

DCC011/DCC222 – Introdução a Banco de Dados

Trabalho Prático

Projeto de Banco de Dados

Professor: Clodoveu Davis

Alunos:

Bernardo Zschaber Morato Nogueira – 2024066121

Cauã Neto Santos Pires – 2024005335

Mateus Matsura Teles Costa – 2024066059

Pedro Rangel Ferreira de Aguiar – 2024065907

Rafael Junqueira Baracat Silva – 2024066024

Renato Vilela de Melo Pacheco – 2024066016

ICEx – UFMG
2025/2

Sumário

1	Introdução	3
2	ER	3
2.1	Estado	4
2.2	Investimento Educacional	4
2.3	Taxa Escolar	5
2.4	Violência	5
2.5	Relações	6
2.6	Dicionário de Dados	6
3	Aquisição dos Dados	8
3.1	Estado	8
3.1.1	Análise dos dados	8
3.2	Taxa Escolar	9
3.2.1	Análise das fontes	9
3.2.2	Dificuldades de obtenção	9
3.2.3	Análise exploratória	10
3.3	Violência	12
3.3.1	Análise das fontes	12
3.3.2	Dificuldades de Obtenção	12
3.3.3	Análise Exploratória	12
3.4	Investimento Educacional	13
3.4.1	Análise das fontes	13
3.4.2	Dificuldades de obtenção	13
3.5	Análise Exploratória	14
4	Análise gráfica	15
4.1	Estatísticas	15
4.1.1	Dados relevantes	15
4.1.2	Outliers	16
4.1.3	Distribuição	17
4.2	Fundamentos da análise	18
4.2.1	Investimento Estatal x Evasão Escolar	19

4.2.2	Evasão Escolar x Taxa de Homicídios	21
4.2.3	Evolução Histórica	22
4.2.4	Correlação Geral	23
5	Conclusão	25

1 Introdução

A Lei de Acesso à Informação é um recurso garantidor da democracia brasileira, já que permite a fiscalização dos órgãos públicos e correlatos a partir da análise dos dados associados a esses. Fundamentando, dessa maneira, a habilitação dos cidadãos brasileiros na percepção do status quo vigente e possibilitando maior precisão das exigências da população em relação às ações públicas.

Nesse viés, visamos aplicar a LIA para identificar e analisar o impacto dos investimentos públicos em educação na retenção dos alunos nas instituições de ensino, especificamente o fenômeno de evasão escolar e sua relação com as ocorrências de violência nos estados brasileiros.

Para tal, em primeiro momento, iremos descrever as fontes onde captamos os dados que irão fomentar nossa análise, além das dificuldades de captação das informações nos recursos públicos. Além disso, será apresentado o esquema Entidade-Relacionamento (ER) que mapeia os dados obtidos em tabelas, que, por sua vez, serão usadas para posterior análise em um ambiente SGBD, especificamente o PostgreSQL. Por fim, serão apresentadas análises sobre os dados adquiridos e a interpretação da relação entre eles a respeito do fenômeno estudado.

2 ER

Nesta seção, é apresentado o diagrama Entidade-Relacionamento (ER) desenvolvido para o projeto. A Figura 1 ilustra a estrutura geral do modelo.

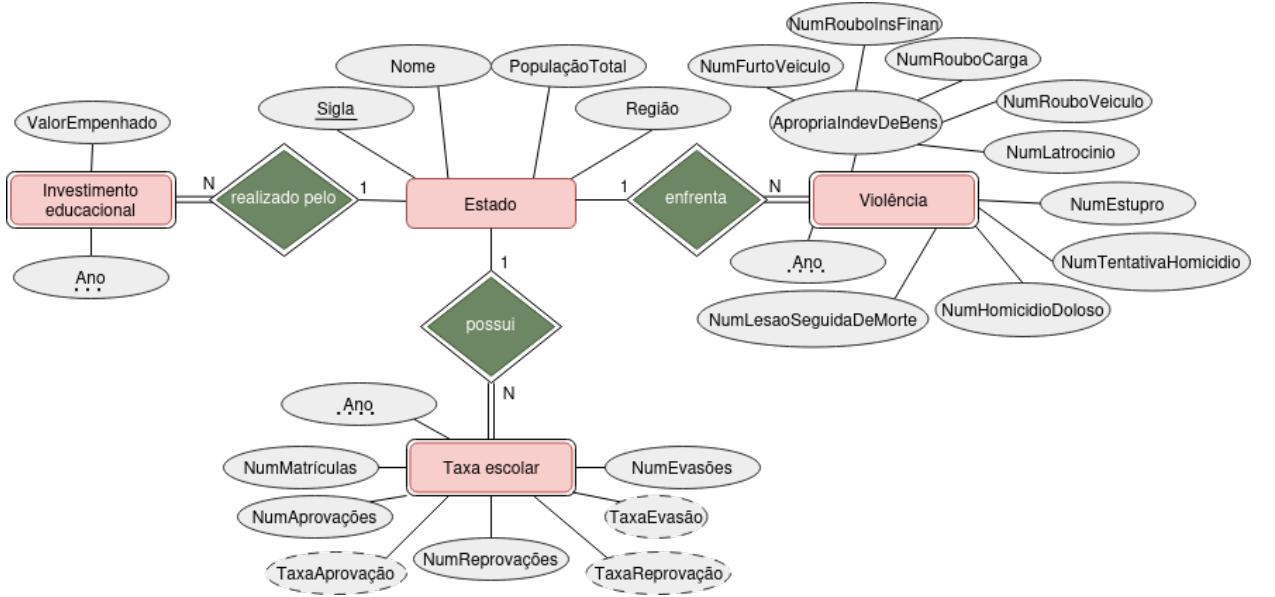


Figura 1: Diagrama Entidade-Relacionamento do sistema.

