU, W.-L.; WANG, Y.-S.; LIN, W.-C. Chess evolution visualization. *IEEE Transactions on Visualization and Computer Graphics*, v. 20, n. 5, p. 702–713, 2014.
- MACEACHREN, A. M.; KRAAK, M.-J. Research challenges in geovisualization. *Cartography and geographic information science*, Taylor & Francis, v. 28, n. 1, p. 3–12, 2001.
- MALETZKE, A. G. *Uma metodologia para extração de conhecimento em séries temporais por meio da identificação de motivos e da extração de características*. Tese (Doutorado) — Universidade de São Paulo, 2009.
- MAYERS, R. See the light: 'biking-with' as an artographic method of public pedagogy. *Journal of Arts & Communities*, 2021.
- MCGARRY, T. Applied and theoretical perspectives of performance analysis in sport: Scientific issues and challenges. *International Journal of Performance Analysis in Sport*, Routledge, v. 9, n. 1, p. 128–140, 2009. Disponível em: <<https://doi.org/10.1080/24748668.2009.11868469>>.
- MUELDER, C.; MA, K.-L. Rapid graph layout using space filling curves. *IEEE Transactions on Visualization and Computer Graphics*, IEEE, v. 14, n. 6, p. 1301–1308, 2008.
- MUNZNER, T. *Visualization analysis and design*. [S.l.]: CRC press, 2014.

NAGPAL, R.; MEHROTRA, D.; BHATIA, P. The state of art in website usability evaluation methods. In: _____. [S.l.: s.n.], 2017. ISBN 9781522519454.

NIELSEN, J. Enhancing the explanatory power of usability heuristics. In: *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems*. New York, NY, USA: Association for Computing Machinery, 1994. (CHI '94), p. 152–158. ISBN 0897916506. Disponível em: <<https://doi.org/10.1145/191666.191729>>.

OLIVEIRA, G. et al. Visualizing running races through the multivariate time-series of multiple runners. In: . [S.l.: s.n.], 2013. p. 99–106.

ORTIGOSSA, E. S. *Visualização de séries temporais multidimensionais: uma análise sobre a qualidade do ar em São CarlosSP*. Tese (Doutorado) — Universidade de São Paulo, 2018.

PAGE, M.; MOERE, A. V. Towards classifying visualization in team sports. In: IEEE. *International Conference on Computer Graphics, Imaging and Visualisation (CGIV'06)*. [S.l.], 2006. p. 24–29.

PERIN, C.; BOY, J.; VERNIER, F. Using gap charts to visualize the temporal evolution of ranks and scores. *IEEE Computer Graphics and Applications*, IEEE, v. 36, n. 5, p. 38–49, 2016.

PERIN, C.; VUILLEMOT, R.; FEKETE, J.-D. Soccerstories: A kick-off for visual soccer analysis. *IEEE Transactions on Visualization and Computer Graphics*, v. 19, n. 12, p. 2506–2515, 2013.

PERIN, C. et al. State of the art of sports data visualization. In: WILEY ONLINE LIBRARY. *Computer Graphics Forum*. [S.l.], 2018. v. 37, n. 3, p. 663–686.

PLAISANT, C. The challenge of information visualization evaluation. In: *Proceedings of the Working Conference on Advanced Visual Interfaces*. New York, NY, USA: Association for Computing Machinery, 2004. (AVI '04), p. 109–116. ISBN 1581138679. Disponível em: <<https://doi-org.ez10.periodicos.capes.gov.br/10.1145/989863.989880>>.

PUCHER, J.; BUEHLER, R. Cycling towards a more sustainable transport future. *Transport reviews*, Taylor & Francis, v. 37, n. 6, p. 689–694, 2017.

QIN, X. Making data visualization more efficient and effective: a survey. *The VLDB Journal*, 2019.

RAUTER, S.; FISTER, I. A collection of sport activity files for data analysis and data mining. University of Ljubljana, 2015.

RIBARSKY, W. et al. Discovery visualization using fast clustering. *IEEE Computer Graphics and Applications*, v. 19, n. 5, p. 32–39, 1999.

RIBECCA, S. Dataviz Catalogue, 2013. Disponível em: <<https://datavizcatalogue.com/index.html>>.

