# JF EM DADOS: DEMOCRATIZANDO INFORMAÇÃO ATRAVÉS DA CIÊNCIA DE DADOS

*Matheus Valentim e Marcello Filgueiras*

## A JFemDados

Com mais de 500 mil habitantes, o município de Juiz de Fora é um dos maiores do estado de Minas Gerais. Pólo educacional local, a cidade tem grande relevância na região, tendo vários municípios de menor porte dependendo diretamente de Juiz de Fora na saúde, na educação e no mercado de trabalho. Mesmo relevante, o município tem uma gestão por vezes omissa a questões relevantes e um debate público incipiente, que usa muito pouco de fontes oficiais e estatísticas.

Nosso objetivo é demonstrar, através de visualizações de dados públicos (mapas, gráficos e tabelas) todo tipo de temática que envolve o município. Objetivamos qualificar o debate das mais variadas maneiras, criando um jornalismo de dados local bem embasado e didático, que nutra o debate municipal com conteúdo quantitativo. Isso porque a Prefeitura, junto de outros órgãos municipais, apesar de disponibilizar os dados, o faz de maneira pouco organizada e de difícil consulta por parte da população. Nosso esforço é na direção contrária: extrair informações públicas e oficiais que ficam "arquivadas" no fundo de bases de dados ou de documentos oficiais e expô-las de uma maneira visual, acessível e didática. Tudo isso tendo como principal insumo a linguagem R, onde fazemos a maioria dos nossos trabalhos e também onde estamos desenvolvendo nosso blog via {blogdown} e um site interativo via {shiny}.

## Nosso workflow e o uso do R

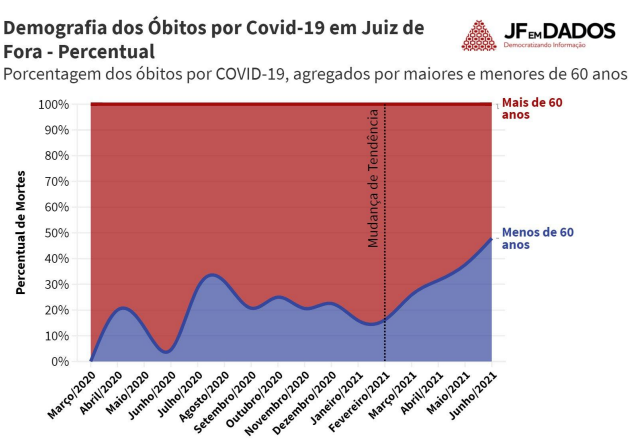
O fluxo de nosso trabalho envolve três grandes etapas: a extração/obtenção dos dados, o tratamento das bases e a disponibilização de visualizações e das bases tratadas para o público.

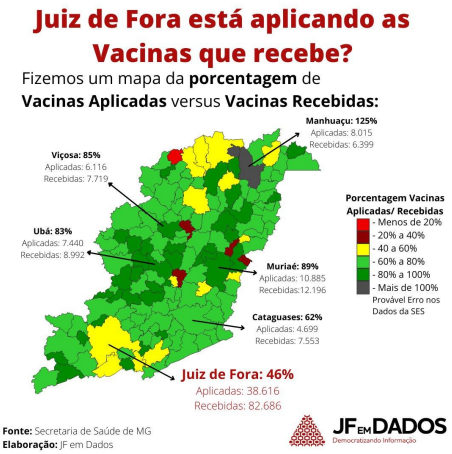
Os dados são extraídos de páginas oficiais de órgãos da cidade, como o site da Prefeitura e o site da Câmara Municipal, e também de páginas de órgãos públicos brasileiros, como IBGE e DataSUS. A extração envolve uso de técnicas de web scrapping através de pacotes como {httr}, {xml2}, {rvest} e {pdftools}, e é feita principalmente nos casos em que as fontes oficiais só disponibilizam pdfs ou tabelas virtuais dos dados, como foi o caso com óbitos e casos de COVID-19. Outra técnica de extração é a utilização da {basedosdados}, filtrando as bases nacionais para a cidade de Juiz de Fora. O tratamento dos dados consiste em transitar as bases para um formato tidy, sempre com o critério de padronizar as variáveis presentes em colunas. A visualização e a disponibilização das bases para o público é talvez a parte mais importante do nosso trabalho. Atualmente, as nossas visualizações são feitas através do {ggplot2} e do Flourish. Para atingir nosso público, temos publicações diretas em mídias sociais, como Twitter e Instagram (@jfemdados), além de uma parceria com a Tribuna de Minas, jornal estadual que publica os dados da cidade diretamente em cadernos impressos e virtuais. Todo nosso material usado, incluindo as bases que criamos no processo, as visualizações e os R scripts usados ficam no nosso Github.

