EXCEL 1 — DATA ANALYSIS BASICS

Computer Practicum EN - WIS - LES

Lecturer: Niki Hrovatin

What Is Microsoft Excel?

Excel is a spreadsheet application for organizing, processing, analyzing, and visualizing data.

- A **worksheet** contains cells arranged in **rows and columns**.
- **Formulas and functions** calculate results from data.
- **Charts** visualize data; **PivotTables** summarize and compare groups.

In this lab: import data → clean → analyze → visualize.

Why Is Tabular Data Useful?

One row = one observation. One column = one variable. One cell = one value.

Day	Temperature (°C)	RH (%)	Fire
01/06	29	57	no
02/06	29	61	no
06/06	31	67	yes

- **Consistent structure:** easy to filter and sort.
- **Calculations:** compare values and compute summaries.
- **Compatibility:** tables can be shared across many programs.

Common Data File Formats

Format	Contents	Typical use
CSV (`.csv`)	Plain-text values separated by commas	Sharing simple tables
TSV (`.tsv`)	Plain-text values separated by tabs	Exporting tabular data
Excel (`.xlsx`)	Worksheets, formulas, charts, formatting	Working in Excel
JSON (`.json`)	Structured data with keys and values	Web services and APIs

Remember: CSV and TSV are text files, not Excel workbooks.

CSV — Comma-Separated Values

A **CSV** is a plain-text file representing a table:

```
day,month,Temperature,RH,Classes
01,06,29,57,notfire
02,06,29,61,notfire
06,06,31,67,fire
```

- **First row:** column names (header).
- **Each next row:** one observation; commas separate fields.
- **No built-in formulas, formatting, or charts.**

Excel: Data → From Text/CSV.

Our Dataset — Algerian Forest Fires

What: daily weather observations and fire-danger indicators.

- **Where:** Bejaia (northeast) and Sidi Bel-Abbes (northwest), Algeria.
- **When:** June–September **2012**.
- **Size:** **244** records in the original dataset; **243** in the cleaned CSV.
- **Outcome:** `Classes` = `fire` or `notfire`.

Questions: Are fire days hotter and drier? Which months and regions have a higher proportion of fire days?

Dataset Columns: What Do They Mean?

Date and classification

- `day`, `month`, `year` — observation date.
- `Region` — Bejaia or Sidi Bel-Abbes.
- `Classes` — `fire` / `notfire` label.

Weather measurements

- `Temperature` ↑ — air temperature (°C).
- `RH` ↓ — relative humidity (%).
- `Ws` ↑ — wind speed (km/h).
- `Rain` ↓ — daily rainfall (mm).

Fire Weather Index (FWI) indicators

- `FFMC` ↑ — dryness of fine surface fuels.
- `DMC` ↑ — dryness of medium-depth organic fuels.
- `DC` ↑ — deeper, long-term drought.
- `ISI` ↑ — potential rate of fire spread.
- `BUI` ↑ — potential available fuel.
- `FWI` ↑ — overall potential fire intensity.

↑ Higher value generally increases fire danger.
↓ Higher value generally reduces fire danger.

Cells, Rows, and Columns

A worksheet is a grid: **columns have letters** and **rows have numbers**.

	A	B	C
1	Day	Temp.	Rain
2	01/06	29	0
3	02/06	31	2

- **Cell B2** = column B, row 2.
- **Select a row / column:** click its **number / letter**.
- **Select a cell:** click it, or enter its address in the **Name Box**.
- More on selection

Selecting Cells and Ranges

A **range** is a group of cells, written using `:`.

Reference	Selection
`B2`	One cell
`A2:C8`	Rectangular range
`B:B`	Entire column B
`2:2`	Entire row 2
`A2:A20`	Cells A2 through A20

How: drag over cells, or type a reference into the **Name Box** (left of the formula bar).

- More on Relative references

Writing Your First Excel Formula

A formula **always starts with =** and can reference other cells.

```
=B2+C2
=(B2+C2)/2
```

- Write the formula in the **result cell**, then press **Enter**.
- Excel displays the **result** in the cell and the **formula** in the formula bar.
- When B2 or C2 changes, the result is **recalculated**.

Arithmetic Operators in Formulas

Operator	Operation	Example
+	Add	=A2+B2
-	Subtract	=A2-B2
*	Multiply	=A2*B2
/	Divide	=A2/B2
^	Power	=A2^2

Use parentheses to control the order:

```
=(A2+B2)/2
```

More on:

- Excell formulas
- Arithmetic operators

Relative vs. Absolute References

When a formula is copied, **relative references move**; **absolute references stay fixed**.

In C2	Copied down to C3
`=A2*B2`	`=A3*B3`
`=A2*$B$1`	`=A3*$B$1`

- A2 — **relative:** column and row can change.
- B1 — **absolute:** always points to cell B1.

Use $ when a constant (e.g., a conversion factor) is stored in a fixed cell.

