

# Título

## 1) Coleta de dados

A Prova Brasil é uma avaliação diagnóstica realizada pelo Instituto Nacional de Estudos e Pesquisas Educacionais Anísio Teixeira (Inep/MEC) com o intuito de avaliar os alunos de ensino básico do quinto ano e do nono ano do Ensino Fundamental. Assim, tem o objetivo de aferir a qualidade do ensino oferecido pelo sistema educacional brasileiro a partir de testes padronizados e questionários socioeconômicos.

A prova é realizada em escolas públicas e particulares. Pode-se declarar que a população alvo são os próprios alunos do quinto e nono anos. Entretanto, não temos o mesmo grupo como população acessível, visto que apenas escolas com mais de 20 alunos matriculados nesses anos são avaliadas, perdendo assim uma parte da população alvo.

As principais características levantadas incluem nome, idade, nota das provas (Português e Matemática), escolaridade do pai e da mãe, sexo, entre outras variáveis. Esses dados são coletados através da prova e de questionários feitos pelos alunos.

## 2) Variáveis

Segue uma tabela com todas as variáveis presentes nas amostras recebidas pelos alunos de matrícula 232014825, 232014905 e 232014880.

## 3) Variáveis categóricas

As variáveis escolhidas para análise foram UF, sexo, cor, área e trabalho. Todas são variáveis qualitativas nominais.

### UF

A variável UF se refere à unidade federativa da localização da escola. Na amostra, existem estudantes de escolas de todas as 27 UFs do país.

Variável	Tipo de variável	Escala	Descrição da variável
Ano	Quantitativa discreta	Nominal	Ano de realização da SAEB
Região	Qualitativa nominal	Nominal	Região onde é localizada a escola
Unidade Federativa	Qualitativa nominal	Nominal	UF da localização da escola
Município	Qualitativa nominal	Nominal	Município onde é localizada a escola
Área	Qualitativa nominal	Nominal	Área de localização do estudante
Dependência administrativa	Qualitativa nominal	Nominal	Âmbito de subordinação da escola
Localização	Qualitativa nominal	Nominal	Localização da escola
Nota em LP	Quantitativa contínua	Intervalar	Nota na prova de língua portuguesa
Nota em MAT	Quantitativa contínua	Intervalar	Nota na prova de matemática
Sexo	Qualitativa nominal	Nominal	Sexo do estudante
Cor	Qualitativa nominal	Nominal	Raça/etnia do estudante
Mês de Nasc.	Qualitativa nominal	Nominal	Mês de nascimento
Ano de Nasc.	Quantitativa discreta	Nominal	Ano de nascimento
Computador	Qualitativa ordinal	Razão	Se o aluno tem computador em casa
Esc. Mãe	Qualitativa ordinal	Ordinal	Escolaridade da mãe
Esc. Pai	Qualitativa ordinal	Ordinal	Escolaridade do pai
Tempo de tela	Qualitativa ordinal	Ordinal	Quanto tempo o estudante gasta em telas eletrônicas
Afazer dom.	Qualitativa ordinal	Ordinal	Tempo gasto fazendo afazeres domésticos
Trabalho	Qualitativa nominal	Nominal	Se o aluno trabalha ou não
Perspectivas	Qualitativa nominal	Nominal	Perspectivas do futuro

**Tabela 1:** Explicação das Variáveis

A partir das informações mostradas pelo mapa e pela tabela, observa-se que as unidades federativas com maior número de estudantes que realizaram a prova são São Paulo, Minas Gerais e Bahia, concentrando 35,53% dos estudantes. Já os estados com menos estudantes foram Roraima e Amapá, com 3 e 4 estudantes, respectivamente.

## Cor

A variável cor representa a etnia de cada estudante. É qualitativa nominal e está dividida em seis categorias: branca, parda, preta, indígena, amarela e não declarada. Segue uma tabela e um gráfico sua distribuição:

