Excelling with Spreadsheets

Storing data in tabular form is a common practice for managing information.
Explore essential tips for organizing spreadsheets to ensure your data is easily accessible and usable by both people and machines.

Some Important Rules

date	time_pst	tideheight_m	location
2024-08-26	06:00	1.4	santa monica
2024-08-26	18:00	1.8	laguna
2024-08-25	NA	2.2	malibu
2024-08-25	12:00	1.3	santa monica
2024-07-26	9:00	2.0	laguna
2024-07-26	06:00	1.8	malibu

- Whenever applicable, follow standards (e.g., ISO date format)
- Indicate in the documentation how missing values are represented
- NA, NULL, or blank are the best options
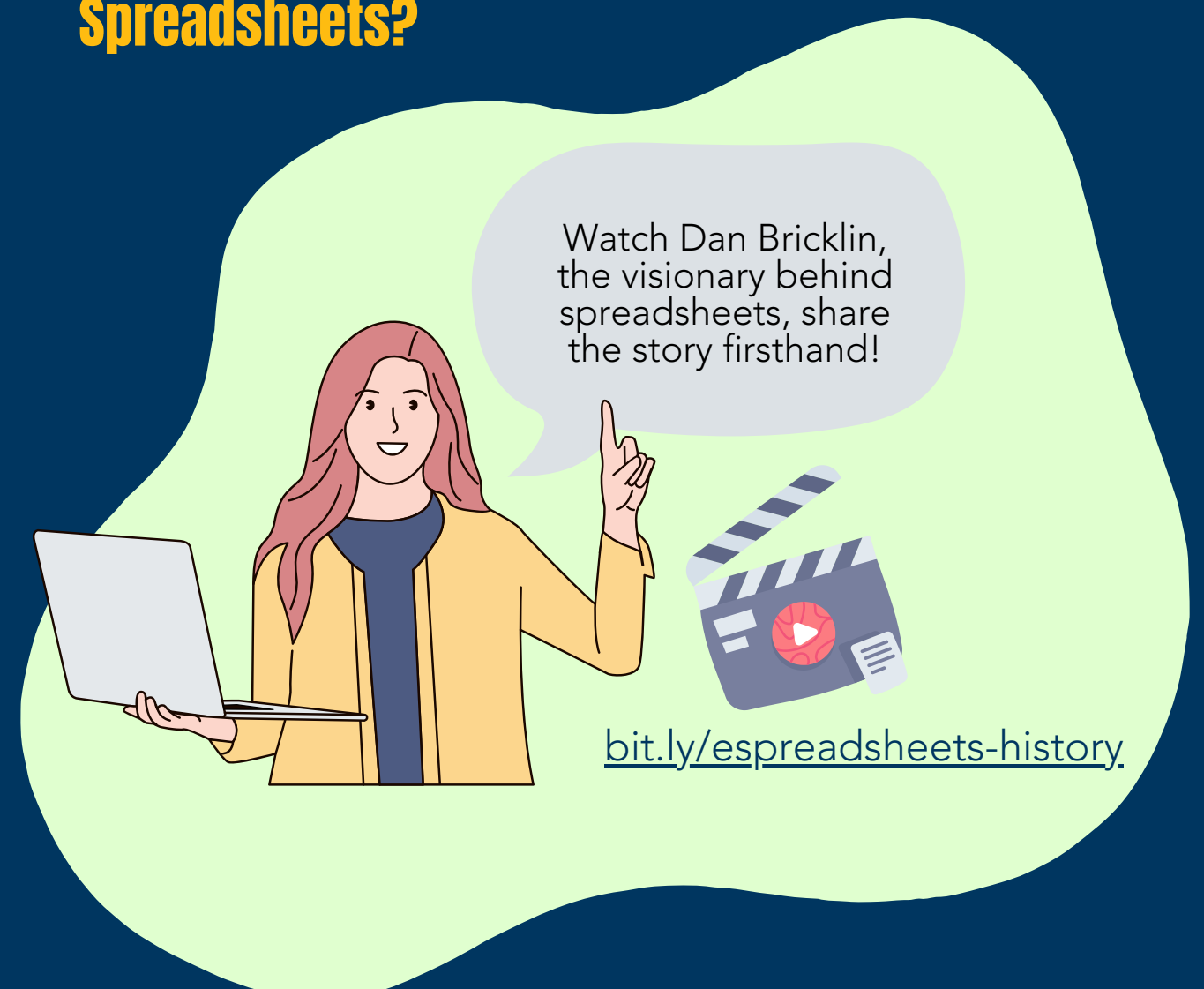- Each cell should contain one piece of information only; keep notes in a separate file

- Use the first row for concise column headers that clearly and uniquely describe the data
- Include units in column headers where relevant to provide data context
- Avoid spaces in column headers, as they cause data processing issues with some programming languages (e.g., R)
- Each row should represent a single record, observation, or case in the dataset
- Ensure consistent data entry and apply quality checks and validation rules to avoid errors

Other Things to "Keep Tabs On"

Safeguard a Copy of the Original Data
Spreadsheet programs are not designed to handle versions effectively. Ensure to create a copy before making any edits to the raw data.

Keep One Table per Sheet
If multiple related spreadsheets must be included in the same file, keep one table per tab, follow a naming convention, and describe their relationships in the data documentation.

Avoid Special Formatting
Refrain from merging cells and avoid relying on colors or text formatting to convey meaning. They can lead to misinterpretation and disrupt automated processing.

Do Not Include Plots
Keep spreadsheets focused on the data needed to reproduce figures and charts.

Consider Including a Data Dictionary
If the meanings and units of the variables are not evident in the spreadsheet, include a dictionary with explanations.

Use Open File Formats
Give preference to lightweight, plain text, non-proprietary formats like CSV for better compatibility and processing.

Choose more Robust Tools for Data Analysis
Spreadsheet programs are excellent for data entry but fall short in data analysis due to scalability, data integrity, version control, and collaboration issues. They also lack transparency because of hidden formulas, which makes error debugging difficult.

Interested in the History of Electronic Spreadsheets?

Watch Dan Bricklin, the visionary behind spreadsheets, share the story firsthand!

bit.ly/espreadsheets-history

Recommended Reading

Broman, K. W., & Woo, K. H. (2018). Data Organization in Spreadsheets. The American Statistician, 72(1), 2–10.
https://doi.org/10.1080/00031305.2017.1375989

UC SANTA BARBARA
Library

Want to learn more?
rds@library.ucsb.edu

www.library.ucsb.edu