

Educação e desigualdade socioeconômica

Análise feita por Rafael da Silva Fernandes

Introdução ao tema

O monitoramento da qualidade da educação básica, por meio de indicadores, é uma atividade essencial para a orientação e a avaliação das políticas públicas educacionais e das formas de gestão dos sistemas de ensino do país. A introdução do Índice de Desenvolvimento da Educação Básica (Ideb), índice que sintetiza indicadores de fluxo (taxa média de aprovação na etapa de ensino, calculada a partir dos dados do Censo Escolar da Educação Básica) e de desempenho (proficiência média padronizada dos alunos, pelo Sistema de Avaliação da Educação Básica), possibilitou, de forma objetiva, que governantes, gestores e sociedade civil acompanhassem o desenvolvimento da educação básica, a partir da mensuração dessas duas dimensões, tanto das escolas quanto das redes de ensino brasileiras.

Em virtude da relação entre o desempenho escolar e o perfil social, econômico e cultural dos alunos, evidenciada em vários países e ao longo de períodos distintos, o Inse será usado com o objetivo de contextualizar os resultados obtidos pelos estabelecimentos de ensino, nas diferentes avaliações e exames realizados pelo Inep (Instituto Nacional de Estudos e Pesquisas Educacionais Anísio Teixeira). Desta feita, será possível conhecer, de forma matizada, as escolas que enfrentam maiores desafios e as que foram bem sucedidas na promoção do ensino e, conseqüentemente, da aprendizagem dos alunos, ao delinear, junto com outros indicadores, em que condições esses processos ocorrem. Por esta razão, tais informações podem subsidiar a formulação e a implementação de um conjunto de políticas e ações governamentais que visam a contribuir com a melhoria do fluxo escolar e da aprendizagem dos alunos, bem como com a diminuição das desigualdades sociais e regionais ainda existentes, ao possibilitar que o poder público apoia, de maneira focalizada, as escolas com mais dificuldades e promova a disseminação de experiências pedagógicas que se revelaram exitosas.

Assim, o nível socioeconômico é considerado um construto latente, que sintetiza de maneira unidimensional informações sobre a escolaridade dos pais e sobre a renda familiar, e o Inse objetiva contextualizar o desempenho das escolas nas avaliações e exames realizados pelo Inep, bem como o seu esforço na realização do trabalho educativo, ao caracterizar, de modo geral, o padrão de vida de seu público, relacionados à respectiva posição na hierarquia social.

O Sistema de Avaliação da Educação Básica (Saeb) é um processo de avaliação somativa em larga escala realizado periodicamente pelo Inep e que permite aos diversos níveis governamentais avaliar a qualidade da educação praticada no país, de modo a oferecer subsídios para a elaboração, o monitoramento e o aprimoramento de políticas baseadas em evidências. O Saeb é composto tanto por testes cognitivos, de língua portuguesa (foco em leitura) e matemática (foco em resolução de problemas), aplicados aos alunos, quanto por

questionários contextuais, que coletam informações sobre escolas, diretores, professores e sobre os próprios alunos. Trata-se, em síntese, de um conjunto de testes e questionários aplicados tanto de modo amostral como censitariamente. Em 2015, os testes censitários foram aplicados aos estudantes do 5º e 9º ano do ensino fundamental regular, das escolas públicas, urbanas e rurais, que tinham 20 ou mais alunos matriculados na série avaliada.

A outra base é fornecida pelo Enem, que, a partir da sua reformulação em 2009, passou a aplicar um conjunto de quatro provas objetivas, referentes às áreas de Linguagens e Códigos, Matemática, Ciências Humanas e Ciências da Natureza, e uma Redação. A inscrição no exame foi voluntária e pôde ser feita por quem estava concluindo ou já concluiu o ensino médio, bem como por aqueles que desejavam obter a certificação para sua conclusão. Em 2015, a participação dos candidatos considerados concluintes do ensino médio no exame, mesmo sendo voluntária, foi de pouco mais de 975 mil - estudantes. O questionário contextual do Enem, que fornece informações sobre o aluno e sua família, faz parte de uma das fases do processo de inscrição, que requer o seu preenchimento. Os objetivos desse exame, por sua vez, são os seguintes: ser utilizado como processo seletivo para o acesso ao ensino superior, servir de parâmetro para reformulações dos currículos do ensino médio, possibilitar a certificação para conclusão desta etapa de ensino e servir de referência para autoavaliação do aluno.

A partir das respostas dos estudantes aos questionários contextuais do Saeb e do Enem, o Inse foi construído. As questões utilizadas dizem respeito à renda familiar, à posse de bens e contratação de serviços de empregados domésticos pela família dos estudantes e ao nível de escolaridade de seus pais ou responsáveis. O universo de referência do Inse, por sua vez, inclui somente os dados de estudantes concluintes regulares que responderam a mais de cinco questões.

Nesse texto, buscaremos entender de que maneira a utilização de dados abertos pode colaborar para a redução de desigualdades no Brasil, tendo como base o Inse. A partir dele, traçamos alguns questionamentos iniciais, que buscaremos responder por meio de dados abertos:

- Escolas que encontram-se na capital do Estado possuem maior Inse do que as localizadas no interior?
- Escolas que encontram-se na zona urbana do Estado possuem maior Inse do que as na zona rural?
- De que forma as diferentes dependências administrativas das escolas influenciam no cálculo do Inse?
- Quais Estados e Regiões são os mais afetados?
- Qual a relação entre o Inse de cada Estado e Região do país e sua atividade econômica?

Para responder a essas perguntas, foram utilizadas as bases de dados abaixo.

Bases de dados

● Inse 2015

Fonte:

<https://www.gov.br/inep/pt-br/aceso-a-informacao/dados-abertos/indicadores-educacionais/nivel-socioeconomico>

As fontes para a construção desse indicador foram os dados dos questionários contextuais dos estudantes, fornecidos pelos Microdados disponibilizados pelo Inep, do Sistema de Avaliação da Educação Básica (Saeb) e do Exame Nacional do Ensino Médio (Enem), referentes ao ano de 2015. Optou-se pelos Microdados tanto por fornecerem informações com o menor nível de agregação, cuja unidade mínima é o indivíduo, quanto porque essas bases incluem um amplo espectro de escolas públicas e privadas.

O universo de referência do Inse inclui somente os dados dos estudantes dessas bases que responderam, ao preencher o questionário contextual, cinco ou mais questões, referentes a:

- posse de bens no domicílio: televisão em cores, tv por assinatura, telefone fixo, telefone celular, acesso a internet, aspirador de pó, rádio, videocassete ou DVD, geladeira, freezer (aparelho independente ou parte da geladeira duplex), máquina de lavar roupa, carro, computador, quantidade de banheiros e quartos para dormir
- contratação de serviços: contratação de serviços de mensalista ou diarista
- renda: renda familiar mensal, em salários mínimos
- escolaridade: escolaridade do pai e escolaridade da mãe

Utilizou-se um modelo de resposta gradual, da Teoria de Resposta ao Item, que permite o uso de informações ordinais e possibilita a estimação do indicador com alguns dados faltantes.

A medida de nível socioeconômico do aluno é expressa em uma escala contínua, com média igual a 50 e desvio padrão igual a 10. Não obstante, como a Teoria de Resposta ao Item é um modelo probabilístico, é preciso ressaltar que a resposta do aluno a uma questão se relaciona com a escala de maneira probabilística, de modo que existe uma possibilidade maior dele assinalar uma dada alternativa, em função do seu nível socioeconômico.

Dicionário de dados

- **CO_ESCOLA:** Código da Escola no CENSO Escolar
- **NOME_ESCOLA:** Nome da Escola no CENSO Escolar
- **CO_UF:** Código da Unidade da Federação
- **NOME_UF:** Nome da Unidade da Federação
- **CO_MUNICIPIO:** Código do Município
- **NOME_MUNICIPIO:** Nome do Município
- **ID_AREA:** Área da Escola (relacionado ao Município):
 - 0 - Interior
 - 1 - Capital
- **TP_DEPENDENCIA:** Dependência Administrativa da Escola:
 - 1 - Federal
 - 2 - Estadual
 - 3 - Municipal
 - 4 - Privada
- **TP_LOCALIZACAO:** Localização da Escola:
 - 1 - Urbana
 - 2 - Rural
- **QTD_ALUNOS_INSE:** Quantidade de Alunos com INSE calculado utilizado para o cálculo das médias por escola
- **INSE_VALOR_ABSOLUTO:** Média do Indicador de Nível Socioeconômico dos alunos da escola
- **INSE_CLASSIFICACAO:** Classificação do Indicador de Nível Socioeconômico em 6 Grupos:
 - Grupo 1 (muito baixo)
 - Grupo 2 (baixo)
 - Grupo 3 (médio baixo)
 - Grupo 4 (médio alto)
 - Grupo 5 (alto)
 - Grupo 6 (muito alto)

Para melhor descrever o nível socioeconômico das escolas, foram criados, a partir da análise de cluster por método hierárquico, seis grupos, de forma que o Grupo 1 congrega as escolas com Inse médio mais baixo e o Grupo 6, com mais alto.

Visualizando a base de dados

	CO_ESCOLA	NOME_ESCOLA	CO_UF	NOME_UF	CO_MUNICIPIO	NOME_MUNICIPIO	ID_AREA	TP_DEPENDENCIA
0	11000058	CENTRO DE ENSINO CLASSE A	11	Rondônia	1100205	1100205	1	4
1	11000171	CENTRO EDUCACIONAL MOJUCA	11	Rondônia	1100205	1100205	1	4
2	11000198	CENTRO EDUCACIONAL OBJETIVO	11	Rondônia	1100205	1100205	1	4
3	11000201	EEEF PROF HERBERT DE ALENCAR	11	Rondônia	1100205	1100205	1	2
4	11000244	COLEGIO DOM BOSCO	11	Rondônia	1100205	1100205	1	4

TP_LOCALIZACAO	QTD_ALUNOS_INSE	INSE_VALOR_ABSOLUTO	INSE_CLASSIFICACAO
1	130	64,17	Grupo 6
1	17	54,96	Grupo 4
1	37	61,17	Grupo 5
1	86	53,46	Grupo 4
1	49	56,88	Grupo 5

Tratamento da base de dados

A primeira letra de cada palavra do nome da Unidade da Federação foi colocada em maiúsculo para ficar de acordo com as demais bases de dados a serem trabalhadas.

Os valores da coluna "INSE_VALOR_ABSOLUTO" estavam originalmente com a notação brasileira para números reais, e por isso a vírgula foi substituída por um ponto e então foram transformados em float, para que fosse possível realizar cálculos.

As colunas "CO_ESCOLA", "NOME_ESCOLA", "CO_MUNICIPIO" e "NOME_MUNICIPIO" não são relevantes para a análise pois contém apenas os códigos de identificação e nomes das escolas e municípios em que o Inse foi calculado.

A coluna "CO_UF" não é importante aqui pois trata-se apenas do código de identificação de cada Unidade da Federação, e trabalharemos em cima das siglas e nomes das mesmas.

