04

Agregando e visualizando nossos dados

Entendendo as agregações

Uma das funcionalidades mais interessantes do ElasticSearch é a capacidade de executar agregações de modo distribuído, combinado com um engine de busca bem flexível, como vimos até agora.

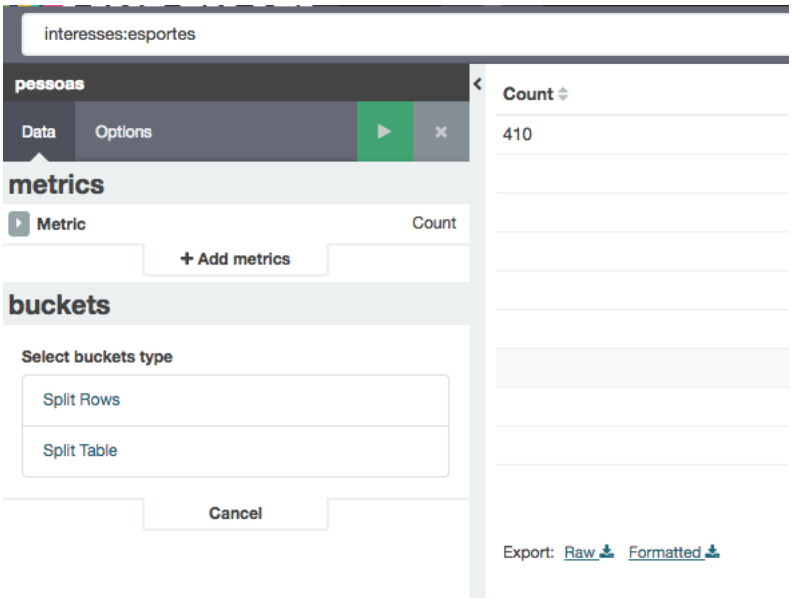
Map-Reduce com Elasticsearch

Caso você já tenha ouvido sobre Map-Reduce (https://pt.wikipedia.org/wiki/MapReduce), pense em um modo de espalhar os documentos que temos em grupos separados (as agregações) e depois reduzi-los a um valor que pode ser a quantidade de documentos deste grupo. A ideia é igual ao uso de *group by* em bancos de dados relacionais.

Nada melhor que um exemplo prático para exemplificar o que foi dito no parágrafo anterior. Queremos saber quais são as 5 formações que mais possuem interessados em esportes, em cada um dos estados do Brasil. Note que queremos uma resposta em alguns segundos, mesmo para um volume de dados na casa de centenas de gigabytes de dados (ou mesmo terabytes).

Para encontrar as 5 formações que possuem mais interessados em esportes por estado, precisamos entender como os documentos devem ser separados em grupos e depois que tipo de redução devemos utilizar.

