

## **Assignment 02: Análise Exploratória de Dados**

Visualização de Dados

Felipe Marques Esteves Lamarca

### **Introdução: Análise de Eleições Legislativas no estado do Rio de Janeiro**

Neste trabalho, proponho uma análise exploratória das Eleições Legislativas realizadas no estado do Rio de Janeiro entre 2008 e 2022, isto é, tanto aquelas que ocorreram em âmbito federal (para eleição de deputados federais e estaduais), como no âmbito municipal (para a eleição de vereadores). Ao longo da exploração dos dados, vamos observar algumas informações importantes que dizem respeito ao sistema político fluminense, em particular, mas também brasileiro, de forma geral: a dificuldade de acesso de minorias a cargos políticos, a importância das campanhas políticas, as limitações de sistemas de eleições proporcionais e outros aspectos.

Algumas questões prévias orientam a análise dos dados. São elas:

**Questão 1:** Existe alguma relação explícita entre as despesas de campanha de um candidato e o número de votos que ele recebeu na eleição?

**Questão 2:** Em sistemas de eleição proporcionais, uma das discussões mais importantes é a respeito do fato de que o número de votos que um partido recebe não se reflete, necessariamente, no número de cadeiras que conquista. De que forma isso afeta o caso fluminense?

**Questão 3:** Houve alguma mudança significativa no número de cadeiras na Câmara dos Deputados obtidas por diferentes partidos ao longo do tempo? Em particular, é possível perceber alguma movimentação do eleitorado entre os espectros políticos no contexto de embate entre bolsonarismo e petismo?

**Questão 4:** Existe alguma tendência de filiação partidária a algum determinado partido de acordo com a renda do candidato? Ou seja, candidatos com os maiores valores declarados em bens são mais propensos a disputar as eleições por um determinado partido?

**Questão 5:** Quais cargos políticos são mais facilmente acessados por minorias representativas? Ou seja, quais cargos políticos são mais propensos a serem ocupados por candidatas mulheres ou pessoas não-brancas, por exemplo?

## O dataset

Para tentar responder estas questões, utilizei uma composição de *datasets* acessados através da Base dos Dados. Especificamente, a base de dados referente às eleições brasileiras é composta por uma série de tabelas, dentre as quais utilizei: *candidatos*, *resultados\_candidatos* e *bens\_candidatos*. Dessa forma, foi possível construir um *dataset* final contendo informações sobre todos os candidatos que disputaram eleições legislativas no estado do Rio de Janeiro entre 2008 e 2022. Abaixo, apresento as primeiras 5 linhas do *dataset* utilizado:

	ano	id_candidato_bd	nome_urna	sigla_partido	cargo	genero	raca	resultado	votos	bens_declarados	despesas_campanha	tipo_eleicao	new_partido
0	2022	1244391	Daniela Do Waguinho	UNIÃO	deputado federal	feminino	branca	eleito por qp	213706	733291.33	3080522.00	federal	UNIÃO
1	2022	596743	General Pazuello	PL	deputado federal	masculino	branca	eleito por qp	205324	1288362.00	944065.37	federal	PR/PL
2	2022	1367243	Taliria Petrone	PSOL	deputado federal	feminino	preta	eleito por qp	198548	19498.36	1575937.20	federal	PSOL
3	2022	723988	Doutor Luizinho	PP	deputado federal	masculino	branca	eleito por qp	190071	3579347.95	1248760.96	federal	PP
4	2022	1278961	Márcio Canella	UNIÃO	deputado estadual	masculino	branca	eleito por qp	181274	3766549.57	1260269.21	federal	UNIÃO

Figura 1 – *Dataset* utilizado

As colunas *tipo\_eleicao* e *new\_partido* são uma primeira manipulação dos dados e foram incluídas programaticamente. O significado da primeira é autoexplicativo: conforme o ano de realização da eleição, determinei se tratava-se de uma eleição federal (para eleição de deputados federais e estaduais) ou de uma eleição municipal (para eleição de vereadores). A segunda coluna foi adicionada por conta de uma característica do sistema político brasileiro: partidos mudam de nome, somem ou se fundem ao longo do tempo. Por isso, em análises de longo prazo onde comparamos duas ou mais eleições, é necessário ter em mente que dois partidos podem, na verdade, ser o mesmo. É o caso do Partido Liberal (PL), na linha de índice 1 da figura, que já foi chamado de Partido da República (PR).

## Data Wrangling

Após os dados terem sido baixados da Base dos Dados, o primeiro passo foi unir tabelas para obter informações importantes. Por exemplo, para desenvolver a **Questão 5**, utilizei dados de geometria que foram unidos à tabela principal a partir do id do município. Essa coluna não consta na tabela anterior por ter sido utilizada na resposta de uma questão específica, que surgiu a partir da análise principal.

Além das duas novas colunas inseridas na tabela, conforme anunciado na seção anterior, uma falha nos dados públicos requiriu um terceiro ajuste. O TSE não coletava a raça dos candidatos antes de 2014, de modo que as análises nesse recorte seriam limitadas. No entanto, sabendo que é comum que candidatos concorram a eleições mais

de uma vez, adicionei a informação de raça a alguns candidatos que concorreram antes de 2014. Naturalmente, isso só era possível para os candidatos que concorreram novamente a partir de 2014.

Além disso, utilizei uma série de métodos de manipulação de dados ao longo da produção das visualizações. Em todos os gráficos foi usado pelo menos um método de agregações simples como média, contagem e pontos de máximo/mínimo, e os gráficos que respondem à **Questão 5** precisaram de uma pivotagem nas tabelas para o cálculo dos percentuais.