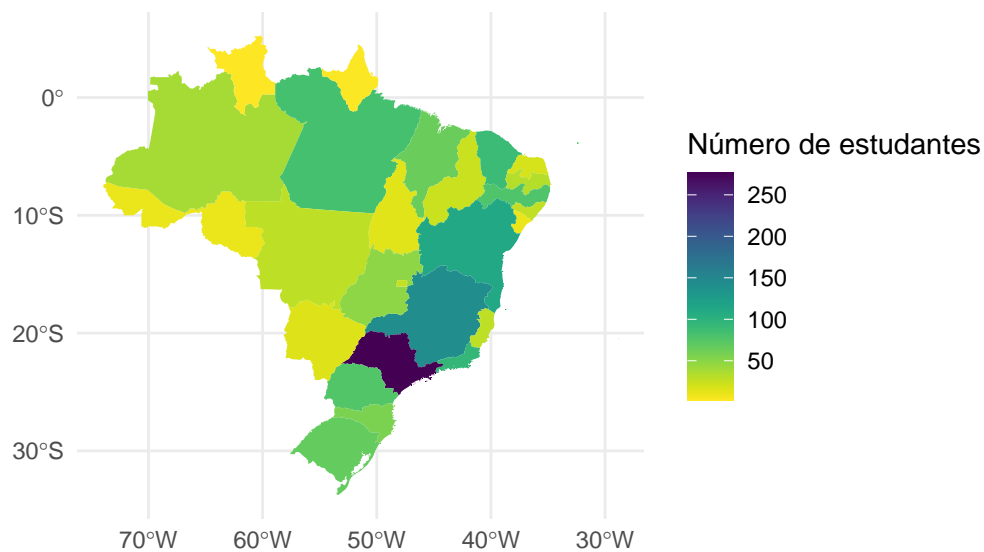
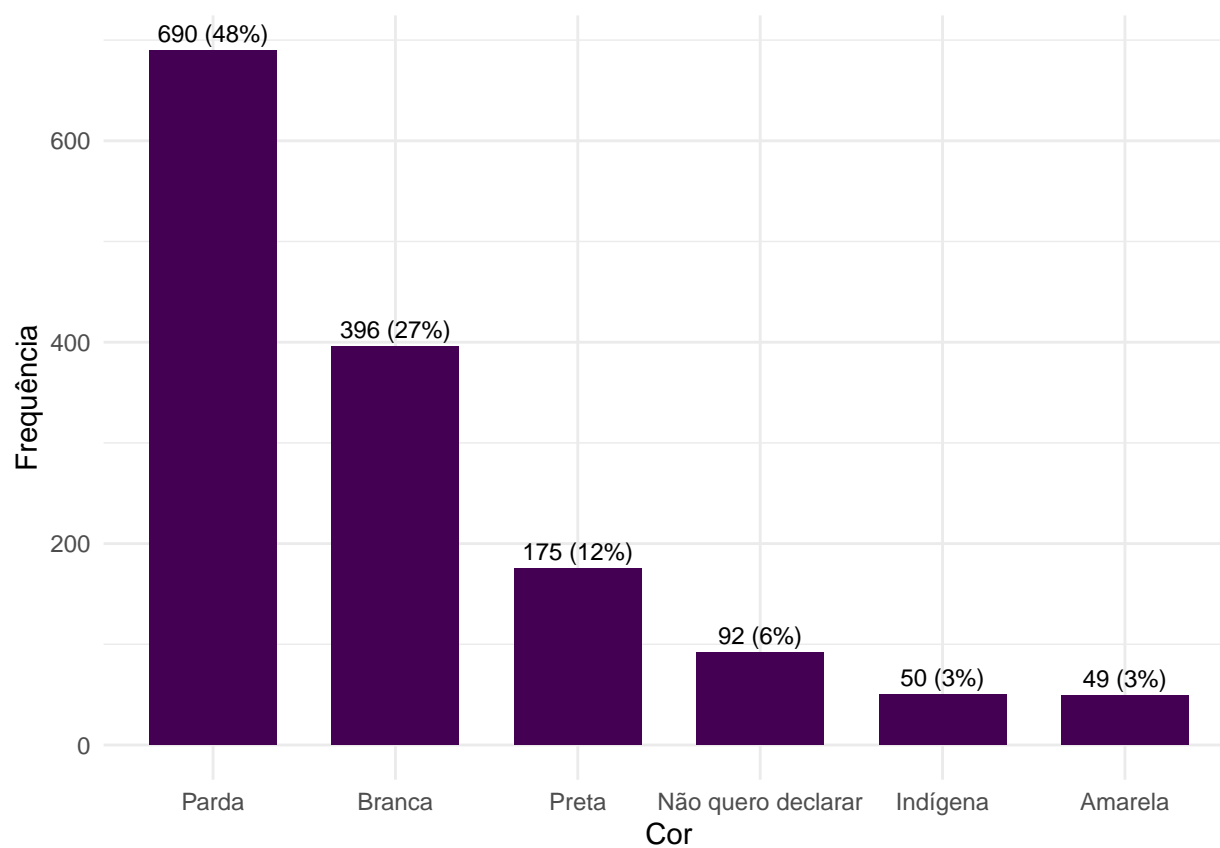