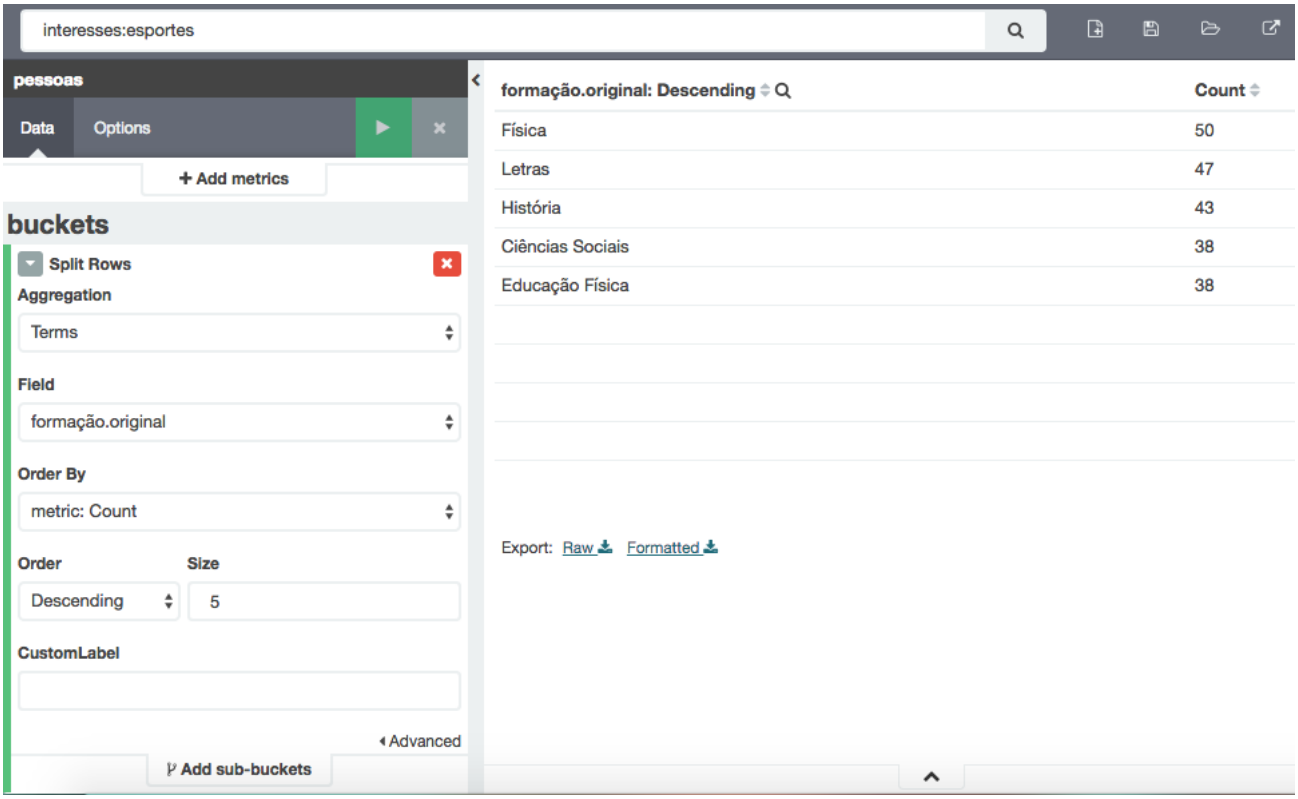
Vamos passo a passo. Queremos saber as 5 formações que mais possuem interessados em esportes. Já sabemos como listar todos os documentos que possuem pessoas interessadas em esportes. Para listar `quantos`, basta contá-los:



Veja que a contagem é a nossa métrica. Dado um conjunto de documentos, aplicamos uma métrica. Neste caso, usamos a métrica **Count**, 410 na imagem.

