

# Conectar a PDFs

---

Bem-vindo a este vídeo sobre o conector de PDF. Você pode baixar os arquivos PDF para acompanhar os procedimentos usando sua própria cópia do Tableau.

Arquivos PDF podem conter dados valiosos para análises no Tableau. O conector de PDF no Tableau foi desenvolvido para transformar esses dados em tabelas. Como os PDFs não têm metadados sobre os dados, talvez você precise executar algumas tarefas após a conexão.

## Conectar tabelas de dados em um arquivo PDF

O Tableau lê tabelas de dados de referência cruzada em PDFs. Idealmente, elas devem ter esta aparência, com colunas e linhas, e uma entrada em cada linha. Os dados sobre as ações começam na página 2. Vamos abrir o Tableau. Quando selecionamos Arquivo PDF no painel Conectar, a janela para seleção de páginas é aberta. Podemos analisar o documento inteiro, uma página específica ou um intervalo de páginas. Analisaremos as páginas 2 a 8, e arrastaremos a primeira página para o canvas. Uma das linhas do cabeçalho do PDF está confundindo o Tableau, mas, se ativarmos o interpretador de dados, os cabeçalhos são exibidos corretamente.

## União de linhas

Uma tabela foi gerada para cada página, mas, como elas são estruturadas com os mesmos cabeçalhos, fica fácil mesclá-las usando o recurso de união. Com a primeira tabela exibida no canvas, arrastaremos todas as outras tabelas para a área de união, posicionando-as abaixo da primeira. Uma nova coluna para o nome da tabela é adicionada, e, se rolarmos para baixo, podemos ver que os dados da página 2 foram perfeitamente alinhados aos dados da página 3.

## Organizar tabelas irregulares

Este PDF tem uma ótima estrutura para importação, mas isso nem sempre acontece. Como regra geral, o Tableau se conectará melhor a PDFs que têm uma estrutura tabular como esta, com apenas uma entrada por linha, sem hierarquias ou cabeçalhos aninhados e sem subtabelas. Mas os PDFs não precisam ser perfeitos para funcionar. Vamos conectar outro arquivo, um exemplo que não seja tão fácil quanto este. A tabela que queremos está na página 14. Lembre-se: o Tableau procura números de página, que podem ou não corresponder à paginação do documento.

Adicionaremos outra fonte de dados e escolheremos a página 14. A opção “Analisar arquivo PDF novamente”, no menu suspenso das conexões de dados, nos permite selecionar as páginas que serão analisadas. Existe apenas uma tabela nesta página, mas há três opções aqui à esquerda. O Tableau detectou três formas possíveis de transformar essa página em tabela. Se arrastarmos uma de cada vez para o canvas, veremos seus conteúdos. (Observação: quando unir tabelas de diferentes páginas, como fizemos antes, se houver várias versões de cada página, verifique se uniu a versão correta da tabela, e não múltiplas versões da tabela de uma mesma página.)

Vamos arrastar a tabela 1 e usar o interpretador de dados. Parece que a tabela 1 está completa, mas, por algum motivo, os anos de 1995 a 1997 foram transformados em uma só coluna. Se quisermos usar essa versão dos dados, podemos organizar os dados utilizando uma divisão personalizada para dividir todas as colunas por espaços e, em seguida, renomear os novos campos com os anos correspondentes.

Mas vamos ver as tabelas 2 e 3. A tabela 2 parece ser a parte inferior da tabela original, e a tabela 3, a parte superior. Prefiro a formatação das colunas dessas duas tabelas, então vamos utilizá-las. Primeiro, vamos unir essas duas tabelas, arrastando a tabela 2 para baixo da tabela 3. Há um problema de correspondência entre as colunas “Inflows” e F1. Se selecionarmos as duas colunas e escolhermos “Mesclar campos não correspondentes”, o problema será corrigido. Vamos renomear a nova coluna como “Water Sources”.

## Lidar com valores nulos

Várias linhas contêm o valor nulo, seja porque um subcabeçalho, como “Change in storage”, foi interpretado como uma linha de dados, ou porque uma única linha, como “abstraction from hydroelectricity generation”, foi interpretada como duas linhas diferentes. Para eliminarmos esses valores nulos, adicionaremos um filtro de fonte de dados. No canto superior direito, clique em Adicionar. Vamos adicionar um filtro. Podemos selecionar qualquer uma dessas colunas que contêm um valor nulo. Vou selecionar F10, clicar em OK e selecionar nulo e Excluir. Clique em OK e em OK novamente. As linhas com nulos foram eliminadas, e ficamos apenas com as entradas de dados.

