Data workshop - Excel

Aileen McDonald
May 2021

Who am I?

- Cohort lead & Data Instructor at CodeClan

  CodeClan is a not-for-profit social enterprise in Scotland. The first and only SQA accredited digital skills academy on a mission to help bridge the digital skills gap in Scotland's growing tech industry.

- Volunteer at DataKindUK

  DataKindUK is a charity that builds data science capacity in social change organisations.

Objectives & approach

Objectives

- Introduction to transferrable Excel skills.

Approach

- Introduce concepts.
- Work through examples together

Resources

- Slides with step by step guide will be available after the talk.

Introduction to Microsoft Excel

Excel is a spreadsheet application developed and published by Microsoft. It is part of the Microsoft Office suite of products (including Word, PowerPoint etc.).

Estimated about 750 million people worldwide use Excel. [1]

Excel organizes data in columns and rows and allows you to manipulate and visualise your data.

Excel skills are hugely valuable and transferrable.

Open ‘excel_workshop_examples1’ file

Foundations

Terminology

Workbook: spreadsheet

Worksheet: tabs within a workbook

Rows: left hand side, numbers

Columns: along top, letters (A-Z, AA-ZZ,...etc)

Cell: row and column combined (e.g. A1, B2)

Area: top left to bottom right cell in area (e.g. A1:C4)

Layout

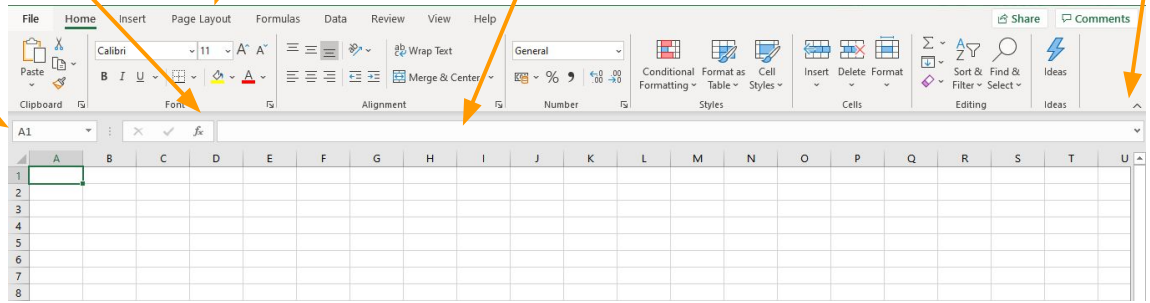
Tabs - Home, Insert etc. (contain different functionality)

Formula button

Formula bar

Pin the toolbar so always showing if it's not already (bottom right of toolbar, click on pin)

Cell/area selected

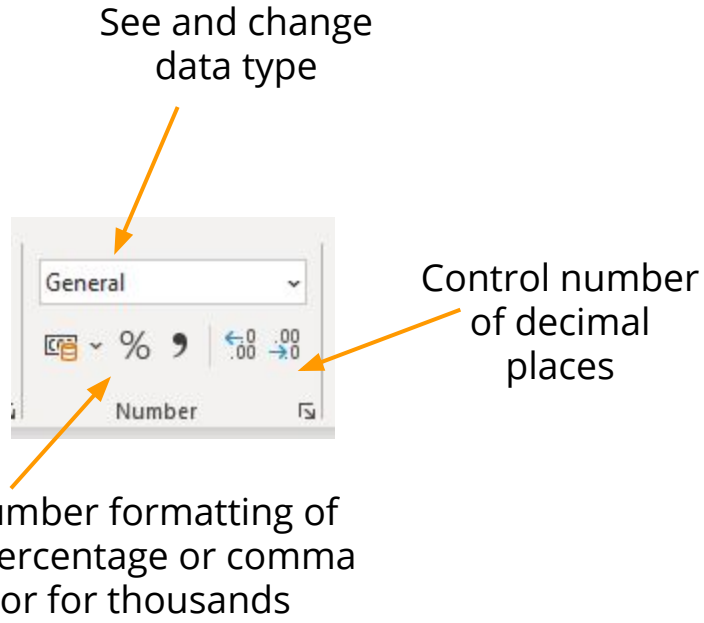


Formatting

Formatting cells

The number 1 displayed as each of the main data types/formats

Number	1
Currency	£1.00
Accounting	£ 1.00
Short date	01/01/1900
Long date	01 January 1900
Time	00:00:00
Percentage	100.00%
Fraction	1
Scientific	1.00E+00
Text	1

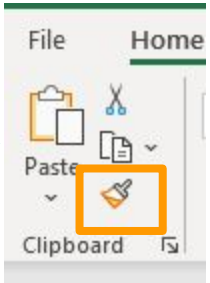


More formats: Right click cell/area -> *'Format'* -> more data types and customisable

Format painter

- If you take time to format cells (perhaps font, colour, data type) you may want to ‘copy’ this formatting to other cells.
- Select area want to copy, click ‘*Format painter*’ and then select area want to apply it to.

