

Glyph-based visualization of health trajectories

Harri Siirtola, Javier Gracia-Tabuenca,
Roope Raisamo
TAUCHI Research Center
Tampere University
Tampere, Finland
firstname.lastname@tuni.fi

Marianna Niemi
Computational Biology
Tampere University
Tampere, Finland
marianna.niemi@tuni.fi

Mary Pat Reeve
Finnish Institute for
Molecular Medicine
University of Helsinki
Helsinki, Finland
mary.reeve@helsinki.fi

Tarja Laitinen
Administration Center
Tampere University Hospital
Tampere, Finland
tarja.laitinen@pshp.fi

Abstract—Whenever a diagnosis is given, a procedure is performed, or a drug is prescribed, it leads to an entry into an electronic health record (EHR) system. Previously, this data was difficult to utilize because of rules regarding confidentiality, but new security approaches and pseudonymization have enabled us to work with this data.

Health-related data is voluminous and complex, and it can be difficult to abstract a meaningful overview. One of the complexities is its longitudinality. Often medical research is cross-sectional – we often take a point in time for analysis, when instead, it might be more beneficial to see the trajectory that led to the point in time.

We are currently developing a trajectory visualization tool for longitudinal electronic health data. It is a web-based tool that interfaces with the OHDSI data infrastructure and visualizes the cohorts and concept sets (groups of medical codes) defined via the OHDSI Atlas GUI.

Currently, our tool is in user testing and it will be deployed to a wider user group during the spring. The user feedback has been positive. Users find the tool especially useful in understanding and debugging their OHDSI Atlas cohort definitions.

Index Terms—health information management; data visualization; glyphs

I. INTRODUCTION

The value of real-world health data has been recognized in clinical decision-making and biomedical research, whether investigator-led or industry-initiated studies. Different health registries and electronic health records (EHRs) provide valuable data to develop better treatments, design medicinal products, and even predict health problems before they arise.

During the last years, access to health data has improved. The change has been possible due to the development of secure analysis environments and improved regulation of data protection (e.g. in Finland, the law of “Secondary use of health and social data”¹). Additionally, there are global efforts to develop a common data model (CDM) for health data, allowing one to perform studies with hundreds of thousands of cases and controls and make discoveries affected by genetic makeup and environmental exposures.

The health data is voluminous, complicated, and longitudinal by nature. As known, we should always visualize the data before analyzing. Without a comprehensive understanding of the data set, it is possible to take an approach that leads to wrong conclusions [1]. We address how to make longitudinal

health data visible and therefore enable better understanding of the proportion of missing/unreliable data, view individual patient profiles and their potential clustering with other similar profiles, and to find outliers and the relations between variables.

In this paper we present an interactive visualization tool for longitudinal health data. This tool is under development in the FinnGen² research project [2] and will eventually be made available as an open-source tool. The aims of the tool are as follows:

- to help understanding the change over time aspect of health register data;
- to visualize a large number of longitudinal cases;
- to facilitate understanding of cohorts and concept sets created within OHDSI Atlas tool.

We call our visualization program *Trajectory Visualization Tool (TVT)*. In the following, we will discuss the previous work in the area, introduce our tool by using mock data and discuss the ideas for further development.

II. BACKGROUND

Visualization of individual patient history has been discussed, for example, in the context of the *LifeLines* tool [3], [4], which has been applied not only to patient history but also to the visualization of a person’s criminal record. *LifeLines* creates interactive timelines to which events have been attached and can be handled differently. The events are divided into different categories that can be hidden and displayed, and the timeline scaling and focus can be changed as needed. *LifeLines* implements Shneiderman’s data visualization mantra (“overview first, zoom and filter, then details-on-demand”, [5]). A newer version, *LifeLines2*, can be used to visualize the timelines of up to a dozen people at the same time [6], and there is also an proposed extension to the technique to show temporal uncertainties [7].

The visualization of large amounts of patient information has been dealt with relatively little in literature. Rind et al. [8] conducted a survey where they went through 14 systems presented in the literature. They evaluated systems with the following criteria: number of data types, multidimensional presentation capability, maximum number of patient records that

¹<https://stm.fi/en/secondary-use-of-health-and-social-data>

²<https://www.finnngen.fi/en>

can be seen at a time, and support for different intent [9]. Of the evaluated systems, only eight supported the visualization of more than one patient record at the same time, and they were also able to present a maximum of ten record variables. Thus, systems developed to visualize large patient cohorts were not found in the mapping. Rind et al. [8] state that the main reason for this is the lack of a benchmark database – it is very difficult to access the right patient information, so you should create an example database that is quite similar to real patient information. It would then be possible to better compare the systems developed for the visualization of patient information. This has been later accomplished in OHDSI project³ [10], [11].

