

Universidade Federal do Paraná
Setor de Ciências Exatas
Departamento de Estatística
Programa de Especialização em *Data Science* e *Big Data*

Henrique Favre

**Aplicativo em Flex Dashboards para Visualização
de Receitas no Contexto de Pesquisa &
Desenvolvimento de Produtos Alimentícios**

**Curitiba
2023**

Henrique Favre

**Aplicativo em Flex Dashboards para Visualização de
Receitas no Contexto de Pesquisa & Desenvolvimento de
Produtos Alimentícios**

Monografia apresentada ao Programa de Especialização em *Data Science* e *Big Data* da Universidade Federal do Paraná como requisito parcial para a obtenção do grau de especialista.

Orientador: Prof. Wagner Bonat

Curitiba
2023

Aplicativo em *Flex Dashboards* para Visualização de Receitas no Contexto de P&D de Produtos Alimentícios

Henrique Favre¹, Wagner Bonat²

¹UFPR: Universidade Federal do Paraná Curitiba (PR) Brasil*

²UFPR: Universidade Federal do Paraná Curitiba (PR) Brasil†

Esse trabalho trata da construção de um aplicativo, feito em R e *Flex Dashboards*, publicado via *RStudio Connect*, que democratiza, para Engenheiros de Alimentos em uma organização de Pesquisa e Desenvolvimento de Produtos Alimentícios, o acesso ao histórico de formulações, por técnicas de visualização de dados e de redução de dimensionalidade, em contraste a um contexto anterior em que o acesso a tais informações dependia da busca individual dos dados. O aplicativo permite ao usuário explorar de forma intuitiva o histórico contido em uma base de dados e obter ideias quanto aos ingredientes já explorados, a combinações de ingredientes inexploradas e de receitas similares já estudadas anteriormente.

Palavras-chave: Flex Dashboards, Visualização, Pesquisa e Desenvolvimento

This paper concerns the building of an app in R and Flex Dashboards, published via RStudio Connect, which democratizes the access to historical data on recipes for Food Engineers in a Research and Development organization, as opposed to a previous context in which the access to such information depended on the individual search of the data. Data visualization and dimensionality reduction techniques are used, allowing the end user to intuitively explore historical information contained in a database while generating insights on which ingredients have been already studied and of similar recipes explored in the past.

Keywords: Flex Dashboards, Visualization, Research and Development

1. Introdução

A um Engenheiro de Alimentos, no contexto de Pesquisa e Desenvolvimento de produtos alimentícios, são designados projetos para a solução de problemas em produtos, tais como: melhoria de qualidade, quando o produto se apresenta defasado, em testes com consumidores, perante competidores; redução de custo de formulação ou processamento, em caso de necessidade de recuperação de margem; e substituição de ingredientes, quando existem mudanças regulatórias ou de cenário comercial.

Cada projeto de desenvolvimento de produto requer, em algum nível, a fabricação e degustação de protótipos, em bancada ou em planta piloto, até que seja encontrada uma solução que satisfaça todas as restrições impostas. Exemplos de restrição incluem, mas não se restringem a:

- Limites regulatórios para o uso de ingredientes específicos, como edulcorantes ou cacau;
- Limites de natureza nutricional, como teores de açúcares, sódio e gordura saturada;
- Limites no consumo de energia ou emissão de gases do efeito estufa;
- Limites de custo.

A fabricação de produtos, mesmo que somente em ambiente piloto, é custosa. Portanto, é de grande valia ao Engenheiro obter acesso ao conhecimento das formulações já realizadas no passado – e seus resultados e dados – visto que podem proporcionar redução no número de amostras necessárias em um projeto corrente.

A adoção de sistemas de gerenciamento de laboratórios (LIMS), de experimentos em laboratório (*Lab Notebooks*) e de execução de manufatura (MES), aliada a associação de bancos de dados que permitam a conexão de receitas, suas condições de processamento e seus resultados, abre a possibilidade para o Engenheiro

*henriquefavre@gmail.com

†wbonat@ufpr.br

tenha acesso a dados cujo acesso antes dependia de buscas manuais em pastas de projeto, físicas ou eletrônicas.

Esse trabalho trata da construção de um aplicativo, em R e *Flex Dashboards* que democratiza, via publicação em *RStudio Connect*, o acesso ao histórico de formulações por técnicas de redução de dimensionalidade e de visualização de dados. O aplicativo permite ao usuário explorar o histórico contido em uma base de dados e obter ideias quanto aos ingredientes já explorados, e combinações de ingredientes inexploradas e a receitas similares.

Os materiais e métodos utilizados são descritos, incluindo pacotes utilizados, metodologias estatísticas, técnicas de visualização de dados, e o uso do *ChatGPT*. Os resultados são então apresentados conforme a ordem em que foram implementados no aplicativo de visualização, e em seguida as conclusões são apresentadas.

2. Materiais e Métodos

Nessa seção são apresentados os pacotes utilizados para desenvolvimento do aplicativo. Os dados disponíveis e utilizados são descritos. As abordagens utilizadas para a sumarização de dados, por meio de metodologias estatísticas e técnicas de visualização de dados são apresentadas.

2.1. Linguagens Utilizadas para a Construção do Painel (*Dashboard*)

Linguagem de Programação R: R [6] é uma linguagem de programação estatística amplamente utilizada em ciência de dados. É uma linguagem de código aberto que oferece uma vasta gama de pacotes e bibliotecas para realizar análises estatísticas avançadas. No contexto deste trabalho, o R foi escolhido como a linguagem principal de programação devido à sua flexibilidade e recursos estatísticos robustos. O R permite a manipulação de dados, cálculos estatísticos e a criação de gráficos de alta qualidade, facilitando a exploração e a visualização das receitas de alimentos de forma eficiente.

Pacote de R *Tidyverse*: *Tidyverse* [10] é uma coleção de pacotes em linguagem R criados com uma gramática intuitiva, que facilita a manipulação, análise e visualização de dados. Dentre seus pacotes, se destacam *dplyr*, *ggplot2* e *tidyr*, que oferecem uma abordagem consistente e intuitiva para tarefas comuns de

ciência de dados. No contexto deste trabalho, o *Tidyverse* foi utilizado para realizar a limpeza e preparação dos dados, e ligação (*join*) de diferentes tabelas, transformando-os em um formato adequado para a visualização das fórmulas. O pacote *ggplot2*, em particular, foi utilizado para criar gráficos personalizados que representam visualmente as fórmulas de maneira clara e informativa.

***Flex Dashboards* [1]:** Esta biblioteca em R permite a criação de *dashboards* interativos e responsivos. Ele oferece uma variedade de recursos para criar layouts flexíveis, incluindo a possibilidade de adicionar gráficos interativos, tabelas e outros elementos personalizados. Neste trabalho, o *Flex Dashboards* foi escolhido como a ferramenta principal para criar o painel de visualização das fórmulas. Com sua facilidade de uso, programação simples e capacidade de personalização, o *Flex Dashboards* permitiu a criação de um painel intuitivo e interativo para apresentar dados de forma acessível aos usuários.