Formatted cells	Unformatted cells
£3.00	3
£4.75	4.75
£3.11	3.109



Formatted cells	Unformatted cells
£3.00	£3.00
£4.75	£4.75
£3.11	£3.11

Dealing with width & height

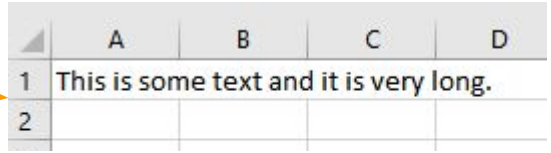
- Dealing with badly adjusted width and heights of columns and rows can make hard to read and explore data.
- In this example:
  - Column A the title is hidden (because of column width)
  - Column C is too wide
  - Row 4 is too tall (unnecessary space)
- '*Format*' -> '*AutoFit*' row height and/or column width to resolve
  - Shortcut - double clicking on edge of row/column header

	A	B	C	D
1	Service U	Name	Age	
2	1	John	33	
3	2	Frank	45	
4	3	Polly	21	
5	4	David	56	
6				
7				

	A	B	C	D
1	Service User ID	Name	Age	
2	1	John	33	
3	2	Frank	45	
4	3	Polly	21	
5	4	David	56	

Dealing with long text entries

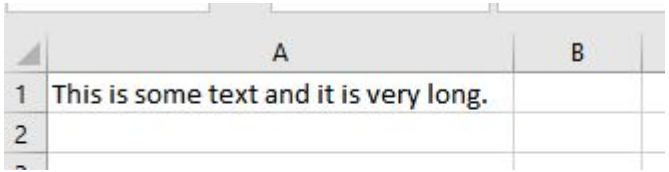
This text sits in column A but runs over into other columns.



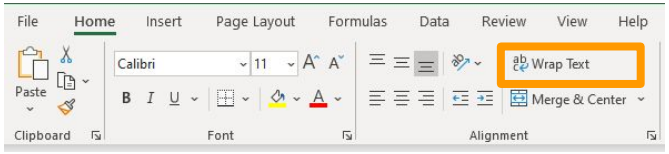
Can double click on edge of column A header -> expands the column

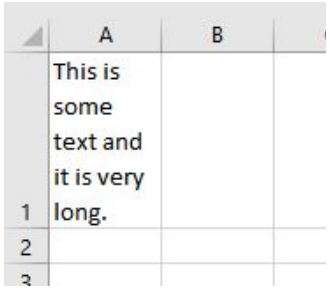
OR

'Format' -> 'Autofit Column Width'



'Text wrap' -> moves text onto new lines inside the cell

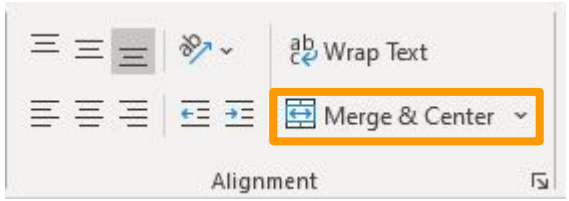




Merge cells

- Can be useful for headers across tables
- **BUT** can cause issues selecting columns and so with formulas (based on columns)

E	F	G
Table 2		
Service	Value	Date
A	5	03/01/2010
B	43	06/04/2019
B	6	15/03/2018
C	8	29/10/2019
C	12	02/06/2019



E	F	G
Table 2		
Service	Value	Date
A	5	03/01/2010
B	43	06/04/2019
B	6	15/03/2018
C	8	29/10/2019
C	12	02/06/2019

Tips - data format/layout

- One data type/variable in a single column
- Don’t have headers half way through
- Don’t have colours to distinguish different category of data (if plan on using any formulas on it, ok if just for visual or manual filters)

Note: these type of tips not an issue if data is just to ‘visualise’ but if plan on calculations etc. then may be an issue

Tips - data format/layout

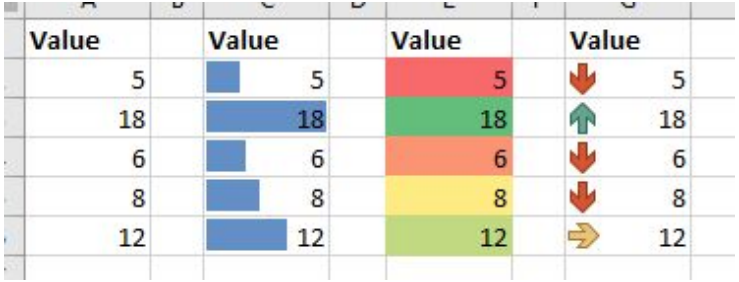
	A	B	C	D	E
1	**Service A**	**Service B**	**Service C**		Legend
2	**Volunteer**	**Volunteer**	**Volunteer**		No longer active
3	Gary	Juila	Erin		
4	David		Frank		
5	Jo				
6	**Staff**	**Staff**	**Staff**		
7	Terry		Heather		
8	Kerri		Ric		
9			Louise		
10					
11					

	Name	Type	Service	Status
2	Gary	Volunteer	A	Active
3	David	Volunteer	A	Active
4	Jo	Volunteer	A	Active
5	Terry	Staff	A	Unactive
6	Kerri	Staff	A	Unactive
7	Juila	Volunteer	B	Active
8	Erin	Volunteer	C	Active
9	Frank	Volunteer	C	Unactive
10	Heather	Staff	C	Active
11	Ric	Staff	C	Active
12	Louise	Staff	C	Active
13				

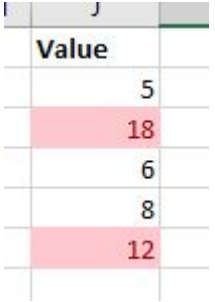
Note: these type of tips not an issue if data is just to ‘visualise’ but if plan on calculations etc. then may be an issue

Conditional formatting

- Select data -> *'Home'* -> '*Conditional Formatting'* -> multiple formatting options
- Options such as colors ('*Colour Scales*'), icons ('*Icon Sets*'), and data bars ('*Data Bars*'), which are based on the cell values compared to the rest of the cell values in the selected area.