Müller et al. have proposed using adaptive and multilevel glyphs [12], [13] for presenting large biomedical data sets. They use more complex visual encoding and 3D elements.

Wang et al. [14] have presented a visual analysis approach to analyze EHRs. Their method can process and display thousands of patient records as informative plots that are based on the Sankey diagram [15]. Their system can be used to visualize, e.g., comorbidity trajectories of patient cohorts.

The Danish Disease Trajectory Browser (DTB) [16] is a tool developed to explore almost 25 years of data from the Danish National Patient Register, containing about 7.2M patients and 122M observations. It can show visually disease progression patterns and combine them into disease trajectory networks, displaying an entire multimorbidity spectrum of a disease in a single connected graph.

III. THE DATA PROTECTION ISSUE

Data protection is a top priority in research systems that process personal health information. Ideally, the data should be completely anonymous, but that goal is difficult to achieve. The practical approach recommended by WHO [17] is to pseudonymize and aggregate the data, and limit the access to systems processing it. This approach has been adopted also in FinnGen: register holders supply the data pseudonymized and it is processed in a system without network connection. The data analysis system can only be accessed after a successful completion of training course on data protection issues. While the pseudonymized personal-level data is available in the system, only aggregated results ($N > 5$) are allowed to be exported (after approval process).

In the analysis of trajectory data, the aggregation of timelines does not make sense, and the data must be aggregated longitudinally, into months and years. In addition, a random but same error of ± 15 days is introduced to all dates of a person. While the data is transformed like this, we can still separate intermittent and continuous drug administration and diagnoses, and detect seasonal changes.

Currently, the data protection requirements allow only users having person-level data access to use TVT. We are exploring other trajectory abstractions to find a representation method that would not have this limitation.

IV. GLYPH TIMELINES

A glyph in TVT is a construct of six rectangles. Color of the rectangle can be freely chosen, and the height of the rectangle depicts the relative frequency of events encoded by the rectangle. Figure 1 shows a simple single-rectangle timeline with the values depicted below the bar.

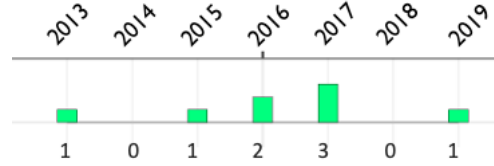


Fig. 1. Representing a simple data trajectory with bars.

The event frequencies in TVT are always aggregated, usually per year, but also per month in some situations. Seeing the exact date of an event is not needed when you are interested in the trend of events.

The event frequencies are scaled within each timeline by default. There is an option to use global scaling which enables easier comparison of frequencies between timelines. However, this is not generally useful as very high-frequency events in one timeline can minimize trends in other timelines so much that they cannot be seen.

Showing the variation of a single value over time is sometimes useful, but generally we want to see an interplay of several variables. The glyph design we propose is a simple one, considering the design space [18], but the data volume requires visual simplicity. TVT uses a six-segment glyph to show the trajectories of six variables. The encoding for a single glyph segment is defined as in Figure 2.

The top-level menu structure of TVT user interface is also shown in Figure 2. The main functions are:

- select a database where the data is extracted from (“Data sources”)
- save and restore a snapshot of the extracted data (“File”)
- assign concept sets, visit types and colors to glyph segments (“Encoding”)
- filter data, both horizontally and vertically (“Filtering”)
- reorder data timelines, both horizontally and vertically (“Reordering”)
- cluster data to gain overview (“Clustering”)

As an example, suppose we are interested in visualizing a cohort of patients having the most common asthma comorbidities as a timeline visualization. Beasley et al. [19] report a dozen asthma comorbidities, and nine more with elevated co-occurrence rate. As we can select six conditions at a time, we could choose chronic rhinitis and sinusitis, reflux, sleep apnea, and hyperventilation. In a real analysis scenario we would detect and select the most frequent comorbidities in OHDSI Atlas, when defining our cohort. In our mock data, we have 639 cases having at least one of the comorbidities observed (2101 observations altogether).