- More on Absolute references

Four Types of Cell References

Reference	What stays fixed?
A1	Nothing — relative
A1	Column A and row 1
$A1	Column A only
A$1	Row 1 only

Copying Formulas with AutoFill

AutoFill copies a formula to adjacent cells.

Cell	Formula
C2	=A2*B1
C3	=A3*B1
C4	=A4*B1



1. Write the formula in C2.
2. Grab the **fill handle** (small square at the cell's lower-right corner).
3. **Drag downward** — relative references update automatically.

Tip: double-click the fill handle to fill alongside continuous neighboring data.

More on Autofil

Basic Excel Functions

A **function** performs a calculation on arguments (values or ranges).

Function	Purpose
`=SUM(B2:B20)`	Add values
`=AVERAGE(B2:B20)`	Mean
`=MIN(B2:B20)` / `=MAX(B2:B20)`	Minimum / maximum
`=COUNT(B2:B20)`	Count numbers
`=COUNTA(B2:B20)`	Count nonempty cells

`B2:B20` is a range — from B2 through B20.

More Excell functions

Conditional Functions: COUNTIF and AVERAGEIF

Analyze observations that satisfy a condition.

```
=COUNTIF(C2:C20;"fire")
=AVERAGEIF(C2:C20;"fire";B2:B20)
```

- **COUNTIF:** number of cells in column C equal to `fire`.
- **AVERAGEIF:** average of B2:B20 **only where** C2:C20 is `fire`.

Important: depending on Excel's regional settings, replace argument commas with semicolons (`;`).

Logical Comparisons and IF

Comparison operators: = equals, <> not equals, > greater, < less, >= and <=.

```
=IF(B2>30;"Hot";"Normal")
```

- If `B2 > 30`, the result is **Hot**.
- Otherwise, the result is **Normal**.

General syntax:

- `=IF(condition; value_if_true; value_if_false)`

Excel Tables: Referencing Columns by Name

An **Excel Table** lets formulas use **column headers** instead of cell coordinates.

1. If needed: select the dataset → **Insert → Table** → check **My table has headers**.
2. Select the Table → **Table Design → Table Name** → rename it `Fires`.

Reference	Selects
`Fires[Temperature]`	All values in the **Temperature** column
`Fires[#Headers]`	The table's **header row**
`[@FWI]`	The **FWI** value in the current row (inside the Table)

Syntax: `TableName[ColumnName]` — automatically includes newly added rows.

Using Table Columns in Formulas

Example: our forest-fire dataset, imported as a Table named `Fires`.

```
=AVERAGE(Fires[Temperature])
=COUNTIF(Fires[Classes];"fire")
=AVERAGEIF(Fires[Classes];$A2;Fires[Rain])
```

- **AVERAGE:** mean temperature across all observations.
- **COUNTIF:** number of days classified as `fire`.
- **AVERAGEIF:** mean rainfall for the class written in cell `$A2` (`fire` or `notfire`).

Advantage: readable formulas that automatically include new Table rows.

Common Formula Mistakes

Problem	What it usually means
`#DIV/0!`	Division by zero
`#REF!`	Invalid cell reference
`#VALUE!`	Unexpected data type
`#####`	Column is too narrow (or invalid date/time)

- **Percentages:** `0.25` can be *formatted* as `25%`.
- **Blank vs. zero:** AVERAGE ignores empty cells but includes zeros.
- **Text numbers:** check data types after importing a CSV.

Charts - From Data to Insights

Charts represent data visually to make comparisons, changes, and proportions easier to understand.

What do you want to know?	Choose a...
How is a whole divided?	Pie chart
How does a value change over time?	Line chart
Which category is higher or lower?	Bar / column chart

Choose the chart based on the question and the type of data.

Pie Chart - Parts of a Whole

Use when: a few categories make up **one whole**.

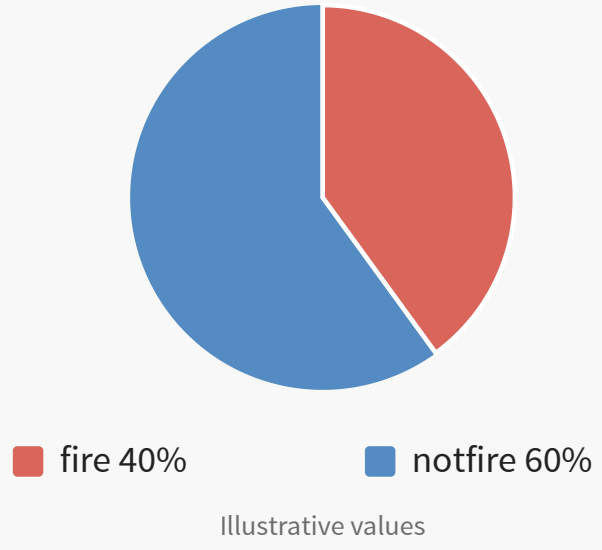
Data: category + count or percentage.

Shows: each category's **share of the total**.

