

Análise de investimentos em transporte e infraestrutura e seus impactos

Arick Jurdan dos Reis

October 29, 2024

Abstract

Esse artigo apresenta uma análise de investimentos em transporte/infraestrutura no Brasil e seus impactos, utilizando ferramentas de linguagem de programação. O objetivo é fazer a compreensão dos dados relacionados a transporte e infraestrutura, como indicadores e estatísticas e descobrir insights para justificar os impactos e fazer previsões.

1 Introduction

Os dados são a base fundamental para fazer uma análise de investimentos, monitoramento e avaliação de políticas de mobilidade e desenvolvimento urbano. Eles fornecem informações sobre os custos e benefícios dos projetos, os impactos na economia e na sociedade, e os riscos e incertezas associados. Sem dados confiáveis, é impossível tomar decisões de investimento informadas e eficazes.

Para analisar os dados, utilizamos a linguagem de programação R. O R é uma linguagem poderosa e versátil para análise de dados, com uma ampla gama de bibliotecas e ferramentas disponíveis. A principal biblioteca utilizada neste estudo foi: ggplot, para visualização dos dados. Para manipular e fazer consultas nas bases de dados utilizamos a linguagem SQL junto com o serviço de armazenamento e análise de dados na nuvem oferecido pelo Google, o BigQuery.

2 Apresentação dos dados

Neste estudo, utilizamos diversas bases de dados para obter informações sobre investimentos em transporte e infraestrutura no Brasil. As principais bases de dados utilizadas foram: [Atlas Esgotos](#), [Indicadores de Mobilidade e Transportes](#), [Frota de veículos no Brasil](#) e [Atlas Esgotos](#). Disponíveis de forma gratuita pela organização [Base dos Dados](#).

2.1 Atlas Esgotos

Desenvolvido sob a coordenação do da Agência nacional de águas(ANA), essa base de dados apresenta informações, diagnósticos e propostas de soluções para todas as 5.570 sedes municipais do País, amparados na avaliação da situação da coleta e tratamento dos esgotos. A cobertura temporal é de 2013 e possui 37 colunas, mas usaremos apenas as mais relevantes para a análise, na Figura 1:

	sigla_uf	populacao	investimento_trat_colet	indice_atendimento	invest_por_habit
1	AC	562843	841152205	0.237	1494.47
2	AL	2437832	3115972924	0.266	1278.17
3	AM	3014220	4305944959	0.230	1428.54
4	AP	658840	1136176495	0.217	1724.51
5	BA	10880101	9595973249	0.420	881.97

Figure 1: agrupamento dos dados: Atlas Esgotos

A seguir, na Figura 2 apresentamos um gráfico, que relaciona os três primeiros campos do dataset. O método que utilizamos para obtenção desse gráfico foi utilizando a função `ggplot()`, e inserimos, como parametro, dois elementos, que representam o gráfico de barras e o grafico de linha.

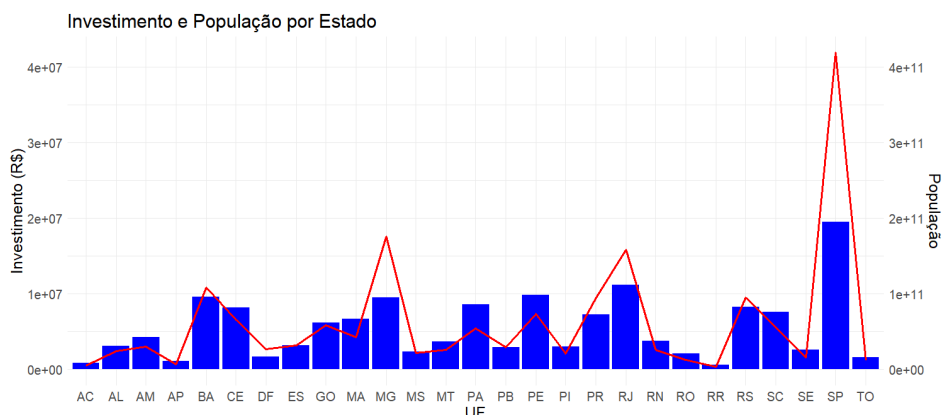


Figure 2: Investimento em Infraestrutura(em azul) e população(em vermelho) por estado

O gráfico revela uma disparidade significativa no investimento e na população entre os estados brasileiros. São Paulo se destaca como o estado com maior investimento e população. Mas essa disparidade é justificada por questões geográficas e histórico de desenvolvimento.

