

Introdução

O conjunto de dados escolhido para este estudo exploratório é o “Billionaires Statistics Dataset”, obtido do Kaggle. Este banco de dados compreende 2640 observações e 35 variáveis, incluindo informações sobre o patrimônio total de cada bilionário, país de origem, idade, gênero e a origem de sua riqueza. O objetivo deste estudo é investigar o comportamento do patrimônio total acumulado em relação a variáveis de diferentes categorias, visando identificar possíveis associações entre elas.

Metodologia

Este conjunto de dados abrange um total de 35 variáveis que armazenam dados sobre bilionários, porém, para deixar o trabalho mais suscinto e organizado, foram escolhidas as variáveis que os estudantes julgaram ser mais interessantes para desenvolver a análise exploratória.

Table 1: Explicação das Variáveis

| Variavel | Explicacao |
|----------------------|--|
| Rank | Posição do bilionário em uma classificação ou lista. |
| FinalWorth | Fortuna total do bilionário em dólares. |
| Category | Categoria à qual o bilionário pertence. |
| PersonName | Nome do bilionário. |
| Age | Idade do bilionário. |
| Country | País de origem do bilionário. |
| City | Cidade associada ao bilionário. |
| Source | Fonte ou origem da riqueza do bilionário. |
| Industries | Setores de indústrias em que o bilionário está envolvido. |
| CountryOfCitizenship | País de cidadania do bilionário. |
| Organization | Organização ou empresa associada ao bilionário. |
| SelfMade | Indica se a fortuna do bilionário foi adquirida por meio de esforços próprios. |
| Status | Status do bilionário. |
| Gender | Gênero do bilionário. |
| Birthdate | Data de nascimento do bilionário. |
| Lastname | Sobrenome do bilionário. |
| Firstname | Nome próprio do bilionário. |
| Title | Título associado ao bilionário. |
| Date | Data associada a alguma informação específica. |
| State | Estado associado ao bilionário. |
| Residence | Região ou estado de residência do bilionário. |
| BirthYear | Ano de nascimento do bilionário. |
| BirthMonth | Mês de nascimento do bilionário. |
| Birthday | Dia de nascimento do bilionário. |
| CPI_Country | Índice de Preços ao Consumidor (CPI) para o país do bilionário. |
| CPI_Change | Mudança no Índice de Preços ao Consumidor (CPI) para o país do bilionário. |
| GDP_Country | Produto Interno Bruto (PIB) do país do bilionário. |
| Education_Enrollment | Taxa bruta de matrícula no ensino superior no país do bilionário. |
| Primary_Education | Taxa bruta de matrícula no ensino fundamental no país do bilionário. |
| Life_Expectancy | Expectativa de vida no país do bilionário. |
| Tax_Revenue | Receita fiscal para o país do bilionário. |
| Total_Tax_Rate | Taxa total de impostos para o país do bilionário. |
| Population_Country | População do país do bilionário. |
| Latitude_Country | Latitude geográfica do país do bilionário. |
| Longitude_Country | Longitude geográfica do país do bilionário. |

Para o estudo das variáveis, foram empregadas medidas de centralidade (como média e mediana) e medidas de dispersão (como desvio padrão e variância). Além disso, foram realizados cálculos estatísticos para verificar a existência de associações entre as variáveis e para quantificar a força dessas associações. A análise é complementada por tabelas e gráficos que proporcionam uma representação visual do comportamento de cada variável, facilitando a interpretação dos resultados. Essa abordagem abrangente visa fornecer uma compreensão profunda das relações entre as variáveis no contexto dos bilionários estudados.

| Medidas | Formula |
|-----------------------|---|
| Média | $\bar{x} = \frac{1}{n} \sum_{i=1}^n x_i$ |
| Mediana | $\begin{cases} x_{(k)} & \text{se } n \text{ é ímpar,} \\ \frac{1}{2}(x_{(k)} + x_{(k+1)}) & \text{se } n \text{ é par,} \end{cases}$ |
| Desvio Padrão | $\sqrt{\frac{\sum_{i=1}^n (x_i - \bar{x})^2}{n-1}}$ |
| Mínimo | $\min(x)$ |
| Máximo | $\max(x)$ |
| Assimetria | $\frac{\frac{1}{n} \sum_{i=1}^n (x_i - \bar{x})^3}{(\frac{1}{n} \sum_{i=1}^n (x_i - \bar{x})^2)^{3/2}}$ |
| Coef. de Pearson | $r = \frac{n \sum xy - (\sum x)(\sum y)}{\sqrt{[n \sum x^2 - (\sum x)^2][n \sum y^2 - (\sum y)^2]}}$ |
| Coef. de Determinação | $R^2 = 1 - \frac{\text{var}(x)}{\text{var}(y)}$ |