Gráfico 1: Mapa



Com base nos dados apresentados, é possível notar que a maioria dos estudantes são pardos, representando 48% do total de estudantes. As etnias com menor número de estudantes foram indígenas e amarelos, representando, juntas, apenas 6% dos estudantes.

Unidade Federativa	Frequência absoluta	Frequência relativa
São Paulo	277	0.185
Minas Gerais	143	0.095
Bahia	113	0.075
Rio De Janeiro	95	0.063
Ceará	90	0.060
Pará	84	0.056
Pernambuco	78	0.052
Paraná	77	0.051
Rio Grande Do Sul	68	0.045
Maranhão	65	0.043
Santa Catarina	57	0.038
Goiás	49	0.033
Amazônas	39	0.026
Paraíba	33	0.022
Espírito Santo	30	0.020
Mato Grosso	30	0.020
Alagoas	28	0.019
Distrito Federal	27	0.018
Piauí	25	0.017
Rio Grande Do Norte	20	0.013
Mato Grosso Do Sul	17	0.011
Tocantins	15	0.010
Acre	11	0.007
Rondônia	11	0.007
Sergipe	11	0.007
Amapá	4	0.003
Roraima	3	0.002

**Tabela 2:** Quantidade de alunos por unidade federativa

## Área

A variável área representa a região de localização da escola do estudante, na capital ou no interior. Segue uma tabela e um gráfico com sua distribuição:

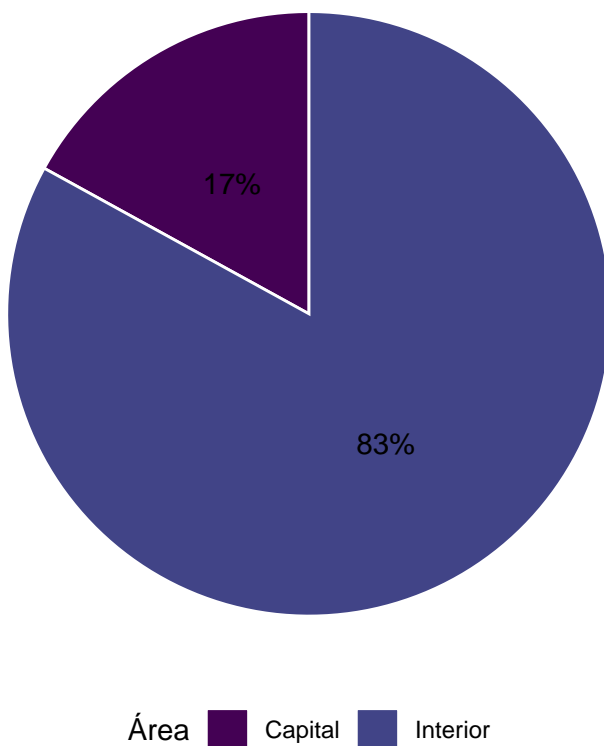
Cor	Frequência absoluta	Frequência relativa
Parda	690	48
Branca	396	27
Preta	175	12
Não quero declarar	92	6
Indígena	50	3
Amarela	49	3