Podemos observar uma relação: o gráfico mostra uma correlação positiva entre investimento e população. Estados com maior investimento tendem a ter maior população, e vice-versa. Essa relação pode ser explicada por diversos fatores, como melhoria da qualidade de vida e atração da mão de obra.

2.2 Taxa de Motorização

Nesse conjunto obtemos dados de taxa de motorização nas 26 capitais de estados brasileiros e Distrito Federal, disponibilizado pela plataforma MobiliDADOS. A taxa de motorização é um indicador que mede a relação entre o número de veículos automotores e a população de uma região. É calculada dividindo o número de veículos por 1.000 habitantes.

Novamente fizemos o agrupamento dos dados durante a consulta e selecionamos os campos que serão utilizados:

	ano	sigla_uf	id_municipio	taxa_motorizacao
1	2001	SE	2800308	299.20969
2	2001	PA	1501402	103.11091
3	2001	MG	3106200	291.75419
4	2001	RR	1400100	180.12220
5	2001	DF	5300108	290.95991
6	2001	MS	5002704	258.65437

Figure 3: agrupamento dos dados: Taxa de Motorização

Para representar a taxa de motorização ao longo do tempo nos estados do Brasil, foi necessário organizar os dados de forma que eles tenham uma coluna para região. Usamos um gráfico de linhas com múltiplas séries temporais, onde cada série representa uma região, na Figura 4.

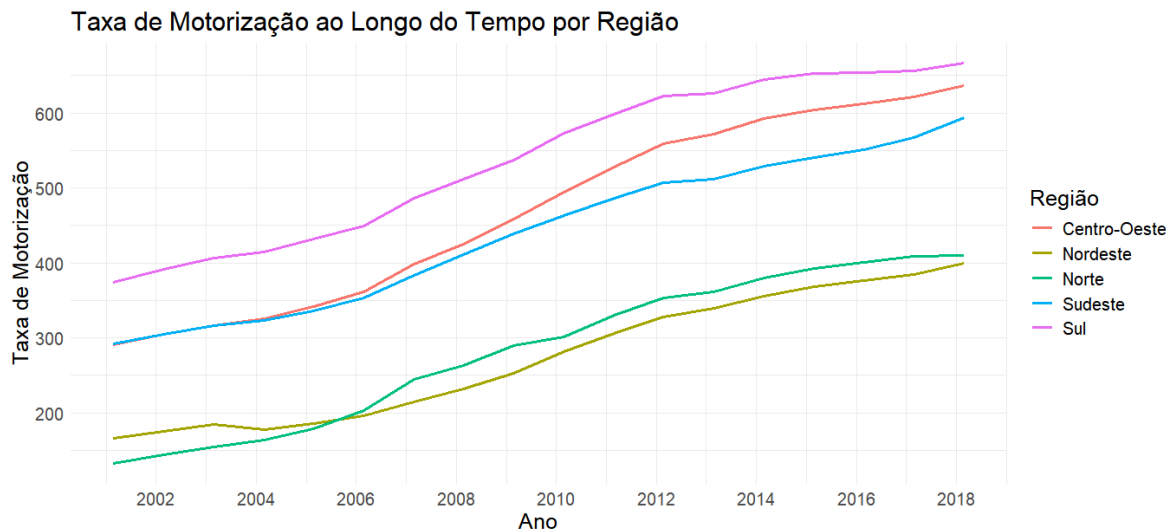


Figure 4: Série temporal: Taxa de Motorização ao longo do tempo por região

O gráfico mostra que a taxa de motorização no Brasil vem crescendo constantemente ao longo dos anos. Esse crescimento pode ser atribuído a um maior investimento em infraestrutura de transporte, que facilita o deslocamento de pessoas e bens. Também é importante considerar impactos negativos desse crescimento, como, por exemplo, a emissão de gases poluentes pelos veículos.

2.3 Emissão de CO2

Esse conjunto nos fornece dados de emissão de CO2 e Material Particulado por habitante resultantes do uso de combustível, nas 26 capitais dos estados brasileiros e Distrito Federal. A cobertura temporal é entre 2007 e 2018.

A análise desses dados serão importantes para entender o impacto causado pelo crescimento da frota de veículos, sendo fundamental para direcionar investimentos em transporte e infraestrutura sustentáveis.