³<http://atlas-demo.ohdsi.org/>

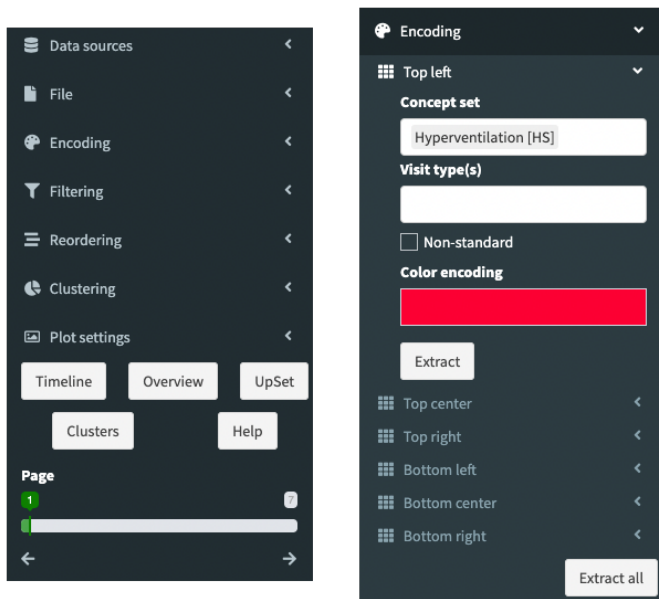


Fig. 2. The top-level of TVT menu structure on the left, and the set encoding for a glyph segment on the right: choose concept set, visit type, and color. Leaving the visit type empty allows all possible visit types. The non-standard checkbox is chosen for concept sets outside the OHDSI Atlas standard ones.

Figure 3 shows one possible glyph encoding. The fifth segment (the black one) is reserved for asthma diagnoses, the others are for comorbidities. For each glyph segment, the top line of text box gives the name of the concept set, the following line lists the included concepts, and “N” and “C” give the number of observed concepts and the number of found cases, respectively. After the concept set name there might be a list of visit types, if limited, otherwise all visit types are searched for. The “visit type” is OHDSI terminology, it can be several things like given diagnosis or drug, procedure made, device used, etc.

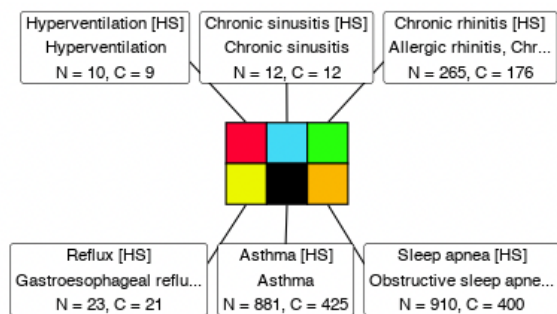


Fig. 3. Glyph legend for asthma comorbidities, ‘N’ is the number of entries extracted, and ‘C’ is the number of cases (timelines) they appeared on.

Seeing six variables on a timeline may seem like a significant limitation. Originally, our tool extracted the data from the raw health register data, and it soon became evident that defining a cohort with this approach is too complicated and limiting. The current approach uses the OHDSI Atlas tool [10] to define cohorts and concept sets. Each glyph segment is

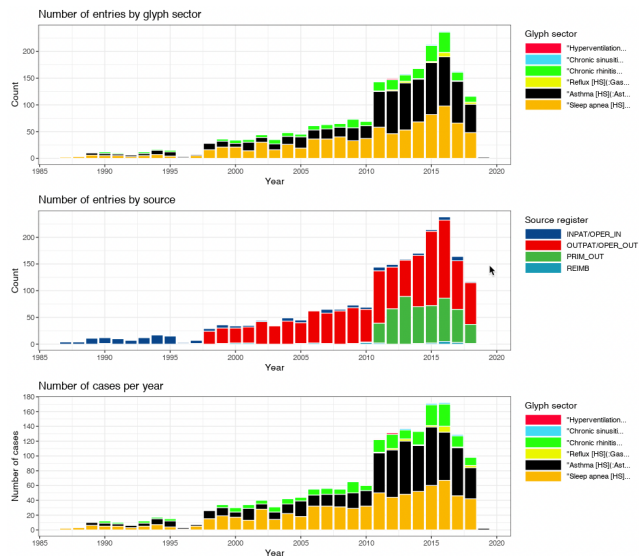


Fig. 4. Overview of data.

assigned to one OHDSI Atlas concept set. A concept set has a considerable expressiveness, and since OHDSI databases implement OMOP Common Data Model (Observational Medical Outcomes, [11]), the data can contain both standard OHDSI concepts and non-standard (usually national) vocabularies to describe the observations. Additionally, a concept-set can provide an abstraction that makes it easier to see the trend of the glyph; for example, one concept could contain all the medical codes from different eras (ICD8, 9, and 10) that refer to the same diagnosis.

Figure 9 shows the current TVT user interface and how the mock asthma comorbidity data appears in it. The *shinydashboard* [20], [21] user interface has a hierarchical menu on the left which can also be hidden. The default view in TVT is a timeline. For each person, there is a solid line extending from the cohort entry to the cohort exit, and there is a dashed line to the birth date. The line’s color encodes the gender (red = female, blue = male).

V. OVERVIEW OF GLYPH TIMELINES

The typical size of a cohort can vary from a few dozens to several thousand cases. TVT shows such cohorts as pages of timelines, displaying 500 cases per page. This is fine if you want to browse and look at individual cases, but it is hard to get an overview of the data. To gain overview, we provide both static and interactive views into data.

