



**Universidade de Brasília  
Departamento de Estatística**

**Dashboard da Graduação do Departamento de Estatística  
Shiny e Streamlit**

**Gabriel Reis Lourenço Nogueira**

Orientador: Prof. Dr. James Matos Sampaio

**Brasília  
2021**



**Gabriel Reis Lourenço Nogueira**

**Dashboard da Graduação do Departamento de Estatística  
Shiny e Streamlit**

Orientador: Prof. Dr. James Matos Sampaio

Relatório apresentado para o Departamento de Estatística da Universidade de Brasília como parte dos requisitos necessários para obtenção do grau de Bacharel em Estatística.

**Brasília  
2021**



# Agradecimentos

Aos meus pais Lusmere e Juarez, que mesmo com minhas duas trocas de cursos na UnB e intercâmbio na China antes de me encontrar na Estatística, permaneceram inabaláveis no apoio e na esperança de que esse momento fosse chegar. Devo tudo o que sou a vocês e não há sorte maior do que a de tê-los como pais.

À Tamara, minha namorada, um exemplo de pessoa em todos os sentidos e o melhor presente que a UnB me deu.

À ESTAT, por todo o crescimento profissional e pessoal proporcionado e, principalmente, pelas pessoas incríveis que colocou em meu caminho, com as quais iniciei a jornada mais desafiante e edificante da minha vida, a empresa Tukey Data.

Ao professor James Matos Sampaio, por ter iniciado esse e outros projetos relevantes e inovadores para o Departamento, e que mesmo com muitas outras responsabilidades, foi um ótimo orientador. E a todos os professores que contribuíram direta ou indiretamente para a minha formação.

Ao Departamento de Estatística, à Universidade de Brasília e todo o seu corpo docente, que sempre proporcionaram um ensino de qualidade.



# Resumo

Na era digital em que vivemos, é essencial o uso de dados e de tecnologia para embasar decisões. No entanto, o Departamento de Estatística da UnB não tem acesso rápido a dados e análises de qualidade sobre seus alunos e suas disciplinas. Este trabalho teve por objetivo a construção de *dashboards* interativos e informativos para o Departamento de Estatística da UnB usando as linguagens de programação R e Python.

Foram utilizados dados de 1994 até 2021 dos resultados de alunos da UnB em disciplinas do Departamento de Estatística e do Departamento de Matemática, com uso das linguagens R e Python e de várias bibliotecas disponíveis para elas no tratamento e na visualização dos dados.

Foram desenvolvidos e comparados um *dashboard Shiny* e um *dashboard Streamlit*, implementados com as linguagens R e Python respectivamente, com gráficos e informações gerais do departamento, dos professores, das disciplinas do Departamento de Estatística e do Departamento de Matemática que são obrigatórias para alunos da Estatística.

Ambos as tecnologias permitiram a construção de painéis interativos informativos e de boa performance, com diferenças na facilidade de desenvolvimento e de customização.

**Palavras-chave:** dashboard, rstats, python, shiny, streamlit.





# Abstract

In the digital age we live in, the use of data and technology to support decisions is essential. However, the UnB Statistics Department does not have quick access to quality data and analysis about its students and its subjects. This paper aimed to build interactive and informative dashboards for the UnB Statistics Department using **R** and **Python** programming languages.

Data from 1994 to 2021 of the results of UnB students in subjects from the Department of Statistics and the Department of Mathematics were analysed, using the **R** and **Python** languages and various libraries available for them in the treatment and visualization of the data.

A **Shiny** dashboard and a **Streamlit** dashboard were developed and compared, implemented with **R** and **Python** languages respectively, with graphs and general information of the department, professors, subjects from the Department of Statistics and from the Department of Mathematics which are mandatory for students of Statistics.

Both technologies allowed the construction of informative and well-performing interactive dashboards, with differences in the ease of development and customization.

**Keywords:** dashboard, rstats, python, shiny, streamlit.



## Lista de Ilustrações

1	Primeiras observações do banco de dados do SIGRA. . . . .	14
2	Primeiras observações do banco de dados do SIGAA. . . . .	14
3	Dados do SIGRAA, com inconsistências nas primeiras observações: escritas diferentes em <i>Disciplina</i> , observações duplicadas por troca de curso. . . . .	15
4	Dados do SIGAA, com padrão diferente para <i>Período</i> , <i>Horário</i> , ausência de <i>CPF</i> e utilização de <i>Frequência</i> em vez de <i>Faltas</i> . . . . .	16
5	Exemplos de elementos da interface de usuário do <i>ui</i> . . . . .	17
6	Dashboard Shiny - Página Inicial (Estatística) . . . . .	21
7	Dashboard Shiny - Disciplinas do bacharelado . . . . .	22
8	Dashboard Shiny - Disciplinas do bacharelado com filtros . . . . .	23
9	Dashboard Shiny - Disciplinas do bacharelado com valores absolutos . . . . .	23
10	Dashboard Shiny - Disciplinas do bacharelado sem menções . . . . .	24
11	Dashboard Shiny - Disciplinas de serviço . . . . .	25
12	Dashboard Shiny - Disciplinas de serviço com valores absolutos . . . . .	25
13	Dashboard Shiny - Disciplinas da Matemática . . . . .	26
14	Dashboard Shiny - Disciplinas da Matemática com valores absolutos . . . . .	27
15	Dashboard Shiny - Professores . . . . .	28
16	Dashboard Shiny - Professores . . . . .	28
17	Dashboard Shiny - Alunos por professor . . . . .	29
18	Dashboard Streamlit - Tela inicial . . . . .	30
19	Dashboard Streamlit - Disciplinas do bacharelado . . . . .	31
20	Dashboard Streamlit - Menções SS das disciplinas do bacharelado . . . . .	32
21	Dashboard Streamlit - Disciplinas de serviço . . . . .	33
22	Dashboard Streamlit - Disciplinas de serviço em números absolutos . . . . .	34
23	Dashboard Streamlit - Disciplinas de serviço em números absolutos . . . . .	35
24	Dashboard Streamlit - Professores . . . . .	36



## SUMÁRIO

<b>1 INTRODUÇÃO</b>	13
<b>2 METODOLOGIA</b>	14
2.1 BANCOS DE DADOS	14
2.2 LEITURA E TRATAMENTO DOS DADOS	15
2.3 DASHBOARD SHINY	16
2.4 DASHBOARD STREAMLIT	19
<b>3 RESULTADOS</b>	20
3.1 DASHBOARD SHINY	20
3.1.1 Estatística	20
3.1.2 Disciplinas do Bacharelado	21
3.1.3 Disciplinas de Serviço	24
3.1.4 Disciplinas da Matemática	26
3.1.5 Professores	27
3.2 DASHBOARD STREAMLIT	29
3.2.1 Estatística	29
3.2.2 Disciplinas do Bacharelado	30
3.2.3 Disciplinas de Serviço	33
3.2.4 Disciplinas da Matemática	34
3.2.5 Professores	35
3.3 COMPARAÇÃO SHINY VS. STREAMLIT	36
3.3.1 Dificuldade e Velocidade de Desenvolvimento	37
3.3.2 Customização	37
3.3.3 Experiência do Usuário	38
3.3.4 Hospedagem	38
<b>4 CONCLUSÃO</b>	39
<b>REFERÊNCIAS</b>	40

## 1 INTRODUÇÃO

Apesar de formar estatísticos e contar com profissionais que são referências nacionais na área, o Departamento de Estatística da Universidade de Brasília (UnB) carece de dados e análises de qualidade sobre as disciplinas ministradas, os professores e o desempenho de seus alunos. Informações essas que são essenciais para uma boa compreensão da realidade do departamento e para a identificação de intervenções necessárias para proporcionar uma melhor experiência para discentes e docentes.

Poderia-se tentar resolver o problema por meio de relatórios longos com gráficos e análises estáticas. Mas o estágio avançado das linguagens de programação e da ciência de dados possibilita o desenvolvimento e a disponibilização de painéis interativos e dinâmicos de dados, ou *dashboards*, que tornam extremamente acessível e intuitivo a manipulação de dados e a consequente revelação de informações valiosas para usuários, mesmo aqueles que não têm muita familiarização com estatística.

Para se chegar a um *dashboard* informativo e esclarecedor, o primeiro passo é se valer de bons dados, o que por si só já representa um enorme desafio, pois o departamento não tem fácil acesso a essas informações. E mesmo quando consegue, os dados são gerados por múltiplos sistemas de características diferentes, com organização não ideal, duplicados e com erros. Portanto foi imprescindível um trabalho inicial profundo de compreensão, limpeza e organização dos dados.

Somente então foi possível partir para a construção do *dashboard*, que se iniciou na escolha entre várias alternativas de linguagens e plataformas disponíveis no mercado. Como o Bacharelado em Estatística da UnB fornece excelente formação na linguagem R, **Shiny**, que é uma plataforma para criação de aplicativos web consagrada no meio científico, é uma opção óbvia. Entretanto é inegável a relevância da linguagem Python para a ciência de dados e suas vantagens em algumas questões quando comparada ao R. Portanto, além do **Shiny**, foi elaborado, também, um painel interativo de dados com **Streamlit**, um *framework* de código aberto para elaboração de aplicativos web em Python.

Nesse contexto, o objetivo do presente trabalho é organizar e apresentar de maneira interativa e elucidativa dados da graduação pertinentes para o Departamento de Estatística da UnB, comparando duas opções de grande relevância atual para criação de *dashboards*: **Shiny** e **Streamlit**.

## 2 METODOLOGIA

A construção de um bom *dashboard* envolve várias etapas. A primeira delas é a obtenção, a exploração e a compreensão das informações que serão exibidas. Em seguida, é necessário tratar os dados para corrigir problemas e colocá-los em formatos adequados para análises e exposições. Só então é possível elaborar painéis interativos e informativos.

### 2.1. BANCOS DE DADOS

Os dados utilizados foram obtidos junto à Secretaria de Administração Acadêmica (SAA) da UnB. São dois bancos principais, gerados por dois sistemas diferentes. O primeiro traz mais de 380 mil observações, referentes ao período de 1994 a 2020 (verão) e fornecidas pelo Sistema de Informações Acadêmicas de Graduação (SIGRA). O segundo, com cerca de 11 mil observações, compreende o período de 2020 a 2021 (1º semestre), quando a UnB migrou para o Sistema Integrado de Gestão de Atividades Acadêmicas (SIGAA).

Ambos os bancos trazem os resultados de alunos da graduação nas disciplinas oferecidas pelo departamento de Estatística e naquelas oferecidas pelo departamento de Matemática que são obrigatórias para alunos do Bacharelado em Estatística. Cada observação identifica o aluno, a disciplina cursada e o seu desempenho. Portanto são informações como nome, matrícula, curso do aluno, porcentagem de faltas e menção final. Traz ainda nome, período e turma da disciplina, assim como o professor que a ministrou. As primeiras observações de cada banco estão expostas nas figuras a seguir, com nome do aluno, matrícula e nome do professor criptografados para preservação de identidade.