Ambiente de desenvolvimento integrado (IDE) *RStudio* [8]: Recentemente rebatizado para *Posit*, o *RStudio* é um ambiente projetado originalmente para trabalhar com a linguagem R. Ele oferece uma interface intuitiva e amigável, facilitando a escrita, a execução e a depuração de código. O *RStudio* possui recursos como realce de sintaxe, autocompletar, gerenciamento de pacotes e integração com controle de versão (Git), o que facilita o desenvolvimento do projeto. Neste trabalho, o *RStudio* foi utilizado como a principal IDE durante o desenvolvimento do *dashboard* de visualização das fórmulas. Seu *browser* integrado permite a rápida alternância entre programação e verificação do resultado, durante a fase de testes do código.

Sistema para controle de versões *Git*: Git [2] é um sistema de controle de versão amplamente utilizado para rastrear e gerenciar alterações no código fonte. Ele permite que várias pessoas trabalhem simultaneamente no mesmo projeto, facilitando a colaboração e o acompanhamento das alterações feitas ao longo do tempo. No contexto deste trabalho, o Git foi utilizado como a ferramenta de controle de versão para o código fonte do projeto. Ele permitiu o rastreamento de alterações, a criação de *branches* para desenvolvimento paralelo e a integração com serviços de hospedagem, como o *GitHub*, para facilitar o compartilhamento e a revisão do código.

Plataforma de compartilhamento de aplicativos *RStudio Connect* [7]: Esse ambiente possui uma plata-

forma de publicação e compartilhamento de aplicativos e *dashboards* desenvolvidos em R. Ele oferece recursos avançados de implantação e gerenciamento de aplicativos, permitindo que os usuários acessem e interajam com os *dashboards* de forma segura e escalável. Neste trabalho, o *RStudio Connect* foi utilizado para implantar e hospedar o painel de visualização das fórmulas. Isso permitiu que o *dashboard* fosse acessível a um público mais amplo, facilitando a disseminação das informações e resultados obtidos.

2.2. Dados Disponíveis

Esse trabalho se trata de uma versão inicial de aplicativo, ainda em estágio de coleta da opinião dos usuários finais. Ademais, o conjunto de dados, mesmo que coletados ao longo de 3 anos, não possuem mais do que 92 receitas. Por essas características, a disponibilização de dados se deu através de planilhas em formato csv, e ligadas (operação *join*) realizada pelos códigos únicos identificadores de receitas e pelas descrições únicas dos ingredientes. As tabelas disponíveis são:

Receitas: um arquivo contendo as receitas, sendo a primeira coluna recipiente de um identificador único e as demais colunas nomeadas pelos nomes de ingredientes, cujo valor em cada linha representa a quantidade porcentual de cada ingrediente na receita. Todos os valores numéricos em cada linha somam, portanto, 100. Valores faltantes são substituídos por zero, justamente por representarem a ausência do ingrediente na receita.

Testes: um arquivo contendo metadados sobre os testes, possuindo uma coluna com o mesmo identificador único do arquivo de receitas, e demais colunas indicando:

- a presença ou ausência de resultados de testes sensoriais naquela receita (o que torna o experimento mais valioso do ponto de vista de obtenção de conhecimento);
- o local onde a receita foi gerada;
- o formato do produto; e
- a marca do produto.

Ingredientes: um arquivo contendo uma associação entre os ingredientes e seus grupos (por exemplo: ácidos, emulsificantes, edulcorantes, corantes, aromas).

2.3. Técnicas Estatísticas

Nessa subseção estão apresentadas as técnicas estatísticas, na ordem em que se sugere que sejam utilizadas

pelo usuário final. Tais visualizações estão dispostas em abas, que organizam o processo de exploração para o usuário final.

Tabelas e Distribuições de Frequência: A tabulação é uma técnica estatística utilizada para resumir dados categóricos por meio da contagem de frequências e proporções agrupando os dados em classes, intervalos ou categorias [3]. Ela permite visualizar a distribuição de uma única variável e é frequentemente representada em forma de tabelas simples, exibindo as categorias e suas respectivas contagens ou porcentagens. A tabulação é uma etapa inicial essencial na análise exploratória de dados e fornece a possibilidade de descobertas sobre a distribuição e a frequência de cada categoria.

A tabulação cruzada é uma extensão da tabulação que busca o resumo simultâneo de duas ou mais variáveis categóricas. A tabulação cruzada permite analisar a relação entre essas variáveis. Os resultados são apresentados em forma de tabelas multidimensionais, exibindo as contagens ou porcentagens para cada combinação de categorias.

Especificamente no aplicativo de visualização de receitas, são apresentadas tabelas que buscam dar ao usuário final a percepção de quantos ingredientes estão contidos na base de dados e quantos deles estão presentes no histórico de receitas com uma maior ou menor variação. A visualização mostra o total de ingredientes únicos, e quantos foram utilizados em receitas de forma fixa, ou variando em dois ou mais níveis. Uma tabela com filtro é disponibilizada para cada ingrediente, mostrando a contagem de vezes em que foi incluído em receitas, em quantos diferentes níveis o ingrediente foi estudado.

Análise de Componentes Principais (ACP): Essa técnica estatística é utilizada para reduzir a dimensionalidade de um conjunto de dados multivariados através da construção de componentes principais, calculados como combinações lineares das variáveis originais [4]. Ela permite identificar os principais padrões de variabilidade nos dados e representá-los em um espaço de menor dimensão.

A ACP é amplamente aplicada em diversas áreas, como análise de dados, aprendizado de máquina e visualização.

Dada a grande quantidade de ingredientes envolvida nos testes, optou-se por uma visualização que dê ao usuário final uma visão relativamente rápida das receitas constantes no histórico de dados, sendo possível a identificação de combinações de ingredientes nunca

explorada – vistas como espaços em branco no gráfico de escores dos componentes principais – ou de ingredientes com alta correlação entre eles – aqueles pertencentes ao mesmo componente, que tendem a ter alto índice de correlação linear entre si devido às propriedades de ACP.

A inclusão de botões de seleção permite ao usuário colorir os pontos conforme os metadados do teste, por exemplo ajudando-os a encontrar regiões com muitos testes porém sem a presença de receitas na base de dados.

Coefficiente r de Correlação Linear de Pearson: o coeficiente r é uma medida estatística que avalia a relação linear entre duas variáveis contínuas. Ele varia de -1 a 1, onde -1 indica uma correlação negativa perfeita, 1 indica uma correlação positiva perfeita e 0 indica ausência de correlação [3]. O coeficiente de *Pearson* é amplamente utilizado em análise exploratória de dados, testes de hipóteses e modelagem estatística.

