



PONTIFÍCIA UNIVERSIDADE CATÓLICA DE GOIÁS

## **Atividade Avaliativa**

### **Introdução a Estatística:**

### **ANÁLISE DE HÁBITOS E DESEMPENHO DE ESTUDANTES**

Rairon Braga, Victor Gabriel

Goiânia – 2025

## **Introdução**

Este trabalho tem como objetivo investigar a relação entre os hábitos diários e o desempenho acadêmico de estudantes com base em um conjunto de dados contendo variáveis como horas de estudo, sono, uso de redes sociais, saúde mental, entre outras. A compreensão dessas relações pode contribuir para estratégias pedagógicas e pessoais mais eficazes, auxiliando estudantes e educadores a promoverem um melhor rendimento.

## **Objetivos**

### **Objetivo Geral:**

Analisar a influência de hábitos pessoais e comportamentais no desempenho acadêmico de estudantes.

### **Objetivos Específicos:**

- Verificar a correlação entre horas de estudo e notas.
- Avaliar o impacto do sono e do uso de redes sociais sobre o desempenho acadêmico.
- Investigar diferenças de desempenho por gênero.
- Examinar a influência da saúde mental e prática de exercícios nas notas.

## **Metodologia**

Este trabalho adota uma abordagem quantitativa e descritiva, fundamentada na análise exploratória de dados (EDA – *Exploratory Data Analysis*). O objetivo é compreender como hábitos de vida impactam o desempenho acadêmico de estudantes, utilizando técnicas estatísticas e de visualização de dados.

### **Coleta de Dados**

O dataset utilizado foi obtido da plataforma Kaggle, contendo 1.000 registros simulados de estudantes. As variáveis analisadas incluem características demográficas (idade, gênero), hábitos de estudo, uso de redes sociais, saúde mental, qualidade da dieta, frequência de exercícios, entre outras, além da variável alvo: nota na prova final.

### **Tratamento dos Dados**

As etapas de tratamento incluíram:

- Importação dos dados utilizando a linguagem R.
- Conversão de variáveis categóricas para o tipo fator.
- Verificação e tratamento de valores ausentes (não encontrados nesse dataset).
- Checagem dos tipos de dados para garantir consistência nas análises.

### **Análise Descritiva**

Foram aplicadas técnicas de estatística descritiva, como:

- Medidas de tendência central: média e mediana.
- Medidas de dispersão: desvio padrão.
- Análises segmentadas por gênero e por categorias de hábitos.

### **Visualização de Dados**

Foram utilizados gráficos diversos para facilitar a interpretação dos dados, tais como:

- Histogramas (distribuição das notas);
- Gráficos de dispersão (relações entre horas de estudo, sono, exercício e notas);
- Boxplots (comparações entre gêneros e categorias de estudo);
- Gráficos de barras (ranqueamento dos alunos com maiores e menores notas);
- Análise de correlação por meio de matrizes.

### **Ferramentas e Softwares**

A análise foi realizada integralmente na linguagem R, utilizando os seguintes pacotes:

- tidyverse: manipulação e visualização dos dados;
- dplyr: transformação dos dados;
- ggplot2: criação de gráficos;
- readr: importação dos dados.

## **6. Metodologia Estatística**

Além das análises descritivas, foi explorada a correlação entre variáveis quantitativas, permitindo identificar relações como:

- O impacto das horas de estudo na nota final;
- A influência de hábitos como sono, saúde mental e atividade física no desempenho acadêmico.

Quando pertinente, foram aplicados modelos lineares simples para observar tendências nas relações entre variáveis.

## **Conclusões**

A partir da análise realizada, foi possível observar de forma clara a influência de diversos fatores comportamentais e de estilo de vida no desempenho acadêmico dos estudantes. A análise estatística e visualização dos dados evidenciaram que hábitos como maior número de horas de estudo, boa qualidade do sono, prática regular de exercícios físicos e melhor saúde mental estão diretamente associados a notas mais altas.

Por outro lado, o excesso de tempo dedicado a redes sociais e plataformas de entretenimento, combinado com baixos índices de sono e saúde mental comprometida, mostrou-se frequentemente relacionado a desempenhos acadêmicos inferiores.