## Análise Exploratória

Dadas as explicações iniciais, podemos proceder à análise exploratória de fato. Como a unidade de análise é primordialmente o candidato, é importante observarmos como o número de candidatos se distribuiu ao longo do tempo. O gráfico abaixo dá conta dessa informação:



A primeira informação importante é que eleições federais e eleições municipais diferem em alta escala no número de candidatos disputando eleições — nas municipais, o número de candidatos é muito maior. Naturalmente, isso deverá ser levado em conta nos gráficos a seguir. Uma hipótese para a essa diferença é que, como candidatos a vereador são votados apenas no âmbito da cidade, as campanhas são provavelmente mais baratas e o incentivo à participação na disputa é maior. Além disso, o número de cargos para vereador é muito maior, já que cada cidade possui sua Câmara Municipal. A ALERJ, por outro lado, dispõe de pouco mais de 40 cadeiras, ao mesmo tempo que a bancada fluminense na Câmara dos Deputados possui 46 assentos. Outra informação que a visualização traz é que o número de candidatos em eleições para vereador cresceu bastante entre 2008 e 2020.

Para testar a hipótese que o gráfico ajuda a levantar, vamos à primeira questão:

1. Existe alguma relação explícita entre as despesas de campanha de um candidato e o número de votos que ele recebeu na eleição?

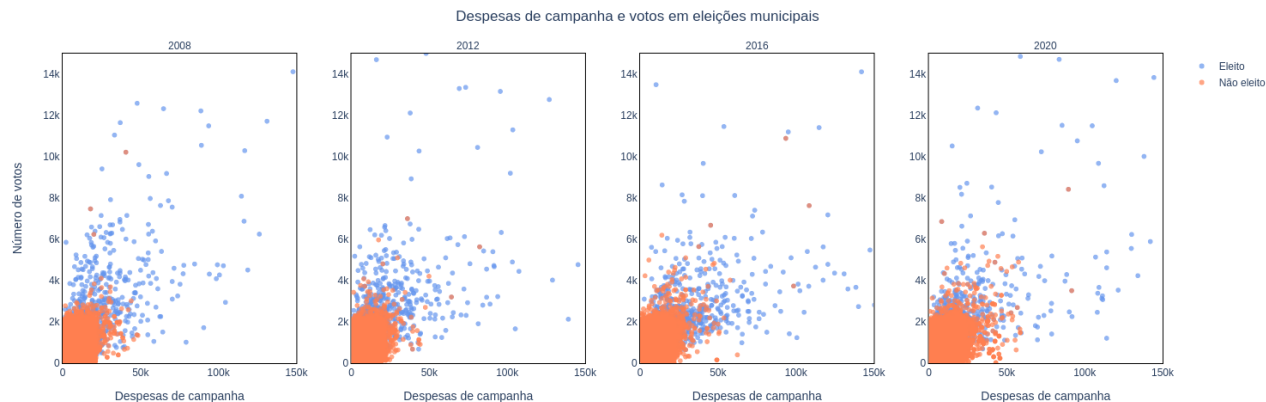
Segundo a literatura em Ciência Política, vários fatores influem no resultado da campanha de um candidato: sua raça, sua idade, seu gênero, seu partido, o valor investido em campanha etc. O *dataset* utilizado permite testar a hipótese dos gastos de campanha, já que possuímos uma coluna com essa informação e outra que indica se o candidato foi eleito ou não. Vamos observar, primeiro, o caso das eleições federais:



Primeiro, é importante comentar a respeito de algumas decisões sobre esse gráfico e os próximos relacionados ao mesmo assunto. Apesar de haver campanhas muito mais caras que R\$ 800 mil, optou-se por limitar a escala do eixo  $x$  para possibilitar uma análise da área do gráfico onde os pontos estão mais concentrados. O mesmo vale para o eixo  $y$  que, no caso, mapeia o número de votos. Outra decisão importante está relacionada ao fato de que há uma massa de candidatos não eleitos: adicionou-se um certo grau de opacidade aos pontos para que se tornasse possível observar os pontos onde há *overplotting*.