Sobre a estrutura geral, o diagrama indica quatro entidades: **Estado**, **Investimento educacional**, **Violência** e **Taxa Escolar** que, juntas, são responsáveis por modelar a análise descrita na introdução 1. A seguir, faremos uma análise dos atributos dessas entidades, suas relações e suas cardinalidades:

2.1 Estado

- Sigla: chave primária da entidade (BA, SE, MG, ...);
- Nome: nome do estado;
- População Total: indica a quantidade de pessoas residentes no estado;
- Região: indica a macroregião do estado;

2.2 Investimento Educacional

Entidade fraca em relação ao Estado.

- Valor empenhado: valor reservado para o investimento educacional;

- Ano: chave parcial da entidade;

2.3 Taxa Escolar

Entidade fraca em relação ao Estado.

- Ano: chave parcial da entidade;
- NumMatrículas: número de estudantes matriculados;
- NumAprovações: número de estudantes aprovados;
- NumReprovações: número de estudantes reprovados;
- NumEvasões: número de estudantes que abandonaram a formação;
- TaxaAprovação: relação entre matrículas e aprovados;
- TaxaReprovação: relação entre matrículas e reprovados;
- TaxaEvasões: relação entre matrículas e evasões;

2.4 Violência

Entidade fraca em relação ao Estado.

- Ano: chave parcial da entidade;
- ApropriaIndevDeBens: número de crimes de natureza "Apropriação indevida de bens";
- NumEstupro: número de vítimas de violência sexual;
- NumTentativaHomicidio: número de atentados contra a vida;
- NumHomicioDoloso: número de atentados contra a vida com intenção de matar;
- NumLesaoSeguidaDeMorte: número de lesões seguidas de morte;

2.5 Relações

- Estado x Investimento Educacional: relação 1-N, um estado pode ter vários investimentos educacionais, e a relação é total para o lado do investimento;
- Estado x Violência: relação 1-N, um estado pode ter vários casos de violência, e é total para o lado de violência;
- Estado x Taxa Escolar: relação 1-N, um estado pode ter várias taxas diferentes, e é total para o lado da taxa;

2.6 Dicionário de Dados

Legenda de Abreviações das Tabelas:

- **PK:** Chave Primária (Primary Key)
- **FK:** Chave Estrangeira (Foreign Key)
- **ref.:** Referencia
- **ODC:** On Delete Cascade
- **(1), (2):** Indicam as partes de uma Chave Primária composta

Atributo	Tipo/Largura	Nulo?	Único?	Restrições
Sigla	char(2)	N	S	PK
Nome	varchar(100)	N	S	—
População Total	BIGINT	N	N	—
Região	varchar(50)	N	N	—

Tabela 1: Dicionário da Relação: **Estado**

Atributo	Tipo/Largura	Nulo?	Único?	Restrições
Ano	number(4)	N	N	PK (1)
Sigla	char(2)	N	N	PK (2), FK ref. Estado(Sigla) (ODC)
Valor empenhado	INT	N	N	—

Tabela 2: Dicionário da Relação: **Investimento Educacional**

Atributo	Tipo/Largura	Nulo?	Único?	Restrições
Ano	number(4)	N	N	PK (1)
Sigla	char(2)	N	N	PK (2), FK ref. Estado(Sigla) (ODC)
Aprovações	INT	N	N	—
Reprovações	INT	N	N	—
Evasões	INT	N	N	—
Matrículas totais	INT	N	N	—

Tabela 3: Dicionário da Relação: **Taxa Escolar**

Atributo	Tipo/Largura	Nulo?	Único?	Restrições
Ano	number(4)	N	N	PK (1)
Sigla	char(2)	N	N	PK (2), FK ref. Estado(Sigla) (ODC)
Estupros	INT	N	N	–
Furtos de veículo	INT	N	N	–
Homicídios dolosos	INT	N	N	–
Lesões corporais seguidas de morte	INT	N	N	–
Roubos a instituições financeiras	INT	N	N	–
Roubos de carga	INT	N	N	–
Roubos de veículo	INT	N	N	–
Latrocínios	INT	N	N	–
Tentativas de homicídio	INT	N	N	–

Tabela 4: Dicionário da Relação: **Violência**

3 Aquisição dos Dados

A aquisição dos dados foi feita com os recursos disponibilizados no arquivo de orientação do trabalho prático e foi associada a múltiplas fontes, com o objetivo de garantir consistência nas informações apresentadas. A carga dos dados foi feita utilizando arquivos .csv e o comando "copy" no PostgreSQL.

3.1 Estado

3.1.1 Análise dos dados

Os dados de população seguiram o censo do IBGE- Instituto Brasileiro de Geografia e Estatística - responsável pela emissão de dados voltados à demografia brasileira.

3.2 Taxa Escolar

3.2.1 Análise das fontes

As fontes dos dados estão vinculadas ao INEP - Instituto Nacional de Estudos e Pesquisas Educacionais Anísio Teixeira - instituição pública associada ao Ministério da Educação, responsável por censos que nos permitem identificar a amplitude dos resultados decorrentes de ações públicas voltadas à educação.

Foram usadas duas fontes principais:

- **Painéis estatísticos censo escolar inepdata:** voltada para a obtenção de dados relacionados ao número de matrículas nas instituições públicas.
- **Taxas de Rendimento Escolar:** voltada para a obtenção de dados para a identificação dos números de retenção, evasão e aprovação dos estudantes.

3.2.2 Dificuldades de obtenção

A obtenção dos dados referentes ao número de matrículas do ensino básico foi dificultada pela limitação nos formatos disponíveis para extração das informações. O único formato acessível estava no Power BI, o que tornou o processo mais custoso, pois foi necessário realizar a captura dos indicadores manualmente.

