



Análise de Dados Médicos

Ferramentas Essenciais no Excel

04.Nov.2024

João Alves
FMUP | Faculdade de Medicina da Universidade do Porto
MEDCIDS | Departamento de Medicina da Comunidade, Informação e Decisão em Saúde

EXCEL: o que é e para que serve?

“O Excel é uma ferramenta incrivelmente poderosa (...) que permite adicionar dados, classificá-los e filtrá-los, inseri-los em tabelas e criar gráficos incríveis.” [Microsoft]

- Organizar dados
- Estruturar tabelas
- Fazer cálculos matemáticos
- Construir gráficos
- Produzir relatórios
- ...



A estrutura funcional da folha de cálculos é a célula

	A	B
1	3,14	
2	pi	
3	=PI()	
4		

Células podem conter:

- números
- texto
- fórmulas

No formato tabular, os dados são dispostos em colunas e linhas

ID	Age	Gender	Weight	Height	BMI
001	34	M	58	169	20,3
002	58	F	82	165	30,1
003	29	M	76	162	29,0
004	42	M	63	182	19,0
005	25	F	78	173	26,1
006	26	M	58	161	22,4
007	50	M	77	184	22,7
008	25	M	66	174	21,8
009	50	F	73	172	24,7
010	63	M	59	164	21,9

Nota: esta tabela contém exclusivamente dados sintéticos gerados aleatoriamente.
Estes dados não representam valores reais de quaisquer pacientes reais.

Os dados são organizados em formato de tabela
⇒ **dados tabulares**

Tipicamente, apresentam-se:

- Variáveis em colunas
- Observações em linhas

Dados tabulares são geralmente gravados em ficheiros XLS ou CSV

	A	B	C	D	E	F
1	ID	Age	Gender	Weight	Height	BMI
2	001	34	M	58	169	20,3
3	002	58	F	82	165	30,1
4	003	29	M	76	162	29,0
5	004	42	M	63	182	19,0
6	005	25	F	78	173	26,1
7	006	26	M	58	161	22,4
8	007	50	M	77	184	22,7
9	008	25	M	66	174	21,8
10	009	50	F	73	172	24,7
11	010	63	M	59	164	21,9

Ficheiros Excel

- Extensões .XLSX ou .XLS (anterior a 2007)
- Outras extensões: .XLSB, .XLSM, ...
- Aceitam valores ou fórmulas
- Aceitam formatação

0	1	2	3	4	5	6	7	8	9	10
ID ,Age ,Gender ,Weight ,Height ,BMI										
001 ,34 ,M ,58 ,169 ,20.3										
002 ,58 ,F ,82 ,165 ,30.1										
003 ,29 ,M ,76 ,162 ,29.0										
004 ,42 ,M ,63 ,182 ,19.0										
005 ,25 ,F ,78 ,173 ,26.1										
006 ,26 ,M ,58 ,161 ,22.4										
007 ,50 ,M ,77 ,184 ,22.7										
008 ,25 ,M ,66 ,174 ,21.8										
009 ,50 ,F ,73 ,172 ,24.7										
010 ,63 ,M ,59 ,164 ,21.9										

Ficheiros CSV (dados separados por vírgula)

- Extensão .CSV (“Comma Separated Values”)
- Ficheiros de texto com valores separados por vírgulas (,)
- Podem usar outros caracteres de separação, como ponto e vírgula (;) ou tabulação (\t)
- Não aceitam fórmulas nem formatação

Ao abrir ficheiros CSV em Excel pode ser necessário converter os dados para células

1. Selecionar coluna A
2. Selecionar no menu Data > Text to Columns
3. Escolher “Delimited” para indicar que os dados são separados por um elemento de delimitação
4. Identificar o elemento de delimitação:
 - a. Semicolon (;
 - b. Comma (,)
 - c. ...
5. Terminar

	A	B	C
1	ID,Age,Gender,Weight,Height,BMI		
2	001,34,M,58,169,20.3		
3	002,58,F,82,165,30.1		
4	003,29,M,76,162,29.0		
5	004,42,M,63,182,19.0		
6	005,25,F,78,173,26.1		
7	006,26,M,58,161,22.4		
8	007,50,M,77,184,22.7		
9	008,25,M,66,174,21.8		
10	009,50,F,73,172,24.7		
11	010,63,M,59,164,21.9		

Ao abrir ficheiros CSV em Excel pode ser necessário converter os dados para células

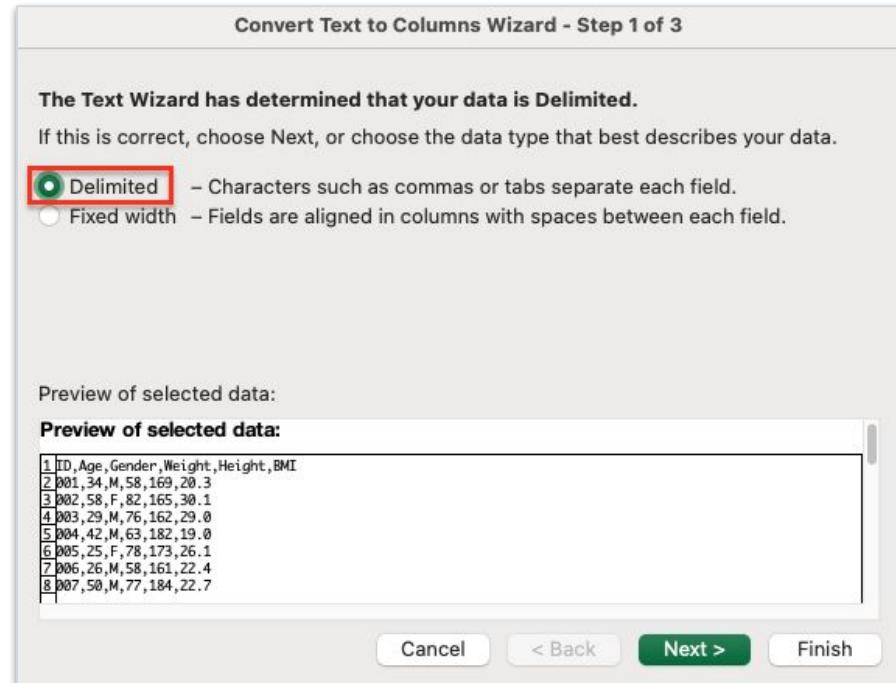
1. Selecionar coluna A
2. Selecionar no menu Data > Text to Columns
3. Escolher “Delimited” para indicar que os dados são separados por um elemento de delimitação
4. Identificar o elemento de delimitação:
 - a. Semicolon (;
 - b. Comma (,)
 - c. ...
5. Terminar



	A	B	C	D	E	F
1	ID,Age,Gender,Weight,Height,BMI					
2	001,34,M,58,169,20.3					
3	002,58,F,82,165,30.1					
4	003,29,M,76,162,29.0					
5	004,42,M,63,182,19.0					
6	005,25,F,78,173,26.1					
7	006,26,M,58,161,22.4					
8	007,50,M,77,184,22.7					
9	008,25,M,66,174,21.8					
10	009,50,F,73,172,24.7					
11	010,63,M,59,164,21.9					

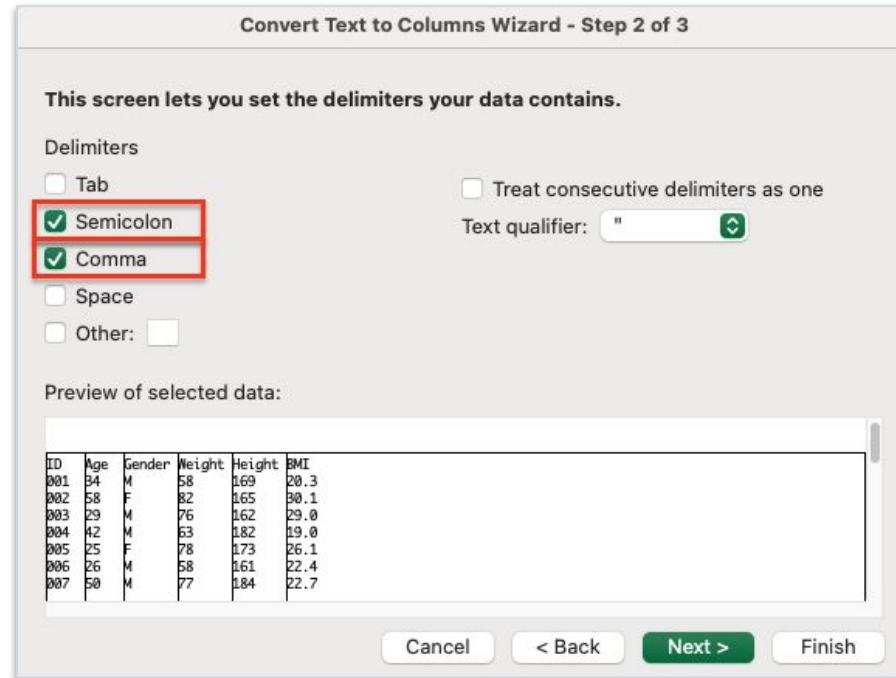
Ao abrir ficheiros CSV em Excel pode ser necessário converter os dados para células

1. Selecionar coluna A
2. Selecionar no menu Data > Text to Columns
3. Escolher “Delimited” para indicar que os dados são separados por um elemento de delimitação
4. Identificar o elemento de delimitação:
 - a. Semicolon (;)
 - b. Comma (,)
 - c. ...
5. Terminar



Ao abrir ficheiros CSV em Excel pode ser necessário converter os dados para células

1. Selecionar coluna A
2. Selecionar no menu Data > Text to Columns
3. Escolher “Delimited” para indicar que os dados são separados por um elemento de delimitação
4. Identificar o elemento de delimitação:
 - a. Semicolon (;
 - b. Comma (,)
 - c. ...
5. Terminar



Ao abrir ficheiros CSV em Excel pode ser necessário converter os dados para células

1. Selecionar coluna A
2. Selecionar no menu Data > Text to Columns
3. Escolher “Delimited” para indicar que os dados são separados por um elemento de delimitação
4. Identificar o elemento de delimitação:
 - a. Semicolon (;
 - b. Comma (,)
 - c. ...
5. Terminar

	A	B	C	D	E	F
1	ID	Age	Gender	Weight	Height	BMI
2	1	34	M	58	169	20.3
3	2	58	F	82	165	30.1
4	3	29	M	76	162	29.0
5	4	42	M	63	182	19.0
6	5	25	F	78	173	26.1
7	6	26	M	58	161	22.4
8	7	50	M	77	184	22.7
9	8	25	M	66	174	21.8
10	9	50	F	73	172	24.7
11	10	63	M	59	164	21.9
12						

Tabelas

Para além de dispor os dados em tabelas simples, é útil organizar dados em **intervalos nomeados**¹ ou em **tabelas dinâmicas**²

1. named ranges
 2. pivot tables

É útil registrar tabelas como intervalos nomeados (named range)

The screenshot shows the Microsoft Excel ribbon with the 'Insert' tab selected. The 'Tables' group in the ribbon has the 'Table' icon highlighted with a red box. Below the ribbon, a data table is displayed with columns labeled ID, Age, Gender, Weight, Height, and BMI. The data rows from 2 to 11 are visible.

	A	B	C	D	E	F
1	ID	Age	Gender	Weight	Height	BMI
2	1	34	M	58	169	20,3
3	2	58	F	82	165	30,1
4	3	29	M	76	162	29,0
5	4	42	M	63	182	19,0
6	5	25	F	78	173	26,1
7	6	26	M	58	161	22,4
8	7	50	M	77	184	22,7
9	8	25	M	66	174	21,8
10	9	50	F	73	172	24,7
11	10	63	M	59	164	21,9
12						

Intervalos nomeados (named ranges) permitem:

- Identificar o conjunto de dados por um nome, em vez de um intervalo de células
eg. **table1** em vez de **A1:F11**
- Identificar a coluna pelo nome da variável
eg. **[@Age]** em vez de **B2:B11**
- Preencher colunas de fórmulas automaticamente
- Criar filtros automaticamente
- Associar “slicers”
- ...

A ordenação e filtragem pode ser regulada nos botões à direita das células de cabeçalho

- Números de linha a azul indicam que estão aplicados filtros à tabela
- Símbolo de漏 no botão indica que um filtro está aplicado nessa coluna
- Símbolo de seta no botão indica que uma ordenação está aplicada à coluna (↑Ascendente / ↓Descendente)
- Pressionando no botão de filtro (▼) permite aceder a várias opções de ordenação e filtragem

The screenshot shows a Microsoft Excel spreadsheet with a data table in rows 1 through 20. The columns are labeled A through I. Row 1 contains the column headers: ID, Age, Gender, Weight, Height, and BMI. The Height column has a blue header bar, indicating a filter is applied. The BMI column has a grey header bar, indicating a sort is applied. Arrows point from the first three bullet points in the list to the blue header bar of the Height column and the grey header bar of the BMI column. The BMI header bar features a downward-pointing arrow and a funnel icon. The Height header bar features a double-headed vertical arrow icon. Row 20 is the last row of data, with row numbers 12 through 19 visible on the left.

E1	A	B	C	D	E	F	G	H	I
1	ID	Age	Gender	Weight	Height	BMI			
2	6	26	M	58	161				
3	3	29	M	76	162				
4	10	63	M	59	164				
6	1	34	M	58	169				
9	8	25	M	66	174				
10	4	42	M	63	182				
11	7	50	M	77	184				
12									
13									
14									
15									
16									
17									
18									
19									
20									

Sort

By colour: None

Filter

By colour: None

Choose One

Search

(Select All)
161
162
164
169
174
182

Auto Apply

Apply Filter Clear Filter

Os comandos de opções de ordenação e filtragem de dados encontram-se no separador “Data”



Ordenação Ascendente/Descendente
para uma única coluna

Ordenação múltipla, para várias
colunas e vários critérios

Criação/remoção de filtros
Limpeza de filtros
Opções avançadas

A formatação condicional permite alterar o aspetto das células em função da informação que contêm

	A	B	C	D	E	F
1	ID	Age	Gender	Weight	Height	BMI
2	1	34	M	58	169	20,3
3	2	58	F	82	165	30,1
4	3	29	M	76	162	29,0
5	4	42	M	63	182	19,0
6	5	25	F	78	173	26,1
7	6	26	M	58	161	22,4
8	7	50	M	77	184	22,7
9	8	25	M	66	174	21,8
10	9	50	F	73	172	24,7
11	10	63	M	59	164	21,9

Opções de formatação condicional:

- Destacar células com valores maiores ou menores que X, ou valores repetidos
- Destacar os Top-N valores, ou valores acima/abaixo da média
- Acrescentar barras com dimensão proporcional ao valor da célula
- Destacar as células em escalas de gradiente de cores
- ...

O mesmo menu permite apagar ou editar as regras de formatação condicional já aplicadas.

Slicers (Segmentação de Dados) fornecem botões que permitem filtrar dados em tabelas nomeadas ou em tabelas dinâmicas

Para inserir slicer:

- Selecionar uma qualquer célula da tabela
- Selecionar menu **Table**
- Escolher **Insert Slicer**

ID	Age	Gender	Weight	Height	BMI
2	58	F	82	165	30,1
5	25	F	78	173	26,1
9	50	F	73	172	24,7

ID	Age	Gender	Weight	Height	BMI
1	34	M	58	169	20,3
3	29	M	76	162	29,0
4	42	M	63	182	19,0
6	26	M	58	161	22,4
7	50	M	77	184	22,7
8	25	M	66		
10	63	M	59		

Home Insert Draw Page Layout Formulas Data Review View Automate Developer Table

Table Name:

Summarise with Pivot Table
 Remove Duplicates
 Convert to Range
 Insert Slicer
 Refresh

Header Row
 First Column
 Total Row
 Last Column
 Banded Rows
 Banded Columns

Uma tabela dinâmica é uma ferramenta avançada para calcular, resumir e analisar dados

Average of Weight Average of Height

	Average of Weight	Average of Height
F	77,7	170,0
M	65,3	170,9
Grand Total	69	170,6

Uma tabela dinâmica permite:

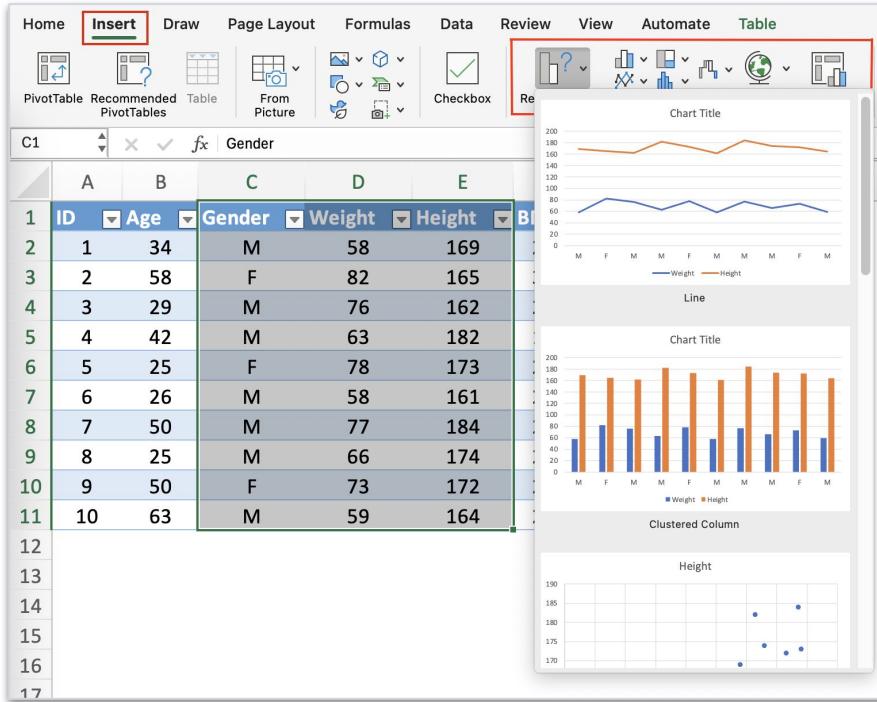
- Selecionar e dispor as variáveis em linha ou em coluna para criar diferentes resumos
- Resumir os dados em diferentes estatísticas sumárias: soma, média, máximo, mínimo, ...
- Aplicar filtros
- Criar gráficos dinâmicos em função dos valores apresentados na tabela

Gráficos

Em Excel, as fórmulas permitem fazer cálculos, manipular conteúdos ou testar condições



A visualização através de gráficos permite um entendimento de dados complexos de forma mais simples e mais rápida

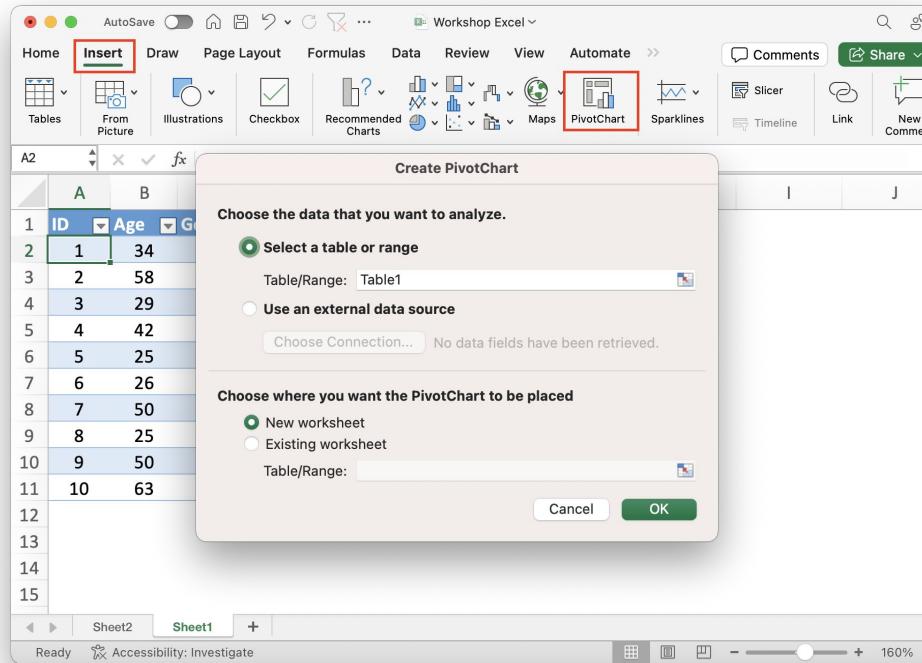


Insert > Recommended Charts

ou, em alternativa, selecionar um tipo de gráfico específico

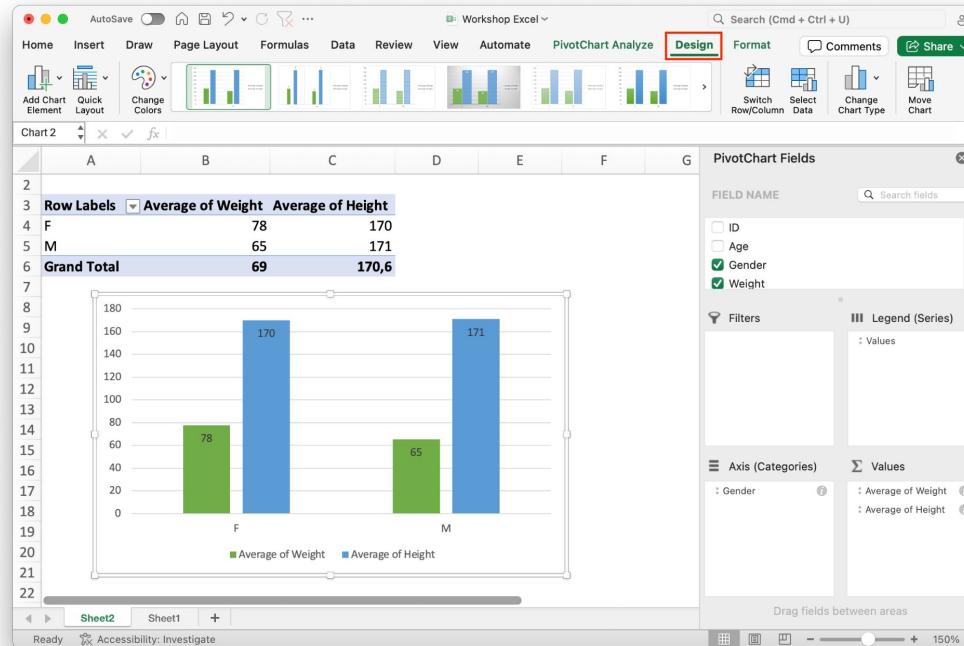
Sugestão: começar por selecionar as colunas com os dados que se pretende visualizar, e só depois inserir um gráfico para essas variáveis.

Gráficos dinâmicos, baseados em tabelas dinâmicas, permitem maior flexibilidade no controlo das variáveis a apresentar

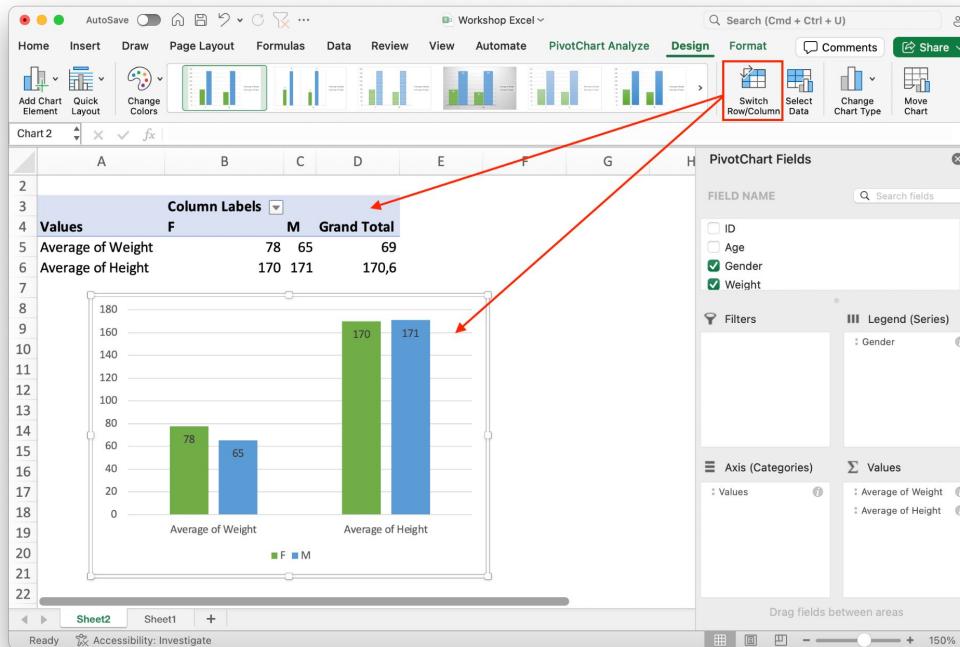


Insert > PivotChart

O menu Design oferece várias opções de configuração de gráficos



O menu Design oferece várias opções de configuração de gráficos

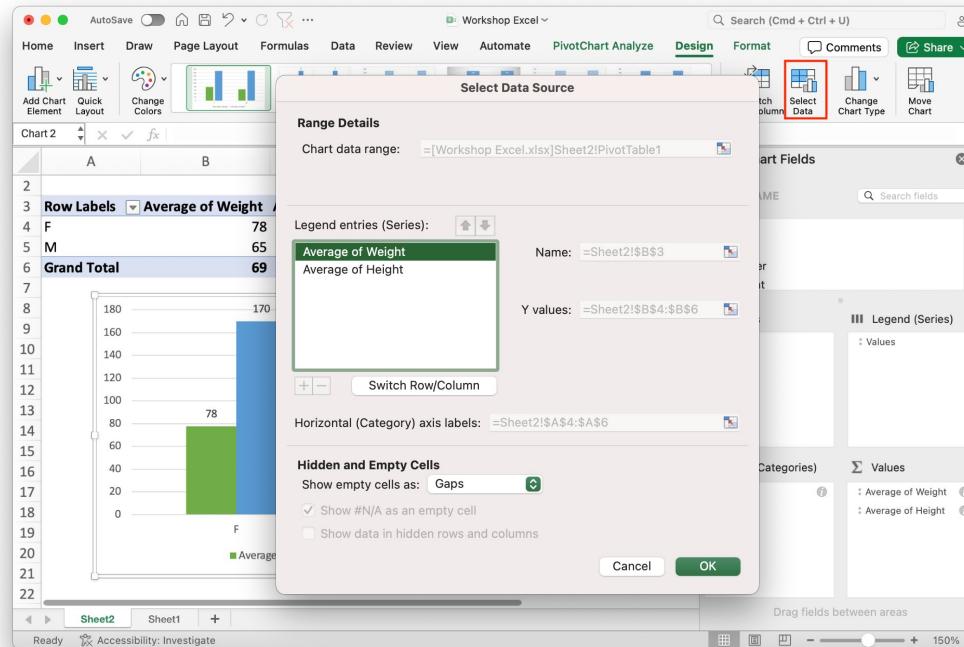


Design > Switch Row/Column

Troca as variáveis das colunas com as das linhas

Nota: Faz uma inversão da posição das variáveis, não faz uma rotação de 90° no gráfico!

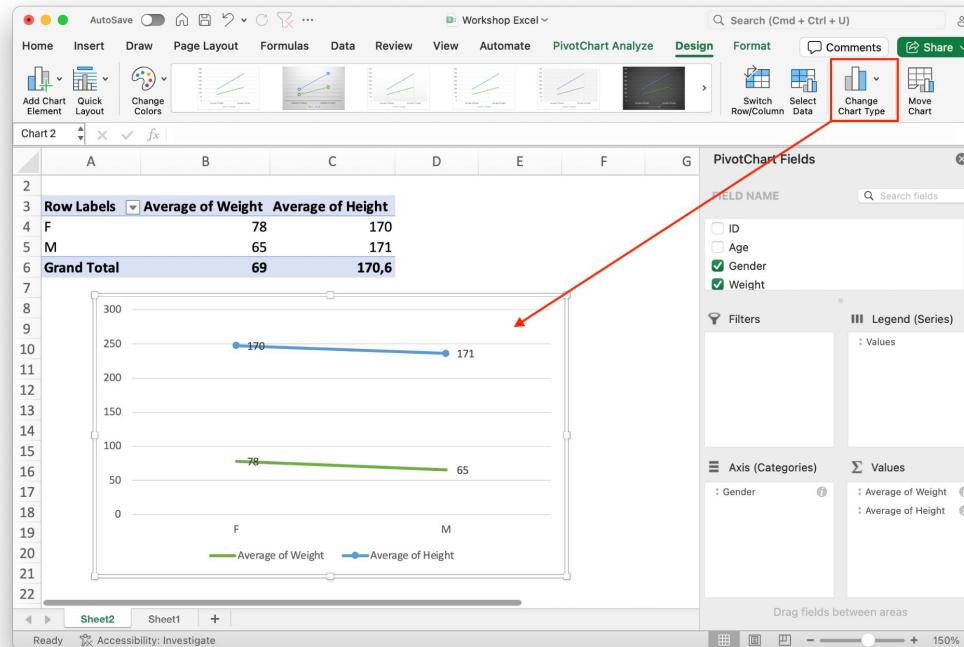
O menu Design oferece várias opções de configuração de gráficos



Design > Select Data

Permite selecionar as séries de dados a incluir no gráfico

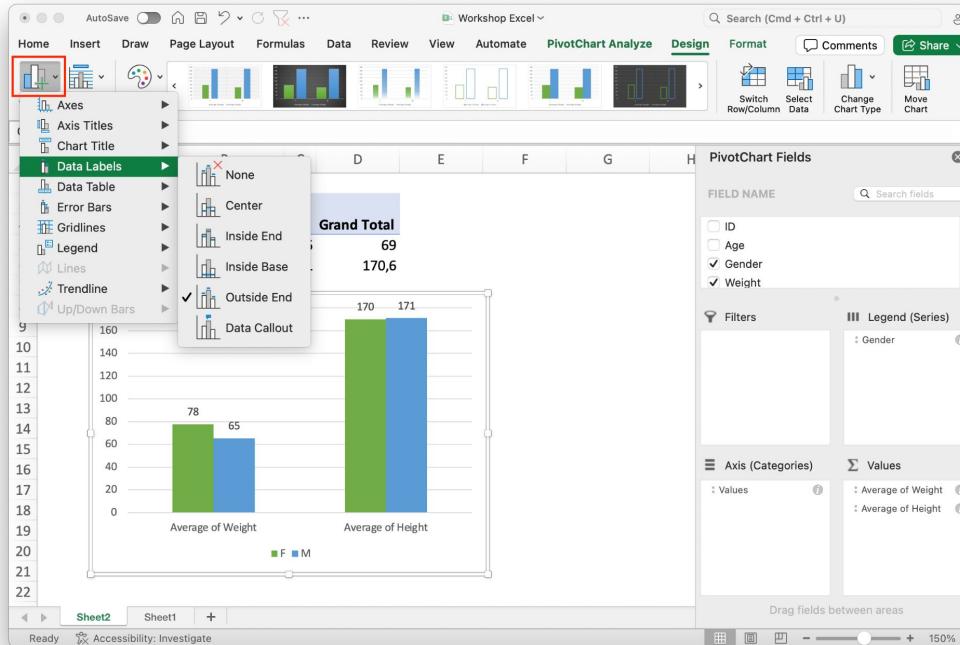
O menu Design oferece várias opções de configuração de gráficos



Design > Change Chart Type

Permite alterar o tipo de gráfico

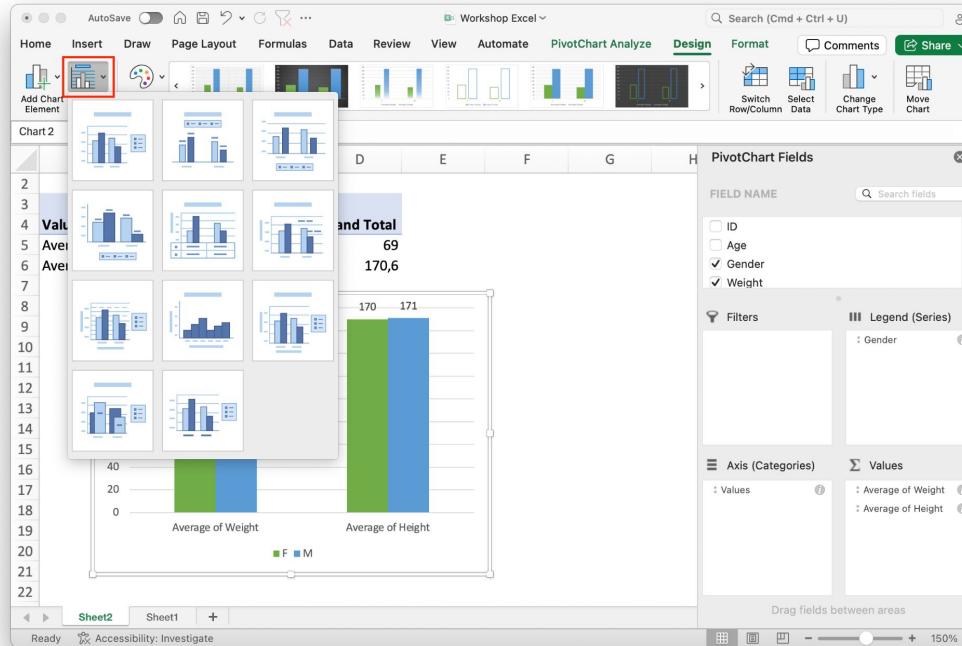
O menu Design oferece várias opções de configuração de gráficos



Design > Add Chart Element

Permite acrescentar, retirar ou mudar a posição de diversos elementos do gráfico

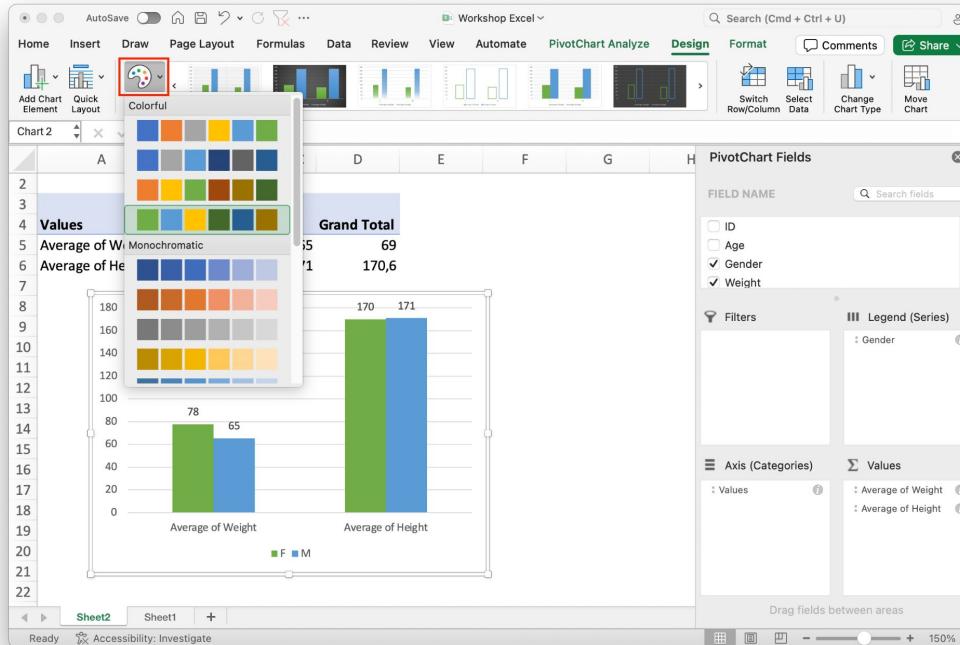
O menu Design oferece várias opções de configuração de gráficos



Design > Quick Layout

Oferece diversas configurações pré-definidas para a disposição dos diferentes elementos do gráfico

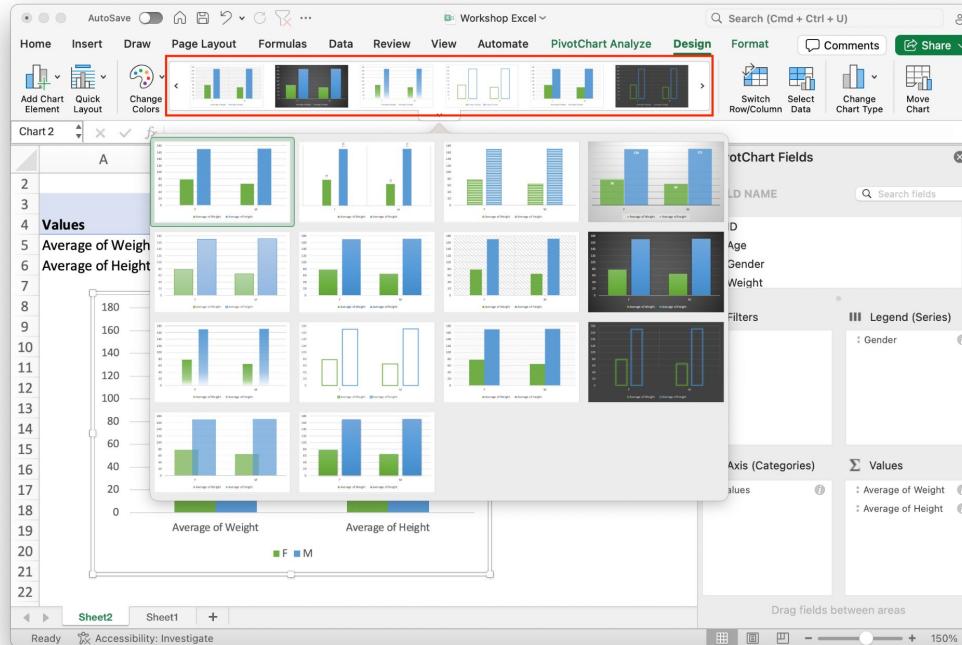
O menu Design oferece várias opções de configuração de gráficos



Design > Change Colors

Oferece diversos esquemas de cores pré-definidos a aplicar a cada série de dados

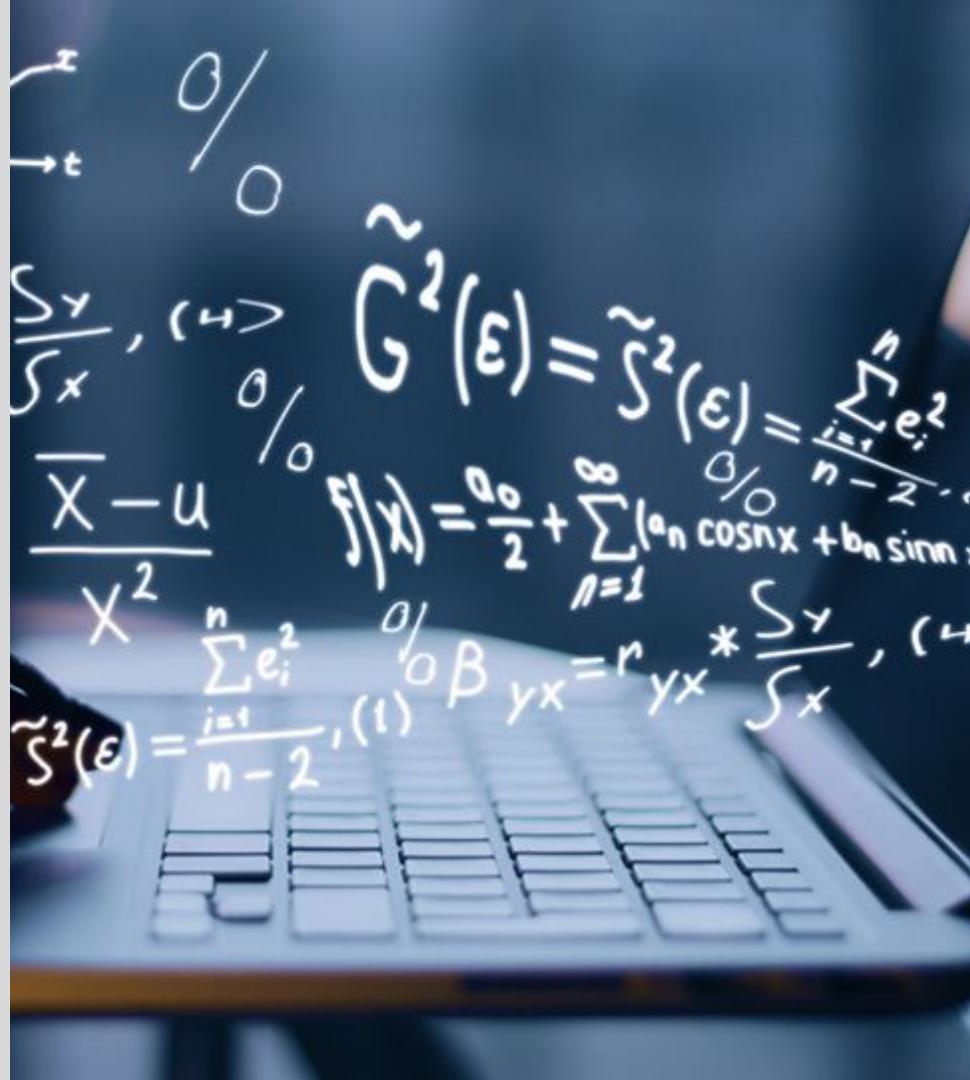
O menu Design oferece várias opções de configuração de gráficos



Permite alterar para diferentes configurações visuais pré-definidas

Fórmulas

Em Excel, as fórmulas permitem fazer cálculos, manipular conteúdos ou testar condições



O Excel usa o sistema de referência por **Coluna-Linha (A1)**, com as colunas indicadas por letras e as linhas indicadas por números

	A	B	C
1			
2			
3			
4			
5			
6			

Diagrama ilustrativo do sistema de referência A1:

- A1: Célula individual na interseção da Coluna A e Linha 1.
- A3: Célula individual na interseção da Coluna A e Linha 3.
- B5: Célula individual na interseção da Coluna B e Linha 5.
- C2: Célula individual na interseção da Coluna C e Linha 2.
- 1: Linha 1, que contém celulas A1, B1, C1.
- A1:A6: Intervalo que inclui todas as células da Coluna A, desde a Linha 1 até a Linha 6.

- A1 refere-se a uma célula
- A1:A6 refere-se a um intervalo de células
- A:A refere-se a toda a coluna A
- 1:1 refere-se a toda a linha 1

exemplo:

- =SUM(B:C)
soma todos os valores contidos nas colunas B a C
- =COUNT(2:2)
conta todos os valores numéricos contidos na linha 2

No sistema de referência por Coluna-Linha (A1) usa-se o símbolo de dólar (\$) para bloquear a coluna (\$A1), a linha (A\$1) ou ambas (\$A\$1)

	A	B	C	D	E	F
1	1	4		1	4	
2	2	5		2	5	
3						
4				1	1	
5				2	2	
6						
7				1	4	
8				1	4	
9						
10				1	1	
11				1	1	

=A1 =B1

=A2 =B2

=\$A1 =\$A1

=\$A2 =\$A2

=A\$1 =B\$1

=A\$1 =B\$1

=\$A\$1 =\$A\$1

=\$A\$1 =\$A\$1

A1 representa uma referência relativa

- Se arrastada ou copiada para baixo, passa a A2
- Se arrastada ou copiada para a direita passa a B1

\$A\$1 representa uma referência absoluta

- Para onde quer que seja arrastada ou copiada, mantém-se como \$A\$1

Anatomia e sintaxe de funções

- A expressão da fórmula começa com sinal de igual
- Um função é seguida de parênteses curvos
- Cada função pode requerer nenhum, um, ou vários argumentos
- Pode conter valores numéricos, texto, referências a outras células ou mesmo outras funções

=TODAY()

=ABS(-3)

=ROUND(3,45; 1)

=UPPER("fmup")

=SUM(A1:A5)

=POWER(PI(); 2)

A screenshot of the Microsoft Excel interface. The ribbon at the top includes Home, Insert, Draw, Page Layout, Formulas, Data, Review, Comments, and Share. The 'Formulas' tab is selected. Below the ribbon is a toolbar with icons for Clipboard, Font, Alignment, Number, Cells, Editing, Sensitivity, and Add-ins. A formula bar shows '=SUM'. The main area shows a worksheet with columns A, C, D, E and rows 1 through 25. Cell A1 contains the formula '='. A callout bubble points to the 'fx' button in the formula bar, with the text 'Pressionar botão **fx**' below it. To the right, the 'Formula Builder' dialog box is open, showing search results for 'round' and detailed information about the 'ROUND' function.

Pressionar botão **fx**

Formula Builder

Q round

All

ROUND

ROUNDDOWN

ROUNDUP

Maths Trig

ROUND

ROUNDDOWN

ROUNDUP

Insert Function

fx ROUND

Rounds a number to a specified number of digits.

Syntax

ROUND(number;num_digits)

- **Number:** is the number you want to round.
- **Num_digits:** is the number of digits to which you want to round. Negative rounds to the left of the decimal point; zero to the nearest integer.

More help on this function

O construtor de fórmulas dá várias opções de ajuda para a busca de funções e composição de fórmulas

Campo de busca de funções

Descrição da função

Sintaxe da função

Descrição dos argumentos usados pela função

Os argumentos de cada função podem ser vistos no construtor de fórmulas ou na caixa de ajuda por baixo da barra de fórmulas

The screenshot shows the Microsoft Excel interface. The ribbon is visible at the top with tabs like Home, Insert, Draw, Page Layout, Formulas, Data, Review, etc. Below the ribbon, the formula bar displays 'SUM' and the beginning of an IF function: '=IF('. The main area shows a 1x9 grid of cells. Cell A1 contains '=IF(' and is currently selected. To the right of the grid, a 'Formula Builder' dialog box is open, showing the structure of the IF function: 'IF Logical_test = logical'. Below this, there are three input fields: 'Value_if_true = any' and 'Value_if_false = any', each with a small icon to its right. Two red arrows point from the text in the first two bullet points to the formula bar and the 'Logical_test' field in the Formula Builder dialog respectively.

Fórmulas matemáticas básicas

	A	B	C	D	E	F	G
1	2	5	7		2	=MIN(A1:C1)	
2							
3	2	5	7		7	=MAX(A3:C3)	
4							
5	2	5	7		3	=COUNT(A5:C5)	
6							
7	2	5	7		14	=SUM(A7:C7)	
8							
9	2	5	7		70	=PRODUCT(A9:C9)	

Fórmulas matemáticas básicas

	A	B	C	D	E	F	G
1	2	5	7		4,6667	=AVERAGE(A1:C1)	
2							
3	2	5	7		2,5166	=STDEV(A3:C3)	
4							
5	3	4	4		4	=MODE(A5:C5)	
6							
7	3	4			81	=POWER(A7; B7)	
8							
9	3	4			81	=A9^B9	

Fórmulas matemáticas para arredondamento

	A	B	C	D	E	F	G
1	9,347				9,3	=ROUND(A1; 1)	
2							
3	9,347				10	=CEILING.MATH(A3)	
4							
5	9,347				9	=FLOOR.MATH(A5)	
6							
7	1 300				1 500	=CEILING.MATH(A7; 500)	
8							
9	1 300				1 000	=FLOOR.MATH(A9; 500)	

Fórmulas com texto

	A	B	C	D	E	F	G
1	FMUP				FM	=LEFT(A1; 2)	
2							
3	FMUP				UP	=RIGHT(A3; 2)	
4							
5	FMUP				MU	=MID(A5; 2; 2)	
6							
7	wOrKsHoP				WORKSHOP	=UPPER(A7)	
8							
9	wOrKsHoP				workshop	=LOWER(A9)	
10							
11	wOrKsHoP				Workshop	=PROPER(A11)	

Fórmulas com texto

	A	B	C	D	E	F	G
1	F	M	U	P		FMUP	=TRIM(A1)
2							
3	TRUP				FMUP		=SUBSTITUTE(A3; "TR"; "FM")
4							
5	X	Y	Z		XYZ		=CONCAT(A5:C5)
6							
7	X	Y	Z		X, Y, Z		=TEXTJOIN(", "; TRUE; A7:C7)

Fórmulas com datas

	A	B	C	D	E	F	G
1					04/11/2024	=TODAY()	
2							
3					04/11/2024 10:15	=NOW()	
4							
5	4	11	2024		04/11/2024	=DATE(C5; B5; A5)	
6							
7	04/11/2024				4	=DAY(A7)	
8							
9	04/11/2024				11	=MONTH(A9)	
10							
11	04/11/2024				2024	=YEAR(A11)	
12							
13	04/11/2024				2	=WEEKDAY(A13; 1)	

Uso de expressões condicionais

	A	B	C	D	E	F	G
1	5	10		FALSE		=A1 = B1	
2							
3	5	10		TRUE		=A3 < B3	
4							
5	5	5		FALSE		=A5 > B5	
6							
7	5	5		TRUE		=A7 >= B7	
8							
9	5	10		menor		=IF(A9 < B9; "menor"; "maior")	
10							
11	5	2		maior		=IF(A11 < B11; "menor"; "maior")	

Contagem e Soma condicional

	A	B	C	D	E	F	G
1	Ano	Relatorio	Valor				
2	2022	Sim	100				
3	2022	Não	200				
4	2023	Sim	300				
5	2023	Não	400				
6	2024	Sim	500				
7	2024	Não	600				

300 =SUMIF(A2:A7; 2022; C2:C7)
(range; criteria; sum_range)

100 =SUMIFS(C2:C7; A2:A7; 2022; B2:B7; "Sim")
(sum_range; criteria_range_1; criteria_1; criteria_range_2;
criteria_2; ...)

COUNT / SUM ⇒ conta / soma todos os valores

COUNTIF / SUMIF ⇒ conta / soma valores que cumpram 1 condição

COUNTIFS / SUMIFS ⇒ conta / soma valores que cumpram múltiplas condições

Procura de valores em tabelas

	A	B	C	D	E	F	G	H	I
1	IMC_li	IMC_ls	Descrição						
2	0	18,49	Baixo Peso	18,5	Peso Normal		=VLOOKUP(E2; Tabela2; 3)		
3	18,5	24,99	Peso Normal	27	Excesso de Peso		=VLOOKUP(E4; Tabela2; 3; TRUE)		
4	25	29,99	Excesso de Peso	27	#N/A		=VLOOKUP(E5; Tabela2; 3; FALSE)		
5	30	34,99	Obesidade Moderada	31	Obesidade Moderada		=XLOOKUP(E7; Tabela2[IMC_li]; Tabela2[Descrição];;-1)		
6	35	39,99	Obesidade Severa						
7	40	100	Obesidade Mórbida						

Nota: privilegiar o uso de **XLOOKUP** em vez de VLOOKUP

Exercícios



Abrir página **data.gov** e, procurar por “**Monthly Counts of Deaths**” de 2014-2019 e descarregar ficheiro de dados CSV

The screenshot shows two browser windows. The left window is the Data.gov homepage, featuring a search bar, a 'Most Viewed Datasets' section, and a 'Recently Added Datasets' section. The right window is a detailed view of a specific dataset titled 'Monthly Counts of Deaths by Select Causes, 2014-2019'. This page includes the dataset's logo (U.S. Department of Health & Human Services), its metadata (last updated April 21, 2022), an 'Access & Use Information' section, and a 'Downloads & Resources' section where a 'Comma Separated Values File' is available for download.

Abrir página **data.gov**
<https://data.gov/>

Procurar conjunto de dados “**Monthly Counts of Deaths by Select Causes, 2014-2019**”
<https://catalog.data.gov/dataset/monthly-counts-of-deaths-by-select-causes-2014-2019-da9df>

Descarregar ficheiro de dados **CSV**
Comma Separated Values File >
Download

1. Carregar ficheiro de dados

The screenshot shows a Microsoft Excel spreadsheet titled "Workshop Excel exercicios". The Data tab is selected. A table is displayed with columns: Jurisdiction of Occurrence, Year, Month, All Cause, Natural Cause, Septicemia, Malignant Neoplasms, Diabetes, and others. The data starts from row 1 and continues down to row 20. The first few rows show data for the United States from 2014 to 2019.

	Jurisdiction of Occurrence	Year	Month	All Cause	Natural Cause	Septicemia	Malignant Neoplasms	Diabetes	Other
1	United States	2014	1	243 298	226 621	3 944	51 101	45 558	1 100
2	United States	2015	1	265 355	247 269	4 194	52 346	46 226	1 100
3	United States	2016	1	245 823	227 341	3 846	51 863	48 258	1 100
4	United States	2017	1	262 832	241 918	4 089	52 120	46 531	1 100
5	United States	2018	1	286 744	265 418	4 502	52 876	46 447	1 100
6	United States	2019	1	257 649	237 219	3 580	52 087	46 739	1 100
7	United States	2014	2	211 980	197 001	3 214	45 558	46 447	1 100
8	United States	2015	2	227 047	211 028	3 559	46 226	46 226	1 100
9	United States	2016	2	230 021	212 140	3 459	48 258	48 258	1 100
10	United States	2017	2	233 819	215 160	3 511	46 531	46 531	1 100
11	United States	2018	2	236 998	218 466	3 660	46 447	46 447	1 100
12	United States	2019	2	232 821	214 074	3 257	46 739	46 739	1 100
13	United States	2014	3	228 477	212 045	3 451	50 646	50 646	1 100
14	United States	2015	3	242 712	224 708	3 723	51 192	51 192	1 100
15	United States	2016	3	244 283	224 696	3 773	51 566	51 566	1 100
16	United States	2017	3	251 732	231 028	3 796	50 785	50 785	1 100
17	United States	2018	3	248 805	228 464	3 673	50 689	50 689	1 100
18	United States	2019	3	254 929	234 074	3 559	51 450	51 450	1 100
19	United States	2014	4	215 600	199 454	3 125	48 304	48 304	1 100

1. Abrir ficheiro CSV no Excel
2. Converter dados para tabela
[Data > Text to Columns]
3. Converter tabela em tabela nomeada
[Insert > Table]
4. Gravar ficheiro como XLSX
[Menu: File > Save As... Excel Workbook (.xlsx)]

Extra

5. Ajustar tamanho de todas as colunas
6. Alterar esquema de cores
[Page Layout > Colors]

2. Ordenação e Filtragem

The screenshot shows a Microsoft Excel spreadsheet titled "Workshop Excel exercicios". The table is titled "Jurisdiction of Occurrence" and contains data from January to December 2014 for the United States. The columns include Jurisdiction of Occurrence, Year, Month, All Cause, Natural Cause, Septicemia, Malignant Neoplasms, and Diabetes Mellitus. The "Natural Cause" column is currently selected. The "View" tab is selected in the ribbon. The status bar at the bottom shows "Ready" and "12 of 72 records found".

	A	B	C	D	E	F	G	H
1	Jurisdiction of Occurrence	Year	Month	All Cause	Natural Cause	Septicemia	Malignant Neoplasms	Diabetes Mellitus
2	United States	2014	1	243 298	226 621	3 944	51 101	
3	United States	2014	2	211 980	197 001	3 214	45 558	
4	United States	2014	3	228 477	212 045	3 451	50 646	
5	United States	2014	4	215 600	199 454	3 125	48 304	
6	United States	2014	5	216 862	199 843	3 105	49 497	
7	United States	2014	6	204 687	187 644	2 947	48 103	
8	United States	2014	7	209 373	192 035	3 112	49 259	
9	United States	2014	8	208 013	190 204	2 931	49 519	
10	United States	2014	9	205 274	188 322	2 918	48 331	
11	United States	2014	10	218 147	200 741	3 203	50 900	
12	United States	2014	11	221 317	204 268	3 192	49 377	
13	United States	2014	12	243 390	225 948	3 798	51 105	

1. Ordenação simples

- Considerando todas as causas de morte, qual o mês com menos morte registadas?
- Qual o mês com mais mortes por Septicemia?

2. Ordenação múltipla: Apresentar os dados por ordem cronológica (ie. ano/mês) [Data > Sort]

- Aplicar **formatação condicional** na coluna "Malignant Neoplasms" para visualizar valores por escala de cores
[Home > Conditional Formatting > Color Scales]

4. Filtro simples: Filtrar registos do ano 2014

- Múltiplos filtros: Filtrar último trimestre dos anos 2018-2019

6. Filtro condicional:

- Quanto meses têm >250.000 mortes registadas?

3. Acrescentar colunas com valores calculados

	A	B	C	D	E	V	W	X
1	Jurisdiction of Occurrence	Year	Month	All Cause	Natural Cause	Unnatural Cause	Unnatural perc	
2	United States	2014	1	243 298	226 621	16 677	6,9%	
3	United States	2015	1	265 355	247 269	18 086	6,8%	
4	United States	2016	1	245 823	227 341	18 482	7,5%	
5	United States	2017	1	262 832	241 918	20 914	8,0%	
6	United States	2018	1	286 744	265 418	21 326	7,4%	
7	United States	2019	1	257 649	237 219	20 430	7,9%	
8	United States	2014	2	211 980	197 001	14 979	7,1%	
9	United States	2015	2	227 047	211 028	16 019	7,1%	
10	United States	2016	2	230 021	212 140	17 881	7,8%	
11	United States	2017	2	233 819	215 160	18 659	8,0%	
12	United States	2018	2	236 998	218 466	18 532	7,8%	
13	United States	2019	2	232 821	214 074	18 747	8,1%	
14	United States	2014	3	228 477	212 045	16 432	7,2%	
15	United States	2015	3	242 712	224 708	18 004	7,4%	
16	United States	2016	3	244 283	224 696	19 587	8,0%	
17	United States	2017	3	251 732	231 028	20 704	8,2%	
18	United States	2018	3	248 805	228 464	20 341	8,2%	
19	United States	2019	3	254 929	234 074	20 855	8,2%	
20	United States	2014	4	215 600	199 454	16 146	7,5%	

- Acrescentar nova coluna com valores de mortes por causas não naturais
Unnatural Cause: =[@[All Cause]] - [@[Natural Cause]]

- Acrescentar coluna com percentagem de mortes por causa não natural
Unnatural perc: =[@[Unnatural Cause]] / [@[All Cause]]

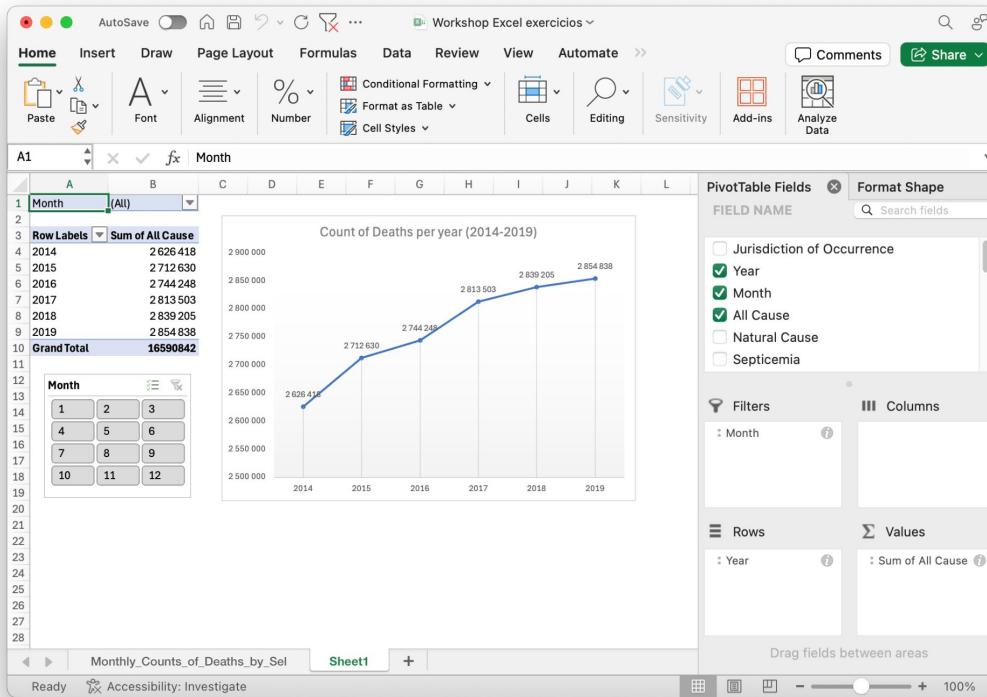
- Formatar valores da coluna "Unnatural perc" para percentagem (%) com 1 casa decimal

- Formatar todos os restantes valores numéricos (exceto ano e mês) para incluir separador de milhares - eg. 243 289

Extra

- Ocultar colunas intermédias (col. F-U)
- Remover linhas de grelha [View > Show > Gridlines]

4. Tabelas dinâmicas (Pivot Table)



- Criar nova tabela dinâmica [Insert > PivotTable]
 - All Cause → Values
 - Year → Rows
 - Month → Columns
- Passar Month da caixa Columns para a caixa Filters; testar filtro (eg. selecionar apenas mês 1)
- Inserir gráfico de linhas, para ver evolução do número de mortes ao longo dos anos

Extra

- Testar opções de formatação do gráfico
- Introduzir Slicer para seleção de meses (Month)

5. Tabela manual vs Tabela dinâmica

The screenshot shows a Microsoft Excel interface with two tables. The top table, labeled 'All Cause', is a manual table with columns for Year, Sum, Count, Average, and SD. The bottom table, also labeled 'All Cause', is a dynamic PivotTable with columns for Year, Sum, Count, Average, and StdDev. The PivotTable Fields pane on the right shows fields for Jurisdiction of Occurrence, Year, Month, All Cause, and Natural Cause, with Year and All Cause selected. The ribbon at the top shows various tabs like Home, Insert, Page Layout, etc., and the status bar at the bottom indicates 'Ready' and 'Accessibility: Investigate'.

	Sum	Count	Average	SD
2014	2 626 418	12	218 668	
2015	2 712 630	12	226 053	
2016	2 744 248	12	228 687	
2017	2 813 503	12	234 459	
2018	2 839 205	12	236 600	
2019	2 854 838	12	237 903	
Overall	16 590 842	72	230 428	15 602

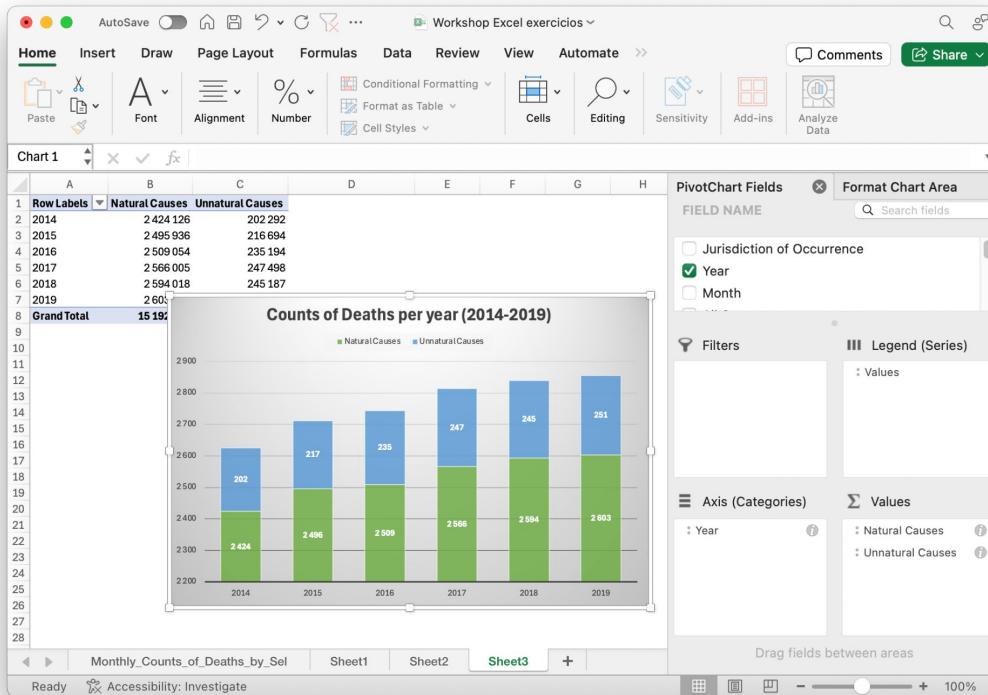
Row Labels	Sum	Count	Average	StdDev
2014	2 626 418	12	218 868	13 325
2015	2 712 630	12	226 053	15 483
2016	2 744 248	12	228 687	12 024
2017	2 813 503	12	234 459	15 083
2018	2 839 205	12	236 600	18 567
2019	2 854 838	12	237 903	12 122
Grand Total	16 590 842	72	230 428	15 602

- Numa nova folha, desenhar manualmente uma tabela com:
 - Linhas: anos de 2014 a 2019
 - Ir acrescentando colunas com fórmulas:

Sum	= SUMIF()
Count	= COUNTIF()
Average	= AVERAGEIF()
StdDev	= STDEV.S()

[Nota: o Excel não tem fórmula para cálculo condicional do desvio padrão!]
- Criar nova tabela dinâmica, a colocar por baixo da tabela manual
 - Year → Rows
 - All Cause → Values (4x)
 - Alterar a estatística sumária das colunas para Sum, Count, Average, Standard Deviation

6. Gráficos dinâmicos (Pivot Chart)

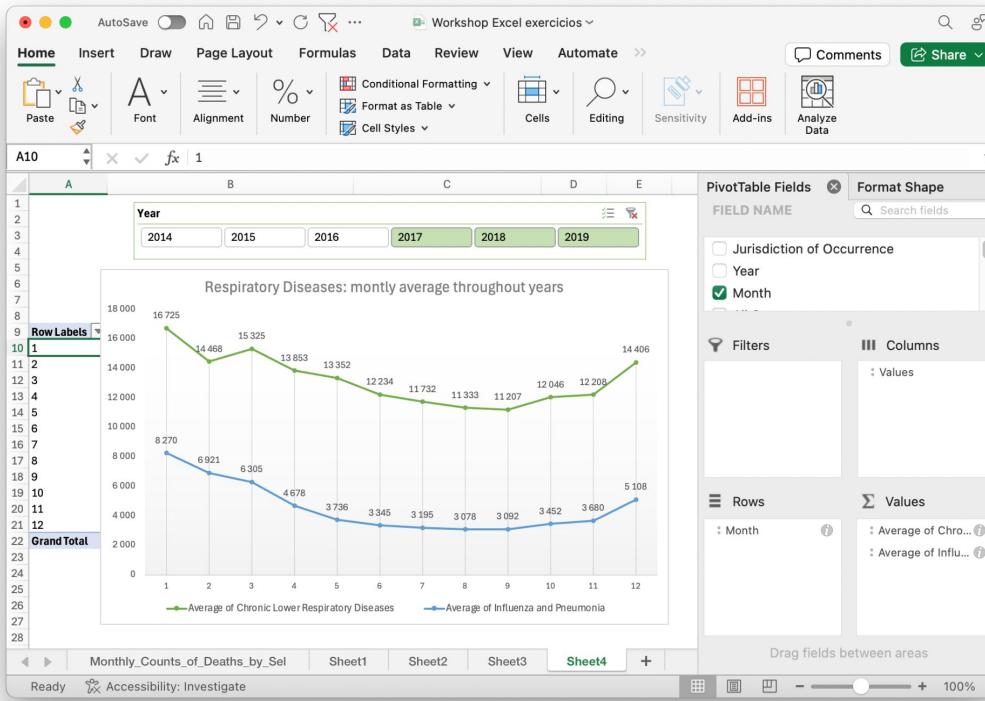


- Criar nova tabela dinâmica [Insert > PivotTable]
 - All Cause → Values
 - Year → Rows
 - Month → Columns
- Passar Month da caixa Columns para a caixa Filters; testar filtro (eg. selecionar apenas mês 1)
- Inserir gráfico de linhas, para ver evolução do número de mortes ao longo dos anos

Extra

- Testar opções de formatação do gráfico
- Introduzir Slicer para seleção de meses (Month)

7. Análise exploratória de dados com tabelas e gráficos dinâmicos



Sugestão

- Usar uma nova tabela dinâmica para explorar a média mensal de ocorrência de morte por doenças respiratórias
 - Apresentar os resultados em gráfico
 - Acrescentar slicer para a variável Year

Bom trabalho!

João Alves

jalves@med.up.pt