

# Análise de Dados do Arquivo hsb2f.csv

Usando o Perplexity IA

João Pedro Albino

06/11/2024

## Questão inicial

**Escrever um programa em linguagem R que leia o arquivo hsb2f.csv, realize uma Análise Exploratória de Dados e mostre o código.**

**Resposta:** Para realizar uma Análise Exploratória de Dados (AED) no arquivo hsb2f.csv, você pode utilizar a linguagem R.

Abaixo está um exemplo de código que lê o arquivo, realiza algumas análises descritivas e visualizações básicas.

## Introdução

Neste documento, realizaremos uma análise exploratória dos dados contidos no arquivo **hsb2f.csv**. O objetivo é investigar as notas de matemática dos estudantes em diferentes grupos demográficos.

## Carregando Pacotes Necessários

```
# Carregar pacotes necessários
library(tidyverse) # Para manipulação de dados e visualização
library(ggplot2)   # Para gráficos
```

## Carregando os Dados

```
# Carregar os dados do arquivo csv
file_path <- "../dados/hsb2f.csv" # Caminho correto
data <- read.csv(file_path, sep = ";")
```

## Estrutura dos Dados

Verificar as primeiras linhas do dataset

```
# Verificar as primeiras linhas do dataset
head(data)
```

```
##      id      genero      raca clasocial tiposcola programa ler escrever matematica
## 1  70 masculino branca      baixa pública básico 57      52      41
## 2 121 feminino branca      média pública técnico 68      59      53
## 3  86 masculino branca      alta pública básico 44      33      54
## 4 141 masculino branca      alta pública técnico 63      44      47
## 5 172 masculino branca      média pública acadêmico 47      52      57
## 6 113 masculino branca      média pública acadêmico 44      52      51
##      ciencias estsociais
## 1      47      57
## 2      63      61
## 3      58      31
## 4      53      56
## 5      53      61
## 6      63      61
```

## Resumo estatístico dos dados

```
summary(data)
```

```
##      id      genero      raca      clasocial
## Min.   : 1.00   Length:200   Length:200   Length:200
## 1st Qu.:50.75   Class :character Class :character Class :character
## Median :100.50   Mode  :character Mode  :character Mode  :character
## Mean    :100.50
## 3rd Qu.:150.25
## Max.    :200.00
##      tiposcola      programa      ler      escrever
## Length:200      Length:200      Min.   :28.00   Min.   :31.00
## Class :character Class :character 1st Qu.:44.00   1st Qu.:45.75
## Mode  :character Mode  :character Median :50.00   Median :54.00
##                                     Mean    :52.23   Mean    :52.77
##                                     3rd Qu.:60.00   3rd Qu.:60.00
##                                     Max.    :76.00   Max.    :67.00
##      matematica      ciencias      estsociais
## Min.   :33.00   Min.   :26.00   Min.   :26.00
## 1st Qu.:45.00   1st Qu.:44.00   1st Qu.:46.00
## Median :52.00   Median :53.00   Median :52.00
## Mean    :52.65   Mean    :51.85   Mean    :52.41
## 3rd Qu.:59.00   3rd Qu.:58.00   3rd Qu.:61.00
## Max.    :75.00   Max.    :74.00   Max.    :71.00
```

## Estrutura dos dados

```
str(data)
```

```
## 'data.frame': 200 obs. of 11 variables:
## $ id      : int 70 121 86 141 172 113 50 11 84 48 ...
## $ genero   : chr "masculino" "feminino" "masculino" "masculino" ...
## $ raca     : chr "branca" "branca" "branca" "branca" ...
## $ clasocial : chr "baixa" "média" "alta" "alta" ...
```

```
## $ tiposcola : chr "pública" "pública" "pública" "pública" ...
## $ programa  : chr "básico" "técnico" "básico" "técnico" ...
## $ ler       : int 57 68 44 63 47 44 50 34 63 57 ...
## $ escrever  : int 52 59 33 44 52 52 59 46 57 55 ...
## $ matematica: int 41 53 54 47 57 51 42 45 54 52 ...
## $ ciencias  : int 47 63 58 53 53 63 53 39 58 50 ...
## $ estsociais: int 57 61 31 56 61 61 61 36 51 51 ...
```

## Análise das Variáveis Categóricas

### 1. Gênero

```
ggplot(data, aes(x = genero)) +
  geom_bar(fill = "steelblue") +
  labs(title = "Distribuição de Gênero dos Estudantes",
       x = "Gênero",
       y = "Frequência") +
  theme_minimal()
```

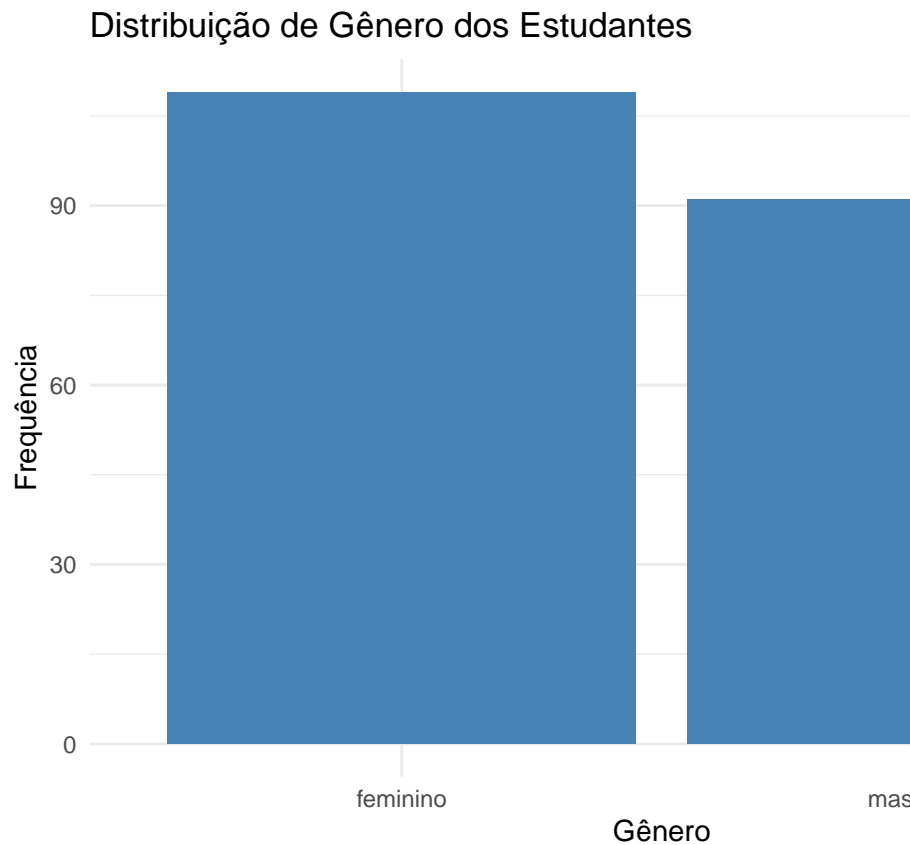


Gráfico de barras para a variável gênero

### 2. Raça

```
ggplot(data, aes(x = raca)) +
  geom_bar(fill = "lightgreen") +
  labs(title = "Distribuição da Raça dos Estudantes",
        x = "Raça",
        y = "Frequência") +
  theme_minimal()
```

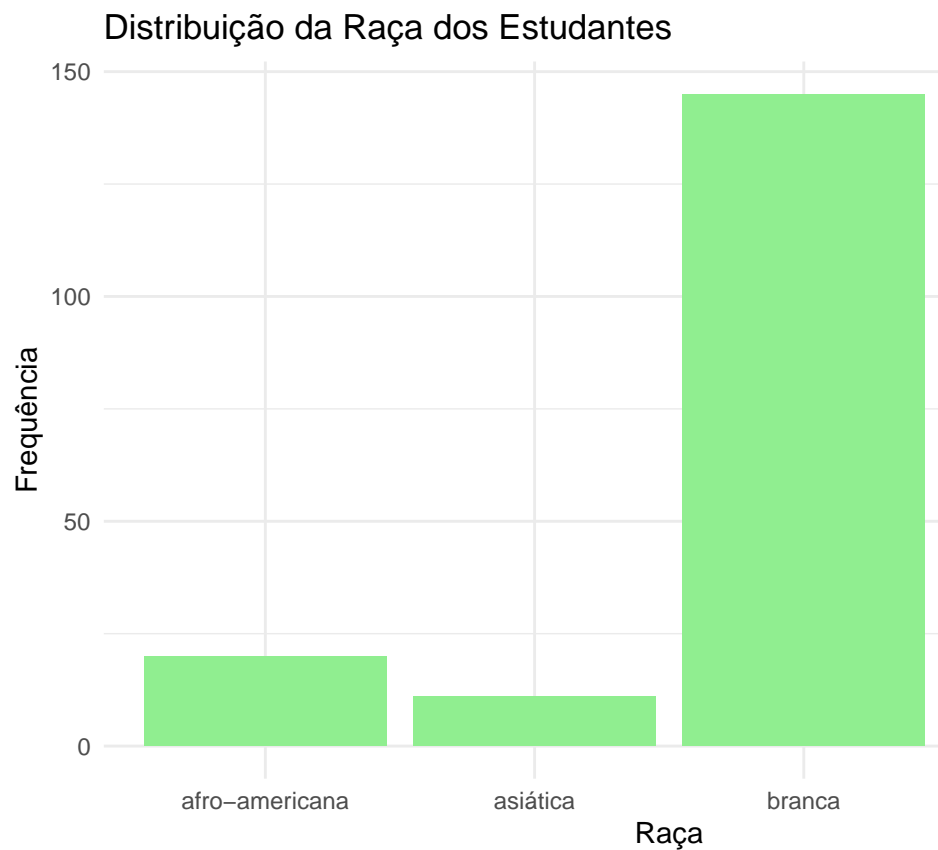


Gráfico de barras para a variável raça

### 3. Classe Social

```
ggplot(data, aes(x = clasocial)) +
  geom_bar(fill = "salmon") +
  labs(title = "Distribuição da Classe Social dos Estudantes",
        x = "Classe Social",
        y = "Frequência") +
  theme_minimal()
```