Gráficos exploratórios: gráficos como o histograma, o gráfico de dispersão e o mapa de calor, são ferramentas visuais para a análise de dados [3].

O histograma é utilizado para visualizar a distribuição dos dados e identificar padrões como concentração e assimetria.

O gráfico de dispersão é utilizado para exibir a relação entre duas variáveis contínuas, permitindo observar a presença de tendências, agrupamentos e *outliers*.

O mapa de calor é utilizado para visualizar matrizes de dados, destacando padrões de valores altos e baixos por meio de cores.

Esses gráficos são amplamente empregados em análise exploratória de dados, permitindo uma compreensão inicial das características e relações presentes nos conjuntos de dados.

Distâncias de Manhattan e Euclidiana: são métricas utilizadas para medir a diferença entre dois pontos em um espaço euclidiano. A distância de *Manhattan* é calculada somando-se as diferenças absolutas entre as coordenadas dos pontos, enquanto a Distância Euclidiana é calculada utilizando o teorema de Pitágoras para obter a distância em linha reta entre os pontos. Essas métricas são comumente utilizadas em análise de dados, agrupamentos, classificação e outras tarefas de aprendizado de máquina.

Nesse projeto, o objetivo é proporcionar ao usuário final, quais receitas constantes do banco de dados são as mais próximas em relação a uma receita de interesse.

2.4. ChatBot (ChatGPT, versão 3.5)

O *ChatGPT* [5] é considerado um *Large Language Model (LLM)*, popularizado no final de 2022. Desde a sua disponibilização do *ChatGPT* ao público, sua adoção na geração de textos levanta questionamentos éticos sobre vício, robustez, confiabilidade e toxicidade (na linguagem) [11]. O uso de LLMs como ferramentas de apoio a programadores também tem sido pesquisado, e as primeiras publicações a esse respeito concluem que o *ChatGPT* supera modelos anteriores por larga margem, e ressaltam a necessidade do aperfeiçoamento, por parte de programadores, das instruções fornecidas (*prompts*) ao modelo, evitando, por exemplo, instruções muito longas [9].

No contexto desse trabalho, o autor pode confirmar tais conclusões. O *ChatGPT* foi amplamente utilizado durante a criação do código, tendo sido ferramenta fundamental para superar a curva de aprendizagem dos pacotes *Tidyverse* e *Flex Dashboards*. Instruções eram fornecidas ao *ChatGPT* e o código gerado então adaptado ao caso particular do projeto. Em sua maioria, os códigos não funcionavam a ponto de se necessitar a busca por outras fontes de informação. Soluções para a manipulação de dados, geradas em linguagem base do R, eram rapidamente substituídas pela sintaxe *Tidyverse*, bastando apenas instruir o modelo com o seguinte *prompt*: "transforme o código anterior em *tidyverse*".

3. Apresentação dos Resultados

A disponibilização da ferramenta aos usuários é feita por meio do *RStudio Connect*, conforme visto na Figura 1.

Com o código apropriadamente disponibilizado em um repositório de software, o *RStudio Connect* é utilizado para publicá-lo, bastando para isso referenciar o repositório e dar um nome ao aplicativo. Os usuários então passam a ter acesso, via browser, ao aplicativo.

Nesse artigo serão apresentados os resultados na sequência em que são apresentados em suas abas, com a ressalva de que dados confidenciais estão propositalmente borrados. A sequência busca seguir uma ordem lógica de apresentação dos resultados ao usuário final.

Dado que os dados históricos são provenientes de diversos estudos realizados em separado, seus resultados não se propõem a cobrir todo o possível espaço de receitas. O aplicativo se propõe a fornecer ao usuário respostas para perguntas como:

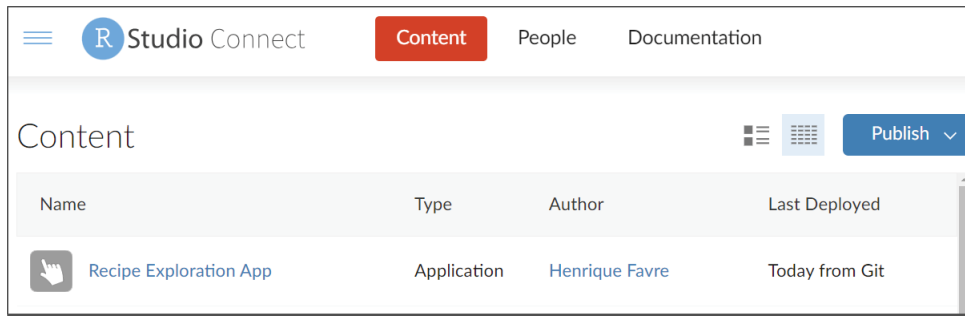


Figura 1: Publicação via RStudio Connect

- Quantos ingredientes foram estudados?
- Quais desses ingredientes foram estudados em diversos níveis?
- De quais ingredientes podemos extrair efeitos com menor risco de confundimento, ou, quais ingredientes têm a variação em seus níveis menos correlacionadas com os níveis de outros ingredientes?
- Quais receitas devo criar com o intuito de preencher melhor o espaço amostral?
- Existem grupos de receitas em um dado espaço das quais nunca coletamos resultados sensoriais ou de consumidor?

3.1. Uso do ChatGPT, versão 3.5

O uso do *ChatGPT* 3.5 permeia a construção do trabalho, mesmo que não percebido no aplicativo ao usuário final. Tal ferramenta permitiu ganho de produtividade ao autor na construção de um total de aproximadamente 1.000 linhas de código em pacotes não antes utilizados pelo autor.

Apesar do *ChatGPT* ser um LLM que processa o português, a escolha do autor foi a de utilizar a língua inglesa, que tende a possuir maior base de informações dos pacotes de R utilizados. Instruções sempre foram fornecidas de forma genérica de modo a não permitir qualquer compartilhamento de dados confidenciais, como receitas, nomes de ingredientes, entre outros.

A Figura 2 mostra um exemplo no uso do *ChatGPT* pelo autor específico, em que se buscava uma fatia de um conjunto de dados conforme os valores em outras colunas. A solução original era proposta em código R base. Apesar de correta, a intenção era realizar as tarefas em *Tidyverse*. Ao *ChatGPT* bastou receber a instrução "pode fazer o mesmo no modo *tidyverse*?" e uma nova solução foi apresentada usando o pacote de interesse.

Pode-se confirmar a importância de uma boa escrita das instruções (*prompts*), tendo as instruções específicas obtido melhores respostas, e instruções muito longas, mesmo que itemizadas, não obtido igual sucesso.