Outro ponto que impactou a busca pelos números de matrículas por unidade federativa (UF) foi a indisponibilidade de dados consolidados sobre reprovações, aprovações e evasões. Inicialmente, esperávamos encontrar essas informações organizadas por UF e de forma compatível com nossa proposta; contudo, os dados disponíveis estavam em formato percentual e divididos apenas entre ensino fundamental e médio. Essa separação por nível de ensino dificultou a obtenção dos totais de matrículas, afinal, elas precisaram ser tratadas individualmente, levando em conta essa divisão. Após o trabalho de coleta manual dos dados — número de matrículas e taxas correspondentes —, foi necessário realizar a manipulação e o tratamento das informações para alcançar os resultados desejados.

A imagem a seguir ilustra o único meio encontrado para a captura do número de matriculados por unidade federativa.

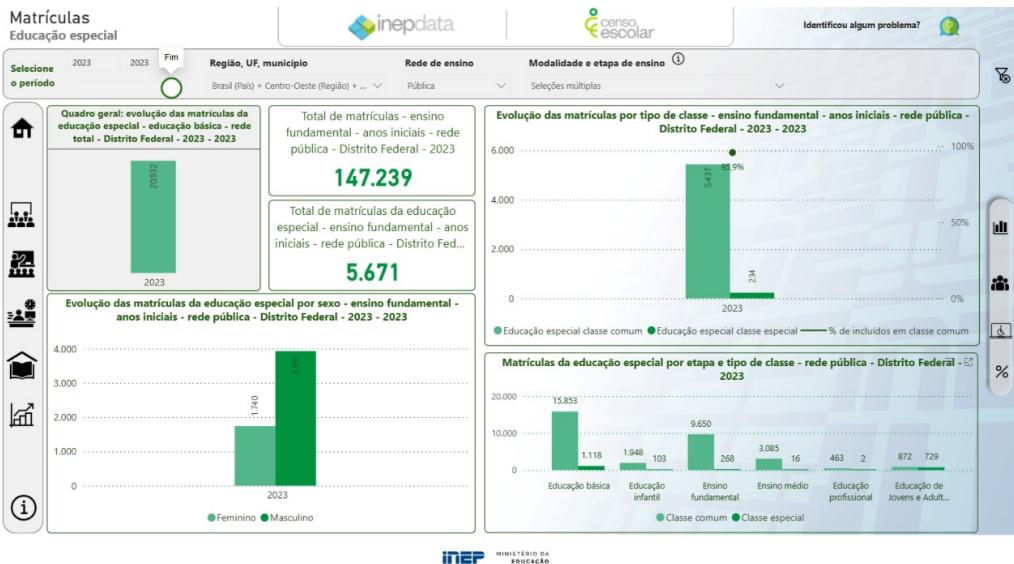


Figura 2: Única formatação dos dados encontrada

3.2.3 Análise exploratória

Entende-se que a necessidade de utilização de duas fontes, para além de garantir uma completude de pesquisa, decorre da carência de dados significativos de apenas uma das partes. Por isso, foi preciso extraír informações de ambas as plataformas e realizar o remanejamento delas em uma planilha para a obtenção dos dados de interesse.

Sobre a eficiência da obtenção dos dados, destacam-se os problemas:

- **Disponibilidade e acesso aos dados:**

- Dificuldade para encontrar dados completos de anos anteriores, com séries históricas incompletas ou inconsistentes;
- impossibilidade de extração direta em formato .csv, limitando a automação da análise; ausência de suporte a API para importação programática dos dados, exigindo extração manual.

- **Usabilidade e desempenho:**

- Sistema lento e suscetível a falhas na atualização ao modificar critérios de seleção;
- Navegação confusa para extração dos dados, com menu extenso e hierarquia pouco intuitiva;
- Suporte limitado a múltiplos filtros simultâneos, restringindo análises mais complexas.

Sobre a qualidade dos dados obtidos, destacam-se os problemas:

• Formatação e compatibilidade:

- Alta granularidade dos dados sem documentação adequada, dificultando a identificação das variáveis relevantes;
- Utilização da extensão .xls (modelo antigo do Excel) em detrimento do formato mais atual .xlsx, com possíveis limitações de compatibilidade.

• Critérios de obtenção distoantes temporalmente:

- Mudança na disposição do conteúdo e layout das planilhas ao longo dos anos, exigindo adaptação constante dos scripts de processamento;
- Alteração nos parâmetros e metodologia do estudo conforme as diferentes edições, comprometendo a comparabilidade entre anos;
- Separação dos dados do ensino fundamental e médio em arquivos distintos, aumentando a complexidade da consolidação

• Exposição e organização:

- Estrutura visual que dificulta a localização dos dados brutos desejados, com informações relevantes dispersas entre múltiplas abas e seções;
- Presença de células mescladas, formatação condicional e outros elementos de apresentação que interferem no processamento automatizado

3.3 Violência

3.3.1 Análise das fontes

Os dados foram extraídos da solução do SINESP - Sistema Nacional de Informações de Segurança Pública -, órgão responsável pela compilação de informações de interesse para consultas operacionais, investigativas e estratégicas sobre segurança pública, crime, sistema prisional e drogas. Nesse sentido, a extração visou captar o aspecto de ações que ferem a integridade física ou material dos afetados, entre elas homicídios, furtos e lesões. Foi usada uma fonte principal:

- **Ocorrências Criminais:** voltada para a obtenção de dados dentro do espectro caracterizado como "violência", seja esta sexual, física ou material.

3.3.2 Dificuldades de Obtenção

A estruturação dos dados promovida pelo SINESP favorece aspectos de análise, dada a característica principal de seu fornecimento voltada à investigação. Por isso, não houveram grandes dificuldades na obtenção e formatação dos dados.