Nome	Matrícula	Curso	Cód. Disciplina	Disciplina	Semestre em que cursou a disciplina	Professor	Turma	Hora Início	Hora Fim	Menção	Faltas
GhJArfSgmoJQngt	^1#868\$	Biblioteconomia	115011	Estatística Aplicada		19941 PuykfgztoufIgxopufHgxhuvg	C	1899-12-31T18:00:00Z	1899-12-31T19:50:00Z	SS	0,06
GhkrfXoigxufVosktzgfLkxkzozfJgfluyzg	*98687\$	Matemática	113042	Cálculo 2		19941 RAoFfJkUroBkoxgfGsuox	E	1899-12-31T19:00:00Z	1899-12-31T20:40:00Z	MM	0,23
GhkrfXoigxufVosktzgfLkxkzozfJgfluyzg	*98687\$	Matemática	113042	Cálculo 2		19941 RAoFfJkUroBkoxgfGsuox	E	1899-12-31T21:00:00Z	1899-12-31T22:40:00Z	MM	0,23
GhkrfXoigxufVosktzgfLkxkzozfJgfluyzg	*98687\$	Matemática	115045	Probabilidade e Estatística		19941 UzgBoufYoeukyftHgxhuvgLornu	D	1899-12-31T21:00:00Z	1899-12-31T22:50:00Z	II	0,2
GhixxRsozgfJkUroBkoxg	*98687#	Engenharia Civil	113051	Cálculo 3		19941 PuykfgkxjufYerBg	C	1899-12-31T10:00:00Z	1899-12-31T11:50:00Z	SS	0,03
GloBgrfRuvkyfSutzkoxu	*87*6%7	Engenharia Civil	113051	Cálculo 3		19941 PuykfgkxjufYerBg	C	1899-12-31T10:00:00Z	1899-12-31T11:50:00Z	II	0,13

Figura 1: Primeiras observações do banco de dados do SIGRA.

Nome	Matrícula	Curso	Disciplina	Semestre	Professor	Turma	Horário	Menção	Frequência
GgxuufTgBgxxuf1gxmg	7366687*99	FISICA	CÁLCULO 1	20201	IgsorgfJkUroBkoxgf1okoxg	G	2N12 3N34 5N34	MM	100,0
GgxuufTgBgxxuf1gxmg	7366687*99	FISICA	CÁLCULO 1	20201	PuykfrAoyfZkxArfIgxokzkxou	G	2N12 3N34 5N34	MM	100,0
GgxuufTgBgxxuf1gxmg	7366687*99	FISICA	CÁLCULO 2	20202	NkxowAkfluyzgfJuyfXkoy	01A02	2N12 3N34	MM	100,0
GgxuufTgBgxxuf1gxmg	7366687*99	FISICA	INTRODUCAO A ALGEBRA LINEAR	20202	GhixxuzfSgygEuyofLgxogfUngyno	G	3N12 5N12	MM	100,0
GhJArfngttgt	8686f8#81	ENGENHARIA	CÁLCULO 1	20202	XoigxufXgsuyfLxgmkrno	28A	248M34	MS	90,0
GhkrfgrkxyufUmrogxo	86868986%	FISICA	CÁLCULO 1	20201	PuykfrAoyfZkxArfIgxokzkxou	06A	2N12 3N34	SR	100,0

Figura 2: Primeiras observações do banco de dados do SIGAA.

Rapidamente foi possível identificar diversas inconsistências entre os bancos de dados, como nomes das disciplinas com formatações diferentes, o horário codificado e

ausência do código da disciplina no banco do SIGAA, percentual de faltas em um e frequência no outro, entre outras. Também foram identificados vários problemas e inconsistências no mesmo banco, como múltiplas inserções para alunos que trocaram de curso, de matrícula ou que tiveram suas menções revisadas e até duplicação de observações pelo número de créditos da matéria.

Assim sendo, torna-se crucial uma análise inicial minuciosa para compreensão e correção dos problemas nos dados e compatibilização entre os dois bancos.

## 2.2. LEITURA E TRATAMENTO DOS DADOS

Para a leitura e tratamento dos dados, foi utilizado tanto a linguagem R versão 4.0.4, por meio do software RStudio, quanto a linguagem Python, versão 3.9.4, por meio do software Visual Studio Code.

São vários os pacotes disponíveis para R que facilitam o processo de tratamento e limpeza dos dados. Entre alguns que foram utilizados estão: `tidyverse`, `lubridate` e `janitor`. Para a correção de inconsistências textuais, foi necessário também conhecimento em expressões regulares e o uso de pacotes como `stringr` e `stringi`. Já em Python, a biblioteca `Pandas` é suficiente para a grande maioria das operações de tratamento e limpeza de dados.

Diversos problemas de inconsistências foram encontrados na análise exploratória dos dados. Entre eles, destaca-se:

- Nomes de disciplinas, professores e alunos com escritas diferentes (com e sem acento, maiúsculas e minúsculas, nomes de disciplinas antigos e atuais, erros de digitação);
- Observações duplicadas por várias razões (troca de curso ou matrícula do aluno, revisão de menção, repetição por número de créditos da disciplina, entre outras);
- Diferenças entre padrões e variáveis utilizadas no sistema antigo (SIGRA) e sistema novo (SIGAA).

Alguns desses problemas podem ser vistos já nas primeiras observações dos bancos de dados:

```
> head(historico)
# A tibble: 6 x 14
  Matrícula CPF Nome Curso `Cód. Disciplina` Disciplina Período `Matrícula Prof` Professor Turma Menção Faltas `Hora início` `Hora Fim`
  <chr> <chr> <chr> <chr> <chr> <chr> <chr> <chr> <chr> <chr> <chr> <chr> <chr> <chr>
1 9465957 - [REDACTED] Bibliotecono_ 115011 Estatística Aplica_ 19941 514730 [REDACTED] C SS 6% 18:00 19:50
2 9320717 - [REDACTED] Matemática 115045 Probabilidade e Es_ 19941 715158 [REDACTED] D II 20% 21:00 22:50
3 9202871 - [REDACTED] Estatística 115932 METODOS ESTATISTIC_ 19941 118036 [REDACTED] A MI 17% 14:00 15:50
4 9319433 - [REDACTED] Engenharia E_ 115045 Probabilidade e Es_ 19941 954527 [REDACTED] C MM 5% 10:00 11:50
5 8804516 [REDACTED] Ciências Bio_ 115070 BIOESTATISTICA 19941 900346 [REDACTED] B MS 8% 08:00 09:50
6 8804516 [REDACTED] Nutrição 115070 BIOESTATISTICA 19941 900346 [REDACTED] B MS 8% 08:00 09:50
```

Figura 3: Dados do SIGRAA, com inconsistências nas primeiras observações: escritas diferentes em *Disciplina*, observações duplicadas por troca de curso.



```
> head(historico_sigaa)
# A tibble: 6 x 11
  'Nome do Aluno'      Matrícula Curso      Disciplina      Período 'Matrícula do Profes...' 'Nome do Professor'      Turma Horário Menção Freqüência
  <chr>                <chr>      <chr>      <chr>          <chr> <chr>      <chr>          <chr> <chr> <chr>
1 180,114,891 ESTATÍSTICA INTRODUÇÃO A PROBABIL... 20,201 2279025 B 2M34 4L MM 100
2 190,041,889 ENGENHARIA MECÂNICA PROBABILIDADE E ESTATÍ... 20,201 1071560 AB 4M12 6L SS 100
3 170,027,864 ESTATÍSTICA MÉTODOS ESTATÍSTICOS 2 20,201 404700 01A 24T23 MM 92
4 190,133,619 ENGENHARIA DE REDES DE CO... PROBABILIDADE E ESTATÍ... 20,201 1071560 AB 4M12 6L MS 100
5 120,049,031 ENGENHARIA ELETRÔNICA PROBABILIDADE E ESTATÍ... 20,201 1953590 BB 2M34 4L SR 100
6 160,063,728 ESTATÍSTICA CONSULTORIA 20,201 NA NA 01A 35M34 SS 100
```

Figura 4: Dados do SIGAA, com padrão diferente para *Período*, *Horário*, ausência de *CPF* e utilização de *Freqüência* em vez de *Faltas*.

Entre os principais tratamentos que foram empregados para corrigir e compatibilizar os dados, destaca-se a padronização dos nomes, com remoção de acentos e transformação em caixa baixa, e a conversão das variáveis no tipo adequado (categórica, numérica, data, etc.). Para compatibilização dos dois bancos (SIGAA e SIGRA), também foi necessário padronizar os nomes das variáveis, decodificar a variável *Horário* do SIGAA e transformar a variável *Freqüência* em *Faltas*.

Após os bancos de dados serem tratados, compatibilizados e juntados, foi feita uma análise minuciosa de dados duplicados, dificultada pela ausência de data ou hora de inserção dos dados e pela profusão de situações diferentes. No caso de duplicações devido a troca de matrícula do aluno, foram mantidas apenas as observações da matrícula mais recente. Quando a repetição foi por conta de troca de curso, conservou-se as observações do curso de semestres mais recentes. Também houve duplicações por conta de revisão de menção, situação em que se optou por manter a menção mais alta.

Por fim, terminada a limpeza dos dados, os acentos foram recolocados e os dados foram convertidos em formatos adequados e agradáveis para exposição nos *dashboards*. Todo o código utilizado no processo pode ser conferido em <https://github.com/unb-est/est-shiny>.

### 2.3. DASHBOARD SHINY

*Shiny* é um pacote de linguagem R lançado em julho de 2012 pela equipe do RStudio que permite o desenvolvimento de aplicativos interativos web. É um pacote projetado para que cientistas de dados possam criar painéis interativos de dados apenas com R, sem necessidade de conhecimento de tecnologias típicas de desenvolvimento web, como HTML, CSS ou Javascript. No entanto, é possível utilizar essas tecnologias para personalizar e aprimorar o aplicativo (WICKHAM, 2020).

A lógica de programação do *Shiny* é basicamente a mesma do desenvolvimento de software ou web padrão, com separação entre o código responsável pela interface do usuário (ou *User Interface* - *UI*), aquilo que o usuário do aplicativo vê, também conhecido como *front-end*, e o código responsável pelos cálculos necessários para manipulação e

visualização dos dados, conhecido como *back-end*.

A interface de usuário é controlada pela função *ui*, onde o programador cria e posiciona os elementos que serão exibidos na interface do aplicativo e com os quais o usuário interage (também conhecidos como *inputs*). São estruturas como botões, caixa de seleção, caixa de texto, etc., exemplificadas na figura a seguir, juntamente com o código de cada uma.

