

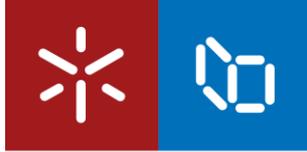


**Universidade do Minho**  
Escola de Letras, Artes e Ciências Humanas

Raquel Rossi Ferreira

**DataViz: Aplicativo em R Shiny para  
Escolha Eficiente de Visualizações  
de Dados**





**Universidade do Minho**  
Escola de Letras, Artes e Ciências Humanas

Raquel Rossi Ferreira

## **DataViz: Aplicativo em R Shiny para Escolha Eficiente de Visualizações de Dados**

Dissertação de Mestrado  
Mestrado em Humanidades Digitais

Trabalho efetuado sob a orientação do  
**Professor Doutor Álvaro Iriarte Sanromán**

junho de 2024

## **DIREITOS DE AUTOR E CONDIÇÕES DE UTILIZAÇÃO DO TRABALHO POR TERCEIROS**

Este é um trabalho académico que pode ser utilizado por terceiros desde que respeitadas as regras e boas práticas internacionalmente aceites, no que concerne aos direitos de autor e direitos conexos.

Assim, o presente trabalho pode ser utilizado nos termos previstos na licença abaixo indicada.

Caso o utilizador necessite de permissão para poder fazer um uso do trabalho em condições não previstas no licenciamento indicado, deverá contactar o autor, através do RepositóriUM da Universidade do Minho.

### **Licença concedida aos utilizadores deste trabalho**



**Atribuição-NãoComercial**

**CC BY-NC**

<https://creativecommons.org/licenses/by-nc/4.0/>

## **AGRADECIMENTOS**

Gostaria de agradecer a minha família por todo o apoio durante o período do mestrado, em especial para a minha mãe que me apoiou desde o início quando eu ainda estava no Brasil e decidi escolher Portugal para continuar meus estudos.

Ao Ronaldo, além de ser um ótimo irmão, nunca se importou de ter que resolver minhas burocracias com os bancos no Brasil para poder me mandar dinheiro todos os meses, e ainda foi me buscar no aeroporto quando fui pra casa no Natal de 2023 com uma placa de “Bem-vinda de volta Rachel”.

Aos meus colegas de mestrado que fizeram essa jornada ser leve e divertida, pois sempre nos apoiávamos nas entregas dos relatórios finais de cada disciplina e tentávamos ajustar com os professores as datas dessas entregas para que ninguém fosse prejudicado.

A Ana, Bruna e Wylliam por fazerem os meus dias mais divertidos, mesmo que a distância. Obrigada por participarem de ligações de vídeo tarde da noite nas sextas-feiras, por participarem através de fotos e vídeos das minhas viagens pela Europa Ocidental no verão e por apoiarem minhas decisões (por mais malucas que fossem). Vocês são os melhores amigos que alguém poderia ter.

Ao meu orientador, professor Álvaro Iriarte Sanromán, muito obrigada por todo o suporte, motivação e ensinamentos.

Por fim, um grande agradecimento a Universidade do Minho, pela oportunidade de ensino, aprendizado e crescimento profissional e pessoal.

## **DECLARAÇÃO DE INTEGRIDADE**

Declaro ter atuado com integridade na elaboração do presente trabalho académico e confirmo que não recorri à prática de plágio nem a qualquer forma de utilização indevida ou falsificação de informações ou resultados em nenhuma das etapas conducente à sua elaboração.

Mais declaro que conheço e que respeitei o Código de Conduta Ética da Universidade do Minho.

## RESUMO

A análise e interpretação de informações quantitativas são fundamentais para a tomada de decisões, tanto no âmbito profissional quanto no pessoal. A apresentação de tabelas, gráficos e diagramas é uma prática cotidiana nos meios de comunicação, nas universidades, nos negócios, na administração, na saúde, etc.

A Estatística é um componente comum na estrutura curricular da maioria dos cursos de graduação, exceto nos cursos de humanidades, sendo essencial na formação de novos profissionais em qualquer campo de estudo.

A importância da aplicação correta da Estatística pode ser observada, por exemplo, através da grande quantidade de equívocos presentes em pesquisas e artigos científicos na área da saúde (Karadeniz et al., 2019). Uma das possíveis razões para tais problemas reside na dificuldade dos estudantes em reconhecer a relevância das disciplinas associadas à Estatística em seus campos de atuação. Isso ocorre juntamente com atitudes negativas em relação à disciplina, deficiências na base matemática dos estudantes, ênfase excessiva nos cálculos e falta de aplicação e prática durante o curso de graduação (Cazorla, 2015). Portanto, é necessário estabelecer uma comunicação mais eficaz entre teoria e prática para melhorar o processo de aprendizagem. Assim, a adoção de métodos pedagógicos eficazes torna-se essencial para capacitar o aluno a ser um agente ativo na sua própria educação (Costa et al., 2019).

Atualmente, também nas ciências humanas é imprescindível ter algumas noções básicas de Estatística e saber configurar gráficos adequados, daí a importância de implementar na oferta educativa (de 1º, 2º e 3º ciclo) unidades curriculares, seminários ou *workshops* sobre métodos de investigação em ciências humanas e sociais, onde os métodos quantitativos ou mistos deverão ter uma relevância particular, permitindo ultrapassar análises baseadas em meras intuições subjetivas.

As máquinas vão permitir-nos fundamentar, apoiar em corpora e em dados, as nossas intuições subjetivas. O uso da tecnologia possibilita a criação de materiais interativos que integram hipertextos, aplicações multimídia, simulações e outros recursos, os quais os estudantes acessam dentro de um contexto que poderia ser descrito como uma atitude de aprendizagem ativa. Velleman e Moore (1996) enfatizam que as tecnologias podem ser empregadas não apenas para simplificar a aplicação de ferramentas estatísticas, mas também no aprimoramento do processo de ensino e aprendizagem da disciplina.

**Palavras-chave:** Visualização de dados, Estatística, Ciências Humanas, Humanidades Digitais

## **ABSTRACT**

The analysis and interpretation of quantitative information are fundamental for decision-making, both professionally and personally. The presentation of tables, graphs and diagrams is an everyday practice in the media, universities, business, administration, healthcare, etc.

Statistics is a common component in the curricular structure of most undergraduate courses, except for humanities courses, and is essential in the training of new professionals in any field of study.

The importance of correctly applying Statistics can be observed, for example, through the large number of mistakes present in research and scientific articles in the area of health (Karadeniz et al., 2019). One of the possible reasons for such problems lies in the difficulty students have in recognizing the relevance of disciplines associated with Statistics in their fields of activity. This occurs along with negative attitudes towards the discipline, deficiencies in students' mathematical base, excessive emphasis on calculations and lack of application and practice during the undergraduate course (Cazorla, 2015). Therefore, it is necessary to establish more effective communication between theory and practice to improve the learning process. The choice of effective pedagogical methods becomes essential to enable students to be active agents in their own education (Costa et al., 2019).

Currently, also in the human sciences it is essential to have some basic notions of Statistics and know how to configure appropriate graphs, hence the importance of implementing in the educational offer (1st, 2nd and 3rd cycle) curricular units, seminars or workshops on research methods in human sciences and social, where quantitative or mixed methods should have particular relevance, allowing analyzes based on mere subjective intuitions to be overcome.

Machines will allow us to substantiate, support our subjective intuitions with corpora and data. The use of technology enables the creation of interactive materials that integrate hypertexts, multimedia applications, simulations and other resources, which students access within a context that could be described as an active learning attitude. Velleman and Moore (1996) emphasize that technologies can be used not only to simplify the application of statistical tools, but also to improve the teaching and learning process of the subject.

**Keywords:** Data Visualization, Statistics, Human Sciences, Digital Humanities

## ÍNDICE

DIREITOS DE AUTOR E CONDIÇÕES DE UTILIZAÇÃO DO TRABALHO POR TERCEIROS	i
AGRADECIMENTOS	ii
DECLARAÇÃO DE INTEGRIDADE	iii
RESUMO	iv
ABSTRACT	v
ÍNDICE	vi
ÍNDICE DE FIGURAS	viii
<b>1 Introdução</b>	<b>1</b>
<b>2 Contextualização</b>	<b>3</b>
2.1 Identificação do problema e objetivos	3
2.2 Contextualização teórica	3
2.2.1 A análise e interpretação baseada em dados	3
2.2.2 A análise quantitativa nas ciências humanas	4
2.2.3 Diferentes formas de apresentar dados	6
2.2.4 Principais erros em gráficos estatísticos	12
2.2.5 O uso de cores	17
2.2.6 Software R, R Studio e Shiny	19
2.2.7 Aplicabilidade do R Shiny no contexto do ensino de visualização de dados	20
<b>3 Metodologia</b>	<b>22</b>
3.1 Sumarização das etapas de trabalho realizado	22
3.2 Criação do wireframe	23
3.3 Desenvolvimento do aplicativo	26
3.4 Questionário	29
3.4.1 Regras adotadas	29

3.4.2 Aplicação do questionário	30
<b>4 Análise e Discussão dos Resultados</b>	<b>32</b>
<b>5 Conclusões</b>	<b>36</b>
<b>BIBLIOGRAFIA</b>	<b>37</b>
<b>ANEXOS</b>	<b>42</b>

## ÍNDICE DE FIGURAS

Figura 1. Exemplo de um gráfico de barras.....	7
Figura 2. Exemplo de um gráfico de barras agrupadas.....	8
Figura 3. Exemplo de um gráfico de setores.....	9
Figura 4. Exemplo de um gráfico de linha.....	10
Figura 5. Exemplo de um gráfico de dispersão.....	11
Figura 6. Gráfico de barras com escala desproporcional.....	13
Figura 7. Gráfico de barras com alturas incorretas.....	14
Figura 8. Gráfico de linhas com mudança no eixo y.....	14
Figura 9. Gráfico de linhas corrigido.....	15
Figura 10. Gráfico de setores com erro.....	16
Figura 11. Gráfico de área invertido.....	17
Figura 12. Gráfico de barras com segmentação por cor.....	18
Figura 13. Mapa representando intensidade de cor.....	19
Figura 14. Etapas do trabalho realizadas.....	22
Figura 15. Esboço da tela inicial do aplicativo.....	23
Figura 16. Primeira versão da tela inicial do aplicativo.....	24
Figura 17. Esboço de uma das abas do aplicativo.....	25
Figura 18. Primeira versão de uma das abas do aplicativo.....	26
Figura 19. Dashboard ilustrando uma boa e má visualização gráfica.....	31
Figura 20. Versão final da tela inicial do aplicativo.....	33
Figura 21. Versão final de uma das abas do aplicativo.....	33
Figura 22. Comentários nos códigos em R.....	34
Figura 23. Comentários nos códigos em Python.....	35

## 1 Introdução

Com os avanços tecnológicos dos últimos anos, as aptidões e habilidades para a formação dos cidadãos se tornam cada vez mais complexas e desafiadoras. Nesse sentido, de acordo com Cazorla, “ser alfabetizado, no mundo atual, significa não apenas o domínio da língua materna, mas também o domínio da linguagem numérica” (Cazorla, 2004, p.1).

Desde que ingressei no ensino superior e tive o primeiro contato com as disciplinas do bacharelado em Estatística, aquela com a qual tive mais afinidade e interesse foi a Estatística Descritiva, onde o objetivo é resumir e descrever os atributos mais importantes dos dados. Os gráficos desempenham um papel fundamental na Estatística Descritiva, pois oferecem uma representação visual dos dados que facilita a interpretação e a compreensão das informações. Eles ajudam a resumir grandes conjuntos de dados de forma visualmente intuitiva, permitindo identificar padrões, tendências e relações entre variáveis de maneira mais simples e eficaz.

A Estatística é uma ferramenta poderosa, mas também pode ser facilmente mal interpretada ou manipulada se não for usada corretamente. Segundo Huff (1954), os números podem ser apresentados de maneira enganosa para distorcer a verdade ou induzir a falsas conclusões. Infelizmente, ao longo dos anos, vimos muitos exemplos de uso inadequado da Estatística em diversos contextos, incluindo a mídia, a política e até mesmo a academia.

O propósito central deste trabalho é abordar e resolver um desafio comum enfrentado pelos alunos de Humanidades: Como escolher o gráfico correto na análise dos dados.

A proposta do desenvolvimento de um aplicativo em R *Shiny* surgiu após ser observado que na disciplina de Análise e Visualização de Dados do mestrado de Humanidades Digitais da Universidade do Minho, os alunos, maioritariamente provenientes de licenciaturas nas áreas de Letras e Ciências Humanas enfrentavam uma certa dificuldade na tomada de decisão para a escolha da visualização gráfica que melhor se adaptava aos dados analisados nos seus relatórios (Chang et al., 2024).

Inicialmente será feita uma breve revisão teórica sobre a análise e interpretação baseada em dados, explorando sua importância em diferentes contextos e sua aplicabilidade nas ciências humanas, com ênfase na análise quantitativa.

Em seguida, será examinada a variedade de formas de apresentar dados, destacando a necessidade de escolher o tipo de gráfico adequado e evitando os principais erros em gráficos estatísticos, como o uso inadequado de cores e a tipologia, que podem comprometer a compreensão dos dados.

Posteriormente, será feita uma breve introdução sobre as vantagens da utilização do R *Shiny* e como se pode usar essa ferramenta interativa no ensino da visualização de dados de forma divertida e eficiente.

A criação do *wireframe* será feita utilizando Figma (Figma, 2024). O desenvolvimento do aplicativo será feito utilizando a biblioteca de funções *Shiny* do *Software R*.

A disponibilização do aplicativo será via servidor do *Shiny* gerando um *link* que ficará disponível tanto para os alunos quanto para os professores que desejem utilizá-lo. Por fim, os resultados serão apresentados, analisando também a eficiência pedagógica do aplicativo, bem como a usabilidade e experiência do usuário, que será avaliada a partir da distribuição de um breve questionário a ser construído no *Google Forms* e disponibilizado para os alunos da edição de 2023/2024 de Análise e Visualização de Dados do mestrado de Humanidades Digitais.

## **2 Contextualização**

### **2.1 Identificação do problema e objetivos**

Uma das disciplinas onde os alunos do ensino superior apresentam maior dificuldade de aprendizagem é a disciplina de Estatística (Felgueiras, 2013), sendo a forma de apresentar a informação por meio de gráficos e tabelas um aspecto que apresenta maior dificuldade, possivelmente devido à falta de consciência da complexidade de escolher o gráfico adequado.

O objetivo deste trabalho é desenvolver uma aplicação *web* que seja simples e intuitiva, visando como público alvo estudantes de Humanidades.

O propósito do aplicativo será orientar o utilizador na seleção mais adequada da visualização gráfica dos dados analisados.