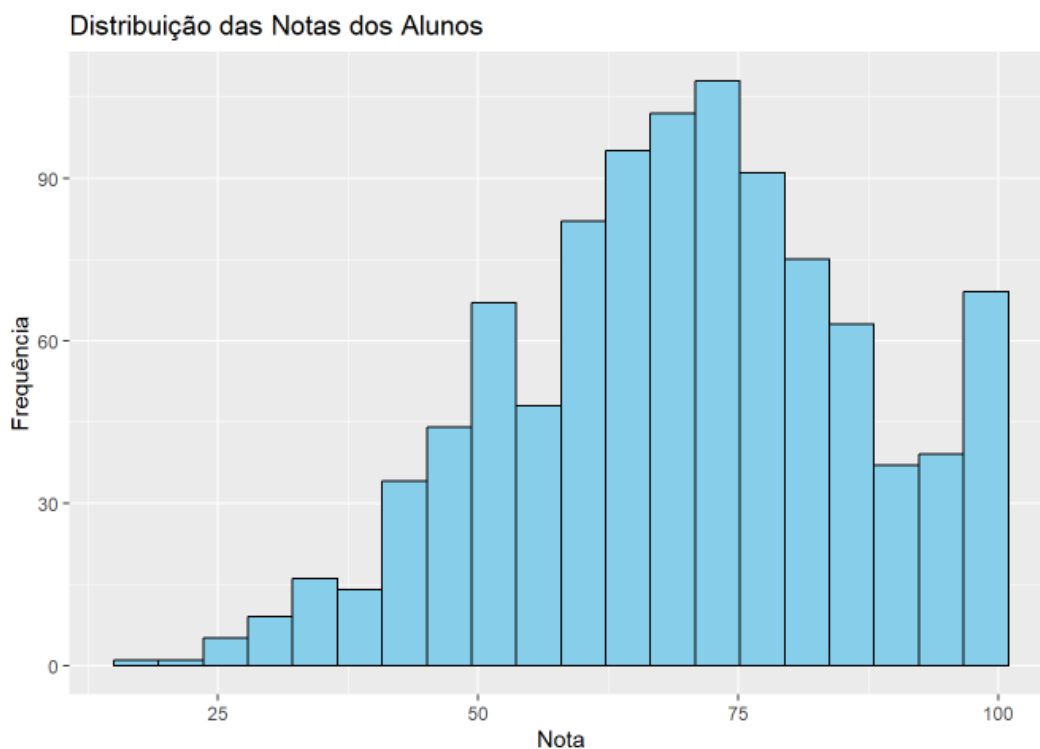
Além disso, foi possível perceber que a ausência de práticas extracurriculares, a má qualidade da internet e baixos níveis de escolaridade dos pais também podem estar associados a variações nas notas, embora esses efeitos sejam mais sutis e dependam de outros fatores associados.

Este estudo, embora baseado em dados simulados, reforça a importância de um equilíbrio saudável entre as atividades acadêmicas, o cuidado com a saúde física e mental e os hábitos de lazer dos estudantes. As informações extraídas podem servir como base para reflexões tanto dos alunos quanto de instituições educacionais, no desenvolvimento de estratégias de apoio e melhoria do desempenho acadêmico.

Por fim, a utilização de ferramentas estatísticas e de análise de dados, como o R, mostrou-se fundamental para explorar, interpretar e visualizar padrões que poderiam passar despercebidos em uma análise superficial, fortalecendo o papel da ciência de dados como apoio à tomada de decisões no âmbito educacional.

## Anexos

### Anexo 01:



#### Análise do Gráfico — Distribuição das Notas dos Alunos (Anexo 01)

O gráfico exibido representa um histograma da distribuição das notas dos alunos. É possível observar que a distribuição das notas não é completamente simétrica, apresentando uma concentração maior de alunos na faixa entre aproximadamente 60 e 80 pontos, indicando que a maioria dos estudantes obteve desempenho de médio para alto.