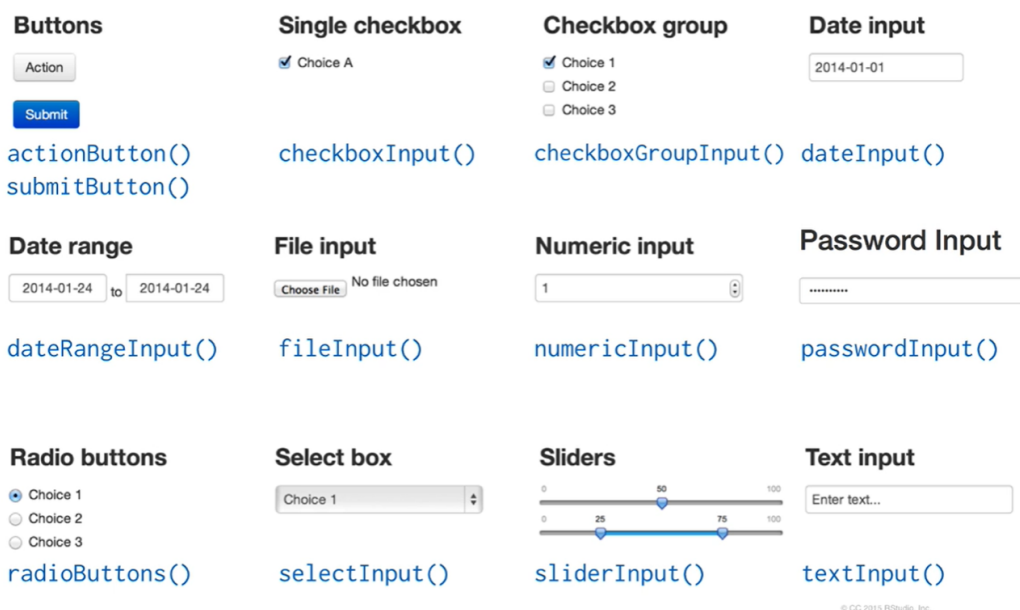


Figura 5: Exemplos de elementos da interface de usuário do *ui*.

Além desses elementos, também é necessário indicar e organizar os espaços para exibição dos gráficos e de imagens na função *ui*. Esses espaços juntamente com os elementos de interação podem ser organizados em uma infinidade de maneiras, com estruturas pré-estabelecidas como painéis e abas, ou livremente em linhas e colunas. Os códigos correspondentes a essa parte externa do aplicativo são bem diferentes da programação comum em R e, portanto, existe uma curva de aprendizado considerável.

Após organizar todos os elementos e espaços da interface do usuário, é necessário programar a manipulação e a visualização dos dados, cujo código deve ficar dentro da função *server* (servidor). Esse é código que faz o “trabalho pesado” de cálculos e manejo de dados, mas que é bem familiar para quem está acostumado a trabalhar com R.

Finalmente, ao executar a função *shinyApp* com os argumentos *ui* e *server*, os elementos da interface de usuário são conectados à parte interna da programação (*back-end*) para produção do aplicativo. Todo o código utilizado pode ser verificado em (<https://github.com/unb-est/est-shiny>).

Para o desenvolvimento do *dashboard* em Shiny, versão 1.6.0, também foi uti-

lizado o `RStudio`, que possibilita tanto a elaboração dos painéis quanto a hospedagem online por meio de sua plataforma *shinyapps.io*.

São inúmeras as possibilidades para edições e customizações adicionais de *dashboards Shiny*. Pode-se optar por utilizar configurações padrões e gráficos estáticos ou, como é o caso deste trabalho, implementar funcionalidades e personalizações mais avançadas por meio de pacotes como `shinyWidgets`, `dashboardthemes` e uso de CSS, bem como gráficos mais dinâmicos e interativos com pacotes como `plotly`.

## 2.4. DASHBOARD STREAMLIT

**Streamlit** é, simplificada, um equivalente de **Shiny** para linguagem **Python**. É um *framework* de código aberto que permite o desenvolvimento de aplicativos web de dados utilizando apenas **Python**. Assim como no **Shiny**, também possibilita o uso de outras tecnologias para aprimoramento e personalização do aplicativo. Foi lançado em março de 2018 e neste trabalho será utilizada a versão 0.80.0.

O **Streamlit** foi desenvolvido para possibilitar o desenvolvimento de aplicativos web em minutos ou poucas horas. Consequentemente, a estrutura e sintaxe de código é muito simples e familiar para quem está acostumado a usar **Python**. Ao contrário do **Shiny**, não há muita diferenciação entre a parte da interface de usuário (*front-end*) e o código da parte interna do aplicativo (*back-end*), com tudo ficando dentro de apenas uma função.

Para criar elementos de interação na interface do usuário e posicioná-los, basta utilizar os métodos adequados da biblioteca *streamlit* importada. Em **Python**, um método é, basicamente, uma função que pertence a um tipo de objeto e acessada por meio do ponto final. Então caso se deseje criar uma caixa de seleção múltipla (*multiselect*) e posicioná-la na barra lateral (*sidebar*) do aplicativo, por exemplo, basta utilizar `streamlit.sidebar.multiselect(...)` e atribuir a uma variável, que armazenará o valor resultante de qualquer interação do usuário com esse elemento.

Para os cálculos e manipulações de dados, não há quase nenhuma novidade para quem sabe realizar análises estatísticas com **Python**. Quando se quer utilizar o resultado de uma interação do usuário com um elemento da interface para determinado cálculo ou manipulação, basta referenciar a variável a que foi atribuída aquele elemento.

No desenvolvimento do *dashboard Streamlit*, foi utilizado o software **Visual Studio Code** versão 1.54.1, que também permite a hospedagem online. Entre as inúmeras bibliotecas disponíveis para construção de gráficos interativos em **Python**, foi utilizado **Plotly**, que também está disponível para **R**.

Visando melhor desempenho e manutenção dos códigos, o que é imprescindível para que o trabalho realizado fique como legado para o departamento, grande esforço foi empreendido na utilização de boas práticas de programação e no uso do **GitHub** para a hospedagem de códigos, arquivos e para o controle de versões, como pode ser conferido em <https://github.com/unb-est/est-streamlit>.

### 3 RESULTADOS

Os resultados obtidos serão apresentados a seguir.

#### 3.1. DASHBOARD SHINY

O *Dashboard Shiny*, apenas com o banco de dados do sistema SIGRA e disciplinas oferecidas pelo Departamento de Estatística, foi iniciado por mim e mais três alunas na disciplina *Tópicos em Estatística 1* do período 1<sup>o</sup>/2020, ministrada pelo professor James e pela professora Ana Maria.

Após os desenvolvimentos em TCC1 e TCC2, o painel interativo traz agora os dados do SIGRA e do SIGAA, assim como disciplinas do Departamento de Matemática cursadas por estudantes do bacharelado de Estatística. O *Dashboard* possui cinco páginas:

1. Estatística: página inicial com resumo de informações gerais do Departamento de Estatística;
2. Disciplinas do Bacharelado: gráficos e informações sobre as disciplinas do departamento cursadas pelos estudantes do bacharelado em Estatística;
3. Disciplinas de Serviço: gráficos e informações sobre as disciplinas do departamento cursadas por alunos de outros cursos;
4. Disciplinas da Matemática: gráficos e informações sobre as disciplinas do Departamento de Matemática obrigatórias para estudantes do bacharelado em Estatística;
5. Professores: com gráficos focados nos professores do departamento.

##### 3.1.1 Estatística

A página inicial (Figura 6), denominada *Estatística*, traz dois gráficos de rosca com a formação dos professores do departamento (doutor ou mestre) e o tipo de disciplina (bacharelado, serviço ou matemática). Logo abaixo, tem-se um gráfico de linha com a quantidade de formandos do Bacharelado em Estatística ao longo do tempo, baseado em dados fornecidos pela secretaria do departamento. Ao se passar o mouse por cima dos pontos, é possível visualizar informações mais detalhadas.

À esquerda dos gráficos, tem-se informações gerais do departamento, como total de estudantes matriculados, de disciplinas ofertadas e de professores, além da aprovação e reprovação média. Esses dados dizem respeito a todos os períodos desde de 1994, mas é possível filtrar para algum período de interesse, por meio da lista suspensa na barra lateral esquerda, que por padrão inicial traz selecionado “Todos”.

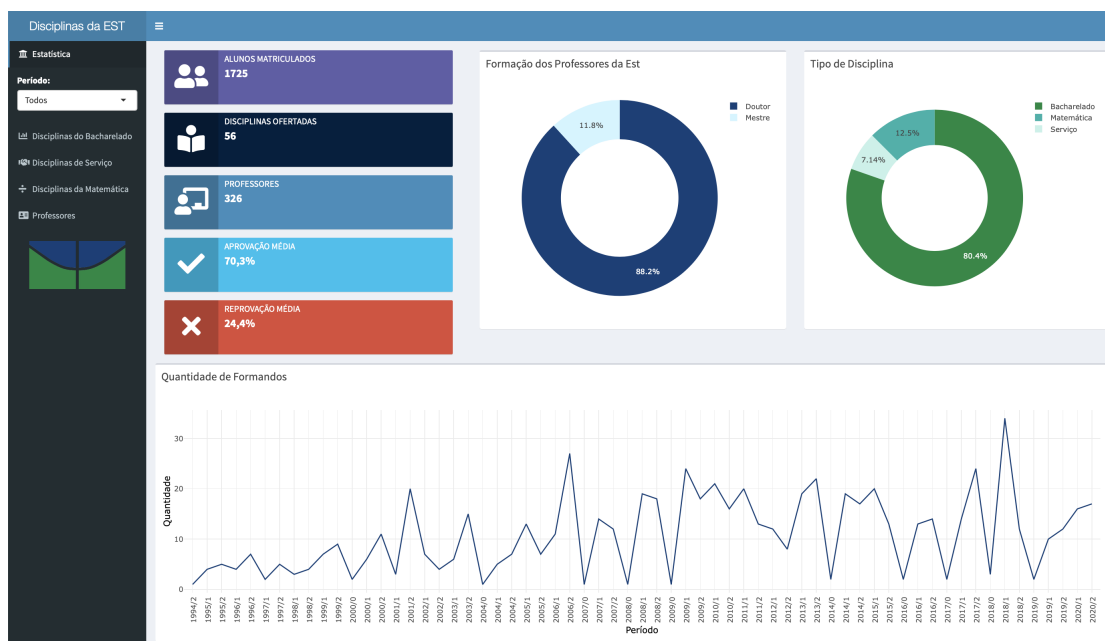


Figura 6: Dashboard Shiny - Página Inicial (Estatística)

### 3.1.2 Disciplinas do Bacharelado

Seguindo a ordem da barra lateral à esquerda, a segunda página (Figura 7) traz dados das disciplinas do bacharelado em Estatística. O destaque é um gráfico de barras empilhadas das menções (cores) obtidas pelos estudantes nas matérias (eixo y), que estão em ordem decrescente de aprovação. Os valores do eixo x representam a proporção de cada menção naquela disciplina. Portanto, é possível observar que a disciplina Laboratório de Estatística 2 é a que possui maior taxa de aprovação, enquanto Modelos Lineares Generalizados é a que possui a menor, considerando-se todos os períodos desde 1994, a menção SR e trancamentos.