- ROMANILLOS, G. et al. Big data and cycling. *Transport Reviews*, Taylor & Francis, v. 36, n. 1, p. 114–133, 2016.
- SADANA, R.; STASKO, J. Designing multiple coordinated visualizations for tablets. In: WILEY ONLINE LIBRARY. *Computer Graphics Forum*. [S.l.], 2016. v. 35, n. 3, p. 261–270.
- SHNEIDERMAN, B. The eyes have it: a task by data type taxonomy for information visualizations. In: *Proceedings 1996 IEEE Symposium on Visual Languages*. [S.l.: s.n.], 1996. p. 336–343.
- SLOCUM, T. A. et al. *Thematic cartography and geovisualization*. [S.l.]: CRC Press, 2014.
- STOUT, D. W. *Social Media Statistics 2022: Top networks by the numbers • Dustin Stout*. 2022. Disponível em: <<https://dustinstout.com/social-media-statistics/>>.
- TAN, D. et al. Adaptivtree: Adaptive tree visualization for tournament-style brackets. *IEEE Transactions on Visualization and Computer Graphics*, v. 13, n. 6, p. 1113–1120, 2007.
- TELEA, A. C. *Data visualization: principles and practice*. [S.l.]: CRC Press, 2014.
- TORRE-ARENAS, I. de la. *Movements and Transformations How Motion Conveys Messages, Improves Storytelling and Captivates Viewers*. Tese (Doutorado) — Northeastern University, 2017.
- TRAINA, A. J. M. *Suporte à visualização de consultas por similaridade em imagens médicas através de estrutura de indexação métrica*. Tese (Doutorado) — ICMC/USP, 2001.
- TUTIDA, S. M. *Mapeamento por cores baseado em aspectos de percepção visual*. Tese (Doutorado) — Universidade de São Paulo, 1998.
- WAGEMANS, J. et al. A century of gestalt psychology in visual perception: Ii. conceptual and theoretical foundations. *Psychological bulletin*, American Psychological Association, v. 138, n. 6, p. 1218, 2012.
- WARD, M. O.; GRINSTEIN, G.; KEIM, D. *Interactive data visualization: foundations, techniques, and applications*. [S.l.]: AK Peters/CRC Press, 2010.
- WARE, C. *Information visualization: perception for design*. [S.l.]: Morgan Kaufmann, 2019.
- WEBER, M.; ALEXA, M.; MÜLLER, W. Visualizing time-series on spirals. In: *Infovis*. [S.l.: s.n.], 2001. v. 1, p. 7–14.
- WOOD, J. Visualizing personal progress in participatory sports cycling events. *IEEE Computer Graphics and Applications*, v. 35, n. 4, p. 73–81, 2015.

Apêndice

A

APÊNDICE

ANEXO 1

Seção 1 de 11

Avaliação da Abordagem de Visualização Esportiva Proposta - CycleVis

O presente formulário está associado à pesquisa de mestrado conduzida pelo discente Iury Gomes de Oliveira, vinculado ao Programa de Pós-Graduação em Computação da Universidade Federal da Bahia (PGComp-UFBA). A finalidade principal deste instrumento consiste em coletar respostas relativas à aplicação da abordagem desenvolvida para a Visualização Esportiva de Ciclismo.

Essa coleta será realizada por meio de tarefas de exploração, questionários de satisfação e solicitações de feedback dos usuários. O foco dessas atividades é a avaliação da interação, percepção e dos aspectos visuais da abordagem proposta.

As tarefas compreendem identificar padrões visuais e a utilização de mecanismos de interação voltados para a extração de informações. O propósito é apurar a eficácia da abordagem e verificar sua aplicabilidade no contexto esportivo de ciclismo.

Informações de contato dos responsáveis pelo acompanhamento da pesquisa:

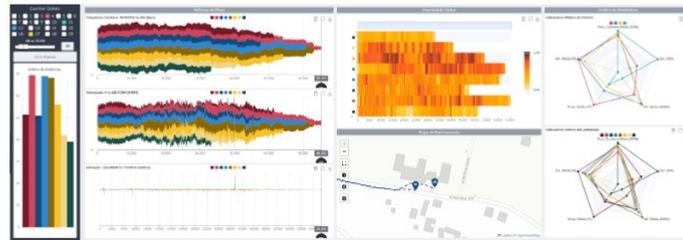
Pesquisador Responsável (mestrando): **Iury Gomes de Oliveira**

Pesquisador Orientador: **Prof. Dr. Danilo Barbosa Coimbra (UFBA)**

Não será necessário informar dados de identificação como nome, endereço ou documentos pessoais, como RG, CPF ou CNH.