A utilização da ferramenta, simples e interativa, facilitará a elaboração de relatórios e trabalhos acadêmicos, especificamente no que se refere à visualização dos dados previamente analisados, permitindo aos utilizadores explorarem novas opções de visualizações em linguagens como R e Python.

### **2.2 Contextualização teórica**

A seguir são apresentadas definições de conceitos importantes relacionados com a análise de dados, seguido do desenvolvimento do tema ao qual este trabalho se propõe.

#### **2.2.1 A análise e interpretação baseada em dados**

A análise e interpretação de dados são habilidades essenciais no contexto atual, onde a quantidade de informações disponíveis é vasta e crescente. Como destacado por Grolemond e Wickham (2017) em "*R for Data Science*", a capacidade de manipular e visualizar dados de maneira eficaz é fundamental para extrair *insights* significativos e tomar decisões informadas. Além disso, conforme abordado por Tufte (1983) em "*The Visual Display of Quantitative Information*", a escolha do tipo adequado de gráfico ou tabela desempenha um papel crucial na comunicação clara e precisa de resultados analíticos.

Conforme destacado por Tufte (1983), os gráficos bem elaborados são frequentemente os instrumentos mais poderosos e simples de informação, capazes de simplificar conceitos complexos e contribuir para importantes descobertas científicas. Eles possuem a capacidade de apresentar padrões e tendências nos dados de maneira intuitiva, permitindo que pesquisadores e profissionais obtenham informações valiosas e façam escolhas fundamentadas.

Além disso, como afirmado por Wilkinson (2005), a visualização de dados por meio de gráficos não apenas simplifica a interpretação, mas também pode revelar relações e fenômenos ocultos nos dados, levando a avanços significativos em uma variedade de campos, desde a medicina até as ciências sociais. Portanto, a criação de gráficos bem projetados é essencial para a análise eficaz e a comunicação clara de informações em diversas áreas do conhecimento.

Nos métodos de ensino contemporâneos, destaca-se a dificuldade recorrente dos estudantes na criação e interpretação de gráficos estatísticos, (Curcio, 1987; Morais, 2010). Esse cenário contrasta com a percepção comumente compartilhada por muitos docentes, os quais tendem a considerar a Estatística “um tema para o qual os alunos são facilmente motivados e em cuja aprendizagem não apresentam grandes dificuldades” (Sousa, 2002, p. 4).

Esse desafio envolve vários aspectos, para além da aprendizagem das medidas utilizadas na estatística descritiva, como a escolha do tipo de gráfico adequado para os dados, a habilidade de criar gráficos visualmente atraentes e informativos, a interpretação precisa dos resultados apresentados e o domínio do *Software* de gráficos utilizado. Além disso, os alunos também precisam ser capazes de comunicar efetivamente suas descobertas por meio dos gráficos, explicando claramente os resultados e como eles se relacionam com a problemática em análise.

Para superar essas dificuldades, é essencial que os educadores forneçam orientação e recursos adicionais, além de oportunidades para os alunos praticarem a criação e interpretação de gráficos em diferentes contextos (Batanero et al., 2011). O *feedback* construtivo sobre os gráficos criados pelos alunos também pode ser útil para ajudá-los a aprimorar suas habilidades nessa área.

Se o processo de ensino for embasado no uso de tecnologias digitais, pode-se observar um potencial motivador para os estudantes, uma vez que essas ferramentas facilitam a prática da aprendizagem colaborativa e o desenvolvimento de atividades em grupo. Essa abordagem permite que os alunos aprendam uns com os outros e explorem, construam e compreendam conceitos ao modelarem o pensamento estatístico (Lopes, 2013).

### **2.2.2 A análise quantitativa nas ciências humanas**

Os programas de graduação geralmente incluem disciplinas de Estatística em suas grades curriculares. Embora essas disciplinas frequentemente abordam conteúdos teóricos, elas servem como alicerce para a aprendizagem de disciplinas mais específicas do curso. Os estudantes de cursos de ciências humanas frequentemente manifestam resistência em relação à disciplina de Estatística devido ao fato de terem que aprender noções de Estatística sem grandes conhecimentos de matemática, associado ao grande

volume de cálculos e fórmulas envolvidos, além da dificuldade em estabelecer uma conexão entre os conceitos estatísticos e sua aplicação em situações práticas (Nolan & Speed, 1999).

A Estatística é frequentemente percebida pelos alunos do ensino superior como uma extensão da matemática, o que pode ser atribuído em parte ao fato de que os conteúdos estatísticos no ensino fundamental e médio são muitas vezes ministrados por professores licenciados em matemática, muitos dos quais podem não ter preparação adequada para o ensino da Estatística, comprometendo assim a aprendizagem em todos os níveis de ensino.

Pesquisas realizadas por Cazorla (2002) revelam que os alunos enfrentam dificuldades mesmo na utilização de ferramentas de análise de dados mais simples, como a representação gráfica de dados. Dancey e Reidy (2019) observam que, embora muitos alunos possam compreender conceitos teóricos complexos, como o conceito de inteligência, lutam para entender os conceitos estatísticos, muitas vezes devido à percepção de que a Estatística é excessivamente centrada em fórmulas matemáticas.

Outros desafios enfrentados pelos alunos de Estatística no ensino superior, conforme identificados por Cazorla (2002), incluem a natureza abstrata dos conceitos estatísticos, a fundamentação teórica na matemática, a carga horária reduzida das disciplinas de Estatística em relação à extensão dos conteúdos abordados e questões afetivas, como a transferência de sentimentos negativos provenientes de experiências anteriores com a disciplina de matemática.

Esses fatores têm um impacto significativo na aprendizagem dos alunos. Assim, os professores de Estatística nos cursos de graduação em ciências humanas e sociais devem refletir sobre o propósito principal da disciplina e desenvolver suas práticas de ensino com foco no desenvolvimento do pensamento estatístico dos alunos, trabalhando suas atitudes e motivação para que sejam capazes de compreender e aplicar os conceitos estatísticos de forma eficaz.

No estudo conduzido por Barroso (2010), o objetivo foi investigar o papel da disciplina Estatística nos cursos das áreas sociais e humanas, além de examinar como os alunos aplicam os conceitos aprendidos em suas pesquisas científicas. A pesquisa adotou uma abordagem que combinou métodos quantitativos e qualitativos, envolvendo a coleta de dados de alunos do 7º e 8º períodos de diferentes cursos, bem como de alguns professores de uma instituição de ensino superior do Brasil.

Ao longo do estudo, foi perceptível a necessidade de repensar o ensino de Estatística no ensino superior, especialmente nas áreas das ciências humanas e sociais. Houve uma reflexão sobre a abordagem pedagógica, buscando fugir das aulas tradicionais focadas em cálculos e fórmulas, e enfatizando mais a discussão dos resultados. Identificou-se a importância de desenvolver o pensamento estatístico dos alunos e de adotar posturas pedagógicas que promovessem a compreensão e aplicação dos conceitos.

### 2.2.3 Diferentes formas de apresentar dados

Devemos reconhecer que, independentemente do grau de aplicação da Estatística em suas futuras carreiras, os estudantes estão se transformando em consumidores de informações estatísticas. Isso se deve ao fato de que todos nós estamos constantemente expostos a uma inundação de dados resumidos em gráficos e tabelas por meio de diversos meios de comunicação (Carver et al., 2016).

É perceptível que os gráficos estatísticos são amplamente utilizados para apresentar dados nos meios de comunicação. No entanto, é importante notar que, muitas vezes, o apelo visual desses gráficos pode ser mais impactante do que os próprios valores apresentados. Isso os torna suscetíveis a possíveis erros e usos indevidos, já que a interpretação visual pode distorcer a realidade dos dados. Portanto, embora os gráficos sejam ferramentas poderosas para comunicar informações de forma acessível, é essencial abordá-los com cautela e garantir que sejam utilizados de maneira precisa e responsável.

Cada modalidade de representação gráfica de dados apresenta uma estrutura composta por elementos distintos, cada um com sua função específica. Entre estes, destacam-se as dimensões visuais como linhas e barras, que servem para expressar os valores dos dados. Adicionalmente, são essenciais os rótulos, compreendendo os nomes dos eixos, o título do gráfico e as unidades de medida pertinentes. Por fim, o fundo do gráfico engloba elementos como cores, grades e figuras de suporte.

A competência em literacia Estatística não se limita apenas ao reconhecimento dos tipos de gráficos utilizados, mas também à habilidade de empregar a terminologia precisa para descrever suas distintas partes e, crucialmente, interpretar os dados que eles representam (Friel et al., 2001).

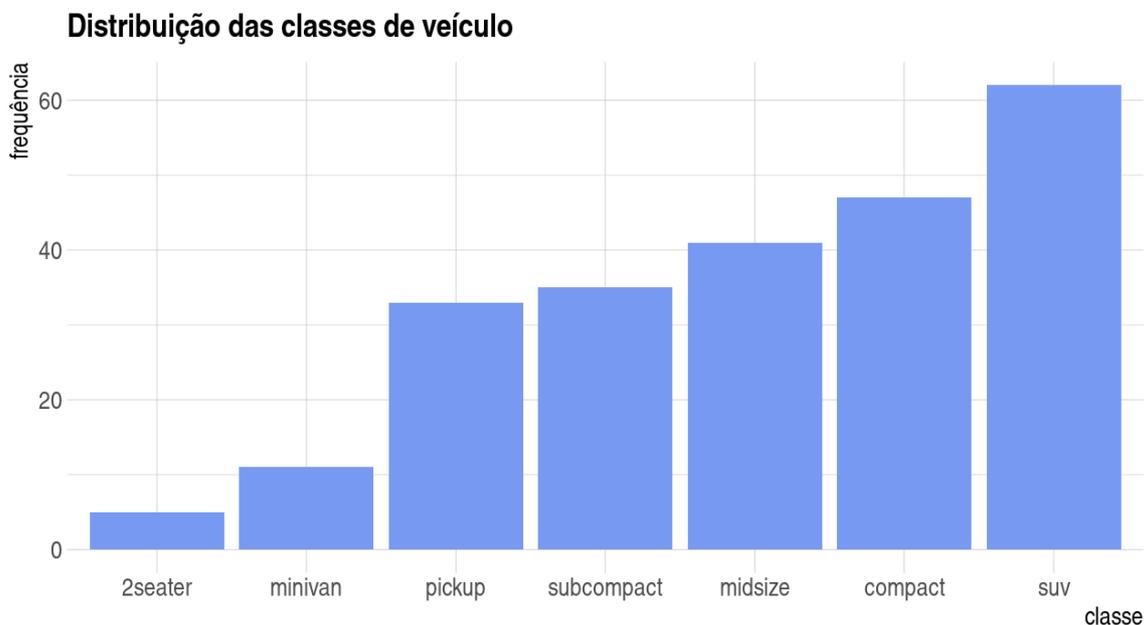
Segundo Prates (2017), gráficos estatísticos devem ser simples e possibilitar uma análise rápida dos dados observados, deve haver clareza possibilitando a leitura e interpretação correta dos valores e deve haver veracidade, expressando a verdade sobre os dados analisados.

A seguir, apresentamos uma breve explicação de uso dos tipos de gráficos estatísticos mais comumente utilizados na mídia e academia. A Figura 1, mostra a distribuição de classes de veículos em um gráfico de barras. A Figura 2, mostra a qualidade do corte de diamante de acordo com a cor em um gráfico de barras agrupadas. A Figura 3, mostra a porcentagem de formandos em cada curso a partir de um gráfico de setores. A Figura 4, mostra a popularidade de nomes americanos nos últimos 30 anos em um gráfico de linha. Por fim, a Figura 5, mostra a relação entre comprimento e largura de sépalas por espécie em um gráfico de dispersão. Os conjuntos de dados utilizados para o desenvolvimento dos gráficos são fictícios ou fazem parte de pacotes disponíveis no *Software R*.

## Gráfico de barras

Um gráfico de barras mostra a relação entre uma variável numérica e uma variável categórica. Cada entidade da variável categórica é representada como uma barra. O tamanho da barra representa seu valor numérico. Neste exemplo, foi utilizado o conjunto de dados mpg disponível no *Software R*, que contém informações sobre o uso de combustível em modelos de carros dos Estados Unidos que tiveram um novo lançamento a cada ano entre 1999 e 2008. O conjunto de dados contém 234 linhas e 11 variáveis, onde a variável classe corresponde ao tipo de carro.

Figura 1. Exemplo de um gráfico de barras.



Fonte: Autor (2024).

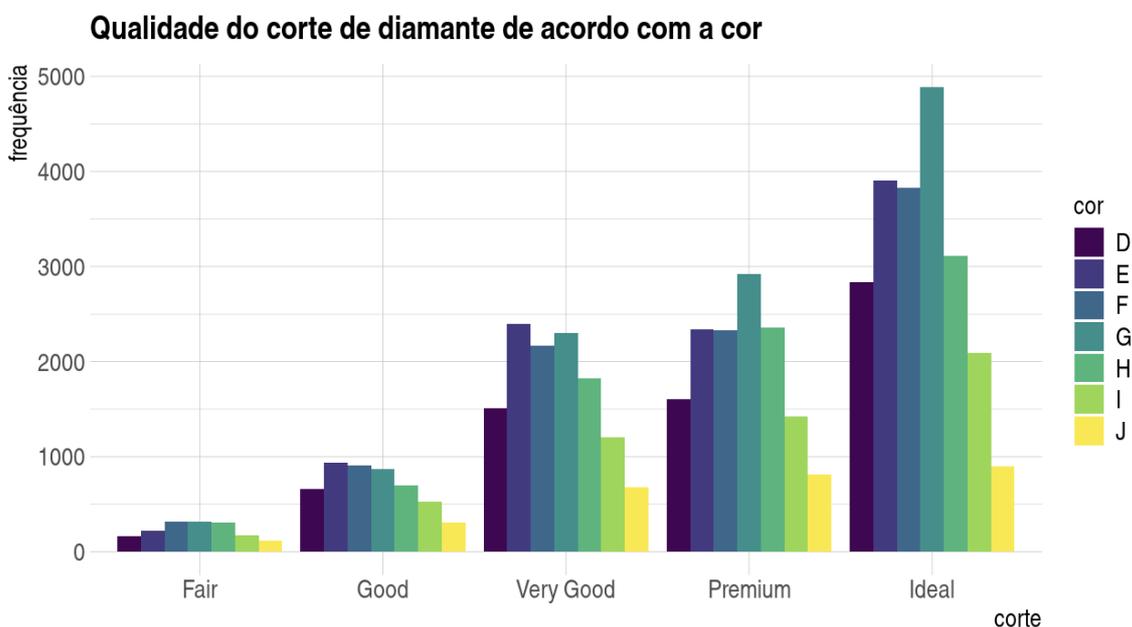
A proposta de visualização desse gráfico é mostrar a distribuição das classes de veículos, ou seja, com esse gráfico podemos perceber que os veículos que tiveram um maior número de lançamento entre os anos de 1999 e 2008 foram os do tipo SUV.