Resultados

1. Relação entre a variável gênero e o patrimônio total

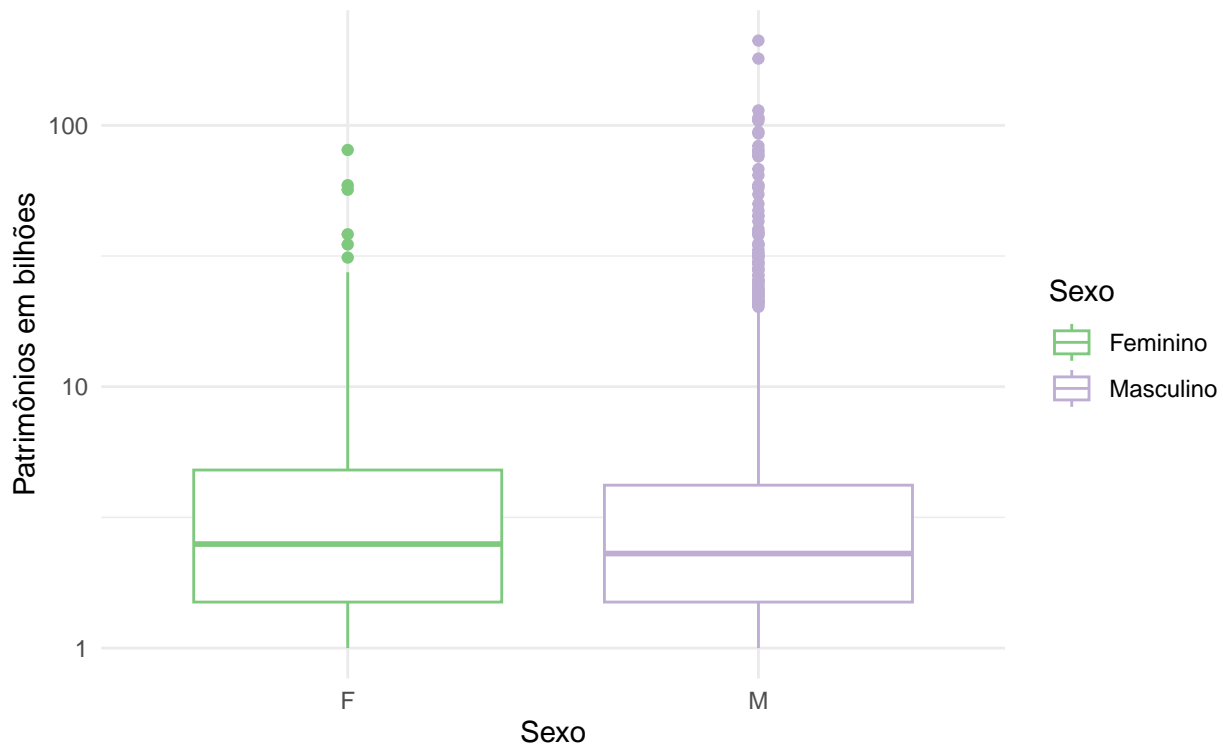


Figure 1: Boxplot: Sexo vs. patrimônio total em bilhões de dólares

O valor de R^2 é: 4.32614e-06

Table 2: Medidas do patrimônio total em Bilhões

| Sexo | Media | Mediana | S | Min | Max | Outliers | n |
|-----------|-------|---------|-------|-----|-------|----------|------|
| Masculino | 4.63 | 2.3 | 10.12 | 1 | 211.0 | 0.092 | 2303 |
| Feminino | 4.57 | 2.5 | 7.61 | 1 | 80.5 | 0.068 | 337 |

Antes de procedermos à análise de associação, é relevante considerar algumas observações sobre as variáveis em questão. O conjunto de dados revela uma disparidade significativa entre o número de observações para o gênero masculino em comparação com o feminino, afinal, neste banco de dados existem 2640 observações e somente 337 delas são referentes a pessoas do gênero Feminino. Embora a disparidade persista, é essencial destacar um aumento significativo, a cerca de 50%, no número de mulheres bilionárias nos últimos anos. Esse aumento supera até mesmo o crescimento observado entre os homens no mesmo período. Assim, embora a maioria dos dados ainda esteja centrada no gênero masculino, a presença feminina registrou um notável e considerável aumento. É crucial destacar que a análise proposta não busca estabelecer causalidade. Desse modo, não se pode prever a realidade das mulheres que não são bilionárias e o porquê disso.