Para otimizar a visualização dos dados, selecionamos duas tabelas: Na primeira temos a emissão por habitante de CO2 e material particulado ao longo do tempo, dividido por região. E na segunda a média de emissão do Brasil

	ano	regiao	f0_	f1_
1	2007	Nordeste	576.8	135.8
2	2007	Norte	886.9	224.6
3	2007	Sudeste	685.3	129.3
4	2007	Centro-Oeste	1155.0	285.8
5	2007	Sul	934.3	172.0
6	2008	Nordeste	621.8	143.2

Figure 5: Por região

	ano	emissao_co2	emissao_mp
1	2007	798.6	184.1
2	2008	830.1	186.7
3	2009	819.0	171.7
4	2010	909.8	172.4
5	2011	962.3	161.8
6	2012	1060.0	158.3

Figure 6: Brasil

Para representar a Emissão de CO2 por habitante ao longo do tempo, usamos um gráfico de linhas com múltiplas séries temporais, onde cada cor representa uma região, e em preto a média brasileira. O método que usamos para plotagem, na linguagem R, consiste em combinar dois gráficos em um único gráfico com as linhas sobrepostas, usando a função `geom_line()` duas vezes dentro do mesmo `ggplot()`.

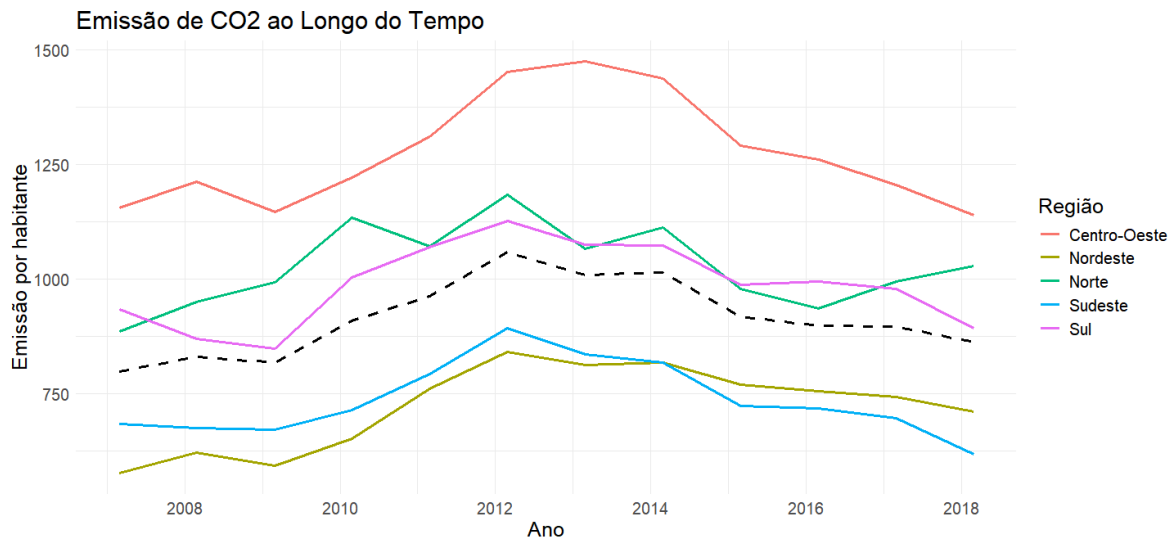


Figure 7: Série temporal: Emissão CO2 ao longo do tempo (em preto a média brasileira)

Nesse gráfico, é possível observar um aumento geral nas emissões durante o período, com algumas regiões apresentando crescimento mais acentuado do que outras.

2.4 Acessibilidade

Por fim, apresentamos os dados do projeto Acesso a Oportunidades, do Instituto de Pesquisa Econômica Aplicada (ipea), que divulga estimativas de acesso a postos de saúde e educação por modo de transporte para as 20 maiores cidades do Brasil, no ano de 2019.

É importante incluir esses dados na análise, pois podem identificar áreas com carência de acesso a serviços de saúde e educação, permitindo que os investimentos em transporte e infraestrutura sejam direcionados para regiões que mais precisam.

	id_municipio	populacao	estabelecimentos_ensino	estabelecimentos_saude	municipio
1	3550308	11039920	2723	891	São Paulo (SP)
2	3304557	6119114	1888	393	Rio de Janeiro (RJ)
3	2927408	2598176	651	344	Salvador (BA)
4	5300108	2537123	688	272	Brasília (DF)
5	2304400	2405318	473	235	Fortaleza (CE)
6	3106200	2352062	555	346	Belo Horizonte (MG)

Figure 8: agrupamento dos dados: Atlas Esgotos

Na figura 9, apresentamos um gráfico de colunas, com barras horizontais, representando a quantidade de postos de ensino e saúde, respectivamente, para cada uma das cidades.