3.3.3 Análise Exploratória

Os dados dispostos pelo SINESP são apresentados de maneira a garantir a qualidade dos dados, uma vez que a granularidade e a consistência das informações são presentes, isso combinado a um acesso facilitado possibilitam uma melhor análise do fenômeno proposto no trabalho. Especificamente, as tabelas disponíveis demonstram granularidade mensal na questão de tempo e apresenta divisão de dados tanto por unidade federativa quanto por município. Destaca-se, por fim, que seria interessante se a formatação dos dados seguisse o padrão-coluna na apresentação dos dados, ao invés do padrão-linha para cada tipo de crime, de modo a reduzir a redundância na coleta.

3.4 Investimento Educacional

3.4.1 Análise das fontes

Os dados foram obtidos do SINCOFI - Sistema de Informações Contábeis e Fiscais do Setor Público Brasileiro- sistema vinculado ao Ministério da Fazenda e responsável pela obtenção de dados financeiros e contábeis, sobretudo para uso de entes federativos. Nesse sentido, a extração visou captar, primeiramente, a relação dos recursos teoricamente alocados para o setor educacional e os investimentos verdadeiramente efetivados, contudo esse dado não consta de maneira plena na fonte. Foi usada uma fonte principal:

- Relatórios Resumidos da Execução Orçamentária: voltada para a obtenção das relações de investimento no setor educacional brasileiro, com uma consulta excessivamente técnica, fruto da natureza dos dados do órgão voltados às políticas públicas.

3.4.2 Dificuldades de obtenção

A respeito dos dados das despesas públicas do país destinadas à educação, de forma similar aos conteúdos anteriores, a extração do tipo de conteúdo que buscamos para a análise também se mostrou desafiadora. Nesse contexto, foi preciso que a pesquisa atravessasse por diversas fontes diferentes do governo, principalmente devido ao formato ou agregação de cada site que em grande maioria não atendiam à demanda do projeto.

Destaca-se, primeiramente, que, ao longo do processo de coleta, entendemos a viabilidade do portal da transparência para a obtenção das informações buscadas. As duas principais categorias de dados que julgamos úteis para a análise são: a agregação das despesas empenhadas e aquelas que foram realmente efetuadas (pagas). Dessa forma, esse portal do governo não contemplava exatamente essa demanda. Na verdade, a base de dados mais próxima do ideal foi um espaço de consulta detalhada que permitia a pesquisa de todos os documentos emitidos relacionados à despesa pública. Com essa fonte, e uma filtragem adequada, o nosso objetivo seria alcançado se fosse possível a coleta dos dados de todos os documentos para a agregação adequada em um formato interpretável para o projeto. No entanto, esse pro-

cesso foi descartado, pois o volume gigantesco de dados não pôde nem ser resgatado da base devido às restrições de download de consultas por número de registros. Essa realidade exemplifica uma forma engenhosa de burlar a Lei do Acesso à Informação, escondendo os dados por meio de uma granularidade altíssima das informações.

Após essa conclusão devastadora, recorremos a outro website do governo para a coleta desses dados, sendo este o *Siconfi*, sistema de dados associado ao tesouro nacional. Por mais que a navegação por esse sistema se mostrou complicada - devido ao teor mais técnico da fonte - a obtenção dos dados foi bem mais agradável por meio de consultas do *Relatório Resumido da Execução Orçamentária (RREO)*, que já permitia uma agregação imediata dos valores no escopo do estado, divididas por ano. Contudo, essa fonte não apresentava de forma alguma os valores pagos de despesa pública, o que também proporciona uma experiência incompleta para o experimento. Como não conseguimos encontrar nenhuma outra fonte comparável para esse elemento do projeto, foi decidido que a análise prosseguirá apenas considerando os valores empenhados de investimento público na educação.

3.5 Análise Exploratória

Para além das dificuldades expostas acima, a qualidade dos dados encontrada no Sinconfi foram suficientes tanto no teor granularidade quanto pelo viés de consistência, além disso não houveram dados faltantes no período analisado. Isso garantiu que as consultas e análises derivadas fossem feitas com maior facilidade.

4 Análise gráfica

4.1 Estatísticas

4.1.1 Dados relevantes

Dimensão	Média	Desvio Padrão	Mínimo	Mediana	Máximo	Amplitude
Investimento por Aluno (R\$)	4522.82	2739.0	1361.77	3677.53	18324.78	16963.01
Taxa de Evasão (%)	2.96	1.69	0.15	2.71	7.84	7.69
Homicídios (por 100 mil hab.)	26.99	11.54	6.1	25.98	59.11	53.0

Tabela 5: Estatísticas Descritivas das Variáveis

Outliers (3 maiores valores por Dimensão):

Investimento por Aluno (R\$):

1. DF (2015): R\$18324.78
2. DF (2022): R\$16923.29
3. DF (2016): R\$16593.76

Taxa de Evasão (%):

1. AL (2015): 7.84 %
2. RN (2021): 7.61 %
3. SE (2015): 7.55 %

Homicídios (p/ 100 mil hab.):

1. SE (2016): 59.11
2. RN (2017): 56.38
3. PE (2017): 54.21

4.1.2 Outliers

Aqui utilizamos os **Boxplots** (Diagramas de Caixa) para identificar quem foge da regra.

- A **Caixa Colorida** contém 50% dos dados (a "maioria normal").
- A **Linha dentro da caixa** é a Mediana (o valor central exato).
- Os **Pontos flutuando** fora das linhas pretas (bigodes) são os **Outliers** (Discrepantes).

