

Visualizing Denmark's Cultural Heritage

Stefan Jänicke

stjaenicke@imada.sdu.dk

IMADA, University of Southern Denmark, Odense, Denmark

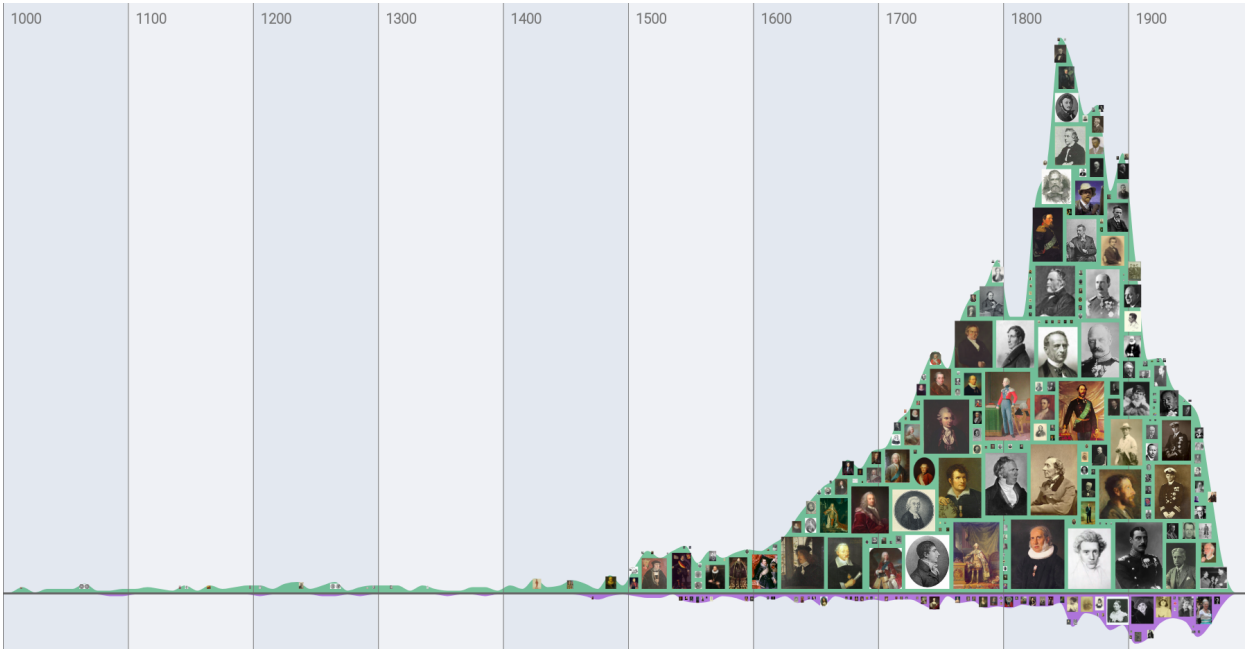


Fig. 1. Timeline depicts the lifetimes of women and men in a digital edition of the Dansk biografisk Lexikon (dDBL).

Abstract— Visualization as a method to reveal patterns in large data sets is a powerful tool to build bridges between data science and other research disciplines. The value of visual design is documented with a showcase on the Dansk biografisk Lexikon. The original version of this article was published in the 2020 November issue of *Aktuel Naturvidenskab* [5].



THE proverb "a picture is worth a thousand words" means that still images embody the power to express complicated concepts, ideas or relationships much more effectively than textual descriptions. The painting *Udslidt* (Figure 2) by social realistic painter Hans Andersen Brendekilde can be seen as an artistic implementation of this proverb. It reflects the hard rural life in late 19th century Denmark, specifically emphasized by the man who died from hard graft, his screaming wife, and the wasteland that frames the scene. Even though the latter sentence provides a detailed description of the problem, the painting does a much better job in conveying the socio-critical message to the observer.