Respostas erradas são frequentes, levando a *bugs*, cabendo ao programador informar o erro gerado e esperar que o *ChatGPT* encontre uma solução. Em alguns casos isso leva a um ciclo de soluções incorretas, a partir de onde se faz necessário a busca em outras fontes de informação.

Mesmo com alguns percalços, o autor recomenda o uso do *ChatGPT* como apoio em programação.

As próximas sub-seções apresentam o aplicativo criado, aba a aba. A interface do aplicativo foi construída em inglês devido à natureza global da empresa para a qual foi criado o aplicativo.

3.2. Visualização do uso de ingredientes (Aba 1)

O painel da primeira aba do aplicativo busca dar uma visão geral da base de dados, especificamente quanto ao total de ingredientes nela contidos e quanto estes de fato foram explorados. Nele são disponibilizadas tabelas cruzadas, filtros e sumários que mostram quantos e quais ingredientes (ou grupos de ingredientes) foram mais explorados em testes anteriores. O painel pode ser visto na Figura 3.

O painel esquerdo mostra as contagens totais de ingredientes contidos na base de dados. Dentre esses, quantos foram estudados em apenas um nível fixo, quantos em dois níveis, e quantos em três ou mais níveis.

Na segunda coluna, uma tabela cruzada permite ao usuário final visualizar as contagens de ingredientes conforme a participação em receitas (desde "menos de 10% do total" até "mais de 30% do total"), e em quantos níveis variaram (1, 2, ou "3 ou mais").

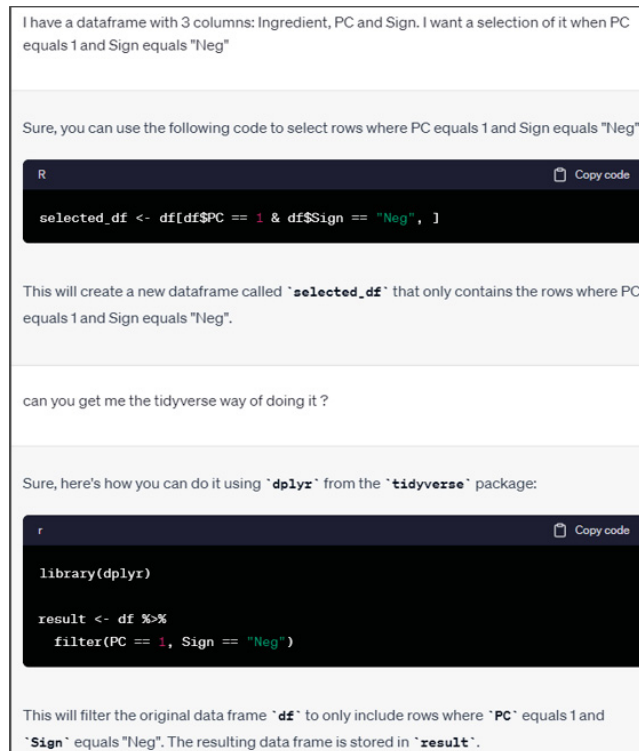


Figura 2: Exemplo do uso do ChatGPT durante a programação do aplicativo

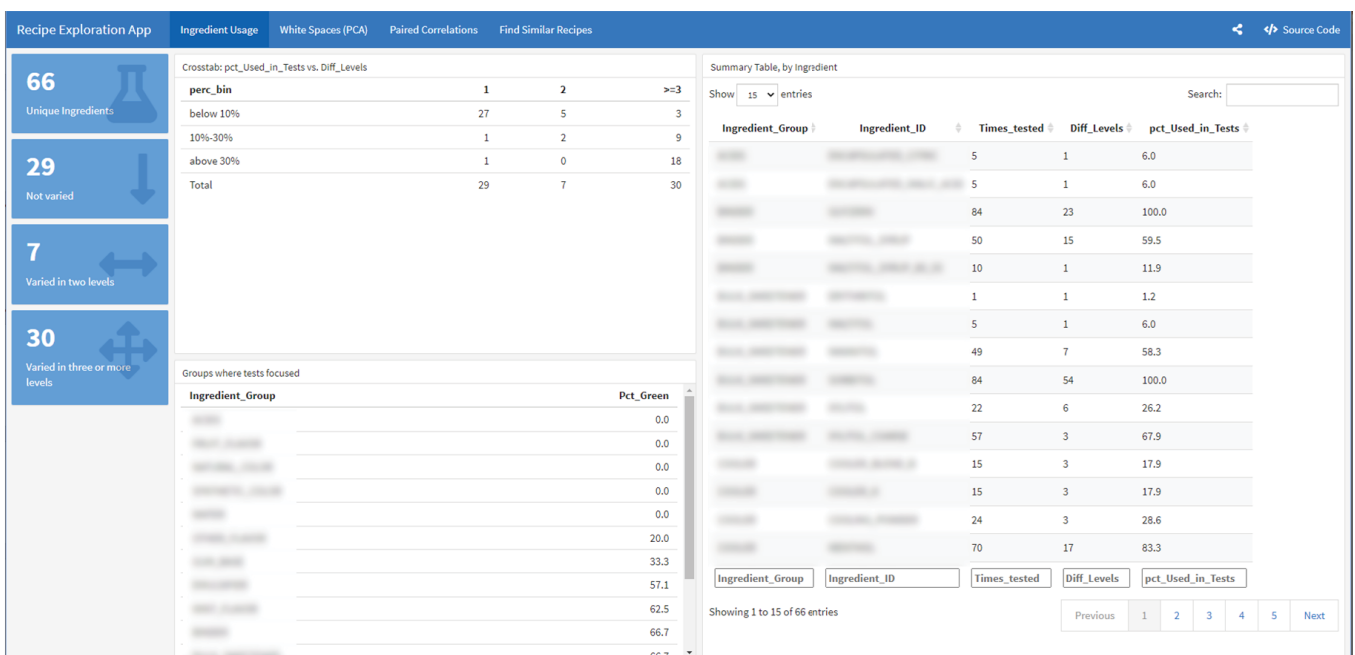


Figura 3: Visualização do uso de ingredientes (Aba 1)

Por convenção, os ingredientes já estudados em dois ou mais níveis, e presentes em mais de 10% dos testes são categorizados pela cor verde. A segunda tabela da segunda coluna mostra, por grupo de ingredientes (provenientes da tabela "Ingredientes"), quais foram aqueles explorados em vários níveis ao longo dos anos,

o que pode indicar ao usuário uma menor necessidade de criação de protótipos.

Na terceira coluna, uma tabela com filtro permite que o usuário explore os dados de utilização por ingrediente, obtendo o total de níveis em que cada ingredi-

ente individual foi explorado e o total de receitas em que foi utilizado.