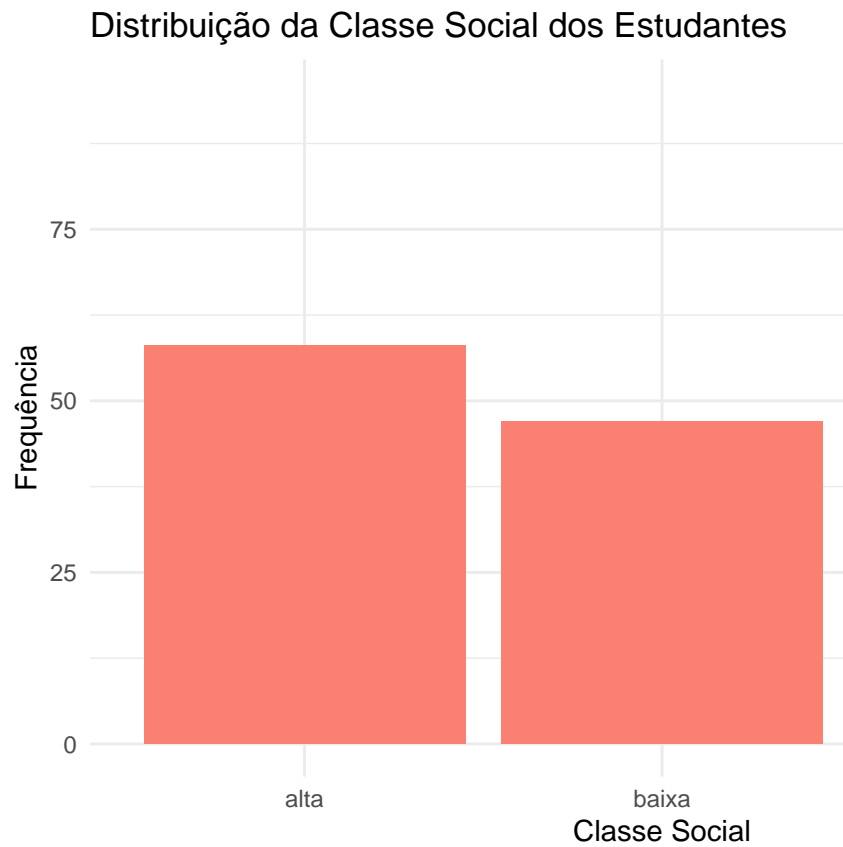


Gráfico de barras para a variável classe social

#### 4. Tipo de Escola

```
ggplot(data, aes(x = tiposcola)) +  
  geom_bar(fill = "orange") +  
  labs(title = "Distribuição do Tipo de Escola dos Estudantes",  
        x = "Tipo de Escola",  
        y = "Frequência") +  
  theme_minimal()
```

Distribuição do Tipo de Escola dos Estudantes

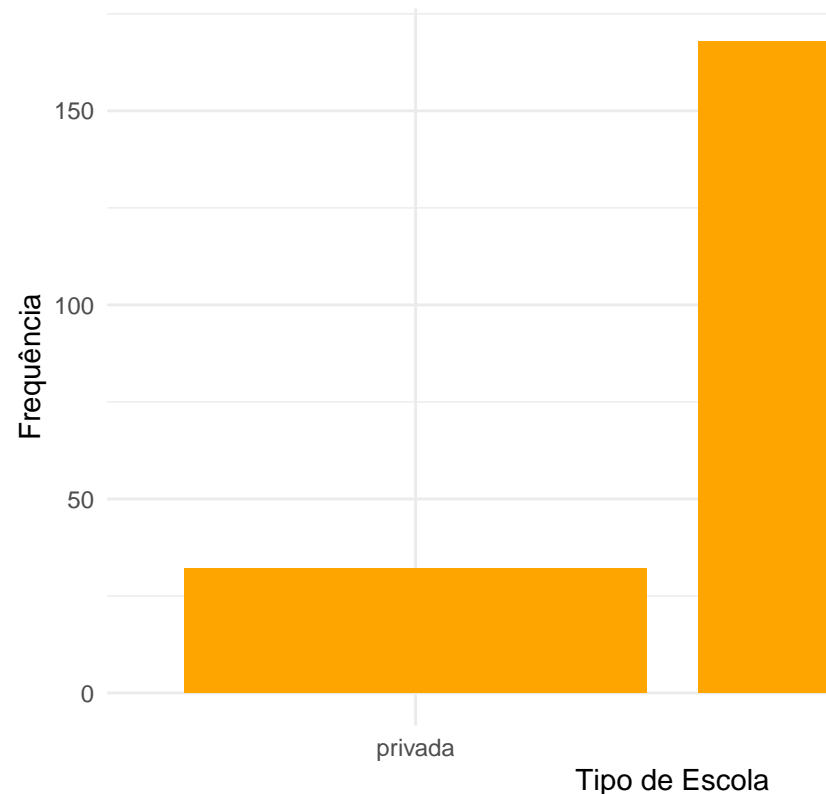


Gráfico de barras para a variável tipo de escola

### Conclusão Preliminar da Análise Descritiva

Os gráficos anteriores mostram a distribuição das variáveis categóricas no dataset `hsb2f.csv`. Essas visualizações ajudam a entender melhor o perfil dos estudantes em relação ao gênero, raça, classe social e tipo de escola.

**Prompt:** Realizar uma análise mais profunda dessas variáveis categóricas. Quais **insights** poderiam ser observados?

*Resposta:*

Para realizar uma análise mais aprofundada das variáveis categóricas do dataset `hsb2f.csv`, podemos explorar como essas variáveis se relacionam com o desempenho acadêmico dos estudantes, especialmente nas notas de matemática.

A seguir, apresento alguns insights que podem ser observados a partir da análise das variáveis categóricas, como gênero, raça, classe social e tipo de escola.

#### 1. Gênero

- **Análise:**

- Comparar as médias das notas de matemática entre meninos e meninas pode revelar se existe uma diferença significativa no desempenho acadêmico entre os gêneros.
- Gráficos de barras ou boxplots podem ser utilizados para visualizar essas diferenças.

- **Insight:**

- Se as meninas apresentarem notas consistentemente mais altas do que os meninos, isso pode indicar um desempenho acadêmico superior em determinadas disciplinas, o que pode ser explorado para entender fatores como motivação e apoio educacional.

## 2. Raça

- **Análise:**
  - Analisar as médias das notas de matemática por raça (branca, afro-americana, hispânica, asiática) pode evidenciar disparidades raciais no desempenho acadêmico.
  - Utilizar gráficos de barras ou gráficos de violino para visualizar a distribuição das notas por raça.
- **Insight:**
  - Se os estudantes brancos tiverem notas significativamente mais altas do que os estudantes afro-americanos e hispânicos, isso pode indicar desigualdades no acesso a recursos educacionais e suporte acadêmico. Essa informação é crucial para desenvolver políticas educacionais que visem reduzir essas disparidades.

## 3. Classe Social

- **Análise:**
  - Comparar as notas de matemática entre diferentes classes sociais (baixa, média, alta) pode mostrar como a classe social impacta o desempenho acadêmico.
  - Gráficos de barras ou boxplots podem ser usados para visualizar as diferenças nas médias.
- **Insight:**
  - Se os alunos de classe alta apresentarem notas muito superiores aos de classe baixa, isso pode sugerir que fatores socioeconômicos influenciam diretamente o desempenho escolar. Intervenções direcionadas podem ser necessárias para apoiar estudantes de classes sociais mais baixas.

## 4. Tipo de Escola

- **Análise:**
  - Comparar as médias das notas de matemática entre alunos de escolas públicas e privadas pode ajudar a entender o impacto do tipo de escola no desempenho acadêmico.
  - Gráficos comparativos podem ser utilizados para visualizar essas diferenças.
- **Insight:**
  - Se os alunos em escolas privadas apresentarem notas significativamente mais altas do que os alunos em escolas públicas, isso pode indicar que escolas privadas oferecem melhores recursos educacionais ou ambientes de aprendizagem. Essa informação pode ser utilizada para justificar investimentos em infraestrutura e recursos nas escolas públicas.

## 5. Interações entre Variáveis

- **Análise Avançada:**
  - Explorar interações entre gênero, raça e classe social em relação às notas de matemática pode revelar padrões mais complexos.
  - Usar modelos estatísticos ou gráficos multifacetados (como facetas) para investigar como essas variáveis interagem entre si.
- **Insight:**
  - Por exemplo, se meninas brancas de classe alta têm notas significativamente melhores do que meninos afro-americanos de classe baixa, isso destaca a necessidade de abordagens diferenciadas para atender às necessidades específicas dos diferentes grupos demográficos.

## Conclusão Preliminar

A análise das variáveis categóricas no dataset `hsb2f.csv` pode fornecer insights valiosos sobre o desempenho acadêmico dos estudantes.

Identificar disparidades e padrões nas notas de matemática em relação ao gênero, raça, classe social e tipo de escola é fundamental para desenvolver estratégias educacionais eficazes e promover a equidade no acesso à educação.

Essas análises devem ser acompanhadas por intervenções práticas que visem melhorar o suporte a grupos sub-representados ou em desvantagem.

**Prompt:** Realize essas análises utilizando a linguagem R.

**Resposta:** Para realizar uma análise mais profunda das variáveis categóricas do dataset `hsb2f.csv` utilizando a linguagem R, você pode seguir o passo a passo especificado.

Esta análise incluirá a leitura dos dados, a exploração das variáveis categóricas e a visualização dos resultados.

## Passo a Passo em R

- Carregar os Pacotes Necessários: Instale e carregue os pacotes `dplyr` e `ggplot2` para manipulação de dados e visualização.
- Ler o Dataset: Carregue o arquivo CSV.
- Explorar as Variáveis Categóricas: Crie gráficos para visualizar a distribuição das variáveis categóricas, como gênero, raça, classe social e tipo de escola.
- Analisar as Médias: Calcule as médias das notas de matemática para cada grupo categórico.
- Visualizar as Médias: Use gráficos para mostrar as médias das notas de matemática em relação às variáveis categóricas.