Ao examinar o boxplot, é evidente a presença de vários outliers na distribuição do patrimônio para o gênero masculino. A caixa correspondente a esse grupo é visualmente menor, sugerindo, inicialmente, uma maior homogeneidade nos dados em comparação com o gênero feminino. Contudo, essa impressão é contrabalançada ao considerarmos o cálculo do desvio padrão. Os resultados indicam que a distribuição do patrimônio para o gênero feminino é mais centralizada, uma vez que apresenta menos outliers, mesmo que a caixa seja mais ampla. É válido observar que ambas as caixas são iniciadas do mesmo ponto, indicando a possibilidade de pelo menos um homem e uma mulher possuírem o mesmo patrimônio.

A análise do coeficiente de determinação (R^2) é crucial para avaliar a existência de associação entre as variáveis. Notavelmente, o valor obtido revela-se muito baixo, indicando uma associação extremamente fraca. Essa análise nos leva a crer que as variáveis gênero e patrimônio total são praticamente independentes uma da outra. Assim, a hipótese de associação não se sustenta, reforçando a ideia de que não há uma relação substancial entre o gênero e o patrimônio total nos dados analisados.

2. Relação entre as indústrias e o patrimônio acumulado dos bilionários

Para estudar a relação entre duas variáveis, é necessário primeiramente representar os dados.

Graficamente pelo box-plot:

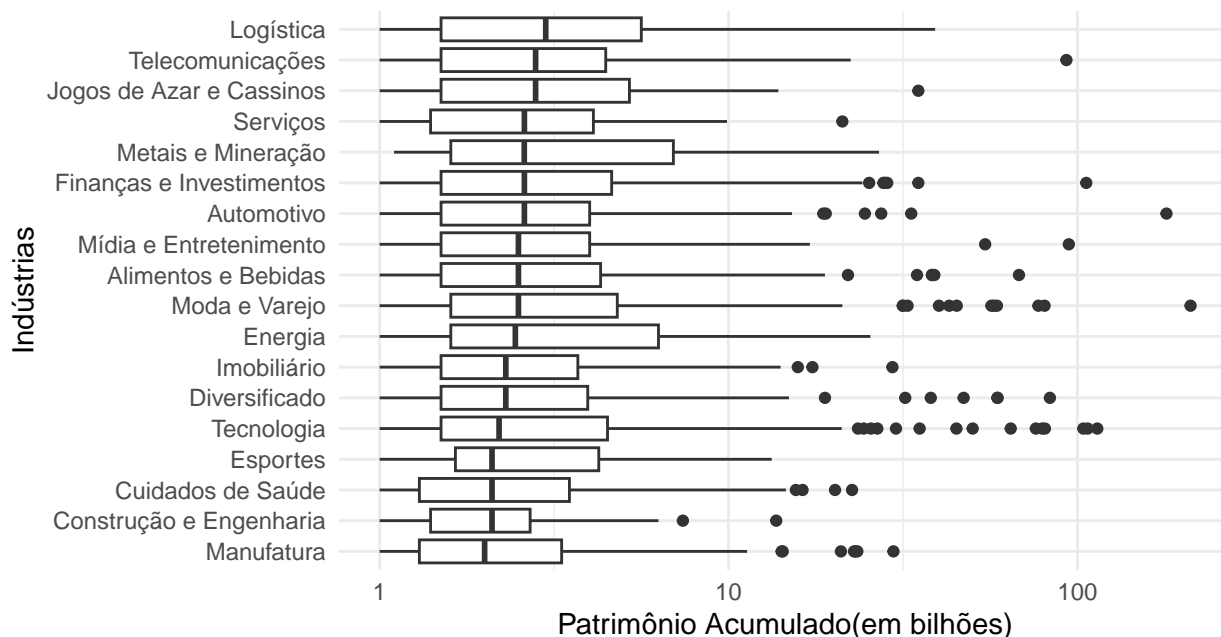


Figure 2: Box-plot do patrimônio acumulado(em log10) por bilionários daquela indústria

Analisando o boxplot, observamos certo comportamento padronizado independente da indústria. Temos a indústria tecnológica com alta quantidade de outliers e logística com maior mediana. Contudo foi-se também calculado o coeficiente de determinação(R^2) para verificar a falta de associação entre as variáveis indústria e patrimônio acumulado. Tal coeficiente é igual a 0.01. Como ele varia de 0 a 1 e o nosso apresenta um valor muito próximo de 0, podemos afirmar a falta de associação. Isso representa que saber ou não a indústria do bilionário, não nos ajuda a saber o patrimônio acumulado dele.