Há também uma quantidade considerável de alunos que atingiram notas próximas de 100, o que pode indicar que existe um grupo de alto desempenho na amostra. Por outro lado, há uma menor frequência de alunos nas faixas mais baixas, com notas abaixo de 40, sugerindo que casos de desempenho muito baixo são menos comuns.

A forma do histograma sugere uma leve assimetria à esquerda, com uma cauda estendida na direção das notas mais baixas, o que indica que embora a maioria esteja concentrada no desempenho médio-alto, existe uma minoria que apresenta dificuldades maiores.

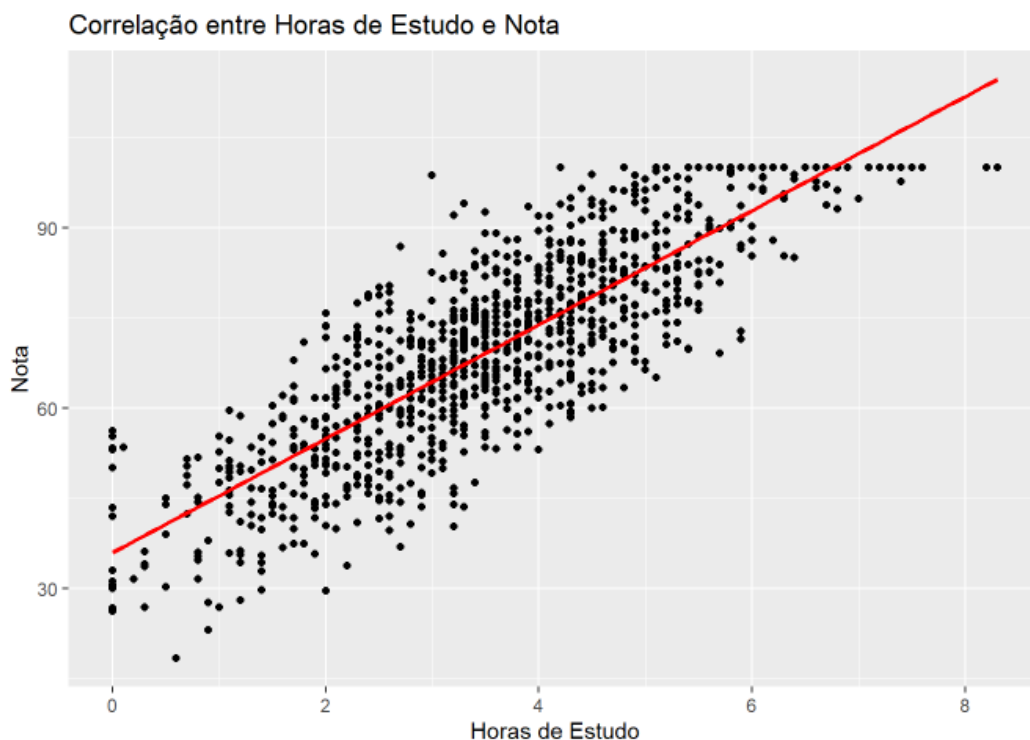
#### Interpretação Geral:

- Desempenho predominante: médio a alto (60–80 pontos).

- Alunos com desempenho muito alto: há um pico próximo de 100, destacando alguns alunos de excelência.
- Desempenho muito baixo: poucos alunos estão abaixo de 40, representando uma minoria.
- Possível distribuição assimétrica negativa, o que é comum em contextos onde parte dos estudantes recebe algum tipo de suporte ou motivação que eleva o desempenho da maioria, enquanto uma pequena parcela apresenta dificuldades.

Este tipo de análise é fundamental para que instituições de ensino possam identificar tanto alunos que necessitam de acompanhamento quanto aqueles que podem ser incentivados a desenvolver ainda mais seu potencial.

#### Anexo 02:



#### Análise do Gráfico — Correlação entre Horas de Estudo e Nota (Anexo 02)

O gráfico de dispersão apresenta a relação entre o número de horas de estudo por dia e as notas dos alunos. É possível observar uma clara relação positiva entre essas duas variáveis, representada pela linha de tendência em vermelho.