Usando buckets

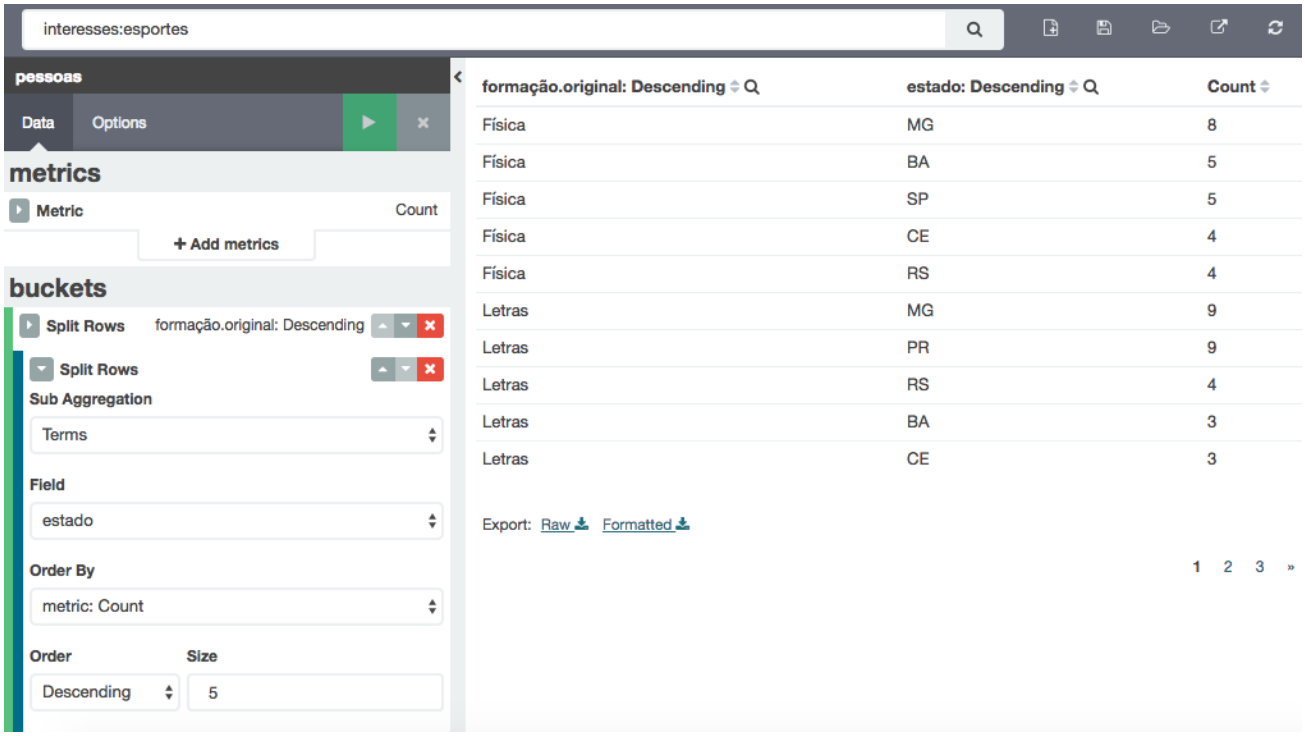
O próximo passo é quebrar esta contagem em grupos ou, como chamado pelo ElasticSearch e Kibana, quebrar em *buckets*. Vamos criar um *bucket* para cada uma das 5 formações com mais interessados em esportes (*Split-Rows -> Size: 5*):



Note que:

- ***Aggregation Terms***: tipo de agregação que usamos para texto.
- ***Field***: nome do atributo a ser usado na agregação. Note que devemos usar os atributos que **não** foram analisados, caso contrário teremos os *tokens* gerados no resultado e não o valor que indexamos. Ainda bem que usamos *multi-fields*!
- ***Order By***: como o *bucket* deve ser ordenado. Neste caso, será ordenado pela quantidade de elementos que possui.
- ***Order***: se queremos os *buckets* com mais elementos no topo (*Descending*) ou com menos elementos (*Ascending*).
- ***Size***: quantos *buckets* queremos.

O próximo e último passo é quebrar cada *bucket* de formação por estado. Para tal, criaremos um *sub-bucket* por estado. Neste exemplo, vamos nos limitar ao top 5 estados:



Note que, uma vez que entendemos os conceitos, fica muito simples encontrar respostas para perguntas que até então pareciam complicadas. Porém, temos um problema. Se com apenas 5 cursos e 5 estados fica difícil visualizar todos os cursos e estados, imagine com todos os 26 estados do Brasil.

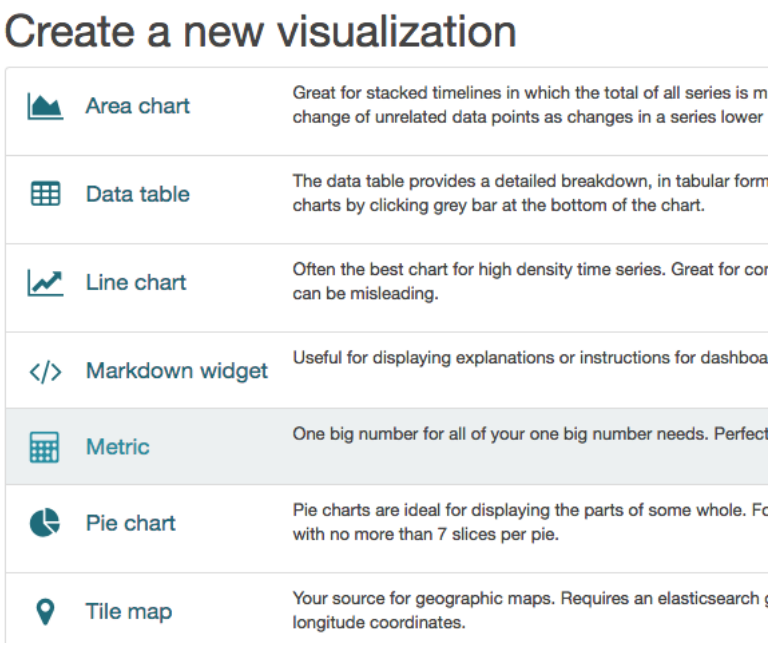
Criando Visualizações e Dashboards

Vamos criar um *dashboard* para podermos visualizar os dados que temos em nosso índice. Nosso objetivo é criar as seguintes visualizações:

- Quantidade de registros que temos em nosso índice.
- Gráfico de pizza com as 5 formações mais populares.
- Gráfico de barra com as 5 formações mais populares distribuídas entre os 26 estados.

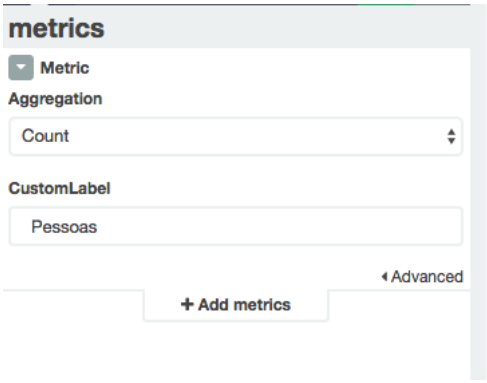
Métrica Count

Para criarmos visualizações, basta irmos em ***Visualize*** e escolher o tipo de visualização que desejamos criar. Como nossa primeira visualização é apenas o valor de uma métrica, devemos escolher ***Metric***, como mostrado a seguir:

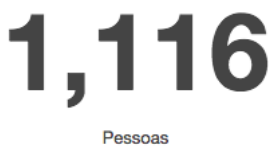


Como não salvamos nenhuma das nossas buscas, escolhemos ***From a new search***. Como só temos um índice, o próprio Kibana faz a escolha para nós. Como queremos apenas o número de registros, a métrica ***count***, que é utilizada como padrão na criação de uma nova visualização, já resolve nosso problema.

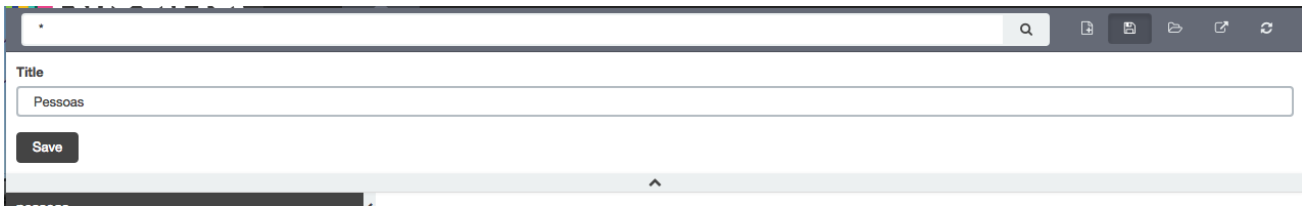
Vamos apenas alterar o rótulo, como mostrado a seguir:



E temos o resultado:



Basta salvarmos nossa visualização e dar um nome a ela:

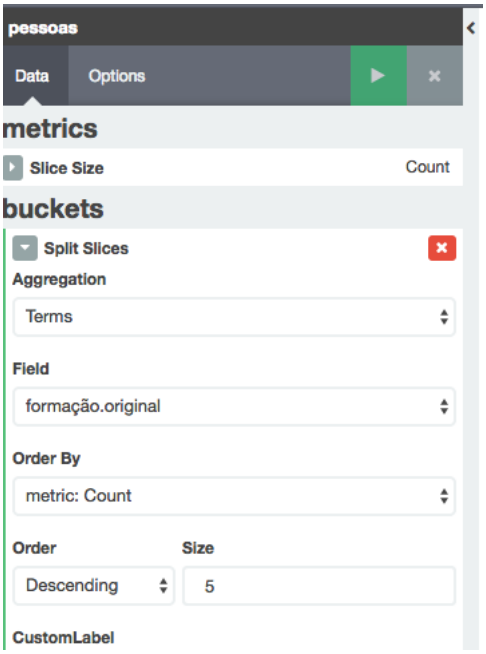


Usando Pie Chart

Vamos criar agora o gráfico de pizza. O processo é o mesmo, porém devemos selecionar *Pie Chart*. A diferença é que precisamos decidir como 'cortar' o gráfico.