Visualization as a method to analyze data displayed on the screen makes use of the human ability to quickly perceive patterns inherent in the data. A visual design process translates a real world concept (e.g., Danish art) into visual representations in order to make the concept easier accessible to the observer. In this context, real world objects (e.g., Danish painters) related to that concept are transformed to data entities (e.g., persons) described by specific attributes (e.g., life data, number of paintings, painting style) and relations among them (e.g., familial, collegial or teacher-student relationships between painters). These attributes and relations are then translated to a visual vocabulary (e.g., a time chart depicting the life data of painters or a social network graph to illustrate relations) capable of reflecting occurring patterns. The visual design process always involves a distortion, thus, needs to be undertaken with care in order to ensure that the observer is able to draw accurate conclusions. However, especially in application sce-

narios involving non-experts in computer science, visualizations are indispensable means for hypothesis verification and generation concerning domain-specific data sets.



Fig. 2. Hans Andersen Brendekilde's *Udslidt* (1889).

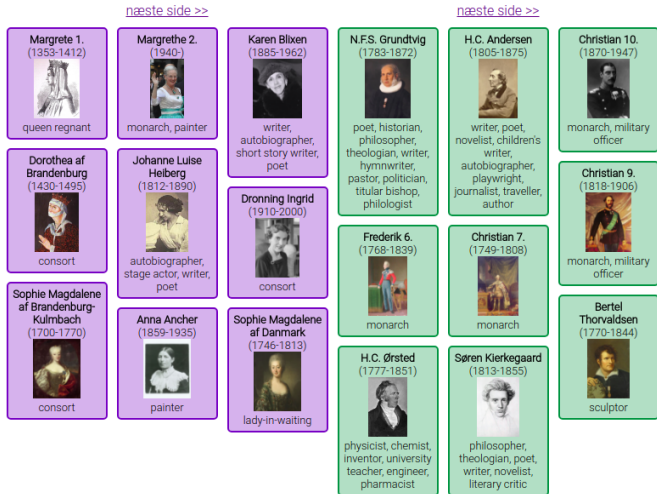


Fig. 3. Numbers of women and men in the dDBL.

In the past ten years, I have been working together with experts from various domains, among which most of my projects were interdisciplinary collaborations with humanities scholars in the context of the digital humanities. As a consequence of the digitization endeavours in the past decades, humanities scholars now have on-demand access to large amounts of cultural heritage data, and the digital humanities as a research domain addresses the need of developing solutions to make sense thereof. Visual data exploration has been proven a valuable research instrument in this context and continues on gaining more and more importance for knowledge discovery [7]. The remainder of this article sheds a light on how visualization can reveal different aspects of Denmark's cultural heritage.

VISUAL EXPLORATION OF THE DANSK BIOGRAFISK LEKSIKON

In her article on cultural heritage, humanities scholar Elena Franchi emphasizes that all people contribute to the culture of the world [3]. The more important a person for the culture heritage is, the more is written about that person and related artifacts. The most influential persons for Denmark's cultural heritage are likely to be found in the Dansk Biografisk Leksikon (DBL). Its third and last edition published between 1979 and 1984 offers biographical information for around 20,000 persons, both Danes and foreigners who significantly contributed to Danish social life and culture. For around 6,000 persons, an entry exists in the Danish Wikipedia [12], which can be seen as a digital DBL edition (dDBL) containing the persons of most contemporary relevance. A variety of structured biographical information is available, amongst which gender, lifetime, place of birth and professions are provided for a large number of persons. The diversity of these attributes provides a suitable basis for a visual exploration scenario, in which we focus on the comparison of women and men in the dDBL.

The gender ratio is depicted in a bar chart view in Figure 3. This first, very basic form of visual data representation reveals the strong gender imbalance in the DBL accurately, as the displayed ratio corresponds to the ratio of the third printed edition that contains around 7-8% women. The bar chart is coupled with a list view that orders all women and men according to their suggested impact on the Danish cultural heritage. This impact is approximated based on the Wikipedia entry of a person: the more text is written about a person, the higher its impact. Clicking an entry redirects to the corresponding Wikipedia entry.

In order to support information seeking [11], three interactive visual interfaces have been designed, all of which support a comparative gender analysis of biographical features.