Pela tabela:

Ademais, foi-se feito uma tabela com certas medidas-resumo para descrever melhor esses dados. A coluna outliers é a porcentagem da distribuição daquela indústria que são outliers.

Table 3: Medidas do patrimônio total em bilhões

| Industria | Media | Mediana | S | Min | Max | Outliers | n |
|--------------------------|-------|---------|-------|-----|-------|----------|-----|
| Automotivo | 7.20 | 2.60 | 21.47 | 1.0 | 180.0 | 0.151 | 73 |
| Metais e Mineração | 6.04 | 2.60 | 7.30 | 1.1 | 27.0 | 0.149 | 74 |
| Diversificado | 4.84 | 2.30 | 9.72 | 1.0 | 83.4 | 0.112 | 187 |
| Manufatura | 3.15 | 2.00 | 3.47 | 1.0 | 29.7 | 0.111 | 324 |
| Moda e Varejo | 6.39 | 2.50 | 16.38 | 1.0 | 211.0 | 0.109 | 266 |
| Alimentos e Bebidas | 4.52 | 2.50 | 7.16 | 1.0 | 68.0 | 0.108 | 212 |
| Logística | 5.99 | 3.00 | 8.61 | 1.0 | 39.1 | 0.100 | 40 |
| Tecnologia | 5.98 | 2.20 | 14.04 | 1.0 | 114.0 | 0.099 | 314 |
| Imobiliário | 3.41 | 2.30 | 3.36 | 1.0 | 29.5 | 0.098 | 193 |
| Construção e Engenharia | 2.63 | 2.10 | 2.18 | 1.0 | 13.7 | 0.089 | 45 |
| Cuidados de Saúde | 3.20 | 2.10 | 3.27 | 1.0 | 22.6 | 0.085 | 201 |
| Jogos de Azar e Cassinos | 4.82 | 2.80 | 6.94 | 1.0 | 35.0 | 0.080 | 25 |
| Mídia e Entretenimento | 4.70 | 2.50 | 11.22 | 1.0 | 94.5 | 0.077 | 91 |
| Esportes | 3.45 | 2.10 | 3.04 | 1.0 | 13.3 | 0.077 | 39 |
| Finanças e Investimentos | 4.31 | 2.60 | 6.93 | 1.0 | 106.0 | 0.073 | 372 |
| Telecomunicações | 6.56 | 2.80 | 16.51 | 1.0 | 93.0 | 0.065 | 31 |
| Energia | 4.53 | 2.45 | 4.80 | 1.0 | 25.5 | 0.050 | 100 |
| Serviços | 3.27 | 2.60 | 3.17 | 1.0 | 21.2 | 0.038 | 53 |

Através da tabela observamos que a indústria automotiva tem a maior média e maior desvio padrão, como também segundo maior valor máximo. Essa indústria também tem a maior quantidade relativa de outliers. Suspeitamos, então de uma associação positiva entre a N(quantidade de observações daquela indústria) e o S(desvio-padrão). Ao representar graficamente, temos:

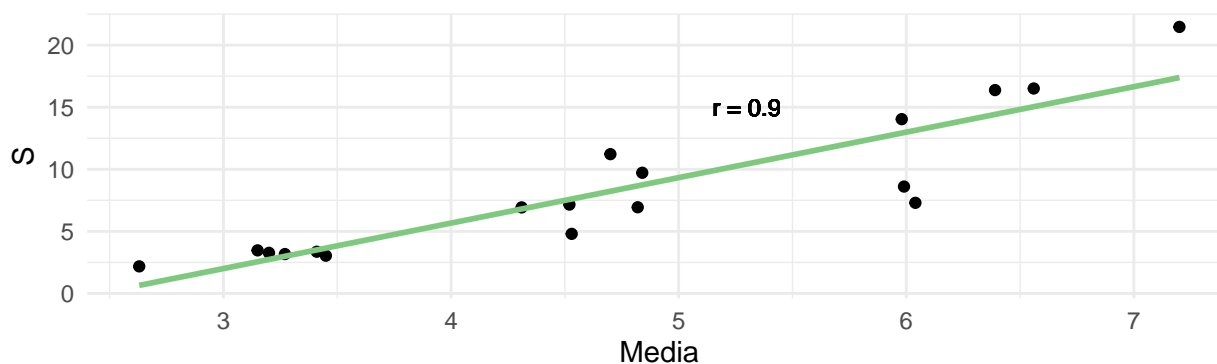


Figure 3: Gráfico de dispersão das médias pelos desvios-padrão das indústrias(em bilhões)