Por fim, a coluna "QTD_ALUNOS_INSE" aborda a quantidade de alunos que responderam o formulário nas suas respectivas escolas, e não é o intuito dessa análise discutir quais regiões possuem mais ou menos alunos.

Por essas razões, as colunas abordadas acima foram removidas do DataFrame a ser estudado.

	NOME_UF	ID_AREA	TP_DEPENDENCIA	TP_LOCALIZACAO	INSE_VALOR_ABSOLUTO	INSE_CLASSIFICACAO
0	Rondônia	1	4	1	64.17	Grupo 6
1	Rondônia	1	4	1	54.96	Grupo 4
2	Rondônia	1	4	1	61.17	Grupo 5
3	Rondônia	1	2	1	53.46	Grupo 4
4	Rondônia	1	4	1	56.88	Grupo 5

● PIB 2015

Fonte:

<https://agenciadenoticias.ibge.gov.br/agencia-sala-de-imprensa/2013-agencia-de-noticias/releases/17999-contas-regionais-2015-queda-no-pib-atinge-todas-as-unidades-da-federacao-pela-primeira-vez-na-serie>

O PIB (Produto Interno Bruto) é a soma de todos os bens e serviços finais produzidos por um país, estado ou cidade, geralmente em um ano. Todos os países calculam o seu PIB nas suas respectivas moedas. Ele mede apenas os bens e serviços finais para evitar dupla contagem. Se um país produz 100 reais de trigo, 200 reais de farinha de trigo e 300 reais de pão, por exemplo, seu PIB será de 300 reais, pois os valores da farinha e do trigo já estão embutidos no valor do pão.

Os bens e serviços finais que compõem o PIB são medidos no preço em que chegam ao consumidor. Dessa forma, levam em consideração também os impostos sobre os produtos comercializados. Dessa forma, o PIB não é o total da riqueza existente em um país. Esse é um equívoco muito comum, pois dá a sensação de que o PIB seria um estoque de valor que existe na economia, como uma espécie de tesouro nacional. Na realidade, o PIB é um indicador de fluxo de novos bens e serviços finais produzidos durante um período. Se um país não produzir nada em um ano, o seu PIB será nulo.

Dicionário de dados

- **Unidade da Federação:** Nome da Unidade da Federação
- **Valor corrente**(R\$ 1 000 000): Valor do PIB em milhões
- **Participação**(%): Porcentagem de participação em relação ao PIB total do Brasil em 2015
- **Posição relativa da variação em volume:** Quanto menor à variação em volume, mais bem colocado está o Estado
- **Variação em volume**(%): Diferença na participação entre os anos de 2002 e 2015

Visualizando a base de dados

	Unidade da Federação	Valor corrente\	(R\$ 1 000 000)	Participação\	(%)	Posição relativa da variação em volume	Variação em volume\	(%)
0	Mato Grosso do Sul	83 082		1,4		1º	-0,3	
1	Roraima	10 354		0,2		2º	-0,3	
2	Tocantins	28 930		0,5		3º	-0,4	
3	Pará	130 883		2,2		4º	-0,9	
4	Distrito Federal	215 613		3,6		5º	-1	

Tratamento da base de dados

A primeira letra de cada palavra do nome da Unidade da Federação foi colocada em maiúsculo para ficar de acordo com as demais bases de dados a serem trabalhadas.

Os valores da coluna "Valor corrente\

(R\$ 1 000 000)" estavam originalmente com o caractere "espaço" como separador para a casa dos milhares, e por isso esse caractere foi removido e então foram transformados em int, para que fosse possível realizar cálculos.

As colunas "Participação\

(%)", "Posição relativa da variação em volume" e "Variação em volume\(%)" foram removidas do DataFrame original pois tratam-se de uma análise de comparação entre os dados de 2002 e 2015, e o intuito desse documento é discutir a desigualdade com base nos dados mais atuais.

	Unidade da Federação	Valor corrente\	(R\$ 1 000 000)
0	Mato Grosso Do Sul	83082	
1	Roraima	10354	
2	Tocantins	28930	
3	Pará	130883	
4	Distrito Federal	215613	

• Relação entre Estados e Regiões do Brasil

Fonte: <https://doutoresdoexcel.com.br/relacao-estados-e-regioes-curso-power-bi/>

Base de dados que contém cada Estado do Brasil, sua sigla, capital e Região à que pertence.

Dicionário de dados

- **SIGLA:** Sigla do Estado
- **ESTADO:** Nome do Estado
- **CAPITAL:** Nome da capital do Estado
- **REGIÃO:** Região à que pertence o Estado

Visualização da base de dados

	SIGLA	ESTADO	CAPITAL	REGIÃO
0	AC	ACRE	RIO BRANCO	NORTE
1	AM	AMAZONAS	MANAUS	NORTE
2	AP	AMAPÁ	MACAPÁ	NORTE
3	PA	PARÁ	BELÉM	NORTE
4	RO	RONDÔNIA	PORTO VELHO	NORTE

Tratamento da base de dados

A primeira letra de cada palavra do nome do Estado e da Região foi colocada em maiúsculo para ficar de acordo com as demais bases de dados a serem trabalhadas.

A coluna "CAPITAL" não será considerada para a análise a ser trabalhada nesse documento, pois contém apenas a capital de cada Estado do Brasil, e o foco aqui é nos Estados de forma geral e suas respectivas Regiões.

	SIGLA	ESTADO	REGIÃO
0	AC	Acre	Norte
1	AM	Amazonas	Norte
2	AP	Amapá	Norte
3	PA	Pará	Norte
4	RO	Rondônia	Norte

● Finbra 2015

Fonte: <https://basedosdados.org/dataset/br-tesouro-finbra>

O Finbra (Finanças do Brasil) é o banco de dados criado pela Secretaria do Tesouro Nacional (STN), em conjunto com a Caixa Econômica Federal, onde apresenta números consolidados de Estados e municípios brasileiros. A coleta é feita por meio de declaração dos próprios municípios. Estes respondem, anualmente, a um formulário denominado Quadro de Dados Contábeis Consolidados (QDCC), com informações extraídas do seu balanço.

Dicionário de dados

- **id_municipio**: ID Município - IBGE 7 Dígitos
- **sigla_uf**: Sigla da Unidade da Federação
- **ano**: Ano
- **coluna**: Coluna
- **conta**: Conta
- **valor**: Valor (R\$)

Visualizando a base de dados

	sigla_uf	valor
0	RO	15054548.34
1	RO	63929927.19
2	RO	4435185.20
3	RO	38340555.45
4	RO	9203085.20

Trabalharemos apenas com os dados referentes ao ano de 2015, cuja conta "Educação" seja considerada uma despesa paga. Cada linha da tabela acima representa um município, mas como o intuito dessa análise é estudar os Estados e suas Regiões, calculamos abaixo quanto cada Unidade da Federação investiu, em milhões de reais.

Tratamento da base de dados

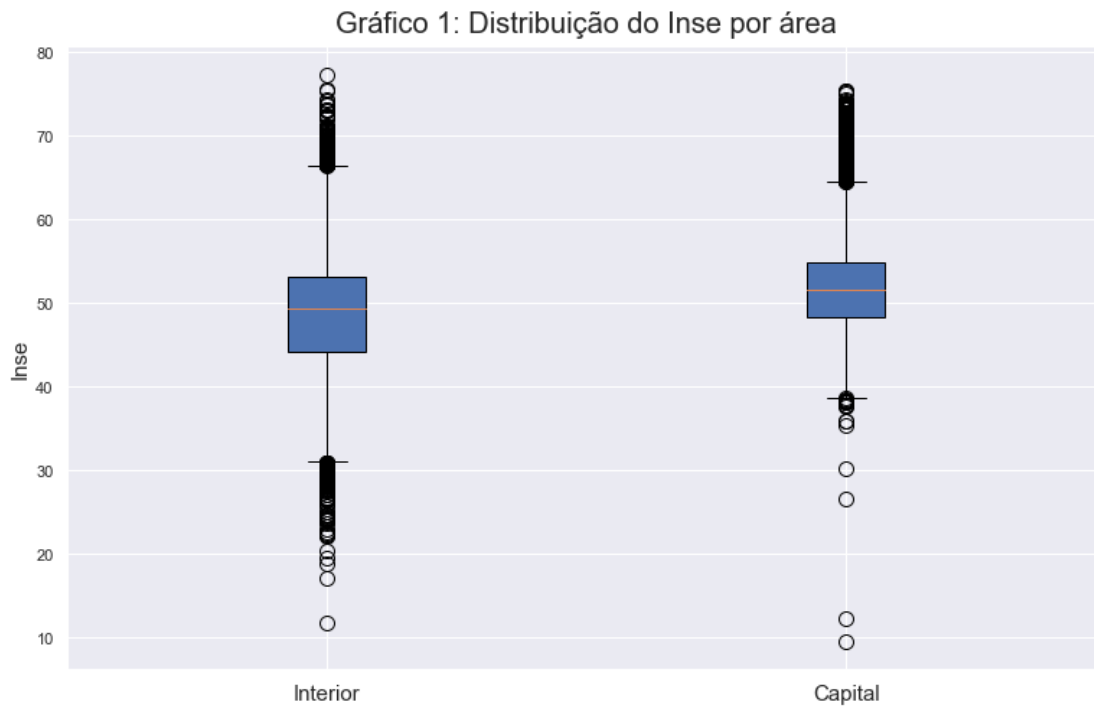
	Estado	Valor (em milhões de reais)
0	Acre	440.460572
1	Alagoas	1908.561488
2	Amazonas	2248.217346
3	Amapá	314.615949
4	Bahia	9582.721678

A partir das bases do Finbra e Relação entre Estados e Regiões do Brasil, iniciamos por construir uma tabela única contendo as informações relativas a cada Estado e o valor (em milhões de reais) total gasto em Educação no ano de 2015.

Análise inicial dos dados - Inse 2015

Para entender melhor como se dá as desigualdades sociais e regionais existentes, iniciamos por obter as distribuições do Inse por cada área, localização administrativa e localização. A seguir, apresentamos três gráficos, relacionados respectivamente às distribuições mencionadas anteriormente.