## Código em R

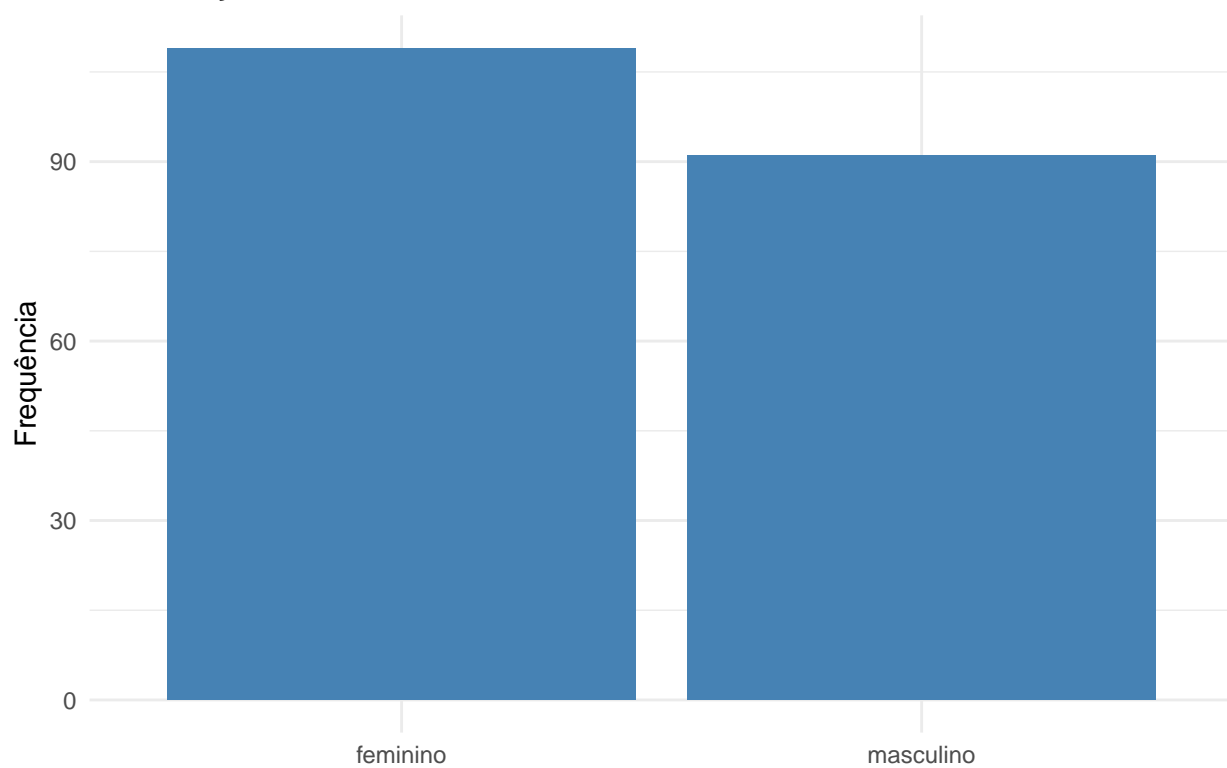
- Carregar os Pacotes Necessários, Ler o Dataset, Verificar as primeiras linhas do dataset (vide acima);

## Análise das variáveis categóricas

```
ggplot(data, aes(x = genero)) +  
  geom_bar(fill = "steelblue") +  
  labs(title = "Distribuição de Gênero dos Estudantes",  
        x = "Gênero",  
        y = "Frequência") +  
  theme_minimal()
```



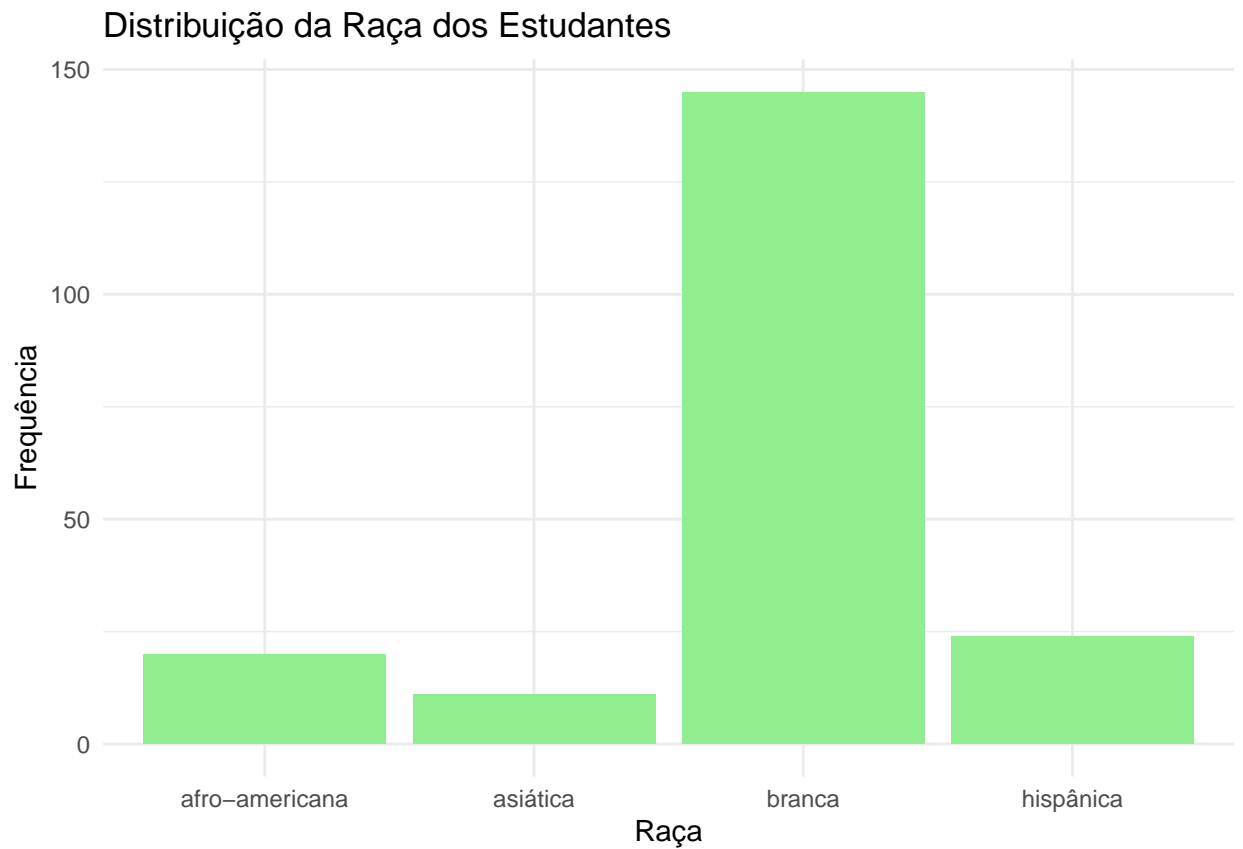
Distribuição de Gênero dos Estudantes



1. Gênero

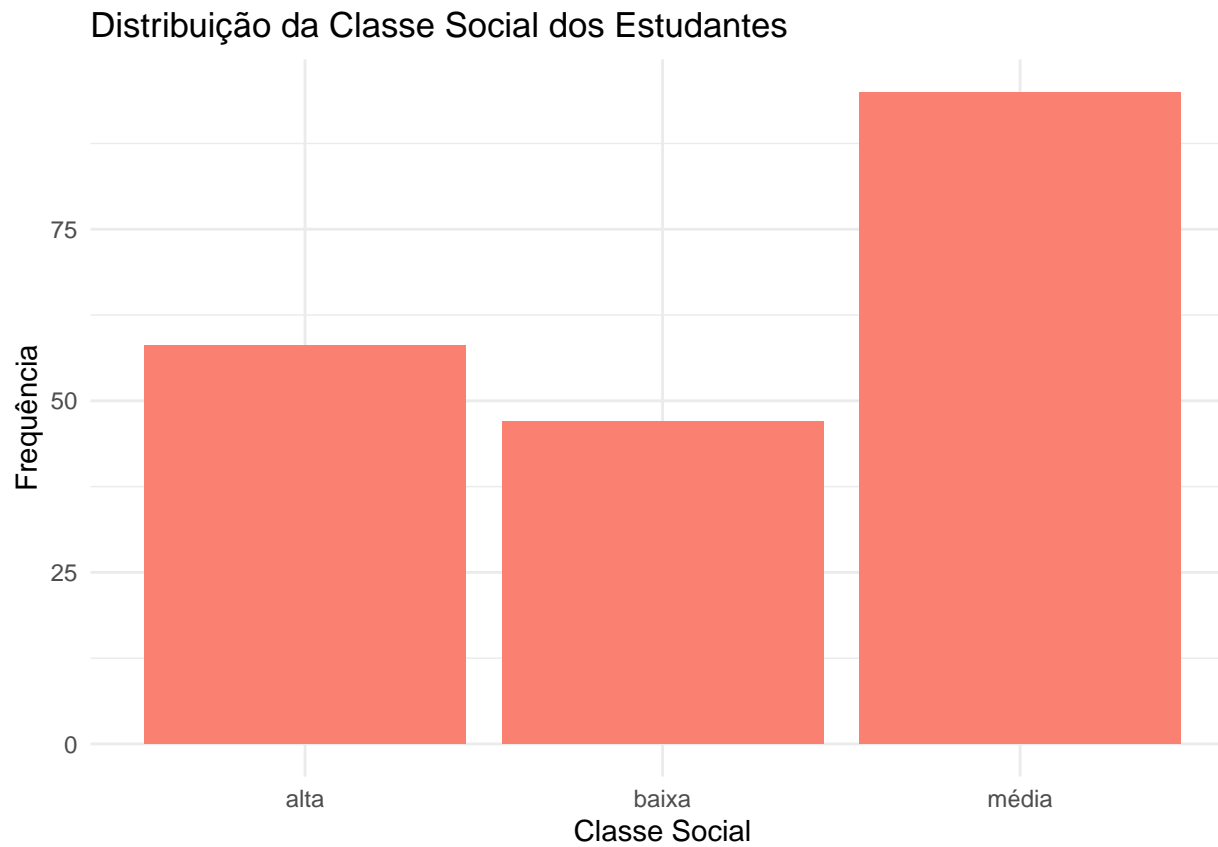
Gênero

```
ggplot(data, aes(x = raca)) +  
  geom_bar(fill = "lightgreen") +  
  labs(title = "Distribuição da Raça dos Estudantes",  
        x = "Raça",  
        y = "Frequência") +  
  theme_minimal()
```