**Tabela 3:** Quantidade de alunos pela cor

Área	Frequência absoluta	Frequência relativa
Capital	255	0.17
Interior	1245	0.83

**Tabela 4:** Quantidade de alunos por área

### Porcentagem de estudantes pela área



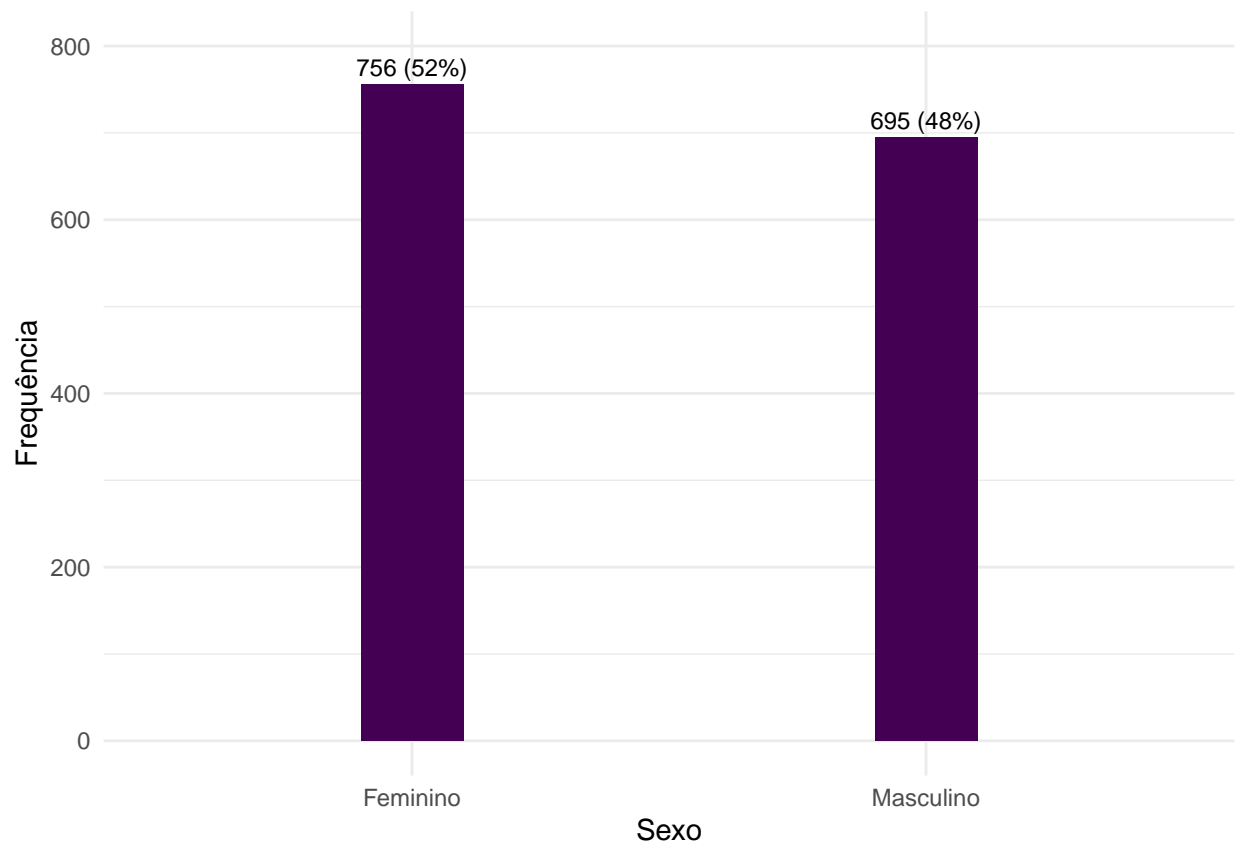
A partir do gráfico, fica claro que há uma proporção muito maior de estudantes de escolas localizadas em cidades de interior do que em cidades de capital.

### Sexo

A variável sexo indica o gênero dos estudantes que realizaram a prova. A seguir, é apresentada sua distribuição:

Sexo	Frequência absoluta	Frequência relativa
Feminino	756	0.52
Masculino	695	0.48