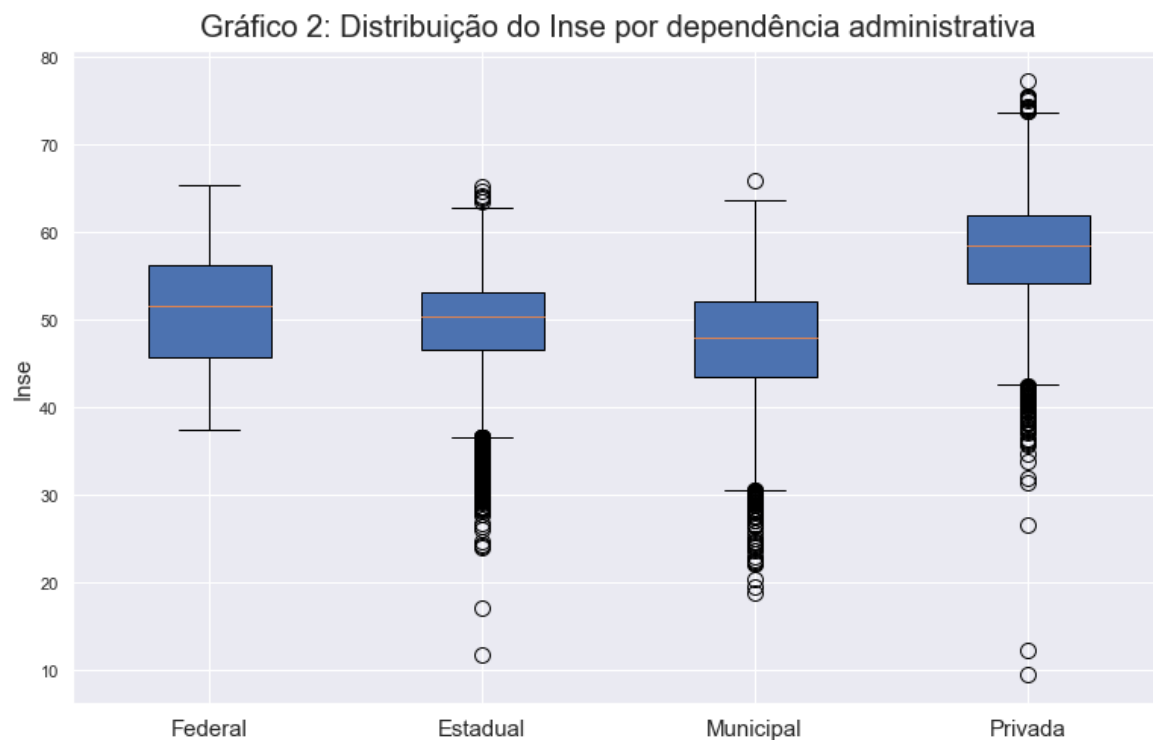
Distribuição do Inse por área



O Gráfico 1 ilustra a distribuição do Inse médio das escolas por área: interior ou capital.

Observa-se que, de modo geral, as escolas na capital possuem um Inse maior do que as escolas no interior, como era de se esperar.

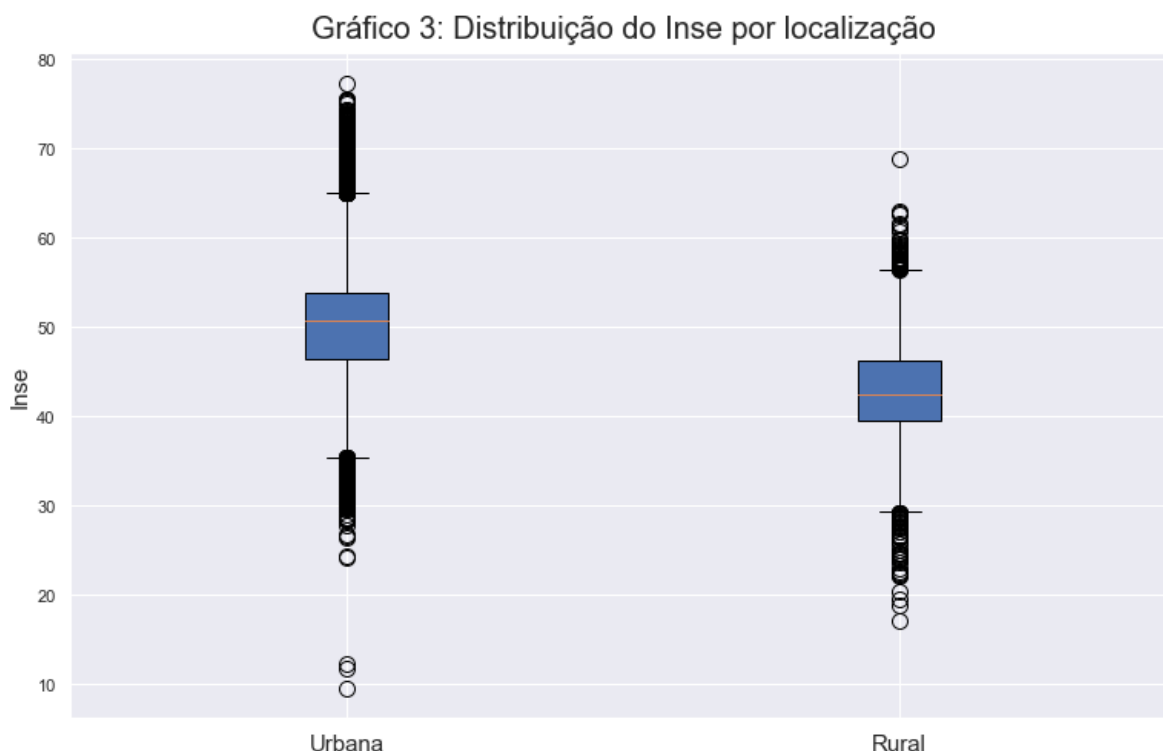
Distribuição do Inse por dependência administrativa



O Gráfico 2 apresenta o Inse das escolas por dependência administrativa.

Observa-se que, em geral, a dependência administrativa com o maior Inse é a das escolas privadas, seguidas pelas escolas federais, estaduais e, por fim, as municipais.

Distribuição do Inse por localização



O Gráfico 3 ilustra a distribuição do Inse médio das escolas por localização: urbana ou rural.

Observa-se que, de modo geral, as escolas urbanas possuem um Inse maior do que as escolas rurais, como era esperado.

Os gráficos apresentados acima são conhecidos como "boxplot" (ou "gráfico de caixa"). O boxplot nos fornece uma análise visual da posição, dispersão, simetria, caudas e valores discrepantes (outliers) do conjunto de dados.

Em relação à posição dos dados, observa-se a linha central do retângulo (a mediana ou segundo quartil).

A dispersão dos dados pode ser representada pelo intervalo interquartílico que é a diferença entre o terceiro quartil e o primeiro quartil (tamanho da caixa), ou ainda pela amplitude que é calculada da seguinte maneira: valor máximo – valor mínimo. Embora a amplitude seja de fácil entendimento, o intervalo interquartil é uma estatística mais robusta para medir variabilidade uma vez que não sofre influência de outliers.

Um conjunto de dados que tem uma distribuição simétrica, terá a linha da mediana no centro do retângulo. Quando a linha da mediana está próxima ao primeiro quartil, os dados são assimétricos positivos e quando a posição da linha da mediana é próxima ao terceiro quartil, os dados são assimétricos negativos.

Já os outliers indicam possíveis valores discrepantes. No boxplot, as observações são consideradas outliers quando estão abaixo ou acima do limite de detecção de outliers.

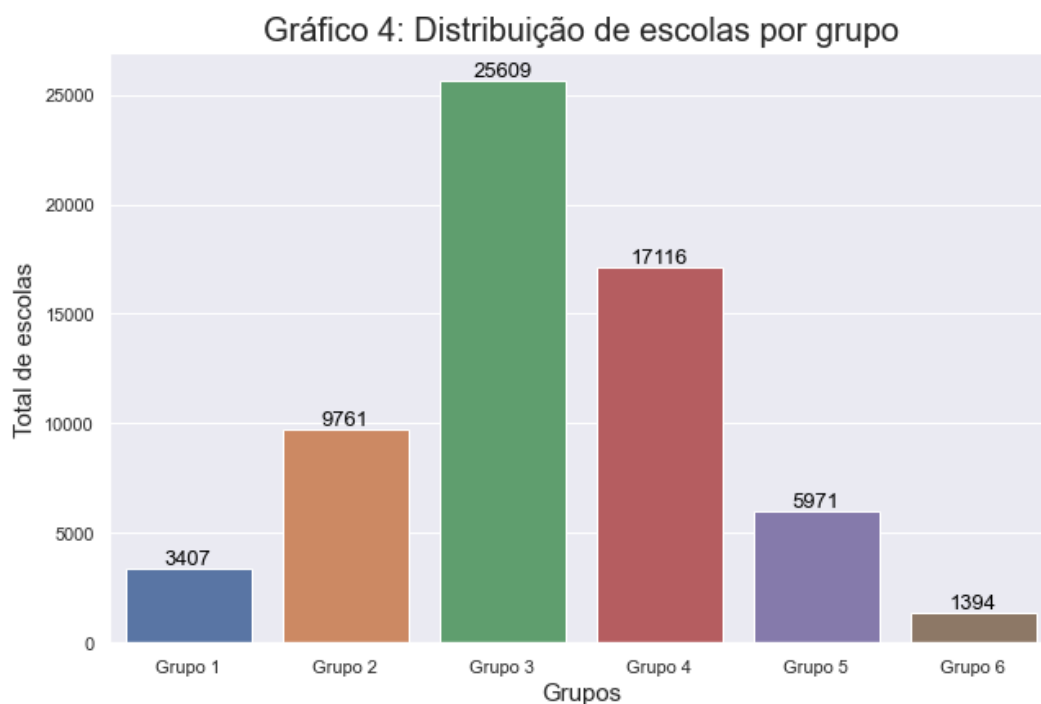
Nos três gráficos apresentados acima, podemos perceber um valor elevado do Inse para escolas que encontram-se na capital, são privadas e estão na zona

urbana do Estado. Além disso, para escolas que encontram-se no interior, são municipais e estão na zona rural do Estado, seus valores de Inse são demasiado baixos.

Dessa forma, enquanto algumas escolas possuem bons resultados, outras várias não. Assim, temos um primeiro indicador para subsidiar a formulação e a implementação de um conjunto de políticas e ações governamentais que visam a contribuir com a melhoria do fluxo escolar e da aprendizagem dos alunos, bem como com a diminuição das desigualdades sociais e regionais ainda existentes, ao possibilitar que o poder público apoie, de maneira focalizada, as escolas com mais dificuldades e promova a disseminação de experiências pedagógicas que se revelaram exitosas.

Vale ainda observar como se dá a distribuição de escolas por grupo, lembrando que há seis grupos, sendo que o Grupo 1 congrega as escolas com Inse médio mais baixo e o Grupo 6, com mais alto.

Distribuição de escolas por grupo



O Gráfico 4 ilustra a quantidade de escolas por grupo.

