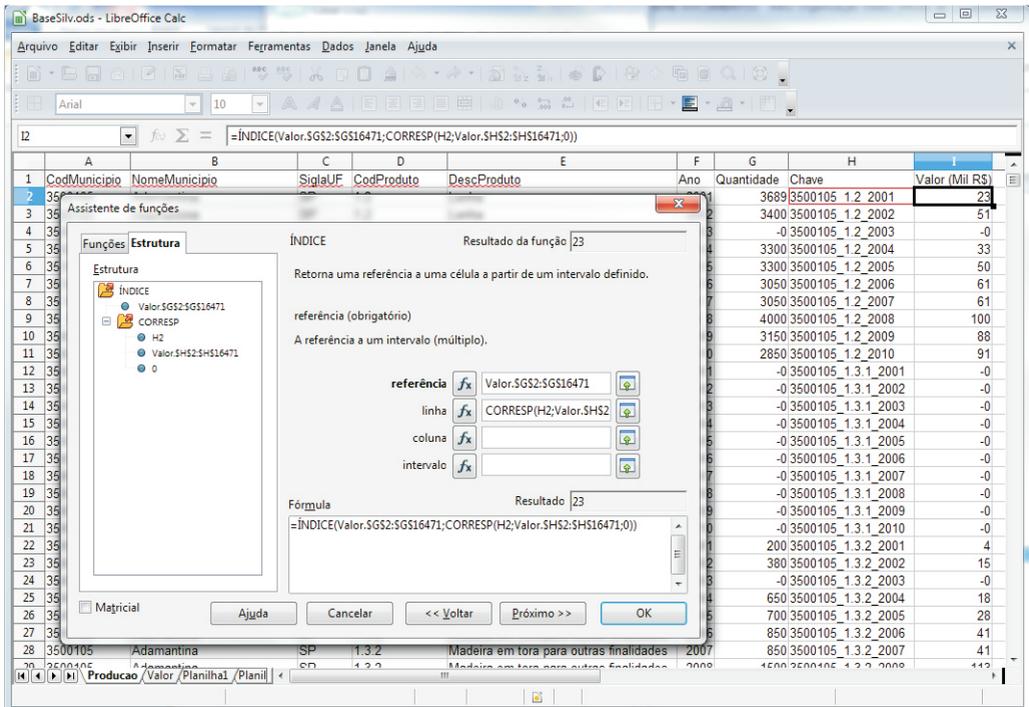


Introdução a organização, tratamento e análise descritiva de bases de dados em planilha eletrônica (Libre Office Calc)



*Empresa Brasileira de Pesquisa Agropecuária
Embrapa Florestas
Ministério da Agricultura, Pecuária e Abastecimento*

Documentos 235

Introdução a organização, tratamento e análise descritiva de bases de dados em planilha eletrônica (Libre Office Calc)

José Mauro Magalhães Ávila Paz Moreira

Embrapa Florestas
Colombo, PR
2012

Embrapa Florestas

Estrada da Ribeira, Km 111, Guaraituba,

83411-000, Colombo, PR - Brasil

Caixa Postal: 319

Fone/Fax: (41) 3675-5600

www.cnpf.embrapa.br

sac@cnpf.embrapa.br

Comitê Local de Publicações

Presidente: Patrícia Póvoa de Mattos

Secretária-Executiva: Elisabete Marques Oaida

Membros: Álvaro Figueredo dos Santos, Antonio Aparecido Carpanezi, Claudia Maria Branco de Freitas Maia, Dalva Luiz de Queiroz, Guilherme Schnell e Schuhli, Luís Cláudio Maranhão Froufe, Marilice Cordeiro Garrastazu, Sérgio Gaiad

Supervisão editorial: Patrícia Póvoa de Mattos

Revisão de texto: Mauro Marcelo Berté

Normalização bibliográfica: Francisca Rasche

1ª edição

Versão digital (2012)

Todos os direitos reservados

A reprodução não-autorizada desta publicação, no todo ou em parte, constitui violação dos direitos autorais (Lei no 9.610).

Dados Internacionais de Catalogação na Publicação (CIP)

Embrapa Florestas

Moreira, José Mauro Magalhães Ávila Paz.

Introdução a organização, tratamento e análise descritiva de bases de dados em planilha eletrônica (Libre Office Calc) [recurso eletrônico] / José Mauro Magalhães Ávila Paz Moreira. - Dados eletrônicos. - Colombo : Embrapa Florestas, 2012.

(Documentos / Embrapa Florestas, ISSN 1980-3958; 235)

Sistema requerido: Adobe Acrobat Reader.

Modo de acesso: World Wide Web.

<<http://www.cnpf.embrapa.br/publica/seriedoc/edicoes/doc235.pdf>>

Título da página da web (acesso em 13 ago. 2012).

1. Planilha eletrônica. 2. Programa de computador. 3. Dado. I. Título.
II. Série.

CDD 005.368 (21. ed.)

Autor

José Mauro Magalhães Ávila Paz Moreira
Engenheiro Florestal, Doutor,
Analista da Embrapa Florestas
josemauro@cnpf.embrapa.br

Apresentação

Vivemos na era da informação, quando dispomos de uma infinidade de bases de dados que podem nos trazer variadas e importantes informações, desde que possamos trabalhá-las e extrair delas as informações que necessitamos. Atualmente temos acesso às ferramentas computacionais (hardware e software) que nos permitem realizar esta tarefa, mas o conhecimento necessário para fazê-lo pode limitar o acesso para a maioria dos usuários.

Na falta de conhecimentos mais especializados em ferramentas de bancos de dados, as planilhas eletrônicas podem, até certo ponto, auxiliar na superação desta barreira ao ser utilizada como ferramenta para a geração dos relatórios desejados. Entretanto, para que as planilhas possam ser utilizadas desta forma, é necessário que o usuário tenha algum conhecimento básico sobre organização e tratamento destes dados neste ambiente e sobre as ferramentas de análise de dados que as mesmas disponibilizam.

A flexibilidade de inserção e organização dos dados é uma das grandes vantagens oferecidas pelas planilhas eletrônicas. Entretanto, esta facilidade de organização leva aos usuários

a adotarem a forma de entrada e apresentação dos dados que melhor lhes convenha, geralmente mais semelhantes aos formulários de coleta de dados ou do relatório que desejam elaborar ao final do trabalho. Contudo, muitas vezes esta forma de organização dos dados pode não permitir o uso das ferramentas da planilha eletrônica para a sua análise de maneira mais eficaz e eficiente.

O objetivo desta publicação é compartilhar alguns conhecimentos a respeito da organização, tratamento e análise descritiva de bases de dados utilizando planilhas eletrônicas, em especial a planilha eletrônica do pacote Libre Office.

Washington L. E. Magalhães
Chefe de Pesquisa e Desenvolvimento

Sumário

Organização de dados	9
Tratamento de bases de dados	19
Utilizando funções da planilha eletrônica	20
Utilizando outras ferramentas da planilha eletrônica	28
União de bases de dados	37
Conceito de relacionamento entre bases de dados	38
Unindo bases de dados com um relacionamento um para um	39
Unindo bases de dados com um relacionamento um para muitos ..	47
Análise descritiva de dados	51
Considerações finais	69
Referências	70

Introdução a organização, tratamento e análise descritiva de bases de dados em planilha eletrônica (Libre Office Calc)

José Mauro Magalhães Ávila Paz Moreira

Organização de dados

Atualmente vivemos na era da informação, onde há disponibilidade de uma grande quantidade de dados, além de ferramentas computacionais (hardware e software) que nos permitem extrair variadas informações das bases de dados disponíveis. Entretanto, a maioria dos softwares estatísticos ou de banco de dados relacionais que lidam com grandes quantidades de dados exige conhecimento especializado para o seu uso, tornando o seu acesso mais difícil para a maioria dos usuários.

Esta barreira pode ser superada, até certo ponto, pelo uso das planilhas eletrônicas, desde que o usuário tenha algum conhecimento básico sobre organização e tratamento destes dados neste ambiente e sobre as ferramentas de análise de dados que as mesmas disponibilizam. Atualmente, as planilhas têm capacidade de lidar com aproximadamente um milhão de registros, e dispõem de ferramentas de fácil uso para a análise preliminar destes dados. Além disso, são mais fáceis de aprender

do que os programas convencionais de estatística, tornando a análise de bases de dados maiores mais acessíveis a usuários finais (NEUFELD, 2003).

A flexibilidade de inserção e organização dos dados é uma das grandes vantagens oferecidas pelas planilhas eletrônicas. Entretanto, esta facilidade de organização leva aos usuários a adotarem a forma de entrada e apresentação dos dados que melhor lhes convenha, geralmente mais semelhantes aos formulários de coleta de dados ou do relatório que desejam elaborar ao final do trabalho. Contudo, muitas vezes esta forma de organização dos dados pode não ser a melhor, por não permitir o uso das ferramentas da planilha eletrônica para a sua análise de maneira mais eficaz, eficiente, e minimizando os erros que possam ocorrer no processo de organização, tratamento e análise da base de dados.

Caso o usuário deseje utilizar o potencial das ferramentas de análise descritiva de dados e facilitar a junção de bases de dados, a organização dos dados na planilha eletrônica deve seguir o mesmo padrão de organização de dados utilizado em qualquer software de estatística, no qual **cada observação da base de dados é inserida em uma linha, e cada coluna representa uma variável.**

O objetivo desta publicação é compartilhar alguns conhecimentos a respeito da organização, tratamento e análise descritiva de bases de dados utilizando planilhas eletrônicas, em especial a planilha eletrônica do pacote Libre Office.

Dados podem ser entendidos como informações numéricas necessárias para nos ajudar a tomar decisões melhor fundamentadas em determinadas situações (LEVINE et al., 2000).

Dados e variáveis

Uma variável pode ser compreendida como uma classificação ou uma medida que se altera para cada caso em estudo (observação) (DUARTE, 2008). Segundo lemma (1992), as variáveis podem ser classificadas em dois tipos: quantitativas, quando descrevem quantidades, podendo ser comparadas a conjuntos numéricos; e qualitativas, que são utilizadas para descrever qualidades, categorias, etc. As variáveis quantitativas podem ser contínuas, quando são utilizadas para descrever dados contínuos, ou seja, que podem assumir qualquer valor no conjunto dos números reais; ou discretas, sendo aquelas que assumem apenas determinados valores no campo dos reais. As variáveis qualitativas podem ser ordinais, quando os valores das suas observações apresentarem um sentido possível de ordenamento, tais como classe de renda, nível de escolaridade, qualidade de um dano em uma árvore ou uma classificação de fuste. Também podem ser classificadas como nominais, quando não houver um sentido de ordenamento nos seus valores, tais como gênero, estado civil, nome de município, nome científico, etc.

As variáveis também podem ser classificadas de acordo com as suas escalas de razão (GUJARATI, 2006), sendo enquadradas nas categorias escalas de razão, de intervalo, ordinal e nominal.

Quando o resultado da razão (x_1/x_2) e distância ($x_1 - x_2$) de valores de variáveis diferentes (x_1 e x_2) resultam em grandezas com significado, então temos duas variáveis com escala de razão. Outra característica destas variáveis é que elas possuem um ordenamento natural dos seus valores ao longo da escala (GUJARATI, 2006).

Variáveis com escalas de intervalo apresentam as características de distância e ordenamento natural entre dois valores distintos. Por exemplo, a distância entre dois períodos de tempo faz sentido, e há um ordenamento natural entre os dois, já a razão entre os mesmos é uma grandeza que não tem sentido.

Variáveis com escala ordinal satisfazem a terceira propriedade da escala de razão (ordenamento natural), tais como classes de renda (alta, média, baixa) ou classificação de danos em árvores (leve, médio e grave).

Variáveis com escala nominal são aquelas que não satisfazem nenhuma das três propriedades da escala de razão, tais como gênero (masculino e feminino), estado civil (casado, solteiro, divorciado).

Compreender como se classificam as variáveis nos auxilia no momento de escolher que tipos de análises ou tratamentos podem ser aplicados para extrair informações a partir das mesmas.

Uma observação da base de dados é composta por uma linha, e o ideal é que ela contenha valores para todas as variáveis armazenadas na base de dados.

Ao longo desta publicação, iremos compartilhar algumas técnicas e ferramentas para organizar e tratar bases de dados em planilhas eletrônicas por meio de um estudo de caso aplicado a bases de dados florestais de fontes secundárias. Iremos construir uma base de dados que contenha as quantidades produzidas e o valor da produção de lenha e madeira em tora dos municípios paulistanos entre o período de 2001 e 2010, e classificá-los de acordo com a mesorregião e microrregião à qual pertencem.

Iniciaremos com a obtenção dos dados de quantidade produzida dos produtos da extração vegetal e silvicultura do site do Instituto Brasileiro de Geografia e Estatística (IBGE). Execute a seguinte sequência de passos:

- 1) Acesse o site do IBGE no endereço www.ibge.gov.br;
- 2) No menu localizado no lado direito, clique na opção Banco de Dados;

3) Clique na opção SIDRA, conforme ilustrado na Figura 1.



Figura 1. Acesso ao banco de dados agregados (SIDRA).(IBGE, 2012).

4) Após entrar na base de dados SIDRA, acesse o tema Silvicultura;

The image shows the SIDRA database interface with a table of economic indicators. The left sidebar has 'Silvicultura' selected. The table has 6 columns and 26 rows of data.

10 Imprensa e reprodução de gravações	-0,04	-0,29	0,21	-2,00	-0,99
19 Fabricação de coque, de produtos derivados do petróleo e de biocombustíveis	0,19	0,74	0,55	-0,36	-0,58
20B Fabricação de sabões, detergentes, produtos de limpeza, cosméticos, produtos de perfumaria e de higiene pessoal	1,89	0,22	1,21	0,13	-0,57
20C Fabricação de outros produtos químicos	2,46	2,84	-0,66	-1,19	-2,06
21 Fabricação de produtos farmoquímicos e farmacêuticos	1,20	-0,17	-0,42	-0,55	0,35
22 Fabricação de produtos de borracha e de material plástico	0,32	0,08	-0,25	0,63	-0,29
23 Fabricação de produtos de minerais não-metálicos	0,80	-0,44	0,69	0,17	0,18
24 Metalurgia	1,19	0,09	-0,88	-0,41	-1,80
25 Fabricação de produtos de metal, exceto máquinas e equipamentos	0,39	3,49	-1,26	0,23	0,92
26 Fabricação de equipamentos de informática, produtos	-1,21	0,57	-0,28	-0,61	0,84

Figura 2. Acesso ao banco de dados agregados (SIDRA). (IBGE, 2012).

5) Acesse a aba intitulada Produção (Figura 3);

The screenshot shows the IBGE SIDRA website interface. The browser address bar displays the URL: www.sidra.ibge.gov.br/bda/silvi/default.asp?t=2&z=t&o=29&u1=1&u2=1&u3=1&u4=1. The page title is "Banco de Dados Agregados". The main navigation bar includes "IBGE Home", "Escreva-nos", "Procurar Tabela", "Lista Conjuntural", "Índice", "Novidades", and "Ajuda". A search box prompts the user to "Digite o nro. da tabela:" with an "OK" button. Below this, there are tabs for "Dados Gerais", "Produção", "Valor", and "Árvores". The "Produção" tab is currently selected. A dropdown menu for "Brasil" is visible on the right side of the navigation bar.

Figura 3. Acesso à aba Produção do tema Silvicultura. (IBGE, 2012).

6) Acesse a tabela de quantidade produzida ao longo dos anos (Figura 4);

The screenshot shows the IBGE SIDRA website interface with the "Quantidade Produzida" table displayed. The browser address bar displays the URL: www.sidra.ibge.gov.br/bda/silvi/default.asp?t=2&z=t&o=29&u1=1&u2=1&u3=1&u4=1. The page title is "Banco de Dados Agregados". The main navigation bar includes "IBGE Home", "Escreva-nos", "Procurar Tabela", "Lista Conjuntural", "Índice", "Novidades", and "Ajuda". A search box prompts the user to "Digite o nro. da tabela:" with an "OK" button. Below this, there are tabs for "Dados Gerais", "Produção", "Valor", and "Árvores". The "Produção" tab is currently selected. A dropdown menu for "Brasil" is visible on the right side of the navigation bar.

Quantidade Produzida						
Brasil						
Grupo do produto	2010	2009	2008	2007	2006	2005
1.1 - Carvão vegetal (Toneladas)	3.448.210	3.378.492	3.975.393	3.806.044	2.608.847	2.526.437
1.2 - Lenha (Metros cúbicos)	49.058.232	41.410.850	42.037.848	39.089.275	36.110.455	35.542.255
1.3 - Madeira em tora (Metros cúbicos)	115.741.531	106.911.408	101.261.900	105.131.741	100.766.899	100.614.643
1.3.1 - Madeira em tora para papel e celulose (Metros cúbicos)	69.778.615	65.345.680	58.181.842	60.964.307	55.114.729	54.698.479
1.3.2 - Madeira em tora para outras finalidades (Metros cúbicos)	45.962.916	41.565.728	43.080.058	44.167.434	45.652.170	45.916.164

Figura 4. Acesso à tabela Quantidade Produzida. (IBGE, 2012).

Para os nossos trabalhos, selecionaremos a base de dados com as seguintes características:

I) Tipo de produto da silvicultura:

- a) Lenha;
- b) Madeira em tora para papel e celulose;
- c) Madeira em tora para outras finalidades.

II) Ano:

- a) Início em 2001 e término em 2010.

III) Unidade territorial:

- a) Todos os municípios do estado de São Paulo [Em São Paulo (549)].

A Figura 5 apresenta a elaboração da consulta desejada.

