BDK12 Data annotation and curation

Units 1 and 2: Data preparation and planning and File and Directory Naming

Exercise:

In a hypothetical study, you are collecting survey data from participants in your community. You will collect structured survey data in a table from 100 participants and this data will include their names, addresses and birthdates.

1. While you are collecting the data, what type of file should you use?
   1. None, it should be handwritten
   2. .docx (MS Word or equivalent)
   3. .xlsx (MS Excel or equivalent)
   4. .mp3 (a voice recording)
2. For long-term preservation and sharing, will you change the file type? Why? What will the new file type be?
3. To share survey questions or other documents, what file type(s) should you use for sharing and preservation?
4. Following the recommended practices for file naming, write examples of file names that you would use for your data files and survey documents.
5. A data dictionary is used to describe data objects so others can understand them, and can enter the data properly and consistently. You can read more about data dictionaries here: https://en.wikipedia.org/wiki/Data\_dictionary

Create a data dictionary for your survey data.

|  |  |  |  |  |
| --- | --- | --- | --- | --- |
| Variable Name | Description | Format | Units | Values |
|  |  |  |  |  |
|  |  |  |  |  |
|  |  |  |  |  |

Variable Name: how it appears in your data

Description: a brief explanation of the variable

Format: using best practices or your own judgement, decide how the data will be recorded. For example, for a date, will the format be YYYY-MM-DD, MM/DD/YY, etc.

Units: units of the variable. For example, if the variable were height, would you be using centimeters or feet/inches.

Values: any restricted values. For example, if you are only collected data about current residents of Oregon, they could not have a non-Oregon address.