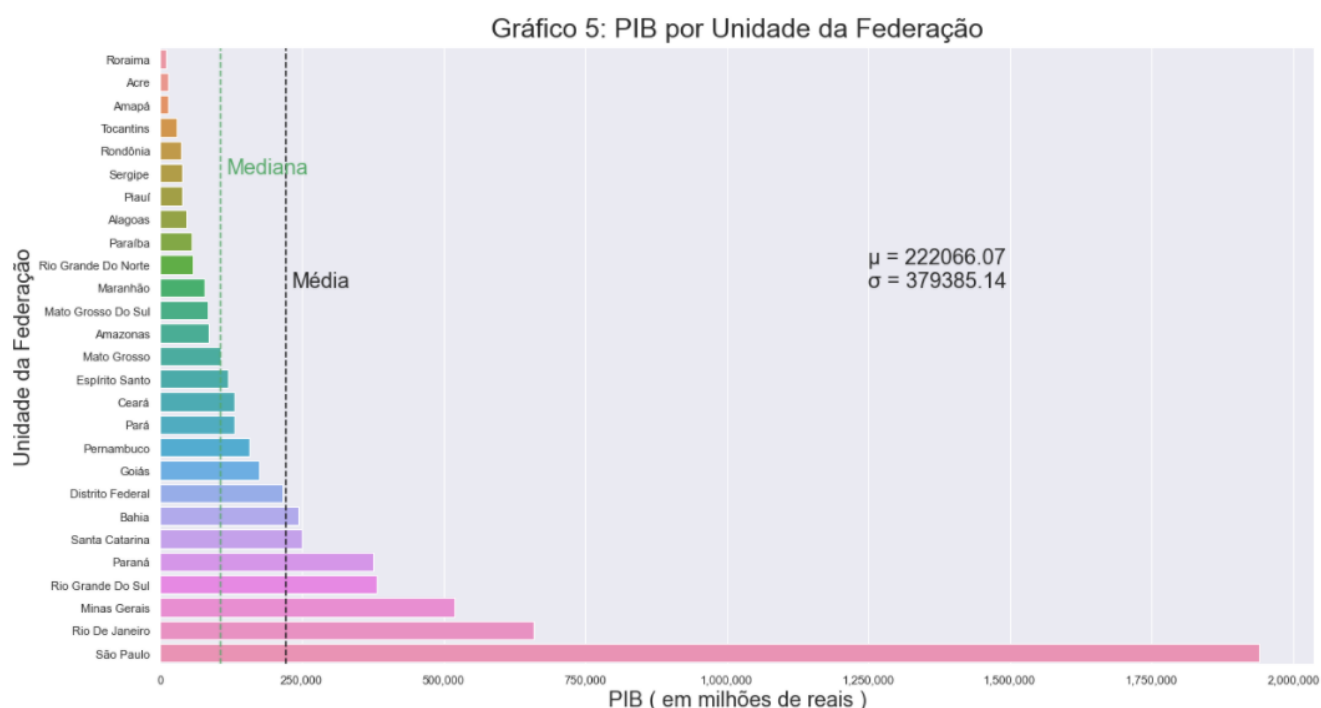
Observa-se que, de modo geral, poucas escolas possuem Inse médio alto, com um total de apenas 1394 no Grupo 6, e que a grande maioria encontra-se no Grupo 3 e abaixo. Temos, portanto, mais um indicador para a necessidade de políticas públicas que sejam focalizadas para as regiões com piores valores de Inse e favoreçam o aumento da igualdade.

Apesar das análises acima serem promissoras, percebe-se a necessidade de obter dados relacionados à economia para o cruzamento de informações. Por essa razão, trabalharemos agora em cima do Produto Interno Bruto (PIB) para estudar as hipóteses já levantadas.

Análise inicial dos dados - PIB 2015

PIB é a sigla para Produto Interno Bruto. Ele é considerado o termômetro da economia: quanto maior o PIB de um país, maior sua atividade econômica – e por sua vez, quanto maior a atividade econômica de um país, mais se consome, vende e investe nele. Apresentamos abaixo gráficos com o valor do PIB por Unidade da Federação e por Região do país no ano de 2015.

PIB por Unidade da Federação



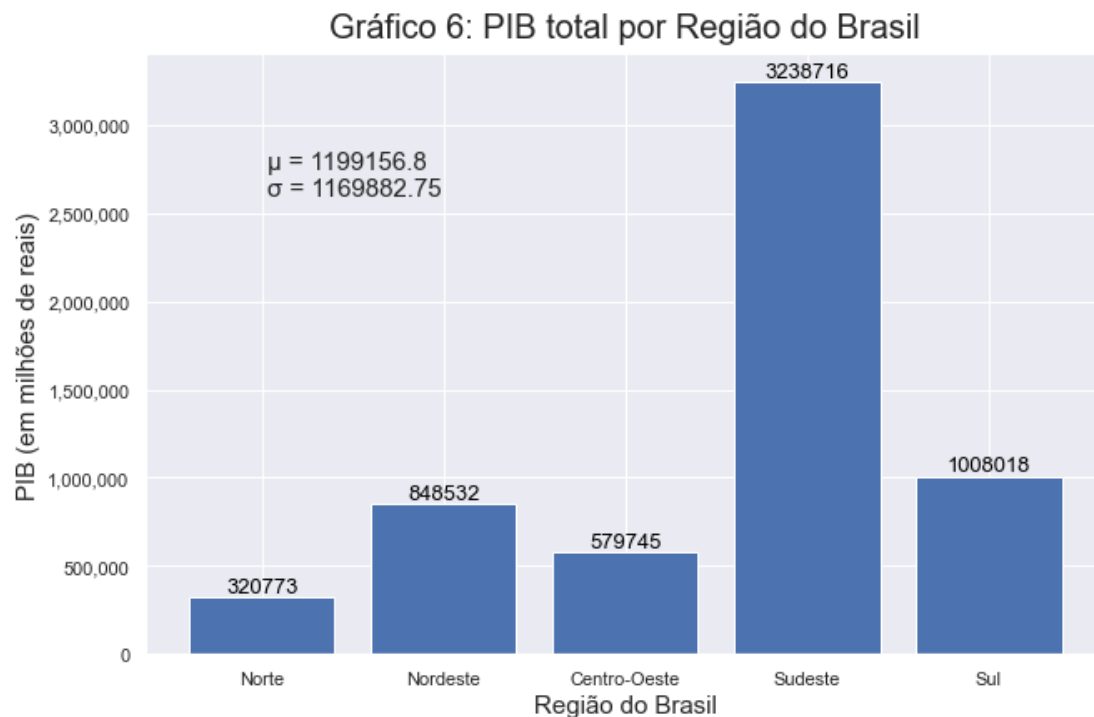
O Gráfico 5 ilustra o PIB total, em milhões de reais, por Unidade da Federação no ano de 2015.

Observa-se que apenas 7 Unidades da Federação encontram-se acima da média. Vale observar também que as três Unidades da Federação com maior PIB são: São Paulo, Rio de Janeiro e Minas Gerais, todas pertencentes à Região Sudeste; e as três com menor PIB são: Roraima, Acre e Amapá, todas pertencentes à Região Norte. Quanto à Região Norte, vale acrescentar também que apenas o Estado do Pará está acima da mediana.

Com um desvio padrão de 379385.14, em milhões de reais, valor superado por apenas por 4 das 27 Unidades da Federação, fica mais do que evidente a desigualdade na distribuição da atividade econômica pelo país.

Mostra-se interessante, dessa forma, compreender como o PIB está distribuído por cada Região do país.

PIB por Regiões do Brasil



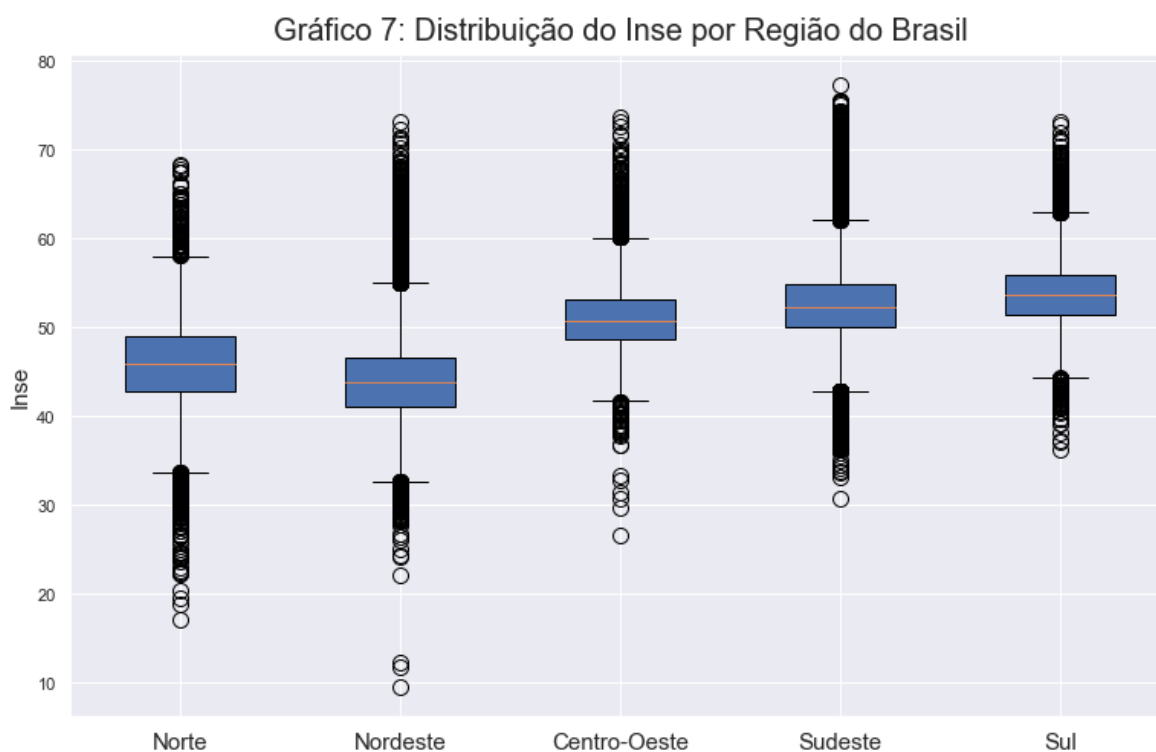
O Gráfico 6 ilustra o PIB total, em milhões de reais, por Região do Brasil no ano de 2015.

Observamos novamente que a Região Sudeste possui o valor mais elevado, e a Região Norte o mais baixo. Temos assim um primeiro indicador, portanto, para a necessidade de políticas públicas que sejam focalizadas para as Regiões com menor atividade econômica.

Avaliando o PIB por Região do país, vamos agora compreender a distribuição do Inse por Região do mesmo para uma melhor análise da desigualdade dos indicadores socioeconômicos entre Estados e Regiões.

Inse por Região do Brasil

Distribuição do Inse por Região do Brasil



O Gráfico 7 ilustra a distribuição do Inse por Região do Brasil.

Observa-se que a região com maior Inse é a Região Sul, e a com menor é a Região Nordeste. Interessante perceber como a distribuição do Inse por Região retirou as Regiões Sudeste e Norte do primeiro e último lugar, respectivamente.

Entretanto, há um outlier na Região Sudeste com o Inse mais elevado de todos, e na Região Nordeste o mais baixo. Podemos observar também que os valores do Inse da região Norte são apenas um pouco maiores do que os da Região Nordeste; e o mesmo pode ser dito entre as Regiões Sudeste e Sul.