The screenshot shows the SIDRA system query interface. At the top, it displays the table name 'Tabela 793 - Quantidade produzida na silvicultura por tipo de produto da silvicultura' and the matrix dimensions 'Matriz multidimensional (1x20x12x3645) com 620.305 valores'. Below this, the 'Variável(1):' field is set to 'Quantidade produzida na silvicultura'. The 'Tipo de produto da silvicultura(9):' dropdown is set to '1.1 - Carvão vegetal (Toneladas)'. The 'Ano(12):' dropdown is set to '2009'. The 'Unidade Territorial(3645):' dropdown is set to 'Em São Paulo(549)'. Under 'Níveis Territoriais', the 'Grande Região(5):' is 'Não', 'Unidade da Federação(25):' is 'Não', 'Mesorregião Geográfica(122):' is 'Não', and 'Microrregião Geográfica(383):' is 'Não'. The 'Município(3109):' is 'Em São Paulo(549)'. The 'Opções de consulta:' section includes 'Visualizar (até 10.000 valores)', 'Preparar para impressão', 'Gerar link para consulta posterior', and 'Gravear' (with a note 'Veja as opções a posteriori efetuadas nos últimos 50 dias'). The 'Arquivo' dropdown is 'ProdSiVSPMunic1', 'Formato' is 'CSV (BR)', 'Modificidade' is 'Imediata (até 20.000 valores)', and 'E-mail' is empty. The 'Incluir nota de Rodapé' and 'Dimensionar com apenas uma seleção' checkboxes are checked. At the bottom, there are 'Reiniciar', 'OK', and '(Utilize Alt-o como atalho para o OK)' buttons.

Figura 5. Construção da consulta sobre produção da silvicultura na base SIDRA. (IBGE, 2012)

Durante o processo para montar a nossa base de dados é importante lembrar dos seguintes aspectos:

1) Marcar a opção “na linha” para cada variável selecionada, isto irá organizar a base de maneira que a variação dos valores de cada variável selecionada (tipo de produto, ano, unidade territorial) varie na linha e não nas colunas, mantendo a nossa base de dados organizada de maneira que cada coluna seja uma variável e cada linha contenha uma observação.

2) Marcar as opções “exibir código” e “exibir nome” na caixa de seleção das unidades territoriais, uma vez que o código de cada município é único. Tal informação irá facilitar o trabalho de junção desta base de dados com outras bases desagregadas por município, tal como o valor da produção.

3) Selecione a opção de gravação mais adequada ao seu caso. Neste exemplo, como a consulta realizada gera menos de 20.000 valores, podemos escolher a opção imediata, e dar um nome a nossa consulta.

4) Após o término da elaboração da consulta, clique no botão OK e a consulta gerada será exibida. Clique no nome do arquivo com o botão direito do mouse, escolha a opção “Salvar link como...” e salve a consulta gerada no seu computador (Figura 6).

Arquivo gravado com sucesso!!!

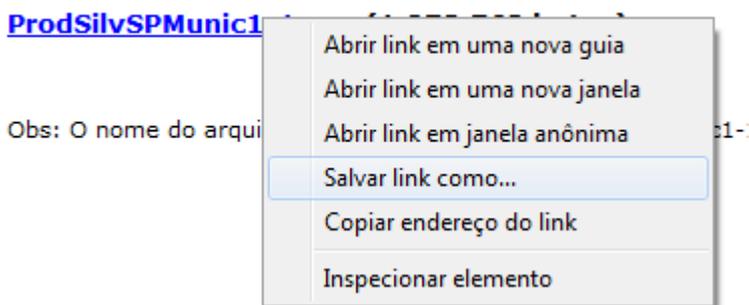


Figura 6. Salvar arquivo contendo a base de dados gerada na base SIDRA.

A base de dados deve ter apenas uma linha de cabeçalho contendo o nome das variáveis, todas as demais linhas devem conter as observações, não podendo haver linhas completamente em branco no meio da base.

- a) Abra a planilha eletrônica (LibreOffice Calc);
- b) vá em Arquivo = > Abrir;
- c) escolha o arquivo gerado pela consulta.

A planilha irá abrir um importador de texto, pois a base está salva utilizando um formato especial de apresentação de dados tabulados em arquivos de texto, que são valores separados por vírgula (comma-separated values – csv). Neste caso, como o idioma da máquina está definido como Português, os valores vêm separados por ponto e vírgula.

Selecione as seguintes opções na tela de importação da planilha:

- a) na seção Importar, escolha “da linha 3”, uma vez que as duas primeiras linhas do arquivo contêm cabeçalhos com informações da tabela e da consulta fornecidas pelo IBGE, e os cabeçalhos das variáveis aparecem na linha 3;
- b) em “opções de separadores”, desmarque todas as outras e selecione a opção “ponto e vírgula” (Figura 7);
- c) Clique em OK.

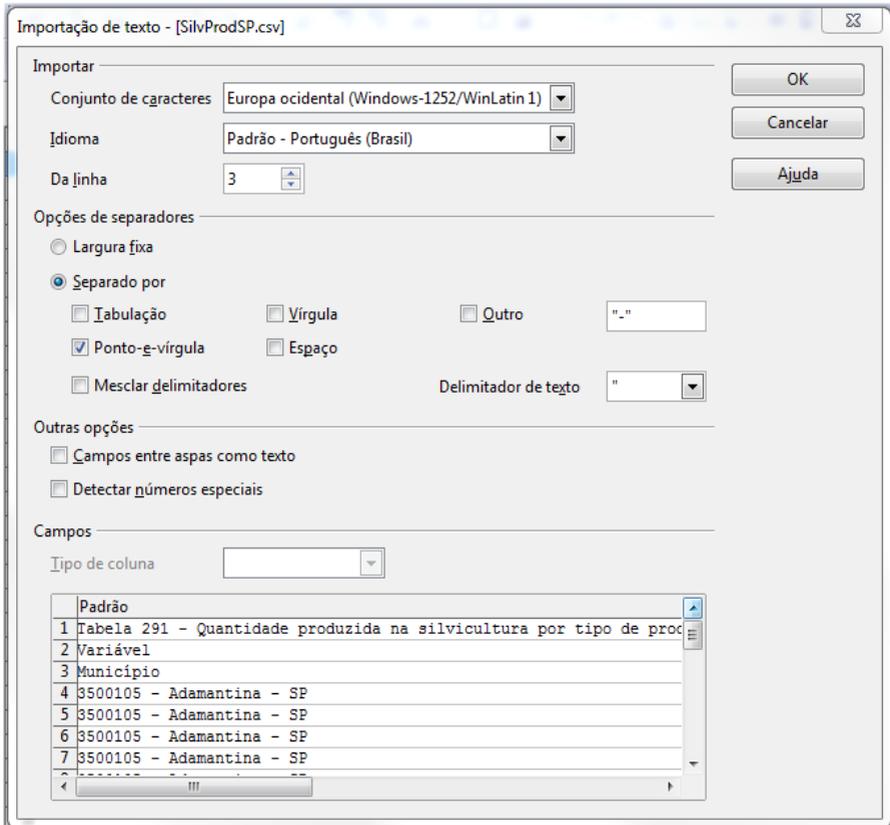


Figura 7. Importando o arquivo csv para a planilha eletrônica Calc do OfficeLibre 3.4.

Ao abrir o arquivo, digite o cabeçalho da variável quantidade produzida na célula D1, exclua as linhas no final da base com as notas de rodapé e a fonte, deixando apenas a base de dados com os cabeçalhos das variáveis na primeira linha e as observações nas linhas restantes. A base deverá ficar como apresentado na Figura 8.

	A	B	C	D
1	Município	Tipo de produto da silvicultura	Ano	Quantidade
2	3500105 - Adamantina - SP	1.2 - Lenha	2001	3689
3	3500105 - Adamantina - SP	1.2 - Lenha	2002	3400
4	3500105 - Adamantina - SP	1.2 - Lenha	2003	-0
5	3500105 - Adamantina - SP	1.2 - Lenha	2004	3300
16467	3557303 - Estiva Gerbi - SP	1.3.2 - Madeira em tora para outras finalidades	2006	-0
16468	3557303 - Estiva Gerbi - SP	1.3.2 - Madeira em tora para outras finalidades	2007	-0
16469	3557303 - Estiva Gerbi - SP	1.3.2 - Madeira em tora para outras finalidades	2008	-0
16470	3557303 - Estiva Gerbi - SP	1.3.2 - Madeira em tora para outras finalidades	2009	-0
16471	3557303 - Estiva Gerbi - SP	1.3.2 - Madeira em tora para outras finalidades	2010	-0

Figura 8. Formato inicial da base de dados de produção da silvicultura.

A nossa base conta com 16.470 registros e uma linha de cabeçalhos, totalizando 16.471 linhas na planilha eletrônica. Salve o arquivo como uma planilha eletrônica [Planilha ODF (.ods)] em um local de sua preferência. O próximo passo é realizar o tratamento dos registros da nossa base de dados para facilitar a sua análise, a sua junção com outras bases de dados e a obtenção de informações.

Tratamento de bases de dados

Optou-se por abordar o tema de tratamento da base de dados em planilha eletrônica nesta publicação devido à necessidade que se encontra de separar ou unir valores das observações para as diferentes variáveis de uma base de dados. Quando a base apresenta poucos registros, tais modificações podem ser realizadas manualmente, registro a registro. Entretanto, quando a base de dados é composta de milhares de registros, torna-

se necessário automatizar esta tarefa, tanto para aumentar a eficiência do processo como para minimizar a ocorrência de erros no resultado final. Neste documento, enfatizaremos o tratamento de bases de dados utilizando funções e ferramentas disponíveis na planilha eletrônica.

Utilizando funções da planilha eletrônica

A variável Município da nossa base de dados de produção da silvicultura possui três informações (variáveis) agregadas em uma única variável: o código do município, o nome do município e a sigla da Unidade da Federação (UF). É interessante separar os três valores em três novas colunas, facilitando a união posterior desta base de dados com outras de interesse que tenham informações discriminadas por município, além de facilitar a obtenção de informações agregadas por UF.

O primeiro passo para se proceder a separação de valores de uma variável é buscar um padrão de separação que se repita em todos os registros da base. O código do município é uma sequência de caracteres dada pelo IBGE, com comprimento de sete caracteres. Então, iremos construir uma fórmula que retorne os sete primeiros caracteres a partir da esquerda da sequência de caracteres da variável município. Insira três novas colunas entre as colunas das variáveis município e tipo de produto da silvicultura, para o armazenamento das variáveis código do município, nome do município e sigla da UF; e renomeie as três colunas, sendo a primeira como CodMunicípio, a segunda como NomeMunicípio e a terceira como SiglaUF. A fórmula a ser inserida na célula B2 é `ESQUERDA(A2;7)`. O resultado da fórmula será exibido na célula B2, sendo o valor 3500105 para o primeiro registro da nossa base. Copie a fórmula para todos os registros ao longo da coluna B para obter os valores dos códigos de municípios em separado. A Figura 9 apresenta a base de dados após estas alterações, e destaca a fórmula utilizada em detalhe.

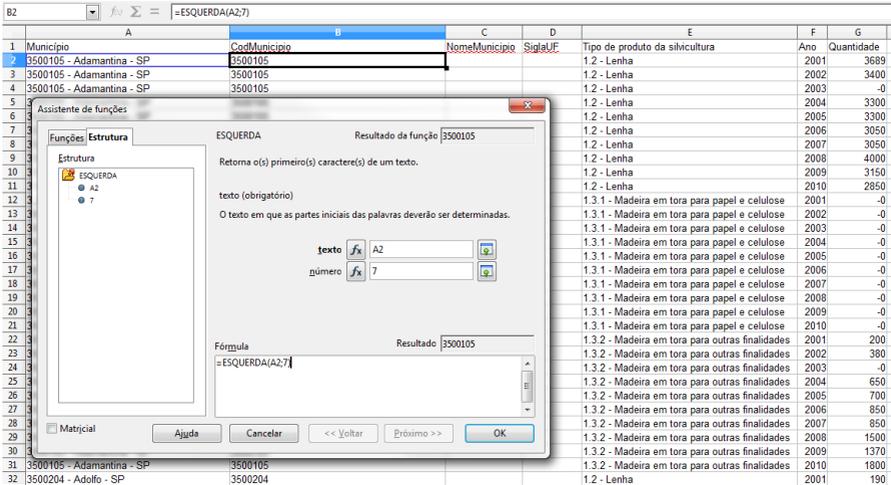


Figura 9. Separando o código dos municípios dos registros da base de dados.

O próximo passo é separar o nome do município de cada registro e inseri-lo na variável NomeMunicípio. Para isto utilizaremos a função EXT.TEXTO() da categoria texto, que retorna uma sequência de caracteres a partir de um texto, onde informamos a função, qual a posição que se inicia a sequência de caracteres que desejamos e o seu comprimento. Poderíamos construir uma fórmula mais complexa que identificasse a posição de início do nome do município e a posição de término do mesmo, sendo o seu comprimento dado pela diferença entre as duas posições.

Entretanto, podemos observar um padrão que foi seguido para construção do nome da variável município, sendo composta pela inserção inicial do código do município com sete caracteres (definido pelo IBGE), seguido da sequência de três caracteres dados por " - ", seguido pelo nome do município, acrescido de uma nova sequência de três caracteres " - ", terminando com a sigla da UF, que apresenta sempre dois caracteres. Desta forma, mesmo que não saibamos o comprimento da sequência de caracteres do nome do município, sabemos que ela será igual ao comprimento total da sequência de caracteres da

variável município menos quinze caracteres (sete do código do município mais seis das duas sequências separadoras acrescidos de dois caracteres da sigla da UF). A Figura 10 apresenta a fórmula utilizada para a obtenção do nome do município dos registros da variável município, a ser inserida na célula C2 [ESQUERDA(A2;11;NÚM.CARACT(A2)-15)]. Perceba que a fórmula se ajustou ao comprimento dos valores das diferentes observações, extraindo sempre o nome correto de cada município.

The image shows a spreadsheet with the following data:

	A	B	C	D
1	Município	CodMunicípio	NomeMunicípio	SiglaUF
2	3500105 - Adamantina - SP	3500105	Adamantina	
3	3500105 - Adamantina - SP	3500105	Adamantina	
4	3500105 - Adamantina - SP			
5	3500105 - Adamantina - SP			
6	3500105 - Adamantina - SP			
7	3500105 - Adamantina - SP			
8	3500105 - Adamantina - SP			
9	3500105 - Adamantina - SP			
10	3500105 - Adamantina - SP			
11	3500105 - Adamantina - SP			
12	3500105 - Adamantina - SP			
13	3500105 - Adamantina - SP			
14	3500105 - Adamantina - SP			
15	3500105 - Adamantina - SP			
16	3500105 - Adamantina - SP			
17	3500105 - Adamantina - SP			
18	3500105 - Adamantina - SP			
19	3500105 - Adamantina - SP			
20	3500105 - Adamantina - SP			
21	3500105 - Adamantina - SP			
22	3500105 - Adamantina - SP			
23	3500105 - Adamantina - SP			
24	3500105 - Adamantina - SP			
25	3500105 - Adamantina - SP			
26	3500105 - Adamantina - SP			
27	3500105 - Adamantina - SP			
28	3500105 - Adamantina - SP			
29	3500105 - Adamantina - SP			
30	3500105 - Adamantina - SP	3500105	Adamantina	
31	3500105 - Adamantina - SP	3500105	Adamantina	
32	3500204 - Adolfo - SP	3500204	Adolfo	
33	3500204 - Adolfo - SP	3500204	Adolfo	

The dialog box for the 'EXT.TEXTO' function is open, showing the formula '=EXT.TEXTO(A2;11;NÚM.CARACT(A2)-15)' and the result 'Adamantina'. The dialog box also shows the function's structure and the text '(obligatório) O texto em que as palavras parciais deverão ser determinadas.'

Figura 10. Obtendo o nome dos municípios dos registros da base de dados.

Para obtenção da sigla da UF, iremos construir uma fórmula para extrair os dois últimos caracteres do texto, ou os dois primeiros caracteres a partir da direita. A função utilizada na fórmula será a função DIREITA() da categoria texto, a qual será inserida na planilha conforme apresentado na Figura 11.

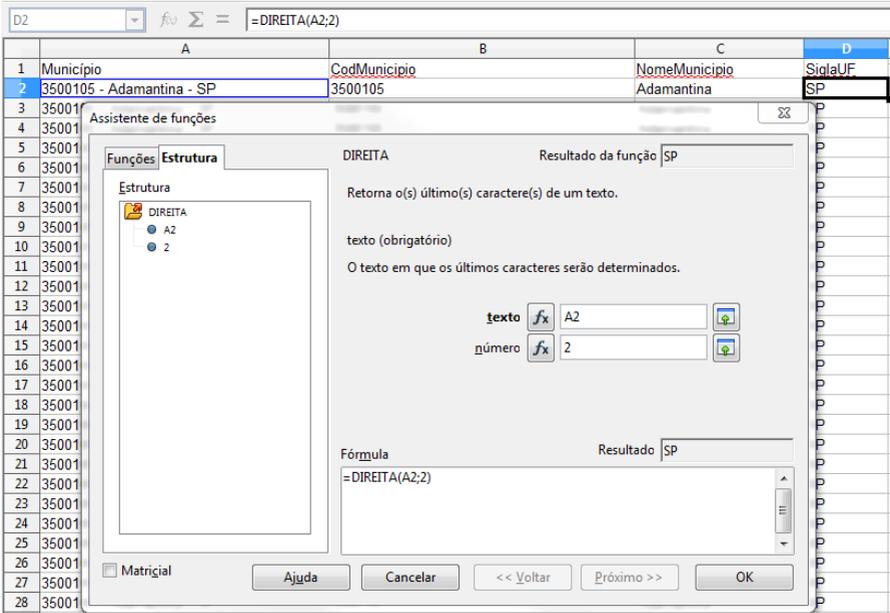


Figura 11. Obtendo a sigla da unidade da federação (UF) dos registros da base de dados.

Após estas modificações, a nossa base de dados deverá ficar conforme apresentado na Figura 12.

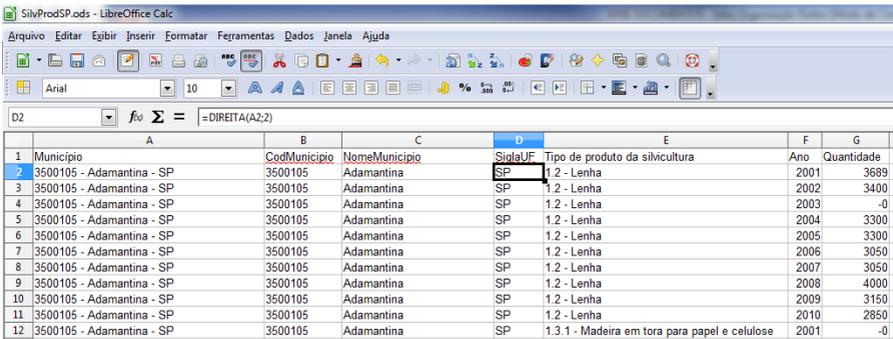


Figura 12. Aparência da base de dados após a separação da variável município.

Para facilitar a junção da base de dados de quantidade produzida da silvicultura com a base de dados de valor da produção da silvicultura, vamos separar os valores de código de produto da silvicultura e descrição do tipo de produto da silvicultura, que estão unidos em uma única variável na coluna E (tipo de produto da silvicultura).

Inicialmente insira duas novas colunas entre as colunas E e F, e as intitule CodProduto e DescProduto, respectivamente. Para buscar um padrão na composição da escrita das observações da variável tipo de produto da silvicultura, vamos utilizar a ferramenta AutoFiltro da planilha eletrônica. Clique em qualquer célula da base de dados, vá no menu Dados => Filtro => AutoFiltro para fazer aparecer as caixas de listagem contendo as opções de filtro para cada variável. Clique na caixa de listagem da variável tipo de produto da silvicultura e observe que apenas três valores distintos são exibidos (Figura 13), de modo que a fórmula a ser elaborada precisará se adequar apenas a estes três registros distintos para que a mesma seja aplicável em toda a base de dados.

	A	B	C	D	E
1	Município	CodMunicípio	NomeMunicípio	SiglaU	Tipo de produto da silvicultura
2	3500105 - Adamantina - SP	3500105	Adamantina	SP	Todas
3	3500105 - Adamantina - SP	3500105	Adamantina	SP	10 primeiros
4	3500105 - Adamantina - SP	3500105	Adamantina	SP	Filtro padrão...
5	3500105 - Adamantina - SP	3500105	Adamantina	SP	- em branco -
6	3500105 - Adamantina - SP	3500105	Adamantina	SP	- preenchida -
7	3500105 - Adamantina - SP	3500105	Adamantina	SP	1.2 - Lenha
8	3500105 - Adamantina - SP	3500105	Adamantina	SP	1.3.1 - Madeira em tora para papel e celulose
9	3500105 - Adamantina - SP	3500105	Adamantina	SP	1.3.2 - Madeira em tora para outras finalidades
10	3500105 - Adamantina - SP	3500105	Adamantina	SP	
11	3500105 - Adamantina - SP	3500105	Adamantina	SP	

Figura 13. Registros únicos de tipos de produto da silvicultura.

Primeiramente observamos que o comprimento da sequência de caracteres que representa o código do produto muda, apresentando comprimentos variados para as diferentes observações da nossa base de dados. Entretanto, todos os códigos de produto terminam no momento em que aparece o primeiro espaço da sequência de caracteres, sendo este um padrão que podemos utilizar para a elaboração da nossa fórmula. Sendo assim, elaboraremos uma fórmula que retorne a sequência de caracteres de comprimento igual à posição do primeiro espaço menos um, contados a partir da esquerda. A fórmula a ser inserida na célula F2 é `ESQUERDA(E2;PROCURAR(" ";E2;1)-1)`. Copie e cole a fórmula para todos os registros da base ao longo da coluna F e teremos os códigos de produto da silvicultura em uma variável, na coluna CodProduto (Figura 14).

Para separação da descrição do tipo de produto da silvicultura, vamos elaborar uma fórmula que retorne a sequência de caracteres com comprimento igual ao comprimento total da sequência de caracteres de cada observação da variável tipo de produto da silvicultura menos o número de caracteres da posição do primeiro espaço menos dois (que são os caracteres referentes ao hífen e ao segundo espaço do conjunto de caracteres que separam as duas variáveis), contando a partir da direita. A fórmula a ser utilizada na célula G2 será a seguinte: `DIREITA(E2;NÚM.CARACT(E2) - PROCURAR(" ";E2;1) - 2)`. O resultado observado após a inserção destas fórmulas em todas as observações da base ao longo da coluna G, está apresentado na Figura 15.

The screenshot shows the Libre Office Calc interface with a spreadsheet and a dialog box for the ESQUERDA function. The spreadsheet has columns A through F. Column A contains municipality names, B contains CodMunicipio, C contains NomeMunicipio, D contains SiglaUF, E contains Tipo de produto da silvicultura, and F contains CodProduto. The dialog box is titled 'Assistente de funções' and shows the 'ESQUERDA' function. The 'texto' field contains 'E2' and the 'número' field contains 'PROCURAR(\" \";E2;1)-1'. The 'Resultado' field shows '1,2'. The formula bar at the bottom of the dialog displays '=ESQUERDA(E2;PROCURAR(\" \";E2;1)-1)'. The dialog also includes a tree view of functions, a 'Matricial' checkbox, and buttons for 'Ajuda', 'Cancelar', '<< Voltar', 'Próximo >>', and 'OK'.

Figura 14. Separação dos códigos de produtos da silvicultura.

The screenshot shows the Libre Office Calc interface with a spreadsheet and a dialog box for the DIREITA function. The spreadsheet has columns A through G. Column A contains municipality names, B contains CodMunicipio, C contains NomeMunicipio, D contains SiglaUF, E contains Tipo de produto da silvicultura, F contains CodProduto, and G contains DescProduto. The dialog box is titled 'Assistente de funções' and shows the 'DIREITA' function. The 'texto' field contains 'E2' and the 'número' field contains 'NÚM.CARACT(E2)-PROG'. The 'Resultado' field shows 'Lenha'. The formula bar at the bottom of the dialog displays '=DIREITA(E2;NÚM.CARACT(E2)-PROG)'. The dialog also includes a tree view of functions, a 'Matricial' checkbox, and buttons for 'Ajuda', 'Cancelar', '<< Voltar', 'Próximo >>', and 'OK'.

Figura 15. Separação da descrição dos produtos da silvicultura.

Para finalizar o tratamento da base de dados de quantidade produzida da silvicultura, vamos eliminar as fórmulas da base e excluir as colunas com as variáveis município e tipo de produto da silvicultura, uma vez que os valores destas variáveis já estão inseridos nas variáveis CodMunicipio, NomeMunicipio, SiglaUF, CodProduto e DescProduto. Copie todos os registros da base

de dados e cole na mesma posição, utilizando a opção “colar especial”, desmarcando as opções “colar tudo” e “fórmulas” (Figura 16). Em seguida exclua as colunas contendo as variáveis Município e Tipo de produto da silvicultura. O resultado final pode ser observado na Figura 17.

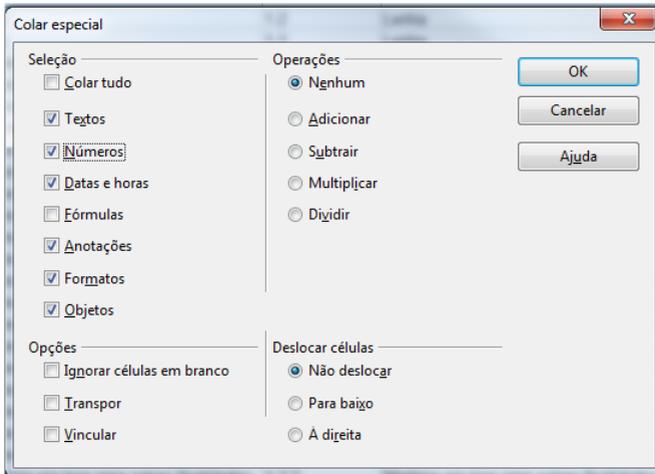


Figura 16. Colar especial para eliminar as fórmulas dos registros da base de dados.

	A	B	C	D	E	F	G
1	CodMunicípio	NomeMunicípio	SiglaUF	CodProduto	DescProduto	Ano	Quantidade
2	3500105	Adamantina	SP	1.2	Lenha	2001	3689
3	3500105	Adamantina	SP	1.2	Lenha	2002	3400
4	3500105	Adamantina	SP	1.2	Lenha	2003	-0
5	3500105	Adamantina	SP	1.2	Lenha	2004	3300
6	3500105	Adamantina	SP	1.2	Lenha	2005	3300
7	3500105	Adamantina	SP	1.2	Lenha	2006	3050
8	3500105	Adamantina	SP	1.2	Lenha	2007	3050
9	3500105	Adamantina	SP	1.2	Lenha	2008	4000
10	3500105	Adamantina	SP	1.2	Lenha	2009	3150
11	3500105	Adamantina	SP	1.2	Lenha	2010	2850
12	3500105	Adamantina	SP	1.3.1	Madeira em tora para papel e celulose	2001	-0

Figura 17. Aspecto final da base de dados de quantidade produzida da silvicultura.

Passaremos agora a efetuar o mesmo tratamento já realizado com a base de dados de quantidade produzida da silvicultura para a base de dados de valor da produção da silvicultura, também obtida no site do IBGE com os mesmos parâmetros utilizados na construção da consulta, com a diferença de que ela será realizada na tabela de valor da produção e não de quantidade produzida.

Utilizando outras ferramentas da planilha eletrônica

A utilização das ferramentas de tratamento de dados disponibilizadas pelas planilhas eletrônicas (no caso o LibreOffice versão 3.4) permitem a separação do conteúdo de variáveis da nossa base de dados de maneira muito mais rápida e eficiente do que aquela realiza por meio de fórmulas. Entretanto, consideramos a apresentação do primeiro método de extrema importância, pois permitiu apresentar o raciocínio de elaboração das fórmulas para obtenção das sequências de caracteres desejadas, possibilitando ao usuário a futura utilização de tais funções para obtenção de subsequências de caracteres que possam envolver outras situações além da separação de valores compostos de uma variável. Além disso, as ferramentas apresentadas a seguir não estavam disponíveis em versões anteriores das planilhas eletrônicas, e a absorção deste conhecimento permitirá ao usuário a separação de variáveis em versões de planilhas eletrônicas mais antigas.

Para obter a base de dados de valor da produção, acesse o SIDRA no site do IBGE seguindo os mesmos passos apresentados nas Figuras 1 e 2, sendo que a aba selecionada será a de valor ao invés de produção (Figura 18). Monte o quadro respeitando as mesmas opções descritas no passo 6 (página 16) e salve o arquivo com o nome de `SilvValorSP.csv` no mesmo local onde foi salvo o arquivo `SilvProdSP.csv`.

Dados Gerais		Produção		Valor		Árvores	
Valor da Produção 				Brasil <input type="text" value="Brasil"/>			
Valor da Produção - Mil Reais							
Brasil							
Grupo do produto	2010	2009	2008	2007	2006	2005	
1.1 - Carvão vegetal	1.685.924	1.494.748	2.024.014	1.585.241	955.382	831.106	
1.2 - Lenha	1.653.710	1.344.232	1.258.005	1.112.621	902.882	827.195	
1.3 - Madeira em tora	7.231.123	6.082.631	5.423.829	5.480.764	5.213.896	4.958.517	
1.3.1 - Madeira em tora para papel e celulose	3.841.347	3.255.171	2.760.266	2.836.487	2.548.486	2.337.218	
1.3.2 - Madeira em tora para outras finalidades	3.389.775	2.827.460	2.663.563	2.644.276	2.665.410	2.621.299	
2 - Outros produtos	139.676	89.344	83.141	99.012	127.200	170.852	
2.1 - Acácia-negra (casca)	9.586	10.554	14.064	18.201	29.841	31.933	
2.2 - Eucalipto (folha)	4.064	2.459	2.245	1.745	3.096	3.701	
2.3 - Resina	126.026	76.331	66.832	79.065	94.263	135.218	
Fonte: Produção da Extração Vegetal e da Silvicultura							

Figura 18. Tabela para consulta dos dados de valor da produção da silvicultura.

Abra o arquivo SilvValorSP.csv e realize a importação dos dados para a planilha eletrônica a partir da linha 3, conforme apresentado na Figura 7. Em seguida, insira o nome do cabeçalho da variável Valor (Mil R\$) na célula D1 e exclua as duas últimas linhas da planilha, pois elas não possuem observações, apenas as notas de rodapé. Salve a planilha com o mesmo nome, mas com uma extensão de planilha eletrônica [Planilha ODF (.ods)] no mesmo local do arquivo SilvProdSP.ods. A forma inicial da nossa base de dados de valor da produção da silvicultura deverá ser semelhante à Figura 19.

	A	B	C	D
1	Município	Tipo de produto da silvicultura	Ano	Valor (Mil R\$)
2	3500105 - Adamantina - SP	1.2 - Lenha	2001	23
3	3500105 - Adamantina - SP	1.2 - Lenha	2002	51
4	3500105 - Adamantina - SP	1.2 - Lenha	2003	-0
5	3500105 - Adamantina - SP	1.2 - Lenha	2004	33
6	3500105 - Adamantina - SP	1.2 - Lenha	2005	50
16464	3557303 - Estiva Gerbi - SP	1.3.2 - Madeira em tora para outras finalidades	2003	-0
16465	3557303 - Estiva Gerbi - SP	1.3.2 - Madeira em tora para outras finalidades	2004	-0
16466	3557303 - Estiva Gerbi - SP	1.3.2 - Madeira em tora para outras finalidades	2005	-0
16467	3557303 - Estiva Gerbi - SP	1.3.2 - Madeira em tora para outras finalidades	2006	-0
16468	3557303 - Estiva Gerbi - SP	1.3.2 - Madeira em tora para outras finalidades	2007	-0
16469	3557303 - Estiva Gerbi - SP	1.3.2 - Madeira em tora para outras finalidades	2008	-0
16470	3557303 - Estiva Gerbi - SP	1.3.2 - Madeira em tora para outras finalidades	2009	-0
16471	3557303 - Estiva Gerbi - SP	1.3.2 - Madeira em tora para outras finalidades	2010	-0
16472				

Figura 19. Formato inicial da base de dados de valor da produção da silvicultura.

Aparentemente, as variáveis código e nome de município e sigla da UF são separadas por um hífen. Para confirmarmos se a nossa suspeita está correta, vamos construir uma fórmula que conte a quantidade de hifens em cada observação da variável município. Se o resultado for dois, então temos um forte indício de que podemos utilizar o hífen como separador entre as variáveis. Como não há uma função que conte a quantidade de um determinado caractere dentro de um texto, utilizaremos uma função que conte o total de caracteres do texto e subtraia este valor da quantidade de caracteres do texto sem os hifens, de modo que o resultado será a quantidade de hifens contida no texto. Para isto, utilizaremos as funções NÚM.CARACT() e SUBSTITUIR() em uma nova coluna, inserida entre as variáveis município e tipo de produto da silvicultura, intitulada verifica (Figura 20).

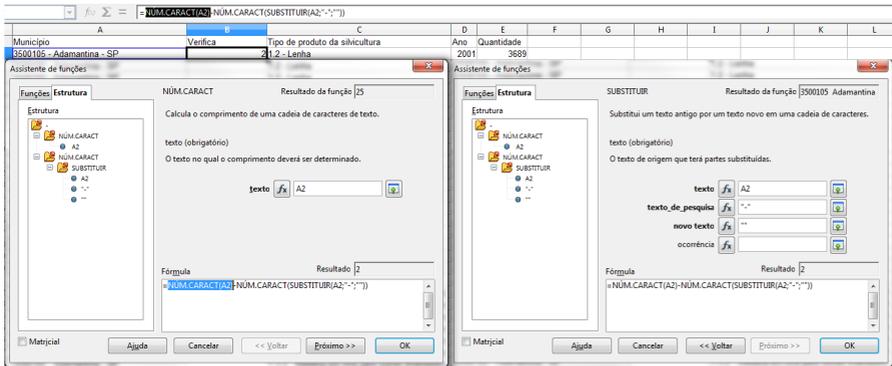


Figura 20. Fórmula para contar o número de hifens dos registros da variável município.

A fórmula apresentada na Figura 20 é composta por dois termos, sendo que o primeiro possui uma função e o segundo duas funções. O primeiro termo calcula o comprimento da cadeia de caracteres contido na célula A2 [NÚM.CARACT(A2)], e o segundo termo é composto de duas funções encapsuladas. Primeiro, substituímos o caractere hífen por um caractere vazio, em seguida calculamos o comprimento da nova cadeia de caracteres sem os hifens [NÚM.CARACT(SUBSTITUIR(A2;"-";""))]. Um argumento de texto (string) sempre deve ser passado para a função entre aspas duplas, no terceiro argumento da função SUBSTITUIR() é um texto vazio. O resultado da nossa fórmula é 2, ou seja, existem dois hifens na célula A2.

Ao copiarmos a fórmula para as demais observações da planilha e utilizarmos a ferramenta de AutoFiltro, podemos verificar se há a ocorrência de registros com mais de dois hifens na nossa base de dados. Ao clicarmos na caixa de listagem contida no cabeçalho da variável verifica, observamos a ocorrência de números 2 e 3, indicando que existem registros cuja cadeia de caracteres da variável município possuem mais de dois hifens, de modo que a utilização do hífen para separar o código do município geraria uma separação errada em alguns registros

da nossa base. Um segundo olhar permite identificar que a sequência de caracteres composta por um espaço, um hífen e outro espaço pode ser utilizada para separar as variáveis.

Para testarmos esta nova hipótese, vamos modificar a nossa fórmula para substituir a sequência de caracteres " - " ao invés de "-". Copie esta nova fórmula para todos os registros e observe a listagem de filtros novamente. Perceba que apenas o número 6 aparece na listagem, ou seja, existem apenas duas sequências de caracteres " - " em todos os registros da nossa base de dados (cada uma com três caracteres). Sendo assim, podemos utilizar esta sequência de caracteres para separar os valores de código de município, nome de município e sigla da UF da nossa base de dados, substituindo-os por um caractere válido como separador de campos na planilha eletrônica.

Os separadores de campos mais utilizados em tabulação de dados em arquivos de texto são apresentados na Figura 21. Como já sabemos que o separador de campos utilizado nas observações da nossa base de dados é o conjunto de três caracteres compostos por " - ", vamos substituí-los por um separador de caracteres válido que possa ser utilizado pela ferramenta de texto para colunas da planilha eletrônica, no caso a sugestão é o ponto e vírgula.

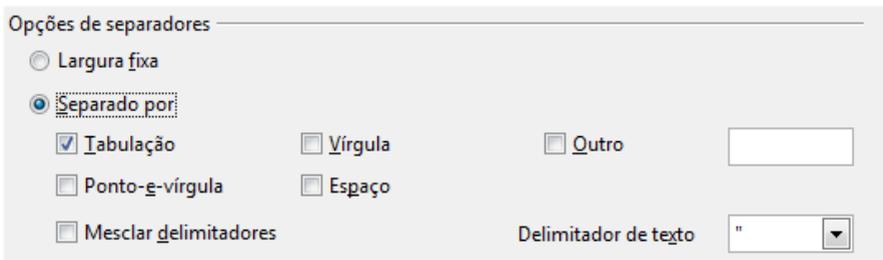


Figura 21. Conjunto de caracteres válidos para a tabulação de dados em arquivos texto.

Inicialmente altere os cabeçalhos das variáveis município e tipo de produto da silvicultura. Substitua o cabeçalho “Município” pelo cabeçalho “CodMunicipio;NomeMunicipio;SiglaUF”, e o cabeçalho “Tipo de produto da silvicultura” por “CodProduto;DescProduto”. Em seguida selecione as colunas A e B, clique no menu Editar => Localizar e Substituir e solicite a substituição dos caracteres “ - ” pelo caractere “;”, conforme ilustrado na Figura 22.

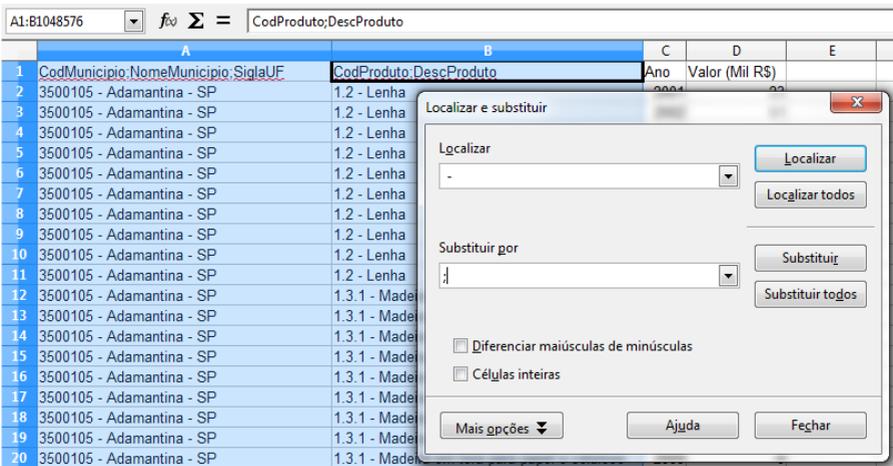


Figura 22. Alterando os caracteres separadores das variáveis.

Após a substituição, insira duas novas colunas entre as colunas A e B. Selecione a coluna A e clique no menu dados => texto para colunas. A inserção das duas novas colunas é necessária porque a ferramenta irá separar o conteúdo da coluna A de acordo com o número de caracteres separadores encontrados em cada observação. Como temos dois pontos e vírgulas em cada linha, os valores contidos na coluna A serão separados em três colunas, sendo a primeira a coluna já ocupada pelos registros (coluna A) e as duas seguintes

pelas colunas imediatamente a direita da coluna A (no caso as colunas B e C). A ferramenta substitui os valores encontrados nas colunas caso estas já estejam sendo utilizadas, de maneira que se não inserirmos duas novas colunas vazias à direita da coluna A, os valores originais das colunas B e C (tipo de produto e ano) serão substituídos, acarretando em perda de dados da nossa base.

Na janela de texto para colunas, execute o seguinte procedimento:

- 1) Escolha o ponto-e-vírgula como opção de separador, desmarcando as outras opções de separadores;
- 2) Na área intitulada campos, selecione a primeira coluna e a opção tipo de coluna será habilitada. Escolha o tipo de coluna texto;
- 3) Repita este procedimento para as outras duas colunas;
- 4) Clique no botão OK para realizar a separação dos valores.

O procedimento acima pode ser observado na Figura 23.

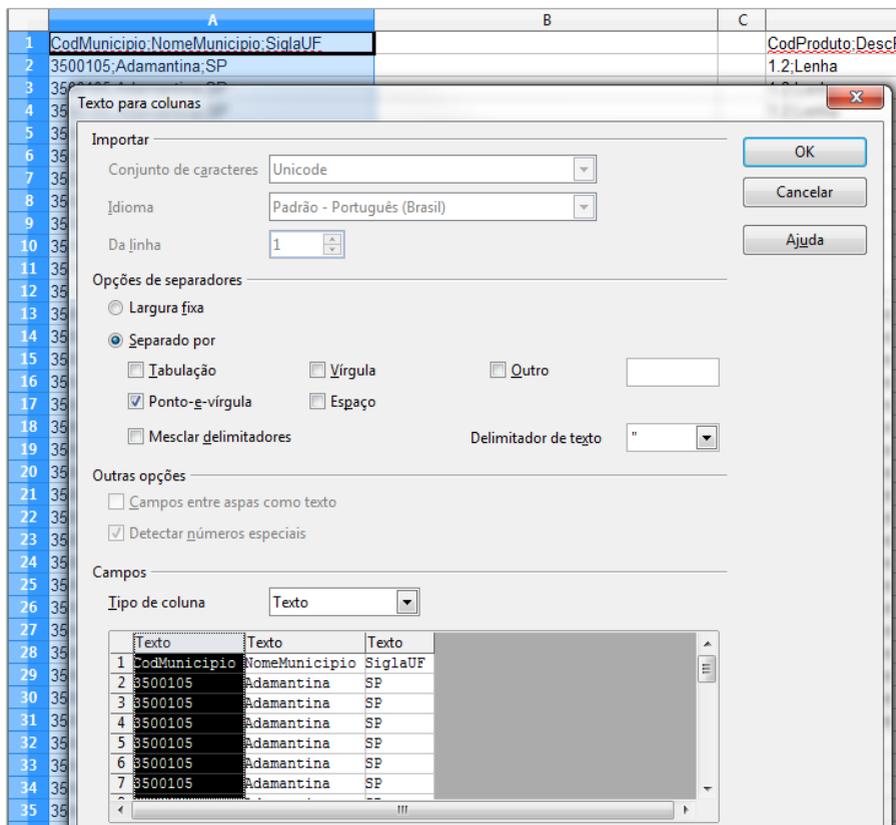


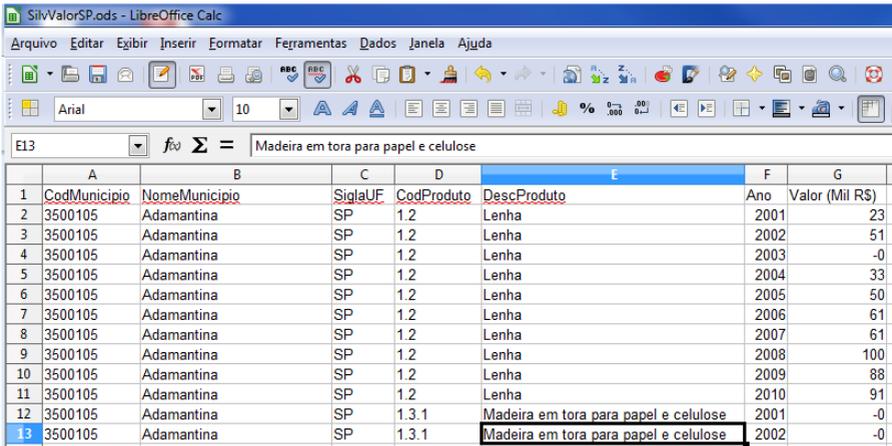
Figura 23. Configurando a janela texto para colunas.

A Figura 24 apresenta o resultado deste procedimento na nossa base de dados, onde os valores da coluna A estão agora divididos nas colunas A, B e C.

	A	B	C	D	E	F	G
1	CodMunicípio	NomeMunicípio	SiglaUF	CodProduto; DescProduto	Ano	Valor (Mil R\$)	
2	3500105	Adamantina	SP	1.2;Lenha	2001	23	
3	3500105	Adamantina	SP	1.2;Lenha	2002	51	
4	3500105	Adamantina	SP	1.2;Lenha	2003	-0	
5	3500105	Adamantina	SP	1.2;Lenha	2004	33	
6	3500105	Adamantina	SP	1.2;Lenha	2005	50	
7	3500105	Adamantina	SP	1.2;Lenha	2006	61	

Figura 24. Resultado da separação dos valores da variável município.

Vamos agora repetir o procedimento para os valores da coluna D, separando o código dos produtos da silvicultura da descrição dos produtos da silvicultura da nossa base de dados. Insira uma coluna em branco entre as colunas D e E, selecione a coluna D e execute a ferramenta texto para colunas, mantendo a mesma configuração utilizada no procedimento anterior. O resultado pode ser observado na Figura 25.



The screenshot shows the LibreOffice Calc interface with the spreadsheet 'SilvValorSP.ods'. The spreadsheet has been updated with an additional column (E) for product description. The data is as follows:

	A	B	C	D	E	F	G
1	CodMunicípio	NomeMunicípio	SiglaUF	CodProduto	DescProduto	Ano	Valor (Mil R\$)
2	3500105	Adamantina	SP	1.2	Lenha	2001	23
3	3500105	Adamantina	SP	1.2	Lenha	2002	51
4	3500105	Adamantina	SP	1.2	Lenha	2003	-0
5	3500105	Adamantina	SP	1.2	Lenha	2004	33
6	3500105	Adamantina	SP	1.2	Lenha	2005	50
7	3500105	Adamantina	SP	1.2	Lenha	2006	61
8	3500105	Adamantina	SP	1.2	Lenha	2007	61
9	3500105	Adamantina	SP	1.2	Lenha	2008	100
10	3500105	Adamantina	SP	1.2	Lenha	2009	88
11	3500105	Adamantina	SP	1.2	Lenha	2010	91
12	3500105	Adamantina	SP	1.3.1	Madeira em tora para papel e celulose	2001	-0
13	3500105	Adamantina	SP	1.3.1	Madeira em tora para papel e celulose	2002	-0

Figura 25. Formato final da base de dados de valor da produção da silvicultura.

A nossa próxima missão será unir as bases de dados de quantidade produzida e valor da produção da silvicultura.

União de bases de dados

Muitas vezes gostaríamos de aumentar o nível de detalhamento das informações obtidas a partir de um conjunto de dados a nossa disposição, sendo necessário inclusive o uso de bases de dados diferentes para a geração das informações desejadas. Em relação ao nosso exemplo, se quisermos saber a evolução do preço médio da madeira em tora para celulose no estado de São Paulo, teríamos que calcular o total anual da produção, o total anual do valor da produção, e dividir o valor pela quantidade em cada ano. Este processo pode parecer simples, mas caso seja necessário obter a série com a evolução do preço médio da madeira para cada produto, ou até por microrregião ou mesorregião do Estado de São Paulo (informação ainda não inserida na nossa base de dados), seria necessário uma série de cálculos e manipulações dos dados na base, que tomariam muito tempo e estariam sujeitos a erros durante o seu processo de execução.

Unir o total resultante das duas bases de dados registro a registro não seria problema, uma vez que teríamos apenas dez registros (um para cada ano). Entretanto, fazê-lo para a base original inteira (mais de 16 mil registros) seria algo impensável e sujeito a erros de digitação, o que poderia comprometer seriamente a qualidade das análises obtidas a partir desta base de dados. Poderíamos, simplesmente, copiar e colar os registros da coluna de valor da produção ao lado da coluna de quantidade produzida, mas se os dados não estiverem exatamente na mesma ordem, estaríamos atribuindo valores da produção que não se referem aos mesmos municípios, produtos e anos das quantidades produzidas, gerando uma verdadeira confusão na nossa base de dados.

As técnicas que serão apresentadas a seguir possibilitam a junção de bases de dados sem a preocupação com a ordem dos

registros, apenas com a correta construção da correlação das bases de dados a serem unidas.

Conceito de relacionamento entre bases de dados

Antes de proceder à união das bases de dados de quantidade produzida e valor da produção da silvicultura, vamos aprender alguns conceitos básicos utilizados no relacionamento de tabelas em bancos de dados relacionais: o conceito de chave (primária e estrangeira) e o conceito de tipos de relacionamento entre tabelas (um para um, um para muitos, muitos para muitos).

Em um banco de dados relacional, as chaves estão diretamente relacionadas com a capacidade de recuperação e manutenção da integridade dos registros. Uma chave primária apresenta o valor ou um conjunto de valores de uma ou mais variáveis que tornam aquele registro exclusivo em toda a tabela, sendo um identificador único do registro (UNIVERSIDADE DO ESTADO DE SÃO PAULO, 2012). As chaves estrangeiras são utilizadas para se estabelecer uma relação entre diferentes tabelas em um banco de dados relacional.

As relações entre tabelas podem ser de três tipos (UNIVERSIDADE DO ESTADO DE SÃO PAULO, 2012):

- Um para um – para cada registro da tabela 1 existe um único registro da tabela 2;
- Um para muitos – para cada registro da tabela 1 existem muitos registros na tabela 2;
- Muitos para muitos – cada registro da tabela 1 pode ter muitos registros na tabela 2 e vice-versa.

As duas bases de dados que queremos unir são um exemplo de um relacionamento um para um, pois para cada registro de quantidade produzida na silvicultura existe apenas um único

registro correspondente de valor da produção. A união da série contendo o índice geral de preços – disponibilidade interna (IGP-DI) com a base de dados de valor da produção seria um exemplo de um relacionamento um para muitos, para cada IGP-DI de cada ano haveriam muitos registros com o mesmo ano na base de dados de valor da produção. Em termos de planilha eletrônica, não há como fazer um relacionamento de muitos para muitos entre duas tabelas com uma única operação. Tal relacionamento resultaria da utilização de duas operações, utilizando as relações um para um ou um para muitos, dependendo do caso. No entanto, não vamos aprofundar na descrição deste tipo de relacionamento nesta publicação.

Em planilhas eletrônicas, as chaves primárias e secundárias não existem e nem estão pré-definidas, de maneira que o usuário que deseje utilizar estes conceitos para unir duas bases de dados deve identificar quais campos na sua tabela tornam os registros exclusivos, sendo esta então a sua chave primária. O importante é definir a chave primária da tabela que possui a(s) variável(eis) que desejamos incorporar à nossa base de dados, para então definir quais variáveis serão utilizadas na tabela da nossa base de dados como chave estrangeira para construção do relacionamento entre as duas tabelas.

Unindo bases de dados com um relacionamento um para um

No caso das bases de dados de quantidade produzida e valor da produção, vamos utilizar a tabela de quantidade produzida como base e trazer para ela a variável valor (Mil R\$) da tabela de valor da produção. Na tabela de valor da produção, a combinação de variáveis que apresenta registros únicos é dada pelo código do município (CodMunicipio), código do produto (CodProduto) e pelo ano. Vamos, então, criar uma variável auxiliar em cada base, que será utilizada como chave no relacionamento das duas tabelas. Em cada base de dados, insira o cabeçalho chave na primeira linha da coluna à direita da última coluna da base de dados. Para criar os valores desta

variável auxiliar vamos utilizar a função texto CONCATENAR(), cujo resultado será a junção dos valores das variáveis selecionadas para compor a base. A inserção da fórmula na base de dados de valor da produção pode ser observada na Figura 26.

Como apresentado na Figura 26, sugere-se a inserção de um caractere que não seja utilizado em nenhum registro das variáveis utilizadas como chave para separar os valores das variáveis no momento de realizar a sua concatenação (neste caso o *underline* “_”). Este procedimento é aconselhável para permitir a visualização individualizada do valor de cada variável no resultado da chave e para impedir possíveis erros na concatenação dos valores. Por exemplo, se estivermos unindo duas variáveis cujos valores do primeiro registro sejam 00010 para a primeira e 809 para a segunda, e ocorra em algum outro registro a combinação de 0001 para a primeira e 0809 para a segunda, ambas resultariam em uma chave com valor 00010809, de maneira que este registro não seria único. A inserção de um caractere separador evita a possibilidade de ocorrência deste erro (o primeiro registro ficaria 00010_809 e o outro, 0001_0809, sendo ambos únicos e atendendo o conceito de chave primária).

The image shows a spreadsheet with the following data in row 2:

A	B	C	D	E	F	G	H
3500105	Adamantina	SP	1.2	Lenha	2001		3500105_1.2_2001

The dialog box 'Assistente de funções' is open, showing the 'Estrutura' tab for the 'CONCATENAR' function. The formula bar shows: `=CONCATENAR(A2;"_";D2;"_";F2)`. The result is `3500105_1.2_2001`. The dialog also shows a list of text fields (texto 1 to texto 4) with their respective values: 'A2', '_', 'D2', and '_'.

Figura 26. Construção da chave na base de valor da produção da silvicultura.

A concatenação dos valores também pode ser realizada por meio do caractere & entre eles. A Figura 27 ilustra esta outra fórmula de concatenação aplicada à construção da chave da base de dados de quantidade produzida (SilvProdSP.ods). Neste caso, o texto explicativo da fórmula não faz sentido, uma vez que utilizamos um caractere especial e não uma função da planilha eletrônica para realizar a concatenação.

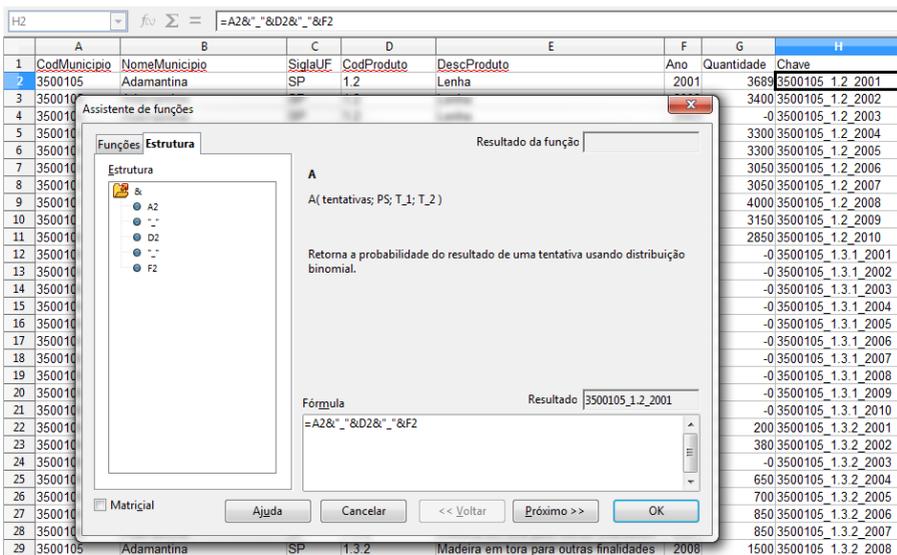


Figura 27. Construção da chave na base de quantidade para produzida da silvicultura

Agora precisamos saber como fazer para que a planilha eletrônica busque o valor da produção da base de dados SilvValorSP.ods, correspondente a cada registro de quantidade produzida da base SilvProdSP.ods. Para isto, utilizaremos duas funções em conjunto, a ÍNDICE() e a CORRESP(), ambas da categoria de funções Planilha.

A função ÍNDICE() da planilha eletrônica solicita quatro argumentos, sendo o primeiro obrigatório e os outros três opcionais, e pelo menos o segundo ou terceiro argumento

devem ser passados para a função. Ela funciona como um jogo de batalha naval, onde definimos um campo de coordenadas discretas (podendo ser um vetor de uma dimensão ou uma matriz de duas dimensões) e informamos à função qual a posição da linha e da coluna que está o valor a ser extraído da matriz. O último argumento é necessário apenas se utilizarmos um intervalo de células múltiplo, que não é o caso aqui. Por exemplo, se quisermos extrair a informação de qual vendedor realizou a terceira venda da Tabela 2, utilizaremos a função ÍNDICE() da seguinte forma: ÍNDICE(A2:D5;3;2), obtendo-se o resultado Marcos. Caso desejarmos o valor da primeira venda faremos ÍNDICE(A2:D5;1;4), e o resultado será 550,00. A função também pode ser utilizada para extrair o valor de um vetor. Para extrair o nome do produto carro, faríamos: ÍNDICE(C2:C5;4). Vale ressaltar que o intervalo de células utilizado para 'Tabela' compreende apenas as linhas com observações, não sendo inclusa a linha com os cabeçalhos das colunas.

Tabela 2. Exemplo de base de dados para uso das funções de Planilha.

	A	B	C	D
1	Mês	Vendedor	Produto	Valor (R\$)
2	Janeiro	João	Bicicleta	550,00
3	Janeiro	Aline	Barco	1.500,00
4	Fevereiro	Marcos	Televisão	4.000,00
5	Fevereiro	João	Carro	35.000,00

No entanto, se precisarmos saber de antemão qual a posição que o registro que queremos buscar ocupa em um vetor ou matriz para então inseri-lo na função ÍNDICE(), esta função não nos ajuda muito. Torna-se necessário o uso de uma função "espiã" que recebe de nós qual a informação que queremos e então se encarrega de descobrir qual a posição que esta informação ocupa na nossa matriz de dados. A nossa função "espiã" é a CORRESP().

A função CORRESP() solicita três argumentos: o valor que se deseja buscar; o vetor onde será realizada a busca; e um parâmetro que indica qual tipo de correspondência deve ser utilizada na busca (-1,0,1). É importante ressaltar que a função retorna o primeiro valor que atender às condições da busca. Se quisermos saber a posição da primeira venda ocorrida no mês de fevereiro na Tabela 2, utilizaríamos a função da seguinte forma: CORRESP("Fevereiro";A2:A5;0), o resultado seria 3. Agora, caso o nosso desejo fosse obter a posição do valor das vendas de João no mês de fevereiro, teríamos um problema, pois se procurássemos por João na coluna de vendedor com a função CORRESP("João";B2:B5;0), o resultado seria 1, mas esta seria a venda de janeiro. Por outro lado, se buscássemos a venda de fevereiro por meio da função CORRESP("Fevereiro";A2:A5;0), o resultado seria 3, mas corresponderia à venda do vendedor Marcos. Para obtermos o registro correto, necessitamos de mais de uma variável para compor chave da nossa tabela, de maneira a tornarmos os registros únicos. Sendo assim, se criássemos uma coluna auxiliar contendo o valor concatenado das observações de mês e vendedor e a utilizássemos como vetor de busca (chave) por meio da função CORRESP("Fevereiro"&"Marcos";'Coluna Auxiliar';0), o valor obtido seria 4, que corresponderia à posição correta da venda de Marcos no mês de fevereiro.

O próximo passo é trazer a coluna de valor da base de dados valor da produção (SilvValorSP.ods) para a base de dados de quantidade produzida da silvicultura (SilvProdSP.ods) de tal maneira que os registros de valor da produção para cada município, produto e ano sejam correspondentes aos registros da base de quantidade produzida. Para não perdermos as bases de dados originais (por medida de segurança), vamos criar uma nova planilha eletrônica intitulada BaseSilv.ods, e copiar as planilhas de produção e valor para este novo arquivo, conforme apresentado na Figura 28.

Vamos então construir uma fórmula que, para cada valor da coluna chave da planilha de produção, procure o valor correspondente na coluna chave da planilha de valor, retornando a posição desse registro na planilha de valor. Então retorne o valor observado na coluna valor que esteja nesta mesma posição na planilha de valor para a planilha de produção. A fórmula que utilizaremos será dada de maneira geral por $\text{ÍNDICE}(\text{'Coluna Valor';CORRESP}(\text{'observação da coluna Chave da planilha Producao';'Coluna Chave da planilha Valor';0}))$ e pode ser observada em detalhe na Figura 29, onde foi configurada para retornar o valor correspondente ao registro inserido na linha 2 da planilha Producao.

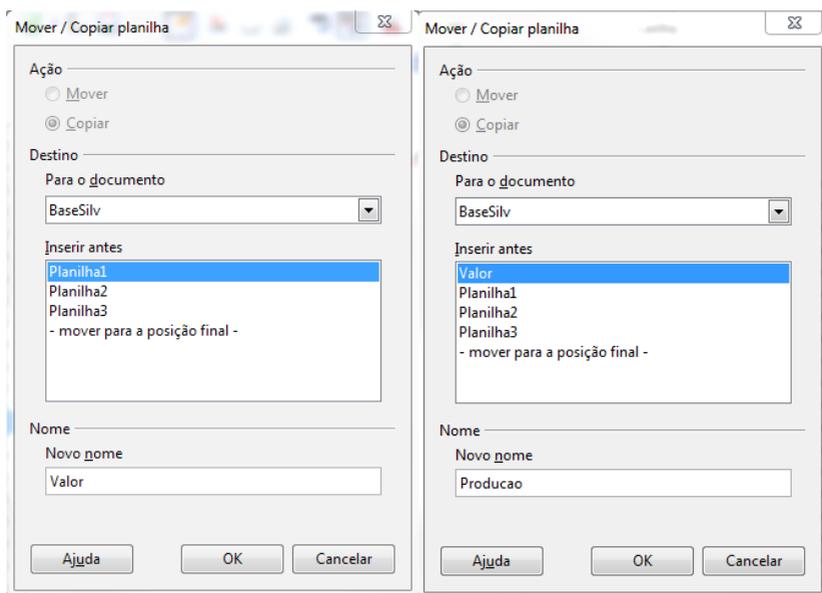


Figura 28. Copiando as bases de dados para um arquivo único BaseSilv.

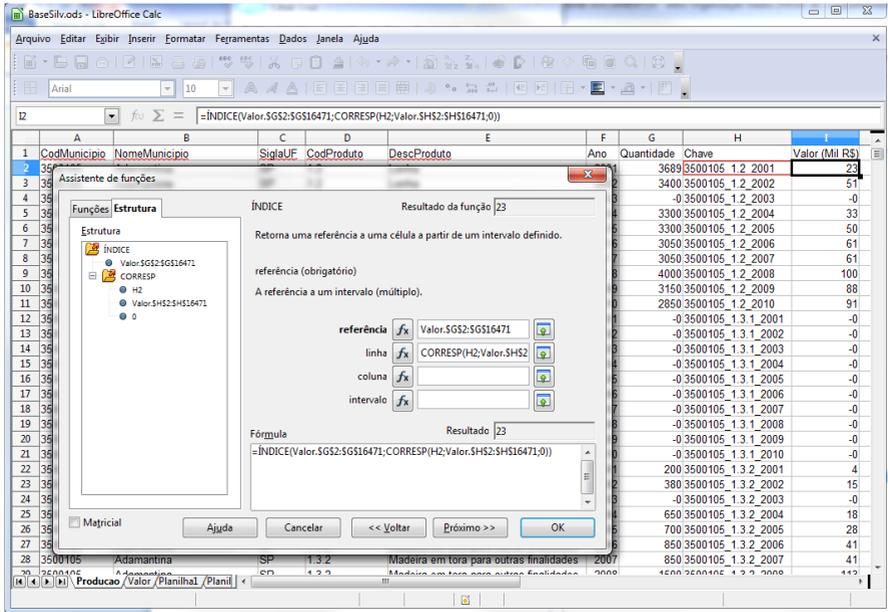


Figura 29. Copiando as bases de dados para um arquivo único BaseSilv.

Na Figura 29 pode-se observar que apenas dois argumentos são passados para a função ÍNDICE(), o vetor que ela deve retornar os resultados (no caso o valor da produção [Valor.\$G\$2:\$G\$16471]) no campo referência, e a posição da observação desejada neste vetor, no campo linha. A função CORRESP() é utilizada para encontrar a posição do registro desejado, onde solicita-se buscar, no vetor da chave da base valor (o segundo argumento da função [Valor.\$H\$2:\$H\$16471]), a posição do registro que tenha exata correspondência (valor 0 no terceiro argumento da função) ao valor 30500105_1.2_2001 (valor da célula H2 – inserida no primeiro argumento da função), que representa a chave do registro desejado na base de Producao. Outro ponto importante é que os intervalos de células correspondentes aos dois vetores devem estar fixados (uso do \$ antes da letra que representa a coluna e o número que representa a linha do endereço de cada célula), de maneira a não se alterarem após executada a cópia da fórmula para os

demais registros da base de Producao. Como desejamos que o valor da chave da planilha Producao se altere à medida que copiamos a fórmula para os demais registros da base de dados de produção, o endereço da célula no primeiro argumento da função CORRESP() deve ser deixado livre (sem \$) (Figura 29).

Após copiarmos a fórmula para os outros 16.469 registros da base de dados de produção, devemos verificar se ocorreu algum erro na busca dos valores. Caso algum valor da chave da base de Producao não seja encontrado no vetor da chave da base de valor, a fórmula retornará o valor de erro #N/DISP, que quer dizer não disponível. Para verificar de maneira rápida se temos algum registro com o valor de erro, habilite o autofiltro da base de Producao e clique na caixa de listagem da variável valor (Mil R\$). Na caixa de listagem aparecem todos os valores não repetidos para os registros de valor, caso haja algum registro com erro, o valor de erro irá aparecer ao final da caixa de listagem. No nosso caso, não apareceu nenhum valor de erro, de maneira que os registros foram importados com êxito. Um último passo é a exclusão das fórmulas da nossa base de dados. Isto é recomendável uma vez que estas fórmulas, mesmo sendo ferramentas poderosas, também exigem uma elevada capacidade de processamento da planilha, o que pode retardar muito o processo de cálculos futuros e tornar o manuseio da planilha algo operacionalmente inviável. Para excluir as fórmulas copie a coluna da variável Valor e cole os valores com a ferramenta Colar Especial não selecionando a opção Fórmulas. Após isto, exclua a coluna Chave da planilha Producao e exclua a planilha Valor, já que ela não é mais necessária uma vez que a sua informação já foi incorporada na planilha Producao. O resultado destas operações na nossa base de dados pode ser observado na Figura 30.

	A	B	C	D	E	F	G	H
1	CodMunicipio	NomeMunicipio	SiglaUF	CodProduto	DescProduto	Ano	Quantidade	Valor (Mil R\$)
2	3500105	Adamantina	SP	1.2	Lenha	2001	3689	23
3	3500105	Adamantina	SP	1.2	Lenha	2002	3400	51
4	3500105	Adamantina	SP	1.2	Lenha	2003	-0	-0
5	3500105	Adamantina	SP	1.2	Lenha	2004	3300	33
6	3500105	Adamantina	SP	1.2	Lenha	2005	3300	50
7	3500105	Adamantina	SP	1.2	Lenha	2006	3050	61
8	3500105	Adamantina	SP	1.2	Lenha	2007	3050	61
9	3500105	Adamantina	SP	1.2	Lenha	2008	4000	100
10	3500105	Adamantina	SP	1.2	Lenha	2009	3150	88
11	3500105	Adamantina	SP	1.2	Lenha	2010	2850	91
12	3500105	Adamantina	SP	1.3.1	Madeira em tora para papel e celulose	2001	-0	-0
13	3500105	Adamantina	SP	1.3.1	Madeira em tora para papel e celulose	2002	-0	-0
14	3500105	Adamantina	SP	1.3.1	Madeira em tora para papel e celulose	2003	-0	-0
15	3500105	Adamantina	SP	1.3.1	Madeira em tora para papel e celulose	2004	-0	-0
16	3500105	Adamantina	SP	1.3.1	Madeira em tora para papel e celulose	2005	-0	-0

Figura 30. Base de dados contendo informações sobre produção e valor da produção.

Unindo bases de dados com um relacionamento um para muitos

Com o objetivo de possibilitar a agregação e posterior análise dos dados por meso e microrregião, e não apenas por Unidade da Federação ou Município, vamos agora incorporar a descrição da Mesorregião e da Microrregião a qual cada município pertence à nossa base de dados. Isto será realizado buscando as informações a partir de uma tabela contendo todos os municípios brasileiros e as suas respectivas Regiões Geográficas, Unidades da Federação, Mesorregiões e Microrregiões. A tabela foi elaborada a partir de um arquivo texto disponibilizado pelo IBGE (2012), o qual contém a distribuição territorial nacional a partir do ano de 2006, e pode ser observada na Figura 31.

	A	B	C	D	E	F	G	H	I	J
1	Região_Cod	Região_Desc	UF_Cod	UF_Desc	Meso_Cod	Meso_Desc	Micro_Cod	Micro_Desc	Munic_Cod	Munic_Desc
2		1 Norte	11	Rondônia	1101	Madeira-Guaporé	11001	Porto Velho	1100205	Porto Velho
3		1 Norte	11	Rondônia	1101	Madeira-Guaporé	11001	Porto Velho	1100338	Nova Mamoré
4		1 Norte	11	Rondônia	1101	Madeira-Guaporé	11001	Porto Velho	1100452	Buritis
5		1 Norte	11	Rondônia	1101	Madeira-Guaporé	11001	Porto Velho	1100700	Campo Novo de Rondônia
6		1 Norte	11	Rondônia	1101	Madeira-Guaporé	11001	Porto Velho	1100809	Candeias do Jamani
7		1 Norte	11	Rondônia	1101	Madeira-Guaporé	11001	Porto Velho	1100940	Cujubim
8		1 Norte	11	Rondônia	1101	Madeira-Guaporé	11001	Porto Velho	1101104	Ikapá do Oeste
9		1 Norte	11	Rondônia	1101	Madeira-Guaporé	11002	Guajará-Mirim	1100080	Costa Marques
10		1 Norte	11	Rondônia	1101	Madeira-Guaporé	11002	Guajará-Mirim	1100106	Guajará-Mirim
11		1 Norte	11	Rondônia	1101	Madeira-Guaporé	11002	Guajará-Mirim	1101492	São Francisco do Guaporé
12		1 Norte	11	Rondônia	1102	Leste Rondoniense	11003	Arquemes	1100023	Arquemes
5559	5	Centro-Oeste	52	Goiás	5205	Sul Goiano	52018	Quirinópolis	5210802	Itajá
5560	5	Centro-Oeste	52	Goiás	5205	Sul Goiano	52018	Quirinópolis	5211305	Itarumã
5561	5	Centro-Oeste	52	Goiás	5205	Sul Goiano	52018	Quirinópolis	5212253	Agua Santa
5562	5	Centro-Oeste	52	Goiás	5205	Sul Goiano	52019	Quirinópolis	5216304	Paranaaguara
5563	5	Centro-Oeste	52	Goiás	5205	Sul Goiano	52018	Quirinópolis	5218508	Quirinópolis
5564	5	Centro-Oeste	52	Goiás	5205	Sul Goiano	52018	Quirinópolis	5220405	São Simão
5565	5	Centro-Oeste	53	Distrito Federal	5301	Distrito Federal	53001	Brasília	5300108	Brasília

Figura 31. Planilha eletrônica contendo as Mesorregiões e Microrregiões brasileiras

Na tabela da Figura 31 observa-se a existência de 5.564 registros, sendo cada registro correspondente a um município brasileiro, cujo código e descrição são apresentados nas colunas I e J da planilha respectivamente. As demais colunas contêm as informações de código e descrição da Região Geográfica (A e B), da Unidade da Federação (C e D), da Mesorregião (E e F) e da Microrregião (G e H). A chave que individualiza os registros desta tabela é o código do município e o seu relacionamento com a nossa base de dados da silvicultura será do tipo um para muitos, uma vez que para cada registro da tabela contendo as divisões territoriais (cada município), temos muitos registros correspondentes na tabela com a base de dados da silvicultura (vários registros com diferentes combinações de produtos e anos para cada município), mesmo que nos interesse apenas um subconjunto dos registros da tabela com as divisões territoriais (os municípios do Estado de São Paulo).

Vamos então inserir duas colunas entre as variáveis CodMunicipio (coluna A) e NomeMunicipio (coluna B) da nossa base de dados de silvicultura, inserindo os cabeçalhos Mesorregião e Microrregião nas células da linha 1 das novas colunas B e C, respectivamente. Copie a planilha contendo as divisões territoriais para o arquivo contendo a planilha da base de dados da silvicultura, para então construir as fórmulas que irão buscar os nomes das Mesorregiões e Microrregiões de cada município.

A fórmula que retornará o nome da Mesorregião deverá buscar, para cada registro da nossa base de dados de silvicultura, a posição do respectivo código de município na base de dados de divisões territoriais, e então retornar para a base de dados da silvicultura o nome da Mesorregião da base de dados de divisões territoriais que se localize na mesma posição (linha) do código de município encontrado. A fórmula que será inserida na célula B2 pode ser observada com detalhe na Figura 32.

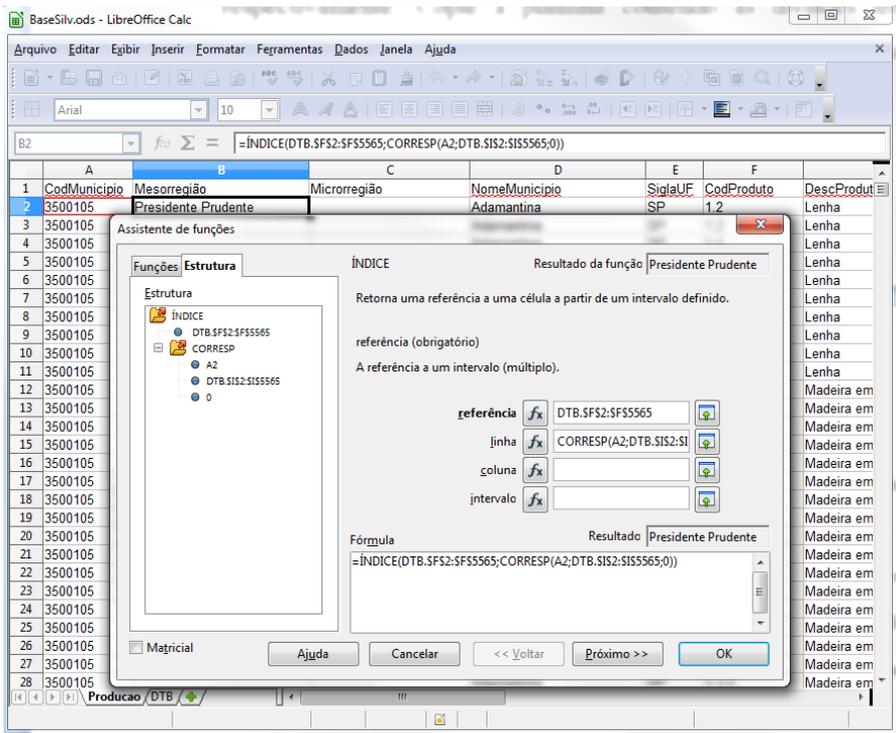


Figura 32. Inserindo a Mesorregião de cada município na planilha de silvicultura.

Agora que temos a base de dados de valor da produção e quantidade produzida por município do Estado de São Paulo, juntamente com as Mesorregiões e Microrregiões de cada município, podemos partir para a análise descritiva dos dados utilizando as ferramentas que a planilha eletrônica nos disponibiliza.

Análise descritiva de dados

Analisar a nossa base de dados utilizando apenas as funções mais comuns da planilha eletrônica, tais como a média e a contagem de registros, seria algo trabalhoso e demandaria muito tempo, principalmente se quisermos analisar a nossa base de dados de maneira mais detalhada, discriminando os resultados por município, mesorregião ou microrregião geográfica. Para aumentar a eficiência e eficácia das nossas análises, iremos lançar mão de uma poderosa ferramenta disponibilizada pela planilha eletrônica, a Tabela Dinâmica.

Segundo a ajuda do LibreOffice, a Tabela Dinâmica “permite combinar, comparar e analisar grandes quantidades de dados. É possível visualizar diferentes resumos dos dados de origem, ver detalhes de áreas de interesse, e pode criar relatórios”. Ela gera tabelas com funções de agrupamento de dados, tais como média, desvio padrão, contagem, variância, em valores absolutos ou percentuais, discriminadas por categorias definidas pelo usuário, e que podem ser alteradas rapidamente. Para criar uma tabela dinâmica no Calc do LibreOffice clique em qualquer célula da nossa base de dados, vá no menu Dados => Tabela Dinâmica => Criar, conforme pode ser observado na Figura 34.

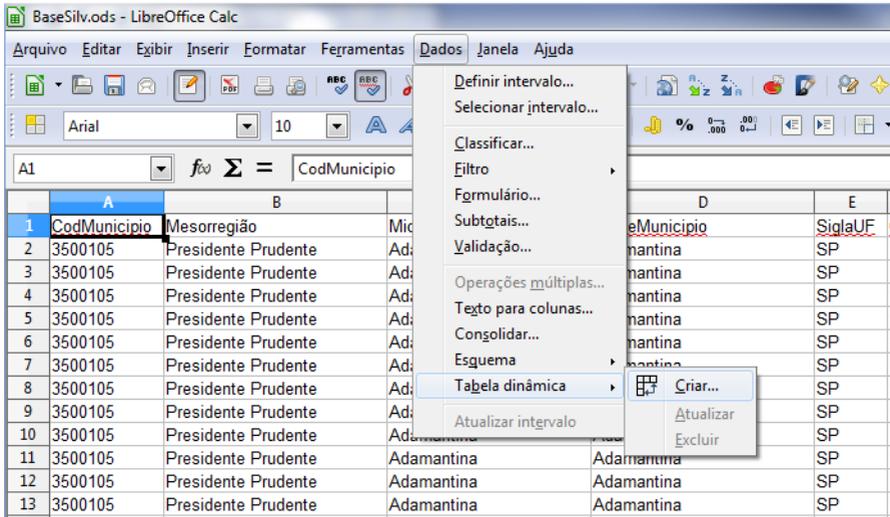


Figura 34. Criando uma tabela dinâmica

Ao aparecer a janela de Seleção de Origem dos dados, deixe marcada a opção Seleção atual e clique em OK. Em seguida irá aparecer a tela para montagem do layout da tabela dinâmica (Figura 35). Os Campos de página são utilizados para filtrar as observações da base de dados utilizadas na realização dos cálculos, os Campos de linha e de coluna são utilizados para discriminar os cálculos realizados, e os Campos de dados é o local onde devem ser inseridas as variáveis que se deseja realizar os cálculos. À direita da tabela aparecem as variáveis disponíveis na nossa base de dados.

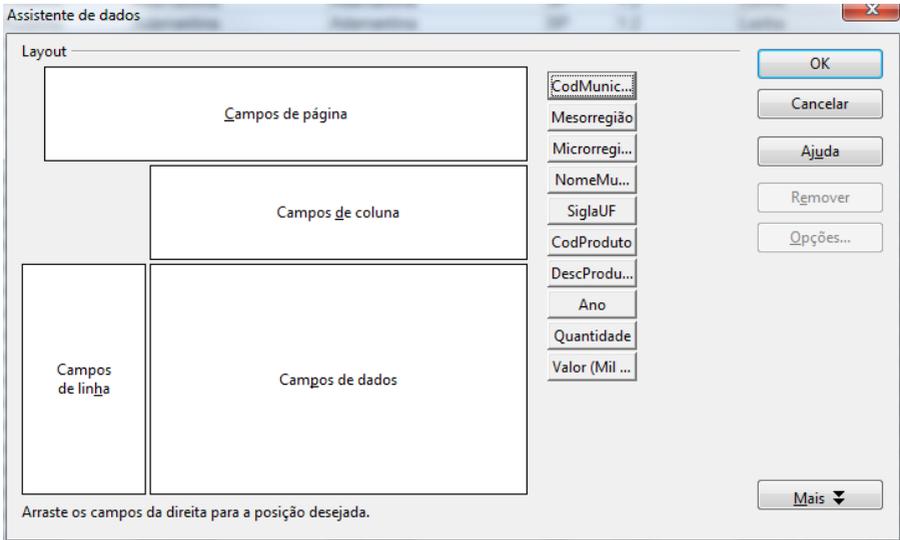


Figura 35. Tela inicial para construção do layout da tabela dinâmica.

Possivelmente, uma primeira informação que gostaríamos de saber seria a quantidade total produzida de cada produto em todo o estado de São Paulo, gerando uma tabela contendo os produtos à direita dos totais (por exemplo) e a quantidade total produzida de cada produto. Para gerar esta tabela, na tela de layout da tabela dinâmica, arraste a variável qualitativa DescProduto para os Campos de Linha e a variável quantitativa Quantidade para os Campos de dados, clique no botão OK (Figura 36).

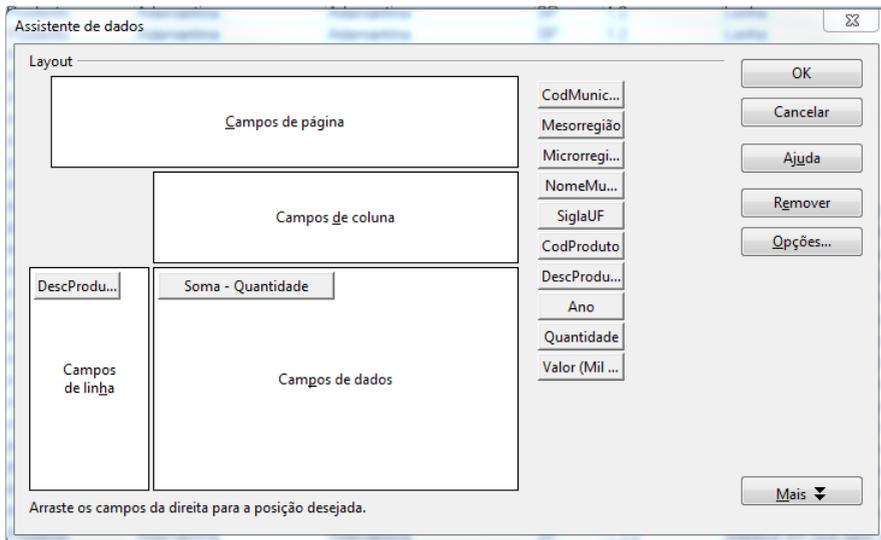


Figura 36. Construção da consulta com o total da quantidade produzida por produto.

Como as quantidades produzidas dos diferentes produtos estão na mesma unidade (m^3), o total gerado no final da tabela é um número que faz sentido, apresentando a quantidade total de madeira em tora produzida no Estado de São Paulo entre 2010 e 2001 (Figura 37). Desta maneira, a atividade de silvicultura nos municípios paulistas produziu mais de 300 milhões de m^3 de madeira bruta ao longo do período analisado.

	A	B
1	Filtro	
2		
3	DescProduto	
4	Lenha	69868670
5	Madeira em tora para outras finalidades	81132721
6	Madeira em tora para papel e celulose	150591621
7	Total Resultado	301593012

Figura 37. Resultado da consulta com o total da quantidade produzida por produto.

Apesar do total produzido no período ser uma informação interessante, também desejamos saber a evolução do total da produção de cada produto ao longo do período analisado, o que resultaria em uma tabela de dupla entrada (DescProduto e Ano). Para construir esta consulta, selecione uma célula da tabela dinâmica, vá no menu Dados => Tabela Dinâmica => Criar para abrir a janela de Layout, e arraste a variável Ano para os Campos de coluna, em seguida clique em ok. O resultado é apresentado na Figura 38.

	Ano	2001	2002	2003	2004	2005	2006	2007	2008	2009	2010	Total Resultado
lenha		7.415.039	7.003.119	7.226.914	6.864.453	6.812.087	7.180.608	7.407.385	6.891.066	6.405.078	6.662.921	69.868.670
Madeira em tora para outras finalidades		6.184.866	7.211.073	7.136.044	8.804.479	8.469.753	8.819.712	8.625.581	8.681.581	8.246.643	8.952.989	81.132.721
Madeira em tora para papel e celulose		11.256.850	13.255.892	13.317.999	14.824.430	15.592.240	19.638.922	17.340.893	14.485.708	13.665.914	17.212.783	150.591.621
Total Resultado		24.856.755	27.470.084	27.680.957	30.493.362	30.874.080	35.639.242	33.373.649	30.058.355	28.317.635	32.628.693	301.593.017

Figura 38. Total da quantidade produzida por produto ao longo do período estudado.

Agora vamos focar apenas na produção de Madeira em tora para papel e celulose, e analisar a evolução da sua produção para cada Mesorregião do estado. Para isto, clique com o botão direito do mouse em qualquer célula da tabela dinâmica e escolha a opção Editar layout (esta é uma outra forma de acessar a janela de Layout). Ao aparecer a janela de Layout da tabela dinâmica arraste a variável DescProduto para os Campos de página e a variável Mesorregião para os Campos de linha (Figura 39).

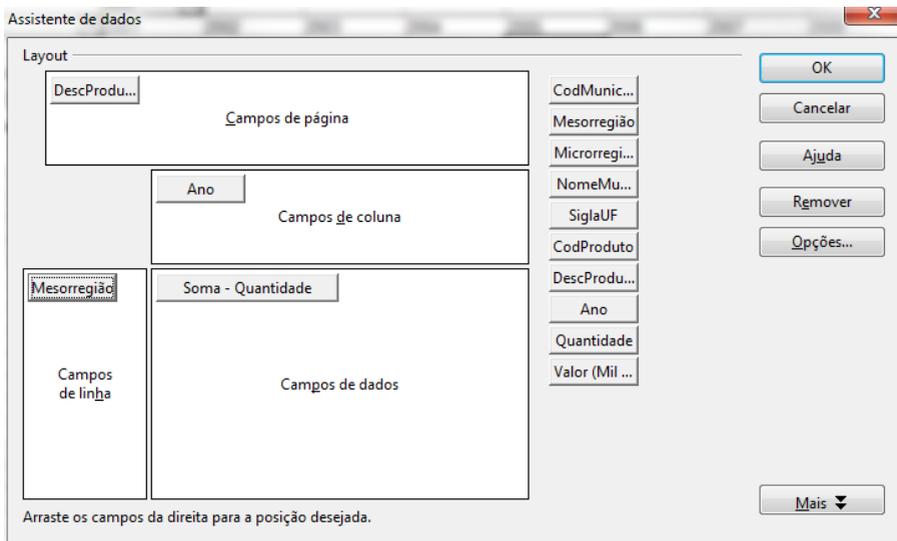


Figura 39. Consulta da evolução da quantidade produzida por mesorregião.

Após o recálculo da tabela dinâmica, a soma da quantidade produzida para cada mesorregião em cada ano poderá ser observada na tabela. Entretanto, o valor está somado para os três tipos de produtos, conforme podemos observar na célula ao lado do campo DescProduto no canto superior esquerdo da planilha. Para considerar apenas os resultados da Madeira em tora para papel e celulose, clique na seta que aparece ao lado do valor da variável DescProduto e escolha a opção Madeira em tora para papel e celulose. Para melhorar a apresentação, escolha todas as células com números e formate-as como números sem casas decimais e com separador de milhar. A tabela contendo a nossa análise está apresentada na Figura 40.

	A	B	C	D	E	F	G	H	I	J	K	L
1	Filtro											
2	DescProduto	Madeira em	tora para papel e celulose									
3												
4	Soma - Quantidade	Ano										
5	Mesorregião	2001	2002	2003	2004	2005	2006	2007	2008	2009	2010	Total Resultado
6	Araçatuba	391.323	60.869	409.875	418.043	629.179	1.375.490	612.831	809.704	118.975	528.651	5.354.940
7	Araraquara	499	677	456	496	549	454	495	517	488	506	5.137
8	Assis	2.086.782	2.154.660	2.414.044	2.223.646	2.350.966	3.354.620	2.315.107	2.419.841	2.143.716	3.461.312	24.924.694
9	Bauru	611.693	1.227.685	730.857	842.948	971.687	721.520	739.060	1.000.840	648.171	826.797	8.321.258
10	Campinas	3.774.232	4.858.836	4.958.420	6.974.636	6.808.645	7.114.039	6.678.139	6.093.770	5.933.234	6.120.202	59.314.153
11	Capetinga	0	0	0	0	0	0	0	0	0	0	0
12	Litoral Sul Paulista	1.409.582	1.508.066	1.460.330	1.481.963	1.471.457	1.549.295	1.480.867	1.458.524	1.489.248	1.606.814	14.916.146
13	Macró Metropolitana Paulista	0	0	0	0	0	0	0	0	0	0	0
14	Marília	0	0	0	0	0	0	0	0	0	0	0
15	Metropolitana de São Paulo	758.917	757.318	488.756	630.525	715.434	718.156	1.589.689	418.699	633.813	688.592	7.399.899
16	Piracicaba	519.383	546.695	502.806	353.933	624.605	1.073.044	1.208.606	249.710	371.723	752.785	6.203.290
17	Presidente Prudente	0	0	0	280	0	0	0	0	0	0	280
18	Ribeirão Preto	451.158	761.883	276.318	902.748	1.188.091	2.144.572	1.533.307	1.289.189	1.289.132	1.646.198	11.482.596
19	São José do Rio Preto	0	160	0	0	0	505	0	0	0	0	665
20	Vale do Paraíba Paulista	1.253.281	1.379.043	2.076.137	995.212	831.627	1.587.227	1.182.782	744.914	1.037.414	1.580.926	12.668.563
21	Total Resultado	11.256.850	13.255.892	13.317.999	14.824.430	15.592.240	19.638.922	17.340.883	14.485.708	13.665.914	17.212.783	150.591.621

Figura 40. Quantidade produzida de Madeira em tora para papel e celulose em cada Mesorregião ao longo do período de tempo analisado.

O resultado obtido (Figura 40) nos informa o valor absoluto da produção de cada mesorregião a cada ano, mas como poderíamos obter a participação da produção das mesorregiões em cada ano e no período total, em percentuais? Normalmente copiaríamos os valores da tabela e calcularíamos o percentual de cada célula em relação ao total da sua coluna (participação de cada mesorregião em cada ano). Entretanto, a tabela dinâmica nos disponibiliza ferramentas para a realização destes cálculos de maneira mais fácil. Clique com o botão direito do mouse em qualquer célula da tabela dinâmica e escolha a opção Editar Layout. Dê um duplo clique na variável Soma – Quantidade que está no campo de dados e a janela Campo de dados irá aparecer. Nela pode-se escolher a função de agregação dos dados que será utilizada (Soma, Contagem, Média, Máx., Min., Produto, etc...), no nosso caso a opção Soma já estará selecionada. Clique no botão Mais para abrir as opções de Valor exibido onde estará selecionada a opção Normal. Altere esta opção para % de coluna e clique em Ok para sair do Campo de dados (Figura 41), e em Ok mais uma vez para encerrar a configuração do layout.

The image shows a spreadsheet with a pivot table and several dialog boxes. The pivot table is located on the right side of the spreadsheet, with columns for years (2009, 2010) and 'Total Resultado'. The 'Assistente de dados' dialog is open, showing the 'Campos de dados' section with 'Soma' selected. The 'Valor exibido' dropdown is set to 'Normal'. The 'Campos de base' dropdown is set to 'Mesorregião'.

	2009	2010	Total Resultado
Araçatuba	0	0	0
Araraquara	118.975	528.651	5.354.940
Assis	488	506	5.137
Bauru	2.143.716	3.461.312	24.924.694
Campinas	648.171	826.797	8.321.258
Itapetininga	5.933.234	6.120.202	59.314.153
Litoral Sul Paulista	0	0	0
Macro Metropolitana Paulista	1.489.249	1.606.814	14.916.146
Manlia	0	0	0
Metropolitana de São Paulo	633.813	688.592	7.399.899
Piracicaba	371.723	752.785	6.203.290
Presidente Prudente	0	0	280
Ribeirão Preto	1.289.132	1.646.198	11.482.596
São José do Rio Preto	0	0	865
Vale do Paraíba Paulista	1.037.414	1.580.925	12.668.563
Total Resultado	13.665.914	17.212.783	150.591.621

Figura 41. Elaboração da consulta de participação anual das mesorregiões na quantidade produzida de Madeira em tora para papel e celulose.

Após o recálculo da tabela dinâmica a opção de filtro na variável DescProduto foi retirada, voltando a utilizar os registros dos três produtos para a realização dos cálculos. Para voltar a obter apenas a participação da Madeira em tora para papel e celulose, filtre a descrição do produto novamente. O resultado obtido (Figura 42) mostra que as principais mesorregiões produtoras de madeira em tora para papel e celulose no estado de São Paulo são Itapetininga, Bauru, Macro Metropolitana Paulista, Vale do Paraíba Paulista e Ribeirão Preto, sendo responsáveis por mais de 80% da produção acumulada ao longo do período de tempo analisado.

	A	B	C	D	E	F	G	H	I	J	K	L
1	Filtro											
2	DescProduto	Madeira em tora para papel e celulose										
3												
4	Soma - Quantidade	Ano										
5	Mesorregião	2001	2002	2003	2004	2005	2006	2007	2008	2009	2010	Total Resultado
6	Araçatuba	0,00%	0,00%	0,00%	0,00%	0,00%	0,00%	0,00%	0,00%	0,00%	0,00%	0,00%
7	Araraquara	3,48%	0,46%	3,08%	2,82%	4,04%	7,00%	3,53%	5,59%	0,87%	3,07%	3,56%
8	Assis	0,00%	0,01%	0,01%	0,00%	0,00%	0,00%	0,00%	0,00%	0,00%	0,00%	0,00%
9	Bauru	18,54%	16,25%	18,13%	15,00%	15,08%	17,08%	13,35%	16,71%	15,69%	20,11%	16,55%
10	Campinas	5,43%	9,26%	5,49%	5,69%	6,23%	3,67%	4,26%	6,91%	4,74%	4,80%	5,53%
11	Itapetininga	33,53%	36,65%	37,23%	47,05%	43,67%	36,22%	38,51%	42,07%	43,42%	35,56%	39,39%
12	Litoral Sul Paulista	0,00%	0,00%	0,00%	0,00%	0,00%	0,00%	0,00%	0,00%	0,00%	0,00%	0,00%
13	Macro Metropolitana Paulista	12,52%	11,38%	10,97%	10,00%	9,44%	7,89%	8,54%	10,90%	10,90%	9,34%	9,91%
14	Manlia	0,00%	0,00%	0,00%	0,00%	0,00%	0,00%	0,00%	0,00%	0,00%	0,00%	0,00%
15	Metropolitana de São Paulo	6,74%	5,71%	3,67%	4,25%	4,59%	3,66%	9,17%	2,89%	4,64%	4,00%	4,91%
16	Piracicaba	4,61%	4,12%	3,78%	2,39%	4,01%	5,46%	6,97%	1,72%	2,72%	4,37%	4,13%
17	Presidente Prudente	0,00%	0,00%	0,00%	0,00%	0,00%	0,00%	0,00%	0,00%	0,00%	0,00%	0,00%
18	Ribeirão Preto	4,01%	5,75%	2,07%	6,09%	7,62%	10,92%	8,84%	8,90%	9,43%	9,56%	7,62%
19	São José do Rio Preto	0,00%	0,00%	0,00%	0,00%	0,00%	0,00%	0,00%	0,00%	0,00%	0,00%	0,00%
20	Vale do Paraíba Paulista	11,13%	10,40%	15,59%	6,71%	5,33%	8,08%	6,82%	5,14%	7,59%	9,18%	8,41%
21	Total Resultado	100,00%	100,00%	100,00%	100,00%	100,00%	100,00%	100,00%	100,00%	100,00%	100,00%	100,00%

Figura 42. Participação das mesorregiões na quantidade produzida de Madeira em tora para papel e celulose em cada ano ao longo do período de tempo analisado.

Vamos concentrar a nossa análise apenas nestas cinco mesorregiões, filtrando as demais do nosso relatório, e observar o comportamento da produção nas microrregiões que compõem cada mesorregião selecionada. Clique na seta que aparece ao lado da célula Mesorregião, desmarque a opção todos e marque estas cinco mesorregiões (Figura 43).

Soma - Quantidade		Ano										Total Resultado
Mesorregião	2001	2002	2003	2004	2005	2006	2007	2008	2009	2010		
Classificar em ordem crescente	2086782	2154660	2414044	2223646	2350966	3354620	2315107	2419841	2143716	3461312	24524694	
Classificar em ordem decrescente	3774232	4858836	4958420	6974636	6808645	7114039	6678139	6093770	5933234	6120202	59314153	
Personalizar a classificação	1409562	1508066	1460330	1481963	1471457	1549295	1480867	1458524	1489248	1606814	14916146	
	451158	761883	276318	902748	1188091	2144572	1533307	1289189	1289132	1646198	11482596	
	1253281	1379043	2076137	995212	831627	1587227	1182782	744914	1037414	1580926	12668563	
	8975035	10662488	11185249	12578205	12650786	15749753	13190202	12006238	11892744	14415452	123306152	

Figura 43. Filtrando as mesorregiões no relatório de tabela dinâmica.

Em seguida, abra o layout da tabela dinâmica e arraste a variável Microrregião para os Campos de linha, posicionando-a logo abaixo da variável Mesorregião. Dessa forma, a tabela dinâmica discriminará primeiramente a mesorregião e, em seguida, as microrregiões que pertencem a cada mesorregião (Figura 44).

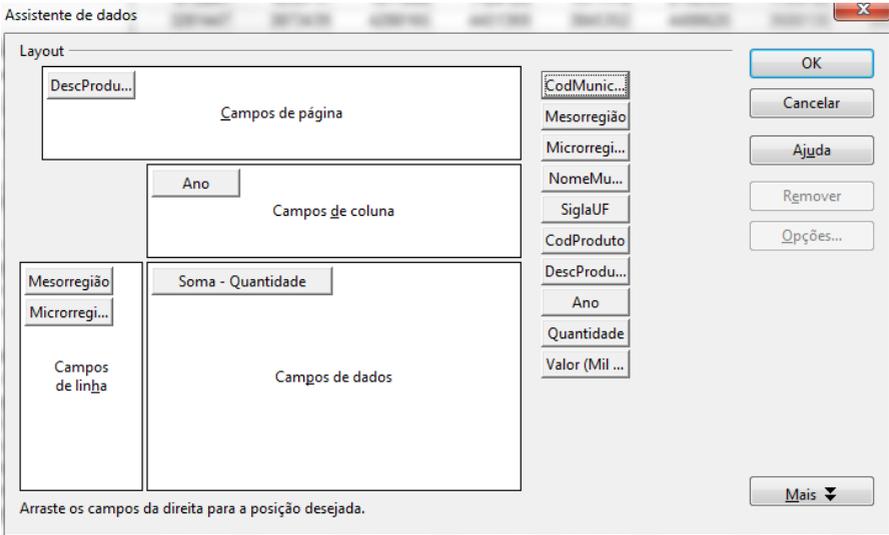


Figura 44. Layout da tabela dinâmica com duas variáveis nos Campos de linha.

Após alterar o layout da tabela dinâmica com duas variáveis nos Campos de linha, filtre novamente a variável DescProduto para Madeira em tora para f... e celulose e formate as células para números com separador de milhar e nenhuma casa decimal após a vírgula. O relatório obtido pode ser observado na Figura 45.

	A	B	C	D	E	F	G	H	I	J	K	L	M
1													
2	Filtro												
3	DescProduto	Madeira em tora para f... e celulose											
4													
5	Soma - Quantidade		Ano										
6	Mesorregião	Microrregião	2.002	2.003	2.004	2.005	2.006	2.007	2.008	2.009	2.010	Total Resultado	
7	Bauru	Awaé	889.773	830.245	672.767	438.973	432.689	893.335	803.005	906.250	642.482	1.253.409	7.753.029
8		Bauru	1.023.054	1.038.562	1.419.646	1.345.814	1.346.243	1.882.461	1.094.791	1.103.968	1.058.993	1.516.999	12.830.531
9		Botucatu	182.908	283.372	321.631	288.175	429.595	436.285	417.311	409.623	442.241	631.157	3.842.298
10		Itai	47	2.481	0	151.484	142.539	142.539	0	0	0	59.747	496.831
11		Lins	0	0	0	0	0	0	0	0	0	0	0
12	Bauru Resultado		2.086.782	2.154.660	2.414.044	2.223.646	2.350.966	3.354.620	2.315.107	2.419.841	2.143.716	3.461.312	24.924.694
13	Itapetininga	Capão Bonito	314.000	1.203.500	1.271.940	1.239.700	1.309.000	1.467.260	1.292.300	1.216.500	1.149.870	1.241.365	11.700.423
14		Itapetininga	2.264.740	2.231.920	2.134.760	4.033.010	3.668.770	3.907.415	3.660.990	3.235.480	3.207.180	3.370.760	31.745.008
15		Itapeva	1.205.452	1.423.416	1.551.730	1.701.926	1.810.875	1.739.370	1.724.849	1.641.790	1.556.184	1.508.087	15.865.719
16		Itatui	0	0	0	0	0	0	0	0	0	0	0
17	Itapetininga Resultado		3.774.232	4.858.836	4.958.430	6.974.636	6.808.545	7.114.039	6.678.139	6.093.170	5.933.234	6.120.203	59.314.153
18	Macro Metropolitana Paulista	Brança Paulista	39.400	87.700	111.800	107.894	99.100	102.080	83.650	101.660	104.409	125.960	963.653
19		Jundiai	0	46	0	0	0	0	0	0	0	0	46
20		Piedade	944.792	951.860	945.795	984.106	962.524	1.020.732	988.335	834.439	927.760	934.720	9.595.053
21		Sorocaba	425.390	468.460	407.745	389.963	409.833	426.483	408.892	422.425	457.079	546.194	4.397.394
22	Macro Metropolitana Paulista Resultado		1.409.582	1.508.066	1.460.330	1.481.963	1.471.457	1.549.295	1.480.867	1.458.524	1.489.248	1.606.814	14.916.148
23	Ribeirão Preto	Barretos	0	0	0	0	0	0	0	0	0	0	0
24		Barretos	163.165	366.542	72.276	0	472.021	867.801	562.772	429.487	26.610	622.326	3.583.000
25		Franca	0	0	0	0	0	0	0	0	0	0	0
26		Jaboticabal	1.539	0	0	10.929	57.317	56.133	0	0	0	0	182.051
27		Ribeirão Preto	286.454	395.341	204.042	891.819	658.753	1.220.638	914.402	859.702	1.262.522	1.023.872	7.717.543
28		São-João-da-Barra	0	0	0	0	0	0	0	0	0	0	0
29	Ribeirão Preto Resultado		451.158	761.883	276.318	902.748	1.188.091	2.144.572	1.533.307	1.289.189	1.289.312	1.646.198	11.482.596
30	Vale do Paraíba Paulista	Bananal	33.051	73.893	85.470	25.561	60.590	60.590	16.622	110.000	151.128	75.959	691.863
31		Campos do Jordão	114.440	22.772	22.877	48.038	18.675	16.058	914	0	0	107.444	350.486
32		Guaratiningá	37.909	42.375	49.994	41.416	60.175	62.778	14.287	12.344	32.740	437.903	782.411
33		Paraibuna/Paratinga	761.349	564.208	521.882	356.253	425.494	701.903	646.952	218.008	301.605	340.173	4.837.828
34		São José dos Campos	306.532	875.795	1.465.714	523.954	206.693	745.898	505.607	404.472	551.940	620.948	6.006.053
35	Vale do Paraíba Paulista Resultado		1.253.281	1,379.043	2,076,337	995,212	831,627	1,981,221	1,182,782	748,914	1,031,414	1,589,926	12,668,563
36	Total Resultado		8,975,035	10,662,488	11,185,249	12,578,205	12,650,786	15,749,153	13,190,202	12,006,238	11,892,744	14,415,452	123,306,152

Figura 45. Resultado da consulta discriminando mesorregião e microrregião.

O último relatório que faremos retornará a evolução da quantidade produzida de Madeira em tora para outras finalidades dos seis municípios paulistas que apresentaram a maior produção considerando apenas os três últimos anos da nossa série de dados (2008 a 2010).

O primeiro passo é encontrar os seis principais municípios produtores de Madeira em tora para outras finalidades considerando apenas os três últimos anos. Abra o editor de layout da tabela dinâmica, arraste as variáveis Mesorregião e Microrregião para fora dos Campos de Linha e a variável NomeMunicípio para os Campos de linha. Arraste também a variável Ano para os Campos de página. Dê um duplo clique na variável Ano após ela estar inserida nos Campos de página, clique no botão Opções e na área intitulada Ocultar itens, marque os anos que não desejamos computar nos nossos cálculos (2001 a 2007). Clique em OK para todas as janelas até sair do editor de layout. Em seguida filtre o valor da variável DescProduto para Madeira em tora para outras finalidades. O resultado pode ser observado na Figura 46.