Figure 4 shows the data breakdown in three different ways. The plot on the top shows the number of observations per glyph sector, basically giving the counts of concept sets (showing an increasing trend of visits). The plot in the middle shows the number of observations per data source (showing that the registries were founded at different time points), and finally, the plot on the bottom shows the total number of cases per year (indicating that the number of individuals is growing over time).

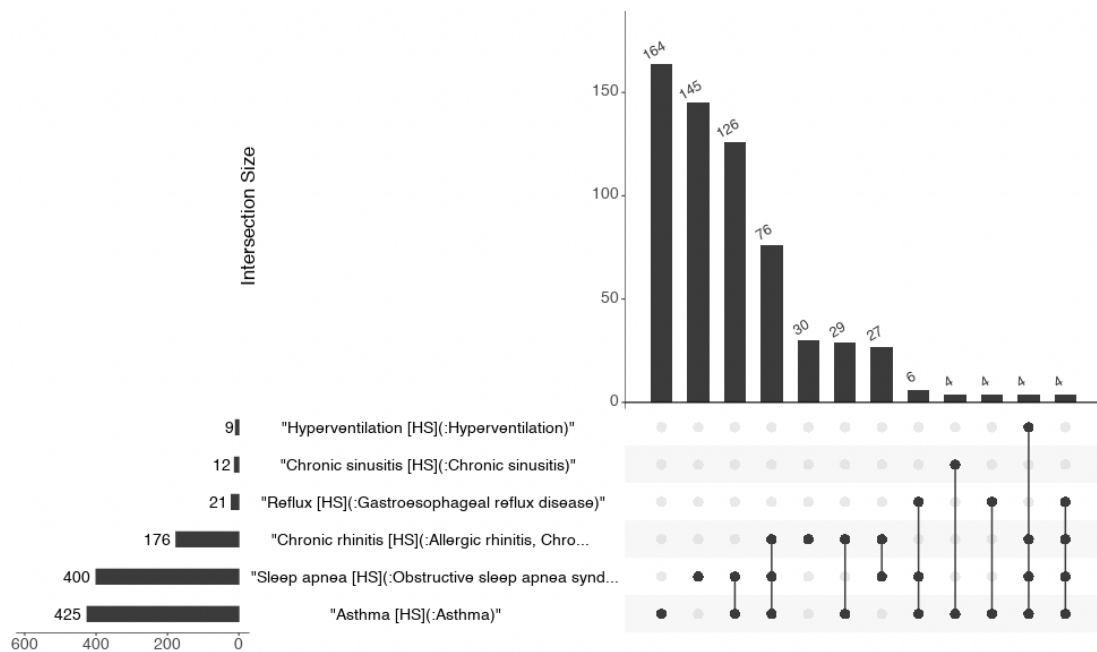


Fig. 5. Upset plot of data. The horizontal bars on the left depict the set sizes, and the vertical bars on the top indicate the size of intersection. The matrix shows the sets in intersection.

Another method to characterize the data is to show how the cases in the glyph sectors intersect (overlap of the comorbidities, Fig. 5). We use UpSet plot [22] to visualize this information. The bars on the left side of the plot show the number of cases in the concept sets of glyph segments. The bars on the top shows the sizes of concept set intersections, and the plot with connected dots indicate the intersection in that column.

Figure 5 indicates that sleep apnea is the most common asthma comorbidity in this (mock) data set, having 126 cases. On the other end, only eight cases have three of these comorbidities.

VI. INTERACTING WITH THE GLYPH TIMELINES

Being able to interact with a visualization has many benefits. If you can interactively manipulate a visualization, there is a good chance that you see something that was not obvious by just looking at it. Shneiderman’s information-seeking mantra [5] recommends “*overview first, zoom and filter, then details-on-demand.*” In addition, Bertin’s classic idea of reordering the data to produce insight-generating views is also a powerful one. He wrote that “*a graphic is never an end in itself; it is a moment in the process of decision-making*” [23, p. 16]. We have applied some of these ideas in TVT (see [19] for a comprehensive list of interaction methods in visualization).

A. Detail view – details-on-demand

Selecting a single timeline from the TVT’s main view (Fig. 9) will open up a detail view (Fig. 6), implementing details-on-demand. In this view, the concepts within the concept sets are split into separate timelines according to source, and the