No gráfico de barras, se os dados não possuírem uma ordem óbvia, a abordagem adotada é ordenar os dados seguindo seus valores. No entanto, é importante considerar o contexto e o objetivo do gráfico ao decidir sobre a ordenação, pois às vezes manter a ordem original dos dados é a abordagem mais apropriada, especialmente se a ordem natural das categorias for importante para a interpretação dos resultados.

## Gráfico de barras agrupadas

Um gráfico de barras exibe o valor de uma variável numérica para diversas entidades. Essas entidades podem ser agrupadas usando uma variável categórica, resultando em um gráfico de barras agrupado. Neste exemplo, foi utilizado o conjunto de dados *diamonds* disponível no *Software R*, que contém informações sobre o preço e outros atributos de aproximadamente 54.000 diamantes. O conjunto de dados contém 53940 linhas e 10 variáveis, onde a variável corte define a qualidade do corte (*Fair*, *Good*, *Very Good*, *Premium*, *Ideal*) e cor define a cor do diamante seguindo uma escala entre D (melhor) a J (pior).

Figura 2. Exemplo de um gráfico de barras agrupadas.



Fonte: Autor (2024).

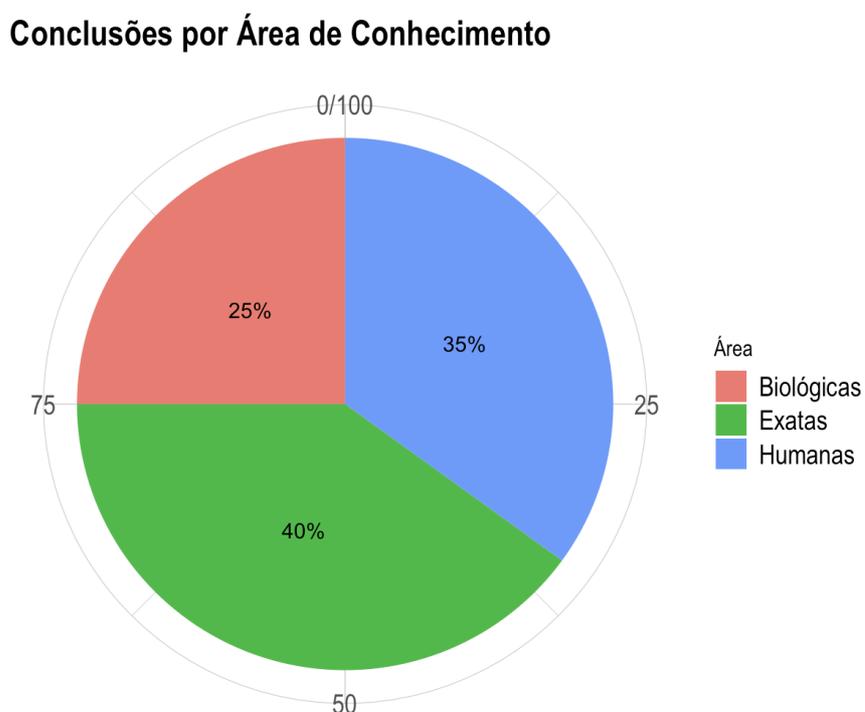
A proposta de visualização desse gráfico é mostrar a qualidade do corte de cada diamante de acordo com a sua cor, ou seja, com esse gráfico podemos perceber que o corte do tipo *Ideal* tem uma quantidade maior de diamantes na cor D (melhor), mas essa quantidade é praticamente a mesma entre diamantes de corte *Very Good* e *Premium* não existindo diferença significativa entre esses dois grupos.

No gráfico de barras agrupadas, é importante selecionar cores distintas para cada grupo, garantindo clareza na visualização. Se houver mais de uma série de dados representada no gráfico, forneça uma legenda clara que explique o que cada cor representa.

## Gráfico de setores

Um gráfico de setores exibe a distribuição proporcional dos elementos que compõem uma série de dados em relação ao total desses elementos. Este tipo de gráfico é especialmente adequado para destacar a importância de um único elemento dentro da série. Neste exemplo, foi utilizado um conjunto de dados fictício com a criação de listas que recebiam os valores com os nomes das áreas de conhecimento (Biológicas, Exatas e Humanas) e seus respectivos percentuais. Esses dados não correspondem a realidade do número de conclusões nessas áreas de conhecimento, os dados foram criados apenas para exemplificar o uso de um gráfico de setores.

Figura 3. Exemplo de um gráfico de setores.



Fonte: Autor (2024).

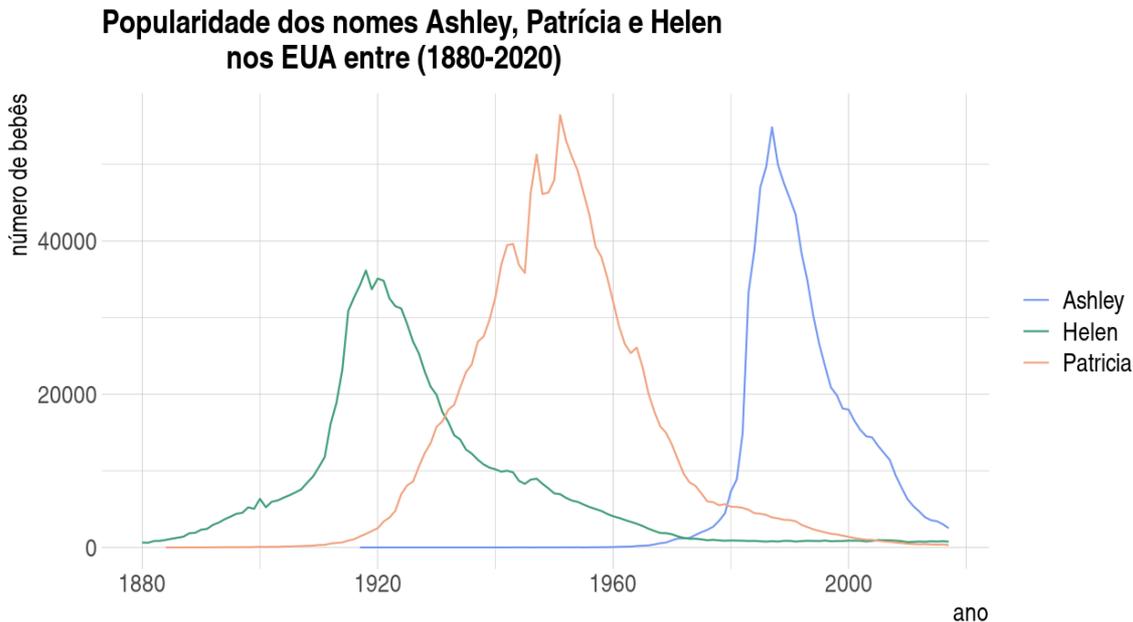
A proposta de visualização desse gráfico é mostrar o percentual de conclusões em cursos de três áreas de conhecimento (Biológicas, Exatas e Humanas), ou seja, com esse gráfico podemos perceber que cursos de áreas de exatas possuem um maior percentual de conclusões quando comparados com cursos de áreas de Humanas e Biológicas, sendo que cursos das áreas de Biológicas possuem o menor percentual de conclusões dos três grupos.

No gráfico de setores, é importante limitar o número de categorias representadas para evitar a sobrecarga visual, não mais que 5 categorias. O uso de cores distintas e uma legenda clara para cada categoria facilita a identificação das partes do gráfico. É essencial que o tamanho de cada setor corresponda com precisão à proporção dos dados que representa, evitando assim distorções da informação.

## Gráfico de linha

Um gráfico de linhas exibe a evolução de uma ou várias variáveis numéricas. Os pontos de dados são conectados por segmentos de linha reta. É semelhante a um gráfico de dispersão, exceto que os pontos de medição são ordenados e unidos com segmentos de linha reta. Neste exemplo, foi utilizado o conjunto de dados *babynames* disponível no *Software R*, que contém informações sobre dados completos de nomes de bebês fornecidos pela Administração do Seguro Social dos Estados Unidos. Os dados incluem todos os nomes com pelo menos cinco usos. O conjunto de dados contém 1.924.665 linhas e 5 variáveis como ano, nome e sexo.

Figura 4. Exemplo de um gráfico de linha.



Fonte: Autor (2024).

A proposta de visualização desse gráfico é mostrar a popularidade de três nomes de bebês entre o período de 1880 e 2020, ou seja, com esse gráfico podemos perceber que o nome Helen foi o mais

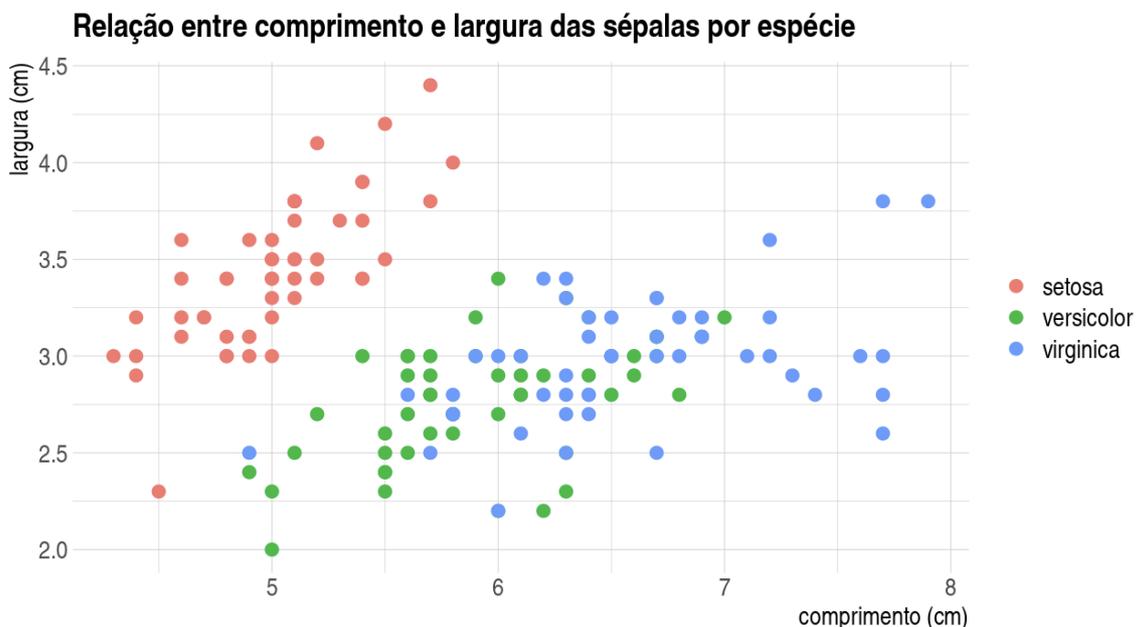
escolhido no século 19 até início do século 20, o nome Patrícia foi muito popular até meados de 1950 e o nome Ashley teve um pico de popularidade no início dos anos 1990.

No gráfico de linhas, é importante limitar o número de linhas representadas no gráfico, destacando apenas as séries de dados mais relevantes ou significativas. O uso de cores distintas para cada série de dados, garantem contraste e legibilidade. É importante que o intervalo de tempo ou variável no eixo x esteja uniformemente espaçado para evitar distorções na representação dos dados.

### Gráfico de dispersão

Um gráfico de dispersão exibe a relação entre duas variáveis numéricas. Para cada ponto de dados, o valor da sua primeira variável é representado no eixo x, o segundo no eixo y. Neste exemplo, foi utilizado o conjunto de dados iris disponível no *Software R*, que contém informações sobre medidas em centímetros das variáveis comprimento e largura da sépala e comprimento e largura da pétala, respectivamente, para 50 flores de cada uma das três espécies de íris (setosa, versicolor e virginica). O conjunto de dados contém 150 linhas e 5 variáveis.

Figura 5. Exemplo de um gráfico de dispersão.



Fonte: Autor (2024).

A proposta de visualização desse gráfico é mostrar a relação entre o comprimento e largura das sépalas das três espécies de íris (setosa, versicolor e virginica), ou seja, com esse gráfico podemos perceber que a espécie virginica é a que possui o maior comprimento e largura de sépalas, enquanto que setosa é a espécie com menor comprimento e largura de sépalas.

No gráfico de dispersão, se o seu conjunto de dados é grande, os pontos do gráfico tendem a se sobrepor, tornando o gráfico ilegível, a solução mais fácil é optar por reduzir o tamanho do ponto. A escolha de cores distintas ajuda a destacar grupos.

#### **2.2.4 Principais erros em gráficos estatísticos**

Segundo Coutinho et al. (2011), compreender e interpretar dados em gráficos é uma tarefa complexa que não surge de forma intuitiva. Arteaga (2011) aponta que muitos erros cometidos por professores ao criar ou interpretar gráficos estão relacionados à falta de compreensão de alguns conceitos estatísticos e como eles se relacionam com o contexto em que o gráfico foi feito.

No artigo de Arteaga et al. (2011), foi examinada a produção de gráficos por parte de futuros professores espanhóis em uma tarefa aberta. Eles identificaram que os erros nos gráficos dos professores originavam-se na falta de compreensão do significado dos números.

Mesmo entre aqueles que estão acostumados com representações estatísticas, como profissionais ou estudantes de Estatística, é comum observar interpretações diversas de gráficos (Monteiro, 2005; Queiroz, 2015; Queiroz et al., 2017). Isso acontece porque a leitura de um gráfico não é fixa, mas sim uma interação dinâmica entre o conhecimento matemático e estatístico de cada pessoa, juntamente com seus valores, crenças e percepções individuais, como defendido por Gal (2002).

Existem várias maneiras pelas quais os gráficos estatísticos podem ser manipulados ou apresentar erros, comprometendo a precisão e a clareza das informações. Estes incluem escalas desproporcionais, onde a relação entre os valores representados não é adequada; alturas incorretas em gráficos de colunas, barras e linhas, distorcendo a comparação entre categorias; início do eixo vertical diferente de zero, o que pode exagerar ou subestimar diferenças entre valores; somatório superior a 100% em gráficos de setores, distorcendo a representação proporcional de cada categoria; inversão do eixo em gráficos de linha, para criar a ilusão de um setor maior do que realmente é; e a união de duas categorias de respostas em um único setor em gráficos de setores, o que pode levar a uma representação enganosa das proporções.