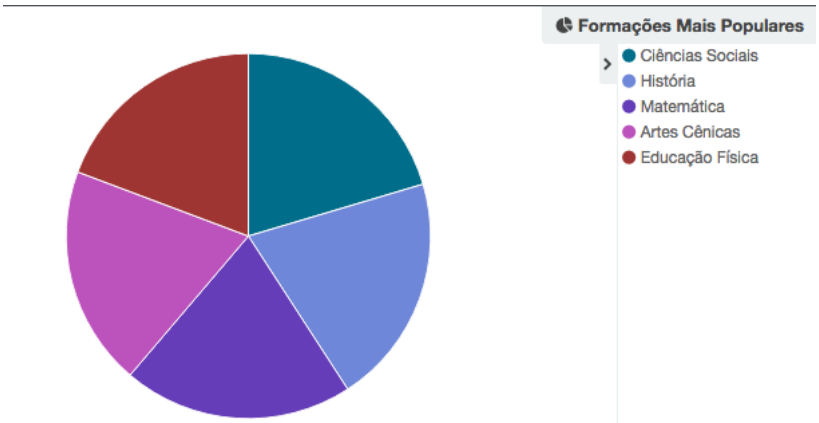
Note que 'cortar', neste contexto, significa criar o *bucket*. Neste caso, queremos apenas um gráfico com o top 5. Logo, devemos utilizar *Split slices* e selecionar o atributo formações.original.

Nosso *size* será 5 e devemos utilizar a ordem *Descending*, pois queremos os mais populares ou com maior `count`, assim como mostrado na figura a seguir:



Importante: Não devemos esquecer de clicar no botão play para ver o resultado da agregação no gráfico.

E temos o gráfico:



Salve o gráfico com o nome **Formações Mais Populares**.

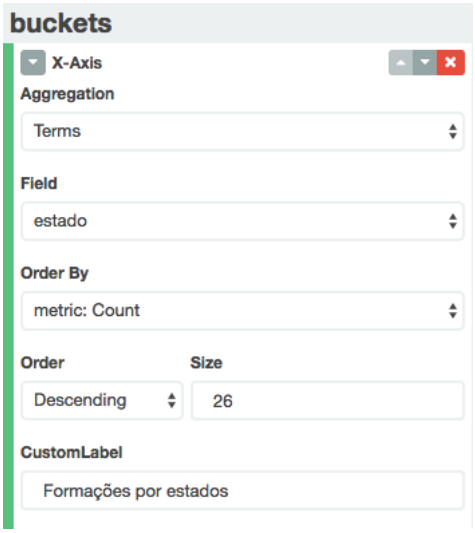
Criando Bar Chart

Por fim, vamos criar o gráfico de barras. O processo é o mesmo, porém devemos selecionar ***Vertical bar chart***. Neste tipo de visualização, temos diferentes maneiras para criar os *buckets* de dados:

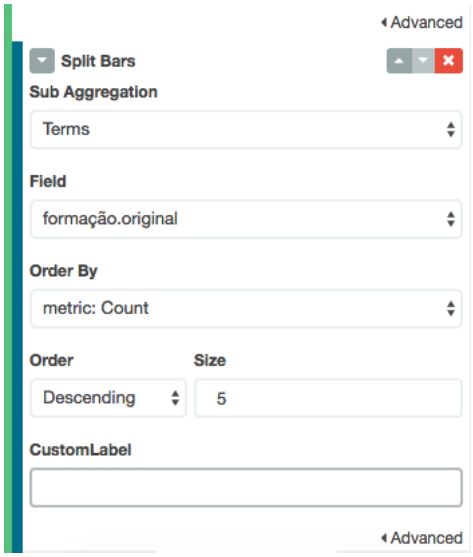
- ***X-Axis***: O rótulo de cada *bucket* criado é utilizado para criar uma barra nova, que será mostrada no eixo X.
- ***Split Bars***: Cada barra já existente no gráfico é particionada pelos rótulos de *bucket* criados.