## 2. Raça

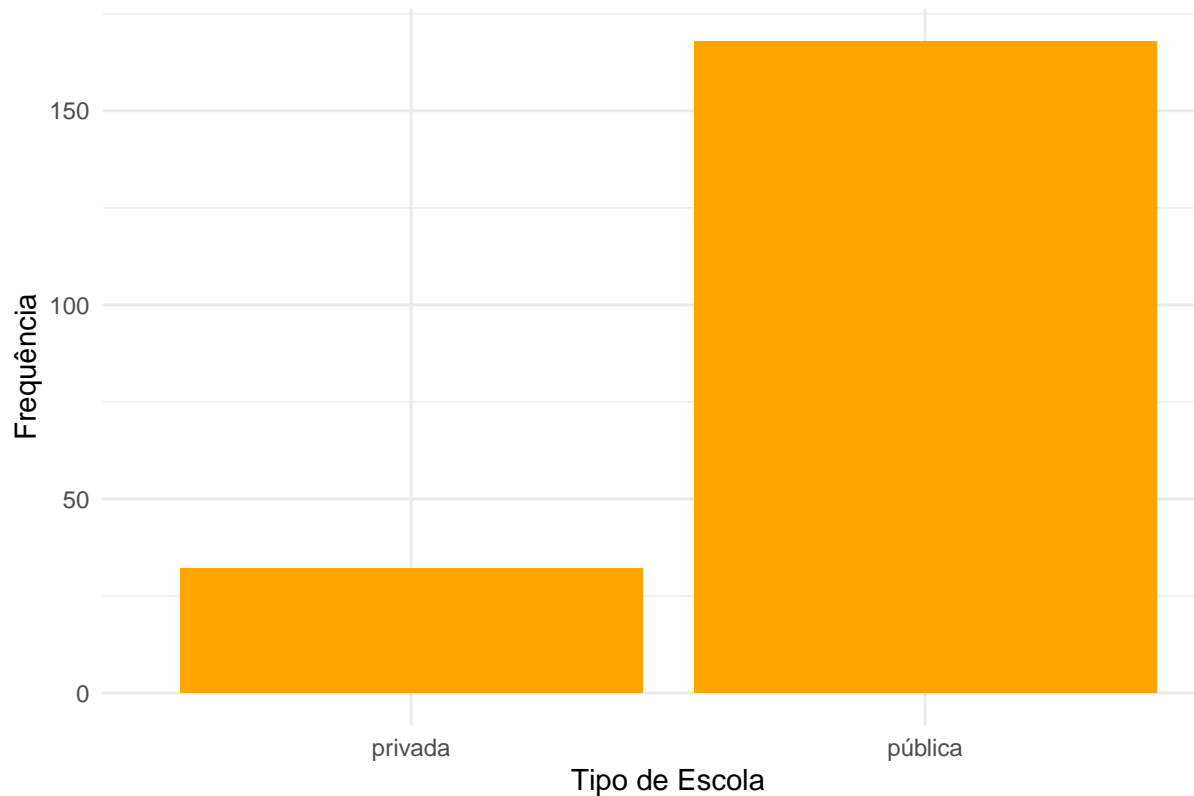
```
ggplot(data, aes(x = clasocial)) +  
  geom_bar(fill = "salmon") +  
  labs(title = "Distribuição da Classe Social dos Estudantes",  
        x = "Classe Social",  
        y = "Frequência") +  
  theme_minimal()
```



### 3. Classe Social

```
ggplot(data, aes(x = tiposcola)) +  
  geom_bar(fill = "orange") +  
  labs(title = "Distribuição do Tipo de Escola dos Estudantes",  
        x = "Tipo de Escola",  
        y = "Frequência") +  
  theme_minimal()
```

Distribuição do Tipo de Escola dos Estudantes



#### 4. Tipo de Escola

Análise das médias das notas de matemática

```
media_genero <- data %>%  
  group_by(genero) %>%  
  summarise(media_matematica = mean(matematica, na.rm = TRUE))  
print(media_genero)
```

#### Médias por Gênero

```
## # A tibble: 2 x 2  
##   genero      media_matematica  
##   <chr>          <dbl>  
## 1 feminino      52.4  
## 2 masculino     52.9
```

```
media_raca <- data %>%  
  group_by(raca) %>%  
  summarise(media_matematica = mean(matematica, na.rm = TRUE))  
print(media_raca)
```

## Médias por Raça

```
## # A tibble: 4 x 2
##   raca          media_matematica
##   <chr>          <dbl>
## 1 afro-americana 46.8
## 2 asiática      57.3
## 3 branca        54.0
## 4 hispânica     47.4
```

```
media_clasocial <- data %>%
  group_by(clasocial) %>%
  summarise(media_matematica = mean(matematica, na.rm = TRUE))

print(media_clasocial)
```

## Médias por Classe Social

```
## # A tibble: 3 x 2
##   clasocial media_matematica
##   <chr>          <dbl>
## 1 alta          56.2
## 2 baixa        49.2
## 3 média        52.2
```

```
media_tipescola <- data %>%
  group_by(tipescola) %>%
  summarise(media_matematica = mean(matematica, na.rm = TRUE))

print(media_tipescola)
```

## Médias por Tipo de Escola

```
## # A tibble: 2 x 2
##   tipescola media_matematica
##   <chr>          <dbl>
## 1 privada      54.8
## 2 pública      52.2
```

## Visualização das médias

```
ggplot(media_genero, aes(x = genero, y = media_matematica, fill = genero)) +
  geom_bar(stat="identity") +
  labs(title = "Média de Matemática por Gênero",
```

```
x = "Gênero",
y = "Média de Matemática") +
theme_minimal()
```

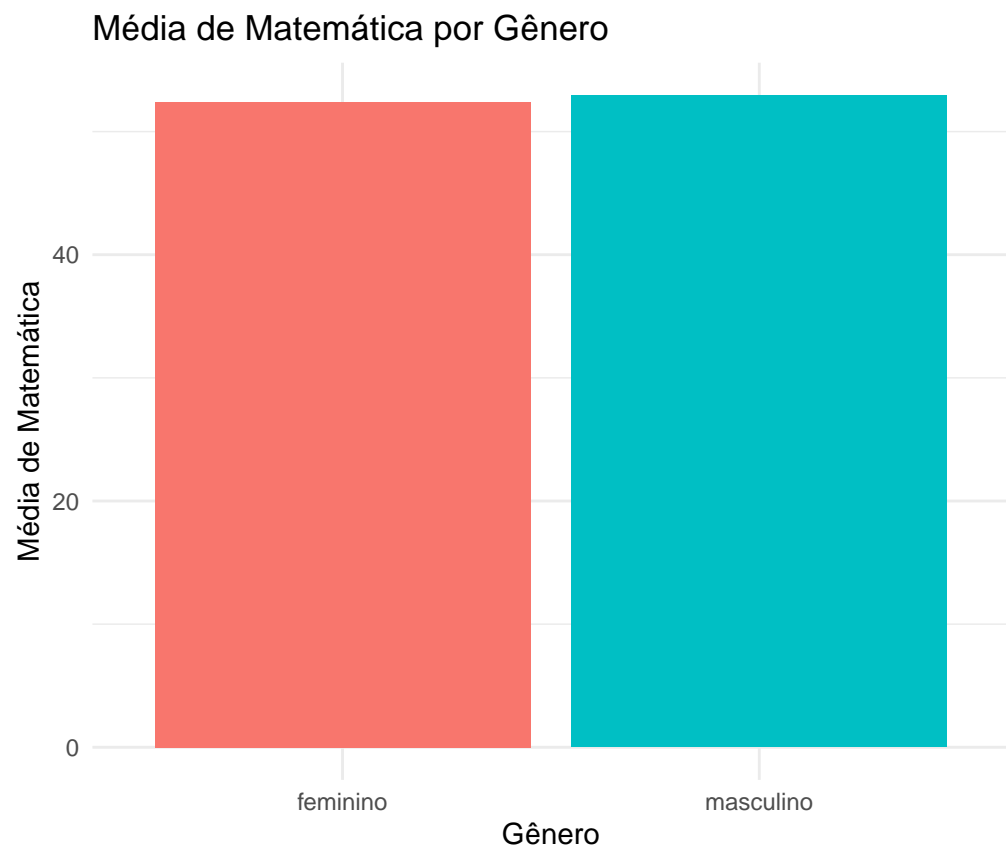


Gráfico das médias por gênero

```
ggplot(media_raca, aes(x = raca, y = media_matematica, fill = raca)) +
  geom_bar(stat="identity") +
  labs(title = "Média de Matemática por Raça",
        x = "Raça",
        y = "Média de Matemática") +
  theme_minimal()
```

