Programming for E&BI: Excel

Christoph Walsh

Introduction

- In this slide deck we will learn some important Excel skills.
- These slides are accompanied by a set of video demonstrations.
- *Note:* this course assumes you already have some basic knowledge of Excel. So we will go through some of the more basic material very quickly.

What is Excel?

- Excel is a *spreadsheet editor*.

- Spreadsheets organize data (numbers and text) in *tabular form*.

  - These are tables of *cells* organized in *rows* and *columns*.

- In a *table*:

  - Columns typically represent a variable types (such as the date, sales, expenditures).
  - Rows typically represent items/records/observations for those variables.

- For example, an ice-cream salesman might record daily sales and the daily temperature in a spreadsheet:

Date	Sales	Temperature
10-08-2023	350	23
11-08-2023	400	26
12-08-2023	380	24

Spreadsheets

- Spreadsheets and Excel are useful for many things:
  - ▶ Organizing and storing many types of data, such as financial information.
  - ▶ Performing statistical analyses on data.
  - ▶ Creating plots and charts of data.

Ancient Spreadsheets

- Spreadsheets are so useful that people have stored data in tabular format for literally thousands of years.

Figure 1: 3700-year-old Babylonian Clay Tablet.

Less-Ancient Spreadsheets

- Until recently, companies recorded sales and costs in accounting ledgers (and some still do).
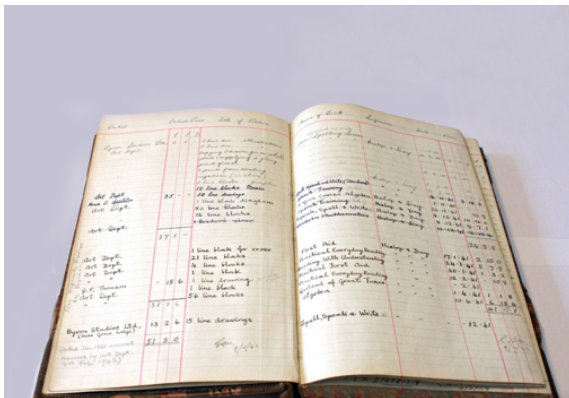

Figure 2: Paper Accounting Ledger.

Digital Spreadsheets

- Spreadsheet software appeared on personal computers as soon they became widespread.

Figure 3: VisiCalc in 1979.

- MS Excel, LibreOffice Calc and Google Sheets are modern versions of these.

Useful Excel Features

- For the rest of this slide deck we will go over a number of useful Excel features.
- Many Excel features we will talk about will also work in other spreadsheet software, like LibreOffice Calc and Google Sheets.

Cell Formats

- When you input data into Excel it tries to guess the intended format.
  - For example, if it is a date, a number, a monetary amount, or just text.
  - If we type in 2/12/23, it guesses that it's a date and formats it to 02/12/2023.
- If we want to change this, we can use the "Format Cells" option.
  - Select the cell(s) we want to change, right-click, and choose "Format Cells…"

Exercises:

1. Type 123.45 in a cell and convert it to a currency (in euros).
2. Type in today's date in the format dd/mm/yyyy and format it to US-style: mm/dd/yyyy.

Cells with Long Streams of Text

- Sometimes cells contain very long pieces of text that run over other columns.
- If we want to change this we have a couple of options.
    1. Use the "Wrap Text" option under Home -> Alignment.
    2. Insert line breaks using Alt + Enter at the points we want the text to run onto a new line.

Exercises:

1. Go to https://loremipsum.io/ and copy the famous (long) Latin filler text (*Lorem ipsum dolor sit amet…*). Add it to a cell and use the wrap text function.
2. Do the same as (1) but manually insert some line breaks instead.

Excel as a Calculator

- We can use Excel as a basic calculator.
- If we wanted to add $2 + 3$, we can go to an empty cell and type =2+3 and Enter. The cell would then turn into 5.
  - In the *Formula bar*, we can still see the original formula =2+3.
- To subtract, we use -. For example, =3-2.
- To multiply, we use *. For example, =2*3.
- To divide, we use /. For example, =6/2.
- To take powers, we use ^. For example, =3^2.
- We can also perform multiple operations together using parentheses. For example, if we want to calculate:

$$\frac{12 + 4}{2 \times 4} = \frac{16}{8} = 2$$

  We would do =(12+4)/(2*4).
- **Exercise:** Use Excel to calculate $\frac{2 \times 4}{2^2}$.

Cell Referencing

- In Excel, rows are referenced by numbers and columns by letters.
- The cell in the 2nd column and 3rd row can be referenced with B3.
- This referencing will be very useful for calculations, and for using functions.
- We can also refer to multiple cells:
  - To refer to rows 3-7 in the 4th column, we write D3:D7.
  - To refer to rows 1-5 in columns 1-2, we write A1:B5.
- We can also refer to cells on other Excel *Sheets*, which are like tabs in the same document.
  - To refer to cell A1 in the sheet called Sheet2, we write Sheet2!A1.