Na Figura 6, o uso de escalas desproporcionais colocou um candidato a prefeito do Estado de São Paulo, no Brasil a frente dos demais candidatos. Ele se colocou a frente em um estudo que apontava os optantes por brancos/nulos como a maior preferência, além de se colocar em escala três vezes maior mediante seu concorrente que está com apenas sete pontos percentuais a menos que ele.

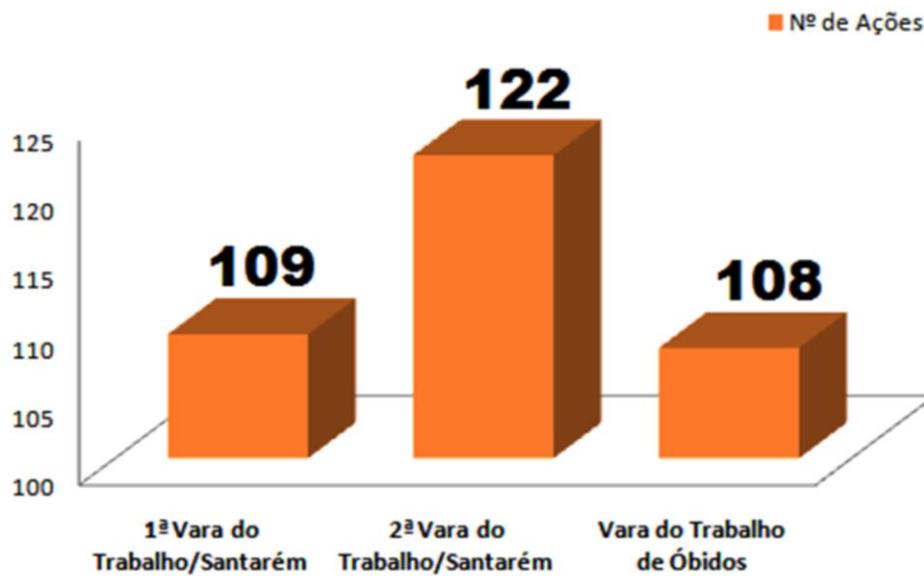
Figura 6. Gráfico de barras com escala desproporcional.



Fonte: Teixeira (2018).

Na Figura 7, além do uso de um gráfico em modelo 3D que dificulta a visualização correta das alturas das barras de cada grupo, o gráfico também não possui rótulos nos eixos x e y. Além disso, o autor do gráfico foi incapaz de avaliar corretamente as proporções das alturas de cada coluna em relação aos rótulos dos dados.

Figura 7. Gráfico de barras com alturas incorretas.

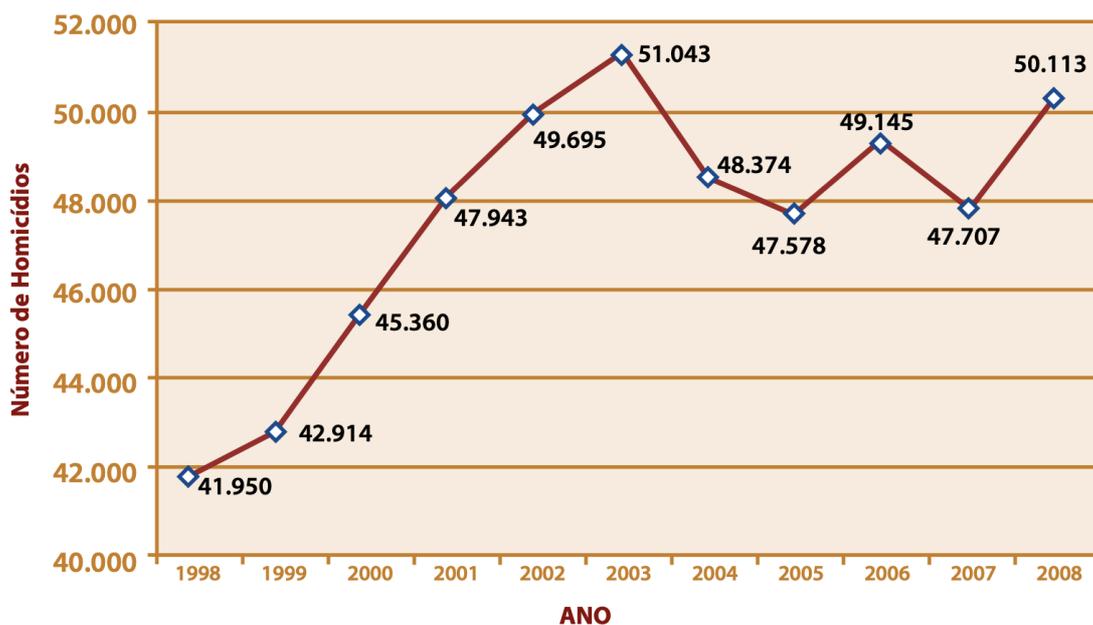


Fonte: Carneiro (2017).

Na Figura 8, a mudança no eixo y foi capaz de tornar uma informação que é estável em algo que estava crescendo de forma alarmante.

Figura 8. Gráfico de linhas com mudança no eixo y.

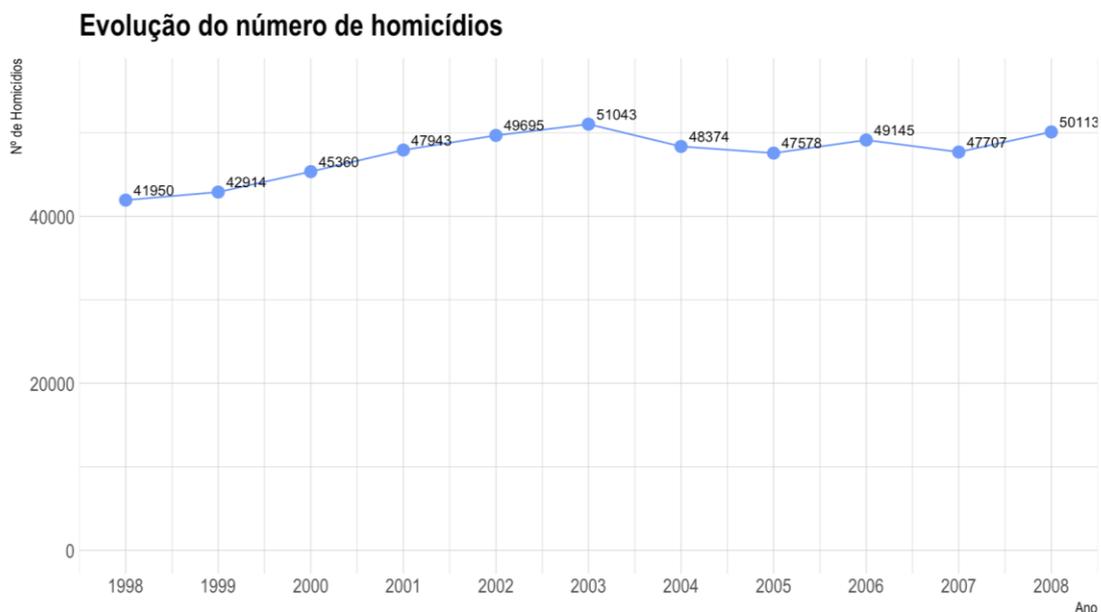
### Gráfico 3.1.1. Evolução do Número de Homicídios. Brasil, 1998/2008.



Fonte: Waiselfisz (2011).

Na Figura 9, podemos verificar o gráfico corrigido. Com o eixo y começando em zero, percebemos que, na verdade o número de homicídios não teve um aumento alarmante como demonstrado no gráfico anterior. Na verdade, esse valor se manteve estável ao longo do período analisado.

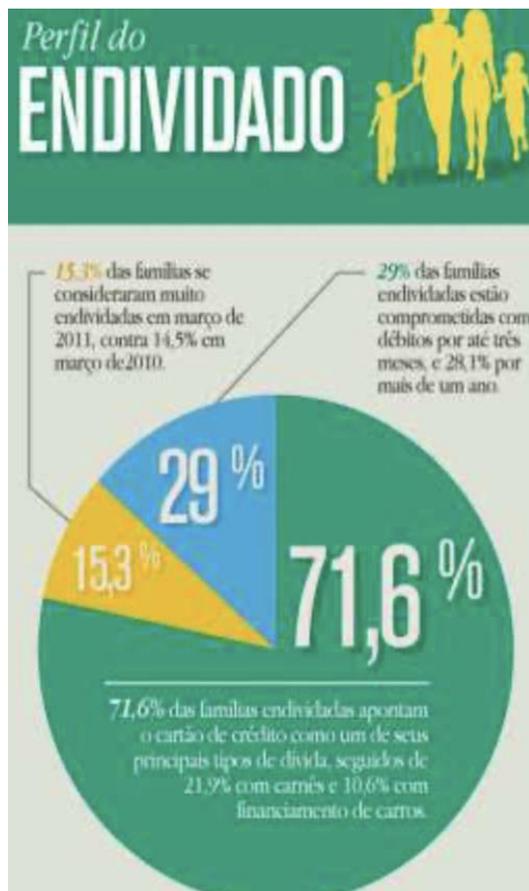
Figura 9. Gráfico de linhas corrigido.



Fonte: Autor (2024).

Na Figura 10, temos um problema de escala, já que 75% deveria ocupar  $\frac{3}{4}$  da área total, e, neste caso, 71,6% ocupa mais do que isso. Além disso, as porcentagens somadas ultrapassam 100%, provavelmente porque a pesquisa permitiu múltiplas respostas ou o gráfico foi criado com dados de perguntas diferentes.

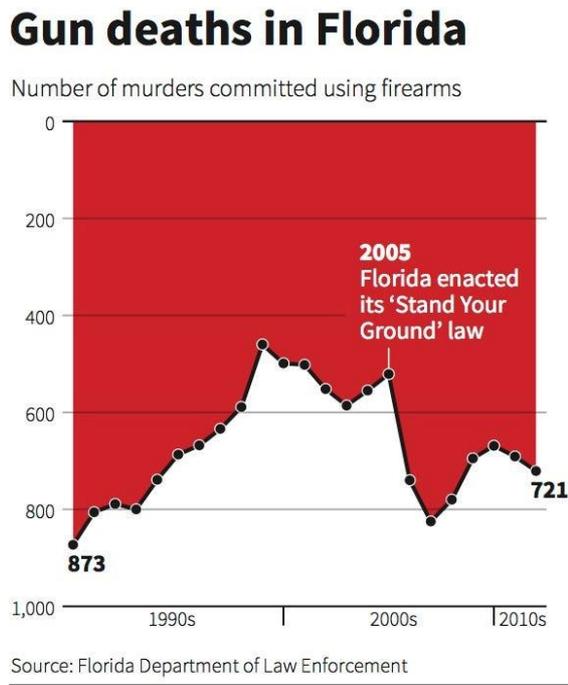
Figura 10. Gráfico de setores com erro.



Fonte: Agência Brasil (2011).

Na Figura 11, se olhar rapidamente talvez não seja perceptível, mas com um olhar mais atento notasse que o eixo y está invertido, então quanto mais para baixo for o valor no gráfico, maior o número.

Figura 11. Gráfico de área invertido.



Fonte: Engel (2014).

## 2.2.5 O uso de cores

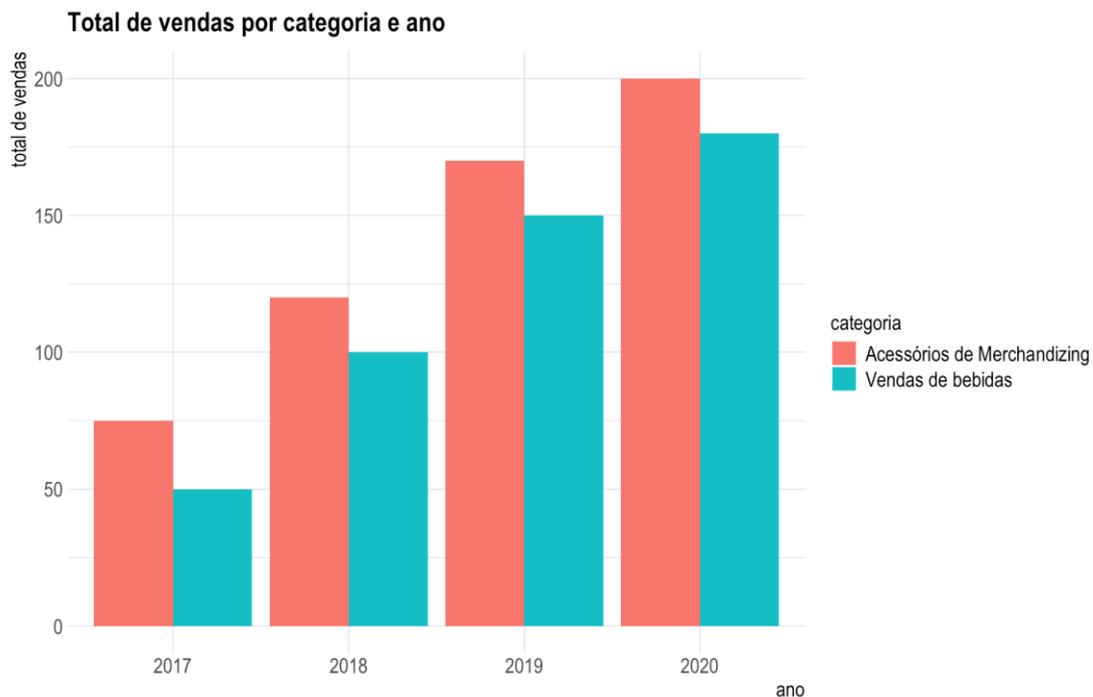
As cores são um componente fundamental na visualização de dados, desempenhando um papel crucial na clarificação dos elementos gráficos e na transmissão de informações. Diversos campos acadêmicos se dedicam ao estudo e ao *design* das cores na comunicação de dados e informações, envolvendo aspectos multidisciplinares que abarcam psicologia, *design*, experiência do usuário (UX) e ciência da computação.

Para Healy (2018) existe uma importância de selecionar esquemas de cores que não apenas sejam esteticamente agradáveis, mas também facilitem a interpretação dos dados pelos espectadores. Ele discute a utilização de esquemas de cores divergentes, sequenciais e qualitativos, e fornece diretrizes práticas sobre quando e como aplicar cada tipo de esquema de cores, considerando o contexto específico da visualização de dados.

A segmentação de dados é uma das formas mais básicas de se utilizar cores e refere-se à prática de dividir conjuntos de dados em categorias ou grupos distintos com base em critérios específicos. O uso eficaz de cores nesse contexto pode ajudar a diferenciar visualmente esses grupos, tornando mais fácil a identificação de padrões e tendências nos dados apresentados. Stone (1996) explora como escolher esquemas de cores apropriados e estratégias de codificação de cores para representar diferentes segmentos de dados de forma clara e intuitiva.