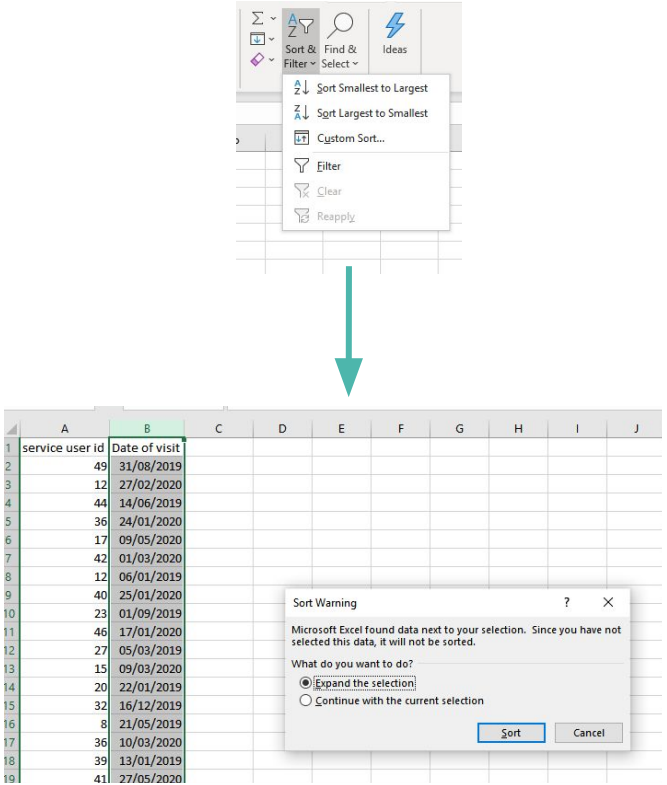
- Can also specify particular formatting rules e.g. highlight if value is over a threshold.
  - In this example have specified want to highlight cells which over value of 10 (and that will be red fill and red text).



Sorting & filtering

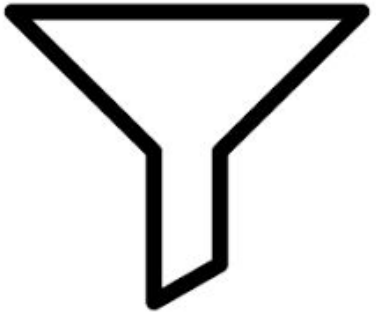
Sorting data - find small/largest

- Select the column you wish to sort the data on and then will get this pop up:
  - This is Excel trying to be helpful - if you just sort based on this column lose relationship between ids and dates
  - Select *'Expand the Selection'* to ensure the relationship is preserved
- If select all columns and sort then it will sort on the left hand side column.
- Can also do '*Custom Sorts*' (e.g. on colour of cell or text A-Z) and can sort on multiple columns



Filtering

- Based on value: manually select which ones to keep/drop.
- Based on value:
  - **Numeric** (greater than a value, between 2 values, above average etc.)
  - **Text** (contains, begins with, equals etc.)
- Based on colour
- Can also sort data from the filter dropdown.



Series

Series

- Select cell(s) and dragging extends patten
- Select single cell and drag
  - If number -> copies number down
  - If date -> adds 1 to date
- Could also use a formula and drag to extend

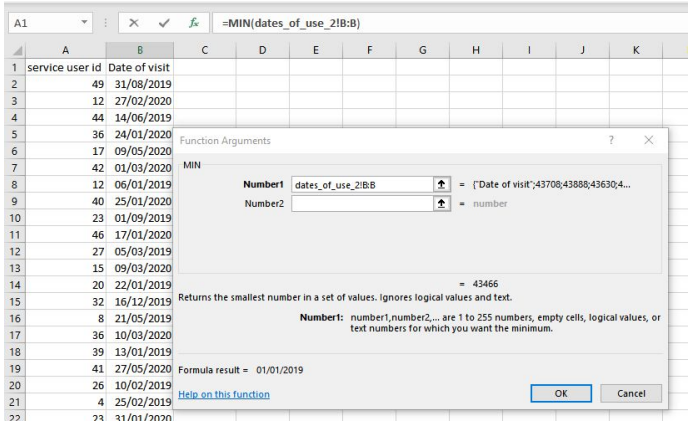
Formulas

Formulas

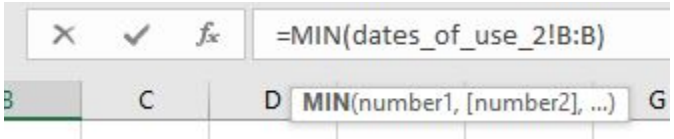
Google formulas (either specific formulas or what what it do do to find name)

Format of formulas
=FORMULA(argument1, argument2,…)

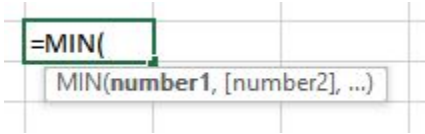
Click on formula button get information the formula and the arguments



Click into formula bar get list of arguments



Start typing and get list of arguments



Calculations & formula

- **Calculations:** multiply(*), divide(/), add(+), subtract(-), power(^) etc.
- **Formulas:** predefined formulas in Excel.

