

Preparatory Notes

Understanding External data

Data most times won't be recorded manually and when they are recorded. You'd rather not do so with R. why? It is an excruciatingly difficult process to create data with R. It is actually way easier to record data with spreadsheets than to record it with R, and this is even when you use the `edit()` function.

Data Sources

When we want to import data into R, we need to know where to get it from. A ***data source*** is the place where the data we want to use originates from. This can be a physical or digital place. A data source could be any of a live measurements from physical devices, a database, a flat file or plain-text file, scraped web data, or any of the innumerable static and streaming data services which abound across the internet. An example of a data source is data.gov, and wikipedia and these are web data. Another example that is actually close to you is your mobile phone. It's a mobile database holding contact information, music data, games, and so on.

Data Format

Data formats represent the form, structure and organization of data. It defines how information is stored, accessed, and interpreted. It's a standardized way to represent data, whether in files or databases, and is crucial for efficient data management and processing. This brings us to another aspect that is important to knowing how to handle data, and that's **file extension**. If you've taken a closer look at the music/audio file on any of your device, you usually see a dot which separates the name of the file and a text which is usually fairly consistent depending on your file organization. This text is regarded to as the **file extension**. This is also true for your video, and picture files. For example, you could see the following extensions:

Table 1: Some of the common multimedia file format including audio, picture and video files.

Audio	Pictures	Video
.mp3	.png	.mp4
.aac	.jpeg	.mov
.flac	.gif	.avi
.wav	.webp	.web,
.ogg	.tiff	.mov

For data set, there are some common file format such as:

- flat file or plain text: `.csv`, `.tsv`, `.txt`, `.rtf`, `.csv2`
- some web data format excluding those above: `.html`, `.json`, `xml`
- spreadsheet: `.xlsx`, `.xls`, `.xslm`, `.ods`, `.gsheet`
- others: `.shp`, `.hdsf`, `.sav`, and many more.

Importing Data

- Plain-text data
- Spreadsheet
- Web API
- Web Scraping
- Introduction to Document Creation
 - Creating reports
 - Creating manuscripts
 - Website creation
- Capstone Project