Our dataset: percentage of `fire` vs `notfire` days.

Remember: all slices add up to **100%**.

More on Pie charts



Line Chart - Changes Over Time

Use when: measurements follow an **ordered sequence**, usually time.

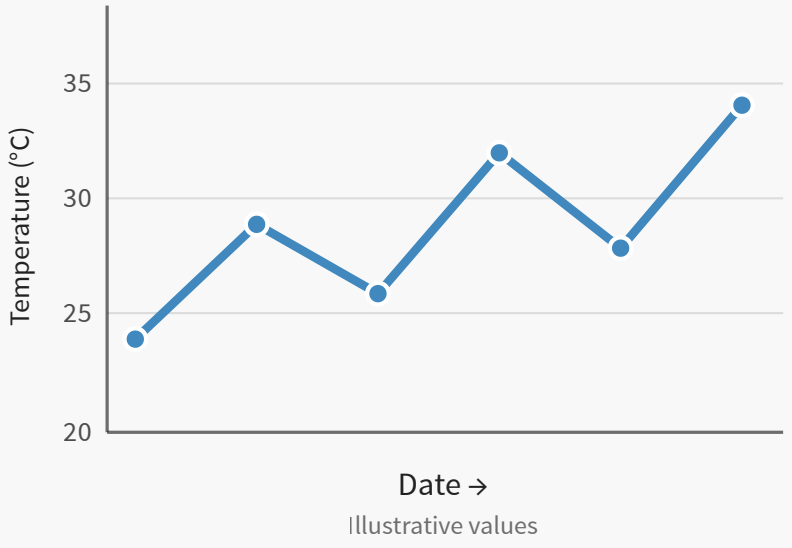
Data: date/time + numerical value.

Shows: trends, peaks, dips, and fluctuations.

Our dataset: daily temperature (in °C) across the observation period.

Read it: horizontal axis = time; vertical axis = value.

More on line charts



Bar / Column Chart - Compare Categories

Use when: comparing **different groups or categories**.

Data: category + numerical value.

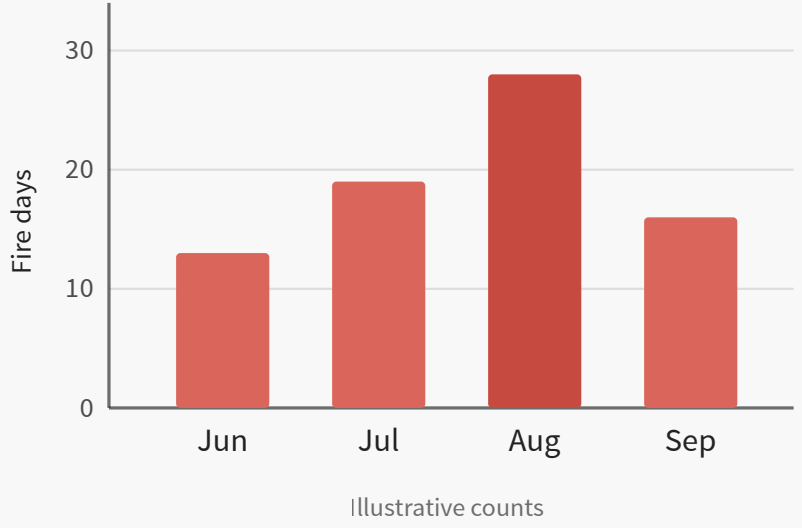
Shows: differences in **counts, averages, or percentages**.

Our dataset: number of fire days per month.

In Excel: *bar* = horizontal; *column* = vertical.

More on:

- Column Charts
- Bar Charts



What Is a PivotTable?

A **PivotTable** turns detailed records into a **summary by category** — no formulas required.

Original observations

Month	Classes
June	fire
June	notfire
July	fire
July	fire

→

PivotTable: count by month

Month	fire	notfire
June	1	1
July	2	0

Key idea: Change the grouping or calculation without altering the original data.

More on Pivot Tables

PivotTable — Four Field Areas

ROWS ↓
Categories listed vertically.
Example: `Month`

COLUMNS →
Categories displayed side by side.
Example: `Classes`

VALUES ∑
Summarize data: **Count, Sum, Average...**
Example: Count of `Classes`

FILTERS ▼
Limit the records included in the view.
Example: `Region`

Create: Select a table cell → **Insert → PivotTable** → choose a location → drag fields into the four areas.

Which months have the most fire days?

Use a pivot table to create this representation:

Month	fire	notfire	Grand Total
giu	10,29%	14,40%	24,69%
lug	15,64%	9,47%	25,10%
ago	20,99%	4,53%	25,51%
set	9,47%	15,23%	24,69%
Grand Total	**56,38%**	**43,62%**	**100,00%**

Excell Tutorial and Exercises

Good tutorial about excell:
https://www.w3schools.com/excel/index.php

Set of exerciese to learn the basics:
https://www.w3schools.com/excel/exercise.php