

---

# Análise de dados sobre deputados

---

Lucas Guimarães Batista

27 de junho de 2021

## 1 INTRODUÇÃO

Uma conversa em tom resignado sobre os rumos futuros do país é tão presente na vida do brasileiro médio que poderia figurar ao lado do Curupira como integrante do folclore. Quem é que nunca ouviu um "isso aqui não tem jeito" para se referir a melhorias na sociedade, não é mesmo? Por um lado, ouvimos essas frases prontas ditas em tom de desaprovação muitas vezes a respeito de decisões políticas tomadas por representantes eleitos. Por outro lado, parece bem aceita a ideia de que a educação e outras formas de investimentos de longo prazo são uma solução para problemas relacionados a representatividade política. Essa proposição parte do princípio de que a sociedade elege aqueles cujos valores mais ressoam consigo, de forma que a representação política é então um espelho da sociedade que a produz. Quis então tratar essas ideias como hipótese e buscar uma forma de testar a veracidade delas com auxílio dos dados disponíveis na Base dos Dados. Tentarei achar relações entre dados que nos permitam responder as seguintes perguntas:

- Quais tendências são manifestadas via voto?
- Há relação entre escolaridade dos representantes eleitos e fatores socio-econômicos da população que o elegeu?
- Se as tendências de voto atuais se mantiverem, como é o cenário futuro?

## 2 DADOS UTILIZADOS

Para fazer a análise, foram utilizados os dados de **Atividade Legislativa** e do **Atlas de Desenvolvimento Humano**. Foram utilizadas as seguintes colunas de cada uma das bases de dados:

Atividade Legislativa	Atlas do Desenvolvimento Humano
sexo	ano
escolaridade_nova	taxa_fundamental_18_mais
ultima_legislatura	taxa_medio_18_mais
	taxa_superior_25_mais

## 3 ANÁLISE

### 3.1 REPRESENTANTES ELEITOS

Começo fazendo uma análise exploratória sobre o dataset de Atividade Legislativa, em especial sobre a tabela relacionada aos deputados. As informações que imediatamente saltam aos olhos são a distribuição de gênero dos deputados e a taxa de escolaridade dos mesmos. Esses dois parâmetros foram utilizados e comparados ao longo das legislaturas desde a redemocratização do Brasil pós ditadura militar até hoje, ou seja, da 48ª legislatura até a 56ª legislatura, a atual. Isso nós dá uma noção da evolução desses dados ao longo do tempo.

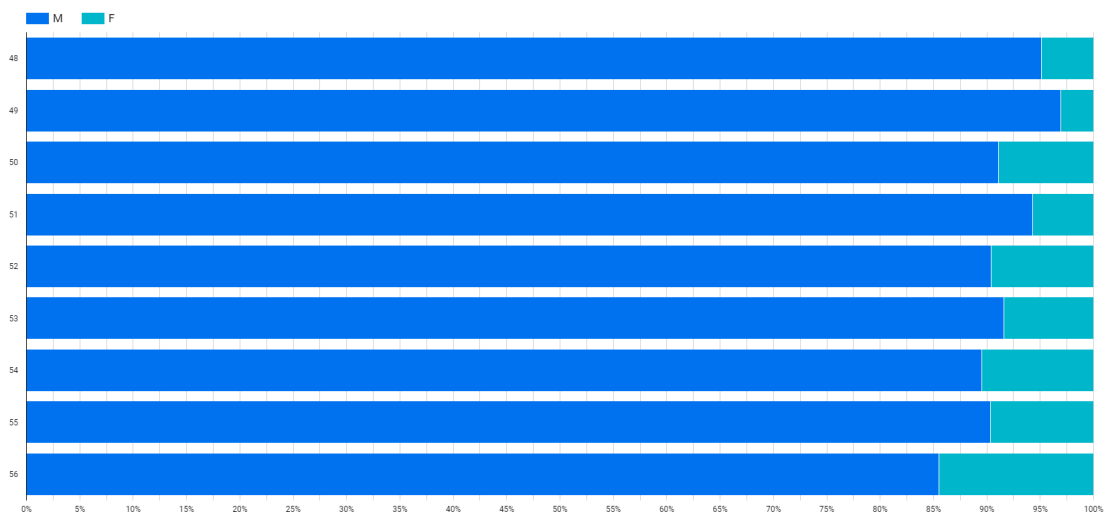


Figura 3.1: Distribuição dos deputados por sexo desde a redemocratização

A *figura 3.1* é uma visualização que mostra que com o passar do tempo mais deputadas mulheres foram eleitas quando levando em consideração a proporção total de eleitos. Isso pode indicar inúmeras coisas: um aumento na participação das mulheres como candidatas na política, uma busca por maior maior representatividade feminina, uma maior aceitação

a mulheres em cargos de poder por parte da sociedade. Possivelmente uma combinação de todos esses fatores, além de outros que nem notamos.