Este formulário está coletando automaticamente os e-mails de todos os participantes. [Alterar configurações](#)

Figura 1 - CycleVis: Uma Abordagem de Visualização Esportiva no Ciclismo



Após a seção 1 Continuar para a próxima seção

Seção 2 de 11

Dados dos Participantes

Descrição (opcional)

Qual a sua idade ? *

- 21 - 25 anos
- 26 - 30 anos
- 31 - 35 anos
- 36 - 40 anos
- 41 - 45 anos
- 46 - 50 anos
- mais de 50 anos

Nível Educacional ? *

- Ensino Médio
- Ensino Superior
- Especialização Latu Sensu
- Mestrado
- Doutorado

Gênero ? *

- Masculino
- Feminino

Você possui experiência com técnicas de Visualização de Dados? *

- Não
- Baixo
- Médio
- Alto

Você pratica ciclismo externamente (rua, parque, ciclovia, rodovia, avenida, competição,...)? *

- Sim
- Não

Caso você pratique ciclismo, com que frequência você costuma pedalar ?

- Todos os dias ou quase todos os dias
- Algumas vezes por semana (2 a 3 vezes)
- Uma vez por semana
- Algumas vezes por mês (1 a 3 vezes)
- Uma a cinco vezes por semestre

Em caso afirmativo na penúltima pergunta, qual é a distância média que você percorre em cada pedalada ?

- Menos de 5 KM
- Entre 5 KM e 15 KM
- Entre 15 KM e 30 KM
- Entre 30 KM e 50 KM
- Mais de 50 KM

Após a seção 2 Continuar para a próxima seção

Seção 3 de 11

Tarefa 01



Objetivo: Avaliar a habilidade do usuário em filtrar e selecionar pedaladas de um ciclista específico.

Antes de iniciar essa Tarefa, aperte F5 para atualizar o navegador

Descrição (opcional)

Selecione os ciclistas com números 4, 12, 17 e 11. Em seguida, estabeleça o intervalo entre 39 e 70 quilômetros (KM).

Depois, clique no botão "Ver Table Lens" para ver detalhes sobre os ciclistas selecionados. Nesta nova tela, selecione:

- 1 - As pedaladas 51.31KM e 69.36 KM do ciclista número 4;
- 2 - As pedaladas 50.98 KM e 68.98 KM do ciclista número 12;

3 - As pedaladas 55.94 KM e 68.24 KM do ciclista número 17;

4 - A pedalada 39.00 KM do ciclista número 11;

5 - Feche a janela do Table Lens e clique em Atualizar Painéis.

Descrição (opcional)

Qual o nível de complexidade dessa tarefa? *

Muito Baixo 1 2 3 4 5 Muito Alto

Após a seção 3 Continuar para a próxima seção

Seção 4 de 11

Tarefa 02

Objetivo: Avaliar a capacidade do usuário identificar os padrões visuais que mapeiam a distância das pedaladas.

Identifique as pedaladas de maior ou menor distância e responda com a afirmação verdadeira.

Descrição (opcional)

Resposta: *

- A pedalada verde corresponde ao ciclista 11 e possui a menor distância
- A pedalada azul escura corresponde ao ciclista 12 e possui a maior distância
- A pedalada vermelha corresponde ao ciclista 17 e possui a maior distância
- A pedalada azul clara corresponde ao ciclista 11 e possui a menor distância

Qual painel visual você mais utilizou para compreender as distâncias das pedaladas?

- Gráfico de Distâncias
- Métricas de fluxo
- Intensidade Global

Qual o nível de complexidade dessa tarefa? *

Muito Baixo 1 2 3 4 5 Muito Alto

Após a seção 4 Continuar para a próxima seção

Seção 5 de 11

Tarefa 03

Objetivo: Avaliar a perspectiva de comparação nas métricas de desempenho considerando múltiplas pedaladas e somente uma pedalada.

