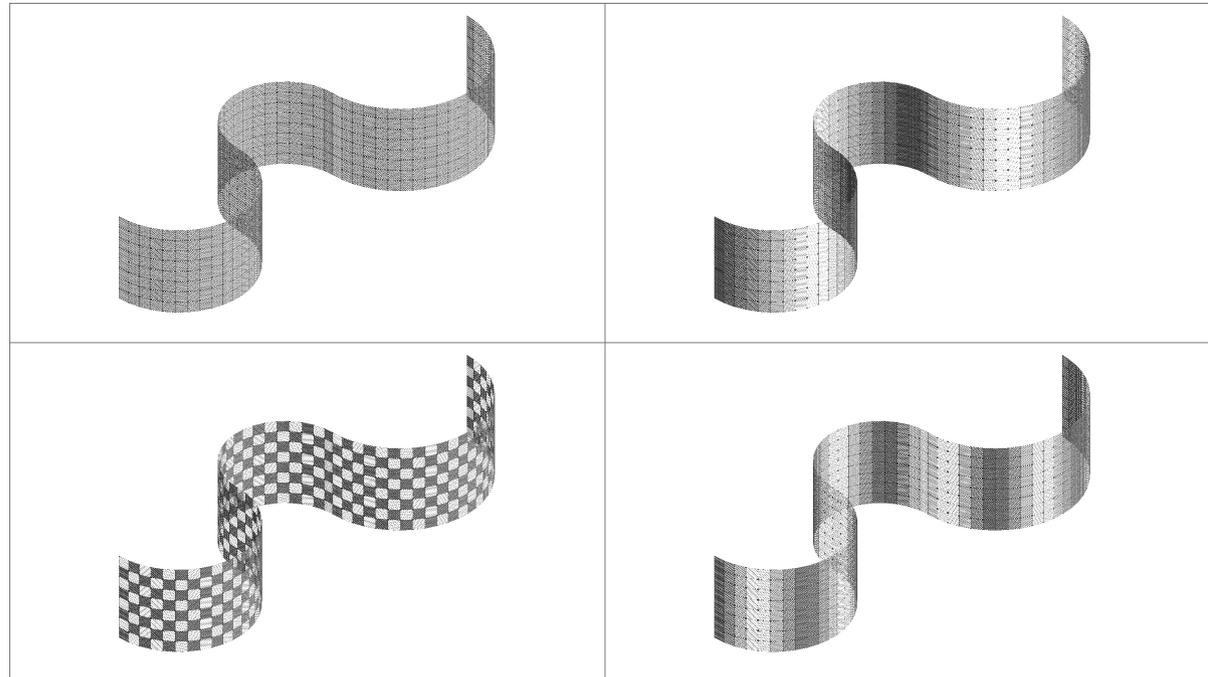


Tópicos em Gestão da Informação II

Aula 02 – Identificando e tratando padrões



Prof. Dalton Martins
dmartins@gmail.com

Gestão da Informação
Faculdade de Informação e Comunicação
Universidade Federal de Goiás

Padrões

- Em muitas ocasiões, recebemos os dados em **formatos que não servem** aos propósitos de análise que desejamos;
- Uma das operações mais comuns em planilhas é **transformarmos** o formato dos dados e formatos que possam ser sistematizados do modo que desejamos;
- Algumas questões envolvem esse processo
 - saber como analisar os dados,
 - onde procurar por padrões que facilitem as transformações
 - como transformar!

Transformações

- Há vários tipos de transformações que podemos fazer nos dados:
 - Mudar o formato de estruturação da informação:
 - De tipo: número para texto, por exemplo;
 - De forma: quebrar em duas ou mais partes e separar em informações independentes, por exemplo;
 - Agrupar a informação em diferentes formas e encontrar padrões na organização dos dados;
 - Contar informações de uma dada categoria;
 - Somar, multiplicar, encontrar a média de uma dada categoria;
 - Entre outros...

Mudando o formato dos dados

- Veremos algumas técnicas a partir de alguns exemplos aplicados:
 - No exemplo a seguir, temos dados em uma coluna de nossa planilha em que o número do CPF e o NOME da pessoa estão juntos.