Logo abaixo, há um gráfico de linha das menções dos estudantes (em porcentagem) ao longo do tempo. É possível identificar picos que correspondem aos períodos de verão em que há ofertas excepcionais de apenas uma matéria. Fora isso, o gráfico mostra um comportamento relativamente estável das menções ao longo do tempo, quando consideradas todas as disciplinas juntas. Mas é possível notar a redução drástica de reprovações nos últimos dois períodos, que correspondem à situação de ensino remoto durante a pandemia, o que já era esperado, uma vez que nesses períodos é permitido retirar disciplinas até o último dia letivo sem nenhum prejuízo curricular.

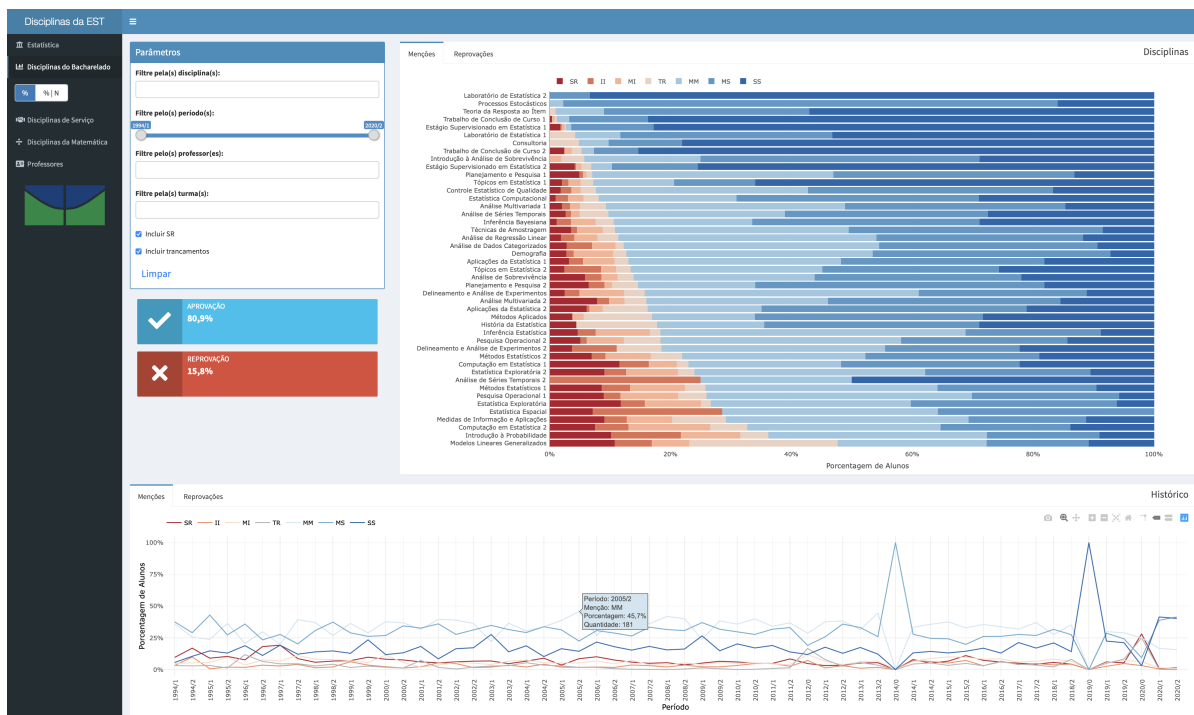


Figura 7: Dashboard Shiny - Disciplinas do bacharelado

Ao se passar o ponteiro do mouse pelos gráficos, é possível observar os valores mais precisos de cada ponto e outras informações. À esquerda do gráfico de barras, há filtros para personalizar a análise. É possível escolher disciplinas para comparação, restringir os períodos, especificar professores, além de incluir ou não a menção SR (sem rendimento) e trancamentos na análise, como se vê na Figura 8. Pode-se ainda remover menções da análise ao se clicar na legenda correspondente.

Na barra lateral esquerda, há um botão que permite alterar os valores exibidos pelos gráficos de porcentagens para valores absolutos (Figura 9). Isso altera significativamente os resultados, dando destaque negativo para aquelas disciplinas obrigatórias dos primeiros semestres do bacharelado, pois são cursadas por muito mais estudantes. No gráfico de linhas, nota-se um aumento mais acentuado no número total de menções de aprovação nos últimos anos.





Caso não se queira tanto detalhamento por menção, cada painel de gráfico possui uma outra aba denominada Reprovações, que mostra as mesmas análises separando-se apenas em reprovação, trancamento e aprovação (Figura 10). No gráfico de linhas, fica ainda mais evidente o aumento no número absoluto de aprovações nos períodos de ensino remoto. Todas as opções de filtros e modificações continuam valendo.

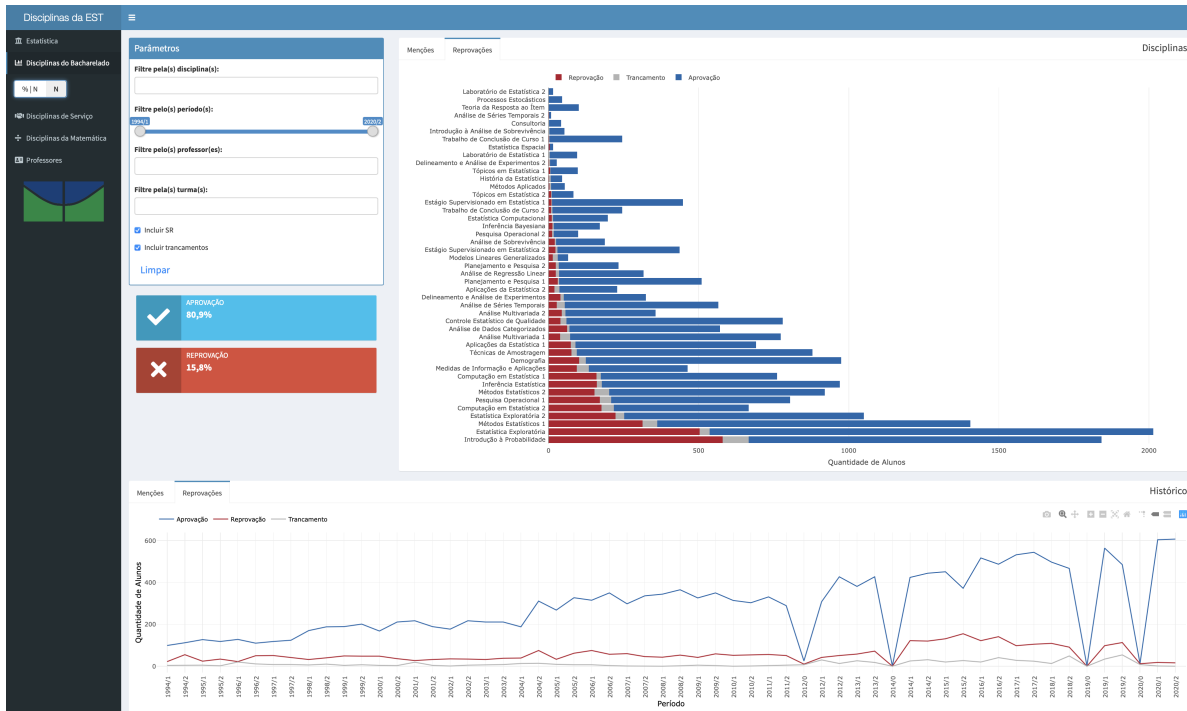


Figura 10: Dashboard Shiny - Disciplinas do bacharelado sem menções

### 3.1.3 Disciplinas de Serviço

A terceira página do *dashboard* (Figura 11) traz as disciplinas de serviço do Departamento de Estatística, que são aquelas oferecidas majoritariamente a outros departamentos. São quatro disciplinas: Probabilidade e Estatística 1 e 2, Bioestatística e Estatística Aplicada.

No quadro abaixo dos filtros, nota-se que a reprovação média dessas disciplinas é maior que a das disciplinas do bacharelado (25,7% contra 15,8%). Também é possível observar que Estatística Aplicada é a disciplina com a menor taxa de aprovação.

Da mesma forma que na página anterior, é possível clicar no botão da barra lateral esquerda para visualizar dados absolutos, como é possível observar na Figura 12. Agora fica evidente porque a disciplina Probabilidade e Estatística 2 tem uma taxa de aprovação muito maior que as outras: o número de estudantes que a cursou é muito menor.

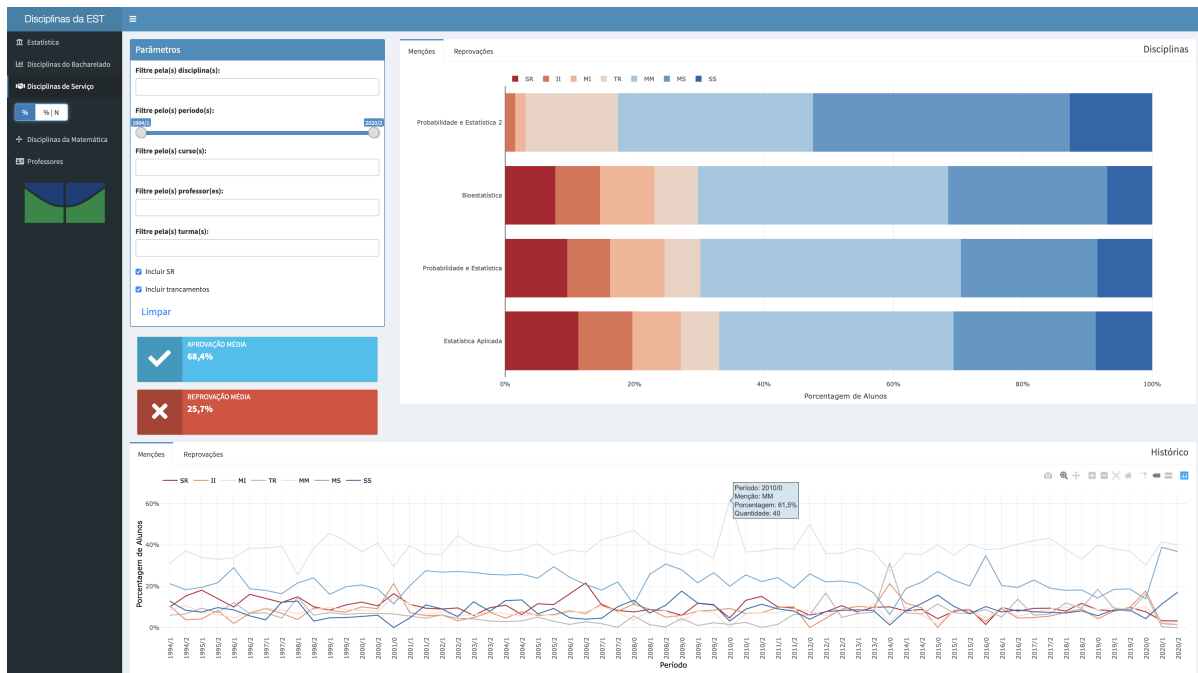


Figura 11: Dashboard Shiny - Disciplinas de serviço



Figura 12: Dashboard Shiny - Disciplinas de serviço com valores absolutos

Com relação aos filtros, a única diferença é que agora também é possível selecionar os cursos de interesse na análise. Os painéis dos gráficos também possuem uma outra aba para análises sem discriminação por menções.