Na Figura 12, o uso de cores distingue as categorias de produtos facilitando a leitura do gráfico. Neste exemplo, foi utilizado um conjunto de dados fictício com a criação de listas que recebiam os valores com os nomes das categorias de produtos (*Acessórios de Merchandising* e *Vendas de bebidas*) e a quantidade vendida no período entre 2017 e 2020. Esses dados não correspondem a realidade do total de vendas desses produtos, os dados foram criados apenas para exemplificar o uso de segmentação de cores no gráfico.

Figura 12. Gráfico de barras com segmentação por cor.



Fonte: Autor (2024).

A representação gráfica da intensidade de um valor pode ser realizada de diversas formas, sendo uma delas através de um mapa. Este tipo de visualização gráfica proporciona uma visão geográfica dos dados, facilitando a compreensão dos padrões de distribuição.

A Figura 13, representa o número de substâncias psicoativas por país na América Latina, regiões com maior número dessas substâncias estão representadas por cores mais escuras, enquanto regiões com menor número possuem cores mais claras.

Figura 13. Mapa representando intensidade de cor.



Fonte: UNODC (2019).

Cores são elementos fundamentais para a visualização de dados, e o uso de ferramentas de visualizações dinâmicas e interativas, como o *Shiny*, potencializa ainda mais essa capacidade. Com o *Shiny*, é possível criar aplicativos *web* interativos que permitem aos usuários explorar os dados de forma personalizada. Essa abordagem não apenas enriquece a experiência do usuário, mas também proporciona uma compreensão mais profunda e significativa dos dados apresentados.

### 2.2.6 Software R, R Studio e Shiny

R é uma linguagem de programação e um ambiente estatístico amplamente utilizado para análise de dados e criação de gráficos. Sua origem remonta ao início dos anos 1990, quando foi concebida por Robert Gentleman e Ross Ihaka, professores da Universidade de Auckland, na Nova Zelândia. Derivada da linguagem S, a criação do R teve como propósito principal a elaboração de um *Software* acessível e gratuito, destinado ao ensino de Estatística aos estudantes da universidade (R Core Team, 2023).

Atualmente, o R é amplamente adotado pela comunidade estatística global devido à sua extensa coleção de funções e pacotes, que abrangem uma ampla variedade de técnicas estatísticas, ferramentas de análise e manipulação de dados.

Sua popularidade é atribuída, em parte, à sua natureza de *Software* livre, que permite acesso aberto e gratuito, além de sua compatibilidade multiplataforma e código aberto, garantindo flexibilidade e acessibilidade para uma ampla gama de usuários e contextos de aplicação.

O RStudio é um ambiente de desenvolvimento integrado de código aberto projetado especificamente para o R, com o propósito de simplificar o desenvolvimento de *scripts* e aplicativos. Além de fornecer uma interface amigável para escrever e executar código R, o RStudio também desempenha um papel significativo na criação e manutenção de novos pacotes. Por exemplo, pacotes amplamente utilizados como o *Shiny* e o *ggplot2* (Chang et al., 2024; Wickham, 2016) foram desenvolvidos e aprimorados com o suporte do RStudio Team (2023). Esses pacotes são importantes em diversas aplicações, incluindo a criação de visualizações interativas e a geração de gráficos estatísticos avançados.

O pacote *Shiny* é um *framework* de aplicações *web* para o R, que oferece uma estrutura *web* sofisticada e eficaz para a construção de aplicativos. Ele permite que os usuários transformem suas análises em aplicativos interativos sem a necessidade de conhecimento prévio em HTML, CSS ou JavaScript. (Chang et al., 2024).

A estrutura de um aplicativo *Shiny* é composta por dois elementos principais: o primeiro é responsável pela construção da interface do usuário (o *ui*), enquanto o segundo realiza as análises e gera os gráficos estatísticos (o *server*). Essas aplicações são interativas, o que significa que os usuários têm a capacidade de modificar as entradas de dados na interface de usuário, e as saídas são atualizadas automaticamente em resposta a essas modificações.

O aplicativo desenvolvido neste trabalho está disponível *online* no servidor do RStudio para aplicativos *Shiny*, conhecido como Shinyapps<sup>1</sup>. Essa plataforma permite que os alunos acessem o aplicativo de qualquer lugar e a qualquer momento, sem a necessidade de configurações adicionais.

### **2.2.7 Aplicabilidade do R Shiny no contexto do ensino de visualização de dados**

A necessidade de automatização de processos em diversas áreas do conhecimento experimentou um rápido crescimento ao longo dos anos, a área da educação não ficou imune a essa tendência, levando a uma maior busca para desenvolver soluções tecnológicas para facilitar o acesso e a comunicação entre professores e alunos. Carrillo & Maldonado (2009) destacam a importância de tecnologias interativas no processo educativo, destacando as vantagens de seu uso no processo de ensino-aprendizagem do ponto de vista dos próprios professores.

---

<sup>1</sup> <https://rachderossi.shinyapps.io/DataViz/>

No que diz respeito às tecnologias interativas e seu impacto no processo de ensino-aprendizagem de Estatística, pesquisas como as conduzidas por Lunsford et al. (2006) e Hagtvedt et al. (2007) evidenciam melhorias significativas no desempenho dos alunos que incorporam soluções tecnológicas durante sua aprendizagem estatística.

Segundo Chance e Rossman (2006), no ensino de Estatística, é importante demonstrar o uso da tecnologia como uma ferramenta para resolver problemas e aproveitar sua capacidade de automatizar cálculos e gráficos, além de facilitar a exploração de conceitos estatísticos por meio de ambientes visuais dinâmicos e interativos.

Sendo assim, o uso do *Shiny* permite aos estudantes a capacidade de manipular visualizações de dados em tempo real, experimentar diferentes abordagens de representação gráfica e interagir diretamente com os resultados. Tal abordagem não apenas torna o processo de aprendizado mais dinâmico e acessível, mas também capacita os alunos a adquirirem habilidades para uma melhor análise e interpretação de dados. A integração do *Shiny* no ensino de Estatística proporciona, portanto, uma oportunidade para promover uma compreensão mais profunda e contextualizada dos conceitos estatísticos, preparando os estudantes para enfrentarem os desafios da análise de dados no mundo real.

### 3 Metodologia

#### 3.1 Sumarização das etapas de trabalho realizado

Figura 14. Etapas do trabalho realizadas.



Fonte: Autor (2024).

A Figura 14 resume as diferentes etapas do trabalho descrito nesta dissertação.

Após uma breve revisão bibliográfica acerca da análise de dados e das dificuldades encontradas pelo seu uso nas ciências humanas, ilustrando as principais formas de apresentação dos dados e os erros cometidos, muitas vezes ocorrendo por uma falta de letramento estatístico, deu-se início ao desenvolvimento do aplicativo.

A segunda etapa, inicia o desenvolvimento da aplicação *web* com a criação do *wireframe* usando o editor gráfico de projetos de *design*, Figma. O uso desta ferramenta facilitou a visualização de um primeiro esboço de como o aplicativo haveria de se parecer, nesta etapa também foi utilizado o conhecimento adquirido nas disciplinas de *design* deste mestrado.

Na terceira etapa, o desenvolvimento do aplicativo em *Shiny*, se mostrou um tanto quanto desafiadora, pois apesar de não necessitar de um vasto conhecimento prévio em HTML, CSS ou JavaScript me levou a um embate de como repassar meu conhecimento sobre visualização de dados sem que houvesse o uso demasiado de textos e de gráficos de difícil entendimento prático.

Na quarta etapa, houve a elaboração de um breve questionário aplicado em uma atividade na aula de Análise e Visualização de Dados aos estudantes da edição 2023/2024 deste mestrado, resultando em um *feedback* construtivo de melhorias no aplicativo, tanto na parte visual quanto na parte dos códigos compartilhados para a reprodução gráfica dos exemplos disponíveis no aplicativo.

Por fim, na última etapa, as respostas obtidas por meio do questionário aplicado serão discutidas e avaliadas.

### 3.2 Criação do wireframe

O processo de criação do aplicativo em *Shiny* se iniciou com um *wireframe* criado a partir da ferramenta Figma versão 116.17.13. Por meio desta ferramenta é possível desenvolver a interface do aplicativo de forma rápida, tornando a visualização do projeto algo fácil.

Desde o início procurei simplificar a interface do aplicativo para que a experiência do usuário fosse a melhor possível, considerando uma boa usabilidade, com cores agradáveis e que fosse algo de uso intuitivo. A disposição do conteúdo na tela segue uma hierarquia visual, com as informações disponíveis da esquerda para direita e botões de ação na parte superior, o que cria um *design* mais lógico, compreensível e que maximiza a experiência do usuário.

Na Figura 15, temos o primeiro esboço da tela inicial do aplicativo. A disposição do conteúdo foi dividida em três “caixas”. Sendo a primeira uma breve explicação do uso do aplicativo, seguido por exemplos ilustrativos do tipo de dados e na última caixa a opção de seleção do tipo de dado que o usuário tem interesse, podendo ser: Numérico, Categórico ou Numérico & Categórico.

Figura 15. Esboço da tela inicial do aplicativo.

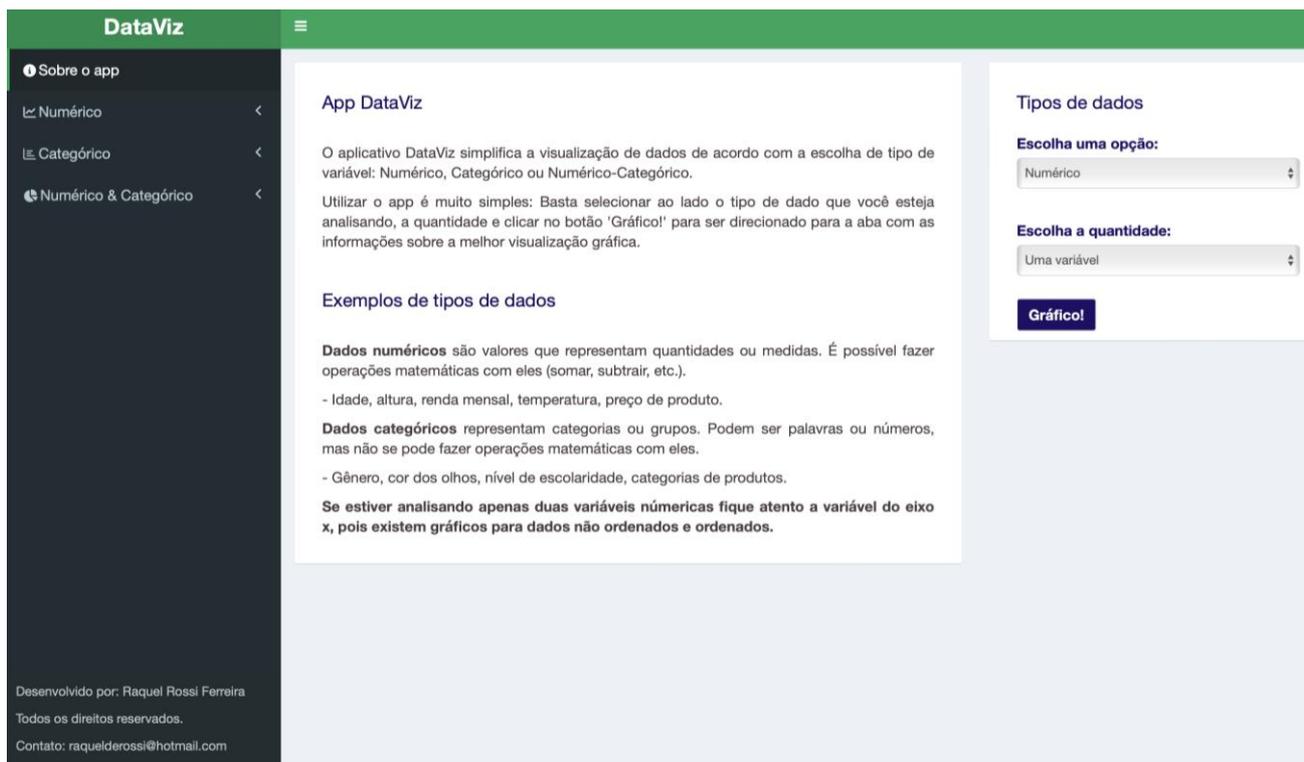


Fonte: Autor (2024).

Na Figura 16, temos a primeira versão da tela inicial, com algumas modificações como o alinhamento do texto, a redução do número de “caixas” para duas, comportando toda a informação textual em apenas uma delas e adicionando mais uma opção de escolha para melhorar a experiência do usuário.

Sendo assim, primeiro o usuário escolherá o tipo de dado com que possui interesse em realizar a visualização gráfica: Numérico, Categórico ou Numérico & Categórico e a seguir irá escolher a quantidade de variáveis a serem analisadas: Uma variável, duas variáveis ou uma combinação de variáveis (caso os dados sejam Numérico & Categórico).

Figura 16. Primeira versão da tela inicial do aplicativo.

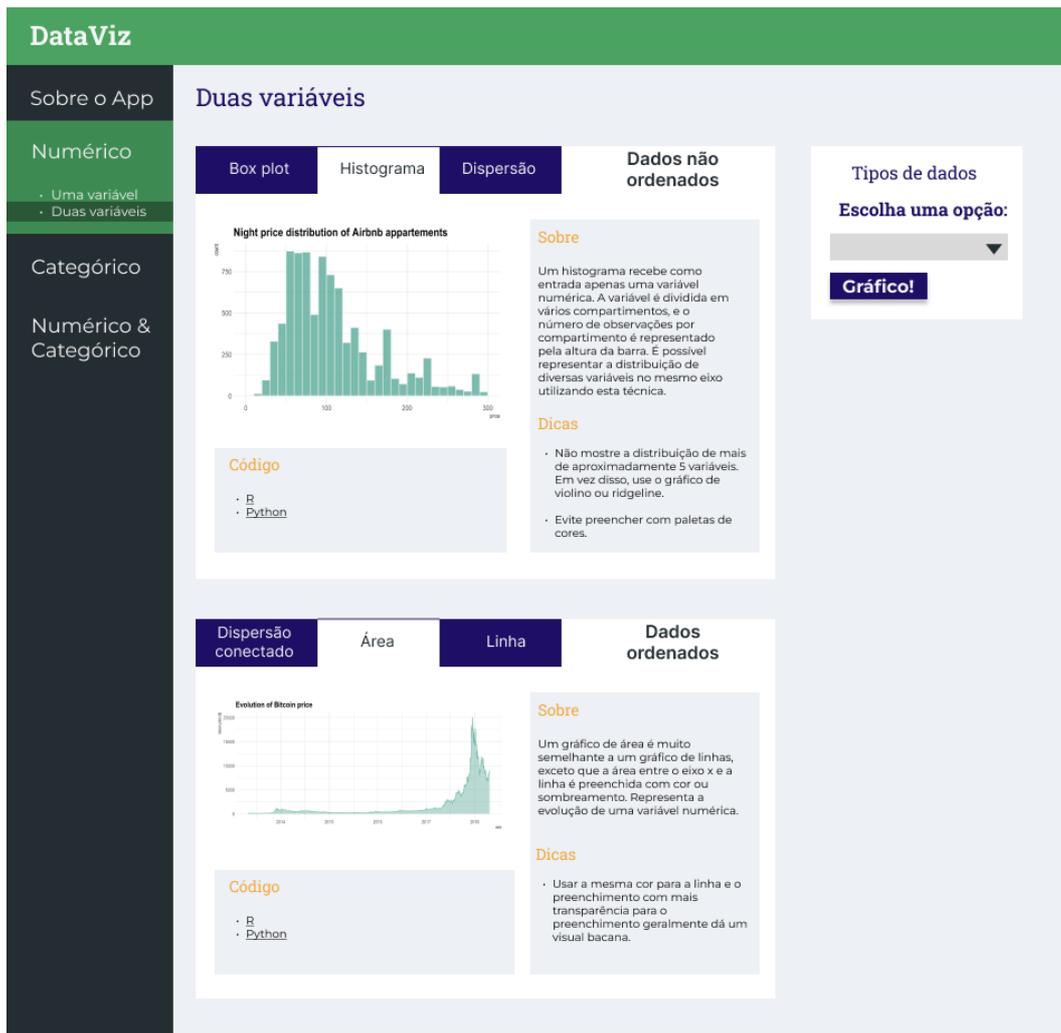


Fonte: Autor (2024).