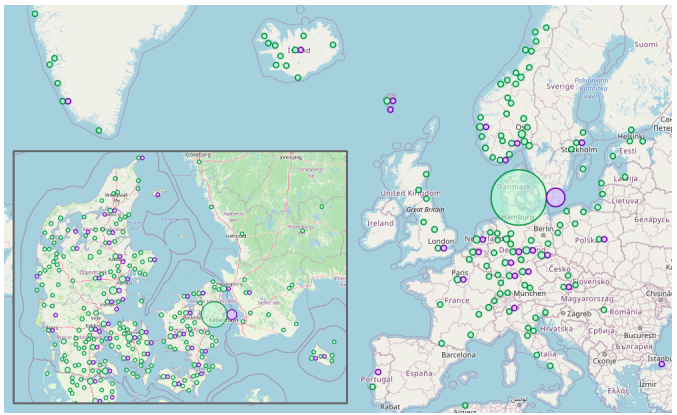


Fig. 4. Map shows places of birth of women and men in the dDBL.

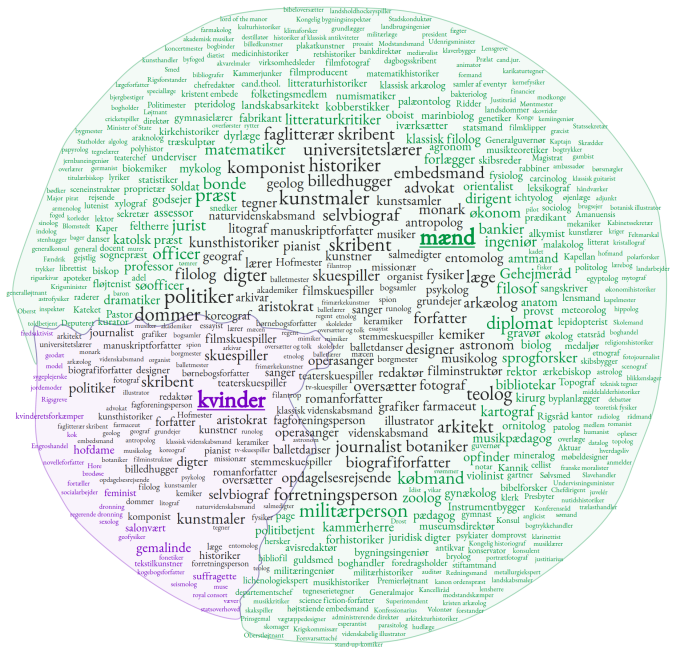


Fig. 5. Tag cloud compares professions of women and men in the dDBL.

A timeline separates women and men in two streams. Persons are grouped into decades using the mean lifetime year (or the date of birth plus twenty years in case of living persons or the date of death in case the date of birth is not known). Thumbnail images placed on the streams inform on the most impactful persons for specific time periods. Figure 1 illustrates the entirety of lifetime data in the dDBL. Whereas the large number of men in the 19th century pictures the Danish Golden Age, a period of exceptional creative production, women seemingly received major attention when the first edition of the DBL was published between 1887 and 1905.

A map [8] uses non-overlapping circles to contrast the places of birth. It has to be said that this information is only provided for around half of the persons. However, a geospatial exploration supports questions like "What persons of historical importance were born in my city?" The size of a circle reflects the amount of persons born in a specific region. Though most persons in the dDBL were born in Denmark as shown in Figure 4, the widespread distribution of dots yet documents an international impact on Denmark's cultural heritage.

A tag cloud visualization [6] illustrates the differences and commonalities among the professions of women and men in the dDBL. While tags colored in purple or green represent professions that were only practiced by women or men, black-colored tags stand for profes-

