

Files, Data, and Information



What's the difference between data and information?

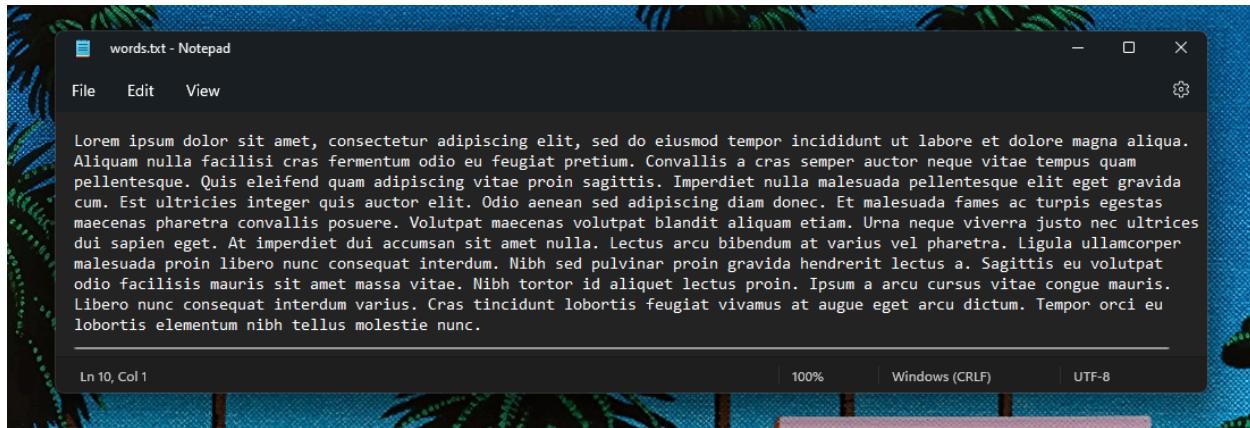
“Data” refers to raw and unorganized facts, numbers, or symbols that don't have any meaning on their own. For example, the number 32 is data, but it doesn't tell us anything meaningful without context. Information, on the other hand, is data that has been processed, organized, and given context and meaning. For instance, “the temperature is 32 degrees Fahrenheit” is information which gives the number 32 context and meaning.

What's a file?

Files are digital containers that store data and information on a computer's hard drive or storage device. These files can hold a variety of data and information and come in different **file formats**. A file format is a way of encoding bits (the ones and zeroes that make up a file) in a particular way so that it can be used for a specific purpose. These formats are associated with a **file type** which tells the user what kind of file it is, and tells the computer how to *open* the file. File types are associated with **extensions**, a series of characters that are added to the end of a file name, separated by a period. For example, a file with the ".docx" extension is a Microsoft Word document file, while a file with the ".jpg" extension is a JPEG image file. File types are information that gives meaning to the data

within a file. Below are some examples of different file types and their extensions.

Text



.txt - A plain text file format that contains only basic text characters and does not include any formatting such as bold, italics, underline, or font size. A .txt file can be opened and read by basically any computer. They are generally small in size, making them easy to transfer and store. They're also easy to create and edit with programs such as Notepad on Windows or TextEdit on Mac.

While the lack of formatting allows for tiny file sizes, it also means the text won't be pretty and can also be difficult to read. Also, .txt files cannot be used to store images, videos, or any other types of data.

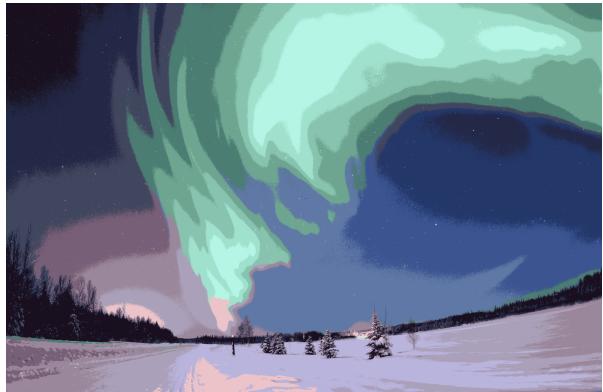
Image

.jpg or .jpeg - Used to store digital images. .jpg files use lossy compression, which means that they are highly compressed in a way that sacrifices image quality for a much smaller file size. .jpg is the most widely supported image format, and it is supported by virtually all web browsers, image editors, and other software.