Na Figura 17, temos o esboço de uma das abas do aplicativo. A proposta é que após o usuário clicar no botão “Gráfico!” será direcionado para a aba correspondente com opções de visualizações gráficas mais adequadas para os dados a serem analisados. Nesta aba o usuário poderá escolher entre visualizações para dados numéricos ordenados e não ordenados, contando com três opções para cada tipo.

Ao clicar nas sub abas com os nomes dos gráficos terá informações como: Uma breve definição daquele gráfico e para que tipo de dados ele se adapta melhor, dicas de como melhorar a visualização, como o uso de cores e limitação de comparação de grupos em uma mesma visualização, além da disponibilidade de reprodução do gráfico a partir de códigos na linguagem R e Python.

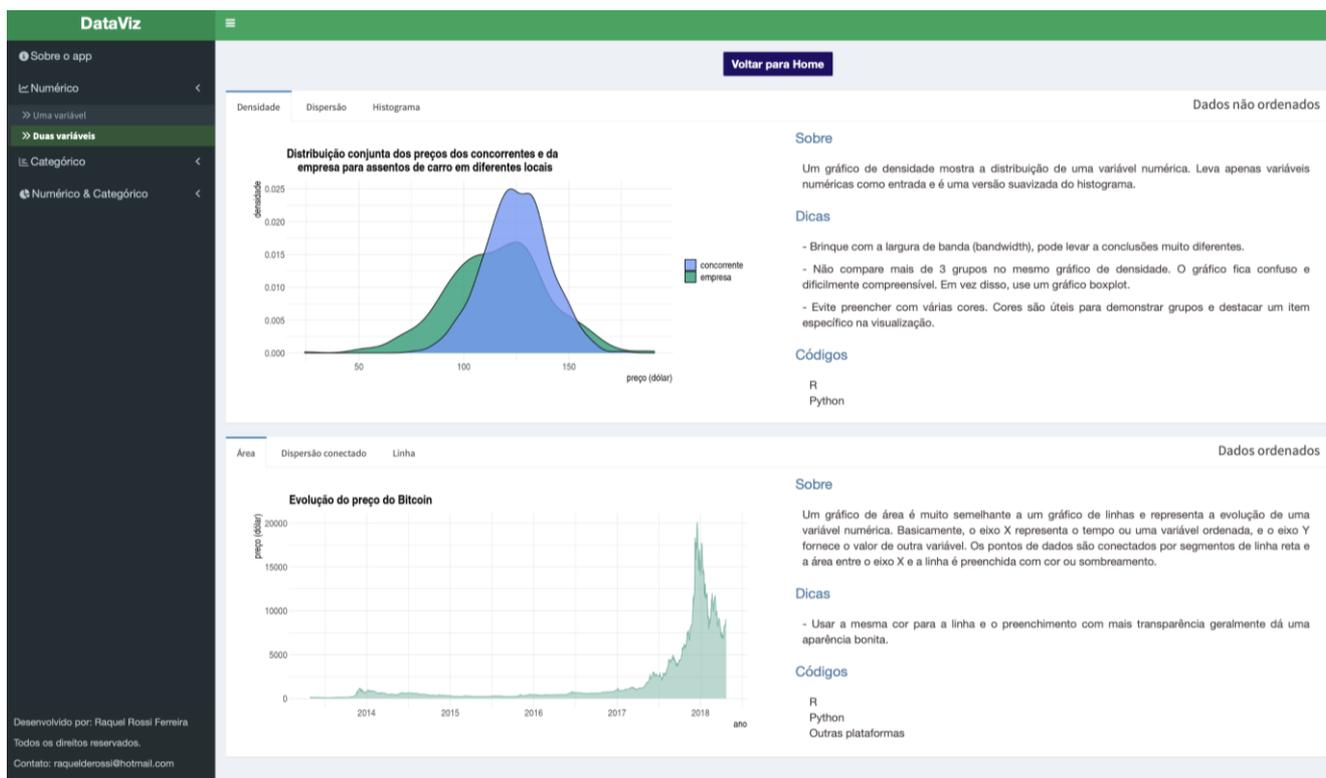
Figura 17. Esboço de uma das abas do aplicativo.



Fonte: Autor (2024).

Na Figura 18, temos a primeira versão desta aba com algumas modificações, foi dado um maior destaque ao gráfico, colocando toda a parte textual alinhada à direita, a caixa que continha informação sobre a opção de seleção do tipo de dado foi substituída por um botão alinhado ao centro na parte superior que ao ser clicado direciona o usuário para a tela inicial e na parte dos códigos, ao clicar na opção desejada abre um *link* para um repositório *online* do GitHub com o passo a passo de como reproduzir o gráfico escolhido.

Figura 18. Primeira versão de uma das abas do aplicativo.



Fonte: Autor (2024).

### 3.3 Desenvolvimento do aplicativo

O aplicativo foi construído usando o *Software* R versão 4.3.1, RStudio versão 2023.12.1 e o pacote *Shiny* versão 1.8.0 (R Core Team, 2023). Os gráficos do aplicativo foram construídos utilizando o pacote de visualização gráfica *ggplot2* versão 3.5.0. Além disso, foram utilizados os pacotes *datasets*, *babynames* e *ISLR* que contêm os conjuntos de dados utilizados para criação dos gráficos.

O uso do *Shiny* permitiu o desenvolvimento do aplicativo com aparência de *dashboard*, o que facilita a navegação entre as abas pelo usuário e a manutenção do código pelo desenvolvedor, pois o *Shiny* permite a criação de funções e a nomeação de variáveis.

O código completo do aplicativo pode ser encontrado disponível *online* no repositório do GitHub<sup>2</sup>, plataforma de hospedagem e compartilhamento de códigos e projetos de programação *online* (GitHub, 2024). As visualizações gráficas que fazem parte do aplicativo foram escolhidas pela sua facilidade de interpretação e provável familiaridade por parte dos usuários.

<sup>2</sup> <https://github.com/rachderossi/App-DataViz>

Sendo assim, as visualizações gráficas escolhidas foram: Área, barras, barras agrupadas, *boxplot*, densidade, dispersão, dispersão conectado, histograma, linha, mapa de calor, nuvem de palavras e pirulito.

Foram utilizadas funções para divisão do conteúdo na tela, como a função “*tabItems*” que serve para criar a navegação por abas, a função “*fluidRow*” é utilizada dentro de “*tabItems*” para criar layouts fluidos e responsivos, esse layout ajusta automaticamente os elementos do aplicativo com base no tamanho da janela do navegador. Para customizar o aplicativo o uso da função “*tags\$div*” foi essencial, pois permitiu criar um elemento div onde foram definidas as classes CSS, como estilos e atributos.

O código com as funções citadas anteriormente, é apresentado na sequência.

```
tabItems(  
  tabItem(tabName = "home",  
    fluidRow(tags$div(class = "custom-box",  
      box(width=8, solidHeader = TRUE,  
      tags$div(class = "title-home", "App DataViz"),  
      br(),  
      tags$div(class = "text-home",  
        style="text-align: justify",  
        tags$p("O aplicativo DataViz simplifica a  
visualização de dados de acordo com a escolha do tipo de variável:  
Numérico, Categórico ou Numérico-Categórico."),  
        tags$p("Utilizar o app é muito simples: Basta  
selecionar ao lado o tipo de dado que você esteja analisando, a  
quantidade e clicar no botão 'Gráfico!' para ser direcionado para a  
aba com as informações sobre a melhor visualização gráfica."))),
```

Na tela inicial o usuário começa por definir os tipos de dados e a quantidade de variáveis que têm interesse em analisar. Ao clicar no botão “Gráfico!” será direcionado para as opções de visualizações gráficas que melhor se encaixam na seleção feita anteriormente. Para o botão foi utilizado a função “*actionButton*” que é usada para criar um botão de ação interativo.

O código com a funções citada anteriormente, é apresentado na sequência. Onde é utilizado um identificador único para o botão e várias propriedades de estilo inline em CSS.

```
actionButton("dashboard_switchTabs", "Gráfico!",  
  style="background-color: #1F0E65;  
  color: #ffff;  
  font-family: 'Montserrat', 'Helvetica Neue',  
'Helvetica', 'Arial', 'sans-serif';  
  font-weight: bold;  
  font-size: 16px;  
  margin-left: 20px;"))),
```

O código utilizado na construção da seleção das abas, é apresentado na sequência. Basicamente o código utiliza a função “*switch*” para selecionar a aba apropriada com base nos valores das variáveis de entrada e, em seguida, atualiza a tabulação selecionada usando a função “*updateTabsetPanel*”.

```
observeEvent(input$dashboard_switchTabs, {
  selected_tab <- switch(input$dashboard_variavel1,
    "Numérico" = switch(input$dashboard_variavel2,
      "Uma variável" = "subMenu1",
      "Duas variáveis" = "subMenu2"
    ),
    "Categórico" = switch(input$dashboard_variavel2,
      "Uma variável" = "subMenu3",
      "Duas variáveis" = "subMenu4"
    ),
    "Numérico & Categórico" = switch(input$dashboard_variavel2,
      "Uma variável num. e uma variável cat." = "subMenu5",
      "Uma variável num. e muitas categóricas" = "subMenu6",
      "Uma variável cat. e muitas numéricas" = "subMenu7"
    )
  )
  updateTabsetPanel(session, "tabs", selected = selected_tab)
})
```

Nas abas onde se encontram as informações sobre os gráficos há sub abas onde o usuário pode navegar e escolher a visualização que mais lhe agrada dentre as disponíveis. Na parte inferior de cada sub aba está disponível em um repositório *online* do GitHub<sup>3</sup> códigos na linguagem R e Python para a reprodução do gráfico. Cada gráfico foi desenvolvido com a sua própria função de saída, facilitando assim a manutenção do código.

O código utilizado na construção de um gráfico é apresentado na sequência. Basicamente o código utiliza a função “*renderPlot*” para gerar um gráfico de barras. No exemplo, ele conta as ocorrências de cada classe de veículo no conjunto de dados mpg, reordena as classes com base em suas frequências e, em seguida, utiliza essas informações para criar o gráfico de barras.

```
output$barPlot <- renderPlot({
  mpg %>%
  count(class) %>%
  mutate(class = reorder(class, n)) %>%
  ggplot(aes(x = class, y = n)) +
  geom_bar(fill = "#6E9AF8", stat = "identity") +
  ggtitle("Distribuição das classes de veículo") +
  xlab("classe") +
  ylab("frequência") +
```

---

<sup>3</sup> <https://github.com/rachderossi/DataViz>

```
theme_ipsum()  
})
```

Após o desenvolvimento do aplicativo, um breve questionário foi elaborado para avaliar sua utilidade e eficácia.

### 3.4 Questionário

Com a finalização do desenvolvimento do aplicativo, foi criado um breve questionário de avaliação (ver Anexo 1) para avaliar a usabilidade e eficácia do mesmo. Esse questionário foi construído na plataforma *Google Forms*, uma plataforma de acesso grátis da Google, de uso e manutenção fácil, que permite extrair dados de forma rápida. A versão final do questionário é constituída por 10 questões, com as seguintes características:

- Texto introdutório que explica o objetivo, o tempo de duração, e a confidencialidade do mesmo;
- Questões de múltipla escolha - Questões: 1, 2, 3, 4, 5, 6, 7, 8;
- Questões de resposta curta - Questões: 9 e 10.

#### 3.4.1 Regras adotadas

O referencial utilizado para orientar a elaboração deste questionário foi o livro *Understanding Your Users: A Practical Guide to User Requirements Methods, Tools, and Techniques* (Courage & Baxter, 2005), especialmente o capítulo 8, do qual foram extraídas algumas regras fundamentais para a construção de questionários *online*:

- Informar no início o propósito do estudo;
- Informar o tempo de duração de preenchimento;
- Dar instruções de preenchimento;
- Indicar a confidencialidade ou anonimato do mesmo;
- Manter o mais curto possível;
- Reduzir o número de respostas abertas;
- Ter atenção ao formato das questões e às palavras utilizadas;
- Estruturar o questionário de forma lógica e sequencial.

