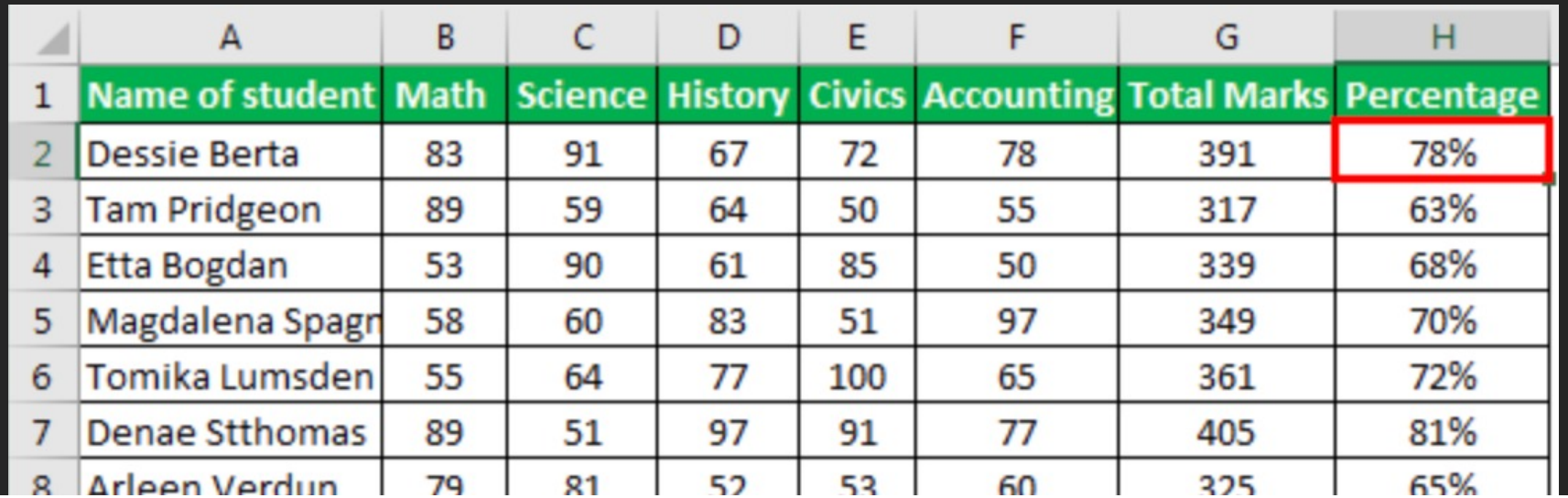
Tabular Data & CSVs

How Data in Data Science Are Commonly Organized and Packaged

Think-Share Activity

- Think of a typical spreadsheet.
- What are the various parts of a spreadsheet?
- What purpose does each part serve?



	A	B	C	D	E	F	G	H
1	Name of student	Math	Science	History	Civics	Accounting	Total Marks	Percentage
2	Dessie Berta	83	91	67	72	78	391	78%
3	Tam Pridgeon	89	59	64	50	55	317	63%
4	Etta Bogdan	53	90	61	85	50	339	68%
5	Magdalena Spagn	58	60	83	51	97	349	70%
6	Tomika Lumsden	55	64	77	100	65	361	72%
7	Denae Stthomas	89	51	97	91	77	405	81%
8	Arleen Verdun	79	81	52	53	60	325	65%

Tabular Data

- A spreadsheet is an example of **tabular data**.
- Tabular data is the **most common** type of data in **data science**.
- Tabular data is structured as a **table** with **rows and columns**.
- Each row represents an **observation**.
- Each column represents a different **variable**.
- The first row often contains **column headings**.

	A	B	C	D	E	F	G	H
1	Name of student	Math	Science	History	Civics	Accounting	Total Marks	Percentage
2	Dessie Berta	83	91	67	72	78	391	78%
3	Tam Pridgeon	89	59	64	50	55	317	63%
4	Etta Bogdan	53	90	61	85	50	339	68%

Q&A Activity

What **application** do we commonly use to create tabular data, and what kind of **file** is the tabular data saved in?

Q&A Activity

What **application** do we commonly use to create tabular data, and what kind of **file** is the tabular data saved in?

Microsoft Excel is probably the most commonly used application.

Excel spreadsheets are commonly saved as XLSX Excel Spreadsheet files.

CSV Files

- Surprise! We often **don't use XLSX** files for our tabular data in data science.
  - XLSX files are proprietary to Microsoft and contain lots of formatting information which is designed to be read by Microsoft Excel.
- Instead we use **CSV files**.
  - CSV stands for **Comma-Separated Values**.
  - CSV files are **plain text** files with the **.csv** file extension.
  - Each **line** of the file is a row.
  - Each column value in a row is **separated by a comma** character.

```
Height,Age,Gender
151.765,63.0,male
139.7,63.0,female
```

- CSV files are convenient for us to read in and process using **computational notebooks**.

How to Make a CSV File in JupyterLab

- Open the Launcher.
- Launch the Text File app.
- Edit the text file with rows of comma-separated values.
  - No spaces around commas.
  - Text values with embedded spaces must be surrounded by quotes.
  - Make sure that you have the right number of values in each row.
- Rename the file, changing the .txt file extension to .csv.
- Right click on CSV file in file explorer to either open with CSVTable app (for viewing only) or with Editor app.

Demo Video: https://youtu.be/NVr8Wf0GS80

Summary

- Tabular data
- Variables and observations
- CSV Files
- How to create CSV files in JupyterLab