É importante analisar os dados de forma crítica e contextualizada, considerando as características de cada cidade (por exemplo, a população) e os diferentes fatores que podem influenciar o acesso a esses serviços. O gráfico apresenta dados para as maiores cidades do Brasil, mas a representatividade pode variar em outras regiões do país.

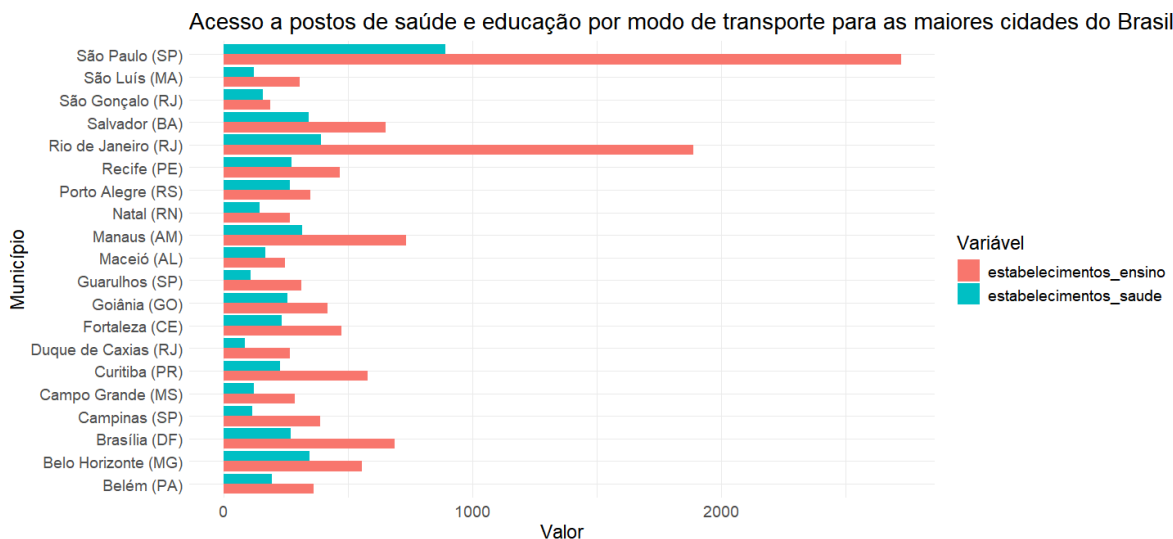


Figure 9: agrupamento dos dados: Atlas Esgotos

3 Análise de dados multiconjuntos:

A análise de dados multiconjuntos é uma técnica poderosa que combina dados de várias fontes para obter uma visão mais completa de um problema.

A partir da apresentação dos dados, surgem diversas perguntas interessantes que podem ser respondidas através de uma análise de dados multiconjuntos. Por exemplo, existe uma relação entre a taxa de motorização de uma região e a sua emissão de CO₂? Como a qualidade da infraestrutura de esgoto em uma área está relacionada com o acesso a serviços de saúde e educação? Existe uma correlação entre o acesso a serviços de saúde e educação em determinadas regiões? Essas são apenas algumas perguntas possíveis e tentaremos responder algumas delas com os conjuntos de dados que temos apresentado.

3.1 Impacto da infraestrutura de esgoto na saúde e educação

O impacto da infraestrutura de esgoto na disponibilidade de postos de saúde e educação é uma questão de grande importância, pois uma infraestrutura eficiente de esgoto pode ter diversos efeitos positivos nessas áreas. A infraestrutura de esgoto desempenha um papel fundamental na promoção da saúde pública, na prevenção de doenças transmitidas pela água e no fornecimento de um ambiente mais seguro e saudável para os moradores.

No entanto, é importante ressaltar que a relação entre a infraestrutura de esgoto, a disponibilidade de postos de saúde e educação não pode ser adequadamente demonstrada nesse artigo através dos dados disponíveis, devido a várias limitações. Uma dessas limitações é que os dados foram coletados de regiões geográficas diferentes e em diferentes faixas temporais. Isso significa que as condições socioeconômicas e do ambiente podem variar entre as regiões e ao longo do tempo, o que dificulta a análise comparativa e a identificação de padrões consistentes.