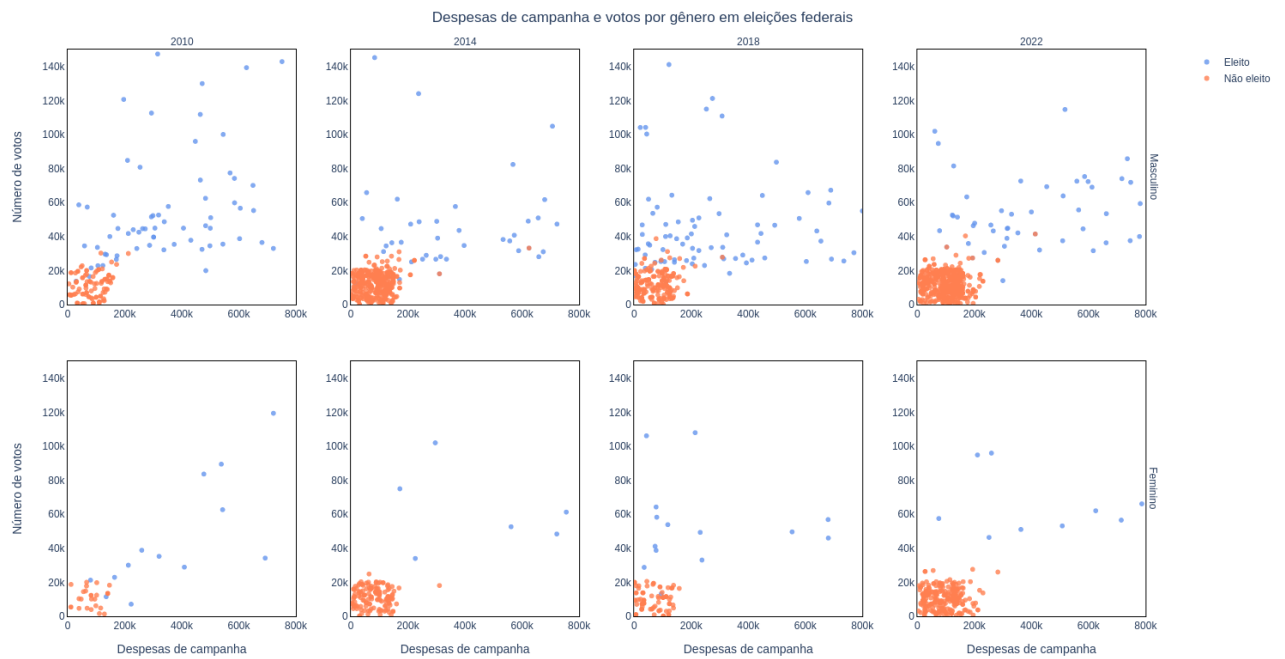
No mais, a análise sugere que, de fato, os gastos de campanha desempenham um papel fundamental no resultado do candidato. Em particular, observe que os candidatos que receberam mais votos, no geral, também são aqueles que tiveram mais gastos em comparação aos candidatos não eleitos. Ainda assim, há aqueles candidatos que tiveram altos gastos de campanha, mas não tiveram a correspondência em número de votos. Há também aqueles que, com poucos recursos de campanha, conseguiram um número alto de votos mas não foram eleitos<sup>1</sup>. Vale a pena observar se esse padrão também é observado no caso das eleições municipais:

<sup>1</sup> Essa característica do sistema eleitoral brasileiro será explorada mais adiante.

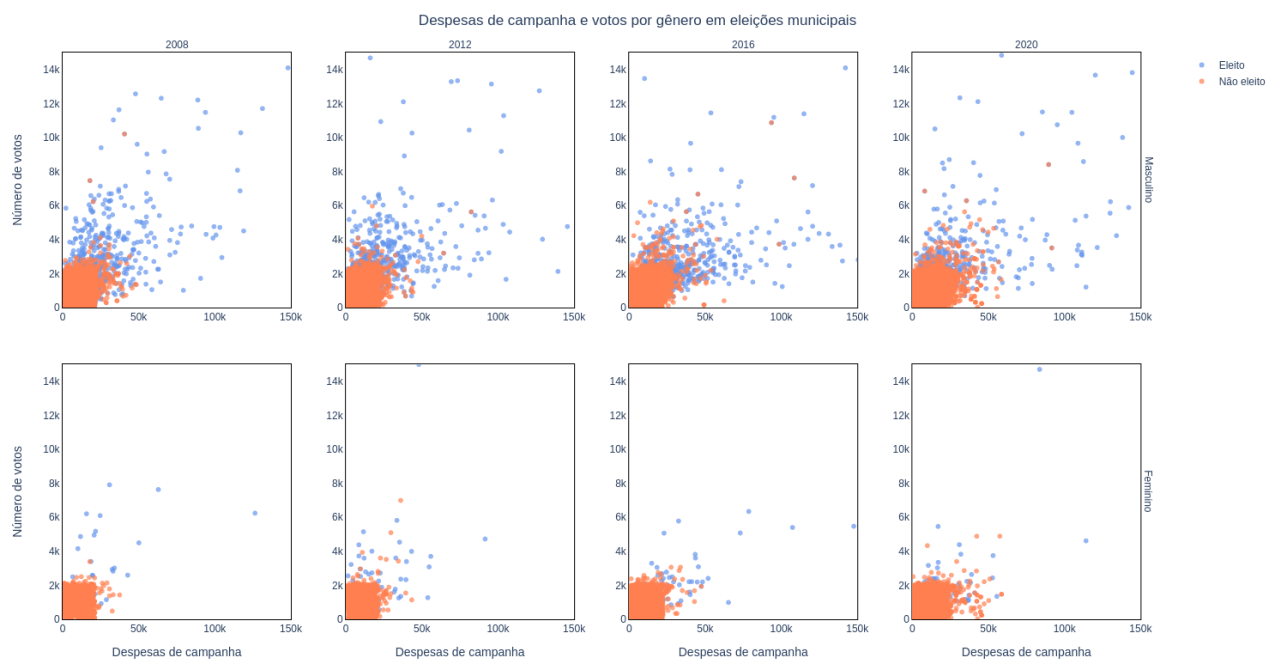


Notavelmente, observamos aqui um número ainda maior de candidatos. Isso é esperado, conforme comentado na introdução deste trabalho. No entanto, vale notar que, diferente das eleições federais, o efeito das despesas de campanha sobre o número de votos é menos direto — há uma quantidade razoavelmente grande de candidatos que conseguiram uma quantidade bastante relevante de votos sem que isso demandasse altos gastos de campanha. Além disso, observe que, de fato, eleições municipais são muito mais baratas que eleições para cargos em instâncias maiores.