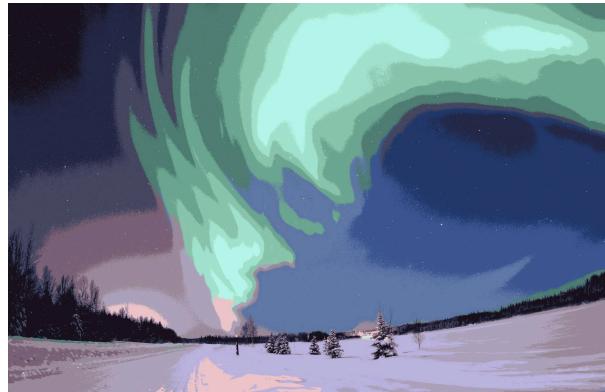
The compression used in .jpg files results in some loss of image quality, particularly in areas of the image with a lot of detail or fine gradations of color. They also don't support transparency, meaning they can't be used to create images with transparent backgrounds.

.png - An image format which uses lossless compression, meaning that image quality is preserved and that the image can be compressed and decompressed without any loss of detail. .png supports transparency, which makes it a popular choice for graphics and icons. It's also often used to store images with a lot of detail or fine gradations of color.

Because of their lossless compression, file sizes tend to be much bigger. Here is a comparison between a .jpg and a .png of nearly identical images.

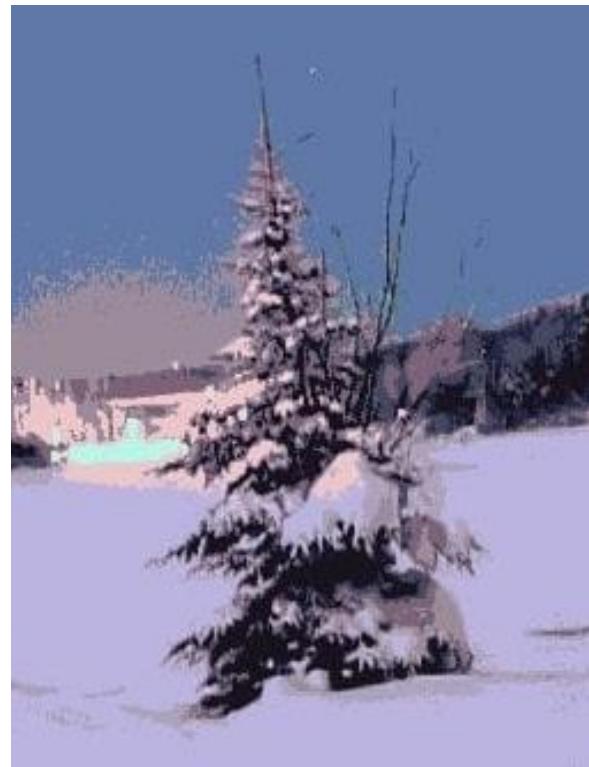
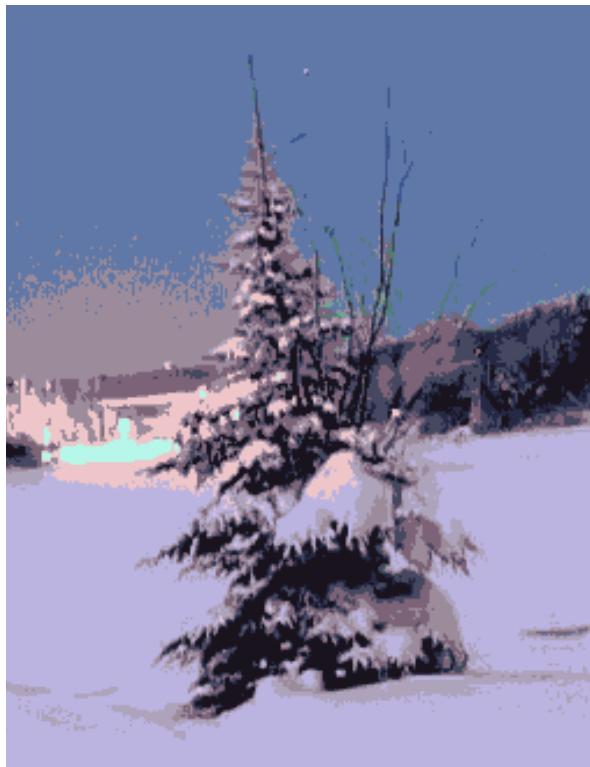


.png - 758 kb



.jpg - 480 kb

As you can see, the difference is barely noticeable until we zoom in:



You might notice some distortions and abnormalities in the .jpg image. These impurities are called artifacts and can get worse over time if a .jpeg image is repeatedly compressed. Common artifacts in JPEG images include blurring, color distortion, and pixelation. These artifacts are especially visible in areas with fine details, such as text or thin lines, and in areas with high contrast.