Credor/Fornecedor
290.414.491-91 - Abilio Wolney Aires Neto
827.354.092-87 - Adailton Ferreira De Araujo
034.983.831-32 - Adalberto Ribeiro Sampaio Junior
003.525.421-10 - Adalto Jose De Souza
917.072.501-20 - Adda Daniela Lima Figueiredo
002.981.061-24 - Adelia Cristina Fernandes Silva
023.760.661-58 - Adesvaldo Jose E Silva Junior
766.664.411-68 - Adjair Maranhao De Sousa

Problemas

- Os dados nesse formato podem gerar alguns problemas:
 - Organizar os nomes em ordem alfabética ou filtrar por letra inicial se torna mais difícil;
 - Organizar os campos apenas por CPF também;
- O ideal seria termos essas informações separadas em duas colunas, desse modo, teríamos maior autonomia sobre lidar com esses dados.

Comandos do Open Office de referência

- NÚM.CARACT
- PROCURAR
- DIREITA
- ESQUERDA

NÚM.CARACT

- A Função NÚM.CARACT apresenta como resultado o tamanho, em número de caracteres, do TEXTO inserido na função como argumento. A estrutura da função é:

=NÚM.CARACT(TEXTO)

O resultado da função =NÚM.CARACT("Teste") é igual a 5.

PROCURAR

- Procura uma cadeia de caracteres de texto dentro de outra. Você também pode definir o local de início da pesquisa. O termo de pesquisa pode ser um número ou qualquer cadeia de caracteres. A pesquisa faz distinção entre maiúsculas e minúsculas.
- Sintaxe: `PROCURAR("ProcurarTexto"; "Texto"; Posição)`
 - ProcurarTexto refere-se ao texto a encontrar.
 - Texto é o texto onde a pesquisa ocorre.
 - Posição (opcional) é a posição no texto a partir da qual a pesquisa começa.
- Exemplo:
 - `=PROCURAR(76;998877665544)` retorna 6.

DIREITA

- Devolve como resultado os N últimos caracteres do texto utilizado como ARGUMENTO.
- O formato da função é: =DIREITA (ARGUMENTO; N)
- O parâmetro N é opcional. Caso o mesmo seja omitido, o valor retornado pela função corresponderá apenas ao último caractere do texto do ARGUMENTO.
- Por exemplo:
 - =DIREITA(“Carro”;2) retornará como resultado o texto “ro”.
 - =DIREITA(“Carro”) retornará como resultado a letra “o”.

ESQUERDA

- Devolve como resultado os N primeiros caracteres do texto utilizado como ARGUMENTO a partir da esquerda.
- O formato da função é: =ESQUERDA (ARGUMENTO; N)
- O parâmetro N é opcional. Caso o mesmo seja omitido, o valor retornado pela função corresponderá apenas ao primeiro caractere do texto do ARGUMENTO.
- Por exemplo:
 - =ESQUERDA(“LibreOffice”;2) retornará como resultado o texto “Li”.
 - =ESQUERDA(“LibreOffice”) retornará como resultado a letra “L”.

Juntando as funções

- Como utilizar essas funções para resolver o problema inicial, ou seja, separar os nomes dos cpfs em nossos dados?
 - O padrão que divide os dois conjuntos de dados é “ - “. A questão é como criar uma estratégia para dividir os dois conjuntos!

Credor/Fornecedor
290.414.491-91 - Abilio Wolney Aires Neto
827.354.092-87 - Adailton Ferreira De Araujo
034.983.831-32 - Adalberto Ribeiro Sampaio Junior
003.525.421-10 - Adalto Jose De Souza
917.072.501-20 - Adda Daniela Lima Figueiredo
002.981.061-24 - Adelia Cristina Fernandes Silva
023.760.661-58 - Adesvaldo Jose E Silva Junior
766.664.411-68 - Adjair Maranhao De Sousa

Juntando as funções

- Para pegar os nomes, faremos:
 - =DIREITA(A2;NÚM.CARACT(A2)-PROCURAR("-";A2)-2)
- Para pegar os CPFs, faremos:
 - =ESQUERDA(A2;PROCURAR("-";A2))

Agrupando dados: encontrando o formato das distribuições de um conjunto de dados

- Vimos na aula passada que para a média ser um bom valor de síntese de conjunto de dados temos de considerar seu desvio padrão:
 - Se o desvio padrão for muito grande (em geral, maior que a média), o conjunto de dados se mostra muito disperso com muitos dados longe da média;
 - Se o desvio padrão for pequeno, o conjunto de dados se mostra menos disperso com muitos dados distribuídos em torno da média.
- Uma boa forma de analisar o conjunto de dados é visualizar sua distribuição.