Dessa maneira, ao calcular o coeficiente de Pearson(r) podemos confirmar associação linear positiva e forte entre a média e o desvio-padrão do patrimônio acumulado dos bilionários separado por indústrias, visto que seu valor varia em modulo de 0 a 1 e como ele está tão próximo de 1, observamos associação linear forte, além disso como ele é positivo, isso indica uma associação linear positiva entre as duas variáveis. Isso mostra que quanto maior o patrimônio acumulado maior médio daquela indústria mais disperso são o patrimônio dos seus bilionários.

3. Relação entre a bilionários empreendedores ou não e seu patrimônio acumulado

É necessário em primeira. instância representar os dados graficamente. Como estamos tratando de uma variável quantitativa para duas categorias diferentes, utilizaremos novamente o boxplot.

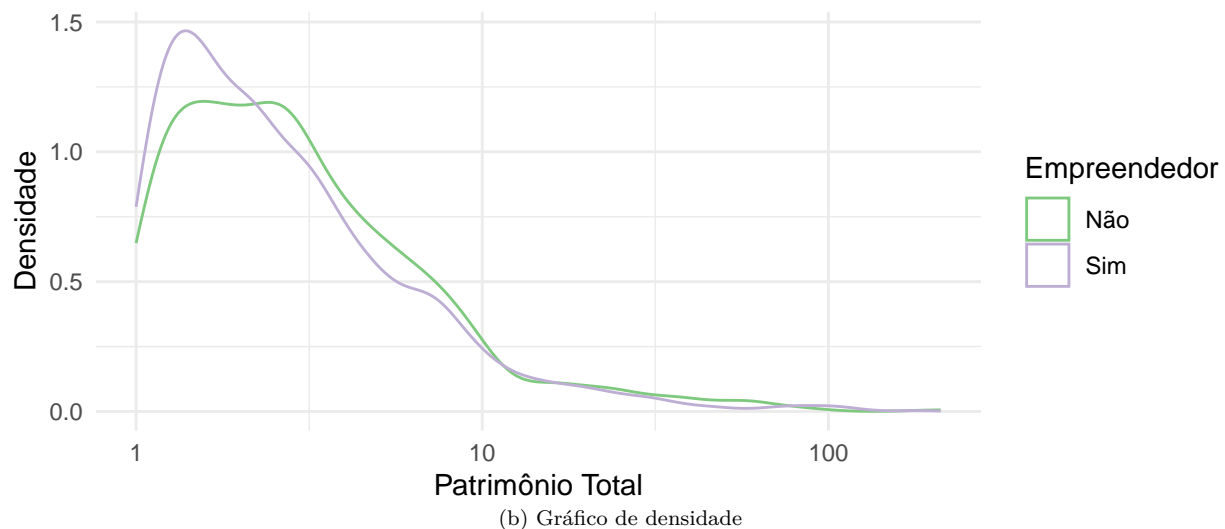
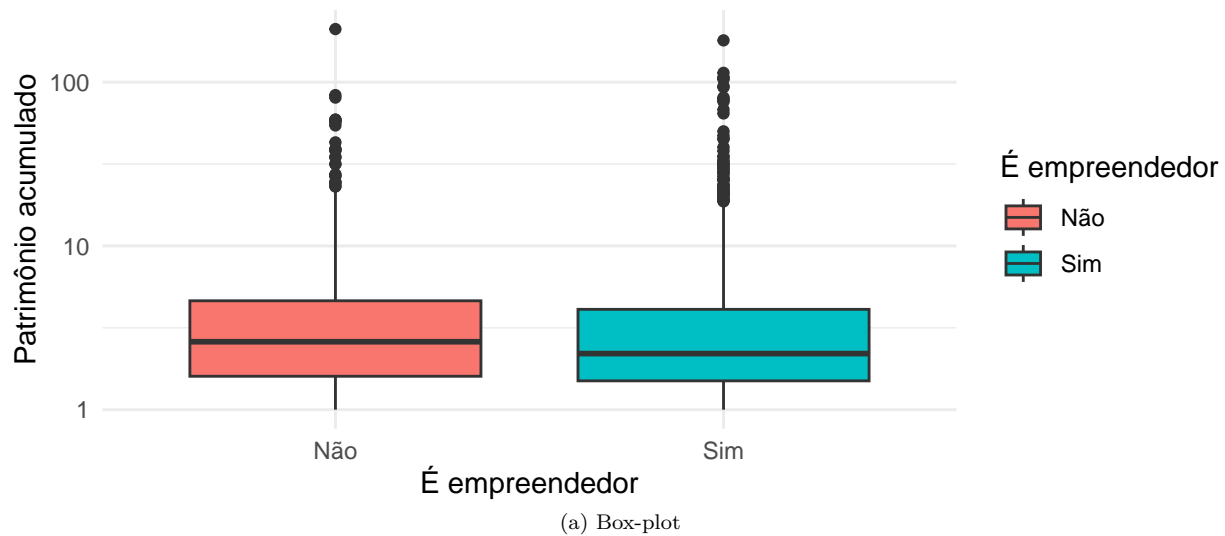