Quanto maior a quantidade de horas dedicadas ao estudo, maior tende a ser a nota obtida pelos alunos. A linha de regressão linear demonstra que existe

**uma tendência consistente: alunos que estudam mais frequentemente alcançam melhores desempenhos acadêmicos.**

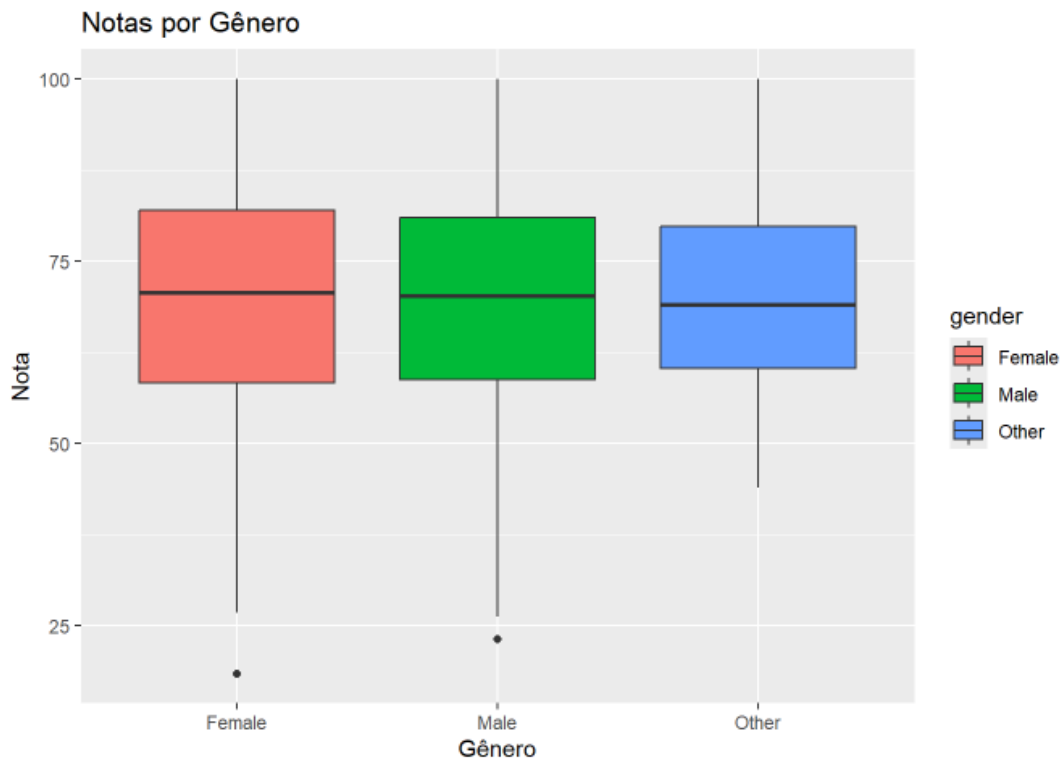
**Observações importantes:**

- **A relação é fortemente positiva, com poucos desvios da linha de tendência, embora haja certa dispersão natural nos dados.**
- **A partir de cerca de 5 horas de estudo diário, as notas começam a se concentrar mais próximas da faixa dos 90 a 100 pontos, indicando que uma dedicação elevada gera ótimos resultados.**
- **Mesmo com baixa quantidade de estudo (entre 0 e 2 horas), alguns alunos conseguem alcançar notas medianas, o que pode ser explicado por fatores como facilidade de aprendizado, qualidade do ensino, ou outros hábitos.**
- **Contudo, a maioria dos pontos abaixo da média (notas inferiores a 60) está associada a menos de 2 horas de estudo por dia, reforçando o impacto negativo da baixa dedicação.**

**Interpretação Geral:**

- **Existe uma forte correlação positiva entre horas de estudo e nota.**
- **Este resultado reforça a importância do hábito de estudo regular para o bom desempenho acadêmico.**
- **A dispersão dos dados indica que, embora estudar mais aumente significativamente a chance de obter melhores notas, existem outros fatores complementares (como sono, saúde mental, alimentação, qualidade da internet, entre outros) que também podem impactar o desempenho.**

### Anexo 03



#### Análise do Gráfico — Notas por Gênero (Anexo 03)

O gráfico de boxplot apresenta a distribuição das notas dos alunos em função do gênero (Female, Male e Other). Esse tipo de gráfico permite visualizar a mediana, os quartis, a dispersão dos dados e possíveis outliers.