Calculations in Excel using Cell Values

- We can use cell references to make calculations.

	A	**B**
1	30	40
2	20	50

- If we want to add A1 and A2 together $(30 + 20)$, we go to an empty cell (for example, A3), and type =A1+A2 and Enter.
  - We could also type = and then click on the cell A1, type +, and then click on the cell A2, and then press Enter.

Functions

- We can perform calculations on a cell or group of cells using in-built *functions*.
- If we want to add numbers in a column to together, we can use the SUM function.
- For example, let's say we want to add the numbers in column A like in the previous slide:

	A	B
1	30	40
2	20	50

- We can go to a new cell, type =SUM(A1:A2) and Enter.
- Here SUM is the *name* of the function. In Excel these are always with capital letters.
- The term in the parentheses is called the *argument* of the function. In this case it's the cells we want to sum.

Some Common Functions to Summarize Values

In what follows below, *cells* refers to a range of cells, like `A1:A2` in the previous slide.

- =AVERAGE(*cells*) – gets the average of *cells*.
- =MEDIAN(*cells*) – gets the median of *cells*.
  - ▶ The median is the value "in the middle" when we sort the numbers.
  - ▶ If there is an even number of numbers it is the average of the two numbers in the middle.
  - ▶ For example, the median of [4, 2, 1, 3] is 2.5, because the average of the two numbers in the middle (2 and 3) is 2.5.
  - ▶ **Exercise:** Populate cells with the numbers [4, 2, 1, 3] and calculate the median.
- =MAX(*cells*) – gets the largest number in *cells*.
- =MIN(*cells*) – gets the smallest number in *cells*.

Adding/Deleting Rows/Columns

- Suppose we have the following table, but we want to add a new row between rows 1 and 2.

	A	B
1	30	40
2	20	50

- We can add 1 new row above row 2 by right-clicking on the "**2**" row index and clicking the `Insert` option.
  - ▶ What was row 2 then becomes row 3, with an empty row in its place.
- We can also add multiple rows at the same time. For example, to add 4 new rows between rows 1 and 2, we do the following steps:
  - ▶ First select rows 2-5. Click on the "**2**" index for row 2, hold `Shift`, then click on the "**5**" index for row 5.
  - ▶ Then right click on one of the selected row indices and press `Insert`. 4 new rows should appear.

Adding/Deleting Rows/Columns

- Adding new columns is very similar. We just right-click on the column index instead. `Insert` will then add a new column to the left.
- **Exercise:** Add 2 columns between `A` and `B` in the previous table.
- To delete a row or column, we select the row/column, right click, and then press the `Delete` option.

Freezing Panes

- Often we have Excel files with many values that we can't see on our screen all at once.
- If you scroll down to examine values, you can no longer see the first row which often has the names of the variables in the columns.
- We can *freeze* the top row so that we always see the top row when we scroll down.
- To do this, we go to "View" → "Freeze Panes" → "Freeze Top Row".
- We can also freeze any selection of cells by using the "Freeze Panes" option.

Sorting Ascending or Descending

Suppose we want to sort the data by the values in column A:

	A	B
1	30	40
2	20	50
3	25	45

- We click on the column (by clicking A).
- We go to the "Data" tab and press the $\frac{A}{Z} \downarrow$ button.
- We tick "Expand the selection". We then get:

	A	B
1	20	50
2	25	45
3	30	40

Sorting Ascending or Descending

- The "Expand the selection" option keeps the cells in B together with their corresponding cell in A, so the cells in B change position as well.
- If we don't want to do that, we tick "Continue with the current selection".
- **Exercise:** Sort the data by the values in column B but *descending* (largest to smallest). Also expand the selection.

Transposing

- Sometimes we have data in a row that we want in a column, or vice versa.
- For example, suppose we have:

	A	B
1	30	40
2	20	50
3	25	45

- But we want to convert it to:

	A	B	C
1	30	20	25
2	40	50	45

Transposing

- We need to select and copy the values we want to transpose.
- We then go to an empty part of the spreadsheet.
- We right click, "Paste Special", and then tick "Transpose".

Filtering

Suppose we have the following data on the sales of 3 different products in 3 different years:

	A	**B**	**C**
1	product	year	sales
2	1	2020	1200
3	1	2021	2400
4	1	2022	1400
5	2	2020	8900
6	2	2021	9700
7	2	2022	7400
8	3	2020	3300
9	3	2021	4400
10	3	2022	2100

If we want to see the sales for all 3 products in 2022 only, we can use the *filter* tool.

