

Organizing data in spreadsheets

Friday, September 24th, 2021
10am PDT/ 11am MDT/ 12pm CDT/ 1pm EDT

Presented by:

Dr. Karl Broman

Karl Broman, PhD
Professor
Department of Biostatistics & Medical Informatics
University of Wisconsin-Madison

There is no fee associated with this webinar, but users are asked to register to receive the Zoom link and password.

Registration: https://bit.ly/osga_2021-09-24



Sponsored by:
The NIDA Center of Excellence in Omics, Systems Genetics, and the Addictome
<http://opar.io>

Webinar Summary: Spreadsheets are widely used software tools for data entry, storage, analysis, and visualization. Focusing on the data entry and storage aspects, this presentation will offer practical recommendations for organizing spreadsheet data to reduce errors and ease later analyses. The basic principles are: be consistent, write dates like YYYY-MM-DD, do not leave any cells empty, put just one thing in a cell, organize the data as a single rectangle (with subjects as rows and variables as columns, and with a single header row), create a data dictionary, do not include calculations in the raw data files, do not use font color or highlighting as data, choose good names for things, make backups, use data validation to avoid data entry errors, and save the data in plain text files.

Broman KW, Woo KH (2018) Data organization in spreadsheets. The American Statistician 78:2–10 (<https://doi.org/gdz6cm>)