## Temáticas abordadas e material produzido

Trazemos desde informações "quentes" e notícias que estão em pauta, como foram casos, ocupação de leitos e vacinação contra COVID-19, até dados ilustrativos "frios" como emissão de poluentes, focos de queimadas na cidade e censo escolar da cidade. Como exemplo desses temas quentes, onde os dados não estavam acessíveis para o público, temos o caso das mortes municipais por COVID-19. Caso emblemático, que envolveu

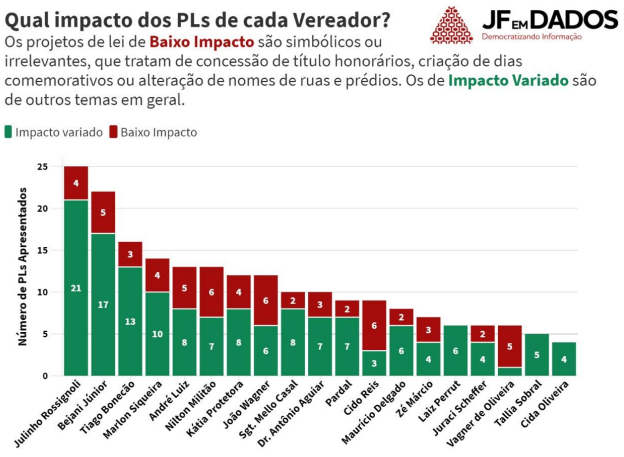
transformar um PDF 46 páginas de texto corrido, em um data frame de 1700 linhas, sendo necessário separar inicialmente cada caso por linhas, e cada variável em colunas específica, com uso denso de regex, {stringr} e {tidyr}. Uma vez separadas, o processo de padronização das variáveis como a variável “comorbidades” foi necessário e feito também com uso de {stringr}. Para a padrozinação de datas, o {lubridate} foi usado. Tudo exigiu cerca de 400 linhas de código. Um RMarkdown desse processo está disponível, bem como o primeiro post do Twitter . A segunda postagem do Twitter destacada reflete a análise da eficiência da ação da prefeitura em aplicar os estoques das vacinas disponibilizadas, comparando com outras cidades do Estado, usando mapas do pacote {geobr} e {sf}.



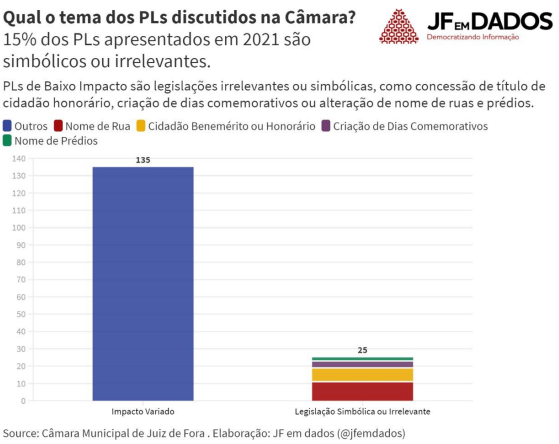




Fonte: Prefeitura de Juiz de Fora • Elaboração: JF em Dados (@jfemdados)

Por fim, apresentamos o nosso carro chefe: a análise da Câmara Municipal. Nesse projeto exploramos a atividade dos parlamentares e transparecemos isso à população. Buscamos demonstrar tanto o que está sendo votado, proposto e discutido por cada um dos vereadores, separando os por tema. Para obter os dados, usamos de web scraping com {httr} e {rvest} para raspar o site da câmara municipal para obter projetos de lei (PLs), moções e requerimentos. Para classificá-los, usamos regex via {stringr} para buscar por temáticas específicas, buscando por palavras ou expressões-chave recorrentes. Além disso, partindo da literatura da ciência política (CUNHA, Patrick Silva. “O Poder Legislativo Municipal: Estrutura, Composição e Produção”), analisamos qualitativamente os projetos que teriam caráter “simbólico”, de pouca relevância prática para a população, tais como criação de dias comemorativos e mudanças em nomes de ruas. O exemplo abaixo pode ser conferido aqui.




Fonte: Câmara Municipal de Juiz de Fora • Elaboração: JF em Dados (@jfemdados)




Source: Câmara Municipal de Juiz de Fora . Elaboração: JF em dados (@jfemdados)