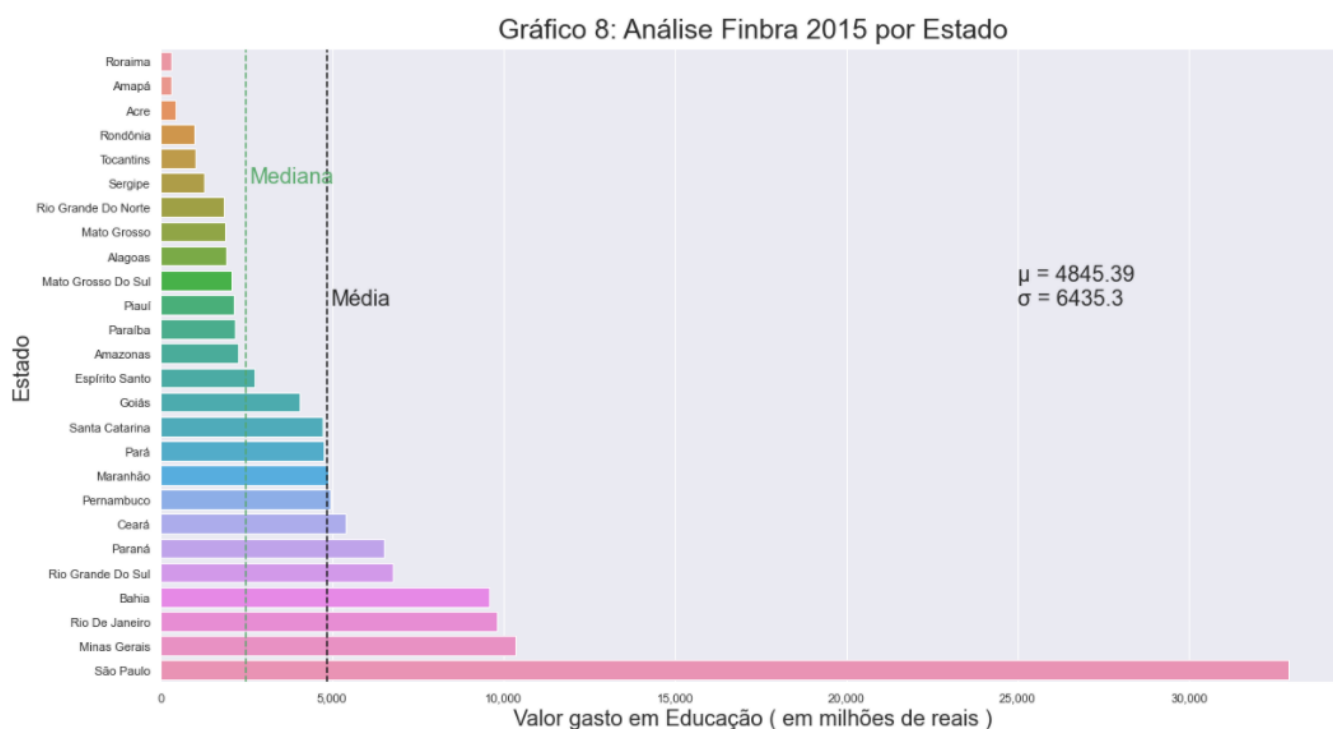
Fica claro, então, que o foco da atividade econômica no país está nas Regiões Sudeste e Sul, enquanto as Regiões Norte e Nordeste - regiões que compreendem um total de 16 das 27 Unidades da Federação - possuem os piores valores de Inse do país, necessitando assim de um conjunto de políticas e ações governamentais para a diminuição das desigualdades sociais e regionais existentes.

Vista as distribuições do PIB e do Inse por Região do Brasil, surge o questionamento: "Quanto cada Estado e Região investiu em educação?".

Análise inicial dos dados - Finbra 2015

Para analisar as despesas, utilizamos a base de dados Finbra. Apresentamos nos gráficos a seguir os gastos relativos à Educação em cada Estado e Região do país no ano de 2015.

Finbra referente à Educação por Estado



O Gráfico 8 ilustra o valor total gasto em educação, em milhões de reais, por Estado do Brasil no ano de 2015.

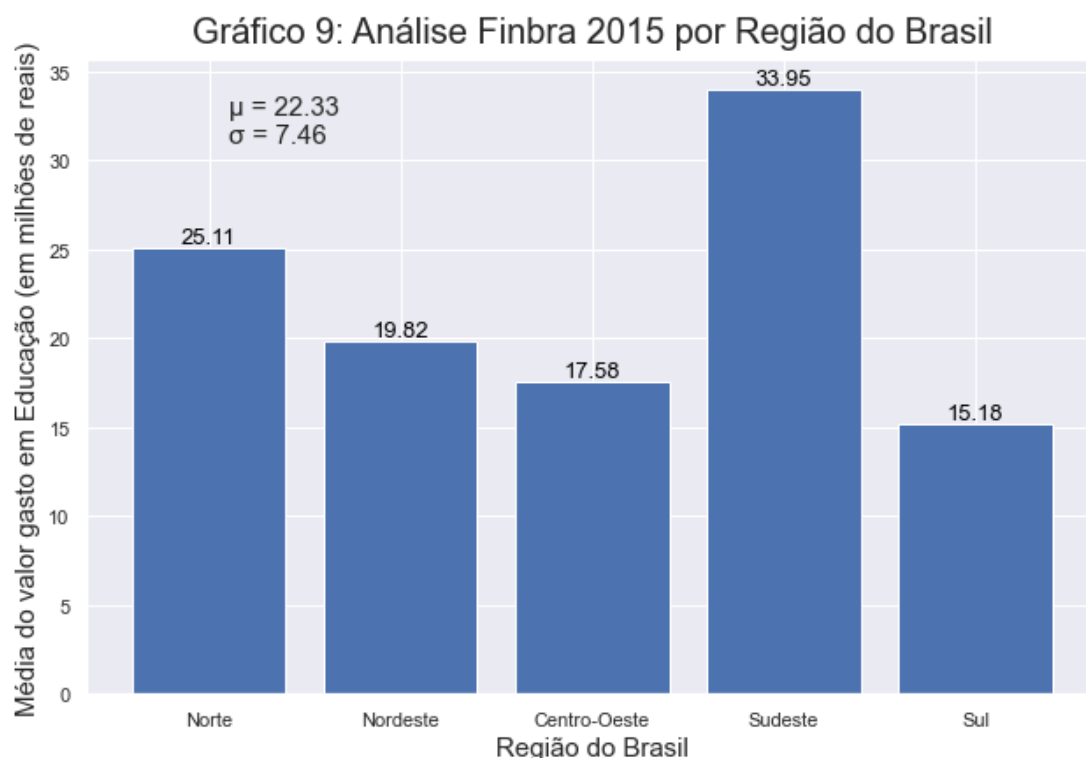
Observa-se que apenas 8 Estados encontram-se acima da média. Vale observar também que os três Estados que mais investiram em educação são: São Paulo, Rio de Janeiro e Minas Gerais, todos pertencentes à Região Sudeste e também os três com maior PIB, como visto no Gráfico 5; e os três Estados que puderem investir menos em Educação são: Roraima, Amapá e Acre, todos pertencem à Região Norte e também os três com menor PIB.

Quanto à Região Norte, vale acrescentar também que nenhum de seus Estados está acima da média, e apenas o Pará está acima da mediana.

Quanto à Região Sul, aquela com maior distribuição de Inse de acordo com o Gráfico 7, e a segunda com maior PIB segundo o Gráfico 6, temos 2 (dois) de seus Estados acima da média: Paraná e Rio Grande do Sul, e pelo gráfico podemos perceber que o Estado de Santa Catarina dista muito pouco da média.

Mais uma vez conseguimos observar a diferença econômica entre os Estados do país, fato claramente constatado pelo desvio padrão de 6435.3, em milhões de reais, valor que mais de 20 Estados não atingiram. Com base nessas informações, é interessante analisar o valor gasto em educação por cada Região do país.

Finbra referente à Educação por Região do país



O Gráfico 9 ilustra a média do valor gasto em educação, em milhões de reais, por Região do Brasil no ano de 2015.

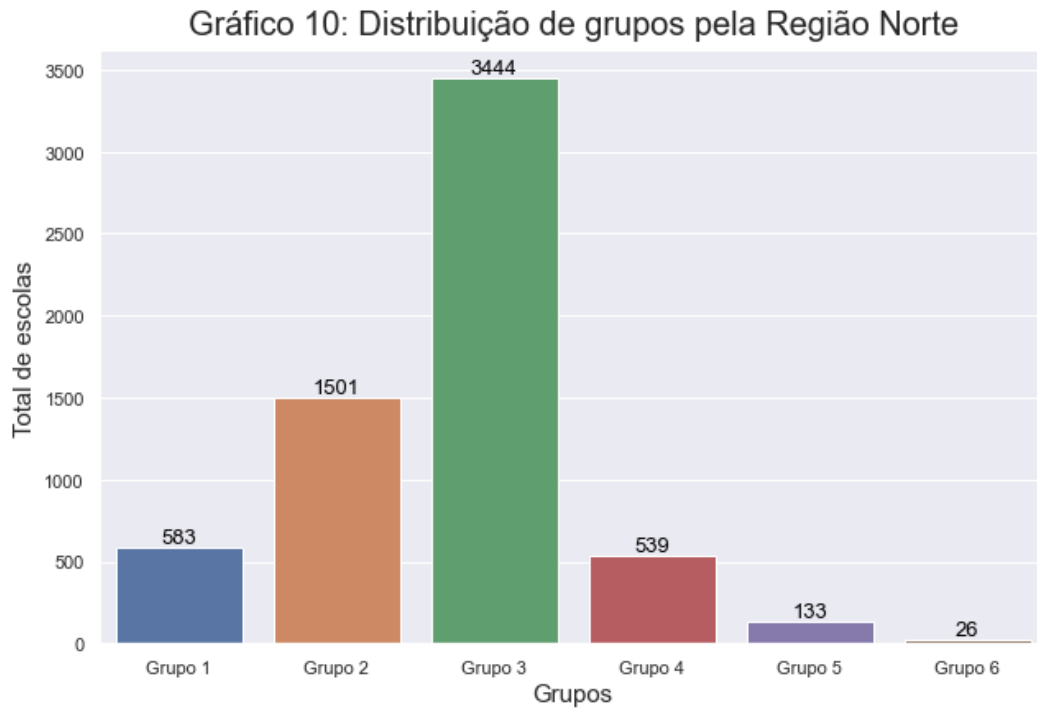
Apesar do gráfico anterior ter mostrado o quanto cada Estado investiu em educação, no gráfico acima percebemos que as regiões Norte e Nordeste foram a segunda e terceira regiões cujos Estados em média depuseram mais dinheiro no assunto, respectivamente. Tal fato pode ter ocorrido por conta de ações governamentais focadas em combater a desigualdade.

Da mesma forma, temos o interessante fato de que os Estados da região Sul são aqueles com a média mais baixa de investimento, apesar de maior distribuição de Inse - como observado no gráfico 7. É possível que os governos dessa região tenham investido há mais tempo nas suas escolas, enquanto as regiões Norte e Nordeste tiveram um começo retardado.

Vimos o quanto cada Estado e Região investiu em educação, mas como todo processo, é algo que ocorre de maneira lenta e duradoura. Por essa razão, queremos analisar como cada Região se encontrava no ano de 2015 e assim surge o questionamento: "Como era a distribuição de grupos por Região do Brasil?".