Filtering

- Go to the "Data" tab and press the "Filter" toolbar button.
- Little down arrows appear on the column names product, year, and sales.
- To see the rows only for 2022, we click on the down arrow next to year.
- We deselect all rows by unticking "Select all" and then selecting "2022".
- We then only see the column names and the values for 2022.
- To go back to normal, you can click "Clear" next to the Filter toolbar button.

Filtering

- You can also try more complex filter operations. For example, show all years bigger or equal to 2021 (both 2021 and 2022).
- We go to the small down arrow next to `Year`, click "Number Filters" $\rightarrow$ "Greater Than Or Equal To".
- We then put 2021 into the box (greater than or equal to 2021).
- We then see the data only for 2021-2022.
- **Exercise:** Filter to see the sales for each year for only product 2.

Autofilling: Sequences

- Often we want to create a sequence in a column like 1, 2, 3, 4, etc
- If we have to do that for many values, it would be a lot of work.
- For common things, Excel can *autofill*.
- We do that as follows:
    1. Type 1, 2, 3 in three neighboring vertical cells.
    2. Select those cells.
    3. Click on the bottom-right of the selection and drag it down for as long as you want the sequence to go.
- This also works for other sequences. For example, 10, 20, 30.
- It also works for repeating the same number (like 1, 1, 1, 1).

Autofilling: Functions

- Often we want to get the sum of values for several columns. For example, for the 3 columns in:

	A	B	C
1	30	20	25
2	40	70	15

- If we have to type =SUM(A1:A2) many times (replacing A with B, C, etc.) it would be a lot of work.
- What we can do is (i) go to cell A3 and type the function =SUM(A1:A2), (ii) copy that cell, (iii) select cells B3 to C3, (iv) Paste (using Ctrl+V).
- This automatically changes the A to B and C in the formulas, giving the sums of the other columns (90 and 40).

Absolute Cell Referencing

- In the previous example, it could happen that we want to sum the values in column `A` using =SUM(A1:A2) and then paste **those** values on B3 and C3. That is, we want 70 repeated 3 times (instead of 70, 90 and 40).
- We can avoid the cell coordinates changing when we paste a function to different places using *absolute cell referencing*.
- We use a dollar symbol in front of the cell coordinates to do this.
- If we write =SUM($A$1:$A$2) in A3 and copy and paste that on B3 and C3, we will see 70 in each cell (instead of 70, 90 and 40).
- We can write $A1 if we want the column part to stay fixed but allow the row part to vary.
- We can write A$1 if we want the column part to vary but keep the row part fixed.

Pivot Tables

Suppose you are a greengrocer and you have data that look like this:

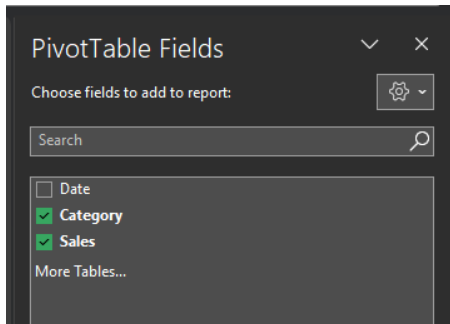|	A	B	C
1	Date	Category	Sales
2	01/10/2023	Fruit	3400
3	01/10/2023	Vegetables	2800
4	01/10/2023	Other	500
5	02/10/2023	Fruit	3600
6	02/10/2023	Vegetables	3000
7	02/10/2023	Other	600
8	03/10/2023	Fruit	3200
9	03/10/2023	Vegetables	1900
10	03/10/2023	Other	450
11	04/10/2023	Fruit	3500
12	04/10/2023	Vegetables	2200
13	04/10/2023	Other	550

Suppose you want to summarize the data on this table. For example:

- The total sales per product category
- The average daily sales per product category
- The total sales per day

We can do this using pivot tables.

Pivot Tables

- Select the table
- Go to Insert → PivotTable
- Choose either new or existing worksheet depending on where you want the new table.

PivotTable Fields ∨ ✕

Choose fields to add to report: ⚙ ∨

Search 🔍

- ☐ Date
- ☑ **Category**
- ☑ **Sales**
- More Tables...

