Critical report of collection

For this dataset I have digitized and datafied my existing collection of books. I have made a conscious choice in the digitization of the collection to only include books that I have gotten of my own accord, meaning that I either bought them myself or received them as a gift resulting from my desire to own the book. This means that any books I have acquired as a result of needing them for classes of my bachelors degree in Cognitive Science are not included in the digital collection. Furthermore, the collection includes books first borrowed and since kept in accordance with the original owner.

The collection of the books themselves has, as mentioned, happened only by my desire to own the book, without any external decision-making agents. Some books have been acquired as a product of recommendation and subsequent desire to own the book after having researched it. The reason for collecting books stems from my enjoyment of reading as well as always having a book to read, so the collection is therefore evergrowing, so I don't run out of books. Another factor is I simply like to own the books I read, so I can read them again, reference them and have them as decoration.

The collection of books is stored in several places on shelves in my apartment. One shelf is dedicated to books with black back covers for uniformity in the expression of the shelf. Most books are stored on a wooden beam situated on a back wall. The books here are organized by several factors. Because the ceiling is situated at an angle, smaller books are placed there. After that books are placed by arrival time, left to right, with the longest owned books to the left and the most previously acquired books to the right. Lastly, some books are stored on a shelf behind my bed, because of their size, making them unable to fit on the beam. These are organized from brightest to darkest cover. One overarching arrangement of the books is that all books that are part of a series or written by the same author, are placed next to each other. Additionally, some books that have distinctive appearances and share this appearance with other books are placed next to each other.

The digitized collection consists of 59 books of varying topics, categories and genres. It includes both fiction and non-fiction.

The digitization and datafication of the collection has been undergone by taking a picture of each book's cover and entering the chosen observations and data for each book in an excel sheet. The images of the books and the excel sheet are kept in a folder together. The chosen characteristics and data for the books are as following:

- Title
- Author
- Color (Front page)

- Topic (What is the books about)
- Category (Fiction/Non-fiction)
- Page count
- Publisher
- Format (Paperback, hardback etc.)
- Genre
- Field (Is there a field of study this relates to)
- Language
- Original publish year
- Condition (1-10, subjective judgment)
- Version publish year
- Part of a series?
- Original language
- ISBN
- Image (The name of the image of the related physical book)

The choice of these characteristics comes down to both what I observed immediately and what I find interesting about the collection. Most of it is meta-data or attribute data, but I believe that some interesting visualizations and insights can come from this. The only truly indexical variable I have included is the ISBN number. Here the ISBN-13 was chosen where it was available, for some only the ISBN-10 was available, this was chosen in absence of an ISBN-13. Furthermore, I believe that more data is better than less data. The reason for excluding some variables like size is that it is both time consuming and intuitively of little value to the digitized collection. The organization of the digitized collection in the excel sheet is determined by my making my way from one side of my apartment to the other and entering the data for each book. I started by the shelf with the black books, then went to the beam and lastly behind the bed. The images are most randomly distributed as they were taken over the course of a longer period and some books were not in their place in the collection, so the image names in the sheet are jumbled.

The dataset includes variables of most categorical nominal variables and continuous interval and ratio variables. The data is mostly structured and is complete in all variables but one, which is the field, as not all the books pertain to a field of study per se. The data is all primary captured data, as I have deliberately chosen what to include in the dataset and none of the data has been derived from the other data.

Mostly the data is gathered in an intuitive "the more data the better" approach, because any unnecessary data can be filtered out later if it turns out to be irrelevant. This is a personal choice for me as I prefer to exclude variables rather than to be missing key information or data, when trying to analyze the data and visualize it.