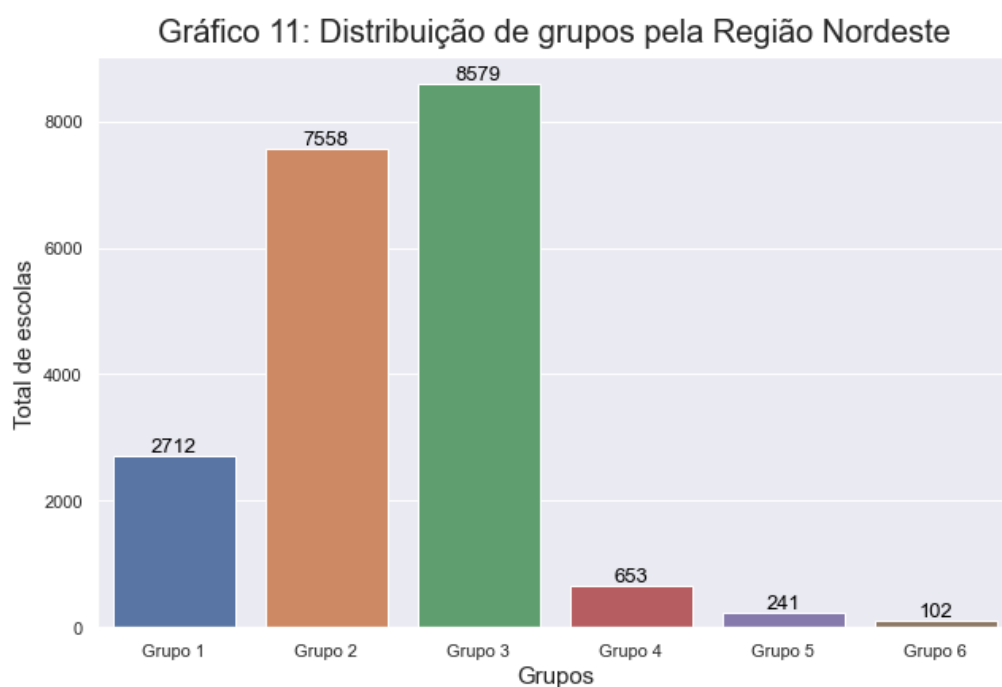
Distribuição de grupos por Região do Brasil

- **Região Norte**



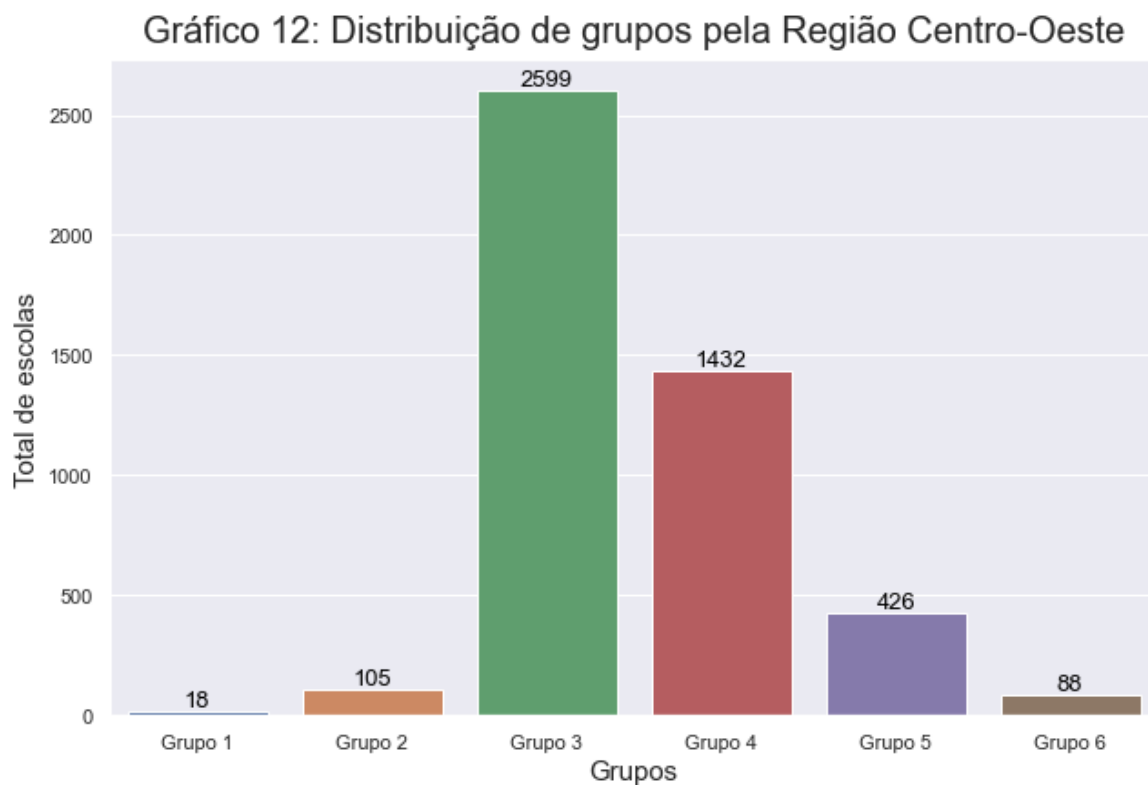
O Gráfico 10 ilustra a distribuição de grupos pela Região Norte do Brasil. Considerando apenas 26 escolas no Grupo 6, enquanto no Grupo 1 temos 583, no Grupo 2, 1501 e no Grupo 3 impressionantes 3444, fica evidente que as escolas da Região Norte enfrentam grandes desafios.

- **Região Nordeste**



O Gráfico 11 ilustra a distribuição de grupos pela Região Nordeste do Brasil. Apenas 102 escolas no Grupo 6, e impressionantes 2712 no Grupo 1, fica evidente que as escolas da Região Norte enfrentam os maiores desafios.

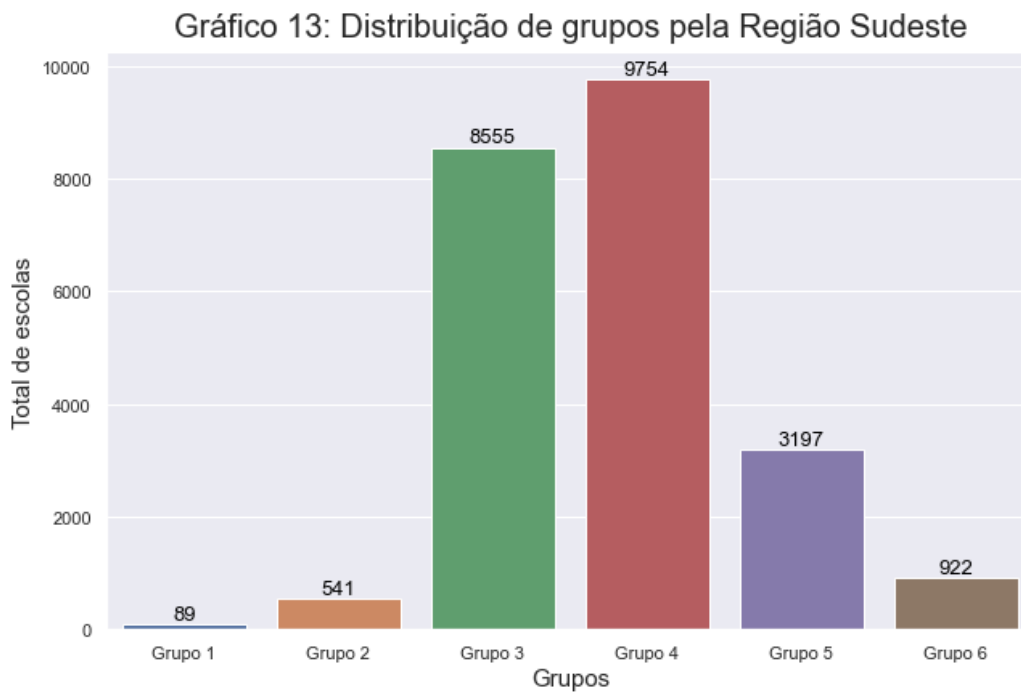
- **Região Centro-Oeste**



O Gráfico 12 ilustra a distribuição de grupos pela Região Centro-Oeste do Brasil.

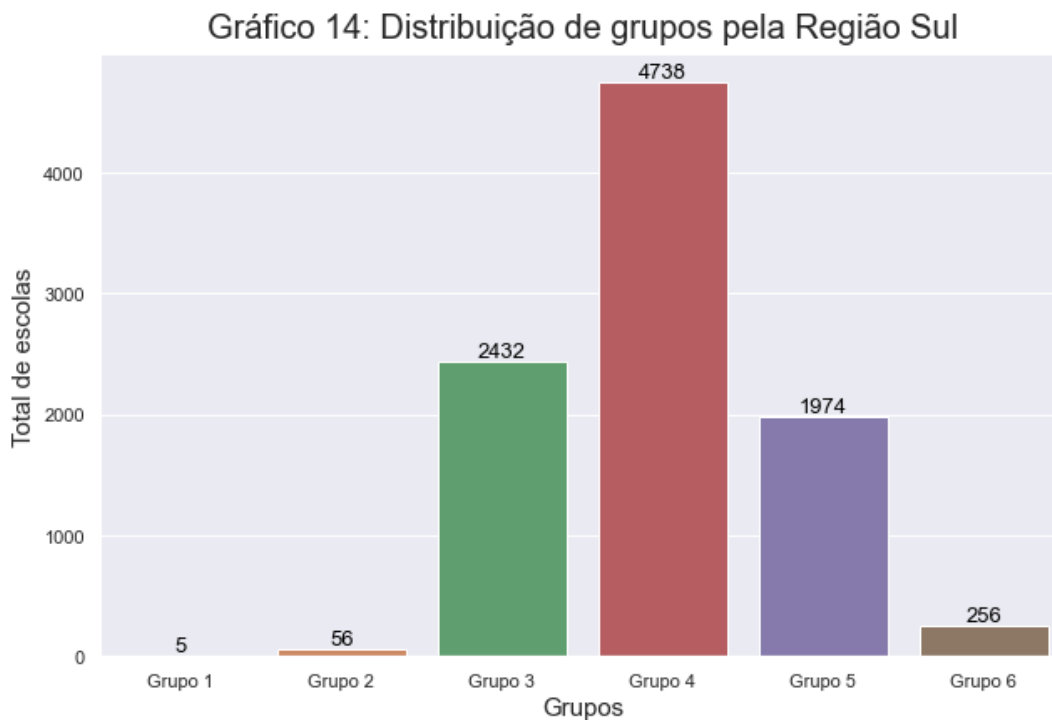
Ao contrário das Regiões apresentadas acima, a Região Centro-Oeste conta com apenas 18 escolas no Grupo 1, e 1432 no Grupo 4, apesar de apresentar um PIB menor do que a Região Nordeste.

- **Região Sudeste**



O Gráfico 13 ilustra a distribuição de grupos pela Região Sul do Brasil. Pela primeira vez, temos a maioria em grupo com Inse médio relativamente alto, com impressionantes 9754 escolas no Grupo 4 e apenas 89 no Grupo 1.