.gif - An image format that supports simple animated images. It uses a lossless compression technique that allows for multiple frames to be combined and stored into a single file.



.gif files also support transparency, making them a popular choice for simple graphics and icons. However, they are limited to 256 colors, which makes them unsuitable for storing high-quality images with a lot of detail. They can also be quite large in file size.

Each of these image file formats has its own advantages and limitations, and the choice of format depends on how the image will be used. You'd want to use .png when image quality is more important, and .jpeg when file size is more important.

Audio

.wav - High-quality uncompressed audio format. .wav files don't lose any sound quality and are compatible with virtually all audio software and hardware, making them a popular choice for audio professionals.

Uncompressed files are very large in size, making them impractical for storing large amounts of audio or for sharing audio over the internet.

.mp3 - MP3 is an audio file format which uses lossy compression, resulting in smaller file sizes compared to .wav files. .mp3 is widely supported by virtually all

audio software and hardware. Their small file size makes them ideal for storing large amounts of audio or for sharing audio over the internet.

Much like a .jpg image, the compression used in .mp3 files results in some loss of audio quality. While it's widely used for music and other forms of audio, it's not ideal for storing high-quality audio, such as audio recorded in a professional studio.

.flac - FLAC (Free Lossless Audio Codec) is a compressed audio format that uses lossless compression to reduce the size of the file without sacrificing audio quality. FLAC files are 50-70% smaller in size than WAV files, and they retain all of the original audio data.

However, not all devices and software support .flac, while almost everything supports .wav. And because FLAC files are compressed, they require more processing power to decode than uncompressed .wav files. This can be an issue for older or less powerful computers or devices. They also cannot be edited in the same way as uncompressed WAV files. This can limit the options available for editing or processing the audio.

Each of these audio file formats has its own advantages and limitations, and the choice of format depends on the specific requirements of the audio being stored. For most users, .mp3 will be more than enough for your audio needs. If you're an audiophile looking to get the most out of your high-quality headphones, .flac is a great option if your software supports it.

Video

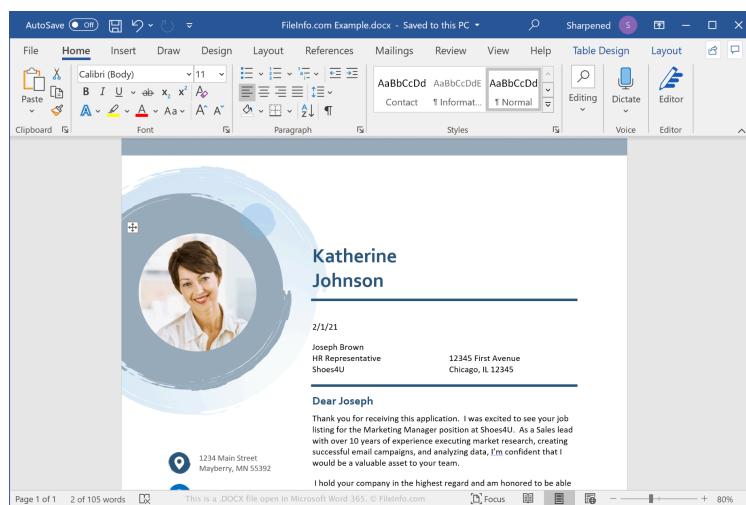
.mp4 - MP4 is a multimedia file format used for video, audio, and images. It is a container format that can store a variety of audio and video **codecs**, which allows it to offer high-quality multimedia playback while keeping the file size relatively small. MP4 files are widely used for streaming and downloading video content over the internet, and they are compatible with most modern devices, including smartphones, tablets, and computers. .mp4 files use lossy compression, which means that they are highly compressed, resulting in smaller file sizes compared to other video formats.

.mov - Developed by Apple and is compatible with both Mac and Windows operating systems. .mov files can be used for streaming videos on the internet and for storing movies and TV shows. Many popular media players, such as QuickTime and VLC, can play .mov files. .mov files can be large in size, particularly if they contain high-quality video and multiple audio and video tracks.

.avi - a multimedia format developed by Microsoft that supports many codecs, making it a versatile format for video playback. However, its usage has declined in recent years because of its large file sizes compared to other formats and limited compatibility with modern devices and web browsers.