3.3. Visualização em dimensionalidade reduzida para busca de espaços inexplorados, via ACP (Aba 2)

No intuito de resumir o máximo de informação a respeito das receitas, a técnica de Análise de Componentes Principais foi aplicada para reduzir a dimensionalidade dos dados, e disponibilizada ao usuário no painel mostrado na Figura 4.

Uma função foi escrita, seguindo os seguintes passos, de forma resumida:

1. ACP (função *prcomp*) na tabela de Receitas, com parâmetros *scale* e *center* iguais a "TRUE";
2. O número total de componentes escolhido como o número par que cobre 85% ou mais da variação presente nos dados;
3. Cada ingrediente é associado a um componente principal conforme seus *loadings* máximos;
4. O gráfico de escores mostra ao usuário cada par de componentes (1 e 2, 3 e 4, 5 e 6 etc.).

Essa visualização permite ao Engenheiro de Alimentos buscar espaços em branco, ou seja, combinações de ingredientes que não possuem qualquer receita estudada. "Botões de rádio" são disponibilizados no painel esquerdo para que o usuário possa colorir os pontos conforme a tabela de Metadados dos testes, verificando quais amostras foram feitas em quais centros de desenvolvimento, se possuem ou não dados sensoriais coletados, a marca do produto e o seu formato.

Também é possível explorar ingredientes que apresentem alto grau de correlação entre si, o que ocorre quando estão presentes nos mesmos componentes, ou descorrelacionados entre si, quando presentes em componentes diferentes.

3.4. Visualização de distribuições e correlações pareadas (Aba 3)

De posse de ingredientes específicos de interesse, o usuário final pode desejar explorar a distribuição nos níveis de utilização de um ingrediente ou a relação pareada entre pares de ingredientes.

A terceira aba apresenta um painel de seleção múltipla de ingredientes ao lado esquerdo e, ao lado direito, uma interface dinâmica que apresenta um tipo de gráfico conforme o número de ingredientes selecionados.

- **Um ingrediente - Histograma:** Permite explorar a distribuição de utilização de cada ingrediente de forma isolada. A figura 5 apresenta um histograma para um ingrediente. De posse desse gráfico, o usuário pode se perguntar em que contexto o ingrediente em questão (um preenchedor de fórmula) foi utilizado em quantidade bem mais baixa que em todas as demais receitas, e qual outro ingrediente foi utilizado para substituí-lo.
- **Dois ingredientes - Gráfico de Dispersão:** Habilita ao usuário explorar a relação entre o uso de dois ingredientes nas receitas do histórico. O recurso *jitter* foi aplicado para espalhar ligeiramente pontos da mesma coordenada, de modo a evidenciar a quantidade total de receitas com a mesma combinação de ingredientes. A figura 6 apresenta um gráfico de dispersão de dois ingredientes, onde se nota, pela presença exclusiva de pontos sobre os eixos vertical e horizontal, que ambos os ingredientes nunca estiveram presentes em uma mesma receita. Tal informação pode estimular o Engenheiro de Alimentos, por exemplo, a realizar um teste com uma mistura dos dois, ou buscar avaliar se o mais caro dos dois pode ser substituído pelo mais barato em seu caso particular.
- **Entre três e seis ingredientes - Matriz de Gráficos de Dispersão:** Permite ao usuário realizar as mesmas investigações do caso de dois ingredientes, porém visualizando mais pares de gráficos e com a adição do valor *r* da Correlação de Pearson, conforme visto na Figura 7.
- **Mais de seis ingredientes - Mapa de Calor:** Ao selecionar mais de seis ingredientes, a matriz de gráficos de dispersão se torna poluída, fazendo sentido remover os detalhes e mostrar somente os coeficientes de Correlação de Pearson em um mapa de calor, onde os tons mais fortes indicam maior força de associação linear entre os níveis de cada par de ingredientes, e a cor representa a natureza da associação linear, sendo positiva quando a célula estiver azul e negativa quando a célula estiver vermelha. Um exemplo do mapa de calor pode ser visto na Figura 8. Adicionalmente, o valor de *r* é mostrado, embora possa não estar perfeitamente visível caso muitos ingredientes sejam selecionados. O mapa de calor, assim como o gráfico de ACP, pode ser uti-