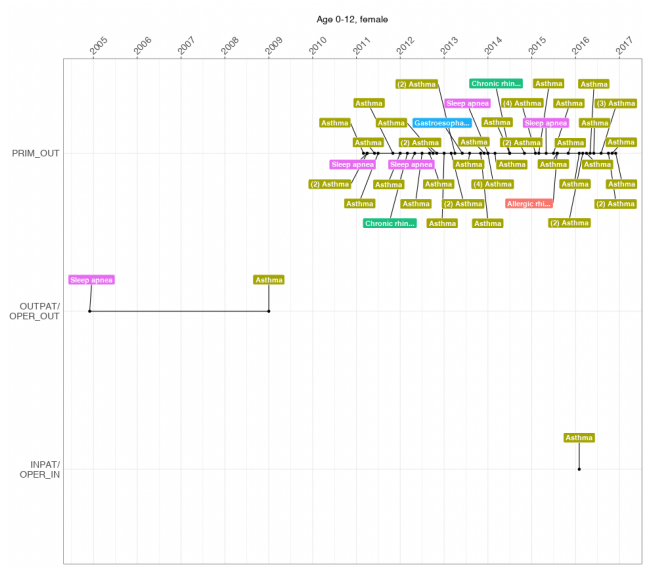


Fig. 6. Detail view of a single timeline. The concepts within concepts sets are split into separate timelines according to data source.

concept occurrences are aggregated per month. The color-coding here is just to distinguish between concept sets – it is not possible to use the color-coding of the glyph, because glyph segment can include several concepts in its concept set.

B. Filter

Figure 7(1) shows the filter menu. The current set of timelines can be filtered according to

- age at event (lower and upper limit)

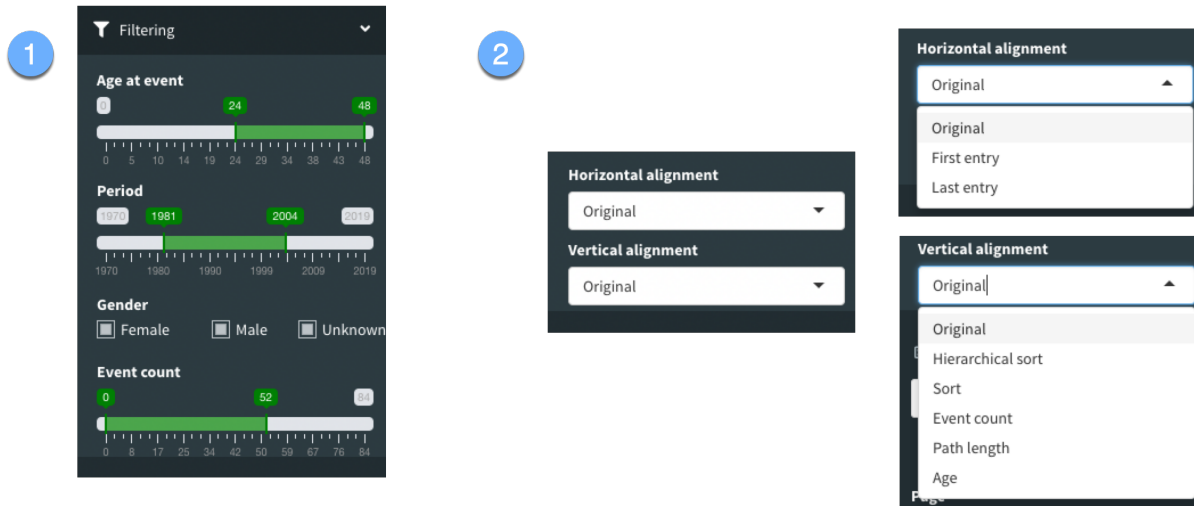


Fig. 7. Filter (1) and Reorder menus (2).

- period (beginning and end)
- gender
- event count (lower and upper limit)

Filtering the timelines also affects other views (Overview, UpSet plot), making it easy to ask what-if questions about the data.

C. Reorder

Figure 7(2) shows the reorder menu, both for horizontal and vertical alignment. Horizontally, the timelines can be made to start or end at the same time point, effectively removing the calendar time. Aligned timelines are easier to compare for patterns. For vertical alignment (or sorting) there are the following options:

- hierarchical sort (for bringing similar timelines together)
- sort (simple, fast sort according to glyph segment values)
- event count
- path length
- age

In Figure 9 the timelines are reordered according to event count, and the top timeline has the highest event count in this cohort.

D. Cluster

The hierarchical sorting of timelines will show some structure of the data but is slow when the amount of data is large, and the view is spread out on several pages. The *Pheatmap* package [24] provides an interesting alternative to get a ‘birds-eye view’ of the data.

With *Pheatmap* package, the clustering process can be divided into phases. In Figure 8(1), one row in the heatmap represents one case (or KMeans-aggregated cluster of cases, if so chosen), and the row is divided into six parts, one for each glyph segment, and time runs from left to right within each part. A cluster dendrogram for the rows is attached on the left side of the heatmap.[⁶]

The heatmap rows (cases) of the heatmap have been pre-processed by grouping similar rows together into clusters of 100 cases with KMeans clustering. This step will expedite the heatmap generation and simplify both the dendrogram and the heatmap.