Frequently used functions
SUM(), AVERAGE(), MAX(), MIN(), COUNT(), COUNTA(), IF(), SUMIF(), COUNTIF(), AVERAGEIF()

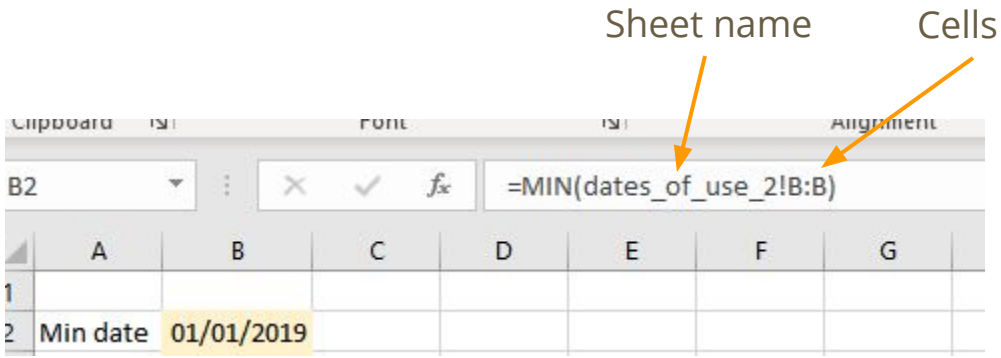
COUNT() (counts numeric cells in the range) and **COUNTA()** (counts all cells in the range) - both useful for finding how many non-blank values there are in a range.

IF() can be used to apply logic to create indicator columns.

SUMIF()/COUNTIF()/AVERAGEIF() used to sum/count/sum values if entries in another column meet some condition.

Formula - find small/largest

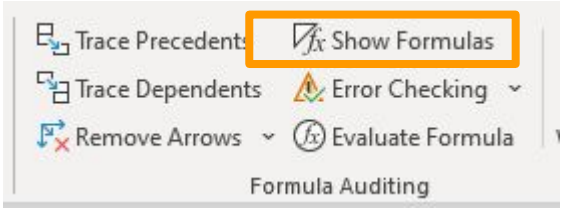
- If add new entry to the dates data then to find max/minimum would need to re-sort.
- **OR** can use formulas to calculate MIN and MAX dates
  - **Use formulas on full column** so that if new entry is added then it is included



Show formula

Can toggle '*Show Formulas*' on and off to see what formulas in your worksheet.

Useful if not sure which cells contain formulas.



Stepping through formulas

Can use *'Evaluate Formula'* to step through formulas to help work out what they are doing.

Useful for errors or generally to understand the calculations better as can help identify where the error is coming from.

Click through using the *'Evaluate'* button to do step-by-step.

References

Relative & absolute reference

Can drag formulas in cells to ‘copy’ the formula down.

But the cell **relative reference** will change.

If don’t want this use ‘$’ to create **absolute references** (which can apply to the row or column or both).

Open ‘excel_workshop_examples2’ file

Index & Match

Index/Match

- Use when want to ‘lookup’ value based on some kind of reference (e.g. id, name etc.).
- Use a combination of functions MATCH() and INDEX()
- Better than VLOOKUP() for number of reasons
    - Speed
    - Issues with VLOOKUP() if insert/delete rows

Index/Match (single column)

1. 2 datasets: each with same reference column (in our case user id)
2. Use **MATCH()** to find the row number that the reference is found in the 2nd dataset.

   =MATCH(cell, area_to_find_match, 0)
3. Use **INDEX()** to find the value in the column of interest for the row number found using **MATCH()**.

   Column wise INDEX only:
   =INDEX(column_to_find_value, row_number)

Index/Match - further

- Can also use to match simultaneously across both rows and columns
- Can also do it across spreadsheets **BUT** be careful if file is moved as will cause issues (as file path will change).
    - **Tip:** do INDEX/MATCH and then copy and paste results as values so no longer linked to other spreadsheet.

#NA error

- If there isn't a match using INDEX/MATCH then will get an **#NA** error.
- **More advanced**: can use **IFNA()** (or **IFERROR()**) function to get round such errors if necessary
    - Can replace with a value - perhaps either a numeric value or a more meaningful message).
    - Resource here for how to use.

Errors

Common errors

Below are some common errors may see when using Excel:

Common Excel Errors		
#N/A	#VALUE!	#NUM!
#NULL!	#DIV/0!	#REF!
#NAME	########	LOGIC?

Resource for how to deal with each of these errors here.

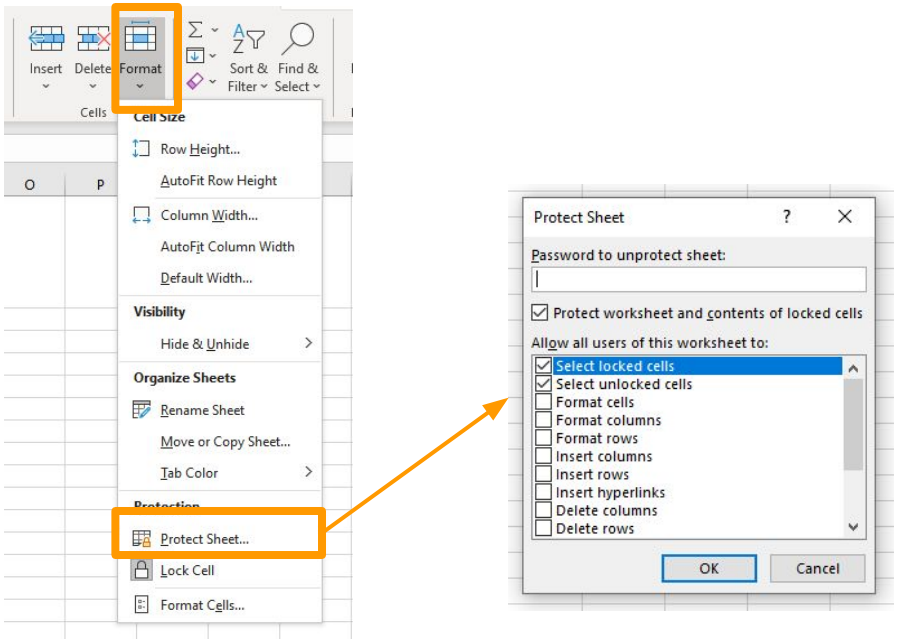
Restricting changes

Restricted changes to sheets and/or cells

Can password protect worksheets and/or specific cells from any changes (or can specify what changes are allowed).