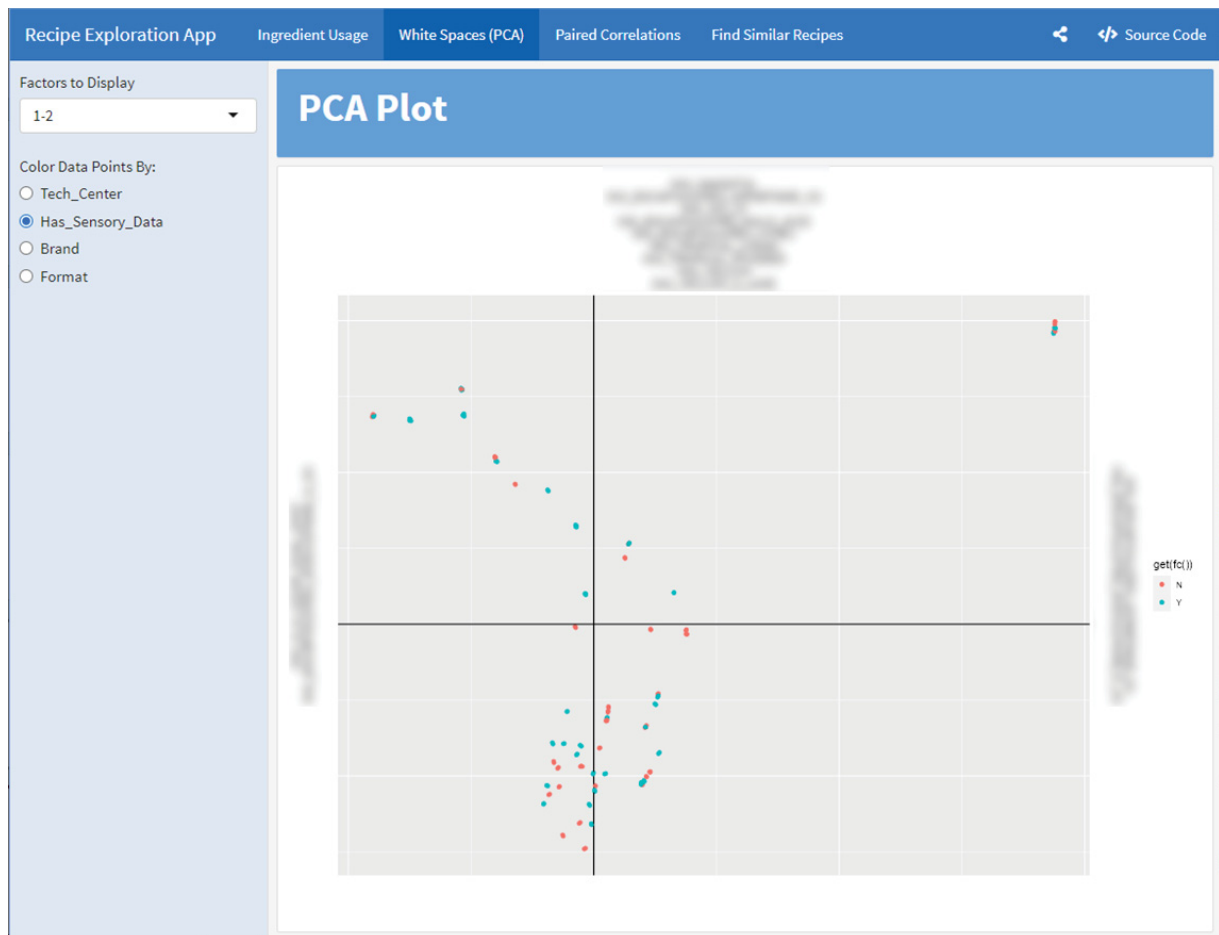


Figura 4: Análise de Componentes Principais (Aba 2)

lizado como uma ferramenta de triagem para o Engenheiro de Alimentos, um primeiro passo antes de focar na análise mais específica de outros pares de ingredientes.

3.5. Buscador de receitas similares e Comparador de Receitas (Aba 4)

O buscador de receitas similares dá a possibilidade ao usuário comparar uma receita própria com as receitas presentes na base de dados, bastando que clique em *browse* e *suba* um arquivo de Excel de formato pré-definido com os dados de sua receita.

A receita inserida aparece na primeira coluna. Na segunda coluna, essa receita é comparada às demais receitas da base usando as distâncias de Manhattan e Euclidiana, ambas calculadas após normalização dos dados para cada ingrediente, de modo a impedir que ingredientes usados em maior quantidade dominassem o cálculo das distâncias. A tabela é ordenada conforme as menores distâncias, sendo as dez mais próximas mostradas ao usuário final.

Possuindo interesse em alguma das receitas visualizadas, o usuário pode copiar o seu código e inseri-lo no painel esquerdo, o que traz na terceira coluna a comparação, ingrediente a ingrediente, de sua receita com aquela do código inserido. A tabela superior mostra todos os ingredientes que estão presentes nos mesmos níveis que a receita proposta e a tabela inferior mostra os ingredientes cujas quantidades diferem entre as duas receitas.

4. Conclusão

A sintaxe da *Flex Dashboards* se mostrou bastante adequada para a construção rápida de um aplicativo que democratiza o conhecimento na organização a respeito de formulações de produtos já realizadas anteriormente, potencializando o trabalho do usuário final.

A solução foi escalada através da plataforma *RStudio Connect*, onde recebe acessos dos usuários interessados em explorar a visualização de dados proporcionada pelo aplicativo. A ferramenta tem recebido sugges-

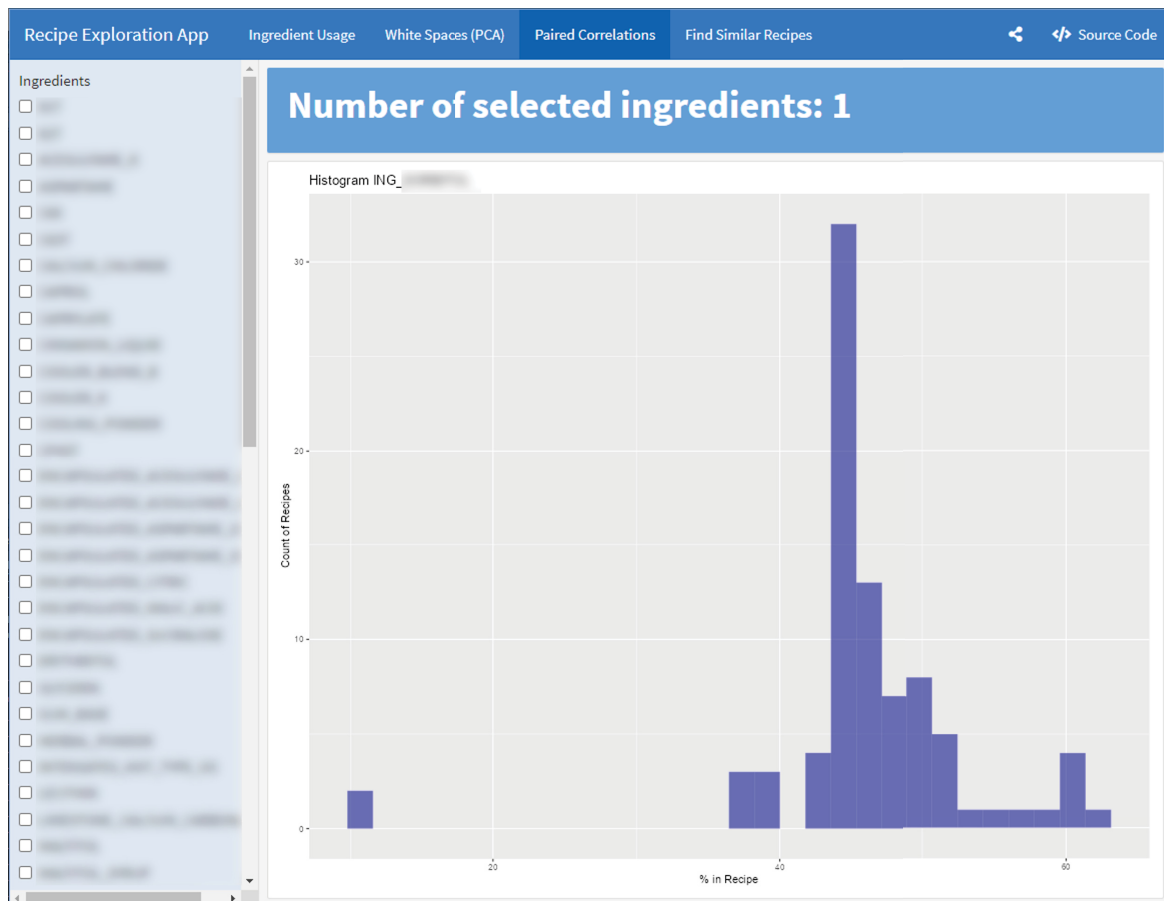


Figura 5: Visualização de distribuições e correlações pareadas (Aba 3) - caso de um ingrediente selecionado

tões de modificação por parte do usuário, o que será realizado pelo uso de metodologias ágeis. Estando em um repositório de Git, ajustes podem ser facilmente priorizados e implementados conforme o retorno de usuários.