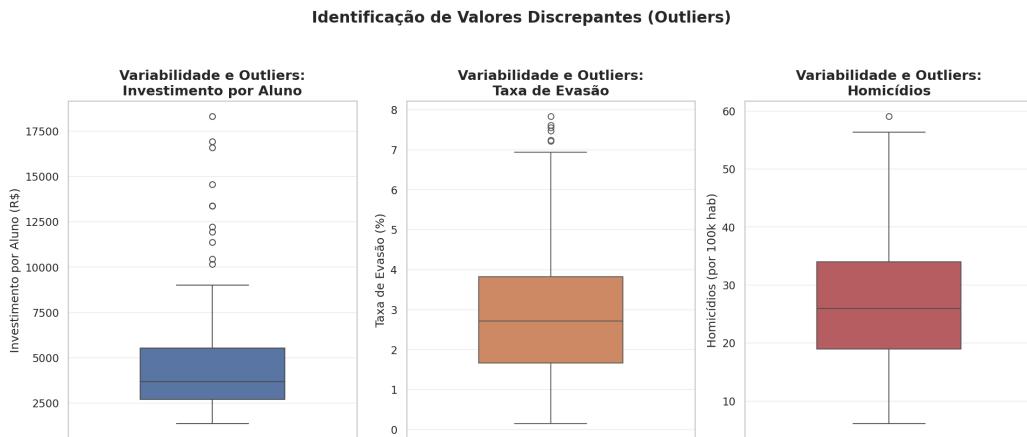


Figura 3: Diagrama de caixa para identificação de Outliers

Análise dos Outliers Identificados:

1. **Investimento (Azul)**: Existem vários pontos acima do limite superior. Isso confirma que alguns estados (provavelmente DF ou SP em anos específicos) investem valores muito acima da realidade nacional.
Decisão Metodológica: Não excluir. Estes não são erros de digitação, mas representam a realidade de entes federativos com maior capacidade de arrecadação.
2. **Evasão (Laranja)**: Há poucos outliers superiores. Isso indica que, embora a média seja 3%, existem momentos de crise aguda onde a evasão dispara para quase 8%.

3. **Violência (Vermelho):** Nota-se pontos isolados na parte superior (acima de 50/60 por 100k hab). Estes representam estados vivendo crises agudas de segurança pública.

4.1.3 Distribuição

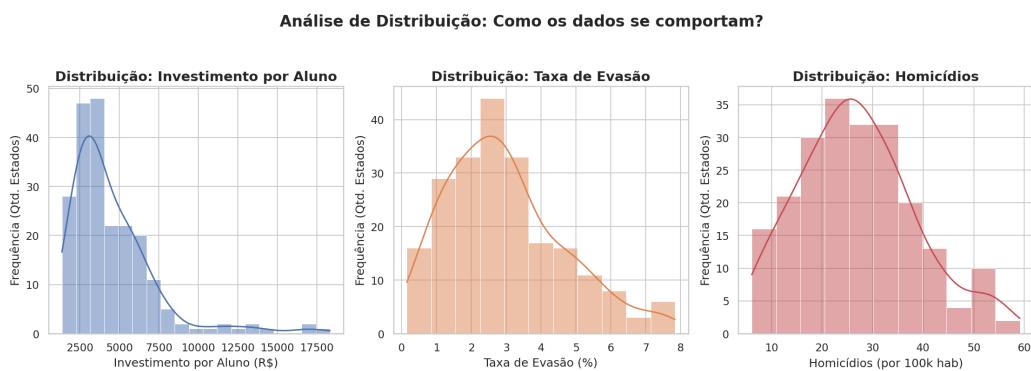


Figura 4: Histogramas de distribuição

1. Investimento por Aluno (R\$)

- **Leitura dos Dados:** A tabela mostra uma **Média** de R\$ 4522.82, mas com um **Desvio Padrão** altíssimo de R\$2739.0. A **Amplitude** (diferença entre o que mais investe e o que menos investe) é brutal: quase 17 mil reais de diferença.
- **Interpretação do Histograma (Gráfico Azul):** O gráfico é **assimétrico à direita** (a cauda longa vai para a direita). *O que isso significa:* A grande maioria dos estados brasileiros se concentra no lado esquerdo do gráfico (baixo investimento, entre R\$ 2.000 e R\$ 6.000). Pouquíssimos estados (as barras pequenas à direita) investem valores altos (acima de R\$ 15.000).
- **Conclusão Teórica:** Há uma profunda **desigualdade orçamentária**. A maioria dos alunos do sistema público conta com recursos limitados, enquanto uma minoria em estados específicos conta com recursos abundantes.

2. Taxa de Evasão (%)

- **Leitura dos Dados:** A **Média** nacional é de 2.96%. No entanto, novamente podemos observar uma forte desigualdade nos dados do país. A **Amplitude** indica que a diferença entre a taxa de evasão do melhor Estado/Ano e o com a pior medição é 7.69%.
- **Interpretação do Histograma (Gráfico Laranja):** A distribuição se aproxima de uma "curva normal"(formato de sino), mas levemente deslocada para a esquerda. *O que isso significa:* A situação "padrão" do Brasil é ter uma evasão entre 2% e 4%. Estados com evasão zero ou evasão acima de 6% são raridades estatísticas.
- **Conclusão Teórica:** O sistema público consegue, na média, reter a maioria dos alunos, mas a existência de casos extremos (máximo de 7,85%) indica falhas graves localizadas de política pública.

3. Homicídios (por 100k habitantes)

- **Leitura dos Dados:** A **Média** é de 26.99. Para contexto, a OMS (Organização Mundial da Saúde) considera taxa acima de 10 como epidemia de violência. O **Mínimo** é 6.1 e o **Máximo** beira 60.
- **Interpretação do Histograma (Gráfico Vermelho):** A distribuição é mais "espalhada"(achatada) que as outras. *O que isso significa:* Não há um padrão tão definido quanto na educação. Temos muitos estados seguros (barra esquerda), muitos violentos (meio) e alguns extremamente violentos (direita). A variabilidade da violência é geográfica e heterogênea.