#### Observações principais:

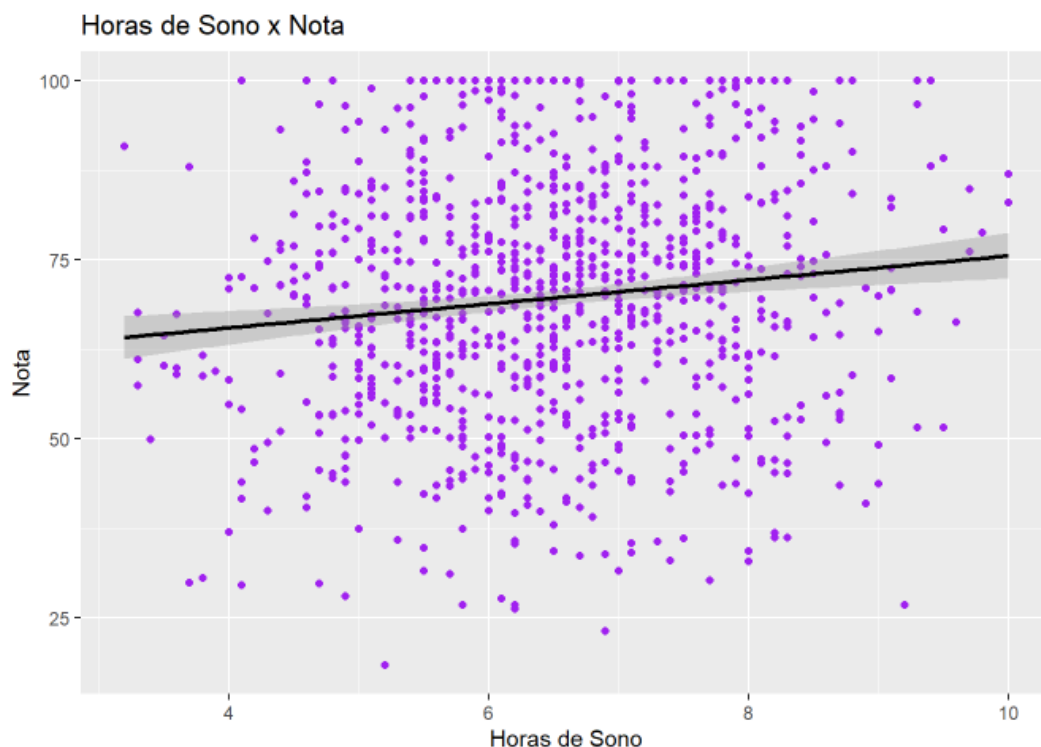
- As medianas das notas são bastante semelhantes entre os três grupos, situando-se em torno de 70 a 75 pontos, o que indica que não há uma diferença significativa no desempenho médio entre os gêneros.
- A distribuição das notas também apresenta padrões similares:
  - A maior parte dos alunos de todos os gêneros está concentrada entre 60 e 85 pontos (faixa do primeiro ao terceiro quartil).
- Existem alguns outliers (valores fora da distribuição esperada) com notas muito baixas, tanto no grupo Female quanto no Male.
- As amplitudes (distâncias entre os valores mínimo e máximo, desconsiderando outliers) são parecidas entre os gêneros, demonstrando que a variabilidade das notas é relativamente constante.

#### Interpretação Geral:



- Não há uma diferença significativa nas notas médias entre os gêneros. Isso sugere que, no contexto analisado, o desempenho acadêmico não sofre influência perceptível do gênero dos alunos.
- A distribuição equilibrada reforça a ideia de que outros fatores, como dedicação, hábitos de estudo e recursos educacionais, são mais determinantes para o desempenho acadêmico do que características demográficas como gênero.