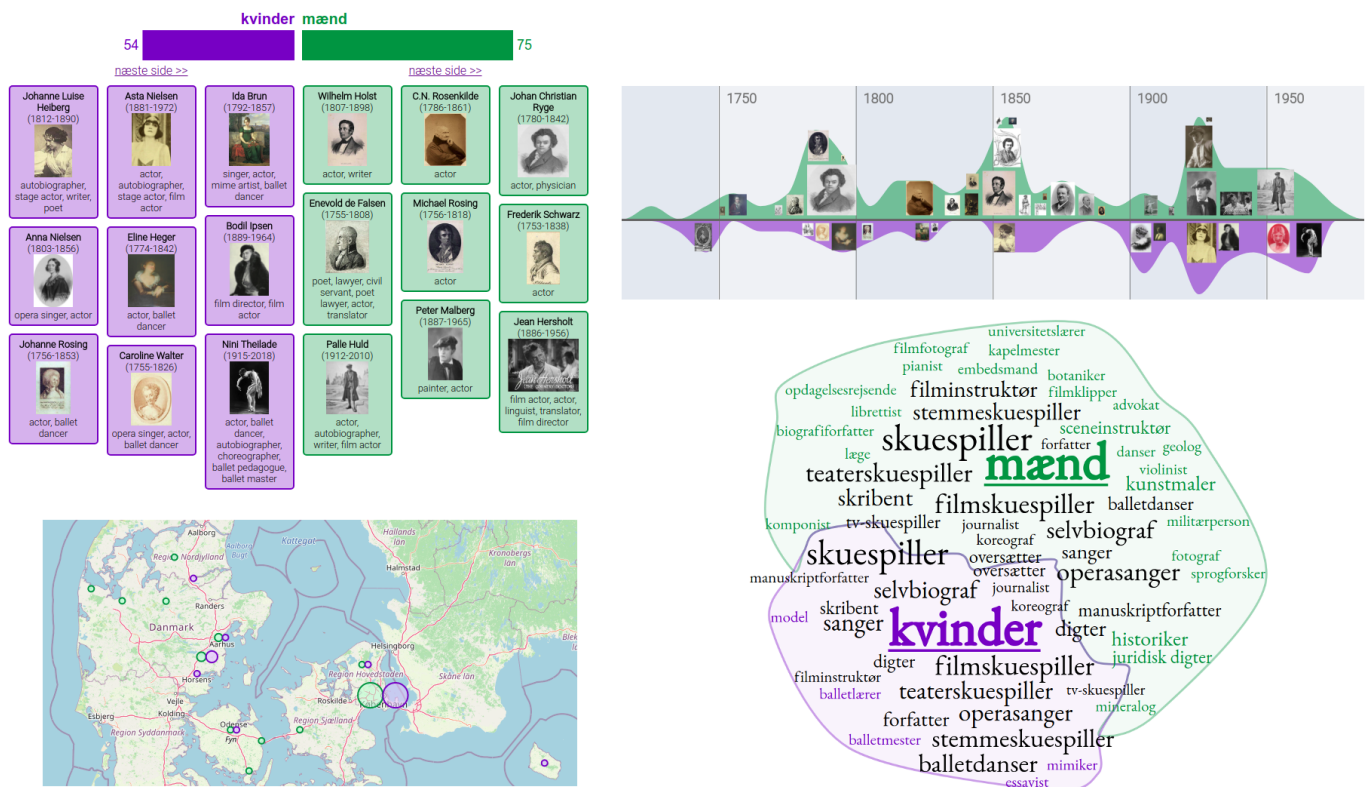


Fig. 6. Balanced gender ratio of 129 persons related to acting.

sions practiced by both groups. Font sizes reflect how often a profession has been practiced. Black tags are further positioned according to their balance between the two groups. If the same number of women and men practiced a profession, tags are positioned closely to the center of the cloud, whereas tags move to the outer bounds when the ratio is imbalanced. At least one profession is provided for 5,127 persons, and a summary of all practiced professions is given in Figure 5.

Each visual interface provides an overview of the distribution of values for the corresponding attribute using the same color scheme, and charges the visual attribute of size to reflect quantities. All interfaces are linked with each other and means of selecting specific attributes via mouse interaction enable a multifaceted visual exploration of the data set. All views are updated only including persons matching the query that can be composed of selections concerning multiple attributes, thereby supporting manifold research inquiries. For example, selecting a time range from 1700 to 2000 in the timeline and professions related to acting in the tag cloud leaves 129 persons with a far more balanced gender ratio. The result is depicted in Figure 6. The map further indicates that acting is a rather urban profession as related persons were typically born in large cities.

This case study on the dDBL exemplifies how visualizations generate humanly perceivable patterns that can guide the user in a visual exploration process. In addition, visualizations are capable of emphasizing important data entities, which are, in the case of the dDBL, persons who significantly contributed to Denmark’s cultural heritage.