### 3.1.4 Disciplinas da Matemática

As últimas disciplinas a serem analisadas são aquelas ofertadas pelo Departamento de Matemática para os estudantes da Estatística. A página 4 do *Dashboard* traz essa análise, mostrada na Figura 13.

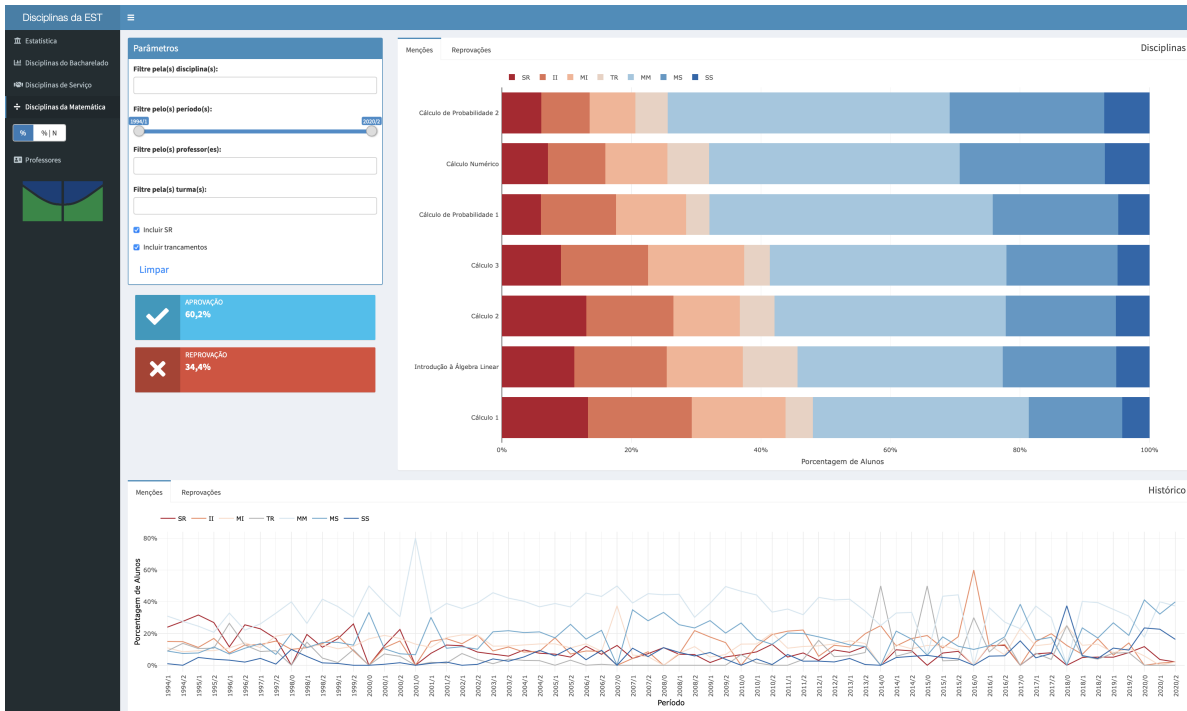


Figura 13: Dashboard Shiny - Disciplinas da Matemática

Imediatamente nota-se que a reprovação média nas disciplinas da Matemática (34,4%) é maior que a das disciplinas de serviço (25,7%) e mais que o dobro que a das disciplinas do bacharelado (15,8%), identificando um grande gargalo para os alunos da Estatística.

Na análise do gráfico de barras, percebe-se que as disciplinais iniciais são aquelas com menor taxa de aprovação. O gráfico de linhas mostra que há grande variação na taxa de cada menção ao longo dos períodos.

Da mesma forma que nas páginas anteriores, é possível visualizar os dados em valores absolutos, como é possível ver na Figura 14. No gráfico de barras, o resultado não muda muito, com as matérias iniciais permanecendo com o maior número de reprovações. No gráfico de linhas, é possível notar que, ao contrário das disciplinas da Estatística, não houve um aumento no número total de aprovados nos períodos de ensino remoto, apesar da redução do número de reprovados, indicando, possivelmente, um grande número de estudantes retirando disciplinas.

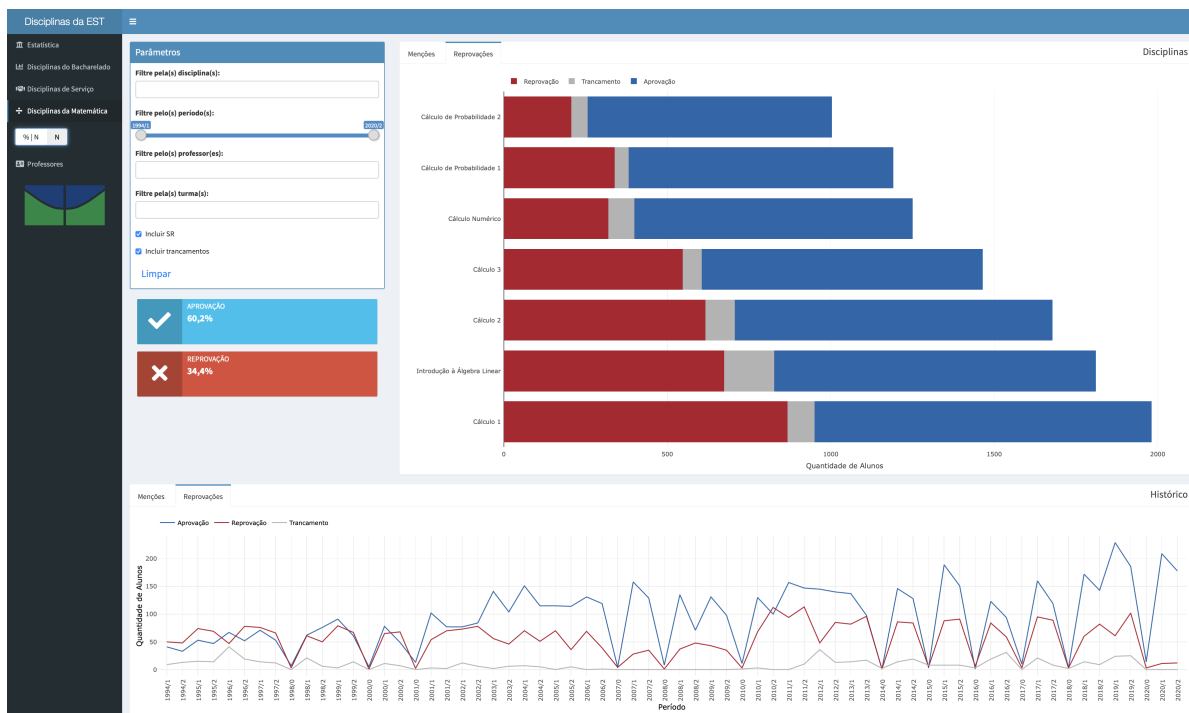


Figura 14: Dashboard Shiny - Disciplinas da Matemática com valores absolutos

### 3.1.5 Professores

A última página traz informações dos professores do departamento. Como é possível observar na Figura 15, o gráfico de barras empilhadas de menções agora mostra os professores no eixo y, cujos nomes foram criptografados para proteção de identidade. Já o gráfico de linhas traz a quantidade média de alunos por professor ao longo dos períodos, com média geral de 68,8.

O padrão inicial é filtrar apenas para professores ativos do departamento, mas isso é facilmente alterado clicando-se na caixa de seleção, resultando na Figura 16. Os outros filtros, seleções e a possibilidade de visualizar valores absolutos são similares aos apresentados anteriormente.

Da mesma forma que nas outras páginas, é possível visualizar o mesmo gráfico de barras sem discriminação de menções (aba Reprovações). No entanto, agora há um terceira aba com a quantidade média de alunos por professor, mostrada na Figura 17.