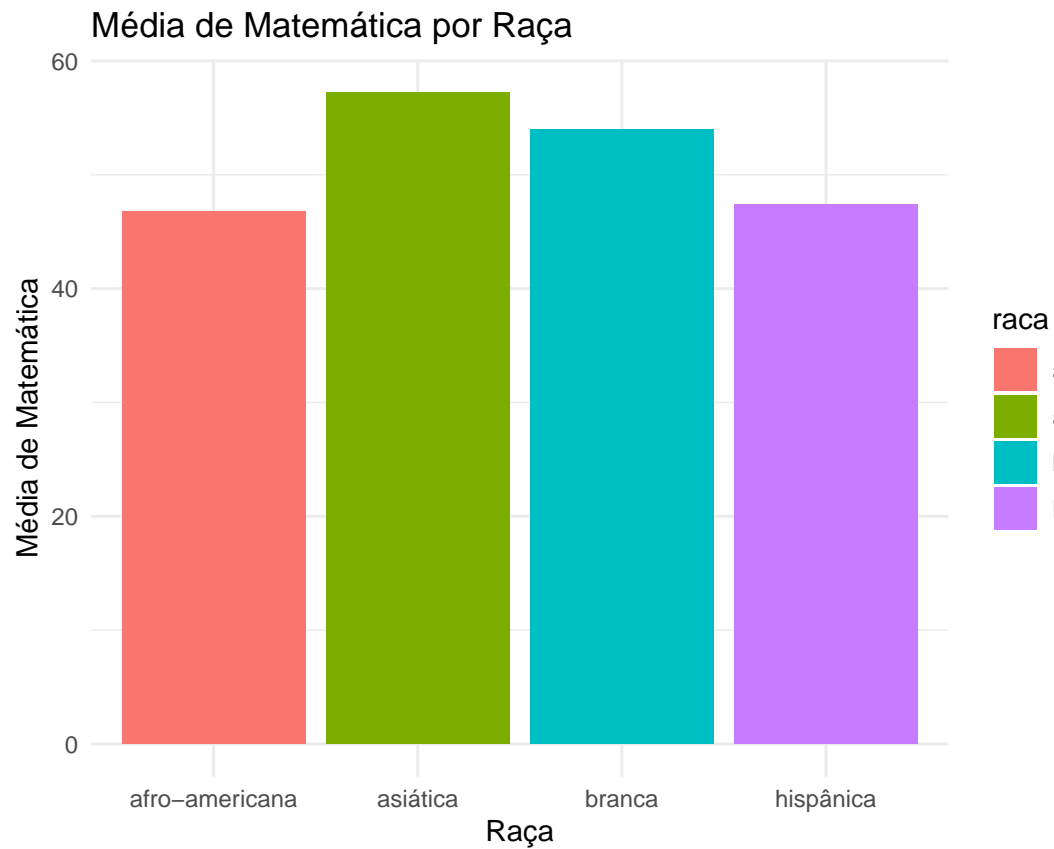


Gráfico das médias por raça

```
ggplot(media_clasocial, aes(x = clasocial, y = media_matematica, fill = clasocial)) +
  geom_bar(stat="identity") +
  labs(title = "Média de Matemática por Classe Social",
       x = "Classe Social",
       y = "Média de Matemática") +
  theme_minimal()
```

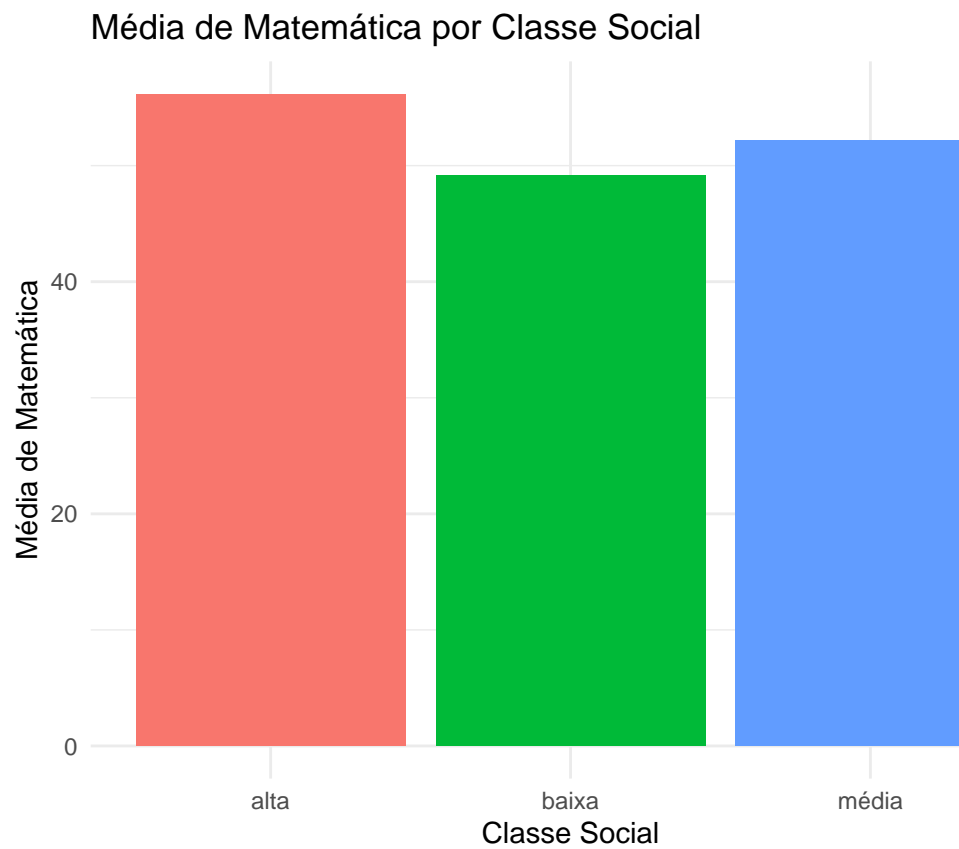
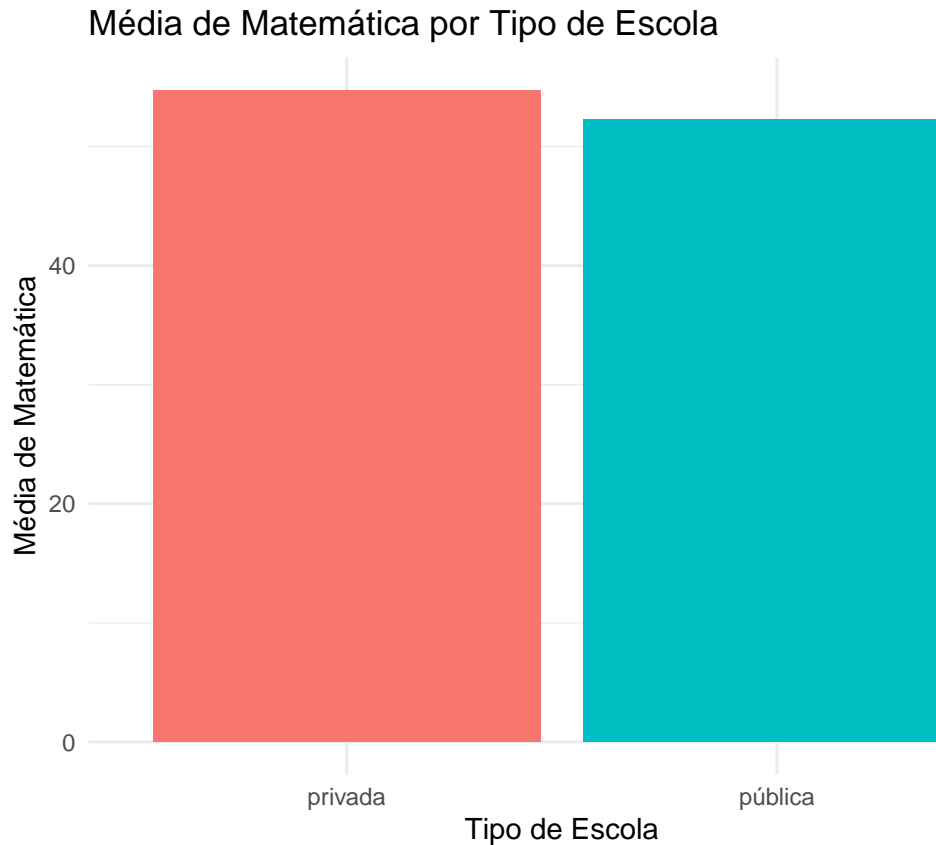


Gráfico das médias por classe social

```
ggplot(media_tipescola, aes(x = tipescola, y = media_matematica, fill = tipescola)) +  
  geom_bar(stat="identity") +  
  labs(title = "Média de Matemática por Tipo de Escola",  
        x = "Tipo de Escola",  
        y = "Média de Matemática") +  
  theme_minimal()
```





**Gráfico das médias por tipo de escola**

**Prompt:** Realizar uma análise mais aprofundada das variáveis categóricas. - Quais insights poderiam ser observados ?

**Resposta:**

Para realizar uma análise mais aprofundada das variáveis categóricas do dataset `hsb2f.csv`, podemos explorar como essas variáveis se relacionam com o desempenho acadêmico dos estudantes, especialmente nas notas de matemática.

A seguir, são apresentadas alguns *insights* que podem ser observados a partir da análise das variáveis categóricas, como `genero`, `raca`, `clasocial` e `tipescola`.

### 1. Variável `genero`

- **Análise:**
- Comparar as médias das notas de matemática entre meninos e meninas pode revelar se existe uma diferença significativa no desempenho acadêmico entre os gêneros. - Gráficos de barras ou boxplots podem ser utilizados para visualizar essas diferenças.
- **Insight:**
  - Se as meninas apresentarem notas consistentemente mais altas do que os meninos, isso pode indicar um desempenho acadêmico superior em determinadas disciplinas, o que pode ser explorado para entender fatores como motivação e apoio educacional.

### 2. Variável `raca`

- **Análise:**
  - Analisar as médias das notas de matemática por raça (branca, afro-americana, hispânica, asiática) pode evidenciar disparidades raciais no desempenho acadêmico.

- Utilizar gráficos de barras ou gráficos de violino para visualizar a distribuição das notas por raça.

- **Insight:**

- Se os estudantes brancos tiverem notas significativamente mais altas do que os estudantes afro-americanos e hispânicos, isso pode indicar desigualdades no acesso a recursos educacionais e suporte acadêmico. Essa informação é crucial para o desenvolvimento de políticas educacionais que visem a reduzir essas disparidades.

### 3. Variável `clasocial`

- **Análise:**

- Comparar as notas de matemática entre diferentes classes sociais (baixa, média, alta) pode mostrar como pertencer a determinada classe social pode impactar o desempenho acadêmico do aluno.
- Gráficos de barras ou boxplots podem ser usados para visualizar as diferenças nas médias.

- **Insight:**

- Se os alunos de classe alta apresentarem notas muito superiores aos de classe baixa, isso pode sugerir que fatores socioeconômicos influenciam diretamente no desempenho escolar. Intervenções direcionadas podem ser necessárias para apoiar estudantes de classes sociais mais baixas.