A DISTANT READING OF HANS CHRISTIAN ANDERSEN’S AND SØREN KIERKEGAARD’S WRITINGS

TWO among the most impactful male persons in the dDBL are Hans Christian Andersen and Søren Kierkegaard, both of which are recognized as the most important Danish writers of the Danish Golden Age. While Andersen is primarily known for his fairy tales, Kierkegaard is perceived as the first existentialist philosopher through his writings. Both shared an ambivalent relationship reaching from contempt to respect, and they also referred to each other in their

works [2]. We perform a distant reading approach to investigate if this ambivalence is reflected in their writings.

Distant reading, a term defined by the digital humanist Franco Moretti [10], describes quantitative approaches to text analysis. Instead of reading texts, in analogy to Information Seeking, texts are regarded from an overview perspective, in which typically frequencies of words are compared. Figure 7 provides such a quantitative view on a fragmented, but representative text corpus including 89 stories (240,734 words) of Andersen (blue) and 12 writings (90,770 words) of Kierkegaard (green). The frequencies of words from both authors are counted, normalized in relation to the numbers of words, stopwords are removed and the 250 most frequently used words from both authors are comparatively visualized. Stopwords are most common words in a language, e.g. short function words like articles or prepositions. They appear frequently in all texts and contribute less to the meaning of it. Without filtering them out, the 250 most frequent words would be mostly stopwords.

On first sight, the two thematic fields are well-reflected by the blue and green sectors of the tag cloud. While blue-colored words like "prinsessen", "slot/tet" or "kejseren" explicitly refer to Andersen’s fairy tales, other words denoting plants, animals or colors accurately picture the scenes in which the stories are told. On the other hand, Kierkegaard’s focus on religion is depicted by frequently used words like "gud", "biskop" or "kristendommen". The shared part in the center of the cloud mostly contains adjectives and function words. However, both writers paint their personal picture of the world as both frequently use the word "verden".

Figure 8 analyzes the context, in which the word appears in more detail using a different visual mapping, adapted from a distance-oriented tag cloud visualization [9]. The blue-bordered sector refers to how Andersen used "verden", the green-bordered sector to how Kierkegaard used it. The color of a word informs on the average distance of words that frequently co-occur with "verden". Red-colored words are typically descriptive and appear very close to verden in the text, e.g., as left or right neighbors. Yellow words typically define