No painel visual **métricas de fluxo**, assinale a **alternativa correta** em relação comportamento global das métricas:

- O ciclista 17 sempre começa as pedaladas com frequência cardíaca maior do que os outros ciclistas
- A velocidade global/média da pedalada do ciclista 11 é maior do que as pedaladas dos outros ciclistas

- O ciclista 11 teve o maior declive (elevação negativa)
- A velocidade global/média da pedalada vermelho escuro é menor do que as outras pedaladas

No painel visual **métricas de fluxo** selecione apenas a pedalada de cor vermelha escura em cada gráfico que mostra a elevação, velocidade e frequência cardíaca.

Em seguida, use a régua e ajuste nos gráficos de elevação, velocidade e frequência cardíaca para a distância de 39800 metros.

Qual a afirmação verdadeira?

- Uma subida na elevação no ponto de 40.000 metros faz a frequência cardíaca aumentar no ponto de 39...
- Um aumento na elevação em 39.800 metros faz a velocidade subir no ponto de 40.000 metros.
- Quando a elevação sobe em 39.800 metros, a velocidade até chegar em 39.800 metros diminui.
- Na elevação em 39800 metros, a a frequência cardíaca cai abruptamente para menos de 100.

Qual o nível de complexidade dessa tarefa? *

	1	2	3	4	5	
Muito Baixo	<input type="radio"/>	Muito Alto				

Após a seção 5 Continuar para a próxima seção

Seção 6 de 11

Tarefa 04

Objetivo: Verificar se o usuário consegue interpretar, compreender e comparar as intensidades globais das pedaladas.

No painel visual de Intensidade Global, assinale a alternativa correta: *

- A pedalada com maior intensidade global é a do ciclista 11
- O ciclista 17 tem pedaladas com maiores intensidades globais do que os outros ciclistas
- A pedalada vermelho escuro não contém a maior intensidade global, ou seja, valor da intensidade global ...
- O último segmento da pedalada azul escuro tem alta intensidade global

Qual o nível de complexidade dessa tarefa? *

	1	2	3	4	5	
Muito Baixo	<input type="radio"/>	Muito Alto				

No painel visual de Intensidade Global, selecione o intervalo de 20.000 metros e 30.000 metros no datazoom e **assinale a alternativa correta:** *

- O ciclista 17 possui pelo menos uma pedalada que consegue chegar no máximo nível de intensidade glo...
- O ciclista 4 pelo menos uma pedalada que consegue chegar no máximo nível de intensidade global
- A pedalada verde possui predominantemente altos níveis de intensidade.
- O ciclista 12 tem níveis de intensidade predominantemente menores que o ciclista 4

Qual o nível de complexidade dessa tarefa? *

	1	2	3	4	5	
Muito Baixo	<input type="radio"/>	Muito Alto				



Seção 7 de 11

Tarefa 05



Objetivo: Avaliar a perspectiva dos indicadores médios na painel visual Gráficos Estatísticos.



Analise o painel visual de Gráficos Estatísticos, em específico os Indicadores Médios dos Ciclistas, e responda à **alternativa correta**:

- A maior elevação média e maior temperatura é do ciclista 4
- O ciclista 11 tem a maior velocidade média e menor duração em minutos
- O ciclista 17 tem a maior frequência cardíaca média
- O ciclista 12 tem a menor temperatura média



Qual o nível de complexidade dessa tarefa? *

	1	2	3	4	5	
Muito Baixo	<input type="radio"/>	Muito Alto				

Analise o painel visual de Gráficos Estatísticos, em específico os indicadores médios das pedaladas, e responda à **alternativa correta**:

- A maior frequência cardíaca média das pedaladas do ciclista 17 é maior que 150 bpm
- O ciclista 12 possui pedalada com temperatura média abaixo de 9 graus Celsius
- A pedalada com menor duração média do ciclista 4 está abaixo de 150 minutos
- Nenhuma das pedaladas tem frequência cardíaca acima de 140 bpm



Qual o nível de complexidade dessa tarefa?

	1	2	3	4	5	
Muito Baixo	<input type="radio"/>	Muito Alto				



Seção 8 de 11

Tarefa 06



Objetivo: Mensurar a competência do usuário em explorar a localização geográfica das pedaladas selecionadas.