#### Anexo 04:



#### Análise do Gráfico — Horas de Sono x Nota (Anexo 04)

O gráfico de dispersão mostra a relação entre as horas de sono e as notas dos alunos, acompanhado de uma linha de tendência (reta de regressão) para indicar o comportamento geral dos dados.

#### Observações principais:

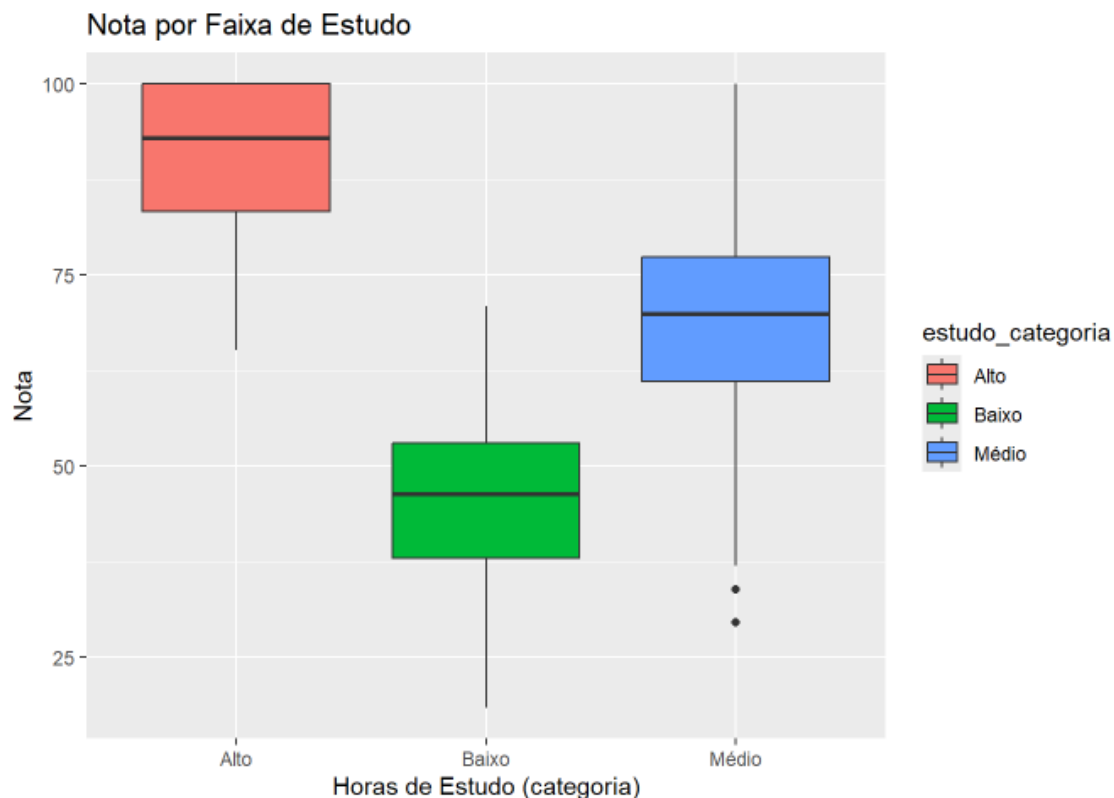
- Existe uma correlação positiva muito fraca entre horas de sono e nota. A linha de tendência levemente inclinada para cima sugere que, em média, alunos que dormem mais tendem a ter notas ligeiramente mais altas.
- Entretanto, os pontos estão muito dispersos em torno da linha, indicando que o sono não é um fator isolado fortemente determinante do desempenho acadêmico.

- Observa-se que alunos com diferentes quantidades de sono podem apresentar tanto notas altas quanto baixas. Há grande variabilidade nas notas independentemente da quantidade de horas dormidas.