O uso do *ChatGPT 3.5* foi ferramenta essencial, na visão do autor, permitindo a retirada de dúvidas e propondo soluções de programação para o caso específico que, mesmo que nem sempre corretas, muitas vezes apontavam um caminho factível. Em comparação, buscas em fóruns online como o *StackOverflow* exigem um maior número de buscas e a adaptação de problemas de outros usuários para o caso particular de interesse, o que nem sempre é viável. Recomenda-se a adoção dessa ferramenta, com a ressalva de que a ela não se devem compartilhar informações de natureza confidencial.

Agradecimentos

Ao DSBD, pelos conhecimentos adquiridos, ao Prof. Wagner Bonat, pela orientação quanto ao uso da fer-

ramenta *Flex Dashboards* e ideias de visualização, e à minha esposa e meu filho, por todo o suporte durante a realização do curso.

Referências

- [1] S. Carson, R. Iannone, JJ Allaire, and B. Borges. *flexdashboard: R Markdown Format for Flexible Dashboards*, 2023. <https://pkgs.rstudio.com/flexdashboard/>, <https://github.com/rstudio/flexdashboard/>.
- [2] B. Chacon, S.; Straub. *Pro git*. Apress, 2014.
- [3] J. Freund and G. Simon. *Estatística Aplicada*. Bookman, Porto Alegre, 2000.
- [4] S.A. Mingoti. *Análise de dados através de métodos de estatística multivariada: uma abordagem aplicada*. Editora UFMG, 2005.
- [5] OpenAI. Chatgpt, 2023. Version GPT-3.5.
- [6] R Core Team. *R: A Language and Environment for Statistical Computing*. R Foundation for Statistical Computing, Vienna, Austria, 2022.
- [7] RStudio. Rstudio connect. <https://www.rstudio.com/products/connect/>, 2021. Version 2022.07.0.
- [8] RStudio Team. *RStudio: Integrated Development Environment for R*. RStudio, PBC., Boston, MA, 2020.

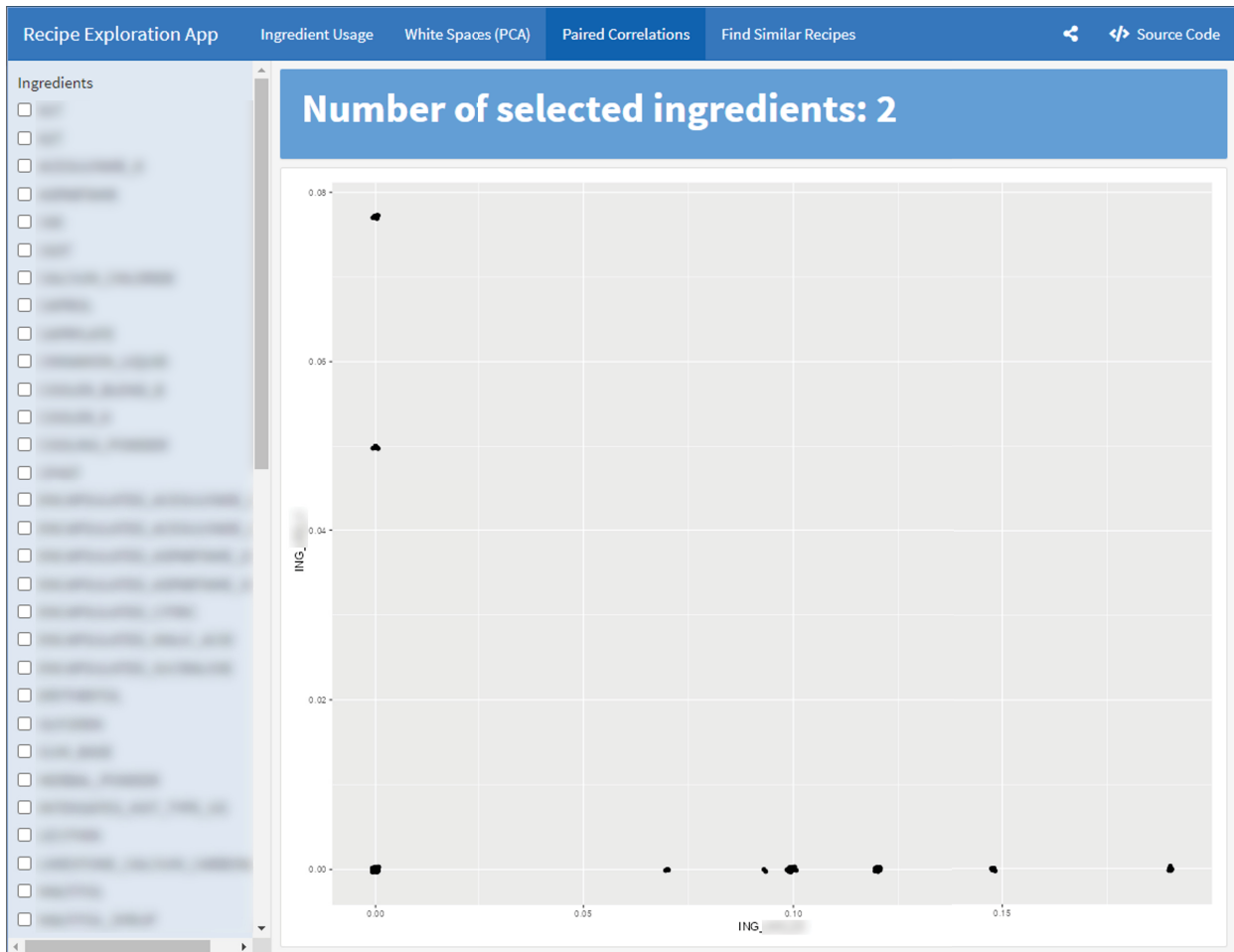


Figura 6: Visualização de distribuições e correlações pareadas (Aba 3) - caso de dois ingredientes selecionados