- ***Split Charts***: O rótulo de cada *bucket* criado é utilizado para criar um novo gráfico.

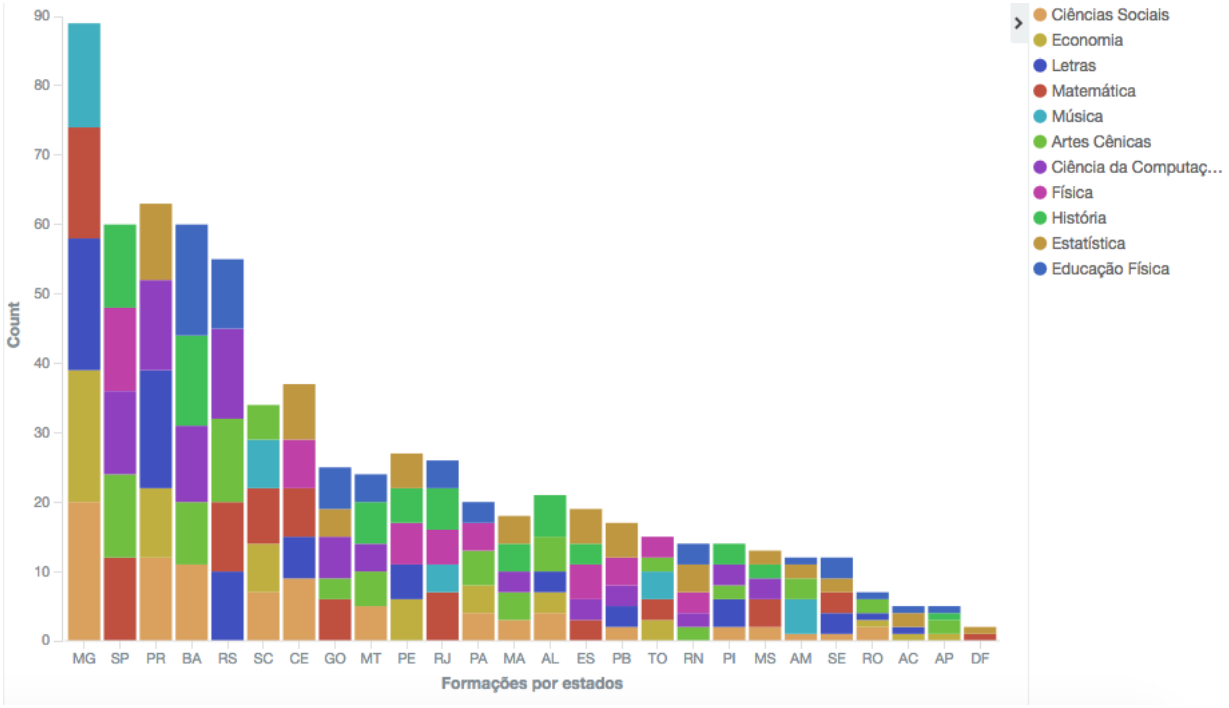
No nosso caso, queremos visualizar as 5 formações mais comuns para cada um dos 26 estados do Brasil. Logo, vamos primeiro criar uma barra para cada estado:



Agora vamos criar outro *bucket* com as 5 formações mais comuns:



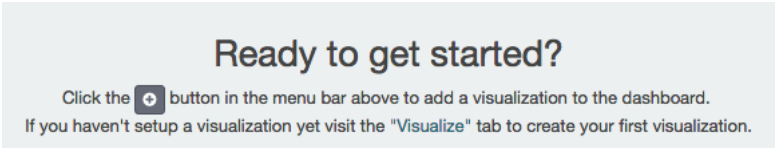
O resultado esperado é:



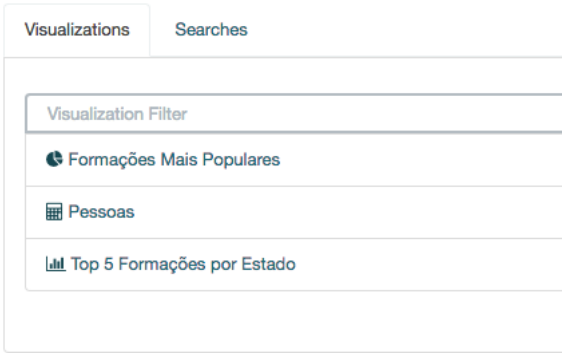
Vamos salvar a visualização com o nome **Top 5 Formações por Estado**.