Figure 4: Gráficos do patrimônio acumulado separado pela variável empreendedor

Através do boxplot, podemos confirmar a presença de múltiplos outliers independente de se o bilionário for empreendedor ou não. Ao observar o gráfico de densidade, suspeita-se de uma assimetria positiva que ao calcularmos, obtemos 9.24 para os empreendedores e 11.04 para os outros. Com isso, confirmamos a presença

de assimetria positiva e forte, dado que seu valor em módulo é maior que 1. Além do mais devemos calcular o coeficiente de determinação novamente para verificar a falta ou não de associação entre as duas variáveis, uma quantitativa e a outra qualitativa. Como o R^2 assume o valor 2×10^{-4} , podemos afirmar uma falta de associação, dado que esse coeficiente varia entre 0 e 1 e encontramos um valor demasiado próximo de 0. Foi-se criada também uma tabela com as medidas-resumo para melhora analisar as categorias.

Table 4: Medidas do patrimônio acumulado em bilhões

| Empreendedor | Media | Mediana | S | Min | Max | Outliers | n |
|--------------|-------|---------|-------|-----|-----|----------|------|
| Não | 4.97 | 2.6 | 10.61 | 1 | 211 | 0.09 | 828 |
| Sim | 4.47 | 2.2 | 9.46 | 1 | 180 | 0.10 | 1812 |

Pela tabela, podemos observar um comportamento muito parecido em quase todos os níveis, apesar da quantidade de bilionários empreendedores ser muito maior. Da mesma maneira, temos os bilionarios não empreendedores com maior valor máximo de patrimônio acumulado.

4. Relação entre os bilionários e os países de residência

O mapa a seguir, que apresenta a quantidade de bilionários em cada país analisado, pode ser utilizado para melhor compreender a distribuição de bilionários ao redor do mundo.