All cells in a sheet are ‘locked’ by default, so once put in password protects whole sheet to the tasks selected in the list.

Instructions on how to only protect certain cells and certain types of cells (e.g. cells that contain formulas) **here**.

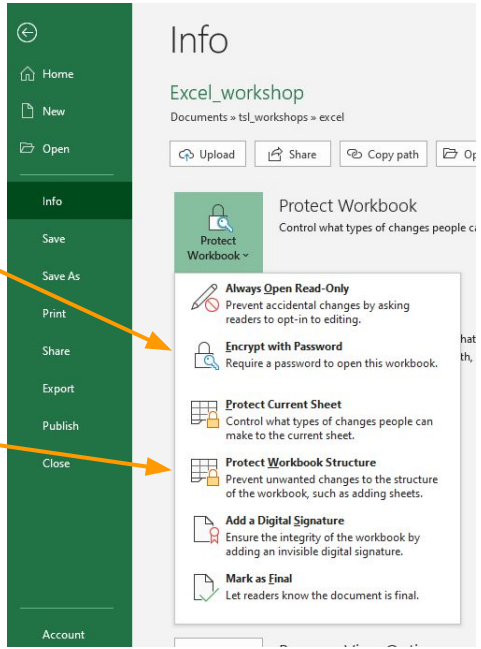


Restricted changes to workbook

'File' -> 'Info' -> 'Protect Workbook'

Can restrict workbook being setting password that is needed to be able to open the workbook.

Can password password protect workbook from 'structure' changes
i.e. insert, delete, rename, move, copy, hide or unhide worksheets.



Audit trails

Audit trail tips - more advanced

Have you ever tried to recreate a figure or graph and you're not sure how that original number was calculated? Clear audit trails and documentation can be helpful to minimise this.

Audit trail is a record of the changes that have been made to a file, and can be helpful to understand how a figure or graph has been calculated.

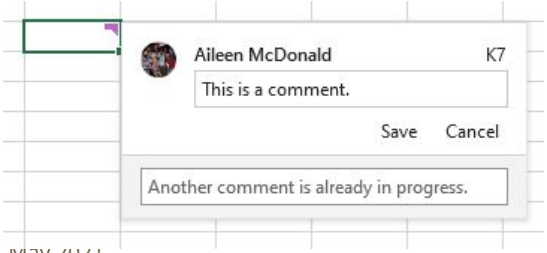
- May want to use 'front page' as instructions (especially if spreadsheet is quite complicated and/or contains a lot of data), list data sources and dates extracted/input
- Can add comments on to cells if have information to highlight specific cells or calculations.

Comments and notes

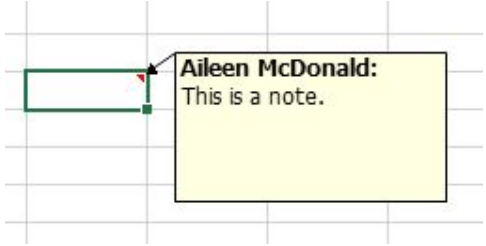
Can use comments or notes to draw attention to entries or outputs in particular cells.

- **Comments:** newer version of Excel have comments that can act as conversations between people, as can ‘reply’ the comments.
- **Notes:** can also be used to make comments about the data (and are the same as comments in older versions of Excel).

Comments



Notes



Audit trail tips - more advanced (cont.)

- Colour coding can be useful for understanding what is hard coded and what is formula, what is ‘output’ (and minimises risk of someone harding coding over a formula) e.g.:
  - Pasted data (from another source): purple
  - Author input: green
  - Formula: light yellow
  - Key output: blue
- Use formulas wherever possible over clickable shortcuts (e.g. sorting columns). Means that there is no audit trail of the steps taken during workflow and makes reproducibility/updates difficult.
- **Advanced:** any checks in Excel should be done via formulas (again so auto update). IF() statements very helpful e.g. IF(E12>5, “ok”, “CHECK”). Avoids manual checks by user every time data is updated.

File naming tips

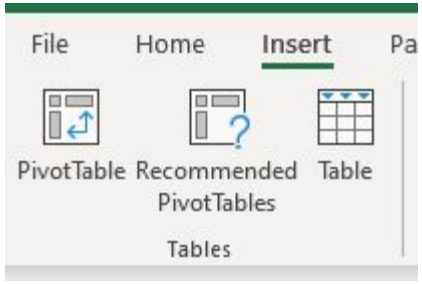
Do you find it difficult to know which is the most up to date version of a spreadsheet? Do you find it difficult to know who is working on which version or what was used in a particular meeting? Do you have files named ‘final’, ‘FINAL’ or ‘FINAL FINAL’?