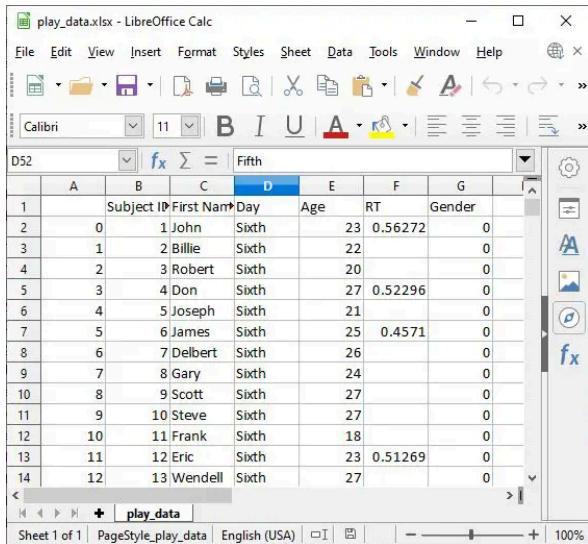
Others

.doc and .docx - Used to store **word-processing documents**, such as letters, resumes, and essays. They are created using Microsoft Word, which is part of the Microsoft Office suite of applications. Microsoft Word provides a wide range of formatting options that are not available with a simple .txt file, including font and paragraph styles, page layout, tables, images, and charts. This makes it easy to create professional-looking documents. While .doc files can be opened using other word processing software, some formatting elements may be lost or changed when the files are opened in a different program. .docx is the newer version, which offers some extra features and is open rather than proprietary.



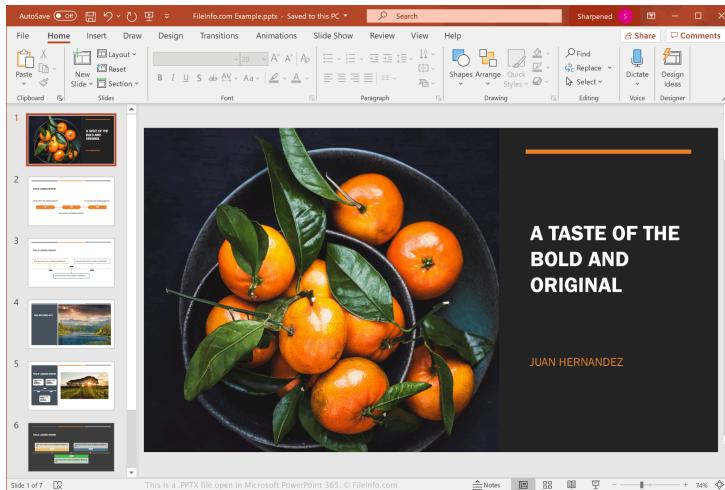
.xls and .xlsx - Used to store **spreadsheets**, such as budgets, financial statements, and data analysis workbooks. They are created using Microsoft Excel, which is also part of Microsoft Office, and are one of the most widely used

spreadsheet formats. Like .doc files, .xls files can be used with other programs but some formatting elements may be lost or changed when the files are opened in a different program.



	A	B	C	D	E	F	G
1	Subject ID	First Name	Day	Age	RT	Gender	
2	0	1 John	Sixth	23	0.56272	0	
3	1	2 Billie	Sixth	22		0	
4	2	3 Robert	Sixth	20		0	
5	3	4 Don	Sixth	27	0.52296	0	
6	4	5 Joseph	Sixth	21		0	
7	5	6 James	Sixth	25	0.4571	0	
8	6	7 Delbert	Sixth	26		0	
9	7	8 Gary	Sixth	24		0	
10	8	9 Scott	Sixth	27		0	
11	9	10 Steve	Sixth	27		0	
12	10	11 Frank	Sixth	18		0	
13	11	12 Eric	Sixth	23	0.51269	0	
14	12	13 Wendell	Sixth	27		0	

.ppt and .pptx - Used to store slideshow presentations. They are created using Microsoft PowerPoint, also part of Microsoft Office. Can be used with other slideshow software, but again, certain elements may be lost or changed when used with something other than PowerPoint.



.pdf - PDF stands for "Portable Document Format". It was created by Adobe as a way to share electronic documents that could be viewed and printed on any computer, regardless of the hardware or software used. All of the fonts, images, and formatting elements needed to display the document are embedded within the file itself. PDF files can contain a wide variety of content, including text, images, and multimedia elements. They can be password-protected and

encrypted, making them a secure way to share sensitive information. PDF files are not always easily editable, though this can be addressed by using third-party software to convert the PDF to a more editable format.