- To get the total sales per category, tick the "Catorgy" and "Sales" columns.
- By default we will already get the desired table.
- Below we see the "Rows" are "Category" and the "Values" are "Sum of Sales"

Row Labels	Sum of Sales
Fruit	13700
Other	2100
Vegetables	9900
Grand Total	**25700**

Pivot Tables

- To instead get the average daily sales per product category:
  - ▶ We click on "Sum of Sales" under $\sum$ Values.
  - ▶ "Value Field Settings"
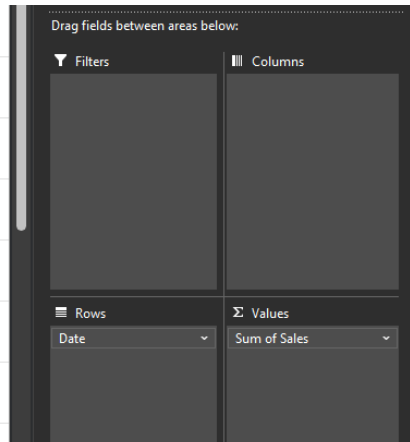 - ▶ Click in "Average".

The result should be:

Row Labels	Average of Sales
Fruit	3425
Other	525
Vegetables	2475
Grand Total	**2141.666667**

☰ Rows		Σ Values	
Category	∨	Average of Sales	∨
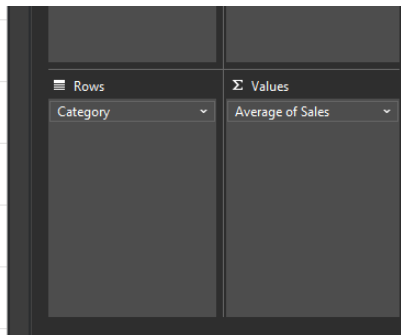
Pivot Tables

- To get the total sales per day:
  - We change the columns to be "Date" and "Sales".
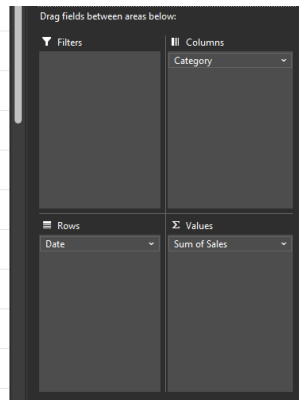 - We change the "Value Field Settings" under $\sum$ Values back to "Sum".

Pivot Tables

- By ticking all of "Date", "Category" and "Sales" we can obtain a more detailed table.
- By dragging "Category" to "Columns" and setting "Rows" to "Date", we can get the sum of sales by date and category simultaneously:

Sum of Sales	Column Labels			
Row Labels	Fruit	Other	Vegetables	Grand Total
01/10/2023	3400	500	2800	6700
02/10/2023	3600	600	3000	7200
03/10/2023	3200	450	1900	5550
04/10/2023	3500	550	2200	6250
Grand Total	13700	2100	9900	25700

Drag fields between areas below:

▼ Filters	⊞ Columns
	Category ▾

≡ Rows	Σ Values
Date ▾	Sum of Sales ▾

VLOOKUP

Suppose you have a table with employee IDs and names, the like the table on the left:

	A	B	C	D	E	F	G	H	I
1	**Raw Data Extract**					**Pay Report**			
2									
3	**Employee ID**	**Last Name**	**First Name**			**Employee ID**	**Pay**	**First Name**	**Last Name**
4	110608	Doe	John			990678	$84,289		
5	253072	Cline	Andy			830385	$137,670		
6	352711	Smith	John			795574	$190,024		
7	391006	Pan	Peter			580622	$122,604		
8	392128	Favre	Bret			549457	$111,709		
9	549457	Elway	John			392128	$85,931		
10	580622	Manning	Eli			391006	$168,114		
11	602693	Vick	Micheal			352711	$89,627		
12	611810	Woods	Tiger			253072	$149,946		
13	612235	Jordan	Micheal			612235	$145,893		
14	795574	Stark	Tony			611810	$64,757		
15	830385	Williams	Prince			602693	$71,478		
16	990678	Pitt	Brad			110608	$131,505		
17									

And you want to use this information to fill in the gaps in the table on the right.

VLOOKUP

- We can do this using the VLOOKUP function.
- More generally, we use VLOOKUP function to find a particular row of a table using a search query (with exact or approximate matching), and to return the value of a different column of that row.
- The VLOOKUP function takes 4 arguments:
  1. The value you want to look up.
  2. The range of values containing the table.
  3. The column number of the table we want to return for the match.
  4. Whether we want an approximate or exact match (TRUE for an approximate match, FALSE for an exact match).

VLOOKUP

In our example:

1. The value we want to look up: We use the employee ID, so F4 for the first employee.

2. The range of values containing the table: This is \$A\$4:\$C\$16, where we use absolute cell referencing as we want to copy the function for the remaining employees, but the table stays fixed.

3. The column number of the table we want to return for the match: This is 3 for the "First Name" and 2 for the "Last Name"

4. Whether we want an approximate or exact match: For employee IDs we should use exact matching, so FALSE.

So: =VLOOKUP(F4; \$A\$4:\$C\$16; 3; FALSE) for "First Name"