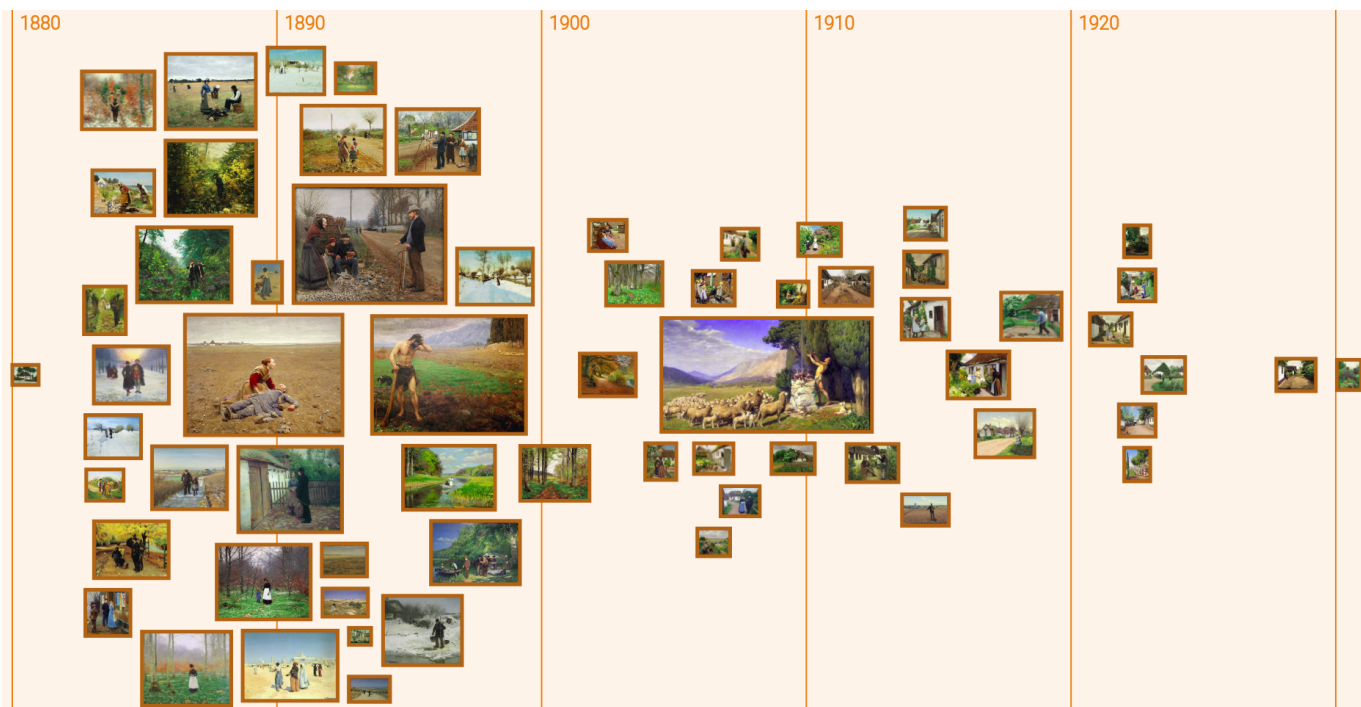


Fig. 10. 64 of Hans Andersen Brendekilde's paintings arranged on a timeline.

ACKNOWLEDGEMENT

I thank my wife Christin for comparing the bodies of work of Hans Christian Andersen and Søren Kierkegaard, and, at the same time, practicing Danish with me.

REFERENCES

- [1] T. Arnold and L. Tilton. Distant Viewing: Analyzing Large Visual Corpora. *Digital Scholarship in the Humanities*, 34(Supplement.1):i3–i16, 2019.
- [2] P. Borberg and T. Lauritsen. Da Kierkegaard røg i flæsket på Andersen, 2013. Videnskab.dk, <https://videnskab.dk/kultur-samfund/da-kierkegaard-rog-i-flaesket-pa-andersen> (Retrieved 2020-05-10).
- [3] E. Franchi. What is cultural heritage?, 2020. Smarthistory, <https://smarthistory.org/what-is-cultural-heritage/> (Retrieved 2020-05-10).
- [4] S. Jänicke. Timages: Enhancing Time Graphs with Iconographic Information. In *LEVIA'18: Leipzig Symposium on Visualization in Applications*, 2018.
- [5] S. Jänicke. Visualisering af Danmarks kulturarv. 5:18–22, 2020.
- [6] S. Jänicke, J. Blumenstein, M. Rücker, D. Zeckzer, and G. Scheuermann. TagPies: Comparative Visualization of Textual Data. In *Proceedings of the 13th International Joint Conference on Computer Vision, Imaging and Computer Graphics Theory and Applications - Volume 3: IVAPP*, pages 40–51. INSTICC, SciTePress, 2018.
- [7] S. Jänicke, G. Franzini, M. F. Cheema, and G. Scheuermann. Visual Text Analysis in Digital Humanities. 36:226–250, 2017.
- [8] S. Jänicke, C. Heine, and G. Scheuermann. GeoTemCo: Comparative Visualization of Geospatial-temporal Data with Clutter Removal based on Dynamic Delaunay Triangulations. In *Computer Vision, Imaging and Computer Graphics. Theory and Application*, pages 160–175. Springer, 2013.
- [9] S. Jänicke and G. Scheuermann. On the Visualization of Hierarchical Relations and Tree Structures with TagSpheres. In *International Joint Conference on Computer Vision, Imaging and Computer Graphics*, pages 199–219. Springer, 2016.
- [10] F. Moretti. *Graphs, Maps, Trees: Abstract Models for a Literary History*. Verso, July 2005.
- [11] B. Shneiderman. The Eyes Have It: A Task by Data Type Taxonomy for Information Visualizations. In *Visual Languages, Proceedings*, pages 336–343, 1996.
- [12] Wikipedia. Kategori:Personer i Dansk Biografisk Leksikon, 2020. https://da.wikipedia.org/wiki/Kategori:Personer_i_Dansk_Biografisk_Leksikon (Retrieved 2020-05-10).