### 3.4.2 Aplicação do questionário

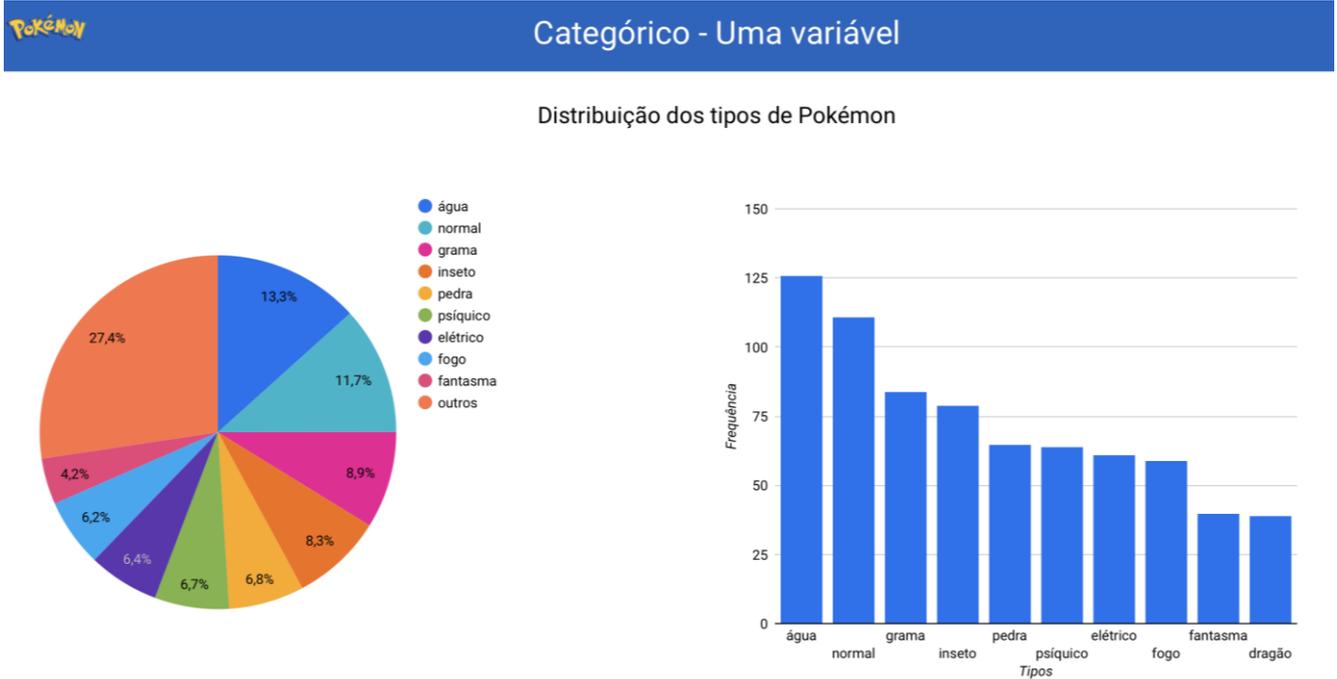
O aplicativo foi apresentado aos alunos durante uma aula da disciplina de Análise e Visualização de Dados do mestrado de Humanidades Digitais, o *link* do aplicativo foi divulgado na plataforma *Blackboard* juntamente com um conjunto de dados fictícios sobre Pokémons para que os alunos testassem as funcionalidades da aplicação e posteriormente respondessem ao questionário.

Na aula explanatória sobre o aplicativo participaram 6 alunos do primeiro ano do mestrado de Humanidades Digitais, durante a aula os alunos tiveram uma breve explicação do objetivo da aplicação e como utilizá-la, e, seguidamente, foram convidados a tentarem reproduzir três visualizações gráficas a partir do conjunto de dados Pokémon, compartilhada via *link* na *Blackboard*. As propostas de visualizações foram as seguintes:

- Gráfico que mostre os tipos de Pokémons;
- Gráfico que mostre os tipos de Pokémons em cada geração;
- Gráfico que relacione o ataque e defesa de cada Pokémon por geração.

Esta atividade durou em torno de 15 minutos e posteriormente houve uma apresentação de um *dashboard*, conforme a Figura 19, ilustrando uma boa e má visualização gráfica e os alunos deveriam indicar qual delas era a correta e o motivo.

Figura 19. Dashboard ilustrando uma boa e má visualização gráfica.



Fonte: Autor (2024).

Por fim, foi solicitado aos alunos que respondessem ao questionário. A seguir suas respostas e contribuições de melhoria serão analisadas com maior detalhe.

## 4 Análise e Discussão dos Resultados

Durante a aula os alunos testaram o aplicativo, sendo que dos 12 alunos matriculados na disciplina 6 responderam as questões através do questionário de avaliação.

Na questão que avalia a usabilidade do aplicativo, 5 alunos (83,30%) disseram que a usabilidade do aplicativo é muito boa e apenas 1 aluno (16,70%) disse achar que o nível de usabilidade do aplicativo é regular.

Sobre compreender o objetivo do aplicativo, 2 alunos (33,30%) disseram compreender completamente o objetivo do aplicativo, enquanto 2 alunos disseram ter compreendido bem e os outros 2 compreenderam razoavelmente.

Observou-se também que 5 alunos (83,30%) disseram que o aplicativo ajudou muito a escolher a visualização gráfica mais adequada e 4 alunos (66,67%) recomendariam este aplicativo para algum colega que futuramente fosse cursar essa disciplina.

Quando questionados sobre o quanto acham que aplicativos interativos, como o aplicativo em questão, contribuem para o aprendizado em disciplinas de análise de dados, 4 alunos (66,70%) disseram que contribui muito e 2 alunos (33,30%) disseram que contribui.

Observou-se também que 4 alunos (66,70%) consideram as recomendações de gráficos fornecidas pelo aplicativo bastante compreensíveis, enquanto 5 alunos (83,30%) consideraram a explicação do conteúdo do aplicativo satisfatória.

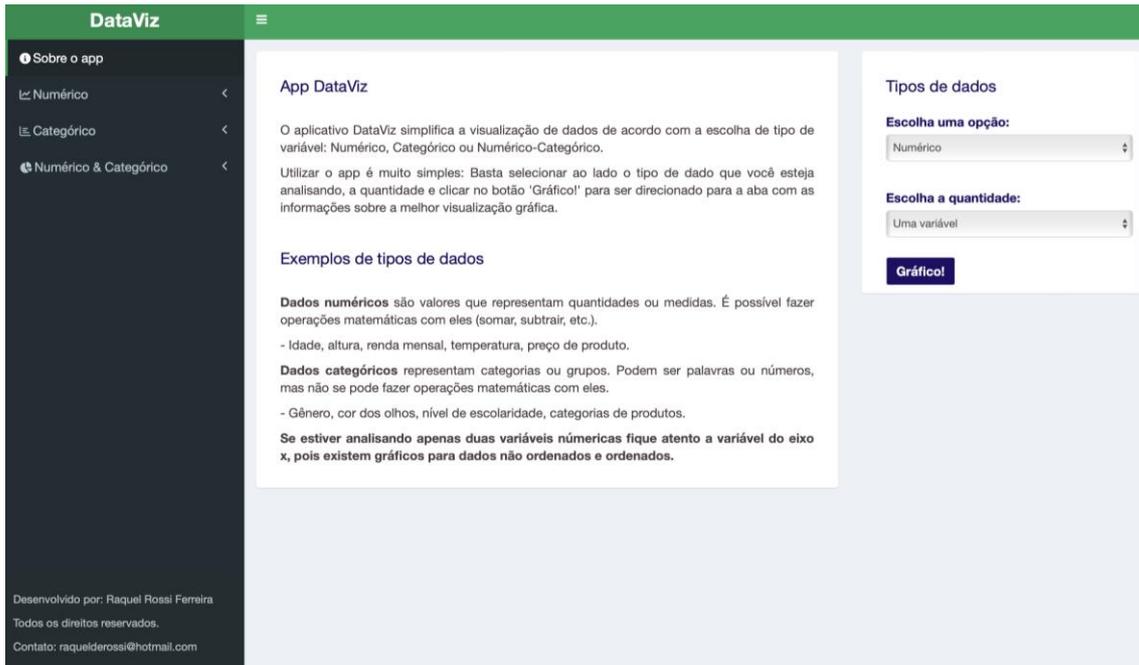
Alguns pontos sinalizados pelos alunos como características que mais gostaram no aplicativo foram a interatividade e facilidade de uso, a possibilidade de conhecer diferentes gráficos e dicas relacionados com a sua utilização, além de saber quais os erros mais comuns na criação de um gráfico e como evitá-los.

Em contrapartida, alguns dos pontos que menos gostaram no aplicativo foi a falta de uma instrução mais detalhada de como reproduzir o gráfico utilizando os códigos em R e Python, pois muitos dos alunos não possuem conhecimentos suficientes nas linguagens de programação citadas. Foram sugeridas alterações na aparência do aplicativo, como justificar o texto e colocar em negrito algumas palavras importantes para dar destaque. Além disso, foi sugerido uma maior variedade de opções gráficas.

A partir das respostas dos alunos, algumas das alterações sugeridas foram feitas. A inclusão de comentários nos códigos em R e Python está mais detalhada, foi criada mais uma opção na seção de códigos nomeada de "Outras plataformas" visando usuários que preferem utilizar o *Looker Studio* ou *Microsoft Excel* para a criação dos gráficos, a aparência do aplicativo foi melhorada, justificando o texto e colocando em negrito palavras importantes.

Nas figuras a seguir é possível visualizar as alterações feitas de acordo com a solicitação dos alunos. Na Figura 20, temos a versão final da tela inicial com o texto justificado e palavras importantes em negrito.

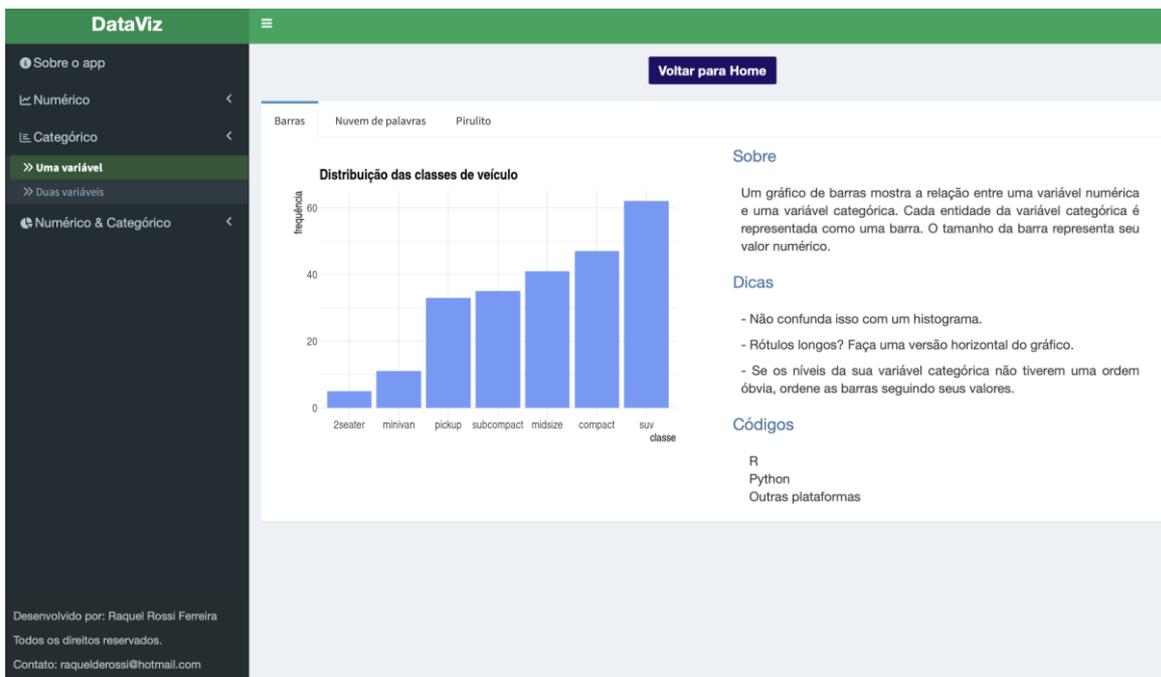
Figura 20. Versão final da tela inicial do aplicativo.



Fonte: Autor (2024).

Na Figura 21, temos a versão final de uma das abas, onde foi criada mais uma opção na seção de códigos.

Figura 21. Versão final de uma das abas do aplicativo.



Fonte: Autor (2024).

Na Figura 22 e na Figura 23, respectivamente, temos inclusão de comentários nos códigos em R e Python, tornando as etapas de reprodução dos gráficos mais simples e detalhadas.

Figura 22. Comentários nos códigos em R.

```
1 library(dplyr)
2 library(ggplot2)
3 library(hrbrthemes)
4 library(readr)
5
6 # Substitua "caminho/para/seu/arquivo.csv" pelo caminho do arquivo CSV em sua máquina
7 # mpg <- read.csv("caminho/para/seu/arquivo.csv", header = TRUE)
8
9 # Criar o gráfico de barras
10 mpg %>%
11   count(class) %>% # Substitua class pela sua variável categórica
12   # Caso exista, substitua n pela sua variável de frequência
13   mutate(class = reorder(class, n)) %>% # Ordenando a variável pela frequência
14   ggplot(aes(x = class, y = n)) +
15   geom_bar(fill = "#6E9AF8", stat = "identity") +
16   ggtitle("Distribuição das classes de veículo") +
17   xlab("classe") +
18   ylab("frequência") +
19   theme_ipsum()
```

Fonte: Autor (2024).

Figura 23. Comentários nos códigos em Python.

```
import pandas as pd
import matplotlib.pyplot as plt

# Substitua "caminho/para/seu/arquivo.csv" pelo caminho do arquivo CSV em sua máquina
# data = pd.read_csv("caminho/para/seu/arquivo.csv")

# A linha 5 substitui essa, pois aqui estou criando dados fictícios
# Define um dicionário com os dados das classes de veículo, pois não estou usando nenhum arquivo
data = {
    'class': ['compact', 'midsize', 'suv', 'midsize', 'compact', 'suv', 'suv', 'compact'],
}

# Caso você já tenha a variável de frequência calculada pode pular para a linha 26
# Cria um DataFrame chamado 'mpg' com os dados fornecidos
mpg = pd.DataFrame(data)

# Conta o número de frequência de cada classe de veículo e armazena em 'class_counts'
class_counts = mpg['class'].value_counts().reset_index()

# Renomeia as colunas para 'class' e 'count'
class_counts.columns = ['class', 'count']

# Ordena as contagens pela frequência
class_counts = class_counts.sort_values(by='count')
# Se a sua variável frequência já estiver calculada:
# data = data.sort_values(by='frequencia')

# Criar o gráfico de barras
plt.figure(figsize=(10, 6))
# class é a variável x e count é a variável y
plt.bar(class_counts['class'], class_counts['count'], color='#6E9AF8')
plt.title('Distribuição das classes de veículo')
plt.xlabel('Classe')
plt.ylabel('Frequência')
plt.xticks(rotation=45)
plt.show()
```

Fonte: Autor (2024).

## 5 Conclusões

O uso de tecnologias na educação veio para facilitar o processo de aprendizagem. Quando aplicadas como um complemento do conteúdo visto em aula, facilitam o processo de ensino-aprendizagem e ajudam a entender os conceitos básicos da disciplina.

Pensando nisso, este trabalho desenvolveu um aplicativo interativo como objeto de aprendizagem. Programado na linguagem R utilizando o pacote *Shiny*. Inicialmente direcionado à disciplina de Análise e Visualização de Dados do mestrado de Humanidades Digitais da Universidade do Minho, para complementar o conteúdo das aulas. O aplicativo aborda uma área específica da Estatística Descritiva que é a criação de gráficos.