4.2 Fundamentos da análise

Para que a análise seja robusta, é fundamental entender como os dados foram tratados e o que as linhas representam:

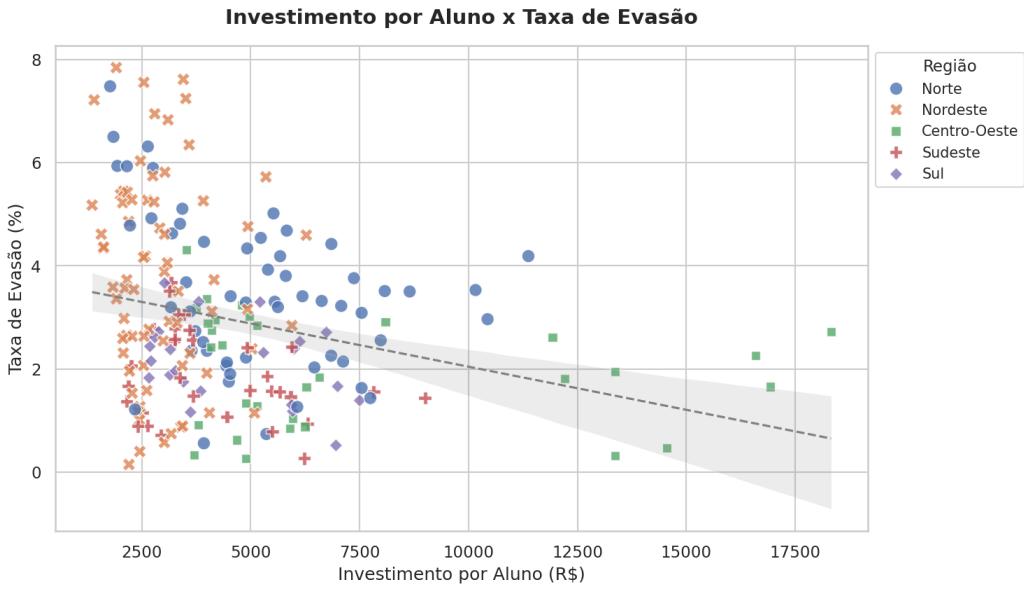
- **Normalização dos Dados:** Todos os gráficos utilizam taxas (por aluno ou por 100 mil habitantes). Isso elimina a distorção populacional

(ex: comparar números absolutos de São Paulo com o Acre seria injusto, tendo em vista a quantidade de habitantes extremamente discrepante).

- **Régressão Linear (A Linha Tracejada):** Nos gráficos de dispersão (bolinhas), a linha tracejada representa o "ajuste linear". Ela é o resultado de um cálculo matemático que busca minimizar a distância entre todos os pontos e uma reta única, buscando representar, por uma linha, a tendência de forma agregada com que os dados se comportam generalizadamente.
 - Se a linha sobe: A correlação é positiva (quando X aumenta, Y aumenta).
 - Se a linha desce: A correlação é negativa (quando X aumenta, Y diminui).

4.2.1 Investimento Estatal x Evasão Escolar

Hipótese Teórica: A falta de empenho estatal (baixo investimento) precariza o processo educacional, resultando em infraestruturas ruins, salários baixos para professores do sistema público e problemas para a qualidade do ensino no geral. Com isso, os alunos são desestimulados ou sequer conseguem condições viáveis de estudo, provocando a evasão.



4.2.2 Evasão Escolar x Taxa de Homicídios

Hipótese Teórica: O jovem fora da escola perde a instrução profissionalizante e a socialização cívica. Sem oportunidades no mercado formal, torna-se vulnerável ao aliciamento pelo crime organizado ou subemprego, aumentando as taxas de violência (vítima ou autor).

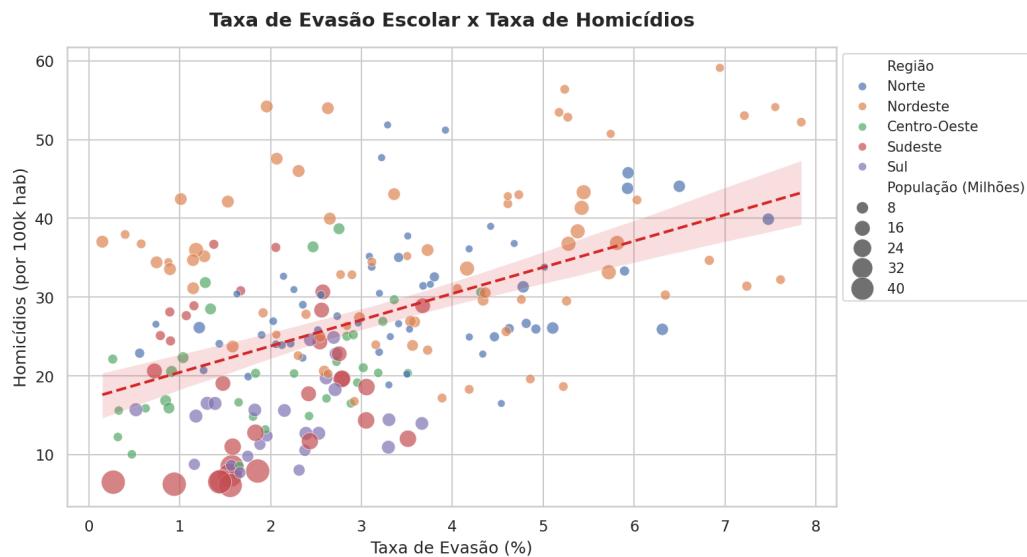


Figura 6: Gráfico representando a relação entre evasão escolar e violência social

Interpretação Visual:

- O tamanho das bolhas representa a população. Note que grandes estados (bolhas grandes cinzas) estão espalhados, mostrando que o problema afeta tanto estados populosos quanto pequenos.
- A linha vermelha é ascendente.