Uma questão que permeia esses dados, mas que não é diretamente tratada pela pergunta inicial, é se há diferenças nessa relação quando separamos os dados pelo gênero do candidato. Em outras palavras, podemos levantar a hipótese de que, quando a candidata é mulher, é necessário que ela gaste, em geral, valores muito altos durante a campanha para ter uma correspondência em votos. Como vimos em um gráfico anterior, isso não é necessariamente verdade para todos os candidatos, já que alguns conseguem, por exemplo, se eleger com gastos menores. Vejamos para as eleições federais:



Nesse caso, uma diferença relevante que se observa é que a quantidade de mulheres candidatas é bem menor em relação aos homens, uma característica sintomática de um sistema que deixa de fora boa parte da população feminina. Essa diferença no número de candidatos dificulta o teste da hipótese levantada pela questão anterior, mas note, ainda assim, que mulheres parecem receber menos votos que os homens, observando atomicamente. O gráfico a seguir é um pouco mais informativo a esse respeito:



Se a hipótese inicial previa que mulheres precisariam gastar mais para obter mais

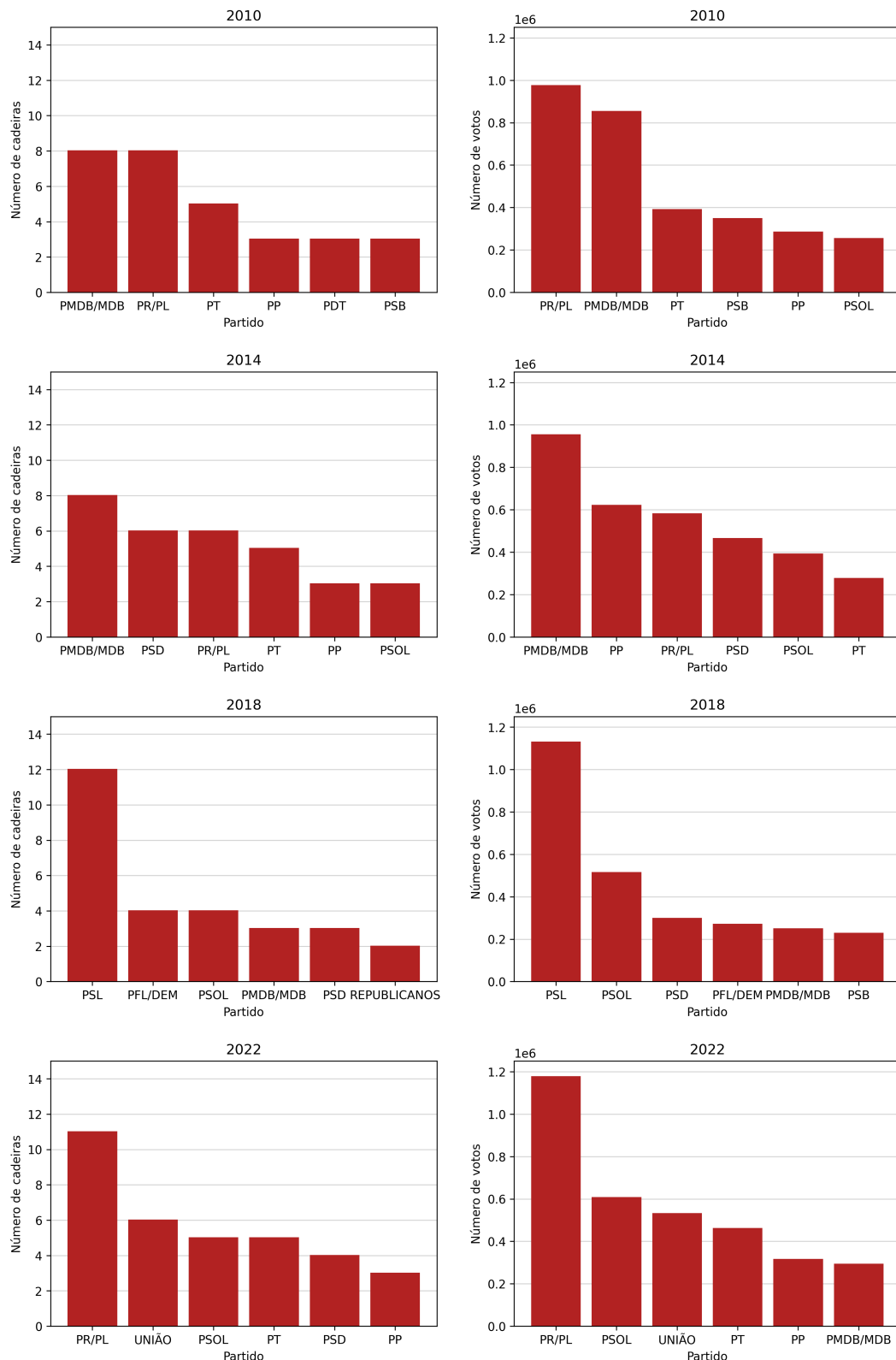
votos, os anos de 2008 e 2012 contrariam essa ideia. Em particular, observamos mulheres que fizeram campanhas relativamente baratas e, mesmo assim, tiveram boa correspondência em número de votos. Essa análise é mais limitada no ano de 2012, quando é possível observar que uma candidata obteve mais de 14 mil votos e, dessa forma, pode ter puxado candidatos com votações menores. Para 2016 e 2020, por outro lado, a hipótese se confirma. Note que há mulheres com gastos de campanha altíssimos, mas com baixa correspondência no número de votos. Enquanto isso, homens com campanhas extremamente baratas conseguem obter muito mais votos com alta frequência.