#### Interpretação Geral:

- Embora o sono tenha uma leve influência positiva nas notas, o efeito é pequeno no conjunto de dados analisado.
- Isso sugere que, embora manter uma boa quantidade de horas de sono seja benéfico para a cognição e desempenho, outros fatores, como hábitos de estudo, preparação e atenção durante as aulas, parecem ter um impacto mais significativo nas notas dos alunos.

#### Anexo 05:



#### Análise do Gráfico — Nota por Faixa de Estudo (Anexo 05)

Este boxplot mostra a distribuição das notas dos alunos de acordo com a categoria de horas de estudo, dividida em três grupos: Baixo, Médio e Alto.

#### Observações principais:

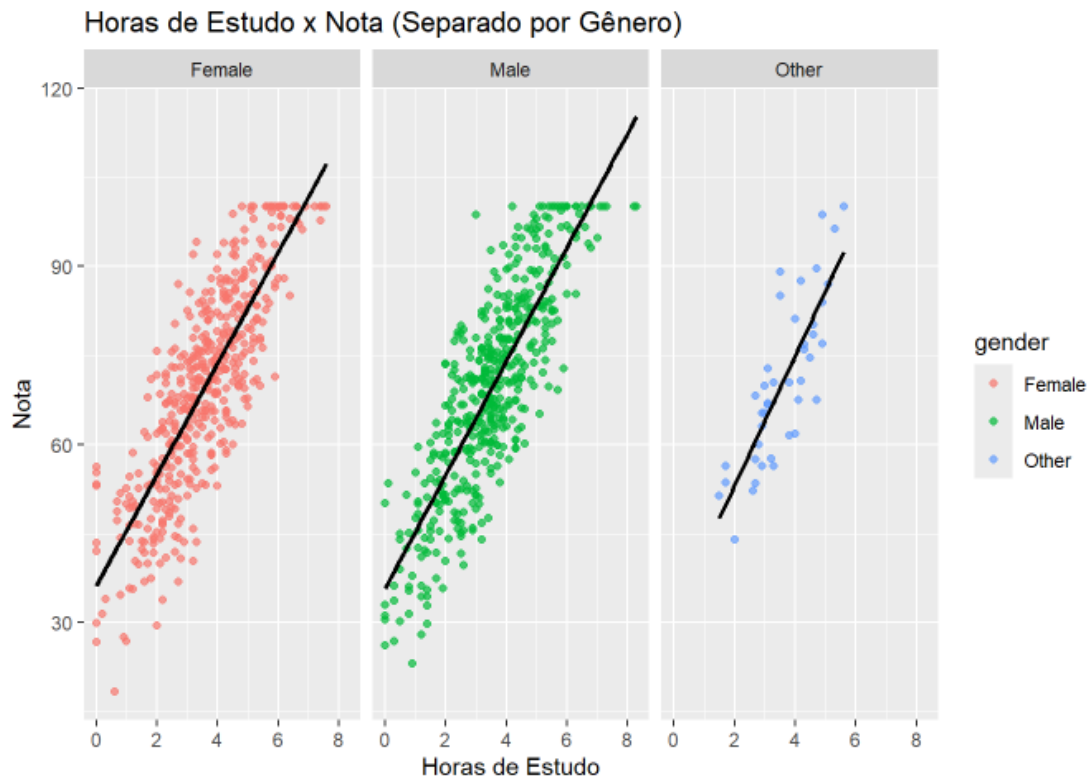
- Alunos com Alto tempo de estudo:
  - Apresentam as maiores notas, com mediana próxima de 95 a 100.

- **A dispersão é pequena, indicando que a maioria desses alunos tem desempenho excelente e consistente.**
- **Alunos com Médio tempo de estudo:**
  - **A mediana está em torno de 70, bem acima dos alunos com baixo tempo de estudo.**
  - **Existe uma variação maior nas notas, mostrando que, apesar de muitos terem notas boas, alguns apresentam desempenhos mais baixos (há outliers).**
- **Alunos com Baixo tempo de estudo:**
  - **São os que têm piores desempenhos, com mediana por volta de 45 a 50.**
  - **A dispersão é grande, com alguns outliers bem baixos, indicando que estudar pouco impacta negativamente as notas.**

**Interpretação Geral:**

- **Existe uma relação clara e direta entre o tempo de estudo e o desempenho acadêmico.**
- **Quanto maior a quantidade de horas de estudo, melhores tendem a ser as notas dos alunos.**
- **Este resultado reforça a importância do hábito de estudo para alcançar um bom desempenho.**

## Anexo 06



### Análise do Gráfico — Horas de Estudo x Nota (Separado por Gênero) (Anexo 06)

Este gráfico de dispersão apresenta a relação entre horas de estudo e nota, separado por gênero: Female (Feminino), Male (Masculino) e Other (Outro).