Analise o painel visual de Mapa de Rastreamento, selecione a **camada de distância** e responda com a **afirmativa correta**:

- A distância da pedalada azul escura para a pedalada amarelo escuro é de 140.25 KM
- Em relação a pedalada azul claro, a pedalada vermelha clara está mais perto do que a vermelha escura
- As pedaladas mais próximas do ciclista 17 são as pedaladas do ciclista 11
- A pedalada do ciclista 11 é a mais próxima da Itália



Qual o nível de complexidade dessa tarefa?

	1	2	3	4	5	
-	-	-	-	-	-	-

Muito Baixo



Muito Alto

Analise o painel visual de Mapa de Rastreamento, selecione a **camada de rotas** e responda com a **afirmativa incorreta**: *

- As pedaladas do ciclista 4 possuem intersecção, ou seja, alguns segmentos destas pedaladas são os m...
- A pedalada vermelha escura é realizada no sentido horário
- Todas as pedaladas estão no formato de circuito, ou seja, todas o ponto de início e final é praticamente ...
- A pedalada amarela clara é realizada no sentido horário

Qual o nível de complexidade dessa tarefa?

Muito Baixo 1 2 3 4 5 Muito Alto

Analise o painel visual de Mapa de Rastreamento, selecione a **camada de intensidade** e responda com a **afirmativa correta**: *

- Na pedala azul escuro é possível notar alta intensidade global na região de Pristava
- As pedaladas do ciclista 4 tem alta intensidade na região de Maribor
- A pedalada amarelo escuro tem alta intensidade na região de Spodnji Poljsnik
- Na visão com todas as pedadalas aparecendo no mapa, a pedalada azul escuro possui maior intensidade

Qual o nível de complexidade dessa tarefa? *

Muito Baixo 1 2 3 4 5 Muito Alto

Após a seção 8 Continuar para a próxima seção

Seção 9 de 11

Tarefa 07

Objetivo: Avaliar a habilidade do usuário em interagir com visualizações coordenadas

Selecione no DataZoom no painel visual da **Intensidade Global** o intervalo de distância entre 20000 e 32000 aproximadamente

Descrição (opcional)

Analise o painel visual de Intensidade Global e o painel visual de Métricas de Fluxo e assinale a alternativa correta *

- Ocorre intensidade global alta na pedala amarelo claro e aumento da velocidade na mesma pedalada pa...
- Ocorre intensidade globais altas nas pedaladas vermelhas e muita elevação nas mesmas pedaladas par...
- A frequência cardíaca da pedalada azul escura é 119 BPM na distância 27400, possuindo Intensidade Gl...
- Existe elevação acima de 30 metros para as pedaladas com intensidade global abaixo de 0.5 no ciclista 4

Qual o nível de complexidade dessa tarefa? *

Muito Baixo 1 2 3 4 5 Muito Alto

Após a seção 9 Continuar para a próxima seção

8 - Você concorda que foi útil utilizar mais de um painel visual para encontrar a informação desejada?

1 2 3 4 5

Discordo totalmente Concordo totalmente



9 - Você acredita que a ferramenta proposta poderia ser utilizada em competições profissionais de ciclismo? *

1 2 3 4 5

Discordo totalmente Concordo totalmente



10 - Você concorda que a ferramenta pode auxiliar no desenvolvimento de estratégias para a atividade do ciclismo? *

1 2 3 4 5

Discordo totalmente Concordo totalmente



11 - Você concorda que a ferramenta proposta poderia ser adaptada para análise de outras modalidades esportivas? *

1 2 3 4 5

Discordo totalmente Concordo totalmente



Após a seção 10 Continuar para a próxima seção ▼

Seção 11 de 11

Feedback



Descrição (opcional)

Por favor, compartilhe seu feedback, impressões, sugestões, críticas e elogios sobre a ferramenta apresentada. *

Texto de resposta longa



Você gostaria de ser informado caso os resultados da pesquisa sejam publicados em um artigo? *

- Sim
- Não



Caso sim na resposta anterior, informe um endereço de e-mail válido para receber os dados da publicação.

Texto de resposta curta