2. Em sistemas de eleição proporcionais, uma das discussões mais importantes é a respeito do fato de que o número de votos que um partido recebe não se reflete, necessariamente, no número de cadeiras que conquista. De que forma isso afeta o caso fluminense?

Vamos contemplar, agora, a segunda questão, que explicita uma das discussões mais debatidas pela Ciência Política. Mesmo que um partido consiga obter mais votos que outro, isso não é garantia de que ele obterá mais cadeiras na Câmara Legislativa em questão. Isso porque, mesmo que as eleições sejam proporcionais, características específicas do sistema eleitoral podem causar distorções. Era o caso, por exemplo, das coligações para eleições proporcionais, que foram proibidas recentemente.

Observe o gráfico abaixo, onde os gráficos da esquerda dizem respeito ao número de cadeiras que o partido fez na bancada fluminense na Câmara dos Deputados, e os da direita informam o número absoluto de votos recebidos pelo partido. Note que os eixos  $x$  são independentes, refletindo justamente o fato de que o sistema eleitoral produz inconsistências:

### O número de votos reflete o número de cadeiras?



Para o ano de 2010, por exemplo, note que PL e PMDB obtiveram o mesmo número de cadeiras, apesar de o PL ter obtido mais votos. No mesmo ano, o PDT esteve entre os 6 partidos com mais cadeiras, mas nem sequer esteve entre os 6 partidos com mais votos. Em 2018, o PSOL obteve quase o dobro de votos do DEM, mas os dois

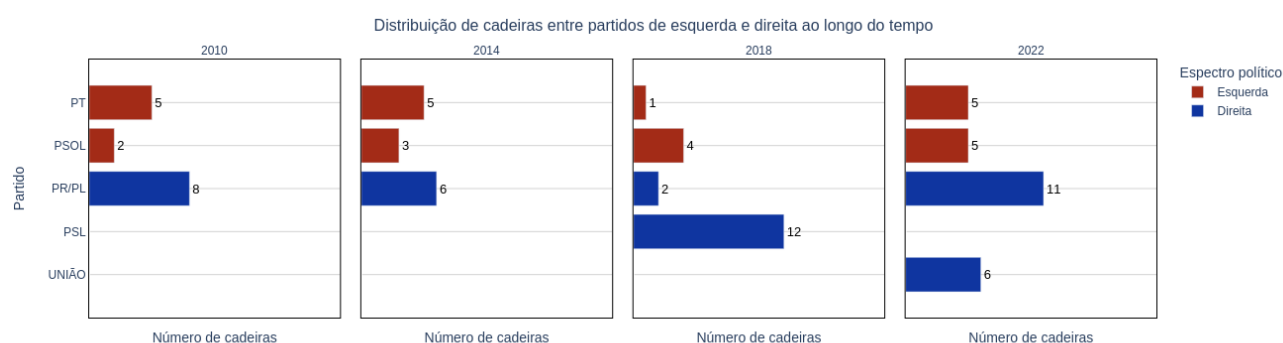


partidos obtiveram o mesmo número de cadeiras na bancada fluminense na Câmara dos Deputados. No caso do estado do Rio de Janeiro em particular, essa característica do sistema parece ter favorecido partidos da direita em detrimento dos partidos de esquerda. Nas últimas eleições, PSOL fez mais votos que o União Brasil, mas este último obteve uma cadeira a mais no cômputo geral.

3. Houve alguma mudança significativa no número de cadeiras na Câmara dos deputados obtidas por diferentes partidos ao longo do tempo? Em particular, é possível perceber alguma movimentação do eleitorado entre os espectros políticos no contexto de embate entre bolsonarismo e petismo?

O estado do Rio de Janeiro é reconhecido por ser conservador do ponto de vista político. Não à toa, é o estado que elegeu Jair Bolsonaro como deputado uma série de vezes. O *dataset* em análise possibilita mapear de que forma o contexto geral da política brasileira de embate entre bolsonarismo e petismo se apresentou na bancada fluminense na Câmara.

Para avaliar isso, vamos analisar o número de deputados federais eleitos por 5 partidos, dois de esquerda (PT e PSOL) e três de direita (PSL, União Brasil e PL). Os partidos estão associados de alguma forma a Lula e Bolsonaro. Mais especificamente, os de esquerda são base de apoio de um, enquanto os de direita são base de apoio do outro. Vejamos:



Entre as eleições de 2010 e 2014, quando o surgimento de Bolsonaro como candidato viável à presidência ainda era longínquo, o PSOL aumentou sua bancada em um candidato e o PL elegeu 2 candidatos a menos. PSL e União Brasil nem existiam ainda. No entanto, após o *impeachment* de Dilma Rousseff e o crescimento do sentimento anti-petista na população fluminense, o PT elege apenas um candidato à bancada e PL e PSL compõem, juntos, 14 eleitos de um total de 46 cadeiras. E, nas últimas eleições, com a reeleição de Lula, PT e PSOL apresentaram aumento no número de eleitos, mas o PL