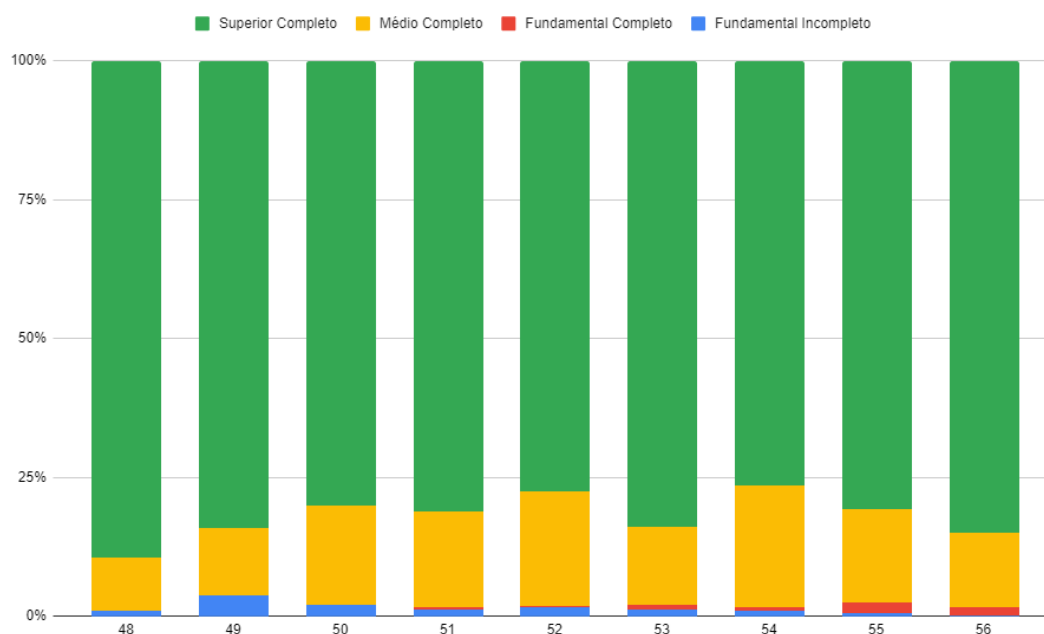


Figura 3.2: Escolaridade dos deputados por legislatura

A *figura 3.2* por si só não traz nenhuma interpretação clara além do mais trivial: deputados federais eleitos têm, em sua vasta maioria, ensino médio completo, com sempre no mínimo 75% deles tendo ensino superior completo. Se movermos nossa atenção para a *figura 3.3* conseguimos extrair um pouco mais de informação. Vemos que atualmente a taxa da população com mais de 25 anos que completou o ensino superior é de menos de 12%, um valor que está muito mais próximo do percentual de deputados federais que possuem até o ensino médio como maior grau de formação educacional. Quando olhamos para as décadas anteriores a disparidade é ainda maior, o que pode apontar para até duas direções. Uma delas é contrária a ideia de que a população elege aqueles com quem ela mais se identifica, a outra é avaliarmos se a comparação de grau de escolaridade entre deputados e a população do país é realmente uma aproximação boa. Embora elas não sejam mutuamente excludentes, estou mais inclinado a acreditar que a única verdadeira é a segunda afirmação. Não acredito que o fato de uma população majoritariamente sem ensino superior completo eleger representantes com ensino superior seja uma razão remotamente válida para acreditar que os eleitores não estão buscando eleger quem eles acreditam que melhor os representam.