Formato da distribuição

- O formato de uma distribuição de dados mostra como os dados estão distribuídos entre o conjunto de valores que eles podem ocupar;
- Permite entender, de forma geral, que características esses dados possuem, permitindo encontrarmos os melhores métodos analíticos para tratá-los;
- Para visualizar o formato de uma distribuição, precisamos:
 - Selecionar uma lista das ocorrências de dados no conjunto;
 - Calcular quantas vezes cada ocorrência aparece no conjunto;
 - Fazer um gráfico para visualizar a distribuição das ocorrências.

Voltando ao Open Office...

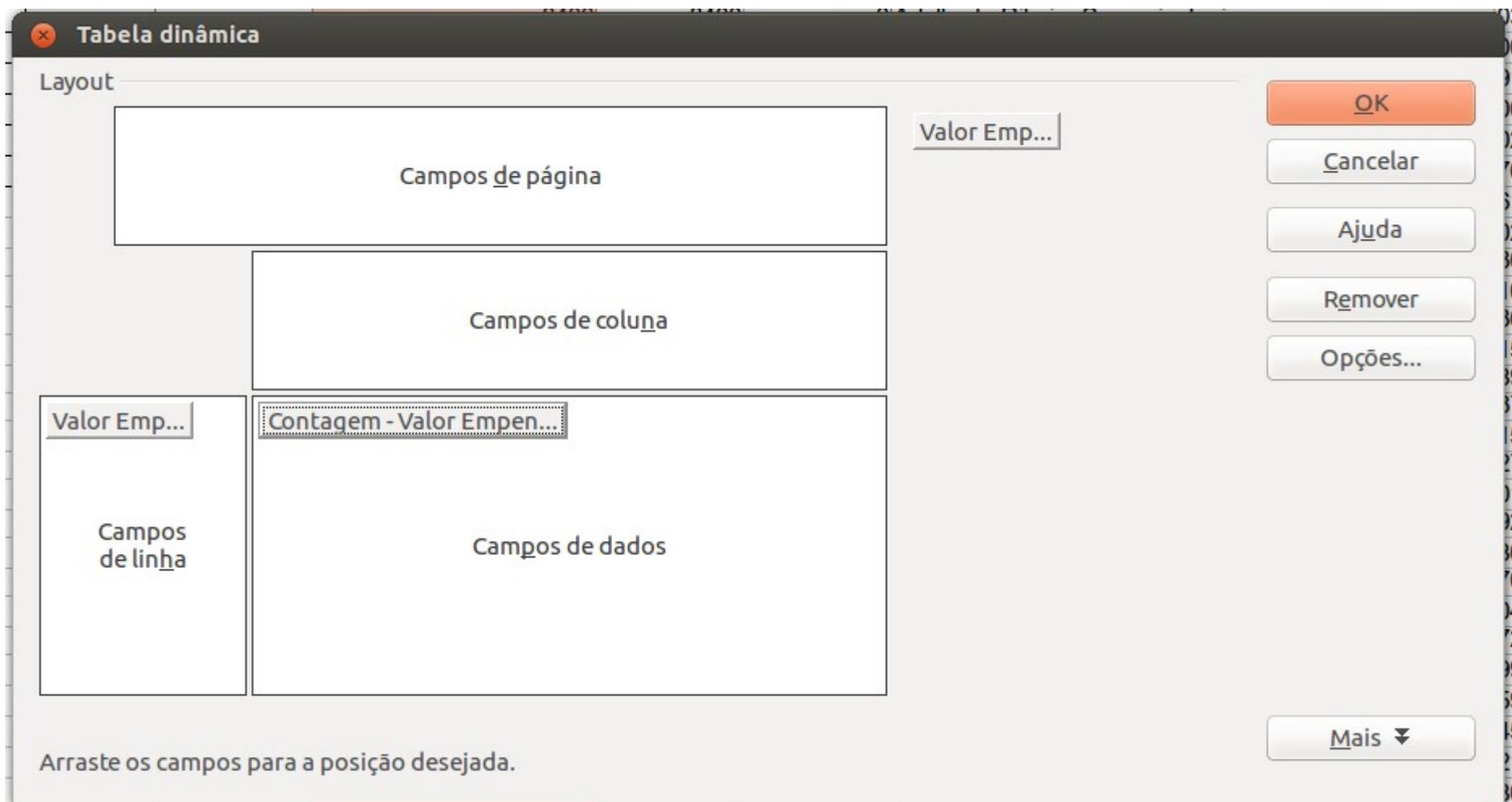
- O Open Office possui um recurso fundamental para análise de dados chamado de Tabela Dinâmica;
- Também conhecido como Assistente de Dados, a Tabela Dinâmica é um recurso do Calc que permite combinar e analisar dados de forma simples e rápida.
- Pode ser usado para criar uma espécie de resumo de um conjunto de dados, a fim de dar uma visão geral sobre a representatividade desses dados.