### 4. Variável `tipescola`

- **Análise:**

- Comparar as médias das notas de matemática entre alunos de escolas públicas e privadas pode ajudar a entender o impacto do tipo de escola no desempenho acadêmico.
- Gráficos comparativos podem ser utilizados para visualizar essas diferenças.

- **Insight:**

- Se os alunos em escolas privadas apresentarem notas significativamente mais altas do que os alunos em escolas públicas, isso pode indicar que escolas privadas oferecem melhores recursos educacionais ou ambientes de aprendizagem. Essa informação pode ser utilizada para justificar maiores investimentos em infraestrutura e recursos nas escolas públicas.

### 5. Interações entre Variáveis

- **Análise Avançada:**

- Explorar interações entre gênero, raça e classe social em relação às notas de matemática pode revelar padrões mais complexos.
- Usar modelos estatísticos ou gráficos multifacetados (como facetas) para investigar como essas variáveis interagem entre si.

- **Insight:**

- Por exemplo, se meninas brancas de classe alta têm notas significativamente melhores do que meninos afro-americanos de classe baixa, isso destaca a necessidade de abordagens diferenciadas para atender às necessidades específicas dos diferentes grupos demográficos.

## Conclusões Preliminares

Uma análise mais aprofundada das variáveis categóricas no dataset `hsb2f.csv` pode fornecer insights valiosos sobre o desempenho acadêmico dos estudantes.

Identificar disparidades e padrões nas notas de matemática em relação ao gênero, raça, classe social e tipo de escola é fundamental para desenvolver estratégias educacionais eficazes e promover a equidade no acesso à educação. Essas análises devem ser acompanhadas por intervenções práticas que visem melhorar o suporte a grupos sub-representados ou em desvantagem.

Citações: [1] <https://ppl-ai-file-upload.s3.amazonaws.com/web/direct-files/40230285/4d49680b-dc42-4851-8c38-40d818b1839b/hsb2f.csv>

*Prompt:* Realize essas análises utilizando a linguagem R.

*Resposta:* Para realizar uma análise mais detalhada das variáveis categóricas do dataset `hsb2f.csv` utilizando a linguagem R, pode-se seguir o passo a passo sugerido abaixo.

A análise incluirá a leitura dos dados, a exploração das variáveis categóricas e a visualização dos resultados.

## Passo a Passo em R

1. **Carregar os Pacotes Necessários:** Instale e carregue os pacotes `dplyr` e `ggplot2` para manipulação de dados e visualização.
2. **Ler o Dataset:** Carregue o arquivo CSV.
3. **Explorar as Variáveis Categóricas:** Crie gráficos para visualizar a distribuição das variáveis categóricas, como gênero, raça, classe social e tipo de escola.
4. **Analisar as Médias:** Calcule as médias das notas de matemática para cada grupo categórico.
5. **Visualizar as Médias:** Use gráficos para mostrar as médias das notas de matemática em relação às variáveis categóricas.

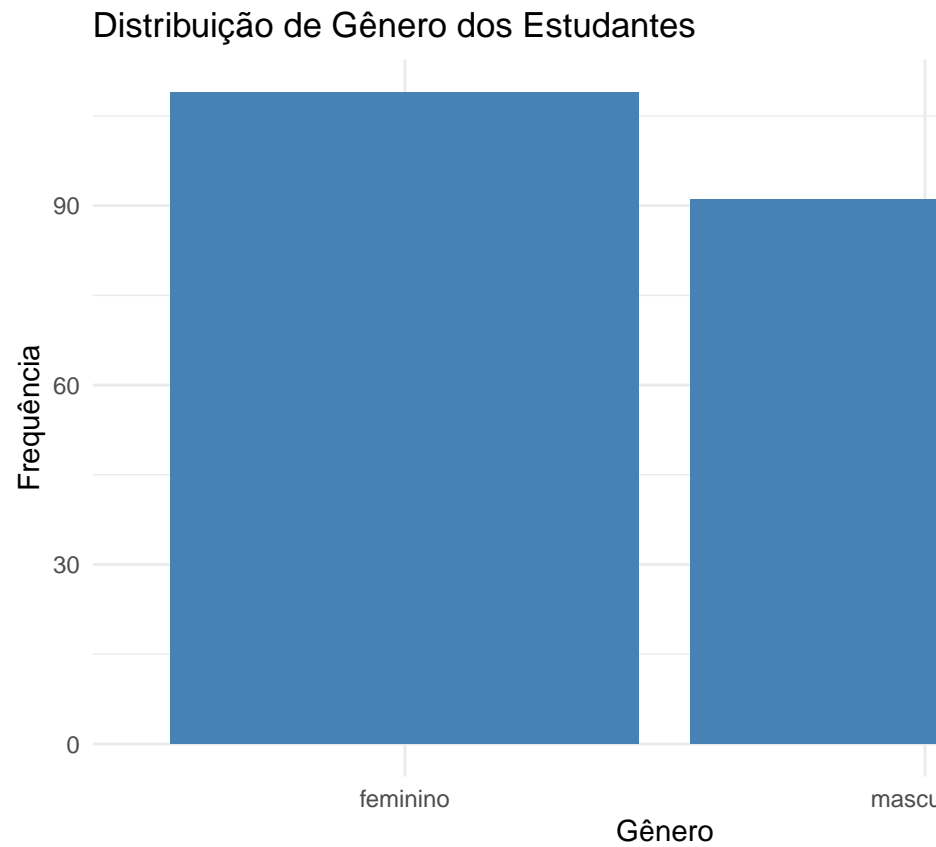
## Código em R

- As etapas Carregar pacotes necessários, Ler o arquivo CSV e Verificar as primeiras linhas do dataset já foram executadas acima!

## Análise das variáveis categóricas

### 1. Gênero

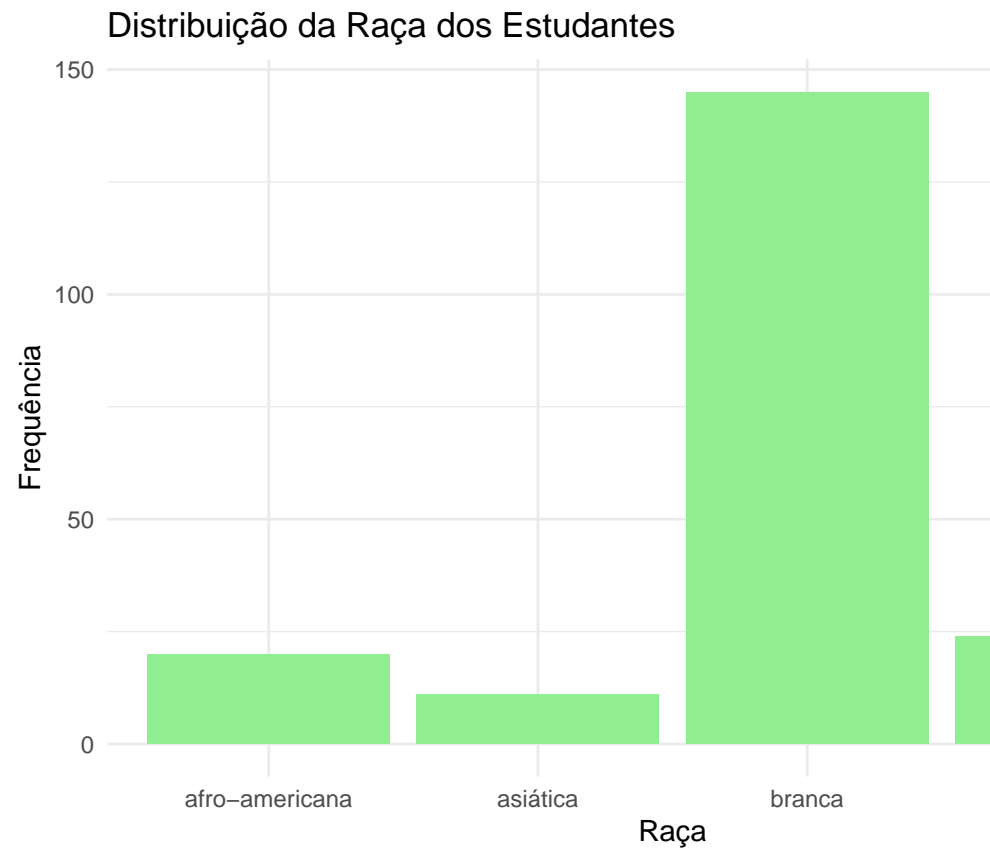
```
ggplot(data, aes(x = genero)) +  
  geom_bar(fill = "steelblue") +  
  labs(title = "Distribuição de Gênero dos Estudantes",  
        x = "Gênero",  
        y = "Frequência") +  
  theme_minimal()
```