**Tabela 5:** Quantidade de alunos pelo sexo



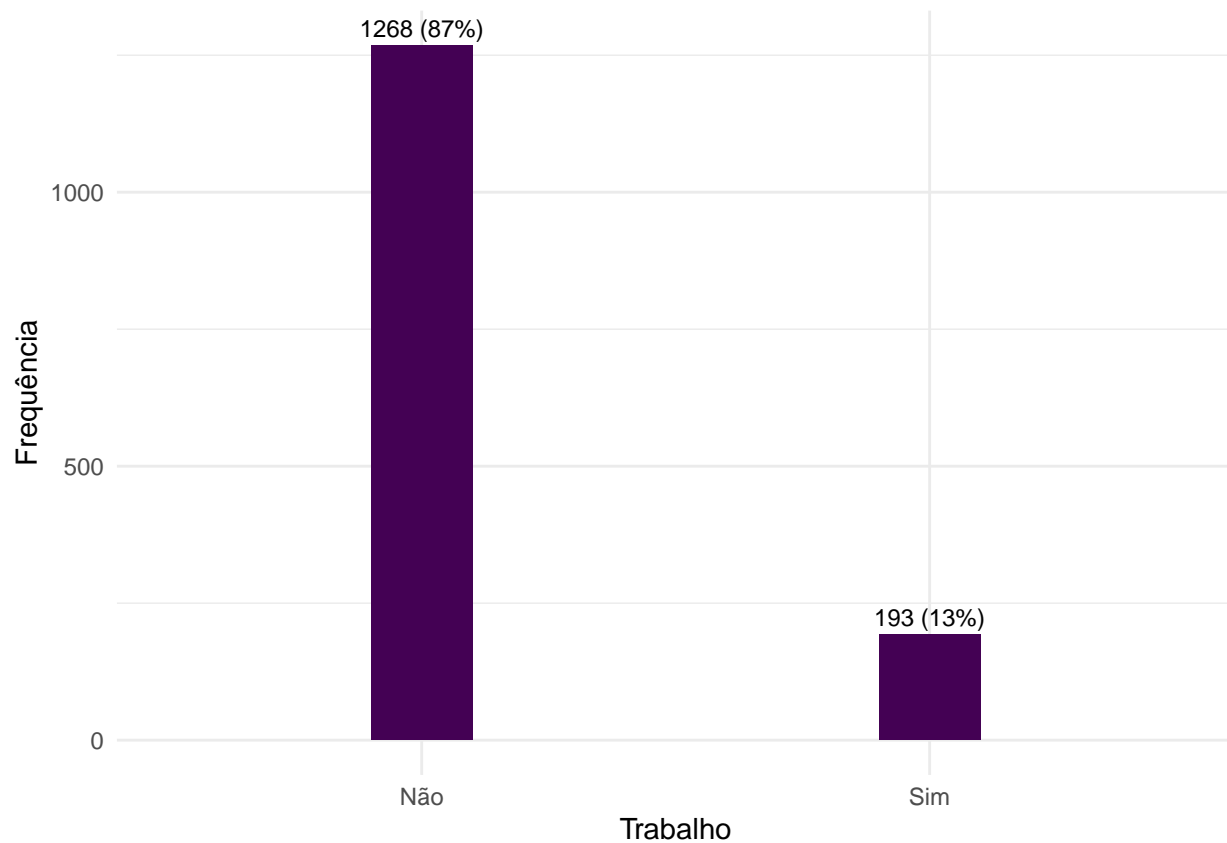
Com base nas informações apresentadas, nota-se que o número de meninas e de meninos que realizaram a prova é semelhante, embora a proporção de meninas seja maior, representando 52% do total de estudantes.

## Trabalho

A variável trabalho indica se o estudante trabalha ou não. A seguir, observa-se sua distribuição:

Trabalho	Frequência absoluta	Frequência relativa
Não	1268	0.87
Sim	193	0.13

**Tabela 6:** Quantidade de alunos pelo trabalho



Analisando os dados apresentados, constata-se que 193 estudantes realizam algum tipo de trabalho, o que representa 13% do total.

#### 4) Variáveis quantitativas

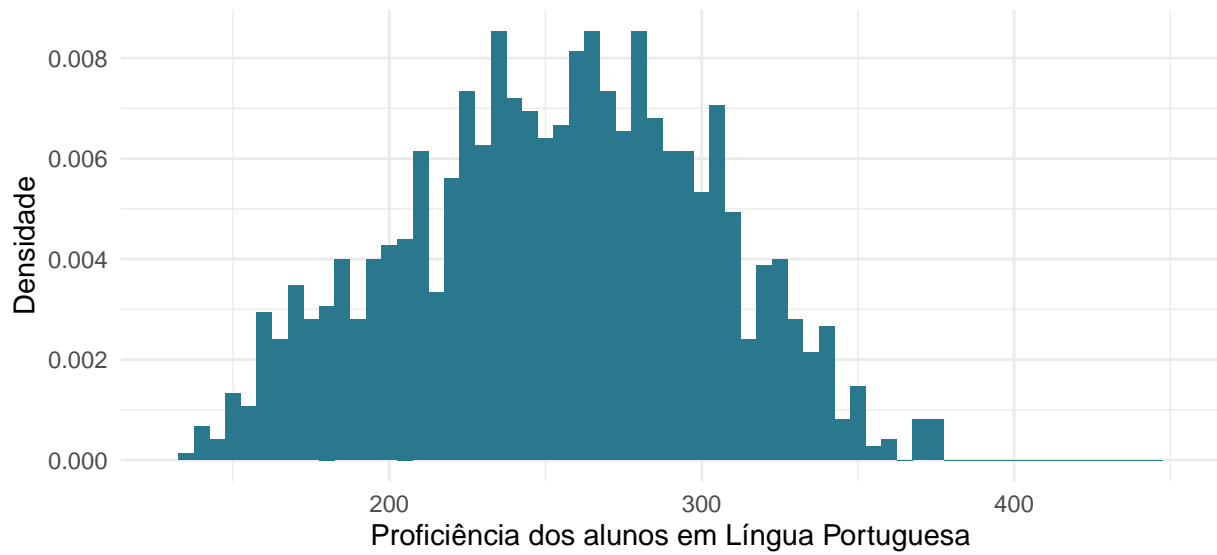
##### Proeficiência do aluno em Língua Portuguesa