Montando a tabela dinâmica

- Se quisermos visualizar a distribuição dos valores “empenhados” em nossa planilha de estudos, teremos de fazer o seguinte:
 - Selecionar a coluna “Valor empenhado”;
 - Ir no menu “Dados”->”Tabela dinâmica”->”Criar”;
 - Vai aparecer a seguinte janela. Escolha “Seleção atual”, pois vai trabalhar apenas nos dados selecionados



Selecionando a estrutura dos dados



- Campos de página: permite criar filtros dinâmicos para o conjunto de dados;
- Campos de coluna: separa os dados em colunas;
- Campos de linha: separa dos dados em linhas;
- Campos de dados: permite as operações de síntese dos dados ali alocados.

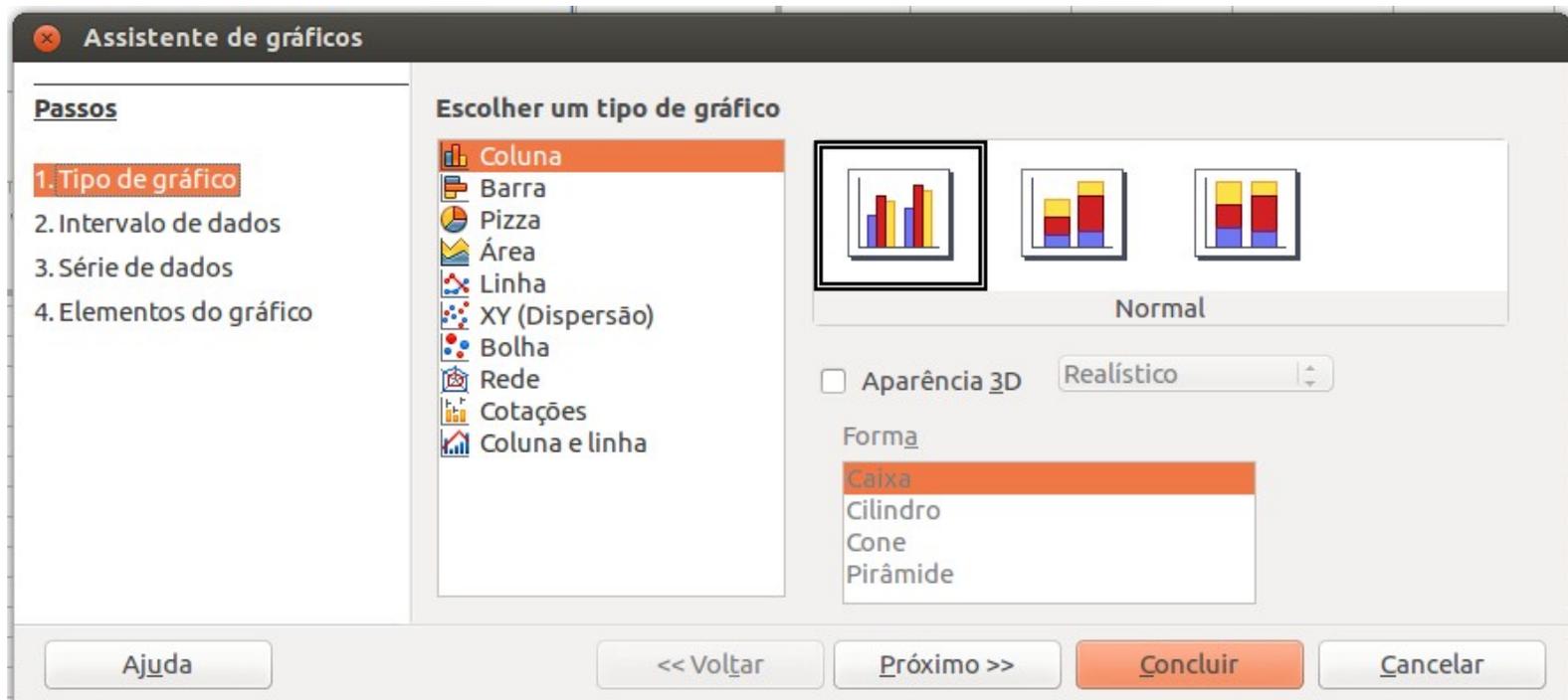
Observando os resultados

- Clicamos em “OK” e o Open Office cria uma nova aba na planilha, onde teremos os dados no seguinte formato, mostrando quantas vezes um valor apareceu no conjunto de dados...

Valor Empenhado	
0	27
550	1
1100	1
1200	8
1300	1
1467,7	1
1486,52	1
1520	1
1600	1
1647	1
1800	1
1860	1
1954,4	1
1990	1
2015	1
2020	1
2060	1
2122	1
2200	1
2276	1
2360	1
2400	39