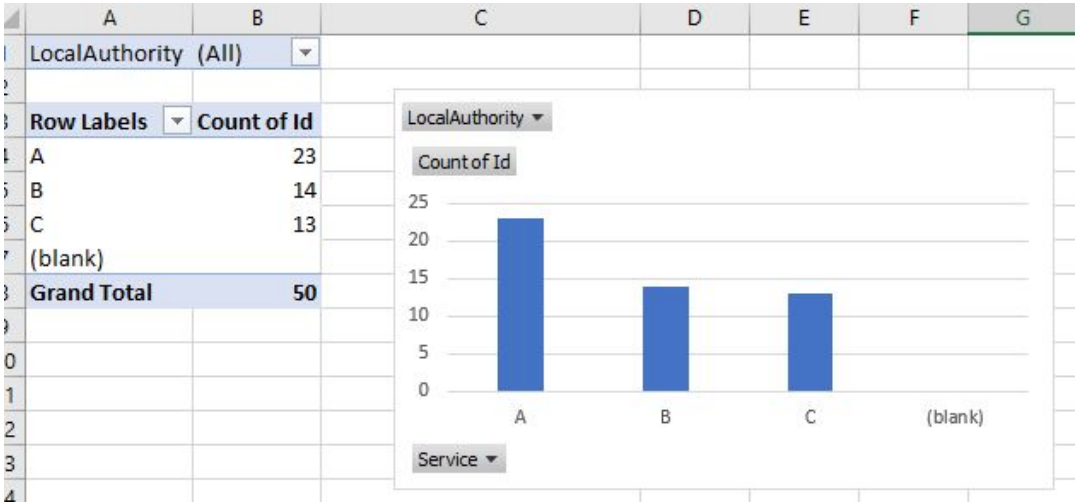
- To minimise confusion can use **file naming conventions (date, name, version)** which is regularly used in heavily audited industries such a finance.
    - *e.g. 20191018_end_of_year_financials_v0.1*
- Version numbers start at v0.1 and then work up to v1, which is the ‘final’ one (perhaps the version that gets presented at a meeting). If any changes are then made after this can start at v1.1 and so on to v2 etc.
- Means know which is the most up to date version is and what version was used on which dates.
- Any graphs or figures used in a presentation or report reference the Excel file name and location where they are from (and date the data used was extracted) in the presentation file.

Pivot tables

Pivot table

- Can summarise/aggregate easily using pivot tables.
- Can also produce graphs of the pivot table results
- If data updates can ‘Refresh’ pivot table and automatically updates





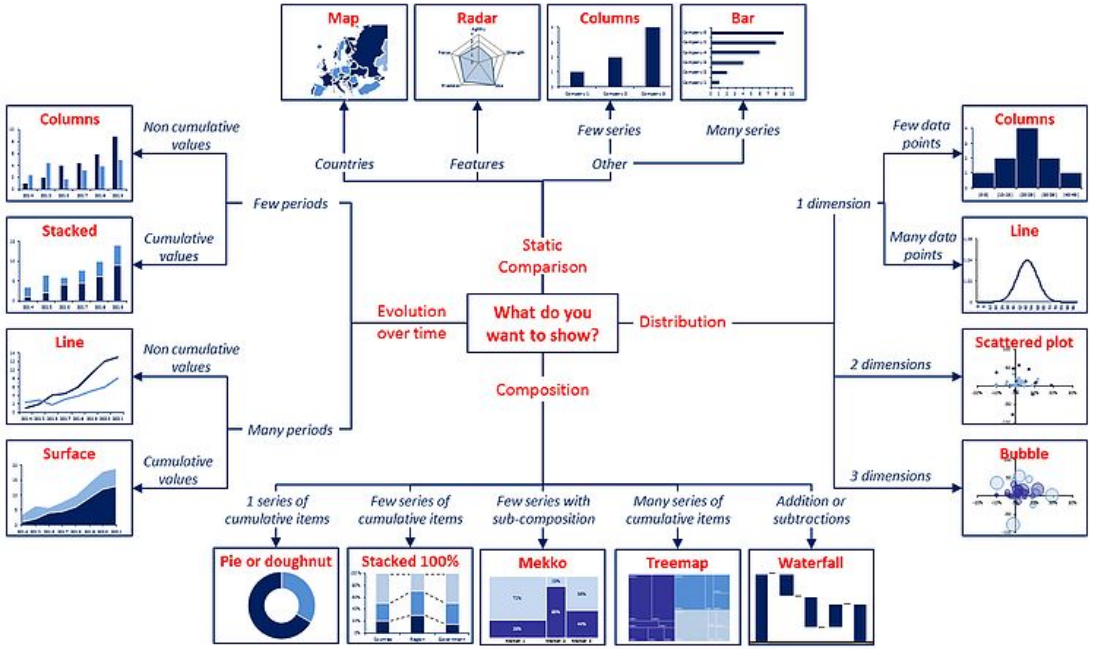
Visuals

Visuals - general

Common charts:

- Line
- Bar (stacked)
- Scatter plot
- Pie chart
- Tree map

More about different charts here.

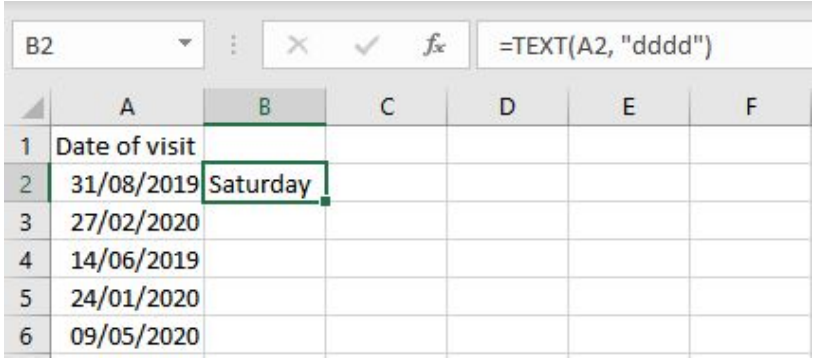


General recommendations on formatting charts here.