Mas vou desfazer isso, porque também vemos que existem vários tipos de fontes de água que são na verdade totais. Queremos filtrar os totais e os nulos de uma só vez. Vá para Filtro > Adicionar e selecione Water Source. Marcaremos as caixas de todos os cabeçalhos ou totais do PDF original ou das linhas que contêm valores nulos. “Hydroelectricity” aparece duas vezes, uma com dados e outra com nulos. Vou mantê-la para não excluir dados. Selecione Excluir e OK.

## Corrigir cabeçalhos e dinamização

Agora todas as linhas têm dados, exceto a linha “hydroelectric” com os nulos, mas corrigiremos isso em breve. Os cabeçalhos não têm nome, exceto esta primeira coluna, mas podemos consultar os valores na tabela do PDF original para ver quais são. As colunas correspondem aos anos de 1995 a 2010. Vou acelerar isso.

Agora podemos dinamizar os dados para criar uma coluna Year (Ano) e uma coluna Million Cubic Meters (Milhões de metros cúbicos). Oculte a coluna com os nomes das tabelas e altere o tipo de dados da coluna Year para data. Altere o tipo de dados da coluna Million Cubic Meters para Número (inteiro). E, como essa coluna já está no seu formato final, filtraremos os valores nulos dela aqui. Em Filtros, selecione Editar, Adicionar, Million Cubic Meters e exclua os nulos.

## Redefinir aliases de membros de um campo

Ainda precisamos corrigir os nomes das fontes de água. Vamos acessar o menu dessa coluna e clicar em Aliases. Aqui, podemos renomear os membros de um campo. É só clicar duas vezes no alias e digitar o nome correto.

- *Este é **discharge from hydroelectricity generation**.*
- *Groundwater*
- ***Abstraction for hydroelectricity***

## Recriar grupos e hierarquias

Por fim, lembro que havia uma estrutura na tabela original no PDF: existem categorias de fontes de água. Podemos criar essa estrutura no painel Dados. Para isso, clicaremos em Planilha 1. Bem, Million Cubic Meters é uma medida, por isso vamos arrastar esse campo para cá.

Agora podemos recriar os grupos. Arraste Water Sources para linhas para criarmos o primeiro grupo. Mantenha a tecla Ctrl pressionada e clique nos membros que você deseja adicionar ao grupo:

- *Abstraction for hydroelectricity*
- *Discharge from hydroelectricity generation,*
- *Evapotranspiration e*
- *To sea and net abstraction*

Clique no ícone de clipe de papel da dica de ferramenta para criar o grupo. Repita esse procedimento para criar outro grupo: *Groundwater, Ice, Lakes and reservoirs, Snow e Soil moisture*

Vamos clicar com o botão direito do mouse neste novo campo do painel Dados e editar o grupo. Primeiro, clicaremos em “Precipitation” e, mesmo sendo apenas um membro, o transformaremos em um grupo. Agora podemos renomeá-los como: *“Inflows”, “Change in storage” e “Outflows”*. E renomearemos o campo com o agrupamento como Categorias.

Se arrastarmos e soltarmos o campo Water Sources original sobre esse novo campo Categorias, é criada uma hierarquia que permite detalhar a exibição.

### **Dicas para trabalhar com PDFs**

As etapas para organizar os dados de um PDF após conectá-lo ao Tableau podem variar, mas espero que este vídeo tenha mostrado algumas ferramentas úteis para você organizá-los e prepará-los para suas análises. Lembre-se que o Tableau pode ter dificuldade para analisar PDFs que contêm: subtabelas, hierarquias nos cabeçalhos, várias linhas de conteúdo que deveriam ser interpretadas como uma única linha. Por fim, observe também que as cores e os sombreamentos de uma tabela podem alterar a interpretação dos seus dados, devido à forma como os PDFs são transformados em tabelas de dados.

### **Conclusão**

Obrigada por assistir ao vídeo de treinamento Conectar a PDFs. Sugerimos que você continue assistindo aos vídeos de treinamento gratuitos para aprender mais sobre como usar o Tableau.