FOSE1025 — Scientific Computing
Week 8 Lecture 1: Transforming Data

Diego Mollá

Department of Computer Science
Macquarie University

FOSE1025 2020H2

Programme

Reading
- Lecture notes

Programme

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
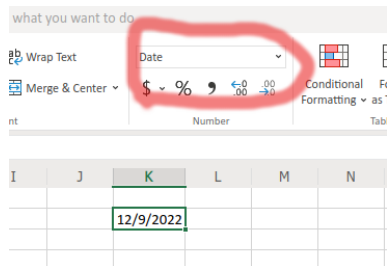
Programme

1. Dates
   - Dates in Excel
   - Dates in MATLAB

2. Long and Wide Formats

Excel Dates Are Not Text or Numbers

Excel does not represent dates and times as text or numbers. They are called "serial numbers" and they represent the number of days since a specific date: 1st January 1900.

Exercise 1

Type 12/9/22 in an Excel cell and observe the format (see screenshot). Change the cell format to "Number". What do you see in the cell?
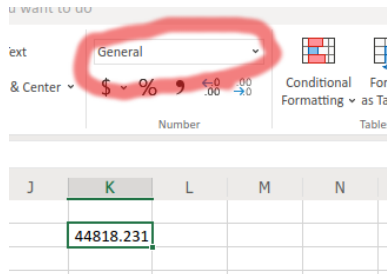
Excel Dates Are Not Text or Numbers

Excel does not represent dates and times as text or numbers. They are called "serial numbers" and they represent the number of days since a specific date: 1st January 1900.

Exercise 2

Type the number 44818.231 in an Excel cell and change the format to "Short Date", what do you see? Change the format now to "Time." What do you see?

Your turn: Participation activity

Complete questions 1 and 2 of this week's lecture participation quiz.

Useful Excel Functions to Manipulate Dates
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
Formatting Dates 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".

Example 1: 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	

Your turn: Participation activity

Complete questions 3 and 4 of this week's lecture participation quiz.

Exercise: Mixed date formats in one column

Create a blank Excel worksheet, import this CSV file, and normalise the dates.

dates.csv

Date , Name , Email , Consultation Times , Zoom
12/01/2020 , Diego Molla−Aliod , diego . molla−aliod@mq . ec
12 May 2020 , Gaurav Gupta , gaurav . gupta@mq . edu . au , ,
15 Apil 2020 , Urvashi Khanna , urvashi . khanna@mq . edu . au
2020−11−23 , Munazza Zaib , munazza−zaib@mq . edu . au , Wed

- If you just double-click on the CSV file and let Excel import the file using defaults, the resulting dates look strange... why?
- Hint: don't let Excel use the General format for the first column.

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, given:
  - year
  - month
  - day
  - hour
  - minute
  - second

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

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')

Your turn: Participation activity

Complete question 5 of this week's lecture participation quiz.

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 E	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.)

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

Long and Wide Formats

- The tables that we are used to see are in the 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)

Processing Tables in Long Format

The lecturer will demonstrate how to use filters and pivot tables to process tables in long format

- Many tables are expressed in long format for some columns.
- Excel does not have a specific tool to process these tables.
- You can use filters to focus on specific values.
- You can also use pivot tables.
- We will see pivot tables more in detail next week, but here we see how to use them to process tables in long format.

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

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 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
L		data	date	param	siteid	
2	1	0	1/1/03	Precipitatior	ACRE	
3	2	0	2/1/03	Precipitatior	AlbertLea	
4	3	11.3199997	3/1/03	Precipitatior	Ames	
5	4	0	4/1/03	Precipitatior	Antigo	
6	5	3.03999996	5/1/03	Precipitatior	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 quiz during your scheduled practical 1 (Friday 6-9pm for external students).
  - You can also find a practice quiz in iLearn. Complete it at your leisure.