Distribuição de gênero dos estudantes

## 2. Raça

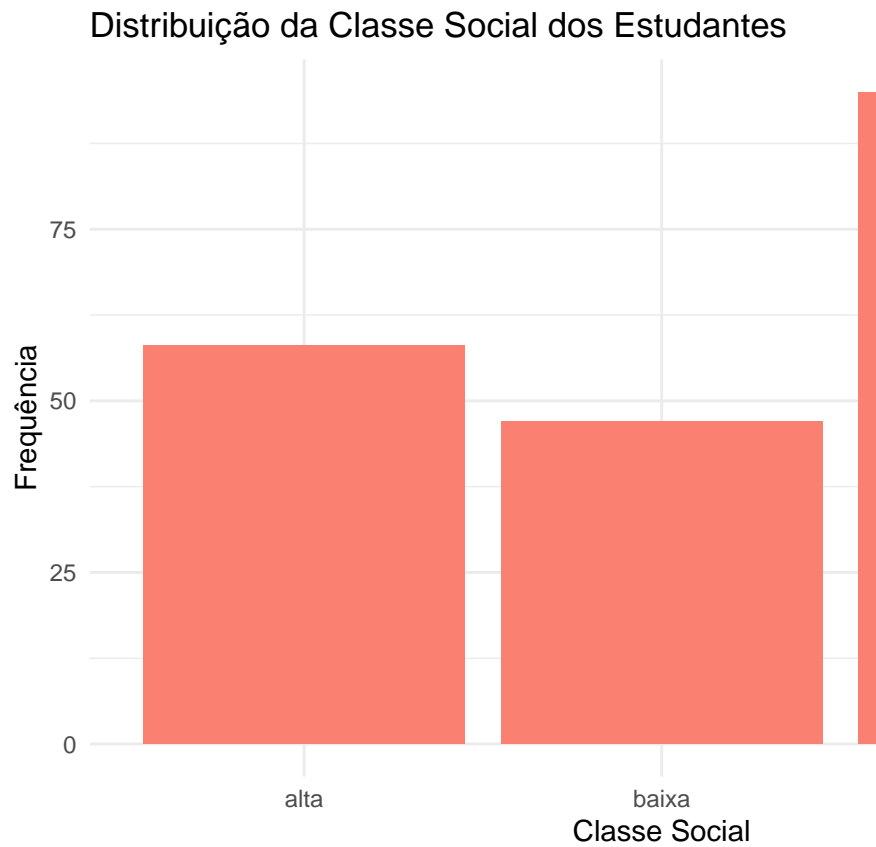
```
ggplot(data, aes(x = raca)) +  
  geom_bar(fill = "lightgreen") +  
  labs(title = "Distribuição da Raça dos Estudantes",  
        x = "Raça",  
        y = "Frequência") +  
  theme_minimal()
```



Distribuição racial dos estudantes

### 3. Classe Social

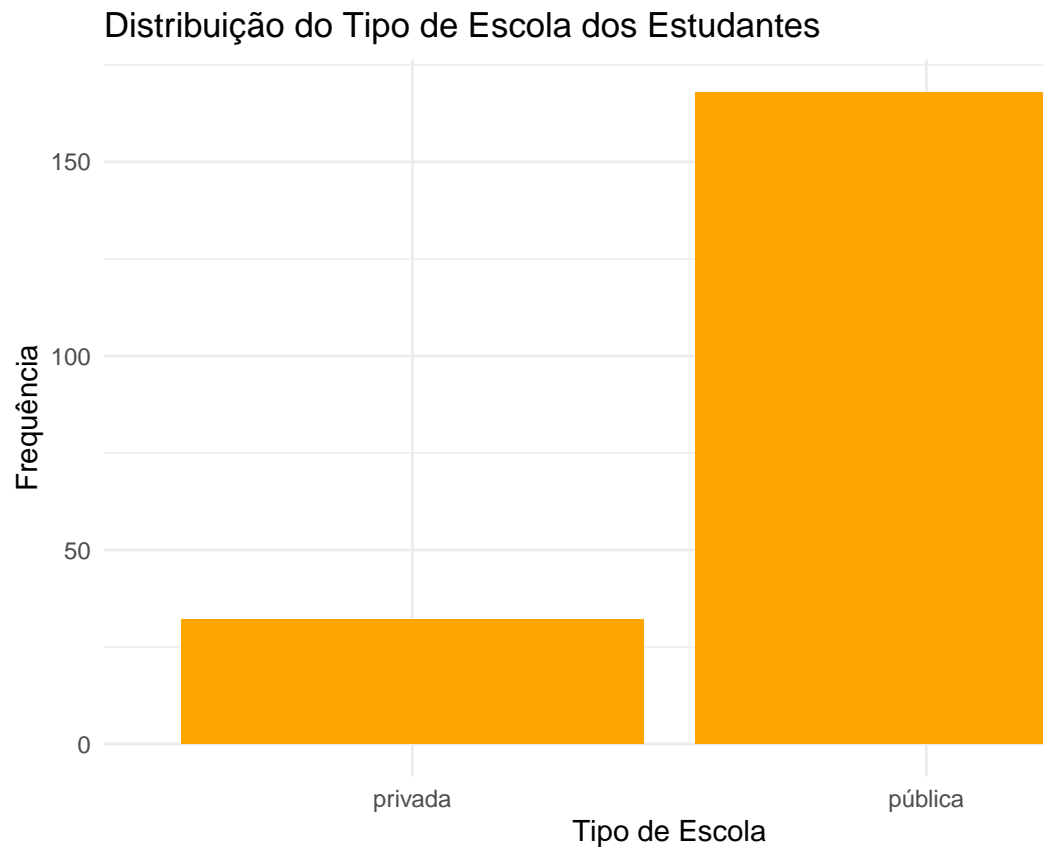
```
ggplot(data, aes(x = clasocial)) +  
  geom_bar(fill = "salmon") +  
  labs(title = "Distribuição da Classe Social dos Estudantes",  
        x = "Classe Social",  
        y = "Frequência") +  
  theme_minimal()
```



Distribuição do nível social dos estudantes

#### 4. Tipo de Escola

```
ggplot(data, aes(x = tiposcola)) +  
  geom_bar(fill = "orange") +  
  labs(title = "Distribuição do Tipo de Escola dos Estudantes",  
        x = "Tipo de Escola",  
        y = "Frequência") +  
  theme_minimal()
```



Distribuição do tipo de escola

Análise das médias das notas de matemática

Calculando as médias das notas de matemática para cada grupo categórico.

```
media_genero <- data %>%
  group_by(genero) %>%
  summarise(media_matematica = mean(matematica, na.rm = TRUE))
print(media_genero)
```

Médias por Gênero

```
## # A tibble: 2 x 2
##   genero   media_matematica
##   <chr>         <dbl>
## 1 feminino      52.4
## 2 masculino     52.9
```

```
media_raca <- data %>%
  group_by(raca) %>%
  summarise(media_matematica = mean(matematica, na.rm = TRUE))
print(media_raca)
```

## Médias por Raça

```
## # A tibble: 4 x 2
##   raca          media_matematica
##   <chr>          <dbl>
## 1 afro-americana 46.8
## 2 asiática      57.3
## 3 branca        54.0
## 4 hispânica     47.4
```

```
media_clasocial <- data %>%
  group_by(clasocial) %>%
  summarise(media_matematica = mean(matematica, na.rm = TRUE))
print(media_clasocial)
```

## Médias por Classe Social

```
## # A tibble: 3 x 2
##   clasocial media_matematica
##   <chr>          <dbl>
## 1 alta          56.2
## 2 baixa        49.2
## 3 média        52.2
```

## Médias por Tipo de Escola

```
media_tipescola <- data %>%
  group_by(tipescola) %>%
  summarise(media_matematica = mean(matematica, na.rm = TRUE))
print(media_tipescola)
```

```
## # A tibble: 2 x 2
##   tipescola media_matematica
##   <chr>          <dbl>
## 1 privada      54.8
## 2 pública     52.2
```

## Visualizando as médias

```
ggplot(media_genero, aes(x = genero, y = media_matematica, fill = genero)) +
  geom_bar(stat="identity") +
  labs(title = "Média em Matemática por Gênero",
       x = "Gênero",
       y = "Média de Matemática") +
  theme_minimal()
```



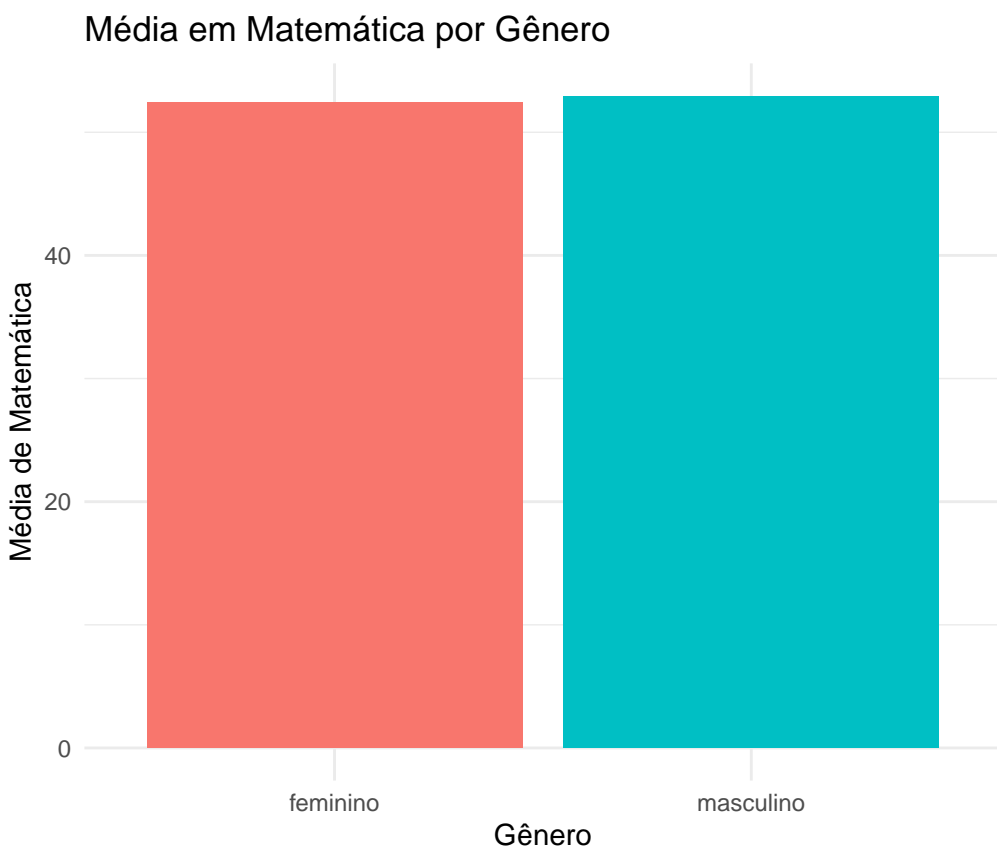
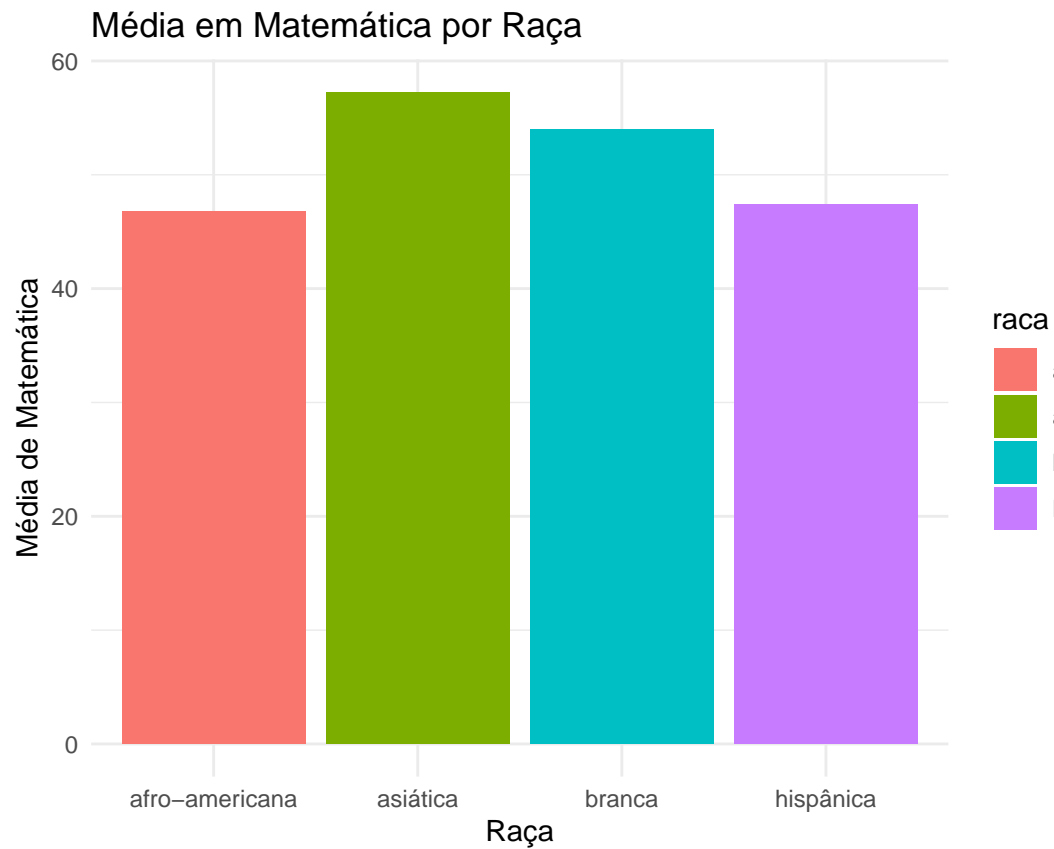


