

# FOSE1025 — Scientific Computing

## Week 8 Lecture 1: Transforming Data

Diego Mollá

Department of Computer Science  
Macquarie University

FOSE1025 2021H2

# Programme

- 1 Dates
- 2 Long and Wide Formats

## Reading

- LinkedIn Learning "Excel 2016, Cleaning up your data", chapter 3.
- MATLAB dates and time.
- LinkedIn Learning "Excel pivot tables for beginners".

# Programme

## 1 Dates

- Dates in Excel
- Dates in MATLAB

## 2 Long and Wide Formats

# Processing Dates

(This part belongs to “cleaning data,” really ...)

- Dates come in many formats, we need to make sure they are in the format we need.
  - dd/mm/yyyy (Australia)
  - dd.mm.yyyy (Germany)
  - mm/dd/yyyy (USA)
  - yyyy/mm/dd (Japan)
  - ...
- If input manually, check if there are errors!
  - 24 Maye 2020

# Programme

- 1 Dates
  - Dates in Excel
  - Dates in MATLAB
- 2 Long and Wide Formats

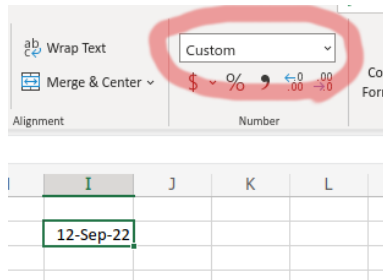
# Excel Dates and Times Are Numbers



Excel represents dates and times as numbers. These are called “serial numbers” and they represent the number of days since a specific date: 1st January 1900.

## Demonstration 1

Type **12 Sep 2022** in an Excel cell and observe how it shows the date (see screenshot). Change the cell format to “Number”. You will see that the cell now displays the number **44816.00**.



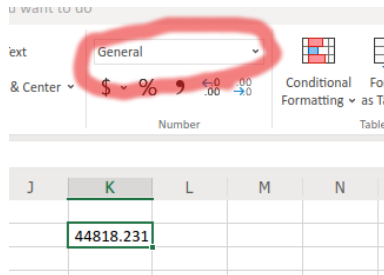
# Excel Dates and Times Are Numbers



Excel represents dates and times as numbers. These are called “serial numbers” and they represent the number of days since a specific date: 1st January 1900.

## Demonstration 2

Type the number **44818.231** in an Excel cell and change the format to “Short Date”. You will see the date **9/14/2022**. Change the format now to “Time.” You will see the time **5:32:38 AM**.



# Useful Excel Functions to Manipulate Dates and Times

## Creating Dates and Times

`DATE(year,month,day)`: Create a date from numbers.

`TIME(hours,minutes,seconds)`: Create a time from numbers.

`DATE(year,month,day) + TIME(hours,minutes, seconds)`: Create a date with time.



# Useful Excel Functions to Manipulate Dates and Times

## From Dates and Times to Text

`TEXT(serial_number,pattern)`

Represent a date as text using a specific pattern. For example, if cell A1 has the formula `=DATE(2020,12,23) + TIME(21,35,12)`:

`TEXT(A1, "dd/mm/yy")` returns the value "23/12/20".

`TEXT(A1, "dd/mm/yyyy hh:mm")` returns the value  
"23/12/2020 21:35".

`TEXT(A1, "dd mmm yyyy hh:mm:ss")` returns the value "23 Dec  
2020 21:35:12" (notice the three "m"?).

`TEXT(A1, "dd mmmm yyyy hh:mm AM/PM")` returns the value  
"23 December 2020 09:35 pm".

# Useful Excel Functions to Manipulate Dates and Times

## From Text to Dates and Times

### DATEVALUE(text)

Convert text into a serial number that represents the date. This function does not convert times, only dates.

`DATEVALUE("12 May 2021")` returns the value 44328.

`DATEVALUE("12 May 2021 3:15pm")` returns the same value 44328.

### VALUE(text)

Convert text into a serial number that represents the date and time.

`VALUE("12 May 2021")` returns the value 44328.

`VALUE("12 May 2021 3:15pm")` returns the same value 44328.64.

## Exercise: Dates in Different Formats

Ch-03.xlsx from <https://www.linkedin.com/learning/excel-2016-cleaning-up-your-data>

- 1 What formula would you type in cell B2?
- 2 What formula would you type in cell D2?

	A	B	C	D
1	Month Year	=DATE(Year, Month, Day)	Year Month	=DATE(Year, Month, Day)
2	10 2016		2016 10	
3	4 2016		2016 4	
4	5 2016		2016 5	
5	9 2015		2015 9	
6	10 2016		2016 10	
7	6 2016		2016 6	
8	4 2015		2015 4	
9	5 2016		2016 5	
10	1 2016		2016 1	
11	12 2015		2015 12	
12	12 2015		2015 10	
13	11 2015		2015 11	
14	8 2016		2016 8	
15	11 2016		2016 11	
16	8 2015		2015 8	
17	8 2015		2015 8	

## Exercise: Mixed date formats in one column

Create a blank Excel worksheet, import this CSV file, and normalise the dates so that they appear as in the screenshot.

`dates.csv`

	A	B	C	D	E	F	G	H
1	Date	Name	Email	Consultati	Zoom			
2	Tuesday, December 1, 2020	Diego Mol	diego.molla-aliod@mq.edu.au					
3	Tuesday, May 12, 2020	Gaurav Gu	gaurav.gupta@mq.edu.au					
4	Wednesday, April 15, 2020	Urvashi Kh	urvashi.kh	Wed 12-1	<a href="https://macquarie.zoom.us/j/472568461">https://macquarie.zoom.us/j/472568461</a>			
5	Monday, November 23, 2020	Munazza Z	munazza-z	Wed 11-1	<a href="https://macquarie.zoom.us/j/267542550">https://macquarie.zoom.us/j/267542550</a>			

# Operating with Excel Dates



## Extraction

The following commands extract parts of a date and time:

- YEAR, MONTH, DAY, HOUR, MINUTE, SECOND
- The result is a number, not a date (i.e. not a serial number)

## Time difference

- The following command can be used to find the difference between two dates:
  - DATEDIF(date1, date2, "y") — difference in years
  - DATEDIF(date1, date2, "m") — difference in months
  - DATEDIF(date1, date2, "d") — difference in days
  - Again, the result is a number, not a date.
- If cells A2 and B2 contain dates, then:
  - B2 - A2 is the time difference in days (and fraction of days).

# Programme

- 1 Dates
  - Dates in Excel
  - Dates in MATLAB
- 2 Long and Wide Formats

# Understanding Dates in MATLAB



<https://au.mathworks.com/help/matlab/date-and-time-operations.html>

- As with Excel, MATLAB has a specific data format for date-time.
- MATLAB's `datetime` allows one to create a date-time. It accepts several formats, including:
  - year, month, day
  - year, month, day, hour, minute, second

```
hello_date = datetime(2020, 7, 3, 18, 30, 23)
hello_date = datetime(2020, 7, 3)
```

MATLAB does not try to guess the meaning of each number.

**They must be placed in the correct order.** Compare these:

```
date1 = datetime(2020, 7, 3)
date2 = datetime(2020, 3, 7)
date3 = datetime(3, 7, 2020)
```

# From Text to Dates and Back



- Sometimes we want to convert a string containing a date (and or time) into MATLAB's date-time, or vice-versa.
- MATLAB's `datetime` can convert from text (and other types) to date.

```
t = datetime('21/09/2020')
```

- MATLAB's `string` converts from date (and other types) to text.

```
w_table.StringDate = string(w_table.Date,  
                             'MM/dd/yyyy')
```



# Text Date Formats in MATLAB

<https://au.mathworks.com/help/matlab/ref/datetime.html#buhzxm1-Format>

- As with Excel, MATLAB allows to read and write dates using different formats.
- MATLAB formats are slightly different from Excel's formats.
- These formats can also be used when importing from CSV files.

## Examples of Formats

Format	Example
dd-MMM-yyyy HH:mm:ss	01-Mar-2000 15:45:17
MM/dd/yyyy	03/01/2000
MM dd yy	03 01 00

## Example: Importing shopping.csv



- The file shopping.csv represents dates using the day and month only, using a specific format of the form “1-Jan”, “2-Jan”, etc.
- In MATLAB, specify the datetime format “dd-MMM” in the “Date” column when you use the data import wizard.

The screenshot shows the MATLAB Data Import Wizard interface. The 'Date' column is selected, and the format 'dd-MMM' is chosen from the dropdown menu. A red circle highlights the 'dd-MMM' option in the dropdown.

	A	B	C	D
	Date	Buyer	Type	Amt
1	Text	Categorical	Categorical	Number
2	Text			Amt
3	Number			\$50
4	Categorical			\$120
5	Datetime			\$10
6	Duration			+ 15 more
7	Set Type for All Selected Columns			
8	6-Jan	Kelly	Cafes	\$
9	7-Jan	Kelly	Books	\$129
10	7-Jan	Dad	Groceries	\$252

# Programme

## 1 Dates

## 2 Long and Wide Formats

- Long and Wide Formats
- Introducing Pivot Tables

# Programme

- 1 Dates
- 2 Long and Wide Formats
  - Long and Wide Formats
  - Introducing Pivot Tables

# Tables as 2D Data

- Remember that tables represent 2-dimensional information.
  - Rows indicate different records.
  - Columns indicate different types of data in the record.
- We can, for example, represent the work address (street, city, postcode, etc) of a group of people.

(file WorkAddresses.xlsx)

First Name	Last Name	Address	City	State	Post	Phone
Deane	Haag	9 Hamilton B	Sydney South	NSW	1235	02-9718-2944
Edelmira	Pedregon	50638 North	Bandy Creek	WA	6450	08-8484-3223
Andrew	Keks	51 Bridge Av	Carwarp	VIC	3494	03-5251-3153
Miesha	Decelles	457 St Sebas	Eltham	VIC	3095	03-5185-6258
Javier	Osmer	6 Ackerman	Doncaster Ea	VIC	3109	03-8369-6924
Kizzy	Stangle	8 W Lake St	Welbungin	WA	6477	08-1937-3980
Sharan	Wodicka	8454 6 17 M	Shenton Park	WA	6008	08-4712-2157
Novella	Fritch	5 Ellestad Dr	Girraween	NSW	2145	02-2612-1455
German	Dones	9 N Nevada	Woronora	NSW	2232	02-2393-3289
Robt	Blanck	790 E Wisco	Woodbury	TAS	7120	03-6517-9318
Rossana	Biler	60481 N Clar	Lee Point	NT	810	08-9855-2125

# Tables as 3D, 4D ... ?

- How would you keep information about the work **and the home address**?
- What if one person has 15 different properties, how do you store the information for all people?
- A solution: Add one column that indicates the type of address.
- (Databases can encode this information more efficiently using relational tables but this is not the topic of this unit.)

	A	B	C	D	E	F	G	H	
	First Name	Last Name	Address Type	Address	City	State	Post	Phone	
1	Deane	Haag	Work	9 Hamilton B	Sydney South	NSW	1235	02-9718-2944	
2	Edelmira	Pedregon	Work	50638 North	Bandy Creek	WA	6450	08-8484-3223	
3	Andrew	Keks	Work	51 Bridge Av	Carwarp	VIC	3494	03-5251-3153	
4	Miesha	Decelles	Work	457 St Sebas	Eltham	VIC	3095	03-5185-6258	
5	Javier	Osmer	Work	6 Ackerman	Doncaster Ea	VIC	3109	03-8369-6924	
6	Kizzy	Stangle	Work	8 W Lake St	Welbungin	WA	6477	08-1937-3980	
7	Sharan	Wodicka	Work	8454 6 17 N	Shenton Park	WA	6008	08-4712-2157	
8	Novella	Fritch	Work	5 Ellestad Dr	Girraween	NSW	2145	02-2612-1455	
9	German	Dones	Work	9 N Nevada	Woronora	NSW	2232	02-2393-3289	
10	Robt	Blanck	Work	790 E Wisco	Woodbury	TAS	7120	03-6517-9318	
11	Rossana	Biler	Work	60481 N Clar	Lee Point	NT	810	08-9855-2125	
12	Deane	Haag	Home	302 N 10th S	Oakleigh Sou	VIC	3167	03-9085-5714	
13	Edelmira	Pedregon	Home	79346 Firest	Gununa	QLD	4871	07-1217-9907	
14	Andrew	Keks	Home	37564 Grace	Salamander	NSW	2317	02-9187-4769	
15	Miesha	Decelles	Home	470 W Irving	Bundaberg N	QLD	4670	07-3963-4469	
16	Javier	Osmer	Home	6 Jefferson S	Middleton	SA	5213	08-5236-2143	

# Long and Wide Formats

- The tables that we are used to see are in a **wide format**.
  - Each column indicates a specific data: name, address, location, temperature, etc.
- For complex data we may want to use a **long format**.
  - One column indicates the type of data.
  - Another column (or columns) indicate the value.

(file weather\_data.csv)

	A	B	C	D	E	F
1		data	date	param	siteid	
2	1	0	1/1/03	Precipitation	ACRE	
3	2	0	2/1/03	Precipitation	Albert Lea	
4	3	11.3199997	3/1/03	Precipitation	Ames	
5	4	0	4/1/03	Precipitation	Antigo	
6	5	3.03999996	5/1/03	Precipitation	Appleton	
7	6	0.49000001	6/1/03	Precipitation	Arlington	
8	7	0	7/1/03	Precipitation	Bean & Beet	
9	8	0	8/1/03	Precipitation	Brookings	
10	9	0	9/1/03	Precipitation	Brownstown	
11	10	0	10/1/03	Precipitation	Columbia	
12	11	0	11/1/03	Precipitation	Crookston	

# Processing Tables in Long Format

The lecturer will demonstrate how to use filters and pivot tables to process tables in long format. File: shopping.csv

- Many tables are expressed in long format for some columns.
- Excel does not have a specific tool to process these tables.
- You have seen how you can use filters to focus on specific values.
- You have also seen how you can use conditional functions to calculate values of one column based on the values of another column.
  - e.g. `=SUMIFS(D:D,G:G,"Fuel")` sums all values in column D such that the cell in row G has the value "Fuel").
- You can also use [pivot tables](#).