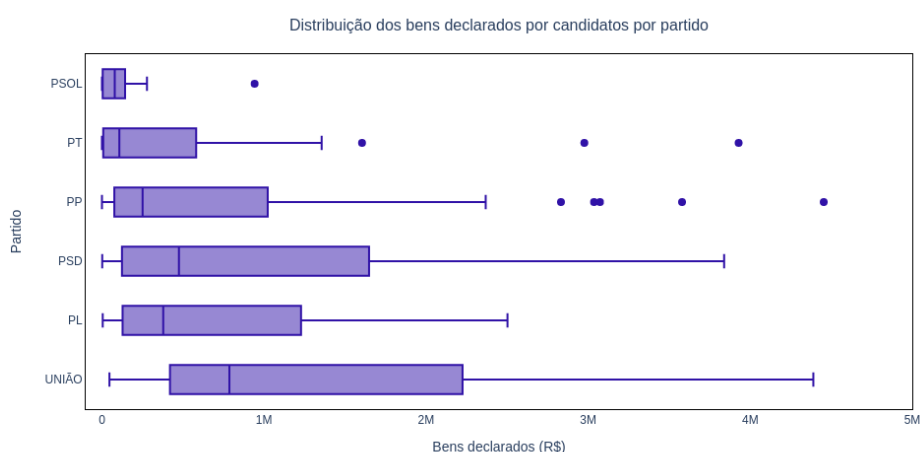
(partido de Bolsonaro), também retoma sua expressão. De forma geral, observa-se que a disputa nacional se afigura no contexto fluminense com certa vantagem para a direita.

A decisão por um gráfico de barras nesse formato se justifica pela facilidade com que é possível observar a variação no número de cadeiras de um partido de um ano para o outro. As cores foram adicionadas justamente para facilitar a apreensão do espectro político do partido em questão, possibilitando uma análise conjunta.

4. Existe alguma tendência de filiação partidária a algum determinado partido de acordo com a renda do candidato? Ou seja, candidatos com os maiores valores declarados em bens são mais propensos a disputar as eleições por um determinado partido?

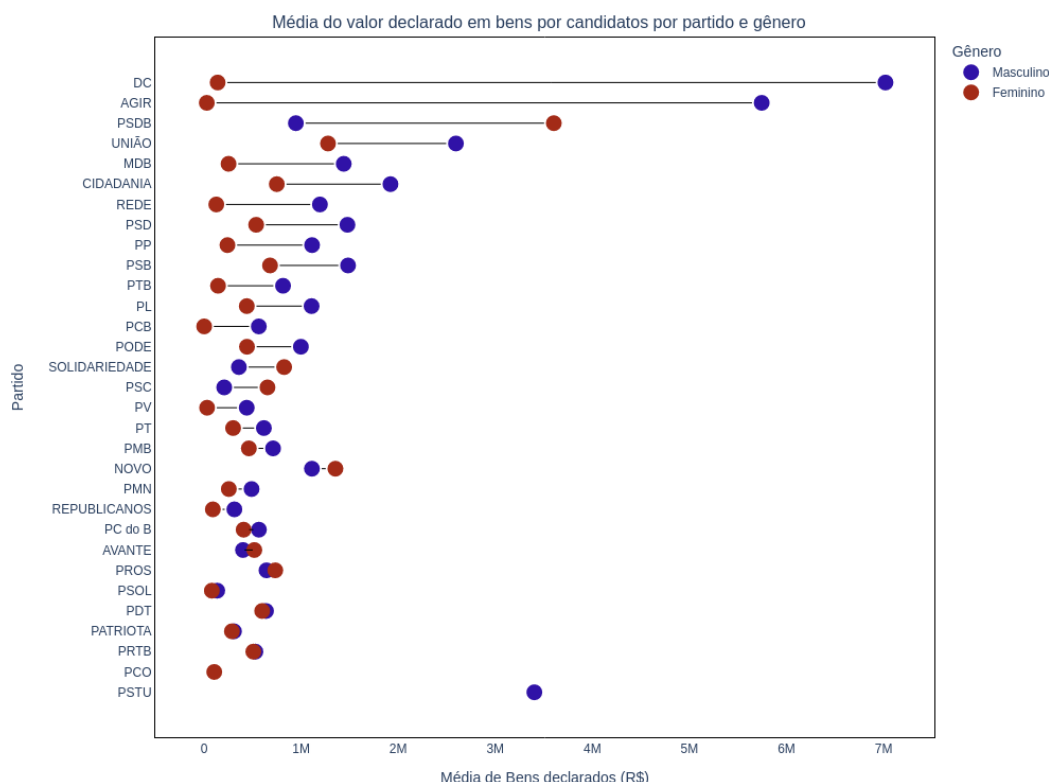
O senso comum faria imaginar que partidos de esquerda tendem a comportar mais candidatos de baixa renda, enquanto partidos de direita tendem a comportar candidatos de renda mais alta. Essa, no entanto, não é necessariamente uma associação verdadeira. Os partidos têm incentivos para aceitar grandes números de candidatos para, por exemplo, tentar ultrapassar a cláusula de barreira ou somar votos suficientes para fazer um número maior de cadeiras.

Mas, mesmo assim, será que pessoas mais ricas tendem a se filiar a partidos de algum espectro ideológico? Como o *dataset* possui informações sobre os bens declarados pelos candidatos, podemos testar essa hipótese. Em particular, vamos avaliar essa informação para os partidos que obtiveram mais votos em 2022, conforme um gráfico já apresentado.



De fato, assumindo que PSOL e PT são partidos de esquerda e que PP, PL, PSD e UNIÃO são partidos de direita, poderíamos de fato concluir que partidos mais à direita comportam mais candidatos ricos. Vale lembrar que esses dados se limitam às eleições para deputados federais no RJ em 2022.