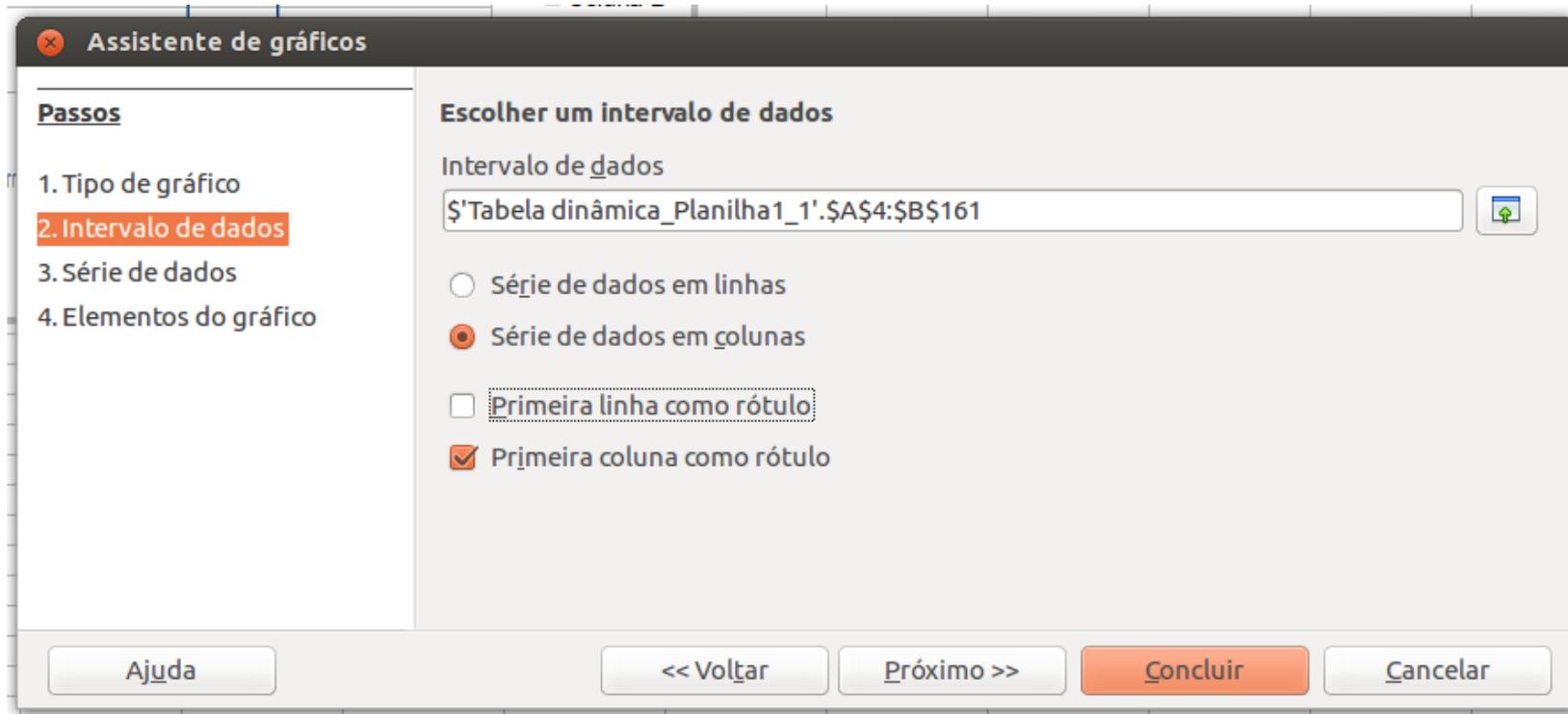
Montando o gráfico

- Selecionados todo o conjunto de dados da nova aba da planilha;
- Clicamos no ícone do gráfico () na barra de ferramentas e veremos a janela a seguir:



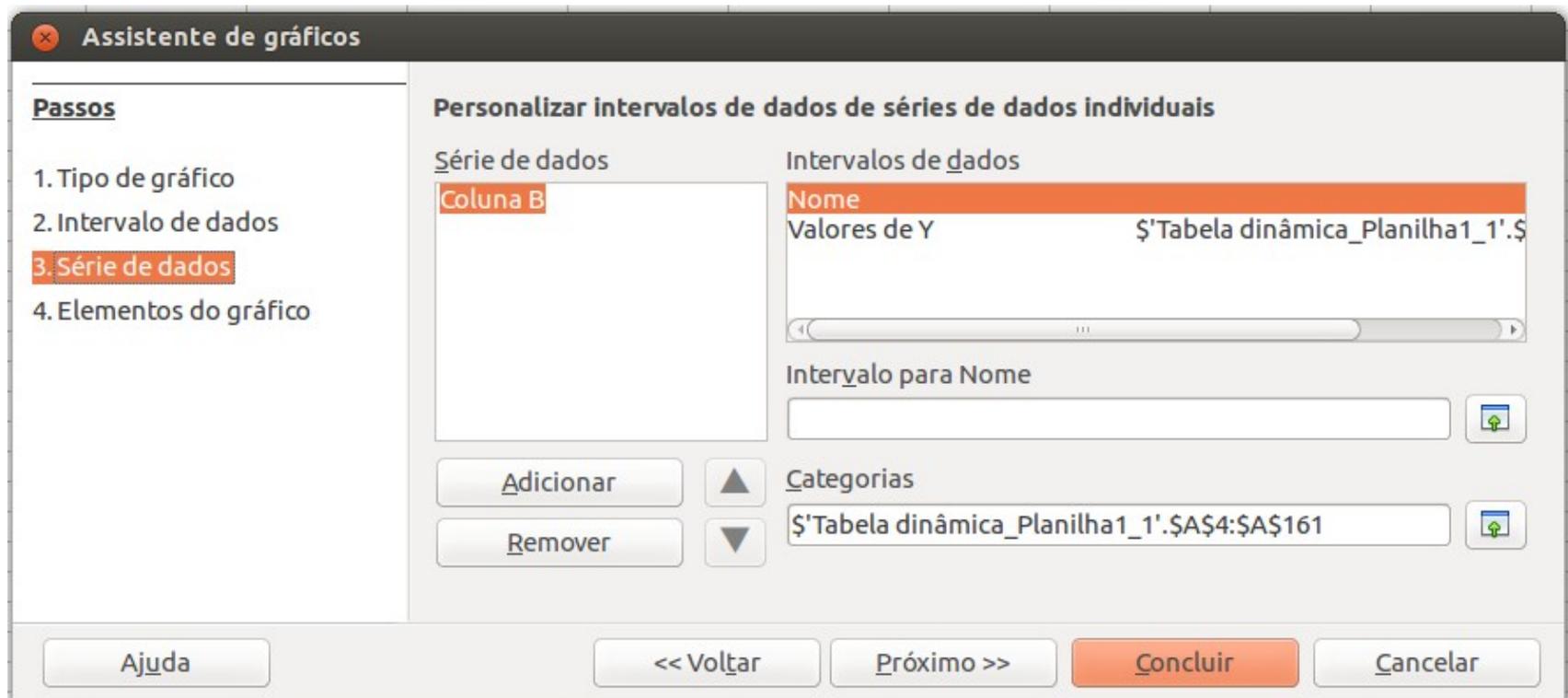
Montando o gráfico

- Selecionamos o tipo “coluna” e clicamos no “Próximo” e veremos a janela a seguir, onde escolheremos “Primeira coluna como rótulo de dados”, clicamos em “Próximo”.