3.2 Correlação entre taxa de motorização e emissão de CO₂

Para determinar se existe uma correlação entre a taxa de motorização e a taxa de emissão de CO₂ ao longo do tempo, usamos um método de cálculo do coeficiente de correlação de Pearson entre as duas séries temporais. Se o coeficiente de correlação for próximo de 1 ou -1, isso indica uma forte correlação positiva ou negativa, respectivamente. Um valor próximo de 0 indica uma falta de correlação linear.

```
# Selecionando as séries temporais relevantes
serie_taxa_motorizacao <- dados_juntos$media_taxa_motorizacao
serie_emissao_co2 <- dados_juntos$f0_
# Calculando a correlação entre as séries temporais
correlacao <- cor(serie_taxa_motorizacao, serie_emissao_co2)
print(correlacao)
```

Para calcular o coeficiente de correlação de Pearson, utilizamos uma função nativa da linguagem R chamada `cor()`. A função `cor()` aceita dois argumentos principais: "a" e "b", que são as duas colunas para as quais calculamos a correlação.

O trecho de código teve a seguinte saída: `[1] 0.4690761`.

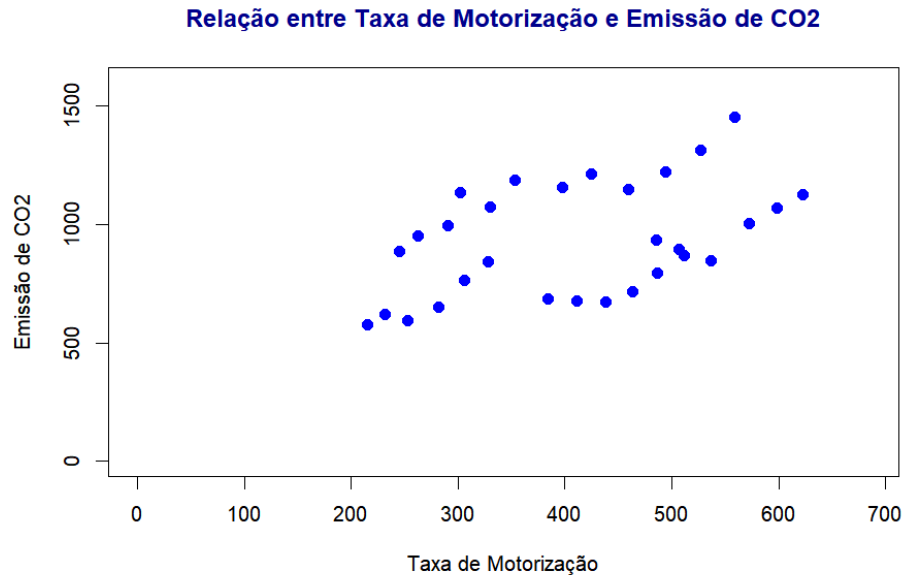


Figure 10: Gráfico de Dispersão: Taxa de Motorização x Emissão de CO2

Na figura 10 exibimos um gráfico, utilizando a função de plotagem nativa do R, para mostrar uma correlação positiva moderada entre a taxa de motorização e a emissão de CO2 ao longo do tempo. Isso significa que, à medida que a taxa de motorização aumenta, a emissão de CO2 também aumenta. Essa correlação é esperada, pois os veículos automotores são uma das principais fontes de emissão de CO2.

4 Conclusão

Ao explorar os dados fornecidos pela Base de Dados foi possível compreender melhor a interação complexa entre o desenvolvimento de infraestrutura, o aumento da motorização e as consequências ambientais associadas.

Quanto ao uso das ferramentas de linguagem de programação R e SQL, posso concluir que foram fundamentais para conduzir uma análise abrangente e rigorosa dos dados. O R me permitiu manipular os datasets, criar visualizações e extrair insights significativos dos conjuntos de dados. Por sua vez, o SQL foi essencial para fazer consultas com filtragem e agrupamento nos grandes volumes de dados.

References

- [1] *LINK DO GITHUB*: https://github.com/Arickjd/PS_UFRJAnalytics2024.
- [2] *Indicadores de Mobilidade e Transportes*. Disponível em: <https://basedosdados.org/dataset/br-mobilidadados-indicadores>. Acessado em: 21/02/2024.

- [3] *Atlas Esgotos*. Disponível em: <https://basedosdados.org/dataset/br-ana-atlas-esgotos>. Acessado em: 21/02/2024.
- [4] *Projeto Acesso a Oportunidades*. Disponível em: <https://basedosdados.org/dataset/br-ipea-acesso-oportunidades>. Acessado em: 21/02/2024.