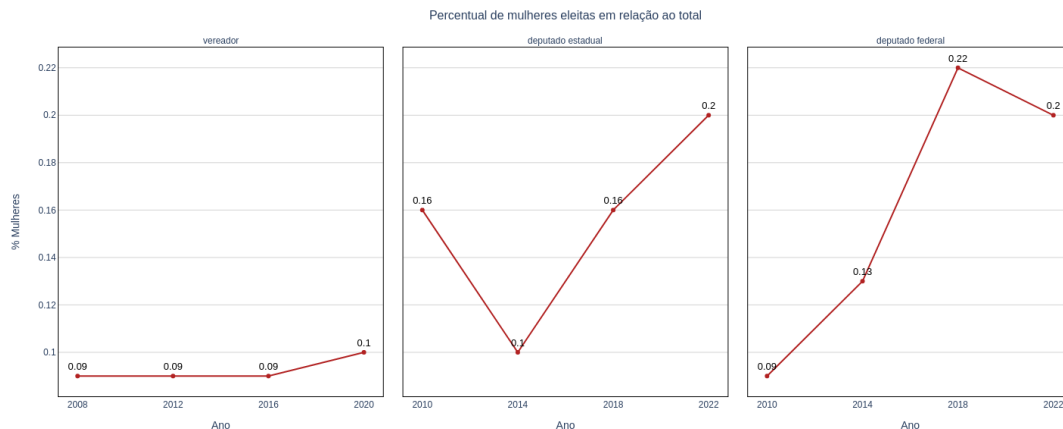
No entanto, esse gráfico não permite que observemos diferenças internas dos partidos no que diz respeito à distribuição da renda dos candidatos de acordo com características mais específicas. Essa é uma questão importante, porque ajuda a definir o perfil dos candidatos dos partidos, por exemplo. O *scatterplot* abaixo faz esse mapeamento no caso de gênero, também para o ano de 2022, onde cada ponto representa a média de bens declarados por homens e mulheres em cada partido:



A análise mais geral mostra que mulheres, na média, declaram valores menores em bens em relação aos homens. A maior diferença média de vantagem para os homens ocorre em partidos mais conservadores, como o Democrata Cristão, Agir e União Brasil. É curioso notar que a média é bem maior para as mulheres no PSDB, o que provavelmente se explica por conta de algum *outlier*.

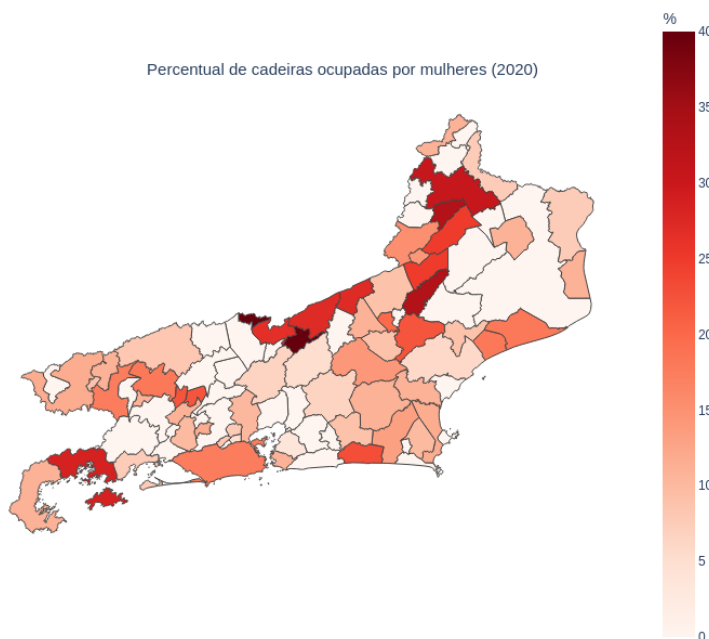
5. Quais cargos políticos são mais facilmente acessados por minorias representativas? Ou seja, quais cargos políticos são mais propensos a serem ocupados por candidatas mulheres ou pessoas não-brancas, por exemplo?

Para avaliar a questão do acesso às minorias, é importante observarmos o percentual em relação ao total de candidatos (eleitos) pertencentes a categorias minoritárias. Por exemplo, vejamos esses percentuais para cada cargo legislativo em relação ao gênero:



Segundo os dados, o cargo legislativo menos ocupado por mulheres é o de vereador. Essa informação é contraintuitiva na medida em que observamos anteriormente que havia mulheres que conseguiam realizar campanhas baratas e mesmo assim se eleger. No entanto, é provável que o número pequeno de cadeiras em Câmaras Municipais de algumas cidades dificulte o acesso de mulheres. Os cargos de deputado estadual e federal ainda são pouco acessados por mulheres, mas houve aumento nos últimos anos.

