Research Data Services

Data Literacy Series, n.11/2022

#COMPRESSING FILES#

While working with large data files, you may need to reduce their size to make them more convenient to store, transmit, and download. Let's understand the bit-rate reduction process and its types.

How does file compression work?

Compression applications use mathematical equations to scan files and look for patterns to extract or replace them with smaller pieces of data or code that take up less space. These tools vary in approach (lossy or lossless), openness, cost, format support, operation system, compression ratio, and speed performance.

Lossy

- Removes details that aren't easily noticed
- Makes files up to 90% smaller
- Reduces the quality of the original version
- Commonly used for audio, images, and videos
- Lossy techniques are format-specific (e.g., MP3, JPEG, and MP4)

VS

Lossless

- Removes some redundant data from a file
- Reduces the size of files by about 50%
- Preserves the quality of the original version
- More often used for textual data
- Zip and Gzip are ubiquitous and are pre-installed on computers

Looking for more robust lossless compression options?

Open/Free & Cross-Platform

PeaZip
Supports over 200 formats and strong encryption

LZ4
Extremely fast compression speed

7zip
High-compression ratio and strong encryption support

For a more comprehensive list of compression applications and lossy and lossless formats, visit: https://en.wikipedia.org/wiki/List_of_archive_formats

Need help with research data? rds@library.ucsb.edu