O aplicativo desenvolvido tem o objetivo de auxiliar na tomada de decisão para a escolha da visualização gráfica que melhor se adapta aos dados analisados pelos alunos. Alguns alunos da disciplina testaram e avaliaram o aplicativo construído. Além disso, mais da metade dos respondentes afirmaram que acreditam que aplicativos interativos contribuem para o processo de aprendizado em disciplinas de análises de dados, o que confirma a importância do uso de tecnologias interativas no processo de ensino-aprendizagem.

Com relação ao questionário, houve um baixo número de respostas, mesmo sendo solicitado aos alunos que comparecessem a aula de explicação do aplicativo e a importância de testar e avaliar a aplicação, pois o mesmo foi pensando para contribuir para o conteúdo da disciplina.

Este trabalho desenvolveu o aplicativo pensando em complementar o material da disciplina de Análise e Visualização de Dados do mestrado de Humanidades Digitais da Universidade do Minho, porém o aplicativo é de livre acesso à comunidade e pode ser utilizado em qualquer disciplina onde os conteúdos contidos no aplicativo sejam abordados, sendo da Universidade do Minho ou de outra instituição.

Pensando em trabalhos futuros, há a possibilidade de melhorar ainda mais este aplicativo com a inclusão de um campo de texto onde o usuário possa escrever a hipótese de pesquisa que tenha interesse e o aplicativo gera um gráfico que faça mais sentido para a hipótese de pesquisa criada, tornando o processo de aprendizagem mais interativo e interessante.

## BIBLIOGRAFIA

Agência Brasil. (2011, abril 28). *Perfil do endividado*. gov.br.

<http://www.brasil.gov.br/economia-c-emprego/2011/08/perfil-do-endividado/perfi-do-endividado>

Arteaga, P. (2011). *Evaluación de conocimientos sobre gráficos estadísticos y conocimientos didácticos de futuros profesores* [Tese de Doutorado, Departamento de Didáctica da Matemática, Universidade de Granada].

<http://www.ugr.es/~batanero/pages/ARTICULOS/arteaga.pdf>

Barroso, J. C. D. G. de A. (2010). *Ensino da estatística em cursos de graduação das áreas das ciências humanas e sociais: desafios e perspectivas*. PUC Minas.

Batanero, C., Díaz, C., Contreras, J. M., & Arteaga, P. (2011). Enseñanza de la estadística a través de proyectos.

In C. Batanero & C. Díaz (Eds.), *Estadística con proyectos* (pp. 9-46). Universidad de Granada.

<https://www.ugr.es/~batanero/pages/ARTICULOS/Libroproyectos.pdf>

Carneiro, J. (2017, abril 9). *Número de Ações movidas contra a empresa Aqua Empreendimentos nas cidades de Santarém e Óbidos, em 2017*. Blog do Jeso. <http://www.jesocarneiro.com.br/negocios/aqua-31-andares-construtora-tem-quase-340-aco-es-na-justica-do-trabalho.html>

Carrillo, J. A. O., & Maldonado, Á. O. (2009). La comunicación didáctica digital en el espacio europeo de educación superior. ICONO 14, *Revista de comunicación y tecnologías emergentes*, 7(3), 64-83.

<https://doi.org/10.7195/ri14.v7i3.303>

Carver, R., et al. (2016). *GAISE College Report ASA Revision Committee, Guidelines for assessment and instruction in statistics education college report 2016*. <http://www.amstat.org/education/gaise>

Cazorla, I. M. (2002). *A relação entre a habilidade viso-pictórica e o domínio dos conceitos estatísticos na leitura de gráficos* [Tese de doutorado, Universidade Estadual de Campinas, Faculdade de Educação].

[https://www.psiem.fe.unicamp.br/pf-psiem/cazorla\\_irenemauroicio\\_d.pdf](https://www.psiem.fe.unicamp.br/pf-psiem/cazorla_irenemauroicio_d.pdf)

Cazorla, I. M. (2004). *Estatística ao alcance de todos*. In Encontro Nacional de Educação Matemática. UFPE.

<http://www.sbem.com.br/files/viii/pdf/12/MC11915634806.pdf>

Cazorla, I. M. (2015). *O ensino de estatística no Brasil*. Sociedade Brasileira de Educação Matemática.  
[http://www.sbem.com.br/gt\\_12/arquivos/cazorla.htm](http://www.sbem.com.br/gt_12/arquivos/cazorla.htm)

Chance, B., & Rossman, A. (2006). *Using simulation to teach and learn statistics*. In Proceedings of the Seventh International Conference on Teaching Statistics (pp. 1-6). Netherlands: International Statistical Institute.  
[https://iase-web.org/documents/papers/icots7/7E1\\_CHAN.pdf?1402524965](https://iase-web.org/documents/papers/icots7/7E1_CHAN.pdf?1402524965)

Chang, W., Cheng, J., Allaire, J., Sievert, C., Schloerke, B., Xie, Y., Allen, J., McPherson, J., Dipert, A., & Borges, B. (2024). *shiny: Web Application Framework for R*. R package version 1.8.1.9001.  
<https://github.com/rstudio/shiny>

Costa, J. D., Santos, W. L., Silva, J. S., & Alves, M. M. S. (2019). Tecnologias e educação: O uso das TIC como ferramentas essenciais para o processo de ensino e aprendizagem. *Brazilian journal of development*, 5(11), 25034-25042. <https://doi.org/10.34117/bjdv5n11-177>

Courage, C., & Baxter, K. (2005). *Understanding Your Users: A Practical Guide to User Requirements Methods, Tools, and Techniques*. <https://doi.org/10.1016/B978-1-55860-935-8.X5029-5>

Coutinho, C. Q. S., Silva, M. J. F., & Almouloud, S. A. (2011). Desenvolvimento do pensamento estatístico e sua articulação com a mobilização de registros de representação semiótica. *Bolema*, 24(39), 495-514.  
<http://www.redalyc.org/pdf/2912/291222099009.pdf>

Curcio, F. (1987). Comprehension of mathematical relationships expressed in graphs. *Journal for research in mathematics education*, 18(5), 382-393. <https://doi.org/10.2307/749086>

Dancey, C. P., & Reidy, J. (2019). *Estatística sem matemática para psicologia* (7ª ed.). Porto Alegre, RS: Penso.

Engel, P. (2014, fevereiro 23). *Gun deaths in Florida increased with "stand your ground"*. Business Insider.  
<https://www.businessinsider.com/gun-deaths-in-florida-increased-with-stand-your-ground-2014-2>

Felgueiras, Ó. (2013). O ensino da estatística no ensino superior em Portugal: o caso de uma professora. In J. M. Contreras, G. R. Cañadas, M. M. Gea y P. Arteaga (Eds.), *Actas de las Jornadas Virtuales en Didáctica de la*

*Estadística, Probabilidad y Combinatoria* (pp. 463-471). Departamento de Didáctica de la Matemática de la Universidad de Granada. <https://dialnet.unirioja.es/descarga/articulo/4770386.pdf>

Figma Design. (2024). *Figma: the collaborative interface design tool*. <https://www.figma.com/>

Friel, S., Curcio, F., & Bright, G. (2001). Making sense of graphs: Critical factors influencing comprehension and instructional implications. *Journal for Research in Mathematics Education*, 32(2), 124-158. <https://doi.org/10.2307/749671>

Gal, I. (2002). Adults' statistical literacy: Meanings, components, responsibilities. *International Statistical Review*, 70(1), 1-51. <https://doi.org/10.2307/1403713>

Github. (2024). *Github: Social coding*. <http://github.com/>

Grolemund, G., & Wickham, H. (2017). *R for data science*. O'Reilly Media.

Hagtvedt, R., Jones, G. T., & Jones, K. (2007). Pedagogical simulation of sampling distributions and the central limit theorem. *Teaching Statistics*, 29(3), 94-97.

Healy, K. (2018). *Data visualization: A practical introduction*. Princeton University Press.

Huff, D. (1954). *How to lie with statistics*. Norton.

Karadeniz, P. G. et al. (2019). Statistical errors in articles published in radiology journals. *Diagnostic and interventional radiology*, 25(2), 102-108. <https://doi.org/10.5152/dir.2018.18148>

Lopes, E. C., (2013). Educação Estatística no Curso de Licenciatura em Matemática. *Boletim de Educação Matemática*, 27(47), 901-915.

Lunsford, M. L., Rowell, G. H., & Goodson-Espy, T. (2006). Classroom research: Assessment of student understanding of sampling distributions of means and the central limit theorem in post-calculus probability and statistics classes. *Journal of Statistics Education*, 14(3). <https://doi.org/10.1080/10691898.2006.11910587>

Monteiro, C. E. F. (2005). *Investigating critical sense in the interpretation of media graphs*. (Tese de Doutorado não publicada). Institute of Education, University of Warwick, Inglaterra.

Morais, P. (2010). *Construção, leitura e interpretação de gráficos estatísticos por alunos do 9.º ano de escolaridade* [Tese de mestrado, Repositório Institucional da Universidade do Minho].

<https://repositorium.sdum.uminho.pt/handle/1822/14944>

Nolan, D., & Speed, T. P. (1999). Teaching Statistics Theory Through Applications. *The American Statistician*, 53(4), 370–375. <https://doi.org/10.2307/2686059>

Prates, W. O. (2017). *Estatísticas para as ciências sociais aplicadas 1*. UFBA, Faculdade de Ciências Contábeis, Superintendência de Educação a Distância.

Queiroz, T. N., Monteiro, C. E. F., Carvalho, L. M. T. L., & François, K. (2017). Interpretation of statistical data: The importance of affective expressions. *Statistics Education Research Journal*, 16(1), 163-180.

<https://doi.org/10.52041/serj.v16i1.222>

Queiroz, T. N. (2015). *Expressões afetivas na interpretação de dados estatísticos*. (Dissertação de mestrado, Programa de Pós-Graduação em Educação Matemática e Tecnológica, Universidade Federal de Pernambuco, Recife, Brasil). <https://repositorio.ufpe.br/handle/123456789/13940>

R Core Team (2023). *R: A language and environment for statistical computing*. R foundation for statistical computing, Viena, Áustria. <https://www.R-project.org>

RStudio Team (2023). *RStudio: Integrated development for R*. RStudio, Inc., Boston.

Sousa, O. (2002). *Investigações estatísticas no 6.º ano*. In GTI (Ed.), *Reflectir e investigar sobre a prática profissional* (pp. 75-97). Lisboa: APM.

Stone, M. (1996). Using color in information display graphics: A review. *Color Research & Application*, 21(1), 31-47. <http://graphics.cs.cmu.edu/nsp/course/15-462/Fall04/slides/maureenStoneSIG2004.pdf>

Teixeira, L. B., (2018, maio 30). *PSDB-SP divulga gráfico desproporcional de Doria e tira do ar após críticas*. UOL Notícias. <https://noticias.uol.com.br/politica/eleicoes/2018/noticias/2018/05/30/psdb-sp-divulga-grafico-desproporcional-de-doria-e-tira-do-ar-apos-criticas.htm>

Tufte, E. R. (1983). *The visual display of quantitative information*. Graphics Press.

United Nations Office on Drugs and Crime. (2019). *Regional drug analysis report - Latin America and Caribbean*. [https://www.unodc.org/documents/scientific/Regional\\_Drug\\_Analysis\\_Report\\_-\\_Latin\\_America\\_and\\_Caribbean\\_2019\\_PT.pdf](https://www.unodc.org/documents/scientific/Regional_Drug_Analysis_Report_-_Latin_America_and_Caribbean_2019_PT.pdf)

Velleman, P. F., & Moore, D. S. (1996). Multimedia for teaching statistics: Promises and Pitfalls. *The American statistician*, 50(3), 217. <https://doi.org/10.2307/2684658>

Waiselfisz, J. J. (2011). *Mapa da violência 2011: Os jovens no Brasil*. São Paulo: Instituto Sangari; Brasília, DF: Ministério da Justiça. <https://flacso.org.br/files/2020/03/MapaViolencia2011.pdf>

Wickham, H. (2016). *ggplot2: Elegant Graphics for Data Analysis*. Springer-Verlag New York. ISBN 978-3-319-24277-4. <https://ggplot2.tidyverse.org>

Wilkinson, L. (2005). *The grammar of graphics*. Springer Science & Business Media

## ANEXOS

### Anexo 1 - Questionário realizado

#### Questionário de avaliação do aplicativo

Este questionário visa coletar a opinião dos alunos da disciplina de análise de dados do mestrado de Humanidades Digitais sobre o aplicativo interativo, a fim de aprimorá-lo.

O questionário tem uma duração aproximada de 5 minutos e as respostas são confidenciais. As perguntas com um asterisco (\*) são de resposta obrigatória.

Obrigada pela sua colaboração!

\* Indica uma pergunta obrigatória

Você conseguiu acessar o aplicativo? \*

- Sim  
 Não

Pergunta 2 de 10

Como você classifica a usabilidade do aplicativo? \*

- Excelente  
 Muito boa  
 Regular  
 Ruim  
 Péssima

Pergunta 3 de 10

Você conseguiu compreender o objetivo do aplicativo? \*

- Compreendi completamente  
 Compreendi bem  
 Compreendi razoavelmente  
 Compreendi um pouco  
 Não compreendi

Pergunta 4 de 10

O aplicativo me ajudou a escolher a visualização gráfica mais adequada. \*

- Concordo totalmente
- Concordo
- Não concordo, nem discordo
- Discordo
- Discordo totalmente

Pergunta 5 de 10

indique o quanto você recomendaria este aplicativo para outros usuários. \*

- Certamente recomendaria
- Provavelmente recomendaria
- Recomendaria ou não
- Provavelmente não recomendaria
- Certamente não recomendaria

Pergunta 6 de 10

Você acha que aplicativos interativos como este contribuem para o aprendizado em disciplinas de análise de dados? \*

- Contribui muito
- Contribui
- Contribui moderadamente
- Contribui pouco
- Não contribui

Pergunta 7 de 10

As recomendações de gráficos fornecidas pelo aplicativo foram compreensíveis? \*

- Muito compreensíveis
- Bastante compreensíveis
- Moderadamente compreensíveis
- Pouco compreensíveis
- Nada compreensíveis

Pergunta 8 de 10

Você acha que a explicação do conteúdo é satisfatória? \*

- Muito satisfatória
- Satisfatória
- Moderadamente satisfatória
- Pouco satisfatória
- Nada satisfatória

Pergunta 9 de 10

O que você mais gostou e/ou menos gostou no aplicativo?

.....

Pergunta 10 de 10

Sugestão, crítica ou comentário:

.....