B11		f ₀	Σ =	28035
	A	B	C	
1	Filtro			
2	DescProduto	Madeira em tora para outras finalidades		
3	Ano	- todas -		
4		- todas -		
5	NomeMunicípio	2008		
6	Adamantina	2009		
7	Adolfo	2010		
8	Aguai			
9	Águas da Prata			
10	Águas de Lindóia			
11	Águas de Santa Bárbara			
12	Agudos			
13	Alambari			
14	Altair			2
15	Altinópolis			0
16	Alto Alegre			1350
17	Alumínio			0
18	Álvares Florence			625
19	Álvares Machado			0

Figura 46. Resultado da consulta discriminando os municípios e os três últimos anos.

O detalhe do filtro da variável Ano foi deixado em destaque na Figura 46 para evidenciar o resultado do ocultamento dos anos entre 2001 e 2007 da variável Ano. Para facilitar encontrarmos os seis maiores produtores dentre os mais de 500 municípios, vamos classificar os municípios por ordem decrescente de quantidade produzida. Selecione as células contendo os nomes dos municípios (deixe de fora o cabeçalho – linha 5 – e o total geral na última linha), acesse o menu Dados => Classificar, selecione a Coluna B e marque a opção Decrescente para a primeira opção de filtro utilizada. Após este procedimento, podemos observar que os seis principais municípios produtores de Madeira em tora para outras finalidades considerando a produção entre 2008 e 2010, os quais são: Itapetininga, Lençóis Paulistas, Cabrália Paulista, Botucatu e São Miguel Arcanjo.

Para filtrar apenas para os municípios que nos interessam poderíamos seguir o mesmo procedimento utilizado para a filtragem dos valores de Ano. Entretanto, fazer isto para mais de quinhentas observações seria algo exaustivo e pouco eficiente. O procedimento sugerido a seguir utiliza uma fórmula para classificar, em uma nova variável diretamente na base de dados, os registros que interessam a nossa análise, e depois utilizar esta variável nos Campos de página para filtrar os registros que desejamos inserir no nosso relatório de tabela dinâmica.

A fórmula que utilizaremos será composta pelas funções SE(), É.NÃO.DISP() e CORRESP(). A lógica da fórmula se baseia no erro que a função CORRESP() retorna ao não encontrar o valor procurado dentro do vetor indicado para busca. Quando isto ocorre, a função CORRESP() retorna o valor #N/DISP, o qual indica que o vetor indicado não contém o valor procurado. A função É.NÃO.DISP() verifica se o argumento inserido é um erro do tipo #N/DISP e retorna o valor VERDADEIRO caso seja, e FALSO caso contrário. A função SE() verifica se o primeiro argumento é VERDADEIRO ou FALSO, caso seja VERDADEIRO

ela retorna o segundo argumento inserido na função, caso seja falso ela retorna o terceiro argumento.

Primeiramente insira uma nova planilha no arquivo da base de dados e copie os seis nomes dos municípios desejados ao longo da primeira coluna, de maneira que os nomes irão ocupar o intervalo de células com endereço A1:A6. Em seguida, insira o cabeçalho **MaiorMunicOF** na primeira linha da coluna a direita da última coluna da nossa base de dados. O terceiro passo é inserir a fórmula a partir da segunda linha, conforme ilustrado na Figura 47.

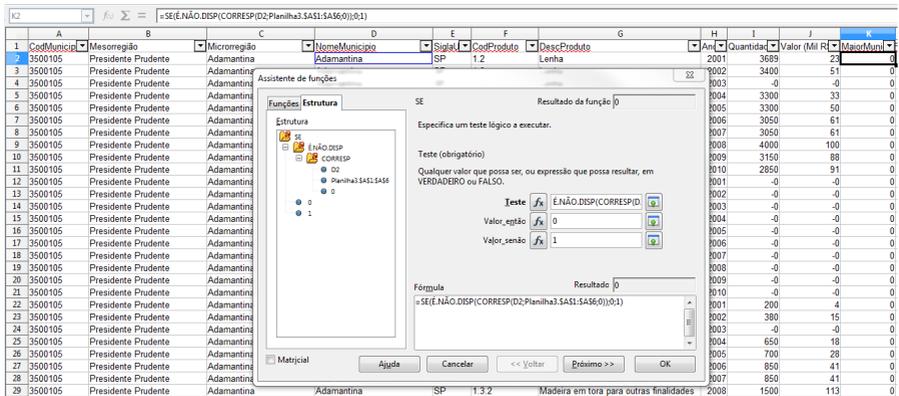


Figura 47. Fórmula para filtragem dos municípios selecionados na base de dados.

Como aumentamos o tamanho da nossa base de dados, a tabela dinâmica não incorpora automaticamente novas linhas ou colunas que venhamos a inserir na nossa base, de maneira que se torna necessário modificar o intervalo de células que alimenta a tabela dinâmica. Vá na tabela dinâmica e abra o layout da tabela, clique no botão **Mais** no canto inferior direito e modifique o intervalo de dados inserido na caixa de seleção "Seleção a partir de", sendo o novo intervalo de dados dado por $\$Producao.\$A\$1:\$K\$16471$. Clique em **Ok**, e abra a janela de layout da tabela dinâmica novamente. A primeira modificação que percebemos é a inserção da variável **MaiorMunicOF** no conjunto das variáveis disponíveis

na nossa base. Dê um duplo clique na variável Ano, clique em Opções e desmarque os anos que foram ocultados para que a tabela dinâmica possa voltar a exibir os valores para todos os anos. Arraste a variável Ano dos Campos de página para os Campos de coluna e insira a variável MaiorMunicOF nos Campos de página. O layout da tabela dinâmica deverá estar conforme apresentado na Figura 48.

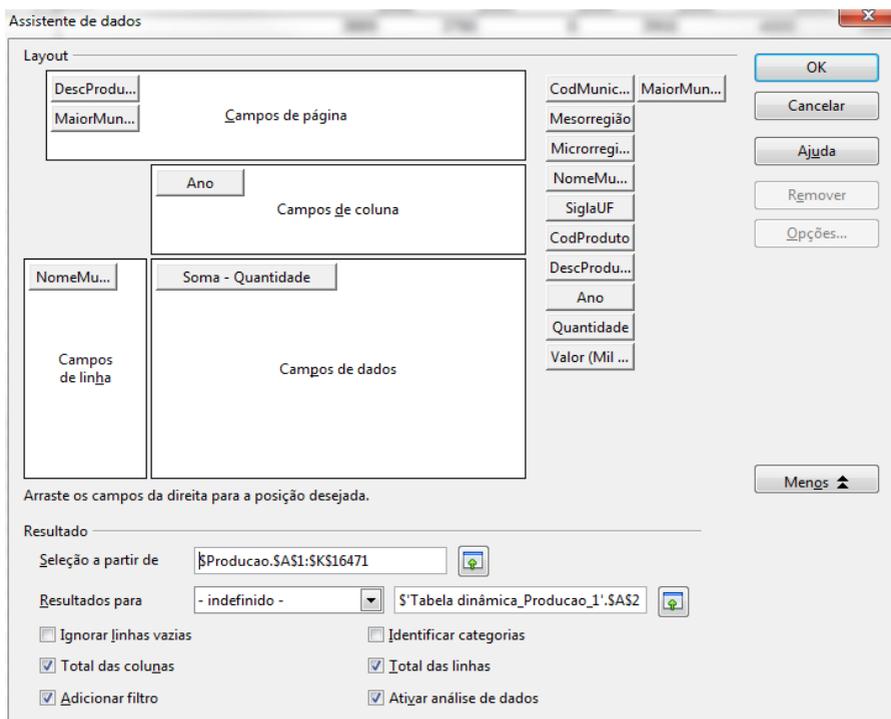


Figura 48. Layout da tabela dinâmica com a variável filtro MaiorMunicOF.

A tabela dinâmica voltará a apresentar a evolução da soma da quantidade produzida dos três produtos da silvicultura para cada um dos 549 municípios em cada ano. Para completar o relatório desejado escolha a opção Madeira em tora para outras finalidades na caixa de listagem da variável DescProduto e o número 1 na caixa de listagem da variável MaioMunicOF.

Formate a exibição dos dados numéricos com separador de milhar e nenhuma casa decimal. O relatório final pode ser observado na Figura 49.

Soma - Quantidade	Ano	2002	2003	2004	2005	2006	2007	2008	2009	2010	Total	Resultado
Boticatu		469.875	623.712	658.665	493.857	497.231	483.619	462.652	448.462	469.203	496.373	5.103.648
Cabrália Paulista		741.301	756.127	778.811	802.175	811.801	840.214	781.399	793.120	815.327	823.480	7.943.755
Itapetininga		815.060	908.610	960.310	1.480.650	1.530.840	1.585.164	1.470.650	1.580.950	1.612.560	1.740.500	13.965.294
Itatinga		165.883	179.516	188.293	429.731	481.502	768.667	754.612	702.574	696.553	698.520	5.091.860
Luís de Paula		360.000	314.974	282.000	539.192	356.366	362.184	900.685	901.320	932.380	934.614	5.883.715
São Miguel Arcanjo		241.950	384.060	333.280	305.840	388.640	330.915	340.300	330.800	334.100	385.600	3.215.485
Total Resultado		2.824.079	2.986.999	3.181.359	4.648.449	3.966.380	4.370.763	4.710.298	4.757.226	4.859.123	5.079.087	40.783.158

Figura 49. Evolução dos seis maiores municípios produtores de Madeira para outros fins entre os anos de 2008 a 2010.

A principal vantagem da fórmula utilizada para filtrar os valores de municípios desejados é a sua flexibilidade, pois se precisarmos aumentar o número de municípios filtrados na nossa análise para cinquenta, precisaríamos apenas modificar o vetor de referência da função CORRESP(), e a fórmula se ajustaria a nova realidade.

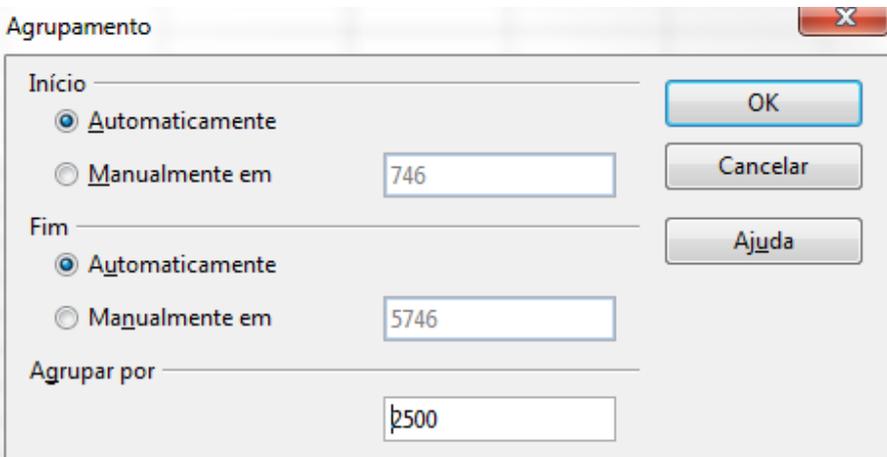
Finalizando a nossa análise dos dados, vamos construir uma tabela com a contagem do número de municípios de acordo com classes de valor da produção de lenha no ano de 2010 com o auxílio da tabela dinâmica. Abra o editor de layout da tabela dinâmica e faça as seguintes modificações:

- 1) Arraste a variável MaiorMunicOF para fora da janela de layout, excluindo-a do relatório a ser calculado;
- 2) Arraste a variável Ano dos Campos de coluna para os Campos de página;
- 3) Arraste a variável Valor (Mil R\$) para os Campos de linha;
- 4) Arraste a variável NomeMunicípio para os Campos de dados e altere a sua função na janela Campo de dados de Soma para Contagem;

- 5) Clique em OK para gerar o relatório;
- 6) Na variável DescProduto selecione a opção Lenha;
- 7) Na variável Ano selecione a opção 2010.

O relatório de tabela dinâmica irá discriminar na primeira coluna todos os registros não repetidos de valor da produção de Lenha no Ano de 2010, variando de zero a 36.856, com a respectiva ocorrência de registros da nossa base em cada valor da produção não repetido.

Clique em qualquer célula de Valor (Mil R\$), vá no menu Dados => Esquema => Agrupar. Para gerar um número razoável de classes, vamos agrupar os nossos dados em classes constantes de 2.500, inserindo este valor no campo Agrupar por (Figura 50).



The image shows a dialog box titled "Agrupamento" (Grouping) with a close button (X) in the top right corner. The dialog is divided into three sections:

- Início** (Start):
 - Automaticamente
 - Manualmente em: 746
- Fim** (End):
 - Automaticamente
 - Manualmente em: 5746
- Agrupar por** (Group by): 2500

On the right side of the dialog, there are three buttons: "OK", "Cancelar" (Cancel), and "Ajuda" (Help).

Figura 50. Agrupando os valores de Quantidade nos Campos de linha.

A Figura 51 apresenta o resultado do nosso relatório de tabela dinâmica, onde podemos perceber que 26 municípios paulistas obtiveram valor da produção igual ou superior a R\$ 2,5 milhões com a venda de lenha de silvicultura no ano de 2010.

Filtro	
DescProduto	Lenha
Ano	2010
Valor (Mil R\$)	
-0-2499	523
2500-4999	13
5000-7499	6
7500-9999	3
10000-12499	2
12500-14999	1
35000-37499	1
Total Resultado	549

Figura 51. Número de município por classes de valor da produção de Lenha em 2010.

Para retirar os municípios que não geraram renda a partir da produção de lenha de silvicultura em 2010 da nossa análise, bem como calcular a frequência relativa de municípios em cada categoria, abra novamente o layout da tabela dinâmica e, na janela de opções da variável NomeMunicípio, clique no botão Mais e selecione a opção “% de coluna”. Clique em OK até executar o relatório, então clique em uma célula de classes de quantidade, vá no menu Dados => Esquema => Agrupar, marque o início como Manualmente em e insira o valor 1, clique em OK. Em seguida filtre novamente a variável DescProduto para Lenha e o Ano para 2010. Na Figura 52 se pode observar que cerca de 30% dos municípios paulista obtiveram renda menor que mil reais com a produção de lenha de silvicultura em 2010, 65% obtiveram entre mil e dois milhões e quinhentos mil reais, e os 5% restantes obtiveram renda bruta acima de dois milhões e meio de reais

Filtro	
DescProduto	Lenha
Ano	2010
Quantidade	
<1	30,24%
1-50000	63,21%
50001-100000	3,46%
100001-150000	0,91%
150001-200000	0,91%
200001-250000	0,73%
250001-300000	0,18%
300001-350000	0,18%
650001-700000	0,18%
Total Resultado	100,00%

Figura 52. Calculando a frequência relativa das classes de Quantidade.

Com estas duas tabelas, teríamos os instrumentos necessários para a construção de um histograma de frequências absolutas e relativas para estudar a distribuição da ocorrência da variável Valor da produção de Lenha dos municípios paulistas em 2010. Para remover a configuração de categorias e voltar a apresentação da tabela dinâmica ao seu formato original, clique em qualquer classe de quantidade, vá no menu Dados => Esquema => Desagrupar.

Considerações finais

A análise de bases de dados maiores, bem como a obtenção de informações conjuntas de bases de dados que não estavam unidas, se torna acessível a uma gama maior de usuários por meio do uso das técnicas e ferramentas de planilhas eletrônicas apresentadas neste documento.

As planilhas eletrônicas disponibilizam outras ferramentas além das que foram apresentadas, que permitem a junção de bases de dados de maneira rápida e eficiente. A tabela dinâmica também possui funcionalidades que podem ser muito úteis dependendo da análise desejada.

O intuito ao se escrever este documento foi iniciar uma exposição introdutória ao tema. O usuário que tenha interesse em se aprofundar nestes conhecimentos pode fazê-lo acessando os vários livros, textos ou sites na internet elaborados sobre o assunto, além dos manuais disponibilizados no LibreOffice.

Referências

DUARTE, M. Conceito de variável. **InfoEscola**: navegando e aprendendo, 15 ago. 2008. Disponível em: <<http://www.infoescola.com/matematica/conceito-de-variavel/>>. Acesso em: 29 mar. 2012.

GUJARATI, D. M. **Econometria básica**. Rio de Janeiro: Elsevier, 2006. 812 p.

IEMMA, A. F. **Estatística descritiva**. Piracicaba, SP: QOP Publicações, 1992. 182 p.

IBGE. Divisão territorial brasileira em 2006. In: _____. **Banco de dados agregados**: sistema IBGE de recuperação automática: Sidra. Disponível em: <<http://www.sidra.ibge.gov.br/bda/territorio/download/default.asp?z=t&o=4>>. Acesso em: 10 abr. 2012.

LEVINE, M. D.; BERENSON, M. L.; STEPHAN, D. **Estatística: teoria e aplicações usando o Microsoft Excel em português**. Rio de Janeiro: LTC, 2000. 811 p.

NEUFELD, J. L. **Estatística aplicada à administração usando o Excel**. São Paulo: Prentice Hall, 2003. 434 p.

UNIVERSIDADE DO ESTADO DE SÃO PAULO. **Treinamento em BrOffice.org base**. Bauru: Faculdade de Ciências, [s.d.]. Disponível em: <www.fc.unesp.br/upload/sti/broffice/BrOffice.org_Base.pdf>. Acesso em: 05 abr. 2012.

Embrapa

Florestas

Ministério da
**Agricultura, Pecuária
e Abastecimento**

GOVERNO FEDERAL
BRASIL
PAÍS RICO É PAÍS SEM POBREZA

CGPE 10050