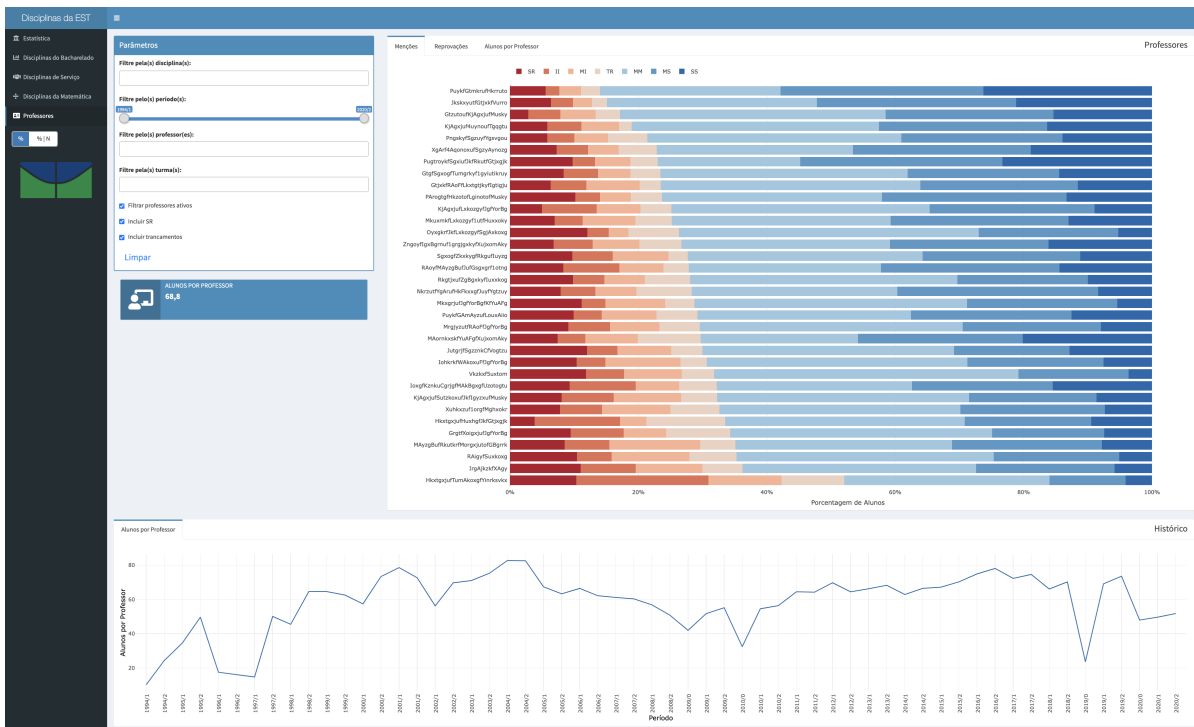


Figura 15: Dashboard Shiny - Professores

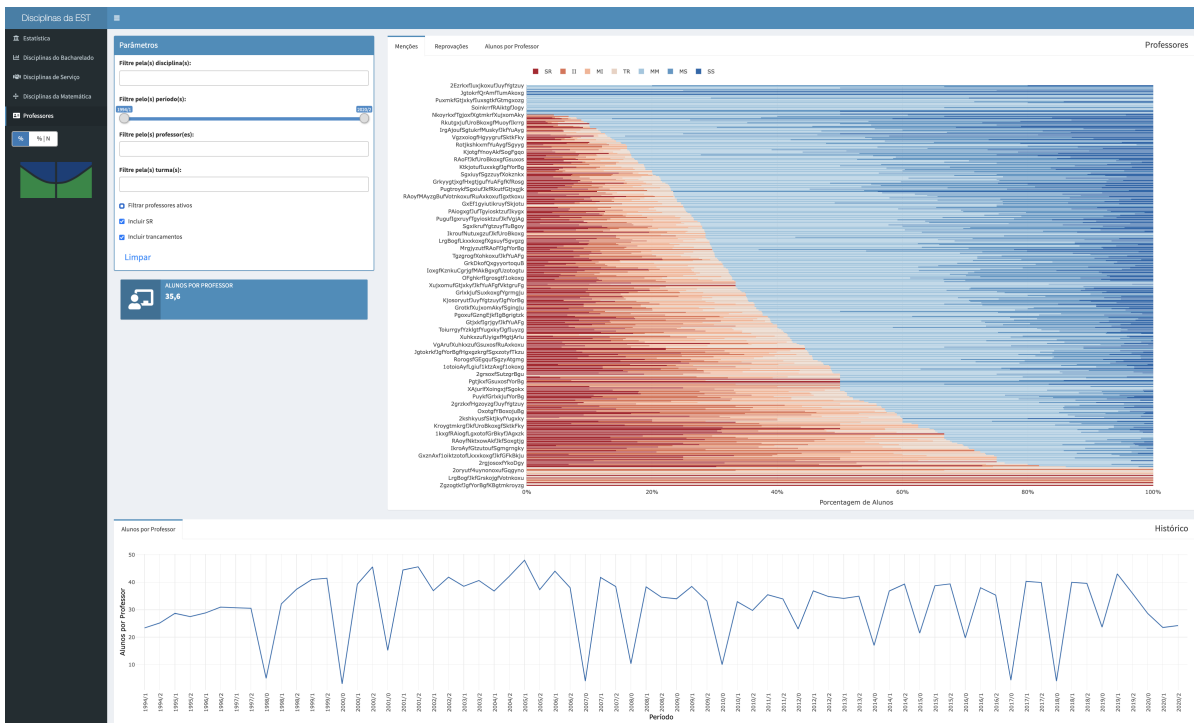


Figura 16: Dashboard Shiny - Professores

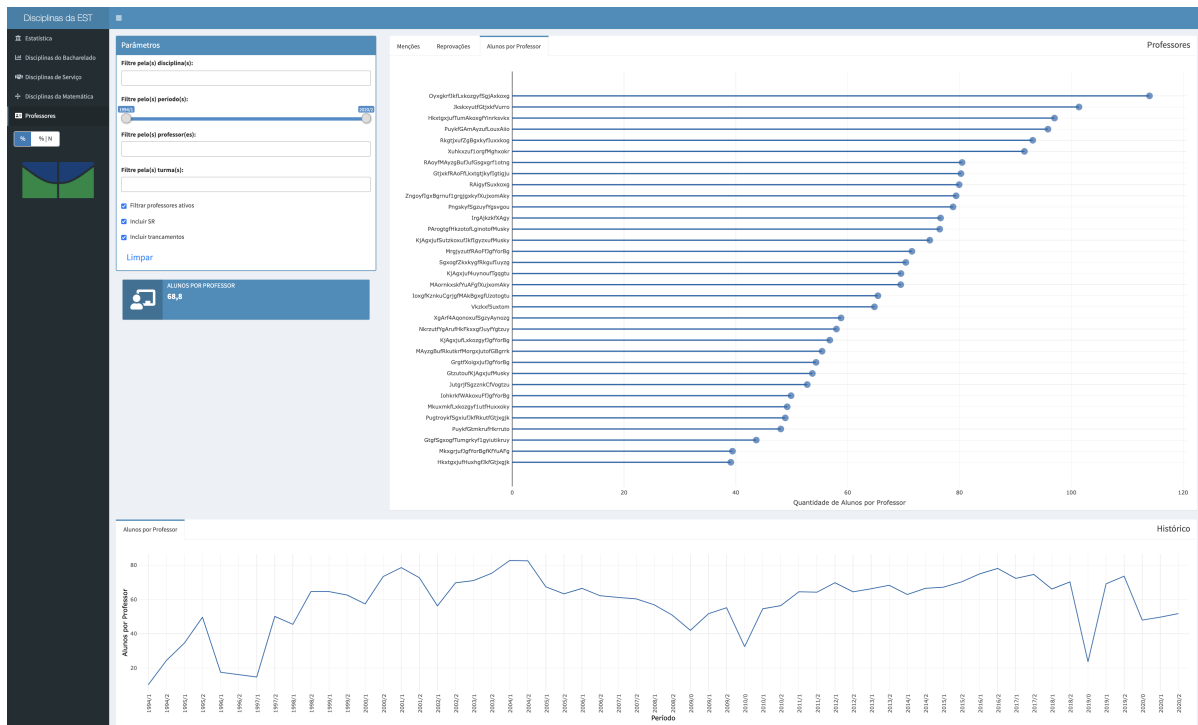


Figura 17: Dashboard Shiny - Alunos por professor

O aplicativo web **Shiny** está disponível em <https://bit.ly/unb-est-app>.

### 3.2. DASHBOARD STREAMLIT

O *Dashboard Streamlit* possui estrutura similar ao *Dashboard Shiny*, com cinco páginas ou telas, baseadas nos mesmos dados:

1. Estatística: página inicial com informações gerais do Departamento de Estatística;
2. Disciplinas do Bacharelado: gráficos das disciplinas do departamento cursadas pelos estudantes do Bacharelado de Estatística;
3. Disciplinas de Serviço: gráficos das disciplinas do departamento cursadas por estudantes de outros cursos;
4. Disciplinas da Matemática: gráficos das disciplinas do Departamento de Matemática obrigatórias para estudantes do bacharelado em Estatística;
5. Professores: gráficos dos professores do departamento.

#### 3.2.1 Estatística

Assim como no **Shiny**, a página inicial do **Streamlit** traz dois gráficos de rosca com a formação dos professores do departamento e o tipo de disciplina, seguidos pelo

gráfico de linhas da quantidade de formandos ao longo do tempo, mostrados na Figura 18.

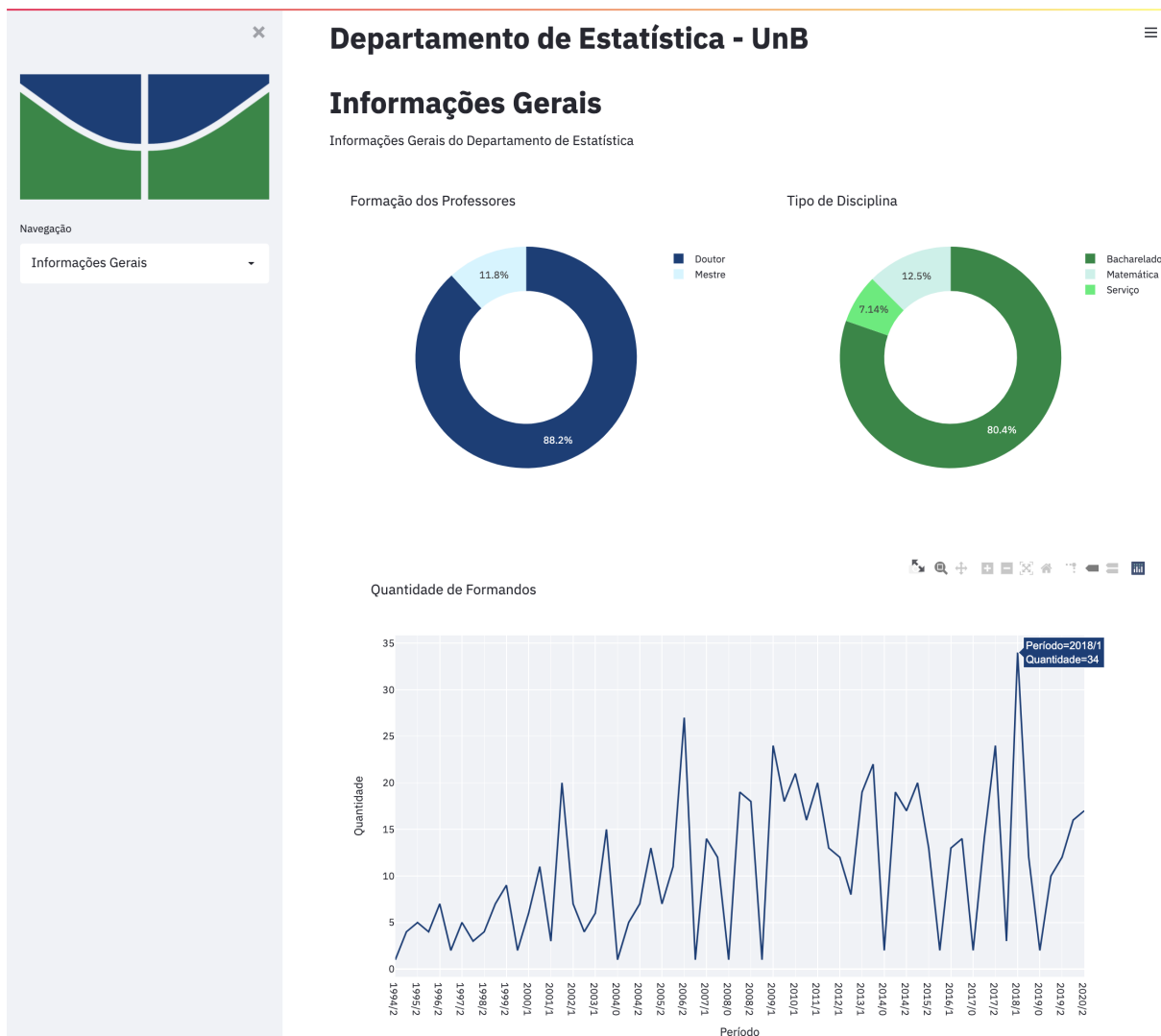


Figura 18: Dashboard Streamlit - Tela inicial

### 3.2.2 Disciplinas do Bacharelado

A navegação entre páginas agora é feita por meio de um menu suspenso na barra lateral esquerda. A segunda opção leva para as informações das disciplinas do bacharelado, que ao ser selecionada, exibe os filtros, o gráfico de barras empilhadas e a linha temporal de menções, apresentados na Figura 19,

Os gráficos possuem as mesmas opções de filtros, interações e modificações do *Dashboard Shiny*, como inclusão ou não de SRs, trancamentos e exibição de gráficos com valores absolutos. Clicando-se duas vezes sobre um valor de menção da legenda, é possível visualizá-la individualmente, pois todos os outros valores são removidos, como se vê na

Figura 20.