Observações principais por grupo:

- **Feminino (Female):**
  - Existe uma forte correlação positiva entre horas de estudo e nota.
  - As notas aumentam de forma consistente conforme as horas de estudo aumentam.
  - A maioria das estudantes se concentra até 6 horas de estudo, e as notas chegam a valores altos (próximo de 100).
- **Masculino (Male):**
  - Também apresenta uma correlação positiva muito forte.
  - A distribuição é mais espalhada em termos de horas de estudo, chegando até quase 8 horas.

- **A relação é semelhante à do grupo feminino, indicando que o impacto das horas de estudo nas notas é consistente entre os gêneros.**
- **Outro (Other):**
  - **Apesar de ter uma quantidade menor de dados, também há uma tendência positiva clara.**
  - **O comportamento segue a mesma lógica dos outros grupos: quanto mais estuda, maior é a nota.**

**Interpretação Geral:**

- **Independente do gênero, quanto mais horas de estudo, melhores são as notas.**
- **A inclinação das retas de tendência é muito semelhante nos três grupos, reforçando que o impacto do hábito de estudo sobre o desempenho acadêmico é universal, sem grandes diferenças entre gêneros.**
- **Este gráfico reforça fortemente a correlação positiva entre dedicação ao estudo e desempenho nas avaliações.**

### Matriz de Correlação

Var	Age	StudyHr	SocMedHr
Age	1.00	0.00	-0.01
StudyHr	0.00	1.00	0.02
SocMedHr	-0.01	0.02	1.00
NetflixHr	0.00	-0.03	0.01
Attend%	-0.03	0.03	0.04
SleepHr	0.04	-0.03	0.02
ExerciseFreq	0.00	-0.03	-0.04
MentalHlth	-0.05	0.00	0.00
ExamScore	-0.01	0.83	-0.17

Var	NetflixHr	Attend%	SleepHr
Age	0.00	-0.03	0.04
StudyHr	-0.03	0.03	-0.03
SocMedHr	0.01	0.04	0.02
NetflixHr	1.00	0.00	0.00
Attend%	0.00	1.00	0.01
SleepHr	0.00	0.01	1.00
ExerciseFreq	-0.01	-0.01	0.02
MentalHlth	0.01	-0.02	-0.01
ExamScore	-0.17	0.09	0.12


Var	ExerFreq	MentHlth	ExamScr
Age	0.00	-0.05	-0.01
StudyHr	-0.03	0.00	0.83
SocMedHr	-0.04	0.00	-0.17
NetflixHr	-0.01	0.01	-0.17
Attend%	-0.01	-0.02	0.09
SleepHr	0.02	-0.01	0.12
ExerFreq	1.00	0.00	0.16
MentalHlth	0.00	1.00	0.32
ExamScore	0.16	0.32	1.00

## Referencias

Dataset adquirido no site kaggle <https://www.kaggle.com/datasets>

**Linguagem R** – R Core Team (2024). *R: A language and environment for statistical computing*. R Foundation for Statistical Computing, Vienna, Austria. URL: <https://www.r-project.org/>

**ggplot2** – Wickham, H. (2016). *ggplot2: Elegant Graphics for Data Analysis*. Springer-Verlag New York. URL: <https://ggplot2.tidyverse.org>

 **dplyr, tidyr, readr, stringr, forcats** – Pacotes do *tidyverse* para manipulação e limpeza de dados.

**corrplot** ou **PerformanceAnalytics** – Pacotes utilizados para visualização de matrizes de correlação.