

## Story-telling interativo com dados de despesa pública: uma análise sobre os Estados brasileiros.

As story-telling interativas permitem a junção de uma síntese narrativa com a exploração e análises avançadas de dados pelo usuário final. No R isso é feito utilizando ao mesmo tempo Rmarkdown e Shiny. O presente trabalho mostra uma construção que apresenta essa abordagem. Trabalha-se aqui com a evolução de dados de despesa pública, notadamente despesas de pessoal, no escopo da federação brasileira. Os dados utilizados são os produzidos para envio de relatórios ao FMI seguindo uma metodologia dessa instituição. A história começa com a possibilidade do usuário montar dinamicamente indicadores de despesa de pessoal e analisar em um mapa de calor o comportamento desse número para todos os Estados brasileiros, ao longo do período de 2010 a 2018. As opções dadas ao usuário permitem que faça análises sobre até doze indicadores diferentes. No gráfico gerado tonalidades vermelhas revelam indicadores mais elevados e o contrário para tonalidades azuis. Inicialmente o nível crítico que divide as cores é dado pela mediana dos valores apresentados. O usuário pode em seguida modificar esse valor.

A narrativa dirigida começa em seguida e analisa três indicadores dos doze possíveis. Encontra-se clusters de Estados formados pelas séries temporais do indicador formado pelo total de despesas de pessoal no critério competência dividido pelo total de receita. Os dois clusters formados são validados em dois gráficos distintos onde fica claro que o algoritmo PAM utilizado gerou uma escolha razoável dos componentes de cada um dos agrupamentos. Em seguida mostra-se uma nova versão do mapa de calor com a evolução temporal do indicador com o foco nos dois clusters formados. A figura 1 mostra esse gráfico.

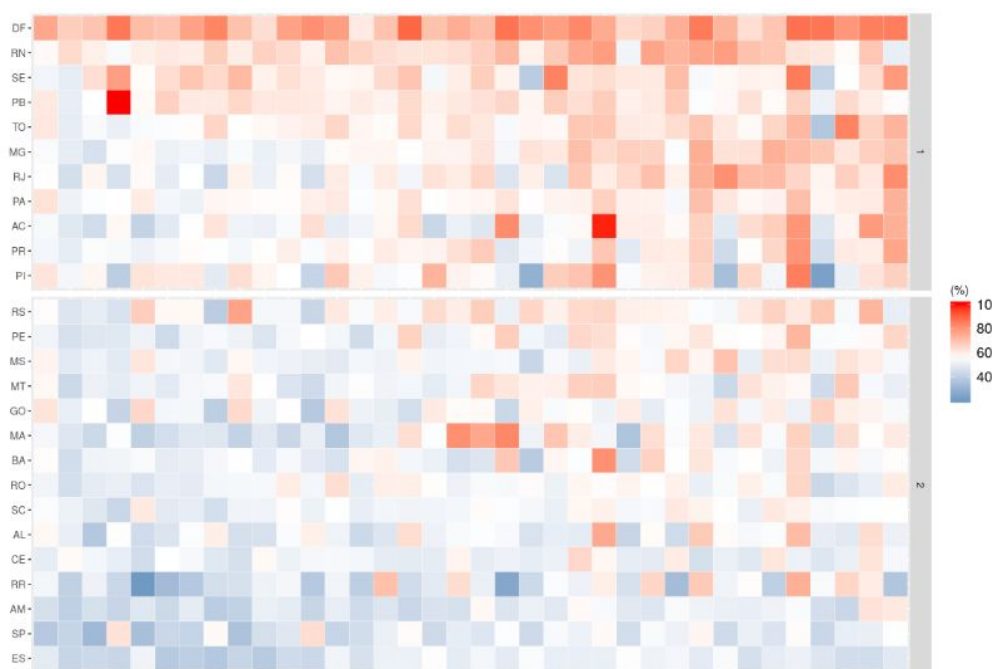


Figura 1: Mapa de calor dos clusters formados pelas séries temporais

Pela Figura 1 o leitor percebe claramente que no cluster 1 o indicador apresenta valores elevados desde o início da série temporal, enquanto que o cluster 2 há um comportamento que revela um início azul e que aos poucos vai tomando tonalidades vermelhas, sem chegar ao final da série temporal no mesmo nível do cluster 1.

A análise temporal dos clusters é submetida a dois novos testes gráficos. Dessa vez pelo agregado anual. Um dos gráficos mostra os dois clusters comparados simultaneamente ano a ano através de box-plot. Fica claro nesse caso pela distribuição dos valores que ao longo dos anos o cluster 1 sempre teve os valores posicionais mais elevados do que o cluster 2. O outro gráfico é em formato de violino e mostra também a distribuição ao longo dos anos, mas nesse caso o cuidado foi de comparar se as médias anuais de cada um dos clusters eram significativamente diferentes. Para isso se usou testes ANOVA. Os violinos de cores iguais apresentam também médias estatisticamente iguais. A função do gráfico em violino foi permitir deixar mais claro ano a ano a distribuição para cada um dos pontos possíveis do indicador. O leitor pode ficar mais confortável com essa visualização em aceitar os resultados das comparações de média dada pelo teste ANOVA.

Para as análises seguintes o indicador inicial é decomposto em dois outros indicadores: despesas com pessoal ativo em relação ao total de receita e despesa com pessoal inativo em relação ao total de receita. Para esses dois indicadores novamente são feitas comparações das médias anuais usando o teste ANOVA. Os quatro gráficos montados para essa análise contam uma história em que para os dois clusters o que vem se destacando é o aumento das despesas com pessoal inativo em relação ao total das receitas.

O estado do Amapá não se enquadra dentro das análises anteriores e recebeu uma sessão específica onde se analisa as quatro séries temporais que podem dar origem aos indicadores trabalhados nessa história. Dois gráfico, um de linha e outro de barras empilhadas trazem visualizações do comportamento dessas séries.

Por fim, a história retorna a possibilidade de interação com usuários. Agora o leitor pode ter acesso a construção de cenários das composições futuras dos clusters através de previsões utilizando ARIMA. Para cada um dos estados, o usuário pode escolher o percentual sobre o valor máximo ou valor mínimo das previsões ou acreditar no valor mais realista. Essa possibilidade deixa na mão do usuário a elaboração de vários conjuntos de cenários baseados em suas próprias expectativas de evolução do indicador para cada um dos estados.

A história se conclui aí. A aposta é que usuários partam de uma exploração inicial de dados, tome contato com uma estrutura narrativa que conduz uma visão específica sobre os dados e se conclui com uma análise mais avançada que permite elaboração sofisticada de cenários. Esses recursos narrativos e analíticos abrem possibilidades para um maior controle social e incrementos em educação fiscal. Por estar em código aberto, espera-se que a partir do exemplo e em breve pela disponibilização dos dados abertos, ocorra o incentivo a novas ressignificações por parte de amplos setores da sociedade civil.

O produto está disponível no seguinte link:

[https://fabdev.shinyapps.io/analise\\_clusters\\_despesas\\_pessoal\\_estado/](https://fabdev.shinyapps.io/analise_clusters_despesas_pessoal_estado/)