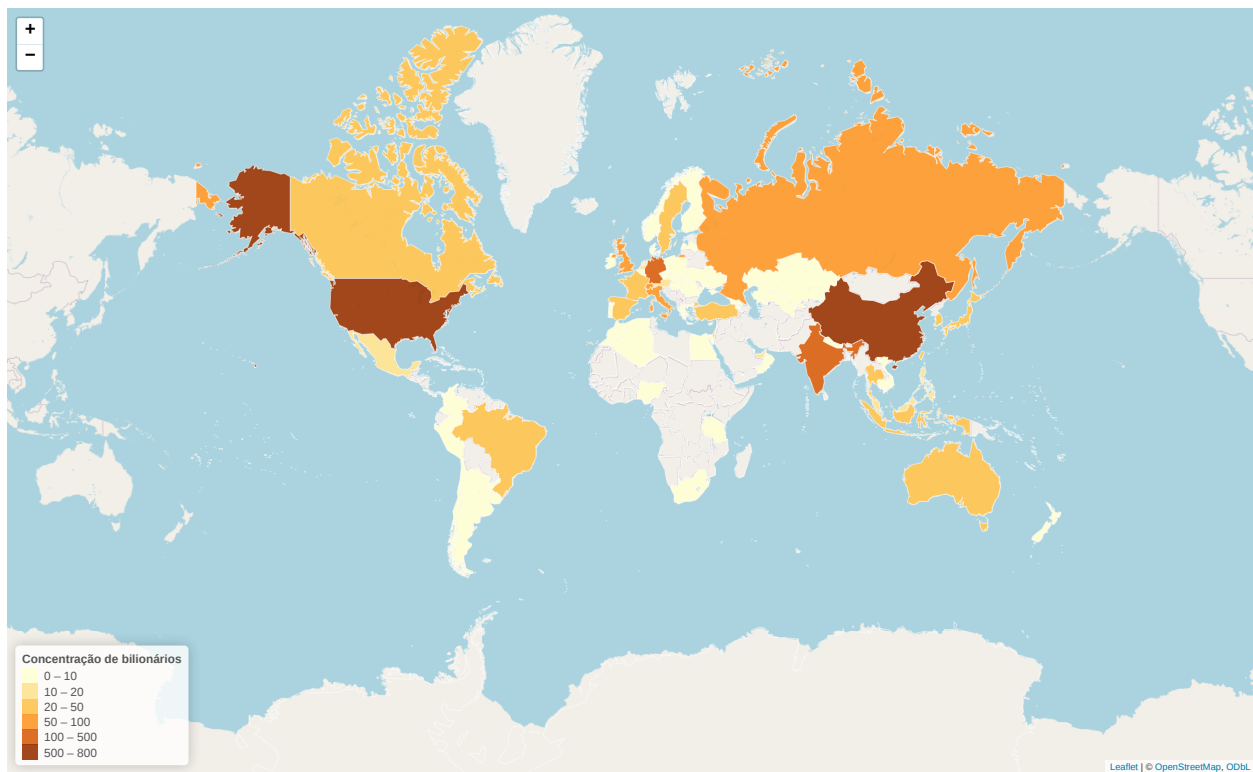


Figure 5: Mapa da distribuição de bilionários no mundo

[Link de uma versão interativa](#)

Table 5: Quinze países com maior número de bilionários

| País | Número de bilionários |
|----------------|-----------------------|
| United States | 754 |
| China | 523 |
| India | 157 |
| Germany | 102 |
| United Kingdom | 82 |
| Russia | 79 |
| Switzerland | 78 |
| Hong Kong | 68 |
| Italy | 55 |
| Singapore | 46 |
| Brazil | 44 |
| Australia | 43 |
| Taiwan | 43 |
| Canada | 42 |
| Japan | 38 |

Com base no mapa e na tabela acima, observa-se que os países com maior número de bilionários são os Estados Unidos, a China e a Índia, de modo que concentram 54,24% dos bilionários do mundo. Ademais, a quantidade de bilionários nos Estados Unidos e na China é significativamente maior do que nos demais países.

5. Relação entre os bilionários e os continentes de residência/natalidade

Ademais, para a análise da quantidade de bilionários por continente, foram feitos os seguintes gráficos:

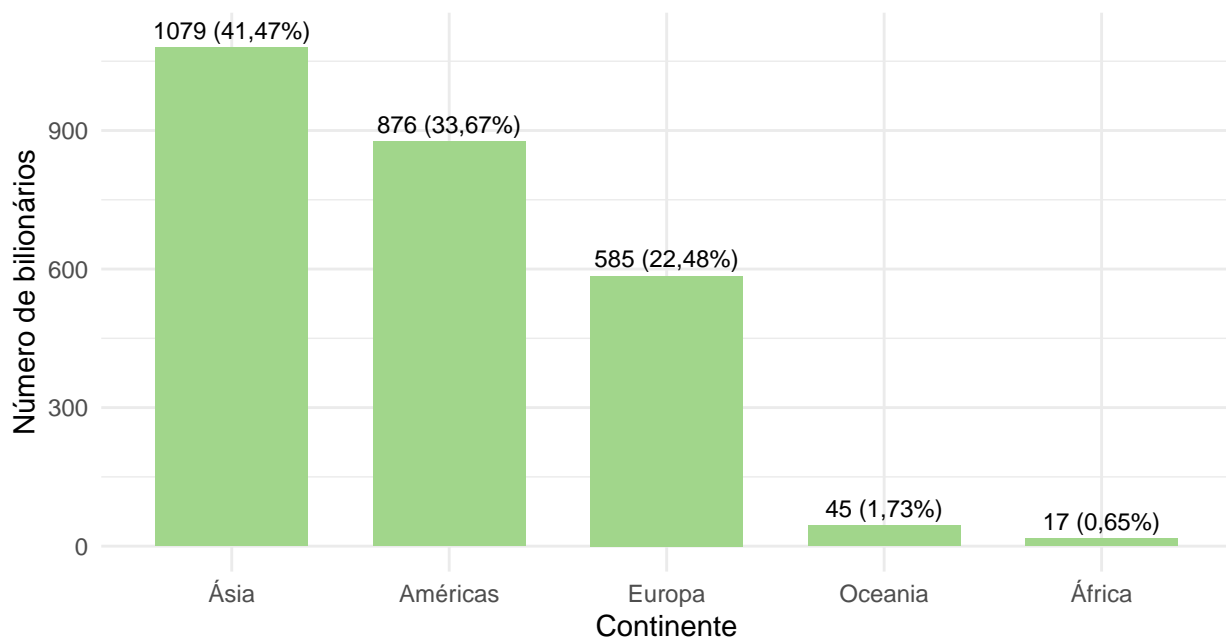


Figure 6: Gráfico de barras do número de bilionários por continente

[Link de uma versão mais completa do gráfico sankey](#)