**Tabela 7:** Distribuição da nota dos alunos

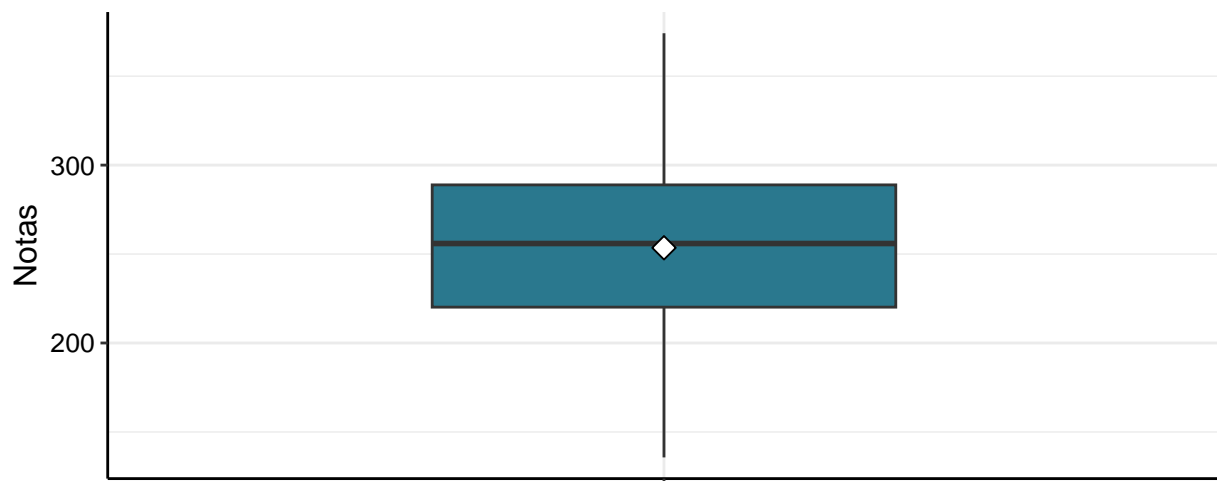
Intervalo de Classe	Frequência
135 [—) 157	27
157 [—) 179	96
179 [—) 201	111
201 [—) 222	162
222 [—) 244	238
244 [—) 266	238
266 [—) 287	238
287 [—) 309	200
309 [—) 331	114
331 [—) 352	59
352 [—) 374	17
<b>Total</b>	<b>1500</b>

**Tabela 8:** Medidas resumo

Média	Mediana	Erro padrão	Coef. de Variação	Coef. de Assimetria de Pearson	Curtose
253.529	255.934	48.805	0.193	-2.817	-0.571