Gráfico das médias por gênero

```
ggplot(media_raca, aes(x = raca, y = media_matematica, fill = raca)) +  
  geom_bar(stat="identity") +  
  labs(title = "Média em Matemática por Raça",  
        x = "Raça",  
        y = "Média de Matemática") +  
  theme_minimal()
```



```
ggplot(media_clasocial, aes(x = clasocial, y = media_matematica, fill = clasocial)) +  
  geom_bar(stat="identity") +  
  labs(title = "Média em Matemática por Classe Social",  
        x = "Classe Social",  
        y = "Média de Matemática") +  
  theme_minimal()
```

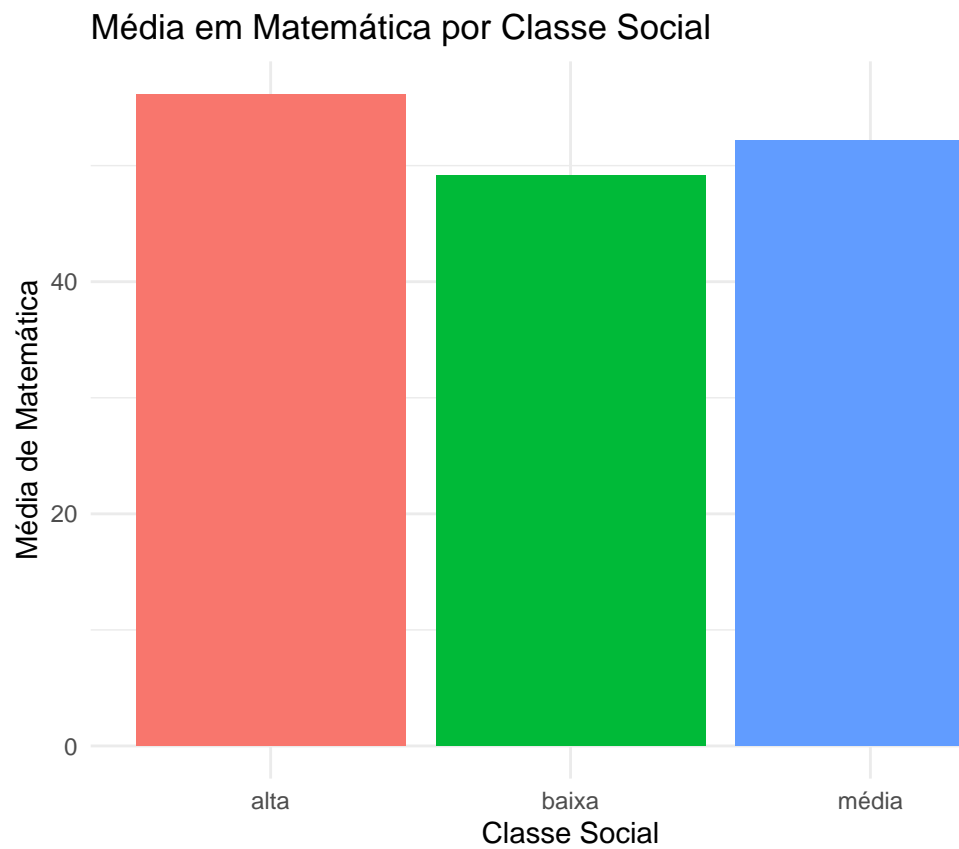


Gráfico das médias por classe social

```
ggplot(media_tipescola, aes(x = tipescola, y = media_matematica, fill = tipescola)) +  
  geom_bar(stat="identity") +  
  labs(title = "Média em Matemática por Tipo de Escola",  
        x = "Tipo de Escola",  
        y = "Média de Matemática") +  
  theme_minimal()
```

Média em Matemática por Tipo de Escola

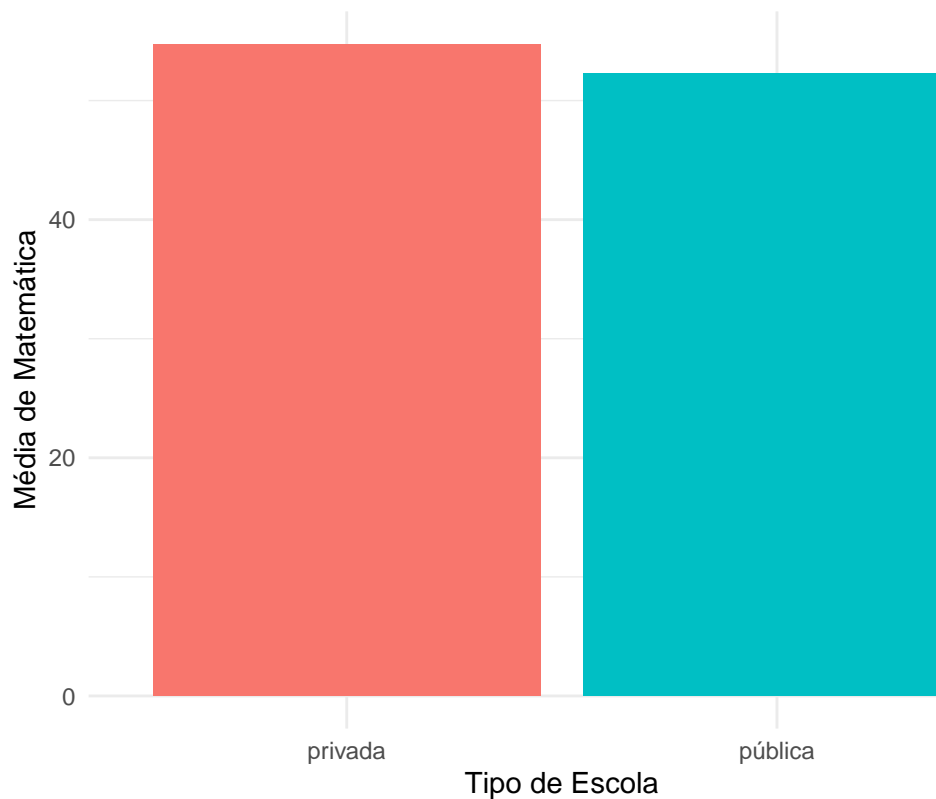


Gráfico das médias por tipo de escola

### Análise e Insights

#### 1. Distribuição do Gênero:

- O gráfico mostrará quantos alunos são do gênero masculino e feminino. Uma distribuição equilibrada pode indicar que ambos os gêneros têm igual acesso à educação.

#### 2. Distribuição da Raça:

- A distribuição racial pode revelar se há diversidade entre os estudantes. Uma predominância de um grupo racial pode indicar desigualdade no acesso à educação.

#### 3. Classe Social:

- A análise da classe social pode mostrar se há uma concentração significativa de alunos em classes sociais baixas ou altas e como isso pode afetar o desempenho acadêmico.

#### 4. Tipo de Escola:

- Comparar escolas públicas e privadas pode revelar diferenças no desempenho acadêmico com base no tipo de instituição.

#### 5. Médias das Notas:

- As médias calculadas para cada variável categórica ajudarão a identificar quais grupos estão se saindo melhor ou pior nas notas de matemática.
- Se um grupo (por exemplo, estudantes brancos ou estudantes de classe alta) tiver notas significativamente mais altas do que outros grupos (como estudantes afro-americanos ou hispânicos), isso pode indicar disparidades que precisam ser abordadas.

Esses insights podem ser usados para desenvolver estratégias educacionais que visem melhorar o desempenho acadêmico em grupos que apresentam desvantagens ou desigualdades no acesso à educação.

Citação: [1] <https://ppl-ai-file-upload.s3.amazonaws.com/web/direct-files/40230285/4d49680b-dc42-4851-8c38-40d818b1839b/hsb2f.csv>