

Master Project: Data-driven Archive Visualization

Huub van de Wetering

Technische Universiteit Eindhoven

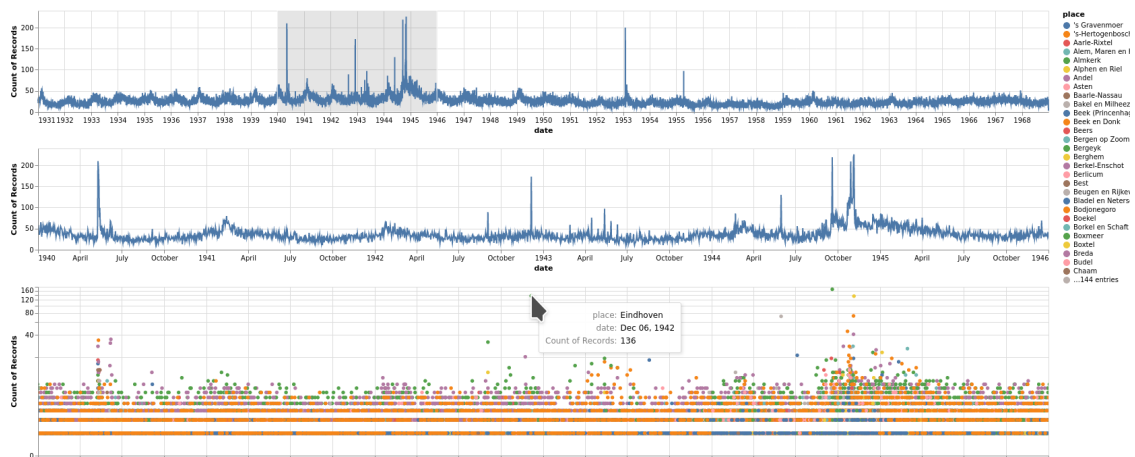


Figure 1: Illustration of daily burial record counts for the province Noord-Brabant in period 1930-1969 (top), in period 1940-1946 (middle, bottom) with counts per city (bottom). The highlighted count refers to December 6 1942 when the Philips factories in Eindhoven got bombed by the Royal Airforce Force resulting in over 130 casualties.

1 Introduction

In the Netherlands there is a rich tradition of preserving archives. For instance, before the actual introduction of 'de burgerlijke stand' (civil status) by Napoleon in 1811, there was already a long tradition of church records in which events as baptism, marriage, and burials were registered. Nowadays, historical records are kept in central archives as for instance BHIC (Brabants Historisch Informatie Centrum,[1]). To facilitate easy access to these archives more and more of them have been scanned, digitized, and made available as open data, see for instance [2]. Digitization is by no means an easy task as it may require reading hand-written text on scans of less well preserved non-standardized documents. The digitized documents are used by both professional researchers and individuals; both can use (online) tools to search records by filtering on attributes of an event, giving them access to detailed information in the databases.

The digitization of the scans is managed and monitored by BHIC but actually done by volunteers. One project of this kind is the digitization of birth, marriage, and burial (BMB) archives of churches. These records originate from books that are organized per archive. Such an archive is typically coupled to a municipality and contains several books, each of which contains the records for a certain time period. For instance, one of the over 300 archives with BMB records in the province of North Brabant has number 1449, is of municipality Sint-Oedenrode, and contains around 40 books of which book with inventory number 30 covers 381 burial records in the period 1806-1810. In total there are at the time of writing 6.363.917 BMB records from 1473 till 1959 digitized. Next to these civil status archives, there are also civil registration archives, and other types of archives in the process of being digitized; see [3].

2 Problem description

Digitization of archives is an ongoing process in which the books in archives are handled in some order. The partial availability of the digitized records poses a problem for both volunteer, researcher and manager. All of them are left in a state in which they have no clear view what has been digitized. Such a view could be on the level of an archive, book, time-period, municipality, etcetera. Using the digitized records as data this problem can be mitigated by designing and implementing an interactive visualization tool.

Using this data also opens the door for additional visualizations. For instance, simply showing frequency counts of record types per day, might reveal interesting information that could trigger professional curiosity of archive researchers. Figure 1 shows such a case in which a visualization shows the death count for the province North Brabant in the period 1930-1969. The peculiarities in the visualizations, like peaks and periodicity or lack of those, may be interesting to examine for the whole period (top view), a selection of that period (middle view), or per municipality (bottom view). Apart from the obvious patterns there might be more hidden patterns in the data that with a suitable interactive (statistical) analysis might come to light.

3 Challenges

Among others, the challenges from a visualization stand point are:

- Handle problems in the data. This may be done by pre-processing, or by simply visualizing these problems.
- Design suitable visualizations and interactions that support users to solve their problems or discover their questions;
- Handle large datasets with millions of records.
- Implement a web-based solution.

4 Project

- The Project is executed at the TU/e.
- The data is available from BHIC either directly (csv) or via open access (XML).
- Last but not least, BHIC is willing to support the student by answering questions on the data and willing to participate in discussing goals and requirements.

References

- [1] Brabants Historisch Informatie Centrum. English - bhic. <http://www.bhic.nl/english>. Accessed: 2022-09.
- [2] Brabants Historisch Informatie Centrum. Stamboom. <http://www.bhic.nl/onderzoeken/stamboom>. Accessed: 2022-09, in Dutch.
- [3] Brabants Historisch Informatie Centrum. Wat zit er in? <https://www.bhic.nl/onderzoeken/wat-zit-er-in>. Accessed: 2022-09, in Dutch.