Montando o Dashboard

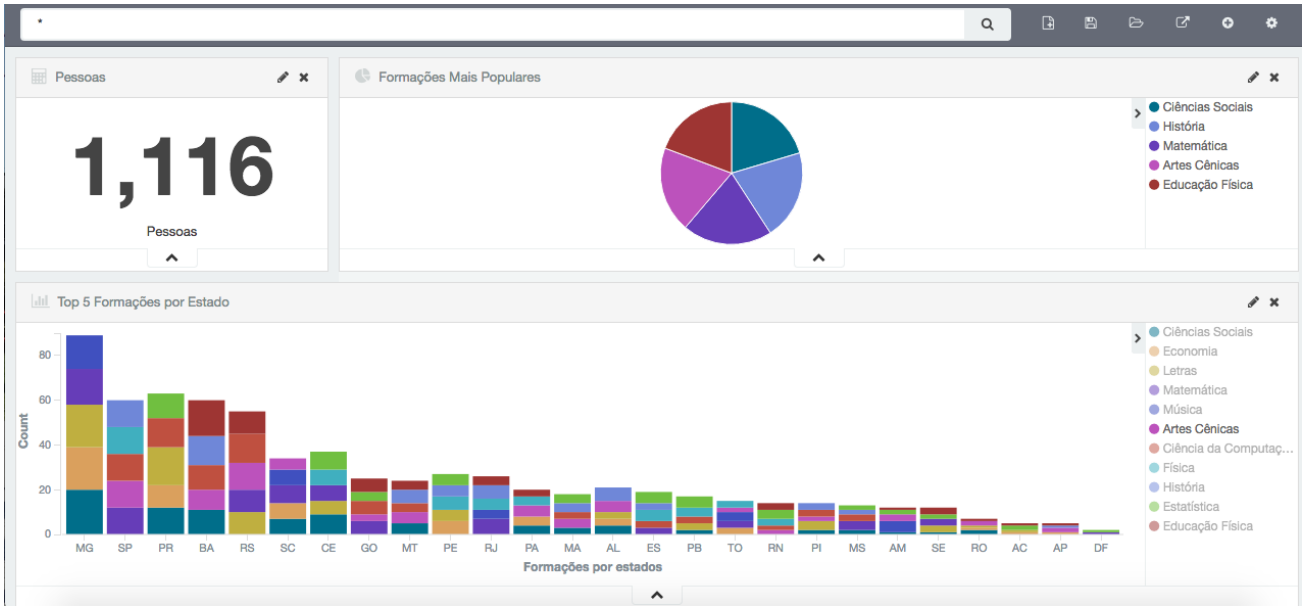
Nosso último passo é criar um *dashboard* para mostrar as visualizações recém criadas. Para tal, basta clicarmos em ***Dashboard***. Note que a mensagem mostrada é bem clara sobre o que devemos fazer para adicionar visualizações:



Basta clicarmos no sinal mostrado na figura abaixo para ver as visualizações que criamos. Note que podemos também adicionar buscas que foram salvas previamente na *tab* ***Discovery***.



Basta agora selecionar as visualizações, uma a uma, que elas serão mostradas no ***Dashboard***. Note que podemos arrastar e dimensionar as visualizações:



Por fim, basta salvarmos o *dashboard* e pronto. Temos uma rápida e interativa maneira de visualizar e interagir com nossos dados.

Repare que, ao clicarmos em alguns dos elementos na tela, como uma fatia do gráfico de pizza, um filtro com o valor selecionado é aplicado em todas as visualizações.

Podemos também usar a barra de filtro para escolher, interativamente o que queremos ver. Veja o exemplo a seguir, onde estamos interessados em filtrar as pessoas que possuem interesse em esportes: