FOSE1025 — Scientific Computing
Week 8 Lecture 1: Transforming Data

Diego Mollá

Department of Computer Science
Macquarie University

FOSE1025 2023H1

Programme

1. Dates

2. Long and Wide Formats

Reading

- These notes
- Readings listed in iLearn — Week 8

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

What might we want to do with dates?

- Sort by date
- Group (or plot) by month, by year, etc
- Determine the time elapsed between two dates or times

Programme

1. Dates
   - Dates in Excel
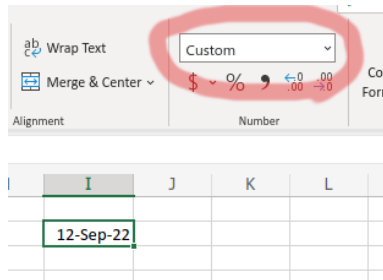   - Dates in MATLAB

2. Long and Wide Formats

Excel Dates and Times Are Numbers 📺

Internally, Excel stores 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.
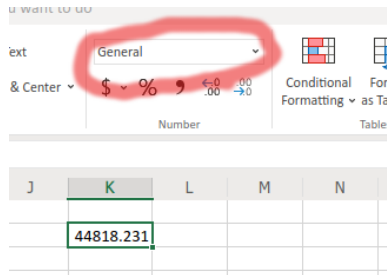
Excel Dates and Times Are Numbers 　📶

Internally, Excel stores 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 ⏭

File: DatesTimes.xlsx

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 text "23/12/20".

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

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

TEXT(A1, "dd mmmm yyyy hh:mm AM/PM") returns the text "23 December 2020 09:35 pm" (now there are four "m").

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 number 44328.

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

VALUE(text)

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

VALUE("12 May 2021") returns the number 44328.

VALUE("12 May 2021 3:15pm") returns the same number 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	9 2015		2015 9	

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

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

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 text 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#buhzxmk-1-Format

- 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.

Programme

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 Wiscor	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.)

Long and Wide Formats

- Most tables that we have seen in this unit so far are in a wide format.
  - Each column indicates 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	AlbertLea	
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	

Processing Excel Tables in Long Format

The lecturer will demonstrate how to use filters and pivot tables to process Excel 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,C:C,"Fuel") sums all values in column D such that the cell in column C 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.

	A	B	C	D	
	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	
	C Jan	Kelly	Cafes	$14	

A Simple Pivot Table

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 that 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
L		data	date	param	siteid	
2	1	0	1/1/03	Precipitation	ACRE	
3	2	0	2/1/03	Precipitation	AlbertLea	
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 do not use conventional date and time expressions. In those cases, Excel and MATLAB may guess the wrong format.
- Both Excel and MATLAB offer functions that can be used to create dates and convert dates to text.
- You need to understand the power of Excel's pivot tables.

What's Next

Assessments this week

- Problem Solver employability hurdle: Friday 28 April

Next Week

- Week 9 lecture: Summarising, Visualising and Analysing Data.