Which day of the week has the most visits?

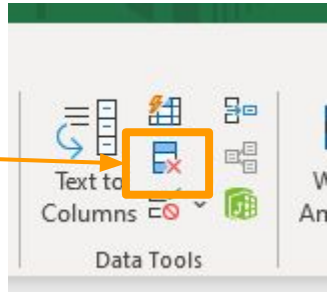
Formula - day of the week

- Google '*day of week from date in Excel*'
- Many resources indicate formula is =***TEXT****(A1, "dddd")*
- Enter this formula in top cell
- Double click bottom right corner of formula cell
    - This copies formula down to end of dates column



Count of each week day

- Get list of unique values (could manually type out days of week) or…
- Select column and copy and paste **as values**
    - Paste as values vs. paste as formula
- *'Data tab' -> 'Delete duplicates'*
- Sort order (custom sort)
- Use **COUNTIF()** function:



> **COUNTIF()** - counts the number of cells in a range, that meets a given criteria
> *COUNTIF(range, criteria)*

- Drag formula down

Visualise results

- Select area -> *'Insert chart'*
- Select chart type
- Think about good practise
  - Remove or reduce salience gridlines (unless feel necessary to read)
  - Ensure axis titles (and units)
  - Chart title

If more entries were to be added in columns A&B (and dragged formula on C) then count and graph would automatically update.

Additional chart tabs

2 additional tabs appear when insert a chart:

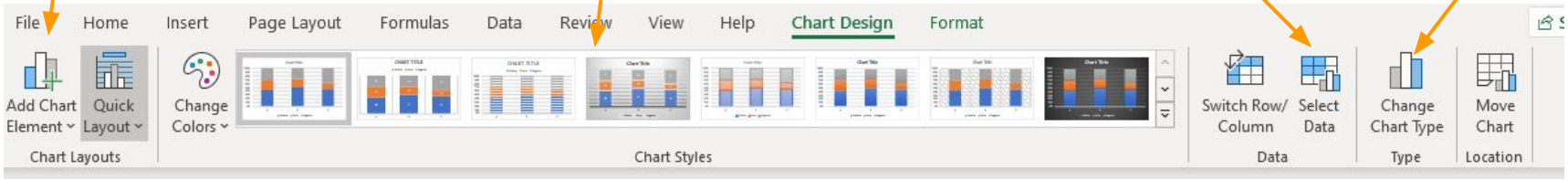
1. **Format:** border, size etc.
2. **Chart Design**: layouts, styles, type etc.

Add title, axis titles, data labels, gridlines etc.

Different default styles.

Edit what data is used in the graph

Change chart type (e.g. line, pie, bar etc.)



Open ‘excel_workshop_dashboard’ file

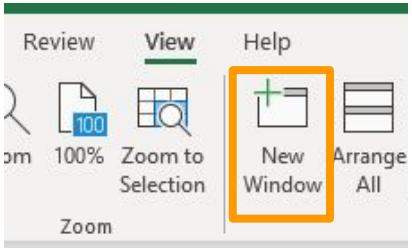
Tips for data visualisation

Data visualisation tips

- No 3D charts - they can skew perception of the visualisation.
- Keep chart and graph headers simple and to the point.
- Make sure have labels on axes - including units!
- Don’t overdo the colours - each colour should have a purpose as can be distracting otherwise.
- Keep consistent with font type and size.
- Be consistent with colours to build association (e.g. blue for service A, green service B etc.).
- Be careful with axes on graphs - usually best to start at 0 so not misleading.
- If showing numbers with decimal points, round to 1 or 2 decimal points depending on the level of detailed required.
- Every graph should have a purpose - people get saturated quickly with too much info. What is the objective of your graph and how can it create value for your organisation?
- Often bar charts can be useful to order of highest to lowest (unless some other ordering is important - such as alphabetical).

Excel tips & tricks

See different tabs within the same workbook



Keep having to switch between tabs within a spreadsheet?

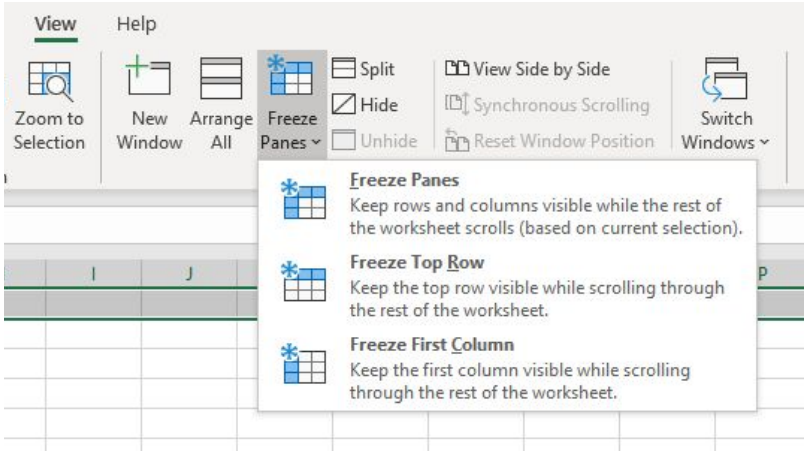
Can use the ‘New window’ functionality which opens up a 2nd version of the spreadsheet which allows you to work/see 2 tabs simultaneously.