Montando o gráfico

- Veremos em seguida a janela a seguir, onde não mexeremos em nenhuma opção pois não vamos modificar o intervalo de dados já selecionado. Clicamos em “Próximo”.



Montando o gráfico

- Preenchemos o título do gráfico e dos seus eixos, para facilitar seu entendimento. Clicamos no botão “Concluir”.

Assistente de gráficos

Passos

1. Tipo de gráfico
2. Intervalo de dados
3. Série de dados
- 4. Elementos do gráfico**

Escolher títulos, legendas e configurações de grade

Título...

Subtítulo

Eixo X

Eixo Y

Eixo Z

Exibir legenda

À esquerda

À direita

Em cima

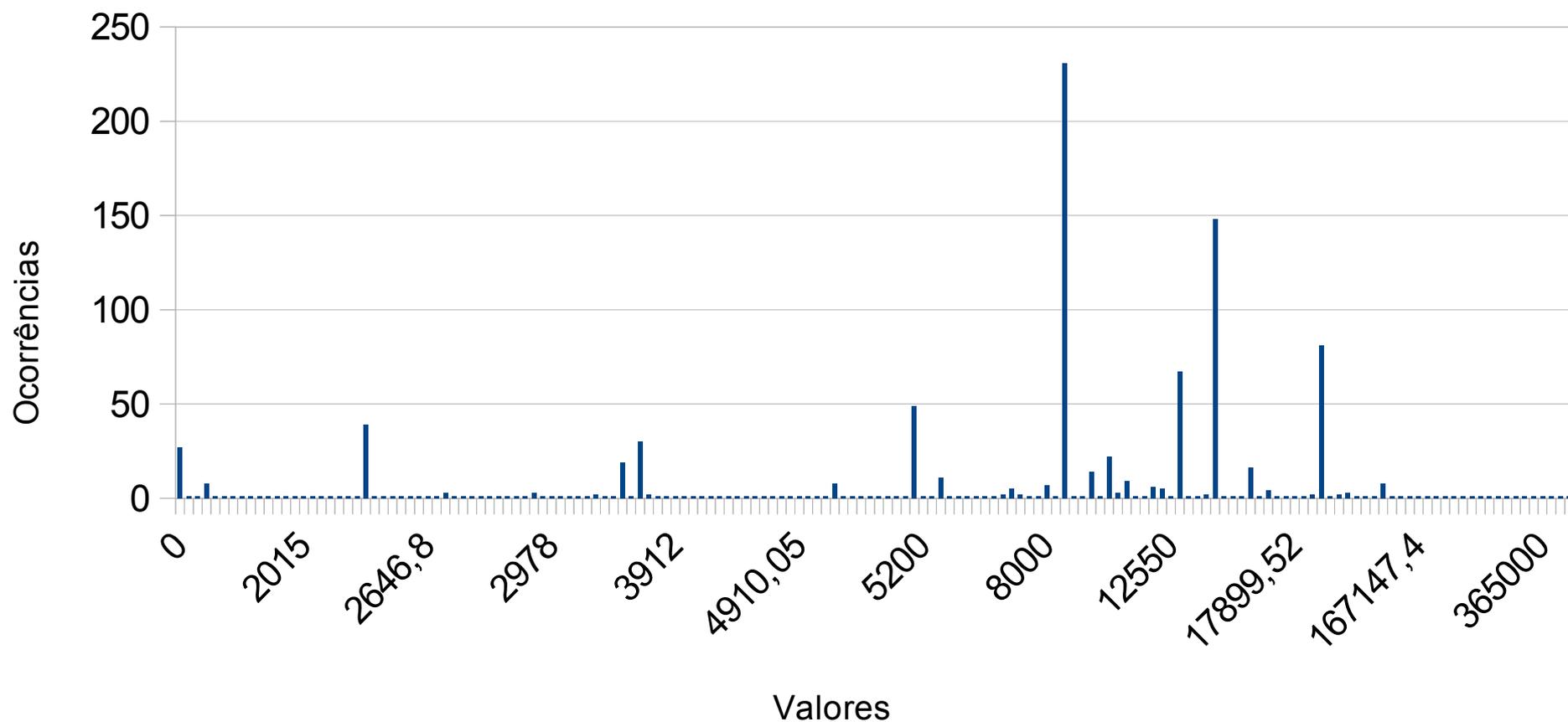
Embaixo

Exibir grades

Eixo X Eixo Y Eixo Z

Montando o gráfico