Archive and Compression formats - These are used to consolidate multiple files into a single file to save space and make them easier to transport or transfer. These file types are typically created and opened using file archiving software, and they can be encrypted and password-protected for added security. Archive and compression software includes [WinZip](#), [7-Zip](#), and [WinRAR](#). Most archive and compression software is designed to work with multiple formats, though some are more widely supported than others.

.zip: ZIP is the most common archive format, widely used in Windows systems. It is capable of compressing files without a significant loss of quality, and it's easy to create and extract ZIP files using built-in features in most operating systems. ZIP files can also be password-protected to ensure that the contents remain secure.

.7z: 7z is a relatively new format that offers higher compression ratios than ZIP. It's an open-source format that can compress large files into smaller archives. However, third-party software like 7-Zip is usually required to create or extract 7z files. And while it offers better compression ratios, it can also be slower to compress and extract files, making it a matter of time vs. space. As of 2023, Windows 11 natively supports 7z.

.rar: RAR is a proprietary archive format developed by the company WinRAR. It is similar to 7z in terms of compression ratio, but is more widely supported. It also offers better error recovery and the ability to split large archives into multiple smaller files, which can be useful for file transfer and storage. However, it requires a license to use beyond a free trial period. As of 2023, Windows 11 natively supports RAR.

.html - A text file that contains code written in Hypertext Markup Language (HTML). It is used to create and structure content for the web and is the standard markup language for creating web pages and apps. HTML allows web developers to add text, images, videos, and other multimedia elements to a web page and define how they are displayed. It also provides a structure for organizing content through the use of headings, paragraphs, lists, and other

formatting elements. HTML files can be viewed in any web browser and are often used in conjunction with Cascading Style Sheets (CSS) and JavaScript to create interactive and dynamic web pages.

.exe - An executable file in Windows that contains code that can be executed by your computer. They are often used to install or run programs on your computer. Be cautious when downloading or opening .exe files, as they can potentially contain harmful malware or viruses that can damage your computer. It is recommended to only download and open .exe files from trusted sources, such as reputable software vendors or websites. Additionally, you should have an antivirus program installed on your computer to help detect and prevent any potential harm from malicious .exe files.

More helpful information

Folders or **directories** typically contain most of the files on your computer. Folders can be nested within other folders to create a hierarchical structure, allowing for more complex organization of files.

Metadata is data that describes other data. It provides additional information about a piece of data, such as when it was created, who created it, what type of file it is, and other details. In digital files, metadata can be embedded in the file itself or stored separately in a database. For example, an .mp3 file can include metadata such as the title of the song, the artist, the album name, the release year, and more. A .jpeg file taken from a digital camera may include information such as the date and time the photo was taken, camera settings (such as shutter speed, aperture, and ISO), GPS location data, and camera model.

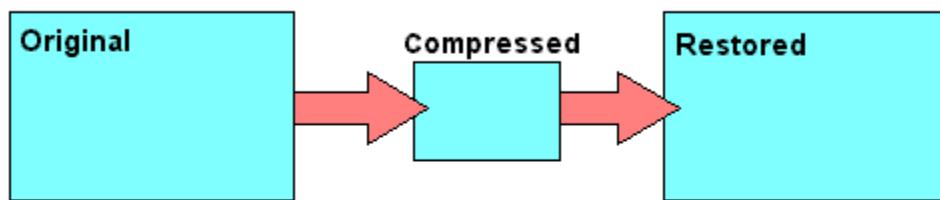
You can use metadata for many purposes such as organizing and managing files, improving searchability, providing information about the origin and history of a file, and ensuring the accuracy and integrity of data. For example, a digital photograph may contain metadata that includes information about the camera used to take the photo, the date and time it was taken, and the location where it was taken. This information can be used to help organize and categorize the photo, make it easier to find using search tools, and verify its authenticity.

Lossy and **lossless** data compression are two different methods for reducing the size of digital files, which was briefly mentioned earlier. Lossy compression

works by permanently removing some of the data from the original file, leading to a smaller file size but also a loss of some quality. Some common examples of lossy file formats include JPEG, MP3, and MP4.

Lossless compression works by reducing the size of a file without any loss of data or quality. The original file can later be perfectly reconstructed from a lossless compressed file. Lossless compression is commonly used for text files, software programs, and digital documents, where every bit of data must be preserved. Some common examples of lossless file formats include PNG, FLAC, and ZIP. The choice between lossless and lossy compression depends on whether the priority is file quality or file size.

LOSSLESS



LOSSY