- **Região Sul**



O Gráfico 14 ilustra a distribuição de grupos pela Região Sul do Brasil.

Com apenas 5 escolas no Grupo 1, 2432 no Grupo 3 e esmagadores 4738 no Grupo 4, a Região Sul é a Região com Inse médio mais elevado.

Apesar das análises serem promissoras, contrasta com o fato da Região Centro-Oeste possuir o segundo menor PIB do Brasil (Gráfico 6), segundo menor investimento em educação (Gráfico 9), mas ter a terceira maior distribuição de Inse médio do país (Gráfico 7) e com apenas 18 escolas no Grupo 1 (Gráfico 12).

Dessa forma, mostra-se a necessidade de obter mais dados para poder cruzar informações, e por essa razão passou-se a incluir outra base, o [Índice de Desenvolvimento Humano \(IDH\) com os dados de 2016](#).

IDH 2016

O Índice de Desenvolvimento Humano (IDH) é uma unidade de medida utilizada para aferir o grau de desenvolvimento de uma determinada sociedade nos quesitos de educação, saúde e renda. A utilização de um indicador que envolvesse outras variáveis que não somente a questão econômica ocorreu pela primeira vez em 1990 pelo Programa das Nações Unidas para o Desenvolvimento (PNUD). Esse indicador foi criado pelo paquistanês Mahbub Ul Haq e pelo indiano Amartya Sen.

A utilização das variáveis educação, saúde e renda permite uma comparação com praticamente todos os países do globo e serve de referência para mensurar a resposta de determinado país frente a essas importantes demandas.

O IDH é uma referência numérica que varia entre 0 e 1. Quanto mais próximo de zero, menor é o indicador para os quesitos de saúde, educação e renda. Quanto mais próximo de 1, melhores são as condições para esses quesitos. No mundo, nenhum país possui o IDH zero ou um.

O Índice de Desenvolvimento Humano Municipal (IDHM) é uma medida composta de indicadores de três dimensões do desenvolvimento humano: longevidade, educação e renda. O índice varia de 0 a 1. Quanto mais próximo de 1, maior o desenvolvimento humano.

O IDHM brasileiro segue as mesmas três dimensões do IDH Global - longevidade, educação e renda, mas vai além: adequa a metodologia global ao contexto brasileiro e à disponibilidade de indicadores nacionais. Embora meçam os mesmos fenômenos, os indicadores levados em conta no IDHM são mais adequados para avaliar o desenvolvimento dos municípios brasileiros. Assim, o IDHM - incluindo seus três componentes, IDHM Longevidade, IDHM Educação e IDHM Renda - conta um pouco da história dos municípios em três importantes dimensões do desenvolvimento humano durante duas décadas da história brasileira.

O IDHM é um índice composto que agrega 3 das mais importantes dimensões do desenvolvimento humano: a oportunidade de viver uma vida longa e saudável, de ter acesso ao conhecimento e ter um padrão de vida que garanta as necessidades básicas, representadas pela saúde, educação e renda.

Vida longa e saudável é medida pela expectativa de vida ao nascer, calculada por método indireto a partir dos dados dos Censos Demográficos do IBGE. Esse indicador mostra o número médio de anos que as pessoas viveriam a partir do nascimento, mantidos os mesmos padrões de mortalidade observados no ano de referência.

Padrão de vida é medido pela renda municipal per capita, ou seja, a renda média de cada residente de determinado município. É a soma da renda de todos os residentes, dividida pelo número de pessoas que moram no município - inclusive crianças e pessoas sem registro de renda. Os dados são do Censo Demográfico do IBGE.

Os três componentes acima são agrupados por meio da média geométrica, resultando no IDHM.

Análise inicial dos dados - IDH 2016

Dicionário de dados

- **Territorialidade:** Nome do Estado
- **Posição IDHM:** Posição do Estado em relação ao IDHM geral
- **IDHM:** Valor do Índice de Desenvolvimento Humano Municipal
- **Posição IDHM Renda:** Posição do Estado em relação ao IDHM quanto à Renda
- **IDHM Renda:** Valor do Índice de Desenvolvimento Humano Municipal quanto à Renda
- **Posição IDHM Educação:** Posição do Estado em relação ao IDHM quanto à Educação
- **IDHM Educação:** Valor do Índice de Desenvolvimento Humano Municipal quanto à Educação
- **Posição IDHM Longevidade:** Posição do Estado em relação ao IDHM quanto à Longevidade
- **IDHM Longevidade:** Valor do Índice de Desenvolvimento Humano Municipal quanto à Longevidade

Visualizando a base de dados

	Territorialidade	Posição IDHM	IDHM	Posição IDHM Renda	IDHM Renda	Posição IDHM Educação	IDHM Educação	Posição IDHM Longevidade	IDHM Longevidade
0	Distrito Federal	1	0,854	1	0,885	2	0,824	1	0,853
1	São Paulo	2	0,831	5	0,851	1	0,837	2	0,807
2	Santa Catarina	3	0,805	3	0,862	3	0,784	4	0,773
3	Rio de Janeiro	4	0,794	4	0,853	7	0,759	4	0,773
4	Paraná	5	0,792	9	0,839	5	0,772	5	0,768

Tratamento da base de dados

A primeira letra de cada palavra do nome do Estado foi colocada em maiúsculo para ficar de acordo com as demais bases de dados.

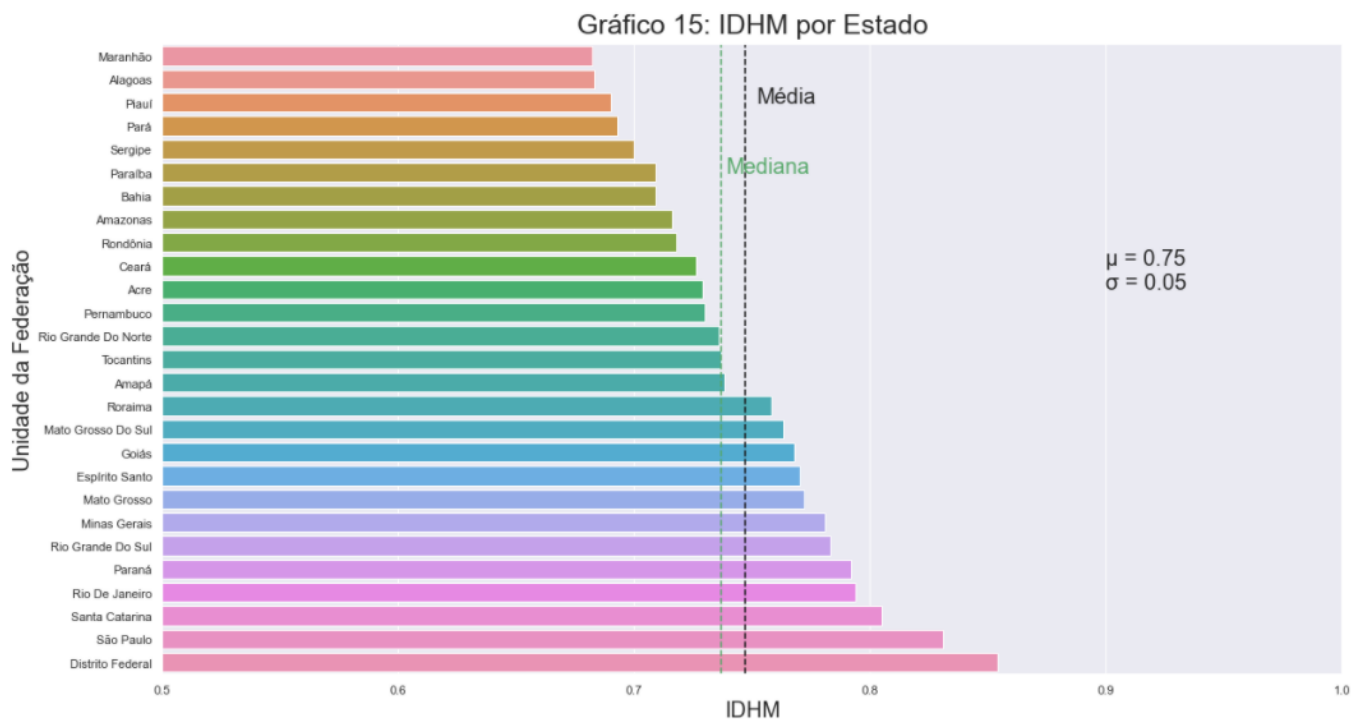
As colunas "Posição IDHM", "Posição IDHM Renda" e "Posição IDHM Educação" foram removidas pois é mais fácil mostrar as posições de cada Estado em relação ao IDHM por meio dos seus valores de IDHM geral e gráficos.

Quanto às colunas "Posição IDHM Longevidade" e "IDHM Longevidade" não são consideradas importantes devido ao fato de que essa análise não abordará o tema "longevidade".

Os valores das colunas "IDHM", "IDHM Renda" e "IDHM Educação" estavam originalmente com a notação brasileira para números reais, e por isso a vírgula foi substituída por um ponto e então foram transformados em float, para que fosse possível realizar cálculos.

	Territorialidade	IDHM	IDHM Renda	IDHM Educação
0	Distrito Federal	0.854	0.885	0.824
1	São Paulo	0.831	0.851	0.837
2	Santa Catarina	0.805	0.862	0.784
3	Rio De Janeiro	0.794	0.853	0.759
4	Paraná	0.792	0.839	0.772

IDHM por Estado



O Gráfico 15 apresenta o IDHM por Estado.

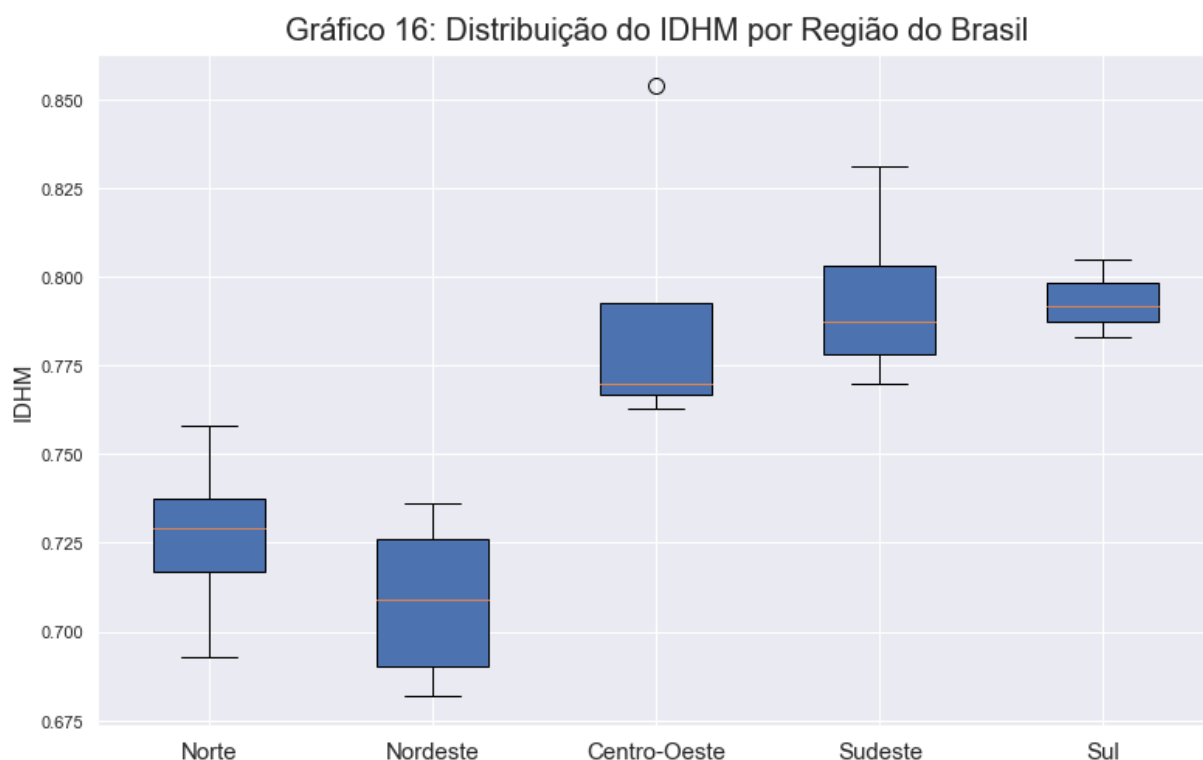
Diferentemente do Gráfico 5, em que era clara a diferença entre os valores de Inse dos Estados, temos no Gráfico 13 uma distribuição muito mais uniforme do IDHM, fato verificado pelo desvio padrão de apenas 0,05.