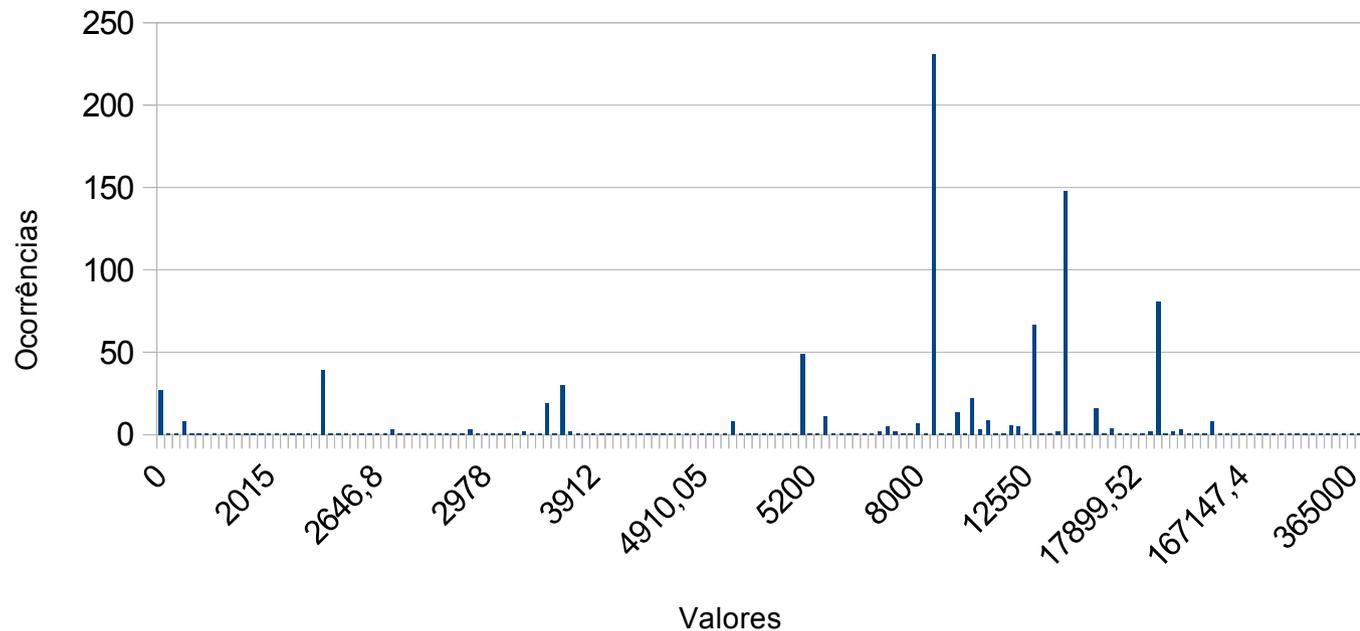
Distribuição de valores empenhados



O que você acha dessa distribuição?

Lembre-se que a média gira em torno de 23500.

Distribuição de valores empenhados



Exercícios

- Encontre os 10 nomes que tiveram maiores valores empenhados;
- Encontre os 10 CPFs que tiveram menores valores empenhados;
- Encontre uma estratégia de busca para identificar automaticamente quais são os nomes que tem CPF e quais CNPJ;
- Monte um gráfico de barras para os 10 nomes que tiveram maiores valores empenhados;
- Crie um gráfico de bolha para visualizar a distribuição de frequência exemplificada nesta aula. Que diferenças você nota entre o gráfico de barras e o de bolha?