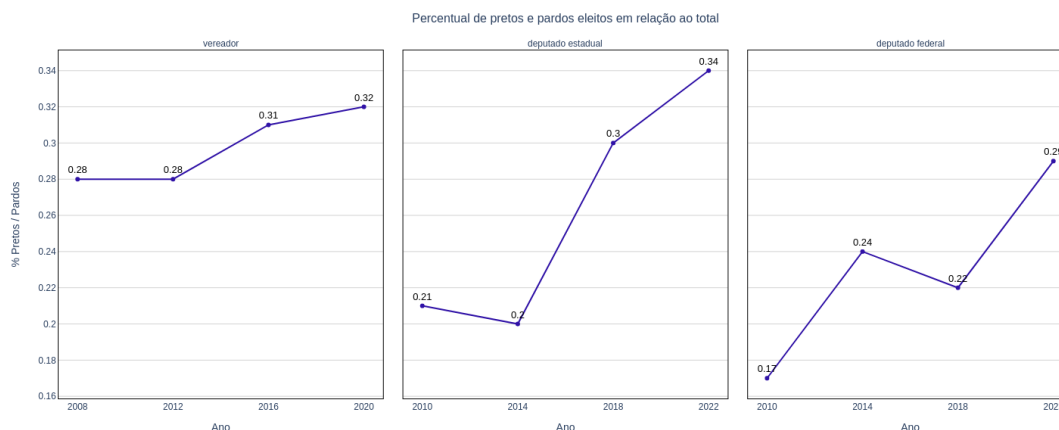
O baixo percentual das mulheres eleitas para vereança fazem levantar uma questão extra. Em quais cidades mulheres são eleitas com mais frequência? Afinal, como comentei, o número de cadeiras provavelmente dificulta o acesso de mulheres a esses cargos. Vamos avaliar essa questão através de um mapa cloroplético, onde a cor da cidade diz respeito ao percentual de mulheres eleitas em relação ao total. Note que o mapa faz o *plot* somente das informações das cidades em 2020.



De fato, note que há muitas cidades onde o percentual de cadeiras ocupadas por mulheres em Câmaras Municipais do estado é 0 ou muito próximo a isso. É provável

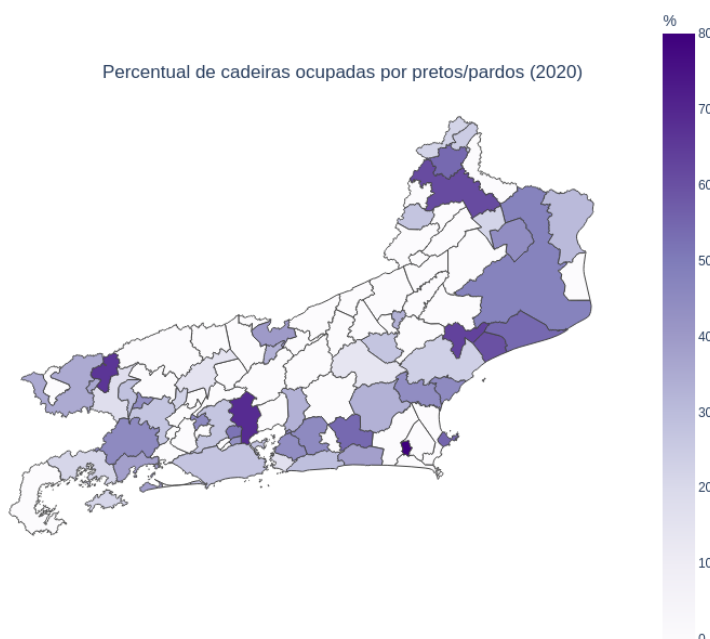
que sejam cidades cujas Câmaras possuem menos cadeiras, aumentando o número de votos necessário para a conquista da eleição. E, como também vimos, mulheres normalmente são menos votadas que homens.

Vamos avaliar agora através da categoria racial:



Apesar de ainda baixos, principalmente levando em consideração as características da população fluminense e brasileira, o percentual de pretos e pardos veio crescendo ao longo dos anos. Em particular, parece que os cargos no nível local (Câmaras Municipais e Assembleia Legislativa) são acessados com menor dificuldade em relação aos cargos federais.

Vale a pena fazer uma *plotagem* de um mapa nos mesmos moldes do anterior:



Observe que novamente há muitas cidades onde pretos e pardos representam um percentual de cadeiras muito próximo a 0, principalmente no interior do estado.

Ainda assim, os percentuais chegam a valores bem maiores em relação ao das mulheres, como a própria escala do gráfico permite observar.

## Considerações finais

De forma geral, o trabalho implementa uma série de visualizações com o objetivo de sanar 5 principais dúvidas, tangenciando outras questões ao longo do processo. Em particular, pudemos observar que há diferenças importantes entre eleições federais e municipais, principalmente no que diz respeito à correspondência entre despesas de campanha e número de votos e quando observamos os dados por gênero. Além disso, notamos que os problemas inerentes a sistemas eleitorais proporcionais afetaram a arena fluminense principalmente favorecendo partidos mais à direita, por vezes obtendo mais cadeiras com quantidades menores de votos.

Além disso, também observamos que o contexto nacional de polarização impactou diretamente a distribuição de cadeiras na bancada fluminense. Partidos associados a Lula ora aumentavam a bancada, ora diminuíam, enquanto partidos associados a Bolsonaro em geral cresceram bastante após seu surgimento como candidato viável. Além disso, partidos de direita tendem a comportar mais candidatos ricos, ao mesmo tempo que apresentam as maiores diferenças entre homens e mulheres no que diz respeito à média de valores declarados em bens.

Finalmente, as visualizações também permitiram avaliar que cargos de vereador são acessados por mulheres com mais dificuldade. Muitas cidades no estado do Rio de Janeiro não elegem uma mulher sequer para as cadeiras disponíveis em suas Câmaras Municipais. Os percentuais de pretos e pardos eleitos também são baixos em todas as instâncias, mas são maiores que os das mulheres.

O objetivo do trabalho foi apresentar essas informações da forma mais clara e direta possível na forma de visualizações de naturezas variadas, contemplando gráficos de barra, linha, *scatterplots*, mapas e *boxplots*. Optou-se também por abordagens um pouco mais avançadas de *data wrangling*, como a pivotagem das tabelas e o preenchimento de informações antigas a partir das informações mais recentes, além de cálculos mais triviais de média e outros agrupamentos.