Finally, in Figure 8(2), the clusters are shown as density plots, allowing one to make observations between clusters. Unfortunately, the mock data does not have interesting subsets in it, but perhaps the idea is communicated. A similar analysis with real data placed 99% of cases into a cluster having all of the comorbidities, a bit over 0.5% into a cluster that did not have sleep apnea and hyperventilation, and the rest into small cluster having only reflux as comorbidity. The parameters for the KMeans clustering and the dendrogram’s cut height are set in the user interface.

VII. DISCUSSION

Our approach to use a glyph-based timeline to represent a trajectory of health data has received positive feedback from the early users. We are currently deploying TVT into wider use within FinnGen community. At the same time, we start one-on-one sessions with volunteer users to improve the usability of TVT. We expect to get new ideas to improve the user interface and develop its functionality. So far, TVT has gained the most positive comments on helping to debug and understand OHDSI Atlas cohorts.

The majority of previous studies have focused on visualizing the individuals in great detail. We have consciously raised the level of abstraction for two reasons. Firstly, we want to obfuscate the data enough for data protection reasons, and thus make the visualization available to wider audience. Secondly, we want to focus on the patterns in the data, instead of raw counts or measures. The goal is to give an approximate answer that is precise enough to see the patterns and understand the data [25].

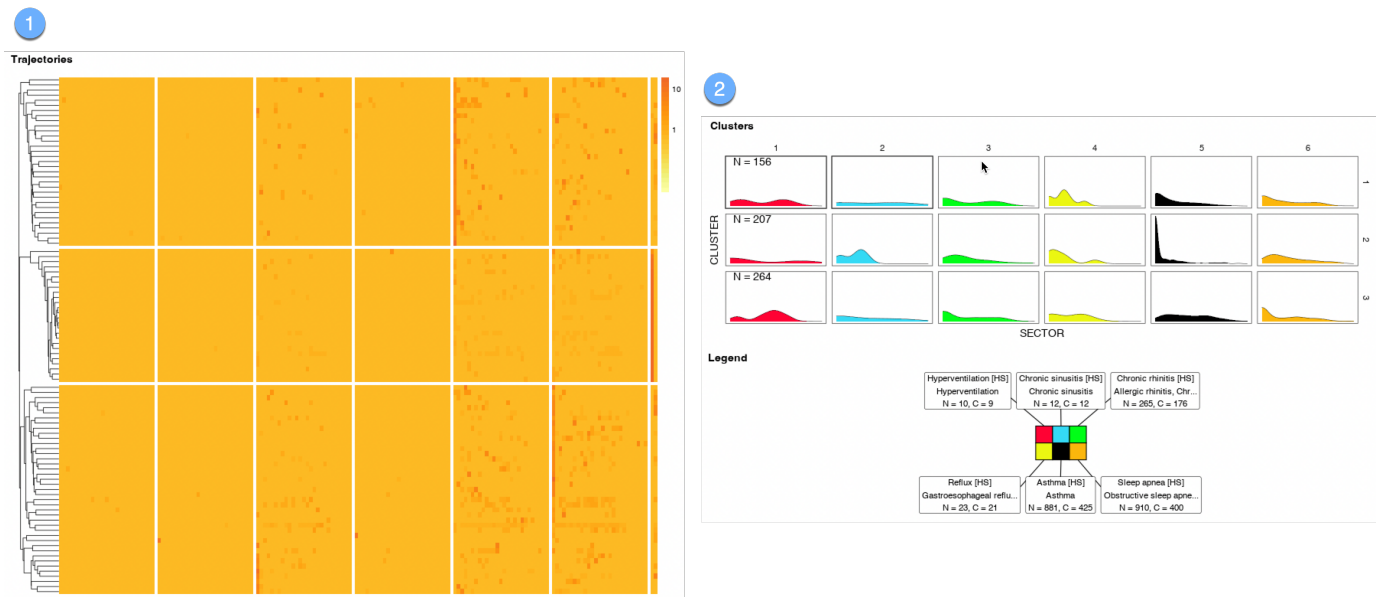


Fig. 8. (1) shows a heatmap of clustered timelines, divided per glyph segment. A line represents one case or a set of cases clustered together with KMeans pre-processing. (2) shows density plots for each glyph segment, divided into clusters.

VIII. CONCLUSION

In cohort-based studies it is essential to understand what you have in your cohort. Without considering the changes over time, the picture of the cohort might be misleading. It is essential to visualize the cohort in sufficient detail to see what's really in there. It is also important to look at the cohort at different levels: an individual, a subset, and the whole cohort level.

In this paper, we have presented an interactive visualization tool for longitudinal health data. What distinguishes our method from the prior work is the ability to visualize really large data sets, having even 40,000 cases or more.

This approach is not limited to health data, and it can be applied to any domain where gaining insight into data trajectories is valuable.