Data quality

- To use formulas such as COUNTIF(), SUMIF() etc. require entries to have same values - different users inputs can make this difficult!
  - E.g. ‘Mon’, ‘mon’, ‘Monday’, ‘monday’, ‘day1’ etc.
- Can use ‘*Data Validation*’ on ‘Data’ tab to:
  - Restrict possible values
  - Only certain data types
  - Data in certain values (e.g. between 2 values etc.)
- Create pop up if invalid input (e.g. ID required to be 7 digits long and user inputs 6 digits)
- More info **here.**

Freeze panes

- Can use when data is very wide or long and want to keep particular rows or columns always in view (when scrolling vertically and/or horizontally
- If it it’s more than 1st row or column then select only the row or column want to freeze up to then select ‘Freeze Pane’)



Tips

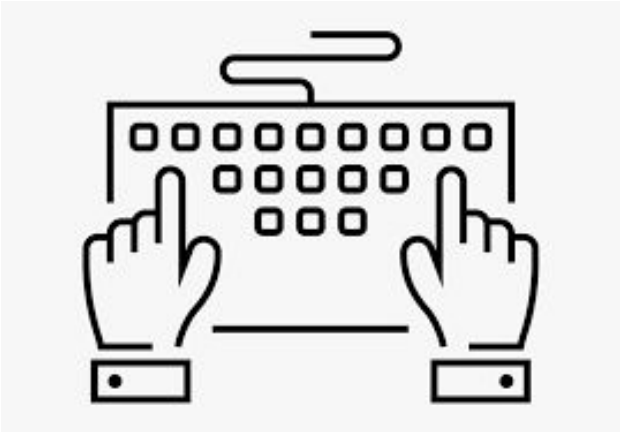
- Start data from top left, cell A1.
- Want to consider - if a new entry was to be made to my raw data would my calculations still work?
- Can be good to have raw data in one sheet any calculations in another.
- Good to have formulas based on columns, rather than areas (avoids issues if new entries added, don't have to update any formulas or charts)
    - e.g SUM(A:A) rather than SUM(A1:A5)
- If using data from another workbook, link the data (rather than copying and pasting) so is updated when other workbook does (and is easy to follow trail)
    - If can't do this then have comment where data is from and date taken.

Useful shortcuts

Ctrl+A: **highlights area**

Ctrl+shift+direction*: **select all row/column** in chosen direction from cell currently in

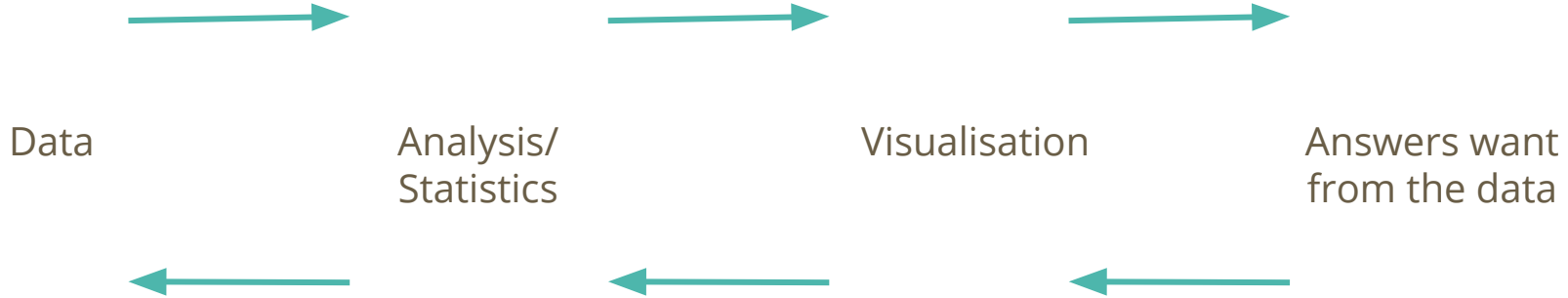
Ctrl+direction*: **jump** to cell that is at **end of the row or column** (in chosen direction from cell currently in)



*direction is arrow (up, down, left, right)

Planning for data visualisation

Workflow



- Think about **end users of the visualisation** & **how they will use the information**.
- Sketch out planned design before begin and what data will need.
    - Lots of free online tools for sketching/designing such as Excalidraw here.

Why are you tracking this on your dashboard?

Questions	Example
What is the metric?	Number of service users.
Why are you tracking this metric?	To ensure that if demand for service is greater than supply we can act to meet demand where possible.
What action will you take if there is a change to the metric?	If it goes above a certain value we will need to increase the capacity for services or recruit more volunteers to meet demand.
Do you have a **benchmark/target** to compare metric against?	Ratio of 10 service users to a volunteer.

Resources

Online learning resources

coursera

Udemy

Online courses

YouTube

Full tutorials and short quick fix videos

Microsoft

Microsoft learning resources

Data to play with...

- UK Data Service here
- Scot Gov. SIMD data here
- Create fake data at Mockaroo.com here
- Scottish Gov. Open data here
- UK Gov. Open data here

BI/Data visualisation tools

Data visualisation tools

- Power BI and Tableau are amongst the most widely used BI (Business Intelligence) tools.

Power BI

Tableau

Benefits of data visualisation tools

- Publish data reports and dashboards online - easy to share.
- Interactive (user can drill into data)
- Hold a lot more information in single view
- Can handle much bigger datasets
- Automatic updates of visuals if new data
- Connecting and visualising many data sources in one place
- Multiple collaborators on single dashboard
- Easy download of visuals for static reports...

...and don't need to be able to code to do all of above.

Thank you

Contact

Email: aileenlmcdonald@gmail.com

Twitter: https://twitter.com/mcd_ails

LinkedIn: https://uk.linkedin.com/in/aileenmcdonald

Scottish Charity Search app: here