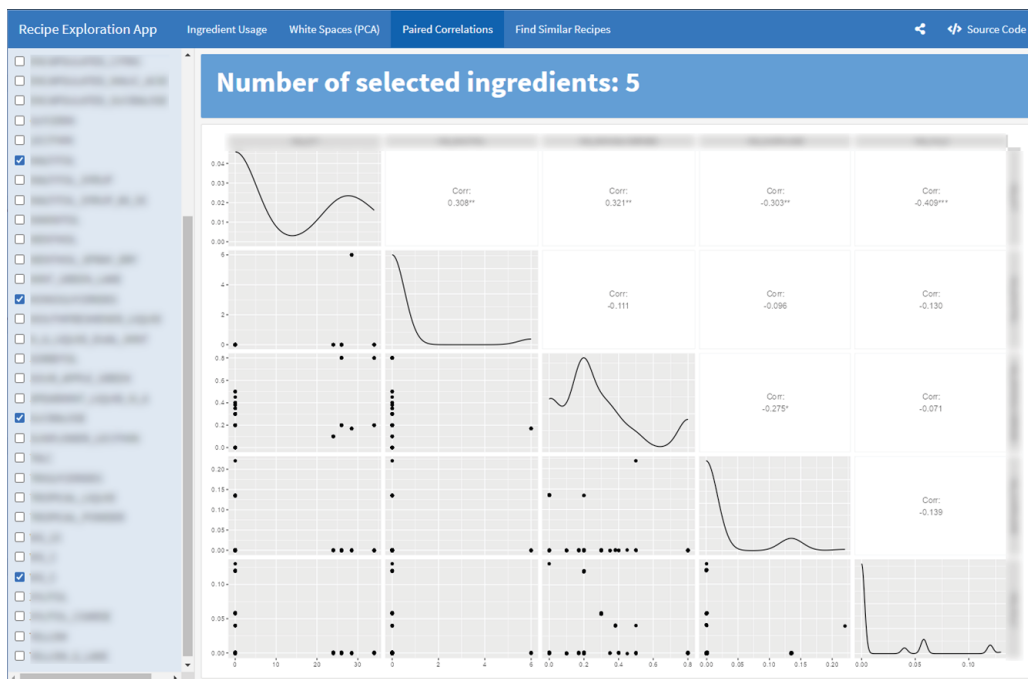


Figura 7: Visualização de distribuições e correlações pareadas (Aba 3) - caso entre três e seis ingredientes selecionados

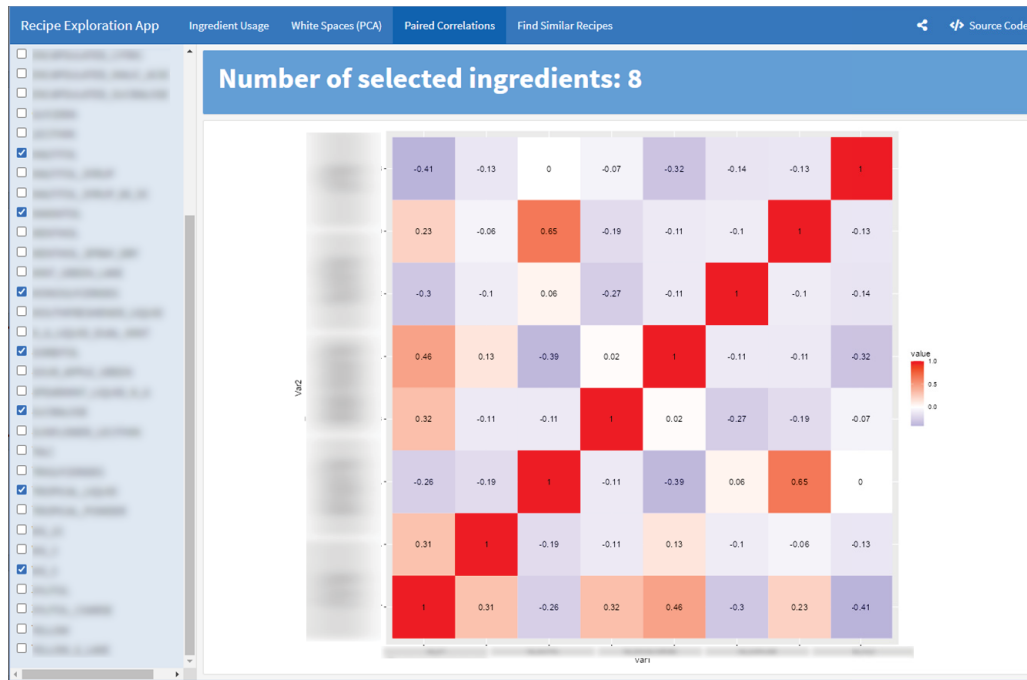


Figura 8: Visualização de distribuições e correlações pareadas (Aba 3) - caso de mais de seis ingredientes selecionados

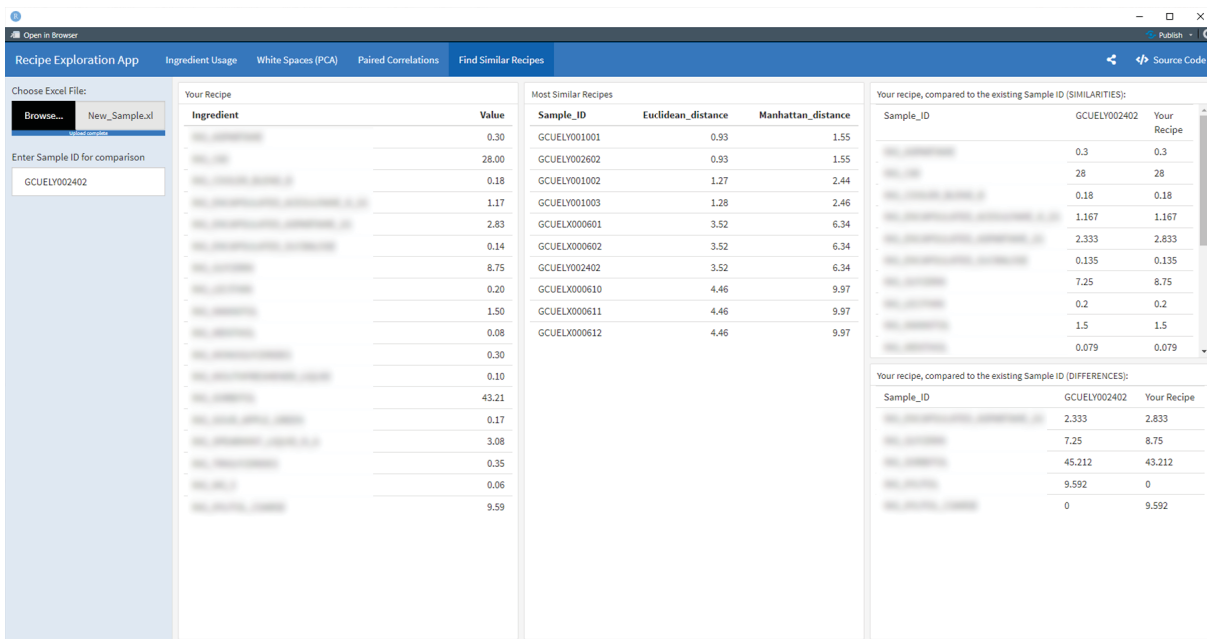


Figura 9: Buscador de receitas similares e Comparador de Receitas (Aba 4)

- [9] H et. al Tian. Is chatgpt the ultimate programming assistant – how far is it?, 2023.
- [10] H. et al. Wickham. Welcome to the tidyverse. *Journal of Open Source Software*, 4(43):1686, 2019.
- [11] Terry Yue Zhuo, Yujin Huang, Chunyang Chen, and Zhenchang Xing. Red teaming chatgpt via jailbreaking: Bias, robustness, reliability and toxicity, 2023.