Conclusão: A hipótese de transitividade é fortalecida. Evasão Escolar causa Violência? Os dados sugerem que sim, ou ao menos que caminham juntas. Estados com evasão próxima a 6-7% apresentam taxas de homicídios drasticamente maiores (40-50 por 100 mil habitantes) do que estados com evasão controlada a 1-2%. Isso valida a tese de que a retenção escolar é, em última

instância, uma política de segurança pública preventiva. Com isso, a transitividade Baixo Investimento -> Alta evasão -> Alta violência torna-se um ponto de análise crucial para a gestão estatal.

4.2.3 Evolução Histórica

Hipótese Teórica: Variações nas políticas educacionais ao longo do tempo impactam a violência com algum atraso ou simultaneidade.

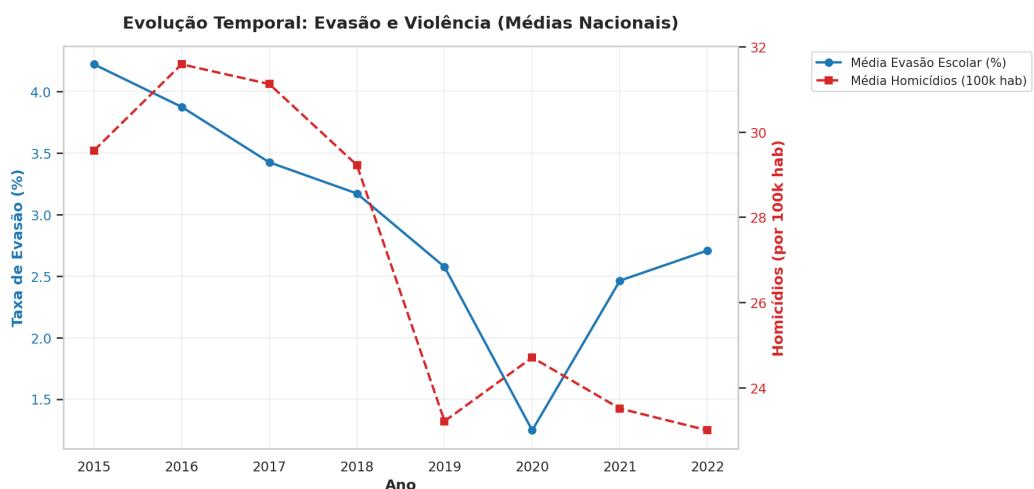


Figura 7: Gráfico que mostra o impacto das medidas governamentais na violência no decorrer do tempo

Interpretação Visual:

- **2015-2019:** Vemos uma queda consistente na taxa de evasão (linha azul) acompanhada, com certa volatilidade, por uma tendência de queda na violência (linha vermelha) a partir de 2017.
- **O Fenômeno 2020 (Pandemia):** Há uma queda brusca na Evasão Escolar em 2020. **Cuidado na análise:** Isso provavelmente não reflete uma melhoria real, mas sim as políticas de "aprovação automática" ou a dificuldade de medir a frequência durante o ensino remoto na pandemia de COVID-19.
- **2021-2022:** A evasão volta a subir (o "rebote" pós-pandemia) e a violência parece estabilizar ou cair levemente, descolando-se um pouco

da curva de evasão nesse curto período, possivelmente devido a outros fatores macroeconômicos ou de segurança pública específicos desse período atípico, ou em razão de problemas recentes de medições dos dados.

Conclusão: A longo prazo, as curvas tendem a "conversar". O esforço nacional para reduzir a evasão (de 4,2% em 2015 para 2,7% em 2022) coincide com um cenário nacional menos violento em 2022 do que era em 2016/2017.