IX. ACKNOWLEDGEMENTS

We want to acknowledge the participants and investigators of FinnGen study. The FinnGen project is funded by two grants from Business Finland (HUS 4685/31/2016 and UH 4386/31/2016) and the following industry partners: AbbVie Inc., AstraZeneca UK Ltd, Biogen MA Inc., Bristol Myers Squibb (and Celgene Corporation & Celgene International II Sàrl), Genentech Inc., Merck Sharp & Dohme LCC, Pfizer Inc., GlaxoSmithKline Intellectual Property Development Ltd., Sanofi US Services Inc., Maze Therapeutics Inc., Janssen Biotech Inc, Novartis AG, and Boehringer Ingelheim International GmbH. Following biobanks are acknowledged for delivering biobank samples to FinnGen: Arctic Biobank [26], Auria Biobank [27], THL Biobank [28], Helsinki Biobank [29], Biobank Borealis of Northern Finland [30], Finnish

Clinical Biobank Tampere [31], Biobank of Eastern Finland [32], Central Finland Biobank [33], Finnish Red Cross Blood Service Biobank [34], Terveystalo Biobank [35] and The Finnish Hematology Registry and Clinical Biobank [36]. All Finnish Biobanks are members of BBMRI.fi infrastructure [37]. Finnish Biobank Cooperative (FINBB) [38] is the coordinator of BBMRI-ERIC operations in Finland. The Finnish biobank data can be accessed through the Fingenious® services [39] managed by FINBB.

TVT is a *shiny* web application [40] based on *shinydashboard* framework [20], [21]. Set of R data science packages, especially *tidyverse* collection [41], are utilized, and all the software tools are based on the open-source programming language *R* [42].

REFERENCES

- [1] F. J. Anscombe, "Graphs in statistical analysis," *The American Statistician*, vol. 27, no. 1, pp. 17–21, 1973. Available: <http://www.jstor.org/stable/2682899>
- [2] M. I. Kurki *et al.*, "FinnGen: Unique genetic insights from combining isolated population and national health register data," *medRxiv*, 2022.
- [3] C. Plaisant, B. Milash, A. Rose, S. Widoff, and B. Shneiderman, "LifeLines: Visualizing personal histories," in *Proceedings of the SIGCHI conference on human factors in computing systems*, 1996, pp. 221–227. Available: <http://www.cs.umd.edu/hcil/lifelines/>
- [4] C. Plaisant and A. Rose, "Exploring LifeLines to visualize patient records," Human-Computer Interaction Laboratory, University of Maryland, 1998. Available: <https://drum.lib.umd.edu/bitstream/handle/1903/455/CS-TR-3620.pdf>



Fig. 9. Asthma comorbidities in TVT. Each glyph timeline represents trajectory of a person

- [5] B. Shneiderman, “The eyes have it: A task by data type taxonomy for information visualizations,” in *VL’96: Proceedings of the 1996 IEEE Symposium on Visual Languages*, 1996, pp. 336–343.
- [6] J. D. Manning, B. E. Marciano, and J. J. Cimino, “Visualizing the data – using Lifelines2 to gain insights from data drawn from a clinical data repository,” *AMIA Summits on Translational Science Proceedings*, vol. 2013, p. 168, 2013.
- [7] W. Aigner, S. Miksch, B. Thurnher, and S. Biffl, “PlanningLines: Novel glyphs for representing temporal uncertainties and their evaluation,” in *Ninth international conference on information visualisation (IV’05)*, 2005, pp. 457–463.
- [8] A. Rind *et al.*, “Interactive information visualization to explore and query electronic health records,” *Foundations and Trends® in Human–Computer Interaction*, vol. 5, no. 3, pp. 207–298, 2013, Available: <https://dl.acm.org/doi/abs/10.1561/11000000039>
- [9] J. S. Yi, Y. ah Kang, J. T. Stasko, and J. A. Jacko, “Toward a deeper understanding of the role of interaction in information visualization,” *Visualization and Computer Graphics, IEEE Transactions on*, vol. 13, no. 6, pp. 1224–1231, 2007, doi: 10.1109/TVCG.2007.70515.
- [10] G. Hripcsak *et al.*, “Observational health data sciences and informatics (OHDSI): Opportunities for observational researchers,” *Studies in health technology and informatics*, vol. 216, p. 574, 2015.
- [11] E. A. Voss *et al.*, “Feasibility and utility of applications of the common data model to multiple, disparate observational health databases,” *Journal of the American Medical Informatics Association*, vol. 22, no. 3, pp. 553–564, 2015.
- [12] H. Muller, H. Maurer, R. Reihs, S. Sauer, and K. Zatloukal, “Adaptive visual symbols for personal health records,” in *2011 15th international conference on information visualisation*, 2011, pp. 220–225.
- [13] H. Müller, R. Reihs, K. Zatloukal, and A. Holzinger, “Analysis of biomedical data with multilevel glyphs,” *BMC bioinformatics*, vol. 15, no. 6, pp. 1–12, 2014.