Figura 19: Dashboard Streamlit - Disciplinas do bacharelado



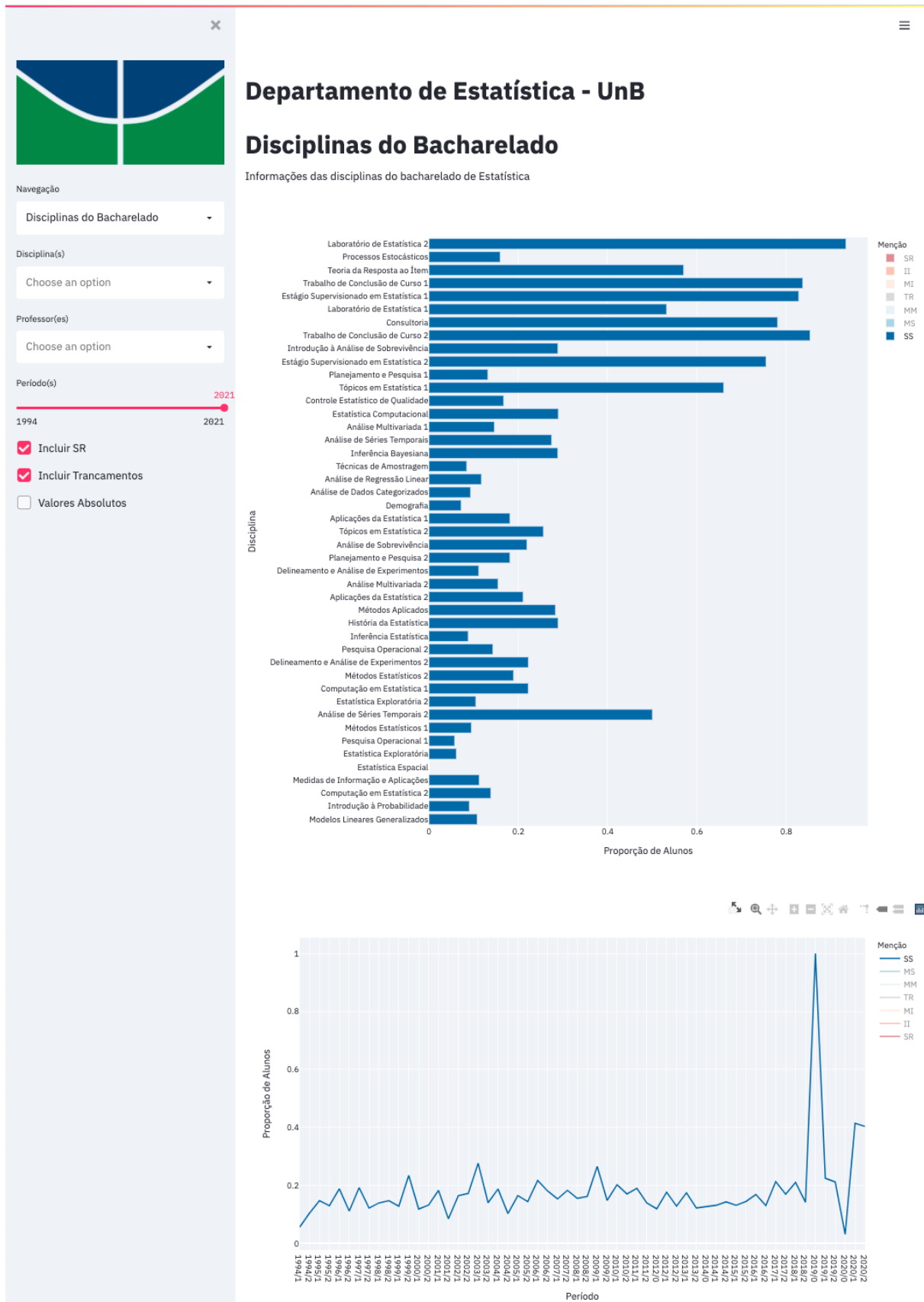


Figura 20: Dashboard Streamlit - Menções SS das disciplinas do bacharelado

### 3.2.3 Disciplinas de Serviço

Seguindo para a terceira opção do menu suspenso de navegação, tem-se as mesmas informações para as disciplinas de serviço do Departamento de Estatística, mostradas na Figura 21. Todas as opções de filtros e análises são análogos à pagina das disciplinas do bacharelado, com adição da opção de selecionar o curso dos estudantes.



Figura 21: Dashboard Streamlit - Disciplinas de serviço

Em vez de um botão para converter os dados visualizados em valores absolutos como no Shiny, isso agora é feito por meio de uma caixa de seleção, que ao ser clicada, exhibe os resultados expostos na Figura 22.



Figura 22: Dashboard Streamlit - Disciplinas de serviço em números absolutos

### 3.2.4 Disciplinas da Matemática

Assim como no *dashboard Shiny*, a quarta página traz informações das disciplinas do Departamento de Matemática que são obrigatórias para estudantes da Estatística. Os resultados estão exibidos na figura a seguir.

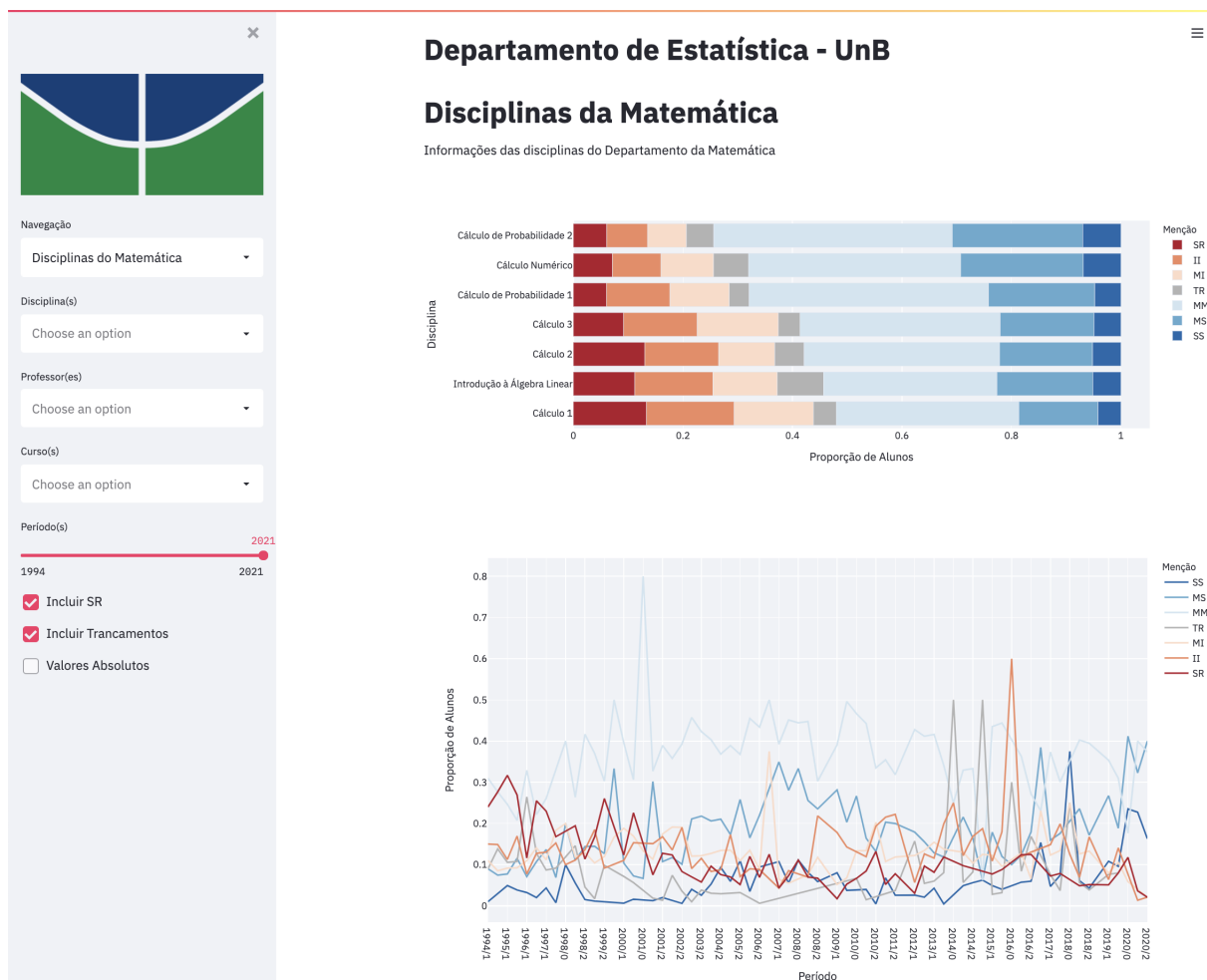


Figura 23: Dashboard Streamlit - Disciplinas de serviço em números absolutos

### 3.2.5 Professores

Por fim, a última página mostra informações dos professores do departamento, com o gráfico de barras empilhadas de menções obtidas por estudantes para cada professor, com nomes mais uma vez criptografados, e o gráfico de linhas do número médio de alunos por professor ao longo dos períodos letivos.