**Gráfico 2:** Histograma das notas de Língua Portuguesa



**Gráfico 3:** Box-plot das notas de LP



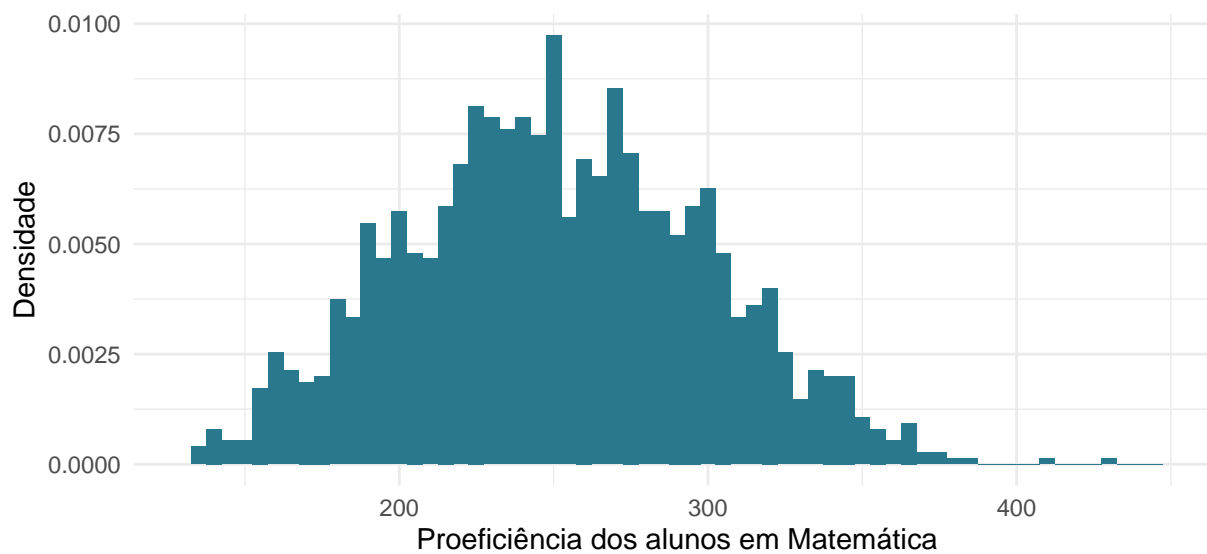
## Proeficiência do aluno em Matemática

**Tabela 9:** Distribuição das notas dos alunos

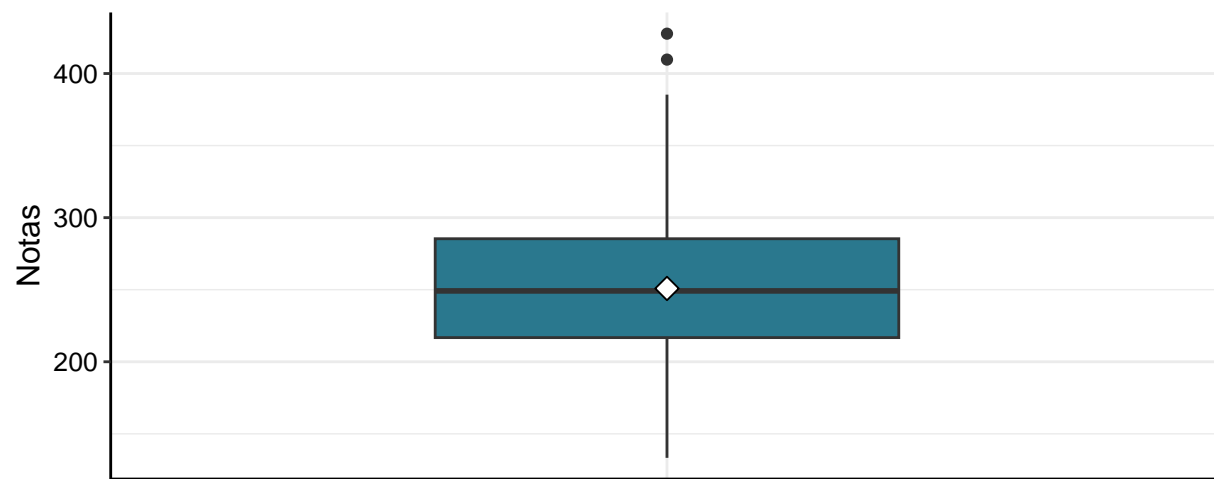
Intervalo de Classe	Frequência
133 [—) 160	37
160 [—) 187	108
187 [—) 214	201
214 [—) 240	299
240 [—) 267	290
267 [—) 294	255
294 [—) 321	193
321 [—) 347	83
347 [—) 374	28
374 [—) 401	4
401 [—) 428	2
<b>Total</b>	<b>1500</b>

**Tabela 10:** Medidas resumo

Média	Mediana	Erro padrão	Coef. de Variação	Coef. de Assimetria de Pearson	Curtose
250.859	249.162	48.681	0.194	3.786	-0.374



**Gráfico 4:** Histograma das notas de matemática



**Gráfico 5:** Box-plot das notas de Matemática