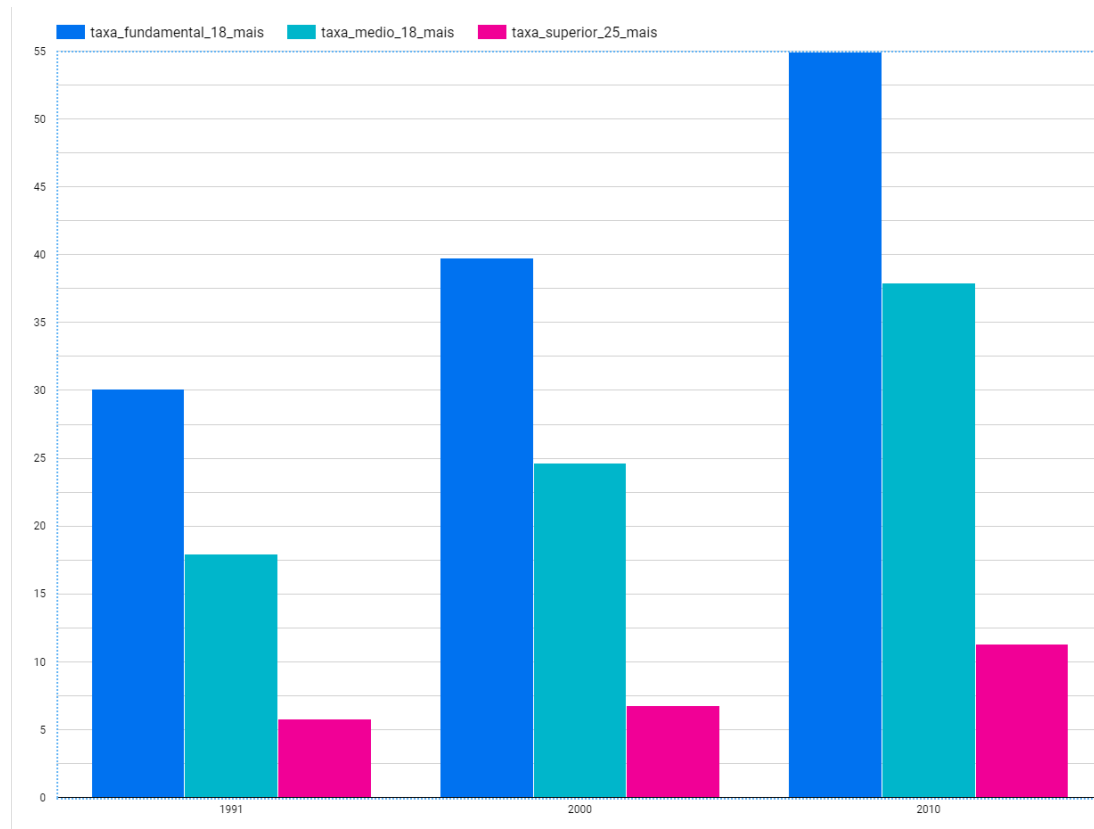


Figura 3.3: Evolução da escolaridade no Brasil

## 4 PREDIÇÃO

Utilizar dados com um enfoque político foi uma escolha feita em função da curiosidade de testar as afirmações de senso comum que costumam ser proferidas pelo povo. Tentei relacionar os poucos dados sobre eleições disponíveis na Base dos Dados com outros que fossem relevantes mantendo o senso crítico e honestidade. Por conta disso, utilizar um modelo de aprendizado de máquina para fazer uma previsão de algum indicador me pareceu algo não só nada trivial como também algo que seria mais prova de técnica do que de conteúdo. Após analisar os inúmeros fatores que implicam em resultados de eleições, acho que seria leviano da minha parte fazer qualquer afirmação tirada de um modelo que tentasse relacionar escolaridade de políticos eleitos com a de seus eleitores. Em função disso, tentei fazer uma previsão que, embora muito mais simples do ponto de vista técnico, parece apresentar um dado muito mais relevante e presente nos debates atuais. Com base na evolução percentual das deputadas mulheres eleitas desde o fim da ditadura militar no Brasil mostrada na *Figura 3.1*, supus um caráter de crescimento linear e implementei um modelo de regressão linear para prever quando que a distribuição de deputados seria igual entre homens e mulheres. Os resultados são mostrados no gráfico a seguir e posteriormente discutidos:

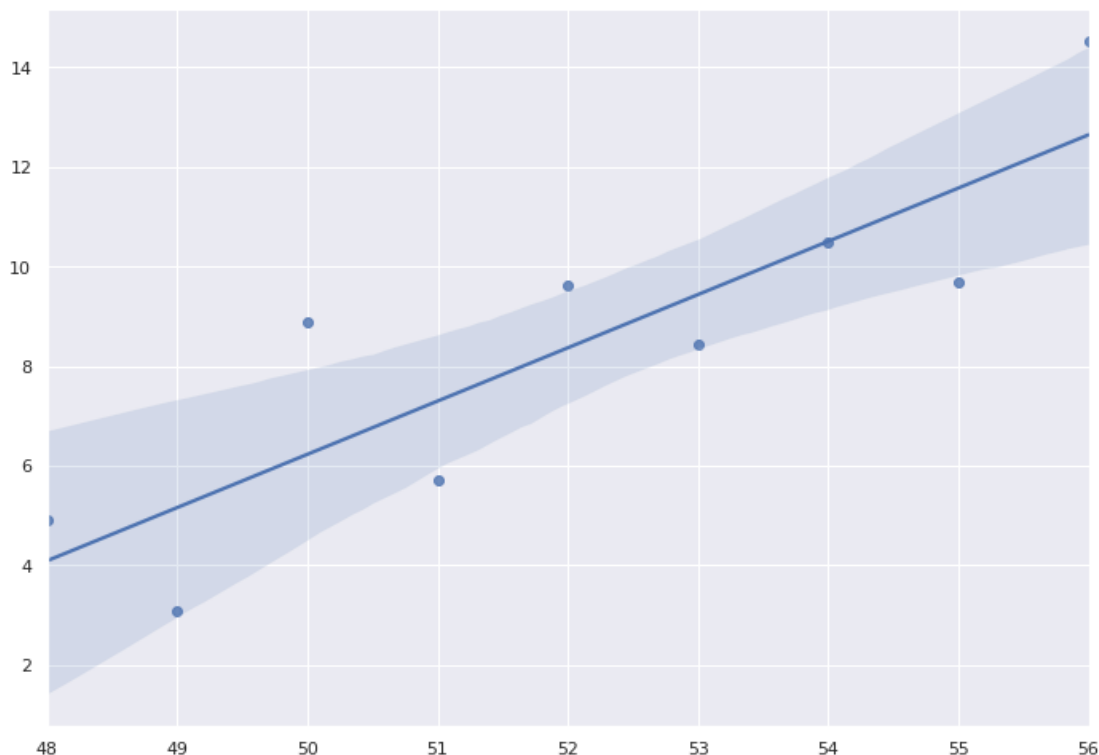


Figura 4.1: Regressão linear para quantidade de deputadas mulheres

A reta gerada tem coeficiente angular  $m = 1.0693$  e coeficiente linear  $n = -47.2386$ . Utilizando-a como aproximação para a tendência crescente de mulheres na política, temos que a participação na câmara será igual entre homens e mulheres na 91ª legislatura. Seguindo as normas eleitorais em vigência, essa legislatura teria início no ano de 2163. Com esse resultado, cabe então analisar criticamente sua validade e suas implicações. Primeiro, parece razoável acreditar que a participação feminina na política seguirá aumentando ao longo do tempo. É uma tendência vista em boa parte dos setores em que seus números são muito menores que os números masculinos. Por outro lado, inúmeros fatores afetam a participação feminina na política, então tentar prever quando e como essa participação aumentará é uma tarefa difícil. A própria decisão arbitrária de tentar prever quando essa divisão chegará em 50% soa forçada. A distribuição entre os sexos no Brasil é de aproximadamente 50%, então poderíamos supor que é o valor esperado, mas no mundo real há muito mais exemplos em que essa distribuição não é respeitada do que exemplos em que ela é. Há também a discussão sobre o valor ridículo que foi encontrado. Mais de 140 anos para ter uma distribuição igual entre os sexos na câmara dos deputados não parece nada razoável, o que pode se dar por algumas razões. A primeira é que muito provavelmente esse crescimento seguirá uma lógica polinomial ou exponencial em vez de uma lógica linear. Além disso, a baixa quantidade de dados disponíveis graças ao período escolhido (de 1985 ao presente) traz consigo uma fragilidade muito grande, tornando qualquer afirmação ou previsão baseada neles bastante suscetível a erros.

## 5 CONCLUSÃO E PRÓXIMOS PASSOS

No fim das contas o trabalho foi um grande aprendizado, tanto sobre trabalhar com dados como sobre os dados em questão. Ter uma ferramenta como o BigQuery a disposição é algo novo para mim e que facilitou imensamente o trabalho de repartir somente os dados relevantes para o projeto em questão. Quanto as perguntas levantadas inicialmente, sinto que consegui achar algumas respostas e trazer um debate sobre as que tangem tendências atuais e futuras de voto. Já sobre a relação entre escolaridade de deputados e seus eleitores, não foi possível fazer essa relação de forma satisfatória, possivelmente mostrando que é uma relação fraca ou não existente. Acho que a minha escolha de focar especificamente nesse conjunto de dados sobre deputados não foi boa porque acabou limitando demais o escopo da análise que eu fui capaz de fazer.

Gostaria, porém, de ter feito um trabalho muito mais aprofundado, levando mais fatores em consideração na minha previsão e de ter gerado visualizações mais interessantes, mas por questões pessoais que me consumiram muito mais tempo que esperado não pude fazer. Dentre as ideias que tive e eu gostaria de ter feito posso listar:

- Gerar uma visualização com o mapa do Brasil separando os dados de escolaridade por estado em vez de generalizá-los para o país todo.
- Buscar fatores mais relevantes que escolaridade para aproximar o grau de representatividade.
- Realizar uma predição mais elaborada, utilizando um modelo de aprendizado de máquina e com algum alvo mais interessante, mas isso se mostrou muito mais difícil do que eu esperava especificamente com esse dataset de deputados.
- Criar um *scraper* que coleta todas as fotos associadas aos deputados de cada legislatura (disponíveis via url no dataset utilizado) e gerar uma imagem do "deputado médio" com auxílio de IA.