# Programme

- 1 Dates
- 2 Long and Wide Formats
  - Long and Wide Formats
  - Introducing Pivot Tables

# Pivot Tables: A Motivational Example

(data from <https://www.linkedin.com/learning/excel-pivottables-for-beginners>)

- Find the total shopping in each category “Fuel”, etc, of file shopping.csv.
- Find the total shopping of each month.
- What shopping per month and per category??
- Pivot tables can help you generate data for all of above and more.

Date	Buyer	Type	Amt
1-Jan	Mom	Fuel	\$50
2-Jan	Mom	Groceries	\$120
3-Jan	Dad	Cafes	\$10
4-Jan	Dad	Fuel	\$40
4-Jan	Kelly	Groceries	\$129
5-Jan	Mom	Cafes	\$12
6-Jan	Kelly	Cafes	\$14

# A Simple Pivot Table



Excel - shopping - Saved

Search (Alt + Q)

File Home Insert Draw Page Layout Formulas Data Review View Automate Help PivotTable Editing Share Comments Catch up

Functions Forms Table PivotTable Picture Shapes Office Add-ins Recommended Column Line Pie Bar Area Scatter Other Charts Hyperlink New Comment

84 fx Books

	A	B	C	D	E	F	G	H	I	J	K	L	M	N	O
1	Buyer	(All)													
2															
3															
4	Sum of Amt	Column Labels													
5	Jan	169	36		271	209	2147	15						2847	
6	Feb	476	59		142	202	2820	15						3714	
7	Mar	160	48		51	329	2348	46		2519				5501	
8	Apr	418	34		307	100	2985	9		3299				7152	
9	May	96	63		240	288	2911	14		2136				5748	
10	Jun	38	145		309	198	2905	86		3352				7033	
11	Jul	60	33		722	228	2834	6		3419				7302	
12	Aug	79	38		143	138	3120	17		3651				7186	
13	Sep	61			163		2377	9		3783				6393	
14	Oct	39			165		3063	13		3492				6772	
15	Nov	67			927	117	2373	10		1030				4524	
16	Dec	328			2627	55	2786	9						5805	
17	Grand Total	1991	456		5739	2192	32669	249		26681				69977	
18															
19															
20															
21															
22															

PivotTable Fields

Choose fields:

- ☐ Date
- ☒ Month
- ☒ Buyer
- ☒ Type
- ☒ Amt

Drag fields between areas below:

**FILTERS**  
Buyer

**COLUMNS**  
Type

**ROWS**  
Month

**VALUES**  
Sum of Amt

Calculation Mode: Automatic Workbook Statistics

Give Feedback to Microsoft

# Anatomy of a Pivot Table

## Filters

- What column to use to filter values.
- Only for columns with categorical data.

## Rows

- What column to use in the rows of the pivot table.
- Only for columns with categorical data.

## Columns

- What column to use in the columns of the pivot table.
- Only for columns with categorical data.

## Values

- What value we want to aggregate.
- Only for columns with numerical data.

# Pivot Tables to Convert from Long to Wide

## Exercise 1 (weather\_data.csv)

What is the average precipitation in Antigo?

- Using AVERAGEIFS
- Using a pivot table

## Exercise 2 (weather\_data.csv)

What is the March-2013 average precipitation in Antigo?

- Using AVERAGEIFS
- Using a pivot table

	A	B	C	D	E	F
1		data	date	param	siteid	
2	1	0	1/1/03	Precipitation	ACRE	
3	2	0	2/1/03	Precipitation	Albert Lea	
4	3	11.3199997	3/1/03	Precipitation	Ames	
5	4	0	4/1/03	Precipitation	Antigo	
6	5	3.03999996	5/1/03	Precipitation	Appleton	

# Take-home Messages

- Both Excel and MATLAB have a specific data type that is used to represent Dates and times.
- Pay attention when importing files that use unconventional date and time expressions. Both Excel and MATLAB may guess the format wrong.
- Both Excel and MATLAB offer functions that can be used to create dates and convert dates to strings.
- Understand the power of Excel's pivot tables.

# What's Next

- Week 9 lecture: Summarising, Visualising and Analysing Data.
- Week 9: in-class test during your scheduled SGTA 1 (Friday 6-9pm for external students).
  - You can also find a practice test in iLearn. Complete it at your leisure.