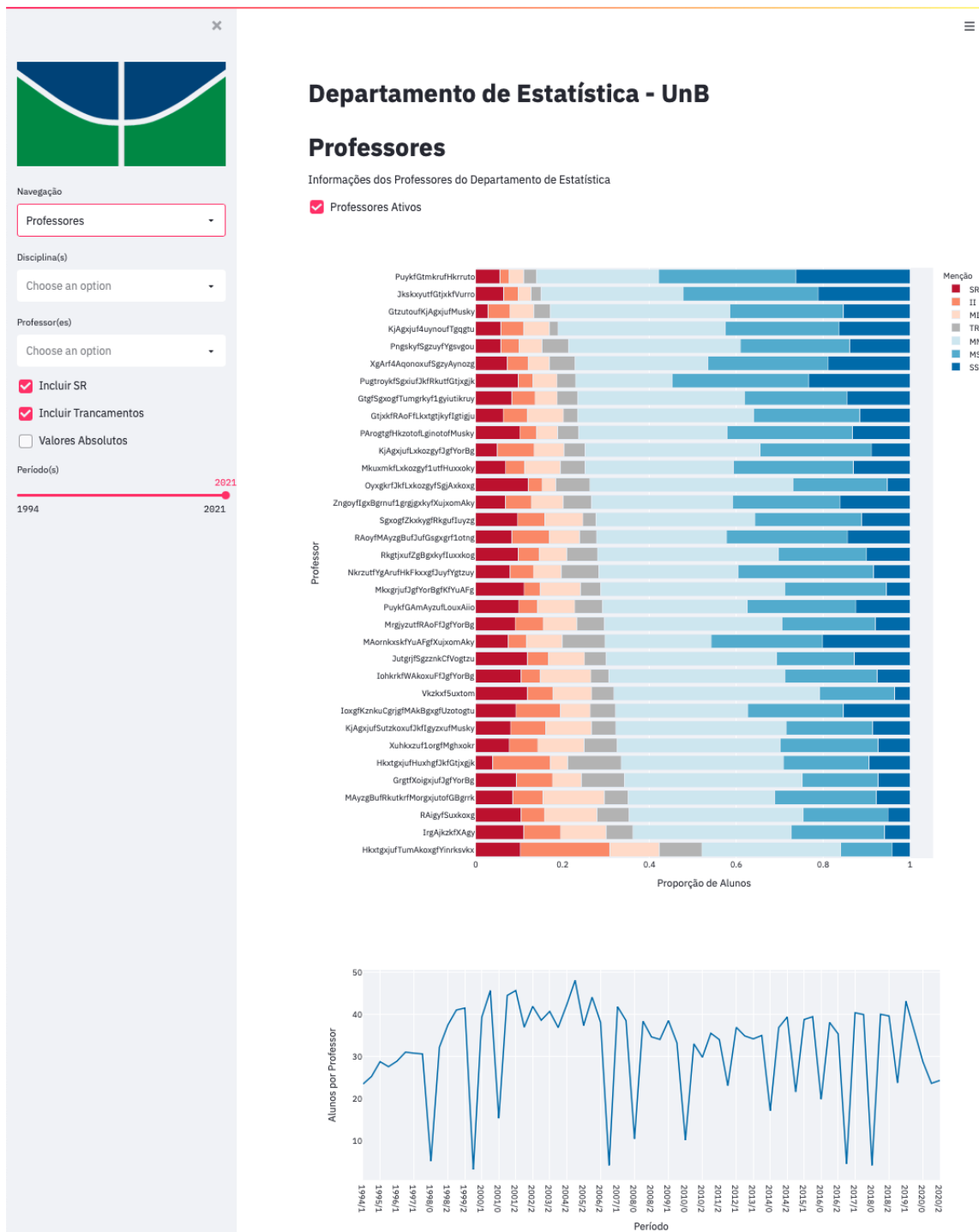


Figura 24: Dashboard Streamlit - Professores

### 3.3. COMPARAÇÃO SHINY VS. STREAMLIT

Os resultados obtidos por ambas as ferramentas foram satisfatórios, mas com diferenças notáveis. As principais distinções encontradas serão discutidas a seguir.

### 3.3.1 Dificuldade e Velocidade de Desenvolvimento

Tanto o **Shiny** quanto o **Streamlit** são *frameworks* que se notabilizaram pela simplicidade de desenvolvimento de *dashboards* para quem já tem experiência nas linguagens *R* e *Python*, respectivamente, sem necessidade de conhecimentos adicionais que são normalmente empregados em desenvolvimento web, como **CSS**, **HTML** e **Javascript**. No entanto, o desenvolvimento em **Streamlit** se mostrou mais fácil para alguém que sabe *Python* do que o desenvolvimento em **Shiny** para alguém que sabe *R*. Pois mesmo sendo bastante experiente em *R* e feito o curso de Estatística com bastante ênfase nessa linguagem, a lógica e a sintaxe para desenvolvimento de **Shiny** é bastante diferente e necessita muito estudo prévio, com uma curva de aprendizagem relativamente alta.

Enquanto no **Streamlit**, mesmo com pouca experiência em *Python*, como a lógica e a sintaxe dos códigos são muito simples e similares a *Python* no geral, não foi necessário muito estudo e a curva de aprendizagem foi menor. Portanto, considera-se que o desenvolvimento de *dashboard Streamlit* é mais fácil que o desenvolvimento de *dashboard Shiny* para funcionalidades básicas.

### 3.3.2 Customização

No quesito personalização, é indiscutível a superioridade do **Shiny**, o que já era esperado, dado o tempo muito maior de mercado e consequente maturidade quando comparado ao **Streamlit**. São inúmeras as bibliotecas desenvolvidas para **Shiny** que permitem, muito facilmente, mudar todo o tema, disposição e funcionalidades do aplicativo, de forma que atingir um visual final extremamente personalizado para as suas necessidades não demanda muito esforço ou conhecimentos adicionais.

O **Streamlit** apresenta opções básicas de customização no próprio aplicativo, ou seja, que o usuário pode alterar, como tema escuro ou claro, cores primárias, secundárias de fundo ou do texto. No entanto, no momento do desenvolvimento, não foi possível encontrar opções prontas de personalização e organização mais avançadas para dar ao aplicativo um visual mais profissional e singular. Isso prejudicou a inclusão de algumas funcionalidades existentes no *dashboards Shiny*, como painéis de gráficos com várias abas. Mas em virtude do tamanho da comunidade em torno de *Python* e a velocidade com que o desenvolvimento de código avança, é possível que não demore para que muitas dessas questões sejam resolvidas.

### 3.3.3 Experiência do Usuário

No que diz respeito a experiência do usuário do aplicativo, principalmente em termos de performance, ambas as ferramentas apresentaram bom desempenho, com os gráficos sendo gerados em poucos segundos na abertura inicial e após o uso de filtros ou na alteração de valores proporcionais para absolutos. O **Streamlit** leva uma pequena vantagem por ser um pouco mais rápido, mas isso pode ser explicado pela menor quantidade de funcionalidades e de customização que o *dashboard Shiny*.

No entanto, ressalta-se que muito esforço foi empregado na otimização do código do **Shiny**, com remoção consciente de todos os cálculos do *front-end* (interface de usuário), resultando em melhor performance do aplicativo. Já no **Streamlit**, por não possuir muita diferenciação entre *front-end* e *back-end*, não foi necessário fazer uma separação rigorosa dos momentos dos cálculos, e mesmo assim o aplicativo se mostrou muito rápido. Portanto, acredita-se que, no geral, o **Streamlit** apresenta uma performance ligeiramente melhor, resultando em uma experiência de usuário um pouco mais agradável.

### 3.3.4 Hospedagem

Tanto o **Shiny** quando o **Streamlit** apresentam soluções convenientes para hospedagem rápida do aplicativo, permitindo o compartilhamento fácil do *dashboard*. O primeiro é possível hospedar na plataforma *shinyapps.io*, que acomoda gratuitamente até 5 aplicativos, com limitação de 25 horas ativas de uso de cada *dashboard*. Já o **Streamlit** permite hospedagem online pelo *github*, com limite de três aplicativos por conta e 1 Gigabyte de memória para cada um.

Ambas as ferramentas também apresentam soluções empresariais pagas, para hospedagem de aplicativos mais profissionais e de uso mais intensivo, com várias opções de planos. Também é possível hospedar em servidores privados, mas nesse caso é necessário maior conhecimento em programação. Para essas soluções mais específicas, acredita-se que o **Streamlit** apresenta a vantagem de utilizar uma linguagem com mais suporte e melhor integração com outras tecnologias.

## 4 CONCLUSÃO

Neste trabalho, foram analisados dados da Universidade de Brasília referentes aos alunos do Departamento da Estatística, gerados pelo SIGRA e pelo SIGRA, dois sistemas de gestão de informações acadêmicas com características diferentes. Os dados tratam dos períodos de 1994 até o 1º semestre de 2021, totalizando quase 400 mil observações.

Grande valor foi obtido já na fase exploratória dos dados, em que surgiram vários desafios para a correção das informações e integração de bases de sistemas diferentes. O uso de dados da UnB desde 1994 implica em muitas alterações de padrões, de sistemas e de pessoas responsáveis pela manutenção dos dados. Isso resulta em uma base grande com problemas diversos de inconsistências e duplicações, sendo necessário um trabalho minucioso para uniformização e adequação dos dados para visualização.

Com esses desafios iniciais superados, o *dashboard Shiny* foi desenvolvido com relativa facilidade, em função da boa formação que o curso de estatística fornece na linguagem R e do início do desenvolvimento na disciplina *Tópicos em Estatística 1*. Já no caso do *dashboard Streamlit*, foi necessário estudo individual extenso da linguagem Python e da tecnologia Streamlit. Felizmente, ambas as ferramentas usufruem de bom suporte e de grandes comunidades.

Os resultados obtidos em ambos os *dashboards* foram satisfatórios, permitindo inúmeras análises rápidas e interativas. Várias descobertas relevantes para o departamento foram possíveis, como a identificação de disciplinas que trazem maior dificuldade para os estudantes, a constatação de que grande parte delas são dos estágios iniciais do curso ou do Departamento da Matemática, além da detecção de diferenças de desempenho de alunos entre professores de uma mesma disciplina. Também foi possível confirmar tendências ao longo do tempo, como a alteração de desempenho de alunos nos períodos de ensino remoto durante a pandemia. São informações como essas que embasam o departamento na tomada de decisões que visem a melhoria da qualidade e da experiência de ensino para todos os envolvidos, objetivo principal do trabalho.

Finalmente, ambas as ferramentas utilizadas cumprem bem seu propósito de permitir a criação rápida de *dashboards* informativos, ágeis e com boa estética. As principais diferenças encontradas estão na dificuldade do desenvolvimento e na personalização do aplicativo. Para pessoas que sabem o básico de R e Python, a curva de aprendizado de Shiny é maior, mas isso é atenuado pela formação em R oferecida pelo bacharelado em Estatística da UnB. Para quem deseja uma maior customização de maneira prática e sem envolver outros conhecimentos, Shiny se mostrou uma alternativa melhor. As duas tecnologias oferecem uma boa experiência para o usuário e diversas opções de hospedagem.



## REFERÊNCIAS

BELORKAR, Abha; GUNTUKU, S.C. **Interactive Data Visualization with Python**. 2nd ed. Birmingham: PACKT, 2020. *E-book*.

FAY, Colin; GIRARD, Cervan; GUYADER, Vincent; ROCHETTE, Sébastien. **Engineering Production-Grade Shiny Apps**. London: Chapman & Hall, 2021. *E-book*. Disponível em: <https://engineering-shiny.org/>. Acesso em: 06 mar. 2021.

GROLEMUND, Garret; WICKHAM, Hadley. **R for Data Science**. Sebastopol: O'REILLY, 2017. *E-book*. Disponível em: <https://r4ds.had.co.nz/>. Acesso em: 20 fev. 2021.

VANDERPLAS, Jake. **Python Data Science Handbook**. Sebastopol: O'REILLY, 2016. *E-book*. Disponível em: <https://jakevdp.github.io/PythonDataScienceHandbook/>. Acesso em: 06 mar. 2021.

WICKHAM, Hadley. **Mastering Shiny**. Sebastopol: O'REILLY, 2020. *E-book*. Disponível em: <https://mastering-shiny.org/index.html>. Acesso em: 03 mar. 2021.