Observamos também que os três Estados com maiores IDHM são: Santa Catarina, Rio de Janeiro e Distrito Federal, todos de Regiões diferentes. Entretanto, os três Estados com piores IDHM são todos da Região Nordeste: Maranhão, Alagoas e Piauí.

Visto a distribuição do IDHM por Unidade da Federação, surge o seguinte questionamento: "Como é a distribuição do IDHM por Região do país"?

IDHM por Região do Brasil

A partir das bases do IDH e Relação entre Estados e Regiões do Brasil, iniciamos por construir uma tabela única contendo as informações relativas a cada Estado e seu IDHM. Buscamos, então, compreender quais Regiões possuem maior e menor IDHM, e que relação esse fato possui com os valores de Inse já apresentados nesse documento.



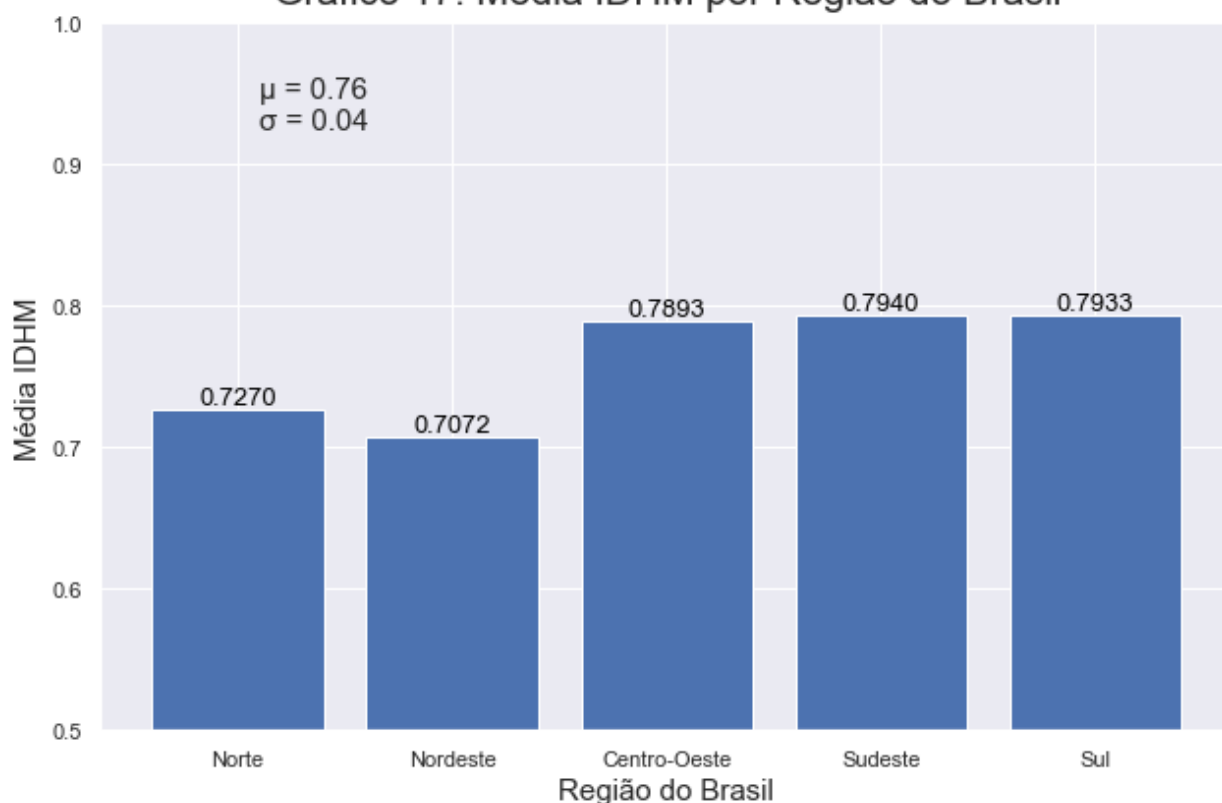
O Gráfico 16 apresenta a distribuição do IDHM por Região do Brasil.

Observa-se a clara discrepância com relação às Regiões Norte e Nordeste às demais.

A Região Nordeste é a Região com IDHM mais baixo, e vimos no Gráfico 7 que é a Região com valores de Inse mais baixos também.

Em realidade, os gráficos 7 e 16 são muito parecidos, pois no Gráfico 7 temos que as Regiões do menor ao maior valor de Inse médio são: Nordeste, Norte, Centro-Oeste, Sudeste e Sul, enquanto no Gráfico 16 temos que as Regiões do menor ao maior valor de IDHM são: Nordeste, Norte, Centro-Oeste, Sul e Sudeste.

Gráfico 17: Média IDHM por Região do Brasil



O Gráfico 17 mostra a média do valor de IDHM por Região do Brasil.

Podemos observar uma diferença muito pequena entre as Regiões Sudeste e Sul, de aproximadamente 0.0007. Entretanto, a diferença entre as Regiões Nordeste e Sudeste é de aproximadamente 0.0868, e tanto a Região Norte quanto a Nordeste encontram-se abaixo da média.

A partir dos gráficos acima, podemos observar que o valor do PIB de cada Região não é fator decisivo para o quão elevado o seu Inse é, tampouco o valor investido em educação, mas sim o IDHM. Ainda que um PIB elevado possa contribuir para melhorias nos índices socioeconômicos da educação, observamos, por exemplo, que a Região Centro-Oeste possui valor de IDHM muito próximo aos das Regiões Sudeste e Sul, mas possui um PIB menor do que a Região Nordeste.

Conclusão

A partir dos dados obtidos, pudemos observar como as desigualdades educacionais e socioeconômicas se manifestam nas diferentes regiões do país. Como caso particular, notamos os casos das regiões Norte e Nordeste, indicando a necessidade de políticas públicas voltadas para esses locais.

Considerando os dados e gráficos apresentados neste documento, fica claro que analisar um conjunto de políticas e ações governamentais para garantir uma Educação de qualidade é algo complexo, que não pode se resumir à atividade econômica e investimentos feitos por cada Estado e Região em um ano. É preciso analisar diversos outros fatores políticos e sociais.

Por fim, esperamos que essa análise, mesmo que simples, sirva para constatar as desigualdades educacional e socioeconômica existentes no Brasil, e quais regiões precisam ser foco de atividades para solucionar essa problemática.

Código Github:

<https://github.com/RafaelxFernandes/PS-UFRJ-Analytica/tree/main/An%C3%A1lise>

Referências:

- <https://medium.com/ufri-analytica/datathon-open-data-day-base-dos-dados-86079c93945f>
- <https://basedosdados.org/dataset/br-inep-indicadores-educacionais>
- https://download.inep.gov.br/informacoes_estatisticas/indicadores_educacionais/2015/nota_tecnica/nota_tecnica_inep_inse_2015.pdf
- https://download.inep.gov.br/educacao_basica/enem/enem_por_escola/2015/nota_tecnica_indicador_nivel_socioeconomico.pdf
- <https://www.ibge.gov.br/explica/pib.php>
- <https://extremepresentation.typepad.com/files/choosing-a-good-chart-09.pdf>
- <https://stackoverflow.com/questions/37593550/replace-method-not-working-on-pandas-dataframe>
- https://matplotlib.org/3.1.1/gallery/pyplots/boxplot_demo_pyplot.html#sphx-glr-gallery-pyplots-boxplot-demo-pyplot-py
- <https://stats.stackexchange.com/questions/3476/how-to-name-the-tick-s-in-a-python-matplotlib-boxplot>
- <https://seaborn.pydata.org/generated/seaborn.barplot.html>
- <https://seaborn.pydata.org/generated/seaborn.countplot.html>
- <https://stackoverflow.com/questions/43214978/seaborn-barplot-displaying-values>
- <https://stackoverflow.com/questions/36780948/seaborn-matplotlib-how-to-repress-scientific-notation-in-factorplot-y-axis>

- <https://stackoverflow.com/questions/46735745/how-to-control-scientific-notation-in-matplotlib>
- <https://pythonforundergradengineers.com/unicode-characters-in-python.html>
- <https://matplotlib.org/stable/tutorials/text/annotations.html>
- <https://pandas.pydata.org/pandas-docs/stable/reference/api/pandas.DataFrame.join.html>
- <https://www.datacamp.com/community/tutorials/joining-dataframes-pandas>
- <https://stackoverflow.com/questions/30228069/how-to-display-the-value-of-the-bar-on-each-bar-with-pyplot-barh>
- <https://datavizpyr.com/sort-bars-in-barplot-using-seaborn-in-python/>
- <http://www.atlasbrasil.org.br/>
- <https://pandas.pydata.org/pandas-docs/stable/reference/api/pandas.DataFrame.drop.html>
- https://pandas.pydata.org/pandas-docs/stable/user_guide/groupby.html
- <https://stackoverflow.com/questions/43131274/how-do-i-plot-two-count-plot-graphs-side-by-side-in-seaborn>
- <https://brasilecola.uol.com.br/o-que-e/geografia/o-que-e-idh.htm>
- <https://www.br.undp.org/content/brazil/pt/home/idh0/conceitos/o-que-e-o-idhm.html>
- <https://operdata.com.br/blog/como-interpretar-um-boxplot/>
- https://www.correiobraziliense.com.br/app/noticia/opiniao/2019/12/16/internas_opiniao.814380/artigo-educacao-e-desigualdade.shtml
- <https://blog.nubank.com.br/o-que-e-pib/>
- <https://tomwebmaraba.wordpress.com/finbra/>