4.2.4 Correlação Geral

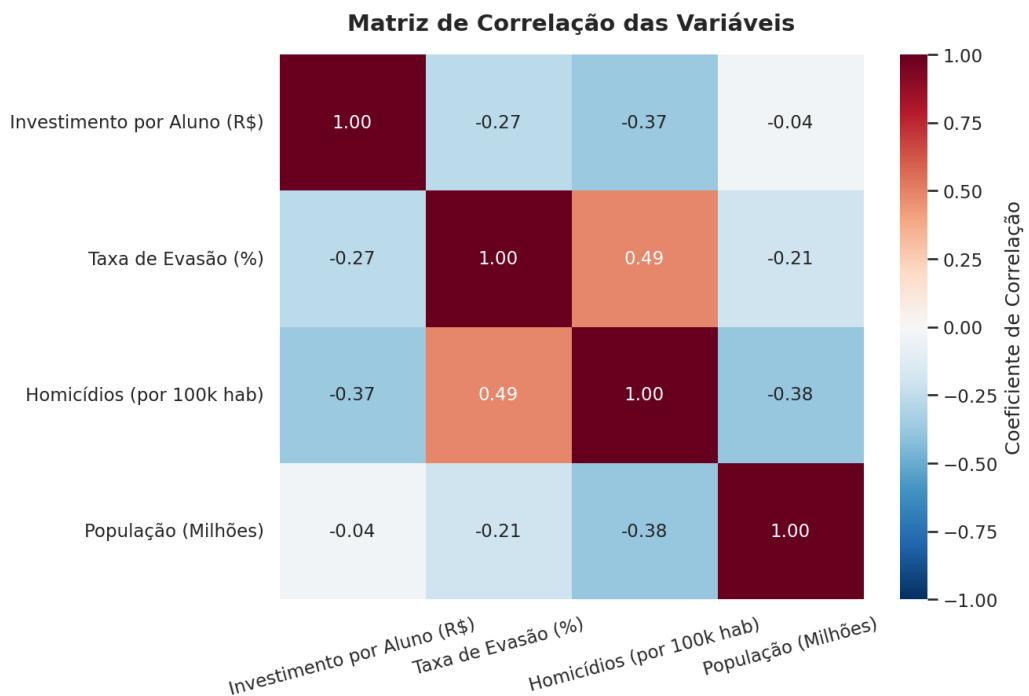


Figura 8: Gráfico representando a correlação geral

Foi calculado o **Coeficiente de Correlação de Pearson** (representado pela letra r). Ele funciona como um "termômetro" que mede a força e a direção da relação linear entre duas variáveis. O coeficiente r responde a duas perguntas fundamentais sobre os dados:

- As variáveis mudam juntas? (Direção)

- **Quão forte é essa conexão?** (Intensidade)

O valor de Pearson **sempre** estará entre -1 e $+1$:

- **-1.0:** Correlação negativa perfeita;
- **< 0 :** Quando uma variável sobe, a outra *tende* a cair. Elas são inversas.
- **$+1.0:$** Correlação positiva perfeita. Quando uma variável aumenta, a outra *sempre* aumenta na mesma proporção.
- **> 0 :** Quando uma variável sobe, a outra *tende* a subir.
- **0.0:** Nenhuma relação linear: Não há padrão. O aumento de uma variável não diz nada sobre a outra.

A matriz representa justamente a correlação entre cada parâmetro. As cores quentes (vermelho) indicam que as variáveis crescem juntas; as frias (azul), que são opostas.

Análise Crítica:

- **Evasão x Homicídios (+0.49):** Esta é a correlação mais forte e preocupante do estudo. O valor positivo de quase 0.5 indica uma conexão substancial: onde o sistema escolar falha em reter o aluno, a violência letal tende a ser maior. Isso corrobora a hipótese de que a escola atua como um fator de proteção social.
- **Investimento x Homicídios (-0.37):** Uma correlação negativa moderada. Estados que investem mais por aluno tendem a ser menos violentos. Isso sugere que o investimento estatal não melhora apenas a educação, mas gera externalidades positivas na segurança pública.
- **População x Indicadores:** Observe que a correlação da população com Evasão (-0.21) e Homicídios (-0.38) é negativa. Isso desafia o senso comum de que "cidades grandes são mais perigosas". Proporcionalmente, estados menores ou menos densos podem sofrer mais com a ineficiência estatal do que os grandes centros que possuem mais recursos e infraestrutura.

Logo, os dados mostram que a evasão escolar e a violência estão estatisticamente associadas. Embora o cálculo de Pearson não prove causalidade direta isoladamente, a forte correlação sustenta a hipótese teórica de que a exclusão escolar contribui para a vulnerabilidade social e criminalidade.

5 Conclusão

Os dados do banco permitem concluir que existe um circuito de causa e efeito infeliz: o baixo investimento estatal acarreta no aumento da probabilidade de evasão escolar, que, por sua vez, apresenta forte correlação positiva com a taxa de homicídios e violências em geral.

Portanto, a transitividade é válida: **A falta de investimento na educação pública é um vetor estatisticamente relevante para o aumento da violência no estado.** A análise gráfica prova, por fim, que a educação não deve ser vista apenas como instrumento pedagógico, mas sim como uma ferramenta importante para o combate à criminalidade e como instrumento garantidor da estabilidade da sociedade contemporânea. A educação permite não somente a reinserção de indivíduos já vulnerabilizados, mas também impede com que esse contingente populacional cresça. Por isso, aponta-se como fundamental para a ação pública o aumento dos investimentos nesse setor em ambientes vulnerabilizados.

Referências

- [1] INSTITUTO BRASILEIRO DE GEOGRAFIA E ESTATÍSTICA. *Censo Demográfico 2022*. Disponível em: <https://www.ibge.gov.br/estatisticas/sociais/trabalho/22827-censo-demografico-2022.html?=&t=resultados>. Acesso em: 29 out. 2025.
- [2] INSTITUTO NACIONAL DE ESTUDOS E PESQUISAS EDUCACIONAIS ANÍSIO TEIXEIRA (INEP). **Painéis estatísticos censo escolar inepdata**. Brasília: Inep, 2025. Disponível em: <https://l1nq.com/YC0Cl>. Acesso em: 29 out. 2025.

- [3] INSTITUTO NACIONAL DE ESTUDOS E PESQUISAS EDUCACIONAIS ANÍSIO TEIXEIRA (INEP). **Taxas de rendimento escolar.** Brasília: Inep, 2020. Disponível em:<https://www.gov.br/inep/pt-br/acesso-a-informacao/dados-abertos/indicadores-educacionais/taxas-de-rendimento-escolar>. Acesso em: 29 out. 2025.
- [4] SISTEMA NACIONAL DE INFORMAÇÕES DE SEGURANÇA PÚBLICA (SINESP). **Ocorrências criminais.** Brasília: Sinesp, 2025. Disponível em: <http://dados.mj.gov.br/dataset/sistema-nacional-de-estatisticas-de-seguranca-publica>. Acesso em: 29 out. 2025.
- [5] BRASIL. Ministério da Fazenda. Secretaria do Tesouro Nacional. Sistema de Informações Contábeis e Fiscais do Setor Público Brasileiro – Siconfi. **Relatórios Resumidos da Execução Orçamentária (RREO).** Disponível em: https://siconfi.tesouro.gov.br/siconfi/pages/public/consulta_finbra_rreo/finbra_rreo_list.jsf. Acesso em: 29 out. 2025.