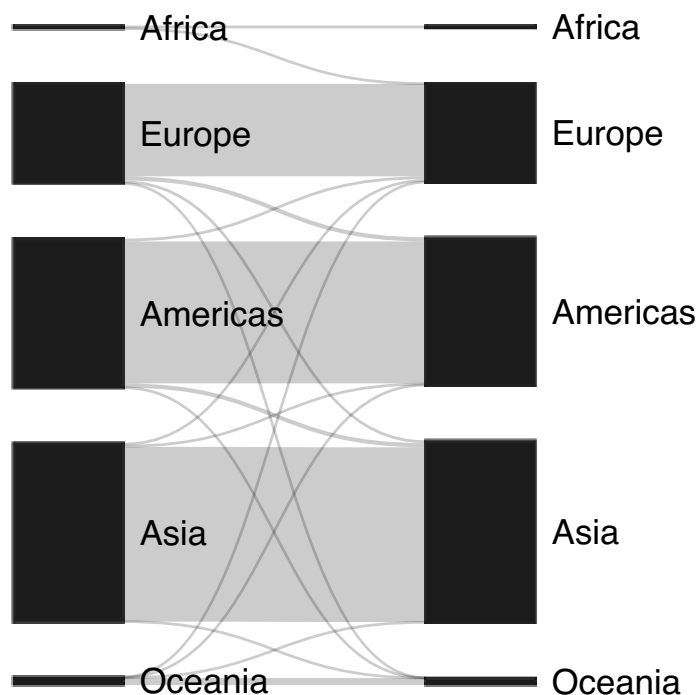


Figure 7: Gráfico sankey da migração da residência dos bilionários

A partir da análise do gráfico de barras, é possível perceber que o continente com maior número de bilionários é a Ásia, com 41,47% do total de bilionários, o que reflete a presença de países com grande concentração de bilionários, como a China e a Índia. Ademais, a África é o país com menor quantidade de bilionários.

Além disso, o gráfico Sankey apresentado mostra o fluxo de bilionários de seu continente de cidadania para o continente que residem atualmente. Com base nele, nota-se pouca migração de bilionários no sentido intercontinental. Os maiores fluxos observados foram da Europa para as Américas e das Américas para a Ásia, ambos de 25 pessoas.

parte do leo

papapapa

Discussão

Pode-se concluir pelos resultados que as análises são insuficientes para mostrar o comportamento de cada variável. As ferramentas disponíveis nos dão um parecer generalizado dos dados que não podem ser extravasados para a população e nem confirmados. Apesar dos coeficientes de variação não terem indicado associação entre a maioria das variáveis estudadas, só haveria a possibilidade de confirmar o comportamento, realizando inferências. Desse modo, as maiores complicações foram acerca da limitação de ferramentas disponíveis para analisar a fundo a existência de associações no banco de dados e em como proceder com a análise exploratória, a partir dos resultados encontrados. Apesar disso tudo, é crucial entender a importância do estudo sobre os bilionários, para que, mesmo de maneira geral, as pessoas entendam como os dados podem se comportar, mesmo que não haja a possibilidade de se fazer a previsão.

Referências bibliográficas

Extração dos dados:

<https://www.linkedin.com/pulse/para-que-serve-um-bilion%C3%A1rio-maite-schneider>

<https://www.forbes.com/real-time-billionaires/#2e5847b53d78>

<https://www.kaggle.com/code/achmadryanfauzi/bilionares-statistic-2023-data-insight>

Medidas estatísticas:

BUSSAB, W. O.; MORETTIN, P. A. Estatística Básica. 9. ed. São Paulo: Pearson, 2017.

Participações

Júlia Lima Nosralla (232014825): Análises 4 e 5, revisão.

Gabriel Henrique Ovidio de Araújo (232014905): Análises 2,3 e relatório.

Leonardo Farias Saraiva (232014834): Análise 6, revisão.

Maria Luiza Moura de Resende Ferro Costa (232014880): Análise 1 e relatório.