- [14] C.-F. Wang, J. Li, K.-L. Ma, C.-W. Huang, and Y.-C. Li, "A visual analysis approach to cohort study of electronic patient records," in *2014 IEEE international conference on bioinformatics and biomedicine (BIBM)*, 2014, pp. 521–528.
- [15] M. H. P. R. Sankey, "Introductory note on the thermal efficiency of steam-engines," *The Engineer*, vol. 86, pp. 236–237, 1898.
- [16] T. Siggaard *et al.*, "Disease trajectory browser for exploring temporal, population-wide disease progression patterns in 7.2 million danish patients," *Nature communications*, vol. 11, no. 1, pp. 1–10, 2020.
- [17] World Health Organization *et al.*, "The protection of personal data in health information systems – principles and processes for public health," World Health Organization. Regional Office for Europe, 2021.
- [18] R. Borgo *et al.*, "Glyph-based visualization: Foundations, design guidelines, techniques and applications," in *Eurographics (state of the art reports)*, 2013, pp. 39–63.
- [19] R. Beasley, A. Semprini, and E. A. Mitchell, "Risk factors for asthma: Is prevention possible?" *The Lancet*, vol. 386, no. 9998, pp. 1075–1085, 2015.
- [20] W. Chang *et al.*, *Shiny: Web application framework for R*. 2021. Available: <https://CRAN.R-project.org/package=shiny>
- [21] W. Chang and B. Borges Ribeiro, *Shinydashboard: Create dashboards with 'shiny'*. 2021. Available: <https://CRAN.R-project.org/package=shinydashboard>
- [22] J. R. Conway, A. Lex, and N. Gehlenborg, "UpSetR: An R package for the visualization of intersecting sets and their properties," *Bioinformatics*, vol. 33, no. 18, pp. 2938–2940, 2017.
- [23] J. Bertin, *Graphics and Graphic Information-Processing*. Berlin: Walter de Gruyter, 1981.
- [24] R. Kolde, *Pheatmap: Pretty heatmaps*. 2019. Available: <https://CRAN.R-project.org/package=pheatmap>
- [25] J. W. Tukey, "The future of data analysis," *The Annals of Mathematical Statistics*, vol. 33, no. 1, pp. 1–67, 1962.
- [26] Arctic Biobank, Available: <https://www oulu.fi/en/university/faculties-and-units/faculty-medicine/northern-finland-birth-cohorts-and-arctic-biobank/information-arctic-biobank>
- [27] Auria Biobank, Available: <https://www.auria.fi/biopankki>
- [28] THL Biobank, Available: <https://www.thl.fi/biobank>
- [29] Helsinki Biobank, Available: <http://www.helsinginbiopankki.fi>
- [30] Biobank Borealis of Northern Finland, Available: <https://www.ppsHP.fi/Tutkimus-ja-opetus/Biopankki/Pages/Biobank-Borealis-briefly-in-English.aspx>
- [31] Finnish Clinical Biobank Tampere, Available: https://www.tays.fi/en-US/Research_and_development/Finnish_Clinical_Biobank_Tampere
- [32] Biobank of Eastern Finland, Available: <https://www.ita-suomenbiopankki.fi/en>
- [33] Central Finland Biobank, Available: <https://www.kssHP.fi/fi-FI/Potilaille/Biopankki>
- [34] Finnish Red Cross Blood Service Biobank, Available: <https://www.veripalvelu.fi/verenluovutus/biopankkitoiminta>
- [35] Terveystalo Biobank, Available: <https://www.terveystalo.com/fi/Yritystietoa/Terveystalo-Biopankki/Biopankki>
- [36] The Finnish Hematology Registry and Clinical Biobank, Available: <https://www.fhrb.fi/>
- [37] BBMRI.fi, Available: <https://www.bbmri.fi>
- [38] Finnish Biobank Cooperative, Available: <https://finbb.fi/>
- [39] Fingenious, Available: <https://site.fingenious.fi/en/>
- [40] W. Chang, J. Cheng, J. Allaire, Y. Xie, J. McPherson, *et al.*, "Shiny: Web application framework for R," *R package version*, vol. 1, no. 5, p. 2017, 2017.
- [41] H. Wickham *et al.*, "Welcome to the tidyverse," *Journal of Open Source Software*, vol. 4, no. 43, p. 1686, 2019, doi: 10.21105/joss.01686.
- [42] R Core Team, *R: A language and environment for statistical computing*. Vienna, Austria: R Foundation for Statistical Computing, 2022. Available: <http://www.R-project.org/>